Model BOSS & eBOSS Luminous Red Galaxies at $0.2 < z < 1.0$ using SubHalo Abundance Matching with 3 parameters


To cite this version:

Jiaxi Yu, Cheng Zhao, Chia-Hsun Chuang, Julian E. Bautista, Ginevra Favole, et al.. Model BOSS & eBOSS Luminous Red Galaxies at $0.2 < z < 1.0$ using SubHalo Abundance Matching with 3 parameters. Monthly Notices of the Royal Astronomical Society, 2022, 516 (1), pp.57-74. 10.1093/mnras/stac2176. hal-03631250

HAL Id: hal-03631250
https://hal.science/hal-03631250
Submitted on 7 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Model BOSS and eBOSS luminous red galaxies at $0.2 < z < 1.0$ using SubHalo Abundance Matching with three parameters

Jiaxi Yu, 1★ Cheng Zhao 1,★ Chia-Hsun Chuang 1,2,★ Julian E. Bautista, 4 Ginevra Favole 1, Jean-Paul Kneib, 1,5 Faizan G. Mohammad, 6 Ashley J. Ross 7, Anand Raichoor 8, Charling Tao, 4 Kyle Dawson 2 and Graziano Rossi 9

1Laboratory of Astrophysics, École Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, CH-1290 Versoix, Switzerland
2Department of Physics and Astronomy, University of Utah, Salt Lake City, UT 84112, USA
3Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 452 Lomita Mall, Stanford, CA 94305, USA
4Aix Marseille University, CNRS/IN2P3, CPPM, F-13007 Marseille, France
5Aix Marseille Université, CNRS, LAM (Laboratoire d’Astrophysique de Marseille) UMR 7326, F-13388 Marseille, France
6Waterloo Centre for Astrophysics, Department of Physics and Astronomy, University of Waterloo, 200 University Ave. W., Waterloo, ON N2L 3G1, Canada
7Department of Astronomy, The Ohio State University, 140 W. 18th Ave., Columbus, OH 43210, USA
8Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
9Department of Astronomy and Space Science, Sejong University, 209, Neungdong-ro, Gwangjin-gu, Seoul, South Korea

Accepted 2022 July 27. Received 2022 July 12; in original form 2022 March 21

ABSTRACT
SubHalo Abundance Matching (SHAM) is an empirical method for constructing galaxy catalogues based on high-resolution N-body simulations. We apply SHAM on the UNIT simulation to simulate SDSS BOSS/eBOSS luminous red galaxies (LRGs) within a wide redshift range of $0.2 < z < 1.0$. Besides the typical SHAM scatter parameter $\sigma$, we include $v_{\text{smear}}$ and $V_{\text{cell}}$ to take into account the redshift uncertainty and the galaxy incompleteness, respectively. These two additional parameters are critical for reproducing the observed 2PCF multipoles on $5 – 25 \, h^{-1} \text{Mpc}$. The redshift uncertainties obtained from the best-fitting $v_{\text{smear}}$ agree with those measured from repeat observations for all SDSS LRGs except for the LOWZ sample. We explore several potential systematics but none of them can explain the discrepancy found in LOWZ. Our explanation is that the LOWZ galaxies might contain another type of galaxies that needs to be treated differently. The evolution of the measured $\sigma$ and $V_{\text{cell}}$ also reveals that the incompleteness of eBOSS galaxies decreases with the redshift. This is the consequence of the magnitude lower limit applied in eBOSS LRG target selection. Our SHAM also set upper limits for the intrinsic scatter of the galaxy–halo relation, given a complete galaxy sample: $\sigma_{\text{int}} < 0.31$ for LOWZ at $0.2 < z < 0.33$, $\sigma_{\text{int}} < 0.36$ for LOWZ at $0.33 < z < 0.43$, and $\sigma_{\text{int}} < 0.46$ for CMaSS at $0.43 < z < 0.51$. The projected 2PCFs of our SHAM galaxies also agree with the observational ones on the 2PCF fitting range.

Key words: methods: numerical – methods: observational – Galaxy: halo – large-scale structure of Universe.

1 INTRODUCTION

Lambda cold dark matter (ACDM) has been the standard cosmological model since the 1990s. Under this framework, observations confirm that dark matter (DM) is the dominant matter component of the Universe (e.g. Planck Collaboration VI 2020). DM particles interact only through gravity and their gravitational evolution is assumed to start from a primordial Gaussian random field with perturbations. If the overdensity arising from the perturbation on small scales is large enough, the infall of matter can overcome the expansion of Universe, and form gravitational-bound clusters, i.e., DM haloes. Perturbations on large scales evolve into web-like structures known as the cosmic web (Bond, Kofman & Pogosyan 1996). Baryonic matter may be captured by haloes with deep potential wells and further evolve into galaxies (Press & Schechter 1974; White & Rees 1978), a good tracer of the invisible cosmic web.

Galaxy surveys explore the large-scale structure of the Universe by measuring the redshifts of millions of galaxies and quasars. Their positions can be used to calculate the two-point correlation function (2PCF) that encodes the history of the universe expansion and the structure growth. The Baryon Oscillation Spectroscopic Survey (BOSS, 2008–2014; Dawson et al. 2012) is the largest project in the third-stage Sloan Digital Spectroscopic Survey 1 (SDSS-III; Eisenstein et al. 2011). It has collected the spectra of more than 1.5 million Luminous Red Galaxies (LRGs; Reid et al. 2016), the brightest red galaxies in the Universe. Its extended version, eBOSS (2014–2020; Dawson et al. 2016) in SDSS-IV (Blanton et al. 2017) has probed another 300,000 LRGs (Ross et al. 2020). Additionally, eBOSS has observed around 270 000 emission-line galaxies (ELGs; 1http://www.sdss.org/.

© 2022 The Author(s)
Published by Oxford University Press on behalf of Royal Astronomical Society
They are bluer, star-forming galaxies that are abundant in the redshift range of $0.5 < z < 2$, where there are active star formation processes (Madau, Pozzetti & Dickinson 1998). With such a big amount of tracers, both projects have achieved a per-cent-level precision in cosmological parameter measurements (Alam et al. 2017, 2021a). Ongoing surveys like the Dark Energy Spectroscopic Instrument (DESI, 2020–2025; DESI Collaboration 2016) is expected to provide tracers in a larger footprint and with higher-resolution spectra, and reach a higher precision in scientific results.

Meanwhile, N-body simulations can solve numerically the gravitational evolution equation of DM in ΛCDM. They are able to accurately describe the DM distribution in the non-linear regime down to individual haloes in a large volume and provide halo properties at any cosmic epoch. Thus N-body simulations help validate cosmological theories that are based on perturbation theories across all the scales, and makes it possible to compare theoretical predictions with observations.

But there is a gap between dark-matter-only N-body simulations and observations. Since baryonic matter interacts with each other through all fundamental forces, their clustering properties do not necessarily follow those of DM on all scales (White & Rees 1978). Moreover, the spatial distributions of different types of galaxies, such as LRGs and ELGs, can also be different in the cosmic web (e.g. Malavasi et al. 2016; Kraljic et al. 2018). To bridge the gap, a galaxy–halo relation that modulates the clustering of haloes to match the observation is of great importance. Because it enables the direct comparison between the theory and the observation via simulation (Wechsler & Tinker 2018).

SubHalo Abundance Matching (SHAM) is a simple and intuitive empirical algorithm to model the galaxy–halo relation. As indicated by the name, SHAM makes use of DM haloes and their substructures that are dubbed subhaloes. The basic assumption of SHAM is a monotonic relation (not necessarily linear) between the galaxy luminosity (or stellar mass) and the halo mass (or the circular velocity; Kravtsov et al. 2004; Vale & Ostriker 2004; Nagai & Kravtsov 2005; Conroy, Wechsler & Kravtsov 2006; Behroozi, Conroy & Wechsler 2010). The algorithm applies the rank-ordering for both haloes and galaxies based on mass-related properties as just mentioned, and assigns galaxies to haloes/subhaloes one by one from the most massive end, until the SHAM galaxy catalogue reaches a desired number density.

Galaxies from this mock catalogue follow the observed or reconstructed luminosity function (e.g. Tasitsiomi et al. 2004; Conroy et al. 2006) or stellar mass function (e.g. Guo et al. 2010; Rodriguez-Torres et al. 2016) by construction. Their two-point statistics are tuned to be consistent with the observed one by introducing a Gaussian scatter between the galaxy luminosity and the halo circular velocity (Tasitsiomi et al. 2004) to include the scatter of the Tully–Fisher relation (Willick et al. 1997; Steinmetz & Navarro 1999). Using the peak maximum circular velocity throughout accumulation history ($V_{\text{peak}}$) instead of the halo mass further improves the performance of SHAM, since the galaxy accretion is free from the mass-stripping effect of subhaloes after reaching $V_{\text{peak}}$ (Trujillo-Gomez et al. 2011). With all the improvements, SHAM catalogues of Rodríguez-Torres et al. (2016) predict the galaxy–halo relation that agrees with the weak-lensing observation from Shan et al. (2017). Other work on SHAM has also included the effect of halo assembly bias and galaxy formation (e.g. Chaves-Montero et al. 2016; Contreras, Angulo & Zennaro 2021). Due to the single-parameter feature and its good agreement on data, SHAM has become a useful tool in cosmological studies.

While most of the SHAM studies choose to fit the projected 2PCF that marginalizes the effect along the line of sight, 2PCF multipole can provide extra information in that direction. Due to the redshift-space distortion (RSD), 2PCF quadrupole in the redshift space is vulnerable to the bias induced by the redshift uncertainty. This uncertainty mainly comes from the LRG redshift determination pipeline that uses broad absorption lines to determine the redshift (Bolton et al. 2012). For eBOSS LRG, Ross et al. (2020) find the uncertainty measured by the repeat observations can be fit by a Gaussian function with a dispersion of 91.8 km s$^{-1}$. This is negligible for large-scale clustering analysis like Baryon Acoustic Oscillation (BAO) at over 80 h$^{-1}$ Mpc. Meanwhile, Smith et al. (2020) demonstrate that the redshift uncertainty of eBOSS quasars influences the quadrupole on $r < 60$ h$^{-1}$ Mpc, affecting SHAM that mainly employs 2PCF on $r < 40$ h$^{-1}$ Mpc.

Moreover, the standard SHAM algorithm described above works well for bright galaxies with complete samples, as it matches halo masses and galaxy masses starting from the most massive ones. When the sample is incomplete due to the survey requirement (e.g. exclude targets with higher luminosity) or when the tracer is absent in massive haloes due to the quenching process (e.g. Kauffmann et al. 2004; Dekel & Birnboim 2006), this implementation can be problematic (Rodríguez-Torres et al. 2016; Favole et al. 2016a).

In this paper, we present a general three-parameter SHAM algorithm for LRGs on 5–25 h$^{-1}$ Mpc that considers the redshift uncertainty effects and the galaxy completeness. In Section 2, we describe the observational data, the N-body simulation and the galaxy mocks for covariance matrices. The calculation of 2PCFs and projected 2PCFs, the redshift uncertainty measurements, and the implementation of SHAM are introduced in Section 3. Section 4 illustrates the good performance of SHAM in fitting the 2PCF, reproducing the redshift uncertainty, the galaxy incompleteness evolution and the projected 2PCF of the observation. Finally, we summarize our findings in Section 5. In our study, we use a flat ΛCDM cosmology with $\Omega_m = 0.31$ and $H_0 = 67.7$ km s$^{-1}$ Mpc$^{-1}$, and the line-of-sight direction of the comoving space is along the Z-axis.

The BOSS and eBOSS SHAM study is part of the final release of the eBOSS measurement. Ross et al. (2020) and Lyke et al. (2020) describe in detail catalogues for the large-scale structure analysis. The mock challenge tasks are completed by Alam et al. (2021b), Avila et al. (2020), Rossi et al. (2021), and Smith et al. (2020) for systematics estimations. Galaxy mocks for covariance matrices used for cosmological analysis are constructed by Lin et al. (2020), Zhao et al. (2021a). The BAO and RSD measurements2 are based on the clustering of LRGs at $0.6 < z < 1.0$ (Bautista et al. 2021; Gil-Marín et al. 2020), ELGs at $0.6 < z < 1.1$ (Raichoor et al. 2020; Tamone et al. 2020; de Mattia et al. 2021), quasars (QSOs) at $0.8 < z < 2$ (Neveux et al. 2020; Hou et al. 2021), and also the Lyα forest at $z > 2.1$ (du Mas des Bouroux et al. 2020). Their cosmological interpretations3 in combination with the final BOSS results and other probes can be found in Alam et al. (2021a). eBOSS data also permit the analysis with cosmic voids (Aubert et al. 2022) and multiple tracers (Wang et al. 2020; Zhao et al. 2021b, 2022).

---


3see https://www.sdss.org/science/cosmology-results-from-eboss and https://svn.sdss.org/public/data/eboss/DR16/cosmomc/tags/v1_0_1/.
2 DATA

2.1 SDSS galaxies

We use the galaxy catalogues from the final data releases of SDSS-III BOSS\(^4\) (DR12; Alam et al. 2015) and SDSS-IV eBOSS\(^5\) (DR16 Ahumada et al. 2020). BOSS LRGs are composed of two subsamples: LOWZ at 0.15 < \(z\) < 0.5 and CMASS at 0.4 < \(z\) < 0.7 (Reid et al. 2016). For eBOSS, its LRG sample has redshift at 0.6 < \(z\) < 1.0 (Ross et al. 2020). Galaxy samples in the catalogues are pre-selected using the photometric information to ensure the observed samples are clean and abundant at a given redshift range for scientific targets. Their spectra are observed by fibers (York et al. 2000; Gunn et al. 2006; Smeel et al. 2013) and processed by the spectroscopic pipeline in order to determine their galaxy type and redshifts (Aihara et al. 2011; Bolton et al. 2012). The catalogues used for clustering analysis (clustering catalogue hereafter) are composed of the 3D position of tracers, as well as several weights to eliminate systematics effects on 2PCF measurements.

2.1.1 Target selection

Target selection using the photometric information includes the signal-to-noise ratio (S/N) selection, the colour selection, the flux limits cut, and the star exclusion (Prakash et al. 2016; Reid et al. 2016). In general, those criteria are meant to have high-quality LRG spectra, obtain the designed redshift range and number density, exclude the low-redshift blue galaxies and ensure high-redshift successful rate, and remove stars from LRG samples, respectively. LOWZ at 0.2 < \(z\) < 0.4 and CMASS \(z\) < 0.6 are expected to be nearly volume-limited, i.e. they are complete (Reid et al. 2016). The study of Leauthaud et al. (2016) confirms the LOWZ completeness but also finds CMASS is not as complete as Reid et al. (2016) describes.

For eBOSS LRG, there is a special magnitude cut, namely

\[
i \geq 19.9, \tag{1}\]

which is set to avoid the BOSS CMASS galaxies (Prakash et al. 2016). This lower limit corresponds roughly to an upper limit of the stellar mass, meaning that the most massive LRGs may have been excluded from the eBOSS samples. We use \(V_{\text{cut}}\) to account for it and we will explain it in detail in Section 3.3.

2.1.2 Repeat observations

Repeat observations from SDSS aim at testing the reproducibility of the spectral measurements and obtaining a higher S/N by coadding multiple spectra (Dawson et al. 2012). We use them to estimate the redshift uncertainty statistically in Section 3.2. The redshifts of repeated samples from BOSS are determined by IDLSPIC2D (Bolton et al. 2012) and those from eBOSS LRGs are determined by REDROCK\(^6\). Less than 5 per cent of galaxies in the clustering catalogues are observed twice, but those galaxies are good representatives of the clustering galaxies (Section 4.2).

2.1.3 Galaxy weights

The 2PCF contains the information of cosmological parameters and structure growth in its amplitude. But the amplitude can be easily modified by various systematics, e.g. photometric systematics, redshift failures, fibre collision effects, and the varying galaxy number density at different redshift (Reid et al. 2016; Ross et al. 2020). The corresponding weights to remove their impacts are: \(w_{\text{phot}}\) for all the photometric effects, \(w_{\text{fail}}\) for the redshift failure rate, \(w_{\text{CP}}\) for the missing close galaxy pairs due to the fibre collision and \(w_{\text{FKP}}\) for minimizing the cosmic variance.

The fibre collision problem arises from the physical size of fibres (62 arcsec) that defines the minimum separation between two fibres. Hence galaxies with small angular distances (i.e. close pairs) cannot be observed in a single exposure. For the BOSS data, the close-pair weight,

\[
w_{\text{CP}} = \frac{N_{\text{CP}} + N_{\text{good}}}{N_{\text{good}}}, \tag{2}\]

is applied to the nearest tracer of collided targets to nullify the fibre collision effects. Here \(N_{\text{CP}}\) is the number of the half-missing tracer pairs and \(N_{\text{good}}\) is the number of tracers with good redshifts (Ross et al. 2020). This nearest-neighbour approach is not able to fully correct the fibre collision effect (Guo, Zehavi & Zheng 2012; Rodriguez-Torres et al. 2016). Nevertheless, our SHAM fitting results are not biased by this effect (see Appendix A).

eBOSS uses a different weighting scheme. The pairwise-inverse-probability (PIP) weighting proposed by Bianchi & Percival (2017) is a better way to account for galaxy pairs with one missing in a single observation and the angular up-weighting (ANG) scheme introduced by Percival & Bianchi (2017) can recover missing galaxy pairs with a distance smaller than the size of a fibre. Mohammad et al. (2020) apply both PIP and ANG weights on eBOSS samples and provide unbiased galaxy clustering down to 0.1 h\(^{-1}\) Mpc.

The FKP weight is obtained by (Feldman, Kaiser & Peacock 1994)

\[
w_{\text{FKP}} = \frac{1}{1 + \bar{\pi}(z)P_0}, \tag{3}\]

where \(\bar{\pi}(z)\) is the number density at redshift \(z\) and \(P_0\) is the amplitude of the observed power spectrum at \(k = 0.15 h\ Mpc^{-1}\). For LRGs in BOSS/eBOSS, we take \(P_0 = 10,000 h^{-1} Mpc^{-1}\) (Reid et al. 2016; Ross et al. 2020).

2.2 N-body simulation: UNIT

Future galaxy surveys will span to larger cosmological volumes with a deeper photometry. The resolution and effective volume of the N-body simulation should also keep up with the improvements. In our study, we use the Universe N-body simulations for the Investigation of Theoretical models from galaxy surveys\(^7\) (UNIT; Chuang et al. 2019). This is a high-resolution and large-effective-volume N-body simulation that uses the fixed-amplitude method to suppress the cosmic variance, thus increasing the effective volume (Angulo & Pontzen 2016). Its effective volume is over 10 times larger than that of BOSS/eBOSS LRGs. UNIT has 4096\(^3\) particles in 1 h\(^{-1}\) Gpc\(^3\) cubic box, and a mass resolution of 1.2 × 10\(^4\) M\(_{\odot}\) h\(^{-1}\). Its cosmological parameters are \(\Omega_m = 0.3089, h = H_0/100\ km\ s^{-1}\ Mpc^{-1} = 0.6774, n_s = 0.9667, \sigma_8 = 0.8147\). The simulation evolves from \(a(t) = 0.01\) (\(z = 99\)) to 1 (\(z = 0\)), and produces 128 snapshots at different redshifts. In our study, we use 12 snapshots from \(z = 0.2760\)

\(^4\)https://data.sdss.org/sas/dr12/boss/lass/
\(^5\)https://data.sdss.org/sas/dr16/eboss/lass/catalogs/DR16/
\(^6\)https://github.com/desihub/redrock.
\(^7\)http://www.unitsims.org/.
to 0.9011, as shown in Table 1, and take one simulation box in each snapshot given the large effective volume of UNIT (see Section 4.1).

### 2.3 Galaxy mocks for covariance matrices

We use galaxy mocks to calculate covariance matrices in Section 3.4. For eBOSS, we use 1200 DR12 PATCHY mocks in each galactic cap. For eBOSS, there are 1000 realizations of EZMOCK mocks for LRGs in each galactic cap. Both of them are able to reproduce the observational two-point and three-point statistics down to non-linear scales.

Kitaura, Yepes & Prada (2014) introduces the PATCHY code to generate galaxy mocks based on the augmented Lagrangian perturbation theory (ALPT). They determine the galaxy properties by apply their own SHAM to the BigMultiDark simulation (Klypin et al. 2016) and fit the observational statistics in both real and redshift space. The clustering properties of PATCHY mocks are then calibrated by those of SHAM catalogues. The fiducial cosmology is a ΛCDM cosmology with $\Omega_m = 0.307115$, $\Omega_b = 0.048206$, $h = 0.6777$, $\sigma_8 = 0.8225$, $n_s = 0.9611$ (Kitaura et al. 2016).

The Effective Zel’dovich approximation mock (EZMOCK) proposed by Chuang et al. (2015) is another fast methodology to generate mock halo or galaxy catalogues, with the same cosmology as PATCHY mocks. EZMOCK relies on the Zel’dovich approximation (Zel’Dovich 1970) to calculate the displacement of DM particles at any redshift, given their initial positions. Tracers are assigned in the density field using both the deterministic and the stochastic bias model (Zhao et al. 2021a). Besides covariance matrix calculations, we also use EZMOCK mocks with and without systematics to study the impact of the biased 2PCF monopole on the projected 2PCF in Section 4.4.

### 3 METHODS

#### 3.1 Galaxy clustering

The 2PCF, denoted by $\xi$, measures the excess probability of finding a galaxy pair compared to a random distribution in a given volume. We use the Landy–Szalay estimator (LS; Landy & Szalay 1993) that minimizes the variances of the measurements:

$$\xi_{LS} = \frac{DD - 2DR + RR}{RR}.$$  

(4)

where galaxy–galaxy pairs (DD), random–galaxy pairs (DR), and random–random pairs (RR) are normalized by the total number of pairs, $\xi$ and the pair counts can be calculated as a function of the pair separation $s$ and the cosine of the angle between $s$ and the line-of-sight ($\beta$), i.e. $\mu = \cos(\beta)$. By weighting the 2D $\xi(s, \mu)$ with Legendre polynomials $P_l(\mu)$, we obtain the 1D $\xi$ multipoles as

$$\xi_l(s) = \frac{2l + 1}{2} \int_{-1}^{1} \xi(s, \mu) P_l(\mu) d\mu.$$  

(5)

In our study, we only consider the monopole $\xi_0$ and quadrupole $\xi_2$ at $s \in [5, 25] h^{-1} \text{Mpc}$. This is because scales larger than $25 h^{-1} \text{Mpc}$ are prone to imaging systematics (Huterer, Cunha & Fang 2013) while scales smaller than $5 h^{-1} \text{Mpc}$ are sensitive to fibre collision effects.

### Table 1. Properties of analysed galaxy samples and their best-fitting SHAM results.

<table>
<thead>
<tr>
<th>Project</th>
<th>Redshift range</th>
<th>$z_{\text{eff}}$</th>
<th>$z_{\text{UNIT}}$</th>
<th>$V_{\text{eff}}$</th>
<th>$10^4 V_{\text{eff}}$</th>
<th>$\sigma$</th>
<th>$\chi^2$/d.o.f.</th>
<th>$V_{\text{coil}}$</th>
<th>$\chi^2$/d.o.f.</th>
<th>Rescaled $\chi^2$/d.o.f.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOWZ</td>
<td>$0.2 &lt; z &lt; 0.33$</td>
<td>0.2754</td>
<td>0.2760</td>
<td>0.29</td>
<td>3.37</td>
<td>0.09$^{+0.06}_{-0.06}$</td>
<td>100.3$^{+7.6}_{-10.4}$</td>
<td>0.0067$^{+0.019}_{-0.0025}$</td>
<td>32/37</td>
<td>31/37</td>
</tr>
<tr>
<td></td>
<td>$0.3 &lt; z &lt; 0.43$</td>
<td>0.3865</td>
<td>0.3941</td>
<td>0.33</td>
<td>2.58</td>
<td>0.20$^{+0.05}_{-0.05}$</td>
<td>73.9$^{+12.3}_{-12.2}$</td>
<td>0.0059$^{+0.026}_{-0.028}$</td>
<td>51/37</td>
<td>50/37</td>
</tr>
<tr>
<td>CMASS</td>
<td>$0.43 &lt; z &lt; 0.51$</td>
<td>0.4686</td>
<td>0.4573</td>
<td>0.47</td>
<td>3.42</td>
<td>0.30$^{+0.05}_{-0.08}$</td>
<td>45.0$^{+20.6}_{-16.6}$</td>
<td>0.0078$^{+0.0074}_{-0.0040}$</td>
<td>41/37</td>
<td>39/37</td>
</tr>
<tr>
<td>LOWZ</td>
<td>$0.51 &lt; z &lt; 0.57$</td>
<td>0.5417</td>
<td>0.5574</td>
<td>0.46</td>
<td>3.63</td>
<td>0.22$^{+0.04}_{-0.04}$</td>
<td>23.2$^{+18.5}_{-15.0}$</td>
<td>0.0144$^{+0.037}_{-0.0036}$</td>
<td>43/37</td>
<td>41/37</td>
</tr>
</tbody>
</table>

**Notes.** The first six columns are the project names of galaxy samples, their redshift ranges, effective redshifts $z_{\text{eff}}$, the closest redshifts of the UNIT snapshot to $z_{\text{eff}}$, effective volumes, and their average number densities. The next four columns are the best-fitting SHAM parameters ($\sigma$, $V_{\text{meas}}$, $V_{\text{coil}}$) and the minimum $\chi^2$/d.o.f. The final column is the rescaled $\chi^2$/d.o.f., taking into account the uncertainty of UNIT simulations. There are redshift bins with two sets of SHAM parameters. The first set (row) is from the three-parameter SHAM and the second set is from two-parameter SHAM without $V_{\text{coil}}$. The two-parameter SHAM can only be applied to complete samples, i.e. LOWZ and CMASS galaxies at $z < 0.6$ (Leauthaud et al. 2016). Redshift bins with only one set of SHAM results are from the three-parameter SHAM.
Even with nearest-neighbour close-pair weights $w_{\text{CP}}$, the biases of the 2PCF quadrupole due to fibre collision effects can be as large as 3σ at $\sim 5 \, h^{-1} \, \text{Mpc}$ (Guo et al. 2012; Rodríguez-Torres et al. 2016). To understand its influence on our SHAM results, we test a fitting range of $[10, 25] \, h^{-1} \, \text{Mpc}$ – which is only mildly affected by fibre collisions – for the BOSS data. It turns out that the SHAM constraints in this range are consistent with those on $s \in [5, 25] \, h^{-1} \, \text{Mpc}$ (see Appendix A). So we use the fitting range $s \in [5, 25] \, h^{-1} \, \text{Mpc}$ for all BOSS/eBOSS samples hereafter.

We use linear $s$ bins with a 1 $h^{-1} \, \text{Mpc}$ interval, and 120 linear $\mu$ bins in the range of $[0, 1]$, for most of the 2PCF measurements in this work. Only for the eBOSS LRGs in different redshift bins, we have eight logarithmic bins for $\xi_0$ and $\xi_2$, respectively.

$\xi$ can also be measured as a function of the distance parallel ($\pi$) and perpendicular ($r_p$) to the line of sight. By integrating $\xi(r_p, \pi)$ with respect to $\pi$, one obtains the projected 2PCF as

$$w_p(r_p) = \int_{\pi_{\text{max}}}^{\pi_{\text{min}}} \xi(r_p, \pi) d\pi.$$  

(6)

We use eight logarithmic bins for $r_p \in [5, 25] \, h^{-1} \, \text{Mpc}$, and $\pi_{\text{max}}$ will be changed according to our needs as introduced in Section 4.4.

FCFC11 (Zhao et al. in preparation) and CORRFUNC PYTHON module (Sinha & Garrison 2019; Sinha & Garrison 2020) are employed to calculate $\xi(s)$ and $w_p$. $\xi(s)$ of eBOSS are computed with PIP + ANG-weighted pair counts (Mhammed et al. 2020).

Our SHAM galaxy catalogue is built from an N-body simulation in a box, so we convert the position of SHAM galaxies from real space to redshift space before calculating the 2PCF with (Kaiser 1987)

$$Z_{\text{redshift}} = Z_{\text{real}} + \frac{v_{\text{pec}} z}{a(t) H(z)},$$

(7)

where $z$ is the coordinate, $z$ is the redshift, and $v_{\text{pec}}$ is the peculiar velocity along the $z$-axis. The estimator used to calculate the SHAM 2PCF is the Peebles–Hauser estimator (Peebles & Hauser 1974):

$$\xi_{\text{PH}} = \frac{DD}{RR} - 1,$$  

(8)

where galaxy–galaxy pairs (DD) is also normalized by the total number of pairs. The random–random (RR) counts follow the analytical expression:

$$RR \approx \frac{4\pi s_{\text{max}}^3 - s_{\text{min}}^3}{3} \frac{1}{V_{\text{box}}} \frac{1}{N_{\mu}}.$$  

(9)

where $s_{\text{max}}$ and $s_{\text{min}}$ are the boundaries of the separation bins, $V_{\text{box}} = 1 \, h^{-3} \, \text{Gpc}^3$ is the volume of the simulation box and $N_{\mu} = 120$ is the number of $\mu$ bins. We do not apply weights to SHAM galaxies as there are no observational systematics and radial selection. The $\xi(s)$ of SHAM galaxies is calculated by the built-in 2PCF calculator in our SHAM implementation code.

3.2 Redshift uncertainty

The redshift uncertainty is an inevitable error while measuring the redshift. It mainly affects the quadrupole of our clustering measurements (Smith et al. 2020), since it can be regarded as random peculiar motions along the line-of-sight. As suggested in Bolton et al. (2012), the redshift determination pipeline is the primary source of the redshift uncertainty of SDSS LRGs, but it is not the dominant source. Hence, using the error given by the pipeline will underestimate the redshift uncertainty. As a result, we estimate it statistically via repeat observations.

The redshift difference between two measurements $\Delta z$ can be converted into $\Delta \nu$ that represents the radial motion as

$$\Delta \nu = \frac{\Delta z}{1 + z} c,$$  

(10)

where $z$ is the redshift of one of measurements with $\text{SPECPRIMARY} = 1$. With jackknife errors, we fit all the $\Delta \nu$ histograms with Gaussian distributions, obtaining the best-fitting dispersion $\sigma_{\Delta \nu}$ as illustrated in Fig. 1. We also quote their standard deviation, $\sigma_{\Delta \nu}$, as a statistical estimate of the redshift uncertainty. In our SHAM algorithm, its clustering effect is quantified by $v_{\text{mean}}$. 

3.3 SHAM implementation

Recent SHAM studies (e.g. Campbell et al. 2018; Granett et al. 2019; Contreras et al. 2021) consistently choose $V_{\text{peak}}$ to represent the halo mass of haloes and subhaloes. The subsequent progress of the SHAM algorithm is to sort the galaxy stellar mass and $V_{\text{peak}}$, which is obtained as

$$V_{\text{peak}} = V_{\text{peak}} S_g,$$  

(11)

where

$$S_g = 1 + N(0, \sigma^2).$$  

(12)

and match them one by one from the largest $V_{\text{peak}}$, until the number density of SHAM galaxies equals to that of the observation. In equation (11), $N(0, \sigma^2)$ represents a Gaussian random distribution centred on zero with $\sigma^2$ variance, where $\sigma$ is the only free parameter in the SHAM. The choice of the Gaussian function is due to the fact that the Tully–Fisher relation, i.e. the galaxy–halo mass relation, has Gaussian-distributed residuals (Willick et al. 1997; Steinmetz & Navarro 1999). We assert

$$S_g = \exp\left(N(0, \sigma^2)\right).$$  

(13)

when the Gaussian scatter is negative. Nevertheless, it does not change the fitting results. There are four prerequisites for the standard SHAM described above: (1) The cosmology of the simulation has to be close to the true one; (2) the simulation should have a good resolution to resolve the subhaloes down to the low-mass end and halo properties should be accurate; (3) the observed clustering measurements should be unbiased; and (4) the observational stellar mass function (SMF) has to be complete in the massive end. If one of these prerequisites is missing, SHAM will not model the observed clustering signal accurately on all scales.

For eBOSS LRGs, there is a rough truncation of the stellar mass in the massive end (Section 2.1.1). A straightforward way to account for this massive-end incompleteness is to discard the most massive haloes. We introduce a new parameter $V_{\text{ceil}}$ to remove the $V_{\text{ceil}}$ per cent of the largest scattered $V_{\text{peak}}$ (i.e. $V_{\text{peak}}$). Since $\sigma$ and $V_{\text{ceil}}$ both modulate the amplitude of 2PCF monopole, they are degenerate.

The redshift uncertainty can add bias to the observed 2PCF quadrupole. Therefore, we introduce an extra Gaussian-random distribution $N(0, v_{\text{mean}}^2)$ in the peculiar velocity of SHAM galaxies as

$$v_{\text{peak}} = v_{\text{pec}}, N(0, v_{\text{mean}}^2).$$  

(14)

The target is to mimic the impact of the measured Gaussian redshift uncertainties in the redshift-space clustering as presented in Fig. 1.

Three-parameter LRG SHAM  61

---

11 https://github.com/cheng-zhao/FCFC.
To conclude, our three-parameter SHAM implementation is as follows:

(i) Scatter the $V_{\text{peak}}$ as suggested in equation (11) with

\[
S_g = \begin{cases} 
1 + N'(0, \sigma^2), & N'(0, \sigma^2) \geq 0 \\
\exp(N'(0, \sigma^2)), & N'(0, \sigma^2) < 0
\end{cases}
\] (15)

(ii) Sort $V_{\text{scat}}^{\text{peak}}$ and discard the most massive $V_{\text{cell}}$ percent of haloes/subhaloes.

(iii) From the remaining catalogue, keep the $N_{\text{galith}}$ largest $V_{\text{scat}}^{\text{peak}}$.

(iv) Assign galaxies in the centre of those haloes/subhaloes, and smear the peculiar velocities of galaxies along the line of sight with equation (14).

(v) Convert the galaxy coordinates from the real space into the redshift space using equation (7).

To model correctly the observed clustering, the redshift of the UNIT snapshot should be close to the effective redshift of the data computed with

\[
z_{\text{eff}} = \frac{\sum z w^2}{\sum w^2}.
\] (16)

where $w$ is the total galaxy weight. For eBOSS (Ross et al. 2020),

\[
w = w_{\text{photosys}} w_{\text{FKP}} w_{\text{tot}} w_{\text{CP}},
\] (17)

while for BOSS the total weight has a different form (Reid et al. 2016):

\[
w = (w_{\text{photosys}} w_{\text{FKP}} w_{\text{tot}} + w_{\text{CP}} - 1).
\] (18)

Then we apply SHAM on this snapshot to find the best-fitting parameters. Besides the 2PCF calculator mentioned in Section 3.1, our code also has a built-in MULTINEST sampler that will be introduced in Section 3.4.

The number of SHAM galaxies $N_{\text{gal}}$ is determined by the observed effective number density $n_{\text{eff}}$ in a given redshift range calculated as

\[
n_{\text{eff}} = \sqrt{\frac{\int n(z)^2 dV}{\int dV}} = \sqrt{\frac{\int n(\chi)^2 \chi^2 d\chi}{\int \chi^2 d\chi}},
\] (19)

where the second equality is due to $dV = A_{\text{eff}} (\chi + d\chi)^3 - \chi^3)/3 = A_{\text{eff}} \chi^2 d\chi + O(d\chi^2)$. $A_{\text{eff}}$ is the effective area of the footprint, and $\chi$ is the comoving distance for an object at redshift $z$. The values are reported in Table 1.

### 3.4 SHAM fitting

The SHAM best-fitting parameters are obtained by minimizing the $\chi^2$ value (i.e. the maximum log-likelihood) between the SHAM and the observational 2PCF. The $\chi^2$ for a given parameter set $\Theta = \{\sigma, v_{\text{smear}}, V_{\text{cell}}\}$ is defined as

\[
\chi^2(\Theta) = (\xi_{\text{data}} - \xi_{\text{model}}(\Theta))^T C^{-1}(\xi_{\text{data}} - \xi_{\text{model}}(\Theta)),
\] (20)
where $\xi = (\xi_0, \xi_2)$ denotes the vector composed of the 2PCF monopole and quadrupole. The vectors $\xi_{\text{data}}$ and $\xi_{\text{model}}$ represent the data vector and the SHAM model 2PCF respectively. In particular, $\xi_{\text{model}}$ is obtained by averaging the 2PCFs of 32 SHAM galaxy realizations generated using the same $\Theta$ with different random seeds. This operation reduces the SHAM statistical uncertainty. $C$ is the unbiased covariance matrix (Hartlap, Simon & Schneider 2007):

$$C^{-1} = C_{\text{mod}}^{-1} = \frac{N_m - N_i - 2}{N_m - 1},$$

(21)

where $N_i$ is the length of $\xi_{\text{data}}$, i.e. the total number of bins used in the 2PCF fitting, and $N_m$ is the number of mocks used to compute the covariance matrix. $N_m = 1200$ for BOSS PATCHY mocks and $N_m = 1000$ for eBOSS EZMOK mocks. $C_{\text{mod}}$ is the covariance matrix calculated as (Zhao et al. 2021a)

$$C_{m,j} = \frac{1}{N_m - 1} \sum_{i=1}^{N_m} [\xi_i(s_i) - \bar{\xi}(s_i)] [\xi_j(s_i) - \bar{\xi}(s_i)],$$

(22)

where $\xi_i$ is the correlation function measured from the $i$th mock, and $\bar{\xi}(s_i)$ is the average of the mock correlation function in a given distance bin $s_i$.

We assume a Gaussian likelihood $L(\Theta)$ for our parameters, which is

$$L(\Theta) \propto e^{-\frac{1}{2} \Theta^T C_{\text{mod}}^{-1} \Theta},$$

(23)

and employ a Monte Carlo sampler MULTINEST12 (Feroz & Hobson 2008; Feroz, Hobson & Bridges 2009; Feroz et al. 2019), an efficient nested sampling technique especially for multimodal posteriors, to constrain $\Theta$. The survival volume (i.e. the prior volume) is defined as (Feroz & Hobson 2008)

$$X(\lambda) = \int_{[\Theta | L(\Theta) = \lambda]} \Pr(\Theta | H) d\Theta,$$

(24)

where $Pr(\Theta | H)$ is the parameter prior, and the evidence integral can be written as

$$Z = \int_0^1 L(\Theta) dX \approx \sum_{i=1}^{M} L_i w_i,$$

(25)

where $i$ is the number of iterations, $w_i = \frac{1}{\lambda} (X_{i-1} - X_{i+1})$, and $X_i$ is a sequence of decreasing values as $0 < X_{M_i} < ... < X_3 < X_1 < X_0 = 1$. During the sampling, live points walk randomly and simultaneously in the parameter space confined by the prior, and some of them will be deactivated if they have the lowest likelihood $L_i$. The sampling terminates if the evidence contribution from the $j$th iteration $\Delta Z_j = L_{\text{max}} X_j$ is lower than a certain threshold (i.e. the tolerance), where $L_{\text{max}}$ is the maximum likelihood among the current set of live points (Feroz et al. 2019). The final results of our SHAM do not change much when changing the tolerance and the initial number of live points. Therefore, for all the tests, we set the tolerance to be 0.5 and the number of particles to be 200, in order to improve the efficiency of the convergence.

The built-in analyser of PYMULTINEST13 (Buchner et al. 2014) is used to extract the median value and the $1\sigma$ limits of the parameters, and the maximum-likelihood $\chi^2$ listed in Table 1. The parameter set with the maximum-likelihood $\chi^2$ is indicated in the posteriors provided by GETDIST (Lewis 2019) in Appendix C.

12 https://github.com/farhanferoz/MultiNest.
13 https://github.com/JohannesBuchner/PyMultiNest.

---

**Figure 2.** The best-fitting three-parameter SHAM 2PCF multipoles compared with those of the LOWZ observations on $5 - 25 h^{-1}$ Mpc. The observational data and the cosmic variance $\epsilon_{\text{obs},2\theta,2}$ are represented by dots with error bars. Solid lines with shades are SHAM 2PCFs and their errors. The first and third row are the monopole and the quadrupole respectively. Residuals normalized by $\epsilon_{\text{obs},2\theta,2}$ are shown in the second and fourth row. The left-hand column is for galaxy samples in redshift slices and the right-hand column is for those in the bulk redshift range. SHAM 2PCFs agree with the observations with rescaled reduced $\chi^2$ values around 1.

---

**4 RESULTS**

4.1 Galaxy clustering fitting

We calibrate our SHAM model with LRGs on $5 - 25 h^{-1}$ Mpc for the full LOWZ, CMASS, and eBOSS LRG samples as well as 9 finer redshift bins as summarized in Table 1. Figs 2–4 show that 2PCFs of the SHAM catalogues agree with those of the observational data in general. Note that there are 32 realizations of SHAM catalogues produced by the maximum-likelihood parameter set, aiming at reducing the statistical error brought by the Gaussian scatters. For this reason, we choose the mean value of their 2PCFs to be the SHAM 2PCF, and re-scale their 2PCF standard deviations by $1/\sqrt{32}$ to obtain the error of the SHAM. We find that the SHAM error is negligible compared to observed statistical error determined by PATCHY EZMOCK mocks. Thus, we decide to ignore it in this study.

The reduced $\chi^2$ values from the fits do not include cosmic variances of the UNIT simulations. To include their influence in the analysis, we divide the reduced $\chi^2$ by

$$\frac{\epsilon_{\text{com}}^2}{\epsilon_{\text{obs}}^2} = 1 + \frac{\epsilon_{\text{UNIT}}^2}{\epsilon_{\text{obs}}^2} \approx 1 + \frac{V_{\text{eff,obs}}}{V_{\text{eff,UNIT}}},$$

(26)

where $\epsilon_{\text{obs}}^2$ and $\epsilon_{\text{UNIT}}^2$ represent the observational variance and the cosmic variance of UNIT simulation, respectively, and $\epsilon_{\text{com}}^2$ is the combination of two errors. $V_{\text{eff,obs}}$ is the effective volume calculated with (Wang, Chuang & Hirata 2013; Reid et al. 2016)

$$V_{\text{eff,obs}} = \sum_i \left( \frac{\bar{n}(z_i) P_0}{1 + \bar{n}(z_i) P_0} \right)^2 \Delta V(z_i),$$

(27)
where $\Delta V(z_i)$ is the volume of the shell at redshift $z_i$ and $\bar{n}(z_i)$ is the mean number density of that shell. The values are listed in the fourth column of Table 1 and we assume $V_{\text{eff,UNIT}} = 10 h^{-3}$ Gpc$^3$ (Chuang et al. 2019). The rescaled $\chi^2$ values are recorded in the final column of Table 1. As the rescaled value is a better indicator of the goodness of the fit than the $\chi^2$ value, our discussions hereafter are all based on the rescaled ones.

The $\chi^2$ value for the bulk CMASS sample is relatively large compared to those of CMASS redshift slices and other bulk samples. This can be attributed to the varying sample completeness of CMASS sample that we will prove later in this paper. The 2PCF quadrupoles of all CMASS SHAM samples are underestimated on 18–25 $h^{-1}$ Mpc, which is consistent with the results of Rodríguez-Torres et al. (2016). Similar (but less obvious) discrepancies are also observed from the eBOSS samples. They may be due to some uncorrected systematics. The monopole disagreement on $r > 10 h^{-1}$ Mpc and the resulting large $\chi^2$ value for eBOSS SHAM at $0.7 < z < 0.9$ might be due to observational systematics as well.

Our measured $\sigma \in [0.17, 0.40]$ for CMASS at $0.43 < z < 0.7$ is consistent with $\sigma = 0.31$ obtained by Rodríguez-Torres et al. (2016). It is worth noting that their scatter considers the incompleteness of the observed stellar mass function, i.e. they obtain the intrinsic scatter $\sigma_{\text{int}}$. But we assume that the observed stellar mass function in a certain range is complete. It means that our $\sigma$ account for both the intrinsic scatter and the observed incompleteness.

Since BOSS samples are complete at $z < 0$ (Leauthaud et al. 2016; Reid et al. 2016), we also apply our SHAM model without $V_{\text{cell}}$ (two-parameter SHAM hereafter) to the same simulation snapshot as the three-parameter version and fit to the same data. The best-fitting 2PCFs and parameter constraints are shown in Fig. 5 and Table 1, respectively. The rescaled reduced $\chi^2$ values for the two- and three-parameter SHAM models are similar for all LOWZ LRGs and CMASS LRGs at $0.43 < z < 0.51$, which confirms their galaxy completeness. The reduced $\chi^2$ for the two-parameter SHAM at $0.51 < z < 0.57$ is significantly larger (the $\chi^2$ difference shows that SHAM with $V_{\text{cell}} = 0$ is rejected more than $4\sigma$), due to the discrepancy of quadrupole on 5–7 $h^{-1}$ Mpc. It also means $\sigma$ and $V_{\text{cell}}$ are not completely degenerate. The monopole of the best-fitting two-parameter SHAM is also systematically lower than that of the observation. The issue in the result of the two-parameter SHAM demonstrate that CMASS at $0.51 < z < 0.57$ is not complete, i.e. $V_{\text{cell}} \neq 0$, in contrast to what has been presented in Reid et al. (2016). This is consistent with the marginalized posterior distributions of $V_{\text{cell}}$ in the three-parameter SHAM (see Fig. C4).

4.2 Redshift uncertainty indicated by $V_{\text{smear}}$

Fig. 6 shows that $\delta_{\Delta v}$ increases monotonically with $z_{\text{cell}}$. This is consistent with the fact that the absorption lines used to determine the redshift are broader at higher redshift than those at lower redshift, leading to larger redshift uncertainties. There is a general consistency between the best-fitting SHAM $V_{\text{smear}}$ and $\delta_{\Delta v}, \sigma_{\Delta v}$, for CMASS. For the eBOSS samples, SHAM $V_{\text{smear}}$ agrees with $\delta_{\Delta v}$, but both of them are systematically larger than $\sigma_{\Delta v}$. It means that we cannot neglect the Gaussian-fitting outliers and their effects on eBOSS quadrupoles. The tail of eBOSS $\Delta v$ distributions also slightly deviates from the Gaussian model, meaning that eBOSS $\sigma_{\Delta v}$ is not a good representative of the observed redshift uncertainty. The best-fitting $V_{\text{smear}}$ values of two-parameter SHAM are lower than those of the three-parameter SHAM, and this difference is larger for complete samples at $z < 0.51$ than that of CMASS at $0.51 < z < 0.57$. This is not a contradiction because the biggest change in $\sigma_{\Delta v}$ posterior happens in CMASS at $0.51 < z < 0.57$ as shown in Figs C7–C9. The small change in $V_{\text{smear}}$ is due to its weaker $V_{\text{cell}}$–$V_{\text{smear}}$ degeneracy.

Discrepancies also exist for LOWZ LRGs at all redshifts and they remain for the two-parameter LOWZ SHAM. To compare the differences in the quadrupole, we replace the best-fitting $V_{\text{smear}}$ values with $\sigma_{\Delta v}$, generate SHAM galaxy catalogues and calculate the 2PCFs. As presented in Fig. 7, their quadrupole differences are larger than $3\sigma$. It means that besides the statistical redshift uncertainty, LOWZ has unknown factors that boost the quadrupole. We investigate various potential causes listed below.

Figure 3. Same figure as Fig. 2, but for CMASS LRGs. The deviation for the bulk CMASS sample (right panels) is relatively large compared to those of the sliced redshift bins. We attribute this difference to its inhomogeneous sample completeness shown in Section 4.1.

Figure 4. Same figure as Fig. 2, but for eBOSS LRGs. The SHAM 2PCFs also agree with the observations except the one at $0.7 < z < 0.9$. 

Downloaded from https://academic.oup.com/mnras/article/516/1/57/6654882 by CNRS - ISTO user on 07 April 2023
Figure 5. 2PCF multipoles of the best-fitting two-parameter SHAM (with $\sigma$ and $v_{\text{sme}}$) compared with those of the observations on 5–25 $h^{-1}$ Mpc as in Fig. 2. The first column is for LOWZ in redshift slices, the second column is for LOWZ at a bulk redshift range, and the third column is for CMASS in redshift slices. SHAM 2PCF at $z < 0.51$ agrees with the observational 2PCF. However, a significant difference in the first two bins of the quadrupole is shown for the CMASS sample at 0.51 < $z$ < 0.57.

Figure 6. Comparison between the measured $v_{\text{sme}}$ from the three- (black dots with error bars) and two-parameter SHAM (open circles with error bars; a horizontal offset is applied to avoid overlapping), the best-fitting Gaussian dispersion $\sigma_{\Delta v}$ (blue line with shades) and the standard deviations $\hat{\sigma}_{\Delta v}$ (red lines with shades) of the $\Delta v$ distribution. The left-hand panel is for the entire sample, while the right panel is for samples in redshift slices. Both the BOSS CMASS and the eBOSS LRG samples show good agreements between $v_{\text{sme}}$ values and the redshift uncertainties measured from repeat samples. However, for the BOSS LOWZ sample, $v_{\text{sme}}$ values are significant larger than those estimated from repeat observations.

4.2.1 Missing subhaloes or velocity bias in the simulation

The host haloes/subhaloes for LRGs have a large mass (e.g. $> 1000$ particles), so we expect they are well resolved. But their peculiar velocity might have biases especially for subhaloes. However, our simulation-based SHAM provides $v_{\text{sme}}$ that agrees with CMASS and eBOSS samples and we do not expect a significant evolution in the velocity bias. Therefore, the properties of haloes and subhaloes should be reliable and do not lead to the $v_{\text{sme}}$ disagreement in LOWZ.

4.2.2 Observational systematics

The LOWZ sample includes SDSS-I/II galaxies with a redshift determination pipeline different from that of SDSS-III BOSS...
While redshift discrepancy. The first, second, and third column represent LOWZ samples at $0.2 < z < 0.33$, $0.33 < z < 0.43$, and $0.2 < z < 0.43$, respectively. The differences between two quadrupoles on $r < 10 h^{-1}$ Mpc are mostly larger than $3\sigma$.

(Bolton et al. 2012). Thus, it is possible that the spectroscopic pipelines yield different uncertainties of redshift measurements. In fact, we find that in the clustering catalogue, 56 percent of galaxies in the North Galactic Cap (NGC) and 90 percent of galaxies in the South Galactic Cap (SGC) are from SDSS-III. The component difference in two galactic caps indicates that we can analyse NGC and SGC separately. Repeated samples are all from SDSS-III. Therefore, the $\Delta v$ from repetitive observations may not be representative of the redshift uncertainty of the LOWZ clustering.

However, as shown in Fig. 8, a high proportion of SDSS-III galaxies in the SGC does not mean a smaller difference between the $\sigma_{\Delta v}$-generated 2PCF and the $v_{\text{smea}}$-generated 2PCF compared to the difference for LOWZ in the SGC. Therefore, the difference in the two spectroscopic pipelines cannot explain the $v_{\text{smea}}$ discrepancy.

4.2.3 Underestimation of redshift uncertainties

Considering the fact that errors from the spectroscopic pipeline underestimate redshift uncertainties, it is difficult to directly prove the representative of samples from the repeat observation. Nevertheless, we can proceed with the photometric information (Appendix B). In the LOWZ sample, we find that the fraction of the blue samples $f_{\text{blue}}$ in repeat observation is smaller than that of the clustering sample, but the differences are minor (see Table B1 and Table B2). Given the similarity in the $f_{\text{blue}}$ and colour–redshift distributions, the sample of the repeat observations are unbiased representatives of galaxies in the clustering catalogue. Even if the repeated samples are biased, this is not the main cause of the discrepancy. Because the redshift uncertainty of LOWZ at lower redshift should be smaller than that of CMASS, but $v_{\text{smea}}$ of LOWZ is systematically larger than that of CMASS.

4.2.4 A more complete model for galaxy assignment

While we are using the same model for CMASS and eBOSS LRGs to describe LOWZ, the discrepancy might indicate that a more complete model is required. For example, the LOWZ sample is composed of a different type of galaxies that has more satellites than CMASS.

In fact, our satellite fraction ($f_{\text{sat}}$) for LOWZ at $0.2 < z < 0.43$ is 12.6 per cent, similar to CMASS $f_{\text{sat}} = 12.3$ per cent at $0.43 < z < 0.7$. They are close to results from the halo occupation distribution fitting with $w_{p}$, another empirical model for the galaxy–halo relation: Parejko et al. (2013), based on DR9, give $12 \pm 2$ per cent for LOWZ at $0.2 < z < 0.4$ and Reid et al. (2014) based on DR10 give $10 \pm 2$ per cent at $0.43 < z < 0.7$. Both SHAM and HOD do not support the explanation of large $f_{\text{sat}}$ in LOWZ. Nevertheless, given the difference in data and fitting scales, $f_{\text{sat}}$ can still be the explanation. Additionally, as pointed out by Ross et al. (2014) and Favole et al. (2016b), we know that the blue tail of CMASS galaxies show a higher quadrupole at $5–40 h^{-1}$ Mpc compared to the dominating redder galaxies, which has 2PCF close to that of the full catalogue. For LOWZ, the ad hoc blue galaxies might play a more important role in the clustering properties. The velocity bias is also found to be different in CMASS and LOWZ samples (Guo et al. 2015; Lange et al. 2022). So it is possible that the LOWZ discrepancy can be mitigated by a better model that considers those factors. We leave it for a future work.

4.3 Galaxy incompleteness indicated by $\sigma$ and $V_{\text{cell}}$

Due to the high degeneracy between $\sigma$ and $V_{\text{cell}}$, the redshift evolutions of $\sigma$ and $V_{\text{cell}}$ are not obvious in the marginalized posterior of eBOSS (Fig. C6). But the 2D $\sigma$–$V_{\text{cell}}$ posterior contours show a clear trend that $\sigma$ or $V_{\text{cell}}$ are smaller at higher redshift.

Thus, given a smaller $\sigma$, we keep more haloes with large $V_{\text{peak}}$ that can be scattered to a certain range of $V_{\text{scatter}}$, leading to a smaller stellar mass incompleteness (see Section 4.1). A smaller $V_{\text{cell}}$ means to remove fewer haloes with the largest $V_{\text{cell}}$, so the sample is more complete at massive-end. This agrees with the minimum magnitude truncation mentioned in Section 2.1.1, since this threshold is expected to remove fewer galaxies at higher redshift (Zhai et al. 2017). To demonstrate this, we plot the completeness–redshift relation in Fig. 9. The completeness is defined as the ratio of the number of eBOSS LRG targets with a complete target selection including (Prakash et al. 2016), to that from a target selection without the $i$-band lower limit $i \geq 19.9$ (equation 1). Since there are no spectroscopic redshift measurements for objects excluded by the eBOSS target selection criteria, we match the eBOSS target
For the complete samples, i.e. LOWZ and CMASS at $0.43 < z < 0.51$, we obtain much tighter constraints on $\sigma$ at different redshifts, as shown in Table 1. Since the $\sigma$ parameter in this study absorbs the incompleteness of stellar mass function, our measurements provide upper bounds of the intrinsic scatter between stellar mass and our halo mass proxy $V_{\text{peak}}$, i.e. $\sigma_{\text{int}} < 0.31$ at $0.2 < z < 0.33$, $\sigma_{\text{int}} < 0.36$ at $0.33 < z < 0.43$, and $\sigma_{\text{int}} < 0.46$ at $0.43 < z < 0.51$ with 95% confidence. We do not consider the measurement at $0.51 < z < 0.57$ to be robust as explained in Section 4.1.

4.4 Consistency check with $w_p$

In principle, when the 2PCF multipoles of a SHAM catalogue agree well with those of the corresponding observational data, the $w_p$ of the two catalogues should also be consistent. To investigate the impacts of potential systematic effects that are usually observed on large scales (Ross et al. 2016), we measure the $w_p$ with two integral constraints for both the best-fitting SHAM catalogues and observations. At first, we use $w_p$ with $\pi_{\text{max}} = 80\, h^{-1}\, \text{Mpc}$ to obtain a good S/N and avoid the noise-dominated large scales (Mohammad et al. 2020). This $\pi_{\text{max}}$ value is a common choice for HOD studies (e.g. Avila et al. 2020). In this case, the $w_p$ of the SHAM for LOWZ samples agrees with the observed $w_p$, while for CMASS and eBOSS samples, the $w_p$ of SHAM catalogues are generally underestimated compared to those of the observations. We then reduce $\pi_{\text{max}}$ to $30\, h^{-1}\, \text{Mpc}$, and the $w_p$ measured from the SHAM and observational catalogues become consistent, as shown in Figs 10–12. Note that $\pi_{\text{max}}$ values are for both observational and SHAM $w_p$.

Potential uncorrected systematics that affect the 2PCF monopole on $r \gtrsim 80\, h^{-1}\, \text{Mpc}$ can explain the inconsistency in $w_p$ with $\pi_{\text{max}} = 80\, h^{-1}\, \text{Mpc}$. The discrepancy on the large scales of the

catalogues with the DECaLS DR9 data (Dey et al. 2019) to use the DECaLS photometric redshift measurements (Zhou et al. 2020) and 78.5 per cent of the eBOSS LRG candidates without the $i$-band lower limit are found in the DECaLS catalogue.

\[\text{http://www.legacysurvey.org/dr9/files}\]
**Figure 10.** The $w_p$ comparison between the best-fitting SHAM catalogues and the observations. The first row and the third row are $w_p$ with $\pi$ integrating up to 80 and 30 $h^{-1}$ Mpc, respectively, and their residuals normalized by the cosmic variance $\epsilon_{\text{obs},w_p}$ is shown in the second and fourth row. $\pi_{\text{max}}$ in the figures is for both observational and SHAM galaxies. The deviation at $r_p$ around 20 $h^{-1}$ Mpc for the bulk sample (0.2 < $z$ < 0.43) decreases while using a smaller $\pi_{\text{max}}$. However, we see that the deviation at 5 $h^{-1}$ Mpc increases, which might indicate systematics at smaller scales.

**Figure 11.** Same as Fig. 10, but for CMASS LRGs. SHAM catalogues start to be more consistent with the observed $w_p$ with $\pi_{\text{max}} = 30$ $h^{-1}$ Mpc compared to $\pi_{\text{max}} = 80$ $h^{-1}$ Mpc. Large deviations are found in the case of $\pi_{\text{max}} = 80$ $h^{-1}$ Mpc. This is consistent with our findings from multipoles that indicate systematics at larger scales, e.g. $r > 18$ $h^{-1}$ Mpc.

**Figure 12.** Same as Fig. 10, but for eBOSS LRGs. Left-hand figures (small redshift bins) show obvious deviations when using $\pi_{\text{max}} = 80$ $h^{-1}$ Mpc, but the deviations are mitigated when using $\pi_{\text{max}} = 30$ $h^{-1}$ Mpc.

2PCF monopole is observed when comparing observations with the galaxy mocks like PATCHY (Kitaura et al. 2016) and EZMOCK (Zhao et al. 2021a). Meanwhile, Huterer et al. (2013) explain that uncorrected photometric systematics can significantly bias clustering on large scales.

To directly demonstrate the consequences of uncorrected systematics on $w_p$ with different $\pi_{\text{max}}$, we compare the 2PCF multiples and projected 2PCFs of EZMOCK mocks without and with observational systematics, respectively. EZMOCK mocks with observational systematics are mocks that mimic the data, including all the systematics mentioned in Section 2.1.3. We do not apply weight corrections, to investigate the potential biases due to systematics. The clustering difference is normalized by the quadrature sum of the standard deviations for EZMOCK mocks with and without systematics. As is shown in Fig. 13, the observational systematics is responsible for the monopole difference, especially at scales larger than 50 $h^{-1}$ Mpc. Consequently, the bias presented in $w_p$ with $\pi_{\text{max}} = 80$ $h^{-1}$ Mpc is much larger than that of $w_p$ with $\pi_{\text{max}} = 20$ $h^{-1}$ Mpc. The result reveals the importance of properly choosing $\pi_{\text{max}}$ to avoid large-scale systematics as much as possible.

**5 CONCLUSION**

SHAM is a powerful method of constructing galaxy catalogue based on high-resolution simulations. We propose a three-parameter SHAM algorithm that is more general for different galaxy surveys. Besides the classical $\sigma$, we introduce $v_{\text{mean}}$ that smears the peculiar velocities of SHAM galaxies to model the effect of the redshift uncertainties, and $V_{\text{coll}}$ that removes the most massive haloes/galaxies to account for the completeness of tracers in the massive end due to the target selection (e.g. luminosity threshold).

We construct SHAM catalogues that reproduce the clustering of BOSS/eBOSS LRGs on $[5, 25] h^{-1}$ Mpc at 0.2 < $z$ < 1.0, based on the UNIT simulations. The 2PCF multipoles and projected 2PCFs of the SHAM catalogues are consistent with those of BOSS/eBOSS.
samples, including the entire LOWZ, CMASS, and eBOSS samples, and nine subsamples in different redshift slices. These results validate our new SHAM model. However, for some of the samples such as CMASS at 0.43 < z < 0.7 and eBOSS LRG 0.7 < z < 0.9, the best-fitting reduced χ² values for the 2PCF multipoles can be larger than 2. This may be due to unknown observational systematics. For the bulk CMASS samples, the inhomogeneity in completeness at different redshift bins can also contribute to the disagreement.

Since the galaxies at z < 0.6 are supposed to be complete (Reid et al. 2016), a SHAM model without V_{cell} should also work for them. This is confirmed by the fitting results except for the CMASS sample at 0.51 < z < 0.57, which is rejected at more than 4σ. Therefore, the BOSS LRGs should have already been incomplete from z ≈ 0.5. This finding is especially important for the reconstruction of the stellar mass function.

Using our SHAM method, the 2PCF bias brought by redshift uncertainties can be quantified through the v_{smear} parameter. Meanwhile, we estimate the redshift uncertainty statistically with the redshift difference Δv obtained from repeat observations. There are two estimators of the statistical redshift uncertainties: (1) the best-fitting Gaussian dispersions σ_{Δv} for the histogram of Δv and (2) the standard deviations of Δv as δ_{Δv}. As expected, both of them increase monotonically with the effective redshift due to broader absorption lines at higher redshifts. Both σ_{Δv} and δ_{Δv} agree with our CMASS v_{smear} measurements. For eBOSS samples, v_{smear} agrees with the δ_{Δv} measurements while it is in tension with σ_{Δv}. This can be explained by the larger outlier fraction of eBOSS histograms. In short, the standard deviations of the redshift differences are in agreement with v_{smear} for both the CMASS and eBOSS LRG samples.

However, for the BOSS LOWZ sample, there are significant disagreements between v_{smear} values and the redshift uncertainties measured from repetitively observed targets. We have looked into some potential sources of systematic biases including the robustness of subhaloes from the UNIT simulations, the SDSS-III spectroscopic pipeline upgrade and the representativeness of the repeat samples. But none of them can lead to the observed discrepancy. Despite the fact that we have similar satellite fraction for LOWZ and CMASS, it is still possible that LOWZ sample is composed of a different subclass of LRGs that has different satellite properties compared to LRGs from CMmass and eBOSS. To validate this, we need a more detailed SHAM model for LOWZ samples. We leave relevant studies to a future work.

We observe the redshift evolution of eBOSS σ−V_{cell} posteriors, indicating the sample is more complete at higher redshift bins. With the photometric redshift from DECaLS, we confirm the target selection criteria with a minimum i-band magnitude cut leads to the completeness evolution. Since the scatter parameter σ in our study include both the intrinsic scatter σ_{int} and the incompleteness of galaxy samples, our two-parameter SHAM measurements provide the constraints on σ_{int} as σ_{int} < 0.31 for LOWZ at 0.2 < z < 0.33, σ_{int} < 0.36 for LOWZ at 0.33 < z < 0.43, and σ_{int} < 0.46 for CMmass at 0.43 < z < 0.51.

The projected 2PCFs of the best-fitting SHAM catalogues are consistent with those of the observations when the integral limit p_{max} is as small as 30 h⁻¹ Mpc. However, w_{p} of SHAM catalogues deviates from the observations when p_{max} = 80 h⁻¹ Mpc. This can be explained by potential uncorrected systematics that mainly affect large scales. A test based on eBOSS mocks with and without systematics also proves that w_{p} evaluated with a p_{max} of around 30 h⁻¹ Mpc is less sensitive to systematic effects, compared to w_{p} with p_{max} = 80 h⁻¹ Mpc. The result also supports the choice of a smaller p_{max} when using w_{p} for SHAM and HOD studies (e.g. Alam et al. 2020).

To conclude, our three-parameter SHAM model works well for LRGs at a wide range of redshift, though there are a few exceptions that may be due to uncorrected observational systematics. Our algorithm is also effective in estimating uncertainties of redshift measurements, and detection of sample incompleteness for most of the SDSS samples. We are going to further improve the SHAM model by accounting for more properties, such as the satellite fraction. In the meantime, we also plan to extend our studies to ELGs. Because their star-forming processes are quenched in massive haloes (Kaufmann et al. 2004; Dekel & Birnboim 2006), the parameter V_{cell} is foreseen to be crucial. Later, we shall also construct multitracer SHAM models for different types of tracers with both auto- and cross-correlations for cosmological analysis.

ACKNOWLEDGEMENTS

JY, CZ, and GF acknowledge support from the SNF 200020_175751 ‘Cosmology with 3D Maps of the Universe’ research grant. GR acknowledges support from the National Research Foundation of Korea (NRF) through Grant No. 2020R1A2C1005655 funded by the Korean Ministry of Education, Science and Technology (MoEST).

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the US Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS website is www.sdss.org.

Figure 13. The effect of a biased 2PCF monopole on w_{p}. The first column represents the 2PCF multipoles of eZMOCK mocks without systematics (black lines with shades) and eZMOCK mocks with systematics (blue lines with shades). The second column is the w_{p} comparison between eZMOCK mocks without systematics (black lines with shades) and eZMOCK mocks with systematics (red lines with shades). The first row is w_{p} with σ integrating to 80 h⁻¹ Mpc and the third row is w_{p} integrating to 20 h⁻¹ Mpc. The second and fourth row for both columns are the difference between their clustering normalized by the quadrature sum of the standard deviations for eZMOCK mocks with and without systematics. The w_{p} difference is suppressed when p_{max} equals a smaller value.
APPENDIX A: THE EFFECT OF FIBRE COLLISION

To avoid potential biases of the clustering measurements due to uncorrected fibre collision effects with nearest-neighbour close-pair weights ($w_{CP}$), we conduct a new SHAM fitting on the monopole and quadrupole at ranges on [10, 25] ($h^{-1}$ Mpc) and [5, 25] ($h^{-1}$ Mpc) (blue).

![Figure A1. The posteriors of SHAM parameters for LOWZ at $0.2 < z < 0.43$, obtained with the fitting of monopole and quadrupole at ranges on [10, 25] (red) and [5, 25] ($h^{-1}$ Mpc) (blue).](Image)

**Figure A1.** The posteriors of SHAM parameters for LOWZ at $0.2 < z < 0.43$, obtained with the fitting of monopole and quadrupole at ranges on [10, 25] (red) and [5, 25] ($h^{-1}$ Mpc) (blue).

**Figure B1.** The colour–redshift diagram for CMASS at $0.43 < z < 0.7$ and the blue galaxy threshold (white dashed lines). Left-hand panel: the colour distribution of galaxies from the clustering catalogue. Right-hand panel: the colour distribution of galaxies from the repeat observation.

**Figure B2.** The same as Fig. B1 for LOWZ with two cuts. The first cut is $(g - i) = 2.35$ (the white solid line) and the second one is $(g - i) = 2.8z + 1.2$ (the white dashed line).

catalogue. The differentiation of the blue and red samples is achieved by applying an ad hoc criterion on the colour–redshift diagram. For example, with a constant colour cut $(g - i) = 2.35$, Favole et al. (2016b) select the blue tail of CMASS samples and find out that they have different clustering properties from the red samples. The magnitudes for colour calculation are the Composite Model Magnitudes (i.e. CMODELMAG; Reid et al. 2016) from the spAll catalogue. We start with using the same cut on both clustering catalogue and repeated samples for LOWZ and CMASS (Figs B1 and B2). The results in Table B1 shows that $f_{\text{blue}}$ in clustering samples are quite close to $f_{\text{blue}}$ of the repeat samples. Despite of the minor difference, for LOWZ, $f_{\text{blue}}$ of the clustering catalogue is always larger than $f_{\text{blue}}$ of the repeated samples, while for CMASS, it is the opposite. Additionally, there is no significant difference in the $\sigma_{\Delta v}$ of blue galaxies and red galaxies that can lead to the disagreement between LOWZ $v_{\text{smear}}$ and LOWZ $\Delta v$.

As shown in Fig. B2, a constant cut cannot help to select the blue tail of LOWZ. A cut varies with the redshift is a better choice. So we apply $(g - i) = 2.8z + 1.2$, and find $f_{\text{blue}}$ drops to the same level as CMASS with a constant cut. The clustering catalogue still has $f_{\text{blue}}$ larger than that of the repeat observations and we have small difference between the red and blue $\sigma_{\Delta v}$ just like our findings in the results of the constant cut. The results can be found in Table B2.

So it proves that the redshift uncertainties measured based on repeat samples should be representative for the clustering sample.

APPENDIX B: BLUE SAMPLES IN LOWZ

In order to indirectly prove the representative of the repeat observation, we choose to study the fraction of blue galaxies, i.e. $f_{\text{blue}}$ in the repeat measurement catalogue and the galaxy clustering
Table B1. The $f_{\text{blue}}$ for the clustering catalogue and the repeat observation, and the $\hat{\sigma}_{\Delta V}$ of the blue and the red galaxies for LOWZ and CMASS samples at different redshift bins using the constant cut $(g-i)=2.35$.

<table>
<thead>
<tr>
<th>$z$ range</th>
<th>$f_{\text{blue}}$ (clt. per cent)</th>
<th>$f_{\text{blue}}$ (repeat per cent)</th>
<th>Red $\hat{\sigma}_{\Delta V}$ (s$^{-1}$ km)</th>
<th>Blue $\hat{\sigma}_{\Delta V}$ (s$^{-1}$ km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2 &lt; $z$ &lt; 0.33</td>
<td>88.9</td>
<td>87.3</td>
<td>29.3 ± 3.4</td>
<td>26.0 ± 2.5</td>
</tr>
<tr>
<td>0.33 &lt; $z$ &lt; 0.43</td>
<td>46.1</td>
<td>44.6</td>
<td>33.1 ± 1.7</td>
<td>33.1 ± 3.6</td>
</tr>
<tr>
<td>0.2 &lt; $z$ &lt; 0.43</td>
<td>67.3</td>
<td>64.6</td>
<td>32.5 ± 1.5</td>
<td>28.8 ± 2.2</td>
</tr>
<tr>
<td>0.43 &lt; $z$ &lt; 0.51</td>
<td>37.3</td>
<td>37.5</td>
<td>46 ± 2.7</td>
<td>48.3 ± 4.7</td>
</tr>
<tr>
<td>0.51 &lt; $z$ &lt; 0.57</td>
<td>36.3</td>
<td>37.4</td>
<td>52.8 ± 3.4</td>
<td>52.6 ± 3.0</td>
</tr>
<tr>
<td>0.57 &lt; $z$ &lt; 0.7</td>
<td>37.0</td>
<td>38.8</td>
<td>61.0 ± 3.6</td>
<td>60.7 ± 4.8</td>
</tr>
<tr>
<td>0.43 &lt; $z$ &lt; 0.7</td>
<td>36.9</td>
<td>38.0</td>
<td>54.0 ± 1.9</td>
<td>54.6 ± 2.8</td>
</tr>
</tbody>
</table>

Notes. The $f_{\text{blue}}$ values of LOWZ clustering catalogue are larger than those of the repeat catalogue, while CMASS samples shows the opposite relation. The difference between the red and blue $\hat{\sigma}_{\Delta V}$ is not significant for LOWZ and CMASS.

Table B2. The same as Table B1 for LOWZ using the redshift cut $(g-i)=2.8z+1.2$.

<table>
<thead>
<tr>
<th>$z$ range</th>
<th>$f_{\text{blue}}$ (clt. per cent)</th>
<th>$f_{\text{blue}}$ (repeat per cent)</th>
<th>Red $\hat{\sigma}_{\Delta V}$ (s$^{-1}$ km)</th>
<th>Blue $\hat{\sigma}_{\Delta V}$ (s$^{-1}$ km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2 &lt; $z$ &lt; 0.33</td>
<td>30.9</td>
<td>29.1</td>
<td>26.8 ± 2.9</td>
<td>25.7 ± 1.7</td>
</tr>
<tr>
<td>0.33 &lt; $z$ &lt; 0.43</td>
<td>30.5</td>
<td>30.2</td>
<td>32.7 ± 1.4</td>
<td>34.1 ± 5.1</td>
</tr>
<tr>
<td>0.2 &lt; $z$ &lt; 0.43</td>
<td>30.7</td>
<td>29.7</td>
<td>30.0 ± 1.5</td>
<td>30.5 ± 3.2</td>
</tr>
</tbody>
</table>

Note. The $f_{\text{blue}}$s of clustering are still larger than those of the repeat catalogue and the difference between the red and blue $\hat{\sigma}_{\Delta V}$ is also small.

APPENDIX C: POSTERIORS OF BOSS AND EBOSS SHAM

The posteriors of all the three-parameter SHAM results are presented in Figs C1–C6. The top panel of each column shows the marginalized posterior of the parameter written at the bottom. The remaining three panels are the 2D posterior contours. Parameters with the maximum likelihood determined by PYMULTINEST\footnote{https://github.com/JohannesBuchner/PyMultiNest.} (Buchner et al. 2014) are marked in the figures. All the posteriors are from converged Monte Carlo chains of MULTINEST. The multimodal 1D posteriors of $\sigma$ and $V_{\text{cell}}$ may be due to the noisy data.

We also present the posteriors of the two-parameter SHAM without $V_{\text{cell}}$ in Figs C7–C9, comparing with those of the corresponding three-parameter SHAM. We keep the posteriors of $V_{\text{cell}}$ for the convenience of visualizing the $V_{\text{cell}}$–$V_{\text{smear}}$ degeneracy.

Figure C1. The posteriors of SHAM parameters for LOWZ samples at $0.2 < z < 0.43$. The black dots in the contours indicate the position of the best-fitting parameters.

Figure C2. The same as Fig. C1 for LOWZ SHAM at $0.2 < z < 0.33$ (magenta) and $0.33 < z < 0.43$ (blue).
Figure C3. The same as Fig. C1 for CMASS SHAM at $0.43 < z < 0.7$.

Figure C4. The same as Fig. C1 for LOWZ SHAM at $0.43 < z < 0.51$ (magenta), $0.51 < z < 0.57$ (blue), and $0.57 < z < 0.7$ (orange).

Figure C5. The same as Fig. C1 for eBOSS LRG SHAM at $0.6 < z < 1.0$.

Figure C6. The same as Fig. C1 for eBOSS LRG SHAM at $0.6 < z < 0.7$ (magenta), $0.6 < z < 0.8$ (blue), $0.7 < z < 0.9$ (orange), and $0.8 < z < 1.0$ (red). The evolution of the $\sigma-V_{\text{cell}}$ contour indicates the incompleteness at higher redshift bins is lower than that of the lower redshift bins, as explain in Section 4.3.
Figure C7. The posteriors of three-parameter SHAM (magenta) compared with two-parameter SHAM (red) for LOWZ samples at $0.2 < z < 0.43$.

Figure C8. The same as Fig. C7 for LOWZ SHAM comparisons at $0.2 < z < 0.33$ (magenta, red) and $0.33 < z < 0.43$ (cyan, blue).

Figure C9. The same as Fig. C7 for CMASS SHAM comparisons at $0.43 < z < 0.51$ (magenta, red) and $0.51 < z < 0.57$ (cyan, blue).

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