The Astrophysical Journal, 928:10 (18pp), 2022 March 20 https://doi.org/10.3847/1538-4357/ac501b

Constructing the Emission-line Galaxy–Host Halo Connection through Auto and Cross Correlations

Hongyu Gao[1], Y. P. Jing[1,2], Yun Zheng[1], and Kun Xu[1]
[1] Department of Astronomy, School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China; ypjing@sjtu.edu.cn
[2] Tsung-Dao Lee Institute, and Shanghai Key Laboratory for Particle Physics and Cosmology, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China
Received 2021 November 21; revised 2022 January 26; accepted 2022 January 27; published 2022 March 22

Abstract

We investigate the [O II] emission-line galaxy (ELG)-host halo connection via auto and cross correlations, and propose a concise and effective method to populate ELGs in dark matter halos without assuming a parameterized halo occupation distribution (HOD) model. Using the observational data from VIMOS Public Extragalactic Redshift Survey, we measure the auto and cross correlation functions between ELGs selected by [O II] luminosity and normal galaxies selected by stellar mass. Combining the stellar–halo mass relation derived for the normal galaxies and the fraction of ELGs observed in the normal galaxy population, we demonstrate that we can establish an accurate ELG–halo connection. With the ELG–halo connection, we can accurately reproduce the auto and cross correlation functions of ELGs and normal galaxies both in real space and in redshift space, once the satellite fraction is properly reduced. Our method provides a novel strategy to generate ELG mock catalogs for ongoing and upcoming galaxy redshift surveys. We also provide a simple description for the HOD of ELGs.

Unified Astronomy Thesaurus concepts: Emission line galaxies (459); Redshift surveys (1378); Galaxy dark matter halos (1880); Dark energy (351); Observational cosmology (1146)

1. Introduction

Distinguishing dark energy models from modified gravity theories requires us to accurately measure the entire evolutionary history of the universe from the matter dominance to the dark energy dominance. By analyzing the clustering of galaxies, we can measure the expansion history and instantaneous expansion rate of the universe from the baryon acoustic oscillation (e.g., Cole et al. 2005; Eisenstein et al. 2005), and the growth rate from the redshift-space distortion (RSD; e.g., Kaiser 1987). The Sloan Digital Sky Survey (York et al. 2000; Gunn et al. 2006) has performed spectroscopic measurement for a large number of galaxies in the low-redshift universe that has been dominated by dark energy. But for distant galaxies with redshift $z > 0.7$, because their continuum gets faint and most of their optical spectral lines are redshifted to the infrared band, it becomes more and more difficult to conduct a large redshift survey at high redshift.

To overcome this difficulty, galaxies with strong [O II] emission lines (ELGs) have naturally become the main target for next generation redshift surveys (Newman et al. 2013; Dawson et al. 2016; DESI Collaboration et al. 2016; Takada et al. 2014). Since the neutral gas can be photoionized by the ultraviolet (UV) radiation of newly formed massive stars and produce [O II] lines, the main population of [O II] emitters is therefore expected to be star-forming galaxies. Although the violent nuclear activities caused by supermassive black holes (SMBHs) also have enough energy to ionize oxygen atoms, the fraction of active galactic nuclei (AGNs) in the [O II] emitters is small (e.g., Comparat et al. 2013). Compared to other spectral features (such as Hα, [O III] and 4000 Å break), the main advantage of the [O II] line is the doublet feature at the wavelengths λ3727, 3729 Å, and it can be detected in the optical window up to redshift 1.6. For example, the Deep Extragalactic Evolutionary Probe 2 (DEEP2; Newman et al. 2013) has measured the spectrum for more than 50,000 galaxies at $z \sim 1$, in which the number of [O II] ELGs is dominant. Currently, the Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration et al. 2016) is conducting spectroscopic observations for more than 17 million [O II] ELGs within $0.6 < z < 1.6$ covering 14,000 deg^2, which makes up for the vacancy of the luminous red galaxy (LRG) sample at $z > 1$. With the help of its near-infrared spectrometers and large-aperture, the Subaru Prime Focus Spectrograph (PFS; Takada et al. 2014) will conduct spectroscopic observations of [O II] ELGs all the way to $z = 2.4$. The combination of these two redshift surveys will increase the number and coverage of observed [O II] ELGs to an unprecedented level.

Before we can extract the cosmological information from the clustering signal of [O II] ELGs, we need to understand the connection between these galaxies and the underlying dark matter. This is also a prerequisite for generating realistic [O II] ELG mock catalogs (e.g., Osato et al. 2022) for these redshift surveys. Halo occupation distribution (HOD) has become one of the most common ways to construct the halo-galaxy connection (e.g., Jing et al. 1998; Ma & Fry 2000; Peacock & Smith 2000; Seljak 2000; Berlind & Weinberg 2002; Zheng et al. 2005, 2007; Zehavi et al. 2011; Zu & Mandelbaum 2015, 2016, 2018; Guo et al. 2016; Rodríguez-Torres et al. 2016; Xu et al. 2016, 2018; Yuan et al. 2018; Wang et al. 2019). Under the HOD framework, the mean occupation number $\langle N(M) \rangle$ of a given galaxy population in a halo is determined by the halo mass. In addition to HOD, the statistics related to the physical properties (such as stellar mass and luminosity) of the galaxies inhabiting a halo of given mass can be described by the conditional luminosity (stellar mass) function (Yang et al. 2003; Cooray 2006; van den Bosch et al. 2007; Yang et al. 2009, 2012; Rodríguez-Puebla et al. 2015;

Guo et al. 2018). Since subhalos can be more accurately resolved as cosmological simulations improve, the (sub)halo abundance matching (AM) method (e.g., Wechsler et al. 1998; Vale & Ostriker 2006; Wang et al. 2006; Behroozi et al. 2010; Guo et al. 2010; Wang & Jing 2010; Simha et al. 2012; Moster et al. 2013; Guo & White 2014; Chaves-Montero et al. 2016; Guo et al. 2016; Wechsler & Tinker 2018; Behroozi et al. 2019; Xu & Jing 2022; Xu et al. 2022) has become a more effective way to link the observable physical quantity (e.g., stellar mass, luminosity) of a galaxy to its host (sub)halo properties (e.g., halo mass, maximum circular velocity). For normal galaxies in a stellar mass-selected galaxy sample, their galaxy–halo connection is relatively easy to understand because the monotonically increasing stellar–halo mass relation (SHMR) indicates that the massive galaxies tend to occupy massive halos, although it might be affected by other properties beyond stellar mass due to galaxy assembly bias (e.g., Cooper et al. 2010; Wang et al. 2013; Zentner et al. 2014; Hearin et al. 2015; Mandelbaum et al. 2016; Hadzhiyska et al. 2020, 2021a; Xu & Zheng 2020; Cui et al. 2021; Zu et al. 2021; Xu & Jing 2022; Zu et al. 2022). However, the situation may become more complicated for ELGs. Since the quench fraction of massive galaxies is relatively higher, galaxies with strong emission lines are not necessarily hosted by massive halos. On the contrary, low-mass galaxies are more likely to have strong star formation processes. Therefore, the probability that a halo hosts an ELG is expected to peak at low mass, and then gradually decreases toward the high-mass end (Geach et al. 2012; Contreras et al. 2013).

Recently, a handful studies have been devoted to studying the [O II] ELG–halo connection in observations (Favole et al. 2016, 2017; Guo et al. 2019; Avila et al. 2020; Okumura et al. 2021). For instance, by simultaneously constraining the SHMR, completeness and quench fraction of the [O II] ELG sample from the extended Baryon Oscillation Spectroscopic Survey (eBOSS; Dawson et al. 2016), Guo et al. (2019) found that the typical host halo mass of eBOSS ELGs is $\sim 10^{12}\,M_{\odot}$ and the satellite fraction is 13%–17%, although the results slightly depend on the assumptions of their quenched fraction model. They also showed that the completeness of eBOSS ELGs is less than 10%, which indicates that only galaxies with the strongest [O II] emissions are selected by eBOSS. Using the [O II] ELG sample identified by the narrow-band (NB) filters at $z > 1$ in the Subaru Hyper-Suprime Cam (HSC) survey, Okumura et al. (2021) found that their angular correlation function can be well fitted by the HOD model proposed by Geach et al. (2012), but the constraints of the model parameters are poor due to the limited data and the large parameter space. From the number densities of the HSC NB ELGs ($\sim 10^{-3}\,\mathrm{Mpc}^{-3}\,h^3$) and eBOSS ELGs ($\sim 10^{-4}\,\mathrm{Mpc}^{-3}\,h^3$), we can easily see that these ELGs are different populations of [O II] emitters. It is important to study how the ELG–halo connection depends on the [O II] luminosity.

Different from previous works, we aim to investigate the ELG–halo connection for different [O II] luminosities by utilizing the cross correlations between ELGs and normal galaxies. Since the host halos of ELGs are widely distributed in mass as we will show, the autocorrelation of ELGs actually mixes the clustering signal of halos with different mass and is therefore difficult to interpret. But for the normal galaxies selected by stellar mass, we already have a relatively clear understanding of their clustering and host halos properties. Furthermore, the number density of normal galaxies is higher, which makes the cross correlation a better measured quantity. Therefore, the cross correlation of an ELG sample with a stellar mass-selected galaxy sample will tell us how ELGs are distributed around halos that are derived from the SHMR of normal galaxies. In this work, we use the galaxy catalog from the VIMOS Public Extragalactic Redshift Survey (VIPERS,[3] Guzzo et al. 2014; Garilli et al. 2014; Scodeggio et al. 2018). Unlike eBOSS. which only selects ELGs with strong emission lines, VIPERS is an *i*-band limit survey and thus contains ELGs with more moderate [O II] luminosity that are also the main targets of DESI and PFS. Different from the HOD modelings mentioned before, we make full use of the ELG-stellar mass relation in observation without establishing a parameterized model. We will demonstrate that by randomly assigning ELGs to dark matter halos according to the SHMR of normal galaxies, we can well repeat the auto and cross correlation functions in both real space and redshift space as long as the satellite fraction is reduced. The method is simple but effective, which could be a starting point for constructing the ELG–halo connection for surveys such as DESI and PFS, and become a test bed for further improving the connection.

The layout of this paper is organized as follows. In Section 2, we describe the observational data and numerical simulation used in this work. In Section 3, we introduce our methods to account for the survey selection effects and to measure correlation functions. The SHMR is derived by the AM method in Section 4. The main results of the ELG–halo connection are presented in Section 5. Finally, we give a brief conclusion in Section 6. Unless otherwise stated, the cosmological parameters used in this paper are $\Omega_{\mathrm{m},0} = 0.268$, $\Omega_{\Lambda,0} = 0.732$, and $H_0 = 100h\ \mathrm{km\ s^{-1}\ Mpc^{-1}} = 71\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$.

2. Galaxy Sample and Simulation Data

We describe the basics of VIPERS and the properties of our [O II] luminosity-selected and stellar mass-selected subsamples. The *N*-body cosmological simulation used in this study is also introduced in this section.

2.1. VIPERS Sample

We use the galaxy catalog of the final public release (PDR-2) (Scodeggio et al. 2018) of VIPERS. This survey overlaps two sky fields W1 and W4 of the Canada–France–Hawaii Telescope Legacy Survey Wide (CFHTLS-Wide[4]), covering about 24 deg^2. The multi-band magnitudes (u, g, r, i, z) of the parent photometric catalog come from the CFHTLS T0005.[5] Ancillary photometric data is supplemented by the VIPERS Multi-Lambda Survey (Moutard et al. 2016), which matched the CFHTLS T0005 catalog with GALEX (Martin et al. 2005) and the VISTA Deep Extragalactic Observations (Jarvis et al. 2013), and provided extra photometry in the near-UV, far-UV, and K_{s} (K_{video}) bands. Galaxies with $i_{\mathrm{AB}} < 22.5$ in the parent catalog satisfying the following color–color criteria:

$$(r - i) > 0.5 \times (u - g) \quad \mathrm{OR} \quad (r - i) > 0.7 \tag{1}$$

[3] http://vipers.inaf.it
[4] http://www.cfht.hawaii.edu/Science/CFHLS/
[5] http://www.cfht.hawaii.edu/Science/CFHLS/T0005/

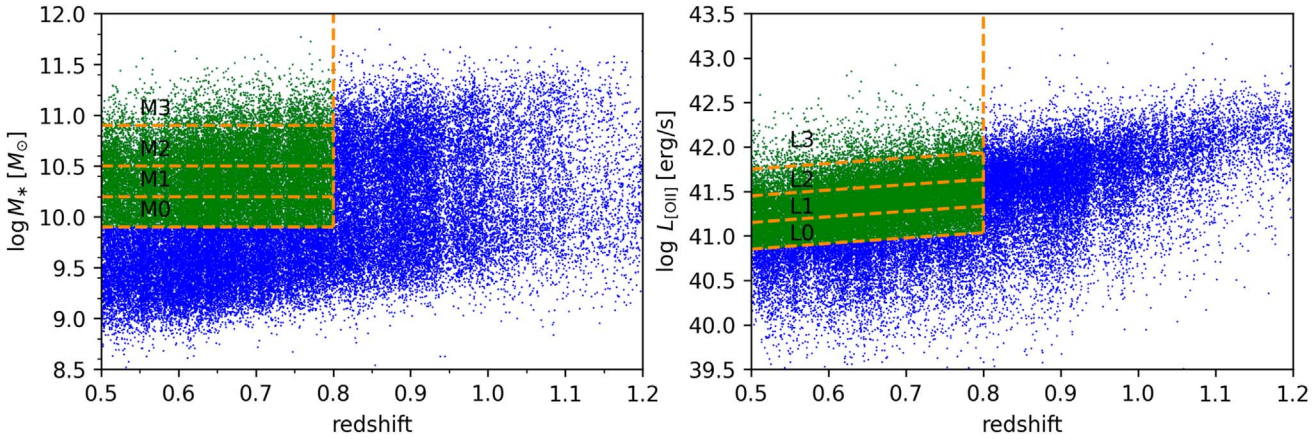

Figure 1. Stellar mass and [O II] luminosity distributions of galaxies in the VIPERS sample. The blue points denote the full galaxy sample while the green points represent the subsamples, whose boundaries are plotted as orange dashed lines.

are selected as the spectroscopic targets. The spectra of about 90,000 galaxies were measured with the VIMOS multi-object spectrograph on the ESO Very Large Telescopes (Le Fèvre et al. 2003). Finally, we only include the VIPERS main galaxy targets (classFlag = 1) with high-quality redshift measurements (zflag $\geqslant 2$) in our research.

2.2. Galaxy Subsamples

For the purpose of analyzing the cross correlations of ELGs and normal galaxies, we divide the galaxy sample in the redshift range $0.5 < z \leqslant 0.8$ into four [O II] luminosity $L_{[\mathrm{O\,II}]}$-selected ($L0$, $L1$, $L2$, and $L3$) and four stellar mass M_*-selected ($M0$, $M1$, $M2$, and $M3$) subsamples. After subtracting the continuum, the [O II] fluxes are measured by fitting a single Gaussian model to the spectrum. The Levenberg–Marquardt algorithm (Levenberg 1944; Marquardt 1963) is adopted to derive the best-fitting [O II] fluxes and their uncertainties. We take the multi-band photometry from the VIPERS Multi-Lambda Survey (Moutard et al. 2016) to model the spectral energy distribution (SED) of galaxies. The LE PHARE (Arnouts et al. 2002; Ilbert et al. 2006) code is used to perform the SED fitting and estimate the physical properties (including stellar mass) of galaxies. More details about the $L_{[\mathrm{O\,II}]}$ measurements and SED template settings can be found in Gao & Jing (2021).[6]

At $0.5 < z \leqslant 0.8$, there are a total of 45,600 galaxies, of which 36,741 have $L_{[\mathrm{O\,II}]} > 0$. The mass and [O II] luminosity distributions of the full sample are displayed as blue points in Figure 1, in which the four M_*-selected and $L_{[\mathrm{O\,II}]}$-selected subsamples are also shown as green points. We present more details of each subsample in Tables 1 and 2. Considering that a galaxy may be contained in both a M_*-selected and a $L_{[\mathrm{O\,II}]}$-selected subsample, we present the fraction of $L_{[\mathrm{O\,II}]}$-selected galaxies included in each M_*-selected subsample in Table 3. This fraction represents the degree of independence of the two subsamples.

Since the stellar mass function (SMF) evolves relatively weakly at $z < 1$ (Pozzetti et al. 2007, 2010; Davidzon et al. 2013), we apply flat stellar mass cuts to construct M_*-selected subsamples. In order to determine the stellar mass completeness limit of the galaxy sample, we follow the same technique proposed by Pozzetti et al. (2010) (see also Davidzon et al. 2013). The 90% stellar mass completeness limit at $z \sim 0.6$ is $10^{9.8}\,M_\odot$, so we take a lower boundary $M_* = 10^{9.9}\,M_\odot$ for the first subsample $M0$.

As for the $L_{[\mathrm{O\,II}]}$-selected subsamples, we adopt redshift evolution cuts to account for the evolution of [O II] luminosity function. Referring to the parameterized evolution model of the characteristic luminosity $L_{[\mathrm{O\,ii}],\star}(z) = L_{[\mathrm{O\,ii}],\star}(0)(1+z)^{\beta_L}$ provided by Comparat et al. (2016), we define the $L_{[\mathrm{O\,II}]}$ cut at z as

$$\log L^{\mathrm{cut}}_{[\mathrm{O\,II}]}(z) = \log L^{\mathrm{cut}}_{[\mathrm{O\,II}]}(z=0.5) + \log\left(\frac{1+z}{1+0.5}\right)^{\beta_L} \quad (2)$$

with $\beta_L = 2.33$. To ensure the completeness of [O II] detection, we set the lower boundary of the first subsample $L0$ as

Table 1
Details of Four Stellar Mass-selected Subsamples

Name	Redshift Range	$\log M_*(M_\odot)$	N_g
$M0$	$0.5 < z \leqslant 0.8$	[9.9, 10.2]	7829
$M1$	$0.5 < z \leqslant 0.8$	[10.2, 10.5]	8355
$M2$	$0.5 < z \leqslant 0.8$	[10.5, 10.9]	8040
$M3$	$0.5 < z \leqslant 0.8$	[10.9, ∞]	1964

Table 2
Details of Four [O II] Luminosity-selected Subsamples

Name	Redshift Range	$\log L_{[\mathrm{O\,II}]}(z=0.5)$ (erg s^{-1})	N_g
$L0$	$0.5 < z \leqslant 0.8$	[40.85, 41.15]	9349
$L1$	$0.5 < z \leqslant 0.8$	[41.15, 41.45]	11721
$L2$	$0.5 < z \leqslant 0.8$	[41.45, 41.75]	6281
$L3$	$0.5 < z \leqslant 0.8$	[41.75, ∞]	1693

Table 3
Fractions of $L_{[\mathrm{O\,II}]}$-selected Galaxies Included in each M_*-selected Subsample

Name	$L0$ Fraction	$L1$ Fraction	$L2$ Fraction	$L3$ Fraction
$M0$	0.210	0.243	0.120	0.039
$M1$	0.181	0.134	0.065	0.020
$M2$	0.172	0.108	0.043	0.016
$M3$	0.149	0.108	0.045	0.024

[6] The cosmological parameters used in the [O II] luminosity computation and SED fitting process are $\Omega_{\Lambda,0} = 0.7$, $\Omega_{m,0} = 0.3$, and $H_0 = 70$ km s^{-1} Mpc^{-1}, which are slightly different from what we adopted in this study. But this does not affect our subsequent analysis because we mainly care about the relative difference between different subsamples rather than their absolute value.

$L_{[\mathrm{O\,II}]}(z=0.5)=10^{40.85}\,\mathrm{erg\,s^{-1}}$, which is above the 5σ detection limit ($F^{\mathrm{lim}}_{[\mathrm{O\,II}]}=3.5\times10^{-17}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$) of VIMOS with an exposure time of 0.75 hr (Comparat et al. 2016).

2.3. Simulation

One of the high resolution N-body cosmological simulations of CosmicGrowth (Jing 2019) is used to study the galaxy–halo connection. This simulation is performed by the P^3M method (Jing & Suto 2002) and has 3072^3 dark matter particles in a $600\,\mathrm{Mpc}\,h^{-1}$ box with the standard Λcold dark matter cosmological parameters: $\Omega_{\mathrm{m}}=0.268$, $\Omega_\Lambda=0.732$, $h=0.71$, $n_{\mathrm{s}}=0.968$, and $\sigma_8=0.83$. The halos and subhalos are identified by the friends-of-friends algorithm (FOF) (Davis et al. 1985) and the Hierarchical-Bound-Tracing algorithm (HBT+) (Han et al. 2012, 2018), respectively.

Although the [O II] ELGs are considered to be more likely to exist in low-mass halos ($\sim10^{12}\,M_\odot$) (e.g., Favole et al. 2016; Guo et al. 2019; Hadzhiyska et al. 2021b; Okumura et al. 2021), the mass resolution $m_{\mathrm{p}}=5.54\times10^8\,M_\odot\,h^{-1}$ of particles in our simulation is sufficient to resolve them. The halo mass M_{h} is defined as its viral mass M_{vir} that is the mass enclosed by a sphere with an average density of $\Delta_{\mathrm{vir}}(z)$ times the critical density of the universe (Gunn et al. 1972; Bryan & Norman 1998). The subhalo accretion mass M_{s} is defined as its virial mass at the last snapshot before merging into the current host halo. In addition, we have carefully dealt with the small subhalos that have been almost or even completely stripped by the tidal force. Using the fitting formula proposed by Jiang et al. (2008), we trace the merger history of those subhalos with less than 20 particles and calculate their merger timescale to judge whether they can remain distinct as subhalos. Finally, the snapshot with $z=0.663$, which is close to the mean redshift of our galaxy sample, is chosen for our analysis. To make a fair comparison with the observations, we have incorporated the RSD effects to these simulated halos (subhalos). We choose the z-axis as the line of sight and define the redshift of the center of the simulation box as 0.663. For a halo (subhalo), the cosmological redshift z_{c} is given according to its comoving distance to the center of the box, while the redshift z_{p} caused by peculiar motion is calculated by v_z/c, where v_z is the velocity in the z-direction and c is the speed of light. Besides, we add to v_z a velocity randomly derived from a Gaussian distribution of the dispersion $\sigma_v=c\sigma_z$, where $\sigma_z=0.00054$ is the typical redshift uncertainty for VIPERS (Scodeggio et al. 2018). Finally, the updated z-axis coordinate of a halo (subhalo) is converted from its observed redshift $z_{\mathrm{obs}}=(1+z_{\mathrm{c}})(1+z_{\mathrm{p}})-1$.

3. Measurement of Galaxy Clustering

In this section, we carefully correct the selection effects in VIPERS and measure the cross (auto) correlation functions for different galaxy subsamples both in redshift space and real space.

3.1. Selection Functions

In order to achieve an accurate measurement of galaxy clustering, we should understand and correct for the selection functions listed below.

1. The survey masks in VIPERS. We can account for these survey masks by applying the same sky geometry to the random sample.
2. Target sampling rate (TSR). Some galaxies in the parent photometric catalog cannot be spectroscopically observed due to the limited number of slits. This effect can bias the targeting of galaxies in the dense region due to the uniform distribution of the slits, and the clustering of galaxies is underestimated. This effect can be corrected by up-weighting $w^{\mathrm{TSR}}=1/\mathrm{TSR}$ (de la Torre et al. 2013; Scodeggio et al. 2018) for each galaxy.
3. Spectroscopic success rate (SSR). It quantifies the probability that the redshift of a galaxy targeted by VIMOS can be successfully measured (i.e., zflag $\geqslant 2$). By exploring the dependence of SSR on the multidimensional parameter space, Scodeggio et al. (2018) evaluates the SSR for each galaxy based on a nearest-neighbor algorithm. Therefore, we also up-weight each galaxy with $w^{\mathrm{SSR}}=1/\mathrm{SSR}$.
4. Slit collisions. Similar to fiber collision, if the distance between the two galaxies is less than the physical size of the silt, only the spectrum of one galaxy can be observed. Additionally, in order to avoid the overlap of spectra in the VIMOS detectors, the spectra of two galaxies with distance below a specific size along the direction perpendicular to the silt cannot be observed at the same time. The combination of these two effects will suppress the clustering of galaxies at small scales. Following the method of Pezzotta et al. (2017), we calculate the angular weights $w^{\mathrm{A}}(\theta)$ using 153 VIPERS mock samples (see Appendix A) to correct the slit collision effect for galaxy pairs.
5. i-band magnitude limit and color sampling rate (CSR). VIPERS adopts the $i^{\mathrm{AB}}<22.5$ flux cut and a color–color cut (Equation (1)) to construct a flux-limited sample at $z>0.5$, which introduces two radial selection functions to the redshift distribution of galaxies. Guzzo et al. (2014) provides an accurate model $\mathrm{CSR}(z)=1/2-1/2\,\mathrm{erf}\,[b(z_{\mathrm{t}}-z)]$ with $b=10.8$ and $z_{\mathrm{t}}=0.444$ for the radial weight $w^{\mathrm{CSR}}=1/\mathrm{CSR}$ to describe the completeness of the color–color selection. To account for the combination of these two radial selection effects, we use the V_{max} method (Cole 2011; de la Torre et al. 2013; Pezzotta et al. 2017; de la Torre et al. 2017; Yang et al. 2020) to generate a smooth redshift distribution for the random sample. We present the detail of the V_{max} method in Appendix B. In Figure 2, we can see that the redshift distribution of the random sample thus generated is well consistent with the observed one.

After considering these selection effects, we can estimate the completeness-corrected number density of our galaxy subsamples through

$$n_{\mathrm{g}}=\sum_i^{N_{\mathrm{g}}}\frac{w_i^{\mathrm{TSR}}w_i^{\mathrm{SSR}}w_i^{\mathrm{CSR}}}{V_{\mathrm{max},i}}, \tag{3}$$

where $V_{\mathrm{max},i}$ is computed with

$$V_{\mathrm{max},i}=\frac{A_{\mathrm{eff}}}{3\times(180/\pi)^2}\times[D^3_{\mathrm{com}}(z_{\mathrm{max},i})-D^3_{\mathrm{com}}(0.5)], \tag{4}$$

where $A_{\mathrm{eff}}=16.324\,\mathrm{deg}^2$ is the effective sky area of VIPERS and D_{com} is the comoving distance.

3.2. Estimation of the Correlation Function

To measure the galaxy clustering in redshift space, we decompose the separation vector $\boldsymbol{s}=\boldsymbol{s_2}-\boldsymbol{s_2}$ of two galaxies into two components r_π and r_{p}. r_π can be obtained by projecting $\boldsymbol{s}$ along the line of sight, $r_\pi=(\boldsymbol{s}\cdot\boldsymbol{l})/|\boldsymbol{l}|$ with $\boldsymbol{l}=(\boldsymbol{s_2}+\boldsymbol{s_2})/2$, and

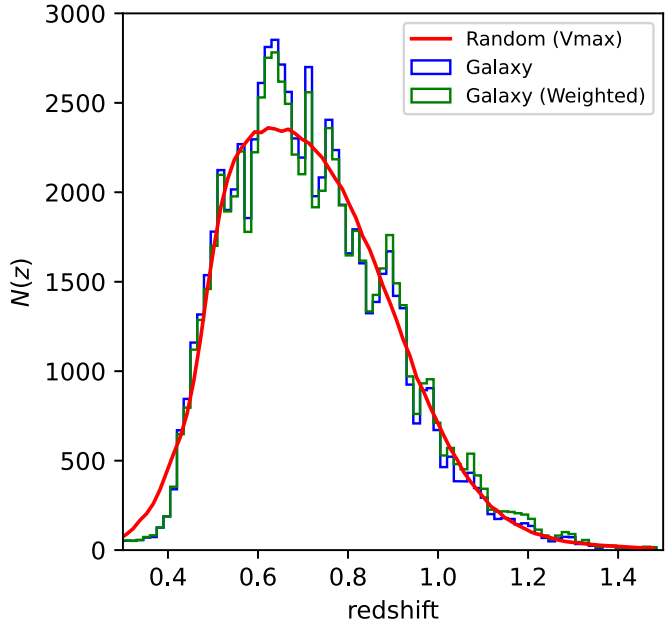

Figure 2. Redshift distributions of the galaxy and random samples. The blue histogram shows the full galaxy sample (W1 and W4 fields) of VIPERS. The distribution of the galaxy sample weighted by $w^{\rm SSR} \times w^{\rm TSR}$ is displayed as the green histogram, which shows a slight difference from the un-weighted one. The red solid line represents the radial distribution of the random sample generated with the $V_{\rm max}$ technique as described in the text. The numbers of weighted galaxies and random points have been properly normalized.

$r_{\rm p}$ is calculated as $\sqrt{s^2 - r_\pi^2}$. We choose 12 $r_{\rm p}$ bins from 0.12–30 Mpc h^{-1} with an equal logarithmic interval and 40 r_π bins from 0–40 Mpc h^{-1} with an equal linear interval. The redshift-space cross (auto) correlation functions for different galaxy subsamples are measured utilizing the Landy–Szalay estimator (Landy & Szalay 1993; Szapudi & Szalay 1998)

$$\xi_{xy}(r_{\rm p}, r_\pi) = \left[\frac{D_x D_y - D_x R_y - D_y R_x + R_x R_y}{R_x R_y}\right], \quad (5)$$

where x, y indicate different samples ($x = y$ for the autocorrelation). The normalized pair counts for galaxy-galaxy, galaxy-random, and random-random are calculated by

$$\begin{aligned} D_x D_y(r_{\rm p}, r_\pi) &= \frac{\sum_{i=1}^{N_{{\rm g},x}}\sum_{j=1}^{N_{{\rm g},y}} w^{\rm A}(\theta_{ij}) w_i^{\rm c} w_j^{\rm c} \Theta_{ij}(r_{\rm p}, r_\pi)}{\sum_{i=1}^{N_{{\rm g},x}}\sum_{j=1}^{N_{{\rm g},y}} w^{\rm A}(\theta_{ij}) w_i^{\rm c} w_j^{\rm c}} \\ D_x R_y(r_{\rm p}, r_\pi) &= \frac{\sum_{i=1}^{N_{{\rm g},x}}\sum_{j=1}^{N_{{\rm r},y}} w_i^{\rm c} \Theta_{ij}(r_{\rm p}, r_\pi)}{N_{{\rm r},y}\sum_{i=1}^{N_{{\rm g},x}} w_i^{\rm c}} \\ D_y R_x(r_{\rm p}, r_\pi) &= \frac{\sum_{i=1}^{N_{{\rm g},y}}\sum_{j=1}^{N_{{\rm r},x}} w_i^{\rm c} \Theta_{ij}(r_{\rm p}, r_\pi)}{N_{{\rm r},x}\sum_{i=1}^{N_{{\rm g},y}} w_i^{\rm c}} \\ R_x R_y(r_{\rm p}, r_\pi) &= \frac{\sum_{i=1}^{N_{{\rm r},x}}\sum_{j=1}^{N_{{\rm r},y}} \Theta_{ij}(r_{\rm p}, r_\pi)}{N_{{\rm r},x} N_{{\rm r},y}}, \end{aligned} \quad (6)$$

where Θ_{ij} is equal to 1 only when a galaxy pair falls into this $(r_{\rm p}, r_\pi)$ bin, and the pair counts have been up-weighted by $w^{\rm c} = w^{\rm SSR} \times w^{\rm TSR}$ and $w^{\rm A}(\theta)$ as mentioned in Section 3.1.

The $\xi_{xy}(r_{\rm p}, r_\pi)$ is integrated along the line of sight to give the real-space projected correlation function (Davis & Peebles 1983)

$$w_{\rm p}^{xy}(r_{\rm p}) = 2\int_0^{r_{\pi,\rm max}} \xi_{xy}(r_{\rm p}, r_\pi) dr_\pi, \quad (7)$$

with $r_{\pi,\rm max} = 40$ Mpc h^{-1}. We also employ the same upper limit of the integration when modeling the $w_{\rm p}^{xy}(r_{\rm p})$ in the simulation to make a fair comparison.

The covariance matrix of the measured $w_{\rm p}^{xy}(r_{\rm p})$ is estimated with the jackknife technique. We divide the entire survey into 24 fields (16 for W1 and 8 for W4) with an area of approximately 1 deg^2 for each field, and the covariance matrix of the measured $w_{\rm p}^{xy}(r_{\rm p})$ can be estimated with

$$C(i,j) = \frac{N_{\rm jack} - 1}{N_{\rm jack}} \sum_{k=1}^{N_{\rm jack}} (w_{{\rm p},i}^k - \bar{w}_{{\rm p},i})(w_{{\rm p},j}^k - \bar{w}_{{\rm p},j}), \quad (8)$$

where $N_{\rm jack} = 24$ is the number of jackknife samples and i (j) denotes the i (j)th $r_{\rm p}$ bin.

In analogy to the way of $w_{\rm p}(r_{\rm p})$ in real space, we also measure the multiple moments of the correlation functions in redshift space. The monopole $\xi_0(s)$, quadrupole $\xi_2(s)$, and hexadecapole $\xi_4(s)$ (Hamilton 1992) are defined as

$$\xi_l(s) = \frac{2l+1}{2}\int_{-1}^{1} \xi(s, \mu) L_l(\mu), \quad (9)$$

where $L_l(\mu)$ is the Legendre function, s is binned from 0.12–30 Mpc h^{-1} with an equal logarithmic interval and μ is binned with a linear width $\Delta\mu = 0.1$.

4. Determining the Stellar–Halo Mass Relation

Before modeling the cross correlations of ELGs with normal galaxies, we first establish the connection between the stellar mass of normal galaxies and their halo mass. To derive the SHMR by the AM method, we use the model proposed by Wang & Jing (2010) and adopt an efficient way (Guo et al. 2016; Zheng & Guo 2016) to calculate the modeled correlation functions.

4.1. Abundance Matching Model

The conditional probability distribution function (PDF) that a galaxy with stellar mass M_* is hosted by a (sub)halo with mass M is assumed to obey a Gaussian distribution

$$p(M_*|M) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\log M_* - \log\langle M_*|M\rangle)^2}{2\sigma^2}\right]. \quad (10)$$

We adopt the parameterized mean relation $\langle M_*|M\rangle$ proposed by Wang & Jing (2010) (see also Wang et al. 2006)

$$\langle M_*|M\rangle = \frac{2k}{\left(\frac{M}{M_0}\right)^{-\alpha} + \left(\frac{M}{M_0}\right)^{-\beta}}, \quad (11)$$

where α and β quantify the slopes of two power-law forms separated at M_0, and k is a normalization constant. In principle, the $p(M_*|M)$ for central and satellite galaxies should be modeled separately to account for possible different formation histories. For the satellites, the current stellar mass depends not only on the accretion mass but also on the subsequent evolution after infalling (Yang et al. 2012). However, the difference of the M_*–M relationship between halo and subhalo is small (Wang & Jing 2010), and the difference should not be important given the current sample size of VIPERS (see

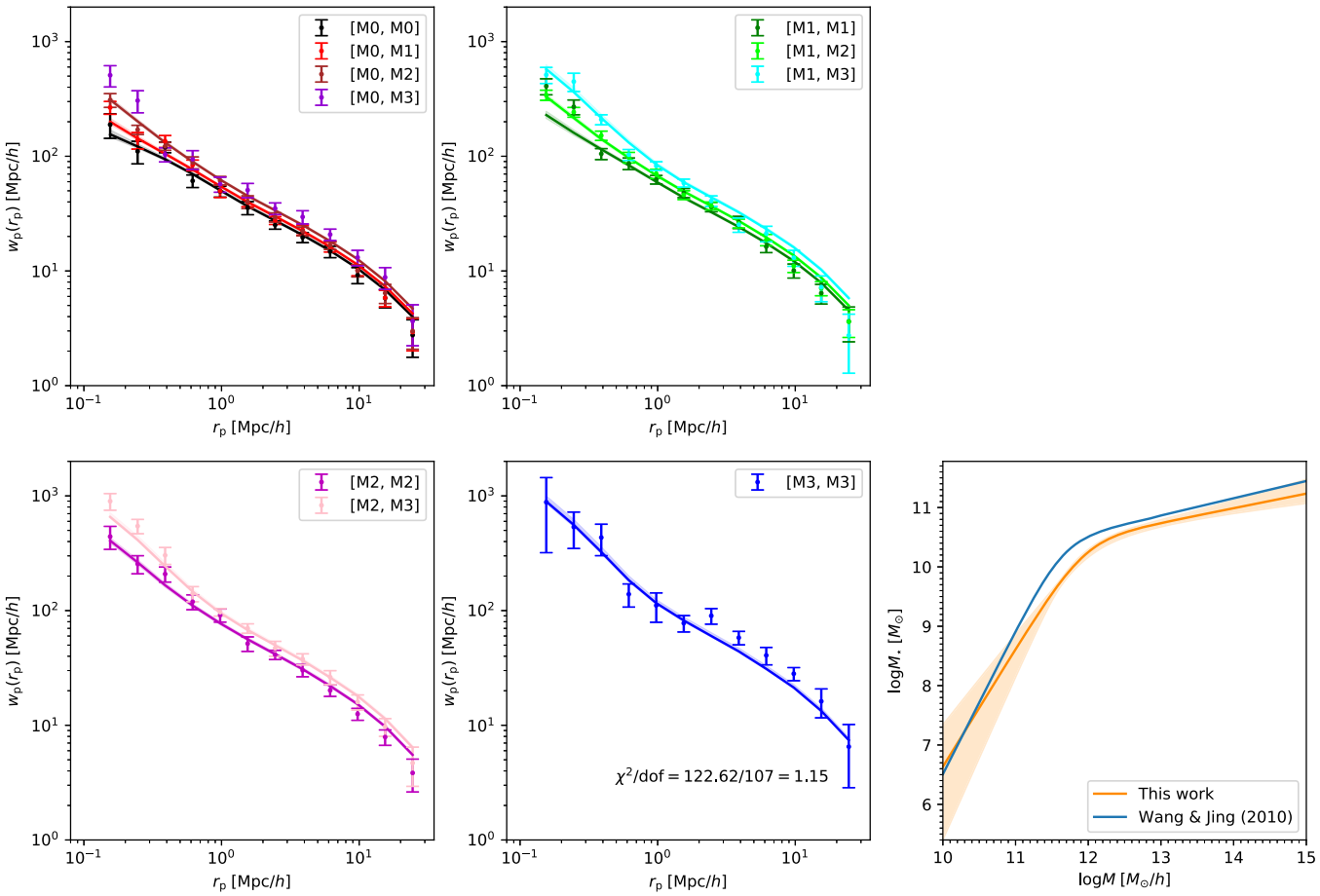

Figure 3. Projected cross (auto) correlation functions for the M_*-selected subsamples, and the SHMR model derived by the AM approach. In the left four panels, the data points with error bars denote the observational measurements. The $\boldsymbol{w}_p$ between different subsamples is marked with different colors. Except for $\boldsymbol{w}_p^{M0M3}$ of [$M0$, $M3$], all observation data are used in the fitting process (see Section 4.3 for details). The best-fitting $\boldsymbol{w}_p$ model as well as its 1σ scatter is plotted as a solid line with a shadowed region. The reduced χ^2 is also denoted in the fourth panel. We present the best-fitting SHMR model in the rightmost panel. The SHMR model derived by Wang & Jing (2010) is also shown as the blue solid line.

below). Therefore, we adopt a unified M_*–M relationship for halos and for subhalos with the same set of parameters: α, β, M_0, k, and σ.

4.2. Tabulated Correlation Functions

When exploring the parameter space, we usually need to populate halos (subhalos) with galaxies based on updated AM model parameters and calculate the correlation functions of modeled galaxies by many times. It will consume a significant amount of CPU time if the correlation functions are not calculated efficiently. Therefore, we extend the tabulated method (Zheng & Guo 2016; Guo et al. 2016) to calculate the cross correlation function in the simulation. The key to this method is to prepare a table for the correlation functions of different halos (subhalos) binned by mass or other physical properties. Different weights are assigned to the tabulated correlation functions according to the AM model, and the combination yields the correlation function of the modeled galaxies. In this way, the halos (subhalos) in our simulation are divided into 500 tiny mass bins with a width of $\Delta \log M = 0.01$ ranging from 10^{10}–$10^{15}\,M_\odot\,h^{-1}$. The correlation functions of halo–halo, halo–subhalo, and subhalo–subhalo for these bins are then measured and organized into three tables each with 500×500 elements. Eventually, the modeled $w_{\rm p,m}^{xy}$ for two galaxy samples x and y is computed by

$$
\begin{aligned}
& w_{\rm p,m}^{xy}(r_{\rm p}) \\
&= \sum_{i,j} \frac{\bar{n}_{{\rm h},i}\bar{n}_{{\rm h},j}}{n_{\rm g,m}^{x} n_{\rm g,m}^{y}} P_x(M_{{\rm h},i}) P_y(M_{{\rm h},j}) w_{\rm p,hh}(r_{\rm p}|M_{{\rm h},i}, M_{{\rm h},j}) \\
&+ \sum_{i,j} \frac{\bar{n}_{{\rm h},i}\bar{n}_{{\rm s},j}}{n_{\rm g,m}^{x} n_{\rm g,m}^{y}} [P_x(M_{{\rm h},i}) P_y(M_{{\rm s},j}) \\
&+ P_y(M_{{\rm h},i}) P_x(M_{{\rm s},j})] w_{\rm p,hs}(r_{\rm p}|M_{{\rm h},i}, M_{{\rm s},j}) \\
&+ \sum_{i,j} \frac{\bar{n}_{{\rm s},i}\bar{n}_{{\rm s},j}}{n_{\rm g,m}^{x} n_{\rm g,m}^{y}} P_x(M_{{\rm s},i}) P_y(M_{{\rm s},j}) w_{\rm p,ss}(r_{\rm p}|M_{{\rm s},i}, M_{{\rm s},j}), \qquad (12)
\end{aligned}
$$

where i, j denote different halo (subhalo) bins. The probabilities that the central and satellite galaxies in the M_*-selected subsample x are hosted by the halo with $M_{{\rm h},i}$ and subhalo with $M_{{\rm s},i}$ are expressed as

$$
\begin{aligned}
P_x(M_{{\rm h},i}) &= P_{{\rm cen},x}(M_{*,x}^{\rm min} < M_* < M_{*,x}^{\rm max}|M_{{\rm h},i}) \\
&= \int_{M_{*,x}^{\rm min}}^{M_{*,x}^{\rm max}} p(M_*|M_{{\rm h},i}) dM_* \qquad (13)
\end{aligned}
$$

and

$$P_x(M_{s,i}) = P_{sat,x}(M_{*,x}^{min} < M_* < M_{*,x}^{max}|M_{s,i}) = \int_{M_{*,x}^{min}}^{M_{*,x}^{max}} p(M_*|M_{s,i})dM_*, \quad (14)$$

where $M_{*,x}^{min}$ ($M_{*,x}^{max}$) is the lower (upper) boundary of the M_*-selected subsample x and $p(M_*|M)$ is the conditional PDF of the stellar mass defined in Equation (10). The modeled number density of the subsample x can be calculated by

$$n_{g,m}^{x} = \sum_i [\bar{n}_{h,i} P_x(M_{h,i}) + \bar{n}_{s,i} P_x(M_{s,i})]. \quad (15)$$

4.3. Fitting Procedure

In the observation, we measure four autocorrelation functions ($\boldsymbol{w}_p^{M0M0}$, $\boldsymbol{w}_p^{M1M1}$, $\boldsymbol{w}_p^{M2M2}$, and $\boldsymbol{w}_p^{M3M3}$), six cross correlation functions ($\boldsymbol{w}_p^{M0M1}$, $\boldsymbol{w}_p^{M0M2}$, $\boldsymbol{w}_p^{M0M3}$, $\boldsymbol{w}_p^{M1M2}$, $\boldsymbol{w}_p^{M1M3}$, and $\boldsymbol{w}_p^{M2M3}$) and four galaxy number densities (n_g^{M0}, n_g^{M1}, n_g^{M2}, and n_g^{M3}) for the stellar mass-selected subsamples. These measured correlation functions are displayed in Figure 3. For the correlation function between subsamples Mi and Mj, we can define its χ^2 as

$$\chi^2_{MiMj} = \sum_{k=1}^{N_{r_p}} \sum_{l=1}^{N_{r_p}} (w_{p,k} - w_{p,m,k}) C_{kl}^{-1} (w_{p,l} - w_{p,m,l}), \quad (16)$$

where $\boldsymbol{w}_p$, $\boldsymbol{w}_{p,m}$, and $\boldsymbol{C}$ denote the observed correlation function $\boldsymbol{w}_p^{MiMj}$, the model prediction $\boldsymbol{w}_{p,m}^{MiMj}$, and the covariance matrix $\boldsymbol{C}^{MiMj}$, respectively. Here, the inverse of covariance matrix $\boldsymbol{C}^{-1}$ is multiplied by a bias-correction factor $(N_{jack} - N_{r_p} - 2)/(N_{jack} - 1)$ (Hartlap et al. 2007), where $N_{jack} = 24$ and $N_{r_p} = 12$ are the number of jackknife subsamples and r_p bins, respectively. Then the total χ^2 is written as

$$\chi^2 = \sum_{i=0}^{3} \sum_{j=i}^{3} \chi^2_{MiMj} + \sum_{i=0}^{3} \frac{(n_g^{Mi} - n_{g,m}^{Mi})^2}{\sigma^2_{Mi}}, \quad (17)$$

where $n_{g,m}^{Mi}$ is the modeled number density of the ith subsample and σ_{Mi} is the field-to-field variation in different jackknife fields. Particularly, since the red satellite galaxies in the $M0$ subsample may be slightly incomplete at $z > 0.7$, the one-halo term of the cross correlation between $M0$ and $M3$, which mainly contains the massive central galaxies, is more likely to be suppressed if the red satellite galaxies in $M0$ are missing. Conservatively, we remove $\boldsymbol{w}_p^{M0M3}$ (corresponding to $i = 0$, $j = 3$ in Equation (17)) in our fitting. In addition, considering the current limited data size, we ignore the covariance between different subsamples and use a total of nine covariance matrices each with 12×12 elements in Equation (16). The degree of freedom is therefore dof $= 12 \times 9 + 4 - 5 = 107$. In Bayesian theory, the posterior distribution is proportional to the likelihood function times the prior of the parameters. We set wide priors for the five parameters: $10 < \log M_0 < 13$, $0.1 < \alpha < 0.5$, $1 < \beta < 5$, $9 < \log k < 12$, and $0 < \sigma < 1$. A Markov Chain Monte Carlo (MCMC) analysis is performed with `emcee` (Foreman-Mackey et al. 2013). The posterior distributions of the model parameters are shown in Figure 4. Overall, all five parameters are well determined.

We present the best-fitting $\boldsymbol{w}_p$ as well as the SHMR as solid lines in Figure 3. The reduced χ^2 is equal to 1.12, indicating a good overall fit. Compared to the SHMR model derived by Wang & Jing (2010) at $z \sim 0.8$ using VVDS observation (Pozzetti et al. 2007), $\log M_0$ is slightly larger and α is slightly smaller in our model ($\log M_0 = 11.64$ and $\alpha = 0.29$ in their unified model). This is partly because there is a degeneracy of $\log M_0$ and α, which is also evident in the $\log M_0$-α contour in Figure 4. Nevertheless, with the larger galaxy sample of VIPERS, we have imposed tighter constraints on the SHMR model at $z \sim 0.6$.

5. Constructing the Halo Occupation of ELGs

In this section, we aim to propose an efficient way to construct the ELG–halo connection. We investigate how to populate the halos with ELGs in the simulation with the measured ELG-stellar mass relation and SHMR. We test our method in both real space and redshift space. We also propose a model for HOD modeling of ELGs.

5.1. ELG-stellar Mass Relation in the Observation

We first measure the fraction of ELGs in the whole population of galaxies as a function of stellar mass in the observation. For each $L_{[O\,II]}$-selected subsample, we divide the galaxies into 20 $\log M_*$ bins ranging from $\log M_* = 8$–$2\,M_\odot$ with a bin width $\Delta \log M_* = 0.2$, and compute the weighted number density $n_g(L_{[O\,II]}, M_*)$ in each bin using Equation (3). Then the fraction of each $L_{[O\,II]}$-selected subsample at a given stellar mass is defined as

$$F_L(M_*) = \frac{n_g(L_{[O\,II]}, M_*)}{n_g(M_*)}. \quad (18)$$

where the number density $n_g(M_*)$ of *all* galaxies for a stellar mass bin can be estimated by integrating the galaxy SMF $\Phi(M_*)$:

$$n_g(M_*) = \int_{\log M_* - \Delta \log M_*/2}^{\log M_* + \Delta \log M_*/2} \Phi(M_*) d \log M_*. \quad (19)$$

Here, we adopt the SMF measured by McLeod et al. (2021) in the redshift range $0.25 < z < 0.75$. McLeod et al. (2021) combined the data from the Hubble Space Telescope CANDELS fields (Grogin et al. 2011; Koekemoer et al. 2011) and other ground-based surveys with deep photometric measurements, and provided the best-fitting parameters of the double Schechter function (Schechter 1976).

In Figure 5, we show the fraction $F_L(M_*)$ for different $L_{[O\,II]}$-selected subsamples. The error bars represent the Poisson errors of the weighted number counts. First, we note that the shapes of $F_L(M_*)$ for the four subsamples are similar, while the locations of the peaks of $F_L(M_*)$ shift slightly from $10^{9.3}$–$10^{9.7}\,M_\odot$ with [O II] luminosity increasing. Furthermore, at the low-mass end, the i-band magnitude limit may have led to a rapid decrease of the number of galaxies, thus causing the rapid drop of $F_L(M_*)$. The gradual decrease of $F_L(M_*)$ at the high mass is expected because more massive galaxies are more likely to stop their star formation and become quiescent. Moreover, the $F_L(M_*)$ of both $L2$ and $L3$ show an upturn at $\log M_* > 11.3\,M_\odot$. This feature might imply that the galaxies at the high-mass end are likely to host AGN, whose violent activities are sufficient

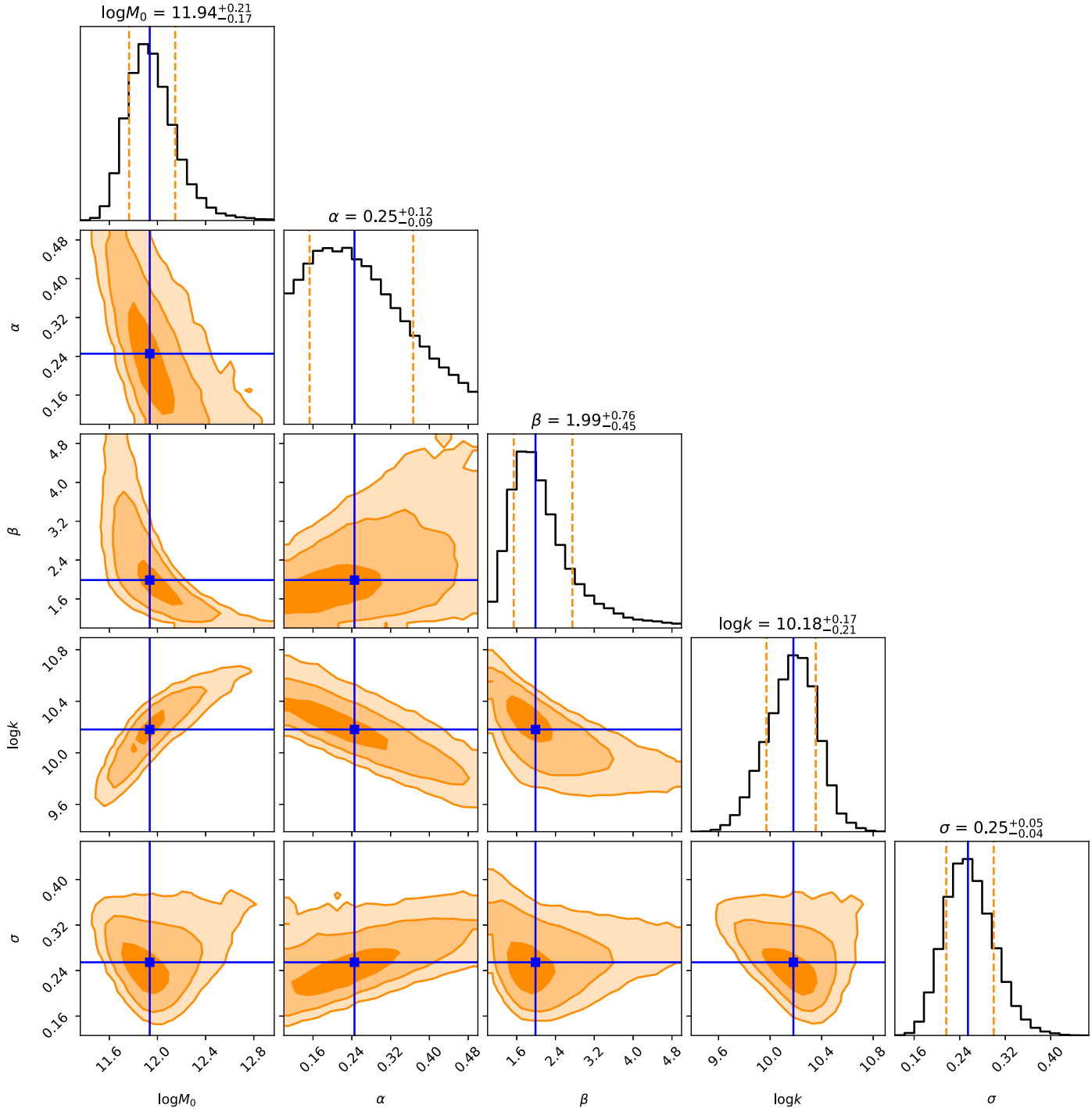

Figure 4. Posterior distributions of the parameters in the unified SHMR model. The 1D PDF of each parameter is plotted as a histogram at the top panel of each column, where the median value and 1σ uncertainty are also labeled. The 2D joint PDF of each parameter pair is shown as a contour with three confidence levels (68%, 95%, and 99%).

enough to generate strong [O II] emissions (e.g., Kocevski et al. 2011). Nevertheless, the AGN contamination cannot significantly affect our results because the number of the massive galaxies with $\log M_* > 11.3\ M_\odot$ is very small (only 12 galaxies in *L*2 and 10 galaxies in *L*3). Instead of using a parameterized model, we linearly interpolate the $F_L(M_*)$- $\log M_*$ relationships to preserve the observed information. Moreover, it is worth mentioning that Guo et al. (2019) points out that the completeness of the ELG sample in eBOSS varies from 1%–10% at different stellar masses, which is comparable to the $F_L(M_*)$ of the *L*2 or *L*3 subsample. This is due to the *g*-band magnitude limit of eBOSS (Raichoor et al. 2017), which causes the majority of the selected ELGs to be luminous ones.

5.2. Populating Halos with ELGs

Combining the ELG-stellar mass relation measured above and the SHMR derived in Section 4, we can connect the ELGs with halos in the simulation. We present the observed autocorrelation functions of the $L_{[\mathrm{O\ II}]}$-selected subsamples and their cross

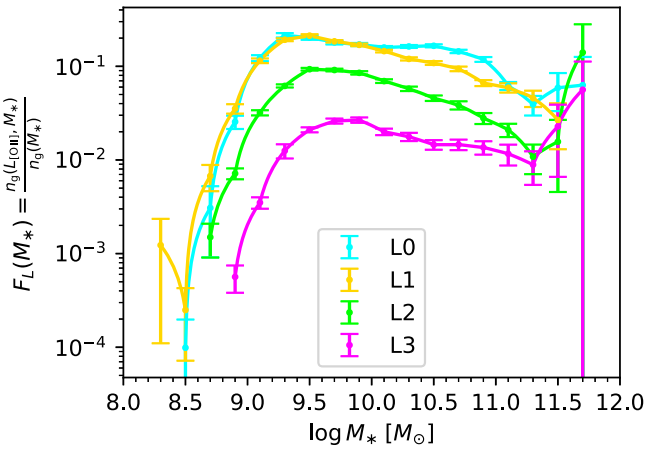

Figure 5. Fractions of the ELGs in the $L_{[\mathrm{O\,II}]}$-selected subsamples, in terms of the whole galaxy population, as a function of stellar mass. The measured $F_L(M_*)$ as well as its Poisson error for different subsamples are denoted as the data points with different colors. The linear interpolation of the observed $F_L(M_*)-\log M_*$ relationships are shown as solid lines with corresponding colors.

correlation functions with the M_*-selected subsamples in Figures 6 and 7. To model these correlation functions, we also adopt Equation (12) to make the calculation more efficient, just changing the $P_{x(y)}(M_{\mathrm{h},i})$ and $P_{x(y)}(M_{\mathrm{s},i})$ in the Equation. For an M_*-selected subsample x, the $P_x(M_{\mathrm{h},i})$ and $P_x(M_{\mathrm{s},i})$ can also be calculated by Equations (13) and (14), respectively. And for an $L_{[\mathrm{O\,II}]}$-selected subsample x, we can define its $P_x(M_{\mathrm{h},i})$ and $P_x(M_{\mathrm{s},i})$ as

$$P_x(M_{\mathrm{h},i}) = P_{\mathrm{cen},x}(L^{\mathrm{min}}_{[\mathrm{O\,II}],x} < L_{[\mathrm{O\,II}]} < L^{\mathrm{max}}_{[\mathrm{O\,II}],x}|M_{\mathrm{h},i}) = \int_{-\infty}^{+\infty} F_{Lx}(M_*)p(M_*|M_{\mathrm{h},i})dM_* \quad (20)$$

and

$$P_x(M_{\mathrm{s},i}) = P_{\mathrm{sat},x}(L^{\mathrm{min}}_{[\mathrm{O\,II}],x} < L_{[\mathrm{O\,II}]} < L^{\mathrm{max}}_{[\mathrm{O\,II}],x}|M_{\mathrm{s},i}) = \int_{-\infty}^{+\infty} F_{Lx}(M_*)p(M_*|M_{\mathrm{s},i})dM_*, \quad (21)$$

where $F_{Lx}(M_*)$ is the fraction of the subsample x at given M_* and $p(M_*|M_{\mathrm{h},i})$ ($p(M_*|M_{\mathrm{s},i})$) is fixed to the best-fitting SHMR derived in Section 4. In the above equations, the sample of ELGs is equivalent to a random selection of the fraction $F_{Lx}(M_*)$ of galaxies from the whole population. In this manner, we calculate the modeled projected cross (auto) correlation functions w_{p} for each $L_{[\mathrm{O\,II}]}$-selected subsample and present them as solid lines in Figure 6, where the four panels represent the four $L_{[\mathrm{O\,II}]}$-selected subsamples. We note that our model overestimates the overall clustering signal, especially at small scales. This may be caused by the assumption that the satellite fraction in each subsample is the same as that of the normal galaxies. However, in the real universe, satellite galaxies are expected to form earlier than central ones, so the probability that they are currently star-forming ELGs is relatively lower. Furthermore, the modeled satellite fractions f_{sat} are displayed as green circles in Figure 8. The value of the satellite fraction is close to 0.3, which is obviously higher than that found in current observational studies (e.g., Guo et al. 2019; Okumura et al. 2021), in which $f_{\mathrm{sat}} < 0.2$.

Motivated by these considerations, we introduce a free parameter f_{sat} to modulate the satellite fraction in our model. Equations (20) and (21) are rewritten as

$$P_x(M_{\mathrm{h},i}) = P_{\mathrm{cen},x}(L^{\mathrm{min}}_{[\mathrm{O\,II}],x} < L_{[\mathrm{O\,II}]} < L^{\mathrm{max}}_{[\mathrm{O\,II}],x}|M_{\mathrm{h},i}) = (1 - f_{\mathrm{sat}})\frac{\bar{n}_{\mathrm{h},i} + \bar{n}_{\mathrm{s},i}}{\bar{n}_{\mathrm{h},i}} \times \int_{-\infty}^{+\infty} F_{Lx}(M_*)p(M_*|M_{\mathrm{h},i})dM_* \quad (22)$$

and

$$P_x(M_{\mathrm{s},i}) = P_{\mathrm{sat},x}(L^{\mathrm{min}}_{[\mathrm{O\,II}],x} < L_{[\mathrm{O\,II}]} < L^{\mathrm{max}}_{[\mathrm{O\,II}],x}|M_{\mathrm{s},i}) = f_{\mathrm{sat}}\frac{\bar{n}_{\mathrm{h},i} + \bar{n}_{\mathrm{s},i}}{\bar{n}_{\mathrm{s},i}} \times \int_{-\infty}^{+\infty} F_{Lx}(M_*)p(M_*|M_{\mathrm{s},i})dM_*, \quad (23)$$

where f_{sat} is the satellite fraction. After the SHMR $p(M_*|M)$ is fixed with the best-fitting parameters shown in Figure 4, the ELG–halo connection can be determined completely by the only one free parameter f_{sat}.

Next, we constrain the parameter f_{sat} by fitting our model with the observed cross (auto) correlation functions. Similar to Equation (17), χ^2 for the ith $L_{[\mathrm{O\,II}]}$-selected subsample Li can be written as

$$\chi^2_{Li} = \chi^2_{LiLi} + \sum_{j=0}^{3}\chi^2_{LiMj}, \quad (24)$$

where we use one autocorrelation function $\boldsymbol{w}_{\mathrm{p}}^{LiLi}$ and four cross correlation functions $\boldsymbol{w}_{\mathrm{p}}^{LiMj}$ to infer the model parameter f_{sat}. The corresponding χ^2_{LiLi} and χ^2_{LiMj} can also be computed in analog to Equation (16). As a result, the dof in our fitting is $\mathrm{dof} = 12 \times 5 - 1 = 59$.

We show the best-fitting f_{sat} as well as the 1σ dispersion of their posterior distributions in Figure 8. It demonstrates that the best-fitting f_{sat} decreases as the $L_{[\mathrm{O\,II}]}$ increases, indicating that the [O II] lines are primarily generated by central galaxies rather than old satellites with little star formation. The value of f_{sat} in our model is also broadly consistent with other observational results (e.g., Guo et al. 2019; Okumura et al. 2021).

The best-fitting w_{p} as well as the 1σ uncertainties are displayed as the solid lines with shadowed regions in Figure 7. The cross correlation functions of the four ELG subsamples are well fitted. It suggests that the SHMR of normal galaxies can be used for ELGs with only the fraction of satellite galaxies reduced. This may indicate that the clustering of normal galaxies in the stellar mass range of ELGs does not depend on the star formation rate. As the ELGs are mostly in the stellar mass range $<10^{10}\,M_\odot$, we expect that normal central galaxies in this mass range at redshift $z \sim 0.7$ are dominantly star-forming galaxies, which supports why we can use the same SHMR. The lower fraction of the satellites indicates that the red satellites should not be included in the ELG sample. Our results are also broadly consistent with the finding of Hadzhiyska et al. (2021b) that DESI-like ELGs have a small assembly bias based on IllustrisTNG simulations. In addition, we note that the observed autocorrelations of $L2$ and $L3$ are slightly higher than

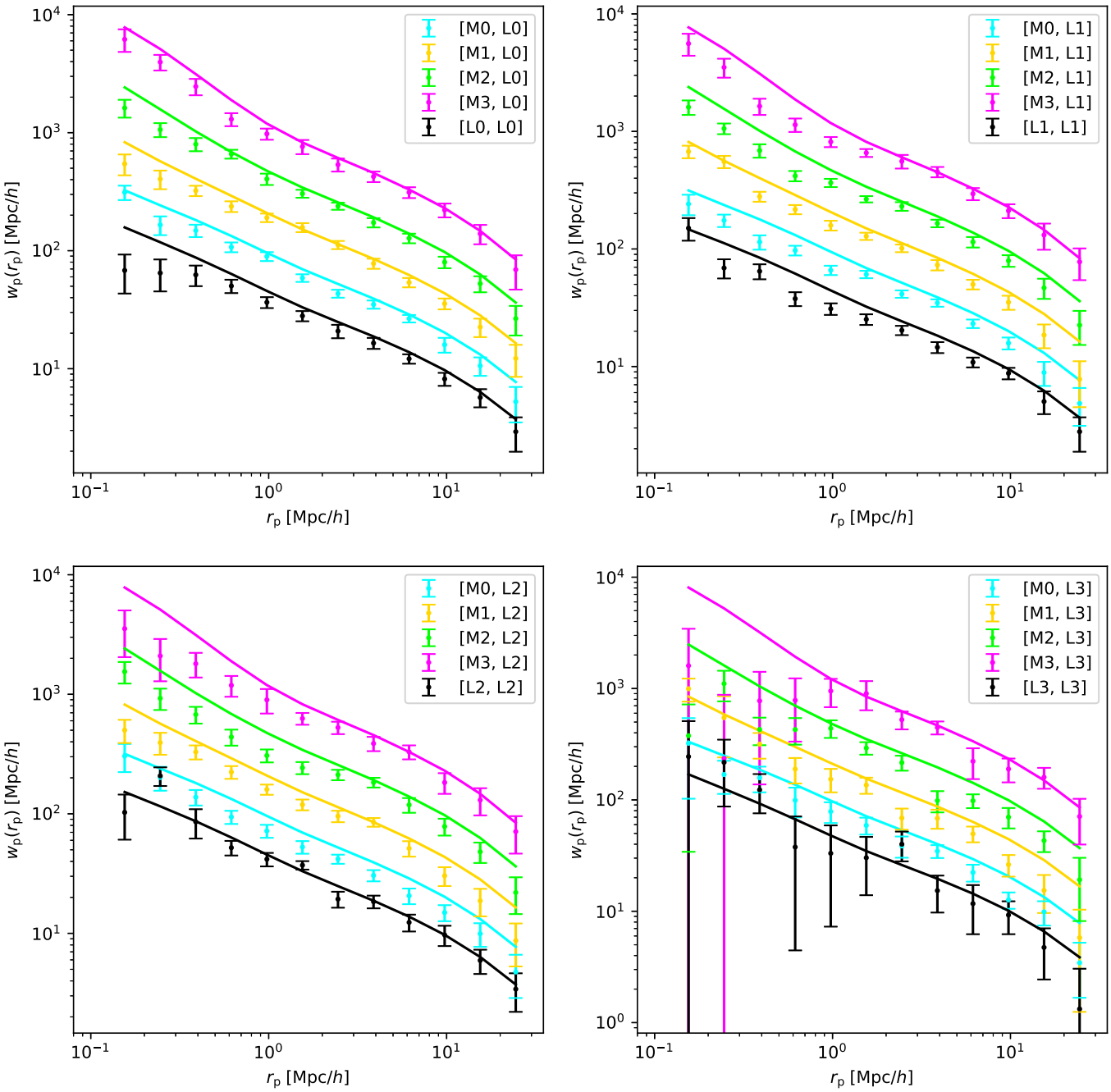

Figure 6. Observed projected cross (auto) correlation functions for the $L_{[O\ II]}$-selected subsamples, compared with the AM model predictions by randomly populating ELGs according to the SHMR and the fraction $F_L(M_*)$. The results of the four subsamples $L0$, $L1$, $L2$, and $L3$ are shown in four panels, respectively. The data points with error bars are measured from VIPERS. The model predictions are plotted as solid lines. Except for the autocorrelations (black), all the other cross correlations have been multiplied by 2^n where n changes with color ($n = 1$ (cyan), 2 (yellow), 3 (lime), and 4 (magenta)) to give a clear illustration.

our model predictions especially at small scales, although the errors are large. In the future, we will carefully investigate this issue using a much larger ELG sample from DESI.

5.3. Predicting the Correlation Functions of ELGs in Redshift Space

We further check the performance of our model predictions for the clustering in redshift space. By replacing $w_p(r_p)$ in Equation (12) with $\xi_0(s)$, $\xi_2(s)$, and $\xi_4(s)$, we can calculate these predicted multipole moments, which are presented in Figure 9 as solid curves. Although we have only fitted the observed real space $w_p(r_p)$, the multipole moments in redshift space predicted by our best-fitting model are also in good agreement with the observations. The ratios of the multiple moments between the observations and the model for the subsample $L0$ (top row in Figure 9) are shown in Figure 10. We can notice that the relative difference between the observed $\xi_0(s)$ and our model is always about 1σ, and less than $\sim$10% for those well-measured data points. For $\xi_2(s)$ and $\xi_4(s)$, although there are larger uncertainties in the measurements, the overall relative difference is still within about the 1σ confidence interval. We omit figures for the other luminosity subsamples, since their ratios have behaviors similar to that shown from the subsample $L0$.

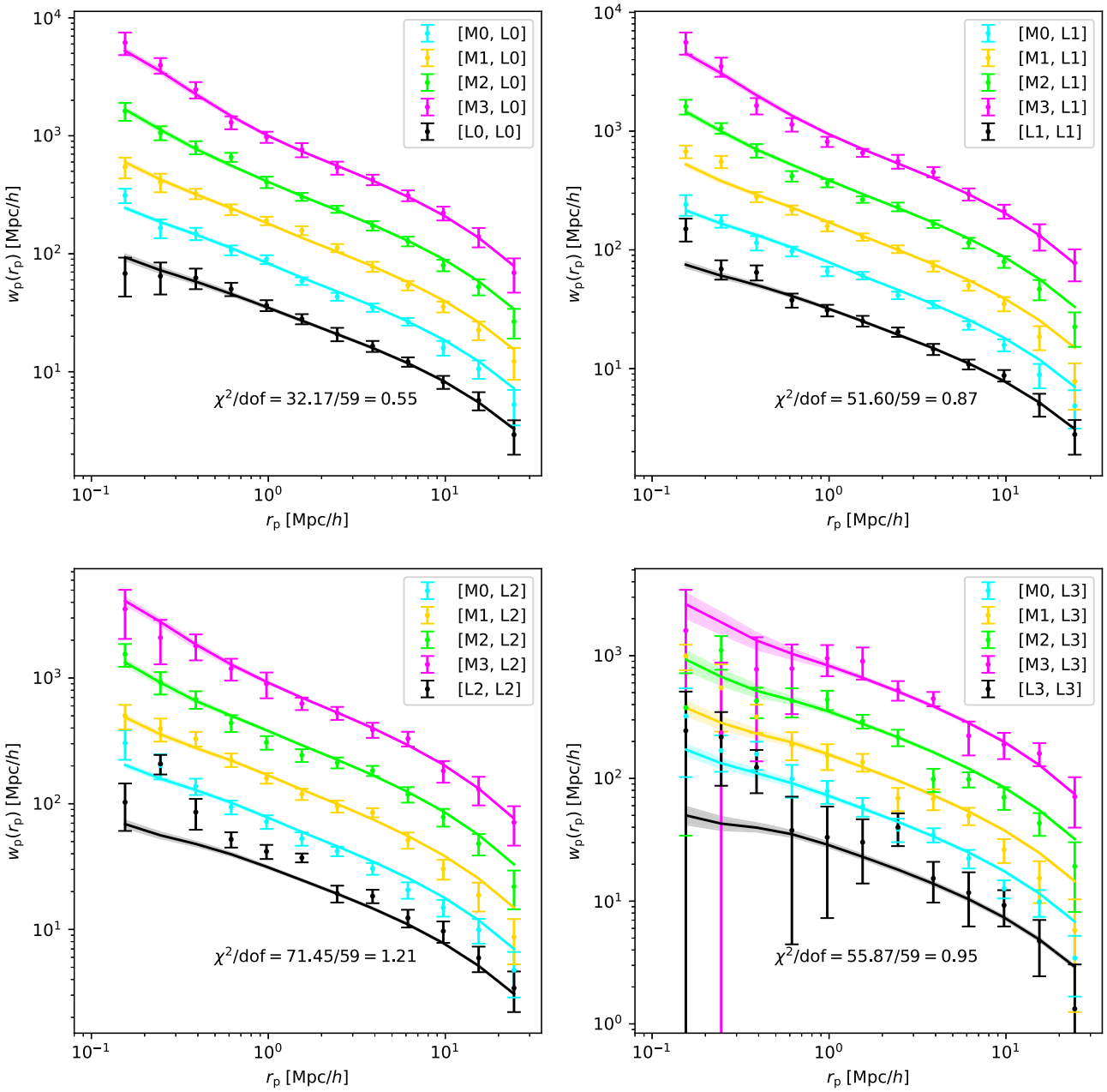

Figure 7. Similar to Figure 6, but the lines are the fitting results of our model with an adjustable satellite fraction f_{sat}. The best-fitting models as well as their 1σ error are plotted as solid lines with shadowed areas. The reduced χ^2 is also marked in each panel.

In general, our model can well reproduce the cross (auto) correlation functions in both real space and redshift space for the ELGs. This method can be easily applied to generate ELG mock catalogs for ongoing spectroscopic surveys such as DESI and PFS.

5.4. Implications for HOD Modeling

In order to compare with the traditional HOD models, we work out the HOD based on our model,

$$\begin{aligned} N_{\mathrm{cen},x}(M_{\mathrm{h}}) &= P_x(M_{\mathrm{h}}) \\ N_{\mathrm{sat},x}(M_{\mathrm{h}}) &= \int P_x(M_{\mathrm{s}}) n_{\mathrm{sub}}(M_{\mathrm{s}}|M_{\mathrm{h}}) dM_{\mathrm{s}} \\ N_x(M_{\mathrm{h}}) &= N_{\mathrm{cen},x}(M_{\mathrm{h}}) + N_{\mathrm{sat},x}(M_{\mathrm{h}}), \end{aligned} \tag{25}$$

where $N_{\mathrm{cen},x}(M_{\mathrm{h}})$, $N_{\mathrm{sat},x}(M_{\mathrm{h}})$, and $N_x(M_{\mathrm{h}})$ are the central, satellite, and total occupation numbers, respectively, in the ELG subsample x, and the probabilities $P_x(M_{\mathrm{h}})$ and $P_x(M_{\mathrm{s}})$ are calculated using Equations (22) and (23) but with a bin width $\Delta \log M = 0.1$. Here, $n_{\mathrm{sub}}(M_{\mathrm{s}}|M_{\mathrm{h}})$ measured from our simulation is the mean subhalo mass function at the given M_{h} bin. The HODs of the four [O II] luminosity-selected subsamples are shown in Figure 11 as solid lines, and the decomposed central and satellite occupation numbers are also displayed as dashed and dotted lines, respectively. Then we compare our model predictions with two recent HOD models of [O II] ELGs. It should be noted that since the HOD depends on the target selections of the ELG samples, we can only qualitatively compare the shapes of these HODs instead of their precise values.

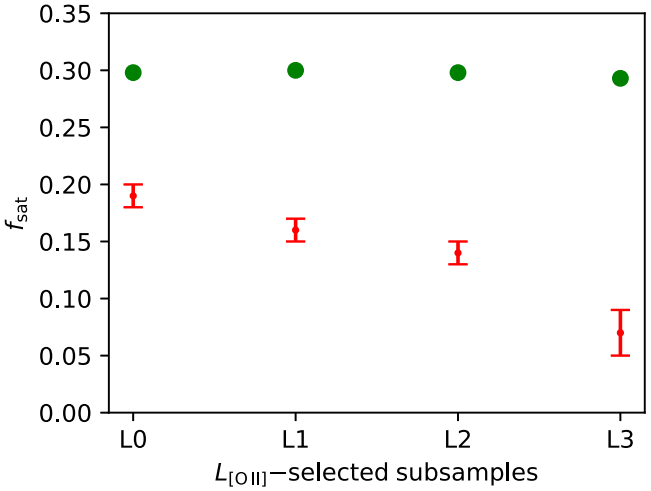

Figure 8. Our best-fitting satellite fractions f_{sat} for the four $L_{[O\,II]}$-selected subsamples (red points with error bars). For comparison, the green circles represent the f_{sat} if each $L_{[O\,II]}$-selected subsample is selected randomly based on the SHMR and $F_L(M_*)$.

The first is the Geach HOD model (Geach et al. 2012). Okumura et al. (2021) has constrained the model parameters based on the HSC NB observations of [O II] emitters at $z = 1.19$ and 1.47, and found that the model can well fit the angular correlation functions of the [O II] emitters. Considering that the parameters are better constrained at = 1.47, we adopt their model parameters at this redshift based on the posterior PDF (see their Table 3) and display the HOD as brown lines in Figure 11. The shape of their N_{cen} at low-mass end is quite similar to ours. However, with the Geach HOD form, N_{cen} in their model tends to be a constant at large halo mass, while our N_{cen} keeps decreasing. Although the HOD at massive end has a relatively small effect on galaxy clustering due to the rapid decline of the halo mass function, our results imply that a decreasing function $N_{cen} = \propto M^{\beta_c}$ ($\beta_c \sim -0.2$) can better describe the massive end of ELG HOD. On the other hand, both the Geach model and ours present a similar power-law form for N_{sat}.

The other is the HOD model (HOD-3 in their paper) proposed by Avila et al. (2020) for the eBOSS ELGs. This model combines a Gaussian function with a decaying power-law form to describe the central occupation. Avila et al. (2020) has constrained this model using the semi-analytical model (SAM) results (Gonzalez-Perez et al. 2018) as well as the eBOSS number density and bias (see their Table 2). We show this model as black curves in Figure 11. Although both models show a continuously reduced N_{cen} toward to the massive end, N_{cen} in the Avila model exhibits a faster decay after the peak. This difference might imply that the AGN feedback mechanism in the SAM (Gonzalez-Perez et al. 2018) is too strong, resulting in quick quenching of galaxies at the massive end. As for N_{sat}, a power-law form can indeed reasonably describe the N_{sat} of ELGs under the current data.

From the above comparison of the three models, we can conclude that since ELGs are mainly the central galaxies with small stellar mass, the form of N_{cen} at large stellar (halo) mass cannot be well constrained with the clustering data of ELGs only. Our results indicate that the following form can better describe the HOD of ELGs:

$$
\begin{aligned}
N_{\rm cen}(M) &= N^{\rm Exp}_{\rm cen,Geach}(M) + N^{\rm Erf}_{\rm cen,Geach}(M) \times \left(1 + \frac{M}{M_{\rm c}}\right)^{\beta_{\rm c}} \\
&= F_{\rm c}^{\rm B}(1 - F_{\rm c}^{\rm A}) \exp\left[-\frac{\log(M/M_{\rm c})^2}{2\sigma^2_{\log M}}\right] \\
&\quad + F_{\rm c}^{\rm A}\left[1 + {\rm erf}\left(\frac{\log(M/M_{\rm c})}{\sigma_{\log M}}\right)\right] \times \left(1 + \frac{M}{M_{\rm c}}\right)^{\beta_{\rm c}} \\
N_{\rm sat}(M) &= N_{\rm sat,Geach}(M) \\
&= F_{\rm s}\left[1 + {\rm erf}\left(\frac{\log(M/M_{\rm min})}{\delta_{\log M}}\right)\right]\left(\frac{M}{M_{\rm min}}\right)^{\alpha_{\rm s}},
\end{aligned}
\tag{26}
$$

where β_c characterizes the decay of N_{cen} at the high-mass end. This HOD preserves most features of the Geach HOD form, but remedies it by introducing a gradually declining function at large mass. On the other hand, the behavior of N_{cen} at large stellar (halo) mass can serve as a sensitive test for feedback models of galaxy formation.

To check the performance of our modified Geach form (Equation (26)), we compare it with our derived HODs of the four $L_{[O\,II]}$ -selected subsamples in Figure 12 (the same as those shown in Figure 11, but shown as hollow circles). We assume that the data points of the derived HODs are equally weighted (assuming a 10% error), and fit them with Equation (26). The corresponding best-fitting results of the modified HOD model are displayed as solid lines. We can see that this HOD model can accurately describe the ELG occupation numbers at all halo mass and at all [O II] luminosity. The parameters of this HOD model are listed in Table 4.

6. Summary

In this work, we constrain the ELG–halo connection using the auto and cross correlation functions of the galaxy subsamples from VIPERS. Combining the SHMR and ELG-stellar mass distribution, we provide a novel method to populate ELGs in cosmological simulations. Our main results are summarized as follows.

1. Using the galaxy catalog from VIPERS, we construct four stellar mass-selected subsamples and four [O II] luminosity-selected subsamples. We also take into account the redshift measurement uncertainty in our *N*-body simulation to make a fair comparison with the observations.
2. Both the angular and radial selection functions of VIPERS have been carefully corrected. Particularly, to account for the radial selection effects caused by the *i*-band limit and the color–color cut, we adopt the V_{max} technique to generate the redshift distribution for the random samples. For all the galaxy subsamples, we measure the projected cross (auto) correlation functions in real space and the multiple moments in redshift space.
3. To determine the SHMR, we apply the AM model proposed by Wang & Jing (2010) to our *N*-body simulation. The theoretical cross (auto) correlation functions of different M_*-selected subsamples are calculated by the tabulated method. We perform an

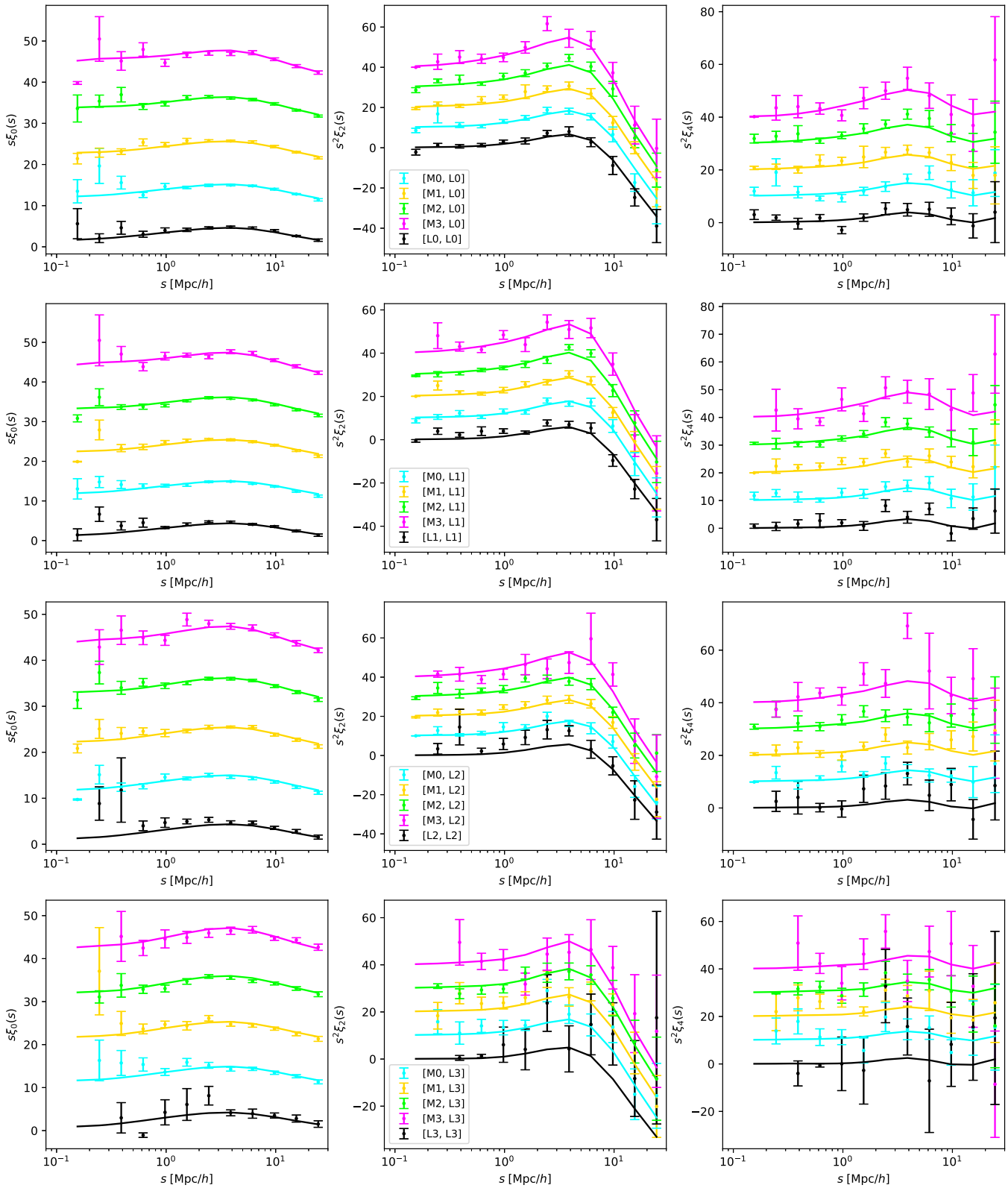

Figure 9. Cross (auto) correlation functions in redshift space for the $L_{[\mathrm{O\ II}]}$-selected subsamples in both observations and our models. The four rows from top to bottom represent the four subsamples *L*0, *L*1, *L*2, and *L*3. The observed monopole $s\xi_0(s)$, quadrupole $s^2\xi_2(s)$, and hexadecapole $s^2\xi_4(s)$ are displayed as data points with error bars in the three columns from left to right, respectively. Except for the black points, all the other data points have been shifted by $10 \times n$ where n changes with color ($n = 1$ (cyan), 2 (yellow), 3 (lime), and 4 (magenta)) to give a clear display. The solid lines represent our model predictions (*not fittings*).

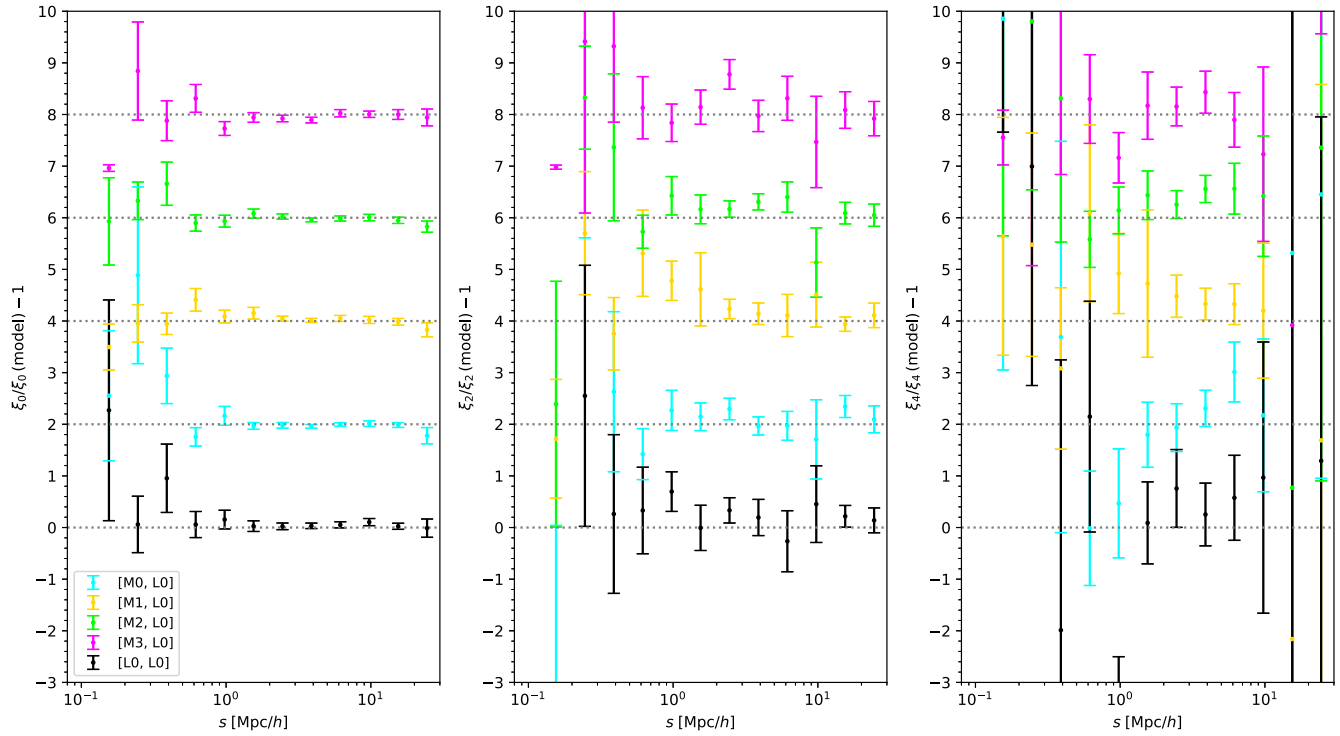

Figure 10. Similar to the top row of Figure 9, but the points with error bars represent the ratios of the multiple moments between the observations and the model predictions for the subsample *L0*. Except for the black points, all the other data points have been shifted by $2 \times n$ where n changes with color ($n = 1$ (cyan), 2 (yellow), 3 (lime), and 4 (magenta)) to give a clear display.

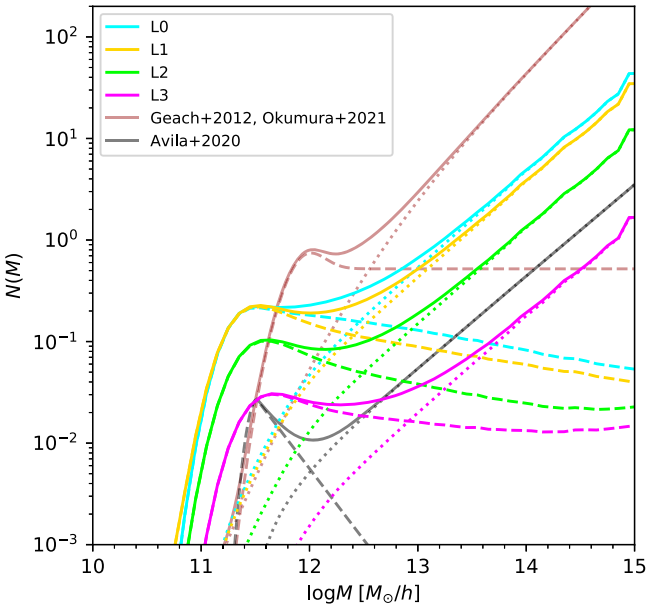

Figure 11. Comparison of different ELG HOD models. The predicted HODs of our four [O II] luminosity-selected subsamples are presented as cyan, yellow, lime, and magenta lines. In addition, we present the Geach HOD model (Geach et al. 2012) with the best-fitting parameters provided by Okumura et al. (2021) as brown lines. The HOD model proposed by Avila et al. (2020) is also shown as black lines. The solid, dashed, and dotted lines denote the total, central, and satellite occupation numbers, respectively.

MCMC analysis to explore the parameters space of SHMR. Our best-fitting SHMR can recover the observational correlation functions well.

4. We measure the ELG fractions $F_L(M_*)$ as a function of stellar mass in the four $L_{[\mathrm{O\ II}]}$-selected subsamples. We demonstrate that the clustering of ELGs can be well matched both in the real space and in the redshift space if we use the above SHMR to assign stellar mass to (sub) halos and then randomly select the ELGs according to their fractions $F_L(M_*)$ at a given stellar mass, as long as the satellite fraction f_{sat} is properly reduced. The method can be applied to constructing mock samples for ongoing and future redshift surveys, such as DESI, PFS, and Euclid.
5. We also derive the halo occupation numbers for the four ELG subsamples, and compare them with some of the previous HOD studies for ELGs. Our results indicate that the Geach form describes well the number of central galaxies at small and typical halo mass, but its constant form overpredicts the number at high halo mass. We propose a modified form, Equation (26), for describing the HOD of ELGs. The behavior at the high mass reflects the feedback processes in galaxy formation. In addition, the power-law form generally describes well the HOD of satellite galaxies.

In short, the cross correlations between ELGs and normal galaxies can play a significant role in constraining the ELG–halo connection. It is worth mentioning that our method can be combined with the Photometric objects Around Cosmic webs (PAC) method (Xu et al. 2022), which utilizes the cross correlation between a special spectroscopic sample (e.g., LRGs, QSOs) and a deep photometric sample, and thus can accurately measure the SHMR (SMF) in a wide stellar mass range. For the ongoing and future spectroscopic surveys, after

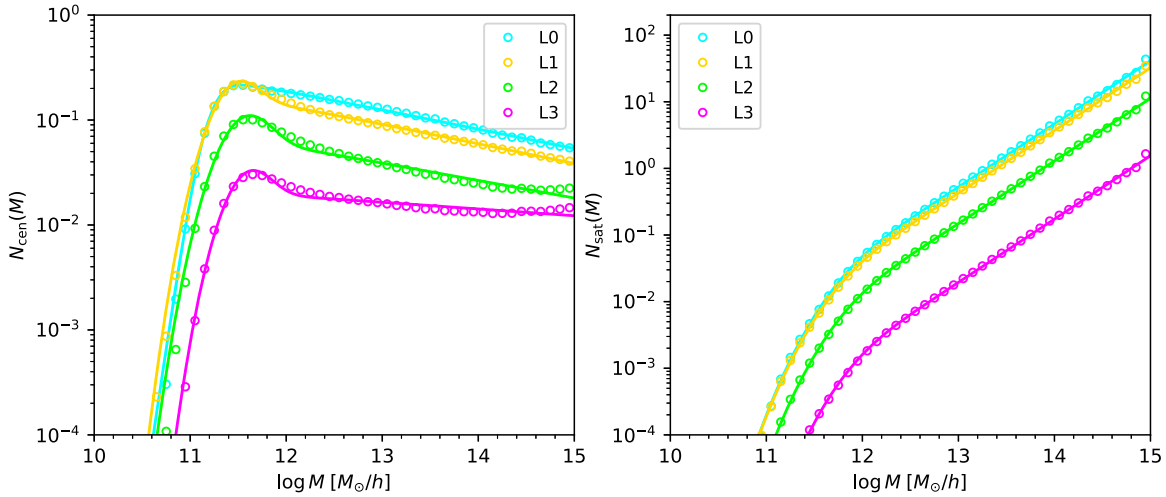

Figure 12. Comparison of the modified Geach HOD form (Equation (26)) with the occupations derived from our mock catalog. The hollow circles denote our derived HODs (same as those shown in Figure 11) of our mock catalogs for the four $L_{[\rm O\ II]}$ -selected subsamples. The best-fitting results are shown as solid lines. The left and right panels correspond to $N_{\rm cen}$ and $N_{\rm sat}$, respectively.

Table 4
Parameters of the Modified Geach Form for the Four Subsamples

Name	$\log M_{\rm c}\ (M_\odot\ h^{-1})$	$\sigma_{\log M}$	$F_{\rm c}^{\rm A}$	$F_{\rm c}^{\rm B}$	$\beta_{\rm c}$	$\log M_{\rm min}\ (M_\odot\ h^{-1})$	$F_{\rm s}$	$\delta_{\log M}$	$\alpha_{\rm s}$
*L*0	11.234	0.206	0.133	0.010	−0.185	11.690	0.015	0.516	0.947
*L*1	11.415	0.224	0.091	0.146	−0.187	11.668	0.012	0.516	0.939
*L*2	11.528	0.241	0.035	0.075	−0.168	11.723	0.005	0.508	0.940
*L*3	11.558	0.217	0.010	0.021	−0.065	11.783	0.001	0.492	0.950

the SHMR is determined using PAC, our method can be further developed and tested, and will provide a novel way to create [O II] ELG mock catalogs.

H.Y.G thanks Xiaokai Chen and Haojie Xu for their kind help. The work is supported by the NSFC (12133006, 11890691, 11621303) and by the 111 project No. B20019. We gratefully acknowledge the support of the Key Laboratory for Particle Physics, Astrophysics and Cosmology, Ministry of Education. This work made use of the Gravity Supercomputer at the Department of Astronomy, Shanghai Jiao Tong University.

This paper uses data from VIPERS. VIPERS has been performed using the ESO Very Large Telescope, under the "Large Programme" 182.A-0886. The participating institutions and funding agencies are listed at http://vipers.inaf.it. Based on observations collected at the European Southern Observatory, Cerro Paranal, Chile, using the Very Large Telescope under programs 182.A-0886 and partly 070.A-9007. Also based on observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/DAPNIA, at the Canada–France–Hawaii Telescope (CFHT), which is operated by the National Research Council (NRC) of Canada, the Institut National des Sciences de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. This work is based in part on data products produced at TERAPIX and the Canadian Astronomy Data Centre as part of the Canada–France–Hawaii Telescope Legacy Survey, a collaborative project of NRC and CNRS. This research uses data from the VIMOS VLT Deep Survey, obtained from the VVDS database operated by Cesam, Laboratoire d'Astrophysique de Marseille, France.

Software: NumPy (van der Walt et al. 2011), SciPy (Oliphant 2007), Matplotlib (Hunter 2007), Astropy (Astropy Collaboration et al. 2013), scikit-learn (Pedregosa et al. 2011), emcee (Foreman-Mackey et al. 2013).

Appendix A
Correction to the Small-scale Clustering

Using the 153 VIPERS mock samples provided by Pezzotta et al. (2017), we compute the angular correlation function $w_{\rm p}(\theta)$ for parent photometric galaxies and $w_{\rm s}(\theta)$ for spectroscopic galaxies which is obtained by applying the silt assign algorithm to the parent catalog. In Figure 13, we show the measurements of $w_{\rm p}(\theta)$ $(w_{\rm s}(\theta))$ in the top panel and $C(\theta) = [1 + w_{\rm s}(\theta)]/[1 + w_{\rm p}(\theta)]$ in the bottom panel. The two turning points of $C(\theta)$ clearly reflect the two typical scales that affect the small-scale clustering. We linearly interpolate the $C(\theta) - \log\theta$ relation and define the angular weight as $w^{\rm A}(\theta) = 1/C(\theta)$.

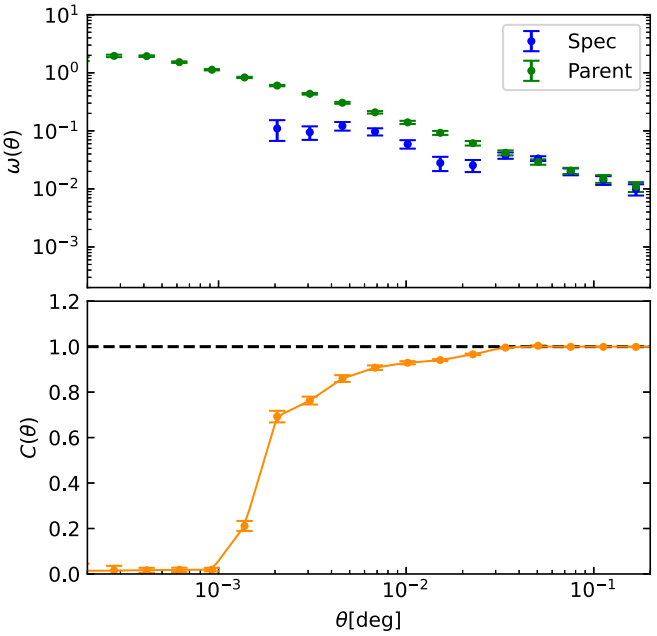

Figure 13. Angular completeness function of VIPERS. In the top panel, the data points with error bar show the angular correlation functions of parent $w_p(\theta)$ (green) and spectroscopic $w_s(\theta)$ (blue) samples measured from 153 VIPERS mock catalogs. The angular completeness function defined as $C(\theta) = [1 + w_s(\theta)]/[1 + w_p(\theta)]$ is plotted as orange points in the bottom panel. The orange solid line represents the linear interpolation of $C(\theta) - \log\theta$ relation.

Appendix B
Generating the Redshift Distribution for a Random Sample

The V_{max} method (Cole 2011; de la Torre et al. 2013, 2017; Pezzotta et al. 2017; Yang et al. 2020) is adopted to produce the radial distribution for random samples. It is in principle much better than the method of randomly shuffling the observed redshifts in generating a random sample. For each galaxy, we convert its z_{max} output by LE PHARE (Arnouts et al. 2002; Ilbert et al. 2006) into V_{max}, which represents the maximum volume of this galaxy that can be observed in VIPERS. Consider that the SSR and TSR have slight impact on the redshift distribution of sample (Pezzotta et al. 2017), we weight each galaxy by $w^{SSR} \times w^{TSR}$. Then we randomly select N_{ran} galaxies based on their probabilities (the probability of selecting different galaxies is not equal due to the weight). For each selected galaxy, we can generate a random point uniformly distributed in its V_{max} and convert the volume V_{ran} of this random point into its redshift z_{ran}. As a result, the original best-fitting SED of the galaxy should be shifted to z_{ran} as a new SED of this random point. We can calculate the u, g, r, i, z magnitudes for this random point and apply the color–color cut (Equation (1)) to it. Figure 14 presents the color space distributions of the random points generated by the above method for the full galaxy sample. We note that the color cut (lime solid line) can clearly distinguish random samples with redshifts lower than (blue points) and higher than (red points) 0.5. The radial distributions of random samples generated by this nonparametric method are shown in Figure 2 for the total sample and in Figures 15 and 16 for the subsamples.

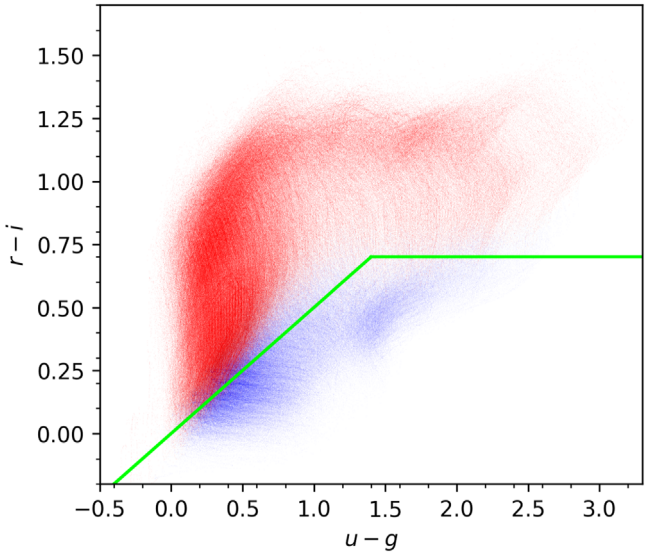

Figure 14. Color space distributions of the random points generated for the full galaxy sample. The blue (red) points denote the random points with $z < 0.5$ ($z > 0.5$). The lime solid line represents the VIPERS color–color cut (Equation (1)) used to exclude the galaxies with $z < 0.5$.

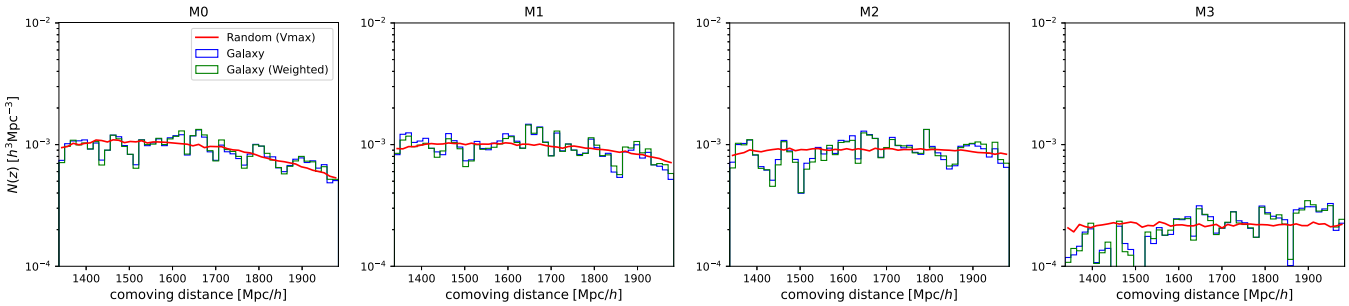

Figure 15. Radial distributions of the M_*-selected subsamples. Similar to Figure 2, the un-weighted (weighted) number densities as function of comoving distance for different subsamples are shown as blue (green) histograms. The radial distributions of corresponding random subsamples are plotted as red solid lines.

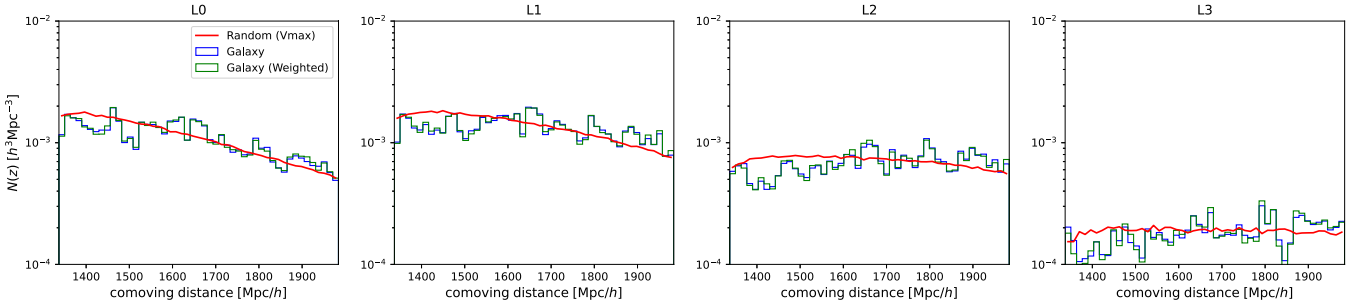

Figure 16. Similar to Figure 15, the radial distributions of the $L_{[\mathrm{O\ II}]}$ -selected subsamples.

ORCID iDs

Y. P. Jing https://orcid.org/0000-0002-4534-3125
Kun Xu https://orcid.org/0000-0002-7697-3306

References

Arnouts, S., Moscardini, L., Vanzella, E., et al. 2002, MNRAS, 329, 355
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Avila, S., Gonzalez-Perez, V., Mohammad, F. G., et al. 2020, MNRAS, 499, 5486
Behroozi, P., Wechsler, R. H., Hearin, A. P., & Conroy, C. 2019, MNRAS, 488, 3143
Behroozi, P. S., Conroy, C., & Wechsler, R. H. 2010, ApJ, 717, 379
Berlind, A. A., & Weinberg, D. H. 2002, ApJ, 575, 587
Bryan, G. L., & Norman, M. L. 1998, ApJ, 495, 80
Chaves-Montero, J., Angulo, R. E., Schaye, J., et al. 2016, MNRAS, 460, 3100
Cole, S. 2011, MNRAS, 416, 739
Cole, S., Percival, W. J., Peacock, J. A., et al. 2005, MNRAS, 362, 505
Comparat, J., Kneib, J.-P., Escoffier, S., et al. 2013, MNRAS, 428, 1498
Comparat, J., Zhu, G., Gonzalez-Perez, V., et al. 2016, MNRAS, 461, 1076
Contreras, S., Baugh, C. M., Norberg, P., & Padilla, N. 2013, MNRAS, 432, 2717
Cooper, M. C., Gallazzi, A., Newman, J. A., & Yan, R. 2010, MNRAS, 402, 1942
Cooray, A. 2006, MNRAS, 365, 842
Cui, W., Davé, R., Peacock, J. A., Anglés-Alcázar, D., & Yang, X. 2021, NatAs, 5, 1069
Davidzon, I., Bolzonella, M., Coupon, J., et al. 2013, A&A, 558, A23
Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, ApJ, 292, 371
Davis, M., & Peebles, P. J. E. 1983, ApJ, 267, 465
Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, AJ, 151, 44
de la Torre, S., Guzzo, L., Peacock, J. A., et al. 2013, A&A, 557, A54
de la Torre, S., Jullo, E., Giocoli, C., et al. 2017, A&A, 608, A44
DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, arXiv:1611.00036
Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al. 2005, ApJ, 633, 560
Favole, G., Comparat, J., Prada, F., et al. 2016, MNRAS, 461, 3421
Favole, G., Rodríguez-Torres, S. A., Comparat, J., et al. 2017, MNRAS, 472, 550
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Gao, H., & Jing, Y. P. 2021, ApJ, 908, 43
Garilli, B., Guzzo, L., Scodeggio, M., et al. 2014, A&A, 562, A23
Geach, J. E., Sobral, D., Hickox, R. C., et al. 2012, MNRAS, 426, 679
Gonzalez-Perez, V., Comparat, J., Norberg, P., et al. 2018, MNRAS, 474, 4024
Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
Gunn, J. E., Gott, J., & Richard, I. 1972, ApJ, 176, 1
Gunn, J. E., Siegmund, W. A., Mannery, E. J., et al. 2006, AJ, 131, 2332
Guo, H., Yang, X., & Lu, Y. 2018, ApJ, 858, 30
Guo, H., Yang, X., Raichoor, A., et al. 2019, ApJ, 871, 147
Guo, H., Zheng, Z., Behroozi, P. S., et al. 2016, MNRAS, 459, 3040
Guo, Q., & White, S. 2014, MNRAS, 437, 3228
Guo, Q., White, S., Li, C., & Boylan-Kolchin, M. 2010, MNRAS, 404, 1111
Guzzo, L., Scodeggio, M., Garilli, B., et al. 2014, A&A, 566, A108
Hadzhiyska, B., Bose, S., Eisenstein, D., & Hernquist, L. 2021a, MNRAS, 501, 1603
Hadzhiyska, B., Bose, S., Eisenstein, D., Hernquist, L., & Spergel, D. N. 2020, MNRAS, 493, 5506
Hadzhiyska, B., Tacchella, S., Bose, S., & Eisenstein, D. J. 2021b, MNRAS, 502, 3599
Hamilton, A. J. S. 1992, ApJL, 385, L5
Han, J., Cole, S., Frenk, C. S., Benitez-Llambay, A., & Helly, J. 2018, MNRAS, 474, 604
Han, J., Jing, Y. P., Wang, H., & Wang, W. 2012, MNRAS, 427, 2437
Hartlap, J., Simon, P., & Schneider, P. 2007, A&A, 464, 399
Hearin, A. P., Watson, D. F., & van den Bosch, F. C. 2015, MNRAS, 452, 1958
Hunter, J. D. 2007, CSE, 9, 90
Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, A&A, 457, 841
Jarvis, M. J., Bonfield, D. G., Bruce, V. A., et al. 2013, MNRAS, 428, 1281
Jiang, C. Y., Jing, Y. P., Faltenbacher, A., Lin, W. P., & Li, C. 2008, ApJ, 675, 1095
Jing, Y. 2019, SCPMA, 62, 19511
Jing, Y. P., Mo, H. J., & Börner, G. 1998, ApJ, 494, 1
Jing, Y. P., & Suto, Y. 2002, ApJ, 574, 538
Kaiser, N. 1987, MNRAS, 227, 1
Kocevski, D. D., Lemaux, B. C., Lubin, L. M., et al. 2011, ApJL, 737, L38
Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
Landy, S. D., & Szalay, A. S. 1993, ApJ, 412, 64
Le Fèvre, O., Saisse, M., Mancini, D., et al. 2003, Proc. SPIE, 4841, 1670
Levenberg, K. 1944, QApMa, 2, 164
Ma, C.-P., & Fry, J. N. 2000, ApJ, 543, 503
Mandelbaum, R., Wang, W., Zu, Y., et al. 2016, MNRAS, 457, 3200
Marquardt, D. W. 1963, J. Soc. Ind. Appl. Math., 11, 431
Martin, D. C., Fanson, J., Schiminovich, D., et al. 2005, ApJL, 619, L1

McLeod, D. J., McLure, R. J., Dunlop, J. S., et al. 2021, MNRAS, 503, 4413
Moster, B. P., Naab, T., & White, S. D. M. 2013, MNRAS, 428, 3121
Moutard, T., Arnouts, S., Ilbert, O., et al. 2016, A&A, 590, A102
Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, ApJS, 208, 5
Okumura, T., Hayashi, M., Chiu, I. N., et al. 2021, PASJ, 73, 1186
Oliphant, T. E. 2007, CSE, 9, 10
Osato, K., Nishimichi, T., & Takada, M. 2022, MNRAS, 511, 1131
Peacock, J. A., & Smith, R. E. 2000, MNRAS, 318, 1144
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, JMLR, 12, 2825, http://jmlr.org/papers/v12/pedregosa11a.html
Pezzotta, A., de la Torre, S., Bel, J., et al. 2017, A&A, 604, A33
Pozzetti, L., Bolzonella, M., Lamareille, F., et al. 2007, A&A, 474, 443
Pozzetti, L., Bolzonella, M., Zucca, E., et al. 2010, A&A, 523, A13
Raichoor, A., Comparat, J., Delubac, T., et al. 2017, MNRAS, 471, 3955
Rodríguez-Puebla, A., Avila-Reese, V., Yang, X., et al. 2015, ApJ, 799, 130
Rodríguez-Torres, S. A., Chuang, C.-H., Prada, F., et al. 2016, MNRAS, 460, 1173
Schechter, P. 1976, ApJ, 203, 297
Scodeggio, M., Guzzo, L., Garilli, B., et al. 2018, A&A, 609, A84
Seljak, U. 2000, MNRAS, 318, 203
Simha, V., Weinberg, D. H., Davé, R., et al. 2012, MNRAS, 423, 3458
Szapudi, I., & Szalay, A. S. 1998, ApJL, 494, L41
Takada, M., Ellis, R. S., Chiba, M., et al. 2014, PASJ, 66, R1
Vale, A., & Ostriker, J. P. 2006, MNRAS, 371, 1173
van den Bosch, F. C., Yang, X., Mo, H. J., et al. 2007, MNRAS, 376, 841
van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, CSE, 13, 22
Wang, L., & Jing, Y. P. 2010, MNRAS, 402, 1796
Wang, L., Li, C., Kauffmann, G., & De Lucia, G. 2006, MNRAS, 371, 537
Wang, L., Weinmann, S. M., De Lucia, G., & Yang, X. 2013, MNRAS, 433, 515
Wang, Z., Xu, H., Yang, X., et al. 2019, ApJ, 879, 71
Wechsler, R. H., Gross, M. A. K., Primack, J. R., Blumenthal, G. R., & Dekel, A. 1998, ApJ, 506, 19
Wechsler, R. H., & Tinker, J. L. 2018, ARA&A, 56, 435
Xu, H., Zheng, Z., Guo, H., et al. 2018, MNRAS, 481, 5470
Xu, H., Zheng, Z., Guo, H., Zhu, J., & Zehavi, I. 2016, MNRAS, 460, 3647
Xu, K., & Jing, Y. 2022, ApJ, 926, 130
Xu, K., Zheng, Y., & Jing, Y. 2022, ApJ, 925, 31
Xu, X., & Zheng, Z. 2020, MNRAS, 492, 2739
Yang, L., Jing, Y.-P., Li, Z.-G., & Yang, X.-H. 2020, RAA, 20, 054
Yang, X., Mo, H. J., & van den Bosch, F. C. 2003, MNRAS, 339, 1057
Yang, X., Mo, H. J., & van den Bosch, F. C. 2009, ApJ, 695, 900
Yang, X., Mo, H. J., van den Bosch, F. C., Zhang, Y., & Han, J. 2012, ApJ, 752, 41
York, D. G., Adelman, J., Anderson, J. E. J., et al. 2000, AJ, 120, 1579
Yuan, S., Eisenstein, D. J., & Garrison, L. H. 2018, MNRAS, 478, 2019
Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2011, ApJ, 736, 59
Zentner, A. R., Hearin, A. P., & van den Bosch, F. C. 2014, MNRAS, 443, 3044
Zheng, Z., Berlind, A. A., Weinberg, D. H., et al. 2005, ApJ, 633, 791
Zheng, Z., Coil, A. L., & Zehavi, I. 2007, ApJ, 667, 760
Zheng, Z., & Guo, H. 2016, MNRAS, 458, 4015
Zu, Y., & Mandelbaum, R. 2015, MNRAS, 454, 1161
Zu, Y., & Mandelbaum, R. 2016, MNRAS, 457, 4360
Zu, Y., & Mandelbaum, R. 2018, MNRAS, 476, 1637
Zu, Y., Shan, H., Zhang, J., et al. 2021, MNRAS, 505, 5117
Zu, Y., Song, Y., Shao, Z., et al. 2022, MNRAS, 511, 1789