**The Hα luminosity and stellar mass dependent clustering of star-forming galaxies at 0.7 < z < 1.5 with 3D-HST**

Callie Clontz\(^1\)*, David Wake\(^2\)†, and Zheng Zheng\(^1\)‡

\(^1\)Department of Physics and Astronomy, University of Utah, 201 James Fletcher Building 115 S. 1400 E., Salt Lake City, UT 84112, USA

\(^2\)Department of Physics and Astronomy, University of North Carolina Asheville, 1 University Heights, Asheville, NC 28804, USA

Accepted XXX. Received YYY; in original form ZZZ

**ABSTRACT**

We present measurements of the dependence of the clustering amplitude of galaxies on their star formation rate (SFR) and stellar mass \((M_*)\) at 0.7 < \(z\) < 1.5 to assess the extent to which environment affects these properties. While these relations are well determined in the local universe, they are much more poorly known at earlier times. For this analysis we make use of the near-IR HST WFC3 grism spectroscopic data in the five CANDELS fields obtained as part of the 3D-HST survey. We make projected 2-point correlation function measurements using ~6,000 galaxies with accurate redshifts, \(M_*\), and Hα luminosities. We find a strong dependence of clustering amplitude on Hα luminosity and thus SFR. However, at fixed \(M_*\), the clustering dependence on Hα luminosity is largely eliminated. We model the clustering of these galaxies within the Halo Occupation Distribution framework using the conditional luminosity function model and the newly developed conditional stellar mass and Hα luminosity distribution model. These show that galaxies with higher SFRs tend to live in higher mass haloes, but this is largely driven by the relationship between SFR and \(M_*\). Finally, we show that the small residual correlation between clustering amplitude and Hα luminosity at fixed \(M_*\) is likely being driven by a broadening of the SFR-\(M_*\) relationship for satellite galaxies.

**Key words:** galaxies: evolution – galaxies: high-redshift – cosmology: large-scale structure

---

**1 INTRODUCTION**

In the cold dark matter (CDM) paradigm of structure formation, the tiny matter density fluctuations resulting from quantum fluctuations during inflation grow over time under the influence of gravitational interactions. Following the formation of dark matter haloes, gas is accreted, cools down, and forms stars to produce galaxies. Galaxies are a natural tracer of the underlying matter distribution to study cosmology. In particular, emission line galaxies, which are mainly star-forming galaxies, have become important targets in current and planned galaxy surveys, such as SDSS-IV eBOSS (Dawson et al. 2016), DESI (DESI Collaboration et al. 2016), and surveys with the Roman Space Telescope (RST; Spergel et al. 2015) and Euclid (Laureijs et al. 2011). In this work, we measure the clustering of Hα-emitting galaxies from Hubble Space Telescope (HST) observations and model the measurements to study the relation between these galaxies and dark matter haloes.

Galaxies are a biased tracer of the underlying matter density field that is connected to cosmology. The galaxy bias encodes information about galaxy formation processes, and a good understanding of it can help tighten cosmological constraints with galaxy clustering. It can be described at the level of individual dark matter haloes as a relation between galaxies and haloes. Such a halo occupation distribution (HOD) formalism specifies how galaxies of a given sample occupy dark matter haloes, which can be constrained by galaxy clustering data (such as the widely used measurements of the two-point correlation function, hereafter the 2PCF).

It has been shown that galaxy properties, including star formation rate (SFR) and stellar mass \((M_*)\), are correlated with their environment (Blanton & Moustakas 2009) and a deeper look into this relationship will facilitate a better understanding of galaxy formation and evolution. The finer details that describe the extent to which environmental processes affect galaxy evolution are still unknown. Processes such as feedback from star formation, black hole accretion, as well as gas cooling each act on various timescales and have different dependencies on galaxy mass and environment. One way to characterise the environment dependence of galaxy properties is to measure the dependence of galaxy clustering on galaxy properties. The HOD modelling of the clustering measurements connects galaxies to their natural environment, the dark matter haloes. The dependence of galaxy clustering on galaxy properties (such as luminosity and colour) has been intensively studied (e.g. Zheng et al. 2007; Zehavi et al. 2011).

The primary goal of this research is to use clustering of star-forming galaxies to investigate the role environment plays in determining the specific SFR (sSFR) and stellar mass of galaxies. It has been shown in the local universe that SFR depends strongly on environment, where galaxies with lower sSFR are found in denser environments (Blanton & Moustakas 2009) and that the clustering amplitude has a strong dependence on SFR (e.g. Li et al. 2008). In this work, we use the data from 3D-HST with well-defined redshifts and photometry to analyse the clustering of Hα-emitting galaxies...
Figure 1. Footprints on the sky of the five survey fields. In each panel, the light blue region shows the footprint, with holes from bright star masks. The orange dots are Hα-emitting galaxies.

and to study the connection between SFR and stellar mass and dark matter haloes at an earlier time.

Our work is similar to previous works by Cochrane et al. (2017) and Cochrane et al. (2018), who studied the clustering of ~ 4,000 Hα emitters at z = 0.8, 1.47, and 2.23 using the narrow-band High-Redshift (Z) Emission Line Survey (HIZELS). We perform similar measurements with the 3D-HST grism spectroscopic survey at 0.7 < z < 1.5 with the common goal of tracking the evolution of the clustering dependence on galaxy properties. The measurements and modelling results provide useful inputs to plan future surveys, such as those with RST (Spergel et al. 2015) and Euclid (Laureijs et al. 2011).

The paper is organised as follows. In Section 2, we introduce the 3D-HST data and the construction of various Hα-emitting galaxy samples. In Section 3, we describe the 2PCF measurements. In Section 4, the dependence of 2PCFs on Hα luminosity and stellar mass is presented and characterised. The halo-based models are introduced and applied to model the 2PCF measurements in Section 5, which include a conditional luminosity function (CLF) model to interpret the dependence of 2PCFs on Hα luminosity and a conditional mass and Hα luminosity distribution model to interpret the joint dependence of 2PCFs on stellar mass and Hα luminosity. We compare our results with previous work in Section 6 and conclude in Section 7.

Throughout this paper, in carrying out the clustering measurements, we adopt a spatially-flat ΛCDM cosmology with density parameters \( \Omega_m = 0.3 \) and \( \Omega_k = 0.7 \). In modelling the clustering, we further adopt the following cosmological parameters, \( H_0 = 100h \text{km s}^{-1}\text{Mpc}^{-1} \) with \( h = 0.68 \), \( n_s = 0.96 \), and \( \sigma_8 = 0.81 \). Haloes are defined as bound regions with mean density 200 times that of the background universe.

2 DATA

In this work, we make use of near-IR HST WFC3 grism spectroscopic data in the five CANDELS fields obtained as part of the 3D-HST survey (Brammer et al. 2012; Momcheva et al. 2016; Skelton et al. 2014). This survey, carried out by the Hubble Space Telescope in 2015, consists of deep near-infrared low resolution WFC3/G141 grism spectroscopic measurements, which are combined with existing deep multi-wavelength photometry to yield accurate redshifts (\( \sigma_z/(1+z) \sim 0.0003 \)), stellar masses, and emission line measurements. The G141 grism provides a spectral resolution \( R \sim 130 \) and covers the wavelength range of 1.1 μm to 1.65 μm. The 3D-HST footprints of the five fields (AEGIS, COSMOS, GOODS-N, GOODS-S, and UDS) cover a total area of ~625 arcmin\(^2\), and are shown in Fig. 1 together with the Hα-emitting galaxies used in our analysis.

In this work we are interested in the Hα emission line which is well detected in the G141 grism spectra over a redshift range of 0.7 < z < 1.5, setting a natural redshift range for our analysis. The 5σ limiting flux of the 3D-HST G141 observations is \( \sim 3 \times 10^{-17} \text{erg s}^{-1}\text{cm}^{-2} \) (Brammer et al. 2012).

To define our master sample of Hα emitting galaxies we select all galaxies with 0.7 < z < 1.5, with clean photometry (use_phot=1), a grism redshift, jh_mag < 26, and Hα flux > 4 \( \times \) \( 10^{-17} \) erg s\(^{-1}\) cm\(^{-2}\). This redshift range corresponds to the wavelength range of the G141 grism where Hα falls and the sensitivity remains high. The other cuts ensure a clean sample of galaxies with robust Hα detections.

To obtain the Hα luminosity of each galaxy, we make a correction to remove the contamination from the adjacent [NII] emission lines that are not separately resolved in the grism spectra. We use the stellar mass – [NII]/Hα flux ratio (stellar mass–metallicity relation) in Wuyts et al. (2014),

\[
\log ([\text{N} \text{II}]/\text{H} \alpha) = \frac{1}{0.57} \left( -0.1 + \log \left( 1 - \exp \left( -\frac{M_*}{10^{10.2} M_\odot} \right)^{0.4} \right) \right)
\]

(1)

In our galaxy sample, the median of the correction factor \( 1/(1 + [\text{NII}]/\text{H} \alpha) \) is about –0.06 dex.

While our sample of Hα emitting galaxies is highly complete in order to accurately measure the clustering we need to account for two observational effects. Firstly, the Hα flux limit means that our Hα luminosity limit increases with increasing redshift. We could deal with this by defining volume-limited samples however that removes a significant fraction of the sample that is already fairly small for 2PCF measurements. Instead we apply a \( V_{\text{max}} \) weight to be used in pair counting for effectively volume-limited measurements.

For both the pair counts in the clustering measurements and number density measurements, we assign to each galaxy a \( 1/\Delta V_{\text{max}} \) weight, which is the comoving volume within which a galaxy is detectable, given its intrinsic luminosity, our sample flux limit and the redshift cuts applied to a given sub-sample. For each galaxy, we use its Hα luminosity and our sample flux limit to obtain the maximum redshift \( z_{\text{max}} \) at which that galaxy would still be included in the sample. For a sample with a redshift range \( z_1 < z < z_2 \), we define \( z_{\text{lower}} = z_1 \) and \( z_{\text{upper}} = \min \{ z_2, z_{\text{max}} \} \) for a given galaxy and compute the comoving volume \( \Delta V_{\text{max}} \) between \( z_{\text{lower}} \) and \( z_{\text{upper}} \). The \( V_{\text{max}} \) weight is then defined as \( 1/\Delta V_{\text{max}} \).

The second effect concerns an increasing level of incompleteness for galaxies close on the sky as a result of overlapping grism spectra. Such contamination is modelled and corrected for in the 3D-HST pipeline, however severely contaminated spectra result in failed grism redshift measurements (Brammer et al. 2012). Such a scale dependent incompleteness can lead to an underestimate in the correlation function on scales where this is important. To correct for
Table 1. Hα Luminosity-Bin and Luminosity-Threshold Galaxy Samples

| Sample | log $L_{H\alpha,\min}$ | log $L_{H\alpha,\max}$ | log $(L_{H\alpha})$ | $z$ | $N_g$ | $N_g,\text{weighted}$ | $n_g$ | $r_0$ | $\gamma$ | $r_0 \{y_{\text{med}}\}$ |
|--------|----------------------|----------------------|-------------------|-----|-------|----------------------|-------|-----|--------|------------------|
| LB1 | 41.10 | 41.30 | 41.21 | 0.82 | 565 | 3942 | 10.29 | 1.34 ± 0.62 | 1.17 ± 0.09 | 2.03 ± 0.19 |
| LB2 | 41.30 | 41.50 | 41.40 | 0.92 | 1162 | 3492 | 9.12 | 2.15 ± 0.47 | 1.28 ± 0.08 | 2.30 ± 0.22 |
| LB3 | 41.50 | 41.70 | 41.60 | 1.04 | 1380 | 2222 | 5.80 | 2.39 ± 0.45 | 1.35 ± 0.09 | 2.24 ± 0.21 |
| LB4 | 41.70 | 41.90 | 41.80 | 1.14 | 1251 | 1520 | 3.97 | 3.24 ± 0.27 | 1.55 ± 0.09 | 2.52 ± 0.32 |
| LB5 | 41.90 | 42.10 | 41.99 | 1.17 | 792 | 932 | 2.43 | 3.66 ± 0.18 | 1.62 ± 0.06 | 2.64 ± 0.29 |
| LB6 | 42.10 | 42.50 | 42.24 | 1.21 | 559 | 646 | 1.69 | 4.43 ± 0.89 | 1.26 ± 0.11 | 4.70 ± 0.51 |
| LT1 | 41.10 | 42.50 | 41.63 | 0.97 | 5709 | 12757 | 33.30 | 2.82 ± 0.16 | 1.44 ± 0.10 | 2.84 ± 0.15 |
| LT2 | 41.30 | 42.50 | 41.73 | 1.04 | 5144 | 8815 | 23.01 | 3.29 ± 0.15 | 1.51 ± 0.03 | 3.30 ± 0.15 |
| LT3 | 41.50 | 42.50 | 41.86 | 1.11 | 3982 | 5322 | 13.89 | 3.22 ± 0.23 | 1.48 ± 0.05 | 3.22 ± 0.22 |
| LT4 | 41.70 | 42.50 | 41.99 | 1.16 | 2602 | 3097 | 8.09 | 3.72 ± 0.17 | 1.50 ± 0.05 | 3.72 ± 0.16 |
| LT5 | 41.90 | 42.50 | 42.11 | 1.19 | 1351 | 1579 | 4.12 | 4.32 ± 0.36 | 1.50 ± 0.09 | 4.32 ± 0.34 |
| LT6 | 42.10 | 42.50 | 42.24 | 1.21 | 559 | 646 | 1.69 | 4.43 ± 0.89 | 1.26 ± 0.11 | 4.73 ± 0.55 |

From top to bottom are four sets of galaxy samples – the Hα-luminosity-bin (LB) samples, the LB samples at lower and higher redshifts (Lz and Hz), the luminosity-threshold (LT) samples, and the LT samples at lower and higher redshifts (Lz and Hz).

For each sample, shown are the minimum, maximum, and mean Hα luminosity (in units of erg s$^{-1}$), mean redshift, total number of galaxies, the sum of the combined weights, and galaxy number density (in units of $10^{-3}$ h$^3$ Mpc$^{-3}$), correlation length ($r_0$) and power-law index ($\gamma$) from the power-law 2PCF fit. The last column shows the value of $r_0$ with $\gamma$ fixed to the median value.

this effect we reassign the weight of any galaxy that failed to get a grism redshift evenly between all the galaxies with grism redshifts that lie within 100 h$^{-1}$kpc.

In the redshift range of $0.7 < z < 1.5$, we construct both luminosity-bin and luminosity-threshold samples of galaxies. The details of the samples can be found in Table 1. Fig. 2 shows the luminosity and redshift ranges of the luminosity-bin samples. To study the evolution of clustering and compare to previous work, we also construct luminosity-bin and luminosity-threshold samples in two additional redshift range, $0.7 < z < 1.1$ and $1.1 < z < 1.5$. We also construct galaxy samples in fine bins of Hα luminosity and stellar mass to study the joint dependence of clustering on both quantities, which will be presented in Section 4.2.

Finally, for measuring the clustering of galaxies, we create a random catalogue for each sample. The angular distribution of the random points matches the footprints of the five survey fields, taking into account the offset in the dispersion of the grism in the spectral direction, and the masking of regions around bright stars (Fig. 1). For the radial distribution of the random points, we make use of the galaxy sample. For each random point, we randomly draw a galaxy from the sample, obtaining its $z_{\text{lower}}$ and $z_{\text{upper}}$. The random point follows a uniform distribution within the comoving volume between $z_{\text{lower}}$ and $z_{\text{upper}}$, from which its radial comoving distance (hence its redshift) is drawn. The random point is assigned the $1/\Delta V_{\text{max}}$ weight and a unity collision weight.

For each galaxy sample, the number of random points is typically 150 times that of the galaxies.

3 TWO-POINT CORRELATION FUNCTION MEASUREMENTS

The 2PCF is a statistical measure of the excess probability of finding, in this case, two galaxies at a given separation when compared to a uniformly distributed random sample. We quantify the clustering of each sample of 3D-HST galaxies using the projected 2PCF.

First, we measure the redshift-space 2PCF for each sample using the Landy-Szalay estimator (Landy & Szalay 1993), as a function of
Table 2. Galaxy Samples in Bins of Stellar Mass and Hα Luminosity

| Sample | log $M_\star_{min}$ | log $M_\star_{max}$ | (log $M_\star$) | log $L_{H\alpha, min}$ | log $L_{H\alpha, max}$ | log $(L_{H\alpha})$ | (z) | $N_\ell$ | $N_\ell$ weighted | $n_\ell$ | $r_0 [\text{Mpc}]$ |
|--------|-----------------|-----------------|----------------|-----------------|-----------------|----------------|-----|-------|----------------|-------|----------------|
| LB1MB1 | 9.20            | 9.44            | 9.31           | 41.10           | 41.64           | 41.39          | 0.90| 429   | 1296           | 3.38  | $3.81 \pm 0.32$ |
| LB1MB2 | 9.44            | 9.78            | 9.59           | 41.10           | 41.64           | 41.45          | 0.93| 442   | 1035           | 2.70  | $3.99 \pm 0.33$ |
| LB1MB3 | 9.78            | 11.50           | 10.26          | 41.10           | 41.64           | 41.45          | 0.94| 429   | 886            | 2.31  | $5.67 \pm 0.58$ |
| LB2MB1 | 9.20            | 9.55            | 9.38           | 41.64           | 41.90           | 41.75          | 1.17| 425   | 531            | 1.39  | $3.67 \pm 0.32$ |
| LB2MB2 | 9.55            | 9.97            | 9.73           | 41.64           | 41.90           | 41.77          | 1.11| 451   | 549            | 1.43  | $3.25 \pm 0.54$ |
| LB2MB3 | 9.97            | 11.50           | 10.42          | 41.64           | 41.90           | 41.78          | 1.09| 437   | 518            | 1.35  | $5.25 \pm 0.45$ |
| LB3MB1 | 9.20            | 9.84            | 9.58           | 41.90           | 42.50           | 42.08          | 1.20| 402   | 468            | 1.22  | $4.02 \pm 0.61$ |
| LB3MB2 | 9.84            | 10.31           | 10.06          | 41.90           | 42.50           | 42.12          | 1.17| 417   | 486            | 1.27  | $3.56 \pm 0.64$ |
| LB3MB3 | 10.31           | 11.50           | 10.65          | 41.90           | 42.50           | 42.16          | 1.18| 403   | 466            | 1.22  | $6.26 \pm 0.46$ |

From top to bottom are three sets of galaxy samples. The first set of samples are constructed by cuts in Hα luminosity bin (LB), and within each LB sample, galaxies are further divided into stellar mass bin (MB) samples. These samples are fitted with $\gamma = 1.45$. The second set of samples are constructed by cuts in stellar mass (MB), and within each MB sample, galaxies are further divided into Hα luminosity bin (LB) samples. These samples are fitted with $\gamma = 1.31$. The third set of samples are similar to the second set, but with a central Hα luminosity-bin sample and Hα luminosity-threshold samples for the upper and lower luminosity samples. No power laws are fitted to these samples. The 2PCF measurements from the third set are used to constrain the conditional stellar mass and Hα luminosity distribution model. For each sample, shown are the minimum and maximum $M_\star$ (in units of $M_\odot$), and mean log $M_\star$, minimum, maximum, and mean Hα luminosity (in units of erg s$^{-1}$), mean redshift, total number of galaxies, the sum of the combined weights, and galaxy number density (in units of $10^{-3} h^3$Mpc$^{-3}$).

**Figure 2.** Luminosity-bin samples and redshift ranges. Galaxies in different Hα luminosity bins are coded with different colours. We choose samples with equal width in Hα luminosity except the uppermost sample which extends to log $L_{H\alpha} = 42.5$. Effectively volume-limited 2PCF measurements are achieved from these flux-limited samples through proper $V_{max}$ weights.

transverse pair separation $r_p$ and line-of-sight pair separation $r_\pi$.

$$\xi(r_p, r_\pi) = \frac{DD - 2DR + RR}{RR}$$  \hspace{1cm} (2)

where DD, DR, and RR are the data-data, data-random, and random-random pair counts within a given $(r_p, r_\pi)$ pair separation bin, normalized by the corresponding total number of pairs, respectively. We set logarithmic bins for $r_p$ centered at log$[r_p/(h^{-1}\text{Mpc})] = -1$ to 1 with bin width $\Delta \log r_p = 0.20$, and linear bins for $r_\pi$ from 0 to $50 h^{-1}$Mpc with bin width $\Delta r_\pi = 0.05 h^{-1}$Mpc.

The redshift-space 2PCF $\xi(r_p, r_\pi)$ is projected along the line-of-sight direction to obtain the projected 2PCF $w_p$, which reduces the redshift-space distortion effect (Davis & Peebles 1983). We have

$$w_p(r_p) = 2 \int_0^{r_{\pi, max}} \xi(r_p, r_\pi)2 \sum_i \xi(r_p, r_{\pi,i}) \Delta r_\pi,$$  \hspace{1cm} (3)

where $r_{\pi,i}$ is the $i$-th $r_\pi$ bin and we take $r_{\pi, max} = 50 h^{-1}$Mpc.

Specifically, we adopt the code Super W of Theta (SWOT; Coupon et al. 2012) for our 2PCF calculations. As we construct each Hα luminosity-bin sample from the flux-limited survey, the maximum redshift depends on luminosity. Each galaxy or random point is assigned the $1/\Delta V_{max}$ weight and collision weight. Following Xu et al. (2016), we modify the SWOT code so that during the pair counting,
each pair is given the larger $1/\Delta V_{\text{max}}$ weight of the two objects, in addition to the product of the collision weights. That is, the DD, DR, and RR pair counts in equation (2) are all with such weights.

The $1/\Delta V_{\text{max}}$ weight ensures that the clustering measurements are effectively volume-limited (e.g. Xu et al. 2016). For each sample of galaxies, the covariance matrix of the projected 2PCF measurement is estimated from 128 jackknife samples.

If the real-space 2PCF $\xi(\mathbf{r})$ is approximated by a power law,

$$
\xi(r) = \left( \frac{r}{r_0} \right)^{-\gamma},
$$

the clustering strength $r_0$ and the power-law index $\gamma$ can be obtained from the power-law fit to $w_p$.

$$
w_p(r_p) = r_p \left( \frac{r_p}{r_0} \right)^{-\gamma} \frac{\Gamma(\gamma/2 - 1/2) \Gamma(1/2)}{\Gamma(\gamma/2)}. \tag{5}
$$

We will first present the clustering results characterised by the power-law fits, before performing the physically motivated halo modelling.

Finally, given the limited volume of the 3D-HST survey, the integral constraint (IC), resulting from taking the measured galaxy number density to be the global mean, may become appreciable on large scales. We estimate the IC in the real-space 2PCF for each galaxy sample by making use of the random catalogue and a model real-space 2PCF,

$$
C_\xi = \frac{\sum_i \xi^\text{mod}(r_i) \text{RR}(r_i)}{\sum_i \text{RR}(r_i)}, \tag{6}
$$

which leads to the corresponding IC in the projected 2PCF

$$
C = 2 \int_{r_{\pi,\text{max}}}^r \xi_\pi \, d\pi = 2C_\xi r_{\pi,\text{max}}. \tag{7}
$$

In the expression, $\text{RR}(r_i)$ is the number of random-random pairs with separation in the $r_i$ bin (linearly spaced bins). The largest separation bin in our calculation is essentially set by the line-of-sight range ($\sim 3000h^{-1}\text{Mpc}$) of the sample volume.

For the model 2PCF $\xi^\text{mod}(r)$, we use that derived from the power-law fit to $w_p$. We apply a cutoff in $\xi^\text{mod}(r)$ beyond $60h^{-1}\text{Mpc}$ to mimic the trend seen in the matter 2PCF, but we verify that this cutoff has little effect in the derived IC. The IC value is then added to $w_p$ to refine the power-law fit and the IC estimate. The procedure is iterated to reach a converged IC. We find that for each sample the value of IC is of the order of $1h^{-1}\text{Mpc}$. This is about 20–35 per cent of the error bars of the data points at the largest scale ($\sim 8h^{-1}\text{Mpc}$) in our $w_p$ measurements, not a significant effect.

4 DEPENDENCE OF GALAXY CLUSTERING ON H_{\alpha} LUMINOSITY AND STELLAR MASS

In this Section, we present the clustering measurements of H_{\alpha}-emitting galaxies. Following the common practice, we characterise the dependence of clustering strength on H_{\alpha} luminosity and stellar mass based on power-law fits to the projected 2PCFs. In the next section, we will interpret the clustering in a more physical and informative way using the halo model.

4.1 Dependence of Clustering on H_{\alpha} Luminosity

The left and right panels of Fig. 3 show the projected 2PCF measurements of our luminosity-bin and luminosity-threshold samples, respectively. Overall, galaxies with higher H_{\alpha} luminosity are more strongly clustered. For the luminosity-bin samples, there is also evidence for the more H_{\alpha} luminous galaxies having steeper clustering profiles on small scales ($\sim 0.3h^{-1}\text{Mpc}$).

To characterise the clustering amplitude we fit the $w_p$ measurement of each sample with a power law and obtain the correlation length $r_0$ and the power-law index $\gamma$ (equations 4 and 5). To better compare the clustering amplitude among different samples, we also redo the power-law fits with index $\gamma$ fixed to the median value of the $\gamma$ free fits, which is 1.31 and 1.49 for the luminosity-bin and luminosity-threshold samples, respectively. These fits are shown as dashed lines in Fig. 3. While a single power law appears to underestimate the clustering amplitude on small scales ($\sim 0.2h^{-1}\text{Mpc}$) for

---

Figure 3. Projected 2PCFs and power-law fits for luminosity-bin samples (left) and luminosity-threshold samples (right). $L_{\alpha}$ is given in units of erg s$^{-1}$. 

Luminosity-Bin Samples

Luminosity-Threshol

samples, respectively. These fits are shown as dashed lines in Fig. 3. While a single power law appears to underestimate the clustering amplitude on small scales ($\sim 0.2h^{-1}\text{Mpc}$) for
the more luminous luminosity-bin samples, it provides a reasonable
description of the large-scale clustering for all the samples.
In Fig. 4 we plot $r_0$ as a function of Hα luminosity for the
luminosity-bin (left) and luminosity-threshold (right) samples with
both $\gamma$ free and fixed cases. In both figures there is a clear dependence
of $r_0$, hence the clustering amplitude, on Hα luminosity, with
the lowest luminosity samples having the weakest clustering.

### 4.2 Dependence of Clustering on Hα Luminosity and Stellar Mass

It has long been known that galaxy clustering depends strongly on
galaxy stellar mass (e.g. Wake et al. 2011), and there is a correlation
between SFR and stellar mass (i.e. the star formation main sequence;
e.g. Li et al. 2008). Here we find that galaxy clustering depends
on Hα luminosity (a proxy for SFR). Is this dependence driven by
the dependence of clustering on stellar mass and the underlying
correlation between SFR and $M_*$? To answer this question, we further
divide the galaxies into different stellar mass bins and construct
luminosity-dependent galaxy samples within the stellar mass bin.
The construction of the galaxy samples in bins of stellar mass is
illustrated in Fig. 5. Five stellar mass bins (indicated by the five vertical
tones) are formed within the range of 9.2 < $\log(M_*/M_\odot)$ < 11.5, with an approximately equal number of galaxies in each bin. The lower bound, $\log(M_*/M_\odot) = 9.2$, is chosen to ensure the completeness in stellar mass, given the survey limit.

Next, to study the clustering dependence on both stellar mass and
Hα luminosity, we use a quantile binning scheme to create samples
of equal size at both fixed Hα luminosity and fixed $M_*$ shown in the
left and right panels of Fig. 6, respectively. See Table 2.

We follow the same procedure as in Sections 2 and 3 to construct
random catalogues and measure the projected 2PCF for each sample. Again, we follow the above procedure and perform power-law fits first with the index $\gamma$ as a free parameter and then set to the median value.

The resulting $w_p$ measurements for the samples in bins of Hα luminosity and $M_*$ are shown in Fig. 7. For the $M_*$-dependent clustering, we only show the inferred $r_0$ from the power-law fits in Fig. 8 (open symbols in the left panel), which increases with stellar mass as expected.

Looking at the bottom row of Fig. 7 there is no monotonic trend
in the clustering amplitude with Hα luminosity when stellar mass is
fixed. In each case the mid-Hα luminosity bin has the lowest
overall clustering amplitude. Conversely, comparing panels left to
right shows a general trend of increasing clustering amplitude with
increasing stellar mass. Turning to the top row in Fig. 7, where
each panel shows the clustering dependence on stellar mass in a bin
of Hα luminosity, the highest stellar mass sample shows a much
higher clustering amplitude in each panel. Looking left to right there
is a small increase in the clustering amplitude in the highest Hα
luminosity bin over the lower two, which show little difference from
each other.

These trends are confirmed by Fig. 8. The left panel shows the clustering strength ($r_0$) as a function of stellar mass for samples of
differing Hα luminosity. There is a clear trend of stronger clustering
for galaxies of higher stellar mass above $M_* \sim 10^{10}M_\odot$, with a
fairly flat trend at lower masses. There is little variation between the
Figure 6. Left: Construction of stellar mass-bin samples at fixed Hα luminosity. The upper limits of each stellar mass bin are chosen to be the 33rd, 66th, and 100th quantiles of galaxies with log$(M_*/M_\odot) > 9.2$, ensuring that each bin has an approximately equal number of galaxies to have good signal-to-noise ratios across our measurements. Right: Construction of Hα luminosity-bin samples at fixed stellar mass using similar quantile binning.

Figure 7. Top row: Projected 2PCFs for stellar mass-dependent samples in bins of Hα luminosity increasing left to right along with power-law models. The slope of the power law fits are fixed to the median value for the Hα Luminosity binned samples. Similarly the models across the bottom panels have a fixed slope of median from the stellar mass binned samples. The positions of the points on the $r_p$ axis are slightly offset for clarity. In each bin of Hα luminosity the highest stellar mass galaxies show the highest clustering amplitude. There is little clear variation in the clustering amplitude as the Hα luminosity increases from the left to right panel. Bottom row: Same as the top row but now for Hα luminosity-dependent samples in bins of stellar mass. The slope of the power law fits are fixed to the median value for the stellar mass binned samples. There is little variation and no clear trend in the clustering amplitude with Hα luminosity in a given bin in stellar mass, but there is an increase in the clustering amplitude as the stellar mass increases left to right. The dashed light grey horizontal line in each panel is shown to make it easier to compare the clustering amplitude between panels.
Dependence of clustering strength ($r_0$) on stellar mass and H\alpha luminosity. Left: $r_0$ as a function of stellar mass in bins of H\alpha luminosity. The three colours represent three fixed H\alpha luminosity bins with each bin split into the highest, lowest and middle thirds by stellar mass. Irrespective of the H\alpha luminosity the clustering amplitude $r_0$ follows the overall trend with stellar mass, indicating little variation of the clustering with H\alpha luminosity at fixed stellar mass. Right: $r_0$ as a function of H\alpha luminosity in bins of stellar mass. The three colours represent three fixed stellar mass bins with each bin split into the highest, lowest and middle thirds by H\alpha luminosity. Again there is at most a weak trend in the H\alpha luminosity dependent clustering within a stellar mass bin but now there is a clear offset between the stellar mass bins indicating a stronger dependence of clustering amplitude on stellar mass at fixed H\alpha luminosity. For comparison, the open circles show the dependence of $r_0$ on stellar mass (left) and H\alpha luminosity (right) inferred from stellar mass-bin and H\alpha luminosity-bin samples. All samples on the left are fitted with $\gamma = 1.45$. Similarly, all samples on the right are fitted with $\gamma = 1.31$.

It is worth noting that there is a weak V-shaped trend of $r_0$ with H\alpha luminosity in each of the stellar mass bin samples, with the middle H\alpha luminosity bin showing a lower clustering amplitude than the upper and lower luminosity bin. If this trend is real, it suggests that galaxies with both the lowest and highest specific SFRs (sSFR) are more strongly clustered than those in the middle. Such a trend could be in part driven by star-forming satellite galaxies having either reduced or enhanced sSFRs compared to central star-forming galaxies of the same stellar mass. That is, the star formation main sequence of satellite galaxies may have a larger scatter in SFR than that of central galaxies. We provide more insights into this trend in Section 5.3 based on the halo modelling results.

While a power-law characterisation of the 2PCFs allow us to study the overall dependence of galaxy clustering on H\alpha luminosity and stellar mass, a halo-based model will provide us a more informative way to interpret the clustering measurements and to study the galaxy-halo connection. We turn to such a model in the following section.

### 5 HALO MODELLING

In this section, we model the clustering measurements within the framework of halo occupation distribution (HOD). With an assumed cosmology, the properties of dark matter haloes, including their mass function and spatial clustering, are readily known. The HOD specifies the relationship between galaxies and dark matter haloes as a function of halo mass. In particular, it parameterises the probability distribution function of finding galaxies in haloes of a given mass, including the mean occupation function. Together with the halo population, the HOD model enables us to calculate the 2PCFs for a sample of galaxies to compare to observational measurements.

#### 5.1 Model Setups

To model the luminosity-dependent clustering, we parameterise the galaxy-halo relation in terms of the H\alpha luminosity distribution of galaxies as a function of halo mass. The halo occupation function of each sample can then be derived by applying the sample luminosity cuts. With such a parameterisation, we are able to simultaneously model the clustering measurements of all the samples with different luminosity cuts.

Similarly, to model the luminosity and stellar mass dependent clustering, the galaxy-halo relation is parameterised by the joint distribution of H\alpha luminosity and stellar mass as a function of halo mass, which allows simultaneous modelling of all the samples of various luminosity and stellar mass cuts.

In this subsection, we present our models for the above two cases, and in the next two subsections we present the modelling results.
5.1.1 Model for the Luminosity Dependent Clustering

As we study luminosity-dependent clustering with luminosity-bin galaxy samples, it is convenient to parameterise the galaxy-halo relation in terms of the conditional luminosity function (CLF; Yang et al. 2003), which describes the luminosity distribution of galaxies as a function of halo mass. The mean occupation function related to the HOD for a given galaxy sample can then be obtained by integrating the CLF over the luminosity range used to construct the sample.

We follow the CLF parameterisation in Yang et al. (2008) and separate it into contributions from central and satellite galaxies. In haloes of a given mass $M_h$, the central galaxy CLF is described by a log-normal distribution,

$$\Phi_{\text{cen}}(L|M_h) \equiv \frac{d(N_{\text{cen}})}{d\log L} = \frac{A_c}{\sqrt{2\pi} \sigma_c} \exp \left[ -\frac{(\log L - \log L_c)^2}{2\sigma_c^2} \right],$$

where $A_c$, $\sigma_c$, and $L_c$ denote the amplitude, width, and the centre of the log-normal function, respectively, with all possibly depending on halo mass. The amplitude $A_c$, which is the integral of the above expression over all luminosities, represents the fraction of central galaxies that are star-forming (with H$\alpha$ emission). By definition, $A_c$ cannot exceed unity.

The CLF of satellite galaxies in haloes of a given mass $M_h$ is parameterised as a Schechter-like function,

$$\Phi_{\text{sat}}(L|M_h) \equiv \frac{d(N_{\text{sat}})}{dL} = \phi^*_s \left( \frac{L}{L^*} \right)^{\alpha_s} \exp \left[ -\frac{\left( \frac{L}{L^*} \right)^{\gamma_s}}{\sigma_s^2} \right].$$

The form deviates from the Schechter function by having a squared term inside the exponential function, which better describes the empirically determined CLF from a galaxy group catalogue (Yang et al. 2008). Here $\phi^*_s$, $\alpha_s$, and $L^*$ are the normalisation, power-law slope at the low-luminosity end, and the characteristic cutoff luminosity.

With the CLF, for a sample of galaxies with luminosity in the range $L_1 < L < L_2$, the mean occupation functions of central and satellite galaxies in haloes of mass $M_h$ can be computed as

$$\langle N_{\text{cen}}(M_h) \rangle = \int_{L_1}^{L_2} \Phi_{\text{cen}}(L|M_h) d\log L,$$

subject to

$$\frac{A_c}{2} \left[ \text{erf} \left( \frac{\log L_2 - \log L_c}{\sqrt{2} \sigma_c} \right) - \text{erf} \left( \frac{\log L_1 - \log L_c}{\sqrt{2} \sigma_c} \right) \right],$$

and

$$\langle N_{\text{sat}}(M_h) \rangle = \int_{L_1}^{L_2} \Phi_{\text{sat}}(L|M_h) dL,$$

subject to

$$\phi^*_s L^*_s \left[ \frac{\alpha_s + 1}{2} \left( \frac{L_1}{L^*_s} \right)^{\gamma_s} - \frac{\alpha_s + 1}{2} \left( \frac{L_2}{L^*_s} \right)^{\gamma_s} \right].$$

where erf$(x) = (2/\sqrt{\pi}) \int_0^x \exp(-t^2) dt$ is the error function and $\gamma(x, s) = \int_s^\infty t^{-1} \exp(-t) dt$ is the lower incomplete gamma function. The expression in equation (13) is for the case with $\alpha_s > -1$. For $\alpha_s < -1$, the integral in equation (12) is evaluated numerically.

In haloes of a given mass, the above CLF forms are described by parameters $A_c$, $L_c$, $\sigma_c$, $\phi^*_s$, $\alpha_s$, and $\Delta \log L_c$. In our model, each of these parameters has a halo mass dependence and redshift dependence. As the median redshifts of our samples are similar, we neglect the redshift dependence in this study. Also given the relatively narrow luminosity range of our samples, we only keep the halo mass dependence for $A_c$, $L_c$, and $\phi^*_s$, assuming they all have a power-law dependence on halo mass. Each of them is parameterised by the value at a pivot mass (chosen to be $10^{11} h^{-1} M_\odot$; with the corresponding parameter labelled with a subscript ‘p’) and the power-law index of the mass dependence. In total, we have 9 parameters to simultaneously model the clustering measurements of all the luminosity-bins samples, which are $A_{c,p}$, $\gamma_A = d \log A_c/d \log M_h$, $L_{c,p}$, $\gamma_L = d \log L_c/d \log M_h$, $\sigma_c$, $\phi^*_{s,p}$, $\gamma_s = d \log \phi^*_s/d \log M_h$, $\alpha_s$, and $\Delta \log L_c$.

Given a set of the 9 CLF parameters and the luminosity cuts ($L_1$, $L_2$) of a galaxy sample, we compute the mean occupation functions of central and satellite galaxies, using equations (10) and (12). Then, with the mean occupation functions, we can calculate the model number density

$$n_g = \int_0^\infty \langle N(M_h) \rangle \frac{dn}{dM_h} dM_h,$$

with $\langle N(M_h) \rangle = \langle N_{\text{cen}}(M_h) \rangle + \langle N_{\text{sat}}(M_h) \rangle$ and $dn/dM_h$ the halo mass function. We follow the method developed in Zheng (2004) and improved in Tinker et al. (2005) to calculate the 2PCF of the galaxy sample. In our work, we adopt the halo definition that the mean density of haloes is 200 times that of the background universe. The number of satellite galaxies in haloes of fixed mass is assumed to follow the Poisson distribution with the mean given by equation (12). The spatial distribution of satellite galaxies within haloes is assumed to be the same as that of the dark matter, following the Navarro-Frenk-White profile (Navarro et al. 1996). The concentration parameter of the profile, $c(M_h) = [c_0/(1 + z)](M_h/M_\odot)^\beta$, with $c_0 = 11$, $\beta = -0.13$, and $M_\odot = 3.79 \times 10^{12} h^{-1} M_\odot$ (the nonlinear mass scale at $z = 0$ for the adopted cosmology).

5.1.2 Model for the Luminosity and Stellar Mass Dependent Clustering

To model the clustering of luminosity-bin samples with a cut in stellar mass, we adopt a formalism similar to that in Xu et al. (2018). To model the dependence of galaxy clustering on luminosity and colour, Xu et al. (2018) develop a global parameterisation of the colour and luminosity distribution as a function of halo mass, namely the conditional colour-magnitude distribution (CCMD), and jointly model the clustering of a large number of galaxy samples defined by cuts in colour and luminosity. In this work, we parameterise the joint distribution of stellar mass and H$\alpha$ luminosity as a function of halo mass, separated into that for central and that for satellite galaxies. This extends the CLF by adding one more dimension.

For the ease of presenting the formalism, we use ‘$x$’ to represent the logarithmic of stellar mass, $x \equiv \log M_*$, and ‘$y$’ the logarithmic of H$\alpha$ luminosity, $y \equiv \log L_{H\alpha}$. For central galaxies, the stellar mass – H$\alpha$ luminosity distribution inside haloes of fixed mass, i.e. the conditional $M_* - L_{H\alpha}$ distribution, is parameterised as a 2D Gaussian distribution,

$$\frac{d^2\langle N_{\text{cen}}(M_h) \rangle}{dx \ dy} = \frac{A_c}{2\pi \sigma_x \sigma_y \sqrt{1 - \rho^2}} \exp \left[ -\frac{Z^2}{2(1 - \rho^2)} \right],$$

with

$$Z^2 = \frac{(x - \mu_x)^2}{\sigma_x^2} + \frac{(y - \mu_y)^2}{\sigma_y^2},$$

$$\rho = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y},$$

and

$$\mu_x = -\frac{1}{\sigma_x^2} \int_{-\infty}^{\infty} \langle N_{\text{cen}}(M_h) \rangle \frac{dn}{dM_h} dM_h,$$

$$\mu_y = -\frac{1}{\sigma_y^2} \int_{-\infty}^{\infty} \langle N_{\text{cen}}(M_h) \rangle \frac{dn}{dM_h} dM_h,$$

$$\sigma_x^2 = \left[ \int_{-\infty}^{\infty} \langle N_{\text{cen}}(M_h) \rangle \frac{dn}{dM_h} dM_h \right] \int_{-\infty}^{\infty} (x - \mu_x)^2 \frac{dn}{dM_h} dM_h,$$

$$\sigma_y^2 = \left[ \int_{-\infty}^{\infty} \langle N_{\text{cen}}(M_h) \rangle \frac{dn}{dM_h} dM_h \right] \int_{-\infty}^{\infty} (y - \mu_y)^2 \frac{dn}{dM_h} dM_h,$$

$$\sigma_{xy} = \left[ \int_{-\infty}^{\infty} \langle N_{\text{cen}}(M_h) \rangle \frac{dn}{dM_h} dM_h \right] \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) \frac{dn}{dM_h} dM_h.$$
where \( M_{h,p} = 10^{11} h^{-1} M_\odot \) is the pivot halo mass and the quantity \( A_{c,p} \) is the value at the pivot halo mass. In this work, we neglect the halo mass dependence and fix \( \gamma_A = 0 \). For the halo mass dependence of the median stellar mass and \( H_\alpha \) luminosity, we follow Xu et al. (2018) to parameterise each to a power law with exponential cutoff at low halo mass. As an example, the median stellar mass has the following form

\[
M_{s,m} = M_{s,t} \left( \frac{M_h}{M_{h,t}} \right)^{\alpha_s} \exp \left[ - \frac{M_h}{M_{h,t}} + 1 \right],
\]

where \( M_{h,t} \) denotes a transition halo mass scale, \( M_{s,t} \) is the median stellar mass at this halo mass scale, and \( \alpha_s \) is the power-law index. It reduces to

\[
\mu_s = \mu_{s,t} + \sigma_s (\log M_h - \log M_{h,t}) + (-M_{h,t}/M_h + 1)/\ln 10,
\]

with \( \mu_{s,t} = \log M_{s,t} \). Similarly, for the median \( H_\alpha \) luminosity, we have the form

\[
\mu_y = \mu_{y,t} + \sigma_y (\log M_h - \log M_{y,t}) + (-M_{y,t}/M_h + 1)/\ln 10.
\]

In total, there are nine free parameters for the conditional \( M_s-L_{H_\alpha} \) distribution of central galaxies. \( A_{c,p}, \mu_{s,t}, M_{s,t}, \alpha_s, \mu_{y,t}, \sigma_y, \sigma_x, \text{and } \sigma_y \).

For satellite galaxies, the conditional stellar mass function (stellar mass function at fixed halo mass) is described by a modified Schechter function (e.g. Yang et al. 2008; Xu et al. 2018),

\[
d(N_{\text{sat}}(M_h))/d\log M_s = \phi_s \left( \frac{M_s}{M_{s,\text{sat}}} \right)^{\alpha_s+1} \exp \left[ \frac{\left( M_{s,\text{sat}} - M_s \right)^2}{2\sigma_s^2} \right],
\]

where \( \phi_s \) is the normalisation, \( \alpha_s \) is the faint-end slope, and \( M_{s,\text{sat}} \) is the characteristic stellar mass. To obtain the parameterisation of the conditional \( M_s-L_{H_\alpha} \) distribution for satellites, we adopt the above form for the stellar mass distribution and assume that the (logarithmic) \( H_\alpha \) luminosity follows a Gaussian distribution at fixed stellar mass (Xu et al. 2018),

\[
\frac{d^2(N_{\text{sat}}(M_h))}{dx \, dy} = \frac{1}{\sqrt{2\pi} \sigma_{y,\text{sat}}^2} \exp \left[ -\frac{(y - \mu_{y,\text{sat}})^2}{2\sigma_{y,\text{sat}}^2} \right] \times \phi_s e^{(\alpha_s+1)(x-x_s)} \exp \left[ -10^{2(x-x_s)} \right],
\]

where \( x_s \equiv \log M_{s,\text{sat}} \). We parameterise \( x_s \) through the stellar mass gap, the difference between the (logarithmic) central and satellite characteristic stellar mass,

\[
x_s = \mu_s - \Delta_{cs}.
\]

The amplitude \( \phi_s \) has a dependence on halo mass

\[
\log \phi_s = \log \phi_{s,p} + \gamma_{\phi}(\log M_h - \log M_{h,p}).
\]

The median \( H_\alpha \) luminosity has a dependence on stellar mass, motivated by the relation on the star formation main sequence,

\[
\mu_y = \mu_{y,p} + \gamma_y(x-x_p),
\]

where the pivot stellar mass is taken to be \( 10^{10} M_\odot \), i.e. \( x_p = 10 \).

In the above, we choose to present the overall framework so that it can be applied to future surveys. In this work, given the sample size and the uncertainties in the measurements, the satellite occupation distribution cannot be tightly constrained. Therefore, we apply strong priors broadly motivated by previous work. For example, we set the gap parameter \( \Delta_{cs} \) to zero and require the conditional stellar mass function of satellites to be higher than 0.1 at \( \log(M_s/M_\odot) = 10 \) in \( \log(M_h/(h^{-1} M_\odot)) = 12 \) haloes (e.g. Leauthaud et al. 2012; Lim et al. 2017). We also require the satellite star formation main sequence

---

**Figure 9.** Illustration of halo modelling results for the luminosity dependent clustering. Top: projected 2PCF measurements (points) and those from the best-fitting model (curves), colour-coded by \( H_\alpha \) luminosity. For clarity, the 2PCFs are each staggered by 0.5 dex. The model reproduces the observed clustering with good accuracy. Bottom: mean occupation function of each sample, inferred from the best-fitting model. The bump represents that of central galaxies, and the power-law-like curve is that of satellite galaxies. Galaxies with increasing \( H_\alpha \) luminosity live in dark matter halos of increasing mass.

Here \( A_c \) is the fraction of haloes at the given mass occupied by star-forming galaxies, \( \mu_x \) and \( \mu_y \) are the mean logarithmic stellar mass and \( H_\alpha \) luminosity in these haloes, \( \sigma_x \) and \( \sigma_y \) are the corresponding standard deviations, and \( \rho (\text{Cov}(x,y)) \) is the coefficient (covariance) of the correlation between logarithmic stellar mass and \( H_\alpha \) luminosity. While all these parameters can depend on halo mass, we find that some of them are not well constrained. Given the limited range of halo mass expected for the samples we consider, we assume no halo mass dependence for \( \sigma_x \) and \( \sigma_y \) and fix \( \rho = 0 \).

The occupation fraction is assumed to have a power-law dependence on halo mass. That is,

\[
\log A_c = \log A_{c,p} + \gamma_A (\log M_h - \log M_{h,p}),
\]

10. Clontz, Wake, & Zheng

MNRA 000, 1–19 (2022)
to be close to the central one, with $\mu_{y,\text{sat}}$ to be within $0.5\sigma_{y,\text{cen}}$ of $\mu_{y,\text{cen}}$ in haloes of $\log(M_h/(h^{-1}M_\odot)) = 12$. With much larger samples from future surveys, such priors would not be necessary and the parameters would be constrained by the data. In total, we have six parameters to describe the conditional $M_h-L_{\text{H}_\alpha}$ distribution of satellite galaxies, $\phi_{s,p}, \gamma_{d}, \alpha_{b,\mu_{y,p}, \text{sat}}, \gamma_{s,p}$, and $\sigma_{y,\text{sat}}$.

With the above parameterisation of the conditional $M_h-L_{\text{H}_\alpha}$ distribution, for a sample defined by cuts in stellar mass and H$\alpha$ luminosity, we can integrate the distribution to obtain the mean occupation functions of central and satellite galaxies for this sample, which can then be used to compute the model 2PCF. As in Xu et al. (2018), this global parameterisation allows us to simultaneously model the 2PCFs of a number of galaxy samples constructed with different stellar mass and H$\alpha$ luminosity cuts.

With the above CLF (Section 5.1.1) or conditional $M_h-L_{\text{H}_\alpha}$ distribution (Section 5.1.2) model setups, for a given set of parameters, we calculate the model prediction for the projected 2PCF $w_p$ and galaxy number density for each galaxy sample and form the $\chi^2$ as

$$\chi^2 = (w_p - w_p^\text{mod})^T C^{-1} (w_p - w_p^\text{mod}) + (n_g - n_g^\text{mod})^T C_n^{-1} (n_g - n_g^\text{mod}),$$

(27)

where $w_p$ and $n_g$ are the vectors of the projected 2PCFs and number densities for either all the luminosity-dependent samples or all the luminosity-and-stellar-mass-dependent samples with the measured values denoted with a superscript ’$\dagger$’. $C$ is the full covariance matrix of the projected 2PCFs, and $C_n$ is the covariance matrix of the number densities. As we have limited area to estimate the 2PCF covariance matrix of each sample using the jackknife method, we neglect the covariance between different samples. For the number density covariance matrix, a 10% uncertainty is assumed for each sample. The final covariance matrix is scaled by $(N - N_\text{d})/(N - 1)$ to account for the mean bias in inverting the matrix (Hartlap et al. 2007), where $N = 128$ is the number of jackknife samples and $N_\text{d}$ is the number of data points. With the above $\chi^2$, the likelihood of the model is proportional to $\exp(-\chi^2/2)$, and a Markov Chain Monte Carlo (MCMC) method is employed to explore the parameter space and to obtain constraints on the model.

5.2 Modelling Results for the Luminosity Dependent Clustering

The constraints on the CLF parameters from simultaneously modelling all the luminosity-bin samples are shown in Fig. A2. In this subsection we focus on the halo occupation functions and several derived quantities.

In the top panel of Fig. 9, the projected 2PCFs (curves) from the best-fitting model are plotted with the measurements (points). For clarity, offsets are added to both the model curves and data points. The model provides a good description of the measurements, with $\chi^2 \approx 21.3$ for 57 degrees of freedom (10 $w_p$ points and 1 number density for each of the six luminosity-bin samples, minus 9 free parameters).

The mean occupation functions derived from the best-fitting model are shown in the bottom panel of Fig. 9. For each sample, the bump at low mass is the mean occupation function of central galaxies. Roughly speaking, central galaxies in our samples reside in haloes of masses $\sim 10^{11.5-12}h^{-1}M_\odot$. In this best-fitting model, nearly all the halos in this mass range are occupied by star-forming galaxies with H$\alpha$ emission. In fact, from the parameter constraints in Fig. A2, the occupation fraction is mostly constrained to be order unity. The amplitude parameter of the central galaxy CLF, $A_{c,p}$ at the pivot mass $10^{11.5}h^{-1}M_\odot$ is about unity or larger than unity. Together with the mass dependence parameters, $\gamma_{c} \equiv \log(A_c)/\log(M_h)$, we can derive the occupation fraction $A_c$ of central galaxies in haloes of any given mass. When this fraction is larger than unity, we set it to be unity in our model. Therefore, the preferred larger-than-unity value of $A_{c,p}$ means unity occupation of central galaxies in low mass haloes, and the data is consistent with such a nearly unity occupation. Only in high mass haloes, the occupation fraction starts to drop.

We note that in Fig. 9 the mean central galaxy occupation function (the bump) for each sample appears to have relatively sharp edges, resulting from the small scatter ($\sigma_{c}$) in central mass at fixed halo mass for the best-fitting HOD used for the illustration. In fact, $\sigma_{c}$ is loosely constrained, with the $2\sigma$ range being $\sim 0.0-0.49$ dex (Fig. A2), which is not reflected in the illustration. The overall trend of the dependence of the occupation on H$\alpha$ luminosity and halo mass, however, is not affected by such uncertainties.

In the bottom panel of Fig. 9, the mean occupation function for satellite galaxies in each sample can be approximately described by a steep power law. It suggests that star-forming satellite galaxies can be found in massive halos. The constraints mainly come from the small-scale clustering, in the one-halo regime. If there were no satellite galaxies, there would be no inter-halo galaxy pairs and the real-space 2PCF would drop to zero below Mpc scale. As a consequence, the projected 2PCF $w_p$ would become flattened below such a scale. The data tend to have a $w_p$ profile increasing toward the smallest scale and a fraction of galaxies can be satellites. From parameter constraints in Fig. A2, we can see that the overall constraints on satellite occupation function is loose — the amplitude $\phi_{s,p}$ at the pivot mass $10^{11}h^{-1}M_\odot$ and the luminosity density gap $\Delta L_{\text{cs}}$ both vary by more than one dex.

The posteriors for derived quantities can be inferred from the MCMC chain. In Fig. 10, the constraints on four derived quantities are shown. The left-most panel is the large-scale galaxy bias factor $b_g$ as a function of H$\alpha$ luminosity, where $b_g$ is the halo bias $b_h$ properly weighted by the mean occupation and halo mass function,

$$b_g = \frac{1}{n_g} \int_0^{\infty} b_h(M_h) \langle N(M_h) \rangle \frac{dn}{dM_h} dM_h.$$  

(28)

Approximately, the galaxy bias factor appears to increase linearly with log $L_{\text{H}_\alpha}$. As the real-space 2PCF scales as $b_g^2$, under the power-law approximation we can connect $b_g$ to the clustering strength $r_0$, $b_g \propto r_0^{\gamma_{b}}$. Our derived $b_g$ trend is consistent with the $r_0$ trend seen in the left panel of Fig. 4.

The second panel in Fig. 10 shows the luminosity dependence of the median mass of haloes hosting central galaxies. The median mass $M_{h,\text{cen}}$ is derived through

$$\int_0^{M_{h,\text{cen}}} \langle N(\text{cen})(M_h) \rangle \frac{dn}{dM_h} dM_h = \frac{1}{2} \int_0^{\infty} \langle N(\text{cen})(M_h) \rangle \frac{dn}{dM_h} dM_h.$$  

(29)

From the dependence, the implied halo mass-H$\alpha$ luminosity relation of central galaxies approximately follows a power law, $M_h \propto L_{\text{H}_\alpha}^{\gamma_{M}}$. The third panel shows the median mass of haloes hosting satellite galaxies, calculated similarly. It is around $10^{13.5-14.0}h^{-1}M_\odot$, but the constraint is not as tight as that for central galaxies, with an uncertainty about 0.5–0.8 dex, showing only a weak luminosity dependence.

The right-most panel in Fig. 10 plots the luminosity dependent fraction of galaxies being satellites. As with the median mass of satellite hosting haloes, the constraint on the satellite fraction is loose (from nearly zero per cent to about five per cent), and it is consistent with no dependence on H$\alpha$ luminosity.
5.3 Modelling Results for the Luminosity and Stellar Mass Dependent Clustering

The formalism of the conditional \( M_\star - L_{H\alpha} \) distribution (Section 5.1.2) can be applied to model the clustering of galaxies samples defined by arbitrary cuts in \( M_\star \) and \( L_{H\alpha} \). To have samples across a large range in stellar mass and \( H\alpha \) luminosity while retaining a reasonable signal-to-noise ratio, we apply cuts to create three stellar mass bins and in each stellar mass bin to construct three \( H\alpha \) luminosity dependent samples. The \( (M_\star/M_\odot) \) range for the three stellar mass bins are 9.1 < \( M_\star < 9.6 \), 9.6 < \( M_\star < 10.1 \), and 10.1 < \( M_\star < 11.5 \), respectively. In each stellar mass bin, the three \( H\alpha \)-luminosity dependent samples have the following ranges of \( \log [L_{H\alpha}/(\text{erg s}^{-1})] \): \( \log L_{H\alpha} < 41.6 \), \( 41.6 < \log L_{H\alpha} < 41.9 \), and \( \log L_{H\alpha} > 41.9 \). See the last set of samples in Table 2.

The \( w_p \) measurements and the number densities of the nine samples are modelled simultaneously with the conditional \( M_\star - L_{H\alpha} \) distribution formalism. Fig. 11 shows the measurements and the model fits. While the model fits appear to be slightly lower in amplitude than the measurements, given the uncertainties in the measurements, the model provides reasonable fits, \( \chi^2 = 22.1 \) for 84 degrees of freedom (99 data points and minus 15 model parameters). The measured 2PCFs show a clear dependence on stellar mass, with stronger clustering for samples of higher stellar mass. For galaxies in each stellar mass bin, the measured 2PCFs show little dependence on \( H\alpha \) luminosity. These trends are well captured by the model.

In Fig. 12, we illustrate the mean occupation functions of central and satellite galaxies in these samples, derived from the best-fitting model of the conditional \( M_\star - L_{H\alpha} \) distribution. In a given stellar mass bin, central galaxies of different \( H\alpha \) luminosity occupy haloes in a similar mass range. As stellar mass increases, the mass scale of hosting haloes also increases. That is, the model shows a tight correlation between central galaxy stellar mass and halo mass, while the correlation between \( H\alpha \) luminosity and halo mass is relatively weak. This can also be seen from the scatter in the central stellar mass and \( H\alpha \) luminosity at fixed halo mass, represented by the parameters \( \sigma_x \) and \( \sigma_y \) in equation (15), with the median value for the former at the level of \( \sigma_x \sim 0.03 \) and the latter at the level of \( \sigma_y \sim 0.35 \) (see Fig. A3). For each sample, the mean satellite occupation function approximately follows a power law, with the satellite fraction around 10–20 per cent but not well constrained (see Table A3). As the 2PCFs on large scales are dominated by contributions from central galaxies, the tight correlation between stellar mass and halo mass explains the main trends seen in Fig. 11.

As with Fig. 9, we caution that the uncertainties in the HOD constraints are not reflected in Fig. 12, which, when considered, would lead to less sharp edges in the mean occupation functions for central galaxies. Nevertheless, the HOD modelling enables us to infer the overall trend of the occupation with stellar mass and \( H\alpha \) luminosity.

To further examine the implied correlation of stellar mass and \( H\alpha \) luminosity with halo mass from the model, in Fig. 13 we show the median galaxy bias and the median halo mass as a function of galaxy stellar mass and \( H\alpha \) luminosity, inferred from all the models in the MCMC chain. For galaxy bias (top panel), except for the low stellar mass end, the contours are more vertical than horizontal, implying that galaxy clustering has a stronger dependence on stellar mass than on \( H\alpha \) luminosity. This is another representation of the trends seen in Fig. 7, Fig. 8 (right panel), and Fig. 11. In the middle panel of Fig. 13, the contours are nearly vertical, demonstrating that the tight correlation is between central galaxy stellar mass (rather than \( H\alpha \) luminosity) and halo mass.

Interestingly, although the contours in the top panel of Fig. 13 are much more vertical than horizontal, they do show a small amount of curvature. That means, at fixed stellar mass the galaxies with the lowest and highest \( H\alpha \) luminosities have a slightly higher galaxy bias. This trend is entirely consistent with the trends shown in the right panel of Fig. 8, where \( r_0 \) shows a V-shaped dependence on \( H\alpha \) luminosity. Since there is virtually no such curvature for central galaxy median halo mass seen in the middle panel of Fig. 13, we can conclude that the dependence of bias on \( H\alpha \) luminosity at fixed stellar mass is most likely caused by the satellite galaxy population. If at fixed stellar mass satellite galaxies had a broader distribution in \( H\alpha \) luminosity, and hence sSFR, than equivalent centrals, we would expect just such a result. Indeed, with our parameterisation, the modelling results show such a trend – from Fig. A3, we have the scatter in \( H\alpha \) luminosity \( \sigma_y \sim 0.35 \) dex (central galaxies) and \( \sigma_y \sim 0.63 \) dex (satellites). Further support comes from the bottom panel of Fig. 13, which shows the satellite fraction in bins of stellar mass and \( H\alpha \) luminosity from the model fit. At fixed stellar mass, the satellite fraction increases at both high and low \( H\alpha \) luminosity, as shown in the bottom panel (also see Table A3). Since satellites are typically hosted by more massive halos than centrals of similar stellar mass, the clustering amplitude (hence galaxy bias) correspondingly increases. As expected, the HOD modelling provides a more informative interpretation of the clustering trend than can be inferred from the power-law fits alone. We caution the reader that the uncertainties on the best fit HOD parameters (Fig. A3) for the satellites, and hence on the satellite fractions (Table A3), are large and that the median values of the derived satellite fractions vary significantly between different HOD model formalisms (Tables A2 and A3), so these trends are tentative.
at best. We highlight them here only because they are consistent with, and provide an explanation for, the V-shaped dependence of $r_0$ on Hα luminosity at fixed stellar mass that we observe.

In Fig. 14, we show the median stellar mass (bottom) and median Hα luminosity (top) of central galaxies as a function of halo mass from the model, following equations (20) and (21). Both are parameterised as a power law toward the high halo mass end and an exponential cutoff toward the low halo mass end. The data (2PCFs and galaxy number densities) lead to good constraints on both relations in haloes around $10^{12} h^{-1} M_\odot$. The constraints become loose at the low and high mass end, as a result of sample limitations. There is a tendency that central Hα luminosity in high mass haloes levels off, indicating less active star formation in central galaxies of high mass haloes. However, the loose constraints prevent us from drawing any robust conclusion.

The solid black curves in Fig. 14 are the predicted relations from the EAGLE hydrodynamic galaxy formation simulation with a box size of 100 Mpc (comoving), i.e. the run named Ref-L0100N1504 (Schaye et al. 2015; McAlpine et al. 2016). For the comparison, we use the halo mass in the EAGLE simulation so that haloes have a mean density 200 times that of the background universe, consistent with what we adopt in this work. We convert the SFR in the simulation to Hα luminosity using the relation in Kennicutt & Evans (2012) and then apply the stellar-mass-dependent dust extinction given by Garn & Best (2010). We note that both the EAGLE simulation and the 3D-HST stellar mass calculation adopt the Chabrier (2003) initial mass function (IMF). For the relation between median stellar mass and halo mass, the EAGLE curve falls right within the 1σ range of our HOD model constraints. Since the EAGLE simulation calibrates its feedback parameters to reproduce the $z \sim 0.1$ stellar mass function, the excellent agreement in the $z \sim 1$ central galaxy stellar mass–halo mass relation with the HOD modelling result is encouraging. For the relation between central galaxy Hα luminosity and halo mass, the EAGLE result has a trend similar to that in the HOD model constraint, but the curve lies slightly out of the 1σ range of the model constraints. In haloes of $\sim 10^{12} h^{-1} M_\odot$, the EAGLE prediction is about 0.15 dex lower than the central constraint of the HOD model. This is in line with the result that the EAGLE simulation underpredicts the $z \sim 1$ SFR function at SFR$\sim10 M_\odot$ yr$^{-1}$ and the cosmic SFR density (Katsianis et al. 2017), when compared with observationally

---

Figure 12. Illustration of the mean occupation function of central galaxies (solid) and satellite galaxies (dashed) as a function of stellar mass and Hα luminosity. We show the prediction from the best-fitting model from jointly modelling the stellar mass and Hα luminosity dependent galaxy clustering based on the conditional $M_\star - L_{H\alpha}$ distribution formalism. From top to bottom panels, the stellar mass of the samples increases, and in each panel colour-coded is the Hα luminosity. Here stellar mass of central galaxies show a tighter correlation with halo mass than Hα luminosity.

Figure 11. Measurements (points) of the stellar mass and Hα luminosity dependent projected 2PCFs and fits (curves) from the conditional $M_\star - L_{H\alpha}$ distribution model. The three panels are for three different stellar mass bins. In each panel the dependence of the 2PCF on Hα luminosity is shown, and the measurements from different Hα luminosity samples are shifted slightly in the horizontal direction for clarity. The clustering shows a clear trend that the amplitude increases with increasing stellar mass, but at fixed stellar mass there is only a weak dependence on Hα luminosity. Although the model predicts slightly lower amplitudes, it provides good fits ($\chi^2 \approx 22.1$ for 84 degrees of freedom) and reproduces the major trends.
Figure 13. Median galaxy bias (top), median mass of haloes hosting central galaxies (middle), and median satellite fraction (bottom), as a function of galaxy stellar mass and Hα luminosity, derived based on the chains from jointly modelling stellar mass and Hα luminosity dependent galaxy clustering based on the conditional $M_\star - L_{H\alpha}$ distribution formalism. In the top panel, except for the lowest stellar mass, the contours are more vertical than horizontal, implying stronger dependence on stellar mass than on Hα luminosity. In the middle panel, the contours are nearly vertical, indicating that for central galaxies stellar mass is more correlated with halo mass than Hα luminosity. The bottom panel shows that at fixed stellar mass satellite fraction is high at both low and high Hα luminosity, which explains the curvature seen in the contours of galaxy bias in the top panel (as well as in the clustering amplitude in the right panel of Fig. 8).

Figure 14. Relation between median Hα luminosity of central galaxies and halo mass (top) and that between median stellar mass of central galaxies and halo mass (bottom). The blue shaded regions are the $1\sigma$ and $2\sigma$ constraints from modelling the stellar mass and Hα luminosity dependent clustering with the conditional $M_\star - L_{H\alpha}$ distribution model, while the red shaded regions in the top panel are those from modelling the Hα luminosity dependent clustering with the CLF model. In each panel, the solid curve is the relation predicted by the EAGLE hydrodynamic galaxy formation model. See the text for detail.

derived values (e.g. from Hα). The likely cause is that the supernova feedback adopted in the EAGLE simulation is too strong (Katsianis et al. 2017). Overall, the HOD modelling results provide useful tests to the galaxy formation model. The broad agreement with the results from the EAGLE simulation, on the other hand, supports our model parameterisation.

The red shaded region in the top panel of Fig. 14 shows the constraints on the dependence of central galaxy Hα luminosity on halo mass from the CLF-based model (Sections 5.1.1 and 5.2). At the low halo mass end, the CLF constraints agree with those from the conditional $M_\star - L_{H\alpha}$ distribution model. Although the uncertainty in the constraints becomes large in high mass haloes, the CLF model tends to have a higher Hα luminosity than the conditional $M_\star - L_{H\alpha}$ distribution model. We note that this is not necessarily a fair comparison—we show here the luminosity only in halos hosting star-forming galaxies. Since the CLF model (red) has a relatively strong dependence of
occupation fraction of star-forming galaxies on halo mass (e.g. with the power-law index of $\sim -2.09$; see Table A1 and Fig. A2), in high mass haloes (e.g. $M_h \sim 10^{12.5} h^{-1} M_{\odot}$) there is a large fraction of haloes with no star-forming galaxies occupied at the centre. If we were to compare the H\alpha luminosity averaged over all haloes (with or without central star-forming galaxies) of fixed mass, we expect to see a better overlap between the red and blue shaded region at the high halo mass end. In addition, the origin of the difference also lies in the difference in the model parameterisation. In the CLF model, central galaxy H\alpha luminosity is assumed to have a power-law dependence on halo mass, while in the conditional $M_* - L_{H\alpha}$ distribution model there is more flexibility to effectively allow the power-law index to change with halo mass. In principle, we could make the CLF model parameterisation more flexible. Since the CLF model already provides good fits to the H\alpha luminosity dependent clustering, we leave it in its current form to show the model degeneracy. With data sets much larger than used here, however, a more flexible form is necessary and we also advocate modelling galaxy clustering within the formalism like the conditional $M_* - L_{H\alpha}$ distribution to efficiently explore the galaxy-halo connection.

6 DISCUSSION

In this work, we construct samples of H\alpha emitting galaxies with $0.7 < z < 1.5$ based on the grism spectra from the 3D-HST survey, study the H\alpha luminosity and stellar mass dependent clustering, and model those clustering measurements using the CLF model and the conditional $M_* - L_{H\alpha}$ distribution model within the HOD formalism. Looking both directly at the measurements and the resulting halo model fits, we see a number of clear results emerging. There is a strong dependence of the clustering amplitude of galaxies on their H\alpha luminosity with more luminous galaxies being more strongly clustered. However, this trend is largely being driven by the correlation of H\alpha luminosity with stellar mass and the well known relationship between stellar mass and clustering amplitude (e.g. Wake et al. 2011; Skibba et al. 2015). The lack of much residual correlation of clustering amplitude with H\alpha luminosity when stellar mass is fixed is clearly demonstrated in Fig. 8 with similar results for the galaxy bias and central halo mass revealed by the halo modelling in Fig. 13.

We do see evidence that there may be a small residual V-shaped H\alpha luminosity dependence to the clustering amplitude when stellar mass is fixed, such that the lowest and highest H\alpha luminosity galaxies in any given stellar mass bin (i.e. highest and lowest sSFR) are slightly more clustered than those in midrange. We do not see any evidence of such a trend when we look at how the typical halo mass of central galaxies depends on stellar mass and H\alpha luminosity in our best fit halo model, with stellar mass being utterly dominant (middle panel of Fig. 13). This implies that this residual V-shaped H\alpha luminosity dependence to the clustering amplitude is being driven by the satellite galaxy population. Such a trend would occur if at fixed stellar mass the satellite fraction is higher at the highest and lowest H\alpha luminosities and hence sSFRs. Our halo modelling provides some indication of the lowest satellite fractions always falling in the central H\alpha luminosity bin when stellar mass is fixed (bottom panel of Fig. 13 and Table A3). This dependence of the satellite fraction on H\alpha luminosity in the best fitting model arises as a result of a broader scatter in H\alpha luminosity at fixed halo mass for star-forming satellites ($\sigma_{\gamma_{\text{sat}}}$) than for star-forming central galaxies. Physically this makes sense as satellite galaxies may experience an initial enhancement of star formation on infall into a cluster (Vulcani et al. 2018, e.g.) followed by a reduction as their gas supply is removed (see Cortese et al. 2021, for a review), leading to a wider range in sSFRs than equivalent central galaxies. With these tantalising results, it is important to note that our constraints on the satellite population are weak at best from the halo modelling, but it is noteworthy that they nicely explain the trend seen in basic clustering amplitude measurements in Fig. 8.

Throughout this work we have treated our sample of star-forming galaxies with $0.7 < z < 1.5$ as a single population and ignored any potential redshift evolution in their properties over this period.\footnote{We do include the redshift evolution in the dark matter halo properties in our halo model, fitting at the mean redshift of each galaxy sample.} We do know that the H\alpha luminosity function is evolving (Sobral et al. 2013) and that there is a small amount of evolution in the star formation main sequence over this redshift range (Whitaker et al. 2014; Schreiber et al. 2015). In Fig. 15 we show measurements of the relations between $r_0$ and H\alpha luminosity in luminosity-threshold (left) and luminosity-bin (middle) samples (Table 1) made exactly as before but split into two redshift ranges, $0.7 < z < 1.1$ and $1.1 < z < 1.5$. While the results at both redshifts are similar, there is an offset such that the lower redshift galaxies have a higher clustering amplitude at the same H\alpha luminosity, except for the highest H\alpha luminosities. Much of the evolution in the H\alpha luminosity function over this redshift range can be explained by pure luminosity evolution, i.e. evolution of the characteristic luminosity $L^*$ (Sobral et al. 2013, Vang et al. in prep), and this is also the case for the clustering. In the right panel of Fig. 15 we show, for the luminosity-bin samples, $r_0$ as a function of H\alpha luminosity divided by $L^*$ calculated at the mean redshift of each sample, where $L^*(z) = 2.63 \times 10^{41} (1+z)^{2.3} \text{erg s}^{-1}$ from the 3D-HST data (Vang et al. in prep). Correcting the H\alpha luminosity by $L^*$ removes practically all the redshift evolution with the two relations now virtually lying on top of each other.

Given there is some redshift evolution in the clustering amplitude at fixed H\alpha luminosity over our redshift range, it is worth considering if that will affect any of our main results. When dividing our samples only by H\alpha luminosity, we do introduce a redshift trend such that the H\alpha luminosity increases so does the mean redshift of the sample, with it going from 0.82 to 1.21 from lowest to highest luminosity respectively. Given the results shown in Fig. 15, the redshift evolution will have a tendency to flatten the relationship between $r_0$ and H\alpha luminosity shown in Fig. 4. There can be a similar tendency in the relation between galaxy bias $b_g$ and H\alpha luminosity and that between median mass $M_{\text{h, cen}}$ of haloes hosting central galaxies and H\alpha luminosity (Fig. 10). However, we expect the effect to be smaller than that on $r_0$. We can see this by noting that the clustering amplitude is proportional to $b_g^2 D(z)^2$ as well as $r_0^2$, where $D(z)$ is the linear growth factor. That is, $b_g \propto r_0^{\gamma/2}/D(z)$. We have $\gamma/2 < 0.7 < 1$ and $D(z)$ decreases with increasing redshift – both factors make $b_g$ less sensitive to the sample redshift than $r_0$, given that we have lower $r_0$ at higher $z$ with fixed H\alpha luminosity. We note that in our HOD modelling the redshift of each sample is adopted to compute halo properties (e.g. halo bias, halo mass function). Therefore, the derived quantities shown in Fig. 10, including $b_g$ and $M_{\text{h, cen}}$, correspond to those at the mean redshift of each sample.

When we bin by stellar mass as well as H\alpha luminosity, it is the stellar mass binning that dominates in determining the mean redshift of a given sample, with the mean redshift increasing by less than 0.3 between the highest and lowest stellar mass samples at fixed H\alpha luminosity. Within a given stellar mass bin the mean redshift changes by at most 0.09 with H\alpha luminosity and so any redshift evolution is negligible. Given that clustering amplitude at fixed stellar mass hardly evolves at all with redshift at these redshifts (e.g. Wake et al.
(hMpc)

Comparison of the clustering strengths inferred from our work with those from Cochrane et al. (2017) for luminosity-threshold and luminosity-bin samples. Left: clustering strength as a function of Hα luminosity threshold. Center: clustering strength as a function of mean Hα luminosity. Right: clustering strength as a function of Hα luminosity normalised by the characteristic luminosity L* of the Hα luminosity function at each redshift. Our results agree within two sigma across the two samples though Cochrane et al. see higher r0 values perhaps due to the differences in the redshifts ranges of our samples. All samples are fitted with a fixed γ of 1.8.

2011; Skibba et al. 2015), we can expect the trends shown in Fig. 8 and results of the halo model fitting in Section 5.3 to be largely unaffected by redshift evolution. To make sure that this is the case we have reproduced Fig. 8 for the two redshift ranges discussed above, confirming that the same trends are observed, but with a larger scatter as would be expected.

There have been previous analyses of the clustering of Hα emitting galaxies both observationally and in simulations. The most comparable to our study are those made with the High-Redshift(z) Emission Line Survey (HiZELS), which is a deep, near-infrared narrow-band Hα survey targeting galaxies in narrow redshift bins at z = 0.8, 1.47, and 2.23 (Geach et al. 2008; Sobral et al. 2009, 2012, 2013). The Hα emitting galaxies are selected based on the narrow-band and broad-band colour excess, with a restframe Hα equivalent width above 25Å. Photometric redshifts are used to ensure that the measured emission line is Hα. Of particular relevance are Cochrane et al. (2017) and Cochrane et al. (2018), which present analyses of the Hα luminosity dependent clustering and Hα luminosity and stellar mass dependent clustering of HiZELS galaxies, respectively. The HiZELS data used in these analyses cover a similar volume as the data we use here. Their two z = 0.8 fields (z = 0.845 ± 0.011 and z = 0.81 ± 0.011) cover a comoving volume of 4.60 × 10^5 h^3 Mpc^3 with their z ≈ 1.47 field covering 2.68 × 10^5 h^3 Mpc^3, compared to the 3.83 × 10^5 h^3 Mpc^3 covered by the 3D-HST survey. 3D-HST extends to lower Hα luminosities than HiZELS, meaning about 50% more galaxies are available to use in our clustering analysis.

In Fig. 15, we compare the clustering strengths r0 from our work with those in Cochrane et al. (2017) for the Hα luminosity-threshold (left) and luminosity-bin samples (middle and right). To enable a closer comparison, we have divided our samples into two redshift bins when measuring their clustering, 0.7 < z < 1.1 and 1.1 < z < 1.5. For the power-law fits, the index is fixed to 1.8 to be consistent with that used in Cochrane et al. (2017). Cochrane et al. (2017) add a dust attenuation correction of A_Hα = 1.0 mag, so we subtract their log L_Hα values by 0.4 dex to be in line with the values adopted in our work.

As can be seen in Fig. 15, our 3D-HST samples are able to reach a lower Hα luminosity than the HiZELS samples at similar redshifts. Looking at the r0 values we measure for our 0.7 < z < 1.1 samples, we see that at lower luminosity they are consistent with those from the HiZELS z ≈ 0.8 samples. As we move to higher luminosities, our r0 values lie below those of HiZELS at z ≈ 0.8, producing a shallower dependence of r0 on Hα luminosity than observed by HiZELS. This shallower slope is not being caused by redshift evolution within our samples, with the mean redshift varying from 0.93 to 0.95 in the overlapping Hα luminosity range of the two surveys, although there is an overall difference of about 0.1 in redshift with HiZELS. For our higher redshift (1.1 < z < 1.5) samples, the r0 values agree with those from HiZELS z ≈ 1.47 samples, but show a much clearer trend as a result of our larger sample size, as well as an extension to lower Hα luminosity. While the results from Cochrane et al. (2017) show a clear evolution of the clustering strength from z ≈ 1.47 to z ≈ 0.8 at all Hα luminosities, we see only a small difference for galaxies of the highest luminosity in our sample. This may partially be caused by the larger redshift difference between the two HiZELS samples, Δz ∼ 0.65, compared to our mean redshift difference of Δz ∼ 0.4, but it could also simply be a reflection of the larger cosmic variance uncertainties in the HiZELS data from their smaller number of independent fields.

Cochrane et al. (2018) measure the Hα luminosity and stellar mass dependent clustering much as we do in Section 4.2, but appear to find quite different results. They conclude that there is evidence for residual clustering variation with stellar mass at fixed Hα luminosity, and no evidence for residual Hα luminosity dependent clustering at fixed stellar mass, such that more luminous galaxies are more strongly clustered. This is quite different to what we concluded from Figs. 8 and 13. Cochrane et al. (2018) also find no dependence of r0 on stellar mass for their Hα emitting galaxies of stellar mass below 4 × 10^10 M_☉, with an increase above that. That is also different to our results (left panel of Fig. 8, open circles), where we see the clustering amplitude increase steadily for masses above 9.6 × 10^9 M_☉.

The most likely cause of these differences are two fold and quite straightforward. Firstly, the HiZELS data has a higher Hα luminosity limit that that of 3D-HST. Our 3D-HST samples go to about 0.3 dex lower in Hα luminosity (e.g. Fig. 2), meaning that we sample below the star formation mass sequence or star formation main sequence (SFMS) over our full stellar mass range. We show in Fig. 16 the distribution of our galaxies in the Hα luminosity–stellar mass plane, along with the same relation (blue curve) derived from fits to the
observed SFMS in Schreiber et al. (2015). To convert the SFR in the Schreiber et al. (2015) SFMS to Hα luminosity, we use the $L_{H\alpha}$--SFR relation in Kennicutt & Evans (2012) and then apply the stellar-mass-dependent dust extinction given by Garn & Best (2010). We sample below the 1σ scatter of the Schreiber et al. (2015) SFMS over virtually our full stellar mass range and below 2σ for galaxies with $M_* > 4 \times 10^9 M_\odot$. In comparison, Cochrane et al. (2018) probe above the SFMS at their lowest stellar masses and do not sample galaxies 2σ below until masses of $8 \times 10^9 M_\odot$. By only sampling the highest Hα luminosity (SFR) galaxies at the lowest stellar masses, as Cochrane et al. (2018) note they likely preferentially selecting satellite galaxies with enhanced star formation (see our Fig. 13). They are also more likely to select galaxies that have measured masses lower than their true values as a result of measurement errors. Both cases will cause the clustering amplitude to be enhanced for the lowest mass galaxies, reducing any dependence on stellar mass.

Likewise if we were to increase our Hα luminosity limit by 0.3 dex to match Cochrane et al. (2018) we would mainly be sampling the higher luminosity rising side of our observed V-shaped trend of $r_0$ with Hα luminosity at fixed stellar mass. That may have led us to conclude that there was evidence of a residual trend of increasing $r_0$ with Hα luminosity independent of stellar mass at the highest Hα luminosities as Cochrane et al. (2018) have done.

It is also important to consider the relatively large uncertainties in the $r_0$ measurements from both studies. While some of the trends are different, a close comparison of the individual measurements in overlapping ranges of mass and Hα luminosity show reasonable agreement given those errors. It will take a larger data set, most likely from Euclid or RST, to resolve these differences.

Coil et al. (2017) and Berti et al. (2019) investigate the sSFR dependent clustering of star-forming and quiescent galaxies at $0.2 < z < 1.2$ from the PRIMUS survey. Coil et al. (2017), looking at the full galaxy population, found some evidence for sSFR dependent clustering in star-forming galaxies with galaxies with lower sSFRs having a somewhat larger clustering amplitude. Our results do not confirm this trend, although we do find our $r_0$ and bias values are entirely consistent with theirs at comparable redshift, stellar masses, and SFRs. The differences may be driven by the choice of Coil et al. (2017) to divide their sample by sSFR over a wide range in mass rather than our choice to use fairly narrow bins in each. Berti et al. (2019) attempt to study the clustering of central galaxies only using an isolation criteria. Interestingly, these largely central only samples show no significant dependence of the clustering amplitude on sSFR for star-forming galaxies at fixed mass. This implies, much as we have found, that any residual dependence of clustering strength on sSFR is being driven by the satellite galaxy population.

7 CONCLUSIONS

In this work, using 3D-HST grism spectroscopic survey data, we study the clustering of Hα-emitting galaxies at $z \sim 1$. The projected 2PCFs of galaxies are measured as a function of Hα luminosity ($L_{H\alpha}$) and stellar mass ($M_*$) and characterised with power-law fits. We develop halo-based models, including the CLF model and the conditional $M_* - L_{H\alpha}$ distribution model, to interpret the clustering measurements and infer the relation between Hα-emitting galaxies and dark matter haloes.

The main results are listed below.

1. We find a clear correlation between Hα luminosity and clustering amplitude, with more luminous galaxies having a higher clustering amplitude.

2. At fixed Hα luminosity galaxies at lower redshift ($0.7 < z < 1.1$) are found to be more clustered than those at higher redshift ($1.1 < z < 1.5$), but the difference in clustering strength between our samples in the two redshift ranges is relatively small.

3. Our measurements of the Hα luminosity dependent clustering broadly agree with those of Cochrane et al. (2017), while we extend to lower luminosities.

4. At fixed Hα luminosity, galaxies with higher stellar mass tend to be more strongly clustered.

5. At fixed stellar mass, clustering strength does not seem to be dependent on Hα luminosity. It implies that the dependence of clustering on Hα luminosity is primarily driven by the relatively tight correlation between Hα luminosity and stellar mass (a.k.a. the star formation main sequence).

6. We use the CLF model to interpret the Hα luminosity dependent clustering. We further develop the conditional $M_* - L_{H\alpha}$ distribution model to jointly model the clustering measurements of galaxies in a series of stellar mass and Hα luminosity bins. The models provide good fits to our measurements.

7. Based on the modelling results, we find that central galaxies with higher luminosity or higher stellar mass reside in haloes of higher mass. Satellite galaxies tend to reside in more massive haloes than similar centrals. The conditional $M_* - L_{H\alpha}$ distribution model shows a tight correlation between central galaxy stellar mass and halo mass and only a weak correlation between central Hα luminosity and halo mass.

8. From the conditional $M_* - L_{H\alpha}$ distribution model, the median galaxy bias appears to have a strong dependence on stellar mass
with at most a weak dependence on Hα luminosity most evident in galaxies of the lowest stellar mass. For central galaxies there is a strong dependence of median mass of hosting haloes on stellar mass and essentially no additional dependence on Hα luminosity. Taken together this implies that there is a small residual dependence of the satellite galaxy halo occupation on Hα luminosity at fixed stellar mass. While the satellite fraction is only loosely constrained with the samples we study, we do see an indication that for galaxies of a given stellar mass the satellite fraction increases for both the high and low Hα luminosity galaxies, which is caused by the environment broadening of the range in SFR hence Hα luminosity of satellite galaxies at fixed stellar mass.

Our results help inform future surveys of star-forming or emission line galaxies (ELGs), such as the High Latitude Spectroscopic Survey (HLSS) of RST (Spergel et al. 2015) and the Euclid survey (Laureijs et al. 2011). They will be useful in designing the surveys and testing analysis pipeline, e.g. by providing inputs for mock construction. The modelling formalism can be applied to these surveys, as well as existing and ongoing surveys, such as ELG samples in eBOSS (Dawson et al. 2016) and in DESI (DESI Collaboration et al. 2016). With large samples from these surveys, the formalism like the conditional \(M_\star-L_{H\alpha}\) distribution will be an efficient model to explore the galaxy-halo relation, and we expect to obtain tight constraints on the relation to learn more about star-forming galaxies.

ACKNOWLEDGEMENTS

We would like to thank the anonymous referee for their helpful and constructive comments. CC acknowledges the support by a department Swigart Summer Research Fellowship. DW acknowledges support from program number HST-AR-13274, provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555. ZZ is supported by NSF grant AST-2007499. The support and resources from the Center for High Performance Computing at the University of Utah are gratefully acknowledged. This work is based on observations taken by the 3D-HST Treasury Program (GO 12177 and 12328) with the NASA/ESA HST, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555.

DATA AVAILABILITY

The data underlying this article are available in The STScI 3D-HST Repository, at https://archive.stsci.edu/prepds/3d-hst/. The measurements and modelling results underlying this article are available either in the article or on reasonable request to the authors.

REFERENCES

Berti A. M., Coil A. L., Hearin A. P., Moustakas J., 2019, ApJ, 884, 76
Blanton M. R., Moustakas J., 2009, ARA&A, 47, 159
Brammer G. B., et al., 2012, ApJS, 200, 13
Chabrier G., 2003, PASP, 115, 763
Cochrane R. K., Best P. N., Sobral D., Smail I., Wake D. A., Stott J. P., Geach J. E., 2017, MNRAS, 469, 2913
Cochrane R. K., Best P. N., Sobral D., Smail I., Geach J. E., Stott J. P., Wake D. A., 2018, MNRAS, 475, 3730
Coil A. L., Mendez A. J., Eisenstein D. J., Moustakas J., 2017, ApJ, 838, 87
Cortese L., Catinella B., Smith R., 2021, Publ. Astron. Soc. Australia, 38, e035
Coupon J., et al., 2012, A&A, 542, A5
DESI Collaboration et al., 2016, arXiv e-prints, p. arXiv:1611.00036
Davis M., Peebles P. J. E., 1983, ApJ, 267, 465
Dawson K. S., et al., 2016, AJ, 151, 44
Garn T., Best P. N., 2010, MNRAS, 409, 421
Geach J. E., Smail I., Best P. N., Kirk J., Casali M., Ivison R. J., Coppin K., 2008, MNRAS, 388, 1473
Hartlap J., Simon P., Schneider P., 2007, A&A, 464, 399
Katsianis A., et al., 2017, MNRAS, 472, 919
Kennicutt R. C., Evans N. J., 2012, ARA&A, 50, 531
Landy S. D., Szalay A. S., 1993, ApJ, 412, 64
Laureijs R., et al., 2011, arXiv e-prints, p. arXiv:1110.3193
Leauthaud A., et al., 2012, ApJ, 744, 159
Li C., Kauffmann G., Heckman T. M., Jing Y. P., White S. D. M., 2008, MNRAS, 385, 1903
Lim S. H., Mo H. J., Lan T. W., Ménard B., 2017, MNRAS, 464, 3256
McAlpine S., et al., 2016, Astronomy and Computing, 15, 72
Momcheva I. G., et al., 2016, ApJS, 225, 27
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
Schaye J., et al., 2015, MNRAS, 446, 521
Schreiber C., et al., 2015, A&A, 575, A74
Selkow R. E., et al., 2014, ApJS, 214, 24
Skibba R. A., et al., 2015, ApJ, 807, 152
Sobral D., et al., 2009, MNRAS, 398, 75
Sobral D., Best P. N., Matsuura Y., Smail I., Geach J. E., Cirasuolo M., 2012, MNRAS, 420, 1926
Sobral D., Smail I., Best P. N., Geach J. E., Matsuura Y., Stott J. P., Cirasuolo M., Kruk J., 2013, MNRAS, 428, 1128
Spergel D., et al., 2015, arXiv e-prints, p. arXiv:1503.03757
Tinker J. L., Weinberg D. H., Zheng Z., Zehavi I., 2005, ApJ, 631, 41
Vulcani B., et al., 2018, ApJ, 866, L25
Wake D. A., et al., 2011, ApJ, 728, 46
Whitaker K. E., et al., 2014, ApJ, 795, 104
Xu H., Zheng Z., Gao H., Zhu J., Zehavi I., 2016, MNRAS, 460, 3647
Xu H., Zheng Z., Gao H., Zu Y., Zehavi I., Weinberg D. H., 2018, MNRAS, 481, 5470
Yang X., Mo H. J., Yang X., Mo H. J., van den Bosch F. C., 2003, MNRAS, 339, 1057
Yang X., Mo H. J., van den Bosch F. C., 2008, ApJ, 676, 248
Zehavi I., et al., 2011, ApJ, 736, 59
Zheng Z., 2004, ApJ, 610, 61
Zheng Z., Coil A. L., Zehavi I., 2007, ApJ, 667, 760

APPENDIX A: BEST-FITTING PARAMETERS AND DERIVED QUANTITIES

In this paper, we model the Hα luminosity-dependent galaxy clustering within the CLF framework and then jointly model the stellar mass and Hα luminosity dependent galaxy clustering within the framework of the conditional stellar mass and Hα luminosity distribution.

For the conditional stellar mass and Hα luminosity distribution, at fixed halo mass, that for central galaxies is parameterised as a 2D Gaussian distribution, which is easy to be visualised. That for satellite galaxies is parameterised through the conditional stellar mass function and a relation between Hα luminosity and stellar mass (i.e. following the star formation main sequence). We use Fig. A1 to provide an illustration of the satellite component in \(10^{12} h^{-1} M_\odot\) haloes. The black curve is the conditional stellar mass function of satellites in these haloes. As we parameterise the Hα luminosity distribution as a function of stellar mass, when applying a cut in Hα luminosity, we obtain the conditional stellar mass function of satellites within the given Hα luminosity bin. The red, orange, and blue curves show the cases for the three Hα luminosity bins used in this paper. The four
Figure A1. Illustration of the conditional stellar mass and Hα luminosity distribution of satellite galaxies in $10^{12} h^{-1} M_\odot$ haloes. The black curve is the conditional stellar mass function (CSMF) of satellites. The red, orange, and blue curves are the CSMF after we apply the Hα luminosity cuts used in this paper. The four vertical lines delineate the three stellar mass bins used in constructing the samples. Integrating each of the red, orange, and blue luminosity distributions model are shown in Table A2 and A3, respectively. For the CSMF model, the best-fitting parameters with 1σ uncertainties and the values of χ² are displayed in Table A1.

The CLF parameterisation and parameter definitions can be found in Section 5.1.1.

Figure A2. The centre panel shows the luminosity distribution model for a given M_\text{h, cen}. The derived quantities for Hα luminosity-bin and luminosity-threshold galaxy samples

Table A2. Derived quantities for Hα luminosity-bin and luminosity-threshold galaxy samples

| Sample  | log M_\text{h, cen} | log M_\text{h, sat} | b_s | f_\text{sat} |
|---------|---------------------|---------------------|-----|-------------|
| LB1     | 11.03*0.06          | 13.49*0.24          | 0.94*0.03 | 0.04*0.02 |
| LB2     | 11.20*0.07          | 13.60*0.24          | 1.03*0.04 | 0.03*0.02 |
| LB3     | 11.36*0.10          | 13.71*0.25          | 1.12*0.05 | 0.03*0.02 |
| LB4     | 11.53*0.12          | 13.84*0.25          | 1.22*0.07 | 0.02*0.02 |
| LB5     | 11.69*0.15          | 13.97*0.26          | 1.30*0.09 | 0.02*0.02 |
| LB6     | 11.85*0.21          | 14.14*0.26          | 1.43*0.13 | 0.02*0.03 |

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

Table A1. Best-fitting CLF parameters for modelling Hα luminosity-bin galaxy samples.

| log A_{c,p} | d log A_{c,p} | log σ_c | log L_{c,p} | d log L_{c,p} | Log φ_{c,p} | d log φ_{c,p} | Δ log L_{cs} | α_c | χ²/d.o.f. |
|-------------|---------------|---------|-------------|---------------|-------------|---------------|-------------|------|-----------|
| 1.70*0.86   | -2.09*0.67    | -0.91*0.39 | 41.10*0.06  | -0.86         | 1.20*0.22   | -0.78         | -0.42       | 2.58*0.53 | -0.72     | -0.54*0.38 | -0.62 | 21.3/57 |

The derived quantities are the median mass M_\text{h, cen} of host haloes for central galaxies, the median mass M_\text{h, sat} of host haloes for satellite galaxies, the galaxy bias factor b_s, and the satellite fraction f_\text{sat}. Halo mass is in units of $h^{-1} M_\odot$. Priors imposed when modelling stellar-mass-Hα-luminosity-bin samples lead to higher satellite fractions than those from modelling luminosity-bin samples. The definitions and properties of the samples are found in Table 1.
Figure A2. Constraints on the CLF model parameters from modelling $L_{H\alpha}$-bin galaxy samples. Here, $\gamma_A \equiv d\log A_c/d\log M_h$, $\gamma_L \equiv d\log L_c/d\log M_h$, and $\gamma_\phi \equiv d\log \phi^*_s/d\log M_h$. Contours in each panel denote the 1$\sigma$ and 2$\sigma$ constraints for the pair of parameters. In each histogram panel, the central vertical line marks the median and the other two indicate the central 68.3 per cent distribution.
Figure A3. Constraints on the parameters of the conditional stellar mass and Hα luminosity distribution model from jointly modelling stellar mass and Hα luminosity dependence galaxy clustering. Here, $\gamma_\phi \equiv \frac{d \log \phi}{d \log M_h}$, and $\gamma_{\mu_{\text{yp, sat}}} \equiv \frac{d \log \mu_{\text{yp, sat}}}{d \log M_h}$. Contours in each panel denote the 1σ and 2σ constraints for the pair of parameters. In each histogram panel, the central vertical line marks the median and the other two indicate the central 68.3 per cent distribution.
Table A3. Derived quantities for stellar-mass-H\(\alpha\)-luminosity-bin galaxy samples

| Sample | \(\log M_{h,\text{cen}}\) | \(\log M_{h,\text{sat}}\) | \(b_g\)     | \(f_{\text{sat}}\)     |
|--------|----------------|----------------|------------|----------------|
| M1L1   | 11.30\(^{+0.11}_{-0.14}\) | 12.02\(^{+0.33}_{-0.38}\) | 1.09\(^{+0.11}_{-0.11}\) | 0.26\(^{+0.22}_{-0.18}\) |
| M1L2   | 11.32\(^{+0.11}_{-0.14}\) | 11.99\(^{+0.28}_{-0.35}\) | 1.17\(^{+0.10}_{-0.09}\) | 0.17\(^{+0.19}_{-0.12}\) |
| M1L3   | 11.34\(^{+0.11}_{-0.14}\) | 11.99\(^{+0.27}_{-0.34}\) | 1.24\(^{+0.17}_{-0.11}\) | 0.24\(^{+0.27}_{-0.17}\) |
| M2L1   | 11.56\(^{+0.10}_{-0.13}\) | 12.18\(^{+0.29}_{-0.26}\) | 1.16\(^{+0.08}_{-0.06}\) | 0.20\(^{+0.13}_{-0.10}\) |
| M2L2   | 11.59\(^{+0.10}_{-0.13}\) | 12.17\(^{+0.26}_{-0.24}\) | 1.23\(^{+0.06}_{-0.05}\) | 0.13\(^{+0.09}_{-0.06}\) |
| M2L3   | 11.61\(^{+0.10}_{-0.14}\) | 12.16\(^{+0.25}_{-0.23}\) | 1.31\(^{+0.09}_{-0.07}\) | 0.17\(^{+0.11}_{-0.08}\) |
| M3L1   | 12.05\(^{+0.13}_{-0.14}\) | 12.52\(^{+0.25}_{-0.27}\) | 1.38\(^{+0.09}_{-0.09}\) | 0.12\(^{+0.09}_{-0.06}\) |
| M3L2   | 12.06\(^{+0.10}_{-0.13}\) | 12.53\(^{+0.23}_{-0.26}\) | 1.49\(^{+0.07}_{-0.08}\) | 0.08\(^{+0.06}_{-0.04}\) |
| M3L3   | 12.11\(^{+0.12}_{-0.15}\) | 12.58\(^{+0.21}_{-0.28}\) | 1.66\(^{+0.09}_{-0.12}\) | 0.13\(^{+0.08}_{-0.07}\) |

Same as in Table A2, but for stellar-mass-H\(\alpha\)-luminosity-bin samples. The definitions and properties of the samples are found in Table 2.