The AEMULUS Project. I. Numerical Simulations for Precision Cosmology

Joseph DeRose[1,2] ⓘ, Risa H. Wechsler[1,2] ⓘ, Jeremy L. Tinker[3], Matthew R. Becker[1,2,4] ⓘ, Yao-Yuan Mao[5] ⓘ,
Thomas McClintock[6], Sean McLaughlin[1,2] ⓘ, Eduardo Rozo[6], and Zhongxu Zhai[3] ⓘ

[1] Kavli Institute for Particle Astrophysics and Cosmology and Department of Physics, Stanford University, Stanford, CA 94305, USA
[2] Department of Particle Physics and Astrophysics, SLAC National Accelerator Laboratory, Menlo Park, CA 94305, USA
[3] Center for Cosmology and Particle Physics, Department of Physics, New York University, 4 Washington Place, New York, NY 10003, USA
[4] Civis Analytics, Chicago, IL 60607, USA
[5] Department of Physics and Astronomy and the Pittsburgh Particle Physics, Astrophysics and Cosmology Center (PITT PACC), University of Pittsburgh, Pittsburgh, PA 15260, USA
[6] Department of Physics, University of Arizona, Tucson, AZ 85721, USA

Abstract

The rapidly growing statistical precision of galaxy surveys has led to a need for ever more precise predictions of the observables used to constrain cosmological and galaxy formation models. The primary avenue through which such predictions will be obtained is suites of numerical simulations. These simulations must span the relevant model parameter spaces, be large enough to obtain the precision demanded by upcoming data, and be thoroughly validated in order to ensure accuracy. In this paper, we present one such suite of simulations, forming the basis for the AEMULUS Project, a collaboration devoted to precision emulation of galaxy survey observables. We have run a set of 75 $(1.05\,h^{-1}\,\mathrm{Gpc})^3$ simulations with mass resolution and force softening of $3.51 \times 10^{10}(\Omega_m/0.3)\,h^{-1}\,M_\odot$ and $20\,h^{-1}\,\mathrm{kpc}$, respectively, in 47 different wCDM cosmologies spanning the range of parameter space allowed by the combination of recent cosmic microwave background, baryon acoustic oscillation, and Type Ia supernova results. We present convergence tests of several observables including spherical overdensity halo mass functions, galaxy projected correlation functions, galaxy clustering in redshift space, and matter and halo correlation functions and power spectra. We show that these statistics are converged to 1% (2%) or to the sample variance of the statistic, whichever is larger, for halos with more than 500 (200) particles, respectively, and scales of $r > 200\,h^{-1}\,\mathrm{kpc}$ in real space or $k \sim 3\,h\,\mathrm{Mpc}^{-1}$ in harmonic space for $z \leqslant 1$. We find that the dominant source of uncertainty comes from varying the particle loading of the simulations. This leads to large systematic errors for statistics using halos with fewer than 200 particles and scales smaller than $k \sim 4\,h\,\mathrm{Mpc}^{-1}$. We provide the halo catalogs and snapshots detailed in this work to the community at https://AemulusProject.github.io.

Key words: large-scale structure of universe – methods: numerical – methods: statistical

1. Introduction

The era of precision cosmology from galaxy surveys is upon us. Galaxy survey data sets have achieved constraining power on a subset of cosmological parameters comparable to measurements of the cosmic microwave background (CMB; Alam et al. 2017; Abbott et al. 2018), but unlike the CMB, these constraints rely on the measurement and modeling of nonlinear structure. In a very real sense, these analyses are already systematics limited, disregarding significant portions of their data in order to mitigate modeling uncertainties. For example, Abbott et al. (2018) limited themselves to scales for which baryonic feedback and nonlinear effects from galaxy biasing could be ignored. Alam et al. (2017), presenting the final analysis of the BOSS galaxy redshift survey, restricted their redshift space distortion measurements to $s > 20\,h^{-1}\,\mathrm{Mpc}$ and $k < 0.15\,h\,\mathrm{Mpc}^{-1}$ in configuration and Fourier space, respectively, to avoid uncertainties in modeling the galaxy velocity field.

Analytic models of these effects for simply selected samples are improving, but even the best models only claim to be accurate to the percent level at $k \sim 0.3\,h\,\mathrm{Mpc}^{-1}$ for matter and halo power spectra before taking into account effects due to hydrodynamics, feedback, and redshift space distortions (Perko et al. 2016; Cataneo et al. 2017). Nonlinear effects are much more difficult to avoid in the halo mass function (HMF), and analytic predictions such as those in Press & Schechter (1974) and Sheth & Tormen (1999) are only accurate at the ~10% level

(Tinker et al. 2008). Depending on the observable, this level of precision is either already a dominant source of error or will be in the very near future (see, e.g., Tinker et al. 2012). While 1% precision in observables is often quoted as a necessary goal, the required precision on predictions for observables is often not this stringent. For instance, McClintock et al. (2019) determined that the precision required for the HMF in order for it to contribute no more than 10% of the total uncertainty in cluster mass calibration for upcoming surveys is 3% at its most demanding.

While analytic methods struggle with nonlinear structure formation, a clear alternative exists in numerical simulations. In the case of gravity, where we have a well-understood standard theory described by general relativity, the effectiveness of simulations is limited only by the coarseness of the discretization allowed by currently available computers. Different algorithms for solving for nonlinear structure growth in dark-matter-only simulations have been shown to produce predictions for the matter power spectrum that are converged at better than the 1% level to $k \sim 1\,h\,\mathrm{Mpc}^{-1}$ (Heitmann et al. 2009; Schneider et al. 2016). It should be noted that these studies are of relative convergence, whereas studies of absolute convergence to the true physical solution still have an open question that likely depends on a better understanding of baryonic physics, neutrinos, and the nature of dark matter itself. Because of the relative successes of the aforementioned simulations, almost all cosmological analyses involving galaxy surveys now

use them in some form (Kitaura et al. 2016; Joudaki et al. 2018; MacCrann et al. 2018).

While great strides have been made in improving their computational efficiency, *N*-body simulations are still relatively expensive. For example, the DS14_A simulation (Skillman et al. 2014), one of the largest simulations run to date with a simulated volume of $(8\,h^{-1}\,\mathrm{Gpc})^3$ and 1.07×10^{12} particles, took approximately 400,000 node-hours, using approximately two-thirds of the TITAN supercomputer for close to two days. While this simulation approaches the volume of many ongoing and upcoming galaxy surveys, it does not even resolve all of the host halos of galaxies in a survey like DES.

Cosmological parameter constraints typically rely on sampling schemes such as Monte Carlo Markov Chains in order to explore parameter space. Modern analyses including cosmological and nuisance parameters numbering in the tens must sample on the order of millions of different cosmologies to reach convergence. Running an *N*-body simulation at each of these steps is not a prospect that will be achievable in the near future; even when considering simulations smaller than DS14_Ahus, there is a need for methodologies that can use relatively few simulations to make robust predictions for the full cosmological parameter space being constrained. Much of the work in this area has been driven by the need for accurate predictions of the matter power spectrum for weak-lensing analyses. For example, the HALOFIT methodology (Smith et al. 2003; Takahashi et al. 2012) fit an analytic expression to a set of *N*-body simulations in various cosmologies to obtain predictions for the matter power spectrum accurate to 5% for $k < 1\,h\,\mathrm{Mpc}^{-1}$ and 10% for $1\,h\,\mathrm{Mpc}^{-1} < k < 10\,h\,\mathrm{Mpc}^{-1}$.

Investigations into more advanced methodologies are ongoing, typically combining algorithms for optimally sampling a chosen cosmological parameter space and a method for interpolating between the observables at the sampled cosmologies. This approach, dubbed cosmic emulation, was first demonstrated for the matter power spectrum in Heitmann et al. (2009). They showed convergence of their simulation results with respect to a number of choices made in solving the *N*-body problem, including mass resolution, force softening, and simulation volume. This work has since been extended to the Friends-of-Friends (FoF) HMF (Heitmann et al. 2016), galaxy correlation function, and galaxy–matter cross-correlation function (Wibking et al. 2019), among other observables. Studies of the convergence of these statistics are not as complete as those for the matter power spectrum. Work toward validating the convergence of these statistics is vital to ensuring the accuracy of predictions built from simulations.

This type of validation is the primary concern of this work. The simulations presented here form the basis for the first set of emulators that is being built as a part of the AEMULUS project, a collaboration focused on the emulation of galaxy survey observables. The goal of the validation presented here is to provide robust convergence estimates for the statistics in question so that they may be properly accounted for in emulators built from these simulations. Emulators for the HMF and redshift space galaxy clustering using these simulations are presented in McClintock et al. (2019) and Zhai et al. (2018), respectively. Additionally, we hope to provide convergence guidelines for future work that simulates the statistics presented here.

In Section 2, we present our cosmological parameter space and the Latin Hypercube (LH) algorithm used to sample from
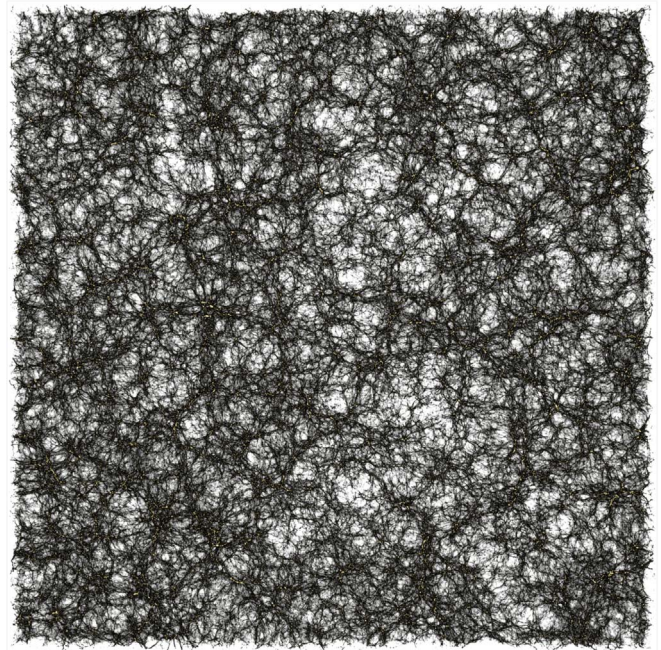

Figure 1. A $50\,h^{-1}\,\mathrm{Mpc}$ thick slice through B25 with density deposition performed as described in Kaehler et al. (2012).

it. In Section 3, we discuss our simulation framework. In Section 4, we discuss the convergence of the observables we emulate with respect to the choices made in our *N*-body solver. In Section 5, we discuss issues related to halo finding and halo definitions, and in Section 6, we compare our simulations to existing emulators. In Section 7, we discuss our plans to release these simulations to the public, and in Section 8 we conclude.

2. Cosmological Parameter Space

The goal of the parameter selection algorithm is to optimally span a large-dimensional space with a limited number of points. Our criterion for optimization is to maximize the accuracy of any scheme to interpolate statistics between the points, which requires the points to be as close to uniformly spaced as possible, while covering as much of the space as possible. We follow the technique outlined in Heitmann et al. (2009), with minor modifications. The process begins with an LH containing $M = 40$ samples of our $N =$ seven-dimensional space. In an LH design, each of the N dimensions is divided into M bins. In each dimension, each of the bins is selected once with no repeats, thus guaranteeing the full range of each parameter within the space is represented sparsely.

A random LH design is not optimally spaced, however. Points can be clumped together, as shown in a two-dimensional projection of our seven-dimensional space in the left-hand side of Figure 2. To quantify the spacing of a given LH, for every point in the space ,we calculate the distance to the closest point in each two-dimensional projection of the space. The quantity of interest is the sum of all minimum distances for all points in all projections. The space is optimal when this quantity is maximized, thus removing any clumping between points and pushing the points to a uniform distribution. To accomplish this, we use an iterative procedure that takes two points from the sample and swaps values in one dimension. If this swapping increases the quantity of interest, the swap is accepted. If it does not, the swap is rejected. This procedure
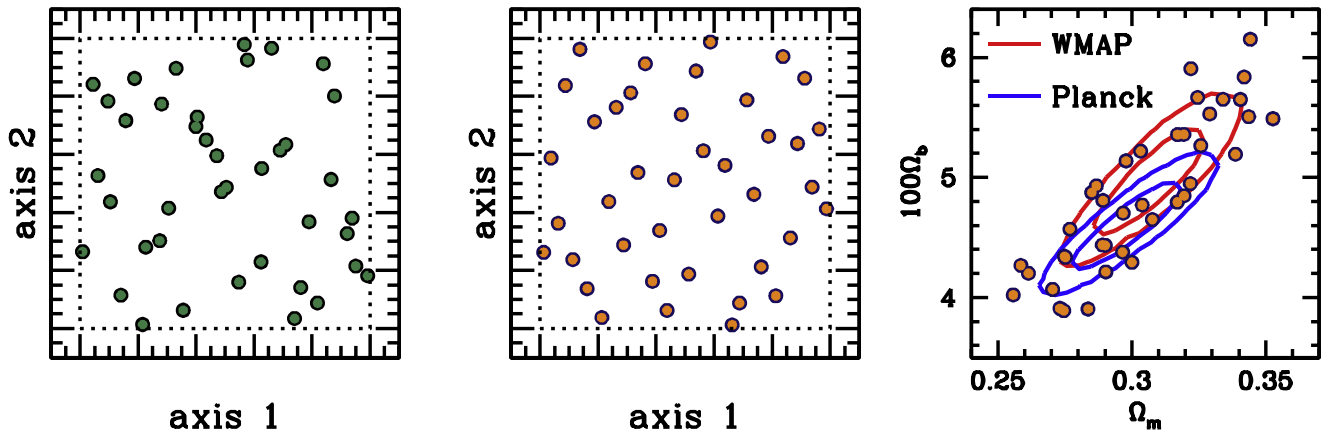
Figure 2. Left panel: two-dimensional projects of a random seven-dimensional Latin Hypercube (LH), with 40 points in total. Middle panel: the same LH, now optimized for more uniform spacing between points. Right panel: the same LH as shown in the middle panel, but now rotated into the eigenspace defined by CMB data. Contours are the WMAP9 and Planck13 joint constraints with BAO and supernovae.

is iterated until convergence. The result of this procedure is shown in the middle panel of Figure 2.

An LH design, by construction, creates a distribution of points in an M-dimensional cube. However, we do have prior knowledge on the distribution of cosmological parameters, and we want the distribution of our points to follow the degeneracies between parameters given current constraints. We use the combination of CMB, baryon acoustic oscillations (BAO), and supernovae (SNe). Specifically, we use the CosmoMC chains produced in the cosmology analysis of the BOSS DR11 BAO analysis (Anderson et al. 2014). Separate chains were run for nine-year *Wilkinson Microwave Anisotropy Probe* (*WMAP*) results (Hinshaw et al. 2013) and for *Planck* 2013 results (Planck Collaboration et al. 2014). Given the differences in these CMB results, as well as our desire for our simulations to span a larger volume of parameter space than current constraints, we combine the chains from *WMAP* and *Planck*. The eigenvalues and eigenvectors of the combined chains are used to set the dimensions of the LH design. The generic LH design has seven]dimensions, with points ranging from [0, 1]. Each of these dimensions is an eigenvector of the cosmological parameter space, and the range [0, 1] maps onto $[-4, 4] \times \sigma_i$, where σ_i is the eigenvalue of vector i. The right-hand panel in Figure 2 shows the generic LH design projected into cosmological parameter space, which now samples the 4σ parameter space allowed by the aforementioned measurements. In this example, we plot Ω_m versus $100\Omega_b$. In this projection, the data points may appear somewhat clumped, but recall that this is an angled projection of the LH. For reference, the 1σ and 2σ contours from the CMB+BAO+SN analyses are presented for both *WMAP* and *Planck*. The sampling of our parameter space that was used in the simulations presented in this paper is shown in Figure 3.

3. *N*-body Simulations

There are three sets of simulations discussed in this work, all run using the L-GADGET2 *N*-body solver, a version of GADGET2 (Springel 2005) modified for memory efficiency when running dark-matter-only simulations. The first of these sets, which we dub "training simulations," is a set of 40 $(1.05\ h^{-1}\ \text{Gpc})^3$ boxes with 1400^3 particles, resulting in a mass resolution of $3.51 \times 10^{10}(\Omega_m/0.3)\ h^{-1}\ M_\odot$. The cosmologies of these simulations, listed in Table 1, are drawn from the

LH discussed in Section 2. These are run with a Plummer equivalent force softening of 20 h^{-1} kpc and a maximum time step of MAX($\Delta \ln a$) = 0.025. We use second-order Lagrangian perturbation theory (2LPT) initial conditions generated at $a = 0.02$ using 2LPTIC (Crocce et al. 2006) with input power spectra as computed by CAMB (Lewis & Bridle 2002), taking $\Omega_\nu = 0$. A slice through one of our simulations is shown in Figure 1.

Each of these 40 simulations is initialized with a different random seed. This is different from the approach taken in some recent simulation suites designed for emulators (e.g., Garrison et al. 2018), but McClintock et al. (2019) and Zhai et al. (2018) showed that this enables our emulators to perform better than the sample variance of our individual simulations, whereas simulations using the same random initial amplitudes and phases are guaranteed to perform only as well as the sample variance of the chosen individual simulation volume. One caveat to this is that methods requiring derivatives of statistics with respect to cosmology may not perform well, due to this choice. Simulations with initial conditions that fix the amplitude of each mode to its expectation value rather than randomly drawing amplitudes from a Rayleigh distribution, as is proper for standard Gaussian initial conditions, have been shown to significantly reduce the variance of many statistics in the linear and semilinear regime without incurring systematic biases (Angulo & Pontzen 2016; Chuang et al. 2018; Villaescusa-Navarro et al. 2018). In the future, we will pursue such methods, but they have not been used for the simulations presented here. We save 10 snapshots at redshifts of $z = \{3.0, 2.0, 1.0, 0.85, 0.7, 0.55, 0.4, 0.25, 0.1, 0.0\}$.

In order to test the accuracy of our emulators, we have also run a set of seven test cosmologies using the same settings as our training simulations. For each test cosmology, we have run five simulations, each with different initial conditions, totaling 35 simulations. We will refer to these as "test simulations" throughout. Five of these cosmologies were chosen to lie along the eigenvector spanning the axis of degeneracy between σ_8 and Ω_m, and the other two were chosen to lie along the second most important eigenvector, perpendicular to the first. The cosmologies of these test simulations are listed in Table 2.

Additionally, we have run a set of simulations varying a number of choices with respect to the L-GADGET2 *N*-body solver, which we will refer to as "convergence test simulations." A few of these simulations were run using a number of
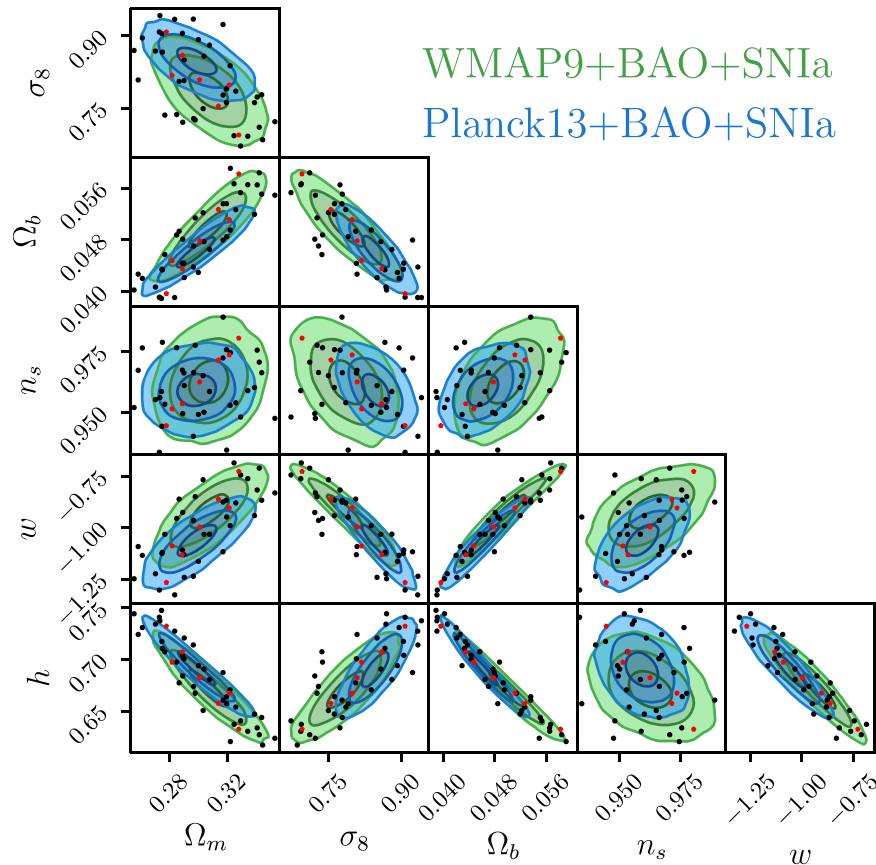
Figure 3. The contours show the 3σ CMB+BAO+SNIa constraints in our parameter space. The 40 training cosmologies and seven test cosmologies are shown in black and red, respectively.

cosmologies, including the Chinchilla cosmology (Lehmann et al. 2017) with $(\Omega_m, h, N_{\mathrm{eff}}, n_s, \sigma_8, w) = (0.286, 0.7, 3.04, 0.96, 0.82, -1)$, which is not used for any of the test or training boxes but is well within our cosmological parameter space. See Table 3 for a summary of the simulations that we have used for these tests.

The names of these simulations all begin with CT to denote that they were run for convergence testing. The first number following CT enumerates the N-body solver parameter set that was used to run the simulation. Various sets of these simulations were run with the same random seed for their initial conditions. The sets with the same seed are demarcated with the same last number in their name, e.g., CT00 and CT60. When necessary, we distinguish between the different cosmologies used by including them in the simulation name. For example, CT00-T00 refers to the simulation run with our fiducial N-body solver parameters using the first random seed for its initial conditions in the T00 cosmology as listed in Table 2.

We have chosen to use volumes of $(400\,h^{-1}\,\mathrm{Mpc})^3$ for the CT simulations rather than the $(1.05\,h^{-1}\,\mathrm{Gpc})^3$ used for the training and test simulations, as changing some of the settings for convergence tests significantly increases the runtime of the simulations. Using a smaller volume allows these simulations to finish in a modest amount of time. We have mitigated the smaller volumes by running four pairs of boxes, CT00, ..., CT03 and CT40, ..., CT43, in three different cosmologies, CHINCH-ILLA, T00, and T04, for our particle-loading test as it is for this

test that we find our largest deviations from convergence, and we wish to constrain these more precisely with better statistics.

We employ the ROCKSTAR spherical overdensity (SO) halo finder (Behroozi et al. 2013) for all of our simulations. ROCKSTAR employs a 6D phase-space FoF algorithm in order to identify density peaks and their surrounding overdensities. We have chosen to use M_{200b} strict SO masses as our fiducial mass definition, where strict refers to the inclusion of unbound particles in the mass estimates of all halos. A discussion of this choice is presented in Section 5. Other than enabling strict SO masses, we have used the default ROCKSTAR settings, choosing the ROCKSTAR softening length to be the same as that used in the N-body solver and leaving on the particle downsampling that ROCKSTAR performs in its initial construction of FoF groups, as we find that this does not affect any of the conclusions presented in this work. Additionally, all results presented here use only host halos, i.e., halos that are not found to lie within a halo with a higher maximum circular velocity.

4. N-body Convergence Tests

The sampling of the parameter space, the effective volume of the training set, and the details of the emulators are all important aspects in determining the final precision and accuracy of our predictions. Equally important is that the observables that are used to train the emulators are converged with respect to all possible choices made when running the simulations. Otherwise, there is a risk of biasing the predictions in ways that are difficult to identify post hoc. For instance, comparison with other predictions is a useful sanity check so long as they agree to

Table 1
The Cosmologies Used in Training Our Emulators, Deemed Training Cosmologies in This Paper

Name	$\Omega_b h^2$	$\Omega_c h^2$	w_0	n_s	$\log 10^{10} A_s$	H_0	N_{eff}
B00	0.0227	0.1141	−0.817	0.9756	3.093	63.37	2.919
B01	0.0225	0.1173	−1.134	0.9765	3.150	73.10	3.174
B02	0.0230	0.1087	−0.685	0.9974	3.094	63.71	3.259
B03	0.0227	0.1123	−0.744	0.9481	3.001	64.04	3.556
B04	0.0221	0.1063	−0.767	0.9651	3.119	65.05	2.664
B05	0.0207	0.1295	−1.326	0.9278	3.024	72.75	2.961
B06	0.0229	0.1115	−0.710	0.9706	3.016	62.70	2.706
B07	0.0228	0.1196	−0.867	0.9663	3.162	64.37	3.939
B08	0.0207	0.1238	−1.164	0.9491	3.147	69.40	3.599
B09	0.0213	0.1158	−0.831	0.9475	3.072	62.36	3.896
B10	0.0219	0.1290	−1.241	0.9610	3.050	72.09	4.236
B11	0.0226	0.1090	−0.861	0.9960	3.158	67.73	2.834
B12	0.0225	0.1168	−0.879	0.9540	3.048	65.38	2.876
B13	0.0219	0.1172	−1.120	0.9788	3.068	71.08	3.004
B14	0.0226	0.1271	−1.117	0.9724	3.094	68.73	2.749
B15	0.0215	0.1285	−1.303	0.9336	3.094	74.10	3.726
B16	0.0218	0.1207	−1.131	0.9662	3.014	70.07	3.769
B17	0.0223	0.1194	−1.248	0.9520	3.035	74.44	3.216
B18	0.0229	0.1157	−1.032	0.9533	3.020	70.75	4.279
B19	0.0224	0.1133	−1.092	0.9673	3.096	72.43	3.684
B20	0.0223	0.1225	−0.990	0.9529	3.120	67.06	3.386
B21	0.0236	0.1172	−0.866	0.9758	3.132	66.39	3.854
B22	0.0215	0.1210	−1.032	0.9586	3.072	68.06	2.621
B23	0.0227	0.1012	−0.566	0.9746	3.019	62.03	3.471
B24	0.0225	0.1103	−0.761	0.9589	3.144	63.03	4.151
B25	0.0209	0.1171	−0.948	0.9345	3.037	65.71	3.089
B26	0.0224	0.1192	−1.125	0.9443	3.128	71.76	2.791
B27	0.0214	0.1134	−0.965	0.9664	3.015	67.39	4.024
B28	0.0217	0.1318	−1.400	0.9586	3.147	74.77	3.811
B29	0.0223	0.1289	−1.236	0.9401	3.159	71.41	3.429
B30	0.0219	0.1239	−1.224	0.9552	3.118	73.43	4.066
B31	0.0212	0.1276	−1.382	0.9561	3.076	73.76	3.344
B32	0.0225	0.1128	−0.926	0.9495	3.043	68.40	3.981
B33	0.0234	0.1150	−0.875	0.9892	3.149	66.05	3.641
B34	0.0228	0.1222	−1.032	0.9500	3.107	69.07	3.131
B35	0.0234	0.1076	−0.613	0.9956	3.140	61.69	3.046
B36	0.0220	0.1213	−1.108	0.9674	3.179	70.41	3.301
B37	0.0229	0.1097	−0.849	0.9776	3.072	66.73	3.514
B38	0.0237	0.1150	−0.955	0.9766	3.054	69.75	4.109
B39	0.0217	0.1201	−0.941	0.9602	3.093	64.70	4.194

Note. Each has one realization with volume $(1050\,h^{-1}\,\text{Mpc})^3$ and $N_{\text{part}} = 1400^3$. Each uses the fiducial settings detailed in Section 3. In particular, they have mass resolutions of $3.51 \times 10^{10}(\Omega_m/0.3)\,h^{-1}\,M_\odot$ and force resolutions of $20\,h^{-1}$ kpc.

Table 2
The Cosmologies Used in the Test Simulations

Name	$\Omega_b h^2$	$\Omega_c h^2$	w_0	n_s	$\log 10^{10} A_s$	H_0	N_{eff}
T00	0.0233	0.1078	−0.727	0.9805	3.039	63.23	2.950
T01	0.0228	0.1128	−0.862	0.9715	3.064	65.73	3.200
T02	0.0223	0.1178	−0.997	0.9625	3.089	68.23	3.450
T03	0.0218	0.1228	−1.132	0.9535	3.114	70.73	3.700
T04	0.0213	0.1278	−1.267	0.9445	3.139	73.23	3.950
T05	0.0218	0.1153	−1.089	0.9514	3.119	69.73	3.700
T06	0.0228	0.1203	−0.904	0.9736	3.059	66.73	3.200

Note. Each has five realizations, each with volume $(1050\,h^{-1}\,\text{Mpc})^3$ and $N_{\text{part}} = 1400^3$ using the fiducial settings detailed in Section 3. In particular, they have mass resolutions of $3.51 \times 10^{10}(\Omega_m/0.3)\,h^{-1}\,M_\odot$ and force resolutions of $20\,h^{-1}$ kpc.

within their purported precision as it is unlikely that both sets of simulations have the same systematic biases, and so their agreement indicates that both predictions are likely converged. However, in the case where such comparisons disagree, it is impossible to determine why unless detailed convergence tests are conducted.

It should be noted again that all of the tests we perform are of relative convergence and not absolute convergence. The

Table 3
Summary of the Boxes Run for Convergence Tests

Name	Cosmology	$N_{\text{realizations}}$	L_{box} (h^{-1} Mpc)	m_{part} (h^{-1} $M_\odot$)	ϵ (h^{-1} kpc)	$\Delta \ln a_{\max}$	a_{start}	α	η
CT0	Chinchilla, T00, T04	3×4	400	$3.30 \times 10^{10}\left(\frac{\Omega_m}{0.286}\right)$	20	0.0250	0.02	0.002	0.0250
CT1	Chinchilla	1	400	3.30×10^{10}	20	0.0250	0.01	0.002	0.0250
CT2	Chinchilla	1	400	3.30×10^{10}	10	0.0250	0.02	0.002	0.0250
CT3	Chinchilla	1	400	3.30×10^{10}	20	0.0250	0.02	0.001	0.0250
CT4	Chinchilla	1	400	3.30×10^{10}	20	0.0125	0.02	0.002	0.0250
CT5	Chinchilla	1	400	3.30×10^{10}	20	0.0250	0.02	0.002	0.0125
CT6	Chinchilla, T00, T04	3×4	400	$4.12 \times 10^{9}\left(\frac{\Omega_m}{0.286}\right)$	20	0.0250	0.02	0.002	0.0250
CT7	T00, ..., T06	7	3000	$2.49 \times 10^{12}\left(\frac{\Omega_m}{0.286}\right)$	20	0.0250	0.02	0.002	0.0250

Note. Columns are simulation name, cosmology, number of different initial condition realizations, box side length, particle mass, Plummer equivalent force softening, maximum time step, starting scale factor, force error tolerance, and time integration error tolerance.

reasons for this are twofold. First, we are knowingly leaving out physics that we believe to be important at some level, such as the effects of baryonic feedback on the matter distribution. Additionally, we do not have an analytic solution toward which we are measuring convergence even for the physics that we have implemented, particularly in the nonlinear regime. There is a growing literature on the possible lack of absolute convergence in N-body simulations in this regime (van den Bosch & Ogiya 2018; van den Bosch et al. 2018), but these issues typically arise when considering dark matter substructure within host halos. Constraining the statistics of substructure is not the goal of the present work, and so we conduct no tests of convergence of such statistics here.

4.1. Measurements

Below we describe our measurements of the following observables:

(a) matter power spectrum, $P(k)$,
(b) three-dimensional matter correlation function, $\xi_{\text{mm}}(r)$,
(c) spherical overdensity HMF, $N(M_{200b})$,
(d) three-dimensional halo–halo correlation function, $\xi_{\text{hh}}(r)$,
(e) projected galaxy–galaxy correlation function, $w_p(r_p)$, and
(f) monopole and quadrupole moments of the redshift space galaxy–galaxy correlation function, $\xi_0(s)$, $\xi_2(s)$.

We briefly detail how we measure each of these statistics and describe our convergence tests in the following subsections.

4.1.1. Matter Power Spectrum

The first statistic we will be interested in is the matter power spectrum, $P(k)$, which is given by

$$\langle \delta(\boldsymbol{k}) \delta(\boldsymbol{k}')^* \rangle = \delta_{\boldsymbol{k},\boldsymbol{k}'} P(k), \tag{1}$$

where $\delta(\boldsymbol{k})$ is the Fourier transform of the matter overdensity field $\delta(\boldsymbol{x}) = (\rho(\boldsymbol{x}) - \overline{\rho})/\overline{\rho}$,

$$\delta(\boldsymbol{x}) = \frac{1}{(2\pi)^3} \int d\boldsymbol{k} \; e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} \; \delta(\boldsymbol{k}), \tag{2}$$

and the angle brackets denote an ensemble average over independent volumes, V. The power spectrum fully describes the statistics of any Gaussian random field and as such is a useful statistic for describing cosmological density fields, which are still Gaussian at most scales until late times.

Because our simulations have periodic boundary conditions, we can estimate the power spectrum using a Fast Fourier Transform (FFT). First, we deposit the density field onto a mesh of dimensions N_{mesh}^3, where $N_{\text{mesh}} = (2L_{\text{box}} k_{\max})/\pi$ using a cloud-in-cell deposition, such that wavenumbers $k \leqslant k_{\max}$ are sampled at or above their Nyquist rate. We take $k_{\max} = 5 \, h \, \text{Mpc}^{-1}$. We then compensate for the mass-deposition window function and average the resulting 3D power spectrum in bins of k, with $dk = L_{\text{box}}/2\pi$. All of these are performed using the PYTHON package NBODYKIT (Hand et al. 2018). We do not perform any shot-noise subtraction, because for the scales we are using, the standard V/N_{part} correction is small, and it is not clear that the correction should necessarily take this form. Unless otherwise noted, this is the only statistic for which we do not include error estimates. At the scales of interest, the errors on $P(k)$ are very small, and estimating them via jackknife as we have done for our other measurements is nontrivial, due to our use of an FFT to measure $P(k)$.

4.1.2. 3D Matter Correlation Function

Because $\delta(\boldsymbol{x})$ is assumed to be a stationary random field, its correlation function is given by the Fourier transform of its power spectrum,

$$\xi(r) = \langle \delta(\boldsymbol{x}) \delta(\boldsymbol{x} + \boldsymbol{r}) \rangle \tag{3}$$

$$= \frac{1}{(2\pi)^3} \int d\boldsymbol{k} \; P(k) \; e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}. \tag{4}$$

We estimate the 3D matter correlation function from our boxes by jackknifing the Landy–Szalay estimator (Landy & Szalay 1993)

$$\hat{\xi}(r) = \frac{DD - 2DR + RR}{RR} \tag{5}$$

where DD, DR, and RR are particle–particle, particle–random, and random–random pair counts normalized by the total number of possible pairs in a given radial bin. We use 27 jackknife regions and downsample the particle distribution by a factor of 100, which we have checked does not affect our results.

Despite the simple relation between the matter power spectrum and 3D correlation function, we check the convergence of both since measurement errors take different forms in the two statistics, e.g., in configuration space, correlations functions are formally only affected by shot noise at $r = 0$,

whereas for power spectra the correction affects all wavenumbers. Additionally, emulators built from these boxes may choose to use one or the other quantity, and determining the scales or wavenumbers where one of these statistics is converged using the other is nontrivial. All pair counting was done using CORRFUNC (Sinha & Garrison 2017).

4.1.3. Spherical Overdensity HMF

In modern theories of ΛCDM galaxy formation, all galaxies are assumed to form within dark matter halos. As such, making converged predictions for the abundance of dark matter halos is of great importance for accurately predicting galaxy statistics. In particular, McClintock et al. (2019) used the simulations presented here to build an emulator for the abundance of SO dark matter halos using $\Delta = 200b$, and so we focus our convergence tests on the statistic used in that work, namely the total number of halos per bin in $\log_{10}(M_{200b})$, $N(M_{200b})$. Additionally, we are interested in only so-called host halos and not subhalos. This is because the galaxy models we employ (in, e.g., Zhai et al. 2018) are based on the Halo Occupation Distribution (HOD) formalism, which has no need for subhalo information, and because the first applications of our HMF emulator will be cosmology constraints using cluster number counts. We estimate $N(M_{200b})$ and its errors in our simulations using a jackknife estimator with 27 jackknife regions.

4.1.4. 3D Halo Correlation Function

The other diagnostic we will use to assess the convergence of our halo populations is the 3D halo correlation function, $\xi_{hh}(r)$. Because the clustering of halos is biased with respect to the matter distribution due to their preferential formation in overdense regions of the matter distribution, convergence of ξ_{hh} at a particular scale, r, does not directly follow from convergence of ξ_{mm} at the same scale, and thus it is important to test for convergence of these separately.

For a discrete field, the two-point correlation function measures the excess probability, relative to a Poisson distribution, of finding two halos at the volume elements dV_1 and dV_2 separated by a distance r (Peebles 1980):

$$dP_{12} = \bar{n}^2[1 + \xi(r)]dV_1 dV_2, \qquad (6)$$

where $\bar{n}$ is the mean number density of the sample. To estimate ξ_{hh}, we again jackknife the Landy–Szalay estimator given by Equation (5), using 27 subvolumes. The measurements of $\xi_{hh}(r)$ presented here are for halos with $M_{200b} > 10^{12} h^{-1} M_{\odot}$, except where otherwise noted.

4.1.5. Galaxy Correlation Functions

In order to calculate galaxy clustering, we employ 10 BOSS massive galaxy sample-like HOD models and populate halos with a galaxy number density of $4.2 \times 10^{-4} (h^{-1} \mathrm{Mpc})^{-3}$ at $z = 0.55$. The typical mass scale M_{min} for the dark matter halo at which half of the halos have a central galaxy is in the range $12.9 < \log M_{min}[h^{-1} M_{\odot}] < 13.5$, and the scatter of halo mass at fixed galaxy luminosity $\sigma_{\log M}$ ranges from 0.05 to 0.5. These models have satellite fractions ranging from 10% to ∼13% and galaxy biases from 2.0 to ∼ 2.13. Satellites are assumed to have a Navarro–Frenk–White profile (Navarro et al. 1996), and their velocity dispersion is assumed to be independent from the position within the halo. More details of this HOD model can be found in Zhai et al. (2018).

The correlation function of the resulting galaxy catalogs is described by the projected correlation function w_p and redshift space multipoles ξ_l. The former is used to mitigate redshift space distortions, probing the real space clustering signal, while the latter can be calculated via Legendre decomposition at a given order l. In the calculation of w_p, the integral along the line-of-sight direction is truncated at $80 h^{-1} \mathrm{Mpc}$, which is large enough to include most of the correlated pairs and produce a stable result. For ξ_l, we perform the decomposition up to $l = 2$ to get the redshift space monopole and quadrupole. Clustering measurements are presented at scales from 0.1 to $50 h^{-1} \mathrm{Mpc}$ and averaged over 10 random realizations of each of the 10 HOD models.

4.2. Convergence Tests

Having described the statistics for which we will check convergence, we now report on the tests that we have performed as well as their results.

4.2.1. Initial Conditions

Since our *N*-body simulations do not start at $a = 0$, we must justify our choice of initial conditions. Two decisions must be made: (1) which analytic prescription to use to generate the initial density and velocity fields and (2) what epoch to generate the initial conditions. For the first, we use 2LPT. For the second, care must be taken to ensure that the analytic treatment used to generate the initial conditions remains accurate for all modes in the simulation until the scale factor at which we start the *N*-body solver. To ensure this, we have chosen a starting time of $a = 0.02$ ($z = 49$). In this section, we show that our observables are robust to this choice by comparing measurements made in a fiducial simulation with the same specifications used in our emulator suite, CT00-CHINCHILLA, to a simulation that has been run with a starting scale factor of $a = 0.01$ ($z = 99$), CT10. It should be noted again that, unless otherwise stated, all convergence tests presented here vary one parameter at a time from the fiducial parameters used for our training and test simulations.

For this test and all following tests, we will only report on deviations from 1% convergence, which exceed 1σ in significance, although we caution that convergence can only be claimed at the 1% level with 68% confidence if the 1σ error bars plotted in each panel fall completely within the gray bands. For this test, a few statistics deviate from convergence by more than this as can be seen in Figure 4. For $M_{200b} < 10^{13} h^{-1} M_{\odot}$, the HMF deviates from convergence. This mass would fall in the lowest mass bin used in McClintock et al. (2019) and is still within the range of halo masses used in HOD models in Zhai et al. (2018). We also find deviations from convergence approaching 1σ in w_p for $\sim r < 300 h^{-1} \mathrm{kpc}$, $P(k)$ for $k \sim 4 h \mathrm{Mpc}^{-1}$ at $z = 0$, and $\xi_{hh}(r)$ for $r \leqslant 1 h^{-1} \mathrm{Mpc}$. The largest scale data point of $\xi_{mm}(r)$ also deviates from 1% convergence for $z = 1$, but this is likely a statistical fluctuation given the convergence of $P(k)$ at large scales.

Possible explanations for the observed deviations from convergence are inaccuracies in 2LPT at these scales in describing nonlinear evolution between $0.01 < a < 0.02$, or, somewhat more likely, this is a manifestation of the discreteness effects discussed in Garrison et al. (2018). If the latter is the case, then the simulation initialized at $a = 0.01$ is actually less accurate than our fiducial simulation as it suffers
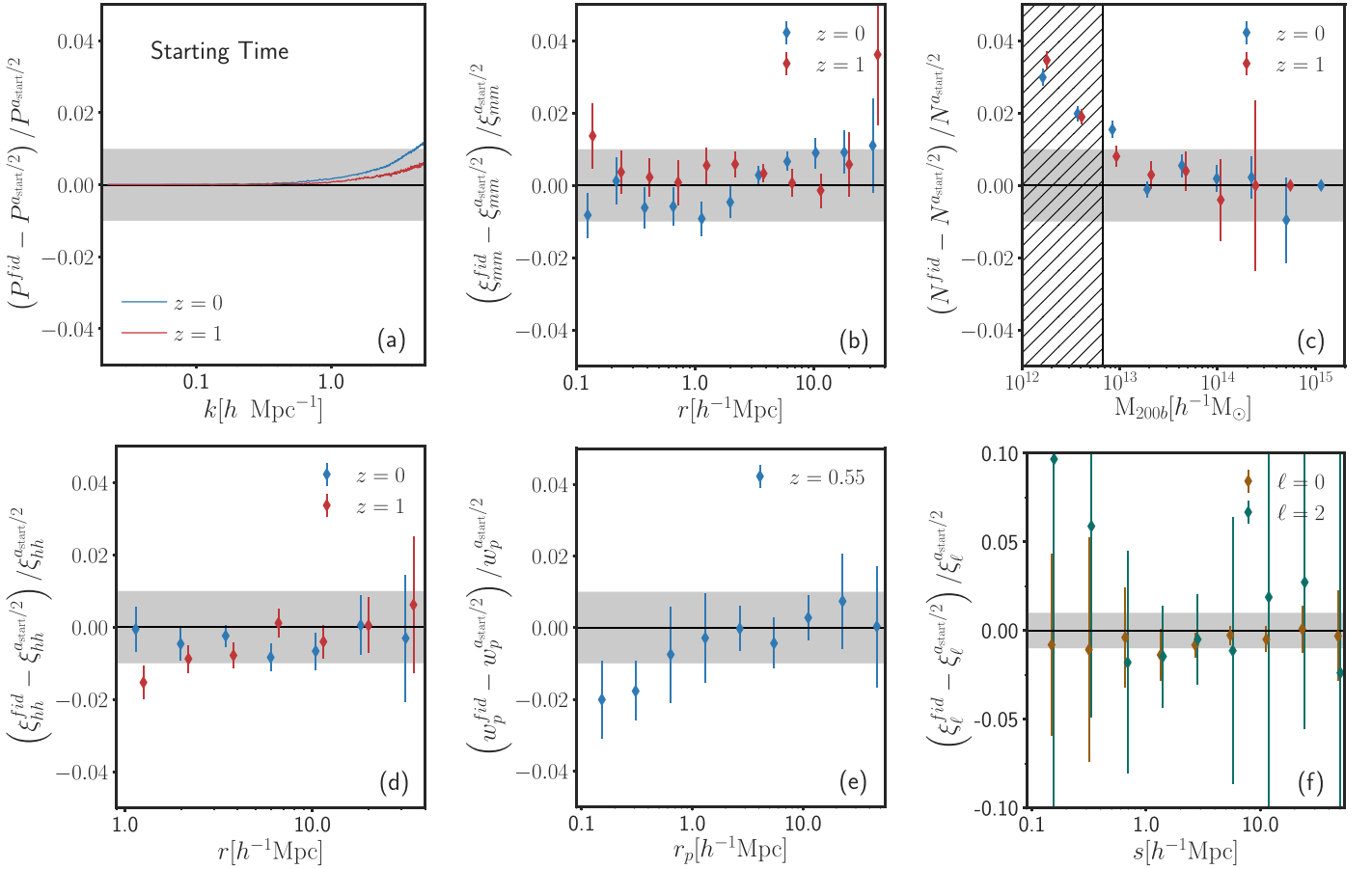
Figure 4. Comparison of a number of observables from CT00-CHINCHILLA, a simulation with our fiducial starting scale factor, $a = 0.02$, and CT10, a simulation with a starting scale factor of $a = 0.01$. The gray band in all figures denotes 1% accuracy. All error bars are calculated via jackknife estimation using the fractional differences in the statistics so as to null sample variance because these simulations are run with the same initial conditions. (a) Matter power spectrum. (b) Matter correlation function. (c) Halo mass function, where the hatched region corresponds to halos with fewer than 200 particles. (d) Halo–halo correlation function for $M_{200b} > 10^{12} \, h^{-1} \, M_\odot$. (e) Galaxy projected correlation function averaged over 10 realizations of 10 different HODs. (f) Redshift space monopole and quadrupole for the same HODs.

from these discreteness effects for a longer period of time. We do not investigate these possibilities further in this work as these inaccuracies are subdominant to the inaccuracies in our simulations, due to particle loading. Future work that we are pursuing with higher resolution simulations will require revisiting these questions.

4.2.2. Force Softening

When solving the N-body problem as an approximation to collisionless dynamics, one must employ a so-called force softening in order to mitigate the effects of unphysical two-body interactions. In L-GADGET2, this is done by representing the single particle density distribution as a Dirac delta function convolved with a spline kernel (Monaghan & Lattanzio 1985) $\delta(\boldsymbol{x}) = W(|\boldsymbol{x}|, 2.8\epsilon)$, where $W(r)$ is given by a cubic spline.

Using this kernel, the potential of a point mass at $r = 0$ for nonperiodic boundary conditions is given by $-Gm/\epsilon$. It is this ϵ that we refer to as the force softening length. Typically, a smaller ϵ yields equations that are closer to those that govern the true universe, but decreasing ϵ by too much at fixed mass resolution will lead to undesirable two-body interactions as mentioned above. There is extensive literature on convergence of various quantities with respect to force softening length (Power et al. 2003, e.g.), but for completeness we investigate

this convergence in the context of the exact statistics that we plan to measure and emulate with this simulation suite.

For our fiducial simulations, we have set $\epsilon = 20 \, h^{-1} \, \mathrm{kpc}$, and for this convergence test we have run an additional simulation, CT20 with $\epsilon = 10 \, h^{-1} \, \mathrm{kpc}$. The results of the comparison between our fiducial simulation, CT00-Chinchilla, and CT20 can be found in Figure 5. The only statistic that deviates from convergence by more than 1%, or sample variance, is $\xi_{mm}(r)$ for $r < 200 \, h^{-1} \, \mathrm{kpc}$. $P(k)$ is converged to the 1% level for the scales measured here, but by $k \sim 3 \, h \, \mathrm{Mpc}^{-1}$ is showing systematic deviations from perfect convergence. This is consistent with the findings of Heitmann et al. (2010), although they only consider $\epsilon > 25 \, h^{-1} \, \mathrm{kpc}$. The deviations seen in ξ_{mm} and $P(k)$ do not appear to have a significant effect on other statistics.

4.2.3. Absolute Force Error Tolerance

Another parameter that governs the accuracy of the gravitational force calculations is how deeply to walk the octree used to partition space when summing the small-scale contributions to the gravitational force on each particle. This is typically referred to as the cell-opening criterion, as it is used to determine whether or not a cell in the tree should be "opened" and traversed. We use the standard L-GADGET2 relative opening criterion, which opens a cell containing mass M,
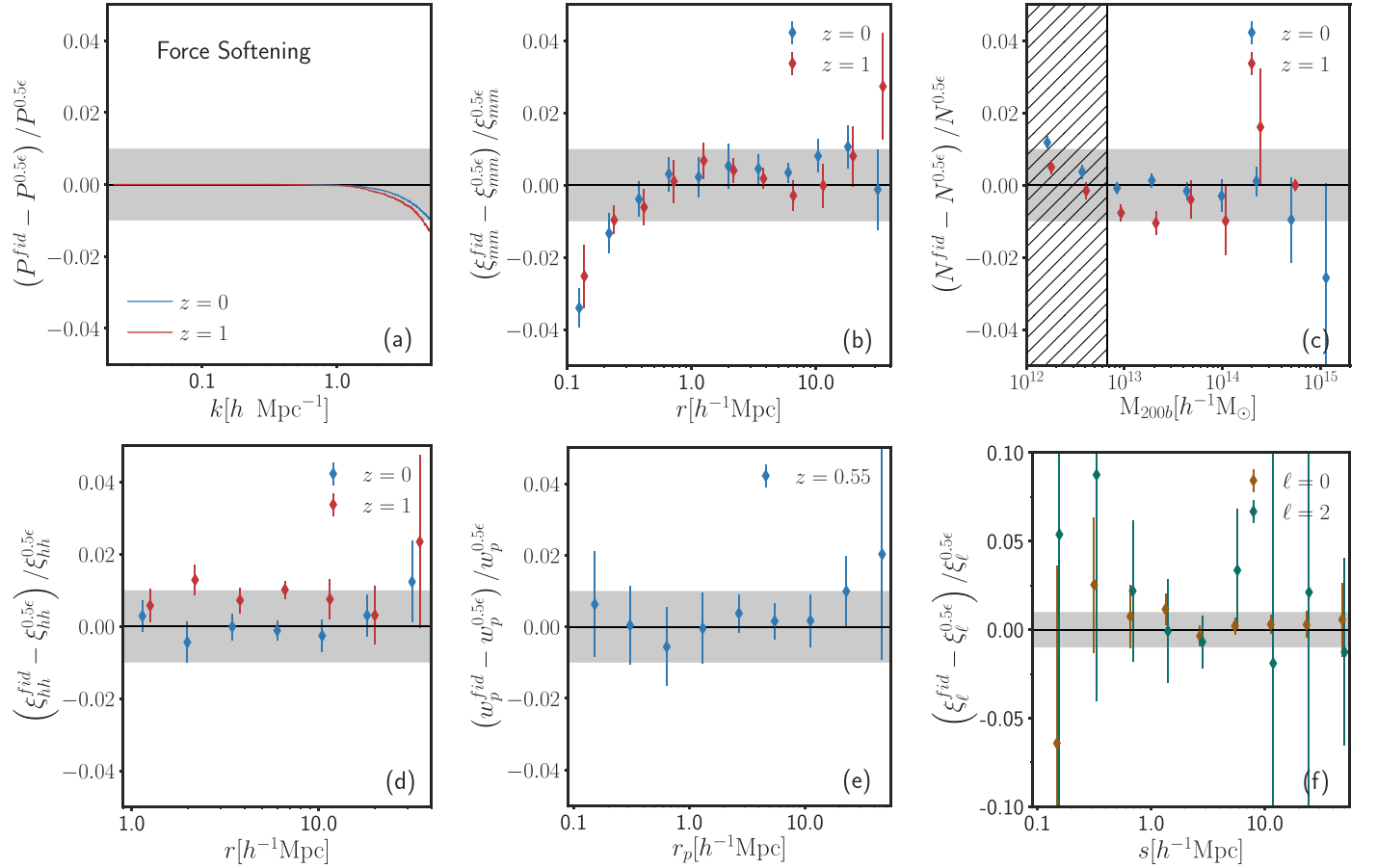
Figure 5. Convergence tests with respect to force softening. Observables measured from a simulation with our fiducial parameters, CT00-CHINCHILLA, are compared to CT20, a simulation with half the force softening: $\epsilon = 10 \ h^{-1}$ kpc. Subpanels are the same as in Figure 4.

extension l at a distance from the point under consideration of r if

$$\frac{GM^2}{r^2}\left(\frac{l}{r}\right)^2 > \alpha |\boldsymbol{a}_{\mathrm{old}}|, \qquad (7)$$

where $|\boldsymbol{a}_{\mathrm{old}}|$ is the magnitude of the acceleration of the particle under consideration in the last time step, and α is a free parameter allowing the tuning of the accuracy. In general, a decreasing α leads to smaller errors in force computation but greater runtime as more nodes in the tree must be opened per time step. Our fiducial runs use $\alpha = 0.002$.

In order to test that our results are converged with respect to this choice, we have run an additional simulation, CT30, with $\alpha = 0.001$. We find no significant deviations from convergence, as can be seen in Figure 6.

4.2.4. Time Stepping

Another significant choice that must be made in the L-GADGET2 algorithm is the maximum allowed time step. The time step for the leapfrog integrator that L-GADGET2 uses is determined by $\Delta \ln(a) = \min[\Delta \ln(a)_{\mathrm{max}}, \sqrt{2\eta\epsilon / |a|}\,]$, where η is the free parameter determining integration error tolerance. Typically, $\Delta\ln(a)_{\mathrm{max}}$ sets the time step at early times when densities are low, and the $\sqrt{2\eta\epsilon / |a|}$ criterion sets the time step in collapsed regions at late times and thus is important in dictating the convergence of halo density profiles (Power et al. 2003).

We have run additional simulations in order to check convergence with respect to time-stepping criteria. In CT40, $\Delta\ln(a)_{\mathrm{max}} = 0.0125$, and in CT50, $\eta = 0.0125$, half of their respective values for our fiducial simulation, CT00. Comparisons of the same measurements detailed in Section 4.1 between CT00-CHINCHILLA and CT40 are shown in Figure 7. No significant deviations are found. The same comparisons were made between CT00-CHINCHILLA and CT50 and were found to be nearly identical, and so we have not included them for conciseness.

4.2.5. Particle Loading

The *N*-body algorithm solves for the evolution of a discretization of the phase-space distribution function of dark matter. Because this phase-space distribution is fundamentally continuous, at least on macroscopic scales, an important parameter governing the accuracy of the algorithm is the number of particles used to sample this distribution function. For the following tests, we have run a set of simulations, CT60, …, CT63, in three different cosmologies, CHINCHILLA, T00, and T04, where we have doubled the number of points with which we sample each spatial dimension, increasing the mass resolution by a factor of 8 from our fiducial settings. Results for the comparison of these simulations with our fiducial set in the Chinchilla cosmology can be found in Figure 8.

The mass function is converged to within 1%, or statistical precision, whichever is larger, for halos that are resolved with 500 particles or more. For masses below this, we observe
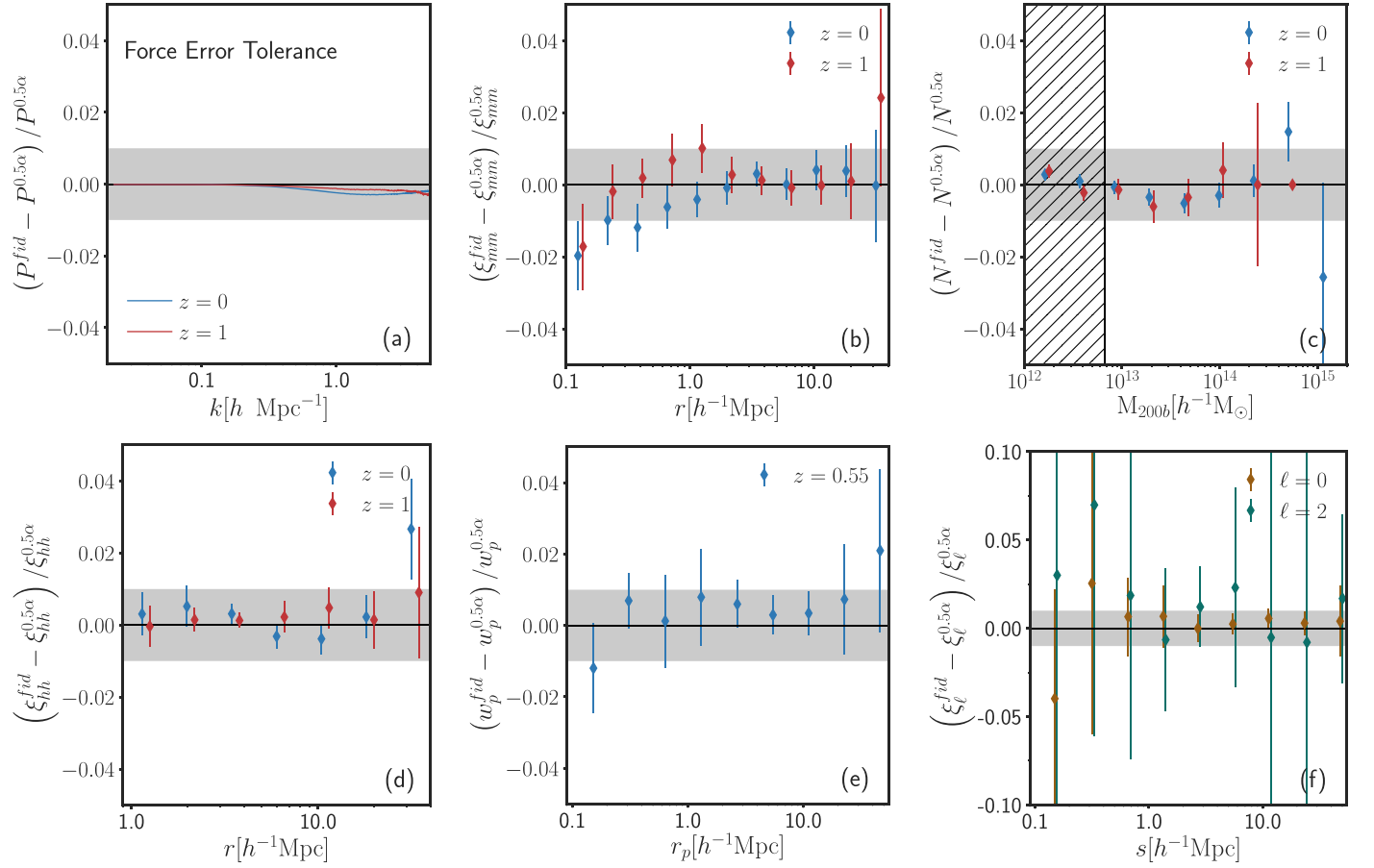
Figure 6. Convergence tests with respect to force error tolerance. Observables measured from a simulation with our fiducial parameters, CT00-CHINCHILLA are compared to CT30, a simulation with half the force error tolerance: $\alpha = 0.002$. Subfigures are the same as in Figure 4.

varying degrees of deviation from convergence, which depend to good approximation on just the number of particles that the halo is resolved with. This can be seen in Figure 9, which demonstrates that bins in particle number show similar behavior for all redshifts except for very low particle numbers at high redshift. Only one cosmology is plotted, but a similar trend holds in the other two cosmologies, despite the three cosmologies spanning a large range in σ_8 and Ω_m. We have fit the following function to the average of these residuals over redshift in order to characterize and correct for them in other works:

$$\frac{N(M_{200b}^{fid}) - N(M_{200b}^{m_{part}/8})}{N(M_{200b}^{m_{part}/8})} = -\exp\frac{-(\log_{10} N_{part} - \log_{10} N_0)}{\sigma_{\log_{10} N}},$$

(8)

where N_{part} is the number of particles corresponding to M_{200b}^{fid}, the halo mass measured in our fiducial simulations. We find $\log_{10} N_0 = 0.25 \pm 0.13$ and $\sigma_{\log_{10} N} = 0.557 \pm 0.046$.

To higher order, the deviations from convergence appear to be dependent on the local logarithmic slope of the mass function, $\Gamma = d\log_{10} N^{fid}/d\log_{10} M_{200b}$, with the worst deviations occurring at low particle number and very steep slopes. This can be seen in Figure 10. Here, we have measured the deviations of our fiducial simulations from convergence as a function of particle number and Γ, where Γ is determined by fitting a quartic spline to $N(M_{200b})$ in the CT0 simulations at all redshifts and taking its logarithmic derivative. We have also

interpolated these measurements in order to make the trends more obvious. Above about 1000 particles, the deviations from convergence of the mass function are less than 1% for all slopes. Below this particle number, there is a trend in error with Γ, leading to the larger errors seen at high redshift in Figure 9. For these reasons, we caution against using the correction as determined above for halos with particle numbers less than 1000 when $\Gamma < -2$.

The deficit of halos that we find in our fiducial simulations compared to the CT6 simulations cannot be explained by the increased Poisson random noise in the mass estimates, as this would lead to an overabundance of halos at a given mass, due to the negative slope of the mass function in a manner analogous to the Eddington bias. Instead, the observed deficit suggests that a bias is being introduced in the density field, which is clear from the deviations observed in $P(k)$, such that low-mass halos are less likely to form in lower resolution simulations.

These errors also propagate into other observables involving halo mass. For instance, ξ_{hh} deviates from convergence by 7.5% when using all halos with $M_{200b} > 10^{12} h^{-1} M_\odot$, but quickly converges as a function of mass as can be seen by the fact that halos with $M_{200b} > 10^{12.5} h^{-1} M_\odot$ only deviate by 3% from the simulations with higher particle loading. Mass cuts above this have noisy ξ_{hh} measurements, and so we cannot make precise statements about their convergence. The galaxy correlation functions are less sensitive to mass resolution at the low-mass end because our HODs are tuned to match the BOSS
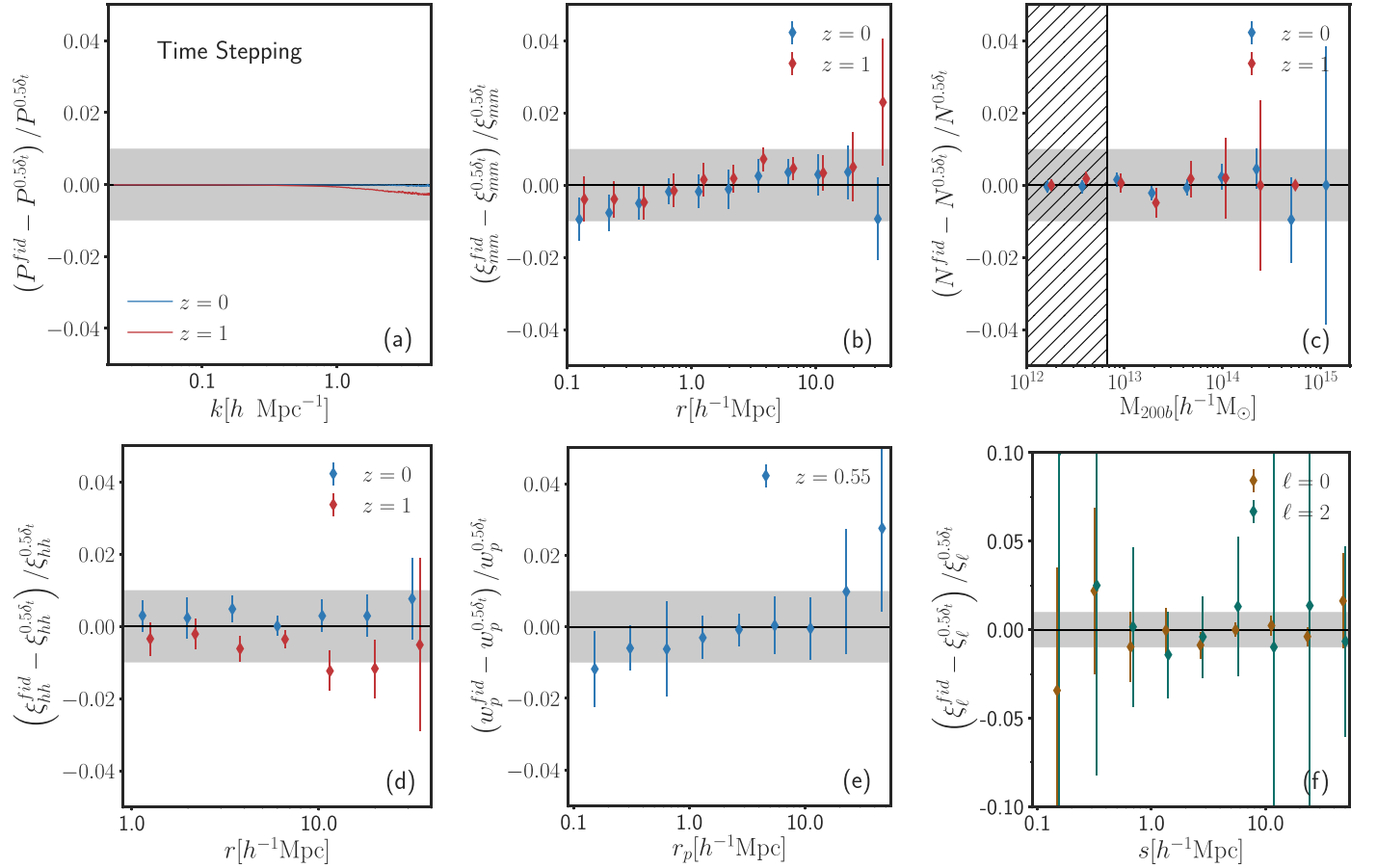
Figure 7. Convergence tests with respect to maximum time step. Observables measured from a simulation with our fiducial parameters, CT00-CHINCHILLA are compared to CT40, a simulation with half the maximum time step: $\Delta \ln(a)_{\mathrm{max}} = 0.0125$. Subfigures are the same as in Figure 4.

massive galaxy sample. This can be seen in Figures 8(e) and (f), where w_p is converged at the 1%–2% level, with the redshift space measurements performing only slightly worse. ξ_{mm} is converged, while $P(k)$ deviates from convergence for $z = 0$ above $k \sim 1.5 \, h \, \mathrm{Mpc}^{-1}$, with a maximal deviation of about 2% at $k \sim 3 \, h \, \mathrm{Mpc}^{-1}$. The deviations from convergence for $P(k)$ are consistent with those found in Schneider et al. (2016): $\sim 1\%$ deviations from convergence for $P(k)$ at $k \sim 1 h \, \mathrm{Mpc}^{-1}$ for an $L_{\mathrm{box}} = 512$, $N_{\mathrm{part}} = 512^3$ simulation.

4.2.6. Finite Box Effects

It is currently beyond the realm of possibility to simulate the entire observable universe at high-enough resolution to be useful. Instead, the common practice in cosmological simulations is to assume periodic boundary conditions with a fundamental mode that is much larger than the scales of interest for the problem at hand. One effect of doing this is that modes larger than the fundamental mode of the box are not included in the growth of structure. Because gravitational collapse is a nonlinear process, the growth of small-scale structure couples to large-scale growth, and thus missing large variances can cause inaccuracies and alter sample variance at smaller scales. Additionally, because our simulations are periodic, only discrete modes, $\mathbf{k} = 2\pi(i, j, k)/L_{\mathrm{box}}$, where $i, j, k \in \mathbb{Z}$, are included in the initial conditions. In order to test the effects of these approximations, we have run a set of much larger, lower resolution simulations, CT70, …, CT76, at the same

cosmologies as our test simulations, where we have $(5 \, h^{-1} \, \mathrm{Gpc})^3$ for each cosmology to compare with. The results of the comparison for the T04 cosmology are shown in Figure 11; the other cosmologies show nearly identical results.

Because the CT7 simulations have worse mass resolution than our test simulations, the analysis in Section 4.2.5 indicates that there should be residual effects in this comparison, due to mass resolution. In order to mitigate the differences arising from mass resolution, we have applied the correction in Equation (8) to both sets of measurements. We find convergence to within sample variance of the test boxes for all masses at both $z = 0$ and $z = 1$, although this is significantly larger than the percent level at $z = 1$ for all masses shown here. We also compared $\xi_{\mathrm{mm}}(r)$ and found no deviations from convergence for $r < 100 \, h^{-1} \, \mathrm{Mpc}$.

5. Halo Finding

In this section, we discuss the sensitivity of our results to choices made with regard to halo finding. We have defined dark matter halos as spherical structures with overdensities of 200 times the background density. This choice is relatively arbitrary, having its basis in simple spherical collapse models that have been shown to be imprecise compared to modern cosmological standards. As such, we discuss the possible impacts that this definition might have on cosmological results obtained from emulators built upon these simulations. Note that we consider the choice of halo finder and the settings used in
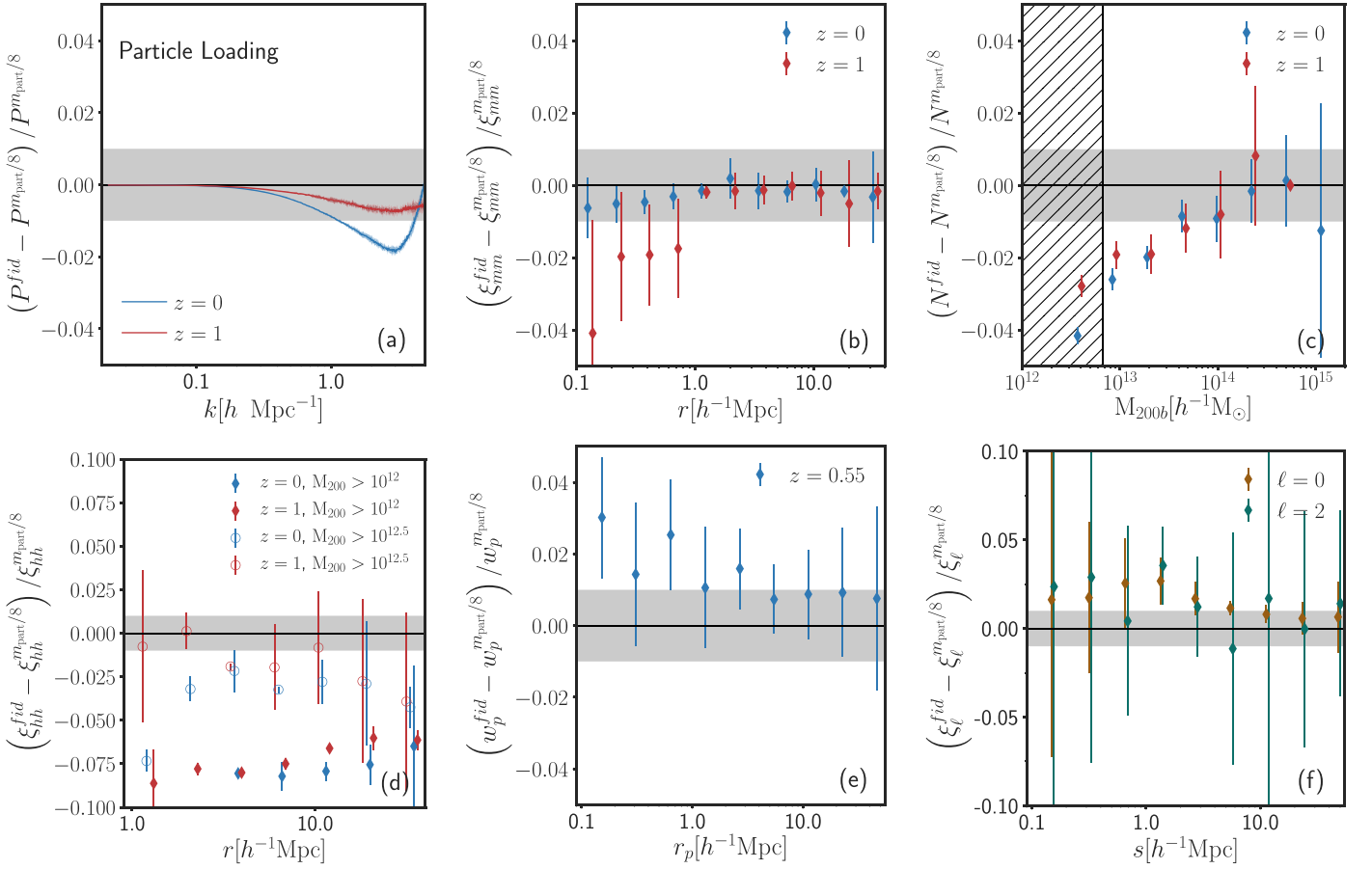
Figure 8. Convergence tests with respect to mass resolution. Subpanels are the same as in Figure 4.
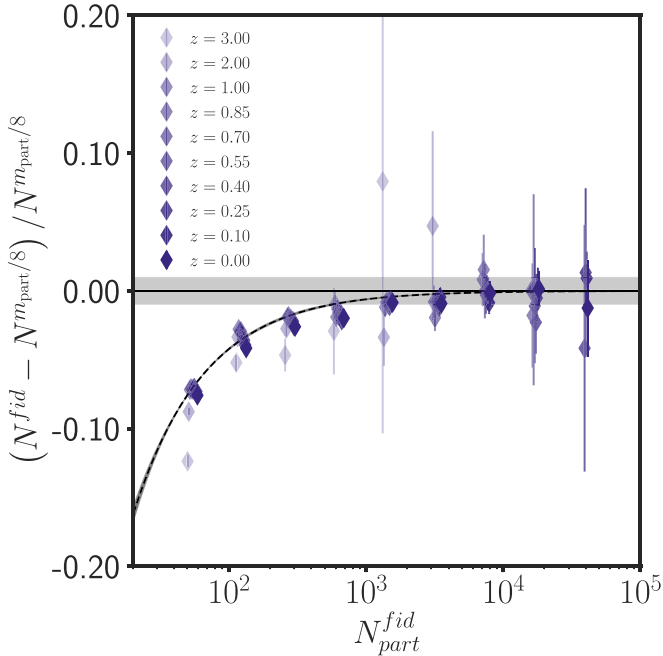

Figure 9. Deviations of the mass functions measured in simulations using our fiducial parameters from simulations with higher mass resolution as a function of redshift. The line is a fit to all of these points in addition to the points for the other two cosmologies that are not shown in this figure.

that halo finder to be part of the mass definition, and as such we do not consider the effects of different halo finders separately.

In the case of using galaxy clustering to constrain cosmology, there is extensive literature on how the choice of halo definition can impact and possibly bias the inferred cosmology. Much of this literature has focused on the effect of secondary parameters on the clustering signals of halos at fixed mass (e.g., Gao et al. 2005; Wechsler et al. 2006; Chue et al. 2018; Mao et al. 2018). This effect propagates differently into galaxy clustering depending on which proxy is then used to assign galaxies to halos (Reddick et al. 2013; Lehmann et al. 2017) and can lead to biases in inferred HOD parameters when ignored (Zentner et al. 2014). Whether these effects lead to biases in inferred cosmology when using HODs and whether these biases can be mitigated through extensions to the HOD model are still open questions.

In the case of the HMF, the situation is equally complicated. We do not directly measure the HMF, unlike the galaxy correlation function, but rather some distribution of observables such as cluster richness or X-ray temperature. In order to constrain cosmology, a mass–observable relation (MOR) must be obtained, and the calibration of this MOR must also assume a halo mass definition. It is imperative that the definition used when constraining the MOR and the definition used for the HMF be the same in order to obtain unbiased cosmological constraints. If the scatter in the MOR is smaller for a particular
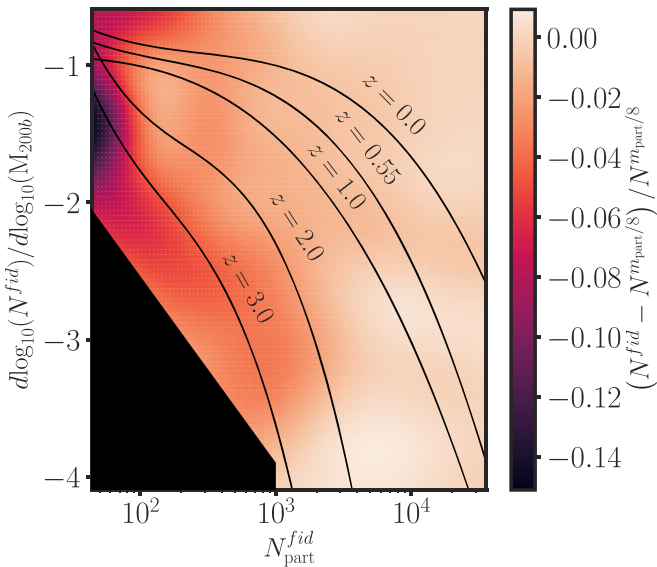
Figure 10. Deviations of the mass functions measured in simulations using our fiducial parameters from simulations with higher mass resolution as a function of particle number and logarithmic slope of the mass function. A clear trend can be seen with the logarithmic slope of the mass function for $N_{part} < 1000$. Black lines show the logarithmic slopes of mass functions measured in the CT00-T00 simulation for different redshifts. The solid black region is beyond where we have any data and so we exclude it from this plot.
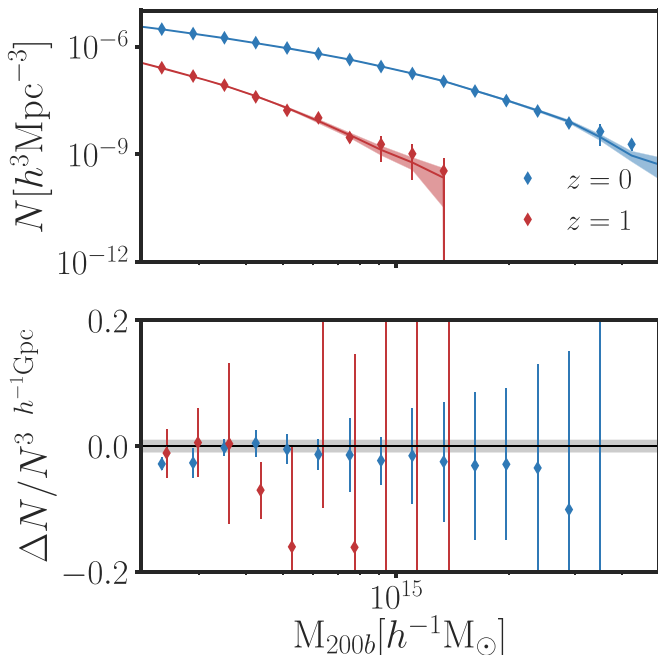
Figure 11. Comparison of the average mass function over the five realizations of T04 with that measured in CT7-T04. The top panel shows the mass functions, where the points are measured from T04 and the lines are measurements from CT72-T04. The bottom panel shows the fractional difference of the mass functions between these simulations. These measurements are consistent with no finite box effects.

mass definition, that definition will yield tighter constraints, but a study of the halo mass definition that minimizes scatter in the MOR is beyond the scope of this paper. A more practical reason for our choice of $\Delta = 200b$ is that it makes the typical radii of cluster mass halos, ~ 0.5–$2\,h^{-1}$ Mpc, significantly larger than the force softening lengths used in our simulations.
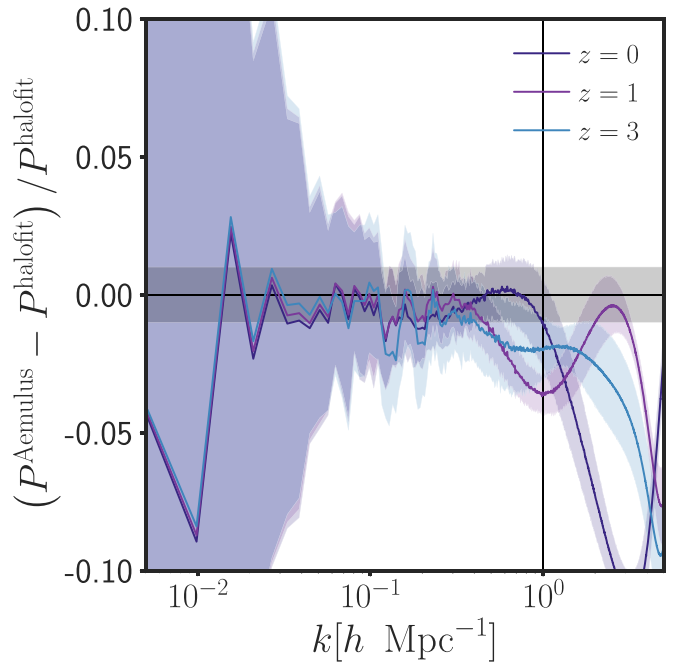

Figure 12. Comparison of the 40 simulation boxes to the Takahashi HALOFIT. The vertical black line demarcates the wavenumber where we expect the effects of mass resolution to become important at the $>1\%$ level. Agreement is to within the reported HALOFIT accuracy.

6. Comparison with Other Emulators and Linear Theory

Having internally validated our simulations, we now compare our measurements to those obtained in other works. Unfortunately, the most precise determination of the matter power spectrum available to date, the Mira–Titan universe emulator (Heitmann et al. 2016; Lawrence et al. 2017), does not cover the same parameter space as our simulations. In particular, they do not include N_{eff} in their parameter space, and varying this can lead to deviations in $P(k)$ on the order of $\sim 10\%$, much greater than the precision at which such a comparison would be relevant. Instead, we have compared our simulations to predictions from the widely used HALOFIT algorithm (Smith et al. 2003; Takahashi et al. 2012), which does span our parameter space. Our simulations are not large enough in volume for precision emulation of the matter power spectrum, but nevertheless we can compare our measured matter power spectra to the HALOFIT predictions for our cosmologies as both an external validation of our simulations and as a further consistency check for the HALOFIT algorithm.

The results of this comparison can be found in Figure 12. Error bars in this figure correspond to the variance of the deviations of our 40 training simulations from HALOFIT. We find better than 1% agreement in the mean deviation until $k \sim 0.3\,h\,\mathrm{Mpc}^{-1}$, but observe maximum errors close to 5%, consistent with the HALOFIT internal error estimation in Takahashi et al. (2012). For scales smaller than $k = 1\,h\,\mathrm{Mpc}^{-1}$, we find large deviations of up to 12%, which are likely due to a combination of inaccuracies in HALOFIT and resolution effects in our simulations. The maximum errors that we observe for $0.1 < k < 1$ are slightly smaller than those reported in Heitmann et al. (2014), but this may be attributable to the differences in the parameter spaces spanned by the two sets of simulations. In future work, we will construct our own
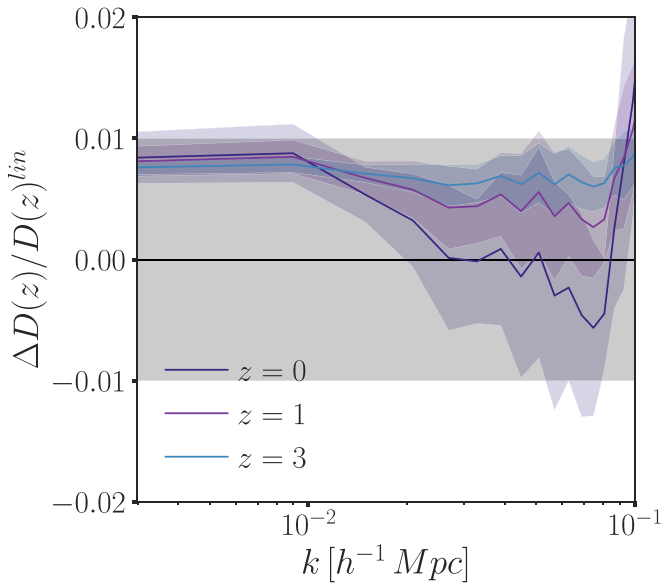
Figure 13. Comparison of the growth factor of our 40 training simulation boxes to that predicted by linear theory. Errors are the standard deviation of this measurement computed over all 40 simulations. We find a discrepancy of 1% on the largest scales measured in our simulations that is nearly independent of cosmology and redshift, suggesting that our simulations reproduce linear theory at this level.

emulator for the matter power spectrum in order to facilitate a direct comparison with the Mira–Titan emulator.

In Figure 13, we test our simulations' ability to reproduce linear growth on large scales. To do this, we measure $P(k)$ in the initial conditions of each simulation and at $z = 0$, 1, and 3 for each of the 40 training simulations. We then take the ratio $P(k, z)/P(k, z = 49)$ and find the fractional difference between this and $D(z)/D(z = 49)$, where $D(z)$ is the linear growth factor for each cosmology at redshift z. Because our initial conditions reproduce linear theory by definition, on large scales this fractional difference equals the accuracy to which our N-body solver reproduces linear theory. As shown in Figure 13, we find about a 1% discrepancy between our simulations and linear theory that is relatively independent of redshift and cosmology. The error bars in that figure are the variance of this measurement over our 40 simulations. There is a mild evolution to larger discrepancies at lower redshifts, which is significant with respect to the standard error on the mean of this quantity averaged over all of our simulations.

7. Data Release

Upon publication of this article, we are making the simulations described here available upon request. This includes the initial conditions, the particle snapshots and halo catalogs at all 10 redshifts described in Section 3, and any measurements used in this paper or McClintock et al. (2019) and Zhai et al. (2018).

We will make the aforementioned data products freely downloadable at https://AemulusProject.github.io at the time this study and its companion papers are published.

8. Discussion and Conclusions

We have presented a new suite of N-body simulations for emulating cosmological observables. The cosmologies of these simulations were sampled from the wCDM 4σ allowed CMB

+BAO+SN parameter space using an orthogonal LH. We investigated the convergence of the following observables with respect to choices made in the L-GADGET2 N-body solver:

matter power spectrum, $P(k)$,
three-dimensional matter correlation function, $\xi_{mm}(r)$,
spherical overdensity HMF, $N(M_{200b})$
three-dimensional halo–halo correlation function, $\xi_{hh}(r)$,
projected galaxy–galaxy correlation function, $w_p(r_p)$, and
monopole and quadrupole moments of the redshift space galaxy–galaxy correlation function, $\xi_0(s)$, $\xi_2(s)$.

We conclude that our observables are converged to 1% accuracy or to the sample variance of our measurements, whichever is larger, with respect to choices made in time stepping and force resolution. Choices with respect to initial conditions lead to minor deviations from 1% convergence for halos resolved with fewer than 200 particles, although further investigation of the source of this discrepancy will be investigated in future work. Our choice of force softening leads to deviations from 1% convergence for scales $r < 200 \, h^{-1}$ kpc.

Particle loading is by far the parameter to which our observables are most sensitive. For halos with greater than 500 particles, we also find convergence at better than the 1% level, or sample variance, but for masses smaller than this, deviations from convergence due to insufficient particle loading increase rapidly. Halos with more than 200 particles, like those used in McClintock et al. (2019), are still converged to better than 2.5%, or sample variance. We have shown that this deviation is largely a function of particle number alone, and have fit this dependence and applied it to build the emulator in McClintock et al. (2019). Additional tests in that study using even higher resolution simulations provide more evidence that this correction is satisfactory for our needs.

We have shown that our HMF predictions are not affected by finite box effects to the precision allowed by sample variance in our test boxes. At $z = 0$, sample variance is smaller than 1% for $M_{200m} \lesssim 4 \times 10^{14} \, h^{-1} \, M_\odot$. The study of this effect in the current work is limited by our inability to use the same initial conditions for the different box sizes necessary for this test, as was done for the rest of the internal convergence tests detailed in this work. As such, these tests are limited by the sample variance in our test boxes, which is greater than percent level for $M_\odot \geqslant 4 \times 10^{14} \, M_\odot$ and $z > 0$. Future efforts are required to ensure that observables are indeed converged with respect to simulation size to the level needed for upcoming surveys.

The matter power spectra in our simulations are consistent with those predicted by the HALOFIT methodology to within their reported errors, but we are unable to compare to the Mira–Titan emulator as our simulations use a different cosmological parameter space. We also show that our simulations reproduce linear growth on the largest scales available in our simulations to 1% accuracy. Tests of this nature are of the utmost importance, and continued work in making them is vital to ensure that emulators of this kind are put to their full use in upcoming analyses.

The work presented here is just the beginning of our effort to contribute high-precision and -accuracy simulations and emulators to the community. Future work will extend our simulation suite significantly, especially with higher resolution simulations that are suited for use with more complete galaxy formation models. Additionally, we plan to expand our

parameter space by including more physics such as neutrino masses, and by expanding the limits of the parameters sampled to include more volume away from CMB+BAO+SN constraints. This is important, as upcoming analyses will attempt to diagnose the tension between different data sets in addition to combining constraints from many different experiments.

The sharing of resources between simulators and the exchange of expertise between simulators, theorists, and observers will be vital in attaining the best possible outcomes for the next generation of surveys. Only a concerted effort from many groups in the domain of cosmic emulation over the next decade will help ensure that stage IV cosmological surveys are not limited by modeling systematics.

ORCID iDs

Joseph DeRose ⓘ https://orcid.org/0000-0002-0728-0960
Risa H. Wechsler ⓘ https://orcid.org/0000-0003-2229-011X
Matthew R. Becker ⓘ https://orcid.org/0000-0001-7774-2246
Yao-Yuan Mao ⓘ https://orcid.org/0000-0002-1200-0820
Sean McLaughlin ⓘ https://orcid.org/0000-0002-7042-6424
Zhongxu Zhai ⓘ https://orcid.org/0000-0001-7984-5476

References

Abbott, T. M. C., Abdalla, F. B., Alarcon, A., et al. 2018, PhRvD, 98, 043526
Alam, S., Ata, M., Bailey, S., et al. 2017, MNRAS, 470, 2617
Anderson, L., Aubourg, É., Bailey, S., et al. 2014, MNRAS, 441, 24
Angulo, R. E., & Pontzen, A. 2016, MNRAS, 462, L1
Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013, ApJ, 762, 109
Blas, D., Lesgourgues, J., & Tram, T. 2011, JCAP, 7, 034
Cataneo, M., Foreman, S., & Senatore, L. 2017, JCAP, 4, 026
Chisari, N. E., Alonso, D., Krause, E., et al. 2018, arXiv:1812.05995
Chuang, C.-H., Yepes, G., Kitaura, F.-S., et al. 2018, arXiv:1811.02111
Chue, C. Y. R., Dalal, N., & White, M. 2018, JCAP, 10, 012
Crocce, M., Pueblas, S., & Scoccimarro, R. 2006, MNRAS, 373, 369
Gao, L., Springel, V., & White, S. D. M. 2005, MNRAS, 363, L66
Garrison, L. H., Eisenstein, D. J., Ferrer, D., et al. 2018, ApJS, 236, 43
Hand, N., Feng, Y., Beutler, F., et al. 2018, AJ, 156, 160
Heitmann, K., Bingham, D., Lawrence, E., et al. 2016, ApJ, 820, 108
Heitmann, K., Higdon, D., White, M., et al. 2009, ApJ, 705, 156
Heitmann, K., Lawrence, E., Kwan, J., Habib, S., & Higdon, D. 2014, ApJ, 780, 111
Heitmann, K., White, M., Wagner, C., Habib, S., & Higdon, D. 2010, ApJ, 715, 104
Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, ApJS, 208, 19
Hunter, J. D. 2007, CSE, 9, 90
Joudaki, S., Blake, C., Johnson, A., et al. 2018, MNRAS, 474, 4894
Kaehler, R., Hahn, O., & Abel, T. 2012, arXiv:1208.3206
Kitaura, F.-S., Rodríguez-Torres, S., Chuang, C.-H., et al. 2016, MNRAS, 456, 4156
Landy, S. D., & Szalay, A. S. 1993, ApJ, 412, 64
Lawrence, E., Heitmann, K., Kwan, J., et al. 2017, ApJ, 847, 50
Lehmann, B. V., Mao, Y.-Y., Becker, M. R., Skillman, S. W., & Wechsler, R. H. 2017, ApJ, 834, 37
Lewis, A., & Bridle, S. 2002, PhRvD, 66, 103511
MacCrann, N., DeRose, J., Wechsler, R. H., et al. 2018, MNRAS, 480, 4614
Mao, Y.-Y., Zentner, A. R., & Wechsler, R. H. 2018, MNRAS, 474, 5143
McClintock, T., Rozo, E., Becker, M., et al. 2019, ApJ, 872, 53
Monaghan, J. J., & Lattanzio, J. C. 1985, A&A, 149, 135
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563
Oliphant, T. E. 2007, CSE, 9, 10
Peebles, P. J. E. 1980, The Large-scale Structure of the Universe (Princeton, NJ: Princeton Univ. Press)
Perko, A., Senatore, L., Jennings, E., & Wechsler, R. H. 2016, arXiv:1610.09321
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, A&A, 571, A16
Power, C., Navarro, J. F., Jenkins, A., et al. 2003, MNRAS, 338, 14
Press, W. H., & Schechter, P. 1974, ApJ, 187, 425
Reddick, R. M., Wechsler, R. H., Tinker, J. L., & Behroozi, P. S. 2013, ApJ, 771, 30
Schneider, A., Teyssier, R., Potter, D., et al. 2016, JCAP, 4, 047
Sheth, R. K., & Tormen, G. 1999, MNRAS, 308, 119
Sinha, M., & Garrison, L. 2017, Corrfunc: blazing fast correlation functions on the CPU, Astrophysics Source Code Library, ascl:1703.003
Skillman, S. W., Warren, M. S., Turk, M. J., et al. 2014, arXiv:1407.2600
Smith, R. E., Peacock, J. A., Jenkins, A., et al. 2003, MNRAS, 341, 1311
Springel, V. 2005, MNRAS, 364, 1105
Takahashi, R., Sato, M., Nishimichi, T., Taruya, A., & Oguri, M. 2012, ApJ, 761, 152
Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, ApJ, 688, 709
Tinker, J. L., Sheldon, E. S., Wechsler, R. H., et al. 2012, ApJ, 745, 16
van den Bosch, F. C., & Ogiya, G. 2018, MNRAS, 475, 4066
van den Bosch, F. C., Ogiya, G., Hahn, O., & Burkert, A. 2018, MNRAS, 474, 3043
van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, CSE, 13, 22
Villaescusa-Navarro, F., Naess, S., Genel, S., et al. 2018, ApJ, 867, 137
Wechsler, R. H., Zentner, A. R., Bullock, J. S., Kravtsov, A. V., & Allgood, B. 2006, ApJ, 652, 71
Wibking, B. D., Salcedo, A. N., Weinberg, D. H., et al. 2019, MNRAS, 484, 989
Zentner, A. R., Hearin, A. P., & van den Bosch, F. C. 2014, MNRAS, 443, 3044
Zhai, Z., Tinker, J., Becker, M. R., et al. 2018, ApJ, 874, 95