Prepared for submission to JCAP

Generating Mock Catalogs for the Baryon Oscillation Spectroscopic Survey: An Approximate N-Body approach

Aaronson Aardvark...e.t.c.

Abstract. We introduce and test an approximate scheme for generating mock catalogs for large-scale structure measurements in galaxy surveys, specializing in this work to the Baryon Oscillation Spectroscopic Survey. things to add later...A brief description of the approximation scheme, tests and the accuracy we reach, and some comments about the timings of the tests and the BOSS samples.

Contents

1 Introduction **1**

2 HACC **2**

3 Convergence Test: Selection of the minimal time steps **2**
3.1 Matching 2
3.1.1 Algorithm 2
3.1.2 Halo Properties 4
3.2 Mass Adjustment 5
3.2.1 Method 5
3.2.2 Observables 7

4 THE BOSS SIMULATIONS **8**
4.1 Simulation Parameters: Nikhil 8
4.2 Building the Galaxy Catalog 8
4.2.1 Halo Occupation Distribution 8
4.2.2 Halo profile 8
4.2.3 Galaxy velocities in halos 9
4.2.4 Galaxy Mocks for BOSS 9

5 Discussion and Summary **9**

1 Introduction

Recently, the necessity and the demand for the large number of large N-body simulations have been increased in astronomy due to the precision required for measurements to understand cosmic acceleration, like the Baryon Oscillation Spectroscopic Survey (BOSS) and the proposed Mid Scale-Dark Energy Spectroscopic Instrument (MS-DESI). As the area covered by those current and future spectroscopic galaxy surveys get larger, galaxy mocks necessary to the cosmological analysis also have to be generated from N-body simulations with the corresponding large volume, which is computationally expensive. Besides the large volume required for the N-body simulations, we need many realizations to achieve the precision required for those measurements.

There are several reasons for why we need large number of N-body simulations. One is to understand and to calibrate systematics caused by non-linear gravitational evolution and galaxy formation. Since those systematics are cosmology dependent, ideally we want to generate galaxy mocks from N-body simulations with various cosmologies. Another reason is to reduce noise for a covariance matrix estimation. Even when 50 realizations are used, the directly estimated covariance matrix is still noisy. Since it is unrealistic to run N-body simulations enough number of times to obtain a smooth directly estimated covariance matrix, there have been several efforts to obtain a smooth covariance matrix mainly in two directions. One approach is to estimate a covariance matrix from theory or by fitting to a modified form of the Gaussian covariance matrix (citations...). There are many different methods to estimate, but a peculiarity common to all the methods included in this approach requires some assumptions to compute the covariance matrix. Those assumptions often prevent those covariance matrices from properly accounting the effects due to non-linear gravitational evolution such as non-gaussianities and non-localities.

Other approach is generating enough number of galaxy mocks by using 2nd Order Perturbation Theory (2LPT) and computing a covariance matrix estimation directly (Scoccimarro and Sheth 2002, Manera et al. 2012, Pinocchio?). Using 2LPT makes generation of galaxy mocks significantly faster than the full N-body simulations, because 2LPT allows us to compute final positions of particles only with their initial positions without any steps in-between. The aim of our method is the same as this

approach in the sense that we want to shorten computational time to generate those mocks. The problem of using 2LPT is that those simulations cannot capture non-linear gravitational evolution and therefore this approach requires to tune halo definitions (i.e., the linking length for the Friends-of-Friends (FOF) algorithm) and halo masses. Also, it is hard to resolve small halos in this approach, and therefore galaxies which reside in those small halos are placed on randomly selected dark matter particles in the mocks. It is, however, important to keep relatively high mass resolution to understand the systematics caused by non-linear gravitational evolution.

Our N-body simulations are consist of two components: a long time step for solving the PM force and a set of short range sub-cycle steps for a direct particle-particle interaction. The idea behind is reducing the number of both time steps as much as we can preserve enough mass resolution to correctly describe the large scale distribution of galaxies.

In the following, we first briefly describe the mechanism of our approximated N-body simulations. In Section 3, we test and compare our method to the full N-body simulation and explain how we calibrate our samples. In Section 4, we populate halos in our samples with galaxies and compute correlation functions based on BOSS Data Release 9/11(?) geometry, and compare it with the observed correlation function.

2 HACC

3 Convergence Test: Selection of the minimal time steps

In this section, we examine how reducing the number of time steps affects on halo properties (i.e., halo mass, position, and velocity) as well as observables such as halo mass functions and halo bias. In order to quantitatively evaluate different time-stepping schemes, we run a set of convergence tests using smaller simulation boxes. We scale down these volumes to $(256h^{-1}\mathrm{Mpc})^3$ with 256^3 particles with the same particle mass as the simulations with $(4000h^{-1}\mathrm{Mpc})^3$. The number of time steps were chosen as 450/5, 300/3, 300/2, 150/3, 150/2, where the first number indicates the number of long time steps and the second number the number of short time steps for each long time step (Can I consider the simulations with 450/5 as the full N-body simulations or not? Also, can/should we compare our results with the full N-body simulations?). To evaluate different time-stepping schemes, we first compare the properties (masses, positions, and velocities) of the individual halos themselves by matching halos in one sample to halos in another. Following to that, we compare statistical descriptions provided by the mass functions and power spectra. We found that differences on halo properties do not significantly affect on those statistical observables.

3.1 Matching

Here, we compare halo properties by matching halos in different samples one by one. We first show our algorithm for identifying the corresponding halos in two different samples and then compare halo mass, position, and velocity for those matched halos. From the quantitative comparison, we find that the samples generated from the simulations with 300 global steps have much less scatter for the baseline of the sample of the 450/5 simulation than the samples with 150 global steps. In addition to that, we see that the differences between sub-cycles are almost negligible.

3.1.1 Algorithm

Since our simulations all start with the same initial conditions, we match halos in different simulations by matching their particle content. Given a halo in simulation A, we consider the halos in simulation B with the corresponding particles. Given this list of possible matches, we match to the halo with the largest number of common particles. To avoid spurious matches, we also require that the fraction of common particles (relative to simulation A) exceeds a threshold. As an example to illustrate how this matching algorithm works, we use the samples from the 300/2 simulation and the 450/5 simulation. Figure 1 shows the cumulative fraction of unmatched halos matching the 450/5 simulation to the 300/2 simulation at $z = 0.15$ with various thresholds. As expected, the unmatched fraction increases

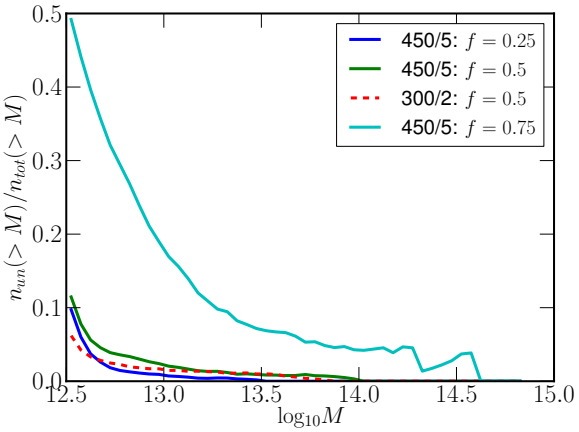

Figure 1. The cumulative fraction of unmatched halos matching the 450/5 simulation to the 300/2 simulation at z=0.15 as a function of halo mass. The solid lines, from top to bottom, correspond to matching thresholds of 75%, 50%, and 25% for the unmatched halos in the 450/5 sample. The dashed line shows the same quantity for the 300/2 sample for a threshold of 50%. As expected, the unmatched fraction increases with decreasing halo mass and increasing threshold. We adopt a threshold of 50% as our default choice. Do we really need the result for the 300/2 simulation?

with increasing threshold and decreasing halo mass. We adopt a threshold of 50% as our default choice.

Since the above matching algorithm is unidirectional, multiple halos in the sample A might be matched to a single halo in the sample B; this happens 1 to 2% of the time with a matching threshold of 50%. We refer to these as multiply-booked halos in what follows. Figure 2 compares halo masses matching the 450/5 simulation to the 300/2 simulation for the case of multiply-booked halos and the rest. The top left panel shows a mass scatter for all the matched halos between those two simulations, while the top right panel shows a mass scatter only for the non multiply-booked halos. The bottom panels show a mass scatter for the case of multiply-booked halos. The bottom left panel shows a mass scatter for individual multiply-booked halos, while we plots a summed halo mass for those corresponding halos in the bottom right panel. As shown in the top left panel, there are low-mass halos in the 450/5 simulation corresponding to high-mass halos in the 300/2 simulation. The same trend is observed for the case of multiply-booked halos, but not for the non multiply-booked halos. Furthermore, those disagreement for halo masses between the two simulations are resolved by adding the corresponding halo masses. This implies that there are multiple halos in the 450/5 simulation which are merged into one halo in the 300/2 simulation.

Figure 3 shows the number densities of the unmatched halos in the 450/5 simulation matching to the 300/2 simulation at $z = 0.15$. There are three reasons that halos are considered as unmatched. First, if particles forming a halo in the sample A do not form a halo in the sample B (i.e., halos in the samples do not share common particles), we consider them as unmatched. Second, if the fraction of common particles over the total number of particles in each halo is less than 50%, we eliminate halos for the case of spurious matching. At last for the case of multiply-booked halos, we remove all but the one with the largest number of common particles. We showed each unmatched number density as a function of halo mass. We only find unmatched halos on low-mass regions for the reason that the halos do not have any common particles. This is because there are some low-mass halos which are identified in one sample but not in another sample due to the way the FOF algorithm define halos. As shown, most of unmatched halos are due to the threshold criterion. We also checked how the number of matched halos is changed as a function of redshift, and we observed that redshift does not affect to the matching algorithm.

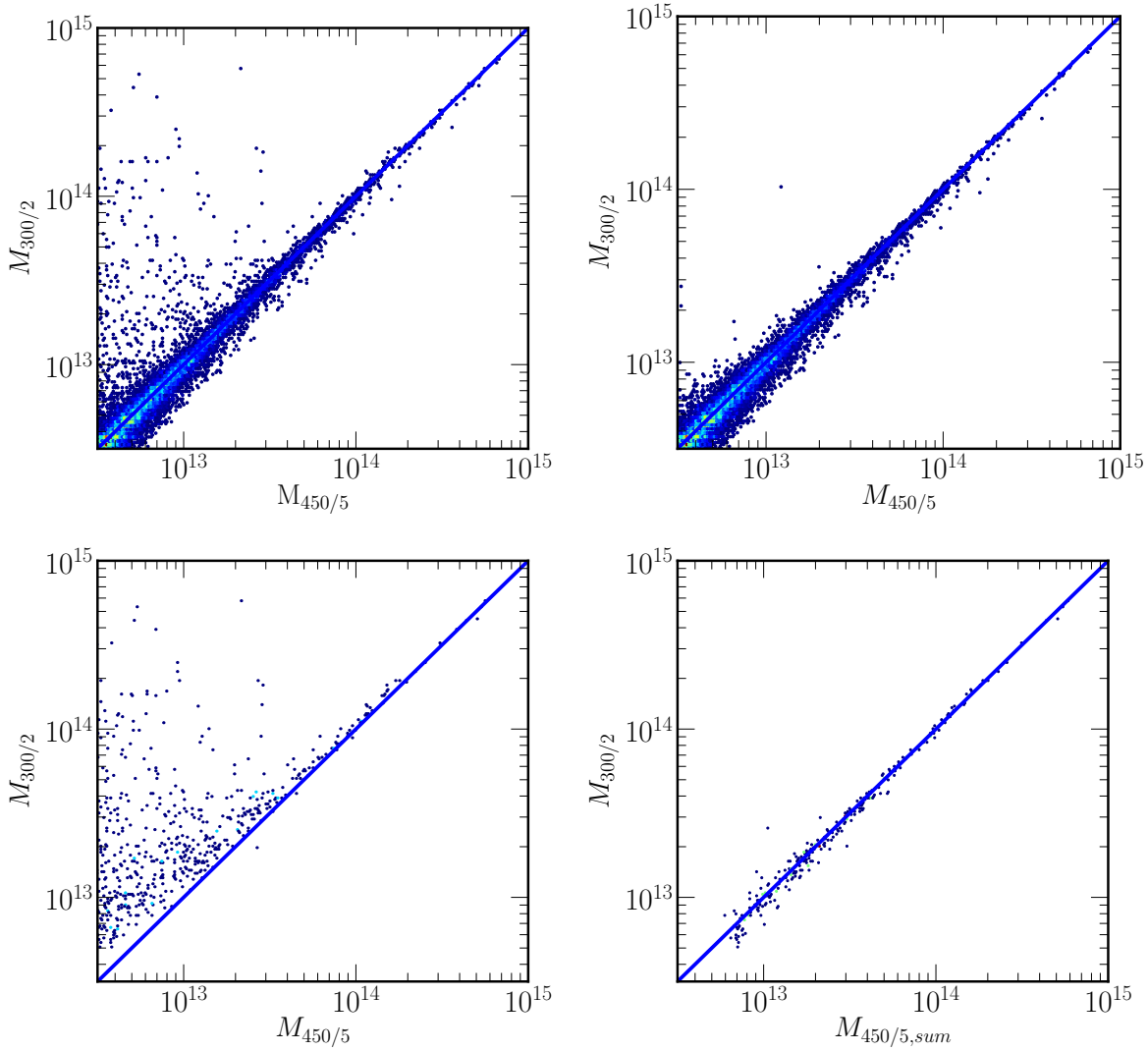

Figure 2. Comparison of halo masses matching the 450/5 simulation (x-axis) to the 300/2 simulation (y-axis) at $z = 0.15$. Panels correspond to halos with different matching criteria: all the matched halos (top left), matched halos having one-to-one correspondence (top right), matched halos not having one-to-one correspondence called "multiply-booked" halos (bottom left), and the "multiply-booked" halos whose corresponding halo masses are added (bottom right). Those panels imply that large mass difference between the 450/5 simulation and the 300/2 simulation shown in the top left panel is mainly because those "multiply-booked" halos in the 450/5 simulation are merged into one halo in the 300/2 simulation due to larger time steps.

3.1.2 Halo Properties

Here, we compare halo properties (i.e., halo mass, position, and velocity) for halos matched to those in the 450/5 simulation. Since we are interested in correctly describing the large-scale distribution of galaxies and it requires to correctly locate dark matter halos in the simulation with correct estimation of halo masses, it is crucial to know how reducing the number of time steps affects on halo properties systematically. The comparison of halo mass for different time-stepping schemes to the 450/5 simulation at $z = 0.15$ is shown in Figure 4. We take all the matched halos whose masses are between $10^{12.5}\mathrm{M}_\odot$ to $10^{13.0}\mathrm{M}_\odot$, $10^{13.0}\mathrm{M}_\odot$ to $10^{13.5}\mathrm{M}_\odot$, and $10^{13.5}\mathrm{M}_\odot$ to $10^{14.0}\mathrm{M}_\odot$, and compute their means and the standard deviations for $\log_{10}(\mathrm{M}/\mathrm{M}_{450/5})$, where $M_{450/5}$ is a halo mass for the

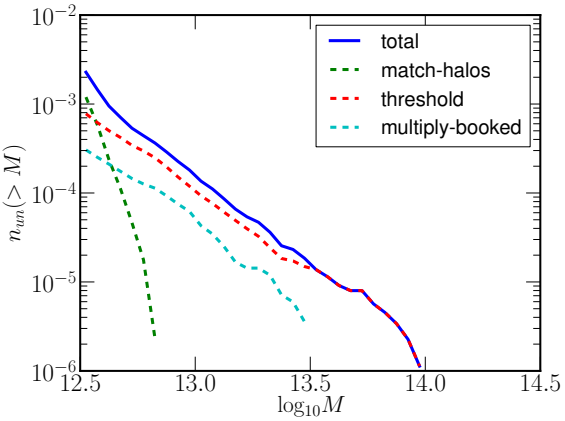

Figure 3. Itemization of unmatched halos shown as cumulative number densities of the unmatched halos from each procedure in the matching algorithm. The solid line is the total and the dashed lines correspond to matching based on particle content (green), elimination due to the matching threshold (red), and elimination of "multiply-booked" halos (cyan). Most large halos being unmatched is due to the threshold.

450/5 simulation and M corresponds to a halo in the samples generated with different time-stepping schemes. Figure 4 shows that halos generated from the simulations with small number of time steps have systematically lower masses than those in the 450/5 simulation. This is partly because we use the same linking length for the FOF algorithm to define halos for all the simulations. Since reducing the number of time steps results in an approximation to the true dynamics of dark matter fields, halos in the 450/5 simulation tend to have tighter and denser structure (show the snapshot?). So, using the same linking length will miss connecting some particles in the simulations with small number of time steps and will result in smaller halos than the corresponding halo in the 450/5 simulation. The panels in Figure 5, from left to right, show the comparison of halo position and velocity for the matched halos at $z = 0.15$. As shown, the 150 global steps have more scatter in the halo positions and the means differ for the velocity difference. For the 300 global steps, the results are significantly improved and the center position is matched in these cases to better than 200 kpc. As is clear from Figure 5, the difference between 3 and 2 sub-cycles is negligible on halo properties. Note that we observe the same trend in halo properties discussed here at different redshifts.

3.2 Mass Adjustment

In the previous subsection, Figure 4 shows that halos generated by the de-tuned simulations have systematically lower masses than the halos in the 450/5 simulation. This suggests the necessity of adjusting halo masses for those cases to the halo masses in the 450/5 simulation. In the following, we describe how we do a mass adjustment and show the resulted observables including mass functions and power spectra.

3.2.1 Method

To calibrate halo masses for the simulations with the reduced number of time steps, we first take all the matched halos between the 450/5 simulation and the de-tuned simulations and compute means for each mass bin. The reason we only take the matched halos is because our purpose of mass adjustment here is to correct systematic mass differences for the halos which are theoretically identical in different samples. After computing the means for each mass bin, we fit those means to a functional shown below so that M_{re} becomes close to the average halo masses for the 450/5 simulation:

$$M_{re} = M(1.0 + \alpha(M/10^{12.0}[\mathrm{M}_\odot])^\beta, \tag{3.1}$$

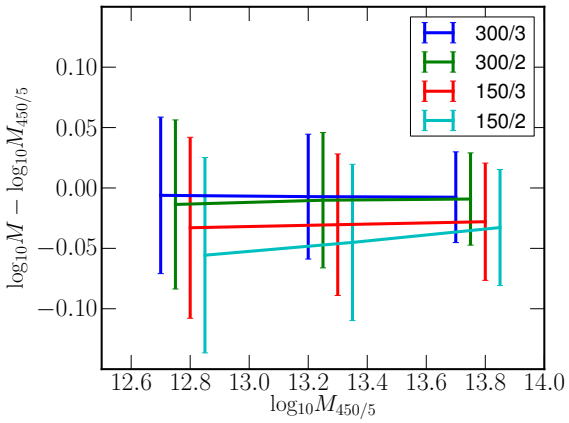

Figure 4. Comparison of halo mass for matched halos between the 450/5 simulation and other time-stepping schemes at $z = 0.15$. We take all the matched halos whose masses are between $10^{12.5}\mathrm{M}_\odot$ to $10^{13.0}\mathrm{M}_\odot$, $10^{13.0}\mathrm{M}_\odot$ to $10^{13.5}\mathrm{M}_\odot$, and $10^{13.5}\mathrm{M}_\odot$ to $10^{14.0}\mathrm{M}_\odot$, and compute the mean and the standard deviation for $\log_{10}(\mathrm{M}/\mathrm{M}_{450/5})$ where $M_{450/5}$ is a halo mass for the 450/5 simulation and M is for the simulations with different number of time steps corresponding to different colors in the plot. This plot shows that halo masses become systematically smaller for the case of small number of time steps than those in the 450/5 simulation.

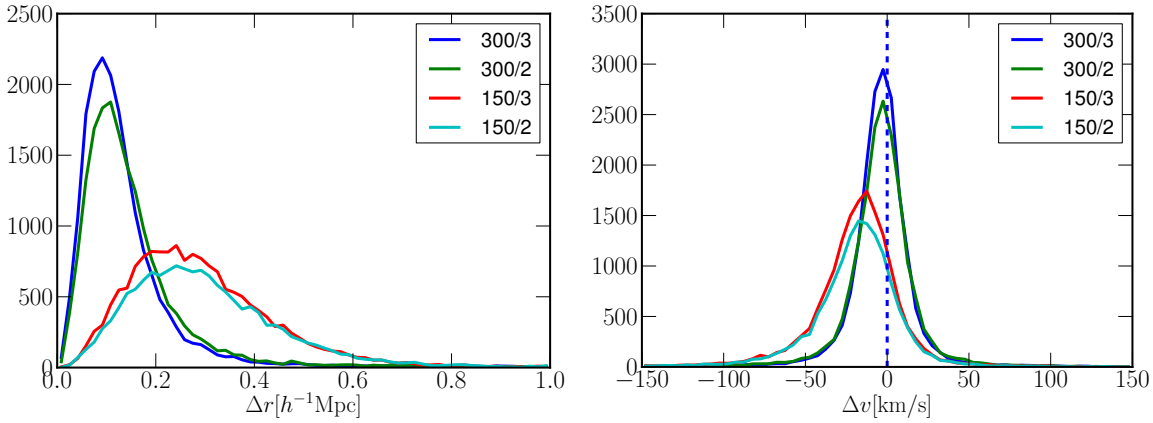

Figure 5. Comparison of matched halos in the different simulations corresponding to the time steps of 300/3 (blue), 300/2 (green), 150/3 (red), and 150/2 (cyan) with respect to the 450/5 simulation. From left to right, we compared halo position and velocity respectively. The agreement between 300 global steps and the 450/5 simulation is considerably good, with little difference from the number of sub-cycles.

	α	β
300/3	0.005	0.175
300/2	0.07	-0.47
150/3	0.101	-0.162
150/2	0.315	-0.411

Table 1. Corresponding α and β in Eqn. 3.1 for the samples with different number of time steps at $z = 0.15$.

where M_{re} is a reassigned halo mass, M is an original halo mass, α and β are free parameters. Correponding α and β for the simulations with different number of time steps at $z = 0.15$ are listed on Table 1.

Those free parameters α and β can be described as a function of redshift. For the case of the

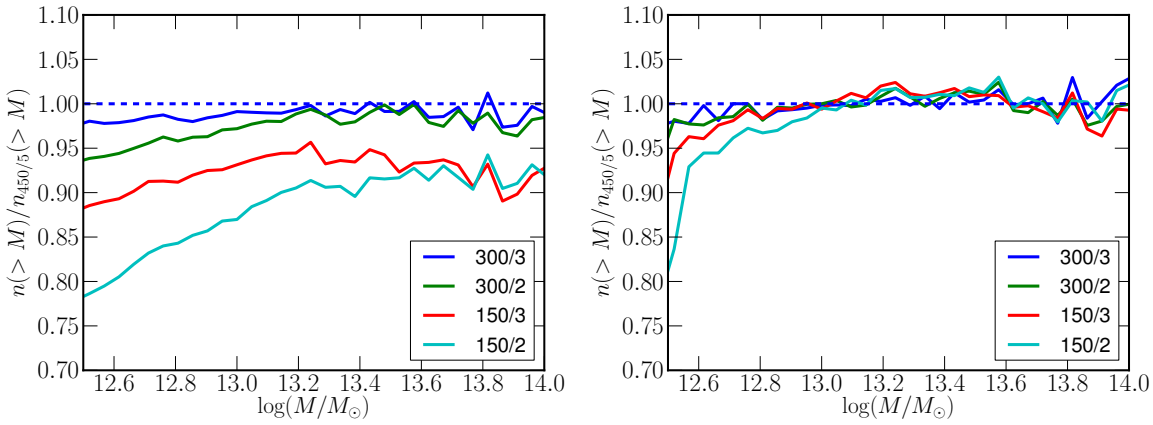

Figure 6. Comparison of cumulative mass functions in different simulations taking the 450/5 simulation as a reference. Lines, from top to bottom, correspond to the simulation with different time steps, 300/3 (blue), 300/2 (green), 150/3 (red), and 150/2 (cyan) respectively. A left panel shows the cumulative mass functions before mass adjustment, and a right panel is after mass adjustment. Those plots indicate that a systematic mass adjustment successfully calibrate halo masses for the de-tuned simulations, even for the case of the 150/2 simulation, which has more than 10% discrepancy before the mass adjustment.

300/2 simulation as an example, we find best fit parameters shown below (any examples or plots for here? How can I tell the readers that this mass adjustment is successful more quantitatively?):

$$\alpha(z) = 0.123z + 0.052, \tag{3.2}$$

and

$$\beta(z) = -0.154z - 0.447. \tag{3.3}$$

3.2.2 Observables

Now, we show mass functions and power spectra for the different simulations after applying the mass adjustment.

We first compute mass functions from outputs of different time-stepping schemes, as shown in Figure 6, where we compare simulations with reduced number of time steps to the 450/5 simulation at $z = 0.15$. In Figure 6, we show the ratio $n(> M)/n_{450/5}(> M)$, where $n_{450/5}(> M)$ is a cumulative mass function for the 450/5 simulation and $n(> M)$ is a cumulative mass function for different time steps shown in different colors. We compare the results before and after mass adjustment (left and right panels respectively). While the mass functions from the 250/3 and 150/2 simulations are suppressed more than10% on all mass ranges before mass adjustment, they are significantly improved after mass adjustment, especially on halo masses greater than $10^{13.0}\mathrm{M}_\odot$. For the simulations with the 300 global time steps, mass adjustment is especially effective on small halo masses.

The next measure of interest is halo-matter cross power spectra between halo and matter density fields, as shown in Figure 7. Figure 7 shows the ratio $P_{hm}/P_{hm,450/5}$ at $z = 0.15$, where $P_{hm,450/5}$ is the cross power spectrum for the 450/5 simulation and P_{hm} is the cross power spectrum for other time steps corresponding to different colors, as labeled in Figure 7. We use the real-space halo density field for the left panel and the redshift-space halo densities for the right panel in Figure 7. For the dark matter density field, we use the output of the 450/5 simulation for all the halo samples. Note that the dark matter density fields are in real-space for both cases. In this way, the ratio $P_{hm}/P_{hm,450/5}$ in real-space is equivalent to the ratio of halo bias between the 450/5 simulation and the simulations with other time-steps. To select halos, we apply the soft-mass cut method using the probability given by

$$< N_{halo}(M) >= \frac{1}{2}\mathrm{erfc}(\frac{\log(\mathrm{M_{cut}}/M)}{\sqrt{2}\sigma}), \tag{3.4}$$

where we set $\mathrm{M_{cut}} = 10^{13.0}[\mathrm{M_\odot}]$ and $\sigma = 0.5$. This probability has a similar form to the halo occupation distribution (hereafter, HOD) technique so that the probability gradually becomes one as increasing halo mass. We use this method to avoid noise from halos scattering across sharp boundaries on halo mass. Note that the errors calculated here are not due to sample variance, because we generate 10 samples from one full sample by using the soft-mass cut method. We see that, as we decrease the number of time steps, the ratio of the cross power spectra increases, especially in redshift-space, we observe large deviations from one on small scales for the 150/2 and 150/3 simulations. This is due to the overall smaller halo velocities for those simulations, which is shown in Figure 5. For the simulations with the 300 global time steps, overall agreements with the 450/5 simulation are almost 1% on any scales in both real-space and redshift-space.

As a conclusion through several convergence tests shown in this section, we find that the observables such as mass functions and power spectra are not affected by the differences on halo properties as much as those systematic differences (better way to phrase?).

4 THE BOSS SIMULATIONS

4.1 Simulation Parameters: Nikhil

describe box size, masses, geometry etc. Show that we can fit in two BOSS volumes per box.

4.2 Building the Galaxy Catalog

4.2.1 Halo Occupation Distribution

To generate galaxy mock catalogs, we use the HOD functional form (citation?) to populate halos with galaxies. Each halo either hosts a central galaxy or does not, while the number of satellites is Poisson distributed about a mean N_{sat}. For each sample, we paametrize $N_{gal}(M) = N_{cen} + N_{sat}$ with five parameters. The HOD functional form gives probabilities for the number of central and satellite galaxies in a halo as a function of halo mass:

$$N_{cen}(M) = \frac{1}{2}\mathrm{erfc}\left[\frac{\ln(M_{cut}/M)}{\sqrt{2}\sigma}\right], \tag{4.1}$$

and

$$N_{sat}(M) = N_{cen}(M)\left(\frac{M - \kappa M_{cut}}{M_1}\right)^{\alpha}, \tag{4.2}$$

where M_{cut}, M_1, σ, κ, and α are free parameters and M is a halo mass. Note that $N_{sat}(M)$ is zero when $M < \kappa M_{cut}$. Satellite galaxies exist only when a central galaxy exists in the halo.

4.2.2 Halo profile

Central galaxies always live at the minimum of the halo potential while satellite galaxies are randomly placed assuming an NFW profile (Navarro, Frenk & White 1996):

$$\rho(r) = \frac{4\rho_s}{\frac{r}{r_s}(1+\frac{r}{r_s})}, \tag{4.3}$$

where r_s is the characteristic radius defined as $r_s = R_{vir}/c$, where R_{vir} is a virial radius for a halo and c is the concentration parameter. There have been several studies how to describe the concentration parameter c as a function of halo mass and redshift (citations). In this paper, we used the formula presented in Klypin et al. 2010:

$$c(M,z) = \frac{c_0}{1+z}(M - M_0)^{-\beta}, \tag{4.4}$$

where$c_0 = 9.60$, $M_0 = 10^{12.}$, and $\beta = 0.75$.

4.2.3 Galaxy velocities in halos

$$< v^2 >= \frac{GM}{R_{vir}} \tag{4.5}$$

4.2.4 Galaxy Mocks for BOSS

5 Discussion and Summary

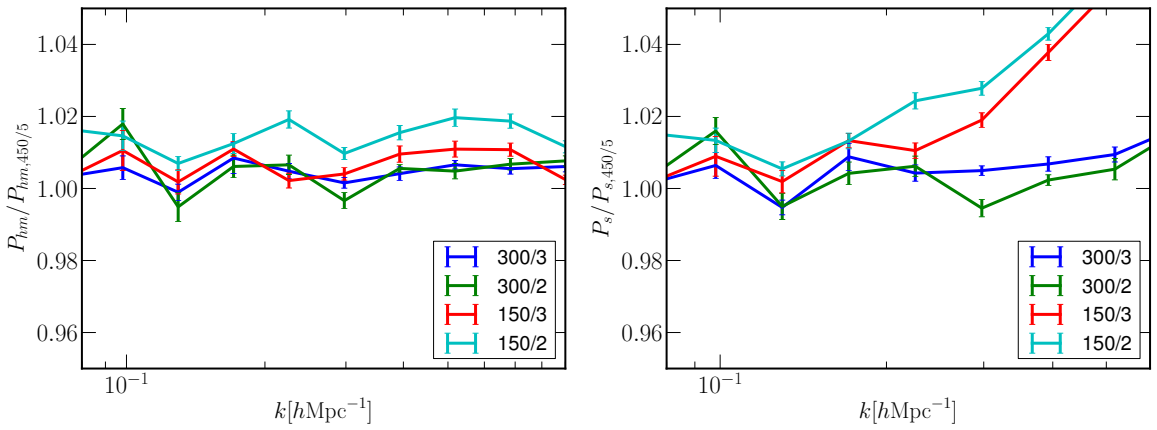

Figure 7. Ratio of halo-matter cross power spectra as a function of time steps with respect to the 450/5 simulation at $z = 0.15$. We use the real-space halo density field for the left panel and the redshift-space halo density field for the right panel, while the dark matter density fields used here are in real-space for both cases. The left panel shows that agreements with the 450/5 simulation are all within 2%. In the right panel, the large discrepancy of the cross power spectra for the simulations with 150 global steps on small scales is mainly due to the systematically small velocities, as shown in Figure 5. Note that the halos are selected based on the soft mass-cut method with $M_{cut} = 13.0$ and $\sigma = 0.5$.

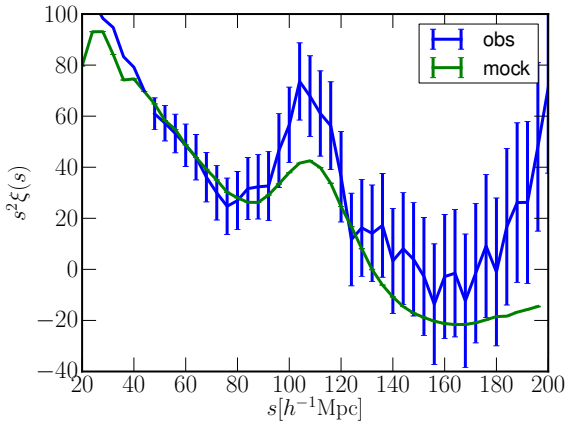

Figure 8. Correlation function monopoles $\xi(s)$ of the mocks (green) and of the CMASS galaxy measured in Anderson et al. (2013) (blue) at $z = 0.55$. The BAO peak of the mock correlation function is smoothed due to the wide window function.