

The impact of baryonic processes on the two-point correlation functions of galaxies, subhaloes and matter

Marcel P. van Daalen,^{1,2★} Joop Schaye,¹ Ian G. McCarthy,³ C. M. Booth⁴
and Claudio Dalla Vecchia⁵

¹*Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands*

²*Max Planck Institute for Astrophysics, Karl-Schwarzschild Straße 1, D-85741 Garching, Germany*

³*Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK*

⁴*Department of Astronomy & Astrophysics, The University of Chicago, Chicago, IL 60637, USA*

⁵*Max Planck Institute for Extraterrestrial Physics, Gießenbachstraße 1, D-85748 Garching, Germany*

Accepted 2014 March 10. Received 2014 March 4; in original form 2013 October 28

ABSTRACT

The observed clustering of galaxies and the cross-correlation of galaxies and mass provide important constraints on both cosmology and models of galaxy formation. Even though the dissipation and feedback processes associated with galaxy formation are thought to affect the distribution of matter, essentially all models used to predict clustering data are based on collisionless simulations. Here, we use large hydrodynamical simulations to investigate how galaxy formation affects the autocorrelation functions of galaxies and subhaloes, as well as their cross-correlation with matter. We show that the changes due to the inclusion of baryons are not limited to small scales and are even present in samples selected by subhalo mass. Samples selected by subhalo mass cluster ~ 10 per cent more strongly in a baryonic run on scales $r \gtrsim 1 h^{-1}$ Mpc, and this difference increases for smaller separations. While the inclusion of baryons boosts the clustering at fixed subhalo mass on all scales, the sign of the effect on the cross-correlation of subhaloes with matter can vary with radius. We show that the large-scale effects are due to the change in subhalo mass caused by the strong feedback associated with galaxy formation and may therefore not affect samples selected by number density. However, on scales $r \lesssim r_{\text{vir}}$ significant differences remain after accounting for the change in subhalo mass. We conclude that predictions for galaxy–galaxy and galaxy–mass clustering from models based on collisionless simulations will have errors greater than 10 per cent on sub-Mpc scales, unless the simulation results are modified to correctly account for the effects of baryons on the distributions of mass and satellites.

Key words: galaxies: formation – cosmology: theory – large-scale structure of Universe.

1 INTRODUCTION

Many cosmological probes are used in order to derive the values of the parameters describing our Universe, often relying on some aspect of large-scale structure. By combining different probes, degeneracies can be broken and the constraints on the numbers that characterize our Universe can be improved. However, observations alone are not enough: strong theoretical backing is needed to interpret the data and to avoid, or at least to reduce, unexpected biases.

Modelling our Universe as a dark matter only Λ cold dark matter (Λ CDM) universe was a reasonable approximation for the interpretation of past data sets. However, over the last few years it has become clear that for many probes this is no longer the case in

the era of precision cosmology: ignoring processes associated with baryons and galaxy formation may lead to serious biases when interpreting data. The existence of baryons and the many physical processes associated with them have been shown to significantly impact, for example, the mass profiles (e.g. Gnedin et al. 2004; Duffy et al. 2010; Abadi et al. 2010; Governato et al. 2012; Martizzi et al. 2012; Velliscig et al. 2014) and shapes of haloes (e.g. Kazantzidis et al. 2004; Tissera et al. 2010; Bryan et al. 2013), the clustering of matter (e.g. White 2004; Zhan & Knox 2004; Jing et al. 2006; Rudd, Zentner & Kravtsov 2008; Guillet, Teyssier & Colombi 2010; Casarini et al. 2011; van Daalen et al. 2011) and, subsequently, weak lensing measurements (e.g. Semboloni et al. 2011; Semboloni, Hoekstra & Schaye 2013; Yang et al. 2013; Zentner et al. 2013), the strong lensing properties of clusters (e.g. Mead et al. 2010; Killedar et al. 2012), and the halo mass function (e.g. Stanek, Rudd & Evrard 2009; Cui et al. 2012; Sawala

*E-mail: daalen@strw.leidenuniv.nl

et al. 2013; Balaguera-Antolínez & Porciani 2013; Martizzi et al. 2013; Velliscig et al. 2014). To complicate matters further, different authors studying the same aspects of galaxy formation often find different and sometimes even contradictory results, depending not only on which physical processes are modelled but also on the choice of numerical code, and particularly on the implementation of subgrid recipes for feedback from star formation and active galactic nuclei (hereafter AGN; e.g. Scannapieco et al. 2012). Until a consensus can be reached, it is therefore important to determine the range of values that observables can take depending on whether certain baryonic processes are included in a model, and the way in which they are implemented.

In this paper, we aim to quantify the effects of baryons and galaxy formation on the two-point real-space correlation function. Specifically, we will investigate how the redshift zero galaxy and subhalo correlation functions and the galaxy–matter cross-correlation, which is observable through galaxy–galaxy lensing, are changed if baryonic processes are allowed to influence the distribution of matter to varying degrees, i.e. using different feedback models. To this end, we will use the reference and AGN models from the Overwhelmingly Large Simulations project (OWLS; Schaye et al. 2010). These were also employed by van Daalen et al. (2011) and we have since repeated them using larger volumes, more particles and a more up-to-date cosmology. The AGN model is particularly relevant, as it has been shown to reproduce many relevant X-ray and optical observations of groups and clusters (McCarthy et al. 2010, 2011; Stott et al. 2012).

Any changes in the clustering of objects brought about by galaxy formation can enter into the correlation function in two ways. The first and most well-established effect is due to a change in the mass of the objects. For example, assuming that higher-mass haloes are more strongly clustered, if supernova feedback systematically lowers the stellar content of haloes, then a model which includes this process is expected to show increased clustering at fixed stellar mass relative to one that does not.¹ Likewise, the clustering of haloes at fixed halo mass is also expected to show increased clustering when efficient feedback is included, due to the total mass of the halo being lowered. Secondly, the positions of galaxies and haloes may shift due to changes in the physics: if the mass within a certain radius around an object changes, then the gravitational force acting on those scales will change as well, affecting the dynamics of nearby galaxies and haloes. Moreover, tidal stripping, and hence also dynamical friction, will affect satellites differently if baryonic processes change the density profiles of either the satellites or the host haloes. We will consider both types of effects here; most importantly, we will disentangle the two and show what effects remain after we account for the change in halo mass, as could be done approximately by selecting samples with constant number density. As we will see, not all shifts in position average out, nor can the modification of the halo profiles be ignored.

Quantifying the significance of the various ways in which clustering measurements may deviate from those in a dark matter only universe is vital for the improvement of current models employed

in clustering studies. Typically, these are based on the distribution of dark matter alone, be they semi-analytical models (see Baugh 2006, for a review), a combination of halo occupation distribution (HOD) and halo models (e.g. Jing, Mo & Börner 1998; Berlind & Weinberg 2002; Cooray & Sheth 2002; Yang, Mo & van den Bosch 2003; Kravtsov et al. 2004; Tinker et al. 2005; Wechsler et al. 2006; van den Bosch et al. 2013) or subhalo abundance matching (SHAM) models (e.g. Vale & Ostriker 2004; Shankar et al. 2006; Conroy, Wechsler & Kravtsov 2006; Moster et al. 2010; Guo et al. 2010; Behroozi, Conroy & Wechsler 2010; Simha & Cole 2013). It is therefore important to investigate which ingredients may currently be missing from such efforts.

The effects of galaxy formation on subhalo–subhalo clustering were previously considered by Weinberg et al. (2008) and Simha et al. (2012). Weinberg et al. (2008) compared the clustering of objects at fixed number density in a dark matter only simulation with a baryonic simulation including weak supernova feedback but no feedback from AGN, and with identical initial conditions. They found that subhaloes cluster more strongly on small scales in the baryonic simulation due to the increased survival rate of baryonic satellites during infall. While we find a similar increase in the autocorrelation of subhaloes on small scales ($r \lesssim r_{\text{vir}}$) – with a corresponding decrease in clustering on slightly larger scales – we point out that such results may be biased, due to the difficulties of detecting infalling dark matter satellites (e.g. Muldrew, Pearce & Power 2011, see our Appendix B).

Simha et al. (2012) extended the work of Weinberg et al. (2008) in several ways, among which are the addition of more effective stellar feedback and the use of the mass of the subhalo at infall, rather than the current mass, when assigning galaxy properties to the subhaloes. They find that the addition of effective feedback causes the discrepancies between clustering in hydrodynamical simulations and results from SHAM to increase. They demonstrate that the two-point correlation function of baryonic subhaloes can be recovered to better than 15 per cent on scales $r > 2 h^{-1}$ Mpc when winds are included, but that the discrepancy at smaller scales in these simulations can be up to a factor of a few. The galaxy correlation function is reproduced much better if the stellar mass threshold is raised; however, as these simulations do not contain any form of feedback that is effective at high stellar masses, we would expect the further addition of a process like AGN feedback to exacerbate the discrepancy between SHAM results and hydrodynamical simulations for massive galaxies.

This paper is organized as follows. We will briefly introduce our simulations and explain how we calculate the relevant quantities in Section 2. Here we will also discuss how we identify the same halo in different simulations, an essential step in order to separate the change in halo mass from other effects. We present our results in Section 3 and summarize our findings in Section 4. Finally, we show the convergence with resolution and box size in Appendix A and consider the fraction of subhaloes successfully linked between simulation in Appendix B.

2 METHOD

2.1 Simulations

We consider three models from the OWLS project (Schaye et al. 2010): *DMONLY*, *REF* and *AGN*. All of these simulations were run with a modified version of `GADGET III`, the smoothed-particle hydrodynamics (SPH) code last described in Springel (2005). We will discuss the models employed briefly below.

¹ Situations in which feedback would have the reverse effect are possible in principle. For example, if the stellar mass–halo mass relation were flat where AGN feedback is important and had a large scatter, then the stellar mass of some galaxies inhabiting such haloes could be lower than that of galaxies in lower-mass haloes. As a result, the most massive galaxies would reside in intermediate-mass haloes. However, such a scenario is not supported by our simulations.

In order to study relatively low-mass objects while also simulating a volume that is sufficiently large to obtain a statistical sample of high-mass objects, we combine the results of simulations with different box sizes. For each model, we ran simulations in periodic boxes of comoving side lengths $L = 200$ and $400 h^{-1}$ Mpc, both with $N^3 = 1024^3$ CDM particles and – with the exception of *DMONLY* – an equal number of baryonic particles. The gravitational forces are softened on a comoving scale of $1/25$ of the initial mean inter-particle spacing, L/N , but the softening length is limited to a maximum physical scale of $1 h^{-1}$ kpc [$L/(100 h^{-1}$ Mpc)]. The particle masses in the baryonic *L200* (*L400*) simulations are $4.68 \times 10^8 h^{-1} M_\odot$ ($3.75 \times 10^9 h^{-1} M_\odot$) for dark matter and $9.41 \times 10^7 h^{-1} M_\odot$ ($7.53 \times 10^8 h^{-1} M_\odot$) for the baryons. We will use the higher-resolution *L200* simulations to study the clustering of galaxies with stellar mass $M_* < 10^{11} h^{-1} M_\odot$ and subhaloes with total mass $M_{\text{sh}} < 10^{13} h^{-1} M_\odot$, while taking advantage of the larger volume of the *L400* simulations to study higher masses. When considering cross-correlations with the matter distribution, resolution is more important than volume, and we use the *L200* simulations at all masses. We discuss our choice of mass limits in Appendix A, where we also show resolution tests. All the simulations we employ in this paper were run with a set of cosmological parameters derived from the *Wilkinson Microwave Anisotropy Probe* (WMAP) 7-year results (Komatsu et al. 2011), given by $\{\Omega_m, \Omega_b, \Omega_\Lambda, \sigma_8, n_s, h\} = \{0.272, 0.0455, 0.728, 0.81, 0.967, 0.704\}$. It is important to note that all simulations with identical box sizes were run with identical initial conditions, which allows us to compare the effects of baryons and galaxy formation for the exact same objects.

The *DMONLY* simulation, as its name suggests, contains only dark matter. This provides us with a useful baseline model for testing the impact of baryon physics.

The *REF* simulation is the reference OWLS model. It includes sub-grid recipes for star formation (Schaye & Dalla Vecchia 2008), radiative (metal-line) cooling and heating (Wiersma, Schaye & Smith 2009a), stellar evolution, mass loss from massive stars and chemical enrichment (Wiersma et al. 2009b) and a kinetic prescription for supernova feedback (Dalla Vecchia & Schaye 2008). The reference simulation is not intended to be the most realistic, but instead includes only those physical processes most typically found in simulations of galaxy formation.

The third and final simulation we consider here, *AGN*, adds feedback from accreting supermassive black holes to the reference simulation. AGN feedback was modelled following the prescription of Booth & Schaye (2009), which built on the model of Springel, Di Matteo & Hernquist (2005). We believe *AGN* to be our most realistic model, as it is the only model that solves the well-known overcooling problem (e.g. Balogh et al. 2001) and that reproduces the observed properties of groups (McCarthy et al. 2010, 2011; Stott et al. 2012). Specifically, this model has been shown to reproduce the gas density, temperature, entropy, and metallicity profiles inferred from X-ray observations, as well as the stellar masses, star formation rates, and stellar age distributions inferred from optical observations of low-redshift groups of galaxies. van Daalen et al. (2011) used this model to show that AGN feedback has a dramatic effect on the clustering of matter; here we wish to investigate whether the effect on the clustering of galaxies and subhaloes is equally important.

2.2 Calculating correlation functions

The correlation function, $\xi(r)$, returns the excess probability, relative to a random distribution, of finding two objects at a given

separation r . It is therefore a measure of the clustering of these objects as a function of scale. As our simulations contain only a moderate number of resolved objects (i.e. galaxies and (sub)haloes), we do not need to resort to approximations that are common in the calculation of two-point clustering statistics. Instead, we can use a parallelized brute force approach in which we obtain the (cross-)correlation function through simple pair counts, using the relation:

$$\xi_{XY}(r) = \frac{DD_{XY}(r)}{RR_{XY}(r)} - 1. \quad (1)$$

Here X and Y denote two (not necessarily distinct) sets of objects (e.g. galaxies and particles or galaxies and galaxies), $DD_{XY}(r)$ is the number of unique pairs consisting of an object from set X and an object from set Y separated by a distance r , and $RR_{XY}(r)$ is the expected number of pairs at this separation if the positions of the objects in these sets were random. As our simulations are carried out with periodic boundary conditions, more complicated expressions involving cross terms of the form $DR_{XY}(r)$ (e.g. Landy & Szalay 1993) are not necessary, nor do we need to actually create random fields; instead, we can simply compute the term in the denominator analytically.

The basic functions that we will consider in this paper are the galaxy autocorrelation function, ξ_{gg} , the galaxy–mass cross-correlation function, ξ_{gm} , the subhalo autocorrelation function, ξ_{ss} , and the subhalo–mass cross-correlation function, ξ_{sm} . We divide galaxies and subhaloes into different bins according to their stellar and subhalo dark matter mass, respectively. When cross-correlating with matter, we weight particles by their mass. To keep the computation time manageable, we use only 25 per cent of all particles for the lowest mass bin of the simulations with $(2 \times) 1024^3$ particles, randomly selected. In all other mass bins, we cross-correlate with the full particle distribution. We have verified that this does not influence our results in any way. Throughout this paper we will focus on the three-dimensional correlation function. We will only show the correlation functions in radial bins where the number of pairs exceeds 10, to prevent our results from being dominated by spurious clumping. We take the position of our objects to be the position of their most-bound particle, and assign each galaxy a mass equal to the total mass in stars in its subhalo. Finally, we confine our analysis to scales $r \lesssim 20 h^{-1}$ Mpc, corresponding to at most 1/10th of box size, in order to avoid the effects of missing large-scale modes.

2.3 Linking haloes between different simulations

As discussed previously, there are two main ways in which the two-point correlation function may be affected by baryonic processes: through changes in the masses of objects, and through shifts in their positions. To disentangle the two effects, we make use of the fact that all OWLS models were run from identical initial conditions, allowing us to identify the same objects in different simulations. In this way we can assign each object in simulation B the mass that the same object possesses in simulation A, thereby isolating the effect of changes in the positions of objects on the clustering signal.

Haloes are identified in our simulations using the Friends-of-Friends algorithm (run on the dark matter particles, with linking length 0.2) combined with a spherical overdensity finder, as implemented in the `SUBFIND` algorithm (Springel et al. 2001; Dolag et al. 2009). For every (sub)halo in simulation A we flag the N_{mb} most-bound dark matter particles, meaning the particles with the highest absolute binding energy. Next, we locate these particles in the other simulations, using the unique number associated with every particle. If we find a (sub)halo in simulation B that contains at least

50 per cent of these flagged particles, a first link is made. The link is confirmed if, by repeating the process starting from simulation B, the previous (sub)halo in simulation A is found.

Here we use $N_{\text{mb}} = 50$, but we have verified that our results are insensitive to this choice (see Velliscig et al. 2014). For haloes with less than N_{mb} dark matter particles, all dark matter particles are used. The fraction of haloes linked quickly increases as a function of mass, reaching essentially unity for sufficiently well-resolved haloes. For all subhaloes employed in this work, the linked fraction of *DMONLY* subhaloes typically exceeds 99 per cent, the exception being the lowest mass bin where the linked fraction is around 98 per cent. However, at small separations the linked fraction can be much smaller. This is explored in more detail in Appendix B.

3 RESULTS

In this section we will explore the effects of baryon physics on the two-point correlation function at redshift zero. We will first consider the galaxy–galaxy and galaxy–matter correlation functions as these are the most directly observable. Since stellar masses are strongly model-dependent, we will switch from galaxies to subhaloes in Section 3.2, which allows us to examine how clustering statistics derived from dark matter only simulations will differ from those including baryons. Finally, in Section 3.3, we will take the change in the mass of subhaloes out of the equation, and consider the change in the correlation function for the exact same objects as a function of the model used.

3.1 Clustering of galaxies

3.1.1 Autocorrelation

In Fig. 1 we plot the galaxy autocorrelation functions, $\xi_{\text{gg}}(r)$, for models *REF* and *AGN* in three different bins of stellar mass, as indicated in the legend. The bottom panel shows the relative difference in the clustering strength of galaxies in these models. Since the clustering of haloes increases with mass, and since AGN feedback reduces the stellar content of massive haloes, one would expect galaxies in the *AGN* simulation to be more strongly clustered at fixed (high) stellar mass. As higher-mass galaxies are expected to host more powerful AGN, this effect is expected to increase with mass. This is indeed what we observe in Fig. 1: as long as we consider sufficiently large scales, galaxies in the *AGN* simulation show increased clustering relative to those in *REF*, and the relative difference between clustering strengths in the two simulations tends to increase with mass. For galaxies with stellar masses $M_* < 10^{10} h^{-1} M_{\odot}$ we expect the effect to be minor, since in such low-mass objects feedback is controlled by stellar rather than AGN feedback in these models (e.g. Haas et al. 2013).

Also indicated in the legend are the number of galaxies in each mass bin for each simulation, the first number corresponding to *REF* and the second to *AGN*. Because AGN feedback systematically lowers the stellar content of massive haloes, and since the number density of haloes decreases with mass, the *AGN* simulation suffers from somewhat worse statistics at high stellar masses than the *REF* simulation. However, this effect is only seen in the highest mass bin, $M_* > 10^{12} h^{-1} M_{\odot}$, and even in this mass range we can still draw robust conclusions for scales $r > 2 h^{-1}$ Mpc.

Note that any two subhaloes must have a finite minimum distance between them in order to, on the one hand, be recognized as separate objects, and on the other, not be tidally destroyed. As we identify galaxies by the subhaloes they occupy, this causes a slight

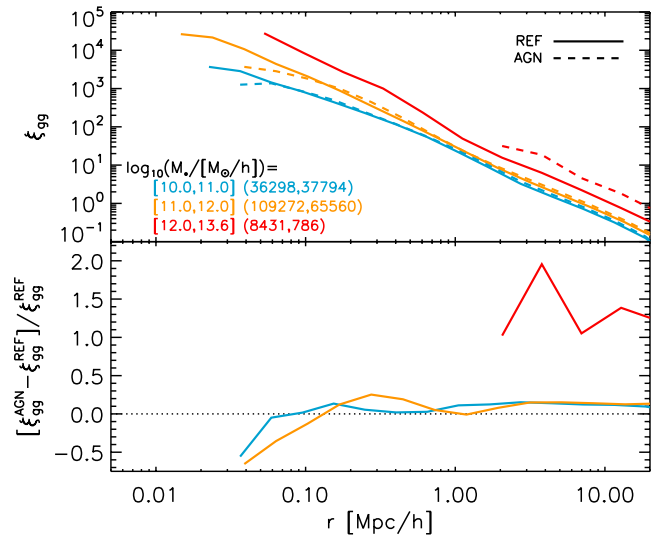


Figure 1. The galaxy autocorrelation function for the *REF* and *AGN* simulations (top), as well as the fractional difference between the two (bottom). Different colours correspond to different stellar masses, as indicated in the legend. The legend also shows the number of galaxies in each bin for each simulation (*REF*, *AGN*). At any mass, galaxies in *AGN* are more highly clustered than those in *REF* on large scales, an effect that increases sharply above $10^{12} h^{-1} M_{\odot}$, where AGN feedback is most important. Note that these effects may be underestimated for the two highest mass bins for reasons discussed in Section 3.1.3. The relative decrease in clustering for the *AGN* simulation on small scales is mostly a numerical effect (see text).

turnover in the galaxy correlation functions on small scales. Since this minimum distance increases with the size and therefore mass of the subhaloes hosting the galaxies, at fixed stellar mass this turnover is seen at larger scales in the model *AGN* than in *REF*. This in turn causes the galaxies in *AGN* to appear less clustered on small scales.

3.1.2 Cross-correlation with matter

Fig. 2 shows the galaxy–matter cross-correlation functions for these simulations, which are relevant for galaxy–galaxy lensing. Due to the high number of particles relative to the number of galaxies, the statistics are significantly improved relative to Fig. 1, and we can see clearly that including AGN feedback greatly increases the clustering of matter and galaxies at fixed stellar mass.² The relative increase of clustering with mass is more strongly scale-dependent than for the galaxy–galaxy case. The relative difference in clustering strength between *AGN* and *REF* is largest around $1 h^{-1}$ Mpc for the most massive galaxies, where galaxies at fixed stellar mass are nearly twice as strongly clustered with matter when AGN are included. At larger scales, *AGN* always shows ~ 50 per cent stronger clustering than *REF* for $M_* > 10^{12} h^{-1} M_{\odot}$. Even for galaxies in the stellar mass range $10^{11} < M_*/[M_{\odot} h^{-1}] < 10^{12}$ we see an increase in clustering of up to 150 per cent around $70 h^{-1}$ kpc, and an offset of ~ 20 per cent at all larger scales.

Interestingly, the relative difference in the galaxy–matter cross-correlation functions between *AGN* and *REF* increases towards

² Note that the number of objects in the two most massive bins, shown in the legend, is lower than for the autocorrelation function. This is because we now use the higher-resolution *L200* for all mass bins, whereas we previously used *L400* for the two highest mass bins to obtain better statistics (see Section 2.1).

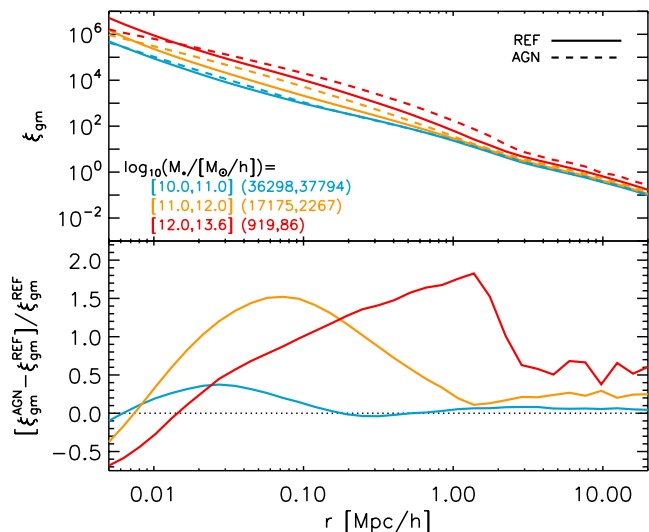


Figure 2. As in Fig. 1, but now showing the galaxy–matter cross-correlation function for the *REF* and *AGN* simulations. Except for sub-galactic scales, AGN feedback tends to increase the clustering of galaxies with matter at fixed stellar mass. Both the overall magnitude of the effect and the length scales over which it occurs increase with stellar mass, and for $M_* > 10^{12} h^{-1} M_\odot$ the increase in clustering with the matter distribution reaches values as high as 180 per cent.

smaller scales before suddenly dropping, causing galaxies to become *less* strongly clustered with the matter distribution in the *AGN* simulation on the very smallest scales probed here. This behaviour is caused by two competing effects, a point we will return to when discussing the subhalo–matter cross-correlation function in the next sections. On the one hand, the lowering of the stellar mass by AGN feedback tends to increase clustering at fixed stellar mass, and more so towards smaller scales, as galaxies of the same stellar mass now inhabit denser environments. On the other hand, as shown in e.g. Velliscig et al. (2014), a large amount of gas – and even dark matter – is removed from the galaxy, and sometimes from the halo entirely, decreasing the density peaks in the matter distribution (see e.g. van Daalen et al. 2011). As we can see from Fig. 2, the latter effect dominates on sub-galaxy scales ($r \lesssim 10 h^{-1} \text{kpc}$).

3.1.3 Caveats

We note that the effect of AGN feedback may be underestimated for massive galaxies due to two effects. The first only applies to the two highest mass bins and only to results based on the *L400* runs (i.e. the autocorrelation functions): the implementation of AGN feedback in these simulations is somewhat resolution dependent, and as a consequence its effect is weaker in the $400 h^{-1} \text{Mpc}$ box than in the $200 h^{-1} \text{Mpc}$ simulation. This is because the seed black holes can only be injected into resolved haloes, which corresponds to a minimum mass, that is eight times higher in the *L400* simulation than in the *L200* simulation (i.e. the difference in mass resolution). The result is that AGN feedback in the $400 h^{-1} \text{Mpc}$ box, used in the two highest mass bins in Figs 1 and 2, may be too weak for galaxies occupying haloes with masses $M \lesssim 10^{13} h^{-1} M_\odot$. In fact, while the effect of resolution is small for galaxies with masses $M_* > 10^{12} h^{-1} M_\odot$, for $10^{11} < M_* / [M_\odot h^{-1}] < 10^{12}$ the effect is significant: when using the higher-resolution *L200* simulation in this mass bin, we find an increase in galaxy–galaxy clustering relative to *REF* of ~ 50 per cent for $r \gtrsim 2 h^{-1} \text{Mpc}$.

The second effect is due to the way stellar mass is estimated in observations, where the use of an aperture excludes intracluster light. For the more massive galaxies in our sample, which host the most powerful AGN, this aperture size is typically significantly smaller than the size of the region containing the stars. However, simulated galaxies are assigned a stellar mass equal to the total mass in stars in its subhalo. The stellar mass of our most massive galaxies is therefore significantly higher than would be estimated observationally. Hence, the strong effects of AGN feedback that we find will be relevant for lower *observed* stellar masses than suggested by our plots.

Regardless, even without taking these effects into account, it is clear that AGN feedback plays an important role in the clustering of galaxies and matter, and should not be ignored in theoretical models that aim to predict $\xi_{\text{gm}}(r)$ to ~ 10 per cent accuracy or better, even when only considering relatively low stellar masses ($M_* = 10^{10}$ to $10^{11} h^{-1} M_\odot$).

At this point it is important to note that although our model *AGN* reproduces the stellar masses of group-sized haloes relatively well (McCarthy et al. 2010, 2011), predicted stellar masses are generally strongly model-dependent, as well as cosmology-dependent. Abundance matching studies, on the other hand, reproduce the stellar mass–halo mass relation by construction (e.g. Moster et al. 2010). Since clustering models typically employ the results from such studies, which in turn rely on dark matter only simulations, it is useful to consider the clustering of the subhaloes that host the galaxies and to select objects by their total subhalo mass, instead of by their stellar mass. This also allows us to consider the effect of galaxy formation relative to a dark matter only scenario. For the remainder of this paper, we will therefore focus on the clustering of subhaloes.

3.2 Clustering of subhaloes

3.2.1 Autocorrelation

The top panel of Fig. 3 shows the subhalo autocorrelation function, $\xi_{\text{ss}}(r)$, for three different simulations: *DMONLY*, *REF* and *AGN*. Different colours indicate different subsamples, selected by the *total* mass of the subhaloes, $M_{\text{sh,tot}}$, though we note that the results would have been very similar had we selected by dark matter mass. The correlation functions are displayed in the top panel, while in the middle panel and bottom panels the baryonic simulations are compared to *DMONLY*. From the top panel we can already see that subhalo clustering in the dark matter only simulation behaves quite differently from that in the baryonic models, especially on small scales ($r \lesssim 1 h^{-1} \text{Mpc}$). Vertical dotted lines indicate the median virial radii³ of subhaloes in each mass bin, which are similar to the scale at which the subhalo correlation functions for *DMONLY* turn over.

At the high-mass end, all three simulations show very similar behaviour. Looking at the middle and bottom panels, where we compare the autocorrelation of subhaloes in *REF* and *AGN* respectively to that in *DMONLY*, we see that all subhaloes in the baryonic simulations are typically ~ 10 per cent more strongly clustered on large scales than their dark matter only counterparts. As we will

³ We computed a characteristic size, r_{vir} , for each subhalo by taking its total mass, $M_{\text{sh,tot}}$, and treating it as the mass within a region with a mean overdensity of $\Delta = 200$ relative to ρ_{crit} (i.e. $r_{\text{vir}} \approx r_{200c}$). For reference, for a typical dark matter halo $r_{500c} \sim 0.65\text{--}0.75 r_{200c}$, where r_{500c} corresponds to the radius out to which the dominant baryonic component (hot gas) of groups and clusters is typically measured (e.g. Vikhlinin et al. 2006).

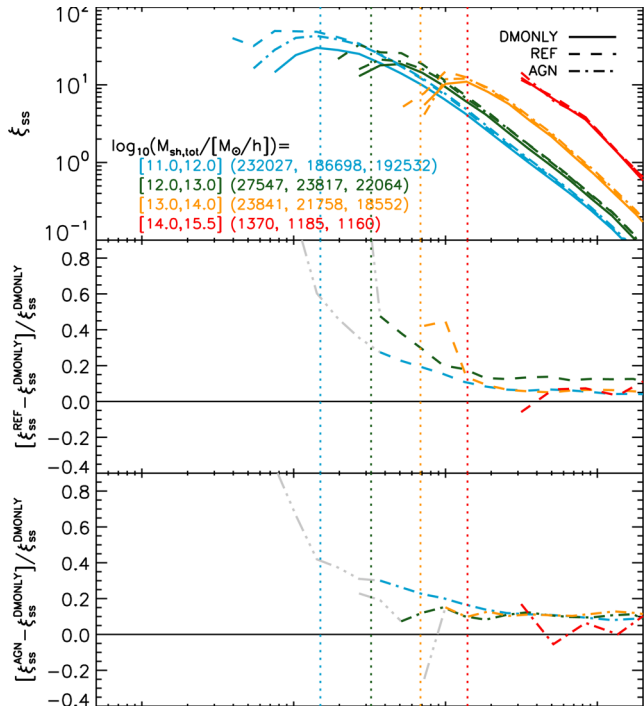


Figure 3. The subhalo autocorrelation function, $\xi_{ss}(r)$, for *DMONLY* (solid), *REF* (dashed) and *AGN* (dot-dashed lines), and the fractional differences between them. Different colours are used for different total subhalo masses, and the number of objects in each bin is indicated in the legend (*DMONLY*, *REF*, *AGN*). Top: the correlation functions for the three simulations. Vertical dotted lines indicate the median r_{vir} of the subhaloes. Middle: the fractional difference of subhalo clustering in *REF* relative to *DMONLY*. The curves are greyed out for radii where they may be biased due to subhalo non-detections (see Appendix B). Bottom: the fractional difference of subhalo clustering in *AGN* relative to *DMONLY*. Both baryonic simulations show increased clustering, and this effect is stronger on smaller scales. Note that the range on the y-axis is much smaller here than in Fig. 1.

demonstrate in Section 3.3, this difference is due to the reduction of subhalo mass caused by baryonic processes. For the larger subhaloes, $10^{13} < M_{\text{sh,tot}}/[M_{\odot} h^{-1}] < 10^{14}$, this offset is somewhat larger when AGN feedback is included, because supernova feedback alone cannot change the subhalo mass by as much as it can for lower halo masses (e.g. Sawala et al. 2013; Velliscig et al. 2014). The offset in clustering strength relative to *DMONLY* of the lowest-mass subhaloes is also slightly increased by the addition of AGN: while the masses of these subhaloes may seem to be somewhat low to be significantly affected by AGN feedback, we should keep in mind that satellite subhaloes may have lost part of their mass through tidal stripping. Moreover, these would correspond to subhaloes of a higher mass in a *DMONLY* simulation, as a significant fraction of the mass has been expelled. Additionally, low-mass subhaloes do not need to host AGN themselves to be affected by them: satellites in groups and clusters are sensitive to changes in the host halo profile and possibly increased stripping caused by the powerful AGN in the more massive galaxies in their environment.

The differences between the baryonic and dark matter only simulations increase rapidly for $r < 2r_{\text{vir}}$, at least for $M_{\text{sh,tot}} < 10^{14} h^{-1} M_{\odot}$. As we can see most easily in the top panel, subhaloes in the *REF* simulation are significantly more clustered on small scales than those in the *AGN* simulation, which seems to contradict the results of the previous section. This is because subhaloes in the *REF* simulation are more compact at fixed mass than those

in the *AGN* simulation, due to the additional form of feedback in the latter which removes more material from the centre and lowers the concentration in the inner parts of the subhaloes. However, the haloes in the *AGN* simulation are still more compact than those in *DMONLY* (see e.g. Velliscig et al. 2014). The increased concentration of subhaloes in baryonic simulations allows them to be identified as separate objects down to smaller scales, and also to withstand the effects of tidal stripping longer than their dark matter only counterparts. Both these effects tend to increase the clustering on small scales. This relative increase in the number density of subhaloes close to the centres of haloes in baryonic simulations was seen before by e.g. Macciò et al. (2006), Libeskind et al. (2010), Romano-Díaz et al. (2010) and Schewtschenko & Macciò (2011; although Romano-Díaz et al. 2010 note that without strong feedback, the effect may be reversed.). On the other hand, baryonic subhaloes are generally less massive when they are centrals, and those that become satellites typically fall in later due to the smaller virial radius of the main halo compared to a pure dark matter run, which means that they should experience less dynamical friction on scales where tidal stripping is not yet important. This is indeed what Schewtschenko & Macciò (2011) find, although this effect cannot be seen for the mass-selected sample shown in Fig. 3 due to the much larger effect of the change in mass.

We explore the clustering behaviour of baryonic satellites in more detail in Section 3.3.1. For now, we note that if our ability to detect baryonic subhaloes down to smaller radii than pure dark matter ones were the dominant cause of an increased number density of subhaloes at small separations in *REF* and *AGN*, this would introduce a bias towards observing a stronger clustering signal in baryonic models on scales $r \lesssim 2r_{\text{vir}}$.⁴ We discuss this possible source of error in Appendix B, and based on the results reported there we have chosen to show the relative differences in clustering as grey dot-dot-dot-dashed curves in Fig. 3 for subhalo masses and scales that may be significantly affected by this bias.

Comparing Figs 1 and 3, we see that the single act of adding AGN feedback affects the clustering of galaxies and subhaloes very differently. For galaxies, a strong increase in clustering is found for the highest-mass galaxies, and on large scales, since the same subhaloes host galaxies with a much lower stellar mass when AGN feedback is added. Low-mass galaxies are, however, not strongly affected by AGN feedback. For subhaloes, on the other hand, we find that the largest effects are found on small scales, and especially at the lowest masses: we find a strong *decrease* in clustering for $r \lesssim r_{\text{vir}}$ when adding AGN feedback to the reference model, regardless of halo mass, and far less change on large scales. These two main differences have two different causes. The large-scale differences between the effect of AGN feedback on galaxies and on subhaloes is that while AGN are powerful enough to quench star formation and to remove a lot of gas from galaxies, thus lowering the stellar mass, they are not powerful enough to significantly change the halo mass. However, as is shown in detail by Velliscig et al. (2014), and as we will also see in the next section, they do have a significant effect on the density profiles of subhaloes, and through this on their distribution. At fixed mass, the subhaloes in *REF* are more compact and more massive than those in *AGN*, causing both the satellite survival rate and the dynamical friction experienced by satellites to increase, which in turn causes the small-scale differences in clustering we just discussed.

⁴ We thank Raul Angulo for pointing out this potential problem.

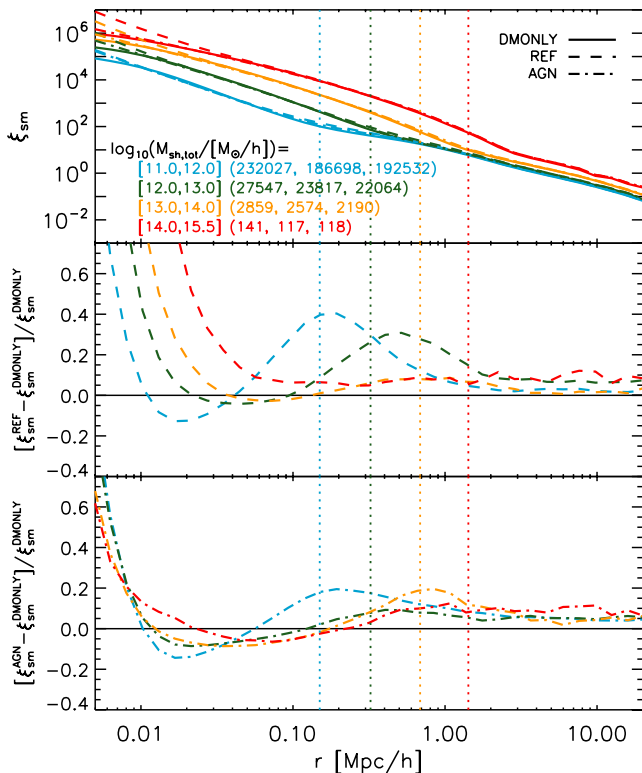


Figure 4. As Fig. 3, but now for the subhalo–mass cross-correlation function, $\xi_{\text{sm}}(r)$. Subhaloes are generally more strongly clustered with matter in the baryonic simulations than in *DMONLY*. The largest differences are found for *REF*, for which $\xi_{\text{sm}}(r)$ can be up to 40 per cent higher on intermediate scales for the lowest-mass subhaloes, and much higher still for any subhalo mass if sufficiently small scales are considered. There is also a constant 5 per cent difference in favour of the baryonic simulations on large scales, regardless of subhalo mass. While the *AGN* model seems to increase clustering at fixed subhalo mass less than *REF*, it does show a stronger decrease in clustering up to scales $r \sim 10^2 h^{-1}$ kpc. Note that in both cases the clustering differences between the models are strongly non-monotonic, which is caused by the interplay between the change in the total subhalo mass and the change in the subhalo mass profiles.

3.2.2 Cross-correlation with matter

We consider the subhalo–mass cross-correlation function in Fig. 4. From the top and middle panels we observe, as was the case for galaxy–galaxy clustering, that on the smallest scales and at fixed total mass, subhaloes cluster far more strongly with matter in the baryonic simulations than in the dark matter only simulations. Additionally, there is a constant 5 per cent offset in favour of baryonic simulations on the largest scales, for all halo masses. The baryonic bias increases as we move from large scales towards the virial radius, but, interestingly, the strength of the effect decreases below scales approximately corresponding to r_{vir} before picking up again at the smallest scales shown. This decrease below r_{vir} even causes the lowest-mass *DMONLY* subhaloes to be more strongly clustered than their *REF* counterparts around $r = 20 h^{-1}$ kpc. For *AGN*, this happens even for the highest-mass subhaloes, and over a larger range of scales.

As we will show in the next section, the strongly non-monotonic behaviour of the relative difference in ξ_{sm} between the baryonic simulations and *DMONLY* is caused by two counteracting effects. On the one hand, the lowered halo masses in the baryonic simulations tend to increase clustering at fixed mass on all scales. On the

other hand, while the dissipation associated with galaxy formation causes the inner halo profile to steepen, increasing clustering on small scales, the associated feedback causes the outer layers of the halo to expand, decreasing clustering on intermediate scales. This effect is stronger when *AGN* feedback is included. Note that we observe similar behaviour for the relative differences between the galaxy–matter cross-correlation functions for *REF* and *AGN*.

Furthermore, by comparing the bottom two panels, we can see that for low halo masses ($M_{\text{sh,tot}} < 10^{12} h^{-1} M_{\odot}$), for which *AGN* feedback is not very important, the small-scale clustering of haloes in *REF* and *AGN* is nearly identical, while subhaloes and matter cluster much more weakly on a range of scales around r_{vir} in *AGN*. On the other hand, for higher-mass haloes ($M_{\text{sh,tot}} > 10^{12} h^{-1} M_{\odot}$), significant differences can be seen down from the smallest scales out to $r \sim 1 h^{-1}$ Mpc. This again confirms the strong effect that *AGN* feedback has on the mass distribution: the higher the mass of the halo, the more important feedback from supermassive black holes is in removing material from the centre. This in turn flattens the mass profiles of the haloes and smooths out the density peaks, decreasing the small-scale lensing signal relative to *REF*.

As we have already pointed out several times, the most important cause of the increase in clustering due to galaxy formation with strong feedback is the lowering of the mass of objects. However, secondary effects, such as the resulting changes in the dynamics and density profiles of haloes, are also expected to be significant. To disentangle these types of effects, we will use our linking scheme to match subhaloes between different simulations, allowing us to see if any significant difference in the clustering remains once the change in mass has been accounted for.

3.3 Accounting for the change in mass

As we are mainly interested in how galaxy formation changes the clustering of objects with respect to a dark matter only scenario, we use the linking algorithm described in Section 2.3 to link subhaloes in *REF* and *AGN* to those in *DMONLY*, and assign all objects the mass of their *DMONLY* counterpart. Note that this means that there are in fact two different *DMONLY* versions of each correlation function: one derived using all subhaloes for which a counterpart was found in *REF*, and one derived using all subhaloes for which a counterpart was found in *AGN*. In practice, however, the linked halo samples are nearly identical, and the resulting correlation functions for *DMONLY* are virtually indistinguishable. We therefore show only one of these in the top panels of Figs 5 and 6, although both are used to determine the differences with respect to *REF* and *AGN*.

3.3.1 Autocorrelation of linked subhaloes

We first consider Fig. 5, where we show the impact of galaxy formation on the clustering of subhaloes once the change in mass has been accounted for. Comparing first the sample sizes (numbers in the legend) to those in Fig. 3, we see that nearly all *DMONLY* subhaloes have a match in each of the baryonic simulations.⁵ Note that the first number in the legend now indicates the sample size of subhaloes linked between *DMONLY* and *REF*, while the second gives the number of subhaloes linked between *DMONLY* and *AGN*.

⁵ As we now select subhaloes by the mass of their *DMONLY* counterpart, the number of subhaloes can only be directly compared to those of *DMONLY* in Fig. 3, not to the number of baryonic subhaloes in Fig. 3.

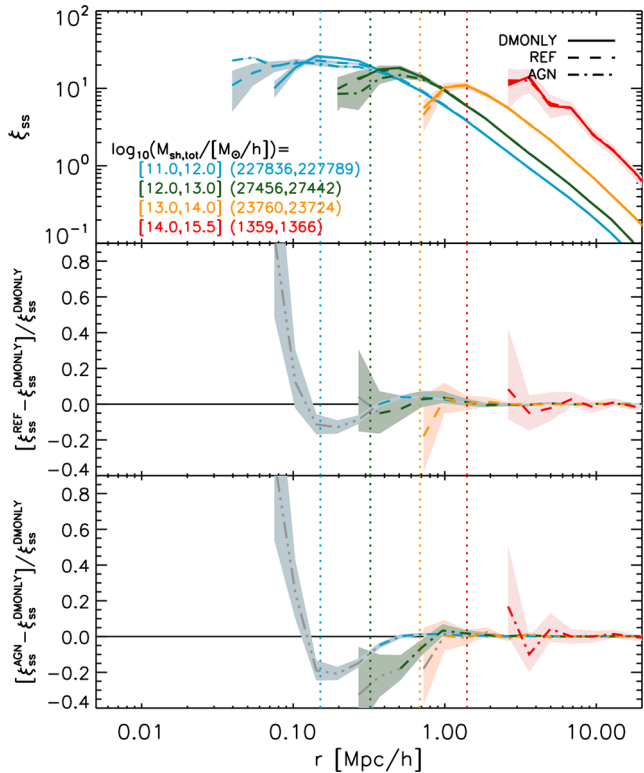


Figure 5. As Fig. 3, but now only showing the autocorrelation functions for subhaloes linked between a baryonic simulation and *DMONLY*, and selected based on their mass in the latter. Relative to Fig. 3, this procedure removes the effects of changes in the subhalo masses. As the numbers in the legend imply, almost the exact same haloes are linked with dark matter only haloes in both cases. The bottom two panels immediately show that in all cases no differences $\gtrsim 5$ per cent in ξ_{ss} remain on scales $r \gg r_{\text{vir}}$, indicating that the differences we saw in Fig. 3 on these scales were due to the masses of the objects changing. For smaller scales, and especially for low-mass subhaloes, the change in dynamics of the objects in the baryonic simulations can have significant effects, which can primarily be seen as a decrease in clustering on scales $r \lesssim 2r_{\text{vir}}$. Shaded areas indicate the regions allowed by 1σ bootstrap errors, which show that the relative small-scale decrease of clustering of low-mass baryonic subhaloes is significant.

We have now also performed 500 bootstrap resamplings for each pair of simulations, and show the 1σ errors derived from these as shaded areas in the figure. As we are now using the exact same (linked) sample of subhaloes for any pair of simulations, we are able to avoid overestimating the errors due to the false assumption that the halo samples of the simulations are independent. Similar errors are expected for Fig. 3.

Comparing the bottom two panels of Fig. 5 to those of Fig. 3, we immediately see that essentially nothing of the ~ 10 per cent difference in the clustering amplitude on large scales remains, confirming that this was solely due to galaxy formation changing the masses of these subhaloes. By accounting for the change in the masses of objects due to the effects of baryon physics, one will therefore automatically obtain the correct autocorrelation function at all halo masses, on scales $r \gg r_{\text{vir}}$.

However, on smaller scales the changes in the dynamics of subhaloes in the baryonic runs become important. This is especially the case for low-mass objects, which are often satellites. As we discussed in Section 3.2.1, Schewtschenko & Macciò (2011) have shown that, initially, satellites in dark matter only simulations move in closer to the centre of the main halo in the same amount of

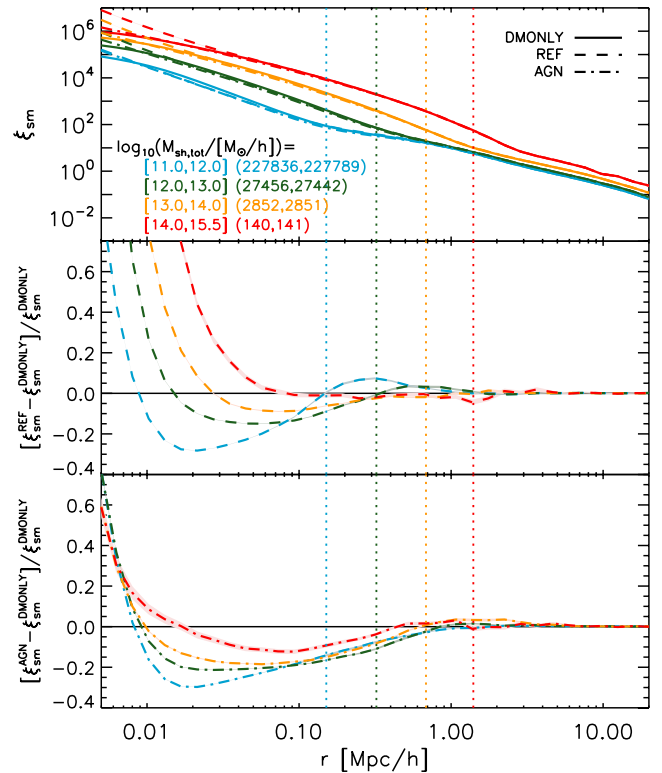


Figure 6. As Fig. 4, but now only showing the cross-correlation functions between matter and subhaloes that have been linked between a baryonic simulation and *DMONLY*, and that have been selected based on their mass in the latter. Relative to Fig. 4, this procedure removes the effects of changes in the subhalo masses, leaving only the effect on the mass profiles and the changes in the positions of the subhaloes. As can be seen from the bottom panel, the change of the mass profile tends to increase the clustering on the very smallest scales (where baryons cool to), but decreases it on intermediate scales (where baryons are evacuated). The latter effect is stronger when AGN feedback is included, and significant over a larger range of scales, for all masses. Shaded areas indicate the regions allowed by 1σ bootstrap errors, which are typically much smaller than the widths of the lines.

time, which is due in part to the decrease in the virial radius of the main halo when baryons are included (also found for baryonic haloes in our simulations, see Velliscig et al. 2014), and in part to the increased dynamical friction experienced by the more massive dark matter satellites. However, as the satellites undergo tidal stripping, baryonic subhaloes are able to retain more of their mass due to their increased concentrations, which causes the situation to reverse on small scales, increasing the number density of baryonic subhaloes relative to pure dark matter ones. This was also found by e.g. Macciò et al. (2006), Libeskind et al. (2010) and Romano-Díaz et al. (2010). However, at the same time one expects to see an increase in the number density – and consequently, the clustering – of baryonic satellite subhaloes at small scales due to the ability to trace baryonic subhaloes longer during infall. This resolution effect could lead to a bias at small separations.

To account for this potential bias, we consider the fraction of subhaloes in *DMONLY* for which a link could be found in *REF* in Appendix B. There we show that the fraction of linked subhaloes decreases strongly on small scales for low-mass subhaloes. Higher-resolution simulations are needed to investigate whether the increased survival rate of baryonic subhaloes, and the resulting increase in clustering seen in Figs 3 and 5 on scales $r \lesssim r_{\text{vir}}$, is physical

or not. We have therefore greyed out the curves in these figures on scales where this bias may play a significant role.

However, even after accounting for this potential bias, interesting differences in clustering remain on scales $r \lesssim 2r_{\text{vir}}$, as Fig. 5 shows. Especially in the *AGN* simulation, subhaloes tend to be ~ 10 per cent less clustered at $r \sim r_{\text{vir}}$. A very small increase in clustering (~ 1 per cent) can be seen on slightly larger scales, $r \sim 3 - 4r_{\text{vir}}$. Both these differences could be explained by the combination of the greater dynamical friction initially experienced by dark matter only subhaloes, together with the delayed infall of baryonic subhaloes. We plan to investigate these effects further in a follow-up paper where we consider the differences in the satellite profiles due to galaxy formation.

Note that small changes in the simulation code (such as changing the level of optimization when compiling the simulation code) can shift the positions of satellite galaxies and subhaloes by small amounts, even if we start from identical initial conditions.⁶ However, as almost all these shifts are random, they average out for two-point statistics. Shifts due to dynamical friction and similar effects acting on satellites are the exceptions, as these tend to systematically move satellite subhaloes closer to their respective centrals.

3.3.2 Cross-correlation with matter

Finally, we consider what remains of the baryonic effects on the subhalo–matter cross-correlation function after accounting for the change in the masses of subhaloes. Here, too, we show 1σ errors in all panels, now derived from 10 000 bootstrap resamplings. In many cases, the errors are smaller than the widths of the lines.

Comparing the bottom panels of Fig. 6 to those of Fig. 4, we see that while the large-scale offset is now completely removed, we are left with a non-negligible effect on scales $r \lesssim 1 h^{-1}$ Mpc for all subhalo masses. This again shows the strong effect that feedback can have on the mass distribution: both supernova and AGN feedback move matter to large scales, decreasing $\xi_{\text{gm}}(r)$. We see that, especially when AGN feedback is included, this can significantly affect clustering out to several times the virial radius, which matches the findings of van Daalen et al. (2011) and Velliscig et al. (2014). Note that this also confirms that the findings of van Daalen et al. (2011), namely that AGN feedback decreases the matter power spectrum at the 1–10 per cent level out to extremely large scales ($r \sim 10 h^{-1}$ Mpc), are caused by the effect (in Fourier space) of a systematic change in the profile of haloes, rather than by AGN somehow having a significant effect the mass distribution out to more than 10 times the virial radius of the haloes they occupy.

There are strong similarities between the relative differences that remain for ξ_{sm} and the relative differences of halo profiles shown in Velliscig et al. (2014) for the same models, leaving no doubts as to the origin of the signal we see here. The strength of the baryonic effect decreases with increasing mass, but is still highly significant at the mass scales of groups and clusters, although it does not extend beyond the virial radius for the highest-mass subhaloes. The lowest-mass subhaloes we consider here experience a maximum decrease in the cross-correlation with matter of 30 per cent, relative to a dark matter only scenario, and even the most massive subhaloes are 10 per cent less strongly clustered with the matter distribution around $r = 100 h^{-1}$ kpc when AGN are included. On the smallest

scales, the increased clustering due to the cooling of baryons still dominates. Note also that the small-scale differences that we found in Fig. 4 between *REF* and *AGN* remain.

These results show us that assigning subhaloes in a dark matter only simulation the masses they would have had if galaxy formation and efficient feedback had been included allows one to obtain the correct clustering predictions on scales $r \gg 1 h^{-1}$ Mpc. However, on smaller scales one cannot correctly predict the cross-correlation with matter, and hence the galaxy–galaxy lensing signal, to better than ~ 10 per cent accuracy without taking into account the change in the mass distribution.

4 SUMMARY

In this work we investigated how the galaxy and subhalo two-point autocorrelation functions and the cross-correlations with the matter, a measure of the galaxy–galaxy lensing signal, are modified by processes associated with galaxy formation. We utilized a set of cosmological, hydrodynamical simulations with models from the OWLS project, run with more particles and an updated cosmology relative to previous OWLS simulations, to examine what the combined effects on the auto- and cross-correlation functions are of adding baryons and radiative (metal-line) cooling, star formation, chemical enrichment and supernova feedback to a dark matter only simulation, as well as the further addition of a prescription of AGN feedback that reproduces observations of groups and clusters. As nearly all clustering models employed in the literature rely on pure dark matter distributions, from either N -body simulations or halo model type prescriptions, it is important to quantify just how important the effects of baryons and galaxy formation are.

Our findings can be summarized as follows.

(i) The stellar masses of galaxies are strongly decreased by (AGN) feedback at fixed subhalo mass, which in turn tends to greatly increase the clustering of galaxies at fixed stellar mass. More importantly for semi-analytical and halo models, the masses of subhaloes are also significantly decreased by the effects of feedback, the result of which is an increase in clustering of ~ 10 per cent on scales $r \gg 1 h^{-1}$ Mpc, for the full range of subhalo masses considered here ($M_{\text{sh,tot}} = 10^{11} - 10^{15.5} h^{-1} M_{\odot}$). This effect is much stronger on smaller scales.

(ii) Both the change in subhalo mass and the modified subhalo profiles act to change the subhalo–matter cross-correlation function by ~ 5 per cent on large scales, and significantly more on sub-Mpc scales. The modulation of the signal is strongly non-monotonic and mass-dependent, with both significant increases and decreases in clustering on different scales.

We used the identical initial conditions of our simulations to link each baryonic subhalo with its dark matter only counterpart, allowing us to effectively exclude the effect of galaxy formation on the change in the masses of these objects. Nearly all subhaloes are successfully matched in this way.

(i) While accounting for the change in mass of subhaloes removes essentially all of the baryonic effects on the autocorrelation of subhaloes on scales $r \gg r_{\text{vir}}$, deviations ~ 10 per cent remain on scales $r \lesssim 2r_{\text{vir}}$, where r_{vir} is the virial radius of the subhalo. We argued that these deviations are mainly caused by the differences in the dynamics of satellites, such as the initially greater dynamical friction experienced by the more massive, recently accreted pure dark matter satellites, and the increased concentration of baryonic subhaloes.

⁶ The *rms* shift in position for subhaloes between *DMONLY* and *AGN* is about $0.04 r_{\text{vir}}$. Similar values are found for shifts between subhaloes in *DMONLY* and *REF*.

(ii) Finally, on scales $r \lesssim 1 h^{-1}$ Mpc strong deviations in the subhalo–matter cross-correlation function remain after accounting for the change in the masses of subhaloes. While on galactic scales ($\lesssim 10 h^{-1}$ kpc) the clustering of subhaloes with matter is always much higher in a baryonic simulation than in the corresponding dark matter only simulation, the inclusion of baryons results in a significant decrease of the cross-correlation for $r \gtrsim 10 h^{-1}$ kpc. These effects are stronger for lower-mass subhaloes, reaching up to 30 per cent for subhaloes with masses $10^{11} < M_{\text{sh,tot}}/[M_{\odot} h^{-1}] < 10^{12}$. When AGN feedback is included, ξ_{sm} decreases by ~ 10 per cent relative to a dark matter only simulation for $r \sim 10^2 h^{-1}$ kpc, even for subhalo masses $M_{\text{sh,tot}} > 10^{14} h^{-1} M_{\odot}$. Mass- and radius-dependent rescalings of halo profiles which extend to several times the virial radius would be needed to account for this effect in dark matter only simulations.

We note that while many of our results rely on a model that includes AGN feedback, other feedback processes may have similar effects on clustering. In principle, any other mechanism that is also effective at high masses, sufficiently reducing the stellar masses of massive galaxies, and allows one to reproduce the global properties of groups and clusters, may show similar effects to those shown here for AGN feedback. For example, a model in which a top-heavy IMF is used in high-pressure environments, such as the OWLS model *DBLIMF*, may have the same qualitative effect on clustering (see e.g. van Daalen et al. 2011).

We stress that while the effects discussed in this paper will certainly need to be modelled in order to achieve the accuracy needed to interpret upcoming cosmological data sets to their full potential, both our knowledge of the relevant physics involved and the currently achievable resolution in cosmological simulations still allow for significant uncertainty in the clustering measures discussed here. The same holds for quantities such as the halo or cluster mass function: much work is yet to be done before we can converge on a realistic prescription of galaxy formation, with uncertainties small enough to match observations in the era of precision cosmology. Although approaches based on dark matter only models, such as semi-analytical modelling or HODs, are able to match the observed galaxy mass function, our results imply that their predictions for galaxy–galaxy and galaxy–mass clustering will have errors greater than 10 per cent on sub-Mpc scales, unless the simulation results are modified to correctly account for the effects of baryons on the distributions of mass and satellites.

ACKNOWLEDGEMENTS

The authors thank Raul Angulo, Marcello Cacciato and Simon White for useful comments and discussions, and Marcello Cacciato also for comments on the manuscript. We would also like to thank the anonymous referee for suggestions that improved this paper. The simulations presented here were run on the Cosmology Machine at the Institute for Computational Cosmology in Durham (which is part of the DiRAC Facility jointly funded by STFC, the Large Facilities Capital Fund of BIS, and Durham University) as part of the Virgo Consortium research programme. This work was sponsored by the Dutch National Computing Facilities Foundation (NCF) for the use of supercomputer facilities, with financial support from the Netherlands Organization for Scientific Research (NWO). We also gratefully acknowledge support from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement

278594-GasAroundGalaxies and from the Marie Curie Training Network CosmoComp (PITN-GA-2009- 238356).

REFERENCES

- Abadi M. G., Navarro J. F., Fardal M., Babul A., Steinmetz M., 2010, *MNRAS*, 407, 435
- Balaguera-Antolínez A., Porciani C., 2013, *J. Cosmology Astropart. Phys.*, 4, 22
- Balogh M. L., Pearce F. R., Bower R. G., Kay S. T., 2001, *MNRAS*, 326, 1228
- Baugh C. M., 2006, *Rep. Progress Phys.*, 69, 3101
- Behroozi P. S., Conroy C., Wechsler R. H., 2010, *ApJ*, 717, 379
- Berlind A. A., Weinberg D. H., 2002, *ApJ*, 575, 587
- Booth C. M., Schaye J., 2009, *MNRAS*, 398, 53
- Bryan S. E., Kay S. T., Duffy A. R., Schaye J., Vecchia C. D., Booth C. M., 2013, *MNRAS*, 429, 3316
- Casarini L., Macciò A. V., Bonometto S. A., Stinson G. S., 2011, *MNRAS*, 412, 911
- Conroy C., Wechsler R. H., Kravtsov A. V., 2006, *ApJ*, 647, 201
- Cooray A., Sheth R., 2002, *Phys. Rep.*, 372, 1
- Cui W., Borgani S., Dolag K., Murante G., Tornatore L., 2012, *MNRAS*, 423, 2279
- Dalla Vecchia C., Schaye J., 2008, *MNRAS*, 387, 1431
- Dolag K., Borgani S., Murante G., Springel V., 2009, *MNRAS*, 399, 497
- Duffy A. R., Schaye J., Kay S. T., Dalla Vecchia C., Battye R. A., Booth C. M., 2010, *MNRAS*, 405, 2161
- Gnedin O. Y., Kravtsov A. V., Klypin A. A., Nagai D., 2004, *ApJ*, 616, 16
- Governato F. et al., 2012, *MNRAS*, 422, 1231
- Guillet T., Teyssier R., Colombi S., 2010, *MNRAS*, 405, 525
- Guo Q., White S., Li C., Boylan-Kolchin M., 2010, *MNRAS*, 404, 1111
- Haas M. R., Schaye J., Booth C. M., Dalla Vecchia C., Springel V., Theuns T., Wiersma R. P. C., 2013, *MNRAS*, 435, 2931
- Jing Y. P., Mo H. J., Börner G., 1998, *ApJ*, 494, 1
- Jing Y. P., Zhang P., Lin W. P., Gao L., Springel V., 2006, *ApJ*, 640, L119
- Kazantzidis S., Kravtsov A. V., Zentner A. R., Allgood B., Nagai D., Moore B., 2004, *ApJ*, 611, L73
- Killedar M., Borgani S., Meneghetti M., Dolag K., Fabjan D., Tornatore L., 2012, *MNRAS*, 427, 533
- Komatsu E. et al., 2011, *ApJS*, 192, 18
- Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, *ApJ*, 609, 35
- Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64
- Libeskind N. I., Yepes G., Knebe A., Gottlöber S., Hoffman Y., Knollmann S. R., 2010, *MNRAS*, 401, 1889
- Macciò A. V., Moore B., Stadel J., Diemand J., 2006, *MNRAS*, 366, 1529
- Martizzi D., Teyssier R., Moore B., Wentz T., 2012, *MNRAS*, 422, 3081
- Martizzi D., Mohammed I., Teyssier R., Moore B., 2013, preprint ([arXiv:1307.6002](https://arxiv.org/abs/1307.6002))
- McCarthy I. G. et al., 2010, *MNRAS*, 406, 822
- McCarthy I. G., Schaye J., Bower R. G., Ponman T. J., Booth C. M., Dalla Vecchia C., Springel V., 2011, *MNRAS*, 412, 1965
- Mead J. M. G., King L. J., Sijacki D., Leonard A., Puchwein E., McCarthy I. G., 2010, *MNRAS*, 406, 434
- Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, *ApJ*, 710, 903
- Muldrew S. I., Pearce F. R., Power C., 2011, *MNRAS*, 410, 2617
- Romano-Díaz E., Shlosman I., Heller C., Hoffman Y., 2010, *ApJ*, 716, 1095
- Rudd D. H., Zentner A. R., Kravtsov A. V., 2008, *ApJ*, 672, 19
- Sawala T., Frenk C. S., Crain R. A., Jenkins A., Schaye J., Theuns T., Zavala J., 2013, *MNRAS*, 431, 1366
- Scannapieco C. et al., 2012, *MNRAS*, 423, 1726
- Schaye J., Dalla Vecchia C., 2008, *MNRAS*, 383, 1210
- Schaye J. et al., 2010, *MNRAS*, 402, 1536
- Schewtschenko J. A., Macciò A. V., 2011, *MNRAS*, 413, 878
- Semboloni E., Hoekstra H., Schaye J., van Daalen M. P., McCarthy I. G., 2011, *MNRAS*, 417, 2020

Semboloni E., Hoekstra H., Schaye J., 2013, MNRAS, 434, 148
 Shankar F., Lapi A., Salucci P., De Zotti G., Danese L., 2006, ApJ, 643, 14
 Simha V., Cole S., 2013, MNRAS, 436, 1142
 Simha V., Weinberg D. H., Davé R., Fardal M., Katz N., Oppenheimer B. D., 2012, MNRAS, 423, 3458
 Springel V., 2005, MNRAS, 364, 1105
 Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726
 Springel V., Di Matteo T., Hernquist L., 2005, MNRAS, 361, 776
 Stanek R., Rudd D., Evrard A. E., 2009, MNRAS, 394, L11
 Stott J. P. et al., 2012, MNRAS, 422, 2213
 Tinker J. L., Weinberg D. H., Zheng Z., Zehavi I., 2005, ApJ, 631, 41
 Tissera P. B., White S. D. M., Pedrosa S., Scannapieco C., 2010, MNRAS, 406, 922
 Vale A., Ostriker J. P., 2004, MNRAS, 353, 189
 van Daalen M. P., Schaye J., Booth C. M., Dalla Vecchia C., 2011, MNRAS, 415, 3649
 van den Bosch F. C., More S., Cacciato M., Mo H., Yang X., 2013, MNRAS, 430, 725
 Velliscig M., van Daalen M. P., Schaye J., McCarthy I. G., Cacciato M., Le Brun A. M. C., Dalla Vecchia C., 2014, preprint (arXiv:1402.4461)
 Vikhlinin A., Kravtsov A., Forman W., Jones C., Markevitch M., Murray S. S., Van Speybroeck L., 2006, ApJ, 640, 691
 Wechsler R. H., Zentner A. R., Bullock J. S., Kravtsov A. V., Allgood B., 2006, ApJ, 652, 71

Weinberg D. H., Colombi S., Davé R., Katz N., 2008, ApJ, 678, 6
 White M., 2004, Astropart. Phys., 22, 211
 Wiersma R. P. C., Schaye J., Smith B. D., 2009a, MNRAS, 393, 99
 Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009b, MNRAS, 399, 574
 Yang X., Mo H. J., van den Bosch F. C., 2003, MNRAS, 339, 1057
 Yang X., Kratochvil J. M., Huffenberger K., Haiman Z., May M., 2013, Phys. Rev. D, 87, 023511
 Zentner A. R., Semboloni E., Dodelson S., Eifler T., Krause E., Hearin A. P., 2013, Phys. Rev. D, 87, 043509
 Zhan H., Knox L., 2004, ApJ, 616, L75

APPENDIX A: CONVERGENCE TESTS

Here we investigate the effects of changing the box size or resolution of the simulations used in this paper on the subhalo autocorrelation function, as this is the main focus of this paper. We will also briefly discuss the effects on the subhalo–matter cross-correlation function.

In Fig. A1, we show the subhalo autocorrelation functions for models *DMONLY* and *REF*. For clarity the correlation functions for the *AGN* model are not shown, but the results are very similar. Contrary to what was done for the figures in the main text, here we do not impose a minimum number of pairs per bin. We vary both the

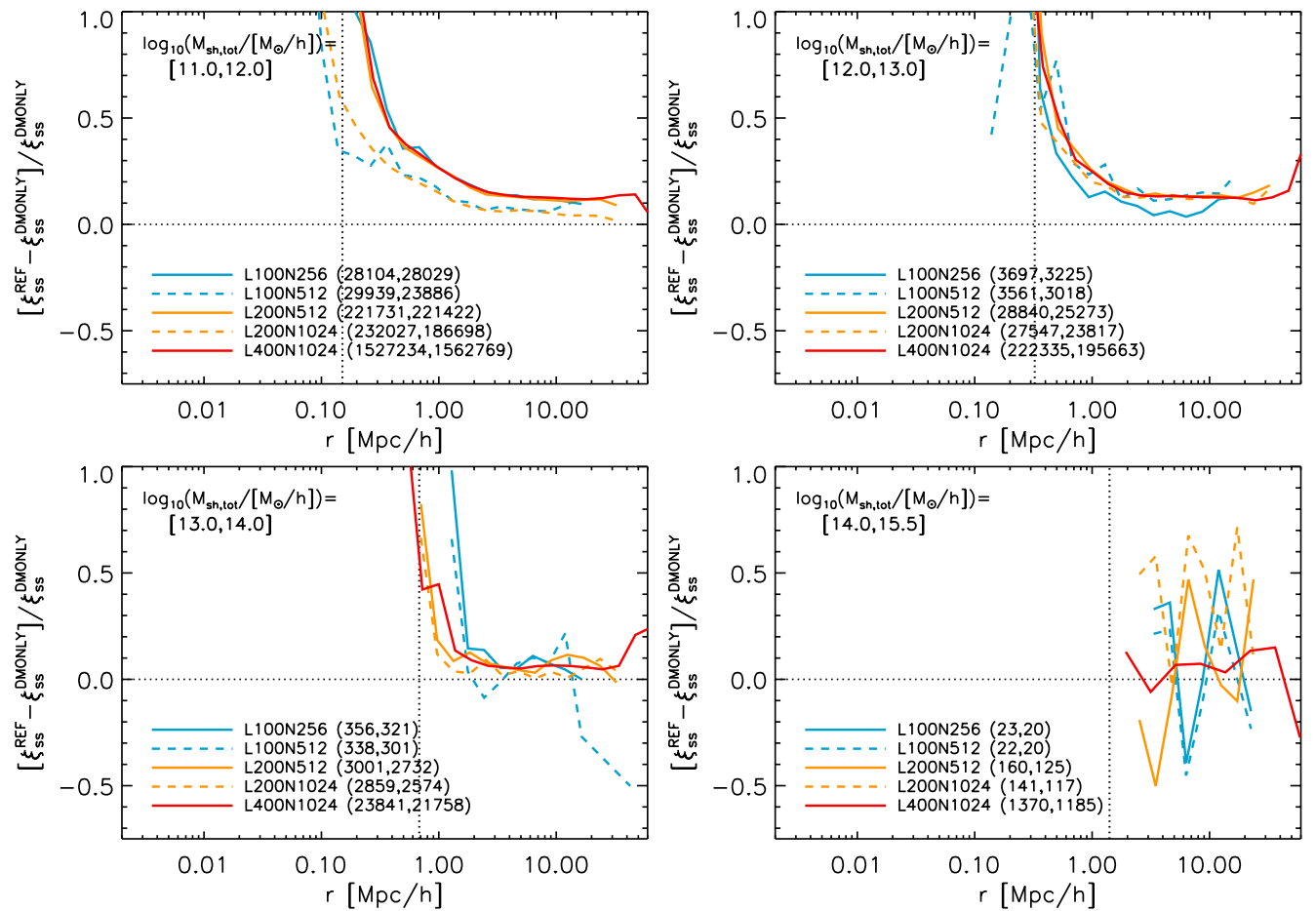


Figure A1. The relative differences in the subhalo autocorrelation functions between models *DMONLY* and *REF*, split by subhalo mass as indicated in the top left of each panel. Contrary to the plots shown in other sections, no minimum number of pairs per bin is imposed. The box sizes and particle numbers, as well as the subhalo numbers for *DMONLY* and *REF*, respectively, are indicated in the legend, and a vertical dotted line indicates the mean virial radius in each mass bin. At fixed resolution (same line style) very little changes, although the effect of the better statistics offered by a larger volume is apparent. At fixed box size (same colour) the results are also very similar, except for the lowest mass bin, where the small-scale clustering is resolution dependent. Note that all simulations show excellent agreement for $10^{12} < M_{\text{sh,tot}}/[M_{\odot} h^{-1}] < 10^{13}$, where neither resolution nor volume is an issue.

box size and particle number in a systematic way: for simulations shown with the same line style (either solid or dashed) we vary the box size at fixed resolution, while for simulations shown with the same colour we vary the resolution at fixed box size.

We first consider the effect of changing the size of the simulated volume. Looking at the solid and dashed lines separately, we can see that very little changes at fixed resolution, except that the results clearly benefit from the better statistics offered by a larger volume. This is noticeable both for the rare high-mass objects, on any scale, and for low-mass objects on the very smallest scales, where very few pairs are found.

If we instead consider each colour of Fig. A1 separately, we see that at fixed box size the results are also very similar. The exception is the lowest mass bin, $10^{11} < M_{\text{sh,tot}}/[M_{\odot} h^{-1}] < 10^{12}$, where the correlation function is clearly resolution dependent when baryons are included. This is because these subhaloes contain only $\sim 10^2$ particles in the low-resolution simulations, which is not quite enough for convergence, especially when feedback processes are included. We have verified that the subhalo mass functions of the highest-resolution simulations shown here are indeed converged using simulations with smaller volumes and higher resolutions (not shown here). The results for the second mass bin on the other hand, $10^{12} < M_{\text{sh,tot}}/[M_{\odot} h^{-1}] < 10^{13}$, are fully consistent between the different resolutions shown here.

We have repeated these same resolution tests for the autocorrelation functions of linked subhaloes, shown in Fig. A2. Here, too, we see that our results are converged for $M_{\text{sh,tot}} > 10^{12} h^{-1} M_{\odot}$.

Based on these tests, we choose to use the higher-resolution *L200N1024* simulations for subhaloes with masses $10^{11} < M_{\text{sh,tot}}/[M_{\odot} h^{-1}] < 10^{13}$, and take advantage of the better statistics offered by the *L400N1024* simulations for subhalo masses $M_{\text{sh,tot}} > 10^{13} h^{-1} M_{\odot}$. Similarly, we opt to use the higher-resolution simulation for the autocorrelation function of galaxies with stellar masses $10^9 < M_{*}/[M_{\odot} h^{-1}] < 10^{11}$, and the larger-volume simulation for galaxies with $M_{*} > 10^{11} h^{-1} M_{\odot}$.

We also verified that the cross-correlation functions shown in this work are sufficiently converged (not shown). For the subhalo–matter (and galaxy–matter) cross-correlation functions, statistics are less of an issue, as the number of particles is the same for the *L200* and *L400* simulations. In other words, while for the autocorrelation functions the number of pairs scales as N_{obj}^2 , the number of pairs for the cross-correlation functions scales as $N_{\text{obj}}N_{\text{part}}$, where $N_{\text{part}} \gg N_{\text{obj}}$. Resolution is still an issue, however: while simulations including baryons always show stronger clustering on galaxy scales than *DMONLY*, the exact scale on which the transition of a relative increase to a relative decrease in clustering occurs depends somewhat on the softening length. Additionally, as we discussed briefly in Section 3.1, the effect of AGN feedback is resolution-dependent in our simulations, due to the fact that seed black holes can only be inserted in resolved haloes. AGN feedback may therefore be weaker at the *L400* resolution than at the *L200* resolution, while the strength of the feedback in the latter was deemed realistic. We therefore choose to use the *L200* simulations at all masses when considering the cross-correlation functions ξ_{gm} and ξ_{sm} , valuing resolution over volume.

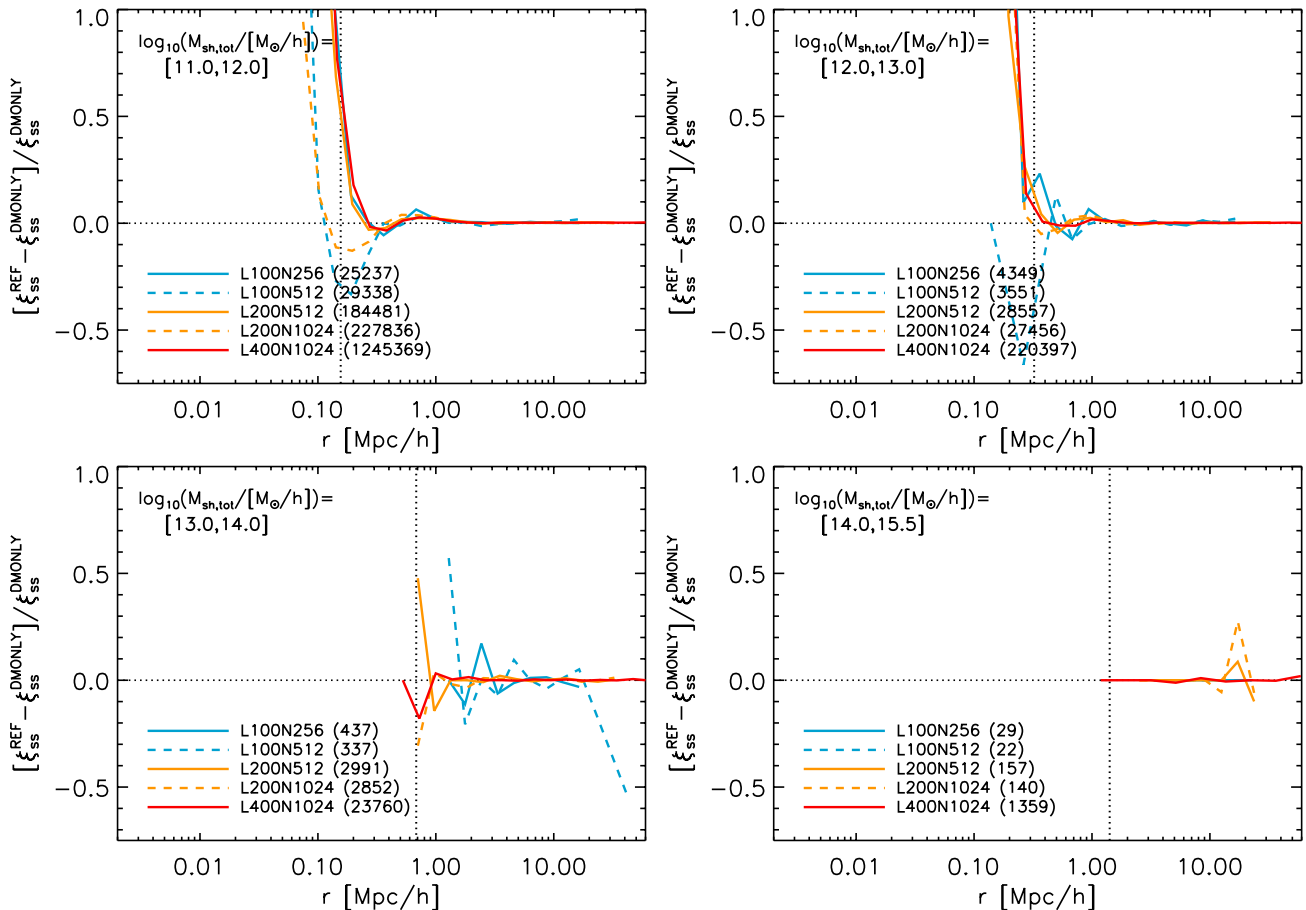


Figure A2. As Fig. A1, but now only showing the autocorrelation functions for subhaloes linked between *REF* and *DMONLY*, and selected based on their mass in the latter. The convergence here is very similar to that seen in Fig. A1.

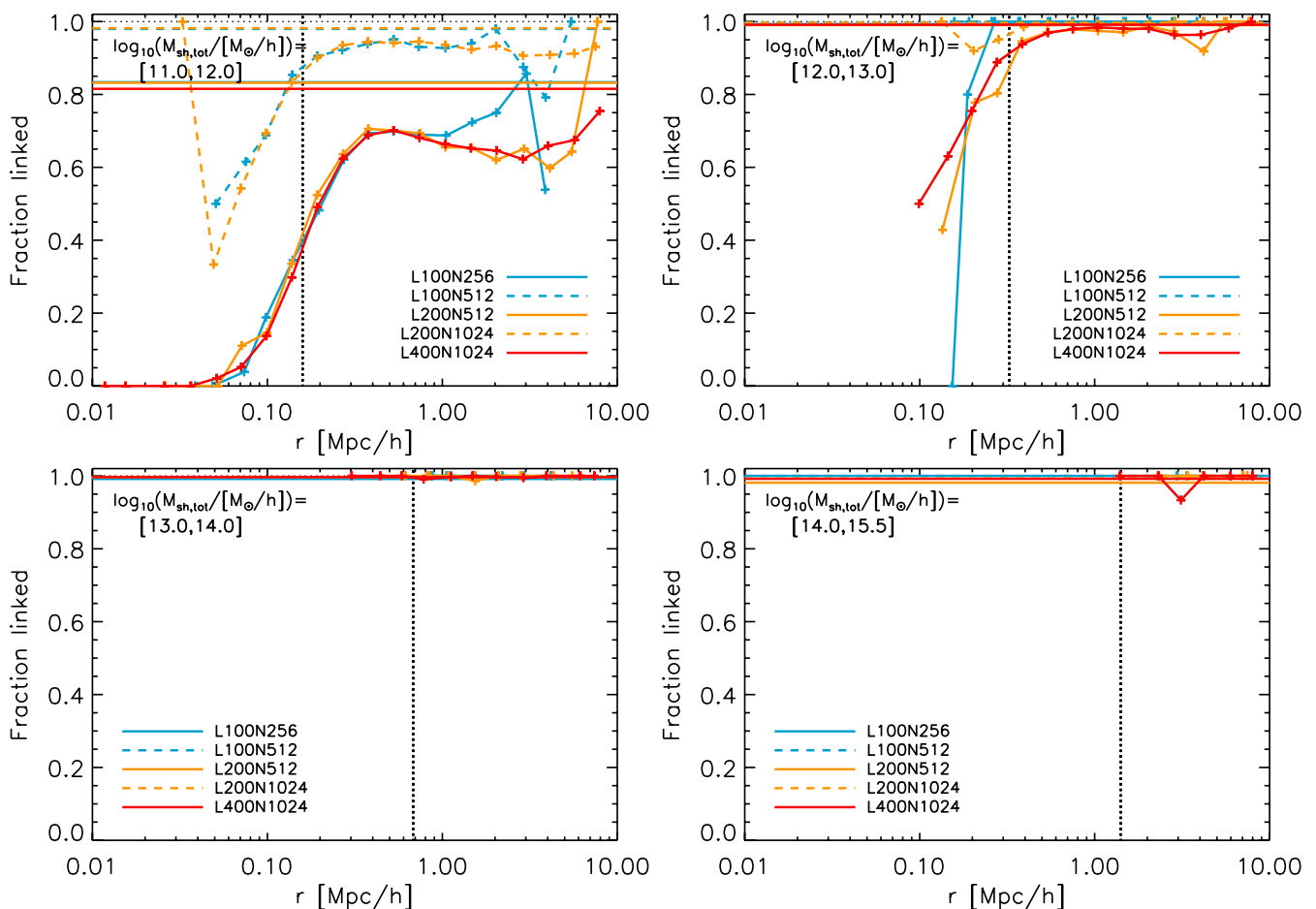


Figure B1. The fraction of subhaloes in *DMONLY* for which a link was found in *REF*, split by subhalo mass as indicated in the top left of each panel. Colours and line styles are as in Fig. A1, and a vertical dotted line once again indicates the mean virial radius in each mass bin. Horizontal lines show the total fraction of linked subhaloes (both centrals and satellites) at the corresponding box size and resolution, while the lines with plot symbols show the fraction of satellite subhaloes linked as a function of radius. For $r \lesssim 2r_{\text{vir}}$ the fraction of linked satellites typically drops sharply as subhaloes are destroyed by tidal stripping or become undetectable. Both the matched satellite and total fractions depend strongly on box size and resolution for subhalo masses $M_{\text{sh,tot}} < 10^{12} h^{-1} M_{\odot}$.

APPENDIX B: LINKED FRACTIONS

Here, we consider the fraction of subhaloes for which a link can be established between *DMONLY* and *REF* as a function of both mass and, in the case of satellites, radius. Both numerical and physical effects play a role here. First, at small radii *SUBFIND* may fail to detect satellite subhaloes even though these have not been fully disrupted yet, due to the high background density of the main halo (e.g. Muldrew et al. 2011). As baryonic subhaloes are typically more concentrated than dark matter only ones, increasing their density contrast, these can be detected down to smaller radii. Secondly, baryonic satellites tend to survive longer than their dark matter only counterparts, as their increased concentration also allows them to better withstand the tidal forces of the main halo (e.g. Macciò et al. 2006). Because of this, our results for linked samples may be biased at radii where a significant fraction of satellite subhaloes is unlinked, as we expect to be better able to detect a pair of identical subhaloes when the baryonic one is located at smaller radii than the dark matter only satellite is located at smaller radii than its baryonic counterpart.

In Fig. B1 we show the fraction of subhaloes in *DMONLY* for which a counterpart is found in *REF*. Once again we do not show a

comparison with *AGN* for clarity, but note that very similar results are obtained.

Horizontal lines show the total fraction of *DMONLY* subhaloes (both centrals and satellites) that is recovered in *REF*, while lines with plot symbols show the fraction of satellites for which a link is found as a function of radius. It is clear that the linked fraction depends heavily on both box size and resolution for $M_{\text{sh,tot}} < 10^{12} h^{-1} M_{\odot}$, although the effect of the box size is only significant for the low-resolution simulations. For the simulation employed in this mass bin throughout the main text of the paper, *L200N1024*, the total fraction of linked subhaloes is around 98 per cent. However, the fraction of linked satellites is significantly lower, especially for radii $r \lesssim 2r_{\text{vir}}$, where the different survival and detection rates of baryonic subhaloes are expected to play a role.

Comparing this panel to the corresponding panel in Fig. A2, we see that the drop in the fraction of matched satellites at small radii corresponds to the strong increase in clustering found for baryonic subhaloes, indicating that this may be a biased result. Similar results are found for satellites with masses $10^{12} < M_{\text{sh,tot}}/[M_{\odot} h^{-1}] < 10^{13}$, although both the total and the satellite linked fractions are much higher than for $10^{12} < M_{\text{sh,tot}}/[M_{\odot} h^{-1}] < 10^{13}$, for all simulations and radii. No drop-off in the linked fraction of satellites is observed at higher masses.

Based on these results, we have chosen to grey out the relative difference curves in the figures showing autocorrelation functions (Figs 3 and 5) on radii where the fraction of linked satellites is < 95 per cent of the total matched fraction. Note that this may not completely remove the possible bias on scales where the satellite contribution dominates the correlation function. Further investigation with higher-resolution simulations is needed to determine whether the upturn observed at small radii is physical or numerical in origin.

Note that the occasional downturn of the linked fraction at relatively large radii, $r \gtrsim 2r_{\text{vir}}$, is due to small-number statistics, as low-mass subhaloes found at these radii are rarely satellites. As the autocorrelation function of linked subhaloes at these radii is dominated by central–central pairs, we do not apply a cut at $r \geq 2r_{\text{vir}}$.

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