

# Linear bias forecasts for emission line cosmological surveys

Alexander Merson<sup>1,2\*</sup>, Alex Smith<sup>3,4</sup>, Andrew Benson<sup>5</sup>, Yun Wang<sup>2</sup>, Carlton Baugh<sup>4</sup>

<sup>1</sup>*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA*

<sup>2</sup>*IPAC, Mail Code 314-6, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA*

<sup>3</sup>*IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France*

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

<sup>5</sup>*Carnegie Observatories, 813 Santa Barbara Street, Pasadena, CA 91101, USA*

## ABSTRACT

We forecast the linear bias for H $\alpha$ -emitting galaxies at high redshift. To simulate a Euclid-like and a WFIRST-like survey, we place galaxies into a large-volume dark matter halo lightcone by sampling a library of luminosity-dependent halo occupation distributions (HODs), which is constructed using a physically motivated galaxy formation model. We calibrate the dust attenuation in the lightcones such that they are able to reproduce the H $\alpha$  luminosity function or the H $\alpha$  cumulative number counts. The angle-averaged galaxy correlation function is computed for each survey in redshift slices of width  $\Delta z = 0.2$ . In each redshift bin the linear bias can be fitted with a single, scale-independent value that increases with increasing redshift. Fitting for the evolution of linear bias with redshift, we find that our Euclid-like and WFIRST-like surveys are both consistent within error with the relation  $b(z) = 0.7z + 0.7$ . Our bias forecasts are consistent with bias measurements from the HiZELS survey. We find that the Euclid-like and WFIRST-like surveys yield linear biases that are broadly consistent within error, most likely due to the HOD for the WFIRST-like survey having a steeper power-law slope towards larger halo masses.

**Key words:** galaxies:formation; cosmology: large-scale structure of Universe; galaxies: statistics; methods:numerical

## 1 INTRODUCTION

Probing the nature of the driving force behind the observed accelerated expansion of the Universe continues to be one of the major goals of modern cosmology. Most of the observational evidence gathered to date is consistent with the theory that the expansion is the result of a mysterious phenomenon known as *dark energy*, although existing observations lack the statistical precision to allow alternative theories to be differentiated and conclusively ruled out. Up and coming cosmological missions aim to make very precise measurements of different cosmological probes of the expansion history of the Universe and the growth rate of cosmic large-scale structure in order to distinguish between the competing theories (e.g. Albrecht et al. 2006; Guzzo et al. 2008; Wang 2008a,b).

For missions such as the ESA-led *Euclid* mission (Laureijs et al. 2011) and the NASA-led *Wide Field Infrared Survey Telescope* mission (WFIRST, Dressler et al. 2012; Green et al. 2012; Spergel et al. 2015), dark energy will be probed using a spectroscopic galaxy redshift survey, which will measure the galaxy clustering and redshift-space distortions of the galaxy distribution, and a photometric survey, which will be used to measure weak gravitational lensing shear of the galaxies. Specifically for the galaxy

clustering measurements, these missions will measure the *baryon acoustic oscillations* (BAO) in the galaxy population, which can be used as a standard ruler to probe the expansion history of the Universe (Blake & Glazebrook 2003; Seo & Eisenstein 2003), and redshift-space distortions, which are sensitive to the growth rate of the large-scale structure (Kaiser 1987; Song & Percival 2009).

The spectroscopic redshift surveys of the Euclid and WFIRST missions will target many tens of millions of emission line galaxies (ELGs), which will be identified using near-IR grism spectroscopy. The primary targets for the spectroscopic galaxy redshift surveys are H $\alpha$ -emitting galaxies between redshifts  $0.9 \lesssim z \lesssim 2$ . The Euclid wide-area survey will cover 15,000 square degrees to an H $\alpha$  line flux limit of  $2 \times 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$  and is expected to detect H $\alpha$ -emitting galaxies in the redshift range  $0.9 \lesssim z \lesssim 1.8$ . In a similar fashion, the WFIRST High Latitude Survey (HLS) will cover  $\sim 2200$  square degrees to a fainter H $\alpha$  line flux limit of  $1 \times 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$  and is expected to detect H $\alpha$ -emitting galaxies in a redshift range of  $1 \lesssim z \lesssim 2$ . Due to their observing strategy, both missions are expected to detect additional emission lines belonging to galaxies outside of these redshift ranges. For instance, both Euclid and WFIRST expect to detect [OIII] emission from galaxies at  $z \gtrsim 2$ . In this work we focus on the redshift surveys of H $\alpha$ -emitting galaxies. The Euclid and WFIRST dark energy missions are designed to be highly complementary: the large

\* E-mail: alex.i.merson@gmail.com

area and shallower depth of Euclid is ideal for optimising statistical precision in cosmological measurements, whilst the smaller area and greater depth of WFIRST is ideal for understanding the systematics that are expected to dominate over measurement uncertainties.

One of the systematics that must be taken into account is *galaxy bias*, which describes how galaxies trace the underlying dark matter distribution. Galaxy formation does not occur uniformly in space, but occurs primarily in the peaks of the matter density field. Galaxies are therefore biased tracers of the density field, sampling only the over-dense regions (e.g. Kaiser 1984; Bardeen et al. 1986; see Desjacques et al. 2018 for a recent review). The bias,  $b$ , of a population of galaxies is defined according to,

$$\xi_{\text{gal}}(r) = b^2(r)\xi_{\text{DM}}(r), \quad (1)$$

where  $\xi_{\text{gal}}(r)$  is the galaxy correlation function and  $\xi_{\text{DM}}(r)$  is the dark matter correlation function, expressed as a function of spatial separation  $r$ . Observational evidence indicates that different galaxy populations, identified for example by luminosity, stellar mass or morphological type, display different clustering amplitudes (e.g. Davis & Geller 1976; Dressler 1980; Guzzo et al. 1997; Norberg et al. 2001, 2002; Zehavi et al. 2002, 2011; Guo et al. 2013; Skibba et al. 2014; Kim et al. 2015; McCracken et al. 2015; Hatfield et al. 2016; Favole et al. 2017; Law-Smith & Eisenstein 2017; Cochrane et al. 2017, 2018; Durkalec et al. 2018). This has been further confirmed with simulations of galaxy formation (e.g. Kauffmann et al. 1997; Benson et al. 2000a; Orsi et al. 2010; Contreras et al. 2013). As such, different galaxy populations are expected to display different bias values. This can be understood in terms of the halo model, which predicts that different galaxy populations commonly reside in dark matter halos of different mass. The hierarchical nature of structure formation means that the bias of galaxies will additionally vary with redshift (e.g. Fry 1996; Tegmark & Peebles 1998; Hui & Parfrey 2008; Basilakos et al. 2012; Mirbabayi et al. 2015). In the non-linear regime the bias is typically scale-dependent, but in the linear regime, at scales  $r \gtrsim 50 h^{-1} \text{Mpc}$ , the bias approaches a constant value (Mann et al. 1998; Peacock & Smith 2000; Seljak 2000; Benson et al. 2000b; Blanton et al. 2000; Berlind & Weinberg 2002; Verde et al. 2002). We refer to the constant bias on large-scales as linear bias,  $b(r) \sim b_{\text{lin}}$ .

To date, many of the clustering analyses of emission line galaxies at high redshift have been carried out using data from the High- $z$  Emission Line Survey (HiZELS, Geach et al. 2008; Sobral et al. 2009, 2013) – a ground-based, narrow band survey capable of detecting emission line galaxies up to  $z \sim 9$ . Using a sample of  $\sim 700$  H $\alpha$ -emitters at  $z = 0.84$ , Sobral et al. (2010) observed that the clustering strength of H $\alpha$ -emitters is strongly dependent on the H $\alpha$  luminosity. Geach et al. (2012) examined the clustering of 370 H $\alpha$ -emitters at  $z = 2.23$  to measure a bias of  $2.4^{+0.1}_{-0.2}$  and estimate that the dark matter halos hosting H $\alpha$ -emitters have a typical mass of  $\log_{10}(M_h/h^{-1}M_\odot) = 11.7 \pm 0.1$ . Geach et al. (2012) additionally construct a parametric model for the *halo occupation distribution* (HOD) of their H $\alpha$ -emitters and attempt to fit this model to their data. Due to the size of their galaxy sample, Geach et al. (2012) were unable to place strong constraints on the HOD parameters. However, despite this, they were able to constrain the effective halo mass of the H $\alpha$ -emitters to be  $\log_{10}(M_{\text{eff}}/h^{-1}M_\odot) = 12.1^{+0.1}_{-0.2}$ . Recently Cochrane et al. (2017) built upon the previous HiZELS results by carrying out an extensive clustering analysis of  $\sim 3000$  H $\alpha$ -emitters at  $z = 0.8$ ,  $\sim 450$  emitters at  $z = 1.47$  and  $\sim 730$  emitters at  $z = 2.23$ , split into both differential and cumulative luminosity bins. Using Markov Chain Monte Carlo sampling,

Cochrane et al. (2017) fit a parametrised HOD to their clustering results to estimate the bias and halo mass of their H $\alpha$ -emitters. They find that the bias of the emitters increases both with redshift and with luminosity, and that the emitters are hosted by halos with mass typically  $\sim 10^{12} h^{-1} M_\odot$ .

In this work we use simulated galaxy catalogues to forecast the linear bias as a function of redshift for H $\alpha$ -emitting galaxies in a Euclid-like redshift survey and a WFIRST-like redshift survey. Forecasts of the galaxy bias are important for two reasons. Firstly, understanding the bias of the H $\alpha$ -emitting galaxies will be crucial for Euclid and WFIRST if we are to correctly infer the dark matter clustering from the galaxy clustering measurements, and thereby accurately estimate cosmological parameters (Gaztañaga et al. 2012; Salvador et al. 2019). Clerkin et al. (2015) provide a comparison of several bias evolution prescriptions and the impact that choice of prescription has on cosmological parameter constraints. Secondly, the *figure-of-merit* of a cosmological mission, which indicates the ability of a mission to successfully measure dark energy, is sensitive to the understanding of galaxy bias. As such forecasting of the bias is vital for helping optimise the observational strategy of a cosmological mission. For example, Orsi et al. (2010) used galaxy mock catalogues built using the GALFORM semi-analytical galaxy formation model to compare effective volumes that could be probed by an H $\alpha$ -selected galaxy survey and an H-band selected galaxy survey. They determine that for an H $\alpha$ -selected galaxy survey to probe an effective volume comparable to that of an  $H_{\text{AB}} = 22$  slit-based survey, the H $\alpha$ -selected would need to probe down to a flux depth of  $10^{-16} \text{erg s}^{-1} \text{cm}^{-2}$ . Orsi et al. (2010) additionally provide some forecasts for the luminosity dependence of the bias of H $\alpha$ -emitters between  $z = 0$  and  $z = 2$ , though due to the limited volume of the underlying N-body simulation they are unable to estimate the bias well into the linear regime ( $r \gtrsim 50 h^{-1} \text{Mpc}$ ).

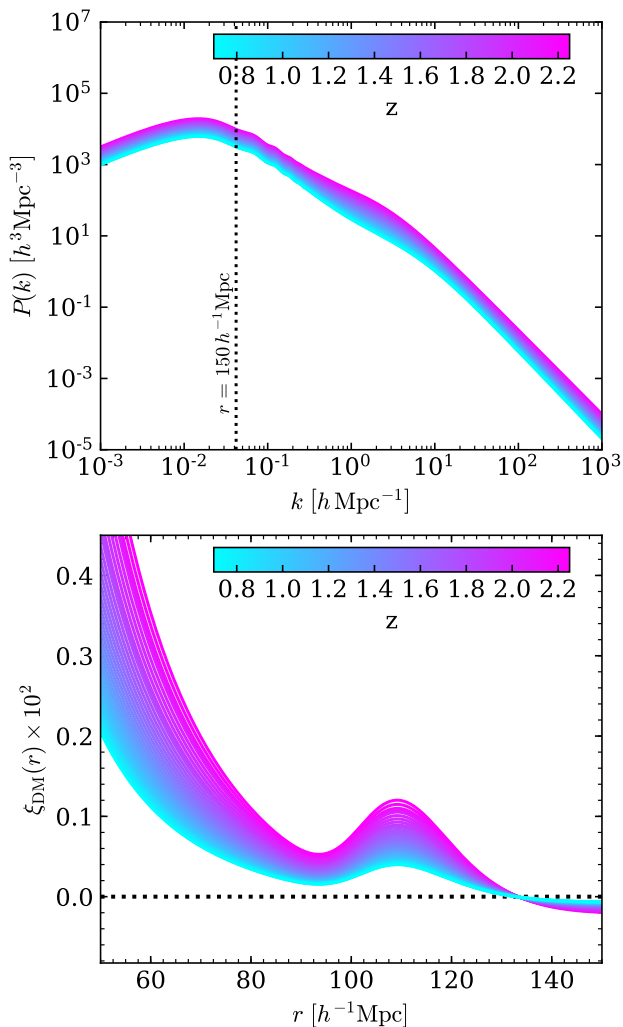
The structure of the paper is as follows: in § 2 we give details of how we construct our lightcone galaxy catalogues; in § 3 we discuss how we calibrate our lightcone catalogues to match the luminosity function and cumulative number counts of H $\alpha$ -emitters; in § 4 we present our linear bias forecasts; in § 5 we compare the linear bias values for the Euclid-like and WFIRST-like surveys; and in § 6 we summarise and conclude.

## 2 MOCK CATALOGUE CONSTRUCTION

For this work we use a pair of mock catalogues built using the methodology of Smith et al. (2017), who populate the dark matter halos of the *Millennium XXL* simulation (MXXL, Angulo et al. 2012) using a set of luminosity-dependent HODs. In this section we provide further details on the construction of these mock catalogues.

### 2.1 The Millennium XXL Simulation

In this work we use the *Millennium XXL* simulation (MXXL, Angulo et al. 2012), which uses  $6720^3 = 303, 464, 448, 000$  particles of mass  $6.17 \times 10^9 h^{-1} M_\odot$  to follow the non-linear, hierarchical growth of dark matter structure within a cubic volume of  $3 h^{-1} \text{Gpc}$ . The cosmology adopted in the MXXL simulation is a  $\Lambda$ CDM cosmology with parameters identical to the cosmological parameters adopted for the *Millennium Simulation* (Springel et al. 2005). These parameters are: a baryon matter density  $\Omega_b = 0.045$ ; a total matter density  $\Omega_m = \Omega_b + \Omega_{\text{CDM}} = 0.25$ , where  $\Omega_{\text{CDM}}$  is



**Figure 1.** Non-linear matter power spectra (top panel) and dark matter two-point correlation function (bottom panel) for an MXXL cosmology for redshifts in range  $0.8 \lesssim z \lesssim 2$ , as indicated by the inset colour bar. All power spectra and correlation functions were computed using the CLASS and HALOFIT functionality in the `NBodyKit` python package (Hand et al. 2018). In the upper panel a vertical dotted line indicates the wavenumber corresponding to a co-moving distance of  $150 h^{-1} \text{Mpc}$ . In the lower panel the BAO peak is clearly visible at approximately  $110 h^{-1} \text{Mpc}$ .

the density in cold dark matter; a dark energy density  $\Omega_\Lambda = 0.75$ ; a Hubble constant  $H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$ , where  $h = 0.73$ ; a primordial scalar spectral index  $n_s = 1$  and a fluctuation amplitude  $\sigma_8 = 0.9$ . This cosmological parameter set is consistent with the first year results from the Wilkinson Microwave Anisotropy Probe (Spergel et al. 2003). We note that this cosmology is discrepant with the latest cosmological parameters measured from the Planck mission (Planck Collaboration et al. 2018), but argue that this should not be of major concern since our ignorance in the modelling of galaxy formation will dominate over any discrepancies in cosmology. In Fig. 1 we show, for reference, the non-linear matter power spectrum and corresponding real-space correlation function as computed for the MXXL cosmology using the open source `Nbodykit`<sup>1</sup> python package (Hand et al. 2018). The BAO peak

<sup>1</sup> <https://github.com/bccp/nbodykit>

is clearly visible and, given the parameters of the simulation, is located at approximately  $110 h^{-1} \text{Mpc}$ .

Particle positions and velocities are stored in 63 snapshots at fixed epochs spaced approximately logarithmically in expansion factor between  $z = 20$  and  $z = 0$ , as was done for the Millennium Simulation (see Lemson & Virgo Consortium 2006). Halo merger trees have also been constructed using the SUBFIND structure finding algorithm (Springel et al. 2001). Smith et al. (2017) additionally constructed friends-of-friends (FOF) merger trees using the most massive subhalo in each FOF group. Further details regarding the MXXL simulation can be found in Angulo et al. (2012).

Smith et al. (2017) constructed a lightcone catalogue of dark matter halos from the MXXL simulation using the methodology described in Merson et al. (2013). An observer is first placed in the simulation box at a randomly chosen position and the simulation box is replicated about the observer so as to generate a sufficiently large volume reaching out to the user-specified redshift. Using the positions of halos and their associated descendants, one can determine the pair of snapshots between which any particular merger tree intercepted the past lightcone of the observer (if it intercepts the lightcone at all). Once the pair of snapshots has been identified, a binary search algorithm is used to interpolate along the trajectory of the halo to compute the exact location at which the halo crosses the lightcone. Interpolation of the halo trajectories between pairs of snapshots is done using cubic interpolation, with the initial position and velocity of a halo and its corresponding descendant used to set the boundary conditions. The lightcone that Smith et al. (2017) built is all-sky and extends out to redshift  $z = 2.2$ .

## 2.2 Populating the MXXL dark matter halos

Ideally we would like to populate the halos of the MXXL simulation with galaxies by running a physically motivated galaxy formation model directly on the MXXL halo merger trees. Unfortunately the mass resolution of the MXXL simulation is too poor, with a minimum resolved halo mass of  $1.22 \times 10^{11} h^{-1} M_\odot$ , such that running a semi-analytical galaxy formation model on the MXXL trees directly would yield an unrealistic galaxy population (e.g. see Angulo et al. 2014). We note, however, that this minimum halo mass is just under an order of magnitude smaller than the typical halo mass for  $H\alpha$ -emitters estimated by Geach et al. (2012).

To populate the MXXL halos we therefore use the pipeline presented in Smith et al. (2017). This pipeline follows the methodology of Skibba et al. (2006), whereby halos are populated with galaxies of different luminosities by random sampling of probability distribution functions created from sets of luminosity-dependent halo occupation distributions (HODs) that evolve with redshift.

Smith et al. (2017) originally developed this pipeline for the purpose of constructing an  $r$ -band selected galaxy catalogue to mimic the Dark Energy Spectroscopic Instrument Bright Galaxy Survey (DESI BGS, DESI Collaboration et al. 2016). For broadband photometrically-selected samples such as this, the HOD is well understood and can be parametrised easily, (e.g. Zehavi et al. 2011; Zhai et al. 2017). However, the HOD of  $H\alpha$ -emitting galaxies is much less well understood, although some parametrisation has been attempted based upon clustering results from HiZELS (Geach et al. 2012; Cochrane et al. 2017). Some attempts have also been made to fit the HOD parameters for simulated star-forming galaxies and simulated [OII]-emitters (Contreras et al. 2013; Gonzalez-Perez et al. 2018; Cochrane & Best 2018).

Rather than adopting any parametrisation for the HODs of  $H\alpha$ -emitting galaxies, we again consider using a semi-analytical

galaxy formation model, but in this case using a model to generate a library of luminosity-dependent HODs that we can interpolate over as a function of luminosity and redshift. We choose to use the open source galaxy formation model GALACTICUS (Benson 2012).

### 2.2.1 The GALACTICUS galaxy formation model

The GALACTICUS<sup>2</sup> semi-analytical galaxy formation model is designed to follow the formation and evolution of a galaxy population within a merging hierarchical distribution of dark matter halos. The astrophysics governing the baryonic processes occurring within the dark matter halos is described using sets of coupled ordinary differential equations (ODEs). These processes include the rate of radiative gas cooling, the quiescent star formation rate, the chemical enrichment of the stellar and gaseous components, as well as the regulation of feedback processes from supernovae and active galactic nuclei. By calling the ODE solver within GALACTICUS at various epochs, one can compute the star formation histories of a population of galaxies from high redshift to the present day. Given a stellar initial mass function (IMF), these histories can be convolved with a single stellar population synthesis model to generate a spectral energy distribution for each galaxy, with which we can compute photometric luminosities for a specified set of filter transmission curves. By default GALACTICUS adopts the Flexible Stellar Population Synthesis code of Conroy et al. (2010), with a Chabrier (2003) IMF.

Emission line luminosities are computed for the GALACTICUS galaxies by interpolating over tabulated libraries generated from the CLOUDY photo-ionisation code (Ferland et al. 2013). These libraries store emission line luminosities as a function of (i) the ionizing continua luminosities for various species (H I, He I and O II), (ii) the hydrogen gas density, (iii) the metallicity of the interstellar medium and (iv) the volume filling factors of H II regions. All of these properties can be computed for the galaxies from GALACTICUS and so by interpolating over the CLOUDY tables in this manner we obtain emission line luminosities that are consistent with the other galaxy properties. Further details regarding the calculation of the emission line luminosities are provided in Merson et al. (2018).

Typically the GALACTICUS model parameters are calibrated to reproduce numerous observational statistics of the galaxy population, in particular those statistics of the local Universe that are the most tightly constrained. The version of GALACTICUS that we use here is the same as the version used in Merson et al. (2018), which employs the standard parameter set. This version of GALACTICUS has been calibrated to reproduce a range of galaxy statistics, particularly those in the local Universe. Emphasis has been placed on reproducing the  $z = 0$  galaxy stellar mass function from the Sloan Digital Sky Survey, as measured by Li & White (2009). Full details of the calibration procedure for GALACTICUS can be found in Benson (2014) and Knebe et al. (2015).

### 2.2.2 Generating a library of luminosity-dependent HODs

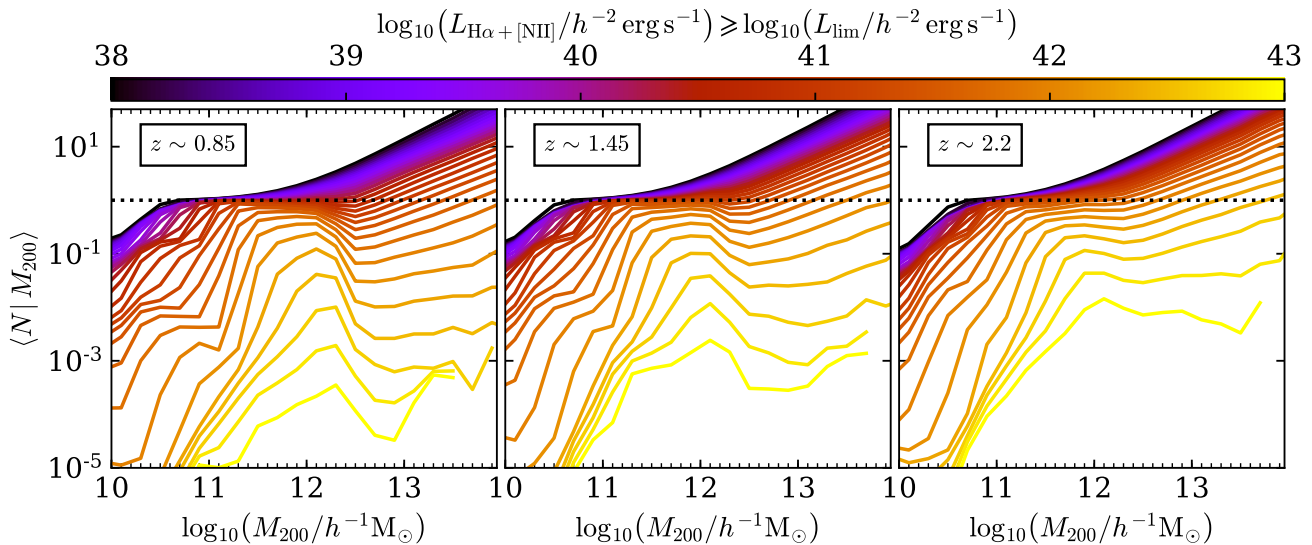
As we have previously mentioned, the poor mass resolution of the MXXL simulation means that we are unable to run GALACTICUS directly on the MXXL merger trees. Instead we run the model on

the halo merger trees of the Millennium Simulation, which we remind the reader has a cosmology identical to that of the MXXL simulation. Populations of galaxies are output at 31 snapshots with redshifts in the range  $0.7 < z < 2.2$  (i.e. completely spanning the redshift ranges of Euclid and WFIRST). For each snapshot we split the galaxies into 31 cumulative luminosity-limited samples, with blended luminosity limits evenly separated in log-space between  $L_{\text{H}\alpha + [\text{NII}]} = 10^{38} h^{-2} \text{erg s}^{-1}$  and  $L_{\text{H}\alpha + [\text{NII}]} = 10^{43} h^{-2} \text{erg s}^{-1}$ . We select galaxies by their blended  $\text{H}\alpha + [\text{NII}]$  luminosity as the resolution of the grisms proposed for Euclid will be too poor to deblend the  $\text{H}\alpha$  line and  $[\text{NII}]$  doublet. See Faisst et al. (2018) and Martens et al. (2019) for discussions on the impact that  $\text{H}\alpha$  and  $[\text{NII}]$  line blending can have on Euclid and WFIRST cosmological measurements. To introduce  $[\text{NII}]$  contamination into the  $\text{H}\alpha$  luminosities from GALACTICUS, we follow the methodology adopted by Merson et al. (2018). The blended luminosities have no dust attenuation applied. For each luminosity-limited sample we construct the HOD by computing the mean number of central and satellite galaxies in halos binned by halo mass,  $M_{200}$ , which we define as the mass within an over-density with an average density corresponding to 200 times the mean density of the Universe. Each HOD is measured in 26 mass bins evenly separated in log-space between  $M_{200} = 10^{9.7} h^{-1} M_{\odot}$  and  $M_{200} = 10^{14.7} h^{-1} M_{\odot}$ . At this stage we now have a tabulated library of luminosity-dependent HODs over which we can interpolate as a function of halo mass, luminosity and redshift. Note that in the library we store three sets of HODs: the HODs for the central galaxies only, the HOD for the satellite galaxies only, and the combined HOD.

In Fig. 2 we show the combined HODs for our 31 blended luminosity-limited samples as measured from running GALACTICUS on three of the redshift snapshots of the Millennium Simulation. The shapes of these HODs are consistent with the shapes of emission line luminosity-selected HODs presented elsewhere in the literature (e.g. Geach et al. 2012; Contreras et al. 2013; Gonzalez-Perez et al. 2018; Cochrane & Best 2018). For the faintest luminosity limits considered, the HODs show a smooth step-like function, similar to the HODs for galaxy samples selected using broadband photometry. For masses below  $10^{11} h^{-1} M_{\odot}$  there is evidence for a secondary step feature, particularly towards lower redshift. However, we note that at such low masses we are approaching the resolution limit of the Millennium Simulation and so the extent to which this secondary step is real is uncertain.

Towards brighter luminosity limits the amplitude of the HOD decreases with the step-like shape of the HOD disappearing as the number of galaxies with bright  $\text{H}\alpha$  luminosity rapidly declines, particularly for very massive halos. For the brightest luminosity limits considered, the distribution of central galaxies is much more peaked, with a maximum value at a halo mass of approximately  $10^{12} h^{-1} M_{\odot}$ . This is understandable as  $\text{H}\alpha$  emission occurs in regions of ongoing star-formation and so  $\text{H}\alpha$  luminosity is correlated with galaxy star-formation rate, which, due to feedback from active galactic nuclei, does not increase monotonically with increasing halo mass. The quenching of satellite galaxies in massive halos also leads to a decline in the number of satellite galaxies with bright  $\text{H}\alpha$  emission. In halos with mass above  $10^{12} h^{-1} M_{\odot}$ , for  $L_{\text{H}\alpha + [\text{NII}]} \gtrsim 10^{41} \text{erg s}^{-1}$  the satellite fraction is below approximately 5 per cent. Indeed the occupation number for satellite galaxies still appears to follow a power law relation, but with a slope that decreases with increasing luminosity. The impact of these quenching mechanisms is that for bright luminosities the HOD turns over for halos with mass  $M_{200} \gtrsim 10^{12} h^{-1} M_{\odot}$ . We can see from Fig. 2 that the turn-over is more significant towards lower redshifts. In

<sup>2</sup> Here we use version 0.9.4 of GALACTICUS, which is publicly available from: [bitbucket.org/galacticusdev/galacticus](https://bitbucket.org/galacticusdev/galacticus). The Mercurial hash I.D. for the particular revision used is: 4787d94cd86e.



**Figure 2.** Halo Occupation Distributions (HODs) as measured from running GALACTICUS on three snapshots of the Millennium Simulation:  $z = 0.85$  (left-hand panel),  $z = 1.45$  (centre panel) and  $z = 2.2$  (right-hand panel). The various colours of the lines indicate the blended  $H\alpha + [NII]$  luminosity limit used to select the galaxies, as indicated in the colour bar. The luminosities have not been attenuated due to dust. The halo mass assumed is the mass within an over-density with average density corresponding to 200 times the mean density of the Universe.

addition we can see that the amplitude of the HOD decreases with increasing luminosity limit, suggesting that galaxies that are bright in  $H\alpha$ -emission are becoming increasingly rare. This change in the amplitude of the HOD, along with the change in the strength of the turnover, is consistent with the observed decline in the global star-formation rate density towards the present day (e.g. Madau et al. 1998; Arnouts et al. 2005; Ly et al. 2007; Shim et al. 2009; Brammer et al. 2011; Madau & Dickinson 2014).

### 2.2.3 Placing galaxies into the MXXL halos

Having generated a library of HODs we now randomly sample the HODs to populate the halos in the MXXL halo lightcone. Note that by doing so, we lose any environmental-dependence (i.e., assembly bias). This does not matter for our present purposes, since we are only interested in measuring the linear bias on large scales.

In order to sample the HODs we will need to interpolate over the library as a function of halo mass, luminosity and redshift. At low masses we can interpolate linearly in log-space between the tabulated values. However, for the most massive halos in the MXXL lightcone, approaching  $10^{15}h^{-1}M_{\odot}$ , we will need to extrapolate beyond the halo mass range of the tabulated HODs. The limited volume of the Millennium Simulation means that our library of tabulated HODs from GALACTICUS extends only to  $M_{200} = 10^{14.7}h^{-1}M_{\odot}$ . In order to extrapolate to the higher halo masses we therefore decide to fit smoothed functional forms to the tabulated occupation distributions for the central galaxies and the satellite galaxies. Details of this fitting procedure are provided in Appendix A.

The functional fits to the tabulated HODs are shown in Fig. 3. In the majority of cases the fits provide a reasonably good description of the tabulated HODs. In some cases, particularly at lower redshifts, the noise in the tabulated HODs for the brightest luminosity limits leads to overlap between the HODs of different luminosity limits. In these instances fits to the HODs will cross such that interpolation between the HODs as a function of luminosity would

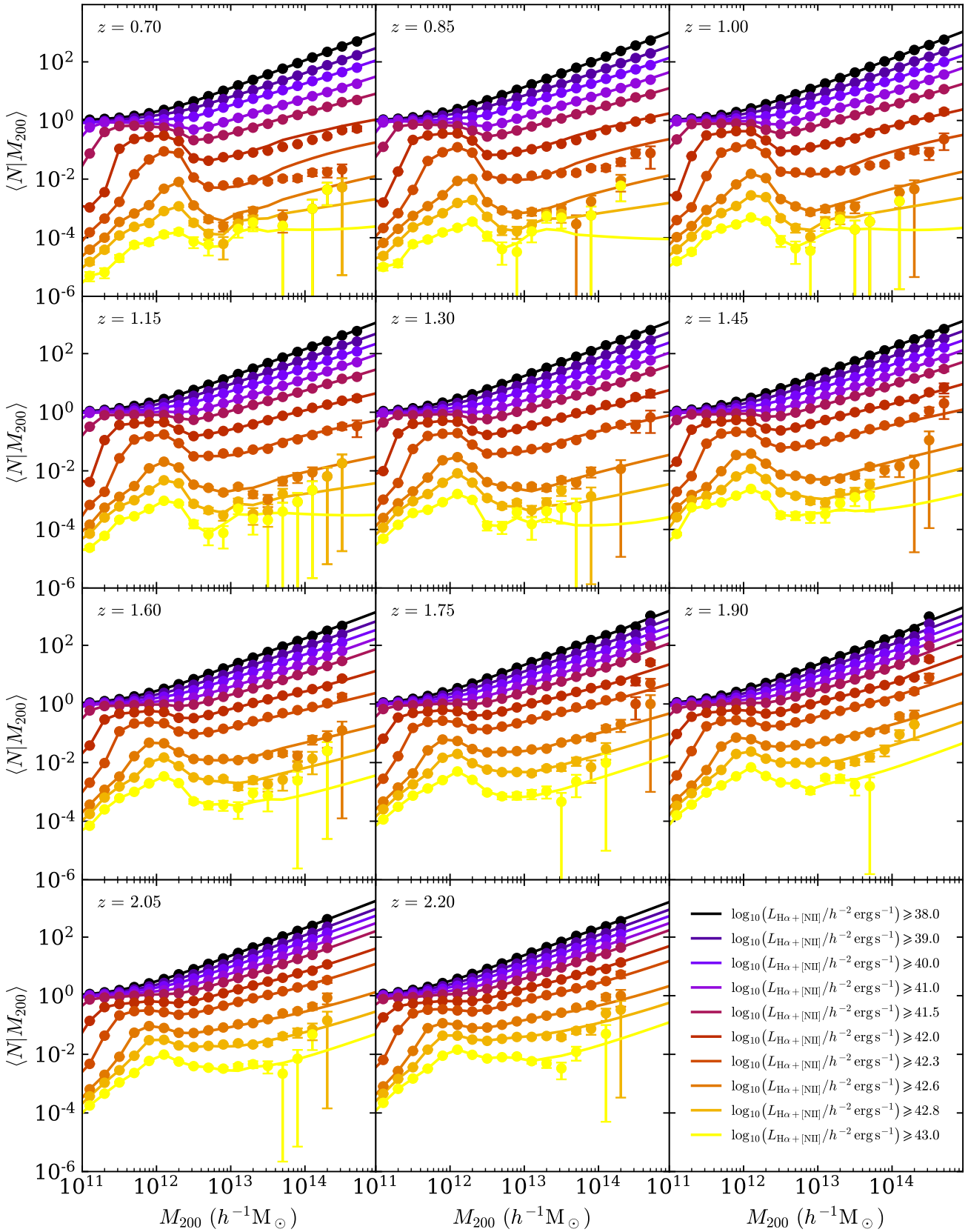
yield a constant HOD with increasing luminosity, which can be a problem in the random sampling of the HODs. This issue can be resolved by applying small vertical shifts of less than 0.3 dex to the amplitude of the HODs of the brightest luminosity limits. Given the rarity of bright  $H\alpha$ -emitting galaxies in massive halos at lower redshift, applying these offsets leads to only negligible differences in the number density of galaxies in our lightcone catalogues (Smith 2018).

Galaxies are then placed into each halo using a methodology presented in Smith et al. (2017) and Smith (2018). The first step in the process is to select a minimum luminosity,  $L_{\min}$ , for the lightcone catalogue. Here we adopt  $L_{\min} = 10^{40}h^{-2}\text{erg s}^{-1}$ , which, at the lowest redshift in the mock, is about an order of magnitude fainter than the luminosity limit for a WFIRST-like selection. Following this we identify all of the halos that could host a central galaxy with  $L_{H\alpha+[NII]} \geq L_{\min}$  by interpolating over the HOD with the corresponding luminosity threshold. We assign a central galaxy to each of these halos for which  $x_1 < \langle N_{\text{cen}}(> L_{\min}|M_{200}, z) \rangle$ , where  $\langle N_{\text{cen}}(> L_{\min}|M_{200}, z) \rangle$  is the mean number of central galaxies brighter than  $L_{\min}$  that could be found in a halo of mass  $M_{200}$  at redshift  $z$ , and  $x_1$  is a random number drawn from a uniform distribution  $0 \leq x_1 \leq 1$ . Central galaxies are placed at the centre-of-mass of the halos, with a velocity equal to the velocity of their host halo. The luminosity of the galaxy,  $L_{\text{cen}}$ , is determined by interpolating over the library of HODs to find the luminosity that satisfies,

$$\frac{\langle N_{\text{cen}}(> L_{\text{cen}}|M_{200}, z) \rangle}{\langle N_{\text{cen}}(> L_{\min}|M_{200}, z) \rangle} = x_2, \quad (2)$$

where  $x_2$  is a second random number, again drawn from a  $0 \leq x_2 \leq 1$  uniform distribution.

The number of satellite galaxies assigned to a halo is determined by drawing from a Poisson distribution with a mean equal to  $\langle N_{\text{sat}}(> L_{\min}|M_{200}, z) \rangle$ , where  $\langle N_{\text{sat}}(> L_{\min}|M_{200}, z) \rangle$  is the mean number of satellite galaxies brighter than  $L_{\min}$  that could be found in a halo of mass  $M_{200}$  at redshift  $z$ . Satellite galaxies



**Figure 3.** Tabulated halo occupation distributions (HODs) from GALACTICUS for 10 luminosity-selected samples of  $\text{H}\alpha$ -emitting galaxies shown for various redshifts between  $z = 0.7$  and  $z = 2.2$ . Solid lines show the functional fits to the HODs. The various colours of the lines indicate the blended  $\text{H}\alpha + [\text{NII}]$  luminosity limit used to select the galaxies, as indicated in the lower right-hand panel. Luminosities have not had any dust attenuation applied. The halo mass assumed is the mass within an over-density with average density corresponding to 200 times the mean density of the Universe.

**Table 1.** Specifications for the lightcone catalogues considered in this work. Note that the Euclid-like lightcone is constructed by applying a completeness limit to the MXXL-15K. The WFIRST-like lightcone is derived from the MXXL-2K lightcone in a similar manner. The sky areas for the MXXL-15K and the Euclid-like lightcones have been rounded to the nearest integer – in reality these lightcones each have an area of  $(3/8) \times 4\pi$  steradians.

Catalogue Name	Flux Limit [erg s <sup>-1</sup> cm <sup>-2</sup> ]	Right Ascension	Declination	Area [deg <sup>2</sup> ]	Redshift Range	Completeness
MXXL-15K	$2 \times 10^{-16}$	$0^\circ \leq \alpha \leq 270^\circ$	$0^\circ \leq \delta \leq 90^\circ$	15470	$0.7 \leq z < 2.2$	1.0
Euclid-like	$2 \times 10^{-16}$	$0^\circ \leq \alpha \leq 270^\circ$	$0^\circ \leq \delta \leq 90^\circ$	15470	$0.9 \leq z < 1.9$	0.45
MXXL-2K	$1 \times 10^{-16}$	$50^\circ \leq \alpha \leq 100^\circ$	$-20^\circ \leq \delta \leq 20^\circ$	2000	$0.7 \leq z < 2.2$	1.0
WFIRST-like	$1 \times 10^{-16}$	$50^\circ \leq \alpha \leq 100^\circ$	$-20^\circ \leq \delta \leq 20^\circ$	2000	$1.0 \leq z < 2.0$	0.7

are placed randomly within the halo following a Navarro et al. (1997) profile and are given random velocities relative to the velocity of the halo. These velocities are drawn from an isotropic Maxwell-Boltzmann distribution with a line-of-sight velocity dispersion,  $\sigma^2 (M_{200})$ , given by,

$$\sigma^2 (M_{200}) = \frac{GM_{200}}{2R_{200}}, \quad (3)$$

where  $R_{200}$  is the radius of a sphere, centred on the halo centre-of-mass, within which the enclosed density is 200 times the mean density of the Universe. The luminosities of the satellite galaxies are assigned using random number generation in the same manner as was done for assigning the luminosities of the central galaxies.

### 2.3 Catalogue specifications

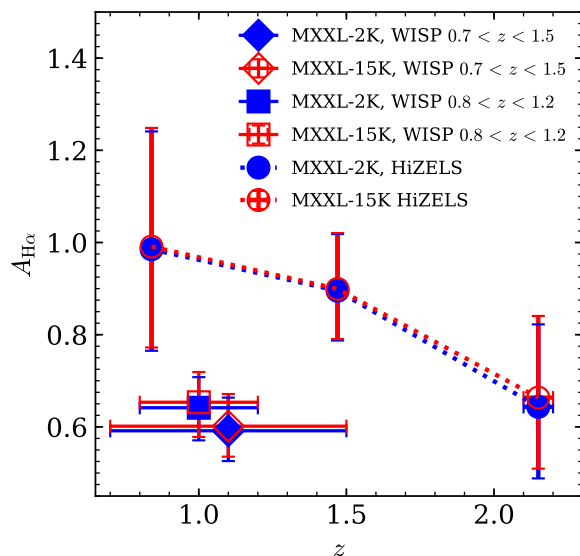
We employ the methodology laid out above to construct two lightcone catalogues: one covering an area of 15 470 deg<sup>2</sup> (MXXL-15K), and another covering a smaller area of 2 000 deg<sup>2</sup> (MXXL-2K). Note that in reality the MXXL-15K lightcone covers  $3/8^{\text{ths}}$  of the sky. The specifications for these two lightcones are shown in Table 1. With these two lightcones we are able to simulate a Euclid-like survey and a WFIRST-like survey (see § 4.1).

At this point we note that due to the wide-angle area of the lightcones and their redshift extent, as well as the finite volume of the MXXL simulation, we expect these lightcones to contain repeated structures. Given the cosmology of the simulation, Smith et al. (2017) determined that an observer placed at the very centre of the box would not observe any repeated structures for  $z \lesssim 0.5$ . Therefore we expect to observe repeated structures, although we do not expect this to significantly affect our clustering measurements. However, we should expect this feature to lead to underestimates in the uncertainties on our clustering measurements, and therefore the uncertainties on the bias estimates. Note that those effects should be negligible, since the mass density variance is very small on scales larger than  $3 h^{-1} \text{Gpc}$ .

## 3 MOCK CATALOGUE CALIBRATION

Before we can compute the galaxy bias it is necessary to calibrate the lightcone catalogues so that they reproduce, as closely as possible, existing observations of H $\alpha$ -emitting galaxies. To calibrate the mock catalogue we apply dust attenuation to the galaxy luminosities, using  $\chi^2$  minimisation to determine the dust attenuation,  $A_{\text{H}\alpha}$ , that leads to the best agreement with a particular observational dataset. The dust attenuation  $A_{\text{H}\alpha}$  relates the un-attenuated luminosity,  $L_{\text{H}\alpha}^0$ , and attenuated luminosity,  $L_{\text{H}\alpha}^{\text{att}}$ , according to,

$$\log_{10} (L_{\text{H}\alpha}^{\text{att}}) = \log_{10} (L_{\text{H}\alpha}^0) - 0.4A_{\text{H}\alpha}. \quad (4)$$

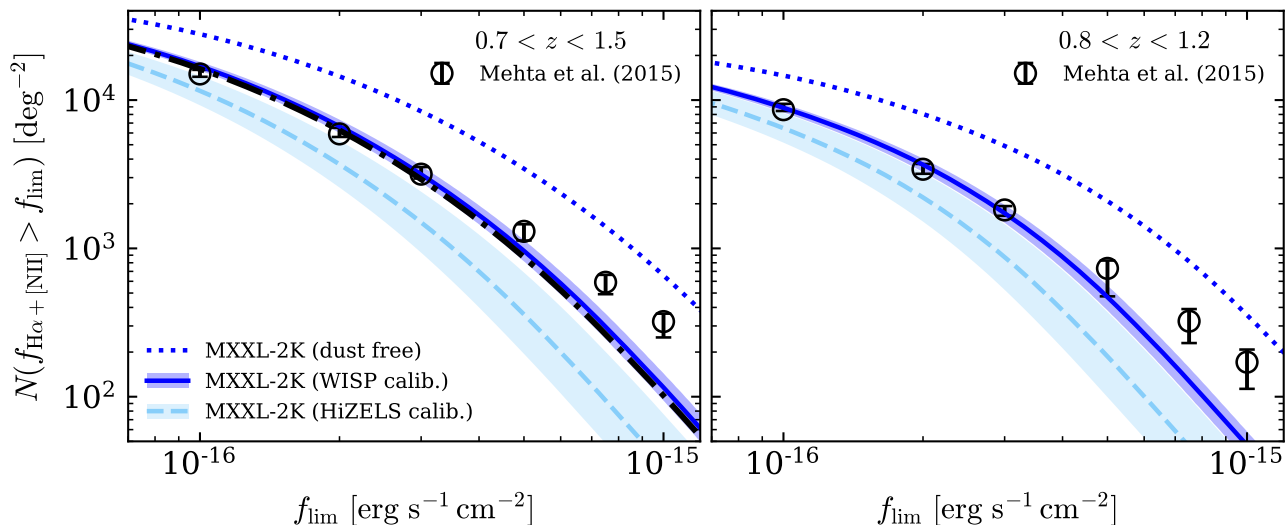


**Figure 5.** Dust attenuation values,  $A_{\text{H}\alpha}$ , recovered from our calibration procedure. Filled symbols correspond to the MXXL-2K lightcone, whilst empty symbols correspond to the MXXL-15K lightcone. Circles show the attenuation values required for a lightcone to reproduce the HiZELS luminosity functions, squares show the attenuation values required to reproduce the WISP number counts in the redshift range  $0.8 < z < 1.2$  and diamonds show the attenuation values required to reproduce the WISP number counts in the redshift range  $0.7 < z < 1.5$ . Uncertainties in the attenuation correspond to the range of  $A_{\text{H}\alpha}$  values whose  $\chi^2$  value is within one standard deviation of the minimum  $\chi^2$  value. Error bars in the redshift direction indicate the width of the redshift range used to compute the counts or luminosity function.

Here we choose to calibrate the catalogues to match the luminosity functions and cumulative number counts of H $\alpha$ -emitters. Being able to reproduce the luminosity function helps ensure that we have the correct number density, whilst reproducing the cumulative number counts confirms that we have the correct total number of galaxies (e.g. Pozzetti et al. 2016; Valentino et al. 2017; Merson et al. 2018).

### 3.1 Observational datasets

For calibration of the MXXL-15K and MXXL-2K lightcones we adopt (i) the cumulative number counts of H $\alpha$ -emitters from the Wide Field Camera 3 (WFC3) Infrared Spectroscopic Parallels Survey (WISP, Atek et al. 2010, 2011) as measured by Mehta et al. (2015), and (ii) the luminosity function of H $\alpha$ -emitters from HiZELS as measured by Sobral et al. (2013).



**Figure 4.** Cumulative number counts measured from the MXXL-2K lightcone over the redshift range  $0.7 < z < 1.5$  (left-hand panel) and the redshift range  $0.8 < z < 1.2$  (right-hand panel). Black data points show the WISP number counts measured by Mehta et al. (2015). In each panel the blue dotted line shows the MXXL-2K number counts assuming no dust attenuation. The blue solid lines show the number counts for the MXXL galaxies when assuming a dust attenuation,  $A_{H\alpha}$ , that leads to the best match to the WISP counts in that redshift range. In the left-hand panel the black dot-dashed line shows the number counts obtained when assuming the value for  $A_{H\alpha}$  that leads to the best match to counts over the redshift range  $0.8 < z < 1.2$ . The light blue dashed lines show the number counts for the MXXL galaxies in the corresponding redshift range assuming a HiZELS-calibrated dust attenuation. Shaded regions show the spread in the counts given the corresponding uncertainty in  $A_{H\alpha}$  (see text for details). All fluxes correspond to blended  $H\alpha + [NII]$  fluxes.

The WISP survey is a pure-parallel near-infrared grism spectroscopic survey that was carried out using the G141 ( $1.2 - 1.7\mu\text{m}$ ,  $R \sim 130$ ) and G102 ( $0.8 - 1.2\mu\text{m}$ ,  $R \sim 210$ ) grisms on the Hubble Space Telescope WFC3. Given the wavelength ranges of these grisms, WISP is able to directly detect the  $H\alpha$  emission line in galaxies out to  $z \lesssim 1.5$ . Note that the Euclid and WFIRST observing strategies will be comparable to the WISP observing strategy. The survey targeted emission line galaxies in several hundred high-latitude fields. Estimates for the stellar mass of a subset of the WISP galaxies analysed by Domínguez et al. (2013), suggest these galaxies have stellar masses in the approximate range  $10^7 h^{-1} M_{\odot} \lesssim M_{\star} \lesssim 10^{11} h^{-1} M_{\odot}$ , assuming a Chabrier (2003) initial mass function. For calibration we use the most recent published cumulative number counts measured by Mehta et al. (2015) who in their analysis used galaxies from 52 separate fields covering an area of  $182 \text{ arcmin}^2$  with a quality flag  $< 16$  down to a flux limit of  $(3 - 5) \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$ . Specifically, the counts that we use correspond to the cumulative counts of  $H\alpha$ -emitters without any correction for  $[NII]$  contamination, i.e. counting the galaxies according to their  $H\alpha + [NII]$  blended flux (see Table 5 of Mehta et al. 2015). We consider the blended flux counts in the redshift ranges  $0.7 < z < 1.5$  and  $0.8 < z < 1.2$ .

In contrast to the WISP survey, HiZELS is a ground-based panoramic survey of emission line galaxies covering several square degrees that was carried out on the United Kingdom Infrared Telescope, the Subaru Telescope and the Very Large Telescope. Custom-made  $NB_J$ ,  $NB_H$ ,  $NB_K$  and NB921 narrow band filters were employed to target emission line galaxies out to  $z \sim 9$ . The specifications of the narrow band filters allowed  $H\alpha$ -emitters to be detected in thin redshift slices centred approximately at  $z = 0.4, 0.84, 1.47$  and  $2.23$ . Further details regarding the survey design can be found in Geach et al. (2008), Sobral et al. (2009), Sobral et al. (2012) and Sobral et al. (2013). Estimates by Cochrane et al. (2018) of the stellar masses of the subset of HiZELS galaxies used in their

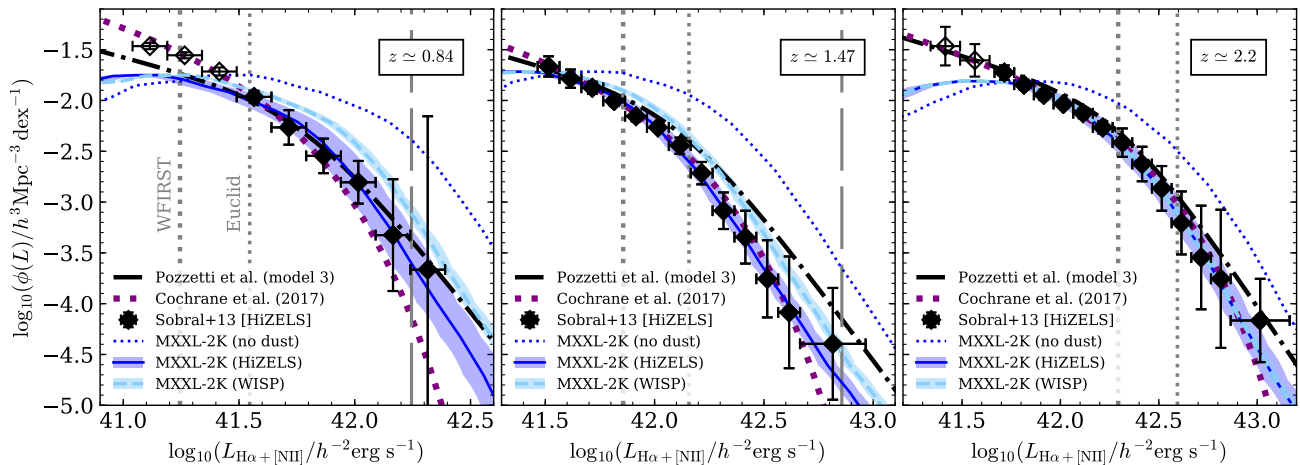
clustering analysis suggest that these galaxies have stellar masses in the approximate range  $10^8 h^{-1} M_{\odot} \lesssim M_{\star} \lesssim 10^{11} h^{-1} M_{\odot}$ , assuming a Chabrier (2003) initial mass function. For our calibration we use the completeness corrected luminosity functions of  $H\alpha$ -emitters presented in Table 4 of Sobral et al. (2013). For their analysis, Sobral et al. (2013) selected  $H\alpha$ -emitters using a combination of colour-colour selections and photometric redshifts when available.

### 3.2 Number counts calibration

As mentioned in the previous subsection, for calibrating the number counts of the lightcone we use the WISP cumulative number counts from Mehta et al. (2015), who measured the number counts in two broad redshift bins,  $0.7 < z < 1.5$  and  $0.8 < z < 1.2$ . For each of these redshift bins we compute the number counts using the dust-free luminosities from the lightcone catalogue and use  $\chi^2$  minimisation to determine the optimum value for  $A_{H\alpha}$  needed to shift these counts into agreement with the observations. Note that this approach assumes that  $A_{H\alpha}$  has a single fixed value for each redshift bin. We note that Domínguez et al. (2013) observe a dust attenuation that varies with observed  $H\alpha$  luminosity, though their attenuation estimates are consistent with a fixed value attenuation that is independent of luminosity (see also Merson et al. 2018).

The results of our calibration procedure are shown in Fig. 4 where we compare the number counts for the MXXL-2K lightcone to the observed cumulative number counts from Mehta et al. (2015) in the redshift ranges  $0.7 < z < 1.5$  (left-hand panel) and  $0.8 < z < 1.2$  (right-hand panel). In each panel the dotted blue line shows the number counts assuming the dust-free luminosities from the lightcone, and the solid blue line shows the number counts for the value of  $A_{H\alpha}$  that leads to the best match to the counts in that particular redshift bin, i.e. the value that minimises the  $\chi^2$  statistic. The blue shaded region shows the range of allowed number





**Figure 6.** Luminosity functions for  $H\alpha$ -emitting galaxies as measured from the MXXL-2K lightcone and compared to the HiZELS luminosity functions (Sobral et al. 2013) measured in three redshift slices centred at:  $z \simeq 0.84$  (left-hand panel),  $z \simeq 1.47$  (centre panel) and  $z \simeq 2.2$  (right-hand panel). In each panel the blue dotted line shows the luminosity functions of the MXXL galaxies computed using the dust-free luminosities. The solid blue line shows the luminosity function for the MXXL galaxies computed assuming the dust attenuation,  $A_{H\alpha}$ , that leads to the best match to the HiZELS luminosity function at that particular redshift. The symbols show the HiZELS observations from Sobral et al. (2013), with the empty symbols corresponding to the luminosity bins that were excluded from the  $\chi^2$  minimisation procedure due to incompleteness of the MXXL-2K lightcone. The light blue dashed line shows the luminosity function obtained assuming the dust attenuation obtained from calibration to the  $0.8 < z < 1.2$  WISP number counts. Shaded regions indicate the range of allowable luminosity functions given the uncertainty on the dust attenuation  $A_{H\alpha}$  (see text for details). The black dot-dashed line shows the luminosity function predicted by the third of the three empirical models presented by Pozzetti et al. (2016). The thick dotted purple line shows the Schechter (1976) fits from Cochrane et al. (2017). All luminosities correspond to blended  $H\alpha + [N_{II}]$  luminosities. The HiZELS and Pozzetti et al. (2016) luminosities have been shifted by one magnitude faintwards to re-introduce dust attenuation and boosted by an additional factor of  $4/3$  to account for  $[N_{II}]$  contamination. Vertical dotted lines indicate the luminosities that correspond to the Euclid and WFIRST flux limits. Vertical dashed lines indicate the luminosities corresponding to a flux of  $1 \times 10^{15} \text{erg s}^{-1} \text{cm}^{-2}$ .

counts given the uncertainty on  $A_{H\alpha}$ , which we define as the range of  $A_{H\alpha}$  values whose  $\chi^2$  value is within one standard deviation of the minimum  $\chi^2$  value. We see that for both redshift bins the calibrated counts from the lightcone are in excellent agreement with the counts, particularly at faint flux limits. Towards brighter flux limits, however, the lightcone under-predicts the number of galaxies. Since the number of bright  $H\alpha$ -emitters is only a very small fraction of the samples for Euclid and WFIRST, we expect this deficit to have negligible effect on our results. Thus, for a Euclid-like flux limit of  $2 \times 10^{-16} \text{erg s}^{-1} \text{cm}^{-2}$ , or a WFIRST-like flux limit of  $1 \times 10^{-16} \text{erg s}^{-1} \text{cm}^{-2}$ , the lightcone shows a deficit of bright galaxies but predicts the correct overall number of galaxies. We do not show the counts for the MXXL-15K lightcone but confirm that they are in equally good agreement with the WISP observations at the MXXL-2K lightcone.

The WISP-calibrated values for  $A_{H\alpha}$  are shown in Fig. 5, where the diamond symbols indicate the calibration results for the  $0.7 < z < 1.5$  redshift bin and square symbols indicate the calibration results for the  $0.8 < z < 1.2$  redshift bin. Filled symbols correspond to the results for the MXXL-2K lightcone and empty symbols correspond to the results for the MXXL-15K lightcone. For both redshift bins the calibration results for the MXXL-2K lightcone are in excellent agreement with the results for the MXXL-15K lightcone. Furthermore, for both lightcones the value for  $A_{H\alpha}$  obtained by calibrating to the  $0.7 < z < 1.5$  counts is consistent within error with the value obtained by calibrating to the  $0.8 < z < 1.2$  counts. Therefore, adopting the calibration result for either redshift bin has negligible impact on the counts. This can be seen in the left-hand panel of Fig. 4 where the black dot-dashed line shows the counts over the redshift range  $0.7 < z < 1.5$  computed assuming the value of  $A_{H\alpha}$  obtained by calibrating to the counts

in the  $0.8 < z < 1.2$  bin. Going forward, we therefore adopt the results from the calibration to the counts in the  $0.8 < z < 1.2$  bin, which we note in the next section are marginally closer in value to the dust attenuation values obtained from calibrating the lightcone luminosity function. The calibration using the luminosity function, which yields the HiZELS attenuation factors, is discussed in the next subsection.

### 3.3 Luminosity function calibration

For calibrating the lightcone luminosity function we use the HiZELS luminosity functions, as presented by Sobral et al. (2013), measured at  $z = 0.84$ ,  $z = 1.47$  and  $z = 2.23$ . For each redshift, we compute the dust-free luminosity function from the lightcone using a thin redshift slice centred on the appropriate HiZELS redshift. Note that since  $z \simeq 2.23$  is beyond the upper redshift extent of our lightcone, in this case we use a thin redshift slice with  $z = 2.2$  as the upper bound. As with the cumulative counts, we use  $\chi^2$  minimisation to determine the optimum value for  $A_{H\alpha}$  needed to shift these luminosity functions into agreement with the observational estimates. Note that prior to our calibration we have adjusted the HiZELS luminosity functions in two ways. Firstly, Sobral et al. (2013) originally corrected for dust attenuation by shifting the  $H\alpha$  luminosity bright-wards by  $A_{H\alpha} = 1 \text{ mag}$  (0.4 dex). Therefore, to ensure that we are comparing to the dust-attenuated HiZELS luminosity function we have removed this shift and made the luminosities fainter by 0.4 dex. Secondly, the HiZELS luminosity functions assumed  $H\alpha$  luminosities that had been corrected to remove  $[N_{II}]$  contamination and so for our calibration we re-introduce the  $[N_{II}]$  contamination. We do this by boosting the HiZELS luminosity function bright-wards by a factor of  $4/3$ . This factor was chosen

to match the median correction of  $[N_{\text{H}\alpha}] / (H\alpha + [N_{\text{H}\alpha}]) \approx 0.25$  reported by Sobral et al. (2013). Note that this correction factor shows some variation with redshift (with this correction being slightly smaller at higher redshift), though this variation is negligible in size.

The calibrated luminosity functions for the MXXL-2K lightcone are shown in Fig. 6. In each panel the blue dotted line shows the luminosity function computed using the dust-free luminosities and the symbols correspond to the observed HiZELS luminosity functions, with the open symbols corresponding to luminosity bins that were excluded in our calibration procedure. These bins were left out following a preliminary calibration that indicated that the MXXL-2K lightcone is incomplete at these faint luminosities and so including these luminosity bins would unfairly bias our calibration results. The two vertical lines indicate the luminosities at each redshift that correspond to the Euclid and WFIRST flux limits. For reference we also show the Schechter (1976) functional fits presented by Cochrane et al. (2017) and the luminosity functions predicted by the third of the three empirical models presented by Pozzetti et al. (2016), both which we have similarly corrected by a factor of  $4/3$  in order to introduce  $[N_{\text{H}\alpha}]$  contamination and applied an attenuation of  $A_{\text{H}\alpha} = 1$  mag.

In each panel the solid blue lines show the MXXL-2K luminosity function for the value of  $A_{\text{H}\alpha}$  that minimised the  $\chi^2$  statistic and leads to the best match to the HiZELS luminosity function at that particular redshift. The blue shaded region shows the range of allowed luminosity functions given the uncertainty on  $A_{\text{H}\alpha}$ , which again is defined as the range of  $A_{\text{H}\alpha}$  values whose  $\chi^2$  value is within one standard deviation of the minimum  $\chi^2$  value. We can see that at each redshift the MXXL-2K lightcone is able to reproduce very well the HiZELS luminosity function brightwards of the Euclid flux limit. For  $z \simeq 1.47$  and  $z \simeq 2.2$  the lightcone provides a good match to the HiZELS luminosity function down to the WFIRST flux limit. At  $z \simeq 0.84$  we see the impact of incompleteness in the lightcone luminosity function as it falls below the observations for luminosities faintwards of the Euclid flux limit. We do not expect this to have a significant impact our clustering analysis however as we note that  $z \simeq 0.84$  is below the lower redshift limit of WFIRST. We do not show the luminosity functions for the MXXL-15K lightcone but can confirm that they are in equally good agreement to the HiZELS observations as the MXXL-2K lightcone.

The dust attenuation values for our luminosity function calibration are also shown with circular symbols in Fig. 5. As with the number counts calibration, the attenuations obtained for the MXXL-15K and the MXXL-2K lightcones are in excellent agreement. The values for  $A_{\text{H}\alpha}$  from our luminosity function calibration are much larger than the values obtained from the WISP number counts calibration, particularly at low redshift. However, we note that at  $z \simeq 0.84$  and  $z \simeq 2.2$  the  $A_{\text{H}\alpha}$  values that we have obtained are consistent with the  $A_{\text{H}\alpha} = 1$  mag that was originally assumed by Sobral et al. (2013). Using the median  $[O_{\text{II}}]/H\alpha$  line ratio as a dust attenuation indicator (see Sobral et al. 2012 for details), Sobral et al. (2013) estimated that their  $z \simeq 1.47$  galaxy sample should have an attenuation of  $A_{\text{H}\alpha} = 0.8$  mag, which, as seen in Fig. 5, is consistent within error with our  $z \simeq 1.47$  calibration result. Additionally, Garn et al. (2010) determined that  $A_{\text{H}\alpha} \approx 1.2$  at  $z \simeq 0.84$ , which is also consistent within the errors with our calibration result at  $z \simeq 0.84$ . At  $z \simeq 2.2$  our luminosity function calibrated attenuation is consistent with attenuation obtained from our calibration to the WISP number counts.

We note that Cochrane et al. (2017) presented updated Schechter (1976) functional fits to the measured HiZELS luminos-

ity functions, which at  $z = 0.84$  differ from the functional fits presented in Sobral et al. (2013). It is clear from Fig. 6, where the Cochrane et al. (2017) fits are shown as thick dotted lines, that calibrating the MXXL lightcones to reproduce the Cochrane et al. (2017) fits would require a dust attenuation that is equal to or stronger than the attenuation required to reproduce the Sobral et al. (2013) measurements. Adopting the Cochrane et al. (2017) fits, instead of the Sobral et al. (2013) measurements, may therefore widen the discrepancy between the dust attenuation needed to reproduce the WISP number counts and the HiZELS luminosity functions.

To see how much of an impact the discrepancy between the WISP-calibrated attenuation and the HiZELS-calibrated attenuation has on the luminosity function, we show in Fig. 6 the luminosity function obtained when assuming the dust attenuation  $A_{\text{H}\alpha}$  calibrated to reproduce the WISP counts in the redshift range  $0.8 < z < 1.2$ . Using the WISP-calibrated attenuation leads to the lightcone over-predicting the luminosity function, particularly for bright luminosities at low redshift. At  $z \simeq 2.2$  the luminosity function assuming the HiZELS-calibrated attenuation and the luminosity function assuming the WISP-calibrated attenuation are virtually identical. For  $z \simeq 0.84$  and  $z \simeq 1.47$  the difference between the HiZELS-calibrated luminosity function and the WISP-calibrated luminosity function is negligible for luminosities corresponding to the Euclid and WFIRST flux limits. We additionally show in Fig. 4 the cumulative number counts obtained when assuming the HiZELS-calibrated dust attenuation. As expected, the HiZELS-calibrated counts are lower than the WISP-calibrated counts, but the discrepancy decreases towards fainter flux limits, such that at the Euclid and WFIRST flux limits the discrepancy is less than a factor of two.

### 3.4 Discussion of calibration results

Why the calibration to the WISP number counts and the calibration to the HiZELS luminosity functions lead to different values for the dust attenuation is not immediately clear. One possibility is that the WISP number counts and the HiZELS luminosity functions are probing different luminosity ranges. As shown in Fig. 4, the WISP cumulative number counts from Mehta et al. (2015) have been measured at flux limits spanning  $1 \times 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$  to  $1 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$ . Note that the fainter limit corresponds to the WFIRST flux limit. These two flux limits have been plotted as vertical dotted and dashed lines in the panels of Fig. 6. We can see that at  $z \simeq 0.84$  and  $z \simeq 1.47$  this flux range covers most of the luminosity range measured by the Sobral et al. (2013) luminosity functions, with only the faintest few luminosity bins outside of the flux range. This would suggest that the small difference in luminosity ranges probed by the cumulative counts and the luminosity functions is less likely to be the cause of the discrepancy. At  $z \simeq 2.2$  the flux range is restricted to brighter luminosities, though at this redshift the WISP-calibrated attenuations and HiZELS-calibrated attenuations are consistent.

The most likely cause of the discrepancy is the different observing strategies and selection functions employed by the WISP and HiZELS surveys. For example, narrow band surveys, such as HiZELS, rely on continuum detection to identify emission lines which means that these surveys typically miss the lowest mass galaxies that spectroscopic grism surveys, such as WISP, are more sensitive to (Mehta et al. 2015). As such, for a given  $H\alpha$  flux limit the galaxies detected by WISP will typically have a lower stellar mass compared to those detected by HiZELS (J. Colbert, pri-

vate communication). Therefore, given the relations observed between stellar mass and dust attenuation  $A_{H\alpha}$  (Garn & Best 2010; Domínguez et al. 2013), we expect the WISP dataset to typically have a weaker dust attenuation, as we can see in Fig. 5. Additionally, the thin-shell volumes probed by narrow band surveys mean that volume densities estimated by such surveys are sensitive to cosmic variance introduced by the presence of large-scale structure (Mehta et al. 2015). As such, cosmic variance may also be contributing to the discrepancy in dust attenuation that our calibration procedure has returned.

Using the HiZELS-calibrated attenuation results in Fig. 5 we can construct a toy model of how the attenuation evolves with redshift, such that we can apply a HiZELS-calibrated attenuation to any galaxy in the mock catalogue by interpolating over these results as a function of redshift. Finally, we note that in our toy model we assume that  $A_{H\alpha}$  depends only on redshift and not additionally on the  $H\alpha$  luminosity. However, for our purposes we argue that adding a luminosity-dependence is not necessary given that we are able to reproduce the HiZELS luminosity functions very well down to the luminosity limits of relevance for Euclid and WFIRST.

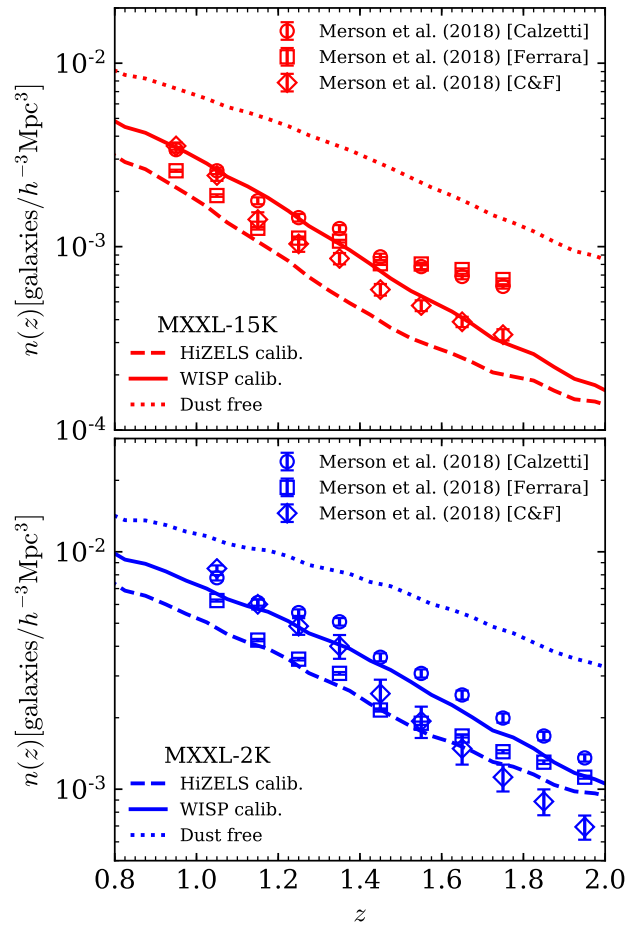
In summary, given the difference in the dust attenuation required to reproduce the WISP number counts and the HiZELS luminosity functions, we choose to carry out our bias forecasts using both a WISP-calibrated version of the lightcones (adopting the values for calibration to the counts in the  $0.8 < z < 1.2$  bin) and a HiZELS-calibrated version of the lightcones.

### 3.5 Galaxy number densities

Before progressing to examine the clustering of the galaxies in the lightcones, we examine the number densities of the galaxies as a check that our calibration yields number densities that are consistent with previous estimates.

In Fig. 7 we show in the upper panel the number densities for the galaxies in the MXXL-15K lightcone, and in the lower panel the number densities for the galaxies in the MXXL-2K lightcone. We have applied a flux limit of  $2 \times 10^{-16} \text{erg s}^{-1} \text{cm}^{-2}$  to the dust-attenuated luminosities in the MXXL-15K lightcone and a flux limit of  $1 \times 10^{-16} \text{erg s}^{-1} \text{cm}^{-2}$  to the dust-attenuated luminosities in the MXXL-2K lightcone. In both panels the solid line shows the number densities when adopting the WISP-calibrated dust attenuations and the dashed line shows the number densities when adopting the HiZELS-calibrated dust attenuations. The dotted line corresponds to applying no dust attenuation. The symbols correspond to the number densities presented in Merson et al. (2018) obtained by applying three different dust attenuation methods to a GALACTICUS lightcone. The dust attenuation methods considered in Merson et al. (2018) were: (i) interpolation over the library of dust curves from Ferrara et al. (1999), (ii) the Charlot & Fall (2000) model where the dust attenuation follows a power law with wavelength, and (iii) the Calzetti et al. (2000) dust screen law.

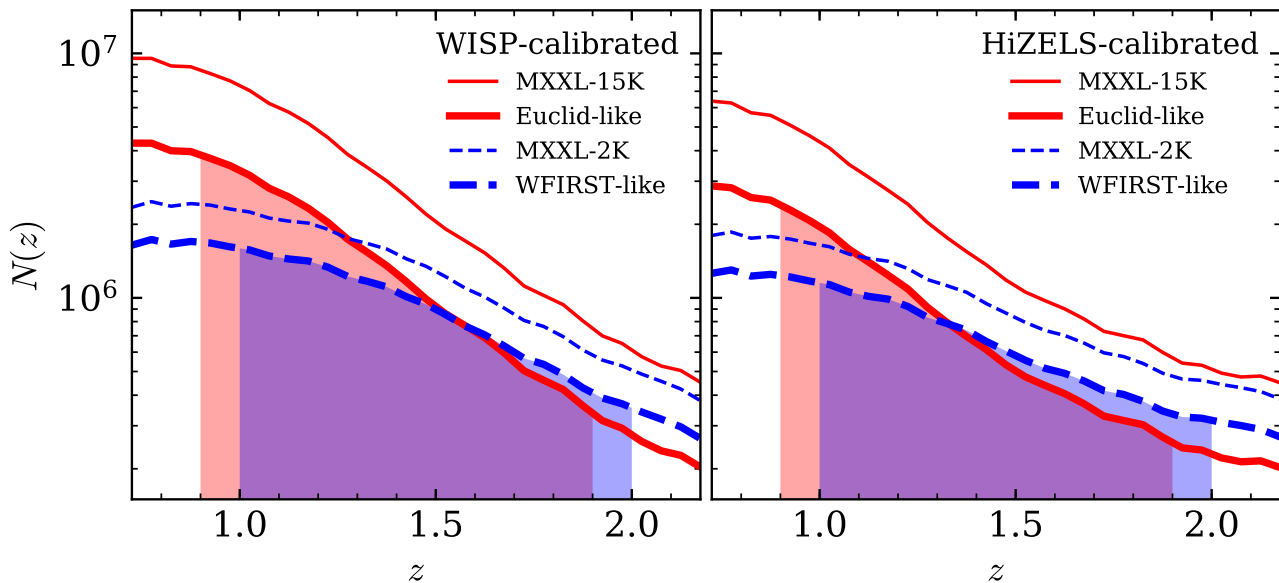
Considering first the MXXL-15K lightcone we see that when adopting a WISP-calibrated dust attenuation, the galaxy number density is consistent with the results from Merson et al. (2018). Adopting a HiZELS-calibrated dust attenuation leads to a number density that is slightly below the Merson et al. (2018) results, but typically remains within approximately a factor of two. The fact that the number densities for the WISP-calibrated attenuations are in better agreement than the HiZELS-calibrated attenuations is not surprising as the GALACTICUS lightcone in Merson et al. (2018) was calibrated to reproduce the WISP number counts from Mehta et al. (2015). Turning to the MXXL-2K lightcone we see that



**Figure 7.** Number density per unit volume of  $H\alpha$ -emitting galaxies in the MXXL-15K lightcone (upper panel) and MXXL-2K lightcone (lower panel). A flux limit of  $2 \times 10^{-16} \text{erg s}^{-1} \text{cm}^{-2}$  has been applied to the dust-attenuated luminosities in the MXXL-15K lightcone and a limit of  $1 \times 10^{-16} \text{erg s}^{-1} \text{cm}^{-2}$  to the dust-attenuated luminosities in the MXXL-2K lightcone. In each panel the solid line shows the number density obtained when adopting the WISP-calibrated dust attenuations, and the dashed line shows the number density obtained when adopting a HiZELS-calibrated dust attenuation. The dotted line shows the number density of the corresponding lightcone when no dust attenuation is applied. The various symbols show the predicted Euclid and WFIRST number densities from the GALACTICUS catalogues presented in Merson et al. (2018), who considered three different dust attenuation methods: Ferrara et al. (1999), Calzetti et al. (2000), and Charlot & Fall (2000, C&F). All flux selections assume a blended  $H\alpha + [N_{II}]$  flux.

the number densities for the WISP-calibrated attenuations and the HiZELS-calibrated attenuations are both consistent with the number densities from Merson et al. (2018), with the difference between the two being consistent in size with the scatter between the number densities predicted for the different dust methods used in Merson et al. (2018). The redshift distributions for the MXXL-15K and MXXL-2K lightcones are provided in Appendix B.

Overall, when we apply the WISP-calibrated and HiZELS-calibrated dust attenuations to the MXXL-15K and MXXL-2K lightcones, the number densities that we obtain are entirely consistent with previous number density estimates.



**Figure 8.** Redshift distributions for our Euclid-like and WFIRST-like surveys, assuming a WISP-calibrated dust attenuation (left-hand panel) and a HiZELS-calibrated dust attenuation (right-hand panel). Thin lines show the redshift distribution of the galaxies in the corresponding MXXL lightcone and thick lines show the redshift distribution for the Euclid-like and WFIRST-like surveys. Specifications of the lightcones and Euclid-like and WFIRST-like surveys are provided in Table 1. The shaded regions indicate the redshift range that is considered in our clustering analysis:  $0.9 \leq z < 1.9$  for our Euclid-like survey and  $1 \leq z < 2$  for our WFIRST-like survey.

#### 4 LINEAR BIAS FORECASTS

Having calibrated our lightcone catalogues we now proceed to make forecasts for the linear bias of  $H\alpha$ -emitters as a function of redshift for a Euclid-like survey and a WFIRST-like survey. To calculate the linear bias we will first need to compute the correlation functions for the galaxies in our catalogues.

##### 4.1 Catalogue preparation

The first step is to build catalogues for our Euclid-like and WFIRST-like surveys. We firstly apply a flux limit to the dust attenuated luminosities from the appropriate lightcone. For our Euclid-like surveys we apply a flux limit of  $2 \times 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$  to select galaxies from the MXXL-15K lightcones and for our WFIRST-like surveys we apply a flux limit of  $1 \times 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$  to select galaxies from the MXXL-2K lightcones.

Secondly, we introduce incompleteness into the surveys by applying the appropriate completeness expected for each redshift survey: 45 per cent for Euclid (Laureijs et al. 2011) and 70 per cent for WFIRST (Spergel et al. 2015). Both surveys have completeness significantly below 100 per cent due to slitless effects. This incompleteness will arise in the Euclid and WFIRST galaxy redshift surveys due to the inability to measure redshifts for faint galaxies with particularly noisy spectra, or for galaxies in high density regions where slitless spectra will overlap and could be confused.

Incompleteness is introduced by using random number generation to discard the appropriate fraction of galaxies. Note that we simply draw random numbers from a uniform distribution and make no attempt to introduce a density-dependent or luminosity-dependent incompleteness. For our Euclid-like surveys we apply a completeness limit of 0.45, consistent with the completeness limit estimated in Laureijs et al. (2011), and reject a random 55 per cent of the galaxies. For our WFIRST-like surveys we apply a complete-

ness limit of 0.7, consistent with the completeness limit provided in Spergel et al. (2015), and reject a random 30 per cent of the galaxies. The specifications for our Euclid-like and WFIRST-like surveys are listed in Table 1 alongside the specifications of the MXXL-15K and MXXL-2K lightcones. In Fig. 8 we show the redshift distributions for our Euclid-like and WFIRST-like surveys, as well as for the MXXL-15K and the MXXL-2K lightcones.

In our clustering analysis we consider two versions of a Euclid-like survey and WFIRST-like survey: one calibrated to reproduce the WISP cumulative number counts, and one calibrated to reproduce the HiZELS luminosity functions. Additionally, for each version we repeat the analysis twice. In the first instance we take the observed redshifts of the galaxies (i.e. with peculiar velocities included), which we shall refer to as *redshift-space*, and in the second instance we take the cosmological redshifts of the galaxies (i.e. with peculiar velocities removed), which we will refer to as *real-space*. Note that galaxies are not observed in real-space, but we include this analysis to show the impact of redshift-space distortions.

##### 4.2 Clustering analysis

For our clustering analysis we split each of our Euclid-like and WFIRST-like surveys into 5 equal redshift bins of width  $\Delta z = 0.2$  and compute the two-point galaxy correlation function in each bin. For the Euclid-like surveys these bins cover the redshift range  $0.9 \leq z < 1.9$ , and for the WFIRST-like surveys these bins cover the redshift range  $1 \leq z < 2$ .

In order to compute the galaxy correlation functions we generate individual catalogues of randoms for each redshift bin. In each instance we generate a sufficiently large number of randoms (at least ten times the number of galaxies in the redshift bin), with the number chosen arbitrarily such that adding any additional ran-

doms does not improve our recovery of the galaxy correlation function. The number of randoms required increases with redshift as the number density of galaxies decreases. For the lower redshifts the number of randoms were also limited by available computing resources. The right ascension and declination of the randoms were found by sampling uniform distributions bounded by the right ascension limits and the cosine of the declination limits of the particular lightcone. To assign redshifts to the randoms we do not sample a uniform distribution as this would not reproduce the correct shape of the galaxy redshift distribution across the bin. Instead we assign redshifts by first replicating the redshifts of the galaxies, until we match the required number of randoms, and then randomly shuffling the galaxy redshifts.

We compute the angle-averaged galaxy two-point correlation function,  $\xi(r)$ , using the open-source TREECORR package (Jarvis et al. 2004), and the Landy & Szalay (1993) estimator,

$$\xi_{LS}(r) = \frac{DD(r) - 2DR(r) + RR(r)}{RR(r)}, \quad (5)$$

where  $DD(r)$  are the weighted galaxy-galaxy pair counts,  $DR(r)$  are the weighted galaxy-random pair counts and  $RR(r)$  are the weighted random-random pair counts as a function of separation  $r$  in units of  $h^{-1}\text{Mpc}$ . For each redshift bin the correlation function is calculated 5 times, each time using a different random seed when applying the survey incompleteness (see § 4.1) and using a different catalogue of randomly generated positions. The correlation function is calculated between  $50 h^{-1}\text{Mpc}$  and  $150 h^{-1}\text{Mpc}$  using 48 linearly spaced bins.

In Table 2 we report for our Euclid-like surveys the average number of galaxies in each redshift bin after introducing incompleteness and the effective redshift of these galaxies. We also report the average ratio of the number of galaxies to the number of random positions. The number of randoms was increased with increasing redshift to account for the decreasing number densities of the galaxy samples. Memory limitations of available computing resources forced us to limit the number of randoms, but in each case we ensured that a sufficient number was used to adequately recover the clustering signal of the galaxies. We report equivalent numbers for our WFIRST-like surveys in Table 3.

The results of our clustering analysis are shown in Fig. 9, which assumes WISP-calibrated surveys, and Fig. 10, which assumes HiZELS-calibrated surveys. In each case the upper grid shows the results for the Euclid-like survey and the lower grid shows the results for the WFIRST-like survey. The top-row of each grid shows the correlation function in each redshift bin, with the darker lines showing the redshift-space result and the fainter lines showing the real-space result. The lines show the average result over 5 repeats (each time changing the random seed) and the error bars showing the standard deviation.

We can see that in each instance we are able to recover the BAO peak in every redshift bin, albeit with increased noise in the highest redshift bin for the Euclid-like surveys. However, in each instance the recovery is significant over the statistical noise, particularly in real-space. For  $r \gtrsim 140 h^{-1}\text{Mpc}$  the correlation functions measured in redshift-space have a higher amplitude than the equivalent measurement in real-space, consistent with the expectations of Kaiser (1987), which are based on linear perturbation theory. The impact of redshift-space distortions (RSDs) leads to some smearing out of the BAO peak.

### 4.3 Linear bias measurements and fitting

#### 4.3.1 Linear bias measurements

Given our galaxy clustering results, we can now use Eq. 1 to compute the linear bias in each redshift bin for our Euclid-like and WFIRST-like surveys. Each galaxy correlation function is divided by the non-linear dark matter correlation function computed at the same effective redshift. The dark matter correlation functions, which we compute using the CLASS and HALOFIT functionality in the NBODYKIT python package assuming an MXXL cosmology (see Fig. 1), are shown with black dashed lines in the panels in the upper row of each grid in Fig. 9 and Fig. 10.

The lower row of panels in each grid in Fig. 9 and Fig. 10 show the measured linear bias as a function of separation. As with the correlation functions, the darker lines correspond to the bias estimated using the redshift-space galaxy correlation functions and the fainter lines correspond to the bias estimated using the real-space galaxy correlation functions. The lines show the bias estimate in each bin of separation averaged over the 5 repeat clustering calculations and the error bars show the standard deviation.

Examining first the bias measurements in real-space, we see that for both the Euclid-like and WFIRST-like surveys the measured bias is consistent with a constant value for scales  $r \lesssim 75 h^{-1}\text{Mpc}$ . For scales larger than this we begin to see deviations away from a constant value, particularly for the WFIRST-like survey. In the  $1.4 \leq z \leq 1.6$  redshift bin for the WFIRST-like analysis, we see a significant negative deviation in the measured bias on scales  $75 h^{-1}\text{Mpc} \lesssim r \lesssim 100 h^{-1}\text{Mpc}$ . The cause of this decrease is not clear and is most likely due to cosmic variance. At the largest scales,  $r \gtrsim 125 h^{-1}\text{Mpc}$ , the bias measurements in each redshift bin grow rapidly as the galaxy and matter correlation functions tend toward zero and ultimately become negative due to the integral constraint. If we move to redshift-space we see a similar behaviour, although the deviations away from a constant on scales  $75 h^{-1}\text{Mpc} \lesssim r \lesssim 125 h^{-1}\text{Mpc}$  are now more pronounced than in real-space. In several instances these scale-dependent deviations display a sinusoidal-like shape and in every instance occur around the BAO scale. Similar sinusoidal features can be seen in Hada & Eisenstein (2019) when the authors examine the difference between the correlation function for a mock galaxy field in redshift-space and the correlation function for the initial linear density field (see also the plots showing the scale dependence of the bias of star forming galaxies in Angulo et al. 2014). It is possible that these deviations arise due to a distortion of the galaxy correlation function relative to the correlation function expected from linear theory. This distortion, which is manifested as a smoothing of the BAO peak and slight shift of the peak away from the linear theory position, is understood to arise from a combination of non-linear collapse, mode coupling, redshift-space distortions, as well as the bias itself (Jeong & Komatsu 2006; Eisenstein et al. 2007; Angulo et al. 2008; Crocce & Scoccimarro 2008; Sánchez et al. 2008; Smith et al. 2008; Anselmi et al. 2016). It is possible to model these effects to correct the shape of the BAO in the linear theory predictions (e.g. Crocce & Scoccimarro 2008; Smith et al. 2008).

#### 4.3.2 Linear bias fitting

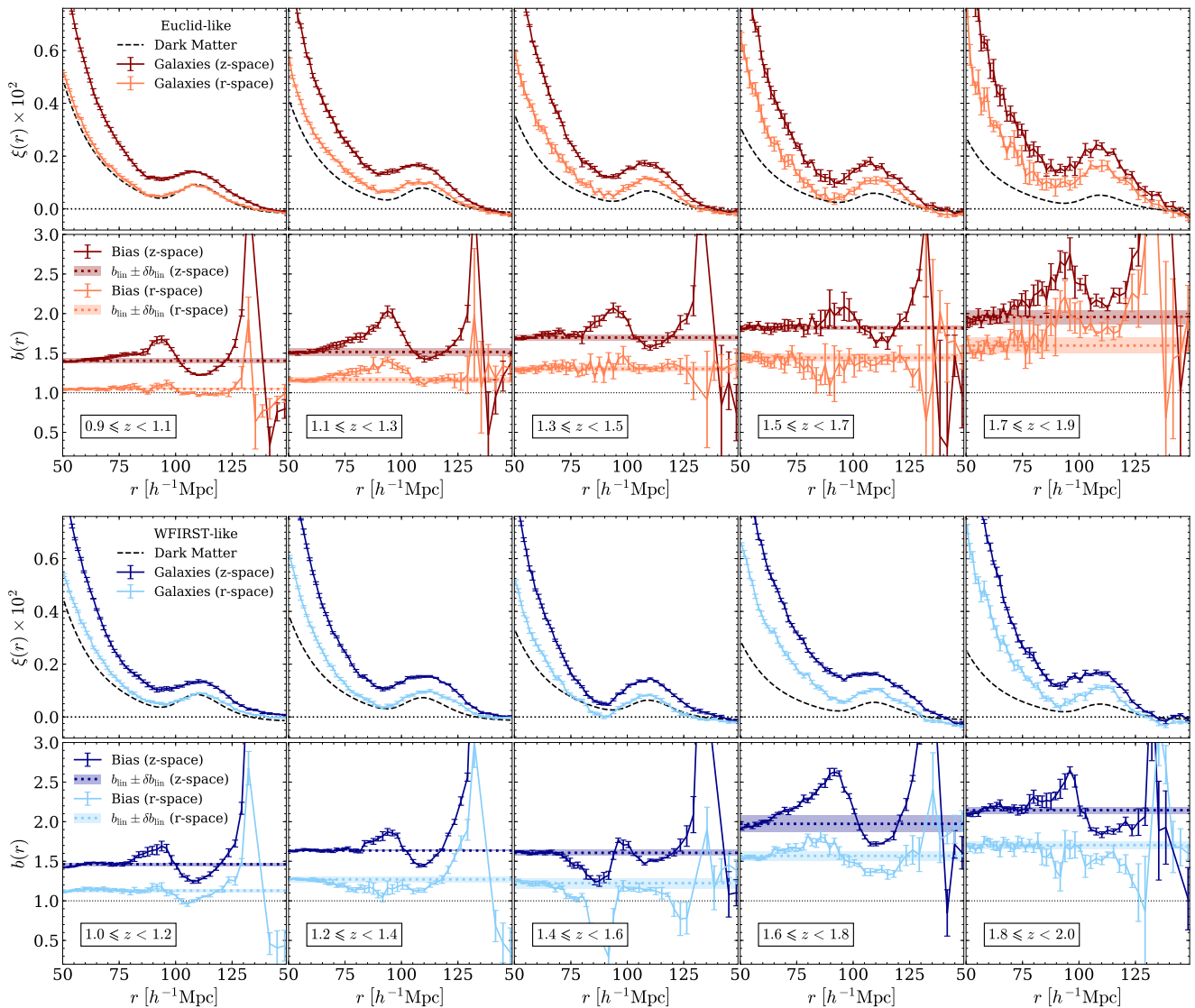
In each redshift bin we attempt to fit the real-space and redshift-space measurements of the bias with a constant, scale-independent value,  $b_{\text{lin}}$ . To do this we use the  $\chi^2$  minimisation to fit a zeroth-order polynomial to the measured bias values  $b(r < r_{\text{cut}})$ . Given

**Table 2.** Catalogue specifications and linear bias fits for our Euclid-like catalogues, measured in redshift-space (upper table) and in real-space (lower table). The properties shown are: the effective redshift of the galaxies in the redshift slice,  $z_{\text{eff}}$ ; the mean number of galaxies used to compute the correlation function,  $\bar{N}_{\text{gal}}$ ; the mean value for the ratio of randoms to galaxies,  $\bar{N}_{\text{ran}}/\bar{N}_{\text{gal}}$ ; the mean and standard deviation in the halo mass,  $M_h$ , for that galaxy sample; and the linear bias fit. The columns show the results for each redshift slice considered. Each table shows the equivalent results when adopting a WISP-calibrated lightcone or a HiZELS-calibrated lightcone.

Calibration	Property	$0.90 \leq z < 1.1$	$1.1 \leq z < 1.3$	$1.3 \leq z < 1.5$	$1.5 \leq z < 1.7$	$1.7 \leq z < 1.9$
<b>Euclid-like (redshift-space)</b>						
WISP	$z_{\text{eff}}$	0.991	1.19	1.38	1.59	1.79
	$\bar{N}_{\text{gal}}$	13175397	8688180	5045749	2912673	1748220
	$\bar{N}_{\text{ran}}/\bar{N}_{\text{gal}}$	10	10	14	25	30
	$\log_{10}(M_h/h^{-1}M_{\odot})$	$11.8 \pm 0.4$	$11.8 \pm 0.3$	$11.8 \pm 0.3$	$11.9 \pm 0.3$	$11.9 \pm 0.3$
	$b_{\text{lin}} \pm \delta b_{\text{lin}}$	$1.40 \pm 0.03$	$1.51 \pm 0.05$	$1.70 \pm 0.04$	$1.82 \pm 0.02$	$1.96 \pm 0.09$
HiZELS	$z_{\text{eff}}$	0.988	1.19	1.39	1.59	1.79
	$\bar{N}_{\text{gal}}$	7785058	4642807	2635798	1687841	1215424
	$\bar{N}_{\text{ran}}/\bar{N}_{\text{gal}}$	10	10	15	25	29
	$\log_{10}(M_h/h^{-1}M_{\odot})$	$11.8 \pm 0.3$	$11.9 \pm 0.3$	$11.9 \pm 0.3$	$11.9 \pm 0.3$	$11.9 \pm 0.3$
	$b_{\text{lin}} \pm \delta b_{\text{lin}}$	$1.42 \pm 0.03$	$1.55 \pm 0.05$	$1.69 \pm 0.04$	$1.88 \pm 0.05$	$1.9 \pm 0.1$
<b>Euclid-like (real-space)</b>						
WISP	$z_{\text{eff}}$	0.991	1.19	1.38	1.59	1.79
	$\bar{N}_{\text{gal}}$	13171311	8682758	5050012	2910250	1748428
	$\bar{N}_{\text{ran}}/\bar{N}_{\text{gal}}$	10	10	15	30	35
	$\log_{10}(M_h/h^{-1}M_{\odot})$	$11.8 \pm 0.4$	$11.8 \pm 0.3$	$11.8 \pm 0.3$	$11.9 \pm 0.3$	$11.9 \pm 0.3$
	$b_{\text{lin}} \pm \delta b_{\text{lin}}$	$1.05 \pm 0.01$	$1.17 \pm 0.04$	$1.30 \pm 0.03$	$1.44 \pm 0.05$	$1.6 \pm 0.1$
HiZELS	$z_{\text{eff}}$	0.988	1.19	1.39	1.59	1.79
	$\bar{N}_{\text{gal}}$	7782514	4639483	2638085	1686470	1215545
	$\bar{N}_{\text{ran}}/\bar{N}_{\text{gal}}$	10	10	14	30	35
	$\log_{10}(M_h/h^{-1}M_{\odot})$	$11.8 \pm 0.3$	$11.9 \pm 0.3$	$11.9 \pm 0.3$	$11.9 \pm 0.3$	$11.9 \pm 0.3$
	$b_{\text{lin}} \pm \delta b_{\text{lin}}$	$1.055 \pm 0.009$	$1.21 \pm 0.04$	$1.32 \pm 0.03$	$1.49 \pm 0.08$	$1.6 \pm 0.1$

**Table 3.** Catalogue specifications and linear bias fits for our WFIRST-like catalogues, measured in redshift-space (upper table) and in real-space (lower table). The meanings of the various properties are the same as in Table 2. The columns show the results for each redshift slice considered. Each table shows the equivalent results when adopting a WISP-calibrated lightcone or a HiZELS-calibrated lightcone.

Calibration	Property	$1.0 \leq z < 1.2$	$1.2 \leq z < 1.4$	$1.4 \leq z < 1.6$	$1.6 \leq z < 1.8$	$1.8 \leq z < 2.0$
<b>WFIRST-like (redshift-space)</b>						
WISP	$z_{\text{eff}}$	1.10	1.29	1.49	1.69	1.89
	$\bar{N}_{\text{gal}}$	5915244	4839323	3576093	2442191	1673105
	$\bar{N}_{\text{ran}}/\bar{N}_{\text{gal}}$	20	20	25	30	40
	$\log_{10}(M_h/h^{-1}M_{\odot})$	$11.7 \pm 0.4$	$11.7 \pm 0.4$	$11.7 \pm 0.4$	$11.8 \pm 0.4$	$11.8 \pm 0.4$
	$b_{\text{lin}} \pm \delta b_{\text{lin}}$	$1.46 \pm 0.02$	$1.63 \pm 0.01$	$1.61 \pm 0.04$	$2.0 \pm 0.1$	$2.15 \pm 0.04$
HiZELS	$z_{\text{eff}}$	1.09	1.29	1.49	1.69	1.89
	$\bar{N}_{\text{gal}}$	4193549	3278049	2341326	1768644	1369560
	$\bar{N}_{\text{ran}}/\bar{N}_{\text{gal}}$	20	20	25	30	40
	$\log_{10}(M_h/h^{-1}M_{\odot})$	$11.7 \pm 0.4$	$11.8 \pm 0.4$	$11.8 \pm 0.3$	$11.8 \pm 0.3$	$11.8 \pm 0.4$
	$b_{\text{lin}} \pm \delta b_{\text{lin}}$	$1.47 \pm 0.02$	$1.64 \pm 0.01$	$1.62 \pm 0.04$	$2.1 \pm 0.1$	$2.14 \pm 0.06$
<b>WFIRST-like (real-space)</b>						
WISP	$z_{\text{eff}}$	1.10	1.29	1.49	1.69	1.89
	$\bar{N}_{\text{gal}}$	5918784	4832447	3578623	2449899	1673483
	$\bar{N}_{\text{ran}}/\bar{N}_{\text{gal}}$	20	20	25	30	40
	$\log_{10}(M_h/h^{-1}M_{\odot})$	$11.7 \pm 0.4$	$11.7 \pm 0.4$	$11.7 \pm 0.4$	$11.8 \pm 0.4$	$11.8 \pm 0.4$
	$b_{\text{lin}} \pm \delta b_{\text{lin}}$	$1.13 \pm 0.02$	$1.27 \pm 0.04$	$1.22 \pm 0.06$	$1.57 \pm 0.06$	$1.70 \pm 0.05$
HiZELS	$z_{\text{eff}}$	1.09	1.29	1.49	1.69	1.89
	$\bar{N}_{\text{gal}}$	4196198	3273552	2342872	1773791	1370214
	$\bar{N}_{\text{ran}}/\bar{N}_{\text{gal}}$	20	20	25	30	40
	$\log_{10}(M_h/h^{-1}M_{\odot})$	$11.7 \pm 0.4$	$11.8 \pm 0.4$	$11.8 \pm 0.3$	$11.8 \pm 0.3$	$11.8 \pm 0.4$
	$b_{\text{lin}} \pm \delta b_{\text{lin}}$	$1.15 \pm 0.02$	$1.28 \pm 0.02$	$1.3 \pm 0.1$	$1.56 \pm 0.07$	$1.72 \pm 0.04$



**Figure 9.** Clustering results and bias fits for a Euclid-like survey with dust-attenuated flux limit  $2 \times 10^{-16} \text{erg s}^{-1} \text{cm}^{-2}$  (upper grid) and a WFIRST-like survey with dust-attenuated flux limit  $1 \times 10^{-16} \text{erg s}^{-1} \text{cm}^{-2}$  (lower grid), constructed from lightcones whose luminosities were attenuated using WISP-calibrated dust attenuation values. In each grid the darker lines show the results in redshift-space, the brighter lines correspond to real-space (i.e. assuming the cosmological redshifts of the galaxies with no peculiar velocity component), and the black dashed lines show the dark matter correlation function at the effective redshift of the bin. In the lower panels of each grid, the horizontal dotted lines and shaded regions correspond to the linear bias fits assuming the appropriate correlation function (see text for details). The redshift range used for selection is shown in the bottom left-hand corner of the lower panels.

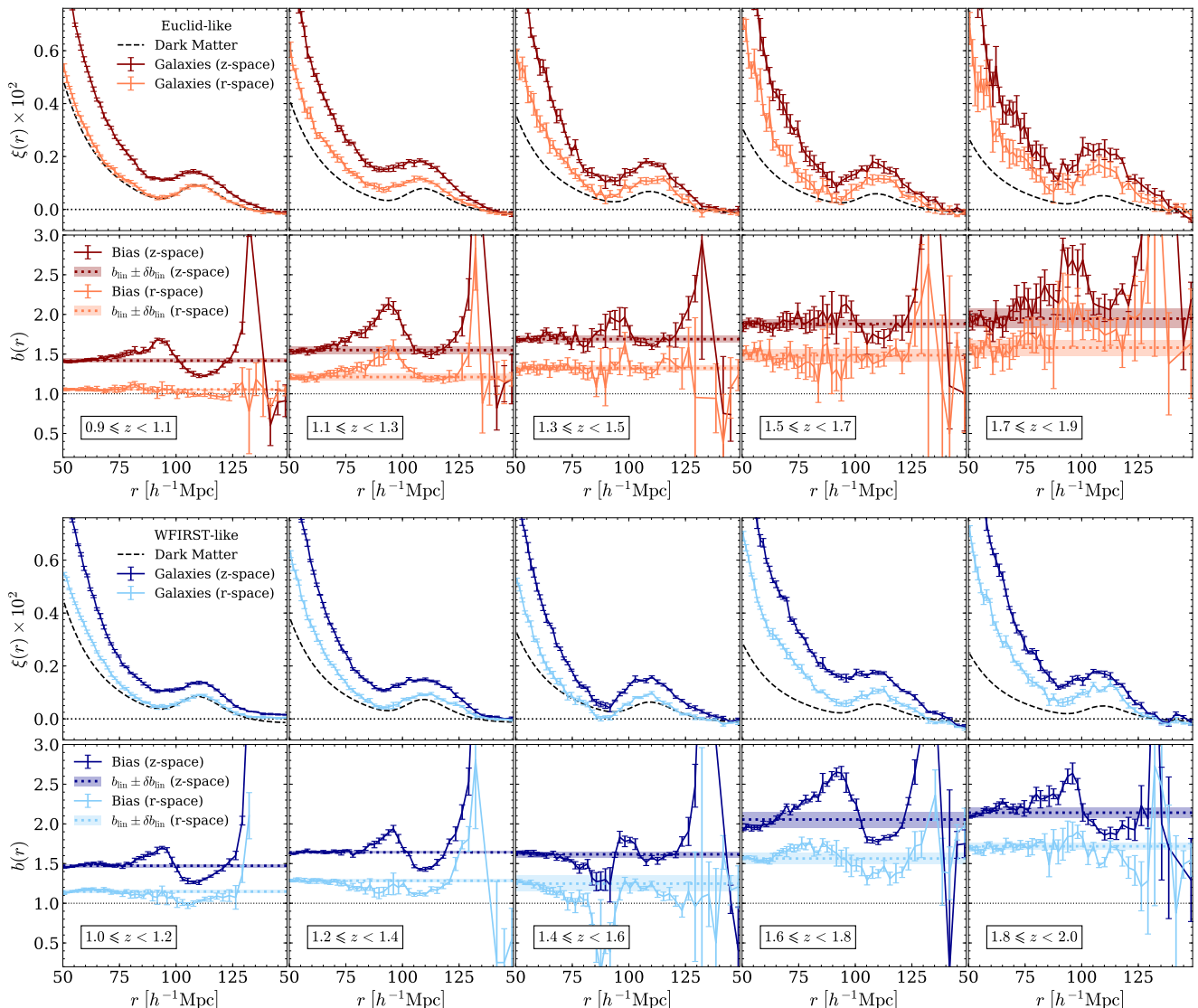
the impact of the large-scale deviations around the BAO scale, care must be taken when selecting the scale at which to fit a constant bias value, unless the BAO distortions are corrected for. We adopt  $r_{\text{cut}} = 75 h^{-1} \text{Mpc}$  for each redshift bin, as below this scale our bias measurements are broadly consistent with a constant value. The results of these fits are shown in Fig. 9 and Fig. 10 as horizontal dotted lines with shaded regions showing an associated uncertainty,  $\delta b_{\text{lin}}$ . The associated uncertainties were originally estimated as the  $1\sigma$  uncertainties on the linear bias fits obtained from the  $\chi^2$  minimisation procedure, which were on the order of 0.1 per cent. These uncertainties appear quite conservative given our measurements of  $b(r)$ . Instead, we estimate the uncertainties on the linear bias fits using the root-mean-square (RMS) of the difference between the

measured values and the fitted value,  $b_{\text{lin}}$ ;

$$\delta b_{\text{lin}} = \sqrt{\frac{1}{N} \sum_{r \leq r_{\text{cut}}} (b(r) - b_{\text{lin}})^2}, \quad (6)$$

where  $N$  is the number of measured bias values with  $r \leq r_{\text{cut}}$ . These RMS uncertainties,  $\delta b_{\text{lin}}$ , which are on the order of 1 per cent, are reported alongside the linear bias values,  $b_{\text{lin}}$ , in Table 2 and Table 3.

Comparing the fitted linear bias values in Table 2 and Table 3 we can see that the bias values for the WISP-calibrated instance and the HiZELS-calibrated instance are in good agreement and consistent within error. The choice of calibration, whether to match the WISP number counts or the HiZELS luminosity function, does not therefore appear to significantly impact the measurement of the bias. Secondly we see that the fitted linear bias values for the



**Figure 10.** Clustering results and bias fits for a Euclid-like survey with dust-attenuated flux limit  $2 \times 10^{-16} \text{erg s}^{-1} \text{cm}^{-2}$  (upper grid) and a WFIRST-like survey with dust-attenuated flux limit  $1 \times 10^{-16} \text{erg s}^{-1} \text{cm}^{-2}$  (lower grid), constructed from lightcones whose luminosities were attenuated using HiZELS-calibrated dust attenuation values. All the lines have the same meanings as in Fig. 9.

Euclid-like and WFIRST-like surveys are also broadly consistent with one another. We will return to this point in our discussion in §5.

#### 4.4 Linear bias evolution with redshift

We now examine how the linear bias of  $\text{H}\alpha$ -emitting galaxies evolves with redshift. We can see from Fig. 9 and Fig. 10 that our fitted values for the linear bias increase with increasing redshift, as we would expect. In Fig. 11 we plot the fits for the linear bias from the individual redshift bins as a function of effective redshift. The filled red diamonds show the fitted bias values for the redshift bins of the Euclid-like survey, whilst the filled blue circles show the fitted bias values for the redshift bins of the WFIRST-like survey. We can now see clearly the consistency between the bias values predicted for the Euclid-like surveys and those predicted for the WFIRST-like surveys.

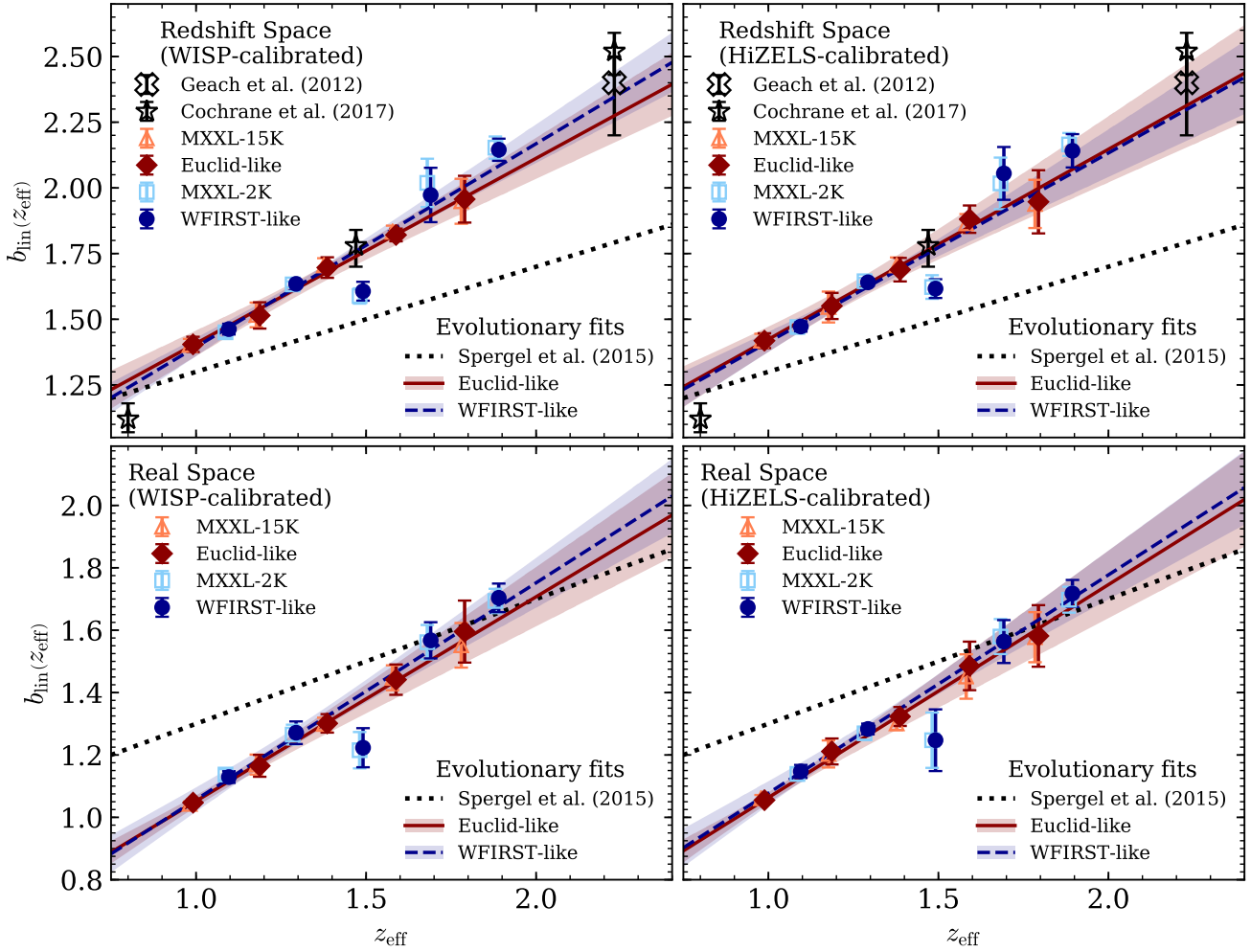
The faint symbols in Fig. 11 show the fitted linear bias values

obtained if we perform the clustering analysis with the MXXL-15K and MXXL-2K lightcones, i.e. assuming that both surveys are 100 per cent complete. This analysis is presented in Appendix C. The level of agreement between the linear bias values from the MXXL-15K and MXXL-2K lightcones and the Euclid-like and WFIRST-like surveys suggests that the introduction of incompleteness does not significantly affect our recovery of the bias.

##### 4.4.1 Evolutionary fits

It is clear from Fig. 11 that we can model the evolution of the linear bias,  $b_{\text{lin}}$  with redshift using a linear relation  $b_{\text{lin}}(z_{\text{eff}}) = m z_{\text{eff}} + c$ , where  $z_{\text{eff}}$  is the effective redshift of the galaxy sample,  $m$  is the gradient and  $c$  is the intercept. We use  $\chi^2$  minimisation to fit linear relations to the linear bias values for the Euclid-like and WFIRST-like surveys, in both redshift-space and real-space and for both the WISP-calibrated values and the HiZELS-calibrated values. The parameters fits, including the  $1\sigma$  and  $2\sigma$  contours, are shown

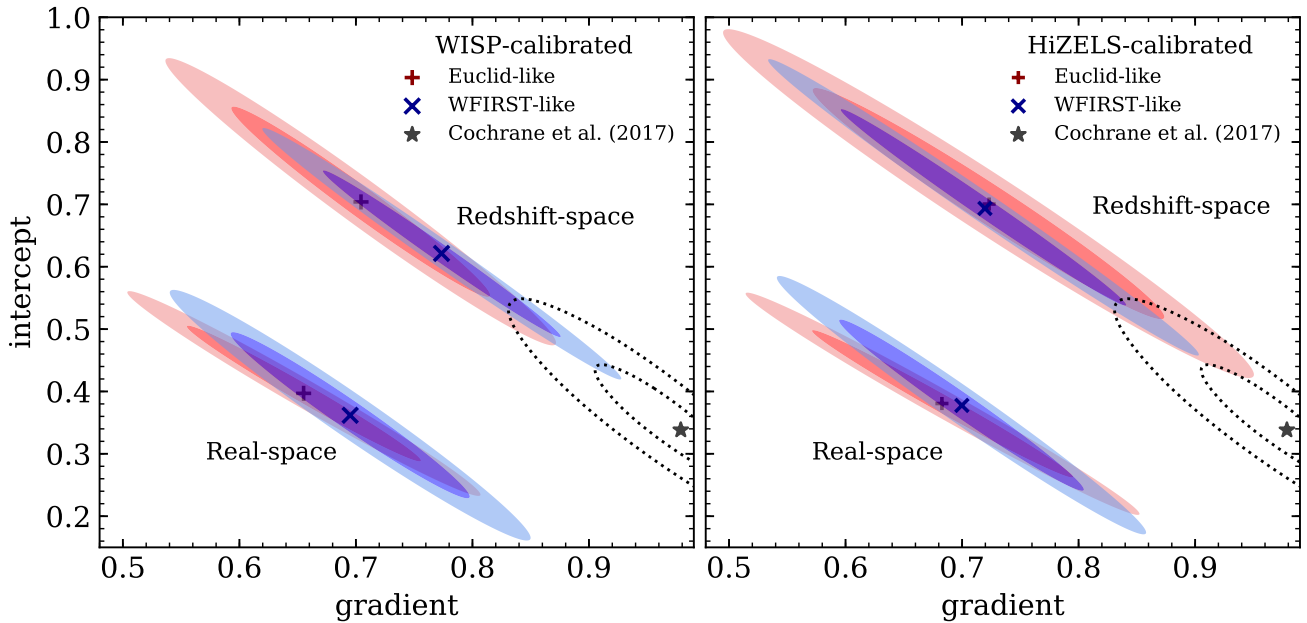




**Figure 11.** Linear bias as a function of effective redshift for both redshift-space (upper panels) and real-space (lower panels) when adopting a WISP-calibrated lightcone (left-hand column) or a HiZELS-calibrated lightcone (right-hand column). Red colours indicate results for our Euclid-like survey and blue colours indicate results for our WFIRST-like survey. The various red and blue symbols show the fitted linear bias values from the clustering analyses in each of the individual redshift bins. Filled symbols show the results for our Euclid-like and WFIRST-like surveys, whilst empty symbols show equivalent results from a clustering analysis of the MXXL-15 and MXXL-2K lightcones, without the introduction of any incompleteness effects (see Appendix C). Linear relations were fit to the bias values using  $\chi^2$  minimisation. The red solid lines show the linear relations for the Euclid-like surveys that minimise the  $\chi^2$  statistic and the red shaded regions correspond to the range of fits that have a gradient and intercept within the  $1\sigma$  uncertainty range. The blue dashed lines and blue shaded regions show the equivalent results for the WFIRST-like surveys. The black cross shows the bias result from Geach et al. (2012), the black stars show the bias results from Cochrane et al. (2017) and the black dotted line shows the linear relation assumed by Spergel et al. (2015).

**Table 4.** Gradient and intercept parameter results obtained from using  $\chi^2$  minimisation to fit linear relations for the redshift evolution of the linear bias for our Euclid-like and WFIRST-like surveys. The upper half of the table shows the parameters fits to our bias measurements in redshift-space, whilst the lower half of the table shows the fits to the bias measurements in real-space. The parameter values stated correspond to the values that minimise the  $\chi^2$  statistic, and the uncertainties correspond to the bounds of the  $1\sigma$  contours as shown in Fig. 12. Parameters are reported for both the WISP-calibrated and HiZELS-calibrated versions of the surveys.

Calibration	Euclid-like		WFIRST-like	
	Gradient	Intercept	Gradient	Intercept
<b>Redshift-space</b>				
WISP	$0.70 \pm 0.11$	$0.70 \pm 0.15$	$0.77 \pm 0.10$	$0.62 \pm 0.13$
HiZELS	$0.72 \pm 0.15$	$0.70^{+0.19}_{-0.18}$	$0.72 \pm 0.12$	$0.69 \pm 0.16$
<b>Real-space</b>				
WISP	$0.66 \pm 0.10$	$0.40 \pm 0.11$	$0.70 \pm 0.10$	$0.36 \pm 0.13$
HiZELS	$0.68 \pm 0.11$	$0.38 \pm 0.12$	$0.70^{+0.10}_{-0.11}$	$0.38 \pm 0.14$



**Figure 12.** Gradient and intercept parameters for the linear relations describing the redshift evolution of the linear bias for our Euclid-like and WFIRST-like surveys. Euclid-like results are coloured red and WFIRST-like results are coloured blue. Symbols indicate the best-fit parameters, corresponding to the minimum  $\chi^2$  fit. Darker shaded ellipses correspond to the  $1\sigma$  uncertainties and lighter shaded ellipses correspond to the  $2\sigma$  uncertainties. The upper pairs of shaded ellipses correspond to the redshift-space results and the lower pairs of shaded ellipses correspond to the real-space results. The left-hand panel shows the results for the WISP-calibrated instance and the right-hand panel shows the results for the HiZELS-calibrated instance. The star and dotted ellipses show the parameter fit and  $1\sigma$  and  $2\sigma$  uncertainties obtained when using  $\chi^2$  minimisation to fit a linear relation to the bias results from Cochrane et al. (2017).

in Fig. 12 and are reported in Table 4. These fits are also shown in Fig. 11 where the solid and dashed lines show the linear fits whose parameters minimise the  $\chi^2$  statistic. The red solid lines correspond to the fits for the Euclid-like survey and the blue dashed lines correspond to the fits for the WFIRST-like survey. The shaded regions in Fig. 11 indicate the range of linear fits that fall within the corresponding  $1\sigma$  contour in Fig. 12.

Comparing the gradients and intercepts fits in Fig. 12 we see that for both the WISP-calibrated and HiZELS-calibrated instances the linear fits for the Euclid-like and WFIRST-like surveys are consistent with one another within the  $1\sigma$  uncertainties. For the HiZELS-calibrated instances the Euclid-like and WFIRST-like fits have gradients and intercepts that are in excellent agreement. For the WISP-calibrated instance the WFIRST-like surveys have a slightly steeper gradient and smaller intercept than the Euclid-like surveys, though the fits are still consistent at the  $1\sigma$  level. In each instance the  $1\sigma$  contours for the WFIRST-like survey are typically narrower than the  $1\sigma$  contours for the Euclid-like survey. The impact of this can be seen by comparing the width of the shaded regions in each panel of Fig. 11. This likely arises due to the individual linear bias values in the Euclid-like surveys having larger uncertainties, particularly at high redshift, compared to the linear bias values for the WFIRST-like surveys. We note however, that the individual linear bias values for the WFIRST-like surveys typically have a larger scatter. For the WFIRST-like surveys there is a visible outlier at  $z_{\text{eff}} \simeq 1.5$  that is most likely due to cosmic variance, though we note that from the WFIRST-like clustering analysis the measured bias shows a significant negative deviation on scales  $75 h^{-1} \text{Mpc} \lesssim r \lesssim 100 h^{-1} \text{Mpc}$  in the  $1.4 \leq z \leq 1.6$  bin. Whilst this deviation, which we discussed in § 4.3, did not impact our fit for the linear bias on scales  $r < 75 h^{-1} \text{Mpc}$ , it may also

signify the presence of cosmic variance. We argue that this outlier is not having a significant impact on our fits in Fig. 11. Overall we see that in redshift-space the Euclid-like and WFIRST-like fits are both consistent within  $1\sigma$  with a linear relation  $b(z) = 0.7z + 0.7$ , in both the WISP-calibrated and the HiZELS-calibrated instances.

#### 4.4.2 Comparison to observations

To compare with linear bias estimates in the literature we show in the upper panels of Fig. 11 the linear bias estimates from Geach et al. (2012) and Cochrane et al. (2017), both of which were obtained from clustering analyses of HiZELS galaxies. The Cochrane et al. (2017) results that we plot correspond to the effective bias results from their ‘full’ samples, with no additional luminosity selection applied. If we extrapolate our linear fits for our Euclid-like and WFIRST-like surveys out to  $z = 2.23$ , we estimate that  $b_{\text{lin}}(z_{\text{eff}} = 2.23) \sim 2.3$ , which is in excellent agreement with the Geach et al. (2012) result of  $b_{\text{lin}} = 2.4_{-0.2}^{+0.1}$ . Indeed, we see in Fig. 11 that the Geach et al. (2012) result falls within the shaded regions corresponding to the  $1\sigma$  uncertainty range for our evolutionary fits.

We see that our linear bias relations are also in excellent agreement with the Cochrane et al. (2017) result at  $z = 1.47$ , though our forecasts suggest a shallower evolutionary relation with redshift, leading to our forecasts predicting a higher bias at  $z = 0.8$  and a lower bias at  $z = 2.23$  compared to the Cochrane et al. (2017) results. This can be seen clearly if we use  $\chi^2$  minimisation to fit a linear relation to the three Cochrane et al. (2017) bias measurements. The best-fitting parameters to the Cochrane et al. (2017) measurements are indicated by the star in Fig. 12, with the dotted lines indicating the  $1\sigma$  and  $2\sigma$  uncertainty contours. The relation that we obtain is

$b(z) = (0.98_{-0.07}^{+0.06})z + (0.34_{-0.09}^{+0.11})$ , where the stated uncertainties correspond to the bounds of the  $1\sigma$  ellipse. We can see from Fig. 12 that the fit to the Cochrane et al. (2017) measurements is consistent with our fits at the  $2\sigma$  level.

Overall our forecasts are in good agreement with the observational bias estimates from HiZELS.

The dotted lines in Fig. 11 correspond to the linear bias relation  $b_{\text{lin}}(z) = 0.4z + 0.9$  that was assumed by Spergel et al. (2015). We find that the Spergel et al. (2015) relation is consistent with our relations in real-space, particularly at  $z \sim 2$ . However, a difference in the slopes leads to our forecasts predicting a smaller bias for  $z \lesssim 2$ . Comparing the Spergel et al. (2015) relation to our relations in redshift-space we see that our relations predict a larger bias for all redshifts  $z > 0.9$ . This is understandable given that the Spergel et al. (2015) relation was based upon the clustering analysis of Orsi et al. (2010) that was carried out in real-space.

## 5 DISCUSSION

As noted previously, by comparing the linear bias fits for our Euclid-like and WFIRST-like surveys, shown in Fig. 9 and Fig. 10, we find that the two surveys yield linear bias values that are broadly consistent and in reasonable agreement with each other. Additionally, Fig. 11 and Fig. 12 show that the redshift evolution of the bias is consistent between the two surveys. At first sight this appears surprising, as one might expect the Euclid-like survey, which has a brighter flux-limit, to have a higher bias. However, consideration of the dark matter halos in which the  $H\alpha$ -emitting galaxies reside may help explain this result.

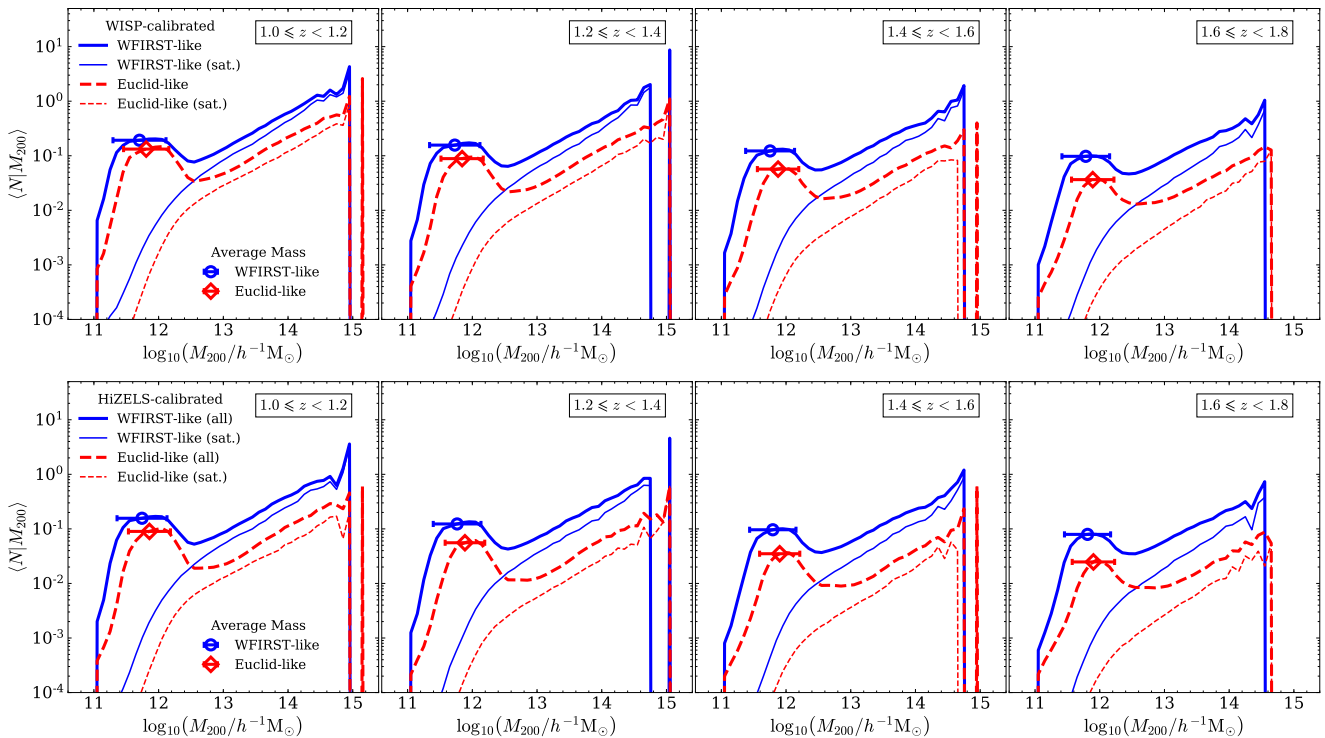
In Fig. 13 we show the HODs for our Euclid-like and WFIRST-like surveys. For the purposes of comparison, we compute the HODs in four redshift bins ( $1.0 \leq z < 1.2$ ,  $1.2 \leq z < 1.4$ ,  $1.4 \leq z < 1.6$  and  $1.6 \leq z < 1.8$ ) that span the redshift range in common to both the Euclid-like and WFIRST-like surveys. The solid blue lines show the HODs for the WFIRST-like galaxies and the red dashed lines show the HODs for the Euclid-like galaxies that fall within the specified redshift bin. Thick lines correspond to the HODs for all galaxies, whilst thin lines correspond to the HODs for satellite galaxies only. The open symbols indicate the mean and standard deviation of the halo masses hosting the Euclid-like and WFIRST-like galaxies. Note that the mean and standard deviations of the halo masses hosting the galaxies used in our clustering analysis are also provided in Table 2 and Table 3. The mean halo masses for the Euclid-like survey and the WFIRST-like survey are consistent within error, with a value  $M_{200} \sim 10^{11.8} h^{-1} M_{\odot}$ . This halo mass is in excellent agreement with the results from Geach et al. (2012), who concluded from their clustering analysis of HiZELS galaxies at  $z = 2.23$  that  $H\alpha$ -emitting galaxies are typically hosted by dark matter halos with mass  $\log_{10}(M_h/h^{-1}M_{\odot}) = 11.7 \pm 0.1$ . Also, based upon their empirical modelling of the stellar mass-halo mass relation, Behroozi et al. (2013) determined that halos of mass  $10^{11.8} h^{-1} M_{\odot}$  are the most efficient at forming stars for all redshifts  $z < 8$ , which also is in very good agreement with our typical halo mass values.

We can see from Fig. 13 that the Euclid-like and WFIRST-like surveys have HODs that are broadly similar in shape, with a peak at  $M_{200} \lesssim 10^{12} h^{-1} M_{\odot}$  and a power-law extending to higher masses. However, the HODs differ in their normalisation. In each redshift bin we see that for any given halo mass the WFIRST-like survey has on average a larger number of galaxies per halo. This is understandable given the deeper flux limit for the WFIRST mission. The

mean halo masses, shown by the open symbols, suggest that the WFIRST-like galaxies are also typically found in lower mass halos than the Euclid-like galaxies. This can also be seen by comparing the peaks in the HODs at  $M_{200} \lesssim 10^{12} h^{-1} M_{\odot}$ , where the peak in the Euclid-like HOD is narrower and skewed towards slightly larger halo mass. We note however that the difference between the mean halo masses for the two surveys is smaller than the widths of the halo mass distributions and so within error the two surveys have an equivalent mean halo mass.

Since the WFIRST-like galaxies are typically placed in lower mass halos we might again expect the WFIRST-like survey to have a smaller linear bias value than the Euclid-like survey. Halo bias models (e.g. Tinker et al. 2010) suggest that at  $M_{200} \lesssim 10^{12} h^{-1} M_{\odot}$  the halo bias is a weak function of halo mass and so the small difference in the mean halo mass between the Euclid-like and WFIRST-like surveys will have only a negligible impact on our linear bias values. However, subtle differences in the shapes of the Euclid-like and WFIRST-like HODs can help explain why the two surveys have consistent linear bias values. Considering the power-law slopes of the HODs, we see that the WFIRST-like HOD has a slightly steeper power-law slope than the Euclid-like HOD. Examining the HODs for satellite galaxies only, shown in Fig. 13 by the thin lines, we see that this is caused by the satellite HOD for the WFIRST-like survey having a steeper power-law slope than the satellite HOD for the Euclid-like survey. As such, as we move towards larger halo mass the typical number of satellite galaxies found in any given halo will grow more rapidly for the WFIRST-like survey than for the Euclid-like survey. This can be seen in Fig. 2 and Fig. 3 where, as we push down to fainter luminosities, the power-law slope of the HOD becomes steeper. Therefore, the number of faint satellites observed in high mass halos will increase more rapidly as we push down to fainter flux limits. Since massive halos are known to be more highly biased, the growing contribution of satellites in high mass halos will lead to a boost in the linear bias of the galaxy sample. The fainter WFIRST-like survey would therefore receive a relatively larger boost in the measured linear bias compared to the Euclid-like survey. As a result, this could lead to the WFIRST-like survey having linear bias that is more consistent with, or potentially greater than, the linear bias of the Euclid-like survey.

We also see that our choice of calibration does not appear to have a large impact on our measurements of the bias, or linear bias fits. We can see in Table 2 or Table 3 that the linear bias fits using a WISP-calibrated lightcone and the fits using a HiZELS-calibrated lightcone are consistent within error with one another, despite the WISP-calibrated lightcones and HiZELS-calibrated lightcones showing a clear difference in the number density of galaxies (see Fig. 7). The difference in the number density of galaxies between the two calibrations is a result of the different dust attenuation values leading to a small difference in the number of galaxies passing the flux selection. The weaker dust attenuation in the WISP-calibrated instance means that additional faint galaxies will be selected, that would not otherwise be selected in the HiZELS-calibrated instance. This is equivalent to having applied a slightly fainter flux limit in the WISP-calibrated instance. Given the luminosity-dependent power-law slope of the HOD, the WISP-calibrated instance will thus include a larger contribution of faint satellites in highly biased, massive halos. As a result, the WISP-calibrated instance will see a slight boost in the measured linear bias relative to the HiZELS-calibrated instance, such that the biases for the two instances are consistent despite the difference in the galaxy number densities.



**Figure 13.** Halo occupation distributions (HODs) for our Euclid-like and WFIRST-like surveys. The upper row of panels show the HODs for the WISP-calibrated instance and the lower row of panels show the HODs for the HiZELS-calibrated instance. For the purposes of comparison, we compute the HODs in the four redshift bins from our WFIRST-like analysis that overlap with the redshift range for our Euclid-like analysis. The blue solid lines show the HODs for the WFIRST-like survey and the red dashed lines show the HODs for the Euclid-like survey. The thick lines show the HODs for all galaxies and the thin lines show the HODs for satellite galaxies only. The symbols show the mean and standard deviation of the halo masses hosting the galaxies in our Euclid-like and WFIRST-like samples.

## 6 CONCLUSIONS & SUMMARY

In this work we forecast the linear bias, as a function of redshift, for  $H\alpha$ -emitting galaxies in a Euclid-like survey and also in a WFIRST-like survey.

To simulate our Euclid-like and WFIRST-like surveys, we use the methodology of Smith et al. (2017), whereby dark matter halos in a lightcone are populated with galaxies by sampling from a set of luminosity-dependent halo occupation distributions (HODs). The lightcone of dark matter halos that we use is the halo lightcone created by Smith et al. (2017) from the Millennium XXL simulation (MXXL, Angulo et al. 2012), which is consistent with the first year results of the Wilkinson Microwave Anisotropy Probe Spergel et al. (2003). This lightcone covers the entire sky out to a redshift of  $z = 2.2$  and has a halo mass resolution of  $M_{200} \geq 10^{11} h^{-1} M_{\odot}$ , where  $M_{200}$  is the mass enclosed within a sphere whose mean density is 200 times the mean density of the Universe.

Unlike broad-band photometric samples, the HOD of  $H\alpha$ -emitters is less well understood and to-date has not been extensively parametrised. As such, instead of using a parametric form for the HOD, we generate a library of HODs as a function of limiting  $H\alpha + [N_{II}]$  luminosity using a physically-motivated galaxy formation model. The model that we use is the open source GALACTICUS semi-analytical model (Benson 2010), which predicts emission line luminosities based upon outputs from the CLOUDY photoionisation code (Ferland et al. 2013). Unfortunately the mass resolution of the MXXL halo merger trees prevents us from running GALACTICUS on the simulation directly. However, we apply GALACTICUS to the Millennium Simulation (Springel et al. 2005),

which has an identical cosmology to the MXXL. We acknowledge that this cosmology is not consistent with current cosmological results from Planck Collaboration et al. (2018), but argue that the uncertainties are most likely dominated by our ignorance of the process of galaxy formation.

Examining the  $H\alpha$  HODs predicted by GALACTICUS, we see that for faint  $H\alpha + [N_{II}]$  luminosity limits the shape of the HOD resembles a smoothed step-like function, whilst towards brighter  $H\alpha + [N_{II}]$  luminosity limits the amplitude of the HOD decreases, with the HOD displaying a broad peak at masses  $M_{200} \sim 10^{12} h^{-1} M_{\odot}$  (Fig. 2). This shape, which is consistent with the shapes HODs of emission line galaxies and star-forming galaxies seen elsewhere in the literature, can be understood as resulting from feedback processes quenching star-formation in more massive halos. Galaxies are placed into the dark matter halos of the MXXL halo lightcone by interpolating over the library of HODs and using random sampling to draw a population of galaxies. Central galaxies are placed at the centre-of-mass of their host halo and move at the halo velocity. Satellite galaxies are placed randomly within the halo following a Navarro et al. (1997) profile, with velocities drawn randomly from an isotropic Maxwell-Boltzmann distribution.

We build two lightcone galaxy catalogues: one covering  $15000 \text{ deg}^2$  with a  $H\alpha + [N_{II}]$  flux limit of  $2 \times 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$ , from which we build our Euclid-like survey, and one covering  $2000 \text{ deg}^2$  with a  $H\alpha + [N_{II}]$  flux limit of  $1 \times 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$ , from which we build our WFIRST-like survey. To calibrate the lightcones so that they have the correct number density and total number of galaxies, we apply dust attenuation to the  $H\alpha + [N_{II}]$

luminosities such that the lightcone luminosity functions are consistent with the luminosity functions from HiZELS (Sobral et al. 2013) and the lightcone cumulative number counts are consistent with the cumulative number counts from the WISP survey (Mehta et al. 2015). The values for the attenuation,  $A_{H\alpha}$ , are determined through a  $\chi^2$  analysis. We find that the lightcones are able to independently reproduce the WISP number counts (Fig. 4) and the HiZELS luminosity functions (Fig. 6), but that different values for  $A_{H\alpha}$  are required to reproduce these datasets (Fig. 5). The cause of this difference is uncertain, but may be due to cosmic variance between the WISP and HiZELS surveys, differences in the selection functions of the surveys or the WISP number counts probing a restricted luminosity range compared to the HiZELS luminosity functions. We therefore proceed with our bias forecasts considering both a WISP-calibrated version and a HiZELS-calibrated version of the each lightcone.

To make bias forecasts we first build our Euclid-like and WFIRST-like surveys by using random sampling to apply incompleteness to the appropriate lightcone. We assume a completeness of 45 per cent for our Euclid-like survey and 70 per cent for our WFIRST-like survey. The TREECORR software (Jarvis et al. 2004) is then used to compute the angle averaged galaxy correlation function in five redshift bins of  $\Delta z = 0.2$  between  $0.9 \leq z < 1.9$  for the Euclid-like survey and between  $1 \leq z < 2$  for the WFIRST-like survey. For each redshift bin, the galaxy correlation function is computed for both WISP-calibrated and HiZELS-calibrated versions of the surveys. The correlation function is computed five times for each bin, each time using a different random seed to apply survey incompleteness. The entire set of calculations is repeated twice, with galaxies selected in redshift-space in the first instance and in real-space in the second instance. By computing the correlation function between  $50 h^{-1}\text{Mpc}$  and  $150 h^{-1}\text{Mpc}$  we are able to recover the BAO peak for each redshift bin (upper panels of Fig. 9 and Fig. 10). Comparing the correlation functions in real-space and redshift-space we see that the correlation function in redshift-space has a higher amplitude, consistent with theory of redshift-space distortions from Kaiser (1987).

To compute the linear bias we divide the galaxy correlation function in each redshift bin by the non-linear dark matter correlation, which we compute at the effective redshift of the galaxies in that bin using the NBODYKIT (Hand et al. 2018) Python package. We find that for scales  $r \lesssim 75 h^{-1}\text{Mpc}$  the measured linear bias in each redshift bin is consistent with a fixed, scale-independent value. However, at larger scales the measured bias displays scale-dependent deviations away from a constant. These sinusoidal-like deviations, which are more pronounced in redshift-space, occur at scales around the BAO scale, which suggests that they could be being caused by distortions of the shape and position of the BAO in the galaxy correlation function relative to the shape and position predicted by linear theory. Such distortions can arise due a combination of non-linear effects, mode coupling, redshift-space distortions, as well as even the bias itself.

Taking the measured linear bias for separations  $r < 75 h^{-1}\text{Mpc}$  we use  $\chi^2$  minimisation to fit a constant, scale-independent value for the linear bias in each redshift bin. Note that the presence of the deviations, due to distortion of the BAO, means that care must be taken when considering the scales at which to fit the linear bias. We estimate the uncertainties on the fitted linear bias values by computing the root-mean-square (RMS) of the difference between the measured and the fitted values (lower panels of Fig. 9 and Fig. 10). The fitted linear bias values increase with increasing redshift, as we would expect.

Using  $\chi^2$  minimisation we fit linear relations to the linear bias as a function of effective redshift (Fig. 11). The best-fitting linear relations for the Euclid-like and WFIRST-like surveys, (shown in Fig. 12 and listed in Table 4), are consistent with one another at the  $1\sigma$  level. For the HiZELS-calibrated instance the best-fitting parameters for the Euclid-like and WFIRST-like surveys are in excellent agreement, though for the WISP-calibrated instance the Euclid-like relation has a shallower gradient. In redshift-space, the best-fitting parameters for the Euclid-like and WFIRST-like surveys are both consistent with the relation  $b(z) = 0.7z + 0.7$ . By extrapolating these linear relations we find that they are in excellent agreement with the bias result from Geach et al. (2012). Our linear bias results are consistent with the results from Cochrane et al. (2017), particularly at  $z = 1.47$ , though our results suggest a shallower slope for the evolution of bias with redshift. Using  $\chi^2$  to fit the Cochrane et al. (2017) bias measurements, we find that our best-fitting parameters are consistent with the Cochrane et al. (2017) fit at a  $2\sigma$  level. Comparing to the linear bias relation assumed by Spergel et al. (2015), we find that our real-space relations are consistent with the Spergel et al. (2015) relation, but in redshift-space our relations predict larger linear bias for all redshifts  $z > 0.9$ .

A surprising result is that our linear bias forecasts for the Euclid-like and WFIRST-like surveys are consistent with one another. This can be seen clearly when the linear bias values are plotted as a function of effective redshift (Fig. 11). We would expect the Euclid-like survey to yield higher bias values due to the brighter flux limit. However, examination of the HODs for the Euclid-like and WFIRST-like surveys shows that the HOD for the fainter WFIRST-like survey has a steeper power-law slope at high masses, leading to a more rapidly growing number of satellite galaxies than in the HOD for the Euclid-like survey (Fig. 13). The greater contribution of faint satellite galaxies in high mass halos can act to boost the bias for the WFIRST-like survey such that it is consistent with or even greater than the bias for the Euclid-like survey. Additionally, by comparing the fitted linear bias values for the WISP-calibrated and HiZELS-calibrated versions of the surveys, we find that the choice of calibration has negligible impact on the linear bias values.

## ACKNOWLEDGEMENTS

We thank the anonymous referee for their many helpful comments and suggestions that greatly enhanced the discussion of this work. We additionally thank Anahita Alvari, Iary Davidzon, Andreas Faisst, Zhongxu Zhai, and the members of the JPL Dark-sector group for various insightful conversations that helped improve this work. AM acknowledges sponsorship of a NASA Postdoctoral Program Fellowship. AM was supported by JPL, which is run under contract by California Institute of Technology for NASA. This work was supported by NASA ROSES grant 12-EUCLID12-0004 and by NASA grant 15-WFIRST15-0008 Cosmology with the High Latitude Survey WFIRST Science Investigation Team (SIT). This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility ([www.dirac.ac.uk](http://www.dirac.ac.uk)). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1, ST/R002371/1 and ST/S002502/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure. Copyright 2019. All rights reserved.

## REFERENCES

- Albrecht A., Bernstein G., Cahn R., Freedman W. L., Hewitt J., Hu W., Huth J., Kamionkowski M. et al, 2006, arXiv:0609591
- Angulo R. E., Baugh C. M., Frenk C. S., Lacey C. G., 2008, MNRAS, 383, 755
- Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012, MNRAS, 426, 2046
- Angulo R. E., White S. D. M., Springel V., Henriques B., 2014, MNRAS, 442, 2131
- Anselmi S., Starkman G. D., Sheth R. K., 2016, MNRAS, 455, 2474
- Arnouts S., Schiminovich D., Ilbert O., Tresse L., Milliard B., Treyer M., Bardelli S., Budavari T. et al, 2005, ApJ, 619, L43
- Atek H., Malkan M., McCarthy P., Teplitz H. I., Scarlata C., Siana B., Henry A., Colbert J. W. et al, 2010, ApJ, 723, 104
- Atek H., Siana B., Scarlata C., Malkan M., McCarthy P., Teplitz H., Henry A., Colbert J. et al, 2011, ApJ, 743, 121
- Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15
- Basilikos S., Dent J. B., Dutta S., Perivolaropoulos L., Plionis M., 2012, Phys. Rev. D, 85, 123501
- Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
- Benson A. J., 2010, Phys. Rep., 495, 33
- , 2012, New A, 17, 175
- , 2014, MNRAS, 444, 2599
- Benson A. J., Baugh C. M., Cole S., Frenk C. S., Lacey C. G., 2000a, MNRAS, 316, 107
- Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000b, MNRAS, 311, 793
- Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
- Blake C., Glazebrook K., 2003, ApJ, 594, 665
- Blanton M., Cen R., Ostriker J. P., Strauss M. A., Tegmark M., 2000, ApJ, 531, 1
- Brammer G. B., Whitaker K. E., van Dokkum P. G., Marchesini D., Franx M., Kriek M., Labbé I., Lee K.-S. et al, 2011, ApJ, 739, 24
- Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
- Chabrier G., 2003, ApJ, 586, L133
- Charlot S., Fall S. M., 2000, ApJ, 539, 718
- Clerkin L., Kirk D., Lahav O., Abdalla F. B., Gaztañaga E., 2015, MNRAS, 448, 1389
- Cochrane R. K., Best P. N., 2018, MNRAS, 480, 864
- Cochrane R. K., Best P. N., Sobral D., Smail I., Geach J. E., Stott J. P., Wake D. A., 2018, MNRAS, 475, 3730
- Cochrane R. K., Best P. N., Sobral D., Smail I., Wake D. A., Stott J. P., Geach J. E., 2017, MNRAS, 469, 2913
- Conroy C., Schiminovich D., Blanton M. R., 2010, ApJ, 718, 184
- Contreras S., Baugh C. M., Norberg P., Padilla N., 2013, MNRAS, 432, 2717
- Crocce M., Scoccimarro R., 2008, Phys. Rev. D, 77, 023533
- Davis M., Geller M. J., 1976, ApJ, 208, 13
- DESI Collaboration, Aghamousa A., Aguilar J., Ahlen S., Alam S., Allen L. E., Allende Prieto C., Annis J. et al, 2016, arXiv:161100036
- Desjacques V., Jeong D., Schmidt F., 2018, Phys. Rep., 733, 1
- Domínguez A., Siana B., Henry A. L., Scarlata C., Bedregal A. G., Malkan M., Atek H., Ross N. R. et al, 2013, ApJ, 763, 145
- Dressler A., 1980, ApJ, 236, 351
- Dressler A., Spergel D., Mountain M., Postman M., Elliott E., Bendek E., Bennett D., Dalcanton J. et al, 2012, arXiv:12107809
- Durkalec A., Le Fèvre O., Pollo A., Zamorani G., Lemaux B. C., Garilli B., Bardelli S., Hathi N. et al, 2018, A&A, 612, A42
- Eisenstein D. J., Seo H.-J., White M., 2007, ApJ, 664, 660
- Faisst A. L., Masters D., Wang Y., Merson A., Capak P., Malhotra S., Rhoads J. E., 2018, ApJ, 855, 132
- Favole G., Rodríguez-Torres S. A., Comparat J., Prada F., Guo H., Klypin A., Montero-Dorta A. D., 2017, MNRAS, 472, 550
- Ferland G. J., Porter R. L., van Hoof P. A. M., Williams R. J. R., Abel N. P., Lykins M. L., Shaw G., Henney W. J. et al, 2013, Rev. Mexicana Astron. Astrofis., 49, 137
- Ferrara A., Bianchi S., Cimatti A., Giovanardi C., 1999, ApJS, 123, 437
- Fry J. N., 1996, ApJ, 461, L65
- Garn T., Best P. N., 2010, MNRAS, 409, 421
- Garn T., Sobral D., Best P. N., Geach J. E., Smail I., Cirasuolo M., Dalton G. B., Dunlop J. S. et al, 2010, MNRAS, 402, 2017
- Gaztañaga E., Eriksen M., Crocce M., Castander F. J., Fosalba P., Martí P., Miquel R., Cabré A., 2012, MNRAS, 422, 2904
- Geach J. E., Smail I., Best P. N., Kurk J., Casali M., Ivison R. J., Coppin K., 2008, MNRAS, 388, 1473
- Geach J. E., Sobral D., Hickox R. C., Wake D. A., Smail I., Best P. N., Baugh C. M., Stott J. P., 2012, MNRAS, 426, 679
- Gonzalez-Perez V., Comparat J., Norberg P., Baugh C. M., Contreras S., Lacey C., McCullagh N., Orsi A. et al, 2018, MNRAS, 474, 4024
- Green J., Schechter P., Baltay C., Bean R., Bennett D., Brown R., Conselice C., Donahue M. et al, 2012, arXiv:12084012
- Guo H., Zehavi I., Zheng Z., Weinberg D. H., Berlind A. A., Blanton M., Chen Y., Eisenstein D. J. et al, 2013, ApJ, 767, 122
- Guzzo L., Pierleoni M., Meneux B., Branchini E., Le Fèvre O., Marinoni C., Garilli B., Blaizot J. et al, 2008, Nature, 451, 541
- Guzzo L., Strauss M. A., Fisher K. B., Giovanelli R., Haynes M. P., 1997, ApJ, 489, 37
- Hada R., Eisenstein D. J., 2019, MNRAS, 482, 5685
- Hand N., Feng Y., Beutler F., Li Y., Modi C., Seljak U., Slepian Z., 2018, AJ, 156, 160
- Hatfield P. W., Lindsay S. N., Jarvis M. J., Häußler B., Vaccari M., Verma A., 2016, MNRAS, 459, 2618
- Hui L., Parfrey K. P., 2008, Phys. Rev. D, 77, 043527
- Jarvis M., Bernstein G., Jain B., 2004, MNRAS, 352, 338
- Jeong D., Komatsu E., 2006, ApJ, 651, 619
- Kaiser N., 1984, ApJ, 284, L9
- , 1987, MNRAS, 227, 1
- Kauffmann G., Nusser A., Steinmetz M., 1997, MNRAS, 286, 795
- Kim J.-W., Im M., Lee S.-K., Edge A. C., Wake D. A., Merson A. I., Jeon Y., 2015, ApJ, 806, 189
- Knebe A., Pearce F. R., Thomas P. A., Benson A., Blaizot J., Bower R., Carretero J., Castander F. J. et al, 2015, MNRAS, 451, 4029
- Landy S. D., Szalay A. S., 1993, ApJ, 412, 64
- Laureijs R., Amiaux J., Arduini S., Auguères J., Brinchmann J., Cole R., Cropper M., Dabin C. et al, 2011, arXiv:11103193
- Law-Smith J., Eisenstein D. J., 2017, ApJ, 836, 87
- Lemson G., Virgo Consortium t., 2006, arXiv:0608019
- Li C., White S. D. M., 2009, MNRAS, 398, 2177
- Ly C., Malkan M. A., Kashikawa N., Shimasaku K., Doi M., Nagao T., Iye M., Kodama T. et al, 2007, ApJ, 657, 738
- Madau P., Dickinson M., 2014, ARA&A, 52, 415
- Madau P., Pozzetti L., Dickinson M., 1998, ApJ, 498, 106
- Mann R. G., Peacock J. A., Heavens A. F., 1998, MNRAS, 293, 209

- Martens D., Fang X., Troxel M. A., DeRose J., Hirata C. M., Wechsler R. H., Wang Y., 2019, *MNRAS*, 485, 211
- McCracken H. J., Wolk M., Colombi S., Kilbinger M., Ilbert O., Peirani S., Coupon J., Dunlop J. et al, 2015, *MNRAS*, 449, 901
- Mehta V., Scarlata C., Colbert J. W., Dai Y. S., Dressler A., Henry A., Malkan M., Rafelski M. et al, 2015, *ApJ*, 811, 141
- Merson A., Wang Y., Benson A., Faisst A., Masters D., Kiessling A., Rhodes J., 2018, *MNRAS*, 474, 177
- Merson A. I., Baugh C. M., Helly J. C., Gonzalez-Perez V., Cole S., Bielby R., Norberg P., Frenk C. S. et al, 2013, *MNRAS*, 429, 556
- Mirbabayi M., Schmidt F., Zaldarriaga M., 2015, *J. Cosmology Astropart. Phys.*, 7, 030
- Navarro J. F., Frenk C. S., White S. D. M., 1997, *ApJ*, 490, 493
- Norberg P., Baugh C. M., Hawkins E., Maddox S., Madgwick D., Lahav O., Cole S., Frenk C. S. et al, 2002, *MNRAS*, 332, 827
- Norberg P., Baugh C. M., Hawkins E., Maddox S., Peacock J. A., Cole S., Frenk C. S., Bland-Hawthorn J. et al, 2001, *MNRAS*, 328, 64
- Orsi A., Baugh C. M., Lacey C. G., Cimatti A., Wang Y., Zamorani G., 2010, *MNRAS*, 405, 1006
- Peacock J. A., Smith R. E., 2000, *MNRAS*, 318, 1144
- Planck Collaboration, Aghanim N., Akrami Y., Ashdown M., Aumont J., Baccigalupi C., Ballardini M., Banday A. J. et al, 2018, arXiv:180706209
- Pozzetti L., Hirata C. M., Geach J. E., Cimatti A., Baugh C., Cucciati O., Merson A., Norberg P. et al, 2016, *A&A*, 590, A3
- Salvador A. I., Sánchez F. J., Pagul A., García-Bellido J., Sanchez E., Pujol A., Frieman J., Gaztanaga E. et al, 2019, *MNRAS*, 482, 1435
- Sánchez A. G., Baugh C. M., Angulo R. E., 2008, *MNRAS*, 390, 1470
- Schechter P., 1976, *ApJ*, 203, 297
- Seljak U., 2000, *MNRAS*, 318, 203
- Seo H.-J., Eisenstein D. J., 2003, *ApJ*, 598, 720
- Shim H., Colbert J., Teplitz H., Henry A., Malkan M., McCarthy P., Yan L., 2009, *ApJ*, 696, 785
- Skibba R., Sheth R. K., Connolly A. J., Scranton R., 2006, *MNRAS*, 369, 68
- Skibba R. A., Smith M. S. M., Coil A. L., Moustakas J., Aird J., Blanton M. R., Bray A. D., Cool R. J. et al, 2014, *ApJ*, 784, 128
- Smith A., 2018, PhD thesis, Durham University
- Smith A., Cole S., Baugh C., Zheng Z., Angulo R., Norberg P., Zehavi I., 2017, *MNRAS*, 470, 4646
- Smith R. E., Scoccimarro R., Sheth R. K., 2008, *Phys. Rev. D*, 77, 043525
- Sobral D., Best P. N., Geach J. E., Smail I., Cirasuolo M., Garn T., Dalton G. B., Kurk J., 2010, *MNRAS*, 404, 1551
- Sobral D., Best P. N., Geach J. E., Smail I., Kurk J., Cirasuolo M., Casali M., Ivison R. J. et al, 2009, *MNRAS*, 398, 75
- Sobral D., Best P. N., Matsuda Y., Smail I., Geach J. E., Cirasuolo M., 2012, *MNRAS*, 420, 1926
- Sobral D., Smail I., Best P. N., Geach J. E., Matsuda Y., Stott J. P., Cirasuolo M., Kurk J., 2013, *MNRAS*, 428, 1128
- Song Y.-S., Percival W. J., 2009, *J. Cosmology Astropart. Phys.*, 10, 004
- Spergel D., Gehrels N., Baltay C., Bennett D., Brekinridge J., Donahue M., Dressler A., Gaudi B. S. et al, 2015, arXiv:150303757
- Spergel D. N., Verde L., Peiris H. V., Komatsu E., Nolte M. R., Bennett C. L., Halpern M., Hinshaw G. et al, 2003, *ApJS*, 148, 175
- Springel V., White S. D. M., Jenkins A., Frenk C. S., Yoshida N., Gao L., Navarro J., Thacker R. et al, 2005, *Nature*, 435, 629
- Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, *MNRAS*, 328, 726
- Tegmark M., Peebles P. J. E., 1998, *ApJ*, 500, L79
- Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, *ApJ*, 724, 878
- Valentino F., Daddi E., Silverman J. D., Puglisi A., Kashino D., Renzini A., Cimatti A., Pozzetti L. et al, 2017, *MNRAS*, 472, 4878
- Verde L., Heavens A. F., Percival W. J., Matarrese S., Baugh C. M., Bland-Hawthorn J., Bridges T., Cannon R. et al, 2002, *MNRAS*, 335, 432
- Wang Y., 2008a, *J. Cosmology Astropart. Phys.*, 5, 021
- , 2008b, *Phys. Rev. D*, 77, 123525
- Zehavi I., Blanton M. R., Frieman J. A., Weinberg D. H., Mo H. J., Strauss M. A., Anderson S. F., Annis J. et al, 2002, *ApJ*, 571, 172
- Zehavi I., Zheng Z., Weinberg D. H., Blanton M. R., Bahcall N. A., Berlind A. A., Brinkmann J., Frieman J. A. et al, 2011, *ApJ*, 736, 59
- Zhai Z., Tinker J. L., Hahn C., Seo H.-J., Blanton M. R., Tojeiro R., Camacho H. O., Lima M. et al, 2017, *ApJ*, 848, 76

## APPENDIX A: COMPONENT FITTING FOR THE HALO OCCUPATION DISTRIBUTIONS

As described in § 2.2.2, we generate a library of luminosity-dependent HODs obtained by running the GALACTICUS galaxy formation model on snapshots from the Millennium Simulation. In order to randomly sample the HODs for halo masses beyond the tabulated range, we choose to fit functional forms to the separate central galaxy and satellite galaxy components.

### A1 Fitting the central galaxy component

Given the occupation numbers of central galaxies,  $N_{\text{cen}}$ , we find that for the faintest luminosity limits  $N_{\text{cen}} \sim 1$  for  $M_{200} > 10^{11} h^{-1} M_{\odot}$ . As discussed in § 2.2.2, for brighter luminosity limits  $N_{\text{cen}}$  becomes peaked around  $10^{12} h^{-1} M_{\odot}$ , with a slight increase towards higher mass. We find that the increase in  $N_{\text{cen}}$  for halo masses above  $10^{14} h^{-1} M_{\odot}$  can be modelled using the linear relation,

$$\log_{10}(-\log_{10}(N_{\text{cen}})) \propto m_{\text{fit}} \log_{10}(M_{200}) \quad (\text{A1})$$

where  $m_{\text{fit}}$  is the fitted slope, which will take a different value for each luminosity limit. Unfortunately, for the brightest luminosity limits the tabulated HODs are quite noisy and there is the potential for the HODs for the different luminosity limits to cross, which must be avoided if we are to draw random samples from the HODs. In order to address these two issues, we take the values for  $m_{\text{fit}}$  and fit a second linear relation such that we have a description for the slope as a function of blended luminosity threshold,  $m_{\text{fit}}(\log_{10}(L_{\text{H}\alpha+\text{[NII]}}))$ . This relation can be then be extrapolated to fit the slopes for luminosities for which the tabulated HOD is noisy and also ensures that the central galaxy occupation distributions for the luminosity limited samples never cross.

The fits to the tabulated occupation numbers of central galaxies are shown in Fig. A1. Whilst these linear relations generally provide reasonably good fits to the tabulated numbers, there are

some small discrepancies, particularly when examining the brightest luminosity limits at lower redshift and higher halo mass. However, we note from Fig. 2 that there are very few bright  $H\alpha$ -emitting galaxies in the most massive halos at these redshifts and so these small discrepancies in the fitting do not have a significant impact on the number densities of central galaxies in our lightcone catalogues.

## A2 Fitting the satellite galaxy component

For the occupation numbers of satellite galaxies, we find that the satellite component of the tabulated HODs are well fit by a double power law. For the faintest luminosity limits a power law fit for masses above  $10^{13}h^{-1}M_{\odot}$  is used to extrapolate the occupation number to higher halo mass.

The fits to the tabulated satellite galaxy occupation numbers are shown in Fig. A2. These fits are able to provide a reasonable match to the tabulated satellite galaxy occupation numbers, even for the brightest luminosity limits.

Just as with the central galaxies, at lower redshifts the satellite occupation numbers for the brightest luminosity limits are quite noisy and power law fits for these bright luminosity limits lead to crossing of the HODs, which we wish to minimise. Therefore for these brightest limits we simply apply a vertical offset to the double power law fitted to occupation numbers for the  $\log_{10}(L_{H\alpha+[NII]}/\text{erg s}^{-1}\text{cm}^{-2}) > 42.14$  limit. These fits, which can be seen in the upper row of panels in Fig. A2, lead to an underestimate of the number of satellite galaxies, though since the fraction of satellite galaxies in these bright luminosity samples is relatively small, these under-estimates do not have a significant impact on the number density of satellite galaxies in our lightcone catalogues.

## APPENDIX B: LIGHTCONE REDSHIFT DISTRIBUTIONS

In Table B1 we provide the redshift distributions,  $dN/dz$ , for the MXXL-15K and the MXXL-2K lightcones. Galaxy counts are shown per unit redshift, per square degree. The MXXL-15K lightcone has an  $H\alpha + [NII]$  blended flux limit of  $2 \times 10^{-16}\text{erg s}^{-1}\text{cm}^{-2}$  and the MXXL-2K lightcone has an  $H\alpha + [NII]$  blended flux limit of  $1 \times 10^{-16}\text{erg s}^{-1}\text{cm}^{-2}$ . The galaxy fluxes have had dust attenuation applied and we provide the galaxy counts for both a WISP-calibrated attenuation and a HiZELS-calibrated attenuation.

## APPENDIX C: LINEAR BIAS FITS FOR MXXL-15K AND MXXL-2K LIGHTCONES

In this section we report the results of the clustering analysis for the MXXL-15K and MXXL-2K lightcones. This analysis is carried out in an identical manner to that for the Euclid-like and WFIRST-like surveys (see § 4) but with the difference that we do not apply any incompleteness to the lightcone catalogues. In other words, this analysis corresponds to making linear bias forecasts for idealised Euclid-like and WFIRST-like surveys that have 100 per cent completeness. The MXXL-15K lightcone, which has a  $H\alpha + [NII]$  blended flux limit of  $2 \times 10^{-16}\text{erg s}^{-1}\text{cm}^{-2}$ , corresponds to an idealised Euclid-like survey, whilst the MXXL-2K lightcone, which has a  $H\alpha + [NII]$  blended flux limit of  $1 \times 10^{-16}\text{erg s}^{-1}\text{cm}^{-2}$ , corresponds to an idealised WFIRST-like survey.

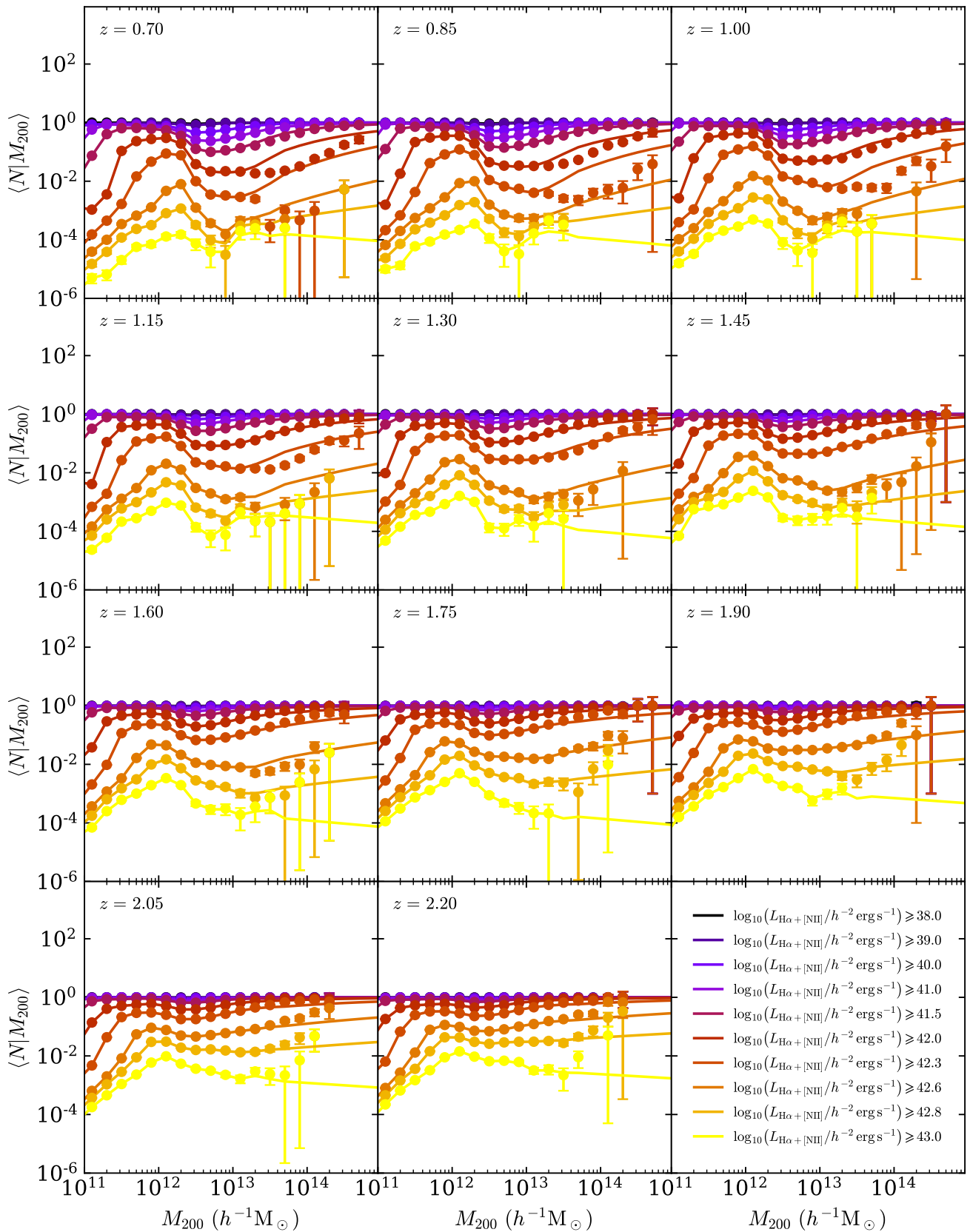
**Table B1.** Sky-averaged redshift distributions,  $dN/dz$ , for galaxies in the WISP-calibrated and HiZELS-calibrated versions of the MXXL-15 and MXXL-2K lightcones. The MXXL-15K lightcone has an  $H\alpha + [NII]$  blended flux limit of  $2 \times 10^{-16}\text{erg s}^{-1}\text{cm}^{-2}$  and the MXXL-2K lightcone has an  $H\alpha + [NII]$  blended flux limit of  $1 \times 10^{-16}\text{erg s}^{-1}\text{cm}^{-2}$ . The redshift column corresponds to the bin centre and the distributions correspond to the number of galaxies per unit redshift per square degree on the sky. Galaxy counts do not include any incompleteness due to instrument-dependent efficiency.

Redshift	MXXL-15K		MXXL-2K	
	WISP	HiZELS	WISP	HiZELS
0.725	12338	8247	23471	18003
0.775	12334	8103	24751	18587
0.825	11484	7392	23701	17539
0.875	11394	7214	24299	17827
0.925	10688	6574	23952	17398
0.975	9983	5943	23100	16732
1.025	9115	5307	22454	16194
1.075	8068	4542	21200	15093
1.125	7435	4027	20606	14489
1.175	6681	3555	20225	14120
1.225	5849	3125	19074	13176
1.275	4993	2630	17489	11882
1.325	4419	2279	16699	11220
1.375	3888	1997	15842	10537
1.425	3345	1766	14444	9477
1.475	2845	1530	13468	8670
1.525	2473	1365	12213	7935
1.575	2212	1259	10957	7357
1.625	1974	1165	10082	7019
1.675	1707	1058	9072	6527
1.725	1446	943	8087	5961
1.775	1323	905	7642	5756
1.825	1213	870	6939	5392
1.875	1041	774	6122	4920
1.925	904	699	5566	4654
1.975	842	685	5270	4599
2.025	743	636	4879	4424
2.075	680	613	4560	4292
2.125	652	619	4240	4139
2.175	583	581	3797	3855

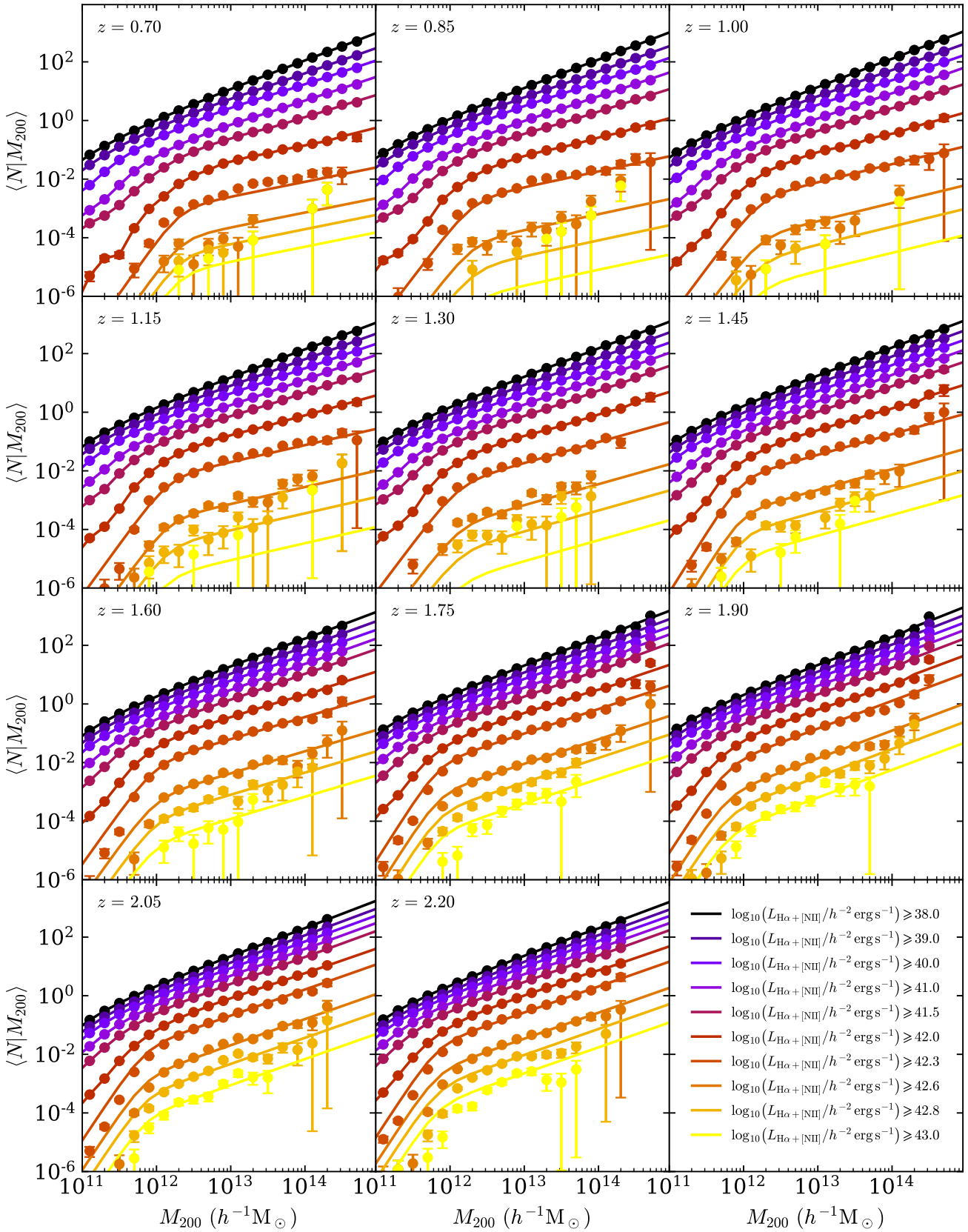
Just as with the clustering analysis for the Euclid-like and WFIRST-like surveys, we divide the MXXL-15K and MXXL-2K lightcones into 5 equal redshift bins of width  $\Delta z = 0.2$  spanning  $0.9 \leq z < 1.9$  for the MXXL-15K lightcone and  $1 \leq z < 2$  for the MXXL-2K lightcone. The correlation function is computed 5 times for each redshift bin, each time using a different random seed to select the random catalogue. We again consider two versions of the lightcones: one calibrated to reproduce the WISP number counts and one calibrated to reproduce the HiZELS luminosity functions. The whole analysis is again repeated twice, firstly selecting the galaxies in redshift-space and secondly selecting the galaxies in real-space. In Table C1 we show for each instance of MXXL-15K analysis the effective redshift, the mean number of galaxies and the mean ratio for the number of galaxies to the number of randoms used for the clustering analysis. Table C2 shows the equivalent numbers for the MXXL-2K analyses.

The galaxy correlation functions for the MXXL-15K lightcone are shown in upper row of each grid in Fig. C1, where the dark red lines show the mean correlation functions in redshift-space (averaged over 5 repeat calculations), and the lighter red lines show





**Figure A1.** Central galaxy occupation distributions for 10 blended luminosity-selected samples from GALACTICUS, plotted at selected redshifts between  $z = 0.7$  and  $z = 2.2$ . The solid lines show the fits to the occupation numbers as discussed in Appendix A1. The various colours of the lines indicate the blended  $H\alpha + [NII]$  luminosity limit used to select the galaxies, as indicated in the colour bar. The halo mass assumed is the mass within an over-density with average density corresponding to 200 times the mean density of the Universe.



**Figure A2.** Satellite galaxy occupation distributions for 10 blended luminosity-selected samples from GALACTICUS, plotted at selected redshifts between  $z = 0.7$  and  $z = 2.2$ . The solid lines show the fits to the occupation numbers as discussed in Appendix A2. The various colours of the lines indicate the blended  $\text{H}\alpha + [\text{NII}]$  luminosity limit used to select the galaxies, as indicated in the colour bar. The halo mass assumed is the mass within an over-density with average density corresponding to 200 times the mean density of the Universe.

the mean correlation functions in real-space (also averaged over 5 repeat calculations). The upper grid assumes a WISP-calibrated version of the lightcone and the lower grid assumes a HiZELS-calibrated version of the lightcone. In every instance the BAO peak is clearly identifiable and we see a larger clustering amplitude in redshift-space compared to real-space, consistent with the expectations from Kaiser (1987). Comparing Fig. C1 with the upper grids of Fig. 9 and Fig. 10 we can see that for the MXXL-15K our correlation functions are less noisy than those for the Euclid-like survey due to the lack of incompleteness. In Fig. C2 we show the equivalent results for the MXXL-2K lightcone where we again are able to identify the BAO peak in each redshift bin, though with perhaps less significance in the highest redshift bins.

The lower row in each grid of panels in Fig. C1 and Fig. C2 show the measured bias,  $b(r)$ , computed from the galaxy and dark matter correlation functions. Comparing our bias measurements to the equivalent measurements for the Euclid-like and WFIRST-surveys (Fig. 9 and Fig. 10) we see again see that the bias measurements are consistent with a constant for scales  $r \lesssim 75 h^{-1} \text{Mpc}$  and that at larger scales we see pronounced scale-dependent deviations away from a constant value. As discussed in § 4.3.1, these deviations are most likely a result of distortions to the BAO caused by a combination of non-linear collapse, mode coupling effects and redshift-space distortions.

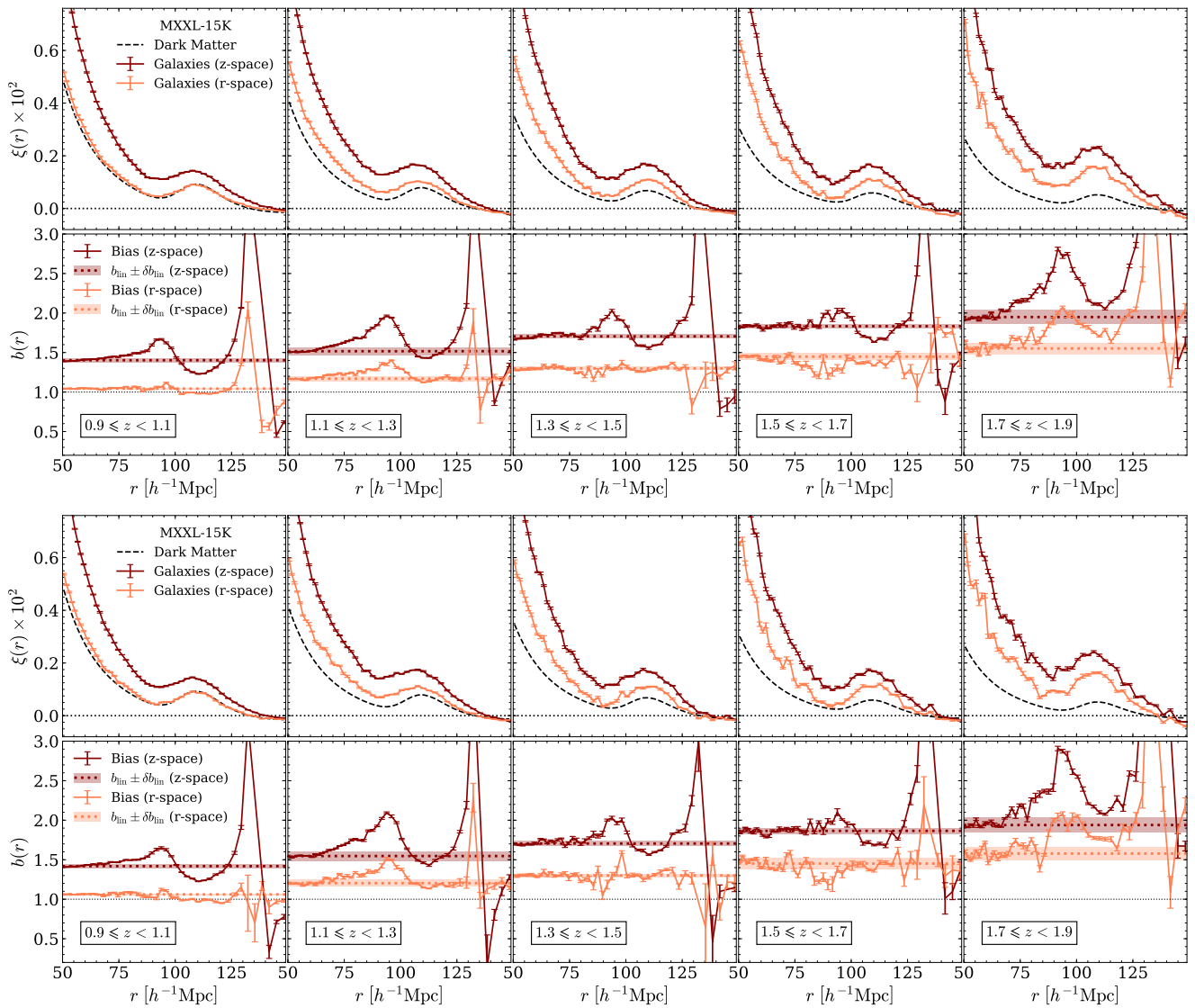
The horizontal dotted lines in the lower row in each grid in Fig. C1 and Fig. C2 show the fit for the linear bias on scales  $r \lesssim 75 h^{-1} \text{Mpc}$ , obtained using  $\chi^2$  minimisation to fit a zeroth order polynomial. The shaded regions show the RMS of the residuals (see §4.3.2 for details). The linear bias fits for the MXXL-15K and MXXL-2K lightcones are reported in Table C1 and Table C2 respectively. For the MXXL-15K and MXXL-2K bias results we again see that the linear bias is larger in redshift-space and increases with increasing redshift. Comparing the linear bias values in Table C1 and Table 2 we see that for the corresponding redshift bins the linear bias values for the MXXL-15K lightcone and the Euclid-like survey are consistent with one another, suggesting that introduction of incompleteness has a negligible impact on our estimates for the linear bias. We also find that there is no difference in the mean halo mass. If we compare the linear bias values in Table C2 and Table 3 we see a similar result for the MXXL-2K lightcone and the WFIRST-like survey.

**Table C1.** Catalogue specifications and linear bias fits for the MXXL-15K lightcone, measured in redshift-space (upper table) and in real-space (lower table). In each instance, the properties shown are: the effective redshift of the slice,  $z_{\text{eff}}$ ; the mean number of galaxies used to compute the correlation function,  $\bar{N}_{\text{gal}}$ ; the mean value for ratio of randoms to galaxies,  $\bar{N}_{\text{ran}}/\bar{N}_{\text{gal}}$ ; the mean and standard deviation in the halo mass,  $M_h$ , for that galaxy sample; and the linear bias fit. The columns show the results for each redshift slice considered. Each table shows the bias fits when adopting a WISP-calibrated version of the lightcone and a HiZELS-calibrated version of the lightcone.

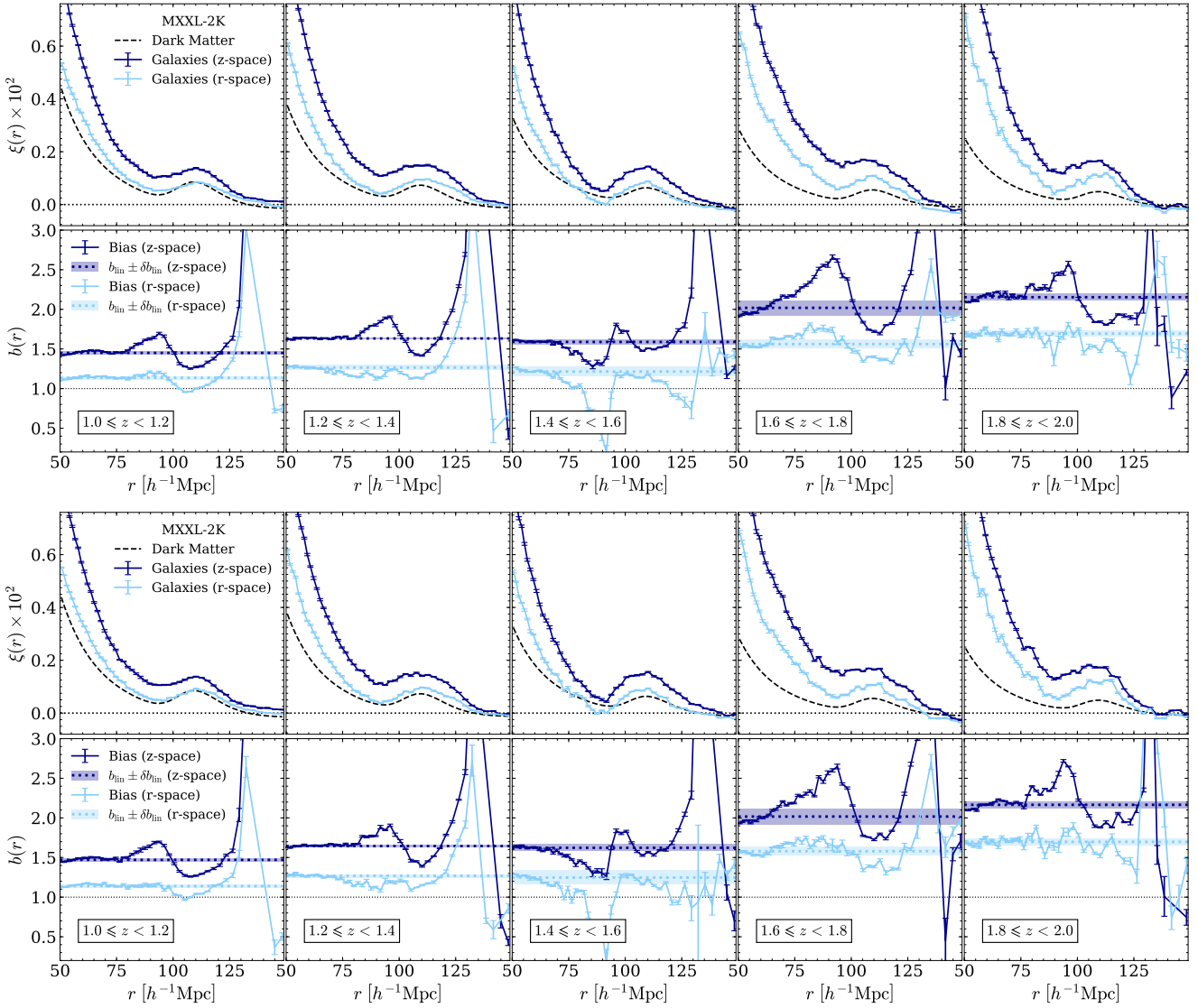
Calibration	Property	$0.90 \leq z < 1.1$	$1.1 \leq z < 1.3$	$1.3 \leq z < 1.5$	$1.5 \leq z < 1.7$	$1.7 \leq z < 1.9$
<b>MXXL-15K (redshift-space)</b>						
WISP	$z_{\text{eff}}$	0.991	1.19	1.38	1.59	1.79
	$\bar{N}_{\text{gal}}$	29281758	19305864	11214261	6472414	3886909
	$\bar{N}_{\text{ran}}/\bar{N}_{\text{gal}}$	10	10	10	15	25
	$\log_{10}(M_h/h^{-1}M_{\odot})$	$11.8 \pm 0.4$	$11.8 \pm 0.3$	$11.8 \pm 0.3$	$11.9 \pm 0.3$	$11.9 \pm 0.3$
	$b_{\text{lin}} \pm \delta b_{\text{lin}}$	$1.40 \pm 0.02$	$1.52 \pm 0.05$	$1.70 \pm 0.03$	$1.83 \pm 0.02$	$1.95 \pm 0.09$
HiZELS	$z_{\text{eff}}$	0.988	1.19	1.39	1.59	1.79
	$\bar{N}_{\text{gal}}$	7785058	4642807	2635798	1687841	1215424
	$\bar{N}_{\text{ran}}/\bar{N}_{\text{gal}}$	10	10	15	25	29
	$\log_{10}(M_h/h^{-1}M_{\odot})$	$11.8 \pm 0.3$	$11.9 \pm 0.3$	$11.9 \pm 0.3$	$11.9 \pm 0.3$	$11.9 \pm 0.3$
	$b_{\text{lin}} \pm \delta b_{\text{lin}}$	$1.42 \pm 0.03$	$1.55 \pm 0.05$	$1.69 \pm 0.04$	$1.88 \pm 0.05$	$1.9 \pm 0.1$
<b>MXXL-15K (real-space)</b>						
WISP	$z_{\text{eff}}$	0.991	1.19	1.38	1.59	1.79
	$\bar{N}_{\text{gal}}$	29272710	19293708	11223651	6467080	3887359
	$\bar{N}_{\text{ran}}/\bar{N}_{\text{gal}}$	10	10	10	15	25
	$\log_{10}(M_h/h^{-1}M_{\odot})$	$11.8 \pm 0.4$	$11.8 \pm 0.3$	$11.8 \pm 0.3$	$11.9 \pm 0.3$	$11.9 \pm 0.3$
	$b_{\text{lin}} \pm \delta b_{\text{lin}}$	$1.043 \pm 0.007$	$1.17 \pm 0.03$	$1.30 \pm 0.02$	$1.45 \pm 0.04$	$1.55 \pm 0.07$
HiZELS	$z_{\text{eff}}$	0.988	1.19	1.39	1.59	1.79
	$\bar{N}_{\text{gal}}$	17295737	10310660	5863143	3747817	2702461
	$\bar{N}_{\text{ran}}/\bar{N}_{\text{gal}}$	10	10	10	15	25
	$\log_{10}(M_h/h^{-1}M_{\odot})$	$11.8 \pm 0.3$	$11.9 \pm 0.3$	$11.9 \pm 0.3$	$11.9 \pm 0.3$	$11.9 \pm 0.3$
	$b_{\text{lin}} \pm \delta b_{\text{lin}}$	$1.06 \pm 0.01$	$1.20 \pm 0.04$	$1.30 \pm 0.02$	$1.45 \pm 0.07$	$1.58 \pm 0.08$

**Table C2.** Catalogue specifications and linear bias fits for the MXXL-2K lightcone, measured in redshift-space (upper table) and in real-space (lower table). The meanings of the various properties are the same as in Table C1. The columns show the results for each redshift slice considered. Each table shows the bias fits when adopting a WISP-calibrated version of the lightcone and a HiZELS-calibrated version of the lightcone.

Calibration	Property	$1.0 \leq z < 1.2$	$1.2 \leq z < 1.4$	$1.4 \leq z < 1.6$	$1.6 \leq z < 1.8$	$1.8 \leq z < 2.0$
<b>MXXL-2K (redshift-space)</b>						
WISP	$z_{\text{eff}}$	1.10	1.29	1.49	1.69	1.89
	$\bar{N}_{\text{gal}}$	8448622	6910632	5108407	3488459	2390012
	$\bar{N}_{\text{ran}}/\bar{N}_{\text{gal}}$	10	20	20	20	30
	$\log_{10}(M_h/h^{-1}M_{\odot})$	$11.7 \pm 0.4$	$11.7 \pm 0.4$	$11.7 \pm 0.4$	$11.8 \pm 0.4$	$11.8 \pm 0.4$
	$b_{\text{lin}} \pm \delta b_{\text{lin}}$	$1.45 \pm 0.02$	$1.63 \pm 0.01$	$1.59 \pm 0.03$	$2.02 \pm 0.09$	$2.15 \pm 0.04$
HiZELS	$z_{\text{eff}}$	1.09	1.29	1.49	1.69	1.89
	$\bar{N}_{\text{gal}}$	5989807	4681804	3344242	2526554	1956709
	$\bar{N}_{\text{ran}}/\bar{N}_{\text{gal}}$	10	20	20	20	30
	$\log_{10}(M_h/h^{-1}M_{\odot})$	$11.7 \pm 0.4$	$11.8 \pm 0.4$	$11.8 \pm 0.3$	$11.8 \pm 0.3$	$11.8 \pm 0.4$
	$b_{\text{lin}} \pm \delta b_{\text{lin}}$	$1.47 \pm 0.02$	$1.64 \pm 0.01$	$1.62 \pm 0.05$	$2.0 \pm 0.1$	$2.17 \pm 0.04$
<b>MXXL-2K (real-space)</b>						
WISP	$z_{\text{eff}}$	1.10	1.29	1.49	1.69	1.89
	$\bar{N}_{\text{gal}}$	8454205	6901558	5112198	3499443	2390544
	$\bar{N}_{\text{ran}}/\bar{N}_{\text{gal}}$	20	20	20	20	20
	$\log_{10}(M_h/h^{-1}M_{\odot})$	$11.7 \pm 0.4$	$11.7 \pm 0.4$	$11.7 \pm 0.4$	$11.8 \pm 0.4$	$11.8 \pm 0.4$
	$b_{\text{lin}} \pm \delta b_{\text{lin}}$	$1.14 \pm 0.02$	$1.27 \pm 0.03$	$1.22 \pm 0.06$	$1.56 \pm 0.06$	$1.69 \pm 0.04$
HiZELS	$z_{\text{eff}}$	1.09	1.29	1.49	1.69	1.89
	$\bar{N}_{\text{gal}}$	5993966	4675862	3346543	2534087	1957573
	$\bar{N}_{\text{ran}}/\bar{N}_{\text{gal}}$	20	20	20	20	22
	$\log_{10}(M_h/h^{-1}M_{\odot})$	$11.7 \pm 0.4$	$11.8 \pm 0.4$	$11.8 \pm 0.3$	$11.8 \pm 0.3$	$11.8 \pm 0.4$
	$b_{\text{lin}} \pm \delta b_{\text{lin}}$	$1.14 \pm 0.02$	$1.27 \pm 0.02$	$1.25 \pm 0.09$	$1.58 \pm 0.06$	$1.70 \pm 0.04$



**Figure C1.** Clustering results and bias fits for the MXXL-15K lightcone. The upper grid of panels shows the results when adopting a WISP-calibrated version of the lightcone and the lower grid of panels shows the results when adopting a HiZELS-calibrated version of the lightcone. In each instance the darker lines show the results in redshift-space, the brighter lines correspond to real-space (i.e. assuming the cosmological redshifts of the galaxies with no peculiar velocity component), and the black dashed lines show the dark matter correlation function computed at the effective redshift of the bin. In the lower panels of each grid, the horizontal dotted lines and shaded regions correspond to the linear bias fits assuming the appropriate correlation function (see text for details). The redshift range used for selection is shown in the bottom left-hand corner of the lower panels.



**Figure C2.** Clustering results and bias fits for the MXXL-2K lightcone. The upper grid of panels shows the results when adopting a WISP-calibrated version of the lightcone and the lower grid of panels shows the results when adopting a HiZELS-calibrated version of the lightcone. All the lines and shaded regions have the same meanings as in Fig. C1. The redshift range used for selection is shown in the bottom left-hand corner of the lower panels.