Revealing the galaxy–halo connection in IllustrisTNG

Sownak Bose [●],[1][★] Daniel J. Eisenstein,[1] Lars Hernquist,[1] Annalisa Pillepich,[2]
Dylan Nelson [●],[3] Federico Marinacci [●],[4] Volker Springel[3] and Mark Vogelsberger [●][5]

[1]*Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA*
[2]*Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany*
[3]*Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, D-85741 Garching, Germany*
[4]*Department of Physics & Astronomy, University of Bologna, via Gobetti 93/2, I-40129 Bologna, Italy*
[5]*Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

ABSTRACT

We use the IllustrisTNG (TNG) simulations to explore the galaxy–halo connection as inferred from state-of-the-art cosmological, magnetohydrodynamical simulations. With the high-mass resolution and large volume achieved by combining the 100 Mpc (TNG100) and 300 Mpc (TNG300) volumes, we establish the mean occupancy of central and satellite galaxies and their dependence on the properties of the dark matter haloes hosting them. We derive best-fitting HOD parameters from TNG100 and TNG300 for target galaxy number densities of $\bar{n}_g = 0.032$ and $\bar{n}_g = 0.016\,h^3\,\mathrm{Mpc}^{-3}$, respectively, corresponding to a minimum galaxy stellar mass of $M_\star \sim 1.9 \times 10^9$ and $M_\star \sim 3.5 \times 10^9\,\mathrm{M}_\odot$, respectively, in hosts more massive than $10^{11}\,\mathrm{M}_\odot$. Consistent with previous work, we find that haloes located in dense environments, with low concentrations, later formation times, and high angular momenta are richest in their satellite population. At low mass, highly concentrated haloes and those located in overdense regions are more likely to contain a central galaxy. The degree of environmental dependence is sensitive to the definition adopted for the physical boundary of the host halo. We examine the extent to which correlations between galaxy occupancy and halo properties are independent and demonstrate that HODs predicted by halo mass and present-day concentration capture the qualitative dependence on the remaining halo properties. At fixed halo mass, concentration is a strong predictor of the stellar mass of the central galaxy, which may play a defining role in the fate of the satellite population. The radial distribution of satellite galaxies, which exhibits a universal form across a wide range of host halo mass, is described accurately by the best-fitting NFW density profile of their host haloes.

Key words: methods: numerical – galaxies: haloes – cosmology: theory – large-scale structure of Universe.

1 INTRODUCTION

In our current theory of structure formation, gravitational instability transforms the initially linear, near homogeneous Universe into the complex network of non-linear structures we observe around us today. The process by which tiny fluctuations in the primordial density field, inferred by microwave background experiments (e.g. Spergel et al. 2003; Komatsu et al. 2011; Planck Collaboration VIII 2016), are amplified into the cosmic web of filaments, voids, walls, and haloes has been captured with impressive accuracy and detail by a rigorous programme of numerical simulations over the last four decades (e.g. Davis et al. 1985; Efstathiou et al. 1985; Springel et al.

2005; Angulo et al. 2012; Prada et al. 2012; Habib et al. 2016; Potter, Stadel & Teyssier 2017). In particular, the statistical agreement between the clustering of galaxies measured in redshift surveys and that predicted by numerical simulations has given credence to the cold dark matter model as the standard paradigm (e.g. Colless et al. 2001; Cole et al. 2005; Eisenstein et al. 2005; Zehavi et al. 2011).

At first order, the clustering of haloes does not trace the total matter distribution exactly, but is instead 'biased' relative to it in a manner that correlates with the mass of the halo: clustering bias increases with increasing halo mass (e.g. Kaiser 1984; Efstathiou et al. 1988; Cole & Kaiser 1989; Bond et al. 1991; Mo & White 1996; Sheth & Tormen 1999). As numerical simulations have allowed us to probe deeper into the non-linear regime, it has become feasible to quantify other properties of haloes – in addition to their mass – that may also determine this bias. The most commonly

explored properties in this context are the large-scale environment in which the halo is embedded, its formation time, concentration, and spin; the dependence of clustering on these secondary quantities is broadly categorized under the umbrella term of 'assembly bias' (e.g. Gao, Springel & White 2005; Harker et al. 2006; Wechsler et al. 2006; Bett et al. 2007; Gao & White 2007; Jing, Suto & Mo 2007; Zentner 2007; Dalal et al. 2008; Mao, Zentner & Wechsler 2018, see Desjacques, Jeong & Schmidt 2018 for a recent review).

Clustering bias is not merely a theoretical concept, but is also imprinted in the observed distribution of galaxies. Correlations between galaxy properties like colour, luminosity, star formation history, and morphology of galaxies and their large-scale environment has long been known from redshift surveys (e.g. Davis & Geller 1976; Dressler 1980; Blanton et al. 2005). Zehavi et al. (2005, 2011) demonstrated a strong luminosity and colour dependence of galaxy clustering in Sloan Digital Sky Survey, while an analogous correlation with age and star formation rate has been established recently by e.g. Wang et al. (2013) and Hearin & Watson (2013). The combined use of photometric galaxy catalogues and weak gravitational lensing by rich clusters has indicated the existence of assembly bias as a function of concentration and assembly history (e.g. Miyatake et al. 2016; More et al. 2016; Montero-Dorta et al. 2017; Niemiec et al. 2018), although it has been suggested that foreground contamination from interloping galaxies introduces systematics that manifest falsely as assembly bias (Busch & White 2017).

On scales smaller than ∼10 Mpc or so, comparison between theory and observation is muddied by the complex interplay between dark matter and baryons. Hydrodynamical simulations have shown that baryonic processes alter the clustering of matter inferred from purely collisionless simulations at the level of tens of per cent and in a scale-dependent fashion (e.g. van Daalen et al. 2014; van Daalen & Schaye 2015; Hellwing et al. 2016; Chisari et al. 2018; Springel et al. 2018). As galaxies are themselves biased tracers of the underlying density field, a faithful assessment of the impact of assembly bias on galactic populations requires a comprehensive model to associate them with their host haloes (see Wechsler & Tinker 2018, for a detailed review of this subject).

A common prescription of this kind, broadly referred to as semi-analytic modelling, follows merger trees to grow galaxies within dark matter haloes using systems of coupled differential equations describing gas cooling, star formation, and feedback (e.g. Kauffmann, White & Guiderdoni 1993; Somerville & Primack 1999; Cole et al. 2000; Croton et al. 2006; Benson 2012; Henriques et al. 2015). An alternative approach involves the statistical assignment of galaxies to haloes, the most prominent of which includes abundance matching (e.g. Mo, Mao & White 1999; Kravtsov et al. 2004; Tasitsiomi et al. 2004; Vale & Ostriker 2004; Behroozi, Conroy & Wechsler 2010; Reddick et al. 2013), empirical models (e.g. Conroy & Wechsler 2009; Behroozi, Wechsler & Conroy 2013; Moster, Naab & White 2013; Rodríguez-Puebla et al. 2016; Behroozi et al. 2019; Tacchella et al. 2018) and Halo Occupation Distributions (HODs; Benson et al. 2000; Peacock & Smith 2000; Berlind & Weinberg 2002; Zheng et al. 2005). Finally, hydrodynamical codes that self-consistently track the evolution of baryons and dark matter allow us to infer the association between galaxies and haloes on sub-Mpc scales; the sophistication and accuracy of these models have improved greatly over the last decade (e.g. Governato et al. 2009; Guedes et al. 2011; Dubois et al. 2014; Hopkins et al. 2014; Vogelsberger et al. 2014b; Schaye et al. 2015; Davé, Thompson & Hopkins 2016; Pillepich et al. 2018a)

Despite the impressive improvement in the performance of hydro-dynamical codes, however, their application to the domain of large-scale structure is limited by the computational cost of simulating such large volumes. Impending surveys like Euclid (Laureijs et al. 2011) and the Dark Energy Spectroscopic Instrument (DESI; Levi et al. 2013) demand the generation of several mock simulations at Gpc scale for an accurate assessment of covariance matrices, systematics induced by cosmological parameters etc. (Percival et al. 2014; O'Connell et al. 2016; Klypin & Prada 2018). The spatial and temporal resolution required by hydrodynamical simulations make this enterprise currently unfeasible. On the other hand, the HOD formalism provides a convenient (and, computationally cheap) framework to 'paint' a mock galaxy population on top of large volume collisionless simulations. Cast in its original form, the HOD formalism is agnostic about any property, apart from mass, when determining the average occupancy of galaxies in haloes; assembly bias is, by construction, excluded. This first-order approximation works surprisingly well: indeed, the clustering of galaxies selected by their luminosity or colour is well reproduced by HODs defined by halo mass only (Tinker et al. 2008; Tinker & Conroy 2009).

In recent times, the development of 'decorated' HODs (e.g. Hearin et al. 2016; Yuan, Eisenstein & Garrison 2018) has augmented the standard HOD paradigm to incorporate scatter that may be introduced by assembly bias. Besides assembly bias, it has also been suggested that the *shape* of the HOD may be sensitive to physical processes associated with galaxy formation operating in host haloes, such as the strength of feedback from active galactic nuclei (AGN; McCullagh et al. 2017). Apart from the demographics of galaxies, assembly bias may also play an important role in predicting differential quantities associated with galaxy formation, such as colour or star formation rate, over scales that extend well beyond the virial radius of host haloes (e.g. Weinmann et al. 2006; Hearin & Watson 2013; Lacerna et al. 2014; Hearin, Watson & van den Bosch 2015; Kauffmann 2015; Bray et al. 2016)

Our goal in this paper is to exploit modern cosmological, hydrodynamical simulations to assess how assembly bias enters the HOD inferred from these simulations. In particular, we use the IllustrisTNG suite of simulations[1] to match haloes in a fully hydrodynamical simulation to their counterparts in a collisionless setup and investigate how the average HOD responds to properties like environment, concentration, formation time, and spin. This exercise is similar in spirit to the recent explorations of Zehavi et al. (2018) and Artale et al. (2018) and complements them with an evaluation of the relative importance of the individual biases. At fixed halo mass, the abundance of subhaloes in dark matter-only simulations are known to vary with formation time and concentration (e.g. Gao et al. 2004).

The layout of this paper is as follows. In Section 2, we present our numerical setup, describing the simulations and the procedures used to explore the HOD in IllustrisTNG. Section 3 presents our main findings. Finally, our conclusions are summarized in Section 4.

2 NUMERICAL PRELIMINARIES

In this section, we briefly introduce the IllustrisTNG magneto-hydrodynamical simulations (Section 2.1), which are used as the theoretical framework for exploring the HOD. We then describe the methodology used to associate galaxies and their host haloes to their counterparts in simulations containing only dark matter

---

[1] http://www.tng-project.org/

(Section 2.2) and how the HODs are then constructed (Section 2.3). Finally, we describe and quantify the list of halo properties used to assess the scatter in the HOD (Section 2.4).

2.1 Simulations

The Next Generation Illustris (IllustrisTNG, hereafter simply TNG) simulations (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018; Springel et al. 2018; Pillepich et al. 2018b) analysed in this paper constitute an ambitious programme of cosmological, hydrodynamical simulations of galaxy formation. The simulations have been performed using the AREPO simulation code (Springel 2010), which employs a hybrid tree/particle-mesh scheme to solve for gravitational interactions of dark matter particles, and a moving, unstructured Voronoi mesh to solve equations of hydrodynamics. A key feature of AREPO is that the moving mesh is adaptive in nature, resolving fluids in regions of high density with many more cells of a smaller size than in low-density environments.

The TNG simulations are specified by a comprehensive galaxy formation model that has been incorporated into AREPO. The details of the physics model are described in detail in Weinberger et al. (2017) and Pillepich et al. (2018a) and, for the sake of brevity, are not repeated here. The TNG model is the direct successor to the original ILLUSTRIS model (Vogelsberger et al. 2013, 2014a,b) with several updates made to the implementation of AGN feedback (Weinberger et al. 2017), the operation of galactic winds (Pillepich et al. 2018a), and the incorporation of magnetic fields (Pakmor, Bauer & Springel 2011; Pakmor & Springel 2013; Pakmor, Marinacci & Springel 2014). This model has been shown to reproduce a wide range of properties of observed galactic populations across cosmic time (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018; Pillepich et al. 2018b; Springel et al. 2018).

In this work, we make use of the TNG100 and TNG300 simulation volumes; both data sets have been made publicly available[2] (Nelson et al. 2019). The TNG100 run consists of a periodic box of length $L_{\rm box} = 75\,h^{-1}$ Mpc ≈ 100 Mpc, containing 2×1820^3 dark matter particles/gas cells, corresponding to an effective mass resolution of $9.44 \times 10^5\,h^{-1}\,{\rm M}_\odot$ in baryons and $5.06 \times 10^6\,h^{-1}\,{\rm M}_\odot$ in dark matter. The maximum physical softening length of dark matter and star particles is set to $0.5\,h^{-1}$ kpc. TNG300 simulates a larger cosmological box of size $L_{\rm box} = 205\,h^{-1}$ Mpc ≈ 300 Mpc with 2×2500^3 resolution elements. The corresponding mass resolution is $7.44 \times 10^6\,h^{-1}\,{\rm M}_\odot$ in baryonic matter and $3.98 \times 10^7\,h^{-1}\,{\rm M}_\odot$ in dark matter. The maximum physical softening length of dark matter and star particles is set to $1.0\,h^{-1}$ kpc.

Initial conditions for both sets of simulations have been generated at $z = 127$ assuming cosmological parameters inferred by *Planck* (Planck Collaboration VIII 2016): $\Omega_0 = 0.3089$ (total matter density), $\Omega_{\rm b} = 0.0486$ (baryon density), $\Omega_\Lambda = 0.6911$ (dark energy density), $H_0 = 67.74$ kms^{-1} Mpc^{-1} (Hubble parameter), and $\sigma_8 = 0.8159$ (linear rms density fluctuation in a sphere of radius $8\,h^{-1}$ Mpc at $z = 0$). Each of the TNG100 and TNG300 simulations have counterparts generated from the same initial phases, but evolved with dark matter only (DMO); these simulations are labelled TNG100-DMO and TNG300-DMO, respectively, and provide the properties of dark matter haloes used to vary the HOD inferred from the runs with full physics (see Section 2.4).

2.2 Identifying and matching haloes

Haloes are identified from the particle distribution first using a 'friend-of-friends' (FOF) algorithm, which connects together dark matter particles separated by at most 20 per cent the mean interparticle separation to form groups (Davis et al. 1985). Particles within each FOF group that are truly gravitationally bound are then identified using the SUBFIND algorithm (Springel et al. 2001), with the requirement that each 'sub'halo contains at least 20 resolution elements, regardless of type. This splits an FOF group into a 'main' halo and its associated subhaloes; a galaxy is then defined by the constituent stars, gas, black holes, and dark matter of a (sub)halo, either central or satellite. Unless otherwise specified, a galaxy is assigned as a member of an FOF group only if it is located within r_{200} of the main halo centre, which we determine in post-processing. Here, r_{200} is defined as the radius within which the mean density of the halo is equal to 200 times the critical density of the universe at the given redshift. In what follows, the mass of a dark matter halo is quoted in terms of M_{200} (i.e. the total mass contained within r_{200}); the stellar mass of a galaxy is the mass contained within a 30 kpc aperture centred on the galaxy.

We establish matches between haloes in the TNG runs with DMO and those with full physics using the procedure outlined in e.g. Lovell et al. (2018) and Bose et al. (2019). First, we consider the 50 most-bound dark matter particles from a candidate halo in the hydrodynamical run, and search for the DMO halo in which there are at least 25 (50 per cent) of these particles. The match is then confirmed by repeating the same process, this time starting with the DMO haloes. More than 97 per cent of haloes more massive than $\sim 10^{11}\,{\rm M}_\odot$ are matched successfully using this bijective scheme. In terms of the HOD, the total number of galaxies assigned to a halo in the DMO simulation is then simply equal to the number of galaxies associated with its match in the hydrodynamical simulation.

2.3 Constructing the halo occupation distribution

To construct the HOD from each simulation, we first rank order galaxies from the hydrodynamical version of each simulation volume in order of decreasing stellar mass. We then select the first N galaxies out of this ranked catalogue, where $N = \bar{n}_g L_{\rm box}^3$ and $\bar{n}_g$ is a free parameter corresponding to the target number density of galaxies we wish to construct the HOD for. For each galaxy that enters the selection, we record its type (central or satellite) and the properties of the DMO halo which its hydrodynamical host is matched to.

To be confident of any conclusions we derive from the simulations, it is important that we limit the galaxy catalogues to entities that are well-resolved. Fig. 1 shows the (cumulative) number density of galaxies at $z = 0$ in TNG100 and TNG300 as a function of stellar mass. The vertical dotted lines mark the stellar mass limit corresponding to 100 particles at the resolution limit of each simulation volume. The horizontal dotted lines, corresponding to $\bar{n}_g = 0.032$ and $\bar{n}_g = 0.016\,h^3$ Mpc^{-3} are, respectively, the target number densities used to construct the HOD from TNG100 and TNG300. These number densities are chosen so as to facilitate comparison with studies by Zehavi et al. (2018) and Artale et al. (2018) who adopt similar values. Correspondingly, the minimum galaxy stellar mass imposed by the number density of each catalogue is $\log[M_\star^{\rm min}/{\rm M}_\odot] = 9.28$ in TNG100 and $\log[M_\star^{\rm min}/{\rm M}_\odot] = 9.55$ in TNG300; unless otherwise stated, galaxies less massive than these limits are not considered in the remainder of our investigation. The comparatively higher mass resolution of TNG100 affords us
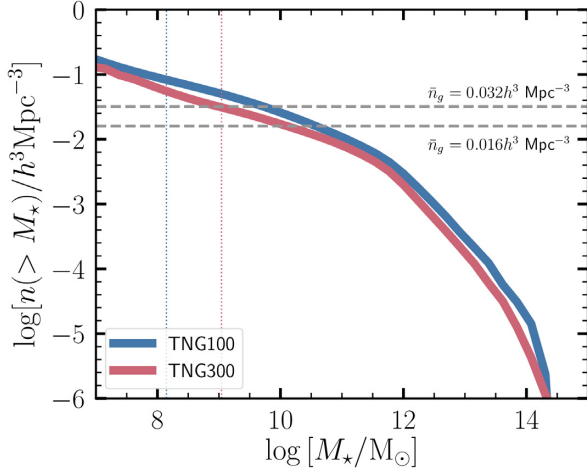
Figure 1. Cumulative number density of galaxies as a function of stellar mass in the TNG100 and TNG300 simulations. The vertical dotted lines mark the mass scale defined by 100 star particles in each simulation, which we choose somewhat arbitrarily as the minimum number of bound stellar elements needed to define a 'galaxy' in each simulation. The galaxy number density, $\bar{n}_g$, used to construct the HOD catalogue in TNG100 is $\bar{n}_g = 0.032\,h^3\,\mathrm{Mpc}^{-3}$, while for TNG300, we use $\bar{n}_g = 0.016\,h^3\,\mathrm{Mpc}^{-3}$; these are denoted by the horizontal dashed lines. In both instances, we are comfortably above the 100 star particle threshold.

the possibility to construct a galaxy sample with a higher number density than with TNG300; on the other hand, the larger volume of TNG300 enables us to better sample the HOD in the regime of rich clusters of galaxies ($\log\,[M_{200}/\mathrm{M}_\odot] \gtrsim 14.5$). Both choices of $\bar{n}_g$ result in catalogues resolved with many more than 100 star particles per galaxy; the results we present are therefore insensitive to this minimum threshold number. Note that the lack of overlap between the two curves presented in Fig. 1 is expected given the comparatively poorer resolution of TNG300 compared to TNG100. In particular, worse resolution results in somewhat lower stellar mass formed at fixed halo mass (see Pillepich et al. 2018b, for a detailed discussion). The conclusions we derive on the HOD are robust to these differences.

Fig. 2 presents the HODs constructed from TNG100 (top panel) and TNG300 (bottom panel) obtained after rank-ordering galaxies and associating them with their host DMO haloes as described in the previous subsection. The HODs are constructed at target number densities of $\bar{n}_g = 0.032$ and $\bar{n}_g = 0.016\,h^3\,\mathrm{Mpc}^{-3}$, respectively, for TNG100 and TNG300. We split the mean occupation per halo mass, $\langle N_{\mathrm{gals}} \rangle$, into contributions from central, $\langle N_{\mathrm{cen}} \rangle$, and satellite galaxies, $\langle N_{\mathrm{sat}} \rangle$, with the values measured from TNG represented by the symbols. On average, every halo in TNG100-DMO more massive than $\log\,[M_{200}^{\mathrm{DMO}}/\mathrm{M}_\odot] \sim 11.5$ hosts at least a central galaxy, while the propensity to contain at least one satellite galaxy appears at around an order of magnitude higher. At the chosen galaxy number density of $\bar{n}_g = 0.016\,h^3\,\mathrm{Mpc}^{-3}$, only one in 10 TNG300-DMO haloes, on average, contain a central galaxy. The mass scales of interest may be better quantified by fitting the simulation results with parametrized versions of the HOD model. The solid lines in Fig. 2 show fits to these populations assuming the 5-parameter HOD model described in Zheng et al. (2005), in which:

$$\langle N_{\mathrm{cen}}(M_h) \rangle = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\log M_h - \log M_{\mathrm{min}}}{\sigma_{\log M}}\right)\right], \tag{1}$$
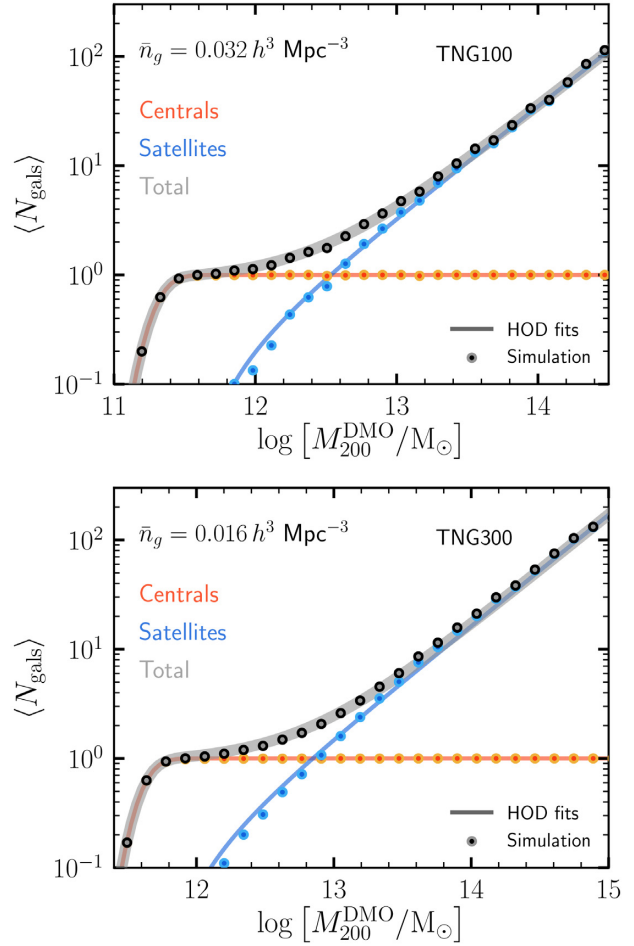

Figure 2. The mean HOD in the TNG100 (top) and TNG300 (bottom) simulations. Each HOD is constructed at fixed number density as indicated in the top left corner of each panel. The mean occupation of central galaxies is shown in orange, satellite galaxies in blue, and the total population in black. Symbols represent the mean occupancy measured in the TNG simulation volumes; the solid curves are fits assuming the Zheng et al. (2005) model.

and

$$\langle N_{\mathrm{sat}}(M_h) \rangle = \left(\frac{M_h - M_{\mathrm{cut}}}{M_1}\right)^{\alpha}. \tag{2}$$

Here, $M_h = M_{200}^{\mathrm{DMO}}$ is the halo mass, M_{min} is the characteristic minimum mass of haloes that host central galaxies, and $\sigma_{\log M}$ is the width of this transition. Furthermore, M_{cut} is the characteristic cut-off scale for hosting satellites, M_1 is a normalization factor and α is the power-law slope. Equations (1) and (2) capture the overall shape of the HOD from our simulations extremely well; the corresponding values for the 5 free parameters of this model ($M_{\mathrm{min}}, \sigma_{\log M}, M_{\mathrm{cut}}, M_1$, and α) are listed in Table 1. The middle row measures these parameters for TNG100 at the same number density as the one adopted for TNG300: while the values are quite similar, there are minor differences that may be attributed to the difference in numerical resolution between the two volumes. Throughout the rest of this paper, we will be concerned with the HOD measured at $z = 0$ only. The redshift evolution of HOD fitting parameters has been studied extensively in recent work by e.g. Contreras et al. (2017) and Smith et al. (2017).

Table 1. Best-fitting HOD parameter values for TNG100 and TNG300 at their respective target galaxy number densities (see main text for descriptions of these parameters). The third column, $\log M_\star^{\min}$, lists the minimum galaxy stellar mass implied by each choice of number density. The main analysis in this paper focuses on (a) and (c) only.

Simulation	Number density, $\bar{n}_g$ (h^3 Mpc^{-3})	$\log M_\star^{\min}$	$\log M_{\min}$	$\sigma_{\log M}$	$\log M_{\mathrm{cut}}$	$\log M_1$	α
(a) TNG100	0.032	9.28	11.292	0.165	11.589	12.483	1.017
(b) TNG100	0.016	9.84	11.587	0.157	11.851	12.795	1.029
(c) TNG300	0.016	9.55	11.601	0.161	11.778	12.809	1.018

2.4 Statistics of the dark halo population

The primary goal of the present investigation is to identify properties of the dark matter halo population (in addition to total mass) that are most informative of the mean (and scatter in) occupancy of galaxies within these haloes in a full hydrodynamical simulation. In other words: what properties of a halo in a DMO simulation determine if its hydrodynamical counterpart is more/less likely to host a central, or contains more/fewer satellite galaxies than the *average* halo at *that mass*.

In order to narrow down the search space for such parameters, it is useful to outline a set of criteria that these quantities should ideally satisfy. First, it is important to avoid spurious correlations in the HOD by considering only properties that have a plausible connection to galaxy occupancy. For example, a halo's epoch of formation is likely to be an informative parameter; its triaxiality, less so (at least directly). We further prioritize halo properties that are well-defined quantitatively, not very sensitive to resolution and ideally measurable at $z = 0$. These conditions are especially pertinent for lower resolution, large volume simulations of large-scale structure where HOD modelling is likely to be applied, but where, for example, the construction of high resolution merger trees is not possible. Note that throughout, we consider haloes more massive than $\log[M_{200}^{\mathrm{DMO}}/\mathrm{M}_\odot] \gtrsim 11.0$ in TNG100 and $\log[M_{200}^{\mathrm{DMO}}/\mathrm{M}_\odot] \gtrsim 11.5$ in TNG300, meaning that halo properties are only measured in FOF groups that are resolved with at least 5000 particles within r_{200}.

With these considerations in mind, we focus on the following set of halo properties:

(i) *Central density*: characterizing the central density of a halo provides a quantitative measure for the depth of the potential well that the central galaxy resides in. A popular way to characterize central density is through the concentration parameter, obtained after fitting an analytic profile to the density distribution, $\rho(r)$, in a dark matter halo. Here, we assume the NFW profile (Navarro, Frenk & White 1996, 1997)

$$\rho(r) = \frac{\rho_s}{\frac{r}{r_s} \left(1 + \frac{r}{r_s}\right)^2} , \qquad (3)$$

where r_s and ρ_s, respectively, are the scale radius and the density of the halo at this radius. The concentration is then given by $c_{\mathrm{NFW}} = r_{200}/r_s$. An unappealing aspect of fitting density profiles to individual haloes is that the procedure requires access to the full particle data set, not to mention the usual vagaries of profile fitting associated with the choice of radial range the analytic form is fit to, the number of particles in the halo, the minimization method used etc. We therefore consider an alternative proxy for the central density, the quantity $\widetilde{V}_{\max} \equiv V_{\max}/(H_0 \, r_{\max})$, where $V_{\max}$ is the peak of the circular velocity profile of the halo, $r_{\max}$ is the radius at which this value is attained, and H_0 makes this quantity dimensionless. As it is defined, $\widetilde{V}_{\max}$ has a one-to-one mapping

to the traditional concentration parameter, c_{NFW}. The benefit of this measure of central density is that the quantities $V_{\max}$ and $r_{\max}$ are typically output by all halo finding algorithms, and neither is obtained through fitting.

(ii) *Angular momentum*: the rotation of a halo is best characterized in terms of its (dimensionless) spin parameter, λ, given by

$$\lambda = \frac{J\sqrt{|E|}}{G M_{200}^{5/2}} , \qquad (4)$$

(Peebles 1969), where J is the magnitude of the angular momentum of the halo and E is its total energy. While the bulk of the angular momentum of a halo is obtained through tides in the linear density field (Doroshkevich 1970; White 1984), the net angular momentum of a halo may be perturbed subsequently through mergers (e.g. Vitvitska et al. 2002; Hetznecker & Burkert 2006). These mergers, in turn, may add to the satellite budget of the halo. We define the spin of a halo using all dark matter particles contained within its r_{200}. A halo's spin is often poorly measured, particularly in low-mass haloes (Bett et al. 2007; Trenti et al. 2010; Benson 2017); by considering only groups with at least 5000 particles within r_{200}, excessively noisy spin measurements are avoided. We have checked explicitly that no haloes with spuriously large values of λ have entered our analysis.

(iii) *Epoch of formation*: we classify haloes in the DMO simulation as early or late-forming using their formation redshift, z_{form}, defined as the epoch by which 50 per cent of the halo's present-day mass is assembled. In practice, this quantity is computed by following the merger tree of each halo along its main progenitor branch; we use merger trees constructed using the SUBLINK algorithm (Rodriguez-Gomez et al. 2015). It is well-established that the redshift of formation and concentration of a halo are correlated, with early-forming haloes exhibiting higher concentrations than their counterparts that collapse later.

(iv) *Environment*: the large-scale environment in which a halo is embedded may be defined in several ways, with origins rooted in percolation analysis (e.g. Zeldovich, Einasto & Shandarin 1982), graph theoretical interpretation of the galaxy distribution (e.g. Colberg 2007), and by computing tesselations or Hessians of the density field (e.g. Aragón-Calvo et al. 2007; Hahn et al. 2007; Sousbie et al. 2008; González & Padilla 2010; Cautun, van de Weygaert & Jones 2013). A full characterization of the cosmic web in the TNG simulations is beyond the scope of this paper; instead, we quantify a halo's large-scale environment using a computationally simpler approach. For each halo in a DMO run, we compute the quantity $\rho_r/\bar{\rho}$, where ρ_r is the mass density in dark matter subhaloes located within a sphere of radius r Mpc from the centre of the halo, while $\bar{\rho}$ is the mean mass density in subhaloes in the entire simulation volume. In computing ρ_r, it is important to exclude subhaloes associated with the halo itself (i.e. those within r_{200}). Furthermore, we only count subhaloes more massive than $10^9 \, \mathrm{M}_\odot$ in bound dark matter mass. As it is defined, a halo with $\rho_r/\bar{\rho} = 1$
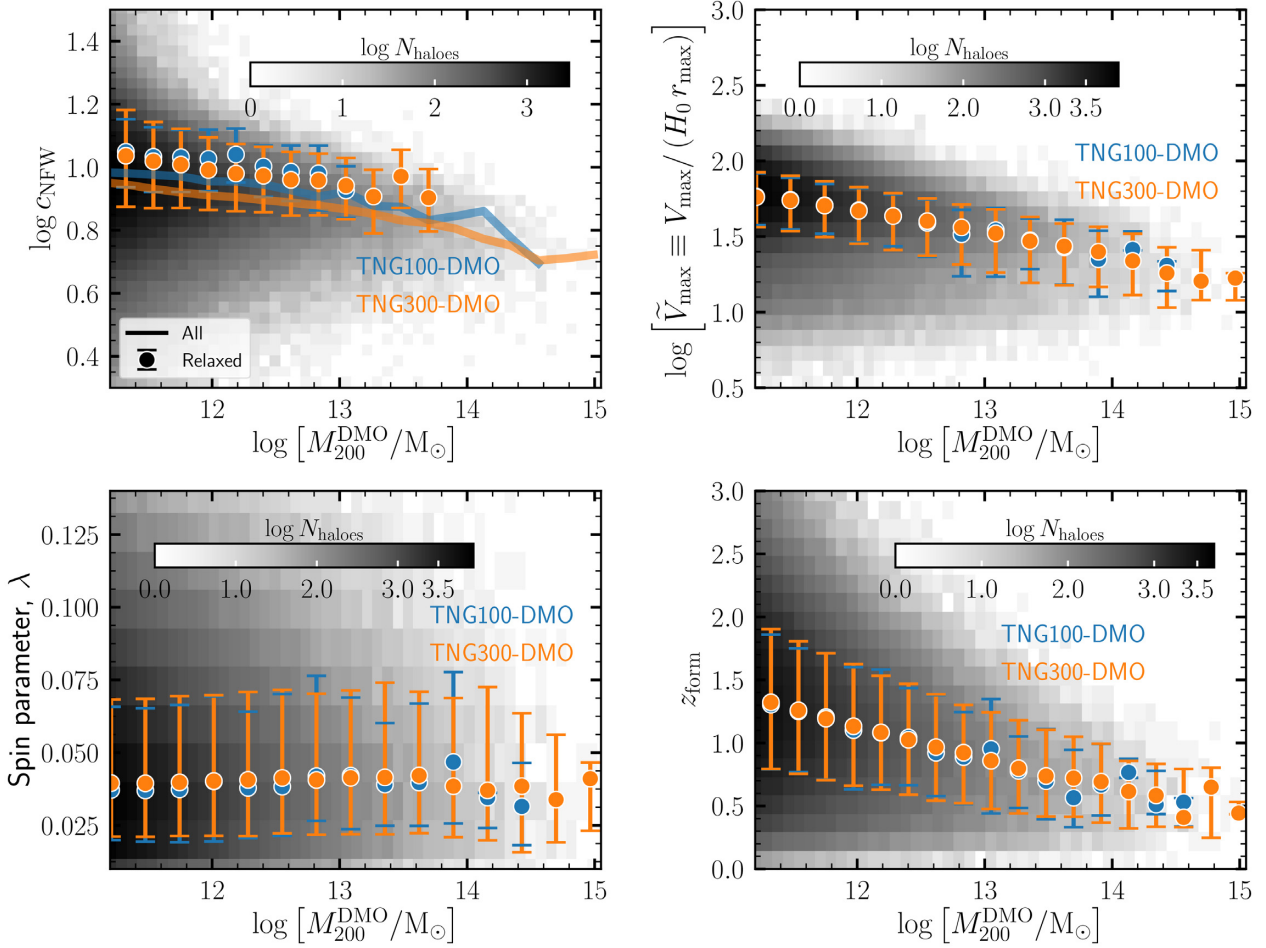
Figure 3. Properties of haloes extracted from the DMO versions of the TNG100 and TNG300 simulation volumes. In each panel, marker symbols show the median relation, while error bars encompass the 16th and 84th percentiles. All relations are shown for $z = 0$. The grey histogram in the background reveals the number of haloes contained in each mass bin for the TNG300 simulation. *Top left panel*: the mass–concentration relation, where the halo concentration, c_{NFW}, is determined by fitting NFW profiles to density profiles of individual dark matter haloes. *Top right panel*: Recasting the mass–concentration relation in terms of the quantity $\widetilde{V}_{max} \equiv V_{max}/(H_0 r_{max})$ (see main text). The behaviour of this relation – in terms of the trend and size of scatter as a function of M_{200}^{DMO} – is qualitatively similar to the traditional mass–concentration relation. *Bottom left panel*: correlation between the dimensionless halo spin, λ, and halo mass. *Bottom right panel*: the halo formation time, z_{form}, as a function of its mass, where the formation redshift is defined as the epoch by which 50 per cent of the halo's present-day mass has been assembled.

lives at roughly the mean density. The definition adopted here is a mass-weighted version of the one used by Artale et al. (2018); the advantage of a mass-weighted assignment is that it is capable of distinguishing between pairs of haloes that share the same number of neighbours within a fixed aperture, but where one halo may be located closer to a more massive entity than the other.

Fig. 3 shows correlations of the central density, spin, and formation redshift with halo mass as determined from TNG100-DMO and TNG300-DMO. The top left and top right panels, respectively, show the mass dependence of the central density defined by the NFW concentration of haloes (c_{NFW}) and our alternative parametrization, the ratio $\widetilde{V}_{max} \equiv V_{max}/(H_0 r_{max})$. Reassuringly, the latter definition shows a similar dependence on mass as the more traditional mass–concentration relation: low-mass haloes have higher concentrations (i.e. higher values of $\widetilde{V}_{max}$) while the typical scatter in each relation also increases with decreasing halo mass. The mass–concentration relation for the subset of dynamically relaxed haloes is represented by the points with error bars in the top left panel. Here, relaxed haloes are identified according to the criteria in Neto et al. (2007),

which defines relaxed haloes as objects where (i) the mass fraction in substructures contained within r_{200} is less than 10 per cent, (ii) the offset between the centre of mass of the halo and its centre of potential does not exceed $0.07 r_{200}$, and (iii) the virial ratio, $2T/|U| < 1.35$, where T is the kinetic energy of particles within r_{200} and U is their gravitational potential energy. The solid lines in this panel show that one measures systematically lower concentrations in haloes that are out-of-equilibrium (e.g. those that are currently undergoing a merger).

The bottom left panel in Fig. 3 shows the median spin–mass relation in TNG. As previous cosmological cold dark matter simulations have shown, there is a very weak (if any) correlation between halo spin and its mass, with a median value of $\lambda \approx 0.033$ across a wide range of halo mass (e.g. Davis et al. 1985; Barnes & Efstathiou 1987; Warren et al. 1992; Cole & Lacey 1996; Mo, Mao & White 1998; Bett et al. 2007; Zjupa & Springel 2017). Finally, the bottom right panel shows the dependence of the redshift of formation on halo mass, showing the behaviour expected of hierarchical structure formation, with low-mass haloes forming earlier than more massive objects. The results from TNG100-DMO and TNG300-DMO are
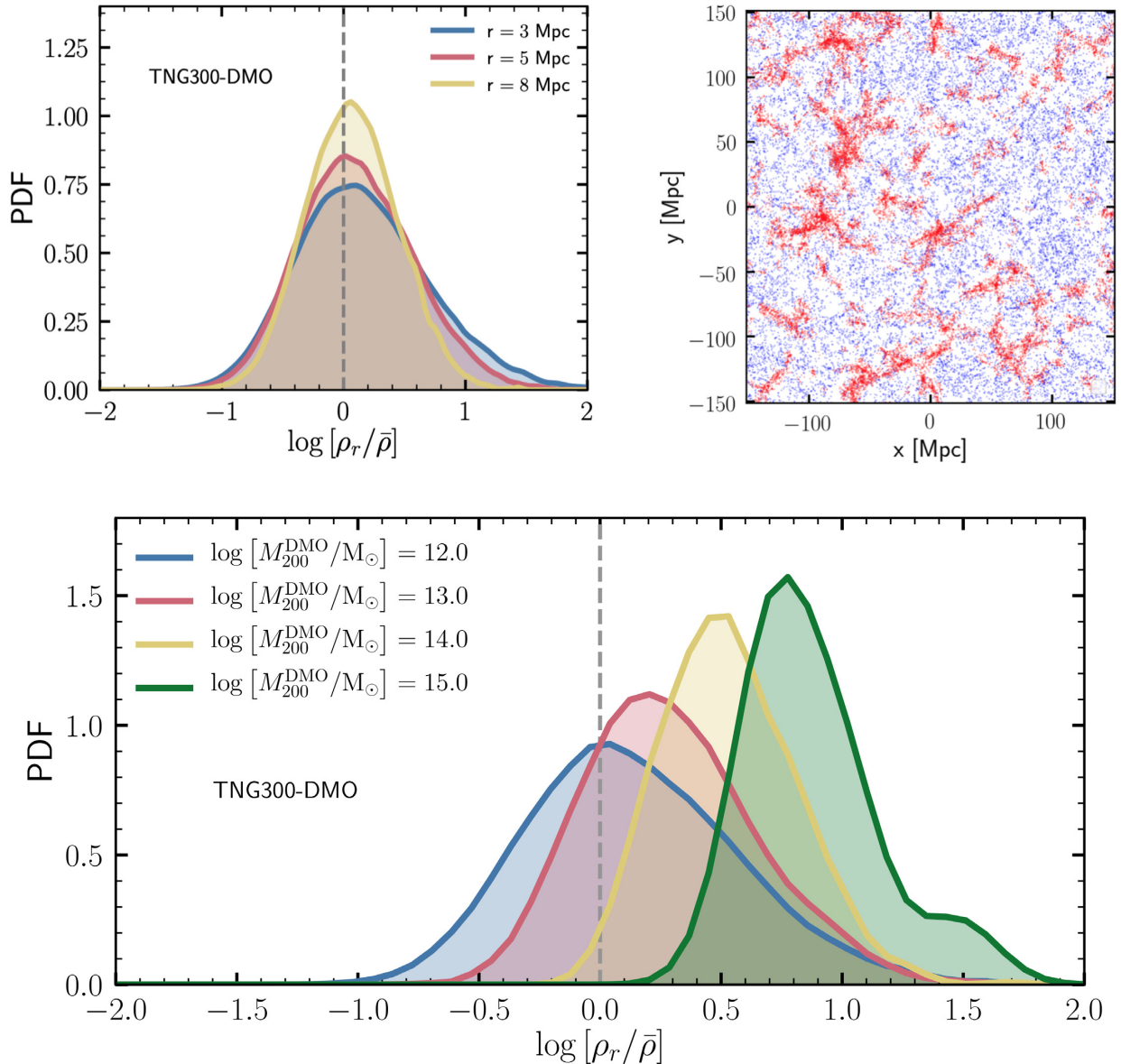
Figure 4. Environmental selection of dark matter haloes from the TNG300 DMO volume. *Top left panel*: Histograms of halo environmental overdensity, $\rho_r/\bar\rho$, for different aperture radius choices (see text for details). *Top right panel*: the locations of haloes existing in the 20 per cent most overdense (red) and 20 per cent most underdense (blue) regions in the TNG300 DMO simulation, defined by the $\rho_r/\bar\rho$ metric, assuming a fiducial filter size of $r = 5$ Mpc. For clarity, we only display the positions of haloes more massive than $M_{200}^{\rm DMO} > 10^{11}$ M$_\odot$ in a region of size $300 \times 300 \times 100$ Mpc. *Bottom panel*: the distribution of environments occupied by haloes split by mass. Again, we have assumed $r = 5$ Mpc.

well-converged over the mass scales in which the two simulations overlap; the combination of the simulations enables us to establish these relations over more than four orders of magnitude (although only a limited range is displayed in Fig. 3).

The top left panel of Fig. 4 shows histograms of halo environmental overdensities as determined by the $\rho_r/\bar\rho$ statistic. Similar to Artale et al. (2018), we have shown the distributions where the radius of the bounding sphere around each halo, $r = 3$, 5, and 8 Mpc. As expected, the distributions peak around $\rho_r/\bar\rho \approx 1$, which corresponds to mean density. The top right panel of this figure shows the locations of haloes existing in the 20 per cent most overdense (red) and 20 per cent most underdense (blue) regions in a slab of size $300 \times 300 \times 100$ Mpc centred on the TNG300-DMO box. In making this projection, we have assumed a filter size of $r =$

5 Mpc, which is our fiducial choice hereafter. The topology of filaments and voids is seen clearly using this simple definition for the large-scale environment. Finally, the bottom panel of Fig. 4 shows the distribution of environmental overdensities in which haloes of different mass live. As one would expect naïvely, more massive haloes exist preferentially in more overdense environments, while lower mass haloes are found in a broader range of overdensities.

3 RESULTS

This section presents the main results of this paper. First, we investigate the variation of the HOD as a function of specific DMO properties. In other words: how does the mean number of centrals, $\langle N_{\rm cen} \rangle$, and satellites, $\langle N_{\rm sat} \rangle$, hosted by haloes *at fixed mass* change

when selecting on properties such as concentration, formation time, spin, and environment (Section 3.1)? We then proceed to investigate the extent to which each of these correlations is independent; this is achieved through the construction of *predicted* HOD catalogues (Section 3.2). Finally, we examine the spatial arrangement of satellite galaxies within haloes, which is another important detail of mock catalogue construction (Section 3.3).

3.1 Responses in the halo occupation distribution

We begin by investigating the response of the HOD to selections on DMO properties at fixed halo mass. In the figures that follow, we showcase these responses by selecting haloes with the 20 per cent highest/lowest values of each DMO property in each bin of halo mass. The results are shown in Figs 5 and 6, and are summarized below:

(i) *Environmental bias*: the top panel of Fig. 5 shows the variation in the mean occupation of galaxies when selecting haloes in the 20 per cent most overdense and 20 per cent most underdense environments (as determined from the corresponding DMO simulation). We find that, at these number densities, the effect of environmental bias is small, particularly in the mean occupation of satellite galaxies. Haloes in the most overdense environments are somewhat more likely to host a central galaxy (with stellar mass $\geq \log M_\star^{\min}$) than those in underdense regions. This is signified by the separation of the blue and red curves below the knee of the HOD. This may be attributed to the increased frequency of mergers experienced by haloes in high density environments (e.g. Fakhouri & Ma 2009). While the effect on $\langle N_{\rm sat} \rangle$ is negligible, we note that the extent to which environmental bias introduces scatter in the mean occupation depends on the precise definition of a satellite galaxy i.e. the bounding radius within which a galaxy must be located in order to be counted as a satellite. As described in Section 2.2, here we have only considered objects located within r_{200}; the effect of varying this choice is presented in Appendix A.

(ii) *Age bias*: the middle panel of Fig. 5 shows the variation in the mean occupation of galaxies when selecting the 20 per cent earliest-forming and 20 per cent latest-forming haloes in each mass bin. The variation induced by the formation time of haloes is much stronger than environment, predominantly in the mean occupation of satellite galaxies. Younger haloes have, on average, more satellites than older ones; in the former case, satellites will have fallen into their host later and thereby undergo fewer orbital passages (during which they could be destroyed) than satellite galaxies in older haloes. The loss of a satellite may be associated to its bound stars being either transferred into the diffuse starlight in the halo or merged on to the central galaxy, but may also result from a satellite halo being disrupted below our minimum total/stellar mass cut implied by the HOD target density threshold. The relations then cross over around the knee of the HOD (where the contribution of $\langle N_{\rm sat} \rangle$ becomes subdominant to $\langle N_{\rm cen} \rangle$). At the lowest masses, early forming haloes are more likely to host a central galaxy through a combination of the added time available to form a central stellar component, and the increased stellar mass deposition from destroyed satellites.

(iii) *Central density bias*: the bottom panel of Fig. 5 shows the variation in the mean occupation of galaxies when selecting the 20 per cent most dense and 20 per cent least dense haloes, as determined by the $\widetilde{V}_{\max}$. At all masses, a more concentrated halo has fewer satellites at present day. This suggests that a halo with a higher central concentration of dark matter has a greater propensity to destroy small satellites through tidal forces; as we

demonstrate later in this section, these highly concentrated haloes also host more massive central galaxies. As in the case of the variation with formation time, the relations cross over at the knee of the HOD: at low-halo mass, a higher concentration halo is on average more likely to host a central galaxy above the minimum stellar mass adopted here ($\log[M_\star^{\min}/{\rm M}_\odot] = 9.28$ for TNG100 and $\log[M_\star^{\min}/{\rm M}_\odot] = 9.55$ for TNG300), owing to its deeper potential well. We have checked explicitly that both the trend and the size of this effect are identical when parameterizing the central density as the NFW concentration of the halo, $c_{\rm NFW}$.

(iv) *Angular momentum bias*: finally, Fig. 6 shows the variation in the mean occupation of galaxies when selecting the 20 per cent fastest rotating and 20 per cent slowest rotating haloes, selected according to their values of λ, the dimensionless spin parameter. The trends with spin may be most readily understood in the context of halo mergers: high angular momentum haloes are predominantly out-of-equilibrium objects that have undergone a relatively recent major merger (e.g. D'Onghia & Navarro 2007); these merger events can drag in satellites to add to the existing population within the halo. As we will demonstrate in the following subsection, the dependence of the HOD on halo spin is captured entirely by its dependence on the central density of haloes.

The strength of the response in the HOD to each property increases with decreasing halo mass; this is because the typical scatter in environment, formation time, concentration, and spin also increases with decreasing halo mass (Figs 3 and 4). We note that the dependence on environment, formation time, and central density have been established previously, such as in Zehavi et al. (2018) in the context of semi-analytic galaxy formation models, and by Chua et al. (2017) and Artale et al. (2018), who investigated these correlations using the ILLUSTRIS (Vogelsberger et al. 2014a) and EAGLE (Schaye et al. 2015) simulations. This work examines these relations in the context of the updated TNG model and is able to extend the analysis into the regime of rich clusters using the TNG300 volume. The qualitative response of the HOD to DMO properties is similar in all cases. In the following subsection, we show that each of these responses is not independent and may, in fact, be explained due to underlying correlations between the DMO properties themselves.

3.2 Independent correlations in the halo occupation distribution

In Section 3.1, we have shown that the mean occupation of galaxies in IllustrisTNG responds sensitively to the properties of the dark matter haloes hosting them. While these correlations are interesting physically, they pose a challenge to theoretical models of the HOD if each correlation is independent: to capture the scatter in the galaxy–halo occupancy, it would be necessary to build in the dependence on environment, concentration, formation time etc. in addition to halo mass. The question then arises as to how much weight one should assign to each of these dependencies.

Of course, each quantity we have examined is not entirely independent of the next. Consider, for example, a halo's large-scale environment. A halo originating from a Lagrangian region located in an overdensity is likely to collapse earlier than one forming in an underdensity (e.g. Sheth & Tormen 2004; Avila-Reese et al. 2005; Zehavi et al. 2018). In haloes that collapse earlier, the central core is assembled at an epoch when the mean density of the universe is higher; this results in a larger value of the central density (or concentration) measured at $z = 0$ (e.g. Li, Mo & Gao 2008; Gil-
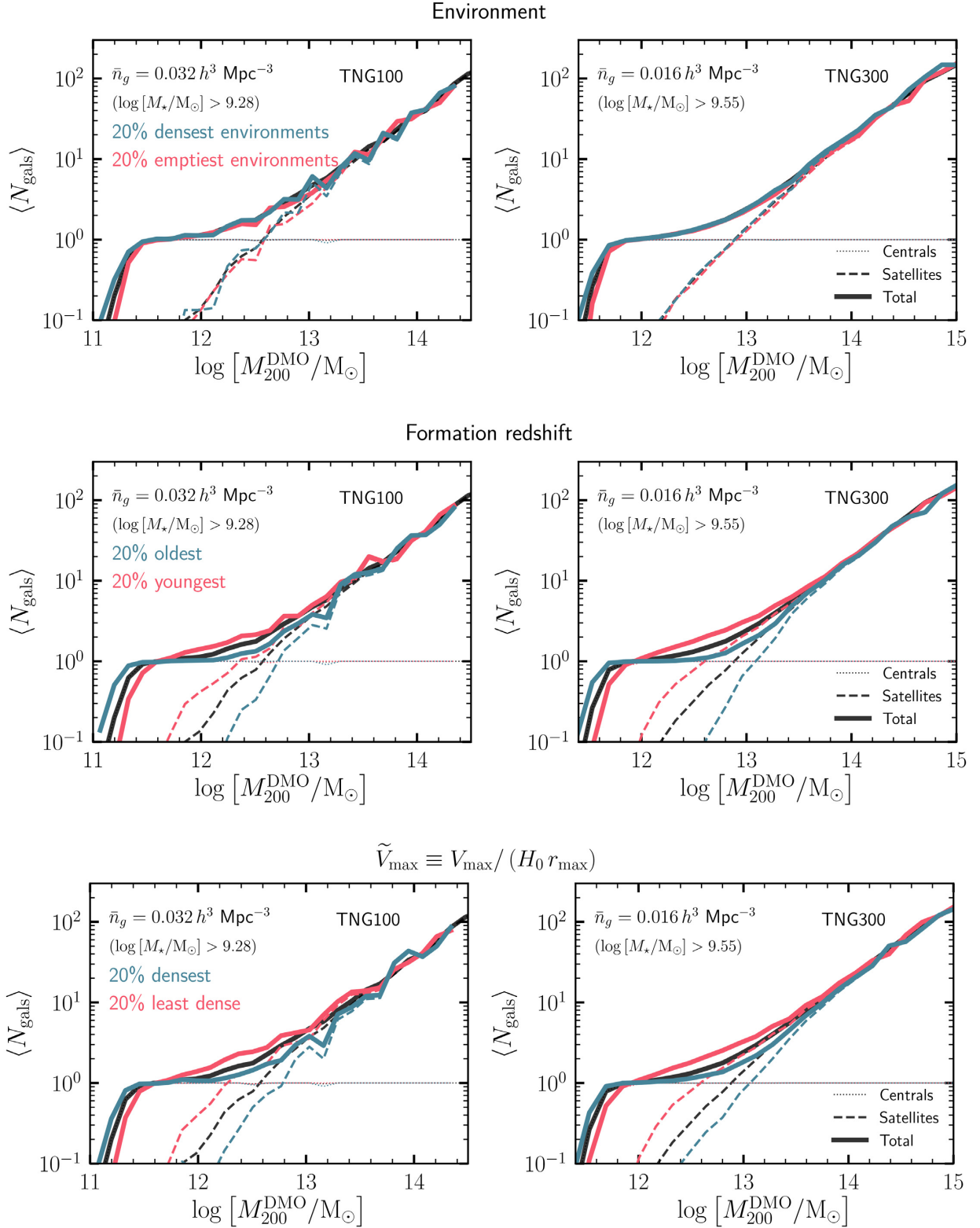
Figure 5. Variation in the mean HOD in the TNG100 (left column) and TNG300 (right column) simulations. Each HOD is constructed at a fixed number density as indicated in the top left corner of each panel. Each row shows the response of the mean HOD after selecting on various DMO properties: environment, formation redshift, and $\widetilde{V}_{\mathrm{max}}$. Centrals are shown in dotted lines, satellites by dashed lines and the total population with solid lines.
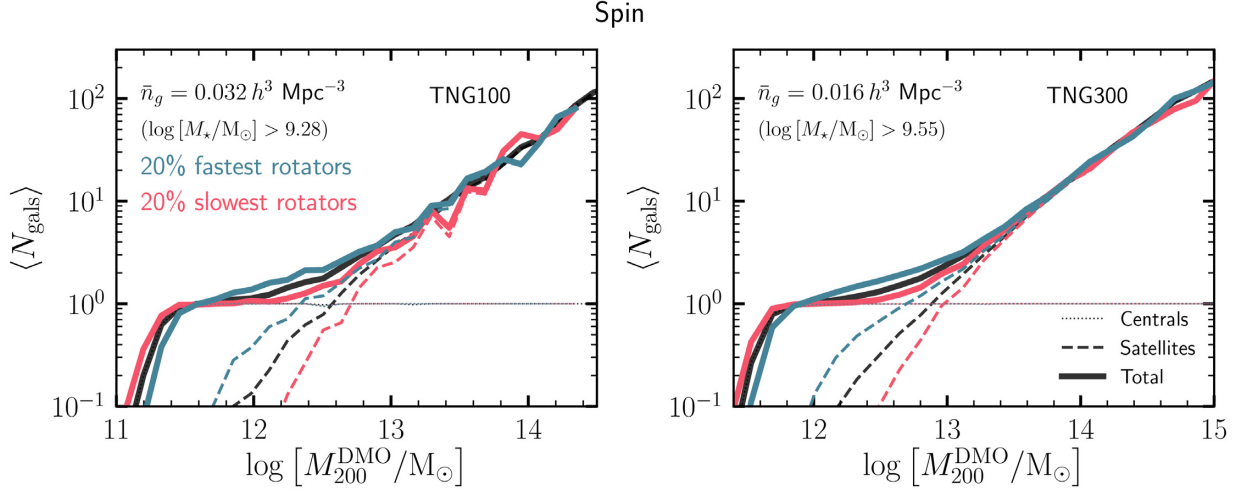
Spin

Figure 6. A continuation of Fig. 5 showing variations of the mean HOD with respect to a halo's dimensionless spin parameter.

Marín, Jimenez & Verde 2011). These correlations between halo properties, albeit weak, may also propagate into correlations in the HOD.

We test the independence of the HOD responses observed in Section 3.1 through the construction of 'synthetic' HODs. First, we construct a grid of dark matter halo mass (M_{200}^{DMO}, which we label as our 'fundamental' variable), a secondary DMO halo property (e.g. environment, which we label as the 'control' variable) and the corresponding number of centrals/satellites hosted by this halo in the full physics run (as determined by the matching procedure described in Section 2.2). Using this space, we construct a spline interpolation function to *predict* the average number of centrals/satellites associated with a halo *given* its mass and the value of the secondary halo property (e.g. the overdensity it lives in). Mathematically, while the *true* HOD measured from IllustrisTNG is given by $\langle N_{gals}|M_{200}^{DMO}\rangle$, the *predicted* HOD is defined by $\langle N_{gals}|M_{200}^{DMO}, \text{control}\rangle$.

Diagrammatically, this procedure is illustrated in Fig. 7. The new HOD, by construction, reproduces the dependence of the mean occupation on the 'control' variable. If after selecting on a *third* DMO property we are able to recover the responses observed in Figs 5 and 6, this third parameter does not give us any more information than the bivariate combination of DMO halo mass and DMO 'control' variable alone. Conversely, a failure to recover the dependency indicates that there is more information to be gained from including a third parameter in the HOD. The methodology we adopt here is inspired by Blanton et al. (2005), who investigated the relationship between environment and broad-band optical properties like luminosity, surface brightness, and colour of galaxies in the Sloan Digital Sky Survey.

The first control variable we consider is environment. As we have described above, we create grids of $N_{cen}(M_{200}^{DMO}, \rho_r/\bar{\rho})$ and $N_{sat}(M_{200}^{DMO}, \rho_r/\bar{\rho})$. We then consider responses of this new HOD to selection on $\widetilde{V}_{max}$ and formation time; the results are shown in Fig. 8. Clearly, the strong correlations with concentration and formation time we have measured previously are no longer reproduced: this suggests that the combined knowledge of the mass and large-scale environment of a halo is unable to predict the dependence of the HOD on other halo properties. In other words, while pairs of halo properties (e.g. environment and formation time) may themselves be correlated, this does not guarantee that one property is successful at predicting the response of the HOD to the other. This conclusion

is akin to the observation made by Mao et al. (2018), who explored the role of halo properties in determining galaxy clustering.

The next control variable we select is halo concentration, parametrized by $\widetilde{V}_{max}$. The HOD predicted by this variable and its associated responses are shown in Fig. 9. We now successfully predict the response of the HOD due to formation time, spin, and velocity dispersion using halo mass and $\widetilde{V}_{max}$ as the fundamental and control variables, respectively. In particular, the qualitative response of the predicted HOD to each additional property is preserved: for example, early-forming haloes are still predicted to form fewer satellites than late-forming ones, as we inferred from the true HODs measured in TNG. There is some residual scatter, particularly in $\langle N_{sat}\rangle$, which the $\widetilde{V}_{max}$-predicted HODs are unable to capture; this dominates in the regime when only one in every few haloes contain a galaxy above the minimum stellar mass threshold.

It is worth reflecting about the physical reasons as to why the quantity $\widetilde{V}_{max}$ captures more general dependencies in the HOD. That this parameter also reproduces the dependence on, say, halo spin, is not trivial, but may be understood as follows. At fixed mass, the distribution of high-spin haloes are thought to be comprised of systems that have recently undergone a major merger (e.g. Vitvitska et al. 2002; Peirani, Mohayaee & de Freitas Pacheco 2004; D'Onghia & Navarro 2007). These out-of-equilibrium systems are also the ones for which we measure preferentially lower concentrations at fixed mass – the subset of 'unrelaxed' haloes represented by the solid lines in Fig. 3 (top left panel). High spin and low concentration are therefore intimately related with one another; both predict reduced satellite counts and therefore affect the HOD in the same way. Indeed, on comparing the bottom row of Fig. 5 (concentration) with Fig. 6 (spin), we find that response of $\langle N_{sat}\rangle$ is similar in both size and shape, but with the colours of the lines reversed.

Ultimately, the efficacy of $\widetilde{V}_{max}$ as a variable for predicting the scatter in the HOD stems from it being a key parameter in quantifying the scatter in the mass of the central galaxy at fixed halo mass. The relationship between the dark matter and stellar content of a halo is a key constraint that galaxy formation models aim to reproduce. Indeed, the *mean* stellar-to-halo mass relation at $z = 0$, inferred from abundance matching, is one of the quantities that the subgrid physics model in TNG is was designed to reproduce, at least in its general shape (Pillepich et al. 2018a). Simulations provide a unique opportunity to explore the scatter around this
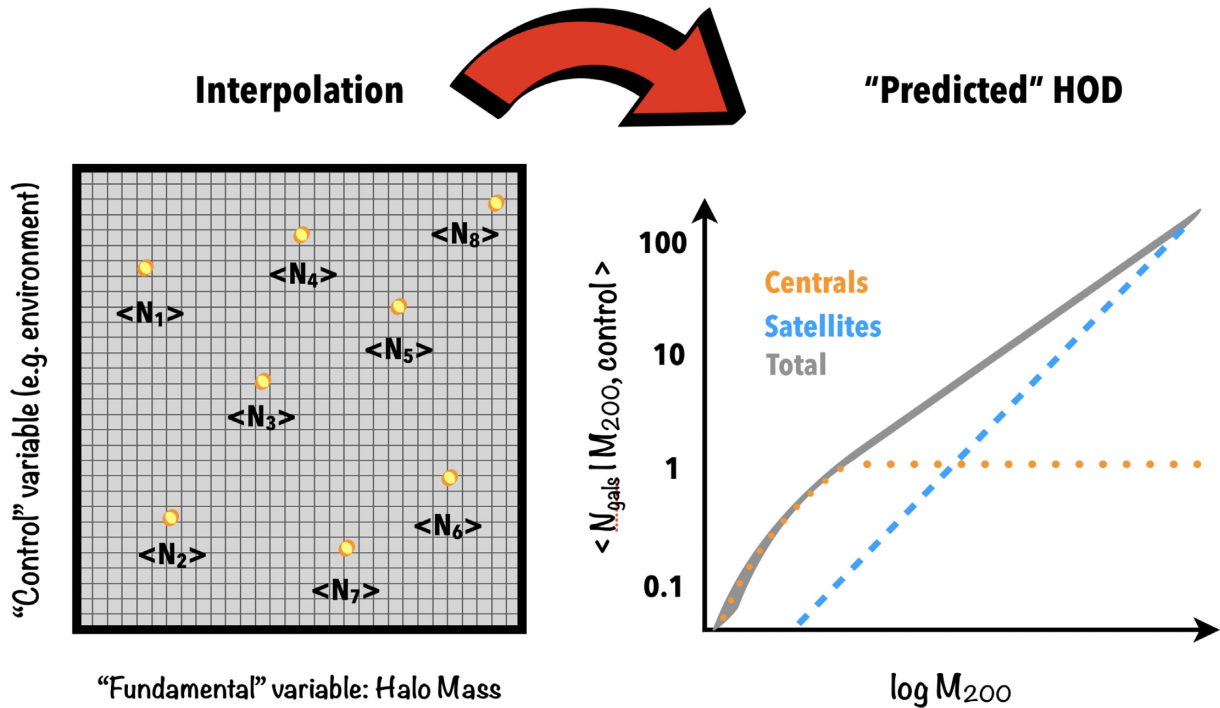
Figure 7. A schematic illustration of how we construct the HOD predicted by pairs of halo properties, by interpolating values of $\langle N_{\mathrm{cen}} \rangle$ and $\langle N_{\mathrm{sat}} \rangle$ on a grid of halo mass and a 'control' variable.

relation induced by properties intrinsic to haloes, the effects of supernova and AGN feedback etc. (e.g. Trujillo-Gomez et al. 2011; Zentner, Hearin & van den Bosch 2014; Matthee et al. 2017; He 2019). Such an exploration is performed more easily in the case of central galaxies, which are free of tidal effects from larger host haloes.

Both idealized and fully hydrodynamical simulations have demonstrated that the presence of a massive central galaxy can significantly deplete a halo's satellite population (e.g. Zentner et al. 2005; D'Onghia et al. 2010; Yurin & Springel 2015; Zhu et al. 2016; Chua et al. 2017; Garrison-Kimmel et al. 2017; Sawala et al. 2017; Richings et al. 2018). This destruction acts preferentially on low-mass satellites on radial orbits with close pericentric passages to the central disc; the depletion also becomes stronger the more massive the central galaxy (Samuel et al. 2019). Properties of dark matter haloes that capture the scatter in the stellar-to-halo mass relation of central galaxies will naturally be good predictors for the variation in the mean occupancy of satellite galaxies.

In Fig. 10, we display the mean and the scatter in the stellar-to-halo mass relation of central galaxies in TNG300.[3] Here, stellar mass is defined as the total mass in star particles contained within a 3D aperture of (physical) size 30 kpc centred on the halo. The dashed lines show shifts in the mean relation (which is shown by the solid white curve) when selecting subsets of the halo population according to their central density only ($\widetilde{V}_{\mathrm{max}}$, left-hand panel) and, in addition, its large-scale environment ($\rho_r/\bar{\rho}$, right-hand panel).

We see clear departures from the mean relation in both panels. In particular, haloes with higher central densities (concentration) host more massive central galaxies; these haloes have deeper potential wells which bring more stellar mass to the centre, and they also form

earlier, allowing more time for the central galaxy to build in mass. The right-hand panel of Fig. 10 shows that the densest haloes that also exist in most overdense environments host even more massive central galaxies, although the dominant contribution comes from the higher concentration of the halo. The correlations we measure between central stellar mass and concentration are consistent with those measured by Matthee et al. (2017) and Artale et al. (2018). This dependence naturally translates into the connection between the HOD and the halo concentration.

We have repeated the exercise of constructing HODs predicted with other DMO properties, using spin, formation time, and velocity dispersion as control variables. Neither spin nor velocity dispersion are as effective as $\widetilde{V}_{\mathrm{max}}$ in recovering correlations in the HOD with a third variable. The formation redshift, z_{form}, is the next best-performing parameter. This comes as no surprise, as several analytic models have demonstrated the intimate link between the z_{form} and the central density of dark matter haloes (e.g. Navarro et al. 1997; Correa et al. 2015; Ludlow et al. 2016). For more generalized applications, however, $\widetilde{V}_{\mathrm{max}}$ is a more convenient second parameter to incorporate into an HOD as it is well-defined: to compute this quantity, there is no need to traverse merger trees in which the number of outputs used to construct the tree affects the accuracy with which z_{form} may be measured. Parametrizing central density as $\widetilde{V}_{\mathrm{max}}$ as opposed to the concentration of an equivalent NFW halo also circumvents the need to access particle data.

3.3 The radial distribution of galaxies in haloes

We conclude our analysis by examining how satellite galaxies in IllustrisTNG are distributed spatially within their host haloes. In the HOD formalism, an appropriate assignment of galaxy positions is vital to recover accurate small-scale ($\lesssim 1$ Mpc) clustering (the so-called 'one-halo' term). Typically, satellite positions are assigned in one of three ways: (i) assuming that satellites trace the dark

---

[3] A comparison between the stellar-to-halo mass relations in TNG and the standard trends inferred from abundance matching has been presented in Pillepich et al. (2018b).
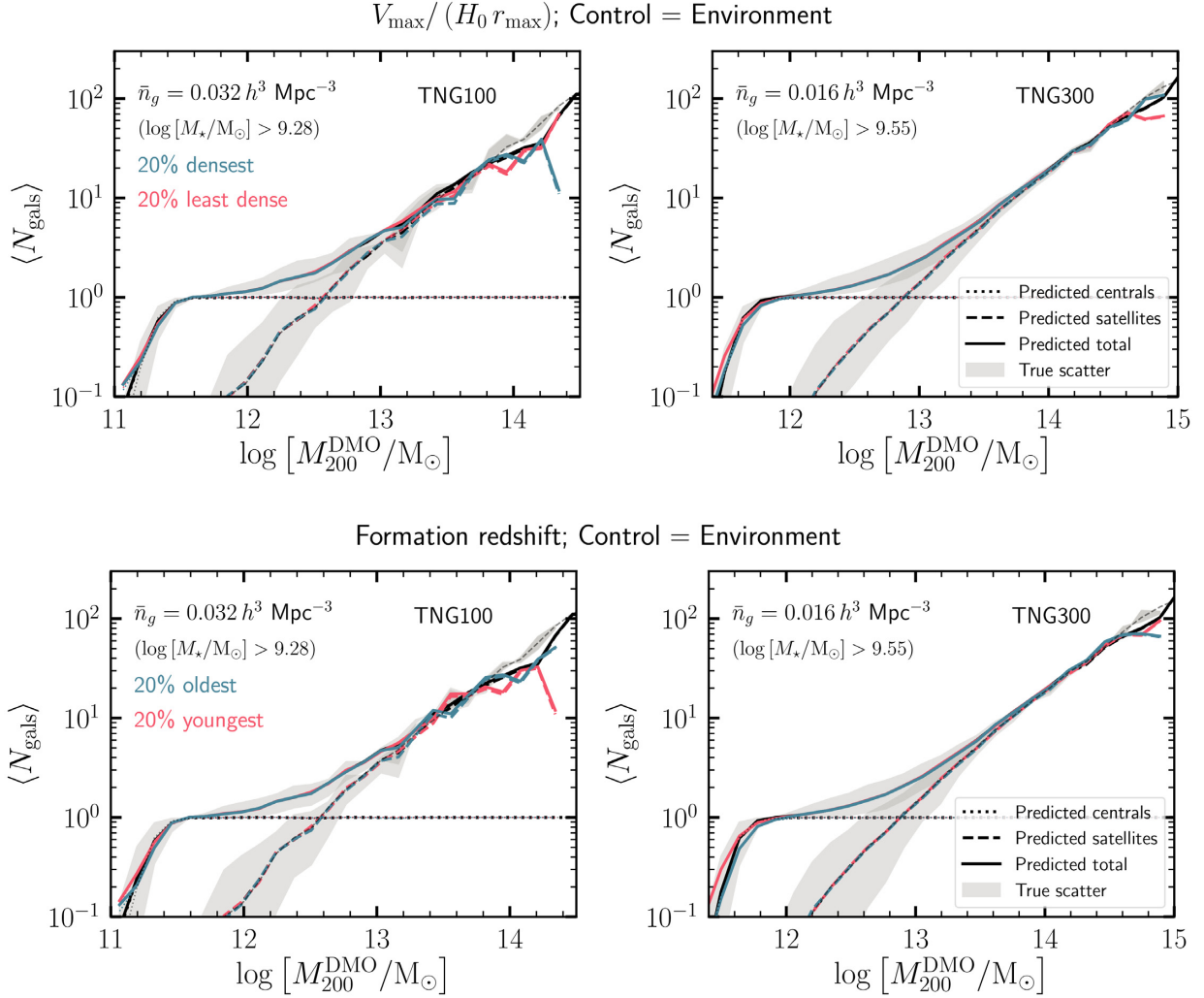
Figure 8. As Fig. 5, but now showing the mean HOD that is *predicted* by a halo's mass, M_{200}^{DMO}, and its large-scale environment in the DMO simulation. The variation of the HOD with respect to selection on environment is reproduced by construction and is not shown here. Predicting the halo occupancy using halo mass and environment alone is unable to reproduce correlations with other halo properties.

matter particles and are distributed according to the best-fitting NFW profile of the host halo; (ii) by assigning satellite positions by placing each one on a randomly selected dark matter particle; (iii) assuming they follow the radial distribution of subhaloes in a DMO simulation. In a hydrodynamical simulation like IllustrisTNG, the orbits of satellite galaxies are evolved self-consistently, providing knowledge of this assignment for free.

In Fig. 11, we present the radial profiles of satellite galaxies measured in TNG100, which offers better mass and spatial resolution than TNG300. In particular, we show the radial number density of luminous galaxies (containing at least 100 star particles), $n(r)$, normalized to the mean number density of satellites identified within r_{200}, denoted as $\langle n_{200}\rangle$. These profiles are represented by the starred symbols, with each panel showing the result for a different range of host halo mass. The normalized profiles take on a universal shape across all halo mass; their shape and normalization are independent of the minimum stellar mass cut adopted in constructing the satellite profiles (see Appendix B). The light grey curve in each panel shows the quantity $n(r)/\langle n_{200}\rangle$ measured for subhaloes (containing more than 100 dark matter particles) in the DMO counterparts of these host haloes. The distribution of subhaloes from TNG100-DMO is

markedly different to the distribution of galaxies, which is much more centrally concentrated. It is well-known that the distribution of subhaloes in a DMO simulation may be fit with an Einasto profile (Einasto 1965):

$$\log\left[\frac{n(r)}{n_{-2}}\right] = -\frac{2}{\alpha}\left[\left(\frac{r}{r_{-2}}\right)^{\alpha} - 1\right],\qquad(5)$$

with a shape parameter, $\alpha = 0.678$, and the parameters n_{-2} and r_{-2} obtained through fitting (e.g. Springel et al. 2008). Furthermore, this shape is independent of subhalo mass (e.g. De Lucia et al. 2004; Gao et al. 2004; Gill, Knebe & Gibson 2004). As shown in Fig. 11, an Einasto profile fit to DMO subhaloes is clearly inaccurate when applied to galaxies.

The dashed orange curve shows an NFW profile where the concentration is set to the median concentration of DMO haloes in this mass bin, and, perhaps surprisingly, provides an excellent match the radial distribution of satellite galaxies in TNG100. To understand why galaxies tend to prefer the more centrally concentrated profile of the total matter, as opposed to the shallower profile of the DMO subhaloes, it is important to identify the subset of subhalo population that is most likely to host a galaxy.
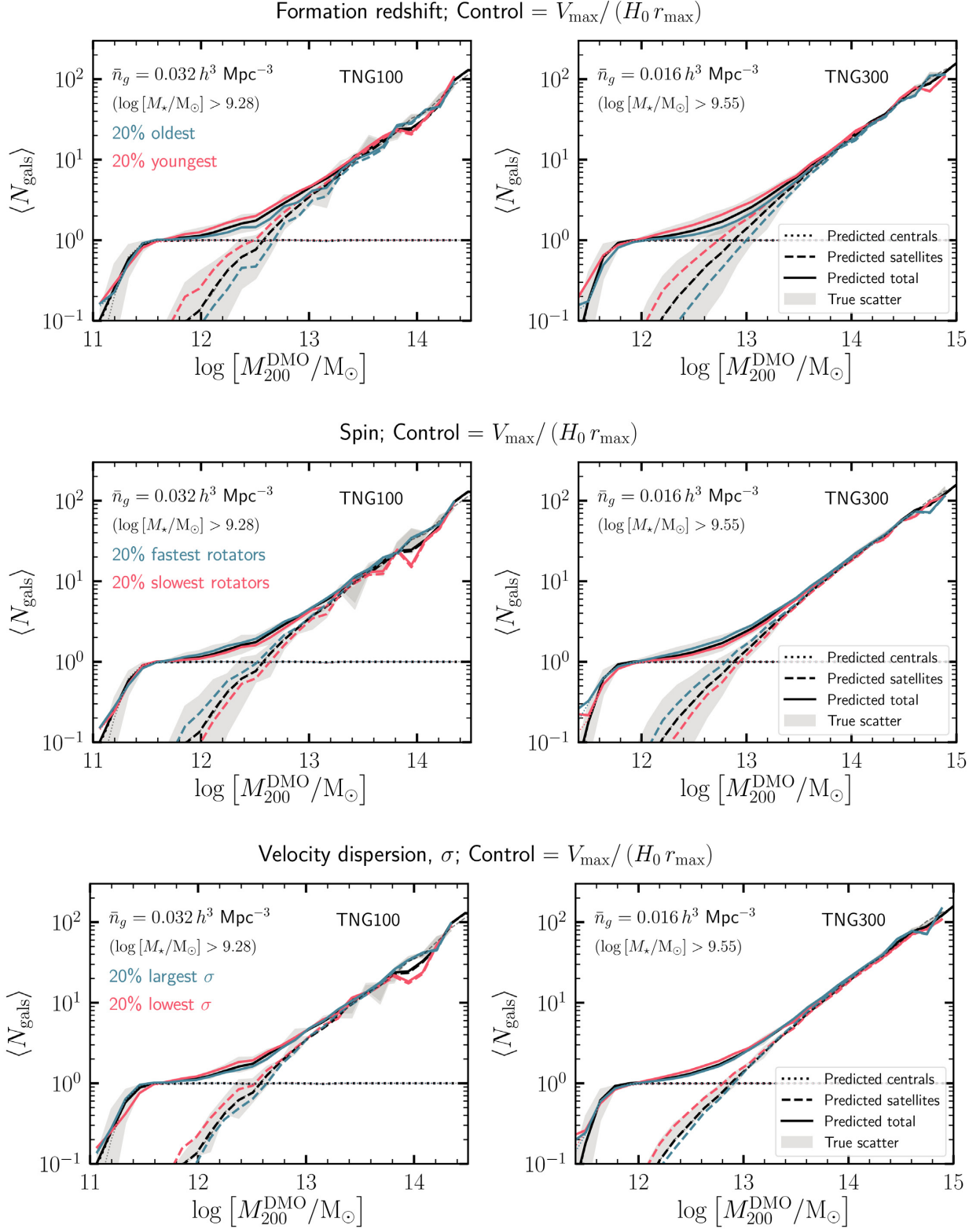
Figure 9. As Fig. 8, but now showing the mean HOD that is *predicted* by a halo's mass, $M_{200}^{\rm DMO}$, and $\widetilde{V}_{\rm max}$ in the DMO simulation. The variation of the HOD with respect to selection on $\widetilde{V}_{\rm max}$ is reproduced by construction. Velocity dispersion, σ, shown in the bottom row, is defined as the one-dimensional dispersion of dark matter particle velocities contained within r_{200}. A predictor based on halo mass and central density is, to a large extent, able to reproduce correlations with other halo properties.
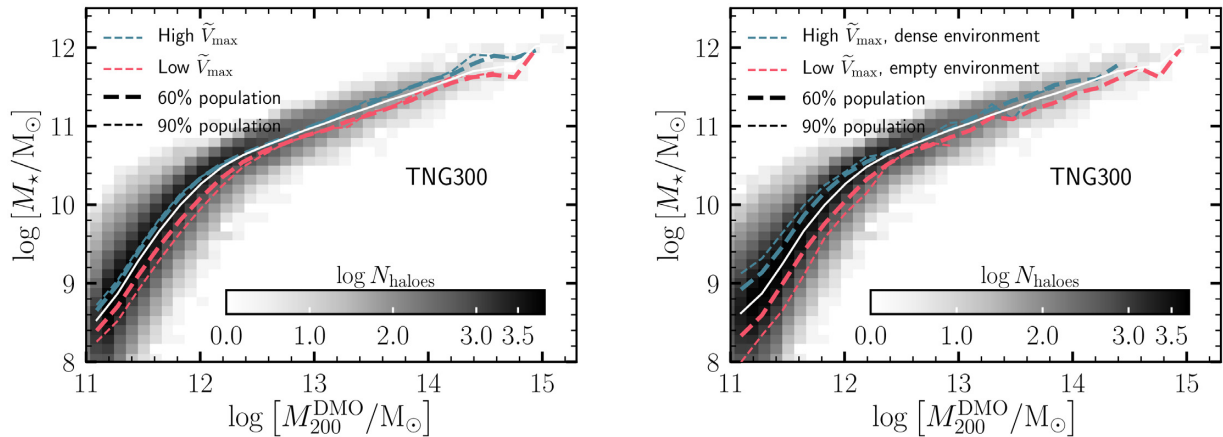
Figure 10. The mean stellar mass, $M_\star$, versus halo mass, M_{200}, relation for central galaxies at $z = 0$ in TNG300. The stellar mass of the central galaxy is defined as the total mass in all star particles identified within a 3D aperture of (physical) size 30 kpc placed at the halo centre. *Left-hand panel*: the blue and red dashed curves, respectively, show the mean stellar mass–halo mass relation for haloes with higher and lower than average values of $\widetilde{V}_{\rm max}$ at fixed halo mass. The thick dashed curves represent the haloes with the 20 per cent largest (blue) values of $\widetilde{V}_{\rm max}$ and 20 per cent smallest (red) $\widetilde{V}_{\rm max}$ (i.e. encompassing 80 per cent of the population); the thin curves show the same for the 5 per cent and 95 per cent cases (i.e. encompassing 90 per cent of the population). The white curve shows the mean stellar-to-halo mass relation. *Right-hand panel*: similar to the panel on the left, but including the environmental overdensity in addition to $\widetilde{V}_{\rm max}$.

For this it is useful to think about the galaxy–halo connection in terms of subhalo abundance matching. In its most traditional form, abundance matching associates the brightest galaxies with the most massive subhalo, defined by values measured at $z = 0$. Unlike independent haloes, subhaloes (the hosts of satellite galaxies) are subject to tidal stripping and therefore lose mass continually after infall by an amount that depends on their orbit, infall mass, and infall time: subhaloes that are more massive prior to infall (and therefore more likely to host a galaxy) undergo greater dynamical friction and sink to the middle of halo, where the mass is stripped more rapidly. Information about the mass being high before infall may therefore be lost through dynamical evolution. Furthermore, substructure finding algorithms are likely to underestimate the mass of subhaloes near the centre of the host halo, or may even fail to detect them altogether.

A more powerful proxy of the potential well associated with a subhalo is provided the quantity $V_{\rm peak}^{\rm DMO}$, defined as the maximum value of $V_{\rm max}$ attained by a DMO subhalo at *any point* in its history (typically just before infall). This quantity, free from the effects of stripping provides a more direct connection between a halo and its galaxy and has indeed been demonstrated to be a more robust metric for subhalo abundance matching (e.g. Conroy, Wechsler & Kravtsov 2006; Trujillo-Gomez et al. 2011; Nuza et al. 2013; Chaves-Montero et al. 2016).

Selecting subhaloes by $V_{\rm peak}^{\rm DMO}$ rather than present-day mass results in considerably different radial distributions, as shown by the dark grey and black curves in Fig. 11. Subhaloes in TNG100-DMO with the highest values of $V_{\rm peak}^{\rm DMO}$ are more centrally concentrated; objects with $V_{\rm peak}^{\rm DMO}/V_{200}^{\rm host} > 0.25$ are sorted nearly identically to the total matter distribution. Note that we do not impose any $z = 0$ particle number cut when constructing the $V_{\rm peak}$-selected radial profiles. These observations are consistent with the findings of previous works by Nagai & Kravtsov (2005), Faltenbacher & Diemand (2006), Kuhlen, Diemand & Madau (2007), and Ludlow et al. (2009). As it has been shown by Springel et al. (2008), the spatial bias of subhalo clustering is camouflaged when these objects are selected by their $z = 0$ mass.

In summary, we find that the radial profiles of galaxies in TNG are described very accurately by the best-fitting NFW profile of the dark matter particles in a DMO simulation. This represents the centrally concentrated distribution of subhaloes with the highest values of $V_{\rm peak}^{\rm DMO}$, which are the entities that are most likely to host a galaxy at present day. On the other hand, the radial profile of DMO subhaloes selected by present-day mass is much shallower than that of galaxies, particularly within $r \lesssim 0.4 r_{200}$. The correct assignment of galaxy positions within their host haloes influences the amplitude of the one-halo term in the two-point correlation function (see e.g. Reddick et al. 2013).

4 CONCLUSIONS

In this paper, we have presented a detailed study of the galaxy–halo connection as revealed by a modern cosmological, hydrodynamical simulation. In particular, we exploit the combination of high-mass resolution and large volume of the 100 and 300 Mpc boxes produced as part of the IllustrisTNG project (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018; Springel et al. 2018; Pillepich et al. 2018b) to explore the HOD.

Our objective has been to identify properties of haloes in a dark matter-only (DMO) simulation that are most predictive of its likelihood to host central/satellite galaxies in a counterpart hydrodynamical simulation. We achieve this by matching haloes more massive than $\sim 10^{11}\,{\rm M}_\odot$ between DMO and full physics versions of each TNG volume, after which we construct galaxy catalogues at fixed number density whilst being careful to include only those galaxies that are well resolved (Figs 1 and 2). In TNG100, we construct the HOD for a number density, $\bar{n}_g = 0.032$ and $\bar{n}_g = 0.016\,h^3\,{\rm Mpc}^{-3}$ in TNG300; we provide fitting formulae for the HODs resulting from these number densities in equations (1), (2), and Table 1. The minimum galaxy stellar mass implied by these number density thresholds is $\log[M_\star^{\rm min}/{\rm M}_\odot] = 9.28$ in TNG100 and $\log[M_\star^{\rm min}/{\rm M}_\odot] = 9.55$ in TNG300. We focus specifically on the following properties and their influence on the mean HOD: concentration (parametrized as the quantity $\widetilde{V}_{\rm max} \equiv V_{\rm max}/(H_0\,r_{\rm max})$),
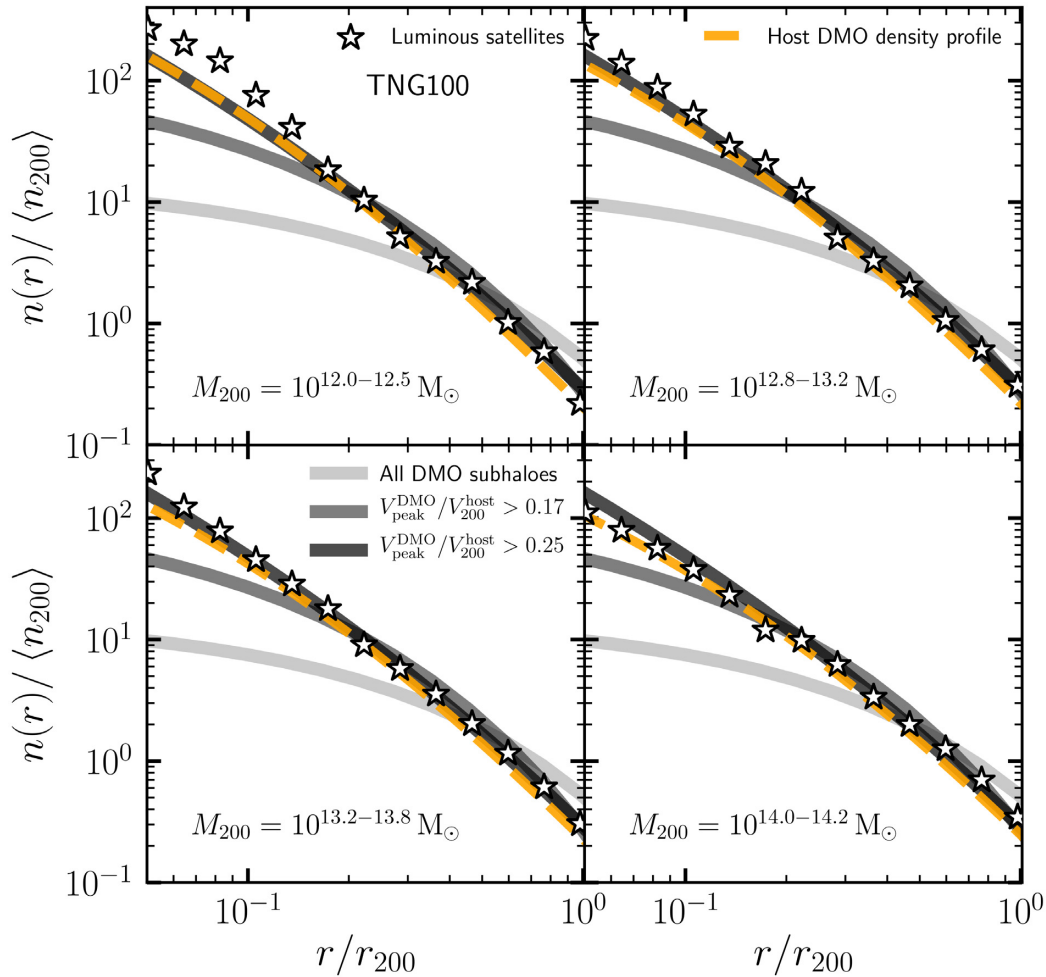
Figure 11. The mean radial number density of luminous satellites normalized to the mean number of luminous galaxies in bins of host halo mass (individual panels). The light grey curve shows the normalized radial profile of dark matter subhaloes contained within hosts of this mass in the DMO simulation; this is a mass-selected sample with the requirement that each subhalo contains at least 100 dark matter particles at present day. The dark grey and black curves show radial profiles of subhalo populations selected according to $V_{\rm peak}^{\rm DMO}/V_{200}^{\rm host}$, the ratio of the subhalo's peak circular velocity to the virial velocity of its host. The $V_{\rm peak}$-selected profiles do not include any $z = 0$ particle number resolution limit. Finally, the dashed orange curve shows an NFW profile fit to the density profile of DMO haloes in this mass range. For clarity, we have displayed the results for the TNG100 simulation only; radial profiles in the TNG300 simulation are qualitatively similar.

formation time, spin (Fig. 3), and environment (Fig. 4). Our findings are summarized below:

(i) The response of the HOD is very weakly correlated with environment, although haloes in high density environments are somewhat more likely to host more satellites than those living in underdensities. Low-mass haloes are more likely to host a central galaxy (with stellar mass $\geq \log M_{\star}^{\rm min}$) if located in a high density region (Fig. 5, top panel). The extent to which environmental bias plays a part in introducing scatter to the HOD depends on the outermost boundary used to define the satellite content of a halo (Appendix A).

(ii) Younger haloes have, on average, more satellites than older ones; these satellites have had more time to orbit the halo and potentially merge on to the central galaxy (Fig. 5, middle panel).

(iii) At all masses, a more concentrated halo has fewer satellites at present day. At low halo mass, a higher concentration halo is on average more likely to host a central galaxy, owing to its deeper potential well (Fig. 5, bottom panel).

(iv) Haloes with larger angular momentum contain more satellites than the average population, and vice versa. Previous works have indicated that high angular momentum haloes have typically undergone a recent merger event; these out-of-equilibrium haloes exhibit lower concentrations and later formation times (Fig. 6).

To test the independence of the HOD response to each variable, we use pairs of halo properties to construct 'controlled' galaxy catalogues: where the mean occupation of centrals and satellites is *predicted* using a combination of halo mass and a second, control variable (Fig. 7). These tests show that:

(v) Despite intrinsic correlation between environment and halo formation time/concentration/spin, an HOD predicted using halo mass and large-scale environment alone is unable to capture the correlations with other halo properties (Fig. 8).

(vi) Among the remaining properties, halo concentration as defined by $\widetilde{V}_{\rm max}$ performs the best in predicting the response of the HOD to tertiary properties like environment, spin, and velocity dispersion (Fig. 9). Formation time performs almost as well, with

the disadvantage that it is not as easily determined using $z = 0$ halo properties.

(vii) The success of concentration as a secondary variable in the HOD is driven by the fact that haloes with steeper inner potentials are more efficient at destroying orbiting satellites. Furthermore, concentration acts as a key secondary parameter in describing the scatter of central galaxy mass at fixed halo mass: haloes with higher concentration host, on average, a more massive central galaxy than those with lower concentration (Fig. 10). A more massive central galaxy may have a higher propensity to destroy orbiting satellites. Furthermore, a satellite orbiting an early forming, high concentration host halo has more time to merge on to the central galaxy by $z = 0$.

(viii) Finally, we measure the radial distribution of luminous satellite galaxies in TNG haloes and find that their profile is described very well by an NFW profile fit to the total matter distribution in the DMO counterparts to these haloes (Fig. 11). This reflects the more centrally concentrated distribution of subhaloes with the highest values of peak circular velocity, $V_{\rm peak}^{\rm DMO}$, which these satellites occupy preferentially. On the other hand, a sample of DMO subhaloes selected by present-day mass exhibits a comparatively shallower profile towards the centre of the host halo.

The HOD formalism provides a convenient framework to understand how the simplest property of a dark matter halo – namely, its mass – may be used to predict the probability for it to contain central and/or satellite galaxies. Cosmological, hydrodynamical simulations that model simultaneously the evolution of dark and baryonic matter provide the ideal test bed for exploring and quantifying the scatter in the HOD as a function of secondary halo properties.

In TNG, we find that halo mass and concentration, defined as the quantity $\widetilde{V}_{\rm max} \equiv V_{\rm max}/(H_0\, r_{\rm max})$, are two particularly prominent variables in determining the occupancy of galaxies in haloes. This combination of parameters is especially enticing as each one is standard output of all halo finding algorithms, thereby circumventing the need to fit density profiles to particle data or traverse merger trees. The physical quantities that these parameters represent are general enough that they may be readily used to construct HOD catalogues for much larger volume simulations as demanded by the next generation of galaxy surveys for which equivalent hydrodynamical simulations will not be computationally feasible. Finally, since halo mass, $V_{\rm max}$ and $r_{\rm max}$ evolve self-consistently according to the underlying theory of gravity, the parametrization we have explored may also be applied to cosmological models beyond traditional ΛCDM.

ACKNOWLEDGEMENTS

REFERENCES

Adhikari S., Dalal N., Chamberlain R. T., 2014, J. Cosmol. Astropart. Phys., 11, 019

Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012, MNRAS, 426, 2046
Aragón-Calvo M. A., Jones B. J. T., van de Weygaert R., van der Hulst J. M., 2007, A&A, 474, 315
Artale M. C., Zehavi I., Contreras S., Norberg P., 2018, MNRAS, 480, 3978
Avila-Reese V., Colín P., Gottlöber S., Firmani C., Maulbetsch C., 2005, ApJ, 634, 51
Barnes J., Efstathiou G., 1987, ApJ, 319, 575
Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
Behroozi P., Wechsler R., Hearin A., Conroy C., 2019, MNRAS, 488, 3143
Benson A. J., 2012, New Astron., 17, 175
Benson A. J., 2017, MNRAS, 471, 2871
Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000, MNRAS, 311, 793
Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
Bertschinger E., 1985, ApJS, 58, 39
Bett P., Eke V., Frenk C. S., Jenkins A., Helly J., Navarro J., 2007, MNRAS, 376, 215
Blanton M. R., Eisenstein D., Hogg D. W., Schlegel D. J., Brinkmann J., 2005, ApJ, 629, 143
Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440
Bose S. et al., 2019, MNRAS, 486, 4790
Bray A. D. et al., 2016, MNRAS, 455, 185
Busch P., White S. D. M., 2017, MNRAS, 470, 4767
Cautun M., van de Weygaert R., Jones B. J. T., 2013, MNRAS, 429, 1286
Chaves-Montero J., Angulo R. E., Schaye J., Schaller M., Crain R. A., Furlong M., Theuns T., 2016, MNRAS, 460, 3100
Chisari N. E. et al., 2018, MNRAS, 480, 3962
Chua K. T. E., Pillepich A., Rodriguez-Gomez V., Vogelsberger M., Bird S., Hernquist L., 2017, MNRAS, 472, 4343
Colberg J. M., 2007, MNRAS, 375, 337
Cole S., Kaiser N., 1989, MNRAS, 237, 1127
Cole S., Lacey C., 1996, MNRAS, 281, 716
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Cole S. et al., 2005, MNRAS, 362, 505
Colless M. et al., 2001, MNRAS, 328, 1039
Conroy C., Wechsler R. H., 2009, ApJ, 696, 620
Conroy C., Wechsler R. H., Kravtsov A. V., 2006, ApJ, 647, 201
Contreras S., Zehavi I., Baugh C. M., Padilla N., Norberg P., 2017, MNRAS, 465, 2833
Correa C. A., Wyithe J. S. B., Schaye J., Duffy A. R., 2015, MNRAS, 452, 1217
Croton D. J. et al., 2006, MNRAS, 365, 11
D'Onghia E., Navarro J. F., 2007, MNRAS, 380, L58
D'Onghia E., Springel V., Hernquist L., Keres D., 2010, ApJ, 709, 1138
Dalal N., White M., Bond J. R., Shirokov A., 2008, ApJ, 687, 12
Davé R., Thompson R., Hopkins P. F., 2016, MNRAS, 462, 3265
Davis M., Geller M. J., 1976, ApJ, 208, 13
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
De Lucia G., Kauffmann G., Springel V., White S. D. M., Lanzoni B., Stoehr F., Tormen G., Yoshida N., 2004, MNRAS, 348, 333
Desjacques V., Jeong D., Schmidt F., 2018, Phys. Rep., 733, 1
Diemer B., Kravtsov A. V., 2014, ApJ, 789, 1
Doroshkevich A. G., 1970, Astrophysics, 6, 320
Dressler A., 1980, ApJ, 236, 351
Dubois Y. et al., 2014, MNRAS, 444, 1453
Efstathiou G., Davis M., White S. D. M., Frenk C. S., 1985, ApJS, 57, 241
Efstathiou G., Frenk C. S., White S. D. M., Davis M., 1988, MNRAS, 235, 715
Einasto J., 1965, Trudy Astrofizicheskogo Instituta Alma-Ata, 5, 87
Eisenstein D. J. et al., 2005, ApJ, 633, 560
Fakhouri O., Ma C.-P., 2009, MNRAS, 394, 1825
Faltenbacher A., Diemand J., 2006, MNRAS, 369, 1698
Fillmore J. A., Goldreich P., 1984, ApJ, 281, 1
Gao L., White S. D. M., 2007, MNRAS, 377, L5
Gao L., White S. D. M., Jenkins A., Stoehr F., Springel V., 2004, MNRAS, 355, 819

Gao L., Springel V., White S. D. M., 2005, MNRAS, 363, L66
Garrison-Kimmel S. et al., 2017, MNRAS, 471, 1709
Gill S. P. D., Knebe A., Gibson B. K., 2004, MNRAS, 351, 399
Gil-Marín H., Jimenez R., Verde L., 2011, MNRAS, 414, 1207
González R. E., Padilla N. D., 2010, MNRAS, 407, 1449
Governato F. et al., 2009, MNRAS, 398, 312
Guedes J., Callegari S., Madau P., Mayer L., 2011, ApJ, 742, 76
Habib S. et al., 2016, New A, 42, 49
Hahn O., Porciani C., Carollo C. M., Dekel A., 2007, MNRAS, 375, 489
Harker G., Cole S., Helly J., Frenk C., Jenkins A., 2006, MNRAS, 367, 1039
He J.-h., 2019, preprint (arXiv:1905.01612)
Hearin A. P., Watson D. F., 2013, MNRAS, 435, 1313
Hearin A. P., Watson D. F., van den Bosch F. C., 2015, MNRAS, 452, 1958
Hearin A. P., Zentner A. R., van den Bosch F. C., Campbell D., Tollerud E., 2016, MNRAS, 460, 2552
Hellwing W. A., Schaller M., Frenk C. S., Theuns T., Schaye J., Bower R. G., Crain R. A., 2016, MNRAS, 461, L11
Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, MNRAS, 451, 2663
Hetznecker H., Burkert A., 2006, MNRAS, 370, 1905
Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, MNRAS, 445, 581
Jing Y. P., Suto Y., Mo H. J., 2007, ApJ, 657, 664
Kaiser N., 1984, ApJ, 284, L9
Kauffmann G., 2015, MNRAS, 454, 1840
Kauffmann G., White S. D. M., Guiderdoni B., 1993, MNRAS, 264, 201
Klypin A., Prada F., 2018, MNRAS, 478, 4602
Komatsu E. et al., 2011, ApJS, 192, 18
Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, ApJ, 609, 35
Kuhlen M., Diemand J., Madau P., 2007, ApJ, 671, 1135
Lacerna I., Rodríguez-Puebla A., Avila-Reese V., Hernández-Toledo H. M., 2014, ApJ, 788, 29
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Levi M. et al., 2013, preprint (arXiv:1308.0847)
Li Y., Mo H. J., Gao L., 2008, MNRAS, 389, 1419
Lovell M. R. et al., 2018, MNRAS, 481, 1950
Ludlow A. D., Navarro J. F., Springel V., Jenkins A., Frenk C. S., Helmi A., 2009, ApJ, 692, 931
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
Mao Y.-Y., Zentner A. R., Wechsler R. H., 2018, MNRAS, 474, 5143
Marinacci F. et al., 2018, MNRAS, 480, 5113
Matthee J., Schaye J., Crain R. A., Schaller M., Bower R., Theuns T., 2017, MNRAS, 465, 2381
McCullagh N., Norberg P., Cole S., Gonzalez-Perez V., Baugh C., Helly J., 2017, preprint (arXiv:1705.01988)
Miyatake H., More S., Takada M., Spergel D. N., Mandelbaum R., Rykoff E. S., Rozo E., 2016, Phys. Rev. Lett., 116, 041301
Montero-Dorta A. D. et al., 2017, ApJ, 848, L2
Mo H. J., White S. D. M., 1996, MNRAS, 282, 347
Mo H. J., Mao S., White S. D. M., 1998, MNRAS, 295, 319
Mo H. J., Mao S., White S. D. M., 1999, MNRAS, 304, 175
More S., Diemer B., Kravtsov A. V., 2015, ApJ, 810, 36
More S. et al., 2016, ApJ, 825, 39
Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121
Nagai D., Kravtsov A. V., 2005, ApJ, 618, 557
Naiman J. P. et al., 2018, MNRAS, 477, 1206
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Nelson D. et al., 2018, MNRAS, 475, 624
Nelson D. et al., 2019, Comput. Astrophys. Cosmol., 6, 2
Neto A. F. et al., 2007, MNRAS, 381, 1450
Niemiec A. et al., 2018, MNRAS, 477, L1
Nuza S. E. et al., 2013, MNRAS, 432, 743

O'Connell R., Eisenstein D., Vargas M., Ho S., Padmanabhan N., 2016, MNRAS, 462, 2681
Pakmor R., Springel V., 2013, MNRAS, 432, 176
Pakmor R., Bauer A., Springel V., 2011, MNRAS, 418, 1392
Pakmor R., Marinacci F., Springel V., 2014, ApJ, 783, L20
Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144
Peebles P. J. E., 1969, ApJ, 155, 393
Peirani S., Mohayaee R., de Freitas Pacheco J. A., 2004, MNRAS, 348, 921
Percival W. J. et al., 2014, MNRAS, 439, 2531
Pillepich A. et al., 2018a, MNRAS, 473, 4077
Pillepich A. et al., 2018b, MNRAS, 475, 648
Planck Collaboration XIII, 2016, A&A, 594, A13
Potter D., Stadel J., Teyssier R., 2017, Comput. Astrophys. Cosmol., 4, 2
Prada F., Klypin A. A., Cuesta A. J., Betancort-Rijo J. E., Primack J., 2012, MNRAS, 423, 3018
Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, ApJ, 771, 30
Richings J., 2018, preprint (arXiv:1811.12437)
Rodriguez-Gomez V. et al., 2015, MNRAS, 449, 49
Rodríguez-Puebla A., Primack J. R., Behroozi P., Faber S. M., 2016, MNRAS, 455, 2592
Samuel J. et al., 2019, MNRAS, 2638, preprint (arXiv:1904.11508)
Sawala T., Pihajoki P., Johansson P. H., Frenk C. S., Navarro J. F., Oman K. A., White S. D. M., 2017, MNRAS, 467, 4383
Schaye J. et al., 2015, MNRAS, 446, 521
Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
Sheth R. K., Tormen G., 2004, MNRAS, 350, 1385
Smith A., Cole S., Baugh C., Zheng Z., Angulo R., Norberg P., Zehavi I., 2017, MNRAS, 470, 4646
Somerville R. S., Primack J. R., 1999, MNRAS, 310, 1087
Sousbie T., Pichon C., Colombi S., Novikov D., Pogosyan D., 2008, MNRAS, 383, 1655
Spergel D. N. et al., 2003, ApJS, 148, 175
Springel V., 2010, MNRAS, 401, 791
Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726
Springel V. et al., 2005, Nature, 435, 629
Springel V. et al., 2008, MNRAS, 391, 1685
Springel V. et al., 2018, MNRAS, 475, 676
Tacchella S., Bose S., Conroy C., Eisenstein D. J., Johnson B. D., 2018, ApJ, 868, 92
Tasitsiomi A., Kravtsov A. V., Wechsler R. H., Primack J. R., 2004, ApJ, 614, 533
Tinker J. L., Conroy C., 2009, ApJ, 691, 633
Tinker J. L., Conroy C., Norberg P., Patiri S. G., Weinberg D. H., Warren M. S., 2008, ApJ, 686, 53
Trenti M., Smith B. D., Hallman E. J., Skillman S. W., Shull J. M., 2010, ApJ, 711, 1198
Trujillo-Gomez S., Klypin A., Primack J., Romanowsky A. J., 2011, ApJ, 742, 16
Vale A., Ostriker J. P., 2004, MNRAS, 353, 189
van Daalen M. P., Schaye J., 2015, MNRAS, 452, 2247
van Daalen M. P., Schaye J., McCarthy I. G., Booth C. M., Dalla Vecchia C., 2014, MNRAS, 440, 2997
Vitvitska M., Klypin A. A., Kravtsov A. V., Wechsler R. H., Primack J. R., Bullock J. S., 2002, ApJ, 581, 799
Vogelsberger M., Genel S., Sijacki D., Torrey P., Springel V., Hernquist L., 2013, MNRAS, 436, 3031
Vogelsberger M. et al., 2014a, MNRAS, 444, 1518
Vogelsberger M. et al., 2014b, Nature, 509, 177
Wang L., Weinmann S. M., De Lucia G., Yang X., 2013, MNRAS, 433, 515
Warren M. S., Quinn P. J., Salmon J. K., Zurek W. H., 1992, ApJ, 399, 405
Wechsler R. H., Tinker J. L., 2018, ARA&A, 56, 435
Wechsler R. H., Zentner A. R., Bullock J. S., Kravtsov A. V., Allgood B., 2006, ApJ, 652, 71
Weinberger R. et al., 2017, MNRAS, 465, 3291

Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., 2006, MNRAS, 366, 2

White S. D. M., 1984, ApJ, 286, 38

Yuan S., Eisenstein D. J., Garrison L. H., 2018, MNRAS, 478, 2019

Yurin D., Springel V., 2015, MNRAS, 452, 2367

Zehavi I. et al., 2005, ApJ, 630, 1

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

Zeldovich I. B., Einasto J., Shandarin S. F., 1982, Nature, 300, 407

Zentner A. R., 2007, Int. J. Mod. Phys. D, 16, 763

Zentner A. R., Berlind A. A., Bullock J. S., Kravtsov A. V., Wechsler R. H., 2005, ApJ, 624, 505

Zentner A. R., Hearin A. P., van den Bosch F. C., 2014, MNRAS, 443, 3044

Zheng Z. et al., 2005, ApJ, 633, 791

Zhu Q., Marinacci F., Maji M., Li Y., Springel V., Hernquist L., 2016, MNRAS, 458, 1559

Zjupa J., Springel V., 2017, MNRAS, 466, 1625

APPENDIX A: ENVIRONMENTAL ASSEMBLY BIAS AND THE DEFINITION OF A SATELLITE

In Fig. 5, we showed that while the HOD does indeed respond to environmental overdensity, the correlation is weak. In this Appendix, we display the relation between the HOD and halo environment for different choices of the boundary used to define the extent of a halo.

Fig. A1 shows the response of the HOD to halo environment (i.e. similar to the top row of Fig. 5) where the boundary of a halo is chosen to be r_{200c} (our fiducial choice, where the enclosed overdensity is equal to 200 times the critical density), r_{200m} (enclosed density is 200 times the mean density), and r_{FOF} (which simply selects the irregular edge defined by dark matter particles linked
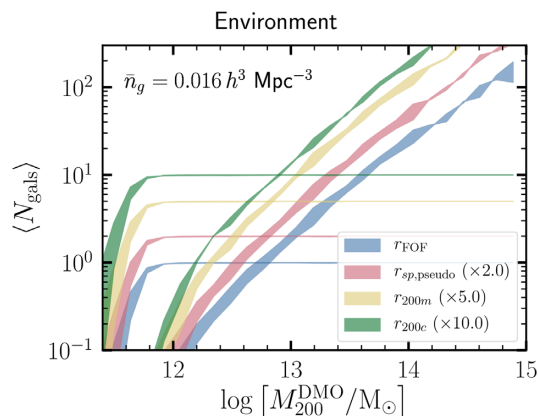
by the FOF algorithm). In general, $r_{200c} < r_{200m} < r_{FOF}$. A satellite galaxy in each catalogue is defined as a galaxy that is located within the corresponding definition of the host halo's boundary, ignoring the SUBFIND-based central versus satellite classifications. The quantity $\rho_r/\bar{\rho}$, which characterizes the environmental overdensity, is adjusted to be consistent with each definition (see Section 2.4 for details). The shaded regions in the figure encompass the variation in the mean occupancy of centrals and satellites for haloes located in the 20 per cent most overdense and 20 per cent most underdense environments – this is simply the area enclosed between the red and blue curves in Fig. 5. The curves corresponding to r_{200m} and r_{200c}, respectively, have been multiplied by factors of 5 and 10 to facilitate comparison.

It is clear from Fig. A1 that the extent to which environmental bias manifests in the HOD depends on the choice of halo boundary. This is true particularly in the case of the mean satellite occupation, which varies more strongly with environment for larger boundary definitions. This highlights the importance of a physically motivated choice for the boundary of a halo, such as the so-called 'splashback' radius (e.g. Fillmore & Goldreich 1984; Bertschinger 1985; Adhikari, Dalal & Chamberlain 2014; Diemer & Kravtsov 2014; More, Diemer & Kravtsov 2015). In Fig. A1, we use $r_{sp,pseudo} = 1.5\,r_{200m}$ as a rough proxy for this radius (Diemer & Kravtsov 2014). The variation of the mean central occupation, on the other hand, is similar for all boundary definitions.

APPENDIX B: NUMERICAL CONSIDERATIONS IN SATELLITE RADIAL PROFILES

In TNG, we find that the normalized radial distribution of galaxies, $n(r)/\langle n_{200}\rangle$, follows a universal profile that is described very well by the best-fitting NFW profile of their host haloes in the DMO simulation (Fig. 11). This Appendix establishes that this observation is independent of the mass range of satellite galaxies.

Fig. B1 repeats the calculation presented in Fig. 11 for different choices for the minimum mass of galaxies used to construct the
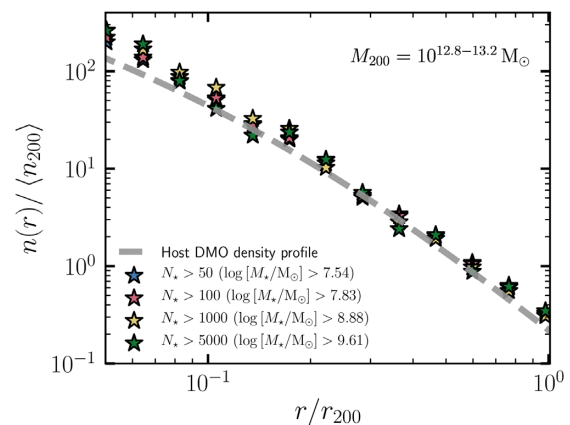

Figure A1. The effect of environment on the HOD for different choices of the maximal radius used to define a satellite galaxy. r_{200c} (r_{200m}) corresponds to the radius within which the mean density of the halo is 200 times the critical (mean) density of the universe; satellites of a halo are defined as the set of galaxies located within this boundary. r_{FOF} refers to the scenario in which a satellite is any galaxy located within the FOF group as a whole. The definition of halo environment for each choice of boundary is adjusted accordingly (see Section 2.4 for details). Finally, $r_{sp,\,pseudo}$ is a proxy for the splashback radius, which we have assumed to correspond to $1.5\,r_{200m}$. Here, we show the results for TNG300 only.

Figure B1. The radial number density profile of luminous satellites extracted from TNG100 for different choices of the minimum number of star particles, $N_\star$, in the galaxy at $z = 0$. The red stars ($N_\star > 100$) reproduce the result presented in the top right panel of Fig. 11; the grey curve is identical to orange curve displayed in that figure. Neither the shape nor the amplitude of the normalized profiles are affected strongly by the star particle threshold.

radial profile. For clarity, we show results solely for the host halo mass range $M_{200} = 10^{12.8-13.2}\,\mathrm{M}_\odot$, although our conclusions are unchanged for other host mass ranges. It is clear from this figure that the profiles for each mass cut is consistent with our fiducial choice of $N_\star > 100$. Reassuringly, both the amplitude and shape of the *normalized* profiles is consistent across all satellite mass ranges, although there is a small deviation between individual profiles towards the centre of the host halo. Fig. B1

suggests that the similarity between the radial profile of galaxies and the host dark matter profile is not biased by the dominance of a particular mass range of satellite galaxies in the radial profile.

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