

Columba: isolated dwarf galaxy populations in diverse cosmological environments simulated with a cold interstellar medium

Jemima M. Briggs^{1,2,★}, Azadeh Fattahi^{1,2,3,★}, Robert A. Crain^{1,★}, Yannick M. Bahe^{1,4,5},
 Sylvia Ploekinger^{1,6}, Matthieu Schaller^{1,7,8}, Alejandro Benítez-Llambay^{1,9}, Evgenii Chaikin^{1,8},
 Alexander J. Richings^{10,11} and Joop Schaye^{10,11}

¹Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK

²Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK

³The Oskar Klein Centre, Department of Physics, Stockholm University, Albanova University Center, SE-106 91 Stockholm, Sweden

⁴School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, UK

⁵Institute of Physics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, CH-1290 Versoix, Switzerland

⁶Department of Astrophysics, University of Vienna, Türkenschanzstrasse 17, A-1180 Vienna, Austria

⁷Lorentz Institute for Theoretical Physics, Leiden University, PO Box 9506, NL-2300 RA Leiden, the Netherlands

⁸Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands

⁹Dipartimento di Fisica G. Occhialini, Università degli Studi di Milano Bicocca, Piazza della Scienza 3, I-20126 Milano MI, Italy

¹⁰Centre for Data Science, Artificial Intelligence and Modelling, University of Hull, Cottingham Road, Hull HU6 7RX, UK

¹¹E. A. Milne Centre for Astrophysics, University of Hull, Cottingham Road, Hull HU6 7RX, UK

Accepted 2026 June 18. Received 2026 June 18; in original form 2026 February 11

ABSTRACT

We introduce a suite of Lambda cold dark matter (Λ CDM) cosmological, hydrodynamical simulations that track the evolution of a large population of dwarf galaxies. The suite comprises zoom-in simulations of 25 spherical, underdense regions of $r = 5$ cMpc, selected to span ≈ 1.5 dex in mean enclosed density, covering voids to filamentary structures, whilst excluding haloes of Milky Way-mass or larger. The simulations achieve a mass resolution of $\sim 10^5 M_\odot$ with a galaxy formation model including cold, dense interstellar gas and whose subgrid stellar feedback efficiency reproduces the $z = 0$ galaxy stellar mass function. We investigate the impact of the cosmic environment on dwarf galaxy formation and evolution. We find that the 5 cMpc environment influences the normalization of the halo and galaxy mass functions, but does not significantly affect the stellar mass–halo mass (SMHM) relation and halo occupation fraction for galaxies with $M_\star = 10^6 - 10^9 M_\odot$. Instead, host halo concentration, estimated from DM-only counterparts, is more important: both the fraction of haloes hosting a resolved galaxy and the scatter about the SMHM relation correlate positively with concentration. Owing to halo assembly bias, concentration also influences galaxy formation times, such that at fixed halo mass more concentrated haloes host galaxies that are both older and more massive. The offset from the mean SMHM relation also anticorrelates with t_{90} , the time at which 90 per cent of a galaxy’s stellar mass has assembled. These correlations between halo properties and galaxy star formation histories present testable predictions for forthcoming observational surveys.

Key words: methods: numerical – galaxies: dwarf – galaxies: evolution – galaxies: formation – galaxies: haloes.

1 INTRODUCTION

Understanding the formation and evolution of dwarf galaxies is a problem at the frontiers of both cosmology and astrophysics. A number of outstanding challenges to the prevailing Lambda cold dark matter (Λ CDM) cosmogony on dwarf galaxy scales are still debated in the literature, such as the ‘missing satellites’ (A. Klypin et al. 1999; B. Moore et al. 1999), ‘cusp-core’ (e.g. R. A. Flores & J. R. Primack 1994), and ‘too big to fail’ (M. Boylan-Kolchin, J. S. Bullock & M. Kaplinghat 2011) problems (see J. S. Bullock & M. Boylan-Kolchin 2017, for a review). These open questions are

often cited as an indication that the Λ CDM paradigm requires revision or extension, but their solution may also follow from an improved understanding of how dwarf galaxies form and evolve, and the concomitant influence of these processes on the structure of dwarf galaxy dark matter (DM) haloes (see e.g. L. V. Sales, A. Wetzel & A. Fattahi 2022).

Dwarf galaxies are of particular interest as ‘laboratories’ for studying the astrophysics of galaxies, as their early formation times and shallow gravitational potentials render them particularly sensitive to the physical processes that regulate galaxy growth. These include both internal mechanisms such as radiative and mechanical feedback from young stars and supernovae (SNe), and external influences such as the ultraviolet (UV)/X-ray ionizing background radiation or interactions with a more massive host galaxy. SN feedback has long been shown to play a

* E-mail: J.M.Briggs@2016.ljmu.ac.uk (JMB); Azadeh.Fattahi@fysik.su.se (AF); R.A.Crain@ljmu.ac.uk (RAC)

particularly important role in removing baryons from low mass haloes and hence reconciling the steep mass function of dark matter haloes with the shallower galaxy stellar mass function (GSMF; e.g. A. Dekel & J. Silk 1986; C. Lacey & J. Silk 1991; R. G. Bower, A. J. Benson & R. A. Crain 2012) in the dwarf galaxy regime. The UV/X-ray ionizing background radiation is expected to evaporate gas from low-mass haloes (those with maximum circular velocity, $V_{\max} \lesssim 30 \text{ km s}^{-1}$) leaving a significant fraction of them devoid of baryons at the present day (T. Quinn, N. Katz & G. Efstathiou 1996; P. R. Shapiro, I. T. Iliev & A. C. Raga 2004; R. A. Crain et al. 2007) and quenching star formation in the small fraction of haloes of this scale that were able to initiate star formation prior to the epoch of reionization (e.g. M. Ricotti & N. Y. Gnedin 2005).

Almost all ultra-faint dwarf galaxies (those with $M_{\star} \lesssim 10^5 M_{\odot}$) are believed to be relics of the early Universe, having formed almost entirely prior to the epoch of reionization (see J. D. Simon 2019, for a review). More massive dwarf galaxies are believed to ‘reignite’ after reionization (A. Benítez-Llambay et al. 2015). These predictions are consistent, at face value, with observations of dwarf galaxies in and around the Local Group. Indeed, ultra-faint dwarf galaxies are mostly old and metal-poor, whereas brighter dwarf galaxies exhibit diverse star formation histories (D. R. Weisz et al. 2011, 2014; C. Gallart et al. 2015).

The sensitivity of Local Group dwarf galaxies to their environment is apparent in the dichotomy of the morphology of dwarf satellite galaxies and more isolated counterparts. The majority of the satellites of the Milky Way (MW) and M31 are gas-poor dwarf spheroidal galaxies, but the isolated dwarf galaxies beyond the immediate environment ($r \gtrsim 300 \text{ kpc}$) of the Milky Way and M31 exhibit significant neutral hydrogen reservoirs and are star forming (see A. W. McConnachie 2012, and references therein). These trends extend to galaxies beyond the Local Group. M. Geha et al. (2012) find a correlation between the fraction of quenched dwarf galaxies and the distance to the nearest massive galaxy using Sloan Digital Sky Survey (SDSS) data, finding that isolated dwarf galaxies in the mass range $10^7 M_{\odot} < M_{\star} < 10^9 M_{\odot}$ are seldom quenched.¹ On the other hand, satellite dwarf galaxies around MW-like hosts are efficiently quenched, as shown by recent results from the Satellites Around Galactic Analogs (SAGA) survey (M. Geha et al. 2024).

Aside from the immediate neighbourhood, the environment defined on larger scales has also been shown to influence properties of galaxies and DM haloes. The clustering of DM haloes depends on secondary halo properties such as their concentration and formation time, in particular for low mass haloes, with more concentrated and older low mass haloes more likely to be found in denser environments, as demonstrated in simulations (L. Gao, V. Springel & S. D. M. White 2005; R. H. Wechsler et al. 2006). Similarly, W. A. Hellwing et al. (2021) predict that low mass ($M_{200} \lesssim 10^{11} M_{\odot}$) haloes in voids exhibit systematically lower concentrations than is typical at fixed mass. The observable properties of galaxies also have been shown to depend on the larger scale environment. For example, I. Zehavi et al. (2011) find that red galaxies in the SDSS redshift survey are more clustered, i.e. live in higher density environments. Additionally, semi-analytical and hydrodynamical simulations show that the overdensity in which a halo resides (on $\simeq 5 \text{ Mpc}$ scales) correlates with the residual of

the stellar mass–halo mass (SMHM) relation about the mean (M. C. Artale et al. 2018; I. Zehavi et al. 2018).

The degree to which environmentally driven biases of DM haloes influence the observable properties of dwarf galaxies remains poorly understood, due to both observational and theoretical challenges. The majority of known low-mass dwarf galaxies are located in the Local Group with the completeness dropping sharply beyond $\sim 3 \text{ Mpc}$. The structural properties of dwarf galaxies beyond the Local Group are also poorly constrained owing to their intrinsic faintness. Recent attempts to extend the observed dwarf galaxy sample, such as the SAGA (M. Geha et al. 2017; Y.-Y. Mao et al. 2021) or Exploration of Local Volume Satellites (ELVES; S. G. Carlsten et al. 2022) surveys, have focused on satellites of MW-like systems, where interaction with the host may significantly influence their properties and hinder investigation of the role of other internal or external physical processes. These observational hurdles have resulted in limited understanding of dwarf galaxies as a population, and even fundamental scaling relations, such as the average stellar mass as a function of halo mass, are poorly constrained below $M_{\star} < 10^8 M_{\odot}$ (e.g. J. I. Read et al. 2017). Nevertheless, this situation is rapidly changing. For example, the Euclid Deep Field, DR1, and the Euclid Wide Survey are estimated to identify approximately 9900, 470 000, and 2.6 million dwarf galaxy candidates (F. R. Marleau et al. 2025), respectively. These developments make it increasingly important to understand the formation of isolated dwarf galaxies in diverse environments, and to develop theoretical predictions that can be confronted with these new data sets.

Cosmological hydrodynamical simulations have proven a powerful means of obtaining insight into the mechanisms that govern the formation and evolution of dwarf galaxies, offering plausible solutions to some of the small-scale tensions faced by the Λ CDM cosmogony (L. V. Sales et al. 2022). Simulations of representative cosmic volumes (side length $L \sim 100 \text{ comoving Mpc}$, hereafter cMpc) have, over the last decade, matured such that they yield galaxy populations with average properties that broadly correspond to those of the observed population (see e.g. R. A. Crain & F. van de Voort 2023). However, the numerical study of dwarf galaxies necessitates high resolution in order to capture the physical mechanisms that govern their evolution and to sample their structure and star formation history (SFH). This unavoidably restricts simulations to small volumes and therefore limits the size of simulated dwarf galaxies samples and the diversity of the simulated environments they are found in.

To achieve high resolution in simulations at a reasonable computational cost, the zoom-in method is commonly adopted (e.g. N. Katz & S. D. M. White 1993), wherein the immediate environment of an individual galaxy is followed with gas dynamics at high resolution, and the remaining cosmological volume is evolved with collisionless dynamics at reduced resolution to ensure the galaxy experiences the correct tidal forces. This method has been used to resimulate Milky-Way and Local Group-like analogues, including populations of field and satellite dwarf galaxies in these environments, in projects such as (but not limited to) Eris (J. Guedes et al. 2011), APOSTLE (T. Sawala et al. 2016), Auriga (R. J. J. Grand et al. 2017), ARTEMIS (A. S. Font et al. 2020), the D.C. Justice League Simulations (E. Applebaum et al. 2021), NIHAO-UHD (T. Buck et al. 2020), Latte (A. R. Wetzel et al. 2016), and ELVIS (S. Garrison-Kimmel et al. 2014). Zoom simulations have also been widely used to follow individual dwarf galaxies in isolation at extremely high resolution, enabling ultra-faint dwarf

¹The study stops at lower mass end of $M_{\star} \sim 10^7 M_{\odot}$ due to incompleteness of data.

galaxies to be simulated e.g. the EDGE project (M. P. Rey et al. 2019, 2020), LYRA (T. A. Gutcke et al. 2021), and dwarf galaxies from FIRE (C. Wheeler et al. 2019).

However, due to their small volume, zoom simulations of individual dwarf galaxies are ill-suited to the study of the influence of diverse cosmic environments. This limitation is best overcome with zoom simulations of regions that are significantly larger than individual galaxies, without incurring an infeasible computational cost and memory footprint. This approach was adopted for the general galaxy population by the Galaxies-Intergalactic Medium Interaction Calculation (GIMIC) simulations (R. A. Crain et al. 2009), which follow the evolution of galaxies within five roughly spherical regions of comoving radius $r = 18 - 25 h^{-1} \text{cMpc}$ drawn from the $L = 500 h^{-1} \text{cMpc}$ Millennium Simulation volume. The regions were chosen to have overdensities of $(-2, -1, 0, +1, +2)$ times the root-mean-square deviation from the mean overdensity, and thus span $\simeq 95$ per cent of the overdensity range in the parent volume whilst simulating only 0.13 per cent of its volume. A similar approach was adopted by the First Light And Reionization Epoch Simulations (FLARES) project (C. C. Lovell et al. 2021) to examine diverse environments in simulations of the epoch of reionization, and by the MARVELOUS project to study dwarf galaxies in the environment defined by a cosmic ‘sheet’ (J. M. Bellovary et al. 2019; F. Munshi et al. 2021).

Here, we introduce the Columba suite of simulations of galaxy formation in the Λ CDM cosmogony, which are designed primarily to examine the influence of large-scale cosmic environment on the evolution of isolated dwarf galaxies, i.e. those not in close proximity to larger structures at the present day. We focus on this subset of dwarf galaxies in order to explore the influence of large-scale environment, and the intrinsic properties of host DM haloes, on the evolution of dwarf galaxies, in the absence of tidal and ram pressure forces exerted by nearby neighbouring galaxies.

Columba adopts a similar approach to GIMIC; the Columba suite comprises zoom simulations of 25 spherical regions of radius $r = 5 \text{cMpc}$ drawn from a large parent volume of $L = 400 \text{cMpc}$. The main motivation for these simulations is studying *isolated* dwarf galaxies in various environments such as voids and filaments. As the simulations focus on isolated dwarf galaxies, these regions do not sample the cosmic density distribution (on the 5cMpc scale of the simulated spheres) uniformly, but preferentially sample underdense regions. This approach enables a large sample of isolated dwarf galaxies, found in regions spanning 1.5 decades in enclosed density, to be followed with resolution better than galaxy formation simulations of large ($L \sim 100 \text{Mpc}$) periodic volumes such as EAGLE (R. A. Crain et al. 2015; J. Schaye et al. 2015; S. McAlpine et al. 2016), IllustrisTNG (D. Nelson et al. 2018; A. Pillepich et al. 2018a; D. Nelson et al. 2019), or SIMBA (R. Davé et al. 2019).

The galaxy formation model used for Columba adopts several subgrid modules developed for the COLd Ism and Better REsolution (COLIBRE) project (E. Chaikin et al. 2026; J. Schaye et al. 2026), combined with elements retained from the earlier Evolution and Assembly of GaLaxies and their Environments (EAGLE) model. Columba therefore shares COLIBRE’s treatments for processes such as turbulence-based star formation, pre-SN feedback, and black hole (BH) seeding and growth, but does not include all of the new physics implemented in COLIBRE. In particular, Columba lacks the live dust model, the turbulent diffusion, and active galactic nucleus (AGN) jet feedback adopted in COLIBRE, and it retains the nucleosynthesis yields and Type Ia SN rates from EAGLE. Columba nevertheless follows many of the key processes that regulate the formation of

the cold, dense interstellar medium, including shielding from UV radiation, the formation and dissociation of molecules, and the radiative cooling that such molecules foster (specific differences relative to the COLIBRE model are detailed in the next section). Cosmological simulations of volumes larger than the environments of individual galaxies have recently begun to treat the cold interstellar medium (ISM) phase explicitly, with notable examples being NewHorizon (Y. Dubois et al. 2021) and FIREbox (R. Feldmann et al. 2023). Our approach complements these studies with an explicit focus on diversity in cosmological environments, which cannot currently be achieved with simulations of uniform resolution (FIREbox) or a single zoom-in region (NewHorizon).

This study serves as a reference for the simulation suite, and presents a number of key findings concerning the connection between the cosmic environment and dwarf galaxy formation and evolution. These findings are in general most clearly demonstrated with tailored simulations of this type. The paper is structured as follows. We begin by presenting in Section 2 our methods, including our strategy for selecting the simulated regions, the generation of the initial conditions (ICs), the numerical and subgrid galaxy formation components of the model used to evolve the ICs to the present day, and our techniques for identifying galaxies and their parent haloes. In Section 3, we confront basic properties of the simulated galaxies with observational measurements to validate our model. We examine the environmental variation of key diagnostics in Section 4, and explore the origin of scatter in the SMHM relation in Section 5. We summarize and discuss our findings in Section 6.

2 METHODS

In this section, we introduce our methods for selecting the simulated regions (Section 2.1) and generating the ICs (Section 2.2), describe the model we use to evolve the ICs to the present day (Sections 2.3 and 2.4), and explain our techniques for identifying galaxies and their parent haloes (Section 2.5).

2.1 Selection of simulated regions

The primary aim of our simulation design is to follow the evolution of dwarf galaxies ($M_* \lesssim 10^{10} M_\odot$) that are isolated at the present day, and at the same time live in a diverse range of large-scale cosmic environments. We identify such environments at $z = 0$ within a parent simulation of $L = 400 \text{cMpc}$ with periodic boundary conditions and its mass distribution realized with $N = 3008^3$ composite particles representing baryons and DM. The simulation adopts the maximum posterior likelihood cosmological parameter values reported by the Dark Energy Survey team from year three data (T. M. C. Abbott et al. 2022, their ‘ $3 \times 2\text{pt} + \text{All Ext.}$ ’ CDM parameters), for which $\Omega_m = 0.306$ yielding a particle mass of $m_{\text{dm}} = 9.22 \times 10^7 M_\odot$. This parent simulation uses the dark matter only initial conditions (ICs) corresponding to the COLIBRE L400m7 volume (J. Schaye et al. 2026), generated with the same cosmology and phases but without the $4 \times$ sampling of dark matter particles used for the COLIBRE ICs. Consequently, the DM particle mass differs from the published L400m7 run by a factor of ≈ 4 . The ICs of this simulation were generated following the same procedure as that adopted by J. Schaye et al. (2023, see their section 2.4) for the FLAMINGO simulations, using the third-order Lagrangian perturbation software MONOFONIC (O. Hahn et al. 2020), which suppress discreteness

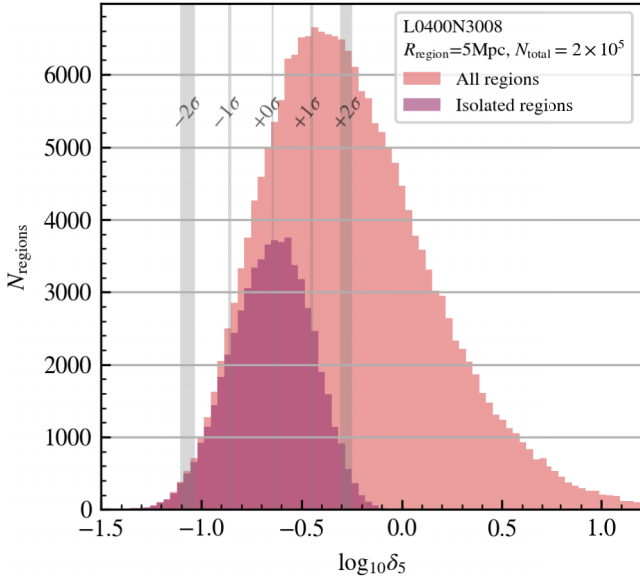


Figure 1. The distribution of mean enclosed density, relative to the cosmic mean density, of $r = 5$ cMpc spheres, $\delta_5 \equiv \rho_5/\bar{\rho}$, selected from a parent cubic volume of side length $L = 400$ cMpc, at $z = 0$. The background distribution in a lighter shade corresponds to 2×10^5 randomly centred spheres, the foreground distribution denotes the subset with no haloes of $M_{200} > 5 \times 10^{11} M_\odot$ within $r = 5$ cMpc of the region centre, and none of mass $M_{200} > 10^{13} M_\odot$ within $r = 10$ cMpc. These criteria necessarily bias this subset towards lower density. The vertical bands denote the densities of this subset corresponding to the $(-2, -1, 0, +1, +2)\sigma$ percentile rank ordering of a Gaussian distribution. The width of the bands corresponds to one percentile about these values.

noise, coupled to the PANPHASIA (A. Jenkins 2013) Gaussian random noise field.² The large-scale phases were fixed to the mean. As for FLAMINGO, we adopt linear power and transfer functions computed with CLASS (J. Lesgourgues 2011; J. Lesgourgues & T. Tram 2011), with separate transfer functions computed and applied for baryons, DM and neutrinos. The ICs were created at $z = 127$.³ and evolved to $z = 0$ assuming collisionless dynamics using the open-source cosmology, gravity, hydrodynamics, and galaxy formation software SWIFT (M. Schaller et al. 2024). The short- and long-range gravitational forces are computed using a fourth-order fast multipole method and a particle-mesh method solved in Fourier space. The influence on large-scale matter clustering due to the non-zero density of massive neutrinos (e.g. J. Lesgourgues & S. Pastor 2006) is modelled with a grid-based linear response method (e.g. Y. Ali-Haïmoud & S. Bird 2013). Haloes are identified in the parent simulation using the VELOCIRAPTOR structure finder (P. J. Elahi et al. 2019), which is further described in Section 2.5.

We elect to simulate spherical regions of $r = 5$ cMpc. The precise size of the regions is not critical, but we find that this choice yields a judicious balance between probing the ‘extremes’ of the density distribution populated by isolated dwarf galaxies, whilst ensuring that even the most underdense regions yield a meaningful population of well-sampled galaxies. A related benefit of

simulating relatively small volumes (compared to typical periodic simulations) is that they can usually be accommodated within the memory footprint (512 gigabytes to 1 terabyte) of the individual compute nodes that comprise many modern high-performance computing facilities, simplifying parallelism.

To select the Columba target regions, we sample the $z = 0$ normalized density⁴, $\delta_5 = \rho_5/\bar{\rho}$, where ρ_5 is the mean mass density within a sphere of radius $r = 5$ cMpc and $\bar{\rho}$ is the mean cosmic mass density, with 2×10^5 spheres randomly centred within the parent volume. The total sampled volume is roughly twice the volume of the parent box and this ensures that most coordinates within the parent volume are sampled at least once. The resulting density distribution is shown as the lightly shaded histogram in Fig. 1. Note that this unbiased sample yields, for a sufficiently large sample of regions, a mean density of $\bar{\rho}$, but as a volume-weighted measure the distribution is not symmetric and the median and peak values are less than unity ($\log_{10} \delta_5^{\text{median}} = -0.304$, $\log_{10} \delta_5^{\text{peak}} \simeq -0.434$).

The darker, foreground histogram shows the density distribution of the subset of regions that are devoid of haloes with mass comparable to or more massive than that of the Milky Way, and which are not in close proximity to a massive halo that would host a group or cluster of galaxies. Specifically, we require that at $z = 0$ there are no haloes of mass⁵ $M_{200} > 5.0 \times 10^{11} M_\odot$ within 5 Mpc of the region centre, and no halo of mass $M_{200} > 10^{13} M_\odot$ within 10 Mpc of the centre. These criteria are satisfied by 5.8×10^4 of the initial 2×10^5 regions (29 per cent), and are necessarily biased towards lower densities, yielding mean, median and peak normalized densities of $\log_{10} \delta_5 = -0.653, -0.645$, and -0.568 , respectively. Vertical bands on Fig. 1 denote for this subsample the densities corresponding to the $(-2, -1, 0, +1, +2)\sigma$ percentile rank ordering of a Gaussian distribution, which have values $\log_{10} \delta_5 = (-1.064, -0.859, -0.645, -0.449, -0.282)$, respectively. The width of the lines corresponds to 1 percentile. For reference, the same percentile ranks for the full sample (lightly shaded histogram) correspond to $\log_{10} \delta_5 = (-0.954, -0.671, -0.304, 0.162, 0.732)$, respectively.

Despite its inherent bias towards low density regions, the sample of regions with no MW-mass haloes nevertheless exhibits significant diversity in the number of haloes enclosed that have dynamical mass comparable to that of the Large Magellanic Cloud (LMC). Such haloes are amongst the most massive permitted by our selection criteria, so the diversity is necessarily greatest for those regions at the upper end of the density distribution. Fig. 2 shows a two-dimensional histogram of the number of regions as a function of enclosed density, δ_5 , and the number of LMC-mass haloes, N_{LMC} , where the latter is defined as $7 \times 10^{10} \leq M_{200}/M_\odot \leq 5 \times 10^{11}$. N_{LMC} correlates positively with δ_5 , varying between zero and 22 in the 5.8×10^4 regions defining our MW-free sample, but with large scatter overall. We characterize the diversity of N_{LMC} as a function of δ_5 for the MW-free sample of regions by computing, for haloes within the 1 percentile-wide density bins denoted by the vertical bands (repeated from Fig. 1), the values corresponding to the Gaussian $(-2, -1, 0, +1, +2)\sigma$ percentile rank ordering of N_{LMC} . These values are denoted by white

²The phase descriptor of this volume is [Panph6,L18,(200557,163876,161484),S1,KK1025,CH3518244376,COLIBRE400].

³Colibre L400m7 ICs are at $z = 63$.

⁴Note that δ is often used to represent the *overdensity*, i.e. $\delta = (\rho - \bar{\rho})/\bar{\rho}$, but here we use $\delta = \rho/\bar{\rho}$ because the large dynamic range of densities we explore is conveniently expressed logarithmically, requiring that $\delta \geq 0$.

⁵Spherical overdensity masses are defined relative to the critical density, ρ_c .

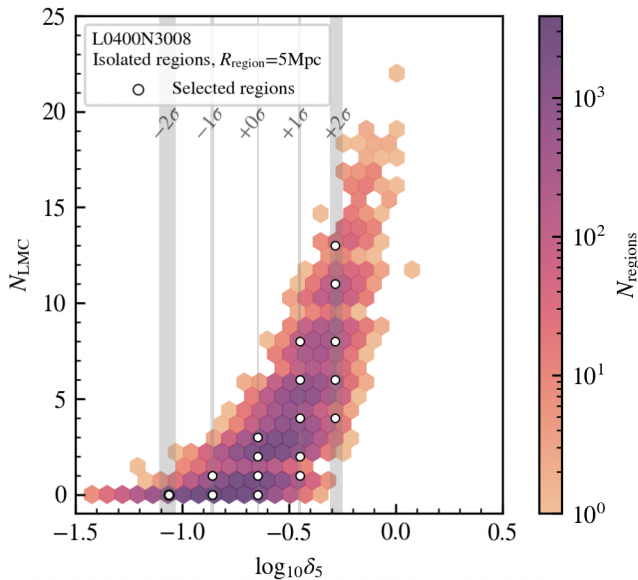


Figure 2. The number of LMC-mass objects ($M_{200} \sim 10^{11} M_{\odot}$) at $z = 0$ within the $r = 5$ Mpc underdense spheres comprising the darker histogram of Fig. 1, as a function of the regions’ normalized density, δ_5 . Hexbins are colour-coded by the number of regions. Denser regions tend to host more LMC-mass haloes, but there is significant scatter in N_{LMC} at fixed density. The vertical bands are repeated from Fig. 1, and indicate the bands of overdensity in which we select our target regions. The white circles show the values of N_{LMC} that correspond to $(-2, -1, 0, +1, +2)\sigma$ percentiles for the distribution of this quantity within each band. Where fewer than five circles per band are shown, more than one of them overlap.

dots overlaid on the vertical bands (where fewer than five dots are visible, two or more of the lowest- σ percentiles correspond to $N_{\text{LMC}} = 0$). For example, in the regions comprising the $+2\sigma$ density sample, the number of LMC-like haloes corresponding to $(-2, -1, 0, +1, +2)\sigma$ are $N_{\text{LMC}} = (4, 6, 8, 11, 13)$, respectively.

We select five regions from each of the $(-2, -1, 0, +1, +2)\sigma$ density bands, requiring that for each band the five regions exhibit N_{LMC} corresponding to the $(-2, -1, 0, +1, +2)\sigma$ values of the corresponding N_{LMC} distribution. For any pair of overdensity and N_{LMC} , we pick a region at random. Our sample therefore comprises 25 spherical regions of $r = 5$ cMpc selected from the $L = 400$ cMpc parent volume, and thus follows a total volume comparable to that of a periodic cube of side $L = 23.5$ cMpc. None of the selected spheres overlap. The centres of the 25 spheres are listed in Table A1, and their projected total mass density⁶ is shown in Fig. 3, with the rows, from top to bottom, corresponding to $(-2, -1, 0, +1, +2)\sigma$ in density, and the columns, from left to right, corresponding to $(-2, -1, 0, +1, +2)\sigma$ in N_{LMC} . The increasing average density from top to bottom is clearly evident.

2.2 Generation of zoomed initial conditions

We generate multiresolution ‘zoomed’ initial conditions of the selected regions embedded within the parent volume using the second-order Lagrangian perturbation theory (2LPT) software

IC_2LPT_GEN (A. Jenkins 2010), again coupled to PANPHASIA. The IC_2LPT_GEN software assumes that the cosmic matter density is comprised only of baryons and DM, therefore the zoomed ICs are generated with cosmological parameters that differ very slightly from those of the parent box, subsuming the matter density of massive neutrinos into that of the dark energy. The cosmological parameters of the zoomed ICs are therefore: $\Omega_{\text{CDM}} = 0.256$, $\Omega_{\text{b}} = 0.0486$, $\Omega_{\Lambda} = 0.695$, $h = 0.681$, $\sigma_8 = 0.807$, and $Y = 0.244$. We also make the simplifying assumption that baryons and DM follow the same total matter power spectrum, which is justified as we focus on spatial scales shorter than that on which the baryon power spectrum is suppressed (e.g. M. P. van Daalen et al. 2011). We generate the linear power spectrum with these parameters using CLASS.

The ICs of each selected region are generated by creating a uniform but multiresolution particle distribution that represents the unperturbed matter field, to which we apply the (linear + 2LPT) density perturbations of the parent volume supplemented by additional small-scale power down to the appropriate Nyquist frequency. The high-resolution component of each particle distribution has the morphology of the $z = 0$ matter distribution within a sphere of $r \simeq 6.5$ cMpc with the same centre as the selected region, when traced back to $z = \infty$. The high-resolution component therefore includes a padding region, corresponding to a spherical annulus of approximate⁷ width $r = 1.5$ cMpc, that ‘protects’ the selected region from (i) contamination by low-resolution composite particles and (ii) the influence of the artificial pressure gradient at the boundary of the high-resolution component. The high-resolution component is populated with four DM particles for every gas particle, in a face-replicating grid similar to that adopted by J. Richings et al. (2021). This yields gas and DM particles of approximately equal mass, suppressing the spurious transfer of kinetic energy that can artificially heat galaxy discs (A. D. Ludlow et al. 2019). The remainder of the $L = 400$ cMpc cube is populated with collisionless composite particles representing baryons and DM, whose mass increases with distance from the high-resolution region (a detailed explanation of this step is given in appendix A1.2 of R. A. Crain et al. 2009). The $z = 127$ density perturbations are applied by displacing the particles from their unperturbed coordinates using the method described by A. Jenkins (2010, 2013). We generate ICs for all 25 selected regions at two resolutions, corresponding to high-resolution particle masses of $\sim 10^5 M_{\odot}$ (‘m5’) and $\sim 10^6 M_{\odot}$ (‘m6’), to enable convergence testing. The initial H and He mass fractions are assumed to be $X_{\text{H}} = 0.756$ and $X_{\text{He}} = 0.244$, respectively. Given the changes between cosmology and IC code, we verified that the zoom ICs faithfully reproduce the structures identified in the parent simulation and preserve the rank ordering of regions by δ_5 and N_{LMC} . We also compare the total enclosed mass in $R = 5$ Mpc at $z = 0$ between the parent volume and ‘m5’ DM-only zoom simulations and find an average agreement of 0.7 per cent and up to 5 per cent across all regions.

2.3 Gravity and hydrodynamics

The zoomed ICs are evolved to $z = 0$ with SWIFT. As per the collisionless simulation of the parent volume, gravitational forces

⁶For superior clarity we show the surface mass density from the resulting DM-only zoom simulations, rather than the low-resolution parent volume.

⁷The high-resolution component’s geometry cannot be specified to arbitrary precision in the unperturbed particle distribution, as it is formed from tiles of particles each in a periodic grid arrangement.

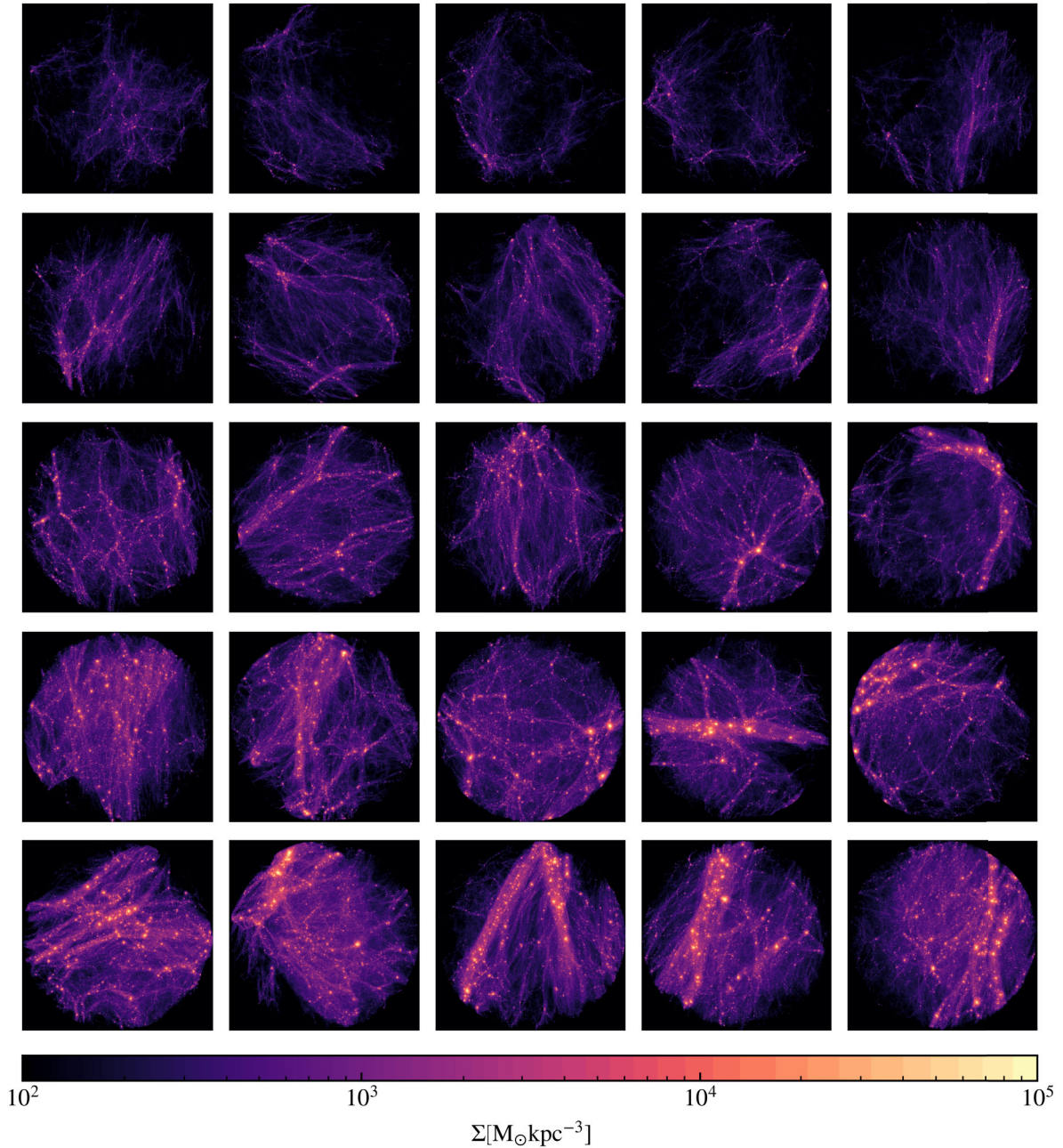


Figure 3. Projections of the total surface mass density in the dark matter only simulations for the 25 selected spherical regions of $r = 5$ cMpc at $z = 0$. From top to bottom each row corresponds to the $(-2, -1, 0, +1, +2)\sigma$ density band, respectively. From left to right, each column corresponds to the $(-2, -1, 0, +1, +2)\sigma$ value of N_{LMC} . For example, the bottom row shows the highest density regions that we simulate (the $+2\sigma$ sample), with median density $\log_{10} \delta_5 = -0.282$ and these regions exhibit, from left to right, $N_{\text{LMC}} = (4, 6, 8, 11, 13)$, respectively.

were solved using a split fast multipole plus particle-mesh approach. The hydrodynamical equations of motion of the gas were solved using the SPHENIX (J. Borrow et al. 2021) density-energy formulation of smoothed particle hydrodynamics (SPH) adopting the quartic spline kernel with 65 weighted neighbours within the kernel support radius (resolution parameter $\eta = 1.2348$). The SPH solver includes artificial viscosity and conduction terms, with built-in limiters to mitigate numerical energy losses. The time-step limiter of F. Durier & C. Dalla Vecchia (2012) is used to ensure accurate evolution of the fluid even during ex-

treme shock events. Time integration uses a leapfrog scheme in which a particle's time-step is the minimum of its gravitational time-step or, if applicable, that dictated by the hydrodynamical Courant–Friedrichs–Lewy condition with parameter $C_{\text{CFL}} = 0.2$. The gravitational time-step of a particle with acceleration a is $\Delta t = (0.025\epsilon/a)^{1/2}$, where ϵ is the gravitational softening length. The time-steps of gas particles cannot exceed that of any of their neighbours by more than a factor of 4. All high-resolution particles adopt the same softening length (see Table 1), and the smoothing length is limited to a minimum $h_{\text{min}}/\epsilon_{\text{gas}} =$

Table 1. The target masses of gas and DM particles at each of the resolution levels we adopt, with corresponding Plummer-equivalent gravitational softening (ϵ_{com}) length and maximum proper value (ϵ_{max}).

Resolution	m_{gas} ($M_{\odot}$)	m_{DM} ($M_{\odot}$)	$\epsilon_{\text{com, gas}}$ (ckpc)	$\epsilon_{\text{max, gas}}$ (pkpc)	$\epsilon_{\text{com, DM}}$ (ckpc)	$\epsilon_{\text{max, DM}}$ (pkpc)
m6	1.84×10^6	2.42×10^6	1.790	0.700	2.091	0.819
m5	2.30×10^5	3.03×10^5	0.895	0.350	1.046	0.410

10^{-5} to avoid artificial runaway collapse in high-density gas (S. Ploekinger et al. 2024).

Stellar and BH particles that form during the simulations do not experience hydrodynamical forces but use an SPH-like kernel to find their neighbours during the operation of subgrid models. BHs use the same $\eta = 1.2348$ kernel as gas particles, whilst stellar particles use a slightly shorter smoothing length ($\eta = 1.1642$). The mass of gas particles can increase due to the donation of mass by neighbouring stellar populations or decrease due to the accretion of mass by neighbouring BHs. Gas particles that exceed their initial mass by a factor of 4 are split into two particles of equal mass, whilst any reaching half of their initial mass become ineligible for mass accretion onto BHs.

2.4 Subgrid models

The simulations appeal to subgrid models to approximate the macroscopic effects of unresolved physical processes. Although we use the same methods as the EAGLE simulations to treat stellar evolution and mass loss, originally developed for the OWLS simulations (J. Schaye et al. 2010), the suite of subgrid models used here otherwise represents a substantial advance with respect to the EAGLE model, with several methods in common with the COLIBRE simulations (E. Chaikin et al. 2026; J. Schaye et al. 2026). In particular, the methods used for star formation and pre-SN feedback, and the seeding and growth of supermassive BHs, are those developed for COLIBRE, and radiative cooling and heating follow a scheme that allows gas to cool to $T = 10$ K. The injection of kinetic and thermal energy from SNe also proceeds similarly to COLIBRE, but differs insofar that a constant fraction of the energy liberated by SNe is coupled to the ISM and a slightly different range of permitted heating temperatures is adopted. Thermal energy from AGN feedback is injected in the same fashion as in COLIBRE, except for the use here of a fixed heating temperature. In contrast to COLIBRE, this model does not include a live dust model, turbulent diffusion, the updated nucleosynthetic yields and SNIa rate prescriptions, nor the jet-mode AGN feedback.

In this section we describe the models used to treat radiative cooling and heating (Section 2.4.1), star formation and the ISM (Section 2.4.2), the evolution of stellar populations and Type Ia SNe (Section 2.4.3), feedback from young stars (Section 2.4.4) and core collapse SNe (Section 2.4.5), the seeding and growth of BHs (Section 2.4.6) and feedback from AGN (Section 2.4.7). As many of the adopted methods are common to the EAGLE and COLIBRE projects and are described in detail elsewhere, we present here only concise descriptions, with an emphasis on differences with respect to the COLIBRE model.

2.4.1 Radiative cooling and heating

Radiative cooling and heating rates for hydrogen, helium, free electrons from these species and H_2 molecules are computed

on-the-fly with the non-equilibrium thermochemistry solver CHIMES (A. J. Richings, J. Schaye & B. D. Oppenheimer 2014a, b). The radiative cooling and heating rates of nine metal species (C, N, O, Ne, Mg, Si, S, Ca, and Fe) and dust are pre-computed and tabulated using CLOUDY version 17.03 (G. J. Ferland et al. 2017), assuming chemical equilibrium. We largely adopt the treatment of the radiation fields (redshift-dependent UV/X-ray background and interstellar radiation field), cosmic ray rates, and shielding column densities of the fiducial model of S. Ploekinger & J. Schaye (2020), but adopt three changes. First, motivated by the observations of B. C. Lacki, T. A. Thompson & E. Quataert (2010), we use a shallower power-law scaling of the cosmic ray ionization rate as a function of the reference column density assumed by the simulations, $\zeta_{\text{CR}} \propto N_{\text{ref}}^1$, rather than $\zeta_{\text{CR}} \propto N_{\text{ref}}^{1.4}$. Second, we add a turbulent pressure component (with a one-dimensional turbulent velocity dispersion of 6 km s^{-1}) to the thermal pressure when computing the Jeans column density. Third, we boost the rate of H_2 formation on dust grains by a factor that increases from 1 to 100 over the density range $n_{\text{H}} = 1 - 100 \text{ cm}^{-3}$. This is motivated by the simulation’s inability to resolve very high density gas clumps in which H_2 would form very efficiently. This boost factor for the H_2 formation rate is not part of the COLIBRE model, which instead tracks the dust abundance using the live dust model of J. W. Trayford et al. (2026) and enhances H_2 via dust growth by accretion. The impact of these changes on the thermal equilibrium temperatures is discussed in detail in S. Ploekinger et al. (2025). We do not impose an artificial pressure floor, and gas is allowed to cool to a minimum temperature of 10 K.

2.4.2 Star formation

We adopt the COLIBRE star formation model detailed by F. S. J. Nobels et al. (2024), with the full set of parameters and implementation as detailed in Section 3.3 by J. Schaye et al. (2026). Gas acquires a non-zero star formation rate (SFR) when it becomes locally unstable against gravitational collapse, specifically when the Jeans mass is smaller than the mass within the SPH kernel. The model has been shown to reproduce the spatially resolved observed Kennicutt–Schmidt relations for neutral, atomic, and molecular gas, and exhibits good convergence over a wide dynamic range in adopted gas mass resolution. Gas is considered unstable when it satisfies the gravitational instability criterion given by equation (6) in J. Schaye et al. (2026) which reduces to $\alpha < \alpha_{\text{crit}} = 1$, where α accounts for thermal and turbulent dispersions. Unstable gas particles are assigned an SFR that follows from the M. Schmidt (1959) law implemented via equation (9) in J. Schaye et al. (2026), with an efficiency per free-fall time of $\epsilon = 0.01$ following observations of star formation within giant molecular clouds (GMCs; e.g. M. R. Krumholz & J. C. Tan 2007). The SFR is then used as a probability for the stochastic conversion of the gas particle into a stellar particle during each time-step of duration Δt .

2.4.3 Stellar evolution and mass loss

We adopt the same R. P. C. Wiersma et al. (2009) stellar evolution and mass loss model used by EAGLE and originally developed for the OverWhelmingly Large Simulations (OWLS), in which stellar particles are treated as simple stellar populations with a G. Chabrier (2003) stellar initial mass function (IMF) defined between the limits 0.1 and $100 M_{\odot}$. Stellar particles inherit the element abundances of their parent gas particle. We track the time-dependent release of 11 species (H, He, C, N, O, Ne, Mg, Si, S, Ca, and Fe) from core collapse and thermonuclear (Type Ia) SNe, and the asymptotic giant branch (AGB) phase of evolved stars. The assumed rate of Type Ia SNe (SNIa) is also very similar to the EAGLE implementation, differing only in the inclusion of an initial delay, t_{delay} , between the formation of a stellar particle and the detonation of the first SNIa. The rate of SNIa per unit initial stellar mass formed is therefore:

$$\dot{N}_{\text{SNIa}}(t) = \mathcal{H}(t - t_{\text{delay}}) \frac{\nu \exp(-2t_{\text{delay}}/\tau)}{\tau} \exp\left(-\frac{t - t_{\text{delay}}}{\tau}\right), \quad (1)$$

where $\mathcal{H}$ is the Heaviside step function, ν is the total number of SNIa per unit initial stellar mass, and $\exp(-t/\tau)/\tau$ is a normalized, empirical delay time distribution function. We use $t_{\text{delay}} = 40$ Myr, $\tau = 2$ Gyr, and $\nu = 1.54 \times 10^{-3} M_{\odot} \text{ yr}^{-1}$. J. Schaye et al. (2015) show that these parameters yield a good match to the observed evolution of the cosmic SNIa rate.

2.4.4 Early stellar feedback

The simulations include three pre-SN feedback processes from massive stars: stellar winds, direct radiation pressure, and H II regions, using the numerical implementations for COLIBRE which are fully described by A. Benítez-Llambay et al. (2026, see also section 3.6 of J. Schaye et al. 2026). These processes are implemented using the Binary Population and Spectral Synthesis (BPASS) stellar evolution and spectral synthesis models (J. J. Eldridge et al. 2017; E. R. Stanway & J. J. Eldridge 2018), adopting a G. Chabrier (2003) IMF defined between 0.1 and $100 M_{\odot}$. Stellar winds are represented by stochastic momentum transfers to nearby gas, radiation pressure by exerting photon momentum on to the gas based on the local optical depth, and H II regions by stochastic ionization and heating of surrounding gas which also temporarily prevents star formation in affected particles.

2.4.5 Core collapse supernovae

The core collapse supernovae (CC SNe) feedback model is thermal-kinetic and designed for use in cosmological simulations that allow gas to cool to $T = 10$ K, partially resolving the cold ISM and hence not needing to impose an effective equation of state. The model injects feedback energy both thermally and kinetically: this combination of modes generates strong galactic winds and the hot ISM through periodic powerful injections of thermal energy, and turbulence in the ISM through frequent lower energy kinetic kicks (E. Chaikin et al. 2023). We broadly follow the approach of COLIBRE (see section 3.7 of J. Schaye et al. 2026) with the exception that we use a fixed feedback energy coupling fraction, $f_E = 1$, rather than a fraction that is a function of the thermal gas pressure. We elect against tuning the value of the feedback coupling efficiency due to the limitations of the current observed dwarf galaxy samples. The energy released by a stellar particle in a time-step is described by equation (15) of

J. Schaye et al. (2026), assuming a G. Chabrier (2003) IMF. A fraction, $f_{\text{kin}} = 0.1$, of this energy is delivered as stochastic kicks and the remainder as thermal energy. The heating temperature for the thermal feedback is a function of gas density as described by equation (18) of J. Schaye et al. (2026) for which we adopt a different normalization density of $n_{\text{H,pivot}} = 0.25 \text{ cm}^{-3}$ and allow the temperature to vary between $10^{6.5} < \Delta T_{\text{SN}} < 10^8 \text{ K}$.

2.4.6 Black holes

Our treatment of the seeding and growth of supermassive BHs broadly follows the methods for COLIBRE detailed in section 3.8 of J. Schaye et al. (2026), with the primary difference being that we retain a fixed BH seed mass across resolutions. The implementation builds on methods introduced by V. Springel, T. Di Matteo & L. Hernquist (2005a) and T. Di Matteo et al. (2008). BHs are treated as collisionless particles seeded within haloes that reach a specified mass threshold, and which grow via gas accretion and mergers with other BH particles. We follow the COLIBRE seeding procedure whereby BH particles are seeded in haloes that do not already have a BH particle once their friends-of-friends (FoF) mass exceeds a mass threshold, for Columba this is equal to $M_{\text{FoF,seed}} = 10^{10} M_{\odot}$. The BH particle is assigned a subgrid mass, m_{BH} , used for calculations of BH properties such as its accretion rate, whilst gravitational interactions are computed using the particle's dynamical mass, $m_{\text{BH}}^{\text{dyn}}$. In Columba, the subgrid mass is initially set to $m_{\text{BH,seed}} = 10^3 M_{\odot}$, irrespective of the simulation resolution. The BH mass accretion rate is estimated using the modified form of the Bondi–Hoyle–Lyttleton formula proposed by M. R. Krumholz, C. F. McKee & R. I. Klein (2006) for use in turbulent media with vorticity, see section 3.8.2 of J. Schaye et al. (2026).

Per Y. M. Bahé et al. (2022, see also section 3.8.3 of J. Schaye et al. 2026), BH particle pairs are merged instantly if their separation becomes less than three gravitational softening lengths, $\Delta r_{\text{BH}} < 3\epsilon_{\text{gas}}$, and their relative velocity satisfies $\Delta v_{\text{BH}} < 2G(M + m)/\Delta r$, where M and m are the dynamical masses of the two BHs. The less massive of the BHs transfers its properties to the more massive BH particle and is then removed from the simulation.

Real BHs are subject to dynamical friction that drains them of orbital kinetic energy and causes inspiral towards the centre of their host galaxy. Our simulations lack the resolution required to accurately model BH dynamical friction, so we model the net effect of the BH moving towards the local gravitational potential minimum following Y. M. Bahé et al. (2022, see also section 3.8.4 of J. Schaye et al. 2026). At each time-step the BH particle is re-positioned to the coordinates of the neighbouring gas particle with the lowest potential. The BH particle is excluded from the calculation of the gravitational potential to avoid becoming trapped by its own potential.

2.4.7 Active galactic nucleus feedback

We use a single injection mechanism for the AGN feedback energy stemming from gas accretion on BHs that is implemented by thermal heating (C. M. Booth & J. Schaye 2009). This follows the methodology detailed in section 3.9 of J. Schaye et al. (2026), with the exception that we adopt a fixed heating temperature rather than scaling it with the BH mass as in COLIBRE. Black holes accumulate feedback energy at a rate proportional to its gas accretion rate (see equation 34 of J. Schaye et al. 2026), with a

coupling efficiency fraction of $\epsilon_f = 0.05$ in Columba. This energy is stored until it is sufficient to heat at least one neighbouring gas particle by a fixed temperature increment of $\Delta T_{\text{AGN}} = 10^9 \text{ K}$, at which point the energy is released and the BH reservoir is reduced accordingly. This thermal heating scheme prevents excessive numerical radiative losses by ensuring that individual feedback events are sufficiently energetic.

2.5 Halo and galaxy identification and sampling

Halos are identified in the simulation outputs using the VELOCIRAPTOR structure finder (P. J. Elahi et al. 2019), which first applies a three-dimensional FoF algorithm to all particles to identify haloes, and then applies a six-dimensional (position and velocity) FoF algorithm to identify substructures. This approach is advantageous as it enables identification of substructures close to halo centres where the density contrast is small. We use the term ‘central galaxies’ (or haloes) to refer to the main substructure of the three-dimensional FoF halo, as identified by VELOCIRAPTOR, while other substructures are considered satellites. Unless otherwise stated our results only consider central galaxies. The centres of haloes and substructures are defined by the coordinates of their most-bound particle.

The properties of haloes and the galaxies they host are computed using the Spherical Overdensity and Aperture Processor (SOAP) (R. McGibbon et al. 2025), by aggregating the properties of the particles of which they are comprised, within various spherical apertures about the halo/substructure centres. Some analyses require that haloes are matched with their counterpart from a collisionless (often termed ‘DM-only’) version of the corresponding simulation, which eliminates the influence of baryons on halo growth and structure. This matching is achieved using the bijective particle matching algorithm which has been previously described by M. Velliscig et al. (2014).

Analyses in subsequent sections require derived quantities, such as the maximum of the circular velocity profile, V_{max} , where the circular velocity profile is $V_c(r) = \sqrt{GM(<r)}/r$, or the circular velocity at the virial radius, $V_{200} \equiv V_c(R_{200})$. We also consider halo concentration, c , which we define as the dimensionless shape parameter of the Navarro–Frenk–White (NFW) density profile (J. F. Navarro, C. S. Frenk & S. D. M. White 1996), where $c_{\text{NFW}} = R_{200}/R_s$ and R_s is the NFW scale radius. Rather than performing parametric density profile fits to recover c_{NFW} , we estimate the (intrinsic) halo concentration using the ratio of V_{max}/V_{200} from the matched counterpart haloes in the collisionless simulations, via

$$\frac{V_{\text{max}}^2}{V_{200}^2} \approx 0.216 \frac{c_{\text{NFW}}}{A(c_{\text{NFW}})}, \quad (2)$$

where,

$$A(c_{\text{NFW}}) \equiv \ln(1 + c_{\text{NFW}}) - \frac{c_{\text{NFW}}}{1 + c_{\text{NFW}}}. \quad (3)$$

J. S. Bullock et al. (2001) advocate the use of this relationship as it enables recovery of a concentration estimate for low-mass haloes that may lack the particle sampling required for a robust density profile fit. We use the collisionless simulations to eliminate the influence of baryon physics on halo concentrations.

Our analyses consider only those haloes/galaxies that reside within the central $r = 5 \text{ cMpc}$ of each region at $z = 0$. We have explicitly checked that these regions are uncontaminated by boundary particles, therefore it is guaranteed that the structures we

examine are themselves uncontaminated. Aggregating the galaxy populations of all 25 simulations at the present day yields a population of $N(M_* > 10^7 M_\odot) = 411$ at m6 resolution and $N(M_* > 10^7 M_\odot) = 448$ at m5 resolution. The m5 resolution simulations yield $N(M_* > 10^6 M_\odot) = 1321$. In this introductory study we do not construct galaxy samples at earlier epochs, at which time the Lagrangian region of the central $z = 0$ sphere does not have a simple morphology, complicating the selection procedure.

3 MODEL VALIDATION AND CONVERGENCE

The macroscopic efficiencies of feedback mechanisms are governed by microphysics acting on spatial scales several orders of magnitude below the resolution scale of cosmological simulations (e.g. S. Orlando et al. 2005). These efficiencies must therefore be implemented on resolved scales by the adopted subgrid models. The appropriate parametrization of the subgrid models is difficult to predict a priori, or to infer from observations, and is resolution dependent. It has therefore become common practice to calibrate the parameters of subgrid feedback models, with the aim of having simulations reproduce judiciously chosen properties of the galaxy population (e.g. M. Hirschmann et al. 2014; M. Vogelsberger et al. 2014; J. Schaye et al. 2015; Y. Feng et al. 2016; M. Tremmel et al. 2017; N. A. Henden et al. 2018; A. Pillepich et al. 2018a; R. Davé et al. 2019; R. Kugel et al. 2023). Such diagnostics commonly include, for example, the galaxy stellar mass function or the inferred stellar mass to halo mass relation, and the central black hole mass–stellar mass relation.

In the absence of authoritative observational constraints on the galaxy stellar mass function in the low-mass regime (though see T. M. Sedgwick et al. 2019), and similarly for the black hole mass–stellar mass relation, we have not tuned the feedback efficiencies or BH seed mass specifically for these simulations. Instead, we adopt fixed values for the fraction of energy from CC SNe ($f_E = 1$) and AGN that couples to the ISM ($\epsilon_f = 0.05$), and the BH seed mass ($m_{\text{BH,seed}} = 10^3 M_\odot$), as noted in Section 2.⁸ Although we did not undertake additional tuning of the parameter choices for these simulations, the underlying subgrid prescriptions inherit parameter choices that were iteratively tested and validated across previous EAGLE (R. A. Crain et al. 2015) and COLIBRE (E. Chaikin et al. 2026) simulations.

Nevertheless, it is imperative to validate the outcome of the simulations via comparison with observational measurements of fundamental galaxy properties, and to assess how robust these fixed quantities are to changes of numerical resolution. We therefore defer examination of large-scale environmental influences on the dwarf galaxy population to Section 4, and first combine the 25 simulations at each resolution to yield a composite galaxy population (albeit one from a collection of regions that are lower density than a random sample, per Section 2.1), in order to examine the GSMF (Section 3.1), stellar masses as a function of halo mass (Section 3.2), and galaxy sizes (Section 3.3).

Our convergence testing is achieved by comparison of the properties of the galaxy population at m5 and m6 resolution using a galaxy formation model with fixed parameters. The comparison thus constitutes a ‘strong scaling test’ in the terminology of J. Schaye et al. (2015, see their section 2.2). Many of the results we

⁸Although we do not tune the energy fractions here, had the simulations yielded unrealistic galaxies, we would clearly have adopted a more refined approach.

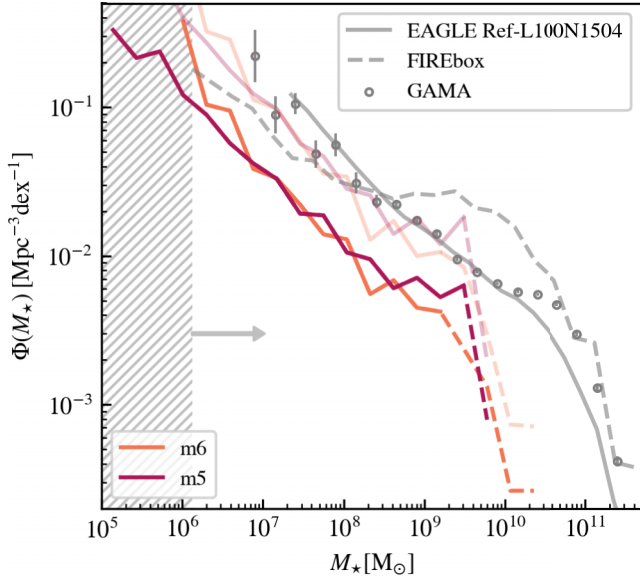


Figure 4. The present-day GSMF of the composite galaxy population from all 25 regions, at m5 (maroon curve) and m6 (orange curve) resolutions, representing a ‘strong convergence’ test. The grey hatched region indicates the stellar mass regime corresponding to 10 or fewer stellar particles (assuming their initial mass) at m5 resolution, the grey arrow extends to the corresponding scale at m6 resolution. The curves are drawn with a dashed line style for bins sampled by fewer than 10 galaxies. Low opacity curves show the GSMFs that result from weighting the number density of galaxies by the reciprocal of their normalized density, δ_5 , enabling comparison with GSMFs from representative volumes. Grey symbols show the GSMF and associated error inferred from the GAMA survey (S. P. Driver et al. 2022). Solid and dashed grey curves show the GSMF of the EAGLE (J. Schaye et al. 2015) and FIREbox (R. Feldmann et al. 2023) simulations, respectively.

examine in Section 4 follow from the intrinsic properties of DM haloes, which do not depend on the implementation of the galaxy formation model, therefore we examine the numerical convergence of the DM halo mass function and the inner structure of DM haloes separately, in Appendix B. Finally, we note that the inclusion of stochastic processes in the adopted subgrid models, such as those governing star formation and feedback, can lead to significant uncertainty in the properties of individual simulated galaxies (e.g. S. Genel et al. 2019; B. W. Keller et al. 2019; J. J. Davies, R. A. Crain & A. Pontzen 2021; J. J. Davies, A. Pontzen & R. A. Crain 2022; J. Borrow et al. 2023). When studying populations of galaxies, as we do here, this uncertainty is manifested primarily as a source of scatter in scaling relations.

3.1 The galaxy stellar mass function

Fig. 4 shows the present-day GSMF of the composite galaxy population formed from all 25 spherical regions of $r = 5$ cMpc. Since we simulate regions that are preferentially low density, one cannot make a direct comparison of the resulting GSMF with observational measurements, nor with GSMFs recovered from simulations of mean-density volumes. We therefore apply a simple correction to the simulated GSMF, rescaling the number density at fixed mass by the reciprocal of the (normalized) mean density of the 25 regions. These rescaled GSMFs are drawn with low-opacity curves, enabling a more meaningful comparison with the GSMF inferred by S. P. Driver et al. (2022) from the fourth data

release of the Galaxy And Mass Assembly (GAMA) survey, shown as white circles. For comparison with the GSMF recovered from a simulation of a large periodic volume, we show the present-day GSMF of the EAGLE Ref-L100N1504 simulation (J. Schaye et al. 2015) as a solid grey curve, and for comparison with a cosmological simulation including an explicit cold ISM phase, we show the FIREbox $z = 0$ GSMF (R. Feldmann et al. 2023) as a dashed grey curve. In both cases, the mass function are shown down to a minimum stellar mass corresponding to 10 stellar particles.

In the regime where galaxies comprise at least 10 or more particles at m6 resolution and the stellar mass bins are sampled by at least 10 galaxies (broadly, $2 \times 10^7 \text{ M}_\odot \lesssim M_* \lesssim 2 \times 10^9 \text{ M}_\odot$), the simulated GSMF exhibits very good convergence between m5 and m6 resolutions, differing by at most 0.22 dex in stellar mass at fixed number density. This convergence behaviour is superior to that exhibited by the EAGLE model and is likely driven by the inclusion of a cold ISM phase, which stabilizes star formation and feedback efficiency across resolutions (see discussion in J. Schaye et al. 2026). The adoption of the thermal-kinetic CC SNe feedback implementation coupled to the isotropic energy distribution method (E. Chaikin et al. 2022, 2023, 2026; J. Schaye et al. 2026) may further contribute to the improved numerical convergence. At the high-mass end of the well-sampled regime of the GSMF (corresponding to an approximate number density of $5 \times 10^{-3} \text{ cMpc}^{-3}$) galaxies reach a slightly greater stellar mass in the m5 simulations, likely due to star formation in the progenitors of these galaxies being able to begin at an earlier epoch, and reflecting the fixed AGN model parameters across resolution, which can influence the efficiency of BH growth and feedback regulation. In the poorly sampled more massive bins, the handful of galaxies tend to reach greater masses in the lower-resolution m6 simulations. While this may primarily reflect stochastic fluctuations, better convergence in this regime could be achieved by calibrating the BH seed mass as a function of resolution for these simulations. Given the small number of massive galaxies sampled in our simulations, we opted against this additional complexity.

The preferential selection of low density regions leads to the unscaled GSMFs residing at a lower number density at fixed stellar mass, relative to the GAMA GSMF. In the regime where simulated galaxies are well resolved and the stellar mass bins well sampled, the rescaled GSMFs of the simulations reproduce the GAMA GSMF well, with the stellar mass at fixed number density differing by $\lesssim 0.22$ dex. This is comparable to the agreement achieved by the EAGLE flagship Ref-L100N1504 simulation, but here this level of agreement extends to stellar masses a factor of 8 lower for the m5 simulations. Even after rescaling, the number density of galaxies of mass $M_* \gtrsim 3 \times 10^9 \text{ M}_\odot$ is significantly lower than inferred by GAMA: this stems from a paucity of massive DM haloes in the selected regions by construction, as can be seen by comparison of the halo mass functions of the regions with that of mean density volumes (see Fig. B1).

3.2 Stellar mass as a function of halo mass

To first order, the abundance and growth of dark matter haloes are simply a function of the assumed cosmological parameters. The stellar mass of the galaxies that forms within them is, however, very sensitive to the baryonic physics included in the simulation. The SMHM scaling relation has emerged as a popular diagnostic, in part because it can be considered as a halo mass-dependent galaxy formation ‘efficiency’ on the scale of individual haloes. We include the SMHM here as a validation diagnostic

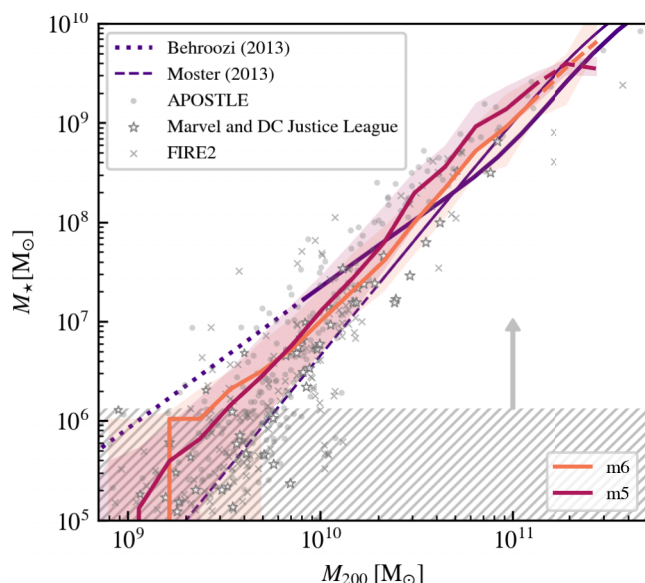


Figure 5. The present-day stellar mass–halo mass scaling relation of central galaxies for the composite population of all 25 regions, at m5 (maroon curve) and m6 (orange curve) resolutions, representing a ‘strong convergence’ test. Line styles and hatching follow the conventions introduced in Fig. 4. Shaded regions denote the 10th–90th percentile scatter in M_* at fixed M_{200} . Indigo curves show the SMHM relations of the semi-empirical models of P. S. Behroozi, R. H. Wechsler & C. Conroy (2013) and B. P. Moster, T. Naab & S. D. M. White (2013), which transition to solid lines in the mass range to which they are fit. Individual central galaxies from various models are shown by the grey points, including APOSTLE (filled circles, T. Sawala et al. 2016), Marvel and DC Justice League (stars, E. Applebaum et al. 2021), and FIRE2 zoom-in simulations (crosses, C. Wheeler et al. 2015; A. Fitts et al. 2017; P. F. Hopkins et al. 2018; C. Wheeler et al. 2019).

as the GSMF can be thought of as a convolution of the DM HMF with the SMHM relation. Moreover, most numerical simulations in the low-mass dwarf galaxy regime maximise spatial resolution by adopting zoomed initial conditions that focus on the environments of individual galaxies, so the SMHM relation is a more useful diagnostic than the GSMF when comparing our simulations to others.

Fig. 5 shows the present-day SMHM of the composite population of field galaxies formed from all 25 simulated regions. Satellite galaxies, whose evolution is potentially affected by ram pressure stripping of gas and mass loss due to tidal stripping, are excluded⁹ in order to illustrate the scatter about the relation due only to processes ‘internal’ to the galaxy’s parent DM halo, however we note that we have not attempted to identify and exclude ‘backsplash galaxies’, i.e. present-day central galaxies that were satellite galaxies at an earlier time. We do not expect a significant population of such galaxies, because there are no massive haloes in our simulated regions. We show the median stellar mass as a function of halo mass in bins of width 0.2 dex, and the corresponding 10th–90th percentile scatter.

The indigo curves denote the present-day SMHM relations from the abundance matching models of P. S. Behroozi et al.

(2013)¹⁰ and B. P. Moster et al. (2013). These models are essentially unconstrained for $M_* \lesssim 10^7 M_\odot$, and in this regime we simply extrapolate the SMHM relation from the resolved regime, drawing the corresponding curves with dotted and dashed line styles, respectively. The uncertainty on the SMHM relation in this regime is a hindering factor in testing predictions for the properties of dwarf galaxies from numerical simulations. L. V. Sales et al. (2022) collate predictions for the SMHM relation from various cosmological hydrodynamical simulations of dwarf galaxies, revealing similar stellar masses at fixed halo mass (for $M_{200} \gtrsim 10^9 M_\odot$, albeit with large scatter) and broad compatibility with extrapolations of the predictions of semi-empirical models. However, the slope and scatter of the SMHM relation recovered from the simulations considered by L. V. Sales et al. (2022) exhibit significant variance. We therefore plot with grey symbols individual central galaxies from high-resolution simulations: filled circles are from the APOSTLE suite (T. Sawala et al. 2016) of zoomed simulations of Local Group (LG)-like environments; open stars are from the Marvel and DC Justice League suites (E. Applebaum et al. 2021) that simulate dwarfs in cosmic sheets and MW-like environments, respectively; crosses denote the Latte (A. R. Wetzel et al. 2016) and ELVIS (S. Garrison-Kimmel et al. 2014) zoom simulations of MW and LG-like environments, in addition to zoom simulations of isolated dwarf galaxies, all evolved with the FIRE-2 model (C. Wheeler et al. 2015, 2019; A. Fitts et al. 2017; P. F. Hopkins et al. 2018).

The good convergence of the GSMF between the m5 and m6 resolutions evident in Fig. 4 translates into similarly good convergence of the SMHM relation. In the regime for which galaxies are resolved by at least 10 stellar particles at m6 resolution ($M_* \gtrsim 2 \times 10^7 M_\odot$) and halo mass bins are sampled by at least 10 galaxies ($M_{200} \lesssim 10^{11} M_\odot$), the scaling relation differs in stellar mass by at most 0.29 dex at fixed halo mass. As one might infer from the analysis of the GSMF, the poorest correspondence is for the relatively massive galaxies, with the m5 simulations exhibiting greater stellar masses at fixed halo mass, until one reaches a mass scale of $M_{200} \sim 10^{11} M_\odot$. Above this mass, the central BHs in these galaxies tend to have greater characteristic masses at m5 resolution, which may contribute to stronger regulation of stellar mass relative to the m6 simulations. However, given the small number of objects in this regime, this should be interpreted with caution. In the low mass regime ($M_{200} \lesssim 10^{10} M_\odot$), simulations of low-mass dwarf galaxies exhibit significant scatter in stellar mass at fixed halo mass, arguably in part owing to the absence of authoritative observational constraints to which the models can be calibrated. The SMHM relations produced by our simulations roughly bisect the extrapolated semi-empirical relations of B. P. Moster et al. (2013) and P. S. Behroozi et al. (2013), and lie within the scatter of the relations traced by the APOSTLE, Marvel, DC Justice League, and FIRE-2 simulations.

3.3 Galaxy sizes

R. A. Crain et al. (2015) showed that it was possible to reproduce accurately the present-day GSMF in a version of the EAGLE simulations adopting an uncalibrated feedback coupling efficiency (i.e. $f_E = 1$), as we adopt here, but at the cost of a poor reproduction of the scaling relation connecting the size and stellar mass

⁹Due to the relatively low mass of our field sample and the resolution limit of the simulations, the fraction of satellites is small; only 11 per cent of all galaxies with $M_* > 10^5 M_\odot$ are satellites.

¹⁰We convert halo masses in this work, which were computed based on the virial overdensity criterion of G. L. Bryan & M. L. Norman (1998), to M_{200} assuming $M_{200} = M_{\text{vir}}/1.2$.

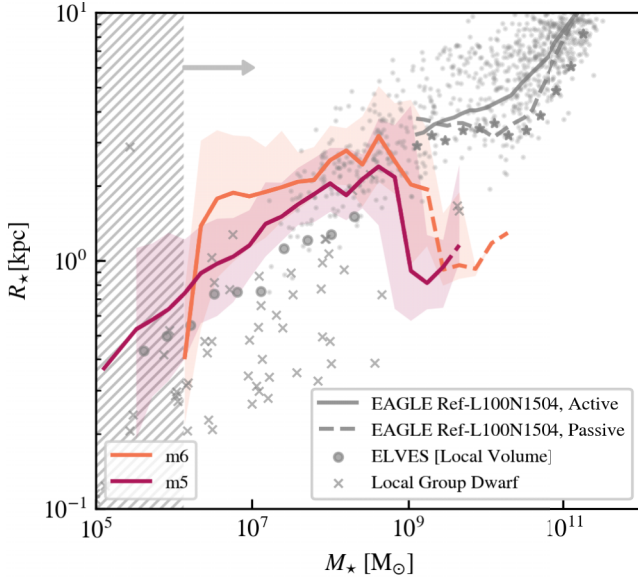


Figure 6. The present-day galaxy size (three-dimensional half-mass radii) – stellar mass scaling relation of the composite galaxy population from all 25 regions, at m5 (maroon curve) and m6 (orange curve) resolutions, providing a ‘strong scaling’ test. Shaded regions denote the 10th–90th percentile scatter in R_* at fixed M_* . Line styles and hatching follow the conventions introduced in Fig. 4. The median sizes of galaxies sampled by 100 or more particles in the EAGLE simulation are shown with grey curves (M. Furlong et al. 2017). Grey points show observational measurements of individual galaxies from I. Trujillo, N. Chamba & J. H. Knapen (2020). Grey star symbols show median measurements from J. A. Hardwick et al. (2022). Circles show medians of the ELVES Local Volume dwarf galaxies from S. G. Carlsten et al. (2021). Crosses show Local Group dwarf galaxies from A. B. Pace (2025). Projected observational sizes have been converted to three-dimensional sizes using $R_{3D} = \frac{4}{3} R_{2D}$.

of galaxies. This shortcoming stemmed primarily from spurious radiative losses in high-density gas, resulting in the formation of unrealistically compact galaxy nuclei. We therefore consider galaxy sizes a critical validation diagnostic, and define size as the three-dimensional stellar half-mass radius, R_* . Fig. 6 shows the median size of all present-day galaxies in our composite sample as a function of stellar mass in bins of width 0.2 dex, and the corresponding 10th–90th percentile scatter. The solid and dashed grey curves show the corresponding scaling relations for active and passive galaxies, respectively, with mass $M_* \gtrsim 10^9 M_\odot$ taken from the EAGLE Ref-L100N1504 simulation, as presented by M. Furlong et al. (2017). We also present observed galaxy sizes for reference, and use the relation $R_{3D} = \frac{4}{3} R_{2D}$ to convert between observed two-dimensional (projected) and three-dimensional sizes (J. Wolf et al. 2010). Faint grey dots denote measurements of projected galaxy sizes from the IAC Stripe 82 Legacy Project by I. Trujillo et al. (2020), grey stars correspond to size measurements of a subsample of xGASS galaxies with robust rotational velocity measurements (J. A. Hardwick et al. 2022), grey circles show local volume galaxies from the ELVES survey (S. G. Carlsten et al. 2021), and grey crosses Local Group field galaxies (A. B. Pace 2025).¹¹

In the mass regime for which galaxies are well resolved in the m5 and m6 realizations, the simulations yield a size–mass

relation that is broadly monotonic up to a mass scale of $M_* \simeq 5 \times 10^8 M_\odot$. In m5, the change in size across two orders of magnitude in mass is comparable to that seen in observational samples, although the simulated galaxies are systematically larger at fixed mass. The visual impression of a flatter relation in the simulations primarily reflects this vertical offset rather than a substantially different slope. Beyond this scale, at both resolutions galaxies quickly become compact, qualitatively similar to the unrealistically compact sizes reported by R. A. Crain et al. (2015) for variants of the EAGLE model that do not scale f_E with the density of natal gas, leading to numerically inefficient feedback due to artificial radiative losses. This likely signals that additional energy input ($f_E > 1$) is required in the progenitors of these more massive present-day galaxies to offset artificial losses. The sizes of the most massive galaxies ($M_* \gtrsim 10^9 M_\odot$ at m5 and $M_* \gtrsim 3 \times 10^9 M_\odot$ at m6) increase mildly as a function of stellar mass, possibly indicating that artificial radiative losses from SN feedback are (at least partially) offset by the onset of effective AGN feedback.

The m6 simulations yield larger galaxies at fixed mass, a result qualitatively consistent with the strong convergence test of galaxy sizes in EAGLE reported by M. Furlong et al. (2017, see their appendix B). At the mass scale corresponding to 10 particles at m6 resolution, $M_* \simeq 2 \times 10^7 M_\odot$, the median galaxy size at m6 resolution (1.92 kpc) is approximately 1.5 times that at m5 (1.39 kpc). The offset declines towards greater masses: at $M_* = 3 \times 10^8 M_\odot$ the corresponding sizes are 2.40 kpc at m6 and 2.22 kpc at m5. This highlights that galaxy sizes require significantly better particle sampling to achieve convergence than stellar masses (see also e.g. M. Furlong et al. 2017; A. Pillepich et al. 2018b; A. D. Ludlow et al. 2019; A. D. Ludlow et al. 2020), and that they are more sensitive to the details of the galaxy formation model than galaxy (stellar) masses.

The simulated galaxies are systematically larger than inferred by S. G. Carlsten et al. (2021) from observations of dwarf galaxies in the Local Group. It is not trivial to compare three-dimensional half-mass radii with isophotal sizes (see e.g. S. Genel et al. 2018; A. de Graaff et al. 2022), and although we convert the observed projected radii into three-dimensional, this conversion remains approximate. Even so, a naïve comparison indicates that in the mass range $2 \times 10^6 M_\odot < M_* < 5 \times 10^8 M_\odot$, galaxies in the m5 simulations are up to twice the size of observed counterparts from S. G. Carlsten et al. (2021) and up to a factor of 3.5 times larger than the binned median of the A. B. Pace (2025) sample. This indicates that the simulations do not reproduce the smaller observed galaxy sizes; even the 10th percentile simulated sizes lie well above the more compact Local Group galaxies in the observational sample.

Although the sizes of galaxies with $M_* \gtrsim 3 \times 10^8 M_\odot$ are clearly inconsistent with observational measurements, warranting care when interpreting related properties such as their SFHs, the simulations reproduce a size–mass relation with a similar slope to observations at lower masses. At fixed stellar mass, the offset in size between the m5 simulations and the S. G. Carlsten et al. (2021) measurements is comparable to, or smaller than, the difference between the I. Trujillo et al. (2020) and S. G. Carlsten et al. (2021) measurements (in the mass regime for which these samples overlap). The simulations therefore differ from the S. G. Carlsten et al. (2021) measurements by less than what might reasonably be inferred as the systematic uncertainty on the latter. We also caution that incompleteness of the data can play a role here as more extended, diffuse dwarf galaxies are harder to discover;

¹¹This study used the Local Volume Database, https://github.com/apace7/local_volume_database.

e.g. ultra diffuse Antlia II around the Milky Way was missing up until recent years and was discovered largely due to *Gaia* proper motion measurements (G. Torrealba et al. 2019). Moreover, the data in the lower mass range are from dwarf galaxies around higher mass hosts which is a different environment from our sample of simulated dwarf galaxies. Overall, we can conclude that while the simulations reproduce the qualitative trend of increasing size with stellar mass below $M_\star \lesssim 5 \times 10^8 M_\odot$, they likely remain systematically too large, particularly relative to Local Group field dwarfs.

4 THE INFLUENCE OF LARGE-SCALE ENVIRONMENT ON ISOLATED DWARF GALAXIES

In this section, we turn to one of the main questions motivating the simulation suite: what is the influence of cosmic environment on the properties of dwarf galaxies and their parent DM haloes? In the light of this question, we examine fundamental properties of our simulated dwarf galaxy population as a function of their environment defined by the mass density within $r = 5$ cMpc spheres, δ_5 . As illustrated by Fig. 3, the five bands of δ_5 (denoted by rows in the figure) broadly correspond to differing characteristic cosmic environments, from void-like regions to filamentary structures. Unless otherwise stated, we focus exclusively on the simulations at m5 resolution in this section.

4.1 The halo mass function

A fundamental prediction of CDM cosmogonies is that the rate at which a region's halo population evolves is dependent on its density (e.g. C. S. Frenk et al. 1988; R. K. Sheth & G. Tormen 2002). We therefore show in the upper panel of Fig. 7 the present-day halo mass functions (HMFs) of our hydrodynamical simulations as a function of δ_5 . We amalgamate the halo populations of the five regions comprising each band of δ_5 (i.e. $-2\sigma, \dots, +2\sigma$), normalizing by their combined volume. The grey curve is a universal mean density HMF fit for the adopted cosmogony of our simulations, created with the COLOSSUS package (B. Diemer 2018) based on the J. Tinker et al. (2008) HMFs assuming spherical overdensity (SO) haloes with mean internal density contrast $\Delta_{\text{crit}} = 200$. Note that their halo centre definition (density peak) differs slightly from the most bound particle centre used in our simulations. This HMF is based on fits to ensembles of collisionless simulations, and does not account for the influence of galaxy formation on the mass and internal structure of haloes; such effects are halo mass dependent, and are the subject of intense study (see e.g. R. Stanek, D. Rudd & A. E. Evrard 2009; A. R. Duffy et al. 2010; M. Velliscig et al. 2014; M. Schaller et al. 2015; S. Bocquet et al. 2016; M. R. Lovell et al. 2018).

As expected, given that the enclosed densities of all simulated regions are less than the cosmic mean density (i.e. $\delta_5 < 1$), the HMFs of all five density bands are shifted to lower number densities at fixed M_{200} than the universal HMFs, with the offset being greatest for the lowest-density band. In the low-mass regime, for which the HMF is well approximated by a power law, the offset in number density at fixed M_{200} is a monotonic function of enclosed density: at $M_{200} = 10^8 M_\odot$, the bands are offset by (1.15, 0.81, 0.61, 0.44, 0.31) dex, respectively, for $(-2, -1, 0, +1, +2)\sigma$ regions, from the J. Tinker et al. (2008) HMF. The influence of the density of the regions is more readily seen in the lower panel, in which the HMFs of the regions comprising each band have

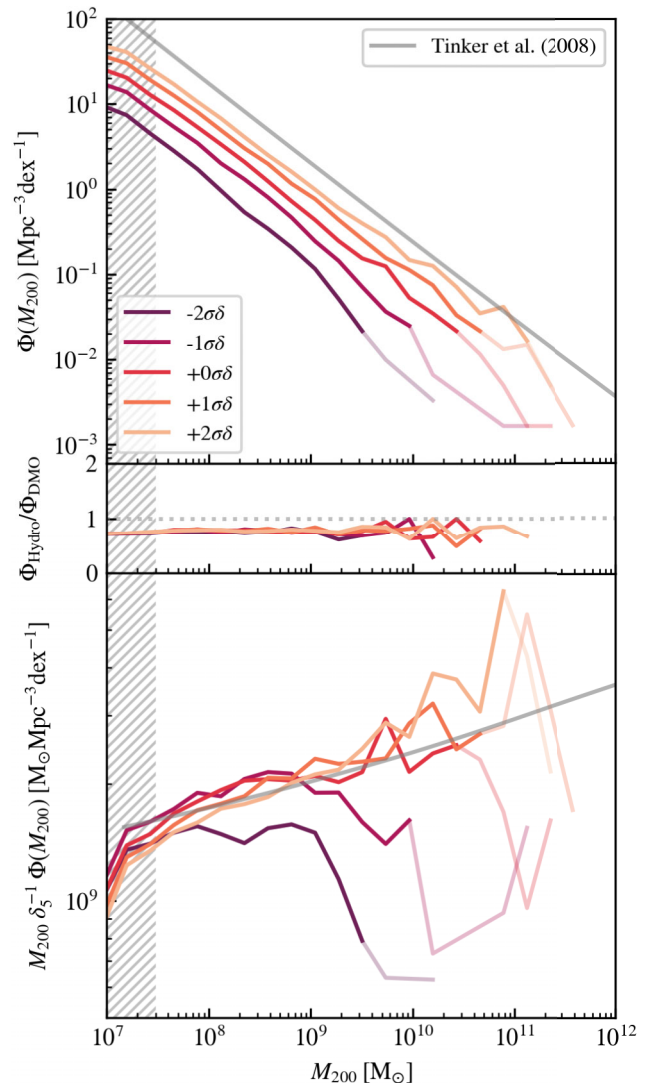


Figure 7. *Upper panel:* The aggregated present-day halo mass functions of the five regions comprising each of the five density bands ($-2\sigma, \dots, +2\sigma$) at m5 resolution, with each band denoted by a coloured curve. Only field (central) objects are considered. Grey hatching indicates the halo mass regime corresponding to fewer than 100 DM particles at m5 resolution. Curves are drawn with a reduced opacity for bins sampled by fewer than 10 haloes. The grey curve shows the halo mass function of J. Tinker et al. (2008) derived from purely collisionless simulations. Higher density regions exhibit a systematically greater number density of haloes at fixed halo mass. *Centre panel:* The ratio of the halo number density at fixed halo mass, $\Phi(M_{200})$, realized in hydrodynamical simulations, to that in their purely collisionless counterpart simulations (DMO). The curve is drawn for bins sampled by at least 10 haloes. The ratio illustrates the suppression of the HMF due to baryonic processes, which systematically reduce the mass of haloes. *Lower panel:* As per the upper panel, but the HMFs have been multiplied by M_{200} to reduce the dynamic range of the y-axis, and normalized by δ_5 to remove the first-order influence of enclosed density on halo number density, highlighting that the offset from the J. Tinker et al. (2008) HMF visible in the upper panel is primarily a consequence of each region's underdensity. The most massive halo formed in each region is a strong function of δ_5 .

been divided by the enclosed density, δ_5 , and multiplied by M_{200} to reduce the y-axis dynamic range, thus better revealing differences in the ‘specific’ halo number density of the density bands. With the influence of density ‘normalized out’, the low-mass power-law regime of the HMFs ($M_{200} < 10^9 M_\odot$) more closely resembles the universal HMF. There remains some scatter about the universal HMF, with a weak systematic trend apparent such that the higher-density regions tend to exhibit a slightly lower specific halo number density than lower-density regions (as also found in the GIMIC simulations; see R. A. Crain et al. 2009). However, the lowest-density region retains a low specific halo number density after normalization, and fully characterising whether there is a genuine systematic trend in specific halo number density as a function of overdensity would require examination of a larger sample of regions. The slope of the low-mass end of the HMF is also broadly insensitive to the density of the region. A. Fattahi, J. F. Navarro & C. S. Frenk (2020) report similar results, finding a correlation with little scatter between the number of low mass haloes and the total mass within a sphere of $r = 3$ cMpc centred on the Local Group-like environments. The lower panel of Fig. 7 also highlights that the characteristic mass above which the HMF ceases to be well described by a power law is broadly a function of δ_5 : more massive haloes in our simulations are only found in the higher density regions, consistent with prior analytical and numerical studies of halo growth (e.g. H. J. Mo & S. D. M. White 1996; R. K. Sheth & G. Tormen 2002).

The scale of the influence of baryonic processes on the HMF is illustrated by the centre panel, which shows the ratio of the halo number density (at fixed halo mass) of the hydrodynamical simulations to that derived from their counterpart simulations with purely collisionless dynamics. In the halo mass regime for which the HMF is well described by a power law, the ratio of the halo number density is approximately $\Phi_{\text{Hydro}}/\Phi_{\text{DMO}} \simeq 0.8$, where $\Phi(M_{200}) \equiv dn(M_{200})/d\log_{10}(M_{200})$. Physically, this effect should more appropriately be considered as a shift in halo mass at fixed number density, i.e. a horizontal rather than vertical shift of the curves. T. Sawala et al. (2013) showed that the reduction of the mass of haloes below $\sim 10^{12} M_\odot$ when baryon physics is included is primarily a consequence of the unbinding of circumgalactic gas due to heating by the metagalactic UV/X-ray background radiation that permeates the cosmos after the epoch of reionization, and the subsequent reduction in the halo mass accretion rates.

4.2 The galaxy stellar mass function

The upper panel of Fig. 8 shows the present-day GSMFs at $m5$ resolution. As in Fig. 7, each coloured curve denotes the GSMF of a different density band, δ_5 , generated by aggregating the galaxy populations of the band’s five constituent simulations. The solid and dashed grey curves show, respectively, the present-day GSMFs of the EAGLE Ref-L100N1504 (J. Schaye et al. 2015) and the FIREbox (R. Feldmann et al. 2023) simulations. Open grey symbols denote the GSMF recovered from the fourth data release of the GAMA survey (S. P. Driver et al. 2022).

The systematic offset of the HMFs as a function of δ_5 shown in Fig. 7 translates into a similar offset of the GSMFs of the regions comprising the five density bands, as the formation of fewer haloes in low-density environments results in the formation of fewer galaxies. Just as the HMFs of the simulations are offset from the mean density HMFs, their GSMFs are also in general offset to lower characteristic number densities than those of the simulations of mean density volumes, and that from GAMA.

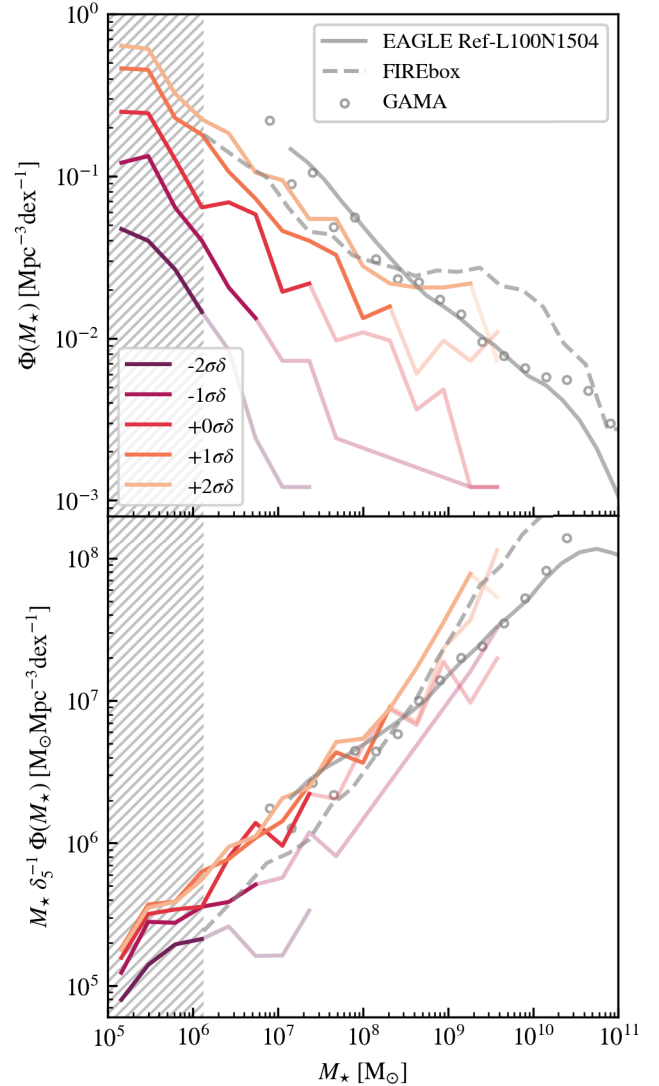


Figure 8. *Upper panel:* The aggregated present-day central galaxy stellar mass functions of the five regions comprising each of the five density bands, with each band denoted by a coloured curve. Grey hatching indicates the stellar mass regime corresponding to fewer than 10 stellar particles at $m5$ resolution. Curves are drawn with a reduced opacity for bins sampled by fewer than 10 galaxies. Grey symbols show the GSMF inferred from the GAMA survey (S. P. Driver et al. 2022). Dotted and dashed grey curves show the GSMF of the EAGLE (J. Schaye et al. 2015) and FIREbox (R. Feldmann et al. 2023) simulations, respectively. In analogy with the HMF, higher density regions exhibit a systematically greater number density of galaxies at fixed stellar mass. *Lower panel:* As per the upper panel, but as in Fig. 7, the GSMFs have been multiplied by M_* , and normalized by δ_5 to remove the effect of enclosed density on halo number density. As with the HMF (Fig. 7) the offset from the GAMA GSMF visible in the upper panel follows primarily from the deviation of each region’s mass density from the cosmic average.

We again examine the effect of environment on the GSMF more closely by showing in the lower panel of Fig. 8 the GSMFs multiplied by the bin mass (in this case, M_*) and normalized by δ_5 . This panel illustrates that the slope of the GSMF in the low-mass regime is broadly insensitive to the environment defined by δ_5 ; in common with the HMF, by ‘normalizing out’ the enclosed density, the GSMFs of each ensemble of simulations comprising

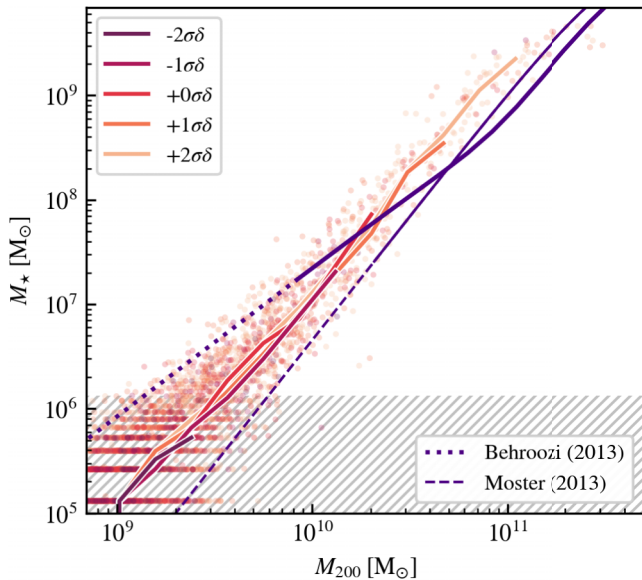


Figure 9. The present-day central stellar mass–halo mass scaling relation, at m5 resolution, of the aggregated central galaxy populations of the five regions comprising each of the five density bands. Coloured curves denote, for each density band, the median stellar mass at fixed halo mass, and include the contribution of haloes without a resolved galaxy. The SMHM relations of the semi-empirical models of P. S. Behroozi et al. (2013) and B. P. Moster et al. (2013) are drawn as dotted and dashed indigo curves, respectively, which transition to solid lines in the mass region to which they are fit. Colours, line styles, and hatching follow the conventions introduced in Fig. 8.

the five density bands more closely resembles the GAMA GSMF, indicating that, to first order, the low normalization of the low-mass regime of the GSMF in low-density regions is a consequence of that lower density. There is however a small residual offset, such that the density-normalized GSMFs of the lowest-density bands (-2σ and -1σ) remain offset from those of the 0σ , $+1\sigma$, and $+2\sigma$ bands. The stellar mass formed per unit total mass is therefore slightly but significantly lower in the lowest density regions. This follows primarily from the analogous differences in halo abundances at the corresponding halo mass scale, seen in the lower panel of Fig. 7: galaxies of $M_* \simeq 10^6 M_\odot$ are hosted by haloes of mass $M_{200} \gtrsim 10^9 M_\odot$, the scale at which clear systematic differences between the density bands become apparent in the density-normalized HMF.

4.3 Stellar mass as a function of halo mass

Fig. 9 shows the present-day SMHM relation of central galaxies in the m5 simulations, with each coloured curve denoting the median central stellar mass at fixed halo mass for the aggregated galaxy population of the five regions comprising each density band, δ_5 . The median is computed from all haloes, including those that do not host a galaxy (i.e. $M_* = 0$), and is drawn for all halo bins sampled by at least 10 galaxies. Individual galaxies are also shown as filled symbols, colour-coded according to the δ_5 of their parent region as per the median relation. The horizontal banding of individual galaxies in the low-mass regime is due to discrete numbers of stellar particles becoming visible on the logarithmic vertical scale.

Aside from the truncation of the curves due to the absence of massive haloes in lower density regions, the median SMHM

relations of the regions comprising each of the δ_5 bands are remarkably similar, in both amplitude and slope. The scatter about the median stellar mass at fixed halo mass is also similar. Cosmological environment on the $r = 5$ cMpc scale defined by the δ_5 bands therefore clearly plays little role in shaping the halo-specific formation ‘efficiency’ (if one defines this as the stellar mass of the central galaxy formed per unit halo mass) of isolated dwarf galaxies. The plot further underscores that the differing normalized stellar mass functions (shown in the bottom panel of Fig. 8) are not a consequence of a systematic influence on this efficiency of the environment defined on this scale.

At first glance, our findings appear to contradict those of I. Zehavi et al. (2018), who examined the influence of halo assembly bias on the galaxy population predicted by a semi-analytic model when applied to the Millennium Simulation (V. Springel et al. 2005b), and those of M. C. Artale et al. (2018, see also T. Meshveliani et al. 2024) who performed a similar study using the EAGLE Ref-L100N1504 simulation. In both cases, the authors report a significant correlation between the median stellar mass (at fixed halo mass) and the total mass density measured within a sphere of $r = 5 h^{-1}$ Mpc, for haloes of mass $M_{200} \gtrsim 10^{11} M_\odot$. This halo mass range is effectively unsampled by our simulations (see Fig. 7) so our comparison does not probe the same fixed halo mass scale examined, nevertheless, we highlight two significant factors that likely contribute to the apparent difference in conclusions.

First, both studies were able to examine a significantly larger dynamic range in density (approximately 3 dex in the Millennium Simulation and 2 dex in EAGLE) than is the case here (approximately 0.8 dex). The positive correlation they find between $M_*(M_{200})$ and the environmental density is particularly apparent as an elevated formation efficiency for galaxies located in high-density regions, which are absent from our sample by construction. Secondly, those studies measure the environmental density in a sphere centred on each galaxy, and thus perform a significantly different test to that presented in Fig. 9. This is particularly true for low-density regions, since by definition their constituent galaxies tend to be located close to the boundary of the region, maximizing the difference between the enclosed density of an arbitrarily centred sphere, and one centred on a galaxy. More direct comparison to the findings of I. Zehavi et al. (2018) and M. C. Artale et al. (2018) can be achieved with halo specific properties, to which we turn to in Section 5. Here, we simply note that our simulations indicate an absence of systematic difference in the typical stellar mass of central dwarf galaxies at fixed halo mass in regions of differing enclosed density on a scale of $r = 5$ Mpc. We cannot rule out an influence of environment, on larger or smaller scales, on dwarf galaxy evolution, though we note that J. F. Wu, C. K. Jespersen & R. H. Wechsler (2024) recently concluded from an analysis of the IllustrisTNG-300 simulation that the galaxy–halo connection is most significantly influenced by the environment measured on a scale of $\simeq 3$ Mpc.

4.4 Halo concentration as a function of halo mass

The intrinsic concentration of DM haloes is anticorrelated with their mass (e.g. J. F. Navarro et al. 1996; A. D. Ludlow et al. 2014; C. A. Correa et al. 2015). This ‘mass–concentration relation’ arises due to the hierarchical nature of structure formation in CDM cosmogonies, with the characteristic density of a halo being proportional to the mean cosmic mass density at its initial collapse time. Low mass haloes therefore generally exhibit higher concentrations, reflecting their collapse at earlier cosmic epochs,

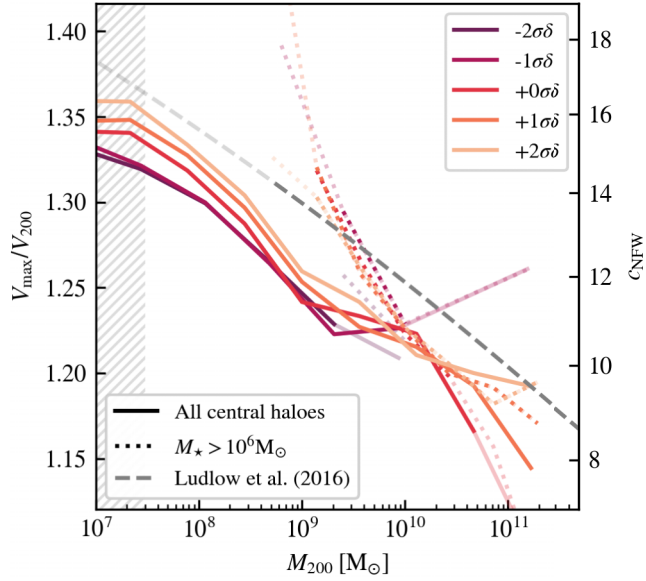


Figure 10. The present-day halo mass–concentration relation, at $m5$ resolution, of the aggregated field halo populations of the five regions comprising each of the five density bands. Solid coloured curves denote, for each density band, the median halo concentration at fixed halo mass, and include the contribution of haloes without a resolved galaxy. The dotted curves show the corresponding relation of only those haloes that host a galaxy of stellar mass $M_* > 10^6 M_\odot$. Halo concentrations are estimated from $V_{\max}/V_{200}$ ratios, as detailed in Section 2.5. The grey dashed line is the fit to the mass–concentration relation presented in A. D. Ludlow et al. (2016) for systems above $M_{200} > 5 \times 10^8 M_\odot$. Colours, line styles, and hatching follow the conventions introduced in Fig. 7. There is a systematic shift in mass–concentration of haloes between different overdensities, especially at the low mass end. Haloes hosting luminous galaxies are more concentrated than average.

making concentration a reasonable proxy for the formation time of a halo (R. H. Wechsler et al. 2002; D. H. Zhao et al. 2009; A. D. Ludlow et al. 2014; C. A. Correa et al. 2015). This implies that regions of differing enclosed density may exhibit halo mass–concentration relations that exhibit systematic offsets from the relation obtained from mean-density volumes.

Fig. 10 shows the present-day median halo concentration at fixed halo mass obtained from the $m5$ simulations. The relation is shown for the population of field haloes identified in the hydrodynamical simulations, but we remind the reader that (per Section 2.5) concentration is measured from each halo’s matched counterpart in the corresponding collisionless simulation, to eliminate the influence of baryons on the inner structure of haloes. Each coloured curve denotes the mass–concentration relation of the galaxy population aggregated from the five simulations comprising each density band, δ_5 , and is drawn with reduced opacity where the $\Delta \log_{10} M_{200} = 0.2$ bin is sampled by fewer than 10 haloes. Dotted curves show the relation recovered from the subset of haloes that host a galaxy of stellar mass $M_* > 10^6 M_\odot$, the threshold corresponding to $\simeq 10$ stellar particles, above which GSMFs of the simulations are reasonably converged (see Section 3.1).

Interestingly, we find that the subset of ‘luminous’ low-mass haloes is strongly biased towards high concentrations, and hence these haloes exhibit a very steep mass–concentration relation. For our choice of stellar mass threshold used to define a ‘luminous’ halo, the mass–concentration relations diverge below

$M_{200} \simeq 3 \times 10^9 M_\odot$, which is the typical mass of haloes hosting galaxies of stellar mass $M_* \simeq 10^6 M_\odot$. Low-mass haloes that host a galaxy therefore have assembly histories that are strongly biased to early formation times. Despite there being a clear influence of environment on the mass–concentration relation of the overall halo population, no such effect is seen for the subset of haloes hosting a galaxy. This implies that halo occupancy is mostly influenced by its assembly history (for which we use concentration here as a proxy) rather than its large-scale environment. From a physical point of view, gas associated with low mass haloes is readily unbound by heating from the metagalactic UV/X-ray background during and subsequent to the epoch of H I reionization. Those present-day low mass haloes whose main progenitor reached a greater mass prior to reionization (i.e. those with an earlier formation time) can more readily form a significant mass of stars by the present day. A similar conclusion was reached by A. Fitts et al. (2017) and M. P. Rey et al. (2019) from analysis of the FIRE and EDGE simulations, respectively.

Our choice of the stellar mass threshold used to define what constitutes a ‘luminous’ galaxy is clearly arbitrary, and follows from resolution considerations. For an alternative choice of threshold, the mass–concentration relations would diverge close to the halo mass that typically hosts galaxies of that stellar mass (see Fig. B2). Relatedly, it is clear that the steep negative gradient of the mass–concentration relation of luminous haloes can be attributed to the steep positive gradient of the low-mass regime of the SMHM relation (see Fig. 5), as small changes in halo mass correspond to large changes in stellar mass. As is also clear from Fig. B2, the inferred mass–concentration relation of luminous haloes is insensitive to resolution, so long as one adopts a stellar mass threshold corresponding to at least 10 stellar particles.

4.5 Halo luminous fraction as a function of mass

An alternative way of viewing the divergence of the mass–concentration relations of dark and luminous haloes, is as a concentration dependence of the luminous fraction of haloes as a function of their mass. The upper panel of Fig. 11 shows the median luminous fraction (where luminous haloes are again defined as those hosting a galaxy of at least $M_* = 10^6 M_\odot$) as a function of halo mass, for the aggregated galaxy populations corresponding to each density band, δ_5 , at $m5$ resolution. The relations recovered from each density band are similar. All haloes with $M_{200} \gtrsim 10^{10} M_\odot$ host a ‘luminous’ galaxy, while there are no luminous galaxies in any halo of $M_{200} \lesssim 10^9 M_\odot$, with the luminous fraction showing a weak dependence on δ_5 close to $M_{200} \lesssim 10^9 M_\odot$. We attribute this difference to the systematic shift in the mass–concentration relation of all haloes at varying δ_5 , apparent in Fig. 10.

The luminous fraction smoothly transitions between $f_{\text{luminous}} = 0$ and 1, with approximately half of haloes with mass $M_{200} \simeq 3 \times 10^9 M_\odot$ being luminous. Unsurprisingly, this is the same mass below which the mass–concentration relation of luminous haloes begins to diverge from the relation of all haloes. As with the latter relation, the position of these curves shifts in response to an alternative choice of the stellar mass threshold used to define a luminous halo (see Fig. B3).

We caution that our definition of the luminous fraction is not equivalent to the absolute halo occupation fraction, and reiterate that our choice of stellar mass threshold ($M_* = 10^6 M_\odot$) is guided by resolution considerations. The luminous fraction based on

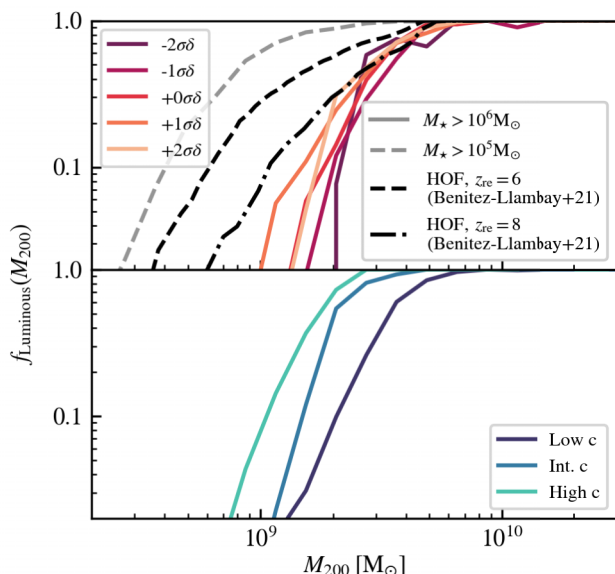


Figure 11. *Upper panel:* The present-day luminous fraction as a function of mass, at m5 resolution, of the aggregated halo population of the five regions comprising the five density bands (coloured curves denote each density band). We deem haloes to be luminous if they host a galaxy with stellar mass $M_* > 10^6 M_\odot$. The luminous fraction is only weakly sensitive to large-scale environment. The grey dashed curve shows the luminous fraction of the aggregated halo population of all 25 simulations (i.e. not split by density band) recovered if one assumes haloes with galaxies $M_* > 10^5 M_\odot$ are luminous. The dashed and dot-dashed black curves show the halo occupation fractions of the semi-analytic model of A. Benítez-Llambay & C. Frenk (2020), assuming an H I reionization redshift of $z = 6$ and $z = 8$. *Lower panel:* The present-day luminous fraction of the total halo population of all 25 simulations at m5 resolution, with haloes binned into the low, intermediate, and high terciles of the intrinsic halo concentration distribution.

this threshold should therefore be considered a lower limit. To highlight the sensitivity of the luminous fraction to a choice of a lower stellar mass fraction, we draw with a dashed grey curve the fraction of haloes, drawn from all 25 simulations at m5 resolution, with at least one stellar particle (broadly corresponding to a ‘galaxy’ mass of $M_* = 10^5 M_\odot$), whilst noting that this test is particularly sensitive to the stochasticity of the simulation’s star formation implementation. We show this curve primarily for illustrative purposes, and to enable comparison with predictions for the absolute halo occupation fraction, which is a highly debated topic in the literature.

Two mass scales influencing the halo occupation fraction are often discussed. Photoheating during the epoch of reionization, and the subsequent presence of a metagalactic UV/X-ray background, unbinds gas from haloes of present-day mass $M_{200} \lesssim 10^{9.5} M_\odot$, suppressing the occupation fraction of less massive haloes below unity. The other mass scale is that corresponding to a halo virial temperature comparable to the hydrogen atomic cooling limit ($T \simeq 7000$ K), which is relevant before the epoch of reionization. This scale is redshift dependent, $M \propto (1+z)^{-3/2}$, and at the present day is $M_{200} \sim 10^9 M_\odot$. Gas associated with haloes of mass below this limit cannot efficiently cool and form stars. A large fraction of haloes with present-day mass $M_{200} < 10^{8.5} M_\odot$ never crossed the atomic hydrogen cooling limit during their formation history and are often predicted to be devoid of stars at the present day. A. Benítez-Llambay & C. Frenk

(2020, hereafter BLF20) construct a semi-analytical model based on these mass scales and realistic halo growth histories to estimate the absolute halo occupation fraction, with cooling, the UV background, and reionization assumptions that differ from those in our simulations. The black dashed and dash-dotted curves in the top panel of Fig. 11 show the halo occupation fraction of this model when the epoch of reionization is assumed to be $z_{\text{reion}} = 6$ and 8, respectively, encompassing what is assumed in our simulations ($z_{\text{reion}} = 7.5$). Despite our simulations’ necessary limitation of a minimum mass threshold of $M_* > 10^5 M_\odot$ (which renders this prediction somewhat of a lower limit for the true occupation fraction), the model of BLF20 predicts noticeably lower occupation fractions than our simulations. This mismatch can be attributed to a number of differences between the two models. The model of BLF20 assumes gas is in hydrostatic equilibrium within spherical dark matter haloes of mean concentration, and that the gas contributes negligibly to the gravitational potential at all radii, i.e. gas does not become self-gravitating. Moreover, the gas in the centre of haloes in their model is assumed to be isothermal at $T \sim 10^4$ K (no cold phase). Other potentially relevant differences are the star formation criteria, assumed cooling rates, and the amplitude of the background UV ionizing radiation. Identifying the source of the difference between our prediction and that of BLF20 in more detail is beyond the scope of this paper and an interesting topic for future work.

The weak dependence of the mass–concentration relation on δ_5 for luminous haloes seen in Fig. 10 indicates that internal halo structure (i.e. assembly-bias related effects) is a more important factor than large-scale environment in governing the stellar mass of haloes. To test this explicitly, in the lower panel of Fig. 11, we revert to our standard threshold of $M_* = 10^6 M_\odot$, and show the median luminous fraction of haloes drawn from all 25 simulations, binned into the lower, central and upper terciles of the intrinsic halo concentration distribution: high concentration haloes have $c_{\text{NFW}} > 18.1$, intermediate concentration haloes have $12.7 < c_{\text{NFW}} < 18.1$, and low concentration haloes have $c_{\text{NFW}} < 12.7$. The luminous fraction changes significantly and systematically with halo concentration. For example, at $M_{200} = 2 \times 10^9 M_\odot$, the occupation fraction changes from $\simeq 10$ per cent for low concentration haloes to $\simeq 80$ per cent for high concentration haloes. The halo masses corresponding to 50 per cent occupancy for low, intermediate, and high concentration bins are $M_{50} = 3.2 \times 10^9 M_\odot$, $2.1 \times 10^9 M_\odot$, and $1.7 \times 10^9 M_\odot$, respectively. These results further underscore that the formation and evolution of low-mass dwarf galaxies are closely tied to their halo assembly history, rather than the large-scale environment.

5 SCATTER IN THE STELLAR MASS–HALO MASS RELATION

The findings presented in the previous section highlight that the influence of the large scale ($\simeq 5$ Mpc) environment on the properties of low-mass central galaxies is driven primarily from differences in the overall halo population (such as the distributions of halo concentration or halo formation time) that emerge in differing environments. In this section, we therefore examine how these key halo properties that mediate the environmental differences correlate with the scatter of the SMHM relation. Section 5.1 focuses on halo concentration and Section 5.2 on galaxy formation times.

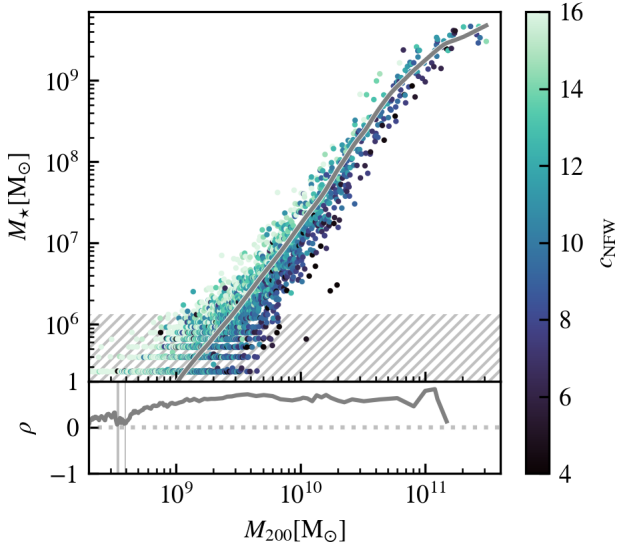


Figure 12. *Top:* The stellar mass versus halo mass (M_{200}) for field galaxies in our 25 simulations at m5 resolution. Each symbol represents an individual galaxy and is coloured by the halo concentration, c_{NFW} , of their dark matter hosts, measured from the V_{max}/V_{200} ratio in the counterpart collisionless simulations. The grey curve shows the median relation computed using the LOWESS method. The scatter at a given halo mass shows a clear trend with concentration, with higher stellar masses in more concentrated haloes. The grey hatched region indicates where galaxies have fewer than 10 star particles. *Bottom:* The running Spearman rank correlation coefficient, ρ , between the residuals Δc_{NFW} and ΔM_* . Positive values close to 1 show a strong positive correlation for all well resolved objects. The shaded regions indicate when the correlation has low significance ($p > 0.01$).

5.1 Correlation with halo concentration

Fig. 12 shows the SMHM relation of the composite galaxy population drawn from all 25 regions at m5 resolution. Symbol colour denotes the intrinsic concentration of the halo, c_{NFW} , and the solid grey curve shows the running median of M_* as a function of M_{200} , computed using the locally weighted scatter plot smoothing method (LOWESS; W. S. Cleveland 1979). We use this method rather than a binned median as it enables the calculation of a median value corresponding to the precise mass of each halo. In the range $M_* = [10^6 - 10^9] M_\odot$, this curve is well fit by a power law of form $M_*/M_\odot = \beta (M_{200}/10^{10} M_\odot)^\alpha$ where $(\alpha, \beta) = (2.17, 1.42 \times 10^7)$, as shown in Appendix D.

Scatter about the SMHM relation at fixed halo mass is strongly and positively correlated with halo concentration such that more concentrated haloes form more massive galaxies. This relation has previously been reported elsewhere for large samples of more massive ($M_{200} \gtrsim 10^{11} M_\odot$) simulated galaxies (e.g. J. Matthee et al. 2017, see also A. D. Montero-Dorta et al. 2020; K. Wang et al. 2026). To quantify the strength of the correlation as a function of halo mass, we compute for each galaxy, i , the offset between the value of $M_{*,i}$ and the LOWESS median of this quantity at $M_{200,i}$, which we define as $\Delta M_{*,i}$. We then compute the equivalent metric for halo concentration, $\Delta c_{\text{NFW},i}$, and compute the Spearman rank correlation coefficient, ρ , of the two quantities in samples defined by a ‘rolling’ window in halo mass. This quantity is plotted in the subpanel of Fig. 12. Solid shading in the panel corresponds to halo mass windows in which the probability of a

chance correlation is high, denoted by Spearman p -value greater than 0.01. A significant positive correlation between stellar mass and halo concentration at fixed halo mass is present for $M_{200} \gtrsim 5 \times 10^8 M_\odot$, though we note that for $M_{200} \lesssim 5 \times 10^9 M_\odot$, the finite resolution of the simulation hinders the sampling of the stellar mass, as is clear from the diagonal hatching in the main panel of the plot.

J. Matthee et al. (2017) attributed the correlation between stellar mass and halo concentration at fixed halo mass to a combination of earlier halo assembly and the increased gravitational binding energy of more concentrated haloes, which reduces the (macroscopic) efficiency of feedback. More recently, K. Wang et al. (2026) argue that, for haloes with $M_{200} \approx 10^{11} - 10^{12} M_\odot$, variations in binding energy are the dominant driver of this correlation. As also noted by A. Fitts et al. (2017), at the low halo masses explored here ($M_{200} \lesssim 10^{10} M_\odot$) photoheating during and after reionization also has a significant influence on the halo’s SFH. The earlier assembly of more concentrated haloes affords them an elevated probability of forming stars prior to the epoch of reionization, and owing to their greater binding energy they lose a smaller fraction of their gas to photoheating. The scatter about the SMHM relation in the simulations increases towards lower halo masses, becoming particularly large below $M_{200} \lesssim 10^{10} M_\odot$. Quantified in 0.2 dex bins, the 10th–90th percentile scatter in M_* at fixed M_{200} increases from 0.71 dex at $M_{200} = 1.7 \times 10^{10} M_\odot$ to 0.85 dex at $M_{200} = 4.5 \times 10^9 M_\odot$. We speculate that this may be a signature of the role played by reionization in exacerbating the dependence of M_*/M_{200} on halo concentration and formation time.

It has long been argued that stellar mass and luminosity should correlate more strongly with the maximum value of a halo’s circular velocity profile than with its halo mass (e.g. D. Nagai & A. V. Kravtsov 2005; C. Conroy, R. H. Wechsler & A. V. Kravtsov 2006), as the maximum circular velocity more directly traces the depth of the central potential well, and hence the ability to retain gas against feedback-driven outflows, independently of the halo’s outer mass distribution. J. Matthee et al. (2017), using the EAGLE simulations, and A. Fitts et al. (2017), using 15 zoom-in simulations of dwarf galaxies from the FIRE simulations, found this to indeed be the case. We therefore show in Fig. 13 the stellar mass, M_* , of field galaxies drawn from the composite galaxy population as a function of the maximum value of their total circular velocity profile, V_{max} . Symbol colour denotes the halo concentration, c_{NFW} and, as in Fig. 12, the solid black curve shows the median $M_* - V_{\text{max}}$ relation computed using the LOWESS method, which can be fit using the function $M_*[M_\odot] = m_0 v^\alpha \exp(-v^\gamma)$ where $v = V_{\text{max}}/50 \text{ km s}^{-1}$ and $(m_0, \alpha, \gamma) = (3.39 \times 10^8 M_\odot, 7.56, 1.59)$. The dashed grey curve shows the corresponding relation inferred from the APOSTLE simulations (A. Fattahi et al. 2018). The two suites yield a similar outcome: over the range $30 < V_{\text{max}} < 100 \text{ km s}^{-1}$, the largest difference between the median values of M_* at fixed V_{max} is 0.41 dex, which is found at the low mass end where the relation is steep.

As expected, the scatter in M_* at fixed V_{max} is far smaller than at fixed M_{200} , with the difference most-readily quantified via the scatter in the x -axis quantity at fixed M_* : in the interval $10^7 < M_*/M_\odot < 10^8$, the 10th–90th percentile scatter at fixed V_{max} is 0.14 dex, whilst the corresponding value at fixed M_{200} is 0.47 dex. The subpanel of Fig. 13 shows that only a small residual correlation between M_* and c_{NFW} at fixed V_{max} remains, and then only for resolved haloes with $V_{\text{max}} \lesssim 50 \text{ km s}^{-1}$. As previously demonstrated by J. Matthee et al. (2017) at larger halo mass scales, halo

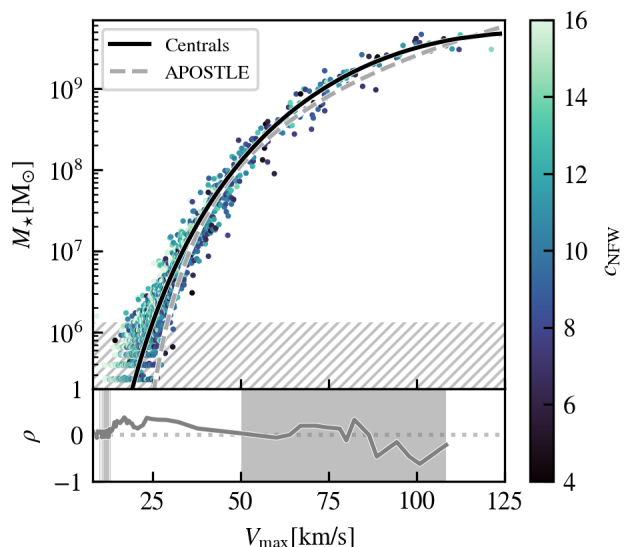


Figure 13. The same as Fig. 12 but showing $V_{\max}$ instead of halo mass. The scatter in stellar mass is smaller, compared to the SMHM relation of Fig. 12. Moreover, unlike the $M_* - M_{200}$ relation, the scatter in the $M_* - V_{\max}$ relation shows no clear correlation with concentration. The running Spearman rank correlation coefficient is shown in the bottom panel. A fit to the median relation is shown in black, following the form $M_*/M_\odot = m_0 \nu^\alpha \exp(-\nu^\gamma)$ where $\nu = V_{\max}/50 \text{ km s}^{-1}$ and $(m_0, \alpha, \gamma) = (3.39 \times 10^8, 7.56, 1.59)$. A fit to the APOSTLE simulation (A. Fattahi et al. 2018) data is shown by the dashed grey curve.

concentration (or, equivalently, halo binding energy) accounts for the bulk, but not all, of the scatter about the SMHM relation.

5.2 Correlation with galaxy formation time

Having established that the scatter about the SMHM relation in the low-mass regime is strongly and positively correlated with halo concentration, a corollary is that the scatter might also correlate with the galaxy’s SFH, owing to the correlation between halo concentration and halo assembly history. We therefore explore the influence of halo concentration on the SFHs of our simulated galaxies by showing the cumulative, normalized SFH of galaxies with present-day stellar mass of $10^7 < M_*/M_\odot < 10^8$ in the left panel of Fig. 14, binned into tertiles of present-day intrinsic halo concentration. The solid curves and shaded regions denote the median and interquartile range of the cumulative SFHs. There is a significant systematic trend such that galaxies hosted by more concentrated haloes typically form their stars earlier. The difference in SFHs is not only present at early times, but persists throughout the assembly of the galaxies; e.g. the cosmic time by which 90 per cent of present-day stellar mass has formed, t_{90} , is 4.4, 7.4, and 10.4 Gyr for the median SFH of galaxies in the high, intermediate, and low concentration samples, respectively. These times are denoted by vertical dashed lines on the figure.

Since halo mass and concentration correlate, we isolate the influence of concentration alone on the previous trend by showing in the right panel of Fig. 14 the SFHs of galaxies hosted by haloes within a narrow range of halo mass, $10^{10} < M_{200}/M_\odot < 2 \times 10^{10}$. The trend between SFH and halo concentration is still evident: high concentration haloes, i.e. earlier-forming haloes host earlier-forming galaxies. The average t_{90} , indicated by the vertical dashed lines, shows a systematic shift to earlier cosmic times as the halo

concentrations increase, corresponding to 5.7, 8.0, and 10.8 Gyr for the high, intermediate, and low concentration bins, respectively. These results emphasize the link between halo assembly and stellar assembly times.

The previous finding indicates a likely correlation between SFH and the scatter about the SMHM relation since, as shown in Section 5, the latter correlates with halo concentration. We explore this in Fig. 15 where we repeat Fig. 12, but here colour-code the symbols by t_{90} . In the regime where galaxy masses, and hence SFHs, are well sampled, there is an evident anticorrelation between the scatter about the SMHM relation and t_{90} .¹² At fixed halo mass, galaxies with greater stellar mass have lower t_{90} and form their stars earlier. The running Spearman rank correlation coefficient between the residuals of stellar mass, $\Delta M_{*,i}$ and $t_{90,i}$, is shown as a function of halo mass in the bottom panel of Fig. 15, which highlights the anticorrelation ($\rho < 0$) for the halo mass interval $4 \times 10^9 \lesssim M_{200}/M_\odot \lesssim 10^{11}$. At greater masses, SFHs are influenced by AGN feedback and the role of concentration is more complicated in such galaxies because it can accelerate the growth of the central BH (e.g. J. J. Davies et al. 2019).

The aforementioned anticorrelation between M_* and t_{90} may appear to contradict the expectation that higher mass dwarf galaxies have more prolonged star formation and hence larger t_{90} (see e.g. C. M. Simpson et al. 2018; R. Digby et al. 2019). However, we emphasize that the anticorrelation applies when considering galaxies at fixed halo mass, and does not apply to the average $M_* - t_{90}$ relation where there is a clear overall trend such that more massive galaxies have a later formation time, except for those galaxies hosted by haloes of mass $M_{200} \gtrsim 10^{11} M_\odot$ whose properties are not accurately reproduced by these simulations (see Section 3). The positive correlation between M_* and t_{90} is evident in Fig. C1.

The star formation history of dwarf galaxies, especially at late cosmic times, is more readily inferred from observations than their halo concentration. Thus, the correlation of the scatter about the SMHM relation and the $t_{90} - M_*$ relation provides a test of our galaxy formation model, one that can, in principle, be corroborated or falsified by observations. We note that the predicted shape of, and scatter about, the SMHM relation are sensitive to the details of the adopted galaxy formation model, but we anticipate that the correlations are reasonably robust to such details because they link properties of DM haloes (halo assembly and concentration) to the coarse-grained stellar assembly history. Even though we do not provide detailed comparison of SFHs or colour of our simulated dwarfs with observations, as it goes beyond the scope of the paper, we emphasize that the trends are expected to be robust. Indeed, S. Y. Kim et al. (2026) report that in the Engineering Dwarfs at Galaxy formation’s Edge (EDGE) simulations the halo mass is more strongly correlated with the average star formation rates of dwarfs than with their absolute stellar mass. This is also motivated by observational constraints, as argued by J. I. Read & D. Erkal (2019) for dwarf galaxies in the Local Group.

6 SUMMARY AND DISCUSSION

We have introduced Columba, a suite of cosmological hydrodynamical simulations of the Λ CDM cosmogony that follow the formation and evolution of isolated dwarf galaxies in diverse

¹²Our results are qualitatively similar if using t_{10} or t_{50} .

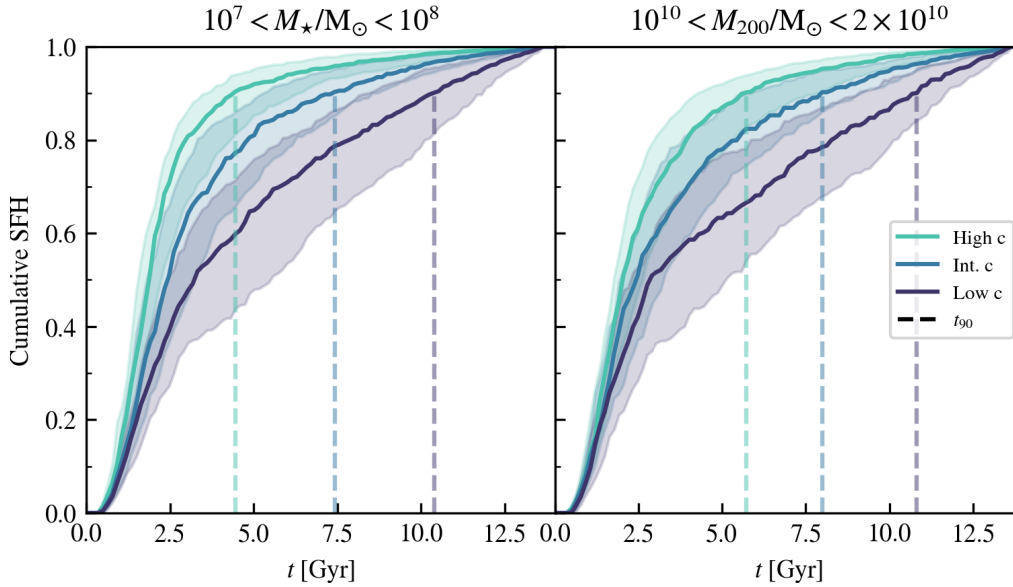


Figure 14. The cumulative star formation histories (SFH) of central galaxies in the simulation suite, divided by their stellar mass at $z = 0$ and their dark matter halo concentration. The left panel corresponds to a present-day stellar mass bin ranging between $10^7 < M_*/M_\odot < 10^8$ and the right panel corresponds to a present-day halo mass bin of $10^{10} < M_{200}/M_\odot < 2 \times 10^{10}$. In both panels, the sample is split into three based on the terciles of the halo concentration distribution. The cumulative SFHs are normalized to the $z = 0$ stellar mass first, and then median and scatter (25th–75th percentiles) are computed and shown by the solid lines and shaded regions, respectively. The median times at which 90 per cent of the final stellar mass has formed, t_{90} , are shown by the vertical dashed lines. Consistent over the full mass range, higher concentration galaxies have earlier SF, reflected by earlier t_{90} . Time and t_{90} correspond to cosmic time.

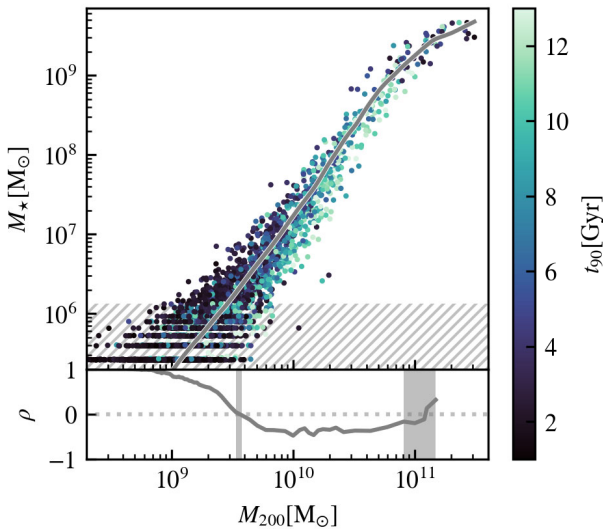


Figure 15. Similar to Fig. 12 but points are colour-coded according to t_{90} , the cosmic time at which 90 per cent of the total stellar mass of a galaxy has formed.

low-density environments. The suite is comprised of 25 zoom-in simulations that follow the evolution of spherical regions with radius $r = 5$ Mpc at $z = 0$, drawn from a collisionless simulation of a periodic volume with side length $L = 400$ cMpc. The regions, which span diverse cosmic environments from voids to filament-like structures, are devoid of massive haloes with present-day mass comparable to or larger than that of the Milky Way. The average density of the target regions at $z = 0$ corresponds to the $(-2, -1, 0, +1, +2)\sigma$ percentiles of the distribution of the

mass density of present-day $r = 5$ Mpc spheres that are devoid of massive haloes. The corresponding overdensities are $\log_{10} \delta_5 = (-1.064, -0.859, -0.645, -0.449, -0.282)$, where $\delta_5 = \rho_5/\bar{\rho}$, ρ_5 is the mean density within a sphere of radius $r = 5$ Mpc and $\bar{\rho}$ is the mean cosmic mass density. The five regions selected at each density, δ_5 , were chosen to further sample the distribution of the number of haloes with mass broadly corresponding to that of the Large Magellanic Cloud, N_{LMC} , at the corresponding density.

The simulations were evolved to the present day with the cosmology, gravity, hydrodynamics, and galaxy formation code *sWIFT* (M. Schaller et al. 2024), with approximately equal mass DM and gas particles, at two levels of resolutions. The first, ‘m6’, corresponds to a gas particle mass of $m_{\text{gas}} = 1.8 \times 10^6 M_\odot$, while the flagship ‘m5’ resolution corresponds to $m_{\text{gas}} = 2.3 \times 10^5 M_\odot$. The simulations adopt a galaxy formation model that incorporates substantial improvements over *EAGLE* and shares several key components with the *COLIBRE* model (E. Chaikin et al. 2026; J. Schaye et al. 2026). These include an explicit model of cold, dense neutral gas, a star formation model that accounts for gas turbulence, a treatment of pre-SN feedback from massive stars, more sophisticated methods of distributing the energy liberated by SNe and AGNs, and updated treatments of BH growth and dynamics. While *Columba* inherits several modules from *COLIBRE*, it does not include a live dust model coupled to radiative cooling, a non-equilibrium H/He network, turbulent diffusion, updated chemical enrichment and SNIa schemes, or the jet-mode AGN feedback channel. The underlying subgrid prescriptions inherit parameter choices that result from testing performed for the *EAGLE* (R. A. Crain et al. 2015) and *COLIBRE* (E. Chaikin et al. 2026) simulation campaigns. As in all cosmological simulations, additional implementation choices such as the feedback heating temperature, and choice of particles selected to be directly influenced by feedback processes also indirectly influence galaxy

properties. Unlike COLIBRE, we also do not vary the subgrid efficiency of feedback from SNe with resolution. Nevertheless, when rescaled to account for their preferential sampling of underdense regions, the simulations reproduce the present-day GSMF in the stellar mass regime they resolve and sample very well, $M_* = 10^6 - 10^9 M_\odot$. The galaxies are slightly larger than their similar-mass counterparts observed in the Local Group, by approximately 60 per cent for the m5 resolution simulations.

We investigate the influence of environment, characterized by δ_5 , on the population of simulated dwarf galaxies. Our main findings are as follows:

(i) Once the difference in the mean densities of the regions is accounted for, the low-mass power-law regime of the HMFs of the simulated regions are very similar, as shown in the bottom panel of Fig. 7. There is a mild second-order effect, such that the density-normalized space-density of haloes at fixed mass ($\phi(M)/\rho_5$) is greater in lower density regions. The characteristic mass scale at which the HMFs transition from a power law to an exponential decline changes systematically with δ_5 , such that lower density regions host fewer high-mass haloes. This characteristic mass scale is $M_{200} \sim 10^9 M_\odot$ in the lowest density regions, affecting the space density of haloes that host dwarf galaxies.

(ii) The HMFs are mildly suppressed compared to their collisionless ('DM-only') counterparts in the low-mass regime as illustrated in the middle panel of Fig. 7. The suppression of the HMF is insensitive to the overdensity of the region; i.e. $\phi_{\text{hydro}}/\phi_{\text{DMO}} \simeq 0.8$, for a halo mass of $M_{200} < 10^9 M_\odot$, for all environments.

(iii) Similar to the HMFs, Fig. 8 shows that the GSMF of regions with different mean density differ in the low-mass regime predominantly in their normalization. The lowest density regions have systematically lower density-normalized GSMFs ($\phi(M_*)/\rho_5$) at all resolved stellar masses. We attribute this difference primarily to the dependence of the HMF break scale on δ_5 .

(iv) The median SMHM relation over the mass scales examined here can be represented by a power law of form $M_*/M_\odot = \beta(M_{200}/10^{10} M_\odot)^\alpha$ where $(\alpha, \beta) = (2.17, 1.42 \times 10^7)$, fitted to the range $M_* = [10^6 - 10^9] M_\odot$. Fig. 9 shows that we find no significant correlation of the median SMHM relation and δ_5 .

(v) Fig. 10 shows that the median halo concentration at fixed halo mass is mildly correlated with δ_5 , consistent with the trend that haloes in denser regions tend to have higher concentrations (see also W. A. Hellwing et al. 2021). However, the median halo concentration of 'luminous' haloes (those hosting a galaxy of $M_* > 10^6 M_\odot$ at m5 resolution) is insensitive to δ_5 . The mass-concentration relation of this population diverges steeply from that of all haloes for $M_{200} \lesssim 3 \times 10^9 M_\odot$ towards higher concentrations.

(vi) The fraction of luminous haloes, displayed in Fig. 11, shows a small (<30 per cent between the regions of highest and lowest δ_5) but systematic correlation with δ_5 , such that the luminous fraction at fixed mass is greater in higher-density regions. We attribute this trend to a strong dependence of the luminous fraction on the intrinsic DM halo concentration, c_{NFW} . For example, for haloes of $M_{200} \simeq 2 \times 10^9 M_\odot$, the luminous fraction differs by 63 per cent between the upper and lower concentration terciles.

Our study highlights the important and closely linked roles of halo assembly and halo concentration on the formation and evolution of dwarf galaxies. We have shown in Fig. 12 that in the low-mass regime ($M_{200} = 10^9 - 10^{11} M_\odot$) the scatter in the $M_* - M_{200}$ relation correlates strongly and positively with the concentration

of a galaxy's host DM halo, when measured using the V_{max}/V_{200} ratio from the matched DM-only counterparts. A corollary of this correlation is that there is less scatter about the $M_* - V_{\text{max}}$ relation (Fig. 13) than about the $M_* - M_{200}$ relation. Scatter about the former relation does not correlate significantly with DM halo concentration. These findings are consistent with those of previous studies examining higher resolution simulations of small samples of dwarf galaxies (e.g. A. Fitts et al. 2017; M. P. Rey et al. 2019).

Studies based on hydrodynamical and/or semi-analytic simulations of large periodic volumes have previously revealed the strong correlation of scatter about the SMHM relation with the formation time or concentration of galaxy host DM haloes on relatively high stellar mass scales ($M_{\text{halo}} \gtrsim 10^{11} M_\odot$; e.g. J. Matthee et al. 2017; M. C. Artale et al. 2018; I. Zehavi et al. 2018; K. Wang et al. 2026). The large samples yielded by our simulations demonstrate that these trends extend into the dwarf galaxy regime as shown in Fig. 12.

In Fig. 14, we show that the SFHs of our simulated dwarf galaxies, either at fixed halo mass or fixed stellar mass, correlate strongly with halo concentrations. Lower concentration haloes, which typically form later, have SFHs delayed by ≈ 5 Gyr in comparison to their high concentration counterparts. This link between SFH and halo concentration (halo assembly) leaves its imprint in the scatter of the SMHM relation. The scatter of the SMHM relation at fixed halo mass is negatively correlated with t_{90} ; i.e. higher stellar mass galaxies form the majority of their stars earlier (smaller t_{90}), compared to their lower stellar mass counterparts at any given halo mass.

C. Gallart et al. (2015) propose two distinct evolutionary paths, fast and slow, based on the observed SFHs of Local Group dwarf galaxies. They hypothesize that these evolutionary paths are associated with the characteristic density of the environment where the dwarfs are formed. We find diversity in the SFH of dwarfs as opposed to two distinct categories. However, we find qualitative agreement with their proposal about the origin of the difference in the SFHs.

The underlying physics behind the correlation of stellar assembly and halo concentration (halo assembly) is driven by a combination of various processes; the gas accretion rate on to haloes, the structure of their gravitational potential, and gas heating during and subsequent to the epoch of reionization. The former two are likely more important at higher halo masses ($M_{200} \gtrsim 10^{11} M_\odot$), whereas processes related to cosmic reionization are thought to be important for less massive haloes ($M_{200} \lesssim 10^9 M_\odot$). The details of these processes and the transition scale between various mechanisms are an interesting subject of investigation for future studies, particularly using simulations that explicitly model radiation transport.

The importance of the correlations we find here between M_* , M_{200} , c_{NFW} , and t_{90} lies in the fact that they link observable properties of dwarf galaxies, i.e. their stellar mass and SFHs, to their DM halo properties which are less readily accessible. In particular, we make clear observable predictions for the scatter of the SMHM relation which can be tested with upcoming measurements of properties of large samples of isolated dwarf galaxies in the local Universe, using *Euclid*, DESI, and the future Rubin and Roman Observatories.

ACKNOWLEDGEMENTS

We thank Adrian Jenkins for generous advice concerning the generation of the initial conditions for this project. JMB and AF

acknowledge support by a UK Research and Innovation Future Leaders Fellowship (grant no. MR/T042362/1). AF was supported by a Wallenberg Academy Fellowship. RAC acknowledges support from Science and Technology Facilities Council (STFC) Small Astronomy Awards ST/Y002474/1 and ST/Y002482/1. SP acknowledges support from the Austrian Science Fund (FWF) through grant no. V 982-N. ABL acknowledges support by the Italian Ministry for Universities (MUR) program ‘Dipartimenti di Eccellenza 2023–2027’ within the Centro Bicocca di Cosmologia Quantitativa (BiCoQ), and support by UNIMIB’s Fonuk redo Di Ateneo Quota Competitiva (project 2024-ATEQC-0050). YMB acknowledges support from United Kingdom Research and Innovation (UKRI) through a Future Leaders Fellowship (grant agreement MR/X035166/1). This work used the Prospero HPC facility at Liverpool John Moores University, and the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1, ST/R002371/1, and ST/S002502/1, Durham University, and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure. This research has made use of NASA’s Astrophysics Data System.

DATA AVAILABILITY

Data produced in this study can be made available upon reasonable request to the authors.

A public version of the SWIFT hydrodynamics code is available at <http://www.swiftsim.com>. The CHIMES non-equilibrium thermochemistry code is publicly available at <https://richings.bitbucket.io/chimes/home.html>.

REFERENCES

- Abbott T. M. C. et al., 2022, *Phys. Rev. D*, 105, 023520
- Ali-Haïmoud Y., Bird S., 2013, *MNRAS*, 428, 3375
- Applebaum E., Brooks A. M., Christensen C. R., Munshi F., Quinn T. R., Shen S., Tremmel M., 2021, *ApJ*, 906, 96
- Artale M. C., Zehavi I., Contreras S., Norberg P., 2018, *MNRAS*, 480, 3978
- Bahé Y. M. et al., 2022, *MNRAS*, 516, 167
- Behroozi P. S., Wechsler R. H., Conroy C., 2013, *ApJ*, 770, 57
- Bellovary J. M., Cleary C. E., Munshi F., Tremmel M., Christensen C. R., Brooks A., Quinn T. R., 2019, *MNRAS*, 482, 2913
- Benítez-Llambay A., Frenk C., 2020, *MNRAS*, 498, 4887 (BLF20)
- Benítez-Llambay A., Navarro J. F., Abadi M. G., Gottlöber S., Yepes G., Hoffman Y., Steinmetz M., 2015, *MNRAS*, 450, 4207
- Benítez-Llambay A. et al., 2026, *MNRAS*, 546, stag268
- Bocquet S., Saro A., Dolag K., Mohr J. J., 2016, *MNRAS*, 456, 2361
- Booth C. M., Schaye J., 2009, *MNRAS*, 398, 53
- Borrow J., Schaller M., Bower R. G., Schaye J., 2021, *MNRAS*, 511, 2367
- Borrow J., Schaller M., Bahé Y. M., Schaye J., Ludlow A. D., Ploekinger S., Nobels F. S. J., Altamura E., 2023, *MNRAS*, 526, 2441
- Bower R. G., Benson A. J., Crain R. A., 2012, *MNRAS*, 422, 2816
- Boylan-Kolchin M., Bullock J. S., Kaplinghat M., 2011, *MNRAS*, 415, L40
- Bryan G. L., Norman M. L., 1998, *ApJ*, 495, 80
- Buck T., Obreja A., Macciò A. V., Minchev I., Dutton A. A., Ostriker J. P., 2020, *MNRAS*, 491, 3461
- Bullock J. S., Boylan-Kolchin M., 2017, *ARA&A*, 55, 343
- Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001, *MNRAS*, 321, 559
- Carlsten S. G., Greene J. E., Greco J. P., Beaton R. L., Kado-Fong E., 2021, *ApJ*, 922, 267
- Carlsten S. G., Greene J. E., Beaton R. L., Danieli S., Greco J. P., 2022, *ApJ*, 933, 47
- Chabrier G., 2003, *PASP*, 115, 763
- Chaikin E., Schaye J., Schaller M., Bahé Y. M., Nobels F. S. J., Ploekinger S., 2022, *MNRAS*, 514, 249
- Chaikin E., Schaye J., Schaller M., Benítez-Llambay A., Nobels F. S. J., Ploekinger S., 2023, *MNRAS*, 523, 3709
- Chaikin E. et al., 2026, *MNRAS*, 548, stag740
- Cleveland W. S., 1979, *J. Am. Stat. Assoc.*, 74, 829
- Conroy C., Wechsler R. H., Kravtsov A. V., 2006, *ApJ*, 647, 201
- Correa C. A., Wyithe J. S. B., Schaye J., Duffy A. R., 2015, *MNRAS*, 452, 1217
- Crain R. A., van de Voort F., 2023, *ARA&A*, 61, 473
- Crain R. A., Eke V. R., Frenk C. S., Jenkins A., McCarthy I. G., Navarro J. F., Pearce F. R., 2007, *MNRAS*, 377, 41
- Crain R. A. et al., 2009, *MNRAS*, 399, 1773
- Crain R. A. et al., 2015, *MNRAS*, 450, 1937
- Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, *MNRAS*, 486, 2827
- Davies J. J., Crain R. A., McCarthy I. G., Oppenheimer B. D., Schaye J., Schaller M., McAlpine S., 2019, *MNRAS*, 485, 3783
- Davies J. J., Crain R. A., Pontzen A., 2021, *MNRAS*, 501, 236
- Davies J. J., Pontzen A., Crain R. A., 2022, *MNRAS*, 515, 1430
- de Graaff A., Trayford J., Franx M., Schaller M., Schaye J., van der Wel A., 2022, *MNRAS*, 511, 2544
- Dekel A., Silk J., 1986, *ApJ*, 303, 39
- Di Matteo T., Colberg J., Springel V., Hernquist L., Sijacki D., 2008, *ApJ*, 676, 33
- Diemer B., 2018, *ApJS*, 239, 35
- Digby R. et al., 2019, *MNRAS*, 485, 5423
- Driver S. P. et al., 2022, *MNRAS*, 513, 439
- Dubois Y. et al., 2021, *A&A*, 651, A109
- Duffy A. R., Schaye J., Kay S. T., Dalla Vecchia C., Battye R. A., Booth C. M., 2010, *MNRAS*, 405, 2161
- Durier F., Dalla Vecchia C., 2012, *MNRAS*, 419, 465
- Elahi P. J., Cañas R., Poulton R. J. J., Tobar R. J., Willis J. S., Lagos C. d. P., Power C., Robotham A. S. G., 2019, *Publ. Astron. Soc. Aust.*, 36, e021
- Eldridge J. J., Stanway E. R., Xiao L., McClelland L. A. S., Taylor G., Ng M., Greis S. M. L., Bray J. C., 2017, *PASA*, 34, e058
- Fattahi A., Navarro J. F., Frenk C. S., Oman K. A., Sawala T., Schaller M., 2018, *MNRAS*, 476, 3816
- Fattahi A., Navarro J. F., Frenk C. S., 2020, *MNRAS*, 493, 2596
- Feldmann R. et al., 2023, *MNRAS*, 522, 3831
- Feng Y., Di-Matteo T., Croft R. A., Bird S., Battaglia N., Wilkins S., 2016, *MNRAS*, 455, 2778
- Ferland G. J. et al., 2017, *Rev. Mex. Astron. Astrofis.*, 53, 385
- Fitts A. et al., 2017, *MNRAS*, 471, 3547
- Flores R. A., Primack J. R., 1994, *ApJ*, 427, L1
- Font A. S. et al., 2020, *MNRAS*, 498, 1765
- Frenk C. S., White S. D. M., Davis M., Efstathiou G., 1988, *ApJ*, 327, 507
- Furlong M. et al., 2017, *MNRAS*, 465, 722
- Gallart C. et al., 2015, *ApJ*, 811, L18
- Gao L., Springel V., White S. D. M., 2005, *MNRAS*, 363, L66
- Garrison-Kimmel S., Boylan-Kolchin M., Bullock J. S., Lee K., 2014, *MNRAS*, 438, 2578
- Geha M., Blanton M. R., Yan R., Tinker J. L., 2012, *ApJ*, 757, 85
- Geha M. et al., 2017, *ApJ*, 847, 4
- Geha M. et al., 2024, *ApJ*, 976, 118
- Genel S. et al., 2018, *MNRAS*, 474, 3976
- Genel S. et al., 2019, *ApJ*, 871, 21
- Grand R. J. J. et al., 2017, *MNRAS*, 467, 179
- Guedes J., Callegari S., Madau P., Mayer L., 2011, *ApJ*, 742, 76
- Gutcke T. A., Pakmor R., Naab T., Springel V., 2021, *MNRAS*, 501, 5597
- Hahn O., Michaux M., Rampf C., Uhlemann C., Angulo R. E., 2020, *Astrophysics Source Code Library*, record ascl:2008.024
- Hardwick J. A., Cortese L., Obreschkow D., Catinella B., Cook R. H. W., 2022, *MNRAS*, 509, 3751
- Hellwing W. A., Cautun M., van de Weygaert R., Jones B. T., 2021, *Phys. Rev. D*, 103, 063517
- Henden N. A., Puchwein E., Shen S., Sijacki D., 2018, *MNRAS*, 479, 5385

- Hirschmann M., Dolag K., Saro A., Bachmann L., Borgani S., Burkert A., 2014, *MNRAS*, 442, 2304
- Hopkins P. F. et al., 2018, *MNRAS*, 480, 800
- Jenkins A., 2010, *MNRAS*, 403, 1859
- Jenkins A., 2013, *MNRAS*, 434, 2094
- Katz N., White S. D. M., 1993, *ApJ*, 412, 455
- Keller B. W., Wadsley J. W., Wang L., Kruijssen J. M. D., 2019, *MNRAS*, 482, 2244
- Kim S. Y. et al., 2026, *MNRAS*, 549, stag825
- Klypin A., Kravtsov A. V., Valenzuela O., Prada F., 1999, *ApJ*, 522, 82
- Krumholz M. R., Tan J. C., 2007, *ApJ*, 654, 304
- Krumholz M. R., McKee C. F., Klein R. I., 2006, *ApJ*, 638, 369
- Kugel R. et al., 2023, *MNRAS*, 526, 6103pace
- Lacey C., Silk J., 1991, *ApJ*, 381, 14
- Lacki B. C., Thompson T. A., Quataert E., 2010, *ApJ*, 717, 1
- Lesgourgues J., 2011, preprint (arXiv:1104.2932)
- Lesgourgues J., Pastor S., 2006, *Phys. Rep.*, 429, 307
- Lesgourgues J., Tram T., 2011, *J. Cosmol. Astropart. Phys.*, 2011, 032
- Lovell C. C., Vijayan A. P., Thomas P. A., Wilkins S. M., Barnes D. J., Irodotou D., Roper W., 2021, *MNRAS*, 500, 2127
- Lovell M. R. et al., 2018, *MNRAS*, 481, 1950
- Ludlow A. D., Navarro J. F., Angulo R. E., Boylan-Kolchin M., Springel V., Frenk C., White S. D. M., 2014, *MNRAS*, 441, 378
- Ludlow A. D., Bose S., Angulo R. E., Wang L., Hellwing W. A., Navarro J. F., Cole S., Frenk C. S., 2016, *MNRAS*, 460, 1214
- Ludlow A. D., Schaye J., Schaller M., Richings J., 2019, *MNRAS*, 488, L123
- Ludlow A. D., Schaye J., Schaller M., Bower R., 2020, *MNRAS*, 493, 2926
- Mao Y.-Y., Geha M., Wechsler R. H., Weiner B., Tollerud E. J., Nadler E. O., Kallivayalil N., 2021, *ApJ*, 907, 85
- Marleau F. R. et al., 2025, preprint (arXiv:2503.15335)
- Matthee J., Schaye J., Crain R. A., Schaller M., Bower R., Theuns T., 2017, *MNRAS*, 465, 2381
- McAlpine S. et al., 2016, *Astron. Comput.*, 15, 72
- McConnachie A. W., 2012, *AJ*, 144, 4
- McGibbon R., Helly J. C., Schaye J., Schaller M., Vandenbroucke B., 2025, *J. Open Source Softw.*, 10, 8252
- Meshveliani T., Lovell M. R., Crain R. A., Pfeffer J., 2024, *MNRAS*, 532, 1296
- Mo H. J., White S. D. M., 1996, *MNRAS*, 282, 347
- Montero-Dorta A. D. et al., 2020, *MNRAS*, 496, 1182
- Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999, *ApJ*, 524, L19
- Moster B. P., Naab T., White S. D. M., 2013, *MNRAS*, 428, 3121
- Munshi F., Brooks A. M., Applebaum E., Christensen C. R., Quinn T., Sligh S., 2021, *ApJ*, 923, 35
- Nagai D., Kravtsov A. V., 2005, *ApJ*, 618, 557
- Navarro J. F., Frenk C. S., White S. D. M., 1996, *ApJ*, 462, 563
- Nelson D. et al., 2018, *MNRAS*, 475, 624
- Nelson D. et al., 2019, *Comput. Astrophys. Cosmol.*, 6, 2
- Nobels F. S. J., Schaye J., Schaller M., Ploekinger S., Chaikin E., Richings A. J., 2024, *MNRAS*, 532, 3299
- Orlando S., Peres G., Reale F., Bocchino F., Rosner R., Plewa T., Siegel A., 2005, *A&A*, 444, 505
- Pace A. B., 2025, *Open J. Astrophys.*, 8, 142
- Pillepich A. et al., 2018a, *MNRAS*, 473, 4077
- Pillepich A. et al., 2018b, *MNRAS*, 475, 648
- Ploekinger S., Schaye J., 2020, *MNRAS*, 497, 4857
- Ploekinger S., Nobels F. S. J., Schaller M., Schaye J., 2024, *MNRAS*, 528, 2930
- Ploekinger S., Richings A. J., Schaye J., Trayford J. W., Schaller M., Chaikin E., 2025, *MNRAS*, 543, 891
- Quinn T., Katz N., Efstathiou G., 1996, *MNRAS*, 278, L49
- Read J. I., Erkal D., 2019, *MNRAS*, 487, 5799
- Read J. I., Iorio G., Agertz O., Fraternali F., 2017, *MNRAS*, 467, 2019
- Rey M. P., Pontzen A., Agertz O., Orkney M. D. A., Read J. I., Saintonge A., Pedersen C., 2019, *ApJ*, 886, L3
- Rey M. P., Pontzen A., Agertz O., Orkney M. D. A., Read J. I., Rosdahl J., 2020, *MNRAS*, 497, 1508
- Richings A. J., Schaye J., Oppenheimer B. D., 2014a, *MNRAS*, 440, 3349
- Richings A. J., Schaye J., Oppenheimer B. D., 2014b, *MNRAS*, 442, 2780
- Richings J., Frenk C., Jenkins A., Robertson A., Schaller M., 2021, *MNRAS*, 501, 4657
- Ricotti M., Gnedin N. Y., 2005, *ApJ*, 629, 259
- Sales L. V., Wetzel A., Fattahi A., 2022, *Nat. Astron.*, 6, 897
- Sawala T., Frenk C. S., Crain R. A., Jenkins A., Schaye J., Theuns T., Zavala J., 2013, *MNRAS*, 431, 1366
- Sawala T. et al., 2016, *MNRAS*, 457, 1931
- Schaller M. et al., 2015, *MNRAS*, 451, 1247
- Schaller M. et al., 2024, *MNRAS*, 530, 2378
- Schaye J. et al., 2010, *MNRAS*, 402, 1536
- Schaye J. et al., 2015, *MNRAS*, 446, 521
- Schaye J. et al., 2023, *MNRAS*, 526, 4978
- Schaye J. et al., 2026, *MNRAS*, 548, stag375
- Schmidt M., 1959, *ApJ*, 129, 243
- Sedgwick T. M., Baldry I. K., James P. A., Kelvin L. S., 2019, *MNRAS*, 484, 5278
- Shapiro P. R., Iliev I. T., Raga A. C., 2004, *MNRAS*, 348, 753
- Sheth R. K., Tormen G., 2002, *MNRAS*, 329, 61
- Simon J. D., 2019, *ARA&A*, 57, 375
- Simpson C. M., Grand R. J. J., Gómez F. A., Marinacci F., Pakmor R., Springel V., Campbell D. J. R., Frenk C. S., 2018, *MNRAS*, 478, 548
- Springel V., Di Matteo T., Hernquist L., 2005a, *MNRAS*, 361, 776
- Springel V. et al., 2005b, *Nature*, 435, 629
- Stanek R., Rudd D., Evrard A. E., 2009, *MNRAS*, 394, L11
- Stanway E. R., Eldridge J. J., 2018, *MNRAS*, 479, 75
- Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, *ApJ*, 688, 709
- Torrealba G. et al., 2019, *MNRAS*, 488, 2743
- Trayford J. W. et al., 2026, *MNRAS*, 545, staf2040
- Tremmel M., Karcher M., Governato F., Volonteri M., Quinn T. R., Pontzen A., Anderson L., Bellovary J., 2017, *MNRAS*, 470, 1121
- Trujillo I., Chamba N., Knapen J. H., 2020, *MNRAS*, 493, 87
- van Daalen M. P., Schaye J., Booth C. M., Dalla Vecchia C., 2011, *MNRAS*, 415, 3649
- Velliscig M., van Daalen M. P., Schaye J., McCarthy I. G., Cacciato M., Le Brun A. M. C., Dalla Vecchia C., 2014, *MNRAS*, 442, 2641
- Vogelsberger M. et al., 2014, *MNRAS*, 444, 1518
- Wang K. et al., 2026, *MNRAS*, 546, stag110
- Wechsler R. H., Bullock J. S., Primack J. R., Kravtsov A. V., Dekel A., 2002, *ApJ*, 568, 52
- Wechsler R. H., Zentner A. R., Bullock J. S., Kravtsov A. V., Allgood B., 2006, *ApJ*, 652, 71
- Weisz D. R. et al., 2011, *ApJ*, 739, 5
- Weisz D. R., Dolphin A. E., Skillman E. D., Holtzman J., Gilbert K. M., Dalcanton J. J., Williams B. F., 2014, *ApJ*, 789, 148
- Wetzel A. R., Hopkins P. F., Kim J.-h., Faucher-Giguère C.-A., Kereš D., Quataert E., 2016, *ApJ*, 827, L23
- Wheeler C., Oñorbe J., Bullock J. S., Boylan-Kolchin M., Elbert O. D., Garrison-Kimmel S., Hopkins P. F., Kereš D., 2015, *MNRAS*, 453, 1305
- Wheeler C. et al., 2019, *MNRAS*, 490, 4447
- Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009, *MNRAS*, 399, 574
- Wolf J., Martinez G. D., Bullock J. S., Kaplinghat M., Geha M., Muñoz R., Simon J. D., Avedo F. F., 2010, *MNRAS*, 406, 1220
- Wu J. F., Jespersen C. K., Wechsler R. H., 2024, *ApJ*, 976, 37
- Zehavi I. et al., 2011, *ApJ*, 736, 59
- Zehavi I., Contreras S., Padilla N., Smith N. J., Baugh C. M., Norberg P., 2018, *ApJ*, 853, 84
- Zhao D. H., Jing Y. P., Mo H. J., Börner G., 2009, *ApJ*, 707, 354

APPENDIX A: PROPERTIES OF THE REGIONS TARGETED FOR RESIMULATION

Key present-day properties of the 25 regions targeted for resimulation, drawn from the parent $L = 400$ cMpc cube, are listed

Table A1. Key properties of the regions within the $L = 400$ cMpc parent volume that were targeted for resimulation. The columns are as described in the text of Appendix A. A visual impression can be obtained by inspection of Fig. 3, for which each row corresponds to a band of δ_5 , and each column to a band of N_{LMC} .

Region index	δ_5 band	N_{LMC} band	$\log_{10}\delta_5$	N_{LMC}	x_c (Mpc)	y_c (Mpc)	z_c (Mpc)
1	-2σ	-2σ	-1.044	0	22.554	61.397	182.388
2	-2σ	-1σ	-1.090	0	285.056	380.053	5.246
3	-2σ	0σ	-1.075	0	170.254	49.515	378.008
4	-2σ	$+1\sigma$	-1.084	0	252.884	61.142	226.887
5	-2σ	$+2\sigma$	-1.053	0	303.710	387.559	259.295
6	-1σ	-2σ	-0.851	0	247.260	223.085	118.065
7	-1σ	-1σ	-0.829	0	360.984	370.986	104.061
8	-1σ	0σ	-0.854	0	113.541	105.269	295.703
9	-1σ	$+1\sigma$	-0.848	1	202.167	22.568	326.103
10	-1σ	$+2\sigma$	-0.859	1	327.347	244.685	199.035
11	0σ	-2σ	-0.638	0	206.470	289.345	337.098
12	0σ	-1σ	-0.635	0	164.727	370.529	389.946
13	0σ	0σ	-0.651	1	60.793	270.998	111.084
14	0σ	$+1\sigma$	-0.636	2	191.408	185.739	41.606
15	0σ	$+2\sigma$	-0.641	4	382.208	57.656	160.808
16	$+1\sigma$	-2σ	-0.436	1	211.498	78.734	381.412
17	$+1\sigma$	-1σ	-0.432	2	77.794	100.334	185.028
18	$+1\sigma$	0σ	-0.445	4	236.746	148.260	82.034
19	$+1\sigma$	$+1\sigma$	-0.444	6	124.015	262.135	382.086
20	$+1\sigma$	$+2\sigma$	-0.442	8	356.896	181.048	185.781
21	$+2\sigma$	-2σ	-0.279	4	134.931	23.480	373.677
22	$+2\sigma$	-1σ	-0.282	6	96.567	46.163	314.870
23	$+2\sigma$	0σ	-0.298	8	101.094	182.843	370.797
24	$+2\sigma$	$+1\sigma$	-0.290	11	211.361	265.825	313.522
25	$+2\sigma$	$+2\sigma$	-0.292	13	51.271	95.723	98.979

in Table A1. The columns of the table give the region index, the ‘bands’ corresponding to the percentile rank ordering of a Gaussian distribution for, respectively, the enclosed normalized density of the $r = 5$ cMpc sphere (δ_5) and the number of LMC-mass haloes within the region (N_{LMC}), the actual values of δ_5 and N_{LMC} , and the Cartesian coordinates of the region centre, (x_c, y_c, z_c), in the coordinate system of the parent volume.

APPENDIX B: NUMERICAL CONVERGENCE TESTS

In this section, we present simple numerical convergence tests of the HMF, the halo concentration–mass relation, and the luminous fraction as a function of halo mass, recovered from the composite halo populations of all 25 simulations at m5 (maroon curves) and m6 (orange curves) resolution.

The convergence of the HMF is shown in Fig. B1. As noted in Section 4.1, the HMFs of the regions are systematically shifted to lower number density at fixed mass relative to the cosmic mean HMF, owing to their relatively low enclosed density. To enable comparison with the cosmic mean HMF of J. Tinker et al. (2008), denoted by the grey curve, the low opacity curves show the HMF recovered when normalizing the number density of haloes from each region by the reciprocal of the region’s enclosed density (δ_5) prior to constructing the composite HMF. This exercise highlights that the renormalized HMFs reproduce the cosmic mean HMF accurately in the mass regime that is well sampled by the low density regions simulated here ($M_{200} \lesssim 10^9 M_\odot$). The figure clearly highlights the excellent convergence of the HMF between the two adopted resolutions: in the mass regime for which haloes are

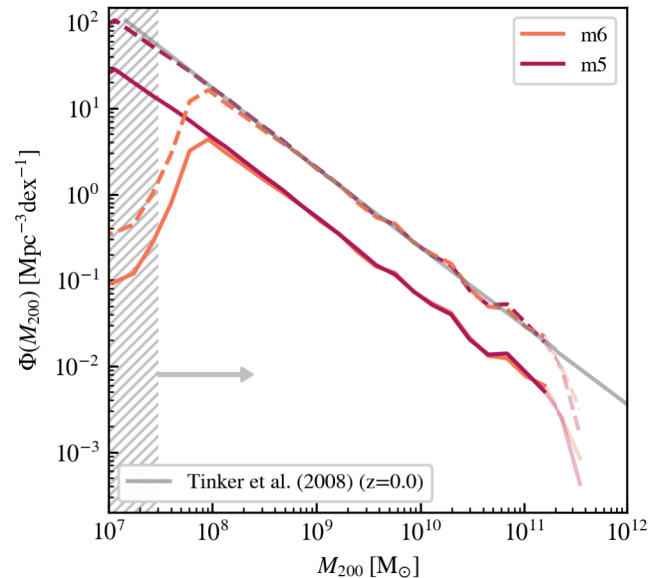


Figure B1. A convergence test of the hydrodynamical HMF with the change in resolution between m6, in orange, and m5, in maroon. The specific resolution details are given in Table 1. Each line combines all 25 regions. At the high-mass end, the lines have lower opacity when the bin contains fewer than 10 haloes. For reference, the grey solid line corresponds to the halo mass function from J. Tinker et al. (2008). The shaded region indicates 100 m5 resolution DM particles and extends to 100 m6 resolution DM particles indicated by the arrow. When the low mean density of our regions are normalized out, the HMFs shift upwards (dashed lines) to match that of J. Tinker et al. (2008) very well.

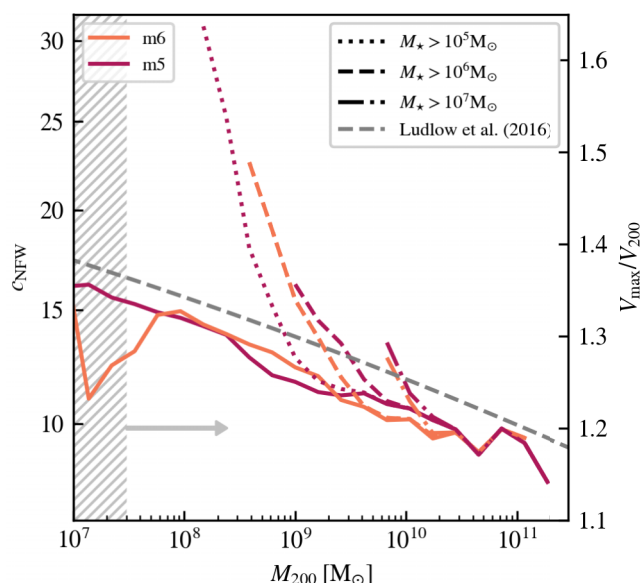


Figure B2. A strong convergence test of the halo mass–concentration relation in the hydrodynamical runs between the m5 resolution, in maroon, and m6, in orange. The NFW concentration parameter is estimated from the ratio of $V_{\max}/V_{200}$ from the matched DMO haloes using equation (2). At the high-mass end, the lines have lower opacity when the bin contains fewer than 10 haloes. Colours and hatching follow the conventions in Fig. B1. We show corresponding relations for haloes that host a ‘luminous’ component, whilst changing the stellar mass threshold for defining luminous: $M_* > 10^5$, 10^6 , and $10^7 M_\odot$ shown as dotted, dashed, and dot-dashed lines, respectively. For the lowest stellar mass threshold, only m5 results are shown since galaxies with $M_* < 10^6 M_\odot$ cannot form in m6 simulations due to its particle mass resolution. The mass–concentration relation from A. D. Ludlow et al. (2016) is shown as a grey dashed line, for reference.

sampled by at least 100 particles at m6 resolution and bins are sampled by at least 10 haloes, the HMFs differ by a maximum of 0.07 dex at fixed halo mass.

Fig. B2 shows the convergence of the halo mass–concentration relation. The relation recovered from the overall halo population (denoted by solid curves) is generally well converged between resolutions, in the mass regime where haloes are sampled by more than 100 particles in both resolutions, differing up to 0.03 dex. The figure also shows the relation recovered from ‘luminous’ haloes, using different threshold stellar masses to define a luminous halo. These are $M_* > 10^5 M_\odot$ (dotted curves), $M_* > 10^6 M_\odot$ (dashed curves), and $M_* > 10^7 M_\odot$ (dot-dashed curves). Using a threshold of $M_* > 10^7 M_\odot$, which corresponds to galaxies sampled by at least 10 and 100 star particles at m6 and m5 resolution, respectively, yields a halo mass–concentration relation that differs by no more than 0.05 dex at fixed halo mass. Use of a lower threshold unsurprisingly yields poorer convergence, owing to inadequate particle sampling of the galaxies.

Fig. B3 shows the convergence of luminous fraction as a function of halo mass, at different resolutions (m5 and m6) and different stellar mass thresholds. The dotted lines correspond to a ‘luminous’ threshold of $M_* > 10^5 M_\odot$, the dashed lines represent $M_* > 10^6 M_\odot$, and the dash-dotted lines represent $M_* > 10^7 M_\odot$.

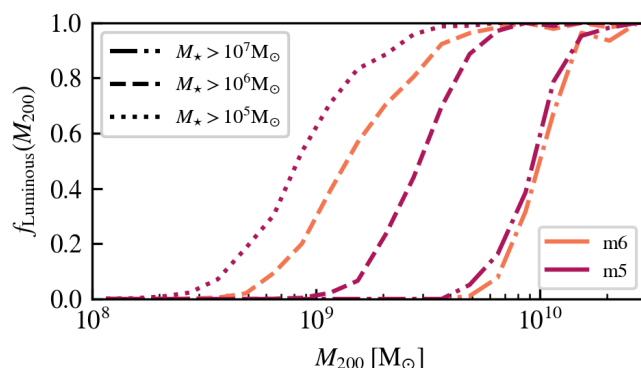


Figure B3. The present-day luminous fraction as a function of halo mass, shown at m5 and m6 resolution in maroon and orange, respectively. We explore the effect of changing the stellar mass threshold defining a ‘luminous’ galaxy: we show $M_* > 10^5$, 10^6 , and $10^7 M_\odot$ with dashed, solid, and dash-dotted lines.

As previously discussed, a threshold of $M_* > 10^6 M_\odot$ is effectively an upper limit for the halo occupation in m6, and similarly $M_* > 10^5 M_\odot$ for m5, formed from one stellar particle. The positions of these curves depend on the choice of stellar mass threshold and reach 100 per cent occupancy at mass scales corresponding to the mass below which the luminous mass–concentration relations diverge from the overall median shown in Fig. B2.

For a threshold of $M_* > 10^7 M_\odot$, the luminous fraction is well converged between resolutions, differing by less than 11 per cent at fixed halo mass. The convergence for a threshold of $M_* > 10^6 M_\odot$ is much poorer, with m6 simulations exhibiting higher luminous fractions differing up to 50 per cent at fixed halo mass. We see a similar poor convergence of the GSMF between m5 and m6 at $M_* < 10^7 M_\odot$ (see, Fig. 8). In this regime, galaxies are sampled by fewer than 10 stellar particles in m6, so that the poor convergence is not surprising.

APPENDIX C: STAR FORMATION HISTORIES AS A FUNCTION OF STELLAR MASS

As discussed in Section 5.2, there is an anticorrelation between the scatter about the median stellar mass at fixed halo mass, and duration of star formation quantified by t_{90} . This may seem at odds with the expectation that more massive dwarf galaxies exhibit more prolonged SFHs. However, this trend applies only when considering galaxies hosted by haloes of fixed mass. Fig. C1 shows t_{90} as a function of stellar mass for all individual simulated galaxies regardless of their halo mass, with the median trend denoted with a solid curve. The overall trend between stellar mass and the duration of the SFH is positive. Each galaxy in the figure is colour-coded by its halo concentration. Consistent with the findings in Section 5.2 the scatter about the median correlates (weakly) with halo concentration: at a fixed stellar mass, galaxies with a delayed SFH and hence a greater value of t_{90} are typically hosted by low concentration, and hence later forming, haloes, illustrating how halo concentration drives the scatter in t_{90} at fixed halo mass.

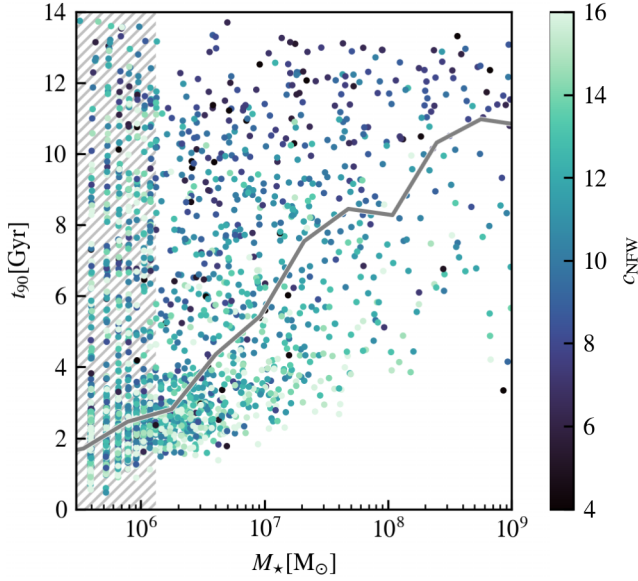


Figure C1. The time it takes for a galaxy to assemble 90 per cent of its total stellar mass, t_{90} , as a function of its present-day stellar mass, M_* . Each galaxy is coloured by its halo concentration, measured from its DM-only counterpart. The median relation is shown by the grey line.

APPENDIX D: STELLAR MASS–HALO MASS RELATION FIT

Fig. D1 shows the present-day SMHM relation of the composite population of field galaxies formed from all 25 simulated regions at m5 resolution. The median SMHM relation can be represented by a power law of the form $M_*/M_\odot = \beta(M_{200}/10^{10}M_\odot)^\alpha$ where $(\alpha, \beta) = (2.17, 1.42 \times 10^7)$, fitted to the range $M_* = [10^6 - 10^9]M_\odot$, as is shown by the grey line in Fig. D1. The semi-empirical models of P. S. Behroozi et al. (2013) and B. P. Moster et al. (2013) are shown in indigo for comparison.

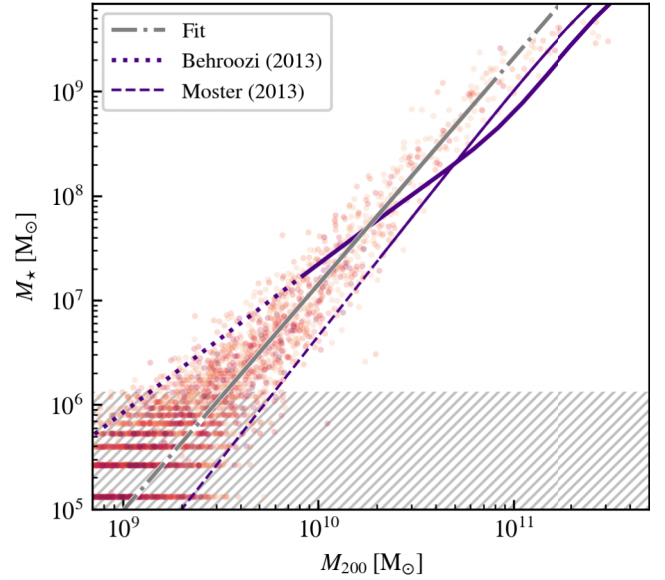


Figure D1. The present-day stellar mass–halo mass scaling relation of central galaxies for the composite population of all 25 regions, at m5 resolution. The grey hatched region indicates the stellar mass regime corresponding to 10 or fewer stellar particles at m5 resolution. The median relation is well fit by a power law of form $M_*/M_\odot = \beta(M_{200}/10^{10}M_\odot)^\alpha$ where $(\alpha, \beta) = (2.17, 1.42 \times 10^7)$, fitted to the range $M_* = [10^6 - 10^9]M_\odot$ and shown by the grey line. Indigo curves show the SMHM relations of the semi-empirical models of P. S. Behroozi et al. (2013) and B. P. Moster et al. (2013), which transition to solid lines in the mass range to which they are fit. All lines are solid for the mass scale they are fit to.

This paper has been typeset from a $\text{\LaTeX}$ file prepared by the author.