

# The UCHUU-GLAM BOSS and eBOSS LRG lightcones: exploring clustering and covariance errors

Julia Ereza <sup>1</sup>★, Francisco Prada,<sup>1</sup> Anatoly Klypin,<sup>2</sup> Tomoaki Ishiyama <sup>3</sup>, Alex Smith <sup>4</sup>,  
Carlton M. Baugh <sup>4</sup>, Baojiu Li <sup>4</sup>, César Hernández-Aguayo <sup>5,6</sup> and José Ruedas<sup>1</sup>

<sup>1</sup>*Instituto de Astrofísica de Andalucía (CSIC), Glorieta de la Astronomía, E-18080 Granada, Spain*

<sup>2</sup>*Department of Astronomy, University of Virginia, Charlottesville, VA 22904, USA*

<sup>3</sup>*Digital Transformation Enhancement Council, Chiba University, 1-33, Yayoi-cho, Inage-ku, Chiba 263-8522, Japan*

<sup>4</sup>*Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*

<sup>5</sup>*Max-Planck-Institut fuer Astrophysik, Karl-Schwarzschild-Str. 1, D-85748 Garching, Germany*

<sup>6</sup>*Excellence Cluster ORIGINS, Boltzmannstrasse 2, D-85748 Garching, Germany*

Accepted 2024 June 18. Received 2024 April 29; in original form 2023 August 23

## ABSTRACT

This study investigates the clustering and bias of Luminous Red Galaxies (LRG) in the BOSS-LOWZ, -CMASS, -COMB, and eBOSS samples, using two types of simulated lightcones: (i) high-fidelity lightcones from UCHUUN-body simulation, employing SHAM technique to assign LRG to (sub)haloes, and (ii) 16 000 covariance lightcones from GLAM-UCHUUN-body simulations, including LRG using HOD data from UCHUU. Our results indicate that UCHUU and GLAM lightcones closely replicate BOSS/eBOSS data, reproducing correlation function and power spectrum across scales from redshifts 0.2 to 1.0, from 2 to 150  $h^{-1}$  Mpc in configuration space, from 0.005 to 0.7  $h$  Mpc $^{-1}$  in Fourier space, and across different LRG stellar masses. Furthermore, comparing with existing MD-PATCHY and EZMOCK BOSS/eBOSS lightcones based on approximate methods, our GLAM-UCHUU lightcones provide more precise clustering estimates. We identify significant deviations from observations within 20  $h^{-1}$  Mpc scales in MD-PATCHY and EZMOCK, with our covariance matrices indicating that these methods underestimate errors by between 10 per cent and 60 per cent. Lastly, we explore the impact of cosmology on galaxy clustering. Our findings suggest that, given the current level of uncertainties in BOSS/eBOSS data, distinguishing models with and without massive neutrino effects on large-scale structure (LSS) is challenging. This paper highlights the UCHUU and GLAM-UCHUU simulations' robustness in verifying the accuracy of Planck cosmological parameters, providing a strong foundation for enhancing lightcone construction in future LSS surveys. We also demonstrate that generating thousands of galaxy lightcones is feasible using  $N$ -body simulations with adequate mass and force resolution.

**Key words:** surveys – galaxies: haloes – large-scale structure of Universe – cosmology: observations – cosmology: theory.

## 1 INTRODUCTION

One of the main goals in cosmology is to ascertain the underlying physical principles governing the formation and evolution of the large-scale structure (LSS) in the Universe. In the current paradigm, this structure originates from primordial density fluctuations seeded on classical scales during inflation, which subsequently grow through gravitational instability. The Baryon Acoustic Oscillations (BAOs) are a unique component of these fluctuations, imprinted on the matter distribution later in cosmic history. They represent the distance travelled by sound waves prior to matter–radiation decoupling [see Cole et al. 2005; Eisenstein et al. 2005, for the first BAO detections using the Sloan Digital Sky Survey (SDSS) and 2dFGRS galaxy redshift surveys].

The latest SDSS-III BOSS (Dawson et al. 2013) and SDSS-IV eBOSS (Dawson et al. 2016) surveys mapped the distribution of

luminous red galaxies (LRG), quasars, and emission-line galaxies to measure the characteristic scale imprinted by the BAO in their clustering signal, and to determine the redshift–distance relation over the redshift interval  $z = 0.2$ – $1$ . These BAO measurements, together with redshift-space distortions (RSD), have allowed the constraint of some cosmological parameters with high precision (see Alam et al. 2021). In particular, for the LRG galaxy samples that are the subject of this investigation, the BOSS Data Release 12 (DR12; Alam et al. 2015) provides the redshifts of 1 million LRG over 10 000 deg $^2$  of the sky, covering the redshift interval from 0.2 to 0.8. Additionally, eBOSS Data Release 16 (DR16; Ahumada et al. 2020) provides a sample with around 180 000 LRG in the redshift interval between  $z = 0.6$  and 1, over an area of 4000 deg $^2$ . These figures will be largely superseded by the ongoing DESI survey, which is expected to measure spectra for an order of magnitude more galaxies, thereby tightening the constraints on cosmological parameters (see DESI Collaboration 2022).

In order to compare a given cosmological model, and in particular, the predictions of the standard Lambda cold dark matter ( $\Lambda$ CDM)

\* E-mail: [julia.fe.1996@gmail.com](mailto:julia.fe.1996@gmail.com)

model with observational data, it is essential to generate high-fidelity galaxy lightcones from cosmological simulations (e.g. de la Torre et al. 2013; Smith et al. 2017; Dong-Páez et al. 2022; Yu et al. 2022). These lightcones must be able to fully capture the properties of the observed galaxy sample, ranging from the angular footprint to clustering. Different methods are available to generate such lightcones. Hydrodynamic simulations are arguably the most accurate way to model the formation and evolution of galaxies. However, these calculations are computationally infeasible for the volumes required in LSS studies (e.g. Frenk et al. 1999; Springel 2005, though see the recent large volume runs in the FLAMINGO suite of Schaye et al. 2023 and the MILLENNIUMTNG run presented by Hernández-Aguayo et al. 2023; Pakmor et al. 2023). A cheaper option is to use a semi-analytical model to predict the galactic content of dark matter (DM) haloes in lightcones (Kitzbichler & White 2007; Merson et al. 2019; Stoppacher et al. 2019; Barrera et al. 2023); however, it is still challenging to tune such models to reproduce clustering measurements as closely as is possible using empirical methods for populating haloes with galaxies.

Dark matter only  $N$ -body simulations are not only a computationally cheaper option, but also enable us to simulate large volumes. These simulations model the gravitational interactions of a system of  $N$  collisionless particles over time and are able to follow the growth of LSS deep into the non-linear regime. To produce galaxy lightcones, the DM haloes extracted from these simulations must be populated with galaxies. Empirical methods are the most popular ones, as they are simple to apply and their parameters can be rapidly chosen to reproduce the clustering and abundance of a target galaxy sample. A common method is the halo occupation distribution, (HOD; Peacock & Smith 2000; Wechsler et al. 2001; Berlind & Weinberg 2002; Kravtsov et al. 2004), which models the probability that a halo of mass  $M_{\text{halo}}$  hosts  $N_{\text{gal}}$  galaxies. These models have a number of free parameters that allow the observed clustering to be replicated with simulated lightcones. Subhalo abundance matching (SHAM; Vale & Ostriker 2006; Behroozi, Conroy & Wechsler 2010; Trujillo-Gomez et al. 2011; Reddick et al. 2013; Guo et al. 2016; Masaki et al. 2023) assumes a one-to-one relation between galaxy luminosity (or stellar mass) and a given halo property, such as halo mass or circular velocity, to which a scatter is applied (e.g. Rodríguez-Torres et al. 2016). Other extensions of the basic SHAM model have also been proposed and applied to build mock catalogues (e.g. Contreras, Angulo & Zennaro 2021; Contreras et al. 2023a, b). The SHAM model relies on the assumption that more luminous (massive) galaxies inhabit more massive haloes. The galaxy lightcones generated using this method reproduce the observed luminosity (or stellar mass) function and, by construction, recover the observed galaxy clustering.

Prior to this study, simulated galaxy lightcones have been generated for both the BOSS and eBOSS-LRG samples. In the context of BOSS, the most commonly used LRG lightcones are the MULTIDARK-PATCHY ones (MD-PATCHY; Kitaura et al. 2016), based on the high-fidelity BIGMULTIDARK lightcone (BIGMD; Rodríguez-Torres et al. 2016). On the other hand, for the eBOSS-LRG, EZMOCK lightcones are employed (Zhao et al. 2021). In the case of BIGMD lightcones, the SHAM method was used to populate LRG within haloes drawn from the BIGMD Planck  $N$ -body simulation (Klypin et al. 2016). Subsequent steps required for generating the lightcones, including those for MD-PATCHY, were carried out using the SURVEY GENERATOR code (SUGAR; as outlined in Rodríguez-Torres et al. 2016). Both the PATCHY and EZMOCK methodologies, primarily designed for covariance matrices, make use of approximate models. The PATCHY code generates fields for dark matter density and peculiar velocity on a mesh, employing Gaussian fluctuations and implement-

ing the Augmented Lagrangian Perturbation Theory (ALPT) scheme (as described in Kitaura, Yepes & Prada 2014). On the other hand, the EZMOCK approach adopts the Zel'dovich approximation (Zel'dovich 1970) to create density fields at specific redshifts. Subsequently, galaxies are added to the density field in both the PATCHY and EZMOCK methods using a model based on the bias of LRG. The parameters of this bias model are determined by fitting the LRG two-point correlation function (2PCF) and three-point correlation function (3PCF). For the PATCHY method, the 2PCF of the high-fidelity BIGMD-BOSS lightcone is fitted, while for the EZMOCK method, the observed eBOSS 2PCF is employed for this fitting process.

There are two primary reasons for generating new simulated lightcones for the BOSS and eBOSS surveys. First, in this work we make use of the UCHUU  $N$ -body cosmological simulation to construct high-fidelity LRG lightcones for both surveys, adopting the SHAM method to populate (sub)haloes with LRG. UCHUU is a large simulation containing a 2.1 trillion particles, set within a  $(2 h^{-1} \text{Gpc})^3$  volume. This unprecedented numerical resolution within such a large volume enables the resolution of DM haloes and subhaloes down to the domain of dwarf galaxies (see Ishiyama et al. 2021 for details). This surpasses the numerical resolution of BIGMD, which was employed for constructing BOSS-LRG lightcones (Rodríguez-Torres et al. 2016). Consequently, we anticipate achieving clustering signals at smaller scales, further enhancing the statistics of low-mass galaxies. In addition, UCHUU adopts the Planck-2015 cosmology (hereafter PL15; Planck Collaboration XIII 2016), making it perfectly well suited for assessing the clustering predictions of the standard  $\Lambda$ CDM model against data from BOSS and eBOSS.

Secondly, alternative methods to  $N$ -body cosmological simulations, such as those adopted to generate the MD-PATCHY and EZMOCK covariance matrices for BOSS and eBOSS, are more efficient but less accurate than  $N$ -body simulations, exhibiting considerably lower precision on scales below  $20 h^{-1} \text{Mpc}$ , as highlighted by our investigation. MD-PATCHY and EZmock are not based on a real  $N$ -body simulation, and they fail to generate a precise matter density field compared to a full  $N$ -body simulation. Consequently, it is unclear whether the resulting galaxy catalogues are feasible in a real universe, raising uncertainty about whether the resulting galaxy lightcones can produce accurate covariance error estimates for the BOSS and eBOSS clustering statistics. Moreover, a recent study by Yu et al. (2023) reported some evidence of model specification errors in MD-PATCHY. Here, we generate BOSS and eBOSS lightcones for estimating covariance error by utilizing GLAM  $N$ -body cosmological simulations that fully encode the nonlinear gravitational evolution (see Klypin & Prada 2018, for details). In contrast to UCHUU, the GLAM simulations can only be used to resolve distinct haloes (not subhaloes), prompting us to employ the HOD approach for populating GLAM haloes with LRG. We obtain this HOD from the galaxy catalogues constructed using UCHUU for BOSS and eBOSS surveys. In addition, previous generation of covariance lightcones required the specification of a substantial number of parameters (five for MD-PATCHY and six for EZmock), resulting in a considerable degree of parameter degeneracy. In the GLAM lightcones, we have only one free parameter: the scatter parameter introduced in the SHAM method applied in UCHUU that quantifies the scatter between the galaxy stellar mass and the proxy of halo mass.

This paper is organized as follows. In Sections 2 and 3, we introduce the BOSS and eBOSS galaxy samples, along with the UCHUU and GLAM simulations used in this study. Section 4 outlines the methodology employed to generate the UCHUU lightcones for BOSS and eBOSS. The construction of the GLAM covariance lightcones is described in Section 5. Our results and their discussion,

**Table 1.** Properties of the BOSS and eBOSS samples studied in this work. The first and second columns provide the sample name and the corresponding redshift range, respectively. The remaining columns list for each sample the (weighted) number of galaxies, the effective area (in  $\text{deg}^2$ ), the galaxy density (in units of  $\text{deg}^{-2}$ ), and the effective volume (in  $h^{-3}\text{Gpc}^3$ ).

Name	$z$ range	$N_{\text{eff}}$	$A_{\text{eff}}$	$N_{\text{eff}}/A_{\text{eff}}$	$V_{\text{eff}}$
LOWZ-N	0.2–0.4	196 186	5790	33.9	0.3633
LOWZ-S	0.2–0.4	91 103	2491	36.6	0.1621
CMASS-N	0.43–0.7	605 884	6821	88.8	1.2777
CMASS-S	0.43–0.7	225 469	2525	89.3	0.4745
COMB-N	0.2–0.75	807 900	5790	139.5	1.7208
COMB-S	0.2–0.75	352 584	2491	141.5	0.7468
eBOSS-N	0.6–1.0	114 004	2476	46.0	0.2584
eBOSS-S	0.6–1.0	71 677	1627	44.1	0.1602

presented in Section 6, are divided into three parts: comparison of observations to our theoretical predictions based on the Planck cosmology as determined from the UCHUU lightcones (Section 6.1), a study of the performance of our GLAM lightcones compared to previous ones (Section 6.2), and an exploration of the effect that cosmology has on the distribution of galaxies using different GLAM runs (Section 6.3).

## 2 GALAXY DATA

We make use of various publicly available observational data sets, including the BOSS-LOWZ, BOSS-CMASS, and eBOSS-LRG samples, encompassing both the Northern (N) and Southern (S) hemispheres. Furthermore, we analyse the combined BOSS-LOWZCMASS samples for both hemispheres. For clarity, we will subsequently refer to these data sets as LOWZ, CMASS, COMB, and eBOSS. Below, we describe in detail each of these samples. A summary of their global properties is given in Table 1.

### 2.1 BOSS samples

We use the final data release of SDSS-III BOSS LRG (DR12; Alam et al. 2017), which contains the redshifts of 1 million LRG. This data set covers an area of approximately  $10\,000\text{ deg}^2$ , divided into two distinct subsets: LOWZ, designed for targeting LRG up to  $z \sim 0.4$ , and CMASS, designed to target massive galaxies within the redshift range of  $0.4 < z < 0.7$ . Additionally, to maximize the effective volume covered by BOSS galaxies, a separate sample is constructed by combining the LOWZ and CMASS samples, denoted as the COMB sample. The spatial distribution and number density of these samples are shown in Figs 1 and 2. The BOSS large-scale structure (LSS) catalogues are publicly available.<sup>1</sup> For a more comprehensive understanding of the LOWZ and CMASS target selection criteria, and the creation of the COMB sample, we refer to Reid et al. (2015).

In these catalogues, as defined in Alam et al. (2017), a total weight is assigned to each galaxy  $i$ :

$$w_{\text{tot},i} = w_{\text{systot},i}(w_{\text{cp},i} + w_{\text{noz},i} - 1), \quad (1)$$

where  $w_{\text{systot}}$  is the total angular systematic weight,  $w_{\text{cp}}$  is the fibre collision correction weight, and  $w_{\text{noz}}$  is the redshift failure weight. This total galaxy weight must be used to obtain an unbiased estimation of the galaxy density field.

The geometry of the sky area covered by each LRG sample is precisely outlined using MANGLE format masks<sup>1</sup>. Regarding the BOSS-COMB sample, the mask is constructed to include every sector (each area covered by a unique set of plates) contained within the CMASS mask into the LOWZ counterpart. Furthermore, information regarding the BOSS fibre completeness is recorded within the MANGLE mask files. The completeness of every observed sector (indexed by  $j$ ) is then defined as follows:

$$C_{\text{BOSS},j} = \frac{N_{\text{obs},j} + N_{\text{cp},j}}{N_{\text{star},j} + N_{\text{gal},j} + N_{\text{fail},j} + N_{\text{cp},j} + N_{\text{missed},j}}, \quad (2)$$

where

$N_{\text{gal},j}$  is the number of galaxies with redshifts from good BOSS spectra.

$N_{\text{star},j}$  is the number of spectroscopically confirmed stars.

$N_{\text{fail},j}$  is the number of objects with BOSS spectra for which either stellar classification or redshift determination failed.

$N_{\text{obs},j}$  is defined as  $N_{\text{obs},j} = N_{\text{gal},j} + N_{\text{star},j} + N_{\text{fail},j}$ .

$N_{\text{cp},j}$  is the number of objects with no spectra, in a fibre collision group where at least one other object of the same target class has a spectrum.

$N_{\text{missed},j}$  is the number of objects with no spectra, and if in a fibre collision group, there are no other objects from the same target class.

On the edges of the survey there are regions of high incompleteness. To avoid these sectors, where a substantial fraction of redshifts is absent, sectors with  $C_{\text{BOSS},j} < 0.7$  are excluded from analysis and are not considered within the scope of this work.

The LSS catalogues for LOWZ and CMASS do not include galaxy stellar masses. We obtain this information from the lightcones presented in Rodríguez-Torres et al. (2016) and Kitaura et al. (2016). The observed LRG stellar mass functions (SMFs) are shown in the left and centre panels of Fig. 3 for LOWZ and CMASS, respectively.

### 2.2 eBOSS LRG sample

We use the final data release of SDSS-IV eBOSS LRG (DR16; Prakash et al. 2016), which contains  $\sim 200\,000$  LRG over  $4000\text{ deg}^2$ . This sample spans a higher redshift range than the CMASS sample, covering  $0.6 < z < 1$  (Ross et al. 2020). The spatial distribution and number density of this eBOSS sample can be seen in Figs 1 and 2. The final LSS catalogues for eBOSS are publicly available.<sup>2</sup>

In these catalogues, the total weight for every galaxy,  $i$ , as described in Ross et al. (2020) is

$$w_{\text{tot},i} = w_{\text{systot},i} w_{\text{cp},i} w_{\text{noz},i}. \quad (3)$$

where  $w_{\text{systot},i}$ ,  $w_{\text{cp},i}$ , and  $w_{\text{noz},i}$  are the same weights as described for BOSS.

The geometry of the sky area covered by the eBOSS sample is detailed within a MANGLE format mask<sup>2</sup>. The completeness for each sector is determined using the same methodology as employed for the BOSS samples. Following the observational systematics processing described in Ross et al. (2020), we apply a spectroscopic completeness threshold of  $C_{\text{eBOSS},j} < 0.5$  to exclude sectors with low completeness.

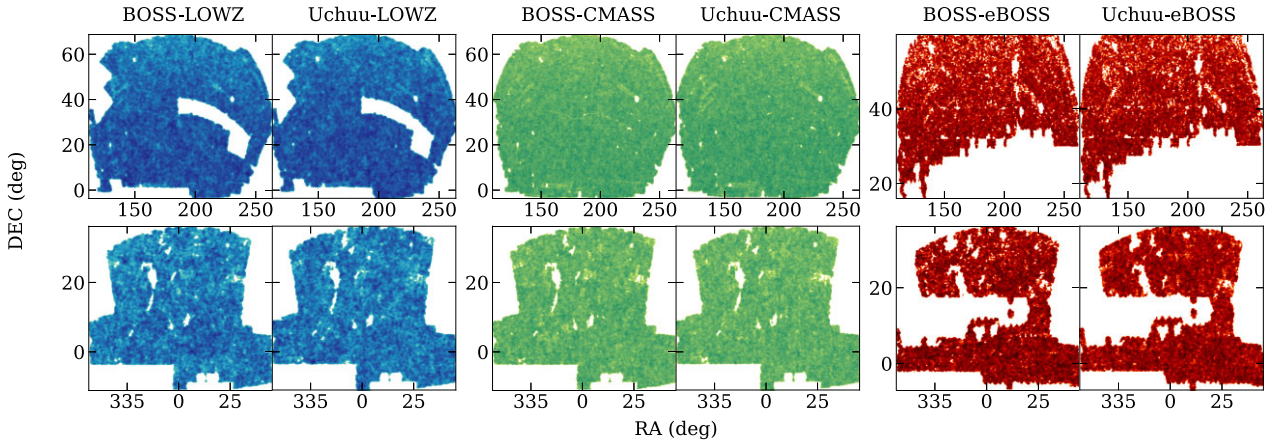
The SMF of the eBOSS galaxies is shown in the right panel of Fig. 3. Stellar masses are obtained from the data outputs provided by Comparat et al. (2017) through the use of the FIREFLY<sup>3</sup> code

<sup>1</sup>Data catalogues and MANGLE masks are available at [SDSS-BOSS Science Archive Server](https://archive.sdss.org/BOSS/).

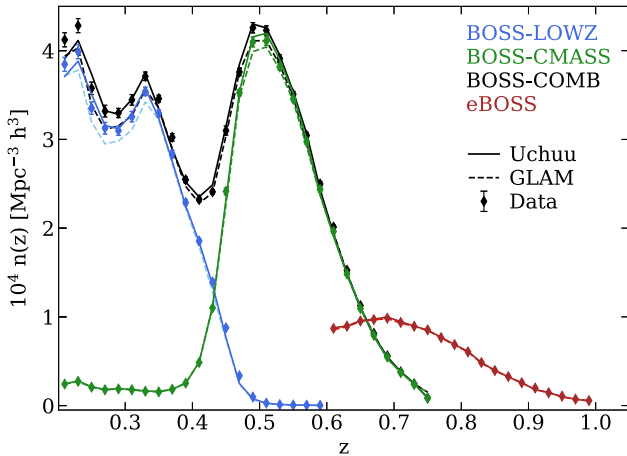
<sup>2</sup>Data catalogues and MANGLE masks are available at [SDSS-eBOSS Science Archive Server](https://archive.sdss.org/eBOSS/).

<sup>3</sup><http://www.icg.port.ac.uk/firefly/>





**Figure 1.** Sky area covered by the LOWZ, CMASS, and eBOSS data catalogues, and by the UCHUU-LOWZ, -CMASS, and -eBOSS lightcones. The upper and lower rows correspond to the Northern and Southern hemispheres, respectively. Colours indicate the angular number density normalized by the densest pixel. Darker colours represent regions of higher density.



**Figure 2.** The comoving galaxy number density in the three samples studied: LOWZ in blue, CMASS in green, and eBOSS in red. The observational estimates are depicted using diamonds along with error bars. The mean  $n(z)$  from both the UCHUU and GLAM-UCHUU lightcones are represented with solid and dashed lines, respectively (following the smoothing method described in the text). The error bars are obtained from the standard deviation of our set of 2000 GLAM-UCHUU lightcones for each of the samples. Additionally, for illustrative purposes, the  $n(z)$  of the BOSS-COMB sample is represented by a solid black line.

applied to the DR16 eBOSS spectra. To avoid any overlap with the BOSS-CMASS LRG, eBOSS adopted a specific magnitude cut-off  $i \geq 19.9$ . This limit results in the most massive LRG being excluded from the eBOSS sample. For details regarding the LRG selection, refer to Prakash et al. (2016).

### 3 COSMOLOGICAL SIMULATIONS

Using the UCHUU and GLAM simulations listed in Table 2, we generate galaxy lightcones as well as lightcones for covariance errors for each of the analysed samples: LOWZ, CMASS, COMB, and eBOSS. The goal of these lightcones is to faithfully reproduce the clustering statistics observed in the survey data.

#### 3.1 The UCHUU simulation

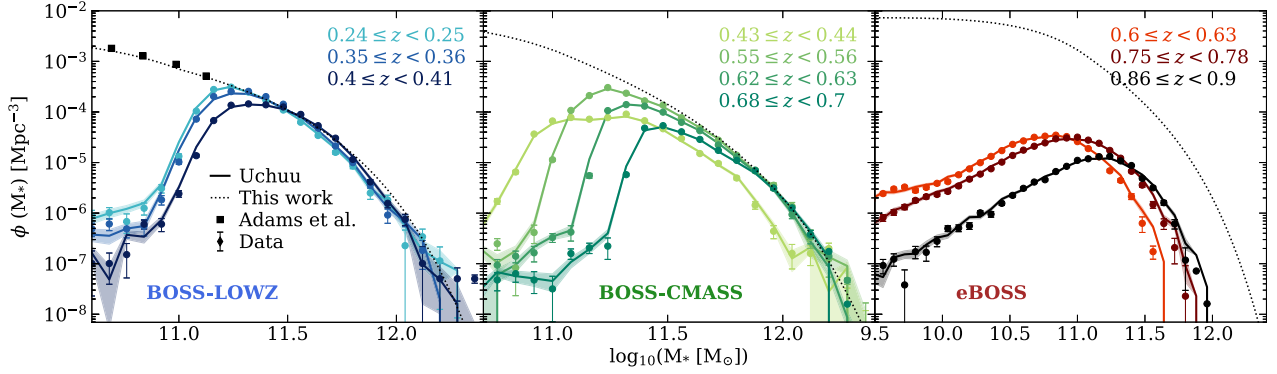
The UCHUU simulation<sup>4</sup> is the largest run within the UCHUU suite (Ishiyama et al. 2021), tracing the evolution of  $12,800^3$  dark matter particles, each with a mass of  $3.27 \times 10^8 h^{-1} M_\odot$ , within a periodic box of  $2.0 h^{-1} \text{Gpc}$ . The cosmological parameters adopted are  $\Omega_{m0} = 0.309$ ,  $\Omega_{b0} = 0.0486$ ,  $\Omega_{\Lambda0} = 0.691$ ,  $h = 0.677$ ,  $n_s = 0.9667$ , and  $\sigma_8 = 0.816$ , representing the best fitting  $\Lambda\text{CDM}$  parameters corresponding to the Planck 2015 cosmology (Planck Collaboration XIII 2016). Employing the TreePM code GREEM (Ishiyama, Fukushima & Makino 2009; Ishiyama, Nitadori & Makino 2012) with a gravitational softening length of  $4.27 h^{-1} \text{kpc}$ , UCHUU tracks the gravitational evolution of particles from redshift  $z = 127$  to  $z = 0$ . A total of 50 particle snapshots were stored over the redshift interval from  $z = 0$  to  $z = 14$ . The identification of bound structures was performed using the ROCKSTAR phase-space halo/subhalo finder (Behroozi, Wechsler & Wu 2013a); subhaloes in UCHUU are 90 per cent complete down to  $V_{\text{peak}} \gtrsim 70 \text{ km s}^{-1}$ , while distinct haloes are 90 per cent complete down to  $V_{\text{peak}} \gtrsim 50 \text{ km s}^{-1}$ . Merger trees were then constructed using the CONSISTENT TREES code (Behroozi et al. 2013b). The properties of the UCHUU simulation are summarized in Table 2.

The high resolution of the UCHUU simulation makes it optimal for generating simulated galaxy lightcones. UCHUU is able to resolve dark matter haloes and subhaloes down to small masses, a key requirement for employing the SHAM. This capability results in high completeness, particularly in terms of the fraction of galaxies resolved at lower redshift. Furthermore, the large volume of UCHUU allows us to study large-scale clustering features, such as the BAO.

#### 3.2 The GLAM simulations

The GLAM simulations (Klypin & Prada 2018) are  $N$ -body cosmological simulations that follow the evolution of  $2000^3$  dark matter particles, each having a particle mass of  $1.06 \times 10^{10} h^{-1} M_\odot$ . All GLAM boxes are  $1.0 h^{-1} \text{Gpc}$  comoving periodic boxes, with  $N_s = 136$  time-steps and a mesh of  $N_g = 5800$  cells per side, resulting

<sup>4</sup>All UCHUU data products are publicly available at [Skies & Universes, \(0:sc\) Uchuu/\(0:sc\)Simulation](https://skies.universesim.org/UCHUU/0:sc/Simulation), including simulated catalogues constructed using various methods (Dong-Páez et al. 2022; Oogi et al. 2022; Aung et al. 2023; Prada et al. 2023).



**Figure 3.** The observed LRG SMF (points) in the LOWZ (left panel), CMASS (middle panel), and eBOSS (right panel) samples. The mean SMF of the UCHUU-LRG lightcones is shown using solid lines. We show the SMF for several redshift bins within the relevant redshift ranges for each sample, using a colour scheme indicated in each plot. The dotted curves show the complete SMFs adopted for the implementation of the SHAM method (see Table 4), while the square points denote the PRIMUS data taken from Moustakas et al. (2013, see Section 4.1). Data error bars represent the standard deviation,  $1\sigma$ , of our set of 2000 GLAM-UCHUU lightcones, while the shaded area for UCHUU represents the error on the mean,  $\sigma/\sqrt{2}$ .

**Table 2.** Numerical and cosmological parameters of the UCHUU and GLAM-UCHUU simulations used in this work. The columns, from left to right, include the simulation name, size of the simulated box  $L_{\text{box}}$  in  $h^{-1}\text{Mpc}$ , the number of particles  $N_{\text{part}}$ , the mass of the simulation particle  $m_{\text{part}}$  in  $h^{-1}M_{\odot}$  units, the cosmological parameters adopted in the simulation ( $h$ ,  $\Omega_{\text{m}0}$ ,  $\Omega_{\Lambda 0}$ ,  $\Omega_{\text{b}0}$ ,  $\sigma_8$ ), the mesh size  $N_{\text{g}}$ , the gravitational softening length  $\epsilon$  in  $h^{-1}\text{Mpc}$  units, the redshift of the initial linear power spectrum  $z_{\text{init}}$ , the number of time-steps  $N_{\text{s}}$ , and number of realizations  $N_{\text{r}}$ .

Simulation	$L_{\text{box}}$	$N_{\text{part}}$	$m_{\text{part}}$	$h$	$\Omega_{\text{m}0}$	$\Omega_{\Lambda 0}$	$\Omega_{\text{b}0}$	$\sigma_8$	$N_{\text{g}}$	$\epsilon$	$z_{\text{init}}$	$N_{\text{s}}$	$N_{\text{r}}$
UCHUU	2 000	12 800 <sup>3</sup>	$3.27 \times 10^8$	0.677	0.309	0.691	0.0486	0.816	468 384	0.00427	127	5 570	1
GLAM-UCHUU	1 000	2 000 <sup>3</sup>	$1.06 \times 10^{10}$	0.677	0.309	0.691	0.0486	0.816	5 800	0.172	104	136	2 000
GLAM-PMILL	1 000	2 000 <sup>3</sup>	$1.06 \times 10^{10}$	0.678	0.307	0.693	0.0481	0.828	4 000	0.250	104	136	100
GLAM-PMILLNOBAO	1 000	2 000 <sup>3</sup>	$1.06 \times 10^{10}$	0.674	0.307	0.693	0.0487	0.828	4 000	0.250	104	136	100
GLAM-ABACUS	1 000	2 000 <sup>3</sup>	$1.09 \times 10^{10}$	0.674	0.315	0.685	0.0493	0.812	7 500	0.133	104	136	100

in a spatial resolution of  $0.172h^{-1}\text{Mpc}$ . The initial conditions are generated using the Zeldovich approximation starting at  $z_{\text{init}} = 104$ . Because of the lower resolution of these simulations compared to UCHUU, the GLAM simulations are only capable of resolving distinct haloes (not subhaloes) with virial masses greater than  $10^{12}h^{-1}M_{\odot}$  (Hernández-Aguayo et al. 2021).

The GLAM simulations have been run assuming different cosmologies. The numerical and cosmological parameters adopted in these simulations are listed in Table 2 and summarized below:

- (i) GLAM-UCHUU: Adopts the same Planck cosmology (PL15) and linear power spectrum as the UCHUU simulation.
- (ii) GLAM-PMILL and -PMILLNOBAO: Adopt the cosmological parameters used in the Planck Millennium simulation (PMILL; Baugh et al. 2019), which uses the best-fitting  $\Lambda\text{CDM}$  parameters from the first Planck 2013 data release (PL13; Planck Collaboration XVI 2014). In PMILLNOBAO, an initial power spectrum that models a matter distribution without baryonic acoustic oscillations is used, leading to the absence of the BAO feature in these realizations.
- (iii) GLAM-ABACUS: Adopts the `abacus_cosm000` cosmology from the ABACUSMILLN-body simulation suite (Maksimova et al. 2021), which includes the effect of massive neutrinos.

## 4 CONSTRUCTING THE UCHUU-LRG LIGHTCONES

We will now outline the procedure employed to create LRG lightcones from the UCHUU simulation. This includes details about the SHAM method, modelling of the LRG stellar mass incompleteness, handling angular masks, and addressing fibre collisions. For each of

the 8 analysed samples, we generate 2 independent lightcones by duplicating the survey footprint across the sky, resulting in a total of 16 lightcones. The properties of the UCHUU lightcones are shown in Table 3.

To capture the redshift evolution of galaxy clustering, we use a series of UCHUU snapshots. For each sample, the redshifts of the boxes used are as follows:

LOWZ:  $z = 0.19, 0.25, 0.30, 0.36, 0.43$   
CMASS:  $z = 0.43, 0.49, 0.56, 0.63, 0.7$   
eBOSS:  $z = 0.63, 0.7, 0.78, 0.86$

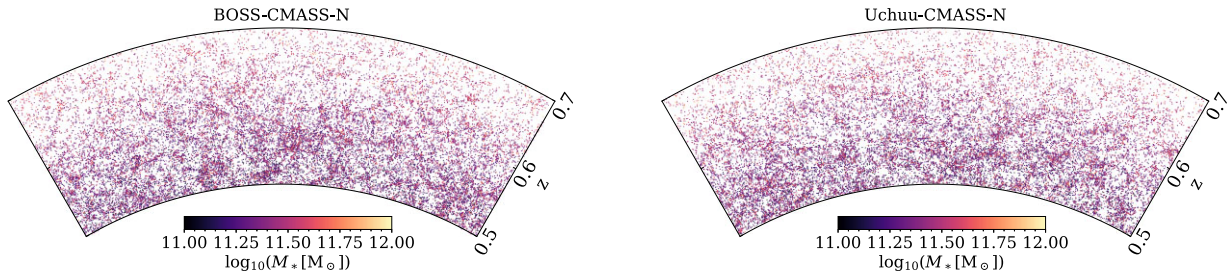
Fig. 4 shows a comparison between a slice of the CMASS-N galaxy distribution in the observations (left panel) and the corresponding UCHUU-LRG lightcone (right panel). Individual galaxies are colour-coded based on their stellar mass, highlighting the consistency between our simulated lightcones and the observed data. The decrease in galaxy density with increasing redshift matches the observed redshift distribution  $n(z)$  in this particular sample (see Fig. 2).

### 4.1 Stellar mass assignment

We adopt the SHAM method to populate galaxies within the dark matter haloes and subhaloes in each UCHUU simulation box. This technique assumes a monotonic relation between the stellar mass of galaxies and a specific property of the associated halo, which is a proxy of its mass, thereby assigning more massive galaxies to more massive (sub)haloes (Guo et al. 2016). In the standard SHAM approach, there is no distinction between haloes and subhaloes, since the assignment is only based on a subhalo property and not the environment of the subhalo (although the method has

**Table 3.** Properties of the UCHUU and GLAM-UCHUU lightcones generated for each observational sample. The first column indicates the simulation used. The second and third columns provide the name and number of lightcones ( $N_{\text{light}}$ ) defining each sample data set. The fourth column gives the redshift range considered for clustering studies. Subsequent columns provide the (weighted) mean number of galaxies, the mean effective area (in  $\text{deg}^2$ ), the mean density of galaxies per unit area (in  $\text{deg}^{-2}$ ), and the mean effective volume (in  $h^{-3} \text{Gpc}^3$ ).

Simulation	Sample	$N_{\text{light}}$	$z$ range	$N_{\text{eff}}$	$A_{\text{eff}}$	$N_{\text{eff}}/A_{\text{eff}}$	$V_{\text{eff}}$
UCHUU	LOWZ-N	2	0.2–0.4	193 973	5751	33.7	0.36
	LOWZ-S	2	0.2–0.4	89 944	2 470	36.4	0.16
	CMASS-N	2	0.43–0.7	602 206	6 798	88.8	1.27
	CMASS-S	2	0.43–0.7	221 921	2 482	89.3	0.47
	COMB-N	2	0.2–0.75	806 468	5 751	140.2	1.71
	COMB-S	2	0.2–0.75	351 348	2 470	142.3	0.74
	eBOSS-N	2	0.6–1.0	113 400	2 462	46.1	0.26
	eBOSS-S	2	0.6–1.0	71 424	1 621	44.1	0.16
GLAM-UCHUU	LOWZ-N	2 000	0.2–0.4	188 160	5 751	32.7	0.36
	LOWZ-S	2000	0.2–0.4	86 194	2470	34.9	0.16
	CMASS-N	2000	0.43–0.7	593 339	6798	87.3	1.26
	CMASS-S	2000	0.43–0.7	217 727	6798	87.7	0.46
	COMB-N	2000	0.2–0.75	787 474	5751	136.9	1.69
	COMB-S	2000	0.2–0.75	341 650	2470	138.3	0.73
	eBOSS-N	2000	0.6–1.0	112 259	2462	45.6	0.25
	eBOSS-S	2000	0.6–1.0	70 510	1621	43.5	0.17



**Figure 4.** Slices through the CMASS-N sample (left), and the UCHUU-CMASS-N lightcone (right), with galaxies colour coded according to their stellar mass as indicated in the legend. The displayed galaxies fall within the range of  $10 < \text{Dec} < 30$  deg, and  $150 < \text{RA} < 210$  deg.

been extended to include such information, see Contreras et al. (2021).

For the SHAM technique, we use the peak value of the maximum circular velocity over the history of the (sub)halo,  $V_{\text{peak}}$ , as a proxy for the mass. This choice of property has been shown to yield galaxy catalogues that exhibit better agreement with observational data (as discussed in Trujillo-Gomez et al. 2011; Reddick et al. 2013; Guo et al. 2016). Several reasons underpin this improvement: (i) circular velocity parameters are typically reached at one-tenth of the (sub)halo’s radius, which provides a more accurate characterization of the scales directly influencing galaxies (Chaves-Montero et al. 2016); (ii)  $V_{\text{peak}}$  is better defined for both haloes and subhaloes when compared to halo mass, as the latter becomes ambiguous for subhaloes. Furthermore, subhalo mass is also influenced by the choice of halo finder, as it depends on a code-defined truncation radius at which the particles are defined as unbound and are removed (Trujillo-Gomez et al. 2011). (iii) the circular velocity is less affected than a mass measurement by the mass-loss experienced by a halo or subhalo after it becomes part of a large structure (Hayashi et al. 2003; Kravtsov et al. 2004).

It is important to note that while the original SHAM approach assumes a monotonic relation between the galaxy stellar mass and  $V_{\text{peak}}$ , observations suggest that this assignment cannot be strictly one to one. To achieve more realistic results, it becomes necessary to introduce scatter into the abundance matching process (e.g. Skibba et al. 2011; Trujillo-Gomez et al. 2011). We adopt the scatter

scheme proposed in Rodríguez-Torres et al. (2016), which defines the scattered  $V_{\text{peak}}$  as the abundance matching parameter:

$$V_{\text{peak}}^{\text{scat}} = (1 + \mathcal{N}(0, \sigma_{\text{SHAM}})) V_{\text{peak}}, \quad (4)$$

where  $\mathcal{N}$  represents a Gaussian-distributed random number with a mean of zero and standard deviation  $\sigma_{\text{SHAM}}$  (the SHAM scatter parameter). The optimal values for the scatter parameters are determined by optimizing the goodness of fit between the 2PCFs measured from the observational samples and those obtained from the UCHUU lightcones. Note that the footprint and number density of these lightcones are set beforehand to match the data, as described in Section 4.3. The mean SHAM scatter parameters over the boxes of each sample are as follows:  $\bar{\sigma}_{\text{LOWZ}} = 0.26$ ,  $\bar{\sigma}_{\text{CMASS}} = 0.30$ ,  $\bar{\sigma}_{\text{eBOSS}} = 0.65$ , all in agreement with previous work (see Nuza et al. 2013; Rodríguez-Torres et al. 2016; Yu et al. 2022). We note that these values increase with increasing redshift. The scatter is not necessarily directly linked to redshift. However, redshift could influence the halo-galaxy connection due to the dynamic processes of galaxy formation and evolution. In addition, as redshift values increase, observations might encounter higher levels of noise and limitations from instruments. The inferred value for the eBOSS scatter parameter seems substantially higher than the values recovered for the other samples that may indicate that it is influenced by the high incompleteness in the SMF for this sample.

In order to implement the SHAM technique in the UCHUU boxes, we require a complete SMF, which accurately represents a complete



**Table 4.** Fitted parameters of the Schechter functions adopted in this work as complete SMFs for the LOWZ, CMASS, and eBOSS samples. The characteristic stellar mass,  $M_*$ , is given in units of  $M_\odot$ , and  $\phi_*$  is given in units of  $\text{Mpc}^{-3} (\log_{10} M_*)^{-1}$ .

Sample	Mass range	$10^3 \phi_*$	$\alpha$	$\log_{10} M_*$
LOWZ	$M_* \leq 10^{11.3}$	2.894	−0.996	10.82
	$M_* > 10^{11.3}$	0.476	−1.557	11.23
CMASS & eBOSS	$M_* \leq 10^{11.0}$	4.002	−0.938	10.76
	$M_* > 10^{11.0}$	0.266	−2.447	11.42

population of galaxies prior to applying any of the magnitude and colour selections specific to the LRG samples considered here. Due to the selection function in the BOSS and eBOSS surveys, the SMF becomes incomplete in these samples at low masses (see Fig. 3). For CMASS and eBOSS, we adopt the Schechter function taken from Rodríguez-Torres et al. (2016) as the complete SMF, which is shown by the dashed line in Fig. 3. These authors combine the CMASS SMF, which is complete at higher stellar masses, with the SMF from the PRIMUS survey at lower masses (Moustakas et al. 2013). In the case of LOWZ, we also incorporate the averaged SMF measurements from PRIMUS for the redshift ranges  $0.2 < z < 0.3$  and  $0.3 < z < 0.5$  (square symbols). Subsequently, we perform a fit to both data sets using a Schechter function. All the fits are shown as dotted lines in Fig. 3. The fitted parameters are given in Table 4.

Once we have the complete SMF fits and the values of  $V_{\text{peak}}^{\text{scat}}$  for each halo and subhalo in UCHUU (i.e. after including the scatter), we are able to apply the SHAM technique. The procedure is as follows:

- (i) Compute the cumulative number density of (sub)haloes as a function of decreasing  $V_{\text{peak}}^{\text{scat}}$ , i.e. the scattered value of the circular velocity.
- (ii) Compute the cumulative number density of galaxies as a function of decreasing stellar mass,  $M_*$ , using the adopted complete SMF.
- (iii) Construct a monotonic relation between the cumulative number density functions from steps (i) and (ii):

$$n_{\text{gal}}(> M_*) = n_{\text{halo}}(> V_{\text{peak}}^{\text{scat}}), \quad (5)$$

which implies that a (sub)halo with  $V_{\text{peak},i}^{\text{scat}}$  will contain a galaxy with stellar mass  $M_{*,i}$ , assigning the most massive galaxy to the (sub)halo with the highest  $V_{\text{peak}}^{\text{scat}}$ . We note that this assignment is monotonic between  $V_{\text{peak}}^{\text{scat}}$  and  $M_*$ , but not between  $V_{\text{peak}}$  and  $M_*$ .

- (iv) Repeat this sequence for every UCHUU box.

The next step is to model the incomplete stellar mass distribution observed in all samples. For UCHUU-LOWZ and -CMASS galaxy boxes, we do this by randomly down-sampling galaxies from the complete SMFs. The approach is different for the case of eBOSS, as the SMF is highly incomplete in this instance for the whole stellar mass range (see the right panel of Fig. 3). To account for this effect, we rely on the results presented in Alam et al. (2020), where they report that the probability of a distinct halo containing a central LRG is at most 30 per cent. Therefore, after applying the SHAM on the UCHUU-eBOSS boxes, we randomly select 30 per cent of the galaxies. The remainder of the observed incompleteness, primarily located at the lower end of stellar masses, is acquired through the same methodology as applied in the LOWZ and CMASS samples. The final SMFs of our UCHUU-LRG lightcones are shown by the solid curves in Fig. 3. We validate our scheme for assigning stellar masses to galaxies by measuring the galaxy clustering in the lightcones as a function of stellar mass, as described below.

## 4.2 Creating the lightcones by joining the cubic boxes

The next step in the production of the UCHUU lightcones involves joining the cubic boxes that have been populated with galaxies, into spherical shells (see Smith et al. 2022, for a detailed description of this method).

First, we position the observer at  $(0, 0, 0) h^{-1} \text{Gpc}$  in comoving coordinates and transform the Cartesian coordinates  $(x, y, z)$  of each galaxy within the cubic box to redshift-space coordinates  $(R_a, \text{Dec}, z)$ , considering the effect of galaxy peculiar velocities. Each box is then divided into spherical shells, where the comoving distance between the observer and the inner/outer edges of the shell corresponds to the redshift that lies midway between this snapshot and the next/previous snapshot. These spherical shells are then joined together to build three full-sky lightcones: UCHUU-LOWZ, -CMASS, and -eBOSS. To cover their sky areas, it becomes necessary to periodically replicate the corresponding galaxy cubic boxes at specific redshift ranges. If the centre of a halo lies within a given spherical shell, the entire halo is included in the lightcone. However, if a halo partially extends into a shell but its centre falls outside the shell, it is excluded from the lightcone.

## 4.3 Radial selection function and angular mask

We randomly downsample the galaxy population in the redshift ranges where the lightcone's  $n(z)$  surpasses the observed one. By doing this, we also implement a smooth transition between the spherical shells, avoiding a sudden step-like behaviour. The redshift distribution of galaxies in the UCHUU lightcones resulting from this process is shown by the solid lines in Fig. 2.

We proceed to match the sky area of the full-sky UCHUU lightcones with that of the observed samples, using the available survey masks (see Sections 2.1 and 2.2). Additionally, we apply various veto masks. These include masks for bright stars, bright objects, and non-photometric conditions (in the real data). Moreover, we incorporate veto masks based on the seeing during the imaging data acquisition and Galactic extinction (avoiding high-extinction areas), see Ross et al. (2017) and Ross et al. (2020) for details. The result of applying these masks and the comparison with observed data is shown in Fig. 1. To improve our statistics and benefiting from our lightcones (without masks) covering the whole sky area, we repeat this procedure once more by shifting the sky position of the masks by 180 deg. By doing this, we are able to obtain two independent UCHUU lightcones for each sample.

As shown in Table 3, the effective area of the UCHUU lightcones is slightly smaller than that of the actual data ( $\sim 1$  per cent lower). The angular geometry of the samples is represented by MANGLE polygons, which lack the finest angular footprint details, contributing to this minor variation in the effective area.

## 4.4 Completeness and fibre collisions

A final step is required to fully replicate the observed samples, which involves the incorporation of survey systematics. This includes the angular completeness of the masks,  $C$ , defined in equation (2). To account for this property, we downsample the regions where the completeness falls below one, retaining only those regions where  $C \geq 0.7$  for UCHUU-LOWZ and -CMASS, and  $C \geq 0.5$  for UCHUU-eBOSS.

Another observational artefact is the so-called fibre collisions. This is due to the finite size of the fibre-head on the plate. If two galaxies are separated by less than 62 arcsec apart (which represents the fibre

collision radius), it is possible that only one of these galaxies will be assigned a spectroscopic fibre, resulting in the inability to determine the redshift of the other galaxy (Reid et al. 2015). The distribution of such closely spaced galaxy pairs is expected to be correlated with galaxy density. To model this effect, we adopt the method described in Guo, Zehavi & Zheng (2012), where the main goal is to divide the complete galaxy sample into two distinct populations:

*Population 1 (P1):* A subsample where no galaxy is within the fibre collision radius of any other galaxy within this same subsample.

*Population 2 (P2):* A subsample including all galaxies that are not of P1. Every galaxy in this population lies within the fibre collision radius of a galaxy in P1. All the galaxies affected by fibre collisions belong to this subsample.

The procedure for creating these two populations is as follows:

- (i) Initially, all galaxies in the lightcone are designated as members of P1.
- (ii) For each pair of collided galaxies in P1, we randomly select one galaxy to be reassigned to P2.
- (iii) For each group of three or more collided galaxies in P1, we chose the galaxy that collides with the most P1 galaxies and assign it to P2.
- (iv) We repeat steps (ii) and (iii) until no galaxies in P1 collides with each other.

Once the two populations have been created, we proceed to randomly select from P2 the galaxies that will not be observed. We do this for all the UCHUU lightcones, resulting in the final percentages of fibre-collided galaxies as follows: 1.3 per cent (2.9 per cent) for LOWZ-N (S) lightcones, 5.0 per cent (4.3 per cent) for CMASS, and 3.4 per cent (3.5 per cent) for eBOSS, which match the values in the observed samples (Reid et al. 2015). Once the unobserved galaxies are determined, we simply assign higher weights to the nearest galaxies that were assigned fibres, accounting for collided galaxies that were not assigned fibres (Nearest Neighbour Weights, NNW).

The NNW method, however, has a drawback: it ignores the correlation between observed and unobserved targets. To address this correlation properly, an approach known as the ‘pairwise inverse probability’ (PIP) weighting scheme was proposed by Bianchi & Percival (2017). This method has been applied to the eBOSS data as demonstrated in Mohammad et al. (2020). While the authors acknowledged the challenges involved in the PIP process, they also noted that approximate methods like NNW yield sufficiently accurate results considering the statistical uncertainties present in the data. Moreover, applying PIP weights for the BOSS samples would require further information that is not publicly available. Given these considerations, we have decided not to apply PIP weights to our lightcones. It is on scales smaller than  $1h^{-1}\text{Mpc}$  where these weights could have the most significant impact. Consequently, we may expect discrepancies in the clustering of galaxies at these pair separations between the results from observations and simulations.

#### 4.5 The UCHUU-COMB lightcone

We now describe the process of generating the UCHUU-COMB lightcones. To achieve this, we combine our UCHUU-LOWZ and -CMASS lightcones and apply the corresponding COMB MANGLE mask. When combining galaxy populations with different clustering amplitudes, it is optimal to assign weights to each sample to account for these differences (Percival, Verde & Peacock 2004). In this work, we count the number of galaxy pairs for various redshift bins within

each sample. We then normalize these values and use them as the weight for galaxies within that redshift range for each sample.

## 5 CONSTRUCTING THE GLAM-UCHUU COVARIANCE LIGHTCONES

To estimate covariance errors, we need to significantly increase the number of simulated lightcones. In this section, we detail the construction of the GLAM-UCHUU lightcones: a set of 16 000 lightcones (2000 lightcones for each of the eight studied samples) derived from the GLAM-UCHUUN-body simulations. The methodology employed for generating these lightcones is described below, while the properties of the produced covariance lightcones are presented in Table 3.

Similarly to the UCHUU lightcones, we use several snapshots from the GLAM simulations to reproduce the redshift-dependent evolution of galaxy clustering. We select the GLAM boxes that closely match the redshifts of the corresponding UCHUU snapshots:

LOWZ:  $z = 0.20, 0.27, 0.31, 0.37, 0.46$

CMASS:  $z = 0.46, 0.49, 0.564, 0.62, 0.72$

eBOSS:  $z = 0.62, 0.72, 0.79, 0.86$

### 5.1 Halo occupation distribution

As described in Section 3.2, the GLAM simulations do not resolve substructure inside distinct haloes. Therefore, the SHAM method discussed in Section 4.1 cannot be directly applied to populate galaxies within haloes in the GLAM simulations. Instead, a statistical method using the halo occupancy of galaxies from the UCHUU LRG catalogues must be employed. The HOD describes the mean number of galaxies within a distinct halo of virial mass  $M_{\text{halo}}$ , which can be written as the sum of the mean number of central and satellite galaxies:

$$\langle N_{\text{gal}}(M_{\text{halo}}) \rangle = \langle N_{\text{cen}}(M_{\text{halo}}) \rangle + \langle N_{\text{sat}}(M_{\text{halo}}) \rangle. \quad (6)$$

In this context, we obtain the HOD from the UCHUU galaxy cubic boxes we construct in Section 4.1. This is possible because each galaxy provides information whether it is a central or satellite, as well as the mass of the distinct halo it inhabits. We then use the set of 2000 GLAM-UCHUU halo cubic boxes available at each redshift to generate galaxy catalogues. In this process, we randomly select galaxies for each distinct halo, following the corresponding HOD as determined from the UCHUU-LRG samples. The UCHUU-LRG HODs for LOWZ, CMASS and eBOSS are shown in Section 6.2. GLAM-UCHUU simulations resolve haloes with masses within  $10^{12}$  and  $10^{15}$ . While the HODs obtained from the UCHUU-LRG samples do not extend significantly outside this mass range, our GLAM-UCHUU boxes end up having a marginal deficit of galaxies compared to UCHUU due to their incompleteness (and not-well resolved haloes) below  $\sim 10^{12.5}$  and above  $\sim 10^{14.5}$ . This effect is visible in Table 3 (and Fig. 2), where the  $\text{GLAM}N_{\text{eff}}(n(z))$  is slightly below that of the data and UCHUU.

### 5.2 Stellar mass assignment and generation of lightcones

The galaxies added into the GLAM-UCHUU boxes do not have assigned stellar masses. Unlike the SHAM approach, which inherently accounts for this property, we must incorporate stellar masses independently. To achieve this, we derive the probability distribution of central and satellite stellar masses from UCHUU galaxy boxes across various ranges of distinct halo masses. These distributions



are then employed to determine the stellar masses of the GLAM galaxies. The resulting SMFs are, as designed, consistent with those of UCHUU.

The subsequent step in generating the GLAM-UCHUU LRG lightcones involves joining the galaxy boxes into spherical shells and smoothing their number density,  $n(z)$ . This procedure is explained in Sections 4.2 and 4.3. By applying the available survey and veto masks, we match the area of the lightcones to that of the observed samples (without shifting the sky position of the masks). As a last step, following the process described in Section 4.4, we account for survey systematics. The creation of the GLAM-UCHUU-COMB lightcones follows the steps described in Section 4.5.

## 6 RESULTS

Here, we present the results of the galaxy clustering analysis conducted both in configuration and Fourier space. These analyses require the application of appropriate weights to both the data and the simulated lightcones.

For the LOWZ, CMASS, and COMB data, we apply the  $w_{\text{tot},i}$  weights as outlined in equation (1). The weights used for the eBOSS data are specified in equation (3). Incorporating the fibre collision weights, described in Sections 4.4 and 5.2, is essential for the UCHUU and GLAM-UCHUU lightcones. To enhance the signal-to-noise ratio of the clustering measurements, galaxies from both the data and lightcones are additionally weighted based on the galaxy number density,  $n(z)$ , using the weights introduced by Feldman, Kaiser & Peacock (1994) (hereafter FKP weights):

$$w_{\text{FKP}} = \frac{1}{1 + n(z)P_{\text{FKP}}}. \quad (7)$$

We set  $P_{\text{FKP}} = 10^4 h^{-3} \text{Mpc}^3$ , which is close to the value of the BOSS and eBOSS LRG power spectrum at  $k = 0.1 h^{-1} \text{Mpc}$ , the scale where we aim to minimize the variance (see Ross et al. 2017, 2020). The  $n(z)$  is estimated for each case from either the data or the lightcones. The random catalogues required for these analyses are obtained from the SDSS-BOSS Science Archive Server and SDSS-eBOSS Science Archive Server. Randoms are weighted only by FKP weights.

The redshift ranges considered for each sample, as well as those shown in the following figures, are listed in Table 1:  $0.2 < z < 0.4$  for LOWZ,  $0.43 < z < 0.7$  for CMASS, and  $0.6 < 1.0$  for eBOSS. In the case of COMB galaxies, we adopt the same analysis scheme as presented by Alam et al. (2017), and consider three redshift ranges:  $0.2 < z < 0.5$ ,  $0.4 < z < 0.6$  and  $0.5 < z < 0.75$ .

### 6.1 Testing the Planck cosmology against the data with UCHUU

In this section, we assess the performance of our theoretical predictions derived from the Planck cosmology, as determined from our UCHUU lightcones, to reproduce the observed data. We start by studying the LOWZ-, CMASS-, and eBOSS-N samples and conclude with an evaluation of the COMB-N+S sample.

#### 6.1.1 Clustering in configuration space

We first study the anisotropic 2PCF in redshift space,  $\xi(s, \mu)$ , which is computed using the Landy & Szalay (1993) estimator. We refer to the observed or simulated data galaxies as  $D$  and the random galaxies as  $R$

$$\xi(s, \mu) = \frac{DD - 2DR + RR}{RR}, \quad (8)$$

where  $s$  represents the redshift-space separation between a pair of objects in units of  $h^{-1} \text{Mpc}$  and  $\mu$  is the cosine of the angle between the pair separation vector and the line of sight. Quantities  $DD$ ,  $DR$ , and  $RR$  are the normalized pair counts of the galaxy–galaxy, galaxy–random, and random–random catalogues, respectively. For a more exhaustive analysis, we decompose  $\xi(s, \mu)$  into its Legendre multipoles:

$$\xi_\ell(s) = \frac{2\ell + 1}{2} \int_{-1}^1 \xi(s, \mu) P_\ell(\mu) d\mu, \quad (9)$$

where  $P_\ell(\mu)$  is the  $\ell$ th Legendre polynomial.

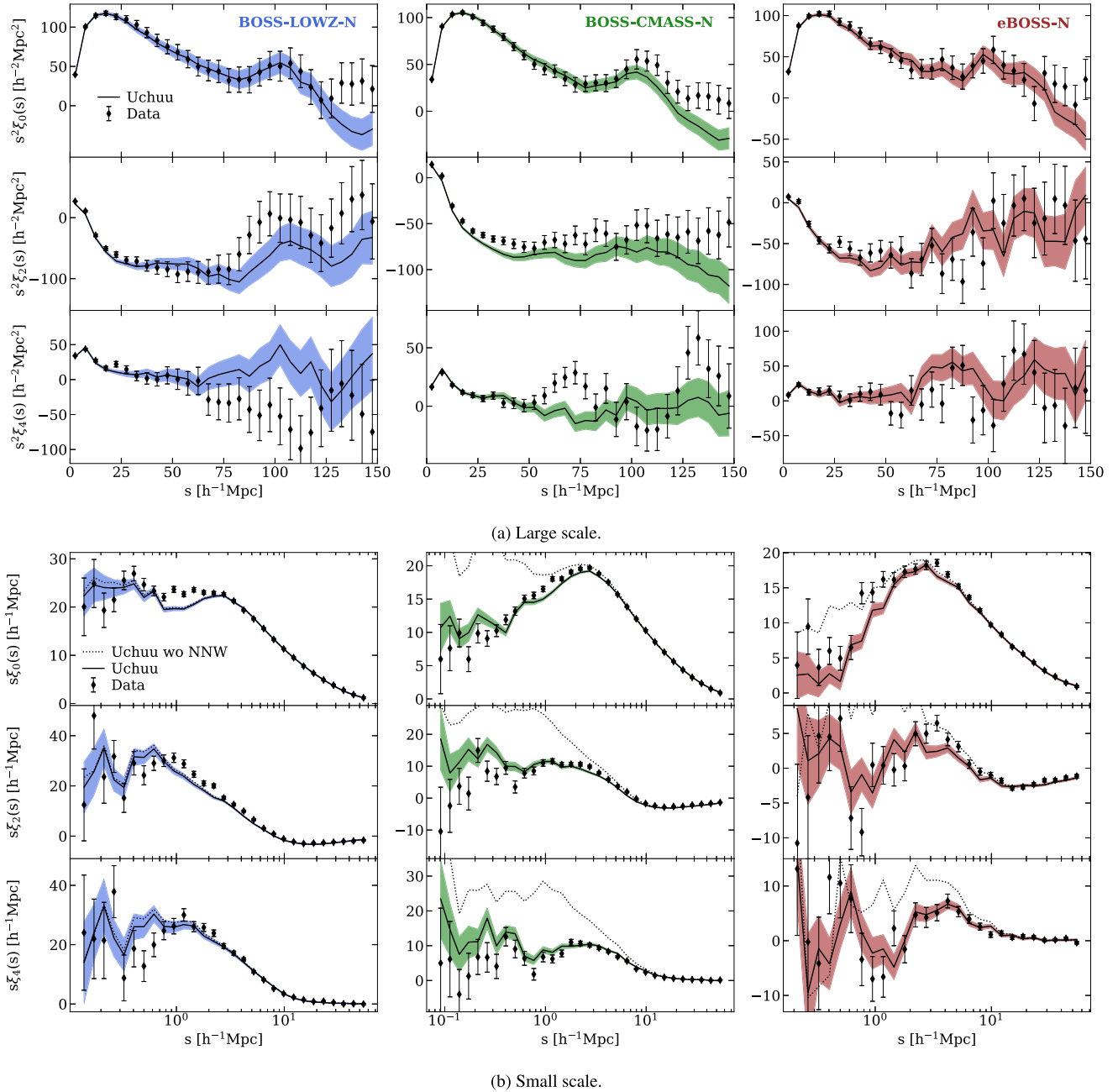
In Fig. 5, we present the monopole ( $\ell = 0$ ), quadrupole ( $\ell = 2$ ), and hexadecapole ( $\ell = 4$ ) moments of the 2PCF for the three samples: LOWZ, CMASS, and eBOSS-N. The results include both observed data and theoretical predictions determined from the mean of the UCHUU lightcones. Overall, our UCHUU lightcones are able to reproduce the clustering measurements obtained from the observed data, showcasing remarkable accuracy in certain samples. Residuals from the 2PCF monopole, denoted as  $r = \xi_0^{\text{UCHUU}} / \xi_0^{\text{data}} - 1$ , maintain values under 2.5 per cent across all samples ranging from the fibre collision scale<sup>5</sup> up to  $25 h^{-1} \text{Mpc}$ . Within the range of  $100 h^{-1} \text{Mpc}$ , residuals remain below 10 per cent for LOWZ and CMASS, and under 15 per cent for eBOSS.

Beyond the BAO scale ( $s \geq 120 h^{-1} \text{Mpc}$ ), significant fluctuations become apparent due to cosmic variance. At the smallest scales, below  $\sim 1 h^{-1} \text{Mpc}$ , deviations between the measurements from the lightcones and the observed data can be attributed to the non-application of the PIP weights in the lightcones, as mentioned in Section 4.4. Notably, there are differences between the results by more than  $1\sigma$  for some scales above  $60 h^{-1} \text{Mpc}$ . Moreover, while our UCHUU lightcones monopole cross zero after BAO peak, the observations do not. It is widely recognized that, at large scales, the measurements are significantly influenced by observational systematics. Huterer, Cunha & Fang (2013) conducted an exhaustive investigation into photometric calibration errors and their repercussions on clustering measurements, illustrating that calibration uncertainties consistently give rise to large-scale power variations. The significance and potential causes of the large-scale excess are studied also in Ross et al. (2017), where they demonstrate that it has no significant impact on BAO measurements. Their findings indicate that these measurements remain resilient, both in lightcones and the data, regardless of whether or not any weights are included.

One may also discern some differences between the results for the multipole moments that cannot be accounted for by the aforementioned reasons, such as the CMASS quadrupole. Exploring variations of the only free parameter in the SHAM,  $\sigma_{\text{SHAM}}$ , does not seem to meaningfully impact the shape of multipoles, suggesting that the observed discrepancies are most likely a result of statistical fluctuation. This is supported by the improved performance of our 2000 GLAM-UCHUU lightcone sets (see Section 6.2.1). All the highlighted trends (including the CMASS quadrupole behaviour) are in agreement with results presented in e.g. Rodríguez-Torres et al. (2016), Kitaura et al. (2016), or Ross et al. (2017).

Using the Landy–Szalay estimator, we now compute the two-dimensional correlation function  $2\text{PCF}_{2d} = \xi(r_p, \pi)$ , where  $r_p$  and  $\pi$  represent the pair separations perpendicular and parallel to the

<sup>5</sup>We define the fibre collision scale as the  $s$ -value at which the ratio between the multipoles of UCHUU with and without NNW is greater than 2 per cent. We set these scales as the minimum, since below this value the accuracy of our results decreases. For 2PCF monopole, it is  $2.3 h^{-1} \text{Mpc}$  in LOWZ,  $4.2 h^{-1} \text{Mpc}$  in CMASS, and  $6.5 h^{-1} \text{Mpc}$  in eBOSS.

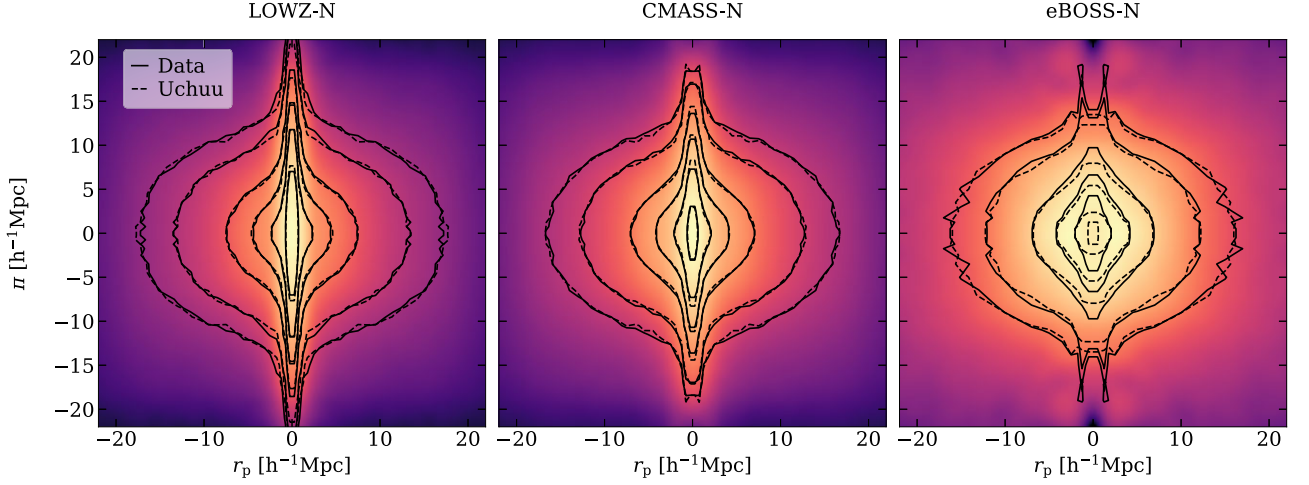


**Figure 5.** Measurements of the monopole, quadrupole, and hexadecapole moments of the 2PCF are presented in the first, middle, and bottom rows, respectively, for CMASS-N, LOWZ-N, and eBOSS-N, arranged from left to right in the columns. Observational measurements are shown by data points, while the solid lines show the mean 2PCF derived from the UCHUU lightcones. Data error bars represent the standard deviation of our set of 2000 GLAM-UCHUU lightcones at the  $1\sigma$  level, whereas the shaded area for UCHUU indicates the error on the mean  $\sigma/\sqrt{2}$ . Panel (a) shows the 2PCF measurements spanning from 0 to  $150 h^{-1}$  Mpc in bins of  $5 h^{-1}$  Mpc. Panel (b) presents the results up to  $60 h^{-1}$  Mpc in bins equally spaced in the logarithm of pair separation. For reference, the mean UCHUU 2PCF without fibre collision correction (NNW) is depicted with dotted lines.

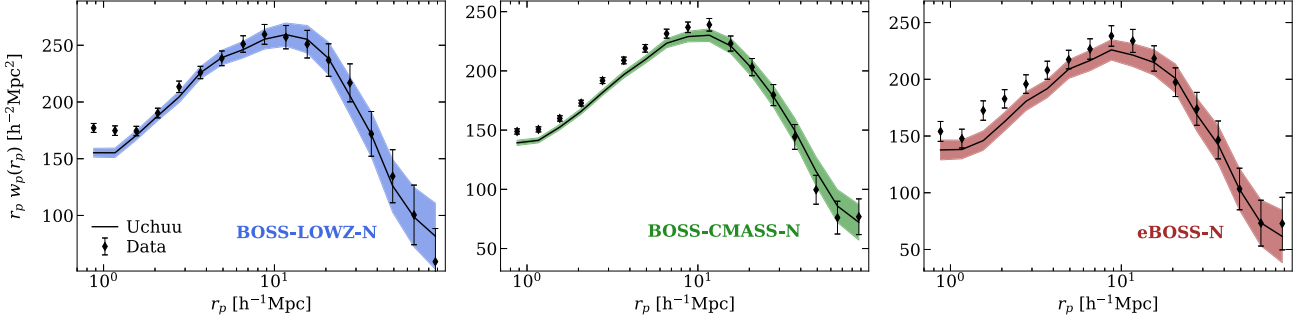
line of sight, respectively. The results are presented in Fig. 6, where the mean 2PCF<sub>2d</sub> derived from UCHUU-LOWZ, -CMASS, and -eBOSS-N lightcones is shown in the left, middle, and right columns, respectively. Solid curves represent the contours estimated from the observed data, while dashed curves depict the contours from UCHUU.

The 2PCF<sub>2d</sub> is the simplest statistical measure of peculiar velocities in cosmological structure. This function remains directionally independent in an isotropic Universe. However, this property does

not hold in redshift space, where only radial separations,  $\pi$ , are distorted by the peculiar velocities of the galaxies. This distortion is commonly referred to as redshift space distortions (RSD). The impact of RSD is particularly prominent on small scales, such as within galaxy groups and clusters, leading to the well-known ‘fingers-of-God’ effect visible in cone plots showing galaxy positions (Peacock et al. 2001). The results shown in Fig. 6 clearly exhibit the presence of RSD, evident by the elongations along the  $\pi$  axis at scales within  $5 h^{-1}$  Mpc. The remarkable agreement between our



**Figure 6.** The mean  $2PCF_{2d}$  from UCHUU-LOWZ-N (left column), -CMASS-N (middle column), and -eBOSS-N lightcones (right column) is plotted as a function of pair separation perpendicular ( $r_p$ ) and parallel ( $\pi$ ) to the line of sight. To illustrate deviations from circular symmetry, the first quadrant is replicated with reflections in both axes. Contours of the  $2PCF_{2d}$  estimated from the observations and from UCHUU are displayed as black solid and dashed curves, respectively. The contour levels shown are [0.5, 0.8, 2, 4, 8, 20].



**Figure 7.** Measurements of the 2PCF for CMASS-N (left column), LOWZ-N (middle column), and eBOSS-N (right column). Measurements from observations are shown by data points, while the solid lines represent the mean 2PCF derived from the UCHUU lightcones. Data error bars represent the standard deviation of our set of 2000 GLAM-UCHUU lightcones,  $1\sigma$ , while the UCHUU shaded area is the error on the mean, calculated as  $\sigma/\sqrt{2}$ . The 2PCF is measured up to  $100 h^{-1} \text{Mpc}$  in equally spaced logarithmic bins.

UCHUU lightcones and the BOSS/eBOSS data is noteworthy. Minor deviations on small scales, related to the non-use of PIP weights, are consistent with the previous results shown in Fig. 5.

As mentioned above, the impact of RSD only affects the pair separation along the line of sight ( $\pi$ ). Therefore, these distortions can be removed by projecting  $\xi(r_p, \pi)$  on to the  $\pi$ -axis (Davis & Peebles 1983; Norberg et al. 2009) as follows:

$$w_p(r_p) = \int_{-\infty}^{+\infty} \xi(r_p, \pi) d\pi. \quad (10)$$

In practice, this integral is truncated at a certain  $\pi_{\text{max}}$ . In this work, we adopt  $\pi_{\text{max}} = 80 h^{-1} \text{Mpc}$ . The projected correlation functions (pCF) are shown in Fig. 7. The agreement between the theoretical predictions determined from the mean of the UCHUU lightcones and the observational estimates aligns with the expectations from Fig. 6.

Finally, we analyse the three-point correlation function (3PCF) in redshift space,  $\zeta(r_{12}, r_{23}, r_{13})$ , which provides a description of the probability of finding triplets of galaxies as a function of their triangular geometry. We compute the 3PCF using ENCORE code<sup>6</sup>, which implements Szapudi & Szalay (1998) estimator:

$$\zeta(r_{12}, r_{23}, r_{13}) = \frac{DDD - 3DDR + 3DRR - RRR}{RRR}, \quad (11)$$

where  $r_{ij} = r_i - r_j$  represents the redshift-space separation between two objects,  $i$  and  $j$ , in units of  $h^{-1} \text{Mpc}$ .

Fig. 8 displays the 3PCF for both observed data and theoretical predictions, for LOWZ, CMASS and eBOSS-N samples. We find a good agreement between our theoretical predictions determined from the mean of the UCHUU lightcones and the observational estimates, with all the data points in agreement within  $1\sigma$  errors.

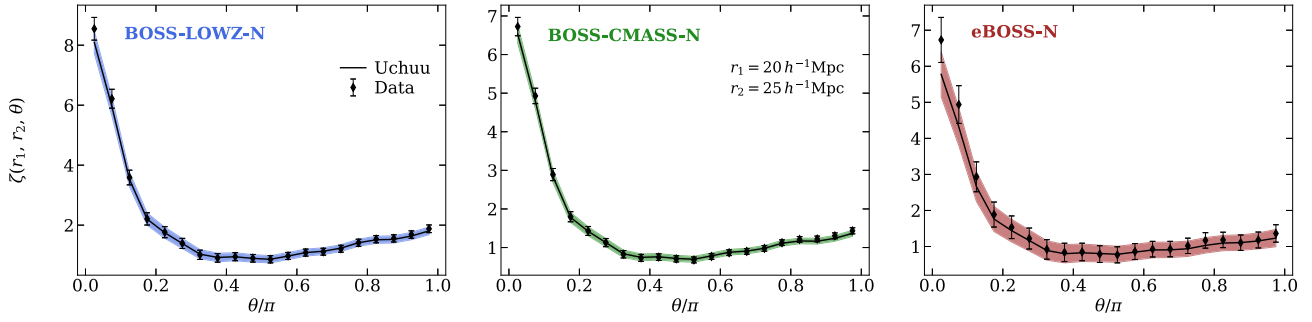
### 6.1.2 Clustering in Fourier space

We measure the power-spectrum monopole,  $P_0(k)$  and quadrupole,  $P_2(k)$ , which can be calculated following equations 2 to 7 from Feldman et al. (1994). These calculations have been carried out using the PYTHON package PYPOWER,<sup>7</sup> which implements the Hand et al. (2017) estimator. This estimator takes coordinate-space positions at mesh nodes and directly computes the normalization factor from the

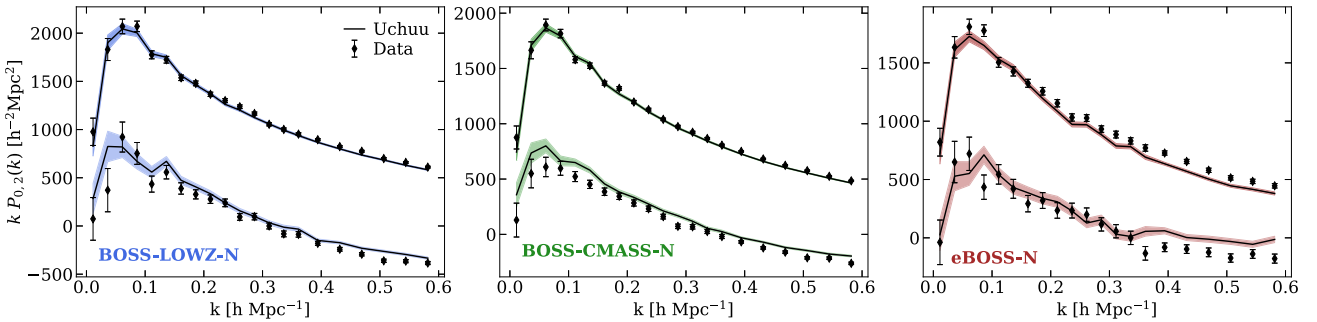
<sup>6</sup><https://github.com/oliverphilcox/encore>

<sup>7</sup><https://pypower.readthedocs.io/en/latest/api/api.html>





**Figure 8.** Measurements of the 3PCF for LOWZ-N (left column), CMASS-N (middle column), and eBOSS-N (right column). Measurements from observations are shown by data points, while the solid lines represent the mean 2PCF derived from the UCHUU lightcones. Data error bars represent the standard deviation of our set of 2000 GLAM-UCHUU lightcones,  $1\sigma$ , while the UCHUU shaded area is the error on the mean, calculated as  $\sigma/\sqrt{2}$ . The 3PCF is presented for  $r_1 = 20h^{-1}\text{Mpc}$  and  $r_2 = 25h^{-1}\text{Mpc}$ , and  $\theta$  is the angle between  $r_1$  and  $r_2$ , in radians.



**Figure 9.** Measurements of the power spectrum monopole (upper curve) and quadrupole (lower curve, lighter shading) for LOWZ-N (left column), CMASS-N (middle column), and eBOSS-N (right column). Observed data measurements are represented by data points. The solid lines represent the mean power-spectrum derived from the UCHUU lightcones. Data error bars represent the standard deviation ( $1\sigma$ ) using the set of 2000 GLAM-UCHUU lightcones, while the UCHUU-shaded area indicates the error on the mean,  $\sigma/\sqrt{2}$ . The power spectrum are measured from  $k = 0.005$  to  $0.6h\text{Mpc}^{-1}$  in  $0.03h\text{Mpc}^{-1}$  bins.

mesh. To mitigate the impact of aliasing, we employ a piecewise cubic spline (PCS) mesh assignment scheme with interlacing, as described by (Sefusatti et al. 2016), and use a grid resolution of  $N_{\text{grid}} = 1024$  in each dimension.

Fig. 9 shows the power spectrum monopole and quadrupole for both observed data and the mean of the UCHUU lightcones, with shot noise subtracted from the  $P_0(k)$  measurements. Our theoretical predictions based on the Planck cosmology successfully reproduce the observed power spectrum monopole, exhibiting residuals below 5.1 per cent for  $k > 0.03$  and  $k < 0.6h\text{Mpc}^{-1}$  in the LOWZ sample, under 4.0 per cent for CMASS, and within 12.5 per cent in eBOSS.

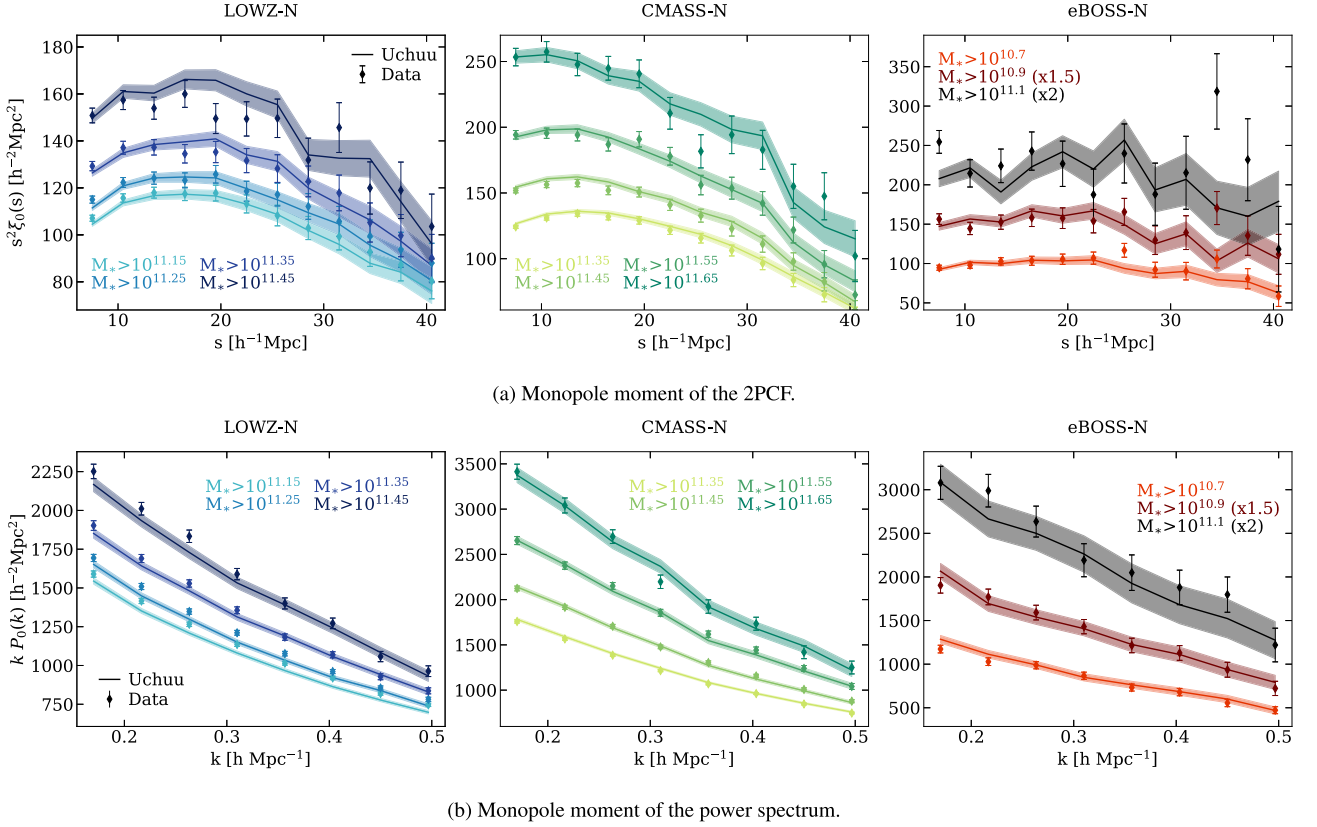
Many of the observed deviations can be attributed to statistical fluctuations. With our extensive set of 2000 GLAM-UCHUU lightcones, we are better equipped to accurately measure the noise (see Section 6.2.2). This enables us to discern where UCHUU predictions align with observations. We find some predictions that slightly differ from the observational data. One instance concerns the UCHUU-eBOSS monopole for  $0.2 < k < 0.4h\text{Mpc}^{-1}$ . Notably, a similar trend has been previously reported for the eBOSS data (see Zhao et al. 2021). Another case involves the quadrupole for  $k > 0.35h\text{Mpc}^{-1}$ . However, as illustrated in Fig. 5(b), effects of fibre collisions on quadrupole extends to scales up to  $15h^{-1}\text{Mpc}$ . This would explain the observed discrepancies in the UCHUU power-spectrum quadrupole at  $k \geq 0.4h\text{Mpc}^{-1}$ , since the accuracy of our results diminishes above these scales (below  $s = 15h^{-1}\text{Mpc}$ ).

### 6.1.3 Dependence of clustering on stellar mass

We now investigate the completeness of the observed galaxy populations by analysing the dependence of the 2PCF and power-spectrum monopoles on stellar mass. The results are shown in Fig. 10, where we observe that, for both the LOWZ and CMASS samples, a significant overlap exists between the observed data and the mean values obtained from the UCHUU lightcones across all the stellar mass thresholds. In agreement with previous studies (see Maraston et al. 2013), these results indicate that the LRG population in both BOSS samples is complete at the massive end. We conclude that the complete SMF assumed in the SHAM method for BOSS accurately describes the real galaxy populations.

For the eBOSS sample, consistent with results from Comparat et al. (2017), we consider an incomplete population of galaxies when generating the UCHUU-eBOSS lightcones. This is achieved by using a complete SMF that goes exceeds the SMF estimated from the data (see Section 4.1). Similar to the other samples, our goal is to assess whether the clustering as a function of stellar mass in UCHUU lightcones agrees with that estimated from the observed data. The results shown in the right panel of Fig. 10 demonstrate the accuracy of the method adopted in this work for modelling the incomplete eBOSS galaxy population. Despite the noise inherent to a low population statistic, there is a good agreement between the data and UCHUU, both samples presenting the same trend with stellar mass.

It is important to note that the eBOSS galaxy mass information was extracted from the Comparat et al. (2017) eBOSS catalogue,



**Figure 10.** Measurements of the 2PCF (a) and the power spectrum (b) monopoles for different stellar mass thresholds in LOWZ-N (left column), CMASS-N (middle column), and eBOSS-N (right column). The various stellar mass thresholds are indicated by the colour scheme in the legend. For a better visualization of the eBOSS sample, we have multiplied the 10.9 and 11.1 stellar mass thresholds measurements by 1.5 and 2, respectively. Observational measurements are shown by data points. The solid lines show the mean derived from the UCHUU lightcones. Data error bars represent the standard deviation of our set of 2000 GLAM-UCHUU lightcones at  $1\sigma$  level, while the shaded area for UCHUU represents the error on the mean,  $\sigma/\sqrt{2}$ . We measure the 2PCF monopole from  $s = 5$  to  $45 h^{-1} \text{ Mpc}$  in  $5 h^{-1} \text{ Mpc}$  bins, and the power spectrum monopole from  $k = 0.15$  to  $0.5 h \text{ Mpc}^{-1}$  in  $0.05 h \text{ Mpc}^{-1}$  bins.

which does not share the exact same footprint or  $n(z)$  as the public eBOSS data and UCHUU lightcones. Additionally, Comparat et al. (2017) does not provide its own random catalogue, leading us to use the public random catalogue for analysing the 2PCF. Again, this catalogue does not present the same footprint and  $n(z)$ . All these differences may introduce errors and discrepancies in the statistics. To mitigate these effects when computing the 2PCF for different stellar mass thresholds, we have selected galaxies from the Comparat et al. (2017) catalogue that satisfy  $\text{Dec} > 42^\circ$  and  $\text{Ra} < 225^\circ$ . This subset aligns more closely with the footprint and  $n(z)$  of public data, as well as the random catalogue. We do implement this filtering in the UCHUU-eBOSS lightcones as well.

#### 6.1.4 Large-scale bias predictions

In Fig. 11, we study the galaxy bias,  $b(s)$ , through our theoretical predictions of the 2PCF monopole. We then proceed to compare these predictions with observed data. For each of the three studied samples, the large-scale bias at their respective median redshift,  $z_{\text{med}}$ , is derived by solving the equation below, which gives a simple linear perturbation theory prediction of the RSD, applicable on large scales (see Kaiser 1987; Hamilton 1998):

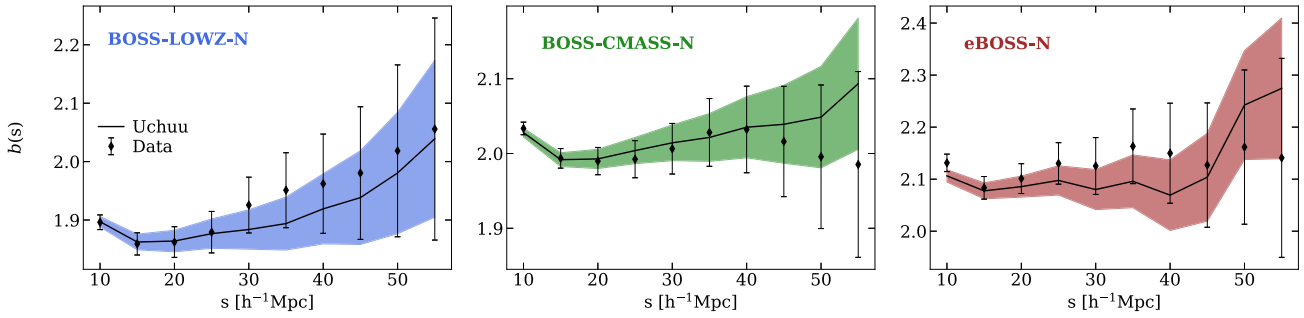
$$\xi_0(s) = b(s)^2 \left( 1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2 \right) \xi_{\text{lin}}(s). \quad (12)$$

In the equation, we have  $\beta = \Omega_{m, z_{\text{med}}}^{0.55}/b(s)$ , where  $\xi_0(s)$  represents the measured 2PCF monopole (derived from data and lightcones in each case), and  $\xi_{\text{lin}}(s)$  corresponds to the dark matter linear 2PCF from the UCHUU simulation at  $z_{\text{med}}$ .

Our UCHUU BOSS/eBOSS lightcones reproduce the  $b(s)$  values from observed data within the uncertainties. The BOSS/eBOSS bias factors,  $b$ , from both observed data and UCHUU lightcones, are presented in Table 5. The bias factors have been obtained as the mean value of  $b(s)$  measured between  $15$  and  $30 h^{-1} \text{ Mpc}$ . This table clearly shows how the bias increases with the median redshift of the sample. This is in agreement with the results reported in Zhou et al. (2021), where a similar behaviour was observed when examining the evolution of large-scale bias with redshift for a DESI-type LRG population selected from the Legacy Survey imaging data set (Dey et al. 2019).

#### 6.1.5 Performance of UCHUU-COMB LRG lightcones

Finally, we assess the ability of our UCHUU-COMB lightcones to reproduce the observed BOSS clustering statistics in both hemispheres, N+S, as well as the clustering evolution with redshift. The monopole, quadrupole and hexadecapole moments of the 2PCF measurements are shown in Fig. 12 for both the observed data and mean UCHUU measurements. We additionally show the GLAM-UCHUU-COMB measurements with a light shaded area to demonstrate their behavior



**Figure 11.** Measurements of the galaxy bias at the median redshift of the CMASS-N (left panel), LOWZ-N (middle panel), and eBOSS-N (right panel) samples are presented. For details on the bias estimation, refer to the text. In the case of the observations, the dark matter clustering is assumed to be the same as that in the UCHUU simulation. Observed data measurements are represented by points. The solid lines show the mean bias derived from the UCHUU lightcones for each of the samples. Data error bars represent the standard deviation of our set of 2000 GLAM-UCHUU lightcones at the  $1\sigma$  level, and the shaded area corresponding to UCHUU indicates the error on the mean,  $\sigma/\sqrt{2}$ .

**Table 5.** Bias factors of the BOSS and eBOSS samples. The first and second columns provide the sample name and its corresponding median redshift. The last two columns display the large-scale bias factors measured from the 2PCF monopole for both the observations and UCHUU lightcones. These bias factors are obtained as the mean value of  $b(s)$  within the range of  $15$  to  $30h^{-1}\text{Mpc}$ .

Sample	$z_{\text{med}}$	$b^{\text{data}}$	$b^{\text{UCHUU}}$
LOWZ-N	0.32	$1.87 \pm 0.03$	$1.87 \pm 0.02$
CMASS-N	0.57	$1.99 \pm 0.02$	$2.00 \pm 0.02$
eBOSS-N	0.70	$2.10 \pm 0.03$	$2.08 \pm 0.02$

and to elucidate the source of potential discrepancies between UCHUU and the data. From left to right, the panels show the redshift intervals over which the 2PCF has been measured:  $0.2 < z < 0.5$ ,  $0.4 < z < 0.6$ , and  $0.5 < z < 0.75$ . Correspondingly, Fig. 13 presents results for the 3PCF. The power-spectrum monopole and quadrupole measurements are shown in Fig. 14 for  $k = 0.005 - 0.6h\text{Mpc}^{-1}$ .

As expected, our UCHUU lightcones are able to fully reproduce with high accuracy the clustering evolution observed in the BOSS N+S data measurements. Residuals from the 2PCF monopole between the CMASS-N fibre collision scale,  $4.2h^{-1}\text{Mpc}$ , and  $25h^{-1}\text{Mpc}$  ( $100h^{-1}\text{Mpc}$ ) exhibit values below 0.9 per cent (8 per cent) for  $0.2 < z < 0.5$  sample, less than 1.5 per cent (15 per cent) for  $0.4 < z < 0.6$  sample, and under 0.9 per cent (10 per cent) for  $0.5 < z < 0.75$  sample. For the 3PCF, all the data points are in agreement within  $1\sigma$  errors with the mean of the UCHUU lightcones. Regarding the power spectrum monopole between  $k > 0.03$  and  $k < 0.6h\text{Mpc}^{-1}$ , the residuals remain within 2.6, 4.5, 3.4 per cent for the three samples (left to right). We interpret the discrepancies in the 2PCF and power spectrum multipoles as arising from the same reasons as discussed in Sections 6.1.1 and 6.1.2. Furthermore, the findings from GLAM-UCHUU validate that certain discrepancies between UCHUU and the data can be ascribed to statistical fluctuations. Our results are in agreement with previous works, both in configuration (see Chuang et al. 2017) and Fourier space (see Beutler et al. 2017).

The results presented in this section allow us to reach the conclusion that our theoretical predictions, based on the standard Planck cosmology model and built by generating lightcones from the UCHUU simulation, exhibit a remarkable ability to reproduce, with a high accuracy, the observed data across all studied samples: LOWZ, CMASS, COMB, and eBOSS, spanning both for Northern and Southern hemispheres. This reaffirms the high quality of the UCHUU simulation and highlights the reliability of our methodology.

Notably, our approach relies on just one free parameter, the scatter  $\sigma_{\text{SHAM}}$ , and covers various aspects, including adopting a complete SMF in the SHAM process and addressing the treatment of galaxy population incompleteness and fibre collision systematics.

## 6.2 Covariance errors from GLAM-UCHUU lightcones

As described in Section 5.1, we populate the GLAM haloes with galaxies using the HOD method. This statistic is obtained from the UCHUU galaxy cubic boxes and is shown in Fig. 15. By construction, the GLAM-UCHUU HOD agrees with that of the high-resolution UCHUU simulation. The forms of the HOD we recover are consistent with those found in previous BOSS and eBOSS LRG HOD studies (e.g. Nuza et al. 2013; Alam et al. 2020).

In this section, we analyse the covariance errors obtained from the GLAM-UCHUU covariance lightcones and present a comparison of these results with MD-PATCHY and EZMOCK. For this analysis, we focus on the CMASS- and eBOSS-N samples, as the Southern hemisphere samples have considerably smaller effective areas compared to the Northern ones. Additionally, the LOWZ and COMB samples have significant discontinuities in their footprints.

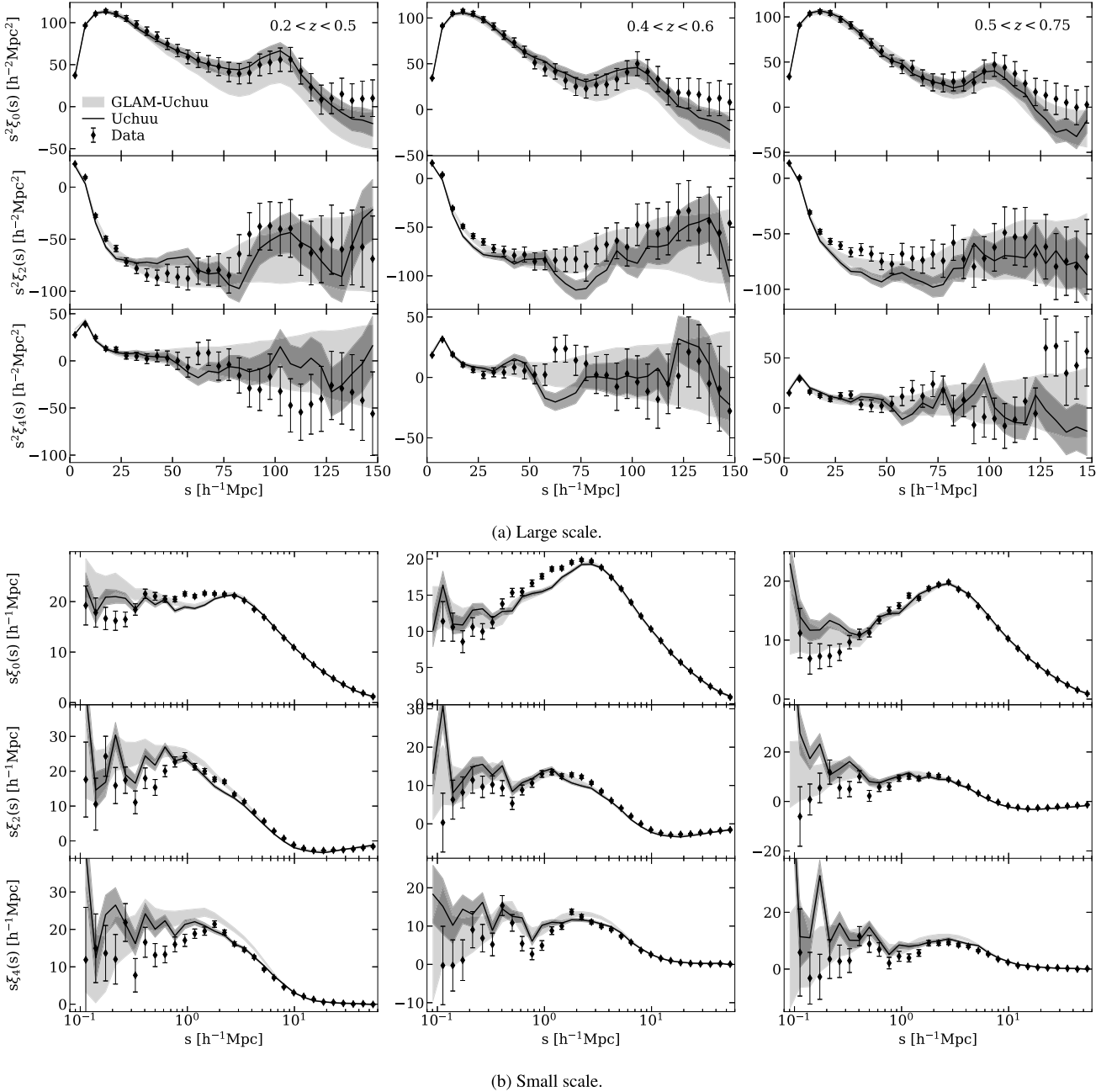
For a fair comparison of GLAM-UCHUU results with the 2048 MD-PATCHY lightcones, we (randomly) select 2000 of those lightcones for the analysis carried out in this section. In contrast, given the availability of only 1000 EZMOCK lightcones, we randomly select an equivalent number from our set of 2000 GLAM-UCHUU lightcones.

### 6.2.1 Covariances in configuration space

We compare the mean 2PCF multipole moments measured from the GLAM-UCHUU runs, along with those from MD-PATCHY and EZMOCK, in Fig. 16. The left column shows the analysis for the CMASS-N sample, while the right column corresponds to the eBOSS-N sample. Overall, GLAM-UCHUU accurately reproduces the observed CMASS-N and eBOSS-N data statistics across all studied multipoles and scales. However, when examining the results of MD-PATCHY and EZMOCK, it becomes evident that their multipole moments exhibit strange features and discrepancies on scales below  $20h^{-1}\text{Mpc}$ . The CMASS-N study also reveals that the resolution of GLAM-UCHUU lightcones is better than that of MD-PATCHY and EZMOCK, reaching down to scales of  $0.1h^{-1}\text{Mpc}$ , in comparison to  $0.3h^{-1}\text{Mpc}$  and  $0.4h^{-1}\text{Mpc}$  for the other lightcones.

Furthermore, the behaviour observed in GLAM-UCHUU confirms that some of the variations previously noted in UCHUU multipole





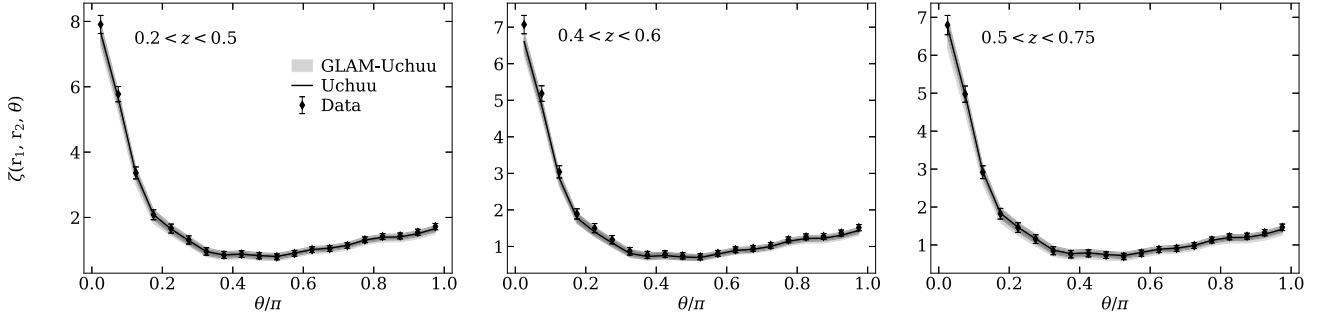
**Figure 12.** Measurements of the monopole, quadrupole and hexadecapole moments of the 2PCF for the COMB-N+S sample in three redshift ranges of interest:  $0.2 < z < 0.5$  (left column),  $0.4 < z < 0.6$  (middle column) and  $0.5 < z < 0.75$  (right column). Observed data measurements are represented as data points. The solid line shows the mean 2PCF derived from UCHUU. We additionally show the standard deviation of GLAM-UCHUU around its mean in a light shaded area. Data error bars represent the standard deviation of our set of 2000 GLAM-UCHUU lightcones,  $1\sigma$ , while the dark shaded area for UCHUU is the error on the mean,  $\sigma/\sqrt{2}$ . Panel (a) shows the 2PCF measurements from 0 to  $150 h^{-1} \text{ Mpc}$  in  $5 h^{-1} \text{ Mpc}$  bins, while panel (b) shows results up to  $60 h^{-1} \text{ Mpc}$  using equally spaced logarithmic bins.

moments can be attributed to statistical fluctuations, as these discrepancies diminish as the number of lightcones increases. For scales below  $60 h^{-1} \text{ Mpc}$ , the 2PCF multipoles of GLAM remain within  $1\sigma$  of the data for both BOSS and eBOSS, whereas UCHUU shows variations between 1 and  $2\sigma$  (see Fig. 5b). The improved consistency of the quadrupole and hexadecapole measurements compared to the monopole, as the number of lightcones increases, could potentially be associated with volumetric effects and harmonic modes within the data. The  $\ell = 2$  and  $\ell = 4$  moments may exhibit greater sensitivity to fluctuations arising from the geometric characteristics

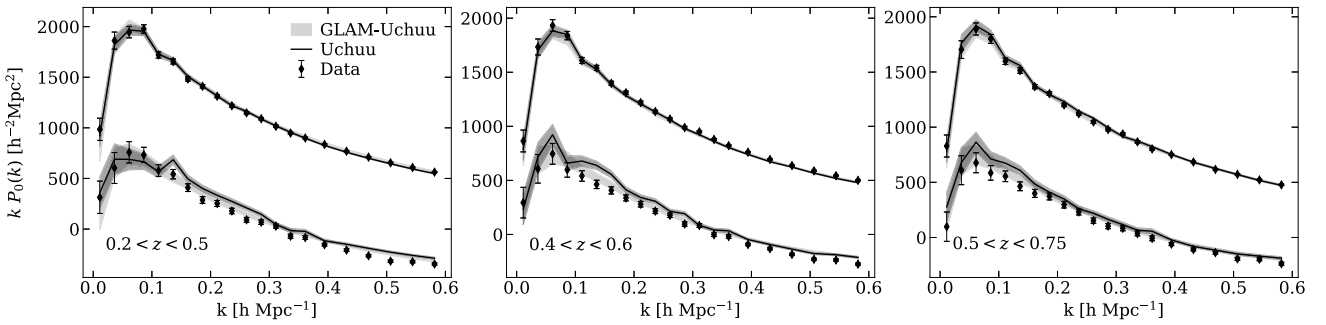
and configuration of the observed volume. Increasing the number of lightcones and averaging the results could help mitigate some of these influences.

The covariance matrix of the 2PCF multipoles,  $\xi_\ell(s)$ , can be defined as follows:

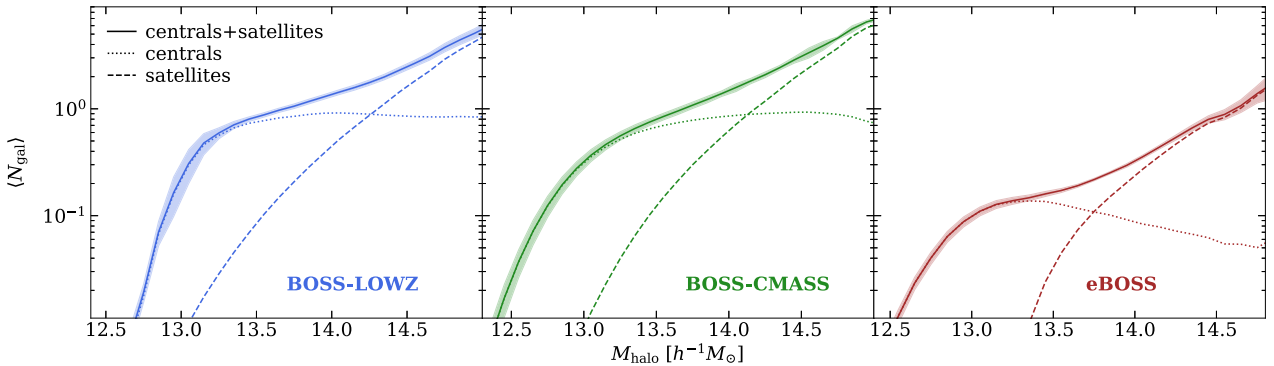
$$\text{Cov}_{\ell\ell'}[s_i, s_j] = \frac{C}{N_m - 1} \sum_p^{N_m} [\xi_\ell^p(s_i) - \bar{\xi}_\ell(s_i)] \times [\xi_{\ell'}^p(s_j) - \bar{\xi}_{\ell'}(s_j)], \quad (13)$$



**Figure 13.** Measurements of the 3PCF for the COMB-N + S sample in three redshift ranges of interest:  $0.2 < z < 0.5$  (left panel),  $0.4 < z < 0.6$  (middle panel), and  $0.5 < z < 0.75$  (right panel). Observed data measurements are represented by data points. The solid line shows the mean power spectrum derived from UCHUU. We additionally show the standard deviation of GLAM-UCHUU around its mean in a light-shaded area. Data error bars represent the standard deviation of our set of 2000 GLAM-UCHUU lightcones,  $1\sigma$ , while the dark-shaded area for UCHUU is the error on the mean,  $\sigma/\sqrt{2}$ . The 3PCF is presented for  $r_1 = 20h^{-1}\text{Mpc}$  and  $r_2 = 25h^{-1}\text{Mpc}$ , and  $\theta$  is the angle between  $r_1$  and  $r_2$ , in radians.



**Figure 14.** Measurements of the power spectrum monopole (upper curve) and quadrupole (lower curve) for the COMB-N+S sample in three redshift ranges of interest:  $0.2 < z < 0.5$  (left panel),  $0.4 < z < 0.6$  (middle panel), and  $0.5 < z < 0.75$  (right panel). Observed data measurements are represented by data points. The solid line show the mean power spectrum derived from UCHUU. We additionally show the standard deviation of GLAM-UCHUU around its mean in a light-shaded area. Data error bars represent the standard deviation of our set of 2000 GLAM-UCHUU lightcones,  $1\sigma$ , while the dark shaded area for UCHUU is the error on the mean,  $\sigma/\sqrt{2}$ . The power spectrum is measured from  $k = 0.005$  to  $0.6 h \text{ Mpc}^{-1}$  in bins of  $0.03 h \text{ Mpc}^{-1}$ .



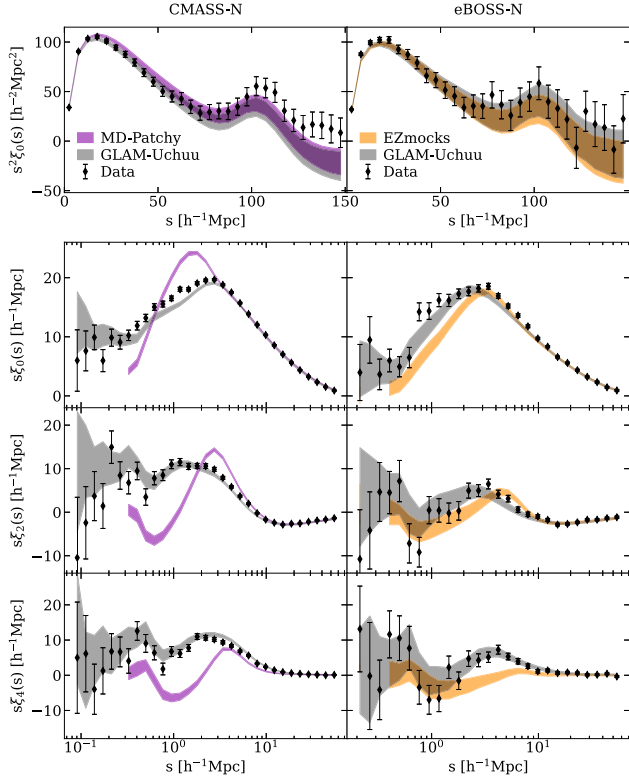
**Figure 15.** The HOD measured in the LOWZ (left panel), CMASS (middle panel), and eBOSS (right panel) samples, as determined from the mean of the UCHUU galaxy cubic boxes. The solid lines represent the combined HOD of centrals and satellites, while the dotted and dashed lines show the mean halo occupancy for centrals and satellites, respectively. The shaded area shows the standard deviation computed using 2000 GLAM-UCHUU lightcones.

where  $N_m$  is the number of lightcones,  $\xi_\ell^p$  indicates the  $\ell$ -order 2PCF multipole of the  $p$ th lightcone, and  $\bar{\xi}_\ell$  denotes the mean 2PCF multipole across all lightcones. The term  $\mathcal{C}$  accounts for the dependence of  $\text{Cov}[s_i, s_j]$  on the size of the simulation box. When the effective volume of the observed sample,  $V_{\text{eff}}$ , is smaller than the simulation volume,  $V_{\text{sim}}$ , then  $\mathcal{C} = 1$ . Conversely, if  $V_{\text{eff}} > V_{\text{sim}}$ , the covariance matrix must be scaled, such that  $\mathcal{C} \neq 1$ . Specifically, for the GLAM-UCHUU-CMASS-N covariance matrix,  $\mathcal{C} \simeq 0.6$  (see Appendix B for details). To further investigate the variations between

lightcones, one can also examine the correlation matrices of the 2PCF multipoles:

$$\text{Cor}_{\ell\ell'}[s_i, s_j] = \frac{\text{Cov}_{\ell\ell'}[s_i, s_j]}{\sqrt{\text{Cov}_{\ell\ell'}[s_i, s_i]\text{Cov}_{\ell\ell'}[s_j, s_j]}}. \quad (14)$$

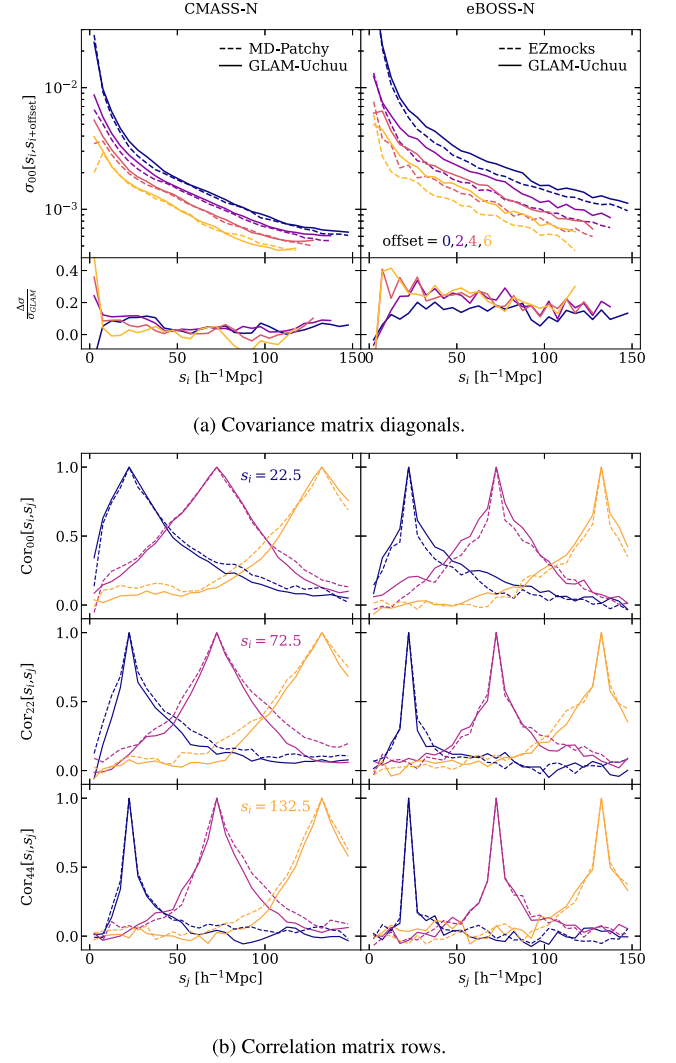
Fig. 17(a) provides a comparison of various components of the 2PCF monopole covariance matrices examined in this study. Specifically, it shows the diagonal and second, fourth, and sixth off-diagonal terms. In the left column, we compare results from GLAM-



**Figure 16.** Measurements of the 2PCF monopole, quadrupole and hexadecapole moments for CMASS-N (left column) and eBOSS-N (right column). The shaded regions represent the standard deviation around its mean for GLAM-UCHUU (in black), MD-PATCHY (purple, left column), and EZMOCK (orange, right column). The points with error bars show the measurements from the observed data, where the errors correspond to the  $1\sigma$  scatter derived from GLAM-UCHUU lightcones. The upper panels show the 2PCF measurements from 0 to  $150 h^{-1} \text{ Mpc}$  in  $5 h^{-1} \text{ Mpc}$  bins. The lower panels show the smaller scales up to  $60 h^{-1} \text{ Mpc}$  in equally spaced logarithmic bins. It is evident that the MD-PATCHY and EZMOCK results exhibit significant discrepancies with the data between  $0.1$  and  $30 h^{-1} \text{ Mpc}$ , further emphasizing the need for generating the GLAM-UCHUU lightcones.

UCHUU-N with those from MD-PATCHY-CMASS-N. The values of the GLAM-UCHUU covariance matrix are consistently higher than those of MD-PATCHY across all slices, with a discrepancy between the two that fluctuates within 10–15 per cent, as shown by the residuals in the bottom panel. We compare GLAM-UCHUU and EZMOCK for the eBOSS-N sample in the right column. The trend is the same: the values of the GLAM-UCHUU elements are higher than those of EZMOCK. The disagreement in this case is larger for some values of  $s$ , reaching up to 30 per cent. Discrepancies within the 10 per cent range are anticipated in both cases, and only discrepancies above this value are considered relevant (see Appendix B). Differences above this percentage are due to the different predictions of each simulation.

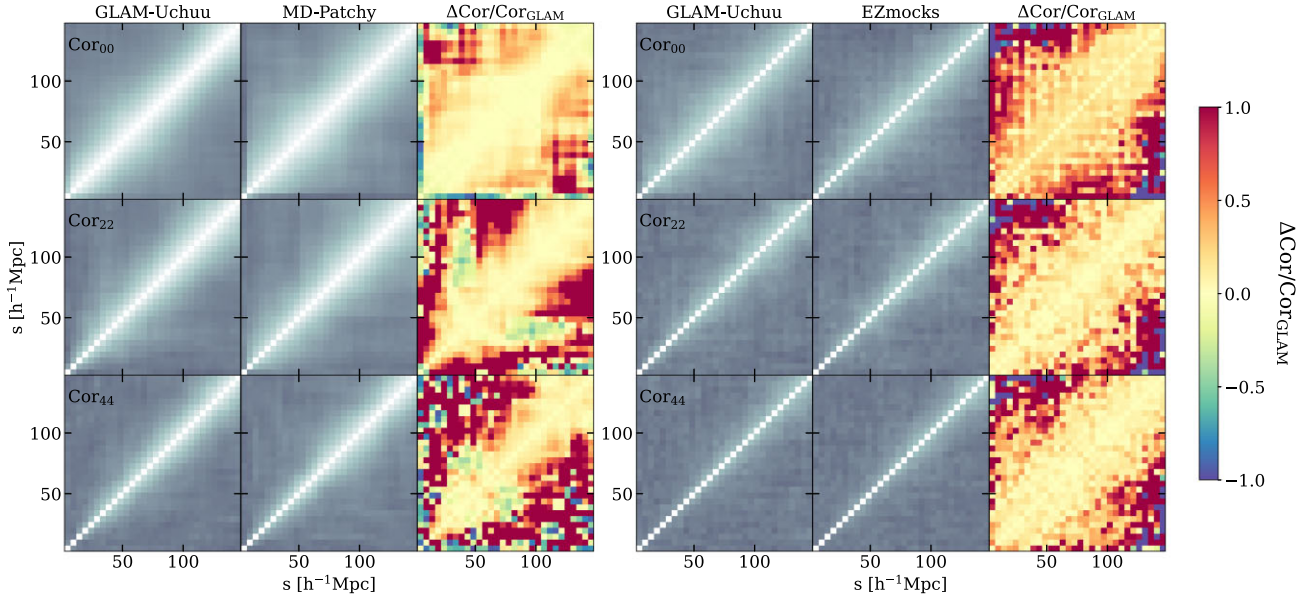
In Fig. 18, we report the derived correlation matrices for the monopole, quadrupole, and hexadecapole moments of the 2PCF. In the left plot, we show the correlation matrices obtained for the CMASS-N sample using GLAM-UCHUU, MD-PATCHY, and the residuals between the two. The right plot follows the same, but for the eBOSS-N sample, replacing MD-PATCHY with EZMOCK. In both cases, the residuals show distinct patterns in  $\text{Cor}_{00}$ ,  $\text{Cor}_{22}$  and  $\text{Cor}_{44}$ . To gain a deeper insight into the degree of correlation and the structure of these matrices, Fig. 17 (b) displays cross-sections through the covariance matrices that highlight the behaviour



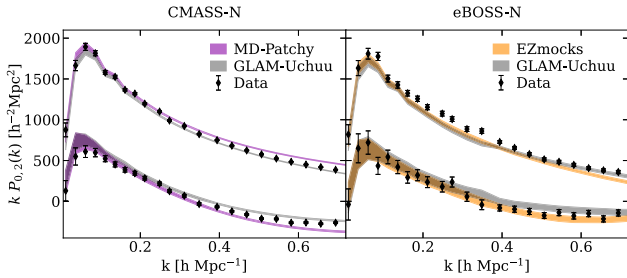
**Figure 17.** Panel (a): Comparison of the covariance matrices of the 2PCF monopole,  $\sigma_{00} = \sqrt{\text{Cov}_{00}}$ , for CMASS-N and eBOSS-N (left and right columns, respectively). For both samples, GLAM-UCHUU is represented by solid lines, while MD-PATCHY (left column) and EZMOCK (right column) are shown with dashed lines. Different colours represent distinct terms of the covariance matrices: diagonal term in blue and second, fourth and sixth off-diagonal terms in purple, pink, and orange, respectively. The bottom panel shows the residuals between covariances. Panel (b): Slices through the correlation matrices of the monopole, quadrupole, and hexadecapole ( $\ell = 0, 2, 4$ ) moments of the 2PCF for CMASS-N and eBOSS-N (left and right columns, respectively). GLAM-UCHUU is depicted with solid lines in both columns. MD-PATCHY (left column) and EZMOCK (right column) are indicated by dashed lines. Different colours represent distinct  $s_i$  values in units of  $h^{-1} \text{ Mpc}$ :  $s_i = 22.5, 72.5$  and  $132.5$  in purple, pink, and yellow, respectively.

of their non-diagonal terms. We make a comparison for three separation values:  $s_i = 22.5, 72.5, 132.5 h^{-1} \text{ Mpc}$ , as indicated by the colour scheme in the plot. In general, we observe a stronger correlation among bins located near the diagonal for the CMASS-N sample compared to the eBOSS-N, resulting in broader peaks in the former case. This correlation tends to weaker as we consider higher order multipoles. When examining individual cases, we notice that the off-diagonal elements of MD-PATCHY display a higher level of correlation across all three multipoles when compared with GLAM-UCHUU. In comparison with EZMOCK, while GLAM-





**Figure 18.** *Left panel:* Correlation matrices (equation 14) of the 2PCF monopole, quadrupole, and hexadecapole moments (first, second, and third rows, respectively) obtained from GLAM-UCHUU (first column) and MD-PATCHY (second column). The residuals,  $r$ , between GLAM-UCHUU and MD-PATCHY, are shown in the third column. Here, a positive value of  $r$  indicates that the values of the GLAM-UCHUU correlation matrix are lower than those of MD-PATCHY, while  $r < 0$  means the opposite. *Right panel:* We examine the eBOSS-N sample covariances. The layout is identical to the left plot, but EZMOCK is used instead of MD-PATCHY.



**Figure 19.** Measurements of the monopole (upper curve) and quadrupole (lower curve) moments of the power spectrum for the CMASS-N (left panel) and eBOSS-N (right panel) samples. The shaded regions represent the standard deviation around the mean for GLAM-UCHUU (in black), MD-PATCHY (purple), and EZMOCK (orange) lightcone. The points with error bars show the measurements from the observed data, where the errors correspond to the  $1\sigma$  scatter derived from GLAM-UCHUU lightcones. We measure the power spectrum over the range  $k = 0.005$  to  $0.72 h \text{ Mpc}^{-1}$ , using  $0.03 h \text{ Mpc}^{-1}$  bins. The results from GLAM-UCHUU show good agreement with the observed clustering statistics for both CMASS-N and eBOSS-N. For  $k > 0.4 h \text{ Mpc}^{-1}$ , the multipoles from MD-PATCHY and EZMOCK begin to deviate from the observed data.

UCHUU presents more highly correlated off-diagonal elements for the monopole, it demonstrates similar results to EZMOCK for the quadrupole and hexadecapole.

### 6.2.2 Covariances in Fourier space

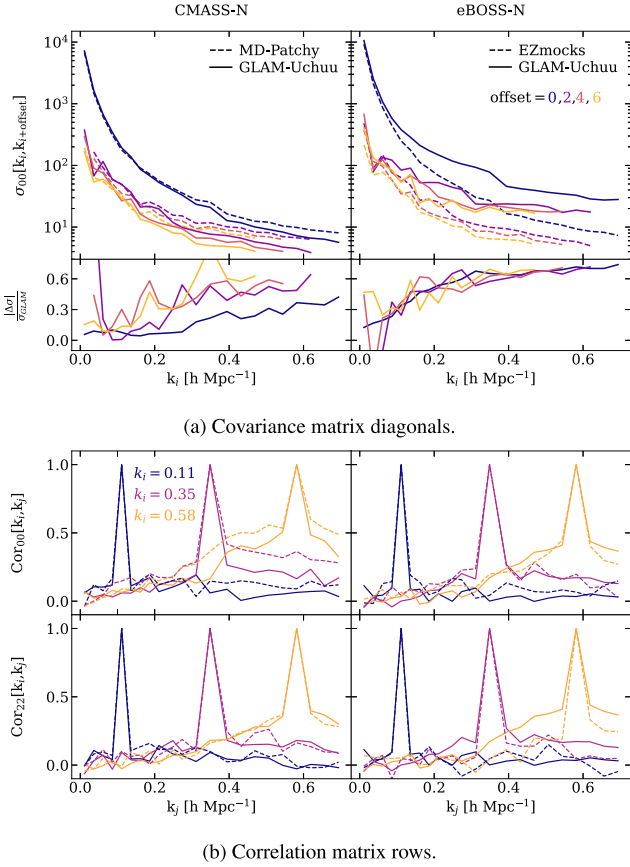
We analyse our GLAM lightcones in Fourier space using the same methods and equations ( $\xi_\ell(s)^p \rightarrow P_\ell(k)^p$ ) as presented in the configuration space section.

Fig. 19 shows the mean monopole and quadrupole moments of the power spectrum measured from GLAM-UCHUU, along with those

from MD-PATCHY and EZMOCK. Error bars represent the uncertainty in the observed data, with errors estimated by the  $1\sigma$  scatter derived from GLAM-UCHUU lightcones. The GLAM-UCHUU results are in good agreement with the observed CMASS-N and eBOSS-N clustering statistics for all the studied multipoles and scales. For  $k$  values above  $0.4 h \text{ Mpc}^{-1}$ , the MD-PATCHY and EZMOCK multipoles deviate from the data. This trend is consistent with that demonstrated in Fig. 16.

In Fig. 20(a), we present the diagonal and second, fourth, and sixth off-diagonal terms of the power spectrum monopole covariance matrices. In the left column, we compare GLAM-UCHUU and MD-PATCHY for the CMASS-N sample, while the right column compares GLAM-UCHUU and EZMOCK for the eBOSS-N sample. The diagonal values of GLAM-UCHUU are slightly higher than those of MD-PATCHY for low  $k$  values. However, this trend inverts for  $k \gtrsim 0.25 h \text{ Mpc}^{-1}$ . When comparing with EZMOCK, GLAM-UCHUU slices are higher at all scales, with a discrepancy between the two that increases with  $k$ , reaching values above 60 per cent for  $k_i \geq 0.3 h \text{ Mpc}^{-1}$ . As in configuration space analysis, discrepancies within 10 per cent are expected, with discrepancies above this value attributed to the distinct predictions of each simulation.

We present the derived correlation matrices for the monopole and quadrupole moments of the power spectrum in Fig. 21. Following the same format as shown in Fig. 18, the left plot displays the correlation matrices obtained for the CMASS-N sample using GLAM-UCHUU and MD-PATCHY. In the right plot, the eBOSS-N sample is analysed with EZMOCK instead of MD-PATCHY. The residuals,  $r$ , are plotted in the third columns, showing again some patterns evident in both monopole and quadrupole moments. We show three cross-sections through these correlation matrices at  $k_i = 0.11, 0.35, 0.58 h \text{ Mpc}^{-1}$  in Fig. 20 (b). In general, the correlation of the analysed cuts are very similar among GLAM-UCHUU, MD-PATCHY and EZMOCK, particularly for the terms near the diagonal. For the bins located farther from the diagonal, noise becomes prominent and the signal amplitude is considerably low, posing



**Figure 20.** Panel (a): Comparison of the covariance matrices of the power spectrum monopole,  $\sigma_{00} = \sqrt{\text{Cov}_{00}}$ , for CMASS-N and eBOSS-N (left and right columns, respectively). For both samples, GLAM-UCHUU is represented by solid lines, while MD-PATCHY (left column) and EZMOCK (right column) are shown with dashed lines. Various colours represent different terms of the matrices: diagonal terms in blue, and second, fourth and sixth off-diagonal terms in purple, pink, and orange, respectively. The bottom panel displays the absolute residuals between the covariances. Panel (b): This panel shows slices through the correlation matrices of the power spectrum monopole and quadrupole for CMASS-N and eBOSS-N (again, left and right columns, respectively). GLAM-UCHUU is depicted with solid lines in both columns. MD-PATCHY (left column) and EZMOCK (right column) are indicated by dashed lines. Different colours correspond to distinct  $k_i$  values in units of  $h\text{Mpc}^{-1}$ :  $k_i = 0.11, 0.35$  and  $0.58$  in purple, pink, and yellow, respectively.

challenges in determining clear trend. In the CMASS monopole case however, we clearly observe that MD-PATCHY display a higher level of correlation than GLAM-UCHUU.

The analysis presented in this section confirms the reliability of the methodology used in generating the GLAM-UCHUU covariance lightcones, as the accurately reproduce the clustering measurements from observations. Comparisons of the GLAM-UCHUU lightcones with results from MD-PATCHY and EZMOCK prove insightful. These approximate approaches exhibit peculiar behaviours in the multipole moments of the 2PCF, clearly demonstrating the advantage of using  $N$ -body simulations over less accurate methods. We have verified that the diagonal elements of the GLAM-UCHUU covariance matrices generally exceed those obtained from MD-PATCHY and EZMOCK. Consequently, the errors we estimate for the observed 2PCFs and power spectrum are larger than previously assumed. However, it is important to note that discrepancies within 10 per cent in the error estimation are anticipated, and only discrepancies above this

threshold are considered relevant as they are, presumably, due to the different predictions of each simulation. The discovery of discrepancies greater than 10 per cent leads to increased uncertainties in cosmological constraints derived from BOSS and eBOSS data compared to those using MD-PATCHY and EZMOCK covariances. This includes impacts on BAO distance estimates and the estimation of cosmological parameters (see Alam et al. 2015, 2021, and references therein), as well as implications for uncovering primordial non-Gaussianity (see Beutler et al. 2017; Merz et al. 2021; Mueller et al. 2022). Certain studies employing MD-PATCHY and EZMOCK covariance matrices have reported tensions in the inferred value of  $\sigma_8$  when compared with Planck measurements (see Wadkar, Ivanov & Scoccimarro 2020; Chen, Vlah & White 2022; Kobayashi et al. 2022). This tension could potentially be mitigated or even resolved with larger uncertainties factored into the estimated parameters.

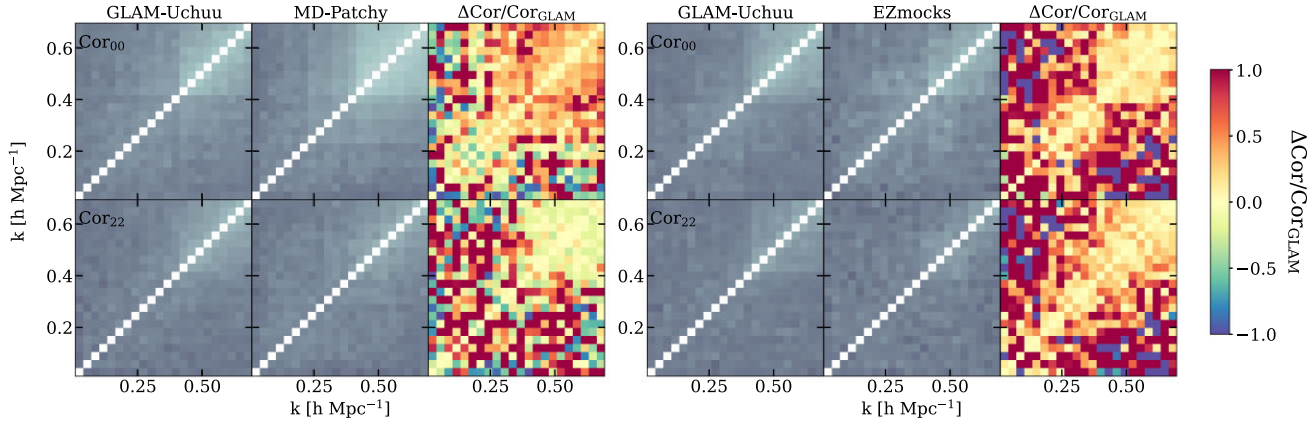
### 6.3 Performance with other cosmologies

Up to this point, we have demonstrated that our UCHUU clustering model, based on standard Planck cosmology, accurately reproduces the observed clustering data while accounting for relevant uncertainties. However, it is also worthwhile to assess the performance of other cosmological models. For this purpose, we use our GLAM simulation runs for different cosmologies as presented in Section 3.2. These models, summarized in Table 2, include GLAM-PMILL, GLAM-PMILLNOBAO, and GLAM-ABACUS. It is important to note that their cosmological parameters do not considerably deviate from the Planck15 cosmology adopted for UCHUU. These newly generated lightcones are produced following the same methodology as the GLAM-UCHUU lightcones, described in Section 5. Regardless of the specific cosmology, galaxies are populated in all GLAM halo cubic boxes using the HODs obtained from the UCHUU galaxy boxes, which are based on the PL15 cosmology. Given the minor difference among the various explored cosmological parameters, creating distinct HODs for each cosmology would yield marginal improvements at the cost of considerable additional complexity.

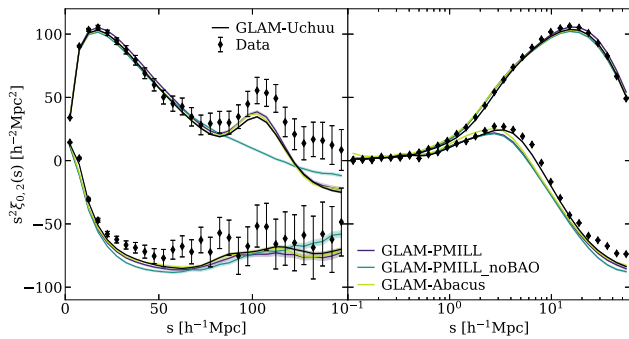
The impact of different cosmologies on the large-scale distribution of galaxies is depicted in Fig. 22, where the mean monopole and quadrupole moments of the 2PCF obtained from the GLAM-PMILL, GLAM-PMILLNOBAO, and GLAM-ABACUS lightcones for the CMASS-N sample are shown. For reference, results from observed data and theoretical predictions calculated from the mean of the GLAM-UCHUU lightcones are also presented. Generally, except for GLAM-PMILLNOBAO, all GLAM lightcones effectively replicate the observational measurements. The absence of the BAO feature in the GLAM-PMILLNOBAO monopole, alongside its BAO-inclusive counterpart GLAM-PMILL, allows us to assess the statistical significance of BAO detection. Despite the measured residuals and associated uncertainties, this study does not definitively distinguish between the presence or absence of massive neutrinos. Although GLAM-UCHUU achieves better results (residuals below 3 per cent up to  $50 h^{-1} \text{Mpc}$ ) compared to GLAM-ABACUS, which includes the effect of massive neutrinos (resulting in residuals of 4 per cent), both models are capable of reasonably reproducing the observed data within the error margins.

## 7 CONCLUSIONS

We have generated and analysed simulated lightcones for luminous red galaxies in BOSS-LOWZ, -CMASS, -COMB and eBOSS samples, all based on the Planck cosmology. Our study encompasses two types of lightcones:



**Figure 21.** *Left panel:* Correlation matrices (equation (14)) for the power spectrum monopole and quadrupole (first and second rows, respectively) obtained from GLAM-UCHUU (first column) and MD-PATCHY (second column). The third column displays the residuals,  $r$ , between GLAM-UCHUU and MD-PATCHY. Here,  $r > 0$  indicates that the values of the GLAM-UCHUU correlation matrix are lower than those of MD-PATCHY, and  $r < 0$  indicates the opposite. *Right panel:* A similar study is conducted for the eBOSS-N sample, using the same layout as the left plot. However, EZMOCK replaces MD-PATCHY in the comparison.



**Figure 22.** Measurements of the monopole and quadrupole moments of the 2PCF for the CMASS-N are presented using different sets of GLAM simulations (see Table 2). The mean 2PCF derived from the GLAM-PMILL lightcones is represented by solid purple lines, GLAM-PMILLnoBAO in blue, GLAM-ABACUS in green, and GLAM-UCHUU in black. The shaded regions indicate the errors of the mean, estimated from the diagonal components of their respective covariance matrices. Observational data are shown as points with error bars, where these bars represent the  $1\sigma$  errors estimated using the GLAM-UCHUU lightcones. The 2PCF measurements span from 0 to  $150 h^{-1}$  Mpc in  $5 h^{-1}$  Mpc bins (left plot), and up to  $60 h^{-1}$  Mpc in equally spaced logarithmic bins (right plot).

(a) High-fidelity LRG lightcones derived from the high-resolution and large-volume UCHUUN-body simulation, which tracks the evolution of haloes and subhaloes. Galaxies are assigned to (sub)haloes using the SHAM technique, which uses one free parameter,  $\sigma_{\text{SHAM}}$  (see equation 4), and the observed stellar mass function. We obtain the corresponding halo occupancy, determining the number of galaxies in each dark matter halo, for these lightcones. This HOD information is then used to populate distinct haloes in lower resolution GLAM-UCHUU simulations.

(b) Covariance LRG lightcones based on two thousand GLAM-UCHUU simulations. Galaxies are assigned to distinct haloes using the HOD method, with HOD parameters obtained from UCHUU lightcones.

It is crucial to highlight that we do not modify our methods to make our results fit the observed clustering of BOSS and eBOSS galaxies. This is in contrast to approximate methods such as MD-PATCHY

and EZMOCK, which employ fitting techniques with numerous free parameters. Our UCHUU and GLAM-UCHUU LRG lightcones are generated without fitting any observed clustering data and without any free parameter: they rely solely on theoretical predictions. The success of this approach depends on high-fidelity cosmological simulations, such as UCHUU. We also demonstrate that generating thousands of galaxy lightcones is feasible using  $N$ -body simulations with adequate mass and force resolution, such as those that we have generated using GLAM and used here. This capability enables us to accurately estimate covariance errors, impact of systematics and identify potential tensions between model and data in future research endeavours.

Throughout this paper, we have studied various aspects of LRG clustering in the BOSS and eBOSS surveys and assess how well our theoretical predictions for Planck cosmology match the data. The main results, based on both our high-fidelity UCHUU and covariance GLAM-UCHUU LRG lightcones, are summarized as follows:

(i) Overall, our UCHUU lightcones accurately reproduces the correlation function and power-spectrum measurements of the observed data for all the studied samples. We find certain discrepancies on large scales, primarily caused by cosmic variance and the substantial impact of observational systematics. On smaller scales, below  $\sim 1 h^{-1}$  Mpc, these deviations can be attributed to the non-application of PIP weights in our analysis. In configuration space, our findings indicate that the residuals of the BOSS 2PCF monopole remain within 10 per cent (15 per cent for eBOSS) up to  $100 h^{-1}$  Mpc and fall below 2.5 per cent up to  $25 h^{-1}$  Mpc. We also find a good agreement for the 3PCF, with our theoretical predictions determined from the mean of the UCHUU lightcones and the observational measurements within  $1\sigma$  errors. Moving to Fourier space, we observe that the power-spectrum monopole residuals between  $k = 0.03$  and  $k = 0.6 h \text{ Mpc}^{-1}$  are below 12.5 per cent for eBOSS, and drop below 5 per cent and 4 per cent for BOSS-LOWZ and -CMASS, respectively.

(ii) From the dependence of clustering on stellar mass, we conclude that the observed LRG population in both BOSS samples is complete at the massive end. We also validate the accuracy of the method adopted in this work for modelling the overall incompleteness of the eBOSS galaxy population.

(iii) The scale-dependent galaxy bias of our UCHUU lightcones is in a good agreement with observations. Additionally, we observe an increase in the bias factor with redshift.



(iv) We check whether our theoretical predictions reproduce the observed evolution of clustering with redshift by analysing the 2PCF, the 3PCF, and the power spectrum of the COMB sample in three redshift bins. In general, we find good agreement between the UCHUU (and GLAM-UCHUU) lightcones and the observed data. We report 2PCF monopole residuals that remain below 1.5 per cent within  $25 h^{-1}$  Mpc, and under 12 per cent within  $100 h^{-1}$  Mpc, and within 4.5 per cent for the monopole power spectrum between  $k = 0.03$  and  $k = 0.6$ . For the 3PCF, we again find good agreement (discrepancies within  $1\sigma$  errors). Any potential discrepancies can be attributed to the same factors mentioned earlier.

(v) Strikingly, the MD-PATCHY and EZMOCK LRG lightcones created for BOSS and eBOSS, respectively, display peculiar features in their clustering multipoles that deviate from observations within  $20 h^{-1}$  Mpc and for  $k > 0.4 h \text{ Mpc}^{-1}$ . In contrast, our  $N$ -body simulation-based lightcones agree with the data even on scales as small  $0.1 h^{-1}$  Mpc.

(vi) When comparing the BOSS-CMASS covariance matrices, we find that GLAM-UCHUU and MD-PATCHY estimates agree, in general, within 15 per cent. On the other hand, EZMOCK estimates for eBOSS deviate up to 60 per cent from GLAM-UCHUU. However, we note that discrepancies of up to 10 per cent are to be expected, with those exceeding this value being attributed to the distinct predictions of each simulation (approximate methods versus  $N$ -body simulations). Regarding the correlation matrices, we notice that GLAM-UCHUU presents lower correlation compared to MD-PATCHY, but similar to EZMOCK.

(vii) Finally, we explore the impact of cosmology on galaxy clustering using various GLAM simulations. Our analysis reveals that it is not feasible to discern between an universe with or without massive neutrinos within the current BOSS/eBOSS uncertainties.

In conclusion, our theoretical predictions, based on the Planck cosmology and derived from the high-fidelity UCHUU lightcones, confirm the accuracy of Planck cosmological parameters in explaining observations from the LSS BOSS/eBOSS surveys, and demonstrate the robustness of the UCHUU simulation. These findings have significant implications for refining lightcone construction methodologies and advancing our understanding of clustering measurements. Moreover, we demonstrate that both MD-PATCHY and EZMOCK LRG lightcones systematically underestimate covariance errors by  $\approx 10$ –60 per cent compared to GLAM-UCHUU. This may have important implications for cosmological parameter inferences derived from BOSS and eBOSS, underscoring the critical importance of utilising high-fidelity simulations to enhance the reliability and accuracy of cosmological analyses.

A similar study, as presented here, can be expanded by using the Modify Gravity version of GLAM (MG-GLAM; see Hernández-Aguayo et al. 2022; Ruan et al. 2022) for analysing BOSS/eBOSS clustering, potentially revealing novel insights into the nature of gravity. In the near future, we plan to generate UCHUU LRG lightcones for the first year of DESI data. We will then construct their covariance matrices within the GLAM-UCHUU framework, as outlined in this paper. This effort will enhance the precision of the cosmological analyses conducted using the DESI data.

## ACKNOWLEDGEMENTS

JE, FP, and AK acknowledge financial support from the Spanish MICINN funding grant PGC2018-101931-B-I00. TI has been supported by IAAR Research Support Program in Chiba University Japan, MEXT/JSPS KAKENHI (Grants JP19KK0344 and

JP21H01122), MEXT as ‘Program for Promoting Researches on the Supercomputer Fugaku’ (JPMXP1020200109 and JPMXP1020230406), and JICFuS. AS, BL, and CMB acknowledge support from the Science Technology Facilities Council through ST/X001075/1. CHA acknowledges support from the Excellence Cluster ORIGINS which is funded by the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) under Germany’s Excellence Strategy – EXC-2094–390783311. JE thanks Oliver Philcox for his assistance in the calculation of the 3PCF.

The Uchuu simulation was carried out on the Aterui II supercomputer at CfCA-NAOJ. We thank IAA-CSIC, CESGA, and RedIRIS in Spain for hosting the Uchuu data releases in the SKIES & UNIVERSES site for cosmological simulations. The analysis done in this paper have made use of NERSC at LBNL and skun6@IAA-CSIC computer facility managed by IAA-CSIC in Spain (MICINN EU-Feder grant EQC2018-004366-P).

This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility ([www.dirac.ac.uk](http://www.dirac.ac.uk)). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1, ST/R002371/1, and ST/S002502/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure.

## DATA AVAILABILITY

The data sets included in this work consist of the following: 2 (lightcones)  $\times$  8 (samples) UCHUU lightcones, the  $2000 \times 8$  GLAM-UCHUU lightcones, the  $100 \times 8$  GLAM-PMILL, GLAM-PMILLNOBAO, and GLAM-ABACUS lightcones, along with the BOSS and eBOSS catalogues. These data sets are publicly available at <https://www.skiesanduniverses.org/Products/>. A comprehensive list and brief description of the catalogue columns can be found in Appendix A.

## REFERENCES

- Ahumada R. et al., 2020, *ApJS*, 249, 3
- Alam S. et al., 2015, *ApJS*, 219, 12
- Alam S. et al., 2017, *MNRAS*, 470, 2617
- Alam S., Peacock J. A., Kraljic K., Ross A. J., Comparat J., 2020, *MNRAS*, 497, 581
- Alam S. et al., 2021, *Phys. Rev. D*, 103, 083533
- Aung H. et al., 2023, *MNRAS*, 519, 1648
- Barrera M. et al., 2023, *MNRAS*, 525, 6312
- Baugh C. M. et al., 2019, *MNRAS*, 483, 4922
- Behroozi P. S., Conroy C., Wechsler R. H., 2010, *ApJ*, 717, 379
- Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013a, *ApJ*, 762, 109
- Behroozi P. S., Wechsler R. H., Wu H.-Y., Busha M. T., Klypin A. A., Primack J. R., 2013b, *ApJ*, 763, 18
- Berlind A. A., Weinberg D. H., 2002, *ApJ*, 575, 587
- Beutler F. et al., 2017, *MNRAS*, 464, 3409
- Bianchi D., Percival W. J., 2017, *MNRAS*, 472, 1106
- Chaves-Montero J., Angulo R. E., Schaye J., Schaller M., Crain R. A., Furlong M., Theuns T., 2016, *MNRAS*, 460, 3100
- Chen S.-F., Vlah Z., White M., 2022, *J. Cosmol. Astropart. Phys.*, 2022, 008
- Chuang C.-H. et al., 2017, *MNRAS*, 471, 2370
- Cole S. et al., 2005, *MNRAS*, 362, 505
- Comparat J. et al., 2017, preprint ([arXiv:1711.06575](https://arxiv.org/abs/1711.06575))
- Contreras S., Angulo R. E., Zennaro M., 2021, *MNRAS*, 508, 175
- Contreras S., Angulo R. E., Chaves-Montero J., White S. D. M., Aricò G., 2023a, *MNRAS*, 520, 489
- Contreras S. et al., 2023b, *MNRAS*, 524, 2489
- DESI Collaboration, 2022, *AJ*, 164, 207

- Davis M., Peebles P. J. E., 1983, *ApJ*, 267, 465
- Dawson K. S. et al., 2013, *AJ*, 145, 10
- Dawson K. S. et al., 2016, *AJ*, 151, 44
- Dey A. et al., 2019, *AJ*, 157, 168
- Dong-Páez C. A. et al., 2024, *MNRAS*, 528, 7236
- Eisenstein D. J. et al., 2005, *ApJ*, 633, 560
- Feldman H. A., Kaiser N., Peacock J. A., 1994, *ApJ*, 426, 23
- Frenk C. S. et al., 1999, *ApJ*, 525, 554
- Guo H., Zehavi I., Zheng Z., 2012, *ApJ*, 756, 127
- Guo H. et al., 2016, *MNRAS*, 459, 3040
- Hamilton A. J. S., 1998, in Hamilton D., ed., *Astrophysics and Space Science Library (ASSL)*, Vol. 231, *The Evolving Universe. Selected Topics on Large-Scale Structure and on the Properties of Galaxies*. Kluwer Academic Publishers, Dordrecht, p. 185
- Hand N., Li Y., Slepian Z., Seljak U., 2017, *J. Cosmol. Astropart. Phys.*, 2017, 002
- Hayashi E., Navarro J. F., Taylor J. E., Stadel J., Quinn T., 2003, *ApJ*, 584, 541
- Hernández-Aguayo C., Prada F., Baugh C. M., Klypin A., 2021, *MNRAS*, 503, 2318
- Hernández-Aguayo C., Ruan C.-Z., Li B., Arnold C., Baugh C. M., Klypin A., Prada F., 2022, *J. Cosmology Astropart. Phys.*, 2022, 048
- Hernández-Aguayo C. et al., 2023, *MNRAS*, 524, 2556
- Huterer D., Cunha C. E., Fang W., 2013, *MNRAS*, 432, 2945
- Ishiyama T., Fukushima T., Makino J., 2009, *PASJ*, 61, 1319
- Ishiyama T., Nitadori K., Makino J., 2012, preprint (arXiv:1211.4406)
- Ishiyama T. et al., 2021, *MNRAS*, 506, 4210
- Kaiser N., 1987, *MNRAS*, 227, 1
- Kitaura F. S., Yepes G., Prada F., 2014, *MNRAS*, 439, L21
- Kitaura F.-S. et al., 2016, *MNRAS*, 456, 4156
- Kitzbichler M. G., White S. D. M., 2007, *MNRAS*, 376, 2
- Klypin A., Prada F., 2018, *MNRAS*, 478, 4602
- Klypin A., Prada F., 2019, *MNRAS*, 489, 1684
- Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, *MNRAS*, 457, 4340
- Kobayashi Y., Nishimichi T., Takada M., Miyatake H., 2022, *Phys. Rev. D*, 105, 083517
- Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, *ApJ*, 609, 35
- Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64
- Li Y., Hu W., Takada M., 2014, *Phys. Rev. D*, 89, 083519
- Maksimova N. A., Garrison L. H., Eisenstein D. J., Hadzhiyska B., Bose S., Satterthwaite T. P., 2021, *MNRAS*, 508, 4017
- Maraston C. et al., 2013, *MNRAS*, 435, 2764
- Masaki S., Kashino D., Ishikawa S., Lin Y.-T., 2023, *MNRAS*, 523, 5280
- Merson A., Smith A., Benson A., Wang Y., Baugh C., 2019, *MNRAS*, 486, 5737
- Merz G. et al., 2021, *MNRAS*, 506, 2503
- Mohammad F. G. et al., 2020, *MNRAS*, 498, 128
- Mohammed I., Seljak U., Vlah Z., 2017, *MNRAS*, 466, 780
- Moustakas J. et al., 2013, *ApJ*, 767, 50
- Mueller E.-M. et al., 2022, *MNRAS*, 514, 3396
- Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, *MNRAS*, 396, 19
- Nuza S. E. et al., 2013, *MNRAS*, 432, 743
- Oogi T. et al., 2023, *MNRAS*, 525, 3879
- Pakmor R. et al., 2023, *MNRAS*, 524, 2539
- Peacock J. A., Smith R. E., 2000, *MNRAS*, 318, 1144
- Peacock J. A. et al., 2001, *Nature*, 410, 169
- Percival W. J., Verde L., Peacock J. A., 2004, *MNRAS*, 347, 645
- Planck Collaboration XVI, 2014, *A&A*, 571, A16
- Planck Collaboration XIII, 2016, *A&A*, 594, A13
- Prada F., Behroozi P., Ishiyama T., Klypin A., Pérez E., 2023, preprint (arXiv:2304.11911)
- Prakash A. et al., 2016, *ApJS*, 224, 34
- Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, *ApJ*, 771, 30
- Reid B. et al., 2016, *MNRAS*, 455, 1553
- Rodríguez-Torres S. A. et al., 2016, *MNRAS*, 460, 1173
- Ross A. J. et al., 2017, *MNRAS*, 464, 1168
- Ross A. J. et al., 2020, *MNRAS*, 498, 2354
- Ruan C.-Z., Hernández-Aguayo C., Li B., Arnold C., Baugh C. M., Klypin A., Prada F., 2022, *J. Cosmol. Astropart. Phys.*, 2022, 018
- Schaye J. et al., 2023, *MNRAS*, 526, 4978
- Sefusatti E., Crocce M., Scoccimarro R., Couchman H. M. P., 2016, *MNRAS*, 460, 3624
- Skibba R. A., van den Bosch F. C., Yang X., More S., Mo H., Fontanot F., 2011, *MNRAS*, 410, 417
- Smith A., Cole S., Baugh C., Zheng Z., Angulo R., Norberg P., Zehavi I., 2017, *MNRAS*, 470, 4646
- Smith A., Cole S., Grove C., Norberg P., Zarrouk P., 2022, *MNRAS*, 516, 4529
- Springel V., 2005, *MNRAS*, 364, 1105
- Stoppacher D. et al., 2019, *MNRAS*, 486, 1316
- Szapudi I., Szalay A. S., 1998, *ApJ*, 494, L41
- Trujillo-Gomez S., Klypin A., Primack J., Romanowsky A. J., 2011, *ApJ*, 742, 16
- Vale A., Ostriker J. P., 2006, *MNRAS*, 371, 1173
- Wadekar D., Ivanov M. M., Scoccimarro R., 2020, *Phys. Rev. D*, 102, 123521
- Wechsler R. H., Somerville R. S., Bullock J. S., Kolatt T. S., Primack J. R., Blumenthal G. R., Dekel A., 2001, *ApJ*, 554, 85
- Yu J. et al., 2022, *MNRAS*, 516, 57
- Yu B., Seljak U., Li Y., Singh S., 2023, *J. Cosmol. Astropart. Phys.*, 2023, 057
- Zel'dovich Y. B., 1970, *A&A*, 5, 84
- Zhao C. et al., 2021, *MNRAS*, 503, 1149
- Zhou R. et al., 2021, *MNRAS*, 501, 3309
- de la Torre S. et al., 2013, *A&A*, 557, A54

## APPENDIX A: CONTENT OF THE PUBLIC CATALOGUES

Below is a list of the columns of each data set, along with a short description.

### A1 UCHUU lightcones

Each UCHUU lightcone has the following columns:

- (i) ra: right ascension (degrees).
- (ii) dec: declination (degrees).
- (iii) z\_cos: cosmological redshift.
- (iv) z\_obs: observed redshift (accounting for peculiar velocities).
- (v) nz: number density of galaxies,  $n(z)$ , at the redshift of the galaxy ( $h^3\text{Mpc}^{-3}$ ).
- (vi) logMst11: logarithm ( $\log_{10}$ ) of the galaxy stellar mass ( $M_\odot$ ).
- (vii) poly\_id: healpix pixel ID containing the galaxy.
- (viii) poly\_w: fibre-collision completeness in the healpix pixel containing the galaxy.
- (ix) w\_cp: fibre collision correction weight (NNW).
- (x) w\_fkp: FKP weight.
- (xi) w\_npof: only in COMB sample lightcones. Extra weight to adjust the contribution of each sample (LOWZ and CMAS).

In addition, the following columns describe the properties of the DM host (sub)haloes of galaxies. These are taken from the original values in the UCHUU simulation halo catalogues:

- (i) Mvir\_all: mass enclosed within the virial overdensity, including unbound particles ( $h^{-1}M_\odot$ ).
- (ii) Rvir: virial radius ( $h^{-1}\text{kpc}$ ).
- (iii) rs: scale radius of a fitted NFW profile (comoving  $h^{-1}\text{kpc}$ ).
- (iv) Cvir: dark matter concentration within Rvir.

- (v)  $v_{\text{peak}}$ : peak value of  $V_{\text{max}}$  over the halo history (physical  $\text{kms}^{-1}$ ).
- (vi)  $\text{id}$ : ID that identify each (sub)halo (unique in the whole simulation).
- (vii)  $\text{pid}$ : ID of the parent central halo for satellite haloes,  $-1$  for central haloes.
- (viii)  $x, y, z$ : 3D position coordinates. The coordinate system of the UCHUU cubic box has been shifted so that the observer is at the origin, and there are periodic replications of the box (comoving  $h^{-1}\text{Mpc}$ ).
- (ix)  $v_x, v_y, v_z$ : 3D velocity (physical  $\text{kms}^{-1}$ ).
- (x)  $v_{\text{rms}}$ : velocity dispersion (physical  $\text{kms}^{-1}$ ).
- (xi)  $v_{\text{los}}$ : velocity vector projected along line of sight, with the observer positioned at the origin ( $\text{kms}^{-1}$ ).

Random catalogues, containing 40 times the number of galaxies of an individual catalogue, are also provided.

## A2 GLAM lightcones

Each GLAM lightcone has the following columns:

- (i)  $\text{gal\_id}$ : indicates whether the galaxy is central or a satellite (one for centrals, two for satellites).
- (ii)  $\text{ra}$
- (iii)  $\text{dec}$
- (iv)  $\text{z\_obs}$
- (v)  $\text{nz}$
- (vi)  $\text{logMstll}$
- (vii)  $\text{poly\_id}$
- (viii)  $\text{poly\_w}$
- (ix)  $\text{w\_cp}$
- (x)  $\text{w\_fcp}$
- (xi)  $\text{w\_ncf}$

In addition, the following columns describe the DM host halo (not subhaloes) properties of galaxies, taken from the original values found in the GLAM simulation halo catalogues,

- (i)  $\text{Mtotal}$ : halo virial mass ( $h^{-1}\text{M}_{\odot}$ ).
- (ii)  $\text{Rvir}$
- (iii)  $\text{rs}$

Random catalogues, containing 30 times the number of galaxies of an individual catalogue, are also provided.

## A3 BOSS and eBOSS data samples

In order to facilitate any analysis, we also make available the observed galaxy samples used in this work. The columns are as follows:

- (i)  $\text{RA}$ : right ascension (degrees).
- (ii)  $\text{DEC}$ : declination (degrees).
- (iii)  $\text{Z}$ : measured redshift, including fibre collisions.
- (iv)  $\text{NZ}$ : number density of galaxies,  $n(z)$ , at the redshift of the galaxy ( $h^3\text{Mpc}^{-3}$ ).
- (v)  $\text{WEIGHT\_CP}$ : fibre collision correction weight.
- (vi)  $\text{WEIGHT\_NOZ}$ : redshift failures weight.
- (vii)  $\text{WEIGHT\_SYSTOT}$ : total angular systematic weight.
- (viii)  $\text{WEIGHT\_FKP}$ : FKP weight.

Random catalogues, containing 40 times the number of galaxies of an individual catalogue, are also provided.

## APPENDIX B: IMPACT OF SIMULATION VOLUME ON COVARIANCE MATRICES

Covariances of correlation functions and power spectrum measure the statistical uncertainties in clustering statistics. They are widely used to estimate statistical errors in measured clustering signals and, ultimately, in cosmological parameters. The main sources of these errors are well understood (e.g. Li, Hu & Takada 2014; Mohammed, Seljak & Vlah 2017; Klypin & Prada 2018, 2019). Three terms define the covariances of the power spectrum: (1) Gaussian fluctuations due to finite number of Fourier harmonics, (2) non-linear fluctuations that are not related to the number of harmonics, and (3) fluctuations due to waves longer than the simulation box. Similar terms exist for the correlation functions, with the first term resulting from Gaussian fluctuations in the number of pairs at a given separation. The first two terms of covariances scale inversely with the volume of the observational sample (Mohammed et al. 2017; Klypin & Prada 2018). The third term, associated with the superscale modes, is very small for simulations with box sizes  $\gtrsim 1h^{-1}\text{Gpc}$  (Klypin & Prada 2018, 2019) and is neglected in this analysis.

In order to re-scale covariances, one must know the effective volume covered by either observational samples or simulated galaxy distributions. While it is straightforward to determine the volume for configurations with volume-limited samples or simulations with cubic boxes, the process becomes more complex with typical observational samples from large-scale galaxy surveys. In these cases, the number density of selected galaxies varies with distance,  $n(z)$ , and the effective volume can be expressed as

$$V_{\text{eff}} = \int \left( \frac{n(z) P_{\text{FKP}}}{1 + n(z) P_{\text{FKP}}} \right)^2 dV, \quad (\text{B1})$$

where  $P_{\text{FKP}} = 10^4 h^{-3}\text{Mpc}^3$ . In this equation, the denominator aims to partially compensate for the incompleteness of the selection function by assigning larger contribution (a higher weight) to pairs of objects with small  $n(z)$  (Feldman et al. 1994). However, this weighting does not change the number of pairs, which ultimately defines the covariance matrix. For instance, even if a pair of galaxies at a large distance is weighted 10 times more than a pair at a small distance, it still counts as only one pair, and its contribution to the overall noise is equivalent to that of the pair at the smaller distance.

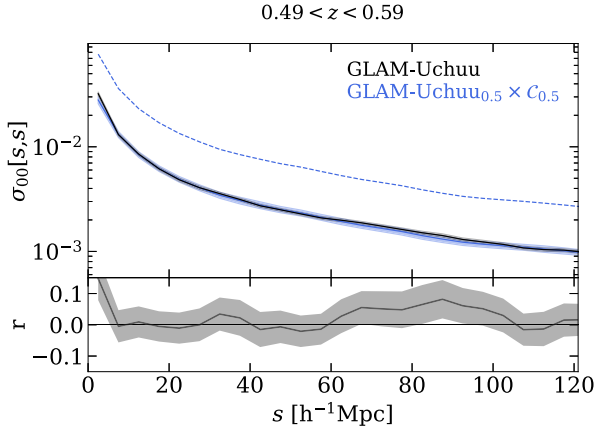
The effective volume for covariances,  $V_{\text{eff, cov}}$ , is defined as a volume that (i) contains the same number of objects as the observed sample,  $N$ ; (ii) has a constant number density; and (iii) provides the same number of pairs as in a random catalogue with number density  $n(z)$ . If  $n_0$  is the average number density in  $V_{\text{eff, cov}}$ , then conditions (i) and (iii) can be expressed as follows:

$$N = n_0 V_{\text{eff, cov}} = \int n(z) dV, \quad \text{and} \quad n_0^2 V_{\text{eff, cov}} = \int n^2(z) dV.$$

By combining these relations, we can determine the effective volume for covariances:

$$V_{\text{eff, cov}} = \frac{\left[ \int n(z) dV \right]^2}{\int n^2(z) dV}. \quad (\text{B2})$$

When the effective volume of a lightcone,  $V_{\text{eff}}$ , is smaller than the volume of the simulation,  $V_{\text{sim}}$ , the resulting covariances remain unaffected by the simulation volume. In this scenario, no re-scaling of covariances is needed. However, if  $V_{\text{eff}} > V_{\text{sim}}$ , box replications are required to cover the desired volume. In such cases, the covariance matrix should be scaled according to equation (13), with  $\mathcal{C}$  defined as  $\mathcal{C} = V_{\text{sim}}/V_{\text{eff, cov}}$ . This section aims to test this re-scaling procedure. To achieve this, we construct 500 lightcones using GLAM-UCHUU



**Figure B1.** Diagonal components of the covariance matrix for the 2PCF monopole,  $\sigma_{00} = \sqrt{\text{Cov}_{00}}$ , of the CMASS-N subsample within  $0.49 < z < 0.59$  are displayed. The black solid curve represents estimates for GLAM-UCHUU. Due to the large volume, these results are not affected by the simulation box size. For reference, we include the GLAM-UCHUU<sub>0.5</sub> covariance without box size correction (blue dashed line). The shaded areas indicate the  $1\sigma$  standard deviation for our set of lightcones. The bottom panel shows the residuals between GLAM-UCHUU and GLAM-UCHUU<sub>0.5</sub>  $\times C_{0.5}$  covariances.

simulations with box sizes  $0.5 h^{-1}\text{Gpc}$  and  $2000^3$  dark matter particles, denoted as GLAM-UCHUU<sub>0.5</sub>. Both GLAM-UCHUU (with  $1h^{-1}\text{Gpc}$  boxes) and GLAM-UCHUU<sub>0.5</sub> simulations are conducted using the same cosmological parameters.

For this test, we estimate the CMASS-N correlation functions for galaxies within the redshift range  $0.49 < z < 0.59$ , equivalent to an effective volume (for covariances) of  $V_{\text{eff}} = 0.55 h^{-3} \text{Gpc}^3$  ( $V_{\text{eff,cov}} = 0.89 h^{-3}\text{Gpc}^3$ ), as defined by equations (B1) and (B2). This  $V_{\text{eff}}$  is smaller than  $V_{\text{sim}} = 1 h^{-3}\text{Gpc}^3$ , so the covariances of GLAM-UCHUU do not need re-scaling. However, it is much larger than  $V_{\text{sim}} = 0.5 h^{-3}\text{Gpc}^3$  of GLAM-UCHUU<sub>0.5</sub> simulations, which require significant re-scaling. The required re-scaling factor for this scenario is calculated to be  $C_{0.5} = 0.5^3/0.89 \approx 0.14$ .

In Fig. B1, we present the diagonal components of the 2PCF monopole covariance matrix,  $\sigma_{00} = \sqrt{\text{Cov}_{00}}$ , for the GLAM-UCHUU and GLAM-UCHUU<sub>0.5</sub> lightcones. For the smaller box simulation, GLAM-UCHUU<sub>0.5</sub>, the covariance without box size correction surpasses that estimated by GLAM-UCHUU. This increase is expected due to the reduced number of independent pairs available in smaller simulation volumes compared to the larger observational sample. After applying the re-scaling factor, we find very good agreement between both simulations, with residuals within  $\sim 5$ – $10$  per cent. Importantly, there's no indication that the re-scaling depends on the distance  $s$ . Despite the covariances of the correlation function changing by almost a factor of 100 from the smallest to the largest scale, the results remain consistent within the  $1\sigma$  statistical error.

For samples discussed in the paper, only CMASS-N requires re-scaling. For this sample, we estimate that the effective volume for covariances is  $V_{\text{eff,cov}} = 1.68 h^{-3}\text{Gpc}^3$ . In this case, we correct the GLAM-UCHUU covariance matrix using factor  $C \approx 0.6$ .

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.