COSMOLOGICAL CONSTRAINTS FROM GALAXY CLUSTERING AND THE MASS-TO-NUMBER RATIO OF GALAXY CLUSTERS: MARGINALIZING OVER THE PHYSICS OF GALAXY FORMATION

Rachel M. Reddick\textsuperscript{1,2}, Jeremy L. Tinker\textsuperscript{1}, Risa H. Wechsler\textsuperscript{1,2}, and Yu Lu\textsuperscript{1,2}

\textsuperscript{1}Kavli Institute for Particle Astrophysics and Cosmology, Physics Department, Stanford University, Stanford, CA 94305, USA; rmredd@stanford.edu, rwechsler@stanford.edu
\textsuperscript{2}SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA
\textsuperscript{3}Center for Cosmology and Particle Physics, Department of Physics, New York University, New York, NY 10003, USA

Received 2013 June 19; accepted 2014 January 20; published 2014 February 21

ABSTRACT

Many approaches to obtaining cosmological constraints rely on the connection between galaxies and dark matter. However, the distribution of galaxies is dependent on their formation and evolution as well as on the cosmological model, and galaxy formation is still not a well-constrained process. Thus, methods that probe cosmology using galaxies as tracers for dark matter must be able to accurately estimate the cosmological parameters. This can be done without knowing details of galaxy formation a priori as long as the galaxies are well represented by a halo occupation distribution (HOD). We apply this reasoning to the method of obtaining $\Omega_m$ and $\sigma_8$ from galaxy clustering combined with the mass-to-number ratio of galaxy clusters. To test the sensitivity of this method to variations due to galaxy formation, we consider several different models applied to the same cosmological dark matter simulation. The cosmological parameters are then estimated using the observables in each model, marginalizing over the parameters of the HOD. We find that for models where the galaxies can be well represented by a parameterized HOD, this method can successfully extract the desired cosmological parameters for a wide range of galaxy formation prescriptions.

Key words: cosmological parameters – galaxies: clusters: general – galaxies: formation – galaxies: halos – large-scale structure of universe

Online-only material: color figures

1. INTRODUCTION

Galaxies trace the dark matter distribution in the universe in a biased and nontrivial way that depends on the properties of the galaxy sample. However, the connection between galaxies and dark matter halos makes determining their bias feasible. Dark matter halos are collapsed, virialized structures that grow from initial density perturbations in the early universe. Galaxies form at the centers of the potential wells of these halos. Therefore, the distribution of galaxies is closely related to the dark matter distribution in the universe. The bias of a galaxy population depends on (1) the cosmology, which determines the statistics of the halo population, and (2) the processes of galaxy formation going on within each halo. This galaxy bias must be accounted for when using measures such as galaxy clustering to infer properties of the dark matter distribution and the underlying cosmology. In particular, differences in predictions for galaxy formation from that of actual galaxies may be instead mistaken for differences in the dark matter and cosmology.

Previous studies have focused on using galaxy clustering alone to determine the cosmology or a parameterized relationship between galaxies and dark matter. A common approach is to use a halo occupation distribution (HOD) to specify the relationship between galaxies and dark matter halos and then to constrain the parameters of this relationship using measurements such as galaxy two-point clustering (e.g., Yang et al. 2008, 2009; Zehavi et al. 2011; Leauthaud et al. 2012, and references therein). Other methods of inferring the relationship between galaxies and dark matter include satellite kinematics (More et al. 2009), galaxy group catalogs (Yang et al. 2009), abundance matching (Reddick et al. 2013 and references therein), counts-in-cells (Reid & Spergel 2009), and numerous applications of weak lensing (e.g., Tasitsiomi et al. 2004; Mandelbaum et al. 2006; Yoo et al. 2006; Cacciato et al. 2009; Sheldon et al. 2009; van Uitert et al. 2011). Several other current studies combine lensing with other measurements to either suppress the effects of small-scale galaxy clustering (e.g., Baldauf et al. 2010; Mandelbaum et al. 2013) or use a conditional luminosity function model to account for it (Cacciato et al. 2013 and references therein).

Parallel to these methods and to take advantage of upcoming galaxy cluster results, here we examine the method described in Tinker et al. (2012). This approach combines a two-point clustering measurement with the mass-to-number ($M/N$) cluster statistic. $M/N$ is the ratio between the mass of the cluster determined via weak lensing and the number of galaxies of some well-defined type it contains, typically specified by cuts in luminosity and color. The clustering is measured on all galaxies in the sample, while $M/N$ isolates only the galaxies in the most highly biased halos. The information in $M/N$ is analogous to that in the mass-to-light ratio ($M/L$) of clusters, which was combined with halo occupation techniques by van den Bosch et al. (2003) and Tinker et al. (2005) to break the degeneracy between cosmology and galaxy bias. In this paper, we will demonstrate that in the absence of observational systematics this approach accurately recovers the cosmological parameters $\Omega_m$ and $\sigma_8$ for a variety of different galaxy formation models. These models include significant differences in the efficiency of galaxy formation, producing different galaxy populations over the same sample of halos. In sum, we show that including the $M/N$ statistic with galaxy clustering allows us to marginalize over galaxy formation and galaxy bias to accurately measure cosmological parameters. Possible additional effects from galaxy survey measurements depend on the data set in question and will be considered in greater detail in future work.
Ongoing surveys, such as the Baryon Oscillation Spectroscopic Survey (BOSS; Dawson et al. 2013) and the Dark Energy Survey (Albrecht et al. 2006) provide, or will provide, galaxies over a large enough volume and depth to make precise statements about their properties. In relation to these large surveys, there have been significant recent improvements in optical galaxy cluster finding (e.g., Rykoff et al. 2012). The improvements include accurate cluster photometric redshifts and richness estimates, both of which are important to accurately determine the dark matter distribution.

We will begin by discussing the different galaxy models we use in Section 2. The $M/N$ method will be briefly reviewed in Section 3. In Section 4 we will present the results of applying this method to our different models. We will summarize our results and discuss future directions in Section 5. Unless stated otherwise, all values are given for $h = 1$, where $h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1})$.

2. MODELS

We use a handful of cosmological simulation boxes with several galaxy formation models applied. First, we use the Consuelo box from the LasDamas suite of $N$-body simulations. The underlying cosmology uses $\Omega_m = 0.25$ and $\sigma_8 = 0.8$. The Rockstar halo finder (Behroozi et al. 2013a) is used for halo identification along with merger trees drawn from Behroozi et al. (2013b). The mass resolution of $\sim 2 \times 10^8$ allows us to resolve halos that typically host galaxies below $L^*$. To this simulation we apply two different methods of generating galaxy models. The first method is a semianalytic model (SAM) based on the methods of Lu et al. (2011, 2012), and the second is an abundance-matching method as described in Reddick et al. (2013). For the first of the two SAMs, we adopt a SAM that includes flexible parameterizations for the most important baryonic processes in galaxy formation to follow the evolution of galaxies in a set of dark matter halo merger trees. The baryonic processes implemented in this SAM are described in Lu et al. (2011, 2012). Different from the previously published versions, the version of the model adopted for this article is applied on a set of halo merger trees extracted from the Consuelo simulation, which is a cosmological $N$-body simulation with subhalos identified. We consider two different versions of this SAM. One version adopts a set of parameters randomly chosen from the posterior distribution of the model that was constrained by the $K$-band luminosity function of local galaxies (Lu et al. 2012). Therefore, the model reproduces the $K$-band luminosity function of low-redshift galaxies. We refer to this version as SAM v1. We note that because the model was constrained on Monte Carlo merger trees generated based on the extended Press–Schechter theory, the model reproduces the constraining data with a larger discrepancy when the SAM is applied to the simulation merger trees.

The second SAM is unconstrained by observational data and is called SAM v2. We arbitrarily change two parameters from SAM v1. Specifically, we increase the mass scale for the halo quenching by one order of magnitude and decrease the supernova feedback mass-loading factor by a factor of two. These modifications result in an enhanced cooling of hot gas in high-mass halos and a reduced star-formation feedback in low-mass halos. The intent is to compare the results from a model with a physically motivated connection between galaxies and dark matter halos (SAM v1) with a model deliberately chosen to differ greatly from physical realism (SAM v2). Thus, the way galaxies form is very different between the two models, producing a dramatically different galaxy–halo connection. The observational measures we use to infer cosmology are substantially different between the two models, even though the halo population is the same. For instance, SAM v2 has many more bright galaxies than SAM v1. A comparison of the spatial distribution of the galaxies in the two SAMs is shown in Figure 1.

| Model | Type | Simulation | Side (Mpc $h^{-1}$) | Luminosity Function Matching | $\sigma$ (dex) | $\chi^2$ Best Fit (dof) | $N$ |
|-------|------|-----------|---------------------|-----------------------------|----------------|------------------------|-----|
| SAM v1 | SAM | Consuelo | 420 | $K$-band | ... | 4.23 | 12 |
| SAM v2 | SAM | Consuelo | 420 | None | ... | 8.72 | 12 |
| SHAM | SHAM | Consuelo | 420 | $r$-band | $v_{\text{peak}}$ | 0.2 | 8.18 | 12 |
| vp0 | SHAM | Bolshoi | 250 | $r$-band | $v_{\text{peak}}$ | 0 | 10.7 | 10 |
| vp1 | SHAM | Bolshoi | 250 | $r$-band | $v_{\text{peak}}$ | 0.1 | 7.89 | 10 |
| vp2 | SHAM | Bolshoi | 250 | $r$-band | $v_{\text{peak}}$ | 0.2 | 7.96 | 10 |
| vp3 | SHAM | Bolshoi | 250 | $r$-band | $v_{\text{peak}}$ | 0.3 | 6.50 | 10 |
| vn2 | SHAM | Bolshoi | 250 | $r$-band | $v_{\text{peak}}$ | 0.2 | 22.7 | 10 |
| Esm | SHAM | Esmeralda | 640 | $r$-band | $v_{\text{peak}}$ | 0.2 | 16.6 | 11 |
| MD | SHAM | MultiDark | 1000 | $r$-band | $v_{\text{peak}}$ | 0.2 | 9.34 | 11 |

Table 1: List of Models

The LasDamas simulations are available at http://lss.phy.vanderbilt.edu/lasdamas/
two different SAMs and the SHAM. This variation, as well as the associated differences in $M/N$ and the clustering, demonstrates how different the galaxy formation is in the three cases and permits a test of whether our method can successfully extract the cosmological parameters from very different models.

We also use the Bolshoi simulation (Klypin et al. 2011), which has the cosmology $\Omega_m = 0.27$ and $\sigma_8 = 0.82$. This simulation has the advantage of probing lower mass subhalos, with the drawback of lower volume relative to Consuelo, i.e., $(250 \text{ Mpc} h^{-1})^3$ rather than $(420 \text{ Mpc} h^{-1})^3$. We will compare a set of models drawn from those presented in Reddick et al. (2013) to demonstrate a smoother variation between HODs than the few extreme cases in Consuelo. The cases we consider will be those matched to the peak maximum circular velocity (or $v_{\text{peak}}$) of the (sub) halos, with scatter ranging from 0 to 0.3 dex in luminosity at fixed $v_{\text{peak}}$ and one model using the maximum circular velocity at the present time ($v_{\text{max}}$) with 0.2 dex scatter. This smoother variation in the galaxy model allows for a more precise test for systematic variation in the inferred cosmology with galaxy formation.

Our final pair of models uses two larger boxes. The Esmeralda simulation, also drawn from the set of LasDamas simulations, has the same cosmology as Consuelo ($\Omega_m = 0.25$ and $\sigma_8 = 0.80$) but a larger volume, $(640 \text{ Mpc} h^{-1})^3$. The MultiDark simulation (Riebe et al. 2013) has a larger volume than Bolshoi $(1000 \text{ Mpc} h^{-1})^3$, but the same cosmology, $\Omega_m = 0.27$ and $\sigma_8 = 0.82$. We run the same SHAM model on both simulations, matching to $v_{\text{peak}}$ with 0.2 dex of scatter. Since both of these simulations with larger volumes have poorer mass resolutions, we will consider galaxy populations at lower number densities (brighter luminosities) where the missing low-mass halos will not have a large impact. We will use these two models as a test of the distinction between cosmologies and to demonstrate changes in the constraints with larger volumes.

For all the Bolshoi and Consuelo models, we use a luminosity threshold of $M_r < -20.5$ to determine our galaxy sample. For the Esmeralda and MultiDark simulations, we use fixed number density cuts of $1.21 \times 10^{-3} \text{ (Mpc} h^{-1})^{-3}$ and $3.00 \times 10^{-4} \text{ (Mpc} h^{-1})^{-3}$, respectively, which correspond to luminosity thresholds of $M_r < -21.0$ and $M_r < -21.447$.

Because the halo model we use is an approximation, there will be physical effects both in the real universe and in our simulations that are not incorporated into the HOD. Using the simulations to make mock galaxy catalogs provides a way to test that the approximations in the HOD do not reduce the efficacy of the HOD approach. These possible physics effects include the following.

1. Scatter in the concentration–mass relationship of dark matter halos. Estimates of this scatter in the literature range from ~0.1 to 0.18 (Bullock et al. 2001; Wu et al. 2013b) for cluster-sized halos, and similar scatter is found in the halo catalogs used in this work. If satellite galaxies follow the same scatter in concentrations as the dark matter, then the distribution of very small-scale pairs may be different from the no-scatter case.

2. Nonspherical halos. Any nonspherical distribution of halo substructures, and thus of subhalos and galaxies, could potentially alter the two-point correlation. Particularly on small scales, additional correlations in shape such as between neighboring halos or within halos can enhance the two-point clustering relative to the case with spherical halos. This is demonstrated by some current studies (e.g., van Daalen et al. 2012; Wu et al. 2013a, and references therein).
3. Deviations of the subhalo distribution from a Navarro–Frenk–White (NFW) profile (e.g., Gao et al. 2008) or from the dark matter concentration–mass relation. As discussed in the previous point, systematic deviations of the satellite galaxy distribution in halos from what is expected will generally produce systematic differences in the clustering predicted from a given HOD, particularly at the smallest scales. Because all of the models used herein place galaxies on resolved subhalos, this effect and the previous one should be included at a realistic level in the models tested.

4. Non-Poisson scatter in the subhalo or satellite galaxy distribution. The distribution for the number of satellites in halos of a given mass is generally described as a Poisson distribution with a mean of $\langle N_{\text{sat}}(M_{\text{vir}}) \rangle$ and second moment $\alpha = \langle N(N - 1) \rangle / \langle N \rangle^2$. Some previous works (Kravtsov et al. 2004; Zheng et al. 2005; Yang et al. 2008) found results consistent with a Poisson distribution, but more recently both Boylan-Kolchin et al. (2010) and Wu et al. (2013a) have found larger scatter for higher satellite numbers.

5. In turn, a difference in this distribution of satellites at low host halo mass or richness has the potential to alter the two-point clustering at small or moderate scales, near the transition between the one-halo and two-halo terms. We expect that the models here span a range of possible impacts for this scatter.

5. Assembly bias. As discussed in, e.g., Gao et al. (2005), Wechsler et al. (2006), Croton et al. (2007), assembly bias alters the distribution of galaxies, even in the case where the mean number of galaxies per halo is held constant. That is, the number of galaxies per halo depends on both the mass of the halo and how it was formed (such as earlier or later formation). In this case, for a given HOD, the $M/N$ measurement would remain the same, but the two-point clustering would differ from that which would be predicted without assembly bias. The extra dependence on formation time or another parameter could potentially bias constraint. We expect that the models here span a range of possible impacts for assembly bias; the abundance-matching models are only weakly dependent on formation history, while the SAMs are more so.

In short, there are a number of approximations in our HOD framework, which make physical assumptions that differ from our simulated catalogs and very likely from the real universe. Thus, in addition to testing different models of galaxy formation, it is necessary to demonstrate that none of these approximations has a deleterious effect on our ability to recover cosmological parameters in an unbiased fashion. In the following section, we will show that we recover the true cosmology in all cases, even when tested with Gpc-sized simulations that yield clustering measurements with errors on the order of a few percent.

3. METHODS

We combine two measurements to reconstruct the cosmology. The first is a projected two-point correlation function (e.g., Zehavi et al. 2005, 2011; Watson et al. 2011; Nuza et al. 2013, and references therein). Our clustering is measured within each simulation box, including redshift–space distortions and integrating up to 40 Mpc $h^{-1}$ along the line-of-sight. We obtain

errors by scaling to the variation in clustering in multiple other simulation boxes of the same resolution (for Consuelo) or in a set of particle mesh boxes of adequate resolution for the scales of interest (Bolshoi). In all cases, we use the full covariance matrices.

The second measurement is the $M/N$ value, which we obtain by first assigning a “richness” to the isolated halos. This richness, which we will label $N_{200}$, is used to emulate optical richness as a mass proxy. $N_{200}$ is not the same as the $N$ used in the $M/N$ ratio. We assume a mass–richness relationship of the following form:

$$\langle \ln N_{200} \rangle = 1.09 + 0.75 \cdot \ln(M_{\text{vir}}/M_{\text{vir},\text{min}}).$$

We use a value of $M_{\text{vir},\text{min}} = 1.44 \times 10^{13} M_{\odot} h^{-1}$ and include a Gaussian scatter of 0.35 in $\ln N_{200}$ (or 0.15 dex) in richness at fixed mass. This is functionally equivalent to that of Rozo et al. (2010) changed to $h^{-1}$ units.

To recover the expected $M/N$ ratio for each bin in richness $N_{200}$, we evaluate an expression similar to that presented in Tinker et al. (2012):

$$M/N = \frac{\sum N_{\text{sat}} \int dM n_h(M) P(N_{200}|M) M}{\sum N_{\text{sat}} \int dM n_h(M) P(N_{200}|M) (N_{\text{sat}})_M M}.$$

Here, $n_h(M)$ is the halo mass function. $P(N_{200}|M)$ is the distribution of richnesses at fixed mass, a log–normal as described above. $(N_{\text{sat}})_M$ is the mean number of satellites above the luminosity threshold in a halo of mass $M$. Covariance matrices are estimated via jackknife resampling of the galaxy simulation.

We simultaneously fit the HOD and the cosmological parameters $\Omega_m$ and $\sigma_8$ using an HOD model combined with the halo mass function of Tinker et al. (2008) and bias functions of Tinker et al. (2010). We fit the two-point clustering and the $M/N$ measurements using a six-parameter HOD of the following form:

$$\langle N \rangle = \langle N_{\text{cen}} \rangle + \langle N_{\text{sat}} \rangle,$$

$$\langle N_{\text{cen}} \rangle = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{\log M_{\text{vir}} - \log M_{\text{min}}}{\sigma_{\log M}} \right) \right] \times \left[ 1 + f_{\text{cen}} \log(M_{\text{vir}}/M_{\text{min}}) \right] \langle N_{\text{cen}} \rangle.$$

$$\langle N_{\text{sat}} \rangle = \left( \frac{M_{\text{vir}}}{M_{\text{threshold}}} \right)^{\alpha_{\text{HOD}}} \exp \left( \frac{M_{\text{cut}}}{M_{\text{vir}}} \right).$$

For the Bolshoi- and Consuelo-based models, we set $f_{\text{cen}} = 0.05$. For the tests on large volumes, we allow $f_{\text{cen}}$ to vary. A test without the $f_{\text{cen}}$ factor indicated that this may bias $\Omega_m$ low by $\sim 10\%–20\%$.

We find that the recovery of the cosmological parameters is insensitive to small differences in the analytic form of the HOD, so long as the HOD model is capable of capturing the important features. As will be demonstrated in the SAM models, differences in the HOD at the low-mass end may be compensated for. However, unusual HODs at the massive end can complicate our procedure. This is especially problematic for HODs where the fraction of halos with a central galaxy brighter than the chosen threshold is significantly less than one. In the models we have considered, this occurs for thresholds with roughly $M_\ast < -21.5$. We discuss the necessity of using a more complex form for the HOD in the Appendix.
To calculate the clustering from the HOD, galaxies are assumed to follow an NFW profile (Navarro et al. 1997; Guo et al. 2012; Tinker et al. 2012) with a concentration of $f_{\text{con}} \times c_{\text{vir}}$, where $c_{\text{vir}}$ is the host halo concentration. The $M/N$ ratio is calculated using the HOD and the $N_{200}$ relationship given in Equation (1). The concentrations are based on the model of Bullock et al. (2001) using the updated parameters from Wechsler et al. (2006). A test using the concentration–mass relation of Prada et al. (2012) did not show any significant difference in the estimation of parameters aside from $f_{\text{con}}$. Therefore, we do not expect that the precise details of the concentration–mass relation will significantly affect our results.

Because the value of $M_{\text{min}}$ is set by the number density of galaxies, we are left with five free HOD parameters and two free cosmological parameters ($\Omega_m$ and $\sigma_8$). Holding the number density fixed is reasonable because the error on the number density is very small ($<1\%$ for Bolshoi and Consuelo, and $~1\%$ for the MultiDark and Esmeralda samples). We will include the number density as a data point, add it to the fit, and set the remaining HOD parameter free in future work with observed data. At present, the error budget is clearly dominated by the $w_p$ and $M/N$ errors. For instance, at $\sim 1$ Mpc $h^{-1}$, the estimated error in the clustering in Consuelo is about $8\%$. For our free parameters, we then use a Markov Chain Monte Carlo (MCMC) method to explore this parameter space with flat priors. For the modeling of the clustering from the HOD, we use the analytical model and halo exclusion given by Tinker et al. (2012).

As described in Tinker et al. (2012), fitting to the clustering alone results in degeneracies between the HOD parameters and the cosmology. Adding the $M/N$ measurement breaks this degeneracy by placing a strong constraint on the power law part of the HOD. This restricts the possible number of satellite galaxies and therefore the small-scale clustering in particular. An example fit to SAM v1 is shown in Figure 2. The $w_p(r_p)$ and $M/N$ measurements on the SAM are clearly well reproduced.

There are some degeneracies that remain in the model. The shape of the cutoff of the HOD at low host halo masses for central galaxies is not always well reproduced, as is clear in Figure 2. In this regime, the $M/N$ ratio provides no constraint, as halos there are too low mass and have too few galaxies for $M/N$ to be reliably measured. Thus, there is a significant degeneracy between the parameters that control the cutoff, which is constrained only by the two-point clustering in the transition region between the one-halo and two-halo terms. Nonetheless, we find that fits to the models are more accurate if all three cutoff parameters ($M_{\text{min}}$, $\sigma$, $M_{\text{in}}$) are included. In the particular case of SAM v1, there is an extra feature in the HOD produced by quenching that is not captured by the form of the HOD we have chosen. A good fit to the $M/N$ and clustering does exist, although the precision of the measured cosmological parameters is reduced. While we do not expect such an unusual HOD to occur physically, it is encouraging that the fit still recovers reasonable cosmological results.

For the large-volume models (using the Esmeralda and MultiDark simulations) prior to running the MCMC chain, we constrain the scale-dependent halo bias. For the other models, we use the parameterization and exact parameters presented in Tinker et al. (2005). That parameterization is given by

$$b^2(M, r) = b^2(M) \left[ 1 + 1.17\xi_{\text{nr}}(r) \right]^{1.49} \left[ 1 + 0.69\xi_{\text{nr}}(r) \right]^{3.09},$$

(6)

where $\xi_{\text{nr}}(r)$ is the nonlinear matter clustering. For all the smaller volume simulations, the numerical factors in this equation are held fixed, which is sufficient for the true HOD measured from the simulation to accurately reproduce the two-point clustering. However, this is not the case for Esmeralda and MultiDark. Therefore, we separately fit these numerical factors in those cases, while holding the HOD and cosmology fixed to the true values. This is necessary to ensure that the modeling of the $w_p(r_p)$ from the true HOD is accurate. The previous study in Tinker et al. (2012) stated that $b(r)$ was one of the main systematic uncertainties in the analysis. Our investigation demonstrates that $b(r)$ is not mass independent. In a future work, we will present a new calibration of scale-dependent bias that fully takes into account the dependence on halo mass.

4. RESULTS

The results using the Consuelo models are shown in Figure 3. Figure 3(a) shows the very different luminosity functions of each model, as well as the different clustering and $M/N$ results for the $M_r < -20.5$ sample. However, all three galaxy models are built on the same halo population. Thus, the differences in clustering and $M/N$ are due only to differences in the halo occupation. The models are well separated in halo occupation, as is shown by the
The constraints in the $\Omega_m - \sigma_8$ plane in Figure 3(d) show the result of applying our method, marginalizing over the HOD. In all three cases, the results are consistent with the true cosmology. The larger area of the SAM v1 contours is largely due to the difficulty in accurately reproducing the nonmonotonic HOD that is present in that model.

The Bolshoi models are shown in Figure 4, in the same format as given in Figure 3. We omit the model with a scatter of 0.1 dex from this plot for the sake of clarity due to its similarity to the zero-scatter case. Because the luminosity function is the same for all of the SHAM galaxy models used on the Bolshoi simulation, we instead show the HODs in Figure 4(a). In all cases, the contours in Figure 4 are consistent with the input cosmology. For this family of models, we also find that the contours are very similar to each other, regardless of the quantity of scatter. This is not surprising, as the HOD for these models is very similar for massive halos, with differences only becoming clear below the halo mass where $M/N$ is measured.

The significantly reduced number of galaxies in the $v_{\text{max}}$ model increases the variance in the $M/N$ measurement, somewhat increasing the area of the contour and altering its shape.

The models using significantly larger volumes in the Esmeralda and MultiDark simulations are shown in Figure 5. We also note here that rather than using the $M/N$ ratio with $N$ as the number of satellites, the analysis on Esmeralda uses the ratio between the mass and the total number of galaxies, $M/N_{\text{tot}}$. This helps avoid some of the complications with deviations of the central HOD from our fitting function, which occurs at relatively high masses for sufficiently bright luminosity thresholds. Ultimately, we expect that using all of a cluster’s galaxies when working with observations will help avoid additional systematic errors introduced by miscentering (e.g., Rozo & Rykoff 2013).

To summarize all of our models, Figure 6 shows the best-fit results and the marginalized errors for all of the models. While in some cases the truth is outside the marginalized 68% confidence limits, the results are consistent in the $\Omega_m - \sigma_8$ plane.

5. SUMMARY AND CONCLUSIONS

We investigate the combination of galaxy clustering and the $M/N$ of galaxies in clusters as a cosmological probe, first introduced by Tinker et al. (2012). Here, we focus on the ability of this method to obtain robust cosmological constraints while marginalizing over galaxy formation. We consider four cosmological boxes with different galaxy formation prescriptions, number densities, clustering properties, luminosity functions, underlying halo occupations, and box volumes. For all 10 models we examine, we are able to recover the correct $\Omega_m$ and $\sigma_8$ values within the 68% contour. This suggests that for data samples of this size, we are effectively able to marginalize over the physics of galaxy formation and its consequent galaxy–halo connection to obtain unbiased cosmological parameters. We suggest that this kind of analysis should become standard procedure for all cosmological constraints that depend on galaxy properties or the galaxy–halo connection.

Of some interest is the dependence of our results on the analytic form of the HOD we choose. Our current model

**Figure 3.** Applying the HOD method to three different galaxy formation models built upon the same $N$-body simulation (Consuelo). Panel (a): luminosity functions. SAM v1 and SAM v2 represent two semianalytic galaxy formation models. SHAM used the abundance-matching method. Inset: luminosity functions used in each model. Axes are $r$-band magnitude and base-10 log of $\Phi$ ($\text{Mpc}^{-1} \text{h}^{-3} \text{mag}^{-1}$). Panel (b): points with error bars show the projected clustering measurements from each model. Curves show best-fit HOD models, where both halo occupation and cosmological parameters are free. Panel (c): $M/N$ measurements, as a function of cluster richness, from the galaxy formation models. Panel (d): constraints on cosmological parameters after marginalizing over HOD parameters. The true cosmology is indicated with the solid square.

(A color version of this figure is available in the online journal.)
Figure 4. Similar to Figure 3, we apply the HOD method to five different abundance-matching models on the same Bolshoi simulation. Because these models use abundance matching, the luminosity functions are identical. The "vp" models use $v_{\text{peak}}$, while "vn" indicates $v_{\text{max}}$. The number indicates the scatter in tenths of dex. The $\sigma = 0.1$ model is very similar to the zero-scatter model and has been omitted for the sake of clarity. Panel (a): HODs for the different models. Panel (b): points with error bars show the projected clustering measurements from each model. Curves show best-fit HOD models, where both halo occupation and cosmological parameters are free. Panel (c): $M/N$ measurements, as a function of cluster richness, from the galaxy formation models. Panel (d): constraints on cosmological parameters after marginalizing over HOD parameters. The true cosmology is indicated with the solid square.

(A color version of this figure is available in the online journal.)

Figure 5. Similar to Figure 4, we apply the HOD method to two different abundance-matching models on two different large-volume simulations. Panel (a): HODs for the different models. Panel (b): points with error bars show the projected clustering measurements from each model. Curves show best-fit HOD models, where both halo occupation and cosmological parameters are free. Panel (c): $M/N$ measurements, as a function of cluster richness, from the galaxy formation models. Panel (d): constraints on cosmological parameters after marginalizing over HOD parameters. The true cosmology for the Esmeralda-based model is shown with a green diamond; the MultiDark cosmology is shown with a black triangle.

(A color version of this figure is available in the online journal.)
is clearly adequate for the galaxy models we present here. However, a more extreme HOD than that of SAM v1, with a much larger decrement in the HOD, was only poorly fit and gave biased cosmological parameters. We expect that such a nonmonotonic HOD is not physical. However, for nonstandard galaxy samples, for example those selected by color or star-formation rate rather than by luminosity, care should be taken to assure that the analytic form of the HOD is flexible enough to model the data.

In future work, we expect to obtain tighter constraints by using multiple different luminosity thresholds on the same data set or by using multiple redshifts. We anticipate applying this method to data from the Sloan Digital Sky Survey, the BOSS Survey, and the Dark Energy Survey. We may apply this method to cluster galaxies only, rather than all galaxies, which may eliminate some of the uncertainties associated with the low host halo mass cutoff in the HOD. Angular clustering, combined with accurate photometric redshifts, will allow application of this method to the large volumes being probed by photometric surveys. We anticipate studying possible biases on cosmology derived from observational methods, such as photometric redshifts and cluster-finding selections. Further work will also be required to show to robustness of this method to photometric redshift errors and over wide redshift ranges.

This work has been supported by the National Science Foundation under NSF-AST-1211838, and by Stanford University through a Stanford Graduate Fellowship to R.M.R. We thank the participants of the Fall 2012 “Small-Scale Structure” workshop at Stanford for useful discussions. We are grateful to Anatoly Klypin for providing access to the Bolshoi and MultiDark simulations, and to our collaborators in the LasDamas collaboration for providing access to the Consuelo and Esmeralda simulations. We thank Peter Behroozi for providing the halo catalogs and merger trees for all four simulations.

APPENDIX

PARAMETERIZATION OF THE HOD

We use a slightly more complex variation on the HOD relative to other works based on the HOD (e.g., Zehavi et al. 2011) because we find that incorrect modeling of the central part of the HOD can cause large deviations in the two-point clustering, particularly for bright thresholds.

An example of this is shown in Figure 7. Here we present a comparison of the true HOD and clustering from the simulation (points) compared against several different instances of the HOD model. In all these cases, we hold the cosmology fixed. Restating our HOD model, we have the following:

\[
\langle N_{\text{cen}} \rangle = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{\log M_{\text{vir}} - \log M_{\text{min}}}{\sigma_{\log M}} \right) \right] \\
\cdot \left[ 1 + f_{\text{cen}} \log \left( \frac{M_{\text{vir}}}{M_{\text{lin}}} \right) \right], \tag{A1}
\]

\[
\langle N_{\text{sat}} \rangle = \left( \frac{M_{\text{vir}}}{M_{\text{1}}} \right)^{\alpha_{\text{HOD}}} \exp \left( -\frac{M_{\text{cut}}}{M_{\text{vir}}} \right). \tag{A2}
\]

This model reduces to the more standard five-parameter model if \( f_{\text{cen}} \) is set to zero, which also ensures that \( M_{\text{lin}} \) no longer has any effect. We demonstrate the difference between these two HOD models as applied to our MultiDark sample in Figure 7.

Our first comparison is between the HOD and HOD+ models, keeping our cosmology fixed to the true values. The “HOD” models are produced by fitting to the HOD directly and then using that specific model to predict the clustering. While the seven-parameter model produces a good fit (HOD+), the other does not. This is because the only way to reproduce the lack of centrals at high halo mass in the five-parameter model is to increase the \( M_{\text{min}} \) threshold and greatly increase the width \( \sigma_{\log M} \).

The clustering produced in this case is severely suppressed due to the inclusion of many low-mass halos, which also forces the number density to be far too high.

We also test fitting each model to the \( w_p(r_p) \) and \( M/N \) statistics, allowing only the HOD parameters to vary. For our best first, we obtain a \( \chi^2 \) of 19 for our seven-parameter model (13 degrees of freedom (dof)). This is a marginally good fit \( (P[\chi^2 < 19] \approx 0.1) \). The model with five parameters has a \( \chi^2 \) of 119 (15 dof), which is clearly not acceptable. Visually,
the latter model is clearly insufficiently clustered on moderate to large scales ($r_p \gtrsim 1 \text{ Mpc} h^{-1}$). Setting the cosmological parameters free in this case would bias $\Omega_m$ low and $\sigma_8$ high relative to the true values.

Figure 7. Comparison of different HOD models. Top: black points show the two-point correlation measured on MultiDark. Lines are the correlation functions calculated from different HODs. The subplot shows the difference between the values measured in the simulation and the clustering predicted from the HOD models. Bottom: black diamonds are the measured central HODs, and the values measured in the simulation and the clustering predicted from the HOD models. The subplot shows the difference between the $\Delta(w_p)$ and $M/N$ with cosmology fixed. Again, the plus denotes a fit using additional parameters. The models with additional parameters clearly provide a better fit to the $w_p(r_p)$ in both cases.

A color version of this figure is available in the online journal.

REFERENCES

Albrecht, A., Bernstein, G., Cahn, R., et al. 2006, arXiv:astro-ph/0609591
Baldauf, T., Smith, R. E., Seljak, U., & Mandelbaum, R. 2010, PhRvD, 81, 063531
Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013a, ApJ, 762, 109
Behroozi, P. S., Wechsler, R. H., Wu, H.-Y., et al. 2013b, ApJ, 763, 18
Boylan-Kolchin, M., Springel, V., White, S. D. M., & Jenkins, A. 2010, MNRAS, 406, 896
Bullock, J. S., Kolatt, T. S., Sigad, Y., et al. 2001, MNRAS, 321, 559
Cacciato, M., van den Bosch, F. C., More, S., Mo, H., & Yang, X. 2013, MNRAS, 430, 767
Cacciato, M., van den Bosch, F. C., More, S., et al. 2009, MNRAS, 394, 929
Croton, D. J., Gao, L., & White, S. D. M. 2007, MNRAS, 374, 1303
Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, AJ, 145, 10
Gao, L., Navarro, J. F., Cole, S., et al. 2008, MNRAS, 387, 536
Gao, L., Springel, V., & White, S. D. M. 2005, MNRAS, 363, L66
Gao, Q., Cole, S., Eke, V., & Frenk, C. 2012a, MNRAS, 427, 428
Klypin, A. A., Trujillo-Gomez, S., & Primack, J. 2011, ApJ, 740, 102
Kravtsov, A. V., Berlind, A. A., Wechsler, R. H., et al. 2004, ApJ, 609, 35
Leauthaud, A., Tinker, J., Bundy, K., et al. 2012, ApJ, 744, 159
Lu, Y., Mo, H. J., Katz, N., & Weinberg, M. D. 2012, MNRAS, 421, 1779
Lu, Y., Mo, H. J., Weinberg, M. D., & Katz, N. 2011, MNRAS, 416, 1949
Mandelbaum, R., Seljak, U., Kauffmann, G., Hirata, C. M., & Brinkmann, J. 2006, MNRAS, 368, 715
Mandelbaum, R., Slosar, A., Baldauf, T., et al. 2013, MNRAS, 432, 1544
More, S., van den Bosch, F. C., Cacciato, M., et al. 2009, MNRAS, 392, 801
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493
Nava, S. E., Sánchez, A. G., Prada, F., et al. 2012, MNRAS, 432, 743
Prada, F., Klypin, A. A., Cuesta, A. J., Betancort-Rijo, J. E., & Primack, J. 2012, MNRAS, 423, 3018
Reddick, R. M., Wechsler, R. H., Tinker, J. L., & Behroozi, P. S. 2013, ApJ, 771, 30
Reid, B. A., & Spergel, D. N. 2009, ApJ, 698, 143
Riebe, K., Partl, A. M., Enke, H., et al. 2013, AN, 334, 691
Rozo, E., & Rykoff, E. S. 2013, arXiv:1303.3373
Rozo, E., Wechsler, R. H., Rykoff, E. S., et al. 2010, ApJ, 708, 645
Rykoff, E. S., Koester, B. P., Rozo, E., et al. 2012, ApJ, 746, 178
Sheldon, E. S., Johnston, D. E., Scanton, R., et al. 2009, ApJ, 703, 2217
Tasitsiomi, A., Kravtsov, A. V., Wechsler, R. H., & Primack, J. R. 2004, ApJ, 614, 533
Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, ApJ, 688, 709
Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, ApJ, 724, 878
Tinker, J. L., Sheldon, E. S., Wechsler, R. H., et al. 2012, ApJ, 745, 16
Tinker, J. L., Weinberg, D. H., Zheng, Z., & Zehavi, I. 2005, ApJ, 631, 41
van Daalen, M. P., Angulo, R. E., & White, S. D. M. 2012, MNRAS, 424, 2954
van den Bosch, F. C., Mo, H. J., & Yang, X. 2003, MNRAS, 345, 923
van Uitert, E., Hoekstra, H., Velander, M., et al. 2011, A&A, 534, A14
Watson, D. F., Berlind, A. A., & Zehavi, I. 2011, ApJ, 738, 22
Wechsler, R. H., Zentner, A. R., Bullock, J. S., Kravtsov, A. V., & Allgood, B. 2006, ApJ, 652, 71
Wu, H.-Y., Hahn, O., Wechsler, R. H., Behroozi, P. S., & Mao, Y.-Y. 2013a, ApJ, 767, 23
Wu, H.-Y., Hahn, O., Wechsler, R. H., Mao, Y.-Y., & Behroozi, P. S. 2013b, ApJ, 763, 70
Yang, X., Mo, H. J., & van den Bosch, F. C. 2008, ApJ, 676, 248
Yang, X., Mo, H. J., & van den Bosch, F. C. 2009, ApJ, 695, 900
Yoo, J., Tinker, J. L., Weinberg, D. H., et al. 2006, ApJ, 652, 26
Zhao, I., Zheng, Z., Weinberg, D. H., et al. 2005, ApJ, 630, 1
Zhao, I., Zheng, Z., Weinberg, D. H., et al. 2011, ApJ, 736, 59
Zheng, Z., Berlind, A. A., Weinberg, D. H., et al. 2005, ApJ, 633, 791