CLIMBER: Galaxy-Halo Connection Constraints from Next-Generation Surveys

Alan N. Pearl,[1] Rachel Bezanson,[1] Andrew R. Zentner,[1] Jeffrey A. Newman,[1] Andy D. Goulding,[2]
Katherine E. Whitaker,[3] Sean D. Johnson,[4] and Jenny E. Greene[2]

[1] *Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA 15260*
[2] *Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544*
[3] *Department of Astronomy, University of Massachusetts, Amherst, MA 01003*
[4] *Department of Astronomy, University of Michigan, Ann Arbor, MI 48109*

ABSTRACT

In the coming decade, a new generation of massively multiplexed spectroscopic surveys, such as PFS, WAVES, and MOONS, will probe galaxies in the distant universe in vastly greater numbers than was previously possible. In this work, we generate mock catalogs for each of these three planned surveys to help quantify and optimize their scientific output. To assign photometry into the UniverseMachine empirical model, we develop the Calibrating Light: Illuminating Mocks By Empirical Relations (CLIMBER) procedure using UltraVISTA photometry. Using the published empirical selection functions for each aforementioned survey, we quantify the mass completeness of each survey. We compare different targeting strategies by varying the area and **targeting** completeness, and quantify how these survey parameters affect the uncertainty of the two-point correlation function. We demonstrate that the PFS and MOONS measurements will be primarily dominated by cosmic variance, not shot noise, motivating the need for increasingly large survey areas. On the other hand, the WAVES survey, which covers a much larger area, will strike a good balance between cosmic variance and shot noise. For a fixed number of targets, doubling the survey area (and halving the completeness) would decrease the uncertainty of the correlation function at intermediate scales by 3%, 24%, and 22% for our WAVES, PFS, and MOONS samples, respectively. Meanwhile, for a fixed survey area, doubling the number of targets improves the same constraints by 14%, 5%, and 2%. All of the utilities used to construct our mock catalogs and many of the catalogs themselves are publicly available.

1. INTRODUCTION

The ΛCDM model of cosmology has been widely accepted for decades, and its parameters are now known to quite high precision (Planck Collaboration et al. 2020). Within this framework, it is assumed that galaxies form inside dark matter halos (White & Rees 1978; Blumenthal et al. 1984). While halos can be accurately modeled through gravity-only simulations (e.g., Bolshoi-Planck; Klypin et al. 2016), the fine details of galaxy formation are strongly influenced by baryonic physics, which poses a serious challenge for theoretical models of galaxy evolution to tackle.

In recent years, several ongoing projects have made great advancements to including baryonic physics in hydrodynamic simulations of galaxies in a cosmological context. These projects include but are not limited to IllustrisTNG (Nelson et al. 2019), Evolution and Assembly of GaLaxies and their Environments (EAGLE; Schaye et al. 2015; Crain et al. 2015), Feedback In Realistic Environments (FIRE; Hopkins et al. 2018), and Simba (Davé et al. 2019). However, it is still computationally prohibitive to resolve the small scales needed to simulate the processes that regulate star formation. Therefore, all of these hydrodynamic simulations still include analytic approximations for these small-scale processes. It is possible to approximate the rates of processes like gas cooling and star formation using semi-analytic models (SAMs; e.g., White & Frenk 1991; Somerville & Davé 2015) which trace dark matter halos through gravity-only simulations and map baryonic physics into these halos using analytic scaling relations. SAMs have contributed significantly to our knowledge of galaxy formation, despite challenges disentangling various physical processes that produce degenerate observations.

Alternatively, many studies of galaxy evolution and cosmology use empirical models to populate galaxies on top of dark matter halos (i.e., the galaxy-halo connection; see Wechsler & Tinker 2018 for an extensive review). Methods such as the halo occupation distribution (HOD; e.g., Berlind & Weinberg 2002; Zheng et al. 2007; Hearin et al. 2016) or abundance-matching (e.g., Kravtsov et al. 2004; Hearin & Watson 2013) are most commonly used to statistically match the stellar masses

of galaxies to the masses of their host halos. Since the number density and clustering of halos are strongly dependent on halo mass (Press & Schechter 1974; Kaiser 1984; Bond et al. 1991; Mo & White 1996; Zentner 2007), these models are informed through observations of the stellar mass function and the two-point correlation function (Zehavi et al. 2005; Reddick 2013; Wang et al. 2019).

Through analytic empirical models, one can infer the stellar-to-halo mass relation (SHMR). The SHMR has provided us valuable insight into the masses of halos that are most efficient at forming stellar mass. At its peak, stellar mass can account for up to roughly 5% of the total mass of Milky Way-mass halos, which is approximately 30% of the cosmic baryon mass fraction. However, this star formation efficiency drops dramatically at both lower and higher halo masses. This is caused by various processes causing star formation to shut off (i.e., quench) through heating or removing the gas that was fueling the star formation. Low-mass quenching is often attributed to stellar feedback (Fierlinger et al. 2016) and satellite stripping (Guo et al. 2011), while high-mass quenching is primarily attributed to active galactic nucleus (AGN) feedback (Fabian 2012). However, the dependence of these processes on secondary halo properties and redshift is poorly constrained by existing datasets.

Most of our constraints on the galaxy-halo connection come from low-redshift surveys such as the Sloan Digital Sky Survey (SDSS; Blanton et al. 2017). While it is commonly assumed that the SHMR does not evolve strongly with redshift, it is particularly difficult to probe the same range of halos at high redshifts because these surveys quickly lose faint, low-mass galaxies and massive galaxies are rare. Extending our empirical constraints on the galaxy-halo connection to the high-redshift universe, where star formation rates (SFRs) were higher and galaxy populations were rapidly evolving, should have profound implications on our knowledge of galaxy evolution.

With the advent of highly multiplexed spectrographs being used on large telescopes, thousands of spectra will be simultaneously measured, a number of spectroscopic surveys will begin to map the distant universe to an unparalleled degree over the next decade. These surveys will probe the evolution of the precise statistical distribution of galaxies at earlier cosmological times than previously possible. Interpreting these types of datasets, however, is particularly challenging due to the systematic sampling that is more easily avoidable in the nearby universe. Utilizing this new information to place constraints on the galaxy-halo connection will require care-

ful planning of survey designs and new theoretical frameworks.

In this paper, we present a procedure for mapping photometric properties (flux and colors) onto physical properties from the UniverseMachine empirical model (Behroozi et al. 2019). We refer to this procedure as Calibrating Light: Illuminating Mocks By Empirical Relations (CLIMBER). We use this procedure to construct mock galaxy catalogs, which we use to investigate the mass completeness and statistical constraints that will be available from several future massively multiplexed spectroscopic galaxy surveys: the Prime Focus Spectrograph Galaxy Evolution Survey (PFS; Takada et al. 2014), the Guaranteed Time Observation Extragalactic Survey of the Multi-Object Optical and Near-infrared Spectrograph for the Very Large Telescope (MOONS; Maiolino et al. 2020), and the Wide Area Vista Extragalactic Survey-Deep (WAVES; Driver et al. 2016). For each survey, we quantify its performance and make recommendations about future extensions to improve its constraining power on the galaxy-halo connection.

This paper is organized as follows: Section 2 explains the procedure we followed to construct our mock galaxy catalog from the fiducial UniverseMachine model (with more details in Appendix A) and discusses selection functions that we place to construct mock surveys of various galaxy populations. In Section 3, we formulate our conservative HOD model. In Section 4, we present mock measurements of number density and the two-point correlation function, which are the primary constraints of this model. In Section 5, we present projected constraints of the two-point correlation function and HOD models through Markov chain Monte Carlo (MCMC) fits, for a variety of survey parameters. We give a brief discussion of our conclusions in Section 6.

The cosmological assumptions used in each step of generating our mock catalog were made self-consistently. Bolshoi-Planck, the UniverseMachine, and all of our following calculations use a Planck-tuned ΛCDM cosmology with parameters given in Table 1. Although the stellar masses and SFRs from UltraVISTA (Muzzin et al. 2013) assumed a slightly different cosmology, their dependence on h has been corrected to match our assumption. Note that all halo masses refer to the virial mass of the halo, and we do not use h-scaled units with the exception of h^{-1} Mpc for distance.

Table 1. Cosmological parameters

Parameter	Value	Description
h	0.678	Hubble parameter
Ω_Λ	0.693	density parameter for dark energy
Ω_m	0.307	density parameter for total matter
Ω_b	0.048	density parameter for baryonic matter
n_s	0.96	normalization of the Power spectrum
σ_8	0.823	amplitude of mass density fluctuation

2. BUILDING THE EMPIRICALLY CALIBRATED MOCK SURVEYS

To construct a realistic mock galaxy catalog, we start from the UniverseMachine (Behroozi et al. 2019) empirical model, which is calibrated to reliably reproduce a very large number of statistics of galaxy populations from $0 < z < 10$. However, this model lacks a crucial element needed to test empirical selection functions: the apparent brightness of each galaxy in the observed-frame wavelengths of photometric filters. We, therefore, calibrate these model galaxies to photometry from the UltraVISTA survey (Muzzin et al. 2013) using a combination of abundance matching and random forest mapping to calculate observed mass-to-light ratios (see Section 2.2).

2.1. *UniverseMachine*

The UniverseMachine (Behroozi et al. 2019) is a sophisticated empirical galaxy-halo connection model of 44 parameters, which were iteratively fit to 1069 observed data points across a redshift range of $0 < z < 10$. It traces each dark matter halo in a gravity-only simulation and assigns a SFR to the galaxy at the center of the halo, assuming that star formation correlates with dark matter assembly. The model thereby tracks the accumulation of stellar mass of each galaxy over its entire formation history. The SFR of each galaxy is drawn from an empirically motivated distribution, which is the sum of two log-normal distributions, representing a quenched and star-forming population. The quenched fraction, as well as the center and width of the star-forming distribution, are parameterized by analytic expressions dependent on halo mass and redshift. In order to impose some assembly correlation, the SFR is not randomly drawn from the model distribution, but instead weighted such that higher SFRs are more likely to be assigned to halos with greater mass accretion rates. For an excellent visual summary of this procedure, see Figure 1 of Behroozi et al. (2019).

The UniverseMachine DR1 derives its halo catalog from the Bolshoi-Planck cosmological N-body simulation (Klypin et al. 2016). The UniverseMachine then provides mock galaxy properties such SFR and stellar mass (note that this is the "live" stellar mass, which is the integral of the star formation history subtracted by the mass returned to the interstellar medium) into each snapshot of this simulation. By piecing together these snapshots, the UniverseMachine has been tuned to reproduce many observables, such as stellar mass functions, two-point correlation functions, the star-forming main sequence, quenched fractions, environmental quenching, and more.

Before using the UniverseMachine to generate mock surveys, one needs to define the empirical properties of each mock galaxy to impose selection functions. This has previously been done by performing stellar population synthesis over each star formation history to fit the UniverseMachine to UV luminosity functions and UVJ quenching classifications. However, the UniverseMachine is only tuned to reproduce global star formation histories. Because our goal is to apply targeting strategies, the distribution of colors and fluxes must be reliable, and we, therefore, empirically calibrate a mapping from stellar mass and SFR (the `obs_sm` and `obs_sfr` columns) to the brightness in various photometric filters, as explained in Section 2.2.

2.2. *CLIMBER*

Most spectroscopic galaxy surveys have well-defined selection functions based on previously taken photometric data. Therefore, we need a method of predicting the observed light of UniverseMachine galaxies at multiple wavelengths to understand the representation of properties of the targeted galaxies (e.g., mass completeness) of these surveys. For this reason, we have developed a procedure to assign mock apparent magnitudes informed by an observational dataset. We refer to this procedure as Calibrating Light: Illuminating Mocks By Empirical Relations (CLIMBER).

In CLIMBER, we utilize the tight correlation between the mass-to-light ratio and color of a galaxy (e.g., Bell & de Jong 2001). Analogously, we map sSFR to the mass-to-light ratio for each mock galaxy via random forest regression. This can be trained by any observational dataset with the desired mass, redshift, and photometric coverage. The data shown in this paper have been calibrated to UltraVISTA photometry (Muzzin et al. 2013), but CLIMBER is a generalizable, flexible procedure that can be applied to any dataset that includes mappings between empirical fluxes and physical galaxy properties.
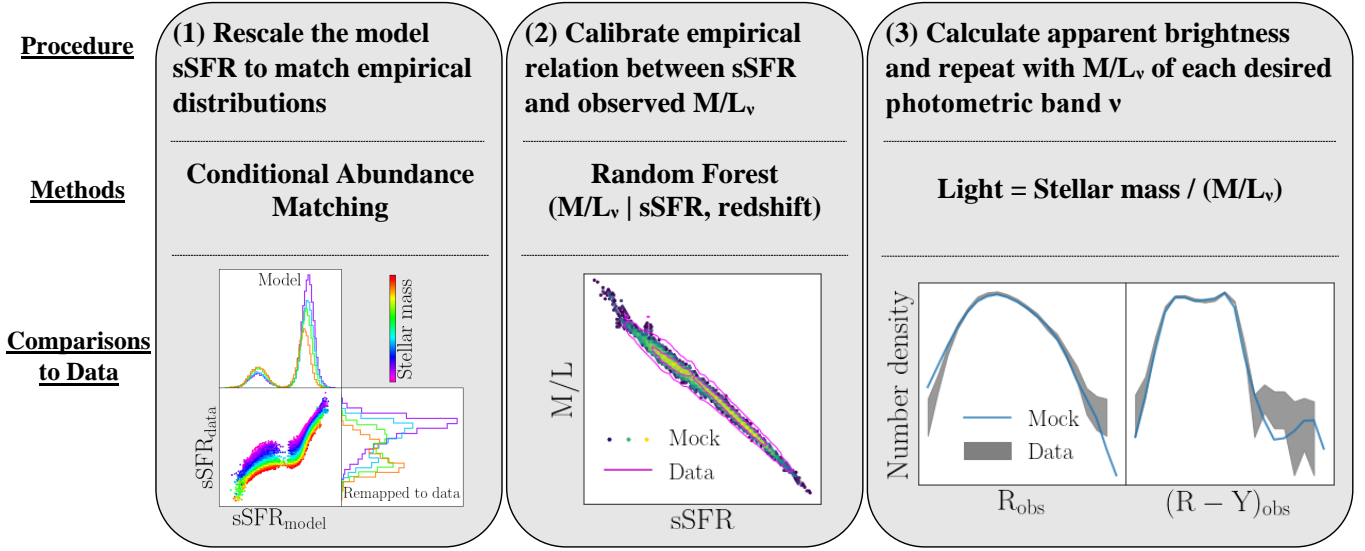
Figure 1. Visualization of our calibration procedure (CLIMBER) developed to assign the brightness and color of each mock galaxy taken from the UniverseMachine empirical model. Note that we abbreviate specific SFR (sSFR; i.e., SFR divided by stellar mass) and stellar mass-to-light ratio (M/L_ν), where ν represents the effective observed-frame frequency of a photometric band.

While the scaling relation between total SFR and stellar mass of star-forming galaxies (the star-forming main sequence) has been measured out to $z \sim 5$ (Kajisawa et al. 2010; Whitaker et al. 2014; Tomczak et al. 2016; Leslie et al. 2020), the evolution of SFRs for galaxies that fall off this relation is poorly constrained due to the difficulty of measuring low SFRs (Leja et al. 2019). For this reason, the UniverseMachine only evolves the star-forming SFR distribution with redshift. In contrast, the specific SFR (sSFR) of each quiescent galaxy in the UniverseMachine was simply drawn from a non-evolving log-normal distribution of $10^{-11.8}$ yr^{-1}±0.36 dex. While this empirically matches the local universe, it likely underestimates SFRs of high-redshift quiescent populations. This assumption does not greatly influence the accumulation of stellar mass modeled in the UniverseMachine, but it presents a problem for assigning the luminosities of quiescent galaxies. We solve this by rescaling the sSFR distributions via conditional abundance matching, as discussed in greater detail in Appendix A.

The most crucial decision one needs to make before running CLIMBER is in choosing an sSFR proxy from the calibration dataset. This proxy must (1) approximately conserve rank-ordering with true sSFR (at fixed stellar mass), (2) produce a tight, negative correlation with mass-to-light ratio, and (3) have a high detection fraction in the full galaxy population – quenched and star-forming galaxies alike. From UltraVISTA, we chose to use the specific ultraviolet SFR (sSFR$_{UV}$), which is the SFR inferred from ultraviolet bands, divided by stellar mass derived from SED fitting. These SEDs were fit by Fitting and Assessment of Synthetic Templates (FAST; Kriek et al. 2009) using the Chabrier (2003) initial mass function, an exponentially declining star formation history, and the Bruzual & Charlot (2003) stellar population synthesis model. Other sSFR proxies may be useful in different datasets. For example, we considered using sSFRs directly from SED fits, but the grid-based values fit by FAST were sampled too sparsely and therefore provide a poor mapping between physical properties and flux.

See Figure 1 for a flow chart visualization of the CLIMBER procedure. To summarize, we first perform conditional abundance matching from the model sSFR to match the empirical sSFR distribution. Then, we train the mapping from sSFR to an observed mass-to-light ratio using random forest. Finally, we convert the UniverseMachine stellar mass values to luminosities via the predicted mass-to-light ratios. For further details and analysis of our procedure, see Appendix A.

2.3. Mock Survey Selections

The product of CLIMBER is a mock realization of the universe in the form of a light cone, which can be iterated over random origins and orientations in the Bolshoi-Planck cube. We conduct mock surveys over these light cones by performing cuts that imitate the selection functions of several next-generation surveys. We then analyze many realizations of each mock survey to try to determine the uncertainty of the number density and two-point correlation function (see Sections 4.1 and 4.2) that will be measured.

In this work, we ignore the intricate details of survey geometries, overlapping pointings, and fiber collisions, which would require targeting strategies that are not yet finalized (however, our mock catalogs will be an extremely useful tool for running targeting simulations and analyzing the systematics they produce). We define our survey geometry by a square in angular coordinates and remove a random subsample to account for **targeting completness**, primarily due to fiber collisions. Our two survey parameters are thus sky area and **targeting** completeness.

We implement this selection using $|\alpha| < \alpha_{max}$ and $|\delta| < \delta_{max}$ where $\alpha_{max} = \delta_{max}$. For small angles, the solid angle area is approximately $\Omega \approx 4\alpha_{max}\delta_{max}$ (in radians/steradians), but to be precise, we calculate α_{max} and δ_{max} by inverting Equation 1.

$$\Omega = \int_{-\alpha_{max}}^{\alpha_{max}} d\alpha \int_{-\delta_{max}}^{\delta_{max}} d\delta \cos(\delta)$$
$$= 4(\alpha_{max})\sin(\delta_{max})$$
(1)

Table 2. Survey parameters

Name	Area (sq. deg)	Completeness	Redshift	Magnitude limits	References
WAVES	66	95%	$0.2 < z < 0.8$	$m_z < 21.25$	Driver et al. (2016)
PFS ($z < 1$)	12	70%	$0.7 < z < 1.0$	$m_Y < 22.5$ & $m_J < 22.8$	Takada et al. (2014)
PFS ($z > 1$)	12	70%	$1.0 < z < 1.7$	$m_J < 22.8$	Takada et al. (2014)
MOONS[a] ($z \sim 1$)	4	72.5%	$0.9 < z < 1.1$	$m_H < 23.0$	Maiolino et al. (2020)
MOONS[a] ($z \sim 1.5$)	4	72.5%	$1.2 < z < 1.7$	$m_H < 23.5$	Maiolino et al. (2020)
MOONS[a] ($z \sim 2$)	4	72.5%	$2.0 < z < 2.6$	$m_H < 24.0$	Maiolino et al. (2020)

[a]These parameters assume that MOONS implements the Xswitch strategy. If the more efficient Stare strategy is used in the VIDEO fields, the total area would increase to 7 sq. deg and the average **targeting** completeness would drop to 71.4%.

To perform any type of scientific study on a sample of galaxies, its selection function must be well understood in terms of physical properties. By imposing the published magnitude limits (see Table 2) of the WAVES, PFS, and MOONS surveys on galaxies in our mock, we can test the fraction of galaxies at a given mass that is included in the selection function to test how well-represented they will be in the survey. For each survey, we show the 90% and 99% mass-completeness limit as a function of redshift in Figure 2. The color-coded bands in this figure enclose the three galaxy populations that we further analyze in this paper by calculating mock observables (Section 4) to constrain our HOD model (Section 3). The mass thresholds and effective redshifts of these samples are listed in Table 3. These cuts are almost entirely above the respective 99% mass-completeness limits, which means these surveys should observe representative samples.

Figure 2 additionally shows the comoving area probed by each survey as a function of redshift, in comparison to that of the Bolshoi-Planck simulation, which is a periodic cube of side length 250 h^{-1} Mpc. Note that WAVES reaches a slightly larger comoving area at the high-redshift end, which may cause the cosmic variance in our mocks to be slightly underestimated, due to the high probability of resampling the same galaxies across realizations. While this should not affect our primary conclusions, a more precise analysis of the cosmic variance in WAVES should use a larger simulation than Bolshoi-Planck.

Table 3. Galaxy samples

Name	Mass threshold ($M_\odot$)	Mass completeness	Redshift range	Effective redshift	Mean sample size
WAVES	10^{11}	99.903%	$0.5 < z < 0.8$	0.647	33,583
PFS	$10^{10.5}$	99.997%	$0.8 < z < 1.2$	0.979	61,307

Table 3 *continued*

Table 3 *(continued)*

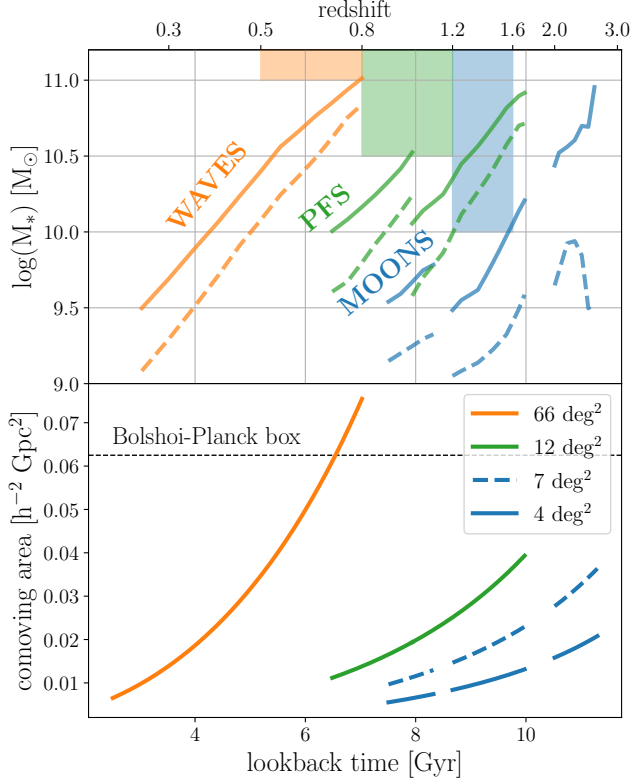| Name	Mass threshold ($M_\odot$)	Mass completeness	Redshift range	Effective redshift	Mean sample size
MOONS	10^{10}	99.744%	$1.2 < z < 1.6$	1.367	41,661

Figure 2. Mass completeness (upper panel) and field size (lower panel) for the targeting strategies (given in Table 2) of PFS, WAVES, and MOONS as a function of redshift. In the upper panel, we include 99% (solid lines) and 90% (dashed lines) completeness limits. These values are averaged over 25 mock catalog realizations. Color-coded bands indicate the mass-complete samples used in this analysis (see Table 3). In the lower panel, we plot the comoving area of each field, with sky areas taken from Table 2.

3. HOD FORMULATION

3.1. *The HOD*

In this paper, we will predict the level of constraints that several upcoming surveys will place on the galaxy-halo connection. To quantify these constraints, we use the halo occupation distribution (HOD), which has been a standard way to measure the galaxy-halo connection in magnitude limited surveys for nearly two decades (Berlind & Weinberg 2002).

The HOD prescribes the mean number of galaxies above a mass or luminosity threshold per halo. This formalism is very popular due to its simplicity and utility for galaxy clustering predictions. We use the HOD
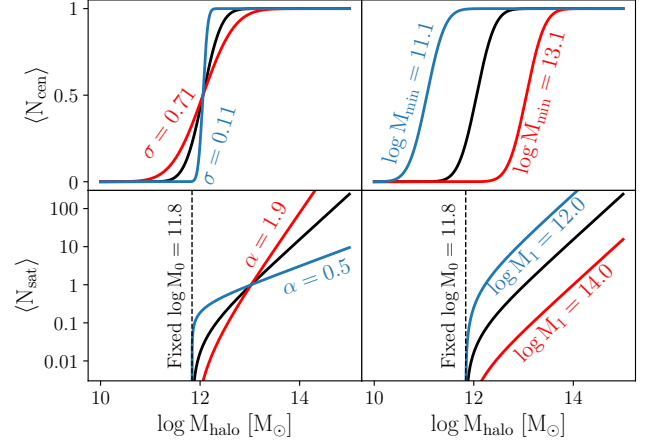
Figure 3. Mean occupation functions of centrals (Equation 2; top panels) and satellites (Equation 3; bottom panels) per halo in our HOD model. The total number of galaxies per halo is the sum of $N_{\rm cen}$ and $N_{\rm sat}$. Black curves are plotted with our fiducial set of parameters for our PFS galaxy sample. We demonstrate the effect of each model parameter by varying them one at a time, as labeled. Note that our conservative HOD model (Section 3.2) always maintains a constant $M_{\rm min}/M_1$ ratio, and conserves total number density by automatically updating $\log M_{\rm min}$ and $\log M_1$ to account for any change in σ or α.

parameter convention introduced by Zheng et al. (2007). Under this formalism, we describe the expected number of central and satellite galaxies per halo above a stellar mass threshold, $M_{*\rm thresh}$, as

$$\langle N_{\rm cen}\rangle = \frac{1}{2}\left(1 + {\rm erf}\left(\frac{\log\left(M_h/M_{\rm min}\right)}{\sigma}\right)\right) \qquad (2)$$

and

$$\langle N_{\rm sat}\rangle = \left(\frac{M_h - M_0}{M_1}\right)^\alpha, \qquad (3)$$

where we do not assume any functional forms for the redshift and $M_{*\rm thresh}$ dependence of the free parameters $M_{\rm min}$, σ, M_0, M_1, and α. Instead, we fit the HOD independently to each galaxy population of interest.

We plot these equations in Figure 3 using the fiducial parameters for our PFS sample and demonstrate how varying these parameters varies the number of galaxies per halo.

The parameters controlling $\langle N_{\rm cen}\rangle$ are the characteristic halo mass $M_{\rm min}$ and the characteristic spread σ. The parameters controlling $\langle N_{\rm sat}\rangle$ are the minimum halo mass M_0, the characteristic halo mass M_1, and the power-law slope α.

While the mean occupation function is a deterministic function of the HOD parameters, note that the number of galaxies assigned to each halo is stochastically drawn from a distribution around this mean. We draw from a Bernoulli distribution (1 or 0) for central galaxies and a Poisson distribution for satellites. Therefore, this induces some stochasticity in the number density and correlation function when populating a simulation of finite volume according to our HOD.

The HOD is constrained by quantities that probe the mass of the underlying halo population: abundance and clustering. To quantify clustering, we measure the two-point correlation function, $w_\mathrm{p}(r_\mathrm{p})$ (see Section 4.2). We quantify abundance with the number density of galaxies above the stellar mass threshold, n, which is related to the SMF (see Section 4.1) by

$$n = \int_{M_{*\mathrm{thresh}}}^{\infty} \Phi(M_*)dM_*. \tag{4}$$

The HOD can be calculated directly from the UniverseMachine by counting the average number of galaxies above the threshold in each halo. We fit Equations 2 and 3 to this calculation in narrow M_h bins to obtain fiducial HOD parameters for our WAVES, PFS, and MOONS galaxy samples. We list each fiducial HOD parameter, as well as the average number density n and satellite fraction f_sat in Table 4.

Table 4. Fiducial HOD parameters (UniverseMachine "truths")

Sample	n (h^3 Mpc^{-3})	M_1/M_min	σ	α	$\log M_\mathrm{min}$	$\log M_1$	$\log M_0$	f_sat
(Parameter type)	(Fixed)	(Fixed)	(Free)	(Free)	(Derived)	(Derived)	(Fixed)	(Derived)
WAVES	9.992×10^{-4}	5.229	0.736	1.291	12.956	13.674	12.176	0.190
PFS	4.838×10^{-3}	8.691	0.407	1.188	12.058	12.997	11.838	0.221
MOONS	7.619×10^{-3}	9.132	0.220	1.196	11.740	12.701	11.655	0.222

3.2. *The Conservative HOD Model*

Combining information from multiple sources, we expect very strong empirical constraints on the number density of most galaxy populations. Even in our mock surveys alone, we measure n to the precision of 1 to 4%, which is an order of magnitude smaller than the fractional error of w_p at large scales. Therefore, if allowed to freely vary, the number density causes a near-degeneracy between the HOD parameters it is sensitive to, increasing the difficulty of calculating constraints through MCMC with little gain. We, therefore, set the fiducial value of n as a hard prior and only consider the HOD parameter-space that conserves the number density from the UniverseMachine. To do this, we integrate Equations 2 and 3 with the halo mass function $\Phi(M_h)$ to obtain

$$n_\mathrm{cen} = \int_{0}^{\infty} \langle N_\mathrm{cen}\rangle \Phi(M_h)dM_h \tag{5}$$

and

$$n_\mathrm{sat} = \int_{0}^{\infty} \langle N_\mathrm{sat}\rangle \Phi(M_h)dM_h, \tag{6}$$

where

$$n = n_\mathrm{cen} + n_\mathrm{sat}. \tag{7}$$

The parameter M_min primarily sets the number density for the centrals and M_1 for the satellites. Since we have removed one degree of freedom by holding the total number density fixed, we are free to combine these two parameters into a single parameter: M_1/M_min. This ratio is directly influenced by the ratio of central to satellite dark matter halos predicted by dark matter simulations. Since this is decided by our cosmological prior, we choose to hold constant the M_1/M_min parameter measured from the UniverseMachine. Then, once a value is chosen for each free parameter, we can individually derive M_1 and M_min by numerically inverting Equation 7.

We remove another degree of freedom in our model by holding the fiducial value of M_0 fixed. This is common practice because the observables that we examine are not sensitive to large changes in this parameter. Therefore, we only tune two free parameters in our conservative HOD model: σ and α. We demonstrate the effect these parameters have on $w_\mathrm{p}(r_\mathrm{p})$ predictions in Figure 5. Increasing σ decreases clustering at all scales, while increasing α increases clustering, especially at small scales. Pushing to smaller-scale measurements will therefore be greatly beneficial in breaking the degeneracy of these parameters.

4. CONSTRAINTS ON THE HOD

Following the standard methodology, we constrain the HOD by empirical measurements of number density and clustering of the galaxy population. Therefore, in this

section, we present mock measurements of the stellar mass function and the projected two-point correlation function.

4.1. Stellar Mass Function

The stellar mass function (SMF), $\Phi(M_*)$, measures the number density of a galaxy population subdivided into bins of stellar mass. This is one of the most direct measurements of the efficiency of galaxy evolution, tracking the overall growth of galaxies over cosmic times from the accumulation of star formation. In terms of the galaxy-halo connection, the SMF provides an estimate for the SHMR, if we assume a good rank-order correlation between stellar and halo mass, as is done by abundance matching models.
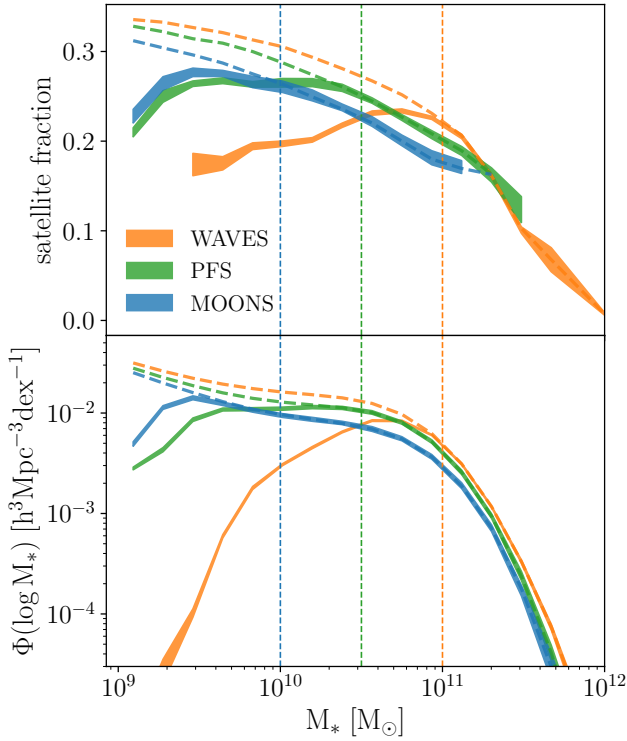
Figure 4. Mock measurements of the stellar mass function (bottom) and satellite fraction (top). We plot WAVES ($0.5 < z < 0.8$) in orange, PFS ($0.8 < z < 1.2$) in green, and MOONS ($1.2 < z < 1.6$) in blue. Each band represents the 1σ confidence region of each mock measurement, inferred by independent realizations (note that the satellite fraction here assumes perfect central/satellite assignment). Vertical dashed lines represent the stellar mass threshold we impose on each survey, while the thick dashed curves represent the true functions without photometric target selection. The observed functions agree very well with the true functions above the respective mass thresholds.

Spectroscopic surveys like PFS, WAVES, and MOONS will improve stellar mass estimates of galaxies

from the respective epochs they are probing due to the significantly increased precision via spectroscopic redshifts, as well as tighter constraints on mass-to-light ratios from stellar ages obtained by stellar population synthesis. This will substantially improve our certainty on the distribution of stellar masses (Muzzin et al. 2009), while abundance measurements will be further solidified by larger photometric surveys. Therefore, from this combination of data sources, we expect tight constraints on the SMF for a wide range of redshifts, especially at $z > 1$, in the coming years.

In each of our redshift samples, we measure the SMF over an ensemble of 25 mock survey realizations to quantify the uncertainty of a single survey. We present the mock SMFs of each survey in the bottom panel of Figure 4. Note that each mock SMF agrees with the truth down to the indicated mass threshold, which is a good sign that the survey samples will be representative of the true galaxy populations. Additionally, the upper panel of Figure 4 shows the satellite fraction as a function of stellar mass, which is also in good agreement down to the completeness limit for each sample.

As discussed in Section 2.3, we are able to quantify the mass completeness of each sample by selecting all mock galaxies over the mass threshold and calculating the fraction of them which are under the survey's magnitude limits. Over the entire redshift range, each sample is well over 99% complete down to its mass threshold (see Table 3).

Note that the SMF is typically used as a direct constraint for the HOD through Equation 4. However, in our conservative model, the SMF is used as a hard prior because we do not allow the total number density of galaxies to vary (see Section 3.2).

4.2. Two-Point Correlation Function

The two-point correlation function, $\xi(r)$, is a canonical constraint on the HOD because it measures the clustering strength of the galaxy population, which is indicative of the clustering strength of the underlying halo population. Since halo clustering is a strong function of halo mass, the two-point correlation function is very sensitive to the typical mass of the halo population (Zehavi et al. 2005; Reddick 2013; Wang et al. 2019). We adopt the projected correlation function, which is defined as

$$w_{\mathrm{p}}(r_{\mathrm{p}}) = 2 \int_0^{\pi_{\max}} \xi(r_{\mathrm{p}}, \pi) d\pi, \qquad (8)$$

where we choose $\pi_{\max} = 50\ h^{-1}$ Mpc. The projected correlation function conveniently integrates out most of the dependence on redshift-space distortions. This is desired because our galaxy-halo connection model has
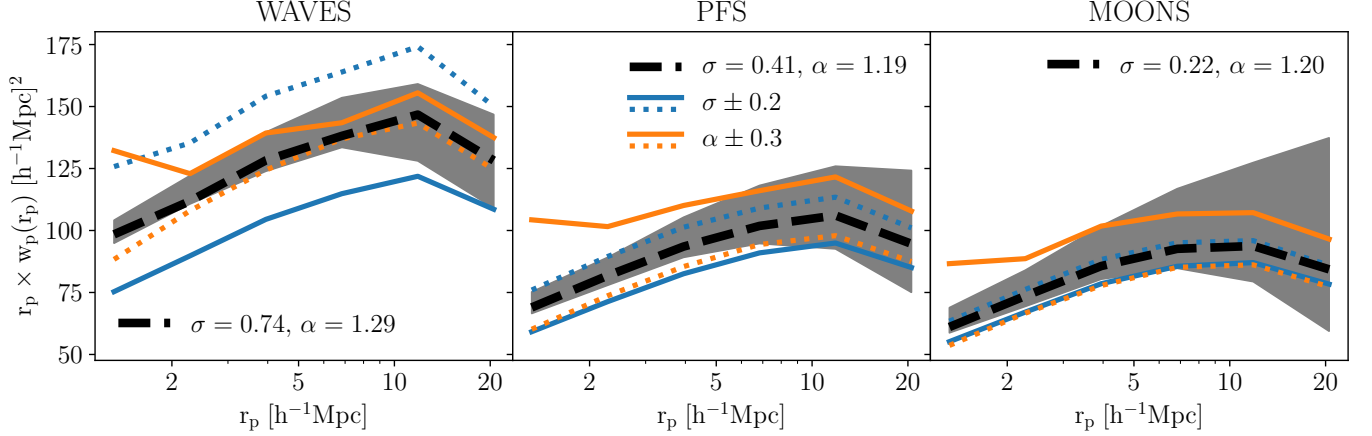
Figure 5. The projected two-point correlation function. Mock 1σ constraints from each survey are given by grey shaded regions. The UniverseMachine "truth" HOD model is represented by the thick black dashed line, while the colored solid and dotted lines show the effect of increasing or decreasing one parameter at a time, respectively. High-mass samples like WAVES are much more sensitive to changes in σ, and low-mass samples like MOONS are more sensitive to changes in α. These parameters produce similar observational effects, but this degeneracy can be broken by probing smaller scales.

no dependence on velocity dispersion, and we do not want our observable to be sensitive to that level of detail.

We perform this computation in six r_p bins with logarithmically spaced edges from 1-27 h^{-1} Mpc. Using a relatively small number of bins here helps reduce the number of realizations needed to calculate the covariance matrix, which must be very precise for our analysis. Due to the systematic sampling caused by fiber collisions in multiplexed spectroscopic surveys, we don't attempt to calculate the two-point correlation function below a scale of 1 h^{-1} Mpc (i.e., 106 arcsec at $z = 0.8$, 78.7 arcsec at $z = 1.2$, and 65.2 arcsec at $z = 1.6$). The fiber positioner patrol diameters of the instruments used by WAVES, PFS, and MOONS will likely be similarly sized, so fiber collisions should not dominate our uncertainty and the effect will be mostly mitigated by revisiting fields multiple times.

We calculate the projected two-point correlation function using the Landy & Szalay (1993) estimator implemented in the `DDrppi`, `DDrppi_mocks`, and `convert_rp_pi_counts_to_wp` functions from the `Corrfunc` package (Sinha & Garrison 2020). We measure $w_\mathrm{p}(r_\mathrm{p})$ from each mock survey in Figure 5, and compare it to various predictions of our conservative HOD model (see Section 3.2). These measurements are sensitive to fairly small variations in σ or α, as can be seen in this figure. However, note that typically higher mass samples (e.g., WAVES) are less sensitive to α and lower mass samples (e.g., MOONS) are less sensitive to σ. This is because halo bias increases more rapidly as a function of mass at higher masses than lower masses, causing variations in high mass halo occupation to be more sensitive to the two-point correlation func-

tion. Conversely, at lower masses, there is less sensitivity to clustering signals except by varying the number of satellites, which dominate the two-point correlation function.

Our model predictions for the projected correlation function are calculated using a periodic box, for which we use the Bolshoi-Planck snapshot whose redshift is closest to the effective redshift for the given sample, as listed in Table 3. We weight our effective redshift (Equation 9) by pair counts, which scales with number times number density (Equation 10).

$$z_\mathrm{eff} = \frac{\int_{z_\mathrm{min}}^{z_\mathrm{max}} z W(z) dz}{\int_{z_\mathrm{min}}^{z_\mathrm{max}} W(z) dz} \tag{9}$$

where

$$W(z) = (dN/dz)n = \frac{(dN/dz)^2}{dV/dz} \tag{10}$$

4.3. *MCMC Fits*

We perform the measurement of $w_\mathrm{p}(r_\mathrm{p})$ (see Section 4.2) on 600 independent realizations of each mock survey, seeded by randomized orientations and origins in the Bolshoi-Planck box, using the `lightcone` code provided in the UniverseMachine package. We then calculate the mean and covariance matrix from these samples and define a six-dimensional multivariate normal likelihood distribution for our six r_p bins of w_p. We then use the `emcee` package to sample the posterior probability distribution of our HOD parameter-space: $\{\sigma, \alpha\}$, with a uniform prior confined to $10^{-5} < \sigma < 5$ and $0.1 < \alpha < 3$. We initialize our MCMC chains very close to the corresponding fiducial parameters given in Table 4, but allow them to run many autocorrelation

lengths to ensure they are well converged, as discussed in Section 5.1.

Note that the $w_{\mathrm{p}}(r_{\mathrm{p}})$ measured by a survey could be obtained more realistically by drawing one of the 600 realizations, rather than using the mean. However, from tests of additional MCMC runs, we confirm that there is no strong bias in the constraining power from using individual realizations instead of using the mean value. Therefore, we define our likelihood using the mean, which is more stable and the only fair comparison between measurements using various survey parameters. Note that cosmic variance and measurement error is still incorporated through the covariance matrix.

5. RESULTS: PREDICTIONS FOR NEXT-GENERATION SURVEYS

5.1. *Forecasts for WAVES, PFS, and MOONS*

Thanks to the small number of free parameters in our model, we obtain favorably high acceptance rates ($\sim 50\%$) and low autocorrelation lengths (~ 50), which helps reduce the time required to run our MCMC chains. To ensure we are not biased by our initial guess, we removed a burn-in of 250 iterations from the beginning of each chain, although this has a very small effect due to the long length of our chains. In each MCMC, we sample 150,000 trial points, yielding posteriors of very high resolution. We present a corner plot of the posterior measured in each mock survey in Figure 6. These posteriors show that our method does a good job of constraining our HOD to a fairly small region in parameter space. The largest difficulty is constraining the α parameter in the WAVES sample due to the high mass centrals dominating both the number density and clustering signal, and a large covariance between σ and α.

Putting together the information from the posteriors from WAVES, PFS, and MOONS, we will be able to constrain the HOD across a wide range of mass and redshift. We compile the predicted constraints on the evolution of these HOD parameters in Figure 7. Certain parameters will still be very poorly constrained using this type of analysis; for example, α in the WAVES sample. This is primarily because the two-point correlation function is most sensitive to σ at high masses and α at low masses, but additional metrics may be able to provide more information (see Appendix B).

Note that the WAVES sample produces very poor constraints on α (the satellite occupation slope parameter), whereas the MOONS sample produces particularly strong constraints on α. This demonstrates a fundamental difficulty of using constraints only from number density and the two-point correlation function. Since the two-point correlation function is primarily sensitive to the most clustered data, it is more informative for satellites at lower mass thresholds and centrals at high mass thresholds (as seen in Figure 5). However, for the WAVES sample, note that σ and α are nearly degenerate, which results in relatively poor constraints for both parameters.

It should be noted that improving our techniques may lead to tighter constraints on these regions of difficulty. First of all, it is possible to partially break the degeneracy between central and satellite clustering by measuring the two-point correlation function to smaller scales (sub-Mpc) using fiber collision corrections. Additionally, alternative statistics like counts-in-cylinders tend to be more sensitive to certain galaxy populations, and therefore provide excellent complementary information to the two-point correlation function (Wang et al. 2019). This will be important for analyzing the real data, but will come at a greater computational cost, particularly because this increases the size of the covariance matrix, and will likely require many more mock realizations to calculate accurately.

5.2. *Measurement Error vs. Survey Parameters*

Telescope time is typically the limiting factor in survey design. To first order, the amount of time a survey requires is roughly proportional to the number of objects observed. It is therefore possible to either scale up the sky area in exchange for a decrease in **targeting** completeness or vice versa **(e.g., if only 30% of the targets were assigned to fibers on the first pointing, one can either repeat the same patch of the sky to double the completeness up to 60%, or instead double the area by observing a different patch of the sky of the same size)**. Increasing the sky area of a survey would increase the volume and therefore decrease the cosmic sample variance; on the other hand, high **targeting** completeness helps mitigate the uncertainty of small-scale pair counts, especially when accounting for fiber collisions (Bianchi & Percival 2017). Although this is not tested in this work, it should also be noted that higher **targeting** completeness should be favored if the goal is to increase the accuracy of identifying central galaxies in group reconstruction (Looser et al. 2021).

Using our mock catalogs to estimate uncertainties, we vary these survey parameters to quantify their effects on constraining power. In Figure 8, we present the dependence of survey parameters on the uncertainty of the projected two-point correlation function. For PFS- and MOONS-like surveys, the only way to greatly reduce uncertainties is by increasing the area (see top row). Increasing the number of targets in the same area makes much more modest improvements (see bottom row). In
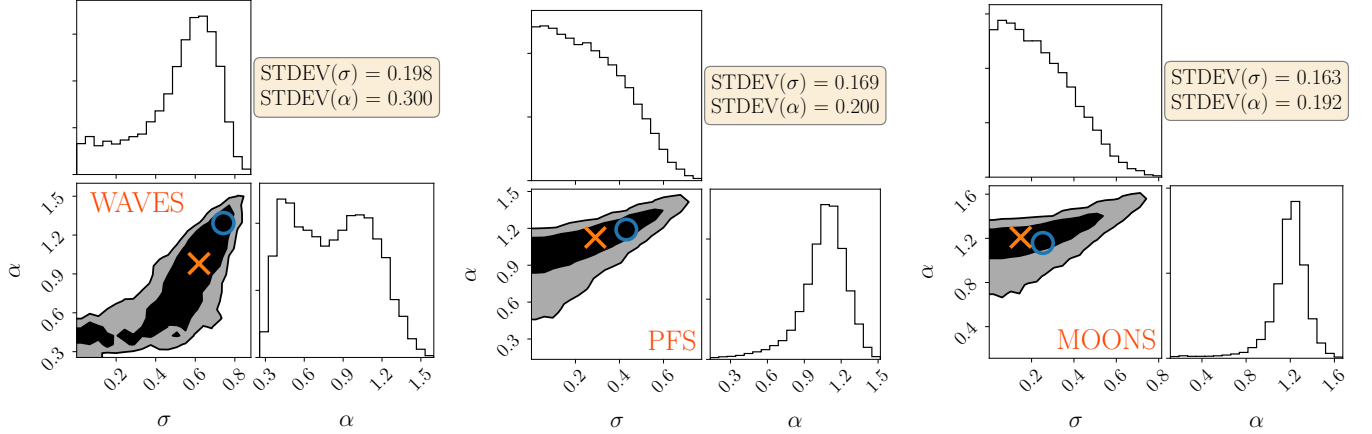
Figure 6. HOD posterior probability distribution measured in WAVES (left), PFS (center), and MOONS (right). Measured by MCMC sampling of our two-parameter conservative HOD model at the effective redshift of each sample, and comparing the predicted $w_{\mathrm{p}}(r_{\mathrm{p}})$ to 600 mock realizations. For each sample, the UniverseMachine truth value is marked with a blue circle and best-fit parameters are marked with an orange X.
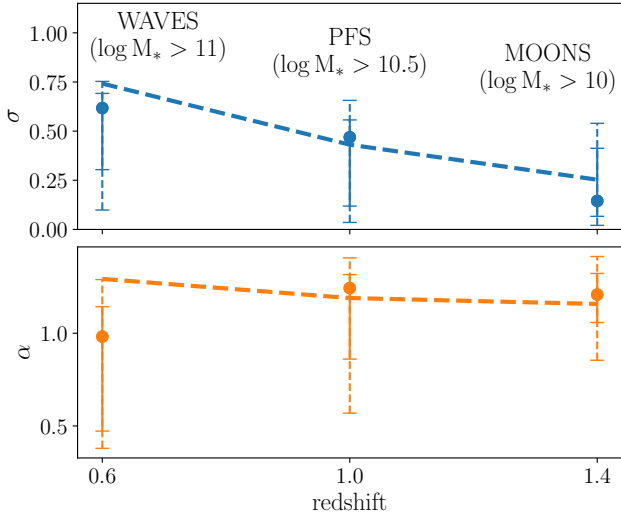

Figure 7. Predicted constraints on the evolution of the HOD. We compile best fit measurements along with 5th, 16th, 84th, and 95th percentile of the posterior (see Figure 6) of σ (upper panel) and α (lower panel). UniverseMachine "truth" values are connected by dashed lines.

other words, smaller area surveys like PFS and MOONS are dominated by cosmic variance, not shot noise. Wider surveys like WAVES (left panel) typically find that the two-point correlation function uncertainty is dominated by cosmic variance on large scales (9-27 h^{-1} Mpc) and shot noise on small scales (1-3 h^{-1} Mpc).

To roughly quantify the effect survey parameters have on correlation function uncertainties, we calculate the percent decrease in w_{p} error at $r_{\mathrm{p}} = 3\text{-}9\ h^{-1}$ Mpc from the true survey parameters for two cases: (1) doubled area and **halved targeting** completeness (conserving the number of targets) and (2) doubled **targeting** com-

pleteness and the same area. **Note that we do not explicitly test doubling these survey parameters, as it is not possible to increase the targeting completeness above 100%. However, it is still useful to consider the effective reduction in cosmic variance vs. shot noise, we make use of** linear regression to fit the slope of the orange lines in Figure 8 to extrapolate these numbers. We find that doubling the survey area decreases the uncertainty of the correlation function at intermediate scales by 3%, 24%, and 22% for our WAVES, PFS, and MOONS samples, respectively. For a fixed survey area, doubling the number of targets improves the same constraints by 14%, 5%, and 2%. These numbers reinforce that the samples probed by PFS and MOONS are dominated by cosmic variance and can only be improved significantly by increasing the observing area.

Given the covariance matrix of $w_{\mathrm{p}}(r_{\mathrm{p}})$ calculated for each set of survey parameters, we also calculate HOD constraints via our MCMC method. We present the HOD constraining power as a function of survey parameters by varying **targeting** completeness and area in Figure 9. In the WAVES survey, we only find very small changes in constraints as we vary the survey parameters. For PFS and MOONS, the difference between Figures 9 and 8 indicates that the covariance of the two-point correlation function with respect to various regions of the universe is significantly different from the covariance due to varying HOD parameters. It appears that this allows the HOD to be somewhat more robust to cosmic variance than w_{p}. Throughout, the constraints on α may be slightly more dependent on survey parameters than the constraints on σ.
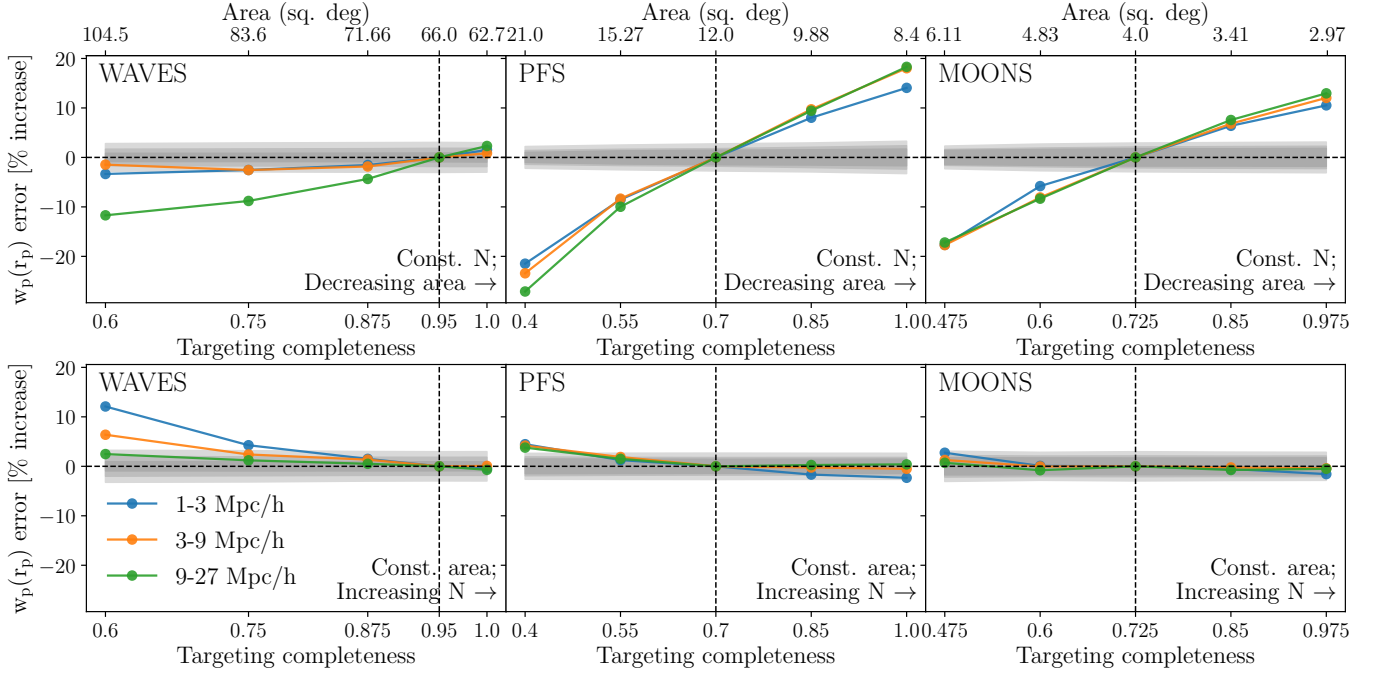
Figure 8. Precision of mock $w_{\rm p}$ measurements, color-coded by $r_{\rm p}$ scale, as a function of **targeting** completeness. Characteristic jackknife uncertainties are shown with grey bands. In the top panels, the sky area is varied to conserve the total number of targets (given in Table 3). In the bottom panels, the sky area is conserved at the true value for each survey, and therefore the number of targets increases with **targeting** completeness. Particularly for the PFS and MOONS samples, there are significantly stronger trends in the top panels. This suggests that the uncertainties in $w_{\rm p}(r_{\rm p})$ are dominated by cosmic variance, as opposed to shot noise.
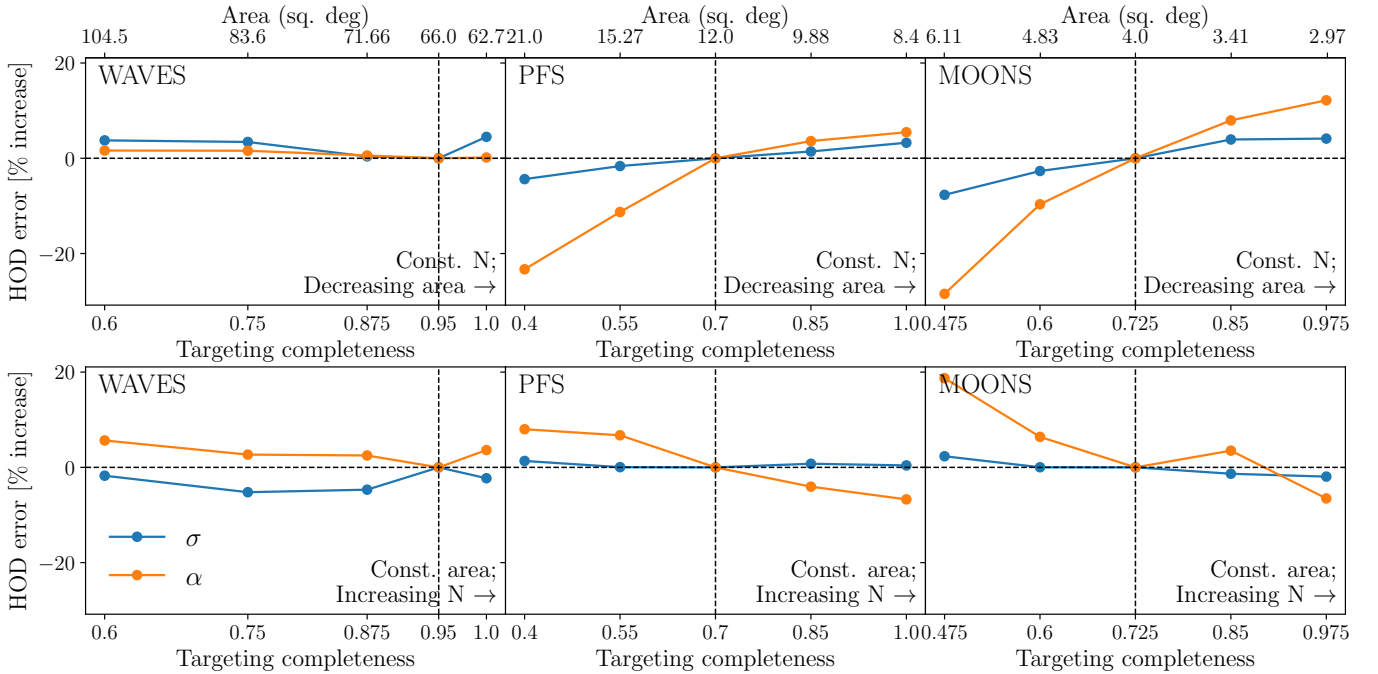
Figure 9. Precision of HOD parameters σ (blue) and α (orange), as a function of **targeting** completeness. The panels are arranged analogously to Figure 8. These uncertainties do not appear to be dominated by cosmic variance as strongly as the two-point correlation function. They are affected just as strongly by shot noise. This demonstrates the importance of propagating uncertainties from the full covariance matrix, rather than just the diagonal components shown in Figure 8.

678   5.3. *Comparisons to Past Surveys*

679   Surveys like WAVES, PFS, and MOONS will be mon-
680   umental in pushing measurements of the two-point cor-

relation function to higher redshifts because they will provide the precise spectroscopic measurements necessary to perform those calculations. There are currently no existing spectroscopic datasets that are comparable in size to the $z > 1$ samples we will obtain from PFS and MOONS.

For the slightly lower redshifts probed by WAVES, the closest comparison would be prism surveys such as the PRIsm MUlti-object Survey (PRIMUS; Coil et al. 2011) or the Carnegie-Spitzer-IMACS (CSI; Kelson et al. 2014). PRIMUS, the larger of these two surveys, has measured the spectra of galaxies out to $z \sim 1.2$ with a redshift precision of $\sigma_z/(1+z) \sim 0.005$. Incorporating all fields that overlap with imaging from the Galaxy Evolution Explorer (GALEX; Martin et al. 2005), Spitzer Space Telescope (Werner et al. 2004), Infrared Array Camera (IRAC; Fazio et al. 2004), and various ground-based surveys, PRIMUS has an area of 5.5 deg^2 and is complete down to similar stellar masses as WAVES (see Moustakas et al. 2013).

We compare mock measurements of the projected two-point correlation function of our WAVES galaxy sample ($0.5 < z < 0.8$ and $\log(M_*/M_\odot) > 11$) for the survey parameters of WAVES and PRIMUS in Figure 10. This plot illustrates the significantly increased precision we can expect to obtain from this sample of galaxies. Additionally, unlike PRIMUS, we assume that the redshift uncertainties in WAVES will be negligible compared to redshift distortions. The additional redshift error from PRIMUS does not greatly contribute to the errorbars of the two-point correlation function, but this does produce a small systematic offset which may further reduce the correlation function's sensitivity to HOD parameters.

However, most of our current understanding of the galaxy-halo connection comes from studies of surveys that either span lower redshifts or rely on photometric redshifts. For example, Zu & Mandelbaum (2015) use photometric redshifts from SDSS (Blanton et al. 2017) and Leauthaud et al. (2012) from COSMOS (Scoville et al. 2007). In Figure 11, we present key findings of these past studies in terms of the SHMR, absolute bias, and satellite fraction of several stellar mass threshold galaxy samples and compare to the same measurements derived from our HOD projections of the WAVES, PFS, and MOONS surveys. These upcoming surveys will push to significantly deeper redshifts than past surveys, with comparable uncertainties.

Measuring the HOD for various surveys at distinct redshifts and stellar mass thresholds has been and will continue to be a powerful tool for studying galaxy evolution as measured by mean halo mass, satellite fraction, and galaxy bias (see Figure 11). Currently, we have lit-
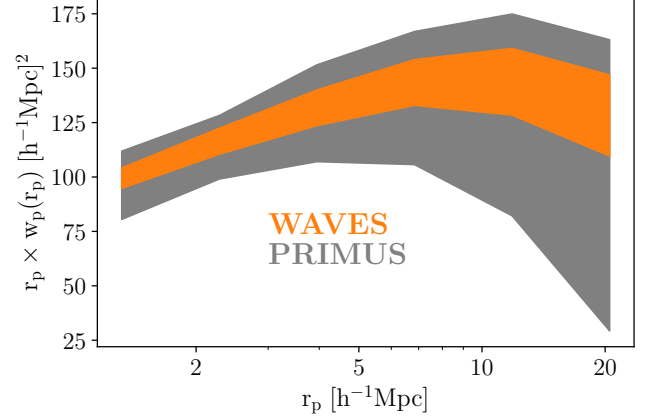
Figure 10. Mock measurements of the projected two-point correlation function for WAVES vs. PRIMUS. Note that we only used 25 independent realizations to quantify the observational error of the PRIMUS measurement, compared to the 600 used for WAVES. The increased precision from WAVES is due to its area (66 deg^2) which is over 10 times that of PRIMUS (5.5 deg^2). A small systematic offset is driven by redshift errors in PRIMUS, which we assume will be negligible compared to velocity distortions in WAVES.

tle evidence for significant redshift evolution in most of these properties. However, more precise and higher redshift measurements of these parameters will give us a much clearer picture of how they may evolve with the age of the universe. Each of these metrics is highly sensitive to various models of galaxy formation, so these new measurements will have a large impact on how we think about the important processes driving the shut down of rapid star formation that occurred at cosmic noon.

6. CONCLUSIONS

In this paper, we present the CLIMBER procedure, which we use to calibrate photometry into the UniversemMachine and similar models. This procedure performs well at reproducing a broad range of properties simultaneously. Twenty-five realizations of the $0.7 < z < 1.7$ mock catalog used for the PFS and MOONS samples in this work are available at https://alanpearl.github.io/#data. Alternatively, you may install the utilities we used to construct these catalogs at https://github.com/AlanPearl/mocksurvey.

We have used our mock catalogs to test the forthcoming generation of massively multiplexed spectroscopic galaxy surveys, which will likely change our understanding of galaxy formation, the galaxy-halo connection, and possibly even cosmology in profound ways. The high-redshift samples being probed will provide new constraints on theories and models which predict how populations evolve with redshift. The UniverseMachine will face new scrutiny in its ability to reproduce cluster-
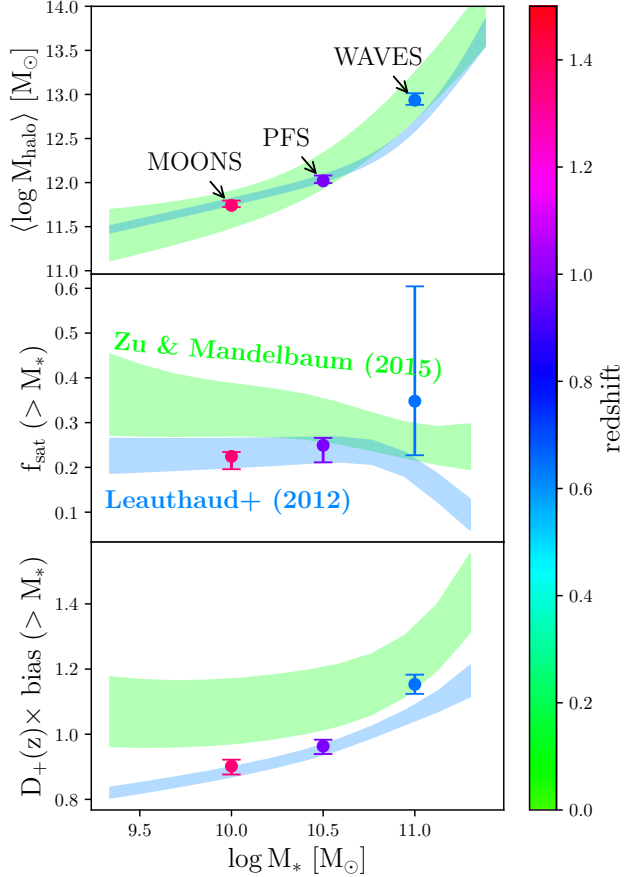
Figure 11. Compiled measurements of the galaxy-halo connection. As a function of stellar mass, we show several studies that have measured the mean halo mass (top panel), satellite fraction above the stellar mass threshold (middle panel), and absolute bias above the stellar mass threshold (bottom panel). Projected HOD analysis from this work for WAVES, PFS, and MOONS is shown by the colored points, which push to significantly higher redshifts with comparable uncertainties to previous studies.

ing and environmental quenching signals in the distant universe. The constraints on parameters of interest in the UniverseMachine will likely tighten significantly to match the new observations. It may even be possible that the UniverseMachine will require reparameterization in order to achieve a good fit to the new data. Either way, analysis of these new surveys will greatly impact our understanding of the evolution of the galaxy population's connection to its dark matter environment over the history of the universe.

Using the two-point correlation function, we have found that surveys such as WAVES, PFS, and MOONS will place new constraints on the galaxy-halo connection. We characterize constraints on the central term of the HOD with the parameters σ and $\log M_{\min}$. The

precision to which we measure these parameters is displayed in Figure 6. We have found that studies of lower mass galaxies, like the MOONS sample, will not achieve strong constraints from $w_{\mathrm{p}}(r_{\mathrm{p}})$ alone and will likely need to use other counts-in-cells statistics that are more sensitive to the weak clustering signal of low-mass centrals.

We characterize constraints on the satellite term of the HOD with the parameters α and $\log M_1$. The precision of these measurements is also displayed in Figure 6. We find that the satellite term of the HOD will be most poorly constrained in high-mass samples, where the clustering signal is dominated by centrals. In this regime, small-scale correlation function measurements are the most sensitive to the satellite occupation of halos. Smaller-scale measurements will be possible from these surveys using fiber collision corrections, but those will likely come at the cost of large shot noise. Therefore, it is still important for follow-up surveys to improve the **targeting** completeness of these galaxy samples and reduce the effect of fiber collisions.

The achievable constraining power of these parameters is dependent on survey parameters, such as **targeting** completeness (which reduces shot noise) and area (which reduces cosmic variance). Our conclusions can be summarized by the following key points:

- The two-point correlation function measurements from PFS and MOONS are both primarily dominated by cosmic variance, rather than shot noise. We have shown that, with fixed sample size, increasing their survey area drastically reduces this uncertainty.

- The HOD constraints from PFS and MOONS are less dominated by cosmic variance. This demonstrates the importance of using the full covariance matrix to calculate HOD constraints.

- From WAVES, there is a more balanced combination of shot noise, which is dominant on small scales ($1-3$ h^{-1} Mpc) and cosmic variance, which is dominant on large scales ($9 - 27$ h^{-1} Mpc). The resulting HOD constraints are not strongly affected by small changes in survey parameters.

Another important survey parameter, which has not been explored in this work, is the number of independent fields. PFS and MOONS are both planning on dividing their survey into several fields, which will slightly mitigate some of their large cosmic variance. Additionally, our predicted future constraints could be improved by supplementing the correlation function with counts-in-cylinders, which is more sensitive to the weak clustering of low-mass centrals. Simulating detailed targeting

strategies may also be important for more precise optimizations of constraining power, as this is necessary to calculate pair counts corrections at very small scales due to fiber collisions. This is important for unbiased estimates of both the two-point correlation function and counts-in-cylinders.

There are sure to be many discoveries in store thanks to this new generation of surveys. This new data will be investigated from all angles of the galaxy-halo connection: empirical models, SAMs, and hydrodynamic simulations alike will be put to the test. We hope that through this combined effort and publicly available tools like our mock catalogs, we will be able to utilize this data to its full potential.

REFERENCES

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123, doi: 10.3847/1538-3881/aabc4f

Behroozi, P., Wechsler, R. H., Hearin, A. P., & Conroy, C. 2019, MNRAS, 488, 3143, doi: 10.1093/mnras/stz1182

Bell, E. F., & de Jong, R. S. 2001, ApJ, 550, 212, doi: 10.1086/319728

Berlind, A. A., & Weinberg, D. H. 2002, ApJ, 575, 587, doi: 10.1086/341469

Bianchi, D., & Percival, W. J. 2017, MNRAS, 472, 1106, doi: 10.1093/mnras/stx2053

Blanton, M. R., Bershady, M. A., Abolfathi, B., et al. 2017, AJ, 154, 28, doi: 10.3847/1538-3881/aa7567

Blumenthal, G. R., Faber, S. M., Primack, J. R., & Rees, M. J. 1984, Nature, 311, 517, doi: 10.1038/311517a0

Bond, J. R., Cole, S., Efstathiou, G., & Kaiser, N. 1991, ApJ, 379, 440, doi: 10.1086/170520

Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000, doi: 10.1046/j.1365-8711.2003.06897.x

Chabrier, G. 2003, PASP, 115, 763, doi: 10.1086/376392

Coil, A. L., Blanton, M. R., Burles, S. M., et al. 2011, ApJ, 741, 8, doi: 10.1088/0004-637X/741/1/8

Crain, R. A., Schaye, J., Bower, R. G., et al. 2015, MNRAS, 450, 1937, doi: 10.1093/mnras/stv725

Davé, R., Anglés-Alcázar, D., Narayanan, D., et al. 2019, MNRAS, 486, 2827, doi: 10.1093/mnras/stz937

Driver, S. P., Davies, L. J., Meyer, M., et al. 2016, in The Universe of Digital Sky Surveys, ed. N. R. Napolitano, G. Longo, M. Marconi, M. Paolillo, & E. Iodice, Vol. 42, 205

Fabian, A. C. 2012, ARA&A, 50, 455, doi: 10.1146/annurev-astro-081811-125521

Fazio, G. G., Hora, J. L., Allen, L. E., et al. 2004, ApJS, 154, 10, doi: 10.1086/422843

Fierlinger, K. M., Burkert, A., Ntormousi, E., et al. 2016, MNRAS, 456, 710, doi: 10.1093/mnras/stv2699

Foreman-Mackey, D. 2016, The Journal of Open Source Software, 1, 24, doi: 10.21105/joss.00024

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306, doi: 10.1086/670067

Guo, Q., White, S., Boylan-Kolchin, M., et al. 2011, MNRAS, 413, 101, doi: 10.1111/j.1365-2966.2010.18114.x

Hearin, A. P., & Watson, D. F. 2013, MNRAS, 435, 1313, doi: 10.1093/mnras/stt1374

Hearin, A. P., Zentner, A. R., van den Bosch, F. C., Campbell, D., & Tollerud, E. 2016, MNRAS, 460, 2552, doi: 10.1093/mnras/stw840

Hopkins, P. F., Wetzel, A., Kereš, D., et al. 2018, MNRAS, 480, 800, doi: 10.1093/mnras/sty1690

Hunter, J. D. 2007, Computing in Science and Engineering, 9, 90, doi: 10.1109/MCSE.2007.55

Kaiser, N. 1984, ApJL, 284, L9, doi: 10.1086/184341

Kajisawa, M., Ichikawa, T., Yamada, T., et al. 2010, ApJ, 723, 129, doi: 10.1088/0004-637X/723/1/129

Kelson, D. D., Williams, R. J., Dressler, A., et al. 2014, ApJ, 783, 110, doi: 10.1088/0004-637X/783/2/110

Klypin, A., Yepes, G., Gottlöber, S., Prada, F., & Heß, S. 2016, MNRAS, 457, 4340, doi: 10.1093/mnras/stw248

Kravtsov, A. V., Berlind, A. A., Wechsler, R. H., et al. 2004, ApJ, 609, 35, doi: 10.1086/420959

Kriek, M., van Dokkum, P. G., Labbé, I., et al. 2009, ApJ, 700, 221, doi: 10.1088/0004-637X/700/1/221

Landy, S. D., & Szalay, A. S. 1993, ApJ, 412, 64, doi: 10.1086/172900

Leauthaud, A., Tinker, J., Bundy, K., et al. 2012, ApJ, 744, 159, doi: 10.1088/0004-637X/744/2/159

Leja, J., Johnson, B. D., Conroy, C., et al. 2019, ApJ, 877, 140, doi: 10.3847/1538-4357/ab1d5a

Leslie, S. K., Schinnerer, E., Liu, D., et al. 2020, ApJ, 899, 58, doi: 10.3847/1538-4357/aba044

Looser, T. J., Lilly, S. J., Sin, L. P. T., et al. 2021, MNRAS, 504, 3029, doi: 10.1093/mnras/stab1061

Maiolino, R., Cirasuolo, M., Afonso, J., et al. 2020, The Messenger, 180, 24, doi: 10.18727/0722-6691/5197

Martin, D. C., Fanson, J., Schiminovich, D., et al. 2005, ApJL, 619, L1, doi: 10.1086/426387

Mo, H. J., & White, S. D. M. 1996, MNRAS, 282, 347, doi: 10.1093/mnras/282.2.347

Moustakas, J., Coil, A. L., Aird, J., et al. 2013, ApJ, 767, 50, doi: 10.1088/0004-637X/767/1/50

Muzzin, A., Marchesini, D., van Dokkum, P. G., et al. 2009, ApJ, 701, 1839, doi: 10.1088/0004-637X/701/2/1839

Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013, ApJS, 206, 8, doi: 10.1088/0067-0049/206/1/8

Nelson, D., Springel, V., Pillepich, A., et al. 2019, Computational Astrophysics and Cosmology, 6, 2, doi: 10.1186/s40668-019-0028-x

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2012, arXiv e-prints, arXiv:1201.0490. https://arxiv.org/abs/1201.0490

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, A&A, 641, A6, doi: 10.1051/0004-6361/201833910

Press, W. H., & Schechter, P. 1974, ApJ, 187, 425, doi: 10.1086/152650

Reddick, R. 2013, arXiv e-prints, arXiv:1304.0240. https://arxiv.org/abs/1304.0240

Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, MNRAS, 446, 521, doi: 10.1093/mnras/stu2058

Scoville, N., Aussel, H., Brusa, M., et al. 2007, ApJS, 172, 1, doi: 10.1086/516585

Sinha, M., & Garrison, L. H. 2020, MNRAS, 491, 3022, doi: 10.1093/mnras/stz3157

Slepian, Z., Eisenstein, D. J., Beutler, F., et al. 2017, MNRAS, 468, 1070, doi: 10.1093/mnras/stw3234

Somerville, R. S., & Davé, R. 2015, ARA&A, 53, 51, doi: 10.1146/annurev-astro-082812-140951

Takada, M., Ellis, R. S., Chiba, M., et al. 2014, PASJ, 66, R1, doi: 10.1093/pasj/pst019

Tomczak, A. R., Quadri, R. F., Tran, K.-V. H., et al. 2016, ApJ, 817, 118, doi: 10.3847/0004-637X/817/2/118

van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, Computing in Science and Engineering, 13, 22, doi: 10.1109/MCSE.2011.37

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261, doi: 10.1038/s41592-019-0686-2

Wang, K., Mao, Y.-Y., Zentner, A. R., et al. 2019, MNRAS, 488, 3541, doi: 10.1093/mnras/stz1733

Wechsler, R. H., & Tinker, J. L. 2018, ARA&A, 56, 435, doi: 10.1146/annurev-astro-081817-051756

Werner, M. W., Roellig, T. L., Low, F. J., et al. 2004, ApJS, 154, 1, doi: 10.1086/422992

Whitaker, K. E., Pope, A., Cybulski, R., et al. 2017, ApJ, 850, 208, doi: 10.3847/1538-4357/aa94ce

Whitaker, K. E., Franx, M., Leja, J., et al. 2014, ApJ, 795, 104, doi: 10.1088/0004-637X/795/2/104

White, S. D. M., & Frenk, C. S. 1991, ApJ, 379, 52, doi: 10.1086/170483

White, S. D. M., & Rees, M. J. 1978, MNRAS, 183, 341, doi: 10.1093/mnras/183.3.341

Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2005, ApJ, 630, 1, doi: 10.1086/431891

Zentner, A. R. 2007, International Journal of Modern Physics D, 16, 763, doi: 10.1142/S0218271807010511

Zheng, Z., Coil, A. L., & Zehavi, I. 2007, ApJ, 667, 760, doi: 10.1086/521074

Zu, Y., & Mandelbaum, R. 2015, MNRAS, 454, 1161, doi: 10.1093/mnras/stv2062
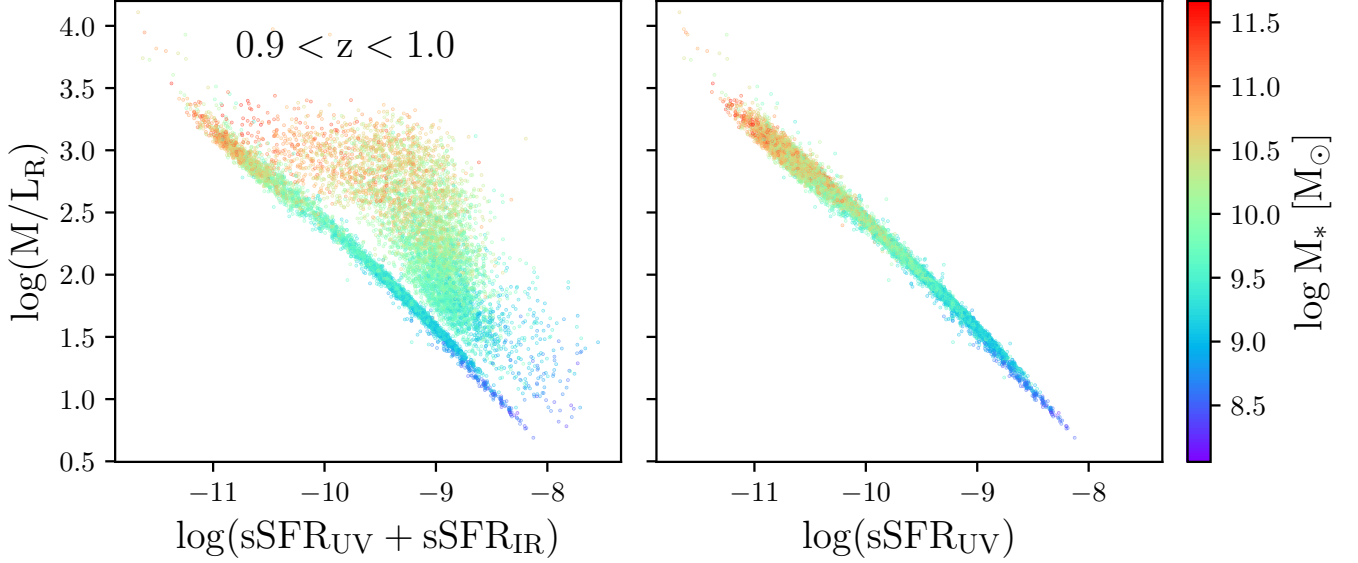
Figure 12. The relation between specific star formation rate and mass-to-light ratio for UltraVISTA galaxies in the redshift slice $0.9 < z < 1.0$. In the left panel, we use total sSFR (UV + IR), and in the right panel, we only use UV sSFR. The two plots share in common all galaxies without any IR detection, but the galaxies with IR detection form a cloud to the right due to their higher total sSFR (removing galaxies without IR detection removes the remaining narrow distribution entirely). We choose not to include sSFR$_{IR}$ in our mass-to-light calibration due to this discontinuity between detections and non-detections.

APPENDIX

A. CLIMBER DETAILS

The goal of Calibrating Light: Illuminating Mocks By Empirical Relations (CLIMBER) is to estimate the luminosity of each mock galaxy in any observed photometric band. Since star-forming galaxies host more young blue stars, color is a smooth function of specific SFR (sSFR). Both of these quantities correlate strongly with the mass-to-light ratio, as shown by Bell & de Jong (2001). The relationship between mass-to-light ratio in each band and sSFR is approximately a power law that can be empirically calibrated.

However, we first need to ensure self-consistency between the model and empirical parameters. In the model, SFRs for the star-forming population are drawn from mass-dependent distribution (the star-forming main sequence) which evolves with redshift. The star-forming main sequence is matched to empirical distributions at similar redshifts to ensure the values are physical. However, for the quiescent population, the SFRs are drawn from a non-evolving log-normal distribution centered around an sSFR of $10^{-11.8}$ yr^{-1}. Given that this is inconsistent with empirical assumptions past $z \sim 0$, we adopt only their rank-ordering. For each mock galaxy, we first map its sSFR to an empirically calibrated value

of ultraviolet sSFR (sSFR$_{UV}$). We choose sSFR$_{UV}$ due to its tight correlation with mass-to-light ratio and reliable measurements in both quiescent and star-forming galaxies.

It would be even better to use the total sSFR by adding infrared (IR) sSFR, but this is not possible for the UltraVISTA dataset because 53% of our training data have no detection in the IR. This creates an artificial discontinuity between IR detections and non-detections, thereby forming two distinct populations in the sSFR-M/L plane (see Figure 12). Including this discontinuity would greatly increase the difficulty in mapping from sSFR to M/L, particularly because it would require knowing the stellar mass to know which population a galaxy is a part of (this is primarily an observational effect in which lower mass galaxies at the same sSFR are less likely to have infrared detections). Including the stellar mass as a feature in our random forest is undesirable because it could have unknown consequences when extrapolating to lower stellar masses. Due to our decision to map sSFR to sSFR$_{UV}$, we warn that our mocks may underpredict the spread in the distribution of M/L at a given sSFR. However, we believe that

this is the best option due to its continuity and accuracy in reproducing magnitude and color distributions.

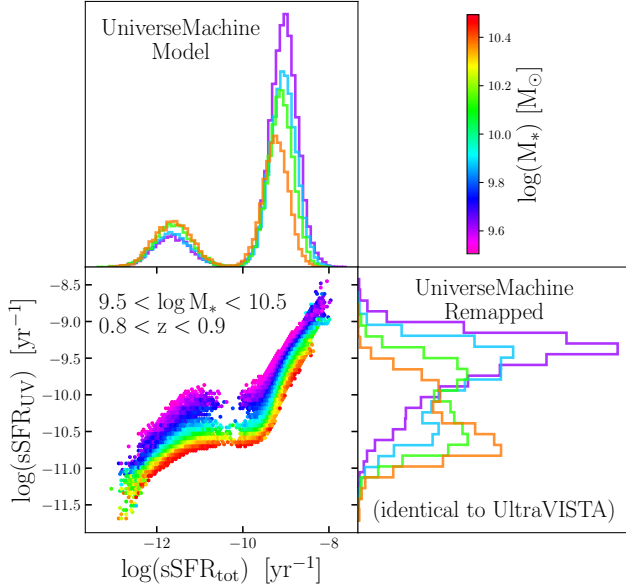**Figure 13.** Conditional abundance-matching mapping from sSFR → sSFR$_{UV}$ for UniverseMachine mock galaxies for a range of stellar masses at the redshift slice $0.8 < z < 0.9$. After being remapped, the distribution is forced to be identical to that of UltraVISTA at fixed stellar mass and redshift. Note that there is a near one-to-one mapping for star-forming galaxies, but the very low sSFRs of quiescent UniverseMachine galaxies are shifted up significantly.
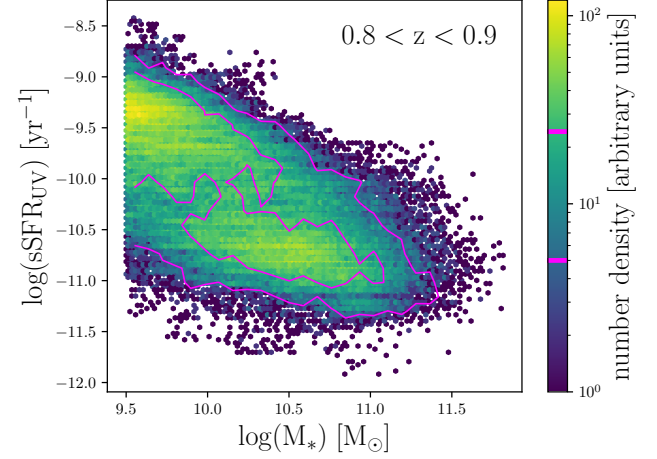
The mapping of sSFR → sSFR$_{UV}$ is not uniform because UV flux decreases with higher dust obscuration, which is strongly dependent on stellar mass (Whitaker et al. 2017). Therefore, we map sSFR → sSFR$_{UV}$ through conditional abundance-matching (CAM) using the code `conditional_abunmatch` from the `halotools` package, which preserves the rank-ordering at fixed stellar mass (tolerance of ~0.05 dex). We iterate this method in fuzzy redshift bins (width of ~0.1) using the code `fuzzy_digitize` from the `halotools` package. We match the distribution to identical photometric redshift bins of the UltraVISTA survey (Muzzin et al. 2013).

A visualization of the CAM mapping from the UniverseMachine sSFR to UltraVISTA-calibrated sSFR$_{UV}$ is shown in Figure 13. **In the lower left panel of this figure, note that this mapping appears nearly linear for star-forming galaxies (sSFR$_{tot}$ $\gtrsim 10^{-10}$), especially at low masses where IR light is insignificant. The large gap between star-forming and quiescent galaxies (sSFR$_{tot}$ $\lesssim 10^{-10}$) is because the UniverseMachine quiescent galaxies' SFR is not fit to data past $z = 0$. By construction, CAM**

Figure 14. The 2D distribution of M_*-sSFR$_{UV}$. Magenta lines represent logarithmic contours of the UltraVISTA training data, while the colored points represent logarithmic counts of UniverseMachine mock galaxies. These distributions are nearly identical by construction, via CAM. Accurately capturing this distribution is crucial for assessing mass completeness under the assumption that sSFR$_{UV}$ controls the mass-to-light ratio.
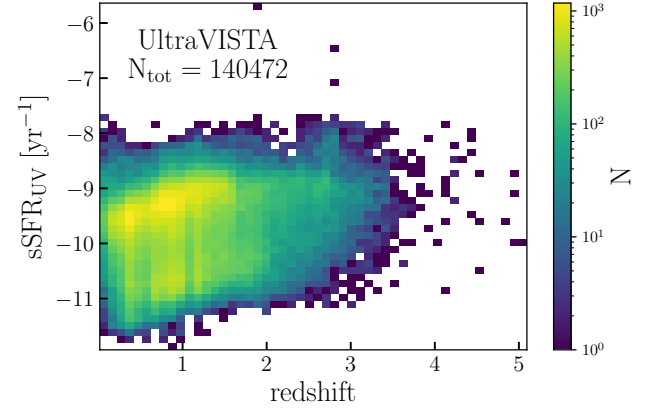

Figure 15. 2D histogram of the UltraVISTA data used to train mass-to-light ratios in our random forest. There are very few regions of parameter space without any data, which makes this an ideal dataset up to $z < 3$.

ensures that the resulting M_*-sSFR$_{UV}$ distribution matches the data it is calibrated to (see Figure 14), which is crucial for assessing mass completeness under the assumption that sSFR$_{UV}$ controls the mass-to-light ratio.

We train the mapping from sSFR$_{UV}$ to mass-to-light ratio using the photometry and FAST stellar masses from UltraVISTA. We plot the feature-space of the UltraVISTA training data in Figure 15. These 140,472 training data leave very few missing regions of feature-

space for $z < 3$, making it an ideal training set for our purposes. The random forest regression (the `RandomForestRegressor` class from the `scikit-learn` package; Pedregosa et al. 2012) is then used to predict $\log(M_*/L)$ from the two features $\{\log \mathrm{sSFR_{UV}}, z\}$. The advantages of this approach are its simplicity, flexibility, and sufficient accuracy in predicting this relation and its intrinsic scatter (see Figure 16 and 17). Additionally, this method automatically includes covariance between mass-to-light ratios of different photometric bands, which is important to accurately capture distributions of color as well as multivariate color-magnitude distributions. While random forests are not particularly good at extrapolation, our mock does not need to extrapolate in feature-space. Since we do not use stellar mass as a predictor of M/L (except indirectly through conditional abundance matching), it generalizes reasonably to masses slightly below the completeness limit of UltraVISTA, although predictions for galaxies far below this limit should be used with care.
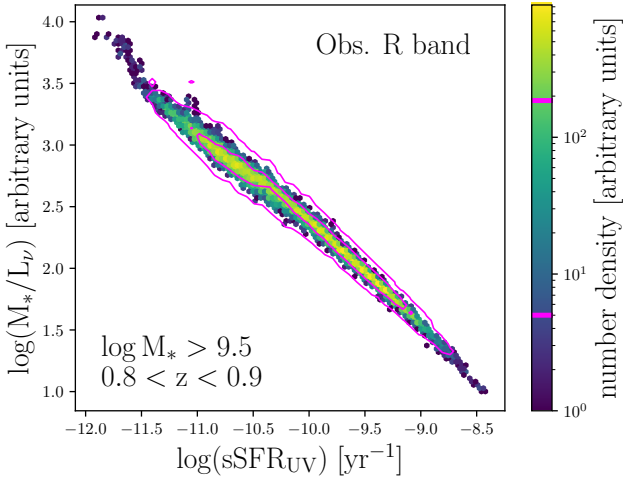

Figure 16. The relation between UV specific star formation rate and mass-to-light ratio in the *observed* R band at the redshift slice $0.8 < z < 0.9$. Magenta lines represent logarithmic contours of the UltraVISTA training data, while the colored points represent logarithmic counts of UniverseMachine mock galaxies which were fit via the Random Forest method described in Appendix A. We present the same relation in all other available photometric bands in Figure 17.

To assess our mock catalog's performance at matching the UltraVISTA photometry, we compare the predicted distributions for a variety of properties, including the colors shown in Figure 18. The mass function and luminosity functions match very well by construction, but we also want to accurately reproduce color distributions, as the selection functions in real surveys will often use color cuts in addition to magnitude cuts. We have tested a variety of choices for the random forest hyperparameters used, but found that the default scikit-learn values performed best at reproducing a broad range of properties simultaneously. The hyperparameter values we use are therefore: `n_estimators` $= 10$, `bootstrap` $=$ `True`, `max_depth` $=$ `None`, `max_features` $=$ `"auto"`, `min_samples_leaf` $= 1$, and `min_samples_split` $= 2$.

We have applied a number of common machine learning validation methods including 5-fold cross-validation testing and learning curve analysis, using the mean absolute deviation (MAD) of colors as our loss function, and find that these hyperparameter values yield a model that is modestly overfitted (cross-validated MAD value of 0.150, versus a training score of 0.058). We calculated an alternative set of optimized hyperparameters by minimizing the mean absolute deviation of the colors (via a random hyperparameter search followed by a smaller, but exhaustive, grid search), which yielded a more converged learning curve, indicating less overfitting (cross-validated MAD value of 0.130, versus a training score of 0.126). However, while the optimized hyperparameters perform better for predictions of individual colors, the M/L values at fixed mass concentrate closely around the mean value rather than capturing the full distribution. This results in worse color distributions (as seen in the bottom panel of Figure 18). Since we are in a regime where the feature-space distribution of the training set is nearly identical to that of the mock (by construction via conditional abundance matching), the overfitting when using the default parameters actually helps us, since it guarantees that the magnitude and color distributions in the mock match those of the training data.

B. ADDITIONAL METRICS

While the two-point correlation function provides quite good HOD constraints, it is also very sensitive to cosmology; particularly the σ_8 parameter, which controls the linear bias of halos. It has been shown (Slepian et al. 2017) that the three-point correlation function would break this degeneracy, at least on linear scales. Calculating the three-point correlation function with existing public codes is too expensive to run at each iteration of an MCMC chain in our analysis. If a more efficient implementation becomes available, we would be interested in including this statistic to compare the added constraining power. This statistic will be especially important for joint analyses of cosmology and the galaxy-halo connection.

Additionally, using the two-point correlation function alone can miss out on important clustering information, as it is primarily driven by the most clustered galaxies,
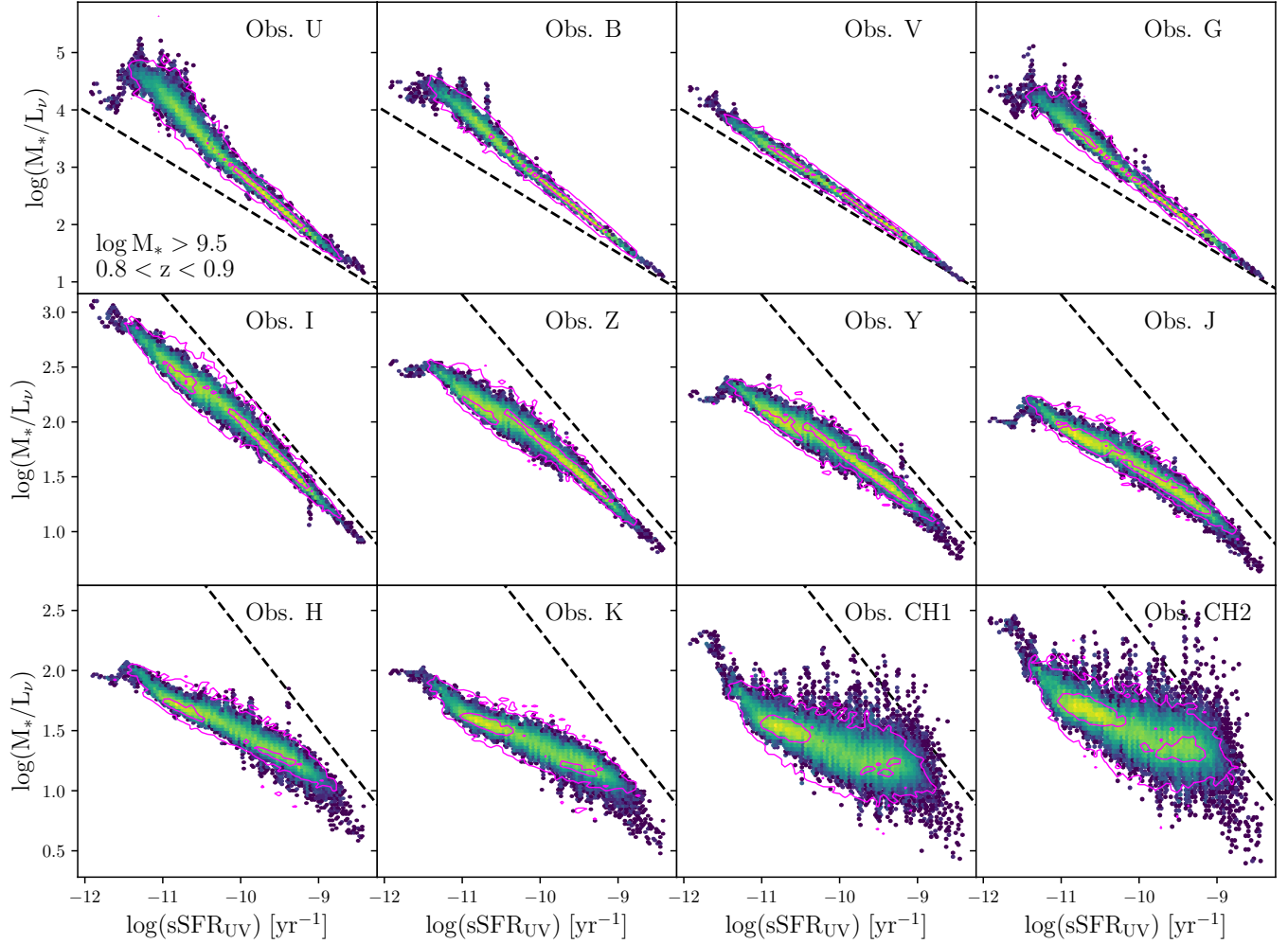
Figure 17. The relation between UV specific star formation rate and mass-to-light ratio in the redshift slice $0.8 < z < 0.9$. Same as Figure 16 but the y-axes represent the mass-to-light ratio in all other available photometric bands. The dashed line in each panel is the best-fit line for the R-band as a reference point.

which is almost always satellites except at the highest masses. One can explicitly measure a signal from central galaxies singling them out via isolation criteria or group catalog reconstruction. One metric in particular that is capable of measuring assembly bias signals is the number of satellites vs. central stellar mass, split into quenched and star-forming populations (shown in private communication with Rodríguez-Puebla). It is unclear how accurately group catalogs can be reconstructed from surveys like PFS and MOONS, which are incomplete due to fiber collisions, and require the use of photometric redshifts to fill in the gaps. This may prove to be a strong argument for extensions to these surveys prioritizing increasing the **targeting** completeness over the area, contrary to the cosmic variance tests herein.

Alternatively, it is possible to supplement the two-point correlation function with other statistics such as counts-in-cylinders (CIC). By tuning the radius of the cylinder and inner annulus, CIC can effectively separate the signal of centrals from satellites, without requiring full group catalog reconstruction. Wang et al. (2019) demonstrate that CIC + $w_{\rm p}(r_{\rm p})$ may be sufficient to constrain assembly bias as well. However, further testing is required for samples affected by fiber collisions like PFS and MOONS.
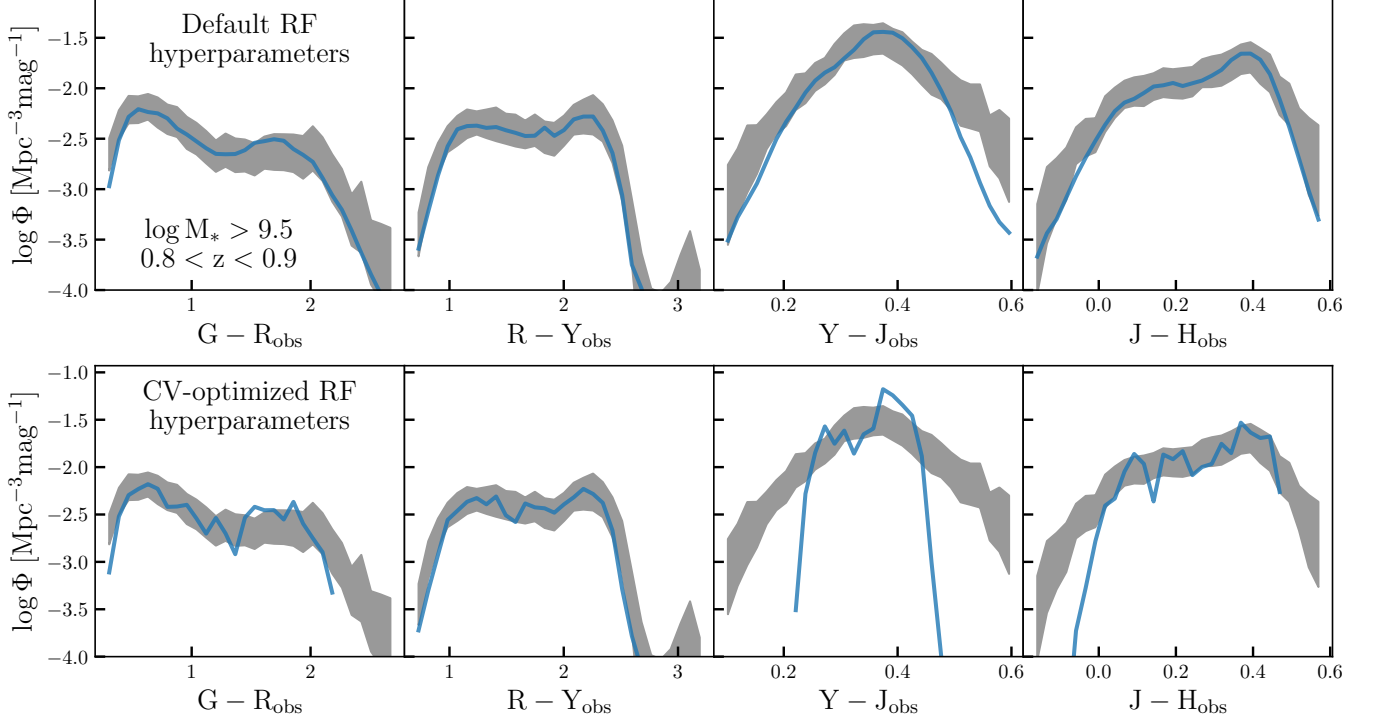
Figure 18. Comparison of the color distributions between UniverseMachine mock galaxies (blue lines) and the UltraVISTA data (grey bands) that they were fit to at the redshift slice $0.8 < z < 0.9$. The fits in the top panel use the default scikit-learn hyperparameters, while those in the bottom panel use cross-validation-optimized hyperparameters. By construction, the optimized hyperparameters predict colors with a lower mean absolute deviation. However, we adopt the default parameters due to their significantly better performance at reproducing the distributions as a whole.