The Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: GLAM-QPM mock galaxy catalogs for the Emission Line Galaxy Sample

Sicheng Lin,† Jeremy L. Tinker,1 Anatoly Klypin,2 Francisco Prada,3 Michael R. Blanton,1 Johan Comparat,4 Kyle S. Dawson,5 Arnaud de Mattia,6 Hélio du Mas des Bourboux,5 Will J. Percival,7,8,9 Anand Raichoor,10 Graziano Rossi,11 Alex Smith,6 Cheng Zhao10

1 Center for Cosmology and Particle Physics, Department of Physics, New York University, 726 Broadway, New York, NY 10003, USA
2 Astronomy Department, New Mexico State University, Las Cruces, NM, USA
3 Instituto de Astrofísica de Andalucía (CSIC), Glorieta de la Astronomía, E-18080 Granada, Spain
4 Max-Planck-Institut für extraterrestrische Physik (MPE), Giessenbachstrasse 1, D-85748 Garching bei München, Germany
5 University of Utah, Department of Physics and Astronomy, 115 S 1400 E, Salt Lake City, UT 84112, USA
6 IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France
7 Waterloo Centre for Astrophysics, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada
8 Department of Physics and Astronomy, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada
9 Perimeter Institute for Theoretical Physics, 31 Caroline St. North, Waterloo, ON N2L 2Y5, Canada
10 Institute of Physics, Laboratory of Astrophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland
11 Department of Physics and Astronomy, Sejong University, Seoul 143-747, Korea

ABSTRACT
We present 2000 mock galaxy catalogs for the analysis of baryon acoustic oscillations in the Emission Line Galaxy (ELG) sample of the Extended Baryon Oscillation Spectroscopic Survey Data Release 16 (eBOSS DR16). Each mock catalog has a number density of $6.7 \times 10^{-4} h^3$ Mpc$^{-3}$, covering a redshift range from 0.6 to 1.1. The mocks are calibrated to small-scale eBOSS ELG clustering measurements at scales of around $10 h^{-1}$ Mpc. The mock catalogs are generated using a combination of GaLAxy (GLAM) simulations and the Quick Particle-Mesh (QPM) method. GLAM simulations are used to generate the density field, which is then assigned dark matter halos using the QPM method. Halos are populated with galaxies using a halo occupation distribution (HOD). The resulting mocks match the survey geometry and selection function of the data, and have slightly higher number density which allows room for systematic analysis. The large-scale clustering of mocks at the baryon acoustic oscillation (BAO) scale is consistent with data and we present the correlation matrix of the mocks.

Key words: large-scale structure of Universe – halo – simulations

1 INTRODUCTION
In modern cosmology, the study of the large-scale structure (LSS) provides key information about the expansion history and growth of structure in the Universe (Davis & Peebles 1983; Eisenstein et al. 2005). Measurements of baryon acoustic oscillations (BAO; Eisenstein & Hu 1998) and redshift space distortions (RSD; Kaiser 1987) require spectroscopic surveys to cover large volumes and to have accurate redshift measurements. In the past, large redshift surveys have included 2dFGRS (Cole et al. 2005), the Sloan Digital Sky Survey (SDSS-II; Eisenstein et al. 2005), 6dFGRS (Beutler et al. 2012), WiggleZ (Blake et al. 2011) and SDSS-III Baryon Oscillation Spectroscopic Survey (BOSS; Dawson et al. 2013). The recently completed extended Baryon Oscillation Spectroscopic Survey (eBOSS; Dawson et al. 2016) is a five year program of the Sloan Sky Digital Survey (SDSS-IV; Blanton et al. 2017) and is aiming at measuring the distance-redshift relation with BAO at the percent-level using various galaxy tracers.

One important question for big redshift surveys is how to determine the uncertainties in the measurements of cosmological parameters. Simulated mock catalogs can be used to estimate the covariance matrix of galaxy clustering and these errors are propagated to cosmological parameter uncertainties by integrating over the parameter likelihood function (Dodelson & Schneider 2013; Percival et al. 2014). This method requires accurate mock catalogs to cover a huge volume and large number density of galaxies as the survey geometry and redshift range grow larger. Determining the uncertainties of a large survey such as eBOSS requires thousands of mock catalogs, which can be computationally expensive. In recent years, several quick methods such as quick particle mesh (QPM; White et al. 2014), effective ZelâĂŹdovich approximation (EZmocks; Chuang et al. 2015) and PerturbAtion Theory Catalog generator of Halo and galaxy Y distributions (PATCHY; Kitaura et al. 2014), have been developed to generate mocks efficiently. EZmocks use an ad-hoc
model to populate galaxies directly on the dark matter field which is generated using the Zeldovich approximation. A similar approach is used by PATCHY but using Augmented Lagrangian Perturbation Theory (ALPT) instead, while the QPM method selects dark matter particles that mimic the statistics of dark matter halos, and then populate galaxies in dark matter halos using a standard halo occupation approach.

We present a set of mock catalogues for the eBOSS ELG sample, which have been tuned to reproduce the small-scale clustering measurements. In this work, we focus on the ELG sample in the $0.6 < z < 1.1$ redshift range. We use the GLAM N-body simulations (Klypin & Prada 2018) and the QPM method to construct the ersatz halo catalogs. Galaxies are then populated using the halo occupation distribution (HOD) model to match the two-point statistics of the eBOSS ELG measurements. We produce a set of 2000 accurate mock catalogs for the estimation of the covariance matrix. The mock catalogs are calibrated to match the ELG clustering at 10 $h^{-1}$Mpc scales.

This study is part of a series of papers of the final eBOSS DR16 data and cosmological measurements. The BAO and RSD results are presented in Bautista et al. (2020) and Gil-Marín et al. (2020) for luminous red galaxies; for emission line galaxies see Raichoor (2020), Tamone et al. (2020) and de Mattia (2020); and see Hou et al. (2020), Neveux et al. (2020) for quasars. The essential data catalogs are presented in Ross et al. (2020) and Lyke et al. (2020), the N-body mocks for systematic errors are presented in Rossi et al. (2020) and Smith et al. (2020), another set of approximate mocks is presented in Zhao et al. (2020). The ELG mock challenge result is presented in Alam et al. (2020), the ELG HOD analysis is presented in Avila et al. (2020). The measurements of BAO in the Ly-α forest is presented in du Mas des Bourboux et al. (2020). Lastly, the cosmological interpretation of the full eBOSS sample can be found in Collaboration et al. (2020).

This paper is organized as follows: Section 2 describes the eBOSS ELG sample used in the analysis. In Section 3, we present our small-scale clustering measurements, with systematic corrections, that we use to calibrate our mock catalogs. The procedure of creating mock catalogs is described in Section 4. Section 5 compares the large-scale clustering of the ELG sample and GLAM-QPM mocks, and we present the covariance matrix. Finally we discuss our results in Section 6. In this paper, the distances are measured in units of $h^{-1}$Mpc with the Hubble constant $H_0 = 100h$ km s$^{-1}$Mpc$^{-1}$. We assume a fiducial $\Lambda$CDM cosmology with parameters $(\Omega_m, h, \Omega_b h^2, \sigma_8, n_s) = (0.307, 0.678, 0.022, 0.828, 0.96)$ for the redshift-distance relationship and mock catalog generation.

2 DATA

The eBOSS survey was conducted using the Sloan Foundation 2.5-m Telescope at Apache Point Observatory (Gunn et al. 2006) to conduct spectroscopic observations. It used the same 1000-fiber spectrographs as BOSS (Smee et al. 2013) to measure four tracers of the underlying dark matter density field. In data release 16, the survey has measured accurate redshifts of 174,816 luminous red galaxies (LRGs) in the redshift range $0.6 < z < 1.0$ (Ross et al. 2020), 173,736 emission line galaxies (ELGs) in the redshift range $0.6 < z < 1.1$ (Raichoor 2020), 343,708 quasars within $0.8 < z < 2.2$ (Ross et al. 2020; Lyke et al. 2020) and 210,005 Ly$\alpha$ quasars within $z > 2.1$ (du Mas des Bourboux et al. 2020). Overall these datasets constrain the redshift-distance relation to 1% – 3% level at 4 redshifts (Dawson et al. 2016; Collaboration et al. 2020).

The eBOSS/ELG program started in September 2016 and it is the first time ELG tracers have been used in SDSS for large-scale clustering measurements. Preliminary work tested the feasibility of using the BOSS spectrograph to conduct ELG observations (Comparat et al. 2013) and the reliability of redshift measurements (Comparat et al. 2015, 2016). The ELG target selection (Raichoor et al. 2017) uses the DECam Legacy Survey (DECaLS; Dey et al. 2018). Its deep imaging data provides the opportunity to reach higher redshift and higher efficiency, defined as percentage of observed ELGs having reliable $z_{spec}$ (Raichoor et al. 2017). The eBOSS/ELG footprint covers an effective area of 369.5 deg$^2$ in the north galactic cap (NGC) and 357.5 deg$^2$ in the south galactic cap (SGC), with an overall number density of 313.0 deg$^{-2}$ (Raichoor 2020).

We use the eBOSS DR16 ELG sample as our dataset. This dataset includes 83,769 galaxies in SGC and 89,967 galaxies in NGC with reliable redshift measurements over the redshift range $0.6 < z < 1.1$. Fig. 1 shows the footprint of ELG SGC and NGC data colored by completeness in each sector, where a sector is the geometric region defined by the unique set of overlapping plates. The completeness in each sector is defined as

$$C = \frac{N_{obs} + N_{cp}}{N_{targ}},$$

where $N_{obs}$ is the number of observed targets per sector, $N_{cp}$ is the number of galaxies with no spectra due to fiber collisions with other targets, and $N_{targ}$ is the total number of targets.

3 GALAXY CLUSTERING

Galaxies are not randomly distributed in the universe. Density perturbations, which are created in the early Universe, evolve under gravitational attraction, and are the seeds of the large-scale structure we see today. In this paper, study the statistics of the ELG galaxy distribution in the eBOSS survey using the two-point correlation function, $\xi(r) = \langle \delta(x)\delta(x+r) \rangle$, which is a measure of the excess probability of finding a pair of galaxies, separated by a distance $r$, compared to if the galaxies were distributed randomly (Peebles 1980). Measurements of clustering on large scales allow us to constrain cosmological parameters via measuring the position of BAO peak and the shape of the clustering signal. On small scales, clustering measurements can be used to probe the relationship between galaxy properties and dark matter halos. Thus, we focus on small-scale clustering to calibrate the adopted HOD model to generate the ELG mocks. We will discuss the HOD model further in Sec. 4.3.

We use the Landy & Szalay estimator (Landy & Szalay 1993) to measure the two-point correlation function,

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

where $DD(r), DR(r)$ and $RR(r)$ are suitably normalized numbers of galaxy-galaxy, galaxy-random, and random-random pairs in each distance separation bin. The distance along the line-of-sight of galaxies is inferred from their redshifts assuming a fiducial cosmological model. The peculiar velocities of galaxies will introduce redshift-space distortions in $\xi(r)$. In order to circumvent the effect of redshift-space distortions, the correlation function is often measured in two dimensions: perpendicular ($r_p$) and along ($\pi$) the line-of-sight. Let $v_1$ and $v_2$ be the position vector of a pair of galaxies in redshift-space, $s = v_1 - v_2$ be the redshift space separation, and $I = (v_1 - v_2)^2/2$ be...
The correlation function into the multipole moments, the integral, as the clustering measurements are noisy for large $n$. In this paper, we choose recovered by integrating over the line-of-sight direction to remove $\pi$.

The projected correlation function (Davis & Peebles 1983) can be measured to define as $\rho = \pi_0 \theta_0 = \sqrt{\pi - \pi^2}$. One can then measure the angular correlation function of galaxies with $\mu = \cos \theta$ the cosine of the angle between the pair separation vector and line of sight. It is often useful to compress the redshift space information in the two-dimensional correlation function into the multipole moments, $\xi(l, \mu), \mu$ are Legendre polynomials.

We consider the effects of imaging systematics, redshift failures and fiber collisions. These are corrected for using a weighting scheme similar to (Anderson et al. 2014), where each ELG is weighed by $\rho = wFD \rho_{\text{sys}} w_{\text{cp}} w_{\text{nzo}}$, where $w_{\text{sys}}$ is the imaging systematics weight, $w_{\text{cp}}$ is the close-pair weight and $w_{\text{nzo}}$ is the redshift failure weight. We describe them in detail below. Unlike Anderson et al. (2014), here we treat $w_{\text{cp}}$ and $w_{\text{nzo}}$ independently.

The systematic weights, $w_{\text{sys}}$, are calculated by using a linear fit of the ELG target density in various photometric variables: galactic extinction, stellar density, HI density, $grz$-band image seeing, and $grz$-band image depth (Raichoor 2020). The close-pair weights, $w_{\text{cp}}$, are calculated by upweighting galaxies in collided pairs by coefficient $N_{\text{target}}/N_{\text{fiber}}$, where $N_{\text{target}}$ is the total number of targets in the collision group and $N_{\text{fiber}}$ is the number of targets in the collision group that has been assigned fibers. The redshift failure weights, $w_{\text{nzo}}$, are the inverse of two fitting functions to the plate-averaged signal-to-noise ratio (pSN) and the fiber position in the focal plane. Details of the systematic model is described in the eBOSS ELG catalog paper (Raichoor 2020).

The FKP weights are defined as $w_{\text{FKP}}(z) = \frac{1}{1 + n(z) P_{\text{FKP}}}$. where $n(z)$ is the weighted number density at redshift $z$ and we choose the same value of $P_{\text{FKP}} = 4000 h^{-3} \text{Mpc}^3$ as in Raichoor (2020). We use FKP weights to account for different number densities of observed ELGs in different redshift intervals.

In addition to using the close pair weights to up-weight galaxies in collided pairs, we also apply an angular weighting to correct the small-scale clustering measurements, using the ratio of the angular correlation functions (Hawkins et al. 2003). This ratio is given by $F(\theta) = \frac{1 + w_{\text{cp}}(\theta)}{1 + w_{\text{l}}(\theta)}$ (8), where $w_{\text{cp}}(\theta)$ is the angular correlation function of galaxies with fibers assigned and $w_{\text{l}}(\theta)$ is the angular correlation function of the parent target samples. We find that the average ratio $F(\theta)$ for angular separations $\theta < 62''$ is around 0.6 and is not sensitive to the value of $\theta$, thus we upweight each galaxy-galaxy pair below $62''$ by $1/F(\theta) = 1.667$.

Fig. 2 presents our results of the small-scale clustering from 0.34$h^{-1}$Mpc to 70$h^{-1}$Mpc in 12 logarithmic bins, after applying the weighting described above. The SGC and NGC are divided into 25 roughly equal spherical areas and the errors are estimated using the jackknife resampling technique. There is no significant difference seen between the clustering of the SGC and NGC.

### 4 Mock Generation

In order to construct accurate mocks to interpret the ELG clustering, we combine GaLaxy Mocks (GLAM) simulations with the Quick Particle-Mesh (QPM) scheme. The whole process can be summarized in the following steps:

- In the first step we run 2000 large GLAM N-body simulations with box size of 3000$h^{-1}$Mpc. This box size is large enough to cover the whole footprint of ELGs up to redshift $z = 1.1$.
- In the second step we apply the QPM code to assign dark matter halos within the density field of the GLAM simulations.
- In the third step we populate the halos with galaxies using a HOD model that is calibrated to reproduce the small-scale clustering measurements of the data.
- In the fourth step, we cut the mock catalogues according to the ELG survey geometry. We compare the large-scale clustering of the mocks with the data.

In the next sub-sections, we will describe the details of generating the mock galaxy catalogs.
4.1 GLAM simulations

GLAM (Klypin & Prada 2018) is a new parallel version of the Particle-Mesh (PM) code that can quickly produce a large number of N-body simulations. We use the GLAM code to generate the matter density field for our mock catalogs since the computational speed is much faster than for QPM simulations. We used MareNostrum-4 computer at Barcelona Supercomputer Center to generate 2000 realizations in the adopted cosmology (see Sec. 1). The volume of each simulation is $3h^{-1}$ Gpc, which is large enough to cover the ELG redshift range $0.6 < z < 1.1$. We used 1500$^3$ particles with a mass per particle of $6.8 \times 10^{11} h^{-1} M_{\odot}$. The simulations were started at $z = 100$ and a constant time-step is used at low redshifts but periodically increases at high redshifts. The total number of time steps is 94, which is large enough to satisfy both accuracy and stability of the integration of the particle trajectories inside dense regions (Klypin & Prada 2018). Under this set of simulation parameters, the total number of CPU hours is 52,000. In order to make the process as efficient as possible, we incorporate QPM as a module in the GLAM code, so halo catalog creation is done on the fly. This procedure prevents extra time consumption due to I/O of large files, and saves disk storage.

4.2 From N-body simulation to halo catalogs

There are many benefits to first create halo catalogs and then generate galaxy mocks using galaxy-halo models. One of which is that we can model multiple target samples with different biases within the halo occupation framework. The other advantage is that we can study and test different galaxy-halo models for the same sample.

Here we use a modified version of the QPM method described in White et al. (2014) to generate halo catalogs from the GLAM simulations. First, we use Fourier methods on a mesh grid to estimate the density field of the GLAM simulations, which is then mapped to a halo mass. We calibrate the mapping scheme so as to match the bias of halos from high resolution simulations (Tinker et al. 2008). We group particles in the GLAM simulations by their density $\mu = \ln(1 + \delta)$ in 8 equally spaced bins, then calculate the bias of each group. The bias is then mapped to a halo mass using the halo bias function, $b(M_h)$, of Tinker et al. (2010). We then fit a smooth function to $b(M_h)$ of the particles. The result fitting function is

$$
\mu(M_h) = 0.5 + 0.1 \log_{10}(M_h/M_0) + \frac{(M_h/M_0)^{0.7}}{1 + (M_h/M_0)^{-0.35}},
$$

where $M_h$ is the halo mass and $M_0 = 10^{13.3} h^{-1} M_{\odot}$ is the transition scale from a logarithmic function to power law. The shape of the function is shown in Fig. 3 compared to the function adopted in White et al. (2014). Note that White et al. (2014) only use halos with $M_h > 10^{13} h^{-1} M_{\odot}$. Compared with White et al. (2014), our fitting function allows us to resolve low-mass halos down to $10^{11} h^{-1} M_{\odot}$. This is important for the ELG samples because ELGs are relatively young galaxies and are more likely to reside in smaller halos compared to luminous red galaxies (Gonzalez-Perez et al. 2018). Our fitting function differs from White et al. (2014) at high masses because they are at different redshifts. We aim to produce mocks at redshift $z = 0.84$ for eBOSS ELG sample, while White et al. (2014) uses redshift $z = 0.55$ for the BOSS LRG sample.

Secondly, we select a subset of particles based on the density to stand in for halos, and these halos are assigned the same positions and velocities as the particles. Particles are sampled using a Gaussian sampling function. A particle with density $\mu_0$ is assigned a halo mass $M_h$ with a probability of

$$
P(\mu_h | \mu_0, \sigma) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left( \frac{(\mu(M_h) - \mu_0)^2}{2\sigma^2} \right),
$$

where the function $\mu(M_h)$ is defined in Eq. 9, and we fix $\sigma = 0.1$. The width $\sigma$ doesn’t have a significant effect for the large-scale bias, as shown in White et al. (2014). Since the total number of particles in each simulation is finite, there are not enough particles at low halo masses. This imposes a mass resolution in our halo catalogs at a mass $M_h > 2 \times 10^{11} h^{-1} M_{\odot}$.

Finally, in order to have the correct halo mass function, we divided the halo masses into $N_b$ bins, where the number $N_b$ is large so that the change in $b(M_h)$ between each bin is small. For each mass bin we calculate the number of particles we need to sample to match the mass function $n(M_h)$ of Tinker et al. (2008), then we loop over all particles, assigning particles as halos with the corresponding probability.

In Fig. 4, we show the average halo mass function from 10 halo
The fitting function used in this work (blue) is approximately logarithmic approximation at low halo masses, transitioning to a power law at the high halo mass end. For comparison, the mapping used in White et al. (2014) is shown in orange. The scatter of the mapping is 0.1 dex, as described in Eq. 10. Only halos with $M_h > 10^{13} h^{-1} M_\odot$ are used in White et al. (2014). The functions differ at high masses since they are for halos at different redshifts.

The 10 catalogs are generated independently from different GLAM simulations. Compared to the QPM mocks used for BOSS LRG sample (Alam et al. 2017), our halo catalogs agree very well to the halo mass function in Tinker et al. (2008) at low halo mass. This verifies that our method yields the correct halo mass function. The small discrepancy at the high halo mass end is due to the power law formalism we choose for $\mu(M_h)$, which means it is less likely to have halos with mass larger than $10^{15} h^{-1} M_\odot$. Since the fraction of ELGs in halos with $M_h \gtrsim 10^{15} h^{-1} M_\odot$ is negligible, this discrepancy will not have a significant impact on our mocks.

In Fig. 5, we compare the halo bias calculated from the halo catalogs with the halo bias of Tinker et al. (2010). The result are in very good agreement for $M_h < 10^{13.5} h^{-1} M_\odot$. The small discrepancy at higher halo mass is due to lack of information of the bias in high density regions: there is insufficient number of particles that have high density to compute the bias. Considering that the mean halo mass of ELGs is around $10^{12} h^{-1} M_\odot$, this discrepancy should be negligible.

### 4.3 Halo occupation distribution for ELGs

We use the halo occupation distribution (HOD) to model galaxy bias (Peacock & Smith 2000; Seljak 2000; Benson et al. 2000; Scoccimarro et al. 2001; Zheng et al. 2009; White et al. 2011; Berlind & Weinberg 2002). The HOD formalism describes the relation between a typical class of galaxies and dark matter halos by the probability $P(N|M)$ that a halo with mass $M$ contains $N$ such galaxies. The population of galaxies can be split into central galaxies, which reside at the center of the halo, and satellite galaxies. Here we assume that

Figure 3. Fitting function describing the relationship between the dark matter overdensity and halo mass, where $\mu = \ln(1 + \delta)$. The fitting function used in this work (blue) is approximately logarithmic approximation at low halo masses, transitioning to a power law at the high halo mass end. For comparison, the mapping used in White et al. (2014) is shown in orange. The scatter of the mapping is 0.1 dex, as described in Eq. 10. Only halos with $M_h > 10^{13} h^{-1} M_\odot$ are used in White et al. (2014). The functions differ at high masses since they are for halos at different redshifts.

Figure 4. The halo mass function from the GLAM-QPM mocks generated in this work (blue points) compared to BOSS LRG mocks (orange points), which are created using the QPM method (White et al. 2014). For both, we show the average mass function of 10 mocks. The errorbars are the standard deviation among the 10 halo catalogs. As a comparison, we also plot the halo mass function from Tinker et al. (2008) at redshift $z = 0.55$ (red) and $z = 0.84$ (blue).

Figure 5. The average halo bias from from 10 GLAM-QPM halo catalogs (blue), where the error bars indicate the standard error on the mean. The orange curve indicates the fitting function of Tinker et al. (2010) at redshift $z = 0.84$. 

MNRAS 000, 1–12 (2020)
the satellite galaxies in each halo follow the same radial distribution as the dark matter, corresponding to a NFW profile (Navarro et al. 1997) where we use the concentration-mass relation of Macciò et al. (2007). The HOD model is a complete description of galaxy bias, i.e., given an HOD model and the halo population from a cosmological model, one can calculate any galaxy clustering statistic on any scale.

HOD modeling has been applied to interpret galaxy clustering in several surveys (Bullock et al. 2001; Moustakas & Somerville 2002; van den Bosch et al. 2003; Zheng 2004; Zehavi et al. 2005; Lee et al. 2006; Zheng & Weinberg 2007; Wake et al. 2008; Blake et al. 2008; Zheng et al. 2009; Parejko et al. 2013; Park et al. 2016; Zhai et al. 2017; Carter et al. 2018). Most of the studies use five parameter HOD models: the occupation function for central galaxies \( N_{\text{cen}} \) is a softened step function with three parameters and the occupation function for satellite galaxies \( N_{\text{sat}} \) is a power law with two free parameters. However, the step function for central galaxies may not model ELGs well, since one would expect that most of the galaxies in higher mass halos are quenched or have attenuated star formation. Therefore it is less likely that central galaxies with strong emission lines will be found in massive halos. Currently, the halo occupation distribution of ELGs are not well understood. Favole et al. (2017) studied the HOD of [OII] emitters in the local universe using the (Sub)Halo Abundance Matching (SHAM) method. Gonzalez-Perez et al. (2018) studied the properties of the host halos of [OII] emitters, and they found that the central galaxy occupation \( \langle N \rangle_{\text{OII cen}} \) can be formalized as the sum of a Gaussian and a step function with amplitude below unity. Zehavi et al. (2011) studied color dependence and they found that the central galaxy occupation \( \langle N \rangle_{\text{cen}} \) is the maximum probability that a central galaxy can be hosted in a halo, and \( \sigma_{\log M} \) is the standard deviation width in log mass and \( f_{\text{max}} \) is the maximum probability that a halo host a central galaxy. For satellite galaxies, \( \alpha \) is the power law slope, \( M_1 \) is the amplitude and \( M_{\text{cut}} \) is the cut-off mass. This formalism is simplified from previous studies, but it is sufficient to model the ELG clustering. The six free parameters in our HOD model are \( \{ f_{\text{max}}, M_{\text{min}}, \sigma_{\log M}, M_{\text{cut}}, M_1, \alpha \} \). The units of \( M_{\text{min}}, M_1 \) and \( M_{\text{cut}} \) are \( h^{-1}M_\odot \), while \( f_{\text{max}} \) and \( \sigma_{\log M} \) are dimensionless quantities. We perform a coarse HOD parameter space search for parameter optimization. A value for each parameter is selected for \( \{ M_{\text{min}}, M_{\text{cut}}, M_1, \alpha \} \) while fixing \( f_{\text{max}} = 0.15 \) and \( \sigma_{\log M} = 0.25 \). The parameter space is shown in Table 1. In subsequent sections we will demonstrate that the ELG clustering is relatively insensitive to the choice of parameters. Thus, we did not explore the whole HOD parameter space, which would be out of the scope of this paper. For the purpose of mock generation, our method is sufficient to model the large-scale bias of ELGs, with reasonable choices for the satellite fraction of the target sample.

There are 54 sets of parameter in total. For each set, we generate a galaxy mock and measure the redshift-space monopole \( \xi_0(r) \), quadrupole \( \xi_2(r) \), as well as projected correlation function \( w_p(r_p) \). We calibrate the HOD by finding the set of parameters which most closely match the clustering of the data at scales between 10 \( h^{-1}\text{Mpc} \) and 30 \( h^{-1}\text{Mpc} \), as one would interpret the linear bias at this scale. At smaller scales, i.e. \( s < 1h^{-1}\text{Mpc} \), the QPM scheme does not model the bias well. We will illustrate this point presently. At larger scales, the uncertainty in the clustering measurements from the data is large due to sample variance, and the clustering measurements are more prone to residual photometric systematics. The best fit HOD parameters that we use to generate ELG mocks are listed in Table 2. We tested the effect of varying \( f_{\text{max}} \) and \( \sigma_{\log M} \), and found that this does not significantly change the clustering measurements.

The occupation function of central galaxies and satellite galaxies is shown in the upper left panel of Fig. 6. The satellite fraction is 17.4\%, in accordance with \( \sim 20\% \) satellite fraction of star forming galaxies in Tinker et al. (2013). This is also in agreement with Favole et al. (2016), where it was found that 22.5 \( \pm 2.5\% \) of ELGs at redshift 0.8 are satellite galaxies. The galaxy number density produced by our HOD is \( 6.7 \times 10^{-4}h^3\text{Mpc}^{-3} \), which is slightly higher than the peak ELG number density \( 6.4 \times 10^{-4}h^3\text{Mpc}^{-3} \). This is intended to be so, as the higher number density allows studies of systematic corrections. This HOD implies that ELGs reside in halos with mass larger than \( 10^{11}h^{-1}M_\odot \), and most of the central galaxies in halos with mass larger than \( 10^{13}h^{-1}M_\odot \) are quenched, which is in agreement with the HOD analysis at redshift \( z \sim 0.85 \) of Tinker et al. (2013). We also compare our result with the HOD study of eBOSS ELGs in Gonzalez-Perez et al. (2018) and Guo et al. (2019) at redshift 0.8 < \( z < 0.9 \). The amplitude of our HOD function is higher than the results of Gonzalez-Perez et al. (2018) and Guo et al. (2019) because we aim to match the peak number density of eBOSS ELG targets for the purpose of producing mock catalogs. In addition their results are for ELGs in the redshift bin 0.8 < \( z < 0.9 \), while we cover a wider redshift range. Fig. 6 also compares the clustering measurements between mocks and data. The clustering is measured from 100 mocks generated with the HOD parameters in Table 2. The errorbars represent the 1\( \sigma \) scatter between mocks. The mocks agree with data at scales from a few \( h^{-1}\text{Mpc} \) to around 20 \( h^{-1}\text{Mpc} \). At scales around \( s \sim 1h^{-1}\text{Mpc} \), the mocks are less clustered than the data. The reason is that the method of sampling the density field
Figure 6. HOD model and galaxy clustering comparison between the GLAM-QPM mocks and ELG data. The upper left panel shows the best fit HOD model for the eBOSS ELG sample (black), where the dashed curve indicates the central HOD, dotted curve the satellite HOD and the solid curve is the total HOD. The satellite fraction of our HOD model is 17.4%. As a comparison, we also show the HOD prediction for eBOSS ELG in the redshift range $0.8 < z < 0.9$ from Gonzalez-Perez et al. (2018) (blue) and Guo et al. (2019) (red). The remaining three panels show the measurements of $w_p(r_p)$, $\xi_0(s)$ and $\xi_2(s)$, respectively. The red curves are measurements from ELG data, averaged between the SGC and NGC, weighted by area. The errorbars are estimated using the jackknife resampling technique, with 25 jackknife samples. The green curves are the mean of the clustering of 100 GLAM-QPM mocks using the best fit HOD parameters in Table 2. The errorbars are the standard deviation, indicating the 1σ scatter.

To find halos does not yield the correct number of halos with small separations, thus leading to a deficit in the correlation function at the transition scale between one-halo and two-halo galaxy pairs. The lower halo mass region of the ELG sample accentuates this effect relative to earlier mocks with luminous red galaxies. Due to this fact, we don’t perform a $\chi^2$ test on different HOD parameters because we don’t want the clustering measurements at small scale ($s \sim 1h^{-1}$Mpc) have undue influence on the selection of HOD parameters.

Instead, we test the impact of HOD parameters to the small-scale clustering of ELGs by changing one parameter around the best-fit parameter set. As shown in Fig. 7, we fix $f_{\text{max}} = 0.15$, $\sigma_{\log M} = 0.25$ and perturb other parameters around our fiducial parameter set in Table 2. The red curve shows the fiducial HOD parameter set, the blue and green curves show the clustering after the perturbation. $M_{\text{min}}$ and $M_1$ affect the clustering of ELGs the most. This is because $M_{\text{min}}$ is the mean halo mass of the Gaussian function for central galaxies, so this parameter determines the halo mass scale that central galaxies reside in, and determines the ELG linear bias. $M_1$ affects the mass of halos that satellite galaxies reside in, so a smaller $M_1$ would result in satellite ELGs being placed in higher mass haloes, increasing the linear bias. There is no significant change in the clustering even if we increase $M_{\text{min}}$ and $M_1$ by a factor of 2. The impact of $M_{\text{cut}}$ and $\alpha$ is even less. We conclude that by choosing suitable HOD parameters, we can produce mock catalogs with small-scale clustering that is in reasonable agreement with the data, and the large-scale bias is relatively insensitive to the parameters chosen.

5 LARGE SCALE CLUSTERING AND COVARIANCE MATRIX

5.1 Large Scale Clustering

We generate 2000 galaxy mocks with the HOD parameters presented in Table 2, and cut the mock to the ELG chunk geometry as well as redshift distribution $n(z)$. We measure the redshift-space monopole and quadrupole up to $200h^{-1}$Mpc using the same method and systematic weights as described in Sec. 3. We choose 40 equally spaced bins for $s$ from $0h^{-1}$Mpc to $200h^{-1}$Mpc.
Figure 7. The impact of varying the fiducial HOD parameters on the redshift-space monopoles and quadrupoles. Redshift-space monopoles are vertically shifted by 50 to conveniently visualize. Our results for the fiducial set of HOD parameters (see Table 2) are shown with red curves, and each panel shows the result of perturbing one parameter at a time. The upper left panel shows the impact of varying $M_{\text{min}}$, and the upper right panel shows the impact of varying $M_1$. The lower panels show the impact of varying $M_{\text{cut}}$ and $\alpha$ on the ELG clustering. The measured clustering of the eBOSS ELGs is indicated by the black points.

The large-scale clustering of the mocks is shown in Fig. 8, in comparison with the data. The shaded area indicates the 1σ and 2σ scatter in the 2000 mocks. The BAO feature can be clearly seen in the mocks, but there is some visible discrepancy between the mocks and data at the BAO scale, and also between the SGC and NGC. We first present the covariance matrix in Sec. 5.2, then test the statistical significance of the differences between SGC and NGC in Sec. 5.3.
Given a set of mock catalogs, the covariance matrix is defined as:

$$C_{ij} = \text{corr}(x_i, x_j) = \frac{\sum_{k=1}^{n_s} (x_i^k - \mu_i)(x_j^k - \mu_j)}{n_s - 1},$$

where $x_i$ is the clustering measurement at the $i$th bin; index $k$ indicates $k$th realization of mocks; $n_s$ is the total number of mocks; $i, j$ are bins of separation; $\mu_i$ is the mean of $x_i^k$.

The correlation matrix $C_{ij}$ is given by

$$C_{ij} = \frac{\sum_{k=1}^{n_s} (x_i^k - \mu_i)(x_j^k - \mu_j)}{\sqrt{\sum_{k=1}^{n_s} (x_i^k - \mu_i)^2 \sum_{k=1}^{n_s} (x_j^k - \mu_j)^2}}.$$

The correlation matrix of the large-scale clustering measurements of the GLAM-QPM mocks is shown in Fig. 9. These correlation matrices are all reasonable: clustering measurements between two bins with distance separated within $\pm 20h^{-1}\text{Mpc}$ are correlated.

In Fig. 10 we show the correlation matrix for the ELG redshift space monopole measurements at small scales. The result is as expected, with a correlation between galaxy pairs on scales $s \gtrsim 2.5h^{-1}\text{Mpc}$, where the pairs are from two distinct halos. Monopole measurements are uncorrelated on scales $s \lesssim 2.5h^{-1}\text{Mpc}$, where galaxy pairs are within the same halo and galaxies are randomly sampled from the halo density profile.

### 5.3 NGC and SGC difference

We use the GLAM-QPM mocks produced in this paper to test the consistency of the NGC and SGC measurements. For the redshift-space monopole at small scales, we measure the cross-

$$\chi^2 = (\hat{s}_{\text{NGC}} - \hat{s}_{\text{SGC}})^T C_{\text{tot}}^{-1} (\hat{s}_{\text{NGC}} - \hat{s}_{\text{SGC}})$$

with 12 data points between 0.34$h^{-1}\text{Mpc}$ and 70$h^{-1}\text{Mpc}$, where the covariance matrix $C_{\text{tot}} = C_{\text{NGC}} + C_{\text{SGC}}$ is estimated using the GLAM-QPM mocks. We measure $\chi^2 / \text{dof} = 22.3 / 12$ and conclude that on small scales, the SGC and NGC clustering measurements are compatible.

In order to see whether the difference between SGC and NGC at large scales is significant, we compute the cross-$\chi^2$ for both the monopole and the quadrupole measurements. We choose 10 linear $s$ bins from 77.5$h^{-1}\text{Mpc}$ to 122.5$h^{-1}\text{Mpc}$, since we’re most interested in the BAO scale at $s \sim 100h^{-1}\text{Mpc}$. The results are $\chi^2 / \text{dof} = 12.53 / 10$ for monopoles and $\chi^2 / \text{dof} = 14.91 / 10$ for quadrupoles. The corresponding p-value is 0.25 and 0.14, meaning that the difference is insignificant and we cannot reject the null hypothesis that the difference between the SGC and NGC is caused by cosmic variance. We also build the $\chi^2$ distribution from our mock sample by selecting 200 NGC mocks and 200 SGC mocks (all from different GLAM simulations) and building the sample distribution from 40,000 $\chi^2$ values from each SGC-NGC pair. Our sample distribution agrees with the $\chi^2$ distribution with $\text{dof} = 10$ perfectly well, indicating that our covariance matrix are valid for robust BAO and RSD analysis, as is done in de Mattia (2020).

### 6 SUMMARY

We present 2000 GLAM-QPM mock catalogs for the eBOSS DR16 ELG sample. We use GLAM simulations to quickly and accurately produce the dark matter density field and the QPM method to assign dark matter halos to particles in the simulation. The haloes are then populated with ELGs using a HOD methodology. We have calibrated the HOD parameters for the eBOSS ELG sample to model the large-scale bias of ELGs. The majority of central galaxies falls in halos with mass between $10^{11}h^{-1}\text{M}_\odot$ and $10^{13}h^{-1}\text{M}_\odot$, and the satellite fraction of our HOD model is 17.4%. The eBOSS ELG survey geometry and radial selection functions are applied to our mocks. This set of mock catalogs is used in the eBOSS ELG RSD analysis in de Mattia (2020).

We’ve shown that the GLAM-QPM mock catalogs agree with ELG data at large scales, in general, within 2$\sigma$ for monopole. For
Figure 9. Correlation matrix of the GLAM-QPM mocks. The upper panel shows the correlation matrix of the monopole, and the lower panel shows correlation matrix of the quadrupole. We use 40 equally spaced bins in $s$ from 0$h^{-1}$Mpc to 200$h^{-1}$Mpc.

quadrupole the ELG SGC data shows higher clustering signal than GLAM-QPM mocks. We examined the cross-$\chi^2$ value for SGC and NGC around BAO scales, and find $\chi^2/dof = 10.09/10$ for the monopole and $\chi^2/dof = 14.86/10$ for the quadrupole. We cannot conclude that the difference between SGC and NGC are due to reasons other than cosmic variance.

The GLAM-QPM mock galaxy catalogs will be published along with eBOSS DR16 galaxy catalog release, and will also be placed at the Skies and Universes site (Klypin et al. 2017) \footnote{www.skiesanduniverses.org}.

**ACKNOWLEDGMENTS**

SL is grateful to the support from CCPP, New York University. J.L.T. and M.R.B. are supported by NSF Award 1615997. F.P. and A.K. acknowledges support from the Spanish MICINN grant GC2018-101931-B-100. G.R. acknowledges support from the National Research Foundation of Korea (NRF) through Grants No. 2017R1E1A1A01077508 and No. 2020R1A2C1005655 funded by the Korean Ministry of Education, Science and Technology (MoEST), and from the faculty research fund of Sejong University. The GLAM simulations used in this paper were done on MareNostrum-4 at the Barcelona Supercomputer Center in Spain.

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS website is www.sdss.org. SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, the Chilean Participation Group, the French Participation Group, Harvard-Smithsonian Center for Astrophysics, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU)/University of Tokyo, Lawrence Berkeley National
Figure 11. cross-$\chi^2$ value from the SGC and NGC clustering measurements (vertical line), calculated using the covariance matrix from the GLAM-QPM mocks, given by $\chi^2 = (\xi_{SGC} - \xi_{NGC})^T C^{-1}_{\xi \xi}(\xi_{SGC} - \xi_{NGC})$. The left panel shows the $\chi^2$ for the monopole and the right panel is for the quadrupole. We choose 10 $r$ bins from $77.5 \ h^{-1}\text{Mpc}$ to $122.5 \ h^{-1}\text{Mpc}$ to study the clustering signal around BAO peak ($r \sim 100 \ h^{-1}\text{Mpc}$). We also show the $\chi^2$ distribution from 40,000 pairs of NGC and SGC mocks (blue histogram), in order to see where the data falls in the distribution. A $\chi^2$ distribution with $\text{dof}=10$ is shown by the green curve.

Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatory of China, New Mexico State University, New York University, University of Notre Dame, Observatório Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

REFERENCES

Alam S., et al., 2017, MNRAS, 470, 2617
Alam S., et al., 2020, submitted
Anderson L., et al., 2014, MNRAS, 441, 24
Avila S., et al., 2020, submitted
Bautista J. E., et al., 2020, submitted
Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000, MNRAS, 311, 793
Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
Beutler F., et al., 2012, MNRAS, 423, 3430
Blake C., Collister A., Lahav O., 2008, MNRAS, 385, 1257
Blake C., et al., 2011, MNRAS, 418, 1707
Blanton M. R., et al., 2017, AJ, 154, 28
Bullock J. S., Wechsler R. H., Somerville R. S., 2002, MNRAS, 329, 246
Carter P., Beutler F., Percival W. J., Blake C., Koda J., Ross A. J., 2018, MNRAS, 481, 2371
Chuang C.-H., Kitaura F.-S., Prada F., Zhao C., Yepes G., 2015, MNRAS, 446, 2621
Cole S., et al., 2005, MNRAS, 362, 505
Collaboration e., et al., 2020, submitted
Comparat J., et al., 2013, MNRAS, 428, 1498
Comparat J., et al., 2015, A&A, 575, A40
Comparat J., et al., 2016, A&A, 592, A121
Davis M., Peebles P. J. E., 1983, ApJ, 267, 465
Dawson K. S., et al., 2013, AJ, 145, 10
Dawson K. S., et al., 2016, AJ, 151, 44
Dey A., et al., 2018, arXiv e-prints,
Dodelson S., Schneider M. D., 2013, Phys. Rev. D, 88, 063535
Eisenstein D. J., Hu W., 1998, ApJ, 496, 605
Eisenstein D. J., et al., 2005, ApJ, 633, 560
favole G., et al., 2016, MNRAS, 461, 3421
Favole G., Rodríguez-Torres S. A., Comparat J., Prada F., Guo H., Klypin A., Montero-Dorta A. D., 2017, MNRAS, 472, 550
Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23
Gil-Marín H., et al., 2020, submitted
Gonzalez-Perez V., et al., 2018, MNRAS, 474, 4024
Gunn J. E., et al., 2006, AJ, 131, 2332
Guo H., et al., 2019, ApJ, 871, 147
Hawkins E., et al., 2003, MNRAS, 346, 78
Hou J., et al., 2020, submitted
Kaiser N., 1987, MNRAS, 227, 1
Kitaura F.-S., Yepes G., Prada F., 2014, MNRAS, 439, L21
Klypin A., Prada F., 2018, MNRAS, 478, 4602
Klypin A., Prada, F., Comparat J., 2017, arXiv e-prints, p. arXiv:1711.01453
Landy S. D., Szalay A. S., 1993, ApJ, 412, 64
Lee K.-S., Giavalisco M., Gnedin O. Y., Somerville R. S., Ferguson H. C., Dickinson M., Ouchi M., 2006, ApJ, 642, 63
Lyke B. W., et al., 2020, submitted
Macció A. V., Dutton A. A., van den Bosch F. C., Moore B., Potter D., Stadel J., 2007, MNRAS, 378, 55
Moustakas L. A., Somerville R. S., 2002, ApJ, 577, 1
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Neveux R., et al., 2020, submitted
Parejko J. K., et al., 2013, MNRAS, 429, 98
Park Y., et al., 2016, Phys. Rev. D, 94, 063533
Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144
Lin et al.

Peebles P. J. E., 1980, The large-scale structure of the universe
Percival W. J., et al., 2014, MNRAS, 439, 2531
Raichoor A., 2020, submitted
Raichoor A., et al., 2017, MNRAS, 471, 3955
Ross A. J., et al., 2020, submitted
Rossi G., et al., 2020, submitted
Scoccimarro R., Sheth R. K., Hui L., Jain B., 2001, ApJ, 546, 20
Seljak U., 2000, MNRAS, 318, 203
Snee S. A., et al., 2013, AJ, 146, 32
Smith A., et al., 2020, submitted
Tamone A., et al., 2020, submitted
Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, ApJ, 688, 709
Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, ApJ, 724, 878
Tinker J. L., Leauthaud A., Bundy K., George M. R., Behroozi P., Massey R., Rhodes J., Wechsler R. H., 2013, ApJ, 778, 93
Wake D. A., et al., 2008, MNRAS, 387, 1045
White M., et al., 2011, ApJ, 728, 126
White M., Tinker J. L., McBride C. K., 2014, MNRAS, 437, 2594
Zehavi I., et al., 2005, ApJ, 630, 1
Zehavi I., et al., 2011, ApJ, 736, 59
Zhai Z., et al., 2017, ApJ, 848, 76
Zhao C., et al., 2020, submitted
Zheng Z., 2004, ApJ, 610, 61
Zheng Z., Weinberg D. H., 2007, ApJ, 659, 1
Zheng Z., Zehavi I., Eisenstein D. J., Weinberg D. H., Jing Y. P., 2009, ApJ, 707, 554
de Mattia A., 2020, submitted
du Mas des Bourboux H., et al., 2020, submitted
van den Bosch F. C., Yang X., Mo H. J., 2003, MNRAS, 340, 771

This paper has been typeset from a TeX file prepared by the author.