The Uchuu–SDSS galaxy light-cones: a clustering, redshift space distortion and baryonic acoustic oscillation study

C A Dong-Páez, A Smith, A O Szewciw, J Ereza, M H Abdullah, C Hernández-Aguayo, S Trusov, F Prada, A Klypin, T Ishiyama, et al.

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **528,** 7236–7255 (2024) https://doi.org/10.1093/mnras/stae062
Advance Access publication 2024 January 9

The Uchuu–SDSS galaxy light-cones: a clustering, redshift space distortion and baryonic acoustic oscillation study

C. A. Dong-Páez [1,2,3★] A. Smith [3,4,5] A. O. Szewciw,[6] J. Ereza [1] M. H. Abdullah [7,8] C. Hernández-Aguayo [9,10] S. Trusov,[11] F. Prada,[1] A. Klypin,[12,13] T. Ishiyama [7] A. Berlind,[6] P. Zarrouk,[11] J. López Cacheiro[14] and J. Ruedas[1]

[1]*Instituto de Astrofísica de Andalucía (CSIC), Glorieta de la Astronomía, E-18080 Granada, Spain*
[2]*Institut d'Astrophysique de Paris, Sorbonne Universités, CNRS, UMR 7095, 98 bis bd Arago, 75014 Paris, France*
[3]*Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*
[4]*IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France*
[5]*Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK*
[6]*Vanderbilt University, 2201 West End Ave, Nashville, TN, 37235 USA*
[7]*Digital Transformation Enhancement Council, Chiba University, 1-33, Yayoi-cho, Inage-ku, Chiba 263-8522, Japan*
[8]*Department of Astronomy, National Research Institute of Astronomy and Geophysics, Cairo 11421, Egypt*
[9]*Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str 1, D-85748 Garching, Germany*
[10]*Excellence Cluster ORIGINS, Boltzmannstrasse 2, D-85748 Garching, Germany*
[11]*Sorbonne Université, Université Paris Diderot, Sorbonne Paris Cité, CNRS, Laboratoire de Physique Nucléaire et de Hautes Energies (LPNHE), 4 place Jussieu, F-75252, Paris Cedex 5, France*
[12]*Astronomy Department, New Mexico State University, Las Cruces, NM 88001, USA*
[13]*Department of Astronomy, University of Virginia, Charlottesville, VA 22904, USA*
[14]*Centro de Supercomputación de Galicia (CESGA), Avenida de Vigo, s/n Campus Sur, E-15705 Santiago de Compostela, Spain*

Accepted 2024 January 5. Received 2023 December 8; in original form 2022 August 1

ABSTRACT

We present the data release of the Uchuu–SDSS galaxies: a set of 32 high-fidelity galaxy light-cones constructed from the large Uchuu 2.1 trillion particles *N*-body simulation using Planck cosmology. We adopt subhalo abundance matching to populate the Uchuu-box halo catalogues with SDSS galaxy luminosities. These box catalogues generated at several redshifts are combined to create a set of light-cones with redshift-evolving galaxy properties. The Uchuu–SDSS galaxy light-cones are built to reproduce the footprint and statistical properties of the SDSS main galaxy survey, along with stellar masses and star formation rates. This facilitates a direct comparison of the observed SDSS and simulated Uchuu–SDSS data. Our light-cones reproduce a large number of observational results, such as the distribution of galaxy properties, galaxy clustering, stellar mass functions, and halo occupation distributions. Using simulated and real data, we select samples of bright red galaxies at $z_{\rm eff} = 0.15$ to explore redshift space distortions and baryon acoustic oscillations (BAO) by fitting the full two-point correlation function and the BAO peak. We create a set of 5100 galaxy light-cones using GLAM *N*-body simulations to compute covariance errors. We report a ~ 30 per cent precision increase on $f\sigma_8$ and the pre-reconstruction BAO scale, due to our better estimate of the covariance matrix. From our BAO-inferred $\alpha_\parallel$ and $\alpha_\perp$ parameters, we obtain the first SDSS measurements of the Hubble and angular diameter distances $D_{\rm H}(z = 0.15)/r_d = 27.9^{+3.1}_{-2.7}$, $D_{\rm M}(z = 0.15)/r_d = 5.1^{+0.4}_{-0.4}$. Overall, we conclude that the Planck Λ CDM cosmology nicely explains the observed large-scale structure statistics of SDSS. All data sets are made publicly available.

Key words: methods: numerical – surveys – galaxies: haloes – dark matter – large-scale structure of Universe.

1 INTRODUCTION

Galaxy redshift surveys, measuring the spatial distribution of galaxies throughout cosmic time, have proven to be key observational probes for constraining cosmological models and astrophysical phenomena. For example, the distribution of galaxies on large scales can be used to infer cosmological parameters (e.g. Alam et al. 2021), as well as to constrain the properties of dark energy that drives the accelerated expansion of the Universe (Riess et al. 1998; Perlmutter et al. 1999). On smaller scales, the galaxy clustering signal encodes information on how galaxies populate dark matter (DM) haloes, on star formation, feedback and other baryonic processes that shape galaxy formation (see Wechsler & Tinker 2018, for a review).

In order to connect galaxy redshift surveys to theoretical predictions, it is essential to generate high-fidelity galaxy catalogues from cosmological simulations that capture the expected properties and clustering of the observed galaxy sample (e.g. de la Torre

★ E-mail: chiandongpaez@gmail.com

et al. 2013; White, Tinker & McBride 2014; Rodríguez-Torres et al. 2016; Lin et al. 2020). Hence, they can be used to increase the amount of information that can be extracted from galaxy survey data. First, the cosmological parameters of the simulation are known exactly. This allows us to assess the systematic and statistical errors of cosmological measurements from galaxy surveys, compute the covariance errors, test the performance of statistical analyses, and aid the theoretical interpretation of survey results. Secondly, high-fidelity catalogues based on simulations allow us to assess the systematics arising from such observational effects as selection function and fibre collisions. These effects are a source of incompleteness in the survey sample. They must be understood in order to minimize their impact in the measured clustering signal (see Smith et al. 2019, and references therein). In particular, modelling fibre collisions is generally a difficult task, since it demands a high-resolution simulation able to resolve the small scales at which fibre collisions operate while having a large enough volume to cover the extent of the survey at hand.

In the last two decades, vast observational efforts such as the 2dF Galaxy Redshift Survey (2dFGRS) (Colless et al. 2001) and the Sloan Digital Sky Survey (SDSS) saga (Alam et al. 2021), have driven most of the major discoveries about the Large-Scale Structure (LSS) of our Universe. The relevance of the SDSS surveys is not purely historical – to date, SDSS continues to be the largest reference galaxy data base, offering sky positions, redshifts, spectra and images for millions of galaxies. New technological advances continue to push the depth and volume of galaxy surveys – the new generation of galaxy surveys, which include the dark energy spectroscopic instrument (DESI) survey (DESI Collaboration et al. 2016), the Large Synoptic Survey Telescope (LSST) (Ivezić et al. 2019), the Subaru Prime Focus Spectrograph (PFS) (Takada et al. 2014) and the Euclid survey (Laureijs et al. 2011), aim to produce unprecedentedly large data sets in an effort to map the Universe to even higher precision.

In order to generate high-fidelity simulated galaxy light-cones for these large surveys, cosmological simulations with high resolution in a large volume are needed. Although it would be desirable to create mocks from hydrodynamical simulations, in which the formation of galaxies is modelled self-consistently by solving the coupled evolution of both baryons and DM, such simulations are complex and computationally expensive. The largest to date do not exceed a few hundred Mpc in box size (e.g. Dubois et al. 2014; Schaye et al. 2015; Pillepich et al. 2018; Springel et al. 2018). Hydrodynamical simulations are also strongly affected by the modelling uncertainties of complex baryonic processes. Thus, one usually resorts to DM-only simulations. Since galaxy formation physics is not included, DM haloes must be populated with galaxies in order to produce the desired simulated galaxy catalogue. This can be done using a variety of methods, each comprising different assumptions about the galaxy–halo connection (Wechsler & Tinker 2018).

A common option to generate galaxies from DM haloes is to use empirically based methods, such as subhalo abundance matching (SHAM; e.g. Marinoni & Hudson 2002; Kravtsov et al. 2004; Vale & Ostriker 2004; Conroy, Wechsler & Kravtsov 2006; Rodríguez-Torres et al. 2016; Safonova, Norberg & Cole 2021). In its simplest form, the main underlying assumption is that every halo and subhalo contains a galaxy and that there are correlations between halo and galaxy properties. For instance, the most massive and luminous galaxies are generally assumed to reside in the most massive haloes. At face value, one can assign galaxies to haloes by generating a rank-ordered relation between observed galaxy luminosities and simulated halo masses. However, such a one-to-one relation between galaxies and haloes is incompatible with observations (e.g. Trujillo-Gomez et al. 2011; Shu et al. 2012) – an intrinsic scatter must be incorporated in the method to produce realistic catalogues. Furthermore, the clustering of haloes is observed to depend on properties other than halo mass, including the halo formation time, concentration, and spin (e.g. Wechsler 2001; Gao, Springel & White 2005; Wechsler et al. 2006). Empirical models such as SHAM or the popular HOD method (e.g. Zehavi et al. 2011) are computationally fast, and are able to work around the uncertainties in the physics of galaxy formation by constraining the model parameters directly with data. Applications of the SHAM method are able to reproduce observed properties of galaxies in large surveys as the luminosity and stellar mass functions or the luminosity and colour-dependent clustering to high accuracy (e.g. Trujillo-Gomez et al. 2011; Rodríguez-Torres et al. 2016).

From real and simulated survey data, the nature of dark energy can be probed for instance by measuring the growth rate of structure, defined in linear theory as $f(a) = \mathrm{dln}D(a)/\mathrm{dln}a$, where a is the scale factor and $D(a)$ is the linear growth function. This parameter can be regarded as a measure of the energy content of the Universe, allowing us to constrain different theories of gravity and dark energy (see e.g. Peebles 1980; Guzzo et al. 2008, for a review). Galaxy peculiar velocities introduce anisotropies in their observed redshifts. This effect, first described in Kaiser (1987), is known as Redshift Space Distortions (RSD). Measurements of RSD in galaxy surveys yield $f\sigma_8$, where $\sigma_8(z)$ is the normalization of the power spectrum at redshift z on a scale of 8 h^{-1} Mpc. Another source of anisotropy comes from the choice of the fiducial cosmology adopted in the clustering analysis, used for converting redshifts and angles to comoving coordinates. If the fiducial cosmology differs from the observed one, galaxy clusters will appear flattened or elongated. The study of this so-called Alcock–Paczynski effect can provide an additional source of cosmological information from the data, as well as a means to validate a cosmological model from a simulated galaxy survey.

In addition to RSD, one can measure the baryon acoustic oscillation (BAO) feature from the two-point clustering statistics in galaxy surveys This allows us to determine the expansion rate of the Universe (e.g. Cole et al. 2005; Eisenstein et al. 2005). Recently, Alam et al. (2021) reported cosmological implications from two decades of SDSS spectroscopic surveys based on clustering measurements from galaxies and quasars in the redshift range $0.1 < z < 3$. This includes the SDSS DR7 main galaxy sample (MGS), at low redshift $z \sim 0.1$, which clustering statistics is compared in this work with the high-fidelity simulated galaxy light-cones built from on our 2.1 trillion N-body Uchuu simulation (Ishiyama et al. 2021).

Our Uchuu–SDSS light-cones are built using the SHAM method to reproduce the basic properties of the SDSS galaxy population, match its sky footprint and selection function, and include the effect of fibre collisions, which facilitates their straightforward comparison with the SDSS MGS data. Furthermore, we study and measure RSD and BAO in the real and simulated data to test the Planck base ΛCDM cosmology model (Planck Collaboration VI 2020). The high resolution and large volume of the simulation allows us to perform high-precision measurements, and model the effect of fibre collisions. The Uchuu–SDSS galaxy catalogues include sky positions, redshifts, *ugriz* apparent and absolute magnitudes stellar masses and star formation rates (SFR), as well as several halo properties. The catalogues are made publicly available at the Skies & Universes website.[1]

[1] http://www.skiesanduniverses.org/Simulations/Uchuu/

The structure of this paper is as follows. In Section 2, we introduce the SDSS MGS data and define the volume-limited galaxy samples used to validate our simulated galaxy catalogues. In Section 3, we describe the Uchuu simulation and the methodology behind the creation of galaxies from the DM halo properties. This includes the construction of light-cones from cubic boxes, the implementation of fibre collisions, and the assignment of additional SDSS galaxy properties. Section 4 presents the basic properties of the Uchuu–SDSS light-cones, compared to the SDSS data. In particular, we explore the statistics of several galaxy properties, the galaxy clustering dependence on luminosity, colour, and stellar mass, and the galaxy HOD. In Section 5, we present our RSD and BAO measurements both in the real and simulated data and compare them with the fiducial Planck cosmology. Finally, in Section 6, we present a summary of our results.

2 SDSS GALAXY SAMPLES

2.1 Parent and volume-limited samples

In this work, we build and compare Uchuu simulated galaxies to observational data from the seventh data release (DR7; Abazajian et al. 2009) of the SDSS (York et al. 2000). More specifically, we make use of the large-scale structure catalogue of the SDSS MGS from the NYU Value Added Galaxy Catalogue (NYU–VAGC; Blanton et al. 2005). We restrict our sample to consist only of galaxies in the contiguous northern footprint in regions which have a completeness of > 90 per cent. This parent sample covers an effective area of $\sim$6511 deg^2 and contains $\sim$ 497 000 galaxies with Petrosian r-band magnitudes in the range $14.5 < r_{\rm pet} < 17.6$. In this sample, ~ 6 per cent of targeted galaxies lack a spectroscopically measured redshift due to fibre collisions. We apply a nearest neighbour correction to these galaxies, assigning to them the redshift of the galaxy with which they collided.

Apparent magnitudes can be linked to absolute magnitudes via the distance modulus equation,

$$r = {}^{0.1}M_{\rm r}^{h} + 5\log_{10} D_{\rm L}(z) + 25 + {}^{0.1}k_{\rm r}(z), \quad (1)$$

where $D_{\rm L}(z)$ is the luminosity distance (in units of h^{-1} Mpc), ${}^{0.1}k_{\rm r}(z)$ is the k-correction, which takes into account the shift in the bandpass with redshift.

Throughout this paper, we denote k-corrected unevolved absolute magnitudes in the r band as ${}^{0.1}M_{\rm r}^{h} \equiv {}^{0.1}M_{\rm r} - 5\log_{10} h$, where the superscript 0.1 indicates that the rest-frame magnitude has been k-corrected to a reference redshift of $z_{\rm ref} = 0.1$ (Blanton et al. 2003a). A similar notation is used for magnitudes in other bands. We also denote the rest-frame $g - r$ colours as ${}^{0.1}(g - r)$, k-corrected to the same reference redshift.

Additionally, the absolute magnitudes are corrected for passive evolution using the model of Blanton (2006). Specifically, the 'evolved' magnitude ${}^{0.1}M_{r}^{h,e}$ is given by

$${}^{0.1}M_{\rm r}^{h,e} = {}^{0.1}M_{\rm r}^{h} + E(z), \quad (2)$$

where the evolution correction $E(z)$ is given by

$$E(z) = q_0(1 + q_1(z - q_{z0}))(z - q_{z0}). \quad (3)$$

This model contains three parameters which we set to values of $q_0 = 2$, $q_1 = -1$, and $q_{z0} = 0.1$.

From this parent sample of galaxies, we construct nine different volume-limited samples, each of which is complete down to a specified r-band magnitude, ${}^{0.1}M_{r}^{\rm max}$, which is k-corrected and corrected for evolution. For our samples, we use the same magnitude thresholds and maximum redshifts as Guo et al. (2015). In all samples, to minimize the impact of peculiar velocities and spurious objects, we keep only galaxies with redshifts $z > 0.02$. In Table 1, for each sample we provide the magnitude threshold, maximum redshift, number of galaxies, number density of galaxies, and effective volume. These quantities can be compared directly to those in table 1 of Guo et al. (2015). We also show the fraction of galaxies in each sample that requires the nearest neighbour correction due to fibre collisions, see Section 3.3.5 for the details.

Table 1. For each SDSS volume-limited sample, the columns list (from left to right): the absolute magnitude threshold, the maximum redshift, the number of galaxies, the galaxy number density (in $10^{-3}\ h^3$ Mpc^{-3}), the effective volume of each sample (in $10^6\ h^{-3}$ Mpc^3), and the fraction of galaxies requiring the nearest neighbour correction. The lower redshift cut for all samples is $z_{\rm min} = 0.02$. Magnitudes of all galaxies are k-corrected and passively evolved to the survey's median redshift of $z = 0.1$ and computed assuming $h = 1$.

${}^{0.1}M_{r}^{\rm max}$	$z_{\rm max}$	N	n_g	$V_{\rm eff}$	$f_{\rm NN}$
−18.0	0.041	35 359	31.95	1.11	0.043
−18.5	0.053	49 272	20.41	2.54	0.046
−19.0	0.064	62 534	14.47	4.55	0.050
−19.5	0.085	112 652	11.09	10.68	0.058
−20.0	0.106	119 734	6.13	20.53	0.062
−20.5	0.132	112 496	3.03	39.03	0.068
−21.0	0.159	71 795	1.13	66.95	0.079
−21.5	0.198	33 505	0.28	125.63	0.099
−22.0	0.245	9820	0.045	218.32	0.149

Finally, in order to study in greater detail the BAO signal in the SDSS MGS, we define a BAO sample in which this signal is enhanced, analogously to Ross et al. (2015). This sample, named SDSSbao, is defined by the cuts,

$$\begin{aligned} &0.07 < z < 0.2 \\ &{}^{0.1}M_{r}^{h} < -21.2 \\ &{}^{0.1}(g - r) > 0.8 \\ &14.5 < r < 17.6 \\ &c > 0.9, \end{aligned} \quad (4)$$

where c is the completeness.

2.2 Stellar masses and star formation rates

Uchuu–SDSS galaxy catalogues list stellar masses M_* and star formation rates SFRs using two independent sources, which are the MPA/JHU catalogue[2] (the Max Planck Institute for Astrophysics and the Johns Hopkins University) and the Granada Group catalogue.[3]

For the MPA/JHU catalogue, M_* was calculated following the methodologies presented in Kauffmann et al. (2003) and Salim et al. (2007). The stellar masses in the MPA/JHU catalogue, are consistent with other estimates (e.g. Taylor et al. 2011; Chang et al. 2015; Duarte & Mamon 2015; Leslie et al. 2016).

For galaxies classified as star-forming (SF) in the MPA/JHU catalogue, SFRs are calculated using the nebular emission lines within the spectroscopic fibre aperture of 3 arcsec as described in Brinchmann et al. (2004). SFRs are calculated using the empirical calibration of H_α emission lines (Kennicutt 1998) and corrected from the dust extinction with the Balmer decrement H_α/H_β (Charlot & Fall 2000), assuming a Kroupa initial mass function (Kroupa 2001).

[2] http://www.mpa-garching.mpg.de/SDSS/DR7/

[3] https://www.sdss.org/dr17/spectro/galaxy_granada/. We use the data from the stellarMassFSPSGranEarlyDust table.

The SFR contribution outside of the fibre is estimated from the galaxy photometry following Salim et al. (2007). Finally, the specific star formation rate (sSFR, defined as SFR/M_*) was calculated by combining the SFR and stellar mass likelihood distributions as outlined in Appendix A of Brinchmann et al. (2004). Throughout the paper, we use the median values of the resulting probability distribution functions as the sSFR of a galaxy.

For the Granada Group catalogue, M_* and SFR are calculated using the SDSS spectroscopic redshift and *ugriz* magnitudes by means of broad-band spectral energy distribution (SED) fitting via Flexible stellar population synthesis technique (FSPS, Conroy, Gunn & White 2009). Our Uchuu–SDSS galaxy catalogues include M_* and sSFR obtained from the early formation time and dust attenuation model (Charlot & Fall 2000) with Kroupa (2001) initial mass functions.

3 CONSTRUCTING THE UCHUU–SDSS CATALOGUES

3.1 The Uchuu Simulation

The Uchuu simulation is a large high-resolution N-body cosmological simulation, the largest simulation in the Uchuu suite (Ishiyama et al. 2021). It follows the evolution of 2.1 trillion ($12\,800^3$) DM particles with particle mass resolution of $3.27 \times 10^8\,h^{-1}M_\odot$ in a $(2.0\,h^{-1}Gpc)^3$ comoving periodic box. The simulation adopts the Planck Λ CDM cosmological parameters: $\Omega_m = 0.3089$, $\Omega_b = 0.0486$, $\Omega_\Lambda = 0.6911$, $h = 0.6774$, $n_s = 0.9667$, and $\sigma_8 = 0.8159$ (Planck Collaboration XIII 2016). Starting at $z = 127$, the subsequent gravitational evolution was solved down to $z = 0$ using the TREEPM code GREEM (Ishiyama, Fukushige & Makino 2009; Ishiyama, Nitadori & Makino 2012), with a gravitational softening length of $4.27\,h^{-1}$kpc.

A set of 50 particle snapshots ranging from $z = 14$ to $z = 0$ were saved, from which bound structures were identified by running the ROCKSTAR phase–space halo/subhalo finder (Behroozi, Wechsler & Wu 2013a). Then merger trees were constructed using the CONSISTENT TREES code (Behroozi et al. 2013b). All Uchuu data products are publicly available.

The characteristics of Uchuu make it ideal for the creation of simulated galaxy catalogues. Its high resolution allows resolving DM haloes down to small halo masses on a large volume. This renders the simulation suitable for the application of SHAM, which requires that subhaloes are identified. The very large volume of Uchuu allows for detailed statistics, as well as the study of large-scale clustering features such as the BAO.

3.2 Box galaxy catalogues

In order to construct Uchuu galaxy light-cones, we first start by applying the SHAM algorithm in simulation boxes. Following the basic principle of SHAM, we assign galaxy luminosities to DM haloes by matching the galaxy luminosity function to the cumulative distribution function of a halo property that serves as a proxy of galaxy stellar mass. A possible choice for this halo property is the maximum circular velocity, V_{max} (e.g. Conroy et al. 2006; Trujillo-Gomez et al. 2011), defined as the maximum value of the halo circular velocity at the redshift of interest,

$$V_{max}(z) = \max\left(\sqrt{\frac{GM(<r,z)}{r}}\right). \quad (5)$$

Using V_{max} generally yields better results compared to the halo mass. Since the maximum velocity is generally reached at smaller scales compared to the halo radius, it characterizes both the halo concentration (Campbell et al. 2018) and the depth of the potential at the typical galactic scales (Chaves-Montero et al. 2016). It is also less affected by the tidal stripping suffered by subhaloes upon being accreted by larger haloes (Hayashi et al. 2003). Another typical choice, which we adopt in this work, is to use the peak circular velocity, V_{peak}, defined as the peak value of V_{max} over the history of the halo. This makes estimates that reproduce even more closely the properties of observed data (e.g. Reddick et al. 2013; Chaves-Montero et al. 2016; Safonova et al. 2021).

The subhaloes in Uchuu are 90 per cent complete down to $V_{peak} \gtrsim 70\,km\,s^{-1}$, while distinct haloes are 90 per cent complete down to $V_{peak} \gtrsim 50\,km\,s^{-1}$. This allows us to reach low-galaxy luminosities in our Uchuu galaxy catalogues.

We match V_{peak} with an observationally motivated target luminosity function, ϕ_{target}. We adopt a recipe closely following Smith et al. (2017), which interpolates between the measured luminosity function from SDSS, ϕ_{SDSS} (Blanton et al. 2003b), at low redshifts and the luminosity function from the GAMA survey, ϕ_{GAMA} (Loveday et al. 2012), at higher redshifts, as follows

$$\phi_{target}({}^{0.1}M_r^h, z) = (1 - w(z))\phi_{SDSS}({}^{0.1}M_r^h, z) + w(z)\phi_{GAMA}({}^{0.1}M_r^h, z), \quad (6)$$

where $w(z)$ is a sigmoid function describing this smooth transition at z = 0.15, which is given by

$$w(z) = \left(1 + e^{-100(z-0.15)}\right)^{-1}. \quad (7)$$

Both luminosity functions are modelled with an evolving Schechter fit

$$\phi(M) = 0.4 \ln 10\, \phi^* \left(10^{0.4(M^*-M)}\right)^{1+\alpha} \exp\left(-10^{0.4(M^*-M)}\right), \quad (8)$$

where the redshift evolution of the Schechter parameters is modelled as

$$\begin{aligned} \alpha(z) &= \alpha(z_0) \\ M^*(z) &= M^*(z_0) - Q(z - z_0) \\ \phi^* &= \phi^*(0)10^{0.4Pz}, \end{aligned} \quad (9)$$

where $z_0 = 0.1$ (Blanton et al. 2003b; Loveday et al. 2012). The corresponding parameters for the SDSS and GAMA functions are shown in Table 2. Note that, this is different from Smith et al. (2017), which used the tabulated measurement from SDSS. In our galaxy assignment algorithm, we use the cumulative LF ($n_g(< {}^{0.1}M_r^h)$), which gives the number density of galaxies brighter than the magnitude threshold.

Starting from the halo V_{peak} values, we use a SHAM algorithm to assign galaxy luminosities by matching the V_{peak} cumulative number density function to our target galaxy luminosity function, with some added intrinsic scatter. We detail below our algorithm, which is based

Table 2. Schechter parameters obtained from a fit to SDSS (Blanton et al. 2003b) and GAMA (Loveday et al. 2012) data. These two luminosity functions are interpolated in order to obtain our target luminosity function (equation 6).

Parameter	SDSS	GAMA
α	−1.05	−1.23
$\phi^*/h^3\mathrm{Mpc}^{-3}$	1.49×10^{-2}	0.94×10^{-2}
$M^*(z_0)$	−20.44	−20.70
P	0.18	1.8
Q	1.62	0.7

on the method introduced in McCullagh et al. (2017); Safonova et al. (2021),

(i) Sort the haloes by $V_{\rm peak}$ in descending order, to compute the $V_{\rm peak}$ cumulative number density function, $n_{\rm h}(>V_{\rm peak})$.
(ii) Assign an 'unscattered' galaxy r-band magnitude, ${}^{0.1}M_r^h$, to each halo by matching the above $V_{\rm peak}$ cumulative halo number density function to the target galaxy luminosity function, preserving the ranking such that large $V_{\rm peak}$ haloes host high-luminosity galaxies.

$$n_{\rm g}(<{}^{0.1}M_r^h) = n_{\rm h}(>V_{\rm peak}) \tag{10}$$

(iii) Define a new 'scattered' value of the magnitude, ${}^{0.1}M_r^{h,\rm scat}$,

$${}^{0.1}M_r^{h,\rm scat} = \mathcal{N}(0,\sigma^2) + {}^{0.1}M_r^h, \tag{11}$$

where $\mathcal{N}(0,\sigma^2)$ is a number drawn from a normal distribution with mean 0 and variance σ^2.
(iv) Sort the haloes by ${}^{0.1}M_r^{h,\rm scat}$, and compute the ${}^{0.1}M_r^{h,\rm scat}$ cumulative distribution, $n_{\rm h}(<{}^{0.1}M_r^{h,\rm scat})$.
(v) Assign the final r-band magnitude by matching the cumulative distribution of ${}^{0.1}M_r^{h,\rm scat}$ to the target cumulative distribution function,

$$n_{\rm g}(<{}^{0.1}M_r^h) = n_{\rm h}(<{}^{0.1}M_r^{h,\rm scat}). \tag{12}$$

In other words, the brightest final luminosities are assigned to the haloes with brightest ${}^{0.1}M_r^{h,\rm scat}$.

All the degrees of freedom in the procedure above are contained in the choice of the scatter parameter σ, which we can regulate. In our case, we reduce the number of tunable parameters of our model and neglect any possible redshift or luminosity dependence of σ by fixing it to a constant value of 0.5 mag. This choice, despite its simplicity, is able to reproduce the observed galaxy clustering while avoiding unphysical features in the $(V_{\rm peak}, {}^{0.1}M_r^h)$ galaxy distribution. This is shown in Fig. 1, where we show the monopole of the two-point correlation function of our Uchuu galaxy catalogue at $z = 0.093$, the catalogue closest to the median redshift of SDSS, compared to the results from our SDSS samples.

Our model is able to recover the SDSS results with good accuracy for a large range of scales and volume-limited samples. The value of σ is calibrated to reproduce the observed SDSS galaxy clustering. For more details about the calibration of σ and its effect on galaxy clustering, we direct the reader to Appendix A. We also recover the target LF by construction.

The resulting relation between $V_{\rm peak}$ and ${}^{0.1}M_r^h$ is shown in Fig. 2 for galaxies with ${}^{0.1}M_{\rm r} < -14.0$ in the $z = 0.093$ box. The large volume of Uchuu allows our mocks to reach a huge dynamical range in galaxy luminosity $-23 < {}^{0.1}M_r^h < -14$.

3.3 Uchuu–SDSS galaxy light-cones

We use a total of 6 snapshots between $z = 0$ and $z = 0.5$, which are separated in redshift by approximately 0.1 ($z_{\rm snap} = 0$, 0.093, 0.19, 0.3, 0.43, 0.49). Light-cones are created from the snapshots by joining them together in spherical shells. An observer is first placed in the box, and the Cartesian galaxy coordinates in each snapshot are converted to equatorial coordinates, and the redshift is calculated, taking into account the line-of-sight velocity. In each simulation snapshot, galaxies in the redshift range $(z_{\rm snap\text{-}1} + z_{\rm snap})/2 < z < (z_{\rm snap} + z_{\rm snap+1})/2$ are selected, which are then joined together to build the light-cone. If the redshift shell is too big to fit inside a single cubic box, periodic replications are applied. In the final light-cone there are no periodic replications below $z = 0.36$.

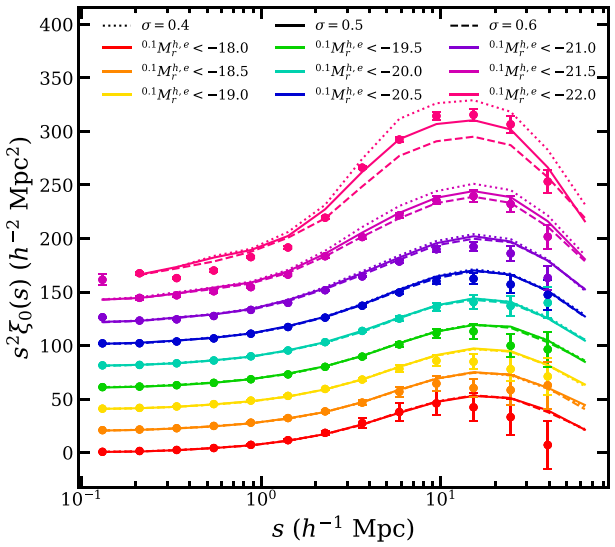

Figure 1. The monopole of the two-point correlation functions from the Uchuu box galaxy catalogue at $z = 0.093$ (solid lines) and the SDSS data (filled symbols) for several volume-limited samples corresponding to luminosity cuts as listed in Table 1. For clarity, lines corresponding to different luminosity cuts have been offset by successive intervals $20\,/\,h^2\,{\rm M}pc^2$, starting from the lowest luminosity sample. The clustering is shown for different values of the scatter parameter $\sigma = \{0.4, 0.5, 0.6\}$, by the dotted, solid, and dashed curves, respectively.

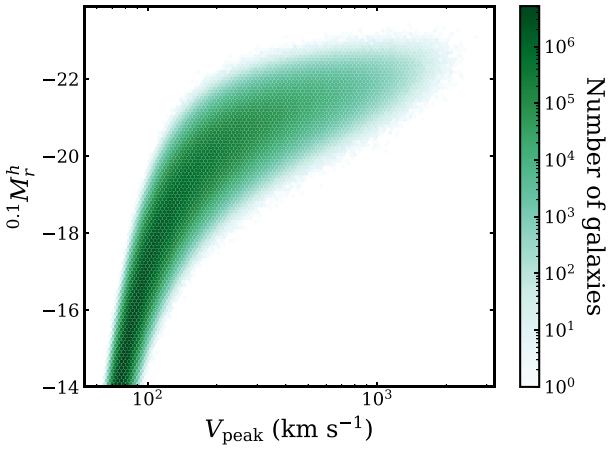

Figure 2. The $(V_{\rm peak}, {}^{0.1}M_r^h)$ distribution in the Uchuu $z = 0.093$ box, assuming a constant scatter parameter $\sigma = 0.5$. Colour indicates the number of galaxies in each hexagonal bin, on a logarithmic scale.

The full-sky light-cone is then cut to the northern contiguous region of the SDSS survey footprint, using a HEALPIX map (Gorski et al. 1999; Blanton et al. 2005; Swanson et al. 2008). The original HEALPIX map has $N_{\rm side} = 512$, but we increase the size of the pixels, using $N_{\rm side} = 128$, and keeping pixels where the completeness in the data is greater than 0.9. This results in a footprint with an area of 7261 deg^2. The SDSS footprint can be replicated across the full sky to create 4 independent SDSS simulated catalogues. By generating light-cones from eight observer positions (with coordinates at either 0 or 1 h^{-1} Gpc, along each of the three dimensions), we are therefore able to create a total of 32 Uchu–SDSS light-cones. The first 8 mocks (constructed with the observer at the origin, and at the centre of the

box) are independent below $z = 0.3$. There is some overlap in the volume between the light-cones at higher redshifts than this, but the fraction of galaxies with $z > 0.3$ is very small (~ 0.1 per cent). We refer to these light-cones as the 8 independent light-cones. The full set of 32 light-cones are independent below $z = 0.175$, and we use all 32 to improve the statistics of our RSD and BAO measurements (see Section 5). The BAO galaxy sample (equation 4) has a maximum redshift of $z = 0.2$, so there is some overlap between mocks, and ~ 25 per cent of galaxies in the sample have $z > 0.175$.

Combining multiple snapshots in this way to construct a light-cone has the issue that there are discontinuities at the boundaries between snapshots. It is possible for the same halo to appear twice at either side of the boundary, or to not appear at all, and the duplicated haloes artificially boost the pair counts on very small scales. We investigate this in Smith et al. (2022a) and find that there is a boost in the real-space clustering on small scales due to this effect. In redshift space, when velocities are included, the effect is greatly reduced. Using snapshots separated by $\sim$0.1 in redshift is a good compromise which adds evolution to the light-cone, without excessively boosting the clustering below 1 h^{-1} Mpc. The clustering measurements on the scales used in a typical RSD analysis ($\gtrsim 20\,h^{-1}$ Mpc) are insensitive to the number of snapshots used.

3.3.1 Magnitude evolution

Each simulation snapshot was constructed to reproduce the target luminosity function at the redshift of the snapshot, $z_{\rm snap}$. In the light-cone, this leads to discontinuities in these properties at the boundaries between snapshots.

In order to create a simulated light-cone with a smoothly evolving luminosity function, and smooth $n(z)$, we rescale the galaxy absolute magnitudes as a function of redshift. The original magnitude assigned to each galaxy is first mapped to the corresponding number density, using the luminosity function at $z_{\rm snap}$. This number density can then be mapped back to an absolute magnitude at the redshift of the galaxy in the light-cone, z, using the target luminosity function at the same redshift, z. By construction, this reproduces the smooth evolution of the target luminosity function.

3.3.2 Colour assignment

After galaxy luminosities have been assigned, we add ${}^{0.1}(g-r)$ colours to our simulated galaxy sample. We use an evolving empirical model for the ${}^{0.1}(g-r)$ colours, based on (Skibba & Sheth 2009; Smith et al. 2017), but improved to better reproduce the colour–magnitude diagram from the GAMA survey, at a range of redshifts (Smith et al. 2022b). The bimodality of the colour probability distribution functions is modelled as a sum of two Gaussian distributions. Defining $p({}^{0.1}(g–r); {}^{0.1}M_r^h, z) \equiv dN/N/d^{0.1}(g–r)$, where N is the number of galaxies,

$$p({}^{0.1}(g-r); {}^{0.1}M_r^h, z) = f_{\rm blue}\,\mathcal{N}_{\rm blue}(\mu_{\rm blue}, \sigma_{\rm blue}) + (1-f_{\rm blue})\,\mathcal{N}_{\rm red}(\mu_{\rm red}, \sigma_{\rm red}), \quad (13)$$

where $\mathcal{N}_{\rm blue}$ and $\mathcal{N}_{\rm red}$ are Gaussian probability distribution functions corresponding to blue and red galaxies, respectively, and $f_{\rm blue}$ is the fraction of blue galaxies. The parameters describing this double-Gaussian distribution all depend on ${}^{0.1}M_r^h$ and z.

At a fixed redshift, the mean and rms of each Gaussian, $\mu_{\rm red}({}^{0.1}M_r^h)$, $\sigma_{\rm red}({}^{0.1}M_r^h)$, $\mu_{\rm blue}({}^{0.1}M_r^h)$ and $\sigma_{\rm blue}({}^{0.1}M_r^h)$, are modelled as broken linear functions. These functions were fit to the colour–magnitude diagram measured in GAMA, in several redshift bins, which we then interpolate between. Basing the colour distributions on the data from GAMA allows us to create simulated galaxy catalogues to very faint magnitudes, which will be useful e.g. for the DESI Bright Galaxy Survey.

However, we find good agreement between these colours and the measurements from the SDSS data. The fraction of blue galaxies at a given luminosity, $f_{\rm blue}({}^{0.1}M_r^h)$, is modelled differently for central and satellite galaxies. Colours are then drawn randomly from the ${}^{0.1}(g-r)$ colour distributions described above. This method is able to reproduce by construction our target colour distributions.

3.3.3 Apparent magnitudes

The apparent r-band magnitude is computed from the absolute magnitude using equation (1). In the r band, we use a set of colour-dependent k-corrections from the GAMA data (see fig. 13 of Smith et al. 2017). These are a set of polynomial k-corrections, measured in several bins of ${}^{0.1}(g-r)$ colour. These k-corrections allow us to calculate apparent magnitudes using the information that was originally available in the simulated catalogue (r-band magnitude and $g-r$ colour). In the other bands, we use the k-corrections of Blanton et al. (2003a).

We have compared the GAMA and SDSS r-band k-corrections, and find that the median k-corrections at each redshift are in good agreement, differing by no more than $\sim$0.01. The 1σ scatter is at a level $<$0.04. At $z = 0.1$, both k-corrections agree exactly, since ${}^{0.1}k(z=0.1) = -2.5\log_{10}(1.1) \approx -0.103$.

3.3.4 Assigning galaxy properties using SDSS data

The method we have used to create the light-cone assigns each galaxy a rest-frame r-band absolute magnitude, ${}^{0.1}M_r^h$, and ${}^{0.1}(g-r)$ colour. In order to assign more properties to the mock, we match galaxies to the SDSS data. We use a k–d tree to find the closest SDSS galaxy, based on z, ${}^{0.1}M_r^h$ and ${}^{0.1}(g-r)$. Each mock galaxy is then assigned the absolute magnitude in the u, g, i, and z band of the closest-matching galaxy, in addition to its stellar mass and specific star formation rate.

The u, g, i, and z band apparent magnitudes are calculated at the redshift of the mock galaxy from the absolute magnitudes, using equation (1) and the SDSS k-corrections.

3.3.5 Modelling fibre collisions

In SDSS, spectroscopic fibres on a single plate cannot be placed closer to each other than the diameter of the fibre plugs. As a result, if two galaxies are in close proximity ($<$55 arcsec), a spectrum can only be obtained for one of them. The tiling of SDSS plates slightly alleviates this problem, with ~ 30 per cent of the SDSS footprint covered by multiple plates. These plate 'overlap regions' allow for spectroscopic redshifts to be obtained for several galaxies that are within 55 arcsec of one another. Still, ~ 6 per cent of targeted galaxies in our SDSS sample lack a spectroscopically measured redshift due to their proximity to a neighbouring galaxy.

To mimic the fibre collision effect of SDSS, when constructing our mock galaxy catalogues, we employ a procedure adapted from Szewciw et al. (2022). First, we link together galaxies into friends-of-friends 'collision groups'. A galaxy is part of a collision group if its angular distance to any galaxy in that group is $<$55 arcsec. Next, we decide whether each galaxy in a group will be 'observed' (i.e. receive its spectroscopic redshift) or 'unobserved'. When making

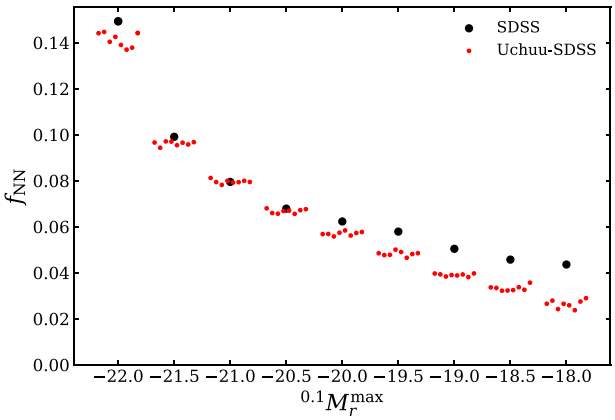

Figure 3. The fraction of galaxies affected by fibre collisions in Uchuu–SDSS and SDSS light-cones. For each of our SDSS volume-limited samples, the bigger black points show the fraction of galaxies receiving the nearest neighbour correction, $f_{\rm NN}$. The smaller red points show $f_{\rm NN}$ when we construct volume-limited samples from each of our eight parent Uchuu–SDSS catalogues.

this choice, we first attempt to maximize the number of galaxies in a collision group that could receive spectroscopic fibres from a single SDSS plate. If there does exist more than one set that maximizes the number of observed galaxies, then we randomly choose one of these sets to be observed. Next, to simulate the effect of SDSS plate overlap regions, we randomly select ~ 30 per cent of the unobserved galaxies to receive their original observed spectroscopic redshifts. The remaining unobserved galaxies are then assigned the redshift of their angular nearest neighbour. Finally, given the new redshifts of these galaxies, we recompute their *k*-corrections (using the colour-dependent *k*-corrections described in Smith et al. 2017) and absolute magnitudes.

It is important to note that this procedure does not fully mimic the role that plate overlap regions play in recovering the redshifts of collided galaxies. In our procedure, the collided galaxies whose redshifts are recovered are chosen at random from the full sky. In SDSS, by contrast, recovered galaxies lie in plate overlap regions and thus are spatially correlated. Furthermore, in SDSS, each overlap region of the sky is covered by a different number of intersecting plates. The number of overlapping plates dictates the number of galaxies whose redshifts can be spectroscopically measured in a given collision group. Our procedure, however, is agnostic with respect to the number of plates required to recover redshifts for any randomly chosen set of collided galaxies.

Despite these differences, we apply the procedure described above to our simulated galaxy catalogues. With this relatively simple procedure described above, the global fraction of galaxies affected by fibre collisions in the simulated light-cone ($f_{\rm NN} \sim 5.2$ per cent) is quite similar to that of SDSS ($f_{\rm NN} \sim 5.9$ per cent). In Fig. 3, we show a comparison of $f_{\rm NN}$ between SDSS (black points) and each of our eight independent Uchuu–SDSS light-cones (red points) for the different volume-limited samples defined in Table 1. In both Uchuu and SDSS, we see the same qualitative trend – $f_{\rm NN}$ increases as we move to more luminous samples. This is expected since more luminous galaxies tend to be more strongly clustered and thus are more likely to be affected by fibre collisions. There is a good agreement for bright volume-limited samples, although for $^{0.1}M_r^{\rm max} \leq -19.5$ Uchuu underestimates $f_{\rm NN}$ by almost a factor of 2.

After running the fibre assignment algorithm on the 32 Uchuu–SDSS light-cones, we evaluate the fibre assignment completeness in healpix pixels with $N_{\rm side} = 512$. The light-cones are then cut to pixels where the average completeness >0.9 (averaged over the 32 light-cones). This results in a final area of 6642 deg^2, which is comparable to the footprint of the SDSS data.

The 32 Uchuu–SDSS light-cone catalogues described in this section, containing *ugriz* magnitudes, stellar masses, star formation, rates and fibre assignment information, are made publicly available at Skies & Universes. A subset of the cubic box catalogues used in the construction of the light-cones, and the companion SDSS data set are also made available. Note that, our released box catalogues also include $^{0.1}(g-r)$ computed using a similar method as described in Section 3.3.2.

4 PROPERTIES OF THE UCHUU–SDSS GALAXIES

4.1 Galaxy properties

In this section, we illustrate the various galaxy properties stored in the Uchuu–SDSS light-cones and compare them with the galaxy properties in the SDSS data.

Fig. 4 shows galaxies in a thin slice of the SDSS catalogue and one of the Uchuu–SDSS light-cones, where each galaxy has been coloured based on its stellar mass, illustrating the similarities between the mock and the data. The density of galaxies falls with increasing redshift since faint galaxies with low stellar masses can only be observed close to $z = 0$, while the brightest galaxies can be observed over the full redshift range. The density of galaxies appears to be in good agreement between the Uchuu–SDSS and SDSS data.

This can be seen quantitatively in Fig. 5, which compares the redshift distribution of the 32 Uchuu–SDSS light-cones with SDSS. The Uchuu–SDSS light-cones are in good agreement with SDSS within the scatter of the 32 light-cones. Both distributions peak at $z \sim 0.08$. At higher redshifts ($z \sim 0.2$), there is a slight excess of galaxies in the Uchuu–SDSS light-cones compared to the data, due to differences in the luminosity function. While the target luminosity function used to construct the simulated light-cones is in good agreement with the SDSS data at low redshifts, we transition to the luminosity function measured in GAMA at higher redshifts. Using the GAMA luminosity function results in a higher number density of galaxies compared to the SDSS measurements, but the SDSS luminosity function is poorly constrained at this redshift.

The absolute magnitudes of galaxies from one of the Uchuu–SDSS light-cones are shown in Fig. 6, compared to SDSS. In the *r*-band magnitudes were assigned to each galaxy to match an evolving target luminosity function from SDSS and GAMA measurements, and we find that this is able to reproduce well the distribution of absolute *r*-band magnitudes in the SDSS data. The magnitudes in the other bands were assigned by matching each simulated galaxy to a galaxy in the data, based on *r*-band magnitude, $g-r$ colour and redshift. By construction, the distribution of these magnitudes is also in good agreement with the SDSS data. The relative difference between the two samples is very small except in the bins containing a small number of galaxies.

The $^{0.1}(g-r)$ colour distributions in the Uchuu–SDSS light-cones are shown in the left-hand column of Fig. 7, as a function of $^{0.1}M_{\rm r}^{h}$, in three narrow redshift bins at $z = 0.05$, $z = 0.1$, and $z = 0.15$. In each redshift bin, the colour distribution is bimodal, with a red sequence of galaxies and a cloud of blue galaxies. The brightest galaxies are red, while fainter galaxies have a higher blue fraction.

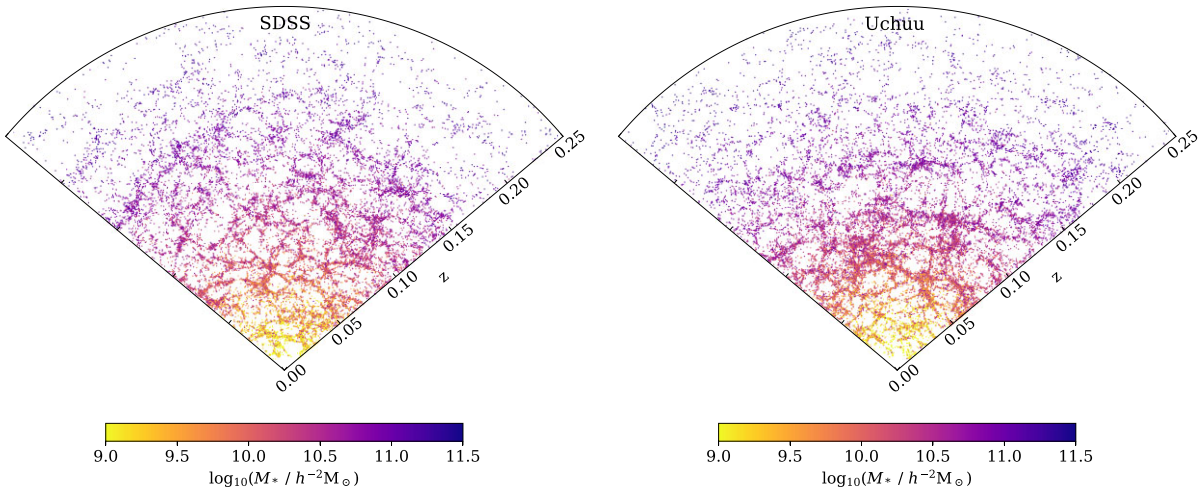

Figure 4. Slice through the SDSS galaxy catalogue (left), and Uchuu–SDSS light-cone (right), where galaxies are coloured by their stellar mass. Galaxies are shown between $35 < \mathrm{dec} < 40$ deg and $140 < \mathrm{RA} < 240$ deg.

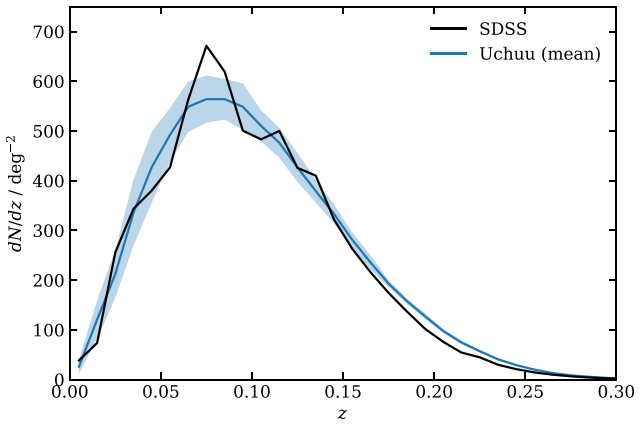

Figure 5. Redshift distribution of galaxies in the Uchuu–SDSS light-cones, compared to the SDSS data, for galaxies with $r < 17.6$. The blue curve is the mean of the 32 simulated light-cones, with the shaded region indicating the 1σ scatter. The black curve shows the redshift distribution of SDSS. Overall, we find good agreement between Uchuu and SDSS, with a small overprediction of the number of galaxies in Uchuu at $z \sim 0.2$ due to differences in the luminosity functions.

The colour distributions show good agreement with the SDSS data, reproducing the same colour evolution with redshift. There is a small discrepancy at low redshifts, where the red sequence is more sloped in Uchuu–SDSS compared to in SDSS since the colours in the Uchuu–SDSS light-cone were tuned to GAMA measurements. The right-hand column shows the same colour distributions but as a function of the stellar mass. Here, we see that the blue galaxies tend to have lower stellar masses than the red galaxies. Again, there is good agreement between the simulated and real data, with a slight discrepancy in the red sequence at low redshifts.

The distribution of star formation rates of galaxies in the light-cone is shown in Fig. 8, in comparison to the SDSS measurements. The upper panel shows stellar mass against sSFR, while $^{0.1}(g-r)$ is plotted against sSFR in the lower panel. There is a clear bimodal distribution of quiescent galaxies with low star formation rates on the left and star-forming galaxies on the right. The quiescent galaxies tend to have higher stellar masses than the star-forming galaxies. Applying a cut of 10^{11} yr^{-1} in sSFR is able to cleanly cut the galaxy catalogue into these two samples, while we can see in the lower panel that a cut in $^{0.1}(g-r)$ would not work as well.

4.2 Luminosity, mass, and colour dependence of clustering

In this section, we compare the clustering in our Uchuu–SDSS light-cones with the observational results from the volume-limited SDSS samples. We measure the first non-zero Legendre multipoles ($\ell = 0, 2, 4$) of the redshift-space two-point correlation function (TPCF), defined as,

$$\xi_\ell(s) = (2\ell+1)\int_0^1 \xi^s(s,\mu)\,L_\ell(\mu)\mathrm{d}\mu, \tag{14}$$

where $s = |\boldsymbol{s}|$, $\mu = \pi/s$ is the cosine of the angle between the line-of-sight direction and the pair separation vector $\boldsymbol{s}$. $\xi^s(s,\mu)$ is the two point correlation function and L_ℓ is the ℓth-order Legendre polynomial. These quantities are computed using the publicly available code FCFC (Zhao et al. 2021).

Fig. 9 compares the monopole, $\xi_0(s)$, quadrupole, $\xi_2(s)$, and hexadecapole, $\xi_4(s)$, of the TPCF of 8 independent Uchuu–SDSS light-cone catalogues against the measurements from the SDSS data set, for the set of volume-limited samples described in Table 1. The clustering of our Uchuu–SDSS galaxies is in good agreement with the data for all the volume-limited samples considered, despite the simplicity of our luminosity assignment model – note that, we neglect any potential dependence of our scatter parameter on redshift or luminosity, and consequently our luminosity assignment model has only one tunable parameter. The agreement is poorer for the $^{0.1}M_r^h < -22$ sample, with the Uchuu–SDSS light-cones underestimating the observed monopole. This suggests that the scatter between our halo mass proxy and galaxy luminosity is likely to decrease for bright galaxies $^{0.1}M_r^h \lesssim -22$, which is in agreement with previous findings (Stiskalek et al. 2021).

Our mocks can also be used to diagnose the contaminating effect of fibre collisions on SDSS measurements. In Fig. 10, we show the monopole of the TPCF down to smaller scales, which are dominated by fibre collisions. Our mock catalogues are able to reproduce the clustering effects of fibre collisions in the SDSS data. By comparing

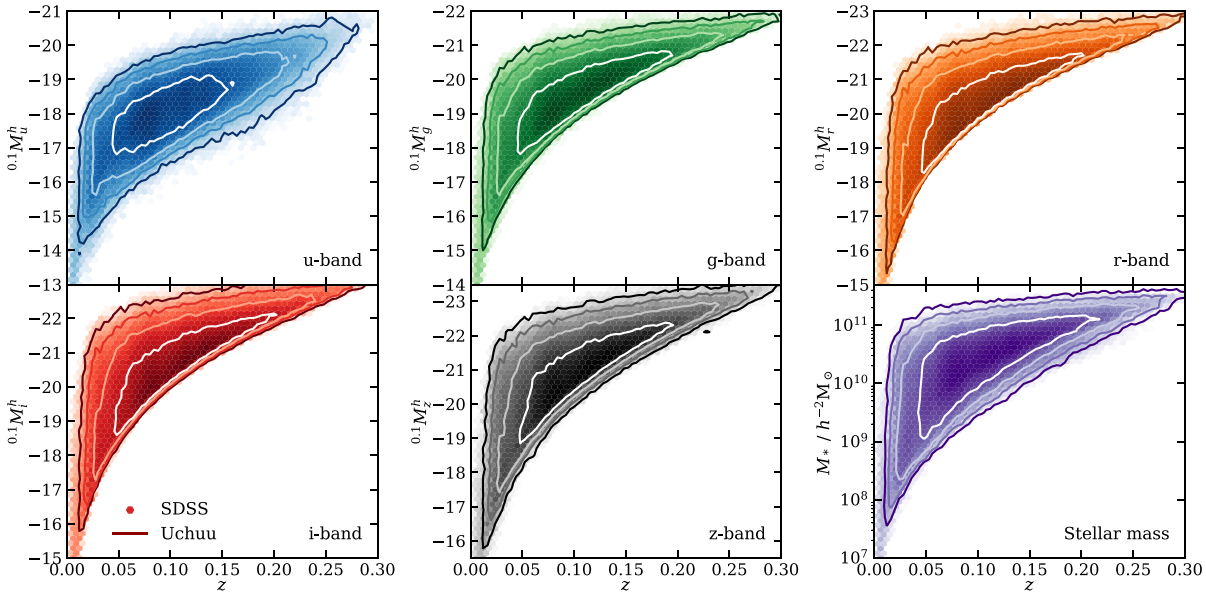

Figure 6. Absolute magnitudes and stellar masses of galaxies in one of the Uchuu–SDSS light-cones (contours), as a function of redshift, compared to the data from SDSS (hexagons), for galaxies with apparent magnitude $r < 17.6$. The different panels show the magnitudes in the *ugriz* bands, as indicated in the legend. The lower right panel shows the MPA stellar masses.

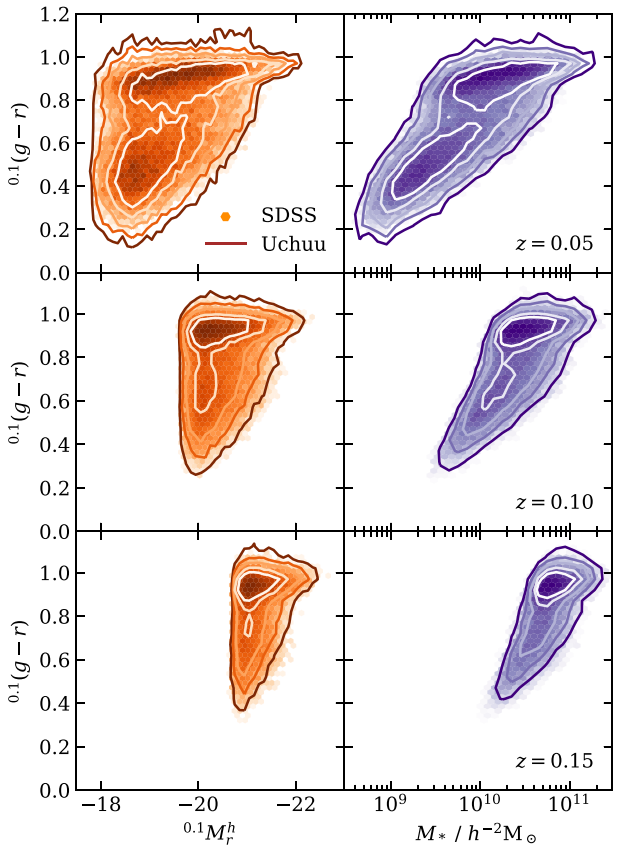

Figure 7. Distribution of $^{0.1}(g-r)$ colours in the Uchuu–SDSS light-cone (contours), compared with the colour distributions of SDSS (hexagons). The left-hand column shows the colour distribution with respect to the *r*-band absolute magnitude, at three different redshifts (from top to bottom, $z = 0.05$, $z = 0.1$, and $z = 0.15$). The right-hand column shows the colours as a function of the MPA stellar mass.

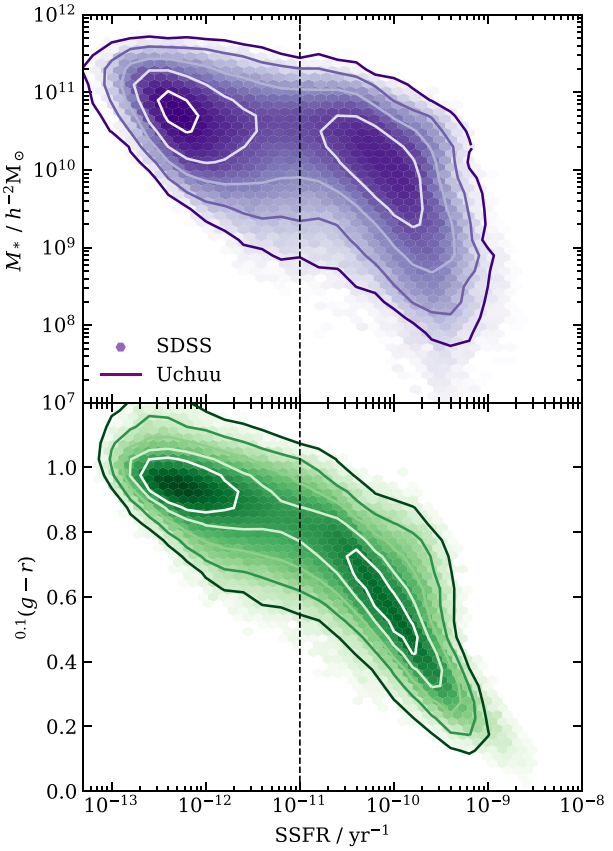

Figure 8. *Top panel*: stellar mass versus sSFR for galaxies in SDSS (hexagons) and in the Uchuu–SDSS light-cone (contours). *Bottom panel*: the $^{0.1}(g-r)$ colour versus sSFR. The vertical dashed line indicates the sSFR cut of 10^{-11} yr^{-1}, which we use to split the bimodal distributions into samples of quiescent and star-forming galaxies.

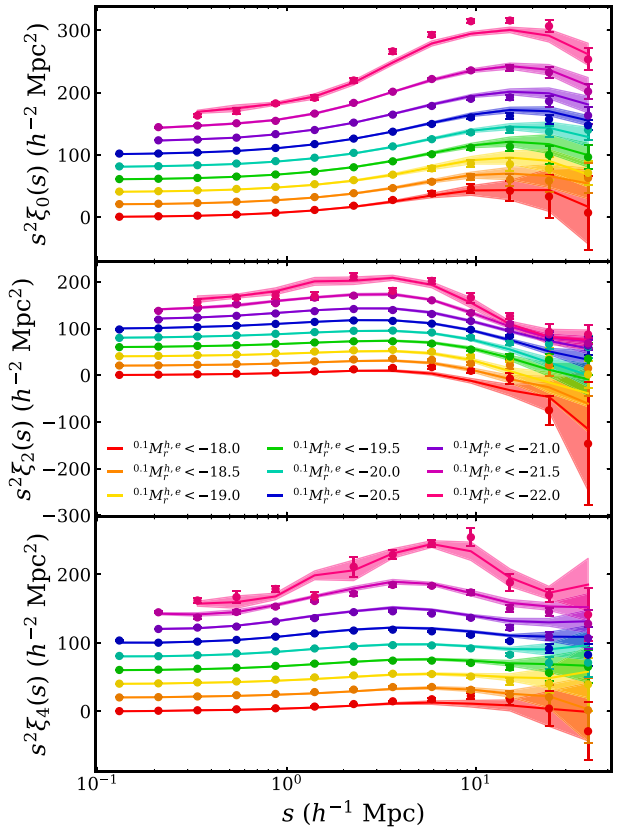

Figure 9. Monopole (top panel), quadrupole (middle panel), and hexadecapole (bottom panel) of the TPCF of Uchuu–SDSS and SDSS galaxies for several volume-limited samples corresponding to luminosity cuts (see Table 1). Solid lines indicate the average of the eight independent Uchuu–SDSS light-cones, while the shaded regions indicate the rms scatter. Data points indicate the clustering measured from the SDSS sample described in Section 2. For clarity, lines corresponding to different luminosity cuts have been offset by successive intervals $20\,/\,h^2\,\mathrm{M}pc^2$, starting from the lowest luminosity sample. Note that, the points that are significantly affected by fibre collision effects have been removed (see Fig. 10).

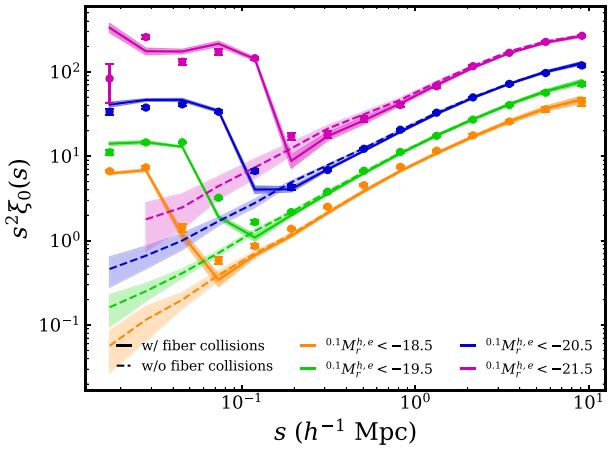

Figure 10. Monopole of the TPCF as in the top panel of Fig. 9, but showing the very small scales dominated by fibre collisions for a subset of the volume-limited samples, with a logarithmic scale. Solid lines correspond to Uchuu–SDSS measurements using fibre-collided redshifts. Dashed lines correspond to measurements from uncollided simulated data. Data points indicate the clustering of the SDSS sample. For clarity, lines corresponding to different luminosity cuts have been offset by successive intervals 0.15 dex, starting from the lowest luminosity sample, and the smallest-s bin for the uncollided $^{0.1}M_r^h < -21.5$ sample has been removed due to high statistical noise. The upturn seen in the clustering signal on the smallest scales is non-physical but an artefact due to the failure of the fiber collision implementation in the SDSS survey.

the measurements from the mock with and without fibre collisions, we are able to discard the SDSS measurements contaminated by fibre collisions in Fig. 9. The regime dominated by fibre collisions corresponds to $s \lesssim 0.1\,h^{-1}$Mpc, although the threshold is larger for brighter samples.

Similarly, Fig. 11 shows the two-point correlation function for several M_* cuts. We find again good agreement between our simulated catalogues and the SDSS data. The poorer agreement for the highest mass cut at $\log_{10}(M_*) > 11$ suggests again a decrease in the scatter in the galaxy–halo connection at the high-mass end, in agreement with previous studies (Behroozi et al. 2019).

In order to study the colour dependence of clustering, we split the sample into two populations of blue and red galaxies. We use a luminosity-dependent colour cut as introduced in Zehavi et al. (2005), (equation 13 in Zehavi et al. 2011).

$$^{0.1}(g-r)_{\rm cut} = 0.21 - 0.03\,{}^{0.1}M_r^{h,e}. \quad (15)$$

Galaxies above the $^{0.1}(g-r)$ cut belong to the red population, while galaxies below the cut belong to the blue population.

Fig. 12 compares the TPCF for blue, red, and all galaxies of Uchuu–SDSS with that of the SDSS data set, for a subset of the volume-limited samples in Table 1. Uchuu–SDSS is in reasonable agreement with the SDSS data, although the agreement is visibly poorer than for the overall volume-limited samples, especially for the brightest samples. This points to limitations due to the simplicity of our colour-assignment algorithm. As described in Section 3.3.1, colours in the light-cone are randomly drawn from our target colour distribution. Our results could be improved, by using information about the halo age to assign colours, thus accounting for assembly bias. This would likely require applying a model involving free parameters, which would need to be finely tuned in order to match the observed colour-dependent clustering.

4.3 Stellar mass function

In order to further validate our simulated galaxy catalogues, we calculate the stellar mass function (SMF) of the SDSS sample (see Section 2.2) and compare it with that from the set of 8 independent Uchuu–SDSS light-cones (see Section 3.3.4). The SMF is estimated by the non-parametric $1/V_{\rm max}$ method widely used in deriving the galaxy luminosity function. To compute the SMF for SDSS and Uchuu samples, we select all galaxies in a redshift range of $0.02 \leq z \leq 0.2$ with stellar masses $M_* \geq 10^9\,h^{-2}\,\mathrm{M}_\odot$ and r-band apparent magnitudes $14.5 \leq r \leq 17.6$.

In Fig. 13 (left panel), we present the SMF obtained from the mean of the 8 Uchuu–SDSS light-cones. Results are compared to SMF derived from the SDSS sample. We also compare our results with that obtained by Moustakas et al. (2013). In Moustakas et al. (2013), M_* was determined utilizing ISEDFIT, a suite of routines used to determine stellar masses, SFRs, and other physical properties of galaxies from the observed broad-band SEDs and redshifts (e.g. Kauffmann et al. 2003; Salim et al. 2007). The Uchuu–SDSS is in

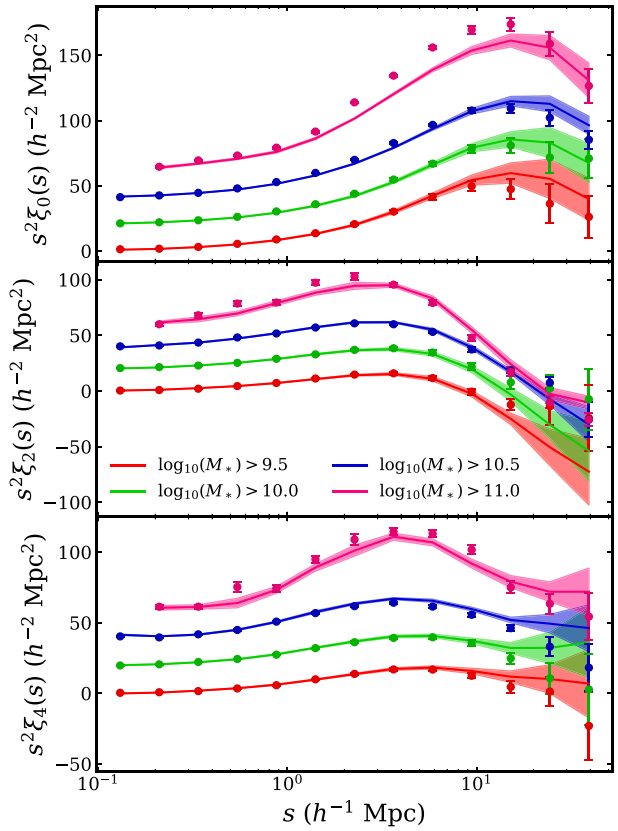

Figure 11. Similar to Fig. 9. Here, the clustering for several M_* cuts is depicted. For clarity, lines corresponding to different cuts have been offset by successive intervals $20\,/\,h^2\,\mathrm{M}pc^2$, starting from the lowest mass sample. Note that, the points that are significantly affected by fibre collision effects have been removed.

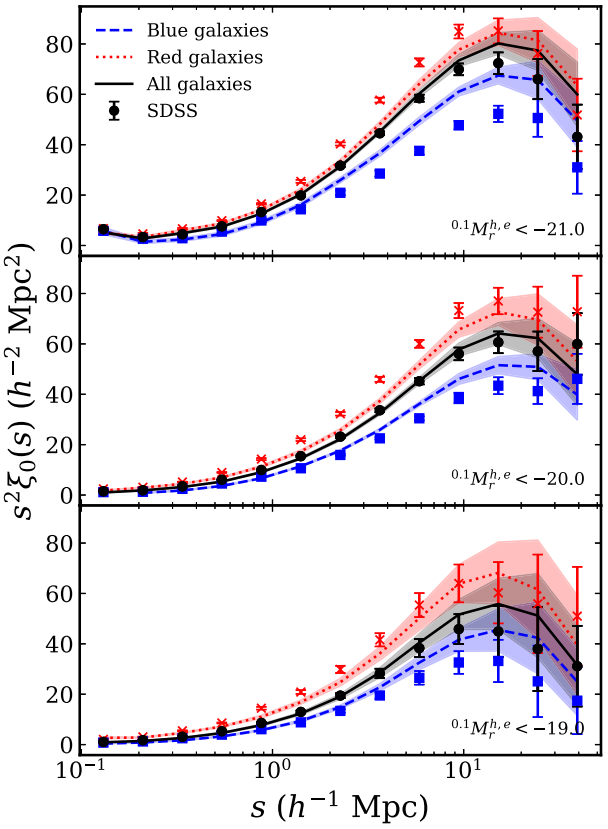

Figure 12. The monopole of the two-point correlation function for SDSS (points with error bars) and Uchuu–SDSS galaxies (solid lines), split into blue (dashed lines, square markers) and red (dotted lines, crosses) galaxies according to equation (15). The dark solid lines indicate the average of the eight independent Uchuu–SDSS light-cones, while the shaded regions show the rms scatter. Each panel corresponds to a volume-limited sample, as indicated in the lower right corner.

reasonably good agreement with both MPA–SDSS and SMF obtained by Moustakas et al. (2013).

The middle and right panels of Fig. 13 show the SMF of quiescent and star-forming galaxies for each data set. We adopt sSFR $= 10^{-11}\,\mathrm{yr}^{-1}$ as the threshold between quiescent and star-forming galaxies. This value corresponds to the minimum between the two peaks of the bimodal sSFR distribution shown in Fig. 8. The Uchuu–SDSS SMFs for quiescent and star-forming galaxies is consistent with those obtained from MPA–SDSS for all masses, although for quiescent galaxies the agreement is slightly poorer compared to the overall sample in the left panel. There is also a good agreement between the Uchuu SMFs and those obtained by Moustakas et al. (2013), with a notable offset at the low- and high-mass ends for quiescent galaxies. One possible reason for this difference is the different methods for the estimation of SFRs used by MPA/JHU and Moustakas et al. (2013), which result in different SFR values.

4.4 Halo occupation distribution

Galaxies are known to be biased tracers of the underlying DM density field. In order to better understand the connection between galaxies and haloes in Uchuu–SDSS catalogues, we investigate the halo occupation distribution. We compute $\langle N_{\rm gal}(>L|M_{\rm halo})\rangle$ – the mean number of galaxies brighter than a given r-band luminosity in a halo with virial mass $M_{\rm halo}$.

The luminosity-dependent HOD is modelled with the functional form described in Zehavi et al. (2011). We write $\langle N_{\rm gal}(>L|M_{\rm halo})\rangle$ as a sum of the mean number of central and satellite galaxies. The mean occupation function of the central galaxies is modelled as a step-like function with a cutoff profile softened to account for the scatter between galaxy luminosity and halo mass, and the mean occupation of satellite galaxies is modelled as a power law modulated by a similar cutoff profile (see Zehavi et al. 2011 for the details). This HOD model has five free parameters: the mass scale, $M_{\rm min}$, and width, $\sigma_{\log M}$, of the central galaxy mean occupation, and the cutoff mass scale, M_0, normalization, M_1', and high-mass slope, α, of the satellite galaxy mean occupation function.

We fit this model to the average HOD obtained from 8 independent Uchuu–SDSS light-cones for the volume-limited samples corresponding to luminosity cuts described in Table 1. The best-fitting HOD parameters are shown in Table 3, along with the SDSS estimates from Zehavi et al. (2011). Fig. 14 shows the mean halo occupation of the Uchuu–SDSS galaxies and the best-fitting HOD models. As seen in the figure, the HOD shifts towards more massive haloes as the luminosity threshold increases – more luminous galaxies occupy more massive haloes.

All our best-fitting parameters, except α, follow an ascending trend as a function of the luminosity threshold. Our results broadly agree with those of Zehavi et al. (2011) over the wide range of brightness thresholds for all volume-limited samples. However, we observe a

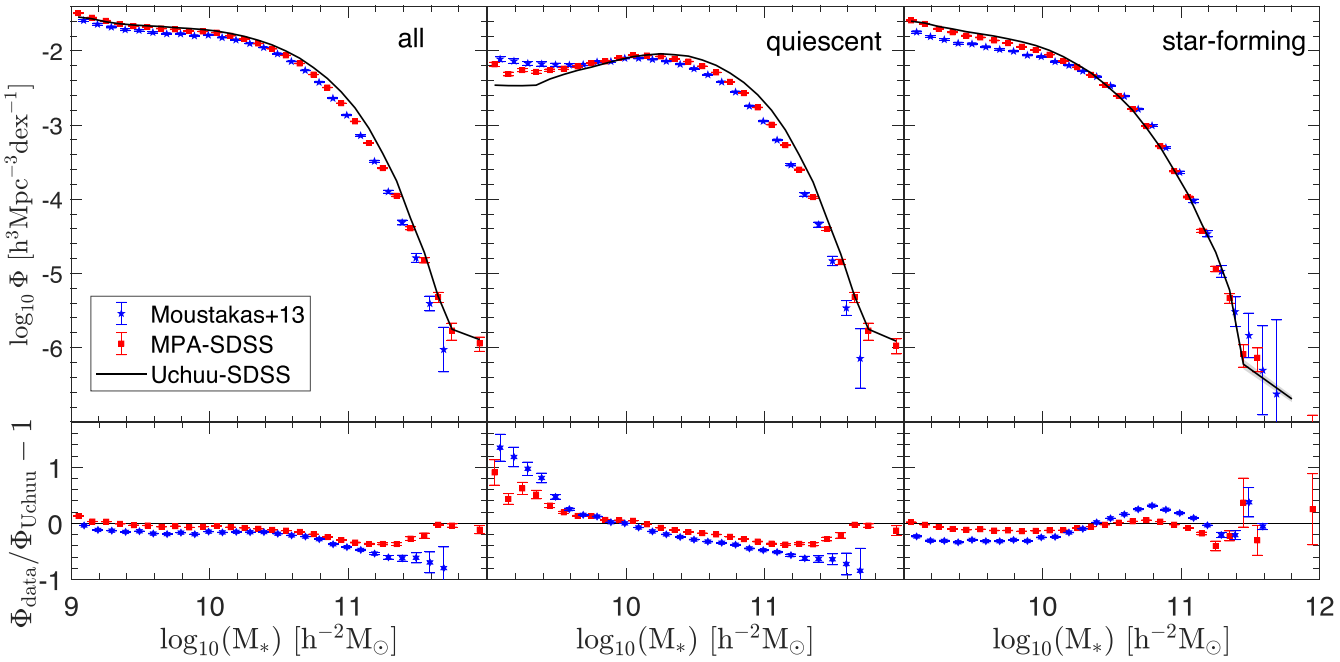

Figure 13. Comparison of the SMF of the Uchuu–SDSS data (the mean of the 8 independent light-cones) with that obtained from the SDSS data with $0.02 \leq z \leq 0.20$, $M_* \geq 10^9\, h^{-2}\,M_\odot$, and $14.5 \leq r \leq 17.6$, in addition to the results from MPA and Moustakas et al. (2013), as shown in the legends. Left, middle, and right panels present the SMF for all, quiescent, and star-forming galaxies, respectively. The lower panels show the relative fraction, $\Phi_{data}/\Phi_{Uchuu} - 1$, of the SMF relative to that from Uchuu–SDSS. The plot shows a reasonable agreement among all three data sets, except for quiescent galaxies at the low- and high-mass ends.

Table 3. Best-fitting HOD parameters for our volume-limited samples. For each luminosity cut the top row indicates the best-fitting parameters for our Uchuu–SDSS light-cones, while the bottom row shows that for the SDSS data as reported in Zehavi et al. (2011). Error bars on our HOD parameters correspond to standard deviation errors on the parameters. Halo masses are in units of $h^{-1}\,M_\odot$.

$^{0.1}M_r^{\max}$	$\log M_{\min}$	$\sigma_{\log M}$	$\log M_0$	$\log M_1'$	α
−18.0	11.34 ± 0.09	0.28 ± 0.20	11.49 ± 0.08	12.57 ± 0.03	0.99 ± 0.03
	11.18	0.19	9.81	12.42	1.04
−18.5	11.46 ± 0.08	0.31 ± 0.16	11.60 ± 0.07	12.65 ± 0.04	0.94 ± 0.03
	11.33	0.26	8.99	12.50	1.02
−19.0	11.62 ± 0.06	0.32 ± 0.12	11.85 ± 0.07	12.79 ± 0.02	0.96 ± 0.02
	11.45	0.19	9.77	12.63	1.02
−19.5	11.80 ± 0.10	0.34 ± 0.16	12.20 ± 0.09	12.90 ± 0.04	0.89 ± 0.03
	11.57	0.17	12.23	12.75	0.99
−20.0	12.05 ± 0.08	0.39 ± 0.11	12.29 ± 0.09	13.16 ± 0.04	0.94 ± 0.04
	11.83	0.25	12.35	12.98	1.00
−20.5	12.37 ± 0.06	0.46 ± 0.06	12.30 ± 0.11	13.51 ± 0.03	0.97 ± 0.04
	12.14	0.17	11.62	13.43	1.15
−21.0	12.81 ± 0.12	0.57 ± 0.09	12.36 ± 0.12	13.81 ± 0.09	0.94 ± 0.14
	12.78	0.68	12.71	13.76	1.15
−21.5	13.37 ± 0.14	0.71 ± 0.09	12.51 ± 0.14	14.18 ± 0.09	0.84 ± 0.16
	13.38	0.69	13.35	14.20	1.09
−22.0	14.03 ± 0.30	0.85 ± 0.22	12.65 ± 0.22	14.72 ± 0.19	0.75 ± 0.30
	14.06	0.71	13.72	14.80	1.35

few discrepancies for some HOD parameters (see Table 3). This is likely due to the difference in the methodologies. While we have computed the halo occupation directly from our values of $^{0.1}M_r^h$ and $M_{\rm halo}$ in the Uchuu–SDSS light-cones, Zehavi et al. (2011) obtains it by fitting the projected correlation function of the observed SDSS data.

5 RSD AND BAO MEASUREMENTS

In this section, we study the BAO signal in the SDSS MGS. For this purpose, we define a BAO sample (SDSSbao, see Section 2.1) in which this signal is enhanced similarly to Ross et al. (2015). We also model the full shape of the TPCF to measure $f\sigma_8$ and obtain the anisotropic BAO distances. While we have the set of 32 Uchuu–SDSS light-cones to compare with the SDSSbao data, in order to further improve our covariance errors on BAO scales, we generate an additional sample of 5100 light-cones, describe below, using a set of lower resolution N-body simulations run with the GLAM code (Klypin & Prada 2018).

In Section 5.1, we describe the method used to construct our 5100 GLAM-SDSSbao light-cones, while their clustering properties

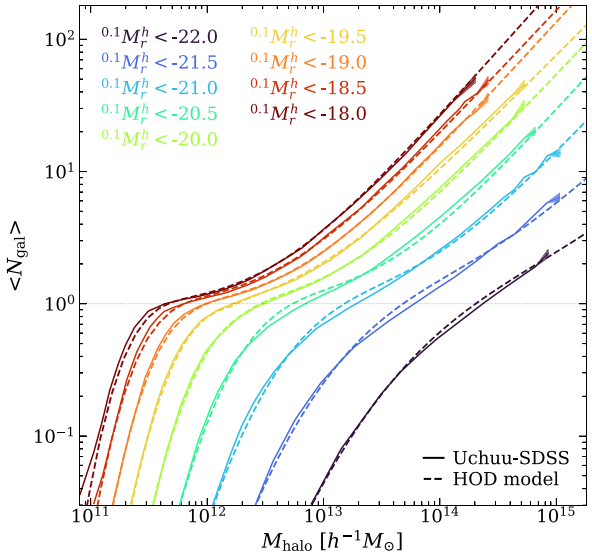

Figure 14. Measured mean halo occupation for the different luminosity threshold samples drawn from 8 independent Uchuu–SDSS light-cones as a function of halo mass. The dashed lines represent the best-fitting models for the halo occupation functions. Best-fitting HOD parameters are shown in Table 3 together with those obtained by Zehavi et al. (2011). The maximum host halo mass in each sample is related to their effective volume (see Table 1).

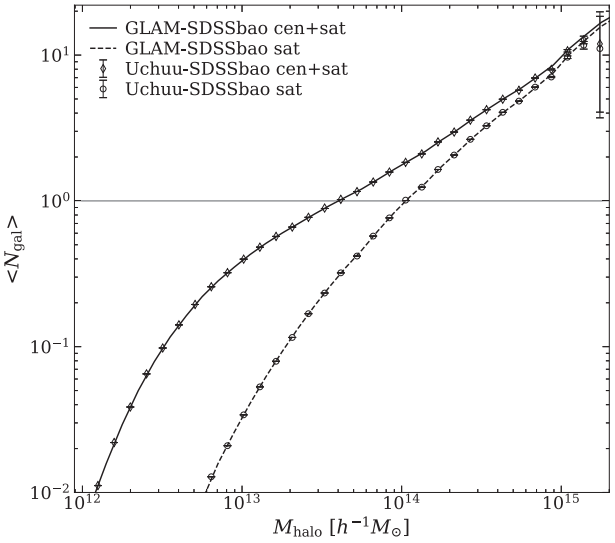

Figure 15. Mean halo occupation distribution from our Uchuu– and GLAM–SDSSbao catalogues. Points represents the values for Uchuu-SDSSbao, diamond for all galaxies, circles for satellite galaxies. Lines refer to the GLAM–SDSSbao results, solid lines for all galaxies, dashed lines for satellite galaxies.

are explored in Section 5.2. Our RSD measurements are presented in Section 5.3. Finally, we measure the isotropic BAO scale in Section 5.4.

5.1 Constructing the GLAM light-cones

We generate 1275 GLAM simulations using the same cosmology and linear power spectrum as the Uchuu simulation. The GLAM simulations follow the evolution of 2000^3 particles of mass $1.07 \times 10^{10}\,h^{-1}\,\mathrm{M}_\odot$ in a cubic box of size 1 h^{-1} Gpc with $N_s = 140$ time-steps, and mesh of $N_g = 5800$. This numerical setup yields a spatial resolution of $\Delta x = 0.172\,h^{-1}$ Mpc. The initial conditions are generated using the Zeldovich approximation starting at $z_{\rm ini} = 105$. The distinct haloes in GLAM are identified with the Bound Density Maximum halo finder (Klypin & Holtzman 1997).

Since the GLAM simulations are unable to resolve substructure inside distinct haloes, the SHAM method introduced in Section 3 cannot be applied, and we thus resort to a statistical HOD method. First, we compute the HOD of the SDSSbao sample by applying the galaxy selection criteria in equation (4) to our 8 independent Uchuu–SDSS light-cones (hereafter, Uchuu–SDSSbao). We then use the 1275 GLAM halo catalogues available at the mean redshift of the BAO sample ($z \sim 0.1$) to generate a galaxy catalogue for each GLAM box by randomly drawing galaxies from the measured halo occupation statistics for each distinct halo. Fig. 15 shows the mean HOD from Uchuu–SDSSbao used to populate with galaxies in the GLAM simulations, along with the resulting mean HOD obtained from the 1275 GLAM galaxy catalogues. By construction, the HOD of Uchuu- and GLAM–SDSSbao galaxies are in agreement. It is important to note that the GLAM simulations are only able to resolve haloes larger than $10^{12}\,h^{-1}\,\mathrm{M}_\odot$. However, the HOD obtained from the high-fidelity Uchuu–SDSSbao light-cones does not extend to masses below this limit. The resulting GLAM galaxy catalogues have an average density of $6 \times 10^{-4}\ h^3\,\mathrm{M}^{-1}\ pc^3$, which corresponds to the average $n(z)$ of the Uchuu–SDSSbao sample.

Once the GLAM haloes for a given box are populated with galaxies, we adjust its number density to match that in the SDSSbao sample. For this step, we use the $n(z)$ presented in Ross et al. (2015) as a reference. Then, we cut it to the northern contiguous region of the SDSS survey footprint. As with the Uchuu light-cones, by replicating the SDSS footprint across the full sky, we can generate a total of 4 independent SDSSbao light-cones from each **GLAM** box, which allows us to create a total of 5100 GLAM–SDSSbao light-cones. We decided not to apply the fibre collision correction to GLAM-SDSSbao since it effect is negligible on the scales we are going to make use of the light-cones (BAO scales).

The mean number density of the GLAM–SDSSbao light-cones is shown in Fig. 16, together with that from Uchuu–SDSSbao and the SDSSbao data. There is a small 20 per cent excess of galaxies in Uchuu–SDSSbao as compared to the data at the high-redshift tail of the distribution. As explained in Section 4.1, this is because in order to build the Uchuu catalogues, we use a target luminosity function that transitions from SDSS to GAMA at $z \sim 0.15$, since the luminosity function from SDSS is poorly constrained at high redshifts. Despite this, the agreement with the SDSS clustering is reasonable as discussed in Section 4.2. A slight deficit of galaxies can also be seen in the GLAM–SDSSbao galaxy number density. The lower resolution of GLAM ($M_{\rm halo} \sim 10^{12}\,h^{-1}\,\mathrm{M}_\odot$) leads to a smaller number of haloes, which is not sufficient to achieve the SDSSbao observed number density at $z \sim 0.15$. Despite the difference in number density, the difference between the clustering in Uchuu– and GLAM–SDSSbao is small (see Fig. 17).

5.2 Two-point correlation function and covariance matrix

In this section, we explore the clustering on the BAO scales in our real and mock data sets.

In order to optimize the BAO clustering signal-to-noise ratio, we weight both galaxies and randoms depending on the galaxy number

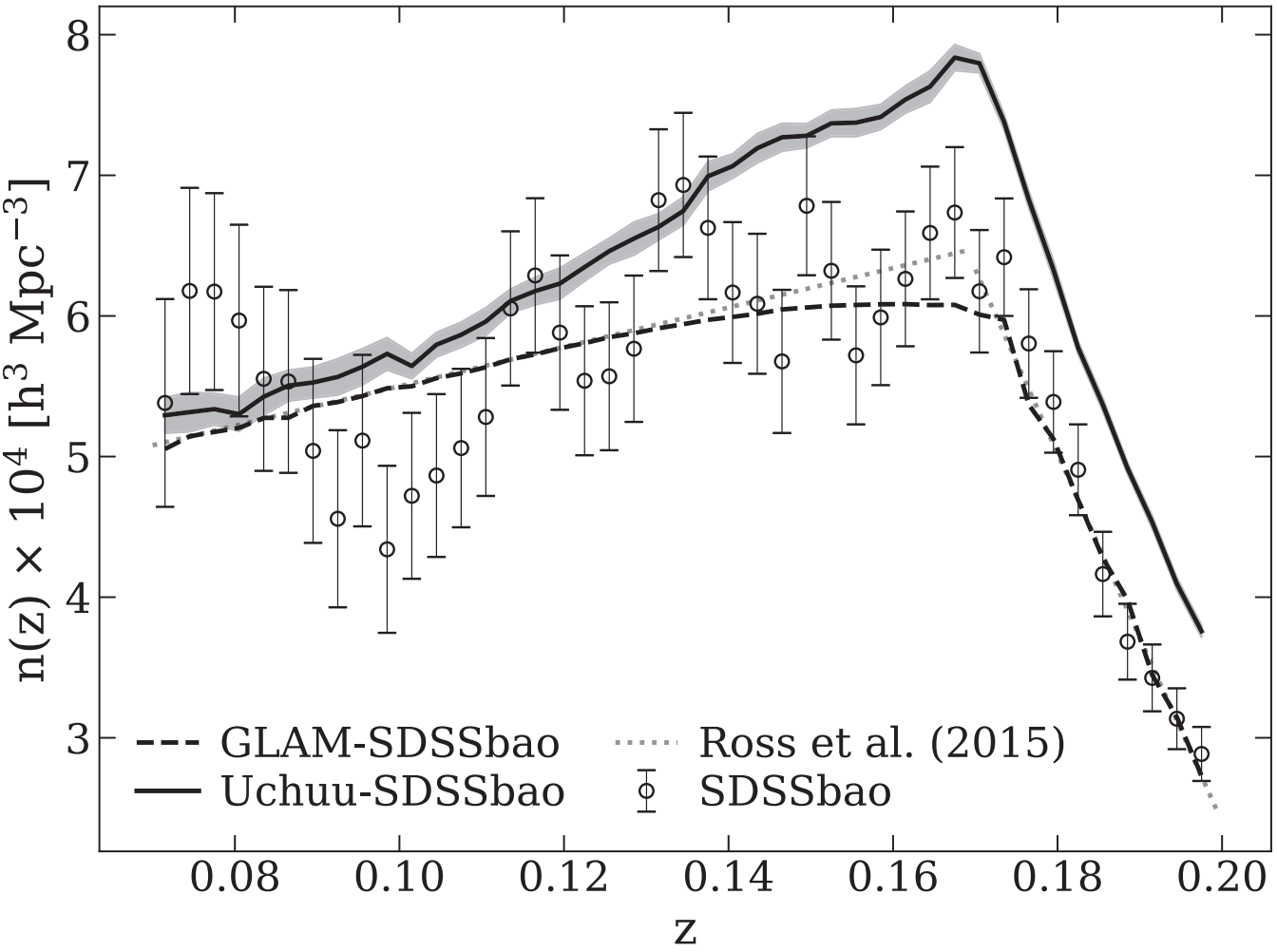

Figure 16. Number density of the SDSSbao data, Uchuu– and GLAM–SDSSbao light-cones. The dashed line represents the mean $n(z)$ of the 5100 GLAM–SDSSbao, while the solid line represents that from the 32 Uchuu–SDSSbao. The standard deviation for each of the samples is shown as a shaded area. The SDSSbao measurments, represented with black dots, and the error bars are estimated from the standard deviation of the GLAM–SDSSbao catalogues. The grey dotted lines shows the fit to SDSSbao data given in Ross et al. (2015). The deviation between GLAM and Ross et al. (2015) at $z \sim 0.15$ arises from the incompleteness of the GLAM haloes due to the lower resolution of the simulations.

density $n(z)$ using FKP weights (Feldman, Kaiser & Peacock 1994; Ross et al. 2015), i.e.

$$w_{\rm FKP} = \frac{1}{1 + P_{\rm FKP} n(z)}. \tag{16}$$

We set $P_{\rm FKP} = 16\,000\ h^{-3}\,{\rm Mpc}^3$, which is close to the measured amplitude at $k = 0.1\,h{\rm M}^{-1}{\rm pc}$.

Further, from the 5100 GLAM–SDSSbao light-cones, we infer the covariance matrix, **C**, defined as

$$C^{\ell_i \ell_j}(s_i, s_j) = \frac{1}{N-1} \sum_{n=1}^{N} [\boldsymbol{\xi}^{(\ell_i)}(s_i) - \bar{\boldsymbol{\xi}}^{(\ell_i)}(s_i)][\boldsymbol{\xi}^{(\ell_j)}(s_j) - \bar{\boldsymbol{\xi}}^{(\ell_j)}(s_j)], \tag{17}$$

where N is the number of light-cones.

Fig. 17 shows the monopole, quadrupole, and hexadecapole of the two-point correlation function of the SDSSbao data, the mean of the 32 Uchuu–SDSSbao light-cones (described in Section 3.3), and the mean of the 5100 GLAM–SDSSbao. We include all bins between 25 and 200 h^{-1} Mpc in 5 h^{-1} Mpc steps. The error bars in the SDSSbao TPCF are calculated from the diagonal elements of the GLAM–SDSSbao covariance matrix. For Uchuu and GLAM, we plot the error in the mean of the correlation functions as shaded regions.

The three data sets show a clear BAO peak at $s \sim 100\,h^{-1}$Mpc. The results from Uchuu– and GLAM–SDSSbao are in good agreement with the SDSSbao observational measurements. The difference between Uchuu–SDSSbao and SDSSbao clustering is of $\chi^2/{\rm dof} = 1.13$, while for Uchuu–SDSSbao and SDSSbao we obtain $\chi^2/{\rm dof} = 1.07$. The GLAM measurements have better statistics than Uchuu, due to the huge number of light-cones in the GLAM suite. Also, Uchuu has a larger number density than GLAM, which translates into a lower clustering amplitude. Note that, the GLAM light-cones, used to estimate uncertainties from its TPCF covariance matrix, match the SDSSbao number density (see Fig. 16).

5.3 RSD measurements

Galaxy clustering analyses that rely on two-point statistics are not sensitive to the growth rate of structure f directly but instead to $f\sigma_8$, where σ_8 is the normalization of the linear power spectrum on a scale of 8 h^{-1} Mpc.

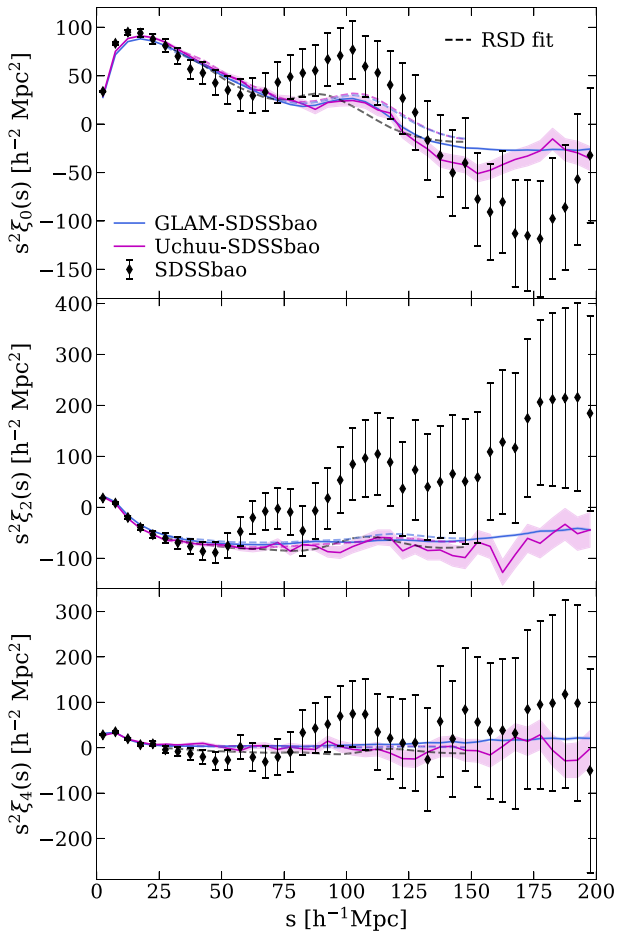

Figure 17. The monopole, quadrupole, and hexadecapole of the two-point correlation function measured in our three SDSS BAO data set: SDSSbao (filled symbols), GLAM-SDSSbao (blue lines) and Uchuu–SDSSbao (pink lines). Errors have been estimated from the covariance matrix of the 5100 GLAM-SDSSbao light-cones. For Uchuu (GLAM) the errors correspond to the error on the mean of the 32 (5100) mocks. Following the same colour code, we also plot the best-fitting RSD model (dashed lines) for each data set, (see Section 5.3).

We measure the linear growth rate of structure, $f\sigma_8$, and the Alcock–Paczynski parameters, $\alpha_\perp$ and $\alpha_\parallel$, of the Uchuu–SDSS light-cones and SDSSbao sample described above using the two-point correlation function, $\xi(s, \mu)$.

The anisotropic Alcock–Paczynski parameters are defined as

$$\alpha_\parallel = \frac{D_{\rm H}(z) r_d^{\rm fid}}{D_{\rm H}^{\rm fid}(z) r_d} \tag{18}$$

and

$$\alpha_\perp = \frac{D_{\rm M}(z) r_d^{\rm fid}}{D_{\rm M}^{\rm fid}(z) r_d}, \tag{19}$$

where $H(z)$ is the Hubble parameter, $D_{\rm M}(z)$ is the angular diameter distance, $D_{\rm H}(z) = c/H(z)$ is the Hubble distance, and r_d is the sound horizon at the drag epoch. Quantities with a 'fid' superscript are calculated for the fiducial cosmology assumed during the analysis, while the quantities without a superscript exist in the true cosmology.

The analysis is performed using the Planck fiducial cosmology, as adopted for Uchuu, to convert the redshifts to comoving distances. If the assumed fiducial cosmology does not match the true cosmology, there is a scaling of the BAO peak position parallel and perpendicular to the line-of-sight (as given in equations (18), 19). Thus, we should recover $\alpha_\perp = \alpha_\parallel = 1$ from the Uchuu– and GLAM–SDSSbao light-cones.

From the covariance matrix $\mathbf{C}$ computed in Section 5.2, we can define the χ^2 statistic as

$$\chi^2 = \frac{N - p - 2}{N - 1} \left(\boldsymbol{\xi}^{\rm Data} - \boldsymbol{\xi}^{\rm Model}\right) C^{-1} \left(\boldsymbol{\xi}^{\rm Data} - \boldsymbol{\xi}^{\rm Model}\right)^T, \tag{20}$$

where p is the number of degrees-of-freedom being fitted and N is the number of GLAM-SDSSbao light-cones, and we have included the Hartlap correction (Hartlap, Simon & Schneider 2007).

Our theoretical model for the TPCF, $\xi^{\rm Model}$, is based on Lagrangian Perturbation Theory (LPT). To model the distribution of galaxies, one also needs to introduce a bias model that connects the matter density, $\delta_{\rm m}$, and the galaxy density, $\delta_{\rm g}$, with two parameters b_1 and b_2 named the linear and quadratic Lagrangian bias, respectively. We can then obtain the model power spectrum, and include the effect of peculiar velocities to account for the RSD effect (Seljak & McDonald 2011).

We also model the Fingers-of-God (FOG) effect in Fourier space, using the phenomenological Lorentz model (Taruya, Nishimichi & Saito 2010):

$$P_{\rm FOG}(\boldsymbol{k}) = \frac{1}{1 + (\boldsymbol{k} \cdot \hat{\boldsymbol{n}}\, \sigma_{\rm FOG})^2/2} P(\boldsymbol{k}). \tag{21}$$

where $P(\boldsymbol{k})$ is the non-linear power spectrum without the FOG effect, $\sigma_{\rm FOG}$ is the one-dimensional velocity dispersion and $\hat{\boldsymbol{n}}$ is the normalized LOS direction vector.

The theoretical power spectrum $P_{\rm FOG}$ is obtained using the `MomentumExpansion` module of VELOCILEPTORS package (for more details, see Chen, Vlah & White 2020; Chen et al. 2021). Finally, we obtain the correlation function by taking the Fourier transform of $P_{\rm FOG}$,

$$\xi(\boldsymbol{x}) = \int d^3\boldsymbol{k} e^{i\boldsymbol{k}\cdot\boldsymbol{x}} P_{\rm FOG}(\boldsymbol{k}). \tag{22}$$

We fit our RSD model to the correlation function multipoles from the three data sets SDSSbao, Uchuu–SDSSbao, and GLAM-SDSSbao in the separation range [25, 145] h^{-1} Mpc, with bins of width 5 h^{-1} Mpc. In addition to the cosmological parameters $f\sigma_8$, $\alpha_\parallel$ and $\alpha_\perp$, we also estimate the Lagrangian biases b_1, b_2 and the Fingers-of-God parameter $\sigma_{\rm FOG}$. Unphysical values of parameters are avoided by setting the priors to $f\sigma_8 > 0$, $b_1 > -1$, and $\sigma_{\rm FOG} > 0$. The first Lagrangian bias is related to the Eulerian bias by $b_{1,\rm Eulerian} = 1 + b_1$. We assume the effective redshift of the SDSSbao sample to be $z = 0.15$.

The correlation function multipoles corresponding to our best-fit models can be seen in Fig. 17. We observe good agreement with Uchuu–SDSSbao and GLAM–SDSSbao measurements. The RSD model predicts the position of the BAO peak in the monopole that is too low as compared with the observed position of the peak. However, this disagreement is within the noise limits.

The minimum of χ^2 is found using IMINUIT (Dembinski et al. 2020), which achieves convergence near the minimum using the first and approximate second derivatives. Errors are estimated from the region of $\Delta\chi^2 = 1$ of the marginalized χ^2 distribution, and they are allowed to be asymmetric. We also run Monte Carlo Markov chains (MCMC) with the EMCEE package (Foreman-Mackey et al. 2013) in order to compute the likelihood surface of our set of fitted parameters. Their convergence is checked with the Gelman–Rubin convergence test (Gelman & Rubin 1992; Brooks & Gelman 1998).

We first test our RSD pipeline on the GLAM–SDSSbao and Uchuu–SDSSbao light-cones. The best-fitting results of the parameters $f\sigma_8$,

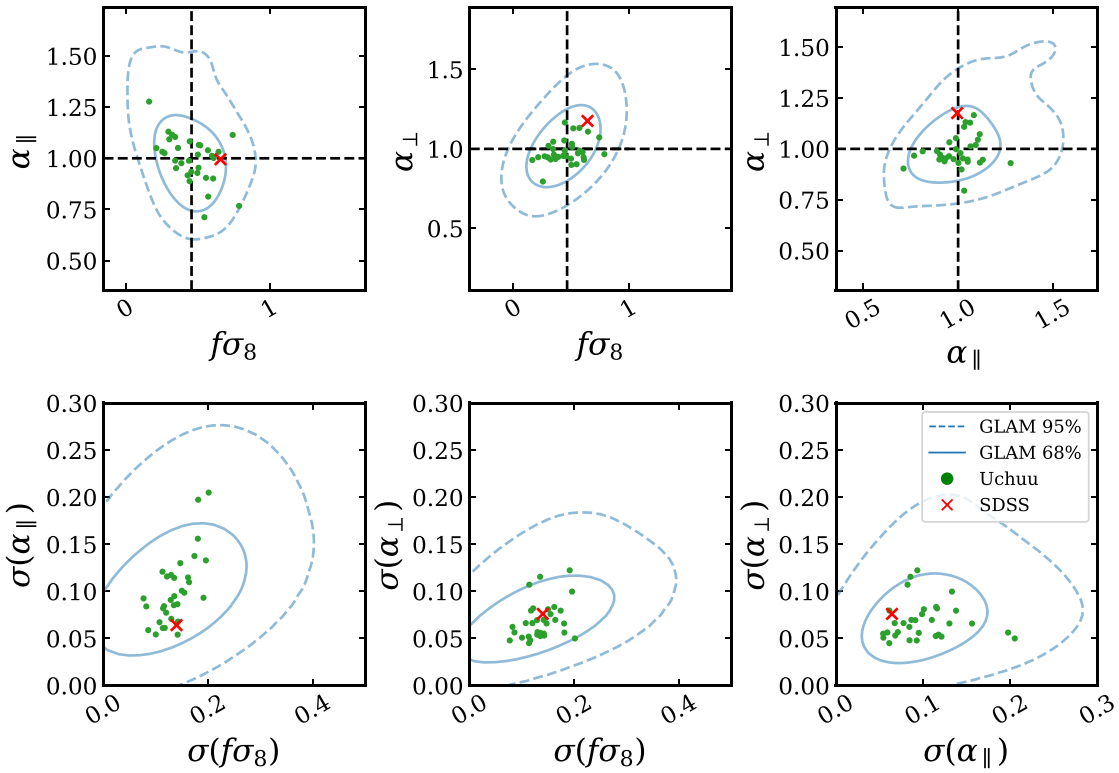

Figure 18. *Top row*: Parameter values of $f\sigma_8$, $\alpha_\parallel$, and $\alpha_\perp$ measured from the GLAM-SDSSbao (blue contours), Uchuu–SDSSbao (green points), and SDSSbao light-cones (red cross). Dashed straight black lines represent the expected values for the fiducial cosmology. *Bottom row*: The corresponding parameter errors.

Table 4. RSD fitted parameters from the SDSSbao data, obtained by χ^2 minimization and using Bayesian MCMC inference. The first two columns correspond to the results from this paper, while the last column shows the values from Howlett et al. (2015) for comparison. Only $f\sigma_8$ and $b_{1,\mathrm{Eulerian}}$ estimated values from Howlett et al. (2015) are found in the literature.

Parameter	χ^2 minimization	MCMC	Reference
$f\sigma_8$	$0.65^{+0.14}_{-0.15}$	$0.60^{+0.15}_{-0.16}$	$0.63^{+0.24}_{-0.27}$
$\alpha_\parallel$	$1.00^{+0.08}_{-0.06}$	$1.04^{+0.14}_{-0.09}$	N/A
$\alpha_\perp$	$1.18^{+0.07}_{-0.09}$	$1.16^{+0.08}_{-0.09}$	N/A
$b_{1,\mathrm{Eulerian}}$	$1.53^{+0.19}_{-0.19}$	$1.59^{+0.22}_{-0.20}$	$1.36^{+0.29}_{-0.26}$
b_2	$-1.8^{+2.7}_{-1.1}$	$-0.8^{+1.7}_{-1.5}$	N/A
$\sigma_{\mathrm{FOG}}[h^{-1}\,\mathrm{Mpc}]$	$0^{+7.4}_{-0.0}$	$4.0^{+2.8}_{-2.7}$	N/A

$\alpha_\parallel$, and $\alpha_\perp$ are summarized in Fig. 18. We confirm that the theoretical model recovers the fiducial Planck values within 1σ.

We then apply the pipeline to the SDSSbao sample. Our results are listed in Table 4. Both χ^2 minimization and MCMC sampling methods provide consistent results, and the small difference in the errors and parameter values are attributed to better treatment of the Fingers-of-God effect with MCMC chains. The parameter distributions for σ_{FOG} are non-Gaussian, as it is restricted to positive values, but the best-fitting value is consistent with 0. Additionally, we compare our obtained $f\sigma_8$ and b_1 with Howlett et al. (2015), finding a good agreement, but we obtain a $\gtrsim 30$ per cent increase in precision on $f\sigma_8$ which can be attributed to our better estimate of the covariance matrix (see Table 4). The results obtained on the mock data are shown in Table 5.

The distribution of parameter values and their uncertainties can be seen in Fig. 18, where the results from the SDSSbao sample together with those from GLAM and Uchuu are shown in blue, orange, and red respectively, and the black lines show the expected values of the parameters for the Planck fiducial cosmology. The results from GLAM– and Uchuu–SDSSbao are consistent with each other and with the SDSSbao data, meaning that the mocks are a fair statistical representation of the data. We confirm that the fits we have performed are valid and have an acceptable χ^2/dof, for GLAM mocks being $\chi^2/\mathrm{dof} = 1.00 \pm 0.18$, for Uchuu mocks being $\chi^2/\mathrm{dof} = 1.07 \pm 0.18$ and for the SDSSbao sample being $\chi^2/\mathrm{dof} = 0.98$. It should be noted, however, that due to a very high number of GLAM mocks, the results from the fits of the mean were expected to be slightly biased due to the precision becoming larger than the systematic uncertainty from our model, which is designed for the analysis of only one realization.

Finally we use our best-fit values of the BAO-inferred $\alpha_\parallel$ and $\alpha_\perp$ parameters to provide a measurement of the Hubble distance and the (comoving) angular diameter distance at the effective redshift of the SDSSbao sample, i.e. $D_{\mathrm{H}}(z = 0.15)/r_d = 27.9^{+2.2}_{-1.8}$, $D_{\mathrm{M}}(z = 0.15)/r_d = 5.1^{+0.3}_{-0.4}$. These results are valuable since they constitute the first two-dimensional distance measurements from SDSS data.

5.4 Isotropic BAO measurements

We measure the isotropic BAO scale from our Uchuu–SDSSbao and the SDSSbao data at $z = 0.15$. The BAO signal can be seen in the correlation function monopole measured from both data sets (see Fig. 17). The BAO scale can be extracted by fitting the monopole of the correlation function to a template that includes the dilation

Table 5. RSD fitted cosmological parameters from the means of Uchuu and GLAM correlation functions, obtained using χ^2 minimization in comparison with the values predicted by the fiducial cosmology.

Parameter	Uchuu	GLAM	Expected value
$f\sigma_8$	$0.44^{+0.02}_{-0.02}$	$0.420^{+0.002}_{-0.002}$	0.46
$\alpha_\parallel$	$0.98^{+0.03}_{-0.03}$	$0.991^{+0.002}_{-0.002}$	1.00
$\alpha_\perp$	$0.98^{+0.02}_{-0.01}$	$0.985^{+0.001}_{-0.001}$	1.00
$b_{1,\mathrm{Eulerian}}$	$1.44^{+0.03}_{-0.03}$	$1.430^{+0.003}_{-0.003}$	N/A
b_2	$-0.8^{+0.2}_{-0.2}$	$-0.641^{+0.03}_{-0.03}$	N/A
$\sigma_{\mathrm{FOG}}[h^{-1}\,\mathrm{Mpc}]$	$0.0^{+1.5}_{-0.0}$	$3.73^{+0.08}_{-0.08}$	N/A

parameter, α,

$$\alpha \equiv \frac{D_{\mathrm{V}}(z) r_{\mathrm{d}}^{\mathrm{fid}}}{D_{\mathrm{V}}^{\mathrm{fid}}(z) r_{\mathrm{d}}}\,, \tag{23}$$

where

$$D_{\mathrm{V}}(z) = \left[cz(1+z)^2 D_{\mathrm{A}}^2(z) H^{-1}(z)\right]^{1/3}\,, \tag{24}$$

is the spherically average distance (Eisenstein et al. 2005); α is related to the anisotropic Alcock–Paczynski parameters $\alpha_\perp$ and $\alpha_\parallel$ given in equations (18) and (19), by $\alpha = \alpha_\parallel^{1/3}\alpha_\perp^{2/3}$.

To obtain the best-fitting α value, we use Bayesian statistics and maximize the likelihood by adopting the BAO model of the monopole of the redshift-space correlation function as presented in Ross et al. (2015). We perform the fit to the simulations and data monopoles on scales $50 < s < 150\ h^{-1}\,\mathrm{Mpc}$, using separation bins of width $\Delta s = 5h^{-1}\,\mathrm{Mpc}$.

The measured and the best-fitting model of the correlation function monopole of the SDSSbao data and the simulated Uchuu and GLAM light-cones are shown in Fig. 19. We find $\alpha = 0.9936 \pm 0.0137$, $\alpha = 1.0056 \pm 0.0008$, and $\alpha = 1.0204 \pm 0.0455$ from Uchuu, GLAM, and the SDSS observations. The precision for the SDSS measurement remains similar if we apply a reconstruction procedure (Eisenstein et al. 2007; White 2015) that partly eliminates the degradation of the BAO signal due to non-linear evolution and RSDs. Our results are in agreement with previous measurements reported by Ross et al. (2015).

From our dilation parameter estimate, we measure a spherically averaged distance $D_{\mathrm{V}}(z = 0.15)/r_{\mathrm{d}} = 4.378 \pm 0.195$. The values of D_{V} from the Uchuu and GLAM light-cone measurements are $D_{\mathrm{V}}(z = 0.15)/r_{\mathrm{d}} = 4.263 \pm 0.059$ and $D_{\mathrm{V}}(z = 0.15)/r_{\mathrm{d}} = 4.314 \pm 0.003$, respectively.

6 SUMMARY

The cosmological interpretation of large galaxy surveys requires the generation of high-fidelity simulated galaxy data. Here, we describe and analyse publicly available Uchuu–SDSS light-cones: a set of simulated SDSS catalogues generated using the Uchuu N-body simulation. Uchuu is a large high-resolution cosmological simulation that follows the evolution of the DM across cosmic time in the Planck cosmology. The Uchuu–SDSS catalogues are tailored to reproduce the sky footprint and galaxy properties of the SDSS MGS observational sample. This facilitates the direct comparison between our simulated light-cones and the observational data to probe the Planck-ΛCDM cosmology model using the large-scale clustering signal.

The rest-frame r-band magnitudes, ${}^{0.1}M_r^h$, are assigned to the Uchuu haloes using the SHAM method, with a simple recipe for the scatter in the galaxy–halo connection. By construction, our scheme reproduces the SDSS and GAMA luminosity functions, and it reproduces to good accuracy the SDSS clustering (see Fig. 1) while using only one free parameter – the scatter. The resulting galaxy catalogues computed at different redshifts are combined in spherical shells and cut to the SDSS sky footprint. By placing observers at different positions in the cubic box, we produce a set of 8 independent Uchuu–SDSS light-cones, with only a small mutual overlap in volume. Additionally, an extended set of 32 light-cones, which overlap for $z > 0.175$, are generated to increase the statistics for our SDSS RSD and BAO measurements. Galaxy colours are also produced using a Monte Carlo method that randomly draws colours from the ${}^{0.1}M_{\mathrm{r}}^h$ and redshift-dependent $g - r$ distribution obtained from the GAMA survey. Redshifts, r-band magnitudes and $g - r$ colours are used to match each simulated galaxy to a galaxy in the SDSS sample. This allows us to assign absolute magnitudes, apparent magnitudes and k-corrections in the u, g, i, and z bands in addition to stellar masses and specific star formation rates. Finally, we implement the effect of fibre collisions by applying a nearest neighbour correction to a set of galaxies situated in close angular proximity.

Our Uchuu–SDSS light-cones are able to recover the galaxy redshift distribution (see Fig. 5) and various galaxy properties from the SDSS survey to very high accuracy (see Figs 6, 7, and 8). We also compute the galaxy correlation functions for several volume-limited samples corresponding to a wide range of luminosities (Fig. 9) and stellar mass cuts (Fig. 11), finding a very good agreement with the SDSS clustering down to the smallest $100\,h^{-1}\,\mathrm{kpc}$ scale. The colour- and stellar-mass-dependent galaxy clustering (see Fig. 12) is in general agreement with the SDSS results. Similarly, the simulated stellar mass function (Fig. 13) is in good agreement with that from SDSS data. We also provide the halo occupation distributions and parameters of the Uchuu–SDSS galaxies (Fig. 14), which are in agreement with previous SDSS analyses.

We explore the RSD and BAO signal in our Uchuu–SDSS light-cones and the SDSS data (see Fig. 17). In order to obtain high-precision covariance matrices for the error estimates of the RSD and BAO measurements, we create a large number of GLAM simulations. We apply a HOD method to populate the GLAM haloes with galaxies, generating a total of 5100 GLAM-SDSSbao light-cones. We measure $f\sigma_8$ and the anisotropic BAO parameters, $\alpha_\perp$ and $\alpha_\parallel$, from a full-shape model fit of the TPCF, finding a very good agreement between the SDSSbao data and our GLAM- and high-fidelity Uchuu–SDSSbao lighcones (see Fig. 18). Our results are given in Table 4. We obtain a $\gtrsim$ 30 per cent increase in precision on $f\sigma_8$, as compared to the previous measurement by (Howlett et al. 2015), which can be attributed to our better estimate of the covariance matrix.

We use our best-fitting values of the BAO-inferred $\alpha_\parallel$ and $\alpha_\perp$ parameters to provide a measurement of the Hubble distance and the (comoving) angular diameter distance at the effective redshift of the SDSSbao sample, i.e. $D_{\mathrm{H}}(z = 0.15)/r_{\mathrm{d}} = 27.9^{+3.1}_{-2.7}$, $D_{\mathrm{M}}(z = 0.15)/r_{\mathrm{d}} = 5.1^{+0.4}_{-0.4}$. We highlight that these distance results are valuable since at present only the spherically averaged distance $D_{\mathrm{V}}(z = 0.15)/r_{\mathrm{d}}$ has been reported by Ross et al. (2015). Finally, we measure the isotropic dilation scale, α, of the BAO signal by fitting a model template of the BAO peak (Ross et al. 2015), obtaining again a good agreement between the simulated and observational SDSS data (Fig. 19).

Based on our results, we conclude that the Planck Λ CDM cosmology nicely explains the observed statistics of the large-scale structure of the SDSS main galaxy survey.

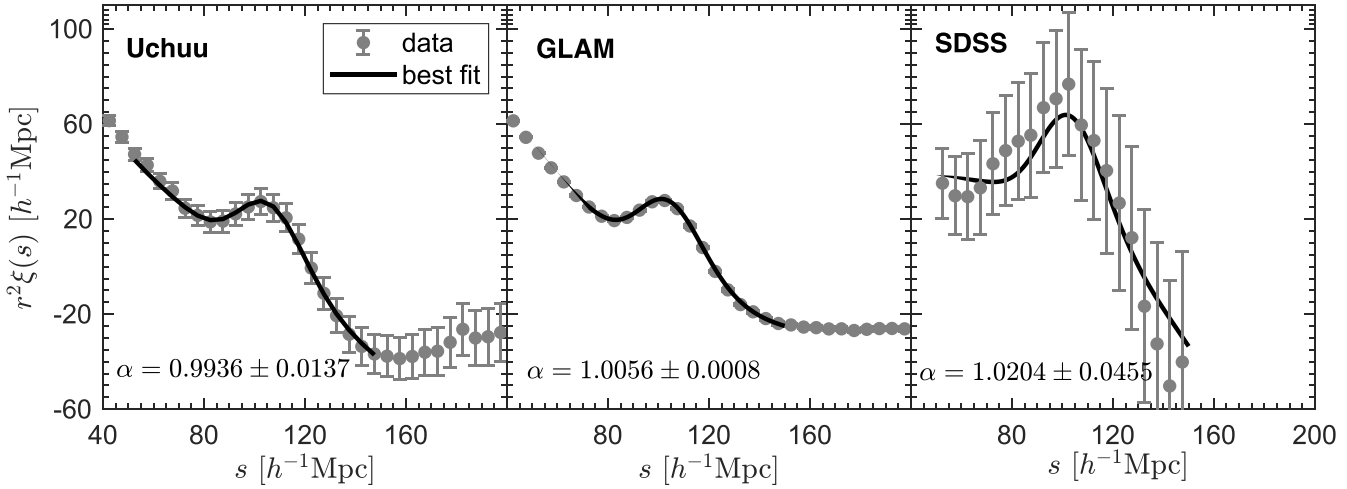

Figure 19. BAO measurements from the correlation function monopole using the Uchuu (left panel) and GLAM (middle panel) simulated light-cones, and SDSS observations (right panel) at $z = 0.15$. Measurements and best-fitting models are shown with error bars and solid lines. We used a bin width of $\Delta s = 5\,h^{-1}$Mpc. The error bars for Uchuu and GLAM were rescaled by the square root of the number of light-cones.

This work shows the great potential of the Uchuu simulation as a canvas for the creation of simulated galaxies (and quasars) for large surveys. The procedures presented in this paper can be readily applied for the creation of high-fidelity light-cones from Uchuu, and covariance errors from GLAM simulations, tailored for upcoming galaxy surveys such as DESI (DESI Collaboration et al. 2016), Euclid (Laureijs et al. 2011), and LSST (LSST Science Collaboration et al. 2009). This will aid the evaluation of analysis pipelines, the assessment of observational biases and systematic effects, and enable cosmological models to be probed.

ACKNOWLEDGEMENTS

We thank Arnaud De-Mattia, Gary Mamon, and Marko Shuntov for helpful discussions. We thank Instituto de Astrofisica de Andalucia (IAA-CSIC), Centro de Supercomputacion de Galicia (CESGA), and the Spanish academic and research network (RedIRIS) in Spain for hosting Uchuu DR1 in the Skies & Universes site for cosmological simulations. The Uchuu simulations were carried out on Aterui II supercomputer at Center for Computational Astrophysics, CfCA, of National Astronomical Observatory of Japan, and the K computer at the RIKEN Advanced Institute for Computational Science. The Uchuu DR1 effort has made use of the `skun@IAA-RedIRIS` and `skun6@IAA` computer facilities managed by the 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). 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. CAD-P, AS, JE, FP, AK, and JR thank the support of the Spanish Ministry of Science and Innovation funding grant PGC2018-101931-B-I00. CAD-P gratefully acknowledges generous funding from the John Simpson Greenwell Memorial Fund. CH-A 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. TI has been supported by IAAR Research Support Program in Chiba University Japan, MEXT/JSPS KAKENHI (Grant Number JP19KK0344, JP21F51024, and JP21H01122), MEXT as 'Program for Promoting Researches on the Supercomputer Fugaku' (JPMXP1020200109), and JICFuS.

DATA AVAILABILITY

The 32 Uchuu–SDSS galaxy light-cones, the 6 Uchuu-box galaxy catalogues at redshifts $z = \{0, 0.093, 0.19, 0.3, 0.43, 0.49\}$, the 5100 Uchuu–SDSSbao galaxy light-cones, and the companion SDSS LSS catalogue used in this work are made available at http://www.skiesanduniverses.org/Simulations/Uchuu/, together with the information on how to read the data. For a list and brief description of the available catalogue columns, please see the supplementary material.

REFERENCES

Abazajian K. N. et al., 2009, ApJS, 182, 543
Alam S. et al., 2021, Phys. Rev. D, 103, 083533
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
Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, MNRAS, 488, 3143
Blanton M. R., 2006, ApJ, 648, 268
Blanton M. R. et al., 2003a, AJ, 125, 2348
Blanton M. R. et al., 2003b, ApJ, 592, 819
Blanton M. R. et al., 2005, AJ, 129, 2562
Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, MNRAS, 351, 1151
Brooks S. P., Gelman A., 1998, J. Comput. Graph. Stat., 7, 434
Campbell D., van den Bosch F. C., Padmanabhan N., Mao Y.-Y., Zentner A. R., Lange J. U., Jiang F., Villarreal A., 2018, MNRAS, 477, 359
Chang Y.-Y., van der Wel A., da Cunha E., Rix H.-W., 2015, ApJS, 219, 8
Charlot S., Fall S. M., 2000, ApJ, 539, 718
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., 2020, J. Cosmol. Astropart. Phys., 2020, 062
Chen S.-F., Vlah Z., Castorina E., White M., 2021, J. Cosmol. Astropart. Phys., 2021, 100
Cole S. et al., 2005, MNRAS, 362, 505

Colless M. et al., 2001, MNRAS, 328, 1039
Conroy C., Wechsler R. H., Kravtsov A. V., 2006, ApJ, 647, 201
Conroy C., Gunn J. E., White M., 2009, ApJ, 699, 486
de la Torre S. et al., 2013, A&A, 557, A54
DESI Collaboration et al., 2016, preprint (arXiv:1611.00036)
Dembinski H. et al., 2024, Zenodo.
Duarte M., Mamon G. A., 2015, MNRAS, 453, 3848
Dubois Y. et al., 2014, MNRAS, 444, 1453
Eisenstein D. J. et al., 2005, ApJ, 633, 560
Eisenstein D. J., Seo H.-J., Sirko E., Spergel D. N., 2007, ApJ, 664, 675
Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Gao L., Springel V., White S. D. M., 2005, MNRAS, 363, L66
Gelman A., Rubin D. B., 1992, Stat. Sci., 7, 457
Gorski K. M., Wandelt B. D., Hansen F. K., Hivon E., Banday A. J., 1999, preprint(astro-ph/9905275)
Guo H. et al., 2015, MNRAS, 453, 4368
Guzzo L. et al., 2008, Nature, 451, 541
Hartlap J., Simon P., Schneider P., 2007, A&A, 464, 399
Hayashi E., Navarro J. F., Taylor J. E., Stadel J., Quinn T., 2003, ApJ, 584, 541
Howlett C., Ross A. J., Samushia L., Percival W. J., Manera M., 2015, MNRAS, 449, 848
Ishiyama T., Fukushige 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
Ivezić Ž. et al., 2019, ApJ, 873, 111
Kaiser N., 1987, MNRAS, 227, 1
Kauffmann G. et al., 2003, MNRAS, 341, 33
Kennicutt R. C., Jr, 1998, ARA&A, 36, 189
Klypin A., Holtzman J., 1997, preprint(astro-ph/9712217)
Klypin A., Prada F., 2018, MNRAS, 478, 4602
Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, ApJ, 609, 35
Kroupa P., 2001, MNRAS, 322, 231
LSST Science Collaboration et al., 2009, preprint (arXiv:0912.0201)
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Leslie S. K., Kewley L. J., Sanders D. B., Lee N., 2016, MNRAS, 455, L82
Lin S. et al., 2020, MNRAS, 498, 5251
Loveday J. et al., 2012, MNRAS, 420, 1239
Marinoni C., Hudson M. J., 2002, ApJ, 569, 101
McCullagh N., Norberg P., Cole S., Gonzalez-Perez V., Baugh C., Helly J., 2017, preprint (arXiv:1705.01988)
Moustakas J. et al., 2013, ApJ, 767, 50
Peebles P. J. E., 1980, The large-scale structure of the universe, Princeton University Press, Princeton, United States.
Perlmutter S. et al., 1999, ApJ, 517, 565
Pillepich A. et al., 2018, MNRAS, 475, 648
Planck Collaboration XIII, 2016, A&A, 594, A13
Planck Collaboration VI, 2020, A&A, 641, A6
Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, ApJ, 771, 30
Riess A. G. et al., 1998, AJ, 116, 1009
Rodríguez-Torres S. A. et al., 2016, MNRAS, 460, 1173
Ross A. J., Samushia L., Howlett C., Percival W. J., Burden A., Manera M., 2015, MNRAS, 449, 835
Safonova S., Norberg P., Cole S., 2021, MNRAS, 505, 325
Salim S. et al., 2007, ApJS, 173, 267
Schaye J. et al., 2015, MNRAS, 446, 521
Seljak U., McDonald P., 2011, J. Cosmol. Astropart. Phys., 2011, 039
Shu Y., Bolton A. S., Schlegel D. J., Dawson K. S., Wake D. A., Brownstein J. R., Brinkmann J., Weaver B. A., 2012, AJ, 143, 90
Skibba R. A., Sheth R. K., 2009, MNRAS, 392, 1080
Smith A., Cole S., Baugh C., Zheng Z., Angulo R., Norberg P., Zehavi I., 2017, MNRAS, 470, 4646
Smith A. et al., 2019, MNRAS, 484, 1285
Smith A., Cole S., Grove C., Norberg P., Zarrouk P., 2022a, MNRAS, 516, 1062
Smith A., Cole S., Grove C., Norberg P., Zarrouk P., 2022b, MNRAS, 516, 4529
Springel V. et al., 2018, MNRAS, 475, 676
Stiskalek R., Desmond H., Holvey T., Jones M. G., 2021, MNRAS, 506, 3205
Swanson M. E. C., Tegmark M., Hamilton A. J. S., Hill J. C., 2008, MNRAS, 387, 1391
Szewciw A. O., Beltz-Mohrmann G. D., Berlind A. A., Sinha M., 2022, ApJ, 926, 15
Takada M. et al., 2014, PASJ, 66, R1
Taruya A., Nishimichi T., Saito S., 2010, Phys. Rev. D, 82, 063522
Taylor E. N. et al., 2011, MNRAS, 418, 1587
Trujillo-Gomez S., Klypin A., Primack J., Romanowsky A. J., 2011, ApJ, 742, 16
Vale A., Ostriker J. P., 2004, MNRAS, 353, 189
Verheijen M. A. W., 2001, ApJ, 563, 694
Wechsler R. H., 2001, PhD thesis, University Of California, Santa Cruz
Wechsler R. H., Tinker J. L., 2018, ARA&A, 56, 435
Wechsler R. H., Zentner A. R., Bullock J. S., Kravtsov A. V., Allgood B., 2006, ApJ, 652, 71
White M., 2015, MNRAS, 450, 3822
White M., Tinker J. L., McBride C. K., 2014, MNRAS, 437, 2594
York D. G. et al., 2000, AJ, 120, 1579
Zehavi I. et al., 2005, ApJ, 630, 1
Zehavi I. et al., 2011, ApJ, 736, 59
Zhao C. et al., 2021, MNRAS, 503, 1149

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX A: THE EFFECT OF SCATTER IN SHAM

As discussed in Section 3.2, our SHAM model has only one free parameter σ, which roughly corresponds to the scatter in ${}^{0.1}M_{\rm r}^{h}$ at a fixed value of our halo mass proxy, *V*peak. In practice, the value of σ affects mainly the amplitude of the TPCF for a given volume-limited sample, results are shown in Fig. 1 for different values of the scatter. Increasing the value of σ introduces a larger number of low-$V_{\rm peak}$ galaxies into the volume-limited samples, which are weakly clustered (and also removes high-$V_{\rm peak}$ galaxies, which are strongly clustered). This has the effect of decreasing the amplitude of the TPCF. The effect of scatter is consequently larger for strongly clustered volume-limited samples, becoming negligible for samples with magnitude threshold ${}^{0.1}M_{r}^{h} > -20$. In order to tune the scatter parameter, we optimize the goodness-of-fit of the TPCF monopole for the Uchuu snapshot at $z = 0.093$ to the SDSS measurements, finding an optimal value of $\sigma = 0.5$. We choose this snapshot since it is the closest to the median SDSS redshift $z = 0.1$.

This value is in agreement with previous observational estimations of the scatter using SHAM. For example, Trujillo-Gomez et al. (2011) estimate the scatter of ${}^{0.1}M_{r}^{h}$ at fixed halo circular velocity, by combining an estimate of the intrinsic scatter of the Tully–Fisher relation (Verheijen 2001) with the scatter resulting from the distribution of dust extinction corrections in SDSS, arriving precisely at a value of 0.5.

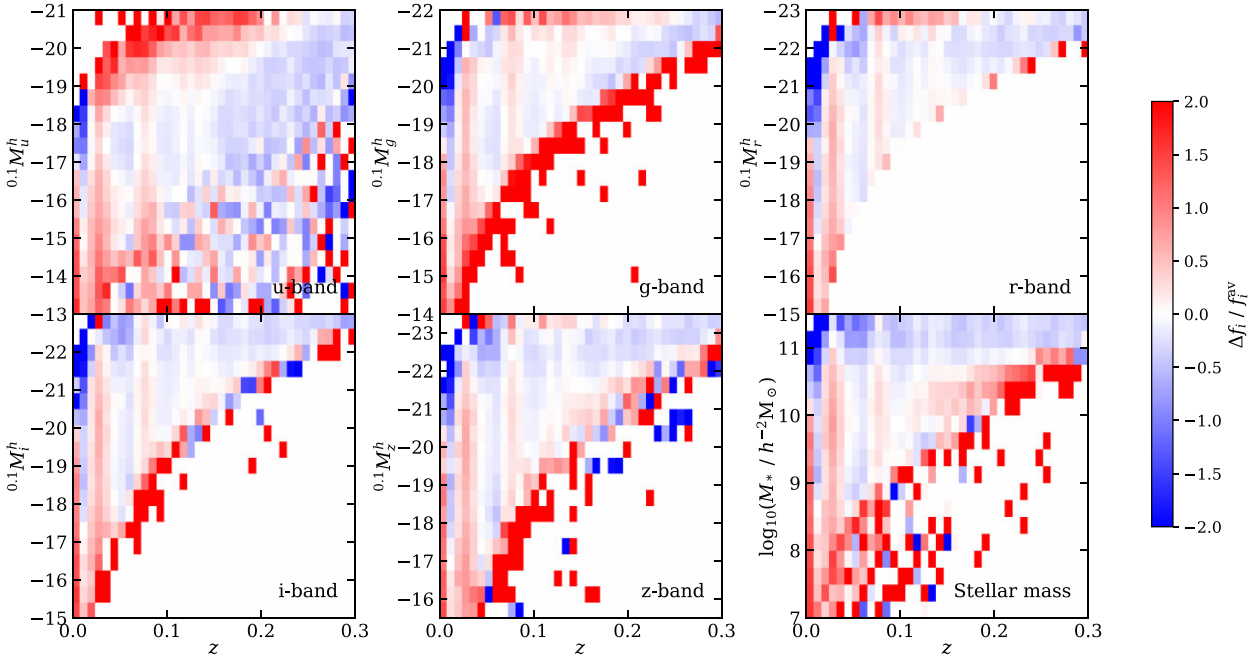

Figure B1. Difference in the magnitude and stellar mass distributions between SDSS and the Uchuu mock. The plot shows $(f_i^{SDSS} - f_i^{Uchuu})/f_i^{av}$, where f_i is the fraction of galaxies in each bin of redshift and magnitude (or stellar mass) from SDSS or Uchuu, and f_i^{av} is the average of these. Blue indicates there are more galaxies in the mock, while red indicates that there are more galaxies in SDSS. In bins where there are many galaxies, the differences are small. Some vertical features can be seen, due to large-scale structure in the catalogues.

APPENDIX B: THE DIFFERENCE BETWEEN THE UCHUU AND SDSS MAGNITUDE DISTRIBUTIONS

To highlight the differences between the Uchuu mock and the SDSS data in terms of the distribution of properties (magnitudes, stellar masses) as a function of redshift, we show the residuals in Fig. B1. The difference between the frequency of galaxies in each bin for SDSS and Uchuu (f_i^{SDSS} and f_i^{Uchuu} respectively), normalized by the average density of the two samples, is very small except for the regions of the very low density of points.

This paper has been typeset from a TEX/LATEX file prepared by the author.