Modelling the orbital histories of satellites of Milky Way-mass galaxies: testing static host potentials against cosmological simulations

Isaiah B. Santistevan [1]⋆†, Andrew Wetzel [1], Erik Tollerud,[2] Robyn E. Sanderson [3,4], Jorge Moreno[4,5] and Ekta Patel[6]

[1]*Department of Physics & Astronomy, University of California, Davis, CA 95616, USA*
[2]*Space Telescope Science Institute, 3700 San Martin Dr, Baltimore, MD 21218, USA*
[3]*Department of Physics & Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*
[4]*Center for Computational Astrophysics, Flatiron Institute, New York, NY 10010, USA*
[5]*Department of Physics and Astronomy, Pomona College, Claremont, CA 91711, USA*
[6]*Department of Physics and Astronomy, University of Utah, 115 South 1400 East, Salt Lake City, Utah 84112, USA*

ABSTRACT

Understanding the evolution of satellite galaxies of the Milky Way (MW) and M31 requires modelling their orbital histories across cosmic time. Many works that model satellite orbits incorrectly assume or approximate that the host halo gravitational potential is fixed in time and is spherically symmetric or axisymmetric. We rigorously benchmark the accuracy of such models against the FIRE-2 cosmological baryonic simulations of MW/M31-mass haloes. When a typical surviving satellite fell in (3.4–9.7 Gyr ago), the host halo mass and radius were typically 26–86 per cent of their values today, respectively. Most of this mass growth of the host occurred at small distances, $r \lesssim 50$ kpc, opposite to dark matter only simulations, which experience almost no growth at small radii. We fit a near-exact axisymmetric gravitational potential to each host at $z = 0$ and backward integrate the orbits of satellites in this static potential, comparing against the true orbit histories in the simulations. Orbital energy and angular momentum are not well conserved throughout an orbital history, varying by 25 per cent from their current values already 1.6–4.7 Gyr ago. Most orbital properties are minimally biased, $\lesssim 10$ per cent, when averaged across the satellite population as a whole. However, for a single satellite, the uncertainties are large: recent orbital properties, like the most recent pericentre distance, typically are ≈ 20 per cent uncertain, while earlier events, like the minimum pericentre or the infall time, are ≈ 40–80 per cent uncertain. Furthermore, these biases and uncertainties are lower limits, given that we use near-exact host mass profiles at $z = 0$.

Key words: methods : numerical – galaxies : kinematics and dynamics – galaxies : Local Group.

1 INTRODUCTION

The satellite galaxies of the Milky Way (MW) and M31 are the most rigorously studied low-mass galaxies, given their proximity to us. The dynamics and evolution of these low-mass galaxies encode rich information about their past and the host halo environment in which they orbit. These low-mass galaxies also differ from the non-satellite galaxies within the Local Group (LG) given that their host galaxies, either the MW or M31, regulated their star formation, and they probe deep within their host potentials. Important questions about their orbital histories include: When did each satellite fall into the MW/M31 halo? How close have they orbited, and when did they experience pericentric passages? How has the mass of the MW/M31 changed over time, and how has it impacted satellite orbits? How well conserved are orbital properties such as energy or angular momentum? Given a near-perfect representation of the host potential, to what extent can we accurately recover orbit properties, such pericentre distances and times?

Satellites in the LG are the only low-mass galaxies for which we can measure their full 6D phase-space coordinates. One historically challenging phase-space component to measure is proper motion. Many studies used the *Hubble Space Telescope* (*HST*) to derive proper motion estimates for galaxies in the LG to estimate their star-formation histories, determine companionship with the Large Magellanic Cloud (LMC), and to study the planar structure of satellites around both the MW and Andromeda (M31) (e.g. van der Marel et al. 2012; Kallivayalil et al. 2013; Sohn et al. 2020; Pawlowski & Sohn 2021). Recent *HST* treasury programs (such as GO-14734, PI Kallivayalil; GO-15902, PI Weisz; GO-17174, Bennet; HSTPROMO) are obtaining proper motions for the remaining satellites; we soon will have orbital dynamics information for all known satellites in the LG.

Furthermore, the *Gaia* space telescope has been revolutionary in providing a wealth of data, such as positions, magnitudes, and proper motions, for over 1 billion sources, including globular clusters and the

satellite galaxies of the MW (Gaia Collaboration 2018). Numerous studies use *Gaia*'s kinematic data to study group infall of satellites, including satellites of the LMC (e.g. Kallivayalil et al. 2018; Fritz et al. 2018a; Patel et al. 2020). Because proper motion measurements improve with multiple observations, many studies now use *HST* and *Gaia* data in conjunction to reduce uncertainties in satellite galaxy proper motions (such as Bennet et al. 2022; del Pino et al. 2022; Warfield et al. 2023). Current observational programs such as the Satellites Around Galactic Analogs (SAGA) survey (Geha et al. 2017; Mao et al. 2021) are observing satellite galaxies around other MW-mass galaxies. JWST soon will be able to obtain proper motions of even more distant low-mass galaxies, beyond the LG (Weisz et al. 2023), and the Vera Rubin Telescope will catalogue more than 10 billion stars within the low-mass galaxies around the MW.

Using the phase-space information of satellite galaxies, in tandem with a model of the Galactic potential, many studies investigate satellite infall and orbital histories. A galaxy becomes a satellite galaxy when it first crosses the virial radius of a more massive halo, which can quench the lower-mass galaxy's star formation (e.g. Gunn & Gott 1972; van den Bosch et al. 2008; Rodriguez Wimberly et al. 2019; Samuel et al. 2022a, b). As satellites reach their closest approach to the host galaxy at pericentre, they orbit in the denser host CGM and feel strong tidal forces ram pressure (e.g. McCarthy et al. 2008; Simons et al. 2020; Martín-Navarro et al. 2021; Samuel et al. 2022a). Many studies use different models for the MW/M31 potential and numerically integrate the orbits for their satellite galaxies to derive orbit properties. However, the results depend strongly on modelling the total mass profile of the MW and M31 (e.g. Fritz et al. 2018a, b; Gaia Collaboration 2018). Another study by Fillingham et al. (2019) jointly used *Gaia* data with the Phat ELVIS (pELVIS; Kelley et al. 2019) dark matter only (DMO) simulations, which include the gravitational effects of a central galaxy, to match simulated satellites to observed satellites in 6D phase space. They then used the distribution of infall times of the matched simulated satellites to infer the infall times for 37 satellites of the MW, which ranged from ≈ 1–11 Gyr ago, similar to other simulation-focused studies (e.g. Wetzel, Deason & Garrison-Kimmel 2015; Bakels, Ludlow & Power 2021; Santistevan et al. 2023). Deriving these infall and orbit history properties is generally difficult, given that we do not know precisely how the mass distribution of the MW or M31 has changed over time.

Stellar streams arise from the disruption of satellite galaxies or globular clusters. Therefore, studying the orbits of streams or globular clusters gives insight into the possible future orbits of satellites that will eventually merge into their host galaxy (e.g. Ibata, Gilmore & Irwin 1994; Majewski, Munn & Hawley 1996; Bullock & Johnston 2005; Price-Whelan et al. 2016, 2019; Bonaca et al. 2021; Panithanpaisal et al. 2021; Ishchenko et al. 2023). Several studies used both the Dark Energy Survey (DES; Dark Energy Survey Collaboration 2016) and the Southern Stellar Stream Spectroscopic Survey (S^5; Li et al. 2019) to discover these small systems and measure their kinematics (e.g. Shipp et al. 2018, 2019; Li et al. 2021, 2022). Comparisons with cosmological simulations in Shipp et al. (2023) suggest that undetected stellar streams may exist around the MW, which the upcoming Vera Rubin Observatory could potentially discover.

Cosmological simulations of MW-mass galaxies allow us to study theoretically the orbital evolution of satellites. Many studies used DMO simulations to understand how subhaloes respond to pericentric events (Robles & Bullock 2021), how subhalo orbits respond to various MW environments (Peñarrubia, Kroupa & Boily 2002; Peñarrubia & Benson 2005; Ogiya, Taylor & Hudson 2021)

and pre-processing and group accretion (Rocha, Peter & Bullock 2012; Wetzel, Deason & Garrison-Kimmel 2015; Li et al. 2020; Bakels, Ludlow & Power 2021). However, many such studies did not account for the important effects of baryons (e.g. Brooks & Zolotov 2014; Bullock & Boylan-Kolchin 2017; Sales, Wetzel & Fattahi 2022).

Of utmost importance in deriving the satellite orbit histories is understanding the mass distribution of the MW and M31. Studies such as Bovy & Rix (2013) and Bovy et al. (2016) focused on fitting and deriving parameters for the disc, such as the scale height and length, while other studies estimated the total mass of the MW or M31 (e.g. Eadie, Springford & Harris 2017; Patel et al. 2018; Eadie & Jurić 2019; Patel & Mandel 2023). One method of defining the total virial mass of a galaxy is by summing the mass within a given radius, such as R_{200m}, the radius that encompasses 200 × the matter density of the Universe (Bryan & Norman 1998). Using constraints from globular cluster kinematics, Vasiliev (2019a) found that the virial mass of the MW is $M_{200m} = 1.2 \times 10^{12} \, M_\odot$, which is in line with the studies mentioned above and with the currently accepted virial mass in the literature of $M_{200m} = 1$–$2 \times 10^{12} \, M_\odot$ (e.g. Bland-Hawthorn & Gerhard 2016). Many studies use the orbits of satellite galaxies to constrain the mass of the MW or M31 (e.g. Evans & Wilkinson 2000; van der Marel & Guhathakurta 2008; Watkins, Evans & An 2010; Irrgang et al. 2013). Patel & Mandel (2023) suggested the mass of M31 to be more massive, $M_{200m} = 2.85$–$3.02 \times 10^{12} \, M_\odot$, from proper motions from *HST* and *Gaia* of satellite galaxies.

Many studies used numerical tools to backward integrate the orbits of satellites, stellar streams, or globular clusters, such as GALPY (Bovy 2015), AGAMA (Vasiliev 2019b), and GALA (Price-Whelan 2017). However, such orbit modelling often makes approximations by keeping the host mass profile fixed over time (e.g. Patel, Besla & Sohn 2017; Fritz et al. 2018a; Fillingham et al. 2019; Pace, Erkal & Li 2022), sometimes varying the MW centre of mass, including an LMC-mass satellite, or including dynamical friction (Weinberg 1986; Lux, Read & Lake 2010; Gómez et al. 2015; Garavito-Camargo et al. 2019; Patel et al. 2020; Garavito-Camargo et al. 2021; Correa Magnus & Vasiliev 2022; Lilleengen et al. 2023).

Lux, Read & Lake (2010) compared various orbit history properties of subhaloes in the DMO Via Lactea I simulation (Diemand, Kuhlen & Madau 2007) to the orbits of MW satellites by using proper motion measurements from the literature and integrating their orbits in fixed potentials. In another study, Arora et al. (2022) compared four models with and without time dependence to investigate the effects of different mass models on stellar streams dynamics in simulations, and found that although most models conserve stream orbit stability, the only model that conserves stability over long periods of time is the time-evolving model. D'Souza & Bell (2022) used two MW-mass host haloes from the ELVIS suite of DMO simulations to test how well orbit modelling recovers the cosmological orbits of subhaloes. Although the majority of dynamical models applied to the MW and M31 assume static host potentials, the fiducial model in D'Souza & Bell (2022) accounted for the true mass growth of each MW-mass host. They compared results from host haloes with and without LMC-mass satellites and showed that orbit modelling better recovers the more recent pericentres and apocentres when compared to the second or third-most recent. They also tested models in which they did not account for any mass growth of the MW-mass host, or the presence of an LMC-like satellite, and found varying degrees of uncertainty associated with each simple model. However, these simulations lacked baryonic physics, including the gravitational effects of a central galaxy, and various works noted the importance of modelling baryonic physics on these scales (e.g. Brooks & Zolotov

2014; El-Badry et al. 2016; Bullock & Boylan-Kolchin 2017; Sales, Wetzel & Fattahi 2022).

In Santistevan et al. (2023), we studied the orbital dynamics and histories of satellite galaxies in the FIRE-2 cosmological zoom-in simulations of MW-mass galaxies. We investigated trends between the present-day dynamical properties, such as velocity, total energy, and specific angular momentum, with the satellite infall times, present-day distance from the MW-mass host, and satellite stellar mass. We also similarly checked for trends with properties at pericentre and found that the most recent pericentre was not the smallest, contrary to the expectation that satellite orbits only shrink over time.

In this paper, we further study the infall and orbital histories of the same satellites. We model an axisymmetric mass profile for each simulated MW-mass host to within a few per cent at $z = 0$, and we backward integrate the orbits of satellites within each one. We then compare these results against the 'true' orbital histories of satellite galaxies in the simulations. Our goal is to quantify rigorously the strengths and limitations of modelling satellite orbits in a static host halo potential, a commonly used technique. Although we focus on satellite galaxies, our results are relevant for orbit models of stellar streams and globular clusters.

Key questions that we address are: (1) How much has the mass profile of a MW-mass host evolved over the orbital histories of typical satellites? (2) How well does orbit modelling in a static axisymmetric host potential recover key orbital properties in the history of a typical satellite? (3) How far back in time can one reliably model the orbital history of satellites in a static axisymmetric host potential?

2 METHODS

2.1 FIRE-2 simulations

We use the cosmological zoom-in baryonic simulations of MW-mass galaxies in both isolated and LG-like environments from the Feedback In Realistic Environments (FIRE) project[1] (Hopkins et al. 2018). We ran these simulations using the hydrodynamic plus N-body code GIZMO (Hopkins 2015), with the mesh-free finite-mass (MFM) hydrodynamics method (Hopkins 2015), and the FIRE-2 physics model that includes several radiative heating and cooling processes such as Compton scattering, Bremsstrahlung emission, photoionization and recombination, photoelectric, metal-line, molecular, fine-structure, dust-collisional, and cosmic-ray heating across temperatures $10-10^{10}$ K (Hopkins et al. 2018). The FIRE-2 physics model also includes the spatially uniform and redshift-dependent cosmic ultraviolet (UV) background from Faucher-Giguère et al. (2009), for which HI reionization occurs at $z_{\rm reion} \approx 10$. Stars form in gas that is self-gravitating, Jeans unstable, molecular (following Krumholz & Gnedin 2011), and dense ($n_{\rm H} > 1000$ cm^{-3}), and represent single stellar populations, assuming a Kroupa (2001) initial mass function. Stars then evolve along stellar population models from STARBURST99 v7.0 (Leitherer et al. 1999), inheriting masses and elemental abundances from their progenitor gas cells. Other stellar feedback processes we implement in the FIRE-2 simulations include core-collapse and white-dwarf (Type Ia) supernovae, stellar winds, and radiation pressure.

We generated cosmological zoom-in initial conditions at $z \approx 99$ within periodic cosmological boxes of comoving length 70.4–172 Mpc, which are large enough to avoid unrealistic periodic

[1]See the FIRE project web site: http://fire.northwestern.edu

Table 1. Properties at $z = 0$ of the 13 MW/M31-mass galaxies/haloes in the FIRE-2 simulations that we analyse, ordered by decreasing stellar mass. Simulations with 'm12' names are isolated galaxies from the Latte suite, while the others are from the 'ELVIS on FIRE' suite of Local Group-like pairs. Columns: host name; $M_{\rm star, 90}$ is the host's stellar mass within $R_{\rm star, 90}$, the disc radius enclosing 90 per cent of the stellar mass within 20 kpc; $M_{\rm 200m}$ is the halo total mass; $R_{\rm 200m}$ is the halo radius; and $N_{\rm satellite}$ is the number of satellite galaxies at $z = 0$ with $M_{\rm star} > 3 \times 10^4$ M$_\odot$ that ever orbited within $R_{\rm 200m}$, totalling 493 across the suite.

Name	$M_{\rm star, 90}$ $(10^{10}\,{\rm M}_\odot)$	$M_{\rm 200m}$ $(10^{12}\,{\rm M}_\odot)$	$R_{\rm 200m}$ (kpc)	$N_{\rm satellite}$	Ref
m12m	10.0	1.6	371	47	A
Romulus	8.0	2.1	406	57	B
m12b	7.3	1.4	358	32	C
m12f	6.9	1.7	380	44	D
Thelma	6.3	1.4	358	34	C
Romeo	5.9	1.3	341	36	C
m12i	5.5	1.2	336	27	E
m12c	5.1	1.4	351	41	C
m12w	4.8	1.1	319	39	F
Remus	4.0	1.2	339	36	B
Juliet	3.3	1.1	321	40	C
Louise	2.3	1.2	333	34	C
m12z	1.8	0.9	307	26	C
Average	5.5	1.4	348	38	-

Note. Simulation introduced in: A: Hopkins et al. (2018), B: Garrison-Kimmel et al. (2019b), C: Garrison-Kimmel et al. (2019a), D: Garrison-Kimmel et al. (2017), E: Wetzel et al. (2016), F: Samuel et al. (2020).

gravity effects on individual MW-mass hosts, using the code MUSIC (Hahn & Abel 2011). We saved 600 snapshots for each simulation with time spacing of ≈ 25 Myr down to $z = 0$, assuming a flat ΛCDM cosmology with the following cosmological parameters consistent with Planck Collaboration (2020): $h = 0.68-0.71$, $\sigma_8 = 0.801-0.82$, $n_{\rm s} = 0.961-0.97$, $\Omega_\Lambda = 0.69-0.734$, $\Omega_{\rm m} = 0.266-0.31$, and $\Omega_{\rm b} = 0.0449-0.048$.

We analyse a similar set of galaxies as Santistevan et al. (2023), only we omit 'm12r', because of its low stellar mass compared to the MW and because we are not able to fit its mass profile to sufficiently high precision (see Appendix A). We also include 'm12z', first introduced in Garrison-Kimmel et al. (2019a). Our sample is from both the Latte suite of isolated MW/M31-mass galaxies, introduced in Wetzel et al. (2016), and the 'ELVIS on FIRE' suite of LG-like MW + M31 pairs, introduced in Garrison-Kimmel et al. (2019a). Table 1 lists several of their properties at $z = 0$, such as stellar mass, $M_{\rm star, 90}$, halo mass, $M_{\rm 200m}$, and radius, $R_{\rm 200m}$, and the number of satellite galaxies at $z = 0$ with $M_{\rm star} > 3 \times 10^4$ M$_\odot$, $N_{\rm satellite}$.

The Latte suite of isolated MW/M31-mass galaxies includes haloes at $z = 0$ with $M_{\rm 200m} = 1-2 \times 10^{12}$ M$_\odot$ with no other haloes of similar mass within $5R_{\rm 200m}$. We also chose m12w to have LMC-mass satellite analogues near $z \approx 0$, and m12z to have a smaller DM halo mass at $z = 0$ (Samuel et al. 2020). Star particles and gas cells are initialized with masses of 7100 M$_\odot$, however, because of stellar mass loss, the typical is ≈ 5000 M$_\odot$. The mass of dark-matter (DM) particles is 3.5×10^4 M$_\odot$ within the zoom-in region. The gravitational softening lengths for star and DM particles are fixed at 4 and 40 pc (Plummer equivalent), respectively, comoving at $z > 9$ and physical thereafter. The gas cells use adaptive force softening, consistent with their hydrodynamic kernel smoothing, down to 1 pc.

The selection criteria for each pair of haloes in the 'ELVIS on FIRE' suite of LG-like galaxies is based on their individual masses ($M_{\rm 200m} = 1-3 \times 10^{12}$ M$_\odot$), combined masses ($M_{\rm tot} =$

$2\text{–}5 \times 10^{12}\,\mathrm{M_\odot}$), their relative separation (600–1000 kpc) and radial velocities ($\upsilon_{\mathrm{rad}} < 0\,\mathrm{km\,s^{-1}}$) at $z = 0$. The mass resolution in the 'ELVIS on FIRE' suite is $\approx 2 \times$ better than the Latte suite, with initial masses of star particles and gas cells of $\approx 3500\text{–}4000\,\mathrm{M_\odot}$.

Our 13 simulated galaxies display broadly consistent properties as similar MW/M31-mass galaxies and exhibit comparable observational properties to the MW or M31, such as: MW/M31-like morphologies (Ma et al. 2017; El-Badry et al. 2018; Garrison-Kimmel et al. 2018; Sanderson et al. 2020) that follow stellar-to-halo mass relations (Hopkins et al. 2018), realistic stellar haloes (Bonaca et al. 2017; Sanderson et al. 2018), and dynamics of metal-poor stars from early galaxy mergers (Santistevan et al. 2021). Each galaxy also hosts a satellite galaxy population with properties comparable to the satellites within the local Universe, such as: stellar masses and internal velocity dispersions (Wetzel et al. 2016; Garrison-Kimmel et al. 2019b), radial and 3-D spatial distributions (Samuel et al. 2020, 2021), star-formation histories and quiescent fractions (Garrison-Kimmel et al. 2019b; Samuel et al. 2022b).

2.2 Halo/galaxy catalogs and merger trees

To generate the (sub)halo catalogs at each of the 600 snapshots, we use the ROCKSTAR 6D halo finder (Behroozi, Wechsler & Wu 2013a) using DM particles only, and we use CONSISTENT-TREES (Behroozi et al. 2013b) to generate merger trees. None of the (sub)haloes that we analyse have any low-resolution DM particle contamination, given the sufficiently large zoom-in volumes.

We briefly review how we implement star particle assignment in post-processing here, but we refer the reader to Samuel et al. (2020) for details. First, we select star particles within $d < 0.8\,R_{\mathrm{halo}}$, out to a maximum distance of 30 kpc, with velocities within $\upsilon < 2\,V_{\mathrm{circ,max}}$ of the (sub)halo's centre-of-mass (COM) velocity. We then keep the star particles within $d < 1.5\,R_{\mathrm{star,90}}$ of the (then) current member stellar population's COM and (sub)halo centre position, where $R_{\mathrm{star,\,90}}$ is the radius that encloses 90 per cent of the stellar mass. Then, we select the star particles with velocities within $\upsilon < 2\,\sigma_{\mathrm{vel,star}}$ of the COM velocity of the member star particles, where $\sigma_{\mathrm{vel,\,star}}$ is the velocity dispersion of the current member star particles. Finally, we iterate on both the spatial and kinematic criteria until the (sub)halo's stellar mass converges to within 1 per cent. This also guarantees that the COM of the galaxy and its (sub)halo are consistent with one another.

We use two publicly available analysis packages: HALOANALYSIS[2] (Wetzel & Garrison-Kimmel 2020a) for assigning star particles to haloes and for reading and analysing halo catalogs/trees, and GIZMOANALYSIS[3] (Wetzel & Garrison-Kimmel 2020b) for reading and analysing particles from Gizmo snapshots.

2.3 Selection of satellites

We select satellites in the same manner as Santistevan et al. (2023). To summarize, we include all satellites at $z = 0$ with $M_{\mathrm{star}} > 3 \times 10^4\,\mathrm{M_\odot}$ that ever orbited within their MW-mass halo's virial radius, R_{200m}. This stellar mass limit corresponds to roughly ≈ 6 star particles, which reasonably resolves the total stellar mass (Hopkins et al. 2018). At our selection threshold of $M_{\mathrm{star}} > 3 \times 10^4\,\mathrm{M_\odot}$, the median peak halo mass is $M_{\mathrm{halo,peak}} \approx 9 \times 10^8\,\mathrm{M_\odot}$ which corresponds to $\gtrsim 2 \times 10^4$ dark-matter particles. Thus, we resolve satellite subhaloes well, to prevent significant numerical disruption according to the

criteria in van den Bosch & Ogiya (2018); see Samuel et al. (2020) for more discussion on our satellite resolution convergence.

We include 'splashback' satellite galaxies that currently orbit outside of the MW-mass halo's R_{200m} but are gravitationally bound to it (e.g. Wetzel et al. 2014). As Table 1 shows, the number of surviving satellites at $z = 0$ per host, including the splashback population, is 26–57, and our sample totals 493 satellites.

To avoid biasing our results to the hosts with more satellites, we oversample the satellites, so that each host contributes a near equal fraction of satellites to the total, to within 5 per cent; see Santistevan et al. (2023) for details.

2.4 Calculating orbit properties

Many dynamical modelling studies implement a static gravitational potential for the host, consisting of a sum of potentials for each component of the galaxy, such as the stellar/gaseous disc, the bulge, the stellar halo, and the dark-matter halo (see for example Kallivayalil et al. 2013; Gómez et al. 2015; Patel, Besla & Sohn 2017). To numerically integrate orbits through time, these studies then often use common numerical tools, such as GALPY, to solve the equations of motion at each timestep.

In our analysis, we backward integrate the orbits of satellite galaxies in mass profiles that we fit to each MW-mass host in the FIRE-2 simulations. In short, we model the mass profiles of the hosts at $z = 0$ with a generalized form of the spherical Navarro-Frenk-White (NFW, Navarro, Frenk & White 1996) density profile using dark matter and hot gas ($T > 10^5\,\mathrm{K}$) particles within $r < 10$ kpc, and *all* particles at $r = 100\text{–}500$ kpc. We model the disc with two double-exponential disc profiles, one for the inner disc (bulge) and one for the outer disc, using star particles and cold gas ($T < 10^5\,\mathrm{K}$). The median fit across all 13 MW-mass hosts is within ≈ 5 per cent of the enclosed mass profiles in the simulations at $r > 10$ kpc out to the virial radius. Thus, *we test orbit modelling under a best-case scenario, at least for a static axisymmetric potential, with near perfect knowledge of mass profile at present day*. We note that we do not model the additional gravitational potential for any given satellite galaxy. See Appendix A for more details on our fits to each MW-mass host.

Next, we select the cylindrical positions (R, ϕ, Z) and velocities (v_{R}, v_ϕ, v_{Z}) of the satellites at the $z = 0$ snapshot and use this 6D phase-space information to initialize their orbits. We then use the galactic dynamics python package GALPY[4] (Bovy 2015) to backward integrate the satellite orbits for 13.8 Gyr within each host's static axisymmetric potential. Because the MW-mass host is the only gravitational potential we account for, we do not include any movement of the MW-mass host throughout the satellite orbit integration. This paper focuses on understanding the base uncertainties in static potential orbit modelling, thus, we do not account for dynamical friction in our model. Including dynamical friction may improve the model orbits, however this is outside of the scope of this paper.

We explore numerous properties of satellite orbits, each of which provides insight into their orbit histories. For example, pericentres occur when the satellite is at its closest approach to the MW-mass host, when the satellite feels the strongest tidal forces and is deepest in the host CGM. Some studies use the first post-infall apocentre distance, also called the turn-around radius, as an alternate definition for the radius of the host galaxy (e.g. More, Diemer & Kravtsov 2015; Diemer 2017). The orbital eccentricity of satellites describes the orbit shape, which can change over time in the simulations, but is

---

fixed in the fixed potential models. The orbital energy is invariant in a time-independent potential, and in a spherically symmetric potential, total angular momentum is invariant. Thus, comparing the evolution of these properties with the simulations informs us to what extent this holds. We calculate the following properties for the satellites in our sample.

Pericentre distance, time, velocity, and number: We define the virial properties at R_{200m}, the radius that encloses $200 \times$ the mean matter density of the Universe. We calculate pericentres in the same manner as Santistevan et al. (2023). First, we track the main progenitor of the satellites back in time through all 600 snapshots using the merger trees, and first confirm that the satellite is within the virial radius, R_{200m}, of the MW-mass host halo at a given snapshot. Then, we find local minima in its galactocentric distance within a ± 20 snapshot window, which corresponds to ≈ 1 Gyr in time. Given the ≈ 25 Myr time spacing between snapshots, we then fit a cubic spline to the distance, time, and velocity arrays in this snapshot window, and save the spline interpolated minimum distances, and the corresponding times and velocities at these pericentres. Finally, we assign the total number of pericentres to a satellite based on the number of times the above criteria are met. As mentioned in Santistevan et al. (2023), we checked our pericentre calculations using a snapshot window of ± 4, 8, 10 snapshots and saw that our fiducial window of ± 20 snapshots best eliminates 'false' pericentres, that is, cases where the pipeline finds a pericentre in either numerical noise or short-lived perturbations in the orbits. Additionally, whether we centre the distances on the satellite or MW-mass host galaxies or DM (sub)haloes does not affect our results.

Apocentre distance: We calculate apocentres similar to the way that we calculate pericentres. We first confirm that the satellite has orbited within R_{200m} of the MW-mass host halo before a given snapshot, however, we do not require it to be within R_{200m} at the snapshot of interest so that we may catch apocentres in the 'splashback' phase. We then find local maxima in the satellite's galactocentric distance within a similar window of ± 20 snapshots. Finally, we similarly fit a cubic spline to the distance and save these values.

Infall time: To calculate infall time, we simply ensure that the satellite is within R_{200m} at a given snapshot, and save the corresponding time that this first happens. In orbit modelling, we calculate infall time in two different ways depending on how we treat R_{200m}. The first method involves using the evolving R_{200m} from the simulations, and finding when the model orbit first crosses this distance, similar to how we calculate infall time for the satellites in the simulations. However, in our second method, we keep the present-day R_{200m} as a fixed quantity over time, that is, $R_{200m}(t_{lb} = 0)$, where t_{lb} is lookback time. We find the instances in which the model orbit first crosses within this fixed distance. In the case that the model orbit was always within $R_{200m}(t_{lb} = 0)$, we set the infall time equal to the beginning of the simulations, 13.8 Gyr ago.

Eccentricity: We approximate orbital eccentricity as:

$$e \approx \frac{d_{apo} - d_{peri}}{d_{apo} + d_{peri}} \qquad (1)$$

where d_{apo} and d_{peri} are the apocentre and pericentre distances, respectively. Defined this way, the eccentricity is a constant for a Keplerian orbit. However, it will vary within our model here, thus we choose adjacent pairs of apocentres and pericentres in the actual integrated orbit from the model to calculate it. We make no distinction in the ordering of pericentre/apocentre combinations, i.e. whether not to choose an apocentre only after pericentre, or pericentre only after an apocentre.

Orbital period: We approximate the orbital period by simply calculating the time between adjacent pericentres. We checked how this compares to the timing between adjacent apocentres and found consistent results.

Specific orbital energy: We take the specific orbital energy of a satellite to be the sum of the kinetic and potential energy per mass at each snapshot. The simulation snapshots store the gravitational potential at the location of each particle, which we use to compute at the location of a satellite. Following Santistevan et al. (2023), we select all star, gas, and dark matter particles within ± 5 kpc of the satellite's virial radius, and use the median potential of these particles, to minimize the satellite's self-potential.

Because we track satellites across time, we must properly normalize the potentials at each snapshot. Our sample includes 3 LG-like pairs of MW/M31-mass hosts, thus, we cannot normalize the potentials at arbitrarily large distances. Therefore, we choose to normalize potentials with:

$$U_{sat}(r, t_{lb}) = U_{sat,snap}(r, t_{lb}) - U_{host,snap}(r = 500 \, kpc, t_{lb})$$
$$- \frac{G \times M(< 500 \, kpc, t_{lb} = 0)}{500 \, kpc} + \frac{G \times M(< 500 \, kpc, t_{lb})}{500 \, kpc} \qquad (2)$$

where $U_{sat, snap}(r, t_{lb})$ is potential of a satellite at a given snapshot, $U_{host,snap}(r = 500 \, kpc, t_{lb})$ is the potential for particles within a spherical shell at $r = 500 \pm 5$ kpc around the MW-mass host, and the last two terms are the analytic gravitational potentials, $G \times M(< r)/r$, for the mass enclosed within 500 kpc at present-day and any given lookback time, respectively. We choose 500 kpc for several reasons: (i) the bound mass for a given MW-mass host does not change by more than a few per cent beyond this, (ii) satellites typically orbit as far as the 'splashback' radius, which we approximate as $\approx 1.5 \, R_{200m} \approx 500$ kpc from a spherical collapse model (e.g. Fillmore & Goldreich 1984; Bertschinger 1985), and (iii) we must choose distances smaller than the separation between the two MW-mass hosts ($\gtrsim 840$ kpc).

For the analytic potentials, we get

$$\Phi = \int_{500 \, kpc}^{\infty} \frac{G \times M(< r)}{r^2} \, dr \qquad (3)$$

and because the enclosed mass does not change significantly at 500 kpc, this integral results in $\Phi \approx -G \times M(< 500 \, kpc)/500 \, kpc$. The last three terms in equation (2) ensure that the potential is properly normalized across different snapshots.

When we examine differences in total energy, we divide by the virial potential of the host halo, $U_{200m,0} \equiv G \times M(r < R_{200m}, t_{lb} = 0)/R_{200m}$, because each host has different M_{200m} and R_{200m}, so this ensures that we compare satellite evolution in a similar manner. However, the snapshots for 'm12z', 'Romulus', and 'Remus' do not have stored potential values, thus, we exclude them when we compare orbital energies.

Specific angular momentum: We calculate a satellite's specific angular momentum at each snapshot with $\ell = r \times v$, where r is the total distance between the satellite and the centre of the MW-mass host, and v is the total velocity of the satellite with respect to the centre of the MW-mass host.

Tidal acceleration: Finally, we calculate the tidal acceleration a satellite feels by taking the derivative of $a = G \times M(< r)/r^2$ with respect to r, where $M(< r)$ is the total enclosed mass of the host within a distance r. We calculate this at every snapshot and save only the maximum $|da/dr|$ that a satellite experienced after first infall, although this almost always coincides with when the satellite is at its closest approach to the MW-mass host.
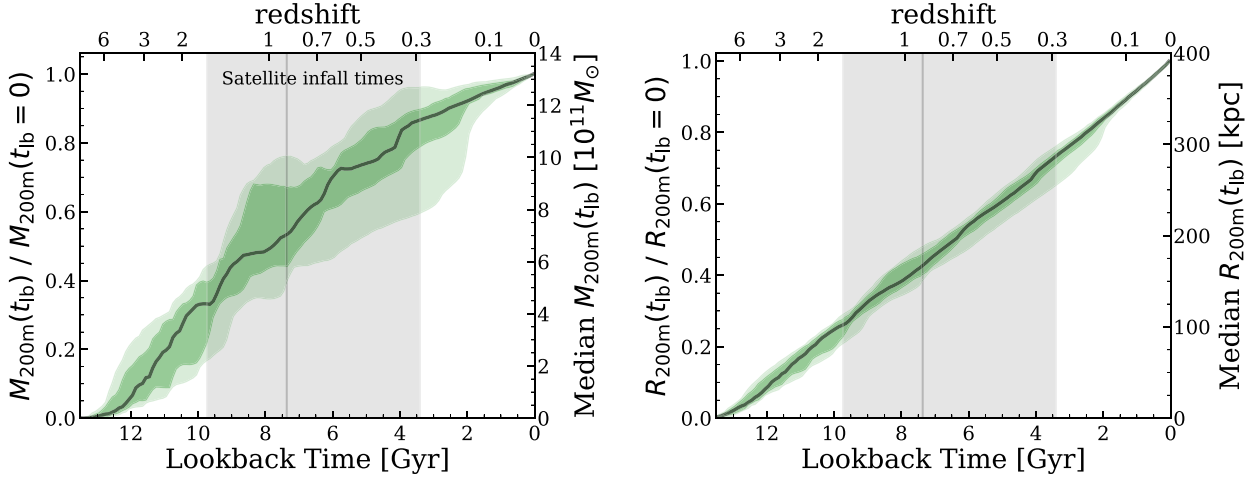
Figure 1. Left: Total (baryonic plus dark matter) mass of an MW/M31-mass halo, M_{200m}, enclosed within R_{200m} as a function of lookback time, t_{lb}. The left axis shows M_{200m} at t_{lb} relative to M_{200m} today, while the right axis converts this to the median M_{200m} across our sample. The black line shows the median, and the dark and light shaded regions show the 68th and 95th percentiles, respectively, across our 13 hosts. The vertical grey line and shaded region show the median and 68th percentile range of the lookback times of infall for surviving satellites. The typical host halo had 54 per cent of its final M_{200m} when a typical satellite fell in ≈ 7.4 Gyr ago, so M_{200m} nearly has doubled over that time. Right: Same, for the growth of the MW/M31-mass halo radius, R_{200m}. R_{200m} shows nearly linear growth with time over the last ≈ 12 Gyr. The typical host halo had 43 per cent of its final R_{200m} when a typical satellite fell in ≈ 7.4 Gyr ago, again highlighting significant change over the orbital history of a typical satellite.

2.5 Disc orientation

We use the 6D phase-space coordinates of satellites at $z = 0$ relative to the MW-mass disc, but the disc precesses over time. In Santistevan et al. (2021), we showed that after the disc stabilizes ≈ 5–11.5 Gyr ago (when the angular momentum vector of the disc stopped rapidly fluctuating in its orientation), the disc continued to precess between 5–130° until $z = 0$. Satellites reach their closest approach to the MW-mass disc when they are at pericentre, so pericentre properties are likely most sensitive to the host disc configuration. Thus, we explore different disc orientations and models to investigate how they affect the resultant satellite orbits in Appendix B. To summarize, in one model we rotate the disc by 90° while keeping the same coordinates for satellites at $z = 0$. In another model, we use a point mass model for the disc. In both models, the median differences between all orbit properties we explore between our fiducial model and these different configurations is small, less than 1 per cent. Therefore the details of the geometric configuration of the galactic potential do not matter much to satellite galaxy orbits. For more discussion about this see Appendix B and Table B1.

3 RESULTS

Many studies integrate satellite orbits using a model of a static axisymmetric MW/M31 potential at $z = 0$, which is unphysical given that the MW evolves over time. Therefore, to provide context for our results on satellite orbits, we first quantify the mass evolution of MW-mass hosts in our simulations, over both long and short timescales, including how this depends on distance. We then explore the extent to which satellites conserve energy and angular momentum. Finally, we explicitly compare results between the simulations and the idealized axisymmetric model. Because many of these distributions are non-Gaussian, throughout we present the median trends across the sample of host galaxies or satellites, as well as the half-width of the 68th or 95th percentile range, which for brevity we refer to as the 1σ and 2σ scatter, respectively.

3.1 Growth of the Milky Way-mass host

3.1.1 Halo virial properties

We define virial properties at R_{200m}, the radius that encloses 200 × the mean matter density of the universe. Fig. 1 shows the evolution of both the total (baryonic plus dark matter) virial mass of the hosts, M_{200m}, and their virial radii, R_{200m}. We show the median trend along with the 68th and 95th percentiles. The left-hand axes show the virial properties normalized to the present day ($t_{lb} = 0$), while the right-hand axes show these in physical units, scaling to the median. The grey line and shaded region also show the median and 68th percentile of the infall times, $t_{infall,MW}^{lb}$, of the satellites in our sample.

The median MW-mass host M_{200m} grew more quickly at earlier times, slowing around 4 Gyr ago. As Santistevan et al. (2020) showed, the fractional stellar mass growth of these MW-mass hosts is broadly consistent with studies based on abundance matching and dark-matter-only simulations (e.g. Behroozi et al. 2013b; Hill et al. 2017). Most satellites fell in 3.4–9.7 Gyr ago, when the median MW-mass host had ≈ 33–86 per cent of its mass at $z = 0$.

Fig. 1 (right) shows the growth of R_{200m} is nearly linear in time, with relatively small fractional scatter. When the typical surviving satellites fell in, the median MW-mass host had 26–73 per cent of its $R_{200m}(z = 0)$. Thus, the MW-mass hosts grew considerably in mass and radius, which affects the orbits of satellites.

3.1.2 Mass within fixed physical radii

Fig. 1 showed the enclosed mass within $R_{200m}(t)$. However, for satellites that already fell into the MW-mass halo, the additional growth $\gtrsim R_{200m}$, may not matter much, given that the orbits depend primarily on just the mass within an orbit. Therefore, Fig. 2 shows the ratio of the enclosed mass within a given fixed physical radius, r, at a given lookback time, t_{lb}, relative to today, $M(< r, t_{lb})/M(< r, t_{lb} = 0)$. We show the median ratios of enclosed mass within $r <$ 50, 100, and 150 kpc, along with the 68th and 95th percentiles for $r < 50$ kpc. Because the satellite orbits are sensitive to the enclosed
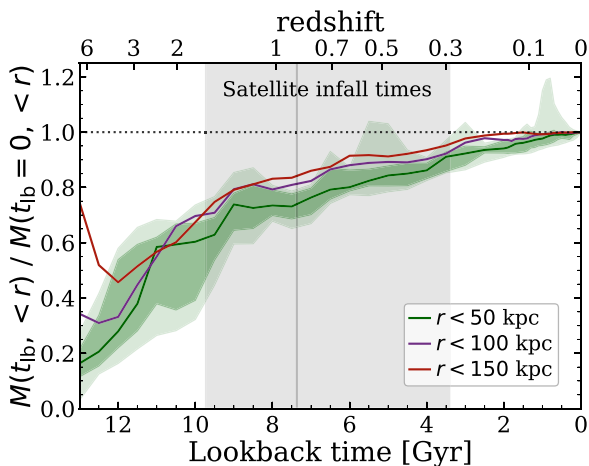
Figure 2. Total mass of the MW/M31-mass host within a fixed physical distance, r, normalized to the value today, versus lookback time, t_{lb}. We show the median mass ratio within $r = 50$, 100, and 150 kpc, and the 68th and 95th percentiles for $r < 50$ kpc. This 'physical' mass growth is less significant than that of M_{200m} (based on an evolving R_{200m}) in Fig. 1. However, *because of baryonic gas cooling, the fractional mass growth of the host is larger at smaller distance.* The mass growth reverses at $t_{lb} \gtrsim 11$ Gyr, when a given distance experienced its initial collapse from the Hubble expansion. For context, the median pericentre distance experienced by surviving satellites is ≈ 50 kpc. A typical satellite fell in 3.4–9.7 Gyr ago (grey shaded region), when the typical enclosed mass was ≈ 61–93 per cent of its value today.

mass, we choose to measure the enclosed mass within these distances because typical pericentre distances for the satellites in our sample are ~ 50 kpc and typical apocentres are ~ 200 kpc.

Because we show enclosed mass within fixed physical radii, the increase with lookback time $\gtrsim 12$ Gyr ago represents the Hubble expansion, prior to the collapse within that radius. The enclosed mass was ≈ 62–91 per cent of its present value during the typical infall times of surviving satellites; larger than for M_{200m} in Fig. 1 during the same time range. The spikes in the 95th percentile range likely arise from massive satellites. The enclosed mass fractions within $r < 100$ kpc and $r < 150$ kpc were larger at nearly all lookback times, meaning that mass has grown fractionally less at larger radii. In other words, the significant growth of the central galaxy, largely via gas cooling/accretion, leads to more significant mass grows at smaller radii.

Fig. 3 (top) shows the evolution of the enclosed mass profile across time from $r = 5$–500 kpc, where r is the satellite distance from the MW-mass host. We first calculate the median profile over all 13 hosts at each distance and snapshot with ≈ 1 Gyr time spacing; see the colourbar for the lookback time of a given mass ratio. Then we normalize the curves for each snapshot to the median profile of the hosts at present day.

In general, at each r the enclosed mass ratio increased over time, and likewise, at each time, the enclosed mass ratio increases with distance. Similar to Fig. 2, the mass growth over time at large r was not as significant compared to small r. For example, typical recent pericentre distances for satellites in the simulations are ≈ 50–60 kpc, and compared to present day, the enclosed mass was only 74 per cent during typical satellite infall times. The median present-day distance of the satellite galaxies in the simulations is around 175 kpc, and at $t_{lb} = 7.4$ Gyr ago, the enclosed mass was ≈ 83 per cent of its mass at $z = 0$. However, near the virial radius and beyond, the enclosed mass was already 97 per cent during typical satellite infall times.
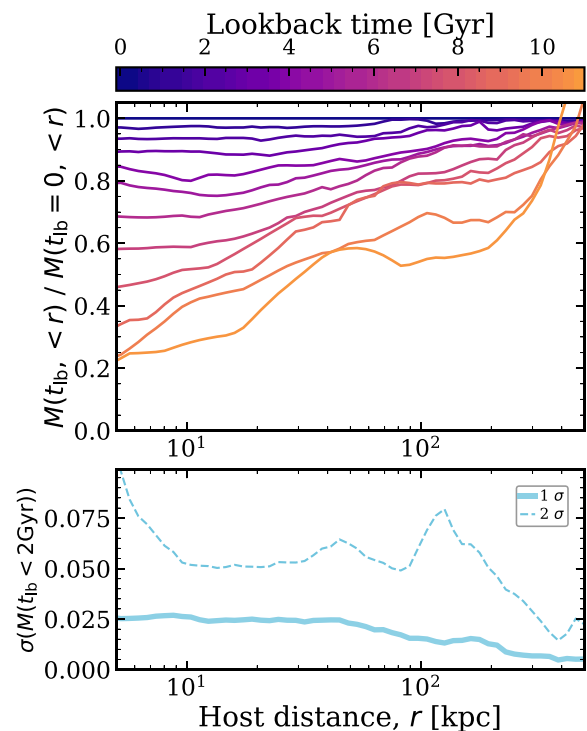

Figure 3. Top: Average long-term evolution of the mass profile (Section 3.1.2). Similar to Fig. 2, the total mass within a given physical distance, r, relative to the value today, but now as a function of r. We show the median across our 13 MW/M31-mass hosts, at various lookback times, t_{lb}, back to 11 Gyr, which encompasses the infall times of >95 per cent of surviving satellites. The enclosed mass increases over time at essentially all r, with the inner halo near the galaxy experiencing the most fractional mass growth. Thus, the approximation of a static halo mass/potential is least accurate for satellites with the smallest pericentres. For context, the median recent pericentre of surviving satellites is ≈ 50 kpc. Bottom: Typical short-term fluctuations in the host halo mass profile (Section 3.1.3). The 1σ and 2σ standard deviation of the fractional change/fluctuation of the total mass of the host over the last 2 Gyr, which is the typical of the orbital timescale of satellites at $r \approx 30$ kpc. These fluctuations in enclosed mass are weaker at larger distances. On average, the host mass growth/fluctuations over the last 2 Gyr are $\lesssim 5$ per cent, though such fluctuations would be even higher for haloes with a massive (LMC- or M33-mass) satellite.

3.1.3 Short-term evolution of host mass

The enclosed mass at any given time is subject to the perturbations of satellite galaxies that are orbiting around and merging within the main host. To examine the variability in the enclosed masses, we calculate a time-averaged enclosed mass profile over the last 2 Gyr, and we normalize the enclosed mass at every snapshot between $t_{lb} = 0$–2 Gyr to this time-averaged one. Fig. 3 (bottom) shows the 1σ and 2σ scatter of these ratios at each distance in the solid and dashed blue lines, respectively, in the bottom panel of Fig. 3.

Similar to the top panel, the 1σ scatter shows larger variability at small r, suggesting that the fractional growth in the inner regions of the MW-mass host is larger than at large r. However, the 2σ scatter does monotonically decrease with distance, but is constant between $r = 10$–100 kpc.

Haloes and their galaxies grow hierarchically over time, and each figure in this section explicitly quantifies this idea in the evolving virial region (Fig. 1), within fixed distance apertures (Fig. 2), and at fixed time (Fig. 3). Modelling the enclosed mass/potential of a host
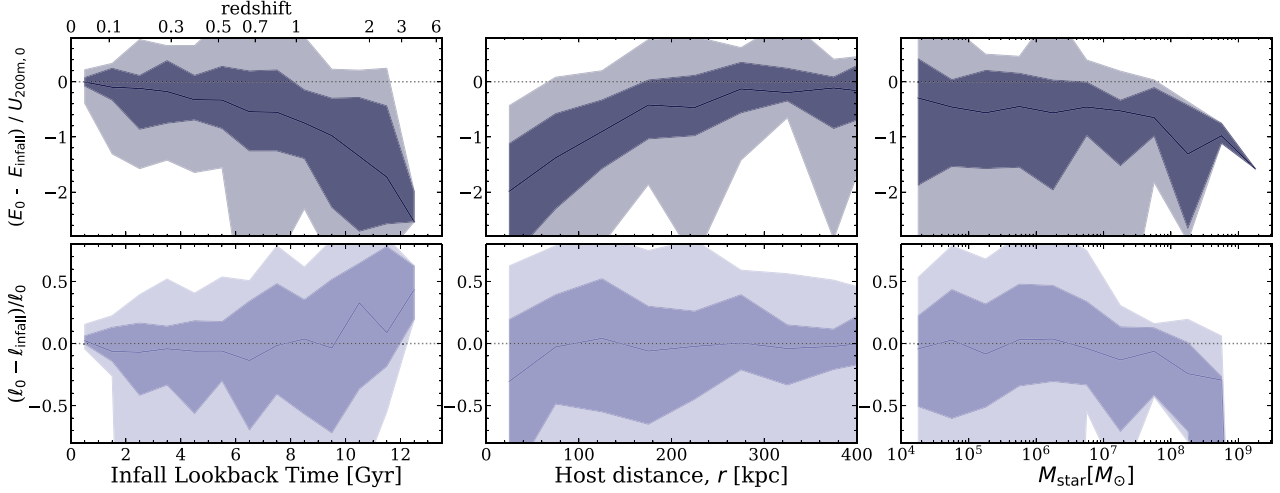
Figure 4. The difference in a satellite's specific orbital energy and specific orbital angular momentum between today and at first infall into the MW-mass halo, versus lookback time of infall (left), current distance from host, r (middle), and satellite M_{star} (right). *Orbit modelling nearly always assumes that these quantities are conserved.* Solid lines show the median for all satellites within the 13 MW-mass hosts, and the dark and light shaded regions show the 68th and 95th percentiles. Top row: Change in specific orbital total energy, E, relative to the virial energy of the MW-mass host halo today, $U_{200m,z=0} = GM_{200m}/R_{200m}$. The median decreases from 0 to -2.5 with increasing infall lookback time, meaning that satellites have lost energy and become more bound since infall, because the host has grown (Figs 1–3). Because r inversely correlates with infall time, the median change in energy increases from -2 to -0.2 with r (middle) and is constant with satellite mass at $M_{star} \lesssim 10^{7.5}\,M_\odot$, but it decreases at higher mass because of dynamical friction. Although E correlates with these properties, the widths of the 68th percentiles span ≈ 1.2–1.5, highlighting the significant variation (and thus uncertainty) for any given satellite. Bottom row: Change in specific orbital angular momentum, ℓ, relative to the value today. The median fractional difference is generally zero for satellites that fell in $\lesssim 9.5$ Gyr ago, but earlier infalling satellites have increased by up to ~ 45 per cent. We find little to no change in the median ℓ with r or M_{star}, except for satellites with $M_{star} \gtrsim 10^7\,M_\odot$, which experienced stronger dynamical friction. The mean widths of the 68th percentiles in the fractional change in ℓ span 70 per cent versus infall time and r, and 100 per cent versus M_{star}. *These uncertainties in both E and ℓ imply that we cannot infer the initial energy or angular momentum of a given satellite's orbit at its time of infall to better than $\gtrsim 2$..*

at $z = 0$ and holding that potential fixed across many Gyr does not account for the real, substantial growth of the host halo environment in which satellites orbit.

3.2 Orbital energy and angular momentum

In a time-independent potential, the specific energy of a satellite's orbit is conserved. Likewise, in a spherically symmetric potential, the specific angular momentum, $\ell = r \times v$, is conserved, while the component of ℓ along the minor axis is conserved in an axisymmetric potential. In Santistevan et al. (2023), we showed that satellites that fell in $\lesssim 10$ Gyr ago conserve their median ℓ across the population, but importantly with large scatter of ≈ 40 per cent. However, we did not examine trends as a function of lookback time across an orbit. We next examine how well orbital energy and angular momentum of a satellite's orbit are conserved. We stress that we show trends across the full population of satellites, including the full range of values of satellites with a particular infall time, distance, and M_{star}, which gives a sense of conservation on a satellite-by-satellite basis.

Fig. 4 (top row) shows the difference in the total orbital energy between present-day and infall into the MW-mass halo. To scale to the characteristic energy of a given halo, we divide these differences by $U_{200m,0} = GM_{200m}/R_{200m}$, the virial gravitational potential of the host halo today. We show these fractional energy differences as a function of the lookback time of satellite infall into MW-mass host, $t_{infall,MW}^{lb}$, distance from the MW-mass host, r, and satellite stellar mass, M_{star}.

Generally, the median fractional difference in specific energy decreases with increasing $t_{infall,MW}^{lb}$ from 0 to -2.5 (top left). Thus, satellites that fell into their MW-mass halo earlier lost more energy,

because the MW-mass host grew in mass by $\gtrsim 20$ per cent (Fig. 2). Next, the median fractional difference in total energy increases with r (top middle), from -2 to roughly -0.2 for satellites near R_{200m}. Satellites that currently orbit at smaller distances have lost more energy since infall, compared to satellites that orbit at larger distances, which only recently fell in. The fractional difference in total energy only weakly depends on M_{star} (top right). Satellites below $M_{star} \lesssim 10^{7.5}\,M_\odot$ have a median fractional energy difference of roughly -0.5, and the fractional energy difference for more massive satellites decreases to as low as -1.6, but we caution that there is only one satellite with $M_{star} > 10^9\,M_\odot$.

Next, we investigate the specific angular momentum of satellite orbits, ℓ. My studies implement a spherically symmetric component to the potential, such as an NFW profile for the DM halo (e.g. Kallivayalil et al. 2013; Patel, Besla & Sohn 2017; Besla, Peter & Garavito-Camargo 2019), in which the total angular momentum is constant. We show in Appendix B that satellite orbits are insensitive to the direction/details of the axisymmetric component of the potential, the disc, so we show results in ℓ and not the component of ℓ along the minor axis of the potential.

In Santistevan et al. (2023), we discussed trends in the angular momentum difference today versus infall, normalized by the angular momentum at satellite infall; $(\ell_0 - \ell_{infall})/\ell_{infall}$. Fig. 4 (bottom row) shows the same difference, only now normalized by the angular momentum at present-day, that is, $(\ell_0 - \ell_{infall})/\ell_0$. Similar to Santistevan et al. (2023), we find weak dependence in the median fractional change in ℓ across the population since infall. The median is as large as ≈ 45 per cent compared to the values at infall. These satellites make up roughly 20 per cent of the total sample. The fractional difference in ℓ shows virtually no dependence with r or

$M_{\rm star}$. The average 1σ scatter is ≈40 and 50 per cent versus r and $M_{\rm star}$.

Our results suggest that satellite *populations* do not show overall conservation of energy or angular momentum. If we focus on the 68th and 95th percentile ranges in each panel, we see cases in which the energies or specific angular momenta for some satellites at present-day are similar to their values at infall, for a large range of infall time, r, and $M_{\rm star}$. However, this is not true for *all* satellites, and the uncertainties in E and ℓ suggest that *one cannot determine the orbital energy or angular momentum of a given satellite at infall to within a factor of $\gtrsim 2$ from present-day measurements*. Rather, satellites commonly lose some orbital energy since infall, likely because of the growth of the MW-mass host potential over time, the effects of dynamical friction on high-mass satellites, and also satellite-satellite interactions that may torque the satellite orbits.

We also investigate the extent to which the kinetic and potential energy components are conserved with respect to infall time, r, and $M_{\rm star}$. Versus infall time, the fractional change in the kinetic energy of satellites that fell in recently is positive and the fractional change in the potential energy was negative, suggesting these satellites are likely on their first infall and nearing their first pericentre. Satellites with larger infall times often had negative fractional changes in kinetic and potential energies, because of both dynamical friction slowing the satellites and the growing host potential over time. Within $r < 100\,{\rm kpc}$, the circular velocity profile rises significantly, and satellites that orbit today have much larger kinetic energies compared to infall. We note no strong trends with $M_{\rm star}$.

Having quantified the change in orbital energy and angular momentum from first infall to today, Fig. 5 quantifies the extent to which a satellite conserves E and ℓ as a function of lookback time. We show the median trend across the population of satellites in the dashed lines. For simplicity, we present the 1σ scatter across the entire population, at a given snapshot in the solid lines. We include only galaxies that were still satellites at a given lookback time, including splashback satellites, so the size of the sample monotonically decreases with lookback time.

Fig. 5 (top) shows the difference in the satellite orbital energy between a given lookback time and today, $E(t) - E_0$, normalized by the MW-mass halo potential today, $U_{200{\rm m}, 0}$. Over the last $\approx 3.5\,{\rm Gyr}$, the median total energy is relatively unchanged, but at earlier times, this fractional difference increases with increasing lookback time from 0 to as large as ≈2 at 11.75 Gyr ago. The fractional change in E reaches 25 per cent at $\approx 4.8\,{\rm Gyr}$ ago, 50 per cent at 7.9 Gyr ago, and 100 per cent at 9.2 Gyr ago.

Although the median fractional energy change over the last 3.5 Gyr is small, this is only a statement about the population, and it does not imply energy conservation for a typical satellite, given the large scatter. The 1σ scatter reaches 25 per cent already at $\approx 1.6\,{\rm Gyr}$ ago, 50 per cent at 6.1 Gyr ago, and 100 per cent 9.1 Gyr ago.

Fig. 5 (bottom) shows the fractional change in the specific angular momentum of satellites, ℓ. Similar to energy, we compute the difference of ℓ at each lookback time with the value today, but now we normalize this difference by ℓ today. The median ℓ is constant for longer, over the last $\approx 6\,{\rm Gyr}$, before which it decreases. This implies that early-infalling satellites gained angular momentum on average, as we showed in Santistevan et al. (2023). The median fractional difference reaches 25 per cent at a much later time of $\approx 11.9\,{\rm Gyr}$ ago, and 50 per cent at 11.4 Gyr ago. Compared to the fractional change in E, the 1σ scatter reaches a given fraction later, 25 per cent at $\approx 4.2\,{\rm Gyr}$ ago, and 50 per cent at 9.3 Gyr ago. We stress again that, although the population median is conserved longer, this does not imply that a given satellite's ℓ is conserved for this long.
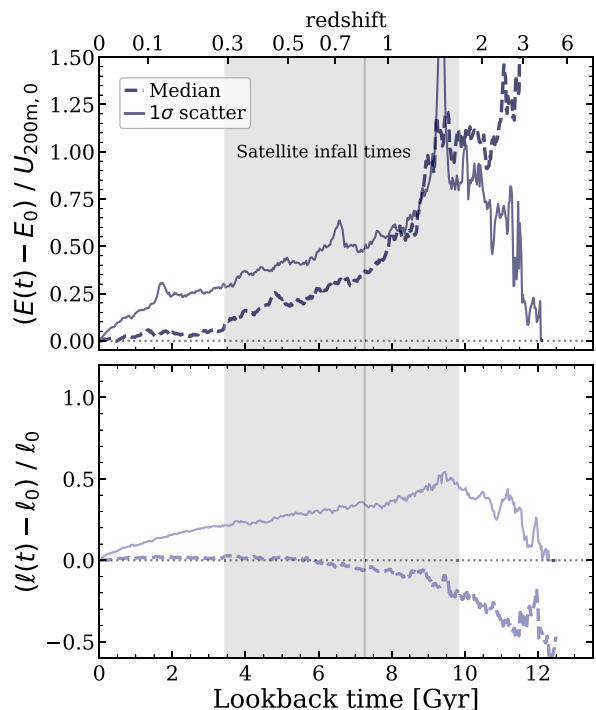

Figure 5. Fractional change in specific energy, E, and specific angular momentum, ℓ, of satellite orbits versus lookback time. Solid line shows the median across all satellites and dashed line shows the 1σ scatter. We only include galaxies that are satellites at a given lookback time, including splashback satellites. The vertical line and shaded region represent the median satellite infall time and 68th percentile range of values, respectively. Top: Specific energy. The median fractional change increases with lookback time, that is, they were less bound at infall than they are today, because the host has grown (Figs 1–3). While the median, which represents the systematic bias across the population, reaches 25 per cent at 4.8 Gyr ago, the 1σ scatter, which represents the uncertainty for a given satellite, reaches 25 per cent already at 1.6 Gyr ago. Bottom row: Specific angular momentum. The median is about zero over the last $\approx 6\,{\rm Gyr}$, but at earlier times the angular momentum decreases, down to -60 per cent, meaning that the specific angular momentum systematically has increased since infall. Again, the 1σ scatter reaches 25 per cent already 4.7 Gyr ago. Thus, while inferences of orbital energy and angular momentum, based on their conservation, are relatively *unbiased* (for the overall population) over the last ≈ 5–8 Gyr, their *uncertainties* for a given satellite are large already $\approx 1\,{\rm Gyr}$ ago.

Rather, the 1σ scatter represents the typical uncertainty for a given satellite.

Figs 4 and 5 show that *neither E nor ℓ are conserved across time*, which agrees with Figs 1–3, and results from the growth and general time dependence of the host halo potential. The 1σ scatter represents the typical uncertainty for a given satellite, which is as large as ≈ 50 per cent for ℓ around 9.3 Gyr ago, and as large as a factor of 2 for E at 9.1 Gyr ago.

3.3 Orbit modelling in static axisymmetric host potential

We now compare orbit properties of satellites from our cosmological simulations to properties derived in an idealized, static, axisymmetric model. As we describe in detail in Appendix A, we fit the present-day host potential, keep it fixed over time, and initialize the satellite orbits at $z = 0$ using the same 6D phase-space coordinates as in the cosmological simulations. Thus, the orbital energy and angular
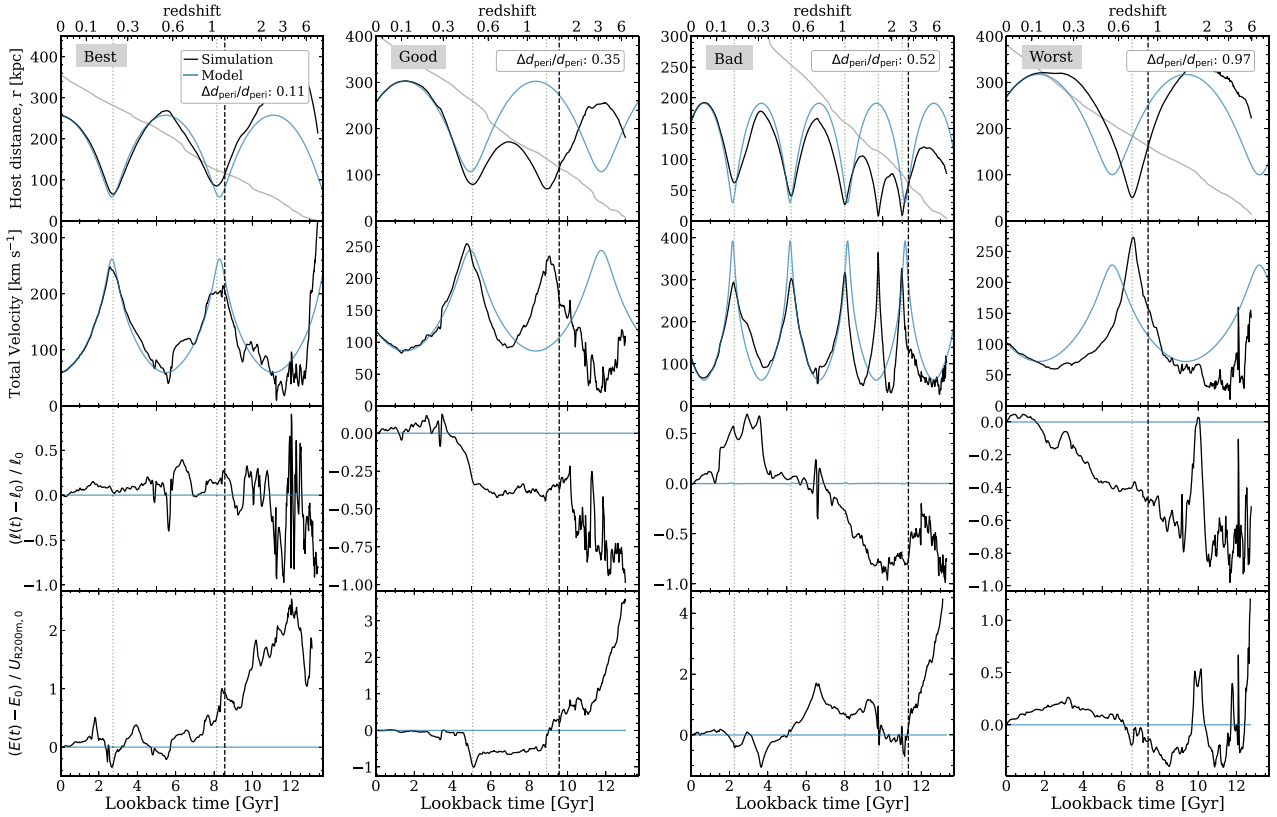
Figure 6. Four case studies of satellite orbital histories. Orbital distance from the host galaxy, r (top row), total velocity (second row), specific angular momentum (third row), and specific total energy (bottom row). We show four satellites based on how well the most recent pericentre agrees between the simulations and orbit modelling, $(d_{\mathrm{peri,\,model}} - d_{\mathrm{peri,\,sim}})/d_{\mathrm{peri,\,sim}}$, with increasing error from left to right (see the legend). Black lines show the simulations, while blue lines show model-based orbits in a static axisymmetric host potential. Black vertical dashed lines show the first infall into the MW-mass halo in the simulation, and in the top row the grey line shows $R_{\mathrm{200m}}(t)$ of the host halo. Green vertical dotted lines show pericentres in the simulation. In the left case study, the orbit model and simulation agree for nearly two full orbits, while the right case study shows agreement for less than half an orbit. Orbit modelling tends to recover the timing of the most recent pericentres better than their distances. As the bottom rows show, specific angular momentum and total energy of the orbit are not generally conserved; see also Figs 4 and 5.

momentum of satellites remain constant, and the satellites orbit periodically across 13.8 Gyr.

Fig. 6 shows four representative satellites: each column shows varying degrees of how well orbit modelling reproduces the most recent pericentre distance. To quantify how well orbit modelling does in reproducing the recent pericentre distance, we measure $\Delta d_{\mathrm{peri}}/d_{\mathrm{peri}} = (d_{\mathrm{peri,\,model}} - d_{\mathrm{peri,\,sim}})/d_{\mathrm{peri,\,sim}}$. From top to bottom, we compare the host-centric distance, the total velocity, specific angular momentum, and specific energy of the orbit.

Orbit modelling agrees well with the simulations during the satellites' recent histories. In the left two columns, orbit modelling recovers the orbits well for one half to two orbits, the third column shows agreement with the timing of the orbit for two and a half orbits but less agreement for the distance and velocity, and the right column show agreement for less than half an orbit.

Even in the left two cases, in which the model does well at reproducing the most recent pericentre distance, orbit modelling does not accurately recover previous pericentres, especially the timing of the pericentres, which continues to become more out of phase with time, likely due to the lack of dynamical friction. The right column shows cases in which the timing of the most recent pericentre is within ≈ 0.5 Gyr but the pericentre distance is off by nearly a factor of 2.

Finally, the third and bottom rows of Fig. 6 show the lack of conservation in specific angular momentum and specific energy for the satellites in the simulations. For each satellite, we show the fractional change in ℓ compared to present-day, that is, $(\ell(t) - \ell_0)/\ell_0$. Even after a satellite falls into its MW-mass halo, its angular momentum can increase or decrease $\gtrsim 50$ per cent over time. The lack of conservation in ℓ is likely a combination of complex processes, including the growth of the MW-mass host, satellite-satellite interactions, mergers, and the non-symmetric potential. In the bottom row, we calculate the fractional change in energy compared to present-day, normalized by the host virial potential energy today, that is, $(E(t) - E_0)/U_{\mathrm{R200m,\,0}}$. Similar to the results in Figs 4 and 5, the specific energy of a satellite orbit decreases over time, primarily because of the growth of the MW-mass host.

In the following subsections, we quantify differences in orbit properties across the entire satellite population. It is worth noting that dynamical friction acts more efficiently at higher masses to rob the satellites of their orbital energy and cause them to merge away (e.g. Boylan-Kolchin, Ma & Quataert 2008). We remind the reader that when interpreting the plots, we do not include a model for dynamical friction.
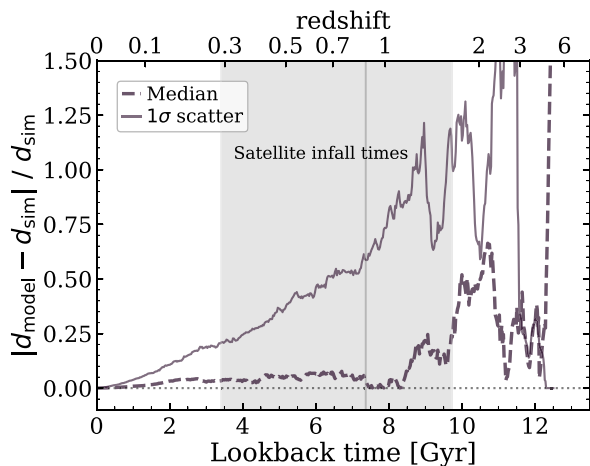
Figure 7. The fractional difference between orbit modelling and the simulations for the host-centric distance versus lookback time. The dashed line shows the absolute value of the median across all satellites and the solid line shows the 1σ scatter of the sample. We choose to show the absolute value of the fractional difference to keep all values positive for visual clarity, however, the median is negative for lookback times $\lesssim 7.5$ Gyr, and positive for earlier times. Similar to Fig. 5, we only include the instantaneous population of satellites for a given lookback time. The median reaches 25 per cent at 9.7 Gyr ago, but the 1σ scatter reaches 25 per cent already at 4 Gyr ago.

3.3.1 Orbital distance

First, Fig. 7 shows the absolute fractional difference in host-centric distance from the simulations versus from orbit modelling, versus lookback time. We show the absolute value of the median to keep all values positive and visual clarity, however, the median is negative for lookback times of $t_{lb} \lesssim 7.5$ Gyr, and positive for $t_{lb} \gtrsim 7.5$ Gyr. Over the last 8.5 Gyr, the median fractional difference is relatively constant at $\lesssim 10$ per cent. Before this, the median then increases to 25 and 50 per cent, around 9.7 and 10 Gyr ago. Prior to ≈ 11.7 Gyr ago, less than 1 per cent of these satellites today were still satellites. The 1σ scatter reaches a 25, 50, and 100 per cent fractional difference at 4, 6.3, and 8.6 Gyr ago.

3.3.2 Virial infall time

Many studies focus on when low-mass galaxies first become satellites, and their properties, such as mass, during infall (for instance Boylan-Kolchin, Besla & Hernquist 2011; Wetzel, Deason & Garrison-Kimmel 2015; Patel, Besla & Sohn 2017). We investigate two ways of calculating the time of first infall into the host halo within orbit modelling, $t_{lb, infall, model}$. First, when a satellite first crossed within the MW-mass halo, accounting for the growth of its R_{200m} over time. Second, we record when a satellite's orbit first crossed within R_{200m} at $z = 0$. Nearly 60 per cent of all satellites in orbit modelling always have orbited within $R_{200m}(t_{lb} = 0)$, so we simply define these infall times to be 13.8 Gyr ago. We take the difference of each $t_{lb, infall, model}$ to the infall time in the simulations, $t_{lb, infall, sim}$ (calculated with an evolving R_{200m}). Fig. 8 shows these differences, versus the infall times in the simulations (left) and versus satellite M_{star} (right).

In the left panel, using an evolving $R_{200m}(t_{lb})$, the median difference between orbit modelling and the simulations is generally within ≈ 2 Gyr. The peak in the curve at 1.5 Gyr is driven by 11 of the 20 satellites in this particular bin in infall time, where orbit

modelling predicts that these galaxies fell in $\gtrsim 5$ Gyr earlier than in the simulations. Similarly, a slightly smaller peak at 7.5 Gyr is caused by the model overpredicting $t_{lb, infall, model}$ for nearly half of the satellites by $\gtrsim 2$ Gyr.

The 68th percentile range is largest for the most recently infalling satellites and decreases with increasing lookback time. Orbit modelling generally overestimates the infall time compared to the simulations, because the model orbits are periodic and more likely to cross the virial radius at earlier times. The model overpredicts the infall time for roughly 65 per cent of all satellites. Even when accounting for the evolving R_{200m}, the 1σ scatter spans $\gtrsim 1$ Gyr, which highlights the large uncertainty in orbit modelling.

When using a fixed $R_{200m}(t_{lb} = 0)$, the median shows relatively good agreement for satellites that fell in within the last ≈ 4 Gyr. Beyond 4 Gyr ago, the median difference in infall time increases until ≈ 7 Gyr ago, where it decreases again. The orbits of these satellites were generally within $R_{200m}(t_{lb} = 0)$ at all times, so the difference between orbit modelling and simulations follows the relation 13.8 Gyr $- t_{lb, infall, sim}$. Of the subset of satellites that fell in between 1–2 Gyr ago, only 2 of the 17 satellites were always within R_{200m}, which is why the 68th percentile dips down close to 0. However, the associated uncertainties in this method of calculating infall time are much worse, and the 1σ scatter reaches as large as ≈ 5.7 Gyr.

Versus M_{star}, both infall metrics follow the same general trends: better agreement for satellites with $M_{star} < 10^7$ M$_\odot$ and larger offsets for higher-mass satellites. The offset between the medians, and 68th percentiles, is roughly 3–4 Gyr between the two infall metrics. The values associated with the fixed $R_{200m}(t_{lb} = 0)$ method skew to larger values, because many of the model orbits always orbited within this distance. The 1σ scatters each span roughly 2.5–3 Gyr, so one cannot accurately determine a satellite galaxy's infall time at any given mass to within ≈ 2.5 Gyr.

3.3.3 Pericentre properties

We next investigate various properties associated with pericentric passages relative to the host galaxy, when the tidal acceleration and ram pressure from the host CGM tend to be strongest.

Most works assume that satellite orbits only *shrink* over time, because of dynamical friction and the time-dependent host potential (e.g. Weinberg 1986; Taylor & Babul 2001; Amorisco 2017), which implies that the most recent pericentre should be the smallest experienced. However, as we showed in Santistevan et al. (2023), for 67 per cent of our satellites with $N_{peri} \geq 2$ the most recent pericentre is not the smallest, because many satellite orbits have grown in pericentre distances over time. Patel et al. (2020) also saw cases in which the most recent pericentre was not the smallest, and suggest that the presence of a massive satellite alone can cause this effect. Therefore, we present trends for the most recent and the minimum pericentres.

Fig. 9 compares pericentre distances, d_{peri}, the velocity at pericentre, v, the number of pericentric passages, N_{peri}, and the timing of pericentres, t_{peri}, versus the lookback time of infall into the MW-mass halo, present-day distance, r, and satellite M_{star}. For each pericentre property, we show the difference between orbit models and simulations, except for pericentre distance, for which we compare the *fractional* difference.

Fig. 9 (top row) shows the trends for pericentre distance, for both the most recent, $d_{peri, rec}$, and the minimum pericentre, $d_{peri, min}$. For a given satellite, all of its pericentre distances are the same in our
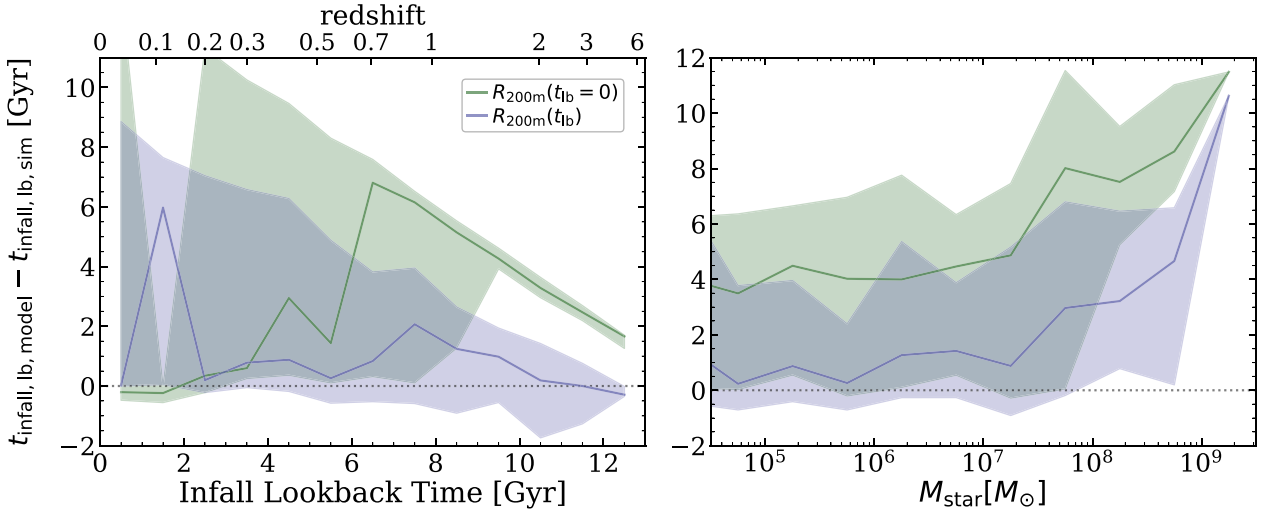
Figure 8. The difference in infall time, from orbit modelling versus the simulations, as a function of infall time in the simulations (left) and satellite $M_{\rm star}$ (right). For orbit modelling, we measure infall time two ways: using a non-evolving host halo radius, $R_{\rm 200m}(t_{\rm lb} = 0)$ (green) and using $R_{\rm 200m}(t_{\rm lb})$ from the simulations (purple). Solid lines show the median and shaded regions show the 68th percentile ranges across our satellites. Left: The median infall time is generally accurate for orbit modelling if using an accurate $R_{\rm 200m}(t_{\rm lb})$, with a median offset of $\lesssim 2\,{\rm Gyr}$ and a typical scatter of 2.4 Gyr. The spike at 1.5 Gyr comes from orbit modelling overpredicting the infall times for most satellites by $\gtrsim 5\,{\rm Gyr}$, given errors in modelling recent apocentre distances. By contrast, orbit modelling using fixed $R_{\rm 200m}(t_{\rm lb} = 0)$ works well for recently infalling satellites but vastly overestimates the lookback time to infall for earlier-infalling satellites, with an average 1σ scatter of 2.8 Gyr. Right: The median and 68th percentile for both infall time metrics show similar trends with satellite mass, offset by $\approx 3\,{\rm Gyr}$. The typical 1σ scatters range from roughly 3 Gyr and 2.5 Gyr when using either the fixed $R_{\rm 200m}(t_{\rm lb} = 0)$ or accurate $R_{\rm 200m}(t_{\rm lb})$, respectively. Orbit modelling fails most significantly for satellites with $M_{\rm star} \gtrsim 10^7\,{\rm M_\odot}$, because dynamical friction has shrunk their orbits.

(static) orbit models. With respect to satellite infall time (left panel), the model recovers the median $d_{\rm peri, rec}$ well, but this is across the entire population of satellites, not for a given satellite. The median fractional difference is within ≈ 20 per cent, and the average 1σ scatter is roughly 19 per cent, smaller than many other properties we present here. Thus, although the median pericentre distance across the population looks reasonable, the prediction for any particular satellite is uncertain by ≈ 20 per cent. Orbit modelling recovers the median $d_{\rm peri, min}$ well for satellites that fell in $\lesssim 5\,{\rm Gyr}$ ago, because all but one of these satellites experienced only one pericentre, so the minimum is the most recent. For satellites that fell in $\gtrsim 5\,{\rm Gyr}$ ago, the median fractional offset in $d_{\rm peri, min}$ diverges from that of $d_{\rm peri, rec}$. Roughly 60 per cent of satellites that fell in $\gtrsim 5\,{\rm Gyr}$ ago experienced multiple pericentres, and because the orbit models only predict a single $d_{\rm peri}$ for a given satellite, these positive values suggest that the satellites in the simulations orbit at closer distances than in our static model. The 1σ scatter for $d_{\rm peri, min}$ reaches 100 per cent around 9.5 Gyr ago, thus one cannot accurately predict the minimum pericentre to within $\gtrsim 2$ for satellites that fell in earlier than this.

In the middle panel, median $d_{\rm peri, min}$ and $d_{\rm peri, rec}$ between the orbit models and simulations show general consistency across all distances. The fractional difference in $d_{\rm peri, rec}$ is $\lesssim 5$ per cent, and $\lesssim 25$ per cent for $d_{\rm peri, min}$. As we will discuss below, ≈ 95 per cent of satellites that currently orbit beyond 300 kpc completed only one pericentre, so the median and 68th percentiles are the same between both $d_{\rm peri, min}$ and $d_{\rm peri, rec}$. Conversely, 2/3 of satellites currently within 300 kpc have $N_{\rm peri} \geq 2$ and, similar to the left panel, the median and scatter for $d_{\rm peri, min}$ increases to positive values, indicating that orbit modelling overpredicts $d_{\rm peri, min}$ for these satellites. For the satellites within 300 kpc, nearly 85 per cent fell into their MW-mass hosts over 5 Gyr ago. As with the left panel, the range in the 1σ scatter is larger for $d_{\rm peri, min}$ than in $d_{\rm peri, rec}$, with average values of 55 per cent and 24 per cent, respectively.

Finally, the median fractional difference in both pericentre metrics shows no dependence on stellar mass for $M_{\rm star} < 10^{8.75}\,{\rm M_\odot}$ (right panel). Lower-mass satellites typically fell in earlier, orbit at closer distances, and completed more pericentres, so their minimum and most recent pericentres are more likely to diverge. Only one satellite in our sample has $M_{\rm star} > 10^9\,{\rm M_\odot}$.

Fig. 9 (second row) shows trends in the total velocity of pericentre, $v_{\rm peri}$, and similar to $d_{\rm peri}$, we show trends in both the minimum, $v_{\rm peri, min}$, and most recent, $v_{\rm peri, rec}$. Again, satellites at small infall time, large r, and high $M_{\rm star}$, only experienced one pericentre, so the median trends in both $v_{\rm peri, min}$ and $v_{\rm peri, rec}$ are the same. The model recovers both the median $v_{\rm peri, min}$ and $v_{\rm peri, rec}$ to within $\approx 30\,{\rm km\,s^{-1}}$ across all infall times and r, and nearly all $M_{\rm star}$. Although the model recovers the median $d_{\rm peri, rec}$ well, it overestimates $v_{\rm peri, rec}$ in all panels presumably because of the lack of dynamical friction and gravitational perturbations from other satellites.

Fig. 9 (third row) compares the number of pericentric passages a satellite experienced, $N_{\rm peri}$. For the model orbits, we only count the number of pericentric passages a satellite experienced since first infall. We count $N_{\rm peri}$ from the two infall metrics in Section 3.3.2: since infall into the MW-mass halo accounting for an evolving $R_{\rm 200m}(t_{\rm lb})$, and since infall while keeping a fixed $R_{\rm 200m}$ $R_{\rm 200m}(t_{\rm lb} = 0)$. We show the *mean* and standard deviation for $N_{\rm peri}$, because it is an integer for a given satellite.

These results for $N_{\rm peri}$ are not particularly sensitive to the two ways in which we calculate first infall. In the left panel, the mean difference in $N_{\rm peri}$ slightly increases with infall time, because orbits in the models are periodic, so longer integration times lead to larger $N_{\rm peri}$. The middle panel shows trends versus host distance, r: for satellites at $\lesssim 250\,{\rm kpc}$, orbit modelling overpredicts $N_{\rm peri}$, because these satellites typically fell into their MW-mass hosts earlier than satellites at larger r. Finally, the mean difference in $N_{\rm peri}$ is generally flat at $M_{\rm star} < 10^{7.5}\,{\rm M_\odot}$, but the difference increases for more
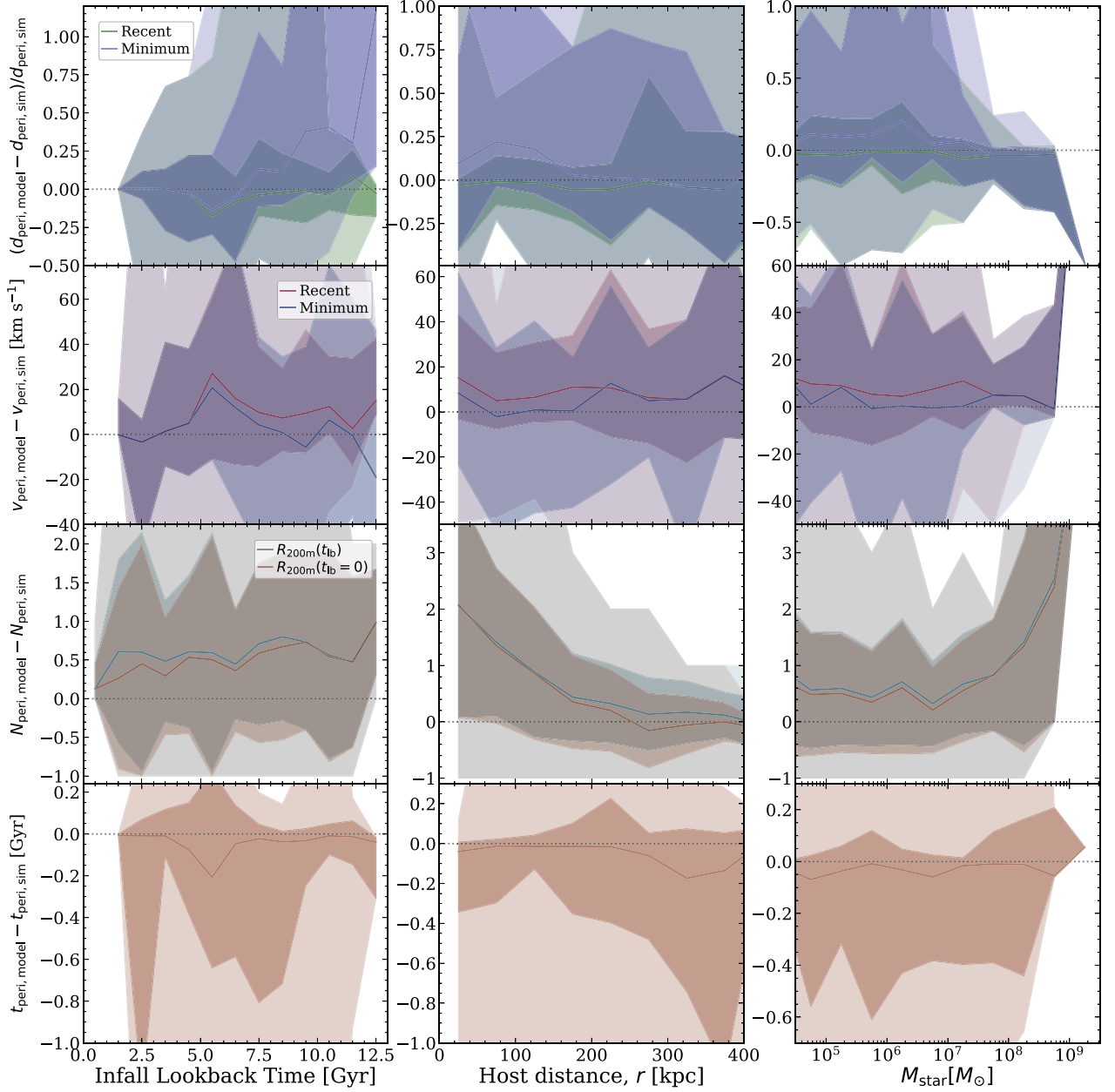
Figure 9. Comparing various properties of orbital pericentres, between the simulations and orbit modelling, for surviving satellites versus their lookback time of infall into the MW-mass halo (left), present-day distance from the MW-mass host, r (middle), and satellite $M_{\rm star}$ (right). Solid lines show the median, and the shaded regions show the 68th and 95th percentiles, across all satellites. Top row: Fractional difference between pericentre distances, $(d_{\rm peri,\,model} - d_{\rm peri,\,sim})/d_{\rm peri,\,sim}$, for both the most recent and the minimum pericentre. Orbit modelling predicts larger minimum pericentres, as high as 100 per cent in the median, for satellites that fell in $\gtrsim 12\,{\rm Gyr}$ ago (left), and within 25 per cent for satellites at small r (middle), and for satellites $\lesssim 10^7\,{\rm M}_\odot$ (right). Second row: Difference between the total velocity at pericentre, $v_{\rm peri,\,model} - v_{\rm model,\,sim}$, for both the most recent and the minimum pericentre. Orbit modelling generally overpredicts both pericentre velocities by ≈ 10–$20\,{\rm km\,s^{-1}}$. Third row: Difference between the mean number of pericentric passages about the MW-mass host, since first crossing the growing host $R_{\rm 200m}(t_{\rm lb})$ and since first crossing $R_{\rm 200m}(t_{\rm lb} = 0)$. This difference slightly increases from ≈ 0 to 1 with $t_{\rm lb}$ (left) and decreases from 2 to 0 with r (middle), because satellites at small distance typically fell in earlier, which means that they orbited longer in the model. Fourth row: Difference between the lookback time of the most recent pericentre, $t_{\rm peri,\,model} - t_{\rm peri,\,sim}$, which shows weak trends with any properties. Although the median trends across the satellite population agree well in most cases, the substantial scatter in all panels implies *significant* uncertainty for a given satellite's orbit history.

massive satellites, given the lack of dynamical friction in the orbit models.

Finally, Fig. 9 (bottom row) compares the timing of just the most recent pericentre, $t_{\rm peri}^{\rm lb}$. The median difference is $\lesssim 0.2\,{\rm Gyr}$ across all three panels. The median is also consistently negative, indicating that orbit modelling predicts more recent pericentres, likely because

in the orbit models the MW-mass host does not reduce in mass going back in time.

Typically maximized near a pericentric passage, satellites feel a tidal acceleration from the host, which strips their mass. We calculate the acceleration to be $a = GM(< r)/r^2$, where $M(< r)$ is the total enclosed mass of the host within a distance r. We then compute the
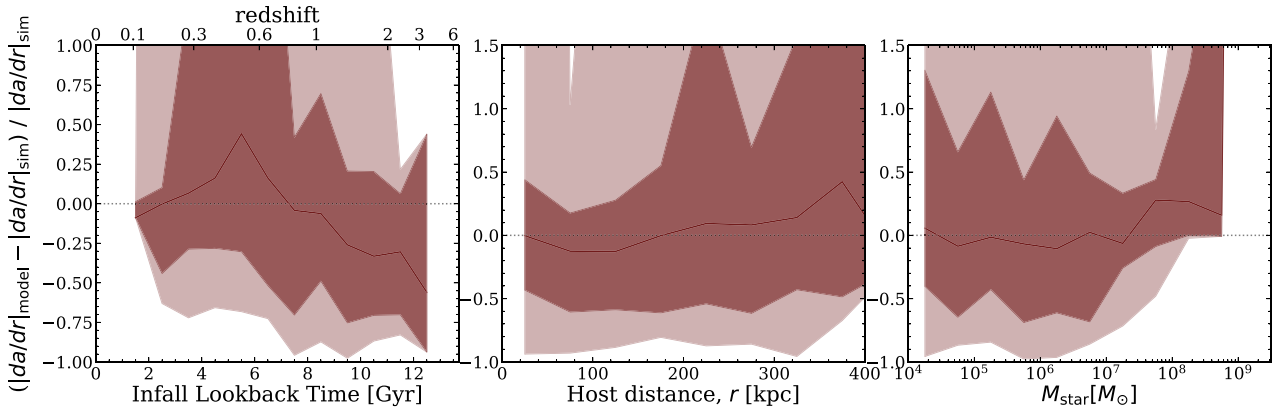
Figure 10. Comparing the maximum tidal acceleration, $|da/dr|$, from the MW-mass host that satellites experienced via orbit modelling versus in the simulations, as a function of lookback time of infall into the MW-mass halo (left), distance from the host, r (middle), and satellite M_{star} (right). Solid lines show the median and the dark and light shaded regions show the 68th and 95th percentiles across all satellites. These trends mirror those for the minimum pericentre distance, $d_{peri, min}$, in Fig. 9 (top). Because this orbit modelling does not account for the growth of the MW-mass host and the orbits are periodic, it increasingly overpredicts the pericentre distance, and underpredicts the tidal acceleration, with increasing $t^{lb}_{infall, MW} \gtrsim 5$ Gyr. The dependence on r and M_{star} are weak. However, the 1σ scatters span more than 50 per cent in each of the panels here, and up to a factor of 2, highlighting the large uncertainties.

derivative with respect to r and save the maximum $|da/dr|$ that a satellite experienced after first infall.

Fig. 10 compares the maximum $|da/dr|$ experienced between the simulation and model. Satellites that fell in $t^{lb}_{infall, MW} = 2.5$–7 Gyr ago typically have larger minimum pericentres in the simulations than in orbit modelling, so the model overpredicts $|da/dr|$ for these satellites by up to 45 per cent. Conversely, satellites that fell in $t^{lb}_{infall, MW} \gtrsim 7$ Gyr ago show larger minimum pericentres in the orbit models, so underpredicts $|da/dr|$ for the earliest satellites by up to 55 per cent. Because the simulations and orbit models agree in $d_{peri, min}$ for satellites that fell in $t^{lb}_{infall, MW} < 2.5$ Gyr ago, the median fractional difference is near zero.

Fig. 10 (middle and right) shows that $|da/dr|$ has little to no dependence on present-day satellite distance or M_{star}. Although the median fractional difference is close to 0 in both panels, the 1σ scatter increases from 0.39 to 2 with r, and the mean scatter versus M_{star} is 73 per cent. In all three panels, the 2σ scatter spans 100 per cent or more, so while the median $|da/dr|$ across the population from orbit modelling is relatively accurate, for any given satellite, orbit modelling over- or underpredicts $|da/dr|$ typically by a factor of ≈ 2.

Finally, Appendix C compares both the timing and distance of pericentres between orbit modelling and the simulations at each previous lookback pericentric event. The bias in both the distance and timing of pericentres increases with increasing lookback pericentre events to roughly 20 per cent in distance, and ≈ 1.5 Gyr in time. The uncertainty in these measurements increases up to the 4th most recent pericentre to ≈ 50 per cent in distance and ≈ 1 Gyr in time. Beyond this, $\lesssim 8$ per cent of satellites experienced 5 pericentres or more.

In summary, we compared various pericentre properties for both the minimum and most recent pericentre events. Across the full sample, the median fractional difference, or bias across the population, for the minimum and most recent pericentre distances are 2.5–6.6 per cent. The bias in the pericentre velocity is within ≈ 20 km s^{-1}, within 0.2 Gyr for the timing of the most recent pericentre, and <2 for the number of pericentric events. Finally, the bias in the maximum tidal acceleration is typically 10's of per cent across infall time, r, and M_{star}. Just as importantly, the typical 1σ scatter, which represents the uncertainty for a given satellite, is significant at ≈ 10–70 per cent.

3.3.4 Apocentre, orbital period, and eccentricity

The apocentre measure how far a satellite orbits from its host, and an orbit spends most of its time near apocentre. Fig. 11 (top) compares trends in the most recent apocentre distance, $d_{apo, rec}$. We only measure an apocentre that occurs after infall into the MW-mass halo. About 68 per cent of our satellites experienced an apocentre; the rest are on first infall.

Versus infall time (top left), 8 satellites fell into the host between $t^{lb}_{infall, MW} = 2$–3 Gyr ago, and orbit modelling generally overpredicts $d_{apo, rec}$ by $\gtrsim 75$ per cent for half of them. The fractional difference in apocentre distance is smaller for earlier-infalling satellites, and the median is ≈ 0.015 with a mean 1σ scatter of 0.06. Fig. 11 (top middle) shows little dependence with r. Satellites that currently orbit at smaller distances generally fell into the MW-mass host earlier, so orbit modelling somewhat underpredicts $d_{apo, rec}$ for satellites within $\lesssim 250$ kpc, similar to how it underpredicted $d_{apo, rec}$ at large $t^{lb}_{infall, MW}$ (top left). Overall, the mean 1σ scatter is ≈ 0.08. Finally, the median fractional difference in apocentre distance decreases weakly with M_{star} (top right). Lower-mass satellites typically fell into their MW-mass halo earlier, and they have smaller fractional differences.

Fig. 11 (middle row) shows trends in the most recent orbital period, t_{orbit}. We define the orbit period as the time difference between the two most recent pericentric passages, and we find nearly identical results using the times between apocentres. 47 per cent of the satellites in our sample experienced 2 or more pericentres in both the simulations and orbit modelling.

In the left panel, satellites that fell in $t^{lb}_{infall, MW} < 4.5$ Gyr ago did not have enough time to undergo 2 pericentres. For earlier infalling satellites, the median difference in t_{orbit} varies by as much as -0.4 Gyr, but the mean across all infall times is -0.15. The difference in t_{orbit} is negligible versus r. Orbit modelling does not account for dynamical friction, therefore, for satellites with $N_{peri} \geq 2$, if orbit modelling underpredicts $d_{peri, rec}$ compared to the simulations, it suggests more bound orbits and smaller t_{orbit} values, which is what we see for these satellites with $M_{star} \gtrsim 10^{6.5}$ M$_\odot$.

Finally, Fig. 11 (bottom row) compares the orbital eccentricities, e. Fig. 11 (bottom left) shows that, for satellites that fell in
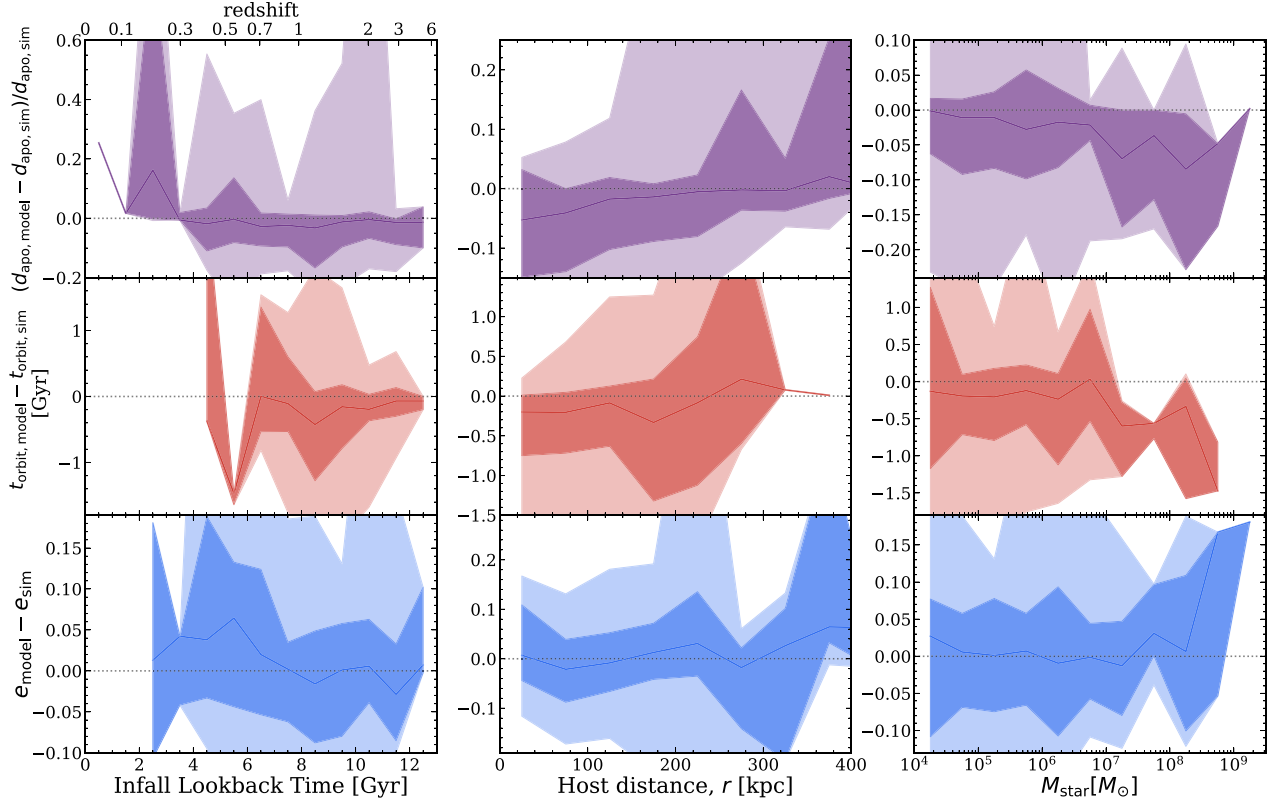
Figure 11. Comparing the most recent apocentre distance, d_{apo}, orbital period, t_{orbit}, and orbital eccentricity, e, from orbit modelling versus the simulations, as a function of lookback time of infall in the MW-mass halo (left), current distance from MW-mass host, r (middle), and satellite M_{star} (right). Solid lines show the median and the dark and light shaded regions show the 68th and 95th percentiles across all satellites. Top row: The most recent apocentre distance. Versus infall time, the median is roughly constant at -2 per cent for satellites that fell in $\gtrsim 3.5$ Gyr ago. The model recovers the median apocentre distance to within ± 5 per cent versus r, but the median decreases slightly with M_{star} from ≈ 0 to 8 per cent at $M_{star} = 10^{8.25}$ M$_\odot$. Although the medians in each panel may only be within a few per cent, the 1σ scatters span ≈ 5–10 per cent or more, highlighting the uncertainty for a given satellite. Middle row: Difference between the most recent orbital time, t_{orbit}, defined as the difference in time between the two most recent pericentres. Because orbit modelling generally underpredicts the recent pericentre lookback times, and because the orbits are periodic, the median difference in t_{orbit} is slightly negative across all panels, and even as low as 0.5–1.5 Gyr for satellites with $M_{star} \gtrsim 10^{7.25}$ M$_\odot$. Bottom row: Difference between most recent orbital eccentricity, $e = (d_{apo} - d_{peri})/(d_{apo} + d_{peri})$. The difference in e varies by at most 0.06 versus $t_{infall,MW}^{lb}$ and r, but satellites with $M_{star} > 10^{8.5}$ M$_\odot$ have differences >0.15. In general, orbit modelling recovers the median properties here within ≈ 7 per cent (see Table 2), though with significant scatter. The 1σ scatters span 8–15 per cent, likely because these properties all depend on pericentre and apocentre events that occur in the recent past.

$t_{infall,MW}^{lb} \lesssim 4$ Gyr ago, orbit modelling recovers $d_{peri,rec}$ well (Fig. 9, top left) and overpredicts $d_{apo,rec}$. Similarly, although orbit modelling recovers $d_{apo,rec}$ well for satellites up to $t_{infall,MW}^{lb} = 7.5$ Gyr ago, it also underpredicts $d_{peri,rec}$, which drives e to be higher in orbit modelling. Orbit models and simulations show similar results for earlier-infalling satellites. The median difference in e is flat with r and $M_{star} < 10^{8.25}$ M$_\odot$.

Again, we compute all results in this subsection based on the most recent pericentre and apocentre, but as Fig. 9 showed, orbit modelling performs worse for earlier properties of an orbit, so comparison of orbital period and eccentricity at earlier stages of these orbits would show even larger disagreements.

3.3.5 Recoverability of orbit properties

We compared 15 properties of satellite orbits in our cosmological simulations against orbit models using static axisymmetric potentials that we fit near-exactly to our hosts at $z = 0$. Table 2 lists the properties that we tested, as well as the median offsets and 1σ and 2σ scatters across our sample of satellites. We compare both

the raw difference of a given orbit property, X, defined as $X_{model} - X_{sim}$, as well as the fractional difference, $(X_{model} - X_{sim})/X_{sim}$. Additionally, we show the fractional change in orbital specific energy since infall relative to the MW-mass potential, $(E_0 - E_{inf})/U_{200m,0}$ and the fractional change in orbital specific angular momentum relative to today, $(\ell_0 - \ell_{inf})/\ell_0$. The orbit models conserve these quantities by definition.

We quantify the goodness of the orbit models in terms of their 'bias' (accuracy) and 'uncertainty' (precision) in the right-most columns of Table 2. The 'bias' describes how well orbit modelling accurately recovers the median orbital property across the satellite population: we define a property to be minimally, moderately, or highly biased if the median fractional offset of the satellite population between orbit modelling and the simulations is <10 per cent, 10–25 per cent, or >25 per cent, respectively. However, even in cases where this bias is small (accuracy is high), orbit modelling can have severe limitations if it cannot model the history of a given satellite to good precision. Thus, we also quantify the 'uncertainty' via how large the scatter in this difference between orbit models and simulations is across the satellite population. We define a property to be minimally, moderately, or highly uncertain if the 1σ scatter is <25 per cent,

Table 2. Comparing the results of orbit modelling in a static axisymmetric host potential against cosmological baryonic simulations.Column list: property name; variable; median offset, 1σ scatter, 2σ scatter.The left columns compare the raw difference, $X_{model} - X_{sim}$, while the middle columns compare the fractional difference, $(X_{model} - X_{sim})/X_{sim}$.Additionally, we describe the strength of the bias (median fractional offset) and uncertainty (1σ scatter of the fractional offset) associated with each property.In the bottom two rows, we show the difference in the orbital energy between present day and the infall time, normalized by the host halo potential energy at present-day, $(E_0 - E_{inf})/U_{200m, 0}$, and the fractional change in the orbital specific angular momentum relative to present-day, $(\ell_0 - \ell_{inf})/\ell_0$.Given that energy and angular momentum are always conserved in the model, we place them below a horizontal line to distinguish them.

Orbital property	Variable	Raw Difference			Fractional Difference			Bias	Uncertainty
		Median offset	1σ scatter	2σ scatter	Median offset	1σ scatter	2σ scatter		
Recent pericentre distance	$d_{peri, rec}$ [kpc]	−1.26	12.1	68.8	−0.025	0.21	1.19	Min	Min
Min pericentre distance	$d_{peri, min}$ [kpc]	3.23	18.5	80.1	0.066	0.53	3.56	Min	High
Lookback time of recent pericentre	$t_{peri, rec}$ [Gyr]	−0.03	0.25	1.72	−0.028	0.09	0.64	Min	Min
Number of pericentres within $R_{200m}(t)$	N_{peri}	0.63	1.14	2.28	0.32	0.72	1.44	High	High
Number of pericentres within $R_{200m, 0}$	$N_{peri, fixed}$	0.53	1.18	2.36	0.24	0.74	1.48	Mod	High
Velocity at recent pericentre	$v_{peri, rec}$ [km s^{-1}]	7.69	26.8	106	0.030	0.10	0.38	Min	Min
Velocity at min pericentre	$v_{peri, min}$ [km s^{-1}]	3.10	43.4	132	0.012	0.17	0.43	Min	Min
Recent apocentre distance	$d_{apo, rec}$ [kpc]	−2.75	12.1	97.3	−0.013	0.06	0.38	Min	Min
Lookback time of infall within $R_{200m}(t)$	t_{inf} [Gyr]	0.57	2.33	5.53	0.09	0.41	2.50	Min	Mod
Lookback time of infall within $R_{200m, 0}$	$t_{inf, fixed}$ [Gyr]	4.17	3.55	8.80	0.44	0.55	2.92	High	High
Recent eccentricity	e_{rec}	0.003	0.07	0.21	0.005	0.15	0.54	Min	Min
Recent period	T_{rec} [Gyr]	−0.19	0.49	1.84	−0.071	0.13	0.41	Min	Min
Max tidal acceleration	$|da/dr|$ [Gyr^{-2}]	−0.07	15.7	353	−0.017	0.66	4.20	Min	High
Energy change since infall	$(E_0 - E_{inf})/U_{200m, 0}$	-	-	-	−0.54	0.82	1.83	High	High
Angular momentum change since infall	$(\ell_0 - \ell_{inf})/\ell_0$	-	-	-	−0.015	0.42	1.68	Min	Mod

25–50 per cent, or >50 per cent. Because bias is more problematic (systematic) than uncertainty, we impose stricter criteria for it.

Fig.12 visually represents these summary results, via the median offsets, and 1σ and 2σ scatters, for the fractional differences between orbit modelling and simulations. We rank order each property independently in each panel.

The median fractional offset (representing the 'bias') of all properties ranges from −0.54 for the fractional change in specific energy to 0.44 for the lookback time of satellite infall using a fixed $R_{200m}(t_{lb} = 0)$. The properties whose median agrees to within ±5 per cent across the population include: the most recent pericentre distance, $d_{peri, rec}$, the lookback time of the most recent pericentre, $t_{peri,rec}^{lb}$, the maximum value of the derivative of the tidal acceleration, $|da/dr|$, the fractional change in the angular momentum relative to today, $(\ell_0 - \ell_{infall})/\ell_0$, the most recent apocentre distance, $d_{apo, rec}$, the eccentricity of the most recent orbit, e_{rec}, and the total satellite velocities at the minimum and most recent pericentres, $v_{peri, min}$ and $v_{peri, rec}$, respectively. Not surprisingly, orbit modelling tends to model/recover *recent* properties of an orbit with the least bias across a population.

Properties that agree moderately, to within 5–10 per cent, include: the distance of the minimum pericentre, $d_{peri, min}$, the lookback time of

infall into the MW-mass host, $t_{infall,MW}^{lb}$, and the most recent satellite orbit period, $t_{orbit, rec}$. Both $t_{infall,MW}^{lb}$ and $d_{peri, min}$ occurred further back in time than the properties that show the least bias.

Finally, the properties that are most systematically biased in orbit modelling are: the number of pericentric passages both with an evolving and fixed R_{200m}, N_{peri} and $N_{peri, fixed}$, the lookback time of infall when keeping a fixed $R_{200m} = R_{200m}(t_{lb} = 0)$, and the change in orbital energy since infall, E_{inf}.

Just as important as examining the bias (offset in the median across the population) is the uncertainty for a given satellite, via the scatter across the population. This ranges across $\approx$0.06–0.82 at 1σ and $\approx$0.38–4.2 at 2σ. Again, properties that occurred more recently generally have smaller 1σ scatter (aside from $v_{peri, min}$).

At best, the uncertainty for a given satellite is 6 per cent in the apocentre distance, and $\gtrsim$ 10 per cent for all other properties. Additionally, these uncertainties reach nearly a factor of $\approx$2 in energy, and the 2σ scatters are $\gtrsim$ 40 per cent.

Because we model the host potential to within a few per cent at z = 0, the uncertainties in Table 2 represent lower limits to the bias/uncertainty in orbit modelling in practice.
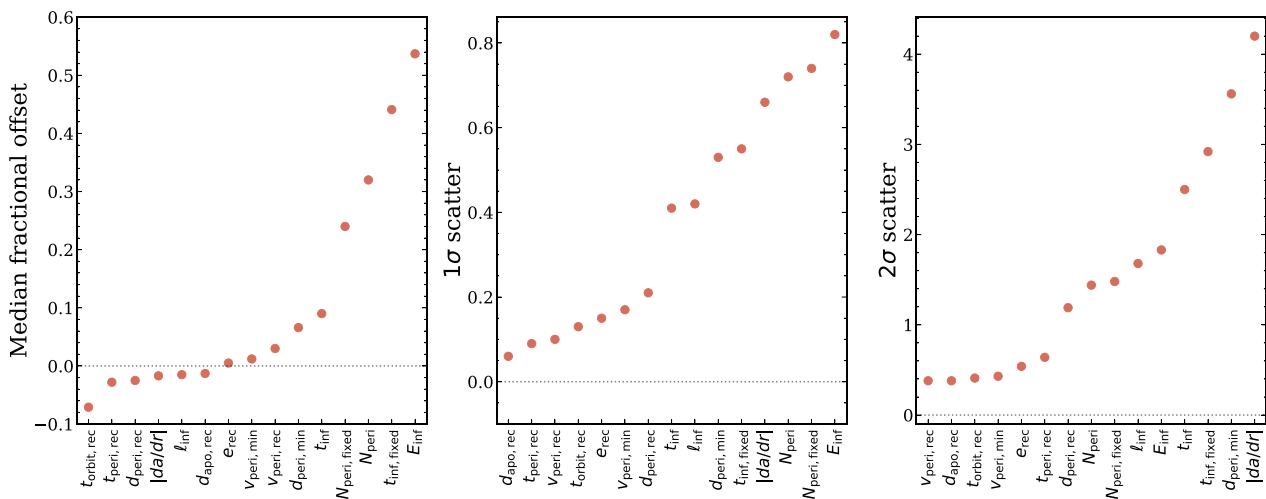
Figure 12. Rank ordering the 15 properties of the orbit histories of satellite galaxies, as Table 2 lists, based on the fractional level of agreement between orbit modelling in a static axisymmetric host potential and the cosmological baryonic simulations. For clarity, we shorten $(E_0 - E_{inf})/U_{200m, 0}$ to E_{inf}, and $(\ell_0 - \ell_{inf})/\ell_0$ to ℓ_{inf}, and we show the absolute change of E_{inf} in the left panel to more easily compare with the other properties. The left panel shows the median offset, in order from negative to positive values, and the middle and right panels show the 1σ and 2σ scatters in order of agreement. Satellite orbit properties that occurred recently, such as the recent pericentre distances/times/velocities, show smaller median offsets, $\lesssim 3$ per cent, compared to properties that occurred further in the past, such as the minimum pericentre distances or infall times, $\gtrsim 5$ per cent. The same is generally true for the 1σ scatter, but not always for the 2σ scatter, despite the strong correlation between the two. Thus, deriving orbit parameters in the *recent* past yields *median* results largely consistent with cosmological simulations, but almost always with significant scatter (uncertainty) always $\gtrsim 10$ per cent, and orbit parameters that occurs further in the past generally suffer from non-trivial bias and *significant* uncertainty. Furthermore, these results are best-case scenarios for static axisymmetric host potentials, because we fit them near-exact to the simulations at $z = 0$.

4 SUMMARY & DISCUSSION

4.1 Summary of results

We compared 15 orbit properties for 493 satellite galaxies around 13 MW-mass hosts in the FIRE-2 suite of cosmological baryonic simulations against orbit histories derived from orbit modelling in a static, axisymmetric potential for the same hosts, to quantify rigorously the accuracy and precision of this orbit modelling technique. Specifically, we fit axisymmetric potentials to each MW-mass hosts at $z = 0$ to within a few per cent, which also means that the uncertainties that we present are lower limits to a more realistic scenario applied to the MW/M31 with uncertainty in the underlying host potential. We now discuss the key questions we raised in the Introduction and our corresponding results.

How much has the mass profile of a MW-mass host evolved over the orbital histories of typical satellites?

(i) Most surviving satellites first fell into their MW-mass halo 3.4–9.7 Gyr ago. During that time, M_{200m} and R_{200m} of the host were 33–86 per cent and 26–73 per cent of their values today (Fig. 1), so they roughly doubled since then.

(ii) Perhaps more relevant for satellite orbits, the total enclosed mass within a fixed physical distance increased meaningfully (Figs 2 and 3). Within 50 kpc, the typical recent pericentre distance of our satellites, the enclosed mass was only ≈ 74 per cent of its present-day value at typical satellite infall times (≈ 7.4 Gyr ago).

(iii) The fractional increase in the enclosed mass of the host is larger at smaller distances (Figs 2 and 3). This is contrary to the expectations of 'inside-out' growth of a dark-matter halo from DMO simulations, where most halo growth occurs at larger radii (e.g. Diemand, Kuhlen & Madau 2007; Wetzel & Nagai 2015). With the inclusion of baryonic physics (most importantly gas cooling),

more physical growth occurs at smaller distances, most relevant for satellites with smaller pericentres.

How well does orbit modelling in a static axisymmetric host potential recover key orbital properties in the history of a typical satellite?

(i) Calculating the infall time of a satellite in the model with a growing $R_{200m}(t)$ yields more consistent results with the simulations, $\lesssim 2$ Gyr offset, compared to using the fixed $R_{200m}(t = 0)$ at present-day, but the 1σ uncertainties in both metrics can be as high as ≈ 5–6 Gyr (Fig. 8).

(ii) Orbit history properties that occurred more recently have smaller fractional offsets and uncertainties than properties that occurred in the past. For instance, the timing and distance of the most recent pericentre have median fractional offsets (uncertainties) $\lesssim 3$ (9–21) per cent, compared to the minimum pericentre distance, which occurred further back in time, which has a fractional offset (uncertainty) of 7 (53) per cent (Figs 8 and 9 and Table 2).

(iii) The orbit properties that orbit modelling recovers best (with the smallest bias and uncertainty) include the distance, timing, and velocity of the recent pericentre, the velocity at the minimum pericentre, the most recent apocentre distance, and the orbit eccentricity and period. The properties that are not recovered well (largest bias and/or uncertainty) include the minimum pericentre distance, number of pericentric passages, the lookback time of infall into MW-mass host halo, maximum strength of the tidal field, and the change in total orbital energy (Fig. 12 and Table 2).

(iv) Even with near-perfect knowledge of the mass distribution/potential at $z = 0$ in the host galaxies, the typical uncertainties in these orbit properties range from 6–82 per cent. Furthermore, the satellite-to-satellite variations in each, that is, the 2σ scatters, are $\gtrsim 40$ per cent. Thus, one cannot recover these orbit properties to within a factor of ≈ 2 or so, which cautions against overgeneralizing

the results for a single satellite from the median trends (Fig. 12 and Table 2).

(v) At fixed mass, the spatial extent or orientation of the host galaxy disc does not significantly affect the orbital properties of the satellites. Compared to a disc that is rotated by 90°, or a point mass disc model, the median offsets between the fiducial disc model are within $\approx 10^{-3}$ per cent, and the widths of the 68th percentiles are less than 6×10^{-2} per cent (Table B1).

How far back in time can one reliably model the orbital history of satellites in a static axisymmetric host potential?

(i) The specific energy and specific angular momentum of an orbit are generally not conserved: uncertainties are less than 25 per cent only back to ≈ 3.1 Gyr. Backward integrating orbits more than ≈ 9 Gyr results in energy uncertainties up to a factor of ≈ 2 or more (Figs 4, 5, and 12 and Table 2).

(ii) At the most recent orbit, the uncertainty in pericentre distance is already ~ 20 per cent, and ~ 200 Myr in pericentre timing (Fig. C1). Subsequent orbits result in larger uncertainties.

4.2 Discussion

4.2.1 Comparison to D'Souza & Bell

Our analysis is closest to that of D'Souza & Bell (2022), who similarly fit symmetric models to MW-mass haloes from the ELVIS suite of DMO simulations (Garrison-Kimmel et al. 2014) to study the uncertainties associated with orbit modelling. First, we note key differences in methods. D'Souza & Bell (2022) used DMO simulations, which neglects the (tidal) acceleration from the central galaxy that modify these orbits and could strip/disrupt satellites that orbit nearby. The internal stellar feedback in a satellite also can reduce the inner density of dark matter and make the satellites more vulnerable to tidal disruption (Bullock & Boylan-Kolchin 2017), although this is a second-order effect (Garrison-Kimmel et al. 2017). Without these baryonic effects and processes, the surviving satellites in DMO simulations typically fell into their MW-mass halo earlier and were able to complete more pericentres, while also orbiting *closer* to the centre of the host with smaller pericentric passages than we showed in Santistevan et al. (2023). However, D'Souza & Bell (2022) did account for the effects of dynamical friction in their model, similar to Patel et al. (2020), which acts to slow satellites down and ultimately merge within the host, which we do not. D'Souza & Bell (2022) also examined the gravitational effects from LMC-like analogues in MW-mass hosts. Because the LMC is so massive, it hosts its own satellite population, and studies suggest that it is on first infall into the MW and near its first pericentre (e.g. Kallivayalil et al. 2013; Deason et al. 2015; Kallivayalil et al. 2018; Patel et al. 2020), so accounting for its gravitational influence on the surrounding satellites is of great interest. Although some hosts in our simulations have LMC-like analogues at previous snapshots (see Samuel et al. 2021; Barry et al. 2023), we do not analyse them specifically.

Another significant difference between D'Souza & Bell (2022) and our analysis is that they account for the true mass growth of the MW-mass host at every snapshot by updating their potentials while keeping the potential fixed between snapshots. We do not account for the mass growth of the MW-mass host, because the majority of orbit modelling studies in the literature implement a fixed mass/potential (e.g. Patel, Besla & Sohn 2017; Fritz et al. 2018a; Fillingham et al. 2019; Pace, Erkal & Li 2022), because we do not know the full mass histories of the MW or M31, and also because the mass assembly history for each MW-mass host in our simulations is unique.

D'Souza & Bell (2022) define a recovered property as being when the absolute value of the fractional difference is less than 30 per cent, that is, $|X_{\rm true} - X_{\rm model}|/X_{\rm true} < 0.3$, and report the fraction of satellites that do not meet this criterion as being 'outliers'. They centre their results on two hosts, 'iDouglas' which is an isolated MW-mass galaxy and 'iOates' which has an LMC analog, but they test orbits in iOates with and without the gravitational contribution from this massive companion. Similar to our analysis, they focus on the timing and distance of various pericentre events, the apocentre distances, and the infall times of satellites, and find that more recent pericentres and apocentres have smaller outlier fractions than pericentres or apocentres that happened at earlier times. In particular, the outlier fractions for the most recent, and second-most recent pericentre distances in the hosts without the additional massive satellite are 31.2–47 and 43.8–69.9 per cent, respectively. Although we do not look specifically at the second-most recent pericentre, we do find that the median fractional offsets and 1σ uncertainties for the most recent pericentre distance, -2.5 and 21 per cent, are smaller compared to the same for the minimum pericentre distance which often occurred ≈ 6 Gyr earlier, 0.066 and 53 per cent. The authors also show that the timing of the most recent pericentre is often better recovered than the distance, with an outlier fraction of 13.8–23.2 per cent. Our results show that the median fractional offset is -2.8, which is comparable to the offset in recent pericentre distance, but with a smaller 1σ uncertainty of 9 per cent. Thus, it is often easier to recover the timing of a pericentre than its distance.

D'Souza & Bell (2022) similarly showed that the distance of the most recent apocentre has a smaller outlier fraction than the most recent pericentre distance, with a value of only 6.2–34.9 per cent. Our work also suggests that the apocentres are easier to recover, with a median fractional offset and 1σ uncertainty of -1.3 and 6 per cent, respectively. They conclude that apocentres are easier to model because they only depend on the binding energy of a satellite galaxy, while the pericentres depend on the angular momentum of a satellite as well as its binding energy. Properties at apocentre also do not intricately depend on the details of the gravitational potential at small distances like pericentres do, and rather, what is more important is modelling the total enclosed mass precisely, as both studies have done. Finally, the authors also calculate the infall times of satellites and find good agreement in their simulations and model, with an outlier fraction of 11.2–28.9 per cent. Although we generally see small median offsets when calculating infall time with an evolving $R_{\rm 200m}(t)$, ≈ 6.7 per cent, the associated uncertainty is high, ≈ 41 per cent, and using a fixed $R_{\rm 200m}(t = 0)$ is worse.

D'Souza & Bell (2022) explored other models for the growth of the MW-mass host, and most comparable to our work is their results in which they keep the mass fixed over time. They report that, depending on the property, using the *static* model in some cases produces the best results as compared with the simulations. However, as one integrates longer in time, the static model becomes less representative of the MW-mass environment, and thus satellite orbit properties that occurred earlier are not modelled as well, which is what we similarly see in the minimum pericentre properties and infall times. The authors explored a model that accounts for the median mass growth of the 48 MW-mass haloes in the ELVIS DMO simulation suite and found that both the static model and median mass growth model reproduce similar results at recent lookback times as well. Finally, they implemented a static model with 40 per cent more mass, and a static model with different halo concentrations, which both returned larger biases, scatters, and outlier fractions. Thus, modelling the mass and shape of the haloes at present-day is important.

Finally, D'Souza & Bell (2022) compared uncertainties in the virial mass of the MW-mass host, and thus uncertainties in the potential, the uncertainty in the 6D phase-space coordinates of the satellites at $z = 0$, the uncertainty in modelling the recent LMC-like accretion, and the uncertainty in modelling the motion of the MW-mass system as it moves throughout the Universe. They concluded that the uncertainties in the recovered orbit history properties when using simple parametric forms of the potential or ignoring the LMC-like contribution to the potential are comparable to the uncertainty in results caused by a ≈ 30 per cent uncertainty in the virial mass of the MW-mass host.

4.2.2 Comparison to other studies

The results in Figs 2 and 3 (top) differ strikingly from those from DMO simulations, where the enclosed mass at smaller radii is set earlier than at larger radii. For example, using the Via Lactea DMO simulation, Diemand, Kuhlen & Madau (2007) showed that the enclosed dark matter mass within 100 kpc assembled prior to $z \approx 1.7$ ($t_{lb} \approx 9.9$ Gyr ago), and grew only by about 10 per cent since then. The enclosed mass at even smaller distances formed earlier. In a related analysis, Wetzel & Nagai (2015), showed that the enclosed dark-matter mass within a fixed physical $r \approx 50$–100 kpc at $z = 1$ was already $\gtrsim 85$–95 per cent of its value at $z = 0$, compared to the mass in stars and gas, which were only ~ 55–70 per cent of the mass at $z = 0$. Because dark matter is collisionless and dissipationless, and because the accretion radius grows over time, dark-matter growth occurs largely 'inside-out'. However, gas can cool over time, which drives the formation of the central galaxy, and leads to more physical mass growth at smaller radii compared with the dark matter.

To derive the orbit histories of satellites of the MW and M31, many studies commonly apply orbit modelling in a static host potential or use simple approximations of the growth of the host. For instance, Kallivayalil, van der Marel & Alcock (2006) used a fixed MW-mass potential, as well as an LMC-like potential, to determine that the SMC was gravitationally bound to the LMC. Kallivayalil et al. (2013) used 3 epochs of *HST* measurements to constrain the LMC's proper motions and suggest that the LMC is likely on its first infall, as previous studies have suggested (e.g. Besla et al. 2007). The authors compared different static models of the MW and models that account for the growth of the MW-mass halo over the last 10 Gyr, but in the evolving models, the authors did not account for the central galaxy. Patel, Besla & Sohn (2017) similarly sought to understand the orbit of the LMC around the MW, as well as the orbit of M33 around M31, and concluded that like the LMC, M33 is likely fell into the M31 halo less than $\lesssim 2$ Gyr ago or so. Although these authors modelled the many components of the main galaxy, they did not account for any time dependence of the host potential.

Patel, Besla & Sohn (2017) also compared the orbits of massive satellite analogues in the DMO Illustris-1-Dark simulation to an NFW model of the MW-mass host halo with a dynamical friction model included. The authors used the $z = 0$ 6D phase-space coordinates for the satellites and integrated their orbits, similar to our pipeline, but for 6 Gyr, and suggest that this type of orbit modelling technique shows good agreement between the two orbits for satellites on first infall and for satellites that recently completed their first pericentre. The orbits that we present in Fig. 6 show good agreement between the simulations and model for recent lookback time as well, however, only the top-left panel shows good agreement for up to 6 Gyr. There are other satellites in our sample that show agreement for this time range, however, we choose to intentionally show test cases in which the model does not do well in this figure as well.

Recent work not only suggests that the LMC has only recently fallen into the MW's halo and is currently near its first pericentre (e.g. Kallivayalil et al. 2009, 2013), but that it has a satellite population of its own (e.g. Deason et al. 2015; Kallivayalil et al. 2018; Patel et al. 2020). Other studies account for the gravitational contribution of the LMC on the other satellites of the MW. Kallivayalil et al. (2018) used *Gaia* data, in conjunction with the DMO Aquarius simulation, to determine which satellites might be gravitationally bound to the LMC. The simulation does have an LMC analogue, with a similar position and velocity at $z = 0$, but does not account for the gravitational effect of the central galaxy. Patel et al. (2020) used newer *Gaia* observations and numerically integrated the orbits of satellites in a model of the potential with a MW, LMC, and SMC component to determine the orbital histories of the LMC and its satellites. Although they keep the potential fixed over time, they only backward integrated the orbits of satellites for ≈ 6 Gyr, given that beyond this time, MW-mass haloes typically have $\lesssim 80$ per cent of their mass at $z = 0$. The authors concluded that the derived orbits for satellites of the LMC strongly depend on whether or not you include the LMC in the global potential, and in some cases, the contribution from the SMC is important as well. Finally, other studies aim to accurately model the LMC potential with basis function expansions fit to simulation data to understand how it affects the MW, and its satellites, as in Garavito-Camargo et al. (2019, 2021); Correa Magnus & Vasiliev (2022), but doing so while accounting for the growth of the MW remains challenging.

Studies such as Fritz et al. (2018a) and Fillingham et al. (2019) provide inferences of the infall times, pericentre and apocentre distances, and orbit eccentricities for all satellites in the MW. Fritz et al. (2018a) used data from *Gaia* DR2, and numerically integrated the orbits of all satellites in GALPY in two different models for the MW potential, which they kept fixed over time. Depending on the mass of the MW, the authors showed that some of the apocentres for the satellites can lie either inside or outside of the virial radius, which has implications for instance in studying how a satellite interacts with the hot gas in the MW halo. Furthermore, the authors suggested that some satellites have pericentres as close as ≈ 20 kpc, where the strong tidal acceleration from the central galaxy is important. As we showed in Figs 1–3, depending on when these pericentres take place, the mass of the host was only a fraction of its mass today. Fillingham et al. (2019) used the Phat ELVIS DMO simulations, which include an analytic disc potential, to statistically sample satellites with similar present-day positions and velocities to make cosmologically informed predictions for what the infall times of the MW's satellites were. The Phat ELVIS simulations account for the growth of the disc potential through abundance matching scaling relations, where each disc has a unique growth rate, and thus, tidally disrupt subhaloes that orbit close to the disc (Kelley et al. 2019).

Most of the MW-mass simulations in the FIRE-2 suite do not have an LMC-like analogue at $z = 0$, but there are at least four analogues at earlier times, within the last 6–7 Gyr ago, and previous works have used these analogues to study both planar configurations of satellites and the effect of LMC-mass satellites on subhalo populations (Samuel et al. 2021; Barry et al. 2023). Given that the LMC recently fell into the MW's dark matter halo, $\lesssim 2$ Gyr ago (e.g. Besla et al. 2007; Kallivayalil et al. 2013; Patel, Besla & Sohn 2017), previous orbit modelling studies that include the effects of the LMC to derive orbits of the MW's satellites work well in this recent regime. On the other hand, because we do not include the effects of an LMC in the host potential, our work informs how well static potential orbit modelling works in the regime before the LMC was a satellite, that is, $\gtrsim 2$ Gyr ago. However, the results suggesting that the LMC

is on first infall and just passed its first pericentre are trustworthy, given that the mass of the MW and its dark matter halo have not significantly changed over the last $\lesssim 2\,\mathrm{Gyr}$. As such, both kinds of studies, both with and without an LMC analogue, complement each other in understanding the full, complex MW formation history.

ACKNOWLEDGEMENTS

DATA AVAILABILITY

The data in the figures above, and the PYTHON code that we used to analyse these data, are available at https://ibsantistevan.wixsite.com/mysite/publications, which uses the publicly available packages https://bitbucket.org/awetzel/gizmo_analysis, https://bitbucket.org/awetzel/halo_analysis, and https://bitbucket.org/awetzel/utilities. The FIRE-2 simulations are publicly available (Wetzel et al. 2023) at http://flathub.flatironinstitute.org/fire which include the star, gas, and dark matter particle data at various redshifts from $z = 0$ to 10, along with the halo merger trees, which include the positions and velocities of all satellites at all 600 snapshots. Additional FIRE simulation data are available at https://fire.northwestern.edu/data. A public version of the GIZMO code is available at http://www.tapir.caltech.edu/~phopkins/Site/GIZMO.html.

REFERENCES

Amorisco N. C., 2017, MNRAS, 464, 2882
Arora A., Sanderson R. E., Panithanpaisal N., Cunningham E. C., Wetzel A., Garavito-Camargo N., 2022, ApJ, 939, 2
Astropy Collaboration, 2022, ApJ, 935, 167
Bakels L., Ludlow A. D., Power C., 2021, MNRAS, 501, 5948
Barry M., Wetzel A., Chapman S., Samuel J., Sanderson R., Arora A., 2023, MNRAS, 523, 428
Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013a, ApJ, 762, 109
Behroozi P. S., Wechsler R. H., Wu H.-Y., Busha M. T., Klypin A. A., Primack J. R., 2013b, ApJ, 763, 18

Bennet P., Alfaro-Cuello M., Pino A. d., Watkins L. L., van der Marel R. P., Sohn S. T., 2022, ApJ, 935, 149
Bertschinger E., 1985, ApJS, 58, 39
Besla G., Kallivayalil N., Hernquist L., Robertson B., Cox T. J., van der Marel R. P., Alcock C., 2007, ApJ, 668, 949
Besla G., Peter A. H. G., Garavito-Camargo N., 2019, J. Cosmology Astropart. Phys., 2019, 013
Bland-Hawthorn J., Gerhard O., 2016, ARA&A, 54, 529
Bonaca A., Conroy C., Wetzel A., Hopkins P. F., Kereš D., 2017, ApJ, 845, 101
Bonaca A. et al., 2021, ApJ, 909, L26
Bovy J., 2015, ApJS, 216, 29
Bovy J., Rix H.-W., 2013, ApJ, 779, 115
Bovy J., Rix H.-W., Schlafly E. F., Nidever D. L., Holtzman J. A., Shetrone M., Beers T. C., 2016, ApJ, 823, 30
Boylan-Kolchin M., Ma C.-P., Quataert E., 2008, MNRAS, 383, 93
Boylan-Kolchin M., Besla G., Hernquist L., 2011, MNRAS, 414, 1560
Brooks A. M., Zolotov A., 2014, ApJ, 786, 87
Bryan G. L., Norman M. L., 1998, ApJ, 495, 80
Bullock J. S., Boylan-Kolchin M., 2017, ARA&A, 55, 343
Bullock J. S., Johnston K. V., 2005, ApJ, 635, 931
Correa Magnus L., Vasiliev E., 2022, MNRAS, 511, 2610
D'Souza R., Bell E. F., 2022, MNRAS, 512, 739
Dark Energy Survey Collaboration, 2016, MNRAS, 460, 1270
Deason A. J., Wetzel A. R., Garrison-Kimmel S., Belokurov V., 2015, MNRAS, 453, 3568
Diemand J., Kuhlen M., Madau P., 2007, ApJ, 667, 859
Diemer B., 2017, ApJS, 231, 5
Eadie G., Jurić M., 2019, ApJ, 875, 159
Eadie G. M., Springford A., Harris W. E., 2017, ApJ, 835, 167
El-Badry K., Wetzel A., Geha M., Hopkins P. F., Kereš D., Chan T. K., Faucher-Giguère C.-A., 2016, ApJ, 820, 131
El-Badry K. et al., 2018, MNRAS, 473, 1930
Evans N. W., Wilkinson M. I., 2000, MNRAS, 316, 929
Faucher-Giguère C.-A., Lidz A., Zaldarriaga M., Hernquist L., 2009, ApJ, 703, 1416
Fillingham S. P. et al., 2019, preprint (arXiv:1906.04180)
Fillmore J. A., Goldreich P., 1984, ApJ, 281, 1
Fritz T. K., Battaglia G., Pawlowski M. S., Kallivayalil N., van der Marel R., Sohn S. T., Brook C., Besla G., 2018a, A&A, 619, A103
Fritz T. K., Lokken M., Kallivayalil N., Wetzel A., Linden S. T., Zivick P., Tollerud E. J., 2018b, ApJ, 860, 164
Gaia Collaboration, 2018, A&A, 616, A1
Garavito-Camargo N., Besla G., Laporte C. F. P., Johnston K. V., Gómez F. A., Watkins L. L., 2019, ApJ, 884, 51
Garavito-Camargo N., Besla G., Laporte C. F. P., Price-Whelan A. M., Cunningham E. C., Johnston K. V., Weinberg M., Gómez F. A., 2021, ApJ, 919, 109
Garrison-Kimmel S., Boylan-Kolchin M., Bullock J. S., Lee K., 2014, MNRAS, 438, 2578
Garrison-Kimmel S. et al., 2017, MNRAS, 471, 1709
Garrison-Kimmel S. et al., 2018, MNRAS, 481, 4133
Garrison-Kimmel S. et al., 2019a, MNRAS, 487, 1380
Garrison-Kimmel S. et al., 2019b, MNRAS, 489, 4574
Geha M. et al., 2017, ApJ, 847, 4
Gómez F. A., Besla G., Carpintero D. D., Villalobos Á., O'Shea B. W., Bell E. F., 2015, ApJ, 802, 128
Gunn J. E., Gott J. Richard I., 1972, ApJ, 176, 1
Hahn O., Abel T., 2011, MNRAS, 415, 2101
Harris C. R. et al., 2020, CoRR, abs/2006.10256
Hill A. R., Muzzin A., Franx M., Marchesini D., 2017, ApJ, 849, L26
Hopkins P. F., 2015, MNRAS, 450, 53
Hopkins P. F. et al., 2018, MNRAS, 480, 800
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Ibata R. A., Gilmore G., Irwin M. J., 1994, Nature, 370, 194
Irrgang A., Wilcox B., Tucker E., Schiefelbein L., 2013, A&A, 549, A137
Ishchenko M., Sobolenko M., Berczik P., Khoperskov S., Omarov C., Sobodar O., Makukov M., 2023, A&A, 673, A152

[5] https://www.craiyon.com

Kallivayalil N., van der Marel R. P., Alcock C., 2006, ApJ, 652, 1213
Kallivayalil N., Besla G., Sanderson R., Alcock C., 2009, ApJ, 700, 924
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
Kelley T., Bullock J. S., Garrison-Kimmel S., Boylan-Kolchin M., Pawlowski M. S., Graus A. S., 2019, MNRAS, 487, 4409
Kroupa P., 2001, MNRAS, 322, 231
Krumholz M. R., Gnedin N. Y., 2011, ApJ, 729, 36
Leitherer C. et al., 1999, ApJS, 123, 3
Li T. S. et al., 2019, MNRAS, 490, 3508
Li Z.-Z., Zhao D.-H., Jing Y. P., Han J., Dong F.-Y., 2020, ApJ, 905, 177
Li T. S. et al., 2021, ApJ, 911, 149
Li T. S. et al., 2022, ApJ, 928, 30
Lilleengen S. et al., 2023, MNRAS, 518, 774
Lux H., Read J. I., Lake G., 2010, MNRAS, 406, 2312
Ma X., Hopkins P. F., Wetzel A. R., Kirby E. N., Anglés-Alcázar D., Faucher-Giguère C.-A., Kereš D., Quataert E., 2017, MNRAS, 467, 2430
Majewski S. R., Munn J. A., Hawley S. L., 1996, ApJ, 459, L73
Mao Y.-Y., Geha M., Wechsler R. H., Weiner B., Tollerud E. J., Nadler E. O., Kallivayalil N., 2021, ApJ, 907, 85
Martín-Navarro I., Pillepich A., Nelson D., Rodriguez-Gomez V., Donnari M., Hernquist L., Springel V., 2021, Nature, 594, 187
McCarthy I. G., Frenk C. S., Font A. S., Lacey C. G., Bower R. G., Mitchell N. L., Balogh M. L., Theuns T., 2008, MNRAS, 383, 593
More S., Diemer B., Kravtsov A. V., 2015, ApJ, 810, 36
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
Ogiya G., Taylor J. E., Hudson M. J., 2021, MNRAS, 503, 1233
Pace A. B., Erkal D., Li T. S., 2022, ApJ, 940, 136
Panithanpaisal N., Sanderson R. E., Wetzel A., Cunningham E. C., Bailin J., Faucher-Giguère C.-A., 2021, ApJ, 920, 10
Patel E., Mandel K. S., 2023, ApJ, 948, 104
Patel E., Besla G., Sohn S. T., 2017, MNRAS, 464, 3825
Patel E., Besla G., Mandel K., Sohn S. T., 2018, ApJ, 857, 78
Patel E. et al., 2020, ApJ, 893, 121
Pawlowski M. S., Sohn S. T., 2021, ApJ, 923, 42
Peñarrubia J., Benson A. J., 2005, MNRAS, 364, 977
Peñarrubia J., Kroupa P., Boily C. M., 2002, MNRAS, 333, 779
Planck Collaboration, 2020, A&A, 641, A6
Price-Whelan A. M., 2017, J. Open Source Softw., 2, 388
Price-Whelan A. M., Johnston K. V., Valluri M., Pearson S., Küpper A. H. W., Hogg D. W., 2016, MNRAS, 455, 1079
Price-Whelan A. M., Mateu C., Iorio G., Pearson S., Bonaca A., Belokurov V., 2019, AJ, 158, 223
Robles V. H., Bullock J. S., 2021, MNRAS, 503, 5232
Rocha M., Peter A. H. G., Bullock J., 2012, MNRAS, 425, 231
Rodriguez Wimberly M. K., Cooper M. C., Fillingham S. P., Boylan-Kolchin M., Bullock J. S., Garrison-Kimmel S., 2019, MNRAS, 483, 4031
Sales L. V., Wetzel A., Fattahi A., 2022, Nat. Astron., 6, 897
Samuel J. et al., 2020, MNRAS, 491, 1471
Samuel J., Wetzel A., Chapman S., Tollerud E., Hopkins P. F., Boylan-Kolchin M., Bailin J., Faucher-Giguère C.-A., 2021, MNRAS, 504, 1379
Samuel J., Pardasani B., Wetzel A., Santistevan I., Boylan-Kolchin M., Moreno J., Faucher-Giguere C.-A., 2022a, MNRAS, 525, 3849
Samuel J., Wetzel A., Santistevan I., Tollerud E., Moreno J., Boylan-Kolchin M., Bailin J., Pardasani B., 2022b, MNRAS, 514, 5276
Sanderson R. E. et al., 2018, ApJ, 869, 12
Sanderson R. E. et al., 2020, ApJS, 246, 6
Santistevan I. B., Wetzel A., El-Badry K., Bland-Hawthorn J., Boylan-Kolchin M., Bailin J., Faucher-Giguère C.-A., Benincasa S., 2020, MNRAS, 497, 747
Santistevan I. B., Wetzel A., Sanderson R. E., El-Badry K., Samuel J., Faucher-Giguère C.-A., 2021, MNRAS, 505, 921
Santistevan I. B., Wetzel A., Tollerud E., Sanderson R. E., Samuel J., 2023, MNRAS, 518, 1427
Shipp N. et al., 2018, ApJ, 862, 114
Shipp N. et al., 2019, ApJ, 885, 3
Shipp N. et al., 2023, ApJ, 949, 44

Simons R. C. et al., 2020, ApJ, 905, 167
Sohn S. T., Patel E., Fardal M. A., Besla G., van der Marel R. P., Geha M., Guhathakurta P., 2020, ApJ, 901, 43
Taylor J. E., Babul A., 2001, ApJ, 559, 716
Vasiliev E., 2019a, MNRAS, 482, 1525
Vasiliev E., 2019b, MNRAS, 482, 1525
Virtanen P. et al., 2020, Nat. Methods, 17, 261
Warfield J. T. et al., 2023, MNRAS, 519, 1189
Watkins L. L., Evans N. W., An J. H., 2010, MNRAS, 406, 264
Weinberg M. D., 1986, ApJ, 300, 93
Weisz D. R. et al., 2023, ApJS, 268, 19
Wetzel A., Garrison-Kimmel S., 2020a, HaloAnalysis: read and analyze halo catalogs and merger trees, Astrophysics Source Code Library, record ascl:2002.014
Wetzel A., Garrison-Kimmel S., 2020b, GizmoAnalysis: read and analyze Gizmo simulations, Astrophysics Source Code Library, record ascl:2002.015
Wetzel A. R., Nagai D., 2015, ApJ, 808, 40
Wetzel A. R., Tinker J. L., Conroy C., van den Bosch F. C., 2014, MNRAS, 439, 2687
Wetzel A. R., Deason A. J., Garrison-Kimmel S., 2015, ApJ, 807, 49
Wetzel A. R., Hopkins P. F., Kim J.-h., Faucher-Giguère C.-A., Kereš D., Quataert E., 2016, ApJ, 827, L23
Wetzel A. et al., 2023, ApJS, 265, 44
del Pino A. et al., 2022, ApJ, 933, 76
van den Bosch F. C., Ogiya G., 2018, MNRAS, 475, 4066
van den Bosch F. C., Aquino D., Yang X., Mo H. J., Pasquali A., McIntosh D. H., Weinmann S. M., Kang X., 2008, MNRAS, 387, 79
van der Marel R. P., Guhathakurta P., 2008, ApJ, 678, 187
van der Marel R. P., Fardal M., Besla G., Beaton R. L., Sohn S. T., Anderson J., Brown T., Guhathakurta P., 2012, ApJ, 753, 8

APPENDIX A: MODELLING HOST MASS PROFILES

We integrate the orbits of satellite galaxies using the galactic dynamics PYTHON package GALPY (Bovy 2015), which allows users to define custom host potentials. We thus fit the enclosed mass profiles of the MW-mass hosts in our simulations as a sum of disc and halo components, to supply to GALPY.

A1 Modelling the disc

We model the MW-mass disc as the sum of two double exponential discs, one for the inner disc/bulge, and one for the outer disc. The density profile of the disc is:

$$\rho(R, Z) = A_{\rm disc}^{\rm inner}\, e^{-R/R_{\rm disc}^{\rm inner} - |Z|/h_z} + A_{\rm disc}^{\rm outer}\, e^{-R/R_{\rm disc}^{\rm outer} - |Z|/h_z} \qquad \text{(A1)}$$

where $A_{\rm disc}^{\rm inner}$ and $A_{\rm disc}^{\rm outer}$ are amplitudes of mass density ($M_\odot\,{\rm kpc}^{-3}$), $R_{\rm disc}^{\rm inner}$ and $R_{\rm disc}^{\rm outer}$ are disc scale radii (kpc), and h_z is the disc scale height (kpc). Given that satellite orbits are more sensitive to the enclosed *mass*, rather than the local density, we first integrate out the vertical component of equation (A1), then we integrate over cylindrical R to obtain the disc enclosed mass given by

$$M(<R) = 4\pi A h_z R_{\rm inner}\left[R_{\rm inner} - e^{-R/R_{\rm inner}}(R_{\rm inner} + R)\right], \qquad \text{(A2)}$$

and a similar term for the outer disc component. We thus ensure that the total enclosed mass agrees to within a few per cent at all relevant radii when integrated over all space. We fit and model the disc with star particles and cold gas ($T < 10^5\,$K) within $R = 0$–$20\,$kpc and $|Z| < 3\,$kpc. Including mass within $|Z| < 4, 5$ did not significantly change the derived parameters.

The orbits of typical satellites are not sensitive to the details of the size, geometry, or orientation of the disc, see Appendix B.

Table A1. Best-fitting parameters to the double-exponential disc profile (equations A1 and A2) and the generalized NFW profile for the halo (equations A3 and A4). Columns: Name of host; halo amplitude, A_{halo}; halo scale radius, a_{halo}; halo inner slope, α; halo outer slope, β; disc inner amplitude, A_{disc}^{inner}; disc inner scale length, R_{inner}; disc outer amplitude, A_{disc}^{outer}; disc outer scale length, R_{outer}; disc scale height, h_z.

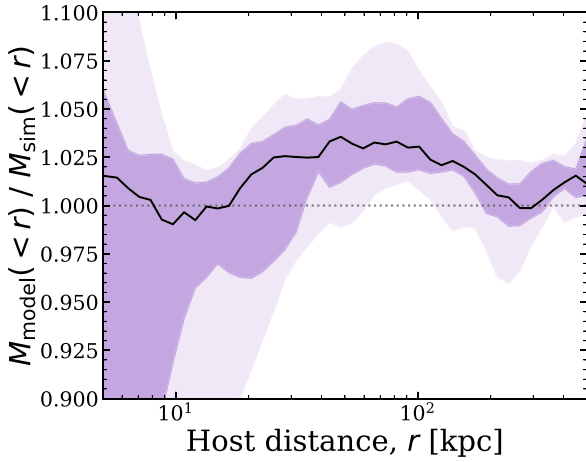| Name	A_{halo} $(10^{11}\,M_\odot)$	a_{halo} (kpc)	α	β	A_{disc}^{inner} $(10^9\,M_\odot\,kpc^{-3})$	R_{disc}^{inner} (kpc)	A_{disc}^{outer} $(10^8\,M_\odot\,kpc^{-3})$	R_{disc}^{outer} (kpc)	h_z (kpc)
m12m	3.78	17.32	1.57	2.78	6.47	0.79	7.98	4.41	0.64
Romulus	5.30	6.65	0.00	2.87	8.31	0.86	2.10	8.04	0.55
m12b	3.84	16.89	1.48	2.82	18.0	0.65	6.08	4.24	0.51
m12f	3.65	14.77	1.45	2.74	8.39	0.84	2.11	7.40	0.54
Thelma	4.49	8.84	0.40	2.90	3.30	1.04	1.98	6.28	0.75
Romeo	8.99	29.91	1.47	3.23	4.61	1.02	2.25	6.64	0.55
m12i	5.70	32.15	1.58	2.99	9.10	0.78	3.87	4.56	0.55
m12c	7.93	24.37	0.90	3.05	7.05	0.71	6.41	3.65	0.57
m12w	2.99	22.87	1.61	2.73	5.53	0.68	11.7	2.19	0.67
Remus	8.32	36.49	1.56	3.21	4.68	0.90	1.60	6.50	0.54
Juliet	2.85	9.57	0.87	2.82	6.97	0.77	1.16	6.32	0.55
Louise	2.30	6.41	0.29	2.76	1.92	0.99	0.54	9.36	0.58
m12z	3.22	33.96	1.58	2.69	0.59	0.40	1.73	2.94	1.54

Figure A1. Ratio of the enclosed total mass in each best-fitting model to that in each simulation, as a function of distance from the centre of the MW-mass host, r, at $z = 0$. We fit a double-exponential profile for the inner and outer discs (equations A1 and A2) and a generalized NFW profile for the halo (equations A3 and A4). Table A1 lists all fit parameters. Black line shows the median and the shaded regions show the 68th and 95th percentile scatter across our 13 hosts. The median agrees to within <4 per cent at all radii and ≲1 per cent for the mass within R_{200m} (344–472 kpc). The ratio shows larger host-to-host scatter at ≲ 10 kpc, but only ≈5 per cent of satellites in our sample orbit this close. This level of agreement ensures that modelling the total (axisymmetric) mass profile of each host at $z = 0$ is not a significant source of error for our orbit modelling.

A2 Modelling the halo

To fit the halo component, we use a generalized form of the spherical Navarro–Frenk–White (NFW, Navarro, Frenk & White 1996) density profile:

$$\rho(r) = \frac{A_{halo}}{4\pi a_{halo}^3} \frac{1}{(r/a_{halo})^\alpha (1 + r/a_{halo})^{\beta-\alpha}} \quad (A3)$$

where A_{halo} is the amplitude ($M_\odot$), a_{halo} is the scale radius (kpc), and α and β are the slopes of the inner and outer density profile. We then integrate the analytic form of $\rho(r)$ to convert this to an enclosed mass profile:

$$M(<r) = \frac{A_{halo}}{3-\alpha} \left(\frac{r}{a}\right)^{3-\alpha} {}_2F_1(3-\alpha, -\alpha+\beta; 4-\alpha; -r/a_{halo})$$
$$(A4)$$

where ${}_2F_1$ is the Gauss hypergeometric function. We model and fit this profile with dark-matter and hot gas ($T > 10^5$ K) within $r < 10$ kpc, and *all* particles (dark matter, all gas, and stars) at $r = 10$–500 kpc. We obtained similar parameters selecting particles within $r < 300, 350, 400$ kpc, but we use the fits out to 500 kpc, because many satellites orbit out to there (see Santistevan et al. 2023). We tried fitting the haloes to a regular NFW profile, where $\alpha = 1$ and $\beta = 3$, but we obtained notably better agreement with the generalized form above.

A3 Total mass profile

We use the parameters we derived from fitting the above analytic density/mass profiles in their respective potentials in GALPY. In GALPY, we specifically use the spherically symmetric 'TwoPowerSphericalPotential' for the DM halo, and axisymmetric 'DoubleExponentialdiscPotential' for the inner and outer discs, and we input all three to define the total potential of a given host, to use to integrate the orbits of its satellites. Table A1 lists the fit parameters.

Fig. A1 shows the ratio of the halo + disc analytic model using the parameters in Table A1 to the enclosed masses in the simulations; a 1:1 ratio is presented as the dotted horizontal line. At all distances we show, we model the median enclosed mass to within 3–4 per cent, and the 68th percentile range is within 5 per cent beyond 10 kpc. Within 10 kpc, the 68th and 95th percentiles of the enclosed mass ratios span both higher and lower than what we present in the y-axis, but only ≈5 per cent of surviving satellites orbited at such small distances.

Individually, we model the two-component disc to within ≈4 per cent at $R = 2$–20 kpc and the halo to within ≈10–20 per cent. Although we double counts the star particles and cold gas between 10 and 20 kpc, the median mass ratio in Fig. A1 is ≲2 per cent within this distance and produces a good fit to the data. One galaxy, m12c, experienced a significant merger (≈3:1 mass ratio) ≈ 9 Gyr ago, thus modelling this host with symmetric analytic models does not capture the complexity in its mass distribution. Thus, the total enclosed mass in m12c is fit to within ≈15 per cent.

Table B1. Similar to Table A1, but now comparing our fiducial model for the host disc to models in which (1) we rotate the disc by 90° or (2) use a simple point mass. We show only the fractional differences for each property X via $(X_{\text{disc, rot}} - X_{\text{disc}})/X_{\text{disc}}$ for the rotated disc model (middle columns), and $(X_{\text{point}} - X_{\text{disc}})/X_{\text{disc}}$ for the point mass model (right columns). Column list: property name, variable, median offset, 1σ scatter. We do not show the minimum pericentre properties given that the orbits in the models are periodic. Values with an asterisk at the end represent cases where we present the mean offset or standard deviation, instead of median and 1σ scatter, because they are discrete quantities or have medians of 0. The overall bias in any orbit property for both alternative models is less than 1 per cent, and the 1σ uncertainties only reach as large as ≈ 7 per cent for the number of pericentric passages. *Thus, the orientation and geometry of the central disc are not important factors in orbit modelling uncertainties.*

Property	Variable	Rotated disc Median offset	1σ scatter	Point mass Median offset	1σ scatter
Recent pericentre distance	$d_{\text{peri, rec}}$	2×10^{-5}	2×10^{-3}	5×10^{-3}	7×10^{-3}
Timing of recent pericentre	$t_{\text{peri, rec}}$	1×10^{-9}	1×10^{-3}	3×10^{-3}	7×10^{-3}
Number of pericentric passages	N_{peri}	$2 \times 10^{-3*}$	$5 \times 10^{-2*}$	$-6 \times 10^{-3*}$	$4 \times 10^{-2*}$
Pericentric passages post-infall at $R_{\text{200m, 0}}$	$N_{\text{peri, fixed}}$	$2 \times 10^{-3*}$	$5 \times 10^{-2*}$	$-8 \times 10^{-3*}$	$7 \times 10^{-2*}$
Velocity at recent pericentre	$v_{\text{peri, rec}}$	-3×10^{-5}	2×10^{-3}	-5×10^{-3}	7×10^{-3}
Recent apocentre distance	$d_{\text{apo, rec}}$	4×10^{-7}	4×10^{-4}	1×10^{-3}	4×10^{-3}
Lookback time of infall	t_{inf}	$-2 \times 10^{-3*}$	1.5×10^{-3}	3×10^{-3}	5×10^{-3}
Lookback time of infall at $R_{\text{200m, 0}}$	$t_{\text{inf, fixed}}$	-1×10^{-7}	2.5×10^{-3}	-2×10^{-8}	3×10^{-3}
Recent orbit eccentricity	e_{rec}	-4×10^{-5}	2×10^{-3}	-1×10^{-3}	5×10^{-3}
Recent orbit period	T_{rec}	3×10^{-8}	2×10^{-3}	7×10^{-3}	1×10^{-2}
Max strength of tidal force	$\lvert da/dr \rvert$	$1 \times 10^{-3*}$	$3 \times 10^{-2*}$	-2×10^{-3}	2×10^{-2}

APPENDIX B: ALTERNATIVE MODELS FOR THE HOST GALAXY POTENTIAL: ROTATED DISC AND POINT MASS

Here, we investigate the results of using different potential models for the central galaxy, including (1) a disc that is rotated by 90° and (2) a point mass. We compare these models with the fiducial disc model in Table A1. We numerically integrate the satellites in these alternative models and calculate the same orbit properties in Table 2. For each property X, we present only the fractional differences between the models, $(X_{\text{disc, rot}} - X_{\text{disc}})/X_{\text{disc}}$ and $(X_{\text{point}} - X_{\text{disc}})/X_{\text{disc}}$, where $X_{\text{disc, rot}}$ refers to the fiducial disc model rotated by 90°, X_{point} refers to the point mass model, and X_{disc} is our fiducial model. Table B1 shows the median offset and width of the 68th percentile. For cases in which the median is zero, we instead show the mean offset and standard deviation.

For the rotated disc model, the median fractional difference across all orbit properties is $\approx 10^{-3}$ or smaller, and the 1σ scatters are less than 5 per cent. We also calculate the median fractional difference only for satellites that orbit closest to the disc, with pericentres < 50 kpc, and the offsets are still $\lesssim 10^{-3}$ per cent. *Thus, we conclude that the orientation of the host galaxy isk is not important for orbit modelling.*

To implement a point mass potential, we use GALPY's 'KeplerPotential', which takes only the mass as the parameter. We integrate the enclosed galaxy mass profile in equation (A2) out to 30 kpc and input this into the point mass potential. Table B1 shows that, even for this simplest-possible model for the central galaxy mass distribution, the median offsets and percentile widths are $\lesssim 7$ per cent. *This reinforces that the details of the mass distribution of the central galaxy are not important; what matters only is modelling the total baryonic mass of the central galaxy.*

We also tried 'shrinking' our fiducial disc model, by reducing the scale radii and scale height to 90, 50, 10, and 1 per cent of their best-fitting values in Table A1, which showed results intermediate between the rotated disc and point mass models.

These results are consistent with the tests of Garrison-Kimmel et al. (2017), who compared the surviving subhalo populations in two of the FIRE-2 MW-mass hosts against DMO simulations with an embedded disc potential. They showed that the number of surviving subhaloes in the simulations with an embedded disc potential agrees well with those in the baryonic FIRE-2 simulations, both of which are much smaller than in a DMO simulation without an embedded disc. They additionally tested various embedded disc potentials by doubling the scale length, fixing the scale height to 1 pc, doubling the total disc mass, including the gas mass to the disc, and implementing a Hernquist sphere instead of their Miyamoto-Nagai disc potential. Most of these alternate disc potentials led to similar results, reinforcing that the details of the shape of the central galaxy potential are less important than simply its overall mass.

APPENDIX C: COMPARING DISCRETE PERICENTRES

To understand better how well orbit modelling and cosmological simulation agree *at each orbit*, Fig. C1 compares the timing and distance of each pericentric event. We select all satellites at a given 'lookback pericentre event', where 1 is the most recent pericentre, 2 is the second-most recent, and so forth. We plot the fractional and raw differences in the pericentre distance and timing.

In our sample, 86 per cent of satellites experienced at least one pericentre, ≈ 47 per cent experienced two or more, and ≈ 30 per cent experienced three or more. The most pericentres experienced is 10, but we cut the figure at 6 pericentres given that only 3.4 per cent experienced more than this.

Focusing on trends with distance, the median fractional difference in the most recent pericentre is only -2.5 per cent, similar to Fig. 9 (top row). As we compare pericentres further back in time, the median decreases to -18 per cent for satellites at their sixth most recent pericentre. However, this is not to say that the model *never* overpredicts the pericentre distances. For the most recent pericentric passage, the model underpredicts the distance in ≈ 60 per cent of the sample, and overpredicts the distance in the other 40 per cent.

Likewise, the 1σ scatter generally increases with increasing lookback pericentre events, where it reaches a max value of ≈ 48 per cent

at the 5th-most recent pericentre. Thus, the median difference between the model and simulation increases with each orbit as we look back in time, but also the uncertainties associated with these pericentre distances increases.

Similar to trends in distance, Fig. C1 (bottom) shows that the median difference in the timing of each pericentre decreases from -30 Myr at the most recent pericentre to -1.2 Gyr for the 6th most recent. Orbit modelling continuously underpredicts pericentre events, presumably because the host mass is unchanging as the satellites orbit. For a given satellite, the more massive host with a deeper gravitational potential will result in a more bound orbit compared to the same satellite in the same MW-mass host with less mass at earlier times. More bound satellites thus orbit deeper in the potential with shorter orbit timescales. Regarding the most recent pericentric passage, the model underpredicts the timing in ≈ 66 per cent of satellites and overpredicts the timing in 30 per cent; the remaining 4 per cent of satellites have nearly identical values. The 1σ scatter spans $\gtrsim 0.5$ Gyr for the second-most recent pericentre and beyond.

These results highlight another aspect of how far back in time orbit modelling reliably works. If one is only interested in deriving the most recent pericentre distance or time, the model recovers the median trends of satellites to within a few per cent, but the uncertainties are non-negligible. For the satellites that orbited more than once, the most recent pericentre is not always the smallest (see Santistevan et al. 2023), therefore if one is interested in how close satellites orbited to their host, then the uncertainty in deriving these distances will be $\gtrsim 35$ per cent, and $\gtrsim 0.5$ Gyr in the timing.
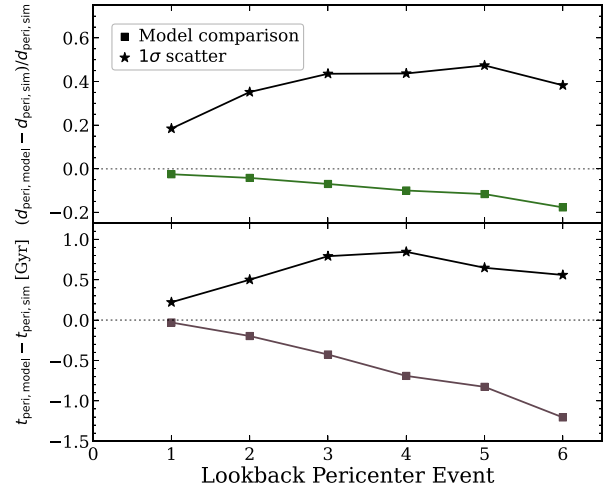

Figure C1. Comparing the timing and distance of each pericentric passage between orbit modelling and simulations, which we measure at each 'pericentre event', where 1 is the most recent pericentre, 2 is the second most recent, and so forth. We select all satellites that experienced (at least) a given number of pericentres, and we show the median and 1σ scatter across all satellites. Top: The median fractional difference in pericentre distance decreases from -2.5 to -18 per cent going back 6 pericentres while the 1σ scatter increases from 18 to 48 per cent. Less than 3 per cent of satellites experienced more than 7 pericentres, in both the simulations and orbit models. Bottom: The median difference in the timing of the pericentric events decreases with lookback pericentre events from -30 Myr for the most recent pericentre to -1.2 Gyr. The 1σ scatter ranges from ≈ 0.2 to 0.85 Gyr, so not only does orbit modelling increasingly systematically underpredict the timing of a pericentre, but also, orbit modelling becomes more uncertain, over longer lookback times.

This paper has been typeset from a TEX/LATEX file prepared by the author.