Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **465,** 2236–2253 (2017)
Advance Access publication 2016 November 10

doi:10.1093/mnras/stw2901

H I-selected galaxies in hierarchical models of galaxy formation and evolution

Anna Zoldan,[1,2]★ Gabriella De Lucia,[2] Lizhi Xie,[2] Fabio Fontanot[2] and Michaela Hirschmann[3]

[1]*Physics Department, Universitá degli Studi di Trieste, Via Valerio 2, I-34127-Trieste, TS, Italy*
[2]*OATS, INAF, Via Bazzoni 2, I-34124-Trieste, TS, Italy*
[3]*Institut d'Astrophysique de Paris, Sorbonne Universités, UPMC-CNRS, UMR7095, F-75014 Paris, France*

Accepted 2016 November 4. Received 2016 November 4; in original form 2016 August 10

ABSTRACT

In this work, we study the basic statistical properties of H I-selected galaxies extracted from six different semi-analytic models, all run on the same cosmological *N*-body simulation. One model includes an explicit treatment for the partition of cold gas into atomic and molecular hydrogen. All models considered agree nicely with the measured H I mass function in the local Universe, with the measured scaling relations between H I and galaxy stellar mass, and with the predicted two-point correlation function for H I-rich galaxies. One exception is given by one model that predicts very little H I associated with galaxies in haloes above $\sim 10^{12}$ $M_{\odot}$: we argue this is due to a too efficient radio-mode feedback for central galaxies, and to a combination of efficient stellar feedback and instantaneous stripping of hot gas for satellites. We demonstrate that treatment of satellite galaxies introduces large uncertainties at low H I masses. While models assuming non-instantaneous stripping of hot gas tend to form satellite galaxies with H I masses slightly smaller than those of centrals with the same stellar mass, instantaneous gas stripping does not translate necessarily in lower H I masses. In fact, the adopted stellar feedback and star formation affect the satellites too. We analyse the relation between H I content and spin of simulated haloes: low-spin haloes tend to host H I-poor galaxies, while high-spin haloes are populated by galaxies in a wide range of H I mass. In our simulations, this is due to a correlation between the initial gas disc size and the halo spin.

Key words: galaxies: evolution – galaxies: formation – galaxies: ISM – galaxies: statistics.

1 INTRODUCTION

Cold gas has a central role in galaxy evolution as it is involved in virtually all processes at play: cooling, star formation, feedback, mergers and processes related to the 'environment' such as ram-pressure and strangulation. The main component of cold gas is neutral hydrogen, in its atomic and molecular forms, each distributed in a 'disc' with a specific density profile (Leroy et al. 2008). Observationally, H I can be observed directly through the 21 cm line, while H_2 must be inferred from the ^{12}CO content through a factor, α_{CO}, whose value and dependence on other physical properties (in particular the gas metallicity) are not well known (Bolatto et al. 2011; Leroy et al. 2011; Magdis et al. 2011; Narayanan et al. 2012; Hunt et al. 2015; Amorín et al. 2016).

Surveys such as the H I Parkes All Sky Survey (HIPASS; Zwaan et al. 2005) or the Arecibo Legacy Fast ALFA survey (ALFALFA; Martin et al. 2010) have recently collected large amounts of data, but since the H I signal is relatively faint, these data are generally limited to low redshift and relatively H I-rich galaxies. These observations have provided detailed information on the basic statistical properties of H I-selected galaxies in the local Universe, in particular the H I mass function (Zwaan et al. 2005; Martin et al. 2010) and the two-point correlation function (2PCF; Meyer et al. 2007; Martin et al. 2012; Papastergis et al. 2013). We live in a very exciting era for H I observations, as new radio facilities such as the Square Kilometre Array (SKA) and its precursors/pathfinders will soon extend significantly the redshift range and dynamical range in H I mass probed (Obreschkow & Rawlings 2009; Kim et al. 2011), providing a new crucial test bed for our models of galaxy formation and evolution. It is therefore important to develop our theoretical tools, and assess how their basic predictions compare with the available data.

Until recently, the interstellar medium (ISM) of galaxies, in both semi-analytic models of galaxy formation and hydrodynamical simulations, was treated as a single phase constituted of gas with temperature below 10^4 K. Stars were assumed to form from this gaseous phase adopting different parametrizations, generally based on some

★E-mail: zoldan@oats.inaf.it

variation of the Kennicutt–Schmidt law (Kennicutt 1989). In the last decade, our understanding of how star formation depends on the local conditions of the ISM has improved significantly thanks to new high-resolution observations of the multiphase gas in samples of nearby galaxies. These data have shown that the density of star formation correlates strongly with the molecular gas density, while there is almost no correlation with the total gas density (e.g. Blitz & Rosolowsky 2006; Leroy et al. 2008, and references therein). These results have triggered significant activity in the theoretical community aimed at including an explicit treatment for the transition from atomic to molecular hydrogen, and molecular hydrogen based star formation laws (Fu et al. 2010; Lagos et al. 2011; Christensen et al. 2012; Kuhlen, Madau & Krumholz 2013). Recent studies have focused on the impact of different physical processes and of cosmology on the H I distribution and properties of H I-selected galaxies (e.g. Obreschkow & Rawlings 2009; Popping et al. 2009; Kim et al. 2011; Davé et al. 2013; Rafieferantsoa et al. 2015; Crain et al. 2016; Kim et al. 2016).

In this paper, we focus on results from semi-analytic models of galaxy formation. While these methods do not include an explicit treatment of the gas dynamics and prevent (at least in their standard implementations) studies of spatially resolved properties of galaxies, their computational cost is typically significantly lower than that of high-resolution hydrodynamical simulations. They therefore offer a flexible and efficient tool to analyse the dependence of model predictions on specific assumptions and parameterizations. We focus here on the basic statistical properties of H I-selected galaxies (i.e. their distributions in mass, their halo occupation distribution (HOD) and their two-point clustering function). To analyse the dependence on the different modelling adopted, we use four different models whose results are publicly available on the Virgo-Millennium Database (Lemson & Virgo Consortium 2006), and other two models recently developed by our group (Hirschmann, De Lucia & Fontanot 2016, and Xie et al. 2016). The latter model, in particular, includes an explicit treatment for partitioning the cold gas associated with model galaxies in H I and H_2. For all other models, as we will discuss below, the H I content of model galaxies has been derived post-processing model outputs.

The layout of the paper is as follows. In Section 2.1, we briefly describe the models used in our study, their backbone simulations and the algorithm we use to build mock light cones from model outputs. In Section 3, we explain how we assign H I masses to model galaxies where this is not an explicit model output, and compare the distribution of H I masses and basic scaling relations predicted by the models with available observational data. In Section 4, we analyse the predicted 2PCF of H I-selected galaxies, and compare model predictions with recent measurements by Papastergis et al. (2013). We discuss the results on the basis of the predicted HOD. In Section 5, we discuss the role of satellite galaxies and study the predicted evolution of their H I content. In Section 6, we study the relation between the H I in galaxies and the physical properties of the hosting dark matter haloes. Finally, in Section 7, we summarize our results and give our conclusions.

2 SIMULATIONS AND GALAXY FORMATION MODELS

In this work, we take advantage of six different semi-analytic models of galaxy formation and evolution that originate from the work of three independent research groups. All models are run on the same *N*-body cosmological simulation, i.e. the Millennium Simulation (Section 2.2), and all follow the principles outlined in White & Frenk (1991) including specific (different) modelling for cooling, star formation, stellar feedback, mergers and starbursts, disc instabilities, chemical enrichment and feedback from active galactic nuclei (AGN). Using models that employ different prescriptions for processes involving cold gas, it is possible to quantify the relative importance in determining the observed relations. At the same time, systematic (i.e. common to all models) disagreements can be used to identify specific aspects of the models that need to be improved. Below, we briefly outline the main differences between the models considered, the main characteristics of the simulations used and describe the software built to construct mock light cones from the available galaxy catalogues.

2.1 Galaxy formation models

The Bower et al. (2006, hereafter B06) model has been developed by the 'Durham group' and is an extension of the GALFORM model published in Cole et al. (2000) and Benson et al. (2003). The model published in De Lucia & Blaizot (2007, hereafter DLB07) has been developed by the 'Munich group' and represents an extension of the model described earlier in Springel et al. (2001), De Lucia, Kauffmann & White (2004b) and Croton et al. (2006). These two models have been the first whose results were made publicly available through a relational data base[1] that we have heavily used for our study.

It was early realized that both these models tend to overpredict the number densities of galaxies smaller than the Milky Way, and the overall fraction of passive galaxies – problems that have turned to be of difficult solution and shared by all hierarchical models published in the last years, including hydrodynamical cosmological simulations (see e.g. De Lucia, Muzzin & Weinmann 2014, and references therein). The other models we use in this study represent two independent branches, both based on subsequent upgrades of the DLB07 model, that provide significant improvements on these problems.

The model described in Guo et al. (2011, hereafter G11) differs from the DLB07 model primarily for a more efficient feedback and a non-instantaneous stripping of the hot reservoir associated with galaxies at the time of infall on larger systems. The modified treatment of satellite galaxies improves the agreement with observational data as for the fraction of active galaxies, while the more efficient feedback brings the predicted mass function in agreement with data in the local Universe. The model, however, still suffers an excess of intermediate- to low-mass galaxies at higher redshift. Henriques et al. (2013) adopt the same physical model as in G11, but include a variation with cosmic time and halo mass of the efficiency with which galactic wind ejecta are re-accreted. They then use Monte Carlo Markov Chain methods to identify the parameter space that allows the measured evolution of the galaxy mass (and luminosity) function from $z = 0$ to $z = 3$ to be reproduced. Henriques et al. (2015, hereafter H15) have later extended this work to the *Planck* first-year cosmology, by 're-scaling' the Millennium Simulation merger trees, as explained in Section 2.2. In this work, we use model outputs based on the rescaled simulation.

The other independent branch we use in our study has been developed by the 'Trieste group', and is provided by the recently published GAlaxy Evolution and Assembly (GAEA) model (Hirschmann et al. 2016), and by an extension of this model (Xie et al. 2016, XBR16). The GAEA model differs from DLB07 primarily for the

[1] http://wwwmpa.mpa-garching.mpg.de/millennium/

inclusion of a sophisticated chemical enrichment scheme that accounts for the finite lifetimes of stars and independent yields from massive stars and both SNII and SNIa, and for an updated stellar feedback scheme, based on the results obtained in the framework of the Feedback In Realistic Environments (FIRE) simulations (Hopkins et al. 2014). The GAEA model includes modifications of the re-incorporation rate of ejected gas similar to those suggested by H15. We refer to the original paper for more details. It is worth stressing that GAEA adopts an instantaneous stripping of the hot gas reservoir associated with infalling galaxies. Yet, Hirschmann et al. (2016) demonstrate that this model is successful in reproducing simultaneously the evolution of the galaxy stellar mass function, the fraction of passive galaxies[2] as a function of stellar mass observed in the local Universe and the observed evolution of the relation between galaxy stellar mass and metallicity content of the gaseous phase.

All models introduced so far treat the gas as a single-phase component, and therefore need to be *post-processed* to infer the H I content associated with each model galaxy, as described in Section 3.1. The XBR16 model represents an update of the GAEA model including a treatment for the partition of the gaseous phase in atomic and molecular gas, based on the empirical relation by Blitz & Rosolowsky (2006). Therefore, among all models considered in our study, this is the only one including a self-consistent treatment of the star formation rate, which depends on the molecular hydrogen content of galaxies. In this model,ç the disc is divided in 21 concentric annuli (in all other models, star formation is modelled assuming a single value for the entire disc), and the star formation rate is estimated in each of the annuli assuming it is proportional to the surface density of molecular gas. A partition of the cold gas into its molecular and atomic components is performed at each time-step (just before star formation takes place) of the simulation, in each annulus. This allows star formation to continue for longer times than other models in the inner parts of the disc. This more sophisticated treatment of the star formation process leads to relevant differences with respect to the GAEA model, as we will show in the next sections.

Among the models used in this paper, B06, DLB07, G11 and H15 are publicly available at the relational data base mentioned above, developed as part of the activities of the German Astrophysical Virtual Observatory (GAVO). The GAEA and XBR16 models are not yet public, but will soon be released on the same data base. Additional models, including more recent versions of the GALFORM model featuring a treatment for the partition of cold gas in atomic and molecular hydrogen, are accessible from an alternative data base server at the Institute for Computational Cosmology.[3]

2.2 Dark matter simulation and merger trees

All models presented in the previous section have been run on the Millennium Simulation (Springel et al. 2005). This is a cosmological *N*-body dark matter only simulation that follows the evolution of $N = 2160^3$ particles of mass $8.6 \times 10^8\, h^{-1}\, \mathrm{M}_\odot$ within a comoving box of $500\, h^{-1}$ Mpc on a side, and with a cosmology consistent with *WMAP*1 data. In particular, the values of the adopted cosmological parameters are: $\Omega_\mathrm{b} = 0.045$, $\Omega_\mathrm{m} = 0.25$, $\Omega_\Lambda = 0.75$, $H_0 = 100h\,\mathrm{Mpc}^{-1}\,\mathrm{km\,s}^{-1}$, $h = 0.73$, $\sigma_8 = 0.9$ and $n_s = 1$. Simulation data were stored at 64 output redshifts, each corresponding to a snapshot number n (with n varying from 0 to 63) through the following formula: $\log(1 + z_n) = n(n + 35)/4200$.

Dark matter haloes were identified using a standard friends-of-friends algorithm with a linking length of 0.2 in units of the mean particle separation. The algorithm SUBFIND (Springel et al. 2001) was then used to identify bound substructures with a minimum of 20 particles, which corresponds to a halo mass resolution limit of $M_h = 1.7 \times 10^{10}\,\mathrm{M}_\odot\, h^{-1}$. Halo merger trees were built tracing the majority of the most bound particles of each subhalo from a given snapshot to the following one. In this way, each halo was assigned a unique descendant.

As mentioned above, H15 is based on a different cosmology: the authors still use the Millennium Simulation but 'rescale' it to the *PLANCK* cosmology (Planck Collaboration et al. 2014) using the technique discussed in Angulo & White (2010), as updated by Angulo & Hilbert (2015). Specifically, the adopted cosmological parameters are: $\Omega_\mathrm{b} = 0.0487$ ($f_b = 0.155$), $\Omega_\mathrm{m} = 0.315$, $\Omega_\Lambda = 0.685$, $h = 0.673$, $\sigma_8 = 0.829$ and $n_s = 0.96$. We note that as found in previous studies (e.g. Wang et al. 2008; Guo et al. 2013), relatively small variations of the cosmological parameters are found to have little influence on the overall model predictions once model parameters are (slightly) modified to recover the same model normalization.

To control the resolution limits of our models, we have also taken advantage of the Millennium II simulation (Boylan-Kolchin et al. 2009, MS-II). The size of this simulation is one fifth that of the Millennium Simulation (i.e. 100 Mpc h^{-1} on a side), but the particle mass is 125 times lower, resulting in a minimum resolved halo mass of $M_h = 1.4 \times 10^8\,\mathrm{M}_\odot\, h^{-1}$. Below, we will use G11, H15 and XBR16 outputs based on the MS-II in order to check the reliability of some relations, in particular those involving dark matter halo masses in the 10^{11}–$10^{12}\,\mathrm{M}_\odot\, h^{-1}$ range.

2.3 Light-cones algorithm

In Section 4, we analyse the two-point correlation function of galaxies selected according to their H I content. We use mock light cones and the projected correlation in order to carry out a straightforward comparison with observations. In addition, the construction of many light cones (some dozens) allows us to give an estimate of the error on the projected 2PCF due to the cosmic variance, as well as of the random noise expected for the real 2PCF.

We have developed an algorithm that creates mock catalogues from the outputs of galaxy formation models. We use these catalogues in the calculation of the 2PCF (Section 4). The approach is very similar to that adopted in the Mock Map Facility developed by Blaizot et al. (2005, MoMaF), and we briefly summarize it next. We refer to the original paper for more details.

As explained above, the outputs of the galaxy formation models used in this study are given by galaxy catalogues stored at a finite number of snapshots, each corresponding to a different redshift z_i. Galaxy catalogues contain a number of physical properties (i.e. masses, metallicities, luminosities, etc.) and consistent redshift and spatial information, i.e. the position of each model galaxy within the simulated box, as well its velocity components. The first problem that arises when constructing mock observations from these types of outputs is that redshift varies continuously along the past light cone while outputs are stored at a finite number of redshifts (in our particular case spaced at approximately 300 Myr intervals out to $z = 1$). The other problem arises because of the need to 'replicate'

[2] The quiescent fraction measured for the lowest mass galaxies is still larger than observational estimates, but improved significantly with respect to predictions from the DLB07 model.
[3] http://galaxy-catalogue.dur.ac.uk:8080/Millennium/

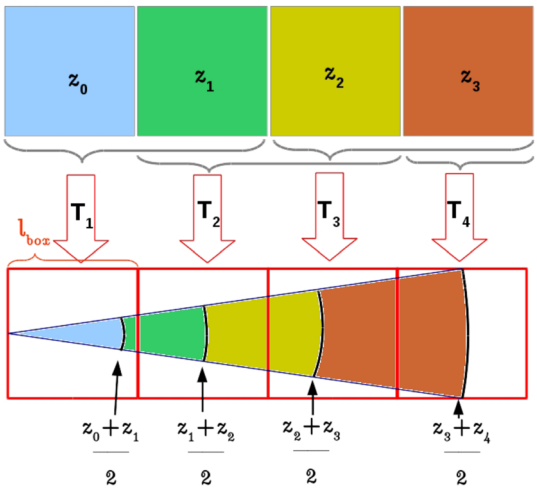

Figure 1. A schematic illustration of our light-cone construction algorithm: a grid of cubic boxes with size length $l_{\rm box}$ (red) is built over the cone, and galaxies inside each box of the grid are randomly shifted/rotated/inverted depending on the specific box they end up in. Galaxies are extracted from the simulation output with redshift closest to the comoving distance to the observer.

the same simulated region of the universe (in our case corresponding to a box of $500\,h^{-1}$ Mpc comoving on a side) several times to fill the light cone.

We try to minimize these problems by building the light cone as follows (we refer to Fig. 1 for a schematic description of our method).

(i) We build a 3D grid made of cubes of the same size $l_{\rm box}$ of the simulation box (the red grid in the figure).

(ii) The cone is placed inside this grid horizontally and moved so as to overlap with the smallest number of grid boxes (the blue cone in the figure).

(iii) The cone is divided into different regions whose edges are given by the average redshift of two subsequent snapshots, namely $z_i^{lc} = (z_i + z_{i+1})/2$ (black arrows in the figure).

(iv) Model galaxies are then placed inside the cone according to the following rules.

(a) Each redshift region z_i^{lc} of the cone is filled with galaxies from the box corresponding to the closest redshift z_i (the correspondence is colour-coded in the figure).

(b) The positions (and velocities) of model galaxies are randomly transformed, in order to avoid replications of structures along the line of sight. A specific random transformation is assigned to each 3D grid box. This is a combination of a random shift, a random rotation of 0, $\pi/2, \pi$ or $3\pi/2$ around a random axis or an inversion of coordinates along a random axis. In the figure, a specific transformation is indicated with a big arrow.

Referring to the specific example shown in Fig. 1, the light-cone element corresponding to the second grid box (from the left) is divided in two redshift regions by the edge $z_2^{lc} = (z_1 + z_2)/2$. The positions (and velocities) of the galaxies in the snapshots corresponding to z_1 and z_2 are transformed using the same specific transformation T_2. The transformed snapshots are then used to fill the grid element considered. Galaxies below z_2^{lc} are extracted from the transformed snapshot z_1, while those above z_2^{lc} are taken from the transformed snapshot z_2.

Adopting this approach, large-scale structures are affected at the edges of the 3D grid boxes, but not at the redshift separation between subsequent snapshots. This allows accurate measurements of the 2PCF.

Blaizot et al. (2005) carried out a detailed analysis of replication effects on the clustering signal. In particular, they quantified the effect due to (i) random tiling and (ii) finite volume. The random tiling approach introduces a negative bias in the mock catalogues because it decorrelates pairs of galaxies when reshuffling them to suppress replication effects. Blaizot et al. (2005) gave a theoretical estimate of the relative error in the number of pairs due to shifting of boxes. For the spatial correlation function, this is found to be less than 10 per cent on scales ranging from 1 to $\sim 10\,h^{-1}$ Mpc. A numerical estimate was also computed by subdividing the volume of the simulation used in their study (a cube of $100\,h_{100}^{-1}$ Mpc on a side) in 8^3 sub-volumes to which translations, rotations and inversions were applied as described above. We have repeated the same exercise by taking advantage of the larger volume of the Millennium Simulation and considering sub-boxes of different sizes ($L_{\rm box} = 250, 125\,{\rm Mpch}^{-1}$). For the analysis presented in the following, light cones are built starting from the entire volume of the MS ($L_{\rm box} = 500\,{\rm Mpch}^{-1}$), and we have verified that the random tiling approach is not expected to significantly affect our measurements.

The finite volume of the simulation prevents us from studying fluctuations larger than $\sim V^{1/3}$. Since in this study we focus on clustering on relatively small scales ($\sim 20\,h^{-1}$ Mpc), we do not expect the finite volume of our simulation to affect significantly our measurements.

3 NEUTRAL HYDROGEN DISTRIBUTION AND SCALING RELATIONS

In this section, we study the distribution of H I masses and basic scaling relations predicted by our models, and compare them with recent observational determinations. Given the resolution mass limit of the Millennium Simulation, all the following analysis is limited to galaxies with $M_* > 10^9\,{\rm M}_\odot$. Below this limit, the sample is incomplete (this is evaluated comparing the galaxy stellar mass function for models based on the MS and on the MS-II). Using the H I mass function, Xie et al. (2016) have determined the completeness limit in H I mass that is assessed to be $M_{\rm HI} \sim 10^{8.8}\,{\rm M}_\odot$. In the following analysis, we study relations involving H I masses slightly below this limit. We checked the influence of incompleteness using the MS-II for the G11, H15 and XBR16 models, and found that, qualitatively, our results do not change.

The following analysis is performed on a sub-volume of the MS of one-sixth, but we have verified that results do not change significantly when considering a larger volume.

3.1 Neutral hydrogen mass estimates

XBR16 is the only model among those considered in our study that accounts self-consistently for a direct dependence of star formation rate on the molecular hydrogen density (Blitz & Rosolowsky 2006; Leroy et al. 2008). This is done by introducing a partition of the cold gas phase in atomic and molecular hydrogen, based on the empirical relation by Blitz & Rosolowsky (2006), at the time of each star formation episode. Therefore, the H I and H_2 content of model galaxies are direct outputs of this model, even if they are not explicitly followed as two separated components. For all the other models, star formation is assumed to depend on the total cold gas content through a classical Kennicutt (1989) relation. In this case,

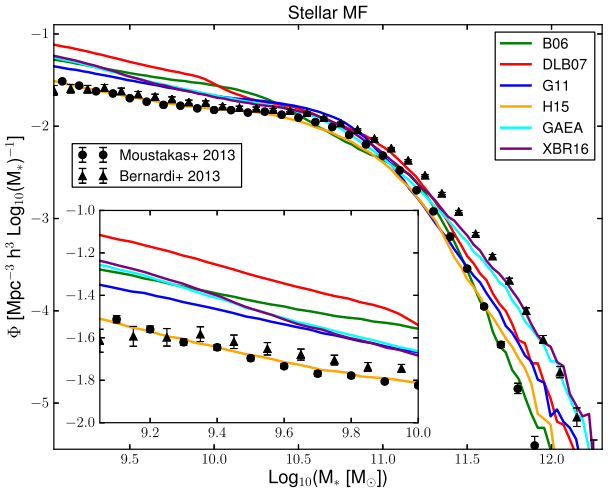

Figure 2. The galaxy stellar mass function predicted by the six models used in our study, compared to observational measurements by Moustakas et al. (2013, circles) and Bernardi et al. (2013, triangles), both based on SDSS. The inset shows a zoom on the low-mass end.

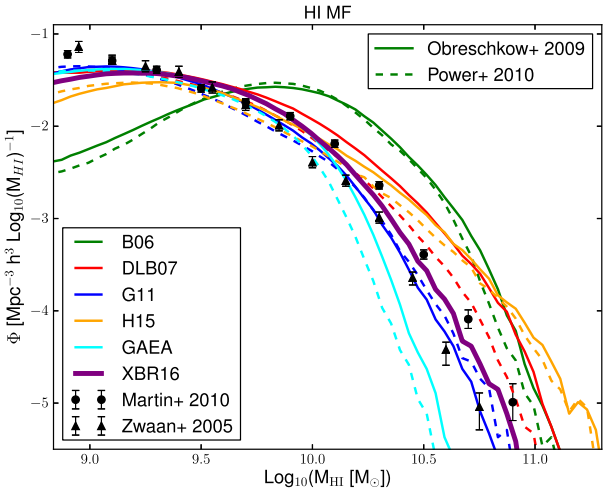

Figure 3. The H I mass function predicted by the models, compared to observational data by Martin et al. (2010, circles) and Zwaan et al. (2005, triangles). Solid lines correspond to the Obreschkow & Rawlings (2009) method to assign H I masses, while dashed lines correspond to the assumption of a constant proportionality between total cold gas and H I mass as in Power et al. (2010).

the H I content of galaxies can be estimated using two different methods.

(i) Combining the observed gas density profile (Leroy et al. 2008) and the dependence of $R_{\rm mol} = M_{H_2}/M_{\rm HI}$ on the disc mid-plane pressure (Blitz & Rosolowsky 2006), Obreschkow & Rawlings (2009) find the following relation:

$$R_{\rm mol} = [3.44 R_c^{-0.506} + 4.82 R_c^{-1.054}]^{-1}.$$

In this equation, $R_c \sim [r_{\rm disc}^{-4} M_{\rm cold}(M_{\rm cold} + 0.4 M_*)]^{0.8}$ is the central $R_{\rm mol}$, $r_{\rm disc}$ is the disc scalelength, $M_{\rm cold}$ the cold gas mass in the disc and M_* the stellar mass of the disc.

(ii) Following Baugh et al. (2004) and Power, Baugh & Lacey (2010), we can assume a direct proportionality between the cold gas content and H I. Helium amounts to about 24 per cent of the total cold gas, while metals represent a negligible fraction. Assuming the H II content is negligible, hydrogen can be divided in ∼71 per cent of H I and ∼29 per cent of H_2, namely $R_{\rm mol} = M_{H_2}/M_{\rm HI} = 0.4$ (Keres, Yun & Young 2003; Zwaan et al. 2005).

Power et al. (2010) analysed the impact of these two approaches on different semi-analytic models (including some used in our study) and found little differences between them at low redshift (differences are larger at high redshift). We confirm that there are only small differences between the H I content of model galaxies estimated using the two methods described above, for all models used in our study. Therefore, unless otherwise stated, we will always compute H I using the Obreschkow & Rawlings (2009) prescription.

3.2 Stellar and H I mass functions

In Fig. 2, we compare the galaxy stellar mass function predicted by the models with observational measurements by Moustakas et al. (2013) and Bernardi et al. (2013). Both determinations are based on the SDSS data, but assume different estimates of stellar mass. In particular, the latter is based on a Sérsic-exponential fit to the surface brightness profiles of galaxies, and translates in significantly higher number densities of massive galaxies. The two measurements highlight that, while uncertainties are relatively small for the intermediate- and low-mass regime, they are much larger above the knee of the mass function. Furthermore, in this mass regime also statistical uncertainties are larger because of the small number of galaxies.

All the models reproduce well the observational measurements around the knee, with a slight underestimation for the B06 and H15 models. The low-mass end of the mass function is well reproduced only by the H15 model, while the largest overprediction is found for the DLB07 and B06 models. As mentioned above, the good agreement found for the H15 model is obtained by construction as these authors tune their model parameters to reproduce the evolution of the galaxy mass function from $z = 0$ to $z = 3$. At high masses, models tend to be closer to the Moustakas et al. (2013) determination up to stellar masses $\sim 10^{11}\,{\rm M}_\odot$, while they basically cover all the range between the two measurements shown for larger masses. For the most massive galaxies, the lowest number densities are obtained by the B06 and H15 models, the largest by GAEA and XBR16.

Fig. 3 shows the H I mass function predicted by the models used in our study, obtained using the methods discussed in Section 3.1 (with the exception of XBR16, that returns a direct estimate of the H I mass). Model predictions are compared to observational data from Zwaan et al. (2005), based on HIPASS, and from Martin et al. (2010), based on the ALFALFA survey.

The two different H I estimates (solid and dashed lines in the figure) return quite similar results, with the Power et al. (2010) prescription slightly underpredicting the H I high-mass end with respect to the method suggested by Obreschkow & Rawlings (2009).

B06 is the model that deviates most from the data. In particular, it overpredicts the number densities of galaxies with high H I masses ($M_{\rm HI} \geq 10^9\,{\rm M}_\odot\,h^{-2}$), and underpredicts it for lower masses. Kim et al. (2011) already discussed this limitation of B06 and proposed a new version of this model tuned to reproduce the HIMF of Zwaan et al. (2005). This modified model also resulted in slightly better agreement with the observed 2PCF based on HIPASS. G11 and XBR16 reproduce quite well the observations, while the GAEA model underestimates the high-mass end. We remind that GAEA and XBR16 are based on the same physical model but the latter includes an explicit treatment for the partition of cold gas in atomic and molecular hydrogen in disc annuli and a star formation law that depends on H_2. The division in annuli translates in the XBR16

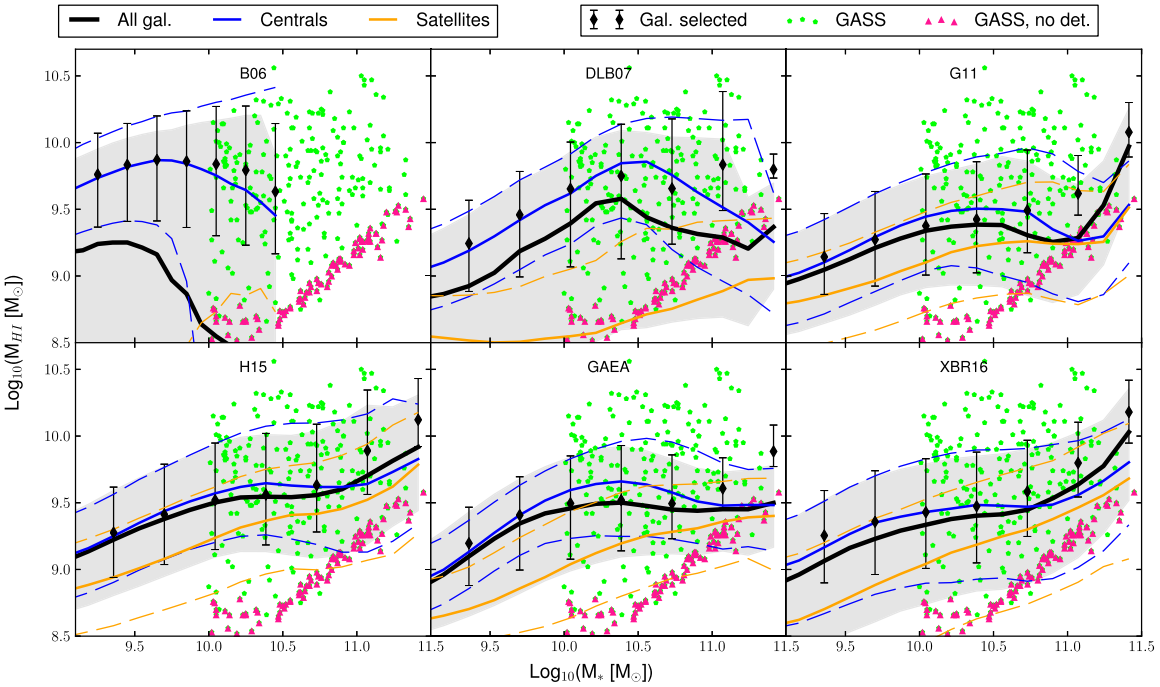

Figure 4. The predicted relation between H I mass and galaxy stellar mass, compared to results from the *GALEX* Arecibo SDSS Survey (symbols). Lines of different colours are used for central galaxies (blue) and satellites (orange), while the black lines and shaded regions correspond to the entire galaxy population. Solid lines correspond to the median of the distributions, while dashed lines (and the shaded region) mark the 1σ distribution. The black diamonds with error-bars show the distribution of all model galaxies when using the same detection threshold of the data. Symbols of different colours are used for detections (green pentagons) and non-detections (upper limits, purple triangles).

model having less H_2 at fixed star formation rate, and this leads to a larger amount of H I (at fixed cold gas). This effect is enhanced for gas-rich galaxies. Finally, we note that the XBR16 model is tuned to reproduce the HIMF.

The slight underestimation of the HIMF at the low-mass end in all the models is due to resolution (Xie et al. 2016).

3.3 Scaling relations

Fig. 4 shows the predicted relation between the H I mass and the galaxy stellar mass for all models used in our study. Solid lines show the median of the distributions, while dashed lines (and the shaded region that corresponds to the black line) mark the 1σ spread. Lines of different colours correspond to centrals (blue) and satellite galaxies (orange). Symbols correspond to observational data from the *GALEX* Arecibo SDSS Survey (Catinella et al. 2013), divided in detected H I (green pentagons) and non-detections, and thus upper limits (purple triangles).

Given the large spread of the observational data and the detection limits, it is hard to use these data to put strong constraints on the models. All models cover the region sampled by the observational data but for the B06 model that predicts no H I in galaxies more massive than $\sim 10^{10}$ M$_\odot$ likely because of too efficient AGN (radio-mode) feedback. We accounted for the same detection limit of observations (for $M_* > 10^{10}$ M$_\odot$, only considered galaxies with $M_{\rm HI}$ larger than the maximum no detection in the observed sample). Results are shown in this case as black diamonds with 1σ error-bars: this narrows the distributions of model galaxies, particularly for the G11, GAEA and XBR16 models. Overall, the agreement between model predictions and observations is good, but some models (e.g. G11, GAEA and XBR16) exhibit a deficit of H I-rich galaxies at any given stellar mass. This remains statistically significant even when applying a detection threshold similar to that of the survey used as a comparison.

As expected, central galaxies always have larger H I masses than satellite galaxies. The difference is largest for the B06 and DLB07 models that adopt an instantaneous stripping of the hot gas reservoir associated with infalling satellites. The G11 and H15 models relax this assumption, which brings the median H I mass of satellite galaxies closer to that of central galaxies. Interestingly, the difference between these medians in the GAEA and XBR16 models, which also assume an instantaneous stripping of the hot gas, is only slightly larger than that found in G11 and H15. As discussed in Hirschmann et al. (2016), this is driven by significantly larger amounts of cold gas at higher redshift and a reduced reheating rate from stellar feedback. Interestingly, at the most massive end, the H I content of satellite galaxies is comparable to that of centrals in all models but B06 and DLB07.

In Fig. 5, we show the cold gas fraction $F_g = M_{\rm cold}/M_*$ as a function of the galaxy stellar mass for all models considered in our study. Symbols show observational measurements by Catinella et al. (2013, GASS survey, pentagons and triangles) and Peeples & Shankar (2011, red points with error bars). The latter are a collection of previous H I measurements from McGaugh (2005), West et al. (2009, 2010) and Leroy et al. (2008). To convert the measurements in cold gas fraction, we have assumed H I represents a constant fraction ($\sim$71 per cent) of the total cold gas available.

As discussed above, the B06 model clearly underpredicts the gaseous content of galaxies of intermediate to high mass. In addition, virtually all gas in this model is associated with central galaxies. Comparing Figs 5 and 4 with the predicted H I mass function for this model (Fig. 3), we can see that the excess of H I-rich galaxies is driven by the high gas fractions of galaxies with low stellar masses. As noted above, even accounting for the same selection limits of

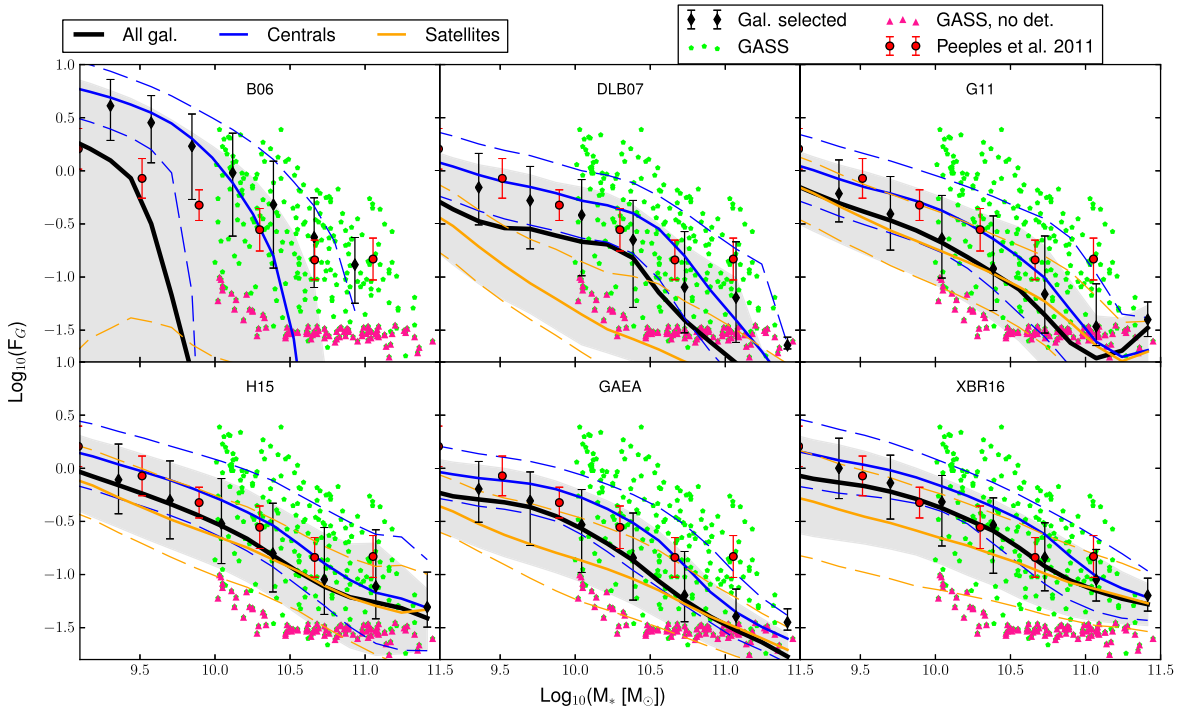

Figure 5. The gas fraction $F_G = M_{\rm cold}/M_*$ of model galaxies as a function of their stellar mass, compared to observational measurements by Catinella et al. (2013, green pentagons and pink triangles), and Peeples & Shankar (2011, red circles with error bars). Lines (solid and dashed), shaded region and diamonds with error-bars have the same meaning as in Fig. 4.

the observations (black diamonds and error-bars), only some of the models predict gas fractions as large as those observed (DLB07, H15, XBR16). Finally, both the DLB07 and the GAEA models underpredict the gas fractions estimated for galaxies with the largest stellar mass. When applying the same selection of the data, however, model predictions appear consistent with observations. It should be noted that observational measurements are sparse and likely incomplete in this mass regime, thus further measurements are necessary to constrain this relation at the highest stellar masses.

4 TWO-POINT PROJECTED CORRELATION FUNCTION OF H I-SELECTED GALAXIES

In this section, we will analyse the clustering properties of model galaxies, selecting them in H I bins. We use, as observational reference, the work by Papastergis et al. (2013, hereafter P13). They used 6123 H I-selected galaxies from the ALFALFA survey, covering a contiguous rectangular sky region of $\sim$1700 deg^2 in the redshift range $z \sim 0.0023$–0.05. They also used an optical sample of 18 516 galaxies in the same volume using the SDSS DR7 (Abazajian et al. 2009), and applying a magnitude cut of $M_r < -17$. As expected and shown in previous studies (e.g. Catinella et al. 2010; Huang et al. 2012), the most luminous galaxies tend to be H I poor, while H I-rich galaxies tend to be the dominant population among galaxies that populate the blue cloud.

Below, we will compare our model predictions with the estimated projected correlation functions for the following H I mass bins: $\log_{10}(M_{\rm HI}\,[{\rm M}_\odot]) \in [8.5;\ 9.5]$, $[9.5; 10.0]$ and $[10.0; 10.5]$. Lower mass bins were considered in P13, but these fall below the resolution limit of our models. The lowest bin considered here is already partially below the completeness limit. We will account for this in the following.

P13 found little evidence of dependence of the clustering signal on the H I content, with some uncertainties for the lowest H I mass bin. For galaxies in this bin, a lower clustering signal is measured with respect to H I richer galaxies, but P13 argue this is not statistically significant due to the lower volume sampled by galaxies in this H I bin.

4.1 The projected correlation function

We compute the 2PCF for all models used in our study taking advantage of the mock light cones described in Section 2.3. To compute the predicted 2PCF, we use the Landy & Szalay (1993) estimator:

$$\xi(r) = \frac{{\rm DD}(r) + {\rm RR}(r) - 2{\rm DR}(r)}{{\rm RR}(r)}$$

where DD(r), RR(r) and DR(r) represent the galaxy–galaxy, random–random and galaxy–random number of galaxy pairs separated by a distance r. In observations, the physical separation r is not directly available, and the observables are the position on the sky and the recessional velocity. Using our mock light cones (Section 2.3), we mimic the data and carry out all calculations in redshift space. We measure the separation among two galaxies as $s = \sqrt{(v_1^2 + v_2^2 - 2v_1 v_2 \cos\theta)}/H_0$, where v_1 and v_2 are the recessional velocities of the galaxies (km s^{-1}), θ is the angle between them in the sky and H_0 is the Hubble constant. Hence, we calculate the correlation function using the separation along the line of sight ($\pi = |v_1 - v_2|/H_0$) and on the sky plane ($\sigma = \sqrt{\pi^2 - s^2}$) to obtain $\xi(\pi, \sigma)$. The projected correlation function used in the following discussion corresponds to

$$w(\sigma) = 2\int_0^\infty \xi(\pi, \sigma){\rm d}\pi.$$

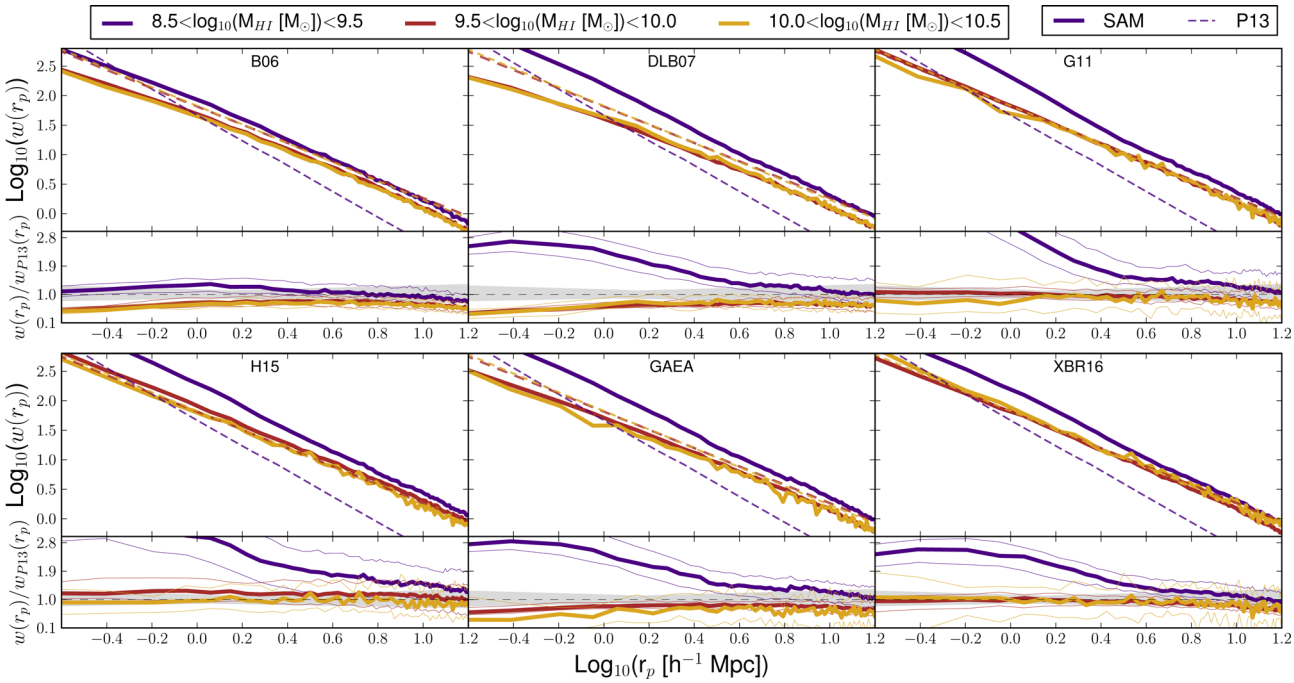

Figure 6. The two-point projected correlation function for different bins in H I mass (solid thick lines of different colours). Each panel corresponds to a different model, and the lower sub-panels show the ratio between the model predictions and the corresponding observational estimates (dashed lines) by P13 for the medium H I mass bin. As discussed in the text, this is equal to the one of the highest H I mass bin and more reliable than the measurements corresponding to the lowest H I mass bin. The 1σ scatter is shown in the lower sub-panels using thin lines for models and a shaded region for the data.

P13 had a non-uniform radial selection function for their sample and accounted for it in their random sample. We do not attempt to mimic the selection function of the data, and simply use for the random catalogue the same smooth redshift distribution of the selected model galaxies.

4.2 Model predictions

Fig. 6 shows the 2PCF for all models considered in this study. P13 found that the middle and highest H I mass bins ($M_{\rm HI} \in [10^{9.5};\ 10^{10}]$ and $[10^{10}; 10^{10.5}]\,{\rm M}_\odot$) have almost the same 2PCF, while they considered the measurements obtained for the lowest H I bin ($M_{\rm HI} \in [10^{8.5};\ 10^{9.5}]\,{\rm M}_\odot$) not reliable due to the small sampling volume. To better show the differences between model predictions and observational measurements, we show in the bottom sub-panels model results in each H I mass bin divided by the observational measurement corresponding to the bin $M_{\rm HI} \in [10^{9.5};\ 10^{10}]\,{\rm M}_\odot$.

For the lowest H I mass bin ($[10^{8.5}; 10^{9.5}]\,{\rm M}_\odot$), a problem is immediately evident: P13 found these galaxies to be less clustered than their H I richer counterparts, but argued that this might be due to finite volume effects. All models considered in this study, in contrast, predict a higher clustering signal for galaxies in this bin. The B06 model is the closest to the observational measurements, but still standing a little above them. All other models largely overpredict the measured clustering signal, particularly at small scales, with G11 and H15 deviating most from the data. As we will see in the following, this H I mass bin is sensitive to various physical prescriptions and to numerical resolution. In addition, it is dominated by satellite galaxies and therefore strongly dependent on the adopted treatment for satellite evolution. We verified how the incompleteness of the model sample influences the final 2PCF of this bin for the G11, H15 and XBR16 models using the MS-II: we find that the predicted clustering signal is lower than that found for the MS, but still a factor of ∼2 larger than that measured.

For H I richer galaxies ($[10^{9.5}; 10^{10}]$ and $[10^{10}; 10^{10.5}]\,{\rm M}_\odot$), B06, DLB07 and GAEA systematically underpredict the clustering signal, while G11, H15 and XBR16 are in good agreement with observational measurements in the corresponding mass bins.

The relatively noisy behaviour of the 2PCF in the highest H I bin is due to the small number of galaxies with such large H I masses.

As noted above, P13 argue that for the lowest H I mass bin, results are affected by smaller sampling volume. We can test the influence of a small sampling volume using our mocks. We show results of this test in Fig. 7. In this case, we use only the XBR16 model (results are similar for the other models) and show the 2PCF only for the lowest H I mass bin. The corresponding model predictions obtained by using only one-sixth of the volume are shown as a dot–dashed line, and are not statistically different from those obtained using the entire volume (solid line).

4.3 Halo occupation distribution

The results in the previous section can be understood by considering the number occupation of haloes of different mass by galaxies with different H I content, i.e. what is typically referred to as HOD. The HOD gives information about the distribution of galaxies in dark matter haloes of different masses, with the possibility to distinguish between centrals and satellites. It can be used to interpret the 2PCF at the scales of the halo dimensions. For this analysis, we use the data from the $z = 0$ snapshots to have a larger statistics.

Results from all models used in this study are shown in Fig. 8. Each panel shows the average number of galaxies with stellar mass larger than $M_* > 10^9\,{\rm M}_\odot$ (approximately corresponding to the resolution limit of the MS) and with different H I mass (different

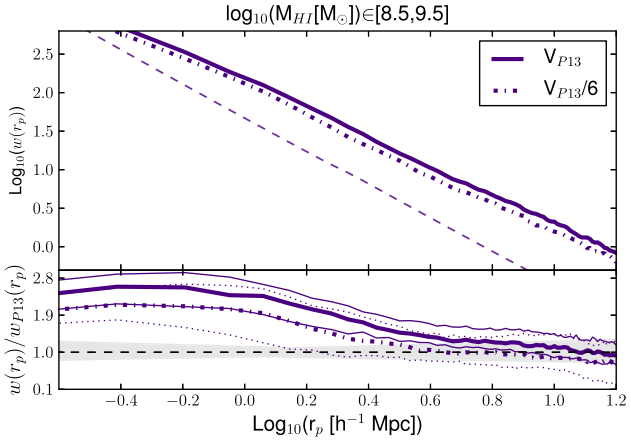

Figure 7. The 2PCF for the lowest H I mass bin for the XBR16 model. The solid line corresponds to model predictions for the full volume considered (i.e. this line corresponds to that shown in Fig. 6). The dot–dashed line corresponds to the results obtained when considering only one-sixth of the volume, as in the observational sample by P13. In the lower sub-panel, we show the ratio between model predictions and the observational estimate of the medium H I bin. Shaded regions and thin lines show the 1σ scatter.

columns) that reside in FoF haloes of mass $M_{200}^{\rm FoF}$. Blue and orange lines correspond to central and satellite galaxies, respectively, while black solid lines show the total. We checked the convergence of the results obtained for low-mass FoF haloes by considering the G11, H15 and the XBR16 model on the MS-II (results are shown as dashed lines in the corresponding rows). The HOD of the central galaxies has a Gaussian shape in B06 (as already noted in Kim et al. 2011), with basically no central galaxies for halo masses larger than $\sim 10^{12.5}$ M$_\odot$ h^{-1}. In the other models, the distributions corresponding to central galaxies are generally broader and extend to significantly larger halo masses than in B06. As noted above, H I in central galaxies of massive haloes is likely suppressed in B06 by efficient radio-mode feedback.

In Fig. 9, we show the mass function of FoF haloes hosting N galaxies with H I mass in different bins (different columns), namely the volume density (per $M_{200}^{\rm FoF}$ bin) of haloes hosting at least N centrals/satellites with a selected H I content. Different rows correspond to $N > 0$, $N > 1$ and $N > 10$. Galaxies are divided in centrals (solid lines) and satellites (dashed lines). Note that the FoF mass function for the central galaxies is shown only in the $N > 0$ row, as each FoF contains only one central galaxy by construction. Only a few haloes have large number of satellites containing large amounts of H I. The GAEA and XBR16 models in particular have significantly larger number of satellite galaxies with modest to significant H I content with respect to the other models. For all other models, the distributions peak at lower halo masses. It is interesting that GAEA and XBR16 predict a significant contribution from the H I gas-rich satellite population albeit assuming an instantaneous stripping of the hot gas associated with infalling galaxies. As noted above and in Hirschmann et al. (2016), this is driven by the modified stellar feedback scheme. For the intermediate and high H I mass bins considered ([$10^{9.5}$; $10^{10.5}$] M$_\odot$), there are less haloes hosting large numbers of galaxies in the GAEA model than in XBR16. We remind that these two models are based on the same physical parametrizations and differ only for the explicit modelling of the partition of cold gas in atomic and molecular gas, the division of the disc in annuli and on the explicit dependence of star formation on the molecular gas content in XBR16.

In the top row of Fig. 9, haloes hosting only one central galaxy dominate the low $M_{200}^{\rm FoF}$ end, and outnumber the FoF with H I-selected satellites up to $M_{200}^{\rm FoF} \sim 10^{12}$ M$_\odot$ h^{-1}. Above this limit, H I-rich satellites become the main contributors.

In Fig. 8, in the lowest H I mass bin ([$10^{8.5}$; $10^{9.5}$] M$_\odot$), satellites outnumber centrals at $M_{200}^{\rm FoF} \sim 10^{12}$ M$_\odot$ h^{-1} in the B06 and H15 models. In DLB07, G11, GAEA and XBR16, the number of satellites becomes larger than the corresponding number of centrals at $M_{200}^{\rm FoF} \sim 10^{11.5}$ M$_\odot$ h^{-1}. All models have more than 10 galaxies per FoF only above $M_{200}^{\rm FoF} \sim 10^{13}$ M$_\odot$ h^{-1} (B06 only at $\sim 10^{14}$). The volume density of such haloes is significant only for the GAEA and XBR16 models. B06 has the lowest number of haloes with more than one satellite in all the H I mass bins, but the difference in the lowest bin is more relevant, in particular compared to the DLB07 model, which instead is aligned with G11 and H15 both in the HOD and in the mass function. This results in different 2PCFs for B06 and the other models.

For the middle H I mass bin ([$10^{9.5}$; 10^{10}] M$_\odot$), we always find less than 10 galaxies per halo in the B06 and DLB07 models. In all other models, haloes with mass larger than $M_{200} \sim 10^{14}$ M$_\odot$ h^{-1} host more than 10 satellite galaxies. However, as for the lowest H I mass bin, the number of these haloes is large only for the GAEA and XBR16 models. In general, B06 and DLB07 have less satellite galaxies than the other models. This can explain the underestimation of the 2PCF: a lower number of satellite galaxies lowers the correlation signal. We remind that these models are characterized by a simple treatment for satellites and gas stripping (see Section 5 for details on satellite evolution) that leads to the well-known problem of too many passive galaxies (Weinmann et al. 2006; Wang et al. 2007; Fontanot et al. 2009). This simplified treatment is, however, assumed also in the GAEA and XBR16 models, in which the effect of the instantaneous stripping of gas is mitigated by a different treatment of stellar feedback.

There are generally very few galaxies in the highest H I mass bin ([10^{10}; $10^{10.5}$] M$_\odot$) considered. In particular, the B06, DLB07 and GAEA models always have only one galaxy per halo below halo masses $M_{200} \sim 10^{14.5}$ M$_\odot$ h^{-1}. For the other models, the number of galaxies becomes larger than one at halo masses larger than $M_{200} \sim 10^{13.7}$ M$_\odot$ h^{-1}. The only model with numerous satellites is XBR16. The difference in number of satellite-rich haloes does not determine an appreciable difference between the XBR16, G11 and H15 2PCFs. The smaller numbers of satellite-rich haloes in the GAEA model result in an underestimated 2PCF. The difference with respect to the XBR16 model can again be ascribed to the approach adopted, based on dividing the star-forming disc in annuli. As commented above, this results in a lower molecular fraction to total cold gas with respect to the GAEA model, and keeps the star formation ongoing in the central regions of the disc for longer times. This implies that satellite galaxies in XBR16 have more H I left than their counterparts in the GAEA model.

5 THE ROLE OF SATELLITE GALAXIES

The results discussed above suggest that satellite galaxies play an important role in the disagreement found between model predictions and observational measurements.

As discussed in Section 2, all models used in our study are built using subhalo-based merger trees extracted from the Millennium Simulation. Dark matter haloes are subject to significant stripping after being accreted on larger systems (e.g. De Lucia et al. 2004a; Gao et al. 2004). At the resolution of the MS, substructures fall below the resolution limit when at distances from halo centre that are

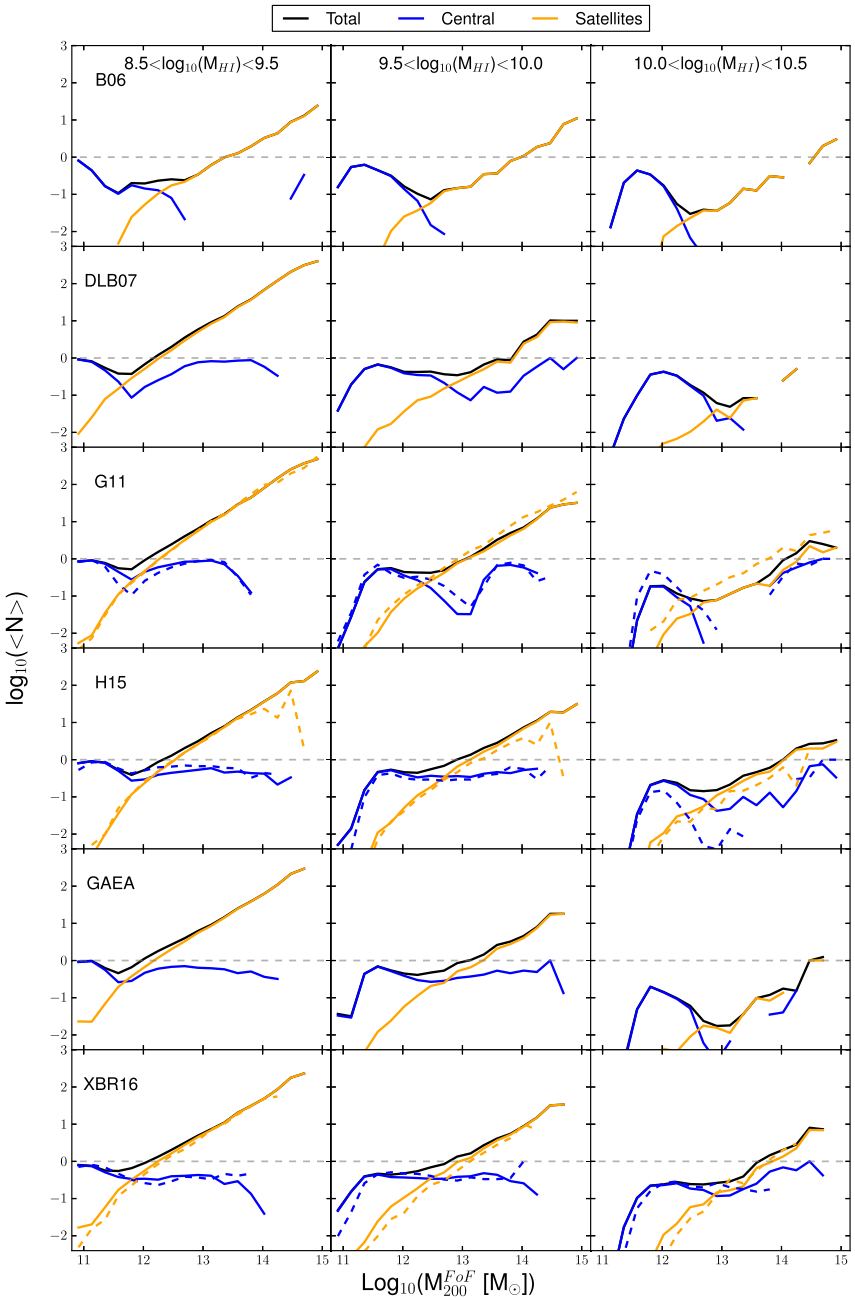

Figure 8. Average number of galaxies with stellar mass larger than $M_* > 10^9$ $M_\odot$ and in different H I mass bins (different columns) for haloes of mass $M_{200}^{\rm FoF}$. Different rows correspond to the different models considered in our study. The colour coding is the same as in Fig. 4: black for all model galaxies, blue for centrals and orange for satellites. In the rows corresponding to the G11, H15 and the XBR16 models, the dashed lines correspond to results based on the MS-II simulation.

typically significantly larger than the separation from which galaxy mergers are expected to occur. All models used in our study assume that galaxies in disrupted subhaloes survive as 'orphan' galaxies. While the specific treatment depends on the model (e.g. G11 and H15 include an explicit treatment for tidal disruption), they all assign to these galaxies a residual merger time that generally depends on the initial orbit and on the mass ratio between the infalling system and the accreting one.

When the infalling process begins, satellites undergo tidal processes that strip away a part or all their hot gas halo. In B06, DLB07, GAEA and XBR16, this process is instantaneous, and all the hot gas is stripped away (in B06, all the gas outside the dynamical radius of the halo) at the infall time (when the galaxy becomes a satellite). In G11 and H15, the stripping is gradual, and the hot gas that remains associated with satellite galaxies can cool providing fresh material for star formation.

The effect of satellite galaxies on the predicted correlation function of H I-selected galaxies is shown in Fig. 10. In this case, we are considering only central galaxies, i.e. we are excluding from galaxy catalogues both 'orphan' galaxies and satellite galaxies associated with distinct dark matter substructures. In all models, the clustering signal becomes weaker, with a shift dependent on the number of the

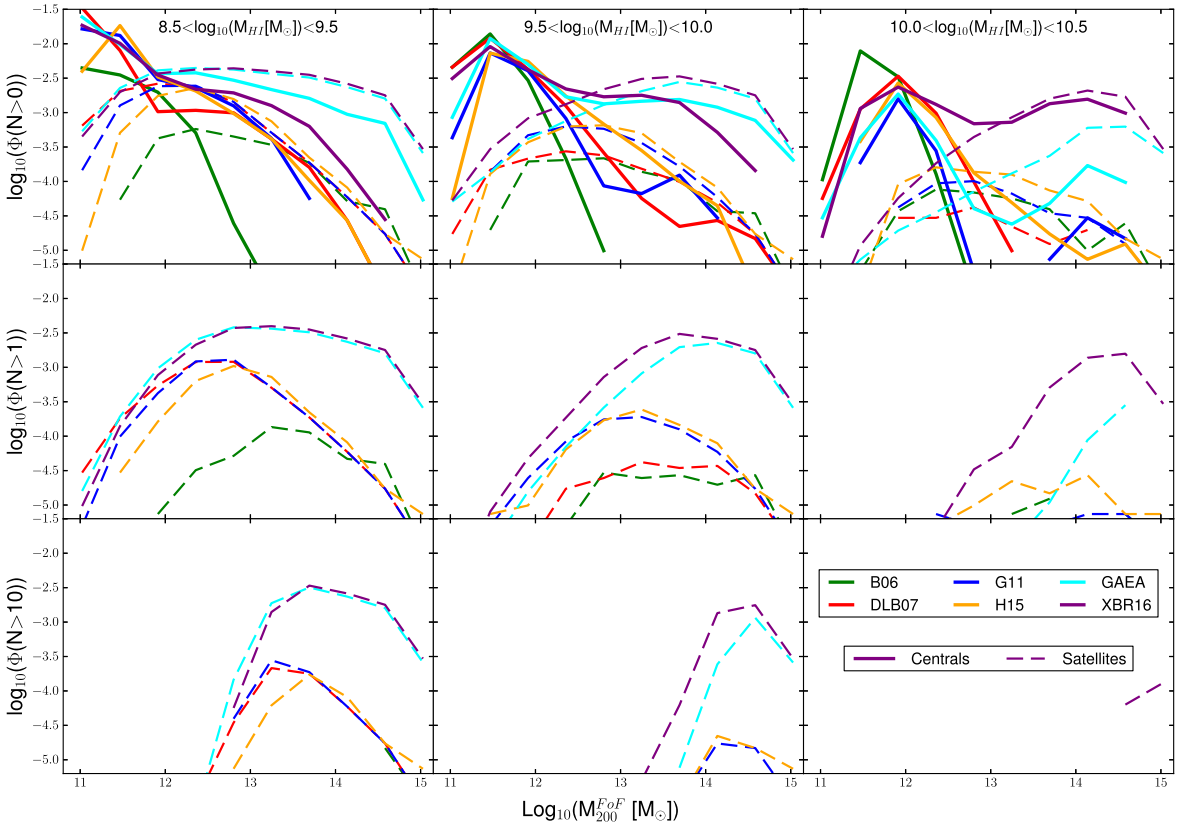

Figure 9. Mass function of FoF haloes hosting at least 1 (top panel), more than 1 (middle panel) and more than 10 (bottom panel) galaxies with H I mass in different bins (different columns). Solid lines are used for central galaxies while long-dashed lines correspond to satellite galaxies.

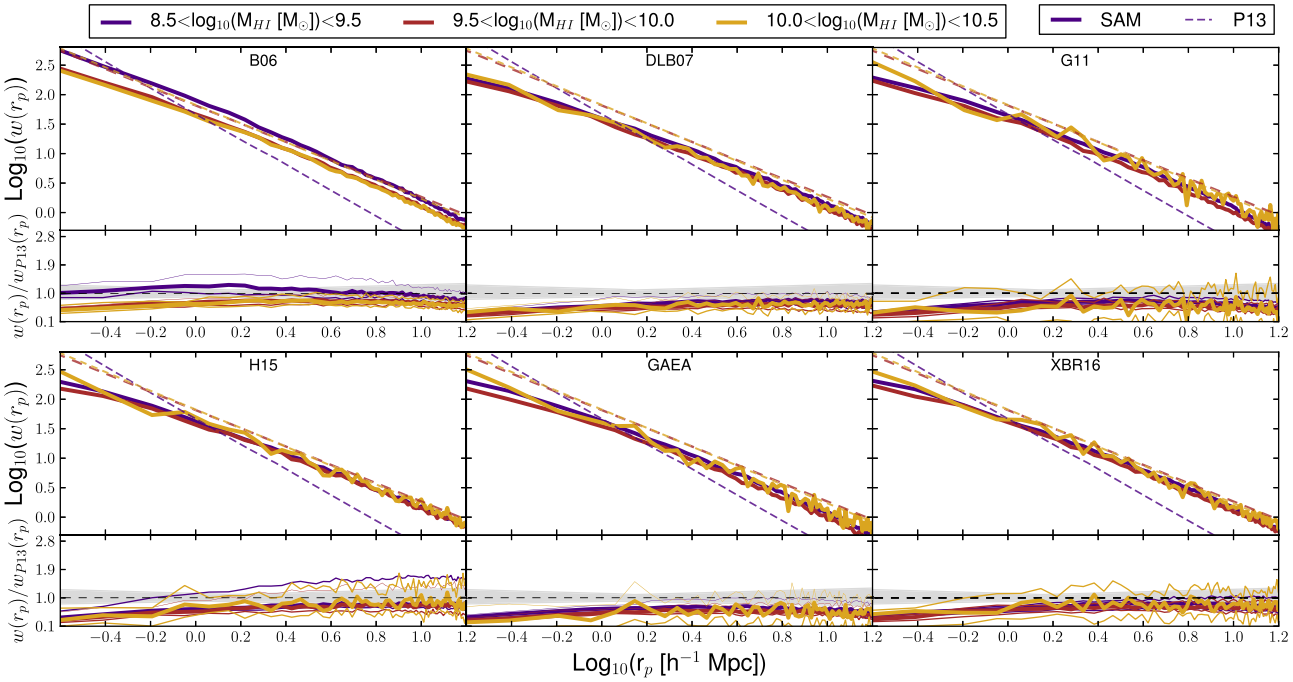

Figure 10. As in Fig. 6, but considering only central galaxies for all models.

satellites in each H I mass bin. For the B06 model, the 2PCF remains the same as if satellites are included. This is expected as Fig. 4 shows that satellite galaxies in this model are typically H I poor. Also for the DLB07 model, small differences are found between the clustering signal predicted including and excluding satellite galaxies in the medium-high H I bins ($[10^{9.5}; 10^{10.5}]\,M_{\odot}$). Again, this is expected because satellite galaxies are H I poor in this model, although the effect is not as strong as for the B06 model.

Table 1. Fraction of central and satellite galaxies in different H I mass bins, for all models considered in this study ($M_* > 10^9$ M$_\odot$).

Model	per cent of centrals/ satellites	per cent of central/satellites with $M_{\rm HI}$ [M$_\odot$] $\in$			
		$[0; 10^{8.5}]$	$[10^{8.5}; 10^{9.5}]$	$[10^{9.5}; 10^{10}]$	$[10^{10}; 10^{10.5}]$
B06	54.3–45.7	7.8–39.5	5.6–3.1	20.4–1.4	16.8–0.6
DLB07	54.4–45.6	1.5–19.8	33.0–24.6	15.6–1.2	4.0–0.1
G11	53.3–46.7	1.9–3.7	34.0–37.6	14.1–4.8	3.2–0.6
H15	61.4–38.6	0.9–2.3	33.7–28.8	20.2–6.3	5.7–1.0
GAEA	59.3–40.7	1.0–7.2	39.7–30.9	16.8–2.6	1.8–0.1
XBR16	57.0–43.0	4.7–12.7	35.0–25.8	15.0–4.1	2.3–0.4

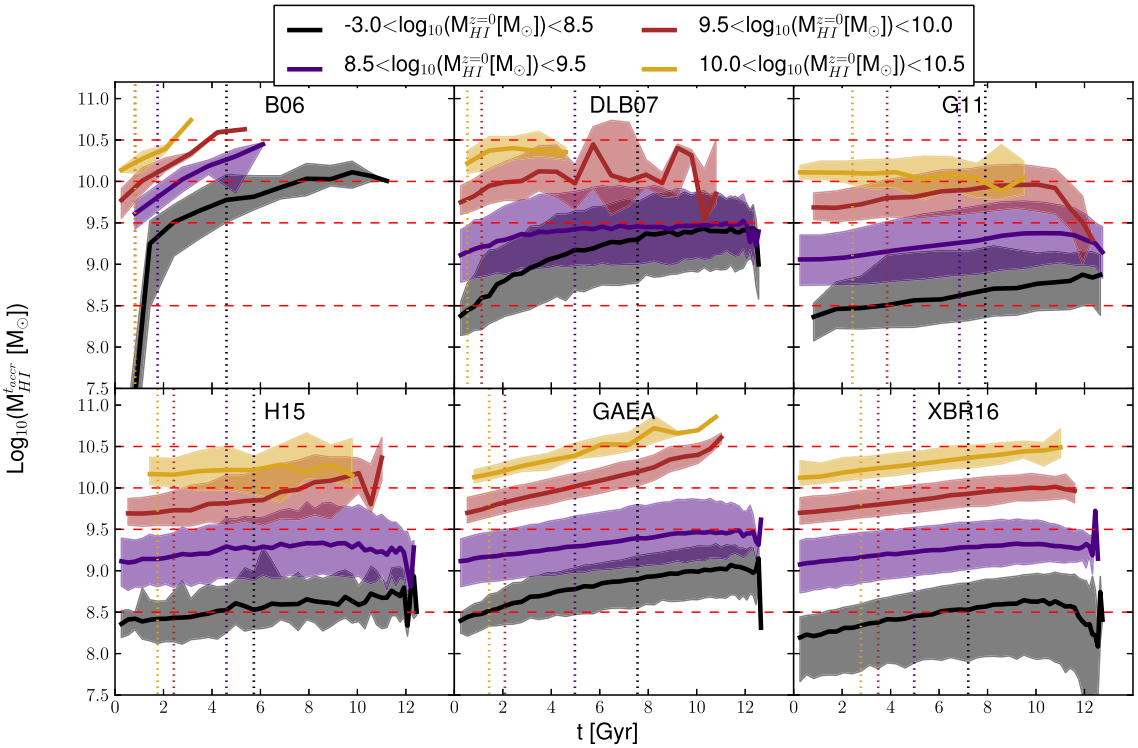

Figure 11. The H I content of satellites at the time of accretion versus the look-back time of accretion. Galaxies are divided according to their final H I content, with different colours representing different H I mass bins. The solid lines correspond to the median of the distributions, while the shaded areas show the one sigma scatter. The vertical dotted lines are the median times of accretion of model galaxies in each H I bin (colour coded as above). The red horizontal dashed lines correspond to the limits of the H I mass bins considered, and are plotted as a reference.

In Table 1, the fraction of satellites in each H I mass bin is listed for each model considered. Satellites account for 39–46 per cent of the total, depending on the model. The majority are in the lowest H I bin ($[10^{8.5}; 10^{9.5}]$ M$_\odot$) considered in this analysis, with the exception of the B06 model where satellites have typically lower H I masses.

Taking advantage of model results, we can also quantify the evolution of the H I content in satellite galaxies. To this aim, we have selected all satellite galaxies at $z = 0$ and followed back in time their main progenitors until they become central galaxies. In Fig. 11, we show the H I mass that satellite galaxies have at the last time they are centrals, as a function of look-back time. Galaxies are split in different bins according to the H I mass at present. We plotted as vertical dotted lines the median times of accretion for each H I bin. The figure shows that, for the two bins corresponding to the largest H I masses, accretion times tend to be lower than for galaxies with lower H I mass. In other words, the H I richest galaxies tend to be accreted later. This is particularly significant for the B06 and DLB07 models that are characterized by the most rapid consumption of the cold gas in satellites. The figure also shows that for some models (e.g. B06, DLB7, GAEA) the slope of the lines tend to be steeper than for the other models, indicating a more rapid depletion of the H I content of model galaxies.

A more direct way to quantify gas depletion in satellite galaxies is to choose a specific redshift of accretion and consider the average evolution of their H I content down to present time. Results of this analysis are shown in Fig. 12 for satellites accreted at $z \sim 1$. Different colours correspond to different H I masses at the time of accretion. The B06 model is characterized by the most rapid depletion rate. Also satellites in the DLB07 model consume their gas rapidly but they tend to flatten when the average H I content in satellites reaches a value $M_{\rm HI} < 10^{8.7}$ M$_\odot$. The figure also clearly shows the different satellite treatment in G11 and H15 compared to GAEA and XBR16: the latter are characterized by a faster depletion rate soon after accretion. In DLB07, G11, H15 and GAEA, the lines tend to flatten after reaching some value. This is due to fact that these models assume a critical surface density of gas for star formation. In the

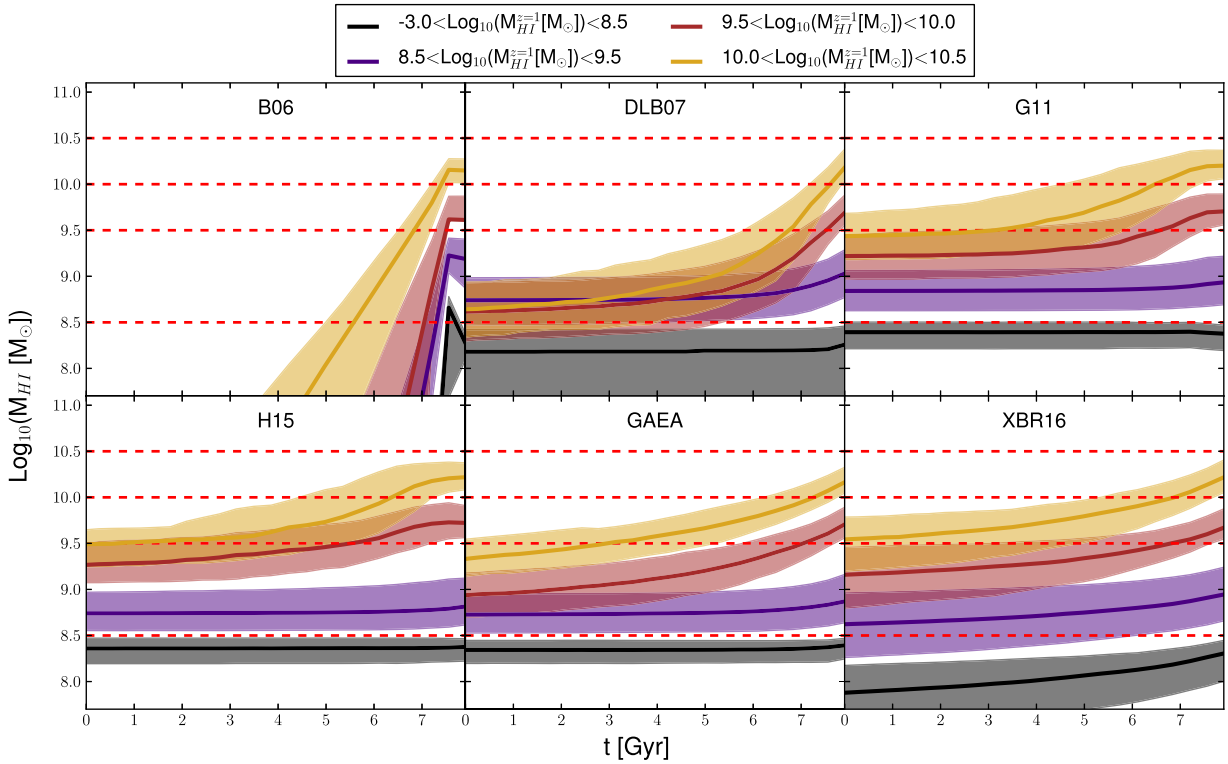

Figure 12. The H I content of satellites accreted at redshift 1 versus the look-back time. Galaxies are divided according to their initial H I content, with different colours representing different H I mass bins at $z = 1$. The solid lines correspond to the median of the distributions, while the shaded areas show the one sigma scatter. The red dashed lines correspond to the limits of the H I mass bins considered, and are plotted as a reference.

XBR16 model, the flattening is less evident. This model does not assume an explicit threshold for star formation but, as discussed above, evaluates the star formation rates in different annuli after estimating the amount of molecular gas available using the (Blitz & Rosolowsky 2006) empirical relation.

6 RELATIONS WITH THE DARK MATTER HALO

In this section, we analyse the relation between H I-selected galaxies and the hosting dark matter haloes. We take advantage of our knowledge of the halo mass to characterize the H I–$M_{\rm halo}$ relation in Section 6.1, and we analyse the dependence of H I content on the dark matter halo spin in Section 6.2.

In the following, for satellite galaxies, we will consider the mass and spin of the parent dark matter halo corresponding to the last time the galaxy was central.

6.1 H I GALAXY CONTENT AND MAXIMUM HALO MASS

P13 used the measured 2PCF for the H I-selected galaxies to estimate the shape and the scatter of the H I-$M_{\rm halo}$ relation. They took advantage of the Bolshoi dark matter only simulation (Klypin, Trujillo-Gomez & Primack 2011), based on a *WMAP7* cosmology, and populated haloes and subhaloes with H I through abundance matching. In their analysis, they linked the current ($z = 0$) H I content of each subhalo, $M_{\rm H\,I}^{z=0}$, to the maximum value of the subhalo mass during its past evolution, $M_{200}^{\rm max}$ (this corresponds with good approximation to the time just before a halo is accreted on a larger structure). The implied assumption is that the H I attached to subhaloes at their maximum mass does not change too much down to redshift 0. They note, but do not discuss further, that while this assumption can be valid for stellar masses, it is generally not a good one for the H I masses, because of ram-pressure stripping and gas consumption due to quiescent star formation in satellite galaxies.

In our work, we take advantage of our knowledge of the H I content of model galaxies both at the time the halo mass was maximum (we assume this corresponds to the last time the galaxy was central), and at redshift 0. We can thus verify the influence of the different satellite treatments on the predicted relation between H I mass and halo mass.

In Fig. 13, we show the relation between the H I mass and the halo mass at the last time the galaxy was central. This figure represents what we would find if we suppose the H I content of galaxies does not change after accretion, as done in P13. Solid lines show the median relation for all galaxies (black), centrals (blue) and today satellites (orange). The 1σ spread is shown by the shaded area (for all galaxies) and the dashed lines (for central and satellites). The relation obtained in P13 is overplotted in red (solid line is the median and dashed lines show the 1σ spread) for a direct comparison. In this figure, the relation obtained for satellite galaxies is very close to that obtained for centrals, with some small differences due to redshift evolution of the H I content of galaxies at the time of accretion (see Fig. 11).

As already noted for the H I–M_* relation (see Section 3.3), B06 fails to reproduce the H I content in medium- to high-mass haloes because satellites are generally too gas-poor and central galaxies are depleted of their cold gas by efficient radio-mode feedback. The other models exhibit a wide scatter in the relation that is almost independent of $M_{200}^{\rm max}$, with values of $\sigma_{\rm H\,I} \sim 0.3-0.5$ dex depending on the model. The shape is qualitatively similar to the one inferred in P13 if we consider H15, GAEA and XBR16,

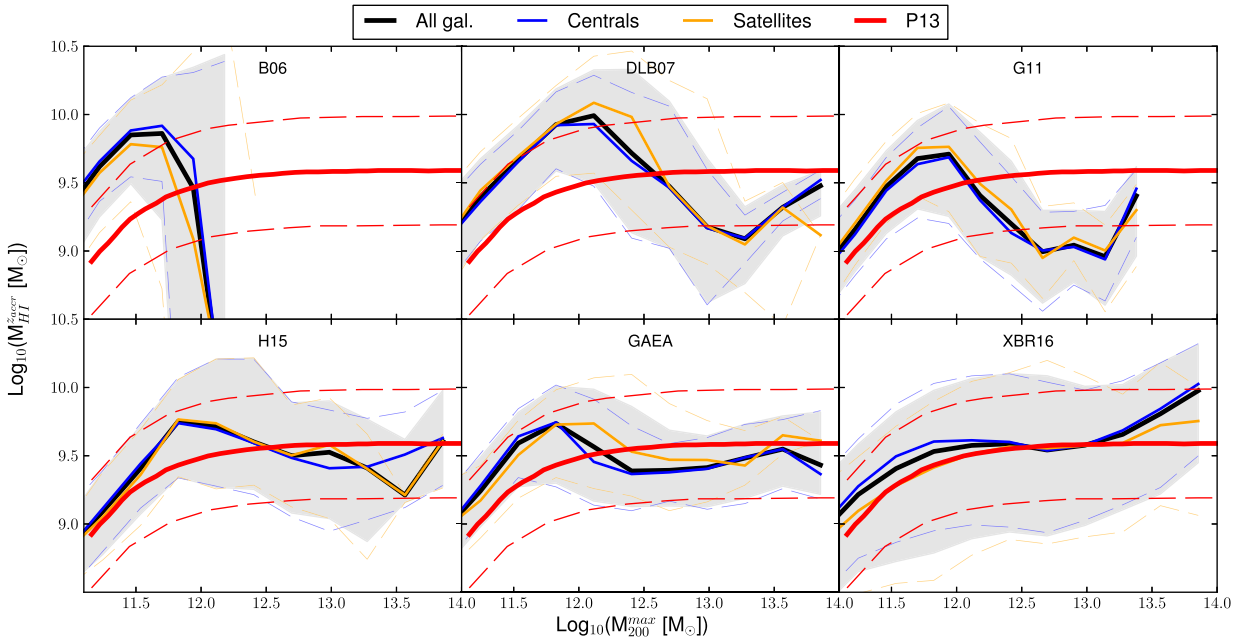

Figure 13. The H I–halo mass relation for the models considered in this study, for central (blue), satellite (orange) and all (black) galaxies. The relation inferred by P13 is shown for comparison (red). The H I mass shown on the y-axis is that measured at the time the parent halo mass peaks (i.e. just before infall). Solid lines show the median values, shaded region corresponds to the 1σ scatter for all galaxies, while dashed lines correspond to the 1σ scatter for centrals and satellites.

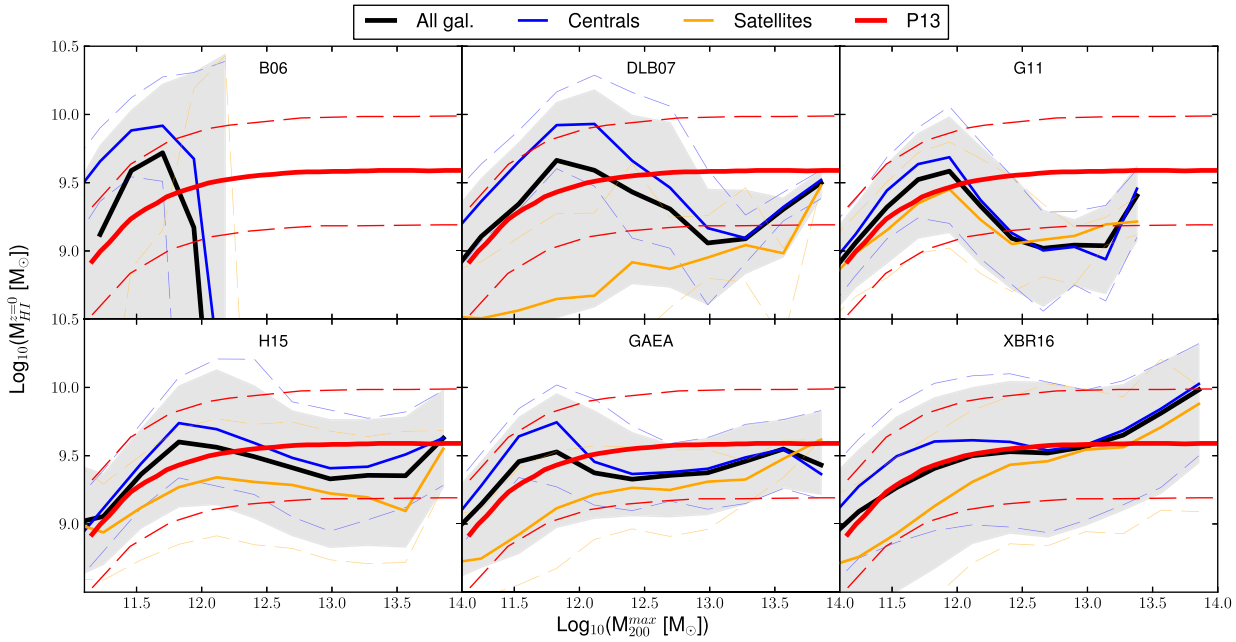

Figure 14. As in Fig. 13, but the H I mass shown on the y-axis is that measured from the models at redshift 0.

while the other models predict more H I for $M_{200}^{\rm max} \lesssim 10^{12}$ M$_\odot$, and less H I than that inferred by P13 in the most massive haloes.

In Fig. 14, we take into account the evolution of H I in satellites, that is, we used for all galaxies their H I content at redshift 0, while $M_{200}^{\rm max}$ corresponds again to the parent halo mass at the last time the galaxy was central. In this figure, we can appreciate a downshift of 0.1–0.4 dex of the relation for $M_{200}^{\rm max} < 10^{12.5-13}$ M$_\odot$. The shift is driven by the satellites with the lowest masses (both halo and H I), whose H I is depleted after accretion because consumed through star formation. The gas depletion increases the scatter of the total relation, with the effect being enhanced for B06, DLB07 and GAEA.

As noted in Section 5, the lowest H I bin considered ($[10^{8.5}; 10^{9.5}]$ M$_\odot$) is dominated by satellite galaxies, and therefore the one most affected by a different treatment for the evolution of these galaxies.

6.2 H I GALAXY CONTENT AND HALO SPIN

A relatively tight correlation is expected between the values of the halo spin and the gas content of galaxies. Huang et al. (2012) studied this relation for the ALFALFA H I-selected galaxies. The spin of the dark matter halo was calculated using the λ estimator proposed by Hernandez et al. (2007), assuming a dark matter isothermal density profile, an exponential surface density profile for the stellar disc and a flat disc rotation curve. Using these assumptions: $\lambda \propto r_{\rm disc}/V_{\rm rot}$, thus the spin depends on the disc scaleradius and the rotational velocity. Huang et al. (2012) found that ALFALFA H I-rich galaxies favour high-λ values. Kim & Lee (2013) simulated the evolution of dwarf-size haloes with varying halo-spin parameters and initial baryon fractions, and found a correlation between disc radius (and therefore gas mass) and λ.

P13 used their measurements of the 2PCF to infer a relation between halo spin and gas content. They divided the haloes of the Bolshoi simulation into three conveniently chosen spin bins: low spin with $\lambda \in [0.002; 0.02]$, medium spin with $\lambda \in [0.02; 0.05]$ and high spin, with $\lambda \in [0.05; 0.20]$. Assigning an H I content to each halo according to the relation presented in Section 6.1, they found that the high- and medium-spin bins have almost the same 2PCFs measured for the H I-selected galaxies, while low-spin haloes have larger clustering signal. Based on these results, they argued that halo spin is the main driver of the H I content of galaxies. We have verified that selecting haloes in the same range of spin values used by P13, we find results consistent with theirs.

Since the models we have used in our study are coupled with a high-resolution cosmological simulation (MS), we can explicitly analyse the correlation between the H I content of model galaxies and the spin of their parent haloes. For the B06 model, information about the spin is not available, so we exclude this model from the following analysis. For central galaxies, we just use the spin of the parent halo. For satellite galaxies, we consider the spin measured for the parent halo at the last time they were central galaxies.

We show the histograms of the spin for some chosen $M_{200}^{\rm max}$ (different columns) and for the usual ranges in H I content in Fig. 15. Dashed lines correspond to the distributions obtained for each H I mass, independently of $M_{200}^{\rm max}$, in order to easily compare each M_{200}^{max} bin to the total. The red vertical lines correspond to the bins used in P13.

The majority of the galaxies with the lowest H I mass ($[10^{8.5}; 10^{9.5}]\,{\rm M}_{\odot}$) are found in haloes with low-spin values ($\lambda \in [0.002; 0.02]$), while the majority of galaxies in the two H I richer bins ($[10^{9.5}; 10^{10.5}]\,{\rm M}_{\odot}$) correspond to mid-spin values ($\lambda \in [0.02; 0.05]$). Interestingly, this distribution is very similar for all models, i.e. the H I-spin correlation/distribution is not dependent on the specific prescriptions of each model. This result can be understood as follows: the halo spin parameter is used to set the initial value of the disc scalelength, and this quantity is then used to compute a density threshold for star formation. In the case of a low-spin halo, the initial radius will be small, and the surface density will be large. The star formation rate is directly dependent on the surface density and thus, for a fixed value of H I, a smaller radius will correspond to a faster H I depletion. This explains why H I-rich galaxies are rare in haloes with low-spin values, and why high-spin (large disc radius) haloes host galaxies with a broad range of H I masses.

The figure shows that, in all models considered in our study, there is no tight correlation between the H I content of galaxies and the spin of their parent haloes: low-spin haloes are more likely populated by galaxies with low H I mass, and H I-rich galaxies are most likely hosted by haloes with large spin values, but high-spin haloes are populated by galaxies in a wide range of H I mass.

7 CONCLUSIONS

In this work, we study the basic statistical properties of H I-selected galaxies extracted from semi-analytic models of galaxy formation, and compare theoretical predictions with available data. In particular, we use four models whose galaxy catalogues are publicly available (B06; DLB07; G11; H15), and two models recently developed by our group (Hirschmann et al. 2016, and Xie et al. 2006). Only one of these models (that described in Xie et al. 2006) includes an explicit modelling for the partition of cold gas into atomic and molecular components, and a molecular hydrogen based star formation law. All models are run on the same cosmological simulation, the Millennium Simulation (Springel et al. 2005).

For those model in which the cold gas is treated as a single star-forming phase, we estimate the H I content of model galaxies a posteriori assuming (as in Obreschkow & Rawlings 2009) that the cold gas is distributed in an exponential disc and that the ratio between molecular and atomic hydrogen is determined by the kinetic gas pressure. All models used in our study include different specific modelling for the various physical processes considered and, in particular, for the evolution of satellite galaxies. We find this to have relevant consequences on model predictions for H I-selected galaxies.

All models considered are in relatively good agreement with the observed local H I mass function, with the exception of the model by B06 that predict too many galaxies with intermediate to large H I mass and too few galaxies with small H I content ($M_{\rm H\,I} < 10^{9.6}{\rm M}_{\odot}$). We find this is due to excessive H I masses in low stellar mass galaxies. This particular model also fails to reproduce the observed scaling relations as it predicts very little H I associated with satellites (likely because of instantaneous stripping of hot gas and efficient stellar feedback), and central galaxies with stellar masses larger than $\sim 3 \times 10^{10}\,{\rm M}_{\odot}$ (because of too efficient radio mode feedback). The observed scaling relations are relatively well reproduced by all other models, but some of them (Guo et al. 2011; Hirschmann et al. 2016, and Xie et al. 2006) exhibit a deficit of H I-rich galaxies ($M_{\rm H\,I} > 10^{10}\,{\rm M}_{\odot}$) at any stellar mass.

The H I content of satellite galaxies varies significantly among the models considered, due to a different treatment for the hot gas associated with infalling galaxies and stellar feedback. As expected, models that assume a non-instantaneous stripping of this hot gas reservoir tend to predict larger H I masses for satellite galaxies. The most massive satellites, in particular, tend to have an average H I content that is very close to that of central galaxies of the same stellar mass. This is due to the fact that these galaxies were accreted relatively recently and evolved as central galaxies for most of their lifetime (De Lucia et al. 2012). Interestingly, assuming an instantaneous stripping of the hot gas at the time of accretion, as in GAEA, does not necessarily imply low H I content for satellite galaxies. As already noted in Hirschmann et al. (2016), this model is characterized by significantly lower fractions of passive (and therefore larger fractions of gas-rich, star-forming) satellites with respect to e.g. the model presented in DLB07. This is a consequence of suppressed and delayed star formation at early times, and leads to larger cold gas fractions at the time of accretion.

Using galaxy catalogues from each model, we have built mock light cones that we have used to analyse how the clustering of H I-selected galaxies compares to recent measurements by P13. In

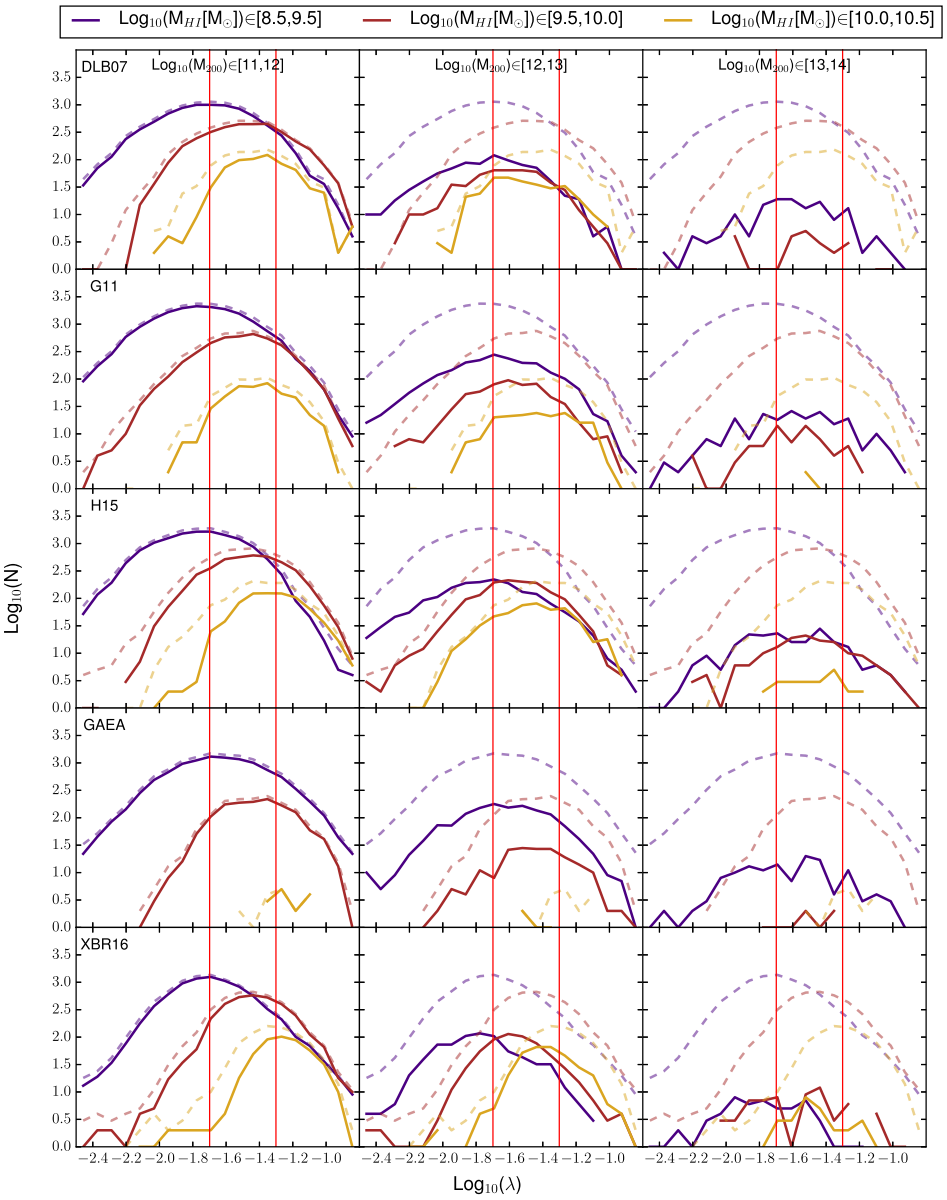

Figure 15. Distribution of halo spin values for galaxies selected in haloes with different $M_{200}^{\rm max}$ (columns) and H I mass (same colours of Fig. 6), for all models used in our study (different rows). The solid lines represent the distribution for each selection, while dashed lines are the total distribution over all the $M_{200}^{\rm max}$ range, reproduced in every panel as reference. The vertical red lines show the spin division considered by P13, and are plotted as reference.

particular, we have considered three H I mass bins: low ($M_{\rm HI} \in [10^{8.5};\ 10^{9.5}]\ {\rm M}_\odot$), intermediate ($M_{\rm HI} \in [10^{9.5};\ 10^{10.}]\ {\rm M}_\odot$) and high ($M_{\rm HI} \in [10^{10};\ 10^{10.5}]\ {\rm M}_\odot$). The lowest H I mass bin is likely affected by limited sampling volume in the observations so, although the data suggest a lower clustering signal for this particular bin, P13 argue that this is not significant and that the two-point clustering function measured for these three bins is not statistically different. In contrast, we find that all models predict for galaxies in the lowest H I bin a clustering signal *higher* than for the H I richer galaxies. For the other two bins ($[10^{9.5}; 10^{10.5}]\,{\rm M}_\odot$), half of the models are in relatively good agreement with data while the other half tend to underpredict slightly the measured clustering. Interestingly, the model by B06 that has the worst performance for the H I mass distribution and scaling relations exhibits the best agreement with data by P13 in the lowest H I bin, with a clustering signal only slightly stronger than that for H I richer galaxies. We show that the lowest H I bin is strongly affected by the adopted treatment for satellite galaxies as this bin is dominated by this galaxy population.

The relation between the H I mass (at the accretion time for satellite galaxies) and halo mass (at its maximum) predicted by all models considered is in quite good agreement with that inferred by P13. Again, the exception is the model by B06 that predicts negligible H I in haloes more massive than $\sim 10^{12}\ {\rm M}_\odot$. This is interesting because, as noted above, the clustering signal predicted by this particular model is the one that is closest to observational measurements. Thus, taken at face value, these results suggest that the clustering of H I-selected galaxies does not provide enough information to constrain the relation between halo mass and H I mass, and that a

crucial element is represented by the evolution of the H I content of satellite galaxies. The scatter of the predicted relation increases in case one considers the H I associated with galaxies at present time, because of gas depletion in satellite galaxies. Specifically, we find that the B06 model exhibits the shortest gas consumption times: a galaxy accreted at $z \sim 1$ with H I mass $\sim 10^{10}\,M_{\odot}$ conserves only about 10 per cent of this gas after 2 Gyr (1 per cent after 3 Gyr). This is likely due to an efficient stellar feedback, coupled with an instantaneous stripping of the hot gas associated with infalling galaxies. The gas consumption time-scale is typically longer in the other models, and the assumption of a gas density threshold for star formation implies that the gas associated with satellite galaxies never falls below such limit. The gas consumption time-scales are longest in the Xie et al. (2006) model where star formation is evaluated in radial annuli, considering the local physical conditions of the ISM.

Finally, we have examined the relation between the H I content of galaxies and the spin of the parent dark matter halo (we have considered the value at the time of accretion for satellite galaxies). We find that low-spin haloes ($\lambda \in [0.002; 0.02]$) are more likely populated by H I-poor galaxies ($[10^{8.5}; 10^{9.5}]\,M_{\odot}$) and H I richer galaxies tend to reside in haloes with large spin ($\lambda \in [0.02; 0.2]$). The scatter, however, is relatively large and haloes with intermediate and large spin values tend to host galaxies with a large dynamic range in H I mass. Interestingly, the distributions are very similar in all models considered, i.e. they are not significantly affected by the specific modelling of the various physical processes affecting the H I content of galaxies. This is somewhat surprising as the spin enters the calculation of the disc radius and this, in turn, affects the star formation rate (and therefore the gas content). In most models, however, the halo spin is only used to determine the initial radius of the gaseous component and does not affect significantly the subsequent evolution. The dependence of the gas initial disc radius on halo spin explains the trends found: an initial small radius and a large gas fraction translates into a high surface density and therefore into a large star formation rate, which consumes the gas rapidly. This explains why haloes with small spin values tend to be associated with gas-poor galaxies.

Our analysis show that different models lead to very similar results for galaxies with intermediate to large H I mass ($[10^{9.5}; 10^{10.5}]\,M_{\odot}$), while significant differences can be found for relatively H I-poor galaxies ($[10^{8.5}; 10^{9.5}]\,M_{\odot}$). As discussed above, this bin is dominated by satellite galaxies and therefore mostly affected by the different treatment for this particular galaxy population. More detailed data in this particular H I mass range are needed to put stronger constraints on galaxy formation models. Dedicated controlled simulations would be useful to quantify the effect of stripping processes in satellites (see e.g. Tonnesen & Bryan 2009). Some attempts in this direction were made using semi-analytic models based on Monte Carlo merger trees (Lagos et al. 2011; Kim et al. 2015) or on the Millennium Simulation II (Kim et al. 2016), and have shown that these models can be used to constrain the physics of low-mass satellites.

At the same time, we need larger statistical samples and the possibility to estimate the galaxy 'hierarchy' (i.e. being a satellite or a central) in observations. New radio instruments, such as SKA[4] and its precursors (Johnston et al. 2008; Booth et al. 2009, for ASKAP and MeerKAT, respectively), will provide valuable data for these analyses.

[4] https://www.skatelescope.org/project/

ACKNOWLEDGEMENTS

The Millennium Simulation data bases used in this paper and the web application providing online access to them were constructed as part of the activities of the German Astrophysical Virtual Observatory (GAVO). This work has been supported by the MERAC foundation and by the PRIN-INAF 2012 grant 'The Universe in a Box: Multi-scale Simulations of Cosmic Structures'. We acknowledge a CINECA award under the ISCRA initiative, for the availability of high-performance computing resources and support. MH acknowledges financial support from the European Research Council via an Advanced Grant under grant agreement no. 321323 NEOGAL.

REFERENCES

Abazajian K. N. et al., 2009, ApJS, 182, 543
Amorín R., Muñoz-Tuñón C., Aguerri J. A. L., Planesas P., 2016, A&A, 588, A23
Angulo R. E., Hilbert S., 2015, MNRAS, 448, 364
Angulo R. E., White S. D. M., 2010, MNRAS, 405, 143
Baugh C. M., Lacey C. G., Frenk C. S., Benson A. J., Cole S., Granato G. L., Silva L., Bressan A., 2004, New Astron. Rev., 48, 1239
Benson A. J., Bower R. G., Frenk C. S., Lacey C. G., Baugh C. M., Cole S., 2003, ApJ, 599, 38
Bernardi M., Meert A., Sheth R. K., Vikram V., Huertas-Company M., Mei S., Shankar F., 2013, MNRAS, 436, 697
Blaizot J., Wadadekar Y., Guiderdoni B., Colombi S. T., Bertin E., Bouchet F. R., Devriendt J. E. G., Hatton S., 2005, MNRAS, 360, 159
Blitz L., Rosolowsky E., 2006, ApJ, 650, 933
Bolatto A. D. et al., 2011, ApJ, 741, 12
Booth R. S., de Blok W. J. G., Jonas J. L., Fanaroff B., 2009, preprint (arXiv:0910.2935)
Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645 (B06)
Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, MNRAS, 398, 1150
Catinella B. et al., 2010, MNRAS, 403, 683
Catinella B. et al., 2013, MNRAS, 436, 34
Christensen C., Quinn T., Governato F., Stilp A., Shen S., Wadsley J., 2012, MNRAS, 425, 3058
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Crain R. A. et al., 2016, MNRAS, preprint (arXiv:1604.06803)
Croton D. J. et al., 2006, MNRAS, 365, 11
Davé R., Katz N., Oppenheimer B. D., Kollmeier J. A., Weinberg D. H., 2013, MNRAS, 434, 2645
De Lucia G., Blaizot J., 2007, MNRAS, 375, 2 (DLB07)
De Lucia G., Kauffmann G., Springel V., White S. D. M., Lanzoni B., Stoehr F., Tormen G., Yoshida N., 2004a, MNRAS, 348, 333
De Lucia G., Kauffmann G., White S. D. M., 2004b, MNRAS, 349, 1101
De Lucia G., Weinmann S., Poggianti B. M., Aragón-Salamanca A., Zaritsky D., 2012, MNRAS, 423, 1277
De Lucia G., Muzzin A., Weinmann S., 2014, New Astron. Rev., 62, 1
Fontanot F., De Lucia G., Monaco P., Somerville R. S., Santini P., 2009, MNRAS, 397, 1776
Fu J., Guo Q., Kauffmann G., Krumholz M. R., 2010, MNRAS, 409, 515
Gao L., White S. D. M., Jenkins A., Stoehr F., Springel V., 2004, MNRAS, 355, 819
Guo Q. et al., 2011, MNRAS, 413, 101 (G11)
Guo Q., White S., Angulo R. E., Henriques B., Lemson G., Boylan-Kolchin M., Thomas P., Short C., 2013, MNRAS, 428, 1351
Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R. E., Guo Q., Lemson G., Springel V., 2013, MNRAS, 431, 3373
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 (H15)
Hernandez X., Park C., Cervantes-Sodi B., Choi Y.-Y., 2007, MNRAS, 375, 163
Hirschmann M., De Lucia G., Fontanot F., 2016, MNRAS, 461, 1760

Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, MNRAS, 445, 581
Huang S., Haynes M. P., Giovanelli R., Brinchmann J., 2012, ApJ, 756, 113
Hunt L. K. et al., 2015, A&A, 583, A114
Johnston S. et al., 2008, Exp. Astron., 22, 151
Kennicutt R. C., Jr1989, ApJ, 344, 685
Keres D., Yun M. S., Young J. S., 2003, ApJ, 582, 659
Kim J.-h., Lee J., 2013, MNRAS, 432, 1701
Kim H.-S., Baugh C. M., Benson A. J., Cole S., Frenk C. S., Lacey C. G., Power C., Schneider M., 2011, MNRAS, 414, 2367
Kim H.-S., Wyithe J. S. B., Power C., Park J., Lagos C. d. P., Baugh C. M., 2015, MNRAS, 453, 2315
Kim H.-S., Wyithe J. S. B., Baugh C. M., Lagos C. d. P., Power C., Park J., 2016, MNRAS, preprint (arXiv:1603.02383)
Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ, 740, 102
Kuhlen M., Madau P., Krumholz M. R., 2013, ApJ, 776, 34
Lagos C. D. P., Baugh C. M., Lacey C. G., Benson A. J., Kim H.-S., Power C., 2011, MNRAS, 418, 1649
Landy S. D., Szalay A. S., 1993, ApJ, 412, 64
Lemson G., Virgo Consortium t., 2006, preprint (arXiv:e-prints)
Leroy A. K., Walter F., Brinks E., Bigiel F., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2782
Leroy A. K. et al., 2011, ApJ, 737, 12
McGaugh S. S., 2005, ApJ, 632, 859
Magdis G. E. et al., 2011, ApJ, 740, L15
Martin A. M., Papastergis E., Giovanelli R., Haynes M. P., Springob C. M., Stierwalt S., 2010, ApJ, 723, 1359
Martin A. M., Giovanelli R., Haynes M. P., Guzzo L., 2012, ApJ, 750, 38
Meyer M. J., Zwaan M. A., Webster R. L., Brown M. J. I., Staveley-Smith L., 2007, ApJ, 654, 702
Moustakas J. et al., 2013, ApJ, 767, 50
Narayanan D., Krumholz M. R., Ostriker E. C., Hernquist L., 2012, MNRAS, 421, 3127
Obreschkow D., Rawlings S., 2009, ApJ, 696, L129
Papastergis E., Giovanelli R., Haynes M. P., Rodríguez-Puebla A., Jones M. G., 2013, ApJ, 776, 43 (P13)
Peeples M. S., Shankar F., 2011, MNRAS, 417, 2962
Planck Collaboration XVI, 2014, A&A, 571, A16
Popping A., Davé R., Braun R., Oppenheimer B. D., 2009, A&A, 504, 15
Power C., Baugh C. M., Lacey C. G., 2010, MNRAS, 406, 43
Rafieferantsoa M., Davé R., Anglés-Alcázar D., Katz N., Kollmeier J. A., Oppenheimer B. D., 2015, MNRAS, 453, 3980
Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726
Springel V. et al., 2005, Nature, 435, 629
Tonnesen S., Bryan G. L., 2009, ApJ, 694, 789
Wang L., Li C., Kauffmann G., De Lucia G., 2007, MNRAS, 377, 1419
Wang J., De Lucia G., Kitzbichler M. G., White S. D. M., 2008, MNRAS, 384, 1301
Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., Croton D. J., Moore B., 2006, MNRAS, 372, 1161
West A. A., Garcia-Appadoo D. A., Dalcanton J. J., Disney M. J., Rockosi C. M., Ivezić Ž., 2009, AJ, 138, 796
West A. A., Garcia-Appadoo D. A., Dalcanton J. J., Disney M. J., Rockosi C. M., Ivezić Ž., Bentz M. C., Brinkmann J., 2010, AJ, 139, 315
White S. D. M., Frenk C. S., 1991, ApJ, 379, 52
Xie L., De Lucia G., Hirschmann M., Fontanot F., Zoldan A., 2016, MNRAS, preprint (arXiv:1611.09372)
Zwaan M. A., Meyer M. J., Staveley-Smith L., Webster R. L., 2005, MNRAS, 359, L30

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