A tale of two populations: the stellar mass of central and satellite galaxies

Eyal Neistein,1⋆ Cheng Li,2 Sadegh Khochfar,1 Simone M. Weinmann,3 Francesco Shankar4 and Michael Boylan-Kolchin5

1Max-Planck-Institute for Extraterrestrial Physics, Giessenbachstrasse 1, 85748 Garching, Germany
2Max-Planck-Institute Partner Group, Key Laboratory for Research in Galaxies and Cosmology, Shanghai Astronomical Observatory, Chinese Academy of Sciences, Nandan Road 80, Shanghai 200030, China
3Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, the Netherlands
4Max-Planck-Institute for Astrophysics, Karl-Schwarzschild-Str. 1, 85748 Garching, Germany
5Center for Galaxy Evolution, 4129 Reines Hall, University of California, Irvine, CA 92697, USA

Accepted 2011 May 27. Received 2011 May 23; in original form 2011 March 17

ABSTRACT

We develop a new empirical methodology to study the relation between the stellar mass of galaxies and the mass of their host subhaloes. Our approach is similar to abundance matching, and is based on assigning a stellar mass to each subhalo within a large cosmological N-body simulation. However, we significantly extend the method to include a different treatment for central and satellite galaxies, allowing the stellar mass of satellite galaxies to depend on both the host subhalo mass, and on the halo mass. Unlike in previous studies, our models are constrained by two observations: the stellar mass function of galaxies, and their autocorrelation functions (CFs). We present results for ∼106 different successful models, showing that the uncertainty in the host subhalo mass reaches a factor of ∼10 for a given stellar mass. Our results thus indicate that the host subhalo mass of central and satellite galaxies is poorly constrained by using information coming solely from the abundance and CFs of galaxies. In addition, we demonstrate that the fraction of stellar mass relative to the universal baryon fraction can vary between different models, and can reach ∼0.6 for a specific population of low-mass galaxies. We conclude that additional observational constraints are needed, in order to better constrain the mass relation between haloes and galaxies. These might include weak lensing, satellite kinematics or measures of clustering other than the CFs. Moreover, CFs based on larger surveys with a better estimate of the systematic uncertainties are needed.

Key words: galaxies: abundances – galaxies: evolution – galaxies: formation – galaxies: haloes – galaxies: luminosity function, mass function – galaxies: statistics – large-scale structure of Universe.

1 INTRODUCTION

The relation between the stellar mass of galaxies and their host dark matter haloes has become a key point of reference for many different theoretical and observational studies. It summarizes in a simple way the complexity of galaxy formation physics when evolved within growing dark matter structure. A special attention was given to the relation between galaxies and their host subhaloes, which are the substructure bound density peaks inside a halo.

The ‘abundance matching’ (hereafter ABM) methodology is an important theoretical tool for constraining the mass relation between galaxies and their host subhaloes (Kravtsov et al. 2004; Vale & Ostriker 2004; Conroy, Wechsler & Kravtsov 2006; Shankar et al. 2006; Behroozi, Conroy & Wechsler 2010; Guo et al. 2010; Moster et al. 2010). In this simple empirical approach one assigns a model galaxy to each subhalo within a cosmological N-body simulation. Assuming there is a one-to-one, monotonic relation between the stellar mass (m*) and the subhalo mass (Minfall, see Section 2.1), the abundance of galaxies and subhaloes can be matched, yielding a unique relation between m* and Minfall. Surprisingly, this simple model provides a good fit to the observed clustering properties of galaxies.

The ABM approach thus offers a practical solution to the relation between subhaloes and galaxies, without going into the complex details of galaxy formation physics. It can be used to constrain the mass relation at various redshifts to predict the star formation rate in galaxies (Conroy & Wechsler 2009), to study the merger rates of galaxies (Hopkins et al. 2010) and to interpret high-resolution hydrodynamical simulations (Sawala et al. 2011). In comparison to
models based on the halo occupation distribution (HOD; Berlind & Weinberg 2002; Cooray & Sheth 2002; Tinker et al. 2005; Zehavi et al. 2005, and references therein), ABM uses explicit information on the location and mass of subhaloes, decreasing the number of free parameters needed in the model.

The success of ABM is intriguing and raises several interesting questions: Is it based on the only possible set of assumptions that can reproduce the abundance and clustering of galaxies? Do we miss models that result in a different \( m_\text{infall} - M_\text{stellar} \) relation? What are the important assumptions made by ABM? How can we explore these assumptions and test to what level they are constrained by observations? In this work we try to address these questions. We specifically focus our results on the freedom in the \( m_\text{infall} - M_\text{stellar} \) relation.

In ABM, the stellar mass is assigned to each subhalo according to \( M_\text{stellar} \). For a satellite subhalo, \( M_\text{stellar} \) is defined as the mass at the last time it was the most massive substructure within its \( \text{FOF} \) group. This is a reasonable assumption because the subhalo mass of satellite galaxies can be significantly stripped after falling into a larger dark matter halo (e.g. van den Bosch, Tormen & Giocoli 2005; Zentner et al. 2005), more so than the stellar mass of its galaxy (for stellar stripping see Monaco et al. 2006; Conroy, Wechsler & Kravtsov 2007; Purcell, Bullock & Zentner 2007; Yang, Mo & van den Bosch 2009a). On the other hand, for a central subhalo, \( M_\text{stellar} \) is defined as its current mass. Wang et al. (2006) found that the relation between \( M_\text{stellar} \) and the stellar mass of galaxies is tight in semi-analytic models (SAMs), justifying the above definition of \( M_\text{stellar} \).

There are various other assumptions made by ABM, which are mainly related to the treatment of satellite galaxies. In a recent paper (Neistein et al. 2011, hereafter Paper I), we have examined these assumptions using the SAM of Neistein & Weinmann (2010). Although ABM models assume that for a given \( M_\text{stellar} \), the stellar mass of central and satellite galaxies is the same, there are various effects within SAMs that violate this assumption. First, the relation between stellar mass and subhalo mass evolves with redshift for central galaxies, affecting satellite galaxies at the time of infall (the typical infall time is a few Gyr ago; Wang et al. 2007). Second, the stellar mass of satellite galaxies might be different already at the time of infall from that of central galaxies at the same time. This is because galaxies that join larger groups as satellites are located in large-scale environment of higher density than the average. Consequently, the dark matter merger histories of these galaxies are already different at early epochs (see also Gao, Springel & White 2005; Harker et al. 2006). Third, once a galaxy becomes a satellite, its stellar mass might still grow. This is especially true for models in which gas stripping in satellite galaxies is modelled on timescales of a few Gyr (Weinmann et al. 2010). All the effects above are consistent with various studies about the properties of central and satellite galaxies (Weinmann et al. 2006a; von der Linden et al. 2007; Khochfar & Ostriker 2008; Pasquali et al. 2009; Skibba & Sheth 2009; Skibba et al. 2011).

In Paper I, we showed that the effects above can change not only the \( m_\text{infall} - M_\text{stellar} \) relation for satellite galaxies but also the autocorrelation function (CF) of galaxies. If satellite and central galaxies with a given \( M_\text{stellar} \) are randomly redistributed in a SAM, the CFS can vary by up to a factor of 4. Even when just redistributing satellite and central galaxies among themselves, the modifications in the CFS can reach a factor of 2 for massive galaxies. We have shown in Paper I that the CFS of SAM galaxies can be reproduced accurately by the ABM approach only when the stellar mass of satellite galaxies is assumed to depend on both \( M_\text{stellar} \) and the host halo mass at \( z = 0 \). This finding is very useful as it saves us the complex modelling of \( m_\text{infall} \) as a function of the various effects mentioned above.

In this paper we make use of the conclusions made in Paper I, and add more ingredients to an ABM model. We would like to study the influence of the new ingredients on the relation between \( M_\text{stellar} \) and \( m_\text{infall} \), while the models are constrained to fit the observational data. We allow the stellar mass of satellite galaxies to depend on both the \( M_\text{stellar} \) and the host halo mass, and to deviate from the behaviour of central galaxies. In addition, we use two ingredients for modelling the location and abundance of satellite galaxies: subhaloes that were lost by the cosmological simulation at high redshift, but can still host galaxies at \( z = 0 \), are identified using an estimate for dynamical friction. The location of unresolved subhaloes is fixed using either the location of the most bound particle of the last identified subhalo, or using an analytical model with a free parametrization.

Once the population of galaxies is broken into two subpopulations, the model cannot be constrained by the abundance of galaxies alone. The number of models that can fit the observed stellar mass function of galaxies is infinite. We therefore check a very large number of models (~10^{12}) and look for those that fit both the observed stellar mass function, and the observed CF of galaxies. We will show that even when using these two constraints, there is a significant amount of freedom in the relation between \( m_\text{infall} \) and \( M_\text{stellar} \). Satellite galaxies might occupy a significantly lower value of \( M_\text{stellar} \) and thus might be more abundant than central galaxies at a fixed \( m_\text{infall} \). The dependence of satellite galaxies on the halo mass is able to compensate for this effect in terms of clustering, and is crucial for fitting the observed CFS.

The observational constraints used in this work include the stellar mass function of galaxies, as derived by Li & White (2009). For the CFS of galaxies we use the methodology presented by Li et al. (2006), using the same stellar masses as in Li & White (2009). These stellar masses are based on redshift and five-band photometric data, as described in detail by Blanton & Roweis (2007). The galaxy sample is based on the final data release (DR7; Abazajian et al. 2009) of the Sloan Digital Sky Survey (SDSS; York et al. 2000). Although there might be both random and systematic deviations in the stellar masses used here, the two observational constraints are self-consistent.

This paper is organized as follows. Section 2 describes in detail the approach we use in this work. We elaborate on the new ingredients used here, and the way we implement the models. In Section 3 we demonstrate the results of our formalism by showing a few different models that fit the data well. Our method for scanning the parameter space is discussed in Section 4, where we explain how the parameter space is sampled, and how we select good models. In this section we also show the properties of the successful models. Lastly, we summarize our results and discuss them in Section 5.

2 THE FORMALISM

In this section we describe the formalism developed for modelling the abundance and clustering of galaxies. In general, we assume that the stellar mass of galaxies depends solely on the properties of the host dark matter haloes. This assumption is similar to the ABM approach (e.g. Vale & Ostriker 2004). However, we significantly extend the ingredients of the model with respect to previous studies. This is motivated by the analysis done in Paper I, based on the SAMs of Neistein & Weinmann (2010). The additional ingredients
are related to the assumptions regarding satellite galaxies: their abundance, location and the effect of the host halo mass on their stellar mass.

We start by summarizing the various ingredients of the formalism, a detailed description of each component is given in the following subsections.

(i) All the subhaloes from a cosmological N-body simulation at \( z = 0 \) are selected.

(ii) In addition, we use subhaloes that have been merged into a bigger structure at high redshift, but might host galaxies that survive until \( z = 0 \). This is due to the effects of dynamical friction and stripping on galaxies, which are not included in the N-body simulation.

(iii) We use a freely tuneable pre-factor in the dynamical friction formula.

(iv) We assign one galaxy with a specific stellar mass (\( m_\ast \)) to each subhalo from the full sample defined in items (i) and (ii).

(v) For central subhaloes within their FOF group, we assume that the stellar mass of galaxies depends on the host subhalo mass at \( z = 0 \) only (\( M_{\text{infall}} \)).

(vi) For satellite subhaloes, we assume that the stellar mass of galaxies depends on two parameters: the host subhalo mass at the time of infall, and the host halo mass at \( z = 0 \).

(vii) We allow the dependence of stellar mass on the subhalo mass to be different for central and satellite subhaloes.

(viii) The location of subhaloes that are not found in the simulation at \( z = 0 \) is set by either the location of their most bound particle, or by an analytical model.

Items (iii), (vi), (vii) and (viii) in the list are new with respect to previous ABM models, and their influence on clustering and abundance of galaxies has not been tested before.

2.1 Subhaloes and merger trees

We use merger trees extracted from the Millennium N-body simulation (Springel 2005). This simulation was run using the cosmological parameters \((\Omega_m, \Omega_\Lambda, h, \sigma_8) = (0.25, 0.75, 0.73, 0.9)\), with a particle mass of \(8.6 \times 10^8 \, h^{-1} \, M_\odot\) and a box size of \(500 h^{-1} \, \text{Mpc}\). The merger trees are based on snapshots spaced by \(250 \, \text{Myr}\), linking subhaloes identified by the SUBFIND algorithm (Springel et al. 2001). Subhaloes are defined as the bound density peaks inside FOF groups (Davis et al. 1985). More details on the simulation and the subhalo merger trees can be found in Springel et al. (2005) and Croton et al. (2006).

The subhalo mass, \( M_\text{h} \), corresponds to the number of particles inside a subhalo, as identified by SUBFIND. The infall mass of the subhalo, \( M_{\text{infall}} \), is defined as

\[
M_{\text{infall}} = \begin{cases} 
M_\text{h} & \text{if central within its FOF group,} \\
M_{\text{h},p}(\text{central}) & \text{otherwise.}
\end{cases}
\]

Here \( M_{\text{h},p}(\text{central}) \) is the lowest redshift at which the main progenitor of the subhalo \( M_\text{h} \) was the most massive within its FOF group, and \( M_{\text{h},p} \) is the main progenitor mass at this redshift.

In addition to \( M_{\text{infall}} \), we will use the halo mass \( M_{200} \), which is defined as the mass within the radius where the halo has an overdensity 200 times the critical density of the simulation. In general, the mass \( M_{200} \) should include both the central subhalo within a group, and all of its satellite subhaloes. However, due to the spherical symmetry forced on \( M_{200} \) and the specific overdensity used here, it might deviate from the exact FOF group mass. In what follows, we will refer to \( M_{200} \) as the ‘halo mass’ to indicate that this mass is computed over a larger spatial region than the subhalo mass. For satellite subhaloes, we assign the same value of \( M_{200} \) as is computed for their central subhalo within the same group. Because of the effects above, the value of \( M_{\text{infall}} \) is often similar or higher than \( M_{200} \) for central subhaloes.

2.2 Satellite subhaloes: dynamical friction and location

Satellite subhaloes are defined as all subhaloes inside a FOF group except the most massive (central) subhalo. Within the N-body simulation, satellite subhaloes lose their mass while falling into a bigger subhalo, and thus might fall below the resolution limit used in the SUBFIND algorithm (20 particles for the Millennium simulation used here). However, the galaxies residing inside these satellite subhaloes might live longer, as they are more dense, and thus less vulnerable to stripping. This effect is significant even for relatively high-resolution cosmological simulations, as was shown in Paper I. It can modify the abundance of subhaloes even at two orders of magnitude above the minimum subhalo mass resolved by the simulation.

In order to take this effect into account, we model the time it takes the galaxy to fall into the central galaxy by dynamical friction. At the last time the satellite subhalo is resolved we compute its distance from the central subhalo \( r_{\text{fall}} \), and estimate the dynamical friction time using the formula of Binney & Tremaine (1987),

\[
t_{\text{df}} = \frac{1.17 \text{V}_c^2 r_{\text{fall}}^3}{G M_{h,1} \ln \left(1 + M_{h,1}/M_{h,2}\right)}.
\]

Here \( M_{h,1} \) is the mass of the central subhalo, \( V_c \) is its virial velocity and \( M_{h,2} \) is the mass of the satellite subhalo. Once a satellite subhalo falls together with its central subhalo into a larger group, we update \( t_{\text{df}} \) for both objects according to the new central subhalo.

The dynamical friction estimate is computed using the mass of subhaloes only, before galaxies are being modelled. Therefore, the formula deviates from its proper definition, as it does not include the effect of the galaxy mass on dynamical friction. However, since this formula uses various simplified assumptions with a larger uncertainty (e.g. the exact trajectories of the satellite subhaloes are ignored, see Boylan-Kolchin, Ma & Quataert 2008), one general constant, \( \alpha_{\text{df}} \), is being used here to absorb all the related uncertainties. SAMs often use \( \alpha_{\text{df}} \sim 2–3 \) (De Lucia & Blaizot 2007; Khochfar & Silk 2009; Neistein & Weinmann 2010) in agreement with more detailed studies of dynamical friction processes (Colpi, Mayer & Governato 1999; Boylan-Kolchin, Ma & Quataert 2008; Jiang et al. 2008; Mo, van den Bosch & White 2010). Since we ignore contribution due to the galaxy mass, we allow more freedom in \( \alpha_{\text{df}} \) as is discussed below.

To summarize, three types of subhalo exist in our models:

(i) central subhaloes: most massive subhaloes within their FOF group;

(ii) satellite subhaloes: all subhaloes except central subhaloes;

(iii) unresolved subhaloes: subhaloes that were last identified at high \( z \), and are added according to the dynamical friction formula above. All the unresolved subhaloes are also satellite subhaloes.

\footnote{Main-progenitor histories are derived by following back in time the most massive progenitor in each merger event.}

\footnote{For central subhaloes, \( M_{200} \) is slightly smaller than \( M_{\text{infall}} \). The average difference is \(-0.08 \, \text{dex}\), with an rms scatter of \(-0.06 \, \text{dex}\).}
The location of unresolved subhaloes is not given by the simulation. We therefore need to estimate the location of the galaxies that are assigned to these subhaloes. This issue was treated by e.g. Croton et al. (2006) who used the location of the most bound particle of the last identified subhalo, as is given by the dark matter simulation at \( z = 0 \). The same location is adopted here as well. However, we would like to take into account the uncertainty involved in this method. Obviously, one dark matter particle inside a collisionless \( N \)-body simulation cannot mimic accurately the location of an extended galaxy, with a possibly very different mass. We therefore use the following analytical model as an additional option:

\[
    r = r_{\text{infall}} (1 - \tau^p)^{1/q}.
\]

Here \( \tau \) is the fraction of time spent out of all the estimated dynamical friction time until \( z = 0 \), \( r \) is the distance we adopt at \( z = 0 \) from the central subhalo and \( p, q \) are constants. In Appendix A we derive this model by following the angular momentum of a particle inside a spherical gravitational potential of the type \( \rho \propto r^{-\gamma} \). We show that for a specific values of \( p \) and \( q \), the model is able to reproduce the location of subhaloes as is modelled by the most bound particle. This parametrization enables us to check various modifications to the location of unresolved subhaloes.

### 2.3 Counting subhaloes

The usual way to implement ABM is to first populate subhaloes with galaxies, and only then to compute the CFs of galaxies. In our approach there are many possible models, mainly depending on the different mass relations for satellite and central galaxies. Since computing the statistics of pairs is the most demanding computational step, it is not possible to scan a significant number of models in the usual technique. Here we present a new way of computing the CFs which is extremely efficient when many models are needed. We first compute the statistics of subhalo pairs, and save them as a function of \( M_{\text{infall}} \) and \( M_{200} \). Only then are the galaxy CFs computed. In this section we explain how the statistics of subhaloes is defined and computed, the next subsection discusses the way this information is used to model galaxies.

We want to constrain our models against the stellar mass function of galaxies. In order to do so we will use the mass function of central and satellite subhaloes:

\[
    \phi_{c}(M_{\text{infall}}) = \frac{1}{V} \frac{dN^c}{d \log M_{\text{infall}}},
\]

\[
    \phi_{s}(M_{\text{infall}}, M_{200}) = \frac{1}{V} \frac{d^2 N^s}{d \log M_{\text{infall}} d \log M_{200}}.
\]

Here \( V \) is the volume of the simulation box, and \( N^c, N^s \) are the numbers of central and satellite subhaloes, respectively. In Fig. 1 we show the two-dimensional mass function \( \phi_s \) for a model with \( \alpha_{\text{df}} = 3 \). A similar behavior as presented here is valid for \( 0.1 \leq \alpha_{\text{df}} \leq 10 \), where the low-mass contour can vary by \( \sim 0.5 \) dex, and the contour of massive subhaloes is hardly affected.

In order to compute the CFs of galaxies, we start by computing the number of subhalo pairs, \( N_p \), for the total sample of subhaloes. In case both subhaloes within the pair are central subhaloes, we count this pair into \( \psi_{cc} \):

\[
    \psi_{cc} (M_{\text{infall}}^1, M_{\text{infall}}^2, r) = \frac{1}{V^2} \frac{d^3 N_{cc}^c}{d \log M_{\text{infall}}^1 d \log M_{\text{infall}}^2 d \log r}.
\]

Here \( M_{\text{infall}}^1, M_{\text{infall}}^2 \) are the infall mass of the first and second subhaloes in the pair, and \( r \) is the distance between these subhaloes within the \( x\text{-}y \) plane. Distances are computed only within the \( x\text{-}y \) plane in order to compute the projected CF, as is described below. In practice we divide the range in \( \log M_{\text{infall}} \) and \( \log (r) \) into 100 and 50 bins, respectively, and save \( \psi_{cc} \) as a multidimensional histogram.

In a similar way we define the pair statistics of central–satellite and satellite–satellite subhaloes:

\[
    \psi_{cs} (M_{\text{infall}}^1, M_{200}^2, r),
\]

\[
    \psi_{ss} (M_{\text{infall}}^1, M_{200}^1, M_{200}^2, r).
\]

Note that for satellite subhaloes the number of pairs is saved as a function of both \( M_{\text{infall}} \) and \( M_{200} \). This is done in order to properly model the dependence of stellar mass on \( M_{200} \). It should be emphasized that the statistics of satellite subhaloes, i.e. \( \phi_s, \psi_{cs}, \psi_{ss} \), depend on the dynamical friction constant, \( \alpha_{\text{df}} \), and on the location we adopt for unresolved subhaloes (\( p \) and \( q \) from equation 3).

### 2.4 Definition of a model – domains in stellar mass

Previous studies have often used an analytic functional form to describe the relation between \( M_{\text{infall}} \) and \( m_* \). For example, Moster et al. (2010) suggested

\[
    m_* = f (M_{\text{infall}}) = c_1 M_{\text{infall}} \left[ c_2 M_{\text{infall}} + M_{\text{infall}}^4 \right]^{-1},
\]

where \( c_i \) are all constants. A straightforward way to extend this approach here would be to parametrize \( f \) as a function of two variables, \( m_* = f (M_{\text{infall}}, M_{200}) \), and to use a different set of parameters for modelling central and satellite galaxies. However, this approach requires a priori knowledge of \( f \), and the resulting solutions might be restricted by the specific functional form chosen. This is especially true when \( f \) is allowed to be different for satellite and central galaxies, so the freedom in its functional shape might be larger than what was found in previous studies (see the examples in Section 3).

Here we suggest a new way to parametrize the relation between \( M_{\text{infall}} \) and \( m_* \), which is motivated by the observational data. The observed CFs of galaxies are computed over mass bins of width 0.5 dex in stellar mass (hereafter ‘domains’). For example, the CF for small-mass galaxies is based on galaxies with stellar mass in the
range

\[ 10^{0.27}, 10^{0.77} \] \( M_\odot \) \( M \). It might well be that a more complex behaviour would add a significant amount of freedom to the models.

In order to model the stellar mass function, we will use the same domains, and demand that the stellar mass function will be reproduced once integrated over each domain. This guarantees that in case it is needed, a detailed solution of the type \( m_\star = f(M_{\text{infall}}, M_{200}) \) exists. However, since our models allow the mass relation above to deviate between central and satellite galaxies, the detailed behaviour of \( m_\star \) within each domain is not well constrained. In general, there might be many different interpolations of the kind \( m_\star = f(M_{\text{infall}}, M_{200}) \) within each domain (between adjacent \( U_i/s \)). These will not change the computed CF, and will fit the observed stellar mass functions. The range in stellar mass for each domain is relatively small, so this freedom is negligible in comparison to the results we will show below.

In general, modelling a scatter in \( m_\star \) for a given \( M_{\text{infall}} \) and \( M_{200} \) is possible within our formalism. However, it demands a detailed knowledge of the functional form, \( m_\star = f(M_{\text{infall}}, M_{200}) \). This addition does not allow us to scan the different models in a very efficient way. We therefore do not treat such a scatter in this work. None the less, the dependence of \( m_\star \) on \( M_{200} \) for satellite galaxies results in a variation in \( m_\star \) as a function of \( M_{\text{infall}} \). This effect will be discussed below.

\[ 3 \text{ A FEW EXAMPLES} \]

In this section we present a few specific models that were found using a large parameter search within our formalism. The details of the search will be presented in the next section. Here we discuss a few examples to demonstrate the parametrization needed for each model, and to highlight the variety of models that are able to fit both the stellar mass function and CFs.

In Figs 2 and 3 we present model A, one example out of a family of models which are extreme with respect to the full population of models. In this model, \( m_\star \) is much higher for satellite subhaloes than for central subhaloes, for a given \( M_{\text{infall}} \). The difference can reach a factor of 100 in stellar mass. As can be seen from Fig. 2, this model includes a relatively strong dependence of \( m_\star \) on \( M_{200} \), a feature that was originally seen in the SAMs analysed in Paper I, although less prominent. We note that previous ABM models were not constructed a full realization of galaxies per each model. This is in contrast to standard ABM models, where a list of all subhaloes within a given simulation is necessary.

2.5 How to compute CFs?

To summarize, each model in this work is defined by the following parameters.

- (a) The value of \( \alpha_{\text{df}} \) used in the dynamical friction formula, as defined by equation (2).
- (b) The values \( p, q \) used in equation (3) for modelling the location of unresolved subhaloes. Alternatively, we use the location as given by the most bound particle.
- (c) For central subhaloes, six values of \( \alpha \) that define the domain boundaries \( U_i \) (equation 13).
- (d) For satellite subhaloes, six values of \( \beta \), and six values of \( \delta \) that correspond to the \( U_i \) boundaries.

Once the parameters \( \alpha_{\text{df}}, p, q \) are chosen, we construct the subhalo statistical functions \( \phi \) and \( \psi \). The parameters in items (c) and (d) above are then used as integration limits for \( \phi \) and \( \psi \). By integrating \( \phi \) we compute the total number of subhaloes within the domain:

\[ N = V \int_{U_1}^{U_2} \phi \ d \log M_{\text{infall}} \]

\[ + V \int_{U_1}^{U_2} \phi \ d \log M_{\text{infall}} d \log M_{200}. \]  

A similar integration of \( \psi \) over each domain results in \( N_{\psi}(r_p) \) – the total number of pairs within each radial bin \( r_p \). The projected CF, \( w_p(r_p) \), is then defined as the deviation in the number of pairs from the average value per volume:

\[ w_p(r_p) = \left[ \frac{L^2 V^2 N_p(r_p)}{A_p} - 1 \right] L. \]  

Here \( A_p \) is the two-dimensional area covered by the radial bin \( r_p \), and \( L \) is the size of the simulation box in \( h^{-1} \) Mpc.

We have tested numerically that our methodology agrees with the standard ABM approach for various different models. In our approach the models are based on quantifying \( \phi \) and \( \psi \) only, so we do not construct a full realization of galaxies per each model. This is in contrast to standard ABM models, where a list of all subhaloes within a given simulation is necessary.
satellite and central galaxies might be very large, more than a factor of 10.

We emphasize that our definition of a ‘satellite’ versus ‘central’ is valid only for subhaloes. It might be that satellite subhaloes will host more massive galaxies than the central subhalo within the same group. In these cases the more massive galaxy in the group might be identified as a ‘central’ galaxy, although its host subhalo is defined as a ‘satellite’ here. In a recent study, Skibba et al. (2011) have estimated the fraction of haloes that host central galaxies that are not the most luminous in their group. Using a group catalogue based on SDSS (Yang et al. 2007) they found that this fraction reaches 25 (40) per cent for low (high) mass haloes. Thus, a different identification of ‘central’ and ‘satellite’ may be used either in observational studies, or in the analysis of hydrodynamical simulations. In our models, the difference between satellite and central subhaloes is defined solely according to the dark matter behaviour, and is motivated by the different merger history of subhaloes, and their final location.

From Fig. 7 it is evident that model C predicts a non-negligible number of low-mass subhaloes that host galaxies with very high stellar masses, including galaxies with $m_\star \sim 0.6 \times 0.17 M_{\text{infall}}$. This shows again that our formalism cannot constrain the mass fraction that is locked in stars, in contrast to what has been claimed using more simplified ABM models (e.g. Behroozi et al. 2010; Guo et al. 2010).

Model C is a good example of a model with a very high abundance of satellite galaxies. Although the number of satellite subhaloes is smaller than central subhaloes at a given $M_{\text{infall}}$, different $m_\star$–$M_{\text{infall}}$ relations for central and satellite subhaloes might result in a high abundance of satellite galaxies. We plot the satellite fraction (the number of satellite galaxies out of all the galaxies of a given $m_\star$) of all the models in Fig. 8. The satellite fraction changes significantly between the models. For model C this fraction reaches unity at low stellar masses, while for models B and D it reaches a maximum value of $\sim 0.3$ and $0.4$, respectively.

Model D is shown in Fig. 9. It is an example for a model in which $m_\star$ for satellites does not strongly depend on $M_{\text{infall}}$. In general, the slope of the $m_\star$–$M_{\text{infall}}$ relation, and the location where it turns over can be related to various processes in galaxy formation physics. Merger rates, cooling, feedback and star formation should all be combined in order to reproduce the observed $m_\star$–$M_{\text{infall}}$ relation (e.g. Shankar et al. 2006). Unless our more extreme models like model D are ruled out by other observational constraints, our results indicate a large amount of freedom in modelling the above processes.

The stellar fractions, $m_\star/M_{\text{infall}}$, for the few models presented here, are plotted in Fig. 10. The halo mass at which the global stellar fraction reaches a maximum can range from $\sim 3 \times 10^{11}$ to $\sim 3 \times 10^{12} M_{\odot}$. In model D this efficiency is approximately constant for satellite galaxies as a function of $M_{\text{infall}}$, showing no global strong peak.

The CFs of all the models are shown in Fig. 4. Models B, C and D fit the observed CFs to a good accuracy, below 0.1 dex rms, while model A is less accurate, reaching a level of 0.2 dex rms. This demonstrates the importance of using this constraint here, in order to narrow down the range of accepted models. As can be seen from Fig. 5, all the models fit the stellar mass function well. Another point that should be emphasized regarding Fig. 4 is the small-scale (< 1 Mpc) clustering of massive galaxies. The CF of different models shows variations at this regime, so it might be that extending the CF of the most massive domain into smaller radii will help to constrain the models. This will require a survey volume much larger than in the current SDSS survey.
Figure 4. The projected CFs derived for models A, B, C and D. Each panel corresponds to galaxies with stellar masses as indicated by the range of log M⊙. Lines show the results of the models. The observational data are using SDSS DR7 with the same technique as in Li et al. (2006), and are shown as error bars. Models B, C and D fit the observational data to a level of 0.1 dex rms, while model A deviates at the level of 0.2 dex (details on the way these errors are computed can be found in Section 4.1).

Figure 5. The stellar mass functions of galaxies using models A, B, C and D (symbols). The observed function derived by Li & White (2009) is plotted in dashed line. Symbols are placed at the centre of each domain.

4 RESULTS

4.1 Search strategy

Each model within our formalism is defined by 21 free parameters, where 18 of them define the domains in stellar masses (\(U_c^i\) and \(U_s^i\) from equation 13) used to compute the CF, one parameter fixes the time-scale for dynamical friction, and two parameters are responsible for the location of unresolved subhaloes. We were able to scan all possible options of the 18 parameters that govern the domains behaviour (within the resolution adopted here). In addition, we have tested five \(\alpha_{df}\) values in the range 0.1–10, and a few options for setting the location of unresolved subhaloes (see Appendix A and Table A1 for more details). In total, we have tested the stellar mass function for \(\sim 10^{12}\) models, approximately \(10^7\) out of them were tested against the observed CFs. More numerical details regarding the search algorithm can be found in Appendix B.

The values of \(\varphi\) and \(\psi\) are saved numerically in fine bins. The choice of bin size is important for several reasons. First, the function \(\psi_{ss}\) depends on five different variables, so the number of bins we use for each variable is limited because of computer memory issues. Secondly, fine bins will require more evaluations of CFs, as the code checks automatically all possible options. This can slow our code dramatically. On the other hand, the bin size limits the search resolution: too big bins will not allow us to find all possible solutions. As a compromise between these requirements we have chosen the following bin sizes: \(M_{\text{infall}}\) is split into bins of 0.02 (0.1) dex for subhaloes with mass smaller (bigger) than \(10^{12}\) M⊙, data per \(r\) and \(M_{200}\) are saved in bins of 0.1 dex. We use small bins for low-mass \(M_{\text{infall}}\) because the typical difference between adjacent \(U_i\) is smaller for small-mass \(M_{\text{infall}}\) (see e.g. model C above).

As described in Section 2.4, the domain boundaries are defined using power-law relations. Writing equation (13) in terms of log mass gives log \(M_{\text{infall}} = \log \beta_i + \delta_i \log M_{200}\). Our search algorithm checks all the possible values of \(\beta_i\) according to the bins of \(M_{\text{infall}}\). \(\delta_i\) is modelled in terms of the line slope, \(\delta_i = \tan \theta_i\). We sample \(\theta_i\) in steps of 6.75, over the range \([-90, 45]\).

Each model is accepted if it follows the conditions below.

(i) The model fits the stellar mass function of Li & White (2009) with an accuracy that is better than 20 per cent. This criterion is
A tale of two populations

Figure 6. Same as Figs 2 and 3 but for model B. Here $\alpha_{\text{df}} = 3$, and the location of unresolved subhaloes is set by the most bound particle.

Figure 7. Same as Figs 2 and 3 but for model C. Here $\alpha_{\text{df}} = 3, p = 0.5, q = 0.8$.

Figure 8. The fraction of satellite galaxies out of all galaxies, $f_{\text{sat}} = \phi_s / (\phi_s + \phi_c)$, for the different models presented here.

applied separately to each domain by integrating the stellar mass function over the domain range ($0.5$ dex in $m_*$). We have checked that the range of models shown below changes in a very minor way, when demanding a better fit that resembles the statistical errors from Li & White (2009) (these are $5, 5, 5, 10$ and $20$ per cent for each domain, in order of increasing mass). Here we choose to use a constant accuracy of $20$ per cent to account for systematic uncertainties in the stellar mass function (comparisons against other measurement can be found in Bernardi et al. 2010; Guo et al. 2010).

(ii) The model fits the logarithm of the observed CF (see Fig. 4) to better than an rms value of $0.1$ dex ($26$ per cent). This estimate is based on all points in the range $0.03 < r < 30 \, \text{Mpc} \, h^{-1}$, sampled in bins separated by $0.2$ dex in $\log(r)$. In order to test the effect of this fit accuracy we also show results using $0.2$ dex deviation. A detailed discussion of these issues is given below.

(iii) The model does not include individual points that deviate from the observed CF to more than a factor of $2$.

(iv) The stellar mass function is fitted also for masses larger than the most massive domain (i.e. for $m_* > 10^{11.77} \, M_\odot$). This domain should include only $23$ galaxies, when using the full Millennium simulation. Since this is a very small number, we allow our models to deviate an additional Poisson error from the nominal value. This means that the number of galaxies with masses larger than $10^{11.77} \, M_\odot$ can have any value in the range $[14, 34]$.
The stellar fraction (i.e. $m_\star/M_{\text{infall}}$) does not exceed the universal fraction of 0.17 for all subhaloes in the sample.

We adopt an accuracy of 0.1 and 0.2 dex for fitting the CFs due to the following reasons. First, the Millennium simulation being used here is based on a cosmological model with $\sigma_8 = 0.9$, higher than the most updated measurements of $\sigma_8 = 0.8$ (Jarosik et al. 2011). This should give rise to some deviation between our models and observations. We find that there is a minimum deviation of 0.08 dex between all the models discussed here, and the observed CFs. Since this deviation is dominated by scales larger than 1 Mpc, it is probably related to the different cosmological model assumed. Secondly, due to the finite bins that are used for saving $\psi$, the accuracy of our search algorithm is approximately 0.03 dex. As a result of the above, our minimum fit criterion is chosen to be 0.1 dex. However, since the error bars presented for the observed CFs are only statistical, it might be that systematic uncertainties would introduce errors that are larger than 0.1 dex. The CFs depend strongly on various galaxy properties like colour, star formation and morphology (Li et al. 2006). Consequently, if the stellar masses of galaxies are systematically biased for galaxies of a given property (for example the high star-forming galaxies), this could change the CFs considerably. We therefore additionally consider a fitting criterion of 0.2 dex to demonstrate the effect of the possible systematic uncertainties.

The minimum subhalo mass ($\sim 6.3 \times 10^{10} M_\odot$) enforced by the Millennium simulation seems to limit our models (see models A and C above). It might be that using simulations of higher resolution will permit more models. The Millennium II simulation (Boylan-Kolchin et al. 2009) is a good candidate for such a study. However, high-resolution simulations are naturally based on much smaller volume than what is being used here, resulting in a smaller statistical sample and non-negligible cosmic variance effects. Since our study is aiming at fitting the CFs to a high accuracy of 0.1 dex, we focus our study only on the Millennium simulation.

4.2 The mass relation

The main results of the parameter search are summarized in Figs 11 and 12. In case we force the models to reproduce the observed CFs to a high accuracy of 0.1 dex, the range of accepted models occupies a region of $\sim 1$ dex in the $m_\star-M_{\text{infall}}$ plane, as shown in Fig. 11. Interestingly, the uncertainty is small for central massive subhaloes. On the other hand, low-mass central subhaloes can host galaxies with very low stellar masses. The range of accepted models increases significantly in Fig. 12, where the accuracy of fitting the CFs is set to 0.2 dex. We have checked that the distributions of models within the plotted envelope in Figs 11 and 12 are roughly uniform, so the range of models is not affected by outliers.

The number of models presented in Figs 11 and 12 is quite large. Each domain includes $\sim 10^5$ successful models in Fig. 11, and $\sim 10^6$ models in Fig. 12. We only take into account domains for which nearby $U_i$ from adjacent domains coincide, and the set of all five domains covers the full mass range. Because of the above, we compute the median $M_{\text{infall}}$ for a given $m_\star$, which does not demand combinations of domains of different $m_\star$. Computing the opposite relation, median $m_\star$ per a given $M_{\text{infall}}$, is much more complicated within our formalism. However, since the relation for central galaxies does not include any scatter, it represents a median $m_\star$ per a given $M_{\text{infall}}$ as well.

In our models, the stellar mass of satellite galaxies depends on their host halo mass, which correlates with the number of...
neighbouring galaxies in observational studies. As a result, the range in the $m_\star - M_{\mathrm{infall}}$ relation changes as a function of the host halo mass. This is shown in Fig. 11 by plotting the envelope of all median relation in the $m_\star - M_{\mathrm{infall}}$ plane, but taking into account only galaxies that reside in clusters (their host halo mass is bigger than $10^{15}\,\text{M}_\odot$). For these satellite galaxies, the range in $M_{\mathrm{infall}}$ for a given $m_\star$ is larger than the one for all galaxies.

Fig. 12 includes the results of weak lensing analysis from Mandelbaum et al. (2006). These estimates are for the host halo mass of central galaxies, and should be compared to the solid lines of our models. The square symbols represent a subsample of early-type galaxies, while the circles correspond to late-type galaxies. As is mentioned in the figure caption, the fraction of late-type galaxies is very small at the most massive stellar mass bin (0.05).

Surprisingly, our results deviate from those of Mandelbaum et al. (2006) at the high-mass end, even when the range of models is large, using an uncertainty of 0.2 dex in matching the CF. This is similar to what was found by previous ABM studies, summarized by Behroozi et al. (2010). Apparently, for a given value of $m_\star$, the weak lensing results constrain the host halo mass by providing mainly an upper threshold (the low values of the host halo often reach the universal baryonic fraction). The opposite seems to be true for central galaxies in our analysis. Deviations between the two studies can be due to various effects.

(i) Mandelbaum et al. (2006) use a specific halo model for computing the lensing signal, which differs from the set of models being used here.

(ii) There is some uncertainty in their study owing to the width of the $M_{\mathrm{infall}}$ distribution for a given $m_\star$.

(iii) Their stellar mass estimates are based on Kauffmann et al. (2003) while the observations of Li & White (2009) use the method of Blanton & Roweis (2007). The difference between these estimates is discussed in both Li & White (2009) and Guo et al. (2010), and its effect on the stellar mass function is probably small. However, it might be that the difference between the methods is more significant for computing the CFs.

(iv) It might be that our range of models is too narrow, or that there are some systematic differences between the two approaches. This could be due to our search resolution, the assumed underlying cosmological models or the assumed initial mass function.

A better, more self-consistent way to compare our results against weak lensing would demand a direct estimate of the lensing signal from our models. This can then be compared to the observed shear signal. Such an analysis is however outside the scope of this work. A recent study by Leauthaud et al. (2011) have used a model based on HOD and demonstrated the strength of applying this additional constraint.

The results of Mandelbaum et al. (2006) indicate that the population of late-type massive central galaxies live inside haloes with lower masses than early-type galaxies. The difference between early- and late-type galaxies indicates that the scatter in the $m_\star - M_{\mathrm{infall}}$ relation might be significant. A large scatter was also found by More et al. (2011), based on satellite kinematics. We do not model such a scatter for central galaxies in this study. Previous works (e.g. Moster et al. 2010) have shown that including a scatter modifies the $m_\star - M_{\mathrm{infall}}$ relation at the massive end, such that for a given $M_{\mathrm{infall}}$, $m_\star$ decrease with increasing scatter. Interestingly, the results of Mandelbaum et al. (2006) reach the universal baryonic fraction, $m_\star/M_{\mathrm{infall}} \sim 0.17$, for part of the galaxy population. This is very similar to what is found here, and can be seen in model C (Fig. 7).

Mandelbaum et al. (2006) define the halo mass as the mass enclosed within a radius that corresponds to 180 times the mean density, this mass is higher than $M_{200}$ (see also Weinmann et al. 2006b), but agrees quite well (<0.1 dex) with $M_{\mathrm{infall}}$ for central subhaloes used here.

\[ \text{Figure 11.} \quad \text{A summary of all the models that fit the observed stellar mass function and the CFs of galaxies at } z = 0. \text{ Lines show the envelope of all relations between } m_\star \text{ and } M_{\mathrm{infall}} \text{ (we use median values of } M_{\mathrm{infall}} \text{ per a given } m_\star \text{ for each model). The results are separated into central and satellite subhaloes. Thin dotted–dashed line corresponds to the median } M_{\mathrm{infall}} \text{ for all satellite galaxies residing in haloes more massive than } 10^{15}\,\text{M}_\odot. \text{ The models reproduce the observed stellar mass function to better than 20 per cent, and the CFs to better than 0.1 dex rms. The thin solid line shows the } m_\star - M_{\mathrm{infall}} \text{ relation using the same behaviour for satellite and central galaxies, with no dependence on } M_{200}. \]

\[ \text{Figure 12.} \quad \text{Similar to Fig. 11 but allowing the models to deviate from the observed CFs by up to 0.2 dex rms. Results from weak lensing analysis by Mandelbaum et al. (2006) are plotted in symbols. Squares and circles refer to the mean value of early- and late-type galaxies, respectively, with error bars that reflect 95 per cent confidence level. The fraction of late-type galaxies out of the full sample is 0.74, 0.60, 0.46, 0.32, 0.20, 0.11 and 0.05 (ordered in increasing } m_\star. \]
3. This shows that the uncertainty in the location set by the most bound particle are plotted in the upper right-hand panel. The middle right-hand panel summarizes models that use the same mass relation for central and satellite galaxies, at a given mass. The lower left-hand panel plots models for which there is no dependence on $M_{200}$. Models using the default value of $\alpha_{df} = 3$ and location given by the most bound particle are plotted in the upper right-hand panel. The middle right-hand panel summarizes models that use the location as given by the most bound particle. The lower right-hand panel summarizes all models with the same dynamical friction constant, $\alpha_{df} = 3$.

In Fig. 13 we plot various subsamples of the models from Fig. 11, showing the influence of different parameters on the relation between $m_*$ and $M_{\text{infall}}$. Our results agree with previous studies in predicting a very tight relation between $m_*$ and $M_{\text{infall}}$. Each panel shows a selection of models out of the full sample, as plotted in Fig. 11. The upper left-hand panel takes into account all models for which the behaviour for satellite and central subhaloes is the same ($\pm 0.2$ dex), and no dependence on $M_{200}$ is allowed. In the middle left-hand panel we plot models with a similar median $M_{\text{infall}}$ for central and satellite galaxies, at a given mass. The lower left-hand panel plots models for which there is no dependence on $M_{200}$. Models using the default value of $\alpha_{df} = 3$ and location set by the most bound particle are plotted in the upper right-hand panel. The middle right-hand panel summarizes models that use the location as given by the most bound particle. The lower right-hand panel summarizes all models with the same dynamical friction constant, $\alpha_{df} = 3$.

It is worth noting that our models do not include solutions that have exactly the same mass relation for central and satellite galaxies, and can still fit the CFs to a level of $0.1$ dex. The minimum deviation between central and satellite galaxies in the upper left-hand panel of Fig. 13 is roughly $0.12$ dex in $M_{\text{infall}}$. This small difference might be an artefact of the underlying cosmological parameters assumed by the Millennium simulation.

The models shown here use the two populations of galaxies to compensate for each other, keeping the overall $m_* - M_{\text{infall}}$ relation relatively similar. This effect is shown in the middle left-hand panel of Fig. 13. Once we force the models to have a similar median $M_{\text{infall}}$ for satellite and central galaxies, the range of models decreases, especially at the high-mass end. It might be that constraints on the differences between $m_*$ of the two populations, for a given $M_{\text{infall}}$, could be enforced ad hoc. However, it is not clear what this limit should be. A reasonable demand would be that the full distribution of the two populations would overlap to some level, allowing a transition of central galaxies into satellites. This condition seems to be fulfilled by our models (see Figs 6, 7 and 9). Since we do not model a distribution for the $m_*$ of central galaxies, we do not explore this issue further here.

The effect of $\alpha_{df}$ and satellite locations is explored in Fig. 13. It seems that both elements add freedom to our models, although the effect of satellite location is slightly stronger. Higher resolution cosmological simulations might help to constrain $\alpha_{df}, p, q$. However, these simulations are based on dark matter only, while the results here are affected strongly by the baryonic components of galaxies (see e.g. Boylan-Kolchin et al. 2008; Jiang et al. 2008).

Lastly, in Fig. 14, we show the results of a parameter search when the reference CFs are not the observed ones, but instead are taken from one model out of the models tested here. The range of accepted models in this case is related to the internal degrees of freedom of the models, and is less related to the choice of the cosmological parameters, the method adopted for computing $m_*$, or to the specific physical assumptions used in our formalism. The fact that Fig. 14 is similar to Fig. 11 indicates that the underlying cosmological model does not affect the range of acceptable models significantly. This is because here the reference CFs are based on the same cosmology, but use a specific model for the relation between $m_*$ and subhaloes. As was pointed out by Cacciato et al. (2009), the effect of the underlying cosmological model should be important when constraints from abundance and clustering of galaxies are combined with galaxy–galaxy lensing measurements. This issue should be further examined using an appropriate $N$-body cosmological simulation.

5 SUMMARY AND DISCUSSION

In this work we have studied the relation between the stellar mass of galaxies ($m_*$) and the mass of their host subhaloes ($M_{\text{infall}}$). Our models are constrained by the abundance of galaxies, and their CF at
z = 0. We have shown that once the population of galaxies is broken into two subpopulations of central and satellite galaxies, the allowed range in the $m_\star$–$M_{\text{infall}}$ relation for each population reaches a factor of $\sim 10$. The range of accepted models depends on the accuracy by which the models reproduce the observed abundance and clustering of galaxies. As was demonstrated in our previous work (Paper I), a different $m_\star$–$M_{\text{infall}}$ relation for central and satellite galaxies is expected in galaxy formation models, although the strength of the deviation cannot be constrained easily.

The shape of the $m_\star$–$M_{\text{infall}}$ relation has a large degree of freedom, resulting in a stellar fraction (the ratio between $m_\star$ and $M_{\text{infall}}$) that can peak anywhere between $3 \times 10^1$ and $3 \times 10^2 M_\odot$. It can even be close to a constant as a function of $M_{\text{infall}}$ for satellite galaxies. Interestingly, the stellar fraction can reach high values, comparable to the maximum universal value of $0.17$, for low-mass satellite galaxies residing in low-mass haloes. Although it seems unlikely that the stellar fraction would exceed $0.17$, our models are not able to constrain this ratio, and instead we are forced to use it as an ad hoc constraint.

Our models are similar to the ABM approach, and are directly relating a galaxy with a specific $m_\star$ to each subhalo within a large $N$-body simulation. However, in comparison to previous studies, we include various additional ingredients for modelling satellite galaxies. Their stellar mass depends on both the host subhalo mass ($M_{\text{infall}}$) and on the halo mass at $z = 0$, and might differ from the stellar mass of central galaxies of the same $M_{\text{infall}}$. We include subhaloes that have merged into more massive subhaloes at high-$z$ but might host galaxies that will survive until $z = 0$, due to long dynamical friction time-scales. The location of these ‘unresolved’ subhaloes is set by either the location of their most bound particle, or by an analytical model. We found that the most important ingredient for fixing the range of accepted models is the way stellar masses for satellite galaxies are modelled.

A model with two populations of galaxies includes a large amount of freedom in matching the observed stellar mass function. We therefore constrain our models also against the observed CF of galaxies. We developed a new formalism to compute the model CF based on the pair statistics of subhaloes. This enabled us to scan systematically a significant part of the parameter space. We have tested $\sim 10^3$ different ABM models, and estimated the CF for $\sim 10^3$ models that showed a good match to the stellar mass function. The accuracy by which we fit the CF affects significantly the range of accepted models within the $m_\star$–$M_{\text{infall}}$ plane. Consequently, a detailed study of the systematic uncertainties involved in this measurement is crucial in order to better constrain our models.

The range allowed in the $m_\star$–$M_{\text{infall}}$ relation might be larger than our prediction due to various effects. There are still a large number of models that we did not test here. These include complicated dependence of $m_\star$ on the halo mass, different estimates for dynamical friction and satellite locations, higher resolution of the underlying $N$-body simulation, different observed stellar mass functions (e.g. Bernardi et al. 2010) and modelling a random scatter in $m_\star$. On the other hand, the range of models might be modified once we use subhaloes from an $N$-body simulation using the most up-to-date cosmological parameters. Previous estimates for the uncertainty in the $m_\star$–$M_{\text{infall}}$ relation were emphasizing the contribution from uncertainty in the stellar mass function, resulting in $0.25$ dex (Behroozi et al. 2010). This uncertainty should be added to what we find here.

In this paper we have reproduced the abundance and clustering of galaxies only at $z = 0$ when constraining the $m_\star$–$M_{\text{infall}}$ relation. It is not likely that data from higher redshifts would limit the $m_\star$–$M_{\text{infall}}$ relation to a narrow range. This is because the same uncertainties discussed here would be valid at high $z$, in addition to the larger observational errors inherent at these redshifts. Moreover, once the $m_\star$–$M_{\text{infall}}$ relation is fixed at some redshift range, it is not straightforward to decide which models violate the physics of galaxy formation by linking galaxies at different epochs. For example, as we showed in Paper I, the $m_\star$–$M_{\text{infall}}$ relation for satellite galaxies at the time of infall might already be different from the relation for central galaxies at the same epoch.

Our empirical results have implications for various aspects of galaxy formation. Guo et al. (2010) and Sawala et al. (2011) argued that stellar fractions at a fixed subhalo mass derived from basic ABM models are systematically lower than the results of detailed hydro simulations. Our results that allow for a much larger spread of stellar mass at a fixed subhalo mass, are less definitive in this respect. This is especially true for dwarf galaxies, for which there is no available clustering data. An additional implication is related to the cumulative energy injected into the interstellar medium in a supernova feedback-constrained scenario. Shankar et al. (2006; see also Dekel & Woo 2003) redefined the expected trend of stellar and halo mass of the type $m_\star \propto f_{\text{snr}} M_{\text{infall}}^{\alpha}$. Here $1 < \alpha < 2$ and $f_{\text{snr}}$ is the fraction of surviving stellar mass. Based on the observed rather low stellar mass in haloes with $M_{\text{infall}} \lesssim 10^4 M_\odot$ inferred from a basic ABM model, they argued that supernova feedback appeared to be insufficient to remove the gas associated with the host halo. The results of this paper, however, show that some ABM models allow for much larger stellar mass fraction in low-mass haloes, thus providing some hints towards the solution of this puzzling issue.

Another possible non-trivial consequence of our results concerns HOD models. As outlined in e.g. Berlind & Weinberg (2002), this class of models is based on a parametrized conditional probability $P(N|M)$ that a dark matter halo of virial mass $M$ contains $N$ galaxies. Adding prescriptions on the spatial distribution of subhaloes derived from accurate $N$-body simulations, HOD models have been rather successful in reproducing the two-point correlation function of different classes of galaxies and at different redshifts. However, once we allow for a different ranking between satellite and central galaxies, as was found in this paper, the shape of $P(N|M)$ might become rather different from its usual functional form. This could induce non-trivial degeneracies in these models, allowing for different occupation distributions that still match the data.

The methodology developed in this work can be used to study the clustering properties of active galactic nuclei, cold gas and subsets of galaxies divided by e.g. luminosity or colour. As was demonstrated here, our method already enabled us to better fit the observed stellar mass functions and CFs of all galaxies, with the current data sets. It might become more important when studying clustering properties of objects which tend to depend more strongly on environment, like the cold gas mass within galaxies. Our method may be able to shed some light on the limitations in modelling the observed high-$z$ galaxies using standard halo models (Quadri et al. 2008; Tinker, Wechsler & Zheng 2010; Wake et al. 2011). In addition, it might be useful for future surveys with high-quality data and a large survey volume, for which the simple ABM approach might not be flexible enough to provide an accurate fit to the data.

Improved constraints on the $m_\star$–$M_{\text{infall}}$ relation might be obtained from observational quantities, other than the stellar mass function and clustering used here. These might include conditional stellar mass functions per halo mass (e.g. Yang, Mo & van den Bosch 2009b), satellite kinematics (More et al. 2011), weak lensing (Mandelbaum et al. 2006; Cacciato et al. 2009), star formation histories and dynamical tracers such as the velocity dispersion and circular velocity (e.g. Dutton et al. 2010). In addition, physically motivated
models like hydrodynamical simulations or SAMs are crucial to obtain insights on the difference between the evolution of satellite and central galaxies. More effort to understand the fundamental relation between the mass of subhaloes and galaxies is clearly needed both on the observational and theoretical side.

ACKNOWLEDGMENTS

We thank the referee for a detailed and constructive report that helped to improve the presentation of this work. We also thank Peter Behroozi, Charlie Conroy, Umberto Maiol, Rachel Mandelbaum, Risa Wechsler and Simon White for useful discussions. 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. FS acknowledges support from the Alexander von Humboldt Foundation. MB-K acknowledges support from the Southern California Center for Galaxy Evolution, a multicampus research programme funded by the University of California office of research.

REFERENCES

Abazajian K. N. et al., 2009, ApJS, 182, 543
Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379
Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
Bernardi M., Shankar F., Hyde J. B., Mei S., Marulli F., Sheth R. K., 2010, MNRAS, 404, 2087
Binney J., Tremaine S., 1987, Galactic Dynamics. Princeton Univ. Press, Princeton, NJ
Blanton M. R., Roweis S., 2007, AJ, 133, 734
Boylan-Kolchin M., Mc C.-P., Quataert E., 2008, MNRAS, 383, 93
Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, MNRAS, 398, 1150
Cacciato M., van den Bosch F. C., More S., Li R., Mo H. J., Yang X., 2009, MNRAS, 394, 929
Colpi M., Mayer L., Governato F., 1999, ApJ, 525, 720
Conroy C., Wechsler R. H., 2009, ApJ, 696, 620
Conroy C., Wechsler R. H., Kravtsov A. V., 2006, ApJ, 647, 201
Conroy C., Wechsler R. H., Kravtsov A. V., 2007, ApJ, 668, 826
Cooray A., Sheth R., 2002, Phys. Rep., 372, 1
Croton D. J. et al., 2006, MNRAS, 365, 11
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
Dekel A., Mendelbaum R., Seljak U., Kauffmann G., Hirata C. M., Brinkmann J., 2006, MNRAS, 368, 715
Mandelbaum R., Seljak U., Kauffmann G., Hirata C. M., Brinkmann J., 2006, MNRAS, 368, 715
Mo H., van den Bosch F. C., White S., 2010, Galaxy Formation and Evolution. Cambridge Univ. Press, Cambridge
Monaco P., Murante G., Borgani S., Fontanot F., 2006, ApJ, 652, 189
More S., van den Bosch F. C., Cacciato M., Skibba R., Mo H. J., Yang X., 2011, MNRAS, 410, 210
Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, ApJ, 710, 903
Neistein E., Weinhmann S. M., 2010, MNRAS, 405, 2717
Neistein E., Weinhmann S. M., Li C., Boylan-Kolchin M., 2011, MNRAS, 414, 1405
Pasquali A., van den Bosch F. C., Mo H. J., Yang X., Somerville R., 2009, MNRAS, 394, 38
Purcell C. W., Bullock J. S., Zentner A. R., 2007, ApJ, 666, 20
Quadri R. F., Williams R. J. J., Lee K., Franx M., van Dokkum P., Brammer G. B., 2008, ApJ, 685, L1
Sawala T., Guo Q., Scannapieco C., Jenkins A., White S., 2011, MNRAS, 413, 659
Shankar F., Lapi A., Salucci P., De Zotti G., Danese L., 2006, ApJ, 643, 14
Skibba R. A., Sheth R. K., 2009, MNRAS, 392, 1080
Skibba R. A., van den Bosch F. C., Yang X., More S., Mo H., Fontanot F., 2011, MNRAS, 410, 417
Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726
Springel V. et al., 2005, Nat, 435, 629
Tinker J. L., Weinberg D. H., Zheng Z., Zehavi I., 2005, ApJ, 631, 41
Tinker J. L., Wechsler R. H., Zheng Z., 2010, ApJ, 709, 67
Vale A., Ostriker J. P., 2004, MNRAS, 353, 189
van den Bosch F. C., Tormen G., Giocoli C., 2005, MNRAS, 359, 1029
von der Linden A., Best P. N., Kauffmann G., White S. D. M., 2007, MNRAS, 379, 867
Wade D. A. et al., 2011, ApJ, 728, 46
Wang L., Li C., Kauffmann G., De Lucia G., 2006, MNRAS, 371, 537
Wang L., Li C., Kauffmann G., De Lucia G., 2007, MNRAS, 377, 1419
Weinhmann S. M., van den Bosch F. C., Yang X., Mo H. J., 2006a, MNRAS, 366, 2
Weinhmann S. M., van den Bosch F. C., Yang X., Mo H. J., Croton D. J., Moore B., 2006b, MNRAS, 372, 1161
Weinhmann S. M., Kauffmann G., von der Linden A., De Lucia G., 2010, MNRAS, 406, 2249
Yang X., Mo H. J., van den Bosch F. C., Pasquali A., Li C., Barden M., 2007, ApJ, 671, 153
Yang X., Mo H. J., van den Bosch F. C., 2009a, ApJ, 693, 830
Yang X., Mo H. J., van den Bosch F. C., 2009b, ApJ, 695, 900
York D. G. et al., 2000, AJ, 120, 1579
Zehavi I. et al., 2005, ApJ, 630, 1
Zentner A. R., Berlind A. A., Bullock J. S., Kravtsov A. V., Wechsler R. H., 2005, ApJ, 624, 505

APPENDIX A: THE LOCATION OF SATELLITE GALAXIES

Here we describe the model we use for fixing the location of unresolved subhaloes. These subhaloes are not identified at $z = 0$, but can still host galaxies according to the dynamical friction estimate.

Consider a massive point particle subject to dynamical friction in an infinite mass distribution with density profile $\rho \propto r^{-\gamma}$. In the limit of a circular orbit for the point particle and an isothermal profile ($\gamma = 2$) for the background density, the decay of the satellite’s orbit due to dynamical friction can be modelled analytically (equations 7–25 of Binney & Tremaine 1987). Assuming that $\gamma$ does not differ much from 2, we can follow the same arguments as Binney & Tremaine and find that the equation governing the decay rate for the specific angular momentum $L$ is

$$\frac{dL}{dt} = \frac{F_r}{M_{sat}} \approx -c \ln M_{sat} \frac{\rho(r)}{\epsilon^2(r)}.$$  \hspace{0.5cm} (A1)
APPENDIX B: SCANNING THE PARAMETER SPACE

As explained in Section 4, each model within our formalism is defined using 18 free parameters that fix the five domains in stellar mass used here. There are additional three parameters that define the dynamical friction time-scale ($\alpha_{df}$), and the location of unresolved subhaloes ($p, q$). Even though this seems to be a huge parameter space, we managed to scan it systematically and with high resolution. In this section we explain the numerical details that allow this parameter search.

(i) Searching the parameter space is done separately for the different domains of stellar mass. This reduces the number of independent parameters to those that define a domain (i.e. six). Only in a later stage we search for solutions that combine a set of five domains that have mutual boundaries.

(ii) We search for solutions one domain after another, starting from the most massive one. At each domain we require that the model boundaries will coincide with an accepted model from the previous domain.

(iii) We first check that the stellar mass function meets the fitting criterion. Only after we have a solution for all the domains we check the fit to the CF.

(iv) Because of the number of bins we adopt for $M_{sat}$ and $\delta_i$, each domain has $10^{10}$ optional combinations of $\alpha_i, \beta_i, \delta_i$. We first compute separately the contributions to the mass function (equation 14) from satellite and central subhaloes ($\phi_s, \phi_c$). Only in a later stage we combine all the $10^{10}$ options to see which of them fits the data.

(v) Before computing the CF we sort the different models such that there will be a minimum difference between neighbouring models in the list. We then compute the CF using the ‘moving average’ scheme. This means that we do not compute the integral of equation (14) for each model, but only compute the incremental changes of this integral when going over the full list of models.

(vi) We compute CF values starting the smallest radius. After the CF is computed for each radius, we check whether the model fits the data. This allows us to quickly reject bad solutions.

Our algorithm is able to compute CF for each domain within a $10^{-2}$ s, using one processor. In total, we have computed the CF for more than $\sim 10^7$ models. The CFs for all models are based on the full sample of subhaloes from the Millennium simulation.

This paper has been typeset from a TeX/\LaTeX file prepared by the author.