Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **510,** 2980–2997 (2022) https://doi.org/10.1093/mnras/stab3587
Advance Access publication 2021 December 21

Galaxy velocity bias in cosmological simulations: towards per cent-level calibration

Dhayaa Anbajagane [1,2★] Han Aung [3] August E. Evrard [2,4] Arya Farahi [5,6] Daisuke Nagai [3,7] David J. Barnes [8] Weiguang Cui [9] Klaus Dolag,[10,11] Ian G. McCarthy [12] Elena Rasia[13,14] and Gustavo Yepes [15]

[1]*Department of Astronomy and Astrophysics, University of Chicago, 5640 S. Ellis Ave, Chicago, IL 60637, USA*
[2]*Department of Physics and Leinweber Center for Theoretical Physics, University of Michigan, Ann Arbor, MI 48109, USA*
[3]*Department of Physics, Yale University, New Haven, CT 06520, USA*
[4]*Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA*
[5]*Michigan Institute for Data Science, University of Michigan, Ann Arbor, MI 48109, USA*
[6]*Department of Statistics and Data Science, The University of Texas at Austin, TX 78712, USA*
[7]*Department of Astronomy, Yale University, New Haven, CT 06520, USA*
[8]*Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[9]*Institute for Astronomy, University of Edinburgh, Edinburgh EH9 3HJ, UK*
[10]*Max-Planck Institut für Astrophysik, Karl-Schwarzschild Str 1, D-85741 Garching, Germany*
[11]*Faculty of Physics, University Observatory Munich, Scheinerstr 1, D-81679 München, Germany*
[12]*Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK*
[13]*INAF, Osservatorio Astronomico di Trieste, via Tiepolo 11, I-34131 Trieste, Italy*
[14]*IFPU – Institute for Fundamental Physics of the Universe, Via Beirut 2, I-34014 Trieste, Italy*
[15]*Departamento de Física Teórica M-8 and CIAFF, Facultad de Ciencias, Universidad Autónoma de Madrid, E-28049 Madrid, Spain*

Accepted 2021 December 6. Received 2021 December 3; in original form 2021 October 9

ABSTRACT

Galaxy cluster masses, rich with cosmological information, can be estimated from internal dark matter (DM) velocity dispersions, which in turn can be observationally inferred from satellite galaxy velocities. However, galaxies are biased tracers of the DM, and the bias can vary over host halo and galaxy properties as well as time. We precisely calibrate the velocity bias, b_v – defined as the ratio of galaxy and DM velocity dispersions – as a function of redshift, host halo mass, and galaxy stellar mass threshold ($M_{\star,\rm sat}$), for massive haloes ($M_{\rm 200c} > 10^{13.5}\,{\rm M}_\odot$) from five cosmological simulations: IllustrisTNG, Magneticum, Bahamas + Macsis, The Three Hundred Project, and MultiDark Planck-2. We first compare scaling relations for galaxy and DM velocity dispersion across simulations; the former is estimated using a new ensemble velocity likelihood method that is unbiased for low galaxy counts per halo, while the latter uses a local linear regression. The simulations show consistent trends of b_v increasing with $M_{\rm 200c}$ and decreasing with redshift and $M_{\star,\rm sat}$. The ensemble-estimated theoretical uncertainty in b_v is 2–3 per cent, but becomes *percent-level* when considering only the three highest resolution simulations. We update the mass–richness normalization for an SDSS redMaPPer cluster sample, and find our improved b_v estimates reduce the normalization uncertainty from 22 to 8 per cent, demonstrating that dynamical mass estimation is competitive with weak lensing mass estimation. We discuss necessary steps for further improving this precision. Our estimates for $b_v(M_{\rm 200c}, M_{\star,\rm sat}, z)$ are made publicly available.

Key words: methods: statistical – galaxies: haloes – galaxies: kinematics and dynamics.

1 INTRODUCTION

Galaxy clusters, and their associated massive dark matter (DM) haloes, contain rich information on the composition and evolutionary history of our Universe. This information can be extracted by connecting the observable properties of clusters to the halo mass function (HMF) – a number density of haloes as a function of halo mass – which then translates into constraints on cosmological parameters. A key component of this inference process is constructing a probabilistic mapping function between cluster observables and the mass of the underlying massive halo. Observable properties with established and understood connections are commonly referred to as halo mass proxies, and there exist many across multiple wavelengths (see Allen, Evrard & Mantz 2011; Pratt et al. 2019, for reviews). In this work, we focus on connecting satellite galaxy kinematics, measured via spectroscopy, to the host halo mass.

The collisionless DM velocity field within a halo is driven by the halo's gravitational potential. When virial equilibrium is satisfied, meaning the average kinetic energy is half the magnitude of the potential energy, an estimate for the total halo mass can be obtained by inferring the average kinetic energy from the satellite galaxy kinematics of a halo. Zwicky (1937) famously used this approach to estimate a halo mass and argue for the existence of dark matter.

★ E-mail: dhayaa@uchicago.edu

Note that this technique of halo mass estimation implicitly assumes that the satellite galaxies fairly trace the DM velocity field. Galaxies, however, are known to be biased tracers of the underlying DM density field (Kaiser 1984; Davis et al. 1985; Bardeen et al. 1986) and of the DM velocity field, as shown by the many works that we detail below. The bias pertaining to the latter case is commonly denoted the velocity bias, b_v, and the focus of this work is the mean b_v of the halo population as a function of halo and galaxy properties.

The velocity bias is linked closely to the physics of galaxy formation, and in particular to when galaxies fall into a DM halo and become subject to dynamical friction, tidal disruption, and other non-linear effects (Carlberg, Couchman & Thomas 1990; Carlberg 1991; Colafrancesco, Antonuccio-Delogu & Del Popolo 1995). Notably, the uncertainty in this bias is the dominant systematic uncertainty in dynamical mass estimation techniques (henceforth F16; Farahi et al. 2016; Sifón et al. 2016); the study of F16 showed that the precision in halo mass is limited to ≈25 per cent due to poor knowledge of b_v.

Previous works have studied the qualitative and/or quantitive trends of the satellite galaxy velocity bias as a function of galaxy properties (such as stellar mass, galaxy luminosity, and redshift) using simulations (Biviano et al. 2006; Lau, Nagai & Kravtsov 2010; Munari et al. 2013; Old, Gray & Pearce 2013; Wu et al. 2013; Ye et al. 2017; Armitage et al. 2018; Ferragamo et al. 2020) as well as observational data (Biviano et al. 1992; Stein 1997; Adami, Biviano & Mazure 1998; Adami et al. 2000; Girardi et al. 2003; Goto 2005; Barsanti et al. 2016; Bayliss et al. 2017; Nascimento, Ribeiro & Lopes 2017). The results of these works are all consistent with brighter galaxies being kinematically cooler – and thus having a *lower* velocity bias – than fainter galaxies.

However, Guo et al. (2015c, hereafter G15), who used measurements of the small-scale redshift-space distortions (RSD) to infer a velocity bias as a function of galaxy luminosity, find that brighter galaxies have a *higher* (not lower) velocity bias than fainter galaxies. Their estimates are thus in tension with the aforementioned studies. For example, Bayliss et al. (2017) use observed spectra from nearly 3000 satellite galaxies in 89 clusters identified via the Sunyaev–Zel'dovich effect, and find that the velocity dispersion – which is the second moment of a velocity field – for brighter galaxies is 11 ± 4 percent lower than the velocity dispersion for the full galaxy population. It is speculated that the discrepancy between G15 and the other works arises because G15 uses all galaxies in the survey volume, and not just satellite galaxies hosted in massive haloes (Ye et al. 2017, see conclusions). Another difference is that G15 use both the one-halo *and* two-halo components of the velocity fields in their RSD analysis, whereas all the cluster-focused studies mentioned above limit themselves to the one-halo component alone.

Given the discrepancy in the G15 result, we require an alternative calibration for the satellite galaxy velocity bias as a function of relevant galaxy/halo properties. The other observational works noted above have studied the *relative* trends of the velocity bias with galaxy luminosity but did not estimate the actual values of b_v. Some simulation studies have estimated both b_v and its dependence on galaxy luminosity, but using alternative methodologies to that used in our work: Lau et al. (2010), Wu et al. (2013) selected the top N galaxies per halo according to $M_{\star,\rm sat}$ and studied the response of the velocity bias to varying N, and Ferragamo et al. (2020) selected the top N per cent of all satellite galaxies in cluster-scale haloes. While these works have shed light on the velocity bias of galaxies within host haloes, they do not provide a function or mapping for the velocity bias given a galaxy stellar mass threshold or galaxy magnitude threshold.

The two simulation-based works that *have* estimated this mapping (Ye et al. 2017; Armitage et al. 2018) are limited in either using a small sample size of only the most massive haloes, or using a single simulation model. The former leads to larger statistical uncertainties in the bias estimates, in addition to being limited to a narrow mass range, whereas the latter cannot quantify the theoretical uncertainty in the velocity bias, i.e. the variation in the velocity bias due to different astrophysical and numerical treatments.

In this work, we use an ensemble of simulations to calibrate the satellite galaxy velocity bias, *including* the relevant theoretical uncertainty, as a function of the galaxy stellar mass threshold, host halo mass, and redshift. We extend on the previous body of work in three different directions: (i) We propose and validate a new likelihood-based estimator for the scaling parameters (normalization, slope, and population intrinsic scatter) of the galaxy velocity dispersion with halo mass in the regime of low-galaxy counts per halo, (ii) We perform a convergence study of the velocity bias, as well as galaxy and DM velocity dispersions, across a suite of cosmological, hydrodynamics simulations, and also an N-body simulation with galaxies painted on using a semianalytical model, and; (iii) Finally, using predictions from the ensemble of simulations, we construct a theoretical prior on the velocity bias that incorporates the *modelling uncertainty* associated with varying numerical and galaxy formation treatments. We then use this prior to refine the mean halo mass estimates previously derived in F16 for SDSS redMaPPer galaxy clusters.

A part of our convergence study focuses on the mass-dependent population statistics – mean and intrinsic scatter – of DM velocity dispersion, and is thus a hydrodynamical counterpart to the original study of Evrard et al. (2008, henceforth E08), who used a large ensemble of mostly N-body simulations to set precise constraints on these quantities. Note also that we previously employed a subset of the simulations used in this work to perform similar convergence tests of mass-dependent population statistics for central and satellite galaxy properties of cluster-scale haloes; Anbajagane et al. (2020) find that the distribution of residuals about the property mean relations share similar functional forms, and that the mean relations themselves have moderate offsets between simulations.

This paper is organized as follows: in Section 2 we describe our simulation ensemble, and in Section 3 we define the relevant halo properties and detail the scaling relation estimators, including the aforementioned ensemble velocity likelihood method. Our results for the galaxy/DM velocity dispersion and the velocity bias are presented in Section 4, while the impact of our work for dynamical mass estimation is both demonstrated and discussed in Section 5. Finally, we conclude in Section 6. Our appendices contain results on resolution effects (Appendix A), additional validation tests of the likelihood model (Appendix B), and the impact of radial aperture choices on the velocity bias (Appendix C).

Throughout this work we use a spherical overdensity definition of halo mass, $M_\Delta = \rho_\Delta[\frac{4\pi}{3}R_\Delta^3]$, with contrast value, $\rho_\Delta = 200\rho_c(z)$, where $\rho_c(z)$ is the critical density at redshift z.

2 DATA

We use samples of haloes realized in the following five simulations: (i) IllustrisTNG, (ii) Magneticum Pathfinder, (iii) a superset of Bahamas and Macsis, (iv) MultiDark Planck 2, and (v) The Three Hundred Project. Key properties of the simulations are summarized in Table 1, and a brief description of each follows. The host halo and satellite galaxy samples of each simulation are also compared in Fig. 1.

Table 1. Simulation characteristics for the $z = 0$ halo samples. Table adapted from Anbajagane et al. (2020) with modifications. Note that for MGTM alone we do not have a $z = 0$ data set and thus characterize the $z = 0.06$ output instead. From left to right, we show: (i) simulation acronym as used in this work, (ii) comoving box size, (iii) cosmic matter density parameter at the present epoch, (iv) Hubble constant, (v) force softening scale, (vi) initial mass of stellar particles, (vii) mass of DM particles, (viii) number of haloes with $M_{200c} > 10^{13.5}\,M_\odot$, (ix) number of satellite galaxies within $r < R_{200c}$ of host haloes with $M_{200c} > 10^{13.5}\,M_\odot$ and above the minimum stellar mass threshold used for each simulation, and (x) empirical sources used for tuning sub-grid parameters of each simulation, which consist of the Galaxy Stellar Mass Function (GSMF), supermassive black hole scaling (SMBH), metallicity scaling (metals), and cluster hot gas mass fraction $<R_{500c}$ (CL $f_{\rm gas}$). All simulations assume a flat ΛCDM cosmology, with $\Omega_\Lambda = 1 - \Omega_m$. See text for references. MDPL2 is an *N*-body, DMO simulation, and does not have stellar particles, and The300 consists of zoom-in (re)simulations of the 324 most massive haloes drawn from MDPL2.

Simulation	L [Mpc]	Ω_m	H_0 [$\frac{\rm km\,s^{-1}}{\rm Mpc}$]	$\epsilon_{\rm DM}^{z=0}$ [kpc]	$m_\star$ [$M_\odot$]	$m_{\rm DM}$ [$M_\odot$]	$N_{\rm haloes}$	$N_{\rm sat,tot}$	Calibration
TNG300	303	0.3089	67.74	1.48	1.1×10^7	5.9×10^7	1146	40436	See Pillepich et al. (2018a)
MGTM	500	0.2726	70.40	5.33	5.0×10^7	9.8×10^8	4207	90255	SMBH, Metals, CL $f_{\rm gas}$
BM	596	0.3175	67.11	5.96	1.2×10^9	6.6×10^9	9430	132334	GSMF, CL $f_{\rm gas}$
MDPL2	1475	0.3071	67.77	7.4	–	2.2×10^9	157051	2339642	See Behroozi et al. (2019)
The300	–	0.3071	67.77	9.6	3.5×10^8	1.9×10^9	3180	156662	See Cui et al. (2018)

The ILLUSTRISTNG project (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018; Springel et al. 2018; Pillepich et al. 2018b) is a follow up to the ILLUSTRIS project (Vogelsberger et al. 2014). It is run with the moving mesh code AREPO (Springel 2010), and includes a full magnetohydrodynamics treatment with galaxy formation models, as detailed in Weinberger et al. (2017) and Pillepich et al. (2018a). We use the highest resolution run from the TNG300 suite for our main analysis, but also utilize the lower resolution runs in an appendix to perform numerical resolution tests. Haloes are identified via a friends-of-friends (FoF) algorithm, and subhaloes via the SUBFIND algorithm (Springel et al. 2001; Dolag et al. 2009) which applies a binding energy condition to link particles to substructure. Our galaxy catalogues and host halo properties are obtained/computed from the public data release[1] (Nelson et al. 2019).

MAGNETICUM PATHFINDER (MGTM; Hirschmann et al. 2014) is a suite of magnetohydrodynamics simulations run using the smoothed particle hydrodynamics (SPH) solver GADGET-3 (last described in Springel 2005). Haloes are identified using a FoF algorithm, and subhaloes are identified using SUBFIND. We make use of the box 2 h run for this work, and the corresponding galaxy catalogues are obtained from the public data base[2] (Ragagnin et al. 2017). Note that we do not have the $z = 0$ catalogue for this simulation, and instead use the $z = 0.06$ catalogue in its place. Results for other redshifts use the correct catalogues. MGTM also has the most different cosmology to all other simulations in the ensemble (see Table 1); it is based on a WMAP7 cosmology (Komatsu et al. 2011) whereas all other runs have used Planck cosmologies (Planck Collaboration XVI XIII 2014, 2016).

BAHAMAS (McCarthy et al. 2017) and its zoom-in companion MACSIS (Barnes et al. 2017) – which we collectively denote with the acronym 'BM' – are hydrodynamics simulations run using a version of GADGET-3 developed independently of the MGTM version. The MACSIS ensemble contains 390 haloes, with each halo first drawn from a parent 3.2 Gpc *N*-body, dark-matter only (DMO) simulation and then re-simulated in individual, separate volumes with a full hydrodynamics prescription aligned with the BAHAMAS treatment. MACSIS extends the high-mass end of BM sample to $M_{200c} \approx 4 \times 10^{15}\,M_\odot$ at $z = 0$. Haloes are once again identified via the FoF algorithm and substructure is identified via SUBFIND.

MULTIDARK PLANCK 2 (MDPL2) is part of the MultiDark suite (Klypin et al. 2016) and is an *N*-body, DMO simulation run using the L-GADGET-2 solver – a memory-efficient variant of GADGET optimized for simulations with a large number of particles. Haloes and subhaloes were identified using the ROCKSTAR halo finder (Behroozi, Wechsler & Wu 2013), which identifies structure in the full 6D position-velocity phase-space as opposed to the 3D position-space used by the other halo/subhalo finders mentioned in this work. While MDPL2 is a DMO simulation with no galaxies (and only subhaloes), galaxies can be 'painted' on to the subhaloes using the assembly histories of the latter. Galaxy catalogues for MDPL2 have been generated using a variety of different semianalytical models (SAMs) of which we use public catalogs[3] from the UNIVERSEMACHINE (UM) prescription (Behroozi et al. 2019). The host halo quantities are obtained from the public data base for MDPL2[4] (Riebe et al. 2011). UM, like many other SAMs, artificially adds so-called 'orphan' galaxies – defined as galaxies whose host subhaloes have been tidally destroyed – back to the catalogue so that the two-point galaxy correlation function in the simulated catalogue matches observational constraints. For these artificially added galaxies, which reside preferentially near the halo core, UM can only approximately evolve their velocities (see appendix B of Behroozi et al. 2019), and so the velocity statistics of this orphan galaxy population can be discrepant from the 'truth'.

THE THREE HUNDRED PROJECT (The300; Cui et al. 2018) is a set of 324 massive haloes that were first identified in the MDPL2 simulation, and then re-simulated within spheres of radius 22 (comoving) Mpc with a full hydrodynamics prescription (Rasia et al. 2015) using the GADGET-X SPH solver (Beck et al. 2016). Haloes and subhaloes are identified with Amiga's Halo Finder (AHF; Knollmann & Knebe 2009), which uses an adaptive mesh refinement grid to represent the density field/contours and also has a binding energy criterion similar to that of SUBFIND. While The300 is mass-complete only above $M_{200c} \approx 10^{15}\,M_\odot$ at $z = 0$, it still resolves many haloes at masses below this mass scale. We continue using all haloes above $M_{200c} > 10^{13.5}\,M_\odot$ and demonstrate in Appendix B1 that the selection effects in the mass-incomplete part of the halo sample do not impact the quantities of interest to us.

2.1 Study limitations

In the strongly non-linear regime of ΛCDM structure formation, verification studies of the statistics from different simulations is an important way to assess modelling uncertainties. While our ensemble of simulations is extensive, with a variety of astrophysical treatments and a moderate range in numerical resolution, we note that in all cases the collisionless dark matter component was evolved using some

[1] https://www.tng-project.org/
[2] http://magneticum.org/data.html
[3] https://www.peterbehroozi.com/data.html
[4] https://www.cosmosim.org/

version of GADGET. IllustrisTNG may seem to be an exception, but its solver, AREPO, inherits its *N*-body methodology from GADGET as well. Previous results have shown that solutions from different *N*-body, gravity-only solvers for the matter power spectrum can differ on their estimates of the small scales ($k > 1\,\mathrm{Mpc}^{-1}$) by more than 3 per cent (see fig. 1 of Schneider et al. 2016). Thus, our ideal simulation ensemble would also include cosmological hydrodynamics simulations based on other *N*-body solvers, such as RAMSES (Teyssier 2002) and PKDGRAV (last described in Stadel 2001), but we are not aware of any such simulations that also contain a large enough population of cluster-scale haloes to be used in this work.

Mansfield & Avestruz (2021) also show that the scaling relations of internal DM halo properties realized by different suites of *N*-body simulations do not always converge to the same result – even those that share the same *N*-body solver, but differ in their choice of control parameters such as force softening scales and DM particle mass – with the level of non-convergence varying according to halo property. We stress that the cluster-scale haloes studied in this work are resolved by at least $N > 10^5$ particles, where Mansfield & Avestruz (2021) find that velocity-based properties of interest are strongly converged. Further evidence of this comes from the nIFTy comparison project, who simulated the same cluster-scale halo using different codes and found that the bulk halo properties – such as velocity dispersion and shapes – agree at the per cent-level across codes (Sembolini et al. 2016a,b). Note that while we also study galaxies resolved by as few as $N \sim 100$ DM particles, we do not focus on the *internal* DM distributions and dynamics of these structures (the primary target for such non-convergence issues) and only concern ourselves with their bulk kinematic properties.

How haloes are identified is also a relevant factor. Throughout our work we rely on catalogues that were constructed by running halo/subhalo finders on a simulation's particle data set. Differences between the finders can impact the identification of objects in the simulations and affect the derived population statistics. Previous comparison studies of commonly used finders have found some significant differences (Knebe et al. 2011; Onions et al. 2012). The simulations in our work use a variety of different finders – three simulations use SUBFIND, one uses AHF and one uses ROCKSTAR. Thus, some part of the simulation-to-simulation variance in population statistics will also come from differences across the finders.

3 METHODS

We employ two different methods to measure the virial (or velocity dispersion) scaling relation for DM and galaxies, respectively. The DM virial scaling (Section 3.1) is obtained by first measuring the DM velocity dispersion for individual haloes and then summarizing the mass-dependent statistics of the population using KLLR, a local linear regression method described further in Section 3.1.1. For the galaxy virial scaling (Section 3.2), on the other hand, the sparseness of the satellite galaxy counts motivates us to employ an ensemble likelihood estimator, described further in Section 3.2.1, that circumvents the need to measure the galaxy velocity dispersion for individual haloes. In Section 3.3, we sub-sample DM particles and verify that the likelihood method returns scaling parameters consistent with the KLLR estimates. The reader wishing to skip the technical details of the measurement can go directly to the definition of velocity bias in Section 3.4.

3.1 Dark matter virial scaling

By convention (Yahil & Vidal 1977), the DM velocity dispersion, $\sigma_{\rm DM}$, of the host halo is defined as the average of the dispersion along the three Cartesian components

$$\sigma^2_{\rm DM} = \frac{1}{3(N_{\rm part}-1)} \sum_{i=1}^{N_{\rm part}} \sum_{j=1}^{3} (v_{ij} - \langle v_j \rangle)^2. \quad (1)$$

Here, $N_{\rm part}$ is the number of DM particles within $R_{\rm 200c}$ of the halo centre, v_{ij} is the velocity of particle i along the $j^{\rm th}$ Cartesian component, and $\langle v_j \rangle$ is the mean velocity of all $N_{\rm part}$ DM particles along that same component. All velocities here are peculiar velocities in the proper frame.

Since the haloes are well-resolved, containing $N \gtrsim 10^4$ DM particles, we measure the DM velocity dispersion for each individual halo and use a localized linear regression approach, described further below, to infer the $\sigma_{\rm DM}$–$M_{\rm 200c}$ scaling relation. The study of E08 showed that scaling DM velocity dispersion with $h(z)M_{\rm 200c}$ (instead of $M_{\rm 200c}$) leads to a more universal relation across different cosmologies and redshifts, and so we employ this effective mass as our independent variable. Here, $h(z) = H(z)/100$ is the dimensionless Hubble parameter. The utility of this $h(z)$ factor in capturing the cosmological dependence of the DM virial relation was recently confirmed by Singh et al. (2020, see their section 4.6).

3.1.1 *Kernel-Localized Linear Regression (KLLR)*

In this work, we employ KLLR[5] (Farahi, Anbajagane & Evrard, in preparation; Farahi et al. 2018a), a localized linear regression method, to derive the scale-sensitive estimates of the mean, slope, and intrinsic scatter of the DM virial relation. The KLLR method is particularly useful for estimating scaling relations in simulations that include baryonic processes. Recently, Anbajagane, Evrard & Farahi (2022) used KLLR to study the population statistics of key DM halo properties – including $\sigma_{\rm DM}$ – across six decades in halo mass and showed that all relations have clear mass-dependent features that originate from galaxy formation processes.

In brief, KLLR applies a weight to the halo population – where the weight is given by a Gaussian kernel in log-mass, $\log_{10}M$, with variance $\sigma^2_{\rm KLLR}$ – and then performs linear regression on the weighted data set. Systematically shifting the center of the kernel provides mass-dependent estimates of the fit parameters. The kernel width is a free parameter and for this work we choose $\sigma_{\rm KLLR} = 0.3$ dex. In general, wider kernels wash out small-scale features while smaller kernels increase the noise of the estimates. Our choice here is optimized to reduce the uncertainty of the estimates while still capturing the relevant parameter evolution with halo mass.

Given that we use a Gaussian kernel, the KLLR estimates at a mass-scale M are still informed by haloes with $M_{\rm 200c} < M$. So, KLLR parameters at the mass threshold of our analysis, $10^{13.5}\,\mathrm{M}_\odot$, must be estimated from a sample that extends sufficiently below this value. We therefore include haloes with masses, $M_{\rm 200c} \geq 10^{13}\,\mathrm{M}_\odot$, which covers the lower $\approx 2\sigma_{\rm KLLR}$ tail of a Gaussian kernel centered on $10^{13.5}\,\mathrm{M}_\odot$. We have confirmed that this choice leads to a negligible edge-effect bias of <0.1 per cent in the expectation value of $\sigma_{\rm DM}$.

3.2 Galaxy virial scaling

The quantity available to spectroscopic measurements is the line-of-sight velocity of galaxies within a cluster. Since the cosmic distances to clusters are typically much larger than their sizes, the radial component of the velocities is close to a simple Cartesian projection.

[5] https://github.com/afarahi/kllr

Measured along a single projection direction, the quantity of interest is the peculiar velocity difference,

$$\Delta v_{ik} = v_i - v_{\mathrm{halo},k}, \tag{2}$$

of a galaxy i lying within R_{200c} of the centre of host halo k. Here, $v_{\mathrm{halo},k}$ is the host halo k's mean velocity along the chosen projection direction, computed as the mass-weighted mean velocity of all matter components (DM and all phases of baryons) within R_{200c}.[6] Note that this halo reference velocity can differ very slightly from the reference used previously for DM, as the latter is computed using only DM particles whereas here we include the baryonic component as well. However, the characteristic scale of this difference – estimated using TNG300 – is small, at $\approx$15 km s^{-1}.

Again, we do not include the effect of the Hubble flow in the velocities; its inclusion changes the scaling relation parameters by $\approx$0.1 per cent and is thus negligible in our analysis. We use all three orthogonal Cartesian axes to triple the number of independent measurements for each host halo. Thus, the effective halo sample size per simulation in analyses of $\sigma_{\mathrm{sat,1D}}$ is three times that shown in Table 1.

3.2.1 *Ensemble Velocity Likelihood (EVL)*

Given the sparseness of satellite galaxy counts per halo, particularly at high galaxy stellar mass thresholds, we do not estimate galaxy velocity dispersion for individual haloes like we do for DM. This is because the standard deviation is a biased estimator of $\sigma_{\mathrm{sat,1D}}$ when the galaxy count per halo is low, which can be a common occurrence in our analyses. Alternative estimators, such as GAPPER and bi-weight, can provide a more accurate estimate of the dispersion (Beers, Flynn & Gebhardt 1990), with the level of accuracy varying across estimators (Ferragamo et al. 2020). However, one must also make robust estimates of the uncertainty in each measurement in order to properly infer scaling parameters via regression. This can be a particularly difficult task when the halo is sampled by only two or three galaxies.

To avoid these sparse-sampling complications, we use an extended version of an ensemble likelihood originally developed by Rozo et al. (2015) to assess cluster membership and then used by F16 and Farahi et al. (2018b) for mass estimation. The basis of the method is an aggregate model for the set of satellite galaxy relative velocities, equation (2), conditioned on host halo mass. For fixed host halo mass and redshift, we write an ensemble population likelihood for the collection of 1D galaxy–halo relative velocities. This likelihood is modelled as a convolution of a Gaussian, representing the thermal bath (or velocity distribution) of a single halo, and a lognormal, representing the range of temperatures (or velocity dispersions) at a given halo mass

$$P(\Delta v_{ik} \mid M_k, \boldsymbol{\theta}) = \int P(\Delta v_{ik} \mid \sigma_k) \times P(\sigma_k \mid M_k, \boldsymbol{\theta})\, \mathrm{d}\sigma_k, \tag{3}$$

where index k runs over all host haloes, and index i runs over the satellite galaxies of host halo k. Here, $P(\Delta v_{ik} \mid \sigma_k)$ is a Gaussian distribution with zero mean and variance, $(\sigma_k)^2$. The distribution, $P(\sigma_k \mid M_k, \vec{\theta})$ is modelled as a Gaussian in $\log_{10}\sigma_k$, with a mean $E[\log_{10}\sigma_k \mid M_k]$ – described below in equations (4) and (5) – and a constant, mass-independent variance, ϵ^2.

The vector, $\vec{\theta} = \{\pi,\ \alpha,\ \epsilon\}$, contains the log-mean normalization, slope and intrinsic scatter of the pure power-law scaling relation

$$E_{\mathrm{lin}}[\log_{10}\sigma_{\mathrm{sat,1D}} \mid M,\ \vec{\theta}] = \pi + \alpha \log_{10}\left(\frac{h(z)M}{10^{14}\,\mathrm{M}_\odot}\right), \tag{4}$$

where $\sigma_{\mathrm{sat,1D}}$ is the satellite galaxy velocity dispersion in km s^{-1}, π is the normalization in decimal log, and α is the slope. We choose 10^{14} M$_\odot$ as the pivot mass scale because it lies close to the midpoint of the halo mass ranges spanned by the different simulations.

We have also extended our model to include a quadratic term with a coefficient, β, that captures any 'running' of the power-law slope with host halo mass

$$E_{\mathrm{quad}}[\log_{10}\sigma_{\mathrm{sat,1D}} \mid M, \vec{\theta}] = E_{\mathrm{lin}}[...] + \beta\left[\log_{10}\left(\frac{h(z)M}{10^{14}\,\mathrm{M}_\odot}\right)\right]^2. \tag{5}$$

However, upon constraining the $\sigma_{\mathrm{sat,1D}}$–M_{200c} relation for different halo populations using equation (5), we find that $|\beta| < 0.01$ and that the parameter's 68 per cent confidence intervals always contain $\beta = 0$. Thus, the data show no preference for the quadratic term and so β is not included in the results shown in Section 4.2 nor in the parameter set, $\vec{\theta}$. On the other hand, our validation tests with DM velocity dispersion, presented below, indicate support for $\beta = 0.01 \pm 0.006$ which deviates moderately from $\beta = 0$ at 1.7σ significance.

In principle, one could also add other halo properties to the scaling relation; studies of the IllustrisTNG simulations find that the scatter in σ_{DM} is strongly correlated with secondary DM halo properties such as concentration and shape (Anbajagane et al. 2022) and so it is plausible, though not necessary, that these secondary properties correlate with $\sigma_{\mathrm{sat,1D}}$ as well. Anbajagane et al. (2022) also show that for the halo mass-scales of our study, the intrinsic scatter in σ_{DM} can be reduced by a factor of two when concentration and velocity anisotropy are included in the regression.

Using equations (3) and (4), we can estimate the scaling relation parameters via Bayesian inference. In our model, posteriors for the parameters are obtained after marginalizing/integrating over the distributions of $\sigma_{\mathrm{sat,1D}}$; one distribution per halo as denoted by the integral in equation (3). For halo samples of $N = 10^3$–10^4, this marginalization step leads to a high-dimensional sampling problem and so we employ the Hamiltonian Monte Carlo method to efficiently sample this space of parameters. Our implementation makes use of existing routines provided by the PyMC3 open-source PYTHON package[7] (Salvatier, Wiecki & Fonnesbeck 2015).

3.3 Validation of the EVL estimator

The mean velocity bias of a halo population depends on the $\sigma_{\mathrm{sat,1D}}$–M_{200c} and σ_{DM}–M_{200c} scaling relations, which are estimated using EVL and KLLR, respectively. Here, we use the DM particles in the TNG300 simulation at $z = 0$ to demonstrate consistency in the scaling parameters returned by the two methods. For our baseline results, we use both a simple least-squares linear regression and KLLR. In this case, σ_{DM} is first computed for individual haloes using all available DM particles in each halo and the regression methods are used to estimate the scaling relation. The EVL method we are validating is the same as in equations (3) and (4) but the inputs are now DM particle velocities, not satellite galaxy velocities. We randomly select 100 DM particles from each halo and input the velocities from all three Cartesian directions since we are computing the *isotropically*

[6] For MDPL2, which is a DMO simulation, v_{halo} is computed using only DM particles.

[7] https://docs.pymc.io/

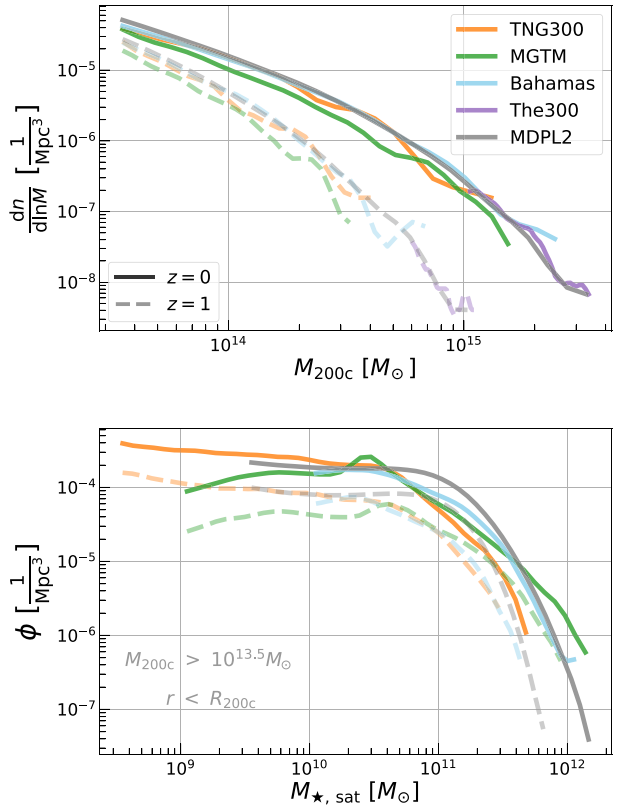

Figure 1. The HMF of host haloes (upper) from the full simulation volume, and the conditional Satellite Galaxy Stellar Mass function (S-GSMF, lower) of all galaxies within R_{200c} of haloes with $M_{200c} > 10^{13.5}\,\mathrm{M}_\odot$, at $z = 0$ (MGTM only is shown at $z = 0.06$). We only show the mass-complete part of The300, and do not include its S-GSMF as it is mass-complete only above $M_{200c} \gtrsim 10^{15}\,\mathrm{M}_\odot$ at $z = 0$. We also only show the BAHAMAS component of BM in both panels, as it is a cosmological halo sample while the complementary MACSIS component is not.

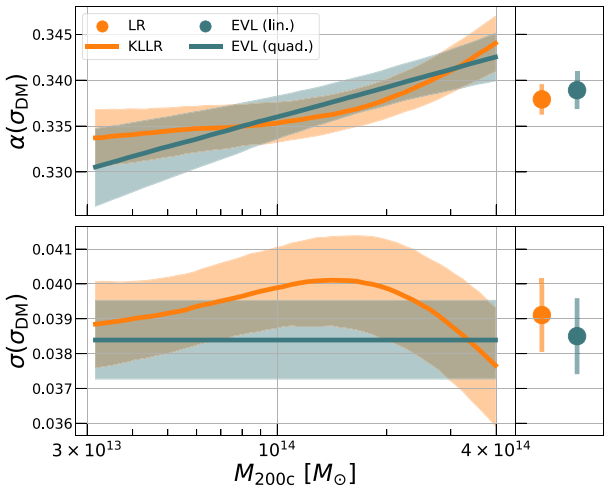

Figure 2. Comparisons of the scaling parameters of DM velocity dispersion with halo mass, extracted via least-squares linear regression (LR, Model I), local linear regression (KLLR, Model III), and EVL with a constant slope (lin., Model II) or a running of the slope with log-mass (quad., Model IV). Models I and II, as well as model III and IV, are in statistical consistency for both the slope (top panel) and the scatter (bottom panel). The mean relations of the pairs of models (in Fig. B2) are also statistically consistent with each other. The rest of this work uses the EVL model with a linear, constant slope. The uncertainties are 68 per cent intervals determined via bootstrap sampling for models I and III, and the marginalized 1D posteriors for models II and IV.

averaged – and not line of sight – DM velocity dispersion. Thus each halo is associated with 300 different velocities.

Fig. 2 shows a comparison between the different models, which are listed in Table 2. The right-hand panels compare the linear regression and linear EVL (models I and II) which return single values of slope and scatter, and the left ones compare KLLR and the quadratic EVL (models III and IV). The slopes (top panels) and the scatter (bottom panel) are in statistical agreement for both pairs of model comparisons. The normalizations are also statistically consistent, as shown and discussed in Appendix B2; this is true even when we input only $N = 10$ DM particles per halo in the EVL method.

To make a fair comparison between these methods, we ensure their estimates are derived from the same sample of haloes, and thus all methods only use haloes with $M_{200c} > 10^{13.5}\,\mathrm{M}_\odot$. This notably leads to an edge effect at the lower mass threshold of the KLLR estimates in Fig. 2, observed as the plateauing of the slope. We reiterate that all fiducial KLLR-based estimates in this work (e.g. Fig. 3) are derived from samples with appropriate halo mass ranges and do not suffer from any edge effects.

3.4 Velocity bias definition

The conventional definition of velocity bias is the ratio of the galaxy and DM velocity dispersions conditioned on host halo mass and redshift. Given the population statistics methods described above, we measure the velocity bias scaling relation, b_v, using the difference in the log-linear virial scaling relations

$$\log_{10} b_v = E[\log_{10}\sigma_{\mathrm{sat,1D}}|M_{200c}, M_{\star,\mathrm{sat}}, z] - E[\log_{10}\sigma_{\mathrm{DM}}|M_{200c}, z], \quad (6)$$

where $E[x]$ represents the expectation value of x derived by either the KLLR or the EVL method. Here, b_v is implicitly a function of M_{200c}, $M_{\star,\mathrm{sat}}$, and z.

Table 2. The methods corresponding to each model used in the validation tests shown in Fig. 2.

Name	Method
Model I	Linear regression (LR)
Model II	Linear EVL, equation (4)
Model III	KLLR (Section 3.1.1)
Model IV	Quadratic EVL, equation (5)

4 VIRIAL SCALING RELATIONS AND VELOCITY BIAS

We first present results for the DM and galaxy velocity dispersion scaling relations in Sections 4.1 and 4.2 before showing the velocity bias scaling relation in Section 4.3. Our scatter is expressed as a natural log – thus, a fractional scatter – and uncertainties on all estimates are 68 per cent confidence intervals, determined from the marginalized posteriors for the likelihood-based estimates (for galaxies) or from 1000 bootstrap resamplings of the KLLR sample (for DM).

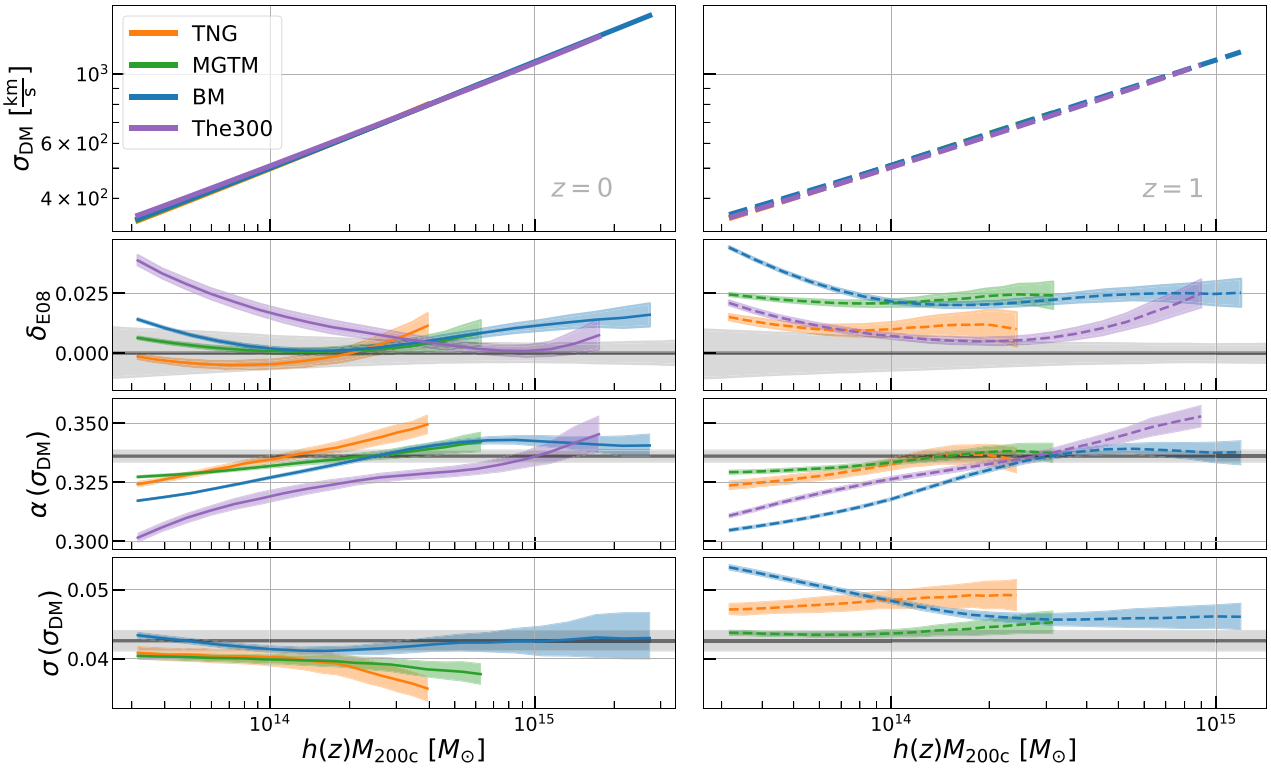

Figure 3. KLLR-derived scaling parameters of the DM velocity dispersion with halo mass at $z = 0$ (left column) and $z = 1$ (right column), from four different simulations (see legend). We follow Evrard et al. (2008, E08) in using $h(z)M_{200c}$ as the mass scale; see text for details. The upper middle panel shows the fractional difference between each simulation and the E08 prediction, $\delta_{\rm E08} = \ln(\sigma_{\rm DM}/\sigma_{\rm DM}^{\rm E08})$, with coloured (grey) bands showing the uncertainty in the simulation (E08) estimates. The slopes (lower middle panel) at low halo masses are shallower than the $\alpha = 1/3$ self-similar expectation, and at high masses agree well with the E08 result, $\alpha = 0.3361 \pm 0.0026$, shown in grey. The scatter (bottom panel) is relatively similar across the simulations and generally consistent with the E08 result, 0.0426 ± 0.0015, also shown in grey. THE300 has a significantly larger scatter at low halo masses due to the environment-selected nature of the low-mass sample, and so we omit it from the bottom panel for clarity but provide additional details in Appendix B1.

4.1 DM velocity dispersion, $\sigma_{\rm DM}$

Under self-similar evolution of haloes in virial equlibirum, the slope of the $\sigma_{\rm DM}$–M_{200c} relation is expected to be $\alpha = 1/3$ (Kaiser 1986; Bryan & Norman 1998). This expectation has since been confirmed for DMO and non-radiative simulations by Evrard et al. (2008), whose meta-analysis found $\alpha = 0.3361 \pm 0.0026$. Simulations with full baryonic physics treatments of galaxy formation find similar results for high-mass haloes (Lau et al. 2010; Armitage et al. 2018), but deviations of order 10 per cent in amplitude are found as one moves towards lower mass haloes (Anbajagane et al. 2022).

In Fig. 3, we show the $\sigma_{\rm DM}$–M_{200c} relation of four simulations for $z = 0$ and $z = 1$. MDPL2 is omitted as we do not have access to the requisite data. Note that we regress against $h(z)M_{200c}$ as, under self-similar evolution, the normalization of $\sigma_{\rm DM}$ when using this effective mass scale should have no redshift evolution (Kaiser 1986; Evrard et al. 2008; Singh et al. 2020). The normalizations are generally in good statistical agreement with E08 at $z = 0$ (upper middle panel, Fig. 3), although there is moderate tension with the BM simulation at the high-mass end.

The300 haloes at $z = 0$ have normalizations of up to 4 per cent higher than E08 for mass scales $M_{200c} < 10^{14.5}\,{\rm M}_\odot$. The sample at these lower halo masses is incomplete, as they only contain haloes that are within 22 Mpc of the 324 most mass haloes from MDPL2 that were selected for resimulation in The300. Thus, these lower mass haloes preferentially lie in regions with strong gravitational tidal fields, and some will have experienced fly-throughs or near encounters with their larger neighbours.

At $z = 1$, discrepancies of up to 2 per cent exist at $M_{200c} > 10^{14}\,{\rm M}_\odot$, and BM deviates by 4 per cent at the lowest masses. These deviations are significant at the $>4\sigma$ level in MGTM and BM, but the TNG300 and The300 normalizations exhibit less tension. We note that E08 quoted a normalization uncertainty of 0.5 per cent at $10^{15}\,{\rm M}_\odot$, rising to 1 per cent at $10^{13.5}\,{\rm M}_\odot$. Oddly, a normalization upturn at low masses is seen at $z = 1$ in BM but not in The300.

The slopes in the lower middle panel show a clear mass-dependence, as anticipated analytically (Okoli & Afshordi 2016), with nearly 5 per cent variation across the whole mass range, and with most of the deviations coming at low halo masses. This justifies our use of the KLLR method over a regular linear regression. The redshift evolution of this feature does have some discrepancies – the BM runs have shallower slopes at higher redshifts, whereas all other simulation populations have slopes either steeper or comparable to their $z = 0$ slopes.

We note the relevance of multiple full physics simulations exhibiting a mass-dependent slope across the range of redshifts probed in this analysis. The study of Anbajagane et al. (2022), which spans six decades in halo mass across three ILLUSTRISTNG simulation volumes, finds that the population statistics of multiple DM properties – velocity dispersion, concentration, halo shapes, formation histories etc. – contain non-monotonic, mass-localized features which originate from the interplay between AGN feedback

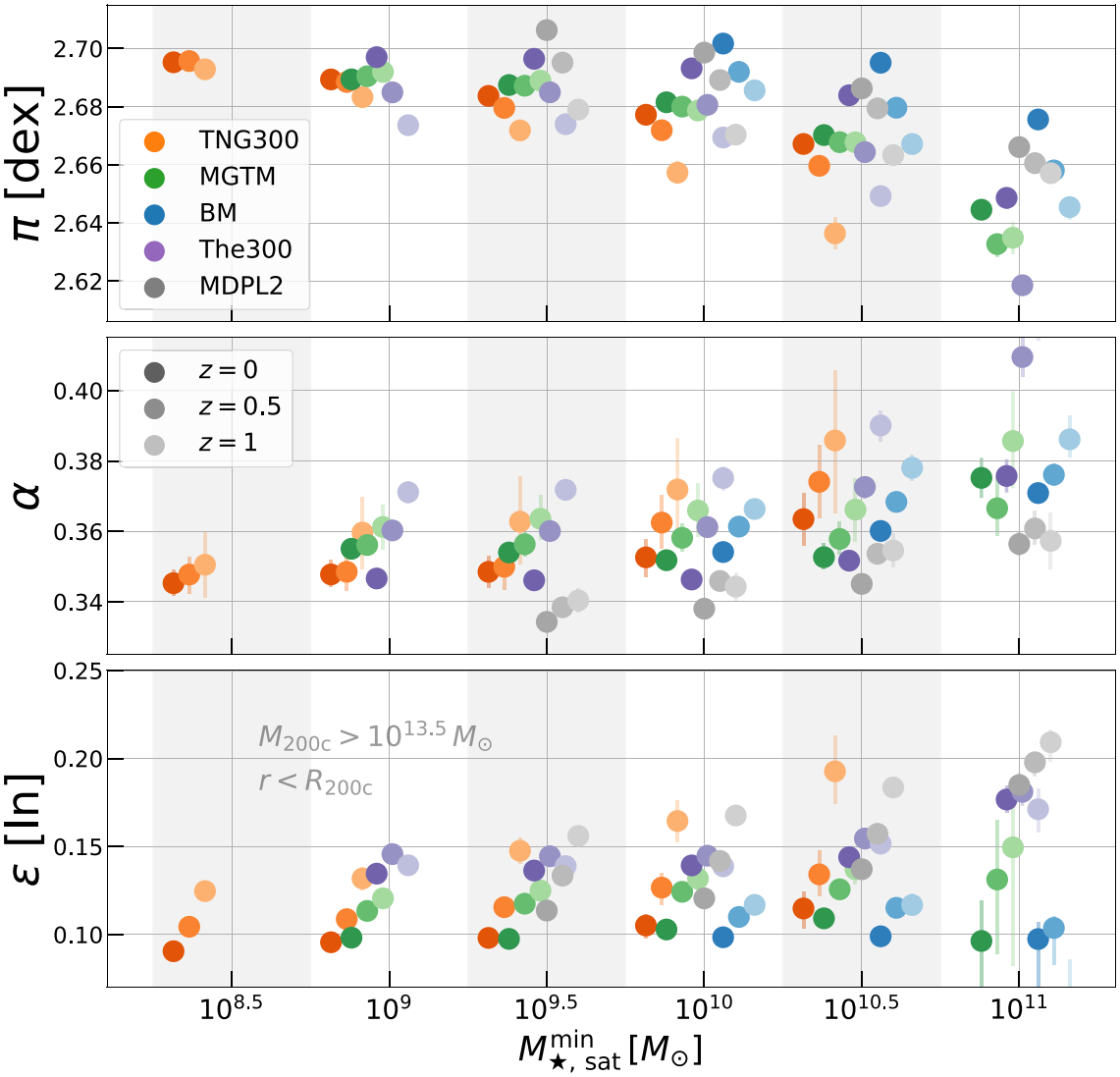

Figure 4. EVL-derived parameters of the $\sigma_{\rm sat,1D}$]–$M_{\rm 200c}$ scaling relation, equation (4), as a function of galaxy stellar mass threshold (denoted here as $M^{\rm min}_{\star,\rm sat}$), for all simulations at redshifts, $z \in \{0, 0.5, 1\}$. We use all galaxies within $R_{\rm 200c}$ of a host halo, and all host haloes with $M_{\rm 200c} > 10^{13.5}\,{\rm M}_\odot$; the exception is MDPL2 whose large halo population is subsampled, while preserving the shape of the HMF, to keep our computation time manageable. All simulations display the 'brighter is cooler' effect; normalizations (top, in decimal log) consistently decrease with increasing $M_{\star,\rm sat}$ threshold, and slopes (middle) consistently increase. The scatter (bottom, in natural log) shows some dependence on $M_{\star,\rm sat}$ threshold, particularly at higher redshifts. Most of the errors bars are smaller than the size of the circles. We artificially offset the points horizontally to enhance their visibility. The brackets in the axis labels of the top and bottom panels denote the log base of the quantities shown in each panel (either natural log, ln, or decimal log, dex).

and gas cooling. For $\sigma_{\rm DM}$, the inclusion of such processes results in the slope decreasing to $\alpha < 1/3$ towards group-scale haloes, and here we confirm consistent behaviour in three other hydrodynamical simulations.

4.2 Galaxy velocity dispersion, $\sigma_{\rm sat,1D}$

In moving to the $\sigma_{\rm sat,1D}$ relation, we examine the linear parameters for a pure power-law assumption for EVL, and utilize a range of stellar mass thresholds. For TNG300 we do not show results for $M_{\star,\rm sat} > 10^{11}\,{\rm M}_\odot$ due to the small sample size. Some simulations are unavailable at low $M_{\star,\rm sat}$ due to resolution limits.

Fig. 4 shows the derived parameters at $z \in \{0, 0.5, 1\}$. All simulations – both hydrodynamics and semianalytical variants – show a clear trend of the scaling relation normalization, π, decreasing as we increase the stellar mass threshold of the galaxy sample (top panel). This is qualitatively consistent with the signal found in the observational and simulation-based works previously noted (discussed further in Section 4.3), and is therefore also in tension with G15. The drop in normalization also steepens considerably beyond $M_{\star,\rm sat} > 10^{10.5}\,{\rm M}_\odot$, and this agrees with similar transition points from previous work[8] – r-band magnitude $M_{\rm r} \approx -21.5$ (Adami et al. 1998, 2000), z-band magnitude $M_z \approx -22$ (Goto 2005), and i-band magnitude $M_i \approx -22$ (Old et al. 2013), where the last result is from simulations whereas the rest are from observations.

[8]For these comparisons, we have used the TNG300 galaxy catalogue to find the absolute magnitudes, in different bands, corresponding to $M_{\star,\rm sat} = 10^{10.5}\,{\rm M}_\odot$.

The physical origin for the dependence of the normalization on $M_{\star,\rm sat}$ requires closer investigation, but dynamical friction (Chandrasekhar 1943) has been shown to reduce the velocities of a satellite galaxy sample over time (Ye et al. 2017; Armitage et al. 2018). Satellite galaxies of a larger mass (or mass-proxy, such as $M_{\star,\rm sat}$) experience stronger dynamical friction, and this naturally leads to the normalization of $\sigma_{\rm sat,1D}$ decreasing towards more massive galaxies. In addition, massive central galaxies are born at rest in their parent haloes, and therefore form a cooler sub-population during mergers compared to their satellite counterparts. After merging, most of the previously central galaxies will be classified as satellites of the larger system, and the most massive of these satellites would have established a lower velocity due to their prior role as centrals in the pre-merger phase.

At fixed $M_{\star,\rm sat}$ and z, the simulations' normalizations vary by about 0.02 dex, or ∼5 per cent. There is some striation between simulations, with BM preferring a higher normalization than most others, possibly due to its lower resolution. In Appendix A, we use multiple TNG300 runs to study resolution effects and show the normalization is amplified by decreased resolution. Such resolution effects also show up in other integral halo properties based on satellite galaxies – Anbajagane et al. (2020) studied the $M_{\star,\rm sat}$-thresholded satellite galaxy counts of massive haloes in multiple simulations (TNG300, MGTM, and BM) and found that increasing resolution can lead to 50−90 per cent more galaxy counts at a given host halo mass.

The normalizations in Fig. 4 also show stronger redshift evolution at higher $M_{\star,\rm sat}$. This is expected as the knee of the S-GSMF evolves more dramatically with redshift compared to the low $M_{\star,\rm sat}$ end. Thus, by fixing the $M_{\star,\rm sat}$ threshold at high values and varying redshifts, we sample significantly different parts of the S-GSMF. Note also that the velocity dispersion is always lower at higher redshifts, and this can be understood as follows. Let us define the 'rank' of a galaxy as its place in the $M_{\star,\rm sat}$-ordered list of galaxies at a given redshift; a rank of 1 implies the galaxy is the most massive in the sample. For a fixed $M_{\star,\rm sat}$ threshold, a low-redshift galaxy sample contains more low-rank galaxies ('low' meaning a larger rank) than the high-redshift sample. This, in the context of more massive galaxies being kinematically cooler, naturally implies that the higher redshift sample will have a lower velocity dispersion. Thus, the physical picture of dynamical friction discussed previously can explain the redshift evolution of the normalization as well.

The slopes of the $\sigma_{\rm sat,1D}$–$M_{\rm 200c}$ relation are constrained within $0.33 \lesssim \alpha \lesssim 0.40$ for all simulations across all redshifts and stellar mass thresholds (middle panel, Fig. 4), and have a redshift evolution in agreement with Munari et al. (2013), who found a range $0.35 < \alpha < 0.37$ over $0 < z < 1$ for $M_{\star,\rm sat} > 10^{9.5}\,{\rm M}_\odot$. Armitage et al. (2018) also computed the slopes at different redshifts, but given the large uncertainties on their estimates (resulting from a small sample size) their results are statistically consistent with no redshift evolution.

Compared to the $\sigma_{\rm DM}$–$M_{\rm 200c}$ relation (Fig. 3), the $\sigma_{\rm sat,1D}$–$M_{\rm 200c}$ relations in all simulations scale more steeply with halo mass. This feature could be due to dynamical friction – a satellite galaxy of a given size experiences *more* dynamical friction from a less massive host halo (see equation 4.2 and Appendix B Lacey & Cole 1993), and this additional host halo mass dependence can increase the slope of the $\sigma_{\rm sat,1D}$–$M_{\rm 200c}$ relation. The cold birth persistence after a merger may also play a more significant role in lower mass haloes. The mismatch of slopes between DM and galaxies naturally results in a halo-mass dependence of the velocity bias (see Section 4.3).

Finally, the scatter depends weakly on $M_{\star,\rm sat}$ (bottom panel, Fig. 4), though this dependence becomes stronger at higher redshifts. In general, the halo samples at $z = \{0, 0.5\}$ find $\epsilon \approx 0.1$ and this is broadly consistent with previous findings that lie in the range $\epsilon \in [0.06, 0.15]$ depending on galaxy stellar mass and redshift (Munari et al. 2013; Armitage et al. 2018).

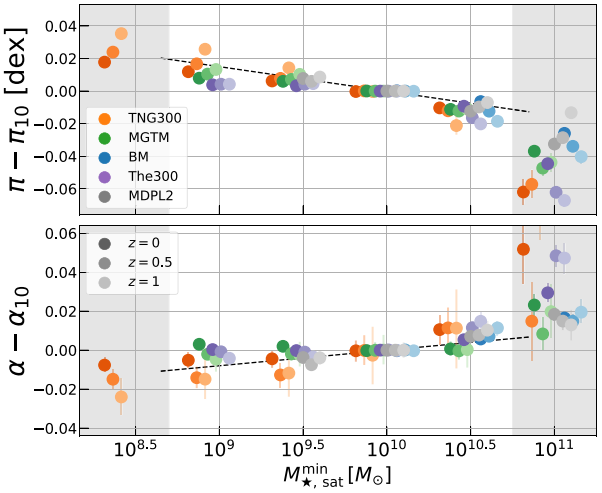

Figure 5. The normalizations (top panel, in decimal log) and slopes (bottom panel) of the $\sigma_{\rm sat,1D}$–$M_{\rm 200c}$ scaling relation, as shown in Fig. 4, but offset from their values at $M_{\star,\rm sat} > 10^{10}\,{\rm M}_\odot$. All sims show a 5–10 per cent difference in both parameters over the three-decades of $M_{\star,\rm sat}$ threshold values used in this work. The dashed black lines are linear fits, equation (7), with parameters given in Table 3. The grey bands show the points that are excluded from the fit.

Table 3. Variation of the scaling relation parameters, π and α, with stellar mass threshold, equation (7).

Parameter	q
Normalization, π	-0.015 ± 0.001
Slope, α	0.008 ± 0.001

4.2.1 *Relative trends with stellar mass threshold*

We next focus on the *relative* trends of EVL parameters with $M_{\star,\rm sat}$. In Fig. 5, the normalizations and slopes of each run and redshift, as shown in Fig. 4, have been normalized to their values for the threshold $M_{\star,\rm sat} > 10^{10}\,{\rm M}_\odot$. At this threshold, all runs have well-constrained estimates for the parameters. The variation in the normalization and slope with $M_{\star,\rm sat}$ threshold is well fit by a linear relation

$$x\left(M^{\rm min}_{\star,\rm sat}\right) - x_{10} = q\,\log_{10}\left(\frac{M^{\rm min}_{\star,\rm sat}}{10^{10}\,{\rm M}_\odot}\right), \tag{7}$$

where x is either π or α and x_{10} is the value for the threshold $M_{\star,\rm sat} > 10^{10}\,{\rm M}_\odot$. Fits are performed only using results for $10^9\,{\rm M}_\odot < M^{\rm min}_{\star,\rm sat} < 10^{10.5}\,{\rm M}_\odot$; at lower masses, we only have estimates for TNG300 while at higher masses the data trends deviate significantly from just a simple linear relationship. The data points are also *not* weighted by their errors during fitting as these errors are set by sample size, and so would cause the fitting procedure to strongly weight larger simulations with more haloes while not accounting for numerical resolution. Instead, in this fit, all simulations are equally weighted regardless of sample size.

The fit parameters for equation (7) are presented in Table 3 and the fits are shown in Fig. 5 as dashed black lines. The fractional scatter about the fit is ≈1 per cent in the region

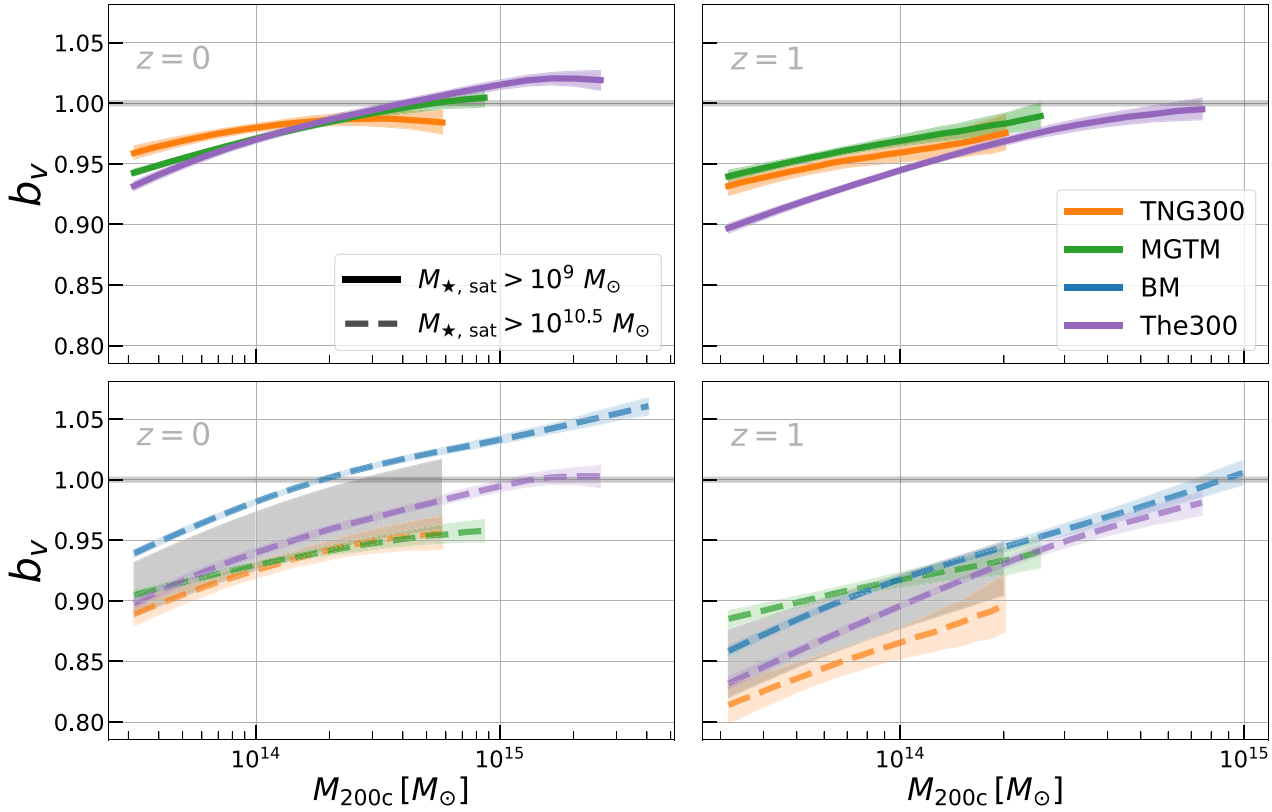

Figure 6. The satellite galaxy velocity bias for four simulations (different colors) as a function of halo mass, galaxy stellar mass threshold (different linestyles), and redshift (different panels). The velocity bias, b_v, increases with host halo mass, and decreases with redshift and galaxy stellar mass threshold. At fixed M_{200c}, $M_{\star,sat}$, and z, the simulations tend to vary by ≈2–3 per cent. The horizontal grey lines in all panels shows the unbiased case, $b_v = 1$, and the grey bands in the bottom row show the 68 per cent confidence interval of our theoretical prior on b_v estimated using all four simulations.

$10^9\,\mathrm{M}_\odot < M^{\mathrm{min}}_{\star,\mathrm{sat}} < 10^{10.5}\,\mathrm{M}_\odot$. Given a velocity bias (or galaxy velocity dispersion) estimate from a specific sample with a $M_{\star,\mathrm{sat}}$ threshold, these fits can be used to 'translate' that constraint to different galaxy stellar mass/magnitude thresholds. Note that our chosen analytical model is simplified, and thus approximate, given it does not include (i) redshift evolution of the parameters, and; (i) quadratic log-linear and higher order terms in the fit, which are particularly relevant towards high $M_{\star,\mathrm{sat}}$.

4.3 Galaxy velocity bias, b_v

The velocity bias, defined in equation (6), is the ratio of the mean scaling relations of galaxy and DM velocity dispersion with halo mass (Sections 4.2 and 4.1). Fig. 6 presents the velocity bias inferred from four simulations, as functions of M_{200c}, for two choices of $M_{\star,\mathrm{sat}}$ threshold, and two choices of redshift, z. MDPL2 is omitted as we did not have the requisite data for the σ_{DM}–M_{200c} relation.

At fixed $M_{\star,\mathrm{sat}}$ and z, the velocity bias increases nearly linearly with $\log_{10}M_{200c}$, with ∼20 per cent variation over the range presented. There is good constistency in values among the simulations. At fixed M_{200c}, $M_{\star,\mathrm{sat}}$, and z, the variation in b_v across simulations is 2–3 per cent for more than 90 per cent of the 3D parameter space, and improves to nearly *percent-level precision* if we consider the three highest resolution simulations (TNG300, MGTM, and The300). In general, this precision degrades the most at regimes of high z, M_{200c}, and/or $M_{\star,\mathrm{sat}}$, where it is amplified by the larger statistical uncertainties due to the smaller sample sizes.

We then construct a theoretical prior for b_v by first representing each simulation's b_v estimate as a Gaussian with a standard deviation given by the statistical uncertainty in b_v. Then, we sum the individual Gaussians to form a multimodal distribution, and compute its mean and standard deviation. These provide the moments for a Gaussian representation of the ensemble-based theoretical prior on b_v. Examples of these priors are shown in the bottom panels of Fig. 6.

The stellar mass dependence of b_v comes solely from the $\sigma_{\mathrm{sat,1D}}$–M_{200c} relation shown in Section 4.2. Previous observational and simulation-based works have studied the dependence of $\sigma_{\mathrm{sat,1D}}$ (and thus, b_v) on different galaxy/subhalo properties and have found similar trends to us. This is because their chosen properties, and subsequent analyses frameworks, are qualitatively related to the $M_{\star,\mathrm{sat}}$ threshold-based analysis we employ here, and we detail these connections below.

Prior observational works all threshold on either absolute magnitudes (Stein 1997; Adami et al. 1998, 2000; Goto 2005; Bayliss et al. 2017; Nascimento et al. 2017), or the difference $m - m_3$, where m is the galaxy apparent magnitude and m_3 is that of the third brightest galaxy in the cluster (Biviano et al. 1992; Girardi et al. 2003; Barsanti et al. 2016). Absolute magnitude thresholds are nearly equivalent to $M_{\star,\mathrm{sat}}$ thresholds given the close link between the two quantities, and a threshold on $m - m_3$ is equivalent to a $M_{\star,\mathrm{sat}}$ threshold that is allowed to vary across host haloes.

Simulation-based works have also used a variety of techniques – Biviano et al. (2006) studied b_v separately for early-type and late-type galaxies, and this corresponds to selections of high $M_{\star,\mathrm{sat}}$ and low $M_{\star,\mathrm{sat}}$, respectively. Lau et al. (2010) and Wu et al. (2013) select

the top N galaxies in each host halo according to $M_{\star,\rm sat}$ which, like $m - m_3$, is equivalent to using an $M_{\star,\rm sat}$ threshold that varies across host haloes. Ye et al. (2017) and Armitage et al. (2018) used differential stellar mass bins instead of cumulative ones but this method can still preferentially select galaxies based on $M_{\star,\rm sat}$. Finally, Ferragamo et al. (2020) took all the satellite galaxies belonging to cluster-scale haloes, rank-ordered them according to galaxy mass, and then selected only the top N per cent. This is equivalent to an $M_{\star,\rm sat}$ threshold that is set by the galaxy number counts. So all of the above works – both observation- and simulation-based – use frameworks that are qualitatively equivalent to the $M_{\star,\rm sat}$ thresholds used here, and thus show the same qualitative trends of more massive, or brighter, galaxies being kinematically cooler than their less massive, or fainter, counterparts.

We also find qualitative consistency with previous simulation-based analyses for the trends of velocity bias as a function of host halo mass (Munari et al. 2013; McCarthy et al. 2017; Ye et al. 2017) and redshift (Lau et al. 2010; Munari et al. 2013). Note also that the range of values we find, $b_v \in [0.8, 1.1]$, overlaps with those from these previous simulation studies (Lau et al. 2010; Munari et al. 2013; Wu et al. 2013; Ye et al. 2017; Armitage et al. 2018; Ferragamo et al. 2020).

5 REVISED MASS SCALE OF LOW-z SDSS CLUSTERS

Uncertainty in cluster mass calibration is a key systematic in cosmological analyses of galaxy cluster abundances (e.g. Murata et al. 2019; Costanzi et al. 2021). Weak lensing mass calibration is the current gold standard of mass calibration techniques (e.g. Bellagamba et al. 2019; McClintock et al. 2019; Miyatake et al. 2019; Murata et al. 2019; Kiiveri et al. 2021; Wu et al. 2021), while dynamical mass calibration using ensemble virial scaling is currently limited by the uncertainties in b_v (F16, Sifón et al. 2016). Increasing the precision of dynamical mass estimation, potentially to the level of weak lensing mass estimation, is also a prerequisite to enabling cluster-based tests of general relativity (e.g. Shirasaki et al. 2021a), which require comparisons of weak lensing and dynamical masses.

In Section 5.1 below, we update the analysis of F16 using our new theoretical prior for b_v, and show that we improve the precision on the mean log-halo mass by a factor of 3. Then, in Section 5.2, we discuss necessary future improvements for further improving the precision of dynamical mass estimates.

We also stress that while we focus on one example (an update to F16) to demonstrate the impacts of our work, improving dynamical mass estimation via our theoretical b_v priors has broader implications for cluster-based science that we do not explore here. For example, Bocquet et al. (2015, see Sections 3 and 5.1) discuss that a 1 per cent prior on the velocity bias improves constraints on both astrophysical and cosmological parameters connected to galaxy clusters by ≈30 per cent. Cluster counts as a function of their galaxy velocity dispersion and redshift has also emerged as an alternative approach for cluster cosmology (Caldwell et al. 2016; Ntampaka et al. 2017; Ntampaka, Rines & Trac 2019; Kirkpatrick et al. 2021), and the velocity bias is a critical component in forward modelling the relevant observable from cosmological parameters.

5.1 Updating the mass–richness normalization of F16

The work of F16 uses a slightly older version of EVL to estimate the the velocity dispersion of galaxies for an SDSS redMaPPer cluster sample, and then employs the velocity bias constraints from G15 to estimate the normalization of the halo mass–optical richness scaling relation $\langle M_{200c} \mid \lambda, z\rangle$, where the richness λ is an observational analog for the counts of red sequence satellite galaxies in a halo. We will henceforth refer to masses estimated using this approach as 'EVL masses.' Our update here replaces the G15 estimates with the theoretical b_v prior estimated in this work and recomputes the F16 normalization. While we focus primarily on updating the normalization, the slope of the scaling relation will also shift as G15 (and thus, F16) assumed that b_v did not vary with halo mass, whereas Fig. 6 shows that there is a significant mass dependence.

The pivot scales of F16 are $z = 0.2$ and $\lambda = 30$ for which the normalization, derived with the G15 estimate of $b_v = 1.05$, is $M_{200c} = 1.56 \pm 0.35 \times 10^{14}\,{\rm M}_\odot$. We are fortunate that we have b_v estimates near this halo mass scale from all four simulations (see Fig. 6). The ensemble-estimated theoretical prior for b_v is most accurate in halo mass ranges where all simulations are available, and gets more limited towards high halo masses, with only two simulations (BM and The300) sampling $M_{200c} > 10^{15}\,{\rm M}_\odot$ at $z = 0$ and $M_{200c} > 10^{14.5}\,{\rm M}_\odot$ at $z = 1$.

To update the F16 mass normalization, we use the Gaussian representation of the theoretical b_v prior as described in Section 4.3 and shown in Fig. 6. The first two moments of the Gaussian prior are interpolated over $M_{\star,\rm sat}$, M_{200c}, and z as necessary. F16 employed an approximate magnitude threshold of $M_{\rm r} \sim -21.5$ for their redMaPPer sample, and this corresponds to a stellar mass of $M_{\star,\rm sat} \approx 10^{10.3}\,{\rm M}_\odot$ in the TNG300 simulation. We tested the inclusion of a 50 per cent uncertainty (≈0.2 dex) on this threshold, and note that our results are not that sensitive to this choice. At a halo mass of $10^{14.2}\,{\rm M}_\odot$ at $z = 0$, the dependence of velocity bias on stellar mass threshold is weak, $b_v \propto (M_{\star,\rm sat})^{-0.015}$. Since the EVL mass estimate scales as $M \propto b_v^{-3}$, a 50 per cent uncertainty in $M_{\star,\rm sat}$ translates to an 2.3 per cent error in the mass scale. This is currently not a significant source of uncertainty and alters our total uncertainty estimates, presented further below, by ≈0.2 per cent.

The velocity bias is also a function of halo mass. Given the stellar mass threshold of $M_{\star,\rm sat} > 10^{10.3}\,{\rm M}_\odot$ at $z = 0.2$, we solve for the mass normalization iteratively. We first make an initial guess for the velocity bias, and derive a mass normalization constraint. Then we compute the simulation ensemble-estimated velocity bias at that mass scale and re-derive the mass normalization. This step is repeated until the normalization converges to within 0.01 per cent, which takes <10 iterations.

The mean and uncertainty of the normalization are determined via Monte Carlo sampling. We first draw a large number of random samples – we choose $N = 10^7$ for which our uncertainty estimates converge to within absolute deviations of 10^{-4} – from Gaussian priors for each of the following: the $\sigma_{\rm sat,1D}$ measurement from F16, the central galaxy velocity bias used in F16 (which comes from G15), the satellite galaxy velocity bias as constrained in this work, the normalization and slope of the $\sigma_{\rm DM}$–M_{200c} relation from E08, and finally, the cosmological parameters Ω_M and h from Planck Collaboration XIII (2016). The mean and variance we use for each Gaussian prior come from the works listed above alongside each quantity. For b_v, these are obtained from the ensemble-estimated theoretical prior. Then, we once again compute the mass normalization iteratively, and the mean and standard deviation of the resulting distribution of 10^7 normalizations is the quoted mean and uncertainty.

Fig. 7 presents two variants of the updated F16 constraints – one where we use all four simulations (TNG300, BM, MGTM, and The300), and one where we exclude BM and use only the three highest resolution ones. The latter is motivated by the resolution

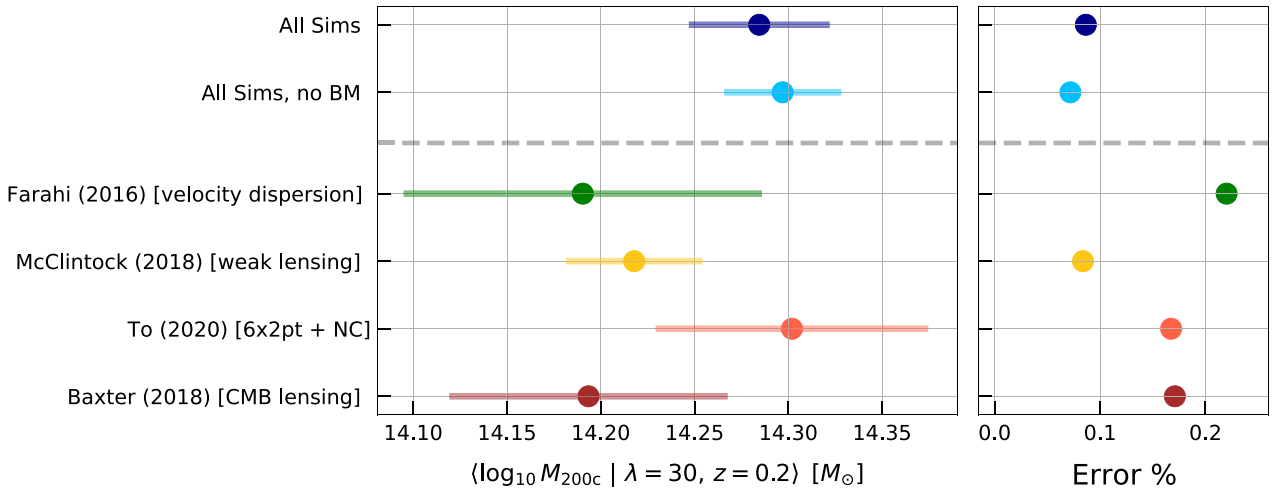

Figure 7. Mass–richness normalization at fixed redshift, $\langle \log_{10} M_{200c} \,|\, \lambda = 30, z = 0.2 \rangle$ derived from different measurement methodologies (left-hand panel, see also Table 4), and the fractional uncertainty on each estimate (right-hand panel). First two points from the top are our updates to F16 using the theoretical b_v prior from either all four cosmological hydrodynamics simulations or the three highest resolution ones. Either estimate reduces the uncertainty from the original F16 estimate by a factor of ≈3. The last three results were evaluated away from their calibrated pivot points and thus include uncertainties from the evolution with redshift and richness as well.

considerations mentioned previously and discussed further in Appendix A. Since the mean velocity bias has declined, to $b_v = 0.98$ compared to the original G15 estimate of $b_v = 1.05$, the inferred mass scale (which goes as $M \propto b_v^{-3}$) rises by nearly 0.1 dex. The 0.07 shift in b_v is within the stated 1σ error of G15 ($\Delta b_v = 0.08$) so the revised mass estimate remains consistent with the wide 68 per cent confidence interval of the original F16 estimate determined using the G15 constraints. The revised EVL normalization also remains consistent with existing weak lensing estimates, including the recent multiprobe estimate of To et al. (2021).

F16 report a slope of $\alpha = 1.31 \pm 0.06_{\rm stat} \pm 0.13_{\rm sys}$ for the mass–richness relation, but this value was derived assuming a constant b_v. At low redshift, the simulations display a weak halo mass dependence, $b_v \propto (M_{200c})^{0.03}$, which would imply a shift of -0.09, meaning a revised slope of $\alpha \approx 1.22$. The shallower slope of the SDSS redMaPPer mass–richness relation still lies between values in the literature derived from different methodologies; Simet et al. (2018) find $\alpha = 1.30 \pm 0.10$ while Murata et al. (2019) find $\alpha = 1.00 \pm 0.05$.

Fig. 7 and Table 4 compare our revised EVL mass normalization with those from previous observational studies. When a published value is quoted at a different richness and redshift, we translate it to the pivot richness and redshift of F16 – $\lambda = 30$ and $z = 0.2$ – while incorporating the extra uncertainty from moving off the fiducial pivot scale due to slope or redshift evolution uncertainties. When other works only quote the richness–mass normalization, $\langle \lambda \,|\, M_{200c}, z \rangle$, we use the formalism of Evrard et al. (2014) to invert the scaling relation and obtain the mass–richness normalization, $\langle M_{200c} \,|\, \lambda, z \rangle$. Many studies also quote their halo mass in M_{200m}, which is defined similar to M_{200c} but now with $\rho_\Delta = 200\rho_m(z)$, where ρ_m is the mean matter density at redshift z. We convert $M_{200m} \rightarrow M_{200c}$ using the COLOSSUS[9] open-source python package (Diemer 2018) while employing the concentration–mass relation from Diemer & Joyce (2019). The uncertainty from the concentration relation is not incorporated into our final estimate.

[9]https://bdiemer.bitbucket.io/colossus/

The right-hand panel of Fig. 7 shows the magnitude of the total uncertainty in the various mass normalizations. The specific values we constrain, as well as those of the comparison works, are found in Table 4. The improved precision on b_v, due to the simulation ensemble-estimated theoretical prior, reduces the original F16 uncertainties by a factor of 3, and makes dynamical mass estimation a competitive technique in determining the normalization of the mass–richness relation.

5.2 Roadmap to more precise EVL mass estimates

To motivate potential improvements in future analysis, we illustrate in Fig. 8 the importance of difference sources of the uncertainty in the EVL mass normalization. The figure shows the fraction of the overall variance contributed by the uncertainty in each individual source. For the purpose of illustration, we assume $z = 0.5$, $M_{\star,\rm sat} > 10^{10}\,{\rm M}_\odot$, and exclude BM when determining the theoretical prior for $b_{v,\rm sat}$. The statistical uncertainty is the SDSS value from F16. The overall fractional uncertainty in the mass scale is $\sigma_{\rm err} = 7.8$ per cent, not including the $M_{\star,\rm sat}$ uncertainties discussed previously.

The halo velocity we employ as the theoretical reference rest frame in equation (2) is unavailable to observers. Instead, F16 uses the central galaxy velocity as a proxy, thereby bringing the central galaxy velocity bias, $b_{v,\rm cen}$, into the analysis. Uncertainty in that component is comparable to the current SDSS statistical error in the velocity dispersion normalization at $\lambda = 30$. So satellite galaxy velocity bias is now a sub-dominant source of systematic uncertainty, while the central galaxy velocity bias becomes the dominant source.

Prospects for improvements to the dominant sources of uncertainty are good. The statistical uncertainty of the measurement will improve just by increasing the size and depth of spectroscopic samples. The original analysis of velocity dispersion scaling with optical richness by Rozo et al. (2015) employed roughly 9000 clusters, each sampled by 20 or more spectroscopic galaxy members, for a sample size of approximately 200 000 galaxies. Recent wide-area imaging surveys, such as the Dark Energy Survey (DES; The Dark Energy Survey Collaboration 2005) and Hyper Suprime-Cam (Aihara et al. 2018), are producing much larger optically selected cluster samples, and

Table 4. The mass–richness normalization from this work and from previous works that estimated the same using different techniques: a slightly older version of EVL (F16), weak lensing (McClintock et al. 2019), CMB lensing (Baxter et al. 2018), and from a combined cosmological analysis of different probes, including cluster-scale haloes (To et al. 2021). The estimates for McClintock et al. (2019) and To et al. (2021) are quoted away their respective pivot points, and thus their errors include additional uncertainties coming from the redshift and richness evolution. The uncertainties at their pivot scales are 5.1 and 4.6 per cent, respectively.

Source	Technique	$\langle \log_{10} M_{200c} \vert \lambda = 30, z = 0.2 \rangle$	Error
This work, All sims	Ensemble velocity likelihood (EVL)	14.28 ± 0.037	8.7 per cent
This work, Sims w/o BM	EVL	14.30 ± 0.031	7.4 per cent
Farahi et al. (2016)	Older EVL	14.19 ± 0.096	22 per cent
McClintock et al. (2019)	Background galaxy weak lensing	14.22 ± 0.035	8.1 per cent
To et al. (2021)	Galaxy/DM clustering + cluster abundance	14.30 ± 0.079	18 per cent
Baxter et al. (2018)	CMB lensing	14.19 ± 0.074	17 per cent

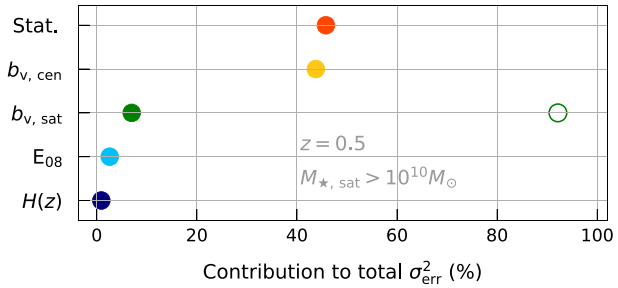

Figure 8. The percentage contribution of each component X to the total *variance* of the mass normalization. From top to bottom X is: (i) the statistical uncertainty from the $\sigma_{sat,1D}$ observational measurement, (ii and iii) the central and satellites galaxy velocity biases, (iv) the $\sigma_{DM}-M_{200c}$ relation from E08, and (v) the cosmology uncertainty from $H(z)$ assuming Planck-like constraints. The open green circle is the previous result using G15. We show results for $z = 0.5$ and $M_{\star,sat} > 10^{10}\,M_\odot$, but the qualitative behaviour is similar across a large part of the parameter space. We have used only the three highest resolution simulations (TNG300, MGTM, and The300) for this analysis. See text for details.

overlapping areas of sky are being probed by Sunyaev–Zel'dovich observations from the Atacama Cosmology Telescope (Choi et al. 2020) and South Pole Telescope (Carlstrom et al. 2011), and also X-ray observations from the *eROSITA* mission (Merloni et al. 2012). Spectroscopic surveys such as the Dark Energy Spectroscopic Instrument (DESI; Dey et al. 2019), Euclid (Laureijs et al. 2011) and, in the longer term, the Nancy Grace Roman Telescope (Akeson et al. 2019) and Extremely Large Telescope MOSAIC (Evans et al. 2015), will produce samples larger by an order of magnitude or more compared to the SDSS analysis of F16.

The central galaxy velocity bias, on the other hand, will need to be studied more extensively via simulations to quantify the theoretical uncertainty by constructing an ensemble-estimated theoretical prior in a manner similar to that of this work. Previous works have calibrated this bias as a function of galaxy and/or host halo properties, but either do not adequately describe galaxy velocities within clusters, i.e. the one-halo term (G15), or are unable to quantify the theoretical uncertainty due to the study being limited to a single simulation (Martel, Robichaud & Barai 2014; Ye et al. 2017).

Other sources of systematic uncertainty, such as miscentering and projection, will also need to be precisely calibrated. A promising approach for the former may be to use multiwavelength observations, such as X-ray and optical (Zhang et al. 2019), to define a well-centred subset of clusters. Application of EVL and other dynamical mass techniques to this subset would produce estimates more reflective of the underlying massive halo population.

Note that the EVL method focuses solely on a line-of-sight velocity dispersion. Other methods, such as those using caustics (Rines et al. 2003; Rines & Diaferio 2006; Gifford, Miller & Kern 2013; Gifford, Kern & Miller 2017), or the Jeans equation (Mamon, Biviano & Boué 2013), also employ the transverse radial distances. Recent deep learning techniques also make full use of the 2D phase space consisting of line-of-sight velocities and transverse radial distances (Ho et al. 2019).

Additionally, while we have discussed and documented the 'brighter is cooler' effect in the context of a velocity dispersion/bias, the effect impacts the full 6D position–velocity phase space of the galaxies. So other relevant features in this phase space – such as the outer caustic surface, or splashback feature (Adhikari, Dalal & Chamberlain 2014; Diemer & Kravtsov 2014; More, Diemer & Kravtsov 2015) – will also be impacted by this effect. Previous studies of both observations and simulations have found that the radial location of the splashback feature, as estimated via the galaxy number density profile, depends on galaxy properties such as colour and mass (Adhikari et al. 2021; Dacunha et al. 2021).

6 CONCLUSIONS

Estimating the mass scale of galaxy clusters from the ensemble velocity statistics of satellite galaxies is a method that is currently limited by uncertainties in how well galaxies trace the DM velocity field. The velocity bias, b_v – which is the ratio of velocity dispersion of satellite galaxies to that of dark matter – is a key source of uncertainty that we address using new statistical methods applied to an ensemble of cosmological hydrodynamics simulations that include an extensive range of galaxy formation physics.

We extract estimates of b_v as a function of host halo mass, satellite galaxy stellar mass threshold, and redshift using a set of four independent cosmological hydrodynamics simulations. This is done using both a local linear regression, as well as a new ensemble velocity likelihood method that is unbiased for low galaxy counts per halo. The collective analysis of the multiple simulations allows us to derive an ensemble-estimated theoretical prior on b_v that quantifies the uncertainty driven by different astrophysical and numerical treatments. Our main results are as follows:

(i) At $z = 0$, the DM velocity dispersion scaling relation is consistent across all simulations at the one percent level and agrees with prior expectations from E08 (Fig. 3), but larger deviations are seen at $z = 1$. The slopes in all simulations have a consistent mass dependence, and are shallower than the self-similar expectation ($\alpha = 1/3$) at halo masses below $M_{200c} < 10^{14}\,M_\odot$.

(ii) The normalization of the galaxy velocity dispersion scaling relation decreases with stellar mass threshold, indicating that more massive galaxies are kinematically cooler than their lighter counterparts (top panel, Fig. 4). The redshift and stellar mass dependence of this feature is consistent across an ensemble consisting of four hydrodynamics simulations and one *N*-body/semianalytical simulation (Fig. 5).

(iii) In all simulations, the slopes of the $\sigma_{\rm sat,1D}$–$M_{\rm 200c}$ scaling relation are greater than the self-similar expectation, and the relation steepens with both stellar mass threshold and redshift (middle panel, Fig. 4).

(iv) The ratio of the $\sigma_{\rm sat,1D}$–$M_{\rm 200c}$ and $\sigma_{\rm DM}$–$M_{\rm 200c}$ scaling relations yields a velocity bias, b_v, that varies as a function of host halo mass, galaxy stellar mass threshold, and redshift (Fig. 6). The simulation-to-simulation variation is 2–3 per cent for more than 90 per cent of the 3D parameter space constituting $M_{\rm 200c}$, $M_{\star,\rm sat}$, and z. However, this reduces to *percent-level precision* when considering only the three highest resolution simulations (TNG300, MGTM, and The300). The uncertainty is larger at higher redshift and higher halo/stellar mass scales where the halo samples are sparse.

(v) We update the mass normalization of optically selected SDSS clusters studied in F16 by using the ensemble-estimated theoretical b_v prior derived in our work. Our more precise estimate improves the uncertainty on the normalization from 22 per cent to 7–8 per cent (Figs 7 and 8), and makes dynamical mass estimation using the ensemble velocity of satellite galaxies a technique that is competitive with weak lensing.

The *trends* in velocity bias discussed in this work are all empirically testable with ongoing spectroscopic campaigns of clusters such as SPIDERS (Kirkpatrick et al. 2021) and DESI. The dependence of $\sigma_{\rm sat,1D}$ on $M_{\star,\rm sat}$ (or more precisely, the galaxy luminosity) has already been observationally studied for many different modestly sized samples of clusters ($N \sim 100$) as was noted before, while such observational studies of the redshift and halo mass trends have not yet been well-explored. The same data sets could be used to derive EVL-based constraints on the mass–richness normalization for cluster samples selected by different methods. Comparisons of precise mass-scale estimates between X-ray, SZ, and optically selected samples would offer insights into sample selection models, the strength of projection effects, and intrinsic covariance among stellar, hot gas and dark matter properties.

Finally, while we have focused on galaxy cluster mass calibration as the premier application of our velocity bias constraints, our results can also be relevant for models of small-scale RSDs measurements (e.g., Tinker 2007; Reid et al. 2014; Guo et al. 2015a, b, c; DESI Collaboration 2016; Yuan, Eisenstein & Garrison 2018; Zhai et al. 2019; Tonegawa et al. 2020; Alam et al. 2021; DeRose, Becker & Wechsler 2021; Shirasaki et al. 2021b; Lange et al. 2022) and more generally, any small-scale *N*-point auto- or cross-correlation function that uses galaxy kinematics as an observational tracer of the DM velocity field. Such cosmological probes will also be highly relevant over the next decade given the expected large sky coverage and redshift range of DESI and future spectroscopic surveys.[10]

ACKNOWLEDGEMENTS

We thank Chun-Hao To and Alexander Knebe for useful discussions. We also thank the IllustrisTNG Team for publicly releasing all TNG simulation data, and Peter Behroozi for doing the same for the MDPL2 UniverseMachine catalogues. We also thank the anonymous referee for helpful comments on the presentation of the methods employed in this work.

DA is supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1746045. AF is partially supported by a Michigan Institute for Data Science (MIDAS) Fellowship. WC is supported by the European Research Council under grant number 670193 and by the Science and Technology Facilities Council (STFC) AGP Grant ST/V000594/1. He further acknowledges the science research grants from the China Manned Space Project with NO. CMS-CSST-2021-A01 and CMS-CSST-2021-B01. KD acknowledges support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC-2094 - 390783311 and through the COMPLEX project from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program grant agreement ERC-2019-AdG 860744. GY acknowledges financial support from the MICIU/FEDER (Spain) under project grant PGC2018-094975-C21.

This research was supported in part through computational resources and services provided by Advanced Research Computing (ARC), a division of Information and Technology Services (ITS) at the University of Michigan, Ann Arbor.

The TNG simulations were run with compute time granted by the Gauss Centre for Supercomputing (GCS) under Large-Scale Projects GCS-ILLU and GCS-DWAR on the GCS share of the supercomputer Hazel Hen at the High Performance Computing Center Stuttgart (HLRS).

The *Magneticum* simulations were performed at the Leibniz-Rechenzentrum with CPU time assigned to the Project 'pr86re'.

The 300 project has received financial support from the European Union's Horizon 2020 Research and Innovation programme under the Marie Sklodowskaw-Curie grant agreement number 734374, i.e. the LACEGAL project. We would like to thank The Red Española de Supercomputación for granting us computing time at the MareNostrum Supercomputer of the Barcelona Supercomputing Center - Centro Nacional de Supercomptacion (BSC-CNS) where most of the 300 cluster simulations have been performed. The MDPL2 simulation has been performed at LRZ Munich within the project pr87yi. The CosmoSim database (https://www.cosmosim.org) is a service by the Leibniz Institute for Astrophysics Potsdam (AIP). Part of the computations with `GADGET-X` have also been performed at the 'Leibniz-Rechenzentrum' with CPU time assigned to the Project 'pr83li'.

The MultiDark Database used in this paper and the web application providing online access to it were constructed as part of the activities of the German Astrophysical Virtual Observatory as result of a collaboration between the Leibniz Institute for Astrophysics Potsdam (AIP) and the Spanish MultiDark Consolider Project CSD2009-00064.

All analysis in this work was enabled greatly by the following software: `PANDAS` (McKinney 2011), `NUMPY` (van der Walt, Colbert & Varoquaux 2011), `SCIPY` (Virtanen et al. 2020), and `MATPLOTLIB` (Hunter 2007). We have also used the Astrophysics Data Service (ADS) and `arXiv` preprint repository extensively during this project and the writing of the paper.

DATA AVAILABILITY

The tables containing the scaling relations for galaxy/DM velocity dispersions and the velocity bias are publicly available at

[10] e.g., ASTRO2020 White Paper: Towards a Spectroscopic Survey Roadmap for the 2020s and Beyond

https://github.com/DhayaaAnbajagane/VelocityBias. We also provide a convenience script that parses the scaling parameter files, and also provides the theoretical b_v prior while being able to interpolate over host halo mass, galaxy stellar mass threshold, and redshift, as needed.

The galaxy and halo catalogs for ILLUSTRISTNG, MAGNETICUM PATHFINDER[11], and UNIVERSEMACHINE are all publicly available at the repositories linked in Section 2. The data for THE300, BAHAMAS, and MACSIS are not available at a public repository, but can be provided on request.

REFERENCES

Adami C., Biviano A., Mazure A., 1998, A&A, 331, 439
Adami C., Holden B. P., Castander F. J., Mazure A., Nichol R. C., Ulmer M. P., 2000, A&A, 362, 825
Adhikari S., Dalal N., Chamberlain R. T., 2014, J. Cosmol. Astropart. Phys., 2014, 019
Adhikari S. et al., 2021, ApJ, 923, 37
Aihara H. et al., 2018, PASJ, 70, S4
Akeson R. et al., 2019, preprint (arXiv:1902.05569)
Alam S., Peacock J. A., Farrow D. J., Loveday J., Hopkins A. M., 2021, MNRAS, 503, 59
Allen S. W., Evrard A. E., Mantz A. B., 2011, ARA&A, 49, 409
Anbajagane D., Evrard A. E., Farahi A., Barnes D. J., Dolag K., McCarthy I. G., Nelson D., Pillepich A., 2020, MNRAS, 495, 686
Anbajagane D., Evrard A. E., Farahi A., 2022, MNRAS, 509, 3441
Armitage T. J., Barnes D. J., Kay S. T., Bahé Y. M., Dalla Vecchia C., Crain R. A., Theuns T., 2018, MNRAS, 474, 3746
Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15
Barnes D. J., Kay S. T., Henson M. A., McCarthy I. G., Schaye J., Jenkins A., 2017, MNRAS, 465, 213
Barsanti S., Girardi M., Biviano A., Borgani S., Annunziatella M., Nonino M., 2016, A&A, 595, A73
Baxter E. J. et al., 2018, MNRAS, 476, 2674
Bayliss M. B. et al., 2017, ApJ, 837, 88
Beck A. M. et al., 2016, MNRAS, 455, 2110
Beers T. C., Flynn K., Gebhardt K., 1990, AJ, 100, 32
Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013, ApJ, 762, 109
Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, MNRAS, 488, 3143
Bellagamba F. et al., 2019, MNRAS, 484, 1598
Biviano A., Girardi M., Giuricin G., Mardirossian F., Mezzetti M., 1992, ApJ, 396, 35
Biviano A., Murante G., Borgani S., Diaferio A., Dolag K., Girardi M., 2006, A&A, 456, 23
Bocquet S. et al., 2015, ApJ, 799, 214
Bryan G. L., Norman M. L., 1998, ApJ, 495, 80
Caldwell C. E., McCarthy I. G., Baldry I. K., Collins C. A., Schaye J., Bird S., 2016, MNRAS, 462, 4117
Carlberg R. G., 1991, ApJ, 367, 385
Carlberg R. G., Couchman H. M. P., Thomas P. A., 1990, ApJ, 352, L29
Carlstrom J. E. et al., 2011, PASP, 123, 568
Chandrasekhar S., 1943, ApJ, 97, 255
Choi S. K. et al., 2020, J. Cosmol. Astropart. Phys., 2020, 045
Colafrancesco S., Antonuccio-Delogu V., Del Popolo A., 1995, ApJ, 455, 32
Costanzi M. et al., 2021, Phys. Rev. D, 103, 043522
Cui W. et al., 2018, MNRAS, 480, 2898
Dacunha T., Belyakov M., Adhikari S., Shin T.-h., Goldstein S., Jain B., 2021, preprint (arXiv:2111.06499)
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
DeRose J., Becker M. R., Wechsler R. H., 2021, preprint (arXiv:2105.12104)
DESI Collaboration, 2016, preprint (arXiv:1611.00036)
Dey A. et al., 2019, AJ, 157, 168
Diemer B., 2018, ApJS, 239, 35
Diemer B., Joyce M., 2019, ApJ, 871, 168
Diemer B., Kravtsov A. V., 2014, ApJ, 789, 1
Dolag K., Borgani S., Murante G., Springel V., 2009, MNRAS, 399, 497
Evans C. et al., 2015, preprint (arXiv:1501.04726)
Evrard A. E., Arnault P., Huterer D., Farahi A., 2014, MNRAS, 441, 3562
Evrard A. E. et al., 2008, ApJ, 672, 122 (E08)
Farahi A., Evrard A. E., Rozo E., Rykoff E. S., Wechsler R. H., 2016, MNRAS, 460, 3900 (F16)
Farahi A., Evrard A. E., McCarthy I., Barnes D. J., Kay S. T., 2018a, MNRAS, 478, 2618
Farahi A. et al., 2018b, A&A, 620, A8
Ferragamo A., Rubiño-Martín J. A., Betancort-Rijo J., Munari E., Sartoris B., Barrena R., 2020, Eur. Phys. J. Web Conf., 228, 00011
Gifford D., Miller C., Kern N., 2013, ApJ, 773, 116
Gifford D., Kern N., Miller C. J., 2017, ApJ, 834, 204
Girardi M., Rigoni E., Mardirossian F., Mezzetti M., 2003, A&A, 406, 403
Goto T., 2005, MNRAS, 359, 1415
Guo H. et al., 2015a, MNRAS, 446, 578
Guo H. et al., 2015b, MNRAS, 449, L95
Guo H. et al., 2015c, MNRAS, 453, 4368 (G15)
Hirschmann M., Dolag K., Saro A., Bachmann L., Borgani S., Burkert A., 2014, MNRAS, 442, 2304
Ho M., Rau M. M., Ntampaka M., Farahi A., Trac H., Póczos B., 2019, ApJ, 887, 25
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Kaiser N., 1984, ApJ, 284, L9
Kaiser N., 1986, MNRAS, 222, 323
Kiiveri K. et al., 2021, MNRAS, 502, 1494
Kirkpatrick C. C. et al., 2021, MNRAS, 503, 5763
Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340
Knebe A. et al., 2011, MNRAS, 415, 2293
Knollmann S. R., Knebe A., 2009, ApJS, 182, 608
Komatsu E. et al., 2011, ApJS, 192, 18
Lacey C., Cole S., 1993, MNRAS, 262, 627
Lange J. U., Hearin A. P., Leauthaud A., van den Bosch F. C., Guo H., DeRose J., 2022, MNRAS, 509, 1779
Lau E. T., Nagai D., Kravtsov A. V., 2010, ApJ, 708, 1419
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Mamon G. A., Biviano A., Boué G., 2013, MNRAS, 429, 3079
Mansfield P., Avestruz C., 2021, MNRAS, 500, 3309
Marinacci F. et al., 2018, MNRAS, 480, 5113
Martel H., Robichaud F., Barai P., 2014, ApJ, 786, 79
McCarthy I. G., Schaye J., Bird S., Le Brun A. M. C., 2017, MNRAS, 465, 2936
McClintock T. et al., 2019, MNRAS, 482, 1352
McKinney W. R., Yashchuk V. V., Goldberg K. A., Howells M., Artemiev N. A., Merthe D. J., Yuan S., 2011, in del Rio M. S., Chubar O., eds, Proc. SPIE Conf. Ser. Vol. 8141, Advances in Computational Methods for X-Ray Optics II. SPIE, Bellingham, p. 81410K
Merloni A. et al., 2012, preprint (arXiv:1209.3114)
Miyatake H. et al., 2019, ApJ, 875, 63
More S., Diemer B., Kravtsov A. V., 2015, ApJ, 810, 36
Munari E., Biviano A., Borgani S., Murante G., Fabjan D., 2013, MNRAS, 430, 2638
Murata R. et al., 2019, PASJ, 71, 107
Naiman J. P. et al., 2018, MNRAS, 477, 1206
Nascimento R. S., Ribeiro A. L. B., Lopes P. A. A., 2017, MNRAS, 464, 183
Nelson D. et al., 2018, MNRAS, 475, 624
Nelson D. et al., 2019, Comput. Astrophys. Cosmol., 6, 2
Ntampaka M., Trac H., Cisewski J., Price L. C., 2017, ApJ, 835, 106
Ntampaka M., Rines K., Trac H., 2019, ApJ, 880, 154

[11]The $\sigma_{\rm DM}$ and $M_{\rm 200c}$ quantities for MAGNETICUM PATHFINDER are not available at the public repository and were provided by one of us for use in this study.

Okoli C., Afshordi N., 2016, MNRAS, 456, 3068
Old L., Gray M. E., Pearce F. R., 2013, MNRAS, 434, 2606
Onions J. et al., 2012, MNRAS, 423, 1200
Pillepich A. et al., 2018a, MNRAS, 473, 4077
Pillepich A. et al., 2018b, MNRAS, 475, 648
Planck Collaboration XVI, 2014, A&A, 571, A16
Planck Collaboration XIII, 2016, A&A, 594, A13
Pratt G. W., Arnaud M., Biviano A., Eckert D., Ettori S., Nagai D., Okabe N., Reiprich T. H., 2019, Space Sci. Rev., 215, 25
Ragagnin A., Dolag K., Biffi V., Cadolle Bel M., Hammer N. J., Krukau A., Petkova M., Steinborn D., 2017, Astron. Comput., 20, 52
Rasia E. et al., 2015, ApJ, 813, L17
Reid B. A., Seo H.-J., Leauthaud A., Tinker J. L., White M., 2014, MNRAS, 444, 476
Riebe K. et al., 2011, preprint (arXiv:1109.0003)
Rines K., Diaferio A., 2006, AJ, 132, 1275
Rines K., Geller M. J., Kurtz M. J., Diaferio A., 2003, AJ, 126, 2152
Rozo E., Rykoff E. S., Becker M., Reddick R. M., Wechsler R. H., 2015, MNRAS, 453, 38
Salvatier J., Wiecki T., Fonnesbeck C., 2015, preprint (arXiv:1507.08050)
Schneider A. et al., 2016, J. Cosmol. Astropart. Phys., 2016, 047
Sembolini F. et al., 2016a, MNRAS, 457, 4063
Sembolini F. et al., 2016b, MNRAS, 459, 2973
Shirasaki M., Egami E., Okabe N., Miyazaki S., 2021a, MNRAS, 506, 3385
Shirasaki M., Huff E. M., Markovic K., Rhodes J. D., 2021b, ApJ, 907, 38
Sifón C. et al., 2016, MNRAS, 461, 248
Simet M., McClintock T., Mandelbaum R., Rozo E., Rykoff E., Sheldon E., Wechsler R. H., 2018, MNRAS, 480, 5385
Singh P., Saro A., Costanzi M., Dolag K., 2020, MNRAS, 494, 3728
Springel V., 2005, MNRAS, 364, 1105
Springel V., 2010, MNRAS, 401, 791
Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726
Springel V. et al., 2018, MNRAS, 475, 676
Stadel J. G., 2001, PhD thesis, University of Washington
Stein P., 1997, A&A, 317, 670
Teyssier R., 2002, A&A, 385, 337
The Dark Energy Survey Collaboration, 2005, preprint (arXiv:astro–ph/0510346)
Tinker J. L., 2007, MNRAS, 374, 477
To C. et al., 2021, Phys. Rev. Lett., 126, 141301
Tonegawa M., Park C., Zheng Y., Park H., Hong S. E., Hwang H. S., Kim J., 2020, ApJ, 897, 17
van der Walt S., Colbert S. C., Varoquaux G., 2011, Comput. Sci. Eng., 13, 22
Virtanen P. et al., 2020, Nat. Methods, 17, 261
Vogelsberger M. et al., 2014, Nature, 509, 177
Weinberger R. et al., 2017, MNRAS, 465, 3291
Wu H.-Y., Hahn O., Evrard A. E., Wechsler R. H., Dolag K., 2013, MNRAS, 436, 460
Wu H.-Y., Weinberg D. H., Salcedo A. N., Wibking B. D., 2021, ApJ, 910, 28
Yahil A., Vidal N. V., 1977, ApJ, 214, 347
Ye J.-N., Guo H., Zheng Z., Zehavi I., 2017, ApJ, 841, 45
Yuan S., Eisenstein D. J., Garrison L. H., 2018, MNRAS, 478, 2019
Zhai Z. et al., 2019, ApJ, 874, 95
Zhang Y. et al., 2019, MNRAS, 487, 2578
Zwicky F., 1937, ApJ, 86, 217

APPENDIX A: RESOLUTION EFFECTS

The simulations we consider in our analysis all have different astrophysical model prescriptions, but they also have different resolutions, and it is difficult to fully disentangle how much each effect contributes to the overall differences we observe between simulations (e.g. Fig. 4). To shed light on the impact of resolution, we use the TNG300 suite, which has three different resolution runs – TNG300-1, TNG300-2, and TNG300-3 – to test the resolution dependence of our results. Conveniently, the resolutions of TNG300-2 and TNG300-3 are approximately that of MGTM and BM, respectively. Note that TNG300-1 is the fiducial run that we have used throughout our main analysis.

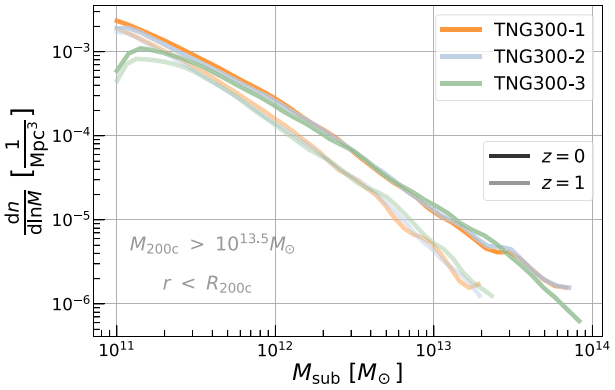

Figure A1. The conditional satellite subhalo mass function for TNG300 runs of different resolution (colours) for two redshifts (tones). We only show subhaloes within R_{200c} of host haloes with $M_{200c} > 10^{13.5}\,\mathrm{M}_\odot$. The two higher resolution runs converge to the same answer across the whole mass range shown here, whereas the lowest resolution run begins diverging below $M_{\rm sub} < 10^{11.5}\,\mathrm{M}_\odot$.

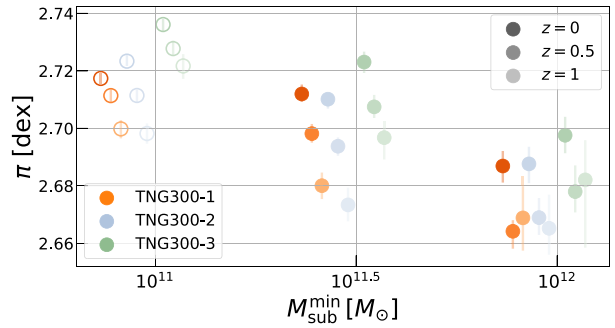

Figure A2. The normalization of the satellite subhalo velocity dispersion scaling with host halo mass, for different subhalo mass thresholds (denoted here by $M_{\rm sub}^{\rm min}$), and for the three different resolutions of the TNG300 suite. The normalizations of the TNG300-1 and TNG300-2 runs are quite similar, whereas the TNG300-3 run has a consistently larger normalization. The results for $M_{\rm sub} > 10^{11}\,\mathrm{M}_\odot$ are shown as open symbols to highlight that the different resolution runs have significant divergence in the satellite subhalo mass function at this subhalo mass, as shown in Fig. A1.

For this analysis, we focus only on the subhaloes, not galaxies. The stellar mass in galaxies is lower in the lower resolution runs (Pillepich et al. 2018b) because the TNG astrophysical model parameters are fixed for all runs and are not re-tuned for each resolution level (Pillepich et al. 2018a). Thus, the three TNG300 runs have different S-GSMFs, and a resolution test that uses galaxy properties would be affected by the absence of the recalibration step. As an alternative, we use the subhalo mass, and limit this study to masses above which the satellite subhalo mass function of the three TNG300 runs, shown in Fig. A1, has approximately converged.

Upon performing EVL for subhaloes, we find clear evidence that the normalizations increase with decreasing resolution (Fig. A2). The TNG300-1 and TNG300-2 runs show consistency amongst one another for all redshifts, but TNG300-3 differs significantly from these two. This is particularly interesting, as it mirrors our main result

where MGTM and TNG300 show similar normalizations, but BM has a significantly higher one. Given that TNG300-2 and TNG300-3 share similar resolution scales with MGTM and BM, respectively, it is possible – though not necessary – that a significant part of the discrepancy in BM arises from just resolution differences. The slopes and scatters of the different runs (not shown here) are statistically consistent with one another.

Note that the velocity bias is a ratio between the galaxy and DM velocity dispersions and thus will be unaffected by resolution *as long as* both dispersions are similarly biased by resolution effects. However, in this section, we are not studying the impact of resolution on b_v, but rather pointing out that simulations that exist in the high-normalization end of our $\sigma_{\rm sat,1D}$ results (Fig. 4) may suffer from potential resolution issues.

APPENDIX B: ADDITIONAL MODEL TESTS

Here, we detail two additional tests of the EVL model. We have assumed throughout our main analysis that our scaling parameters do not run with halo mass, and we validate this assumption in Section B1. Next, in Section B2, we demonstrate the power of our likelihood method in obtaining constraints even from very sparsely populated data

B1 Mass-dependence of galaxy EVL scaling parameters

In our likelihood model, we assume that the slope α and scatter ϵ are independent of host halo mass $M_{\rm 200c}$. We explicitly test this assumption by measuring the scaling relation parameters only using host haloes within mass bins of width 0.2 dex. The results are shown in Fig. B1, where the normalization is still quoted at a scale of $h(z)M_{\rm 200c} = 10^{14}\,{\rm M}_\odot$. There are two sets of grey symbols (open and closed) – the closed grey symbols are results from the fiducial MDPL2 sample while the open grey symbols are results from the MDPL2 sample that has been modified to replicate the incomplete mass function of THE300. The latter is constructed by selecting all MDPL2 haloes of masses below which The300 sample is incomplete, and then preferentially selecting only those MDPL2 haloes that are within 22 (comoving) Mpc of the larger haloes from the mass-complete part of the MDPL2 halo sample. We have verified the consistency between the mass functions of The300 and the modified MDPL2 sample.

From our analysis we are able to make two claims: (i) TNG300, MGTM, and BM show no clear mass dependence in the parameters. The variations are either stochastic, or are within the errorbars of the measurements. (ii) The increased scatter in The300 at low halo mass can be mimicked by a similar sample constructed using MDPL2. While we cannot make any robust claim on the *cause* of the increased scatter, this result implies the underlying cause is at least *correlated* to a selection effect on the local environment of the low-mass haloes.

The slopes and normalizations shown in Fig. B1 are all statistically consistent with one another, with no preference for any mass evolution. Note that this is the case even for the modified MDPL2 sample. We take this as validation that THE300 sample's mean $\sigma_{\rm sat,1D}$–$M_{\rm 200c}$ relation is quite insensitive to the environment-based sample selection for low-mass haloes, whereas the scatter is clearly impacted by this. However, given the velocity bias only depends on the mean relation – and not the scatter – we continue using the entire mass range of THE300 in our main analysis.

B2 Sparse sampling

In our main analysis, we showed that the likelihood estimator provides result consistent with those from commonly used regression methods, in the limit $N_{\rm part} \gg 1$ (Fig. 2). Here, we test how well the likelihood method does when we downsample the data set. We perform the exact same analysis as in Fig. 2, but instead of using $N_{\rm part} = 100$ DM particles, we use only $N_{\rm part} = 10$ particles. We also focus on the mean relation, $\sigma_{\rm DM}$–$M_{\rm 200c}$ instead of its slope and scatter.

We find that the results from the likelihood estimator are statistically consistent with those from KLLR (Fig. B2). Even the upper

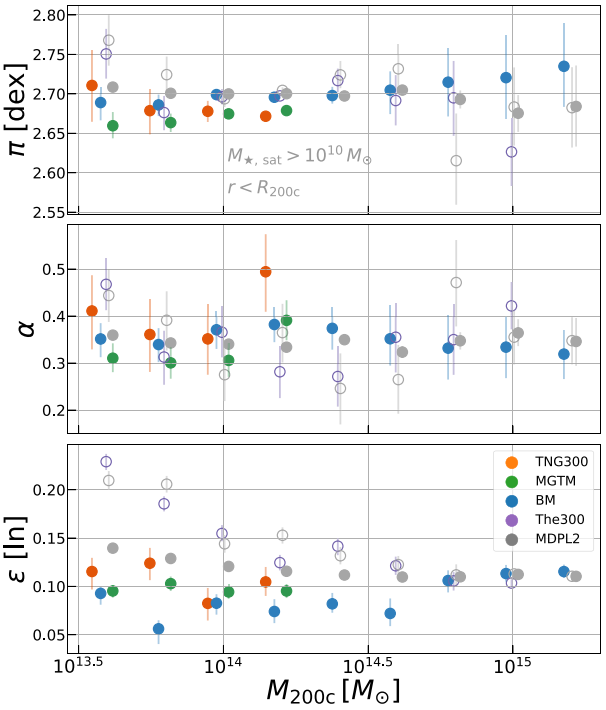

Figure B1. Evolution of the scaling parameters as a function of *halo* mass. We choose bins of width $\Delta M = 0.2$ dex centered on an $M_{\rm 200c}$ mass scale and compute the parameters using only haloes in those bins. Cosmologically mass-complete (mass-incomplete) simulations are shown with closed (open) circles. MDPL2 has both the original mass-complete sample, and a modified sample that mimics the mass-incompleteness resulting from the zoom-in nature of The300 simulations. The normalizations and slopes are statistically consistent with no mass-dependence. For the scatter, however, THE300 shows a strong dependence on halo mass, and the 'incomplete' MDPL2 sample captures this behaviour very well.

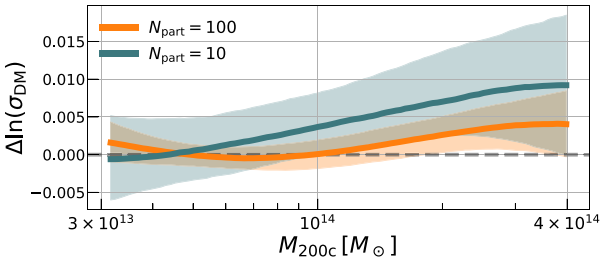

Figure B2. The fractional difference between the EVL-derived and KLLR-derived estimate of the $\sigma_{\rm DM}$–$M_{\rm 200c}$ relation. For the likelihood, we present two versions – one where we use $N_{\rm part} = 100$ particles, and another where we use only $N_{\rm part} = 10$ particles.

bounds of the $N_{\rm part} = 10$ result are at most within 1.5 per cent of the KLLR estimate.

APPENDIX C: SENSITIVITY TO APERTURE

For our main analysis, we have computed the velocity dispersion (both for galaxies and DM) with an aperture defined by the radius $R_{\rm 200c}$. While this is a common choice for observational work on galaxy velocity dispersions (e.g. Sifón et al. 2016, F16), it is still a potential free parameter in future analyses. Here, we redo our analysis of TNG300 but replace this aperture choice with two other commonly used radii for cluster-scale haloes, $R_{\rm 500c}$ and $R_{\rm 200m}$. Both differ from $R_{\rm 200c}$ in only the density contrast used in the radius definition, with $R_{\rm 500c}$ using $\rho_\Delta = 500\rho_c(z)$, and $R_{\rm 200m}$ using $\rho_\Delta = 200\rho_m(z)$. Note that our mass variable continues to be $M_{\rm 200c}$, as we are not studying the impact of changing the spherical overdensity definition, but only of changing the satellite galaxy (and DM particle) sample.

We find that using smaller apertures leads to a higher velocity bias (Fig. C1), and this trend was noted in previous studies of both observations (Sifón et al. 2016) and simulations (Lau et al. 2010; Armitage et al. 2018; Ferragamo et al. 2020). We also find that the relative difference between the b_v from different apertures can depend on $M_{\star,\rm sat}$, $M_{\rm 200c}$, and z. This dependence arises from changes in the normalization and slope of the $\sigma_{\rm sat,1D}$–$M_{\rm 200c}$ relation due to varying the aperture of the measurements. Note also that at high redshift, where we have $\rho_m(z) \approx \rho_c(z)$ and so $R_{\rm 200m} \approx R_{\rm 200c}$, the bias computed within both radii are statistically consistent as expected.

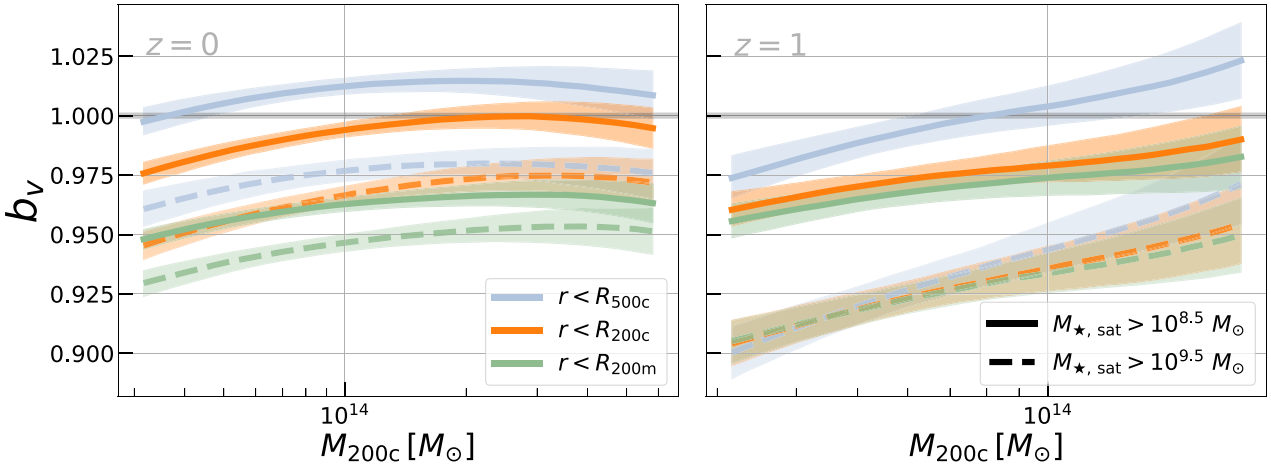

Figure C1. Similar to Fig. 6, but for a single TNG300 sample with different aperture choices. Smaller apertures lead to a higher b_v, and the relative differences between the b_v of different apertures can depend on all of $M_{\star,\mathrm{sat}}$, M_{200c}, and z.

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