

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} and C. L. Hale¹

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*

Accepted 2018 March 23. Received 2018 February 3; in original form 2017 February 10

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 ($b \sim 8$) 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 dust obscuration in the early Universe – 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. There is a slight tension between the number densities and clustering measurements, which 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 ~ 1.8 times greater than UDS) is shown to be consistent with the cosmic variance implied by the clustering measurements.

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

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. Bouwens et al. 2016; McLeod, McLure & Dunlop 2016; Oesch 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 intergalactic medium (IGM) is essentially opaque to photons with wavelengths shorter than the ‘Lyman break’ (1216 Å, 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 1990s (Guhathakurta, Tyson & Majewski 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 number of sources with the *Hubble Space Telescope* (HST) in the late 1990s and 2000s (e.g. Giavalisco 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, Almaini & Foucaud 2013), and more recently 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 led to advances in the

* E-mail: peter.hatfield@physics.ox.ac.uk

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 (Clay et al. 2015; Lacey et al. 2016); and can be readily linked theoretically to the dark matter equivalent, the halo mass function [HMF; the comoving density of dark matter haloes as a function of halo mass, see Murray, Power & Robotham (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 parametrization, 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 normalization. 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; Bowler et al. 2015; Finkelstein 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) and McLeod et al. (2016).

1.2 Clustering

Galaxies are formed and live in dark matter haloes, 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 haloes because such haloes are rarer (Vale & Ostriker 2004). Abundance matching however can only ever give an incomplete account of the connection due to three complications. First, haloes 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-haloes and isolated haloes. 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 per cent of the time, a straightforward abundance matching will erroneously place these sources with a different appearance in more massive haloes 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. Paczynski 1987; Soucaill et al. 1988; Jullo et al. 2007) or weak (galaxy–galaxy lensing, e.g. Brainerd, Blandford & Smail 1996; Bartelmann & Schneider 2001; Mandelbaum et al. 2014; Coupon et al. 2015), 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 (two-point statistics) of the galaxies alongside the number counts (one-point statistics). This can then be linked to models/our theo-

retical understanding of structure formation to estimate the typical environment of the galaxies. One popular framework for modelling galaxy clustering is the ‘halo occupation distribution’ (HOD) model (e.g. Benson et al. 2000; Ma & Fry 2000; Seljak 2000; Berlind & Weinberg 2002; Zehavi et al. 2005), which models the non-linear clustering of galaxies within individual haloes, and the large-scale clustering of the haloes 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. 2016) and $z = 0.5-2$ (Coupon et al. 2015; McCracken 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) (where it has also been possible to compare to clustering predictions from hydrodynamical cosmological predictions, e.g. the Bluetides simulation, see Bhowmick et al. 2017). It is crucial however to understand the relationship at the massive/most luminous end, as this is where active galactic nucleus driven feedback may have a role (Bower et al. 2006; Croton et al. 2006; Silk 2010). 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.

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 for our understanding of *reionization* (see Zaroubi 2013; Natarajan & Yoshida 2014 for an observational and theoretical review, respectively). The current best constraint on the average reionization redshift from the Planck mission is $z = 7.8-8.8$ (Adam et al. 2016), with many probes of the epoch (e.g. Pentericci et al. 2014; Becker, Bolton & Lidz 2015) suggesting that some parts of the Universe could still be undergoing reionization by $z \sim 6$. McQuinn et al. (2007) suggest that differences in the clustering of LBGs and Lyman α emitting galaxies (LAEs) could give an insight into the possible ‘patchy’ nature of reionization. The Lyman α line is suppressed if the source is in a largely neutral region which biases the observations of LAEs towards large ionized H II 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 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.

1.3 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 utilizing 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 $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ 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 μm) across two main fields (see Fig. 1) was used to select the sample; we summarize 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 six 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 data sets 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 summarize 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 $-22.7 < M_{\text{UV}} < -20.5$ with $5.5 < z < 6.5$ over 1.35 deg^2 . We take our final sample as all 161 sources with $M_{\text{UV}} < -21.125$, as the sample completeness drops off rapidly faintwards of this value, as discussed in Bowler et al. (2015), see their fig. 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 characterize 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 – we calculate a model spatial correlation function, and project to an angular correlation function.

3.1 The angular correlation function

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

$$dP = \sigma(1 + \omega(\theta)) d\Omega, \quad (1)$$

where dP is the probability of finding two galaxies at an angular separation θ , σ is the surface number density of galaxies, and $d\Omega$ is the solid angle. This is most commonly estimated by calculating $DD(\theta)$, the normalized 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

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

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.

Uncertainty on measurements of the correlation function can be broken down into (i) ‘Poisson-like’ uncertainty, which depends on the number density of the population tracing the underlying density field, and (ii) ‘cosmic variance’ uncertainty, which depends on how representative the volume under consideration is of structures on the scale of interest (see also Section 5.4). Estimates of uncertainty can be either ‘internal’ (based solely on the data itself), ‘external’ (based on comparison with mock catalogues from simulations or similar), or analytic (based on analytic equations from the theory of clustering statistics, typically using third or fourth order statistics).

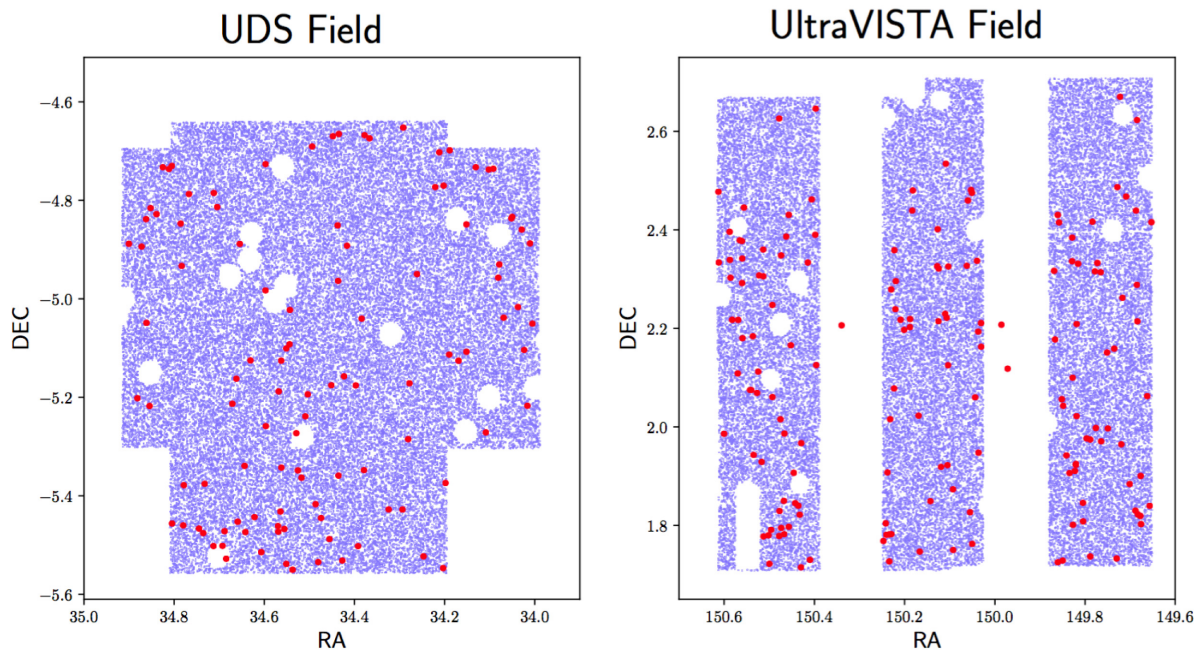


Figure 1. The geometry of the UDS and UltraVISTA fields. The red points are the galaxy locations from Bowler et al. (2015). The blue points are the random points chosen to cover the fields used for the construction of RR for the calculation of the correlation function. The three galaxies in the UltraVISTA 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 figs 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.

In this work uncertainties are calculated with the internal ‘bootstrap-object resampling’ method, which samples the individual galaxies with replacement from the data set, from which we recalculate the correlation function (see Ling, Barrow & Frenk 1986). Repetition of this process produces multiple ‘realizations’ of the correlation function, from which the covariance matrix of the $\omega(\theta)$ values can be estimated. In terms of ‘Poisson-like’ uncertainty, it is possible to calculate the error bars from Poisson uncertainty on the DD values, but Cress et al. (1996) and Lindsay, Jarvis & McAlpine (2014) found errors calculated in this manner were a factor of 1.5–2 smaller than those estimated with bootstrap-object resampling. It is particularly important 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.

The other important internal approaches to estimating the uncertainty on measurements of the correlation function involve dividing the field into N sub-regions. ‘Bootstrap-volume’ resampling (e.g. Maddox et al. 1990) then samples M of these sub-regions (with replacement) repeatedly, calculating the correlation function each time, similarly to bootstrap-object resampling. The ‘jack-knife’ method (Shao 1986) measures how the estimate of the correlation function varies when individual sub-regions are removed. Bootstrap-object estimates are known to underestimate the true cosmic variance uncertainty, e.g. Fisher et al. (1994). Mo, Jing & Boerner (1992) found that analytic and external mock estimates of uncertainty agreed well on all scales, but that bootstrap-object estimates only agreed with analytic estimates on small scales, underestimating the true uncertainty by ~ 35 per cent on scales comparable to the size of the field. Volume resampling methods conversely can

slightly overestimate the true uncertainty; Norberg et al. (2009) discuss the advantages and disadvantages of bootstrap-volume and jackknife, finding that both overestimate the size of error bars by about ~ 40 per cent on scales relevant to this work when compared to external estimates based on mocks (which are beyond the scope of this work, but are typically considered to be closer to the ‘true’ uncertainty). In terms of sub-dividing a survey into smaller volumes, Cabre et al. (2007) note that irregular shapes can jeopardize the estimation of correlation function uncertainty estimates for internal methods, and any subdivision of our survey geometry would certainly have very irregular shapes. In summary, bootstrap-object uncertainty estimates capture ‘Poisson-like’ uncertainty well, but imperfectly captures ‘cosmic variance’ uncertainty compared to bootstrap-volume methods. However given the difficulties of doing volume resampling with an irregular field geometry, and, perhaps most importantly, that our correlation function is calculated treating two separate highly independent fields as one, mitigating ‘cosmic variance’ uncertainty, we conclude that bootstrap-object uncertainty estimates could slightly underestimate the error bars on our measurements of the correlation function on large scales, but are likely sufficiently accurate for our purposes.

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 parametrized 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 ‘one-halo’ term, describing the small-scale clustering of galaxies within an individual halo, and a ‘two-halo’ term, describing the clustering of the haloes themselves. The ‘one-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, Frenk & White 1996) profile. The two-halo term is constructed by scaling the dark matter linear correlation function by the weighted-average halo bias of the host haloes.

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

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

the total number of central galaxies as

$$\langle N_{\text{cen}}(M_h) \rangle = \frac{1}{2} \left(1 + \text{erf} \left(\frac{\log_{10} M_h - \log_{10} M_{\min}}{\sigma_{\log_{10} M}} \right) \right), \quad (4)$$

and the total number of satellites as

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

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. 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, Wechsler & Kravtsov 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 two-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. Coupon et al. 2012; Hatfield et al. 2016).

We investigate the consequences of allowing various parameters to be fixed or free in the fitting process. As per equations 54 and 55 in 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):

$$\log(M_1/M_\odot) = 1.18 \log(M_{\min}/M_\odot) - 1.28, \quad (6)$$

$$\log(M_0/M_\odot) = 0.76 \log(M_1/M_\odot) + 2.3. \quad (7)$$

In this work we use the halo mass function of Behroozi, Wechsler & Conroy (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 DC, 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 DC of 10 per cent (i.e. it is only in its bright phase 10 per cent 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, haloes.

Without incorporating the DC, 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 DC to give the model comoving density:

$$n_{\text{gal}} = \text{DC} \times \int_0^\infty \text{HMF}(M_h) \times \langle N_{\text{tot}}(M_h) \rangle dM_h, \quad (8)$$

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 preparation) to calculate the spatial correlation function. We then project this to an angular correlation function (as per Limber 1954, using a redshift distribution derived

¹<https://github.com/steven-murray/halomod>

by smoothing the point estimates of the LBG redshifts), and subtract off the numerical approximation of the integral constraint to get our final model correlation function.

We use EMCEE² (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

$$\chi^2 = \frac{[\log n_{\text{gal}}^{\text{obs}} - \log n_{\text{gal}}^{\text{model}}]^2}{\sigma_{\log n}^2} + \sum_{i,j} [\omega^{\text{obs}}(\theta_i) - \omega^{\text{model}}(\theta_i)] [C_{i,j}^{-1}] \times [\omega^{\text{obs}}(\theta_j) - \omega^{\text{model}}(\theta_j)], \quad (9)$$

where $n_{\text{gal}}^{\text{obs}}$ is the observed galaxy number density, $n_{\text{gal}}^{\text{model}}$ 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 (calculated as per Trenti & Stiavelli 2008), 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.

In addition to the uncertainty on the number counts and the covariance of the clustering measurements, there is in general ‘cosmic covariance’ between observed number counts and measured clustering amplitude. For an overdense region of the Universe, typically the number counts and measured clustering amplitude will both be higher than the global value, and vice versa for an underdense region (e.g. see Lacasa & Rosenfeld 2016). This covariance would typically lead to an underestimate of the true uncertainties on our inferred HOD parameters. However, the fact that our main analysis is performed jointly over two highly independent fields (one of which is known to be overdense and the other underdense) mitigates the impact of this effect; we conclude not accounting for this covariance is unlikely to affect our key results.

When the parameters are free, we use a uniform prior over $10 < \log_{10}(M_{\text{min}}/M_{\odot}) < 13$, $\log_{10}(M_{\text{min}}/M_{\odot}) < \log_{10}(M_1/M_{\odot}) < 14$ (uniform in log space), and $0 < \text{DC} < 1$. We used 20 walkers with 1000 steps, which have starting positions drawn uniformly from the prior. When we fit the bias b directly (Model E. in Section 4.2) we use a uniform prior over $1 < b < 30$.

We use 500 000 random data points in this study. As per Hatfield et al. (2016) and Hatfield & Jarvis (2016), we use 100 bootstrap-object resamplings to estimate the uncertainty at the 16th and 84th percentiles of the resampling (corresponding to plus/minus one standard deviation). For $n_{\text{gal}}^{\text{obs}}$, we use the value obtained when integrating the luminosity function brightwards to infinity from $M_{\text{UV}} = -21.125$ (as 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 $n_{\text{gal}}^{\text{obs}} = 4.1 \times 10^{-5} \text{ Mpc}^{-3}$, for the fields combined. This is from the best-fitting double power-law model in Bowler et al. (2015). Using the best-fitting Schechter function gives the marginally lower value of $n_{\text{gal}}^{\text{obs}} = 3.8 \times 10^{-5} \text{ Mpc}^{-3}$. 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).

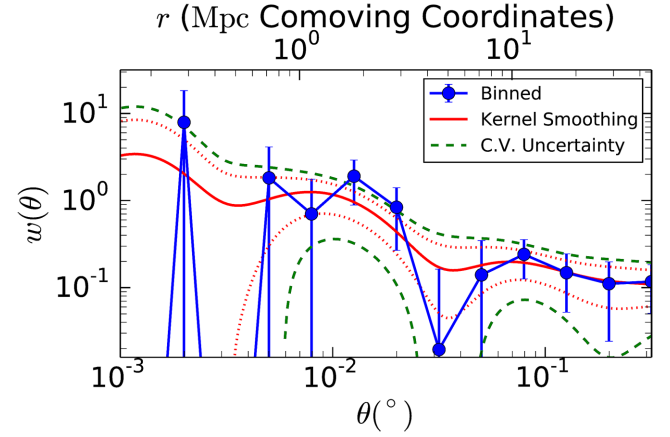


Figure 2. The angular correlation function for our sample of bright ($M_{\text{UV}} < -21.125$) $z \sim 6$ LBGs from Bowler et al. (2015). The figure shows the correlation function estimated both with a binning method (blue points), and a kernel smoothing method (red curve, with dotted lines showing the error on the measurements). The error bars on the measurements from both the binning method and the kernel smoothing method are calculated from the 16th and 84th percentiles of the set of bootstrapped measurements. The dashed green curves show the ‘cosmic variance corrected’ uncertainty on the kernel smoothed clustering measurements, used and discussed in Section 5.4.1.

4 RESULTS

4.1 Clustering measurements

Fig. 2 shows the angular correlation function of the full sample over the range $10^{-3} < \theta/\text{deg} < 10^{-0.5}$, 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 10 angular scales calculated for the bins, as our final measurements.³ Fig. 2 also shows the estimate of the cosmic variance on the clustering measurements used in Section 5.4.1.

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

³With 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.

²<http://dan.iel.fm/emcee/current/>

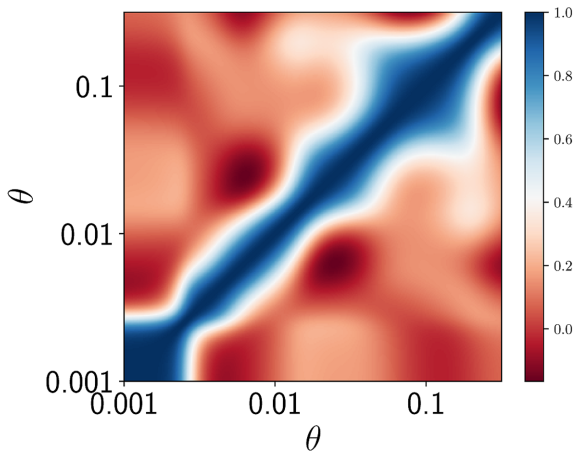


Figure 3. The correlation coefficients (covariance normalized) of our measurements. Blue values are positive correlations, and red values are negative values. The blue diagonal corresponds to the standard deviation measurements (a random variable is always perfectly correlated with itself).

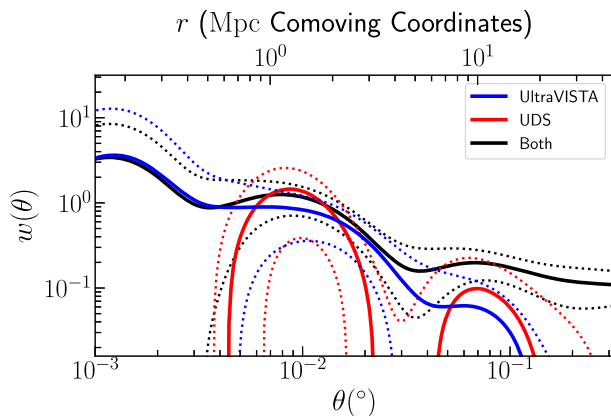


Figure 4. The angular correlation function of our LBG sample, calculated for the UltraVISTA field (blue), the UDS field (red), and the fields collectively (black).

the fitting process. We show the correlation matrix (the covariance matrix with each value normalized 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.

We also calculate the correlation function separately for the UDS and UltraVISTA fields individually, shown in Fig. 4. The single field correlation functions were calculated based solely on each individual field – random data points and RR are re-calculated for each sub-field, DD and DR normalization is based just on the galaxy number counts for the separate fields, etc. Small number statistics make the correlation function for each field extremely noisy, particularly for the UDS field where we only have 64 LBGs with $M_{UV} < -21.125$ in the sample. The error bars are smaller for the measured correlation function when both fields are used, than for the individual fields, as expected. It is clear that all small-scale pairs are from the UltraVISTA field – the UDS correlation function has no power on scales smaller than $\sim 10^{-2.5}$ deg.⁴ On scales of $\sim 10^{-2}$ – 10^{-1} deg the

amplitudes of the UDS and UltraVISTA correlation functions are roughly consistent with each other, and with the correlation function calculated for the fields together.⁵ Finally on scales $> 10^{-1}$ deg, the correlation function is moderately larger for the fields considered together than for the fields considered individually. This is because with just one field, essentially there is only access to \sim one large mode of fluctuation, so variance in the large-scale modes cannot be observed. However with more than one field these large-scale variations can now be measured. Mathematically this behaviour can be seen as a result of the use of the Landy & Szalay (1993) estimator – DR is moderately large on the large scales ($\gtrsim 0.1$ deg) for each individual field, but is smaller when the fields are considered collectively, as more of the randomly placed ‘dummy’ galaxy points are in the UDS field but more of the real galaxy points are in UltraVISTA. The estimate of the correlation function on larger scales when using the earlier Peebles & Hauser (1974) estimator $\omega = \frac{DD}{RR} - 1$ is much lower and essentially is the weighted average of the measurements from each individual field (suggesting that the value of DR is critical). Equivalently, the integral constraint (which starts to become significant on scales comparable to the maximum length scale accessed by a survey) is smaller when the fields are considered together, as there are large numbers of dummy–dummy pairs with one member of the pair in one field, and one in the other field, which contribute to the normalizing denominator, but not the numerator, of the expression for the integral constraint in Roche & Eales (1999) that we use in this work.

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 as follows:

- A. M_{\min} , M_1 , and DC free (e.g. one-halo and two-halo amplitudes and number counts all free)
- B. M_1 fixed as function of M_{\min} as per Section 3.2 (e.g. one-halo amplitude as fixed function of two-halo amplitude)
- C. M_1 fixed as function of M_{\min} , DC = 0.6 (e.g. one-halo amplitude as fixed function of two-halo amplitude, DC fixed at the Harikane et al. 2016 value)
- D. M_1 fixed as function of M_{\min} , DC = 1 (e.g. one-halo amplitude as fixed function of two-halo amplitude, no DC – galaxies ‘on’ at all times)

In addition to these four models we also compute the following:

- E. Galaxy bias from a pure bias model, e.g. fit for b where $\xi_{\text{model}} = b^2 \xi_{\text{DM}}$

- F. Halo masses for the most straightforward 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 haloes greater than that mass is equal to the comoving number density of the galaxy sample.

The results from these six models are shown in Table 1. Fig. 5 shows the data and the best-fitting models. We show the posterior from one fitting in Fig. 6 for illustrative purposes. It is clear that the

outputs $\omega = -1$ for every bootstrapped catalogue (as zero pairs cannot be bootstrapped into a non-zero number of pairs), as opposed to more realistically giving no constraints.

⁵The UDS field correlation function has a ‘dip’ relative to the UltraVISTA and ‘Both’ correlation functions at about 0.02 – 0.04° . We are not aware of any systematic (instrumental or in observing strategy) that inhibits finding galaxy pairs at this angular separation, so believe this is likely a $\sim 1.5\sigma$ chance fluctuation, as opposed to a significant discrepancy.

⁴This illustrates a weakness of the random catalogue approach to estimating the correlation function – when zero pairs are observed, this approach

Table 1. Our constraints on the HOD parameters from the MCMC fitting. Also shown are the corresponding satellite fractions (f_{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 10). 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 DC 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. 6. The lower luminosity parameter values are taken directly from Harikane et al. (2016), apart from satellite fraction, which we calculate.

Model	M_{UV}	$\log M_{\text{min}}$	$\log M_1$	$\log M_0$	α	σ	DC	$10^2 f_{\text{sat}}$	b	$\chi^2/\text{d.o.f.}$
A. 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
B. 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
C. 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
D. 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
E. Bias only	-21.125	NA	NA	NA	NA	NA	NA	NA	$11.4^{+6.5}_{-6.7}$	1.0
F. Matching only	-21.125	11.51	NA	NA	NA	(0)	(1)	(0)	NA	NA
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

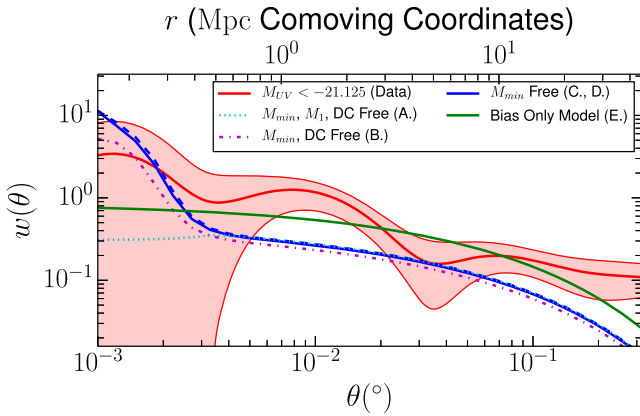


Figure 5. Comparison of our measurements (red curve and shaded area) with the five different clustering models we fit. The two blue curves correspond to models with only M_{min} free, the full line has DC = 0.6 (C.) and the dashed curve DC = 1 (D.). On linear scales all models are very similar, apart from the bias only model (E.), as the number density constraint is restricting the model from going too high. Only the M_1 free model (A.) allows the small-scale amplitude to vary independently.

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 $M_{\text{min}} \sim 10^{11.5} M_{\odot}$, e.g. our galaxies are hosted by haloes of that mass and above. It also seems that the satellite fraction is at most a few per cent, 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 haloes hosting multiple galaxies). Although models with a range of DCs are consistent with the data, the observations seem to favour DC = 1 for Model A., the model with the most freedom.

In general fits were of acceptable quality (see the χ^2 values in Table 1). We discuss some mild tensions in the data more in Section 5.1, but summarize the results of each model here. The M_{min} , M_1 , and DC free model (A.) is free to go to high masses until tension between model and measured number density stops 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 (B., C., D.) 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} , B.), the model actually prefers to go even lower than the abundance matching halo mass, and uses the DC to reach agreement with the number counts. However when these models have the DC fixed (C., D.), 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 (C.) to agree with the observed number counts, the intrinsic number counts must be higher than for DC = 1 (D.), forcing the model to prefer slightly lower halo masses. The ‘pure bias’ model (E.) was able to fit the clustering data well, but with a higher bias, and much larger uncertainty. This illustrates that the clustering is a comparatively weak constraint in models A.–D., and that much of the constraint on the model is coming from the number counts. The reason that models A.–D. are able to achieve stronger constraints on the bias than Model E. is that the number counts are (implicitly) being mapped on to a halo mass via abundance matching, and then on to a bias by the halo bias model, as opposed to the bias being measured to that precision directly (this can only happen because the satellite fraction is so low; at lower redshift number counts do not map on to bias as cleanly as the satellite fraction is in general non-trivial). Essentially the number counts set a maximum possible bias, and the clustering gives a weak constraint pulling the estimate up to that limit. There is a mild mismatch between the clustering and the number counts – haloes of the halo mass implied by the (directly measured) bias are far rarer than the observed galaxies are.⁶ This is in principle problematic,⁷ but the issue has low statistical significance in the case discussed here. In addition, the fact that no one-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 the normalization constraint, which overestimates the abundance of the HMF by a factor of 1.7 at $z = 4$ (Harikane, private communication, see also Harikane et al. 2017). So the results of Harikane et al. (2016) are likely more consistent with a DC of 1 (rather than the fixed value of 0.6 that they use in their analysis). Table 2 shows the observed comoving number densities, compared

⁶This 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).

⁷This 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.

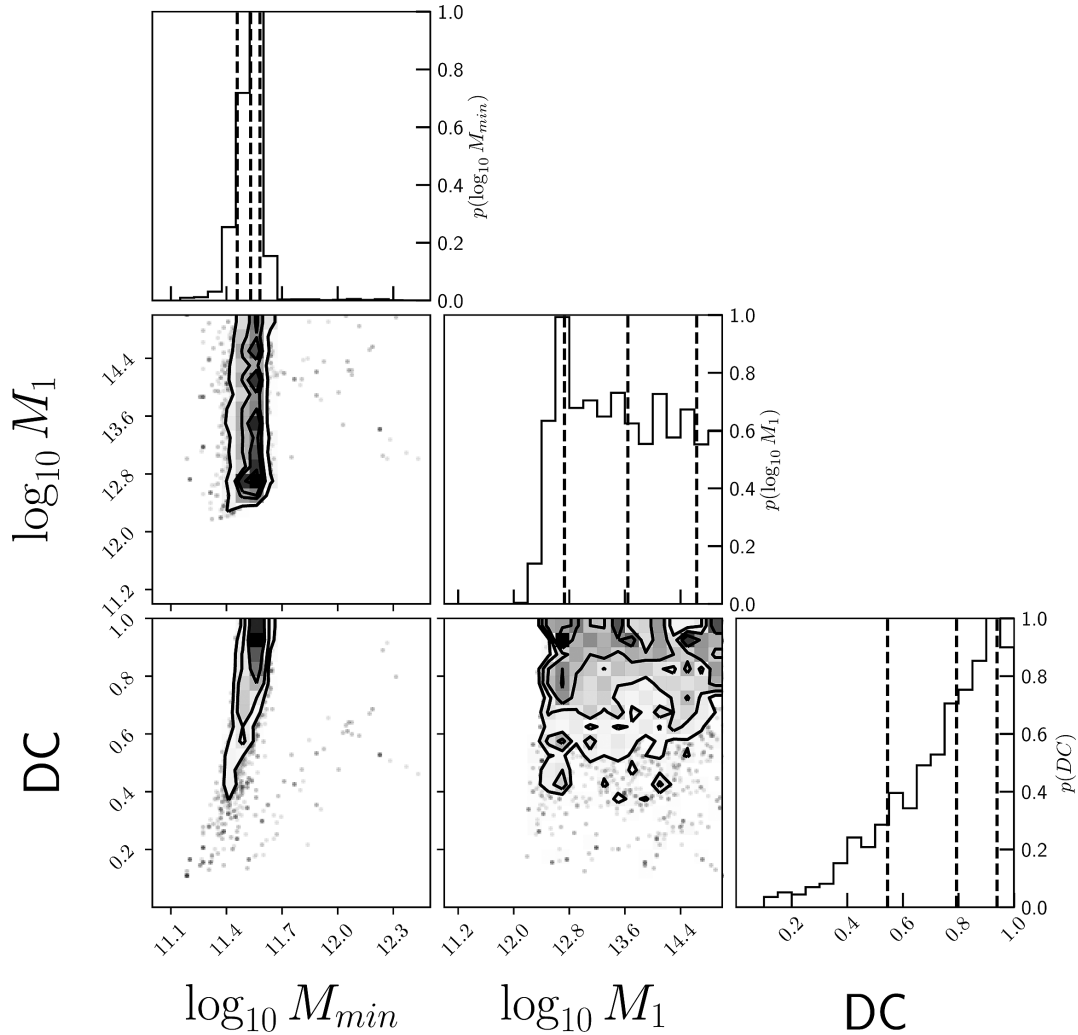


Figure 6. Triangle plot of our posterior from our MCMC fitting for the HOD model with M_{\min} , M_1 , and DC free (masses in log base 10 Solar mass units). The top sub-figure in each column shows the 1D marginalized posteriors (probability density functions scaled to have a peak of unity) of each parameter, with vertical dashed lines showing the 16th, 50th, and 84th percentiles. The other three sub-plots show each possible 2D marginalized posterior.

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 (Mpc^{-3}), (4) the minimum halo mass in the most straightforward abundance matching scheme (log base ten Solar mass units), (5) the model comoving number density (Mpc^{-3}) of the best fit model HOD in this work (A. shown) and Harikane et al. (2016) without incorporating DC, (6) the corresponding minimum halo mass from the HOD model (log base 10 Solar mass units).

Data	M_{UV}	$n_{\text{g}}^{\text{observed}} (\text{Mpc}^{-3})$	$\log M_{\text{min}}^{\text{matched}}$	$n_{\text{g}}^{\text{model}} (\text{Mpc}^{-3})$	$\log M_{\text{min}}^{\text{model}}$
Bowler15	-21.125	4.1×10^{-5}	11.51	6.8×10^{-5}	$11.53^{+0.05}_{-0.07}$
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}	$11.03^{+0.05}_{-0.18}$

with the comoving number densities implied by their HOD models (alongside corresponding figures from this work).

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 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 $10^{-3.25} < \theta/\text{deg} < 10^{-1.25}$, compared with our $10^{-3} < \theta/\text{deg} < 10^{-0.5}$, and for fainter rest-frame absolute magnitudes of $-20.5 < M_{\text{UV}} < -19$ compared with our $-22.7 < M_{\text{UV}} < -21.125$ sample (see Fig. 7). 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. 8. 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

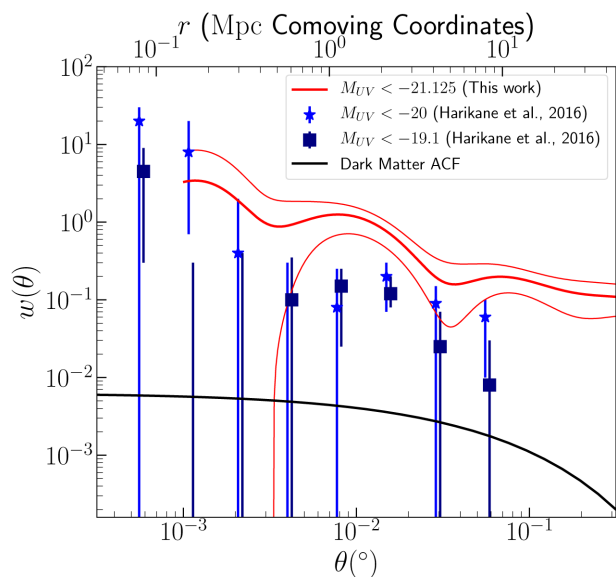


Figure 7. Comparison of our measurements (red curve, 1σ uncertainties in the lighter curves), the lower luminosity Harikane et al. (2016) measurements, and the dark matter angular correlation function (black curve).

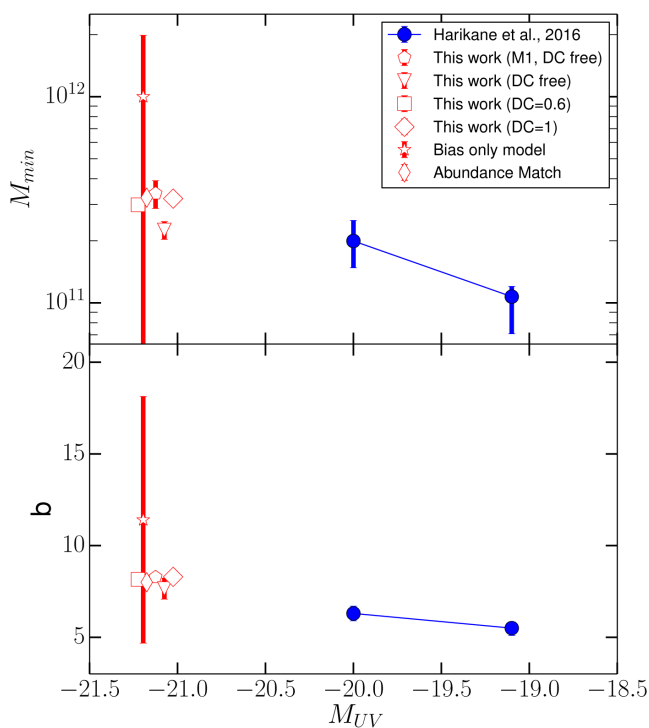


Figure 8. Comparison of our results with comparable measurements of lower luminosity LBGs from Harikane et al. (2016). Top plot: M_{\min} as a function of absolute UV luminosity threshold (in units of Solar mass). Bottom plot: galaxy bias as a function of absolute UV luminosity threshold. The results from our six different models are shown for comparison (x-axis values slightly offset for each model for clarity).

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 DC. The comoving space density of the galaxies in our sample is $4.1 \times 10^{-5} \text{ Mpc}^{-3}$, compared with $3.8 \times 10^{-4} \text{ Mpc}^{-3}$ for the most luminous $z = 6$ Harikane et al. (2016) sample. Harikane et al. (2016) do not measure the DC for this sample, but assume it to be equal to 0.6. As an illustrative example, a DC of 0.6 for the Harikane et al. (2016) sample would mean an actual underlying population comoving density of $6.3 \times 10^{-4} \text{ Mpc}^{-3}$. 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 per cent of its time in this superenhanced 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-luminosity objects in very dense environments. Although the measurement of the correlation function for the UDS field is too poor to make a measurement of the bias, the bias (Model E.) from just the UltraVISTA field is $b = 11.1^{+7.6}_{-6.8}$, consistent with this result. In future larger data sets it will be important to perform the full HOD analysis for separate fields individually to understand how cosmic variance impacts inferred HOD parameters. Planned work with $z \sim 6-7$ LBG samples selected in the VIDEO survey (which goes to comparable depths to the analysis in this work) will allow this comparison to be done with 12 deg^2 of data over three roughly equally sized independent fields.

Bowler et al. (2014, 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. Our clustering results show 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 starbursts. This result gives extra support to the claim that the high-redshift evolution of the high-luminosity end of the luminosity function is determined by dust extinction, cooling rate of gas in haloes, or feedback processes, as opposed to how rare its episodes of high star formation are.

Preliminary results from the Great Optically Luminous Dropout Research Using Subaru HSC LBG survey (GOLDRUSH, Ono et al. 2017) have started to give complementary results over an even larger volume, with deep optical imaging over 100 deg^2 . Ono et al. (2017) suggests that at these redshifts the high-luminosity end of the luminosity function is indeed best fit by a double power law, supporting Bowler et al. (2014, 2015). Similarly, Harikane et al. (2017) make clustering measurements in GOLDRUSH that suggest clustering amplitude continues to increase at high luminosities, and that the ratio of luminosity to halo mass remains high at high halo masses at $z \sim 6$. However their results currently only use optical imaging (unlike our sources, which were selected using near-infrared bands), which makes processes like removing brown dwarf contamination more complex.

5.2 Apparent lack of a one-halo term

Models using extrapolated values of M_1 (B., C., D.) 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 haloes (see Fig. 5). We do not observe this in the data, in contrast to Harikane et al. (2016), see Fig. 7. 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, Srianand & Subramanian (2013), who also observe a lack of a one-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 $M_{\min} \sim 10^{11.4} M_{\odot}$, then we would expect $M_1 \sim 10^{12.2} M_{\odot}$, e.g. only haloes with $M > 10^{12.2} M_{\odot}$ host more than one of our bright sample. The comoving density of $M > 10^{12.2} M_{\odot}$ haloes is $5.2 \times 10^{-7} \text{ Mpc}^{-3}$ and the comoving volume probed by the observations is $\sim 1.7 \times 10^7 \text{ Mpc}^3$. This means that the expected numbers of $M > 10^{12.2} M_{\odot}$ haloes in the volume surveyed is ~ 10 . Just these 10 would give ~ 10 close pairs (in addition a few more would be expected from projection effects), which is more than twice the four close (10^{-3} – $10^{-2.5}$ deg) pairs observed here, and would push the small-scale correlation function up. However these haloes will be extremely biased, much more than the $\sim 10^{11.4} M_{\odot}$ haloes. Conceivably for an extreme case, it could be that if our observations were repeated 10 times, we would find that nine times no $M > 10^{12.2} M_{\odot}$ haloes were observed, and the 10th 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 10th 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 one-halo term is dominated by contributions from very massive haloes, which are the most biased, so there is the most cosmic variance on small scale measurements of the correlation function. This is also consistent with our less dense field (UDS), having no power on small scales. It appears to be the case that neither of our fields are overdense enough to sample the highly biased sample of massive haloes at this redshift 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 one-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 haloes required to host multiple galaxies is less ex-

treme 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 haloes necessary for satellites.

5.3 Mild mismatch between the number counts and bias measurements

As discussed in Section 4.2, the bias obtained by fitting just to the clustering (Model E.) is higher than that obtained when fitting to both the counts and the clustering. The core of the discrepancy is that to obtain the directly measured bias of ~ 11 , the galaxies would need to be in haloes of minimum mass $M_{\min} \sim 10^{12} M_{\odot}$. This corresponds to a comoving density of $\sim 2 \times 10^{-6} \text{ Mpc}^{-3}$, compared with an observed number density of $4.1 \times 10^{-5} \text{ Mpc}^{-3}$, i.e. approximately a factor of 20 lower than the observed value. Equivalently, plain abundance matching (Model F.) would suggest a minimum mass $M_{\min} \sim 10^{11.5} M_{\odot}$ (see Table 2) 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 DC of 1 was needed for their LBG sample, and that even then the measured bias was slightly inconsistent with the number density.

Although of comparatively low statistical significance, we consider some possible explanations for this discrepancy:

(i) Contaminants: 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. In addition, Bowler et al. (2014, 2015) investigated the possibility of contamination from lower luminosity sources being magnified by gravitational lensing. They found that lensing was not a significant factor in either the selection of the sources, or the constraints on the luminosity function, based on the non-detection of strong lensing sources along the lines of sight, and the fact that the sources have no excess lensing over random locations in the field.

(ii) More complex galaxy–halo relations: halo occupancy depends on more than just halo mass. 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, Behroozi & van den Bosch 2016, where the bias of haloes depends on halo assembly history as well as mass), or a dependence on the large-scale linear density field. Galaxy–galaxy lensing can in principle observationally break these degeneracies (e.g. if sources are in older, lower mass haloes, 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.

(iii) Uncertainty in knowledge of the high-redshift dark matter distribution: that the halo mass function, etc. used in the HOD model are flawed. Tinker et al. (2008, 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 per cent at $z = 6$ for $M \sim 10^{11.3} M_\odot$ haloes). 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 per cent, which would require around a 25 per cent 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 haloes, 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, Peacock & Heavens 1998). Introducing a functional form for scale-dependent bias can model some of these effects and Jose, Lacey & Baugh (2016) conclude that quasi-linear clustering has the largest effect at high redshift ($z > 2$), and high 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 overestimate halo mass by up to a factor of ten if unaccounted for.

In summary there are a few possible ways for the measured bias to be higher than the number counts would seem to permit. Significant contamination seems unlikely, and uncertainty in the modelling of the distribution of dark matter in the high-redshift Universe would seem a simpler issue to consider before invoking more complex galaxy–halo connections. Although the discrepancy has a low statistical significance, it seems that incorporation of a high-redshift halo bias calibration, or quasi-linear bias could be worthwhile in future analyses.

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/underdense, 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 one-point statistics, but different two-point statistics. Thus we can use the two-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 one-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).

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

$$\bar{\omega}(A) = \frac{\int_A \int_A \omega(|\vec{\theta}_i - \vec{\theta}_j|) d^2\theta_i d^2\theta_j}{\int_A \int_A d^2\theta_i d^2\theta_j}, \quad (10)$$

where A is the angular region of the field, ω is the two-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-fitting model from Section 4.2), and then taking the average.

We use the formalism of Trenti & Stiavelli (2008) that summarizes well established results from Peebles (1980), Peebles (1993), Newman & Davis (2002), and Somerville et al. (2004) that conclude

$$\bar{\omega}(A) = \frac{\langle N^2 \rangle - \langle N \rangle^2}{\langle N \rangle^2} - \frac{1}{\langle N \rangle}, \quad (11)$$

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

$$\langle N^2 \rangle - \langle N \rangle^2 = \langle N \rangle + \bar{\omega}(A) \langle N \rangle^2, \\ \sigma_{\text{total}}^2 = \sigma_{\text{Poisson}}^2 + \sigma_{\text{CV}}^2, \quad (12)$$

where $\sigma_{\text{total}} = \sqrt{\langle N^2 \rangle - \langle N \rangle^2}$ is the total standard deviation on measurements of number counts, $\sigma_{\text{Poisson}} = \sqrt{\langle N \rangle}$ is the Poisson standard deviation and $\sigma_{\text{CV}} = \sqrt{\bar{\omega}(A) \langle N \rangle}$ 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 $\sigma_{\text{CV}} = b\sqrt{\omega_{\text{DM}}(A) \langle N \rangle}$ in the ‘pure-bias’ case where $\omega = b^2 \omega_{\text{DM}}$, 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 underdensity. 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 summarizes our results when applied to the UltraVISTA and UDS fields for our $z \sim 6$ samples (using our best-fitting pure bias model, Model E.). The most important columns to compare are the ρ_a (the actual surface density of galaxies in the field), ρ_e (the expected surface density there would be if both fields had the average density) and σ_{total} the standard deviation on the surface density measurement including the Poisson and cosmic variance implied by our clustering measurements. The observed over/underdensities of galaxies in the UltraVISTA and UDS fields are 1.75σ and 1.5σ deviations from the model value respectively. For comparison, a straightforward estimation of σ_{Total} using the conventional expression for the standard deviation would suggest $\sigma_{\text{Total}} \approx 57\text{deg}^{-2}$ for the individual sub-fields, substantially higher than our quoted values. This suggests either that our estimates of the cosmic variance in Table 3 are underestimates, or that our fields are extreme over/underdensities. As noted in Bowler et al. (2014), they are the most over- and underdense respectively of the five CANDELS fields, which lends support to the latter interpretation – the observed densities appear to be unusual/extreme over/underdensities. For reference, the comoving number density of Model A. (see Table 2) would correspond to a source surface density of $\sim 169\text{deg}^{-2}$ when including DC. To summarize, at these redshifts, both UltraVISTA and UDS appear to be moderate, but not unreasonable, over- and underdensities respectively. Future work will seek to also better understand a) the cosmic variance uncertainty on our clustering measurements (imperfectly accounted for by our bootstrap-object approach) alongside the cosmic variance on the number counts, and b) the impact of cosmic variance on inferred HOD parameters by modelling the measurements separately in each of the VIDEO fields.

5.4.1 Estimating the cosmic variance on the clustering measurements

Our results in Table 3 suggest that cosmic variance still contributes substantially to the uncertainty on our number counts, even though we are using two highly independent fields. This suggests the uncertainties on our clustering measurements could also be substantially underestimated, leading to underestimates of the uncertainty on inferred parameters. As discussed earlier, a full examination of the role of cosmic variance on the clustering measurements is beyond the scope of this work. However, to gain an approximate estimate of the impact of cosmic variance on the inferred parameters, we make the simple ansatz of scaling the clustering covariance matrix by the amount the counts uncertainty is increased by the presence of cosmic variance, i.e. $(\sigma_{\text{Total}}/\sigma_{\text{Poisson}})^2 \approx 3.2$ (shown with dashed green lines in Fig. 2). We then repeat the analysis in Section 4 using these ‘cosmic variance corrected’ uncertainties. We find that for Model E. (bias only), this correction alters the inferred bias from $b = 11.4^{+6.5}_{-6.7}$ to $b = 13.0^{+8.5}_{-8.0}$. For Model A., $\log_{10} M_{\text{min}} = 11.53^{+0.05}_{-0.07}$ becomes $\log_{10} M_{\text{min}} = 11.40^{+0.06}_{-0.11}$ and still favours DC ≈ 1 and $\log_{10} M_{\text{min}} > 12.2$ (low satellite fraction). The true uncertainties on our inferred parameters are likely somewhere between those in Section 4 and those with the ‘corrected’ uncertainties (as bootstrap objects does not *completely* neglect cosmic variance) – but these results using the corrected uncertainties illustrate that our key conclusions are likely unaltered even when cosmic variance on clustering measurements is accounted for.

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 an HOD scheme.

The key conclusions of this work are as follows:

(i) Bright LBGs ($M_{\text{UV}} \leq -21$) appear to be highly biased ($b \sim 11 \pm 7$ based on the clustering, favouring $b \sim 8$ if the number counts are also taken into account) objects in dense environments, as opposed to being rare temporal episodic incarnations of fainter galaxies ($M_{\text{UV}} \sim -19$). Our results have important implications for the physical origin of the observed steepening of the bright end of the ultraviolet 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 $z \sim 6$. This measured high bias potentially suggests that the bright-end of the luminosity function at $z \sim 6$ could be determined by feedback processes or dust obscuration, rather than DCs.

(ii) The bias based on the clustering is higher than that suggested by the number counts, potentially suggesting that some modification to our knowledge of the high-redshift dark matter distribution could be 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.

(iii) 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 reasonably extreme $\sim 1.6\sigma$ over/underdensities.

(iv) We do not require DC to explain our observations (equivalently DC ~ 1), and the satellite fraction of the sources is very small, at most a few per cent.

In the next few years, deep, wide surveys such as VIDEO and the VISTA Extragalactic Infrared Legacy Survey (VEILS, Hönig et al. 2017), 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 1000s of LBGs (Bowler et al. 2017), which will reveal how the measured large scale structure of LBGs and LAEs relates to reionization.

ACKNOWLEDGEMENTS

The authors thank the anonymous referee for the comments that have improved the quality of this paper. PH and CH wish to acknowledge support provided through STFC studentships. PH wishes to thank the Rector and Fellows of Lincoln College for support through the Graduate Research Fund. Many thanks to Yuichi Harikane and Olmo Piana 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

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_a), the actual angular galaxy density in the field (ρ_a , in deg^{-2}), the expected number of galaxies in the field if it had the mean density based on the two fields combined (N_e), the expected angular galaxy density in the field – all identical figures as we are considering deviations from the mean density (ρ_e , in deg^{-2}), the expected value of the correlation function over the field ($\bar{\omega}$), standard deviation of the surface density estimate from Poisson statistics based on the square root of N_e (σ_{Poisson} , in deg^{-2}), standard deviation on the surface density from cosmic variance, estimated from our clustering measurements (σ_{CV} , in deg^{-2}), our Poisson and cosmic variance errors added in quadrature (σ_{Total} , in deg^{-2}).

Field	Area (deg^2)	N_a	ρ_a (deg^{-2})	N_e	ρ_e (deg^{-2})	$\bar{\omega}$	σ_{Poisson} (deg^{-2})	σ_{CV} (deg^{-2})	σ_{Total} (deg^{-2})
UDS	0.74	64	86	92	124	0.031	13	22	25
UltraVISTA	0.62	103	166	77	124	0.026	14	20	24
Total	1.35	167	124	167	124	0.014	10	15	18

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

Planck Collaboration XLVII, 2016, *A&A*, 596, A108
 Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Giallongo E., 1999, *MNRAS*, 310, 540
 Bahcall N. A., Soneira R. M., 1983, *ApJ*, 270, 20
 Barone-Nugent R. L. et al., 2014, *ApJ*, 793, 17
 Bartelmann M., Schneider P., 2001, *Phys. Rep.*, 340, 291
 Becker G. D., Bolton J. S., Lidz A., 2015, *PASA*, 32, e045
 Behroozi P., Silk J., 2016, preprint ([arXiv:1609.04402](https://arxiv.org/abs/1609.04402))
 Behroozi P. S., Wechsler R. H., Conroy C., 2013, *ApJ*, 770, 57
 Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000, *MNRAS*, 311, 793
 Berlind A. A., Weinberg D. H., 2002, *ApJ*, 575, 587
 Bertin E., Arnouts S., 1996, *A&AS*, 117, 393
 Beutler F. et al., 2011, *MNRAS*, 416, 3017
 Bhowmick A. K., Di Matteo T., Feng Y., Lanusse F., 2017, *MNRAS*, 474, 5393
 Bouwens R. J., Illingworth G. D., Franx M., Ford H., 2007, *ApJ*, 670, 928
 Bouwens R. J. et al., 2015, *ApJ*, 803, 34
 Bouwens R. J. et al., 2016, *ApJ*, 830, 67
 Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, *MNRAS*, 370, 645
 Bowler R. A. A. et al., 2014, *MNRAS*, 440, 2810
 Bowler R. A. A. et al., 2015, *MNRAS*, 452, 1817
 Bowler R. A. A., Dunlop J. S., McLure R. J., McLeod D. J., 2017, *MNRAS*, 466, 3612
 Brainerd T. G., Blandford R. D., Smail I., 1996, *ApJ*, 466, 623
 Cabre A., Fosalba P., Gaztanaga E., Manera M., 2007, *MNRAS*, 381, 1347
 Clay S. J., Thomas P. A., Wilkins S. M., Henriques B. M. B., 2015, *MNRAS*, 451, 2692
 Conroy C., Wechsler R. H., Kravtsov A. V., 2006, *ApJ*, 647, 201
 Cooray A., Sheth R., 2002, *Phys. Rep.*, 372, 1
 Coupon J. et al., 2012, *A&A*, 542, A5
 Coupon J. et al., 2015, *MNRAS*, 449, 1352
 Cress C. M., Helfand D. J., Becker R. H., Gregg M. D., White R. L., 1996, *ApJ*, 473, 7

Croton D. J. et al., 2006, *MNRAS*, 365, 11
 Dunlop J. S. et al., 2013, *MNRAS*, 432, 3520
 Finkelstein S. L. et al., 2015, *ApJ*, 810, 71
 Fisher J., Davis K. B., Strauss M., Yahil M. A., Huchra A., 1994, *MNRAS*, 266
 Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
 Furusawa H. et al., 2008, *ApJS*, 176, 1
 Giallisco M., 2002, *ARA&A*, 40, 579
 Gott J. R., Choi Y.-Y., Park C., Kim J., 2009, *ApJ*, 695, L45
 Guhathakurta P., Tyson J. A., Majewski S. R., 1990, *ApJ*, 357, L9
 Guo H. et al., 2016, *MNRAS*, 459, 3040
 Harikane Y. et al., 2016, *ApJ*, 821, 123
 Harikane Y. et al., 2017, *PASJ*, 70, S11
 Hartley W. G., Almaini O., Foucaud S., 2013, in *Astrophysics and Space Science Proceedings*, Vol. 37, Thirty Years of Astronomical Discovery with UKIRT. Springer Science+Business Media, p. 309
 Hatfield P. W., Lindsay S. N., Jarvis M. J., Häußler B., Vaccari M., Verma A., 2016, *MNRAS*, 459, 2618
 Hearin A. P., Behroozi P. S., van den Bosch F. C., 2016, *MNRAS*, 461, 2135
 Hönig S. F. et al., 2017, *MNRAS*, 464, 1693
 Ilbert O. et al., 2006, *A&A*, 457, 841
 Jarvis M. J., Hatfield P. W., 2017, *MNRAS*, 3570
 Jarvis M. J. et al., 2013, *MNRAS*, 428, 1281
 Johnston R., 2011, *A&AR*, 19, 41
 Jose C., Srianand R., Subramanian K., 2013, *MNRAS*, 435, 368
 Jose C., Lacey C. G., Baugh C. M., 2016, *MNRAS*, 463, 270
 Jose C., Baugh C. M., Lacey C. G., Subramanian K., 2017, *MNRAS*, 465, 111
 Jullo E., Kneib J.-P., Limousin M., Elíasdóttir Á., Marshall P. J., Verdugo T., 2007, *New J. Phys.*, 9, 447
 Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlobber S., Allgood B., Primack J. R., 2004, *ApJ*, 609, 35
 Lacasa F., Rosenfeld R., 2016, *J. Cosmol. Astropart. Phys.*, 2016, 005
 Lacey C. G. et al., 2016, *MNRAS*, 462, 3854
 Laigle C. et al., 2016, *ApJS*, 224, 24
 Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64
 Lawrence A. et al., 2007, *MNRAS*, 379, 1599
 Limber D. N., 1954, *ApJ*, 119, 655
 Lindsay S. N., Jarvis M. J., McAlpine K., 2014, *MNRAS*, 440, 2322
 Ling E. N., Barrow J. D., Frenk C. S., 1986, *MNRAS*, 223, 21P
 Ma C.-P., Fry J. N., 2000, *ApJ*, 543, 503
 Madau P., Dickinson M., 2014, *ARA&A*, 52, 415
 Maddox S. J., Efstathiou G., Sutherland W. J., Loveday J., 1990, *MNRAS*, 242, 43P
 Mandelbaum R. et al., 2014, *ApJS*, 212, 5
 Mann R. G., Peacock J. A., Heavens A. F., 1998, *MNRAS*, 293, 209
 McCracken H. J. et al., 2012, *A&A*, 544, A156
 McCracken H. J. et al., 2015, *MNRAS*, 449, 901
 McLeod D. J., McLure R. J., Dunlop J. S., 2016, *MNRAS*, 459, 3812
 McLure R. J., Cirasuolo M., Dunlop J. S., Foucaud S., Almaini O., 2009, *MNRAS*, 395, 2196
 McLure R. J. et al., 2013, *MNRAS*, 432, 2696

- McQuinn M., Hernquist L., Zaldarriaga M., Dutta S., 2007, *MNRAS*, 381, 75
- Mo H. J., Jing Y. P., Boerner G., 1992, *ApJ*, 392, 452
- Moster B. P., Somerville R. S., Maibetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, *ApJ*, 710, 903
- Murray S. G., Power C., Robotham A. S. G., 2013, *MNRAS*, 434, L61
- Natarajan A., Yoshida N., 2014, *Prog. Theor. Exp. Phys.*, 6B112
- Navarro J. F., Frenk C. S., White S. D. M., 1996, *ApJ*, 462, 563
- Newman J. A., Davis M., 2002, *ApJ*, 564, 567
- Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, *MNRAS*, 396, 19
- Oesch P. A. et al., 2016, *ApJ*, 819, 129
- Oke J. B., Gunn J. E., 1983, *ApJ*, 266, 713
- Ono Y. et al., 2017, *PASJ*, 70, S10
- Ouchi M. et al., 2009, *ApJ*, 706, 1136
- Ouchi M. et al., 2010, *ApJ*, 723, 869
- Paczynski B., 1987, *Nature*, 325, 572
- Peebles P. J. E., 1980, *The Large-Scale Structure of the Universe*. Princeton Univ. Press, Princeton, NJ
- Peebles P. J. E., 1993, *Principles of Physical Cosmology*. Princeton Univ. Press, Princeton, NJ
- Peebles P. J. E., Hauser M. G., 1974, *ApJS*, 28, 19
- Pentericci L. et al., 2014, *ApJ*, 793, 113
- Roche N., Eales S. A., 1999, *MNRAS*, 307, 703
- Schechter P., 1976, *ApJ*, 203, 297
- Scoville N. et al., 2007, *ApJS*, 172, 38
- Seljak U., 2000, *MNRAS*, 318, 203
- Shao J., 1986, *Ann. Stat.*, 14, 1322
- Shapley A. E., 2011, *ARA&A*, 49, 525
- Silk J., 2010, in Carignan C., Combes F., Freeman K. C., eds. *Proc. IAU Symp.* 6, 273
- Somerville R. S., Lee K., Ferguson H. C., Gardner J. P., Moustakas L. A., Giavalisco M., 2004, *ApJ*, 600, L171
- Soucail G., Mellier Y., Fort B., Mathez G., Cailloux M., 1988, *A&A*, 191, L19
- Stark D. P., 2016, *ARA&A*, 54, 761
- Stark D. P., Ellis R. S., Bunker A., Bundy K., Targett T., Benson A., Lacy M., 2009, *ApJ*, 697, 1493
- Steidel C. C., Hamilton D., 1992, *AJ*, 104, 941
- Steidel C. C., Giavalisco M., Pettini M., Dickinson M., Adelberger K. L., 1996, *ApJ*, 462, L17
- Tegmark M., 2003, American Physical Society, April Meeting, 2003, Philadelphia, Pennsylvania, MEETING ID: APR03, abstract #T1.003
- Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, *ApJ*, 688, 709
- Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, *ApJ*, 724, 878
- Trenti M., Stiavelli M., 2008, *ApJ*, 676, 767
- Vale A., Ostriker J. P., 2004, *MNRAS*, 353, 189
- White S. D. M., 1979, *MNRAS*, 186, 145
- Zaroubi S., 2013, in Wiklind T., Mobasher B., Bromm V., eds. *As-trophysics and Space Science Library*, Vol. 396, The First Galaxies, Springer-Verlag, Berlin, p. 45
- Zehavi I. et al., 2005, *ApJ*, 621, 22
- Zheng Z. et al., 2005, *ApJ*, 633, 791

This paper has been typeset from a \LaTeX file prepared by the author.