Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **490,** 3532–3544 (2019) doi:10.1093/mnras/stz2790
Advance Access publication 2019 October 4

Extensions to the halo occupation distribution model for more accurate clustering predictions

Esteban Jiménez [1]★ Sergio Contreras [2]★ Nelson Padilla,[1,3] Idit Zehavi,[4] Carlton M. Baugh[5] and Violeta Gonzalez-Perez [6,7]

[1]*Instituto de Astrofísica, Pontificia Universidad Católica de Chile, Santiago 8970117, Chile*
[2]*Donostia International Physics Center (DIPC), Manuel Lardizabal pasealekua 4, Donostia E-20018, Basque Country, Spain*
[3]*Centro de Astro-Ingeniería, Pontificia Universidad Católica de Chile, Santiago 8970117, Chile*
[4]*Department of Physics, Case Western Reserve University, Cleveland, OH 44106, USA*
[5]*Department of Physics, Institute for Computational Cosmology, Durham University, South Road, Durham DH1 3LE, UK*
[6]*Institute of Cosmology & Gravitation, University of Portsmouth, Dennis Sciama Building, Portsmouth PO1 3FX, UK.*
[7]*Energy Lancaster, Lancaster University, Lancaster LA14YB, UK*

Accepted 2019 September 30. Received 2019 September 24; in original form 2019 June 10

ABSTRACT

We test different implementations of the halo occupation distribution (HOD) model to reconstruct the spatial distribution of galaxies as predicted by a version of the L-GALAXIES semi-analytical model (SAM). We compare the measured two-point correlation functions of the HOD mock catalogues and the SAM samples to quantify the fidelity of the reconstruction. We use fixed number density galaxy samples selected according to stellar mass or star formation rate (SFR). We develop three different schemes to populate haloes with galaxies with increasing complexity, considering the scatter of the satellite HOD as an additional parameter in the modelling. We modify the SAM output, removing assembly bias and using a standard Navarro–Frenk–White density profile for the satellite galaxies as the target to reproduce with our HOD mocks. We find that all models give similar reproductions of the two-halo contribution to the clustering signal, but there are differences in the one-halo term. In particular, the HOD mock reproductions work equally well using either the HOD of central and satellites separately or using a model that also accounts for whether or not the haloes contain a central galaxy. We find that the HOD scatter does not have an important impact on the clustering predictions for stellar mass-selected samples. For SFR selections, we obtain the most accurate results assuming a negative binomial distribution for the number of satellites in a halo. The scatter in the satellites HOD is a key consideration for HOD mock catalogues that mimic ELG or SFR-selected samples in future galaxy surveys.

Key words: galaxies: evolution – galaxies: formation – galaxies: haloes – galaxies: statistics – cosmology: theory – large-scale structure of universe.

1 INTRODUCTION

In the current cosmological paradigm, the Universe is composed of a filamentary network of structures shaped by gravity. In this framework, dark matter haloes correspond to overdense regions that evolve by gravitational instability due to mergers and interactions with other haloes. Galaxy formation occurs inside haloes where baryons collapse in the gravitational potentials and the condensation of cold gas allows the formation of stars and the evolution of galaxies (White & Rees 1978). A detailed description of the halo–galaxy connection enables us to use the galaxies to constrain the cosmological model.

The evolution of dark matter haloes can be followed, to high accuracy, using N-body simulations, which use a set of cosmological parameters as inputs. In contrast, the evolution of galaxies in haloes involves many physical processes that are still poorly understood. The fate of baryons within dark matter haloes has been modelled using different approaches. For example, hydrodynamical simulations provide an insight into the formation and evolution of galaxies for volumes with a comoving side of around 100 h^{-1}Mpc (e.g. Vogelsberger et al. 2014; Schaye et al. 2015). However, these models are computationally expensive and cannot be run over the large volumes needed for cosmological studies. Alternatively, the

★ E-mail: enjimenez@uc.cl (EJ); stcontre@uc.cl (SC)

effect of baryons can be probed in such large volumes using semi-analytical models (SAMs) of galaxy formation. These start from haloes extracted from a large volume dark matter only simulation and use simplified physical models of the processes that shape the evolution of baryons (Cole et al. 2000; Baugh 2006; Benson 2010; Somerville & Davé 2015). Hence, SAMs make predictions for the abundance and clustering of galaxies that can be compared and tested with large surveys.

Another way to describe the galaxy population is with the halo occupation distribution (HOD) framework (Benson et al. 2000; Peacock & Smith 2000; Scoccimarro et al. 2001; Yang, Mo & van den Bosch 2003). This is an empirical approach that provides a relation between the mass of haloes and the number of galaxies hosted by them. This is expressed as the probability distribution $P(N|M_{\rm h})$ that a halo of virial mass $M_{\rm h}$ hosts N galaxies, which satisfy some selection criteria. This approach provides insight into the halo–galaxy connection and can be used to study galaxy clustering (Berlind & Weinberg 2002; Zheng et al. 2005; Conroy, Wechsler & Kravtsov 2006; Zehavi et al. 2011; Wechsler & Tinker 2018). Furthermore, the HOD parameters can be tuned in detail because they only aim to reproduce a limited set of observables such as the galaxy number density and clustering. Thus, HOD modelling is one of the most efficient ways to populate very large volumes or to produce many realizations required, e.g. estimating covariance matrices using mock galaxy catalogues (e.g. Norberg et al. 2009; Manera et al. 2013). These mock catalogues can then be used to test and develop new algorithms that will be used for the next generation of surveys.

The study of star-forming emission line galaxies (ELGs) has gained interest over the last decade as they will be targeted by surveys such as *Euclid* and the Dark Energy Spectroscopic Instrument (DESI) surveys (Laureijs et al. 2011; DESI Collaboration 2016). The luminosity of an emission line depends on a number of factors, including the star formation rate (SFR), gas metallicity, and the conditions in the H II regions (e.g. Orsi et al. 2014). Even though ELGs samples are related to star formation, they are not the same as SFR-selected samples. Still, a similar HOD approach can be used to study both galaxy populations (Geach et al. 2012; Cochrane et al. 2017; Cochrane & Best 2018). In particular, the shape of the HOD in SFR-selected samples is more complex than the case of the more widely studied stellar mass-selected samples (e.g. Contreras et al. 2013; Gonzalez-Perez et al. 2018). For example, the occupation function of central galaxies in ELG samples does not follow the canonical step-like form. Accurate modelling of the HOD will provide the more realistic mock catalogues needed for the analysis of future observational samples.

Here, we use the HOD formalism to test three different ways to populate dark matter haloes with galaxies. The prescriptions of these models aim to replicate as accurately as possible the target galaxy populations of a SAM sample. The comparison between the galaxy population in the mock catalogues and SAM samples is done via the analysis of their two-point correlation function (2PCF), which is related to the power spectra of density fluctuations and is sensitive to cosmology (e.g. DeRose et al. 2019). We also include the scatter of the HOD of satellites in our modelling, and quantify the impact of using this additional parameter on the clustering.

The outline of this paper is as follows. The definition of galaxy samples used and the basic properties of the N-body simulation and the SAM are given in Section 2. The correlation functions and the HODs of the samples are presented in Section 3. In Section 4, we introduce the HOD models used to build the mock catalogues and the recipes employed to perform this procedure.

Table 1. The first column shows the abundance of galaxies in the three density samples used here. The second and third columns show the cuts applied to G13 galaxies in stellar mass and star formation rate, respectively, to achieve these abundances.

$n/h^3{\rm Mpc}^{-3}$	$M^*_{\rm min}/h^{-1}M_\odot$	$SFR_{\rm min}/{\rm yr}^{-1}M_\odot$
$10^{-3.0}$	5.95×10^{10}	5.25
$10^{-2.5}$	3.38×10^{10}	2.53
$10^{-2.0}$	1.25×10^{10}	0.70

The main results and analysis are discussed in Section 5, while in Section 6 we present our conclusions. Appendix A shows the predicted occupation functions for a particular HOD model.

2 SIMULATION DATA

In this section, we give a brief overview of the galaxy formation model used (Section 2.1) and the N-body simulation in which it is implemented (Section 2.2).

2.1 Galaxy formation model

A galaxy formation model needs to take into account a variety of physical processes such as radiative cooling of gas; AGN, supernovae, and photoionization feedback; chemical evolution; star formation; disc instabilities; collapse and merging of dark matter haloes; and galaxy mergers. These affect the fate of baryons in haloes, which lead to the formation and evolution of galaxies. Several physical processes such as star formation and gas cooling are not fully understood due to their complexity. As a consequence, a set of free parameters are used in the equations that model these processes. These free parameters are tuned in order to reproduce observations such as the luminosity functions, colours, and the distribution of morphological types. In this context, different SAMs usually have their own implementations to model these physical processes, predicting different galaxy populations. Here, we use the outputs at $z = 0$ from the SAM of Guo et al. (2013) (hereafter G13), which is a version of the L-GALAXIES code from the Munich group (De Lucia, Kauffmann & White 2004; Croton et al. 2006; De Lucia & Blaizot 2007; Guo et al. 2011; Henriques et al. 2013). The outputs are publicly available from the Millennium Archive.[1]

The samples used here are defined according to three different number densities where we rank the galaxies in the SAM by their stellar mass or SFR in a decreasing way (hereafter, the SAM samples). These samples are useful in order to compare with observational catalogues with similar space densities. Table 1 shows the three number densities and the cuts in stellar mass and SFR used in each case.

2.2 The Millennium simulation

The distribution of dark matter haloes used in this work is drawn from the Millennium-WMAP7 simulation (Guo et al. 2013), which is identical to the Millennium Simulation (Springel et al. 2005), but with updated cosmological parameters that match the results from the WMAP7 observations. This version assumes a flat ΛCDM universe considering $\Omega_m = 0.27$, $\Omega_\Lambda = 0.73$, $h = H_0/(100\ {\rm km\ s^{-1}\ Mpc^{-1}}) = 0.704$, and $\sigma_8 = 0.81$. The simulation

[1] http://gavo.mpa-garching.mpg.de/Millennium/

was carried out in a box size of 500 h^{-1}Mpc following 2160^3 particles of mass $9.31 \times 10^8 h^{-1}\mathrm{M}_\odot$. The run produced 61 simulation snapshots from $z = 50$ up to $z = 0$. G13 use a friends-of-friends group finding algorithm (FOF) to identify dark matter haloes in each snapshot (Davis et al. 1985) and then run SUBFIND to identify the subhaloes (Springel et al. 2001). Halo merger trees are constructed for each output and track the evolution of haloes through cosmic time. These trees are the starting points for the SAM.

3 CHARACTERIZATION OF THE SAM GALAXY SAMPLES

This section introduces the statistics used to characterize the distribution of galaxies, starting with the measurement of the correlation function (Section 3.1), the form of the HOD predicted by the SAM (Section 3.2) and the scatter in the HOD (Section 3.3).

3.1 Clustering measurement: two-point galaxy correlation function

The spatial 2PCF, $\xi(r)$, measures the excess probability of finding a pair of galaxies at a given separation with respect to a random distribution. We compute the 2PCF of the galaxy samples with the Corrfunc code (Sinha & Garrison 2017).

Fig. 1 shows the 2PCF of the stellar mass (top) and SFR (bottom) selected samples for the different space densities. For the former, the amplitude of the clustering increases with decreasing number density, as we consider more massive galaxies. The impact of the inclusion of these massive galaxies is stronger at small scales and is weaker at large scales. In contrast, for the SFR-selected galaxies the amplitude of the 2PCF for the different samples remains largely unchanged except for small scales ($\sim 0.1\,h^{-1}$Mpc), where the satellite–satellite pairs make an important contribution to the clustering amplitude. For both selections, the satellite fraction increases with increasing number density.

In the 2PCF, we can distinguish between the contribution from galaxy pairs in the same halo and from different haloes. The former are the main contributors to the amplitude of the 2PCF on small scales, namely the one-halo term that dominates up to $\sim 1\,h^{-1}$Mpc, while galaxy pairs between different haloes contribute mostly to the two-halo term that determines the clustering on large scales. In this regime, the total number of galaxies in the halo, regardless of whether they are satellites or the central, drives the amplitude of the clustering, acting as a weighting for the bias of each halo in computing an overall 'effective' bias for the sample (see e.g. Baugh et al. 1999). The one-halo term is sensitive not only to the number of satellites, but also to their spatial distribution.

3.2 The halo occupation function predicted by the SAMs

The galaxy populations in the SAMs depend on the choices adopted for the modelling of the baryonic processes. Hence, depending on the SAM employed, different galaxy catalogues with different luminosity functions, stellar mass functions, or correlation functions can be obtained for the same dark matter simulation. For example, Contreras et al. (2013) studied the effects on the clustering predicted from different SAMs and found some differences particularly in galaxy samples selected by SFR and cold gas mass. Moreover, they show that the shapes of the HODs are model-dependent, which reflects the differences in the implementation of physical processes in each SAM. For example, the specific modelling of dynamical friction affects the satellite population in SAMs. Here, we are not interested in the detailed shape of the HOD predicted by a particular SAM, but on how best to use the occupation functions to populate dark matter haloes with galaxies to produce a similar spatial distribution to that resulting from a SAM.

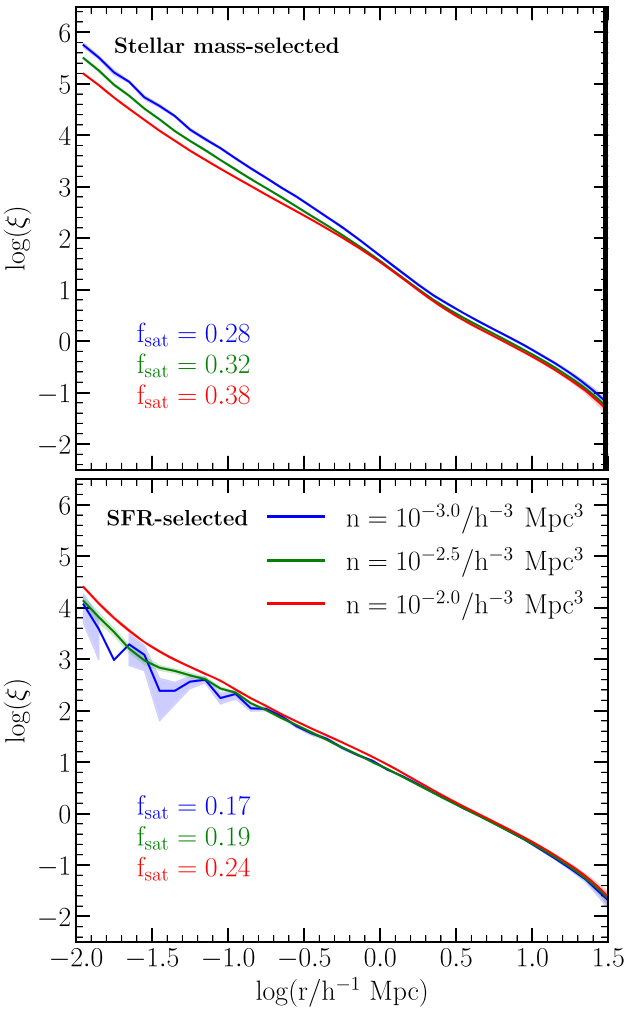

Figure 1. Two-point correlation functions ($\xi(r)$) of different galaxy samples from G13 and defined in Table 1. (*Top*) stellar mass and SFR-selected samples (*bottom*). Colours indicate each sample as labelled in the bottom panel. The fraction of satellites in each sample is shown in both panels, with the colour indicating the sample number density. The shaded regions represent jackknife errors calculated using 10 subsamples.

The HOD is usually broken down into the contribution from central and satellite galaxies. Fig. 2 shows these two components for stellar mass and SFR-selected samples with the same number density for the G13 SAM. Here, each HOD is computed in bins of width 0.08 dex in the logarithm of the halo mass where the position of each $\langle N \rangle$ value is plotted at the median value within each bin. The striking difference in the shape of the HOD of centrals between the two selections is due to the different galaxies that are included. Massive centrals tend to be red galaxies hosted by massive dark haloes. Such centrals are included in stellar mass-selected samples but not when selecting by SFR. The galaxies in the SFR samples correspond mainly to blue star-forming galaxies excluding luminous red galaxies with high stellar mass but low SFR. It is noteworthy that the fraction of haloes that contain a central passing the SFR selection never reaches unity for the sample plotted in

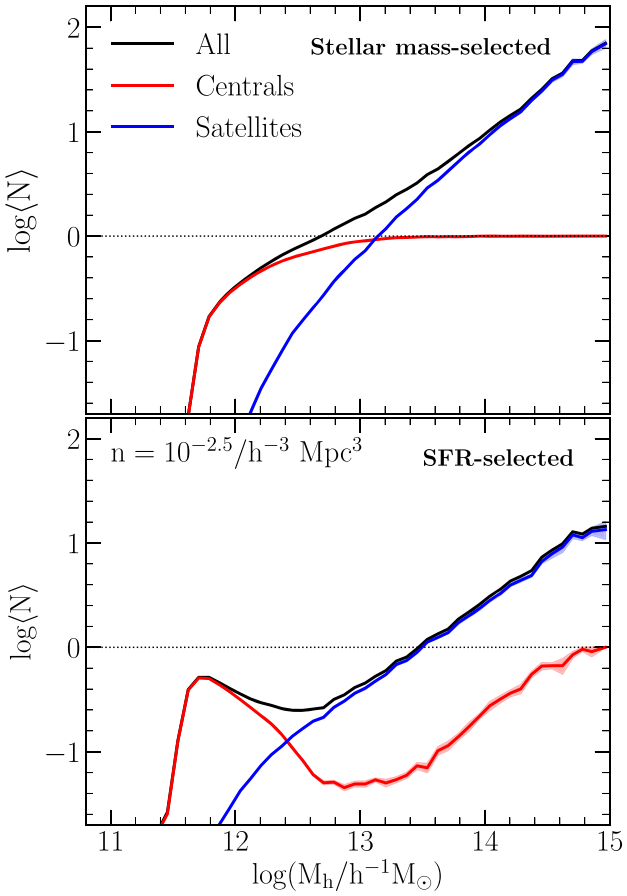

Figure 2. The HOD predicted by G13 for stellar mass (*top*) and SFR-selected samples (*bottom*), for a number density of $10^{-2.5}h^3\mathrm{Mpc}^{-3}$. The black lines show the HOD for the full sample, and the red and blue lines indicate the HOD for central and satellite galaxies, respectively. The horizontal black dotted line shows an average occupation value of unity.

Fig. 2. These features of the HOD of SFR galaxies have been noted in SAMs before (e.g. Contreras et al. 2013, 2019; Gonzalez-Perez et al. 2018) and inferred for blue galaxies in the SDSS (e.g. Zehavi et al. 2011). This shows that a significant number of haloes in the SFR-selected samples do not host a central that is sufficiently highly star forming to meet the SFR threshold. The same situation is found in the other number density samples. Note that, in observational samples, the ranking of galaxies in order of their emission line luminosity may not correspond to the ranking in SFR due to dust attenuation, which means that the highest SFR galaxies may not necessarily have the brightest emission lines (e.g. Hicks et al. 2002; Ly et al. 2012). Moreover, dust attenuation is more significant for very massive galaxies (Sobral et al. 2016), which may include a fraction of ELGs.

We estimate the uncertainties of the HOD values and clustering measurements using jackknife resampling (Norberg et al. 2009), dividing the simulation volume into 10 slices. We use the position of the centre of the potential of haloes to classify the galaxies within each halo. For the HOD values in Fig. 2, errors, shown as the shaded areas, are negligible for all halo masses except at the high-mass end and for the HOD of centrals selected by SFR.

Because of the simple relation between halo mass and occupation number, the HOD represents a useful approach for the construction of mock galaxy catalogues. Here, we have described the first moment of the HOD, the main ingredient of mock-building recipes. Nevertheless, it is important to also consider the second moment, i.e. the dispersion in the HOD of satellites.

3.3 The predicted dispersion in the halo occupation number

When the simplest HOD approach is used to build mock catalogues, the mean of the distribution is the main parameter. Central galaxies are assumed to follow a nearest integer distribution where the mean $\langle \mathrm{N_{cen}} \rangle$ is between zero and one. For satellites, a Poisson distribution with mean $\langle \mathrm{N_{sat}} \rangle$ is the most widely assumed distribution (e.g. Kravtsov et al. 2004; Zheng et al. 2005).

In G13, satellites are classified as type-1 if they are hosted by a resolved subhalo, and type-2 or orphans if the subhalo has been destroyed by tidal effects and is no longer identified. Boylan-Kolchin et al. (2010) found that the number of low-mass subhaloes in main haloes in the Millennium-II Simulation (Boylan-Kolchin et al. 2009) is well described by a negative binomial distribution, which corresponds to a super-Poissonian statistic as its scatter is larger than a Poisson distribution. This suggests that the type-1 satellite population can also be described by this distribution. Based on the outputs from the SAM presented in Jiang & van den Bosch (2016), and using the Bolshoi (Klypin, Trujillo-Gomez & Primack 2011) and MultiDark (Prada et al. 2012) simulations, Jiang & van den Bosch (2017) showed that ignoring this non-Poissonity in the HOD of subhaloes results in systematic errors in the predicted clustering of galaxies. Here, we extend the application of the negative binomial distribution by checking whether the HOD of G13 satellites, which is including type-1 and type-2, is well described by this statistic. We expect that the HOD scatter is model dependent because of the different treatments of dynamical friction. Moreover, as some galaxy properties, such as SFR and stellar mass, have a model-dependent scatter it is reasonable to assume the same for HODs. For example, Guo et al. (2016) showed that different galaxy formation models do not have the same dispersion in the stellar mass–halo mass relation. Therefore, our results are specific to the G13 model. It is likely that a different SAM would require an adjustment to the value of β (defined next) to describe the scatter of the satellite HOD. Nevertheless, we expect our general results to hold for any SAM applied in the Millennium-WMAP7 simulation, and that the satellite distribution displays more scatter than Poisson.

The Poisson and negative binomial distributions differ in their shapes, so it is useful to parametrize the departure from the Poisson scatter. We use the parameter β to denote this departure. For a Poisson distribution, the variance is given by the mean value of the random variable, namely $\langle \mathrm{N_{sat}} \rangle$, with the standard deviation given by $\sigma = \sqrt{\langle \mathrm{N_{sat}} \rangle}$. The negative binomial distribution has the same mean as the Poisson distribution, but a larger scatter that can be expressed as

$$\sigma_{\mathrm{NB}} = \sigma + \beta\sigma, \tag{1}$$

where $0 < \beta < 1$. Then, β indicates the fractional change in the variance with respect to the Poisson standard deviation σ. Under this definition, when $\beta = 0$ the distribution is Poissonian and if $\beta = 1$ the standard deviation is twice that from a Poisson distribution.

The probability function of the negative binomial distribution is given by

$$P(N\,|r,\,p) = \frac{\Gamma(N+r)}{\Gamma(r)\Gamma(N+1)}p^r(1-p)^N. \tag{2}$$

Here, $\Gamma(x) = (x+1)!$ is the gamma function. The parameters r and p are determined by the first moment $\langle N \rangle$ and second moment σ^2

of the distribution:

$$p = \frac{\langle N \rangle}{\sigma^2_{\rm NB}}, \quad r = \frac{\langle N \rangle^2}{\sigma^2_{\rm NB} - \langle N \rangle}. \tag{3}$$

Thus, we can control the width of the negative binomial distribution through the parameter β and compute the value of $\sigma^2_{\rm NB}$.

4 GENERATING HOD MOCK CATALOGUES

We now describe the procedure followed to build HOD mock galaxy catalogues using the HODs of the SAM samples. Section 4.1 presents the three methods we use to populate haloes with galaxies. In Section 4.2, we specify the treatment of the scatter in the HOD of satellites. Section 4.3 explains how we impose a standard Navarro–Frenk–White (NFW) density profile for satellites. Section 4.4 presents the impact of assembly bias in the SAM samples and explains why it must be removed from the SAM in order to compare with the HOD mock catalogues. Finally, in Section 4.5 we discuss the treatment of the radial distribution of satellite galaxies within haloes.

4.1 The HOD models used to build mocks

We test three different HOD schemes of increasing complexity. This helps us to understand the level of complexity needed to obtain accurate clustering predictions. Each model uses occupation functions obtained from linear interpolations of the HOD values in each halo mass bin, as in Fig. 2. The distribution of galaxies can be nearest integer (centrals only) and Poisson or negative binomial (satellites).

4.1.1 1-HOD

The 1-HOD model builds mock catalogues using the HOD of all galaxies from the SAM sample (the black solid lines in Fig. 2) including both centrals and satellites. The model assumes either a Poisson or negative binomial distribution for the occupation number. We adopt a Monte Carlo approach to obtain the final number of galaxies.

This approach does not distinguish between centrals and satellites. If the model predicts that $\rm N \geq 1$, we assume that this halo hosts a central and $\rm N_{sat} = N - 1$. Because of this, the number of centrals and satellites in the 1-HOD mock catalogues can be notably different with respect to the SAM samples where there are haloes with satellites but no central. Moreover, the HODs of these two separate components in the mock catalogues are completely different with respect to the HODs of the SAM samples (see Appendix A). Even though the total number of galaxies in these mock catalogues is essentially the same as in the SAM samples, the 1-HOD removes information about the central/satellite nature of galaxies.

The 1-HOD does not reproduce the galaxy populations in the SAM samples. However, we also compute the clustering of the 1-HOD mocks to emphasize the differences in clustering predictions when using more complex HOD modelling, like the ones shown next.

4.1.2 2-HOD

The 2-HOD model uses the HOD of centrals and satellites separately, i.e. the red and blue solid lines in Fig. 2, respectively. Thus, a particular distribution can be assumed for each component and the modelling is done independently for each one. For centrals, we assume the nearest integer distribution and for satellites we use the Poisson and negative binomial distributions defined by different β parameters starting from $\beta = 0$ (see Section 4.2 to see how we choose the best β).

By construction, this scheme reproduces practically the same number of central and satellites as the SAM samples. Note that in a non-negligible number of realizations it is possible to get haloes without a central, as is found in SAM samples when cuts are applied to the original SAM catalogue. This is more likely for haloes with masses for which $\langle \rm N_{cen} \rangle < 1$, which is more frequently found in SFR-selected samples. The 2-HOD model recovers haloes with satellites but no centrals, as in the SAM samples.

4.1.3 4-HOD

This model contains more information about the galaxy population of the SAM sample than the 2-HOD model. The 4-HOD requires us to store the number of haloes that host a central ($\rm N_{cen}$) and the number of haloes that do not host a central ($\rm N_{nocen}$) as a function of halo mass. Under this definition, the total number of haloes in the volume is the sum of both quantities. Furthermore, the 4-HOD also needs knowledge of the number of satellites in haloes with a central ($\rm N_{sat\_cen}$) and without a central ($\rm N_{sat\_nocen}$). Thus, the total number of satellites is the sum of these two quantities. With these definitions, we build new HODs for satellites that take into account the population of centrals in the SAM samples. The SAM samples contain haloes with satellites but no centrals. This is more common in SFR-selected samples. Indeed, the HOD of centrals in these samples indicates that a large number of haloes do not host a central (see Fig. 2), and the 4-HOD takes this feature into account.

We then define the satellite occupation functions conditioned on whether or not haloes host a central. With the four quantities explained above, we can define the conditional HODs:

$$\langle \rm N_{sat\_cen}(\mathit{M}_h) \rangle = \frac{\rm N_{sat\_cen}}{\rm N_{cen}}(\mathit{M}_h), \tag{4}$$

$$\langle \rm N_{sat\_nocen}(\mathit{M}_h) \rangle = \frac{\rm N_{sat\_nocen}}{\rm N_{nocen}}(\mathit{M}_h). \tag{5}$$

Fig. 3 shows the conditional HODs where the main differences are observed at low halo masses. It can be seen that the ratio is equal to 1 for high halo masses ($M_h \gtrsim 10^{13} h^{-1}\rm M_\odot$). Similar trends are found for the other density samples. We suspect that quenching processes in satellites are less important than in centrals because whether or not the haloes contain a central, they host, on average, the same number of satellites. Even though the ratio between these two HODs is close to unity, it is the galaxies hosted by low mass haloes ($\approx 10^{12} h^{-1}\rm M_\odot$) that dominate the amplitude of clustering.

The conditional HODs are well fitted by a negative binomial distribution, including the HODs of the other number density samples. The 4-HOD method uses a Monte Carlo approach to decide if a halo hosts a central galaxy. Depending on this outcome, one of the two conditional HODs is then chosen to obtain the number of satellites.

4.2 Treatment of scatter in the HOD of satellites

A Poisson distribution is fully described by its first moment. In the case of satellites this is $\langle \rm N_{sat} \rangle$. If the distribution of the number of satellites follows instead a negative binomial distribution, an additional parameter β is needed that specifies the increase in the scatter with respect to a Poisson distribution (see equation 1). We fix the β value so that we reproduce as closely as possible the scatter

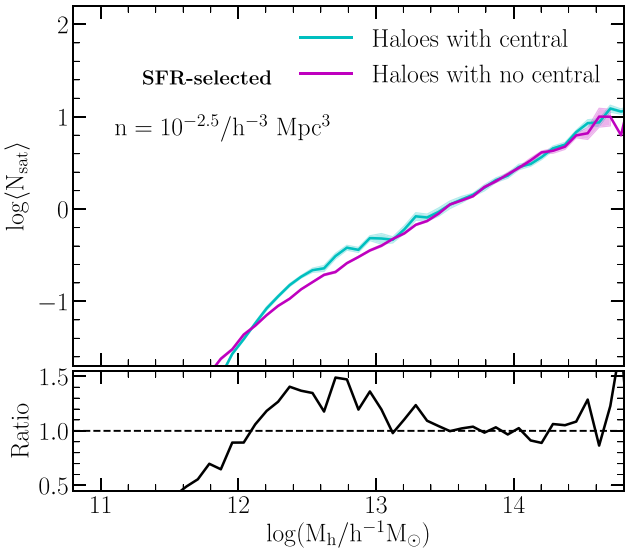

Figure 3. Conditional HODs from the 4-HOD method for an SFR-selected sample with number density $10^{-2.5}h^3\,\mathrm{Mpc}^{-3}$. *Top:* Average number of satellites in haloes with a central (cyan) and without a central (magenta). *Bottom:* Ratio of the two HODs shown in the upper panel. The shaded regions represent jackknife errors calculated using 10 subsamples.

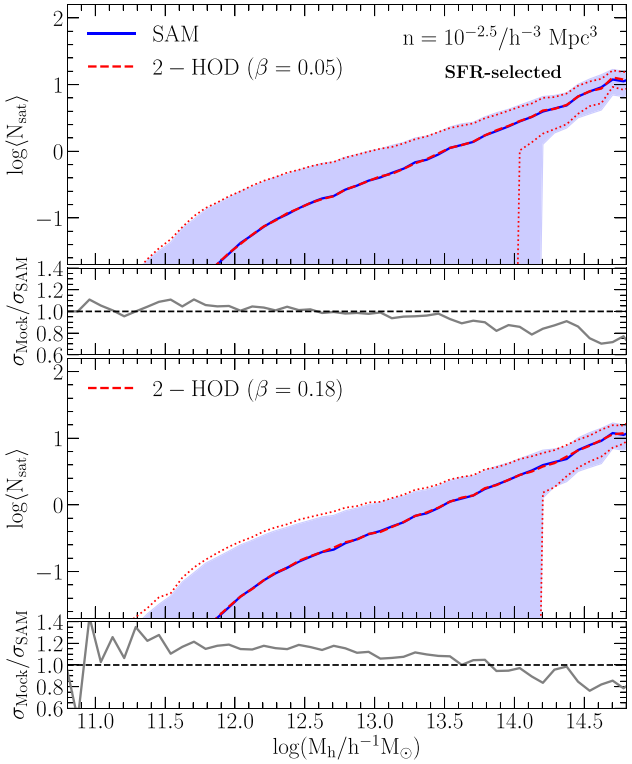

Figure 4. The HOD of satellites in a SAM sample (dashed blue) and a 2-HOD mock catalogue (solid red) contrasting two values of the parameter β that controls the scatter (see equation 1): $\beta = 0.05$ (*top*) and $\beta = 0.18$ (*bottom*). The shaded regions show the HOD scatter and the red dotted lines correspond to the scatter in the HOD mocks. The subpanels show the ratios between the HOD scatter of the mocks and SAM sample. Note that it is not possible to visually distinguish a Poisson scatter from the β scaled versions plotted in the main panels, but this choice would lead to a larger ratio of variances than the range plotted in the lower subpanels.

of the HOD of satellites in a given SAM sample. Fig. 4 shows the scatter of the HOD of satellites in SAMs and 2-HOD mock catalogues for two illustrative β values in an SFR-selected sample. This shows that a small but non-zero β is required to reproduce the HOD scatter of the SAM sample. The same is found for the other number density samples, and for the conditional HODs. Note that when using a larger β value, as shown in the bottom panel of Fig. 4, the HOD scatter is overpredicted for a wide range of halo masses. We do not perform this analysis for the 1-HOD model as satellites are not treated independently in this case.

It is not possible to replicate the HOD scatter in SAMs more closely as this would require β to be a function of mass. Instead, we assume a constant scatter for the HOD of satellites using the same β for all halo mass bins. The accuracy of the β values used is judged by checking the quality of the resulting mocks via comparison of their 2PCFs with the clustering of the shuffled-NFW samples (see Section 4.5 below for the definition of this catalogue). We show in Section 5 that when the scatter of the SAM and HOD mocks are matched up to $M_\mathrm{h} \lesssim 10^{13.5}h^{-1}\mathrm{M}_\odot$ ($\beta = 0.05$ for SFR-selected samples), we obtain the most accurate clustering predictions. In contrast, using larger values for β worsens the predictions (as does using $\beta = 0$, which corresponds to Poisson scatter).

The satellite HOD is well described by the negative binomial distribution for a wide range of halo masses. Fig. 5 shows the satellite PDF in a particular mass bin for a stellar mass and an SFR-selected sample. We show negative binomial distributions defined by $\beta = 0.08$ and $\beta = 0.05$. In order to compute the satellite distributions, we split satellites according to whether or not their haloes host a central galaxy, which is relevant for the 4-HOD model. The satellite distribution matches with the negative binomial when most of haloes in the bin are included. A similar close match is found when comparing with Poisson distributions ($\beta = 0$). Note that in the SFR-selection case most of the haloes do not host a central galaxy, as the HOD of centrals in that bin suggests. The opposite behaviour is observed when selecting by stellar mass.

4.3 The radial distribution of satellite galaxies in haloes

The number of satellites in the mock catalogues is obtained from the adopted HOD model (see Sec 4.1). Their positions in haloes are set according the standard NFW density profile (Navarro, Frenk & White 1996), which requires two parameters, the concentration and scale radius. The former depends on halo mass and the latter is a function of the virial radius. For simplicity, we assume that all haloes in the simulation volume have the same concentration parameter $c = 13.98$, which corresponds to the concentration of a halo at redshift $z = 0$ and mass $M_\mathrm{h} = 10^{12.5}h^{-1}\mathrm{M}_\odot$. We do not use a more realistic model for concentration as we are interested in comparing the HOD models rather than obtain a realistic redistribution of satellites. We impose that the maximum distance from a satellite to the halo centre is two virial radii that depends on the halo mass. This defines the NFW mass profile used to obtain the satellite distances by a Monte Carlo approach. We modify the SAM output to impose a similar satellite distribution as described next (Section 4.5).

4.4 Removing assembly bias from the SAM output

The clustering of dark matter haloes depends on additional properties besides mass (e.g. Gao, Springel & White 2005; Wechsler et al. 2006; Gao & White 2007; Lacerna & Padilla 2011) . For example, Gao et al. (2005) showed that the clustering of low-mass haloes depends on their formation redshift and other works have found

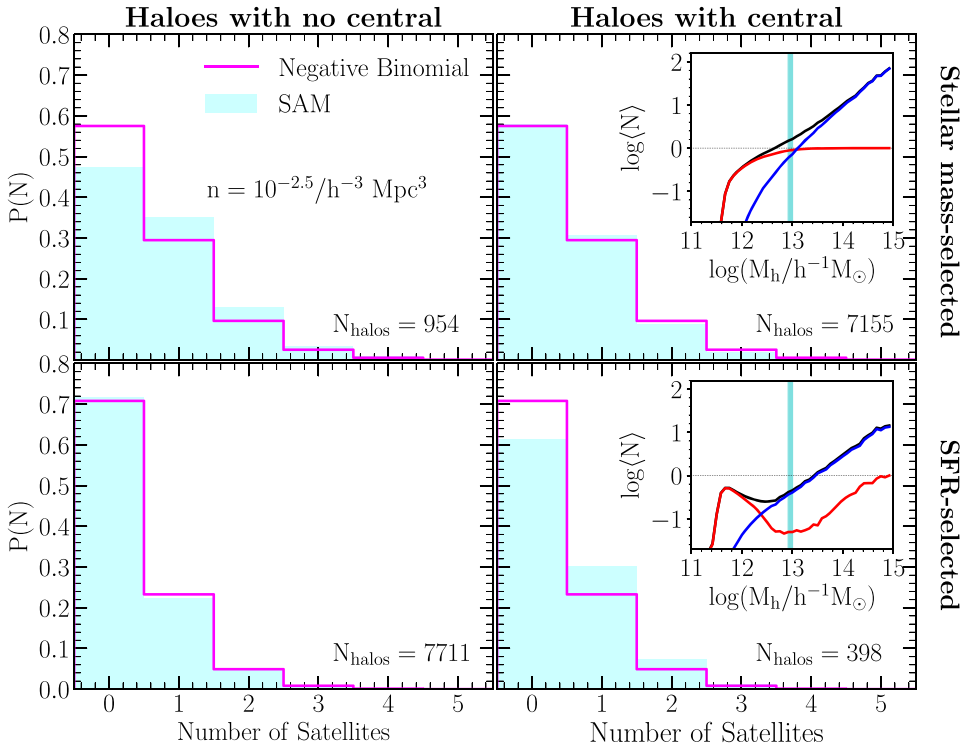

Figure 5. Probability distributions of satellites (the cyan histograms) that are hosted by haloes with masses within the blue shaded (vertical) mass range of the HODs shown in the insets. The galaxies are selected by stellar mass (*top*) and SFR (*bottom*), with a number density of $10^{-2.5}\,h^3\,\mathrm{Mpc}^{-3}$. *Left:* Haloes without centrals (i.e. its central did not meet the stellar mass/SFR-selection criteria). *Right:* Same as left-hand panel but considering haloes with centrals. Note the high probability of finding haloes without centrals meeting the SFR cut, which is expected by the low value of $\langle N_{cen}\rangle$ in the mass range analysed and as is shown in the inset. The distributions are well described by negative binomial distributions (magenta) in the cases when most of haloes in the bin are included (i.e. top right-hand panel and bottom left-hand panel). The negative binomial distributions shown here are obtained using a scatter that is 5 per cent larger than that from a Poisson distribution in the SFR-selected sample, and 8 per cent larger in the stellar mass selection case.

dependencies on concentration and subhalo occupation number (e.g. Wechsler et al. 2006) among other secondary properties. This additional contribution to the halo clustering is commonly known as halo assembly bias and potentially affect the galaxy clustering amplitude on large scales.

The dependence of the HOD on secondary halo properties can also have an impact on the galaxy content of haloes (see Artale et al. 2018; Zehavi et al. 2018; Contreras et al. 2019). For example, using a SAM, Contreras et al. (2019) show that stellar mass-selected galaxies hosted by the most early-formed haloes have a larger clustering amplitude than galaxies in late-forming haloes. This is caused by the occupancy variation between haloes of the same mass but different secondary properties as formation time or environment. The combined effect from halo assembly bias and occupancy variation produces a total impact on galaxy clustering at large scales called galaxy assembly bias.

We use the standard HOD approach that considers only halo mass as the variable regulating the galaxy population. Note that there are some works extending this approach by considering additional halo properties to account for assembly bias (e.g. Hearin, Behroozi & van den Bosch 2016). SAMs include assembly bias because they follow the evolution of baryons in halo merger histories that are shaped by the large-scale environment in the N-body simulation. Namely, SAMs include a dependence on secondary halo properties as these affect the halo merger history and the evolution of galaxies that live within them. Thus, in order to compare the clustering between SAM samples and HOD mocks that use only halo mass as input, it is necessary to remove the galaxy assembly bias signal in the former samples. From this comparison, we can determine the best methodology to produce HOD mock catalogues.

Galaxy assembly bias can be eliminated from SAM samples through the shuffling technique (see Yoo et al. 2006; Croton, Gao & White 2007). This consists of randomly exchanging the galaxy populations between haloes of the same mass, thus removing any connection to the assembly history of the haloes. This procedure does not change the distances from satellites to their central galaxy in each halo. In clustering terms, the one-halo term of this 'shuffled' catalogue is the same as the original SAM sample but its two-halo term is different because assembly bias is not present in the shuffled samples. If the SAM samples did not have assembly bias, we would measure the same 2PCFs for their shuffled samples as measured for the original output.

Note that there are other effects within haloes like galactic conformity, i.e. the correlation of specific SFRs, gas content, colours, and morphologies between centrals and satellites (e.g. Weinmann et al. 2006; Hearin et al. 2016; Lacerna et al. 2018). To study this effect, and other features like satellite alignments, one should consider alternative shuffling schemes where satellites are moved to haloes of the same mass independent of their centrals (e.g. Zu et al. 2008; Zentner, Hearin & van den Bosch 2014). For our purposes, shuffling the satellites together with the centrals suffices to remove the galaxy assembly bias signature in the SAM samples.

Fig. 6 shows the correlation functions of a SAM sample and its shuffled version, for both the stellar mass and the SFR-selected

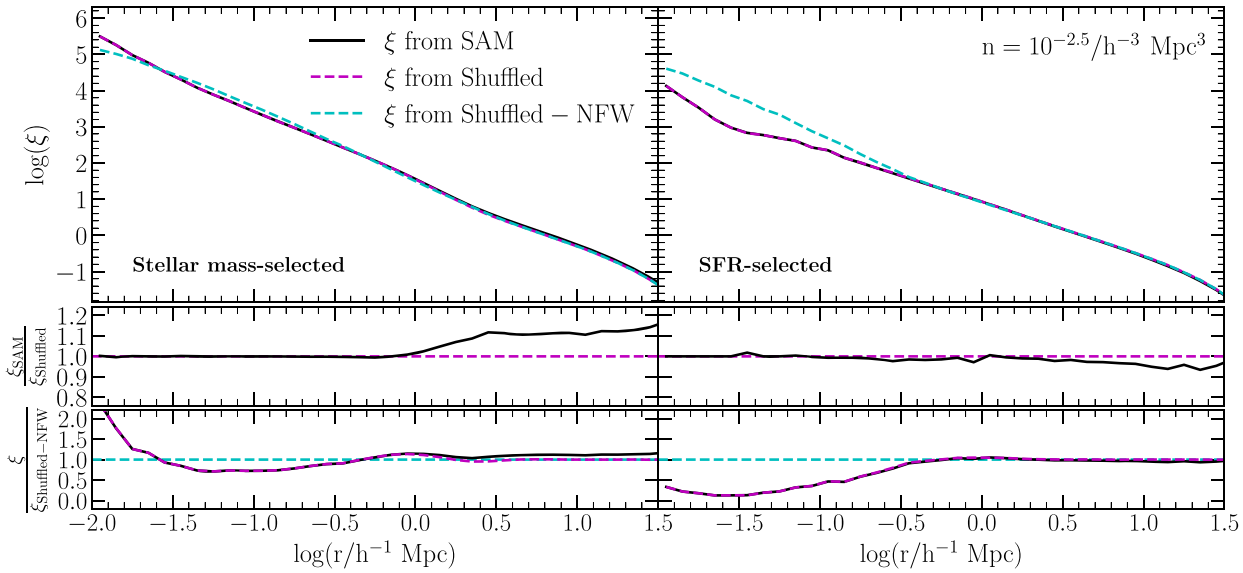

Figure 6. *Top:* Correlation functions of SAM samples (dotted black in the main panel, solid black in the subpanel), and their modifications: the shuffled (dashed magenta) and the shuffled-NFW samples (dashed cyan). The galaxy samples are selected by stellar mass (*left*) and SFR (*right*), with a number density of $10^{-2.5}\,h^3$ Mpc^{-3}. *Middle:* Ratios between the 2PCF of the SAM and shuffled samples. Differences at large scales are signatures of assembly bias. Note that, by construction, the ratio at small scales is equal to 1. *Bottom:* Ratios of the 2PCF with respect to the 2PCF from the shuffled-NFW samples. The differences in the one-halo term below 1 h^{-1}Mpc indicates the departure of the satellite profiles from a NFW.

samples. The assembly bias signature, shown in the middle panels, is evident in the clustering differences between these two catalogues at large separations. We also show the 2PCF of a modified shuffled sample that will be introduced next. The assembly bias signatures remain unchanged for the other samples, but they are noisier for the lowest number density samples as they contain fewer galaxies.

It can be seen that assembly bias increases the clustering for stellar mass-selected samples. Zehavi et al. (2018) explain that this effect is a consequence of the occupancy variation of centrals that tend to be hosted by early-formed haloes, which have a larger clustering amplitude. SFR-selected samples, on the other hand, show a decreased clustering amplitude. For the intermediate galaxy density sample, the assembly bias enhances the two-halo term by ~ 12 per cent for the stellar mass-selected sample and suppresses the amplitude in the SFR selection case by ~ 4 per cent. The enhancement of clustering amplitude for the other stellar mass selections remains similar. For the SFR selections, we observe that suppression of clustering becomes weaker for higher density samples. Indeed, assembly bias can change the sign of the effect, where galaxies in the original samples are less clustered than in their shuffled versions. We see this effect for the highest density sample, which is consistent with the results from Contreras et al. (2019). They found that depending on the redshift and the number density, the galaxy assembly bias can be weaker or even change the sign of its effect.

4.5 The shuffled-NFW target catalogue: changing the satellite distribution in the SAM

The radial profile of satellites in the G13 SAM deviates from the standard NFW profile of dark matter within haloes because the SAM associates galaxies with subhaloes (or a proxy, such as the most bound particle, in the case of subhaloes that are no longer resolved). The radial profile of subhaloes is different from that of the dark matter (see e.g. Angulo et al. 2009). The choice of which subhaloes (and former subhaloes) are associated with galaxies is driven by the galaxy formation model, which determines the luminosity of any galaxy associated with a subhalo and whether or not it has merged due to dynamical friction (only type 2 satellites, those that no longer have a resolved subhalo associated with them, are considered as candidates for galaxy mergers).

The final step before testing the accuracy of the HOD models is to modify the shuffled SAM catalogue to force the satellites in each halo to follow an NFW profile. We call the result the shuffled-NFW catalogue. Because satellite galaxies in the SAMs and shuffled samples do not follow an NFW profile, the one-halo term of their 2PCFs are different to the one-halo term of the shuffled-NFW sample, as shown in the bottom panel of Fig. 6. The shuffled-NFW catalogue does not contain assembly bias, and the satellites follow the *same* NFW profile as adopted in the HOD mocks. We note that the shuffled-NFW is not intended to be the 'best' prediction of galaxy clustering. Instead, it is the target sample for the HOD mocks that has a controlled one-halo clustering pattern to facilitate testing.

To examine the departure from an NFW profile in the SAM, we also produce a SAM–NFW catalogue where satellites in the SAM are forced to follow the same NFW profile as used in the HOD mocks. For this catalogue, we update the satellite positions in the SAM samples according the same NFW density profile used to produce the HOD mocks. Note that this SAM–NFW is different from the shuffled and shuffled-NFW catalogues mentioned above.

5 TESTING THE ACCURACY OF THE HOD MODELS

5.1 Satellite radial distributions and clustering of HOD mocks

In Fig. 7, we show the satellite profiles in the SAM samples, for stellar mass and SFR-selected samples separated into the

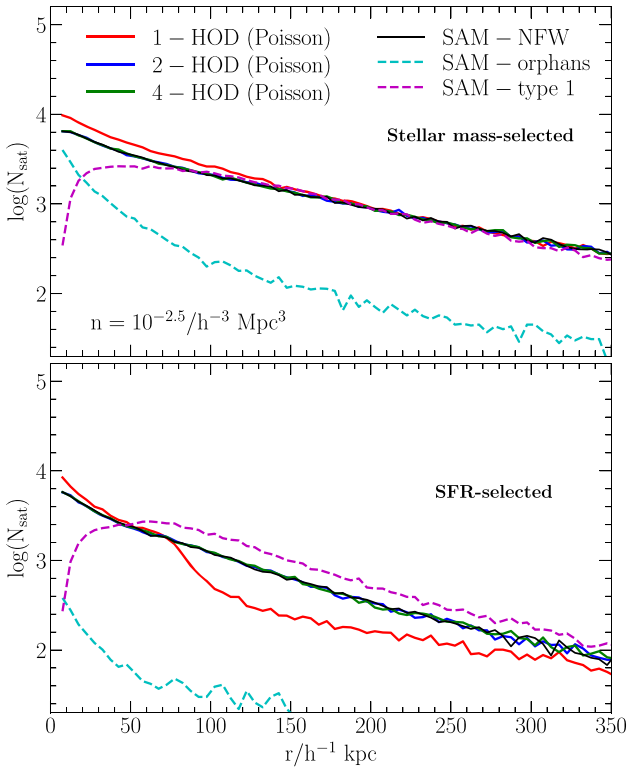

Figure 7. Profile of satellites hosted by subhaloes (dashed magenta) and orphans (dashed cyan) in a stellar mass (*top*) and SFR-selected sample (*bottom*). The lines show the SAM with an NFW imposed for all satellites (solid black), and HOD mocks built by the 1-HOD (solid red), 2-HOD (solid blue), and 4-HOD models (solid green) where the HODs of satellites are described by the Poisson distribution.

contributions from type-1 satellites and from orphan galaxies. It can be seen that the NFW profile (represented by the SAM–NFW catalogue) is different from the profile of type-1 satellites and orphans, particularly for the SFR-selected sample. The profiles in the 2-HOD and 4-HOD mock catalogues are also shown, and they match with the NFW as expected from the construction of the HOD mock. Both models also reproduce the NFW profile in the other number density samples.

The masses of host haloes of 1-HOD satellites do not correspond with the masses in the original SFR-selected samples (see Fig. A1). Thus, the virial radii of these haloes define NFW density profiles that are different from the profiles in the other models. This has an impact on the positions of satellites generating the striking difference with NFW in Fig. 7 for the SFR selection. The same occurs for the stellar mass selections but it is less extreme than the SFR case.

The HOD models predict different galaxy populations for the G13 SAM samples. Table 2 shows the satellite fraction of the SAM samples and HOD mocks built by the three different models, assuming a Poisson distribution for the HOD of satellites. Note that the 2-HOD and 4-HOD reproduce faithfully almost the same satellite fraction as in the SAM samples because of the separate HOD modelling of centrals and satellites.

We check the accuracy of each HOD model by comparing the HOD mocks with the shuffled-NFW sample via their 2PCFs. Fig. 8 shows the clustering of the shuffled-NFW and the HOD mocks built using the HOD models described in Section 4.1. These particular models assume a Poisson distribution for the HOD of satellites.

Table 2. Satellite fractions of the galaxy samples used. The first column indicates their number densities. Column 2, 3, 4, and 5 show the satellite fraction in the SAM samples and in the HOD mock built using the 1-HOD, 2-HOD, and 4-HOD models, respectively.

	Stellar mass			
$n/h^3\,\mathrm{Mpc}^{-3}$	SAM	1-HOD	2-HOD	4-HOD
$10^{-3.0}$	0.280	0.317	0.279	0.280
$10^{-2.5}$	0.322	0.405	0.324	0.324
$10^{-2.0}$	0.381	0.517	0.383	0.383
		SFR		
$n/h^3\,\mathrm{Mpc}^{-3}$	SAM	1-HOD	2-HOD	4-HOD
$10^{-3.0}$	0.171	0.084	0.175	0.172
$10^{-2.5}$	0.195	0.192	0.197	0.197
$10^{-2.0}$	0.244	0.334	0.246	0.246

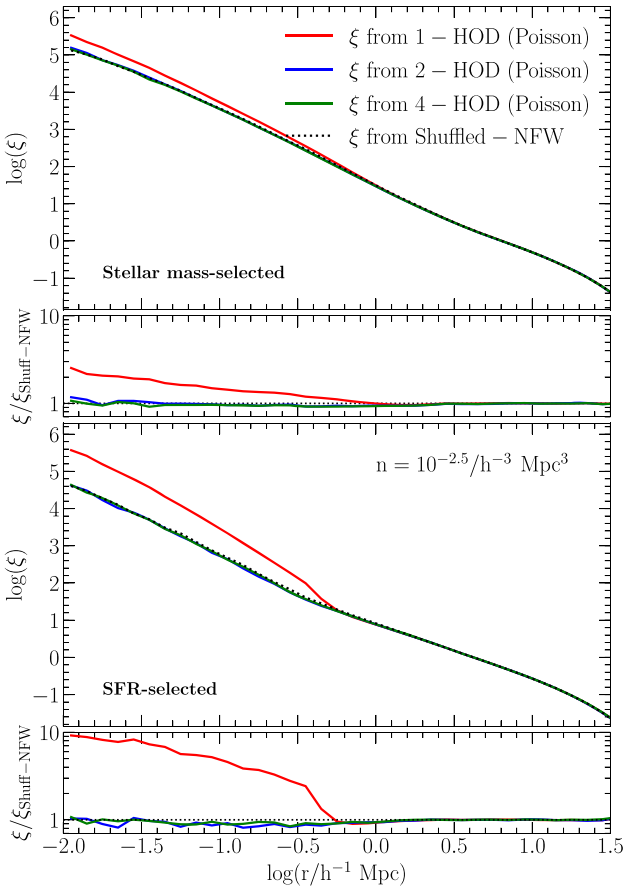

Figure 8. Clustering of HOD mock catalogues of stellar mass (*top*) and SFR-selected samples (*bottom*) for a number density of $n = 10^{-2.5}/h^{-3}\mathrm{Mpc}^3$. The mocks are built using the 1-HOD (red), 2-HOD (blue), and 4-HOD (green) models, assuming a Poisson distribution for the HOD of satellites. The clustering of the shuffled catalogue with an NFW profile is shown as the dotted line. Subpanels show the ratios of the 2PCF of the mocks with respect to the 2PCF of the shuffle-NFW catalogue.

It can be seen that the three schemes produce accurate clustering predictions on large scales. On small scales, the 2-HOD and 4-HOD models produce similar accurate results while the 1-HOD shows striking differences. These deviations come from the overprediction of the number of satellites in the stellar mass-selected samples. For the SFR selection cases, the difference is due to the notably

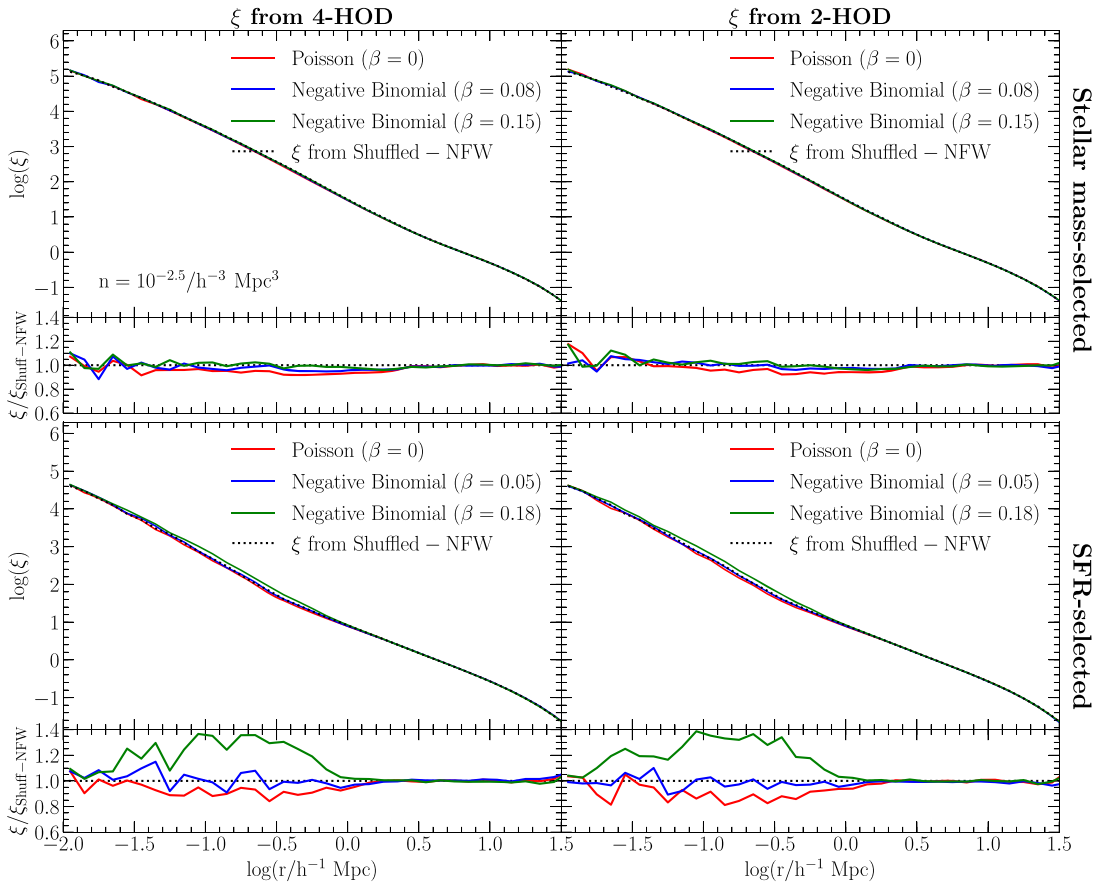

Figure 9. 2PCF of HOD mocks built using the 2-HOD (*right*) and 4-HOD (*left*) models, for stellar mass (*top*) and SFR-selected samples (*bottom*). The HOD models are used to build mock catalogues assuming a Poisson distribution (solid red) and negative binomial distributions of $\beta = 0.05$ (solid blue) and $\beta = 0.18$ (solid green) for the HOD of satellites. The clustering of the shuffled-NFW samples (dotted black) is shown in each panel. The ratios between the clustering of HOF mocks and shuffled-NFW are shown in the subpanels.

different occupation function of central and satellites in the 1-HOD mock (see Appendix A). The inaccuracy of the 1-HOD modelling is also present for the other number density samples too, whereas the 2-HOD and 4-HOD models produce similar quality results to the one shown here. As the 2-HOD and 4-HOD models are clearly the best, we drop the 1-HOD model henceforth. Fig. 8 reflects the importance of separating between centrals and satellites in the HOD modelling. Then, the 1-HOD should be avoided when making HOD mock catalogues.

5.2 Impact of the assumed HOD scatter

To study the impact of the scatter of the HOD on the clustering, we consider different dispersions for the negative binomial distribution in the construction of HOD mocks. Fig. 9 shows the 2PCF of HOD mocks using different β values. For the stellar mass-selected samples, the scatter of the HOD does not have a significant impact.

For the SFR selection, the amplitude of the clustering on small scales is very sensitive to the scatter in the number of satellites. We find that increasing β changes the amplitude of the one-halo term. When we split the contribution to the clustering from low- and high-mass haloes, we observe that the scatter mainly impacts the one-halo term of low-mass haloes. This feature is present in HOD mocks built using the 2-HOD and 4-HOD methods. This is reproduced by both HOD models, indicating that this is a feature particular to the SFR-selected samples.

The most accurate clustering reconstructions for the G13 samples are obtained when we use the 2-HOD or 4-HOD to build mock catalogues assuming a negative binomial distribution for the HOD of satellites. Note that clustering predictions from both models do not show significant differences.

Fig. 10 shows the particular results from the 4-HOD modelling for all the space density samples. For the G13 SFR-selected samples, the 4-HOD (and the 2-HOD) modelling produces the best results when $\beta = 0.05$, which corresponds to a distribution slightly wider than a Poisson distribution. For the case of stellar mass-selected samples, the best reproduction is obtained with $\beta = 0.08$. Using instead the Poisson distribution (i.e. $\beta = 0$) produces worse results for both selections particularly in the one-halo regime.

For SFR selections, when using $\beta = 0.05$ and $\beta = 0$, the departures from the shuffled-NFW catalogues are below ~ 8 per cent and ~ 15 per cent, respectively. It can be seen that the dispersion of the 2PCFs becomes important in the lowest number density sample. However, the assumption of the negative binomial distribution still produces better results, especially in the transition from the one- to the two-halo term.

For the stellar mass selection cases, the impact on clustering when using different β values is much less significant. Indeed, increasing the HOD scatter produces negligible changes in the amplitude of

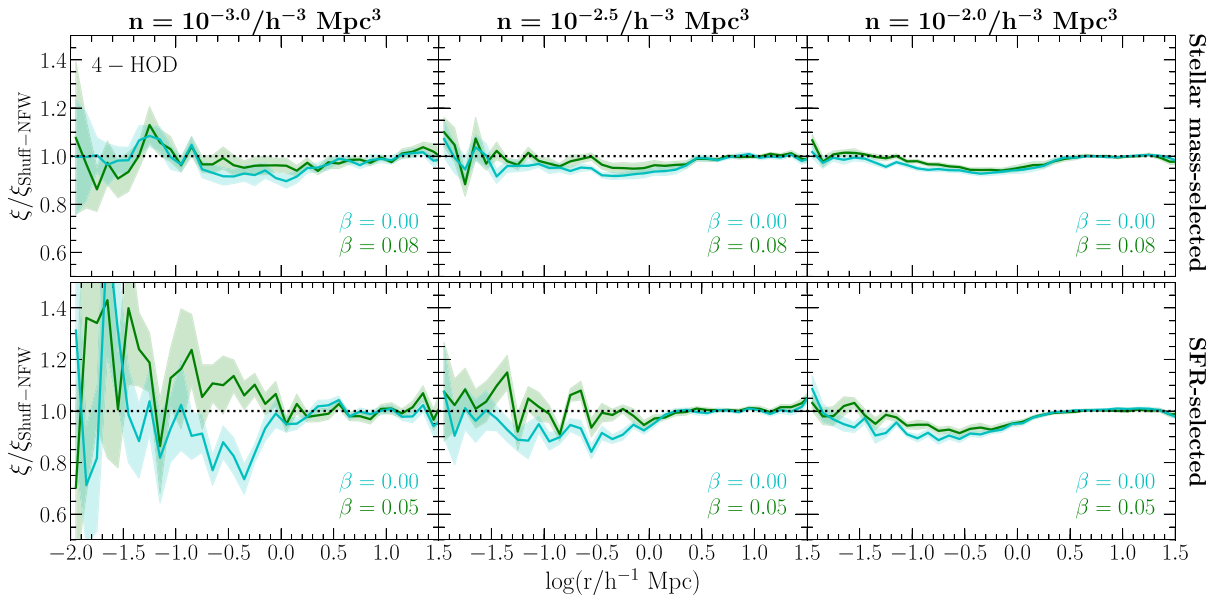

Figure 10. Ratios between the 2PCFs of mock catalogues, constructed with the 4-HOD method, and the shuffled-NFW catalogue. The HOD of satellites in the HOD mocks follows either the negative binomial (green) or Poisson distributions (cyan), with the colour indicating the value of β. We show results for stellar mass (*top*) and SFR-selected samples (*bottom*). Number densities increase from left to the right as labelled at the top of each column. The shaded regions represent jackknife errors calculated using 10 subsamples.

the one-halo term, either from low or high-mass haloes. Comparing with results for SFR selections, we suspect that a small (large) impact on clustering may be related to a high (low) occupation number in low-mass haloes ($M_{\rm h} \lesssim 10^{13} h^{-1} {\rm M}_{\odot}$, see Fig. 2).

The weak relation between clustering and HOD scatter, shown in Fig. 9, suggests that it is not necessary to include additional scatter in the construction of HOD mock catalogues for stellar mass selections. To compare with the Poisson distribution, we show also the clustering prediction for the stellar mass samples using $\beta = 0.08$, for which we obtain the best result.

Satellites in the G13 SAM are well described by a non-Poisson distribution. This is consistent with the HOD of subhaloes found in Boylan-Kolchin et al. (2010). The recipes that build mock catalogues of SFR-selected samples using the HOD approach must undertake an analysis of the scatter of the HOD of satellites as it impacts the clustering. This analysis will provide the best β to construct a HOD mock of a particular sample. For stellar mass-selected samples, the HOD scatter has a weak impact on clustering, so the same analysis is not necessary in the context of HOD mock catalogues.

6 CONCLUSIONS

The next generation of surveys will measure the clustering of the galaxy distribution over a wide range of redshifts. Mock catalogues have proven to be important tools in preparation for this because of their multiple applications including error analysis, data interpretation, and survey planning. SAMs are a physical approach to obtain such mocks, but sometimes their direct application to a simulation is not possible due to the limited resolution of the halo merger trees (Angulo et al. 2014) or the trees may not be available, as in the case of the *Euclid* flagship simulation (Potter, Stadel & Teyssier 2017). Even if the trees were available at the required resolution, the sheer number of haloes in a giga-parsec side N-body simulation may preclude a direct calculation with a SAM.

The HOD model provides a simple yet efficient way to construct mock catalogues. The modelling consists of using a probability distribution to obtain the number of galaxies hosted in a halo of a particular mass. This simple method allows us to create large sets of mock catalogues for huge cosmological simulations. This is useful in the context of ELGs, as they are targets in current and coming surveys.

To determine the level of complexity needed to produce accurate mock catalogues, we test different HOD models. The 1-HOD uses the HOD of all galaxies making no distinction between centrals and satellites while the 2-HOD uses the HOD of these two components separately. The 4-HOD stores additional information about whether or not haloes host a central, and it constructs conditional HODs for satellites taking into account this information.

Because SAMs include assembly bias by construction, and in their simplest form HOD mocks do not, we remove the assembly bias from the G13 SAM samples by shuffling the galaxy populations among haloes of the same mass creating the shuffled catalogue. This allows us to make a direct comparison between the clustering of our mocks and the SAMs from which we extract the HOD measurements. For example, we find that, for the intermediate galaxy density sample in the G13 SAM, the assembly bias affects the 2-halo term of the 2PCF of stellar mass-selected galaxies increasing the amplitude by ~ 12 per cent. For the SFR-selected galaxies, in contrast, the assembly bias suppresses the clustering by ~ 4 per cent. We also impose the standard NFW profile for satellites in the shuffled catalogue as is done for satellites in the HOD mock catalogues. We can then check the accuracy of the HOD models through a comparison between the 2PCFs of the HOD mocks and the shuffled-NFW catalogues.

The 2-HOD and 4-HOD produce the best mock catalogues as their 2PCFs are in close agreement with the clustering of the shuffled-NFW sample. We obtain the best results using a negative binomial distribution for the (conditional) HOD (see equation 1); in previous works this was commonly considered to be a Poisson distribution.

This is consistent with the subhalo HOD found in Boylan-Kolchin et al. (2010) using the Millennium-II simulation. Furthermore, we found that the assumption of this non-Poisonnian HOD changes the galaxy clustering. Previously, Jiang & van den Bosch (2017) found a similar result using subhaloes from the Bolshoi, the MultiDark N-body simulations, and the SAM presented in Jiang & van den Bosch (2016).

The scatter of the HOD of satellites in G13 is reproduced by a negative binomial distribution up to halo masses of $M_{\rm h} \lesssim 10^{13.5} h^{-1}{\rm M}_{\odot}$. The galaxies in this halo mass range dominate the amplitude of the 2PCF. We quantify the departure from the Poisson distribution with the parameter β (see equation 1). We obtain the best clustering predictions for SFR-selected samples using $\beta = 0.05$ and $\beta = 0.08$ for stellar mass-selected samples. These correspond to negative binomial distributions slightly wider than Poisson. Because of the specific modelling of different SAMs, we expect that the best β values for each sample are model dependent. For stellar mass-selected samples, we find that the HOD scatter has a weak impact on clustering, making the addition of this additional parameter unnecessary in the context of mock catalogues. As a first approach, we suggest choosing β such that the HOD modelling reproduces as closely as possible the HOD scatter of a given SAM sample (see Fig. 4). Then, by analysing the 2PCF of the resulting HOD mock, we may consider another β value that fits the HOD scatter in a particular halo mass regime (for $M_{\rm h} \lesssim 10^{13.5} h^{-1}{\rm M}_{\odot}$ in this work).

The analysis of the HOD of satellites is important because the width of the distribution (determined by the β parameter) has a large impact on the one-halo term of the 2PCF of mock catalogues that emulate SFR-selected sample and ELG samples. If we consider the Poisson distribution for the HOD of satellites ($\beta = 0$) the 2PCF of the mock catalogues is underestimated with respect to the clustering of the shuffled NFW. In contrast, using the negative binomial distribution increases the amplitude of clustering in the one-halo regime. If we assume a value of β larger than the one present in the distribution of number of satellites, the clustering on small scales is further overestimated. We highlight the importance to perform a careful analysis of the satellite HOD if the HOD framework is used to produce mock catalogues, for ELGs or star-forming galaxies, following a particular model or observation.

ACKNOWLEDGEMENTS

We thank the anonymous referee for insightful comments that helped improve the presentation of this paper. This work was made possible by the efforts of Gerard Lemson and colleagues at the German Astronomical Virtual Observatory in setting up the Millennium Simulation data base in Garching. EJ, SC, NP, and IZ acknowledge the hospitality of the ICC at Durham University. EJ acknowledges support from 'Centro de Astronomía y Tecnologías Afines' BASAL 170002. SC acknowledges the support of the 'Juan de la Cierva formacion' fellowship (FJCI-2017-33816). NP acknowledges support from Fondecyt Regular 1191813. IZ acknowledges support by NSF grant AST-1612085. This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement no. 734374. The calculations for this paper were performed on the Geryon computer at the Center for Astro-Engineering UC, part of the BASAL PFB-06, which received additional funding from QUIMAL 130008 and Fondequip AIC-57 for upgrades.

REFERENCES

Angulo R. E., Lacey C. G., Baugh C. M., Frenk C. S., 2009, MNRAS, 399, 983
Angulo R. E., White S. D. M., Springel V., Henriques B., 2014, MNRAS, 442, 2131
Artale M. C., Zehavi I., Contreras S., Norberg P., 2018, MNRAS, 480, 3978
Baugh C. M., 2006, Rep. Prog. Phys., 69, 3101
Baugh C. M., Benson A. J., Cole S., Frenk C. S., Lacey C. G., 1999, MNRAS, 305, L21
Benson A. J., 2010, Phys. Rep., 495, 33
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
Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, MNRAS, 398, 1150
Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., 2010, MNRAS, 406, 896
Cochrane R. K., Best P. N., 2018, MNRAS, 480, 864
Cochrane R. K., Best P. N., Sobral D., Smail I., Wake D. A., Stott J. P., Geach J. E., 2017, MNRAS, 469, 2913
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Conroy C., Wechsler R. H., Kravtsov A. V., 2006, ApJ, 647, 201
Contreras S., Baugh C. M., Norberg P., Padilla N., 2013, MNRAS, 432, 2717
Contreras S., Zehavi I., Padilla N., Baugh C. M., Jiménez E., Lacerna I., 2019, MNRAS, 484, 1133
Croton D. J., Gao L., White S. D. M., 2007, MNRAS, 374, 1303
Croton D. J. et al., 2006, MNRAS, 365, 11
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
De Lucia G., Blaizot J., 2007, MNRAS, 375, 2
De Lucia G., Kauffmann G., White S. D. M., 2004, MNRAS, 349, 1101
DeRose J. et al., 2019, ApJ, 875, 69
DESI Collaboration, 2016, preprint (arXiv:1611.00036)
Gao L., White S. D. M., 2007, MNRAS, 377, L5
Gao L., Springel V., White S. D. M., 2005, MNRAS, 363, L66
Geach J. E., Sobral D., Hickox R. C., Wake D. A., Smail I., Best P. N., Baugh C. M., Stott J. P., 2012, MNRAS, 426, 679
Gonzalez-Perez V. et al., 2018, MNRAS, 474, 4024
Guo Q. et al., 2011, MNRAS, 413, 101
Guo Q., White S., Angulo R. E., Henriques B., Lemson G., Boylan-Kolchin M., Thomas P., Short C., 2013, MNRAS, 428, 1351(G13)
Guo Q. et al., 2016, MNRAS, 461, 3457
Hearin A. P., Behroozi P. S., van den Bosch F. C., 2016, MNRAS, 461, 2135
Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R. E., Guo Q., Lemson G., Springel V., 2013, MNRAS, 431, 3373
Hicks E. K. S., Malkan M. A., Teplitz H. I., McCarthy P. J., Yan L., 2002, ApJ, 581, 205
Jiang F., van den Bosch F. C., 2016, MNRAS, 458, 2848
Jiang F., van den Bosch F. C., 2017, MNRAS, 472, 657
Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ, 740, 102
Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, ApJ, 609, 35
Lacerna I., Padilla N., 2011, MNRAS, 412, 1283
Lacerna I., Contreras S., González R. E., Padilla N., Gonzalez-Perez V., 2018, MNRAS, 475, 1177
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Ly C., Malkan M. A., Kashikawa N., Ota K., Shimasaku K., Iye M., Currie T., 2012, ApJ, 747, L16
Manera M. et al., 2013, MNRAS, 428, 1036
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, MNRAS, 396, 19

Orsi Á., Padilla N., Groves B., Cora S., Tecce T., Gargiulo I., Ruiz A., 2014, MNRAS, 443, 799
Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144
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
Schaye J. et al., 2015, MNRAS, 446, 521
Scoccimarro R., Feldman H. A., Fry J. N., Frieman J. A., 2001, ApJ, 546, 652
Sinha M., Garrison L., 2017, Corrfunc: Blazing fast correlation functions on the CPU (ascl:1703.003), http://adsabs.harvard.edu/abs/2017ascl.soft03003
Sobral D., Stroe A., Koyama Y., Darvish B., Calhau J., Afonso A., Kodama T., Nakata F., 2016, MNRAS, 458, 3443
Somerville R. S., Davé R., 2015, ARA&A, 53, 51
Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726
Springel V. et al., 2005, Nature, 435, 629
Vogelsberger M. et al., 2014, MNRAS, 444, 1518
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
Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., 2006, MNRAS, 366, 2
White S. D. M., Rees M. J., 1978, MNRAS, 183, 341
Yang X., Mo H. J., van den Bosch F. C., 2003, MNRAS, 339, 1057
Yoo J., Tinker J. L., Weinberg D. H., Zheng Z., Katz N., Davé R., 2006, ApJ, 652, 26
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
Zentner A. R., Hearin A. P., van den Bosch F. C., 2014, MNRAS, 443, 3044
Zheng Z. et al., 2005, ApJ, 633, 791
Zu Y., Zheng Z., Zhu G., Jing Y. P., 2008, ApJ, 686, 41

APPENDIX A: 1-HOD OCCUPATION FUNCTIONS

The 1-HOD model uses the HOD of all SAM galaxies in a sample to produce mock catalogues. This does not necessarily reproduce the HOD of central and satellites separately.

The 1-HOD assumes a distribution for the full HOD, which we take to be Poisson or a negative binomial. However, centrals follow the nearest integer distribution, whereas satellites follow a Poisson or negative binomial distribution. As this distinction is not made, the predicted HODs of centrals and satellites in the resulting mock are different with respect to the original SAM samples. Fig. A1 shows that 1-HOD tends to put satellites in very low mass haloes. Moreover, the occupation function of centrals in the SFR selection is overpredicted over a wide halo mass range.

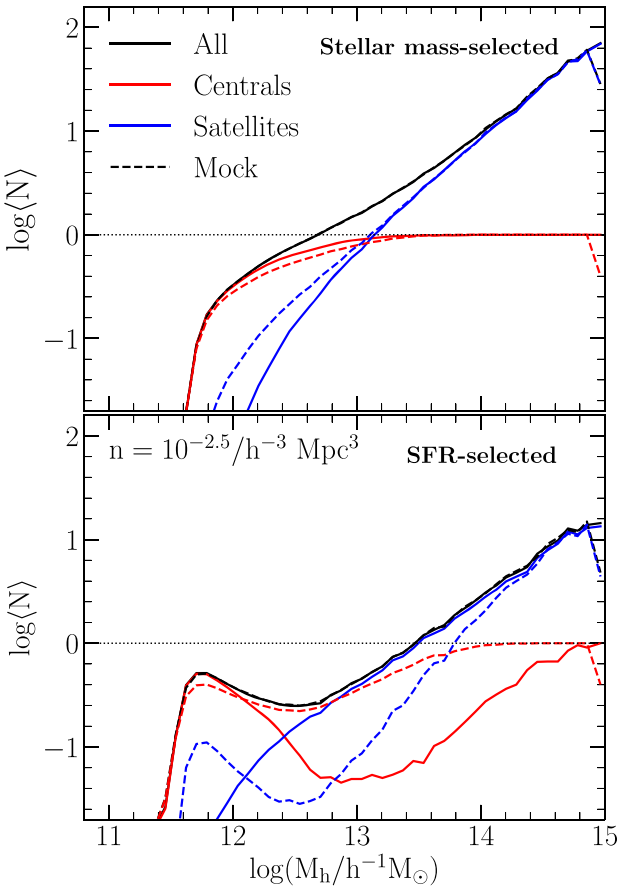

Figure A1. Same as Fig. 2 with the addition of the HODs of the 1-HOD mock catalogues (dashed lines). Solid lines correspond to the HOD of SAM samples. The 1-HOD model reproduces the HOD of all galaxies but not the HOD of central and satellites separately.

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