Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **499,** 2104–2115 (2020) doi:10.1093/mnras/staa3027
Advance Access publication 2020 October 1

Are Ly α emitters segregated in protoclusters regions?

Tomás Hough [1,2,3]★ Siddhartha Gurung-López,[4] Álvaro Orsi,[5,6] Sofía A. Cora [1,2,3] Cedric G. Lacey [7] and Carlton M. Baugh[7]

[1]*Instituto de Astrofísica de La Plata (CCT La Plata, CONICET, UNLP), Observatorio Astronómico, Paseo del Bosque, B1900FWA La Plata, Argentina*
[2]*Facultad de Ciencias Astronómicas y Geofísicas, Universidad Nacional de La Plata, Observatorio Astronómico, Paseo del Bosque, B1900FWA La Plata, Argentina*
[3]*Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Rivadavia 1917, Buenos Aires, Argentina*
[4]*Department of Physics, Missouri University of Science and Technology, 1315 N. Pine Street, Rolla, MO 65409, USA*
[5]*Centro de Estudios de Física del Cosmos de Aragón, Plaza San Juan 1, piso 2, E-44001 Teruel, Spain*
[6]*PlantTech Research Institute Limited, South British House, 4th Floor, 35 Grey Street, Tauranga 3110, New Zealand*
[7]*Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham, DH1 3LE, UK*

Accepted 2020 September 22. Received 2020 September 18; in original form 2020 May 8

ABSTRACT

The presence of neutral hydrogen in the interstellar medium (ISM) and intergalactic medium (IGM) induces radiative transfer (RT) effects on Ly α photons that affect the observability of Lyman alpha emitters (LAEs). We use the GALFORM semi-analytic model of galaxy formation and evolution to analyse how these effects shape the spatial distribution of LAEs with respect to H α emitters (HAEs) around high-density regions at high redshift. We find that when a large sample of protoclusters is considered, HAEs showing also Ly α emission (HAEs + LAEs) populate the same regions as those that do not display the Ly α line at $z = 2.2$. We compare against the protocluster USS1558-003, one of the most massive protoclusters located at $z = 2.53$. Our results indicate that the strong depletion of HAEs + LAEs present in the high-density regions of USS1558-003 may be due to cosmic variance. We find that at $z = 2.2$ and $z = 3.0$, RT of the ISM produces a strong decline (30–50 per cent) of the clustering amplitude of HAEs + LAEs with respect to HAEs towards the protoclusters centre. At $z = 5.7$, given the early evolutionary state of protoclusters and galaxies, the clustering of HAEs + LAEs has a smaller variation (10–20 per cent) towards the protoclusters centre. Depending on the equivalent width and luminosity criteria of the emission-line galaxy sample, the IGM can have a mild or a null effect on galaxy properties and clustering in high-density regions.

Key words: Radiative transfer – methods: numerical – Galaxies: clusters: general – Galaxies: high-redshift – Galaxies: clusters: intracluster medium.

1 INTRODUCTION

The properties of the large-scale environment in which protoclusters are embedded is crucial to determine how they will evolve into massive galaxy clusters ($M_\star > 10^{14}\,\mathrm{M}_\odot$) at the present time. Protoclusters at high redshift ($z \geq 2$) are identified as overdense regions of galaxies and gas, usually associated to radio galaxies (e.g. Le Fevre et al. 1996; Pentericci et al. 1997; Venemans et al. 2002, 2007; Hayashi et al. 2012; Orsi et al. 2016), quasars (e.g. Wold et al. 2003; Kashikawa et al. 2007; Overzier et al. 2009; Adams et al. 2015), or other massive objects (e.g. submillimetre galaxies, Ly α blobs).

Emission-line galaxies (ELGs) are star-forming galaxies whose spectra contain intense nebular emission lines. As the characteristic intensity of lines in emission allows detection and precise redshift, ELGs are often used to detect matter overdensities at high redshift. This helps constrain their spatial distribution over a small slice of cosmic volume.

Among ELG, those that have a detectable Ly α emission line ($\lambda = 1216$ Å) or a detectable H α emission line ($\lambda = 6563$ Å) are referred to as Lyman alpha emitters (LAEs) and H α emitters (HAEs), respectively. In a star-forming galaxy, these lines have the same astrophysical origin, i.e. they are produced when the ionizing emission of young and massive stars is absorbed by atomic hydrogen regions located in the interstellar medium (ISM). The recombination of these atoms leads to the emission of both Ly α and H α photons (Orsi, Lacey & Baugh 2012; Dijkstra 2017). Furthermore, Ly α photons are absorbed and scattered by the ISM, the circumgalactic medium (CGM), and the intergalactic medium (IGM) through complex radiative transfer (RT) processes that affect the LAEs observed properties (Orsi et al. 2014; Gurung-López et al. 2019a). However, due to the small cross-section of the interaction between Hα photons and neutral hydrogen atoms, HAEs are largely unaffected by these effects, making these galaxies excellent tracers of instantaneous star formation rate (SFR) (Kennicutt 1998; Calzetti 2013). Dual emitter surveys comprising both H α and Ly α emission at the same redshift are a powerful tool to understand the intrinsic properties of HAEs and LAEs, and provide insights on the intrinsic and observed Ly α luminosity functions. Although the fraction of dual emitters (HAEs + LAEs) depends on the survey depths (Shimakawa et al. 2017b), several studies report on the low fraction of Ly α photons that escape from HAEs, from ∼5 per cent (Hayes et al. 2010; Matthee

★ E-mail: tomashough@gmail.com

et al. 2016) up to ∼37 per cent (Sobral et al. 2017). Escape fraction is found to strongly anticorrelate with dust extinction and SFR, and only weakly with stellar mass (Hayes et al. 2011; Matthee et al. 2016). We remark that dual emitter surveys can be performed on a narrow redshift range ($z \sim 2.2-2.5$) from ground-based observatories.

The spatial distribution of ELGs in overdense regions, can be used to infer the underlying dark matter (DM) distribution, and how galaxy properties relate to the environment in which they reside (Mo et al. 2004; Cooray 2005; Overzier et al. 2006; Orsi et al. 2016; Ota et al. 2018; Shi et al. 2019; Gonzalez-Perez et al. 2020). In particular, the way in which the environment of protoclusters affects the emission of ELG at high redshift is still a matter of debate (see Overzier 2016, for a review). Several authors have reported on different galaxy populations residing in high-density regions. For instance, using data from the VLT FORS fields, Venemans et al. (2002, 2007) found that LAEs are (relatively) randomly distributed in the protocluster TN-J1338, located at $z \sim 4.1$, while Overzier et al. (2008) found that LAEs seem to prefer regions devoid of UV-selected Lyman Break Galaxies. Hayashi et al. (2016) studied the properties of HAEs that trace the rich protocluster USS1558-003, located around a radio galaxy at $z = 2.53$ (Hayashi et al. 2012). They found that HAEs with $M_\star > 10^{10}\,\mathrm{M}_\odot$ are located in the SFR $- M_\star$ main sequence of star-forming field galaxies, while some HAEs with $M_\star < 10^{9.3}\,\mathrm{M}_\odot$ are deviated upward the main sequence, with SFRs consistent with starburst galaxies. Shimakawa et al. (2017a, hereafter S17) analysed the Ly α emission of the HAEs in the protocluster USS1558-003 and found a clear lack of LAEs in dense regions traced by HAEs, and suggested that an excess of dust and gas accreted in cold streams might prevent the escape of Ly α photons from the core of the protocluster. This dual emission analysis has been also performed in the SpiderWeb protocluster, located around the PKS 1138 − 262 radio galaxy at $z = 2.16$; this protocluster presents a concentration of HAEs that increases towards the radio galaxy, while LAEs do not (Kurk et al. 2004).

A way to quantify how galaxies are distributed around a central object is computing the cross-correlation function. At high redshift, the cross-correlation function $\xi_{\rm cc}$ between overdensity tracers (radio galaxies and quasars) and ELGs can offer different information on small and large scales. Orsi et al. (2016) found that, at large scales ($r \gtrsim 10\mathrm{Mpc}\,h^{-1}$), the amplitude of $\xi_{\rm cc}$ for H α and Ly α emitters is larger when the central objects are radio galaxies, because they inhabit more massive haloes. At small scales, faint LAEs ($\mathrm{L}\alpha > 10^{41}\mathrm{erg\,s^{-1}}\,h^{-2}$) have higher $\xi_{\rm cc}$ than bright LAEs ($\mathrm{L}\alpha > 10^{42}\mathrm{erg\,s^{-1}}\,h^{-2}$), because active galactic nucleus (AGN) feedback prevent starburst galaxies to dominate the galaxy abundance at small separations. Recently, Gurung-López et al. (2020) found that the presence of the IGM induces a scale-dependent effect on the autocorrelation function $\xi(\mathrm{r})$ of LAEs at $z = 5.7$, where the shape of $\xi(\mathrm{r})$ becomes broader at the baryon acoustic oscillation scale, and the maximum is displaced by $\sim 1\,\mathrm{cMpc}\,h^{-1}$. If the presence of the IGM had an impact on the observed spatial distribution of LAEs in high-density environments, it could produce misleading conclusions on the interpretation of clustering data of future surveys such as Hobby Eberly Telescope Dark Energy Experiment (HETDEX; Hill et al. 2008), Dark Energy Spectroscopic Instrument (DESI; Levi et al. 2013).

In this work, we use the GALFORM semi-analytic model of galaxy formation and evolution (Cole et al. 2000; Lacey et al. 2016; Baugh et al. 2019) to explore the spatial distribution of HAEs and LAEs around protoclusters, at redshifts up to $z \lesssim 6$, and evaluate the impact of the IGM in such distributions.

In Section 2, we describe the semi-analytic model and dark-matter-only simulation on which the model is applied, along with a brief description of our theoretical approach of the RT process that takes place in both the ISM and IGM. In Section 3, we analyse the spatial distribution of LAEs and HAEs in high-density environments. The impact of the IGM on the clustering of LAEs at small scales around protoclusters is detailed in Section 4. Our conclusions are summarized in Section 5.

2 THEORETICAL APPROACH

The construction of a realistic synthetic population of galaxies in a cosmological context requires a set of numerical tools that combines the cosmological framework with baryonic physics, which rules the intrinsic and observable properties of galaxies. This is achieved by combining a cosmological DM simulation with a semi-analytic model of galaxy formation, and open-source software that incorporate ISM and IGM radiative transfer effects.

(i) **Dark-matter-only simulation.** The P-Millennium (Baugh et al. 2019) is a state-of-the-art dark-matter-only *N*-body simulation that models the hierarchical growth of structures in the Lambda cold dark matter (ΛCDM) scenario. It uses the Planck cosmology: $\mathrm{H}_0 = 67.77\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_\Lambda = 0.693$, $\Omega_\mathrm{M} = 0.307$, and $\sigma_8 = 0.8288$ (Planck Collaboration 2016). The box size is 542.16 $\mathrm{cMpc}\,h^{-1}$ and the particle mass $M_\mathrm{p} = 1.061 \times 10^8\,\mathrm{M}_\odot\,h^{-1}$ (5040^3 dark matter particles). The DM halo merger trees are constructed from the SUBFIND subhaloes using the DHALOS algorithm described in Jiang et al. (2014). Haloes that contain at least 20 particles are retained, corresponding to a halo mass resolution limit of $2.12 \times 10^9\,\mathrm{M}_\odot\,h^{-1}$.

(ii) **Semi-analytic model.** We use the GALFORM semi-analytic model of galaxy formation and evolution. In short, GALFORM initially populates DM haloes with gas. Then, tracking the merger history of haloes, the gas is evolved including several physical mechanisms: (i) shock heating and radiative cooling of gas inside haloes; (ii) formation of a galactic disc with quiescent star formation; (iii) triggering of starburst episodes in bulges due to disc instabilities and mergers; (iv) active galactic nuclei, supernovae and photoionization feedback to regulate the SFR; (v) the chemical evolution of gas and stars.

GALFORM computes the H α and Ly α luminosities of galaxies from the total production rate of hydrogen ionizing photons (Lyman continuum photons). This is obtained by integrating the composite spectral energy distribution (SED) of each galaxy over the extreme-UV continuum down to the Lyman break at $\lambda = 912$ Å. Then, by assuming that all of these ionising photons are absorbed within the ISM of the galaxy and that no direct recombination into ground state takes place (case B recombination), a fraction of Lyman continuum photons is converted into different line fluxes (Osterbrock 1989; Dijkstra 2014).

On one hand, H α emission can suffer dust attenuation. GALFORM includes a two-step dust attenuation: one for the emission of stars that are still inside their birth cloud, and one for the emission that emerges from molecular clouds and stars located outside the clouds, which are affected by diffuse dust component present in the disc/bulge components of the galaxy. This model includes diffuse dust attenuation at 14 bands, including the R band (centred at 6594 Å). We refer the reader to appendix A of Lacey et al. (2016) for more details. On the other hand, the intrinsic luminosity of Ly α photons is expected to be reduced by both the scattering they suffer by neutral hydrogen atoms in the ISM and IGM, and their absorption by dust grains.

(iii) **ISM radiative transfer model.** Ly α photons are assumed to escape the galaxy through outflows. The outflow is characterized by an expansion velocity, hydrogen column density, and dust optical depth, which depend upon the galaxy properties. The outflow velocity is computed as:

$$V_{exp,c} = \kappa_{V,c} SFR_c \frac{r_c}{M_\star}, \quad (1)$$

where the subscript c denotes the galaxy component (disc or bulge), SFR_c is the star formation rate in $M_\odot Gyr h^{-1}$, r_c is the half stellar mass radius in $Mpc h^{-1}$, $M_\star$ is the total stellar mass of the galaxy in $M_\odot h^{-1}$, and $\kappa_{V,c}$ are free dimensionless parameters that regulate the efficiency of gas ejection. The neutral hydrogen column density of the outflows is computed for each component as:

$$N_{H,c} = \kappa_{N,c} \frac{M_{cold,c}}{r_c^2}, \quad (2)$$

where $M_{cold,c}$ is the cold gas mass of each component in $M_\odot h^{-1}$ units, and $\kappa_{N,c}$ are free parameters calibrated for each component and redshift. Finally, the optical depth of dust absorption is computed as:

$$\tau_{a,c} = (1 - A_{Ly_\alpha}) \frac{E_\odot}{Z_\odot} N_{H,c} Z_c, \quad (3)$$

where $E = 1.77 \times 10^{-21}$ cm^{-2} is the ratio τ_a/N_H for solar metallicity, $A_{Ly\alpha}$ is the albedo at Lyα wavelength, the solar metallicity is $Z_\odot = 0.02$ and Z_c is the metallicity of the cold gas of each component.

Then, in order to compute the escape fraction (f_{esc}), we use FLaREON (Gurung-López, Orsi & Bonoli 2019b), an open PYTHON package based on a Monte Carlo RT code (Orsi et al. 2012) that predicts the Ly α line profiles and escape fractions of photons in outflows of different characteristics. The FLaREON code includes three different gas outflow geometries: Thin Shell, Galactic Wind (both with spherical symmetry but different neutral hydrogen density profiles), and Bicone (see Gurung-López et al. 2019b, for further details). In this work, we use the Thin Shell geometry to compute the ISM transmission, where the hydrogen column density of the outflow is described by equation (2). The Thin Shell geometry reproduces better the observed properties of LAEs, like the dependence of the offset of the peak of the Ly α line on stellar mass, SFR and EW (Gurung-Lopez et al. 2020b, in preparation).

(iv) **IGM radiative transfer model.** While inside galaxies the losses of Ly α flux are due to dust absorption, photons in the IGM are scattered out of the line of sight by the neutral hydrogen.

We estimate the IGM transmission for every galaxy depending on the local environmental properties, such as density, velocity, and ionization state of the IGM. In the simulation, the IGM is distributed in cosmological boxes of $\sim$2cMpch^{-1} a side, with its density determined according to the DM content inside the box.

As a first approximation, the IGM absorbs photons with wavelengths shorter than 1216 Å. Moreover, as galaxies lie in overdense regions, the IGM opacity is higher close to the galaxy, causing the drop in the transmission close to Ly α wavelength. Then, the IGM transmission flattens to the IGM cosmic transmission. Additionally, the mean number density of neutral hydrogen atoms in the IGM increases with redshift up to $z \lesssim 6$, leading to an increase of the optical depth as well (see McQuinn 2016, for a review). The average transmission of the IGM results of 85 per cent (40 per cent) for $z = 2$ ($z = 4$) (Dijkstra 2014), and drops below 1 per cent at $z = 5.7$ for higher frequencies than that of Ly α (Gurung-López et al. 2019b).

3 LAE DEPLETION AT HIGH DENSITIES

S17 studied the Ly α emission of HAEs located in USS1558-003, the richest protocluster known at $z \sim 2.5$, with an estimated dynamical mass of $\sim 10^{14}$ M$_\odot$, consistent with a progenitor of a massive cluster ($\sim 10^{15}$ M$_\odot$) in the local Universe (Shimakawa et al. 2014). They found that LAEs tend to avoid high-density regions traced by HAEs, and that denser regions present lower Ly α escape fractions. This could be produced by a gaseous and dusty component covering the protocluster core, and it is not clear whether this should be expected as systematic for other protoclusters, or is the result of a particular conjunction of intrinsic characteristics.

Motivated by these observational results, we use GALFORM semi-analytic model to study the spatial segregation of LAEs relative to HAEs in a wide sample of simulated protoclusters. As in S17, we study the relation between the spatial distribution of galaxies and the local density of galaxies that exhibit H α and Ly α emission simultaneously (HAEs + LAEs). In their work, S17 performed Ly α imaging using the NB428 narrow band filter (central wavelength of 4297 Å and FWHM of 84 Å) of the Subaru Prime Focus Camera of the Subaru Telescope. The full width at half-maximum (FWHM) that was used allows the detection of LAEs with $z = 2.53 \pm 0.03$, or 21.3 Mpc uncertainty depth. S17 included the observations in H α performed by Hayashi et al. (2016), and the ELG sample consisted of 104 HAEs, with 13 of those galaxies also presenting Ly α emission.

We create mock catalogues of protoclusters at $z = 2.2$ with the same spatial constraints as the one observed by S17. To do this, from all our central galaxies, we consider radio-galaxy candidates as a protocluster centre at high redshift. Radio galaxies were selected according to the halo mass function (Orsi et al. 2016). At $z = 2.2$, we have 1048 protoclusters candidates with $M_{halo} > 10^{13.2}$ M$_\odot h^{-1}$. The mean density of galaxies inside 2 cMpc of these objects spans between 10 and 400 times the mean density of objects in our simulation.

The distance to the Nth neighbour is commonly used as a proxy for local density, and has the advantage of not assuming an underlying geometry (Baldry et al. 2006; Bluck et al. 2020). In the case of S17, they define the mean projected distance $< a >_{Nth} = 2 \times (\pi \sum_{Nth})^{-0.5}$, where $\sum_{Nth} (= N/\pi r_{Nth}^2)$ is the number density of galaxies within the r_{Nth} radius. This is the distance to the $(N-1)$th neighbour from each galaxy, and $N = 5$.

By applying emission-line equivalent width, EW, and luminosity limits, we define two samples of simulated ELG around protoclusters at $z = 2.2$:

(i) Flux limited sample (referred to as FL): We consider the same EW and luminosity limits as in S17 (see also Hayashi et al. 2016; Shimakawa et al. 2017b, for more details). In this case, HAEs have line emission widths $EW_{H\alpha} > 18.6$ Å and luminosity $L_{H\alpha} > 4.35 \times 10^{41}$ erg s^{-1}, while HAEs + LAEs are imposed to have also $EW_{Ly\alpha} > 15$ Å and $L_{Ly\alpha} > 4.4 \times 10^{41}$ erg s^{-1} (i.e. HAEs + LAEs satisfy both luminosity and EWs cuts).

(ii) Fixed number density sample (referred to as FN): We impose luminosity limits that allow us to match the observed surface density of HAEs and HAEs + LAEs as in S17. In this case, HAEs have $L_{H\alpha} > 10^{41}$ erg s^{-1}, while HAEs + LAEs are imposed to have also $L_{Ly\alpha} > 1.5 \times 10^{42}$ erg s^{-1}. The EW limits are equivalent as in the first sample. With these limits, our protocluster candidates have a median of 90 HAEs and 13 HAEs + LAEs.

For both samples, ELGs located inside a $2.0 \times 3.5 \times 21.5$ Mpc box centred on each protocluster centre are considered mem-

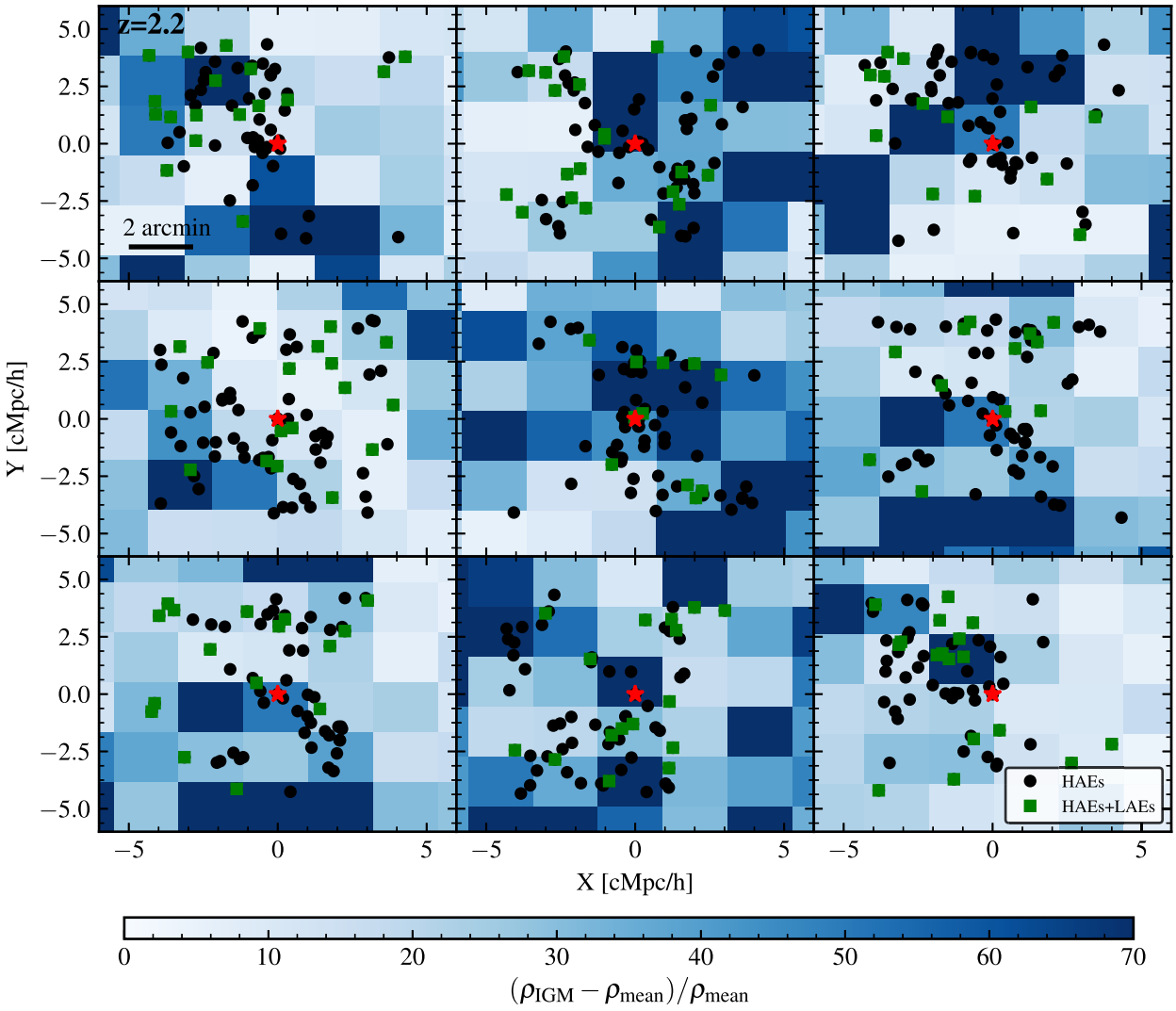

Figure 1. Spatial distribution of HAEs (black dots) and HAEs + LAEs (green squares) in nine simulated protoclusters at $z = 2.2$. Positions are given in co-movil coordinates. Each protocluster has $M_{halo} > 10^{13.7}\,M_{\odot}h^{-1}$ (the red star indicates the central radio galaxy). The spatial constraints, EW and luminosity criteria are adopted from S17 (FL sample). The coloured squares represent the maximum value of the IGM overdensity (considering the extension of the protocluster along the z coordinate) with respect to the mean density of IGM in the complete simulation. We highlight that the 3 protoclusters of the upper panel are those that present a HAEs + LAEs depletion similar to what is present in USS1558-003 protocluster, located at $z = 2.53$ (S17).

bers of the sample. These spatial constraints correspond to those of S17.

Although the protoclusters in both our FN and FL samples span halo masses between $10^{13.2}\,M_{\odot}\,h^{-1}$ and $10^{14.2}\,M_{\odot}\,h^{-1}$ (a mass range that comprises the value of the dynamical mass of the protocluster USS1558-003), none of the protoclusters in the FL sample reaches the number density of HAEs in USS1558-003.

Reproducing high-density environments at high redshift is challenging for simulations, mainly due to failures in capturing correctly the baryonic physics that are involved in the assembly of stellar mass with cosmic time. To approximate this high-density environment, we use only the 30 most massive protoclusters of our FL sample, with halo masses above $10^{13.7}\,M_{\odot}\,h^{-1}$ and an average of 45 HAEs and 10 HAEs + LAEs. The fact that the number density of HAEs in these protoclusters do not reach the number density of USS1558-003 could induce a bias in our analysis: a higher number of HAEs would certainly decrease $< a >_{5th}$, affecting the spatial distribution of both populations. However, we consider that the environment in which galaxies reside is well characterized, as the simulated DM halo masses encompasses the inferred M_{halo} of USS1558-003, and we highlight that the analysis of both FN and FL samples points towards the same conclusion.

In Fig. 1, we show the spatial distribution of HAEs and HAEs+LAEs in a subsample of the 30 most massive protoclusters of the FL sample. In coloured squares, we represent the maximal IGM overdensity (along the line of sight, in the volume occupied by the protocluster) with respect to the median IGM density in the complete simulation. The protoclusters of the upper row are those that present a depletion of HAEs + LAEs in high local densities ($< a >_{5th} < 0.3$), similar to what is found in S17, as we discuss in Fig. 2. While HAEs in most simulated protoclusters seem to distribute around the central radio galaxy, in some cases the distribution have a clear offset of $\sim$2 arcmin, similar to S17. The middle and right-hand panels of the bottom row are examples of such protoclusters. We note the presence of a correlation between IGM and HAEs overdensities, despite the spatial fluctuations of the IGM density. As can be appreciated, the IGM model consists of homogeneous cubes of $\sim$2 cMpc a side, where the IGM density is associated with the DM content of the

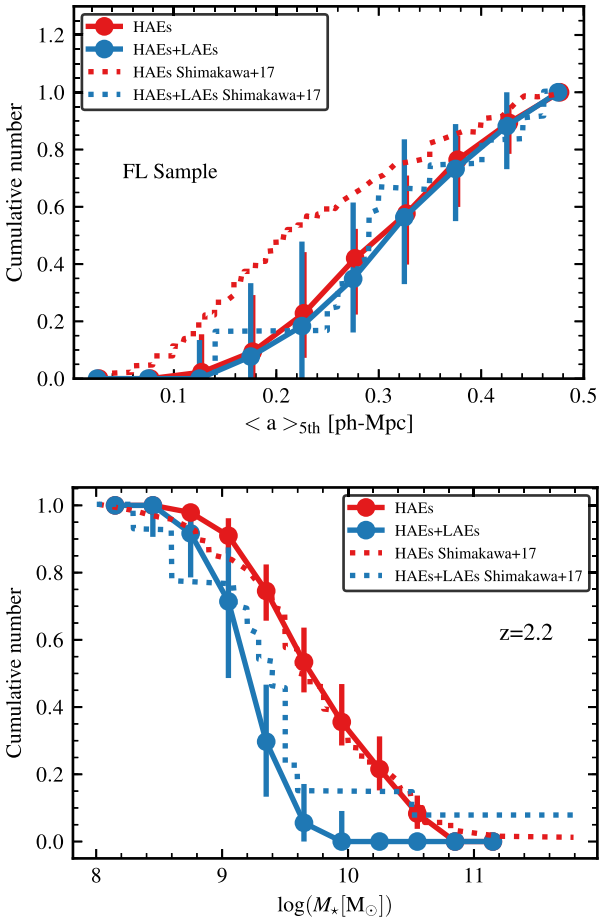

Figure 2. Cumulative number of HAE and HAEs + LAE in terms of the mean projected distance $< a >_{5th}$ (upper panel) and stellar mass content (lower panel) for the FL sample. Solid lines represent the median and error bars denote the 10–90th percentiles for HAEs (red) and HAEs + LAEs (blue) for the 30 most massive protoclusters selected at $z = 2.2$, who present halo masses above $10^{13.7}\,M_{\odot}\,h^{-1}$. Dotted lines represent the behaviour of the USS1558-003 protocluster, located at $z = 2.53$. The selections of synthetic HAEs and HAEs + LAEs match the observational conditions of S17.

box; this might result as a rough approximation. However, given the spatial constraints imposed by the observations, each simulated protocluster comprises ∼640 IGM cubes and, for each galaxy, the transmission of the Ly α line is computed through the cubes along the line of sight.

A more realistic IGM model would need a more refined grid, which would produce prohibitively expensive computational time due to the large size of the DM simulation box.

For the FL sample, we compute the mean projected distance $<a>_{5th}$ for every HAE and compare the median cumulative number of HAEs and HAEs + LAEs in the upper panel of Fig. 2. Error bars represent 10–90th percentiles. We find a clear discrepancy with observations: while S17 find that HAEs+LAEs avoid the densest regions, our analysis indicate that, on average, HAEs + LAEs inhabit the same regions as HAEs. We notice that some individual protoclusters present depletion of LAEs, while others are depleted of HAEs in high-density regions. This results in similar median behaviour, but with a relatively high dispersion when the 30 protoclusters are considered.

In general, galaxies that inhabit dense environments tend to be more massive than field galaxies, at low and intermediate redshifts (Baldry et al. 2006; Darvish et al. 2015). At $z \gtrsim 2$, HAEs in protoclusters have been found to be more massive than HAEs in the field. We note that our simulated HAEs present an increase in stellar mass towards the protocluster centre: HAEs present between 1.3 and 3 times the stellar mass of HAEs located in average regions, in consistency with observations (Hatch et al. 2011; Koyama et al. 2013). In the lower panel of Fig. 2, we present the stellar mass distribution of both populations of the FL sample, showing a remarkable agreement with the observational data given by S17. Ly α RT favours the escape of Ly α photons from galaxies with lower stellar mass, dust content, and SFRs than HAEs with the same line luminosity (Guaita et al. 2011; Orsi et al. 2016; Shimakawa et al. 2017b), and is further detailed in Section 4. Consistently, our HAEs + LAEs have lower stellar masses than HAEs, and start to accumulate at $\log(M_{\star}/M_{\odot}) \sim 9.6$, while HAEs start to accumulate at $\log(M_{\star}/M_{\odot}) \sim 10.8$. This means that although the FL sample has lower number density of HAEs than USS1558-003, the sample is thus complete as the stellar masses of both HAEs and HAEs + LAEs are in accordance with observations.

In order to explore environments with similar ELG number density than USS1558-003, we perform the same analysis for the FN sample, which considers the complete 1048 protocluster candidates at $z = 2.2$. In this case, for a given value of the mean local density $< a >_{5th}$, the number density of HAEs increases with respect to the one obtained with the FL sample, achieving a better agreement with S17, as shown in the upper panel of Fig. 3. Nevertheless, HAEs + LAEs do not seem to specifically avoid the regions traced by HAEs in the FN sample either. In fact, some protoclusters follow the observational trend, while others present the opposite behaviour, as can be appreciated from Fig. 4. To quantify the depletion of HAEs or LAEs in individual protoclusters, we compute a two-sample Kolmogorov–Smirnov (KS) test between the cumulative distributions of HAEs and HAEs + LAEs. In Fig. 4, we show which separations are more likely to occur in the FN sample. We assign positive values for protoclusters that present HAEs+LAEs depletion, and negative values for protoclusters that show HAEs depletion. It is clear that a small depletion of HAEs+LAEs is the most likely scenario (62 per cent present HAE+LAE depletion). But a similar depletion of HAEs is also found, which results in a statistically negligible depletion of HAEs + LAEs with respect to HAEs when all protoclusters are averaged.

The lower panel of Fig. 3 shows that both HAEs and HAEs+LAEs in the FN sample have lower stellar masses than observed. This is due to the fact that in the FN sample we are considering HAEs with lower luminosity than in the FL sample, so their stellar masses are also lower. Besides, although HAEs start to accumulate at higher stellar mass than HAEs + LAEs, this trend is not as steep as observed.

As suggested by S17, the accretion of cold streams that supply the protocluster core with H I gas could prevent Ly α photons from escaping from the dense regions of USS1558-003. We find that when a large sample of protoclusters is considered, this characteristic signature should not be expected to be as violent as in S17, although a small segregation of LAEs is the most likely scenario. Probably, the result found by S17 corresponds to a particular situation. If the inflow of cold gas occurs along a stream in the line of sight, it could enhance the dispersion of Ly α photons and diminish the number of HAEs + LAEs detected in dense regions. In Fig. 4, the dotted vertical line represents the maximum separation in USS1558-003. Among the simulated protoclusters that present HAEs + LAEs depletion, we find that ∼10 per cent follow the observational trend of S17, suggesting that their result is subject to cosmic variance. However,

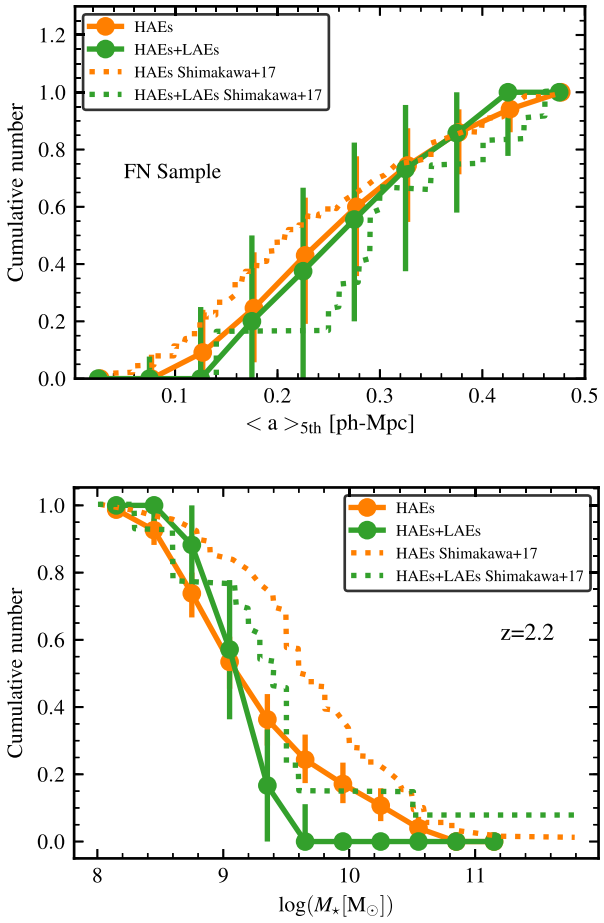

Figure 3. Same as Fig. 2, but for the FN sample. Solid lines represent the median and error bars denote the 10–90th percentiles of HAEs (orange) and HAEs + LAEs (green) for 1048 protoclusters selected at $z = 2.2$, who present halo masses above $10^{13.2}$ M$_\odot$ h^{-1}. Dotted lines represent the behaviour of the USS1558-003 protocluster. The selection of HAEs and HAEs + LAEs was defined to match the surface density of ELGs in USS1558-003 (S17).

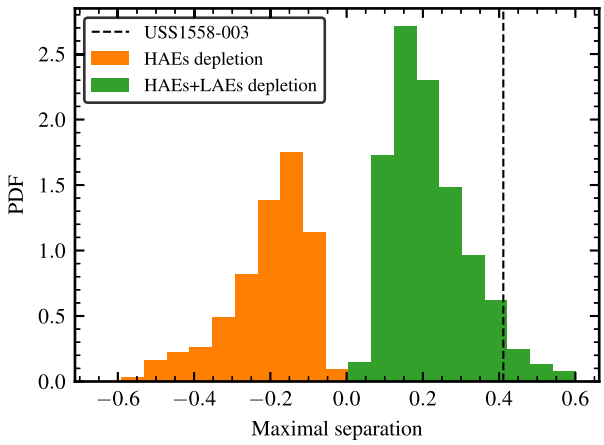

Figure 4. Histogram of the KS tests between cumulative numbers of HAEs and HAEs + LAEs for all simulated protocluster considered in Fig. 3 (FN sample). Positive values (in green) are associated to HAEs + LAEs depletion (as in S17), while negative values (in orange) are associated with HAEs depletion. Near ∼10 per cent of the simulated protoclusters have distances consistent with USS1558-003, represented by the black vertical line.

only in ∼1 per cent of the protoclusters the null hypothesis can be rejected at a 95 per cent confidence level.

It is noteworthy that the results obtained with the FN sample are independent of the halo mass limit of our protocluster candidates, as we find the same general behaviour when selecting haloes with $\log(M_{\rm halo}[{\rm M}_\odot\, h^{-1}]) > 13.5$ and $\log(M_{\rm halo}[{\rm M}_\odot\, h^{-1}]) > 14.0$. Moreover, the protoclusters that present Ly α depletion consistent with S17 show no specific signature in their intrinsic properties, such as sSFR, metallicity or DM content, with respect to those protoclusters that do not present Ly α depletion, or even with those that present H α depletion. This analysis sustains the hypothesis that a specific environmental effect could be producing the observed Ly α depletion.

Due to the powerful multiwavelength emission of AGNs, a certain degree of AGN contamination in NB observational samples of ELGs is expected. For instance, Sobral et al. (2016) studied a sample of 59 high luminous HAEs ($L_{\rm H\alpha} > 10^{42}$ erg s^{-1}) in the redshift range $0.8 < z < 2.23$, and found that ∼30 per cent of the galaxies hosts AGNs, but the AGN fraction increases with H α luminosity, and has little to no dependence on redshift. In a sample of 188 LAEs located at $z = 2.23$, Sobral et al. (2017) showed that only for L $> 10^{43}$ erg s^{-1} the luminosity function is dominated by X-ray AGNs. In the particular case of USS1558-003 protocluster, Macuga et al. (2019) found that a surprisingly low fraction (only ∼ 2 per cent) of the $H\alpha$ emitters are X-ray AGNs. In our model, both H α and Ly α luminosities are powered only by the star-forming regions inside galaxies, i.e. our HAEs and HAEs + LAEs samples are not contaminated with AGNs, which result in a particularly good comparison with ELGs in USS1558-003 protocluster.

4 THE IMPACT OF THE IGM ON THE CLUSTERING AT SMALL SCALES

In the ΛCDM paradigm, the density of the IGM is higher around massive structures, increasing the probability of scattering Ly α photons that escape from star-forming galaxies. Gurung-López et al. (2020) used the semi-analytic model and RT tools detailed in Section 2 to explore the coupling between the IGM and the Ly α observability, and found that the IGM modifies the clustering amplitude of LAEs on scales > 20 cMpc h^{-1}, while at scales < 5 cMpc h^{-1} the clustering of LAEs seems unaffected by the presence of the IGM. Although Gurung-López et al. (2020) studied the clustering of LAEs in a wide variety of scales, they did not take into account the local environment in which galaxies reside. The impact of IGM on the clustering of LAEs at small scales might differ in more extreme environments.

In this section, we explore the coupling between IGM and HAEs+LAEs in high-density environments by comparing the clustering of HAEs and HAEs + LAEs at small scales (< 10 cMpc h^{-1}) for the model with and without IGM included (IGM and noIGM model, respectively). We select central galaxies as protoclusters centres according to the host halo mass distribution at $z = 2.2$, $z = 3.0$, and $z = 5.7$. The selection results in 1048, 579, and 1564 candidates with halo masses of $M_{\rm halo} > 10^{13.2}$ M$_\odot$, $M_{\rm halo} > 10^{13}$ M$_\odot$, and $M_{\rm halo} > 10^{12}$ M$_\odot$ for $z = 2.2$, $z = 3.0$, and $z = 5.7$, respectively. These limits correspond to the peak of the host halo mass distribution of radio galaxies (Orsi et al. 2016). The limit for halo mass of central galaxies at $z = 5.7$ is somewhat arbitrary, but our findings are insensitive to these specific values; we arrive to the same conclusions when repeating our analysis varying the limits in 0.2 dex.

It is worth mentioning that at $z = 2.2$, the protoclusters candidates are equivalent to those of Section 3. We select HAEs and HAEs + LAEs inside a spherical shell of radius $r = 10$ cMpc h^{-1}

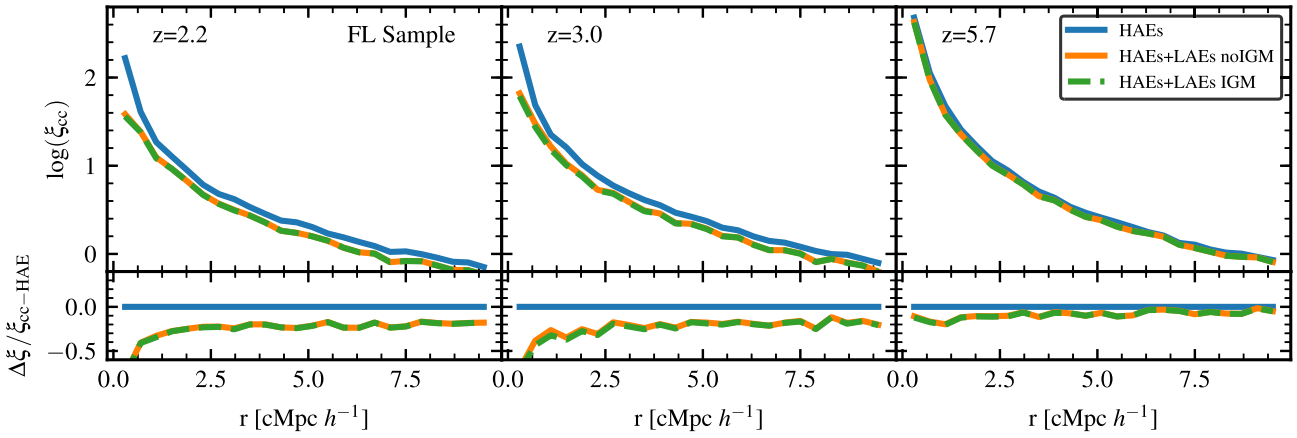

Figure 5. Cross-correlation functions for HAEs and HAEs + LAEs for $z = 2.2$ (left-hand panel), $z = 3.0$ (middle panel), and $z = 5.7$ (right-hand panel). The galaxies are classified as HAEs or LAEs following S17, where HAEs have EW > 18 Å and $L_{H\alpha} > 4.35 \times 10^{41}$ ergs^{-1}, and HAEs + LAEs are imposed to have also $L_{Ly\alpha} > 4.4 \times 10^{41}$ erg s^{-1} and EW > 15 Å (FL Sample defined in Section 3). Solid blue lines represent HAEs, which are not affected by IGM at any redshift. Dashed green and solid orange lines represent HAEs + LAEs from the model with and without IGM effect, respectively.

from the central object. To define the ELGs samples at different redshifts, we maintain the same luminosity and EW criteria that define FL and FN samples in Section 3. We caution the reader that, in this section, the spatial constraints to select ELGs are modified with respect to Section 3, thus the samples at $z = 2.2$ do not match exactly. However, the properties of both samples discussed further in this section are insensitive to the different spatial constraints aforementioned. Hence, we will keep the sample nomenclature defined in Section 3, specifying the redshift considered when necessary.

The number density of objects is computed as the sum over all protoclusters of HAEs and HAEs + LAEs with protocluster-centric distances less than 10 cMpc h^{-1}, divided by the sum of the volumes of the protoclusters. The values for both samples at all redshifts considered are listed in table 1.

We quantify the clustering as the cross-correlation function between halo mass selected central objects and ELGs, ξ_{cc}. This is estimated as

$$\xi_{cc}(r) = \frac{DD(r)}{N_c n_{gal} \Delta V(r)} - 1, \quad (4)$$

where $DD(r)$ is the total number of ELGs around central objects at a distance $r \pm \Delta r/2$, N_c is the total number of protocluster candidates at the corresponding redshift, n_{gal} is the mean number density of ELGs in the box, and $\Delta V(r)$ is the volume of a spherical shell of radius r and width Δr. This width corresponds to the bin size used to compute ξ_{cc}. As our simulation box is periodic, the pair counts are not affected by edge effects, so the use of estimators that rely on random set of objects is not necessary.

In Fig. 5, we show the clustering of the HAEs and HAEs + LAEs in the FL sample at three different redshifts, and compute the relative difference with respect to the clustering of the HAEs population as $\Delta\xi = (\xi_{cc} - \xi_{cc\text{-HAE}})/\xi_{cc\text{-HAE}}$. We find that for $z = 2.2$ and $z = 3.0$, HAEs + LAEs are ∼50 per cent less clustered than HAEs in the core of the protoclusters, and ∼20 per cent less clustered from $r \sim 2.5$ cMpc h^{-1} to the outskirts of the protocluster (left and middle panel of Fig. 5). At $z = 5.7$, the clustering of HAEs + LAEs is ∼15–20 per cent smaller in the core, and $\lesssim$ 10 per cent for $r \gtrsim$ 2 cMpc h^{-1}. In all cases, the results from the IGM model are basically indistinguishable from those obtained from the noIGM model: Given the combination of Ly α and H α luminosity thresholds, the vast majority of HAEs+LAEs in the noIGM model are also classified as HAEs + LAEs in the IGM model, hence their clustering are almost identical.

In Fig. 6, we analyse the metal content of the cold gas phase, Z_{gas}, SFR, and M_{halo} of ELGs located at distances of r < 10 cMpc h^{-1} from the protoclusters centres at $z = 2.2$, $z = 3.0$, and $z = 5.7$ for the FL sample. Observations suggest that Ly α emission is strongly dependent on the stellar mass, $M_\star$, of galaxies. Massive star-forming systems have higher gas mass content (Kereš et al. 2005) (which increases the scattering of Ly α photons), and have been forming stars longer, which leads to higher metal and dust content in the ISM. In fact, the anticorrelation between f_{esc} of Ly α photons with $M_\star$ and SFR has been reported at $z \sim 2$ (Matthee et al. 2016) and $3 < z < 4.6$ (Oyarzún et al. 2017).

In our case, both HAEs and HAEs + LAEs have low Z_{gas} at $z = 5.7$, and their probability density function (PDF) are very similar. This allows the escape of Ly α photons from both populations and promotes a selection of galaxies with similar SFR (lower left and middle panels of Fig. 6). HAEs + LAEs tend to inhabit haloes slightly less massive than HAEs, hence their clustering is smaller in the central and outer regions of the protoclusters, as it was already mentioned (right-hand panel of Fig. 5). The chemical enrichment of the ISM due to stellar evolution increases the metal content within galaxies with cosmic time, and thus, at lower redshifts, Ly α photons are more likely to escape from galaxies with intermediate metallicities ($-3 < \log(Z_{gas}/Z_\odot) < -2$) and intermediate SFRs ($9.0 < \log(\mathrm{SFR}\,[M_\odot\,\mathrm{Gyr}^{-1}\,h^{-1}]) < 9.8$), as we show respectively in the upper and middle panels of the left and middle columns in Fig. 6. In our model, HAEs + LAEs with intermediate Z_{gas} and intermediate SFR statistically inhabit DM haloes clearly less massive ($10.9 < \log(M_{halo}[M_\odot]) < 11.5$) than HAEs ($10.9 < \log(M_{halo}[M_\odot]) < 12.0$), both at $z = 2.2$ and $z = 3.0$. Hence, the clustering amplitude with respect to HAEs is lower.

We conclude that for the FL sample, RT processes that occur inside HAEs + LAEs raise a selection effect over galactic properties, which results in a strong decrease in the clustering amplitude with respect to the HAEs population at $z = 2.2$ and $z = 3.0$. We emphasize that IGM produces no enhancement of this effect, not even at high redshift.

In the case of the FN sample, when we lower the H α luminosity limit for HAEs and raise the Ly α luminosity limit for HAEs+LAEs with respect to the FL Sample, we find that the IGM effect becomes

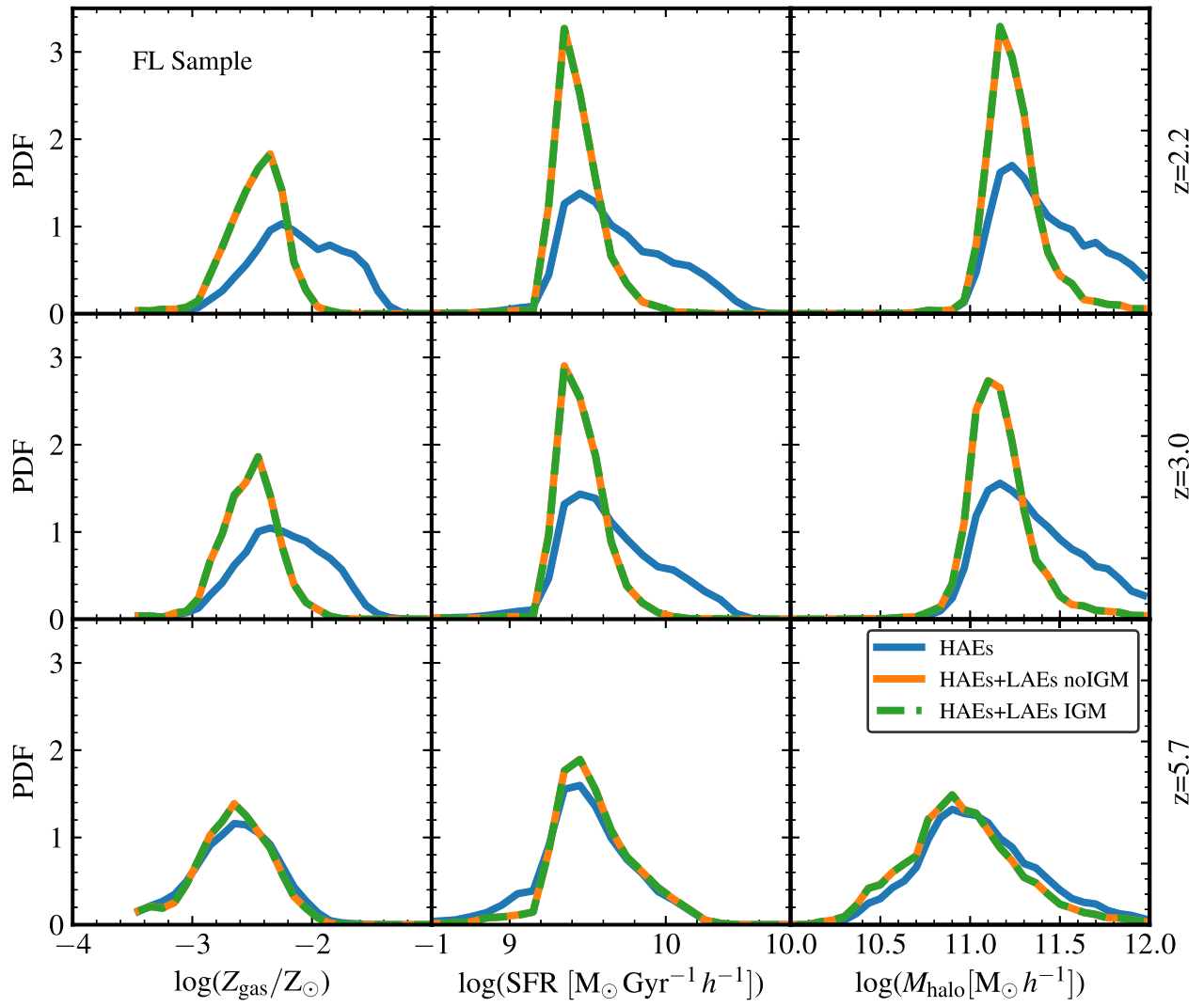

Figure 6. Cold gas metallicity (left column), star formation rate (middle column), and mass of DM halo (right column) for ELGs located at distances of r < 10 cMpc h^{-1} from protoclusters centre. Protoclusters are selected according to their DM halo mass, at three different redshifts: $z = 2.2$ (upper panels), $z = 3.0$ (middle panels), and $z = 5.7$ (lower panels). In each case, HAEs have EW > 18 Å and $L_{\mathrm{H}\alpha} > 43.5 \times 10^{41}$ erg s^{-1}, and HAEs + LAEs are imposed to have also $L_{\mathrm{Ly}\alpha} > 4.4 \times 10^{41}$ erg s^{-1} and EW > 15 Å (as in the FL Sample of Section 3). The properties of HAEs+LAEs are the same for the simulation with and without IGM effect included, hence the presence of the IGM has a negligible impact on the HAEs + LAEs population of the FL sample at all redshifts.

noticeable. The number density of HAEs + LAEs in the IGM model results lower than in the noIGM model, as can be noted in the second column of Table 1. In this case, a small difference arises in the clustering of HAEs + LAEs between both models (Fig. 7). At $z = 2.2$ and $z = 3.0$, HAEs + LAEs are less clustered than HAEs, as in the FL sample. For $r \leq 4$ cMpc h^{-1}, the IGM diminishes the clustering of HAEs + LAEs, and the slope is also less pronounced than in the FL sample.

At $z = 2.2$ and $z = 3.0$, HAEs still have higher gas metal content than HAEs + LAEs, as the FL sample (left column of Fig. 8). HAEs have $3 \lesssim \log(Z_{\mathrm{gas}}) \lesssim -1.5$, while HAEs + LAEs have $-3 \lesssim \log(Z_{\mathrm{gas}}) \lesssim -2$, and no dependence with the IGM is noticed. But the low H α luminosity threshold results in HAEs with low SFR, and the peak of the PDF is approximately $10^{8.7}$ M$_\odot$ Gyr^{-1} h^{-1} (middle column of Fig. 8), while the peak of the PDF for HAEs + LAEs is approximately $10^{9.3}$ M$_\odot$ Gyr^{-1} h^{-1}. However, HAEs can reach SFR $\sim 10^{10.4}$ M$_\odot$ Gyr^{-1} h^{-1}, 0.5 dex higher than HAEs + LAEs, which have values restricted to $10^9 \lesssim$ SFR [M$_\odot$Gyr^{-1} h^{-1}] $\lesssim 10^{9.9}$. The presence of the IGM only slightly decreases the SFR of HAEs + LAEs.

A similar behaviour is found for the mass of the DM haloes. The peak of the PDF for HAEs is approximately $10^{10.9}$ M$_\odot$ h^{-1}, while for HAEs + LAEs it is approximately $10^{11.1}$ M$_\odot$ h^{-1} (right column of Fig. 8). However, HAEs can inhabit more massive haloes than HAEs + LAEs, as they can reach values up to 10^{12} M$_\odot$ h^{-1}, while HAEs + LAEs reach up to $10^{11.5}$M$_\odot$ h^{-1}. More massive DM haloes are expected to locate towards the centre of the protocluster (Orsi et al. 2016), hence those galaxies in the FN sample that inhabit the most massive haloes and have higher SFRs dominate the clustering behaviour.

At $z = 5.7$, the FN sample have HAEs + LAEs that present higher metal content, higher SFR and inhabit more massive DM haloes than HAEs (lower panels of Fig. 8), hence the clustering of HAEs + LAEs results to be ∼30–10 per cent higher for $r < 2$ cMpc h^{-1} with respect to HAEs, for the model without IGM included. When IGM is considered, the clustering of HAEs + LAEs results in 20–10 per cent higher

Table 1. Number density of ELG around protoclusters, when using different EW, H α, and Ly α luminosity limits. In the **FL Sample**, HAEs have $L_{H\alpha} > 4.35 \times 10^{41}$ erg s^{-1}, while HAEs + LAEs are imposed to have also $L_{Ly\alpha} > 4.4 \times 10^{41}$ erg s^{-1}. In the **FN Sample**, HAEs have $L_{H\alpha} > 10^{41}$ erg s^{-1}, while HAEs + LAEs are imposed to have also $L_{Ly\alpha} > 1.5 \times 10^{42}$ erg s^{-1}. In both cases, HAEs have EW > 18 Å and LAEs have EW > 15 Å. The number density is computed as the number of galaxies inside 10 cMpc h^{-1} of all protoclusters, divided by the volume of all protoclusters at each redshift.

	FL Sample [$\times 10^{-3}$ cMpc$^{-3}h^{-3}$]			FN Sample [$\times 10^{-3}$ cMpc$^{-3}h^{-3}$]		
	HAEs	HAEs + LAEs noIGM	HAEs + LAEs IGM	HAEs	HAEs + LAEs noIGM	HAEs + LAEs IGM
$z = 2.2$	19.9	4.97	4.95	60.4	9.09	8.07
$z = 3.0$	26.7	7.79	7.68	76.9	13.3	10.3
$z = 5.7$	15.8	7.85	7.81	45.5	17.6	14.8

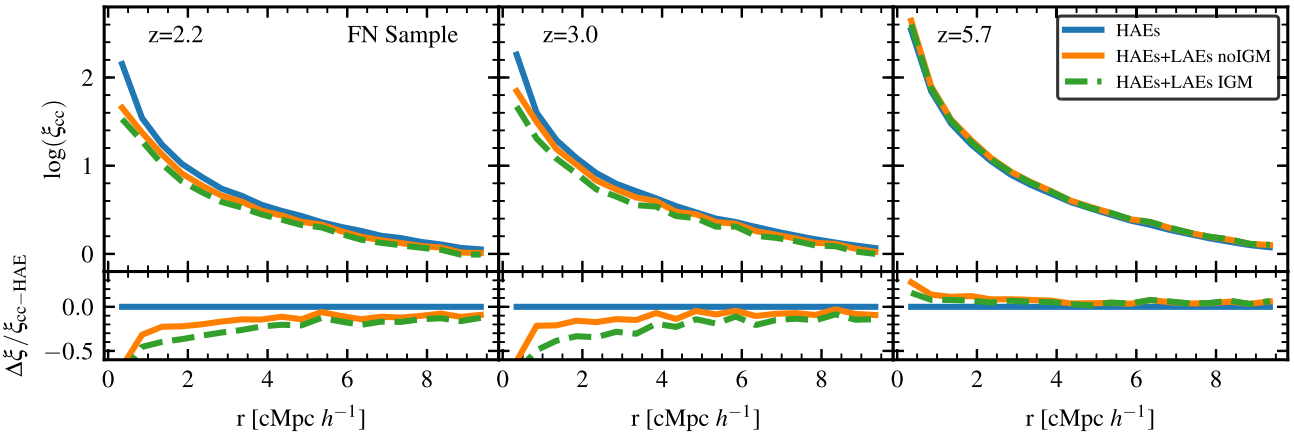

Figure 7. Same as Fig. 5. In this case, HAEs have EW > 18 Å and $L_{H\alpha} > 10^{41}$ erg s^{-1}, and HAEs + LAEs are imposed to have as $L_{Ly\alpha} > 1.5 \times 10^{42}$ erg s^{-1} and EW > 15 Å (as in FN Sample of Section 3).

than HAEs for $r < 2$ cMpc h^{-1}. For $r > 2.5$ cMpc h^{-1}, the clustering of HAEs and HAEs + LAEs is very similar for both models.

Figs 6 and 8 allow us to conclude that the radiative processes that take place inside galaxies shape the observable properties of ELGs, while IGM has only a minor impact. The IGM density (computed as described in Fig. 1) between 8 and 10 cMpch^{-1} from the protocluster centre spans between ∼5 and ∼12 times the mean IGM density of the simulation, reflecting that protoclusters are embedded in an extensive and overdense matter distribution. However, if we restrict to ELGs located at $r < 5$ cMpch^{-1}, we still do not appreciate a substantial difference in the properties for the simulations with and without IGM. The effect of the IGM on the Ly α transmission depends on the density of the IGM in which the ELGs reside. In particular, Gurung-López et al. (2020) ranked their LAE samples according to the IGM density where they reside, and split them into 3 sub-samples: underdense (below the percentile 33 of density), intermediate (between the 33rd and 66th percentiles), and overdense (above the 66th percentile). They show that at $z = 2.2$ ($z = 3.0$), the transmission for wavelengths bluer than Ly α ($\lambda \sim 1214$ Å) is 0.9 (0.4), 0.85 (0.2), and 0.8 (0.1) for underdense, intermediate, and overdense environments, respectively. For $z = 5.7$, the transmission remains below 1 per cent even in underdense regions. This means that IGM is attenuating LAEs specifically in overdense regions at $z = 2.2$ and $z = 3.0$, while at $z = 5.7$ the effect is the same throughout all environments. When this is combined with the permissive Ly α luminosity limit of the FL sample, the vast majority of HAEs + LAEs remain in the sample when the IGM is included. In the case of the FN sample, the low H α luminosity limit results in higher number densities (with respect to the FL sample) for HAEs and HAEs + LAEs in both models, as can be appreciated in Table 1, but given the restrictive Ly α luminosity limit of the FN sample, a higher proportion of HAEs + LAEs are excluded when the IGM is included.

From the results of our models, we conclude that the clustering of ELGs in high-density environments is clearly dominated by the RT processes inside galaxies, and the IGM plays a secondary role in decreasing the clustering, even at high redshift. A shallow and wide survey, targeting bright LAEs, is more prone to detect galaxies affected by the IGM rather than a deep small survey.

5 DISCUSSION AND CONCLUSIONS

We study the possible spatial segregation of LAEs with respect to HAEs around a wide sample of protoclusters, following the observational work performed by S17 on the protocluster USS1558-003, located at $z = 2.53$. With this aim, we create catalogues of ELGs that include Ly α radiative transfer of both the ISM and IGM, by combining a cosmological DM simulation (P-Millennium) with a semi-analytic model of galaxy formation (GALFORM). We define two samples of HAEs and HAEs + LAEs at $z = 2.2$. On the one hand, we built a sample designed to reproduce the same constraints imposed by the observational work (FL sample). In this sample, HAEs have line emission widths EW$_{H\alpha} > 18.6$ Å and luminosity $L_{H\alpha} > 4.35 \times 10^{41}$ erg s^{-1}, while HAEs + LAEs are imposed to have also EW$_{Ly\alpha} > 15$ Å and $L_{Ly\alpha} > 4.4 \times 10^{41}$ erg s^{-1}. On the other hand, we consider a sample designed to reproduce the observed surface density of HAEs and HAEs + LAEs (FN sample). In this sample, HAEs have $L_{H\alpha} > 10^{41}$ erg s^{-1}, while HAEs + LAEs have also $L_{Ly\alpha} > 1.5 \times 10^{42}$ erg s^{-1}, and we maintain the same EW cut as for the FL Sample. We also explore how the radiative transfer of

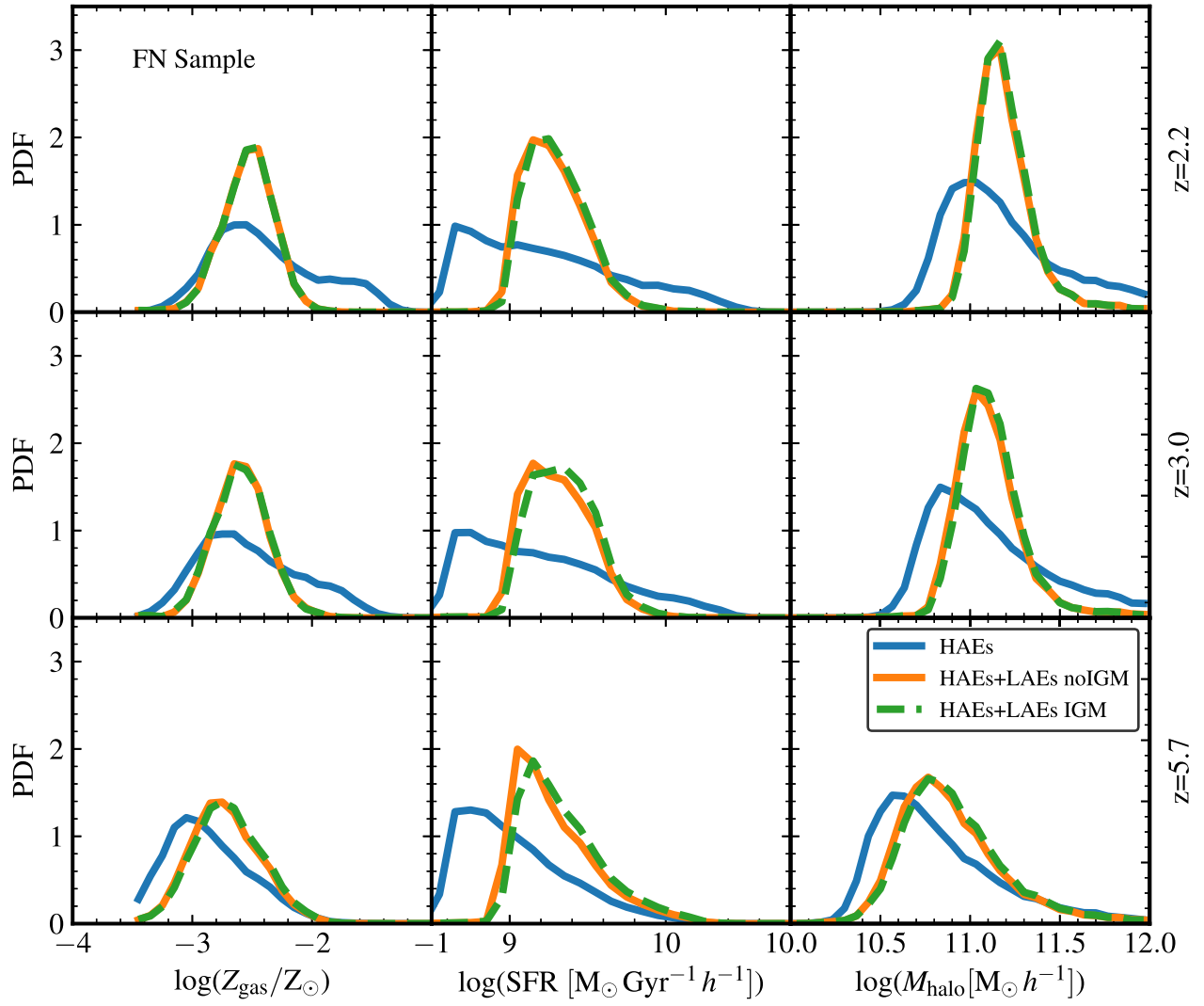

Figure 8. Same as Fig. 6. In this case, HAEs have EW $>$ 18 Å and $L_{\mathrm{H}\alpha} > 10^{41}$ erg s^{-1}, and HAEs + LAEs are imposed to have as $L_{\mathrm{Ly}\alpha} > 1.5 \times 10^{42}$ erg s^{-1} and EW $>$ 15 Å (as in FN Sample of Section 3).

the IGM affects the clustering of FL and FN samples at $z = 2.2$, $z = 3.0$, and $z = 5.7$, by comparing models with and without IGM radiative transfer effect included. Our main results are summarized as follows:

(i) We average the behaviour of simulated protoclusters at $z =$ 2.2 for both FL and FN samples, and do not find the high depletion of HAEs + LAEs in the densest regions of protoclusters present in USS1558-003.

(ii) Only ∼10 per cent of the simulated protoclusters are in consistency with the high HAEs + LAEs depletion present in USS1558-003, suggesting that the observational result could be subject to cosmic variance.

(iii) We analyse the clustering of ELGs in the FL and FN samples up to 10 cMpc h^{-1} from protoclusters centre. We find that RT processes inside galaxies create selection effects over galaxy properties for both samples. For the FL sample, HAEs + LAEs tend to have lower SFRs, lower metallicities, and inhabit less massive haloes than HAEs at $z = 2.2$ and $z = 3.0$.

(iv) In the FL sample, the clustering of HAEs + LAEs turns out to be ∼50 per cent lower than that of HAEs in the core of protoclusters ($r < 1$ cMpc h^{-1}), and ∼20 per cent lower in the outskirts ($r > 2.5$ cMpc h^{-1}). For $z = 5.7$, the clustering of HAEs + LAEs is between ∼10 and 20 per cent smaller than HAEs in the protocluster core, and less than 10 per cent smaller for $r > 2$ cMpc h^{-1}.

(v) The properties of HAEs + LAEs in the FL sample are not affected by the presence of the IGM in the model, hence the clustering of both models are almost identical. This indicates that in a survey with the capability to detect HAEs and LAEs with EW $\gtrsim$ 15 Å and $L \gtrsim 4.4 \times 10^{41}$ erg s^{-1} at $z \leq 5.7$ for both H α and Ly α, the clustering of HAEs + LAEs should not be affected by the presence of the IGM.

(vi) In the case of the FN sample, the low H α luminosity threshold allows the inclusion of HAEs with lower SFRs and less massive DM haloes than the FL sample. Nevertheless, near the protocluster centre, the clustering is dominated by massive DM haloes, which tend to be HAEs hosts. The clustering of HAEs + LAEs is ∼40 per cent lower than HAEs in the core of protoclusters, and ∼15 per cent lower for $r > 4$ cMpch^{-1} at $z = 2.2$ and $z = 3.0$.

The presence of the IGM results in a lower clustering amplitude and less pronounced slope at $r < 4$ cMpc h^{-1} with respect to the noIGM model.

(vii) Due to the restrictive Ly α luminosity threshold of the FN sample, at $z = 5.7$, HAEs+LAEs tend to have higher SFRs and metallicities, and inhabit more massive DM haloes than pure HAEs, although the difference is small. This results in a ∼5–30 per cent higher clustering amplitude for HAEs + LAEs at $r < 2$ cMpc h^{-1}. When the IGM is included, the clustering varies only from ∼5 to 15 per cent in the central region. For $r > 2$ cMpc h^{-1}, the clustering of HAEs and HAEs + LAEs is very similar.

S17 suggest that the accretion of cold gas streams directly into the core of the protocluster could prevent Ly α photons from escaping, resulting in a lack of LAEs in high-density regions. Although gas accretion along the line of sight could enhance the depletion, we claim that, on average, HAEs and HAEs + LAEs can trace similar local densities at $z = 2.2$. Moreover, S17 suggest that, as mean projected distances are small in high-density regions, Ly α photons that escape from HAEs may penetrate the CGM of a foreground galaxy increasing the depletion. Although our model does not include RT of an extended CGM component around each galaxy, we expect this effect to be small. Observations suggest that ionized, dense ($n > 1\,\mathrm{cm}^{-3}$), and relatively cold ($\mathrm{T} \sim 10^4$ K) reservoirs of gas should surround massive galaxies at $z \geq 2$ (Cantalupo 2017). In some cases, such as radio-loud and radio-quiet quasars, the Ly α emission from these gaseous haloes can be traced out to 100 kpc of galactic radii. However, in faint LAEs (with surface brightness $\mathrm{SB} \gtrsim 4 \times 10^{21}$ ergs^{-1}cm^{-2}arcsec^{-2}) Ly α haloes can reach ∼ 60 kpc of galactic radii (see Wisotzki et al. 2018; Witstok et al. 2019, and references therein). The properties of the CGM around high redshift galaxies are still matter of debate, and the results of our model are not sensible to the effect that CGM might produce on background LAEs. Moreover, the resolution of our IGM model presumes a limitation to our analysis. A reduction of the grid by a factor of 4 would be ideal to reach the scale of the local density $< a >_{5\mathrm{th}}$ at which S17 finds the depletion of HAEs + LAEs, thus allowing to confirm our results. Multi narrow-band surveys such as Javalambre-Physics of the Accelerated Universe Astrophysical Survey (J-PAS; Benitez et al. 2014) will provide a large ELGs sample at $z \sim 2$, where our model suggests that LAEs can inhabit high densities traced by HAEs.

The properties and clustering of HAEs + LAEs residing in high-density regions depend mainly on the RT effects that happen inside them. IGM presence results in a second order effect, which depends on the EW and luminosity criteria that defines the ELG sample. The next generation of multiwavelength imaging surveys will be able to characterize high redshift environments of overdense regions with unprecedented detail. In particular, a number of these will rely on ELGs to map matter distribution. Spectroscopic surveys such as HETDEX (Hill et al. 2008) and DESI (Levi et al. 2013) have sufficient spectral resolution to probe the scales on which our model predicts that RT of the ISM induces a selection effect on ELGs properties, giving rise to a different clustering amplitude.

ACKNOWLEDGEMENTS

We thank the referee for his/her useful comments, that helped improve this manuscript. TH acknowledges *Consejo Nacional de Investigaciones Científicas y Técnicas* (CONICET), Argentina, for their supporting fellowships. AO acknowledges support from project AYA2015-66211-C2-2 of the Spanish *Ministerio de Economía, Industria y Competitividad.* SAC acknowledges funding from CONICET (PIP-0387), and Universidad Nacional de La Plata (G11-150), Argentina. This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Sklodowska-Curie grant agreement No. 734374; project acronym: LACEGAL. The programmes and figures had been developed under PYTHON language (Van Rossum & Drake 2009), using specially NUMPY (Van Der Walt, Colbert & Varoquaux 2011) and MATLOTLIB (Hunter 2007).

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Adams S. M., Martini P., Croxall K. V., Overzier R. A., Silverman J. D., 2015, MNRAS, 448, 1335
Baldry I. K., Balogh M. L., Bower R. G., Glazebrook K., Nichol R. C., Bamford S. P., Budavari T., 2006, MNRAS, 373, 469
Baugh C. M. et al., 2019, MNRAS, 483, 4922
Benitez N. et al., 2014, preprint (arXiv:1403.5237)
Bluck A. F. L., Maiolino R., Sánchez S. F., Ellison S. L., Thorp M. D., Piotrowska J. M., Teimoorinia H., Bundy K. A., 2020, MNRAS, 492, 96
Calzetti D., 2013, Star Formation Rate Indicators. p. 419
Cantalupo S., 2017, in Fox A. J., Davé R., eds, Astrophysics and Space Science Library, Vol. 430, Gas Accretion on to Galaxies. Springer-Verlag, Berlin, p. 195
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Cooray A., 2005, MNRAS, 363, 337
Darvish B., Mobasher B., Sobral D., Scoville N., Aragon-Calvo M., 2015, ApJ, 805, 121
Dijkstra M., 2014, Publ. Astron. Soc. Aust., 31, e040
Dijkstra M., 2017, preprint (arXiv:1704.03416)
Gonzalez-Perez V. et al., 2020, MNRAS, 498, 1852
Guaita L. et al., 2011, ApJ, 733, 114
Gurung-López S., Orsi Á. A., Bonoli S., Baugh C. M., Lacey C. G., 2019a, MNRAS, 486, 1882
Gurung-López S., Orsi Á. A., Bonoli S., 2019b, MNRAS, 490, 733
Gurung-López S., Orsi Á. A., Bonoli S., Padilla N., Lacey C. G., Baugh C. M., 2020, MNRAS, 491, 3266
Hatch N. A., Kurk J. D., Pentericci L., Venemans B. P., Kuiper E., Miley G. K., Röttgering H. J. A., 2011, MNRAS, 415, 2993
Hayashi M., Kodama T., Tadaki K.-i., Koyama Y., Tanaka I., 2012, ApJ, 757, 15
Hayashi M., Kodama T., Tanaka I., Shimakawa R., Koyama Y., Tadaki K.-i., Suzuki T. L., Yamamoto M., 2016, ApJ, 826, L28
Hayes M. et al., 2010, Nature, 464, 562
Hayes M., Schaerer D., Östlin G., Mas-Hesse J. M., Atek H., Kunth D., 2011, ApJ, 730, 8
Hill G. J. et al., 2008, ASP Conf. Ser. Vol. 399, Panoramic Views of Galaxy Formation and Evolution. Astron. Soc. Pac., San Francisco, p. 115
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Jiang L., Helly J. C., Cole S., Frenk C. S., 2014, MNRAS, 440, 2115
Kashikawa N., Kitayama T., Doi M., Misawa T., Komiyama Y., Ota K., 2007, AJ, 663, 765
Kennicutt, Robert C. J., 1998, ARA&A, 36, 189
Kereš D., Katz N., Weinberg D. H., Davé R., 2005, MNRAS, 363, 2
Koyama Y., Kodama T., Tadaki K.-i., Hayashi M., Tanaka M., Smail I., Tanaka I., Kurk J., 2013, MNRAS, 428, 1551
Kurk J. D., Pentericci L., Röttgering H. J. A., Miley G. K., 2004, A&A, 428, 793
Lacey C. G. et al., 2016, MNRAS, 462, 3854
Le Fevre O., Deltorn J. M., Crampton D., Dickinson M., 1996, ApJ, 471, L11
Levi M. et al., 2013, preprint (arXiv:1308.0847)
Macuga et al., 2019, ApJ, 874, 54
Matthee J., Sobral D., Oteo I., Best P., Smail I., Röttgering H., Paulino-Afonso A., 2016, MNRAS, 458, 449

McQuinn M., 2016, ARA&A, 54, 313
Mo H. J., Yang X., van den Bosch F. C., Jing Y. P., 2004, MNRAS, 349, 205
Orsi A., Lacey C. G., Baugh C. M., 2012, MNRAS, 425, 87
Orsi Á., Padilla N., Groves B., Cora S., Tecce T., Gargiulo I., Ruiz A., 2014, MNRAS, 443, 799
Orsi Á. A., Fanidakis N., Lacey C. G., Baugh C. M., 2016, MNRAS, 456, 3827
Osterbrock D. E., 1989, Astrophysics of Gaseous Nebulae and Active Galactic Nuclei. University Science Books, Mill Valley, CA, p. 408
Ota K. et al., 2018, ApJ, 856, 109
Overzier R. A., 2016, A&AR, 24, 14
Overzier R. A. et al., 2006, ApJ, 637, 58
Overzier R. A. et al., 2008, ApJ, 673, 143
Overzier R. A., Guo Q., Kauffmann G., De Lucia G., Bouwens R., Lemson G., 2009, MNRAS, 394, 577
Oyarzún G. A., Blanc G. A., González V., Mateo M., Bailey, John I. I., 2017, ApJ, 843, 133
Pentericci L., Roettgering H. J. A., Miley G. K., Carilli C. L., McCarthy P., 1997, A&A, 326, 580
Planck Collaboration, 2016, A&A, 594, A13
Shi K. et al., 2019, ApJ, 879, 9
Shimakawa R., Kodama T., Tadaki K. I., Tanaka I., Hayashi M., Koyama Y., 2014, MNRAS, 441, L1
Shimakawa R. et al., 2017a, MNRAS, 468, L21
Shimakawa R. et al., 2017b, MNRAS, 468, 1123
Sobral D., Kohn S. A., Best P. N., Smail I., Harrison C. M., Stott J., Calhau J., Matthee J., 2016, MNRAS, 457, 1739
Sobral D. et al., 2017, MNRAS, 466, 1242
Van Der Walt S., Colbert S. C., Varoquaux G., 2011, Comput. Sci. Eng., 13, 22
Van Rossum G., Drake F. L., 2009, Python 3 Reference Manual. CreateSpace, Scotts Valley, CA
Venemans B. P. et al., 2002, ApJ, 569, L11
Venemans B. P. et al., 2007, A&A, 461, 823
Wisotzki L. et al., 2018, Nature, 562, 229
Witstok J., Puchwein E., Kulkarni G., Smit R., Haehnelt M. G., 2019, preprint (arXiv:1905.06954)
Wold M., Armus L., Neugebauer G., Jarrett T. H., Lehnert M. D., 2003, AJ, 126, 1776

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