Mon. Not. R. Astron. Soc. **000**, 000–000 (0000) Printed 9 December 2015 (MN LATEX style file v2.2)

Halo Definition and Environmental Effects

Antonio Villarreal[1⋆], Andrew R. Zentner[1]†, Christopher W. Purcell[2]‡
[1]*Department of Physics and Astronomy & Pittsburgh Particle Physics, Astrophysics, and Cosmology Center (Pitt-PACC), University of Pittsburgh, Pittsburgh, PA*
[2]*Department of Physics and Astronomy, West Virginia University, Morgantown, WV*

In preparation

ABSTRACT
Recent work has shown the importance of environment to the properties of dark matter halos. This brings conflict to standard implementations of the halo model and excursion set theory which assume that the properties of a population within the halo is determined by the mass of the halo alone. We seek to find a definition of the size of a halo that allows us to minimize the impact of assembly bias on halo model calculations. We analyze the dependence on environment of our properties using the method of marked correlation functions for several different halo definitions, utilizing the Diemer et al simulations. We find that environmental dependencies are dramatically different as we vary the definition of the halo radius in terms of the overdensity Δ. At large length scales, we find that the majority of assembly bias is removed through suitable redefinition of Δ. We are able to determine that the majority of the reduction in assembly bias is related to the elimination of host halos that would cease to be hosts in catalogs at lower values of Δ. Further, we analyze how different mass cuts affect this methodology. We note that unresolved halos leads to assembly bias being missed and that the most massive halos seem to exhibit minimal assembly bias. We further note that the choice of halo definition can induce assembly bias and consider how this may be interpreted in the context of previous results in the literature.

Key words: dark matter – galaxies: halos – galaxies: formation – large-scale structure of universe – methods: numerical

1 INTRODUCTION

[ARZ: We will need to work on the introduction considerably as we get a better handle on the final results. The first two paragraphs can probably be combined into a single shorter paragraph. I also like to end the first paragraph by telling the reader what it is that we aim to do in the paper. The rest of the introduction will need be be completely rewritten in a more professional manner. However, let's get the middle of the paper complete before we work a lot on the introduction and conclusion sections.]

In the current concordance cosmology, the creation of observed galaxies and clusters is often seen as arising from the hierarchical mergers of dark matter halos.(White & Rees 1978). Being able to model the properties of dark matter halos and the galaxies within gives us a potential probe for the physical processes that go into galaxy formation. The excursion-set formalism of galaxy clustering (Bond et al. 1991; Lacey & Cole 1993; Somerville & Kolatt 1999; Zentner 2007) and the standard halo model of galaxy clustering (Seljak 2000; Peacock & Smith 2000; Scoccimarro et al. 2001; Berlind & Weinberg 2002; Bullock et al. 2002; Cooray & Sheth 2002) both can help us in this task, but rely on underlying asumptions. The first is that the statistics of the objects within a dark matter halo is a function of the mass alone. The second is that the clustering of dark matter halos is a function of mass as well. In this paper we propose a simple redefinition of halo size that will help correct for inaccuracies in these two assumptions.

It has previously been demonstrated that the clustering of halos is dependent on not only the halo mass, but also on the formation time of the halo (Sheth & Tormen 2004; Gao et al. 2005; Croton et al. 2007). Further-

⋆ E-mail: asv13@pitt.edu
† E-mail: zentner@pitt.edu
‡ E-mail: cwpurcell@mail.wvu.edu

more, it has been shown that the clustering of a given halo is dependent on the halo concentration (Wechsler et al. 2006). This necessitates corrections be made to the standard implementations to account for this. More complicated methodologies have been made to extend to using merger histories directly from simulation (Dvorkin & Rephaeli 2011), as well as to account for the dependence due to concentration (Gil-Marín et al. 2011). The relationship that clustering has to the properties of the halo is commonly referred to assembly bias or environmental effects.

Our method of halo redefinition is motivated by the size of a halo not necessarily being a well-motivated property. What is often referred to as the "virial radius" will not contain all gravitational bound dark matter particles in the halo (Kazantzidis et al. 2006). Rather, it is a matter of convention that does not have a common definition across the literature. For example, some studies have chosen to use halo radius defined with respect to the critical mass density of the universe, while many simulation papers choose to use the mean background mass density of the simulation. Furthermore, the overdensity often used can typically vary from 178 (from basic spherical collapse) to 200 (a common simulation definition) to 337 ("virial" in ΛCDM). This can lead to a factor of a few difference in the halo radius just as a matter of definition. We choose to define a halo radius in a way that encompasses the nearby environmental effects. These may be due to large scale structure or driven by baryonic physics. Using a simulated box, we can then test how the redefinition of the halo size affects the relationship between the clustering and the properties of the halo. In the case that the properties of the halo become independent of the clustering, it is possible to utilize standard implementations of the halo model without necessitating more complicated theory.

In § 2 of this paper, we discuss the simulated cosmological boxes that we utilize for our statistics and the halo finder utilized. In § 3, we consider the properties of interest within our halo simulation and their standard definitions. In § 4, we discuss the statistics that we have used in order to test for environmental effects and the removal of known mass scaling from these statistics. In § 5, we present our results on our tests and consider how each definition affects assembly bias. In § 6, we discuss the significance of reducing environmental effects through a redefinition of halo environments and discuss possible applications of our methodology. We also consider the nature of assembly bias as a function of halo definition.

2 SIMULATION DATA AND HALO FINDING

In order to study the effects of halo redefinition, we make extensive use of three simulation boxes. The Diemer et al simulations each utilize a Planck best fit cosmology with $\Omega_{\rm M} = 0.32$, $\Omega_\Lambda = 0.68$, and $h_\circ = 0.67$. We use three simulation boxes with comoving sizes of 125, 250, and 500 h^{-1}Mpc respectively. The particle masses are 1.61555×10^8, 1.29244×10^9, and 1.03395×10^{10} $h^{-1}{\rm M}_\odot$ respectively with a total of 1024^3 particles in each simulation. We will refer to each simulation as L0125, L0250, or L0500 for the remainder of the paper. This set of simulations allows us to probe the resolution effects on halo finding, as each box has a factor of eight difference in effective resolution from the next box. In particular, L0125 has the least massive fully resolved halos as a result of the smaller particle mass, while L0500 has the most robust statistics for the most massive halos as a result of the larger simulation volume.

For halo finding we use the ROCKSTAR halo finder, which works on the phase space algorithm described in Behroozi, Wechsler & Wu (2013). In short, ROCKSTAR determines the initial groupings using a Friends-of-Friends algorithm in phase space, before then applying the spherical overdensity case in order to determine the properties of interest. Unbound particles are removed prior to the calculation of halo mass and other properties of interest. In addition, we take interest of the shape of the halo, which is determined through the sorted eigenvalues of the inertia tensor, with principal ellipsoid axes defined such that $a > b > c$. Some of these parameters are determined after fitting the particles to the Navarro-Frenk-White profile (Navarro, Frenk & White 1997).

[ARZ: Try rewriting this paragraph. There are many problems here. First, halo size is not hard to define, it is just that the definitions in common use are not particularly well motivated. Second, your statement of how a halo is defined is incorrect. There is no assumption made about the distribution of mass in the halo. The definition is with respect to the mean density within a given radius.] While the definition of halo size is fairly straightforward, the individual choices that make up the calculation are not particularly well-motivated. While the virial radius of a halo may be one natural choice, simulations have shown that it often misses gravitationally bound particles (Kazantzidis, Zentner & Kravtsov 2006). Halo size has become a matter of definition, with $\Delta = 200$ being the most common choice within the literature. However, it is not uncommon to see calculations carried out utilizing $\Delta = 178$, as derived through spherical collapse models, or $\Delta = 337$, as calculated in the case of ΛCDM. Given the lack of a well-motivated standard, we choose to redefine the halo in terms of the overdensity parameter, Δ, by defining the halo radius, r_Δ, as follows:

$$\bar{\rho}(r_\Delta) = \Delta \rho_{\rm m} \tag{1}$$

Here, $\rho_{\rm m}$ is the mean background mass density of the simulation box, while $\bar{\rho}$ is the mean density within r_Δ. It should be noted that the use of the critical density or the mean background mass density is interchanged within the literature. The difference between these two values can be as high as a factor of a few, which will then carry forward into the calculation. We allow the overdensity parameter Δ to vary over the range of the fiducial definition of $\Delta = 200$ down to values as low as $\Delta = 10$. As discussed above, this change is motivated due to the fact that the definition of halo size is primarily a matter of convention.

3 HALO PROPERTIES

[ARZ: This section needs to be rewritten to include more details and a more professional style. I will try to outline this section as it should be written here.]

We explore the clustering of halos as a function of a number of halo properties that are commonly explored in the literature and can be easily measured for a halo from an individual simulation snapshot. The first of these is the halo concentration parameter,

$$c_{\rm NFW} = \frac{r_\Delta}{r_{\rm s}}, \tag{2}$$

where r_Δ is the radius of the halo given an overdensity parameter Δ used to define the halo and $r_{\rm s}$ is the halo scale radius. We also use an additional measure of the concentration of the halo density profile,

$$c_{\rm V} = \frac{V_{\rm max}}{V_\Delta}, \tag{3}$$

where $V_{\rm max}$ is the maximum circular velocity achieved within the halo and V_Δ is the circular velocity at the halo radius, r_Δ. The quantity $c_{\rm V}$ can be measured from simulations without any need for fitting halo density profiles to determine scale radii and is therefore robust to choices of fitting methods.

Halo concentrations are useful to explore for these purposes for a number of reasons. First of all, environment dependence of concentrations is of direct interest in modeling gravitational lensing statistics. Secondly, concentrations can be measured from individual halo snapshots relatively easily yet halo concentrations are known to be strongly correlated with the formation histories of dark matter halos with earlier forming halos having higher concentrations at fixed halo mass **[ARZ: Cite papers here like Wechsler et al. 2002]**. As such, exploring concentrations dependence of halo clustering can help quantify some of the halo age dependence of clustering without the need for constructing merger trees. This is particularly important in the present study in which the halo finding is performed repeatedly and constructing a merger tree for each run of the halo finder with different Δ can be prohibitive. We will explore measures of halo age directly in a forthcoming follow-up study dedicated to halo formation histories.

Figure 1 shows the mean $c_{\rm NFW}$-M_Δ relation for halos defined with $\Delta = 200$ in L0125, L0250, and L0500. For each simulation, we consider halos only above a minimum mass, as shown in Fig. 1, to ensure that concentration measurements are not compromised by resolution. Likewise, Figure 2 shows the analogous relation for the alternative definition of halo concentration, $c_{\rm V}$.

Figure 3 shows the relationship between $c_{\rm NFW}$ and $c_{\rm V}$ for quantifying halo concentration on a halo-by-halo basis. Generally the two proxies for concentration are related to each other in a simple manner with relatively small scatter indicating that these two quantities largely contain the same information about each halo.

In addition to halo concentrations, we explore halo clustering as a function of a variety of other halo properties. We explore halo clustering as a function of halo shape quantified by the ratio of the minor and major axes length,

$$s = \frac{c}{a}, \tag{4}$$

where a is the major axis length and c is the minor axis length. The mean relations of halo shapes as a function of halo mass for $\Delta = 200$ is shown in Figure 4 along with the mass thresholds selected to ensure that our results are not compromised by resolution.

We also explore halo clustering as a function of halo spin quantified by the spin parameter λ as introduced by

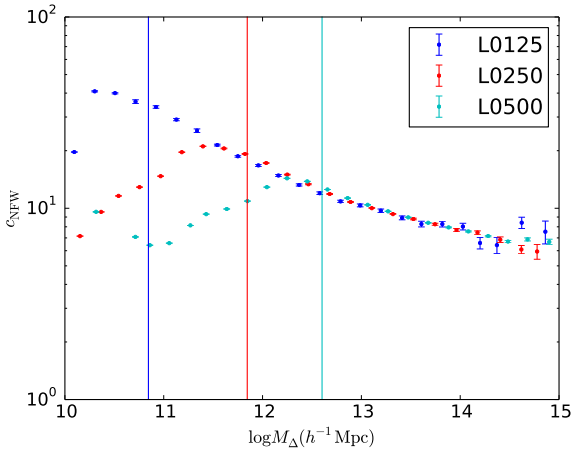

Figure 1. An example of the relationship between the NFW concentration and halo mass for each of our simulations for $\Delta = 200$. The chosen lower limit on halo mass for our sample is marked as a blue, red, or cyan line for L0125, L0250, and L0500 respectively. At lower mass, halos are likely ill defined due to resolution limits.

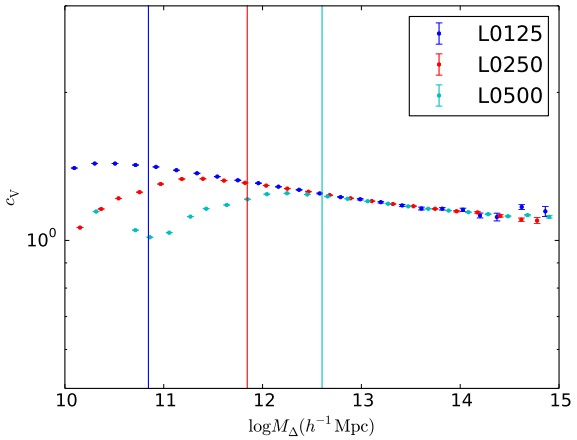

Figure 2. An example of the relationship between the velocity ratio concentration and halo mass for each of our simulations for $\Delta = 200$. The chosen lower limit on halo mass for our sample is marked as a blue, red, or cyan line for L0125, L0250, and L0500 respectively. At lower mass, halos are likely ill defined due to resolution limits.

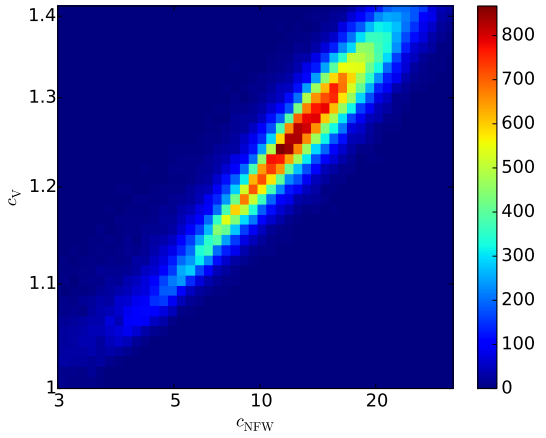

Figure 3. The relationship between the two different marks of concentration, using halos in L0250.

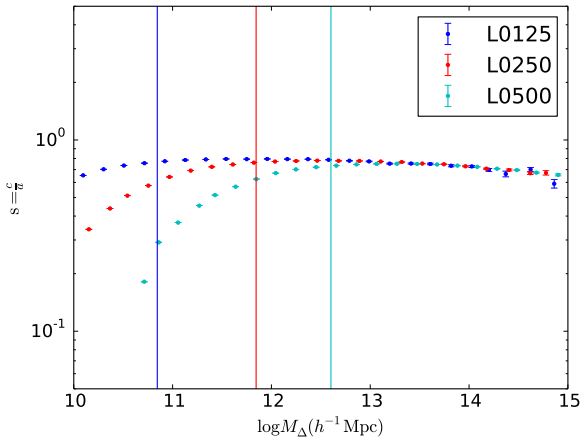

Figure 4. An example of the relationship between halo shape and halo mass for each of our simulations for $\Delta = 200$. The chosen lower limit on halo mass for our sample is marked as a blue, red, or cyan line for L0125, L0250, and L0500 respectively. At lower mass, halos are likely ill defined due to resolution limits.

(Peebles 1969),

$$\lambda = \frac{J\sqrt{|E|}}{GM_\Delta^{2.5}} \tag{5}$$

where J is the halo angular momentum, E is the total energy of the halo, and M_Δ is the mass at the halo radius, r_Δ. The mean relations of halo spins as a function of halo mas for $\Delta = 200$ is shown in Figure 5 along with the mass thresholds selected to ensure that our results are not compromised by resolution.

In practice, the mean relations between the various halo properties and the mass thresholds must be determined for each combination of simulation, halo property (e.g., $c_{\rm NFW}$ or s), and halo definition (e.g., value of Δ). We use thresholds determined by the particular case under consideration so as to ensure that resolution limitations do not drive any of our primary results. A higher

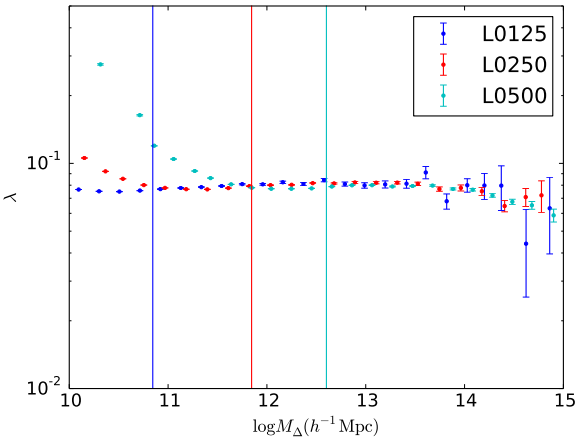

Figure 5. An example of the relationship between halo spin λ and halo mass for each of our simulations for $\Delta = 200$. The chosen lower limit on halo mass for our sample is marked as a blue, red, or cyan line for L0125, L0250, and L0500 respectively. At lower mass, halos are likely ill defined due to resolution limits.

mass threshold necessarily means that we do not have any issues due to halo resolution, but reduces the statistics. In addition, it should be noted that halo assembly bias is weakest in the most massive halos, which may reduce the potential signal at high masses. As an example, we summarize some of the mass thresholds we have used for a set of Δ values in Table 1.

4 METHODOLOGY

There are several well understood effects in our simulation that must be accounted for before we can draw any conclusions. The first is the matter of simulation resolution. In each of our generated halo catalogs, we can anticipate finding artificial halos. While existing under our constraint of having only a set overdensity, they will have properties that conflict with well known trends that have been shown in prior works, such as concentration decreasing as a function of halo mass (Wechsler et al. 2006). As shown in Figure 1, we can roughly identify the region in which these artificial halos become significant by determining where the monotonic relationship between concentration and mass breaks at low mass. In order to avoid our halo finding statistics being thrown off by these artificial halos, we set a lower mass threshold, limiting the sample to only those halos that are physically significant. We have chosen our threshold to be set by L0250 in order to minimize the shift to the remaining catalogs.

Another well known effect is the scaling of our properties as a function of halo mass, as well demonstrated within the literature (Duffy et al. 2008). We are interested in the clustering behavior beyond this well known effect, so we seek to remove this scaling. We take all host halos of interest and sort them by their halo mass. Each set of halo properties are then placed in bins of equal population in order to assure that enough data points exist for

Table 1. Minimum mass thresholds for each of our analyses depending upon the value of the overdensity, Δ, used to define the halos. In the columns below the values of Δ we show the minimum host halo masses considered in units of $h^{-1}\mathrm{M}_{\odot}$.

Cutoff Name	$\Delta = 200$	$\Delta = 100$	$\Delta = 75$	$\Delta = 50$
low mass	7×10^{10}	8×10^{10}	9×10^{10}	1×10^{11}
mid mass	7×10^{11}	8×10^{11}	9×10^{11}	1.5×10^{12}
high mass	4×10^{12}	5×10^{12}	6×10^{12}	7×10^{12}

robust statistics. We then normalize each mark in a given bin to the mean value. The end result is halo properties that do not change as a function of mass, allowing us to more accurately analyze clustering behavior within our simulation.

To normalize the satellite number we follow the prescription of Wechsler et al. (2006). In addition to the mass cutoffs on our data, we eliminate ill-resolved satellites by choosing a cutoff in $v_{\rm max}$ for the host halo. We then attempt to match subhalos to this mass halo, making a secondary cutoff in the case that the value of $v_{\rm max,sub}/v_{\rm max,host}$ is not above a threshold value. We set these two parameters such that there are no isolated or poorly resolved subhalos in our data sample.

In order to test for environmental effects, we choose to utilize two primary methods. The first is to look at the two-point correlation functions of the host halo catalogs. We take the difference between the samples of the top 20% and bottom 20% in normalized concentration, and then normalize these results by the overall correlation function.. Assuming that each halo was statistically similar, regardless of host halo clustering, we would expect that this statistic would approach zero at all scales.

The second method is to use a weighted correlation function, often referred to in the literature as the marked correlation function (MCF). Our marks in this case are the log values of the halo concentration proxies described previously, the halo shape, and the halo satellite number. The use of this tool in measuring clustering has been shown previously in Wechsler et al. (2006) or Harker et al. (2006). We follow the methodology of Wechsler et al. (2006) in defining the MCF as

$$\mathcal{M}_m(r) \equiv (\langle m_1 m_2 \rangle_p(r) - \langle m \rangle^2 / \mathcal{V}(m), \qquad (6)$$

where m is the value of the mark for a given halo, $\langle m \rangle$ the mean and $\mathcal{V}(m)$ the variance of the mark. We then compare our MCFs to an error bar representing the approximate 2-sigma confidence region generated by randomizing the marks among our halos 400 times and taking the 10th lowest and 390th highest values of the mark among the randomizations. In the case of no environmental effects upon the property of interest, one would expect that the MCF would fall within the region bounded by these limits.

5 RESULTS AND DISCUSSION

Ultimately, as a result of the limited nature of simulation data, we must make some lower mass cutoff in our study of halo properties. Table 1 describes where we have chosen to place these cuts in a variety of different measures. The "mid mass" set of cuts has been chosen for our preliminary analysis choice, as it gives us a healthy balance between mass resolution and robust statistics. Figure 1 shows an example of this cutoff in the $\Delta = 200$ case as a red line. We do not include L0500 in this initial analysis, as that simulation holds several bins of halos deviating significantly from a monotonic mass-concentration relation. This may be a result of including halos that are unphysical, but still meets the halo-finding algorithm's requirements. The remaining two cuts have been chosen to explore any resolution related effects. The "low mass" cuts explore the lowest possible mass cutoff that excludes ill resolved halos in the Diemer et. all simulations, while the "high mass" set of cuts represent the mass cutoff chosen to avoid contamination due to resolution effects in our largest simulation box.

The left column of Figure 6 demonstrates the results of our method as we change the value of Δ. It can be seen that L0125 and L0250 both minimize the difference between the two correlation functions at large scales between $\Delta = 75$ and $\Delta = 50$. It should be noted that there is effectively a "sweet spot" effect, where decreasing the value of Δ (or correspondingly, the size of an average halo) will start to give increasing amounts of power in the correlation function to the least concentrated halos - essentially introducing a new assembly bias. The physical nature of this is of interest to us, and we attempt to present an intuitive sense of the effect further on.

The NFW defined concentration MCF is shown in the left column of Figure 7. It can been that as we decrease the value of the overdensity parameter Δ to lower values down to $\Delta = 50$. It can be seen that at scales of $r > 10\ h^{-1}$Mpc , environmental effects are removed between $\Delta = 75$ and $\Delta = 50$ in the Diemer et al. boxes. This is repeated for the velocity ratio defined concentration in the left column of Figure 8. It again demonstrates that the same values of Δ are preferred for the removal of assembly bias as with the NFW concentration. This seems to indicate that our methodology should be effective at minimizing the impact of assembly bias at large scales.

The left column of Figure 9 demonstrates a case in which the halo property does not have environmental dependence removed under our simple method of halo redefinition. In particular, with decreasing values of Δ, we only serve to increase the strength of this environmental dependence. This result seems to indicate a process occurring to drive the shape of the halo is on scales that are significantly larger than the size of the halo, such that our method will not be able to encompass these effects. A similar result is located in both the left columns of Figure 10 and Figure 11, in which the spin parameter mark

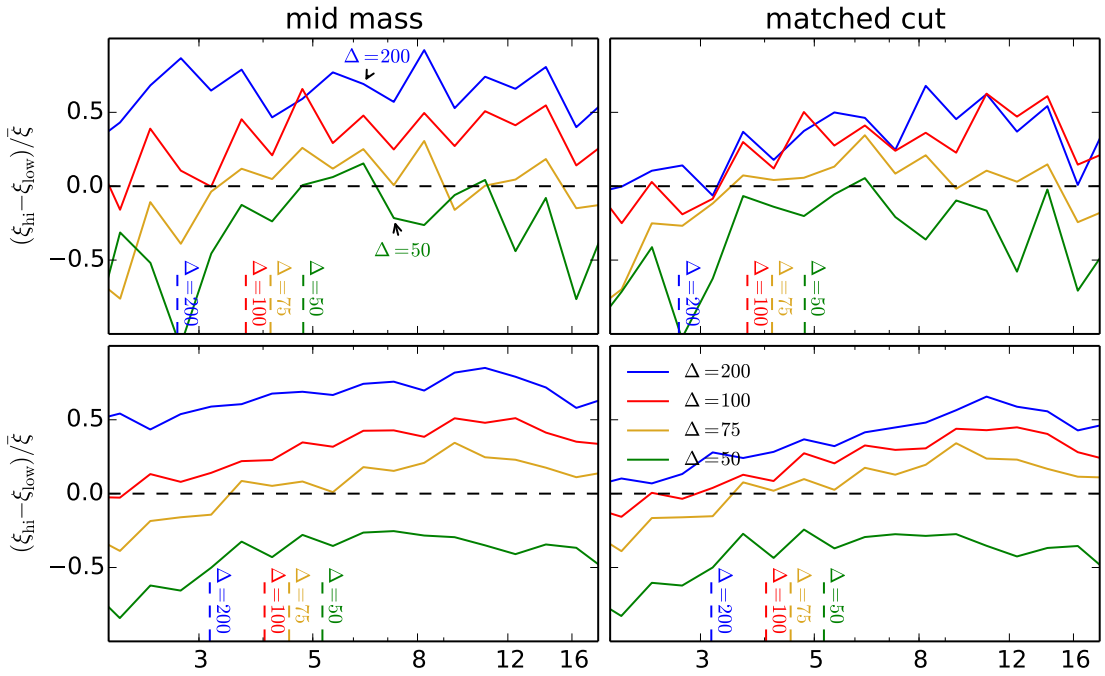

Figure 6. The difference of the correlation function for only the top 20% most concentrated halos and the bottom 20% in concentration, normalized by the overall correlation function of the entire sample. From top to bottom, we show the results for L0125 and L0250 respectively. The left column utilizes all host halos in the mass cut. The right column contains only host halos that exist inside of our best fit value of Δ, as explained in the text. The dashed lines along the bottom denote the largest halo radius for a given value of the overdensity parameter.

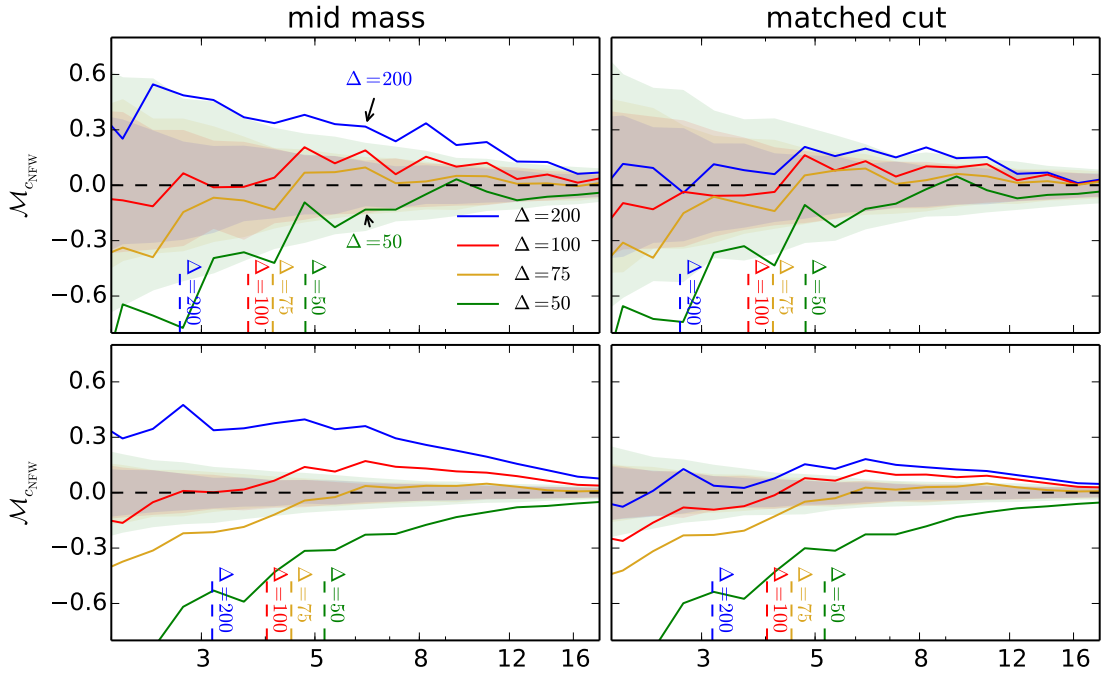

Figure 7. The marked correlation function for the concentration defined according to the NFW profile. From top to bottom, we show the results for L0125 and L0250 respectively. The left column utilizes all host halos in the mass cut. The right column contains only host halos that exist inside of our best fit value of Δ, as explained in the text. The shaded bands represent 2-sigma confidence regions generated by randomization of the marks. The dashed lines along the bottom denote the largest halo radius for a given value of the overdensity parameter.

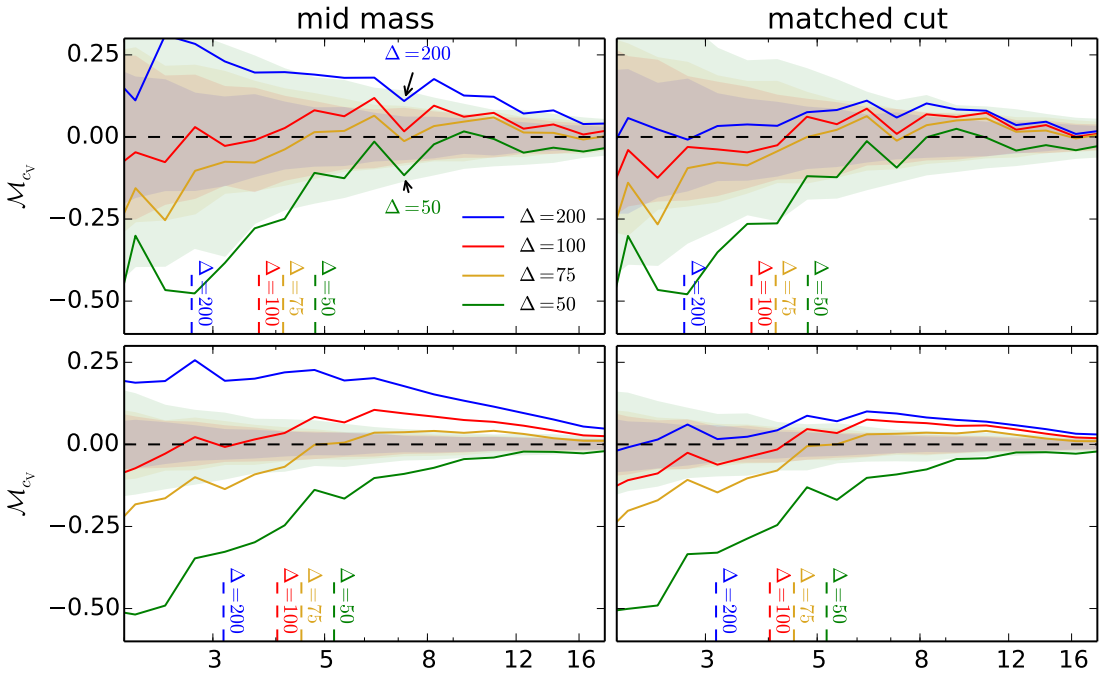

Figure 8. The marked correlation function for the concentration defined according to the velocity ratio. From top to bottom, we show the results for L0125 and L0250 respectively. The left column utilizes all host halos in the mass cut. The right column contains only host halos that exist inside of our best fit value of Δ, as explained in the text. The shaded bands represent 2-sigma confidence regions generated by randomization of the marks. The dashed lines along the bottom denote the largest halo radius for a given value of the overdensity parameter.

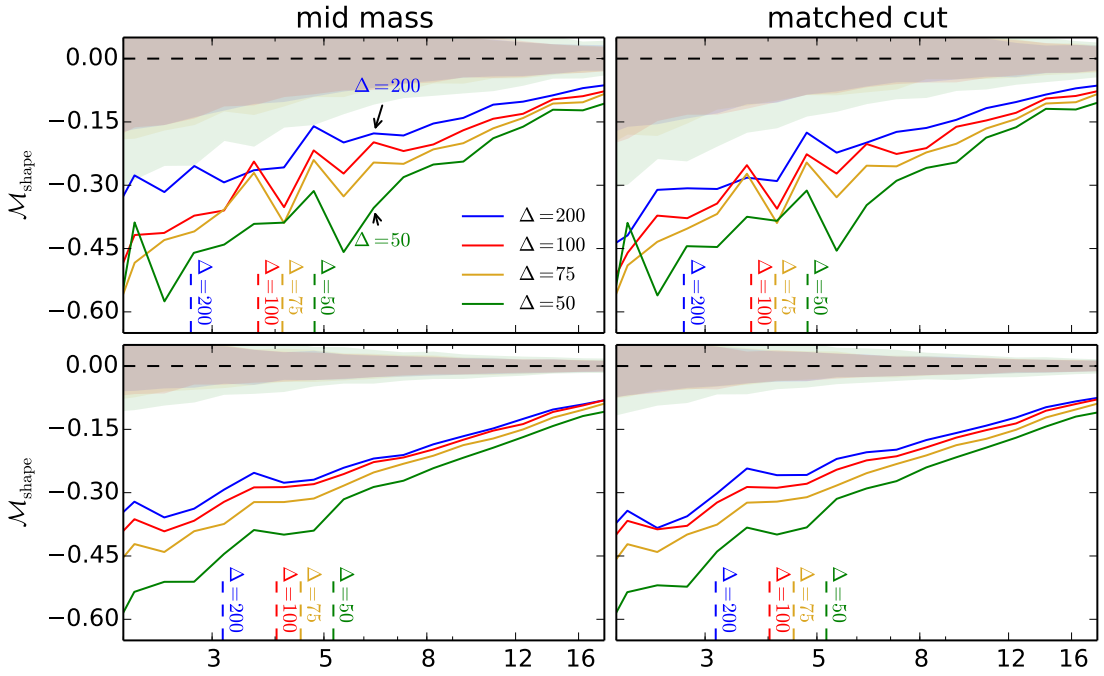

Figure 9. The marked correlation function for the shape of the halo. From top to bottom, we show the results for L0125 and L0250 respectively. The left column utilizes all host halos in the mass cut. The right column contains only host halos that exist inside of our best fit value of Δ, as explained in the text. The shaded bands represent 2-sigma confidence regions generated by randomization of the marks. The dashed lines along the bottom denote the largest halo radius for a given value of the overdensity parameter.

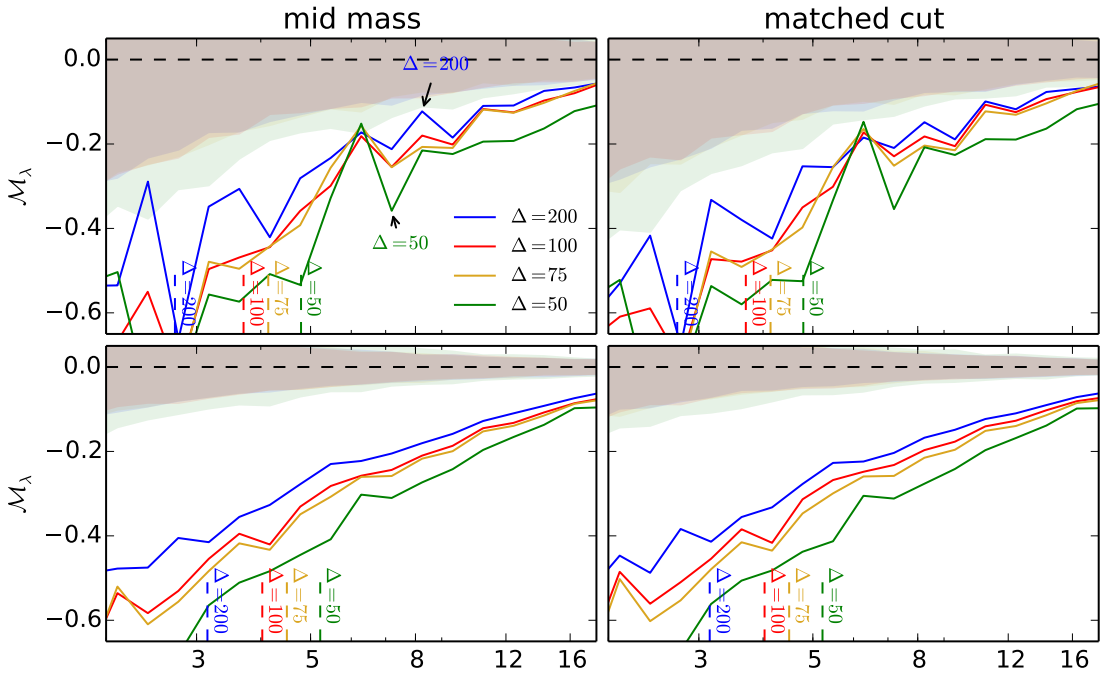

Figure 10. The marked correlation function for the spin parameter of the halo. From top to bottom, we show the results for L0125 and L0250 respectively. The left column utilizes all host halos in the mass cut. The right column contains only host halos that exist inside of our best fit value of Δ, as explained in the text. The shaded bands represent 2-sigma confidence regions generated by randomization of the marks. The dashed lines along the bottom denote the largest halo radius for a given value of the overdensity parameter.

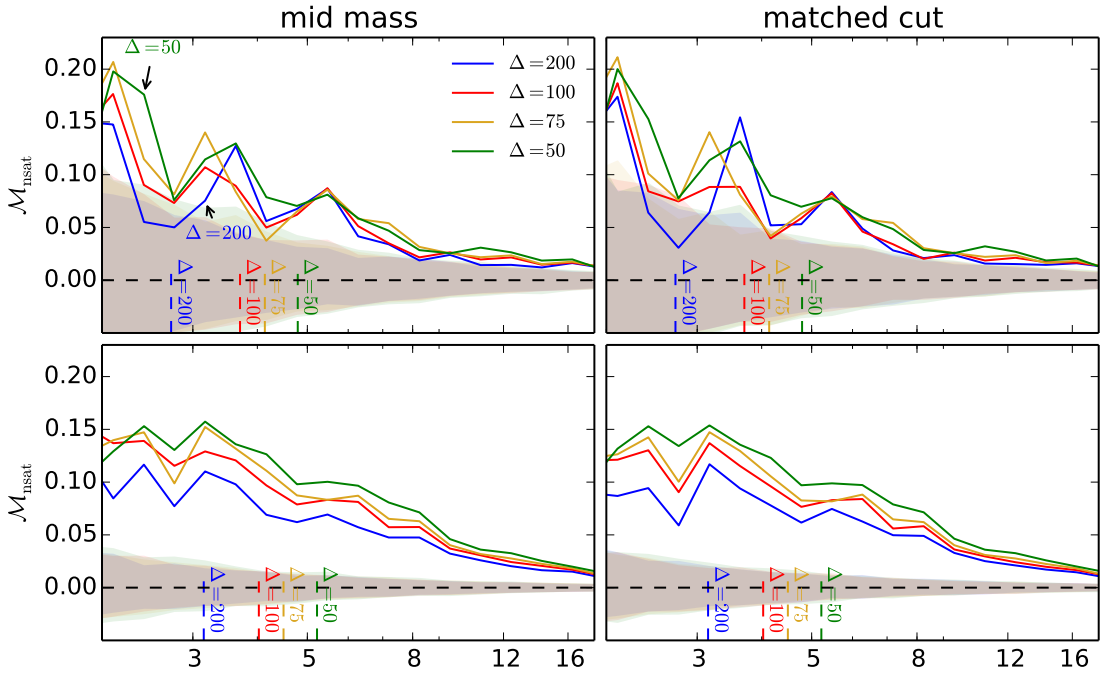

Figure 11. The marked correlation function for the number of satellites for a host halo. From top to bottom, we show the results for L0125 and L0250 respectively. The left column utilizes all host halos in the catalog. The right column contains only host halos that exist inside of our best fit value of Δ, as explained in the text. The shaded bands represent 2-sigma confidence regions generated by randomization of the marks. The dashed lines along the bottom denote the largest halo radius for a given value of the overdensity parameter.

and satellite number mark experience the exact same effect.

Overall, while some statistics of the halo can be made to become independent of environmental effects, there exist several that cannot. These environmental effects are not negligible when considering applications of the halo model. In contrast, if one is only interested in the concentration of the halo, it is feasible to take advantage of our methodology in order to remove environmental effects. In contrast, it is not possible to apply the halo model directly to statistics such as the number of satellite halos or the shape without accounting for these environmental effects. We do not yet have a full understanding of what leads to the visible effects that we see on some of these marks, but we present a series of cases in which an intuitive understanding may be gained as to the root causes.

We would also like to study our sample for some insight as to how this process of correction is actually functioning. For example, we would like to know if it is a more careful selection of halos in our simulation that results in the removal of this apparent assembly bias or if it is the increased noise in the statistics as a result of reducing the number of halos. There is also the possibility of halo statistics changing as a result of the new halo definitions to be considered. One can imagine that the addition of a large amount of mass toward the edge of the halo would reasonably change measures of concentration, shape, and spin. To test some of these possibilities, we run an algorithm to match halos across our multiple catalogs. We identify halos that are within a distance tolerance that we scale such that there are no identifications of multiple halos to the same object in another catalog. Catalogs are all matched against a single target catalog. This target catalog is chosen to fit the value of Δ that we find best removes assembly bias in large scales for the concentration marks chosen. For the Diemer et al. simulations, we find this to be a value of approximately $\Delta = 70$. Once the catalogs have been matched against this primary catalog, we rerun our marked correlation function calculations using only host halos identified across both catalogs. This tests if the exclusion of halos that are subsumed in the lower Δ catalog is the primary cause of our results as well as those halos that are commonly referred to as backsplash halos.

We can see the results of this test in Figure **??** through Figure **??**. Comparing in each case, the $\Delta = 200$ result when utilizing a catalog matched to our best fit Δ demonstrates less assembly bias than the full catalog for the mass cut would demonstrate for our concentration marks. While the magnitude of this result does vary with scale, upwards of 80% of assembly bias can be accounted for as a result of this selection effect. Furthermore, most of the removed assembly bias occurs on scales that are comparable to the increase in the halo radius between the samples. This seems to indicate that there is significant contamination of host halo statistics by backsplash halos and other halos that would be subsumed into a common host given an increase in halo size.

However, it is worth noting that statistics such as spin parameter, shape, and satellite numbers do not benefit from this selection effect. Assembly bias in these statistics remains nearly identical, suggesting that the selection effects that preferentially account for backsplash halos do not impact these marks. This does serve as a consistency check: given that our methodology does not seem to have a positive impact on the behavior of these marks, the fact that this selection criteria would have positive impact on the data would be an oddity. Instead, we have another confirmation as to the intrinsic ties between these halo statistics and environment.

Given that the removal of assembly bias is not being driven by introducing noice, we now will attempt to determine why the different marks exhibit different behaviors. The first interesting feature is how well our two separate definitions of concentration interact with each other. In the case that the halo can be described well by an NFW profile, one expects a direct relationship between the NFW defined halo concentration and the velocity ratio defined halo concentration. While some variation can be expected due to halos not perfectly being fit by an NFW profile, we do see that the features in one concentration proxy are mirrored in the other. This allows for the two concentration markers to support each other well with regards to our ability to remove environmental effects on large scales.

Halo shape and satellite number are statistics that do not end up having their environmental effects removed and can even be made more prominent by our methodology. One intuitive way to consider the former statistic is in the context of the cosmic web. Studies have shown a statistically significant alignment between filaments and satellite galaxy position (Tempel et al. 2015; Velliscig et al. 2015). Our method then expands the halo radius and subsumes material that was previously outside of the halo. A simple graphic illustrates this potential effect in Figure 12. As there is a preferential distribution of these satellites that are being subsumed into the halo, this would serve to induce a shape to the halo that would then be determined by Rockstar. In addition, as our satellite number is chosen by those halos within the halo radius, we anticipate that the most clustered regions would see the largest increase in satellite count and thus see an increase in the satellite number mark.

The "sweet spot" behavior of this method is also of interest to us. The halo redefinition process that we use serves to decrease halo clustering for the most concentrated halos and increase halo clustering for the least concentrated halos. In the case of the high concentration cut, the reduced clustering can be seen as a result of halo exclusion. As we exchange collections of tightly clustered smaller halos for a single larger halo, then the most prominent contribution of the two-point correlation function will be reduced at all scales. Furthermore, it suggests that the assembly bias that we witness is correlated with the choice of halo definition. This effect may be necessary to explaining the mixed results in the field at determining just how strong assembly bias is.

Given the differences between the simulation results and with the data available to us with the Diemer et al. simulations, we then explore the effect of the choice in mass cut. The low mass cut effects are shown in the Figure 13 through Figure 18. This cut explores the area in which we are including ill-resolved halos potentially into

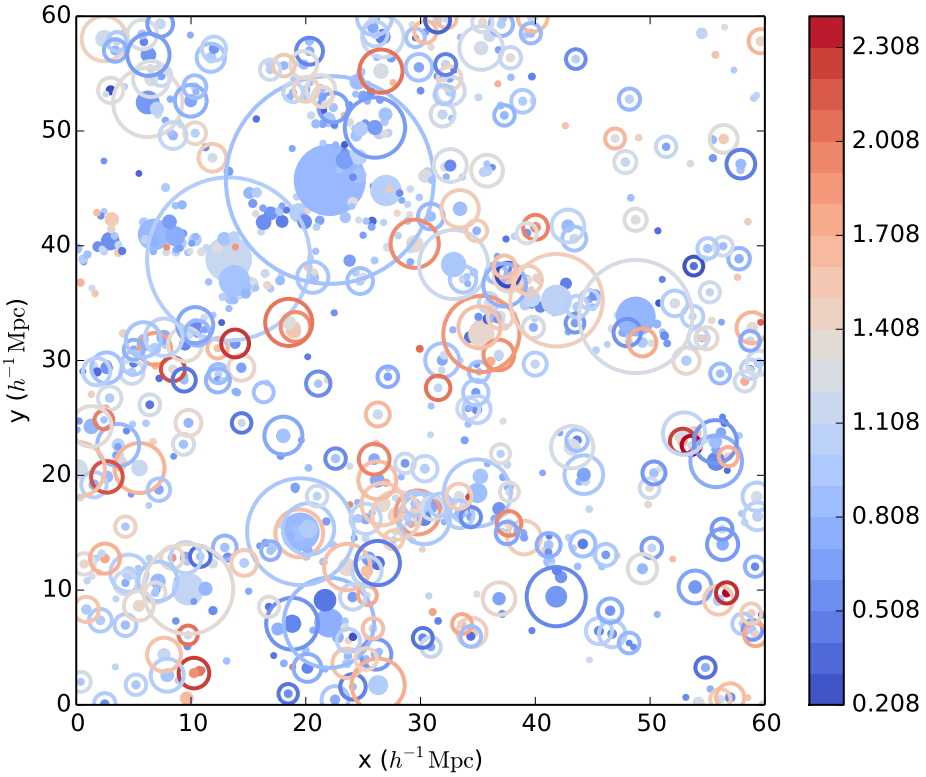

Figure 12. A 20 h^{-1}Mpc deep cut of L0250 along the z-axis. This zoom-in demonstrates the process that decreases the shape parameter as a function of clustering. The size of each circle represents the projection of a spherical dark matter halo with a given halo radius onto the x-y plane. Filled circles use the $\Delta = 200$ catalog and unfilled circles use the $\Delta = 10$ catalog in order to make the effect more visible. Color scale refers to the log shape mark, normalized by halos of the same mass.

the simulation. Namely, utilizing this mass cut on L0250 data includes halos that do not fit the form of the expected monotonic halo mass-concentration relation. We can determine several facts from this exercise in resolution testing. The first is that the general behavior of the marks seems to be the same regardless of the influence of the resolution effects. Decreasing the value of Δ still moves our marked correlation functions in the same directions as the previous mass cut, although often from a very different starting location. Even with potentially ill resolved objects, the method can be brought to bear upon the problem - potentially a concern if the simulation's ability to resolve a particular statistic is questionable. Also noteable is that the level of the assembly bias is reduced in the case of this lower mass cut in L0250 when compared to L0125. This is distinctly different from our previous example, in which both L0125 and L0250 both contain only well resolved halos within the mass cut and have nearly identical assembly bias at every scale, aside from simulation noise. This seems to imply that including unphysical halos in our sample makes it difficult to determine the actual assembly bias at work.

We can repeat this exercise by looking at a different set of mass cuts. In this case we will utilize our highest mass cut on L0250and L0500. This case does not include poorly resolved halos as we were including in the previous example. Instead, the smaller sized box runs the risk of having less of the most massive halos, resulting in having a larger variance for objects at high mass. The results of this are shown in Figure 19 through Figure 24. There are several key observations to be taken away from this series of plots. The first is the fact that like in our previous example, we see no significant change in assembly bias when only including well resolved halos and see a considerable change when we include unresolved halos in L0500. Of greater import is the fact that we see that in several cases, when interest is only in the most massive halos we find that there is next to no assembly bias. Changing our definition of the halo radius in these cases can even induce halo assembly bias. Given the previous discussion regarding the differences in halo radius definitions chosen (e.g. critical density versus mean density), this implies that we can have very different measurements of this effect solely based on halo definition - a fact that has yet to be explored thoroughly within the literature.

6 CONCLUSIONS

We have looked at how to use CFs and MCFs in order to analyze the environmental effects upon the properties of the halo. We have suggested a method of removing the

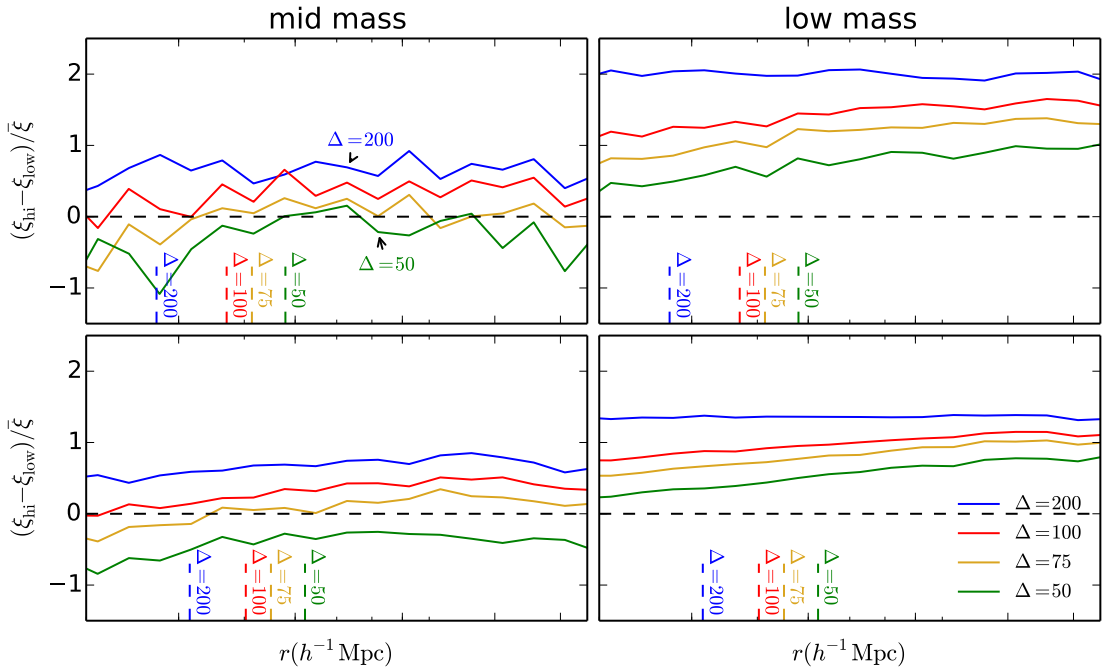

Figure 13. The difference of the correlation function for only the top 20% most concentrated halos and the bottom 20% in concentration, normalized by the overall correlation function of the entire sample. The top row uses L0125 data while the bottom row uses L0250 data. The left column utilizes the "mid mass" cutoff, while the right column demonstrates the "low mass" cutoff. The dashed lines along the bottom denote the largest halo radius for a given value of the overdensity parameter

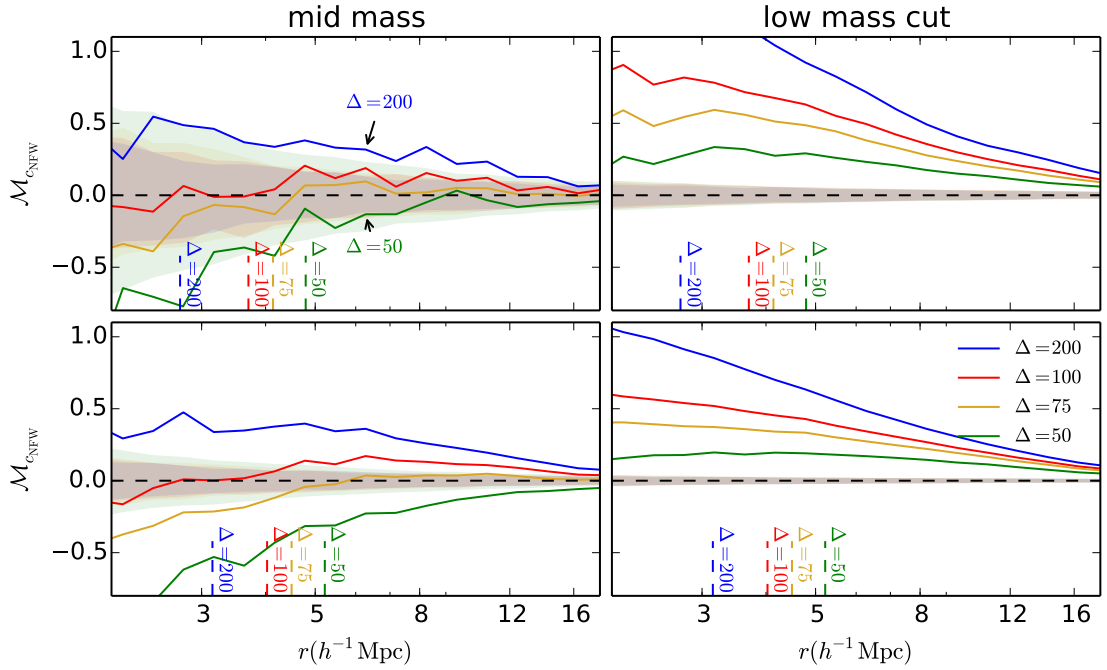

Figure 14. Comparison of the marked correlation function for the concentration defined according to the NFW profile between the "mid mass" cutoff (left column) and the "low mass" cutoff (right column). The top row uses L0125 data while the bottom row uses L0250 data. The shaded bands represent 2-sigma confidence regions generated by randomization of the marks. The dashed lines along the bottom denote the largest halo radius for a given value of the overdensity parameter.

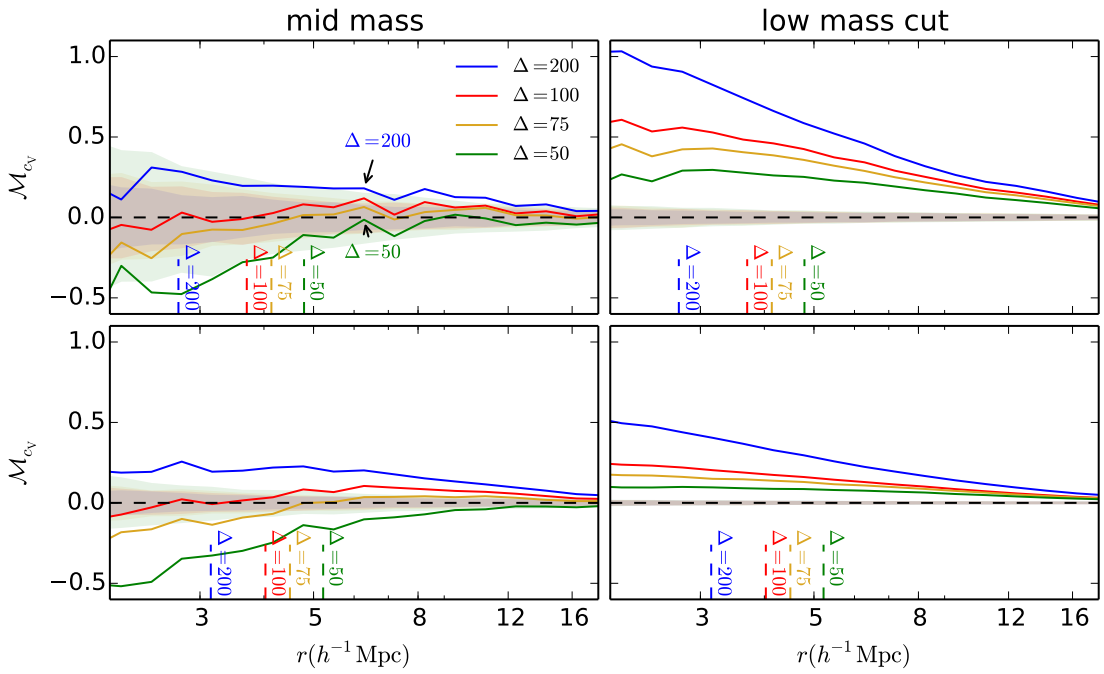

Figure 15. Comparison of the marked correlation function for the concentration defined according to the velocity ratio between the "mid mass" cutoff (left column) and the "low mass" cutoff (right column). The top row uses L0125 data while the bottom row uses L0250 data. The shaded bands represent 2-sigma confidence regions generated by randomization of the marks. The dashed lines along the bottom denote the largest halo radius for a given value of the overdensity parameter.

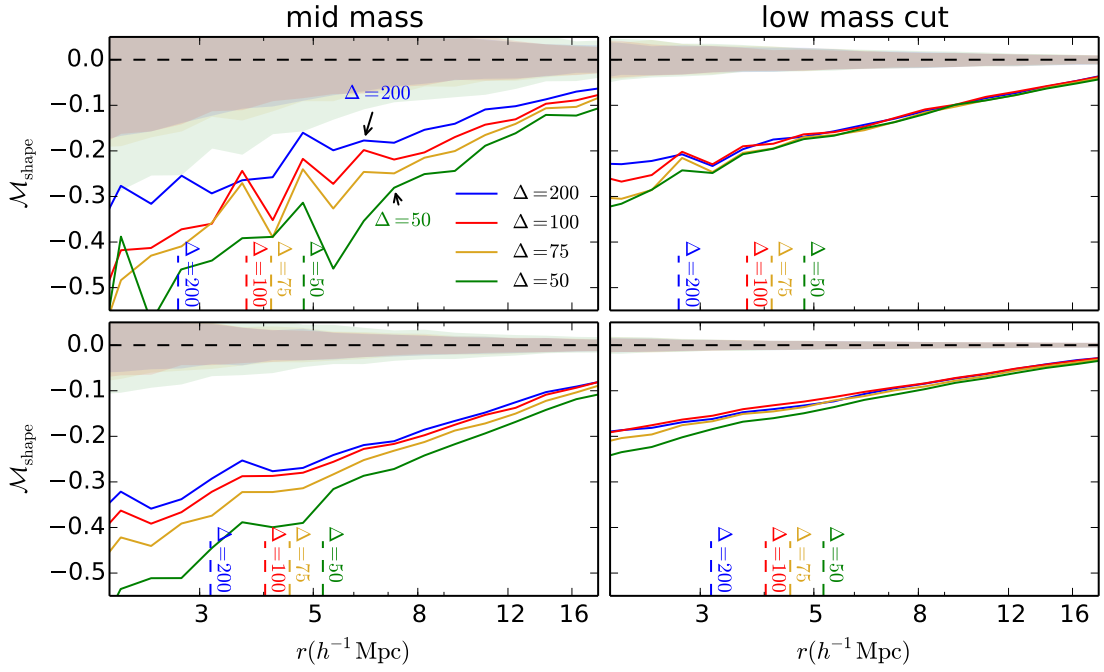

Figure 16. Comparison of the marked correlation function for the shape of the halo between the "mid mass" cutoff (left column) and the "low mass" cutoff (right column). The top row uses L0125 data while the bottom row uses L0250 data. The shaded bands represent 2-sigma confidence regions generated by randomization of the marks. The dashed lines along the bottom denote the largest halo radius for a given value of the overdensity parameter.

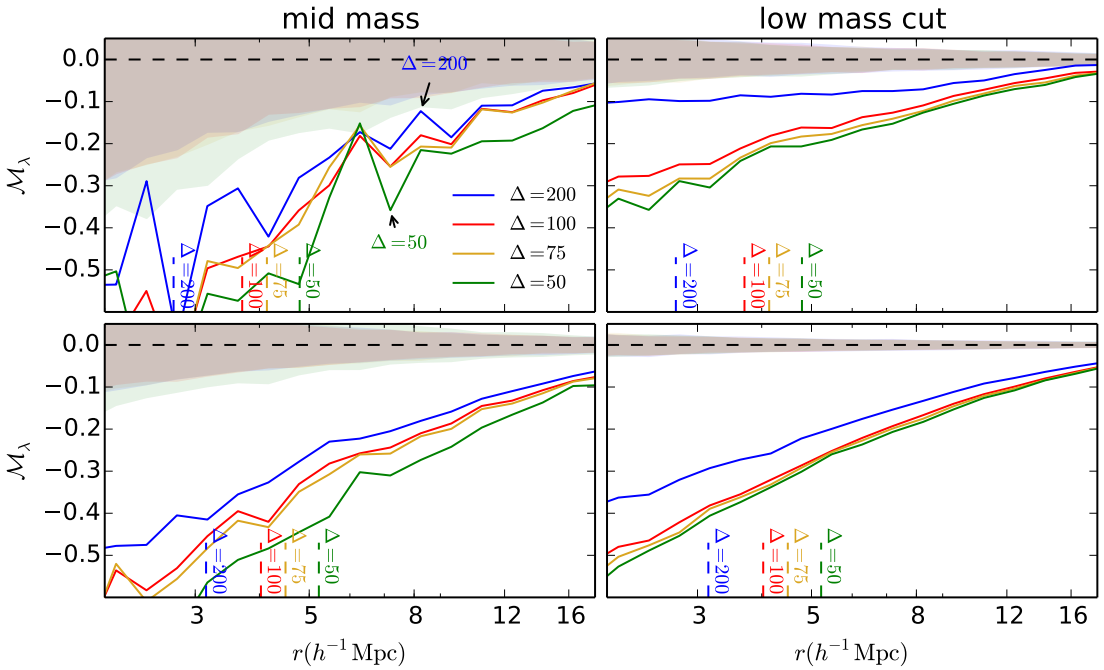

Figure 17. Comparison of the marked correlation function for the spin of the halo between the "mid mass" cutoff (left column) and the "low mass" cutoff (right column). The top row uses L0125 data while the bottom row uses L0250 data. The shaded bands represent 2-sigma confidence regions generated by randomization of the marks. The dashed lines along the bottom denote the largest halo radius for a given value of the overdensity parameter.

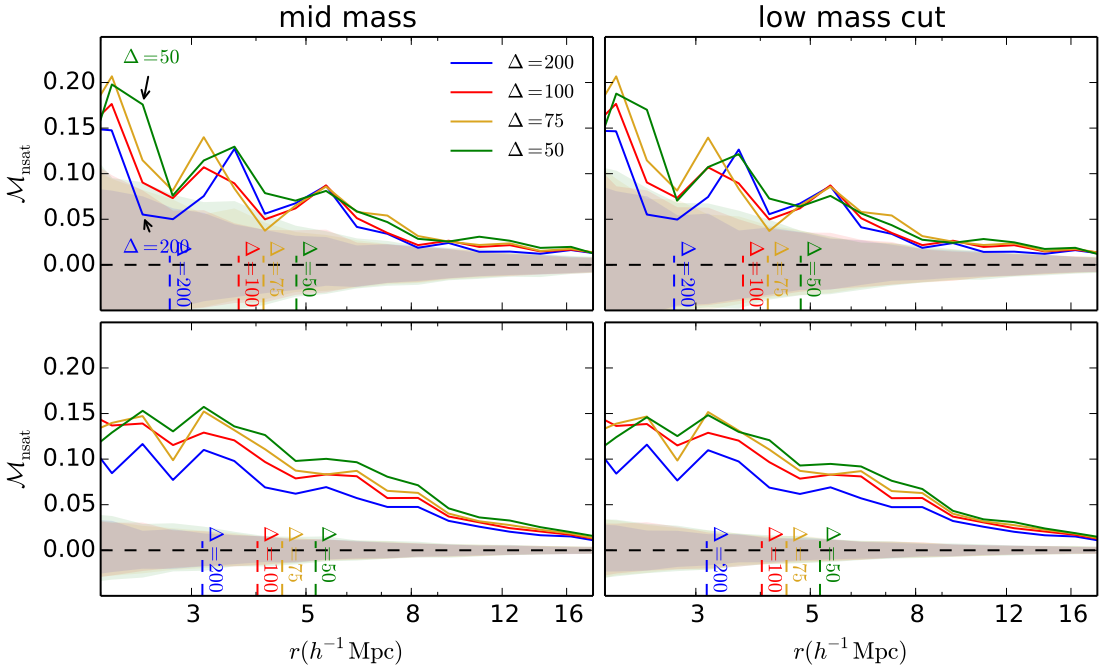

Figure 18. Comparison of the marked correlation function for the satellite number between the "mid mass" cutoff (left column) and the "low mass" cutoff (right column). The top row uses L0125data while the bottom row uses L0250data. The shaded bands represent 2-sigma confidence regions generated by randomization of the marks. The dashed lines along the bottom denote the largest halo radius for a given value of the overdensity parameter.

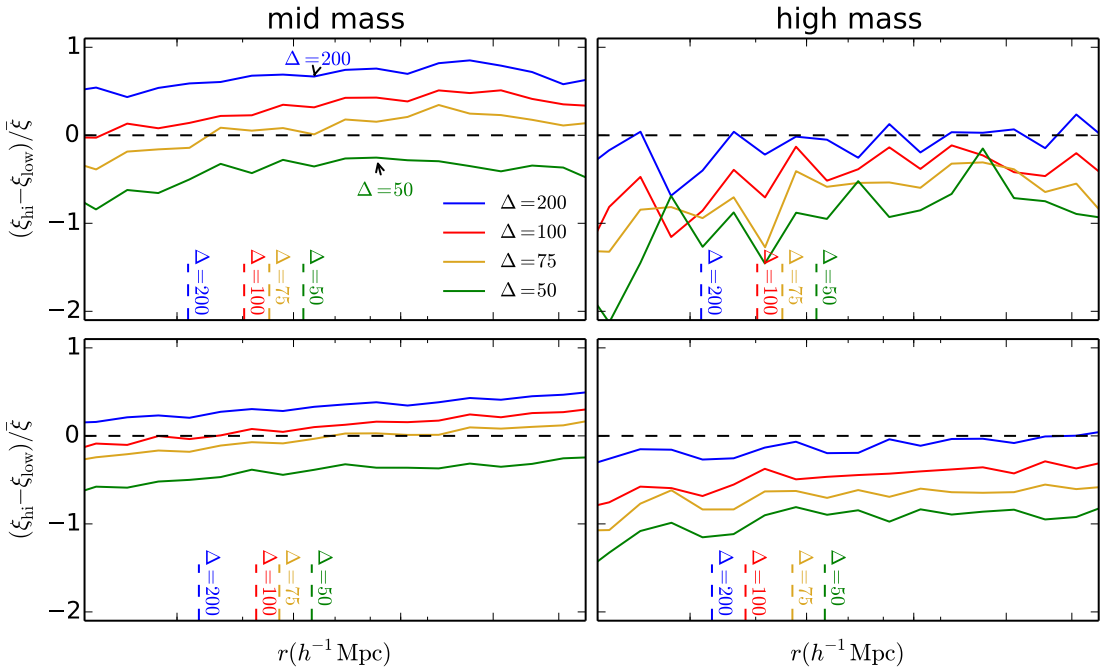

Figure 19. The difference of the correlation function for only the top 20% most concentrated halos and the bottom 20% in concentration, normalized by the overall correlation function of the entire sample. The top row uses L0250 data while the bottom row uses L0500 data. The left column utilizes the "mid mass" cutoff, while the right column demonstrates the "high mass" cutoff. The dashed lines along the bottom denote the largest halo radius for a given value of the overdensity parameter

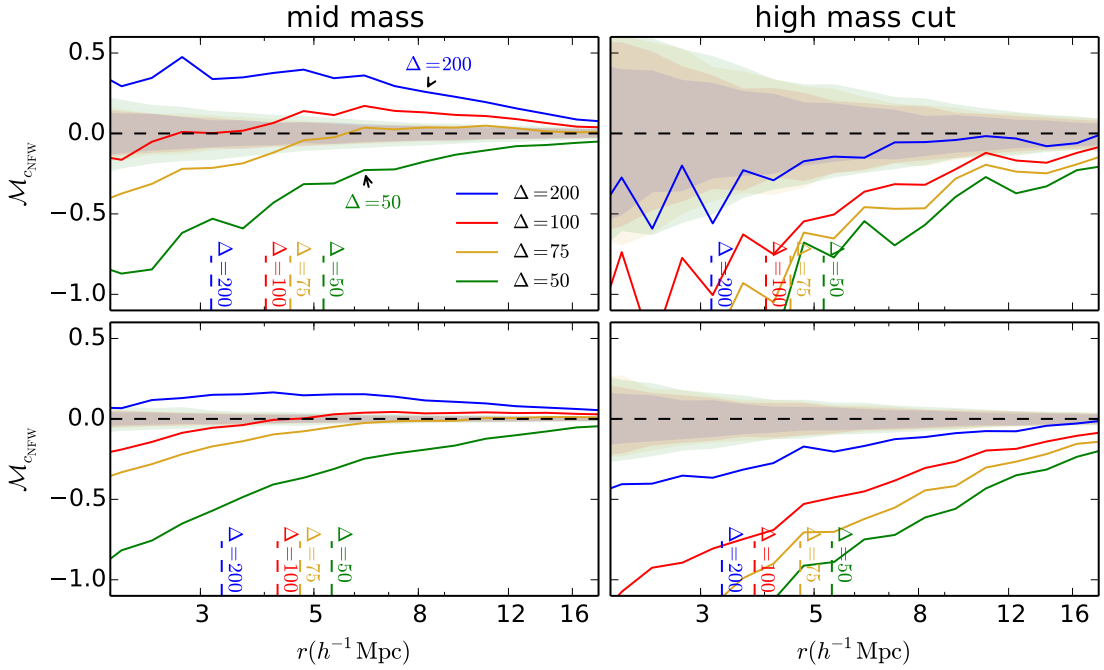

Figure 20. Comparison of the marked correlation function for the concentration defined according to the NFW profile between the "mid mass" cutoff (left column) and the "high mass" cutoff (right column). The top row uses L0250 data while the bottom row uses L0500 data. The shaded bands represent 2-sigma confidence regions generated by randomization of the marks. The dashed lines along the bottom denote the largest halo radius for a given value of the overdensity parameter.

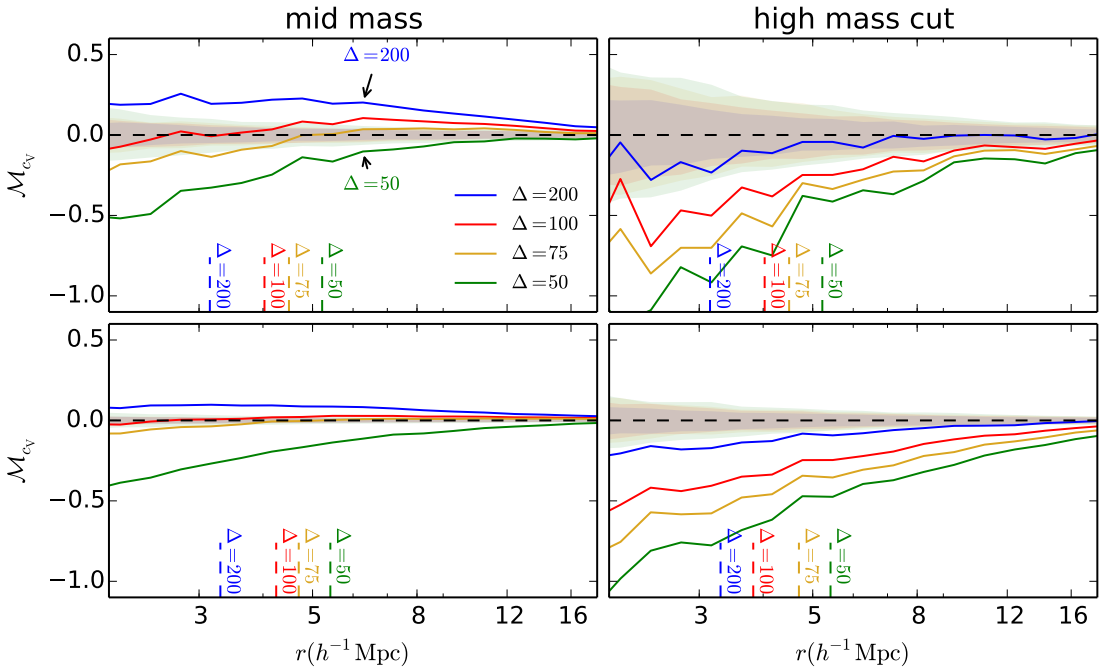

Figure 21. Comparison of the marked correlation function for the concentration defined according to the velocity ratio between the "mid mass" cutoff (left column) and the "high mass" cutoff (right column). The top row uses L0250 data while the bottom row uses L0500 data. The shaded bands represent 2-sigma confidence regions generated by randomization of the marks. The dashed lines along the bottom denote the largest halo radius for a given value of the overdensity parameter.

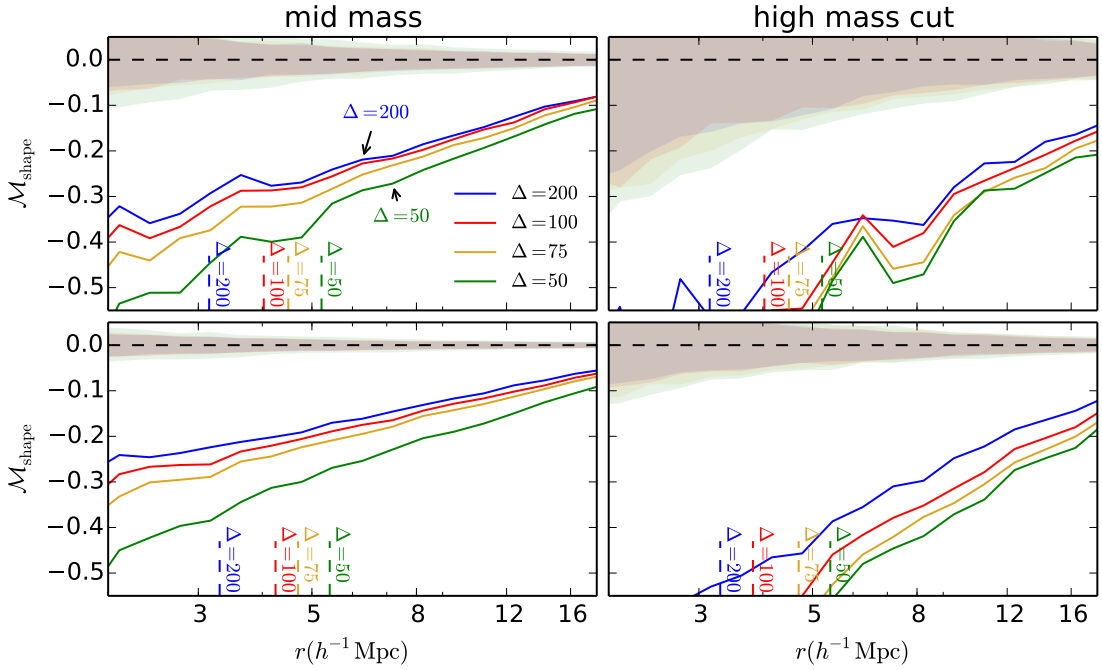

Figure 22. Comparison of the marked correlation function for the shape of the halo between the "mid mass" cutoff (left column) and the "high mass" cutoff (right column). The top row uses L0250 data while the bottom row uses L0500 data. The shaded bands represent 2-sigma confidence regions generated by randomization of the marks. The dashed lines along the bottom denote the largest halo radius for a given value of the overdensity parameter.

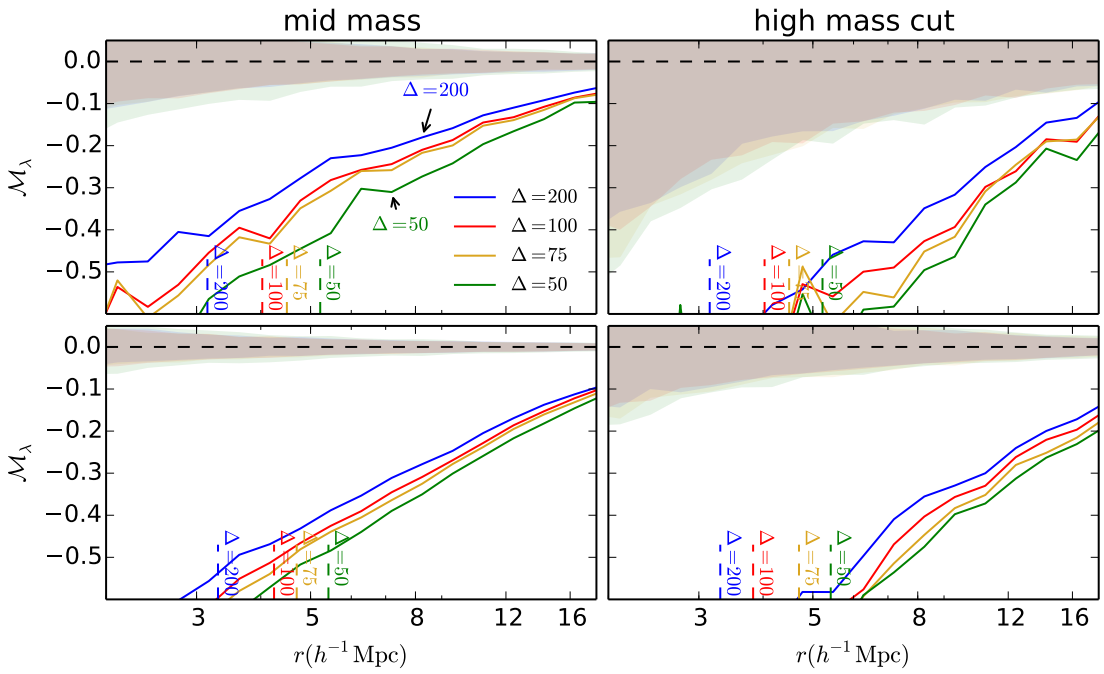

Figure 23. Comparison of the marked correlation function for the spin of the halo between the "mid mass" cutoff (left column) and the "high mass" cutoff (right column). The top row uses L0250 data while the bottom row uses L0500 data. The shaded bands represent 2-sigma confidence regions generated by randomization of the marks. The dashed lines along the bottom denote the largest halo radius for a given value of the overdensity parameter.

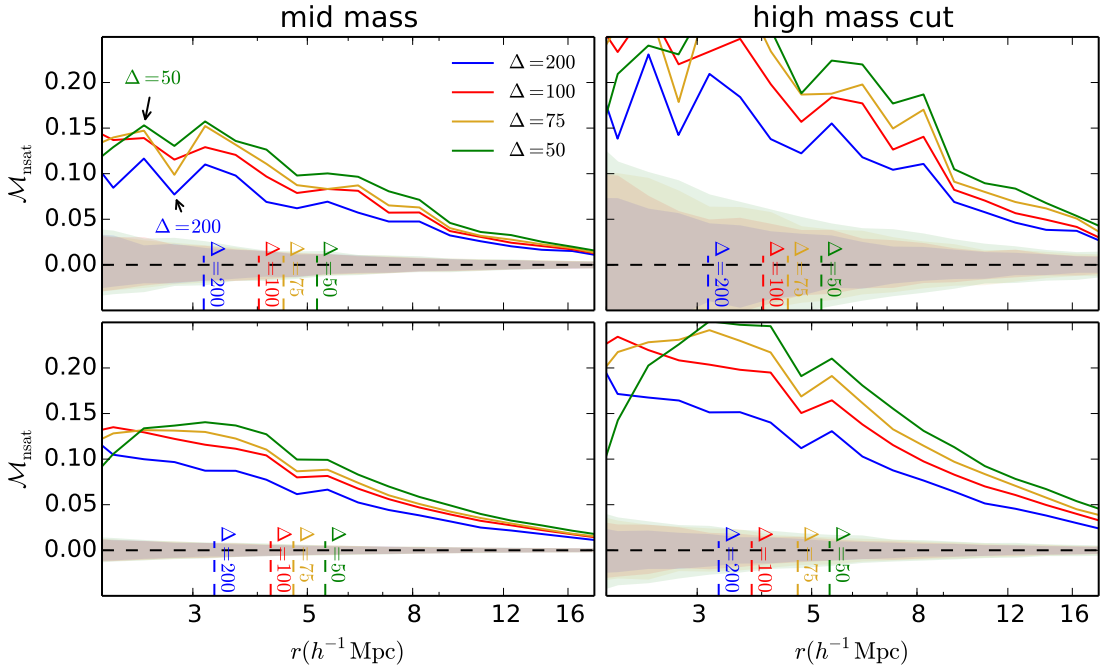

Figure 24. Comparison of the marked correlation function for the satellite number between the "mid mass" cutoff (left column) and the "high mass" cutoff (right column). The top row uses L0250 data while the bottom row uses L0500 data. The shaded bands represent 2-sigma confidence regions generated by randomization of the marks. The dashed lines along the bottom denote the largest halo radius for a given value of the overdensity parameter.

mass dependence that is not subject to the small number statistics at large halo masses. Taking our various tests, we then apply a change to the threshold density Δ in an attempt to remove the effect that environment has upon these properties. We come to the following conclusions from our simulation data.

- Our halo redefinition method does not cause any substantial breakdown in the ROCKSTAR halo finding algorithm, though this may not be the case for every halo finding methodology. This is something that should be considered prior to utilization of this method, unless working directly from particle data. As our initial halo sizes and locations are determined through spherical overdensities, it cannot be assumed that starting from a FoF grouping and then determining values through particle data directly will produce identical results. Similarly, different cosmologies may remove environmental effects at different scales.
- When looking at the two-point correlation function, there appears to be a "sweet spot" that appears to remove environmental effects the most efficiently. Going beyond that seems to reintroduce environmental effects, possibly as an extreme side effect of halo exclusion.
- For our marked correlation functions we see that both proxies of concentration that we use as marks show significant removal of environmental effects at large scales for similar values of the overdensity parameter Δ. In cases where one is only interested in the concentration of dark matter halos and large scales (or correspondingly small values of k), this method will allow you to compensate for bias that environment could introduce to calculations dependent upon the halo model. This may prove valuable for calculations such as that of the shear power spectrum calculated through weak lensing.
- The environmental effects on the shape of the host halo and the satellite number of the host halo cannot be removed regardless of the chosen redefinition of Δ. We propose that this may be intrinsically tied to the nature of the filaments, whose effects cannot be removed by a simple redefinition of the halo radius.
- This method is definitively related to the mass of the halos that are being observed. Furthermore, it appears that the majority of the reduction in assembly bias is tied to the exclusion of halos from the catalog as a result of being subsumed into larger halos. This information does not seem to be contradictory; it can be intuitively understood that the region about the most massive halos will be different than the region around the least massive halos, leading to a different frequency at which halos are being excluded. It does however warrant that careful consideration be given to the sample of halos that are of interest.
- The selection of halo size is intrinsically related to the assembly bias and varies across scales. This might help to resolve contradictory results in the search for halo assembly bias in the literature.

This methodology, while certainly not perfect in accounting for assembly bias, may be of significance when applied to galaxy formation models and give insight into seemingly conflicting results. Provided that the properties of interest in a given model behave well under our redefinition, it will allow us to create better mock galaxy catalogs without resorting to more complicated models requiring halo formation histories - giving us another powerful tool to test observation against.

There remain possible uncertainties to study in the future. One possible area of follow-up is the matter of simulation cosmology, which is not explored in this text. It is possible that the choice of cosmology may change observed assembly bias as a function of the halo masses, something that our methodology should be capable of observing. Furthermore, we can determine if the choice of halo size that best reduces assembly bias is a function of the chosen cosmology. This may be of interest in attempting to determine signatures of assembly bias in observational samples in the future.

ACKNOWLEDGMENTS

We are grateful to many people.

REFERENCES

Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013, Astrophys. J. , 762, 109
Berlind A. A., Weinberg D. H., 2002, Astrophys. J. , 575, 587
Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, Astrophys. J. , 379, 440
Bullock J. S., Wechsler R. H., Somerville R. S., 2002, MNRAS , 329, 246
Cooray A., Sheth R., 2002, Phys. Rep., 372, 1
Croton D. J., Gao L., White S. D. M., 2007, MNRAS , 374, 1303
Duffy A. R., Schaye J., Kay S. T., Dalla Vecchia C., 2008, MNRAS , 390, L64
Dvorkin I., Rephaeli Y., 2011, MNRAS , 412, 665
Gao L., White S. D. M., Jenkins A., Frenk C. S., Springel V., 2005, MNRAS , 363, 379
Gil-Marín H., Jimenez R., Verde L., 2011, MNRAS , 414, 1207
Harker G., Cole S., Helly J., Frenk C., Jenkins A., 2006, MNRAS , 367, 1039
Kazantzidis S., Zentner A. R., Kravtsov A. V., 2006, Astrophys. J. , 641, 647
Lacey C., Cole S., 1993, MNRAS , 262, 627
Navarro J. F., Frenk C. S., White S. D. M., 1997, Astrophys. J. , 490, 493
Peacock J. A., Smith R. E., 2000, MNRAS , 318, 1144
Peebles P. J. E., 1969, Astrophys. J. , 155, 393
Scoccimarro R., Sheth R. K., Hui L., Jain B., 2001, Astrophys. J. , 546, 20
Seljak U., 2000, MNRAS , 318, 203
Sheth R. K., Tormen G., 2004, MNRAS , 350, 1385
Somerville R. S., Kolatt T. S., 1999, MNRAS , 305, 1
Tempel E., Guo Q., Kipper R., Libeskind N. I., 2015, ArXiv e-prints
Velliscig M., Cacciato M., Schaye J., Bower R. G., Crain R. A., van Daalen M. P., Dalla Vecchia C., Frenk C. S., Furlong M., McCarthy I. G., Schaller M., Theuns T., 2015, ArXiv e-prints

Wechsler R. H., Zentner A. R., Bullock J. S., Kravtsov A. V., Allgood B., 2006, Astrophys. J. , 652, 71
White S. D. M., Rees M. J., 1978, MNRAS , 183, 341
Zentner A. R., 2007, International Journal of Modern Physics D, 16, 763