The Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: N-body Mock Challenge for Galaxy Clustering Measurements

Graziano Rossi, Peter D. Choi, Jeongin Moon, Julian E. Bautista, Hector Gil-Marín, Romain Paviot, Mariana Vargas-Magaña, Sylvain de la Torre, Sebastien Fromenteau, Ashley J. Ross, Santiago Ávila, Etienne Burtin, Kyle S. Dawson, Salman Habib, Katrin Heitmann, Jiamin Hou, Eva-Maria Mueller, Will J. Percival, Alex Smith, Cheng Zhao, Gong-Bo Zhao, Stéphanie Escoffier

1 Department of Physics and Astronomy, Sejong University, Seoul, 143-747, Korea
2 Institute of Cosmology & Gravitation, Dennis Sciama Building, University of Portsmouth, Portsmouth, PO1 3FX, UK
3 Institut de Ciències del Cosmos, Universitat de Barcelona, ICCUB, Martí i Franquès 1, E08028 Barcelona, Spain
4 Institut d’Estudis Espacials de Catalunya (IEEC), E08034 Barcelona, Spain
5 Aix Marseille Univ, CNRS, CNES, LAM, Marseille, France
6 Instituto de Física, Universidad Nacional Autónoma de México, Apdo. Postal 20-364, Ciudad de México, México
7 Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México, Av. Universidad s/n, 62210 Cuernavaca, Morelos, Mexico
8 Center for Cosmology and Astro-Particle Physics, Ohio State University, Columbus, OH 43210
9 Universidad Autónoma de Madrid, 28049, Madrid, Spain
10 Instituto de Física, Universidad Autónoma de Madrid, 28049 Madrid, Spain
11 CEA, Centre de Saclay, Irfu/SPP, F-91191 Gif-sur-Yvette, France
12 Department of Physics and Astronomy, University of Utah, Salt Lake City, UT 84112, USA
13 High Energy Physics Division, Argonne National Laboratory, Lemont, IL 60439, USA
14 Computational Science Division, Argonne National Laboratory, Lemont, IL 60439, USA
15 Max-Planck-Institut für Extraterrestrische Physik, Postfach 1312, Giessenbachstrasse 1, 85748 Garching bei München, Germany
16 Sub-department of Astrophysics, Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH
17 Waterloo Centre for Astrophysics, University of Waterloo, Waterloo, ON N2L 3G1, Canada
18 Department of Physics and Astronomy, University of Waterloo, Waterloo, ON N2L 3G1, Canada
19 Perimeter Institute for Theoretical Physics, 31 Caroline St. North, Waterloo, ON N2L 2Y5, Canada
20 Laboratory of Astrophysics, École Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland
21 National Astronomy Observatories, Chinese Academy of Science, Beijing, 100012, P. R. China
22 College of Astronomy and Space Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

20 July 2020
ABSTRACT

We develop a series of $N$-body data challenges, functional to the final analysis of the extended Baryon Oscillation Spectroscopic Survey (eBOSS) Data Release 16 (DR16) galaxy sample. The challenges are primarily based on high-fidelity catalogs constructed from the Outer Rim simulation – a large box size realization ($3h^{-1}$Gpc) characterized by an unprecedented mass resolution, down to $1.85 \cdot 10^{9} h^{-1} M_{\odot}$. We generate synthetic galaxy mocks by populating Outer Rim halos with a variety of halo occupation distribution (HOD) schemes of increasing complexity, spanning different redshift intervals. We then assess the performance of three complementary redshift space distortion (RSD) models in configuration and Fourier space, adopted for the analysis of the complete DR16 eBOSS sample of Luminous Red Galaxies (LRGs). We find that all the methods are mutually consistent, with comparable systematic errors on the Alcock-Paczynski parameters and the growth of structure, and robust to different HOD prescriptions – thus validating the robustness of the models and the pipelines for the baryon acoustic oscillation (BAO) and full shape clustering analysis. In particular, all the techniques are able to recover $\sigma_8$ and $\alpha_\perp$ to within $0.9\%$, and $f\sigma_8$ to within $1.5\%$. As a by-product of our work, we are also able to gain interesting insights on the galaxy-halo connection. Our study is relevant for the final eBOSS DR16 ‘consensus cosmology’, as the systematic error budget is informed by testing the results of analyses against these high-resolution mocks. In addition, it is also useful for future large-volume surveys, since similar mock-making techniques and systematic corrections can be readily extended to model for instance the DESI galaxy sample.

Key words: methods: analytical, statistical, numerical – galaxies: formation, clustering — cosmology: theory, large-scale structure of Universe

1 INTRODUCTION

The Sloan Digital Sky Survey (SDSS; York et al. 2000), currently in its fourth generation (SDSS-IV; see Blanton et al. 2017 for a review), has established a remarkable legacy in astronomy and set new standards for precision cosmology. A key component of the SDSS-IV, the Extended Baryon Oscillation Spectroscopic Survey (eBOSS; Dawson et al. 2016) is now releasing the final cosmological catalogs (Lyke et al. 2020; Raichoor et al. 2020; Ross et al. 2020) with the Data Release 16 (DR16), summarizing the efforts of more than 10 years of operations. eBOSS spectroscopically targets four distinct astrophysical populations: luminous red galaxies (LRGs, the primary focus of this work), emission line galaxies (ELGs), clustering quasars (QSOs), and the Lyman-α (Lyα) forest of quasars at high redshift. In a novel and yet uncharted redshift interval, eBOSS has built the most complete, unprecedented large volume map of the universe usable for large-scale structure (LSS) to date.

Exquisite high-quality data from the SDSS have been pivotal in firmly establishing the standard minimal six-parameter concordance cosmological scenario dominated by cold dark matter (CDM) and a dark energy (DE) component in the form of a cosmological constant $\Lambda$, known as the $\Lambda$CDM model. Traditionally, this has been achieved by using the baryon acoustic oscillation (BAO) feature as measured in galaxy and quasar clustering, to estimate the angular distance $D_M$ and the Hubble parameter $H$, as well as their product from the Alcock-Paczynski effect (AP; Alcock & Paczynski 1979), and the growth of structure quantified by $f\sigma_8(z)$ from redshift-space distortions (RSD) – with $f(z)$ the logarithmic growth rate of the linear fluctuation amplitude with respect to the expansion factor, and $\sigma_8(z)$ the normalization of the linear theory matter power spectrum at redshift $z$ via the rms fluctuation in $8 h^{-1}$Mpc spheres. Since the very first BAO detections (Colless et al. 2003; Cole et al. 2005; Eisenstein et al. 2005), measurements of the BAO peak have been sharpening and expanding in redshift range, allowing for multiple accurate cosmological constraints and solid confirmations of the $\Lambda$CDM framework. Noticeably, the eBOSS team has recently presented the first measurement of the BAO signal in a novel uncharted redshift range ($0.8 < z < 2.2$) using the clustering properties of 147,000 new quasars (Ata et al. 2018), and reported a BAO detection with a significance $> 2.8\sigma$ along with detailed high-$z$ distance measurements (within $3.8\%$), a remarkable result that confirms and extends the validity of the standard $\Lambda$CDM cosmological model to an unprecedented large-volume.

To this end, multiple techniques involving RSD methods and clustering estimators along with BAO reconstructions in configuration or Fourier space are generally adopted for the analysis of the various LSS tracers, to extract cosmological information. The most up-to-date SDSS results involving LRGs can be found in Beutler et al. (2017a,b), Gil-Marín et al. (2017, 2018), Bautista et al. (2018), Mueller et al. (2018), Vargas-Magaña et al. (2018), Wang et al. (2018b), Icaza-Lizazola et al. (2019), and Zheng et al. (2019). Regarding ELGs, one of the novelties in eBOSS, recent studies have been carried out by Comparat et al. (2016), Raichoor et al. (2017), Guo et al. (2019). For the QSO population, see e.g. Gil-Marín et al. (2018), Hou et al. (2018), Wang et al. (2018a), Zarrrouk et al. (2018), Zhu et al. (2018), Zhao et al. (2019); and for Lyα QSOs see Blomqvist et al. (2019) and de Sainte Agathe et al. (2019).
Traditionally, all the main results from different SDSS tracers are eventually combined in a ‘consensus’ publication (e.g. Aubourg et al. 2015; Alam et al. 2017; Ata et al. 2018), and confronted with measurements from other state-of-the-art surveys – such as Planck (2018). This consensus is then of utmost importance, as it represents a legacy for the entire science community. We are now releasing the final eBOSS DR16 consensus analysis that summarizes the full impact of the SDSS spectroscopic surveys on the cosmological model (Collaboration et al. 2020), which encapsulates all the supporting clustering measurements presented in Bautista et al. (2020) and Gil-Marín et al. (2020) for LRGs, Hou et al. (2020) and Neveux et al. (2020) for QSOs, de Mattia et al. (2020) and Tamone et al. (2020) for ELGs, as well as du Mas des Bourboux et al. (2020) for the Lyα forest.

In this respect, quantifying the systematic error budget in RSD methods and BAO clustering estimators for all the eBOSS tracers as well as characterizing the robustness of the analysis pipelines are essential tasks, in order to obtain unbiased cosmological parameters, accurate $f \sigma_8$ constraints, and reliable consensus likelihoods. This is indeed the central aim of our work: here we focus on galaxies, and assess the performance and robustness of the BAO fitting methods and of three complementary RSD full shape (FS) models in configuration and redshift space, adopted in Bautista et al. (2020) and Gil-Marín et al. (2020) for the analysis of the complete DR16 eBOSS LRG sample – briefly described in Section 5. See also Smith et al. (2020) for an analogous effort on the QSO sample, and Alam et al. (2020), Avila et al. (2020), and Lin et al. (2020) for ELGs.

With this primary goal in mind, we devise a targeted galaxy mock challenge. In embryonic form, a similar mini-challenge was already present in the consensus eBOSS Data Release 12 (DR12) LRG analysis (Alam et al. 2017, see their Section 7). Here we expand on that, and carry out a more systematic investigation. Specifically, in our challenge (detailed in Section 6) we test the performance of BAO/RSD LRG fitting techniques against different galaxy population schemes and bias models having analogous clustering properties, with the main objective of validation and calibration of such methods and the quantification of theoretical systematics.

Assessing the robustness and accuracy of RSD models is only possible via high-fidelity ($N$-body-based) synthetic realizations. In this work, we construct new heterogeneous sets of galaxy mocks from the best simulation currently available, the Outer Rim (Heitmann et al. 2019, see Section 4) – a large box size run ($3h^{-1}$Gpc) characterized by an unprecedented mass resolution, down to $1.85 \cdot 10^8 h^{-1} M_\odot$. We base our methodology on Halo Occupation Distribution (HOD) techniques, in an increasing level of complexity (as thoroughly explained in Section 3): in particular, moving from the most conventional HOD framework, we explore more sophisticated scenarios able to distinguish between quiescent and star-forming galaxies and with the inclusion of assembly bias, that generalize further the standard HOD framework. We also exploit a small homogeneous set of cut-sky mocks (the $N$-series) – which has been previously used in the SDSS DR12 galaxy clustering analysis (Alam et al. 2017) – to address the impact of cosmic variance and related theoretical systematics, and make use of a new series of DR16 EZmocks (Zhao et al. 2020) for determining the rescaled covariance matrices functional to all the analyses. The mock-making procedure is explained in detail in Section 4.

By confronting the different BAO and RSD LRG fitting techniques on a common ground against a subset of those high-fidelity mocks having different HOD prescriptions, we are thus able to assess their performance, quantify the systematic errors on the Alcock-Paczynski parameters and the growth of structure, and eventually confirm the effectiveness of the LRG analysis pipelines. In particular, we anticipate that we find all the methods mutually consistent, and robust to different HOD prescriptions, validating the models used for the LRG clustering analysis.

Furthermore, the mock challenge developed here is suitable to a number of applications. Beside being directly useful for the final eBOSS DR16 ‘consensus cosmology’ (Collaboration et al. 2020), as the systematic error budget for the ultimate $f \sigma_8$ constraint are informed by testing the results of analyses against these high-resolution mocks, our work may be relevant for future large-volume surveys. For example, similar mock-making techniques and systematic corrections can be readily extended to model the Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration et al. 2016) and the Large Synoptic Survey Telescope (LSST; Ivezić et al. 2019) galaxy samples.

The layout of the paper is organized as follows. Section 2 briefly describes the eBOSS DR16 data release, and the final LRG sample. Section 3 provides the theoretical foundation for modeling the galaxy-halo connection, and explains the different HOD schemes adopted in the mock-making procedure. Section 4 describes the tools and methodology used to construct high-fidelity mocks. Section 5briefly presents the different RSD models, that are described in depth in companion papers. Section 6 shows selected results from the mock challenge, and compares the various LRG BAO and RSD models in configuration and Fourier space. Section 7 presents the global error budget for the completed eBOSS DR16 LRG sample, with a primary focus on theoretical systematics. Finally, we conclude in Section 8, where we summarize the main findings and indicate future avenues.

Throughout the paper and if not specified otherwise, all numerical values of length and mass are understood to be in $h = 1$ units.

2 SDSS-IV EBOSS AND DR16 LRG SAMPLE

2.1 SDSS Legacy and eBOSS

SDSS observations, carried out on the 2.5-meter Sloan Foundation telescope at Apache Point Observatory (Gunn et al. 2006), first begun in July 2014 (SDSS-I and SDSS-II; York et al. 2000). Since then, thanks to the remarkable efforts of more than 10 years of operations, the survey has evolved till its current fourth generation (SDSS-IV), collecting an increasing number of high-quality data for high-precision cosmology – outperforming on the targets that drove the initial survey design. eBOSS, a key component of the SDSS-IV and ranked in the highest tier in the 2018 DOE-HEP Portfolio Review, is a continuation of the Baryon Oscillation Spectroscopic Survey (BOSS) – part of the SDSS-III (Eisenstein et al. 2011) – and a pre-cursor for DESI (DESI Collaboration et al. 2016). eBOSS lies at the leading edge of cosmological experimentation: by spectroscopically targeting four
distinct astrophysical populations in a unique redshift interval, eBOSS has built the largest volume and most complete map of the Universe to date of any redshift survey. The primary innovation in eBOSS is extending BAO measurements with ELGs and a much larger number of quasars, enabling a percent-level measurement in the critical epoch of transition from deceleration to acceleration (i.e., $0.8 < z < 2.2$). This is why the eBOSS data set allows exploration of DE in epochs where no precision cosmological measurements currently exist (improving the DE Figure of Merit by a factor of 3), by addressing three Particle Physics Project Prioritization Panel (P5) science drivers and pursuing four key goals: BAO measurements of the Hubble parameter and distance as a function of $z$, RSD measurements of the gravitational growth of structure, constraints on the neutrino mass sum, and constraints on inflation through measurements of primordial non-Gaussianity. In particular, the exquisite BAO and RSD measurements that eBOSS provide (see e.g. Collaboration et al. 2020) are key for DE and gravity studies. Moreover, eBOSS has the spectroscopic capabilities to complement and enhance other current and future cosmological probes, representing a strategic asset and a pathfinder for upcoming experiments.

2.2 The eBOSS DR16 LRG Sample

The LRG spectroscopic sample allowed the first SDSS detection of the BAO peak in the galaxy large-scale correlation function (Eisenstein et al. 2005). Since its original version, comprised by 46,748 LRGs over 3816 deg$^2$ at $0.16 < z < 0.47$, the SDSS LRG catalog has considerably grown both in size and redshift depth, thanks to over a decade of observations. In particular, BOSS was designed to measure BAOs with LRGs over the redshift range $0.2 < z < 0.75$, while eBOSS increases the redshift coverage up to $z = 1$. With the final eBOSS DR16, completed on March 1, 2019, the LRG eBOSS-only released sample contains 174,816 galaxies with good redshifts in the interval range $0.6 < z < 1.0$, with an effective redshift $\bar{z} = 0.698$, spanning a total area of 4104 deg$^2$ and an effective volume of $1.241 \times 10^6$ Gpc$^3$. LRG targets were selected using optical and infrared imaging over 7500 deg$^2$ angular footprint, using photometry with updated calibration (Dawson et al. 2016): full details of LRG selections are provided in Prakash et al. (2016). To this end, Bautista et al. (2018) recently demonstrated that the sample is well-suited for LSS studies. The final DR16 eBOSS-only LRG sample is combined with the high redshift tail of the BOSS galaxy sample (CMASS), in order to provide one catalog of luminous galaxies with $z > 0.6$. Overall, BOSS CMASS galaxies make up slightly more than half of the total sample, and the area they occupy is more than twice that of eBOSS LRGs – over an effective volume of $1.445 \times 10^6$ Gpc$^3$, hence the total effective volume of the combined DR16 LRG sample is $2.654 \times 10^6$ Gpc$^3$. Most of the BOSS CMASS footprint was re-observed by the eBOSS LRG program, which covered 37% and 65% of the original Northern Galactic Cup (NGC) and Southern Galactic Cup (SGC) CMASS areas, respectively. The projected number density of galaxies with $0.6 < z < 1.0$ is more than twice as high for the eBOSS LRGs ($44 \, \text{deg}^{-2}$ compared to $21 \, \text{deg}^{-2}$).

Regarding redshift assignment, a different philosophy both for redshift estimates and spectral classification has been designed specifically for the eBOSS clustering catalogs, motivated by new challenges due to low signal-to-noise eBOSS galaxy spectra. In fact, previous routines used for BOSS were not optimized for the fainter and higher redshift LRG galaxies that comprise the eBOSS LRG sample, and therefore a new approach and software development to provide accurate redshift estimation (indicated as REDROCK$^1$) was necessary. As a result, the redshift completeness approaches 98% for the eBOSS LRG sample with a rate of ‘catastrophic failures’ estimated to be less than 1% – hence such redshift failures are not a concern for the LRG sample.

About sector completeness (a sector being an area covered by a unique set of plates), for the eBOSS LRG sample the 100% completeness requirement was relaxed, to increase the fiber efficiency and total survey area. To this end, the completeness of the eBOSS LRG sample exceeds 95% in every relevant chunk (i.e. an area tiled in a single software run) of the survey, where completeness statistics are determined on a per-sector basis.

A technical description of the LRG observational strategy, and on how spectra are turned into redshift estimates, can be found in Ross et al. (2020). Extensive details on the LRG catalog creation, observing strategy, matching targets and spectroscopic observations, veto masks, etc., as well as observational effects such as varying completeness, collision priority, close pairs, redshift failures, systematics related to imaging and their correction are also presented in Ross et al. (2020).

3 MODELING THE GALAXY-HALO CONNECTION: THEORETICAL BACKGROUND

In this section, we provide a concise overview of the theoretical formalism underlying our mock-making procedure, in an increasing level of complexity. Starting from the most conventional HOD approach, we then consider more sophisticated scenarios able to distinguish between quiescent or star-forming galaxies, and with the inclusion of assembly bias – that generalize further the standard HOD framework.

3.1 HOD Modeling: Basics

The Halo Occupation Distribution (HOD; see e.g., Peacock & Smith 2000; Seljak 2000; Scoccimarro et al. 2001; Berlind & Weinberg 2002; Kravtsov et al. 2004; Zheng et al. 2005 for pioneering works, and e.g., Guo et al. 2016; Yuan et al. 2018; Alpaslan & Tinker 2019; Tinker et al. 2019 for more recent implementations and extensions) is a popular framework able to establish a statistical connection between galaxies and dark matter halos bypassing the complex galaxy formation physics, useful to inform models of galaxy formation, interpret LSS measurements, and eventually constrain cosmological parameters. The core assumption underlying any HOD modeling is that all galaxies reside in dark matter halos, and that halos are biased tracers of the dark matter density field. In this regards, knowledge of how galaxies populate, and are distributed within, dark matter halos enables

1 See github.com/desihub/redrock
a complete description of all the statistics of the observed galaxy distribution.

In the most conventional HOD formulation, the central quantity is the probability distribution function (PDF) of galaxies within halos $P(N|_{\text{h}}M_h)$, namely the probability that a halo of mass $M_h$ hosts $N_g$ galaxies in a pre-defined sample. Galaxies are further split into centrals and satellites, and conventionally the occupation statistics of central galaxies $\langle N_{\text{cen}} \rangle$ are modeled separately from satellites $\langle N_{\text{sat}} \rangle$, so that:

$$\langle N_g | M_h \rangle = \langle N_{\text{cen}} | M_h \rangle + \langle N_{\text{sat}} | M_h \rangle.$$  

In the standard mass-only ‘ansatz’ (i.e., halo bias $b_h$ is only a function of halo mass), central galaxies are commonly assumed to reside at the center of their host halos, inheriting the corresponding halo velocity and concentration values, while satellite galaxies are typically designed to follow a radial number density distribution that traces the NFW density distribution (Navarro et al. 1997) of the underlying dark matter halo. Hence, the starting point for any HOD-style model is choosing an analytic form for $\langle N_{\text{cen}} \rangle$ and $\langle N_{\text{sat}} \rangle$. In the vast majority of HOD studies it is assumed that centrals and satellite HODs are completely uncorrelated, so that $\langle N_{\text{cen}}N_{\text{sat}} | M_h \rangle = \langle N_{\text{cen}} | M_h \rangle \langle N_{\text{sat}} | M_h \rangle$. This means that satellites have no knowledge of the central galaxy occupation of their host halo. Moreover, motivated by the occupation statistics of subhalos in high-resolution $N$-body simulations, the PDF of satellite occupation is commonly assumed to be Poissonian, so that $\langle N_{\text{sat}}(N_{\text{sat}} - 1) | M_h \rangle = \langle N_{\text{sat}} | M_h \rangle^2$.

The widespread success of the standard HOD framework in interpreting galaxy clustering statistics, the galaxy-halo connection, and testing cosmology at small scales is not free from several drawbacks. There are in fact a number of simplifying assumptions that enter in the conventional HOD modeling, and may represent a limitation of its predictive power – see for example the recent interesting study by Hadzhiyska et al. (2020).

To start with, while certainly halo mass is the dominant parameter governing the environmental demographics of galaxies, in reality semi-analytical models and hydrodynamical simulations of galaxy formation and evolution predict significant correlations between galaxy properties and halo properties other then mass (i.e., halo formation time, concentration, halo spin, merger history, star formation rate, etc.). Such dependence of the spatial distribution of dark matter halos upon properties besides mass in generically referred to as \textit{halo assembly bias}, and assembly bias has been detected for example in an SDSS sample of galaxy clusters (Miyatake et al. 2016). It has been shown that ignoring assembly bias in HOD modeling yields constraints on the galaxy-DM connection that may be plagued by significant systematic errors (Yang et al. 2006; Blanton & Berlind 2007; Zentner et al. 2014). In addition, while in standard HOD studies centrals and satellite HODs are completely uncorrelated, likely, a degree of central-satellite correlation is always present: such correlations are induced by interesting astrophysics, rather than being simply a nuisance systematics. In fact, the correlation encodes the extent to which the properties of satellite galaxies (stellar mass, color, etc.) may be correlated with the properties of its central galaxy at fixed halo mass (i.e., galactic cannibalism or conformity). Moreover, central galaxies may not be located at the central (minimum) of the halo potential well, the occupation statistics of subhalos in host halos of fixed mass has been shown to deviate from a Poisson distribution especially in the limit where the first occupation moment become large, the satellite distribution may not track the NFW spatial profile of the dark matter halo, and much more – see for example Yuan et al. (2018) for an extensive discussion on such challenges.

Within this complex framework, the main goal of our study is to produce a series of synthetic galaxy catalogs spanning a variety of HODs, in order to assess the robustness of different fitting methodologies relevant for LRG clustering. In this respect, the primary focus is not to improve the HOD modeling and the galaxy-halo connection. However, we have devised the mock challenge in an increasing order of HOD complexity by exploring various methodologies, so that our study may be helpful for ameliorating the galaxy-halo con-

Figure 1. HOD shapes in the Zheng et al. (2007) model, used for the production of galaxy mocks. In the various panels, dotted lines describe the central occupation statistics (Equation 2), dashed lines are used for the satellite occupation statistics (Equation 3), and solid lines represent the composite HODs. Three luminosity thresholds are considered, corresponding to different HOD parameter choices, as reported in Table 1. See the main text for more details.
connection in future works. Motivated by these reasons, we start from the simplest and most conventional HOD framework, and gradually increase the complexity till considering models with assembly bias, particularly useful in exploring intermediate correlations between central-satellites, as well as more generalized HOD approaches. As a byproduct of our work, we are thus able to draw some interesting conclusions regarding the galaxy-halo connection, based on our high-fidelity mocks.

In this work, unless specified otherwise, we always consider two galaxy populations (referred as centrals and satellites); moreover, we generally assume that the central phase space model requires central galaxies to be located at the exact center of the host halo with the same halo velocity, and that the satellite phase space model follows an unbiased NFW profile with a phase space distribution of mass and/or galaxies in isotropic Jeans equilibrium, where the concentration of galaxies is identical to the one of the parent halo.

### 3.2 Traditional HOD: Zheng Model

The most traditional composite HOD model is the one first proposed by Zheng et al. (2007); it represents the backbone for any other HOD framework, and the starting point of this work. The central occupation statistic \( N_{\text{cen}} \) is described by a nearest integer distribution with first moment given by an error function introduced by Zheng et al. (2005), namely:

\[
\langle N_{\text{cen}}(M_h) \rangle = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{\log(M_h) - \log(M_{\text{min}})}{\sigma_{\log M}} \right) \right]
\]

where \( M_h \) is the halo mass, \( M_{\text{min}} \) is the characteristic minimal mass for a halo to host a central galaxy above a luminosity threshold, and \( \sigma_{\log M} \) is the rate of transition from \( \langle N_{\text{cen}} \rangle = 0 \) to \( \langle N_{\text{cen}} \rangle = 1 \), representing the width of the cutoff profile. Hence, central galaxies are characterized just by two HOD parameters. The satellite occupation statistic \( N_{\text{sat}} \) is represented by a Poisson distribution with first moment given by a power law that has been truncated at the low-mass end (Kravtsov et al. 2005), and described by three parameters:

\[
\langle N_{\text{sat}}(M_h) \rangle = \frac{M_h - M_0}{M_1}^\alpha
\]

\[\text{Table 1. HOD parameters adopted for the Zheng et al. (2007) model, corresponding to different ‘luminosity threshold’ values.}\]

| Threshold | \( \log M_{\text{min}} \) | \( \sigma_{\log M} \) | \( \alpha \) | \( M_0 \) | \( M_1 \) |
|-----------|------------------|------------------|--------|------|------|
| Th1 (\( M_t = -19 \)) | 11.60 | 0.26 | 1.02 | 11.49 | 12.83 |
| Std (\( M_t = -20 \)) | 12.02 | 0.26 | 1.06 | 11.38 | 13.31 |
| Th2 (\( M_t = -21 \)) | 12.79 | 0.39 | 1.15 | 11.92 | 13.94 |

where \( \alpha \) is the power law slope of the relation between halo mass and \( \langle N_{\text{sat}} \rangle \), \( M_0 \) a low-mass cutoff in \( \langle N_{\text{sat}} \rangle \), and \( M_1 \) a characteristic halo mass when \( \langle N_{\text{sat}} \rangle \) begins to assume a power law form. Note that the previous distribution can be optionally modulated by the central distribution \( \langle N_{\text{cen}} \rangle \). Redshift has no impact on this model.

Following HALOTOOLS conventions (Hearin et al. 2017; see Section 4.4), the setting of these parameters in our mock-making procedure is controlled by a luminosity threshold, intended as the r-band absolute magnitude of the luminosity of the galaxy sample. The HOD parameters used in our modeling are those of Table 2 in Zheng et al. (2007), and conveniently reported in Table 1 as a function of threshold: specifically, we consider three thresholds in this work, referred globally as ‘threshold 1’ (Th1; \( M_t = -19 \)), ‘standard’ (Std; \( M_t = -20 \)), and ‘threshold 2’ (Th2; \( M_t = -21 \)); the latter one is closer to the characteristics of the eBOSS DR16 LRG sample. Since the meaning of ‘threshold’ in the Zheng model is effectively different from that of the other HOD models considered (based on stellar mass rather than luminosity, with some rough correspondence, see again Section 4.4), we opted to keep the Zheng framework separate from the others, in order to avoid confusion; we also notice that in general stellar mass is a more faithful tracer of the halo mass than galaxy luminosity. The shapes of the Zheng HODs used in this work are shown in Figure 1, for the three conventional choices of HOD parameters corresponding to the previously mentioned threshold values (see Table 1). Note also that in our HOD modeling we do not modulate the satellite distribution by the central one, unless specified otherwise.

### 3.3 Adding the SHMR Complexity: Leauthaud Model

The second model we consider is the Leauthaud prescription (Leauthaud et al. 2011, 2012), a composite HOD framework that extends the standard Zheng formalism by including a parameterization of an underlying stellar-to-halo mass relation (SHMR), which specifies the mean mass of a galaxy as a function of halo mass. The main assumption in this picture is that the SHMR is valid only for central galaxies, as satellites and centrals experience distinct stellar growth rates and so it is necessary to model them separately. To this end, the conditional stellar mass function (CSMF) – denoted as \( \Phi(M_* | M_h) \), with \( M_* \) the mass in stars and \( M_h \) the mass of the parent halo – is divided into centrals and satellites, namely \( \Phi(M_* | M_h) = \Phi_{\text{cen}}(M_* | M_h) + \Phi_{\text{sat}}(M_* | M_h) \). Moreover, \( \Phi_{\text{cen}}(M_* | M_h) \) is modeled stochastically as a log-normal PDF with a log-normal scatter \( \sigma_{\log M_*} \), and normalized to unity. The exact form is (Leauthaud et al. 2012):

\[
\Phi_{\text{cen}}(M_* | M_h) = \frac{1}{\ln(10)\sigma_{\log M_*}\sqrt{2\pi}} \exp \left[ -\frac{\left( \log M_* - \log(f_{\text{SHMR}}(M_h)) \right)^2}{2\sigma_{\log M_*}^2} \right]
\]

where \( f_{\text{SHMR}} \) is the logarithmic mean of the stellar mass given the halo mass for the \( \Phi_{\text{cen}} \) distribution function. Equation (4) incorporates the scatter associated with the determination of stellar masses, as well as the intrinsic scatter in stellar mass at fixed halo mass due to astrophysical processes. The functional form for \( f_{\text{SHMR}} \) is described by 5 shape parameters.
parameters \((M_0, M_1, \beta, \delta, \gamma)\), and defined via its inverse function as (Behroozi et al. 2010):

\[
\log[f_{\text{SHMR}}(M)] = \log(M_1) + f_{\text{low}}(M/M_0) + f_{\text{high}}(M/\delta)
\]

\[
= \log(M_1) + \beta \log \left(\frac{M}{M_0}\right) + \frac{\langle M_h/M_0 \rangle^\beta}{1 + \langle M_h/M_0 \rangle^\gamma} - \frac{1}{2},
\]

(5)

with \(f_{\text{low}}\) and \(f_{\text{high}}\) the low and high mass parts of the SHMF, respectively. In the previous expression, \(M_0\) and \(M_1\) are characteristic stellar and halo masses — respectively — in the \(\langle M_\star, M_h \rangle\) map, \(\beta\) is a low-mass slope of the \(\langle M_\star, M_h \rangle\) map, \(\delta\) is the high-mass slope of the same map, and \(\gamma\) represents the transition between the low- and high-mass behavior of the \(\langle M_\star, M_h \rangle\) mapping. The redshift evolution of the SHMR is modeled by allowing the parameters that define \(f_{\text{SHMR}}\) in Equation (5) to vary linearly with the scale factor \(a\) as:

\[
\log[M_1(a)] = \log(M_1(0)) + \log(M_1(a)/M_1(0)) - 1
\]

(6)

\[
\log[M_0(a)] = \log(M_0(0)) + \log(M_0(a)/M_0(0)) - 1
\]

(7)

\[
\beta(a) = \beta_0 + \beta(a - 1)
\]

(8)

\[
\delta(a) = \delta_0 + \delta(a - 1)
\]

(9)

\[
\gamma(a) = \gamma_0 + \gamma(a - 1)
\]

(10)

Therefore, the model for the SHMR effectively requires 10 parameters. For our modeling, we adopt the same literature values as in the second column of Table 2 in Behroozi et al. (2010) at \(z = 0\), and also reported in the upper part of Table 2 for convenience. Furthermore, Equation (4) requires a functional form for the lognormal scatter \(\sigma_{\log M}\) in the SHMR; motivated by Leauthaud et al. (2012), who found that a halo mass-varying scatter produced no better fit than a model with constant scatter, we assume a constant scatter in this work – which can be thought as the sum in quadrature of an intrinsic component plus a measurement error component. Specifically, we set \(\sigma_{\log M} = 0.20\) in our modeling for all the redshifts considered.

The top panel of Figure 2 provides an example of the effects of varying in turn the 5 main parameters that control the SHMF (Equation 5) at \(z = 0\) by 10% and up to 50% — as specified in the plot with different line styles and colors, when the lognormal scatter is kept constant. The baseline parameters are fixed as in Behroozi et al. (2010) at \(z = 0\) (solid black line, and upper part in Table 2). The bottom panel of the same figure displays the SHMF underlying the Leauthaud model at \(z = 0\), along with its redshift evolution for the two main redshift snapshots considered in our mock-making procedure (\(z = 0.695\) and \(z = 0.865\), respectively — see Section 4): we use these SHMRs in our modeling.

The key difference with respect to the Zheng model relies in the assumption that the stellar mass, rather than the galaxy luminosity, is used to implement the HOD as a more reliable tracer of the halo mass. To this end, for a volume-limited sample of galaxies such that \(M_* > M_{\text{thr}}\), with \(M_{\text{thr}}\) a galaxy threshold mass, the central occupation function \(\langle N_{\text{cen}}(M_h|M_{\text{thr}})\rangle\) is fully specified given \(\Phi_{\text{cen}}(M_h|M_{\text{thr}})\) according to (Leauthaud et al. 2011):

\[
\langle N_{\text{cen}}(M_h|M_{\text{thr}})\rangle = \int_{M_{\text{thr}}}^{\infty} \Phi_{\text{cen}}(M_h|M_{\text{thr}}) \, dM_h.
\]

(11)

Assuming that \(\sigma_{\log M}\) is constant, the previous expression can be readily integrated and becomes:

\[
\langle N_{\text{cen}}(M_h|M_{\text{thr}})\rangle = \frac{1}{2} \left[ 1 - \text{erf} \left( \frac{\log(M_{\text{thr}}) - \log[f_{\text{SHMR}}(M_h)]}{\sqrt{2}\sigma_{\log M}} \right) \right].
\]

(12)

Equation (12) represents a generalization of the Zheng HOD formula (Equation 2), and it is controlled by 5 parameters that enter in the SHMR: plus 1, if we allow for a varying scatter in the SHMR, and plus additional 5 parameters if we also consider redshift evolution in the SHMR. Interestingly,
Equation (2) – i.e. the Zheng model – can be readily recovered from (11) as a limiting case by assuming a constant scatter in the SHMR and by setting \( f_{\text{SHMR}} \) to be a power law; however, this latter assumption is not realistic.

Regarding the satellite occupation function, in the Leauthaud model it is parameterized as a power law of host mass with an exponential cutoff, and can be optionally scaled by \((N_{\text{cen}})\) (this is not done in our case). Specifically:

\[
\langle N_{\text{sat}}(M_{\text{th}} | M_{\text{th}}^{\text{thr}}) \rangle = \left( \frac{M_{\text{h}}}{M_{\text{th}}} \right)^{\alpha_{\text{sat}}} \exp \left( - \frac{M_{\text{cut}}}{M_{\text{h}}} \right)
\]

(13)

where \( M_{\text{th}} \) defines the amplitude of the power law and \( M_{\text{cut}} \) sets the scale of the exponential cutoff; halos with masses \( M_{\text{h}} < M_{\text{th}} \) are extremely unlikely to host a satellite galaxy. Instead of simply modeling \( M_{\text{sat}} \) and \( M_{\text{cut}} \) as constant factors of \( f_{\text{SHMR}}(M_{\text{th}}^{\text{thr}}) \), flexibility is added by enabling \( M_{\text{sat}} \) and \( M_{\text{cut}} \) to vary as power law functions of \( f_{\text{SHMR}}(M_{\text{th}}^{\text{thr}}) \):

\[
\frac{M_{\text{sat}}}{M_{12}} = B_{\text{sat}} \left[ \frac{f_{\text{SHMR}}(M_{\text{th}}^{\text{thr}})}{M_{12}} \right]^{\beta_{\text{sat}}}
\]

(14)

\[
\frac{M_{\text{cut}}}{M_{12}} = B_{\text{cut}} \left[ \frac{f_{\text{SHMR}}(M_{\text{th}}^{\text{thr}})}{M_{12}} \right]^{\beta_{\text{cut}}}
\]

(15)

where \( M_{12} = 10^{12} M_{\odot} \). Hence, satellite occupation statistics, independent of binning schemes and parameterized with threshold samples, are modeled by 5 parameters: \( \alpha_{\text{sat}}, \beta_{\text{sat}}, B_{\text{sat}}, \beta_{\text{cut}}, B_{\text{cut}} \). In detail, \( \alpha_{\text{sat}} \) is the power law slope of the relation between halo mass and the satellite mean occupation function \( \langle N_{\text{sat}} \rangle \), \( \beta_{\text{sat}} \) and \( B_{\text{sat}} \) control the amplitude of the power law slope of \( \langle N_{\text{sat}} \rangle \), and \( \beta_{\text{cut}} \) and \( B_{\text{cut}} \) control the low-mass cut off in \( \langle N_{\text{sat}} \rangle \). These satellite parameters are fixed as in Table 5 of Leauthaud et al. (2011) for the first redshift bin, and also reported in the bottom part of Table 2 for convenience.

In summary, the Leauthaud framework is completely determined by 11 HOD parameters: 6 controlling the central galaxies and 5 for satellite galaxies, plus additional 5 if we also take into account the redshift evolution of the SHMR. In this framework, ‘threshold’ should be intended as the minimal stellar mass of the galaxy sample, rather than galaxy luminosity.

The shapes of the HODs in the Leauthaud model adopted in this work are shown in the top panels of Figures 4 and 5, for the 2 main redshift intervals considered here \((z = 0.695 \text{ and } z = 0.865, \text{ respectively})\) and for 3 thresholds in mass, indicated as ‘Threshold 1’ (Th1; \( M_{\text{th}}^{\text{thr}} = 10^{10} h^{-1} M_{\odot} \)), ‘Standard’ (Std; \( M_{\text{th}}^{\text{thr}} = 10^{10.5} h^{-1} M_{\odot} \)), and ‘Threshold 2’ (Th2; \( M_{\text{th}}^{\text{thr}} = 10^{11} h^{-1} M_{\odot} \)). In particular, Th2 is the one closer to the eBOSS LRG sample, and we mainly focus on this mass interval in our analysis. Note finally that in the Leauthaud framework \( \langle N_{\text{sat}}(M_{\text{th}} | M_{\text{th}}^{\text{thr}}) \rangle \) depends on \( \langle N_{\text{cen}}(M_{\text{th}} | M_{\text{th}}^{\text{thr}}) \rangle \), indicating that in this particular model the occupation statistics of centrals and satellites are correlated.

### 3.4 HOD with Color/SFR: Tinker Model

The next HOD framework we consider has been introduced by Tinker et al. (2013): it is an extension of the Leauthaud formalism previously described, to samples defined by both stellar mass and star formation (SF) activity. In this context, 

| LEAUTHAUD MODEL |
|------------------|
| **Centrals** |
| \( \log[M_{1/0}] \) | 12.35 |
| \( \log[M_{1/1}] \) | 0.30 |
| \( \log[M_{0/0}] \) | 10.72 |
| \( \log[M_{0a}] \) | 0.59 |
| \( \beta_{\text{sat}} \) | 0.43 |
| \( \beta_{\text{cut}} \) | 0.18 |
| \( \delta_{\text{sat}} \) | 0.56 |
| \( \delta_{\text{cut}} \) | 0.18 |
| \( \gamma_{\text{cut}} \) | 1.54 |
| \( \sigma_{\log M_{\text{st}}} \) | 2.52 |

| **Satellites** |
| \( \alpha_{\text{sat}} \) | 1.0 |
| \( \beta_{\text{sat}} \) | 0.859 |
| \( B_{\text{sat}} \) | 10.62 |
| \( \beta_{\text{cut}} \) | -0.13 |
| \( B_{\text{cut}} \) | 1.47 |

Figure 3. SHMFs for central galaxies adopted in the Tinker model and in our mock-making procedure at \( z = 0 \) (solid lines), \( z = 0.695 \) (dotted lines), and \( z = 0.865 \) (dashed lines). Active galaxies are displayed in blue, while quiescent galaxies are represented in brown.
galaxies can be roughly categorized into the star-forming sequence of blue, disky, gas-rich galaxies, and the quiescent, ellipsoidal galaxies with old stellar populations and red colors: the bimodality is firmly in place at $z = 1$ (Tinker et al. 2013, 2019). Therefore, in this model galaxies are divided into quiescent, to indicate galaxies that have little to no star formation and are intrinsically located on the red sequence, and the set of star-forming galaxies. Hence, the Tinker model represents a minimal modification of the Leauthaud prescription to adapt it to passive and SF subsamples of galaxies. The HOD behavior is in fact governed by an assumed underlying SHMR as first introduced in Behroozi et al. (2010), but that is now distinct for star-forming and quiescent populations: each subsample will then have a separate $f_{\text{SHMR}}$ with 2 different sets of HOD parameters related to their specific SHMRs. A constant scatter $\sigma_{\log M}$ in the SHMR is adopted here, but the scatter is different and independent for passive and SF central galaxies. The main difference with respect to the Leauthaud formalism is the following requirement:

$$\int \left\{ f_q(M_h) \Phi_{\text{cen}}(M_{\ast} | M_h) + [1 - f_q(M_h)] \Phi_{\text{cen}}^{SF}(M_{\ast} | M_h) \right\} dM_\ast = 1$$

(16)

where $f_q(M_h)$ is a function specifying the fraction of times that a halo of mass $M_h$ contains a quenched central galaxy (independent of galaxy mass), and $\Phi_{\text{cen}}$ is the conditional stellar mass function for central quiescent or star-forming galaxies, each normalized to unity. The function $f_q(M_h)$ does not have a parametric form, but five halo mass points are chosen at which to specify $f_q(M_h)$ and smoothly interpolate between them, where the 5 masses are evenly spaced in $\log M_h$. Moreover, to avoid explicit dependencies of HOD parameters on bin sizes, all HODs are defined as threshold quantities, which provides maximal flexibility.

Figure 3 shows the SHMFs adopted in the Tinker model, as well as in our mock-making procedure, along with their redshift evolution: solid lines refer to $z = 0$, dotted and dashed lines are at $z = 0.695$ and $z = 0.865$, respectively. Active galaxies are represented in blue, while quiescent galaxies are displayed in brown.

As the Tinker model inherits almost all the features and methods of the Leauthaud framework, the HOD for centrals is the same as in Equation (12), but the parameters of the $f_{\text{SHMR}}$ are independent for each subsample. Moreover, for red central galaxies, the HOD is multiplied by $f_q(M_h)$, and by $1 - f_q(M_h)$ for SF central galaxies.

The occupation statistics of satellite galaxies as a function of halo mass are similar to those of Leauthaud et al. (2011), although the satellite occupation of passive and star-forming galaxies subsamples are treated independently. Hence, a modification is introduced in order to produce a proper cutoff scale by including $f_{\text{SHMR}}$ to the numerator in the exponential cutoff, so that:

$$\langle N_{\text{sat}}(M_h | M_{\ast}^{\text{th}}) \rangle = \left( \frac{M_h}{M_{\ast}} \right)^{\alpha_{\text{sat}}} \exp \left[ - \frac{M_{\ast}}{C_{\text{sat}}} + \int f_{\text{SHMR}}^{-1}(M_h) \right]$$

(17)

This guarantees that satellite occupation fully cuts off at the same halo mass scale as central galaxies of the same mass. In addition, while in Leauthaud et al. (2011) $\alpha_{\text{sat}} = 1$, here the fraction of satellites that are star forming depends on halo mass, so $\alpha_{\text{sat}} = 1$ is allowed to be free for both passive and star-forming subsamples.

In summary, the Tinker model is characterized by 27 free parameters: 11 are needed for the composite HOD of a given subsample (5 for the central SHMR, one additional for the SHMR scatter, plus 5 for the satellite occupation statistics), and 5 pivot points are necessary to specify $f_q(M_h)$. Each set of 27 parameters describes the galaxy-halo relation at a given redshift, and clearly additional quantities are required to characterize the redshift evolution. In this work, we adopt literature values from the lowest redshift bin in Table 2 of Tinker et al. (2013), as reported in Table 3.

The central panels of Figures 4 and 5 show the shapes of the HODs in the Tinker model adopted in this work, for the same 3 thresholds in mass described before for the Leauthaud framework, at $z = 0.695$ and $z = 0.865$, respectively, and also distinguishing between centrals and satellites. 

| Table 3. HOD parameters for central and satellite galaxies assumed for the calibration of the Tinker et al. (2013) model. |
|-----------------|-----------------|
| **TINKER MODEL** | **CENTRALS** |
| $\log [M_1, \text{active}]$ | 12.56 |
| $\log [M_1, \text{quiescent}]$ | 12.08 |
| $\log [M_0, \text{active}]$ | 10.96 |
| $\log [M_0, \text{quiescent}]$ | 10.70 |
| $\beta_{\text{active}}$ | 0.44 |
| $\beta_{\text{quiescent}}$ | 0.32 |
| $\delta_{\text{active}}$ | 0.52 |
| $\delta_{\text{quiescent}}$ | 0.93 |
| $\gamma_{\text{active}}$ | 1.48 |
| $\gamma_{\text{quiescent}}$ | 0.81 |
| $\sigma_{\log M_\ast, \text{active}}$ | 0.21 |
| $\sigma_{\log M_\ast, \text{quiescent}}$ | 0.28 |
| **SATELLITES** | |
| $\alpha_{\text{sat,active}}$ | 0.99 |
| $\alpha_{\text{sat,quiescent}}$ | 1.08 |
| $\beta_{\text{sat,active}}$ | 1.05 |
| $\beta_{\text{sat,quiescent}}$ | 0.62 |
| $B_{\text{sat,active}}$ | 33.96 |
| $B_{\text{sat,quiescent}}$ | 17.9 |
| $\beta_{\text{cut,active}}$ | 0.77 |
| $\beta_{\text{cut,quiescent}}$ | -0.12 |
| $B_{\text{cut,active}}$ | 0.28 |
| $B_{\text{cut,quiescent}}$ | 21.42 |

MNRAS 000, 1–35 (2020)
tive and quiescent galaxy HODs are represented by different colors, as indicated in the panels, and the global HODs are also displayed. As noted by Tinker et al. (2013) and also evident from our figures, the number of quiescent satellites exhibits minimal redshift evolution; all evolution in the red sequence is due to low-mass central galaxies being quenched of their star formation. Moreover, the efficiency of quenching star formation for centrals increases with cosmic time, while the mechanisms that quench the star formation of satellite galaxies in groups and clusters is losing efficiency.

3.5 Decorated HOD: Hearin Model

The fourth galaxy-halo prescription we consider is the Hearin model (Hearin et al. 2016), a decorated HOD framework designed to account for assembly bias, that naturally extends the standard HOD approach, minimally expands the parameter space, and maximizes the independence between traditional and novel HOD parameters. The model builds on early work by Wechsler et al. (2006). The halo occupation statistics are described in terms of two halo properties rather than just one, and the extra degree of freedom has relevant impact on galaxy clustering. The formalism of the model is general and flexible, with parametric freedom, and it can be applied to any halo property in addition to halo mass; it is also readily extendable to describe HODs that depend upon numerous additional halo properties. Interestingly, the decorated HOD formalism allows one to characterize and quantify the degree of central-satellite correlation at fixed halo mass, which is an indication of compelling astrophysics – such as galactic cannibalism or conformity. We refer the reader to Hearin et al. (2016) for extensive modeling details, and report here only a few key aspects relevant to our work.

In particular, the core idea is based on the principle of ‘HOD conservation’, which preserves the moments of the standard HOD formalism: namely, it is required that the marginalized moments of a new decorated framework are equal to those of the standard HOD model, in order to minimize the modifications needed for assembly bias. In this regard, any model $P_{\text{dec}}(N, M_h, x)$ with marginalized moments that satisfy the HOD conservation preserves the moments of $P_{\text{std}}(N, M_h)$, where $P_{\text{std}}$ is the occupation statistics of a standard HOD, $P_{\text{dec}}$ is the occupation statistics of a decorated HOD model, and $x$ represents a secondary halo property such that the HOD depends both on $x$ and $M_h$, and the clustering of halos depends upon $x$. Within this formalism, a standard HOD is recovered in the limiting condition that the strength of the assembly bias is zero.

For central galaxies, in order to construct decorated HOD models that preserve the full $P_{\text{std}}(N_{\text{cen}}, M_h)$ one just needs to ensure that the first order distribution function $\delta N_{\text{cen}}$ satisfies the integral relation (Hearin et al. 2016):

$$\int \delta N_{\text{cen}}(M_h, x) P(x|M_h) dx = 0.$$  \hspace{1cm} (18)

For satellites the situation is more complex, as it is not possible to conserve the HOD under the assumption that both $P_{\text{sat}}(N_{\text{sat}}, M_h)$ and $P_{\text{std}}(N_{\text{sat}}, M_h, x)$ obey Poisson statistics – as typically done: intuitively, this is because there is an additional source of variance associated with the allocation of satellites into sub-populations at a given halo mass. Moreover, under HOD conservation, the average number of central-satellite pairs in massive halos for a decorated model is identical to that of its standard baseline model, except for the narrow range in halo masses for which $0 \approx \langle N_{\text{cen}}|M_h \rangle \leq 1$.

For our purposes, we consider the Hearin model in its simplest formulation, by assuming two discrete halo sub-populations with different occupation statistics at fixed mass; this is essentially a perturbation of the Leauthaud et al. (2011) formalism, with the addition of assembly bias both in centrals and satellites. We choose the halo NFW concentration as the secondary halo property ($x$) used to modulate the assembly bias. Specifically, the first halo sub-population (indicated as ‘type 1’ halos) contains a fraction $P_1$ of all halos at fixed mass, for which $x > \bar{x}(M_h)$; the second sub-population (‘type 2’ halos) contains $P_2 = 1 - P_1$ of all halos at fixed mass, for which $x < \bar{x}(M_h)$. The halo population is split into the $P_1$ percentile of highest-concentration halos, and assigned a satellite galaxy occupation enhancement, while the remaining $P_2 = 1 - P_1$ percentile of lowest-concentration halos receive a satellite galaxy occupation decrement. Essentially, we require halos at fixed mass above- or below-average concentration to have above- or below-average mean occupation. For simplicity, we assume a 50/50 split at each halo mass based on the conditional secondary percentiles; halos within the top 50 per cent of concentration at fixed $M_h$ are assigned to the first subpopulation, and the remaining to the second population (so $P_1 = P_2 = 0.5$). The strength of assembly bias in the occupation statistics of centrals and satellite galaxies is modulated with two free parameters $A_{\text{cen}}$ and $A_{\text{sat}}$ respectively, where $-1 \leq A_{\text{cen}} \leq 1$ and $-1 \leq A_{\text{sat}} \leq 1$. With this choice, a positive value for $A_{\text{bias}}$ implies that halos with above-average concentration have boosted galaxy occupations; note also that more positive values of $A_{\text{bias}}$ correspond to models in which more concentrated halos host more galaxies relative to less concentrated halos of the same mass. When both of these parameters are set to zero, the model is formally equivalent to the baseline ‘no assembly bias model’ of Leauthaud et al. (2011). We consider a constant assembly bias strength at all masses for simplicity, and the sign convention is to choose type-1 halos in the upper percentile of the secondary property. Moreover, we assume that both $P_{\text{sat}}(N_{\text{sat}}, M_h)$ and $P_{\text{dec}}(N_{\text{sat}}, M_h, x)$ are Poisson distributions, so that the decorated HOD is entirely specified by $A_{\text{cen}}$ and $A_{\text{sat}}$.

In our mock-making procedure, we consider two cases for the strength of assembly bias related to centrals and satellites: in the first case (more conservative), we simply set $A_{\text{cen}} = A_{\text{sat}} = 0.5$, so the strength of assembly bias is equal for both centrals and satellites, with the boost to their mean occupation equal to 50% of the maximum allowable strength at each mass; in the second case (less conservative), we set different assembly bias strengths for centrals and satellites, namely $A_{\text{cen}} = 1.0$ and $A_{\text{sat}} = 0.2$. This latter choice is shown in the lower panels of Figures 4 and 5, where we display the shapes of the Hearin HODs for the upper- and lower-percentile split in halo concentration, respectively, as indicated in the plot. In this case, the satellite HODs are also modulated by their corresponding central distributions. The same 3 thresholds in mass described before for the Leauthaud framework are adopted, at $z = 0.695$ and $z = 0.865$, and as usual we display the central occupation statistics (dotted lines), satellite occupation stas-
Figure 4. HOD shapes adopted in our mock-making procedure, at $z = 0.695$, for 3 thresholds in mass, denoted as ‘Thres 1’ ($M_{\text{thr}}^\ast = 10^{10.5} h^{-1} M_\odot$), ‘Standard’ ($M_{\text{thr}}^\ast = 10^{10.5} h^{-1} M_\odot$), and ‘Thres 2’ ($M_{\text{thr}}^\ast = 10^{11} h^{-1} M_\odot$). Top panels display the various HODs in the Leauthaud model. Central panels show the Tinker model, where active and quiescent galaxy HODs are represented by different colors, as indicated in the figure. Bottom panels are for the Hearin HODs, where galaxies are split into upper- and lower-percentiles in terms of halo concentration, respectively, with different assembly bias strength for centrals and satellites ($A_{\text{cen bias}} = 1.0$ and $A_{\text{sat bias}} = 0.2$). In all the plots, the central occupation statistics is displayed with dotted lines, the satellite occupation statistics with dashed lines, and the global HOD shapes with solid lines.

As noted by Hearin et al. (2016, 2017) and Tinker et al. (2019), assembly bias can enhance or diminish the clustering on large scales, but in general it increases the clustering on scales below Mpc – being qualitatively different at large and small scales. Also, assembly bias in satellites versus centrals imprints a distinct signature on galaxy clustering as well as lensing, and the degree to which assembly bias alters galaxy clustering statistics can be quite sensitive to the underlying baseline mass-only HOD of the galaxy population under consideration. In particular, the impact of assembly bias on galaxy clustering is quite sensitive to the steepness of the transition from $\langle N_{\text{cen}} | M_h \rangle_{\text{std}} = 0$ at low host masses to $\langle N_{\text{cen}} | M_h \rangle_{\text{std}} = 1$ at high host masses. This steepness is controlled by the level of stochasticity in the central galaxy stellar mass at fixed halo mass, parameterized in our baseline model by $\sigma_{\text{log} M_*}$. Note that changing the values of $A_{\text{bias}}$...
Figure 5. Same as Figure 4, but at $z = 0.865$.

4 MODELING THE GALAXY-HALO CONNECTION: TOOLS AND METHODOLOGY

In this section, we briefly describe the tools and methodologies behind our mock-making procedure, the main $N$-body simulation used, and the pipeline to produce novel heterogeneous sets of Outer Rim-based galaxy catalogs.

4.1 Outer Rim Mocks

The baseline simulation used for all our mock-making procedure is the Outer Rim (OR) run, extensively described in Heitmann et al. (2019). The simulation has been developed along the glorious tradition of the Millennium simulation (Springel 2005), with similar mass resolution but a volume coverage increase by more than a factor of 200. Currently, the Outer Rim is among the largest high-resolution gravity-only $N$-body simulations ever performed, spanning a $(3h^{-1}\text{Gpc})^3$ volume, and characterized by an unprecedented mass resolution (down to $1.85 \cdot 10^9 h^{-1}M_\odot$) evolving 1.07 trillion particles – i.e., $10^{24}$. The actual size of the simulation was chosen to cover a volume large enough to enable synthetic sky catalogs for eBOSS, DESI and LSST, while main-
Length units displayed in the left panel are in $h^{-1}$Mpc, while the middle panel is a progressive zoom into a $50 \times 50$ $h^{-1}$Mpc$^2$ block. Points the figure are FOF halos, color coded by their mass. Zooming into a smaller $7 \times 7$ $h^{-1}$Mpc$^2$ inset of the halo catalog, the right panel displays the ellipsoidal shape of a halo of mass $4.938 \times 10^{13}$ $h^{-1}$M$_\odot$ contained inside that area, rendered with a 1% random particle subsample. Length units displayed in the left panel are in $h^{-1}$Mpc. The impressive resolution of the simulation, down to $1.85 \times 10^9 h^{-1} M_\odot$, allows one to resolve accurately also relatively low-mass halos.

For a given redshift, our halo catalog (split into 110 subfiles, stored in a way such that they are not contiguous volumes) contains the number of particles in halo (Halo Count), the halo ID (Halo Tag), the halo FOF mass in $h^{-1}$M$_\odot$ units, the comoving halo center positions from the potential minimum (in $h^{-1}$Mpc), and the comoving peculiar velocities of halo centers (in km/s). Note that the halo center is defined by its potential minimum most bound particle, since accurate center-finding is important for measuring the halo concentration, for halo stacking, and for placing central galaxies from HOD modeling.

Figure 6 visualizes a small portion of the Outer Rim halo catalog at $z = 0.865$. Specifically, the left panel shows a $100 \times 100$ $h^{-1}$Mpc$^2$ projection along x and y and across z, with thickness $\Delta z = 50$ $h^{-1}$Mpc, while the middle panel is a progressive zoom into a $50 \times 50$ $h^{-1}$Mpc$^2$ block. Points the figure are FOF halos, color coded by their mass. Zooming into a smaller $7 \times 7$ $h^{-1}$Mpc$^2$ inset of the halo catalog, the right panel displays the ellipsoidal shape of a halo of mass $4.938 \times 10^{13}$ $h^{-1}$M$_\odot$ contained inside that area, rendered with a 1% random particle subsample. Length units displayed in the left panel are in $h^{-1}$Mpc. The impressive resolution of the simulation, down to $1.85 \times 10^9 h^{-1} M_\odot$, allows one to resolve accurately also relatively low-mass halos.

Figure 6 visualizes a small portion of the Outer Rim halo catalog at $z = 0.865$. Specifically, the left panel shows a $100 \times 100$ $h^{-1}$Mpc$^2$ projection along x and y and across z, with thickness $\Delta z = 50$ $h^{-1}$Mpc, while the middle panel is a progressive zoom into a $50 \times 50$ $h^{-1}$Mpc$^2$ block. Points the figure are FOF halos, color coded by their mass. Zooming into a smaller $7 \times 7$ $h^{-1}$Mpc$^2$ inset of the halo catalog, the right panel displays the ellipsoidal shape of a halo of mass $4.938 \times 10^{13}$ $h^{-1}$M$_\odot$ contained inside that area, rendered with a 1% random particle subsample. Length units displayed in the left panel are in $h^{-1}$Mpc. The impressive resolution of the simulation, down to $1.85 \times 10^9 h^{-1} M_\odot$, allows one to resolve accurately also relatively low-mass halos.

For a given redshift, our halo catalog (split into 110 subfiles, stored in a way such that they are not contiguous volumes) contains the number of particles in halo (Halo Count), the halo ID (Halo Tag), the halo FOF mass in $h^{-1}$M$_\odot$ units, the comoving halo center positions from the potential minimum (in $h^{-1}$Mpc), and the comoving peculiar velocities of halo centers (in km/s). Note that the halo center is defined by its potential minimum most bound particle, since accurate center-finding is important for measuring the halo concentration, for halo stacking, and for placing central galaxies from HOD modeling.

For a given redshift, our halo catalog (split into 110 subfiles, stored in a way such that they are not contiguous volumes) contains the number of particles in halo (Halo Count), the halo ID (Halo Tag), the halo FOF mass in $h^{-1}$M$_\odot$ units, the comoving halo center positions from the potential minimum (in $h^{-1}$Mpc), and the comoving peculiar velocities of halo centers (in km/s). Note that the halo center is defined by its potential minimum most bound particle, since accurate center-finding is important for measuring the halo concentration, for halo stacking, and for placing central galaxies from HOD modeling.

For a given redshift, our halo catalog (split into 110 subfiles, stored in a way such that they are not contiguous volumes) contains the number of particles in halo (Halo Count), the halo ID (Halo Tag), the halo FOF mass in $h^{-1}$M$_\odot$ units, the comoving halo center positions from the potential minimum (in $h^{-1}$Mpc), and the comoving peculiar velocities of halo centers (in km/s). Note that the halo center is defined by its potential minimum most bound particle, since accurate center-finding is important for measuring the halo concentration, for halo stacking, and for placing central galaxies from HOD modeling.

For a given redshift, our halo catalog (split into 110 subfiles, stored in a way such that they are not contiguous volumes) contains the number of particles in halo (Halo Count), the halo ID (Halo Tag), the halo FOF mass in $h^{-1}$M$_\odot$ units, the comoving halo center positions from the potential minimum (in $h^{-1}$Mpc), and the comoving peculiar velocities of halo centers (in km/s). Note that the halo center is defined by its potential minimum most bound particle, since accurate center-finding is important for measuring the halo concentration, for halo stacking, and for placing central galaxies from HOD modeling.

For a given redshift, our halo catalog (split into 110 subfiles, stored in a way such that they are not contiguous volumes) contains the number of particles in halo (Halo Count), the halo ID (Halo Tag), the halo FOF mass in $h^{-1}$M$_\odot$ units, the comoving halo center positions from the potential minimum (in $h^{-1}$Mpc), and the comoving peculiar velocities of halo centers (in km/s). Note that the halo center is defined by its potential minimum most bound particle, since accurate center-finding is important for measuring the halo concentration, for halo stacking, and for placing central galaxies from HOD modeling.
4.2 Nseries Mocks

In addition to the heterogeneous sets of high-fidelity Outer Rim mocks developed in this work, we also exploit a small homogeneous set indicated as the Nseries, which has been previously used in the SDSS DR12 galaxy clustering analysis (Alam et al. 2017). The homogeneous set, comprised of 84 mocks in total, is particularly suitable to address cosmic variance and modeling systematics at the sub-percent level, since all mocks have the same underlying galaxy bias model built upon the same cosmology, but each mock is a quasi-variance and modeling systematics at the sub-percent level, 84 mocks in total, is particularly suitable to address cosmic variance, as well as LSS. Moreover, these mocks are cut sky: they have the same angular and radial selection function as the NCG DR12 CMASS sample within the redshift range $0.43 < z < 0.70$, and therefore they include observational artifacts closer to the eBOSS DR16 sample. The N-body simulations from which these cut-sky mocks were created have been produced with GADGET2 (Springel 2005), with input parameters to ensure sufficient mass and spatial resolution to resolve the halos that BOSS galaxies occupy. Specifically, the Nseries cosmology is characterized by $\Omega_m = 0.286$, $h = 0.7$, $\Omega_b = 0.047$, $\sigma_8 = 0.820$, and $n_s = 0.96$. The main difference with respect to the Outer Rim mocks – apart from being cut sky and not built on periodic cubes – is that the Nseries derive from multiple realizations of the dark matter field on a larger volume with different random seeds (i.e., a series of N-body simulations identical in all but in the initial random seed), which allows one to address the impact of cosmic variance: this is not achievable with only a halo catalog nor a single N-body simulation at hands.

4.3 EZmocks

For determining the rescaled covariance matrices functional to the subsequent analyses, we also make use of a new series of DR16 EZmocks, thoroughly described in Zhao et al. (2020). These large number of galaxy catalogs (1000 per tracer), having accurate clustering properties, are generated with a complex methodology built around the Zel’dovich approximation (Zel’’dovich 1970), and effectively including stochastic scale-dependent, non-local, and non-linear biasing contributions; extensive details on the methodology can be found in the first release paper by Chuang et al. (2015). Non-local effects, such as tidal fields not included in linear Lagrangian Perturbation Theory (LPT) or other biasing contributions, are effectively included in both the scatter relation and the tilting of the initial power spectrum. The missing power towards small scales of perturbative approaches is included in the modulation of the initial power spectrum, when fitting for the resulting halo populations. These mocks have accurate clustering properties – nearly indistinguishable from full N-body solutions – in terms of the one-point, two-point, and three-point statistics. The underlying cosmology is based a flat CDM model, with $\Omega_m = 0.307115$, $\Omega_b = 0.048206$, $h = 0.6777$, $\sigma_8 = 0.8225$, and $n_s = 0.9611$. Specifically for LRGs, they contain the complexity of blending the CMASS plus eBOSS LRG samples, as well as all the realistic effects of mask, cut sky, and observational systematics (i.e., fiber completeness, spectroscopic success rate, redshift failures, photometric systematics). For our analysis, we adopt dedicated cubic EZmocks rather than the cut sky set for covariance estimations, to comply with the characteristics of the high-fidelity Outer Rim-based realizations. The EZmocks are extensively used in all the supporting eBOSS DR16 papers and in the final eBOSS consensus analysis. For additional technical details, we refer the reader to the companion paper by Zhao et al. (2020).

4.4 Galaxy Mock-Making Procedure: Methods

Our synthetic high-fidelity galaxy mocks are primarily produced exploiting the standard HALOTOOLS framework (Hearin et al. 2017), and by introducing a number of customizations depending on the desired challenge and model explored (see Section 6, as well as the previous theoretical part). In particular, we interface HALOTOOLS capabilities with the Outer Rim halo catalog at different redshifts. HALOTOOLS is an open-source, community-driven Python powerful package for studying the galaxy-halo connection, which provides a highly modular, object-oriented platform for building HOD models, so that individual modeling features can easily be swapped in and out. This modularity facilitates rigorous study of all the components that makes up a halo occupation model, and has been designed from the ground-up with assembly bias applications in mind. In this view, although our main products are based on the Outer Rim simulation, following the HALOTOOLS philosophy the pipelines developed here are written in a general and flexible manner, so that any type of customization is readily achievable with minimal efforts and modifications; hence, our modular-approach procedure is quite general, and readily applicable to any halo catalog and survey design in mind. The concept of generality and reusability of the code is in fact what has driven this design from the start. In this view, although we limit here our modeling approach to HOD-based techniques (mainly due to limitations in our available halo catalog products), we plan to pursue subhalo and abundance matching methods in follow-up studies, using the same code and modular structure.

Adopting HALOTOOLS conventions, three main primary keyword arguments are used to customize all the instances retrieved by the mock factory, common to all the different HOD models developed here (besides specific HOD parameters), namely: redshift, threshold, and modulate with cenocc – the latter being the modulation of the satellite distribution with the central one. In our mock-making procedure, except for the Hearin framework, all other satellite HODs are not modulated by their corresponding central distributions. Also, as previously mentioned, we treat the conventional Zheng model separately since its HOD is effectively redshift-independent and the meaning of ‘threshold’ in the model is based on luminosity rather than stellar mass, unlike for the other 3 frameworks considered (i.e., Leauthaud, Tinker, Hearin). Specifically, HALOTOOLS is used to populate dark matter halos in the Outer Rim simulation with galaxies having a stellar mass $M_* > 10^{10.0} h^{-1} M_\odot$ (‘Threshold 1’), $M_* > 10^{10.5} h^{-1} M_\odot$ (‘Standard’), and $M_* > 10^{11.0} h^{-1} M_\odot$ (‘Threshold 2’). This roughly corresponds to ‘Threshold 1’ ($M_t = -19$), ‘Standard’
A composite HOD model is fully defined once one specifies the occupation statistics and phase space prescription for centrals and satellites. The theoretical formalism related to each individual model has been presented in Section 3. The corresponding numerical implementation is briefly explained in what follows – noting that dealing with the Outer Rim simulation poses several non-trivial challenges in handling massive datasets. Specifically, at the highest level, we select the redshift, threshold, and number of desired mocks to produce. Then, to populate halos with central galaxies we first calculate the value of \( \langle N_{\text{cen}} \rangle \) for every halo in the simulation according to the HOD formulas in our different prescriptions (Section 3). For every halo in the simulation, we draw a random number \( r \) from \( U[0,1] \), a uniform distribution between zero and unity. For all halos with \( r \leq (N_{\text{cen}}) \), we place a central galaxy at the halo center, leaving all other halos devoid of centrals. Populating satellites is more complicated, because the spatial distributions are nontrivial. The first step is similar to that of the centrals, namely compute \( \langle N_{\text{sat}} \rangle \) for every halo using our specified formulas for a given HOD model (see again Section 3). For each halo, the number of satellites that will be assigned to the halo is then determined by drawing an integer from the assumed satellite occupation distribution \( p(N_{\text{sat}}|M) \) or \( p(N_{\text{sat}}|M_{\text{tot}} \cdot x) \). Satellites are modeled as being isotropically distributed within their halos according to a NFW profile with concentration equal to the parent halo, using the Dutton-Macciò model (Dutton & Macciò 2014). Monte Carlo realizations of both radial and angular positions are generated via the method of inverse transformation sampling. Briefly, first one generates realizations of points uniformly distributed on the unit sphere. These halocentric \((x,y,z)\) coordinates are then multiplied by the corresponding realization of the radial position \( r \), which is determined as follows: first, calculate \( P_{\text{NFW}}(< \tilde{r}|c) \) where \( c \) is the concentration, \( \tilde{r} = r/R_\text{vir} \) is the scale radius, \( R_\text{vir} \) the virial radius of the halo, and \( P_{\text{NFW}}(< \tilde{r}|c) \) is the cumulative probability distribution function of the mass profile of a NFW halo:

\[
P_{\text{NFW}}(< \tilde{r}|c) = \frac{M_{\text{NFW}}(< \tilde{r}|c)}{M_{\text{tot}}} = \frac{g(c\tilde{r})}{g(c)}
\]

where

\[
g(x) = \ln(1 + x) - \frac{x}{1 + x}.
\]

Then, for a halo with concentration \( c \) populated by \( N_{\text{sat}} \), draw \( N_{\text{sat}} \) random numbers \( p \) from \( U[0,1] \). Each value of \( p \) is interpreted as a probability where the corresponding value for the scaled radius \( \tilde{r} \) comes from numerically inverting \( p = P_{\text{NFW}}(< \tilde{r}|c) \). Scaling the \((x,y,z)\) points on the unit sphere by the value \( r \) gives the halocentric position of the satellites.

All these high-fidelity mocks have been produced at the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 using the Cori supercomputer, a Cray XC40 with a peak performance of about 30 petaflops. Cori is comprised of 2,388 Intel Xeon “Haswell” processor nodes, and 9,688 Intel Xeon Phi “Knight’s Landing” (KNL) nodes. The system also has a large Lustre scratch file system and a first-of-its kind NVRAM “burst buffer” storage device. We devised new customized scripts and pipelines to produce such mocks on Cori, exploiting especially the multi-thread architecture. Our mock-making code/pipeline is memory efficient and optimized to the machine. Some additional supporting numerical work has also been carried out using the Korea Institute of Science and Technology Information (KISTI) supercomputing infrastructure.

In closing this part, we note that our main goal in the cubic N-body-based mock-making production and in the related mock challenge is to test the validity and robustness of different BAO and RSD fitting techniques on a common ground against a series of different HOD prescriptions, and validate the clustering analysis pipelines and the various RSD models. Hence, we are not concerned with reproducing exactly all the features of the eBOSS DR16 LRG sample, and this is why the various HOD parameters that enter in the models outlined in Section 3 have been maintained to their corresponding literature values. Realistic observational artifacts related to the LRG sample, such as cut sky, matching number density, observational systematics, etc., are instead part of the EZmock release (Zhao et al. 2020).

5 ANALYSIS: METHODOLOGY

In this section, we briefly describe the three configuration and Fourier space techniques used in the analysis of the challenge mocks, based on three different RSD analytical models – exploiting the full shape (FS) information in the correlation function or power spectrum. The detailed BAO modeling is instead described in our LRG companion papers. All these methods are adopted in the main analysis of the final eBOSS DR16 LRG sample.

5.1 CLPT-GS

The CLPT-GS-based method is a combination of the Convolutional Lagrangian Perturbation Theory (CLPT) and the RSD Gaussian Streaming (GS) formalism, originally developed by Reid & White (2011), Carlson et al. (2013), and Wang et al. (2014). CLPT provides a non-perturbative resummation of Lagrangian perturbation to the two-point statistic in real space for biased tracers. In particular, the two-point correlation function is expanded in its Lagrangian coordinates considering the LRG tracer to be locally biased with respect to the initial CDM overdensity, and the expansion is performed over different orders of the Lagrangian bias function. The key equation for the two-point correlation \( \xi_{\text{LRG}}(r) = \langle \delta_{\text{LRG}}(x)\delta_{\text{LRG}}(x + r) \rangle \), with \( \delta \) the Lagrangian and Eulerian coordinates, respectively, \( \delta \) the overdensity, and \( r \) the LRG separation, is:

\[
1 + \xi_{\text{LRG}}(r) = \int M(r, \mathbf{q}) \, d\mathbf{q}.
\]

where \( M(r, \mathbf{q}) \) is the convolution kernel taking into account the displacements and bias expansion up to its second derivative term. The bias derivative terms are computed using a linear power spectrum, obtained with CAMB (Lewis et al. 2000) for a fixed cosmology – namely, the fiducial cosmology of the analysis. The peculiar velocity effect on
clustering statistic is also modeled, and the pairwise velocity distribution \( v_{12} \) and velocity dispersion \( \sigma_{12} \) are given by (Wang et al. 2014):

\[
v_{12}(r) = \frac{1}{1 + \xi_{\text{LRG}}(r)} \int M_1(r, q) dq \tag{22}
\]

and

\[
\sigma_{12}(r) = \frac{1}{1 + \xi_{\text{LRG}}(r)} \int M_2(r, q) dq \tag{23}
\]

where the kernels \( M_1(r, q) \) and \( M_2(r, q) \) also depend on the first two derivatives of the Lagrangian bias, which are free parameters in the model, in addition to the growth factor. CLPT generates more accurate multipole than linear theory and even the Lagrangian Resummation Theory (LRT; Matsubara 2008), but a better performance is needed in order to study the smaller scales of quadrupoles. To achieve such precision, the real space CLPT models of the two-point statistics are mapped into redshift space following the Gaussian Streaming Model (GSM) formalism proposed by Reid & White (2011). In particular, the pairwise velocity distribution is assumed to have a Gaussian shape dependent on both the angle \( \mu \) between the separation vector and the line-of-sight (LOS), and the LRG separation \( r \) in its parallel (\( r_\parallel \)) and perpendicular (\( r_\perp \)) components with respect to the LOS.

The main equation for the correlation function is given by:

\[
1 + \xi_{\text{LRG}}(r_\perp, r_\parallel) = \frac{1}{2\pi(\sigma^2_{\text{LRG}}(r, \mu) + \sigma_{\text{FOG}}^2)} \left[ 1 + \xi_{\text{LRG}}(r) \right] \exp \left[ -\frac{r_\parallel^2 - y - \mu r_\parallel(\mu, \mu)^2}{2(\sigma^2_{\text{LRG}}(r, \mu) + \sigma_{\text{FOG}}^2)} \right] dq,
\]

where \( \xi_{\text{LRG}}(r) \), \( \xi_{\text{LRG}}(r_\perp, r_\parallel) \), and \( \sigma_{\text{LRG}}(r) \) are computed from CLPT as previously indicated, and \( \sigma_{\text{FOG}} \) is the Fingers of God (FoG) parameter to account for an additional contribution to the velocity dispersion given by satellite galaxies. For the RSD model, the Alcock & Paczynski (1979) effect implementation follows that of Xu et al. (2013). The Alcock-Paczynski distortions are modeled through the \( \alpha \) and \( \epsilon \) parameters, which characterize respectively the isotropic and anisotropic distortion components.

With this technique, the FS RSD analysis in configuration space is performed, and for a given cosmology the model has 4 free parameters, namely \( \{ f, \sigma_8, F', F'' \} \), with \( f \) the linear growth factor and \( F' \) and \( F'' \) the first and second derivatives of the Lagrangian bias function \( F \). For extensive details on this method see Icaza-Lizaola et al. (2019) and Bautista et al. (2020).
5.3 TNS in Fourier Space

While the previous techniques are used to carry out the analysis of the LRG sample in configuration space, the method described here—also based on the TNS model and indicated as ‘\(P_{2\text{TNS}}\)—is performed in Fourier space. To this end, the modeling of the BAO signal within this framework—along with the BAO fitting procedure in Fourier space—are described in Gil-Marín et al. (2020). Here, we briefly illustrate only the strategy adopted for the RSD and Alcock-Paczyński analysis, exploiting the FS information in the power spectrum.

Specifically, the FS formalism employed to describe power spectrum multipoles is the same as the one previously used in BOSS and eBOSS studies (Gil-Marín et al. 2016) and quasars (Gil-Marín et al. 2018). We previously used in BOSS and eBOSS studies for galaxies (Gil-Marín et al. 2020). Here, we briefly illustrate as ‘described here—also based on the TNS model and indicated analysis of the LRG sample in configuration space, the method prescribed in Gil-Marin et al. (2020). Here, we briefly illustrate as ‘described here—also based on the TNS model and indicated

\[
\ell P_{g}(\ell) \equiv \frac{2}{2\pi} \frac{1}{2\ell a_{\ell}^{2}} \int_{-1}^{1} d\mu L_{\ell}(\mu) F_{g}^{(\ell)}(k, \mu, \mu') P_{g}(k, \mu). 
\]  

where explicit expressions for \(k'(k, \mu)\) and \(\mu'(\mu)\) are given in Gil-Marín et al. (2020). We also consider that the shot noise contribution in the power spectrum monopole may differ from a Poisson sampling prediction, and parameterize this potential deviation with a free parameter \(\langle A_{\text{noise}} \rangle\), which modifies the shot noise amplitude without introducing any scale dependence. By default, our measured power spectrum monopole has a fixed Poissonian shot noise contribution subtracted, whereas this is not the case for higher multipoles.

With this technique, the full shape RSD analysis in Fourier space is performed, and for a given cosmology the model has 7 free parameters, namely \((a_{0}, a_{1}, f, r_{\text{NS}})\) and \((b_{1}, b_{2}, \sigma_{\text{FoG}}, \sigma_{\text{TNS}})\). Note that while the BAO analysis consists of using a fixed and arbitrary template to compare the relative BAO-peak positions in the power spectrum multipoles, the FS analysis allows for a full modeling of the shape and amplitude of the power spectrum multipoles, taking into account DM non-linear effects, galaxy bias and RSDs. For extensive details on this method see Gil-Marín et al. (2020).

6 THE GALAXY MOCK CHALLENGE

In this section, we present the main outcomes of the galaxy mock challenge. After a brief description of the available products and some general properties of the mocks, we show selected results in configuration and Fourier space. We eventually compare the complementary BAO/RSD models adopted for the analysis of the complete DR16 eBOSS LRG sample, assessing the theoretical systematic budget. Our findings demonstrate that all the methods are mutually consistent, with comparable systematic errors on the Alcock-Paczyński parameters and the growth of structures, and robust to different HOD prescriptions—thus validating the clustering analysis pipelines.

6.1 Mock Products

For the galaxy mock challenge, we devised three sets of heterogeneous Outer Rim-based galaxy mocks (indicated as ‘Challenge Set 1’, ‘Challenge Set 2’, ‘Challenge Set 3’, respectively).\(^5\) These are cubic mocks, in the Outer Rim cosmology, obtained by populating Outer Rim halo catalogs with galaxies as explained in Section 4.1. Details regarding each set are provided next.

Specifically, ‘Challenge Set 1’ (HOD VARIATIONS) contains a total of 3240 mocks (1620 at \(z = 0.695\), and 1620 at \(z = 0.865\)), grouped into 4 model categories according to the underlying HOD scheme (i.e., Zheng, Leauthaud, Tinker, Hearin); each model category consists of 3 ‘flavors’, denoted as ‘Standard’ (Std), ‘Threshold 1’ (Th1), and ‘Threshold 2’ (Th2). As explained in Sections 3 and 4.4, the meaning of ‘flavor’ is related to the key parameter ‘threshold’, which globally sets all the individual HOD parameters as

\[ V_{\text{gal}}(r, \mu) = \frac{2}{2\pi} \frac{1}{2\ell a_{\ell}^{2}} \int_{-1}^{1} d\mu L_{\ell}(\mu) \int_{-1}^{1} \int_{-1}^{1} d\mu' F_{g}^{(\ell)}(k', \mu, \mu') P_{g}(k, \mu). \]  

where explicit expressions for \(k'(k, \mu)\) and \(\mu'(\mu)\) are given in Gil-Marín et al. (2020). We also consider that the shot noise contribution in the power spectrum monopole may differ from a Poisson sampling prediction, and parameterize this potential deviation with a free parameter \(\langle A_{\text{noise}} \rangle\), which modifies the shot noise amplitude without introducing any scale dependence. By default, our measured power spectrum monopole has a fixed Poissonian shot noise contribution subtracted, whereas this is not the case for higher multipoles.

\[ V_{\text{gal}}(r, \mu) = \frac{2}{2\pi} \frac{1}{2\ell a_{\ell}^{2}} \int_{-1}^{1} d\mu L_{\ell}(\mu) \int_{-1}^{1} \int_{-1}^{1} d\mu' F_{g}^{(\ell)}(k', \mu, \mu') P_{g}(k, \mu). \]  

where explicit expressions for \(k'(k, \mu)\) and \(\mu'(\mu)\) are given in Gil-Marín et al. (2020). We also consider that the shot noise contribution in the power spectrum monopole may differ from a Poisson sampling prediction, and parameterize this potential deviation with a free parameter \(\langle A_{\text{noise}} \rangle\), which modifies the shot noise amplitude without introducing any scale dependence. By default, our measured power spectrum monopole has a fixed Poissonian shot noise contribution subtracted, whereas this is not the case for higher multipoles.
best fit realizations from the corresponding literature dictionary of each HOD model (unless specific customizations are introduced). At a given redshift, we produced 135 mocks per model flavor, by populating the full 3$h^{-1}$Gpc Outer Rim periodic halo catalog box 5 times, and then cutting each full box into 27 subcubes of $1h^{-1}$Gpc side and rescaling the various spatial positions accordingly. This means that effectively we have 27 fully independent (i.e., not sharing the same DM field) mocks per realization, and each of these 27 mocks will have 5 different replicas. According to the modeling explained in Section 3, central galaxies are always located at the center of their parent halos with identical velocities, while the satellite population is statistically different in all the realizations – assuming a NFW profile. We then add RSDs to each individual mock in two different ways: radially, or with the usual plane-parallel approximation. This is the primary set considered in the following analysis.

‘Challenge Set 2’ (REDSHIFT EVOLUTION) is similar to the previous one, but now the redshift evolution is taken into account for one threshold flavor and 3 different HOD prescriptions. In detail, we consider 7 redshift intervals, namely $z = (0.402, 0.502, 0.618, 0.695, 0.779, 0.865, 1.006)$, and produced a set of 2835 mocks (135 x 7 x 3) with 3 HOD schemes (Leauthaud, Tinker, Hearin), for the ‘Threshold 1’ flavor. Even in this case, we consider subcubes of $1h^{-1}$Gpc side.

Finally, ‘Challenge Set 3’ (HOD VARIATIONS / LARGE BOX) is similar to the first one, but in this case we exploit the full Outer Rim box ($3h^{-1}$Gpc) with periodic boundary conditions rather than subcubes, and produced 100 realizations per flavor for all the HODs and thresholds considered in the first set – for a total of 2400 mocks. If desirable, these large-box realizations can be casted into smaller pseudo-independent mocks by performing cuts along different directions of the boxes, and also via the inclusion of partial overlaps in order to maximize the effective volume.

Regardless of the specific set, each mock contains the following information: galaxy spatial positions (in $h^{-1}$Mpc), galaxy velocities (in comoving km/s), the galaxy type (central, satellite), the number of centrals that a halo hosts (either 0 or 1), the number of satellites per halo, the global ID of the halo a galaxy belongs to, the halo mass and virial radius, the central star formation designation for some models (active, quiescent), the number of active or quiescent satellites, and the percentile spit in concentration for models with assembly bias.

A summary of all the synthetic products available, categorized by HOD and redshift, is provided in Table 4. While only a subset of these mocks is used for testing the BAO templates and the RSD models adopted for the characterization of LRG clustering systematics, with this work we release the entire suite of products. Moreover, in addition to the homogeneous Outer Rim mocks, as detailed in Section 4.2 we also exploit 84 homogeneous cut-sky Nseries mocks, which have been previously used in the SDSS DR12 galaxy clustering analysis (Alam et al. 2017), to address cosmic variance in the various methods – since the Nseries derive from multiple realizations of the dark matter field with different random seeds. Here we show only one global Nseries application, while in Gil-Marin et al. (2020) and Bautista et al. (2020) those mocks are extensively used to assess systematics re-

---

**Table 4.** List of Outer Rim synthetic products developed for the galaxy mock challenge.

| Set               | HOD Style | HOD Flavor | Redshift | Box [$h^{-1}$Gpc] | Total Mocks |
|-------------------|-----------|------------|----------|-------------------|-------------|
| Challenge Set 1   | Zheng07   | Th1, Std, Th2 | 0.695, 0.865 | 1.0            | 810         |
|                   | Leauthaud11 | Th1, Std, Th2 | 0.695, 0.865 | 1.0            | 810         |
|                   | Tinker13  | Th1, Std, Th2 | 0.695, 0.865 | 1.0            | 810         |
|                   | Hearin15  | Th1, Std, Th2 | 0.695, 0.865 | 1.0            | 810         |
|                   | Challenge Set 2 | Leauthaud11 | Th1 | 0.402, 0.502, 0.618, 0.695, 0.779, 0.865, 1.006 | 1.0 | 945 |
|                   | Tinker13  | Th1 | 0.402, 0.502, 0.618, 0.695, 0.779, 0.865, 1.006 | 1.0 | 945 |
|                   | Hearin15  | Th1 | 0.402, 0.502, 0.618, 0.695, 0.779, 0.865, 1.006 | 1.0 | 945 |
| Challenge Set 3   | Zheng07   | Th1, Std, Th2 | 0.695, 0.865 | 3.0            | 600         |
|                   | Leauthaud11 | Th1, Std, Th2 | 0.695, 0.865 | 3.0            | 600         |
|                   | Tinker13  | Th1 | 0.695, 0.865 | 3.0            | 600         |
|                   | Hearin15  | Th1 | 0.695, 0.865 | 3.0            | 600         |
Figure 7. Example of the distribution of satellite galaxies within the same and randomly chosen Outer Rim halo, at $z = 0.695$, according to the different HOD prescriptions presented in Section 3. The plot represents a spatial projection $(x - y)$ of the halo, and its spherical shape as determined by its virial radius is also indicated in the various panels. Clockwise, starting from the upper left corner, the Zheng, Leauthaud, Hearin, and Tinker models are shown, respectively.

Table 5. Number densities of the challenge mocks listed in Table 4, ordered per HOD type and flavor.

| Number Density $[10^{-4}(h^3\text{Mpc}^{-3})]$ | HOD Model | Th1 | Std | Th2 |
|-----------------------------------------------|-----------|-----|-----|-----|
| Zheng07                                      | 67.05     | 34.26 | 5.64 |     |
| Leauthaud11                                  | 50.55     | 11.85 | 0.69 |     |
| Tinker13                                     | 26.37     | 8.12  | 0.79 |     |
| Hearin15                                     | 44.93     | 10.82 | 0.69 |     |

related to each individual fitting method in configuration or Fourier space, respectively.

6.2 Challenge Mocks: General Properties

Before moving on to the BAO and RSD analyses related to the galaxy mock challenge, we first provide a global view of the Outer Rim products listed in Table 4. In Figure 7, we show an example on how satellite galaxies are distributed within the same Outer Rim halo, according to the different HOD prescriptions presented in Section 3, to convey some intuition on the galaxy-halo connection modeling and the spatial location of satellites. The plot displays the x-y spatial projection at $z = 0.695$ of a randomly chosen halo, with its spherical shape determined by its virial radius. The standard Zheng model is shown in the upper left panel, and clockwise the Leauthaud, Hearin, and Tinker models are displayed, respectively. While in all the HOD schemes the satellite phase space statistics follow an unbiased NFW profile with a phase space distribution in isotropic Jeans equilibrium and galaxy concentration identical to that of the parent halo, more sophisticated frameworks such as the Tinker model (lower left corner) are able to distinguish between active and quiescent populations (indicated with different colors in the panel), thus providing additional useful physical insights.

Table 5 reports the number densities of the challenge mocks, expressed in units of $10^{-4}(h^3\text{Mpc}^{-3})$, and ordered by HOD type and flavor. The ‘Th2’ models of Leauthaud, Tinker, and Hearin are those characterized by a number density closer to the eBOSS LRG sample (see e.g. Figure 1 in Gil-Marín et al. 2020 and Figure 1 in Bautista et al. 2020 for details), and are extensively used in the next sections to test the LRG BAO and RSD analysis pipelines. The other threshold levels are more suitable for example for assessing the details of the satellite distributions and for studying the galaxy-halo connection, particularly in relation to ELGs; they are only marginally explored in this work, although of high interest, and left to future applications.

Figure 8 shows examples of the three-dimensional galaxy clustering quantified by the 2-point spatial correlation $\xi(r)$, as a function of separation $r$. The various measurements are performed at $z = 0.695$ (top panels) and at $z = 0.865$ (bottom panels), for the mocks belonging to ‘Challenge Set 1’. From left to right, the ‘Th1’, ‘Std’, and
Figure 8. Clustering properties of the ‘Challenge Set 1’. Examples of 2-point spatial correlation functions computed at $z = 0.695$ (top panels) and at $z = 0.865$ (bottom panels), for the three threshold levels denoted as ‘Th1’, ‘Std’, and ‘Th2’ – in decreasing number density order, from left to right. The Leauthaud (blue), Tinker (red), and Hearin (green) HOD models are displayed with different colors. They are characterized by approximately similar clustering properties, at a fixed threshold level. See the main text for more details.

Figure 9. Spatial clustering of satellite galaxies at $z = 0.865$, split by active/star-forming (blue) and quiescent (red) – as described by the Tinker model for the ‘Std’ threshold level. This more complex HOD framework provides interesting insights on the galaxy-halo connection physics.

‘Th2’ flavors are displayed, respectively. The HOD models of Leauthaud (blue), Tinker (red), and Hearin (green) are displayed with different colors: they are characterized by approximately similar clustering properties, at a fixed threshold level. Each measurement represents an average over 135 mocks, according to the specific HOD style and flavor, and errorbars are 1σ variations. The effect of a decreasing number density (from left to right) is of an overall increase in the clustering and BAO peak amplitude, with a relatively small redshift dependence.

Finally, Figure 9 displays an example of the 2-point spatial clustering of satellites, split by active/star-forming (blue) and quiescent (red) galaxies, at $z = 0.865$ for the ‘Std’ threshold level (see Table 5). The Tinker HOD formalism used in this work – primarily for LRG studies – provides also interesting applications to ELGs that go beyond the scope of this paper, but that highlight the flexibility of our mock-making procedure and mock products in exploring the physics of the galaxy-halo connection.

6.3 Galaxy Mock Challenge: BAO Analysis and HOD Systematics

Approximate catalogs such as the EZmocks (Zhao et al. 2020) are in principle sufficient for covariance estimates and for quantifying systematic biases in BAO studies, while the analysis of the PS of the correlation function and power spectrum requires high-fidelity ($N$-body-based) mocks to precisely test the modeling. Nevertheless, using high-resolution mocks, we are able to characterize the impact of systematics in HOD modeling both on BAO and RSD constraints with high-accuracy. Specifically, the main goals are to quantify possible effects induced by different galaxy HOD schemes on the cosmic growth rate and obtain useful information on parameter inference based on HOD variations, to assess the impact of an arbitrary choice of the BAO reference template on the inferred cosmological parameters, and more generally...
to determine the theoretical systematic budget and validate the clustering analysis pipelines.

The standard procedure common to all BAO fitting methods is to assume a fixed and arbitrary template, and compare the relative BAO peak positions in the correlation function or power spectrum multipoles. Reconstruction techniques such as those presented in Burden et al. (2014, 2015) are then applied to the density field, in order to remove a fraction of the RSDs and the nonlinear motions of galaxies. The BAO feature in the 2-point statistics (both in configuration and Fourier space) is sharpened, increasing the precision of the measurement of the acoustic scale.

The BAO scale measurement in configuration space adopted here is the same as the one described in previous SDSS publications (i.e., Alam et al. 2017; Ata et al. 2018; Bautista et al. 2018), and thoroughly illustrated in our companion paper Bautista et al. (2020), while the modeling of the BAO signal along with the BAO fitting procedure in Fourier space are explained in detail in Gil-Marin et al. (2020). In particular, for the latter case, the power spectrum anisotropic signal is modeled in order to measure the BAO peak position and marginalize over the broadband information – taking into account the BAO signal both in the radial and transverse line-of-sight (LOS) directions. Generally, BAO results are obtained from pre- and post-reconstructed data, while RSD results use only the non-reconstructed sample. In the following analyses, we assume standard dependency of the growth rate $f$, and adopt a smoothing scale of 15 $h^{-1}$Mpc. Whenever required, galaxy redshifts are converted into radial comoving distances for clustering measurements, using the cosmological parameters of the OR simulation. As shown in Bautista et al. (2020), the analysis methodology is insensitive to the choice of a fiducial cosmology.

Figure 10 is an example of the redshift-space galaxy clustering (monopole and quadrupole), along with corresponding BAO fits, as inferred from the average of 27 OR-based mocks at $z = 0.695$ having different HOD schemes and a “flavor” corresponding to ‘Th2’. From left to right, the Leauthaud, Tinker, and Hearin models are displayed, respectively. Top panels are for the pre-reconstructed fields, while bottom panels refer to post-reconstructed fields. Specifically, for cosmological analyses the information contained in the anisotropic 2-point correlation function $\xi(s, s')$ – decomposed into polar coordinates $(s, \mu)$ aligned with the LOS direction, with $\mu$ the cosine of the angle between the LOS and separation vector directions, and $s$ the norm of the galaxy separation vector $s$ – is compressed into the correlation function multipole moments $\xi_\ell$, obtained by decomposing $\xi(s, s')$ on the basis of Legendre polynomials $P_\ell$ as:

$$\xi_\ell(s) = (2\ell + 1) \sum_i \xi(s, \mu_i) P_\ell(\mu_i) \Delta \mu.$$  \hspace{1cm} (32)

In the previous expression, only even multipoles do not vanish, and the correlation function is binned according to the absolute value of $\mu$. In our analyses, we only consider the $\ell = 0, 2, 4$ moments, namely monopole, quadrupole, and hexadecapole (whenever specified), and $\xi(s, s')$ is quantified with the classical Landy & Szalay (1993) estimator. The pair counts are binned into $5h^{-1}$Mpc bins in separation and 0.01 in $\mu$. In the panels of Figure 10, the BAO feature is clearly seen at $s \approx 100h^{-1}$Mpc, as well as the impact of the reconstruction procedure: the BAO feature appears in fact much sharper in the bottom panels. The various fits shown in the figure are obtained with the BAO technique adopted in Bautista et al. (2020) for the analysis of the correlation function in configuration space, and correlation function multipoles $\xi_\ell(s)$ are rendered as a function of separations $s$ relevant for BAO ($30 \leq s \leq 180h^{-1}$Mpc), starting from the modeling of the redshift-space anisotropic power spectrum. In particular, as explained in Bautista et al. (2020), the nonlinear broadening of the BAO peak is modeled by multiplying the “peak-only” power spectrum $P_{\text{peak}}$ by a Gaussian term with $\Sigma^2_{\text{ drag}}(\mu) = \Sigma^2_{\text{drag}} + \Sigma^2_{\text{intrinsic}}(1-\mu^2)$, with $\Sigma_{\text{drag}}$ and $\Sigma_{\text{intrinsic}}$ the BAO damping terms, and the non-linear random motions on small scales are rendered with a Lorentzian term parametrized by $\Sigma_s$. When performing fits to the multipoles of a single realization of the survey, the values of $(\Sigma_\theta, \Sigma_\perp, \Sigma_s)$ are maintained fixed to improve convergence. Moreover, the BAO peak position is parameterized via two dilation parameters that scale separations into transverse ($\alpha_\perp$) and radial ($\alpha_\parallel$) directions. These quantities are related to the comoving angular diameter distance $D_M(z) = (1 + z)D_A(z)$ and to the Hubble distance $D_H(z) = c/H(z)$ as:

$$\alpha_\parallel = \frac{D_M(z_{\text{eff}})/H_{\text{drag}}}{D_M(z_{\text{eff}})/H_{\text{drag}}}.$$  \hspace{1cm} (33)

| CF [Th2] | Leauthaud | Tinker | Hearin |
|----------|-----------|--------|--------|
|          | BAO       | Post-Rec |
| $a_\perp$ | 0.9990 ± 0.0080 | 0.9922 ± 0.0073 | 1.0074 ± 0.0083 |
| $a_\parallel$ | 1.0084 ± 0.0164 | 1.0234 ± 0.0147 | 1.0040 ± 0.0157 |
| $\Sigma_\perp$ | 6.7663 ± 1.1320 | 7.6741 ± 1.2932 | 7.4335 ± 1.0760 |
| $\Sigma_\parallel$ | 10.5333 ± 1.5795 | 8.4652 ± 1.5436 | 9.5687 ± 1.6354 |
| $\beta$ | 0.2685 ± 0.0965 | 0.2404 ± 0.1017 | 0.2137 ± 0.0910 |
| $b$ | 2.6748 ± 0.1315 | 2.4769 ± 0.1393 | 2.7525 ± 0.1339 |
| $\chi^2$ | 129.4 | 115.4 | 126.0 |

Table 6. BAO fits to the average pre- and post-reconstructed 2PCFs for different HOD prescriptions over 27 corresponding Outer Rim mock realizations, at $z = 0.695$, with the ‘Th2’ flavor – as displayed in Figure 10.
Figure 10. Monopole and quadrupole of the average 2PCFs as computed from a subset of 27 OR-based mocks per HOD type, and fits of the BAO feature as seen in the correlation function multipoles. Top panels show results for the pre-reconstruction case, bottom panels refer to the reconstructed density field. The HOD models of Leauthaud, Tinker, and Hearin are shown – from left to right, respectively – for the ‘Th2’ flavor at $z = 0.695$. Note that the BAO feature (around 100$ h^{-1}$Mpc) appears much sharper after application of the reconstruction procedure, as expected. Results of these fits are reported in Table 6.

\[
a_{\parallel} = \frac{D_H(z_{\text{eff}})}{D_H(z_{\text{drag}})} \frac{r_{\text{drag}}}{r_{\text{drag}}},
\]

with $r_{\text{drag}}$ the comoving horizon scale at the drag epoch. Fits on mock multipoles are performed – including hexadecapole, which however does not add extra information. In the procedure, BAO broadband parameters are let free while both dilation parameters are allowed to vary between 0.5 and 1.5. A total of 9 parameters are fitted simultaneously. Table 6 contains the results of such BAO fits in configuration space, where in particular $b$ is the linear bias and $\beta = f/b$ is the RSD parameter. The covariance matrix used for the fit is obtained from 1000 EZmocks, properly rescaled by the difference in particle number to match the characteristics of the OR-based mocks. Note that expected statistical errors in the eBOSS LRG data sample are of the order of $\sim 1.9\%$ for $a_{\perp}$ and $\sim 2.6\%$ for $a_{\parallel}$, and that reconstruction improves constraints on $a_{\perp}$ and $a_{\parallel}$, as expected.

Figure 11 shows examples of redshift-space galaxy
Figure 11. Monopole and quadrupole of the average power spectra as computed from a subset of 27 OR-based mocks per HOD type, and fits of the BAO feature as seen in the power spectrum multipoles (solid lines). The corresponding no-wiggle model is also reported in the figure, with dotted lines. Top panels show results for the pre-reconstruction case, bottom panels refer to the reconstructed density field. The HOD models of Leauthaud, Tinker, and Hearin are shown – from left to right, respectively – for the ‘Th2’ flavor at $z = 0.695$. Results of these fits are reported in Table 7.

Power spectra as computed from the average of 27 OR-based mocks, each set being characterized by a different HOD scheme, at $z = 0.695$ for the ‘Th2’ flavor. The plot represents the analogous, in Fourier space, of the previous correlation function estimates in configuration space. The pre- (top panels) and post-reconstructed (bottom panels) monopoles and quadrupoles of the power spectra are shown, for the Leauthaud, Tinker, and Hearin models – from left to right, respectively. Fits are obtained with the BAO theoretical model of Gil-Marin et al. (2020), considering wave numbers between $0.02 \leq k [\text{h}\text{Mpc}^{-1}] \leq 0.30$, and the corresponding results are reported in Table 7. Unlike correlation function calculations, Discrete Fourier Transform (DFT) methods used to compute the power spectrum multipoles are quite sensitive to the assumption of periodic boundary conditions, and therefore a procedure denoted as ‘padding’ is applied in this process, to mitigate non-periodicity effects. The detailed effects of non-periodicity on BAO measurements are discussed in Gil-Marin et al. (2020). In particular, results of such analyses show that no significant changes are observed...
parameters are computed in the OR cosmology, at the expected values for the dilation power spectrum for different HOD prescriptions over 27 corresponding Outer Rim mock realizations, at $z = 0.695$, with the ‘Th2’ flavor – as displayed in Figure 11.

| PS [Th2] | Leauthaud | Tinker | Hearin |
|----------|-----------|--------|--------|
| **BAO**  | **Pre-Rec** |        |        |
| $\alpha_\perp$ | 1.0028 ± 0.0107 | 0.9953 ± 0.0124 | 1.0108 ± 0.0099 |
| $\alpha_\parallel$ | 0.9885 ± 0.0178 | 1.0023 ± 0.0177 | 0.9779 ± 0.0143 |
| $\Sigma_{\perp}$ | 4.2992 ± 1.8875 | 5.9568 ± 2.0364 | 4.1865 ± 1.8043 |
| $\Sigma_\parallel$ | 9.6867 ± 1.8250 | 7.5044 ± 2.2301 | 7.0045 ± 2.0438 |
| $\chi^2$ | 44.3 | 34.5 | 57.8 |

| **BAO**  | **Post-Rec** |        |        |
|----------|--------------|--------|--------|
| $\alpha_\perp$ | 0.9976 ± 0.0075 | 0.9975 ± 0.0088 | 1.0122 ± 0.0075 |
| $\alpha_\parallel$ | 0.9938 ± 0.0104 | 0.9976 ± 0.0121 | 1.0002 ± 0.0113 |
| $\Sigma_{\perp}$ | 15.0 ± 1.5 | 15.0 ± 1.5 | 15.0 ± 1.5 |
| $\Sigma_\parallel$ | 1.0585 ± 0.7909 | 1.3377 ± 0.9904 | 1.1071 ± 0.8221 |
| $\chi^2$ | 94.4 | 69.0 | 120.1 |

in terms of $\alpha_\perp$, while shifts at the level of 2 – 3% can be systematically seen in $\alpha_\parallel$ if padding is not applied. Hence, non-periodic effects are relevant in determining $\alpha_\parallel$, but they do not impact significantly $\alpha_\perp$. Moreover, no relative shifts in any of the $\alpha$ parameters are seen when the HOD model or flavor is varied – as we show next.

Figure 12 summarizes and confronts the performance of the BAO fitting technique adopted in the analysis of the final eBOSS LRG sample, in configuration and Fourier space, with respect of variations in the underlying HOD model. For each mock realization, at a fixed HOD scheme and threshold flavor, the correlation function and power spectrum are computed along with their multipoles, respectively. Subsequently, fits for the BAO peak position are performed – both to the pre- and post-reconstructed synthetic catalogs – to determine the dilation parameters $\alpha_\parallel$ and $\alpha_\perp$ and their corresponding errors. The expected values for the dilation parameters are computed in the OR cosmology, at the effective redshift of $\bar{z} = 0.695$. In addition, fits to the average multipoles of a given set of mocks (27 mocks per set) are also carried out, to probe biases in a very high precision configuration. In particular, the BAO pipeline on the power spectrum monopole and quadrupole is run in the interval $0.02 \leq k [h \text{ Mpc}^{-1}] \leq 0.30$. The various BAO fittings are performed by fixing the BAO damping parameters ($\Sigma_\parallel$, $\Sigma_\perp$) at their best-fitting values on the mean of the pre- and post-reconstructed mocks, and the analysis is done in terms of the scaling parameters $\alpha_\perp$ and $\alpha_\parallel$. The various covariances used in the analysis of the OR-based mocks are derived from the EZMocks, properly rescaled to account for differences in number density (see Table 5). As explained in detail in Bautista et al. (2020), the final BAO model is a combination of the cosmological multipoles $\xi_\ell$ and a smooth function of separation, which accounts for unknown systematic effects in the survey that can potentially contaminate the results.

Table 8 contains the results when fitting the mean corre-lation functions and power spectra (rows labeled ‘Mean’) of a set of 27 independent realizations of the OR-based mocks with different HOD prescriptions (‘Challenge Set 1’), as well as the mean of the fits of individual realizations (rows labeled ‘Individual’). Those data are shown in Figure 12, where each sub-panel displays the difference between the measured $\alpha_\parallel$ and $\alpha_\perp$: their expected value are inferred for the pre- (left panels) and post-reconstructed (right panels) catalogs. Mean estimates are displayed in red with filled rectangles as derived from configuration space techniques, and in blue with filled triangles as determined with Fourier space techniques. The associated errors are consistently the errors of the mean, obtained by rescaling the related EZMocks covariance by the number of realizations, $N_{\text{OR}} = 27$. Therefore, these errors are a factor of $\sqrt{N_{\text{OR}}}$ smaller than the error one would obtain for a single realization of these mocks. Analogous empty symbols are used to display the same corresponding individual measurements for pre-and post-reconstruction catalogs, respectively. In this case the error associated is the root mean square (rms) of all the individual fits, scaled by the square root of the number of realizations ($\sqrt{N_{\text{OR}}}$). The second and fourth sub-panels in Figure 12 show the difference between the measured and the expected values of $\alpha_\parallel$ and $\alpha_\perp$ in terms of number of statistical $\sigma$ of the error of the mean, and the $\text{rms}/\sqrt{N_{\text{OR}}}$. The horizontal grey bands highlight the $1\sigma$ error level. In general, considering fits to the mean over 27 realizations, the reported dilation parameters for all the different HODs are consistent with their expected values within 0.8% for $\alpha_\perp$ and 1.2% for $\alpha_\parallel$. Recall that the expected statistical errors in the eBOSS LRG data sample are of the order of ~ 1.9% for $\alpha_\parallel$ and ~ 2.6% for $\alpha_\perp$. From the $N$-body mocks we do not observe any significant BAO peak position shift with respect to their corresponding expected value in any of the post-reconstructed catalogs analyzed. The BAO pipeline performs well with different HOD models; some fluctuations are present, but their values lie always below the ±2$\sigma$ limit, hence the shifts are not significant. Overall, we do not detect any relative systematics due to different HOD modeling, although the statistical precision of the Outer Rim-based mocks is comparable to the statistical precision of the LRG sample (see Section 7). Interestingly, from Figure 12 it is evident that the reconstruction procedure (right panels) generally ameliorates the agreements of the $\alpha$-parameters with their expected values. Moreover, it is also worth noticing that for the average values most of the detected discrepancy (after reconstruction) arises from the Hearin HOD model; this is not unexpected, since we have considered a quite extreme case of assembly bias both in the central and satellite galaxy population – as explained in Section 3.

Finally, Figure 13 provides a different and interesting insight into the galaxy-halo connection, and also shows a remarkable consistency between BAO techniques in configuration and Fourier space. Specifically, the figure displays the mean of the BAO fits obtained by averaging the 27 OR mocks at ‘Th2’ having different HODs (filled rectangles and triangles, as in Figure 12), but now compared with the average across all the HOD models – indicated with identi-
Figure 12. Performance of the BAO fitting methods in configuration and Fourier space, with respect of variations in the underlying HOD model. Fits to individual mock realizations as well as to the mean of a set of 27 independent realizations of the OR mocks (from the ‘Challenge Set 1’) are carried out (see Table 8). The difference between the measured $\alpha_\parallel$ and $\alpha_\perp$ from pre- (left panels) and post-reconstructed (right panels) catalogs are displayed with different symbols and colors, as indicated in the figure. This BAO fitting methodology is adopted in the analysis of the final eBOSS LRG sample. Overall, we do not detect any significant systematics due to different HOD prescriptions. See the main text for more details.

cal corresponding open symbols. The shaded areas represent the 1% error level on the $\alpha$ parameters. This allows one to disentangle the systematics introduced by the HOD modeling (open symbols) versus the theoretical systematics related to BAO fitting methodologies (filled symbols). After application of the BAO reconstruction procedure, all of the measurements are within 0.5 – 1.2% of their expected values, thus below the statistical precision of the eBOSS LRG sample. Hence, for BAO-only fitting methods, both modeling and HOD systematics are subdominant to the global systematic error budget and the BAO analysis is unbiased. However, sub-percent level corrections may become relevant for future surveys like DESI, that are expected to achieve sub-percent statistical precision on the galaxy sample.

6.4 Galaxy Mock Challenge: RSD Analysis and HOD Systematics

In Section 5, we have briefly described the three RSD theoretical models adopted for the analysis of the final eBOSS DR16 LRG sample. Here, we confront those models and show that they are mutually consistent, with comparable systematic errors on the Alcock-Paczynski parameters and the growth of structure – as well as robust to different HOD prescriptions. While for the previous BAO-only analysis simply the BAO peak position has been taken into account, here we consider a full modeling of the shape and amplitude of the correlation function and power spectrum multipoles, including nonlinear DM effects, galaxy bias, and RSDs. We generically refer to this methodology as the ‘full shape’ (FS) analysis. Quantitative investigations involving the correlation function or power spectrum FS require high fidelity $N$-body-based mocks to test and validate the underlying RSD models, and typically such analyses are only performed over pre-reconstructed synthetic catalogs. The overall aim is to quantify the impact of different HOD prescriptions used to populate simulated halos with galaxies on RSD constraints. Our primary focus here is thus on modeling and HOD systematics, while Icaza-Lizaola et al. (2019), Bautista et al. (2020), and Gil-Marin et al. (2020) also examined the impact of the choice of scales in the fits and the choice of a fiducial cosmology. In particular, their conclusions (directly relevant for the current work) suggest that the most robust results and optimal configuration for the FS analysis of the correlation function are obtained with monopole and quadrupole in the range $20 \leq s [h^{-1}\text{Mpc}] \leq 130$, and hexadecapole in the interval $25 \leq s [h^{-1}\text{Mpc}] \leq 130$ for the TNS model in configuration space, while using the interval $25 \leq s [h^{-1}\text{Mpc}] \leq 130$ for all the moments when considering CLPT-GS. Regarding power spectrum computations, the optimal range of scales are $0.02 \leq k [h/\text{Mpc}] \leq 0.15$, and results include the hexadecapole. In what follows, the analysis is carried out in the Outer Rim fiducial cosmology, and the monopole, quadrupole, and hexadecapole ranges are those previously specified – always set for optimal performance. Moreover, for the analysis of the challenge mocks in Fourier space, a procedure called padding is applied, in order to prevent the impact of non-periodicity to affect results when applying the discrete Fourier transform. Also, the mock covariances adopted in these investigations are derived from a set of 1000 EZmocks, and properly rescaled by the difference in particle number.

Figure 14 summarizes the main results of the RSD FS analysis, confronting the three different modeling tech-
Insights on the galaxy-halo connection, obtained by confronting BAO fitting methods in configuration and Fourier space (see Table 8). Filled rectangles and triangles in the various panels are mean estimated of the BAO fits obtained by averaging the 27 OR mocks at ‘Th2’ having different HODs, while identical open symbols show averages across all the HOD models. After application of the BAO reconstruction procedure, all of the measurements are within $0.5 - 1.2\%$ of their expected values (i.e. the shaded green areas in the figure highlight the $1\%$ error level), thus below the statistical precision of the eBOSS LRG sample.

Figure 13. Insights on the galaxy-halo connection, obtained by confronting BAO fitting methods in configuration and Fourier space (see Table 8). Filled rectangles and triangles in the various panels are mean estimated of the BAO fits obtained by averaging the 27 OR mocks at ‘Th2’ having different HODs, while identical open symbols show averages across all the HOD models. After application of the BAO reconstruction procedure, all of the measurements are within $0.5 - 1.2\%$ of their expected values (i.e. the shaded green areas in the figure highlight the $1\%$ error level), thus below the statistical precision of the eBOSS LRG sample.

In general, CF-TNS seems to imply slightly larger errors than CLPT-GS. Comparing fits on the mean with the mean of individual fits, for $a_\perp$ and $f\sigma_8$ there is good agreement in values and errors. The most significant differences are found for the best fits of $a_\parallel$, but this comes with no surprise: it is in fact expected that fits on individual mocks are dominated by the low signal to noise, as the effective volume of a single OR mock is $1.10 \, \text{Gpc}^3$ – thus relatively small. Also, for $a_\parallel$, the fit of the mean is a more robust estimate of potential biases. From the results of fitting the mean, since we do not observe biases larger than $2\sigma$, we conclude that different HODs do not have a significant impact in the fits even from a FS RSD analysis. This type of systematics is always below the statistical error of the LRG sample.

It is also quite interesting to notice the remarkable consistency and (nontrivial) agreement between RSD FS tech-
Table 8. Performance of the BAO templates on OR-based mocks, for the ‘Th2’ flavor and different HOD models. Mocks are analyzed in their own Outer Rim cosmology, so the expected values for both $\alpha_1$ and $\alpha_2$ are 1. For each set of mocks, the results from pre- and post-reconstruction catalogs are presented. We report both the results of fitting the mean of all the mocks, indicated with ‘Mean’, and the mean of individual fits on the mocks, indicated as ‘Individual’. For the fit to the mean, the error quoted is the 1σ of the error on this fit, where the covariances are scaled by the $27 \text{OR}$ realizations per HOD used to compute the mean. For the mean of individual best-fits, the error quoted is the rms divided by $\sqrt{N_{\text{OR}}}$, where $N_{\text{OR}} = 27$. The average of the best-fits is then performed over $N_{\text{OR}}$. Consequently, the errors of ‘Mean’ and ‘Individual’ are comparable. Results of these fits are displayed in Figures 12 and 13.

| BAO Analysis Type | HOD Type | HOD Flavor | Analysis Details | $\alpha_1 - \alpha_1^{\exp}$ | $\alpha_2 - \alpha_2^{\exp}$ | $N_{\text{tot}}/N_{\text{or}}$ |
|-------------------|----------|------------|-----------------|-----------------------------|-----------------------------|-----------------------------|
| Configuration Space [Pre-Rec] | Leauthaud | Th2 | Mean | 0.0084 ± 0.0164 | −0.0010 ± 0.0080 | 1/1 |
| | Leauthaud | Th2 | Individual | 0.0283 ± 0.0197 | −0.0072 ± 0.0123 | 27/27 |
| | Tinker | Th2 | Mean | 0.0234 ± 0.0147 | −0.0078 ± 0.0073 | 1/1 |
| | Tinker | Th2 | Individual | 0.0166 ± 0.0172 | 0.0000 ± 0.0104 | 27/27 |
| | Hearin | Th2 | Mean | 0.0040 ± 0.0157 | 0.0074 ± 0.0083 | 1/1 |
| | Hearin | Th2 | Individual | 0.0035 ± 0.0185 | 0.0090 ± 0.0089 | 27/27 |
| Configuration Space [Post-Rec] | Leauthaud | Th2 | Mean | −0.0063 ± 0.0084 | 0.0045 ± 0.0056 | 1/1 |
| | Leauthaud | Th2 | Individual | −0.0050 ± 0.0108 | 0.0053 ± 0.0063 | 27/27 |
| | Tinker | Th2 | Mean | −0.0024 ± 0.0090 | 0.0014 ± 0.0060 | 1/1 |
| | Tinker | Th2 | Individual | 0.0078 ± 0.0103 | −0.0017 ± 0.0065 | 27/27 |
| | Hearin | Th2 | Mean | 0.0115 ± 0.0094 | 0.0066 ± 0.0057 | 1/1 |
| | Hearin | Th2 | Individual | 0.0091 ± 0.0097 | 0.0087 ± 0.0054 | 27/27 |
| Fourier Space [Pre-Rec] | Leauthaud | Th2 | Mean | −0.0114 ± 0.0178 | 0.0028 ± 0.0107 | 1/1 |
| | Leauthaud | Th2 | Individual | 0.0000 ± 0.0180 | 0.0110 ± 0.0120 | 27/27 |
| | Tinker | Th2 | Mean | 0.0023 ± 0.0177 | −0.0047 ± 0.0124 | 1/1 |
| | Tinker | Th2 | Individual | −0.0110 ± 0.0160 | 0.0230 ± 0.0110 | 27/27 |
| | Hearin | Th2 | Mean | −0.0021 ± 0.0143 | 0.0108 ± 0.0099 | 1/1 |
| | Hearin | Th2 | Individual | −0.0210 ± 0.0160 | 0.0160 ± 0.0110 | 27/27 |
| Fourier Space [Post-Rec] | Leauthaud | Th2 | Mean | −0.0062 ± 0.0104 | −0.0024 ± 0.0075 | 1/1 |
| | Leauthaud | Th2 | Individual | 0.0020 ± 0.0130 | −0.0093 ± 0.0074 | 27/27 |
| | Tinker | Th2 | Mean | −0.0024 ± 0.0121 | −0.0025 ± 0.0088 | 1/1 |
| | Tinker | Th2 | Individual | 0.0038 ± 0.0097 | −0.0006 ± 0.0072 | 27/27 |
| | Hearin | Th2 | Mean | 0.0002 ± 0.0113 | 0.0122 ± 0.0075 | 1/1 |
| | Hearin | Th2 | Individual | 0.0090 ± 0.0150 | 0.0167 ± 0.0061 | 27/27 |

Techniques in configuration and Fourier space. To appreciate this more clearly, in Figure 15 we report the mean values of the three different techniques (CLPT-GS, CF-TNS, and P$_{\chi}$-TNS) with respect to the individual HOD models of Leauthaud, Tinker, and Hearin averaged over 27 mocks at 'Th2' (filled symbols in the various panels), as well as averages across all the three HOD models (corresponding open symbols). The shaded green areas in the figure highlight the 1% error level for $\alpha_1$ and $\alpha_2$, and the 3% error level in $f\sigma_8$. The plot is the analogues of Figure 13, but now for the RSD FS analysis, and once again it allows one to separate the systematics introduced by the HOD modeling (open symbols) versus the theoretical systematics related to individual RSD FS fitting methodologies (filled symbols). In general, as demonstrated by the plot, the HOD systematics is within the ~1% level, even smaller than the modeling systematics, and always below the statistical precision of the eBOSS LRG sample. The modeling systematics instead could reach the percentage level particularly in $\alpha_1$ and $f\sigma_8$, and could represent a dominant source of systematics. From a more extensive FS analysis, we thus conclude that while the HOD systematics is subdominant to the global systematic error budget, the modeling systematics should be taken into account – although both are below the statistical precision of the eBOSS LRG sample. Moreover, from the FS study, we conclude that the different methodologies adopted for the analysis of the final eBOSS LRG sample are mutually consistent and robust, thus validating the clustering analysis pipelines.

6.5 Galaxy Mock Challenge: Modeling Systematics

The heterogeneous set of Outer Rim mocks previously adopted for assessing possible systematic effects in the galaxy-halo connection and imprecisions related to the galaxy clustering modeling (i.e., impact of HODs on BAO and RSD methods, and RSD modeling systematics) is suboptimal in accuracy at the sub-percent level, although this type of accuracy is well-below the statistical sensitivity of the eBOSS LRG sample. This is mainly because of the relatively small effective volume spanned by each individual independent mock, due to the limitations posed by having only a single Outer Rim halo catalog at $z = 0.695$ combined with the constraints intrinsic to the LRG modeling (see Section 3). For this reason, only 27 mocks were used in the previous analyses, as fully independent realizations (i.e. not sharing the same DM field) are required to properly assess cosmic variance. In terms of errorbars, the resolution limit of the OR mocks is in fact around the ~1–2% level. In order to evaluate the performance of the BAO and RSD modeling at a
Figure 14. Full shape RSD analysis: main results on the Alcock-Paczynski parameters and the growth of structure. The three techniques adopted for the analysis of the final eBOSS LRG sample, two in configuration space (CLPT-GS and CF-TNS) and one in Fourier space ($P_k$-TNS), are confronted on a series of Outer Rim mocks having different HOD prescriptions. From left to right, the Leauthaud, Tinker, and Hearin models corresponding to ‘Th2’ are analyzed, as they are closer to the characteristics of the eBOSS LRG sample. Individual fits on each of the 27 realizations per model are performed, as well as fits on the mean of the mocks, allowing one to obtain accurate estimates of $\alpha_\parallel$, $\alpha_\perp$, and $f_{\sigma_8}$. Scatter plots in terms of $\sigma$-deviations are also shown. The corresponding numerical results are reported in Table 9. Errorbars follow the same conventions as in Figure 12.

sub-percent level, a more suitable choice is to abandon a single simulation – although of exquisite mass-resolution such as the Outer Rim – and opt instead for multiple realizations of the same box (i.e., identical initial conditions in all but the random seeds) at a lower mass-resolution and with a larger effective volume.\(^6\) This is the logic beyond the Nseries, a small homogeneous set of 84 pseudo-independent mocks constructed from 7 independent periodic boxes of $2.6h^{-1}\text{Gpc}$ side, projected through 12 different orientations and cuts per box; the mass resolution of these periodic boxes is much lower than that of the Outer Rim run, but it is still sufficient for resolving LRG-type halos ($1.5 \times 10^{11} h^{-1} M_\odot$, with $2048^3$ particles per box). The global effective volume spanned is $84 \times 3.67 \text{[Gpc]}^3$. The Nseries are characterized by the same underlying galaxy bias model built upon the same cosmology, but each mock is a quasi-independent realization – thus not sharing exactly the same LSS – and including observational artifacts closer to the eBOSS DR16 sample, with similar angular and radial selection function of the observed sample. The HOD used is targeted to BOSS CMASS galaxies, at an effective redshift of $\bar{z} = 0.56$. Although this set was originally devised for BOSS galaxies, it is still useful for evaluating modeling systematics at the sub-percentage level also for eBOSS tracers. This is why the Nseries is extensively used in Bautista et al. (2020) and in Gil-Marin et al. (2020) for addressing the modeling systematics related to each complementary analysis in configuration and Fourier space, respectively. Here, we show only an interesting combined example, confronting the fitting methodologies in configuration and Fourier space and the performance of the RSD models previously introduced in Section 5.

\(^6\) Another alternative would be to pursue instead a subhalo-type modeling approach, rather than the more traditional HOD framework, but we do not have access to full merger trees from the Outer Rim simulation.
Figure 15. HOD versus modeling systematics, from a RSD FS analysis in configuration and Fourier space. Filled symbols in the various panels are mean values obtained by fitting the mean of 27 Outer Rim mocks at ‘Th2’ for three different HOD prescriptions (Leauthaud, Tinker, and Hearin frameworks – from left to right, respectively, see Table 9), using three RSD models (CLPT-GS, CF-TNS, and Pk-TNS). Corresponding open symbols are averages across all the three HOD recipes. The shaded green areas in the figure represent the 1% error level for $\alpha_\parallel$ and $\alpha_\perp$, and the 3% error level for $f_{\sigma_8}$. See the main text for more details.

Figure 16 provides a summary result obtained by running the various BAO and RSD FS analysis pipelines on the average of 84 Nseries mocks. Specifically, the left panels display the Alcock-Paczynski parameters derived from a BAO-only fit in configuration and Fourier space, respectively. Filled symbols are used for pre-reconstructed catalogs, while open symbols refer to post-reconstructed catalogs. The left panels show analogous quantities, as well as the growth of structure in terms of $f_{\sigma_8}$, derived from RSD FS fits. In this case the analysis is performed only over pre-reconstructed catalogs, and carried out in the Nseries cosmology (see Section 4.2). Results from the three different RSD models – two in configuration space (CLPT-GS and CF-TNS) and one in Fourier space ($P_k$-TNS), all set for optimal performance as explained before, including the hexadecapole – are displayed with filled symbols. The corresponding numerical values are reported in Table 10. The gray areas in the figure highlight the 0.5% error level for the Alcock-Paczynski parameters, and the 1.0% error level for $f_{\sigma_8}$. As clearly seen, all the different methods are mutually consistent, showing a remarkable accuracy in recovering the expected cosmological parameters ($\alpha_\parallel, \alpha_\perp, f_{\sigma_8}$) at an exquisite level of precision, within at worst 0.9% of their expected values for the $\alpha$’s and within 1.5% for $f_{\sigma_8}$ – as a conservative estimate. These results can be compared with analogous measurements performed on the Outer Rim mocks displayed in Figures 13 and 15. Although here at a sub-percentage level precision, results from the two different sets of mocks are consistent: we observe a similar trend at a fixed HOD recipe, indicating an impressive level of consistency between techniques in configuration and Fourier space. Clearly, the modeling systematics is addressed here with higher accuracy, showing deviations at the sub-percent level for the Alcock-Paczynski parameters and $f_{\sigma_8}$. While this type of systematics may be a dominant source of error in the global systematic budget (despite sub-percent deviations), the LRG sample is primarily dominated by the statistical error of the data.
Table 9. Performance of the RSD FS methods evaluated on the OR-based mocks, for the ‘Th2’ flavor with different HOD models. Mocks are analyzed in their own Outer Rim cosmology, so the expected values are 1 for the $\alpha$ parameters and $f\sigma_8^{\exp} = 0.447$. For each set of mocks, the results from pre-reconstructed catalogs are presented. We report both the results of fitting the mean of all the 27 realizations per HOD, indicated with ‘Mean’, and the mean of individual fits on the mocks, indicated as ‘Individual’. For the fit to the mean, the error quoted is the $\sigma_r$ of the error on the fit, where the covariances are scaled by the 27 realizations. For the mean of individual best-fits, the error quoted is the rms divided by the number of realizations. These results are visualized in Figures 14 and 15.

| RSD Analysis Type     | HOD Type | HOD Flavor | Analysis Details | $\alpha_\parallel - \alpha_\parallel^{\exp}$ | $\alpha_\perp - \alpha_\perp^{\exp}$ | $f\sigma_8 - f\sigma_8^{\exp}$ | $N_{\text{det}}/N_{\text{tot}}$ |
|-----------------------|----------|------------|------------------|------------------------------------------|----------------------------------|---------------------------------|-------------------------------|
| Configuration Space   | Leauthaud| Th2        | Mean             | $-0.0189 \pm 0.0114$                      | $0.0010 \pm 0.0067$              | $0.0115 \pm 0.0156$              | $1/1$                         |
| [Pre-Rec]             | Leauthaud| Th2        | Individual       | $-0.0069 \pm 0.0159$                      | $-0.0052 \pm 0.0092$             | $0.0119 \pm 0.0209$              | $27/27$                       |
| CLPT-GS               | Tinker   | Th2        | Mean             | $-0.0211 \pm 0.0117$                      | $0.0022 \pm 0.0077$              | $0.0066 \pm 0.0178$              | $1/1$                         |
|                       | Tinker   | Th2        | Individual       | $-0.0096 \pm 0.0035$                      | $-0.0014 \pm 0.0093$             | $0.0193 \pm 0.0193$              | $27/27$                       |
|                       | Hearin   | Th2        | Mean             | $-0.0158 \pm 0.0109$                      | $0.0008 \pm 0.0067$              | $0.0187 \pm 0.0178$              | $1/1$                         |
|                       | Hearin   | Th2        | Individual       | $-0.0053 \pm 0.0176$                      | $-0.0065 \pm 0.0089$             | $0.0267 \pm 0.0250$              | $27/27$                       |
| Configuration Space   | Leauthaud| Th2        | Mean             | $-0.0012 \pm 0.0110$                      | $-0.0046 \pm 0.0052$             | $0.0260 \pm 0.0146$              | $1/1$                         |
| [Pre-Rec]             | Leauthaud| Th2        | Individual       | $0.0259 \pm 0.0122$                       | $-0.0001 \pm 0.0069$             | $0.0305 \pm 0.0168$              | $27/27$                       |
| CF-TNS                | Tinker   | Th2        | Mean             | $0.0144 \pm 0.0102$                       | $0.0014 \pm 0.0059$              | $0.0130 \pm 0.0140$              | $1/1$                         |
|                       | Tinker   | Th2        | Individual       | $0.0375 \pm 0.0126$                       | $-0.0049 \pm 0.0076$             | $0.0002 \pm 0.0157$              | $27/27$                       |
|                       | Hearin   | Th2        | Mean             | $0.0056 \pm 0.0098$                       | $0.0014 \pm 0.0062$              | $0.0227 \pm 0.0152$              | $1/1$                         |
|                       | Hearin   | Th2        | Individual       | $0.0291 \pm 0.0127$                       | $-0.0022 \pm 0.0077$             | $0.0205 \pm 0.0168$              | $27/27$                       |
| Fourier Space         | Leauthaud| Th2        | Mean             | $0.0034 \pm 0.0038$                       | $-0.0111 \pm 0.0094$             | $-0.0040 \pm 0.0200$             | $1/1$                         |
| [Pre-Rec]             | Leauthaud| Th2        | Individual       | $0.0610 \pm 0.0140$                       | $-0.0195 \pm 0.0087$             | $0.0060 \pm 0.0160$              | $27/27$                       |
| $P_k$-TNS             | Tinker   | Th2        | Mean             | $0.0060 \pm 0.0144$                       | $-0.0177 \pm 0.0107$             | $-0.0070 \pm 0.0210$             | $1/1$                         |
|                       | Tinker   | Th2        | Individual       | $0.0970 \pm 0.0240$                       | $-0.0047 \pm 0.0097$             | $0.0140 \pm 0.0220$              | $27/27$                       |
|                       | Hearin   | Th2        | Mean             | $-0.0104 \pm 0.0129$                      | $-0.0020 \pm 0.0089$             | $0.0190 \pm 0.0190$              | $1/1$                         |
|                       | Hearin   | Th2        | Individual       | $0.0450 \pm 0.0170$                       | $0.0001 \pm 0.0093$              | $0.0260 \pm 0.0220$              | $27/27$                       |

Table 10. Modeling systematics related to BAO and RSD methodologies, addressed with the Nieres. These numerical results are shown in Figure 16.

| Analysis Type         | Analysis Space | Analysis Method | Fitting Model | $\alpha_\parallel - \alpha_\parallel^{\exp}$ | $\alpha_\perp - \alpha_\perp^{\exp}$ | $f\sigma_8 - f\sigma_8^{\exp}$ |
|-----------------------|----------------|-----------------|---------------|------------------------------------------|----------------------------------|---------------------------------|
| BAO [Pre-Rec]         | Configuration Space | CP – BAO Peak | BAO Template | $0.0014 \pm 0.0045$                      | $0.0059 \pm 0.0023$              | –                               |
| BAO [Pre-Rec]         | Fourier Space   | PS – BAO Peaks | BAO Template | $-0.0045 \pm 0.0041$                      | $-0.0021 \pm 0.0020$             | –                               |
| BAO [Post-Rec]        | Configuration Space | CP – BAO Peak | BAO Template | $0.0031 \pm 0.0024$                      | $0.0023 \pm 0.0015$              | –                               |
| BAO [Post-Rec]        | Fourier Space   | PS – BAO Peaks | BAO Template | $-0.0048 \pm 0.0019$                      | $0.0005 \pm 0.0010$              | –                               |
| RSD [Pre-Rec]         | Configuration Space | CP – Full Shape | CLPT-GS      | $-0.0090 \pm 0.0030$                      | $0.0020 \pm 0.0020$              | $-0.0060 \pm 0.0050$            |
| RSD [Pre-Rec]         | Configuration Space | CP – Full Shape | CF-TNS       | $-0.0050 \pm 0.0030$                      | $-0.0020 \pm 0.0020$             | $0.0060 \pm 0.0040$             |
| RSD [Pre-Rec]         | Fourier Space   | PS – Full Shape | $P_k$-TNS    | $0.0016 \pm 0.0032$                       | $-0.0095 \pm 0.0020$             | $-0.0038 \pm 0.0041$            |

7 SYSTEMATIC ERROR BUDGET

Finally, we address here the LRG global error budget with a major focus on theoretical systematics, and also summarize the previous mock challenge results in term of biases in the estimation of $\alpha_\parallel$, $\alpha_\perp$, and $f\sigma_8$.

7.1 Global Error Budget

In our companion papers Bautista et al. (2020) and Gil-Marin et al. (2020), besides modeling and HOD imperfections, detailed investigations regarding the impact of a fiducial cosmology, the optimal fitting range of scales, effects on non-periodicity, and observational artifacts such as redshift failures, completeness, close-pair collisions, and radial integral constraint (de Mattia et al. 2020) are carried out in configuration and Fourier space, respectively, and the associated errors are carefully quantified using all the available types of mocks. In the following, we indicate the contribution of all these additional systematics as $\sigma_{\text{other}}^\text{sys}$, while we use $\sigma_{\text{model}}^\text{sys}$ for denoting the theoretical systematics ascribed to imperfections in the RSD modeling. Adopting similar conventions as in the companion papers, for a given cosmological parameter $x_p$ measured with error $\sigma_p$ whose reference value is $x_p^\text{ref}$, the systematic error assigned is:

$$\sigma_{p,\text{sys}} = 2\sigma_p \quad \text{if} \quad |x_p - x_p^\text{ref}| < 2\sigma_p;$$

(35)

$$\sigma_{p,\text{sys}} = |x_p - x_p^\text{ref}| \quad \text{if} \quad |x_p - x_p^\text{ref}| \geq 2\sigma_p.$$  

(36)

In essence, anything above the $2\sigma$ level is considered as a detected systematics (corresponding to a 95% confidence level
on the mean of the mocks), and the maximal value is always used as a conservative choice. The statistical properties of the LRG sample are also characterized in Bautista et al. (2020) and in Gil-Marín et al. (2020), and the consensus statistical error related to each individual method is denoted here as $\sigma_{\text{stat}}$. Recall again that from a joint BAO and RSD FS analysis, both in configuration and Fourier space, the statistical consensus errors are 1.9% on $\alpha_\perp$ and 2.6% on $\alpha_\parallel$, respectively. In the subsequent analysis, we always consider fits to the mean, as they are less sensitive to noise effects compared to individual fits, and only focus on RSD FS results.

Table 11 summarizes the global error budget for the eBOSS DR16 LRG sample. Here, the modeling systematics is inferred from the $N$SERIES (see Table 10) and indicated as $\sigma_{\text{model, NS}}$: we explain later on the reason behind this choice, and why the Outer Rim contribution is not included here. The comprehensive systematic error budget intrinsic to each RSD method ($\sigma_{\text{syst}}$) is simply obtained by summing in quadrature the modeling and additional systematics, and the total error budget $\sigma_{\text{tot}}$ is also derived in quadrature from the contributions of $\sigma_{\text{syst}}$ and $\sigma_{\text{stat}}$. In the table, we provide some useful ratios as well, that allow one to directly compare the contribution of systematics or statistics to the total error estimate. While the BAO-only pipeline is essentially unbiased, from the RSD FS analyses we conclude that systematic errors account for a significant fraction of the total error budget, contributing up to 50% (or more) to the uncertainties associated with the Alcock-Paczynski parameters and the growth of structures (see the various ratios). The impact in

![Figure 16. Comparing modeling systematics in BAO and RSD methods, estimated from 84 $N$SERIES mocks. Left panels show the Alcock-Paczynski parameters derived from BAO-only fits in configuration and Fourier space, respectively. Right panels refer to RSD full shape analyses. Filled symbols are used for pre-reconstructed catalogs, while open symbols refer to post-reconstructed catalogs. Gray areas in the figure highlight the 0.5% error level on $\alpha$ and $\alpha_\perp$, and the 1.0% error on $f\alpha_\perp$. These numerical results are reported in Table 10. All the different methods adopted for the clustering analysis of the eBOSS LRG sample are mutually consistent, showing a remarkable accuracy in recovering the expected cosmological parameters at an exquisite level of precision.](image)

| RSD-FS Analysis | Global Syst. |
|-----------------|--------------|
| Error Type      | Model        | $\sigma_{\alpha}$ | $\sigma_{\alpha_\perp}$ | $\sigma_{\text{tot}}$ |
| RSD Modeling    | CLPT-GS      | 0.0090            | 0.0040             | 0.0100             |
| $[\sigma_{\text{model,NS}}]$ | CLPT-GS | 0.0060            | 0.0040             | 0.0080             |
| RSD Additional  | CLPT-GS      | 0.0156            | 0.0127             | 0.0220             |
| $[\sigma_{\text{other}}]$ | CLPT-GS | 0.0153            | 0.0112             | 0.0216             |
| RSD Systematics | CLPT-GS      | 0.0180            | 0.0133             | 0.0242             |
| $[\sigma_{\text{stat}}]$ | CLPT-GS | 0.0164            | 0.0119             | 0.0230             |
| RSD Total       | CLPT-GS      | 0.0280            | 0.0200             | 0.0450             |
| $[\sigma_{\text{tot}}]$ | CLPT-GS | 0.0310            | 0.0180             | 0.0400             |
| RSD Statistical | CLPT-GS      | 0.0280            | 0.0200             | 0.0450             |
| $[\sigma_{\text{stat}}]$ | CLPT-GS | 0.0360            | 0.0270             | 0.0420             |
| RSD Total       | CLPT-GS      | 0.0333            | 0.0240             | 0.0511             |
| $[\sigma_{\text{tot}}]$ | CLPT-GS | 0.0351            | 0.0216             | 0.0462             |
| $\sigma_{\text{model,NS}}/\sigma_{\text{stat}}$ | CLPT-GS | 0.3214            | 0.2900             | 0.2222             |
| $\sigma_{\text{other}}/\sigma_{\text{stat}}$ | CLPT-GS | 0.4935            | 0.4222             | 0.2000             |
| $\sigma_{\text{sys}}/\sigma_{\text{tot}}$ | CLPT-GS | 0.5571            | 0.6350             | 0.4889             |
| $\sigma_{\text{model,NS}}/\sigma_{\text{tot}}$ | CLPT-GS | 0.3250            | 0.2518             | 0.3960             |
| $\sigma_{\text{other}}/\sigma_{\text{tot}}$ | CLPT-GS | 0.6432            | 0.6658             | 0.5370             |
| $\sigma_{\text{sys}}/\sigma_{\text{tot}}$ | CLPT-GS | 0.5301            | 0.6607             | 0.5758             |
| $\sigma_{\text{model,NS}}/\sigma_{\text{tot}}$ | CLPT-GS | 0.3704            | 0.4327             | 0.4175             |
| $\sigma_{\text{other}}/\sigma_{\text{tot}}$ | CLPT-GS | 0.4636            | 0.5191             | 0.4679             |
| $\sigma_{\text{sys}}/\sigma_{\text{tot}}$ | CLPT-GS | 0.3048            | 0.2311             | 0.3406             |
| $\sigma_{\text{model,NS}}/\sigma_{\text{tot}}$ | CLPT-GS | 0.2703            | 0.1854             | 0.1733             |
| $\sigma_{\text{other}}/\sigma_{\text{tot}}$ | CLPT-GS | 0.1710            | 0.1854             | 0.1733             |
| $\sigma_{\text{sys}}/\sigma_{\text{tot}}$ | CLPT-GS | 0.1667            | 0.3229             | 0.1802             |
| $\sigma_{\text{model,NS}}/\sigma_{\text{tot}}$ | CLPT-GS | 0.4686            | 0.5286             | 0.4307             |
| $\sigma_{\text{other}}/\sigma_{\text{tot}}$ | CLPT-GS | 0.4361            | 0.5191             | 0.4679             |
| $\sigma_{\text{sys}}/\sigma_{\text{tot}}$ | CLPT-GS | 0.3948            | 0.2711             | 0.3406             |
| $\sigma_{\text{model,NS}}/\sigma_{\text{tot}}$ | CLPT-GS | 0.5410            | 0.5542             | 0.4731             |
| $\sigma_{\text{other}}/\sigma_{\text{tot}}$ | CLPT-GS | 0.4684            | 0.5513             | 0.4990             |
| $\sigma_{\text{sys}}/\sigma_{\text{tot}}$ | CLPT-GS | 0.3474            | 0.3971             | 0.3853             |
| $\sigma_{\text{model,NS}}/\sigma_{\text{tot}}$ | CLPT-GS | 0.8410            | 0.8324             | 0.8810             |
| $\sigma_{\text{other}}/\sigma_{\text{tot}}$ | CLPT-GS | 0.8835            | 0.8343             | 0.8666             |
| $\sigma_{\text{sys}}/\sigma_{\text{tot}}$ | CLPT-GS | 0.9378            | 0.9178             | 0.9228             |
Table 12. Modeling systematics derived from RSD FS analyses of the ‘Th2’ Outer Rim challenge mocks (see Tables 4 and 5).

| RSD-FS Analysis | Model Syst. | $\sigma_{\alpha_1}$ | $\sigma_{\alpha_2}$ | $\sigma_{\text{f}_{\text{RSD}}}$ |
|-----------------|-------------|---------------------|---------------------|-----------------|
| RSD Modeling    | CLPT-GS     | 0.0228              | 0.0134              | 0.0330          |
| $[\sigma_{\text{model,OR}}]_{\text{sys, LH}}$ | CF-TNS     | 0.0220              | 0.0104              | 0.0292          |
|                 | $P_\delta$-TNS | 0.0276              | 0.0188              | 0.0400          |
| RSD Modeling    | CLPT-GS     | 0.0234              | 0.0154              | 0.0356          |
| $[\sigma_{\text{model,OR}}]_{\text{sys, TK}}$ | CF-TNS     | 0.0204              | 0.0118              | 0.0280          |
|                 | $P_\delta$-TNS | 0.0288              | 0.0214              | 0.0420          |
| RSD Modeling    | CLPT-GS     | 0.0218              | 0.0134              | 0.0356          |
| $[\sigma_{\text{model,OR}}]_{\text{sys, HE}}$ | CF-TNS     | 0.0196              | 0.0124              | 0.0304          |
|                 | $P_\delta$-TNS | 0.0258              | 0.0178              | 0.0380          |
| RSD Modeling    | CLPT-GS     | 0.0227              | 0.0141              | 0.0347          |
| $[\sigma_{\text{model,OR}}]_{\text{sys}}$ | CF-TNS     | 0.0207              | 0.0115              | 0.0292          |
|                 | $P_\delta$-TNS | 0.0274              | 0.0193              | 0.0400          |
| $\sigma_{\text{model,OR}}_{\text{sys, LH}} / \sigma_{\text{stat}}$ | CLPT-GS     | 0.8143              | 0.6700              | 0.7334          |
|                 | CF-TNS     | 0.7097              | 0.5778              | 0.7300          |
|                 | $P_\delta$-TNS | 0.7667              | 0.6963              | 0.9524          |
| $\sigma_{\text{model,OR}}_{\text{sys, TK}} / \sigma_{\text{stat}}$ | CLPT-GS     | 0.8357              | 0.7700              | 0.7911          |
|                 | CF-TNS     | 0.6581              | 0.6556              | 0.7000          |
|                 | $P_\delta$-TNS | 0.8000              | 0.7926              | 1.0000          |
| $\sigma_{\text{model,OR}}_{\text{sys, HE}} / \sigma_{\text{stat}}$ | CLPT-GS     | 0.7786              | 0.6700              | 0.7911          |
|                 | CF-TNS     | 0.6323              | 0.6889              | 0.7600          |
|                 | $P_\delta$-TNS | 0.7167              | 0.6593              | 0.9048          |
| $\sigma_{\text{model,OR}}_{\text{sys}} / \sigma_{\text{stat}}$ | CLPT-GS     | 0.8095              | 0.7033              | 0.7718          |
|                 | CF-TNS     | 0.6667              | 0.6407              | 0.7300          |
|                 | $P_\delta$-TNS | 0.7611              | 0.7160              | 0.9524          |
| $\sigma_{\text{model,OR}}_{\text{sys, LH}} / \sigma_{\text{model, OR}}_{\text{sys}}$ | CLPT-GS     | 2.5334              | 3.3500              | 3.3000          |
|                 | CF-TNS     | 3.6667              | 2.6000              | 3.6500          |
|                 | $P_\delta$-TNS | 4.3125              | 1.9789              | 4.8781          |
| $\sigma_{\text{model,OR}}_{\text{sys, TK}} / \sigma_{\text{model, OR}}_{\text{sys}}$ | CLPT-GS     | 2.6000              | 3.8000              | 3.5600          |
|                 | CF-TNS     | 3.4000              | 2.9500              | 3.5000          |
|                 | $P_\delta$-TNS | 4.5000              | 2.2526              | 5.1219          |
| $\sigma_{\text{model,OR}}_{\text{sys, HE}} / \sigma_{\text{model, OR}}_{\text{sys}}$ | CLPT-GS     | 2.4222              | 3.3500              | 3.5600          |
|                 | CF-TNS     | 3.2667              | 3.1000              | 3.8000          |
|                 | $P_\delta$-TNS | 4.0312              | 1.8737              | 4.6341          |
| $\sigma_{\text{model,OR}}_{\text{sys, LH}} / \sigma_{\text{model, OR}}_{\text{sys}}$ | CLPT-GS     | 2.5185              | 3.5167              | 3.4733          |
|                 | CF-TNS     | 3.4445              | 2.8834              | 3.6500          |
|                 | $P_\delta$-TNS | 3.2812              | 2.0351              | 4.8780          |

The determination of $\alpha_1$ and $\alpha_2$ is at the $\sim 1.0\%$ level, and it can reach even $\sim 1.5\%$–$2.0\%$ for $f_{\text{RSD}}$. From configuration space analyses, the most relevant contribution to systematics is caused by observational artifacts. In Fourier space, the most dominant systematic is arising from the assumption of a reference cosmology, that can bias in particular the estimation of $f_{\text{RSD}}$ up to $2.0\%$. All of the other effects, including modeling systematics, are within the $1.0\%$ range or below. Eventually, systematic errors are added only to the diagonal of the covariance of each measurement, assuming that all the contributions to systematics are independent.

### 7.2 Impact of Modeling Systematics

The modeling systematics estimated from ‘Th2’ Outer Rim mocks analyzed in Section 6 are detailed in Table 12, where we list all the contributions inferred from the individual HODs of Leauthaud (LH), Tinker (TK), and Hearin (HE), respectively, as well as the combined theoretical systematics derived by simply averaging those contributions ($\sigma_{\text{model,OR}}$). We also report some useful ratios, for the ease of comparison. Not surprisingly, the modeling systematics obtained from Outer Rim mocks are much larger than those derived from the NSERIES. The reason is related to the difference in effective volume, combined with the limited number of fully independent realizations available (27 synthetic catalogs per flavor). In fact, the global effective volume of ‘Th2’ Outer Rim mocks is $29.7\ Gpc^3$, about 11 times bigger than the combined CMAS plus eBOSS LRG sample, but $\sim 10.27\ times$ smaller than the global effective volume spanned by the NSERIES – which is $308.28\ Gpc^3$, and thus 113 times larger than the combined DR16 LRG sample. In this respect, the statistical threshold of the NSERIES is at the $0.1\%–0.5\%$ level, while the resolution of Outer Rim mocks is around $1.0\%–1.5\%$. As evident from Table 12, the modeling systematics inferred from Outer Rim mocks is closer to the statistical error of the LRG sample (see the various ratios $\sigma_{\text{sys, LH}} / \sigma_{\text{stat}}$), and a factor $\sim 2–5$ times bigger than uncertainties derived from the NSERIES. The size of the errorbars simply scales with the global effective volume and the number of available realizations, as clearly highlighted by the various ratios $\sigma_{\text{model,OR}} / \sigma_{\text{stat}}$, $\sigma_{\text{sys, TK}} / \sigma_{\text{stat}}$, $\sigma_{\text{sys, HE}} / \sigma_{\text{stat}}$, and $\sigma_{\text{model,OR}} / \sigma_{\text{model,NS}}$.

To this end, we have constructed new heterogeneous -body data challenge with the aim of testing and validating the robustness of the LRG clustering pipelines of Bautista et al. (2020) in configuration space, and of Gil-Marín et al. (2020) in Fourier space. We have also quantified the theoretical systematics related to BAO and RSD fitting methodologies, and the bias intrinsic to the modeling of the galaxy-halo connection.

To this end, we have constructed new heterogeneous galaxy mocks from the Outer Rim simulation spanning different redshift intervals, using a variety of HOD schemes of

### 8 CONCLUSIONS AND OUTLOOK

In support of the final analysis of the eBOSS DR16 galaxy sample, we have carried out an extensive N-body data challenge with the aim of testing and validating the robustness of the LRG clustering pipelines of Bautista et al. (2020) in configuration space, and of Gil-Marín et al. (2020) in Fourier space. We have also quantified the theoretical systematics related to BAO and RSD fitting methodologies, and the bias intrinsic to the modeling of the galaxy-halo connection.

To this end, we have constructed new heterogeneous galaxy mocks from the Outer Rim simulation spanning different redshift intervals, using a variety of HOD schemes of
increasing complexity. The theoretical foundation for modeling the galaxy-halo connection is laid out in Section 3, and the mock-making procedure is explained in detail in Section 4. Moving from the most conventional HOD approach, we have considered more sophisticated scenarios able to distinguish between quiescent or star-forming galaxies, and with the inclusion of assembly bias that generalize further the standard HOD framework. Our Outer Rim-based mocks cover a range of number densities and effective volumes, and are well-suited for a variety of studies. In this work, we have mainly focused on a subset at $z \approx 0.69$, with characteristics closer to the eBOSS LRG sample (i.e., ‘Th2’ flavor with the Leauthaud, Tinker, and Hearin prescriptions). We have also briefly exploited a small homogeneous synthetic set (the Nseries), which has been previously used in the SDSS DR12 galaxy clustering analysis and is more suitable to assess theoretical systematics at the sub-percent level, thanks to a larger effective volume.

In our challenge, detailed in Section 6, we have tested the performance of BAO and RSD fitting techniques against different galaxy population schemes and bias models having analogous clustering properties, with the main objective of validation and calibration of such methods and the quantification of theoretical systematics. The mock products have allowed us to confront on a common ground and assess the performance of the BAO fitting methodologies for the LRG sample, and of three complementary RSD models in configuration and Fourier space – denoted as CLPT-GS, CF-TNS, and $P_{\perp}$-TNS, respectively. Overall, we have found a remarkable agreement at the sub-percent level between different techniques in configuration and Fourier space (see in particular Figures 13, 15 and 16), along with an impressive level of consistency among BAO fitting and reconstruction procedures and from all the RSD models used in FS analyses. All of the methods performed equally well, with comparable errors on the Alcock-Paczynski parameters and the growth of structure. Moreover, reconstruction significantly improved the constraints on both $\sigma_8$ and $\sigma_z$. We have thus validated the robustness of the LRG clustering analysis pipelines.

Regarding systematics and the global error budget (Section 7), we have found that the impact of different HOD prescriptions is always sub-dominant to the total systematics, and that modeling systematics in the estimation of $\sigma_8$ and $f\sigma_8$, although at worst around $\sim 1.5\%$, may be a dominant source of error in the comprehensive quantification of systematics. In particular, from the analysis in configuration space of pre-reconstructed mocks (considering only fits to the mean), biases in the recovered $\alpha$ values reach up to 0.5% in $\sigma_2$ and 1.0% in $\sigma_{\parallel}$. After reconstruction, there is a reduction of the biases to less than 0.2%, hence the BAO analysis is unbiased. For RSD analyses in configuration space, the most significant contribution to systematic errors arises from observational effects. From the Fourier space methodology, for the post-reconstruction BAO analysis we detected a 0.5% systematic shift induced by modeling systematic on $\sigma_8$, and none for $\sigma_{\parallel}$, with a resolution limit of 0.2% for the Nseries mocks. The systematic shift is of order 1.5% for the FS analysis from Outer Rim mocks instead. Moreover, we did not detect any significant relative shift on the cosmological parameters when either the HOD model or the flavor is varied. Such results put constrains in the upper limit of systematic errors in the modeling, as a result of different HODs, with upper limits of order $0.5 - 1.1\%$ systematic shifts. In any case, both HOD and modeling systematics are below the statistical error of the eBOSS LRG data. The expected statistical errors in the eBOSS LRG data sample are in fact of the order of $\sim 1.9\%$ for $\sigma_8$, and $\sim 2.6\%$ for $\sigma_{\parallel}$. Eventually, these systematic corrections in the Alcock-Paczynski parameters and the growth of structure are combined with additional sources of systematics (Table 11), and such errors are accounted for in the final consensus results (Collaboration et al. 2020) from the analysis of the LRG DR16 galaxy sample.

In addition, we were also able to gain interesting insights on the galaxy-halo connection. Thus, our work may be useful for future applications within the HOD framework. Finally, our analysis provides a global and complementary perspective of the systematic studies carried out in Bautista et al. (2020) in configuration space, and in Gil-Marín et al. (2020) in Fourier space: their overall agreement at such level of precision is remarkable.

Quantifying the modeling systematics in BAO clustering estimators and in RSD methods for all the eBOSS tracers, as well as characterizing the robustness of the analysis pipelines, are essential tasks in order to obtain unbiased cosmological parameters, accurate $f\sigma_8$ constraints, and reliable consensus likelihoods. In this respect, besides being relevant for the final eBOSS DR16 ‘consensus cosmology’ – as the systematic error budget is informed by testing the results of analyses against these high-resolution mocks – our study represents also a testbed for future large-volume surveys. In particular, similar mock-making techniques and systematic corrections can be readily extended to model for instance the DESI galaxy sample, and we expect that more extensive mock challenges along these lines will be necessary and progressively relevant in the next few years. In fact, mock challenges designed to validate data analysis pipelines and assess the impact of systematics in massive datasets are becoming increasingly important for large-volume surveys – see for example the recent works by MacCraith et al. (2018) for the Dark Energy Survey (DES; The Dark Energy Survey Collaboration 2005), and by Sánchez et al. (2020) for LSST (Ivezic et al. 2019). In this view, while the sub-percent level theoretical systematic corrections quantified in this study may not be relevant for the current state-of-the-art (as they are always inferior to the statistical precision of the data), soon they will become relevant for DESI and LSST, that are expected to achieve sub-percent statistical precision on the galaxy sample; for such surveys, it will be crucial to control the systematics at an extremely low level. In addition, our flexible and highly modular pipeline for building complex HODs offers several directions of extension, as well as applications that go beyond the modeling of LRGs – toward more elaborated galaxy-halo connection physics, particularly in relation to ELGs.

DATA AVAILABILITY
All of the SDSS mock products developed in this study, listed in Table 4, are stored at the National Energy Research Scientific Computing Center (NERSC) and are available upon request. The Outer Rim halo catalogs used to
produce the Outer Rim-based mocks are publicly available at https://cosmology.alcf.anl.gov.

ACKNOWLEDGEMENTS

G.R., P.D.C., and J.M. acknowledge support from the National Research Foundation of Korea (NRF) through Grants No. 2017R1E1A101077508 and No. 2020R1A2C1005655 funded by the Korean Ministry of Education, Science and Technology (MoEST), and from the faculty research fund of Sejong University. Some numerical work performed in this paper was carried out using the Korea Institute of Science and Technology Information (KISTI) supercomputing infrastructure under allocations KSC-2018-G3-0008 and KSC-2018-T1-0061, and with computing resources (Xeon Silver 4114 master node and Xeon Gold 6126 computing node architecture) at Sejong University. HGM acknowledges support from la Caixa Foundation (ID 100010434) which code LCF/BQ/PI18/11630024.

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 web site 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, the Korean Participation Group, Lawrence Berkeley National 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 Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatário 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.

In addition, this research relied on resources provided to the eBOSS Collaboration by the National Energy Research Scientific Computing Center (NERSC). NERSC is a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231.

REFERENCES

Alam S., et al., 2017, MNRAS, 470, 2617
Alam S., et al., 2020, in preparation
Alcock C., Paczynski B., 1979, Nature, 281, 358
Alpaslan M., Tinker J. L., 2019, arXiv e-prints, p. arXiv:1911.04509
Assassi V., Simonović A. M., Zaldarriaga M., 2017, Journal of Cosmology and Astroparticle Physics, 11, 054
Ata M., et al., 2018, MNRAS, 473, 4773
Aubourg E., et al., 2015, Phys. Rev. D, 92, 123516
Avila S., et al., 2020, in preparation
Baldauf T., Seljak U., Desjacques V., McDonald P., 2012, Physical Review D, 86, 083540
Bautista J. E., et al., 2018, ApJ, 863, 110
Bautista J., et al., 2020, submitted
Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379
Bel J., Pezzotta A., Carbone C., Sefusatti E., Guzzo L., 2019, A&A, 622, A109
Belínd A. A., Weinberg D. H., 2002, ApJ, 575, 587
Beutler F., et al., 2014, MNRAS, 443, 1065
Beutler F., et al., 2017a, MNRAS, 464, 3409
Beutler F., et al., 2017b, MNRAS, 466, 2242
Blanton M. R., Belínd A. A., 2007, ApJ, 664, 791
Blanton M. R., et al., 2017, AJ, 154, 28
Blomqvist M., et al., 2019, Astronomy and Astrophysics, 629, A86
Bolton A., Percival W. J., Manera M., Vargas Magana M., Ho S., 2014, Monthly Notices of the Royal Astronomical Society, 445, 3152
Bolton A., Percival W. J., Howlett C., 2015, Monthly Notices of the Royal Astronomical Society, 453, 456
Carlson J., Reid B., White M., 2013, MNRAS, 429, 1674
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
Colless M., et al., 2003, arXiv e-prints, pp astro-ph/0306581
Comparat J., et al., 2016, Astronomy and Astrophysics, 592, A121
DESI Collaboration et al., 2016, arXiv e-prints, p. arXiv:1611.00036
Dawson K. S., et al., 2016, AJ, 151, 44
Dutton A. A., Macciò A. V., 2014, MNRAS, 441, 3359
Eisenstein D. J., et al., 2005, ApJ, 633, 560
Eisenstein D. J., et al., 2012, The Astronomical Journal, 142, 72
Gil-Marín H., Wagner C., Verde L., Porciani C., Jimenez R., 2012, J. Cosmology Astropart. Phys., 2012, 029
Gil-Marín H., et al., 2016, Monthly Notices of the Royal Astronomical Society, 460, 4188
Gil-Marín H., Percival W. J., Verde L., Brownstein J. R., Chuang C.-H., Kitaura F.-S., Rodríguez-Torres S. A., Olmstead M. D., 2017, MNRAS, 465, 1757
Gil-Marín H., et al., 2018, MNRAS, 477, 1604
Gil-Marín H., et al., 2020, submitted
Gunn J. E., et al., 2006, AJ, 131, 2332
Guo H., et al., 2016, ApJ, 831, 3
Guo H., et al., 2019, ApJ, 871, 147
Habib S., et al., 2016, New Astron., 42, 49
Hadzhiyska B., Bose S., Eisenstein D., Hernquist L., Spergel D. N., 2020, MNRAS, 493, 5506
Hearin A. P., Zentner A. R., van den Bosch F. C., Campbell D., 2017, MNRAS, 465, 1757
Hearin A. P., et al., 2017, MNRAS, 477, 1604
Hearin A. P., Zentner A. R., van den Bosch F. C., Campbell D., Tollerud E., 2016, MNRAS, 460, 2552
Hearin A. P., et al., 2017, AJ, 154, 190
Heitmann K., et al., 2019, arXiv e-prints, p. arXiv:1909.07742
Ivezic Z., et al., 2019, ApJ, 873, 111
Kaiser N., 1987, MNRAS, 227, 1
Kamata T., et al., 2011, ApJS, 192, 18
Kravtsov A. V., Belínd A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B. O., Primack J. R., 2004, ApJ, 609, 35
Kravtsov A. V., Nagai D., Vikhlinin A. A., 2005, ApJ, 625, 588

MNRAS 000, 1–35 (2020)
