



A lightcone catalogue from the Millennium-XXL simulation

Alex Smith,^{1★} Shaun Cole,¹ Carlton Baugh,¹ Zheng Zheng,² Raúl Angulo,³
Peder Norberg^{1,4} and Idit Zehavi⁵

¹*Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK*

²*Department of Physics and Astronomy, University of Utah, 115 South 1400 East, Salt Lake City, UT 84112, USA*

³*Centro de Estudios de Física del Cosmos de Aragón, Plaza San Juan 1, Planta-2, E-44001 Teruel, Spain*

⁴*Centre for Extragalactic Astronomy, Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK*

⁵*Department of Astronomy, Case Western Reserve University, Cleveland, OH 44106, USA*

Accepted 2017 June 7. Received 2017 May 16; in original form 2017 January 23

ABSTRACT

Future galaxy surveys require realistic mock catalogues to understand and quantify systematics in order to make precise cosmological measurements. We present a halo lightcone catalogue and halo occupation distribution (HOD) galaxy catalogue built using the Millennium-XXL simulation. The halo catalogue covers the full sky, extending to $z = 2.2$ with a mass resolution of $\sim 10^{11} h^{-1} M_{\odot}$. We use this to build a galaxy catalogue, which has an r -band magnitude limit of $r < 20.0$, with a median redshift of $z \sim 0.2$. A Monte Carlo HOD method is used to assign galaxies to the halo lightcone catalogue, and we evolve the HODs to reproduce a target luminosity function; by construction, the luminosity function of galaxies in the mock is in agreement with the Sloan Digital Sky Survey (SDSS) at low redshifts and the Galaxy and Mass Assembly (GAMA) survey at high redshifts. A Monte Carlo method is used to assign a $^{0.1}(g - r)$ colour to each galaxy, and the colour distribution of galaxies at different redshifts agrees with measurements from GAMA. The clustering of galaxies in the mock for galaxies in different magnitude and redshift bins is in good agreement with measurements from SDSS and GAMA, and the colour-dependent clustering is in reasonable agreement. We show that the baryon acoustic oscillation can be measured in the mock catalogue, and the redshift-space distortions are in agreement with measurements from SDSS illustrating that this catalogue will be useful for upcoming surveys.

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

1 INTRODUCTION

Upcoming galaxy surveys, such as the Dark Energy Spectroscopic Instrument (DESI) survey (DESI Collaboration et al. 2016a,b) and *Euclid* (Laureijs et al. 2011), aim to measure the expansion history of the Universe and the growth of cosmic structures. Measurements of galaxy clustering, redshift-space distortions and weak lensing will test general relativity, constrain theories of dark energy and give us precise cosmological constraints.

In order to reach the high precision required to meet these aims, it is necessary to understand and quantify the systematic uncertainties in measurements from surveys, which requires the use of accurate mock catalogues (Baugh 2008). Since a mock catalogue has a known cosmology and the ‘true’ value of a statistic can be

measured directly, they can be used to develop and test the analysis tools that will be used in real observations.

Mocks are also required to test observational strategies and quantify the resultant levels of sample incompleteness. It is often not possible to assign a fibre to every galaxy due to mechanical constraints on fibre positioning (e.g. Hawkins et al. 2003; Guo, Zehavi & Zheng 2012; Burden et al. 2017; Hahn et al. 2017; Pinol et al. 2017) and even if a fibre is assigned, a redshift measurement can fail if the galaxy has weak emission lines or low surface brightness. This incompleteness may have a significant effect on clustering measurements, and therefore in order to make precise baryon acoustic oscillation (BAO) and redshift-space distortion measurements, it is important that this incompleteness is well understood, and that methods are developed and tested in order to mitigate these effects on the measured clustering. The differences in the clustering statistics expected in viable models are small, making it essential that systematics like these are understood.

* E-mail: a.m.j.smith@durham.ac.uk

Mock catalogues that have realistic galaxy clustering can be created from cosmological simulations. In order to see the BAO peak in clustering measurements, at a scale of the order of $100 h^{-1}$ Mpc, these simulations need to have a very large box size of the order of a few Gpc. Running a hydrodynamical simulation that has both the large volume needed to model such scales and the resolution to produce faint galaxies down to the flux limit of the survey is infeasible, due to the large computational expense. Dark-matter-only simulations are much less expensive. There are several schemes that can be used to populate haloes in a dark-matter-only simulation with galaxies. These include the halo occupation distribution (HOD; e.g. Peacock & Smith 2000; Seljak 2000; Scoccimarro et al. 2001; Berlind & Weinberg 2002; Kravtsov et al. 2004; Zheng et al. 2005), which describes the probability of a halo with mass M contains N galaxies with a certain property; the closely related conditional luminosity function (CLF; e.g. Yang, Mo & van den Bosch 2003), which specifies the luminosity function of galaxies at each halo mass; subhalo abundance matching (SHAM; e.g. Vale & Ostriker 2004; Conroy, Wechsler & Kravtsov 2006), which assumes a correlation between halo or subhalo properties (e.g. mass or circular velocity) and galaxy properties (e.g. luminosity or stellar mass); and semi-analytic models (SAMs; e.g. Baugh 2006; Benson 2010; Somerville & Davé 2015), which use analytic prescriptions to model the formation and evolution of galaxies.

In order to apply a SAM to a simulation, high-resolution merger trees are needed, and these are difficult to construct for large volume simulations. However, there are approaches that can augment the resolution of the simulation merger trees (e.g. de la Torre & Peacock 2013; Angulo et al. 2014; Benson, Cannella & Cole 2016). The SHAM prescription assigns galaxies to subhaloes, requiring a complete subhalo catalogue. Since subhaloes are disrupted when they undergo mergers, this catalogue will only be complete for large subhaloes with thousands of particles, and so a very high resolution simulation is needed to resolve the low-mass subhaloes that will be populated by faint galaxies. The HOD, on the other hand, can be applied to a lower resolution simulation, since satellite galaxies can be placed around the central galaxy following an analytic distribution, without knowledge of the subhaloes. The HOD method can also be applied to simulations in which the underlying cosmology has been rescaled (e.g. Angulo & White 2010).

Ideally, these methods would be used to populate a halo lightcone that is the direct output from a simulation. However, most simulations do not output lightcones, but output snapshots at discrete times. Typically when a HOD method is used, it is applied to a single snapshot. However, this means that the halo bias is constant, and so the clustering of haloes does not evolve with redshift in the mock. Multiple snapshots can be joined together to create a lightcone, but this leads to discontinuities at the boundaries; the same halo could appear twice on either side of the boundary, or not at all (e.g. Fosalba et al. 2015).

The standard abundance matching and HOD schemes do not incorporate evolution. Attempts have been made to extend the abundance matching scheme, such as the one reported by Moster, Naab & White (2013), which reproduces the observed stellar mass function at different redshifts. There is currently no complete model for HOD evolution, as this evolution would depend on the galaxy sample under consideration. Contreras et al. (2017) use the HODs produced in SAMs to build a simple parametrization of the evolution of the HOD parameters.

Here, we describe a HOD method that we use to populate haloes over a range of redshifts from the Millennium-XXL (MXXL) simulation with galaxies. We first create a halo lightcone catalogue from

the simulation by interpolating the positions of haloes between snapshots, which is then populated with galaxies using HODs, reproducing the observed clustering from the Sloan Digital Sky Survey (SDSS; Abazajian et al. 2009) and the Galaxy and Mass Assembly (GAMA) survey (Driver et al. 2009, 2011; Liske et al. 2015).

This paper is organized as follows. In Section 2, we describe the MXXL simulation and outline the method for generating the halo lightcone catalogue. In Section 3, we describe the HOD model, and our method of evolving the HODs with redshift. In Section 4, we outline the method used to populate the halo lightcone with galaxies, and the method used to assign each galaxy a $^{0.1}(g-r)$ colour. In Section 5, we give examples of potential applications of the mock catalogue.

2 HALO LIGHTCONE CATALOGUE

2.1 The MXXL simulation

The MXXL simulation (Angulo et al. 2012b) is a large dark-matter-only N -body simulation in the same family as the Millennium simulation (Springel et al. 2005). The volume of MXXL is 216 times larger than Millennium, with a box size of $3 h^{-1}$ Gpc, and the particle mass is $6.17 \times 10^9 h^{-1} M_{\odot}$, with a force softening of 13.7 kpc. MXXL adopts a Λ cold dark matter (Λ CDM) cosmology with the same 1-yr *Wilkinson Microwave Anisotropy Probe* (WMAP-1) cosmological parameters as the Millennium simulation, $\Omega_m = 0.25$, $\Omega_{\Lambda} = 0.75$, $\sigma_8 = 0.9$, $h = 0.73$ and $n = 1$ (Spergel et al. 2003). The initial conditions were set at a starting redshift of $z = 63$, and the simulation was evolved to $z = 0$ with 63 outputs. The large volume of the simulation means that it can be used to study features such as BAOs and redshift-space distortions with good statistics.

2.2 Merger trees

We use the halo merger trees computed by Angulo et al. (2012b). Haloes were found using a friends-of-friends (FOF) algorithm (Davis et al. 1985), and bound subhaloes were identified using SUBFIND (Springel et al. 2001). Halo merger trees were built by identifying the unique descendant of each subhalo at the subsequent snapshot. For each subhalo, the 15 most bound particles were found, and the subhalo at the next snapshot that contains the greatest number of these particles was defined as the descendant. In the case where two subhaloes contain equal numbers of these particles, the subhalo with the greatest total binding energy was chosen (Angulo et al. 2012a).

2.3 Constructing the halo lightcone catalogue

The full sky halo lightcone catalogue is created using the standard interpolation method (e.g. Merson et al. 2013, but applied to haloes rather than galaxies). An observer is first placed randomly inside the MXXL box. If the observer happened to be placed at the centre, haloes at the edge of the box would have a redshift $z \sim 0.5$; multiple periodic replications of the box must therefore be used in order to construct a catalogue that goes to redshifts higher than this. This replication is done without any artificial rotation or translation in order to prevent the introduction of discontinuities. The positions and velocities of each halo at each snapshot are used to interpolate their trajectories through the simulation. From the position and redshift of a halo at two adjacent snapshots, it can be determined whether the halo crossed the observer's lightcone; if it has, a binary search

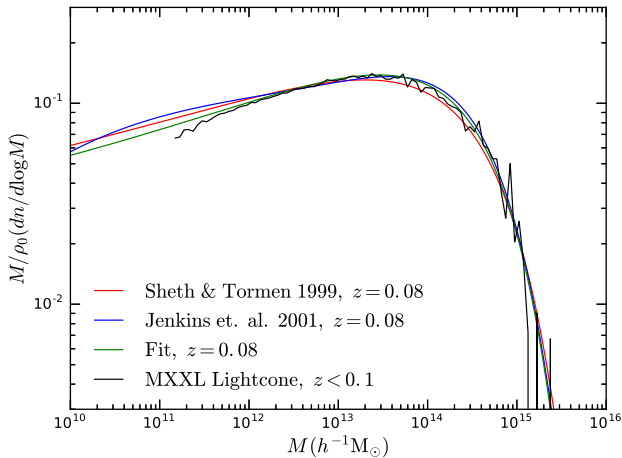


Figure 1. Mass function of the halo lightcone catalogue for $z < 0.1$ (black) compared to the analytic mass function of Sheth & Tormen (1999) (red), Jenkins et al. (2001) (blue) and our fit to the MXXL mass function (green) at the median redshift $z = 0.08$. Halo masses are defined as M_{200m} , and have been interpolated linearly between simulation snapshots.

algorithm is used to find the interpolated position (and velocity) at the redshift where it crosses.

The HOD method of creating the galaxy catalogue (Section 3) assigns galaxies to FOF groups. Since the merger tree is defined for SUBFIND subhaloes, we need to infer the merger tree for the FOF haloes. To do this, we make the assumption that the position (and velocity) of the main subhalo (i.e. the most massive subhalo) in each FOF group is the same as that of the FOF group itself. The descendant FOF group is then found from the descendant of the main subhalo. To interpolate the position (and velocity) of each subhalo, we use cubic interpolation (i.e. use a cubic polynomial to describe the path of the halo in each dimension, using the positions and velocities at the previous and next snapshot as boundary conditions).

We use a halo mass definition of M_{200m} (the mass enclosed by a sphere, centred on the halo, in which the average density is 200 times the mean density of the Universe), as stored in the MXXL output for each FOF group. Since the number of galaxies in each halo depends on its mass, M_{200m} must be interpolated between snapshots. Below $z = 2$, the simulation snapshots are approximately spaced linearly with expansion factor. We use the descendant of the most massive subhalo to find the descendant of each halo, and then interpolate linearly in mass between snapshots, finding the mass at the redshift at which it crosses the lightcone. In the case that two or more haloes merge between snapshots, the total mass of the haloes is interpolated linearly, and each halo is assigned a constant fraction of the total mass. If the halo is not the most massive progenitor, a random time between snapshots is chosen for the merger to take place. If the halo crosses the observer's lightcone after this time, the merger has happened, and the interpolated mass of the halo is transferred to the most massive progenitor.

The mass function of the halo lightcone at low redshifts $z < 0.1$ is shown in Fig. 1 and compared to the Sheth & Tormen (1999) and Jenkins et al. (2001) analytic mass functions. At high masses, the mass function of the lightcone catalogue is in reasonable agreement with Sheth & Tormen (1999) and Jenkins et al. (2001), but there is a lower abundance of less massive haloes. This difference is because haloes in the simulation are identified using a FOF algorithm, and not a spherical overdensity (SO) finder. However, a SO mass is calculated for each FOF halo (M_{200m}). Any small overdensities close to a large FOF group would be identified as part of the large

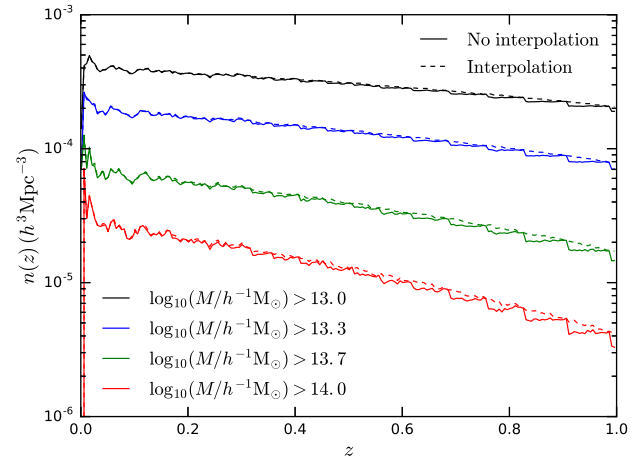


Figure 2. Number density of haloes in the halo lightcone catalogue as a function of redshift for haloes with mass M_{200m} greater than several thresholds, as indicated by the colour. Solid lines are where the halo mass has been kept frozen between snapshots, and dashed lines are where the mass has been interpolated.

FOF group, and therefore these would be missing from the halo catalogue.

In order to add haloes to the catalogue that are below the MXXL mass resolution (Section 2.5), and to evolve the HODs with redshift (Section 3.2), it is useful to have a smooth function that is in close agreement with the actual mass function of the catalogue. For this, we take a mass function with the same form as Sheth & Tormen (1999), but refit the parameters to the MXXL mass function. This fit is shown as the green curve in Fig. 1, which is in better agreement with the MXXL mass function at low masses, and is close to the fit given in equation (2) of Angulo et al. (2012b). The MXXL mass function peels away from this fit slightly at masses close to the resolution limit, but is complete for masses greater than $\sim 10^{12} h^{-1} M_{\odot}$.

The number density of haloes as a function of redshift in the halo lightcone catalogue is shown in Fig. 2 for several mass thresholds. If halo masses are kept fixed between snapshots, step features can be seen, since the mass function is being kept frozen and will only change at the next snapshot. These features are most apparent for the highest mass threshold, for which the number density decreases more rapidly at high redshifts. Mass interpolation greatly reduces these features.

The large-scale real-space correlation function of the lightcone catalogue for FOF groups with masses $M_{200m} > 3 \times 10^{12} h^{-1} M_{\odot}$ and $z < 0.5$ is shown in Fig. 3. This redshift limit avoids structures being repeated due to periodic replication of the box.¹ The BAO peak can be seen clearly in the clustering of haloes.

2.4 Caveats

The interpolation scheme uses the position and velocity of a halo at two snapshots as boundary conditions in order to find the path the halo moved through in the simulation. If two haloes merge together, there is not enough information to determine when this occurs, so we assume they merge at a random time. If a new halo forms, or

¹ Lightcones with a wide opening angle, or directed along the principal axes of the simulation, which extend beyond $z = 0.5$ will contain repeated structures. This will result in clustering errors being underestimated.

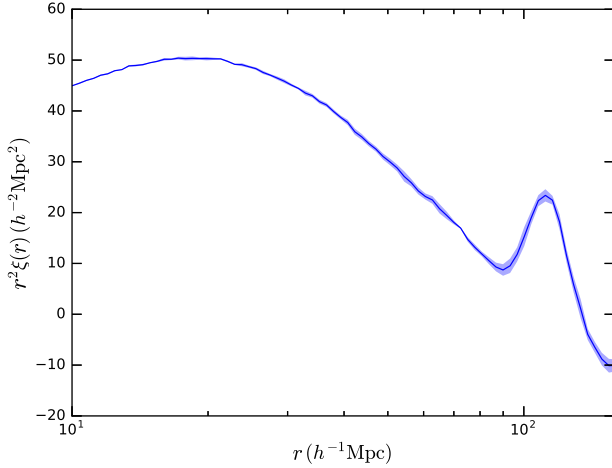


Figure 3. Real-space correlation function, scaled by r^2 , of the halo lightcone catalogue for haloes with $M_{200m} > 3 \times 10^{12} h^{-1} M_{\odot}$ and $z < 0.5$. The blue shaded area shows the error on the mean in the clustering, calculated from four quadrants of the sky.

drops below the resolution limit, we assume this happens exactly on a snapshot.

To construct the merger trees, the descendant of a subhalo is defined as the subhalo that contains the majority of its 15 most bound particles (Angulo et al. 2012a). However, it is likely that some of the particles of the descendant subhalo were not in its progenitor, and vice versa. All of these particles are used to calculate the position and velocity of the subhalo, which can occasionally lead to jumps in the position of a subhalo that are inconsistent with its velocity.

Sometimes, a halo can be lost by the halo finder at one snapshot, but is then found again at a later snapshot. This can happen if a small halo passes very close to a more massive halo at one snapshot; the SUBFIND algorithm can fail to identify the small halo as the algorithm finds that its particles are bound to the massive halo. The MXXL merger trees we use do not make any attempt to add in these haloes lost by SUBFIND. However, since we use the most massive subhalo in a FOF group to trace the FOF merger trees, and use M_{200m} as the mass definition, our results should not be affected much by small subhaloes being lost by the halo finder.

2.5 Haloes below the mass resolution

Populating the resolved haloes in the MXXL halo lightcone with galaxies will result in incompleteness in a magnitude-limited galaxy catalogue at low redshifts. This is because intrinsically faint galaxies that are sufficiently close to the observer to be bright enough to be included in the catalogue occupy haloes that fall below the MXXL mass resolution. We use our fit to the MXXL halo mass function in order to add these haloes into the lightcone catalogue, and position them randomly in the catalogue so that they are unclustered. Other methods for augmenting the halo catalogue exist (e.g. de la Torre & Peacock 2013; Angulo et al. 2014; Benson et al. 2016), but we find that this simple method is able to bring the dN/dz of galaxies in the catalogue into better agreement with the measured dN/dz from GAMA, while only having a very small effect on the measured clustering.

The redshift distribution of the haloes that need to be added to the lightcone catalogue can be calculated from the integral,

$$\frac{dN}{dz} = \int_{M_{\min}(z)}^{M_{\max}} n_{\text{unres}}(M, z) \frac{dV}{dz} dM, \quad (1)$$

where $n_{\text{unres}}(M, z)$ is the number density of unresolved haloes, dV/dz is the comoving volume per unit redshift, $M_{\min}(z)$ is the minimum halo mass that could host a galaxy brighter than the faintest observable galaxy in the survey at that redshift and $M_{\max} = 10^{12} h^{-1} M_{\odot}$ is the mass at which the MXXL mass function is judged to be complete. If the survey is flux limited, then the faintest observable galaxy at each redshift is set by an apparent magnitude threshold; for our mock catalogue we set this threshold to $r = 20$, as this is the magnitude threshold for the DESI Bright Galaxy Survey (BGS; DESI Collaboration et al. 2016a). The number density of unresolved haloes is given by $n_{\text{unres}}(M, z) = n_{\text{fit}}(M, z) - n_{\text{res}}(M, z)$, where $n_{\text{fit}}(M, z)$ is our fit to the number density of haloes in the lightcone, extrapolated to low masses, and $n_{\text{res}}(M, z)$ is the number density of haloes resolved in MXXL. We model the mass function of resolved haloes by multiplying the fit to the mass function by a cut-off at the mass resolution limit of $M_{200m} \sim 10^{11} h^{-1} M_{\odot}$,

$$n_{\text{unres}}(M, z) = [1 - \text{cut}(M, z)] n_{\text{fit}}(M, z), \quad (2)$$

where a good approximation to the cut-off is given by

$$\text{cut}(M, z) = 10^{(-z-2)(\log_{10}(M/h^{-1} M_{\odot})-11)^{0.6}}. \quad (3)$$

In order to add unresolved haloes to the catalogue, we first randomly draw a redshift for each unresolved halo from the dN/dz distribution defined in equation (1). The mass of each halo is then randomly assigned using the mass distribution at the redshift of the halo defined by $n_{\text{unres}}(M, z)$. The haloes are then randomly positioned uniformly on the sky. Since the unresolved haloes are randomly positioned so that they are unclustered, redshift-space distortions do not affect their clustering, and so we set the velocity of each of these haloes to zero. A random concentration is also assigned from the mass–concentration relation of MXXL (with scatter), extrapolated to lower masses.

While the introduction of unclustered haloes only has a small effect on the two-point correlation function, other statistics might also change, for example, density estimators and void statistics. We have not checked the size of this effect, but the final galaxy catalogue includes a flag that indicates whether a galaxy lives inside one of these haloes, enabling these galaxies to be removed when calculating other statistics.

3 HALO OCCUPATION DISTRIBUTION

Galaxies are biased tracers of the underlying dark matter density field. The HOD describes this bias using the probability that a halo of mass M contains N galaxies with a certain property, $P(N|M)$, providing a physical interpretation of galaxy clustering measurements.

The mean number of galaxies in a halo with mass M which are brighter than some luminosity threshold, L , can be written as a sum of central and satellite galaxies (e.g. Zheng et al. 2005),

$$\langle N_{\text{gal}}(>L|M) \rangle = \langle N_{\text{cen}}(>L|M) \rangle + \langle N_{\text{sat}}(>L|M) \rangle. \quad (4)$$

We use central and satellite occupation functions of the same form as Zehavi et al. (2011). The mean number of central galaxies brighter than L is described by a smoothed step function,

$$\langle N_{\text{cen}}(>L|M) \rangle = \frac{1}{2} \left[1 + \text{erf} \left(\frac{\log M - \log M_{\min}(L)}{\sigma_{\log M}(L)} \right) \right], \quad (5)$$

where $\text{erf}(x) = 2\pi^{-1/2} \int_0^x e^{-x'^2} dx'$ is the error function. The parameter M_{\min} is the halo mass for which half of haloes contain a galaxy brighter than L , and $\sigma_{\log M}$ sets the width of the step. For $M \gg M_{\min}$, $\langle N_{\text{cen}}(>L|M) \rangle = 1$, while for $M \ll M_{\min}$, $\langle N_{\text{cen}}(>L|M) \rangle = 0$. The

mean number of satellites per halo brighter than L is given by a power law,

$$\langle N_{\text{sat}}(>L|M) \rangle = \langle N_{\text{cen}}(>L|M) \rangle \left(\frac{M - M_0(L)}{M'_1(L)} \right)^{\alpha(L)}, \quad (6)$$

where M_0 is the cut-off mass scale, M'_1 the normalization and α the power-law slope. M'_1 is different from M_1 , the mass of a halo which on average contains one satellite, although the two quantities are related.² The power law is also multiplied by the central occupation function, which ensures that the brightest galaxy in the halo is the central; there cannot be a satellite brighter than L without there first being a central galaxy brighter than L .

3.1 HODs at redshift 0.1

We use HOD parameters calculated from the SDSS using the procedure of Zehavi et al. (2011). These are calculated for different luminosity threshold galaxy samples, using a Markov chain Monte Carlo (MCMC) code to find the best-fitting HOD parameters that reproduce the measured projected correlation functions within the SDSS uncertainties. Since the cosmology of the MXXL simulation is different from that used by Zehavi et al. (2011), the parameter fitting was redone using the Millennium cosmology. The SDSS HODs use the absolute r -band magnitude of each galaxy, k -corrected to a reference redshift of $z_{\text{ref}} = 0.1$ (see Section 4.3), which is the median redshift of the survey. We denote absolute magnitudes k -corrected to this redshift as $^{0.1}M_r$. Absolute magnitudes written as $^{0.1}M_r$ assume $h = 1$.

The best-fitting HOD parameters, in Millennium cosmology, are shown by the points in Fig. 4. We do not show the errors as they are misleading, due to the probability distributions being highly asymmetric. Projecting these asymmetric probability distributions to one-dimensional errors can lead to the best-fitting values of some of the parameters being outside the error bars, particularly for $\sigma_{\log M}$.

For each HOD parameter, a least-squares routine is used to fit a function that describes the variation with luminosity, which are shown by the dashed lines in Fig. 4. The top panel shows $M_{\min}(L)$ and $M'_1(L)$, for which we fit curves of the same functional form as equation (11) from Zehavi et al. (2011),

$$L/L_* = A \left(\frac{M}{M_t} \right)^{\alpha_M} \exp \left(-\frac{M_t}{M} + 1 \right), \quad (7)$$

where A , M_t and α_M are free parameters. We fit a power law to $M_0(L)$ (second panel). This is a poor fit for the points at $^{0.1}M_r = -18$, which is over three orders of magnitude lower than the fit, and $^{0.1}M_r = -19$, which is 20 orders of magnitude lower. However, increasing the value of this parameter by many orders of magnitude has a very small effect on the shape of the HODs. This is because the occupation function of central galaxies adds a second cut-off to equation (6); if M_0 is below this cut-off, it will not affect the shape of the HODs. The parameter $\alpha(L) \sim 1$ at low luminosities, but increases for the highest luminosity samples (third panel). We fit a linear relation, which smoothly transitions to $\alpha = 1$ at low luminosities. $\sigma_{\log M}(L)$ (bottom panel) is small at low luminosities, with a step up to ~ 0.7 for the two brightest samples. We fit a sigmoid function to $\sigma_{\log M}$, where the width of the step is set such that the HODs do not overlap.

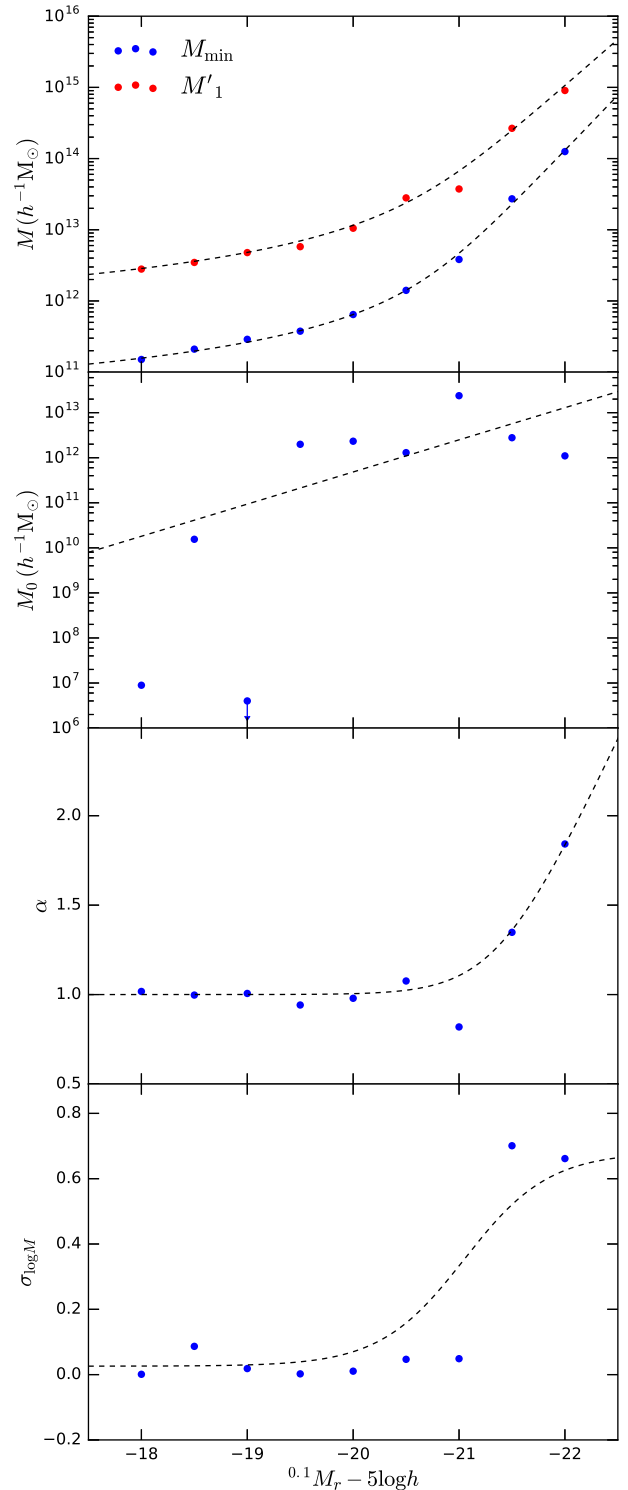


Figure 4. Best-fitting HOD parameters to the SDSS volume-limited samples in Millennium cosmology (points), and smooth functions fitted to these points (dashed lines), as a function of magnitude. Top panel: M_{\min} (blue) and M'_1 (red). Second panel: M_0 . Third panel: α . Bottom panel: $\sigma_{\log M}$. The $^{0.1}M_r - 5 \log h = -19$ sample has $M_0 = 10^{-10.2} h^{-1} M_{\odot}$, but M_0 is poorly constrained. Errors are not shown as they are misleading, due to the highly asymmetric probability distributions.

² Since the satellite occupation function is modified by the centrals, the relation $M_1 = M'_1 + M_0$ is not exact.

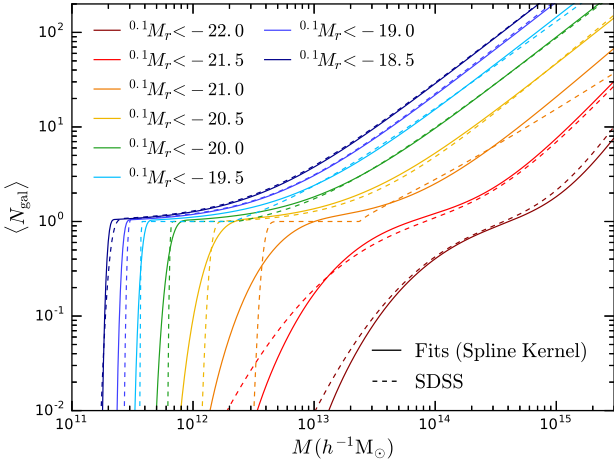


Figure 5. Mean halo occupation functions for luminosity threshold samples, as described by equations (4)–(6), using SDSS HOD parameters in the Millennium cosmology (dashed lines) and our fits to the HOD parameters, using equation (10) in place of equation (5) to describe the contribution from central galaxies (solid lines). Colours indicate the luminosity threshold, as shown by the legend.

The large step in $\sigma_{\log M}(L)$ means that, as the luminosity threshold is increased, there is a rapid jump in the amount of scatter in the luminosities of central galaxies. This results in overlapping HODs, as can be seen for the $^{0.1}M_r < -21$ and $^{0.1}M_r < -21.5$ samples in Fig. 5. For two luminosity thresholds L_1 and L_2 , where $L_1 < L_2$, it must be true that $\langle N_{\text{gal}}(>L_1|M) \rangle \geq \langle N_{\text{gal}}(>L_2|M) \rangle$ since all galaxies brighter than L_2 are also brighter than L_1 . However, if the two occupation functions cross, then this condition is not satisfied for haloes below the mass at which they cross. We therefore must model the HODs such that there is no overlap. There exist HOD frameworks in which the occupation functions cannot overlap (see e.g. Leauthaud et al. 2011), but since we are using HOD parameters obtained using the standard HOD framework, we make a small modification to these HODs to prevent any crossing, as set out below.

Equation (5) assumes that the scatter set by the parameter $\sigma_{\log M}(L)$ is Gaussian. Since a Gaussian function has a long tail that extends to infinity, there will always be an overlap between HODs if $\sigma_{\log M}(L_2) > \sigma_{\log M}(L_1)$. We instead approximate the Gaussian by using a spline kernel (Schoenberg 1946),

$$\begin{aligned} 1 - 6|x|^2 + 6|x|^3 & \quad |x| \leq 0.5, \\ \text{spline}(x) = 2(1 - |x|)^3 & \quad 0.5 < |x| \leq 1, \\ 0 & \quad |x| > 1, \end{aligned} \quad (8)$$

which has $\text{spline}(0) = 1$, $\text{mean} = 0$, $\text{variance} = 1/12$ and $\text{spline}(x) = 0$ for $|x| > 1$. This function can be rescaled and normalized to approximate any Gaussian of mean μ and variance σ^2 as

$$S(x) = \frac{4/3}{\sigma\sqrt{12}} \text{spline}\left(\frac{x - \mu}{\sigma\sqrt{12}}\right). \quad (9)$$

The HOD for central galaxies can therefore be written as

$$\langle N_{\text{cen}}(>L|M) \rangle = \frac{1}{2} \left[1 + F\left(\frac{\log M - \log M_{\min}(L)}{\sigma_{\log M}(L)}\right) \right], \quad (10)$$

where $F(x) = 2 \int_0^x S(x') dx'$. The best-fitting values of $\sigma_{\log M}$ (shown by the points in the bottom panel of Fig. 4) suggest a sharp step between $^{0.1}M_r = -21$ and $^{0.1}M_r = -21.5$. Even using equation (10), the HODs will still overlap with this abrupt step, but

they will not overlap if the step is gradual, unlike equation (5). We make the step in $\sigma_{\log M}(L)$ as narrow as we can while preventing the crossing of HODs (shown by the dashed curve).

The HODs using the SDSS HOD parameters, and our fits, are shown in Fig. 5. Our fits produce halo occupation functions that are in reasonable agreement with the SDSS HODs, with the exception of the $^{0.1}M_r < -21$ and $^{0.1}M_r < -21.5$ samples, where the width of the step set by the parameter $\sigma_{\log M}$ is too broad and narrow, respectively. This is necessary to prevent the HODs from overlapping. The SDSS HOD for the $^{0.1}M_r < -21$ sample appears to have a sharp transition from centrals to satellites, which is due to a large value of M_0 compared to M_1 .

3.2 Redshift evolution

In order to evolve the occupation functions with redshift, we first choose a target luminosity function, $\phi_{\text{target}}(L, z)$, that we would like the galaxies in the mock catalogue to reproduce as a function of redshift. This luminosity function defines a mapping between a luminosity threshold L at redshift z , and the number density of galaxies brighter than this, $n_{\text{gal}}^{\text{target}}(>L, z)$.

For a given HOD, the number density of galaxies brighter than L can be calculated from the integral

$$n_{\text{gal}}(>L, z) = \int n_{\text{halo}}(M, z) \langle N(>L|M, z) \rangle dM, \quad (11)$$

where $n_{\text{halo}}(M, z)$ is the number density of haloes of mass M at redshift z , and $\langle N(>L|M, z) \rangle$ is the halo occupation function at redshift z . The HODs must evolve with redshift such that the condition $n_{\text{gal}}(>L, z) = n_{\text{gal}}^{\text{target}}(>L, z)$ is satisfied.

Since the target luminosity function defines a mapping between a luminosity threshold and the number density of galaxies, the occupation functions can be rewritten as a function of number density, $n_{\text{gal}}: \langle N(>L|M, z) \rangle \equiv \langle N(n_{\text{gal}}|M, z) \rangle$. The shape of the HOD could evolve in a complex way, but for simplicity we keep the shape of the occupation function fixed for constant galaxy number density, but slide the HODs along the halo mass axis such that the target luminosity function is achieved. That is, the HOD parameters $\sigma_{\log M}(n_{\text{gal}}, z)$ and $\alpha(n_{\text{gal}}, z)$ are kept constant, but the three mass parameters M_{\min} , M_0 and M_1 are all multiplied by some factor f ,

$$M_{\text{HOD}}(n_{\text{gal}}, z) = f(n_{\text{gal}}, z) M_{\text{HOD}}(n_{\text{gal}}), \quad (12)$$

where M_{HOD} is one of the HOD mass parameters. The value of f required to achieve the target luminosity function is found by finding the root of the equation

$$n_{\text{gal}}(>L, f(z)) - n_{\text{gal}}^{\text{target}}(>L, z) = 0. \quad (13)$$

At high redshifts, the target luminosity function we use is the evolving Schechter function fit to the luminosity function estimated from the GAMA survey. The Schechter function can be written in terms of magnitudes as

$$\phi(M) = 0.4 \ln 10 \phi^* (10^{0.4(M^* - M)})^{1+\alpha} \exp(-10^{0.4(M^* - M)}), \quad (14)$$

where ϕ^* is the normalization, M^* is a characteristic magnitude and α is the faint-end slope. For GAMA, Loveday et al. (2012, 2015) model the evolution of the Schechter parameters with redshift as

$$\begin{aligned} \alpha(z) &= \alpha(z_0), \\ M^*(z) &= M^*(z_0) - Q(z - z_0), \\ \phi^*(z) &= \phi^*(0) 10^{0.4Pz}, \end{aligned} \quad (15)$$

where Q parametrizes the evolution in luminosity, P parametrizes the evolution in number density and $z_0 = 0.1$ is the same reference

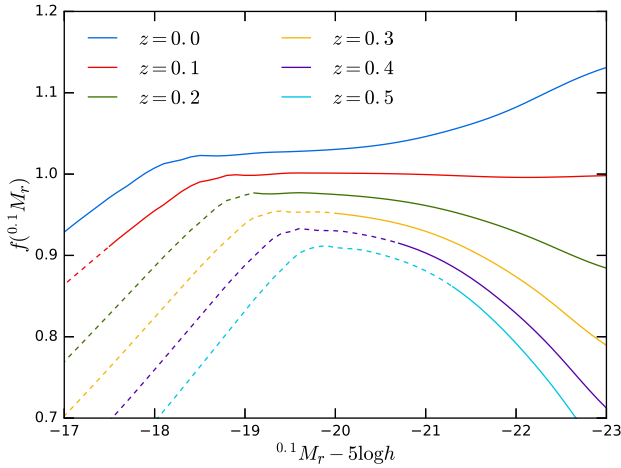


Figure 6. Evolution parameter, f , as a function of magnitude for different redshifts, as indicated by the colour. This is the factor by which the HOD mass parameters are multiplied in order to achieve the galaxy number density set by the target luminosity function. Dashed lines indicate absolute magnitudes that correspond to apparent magnitudes that are fainter than the $r = 20$ limit at that redshift.

redshift as used for the k -corrections (Section 4.3). The faint-end slope is kept constant with redshift since there is not enough data to constrain it at high redshifts. We use the evolving Schechter function, ϕ_{GAMA} , from Loveday et al. (2012) with $P = 1.8$ and $Q = 0.7$.

However, the shape of the GAMA Schechter luminosity function is slightly different from the SDSS luminosity function. Using it as the target at all redshifts would result in the evolution parameter $f \neq 1$ at $z = 0.1$, meaning that the HODs would change from the HODs measured from SDSS. In order not to change the HODs at $z = 0.1$, we use the luminosity function from SDSS, ϕ_{SDSS} , as the target at low redshifts. The SDSS target luminosity function we use is the result of the integral

$$\phi_{\text{SDSS}}(>L) = \int n_{\text{halo}}(M) \langle N(>L|M) \rangle dM, \quad (16)$$

where $n_{\text{halo}}(M)$ is the number density of haloes at $z = 0.1$, and $\langle N(>L|M) \rangle$ is the (unevolved) occupation function. The result of this integral is close to the Blanton et al. (2003) luminosity function for absolute magnitudes brighter than $^{0.1}M_r = -19$, and by definition $f = 1$, so the HODs remain unchanged from SDSS at this redshift. However, at magnitudes fainter than $^{0.1}M_r = -19$, the result of this integral is very flat, while the Blanton et al. (2003) SDSS luminosity function is steeper; at $^{0.1}M_r = -17$ they differ by a factor of ~ 2 . We therefore smoothly transition to the Blanton et al. (2003) luminosity function at $^{0.1}M_r = -19$. This is then extrapolated to fainter magnitudes with a power law.

We interpolate the target luminosity function from ϕ_{SDSS} at low redshifts to ϕ_{GAMA} at high redshifts,

$$\phi_{\text{target}}(M, z) = (1 - w(z))\phi_{\text{SDSS}}(M, z) + w(z)\phi_{\text{GAMA}}(M, z), \quad (17)$$

where the transition between $0.1 < z < 0.2$ is set by the sigmoid function

$$w(z) = (1 + e^{-100(z-0.15)})^{-1}. \quad (18)$$

The evolution parameter, f , for this target luminosity function is shown in Fig. 6 as a function of magnitude for different redshifts. At $z = 0.1$, f is close to 1, by definition. However, it is not exactly

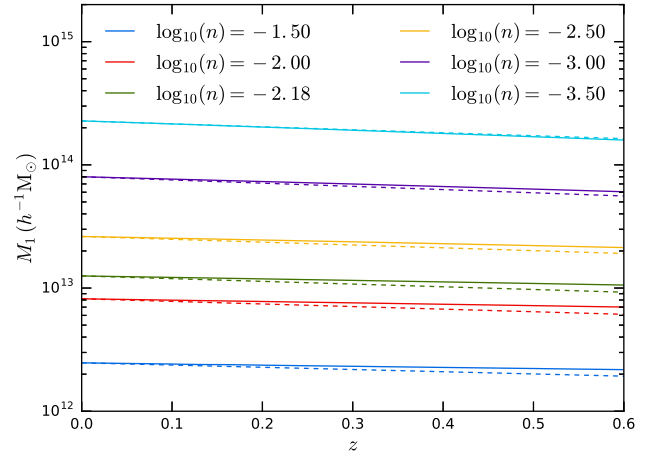


Figure 7. Evolution of the HOD parameter M_1 with redshift for galaxy samples of a fixed number density, where number densities, n , are in units of $h^3 \text{Mpc}^{-3}$. Solid lines show the evolution in the mock, as determined from the target luminosity function. Dashed lines start at the same $M_1(z = 0)$ as in the mock, but show the evolution found in Contreras et al. (2017), as predicted from the Gonzalez-Perez et al. (2014) version of the GALFORM semi-analytic galaxy formation model.

1 because the function $w(z)$, which sets the transition between the two target luminosity functions is close to, but not exactly 0 at $z = 0.1$. At $z = 0.1$, f is equal to 1 to within 1 per cent. Fainter than magnitude -19 , $f(z = 0.1) < 1$. At these faint magnitudes, the target luminosity function is transitioning to the Blanton et al. (2003) luminosity function. Keeping $f(z = 0.1) = 1$ at all magnitudes produces a luminosity function that, while being close to SDSS at the bright end, is too flat at the faint end, so this transition is required to bring the luminosity function of the mock into better agreement with the data.

The evolution of the parameter M_1 implied by this evolution of f is shown in Fig. 7 for galaxy samples of a fixed number density, up to $z = 0.6$. Since the shape of the HODs are kept fixed for a fixed number density, but the HODs are evolved along the mass axis, the other mass parameters M_{min} and M_0 show the same evolution, while $\sigma_{\log M}$ and α are held constant. For comparison, we also show the evolution reported in Contreras et al. (2017) from their fit to the evolution found in the Gonzalez-Perez et al. (2014) version of the GALFORM semi-analytic galaxy formation model (Cole et al. 2000). We find that M_1 decreases slightly with redshift, in remarkably close agreement with what is found in Contreras et al. (2017), although the highest number density samples show slightly less evolution. By construction, the ratio of the parameters M_1/M_{min} is kept constant in the mock. This is in contrast to the behaviour found in Contreras et al. (2017), where they reported that this ratio decreases over the same redshift range. The amount by which this ratio decreases depends on the semi-analytic model used, and on the number density of galaxies; at most it decreases by ~ 50 per cent. The evolution of the HOD model could be extended to include this change in M_1/M_{min} , but we find that simply keeping the mass ratio fixed produces a good match to the measured clustering (see Fig. 12).

4 MOCK GALAXY CATALOGUE

Now we describe in detail the HOD method used to populate the halo catalogue with galaxies, assign each galaxy a luminosity and

$^{0.1}(g-r)$ colour and compare the resultant clustering in the mock with measurements from SDSS and GAMA. These details can be skipped by the reader, but we give a brief summary below.

Section 4.1 describes the HOD method for populating the halo lightcone with galaxies with luminosities. This Monte Carlo method is based on Skibba et al. (2006), but extended to an evolving five-parameter HOD. The number and luminosity of galaxies in each halo are randomly generated such that the input HODs are reproduced. Central galaxies are assigned the position and velocity of the halo, and satellites are randomly positioned around the central, following a Navarro–Frenk–White (NFW) density profile, and assigned a random virial velocity.

The method for assigning a $^{0.1}(g-r)$ colour to each galaxy is described in Section 4.2. This is based on Skibba & Sheth (2009), and randomly assigns a colour from a parametrization of the SDSS colour–magnitude diagram. Section 4.2 describes our modification to the parametrization of the colour–magnitude diagram, which includes evolution, and is in agreement with measurements from GAMA. The colour assigned to each galaxy depends only on its luminosity, its redshift and whether it is a central or satellite galaxy; there is no explicit dependence on halo mass. Colour-dependent k -corrections derived from GAMA are described in Section 4.3. The colour-dependent clustering of galaxies in the mock is shown in Section 4.4.

4.1 Constructing the galaxy catalogue

We use a modified version of the method of Skibba et al. (2006) to populate the halo lightcone catalogue with galaxies, and to assign each galaxy an r -band absolute magnitude, k -corrected to $z = 0.1$. Skibba et al. (2006) use a three-parameter HOD in which the occupation function of central galaxies is simply a step function; we have extended this method in order to reproduce the five-parameter HOD given by equations (6) and (10), which adds scatter to the luminosity of central galaxies, as required by the SDSS clustering data. We also use the fits to the HOD parameters as a function of luminosity as described in Section 3.1 and shown in Fig. 4. To be consistent with the mass definition used in Zehavi et al. (2011), we take the halo mass to be M_{200m} , i.e. the mass enclosed by a sphere in which the average density is 200 times the mean density of the Universe.

For each halo, a number, x , is randomly drawn from the spline kernel probability distribution, $S(x)$ (equation 9), with $\mu = 0$ and $\sigma = 1$. This introduces the scatter in the luminosity of the central galaxy, relative to the average luminosity in a halo of this mass. The luminosity L that is required to produce this scatter is found by solving $x\sigma_{\log M}(L)/\sqrt{2} = \log M - \log M_{\min}(L)$, where the factor of $\sqrt{2}$ comes from how $\sigma_{\log M}$ is defined. Finally, the central galaxy is positioned at the centre of the halo, with the same velocity.

To populate a halo with satellite galaxies, a minimum luminosity, L_{\min} , must first be chosen. We vary L_{\min} with redshift, choosing it to be slightly fainter than the luminosity corresponding to $r = 20$. This ensures that the final mock catalogue is complete to $r = 20$ at all redshifts, while preventing galaxies that are too faint to be observed being unnecessarily added to the catalogue. The number of satellite galaxies to be added to each halo is drawn from a Poisson distribution with mean $\langle N_{\text{sat}}(>L_{\min}|M) \rangle$, which is given by equation (6). For each satellite, a uniform random number $0 < u < 1$ is drawn, and the luminosity is found such that $\langle N_{\text{sat}}(>L|M) \rangle / \langle N_{\text{sat}}(>L_{\min}|M) \rangle = u$. The satellite galaxies are assigned a random virial velocity, rel-

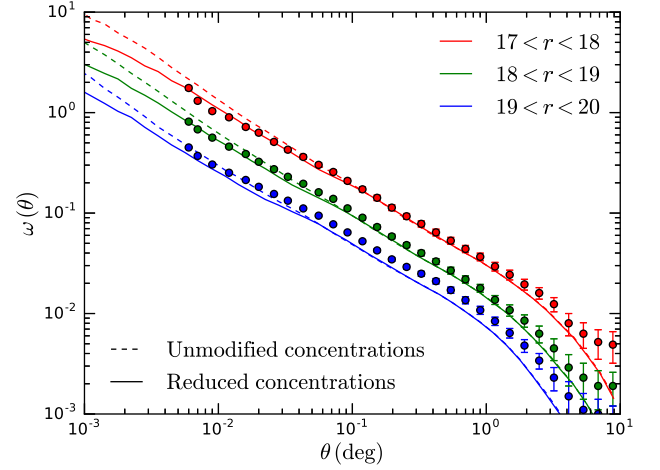


Figure 8. Angular clustering of galaxies in the mock catalogue in bins of apparent magnitude, as labelled (coloured lines). Points with error bars show the angular clustering of galaxies measured in the SDSS (Wang et al. 2013). Dashed curves show the angular clustering where satellite galaxies are positioned such that they follow an NFW density profile with the same, unmodified concentration as the halo. Solid curves show the resulting angular clustering when halo concentrations are reduced by a factor of 2.

ative to the velocity of the central galaxy, which is drawn from a Maxwell–Boltzmann distribution with a line-of-sight velocity dispersion:

$$\sigma^2(M) = \frac{GM_{200m}}{2R_{200m}}, \quad (19)$$

where R_{200m} is the radius of the sphere, centred on the halo, in which the enclosed density is 200 times the mean density of the Universe. Finally, the satellite galaxies are positioned randomly around the centre of the halo such that they follow an NFW (Navarro, Frenk & White 1997) density profile, which is truncated at R_{200m} . We find that using the same concentration, c , as the halo, calculated from $c = 2.16R_{200m}/R_{V_{\max}}$, where $R_{V_{\max}}$ is the radius at which the maximum circular velocity occurs, produces angular clustering that is too strong at small angular scales compared to SDSS (Wang, Brunner & Dolence 2013). This can be improved by reducing the concentration of all haloes by a factor of 2 (see Fig. 8). We therefore use these reduced concentrations when positioning satellite galaxies inside each halo.

The HODs from Zehavi et al. (2011) were fit to the projected correlation functions, using the mass–concentration relation of Bullock et al. (2001), modified to be consistent with their mass definition. This mass–concentration relation is close to what is seen in MXXL. However, modifying the concentrations only has a small effect on the one-halo term of the projected correlation functions. Down to separations of $0.1 h^{-1}$ Mpc, the clustering in the mock catalogue only changes by a small amount. It is the change in the clustering at physical scales smaller than this which causes the small-scale angular clustering to improve, and this is below the scale at which the projected correlation functions were measured in SDSS. The angular correlation function, $\omega(\theta)$, in the mock catalogue is shown in Fig. 8 for galaxies in bins of apparent magnitude, compared to the angular clustering measured in SDSS (Wang et al. 2013). Solid lines show the clustering in the mock with concentrations reduced by a factor of 2, which is in good agreement with SDSS down to a small angular separation of 20 arcsec, although for the faintest sample the clustering is a little low. Using unmodified concentrations result in $\omega(\theta)$ having a slope that is steeper than the SDSS measurements,

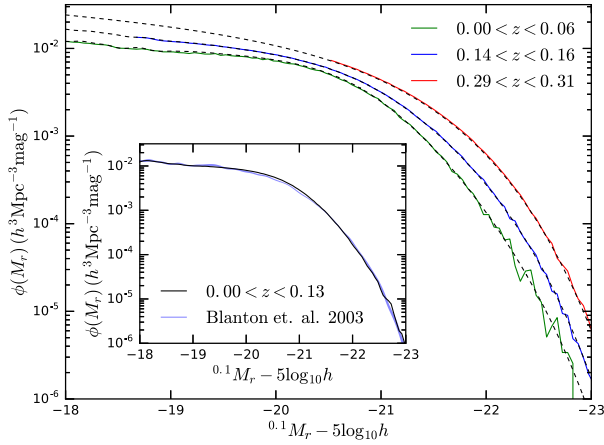


Figure 9. The r -band luminosity function of galaxies in the mock catalogue in different redshift bins, as indicated by the legend. Dashed lines indicate the target luminosity function at the median redshift in each bin, which transitions from the SDSS luminosity function at $z < 0.1$ to the GAMA luminosity function at $z > 0.2$. The smaller panel shows the luminosity function in the mock catalogue over the redshift range $0 < z < 0.13$, compared to the SDSS luminosity function of Blanton et al. (2003).

shown by the dashed curves, resulting in clustering that is too strong at small angular scales. The introduction of unclustered haloes below the mass resolution has the effect of reducing the clustering in the mock, but as we show later in Section 4.1.3, this effect is small.

4.1.1 The luminosity function of the mock

The Petrosian r -band luminosity function of the galaxy catalogue is shown in Fig. 9 for galaxies in three redshift bins. The dashed lines show the target luminosity at the median redshift of each bin, showing that this evolving target luminosity function is reproduced in the mock catalogue. The smaller panel in Fig. 9 compares the luminosity function in the mock at low redshifts with the Blanton et al. (2003) luminosity function from SDSS. Brighter than $^{0.1}M_r = -19$, the luminosity function in the mock is in good agreement with SDSS, which indirectly shows that the mass function of the MXXL lightcone is close to the Jenkins et al. (2001) mass function assumed by Zehavi et al. (2011), and our fits to the HOD parameters as a function of luminosity are a good approximation to the actual values. Fainter than $^{0.1}M_r = -19$, the luminosity functions agree by construction.

4.1.2 The redshift distribution of the mock

The redshift distribution of galaxies brighter than an apparent magnitude limit of $r = 19.8$ is shown in Fig. 10 (see Section 4.3 for the k -corrections used), and compared to the GAMA survey. The dN/dz of the mock catalogue is in good agreement with GAMA, within 15 per cent of the fitted curve at most redshifts. Without adding in the low mass, unresolved haloes at low redshifts (see Section 2.5), there is a deficit in the dN/dz for $z \lesssim 0.2$ (dashed red curve); adding in these haloes increases the number of low-redshift haloes, bringing the dN/dz into better agreement with GAMA (solid red curve).

4.1.3 Clustering of the mock

Projected correlation functions of galaxies in the mock catalogue are shown by the solid curves in Fig. 11 for different luminosity

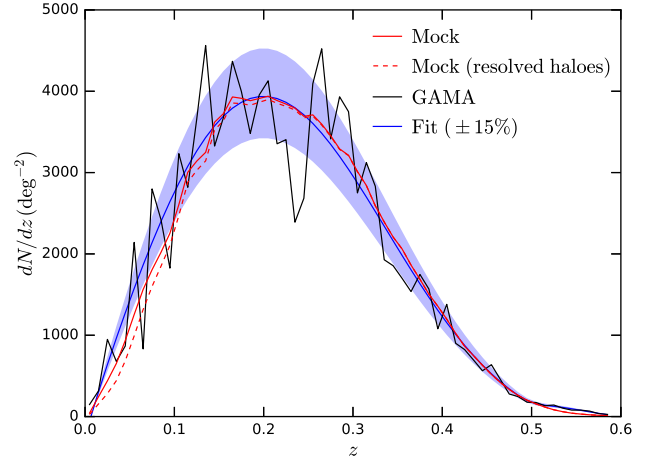


Figure 10. dN/dz of galaxies in the mock catalogue with $r < 19.8$ (red), compared to GAMA (black). The solid red curve shows the redshift distribution of all galaxies, including those residing in unresolved haloes below the MXXL mass resolution, while the dashed red curve only includes galaxies residing in resolved haloes. The blue curve shows a fit to the GAMA dN/dz , where the shaded region indicates ± 15 per cent.

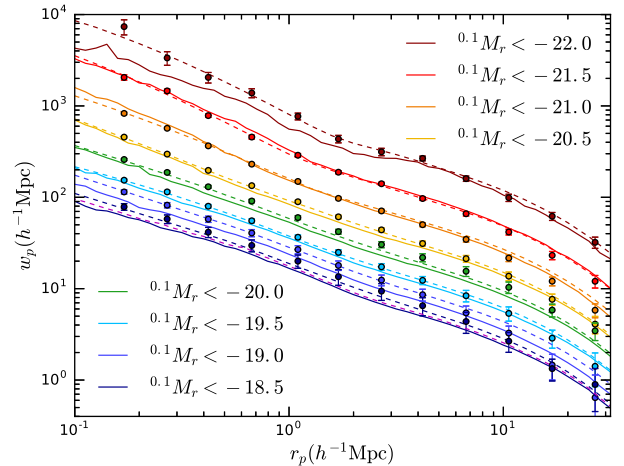


Figure 11. Projected correlation functions from the galaxy catalogue (solid lines), compared to the projected correlation functions from SDSS (Zehavi et al. 2011) (points with error bars) and the projected clustering predicted using the best-fitting HODs in Millennium cosmology (dashed lines), for different luminosity threshold samples, as indicated by the legend. For the $^{0.1}M_r < -18.5$ sample, we also show the projected clustering in the galaxy catalogue omitting all galaxies that reside in unresolved, unclustered haloes (magenta dashed line). For clarity, the results have been offset by successive intervals of 0.15 dex, starting at the $^{0.1}M_r < -20.5$ sample.

threshold samples at $z \sim 0.1$, where we have calculated the two-point correlation functions using the publicly available code CUTE (Alonso 2012).³ These are compared to the measured clustering from SDSS (points with error bars), and the clustering predicted by the best-fitting HODs (dashed lines). We use the same redshift ranges as the SDSS volume-limited luminosity threshold samples (see table 2 in Zehavi et al. 2011). To be consistent with the definition of magnitude used in Zehavi et al. (2011), magnitudes are evolved to $z = 0.1$ using the evolution model $E(z) = Q_0(1 + Q_1(z - z_0))(z - z_0)$, where $Q_0 = 2$, $Q_1 = -1$ and $z_0 = 0.1$. The clustering in

³ <http://members.ift.uam-csic.es/dmonge/CUTE.html>

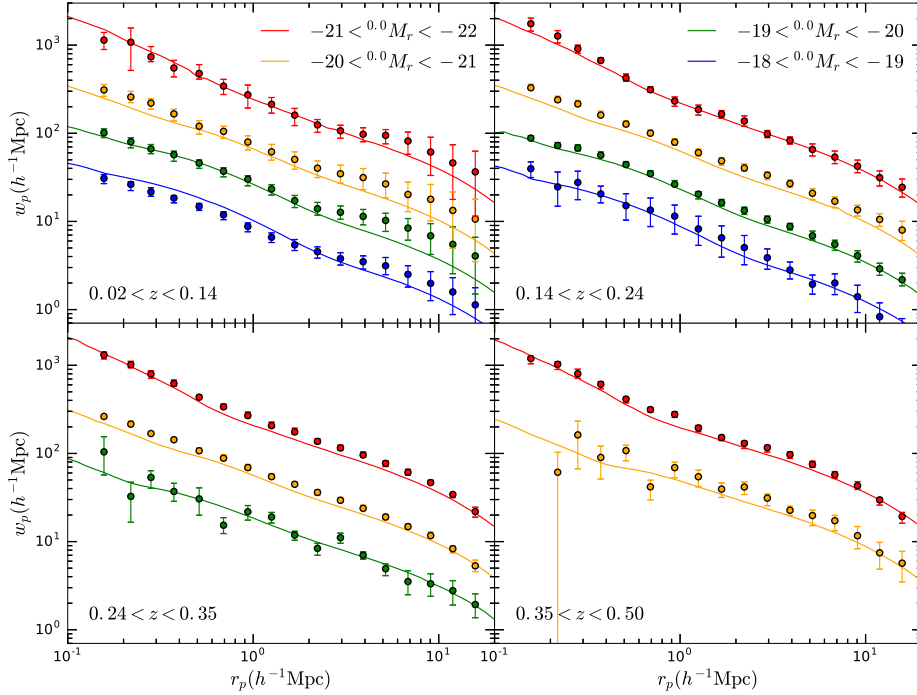


Figure 12. Projected correlation functions in different redshift bins for galaxies in the mock catalogue (lines). Points with error bars show the clustering of galaxies from GAMA (Farrow et al. 2015). Different colours indicate bins in $^{0.0}M_r$ absolute magnitude. Lines are offset by 0.4 dex relative to the $-21 < ^{0.0}M_r < -20$ samples, for clarity.

our galaxy catalogue is in reasonable agreement with the projected correlation functions measured from SDSS.

The small differences in the large-scale clustering between the mock catalogue and the clustering predicted by the best-fitting HODs can be understood by comparing the HODs in Fig. 5. For example, the $^{0.1}M_r < -19$ sample is slightly less clustered in the mock. The fit to the HOD has a smaller M_{\min} than the best-fitting SDSS HOD, meaning that this sample contains more low-mass haloes. These haloes are less biased, and therefore the clustering is reduced compared to SDSS. The $^{0.1}M_r < -22$ sample contains more high-mass haloes, and should therefore be more clustered than SDSS, but the opposite is seen. This is because the brightest samples cover a wider redshift range, and are affected more by the evolution of the HODs.

The clustering of galaxies in the mock catalogue is also affected by the introduction of haloes below the MXXL mass resolution, which are unclustered. Adding these haloes will therefore have the effect of reducing the measured galaxy clustering. The galaxies that reside in these haloes are faint, and have low redshifts, and so the faintest galaxy samples in Fig. 5 are affected by this more than the bright samples. For the $^{0.1}M_r < -18.5$ sample, we illustrate the size of this effect: the magenta dashed curve shows the projected correlation function with galaxies residing in unresolved haloes omitted. Including these galaxies reduces the clustering, but only by a very small amount.

We have checked that if we modify our fits to the HOD parameters to agree exactly with the best-fitting SDSS parameters for one of the magnitude thresholds, and do not evolve the HODs, we reproduce the SDSS correlation functions very closely for that magnitude limit.

In Fig. 12 we show the projected correlation functions in the mock catalogue at high redshifts, compared to the clustering measured in GAMA by Farrow et al. (2015), which has a high completeness of

galaxy pairs (Robotham et al. 2010). Here, the magnitude ranges are defined for magnitudes k -corrected to a reference redshift of $z = 0$, denoted as $^{0.0}M_r$, which have also had evolutionary corrections applied. To be consistent with Farrow et al. (2015), we use the same evolutionary correction $E(z) = -Q(z - z_{\text{ref}}) = -1.45z$ for this comparison. We find that the clustering at high redshifts is in good agreement with the clustering seen in GAMA.

4.2 Assigning colours

We use the method of Skibba & Sheth (2009) to assign each galaxy a $^{0.1}(g - r)$ colour, where g and r are SDSS Data Release 7 (DR7) model magnitudes (Abazajian et al. 2009; Baldry et al. 2010). This method parametrizes the red and blue sequence of the colour-magnitude diagram as two Gaussians with a mean and rms that are linear functions of magnitude. A galaxy is randomly chosen to be red or blue, then a colour is drawn from the appropriate Gaussian. We have modified the parametrization of the colour-magnitude diagram given in Skibba & Sheth (2009) to bring the faint end into agreement with the colour-magnitude diagram from GAMA, and to add evolution. For those interested, this is described in detail below. For clarity, the 0.1 superscript has been omitted from the equations in the following subsections.

4.2.1 Low redshift

For redshifts $z < 0.1$, we use the parametrization used by Skibba & Sheth (2009) that produces a good approximation to the SDSS colour-magnitude diagram, which we summarize below. However, we make some slight modifications to the parametrization to also bring it into agreement with GAMA at faint magnitudes.

The mean and rms of the red and blue sequences are given by

$$\langle g - r | M_r \rangle_{\text{red}}^{\text{Skibba}} = 0.932 - 0.032(M_r + 20),$$

$$\text{rms}(g - r | M_r)_{\text{red}}^{\text{Skibba}} = 0.07 + 0.01(M_r + 20) \quad (20)$$

and

$$\langle g - r | M_r \rangle_{\text{blue}}^{\text{Skibba}} = 0.62 - 0.11(M_r + 20),$$

$$\text{rms}(g - r | M_r)_{\text{blue}}^{\text{Skibba}} = 0.12 + 0.02(M_r + 20). \quad (21)$$

The total fraction of galaxies that are blue is also parametrized as a linear function of magnitude, given by

$$f_{\text{blue}}^{\text{Skibba}}(M_r) = 0.46 + 0.07(M_r + 20). \quad (22)$$

These relations from Skibba & Sheth (2009) produce a colour–magnitude diagram that is in good agreement with SDSS at the bright end. However, the faint end does not agree with what is seen in GAMA (e.g. the first panel in fig. 6 of Loveday et al. 2012). First, at $^{0.1}M_r = -16$, all the galaxies should lie on the blue sequence, while the fraction of galaxies that are blue given by equation (22) is 0.74. At faint magnitudes, we instead use a blue fraction given by

$$f_{\text{blue}}^{\text{faint}}(M_r) = 0.4 + 0.2(M_r + 20), \quad (23)$$

so the total fraction of blue galaxies is

$$f_{\text{blue}}(M_r) = \max\{f_{\text{blue}}^{\text{faint}}(M_r), f_{\text{blue}}^{\text{Skibba}}(M_r)\}; \quad (24)$$

$f_{\text{blue}}(M_r)$ is capped so it is always in the range $0 \leq f_{\text{blue}}(M_r) \leq 1$. Another issue with the parametrization of Skibba & Sheth (2009) is that faint galaxies that lie on the blue sequence are too blue in comparison to the galaxies in GAMA. At $^{0.1}M_r = -18.7$, we transition to a flatter blue sequence, given by

$$\langle g - r | M_r \rangle_{\text{blue}}^{\text{faint}} = 0.4 - 0.03(M_r + 16). \quad (25)$$

If the fraction of satellite galaxies that are blue, $f_{\text{sat}}^{\text{blue}}(M_r)$, is specified, then the mean colour of satellite galaxies is given by

$$\langle g - r | M_r \rangle_{\text{sat}} = f_{\text{sat}}^{\text{blue}}(M_r) \langle g - r | M_r \rangle_{\text{blue}} + (1 - f_{\text{sat}}^{\text{blue}}(M_r)) \langle g - r | M_r \rangle_{\text{red}}. \quad (26)$$

Conversely, equation (26) can be rearranged, and the mean satellite colour can be used to specify the fraction of satellites that are blue,

$$f_{\text{sat}}^{\text{blue}}(M_r) = \frac{\langle g - r | M_r \rangle_{\text{sat}} - \langle g - r | M_r \rangle_{\text{red}}}{\langle g - r | M_r \rangle_{\text{blue}} - \langle g - r | M_r \rangle_{\text{red}}} \quad (27)$$

(equation 8 from Skibba & Sheth 2009, but for blue galaxies). The average colour of a satellite galaxy is parametrized by Skibba & Sheth (2009) as

$$\langle g - r | M_r \rangle_{\text{sat}}^{\text{Skibba}} = 0.83 - 0.08(M_r + 20). \quad (28)$$

Modifying the mean satellite colour has the effect of changing the strength of the colour-dependent clustering. We find that we get a better agreement with the clustering in SDSS by modifying the mean satellite colour to

$$\langle g - r | M_r \rangle_{\text{sat}} = 0.86 - 0.065(M_r + 20). \quad (29)$$

At, for example, $^{0.1}M_r = -16$, the fraction of blue satellites given by equations (27) and (29) is less than 1, meaning that some satellite galaxies are red. However, all galaxies at this magnitude should lie on the blue sequence, as determined from equation (24). In order to achieve the correct f_{blue} from equation (24), the fraction of satellites that are blue must be $(f_{\text{blue}} - f_{\text{cen}})/f_{\text{sat}}$ if all central galaxies are on the blue sequence. If the value of $f_{\text{sat}}^{\text{blue}}$ is greater than this, it is still possible to get the correct f_{blue} by making central galaxies red, but $f_{\text{sat}}^{\text{blue}}$ cannot be less than this. To ensure that at faint magnitudes we

get the total fraction of blue galaxies given by equation (24), we take the fraction of blue satellites to be

$$f_{\text{sat}}^{\text{blue}}(M_r) = \max \left\{ f_{\text{sat}}^{\text{blue}}(M_r), \frac{f_{\text{blue}}(M_r) - f_{\text{cen}}(M_r)}{f_{\text{sat}}(M_r)} \right\}. \quad (30)$$

The fraction of central galaxies that are blue can then be determined from $f_{\text{blue}}(M_r)$ and $f_{\text{sat}}^{\text{blue}}(M_r)$. However, Skibba & Sheth (2009) erroneously state that the fraction of central galaxies that are blue is

$$f_{\text{cen}}^{\text{blue}}(M_r) = f_{\text{blue}}(M_r)/f_{\text{cen}}(M_r), \quad (31)$$

where $f_{\text{cen}}(M_r)$ is the fraction of galaxies that are centrals. Equation (31) is only true if all satellite galaxies are red; since a significant fraction of faint satellites is blue, the fraction of blue central galaxies needs to be reduced to ensure we get the correct total fraction of blue galaxies given by equation (22). This is achieved by changing equation (31) to

$$f_{\text{cen}}^{\text{blue}}(M_r) = \frac{f_{\text{blue}}(M_r) - f_{\text{sat}}^{\text{blue}}(M_r)(1 - f_{\text{cen}}(M_r))}{f_{\text{cen}}(M_r)}. \quad (32)$$

For each galaxy, a uniform random number x is drawn in the interval $0 < x < 1$. For central galaxies, if $x < f_{\text{cen}}^{\text{blue}}(M_r)$ (given by equation 32), the galaxy is blue, and a colour is drawn randomly from the Gaussian distribution defined by equation (21), otherwise it is red, and the colour is drawn from equation (20). Similarly, satellite galaxies are assigned to the blue sequence if $x < f_{\text{sat}}^{\text{blue}}(M_r)$, and the red sequence otherwise.

4.2.2 Evolution of colours with redshift

The colour–magnitude diagram evolves with redshift, as seen for example in fig. 6 of Loveday et al. (2012) from GAMA. We therefore need to evolve the expressions given in Section 4.2.1 in order to produce a mock that has a realistic distribution of colours at these redshifts. In the GAMA data at high redshifts, only the brightest tip of the red and blue sequences can be seen, making it difficult to constrain their slopes. We therefore keep the slope of the red and blue sequence fixed with redshift.

We keep the red and blue sequences fixed at $z < 0.1$, and evolve them with redshift as

$$\begin{aligned} \langle g - r | M_r \rangle_{\text{red}}(z) &= \langle g - r | M_r \rangle_{\text{red}} - 0.18(\min\{z, 0.4\} - 0.1), \\ \text{rms}(g - r | M_r)_{\text{red}}(z) &= \text{rms}(g - r | M_r)_{\text{red}} \\ &\quad + 0.5(z - 0.1) + 0.1(z - 0.1)^2 \end{aligned} \quad (33)$$

and

$$\begin{aligned} \langle g - r | M_r \rangle_{\text{blue}}(z) &= \langle g - r | M_r \rangle_{\text{blue}} - 0.25(\min\{z, 0.4\} - 0.1), \\ \text{rms}(g - r | M_r)_{\text{blue}}(z) &= \text{rms}(g - r | M_r)_{\text{blue}} + 0.2(z - 0.1), \end{aligned} \quad (34)$$

respectively, where we stop evolving the mean of the sequences above $z = 0.4$ in order to prevent too many high-redshift galaxies being assigned as blue.

The mean satellite colour is also evolved as

$$\langle g - r | M_r \rangle_{\text{sat}}(z) = \langle g - r | M_r \rangle_{\text{sat}} - 0.18(z - 0.1), \quad (35)$$

and the fraction of blue galaxies is evolved as

$$f_{\text{blue}}(M_r)(z) = 0.2M_r + 4.4 + 1.2(z - 0.1) + 0.5(z - 0.1)^2 \quad (36)$$

Fig. 13 shows the distribution of colours in the mock catalogue compared to GAMA for galaxies in different redshift and magnitude bins. Our parametrization of the colour evolution is able to

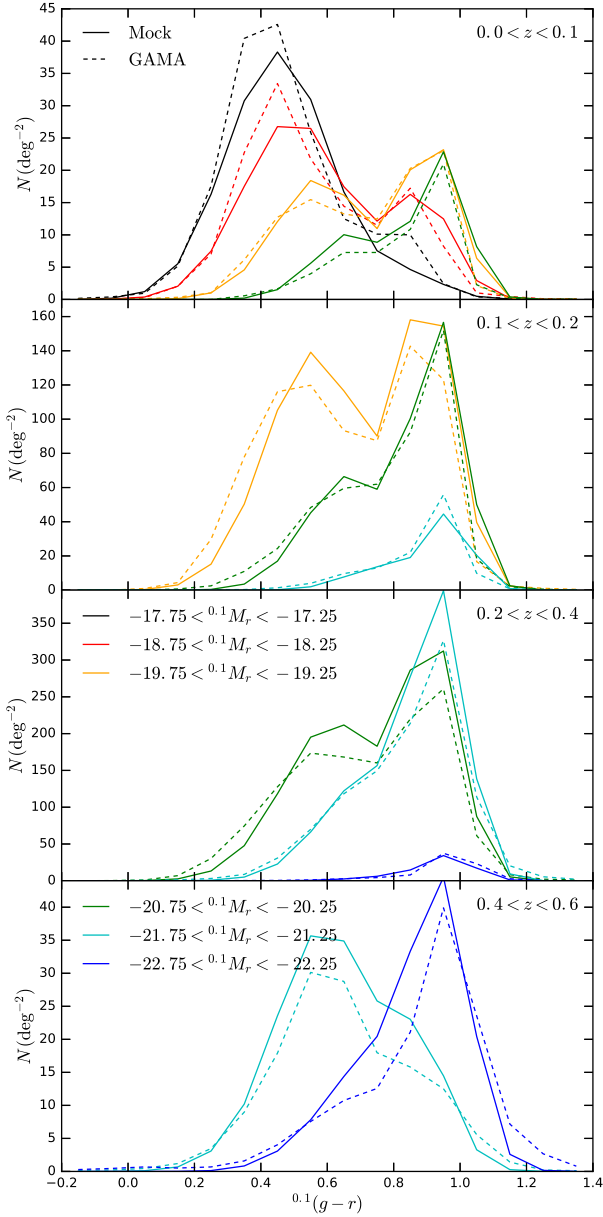


Figure 13. Distribution of $^{0.1}(g-r)$ colours in the mock catalogue (solid lines) compared to GAMA (dashed lines). Each panel shows the colour distributions of galaxies in a certain redshift range. Different ranges in absolute magnitude are indicated by the colour of the line, as shown in the legend, which is split over several panels.

produce a good approximation to the GAMA colour distributions at all redshifts.

To evolve the luminosity function, we have assumed a fixed Q parameter for all galaxies. We note that Loveday et al. (2012, 2015) hint that red and blue galaxies evolve differently, with different Q_{red} and Q_{blue} . However, the assumption of fixed Q with this parametrization is able to reproduce the observed colour-magnitude diagram.

4.3 Colour-dependent k -corrections

In the mock catalogue, we use the HOD method to assign each galaxy an r -band absolute magnitude $^{0.1}M_r$, and the method outlined above to randomly generate a $^{0.1}(g-r)$ colour. However, the apparent magnitude, r , is the quantity that would be measured di-

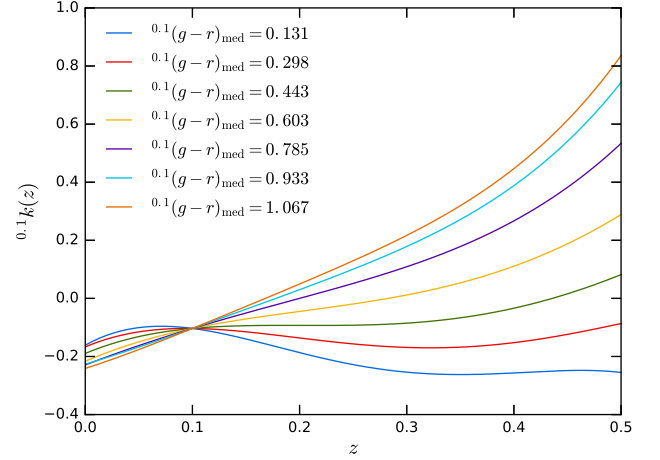


Figure 14. Median $^{0.1}(g-r)$ colour-dependent k -correction for galaxies in GAMA as a function of redshift, in seven equally spaced bins of colour. The colour of each line indicates the colour bin; the median colour is indicated in the legend.

rectly by the survey, and this is related to the absolute magnitude, $^{0.1}M_r$, through the equation

$$^{0.1}M_r - 5 \log_{10} h = r - 5 \log_{10}(d_L(z)) - 25 - ^{0.1}k(z), \quad (37)$$

where $d_L(z)$ is the luminosity distance in units of h^{-1} Mpc, and $^{0.1}k(z)$ is the k -correction. The superscript 0.1 denotes that the magnitude has been k -corrected to a reference redshift of $z_{\text{ref}} = 0.1$. In order to calculate an apparent magnitude for each galaxy in the mock catalogue, we use colour-dependent k -corrections derived from the GAMA survey, similar to those given in table 1 of McNaught-Roberts et al. (2014), except for k -correcting to $z_{\text{ref}} = 0.1$, rather than $z_{\text{ref}} = 0$.

The k -correction for each individual galaxy in GAMA is fit with a fourth-order polynomial of the form

$$^{0.1}k(z) = \sum_{i=0}^4 A_i (z - 0.1)^{4-i}. \quad (38)$$

The median k -correction is then found in seven equally spaced bins of $^{0.1}(g-r)$ colour. Strictly speaking, the constant term in equation (38) should have the value $A_4 = -2.5 \log_{10}(1 + z_{\text{ref}})$ (Hogg et al. 2002; McNaught-Roberts et al. 2014) do not require this, but they end up with values of A_4 close to 0 at their $z_{\text{ref}} = 0$. We force our k -corrections to have the value $A_4 = -2.5 \log_{10}(1.1)$ at our $z_{\text{ref}} = 0.1$, but this only has a small effect on the k -corrections.

Using seven distinct k -corrections based on colour leads to artificial features being added to the mock catalogue; for example step features can be seen in the colour-magnitude diagram at the boundaries between colour bins. In order to remove these features, we interpolate the k -corrections between the median colour in each bin.

Fig. 14 shows the polynomial fits to the k -corrections as a function of redshift. By definition, all the curves cross at $^{0.1}k(z = 0.1) = -2.5 \log_{10}(1.1) \approx -0.103$. The polynomial coefficients are shown in Table 1.

4.4 Colour-dependent clustering in the mock

The projected correlation function of galaxies in the mock at low redshifts, split by red and blue galaxies, is shown in Fig. 15 for different bins in absolute magnitude and compared to the clustering

Table 1. Polynomial coefficients of the median k -corrections of galaxies in GAMA in equally spaced bins of $^{0.1}(g-r)$ colour, as defined in equation (38). $^{0.1}(g-r)_{\text{med}}$ is the median colour in each bin, and A_i are the polynomial coefficients. The constant term $A_4 = -2.5 \log_{10}(1.1) \approx -0.103$, as described in the text.

$^{0.1}(g-r)_{\text{med}}$	A_0	A_1	A_2	A_3
0.131	-45.33	35.28	-6.604	-0.4805
0.298	-20.08	20.14	-4.620	-0.04824
0.443	-10.98	14.36	-3.676	0.3395
0.603	-3.428	9.478	-2.703	0.7646
0.785	6.717	3.250	-1.176	1.113
0.933	16.76	-2.514	0.3513	1.307
1.067	20.30	-4.189	0.5619	1.494

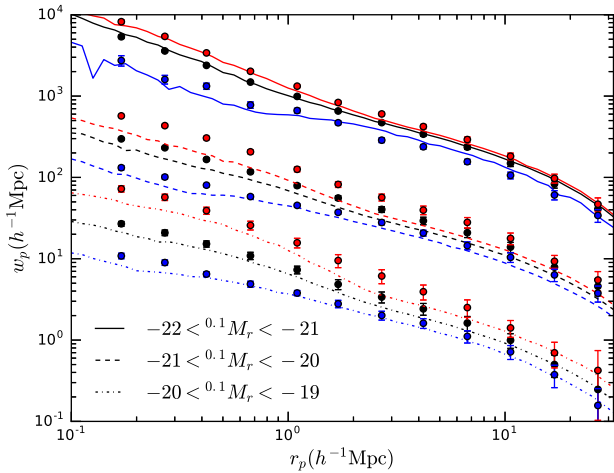


Figure 15. Projected correlation functions of red and blue galaxy samples in the mock catalogue at low redshifts (lines) compared to the SDSS volume-limited samples (points with error bars; Zehavi et al. 2011) for different magnitude bins. The clustering of all galaxies in a sample is shown in black, while clustering for red and blue galaxies, defined by the colour cut $^{0.1}(g-r)_{\text{cut}} = 0.21 - 0.03^{0.1}M_r$, is shown in red and blue, respectively. Line style indicates the magnitude bin, as shown by the legend. For clarity, magnitude samples are successively offset by 1 dex from the $-21 < ^{0.1}M_r < -20$ samples.

in the corresponding volume-limited samples from SDSS (Zehavi et al. 2011), where the red and blue samples are defined using the same colour cut as their equation (13). In the SDSS data, red galaxies are clustered more strongly than blue galaxies, since red elliptical galaxies are more likely to reside in more massive haloes, which are more strongly biased (Eisenstein et al. 2005). As the samples get fainter, the strength of the colour dependence becomes stronger. These trends are reproduced in the mock catalogue, using the modified satellite colour in equation (29).

Projected correlation functions for red and blue galaxies are also shown for higher redshift galaxies in Fig. 16, compared with the clustering seen in GAMA (Farrow et al. 2015). The galaxy samples are defined using the same $^{0.0}M_r$ magnitude ranges as fig. 14 of Farrow et al. (2015), and using the same $^{0.0}(g-r)$ colour cut (their equation 4), where the superscript 0.0 denotes that these magnitudes are k -corrected to a reference redshift of $z_{\text{ref}} = 0$. The clustering of the red and blue galaxies in the mock is in reasonable agreement with the GAMA data.

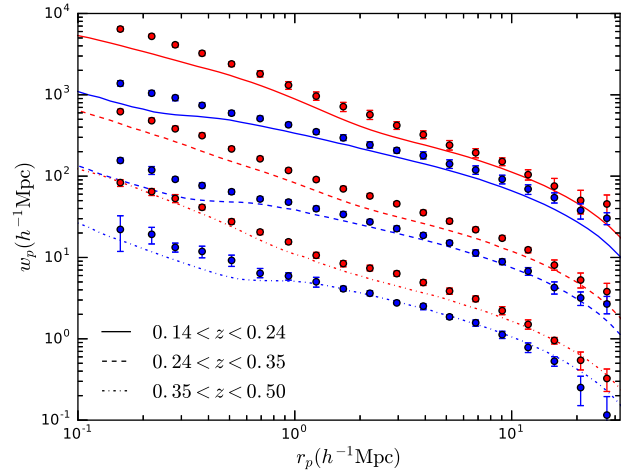


Figure 16. Projected correlation functions of red and blue galaxy samples at high redshift in the mock catalogue (lines), compared to GAMA (points with error bars; Farrow et al. 2015). Red and blue lines indicate red and blue galaxy samples, where the colour cut is defined as $^{0.0}(g-r)_{\text{cut}} = -0.03(^{0.0}M_r + 20.6) + 0.678$. For each sample of galaxies in the mock, the same $^{0.0}M_r$ magnitude range is used as the GAMA galaxy sample. The style of the line indicates the redshift range, as shown in the legend. Redshift samples are successively offset from the $0.24 < z < 0.35$ samples by 1 dex for clarity.

5 APPLICATIONS

As shown in Section 4, the galaxies in the mock catalogue have realistic clustering, which is in agreement with measurements from SDSS and GAMA. The galaxies also have a realistic distribution of $^{0.1}(g-r)$ colours at different redshifts. Future surveys, such as DESI and *Euclid*, aim to make measurements of the BAO and redshift-space distortions, which probe larger scales than have been considered so far. Here, we show as an example some of the measurements that can be made using this mock catalogue.

5.1 BAO

As described in Section 2.1 and shown in Fig. 3, the large box size of the MXXL simulation enables the BAO feature to be seen clearly in the clustering of haloes. Here, we show that the BAO can also be seen in measurements of the redshift-space galaxy clustering. Fig. 17 shows the large-scale redshift-space correlation function for several apparent magnitude threshold galaxy samples, using a redshift weighting $w(z) = 1/(1 + 4\pi J_3 \bar{n}(z))$ (Efsthathiou et al. 1990), where $\bar{n}(z)$ is the number density of galaxies in the sample at redshift z , and $J_3 = \int \xi r^2 dr$, where we have assumed $4\pi J_3 = 3 \times 10^4 h^3 \text{ Mpc}^{-3}$. The BAO peak can be seen in all samples, but the errors in the correlation function are largest for the $r < 18.0$ sample. The $r < 20.0$ sample contains fainter galaxies, and covers a larger volume, which greatly reduces the errors. For comparison, the crosses with error bars show measurements of the BAO from the Baryon Oscillation Spectroscopic Survey (BOSS; Reid et al. 2016) for galaxies in the redshift range $0.2 < z < 0.5$ (Ross et al. 2017). The BAO scale in the mock catalogue is ~ 7 per cent larger than is measured in BOSS. This is consistent with the difference in cosmology between that used in the MXXL simulation and the best fit to observations, including the BOSS results, and is mostly driven by the difference between $\Omega_m = 0.25$ in MXXL and $\Omega_m = 0.31$ in the *Planck* cosmology. The amplitude of the BAO peak in the mock catalogue also

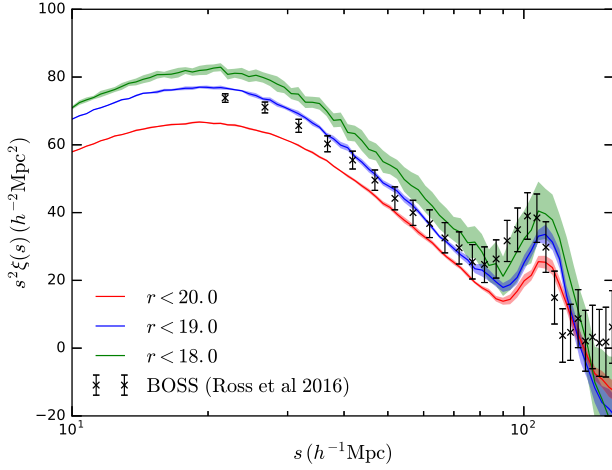


Figure 17. Large-scale redshift-space correlation function in the galaxy catalogue, scaled by s^2 , for different apparent magnitude threshold samples, as indicated by the colour. For each sample, the solid curve is the clustering calculated over the full sky, and the shaded area is the error on the mean of four quadrants. Crosses with error bars indicate the measured clustering from BOSS in the redshift range $0.2 < z < 0.5$ (Ross et al. 2017), divided by a factor of 1.5, to make the comparison easier.

differs with the BOSS results, but we have not made any attempt to match the BOSS colour selection.

5.2 Redshift-space distortions

The two-point correlation function $\xi(s, \mu)$, in bins of s and μ , can be decomposed into multipoles (Hamilton 1992)

$$\xi(s, \mu) = \sum_l \xi_l(s) P_l(\mu), \quad (39)$$

where s is the separation between a pair of galaxies in redshift space, $\mu = \cos \theta$ is the cosine of the angle between the vector s and the line of sight and $P_l(\mu)$ is the l th-order Legendre polynomial. The multipoles can be determined from the measured $\xi(s, \mu)$ by evaluating the integral

$$\xi_l(s) = \frac{2l+1}{2} \int_{-1}^1 \xi(s, \mu) P_l(\mu) d\mu. \quad (40)$$

Because of the symmetry $\xi(s, \mu) = \xi(s, -\mu)$, all odd-numbered terms are zero, and in linear theory, it is only the monopole, $\xi_0(s)$, quadrupole, $\xi_2(s)$ and hexadecapole, $\xi_4(s)$, that are non-zero. The amplitude of these multipoles depends on the strength of the redshift-space distortions, and can provide a way to measure $f(z)\sigma_8(z)$ (Samushia, Percival & Raccanelli 2012).

The multipoles of the redshift-space correlation function of galaxies in the mock catalogue are shown in Fig. 18 for different volume-limited magnitude threshold samples, and compared to measurements of clustering from SDSS (Guo et al. 2015). The monopole and quadrupole show reasonable agreement with the SDSS measurements, although the amplitude of the hexadecapole is a little high for some of the samples. Overall, the redshift-space distortions in the mock catalogue look reasonably realistic, showing that the catalogue will be useful for future surveys that will take redshift-space distortion measurements. We have extended the predictions beyond the range of the SDSS results, where they are easier to model and can be probed by surveys like DESI and *Euclid*.

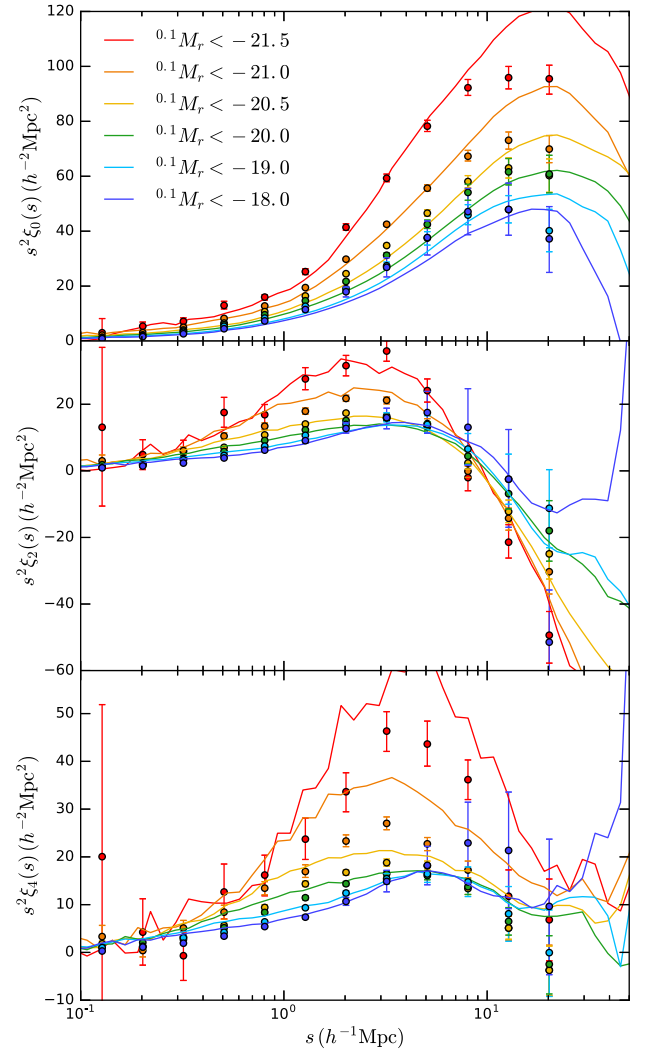


Figure 18. Monopole, $\xi_0(s)$ (top panel), quadrupole, $\xi_2(s)$ (middle panel), and hexadecapole, $\xi_4(s)$ (bottom panel), of the redshift-space two-point correlation function for different volume-limited samples (solid lines), where the colour indicates the magnitude threshold. Points with error bars show the measured clustering from SDSS (Guo et al. 2015).

6 CONCLUSIONS

For upcoming galaxy surveys, such as DESI and *Euclid*, it is important to have realistic mock catalogues in order to test and verify analysis tools, assess incompleteness and determine error covariances. The mock catalogues can also be used to make predictions and set expectations in advance of the first data from the survey.

We have outlined a method for creating a mock catalogue from the MXXL simulation. We first created a halo lightcone catalogue from the simulation, which we then populated with galaxies using a HOD scheme.

The halo lightcone catalogue is created from the simulation by finding the interpolated values of the position, velocity and mass of each halo at the redshift at which it crosses the observer's lightcone. The halo lightcone catalogue covers the full sky, and extends to redshift $z = 2.2$ with a mass resolution of $\sim 10^{11} h^{-1} M_\odot$. Extending the catalogue to high redshifts requires multiple periodic replications of the MXXL box; these replications are only necessary to extend to redshifts greater than $z \sim 0.5$.

The halo catalogue is populated with galaxies using a Monte Carlo method, which randomly assigns galaxies with luminosities to dark matter haloes, following a HOD. This is an extension of the method outlined in Skibba et al. (2006) to a five-parameter HOD. Galaxies in the mock catalogue are also assigned a $^{0.1}(g-r)$ colour, based on the Monte Carlo method of Skibba & Sheth (2009). A galaxy is assigned a colour based on its luminosity, redshift and whether it is a central or a satellite galaxy, which is randomly drawn from a parametrization of the SDSS colour–magnitude diagram.

The values of the HOD parameters we use are based on the best-fitting HODs that reproduce the measured clustering from SDSS (Zehavi et al. 2011), but in Millennium cosmology. In the standard five-parameter HOD model, the parameter $\sigma_{\log M}$ introduces Gaussian scatter in the luminosity of central galaxies for haloes of a fixed mass, which can lead to the unphysical crossing of HODs in different luminosity bins. We modify the HOD model so that this scatter follows a pseudo-Gaussian spline kernel, which prevents this crossing. The HODs are evolved with redshift such that they reproduce a target luminosity function, which is chosen to be the SDSS luminosity function at low redshift, and the luminosity function from GAMA at high redshifts. For a sample of galaxies with a fixed number density, the shape of the HOD is kept fixed with redshift, and the mass HOD parameters are all multiplied by the factor f that is required to produce the correct number density. By construction, the mock catalogue reproduces the SDSS and GAMA luminosity functions, and the ratio of the HOD parameters M_1/M_{\min} is constant with redshift.

We modify the parametrization of the colour–magnitude diagram outlined in Skibba & Sheth (2009), and add evolution, such that the distribution of $^{0.1}(g-r)$ colours in the mock catalogue agrees with measurements from GAMA at different redshifts.

The galaxy catalogue is a flux-limited mock galaxy catalogue, covering the full sky with an r -band magnitude limit of $r < 20.0$ and median redshift $z \sim 0.2$. The angular and projected correlation functions of galaxies in the mock show good agreement with measurements from SDSS and GAMA, and the colour dependence of the clustering is reasonable. The BAO peak can be seen in the large-scale clustering of galaxies in the mock, and galaxies show realistic redshift-space distortions, making this mock useful for upcoming surveys that will measure these.

Here we have presented one mock galaxy catalogue, but to enable model inferences and place tight constraints on cosmological parameters, error covariances need to be determined. This requires the use of many mock catalogues, of the order of 100s to 10 000s. This could be achieved by coupling the HOD component of the mock with an approximate but fast method of generating halo catalogues (e.g. Manera et al. 2013; Monaco et al. 2013; Tashev, Zaldarriaga & Eisenstein 2013; White, Tinker & McBride 2014; Avila et al. 2015; Chuang et al. 2015; Kitaura et al. 2015).

ACKNOWLEDGEMENTS

The authors would like to thank Sam Lyddon for making preliminary versions of several of the figures, and Jon Loveday for useful discussions. We thank the referee, Darren Croton, for several insightful comments. This work was supported by the Science and Technology Facilities Council (ST/M503472/1). AS, SC, CB and PN acknowledge the support of the Science and Technology Facilities Council (ST/L00075X/1). ZZ was partially supported by NSF grant AST-1208891 and NASA grant NNX14AC89G. RA acknowledges support from AYA2015-66211-C2-2. PN acknowledges the support of the Royal Society through the award of a University

Research Fellowship and the European Research Council, through receipt of a Starting Grant (DEGAS-259586). IZ was supported by NSF grant AST-1612085, a CWRU Faculty Seed Grant and by ERC starting grant DEGAS-259586. IZ also acknowledges the hospitality of the ICC at Durham University.

GAMA is a joint European–Australasian project based around a spectroscopic campaign using the Anglo-Australian Telescope. The GAMA input catalogue is based on data taken from the Sloan Digital Sky Survey and the UKIRT Infrared Deep Sky Survey. Complementary imaging of the GAMA regions is being obtained by a number of independent survey programmes including *GALEX* MIS, VST KiDS, VISTA VIKING, *WISE*, Herschel-ATLAS, GMRT and ASKAP providing UV to radio coverage. GAMA is funded by the STFC (UK), the ARC (Australia), the AAO and the participating institutions. The GAMA website is <http://www.gama-survey.org/>.

This work used the DiRAC Data Centric system at Durham University, operated by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). This equipment was funded by BIS National E-infrastructure capital grant ST/K00042X/1, STFC capital grants ST/H008519/1 and ST/K00087X/1, STFC DiRAC Operations grant ST/K003267/1 and Durham University. DiRAC is part of the National E-Infrastructure.

REFERENCES

- Abazajian K. N. et al., 2009, *ApJS*, 182, 543
- Alonso D., 2012, preprint ([arXiv:1210.1833](https://arxiv.org/abs/1210.1833))
- Angulo R. E., White S. D. M., 2010, *MNRAS*, 405, 143
- Angulo R. E., Springel V., White S. D. M., Cole S., Jenkins A., Baugh C. M., Frenk C. S., 2012a, *MNRAS*, 425, 2722
- Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012b, *MNRAS*, 426, 2046
- Angulo R. E., Baugh C. M., Frenk C. S., Lacey C. G., 2014, *MNRAS*, 442, 3256
- Avila S., Murray S. G., Knebe A., Power C., Robotham A. S. G., Garcia-Bellido J., 2015, *MNRAS*, 450, 1856
- Baldry I. K. et al., 2010, *MNRAS*, 404, 86
- Baugh C. M., 2006, *Rep. Progress Phys.*, 69, 3101
- Baugh C. M., 2008, *Philos. Trans. R. Soc. Lond. Ser. A*, 366, 4381
- Benson A. J., 2010, *Phys. Rep.*, 495, 33
- Benson A. J., Cannella C., Cole S., 2016, *Comput. Astrophys. Cosmol.*, 3, 3
- Berlind A. A., Weinberg D. H., 2002, *ApJ*, 575, 587
- Bernyk M. et al., 2016, *ApJS*, 223, 9
- Blanton M. R. et al., 2003, *ApJ*, 592, 819
- Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001, *MNRAS*, 321, 559
- Burden A., Padmanabhan N., Cahn R. N., White M. J., Samushia L., 2017, *J. Cosmol. Astropart. Phys.*, 03, 001
- Chuang C.-H., Kitaura F.-S., Prada F., Zhao C., Yepes G., 2015, *MNRAS*, 446, 2621
- Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, *MNRAS*, 319, 168
- Conroy C., Wechsler R. H., Kravtsov A. V., 2006, *ApJ*, 647, 201
- Contreras S., Zehavi I., Baugh C. M., Padilla N., Norberg P., 2017, *MNRAS*, 465, 2833
- Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, *ApJ*, 292, 371
- de la Torre S., Peacock J. A., 2013, *MNRAS*, 435, 743
- DESI Collaboration et al., 2016a, preprint ([arXiv:1611.00036](https://arxiv.org/abs/1611.00036))
- DESI Collaboration et al., 2016b, preprint ([arXiv:1611.00037](https://arxiv.org/abs/1611.00037))
- Driver S. P. et al., 2009, *Astron. Geophys.*, 50, 5.12
- Driver S. P. et al., 2011, *MNRAS*, 413, 971
- Efstathiou G., Kaiser N., Saunders W., Lawrence A., Rowan-Robinson M., Ellis R. S., Frenk C. S., 1990, *MNRAS*, 247, 10p
- Eisenstein D. J., Blanton M., Zehavi I., Bahcall N., Brinkmann J., Loveday J., Meiksin A., Schneider D., 2005, *ApJ*, 619, 178
- Farrow D. J. et al., 2015, *MNRAS*, 454, 2120

Fosalba P., Crocce M., Gaztañaga E., Castander F. J., 2015, MNRAS, 448, 2987

Gonzalez-Perez V., Lacey C. G., Baugh C. M., Lagos C. D. P., Helly J., Campbell D. J. R., Mitchell P. D., 2014, MNRAS, 439, 264

Guo H., Zehavi I., Zheng Z., 2012, ApJ, 756, 127

Guo H. et al., 2015, MNRAS, 453, 4368

Hahn C., Scoccimarro R., Blanton M. R., Tinker J. L., Rodriguez-Torres S., 2017, MNRAS, 467, 1940

Hamilton A. J. S., 1992, ApJ, 385, L5

Hawkins E. et al., 2003, MNRAS, 346, 78

Hogg D. W., Baldry I. K., Blanton M. R., Eisenstein D. J., 2002, preprint (arXiv:astro-ph/0210394)

Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, MNRAS, 321, 372

Kitaura F.-S., Gil-Marín H., Scóccola C. G., Chuang C.-H., Müller V., Yepes G., Prada F., 2015, MNRAS, 450, 1836

Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, ApJ, 609, 35

Laureijs R. et al., 2011, preprint (arXiv:1110.3193)

Leauthaud A., Tinker J., Behroozi P. S., Busha M. T., Wechsler R. H., 2011, ApJ, 738, 45

Liske J. et al., 2015, MNRAS, 452, 2087

Loveday J. et al., 2012, MNRAS, 420, 1239

Loveday J. et al., 2015, MNRAS, 451, 1540

McNaught-Roberts T. et al., 2014, MNRAS, 445, 2125

Manera M. et al., 2013, MNRAS, 428, 1036

Merson A. I. et al., 2013, MNRAS, 429, 556

Monaco P., Sefusatti E., Borgani S., Crocce M., Fosalba P., Sheth R. K., Theuns T., 2013, MNRAS, 433, 2389

Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121

Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493

Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144

Pinol L., Cahn R. N., Hand N., Seljak U., White M., 2017, J. Cosmol. Astropart. Phys., 04, 008

Reid B. et al., 2016, MNRAS, 455, 1553

Robotham A. et al., 2010, Publ. Astron. Soc. Aust., 27, 76

Ross A. J. et al., 2017, MNRAS, 464, 1168

Samushia L., Percival W. J., Raccanelli A., 2012, MNRAS, 420, 2102

Schoenberg I. J., 1946, Q. Appl. Math., 4, 45

Scoccimarro R., Sheth R. K., Hui L., Jain B., 2001, ApJ, 546, 20

Seljak U., 2000, MNRAS, 318, 203

Sheth R. K., Tormen G., 1999, MNRAS, 308, 119

Skibba R. A., Sheth R. K., 2009, MNRAS, 392, 1080

Skibba R., Sheth R. K., Connolly A. J., Scranton R., 2006, MNRAS, 369, 68

Somerville R. S., Davé R., 2015, ARA&A, 53, 51

Spergel D. N. et al., 2003, ApJS, 148, 175

Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726

Springel V. et al., 2005, Nature, 435, 629

Tassev S., Zaldarriaga M., Eisenstein D. J., 2013, J. Cosmol. Astropart. Phys., 6, 036

Vale A., Ostriker J. P., 2004, MNRAS, 353, 189

Wang Y., Brunner R. J., Dolence J. C., 2013, MNRAS, 432, 1961

White M., Tinker J. L., McBride C. K., 2014, MNRAS, 437, 2594

Yang X., Mo H. J., van den Bosch F. C., 2003, MNRAS, 339, 1057

Zehavi I. et al., 2011, ApJ, 736, 59

Zheng Z. et al., 2005, ApJ, 633, 791

APPENDIX A: DATABASE

The full sky halo and galaxy lightcone catalogues are made publicly available on the Theoretical Astrophysical Observatory data base⁴ (Bernyk et al. 2016). The catalogues are also available at: <http://icc.dur.ac.uk/data/>.

⁴ <https://tao.asvo.org.au/tao/>

A1 Halo catalogue

The halo catalogue contains a total of 5.1 billion haloes out to a redshift of $z = 2.2$, and contains the following halo properties.

- (i) z_{obs} , the observed redshift, which takes into account the peculiar velocity of the halo.
- (ii) z_{cos} , the cosmological redshift, which ignores the effect of the peculiar velocity.
- (iii) Right ascension, in degrees.
- (iv) Declination, in degrees.
- (v) $M_{200\text{m}}$, the mass enclosed by a sphere in which the average density is 200 times the mean density of the Universe, interpolated to the redshift at which the halo crosses the lightcone, in units of $10^{10} h^{-1} M_{\odot}$.
- (vi) M_{200c} , the mass enclosed by a sphere in which the average density is 200 times the critical density of the Universe, interpolated to the redshift at which the halo crosses the lightcone, in units of $10^{10} h^{-1} M_{\odot}$.
- (vii) V_{max} , the maximum circular velocity, in units of km s^{-1} .
- (viii) $R_{V_{\text{max}}}$, the radius at which V_{max} occurs, in $h^{-1} \text{Mpc}$.
- (ix) $\sigma_{R_{200\text{m}}}$, velocity dispersion of particles within $R_{200\text{m}}$, in units of km s^{-1} .
- (x) Snapshot number in the MXXL simulation.
- (xi) Halo id in the MXXL simulation.

A2 Galaxy catalogue

The galaxy catalogue contains 58.1 million galaxies with $r < 20$, out to redshift $z = 0.8$, and contains the following properties.

- (i) z_{obs} , the observed redshift, which takes into account the peculiar velocity of the galaxy.
- (ii) z_{cos} , the cosmological redshift, which ignores the effect of the peculiar velocity.
- (iii) Right ascension, in degrees.
- (iv) Declination, in degrees.
- (v) $M_{200\text{m}}$ of the host halo, interpolated to the redshift at which the halo crosses the lightcone, in units of $10^{10} h^{-1} M_{\odot}$.
- (vi) Apparent r -band magnitude.
- (vii) $^{0.1}M_r - 5 \log h$, the rest-frame absolute r -band magnitude, k -corrected to a reference redshift of $z_{\text{ref}} = 0.1$, with no evolutionary correction.
- (viii) $^{0.1}(g - r)$ colour, k -corrected to a reference redshift $z_{\text{ref}} = 0.1$.
- (ix) A flag indicating whether the galaxy is a central or a satellite, and whether it is in a resolved or unresolved halo.
- (x) Snapshot number in the MXXL simulation.
- (xi) Halo id in the MXXL simulation.