Mon. Not. R. Astron. Soc. **000**, 1–18 (2015) (MN LATEX style file v2.2)

The environment and host haloes of the brightest $z \sim 6$ Lyman-break galaxies

P. W. Hatfield[1⋆], R.A.A. Bowler[1], M.J. Jarvis[1,2], C.L. Hale[1]
[1] *Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford, OX1 3RH, UK*
[2] *Department of Physics, University of the Western Cape, Bellville 7535, South Africa*

ABSTRACT

By studying the large-scale structure of the bright high-redshift Lyman-break galaxy (LBG) population it is possible to gain an insight into the role of environment in galaxy formation physics in the early Universe. We measure the clustering of a sample of bright ($-22.7 < M_{UV} < -21.125$) LBGs at $z \sim 6$ and use a halo occupation distribution (HOD) model to measure their typical halo masses. We find that the clustering amplitude and corresponding HOD fits suggests that these sources are highly biased ($b \sim 10$) objects in the densest regions of the high-redshift Universe. Coupled with the observed rapid evolution of the number density of these objects, our results suggest that the shape of high luminosity end of the luminosity function is related to feedback processes or the onset of dust obscuration - as opposed to a scenario where these sources are predominantly rare instances of the much more numerous $M_{UV} \sim -19$ population of galaxies caught in a particularly vigorous period of star formation. Despite investigating several variations on the model, we struggle to simultaneously fit both the number densities and clustering measurements. We interpret this as a signal that a refinement of the model halo bias relation at high redshifts or the incorporation of quasi-linear effects may be needed for future attempts at modelling the clustering and number counts. Finally, the difference in number density between the fields (UltraVISTA has a surface density$\sim$ 1.8 times greater than UDS) is shown to be consistent with the cosmic variance implied by the clustering measurements.

Key words: galaxies: evolution – galaxies: star-formation – galaxies: high-redshift – techniques: photometric – clustering – LBGs

1 INTRODUCTION

1.1 Lyman-break Galaxies

The study of Lyman-break Galaxies (LBGs) is a long established probe of the high-redshift Universe (the first few billion years), with samples of many hundreds of star-forming galaxies now known to $z \sim 10$ (e.g. Oesch et al. 2016; Bouwens et al. 2016; McLeod et al. 2016). LBGs are particularly useful as it is possible to establish their photometric redshift to reasonable accuracy in a luminosity regime where spectroscopic confirmation is challenging (e.g. Pentericci et al. 2014). The neutral gas in the inter-galactic medium (IGM) is essentially opaque to photons with wavelengths shorter than the 'Lyman Break' (1216Å, in the far ultraviolet). The source therefore appears faint bluewards of this wavelength, but retains its original luminosity redwards, creating a sharp drop in luminosity. When this spectrum is then redshifted, the location of the break provides a clear spectral feature with which to select galaxies to high-redshifts using broad-band filters.

The technique, originally developed in the early 1990's (Guhathakurta et al. 1990; Steidel & Hamilton 1992; Steidel et al. 1996) in the context of $z \sim 3$ galaxies, where the Lyman break is shifted into visible wavebands, first started providing large numbers of sources with the *Hubble Space Telescope* (HST) in the late 1990's and 2000's (e.g. Giavalisco & Mauro 2002; Bouwens et al. 2007; Dunlop et al. 2013). More recently, the approach is being used to push scientific boundaries at $z \sim 6-9$ where the break is shifted into the near-infrared (see Stark 2016 for a review). Wide field surveys like the United Kingdom Infrared Telescope (UKIRT) Infrared Deep Sky Survey (UKIDSS, in particular the Ultra Deep Survey, UDS, Hartley et al. 2013), and more recently

⋆ peter.hatfield@physics.ox.ac.uk

public surveys on the Visible and Infrared Survey Telescope for Astronomy (VISTA) such as the UltraVISTA survey in the COSMOS field (McCracken et al. 2012) and the VISTA Deep Extragalactic Observations (VIDEO) survey (Jarvis et al. 2013) give access to the deep NIR images of the sky needed for detecting statistically significant samples of the brightest LBGs, which has lead to advances in the understanding of their star formation rates and number densities beyond the break in the luminosity function.

A key observable that can be calculated for LBG surveys (and galaxy surveys in general), is the luminosity or mass function, the comoving number density of galaxies as a function of absolute luminosity or stellar mass (see Johnston 2011 for a review). Measuring and understanding the evolution of luminosity functions with redshift allows us to trace the build-up and evolution of galaxies through cosmic time (Madau & Dickinson 2014); is a key way to compare cosmological simulations of structure formation to observations (Lacey et al. 2016; Clay et al. 2015); and can be readily linked theoretically to the dark matter equivalent, the halo mass function (HMF; the comoving density of dark matter halos as a function of halo mass, see Murray et al. (2013) for a review of current constraints). Luminosity functions are typically observed to have the form of a Schechter function: $n(L) = \phi^*(L/L^*)^\alpha \exp(-L/L^*)$ (Schechter 1976). In this parametrisation, α describes the power law behaviour of number density at the low-luminosity end, L^* is the transition luminosity to the high luminosity exponential cutoff, and ϕ^* is a normalisation. The rest frame UV luminosity function for $z \sim 4-8$ has been determined by several studies (e.g. recent work by McLure et al. 2013; Bouwens et al. 2015; Finkelstein et al. 2015; Bowler et al. 2015) with broad agreement. The highest redshift constraints on the LBG luminosity function are currently at $z \sim 9-10$ e.g. Bouwens et al. (2015, 2016); McLeod et al. (2016).

1.2 Clustering

Galaxies are formed and live in dark matter halos, and the environment of the host halo is believed to be of critical importance for the formation of the resident galaxies (Cooray & Sheth 2002). One way of obtaining information about the galaxy-halo connection is 'abundance matching' - matching the galaxy comoving number density value to the halo mass that is predicted to have the same number density by theoretical considerations of the halo mass function/N-body simulations e.g. rarer galaxies are associated with more massive halos because such halos are rarer (Vale & Ostriker 2004). Abundance matching however can only ever give an incomplete account of the connection due to three complications. Firstly, halos can host multiple galaxies, this can be partially mitigated through *sub*-halo matching (Moster et al. 2010), but this assumes that the occupation statistics are the same for sub-halos and isolated halos. Secondly, scatter in the halo mass to galaxy mass/luminosity relation is not captured by abundance matching. Finally, variations in observational properties of a single population can bias results, in particular orientation or temporal effects e.g. if a given population has a different appearance 10 percent of the time, a straightforward abundance matching will erroneously place these sources with a different appearance in more massive halos as they are rarer, even though they are the same object as the underlying population. For this reason, other probes of the galaxy halo connection are needed. Information from lensing is very effective, either strong (e.g. Jullo et al. 2007) or weak (galaxy-galaxy lensing, e.g. Coupon et al. 2015; Mandelbaum et al. 2013), but requires the background sources to be at even higher redshift than the lenses, making it unfeasible for high ($z > 2$) galaxies.

One viable and popular approach is to measure the clustering (2-point statistics) of the galaxies alongside the number counts (1-point statistics). This can then be linked to models/our theoretical understanding of structure formation to estimate the typical environment of the galaxies. One popular framework for modelling galaxy clustering is the 'Halo Occupation Distribution' model (Zehavi et al. 2005), which models the non-linear clustering of galaxies within individual halos, and the large scale clustering of the halos simultaneously, giving information about how many galaxies are in each halo as a function of halo mass. The HOD model has been applied extensively at $z = 0$ (Guo et al. 2015) and $z = 0.5-2$ (McCracken et al. 2015; Coupon et al. 2015; Hatfield et al. 2016) where large galaxy samples are available. In the more uncertain high-redshift regime, the HOD model has recently been applied to low-luminosity LBG galaxies at $z = 4-7$ by Harikane et al. (2016). It is crucial however to understand the relationship at the massive/most luminous end, as this is where AGN-driven feedback may have a role (Silk & Joseph 2011). There are preliminary hints that redshifts of $z = 6-7$ may mark the onset of quenching (Bowler et al. 2014, 2015), so this is a vital time period for galaxy evolution in the history of the Universe.

1.3 Probes of Reionization

As well as being a crucial period for galaxy formation (see Shapley 2011 for a review), understanding large-scale structure/clustering at $z = 5-8$ is also important cosmologically, in particular for our understanding of *reionization*. In the standard cosmological model, the Universe was initially an ionised plasma, that during *recombination* at $z \sim 1000$ cooled sufficiently for protons and electrons to combine to form neutral atoms. However this medium must have been reionized by the first stars and galaxies at some point between then and $z \sim 6$ to produce the ionized intergalactic medium we see today (see Zaroubi 2013; Natarajan & Yoshida 2014 for an observational and theoretical review respectively). However there is still debate in the literature about which sources had the energy to do this e.g. massive galaxies, AGN, or something else (Grissom et al. 2014; Robertson et al. 2015; Bouwens et al. 2015). Furthermore, most models of reionzation are 'patchy' e.g. non-instantaneous, with some parts of the Universe being reionised earlier than others (Becker et al. 2015; Doré et al. 2007). The current best constraint on the average reionization redshift from the Planck mission (based on the mea-

sured optical depth) is $z = 7.8 - 8.8$ (Adam et al. 2016), with many probes of the epoch (e.g. Pentericci et al. 2014; Becker et al. 2015) suggesting that some parts of the Universe could still be undergoing reionization by $z \sim 6$. Uncovering the cause and nature of reionization is likely to require understanding how the reionization power spectrum interweaves with galaxy large scale structure, so it is essential to build up our understanding of the large scale structure (LSS) of the galaxy population at these redshifts. For example, McQuinn et al. (2007) suggest that differences in the clustering of LBGs and Lyman-alpha Emitting galaxies (LAEs) could give an insight into the possible 'patchy' nature of reionization. The Lyman-alpha line is suppressed if the source is in a largely neutral region which biases the observations of LAEs towards large ionized HII regions. The result is a larger 'observed' clustering for LAEs than the 'intrinsic' clustering of the underlying objects - effectively neutral regions obstruct the line of sight in a way that enhances the clustering of LAEs. LBGs and other probes of the high redshift galaxy population however do not receive such an effect on their clustering, so a boost in the clustering of LAEs relative to LBGS (properly controlling for other variables) could be indicative of reionization. Such an effect is yet to be conclusively measured, e.g. Ouchi et al. (2010) find little evidence at $z = 6.6$ with 207 LAEs observed with the Subaru telescope, but this approach and others like it are likely to give improved constraints to the nature of the epoch of reionization over the coming years.

1.4 Objectives of this work

LBG studies can be informally divided into analyses of 'faint' galaxies (in extremely deep, but narrow surveys), and 'bright' galaxies (in slightly less deep, but extremely wide surveys). Harikane et al. (2016) provide an analysis of the clustering of relatively faint LBGs found within HST deep surveys at $z = 4 - 7$. In this study we seek to extend these measurements to brighter luminosities by utilising wider area surveys. To do this, we measure and model the clustering of the Bowler et al. (2015) high luminosity $z \sim 6$ sample, which covers the degree-scale UDS and UltraVISTA fields. A clustering analysis of a subset of the UDS sample has been performed in McLure et al. (2009), who modelled the correlation function with a single power law, concluding the sources are in dark matter haloes of masses $10^{11.5-12} M_{\odot}$. In this study we perform a similar analysis, but extend to a full HOD model, including an additional field and using deeper data available. Using this enlarged sample, we are then able to discuss what our results mean for feedback processes, models of structure formation, and cosmic variance at high redshift. While samples of bright galaxies do exist at $z > 6.5$ (Bowler et al. 2015), they are too small to provide constraints on the clustering, and hence we limit our analysis to $z \sim 6$.

The structure of this paper is as follows. In Section 2 we describe the sample of LBGs used in this study. In Section 3 we discuss how we measured the correlation function in the sample and constructed our halo occupation distribution models and fitting process. The results are presented in Section 4. In Section 5 we discuss our results, linking them to the literature, and interpreting the cosmic variance between the fields in light of our measurements. Magnitudes are given in the AB system (Oke & Gunn 1983) and all calculations are in the concordance cosmology $\sigma_8 = 0.8$, $\Omega_{\Lambda} = 0.7$, $\Omega_m = 0.3$ and $H_0 = 70$ km s^{-1}Mpc^{-1} unless otherwise stated.

2 DATA AND SAMPLE SELECTION

In this study we use the high luminosity Lyman break galaxy sample of Bowler et al. (2015). Deep optical and infrared data (spanning wavelengths of $0.3 - 2.5\mu$m) across two main fields (see Fig. 1) was used to select the sample; we summarise the observations and selection criteria below, but see Bowler et al. (2015) for a more in depth description.

2.1 UltraVISTA/COSMOS

UltraVISTA (McCracken et al. 2012; Laigle et al. 2016) is the deepest of the 6 public surveys on the VISTA telescope, providing $YJHK_s$ near infra-red data covering the Cosmic Evolution Survey (COSMOS) field (Scoville et al. 2007). The 'paw-print' focal plane of VISTA and the survey observing strategy give a continuous 'deep' field, with discontinuous 'ultra-deep' stripes across it that receive more observing time. Bowler et al. (2015) also used u^*, g, r and i optical data from the T0007 release of CFHTLS in the D2 field, as well as Subaru/SuprimeCam z'-band imaging. The maximal area of overlap of these datasets is in the one square degree of CFHTLS, of which 0.62 deg^2 has ultra-deep UltraVISTA data, and 0.38 deg^2 shallower UltraVISTA. The majority of the sample is in the ultra-deep field and hence for our purposes here we only use the ultra-deep 0.62 deg^2 (see Fig. 1).

2.2 UDS

For the UKIDSS UDS field, Bowler et al. (2015) used B, V, R, i and z' data from the Subaru XMM-Newton Deep Survey (SXDS, Furusawa et al. 2008), and J, H and K band data from the DR10 of the UKIDSS UDS (Lawrence et al. 2007). Again separate z'-band data from Subaru/SuprimeCam was obtained, and in addition, Y band data from the VIDEO survey (Jarvis et al. 2013) was also used. The total overlapping area is 0.74 deg^2 (see Fig. 1).

2.3 Candidate Selection

Again, Bowler et al. (2015) describes the full sample selection, but we summarise the process here. Sources were detected with SExtractor (Bertin & Arnouts 1996), and photometric redshifts were determined with LePhare (Arnouts et al. 1999; Ilbert et al. 2006). Contaminant populations (low redshift interlopers and brown dwarfs) were removed in the SED fitting process. This leaves 156 and 107

$5.5 < z < 6.6$ galaxies in the UltraVISTA and UDS fields respectively. The UltraVISTA field was found to have a higher surface density than the UDS field (by a factor of ~ 1.8); potential causes for this, including lensing and cosmic variance are discussed in section 7 of Bowler et al. (2015).

This process gives in total 263 LBGs in the range $-22.7 < M_{UV} < -20.5$ with $5.5 < z < 6.5$ over 1.35 deg^2. We take our final sample as all 161 sources with $M_{UV} < -21.125$, as the sample completeness drops off rapidly faintwards of this value, as discussed in Bowler et al. (2015), see their figure 6, but is fairly constant with magnitude brightwards of this value.

3 CORRELATION FUNCTIONS AND HOD MODELLING

There is a large selection of statistical measurements that can be used to characterise the clustering of extragalactic sources and large-scale structure, including nearest neighbour (Bahcall & Soneira 1983), genus (Gott et al. 2009), power spectrum (Tegmark & Max 2003) and counts in cells (White 1979). In this study we measure and model the two-point correlation function, the excess probability of how much more likely two galaxies are to be at a given separation than a random uniform distribution (this statistic can be linked to other measurements e.g. counts in cells statistics are 'averaged' correlation functions, and the power spectrum is the Fourier transform of the correlation function).

The underlying meaningful physical relation is the full three dimensional spatial correlation function; however we only have the observables of angular separations and relatively coarse redshift information. Limber Inversion (Limber 1954) provides a way to connect the two. The two main approaches to connect the observables to the spatial correlation function are to either calculate the angular correlation function, and compare to angular projections of the model, or to use the redshift information to form the projected correlation function in both transverse and longitudinal directions (incorporating redshift space distortions, which are normally integrated out), see Davis & Peebles (1983) and Fisher et al. (1994). Here we measure and model the angular correlation function, as calculating the projected correlation function requires precise knowledge of the redshifts of the sample to avoid being biased, and is in general more appropriate for surveys with spectroscopic redshifts.

3.1 The Angular Correlation Function

The angular two-point correlation function $\omega(\theta)$ is defined by:

$$dP = \sigma(1 + \omega(\theta))d\Omega, \tag{1}$$

where dP is the probability of finding two galaxies at an angular separation θ, σ is the surface number density of galaxies, $d\Omega$ is the solid angle. This is most commonly estimated by calculating $DD(\theta)$, the normalised number of galaxies at a given separation in the real data, and $RR(\theta)$, the corresponding figure for a synthetic catalogue of random galaxies identical to the data catalogue in every way (i.e. occupying the same field) except position. We use the Landy & Szalay (1993) estimator:

$$\omega(\theta) = \frac{DD - 2DR + RR}{RR}, \tag{2}$$

which also uses $DR(\theta)$, data to random pairs, as it has a lower variance (as an estimator) and takes better account of edge effects.

Uncertainties are calculated with 'bootstrap resampling', which samples the galaxies with replacement from the dataset, from which we recalculate the correlation function (see Ling et al. 1986). Repetition of this process produces multiple 'realisations' of the correlation function, from which the covariance matrix of the $\omega(\theta)$ values can be estimated. It is possible to calculate the error bars from Poisson uncertainty on the DD values, but Cress et al. (1996) and Lindsay et al. (2014) found errors calculated in this manner were a factor of 1.5 to 2 smaller than those estimated with bootstrap. It is particularly key to account for covariance between adjacent angular space bins in the small-number counts regime here as each galaxy will contribute to multiple bins. In this paper we use 100 bootstrap resamplings to estimate the uncertainty at the 16th and 84th percentiles.

For the construction of our random catalogue we created a mask over the fields to exclude image artefacts and foreground stars. Five galaxies in the UDS field were found to be within the masked area. Although the mask may be a little conservative, it is likely that our measurements of clustering in the vicinity of these sources will be heavily biased by the presence of the artefact being masked, so we do not use these five galaxies when calculating the correlation function (although it makes very little difference to our analysis). The fact that the survey area has a finite area gives a negative offset to the true correlation function, usually known as the integral constraint. As per Beutler et al. (2011) and Hatfield et al. (2016), we calculate the integral constraint using the numerical approximation of Roche & Eales (1999) and treat it as part of the model when fitting parameters. In this paper we calculate the correlation function with both the binning method (DD and RR are how many galaxy pair separations in each angular scale bin) and the continuous estimation/kernel smoothing method described in Hatfield et al. (2016).

3.2 Halo Occupation Distribution modelling

Halo Occupation Modelling is an increasingly popular way of modelling galaxy clustering measurements. We do not describe the full details of the scheme here, see Coupon et al. (2012) and McCracken et al. (2015) for a more complete breakdown. A given set of galaxy occupation statistics is given, usually parametrised by 3-5 numbers, e.g. the number of galaxies in a halo as a function of halo mass. The model correlation function is broken down to a '1-halo' term,

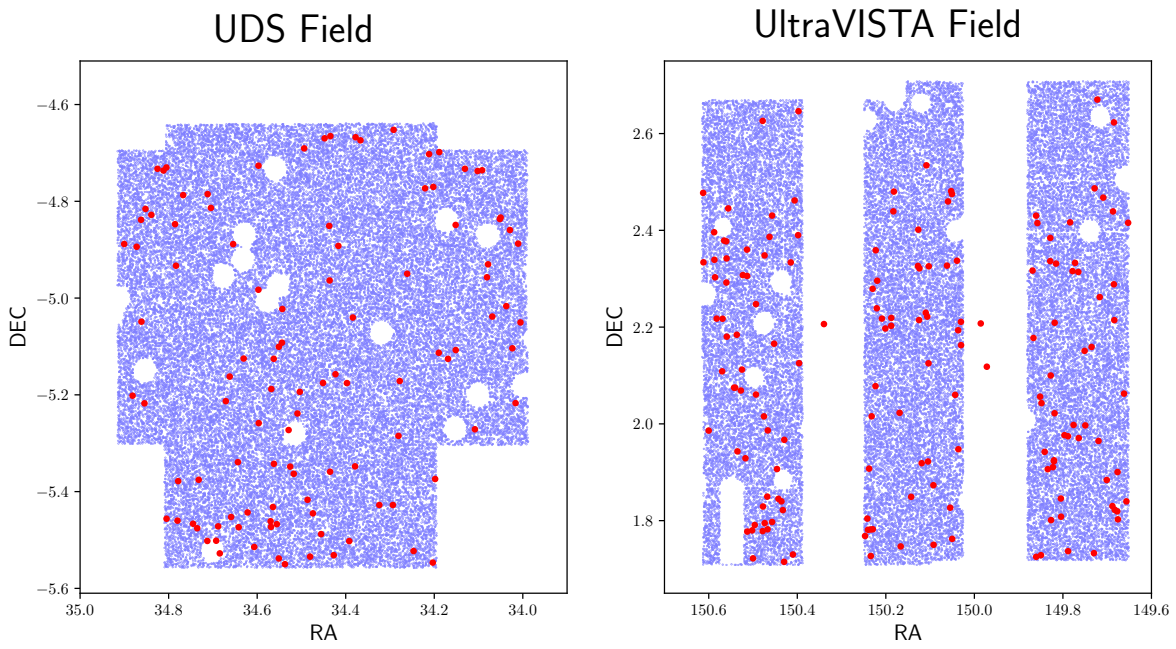

Figure 1. The geometry of the UDS and UltraVISTA fields. The red points are the galaxy locations from Bowler et al. (2015). The blue points are the random points chosen to cover the fields used for the construction of RR for the calculation of the correlation function. The three galaxies in the UltraVISTA field that are not surrounded by blue points are the $z = 6$ sources detected in the deep (as opposed to ultra-deep) part of the UltraVISTA field, that we do not include in this study. The overall shape of the fields is predominantly determined by the part of the sky that the multiple different surveys overlap in. The small scale gaps and holes are foreground stars and detector artefacts etc. See figures 1 and 2 of Bowler et al. (2014) to see how the irregular footprints arise from the intersection of the sky patches covered by different surveys.

describing the small-scale clustering of galaxies within an individual halo, and a '2-halo' term, describing the clustering of the halos themselves. The '1-halo' term is constructed by convolving the profile of galaxies within a halo with itself, weighting by the number of galaxies in the halo, and then integrating over all halo masses. The profile is usually taken to be one galaxy at the centre of the halo (the 'central') and all other galaxies tracing a Navarro-Frank-White (NFW; Navarro et al. 1996) profile. The 2-halo term is constructed by scaling the dark matter linear correlation function by the weighted-average halo bias of the host halos.

The most general HOD parametrisation commonly used is that of Zehavi et al. (2005), that gives the total number of galaxies in a halo as:

$$\langle N_{tot}(M_h)\rangle = \langle N_{cen}(M_h)\rangle + \langle N_{sat}(M_h)\rangle, \quad (3)$$

the total number of central galaxies as:

$$\langle N_{cen}(M_h)\rangle = erf\left(\frac{M_h - M_{min}}{\sigma_{\log M}}\right), \quad (4)$$

and the total number of satellites as:

$$\langle N_{sat}(M_h)\rangle = \langle N_{cen}(M_h)\rangle \left(\frac{M_h - M_0}{M_1}\right)^{\alpha}. \quad (5)$$

This model has five parameters; M_{min} describes the minimum halo mass required to host a central galaxy, $\sigma_{\log M}$ describes how sharp this step jump is (equivalently to the central to halo mass scatter), M_0 is a halo mass below which no satellites are found, and M_1 is the scale mass for accumulating satellites (M_0 is typically a lot smaller than M_1, so M_1 is commonly said to be the halo mass at which the first satellite is accreted, although analytically they are very slightly different - this is the difference between M_1 and M_1' used by some authors). The power law index α describes how the number of satellites grows with halo mass. The random variables of the number of galaxies of the sample under consideration in a halo as a function of halo mass are N_{cen} and N_{sat} (e.g. they are not absolute values). The halo has a central galaxy (of the given galaxy population) with a probability given by Equation 4 and no central with the complementary probability (which is why the expected value for the number of centrals is given by Equation 4). N_{sat} is zero if N_{cen} is zero, and Poisson distributed with mean $\left(\frac{M_h - M_0}{M_1}\right)^{\alpha}$ if $N_{cen} = 1$, giving the expectation in Equation 5. However, although we have the largest sample of bright

LBGs at these redshifts, this is still only a comparatively small sample for HOD modelling. Thus, in order to reduce the number of parameters in the model (six once duty cycle is included, see Section 3.3), we fix some as functions of others.

As per Harikane et al. (2016) we fix $\sigma_{\log M} = 0.2$. The assumptions that go into this choice however are based on results at much lower redshifts (Kravtsov et al. 2004; Zheng et al. 2005; Conroy et al. 2006) which do not necessarily hold at these early times, when the luminosity-halo scatter is fairly unconstrained. Indeed Hatfield et al. (2016) found a scatter of ~ 0.6 consistent with the data at $z \sim 1$. However fortunately for our purposes (unfortunately from the perspective of using clustering to infer the scatter) the 2-point statistics have very little dependence on the scatter. Hence our conclusions do not alter dramatically with choice of $\sigma_{\log M}$, and so we fix it as the same as the Harikane et al. (2016) value for ease of comparison.

We additionally fix $\alpha = 1$; this is both the fiducial value (it is logical to expect that once in the most massive halo regime that the number of satellites scales linearly with the halo mass, as the bulk of the halo mass will have been accreted), as well as the result found by most measurements at moderate ($z < 2$) redshift (e.g. Hatfield et al. 2016; Coupon et al. 2012).

We investigate the consequences of allowing various parameters to be fixed or free in the fitting process. Again as per Harikane et al. (2016), if not free, M_1 and M_0 are fixed as functions of M_{min} following the $z = 0 - 5$ results of Conroy et al. (2006):

$$\log M_1 = 1.18 \log M_{min} - 1.28, \tag{6}$$

$$\log M_0 = 0.76 \log M_1 + 2.3. \tag{7}$$

In this work we use the halo mass function of Behroozi et al. (2013), and the halo bias of Tinker et al. (2010).

3.3 Duty Cycle

The role of a duty cycle (DC) is the main difference to be incorporated when modelling LBG galaxies at high redshift compared with studies in the local Universe. Clustering analyses of LBGs typically find that there is a mismatch between the measured number density, and the number density implied by the clustering (Ouchi et al. 2010). This is in agreement with current understanding of galaxies at these redshifts that suggest that star formation may be highly episodic e.g. Stark et al. (2009). Typically the occupation statistics model implied by fitting *only* to the clustering will suggest a larger comoving number density than is observed in the luminosity function. This discrepancy is typically explained by invoking a *duty cycle*, that the observed LBGs have luminosities that vary dramatically in time, and are being observed only when in a bright phase. This illustrates the importance of understanding clustering alongside the number counts. With a duty cycle of 10 percent (i.e. it is only in its bright phase 10 percent of the time), the underlying galaxy appears 10 times rarer than it actually is. A straight abundance matching in this scenario would then mistakenly put them in rarer, and thus more massive, halos.

Without incorporating the duty cycle, the implied comoving number density is the mean number of galaxies in a halo, times the halo mass function, integrated over all halo masses. This number is then multiplied by the duty cycle to give the model comoving density:

$$n_{gal} = \mathrm{DC} \times \int_0^\infty \mathrm{HMF}(M_h) \times \langle N_{tot}(M_h) \rangle \mathrm{d}M_h, \tag{8}$$

where HMF is the halo mass function and DC is the duty cycle.

3.4 MCMC Fitting

To compare with observations, we use the Halomod [1] code (Murray, Power, Robotham, in prep.) to calculate the spatial correlation function. We then project this to an angular correlation function (as per Limber 1954), and subtract off the numerical approximation of the integral constraint to get our final model correlation function.

We use Emcee [2] (Foreman-Mackey et al. 2013) to provide a Markov Chain Monte Carlo sampling of the parameter space to fit our correlation function. We use a likelihood of:

$$\chi^2 = \frac{[\log n_{\mathrm{gal}}^{\mathrm{obs}} - \log n_{\mathrm{gal}}^{\mathrm{model}}]^2}{\sigma^2_{\log n}} \tag{9}$$

$$+ \sum_{i,j} [\omega^{\mathrm{obs}}(\theta_i) - \omega^{\mathrm{model}}(\theta_i)][C^{-1}_{i,j}][\omega^{\mathrm{obs}}(\theta_j) - \omega^{\mathrm{model}}(\theta_j)], \tag{10}$$

where $n_{\mathrm{gal}}^{\mathrm{obs}}$ is the observed galaxy number density, $n_{\mathrm{gal}}^{\mathrm{model}}$ is the model galaxy number density, $\sigma_{\log n}$ is the error on the log of the number density including both Poisson noise and cosmic variance, θ_i are the angular scales we fit over, ω^{obs} is the observed angular correlation function, ω^{model} is the angular correlation function of a given model, and $C_{i,j}$ is the covariance matrix of the measurements of the correlation function from the bootstrapping.

When the parameters are free, we use a uniform prior over $10 < \log_{10}(M_{\mathrm{min}}/M_\odot) < 13$, $\log_{10}(M_{\mathrm{min}}/M_\odot) < \log_{10}(M_1/M_\odot) < 14$ (uniform in log space) and $0 < \mathrm{DC} < 1$. We used 20 walkers with 1000 steps, which have starting positions drawn uniformly from the prior.

We use 500,000 random data points in this study. As per Hatfield et al. (2016) and Hatfield & Jarvis (2016), we use 100 bootstrap resamplings to estimate the uncertainty at the 16th and 84th percentiles of the resampling. For $n_{\mathrm{gal}}^{\mathrm{obs}}$, we use the value obtained when integrating the luminosity function brightwards to infinity from $M_{UV} = -21.125$ (as

[1] https://github.com/steven-murray/halomod
[2] http://dan.iel.fm/emcee/current/

opposed to the number obtained by dividing the number of sources by the volume probed) as the luminosity function already has incompleteness factored in. This equates to $n_{\rm gal}^{\rm obs} = 4.1 \times 10^{-5}{\rm Mpc}^{-3}$, for the fields combined. This is from the best-fitting double power law model in Bowler et al. (2015). Using the best-fitting Schecter function gives the marginally lower value of $n_{\rm gal}^{\rm obs} = 3.8{\times}10^{-5}{\rm Mpc}^{-3}$. Changing from one value to the other does not impact our conclusions. The main sources of incompleteness are blending with foreground sources, and misclassification of true $z \sim 6$ LBGs as dwarf stars or lower redshift contaminants, see Bowler et al. (2015).

4 RESULTS

4.1 Clustering Measurements

Fig. 2 shows the angular correlation function of the full sample over the range $10^{-3} < \theta/{\rm deg} < 10^{-0.5}$, estimated with both the binning approach (where galaxy pair separations are counted in discrete angular ranges) and the kernel smoothing method (where the distribution of galaxy pair separations is smoothed to produce a continuous estimation of the correlation function in the angular range under consideration). They (as expected) agree well, and produce the familiar approximate power law $\sim \theta^{-0.8}$, although the kernel smoothing method is able to cope better with bins that contain a small number of pairs. For the rest of our analysis, we take the value of the smoothed correlation function, at the ten angular scales calculated for the bins, as our final measurements [3].

In general, measurements of clustering at different scales will be covariant as individual galaxies contribute multiple times to DD, usually at different scales. Furthermore, extra care with covariances is needed when using the kernel method, as a given galaxy pair contributes at a range of scales (this can be mitigated by picking measurements larger than the smoothing scale, but is important to keep track of here as we are in a low-data regime). Therefore we also construct the covariance matrix from the bootstrapped samples for our measurements, in order to account for these covariances in the fitting process. We show the correlation matrix (the covariance matrix with each value normalised by the standard deviation of each measurement) in Fig. 3. Not taking covariances into account would be the equivalent of ignoring the off-diagonal values, which are non-negligible, particular at the large and very small scales.

[3] With the continuous estimation of the correlation function, one can in principle extract an estimate of the correlation function at an arbitrary number of angular scales in the range probed. However this gives dramatically diminishing returns as adjacent measurements would be increasingly covariant e.g. one could take the estimate of the correlation function at 1000 points in the angular range for which we estimate the correlation function, but adjacent points would be almost perfectly correlated and no extra information would be gained.

4.2 Modelling Results

We carry out several MCMC fits to the data with the HOD model as per Section 3.4, with four variations. The four scenarios considered were:

- M_{min}, M_1 and DC free (e.g. 1-halo and 2-halo amplitudes and number counts all free)
- M_1 fixed as function of M_{min} as per Section 3.2 (e.g. 1-halo amplitude as fixed function of 2-halo amplitude)
- M_1 fixed as function of M_{min}, DC = 0.6 (e.g. 1-halo amplitude as fixed function of 2-halo amplitude, duty cycle fixed at the Harikane et al. 2016 value)
- M_1 fixed as function of M_{min}, DC = 1 (e.g. 1-halo amplitude as fixed function of 2-halo amplitude, no duty cycle - galaxies 'on' at all times)

In addition to these four models we also compute:

- Halo masses for the most straightforwards abundance matching scheme e.g. M_{min} for a sample of galaxies above a given luminosity threshold is the halo mass such that the comoving number density of halos greater than that mass is equal to the comoving number density of the galaxy sample (see Table 2)
- Galaxy bias from a pure bias model e.g. fit for b where $\xi_{\rm model} = b^2 \xi_{\rm DM}$.

The results from these 6 models are shown in Tables 1 and 2. Fig. 4 shows the data, the best-fit models. We show the posterior from one fitting in Fig. 5 for illustrative purposes. It is clear that the amplitude of the correlation function is roughly two orders of magnitude larger than the dark matter correlation function in the linear regime, corresponding to a very high bias. Most of our models suggest that $M_{min} \sim 10^{11.5} M_{\odot}$ e.g. our galaxies are hosted by halos of that mass and above. It also seems that the satellite fraction is at most a few percent, which suggests that at most 5-6 galaxies in our sample are satellites (in the scenario that these sources were the same underlying population as the lower luminosity LBGs, the satellite fraction could have been higher as a non-trivial portion would have been from halos hosting multiple galaxies).

It becomes harder however to make statements beyond these basic claims because the HOD fits are only of moderate quality (see the χ^2 values in Table 1). In general fits decreased in quality as more parameters were fixed, as expected. We discuss the tensions more in section 5.1, but summarise the results of each model here. The M_{min}, M_1 and DC free model is free to go to high masses until tension between model and measured number density stop it from going higher. This model can also take M_1 extremely high, to bring the amplitude of the small scale clustering down to match the data. Models that do not have M_1 free cannot vary their small scale behaviour freely. This forces their halo masses down, as the small scale behaviour grows rapidly with M_{min}; if they went higher the disagreement on small scales would become much larger. When DC is free (and M_1 is fixed as a function of M_{min}), the model actually prefers to go even lower than the abundance matching halo

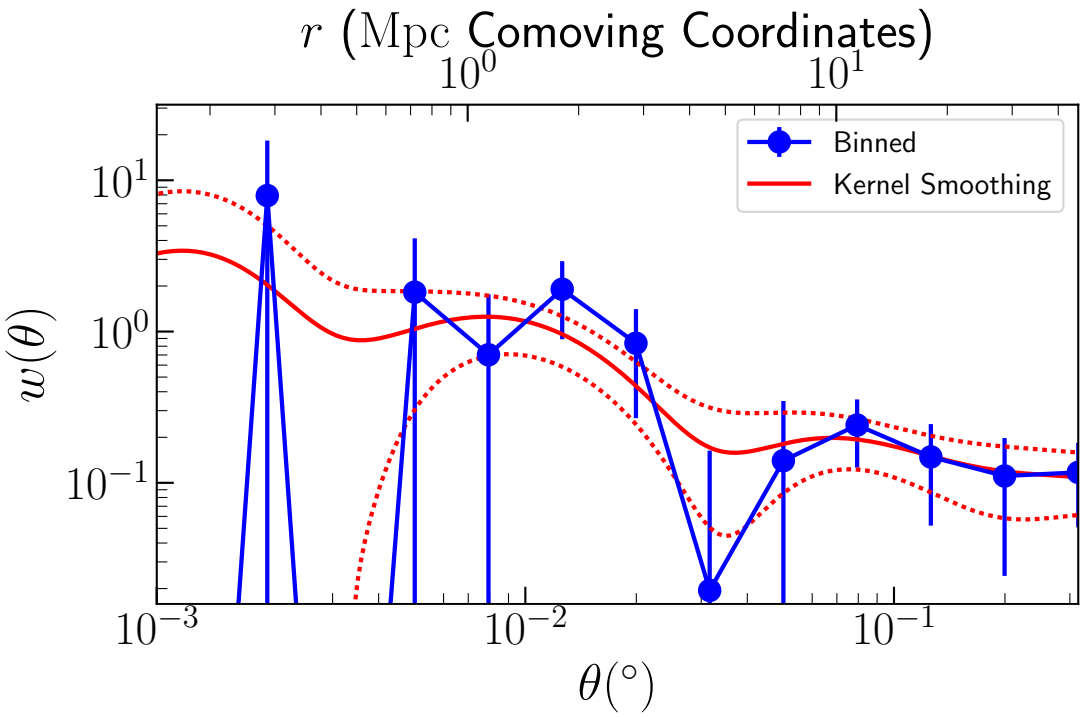

Figure 2. The angular correlation function for our sample of bright ($M_{UV} < -21.125$) $z \sim 6$ LBGs from Bowler et al., (2015). The figure shows the correlation function estimated both with a binning method (blue points), and a kernel smoothing method (red curve, with dotted lines showing uncertainty). Where the binned correlation function dips to negative values corresponds to where there were no galaxy pairs in the bin.

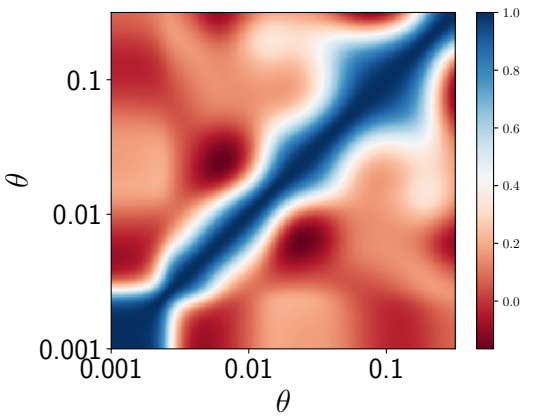

Figure 3. The correlation coefficients (covariance normalised) of our measurements. Blue values are positive correlations, red values are negative values. The red diagonal corresponds to the standard deviation measurements (a random variable is always perfectly correlated with itself).

mass, and uses the duty cycle to reach agreement with the number counts. However when these models have the duty cycle fixed, they cannot do this, so the trade off between agreeing with small scale clustering and the number counts sets the halo mass. For $DC = 0.6$ to agree with the observed number counts, the intrinsic number counts must be higher than for $DC = 1$, forcing the model to prefer slightly lower halo masses. Conversely the 'pure bias' model was able to fit the clustering data well. This suggests that the reason for the poor quality fit is a mismatch between the clustering and the number counts - halos of the halo mass implied by the bias are far rarer than the observed galaxies are [4], which is problematic[5] - although part of the raison d'etre of HOD schemes is to understand how multiple galaxies can occupy the same halo, which would allow the number of galaxies to be greater than the number of halos, this is, as discussed, a few percent effect, as opposed to a factor of ten effect. In addition, the fact that no 1-halo term emerges is slightly anomalous. We note that Harikane et al. (2016) use fitting formulae of the HMF in Tinker et al. (2010) directly without

[4] This problem would have been even worse if the HMF had been taken from Tinker et al. (2010) without the high redshift correction of Behroozi et al. (2013).
[5] This is the opposite problem to what the duty cycle is invoked to solve - duty cycles in clustering studies of LBGs solved the issue of number counts being lower than implied by clustering, the problem here is the number counts are higher.

the normalization constraint, which overestimates the abundance by a factor of 1.7 at $z = 4$ (Y. Harikane 2017, private communication). So the results of Harikane et al. (2016) are likely more consistent with a duty cycle of 1 (rather than the fixed value of 0.6 that they use in their analysis).

5 DISCUSSION

5.1 The link between low- and high-luminosity galaxies and their haloes at $z \sim 6$

The most relevant previous study to compare our results to is Harikane et al. (2016), which presented clustering measurements and HOD fits to $z \sim 6$ LBG galaxies, but on smaller angular scales of approximately $10^{-3.25} < \theta/\mathrm{deg} < 10^{-1.25}$, compared with our $10^{-3} < \theta/\mathrm{deg} < 10^{-0.5}$, and for fainter rest frame absolute magnitudes of $-20.5 < M_{UV} < -19$ compared with our $-22.7 < M_{UV} < -21.125$ sample (see Fig. 6). Thus our results combined with Harikane et al. (2016) describe LBG clustering over almost three orders of angular scale and a factor of 40 in luminosity.

Our bias and halo mass results compared with the results of Harikane et al. (2016) are shown in Fig. 7. Although only moderate quality fits (possible reasons for which are discussed in the subsequent sub-sections), all our fitted models suggest our galaxy sample has a substantially higher typical host halo mass and galaxy bias than the lower luminosity samples in Harikane et al. (2016). This higher bias is evident by directly comparing the two measurements of the correlation function. Our sample has an amplitude ~ 3 times higher than the Harikane et al. (2016) bright ($M_{UV} < -20$) sample with $\omega(0.01^\circ) \sim 0.2$, and our measured bias is a factor of 1.7 greater than that measured for the lower luminosity Harikane et al. (2016) sample (as $\omega \propto b^2$). In general higher luminosity and higher stellar mass galaxy samples have higher biases, but it is important to note that it was not a foregone conclusion to measure a bias this high. It was entirely possible that our $M_{UV} \sim -21.5$ sample could have been the same (or largely the same) population as the sample of Harikane et al. (2016), just observed during a particularly vigorous but rare burst of star formation. If that had been the case, we would have measured a lower clustering amplitude, and inferred a much lower duty cycle. The comoving space density of the galaxies in our sample is 4.1×10^{-5} Mpc^{-3}, compared with 3.8×10^{-4} Mpc^{-3} for the most luminous $z = 6$ Harikane et al. (2016) sample. Harikane et al. (2016) do not measure the duty cycle for this sample, but assume it to be equal to 0.6. As an illustrative example, a duty cycle of 0.6 for the Harikane et al. (2016) sample would mean an actual underlying population comoving density of 6.3×10^{-4}Mpc^{-3}. If our sample was part of the same population, that would correspond to $DC = 0.06$ (in other words that the fainter population spends approximately 6 percent of its time in this super-enhanced state of star formation). However the amplitude of the clustering rules this out and our $M_{UV} < -21.125$ sample is comprised of continuously high-luminousity objects in very dense environments.

5.2 Apparent lack of a 1-Halo term

Models using extrapolated values of M_1 suggest that at scales of $10^{-2.5}$deg and smaller there should be a sharp upturn in the value of the correlation function as the observations start to probe clustering of multiple galaxies within individual halos (see Fig. 4). We do not observe this in the data, in contrast to Harikane et al. (2016), see Fig. 6. A direct interpretation of this would be that M_1 just increases much faster than the extrapolation of Equation 6 i.e. the satellite fraction drops off extremely fast and an unfeasibly large (for this redshift) halo is needed to host two of these sources. Another possibility is suggested by Jose et al. (2013), who also observe a lack of a 1-halo term in clustering measurements of $z \sim 3-7$ LAEs. Their proposed solution was that halo occupancy behaved in a sub-Poissonian manner, and they found that a modified distribution (see their equation 15) was able to reproduce the measurements. However, we suggest that there are good reasons to believe that there is strong cosmic variance on our small scale measurements that is not accounted for in the bootstrap uncertainties, making it hard to make direct inferences about the satellite population of these galaxies.

For single contiguous field observation, cosmic variance is smaller on small scales than on large scales in a limited sense, simply because one observes more instances of small scale structure. However these are not independent instances of small-scale structure, as they all come from the same large-scale density field. To illustrate this, suppose our sources have $M_{\mathrm{min}} \sim 10^{11.4} M_\odot$, then we would expect $M_1 \sim 10^{12.2} M_\odot$ e.g. only halos with $M > 10^{12.2} M_\odot$ host more than one of our bright sample. The comoving density of $M > 10^{12.2} M_\odot$ halos is 5.2×10^{-7}Mpc^{-3} and the comoving volume probed by the observations is $\sim 1.7 \times 10^7$Mpc^3. This means that the expected numbers of $M > 10^{12.2} M_\odot$ halos in the volume surveyed is ~ 10. Just these ten would give ~ 10 close pairs (in addition a few more would be expected from projection effects), which is more than twice the 4 close ($10^{-3} - 10^{-2.5}$deg) pairs observed here, and would push the small scale correlation function up. However these halos will be extremely biased, much more than the $\sim 10^{11.4} M_\odot$ halos. Conceivably for an extreme case, it could be that if our observations were repeated 10 times, we would find that 9 times no $M > 10^{12.2} M_\odot$ halos were observed, and the tenth time a very overdense region is observed, which has 100 in. In the first nine cases, no satellites would be observed, leading to a flat correlation function to small radii that we see in our observations, and the tenth time an overestimate of the satellite fraction is measured. Therefore we would expect there to be very substantial cosmic variance on our measurements of the correlation function on small scales - variance that is not incorporated into our errors on our clustering measurements. Essentially the 1-halo term is dominated by contributions from very massive halos, which are the most biased, so there is the most cosmic variance on small scale measurements of the correlation function. It appears to be the case that neither of our fields are overdense enough to sample the highly biased sample of massive halos at this red-

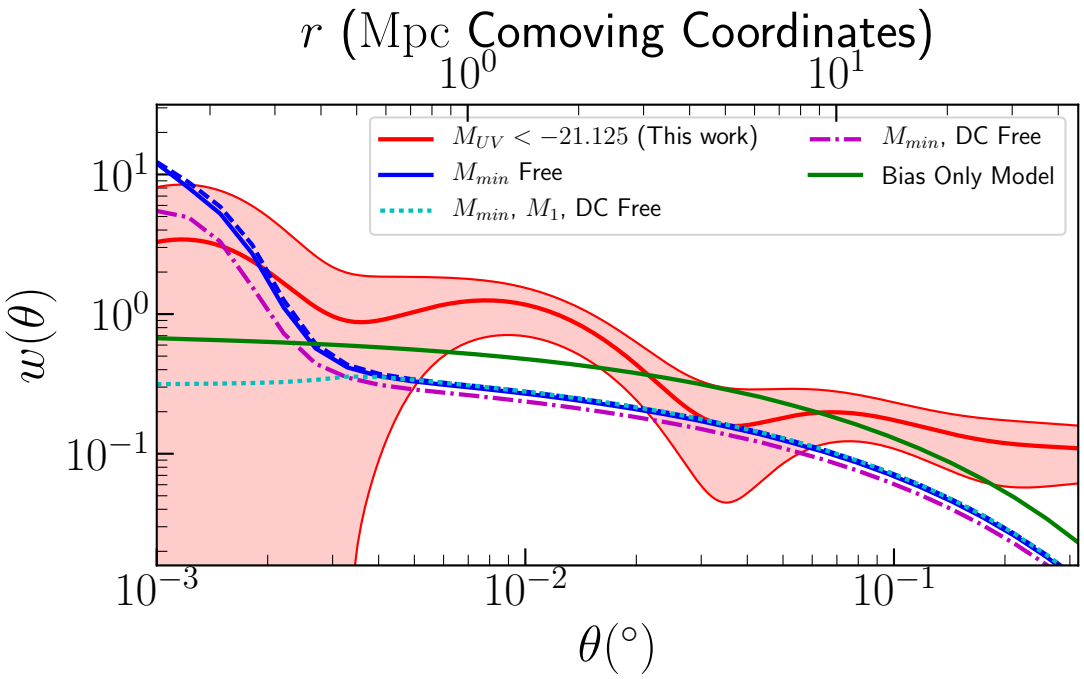

Figure 4. Comparison of our measurements (red curve and shaded area) with the five different models we fit. The two blue curves correspond to models with only M_{min} free, the dashed curve has $DC = 1$ and the full line $DC = 0.6$. On linear solutions all models are very similar, apart from the bias only model, as the number density constraint is restricting the model from going too high. Only the M_1 free model allows the small-scale amplitude to vary independently.

Table 1. Our constraints on the HOD parameters from the MCMC fitting. Also shown are the corresponding satellite fractions ($f_{\rm sat}$) and galaxy biases (b) and fit reduced χ^2 of the samples. Quantities in brackets are either fixed in the model, or fixed as a function of other parameters in the model. Masses are in Solar mass units (log base ten). Note that values and error bars quoted are the 16th, 50th and 84th percentiles of the posterior, as opposed to the peak values. This makes very little difference apart from the posterior for the duty cycle value for the M_{min}, M_1 and DC free model, which is peaked at DC=1 and hence only has one tail, see Fig. 5. The lower luminosity parameter values are taken directly from Harikane et al. (2016), apart from satellite fraction, which we calculate.

Model	M_{UV}	$\log M_{min}$	$\log M_1$	$\log M_0$	α	σ	DC	$10^2 f_{sat}$	b	χ^2/d.o.f.
M_1, DC free	-21.125	$11.53^{+0.05}_{-0.07}$	$13.64^{+0.99}_{-0.91}$	$(12.67^{+0.75}_{-0.69})$	(1)	(0.2)	$0.79^{+0.15}_{-0.25}$	< 0.2	$8.28^{+0.23}_{-0.32}$	1.3
DC free	-21.125	$11.35^{+0.13}_{-0.05}$	$(12.12^{+0.16}_{-0.05})$	$(11.51^{+0.12}_{-0.04})$	(1)	(0.2)	$0.36^{+0.3}_{-0.12}$	$3.87^{+0.24}_{-0.64}$	$7.65^{+0.57}_{-0.18}$	1.5
DC = 0.6	-21.125	$11.48^{+0.02}_{-0.02}$	$(12.26^{+0.03}_{-0.03})$	$(11.62^{+0.02}_{-0.02})$	(1)	(0.2)	(0.6)	$3.29^{+0.11}_{-0.1}$	$8.16^{+0.1}_{-0.1}$	1.3
DC = 1	-21.125	$11.51^{+0.02}_{-0.02}$	$(12.3^{+0.03}_{-0.02})$	$(11.65^{+0.02}_{-0.02})$	(1)	(0.2)	(1)	$3.16^{+0.08}_{-0.1}$	$8.3^{+0.12}_{-0.08}$	1.9
Bias Only	-21.125	NA	NA	NA	NA	NA	NA	NA	$10.86^{+0.1}_{-0.2}$	0.8
Harikane16	-20.0	$11.30^{+0.10}_{-0.13}$	$(12.06^{+0.07}_{-0.16})$	$(11.47^{+0.05}_{-0.12})$	(1)	(0.2)	(0.6)	5.0	$6.3^{+0.4}_{-0.4}$	0.5
Harikane16	-19.1	$11.03^{+0.05}_{-0.18}$	$(11.75^{+0.20}_{-0.29})$	$(11.23^{+0.15}_{-0.22})$	(1)	(0.2)	(0.6)	7.1	$5.5^{+0.2}_{-0.4}$	1.4

shift which could be massive enough to host multiple bright LBGs, and instead our small scale measurements are dominated by the angular projection of the linear clustering (e.g. objects near in angular space by chance, but not near in physical space).

The Harikane et al. (2016) measurements however do have a prominent 1-halo term. We suggest that the reason for this may lie in the fact that a) they are at lower luminosities, so the cosmic variance on halos required to host multiple galaxies is less extreme and b) their correlation functions are measured from galaxies in seven different fields (rather than our two), so they had a greater chance of observing a dense field that had the massive halos necessary for satellites.

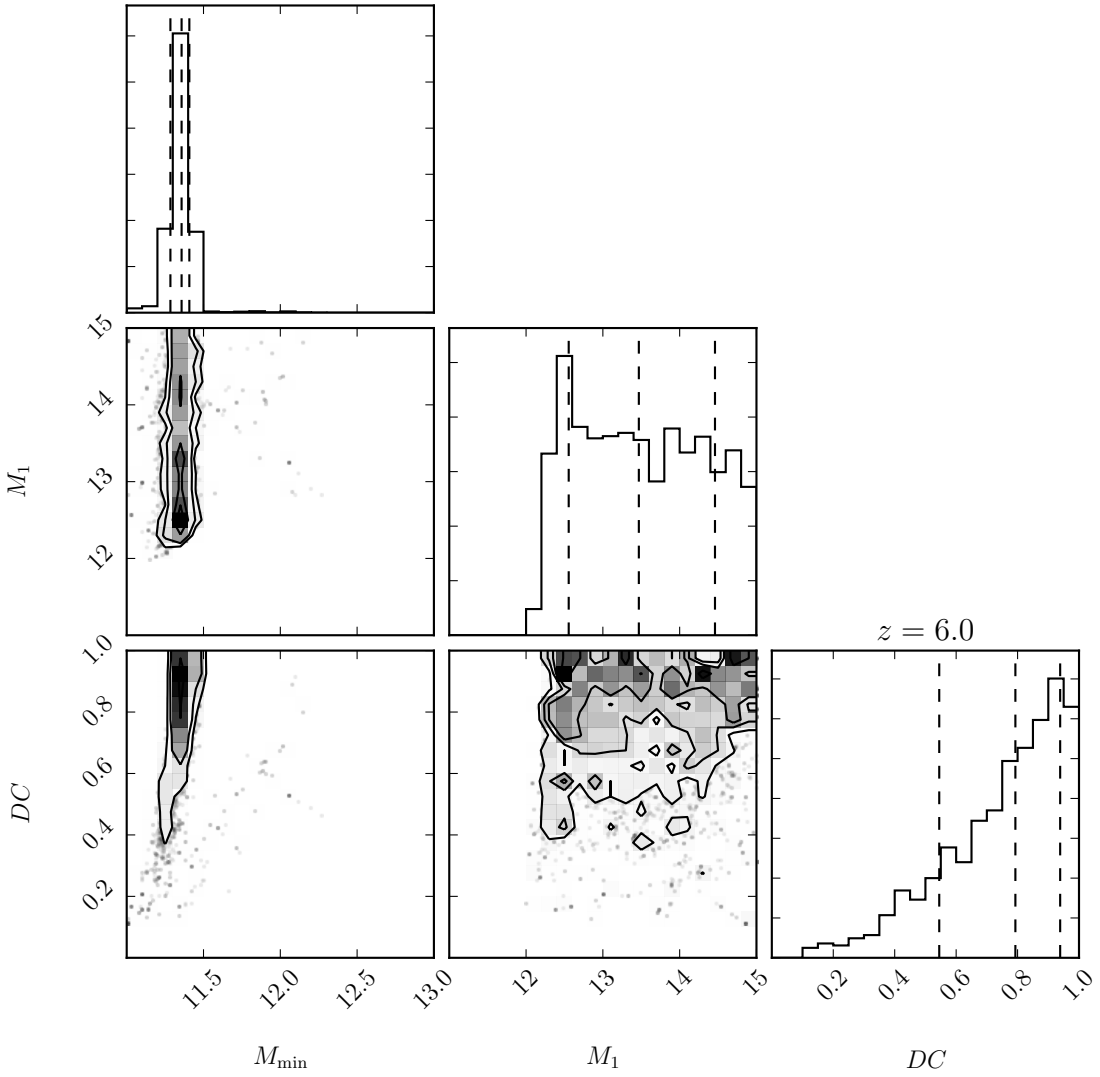

Figure 5. Triangle plot of our posterior from our MCMC fitting for the HOD model with M_{min}, M_1 and DC free (masses in log base ten Solar mass units). Dashed lines on the one dimensional single parameter plots are 16th, 50th and 84th percentiles.

Table 2. Comparison of abundance matching results to clustering fits for our data. Columns are (1) LBG sample used, (2) LBG threshold absolute magnitude, (3), observed comoving number density ($\rm Mpc^{-3}$), (4) the minimum halo mass in the most straightforwards abundance matching scheme (log base ten Solar mass units), (5) the model comoving number density ($\rm Mpc^{-3}$) of the best fit HOD models in this work and Harikane et al., (2016) without incorporating duty cycle, (6) the corresponding minimum halo mass from the HOD model (log base ten Solar mass units).

Data	M_{UV}	$n_g^{\rm observed}$ ($\rm Mpc^{-3}$)	$\log M_{min}^{\rm matched}$	$n_g^{\rm model}$ ($\rm Mpc^{-3}$)	$\log M_{min}^{\rm model}$
Bowler15	-21.125	4.1×10^{-5}	11.51	8.9×10^{-6}	$11.53^{+0.05}_{-0.07}$
Harikane16	-20.0	3.8×10^{-4}	11.09	2.1×10^{-4}	$11.30^{+0.10}_{-0.13}$
Harikane16	-19.1	13.4×10^{-4}	10.79	7.3×10^{-4}	$11.03^{+0.05}_{-0.18}$

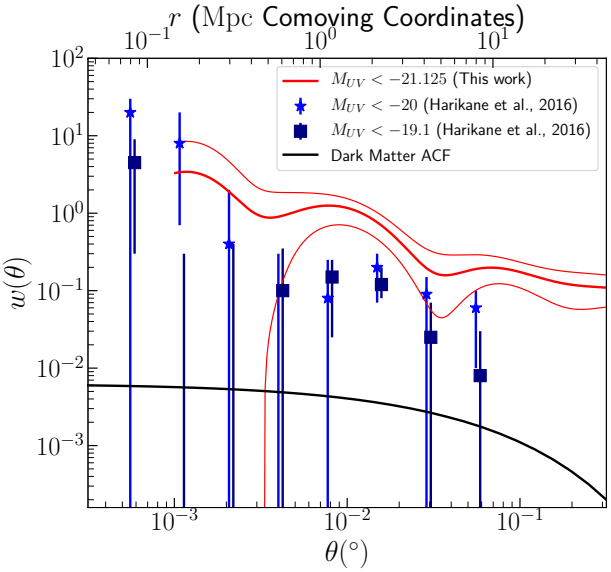

Figure 6. Comparison of our measurements (red curve, 1-σ uncertainties in the lighter curves), the lower luminosity Harikane et al., (2016) measurements, and the dark matter angular correlation function (black curve).

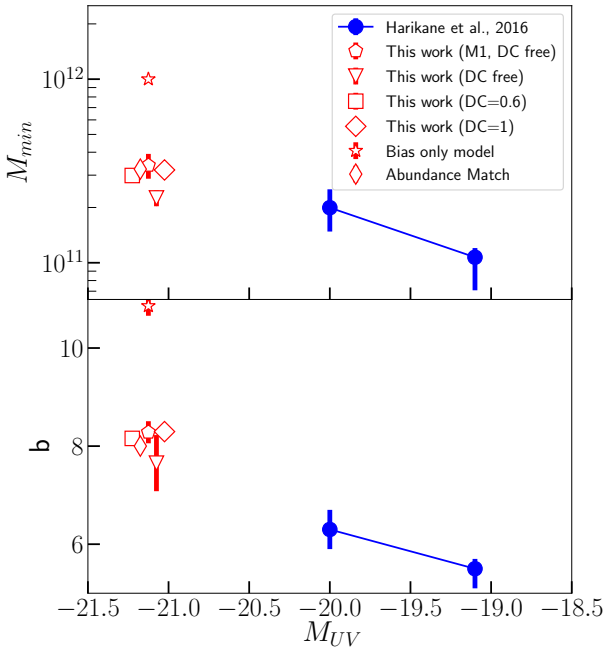

Figure 7. Comparison of our results with comparable measurements of lower luminosity LBGs from Harikane et al. (2016). Top plot: M_{min} as a function of absolute UV luminosity threshold (in units of Solar mass). Bottom plot: galaxy bias as a function of absolute UV luminosity threshold. The results from our six different models are shown for comparison (x-axis values slightly offset for each model for clarity).

5.3 Mismatch between the number counts and bias measurements

As discussed in Section 4.2, it was not possible to obtain a good HOD fit to the number counts and clustering. A good fit was only obtained for the clustering measurements with a pure bias model. The core of the discrepancy is that to obtain the directly measured bias of ~ 10, the galaxies would need to be in halos of minimum mass $M_{\rm min} \sim 10^{12} M_{\odot}$. This corresponds to a comoving density of 2.8×10^{-6}Mpc^{-3}, compared with an observed number density of 4.1×10^{-5}Mpc^{-3} i.e. approximately a factor of 15 lower than the observed value. Similarly, plain abundance matching would suggest minimum mass $M_{\rm min} \sim 10^{11.5} M_{\odot}$ corresponding to a bias of ~ 8. We note that Barone-Nugent et al. (2014) report a very similar issue at $z \sim 7.2$, where they found that a duty cycle of 1 was needed for their LBG sample, and that even then the measured bias was slightly inconsistent with the number density. We discuss in this sub-section possible explanations for this discrepancy.

5.3.1 Contaminants

It is possible that some of the sources used for our clustering measurements are not truly $z \sim 6$ LBGs, but are instead brown dwarfs or galaxies at other redshifts. This is unlikely to be the cause of the discrepancy, as Bowler et al. (2015) had access to photometry across a very large range of wavelengths and performed extensive testing with brown dwarf templates to rule out substantial contamination. Furthermore, stellar contamination would actually reduce the clustering amplitude as stars are unclustered and have no physical correlation with the galaxies. This is also in general the case for contamination from galaxies at other redshifts, which would not be spatially correlated with the region of space probed at $z \sim 6$. One possible exception to this is if a substantial proportion of the sources are actually at $z \sim 1.3$ (the redshift degenerate with $z \sim 6$ when it is difficult to distinguish between the two spectral breaks in SED modelling), in which case we would effectively be measuring the correlation function at $z \sim 1.3$, which would render all modelling so far void. However we dismiss this possibility - although it is possible there are a few $z \sim 1.3$ interlopers in the sample, it seems extremely unlikely that they make up a substantial proportion of the sample a) because of our confidence in the template fitting, and b) because of the deep multi-wavelength data used in Bowler et al. (2015). We conclude

that contamination is unlikely to be a substantial factor in the clustering-number density discrepancy.

5.3.2 More complex galaxy-halo relations

HOD modelling is predicated on the principle that the only thing that determines the galaxy content of a halo is the mass of the halo - if this is violated, then in general more complex relations between the galaxy-halo relation and clustering measurements are possible. Well known cases include assembly bias (see Hearin et al. 2015, where the bias of halos depends on halo assembly history as well as mass), a requirement of a nearby halo to interact with (Cen & Safarzadeh 2015), or a dependence on the large scale linear density field. These effects are potentially plausible - it is possible to imagine a galaxy having a brief starburst (which we then observe as an ultra-bright source) as a result of an interaction in a denser region of the Universe, or for the amount of gas left for star formation in a halo to be related to the age of the halo. However it is in general very hard to distinguish between these different effects if one only has access to the large-scale linear bias (which is effectively a one dimensional measurement). Galaxy-galaxy lensing can in principle observationally break these degeneracies (e.g. if sources are in older, lower mass halos, their clustering will reflect the assembly-dependent bias of the hosts, but the lensing will reflect just the mass), but this is likely to never be possible at these redshifts as it requires a high number density of even higher redshift sources. It seems likely that comparison with simulations is the only way to investigate the viability of such underlying processes.

5.3.3 Uncertainty in knowledge of the high-redshift dark matter distribution

A key input to HOD modelling is our knowledge of the spatial distribution of the underlying dark matter, in particular the HMF and halo bias as a function of halo mass (which comes predominantly from N-body simulations). If the dark matter model used is incorrect, then the conclusions from HOD modelling will also in general be incorrect. As shown in Table 1, abundance matching suggests that the sources are in $> 10^{11.5} M_{\odot}$ halos. However, the bias from the clustering would suggest that they are in $> 10^{12} M_{\odot}$ halos, which are a factor of twenty rarer. Tinker et al. (2008) and Tinker et al. (2010) found model HMFs and halo biases from N-body simulations at redshifts of $z = 0 - 2.5$. Behroozi et al. (2013) then introduced a high-redshift calibration to the Tinker et al. (2010) HMF, extending the validity to $z \sim 8$ (representing an increase of approximately 20 percent at $z = 6$ for $M \sim 10^{11.3} M_{\odot}$ halos). Hence confidence in N-body simulations makes a correction factor of ~ 20 appears implausible, within the current structure formation paradigm. However, Behroozi et al. (2013) did not calibrate the high redshift halo bias, so we are effectively using biases at $z \sim 2.5$ extrapolated to $z \sim 6$. The excess in the clustering amplitude is only around a factor of 50 percent, which would require around a 25 percent correction in bias. Thus we suggest that our results could potentially be explained with a high-redshift calibration to the halo bias function that steepens it at the high mass end. See also Behroozi & Silk (2016) for a discussion on this direction of inference e.g. how high redshift stellar-mass functions can give information on the high-redshift HMF.

An alternate potential correction to our understanding of the distribution of dark matter is the incorporation of 'quasi-linear effects'. HOD modelling makes a binary division between non-linear clustering within halos, and large-scale linear bias. However this transition is gradual, not sharp, and bias can be scale-dependent on up to 10 Mpc scales (relevant for scales probed with our observations), although it always tends to a constant value at large scales (Mann et al. 1997). Introducing a functional form for scale-dependent bias can model some of these effects and Jose et al. (2016) conclude that quasi-linear clustering has the largest effect at high redshift ($z > 2$), and high hight halo mass. In particular, Jose et al. (2017) note a similar discrepancy to ours at $3 < z < 5$, and find that quasi-linear effects can cause one to over estimate halo mass by up to a factor of ten if unaccounted for. They give γ values (the correction to the bias) of order 30-40 percent - around the size of our inconsistency - at the relevant masses and scales relevant for our analysis, so it seems plausible that incorporating quasi-linear effects could solve our discrepancy, and be necessary for future analysis. Jose et al. (2017) also show that quasi-linear effects make the transition into the 1-halo term less sharp, which could explain why we do not observe one, as discussed in section 5.2. Quasi-linear effects are sub-percent at lower redshift, so HOD modelling at lower redshift is not invalidated (Van den bosch et al. 2013).

5.3.4 Modification of Our understanding of high redshift structure formation

Alternatively, it may be the case that N-body simulations do not correctly capture the physics of early structure formation, in a way that no calibration will be able to account for. The potential issue of 'too many' high mass/luminosity galaxies has been identified by Steinhardt et al. (2016). They summarise results that suggest that the best constraints on the HMF at $z = 4-10$ inferred from observations is dramatically higher than the HMF from ΛCDM, with the discrepancy getting worse towards high halo masses and higher redshifts. Although in a challenging observational regime, they suggest that the observations show that current theories of structure assembly at $z > 4$ could be flawed.

Although a possibility, our results are not in sufficient disagreement with models to warrant support of this hypothesis yet - it is necessary to explore the much more likely possibilities of the high-redshift halo bias needing calibration and quasi-linear issues before considering more dramatic changes to theories of structure formation.

5.4 Estimating Cosmic Variance

Cosmic variance is a term that can be used to refer to a number of related but subtly different effects. The specific context in which we use the term here is that many extragalactic statistical measurements vary by more than sample variance between different fields because of large scale structure. As noted in Bowler et al. (2015), the number density of our two fields varies by much more than sample variance assuming a Poisson distribution. This is a consequence of large-scale structure, which our clustering measurements quantify. These clustering measurements can be linked back to the number count estimates to see if the cosmic variance observed is consistent with the clustering measurements, or if one of the fields is over/under dense, even accounting for large-scale structure. Understanding cosmic variance can be important for correctly connecting high redshift observations of galaxies with our understanding of reionization e.g. Ouchi et al. (2009).

Note that in general it is possible for two populations to have the same average number counts, but different cosmic variances - this occurs when they have the same 1-point statistics, but different 2-point statistics. Thus we can use the 2-point statistics to refine the estimate of cosmic variance in Bowler et al. (2015) who used the Trenti Cosmic Variance calculator (Trenti & Stiavelli 2008) - which only uses 1-point statistics. A clustered and unclustered population of the same number density will have substantial and zero cosmic variance respectively (both will have Poisson variance). Also 2-point statistics only give the *variance* of the full probability distribution of counts in a field. Higher order statistics (n-point correlation function etc.) are needed to fully probe the full distribution. Similarly, 3-point statistics are needed to quantify the cosmic variance on measurements of 2-point statistics, 4-point statistics are needed to quantify the cosmic variance on measurements of 3-point statistics, and so on ad infinitum. Alternatively, more complex cosmic variance behaviours can be studied with mock catalogs from cosmological simulations of structure formation (Trenti & Stiavelli 2008).

The cosmic variance is related to the *expected value* of the correlation function in the geometry of the field, that is to say the expected value of the correlation function at the separation of two points randomly selected in the field. Analytically we can write the expectation as:

$$\bar{\omega}(A) = \frac{\int_A \int_A \omega(|\vec{\theta_i} - \vec{\theta j}|) d^2\theta_i d^2\theta_j}{\int_A \int_A d^2\theta_i d^2\theta_j}, \tag{11}$$

where A is the angular region of the field, ω is the 2-point correlation function, $\bar{\omega}(A)$ is the expectation of the correlation function in that field, θ_i and θ_j are points in the field, $|\vec{\theta_i} - \vec{\theta j}|$ is their angular separation, and the integrals are double integrals over the area of the field. We calculate this numerically by sampling 100,000 pairs of points in the field, calculating their angular separation, finding the value of the correlation function at that angular scale (with the best fit model from Section 4.2), and then taking the average.

Trenti & Stiavelli (2008) summarise results from Peebles (1993); Newman & Davis (2001) and Somerville et al. (2003) that conclude:

$$\bar{\omega}(A) = \frac{\langle N^2 \rangle - \langle N \rangle^2}{\langle N \rangle^2} - \frac{1}{\langle N \rangle} \tag{12}$$

where N is the random variable of the number of objects in a field. This can be rearranged to the form:

$$\langle N^2 \rangle - \langle N \rangle^2 = \langle N \rangle + \bar{\omega}(A) \langle N \rangle^2, \tag{13}$$

$$\sigma^2_{total} = \sigma^2_{poisson} + \sigma^2_{CV},$$

where $\sigma_{total} = \sqrt{\langle N^2 \rangle - \langle N \rangle^2}$ is the total standard deviation on measurements of number counts, $\sigma_{Poisson} = \sqrt{\langle N \rangle}$ is the Poisson standard deviation and $\sigma_{CV} = \sqrt{\bar{\omega}(A)} \langle N \rangle$ is the standard deviation from cosmic variance e.g. total standard deviation is the Poisson and cosmic variance standard deviations added in quadrature. The standard deviation from cosmic variance reduces to $\sigma_{CV} = b\sqrt{\bar{\omega}_{DM}(A)} \langle N \rangle$ in the 'pure-bias' case where $\omega = b^2 \omega_{DM}$, where b is the bias and ω_{DM} is the dark matter angular correlation function.

This formalism has all the properties one would expect from cosmic variance. Cosmic variance is higher when sources are more clustered (further away from uniform). Cosmic variance becomes lower as the size of the field increases, as the correlation function is sampling larger scales, where the function has a lower value, a consequence of the fact that a larger range of environments are being probed. A more subtle effect is that cosmic variance also varies with field shape, as well as size. The average length scale probed for a circle is a lot smaller than for a long thin rectangle of the same area (for example), corresponding to a higher average correlation function value, and greater cosmic variance. This can be interpreted (as described in Trenti & Stiavelli 2008) as a consequence of the fact that a more compact field geometry is predominantly sampling the same environment, be it an over- or under-density. However a long thin geometry is sampling from a large range of environments, and overdensities and underdensities are more likely to cancel out. The formalism for describing cosmic variance here also works when the field is disconnected. If the 'field' is actually two disconnected subfields separated by a vast distance in the sky, when calculating the average of the correlation function over this field, half the time the two points will be in different sub-fields, and the value of the correlation function on this scale will be effectively zero. This halves the value of $\bar{\omega}(A)$, effectively reducing cosmic variance contribution by a factor of $\sqrt{2}$ as completely different regions of the Universe are being probed.

Table 3 summarises our results when applied to the UltraVISTA and UDS fields for our $z \sim 6$ samples (using our best fit pure bias model). The most important columns to

Table 3. Actual and expected number of galaxies in each field. The columns are: field used, the field angular area (in deg^2), the actual number of galaxies in the field ($N_{\rm a}$), the actual angular galaxy density in the field ($\rho_{\rm a}$, in deg^{-2}), the expected number of galaxies in the field if it had the mean density ($N_{\rm e}$), the expected angular galaxy density in the field - all identical figures as we are considering deviations from the mean density ($\rho_{\rm e}$, in deg^{-2}), standard deviation from Poisson statistics, equal to the square root of $N_{\rm e}$ ($\sigma_{Poisson}$), standard deviation from cosmic variance, estimated from our clustering measurements (σ_{CV}), our Poisson and cosmic variance errors added in quadrature (σ_{Total})

Field	Area (deg^2)	$N_{\rm a}$	$\rho_{\rm a}$ (deg^{-2})	$N_{\rm e}$	$\rho_{\rm e}$ (deg^{-2})	$\bar{\omega}$	$\sigma_{Poisson}$	σ_{CV}	σ_{Total}
UDS	0.74	64	86	92	124	0.027	10	15	18
UltraVISTA	0.62	103	166	77	124	0.023	9	12	15
Total	1.35	167	124	167	124	0.013	13	19	23

compare are the N_a (the actual number of galaxies in the field), N_e (the expected number there would be if both fields had the average density) and σ_{total} the standard deviation on counts including Poisson and cosmic variance implied by our clustering measurements. In both cases, the observed number of galaxies in each field is approximately a 1.5σ deviation from the model value (as noted in Bowler et al. 2014, they are the most over- and under-dense respectively of the five CANDELS fields). Thus at these redshifts, both UltraVISTA and UDS appear to be moderate, but not unreasonable, over and under-densities respectively.

5.5 The onset of quenching

Bowler et al. (2014) and Bowler et al. (2015) report a rapid evolution in the high luminosity end of the luminosity function at $z = 6 - 7$, a transition from a power-law drop off to an exponential cut-off, interpreted as the onset of quenching or dust obscuration. We show in Fig. 8 the luminosity to halo mass ratios implied by the Bowler et al. (2015) luminosity functions and the simplest possible abundance matching scheme:

$$M_h(M_{UV}) = \psi^{-1}_{HMF}(\psi_{lum}(M_{UV})),$$

where M_h is the halo mass corresponding to the magnitude M_{UV}, $\psi_{HMF}(M_h)$ is the comoving number density of halos greater than mass M_h and $\psi_{lum}(M_h)$ is the comoving number density of LBGs brighter than magnitude M_{UV}. This is a very simple model (more complex abundance matching schemes do exists e.g. SHAM, Guo et al. 2015), and ignores complexities such as scatter, satellites and duty cycle, but illustrates the argument in Bowler et al. (2014) that $z = 6 - 7$ is the onset of quenching. At the low luminosity end, the luminosity to halo mass ratio drops with time, as Harikane et al. (2016) find using clustering (see their figure 10). At the high mass end, the ratio is fairly constant, before dropping off towards low redshift e.g. Behroozi et al. (2013). The unphysical rise in the $z = 7$ luminosity to halo mass ratio at the high luminosity end is a result of the fact that the luminosity function cannot (as noted in Bowler et al. 2014) continue as a power law to even brighter magnitudes, as it would quickly be dramatically higher than the

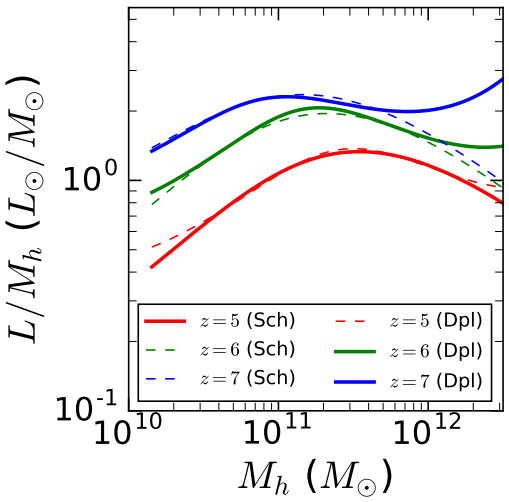

Figure 8. Luminosity (monochromatic luminosity at 1500$\AA$; that is to say, λf_λ where λ is the wavelength and f_λ is the flux per wavelength) to halo mass ratios as a function of halo mass, derived abundance matching Bowler et al., (2015) luminosity functions and Behroozi et al., (2013) halo mass functions. The results are shown for both Schechter function fit and the double power law fit, for $z = 5$, $z = 6$ and $z = 7$. In bold are the best fitting functions, double power law for $z = 6$ and $z = 7$, Schechter for $z = 5$, and the dashed line shows the alternative.

HMF. More realistically, in a scenario with no high luminosity/mass end quenching, the luminosity function would drop off at the same rate as the HMF (which doesn't drop off as fast as a Schechter function). If at higher luminosities it really does continue as a power law, then the LBGs would become much more numerous than their corresponding halos, and almost certainly these bright objects would be rare phases in the duty cycle of a more common population. Davidzon et al. (2017) make a similar argument for the onset of quenching, using stellar mass estimates from UltraVISTA-DR2, SPLASH and Hyper-Suprime-Cam data in the COSMOS field, except finding the transition at $z \sim 3$ rather than $z \sim 6$.

The fact that these galaxies at the bright end of the luminosity function truly are in the densest regions of the Universe as opposed to less biased objects caught in extremely rare massive star-bursts, supports the explanation that the drop off in the luminosity to halo mass ratio at the high-mass end is essentially still operational at $z = 6$. If the sample had been just rare episodes of vigorous star formation, then the interpretation of the steepness/drop off rate at the bright end of the luminosity function would be different - it would instead be dominated by the distribution of star-formation rates a given population has, and how rare its episodes of high star formation are, as opposed to the luminosity function being dominated by a modified halo mass function.

To determine if $z \sim 6-7$ is the onset of mass quenching, a similar analysis to this work of luminous $z = 7$ LBGs would need to be performed to see if the luminosity to halo mass ratio doesn't drop off. There are three main possibilities at $z \sim 7$:

- $M_{UV} \sim -22$ LBGs have the same host halo mass as $M_{UV} \sim -20$ objects; this would suggest they are the same population of objects at different points in their duty cycle
- $M_{UV} \sim -22$ LBGs have the same host halo mass as they do at $z \sim 6$. This would suggest that the luminosity to halo mass ratio at the high luminosity end doesn't change much over $z \sim 6-7$, which would not support $z \sim 6-7$ being the onset of mass quenching/dust obscuration
- $M_{UV} \sim -22$ LBGs have a lower host halo mass as they do at $z \sim 6$, but still higher than the galaxies with $M_{UV} \sim -20$, in such a way that the luminosity to halo mass ratio was constant as a function of halo mass. This would be supportive that $z \sim 6-7$ was indeed the onset of quenching (or dust obscuration).

6 CONCLUSIONS

We have used the largest existing sample of extremely bright Lyman-break galaxies at $z \sim 6$ to investigate their large scale structure and links to the possible onset of feedback quenching or dust obscuration at this redshift. This sample (detailed in Bowler et al. 2015) of 263 LBGs was selected in the UltraVISTA/COSMOS and UDS/SXDS fields, using deep optical and near-infrared data required to distinguish the galaxies from contaminant populations. The method we used to study the connection between the galaxies and their host halo was to measure their clustering with the angular correlation function, and model these measurements with a HOD scheme.

The key conclusions of this work are:

- Bright LBGs ($M_{UV} \leqslant -21$) appear to be highly biased ($b \sim 10$) objects in dense environments, as opposed to being rare temporal episodic incarnations of fainter galaxies ($M_{UV} \sim -19$). This suggests that the bright-end of the luminosity function at $z \sim 6$ is determined by feedback processes or dust obscuration, rather than duty cycles. Our results have important implications for the physical origin of the observed steepening of the bright end of the ultra-violet luminosity function between $z \sim 6$ and $z \sim 7$ (Bowler et al. 2014, 2015) - which in a straightforward abundance matching scheme would imply a dramatically increased luminosity to halo mass ratio at $z \sim 7$ to $z \sim 6$.
- We find a tension between the observed number counts and bias, that suggests that some modification to our knowledge of the high-redshift dark matter distribution is needed This is most likely to be the incorporation of quasi-linear effects (as described in Jose et al. 2017), or possibly a minor calibration upwards of halo bias at high redshift.
- Although number counts within each field differ by far more than Poisson sample variance, estimates of the cosmic variance from the clustering would suggest that both fields are only moderate 1.5-σ over/under densities.
- We do not require duty cycle to explain our observations (equivalently $DC \sim 1$), and the satellite fraction of the sources is very small, at most a few percent

In the next few years, deep, wide surveys such as VIDEO and the VISTA Extragalactic Infrared Legacy Survey (VEILS, Hönig et al. 2016), which will extend the area of VIDEO, will provide improved constraints on the luminosity function and clustering of high-redshift galaxies, and allowing extension to even more luminous LBGs. By the mid 2020s it should be possible to use the *EUCLID* space telescope mission to do this with 1,000s of LBGs (Bowler et al. 2017), which will reveal how the measured large scale structure of LBGs and LAEs relates to reionization.

ACKNOWLEDGEMENTS

The first author wishes to acknowledge support provided through an STFC studentship, and the Rector and Fellows of Lincoln College for support through the Graduate Research Fund. Many thanks to Yuichi Harikane for useful discussions on high redshift clustering, and to Steven Murray for advice for using HALOMOD. This work was supported by the Oxford Centre for Astrophysical Surveys which is funded through generous support from the Hintze Family Charitable Foundation, the award of the STFC consolidated grant (ST/N000919/1), and the John Fell Oxford University Press (OUP) Research Fund. Based on data products from observations made with ESO Telescopes at the La Silla or Paranal Observatories under ESO programme ID 179.A- 2006. Based on observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/IRFU, at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council (NRC) of Canada, the Institut National des Science de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. This work is based in part on data products produced at Terapix available at the Canadian Astronomy Data Centre as part of the Canada-France-Hawaii Telescope Legacy Survey, a collaborative project of NRC and CNRS.

References

Adam R. et al., 2016, Astronomy & Astrophysics, 596, A108

Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Giallongo E., 1999, Monthly Notices of the Royal Astronomical Society, 310, 540

Bahcall N. A., Soneira R. M., 1983, The Astrophysical Journal, 270, 20

Barone-Nugent R. L. et al., 2014, The Astrophysical Journal, 793

Becker G. D., Bolton J. S., Lidz A., 2015, Publications of the Astronomical Society of Australia, 32, e045

Behroozi P., Silk J., 2016, MNRAS, 000, 0

Behroozi P. S., Wechsler R. H., Conroy C., 2013, The Astrophysical Journal, 770, 57

Bertin E., Arnouts S., 1996, SExtractor: Software for source extraction

Beutler F. et al., 2011, Monthly Notices of the Royal Astronomical Society, 416, 3017

Bouwens R. J., Illingworth G. D., Franx M., Ford H., 2007, The Astrophysical Journal, Volume 670, Issue 2, pp. 928-958., 670, 928

Bouwens R. J. et al., 2015, The Astrophysical Journal, 803, 34

Bouwens R. J. et al., 2016, The Astrophysical Journal, 830, 67

Bowler R. A. A. et al., 2015, Monthly Notices of the Royal Astronomical Society, 452, 1817

Bowler R. A. A., Dunlop J. S., McLure R. J., McLeod D. J., 2017, Monthly Notices of the Royal Astronomical Society, 466, 3612

Bowler R. A. A. et al., 2014, Monthly Notices of the Royal Astronomical Society, 440, 2810

Cen R., Safarzadeh M., 2015, The Astrophysical Journal Letters, 798, 38

Clay S. J., Thomas P. A., Wilkins S. M., Henriques B. M. B., 2015, Monthly Notices of the Royal Astronomical Society, 451, 2692

Conroy C., Wechsler R. H., Kravtsov A. V., 2006, The Astrophysical Journal, 647, 201

Cooray A., Sheth R., 2002, Physics Reports, 372, 1

Coupon J. et al., 2015, Monthly Notices of the Royal Astronomical Society, 449, 1352

Coupon J. et al., 2012, Astronomy & Astrophysics, 542, A5

Cress C. M., Helfand D. J., Becker R. H., Gregg M. D., White R. L., 1996, The Astrophysical Journal, 473, 7

Davidzon I. et al., 2017, ArXiv eprints

Davis M., Peebles P. J. E., 1983, The Astrophysical Journal, 267, 465

Doré O., Holder G., Alvarez M., Iliev I. T., Mellema G., Pen U.-L., Shapiro P. R., 2007, Physical Review D, 76, 043002

Dunlop J. S. et al., 2013, Monthly Notices of the Royal Astronomical Society, 432, 3520

Finkelstein S. L. et al., 2015, The Astrophysical Journal, 810, 71

Fisher J., Davis K. B., Strauss M., Yahil M. A., Huchra A., 1994, Monthly Notices of the Royal Astronomical Society, 266

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, Publications of the Astronomical Society of the Pacific, 125, 306

Furusawa H. et al., 2008, The Astrophysical Journal Supplement Series, 176, 1

Giavalisco M., Mauro, 2002, Annual Review of Astronomy and Astrophysics, 40, 579

Gott J. R., Choi Y.-Y., Park C., Kim J., Gott Iii J. R., Choi Y.-Y., Park C., Kim J., 2009, The Astrophysical Journal, 695, L45

Grissom R. L., Ballantyne D. R., Wise J. H., 2014, Astronomy & Astrophysics, 561, A90

Guhathakurta P., Tyson J. A., Majewski S. R., 1990, The Astrophysical Journal, 357, L9

Guo H. et al., 2015, arxiv, 18, 1

Harikane Y. et al., 2016, The Astrophysical Journal, 821, 123

Hartley W. G., Almaini O., Foucaud S., 2013, in Thirty Years of Astronomical Discovery with UKIRT, Astrophysics and Space Science Proceedings, Volume 37. ISBN 978-94-007-7431-5. Springer Science+Business Media Dordrecht, 2013, p. 309, Vol. 37, pp. 309–321

Hatfield P. W., Jarvis M. J., 2016, arxiv

Hatfield P. W., Lindsay S. N., Jarvis M. J., Häußler B., Vaccari M., Verma A., 2016, Monthly Notices of the Royal Astronomical Society, 459, 2618

Hearin A. P., Behroozi P. S., van den Bosch F. C., 2015, arXiv:1504.05578, 000, 13

Hönig S. F. et al., 2016, Monthly Notices of the Royal Astronomical Society, Volume 464, Issue 2, p.1693-1703, 464, 1693

Ilbert O. et al., 2006, Astronomy and Astrophysics, 457, 19

Jarvis M. J. et al., 2013, Monthly Notices of the Royal Astronomical Society, 428, 1281

Johnston R., 2011, The Astronomy and Astrophysics Review, Volume 19, article id.41, 19

Jose C., Baugh C. M., Lacey C. G., Subramanian K., 2017, Mon. Not. R. Astron. Soc, 000, 0

Jose C., Lacey C. G., Baugh C. M., 2016, Mon. Not. R. Astron. Soc, 000, 0

Jose C., Srianand R., Subramanian K., 2013, Monthly Notices of the Royal Astronomical Society, Volume 435, Issue 1, p.368-377, 435, 368

Jullo E., Kneib J.-P., Limousin M., Elíasdóttir Á., Marshall P., Verdugo T., 2007, New Journal of Physics, Volume 9, Issue 12, pp. 447 (2007)., 9, 447

Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlober S., Allgood B., Primack J. R., 2004, The Astrophysical Journal, 609, 35

Lacey C. G. et al., 2016, Monthly Notices of the Royal Astronomical Society, 462, 3854

Laigle C. et al., 2016, The Astrophysical Journal Supplement Series, 224, 24

Landy S. D., Szalay A. S., 1993, The Astrophysical Journal, 412, 64

Lawrence A. et al., 2007, Monthly Notices of the Royal

Astronomical Society, 379, 1599
Limber D. N., 1954, The Astrophysical Journal, 119, 655
Lindsay S. N., Jarvis M. J., McAlpine K., 2014, Monthly Notices of the Royal Astronomical Society, 440, 2322
Ling E. N., Barrow J. D., Frenk C. S., 1986, Monthly Notices of the Royal Astronomical Society, 223, 21P
Madau P., Dickinson M., 2014, Annual Review of Astronomy and Astrophysics, 52, 415
Mandelbaum R. et al., 2013, The Astrophysical Journal Supplement, Volume 212, Issue 1, article id. 5, 28 pp. (2014)., 212
Mann B., Peacock J., Heavens A., 1997, Monthly Notices of the Royal Astronomical Society, vol. 293, p. 209, 293, 209
McCracken H. J. et al., 2012, Astronomy & Astrophysics, 544, A156
McCracken H. J. et al., 2015, Monthly Notices of the Royal Astronomical Society, 449, 901
McLeod D. J., McLure R. J., Dunlop J. S., 2016, Monthly Notices of the Royal Astronomical Society, Volume 459, Issue 4, p.3812-3824, 459, 3812
McLure R. J., Cirasuolo M., Dunlop J. S., Foucaud S., Almaini O., 2009, Monthly Notices of the Royal Astronomical Society, 395, 2196
McLure R. J. et al., 2013, Monthly Notices of the Royal Astronomical Society, 432, 2696
McQuinn M., Hernquist L., Zaldarriaga M., Dutta S., 2007, Monthly Notices of the Royal Astronomical Society, 381, 75
Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, The Astrophysical Journal, 710, 903
Murray S. G., Power C., Robotham A. S. G., 2013, Monthly Notices of the Royal Astronomical Society: Letters, 434, L61
Natarajan A., Yoshida N., 2014, Progress of Theoretical and Experimental Physics, 2014, 6B112
Navarro J. F., Frenk C. S., White S. D. M., 1996, The Astrophysical Journal, 462, 563
Newman J. A., Davis M., 2001, The Astrophysical Journal, Volume 564, Issue 2, pp. 567-575., 564, 567
Oesch P. A. et al., 2016, The Astrophysical Journal, Volume 819, Issue 2, article id. 129, 11 pp. (2016)., 819
Oke J. B., Gunn J. E., 1983, The Astrophysical Journal, 266, 713
Ouchi M. et al., 2009, The Astrophysical Journal, 706, 1136
Ouchi M. et al., 2010, The Astrophysical Journal, 723, 869
Peebles P. J. E., 1993, Physics today, vol. 46, no. 11, p. 87 (1993), 46, 87
Pentericci L. et al., 2014, The Astrophysical Journal, Volume 793, Issue 2, article id. 113, 10 pp. (2014)., 793
Robertson B. E., Ellis R. S., Furlanetto S. R., Dunlop J. S., 2015, The Astrophysical Journal Letters, Volume 802, Issue 2, article id. L19, 5 pp. (2015)., 802
Roche N., Eales S. A., 1999, Monthly Notices of the Royal Astronomical Society, 307, 703
Schechter P., 1976, The Astrophysical Journal, 203, 297
Scoville N. et al., 2007, The Astrophysical Journal Supplement Series, 172, 38
Shapley A. E., 2011, Annu. Rev. Astron. Astrophys, 49, 525
Silk J., Joseph, 2011, Tracing the Ancestry of Galaxies (on the land of our ancestors), Proceedings of the International Astronomical Union, IAU Symposium, Volume 277, p. 273-281, 277, 273
Somerville R. S., Lee K., Ferguson H. C., Gardner J. P., Moustakas L. A., Giavalisco M., 2003, The Astrophysical Journal, Volume 600, Issue 2, pp. L171-L174., 600, L171
Stark D. P., 2016, Annual Review of Astronomy and Astrophysics, 54, 761
Stark D. P., Ellis R. S., Bunker A., Bundy K., Targett T., Benson A., Lacy M., 2009, The Astrophysical Journal, Volume 697, Issue 2, pp. 1493-1511 (2009)., 697, 1493
Steidel C., Giavalisco M., Pettini M., Dickinson M., Adelberger K., 1996, Astrophysical Journal Letters v.462, p.L17, 462, L17
Steidel C. C., Hamilton D., 1992, The Astronomical Journal, 104, 941
Steinhardt C. L., Capak P., Masters D., Speagle J. S., 2016, The Astrophysical Journal, 824
Tegmark M., Max, 2003, American Physical Society, April Meeting, 2003, April 5-8, 2003 Philadelphia, Pennsylvania, MEETING ID: APR03, abstract #T1.003
Tinker J. L., Kravtsov A. V., Klypin A., Abazajian K., Warren M. S., Yepes G., Gottlober S., Holz D. E., 2008, The Astrophysical Journal, Volume 688, Issue 2, article id. 709-728, pp. (2008)., 688
Tinker J. L. J., Robertson B. E. B., Kravtsov a. A. V., Klypin A., Warren M. S. M., Yepes G., Gottlober S., Gottlöber S., 2010, The Astrophysical Journal, 724, 11
Trenti M., Stiavelli M., 2008, The Astrophysical Journal, 676, 767
Vale A., Ostriker J. P., 2004, Monthly Notices of the Royal Astronomical Society, 353, 189
Van den bosch F. C., More S., Cacciato M., Mo H., Yang X., 2013, Monthly Notices of the Royal Astronomical Society, 430, 725
White S. D. M., 1979, Monthly Notices of the Royal Astronomical Society, 186, 145
Zaroubi S., 2013, in The First Galaxies, Astrophysics and Space Science Library, Volume 396. ISBN 978-3-642-32361-4. Springer-Verlag Berlin Heidelberg, 2013, p. 45, Vol. 396, pp. 45–101
Zehavi I. et al., 2005, The Astrophysical Journal, 621, 22
Zheng Z. et al., 2005, The Astrophysical Journal, 633, 791

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