

THE IMPACT OF COSMIC VARIANCE ON CONSTRAINING THE MASS AND OCCUPATION FRACTION OF DARK MATTER HALOS HOSTING LYMAN- α EMITTERS AT $Z \sim 3$

JAIME E. FORERO-ROMERO¹ AND JULIAN E. MEJÍA-RESTREPO^{2,3}

¹ Departamento de Física, Universidad de los Andes, Cra. 1 No. 18A-10, Edificio Ip, Bogotá, Colombia

² Departamento de Astronomía, Universidad de Chile, Camino el Observatorio 1515, Santiago, Chile

³ FAcCom-Instituto de Física-FCEN, Universidad de Antioquia, Calle 70 No. 52-21, Medellín, Colombia

Submitted for publication in *ApJL*

ABSTRACT

We study the impact of cosmic variance in constraining the mass and occupation fraction of dark matter halos hosting Ly α Emitting Galaxies at high redshift. We use an N-body simulation to construct mock fields with the same typical size of observed fields at $z = 3.1$ to match the observed number density distribution across fields and the angular correlation function. In our model a dark matter halo with mass in the range $M_{\min} < M_h < M_{\max}$ can only host one detectable LAE with a probability $0.1 \leq f_{\text{occ}} \leq 1.0$. Our analysis shows that the clustering and number density information are insufficient to impose a tight constraint on the occupation fraction. On the other hand, the minimum mass and maximum mass can be constrained to the range $M_{\max} < 10^{12} h^{-1} M_{\odot}$ and $10^{10.2} h^{-1} M_{\odot} \leq M_{\min} \leq 10^{11.5} h^{-1} M_{\odot}$. Additionally, we find that the consistent models have a narrow mass range, $\Delta M \equiv \log_{10} M_{\max} - \log_{10} M_{\min}$, smaller than 1.0 dex, with a majority in the range $\Delta M < 0.5$ dex. These results fall into constraints already presented in the literature. However, the wide range of values for the occupation fraction and the narrow mass range ΔM are novel results allowed by taking into account the influence of cosmic variance on the statistics derived from observations.

Subject headings: cosmology: theory cosmology: large-scale structure of universe galaxies: formation galaxies: high-redshift galaxies: statistics galaxy: haloes

1. INTRODUCTION

Lyman- α emitting galaxies (LAEs) are helpful in a diverse range of subjects in extragalactic astronomy. LAEs can be used as probes of reionization (Dijkstra et al. 2011), tracers of large scale structure (Koehler et al. 2007), signposts for low metallicity stellar populations, markers of the galaxy formation process at high redshift (Dayal et al. 2009; Forero-Romero et al. 2012) and tracers of active star formation (Guaita et al. 2013).

Capitalizing LAEs observations requires an understanding of their place in the context of a given structure formation model in a cosmological context. Under the current paradigm the dominant matter content of the Universe is to be found in dark matter (DM) and each galaxy is thought to be hosted by larger dark matter structure known as a halo. (Peebles 1980; Springel et al. 2005).

Galaxy formation models find that halo mass predicts with high accuracy galactic properties such as stellar mass and star formation rate (Behroozi et al. 2013b). This suggests that the physical processes that regulate the star formation cycle are dependent on halo mass. For that reason, finding the typical dark matter halo mass hosting LAEs represents an advance to understand the nature of this galaxy population in the context of Lambda Cold Dark Matter (Λ CDM) paradigm.

Some theoretical attempts to solve this problem using an ab-initio approach. They start from the DM halo population to infer the intrinsic star formation rates and Ly α luminosities. From these values they estimate the amount of Ly α photons that escape each galaxy and compute the observed luminosity for each galaxy. These models can predict different observables including: the

luminosity function, correlation function and the equivalent width distributions. Such modelling has been implemented in semi-analytic models (Garel et al. 2012; Orsi et al. 2012) and full N-body hydrodynamical simulations (Laursen & Sommer-Larsen 2007; Dayal et al. 2009; Forero-Romero et al. 2011; Yajima et al. 2012).

However, these calculations involve many uncertain steps such as the estimation of the escape fraction, that is the fraction of Ly α photons that escape the galaxy to the observer. Given the resonant nature of the Ly α line, the escape fraction is sensitive to the dust contents, density, temperature, topology and kinematics of the neutral Hydrogen in the interstellar medium (ISM). Solving the radiative transfer of Ly α photons in the ISM requires Monte Carlo simulations. The process of finding a consensus on the expected value for the Ly α escape fraction in high redshift galaxies is still matter of ongoing research (Neufeld 1991; Verhamme et al. 2006; Forero-Romero et al. 2011; Dijkstra & Kramer 2012; Laursen et al. 2013; Orsi et al. 2012).

A different approach to infer the typical mass of halos hosting LAEs is based on the spatial clustering information. This approach uses the fact that in CDM cosmologies the spatial clustering of galaxies on large scales is entirely dictated by the halo distribution (Colberg et al. 2000), which in turn has a strong dependence on halo mass. Using measurements of the angular correlation function of LAEs, observers have put constraints on the typical mass and occupation fraction of the putative halos hosting these galaxies (Hayashino et al. 2004; Gawiser et al. 2007b; Nilsson et al. 2007; Ouchi et al. 2010). In these studies the observations are done on fields of ~ 1 deg² and the conclusions derived on the halo host mass

do not delve too deeply into the possible impact of cosmic variance.

Recently Yamada et al. (2012) observed a wide area of 2.4 deg^2 under homogeneous instrumental and data reduction conditions. This data set is constructed from 12 different sub-fields that allows us to use clustering statistics and the cosmic variance among fields to constrain the mass and occupation fraction of halos hosting LAEs.

In this paper we investigate the impact of cosmic variance in constraining the mass and occupation fraction of halos hosting LAEs. We do this by constructing mock catalogs from cosmological simulations. Our method populates each halo in the simulation with a LAE without predicting a $\text{Ly}\alpha$ luminosity. This bypasses all the physical uncertainties associated to star formation and radiative transfer. We build mock surveys following the geometry of the observed fields by Yamada et al. (2012) to compared them against observations in terms of the number density distribution and the angular correlation function. This allows us to find a range of parameters of our model that are consistent with observations while taking into account cosmic variance.

This paper is structured as follows. In the next section we present the simulation and the model used to produce the mock catalogs. We also list the criteria used to compare the mocks against observations. In §3 we present the main results for the halo mass and occupation fraction. We continue in §4 with a discussion under the light of other constraints already presented in the literature and theoretical considerations of galaxy formation studies. Finally, we present our conclusions in §5.

Throughout this paper we assume a Λ CDM cosmology with the following values for the cosmological parameters, $\Omega_m = 0.27$, $\Omega_\Lambda = 0.73$ and $h = 0.70$, corresponding to the matter density, vacuum density and the Hubble constant in units of $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

2. METHODOLOGY

Our method is based on the comparison of observations and mock catalogs. We use two different kinds of statistics to perform the comparison: (i) the distribution of the surface number density across fields and (ii) the angular correlation function.

In the next subsections we describe in detail the four key elements in our work-flow. First we present (in two sub-sections) the observations we take as a benchmark. Second, we describe the main characteristics of the N-body simulation and the halo catalogs we use. Third, we recount the important parameters of the simplified model that we use to populate the halo catalogs with LAEs. Finally, we describe some of the statistical tests we adopt to compare observations and mocks.

2.1. Observational constraints

The primary observational constraint we use in this paper is the LAE number density information at $z = 3.1$ obtained by the panoramic narrow-band survey presented by Yamada et al. (2012) from a survey conducted with the Subaru 8.2m telescope and the Subaru Prime Focus Camera, which has a field of view covering $34 \times 27 \text{ arcmin}$, corresponding to a comoving scale of $46 \times 35 \text{ Mpc } h^{-1}$ at $z = 3.09$. The narrow band filter used in the survey is centered at 4977 \AA with 77 \AA width, corresponding

to the redshift range $z = 3.062\text{--}3.125$ and $41 \text{ } h^{-1} \text{ Mpc}$ comoving scale for the detection of the Lyman- α line centered at $z = 3.09$. The authors reported a total 2161 LAEs with an observed equivalent width, in the observer frame, larger than 190 \AA over a total survey area of 2.42 deg^2 that includes 12 sub-fields, this corresponds to an average surface number density of $0.24 \pm 0.01 \text{ arcmin}^{-2}$.

The survey covered four independent fields:

1. The SSA22 field of 1.38 deg^2 with 1394 detected LAEs. This large field is composed by 7 sub-fields.
2. The Subaru/*XMM-Newton* Deep Survey (SXDS)-North, -Center and -South fields, with a total of 0.58 deg^2 and 386 LAEs (3 sub-fields).
3. The Subaru Deep Field (SDF) with 0.22 deg^2 and 196 LAEs (1 sub-field), and
4. The Great Observatory Optical Deep Survey (GOODS-N) field with 0.24 deg^2 and 185 LAEs (1 sub-field).

There is abundant observational work done on LAEs at redshift $z = 3.1$ (Kudritzki et al. 2000; Matsuda et al. 2005; Gawiser et al. 2007a; Nilsson et al. 2007; Ouchi et al. 2008). However, we decide to focus on the data from Yamada et al. (2012) because it has the largest covered area with homogeneous instrumentation conditions (telescope, narrow band filter), data reduction pipeline and conditions to construct the LAE catalog. This ensures that the number density variations among fields are *not* due to different observational conditions or criteria to construct the catalogs.

A secondary benchmark is the angular correlation function (ACF). Yamada et al. (2012) does not report an ACF measurement for their fields. Instead we use the results by Ouchi et al. (2008) who reported the ACF over a region of 1 deg^2 over the SXDS field.

There are some differences between Ouchi et al. (2008) observations and those by Yamada et al. (2012). The details in the color selection, corresponding limiting luminosities and EW thresholds are different in these references. In the case of Ouchi et al. (2008) the number density is $0.099 \pm 0.005 \text{ arcmin}^{-2}$, which is 50% lower than the median value of 0.20 arcmin^{-2} for the general fields in Yamada et al. (2012). In spite of these differences Ouchi et al. (2008) provides the most similar conditions to Yamada et al. (2012).

2.1.1. The adequacy of including SSA22 as an observational benchmark

The field SSA22 has been known to harbor a significant galaxy overdensity. Yamada et al. (2012) estimates that the densest sub-field is likely to be a rare density peak with $3 - 4\sigma$ significance on the scale of $\sim 60 \text{ } h^{-1} \text{ Mpc}$.

All the other sub-fields in SSA22 are average. They have a similar number density as other blank fields. The left panel in Figure 1 shows the integrated distribution of LAEs number density in all the fields reported in Yamada et al. (2012). It is evident that there is only one sub-field that stands out as an outlier in the distribution, corresponding to the densest field in SSA22. All the other fields form a continuous distribution in number density.

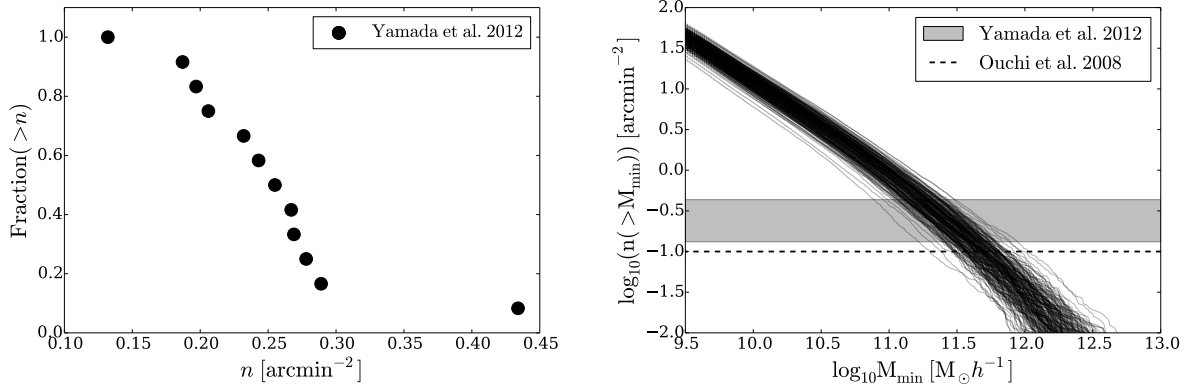


FIG. 1.— Left panel. Cumulative distribution of LAE number densities in all the fields observed by Yamada et al. (2012). The point with the highest surface density corresponds to the densest sub-field in the SSA22 field. Right panel. Surface density of dark matter halos as a function of a minimum halo mass. Each line represents one of the 210 volumes of dimensions $46 \times 35 \times 41 h^{-3} \text{ Mpc}^3$ in the Bolshoi simulation. The horizontal gray band represents the range of surface densities observed for LAEs at $z = 3.1$ as reported by Yamada et al. (2012) and the dashed line the observational results by Ouchi et al. (2008).

Therefore, using the whole SSA22 region as a benchmark in the number density distribution does not impose any bias. However, one has to keep in mind that it is not plausible that the full number density distribution from a simulation will include a very dense field. For this reason we do not require the presence of a very dense field such as SSA22-1 to judge a mock distribution as successful. The statistical test we use to compare mock distributions against observations does not require such a perfect match.

2.2. Simulation and halo catalogs

The Bolshoi simulation (Klypin et al. 2011) we use in this paper was performed in a cubic volume of $250 h^{-1} \text{ Mpc}$ on a side. The dark matter distribution is sampled using 2048^3 particles, which translates into a particle mass of $m_p = 1.35 \times 10^8 h^{-1} M_{\odot}$. The cosmological parameters are consistent with a WMAP5 and WMAP7 data with a matter density $\Omega_m = 0.27$, cosmological constant $\Omega_{\Lambda} = 0.73$, dimensionless Hubble constant $h = 0.70$, slope of the power spectrum $n = 0.95$ and normalization of the power spectrum $\sigma_8 = 0.82$ (Komatsu et al. 2009; Jarosik et al. 2011).

We use halo catalogs constructed with a Friend-of-Friends (FOF) algorithm with a linking length of 0.17 times the inter-particle distance. The catalogs were obtained from the publicly available Multidark database¹ (Riebe et al. 2013). For each halo in the box we store its comoving position in the box (3-D coordinates) and FOF mass. We focus our work on halos more massive than $1 \times 10^{10} h^{-1} M_{\odot}$ that are resolved with at least 70 particles, the reasons for this choice are explained in the next sub-section.

2.3. A model to populate halos with LAEs

We assume that a dark matter halo can only host one detectable LAE at most. This is consistent with theoretical analysis of the correlation function (Jose et al. 2013) and observations that confirm a lack of class pairs in LAEs Bond et al. (2009).

There are three parameters that decide whether a halo host a LAE: the lower and upper bounds for the mass range ($M_{\min} < M_h < M_{\max}$) and the fraction (f_{occ}) of such halos that host a detectable LAE. A physical interpretation of the occupation fraction f_{occ} convolves two phenomena: the actual presence of a star forming galaxy in a halo and its detectability as a LAE.

Our model does not assign a luminosity or escape fraction for each LAE in order to maintain theoretical uncertainties to a minimum. With this flexibility we can explore a wide range of possible masses for the host halos without any strong theoretical prejudice regarding the details of star formation and Ly α escape fraction in high-redshift galaxies.

In what follows we note by the letter \mathcal{M} a model defined by a particular choice of the three scalar parameters M_{\min} , M_{\max} and f_{occ} . For each model \mathcal{M} we create a set of mock fields from disjoint volumes in the simulation. Each volume has the same geometry probed by Suprime-CAM and the narrow band filter, namely rectangular cuboids of dimensions $46 \times 35 \times 41 h^{-3} \text{ Mpc}^3$ where the last dimension goes in the redshift direction. This corresponds to a total area of 880 arcmin^2 in each mock field. We construct a total $5 \times 7 \times 6 = 210$ of such volumes from a snapshot in the Bolshoi simulation. In each mock field a LAE is assigned to the position of a dark matter halo if the halo mass is in the range allowed by the model $M_{\min} < M_h < M_{\max}$ and a random variable taken from an homogeneous distribution $0 \leq \xi < 1$ is smaller than the occupation fraction $\xi < f_{\text{occ}}$.

Next we construct mock surveys by making groups of 12 mock fields out of the 210 available volumes. In total 15 mock surveys are constructed for each model \mathcal{M} . The grouping of the 12 mock fields into a mock catalog is done in two different ways. The first is called *match* because it follows the clustering of the observational fields. From the 12 mock fields, 7 are constructed from contiguous fields in the simulation to mimic the SSA22 region, 3 are also contiguous between them but not to the first 7 fields to mimic the SXDS fields and finally 2 non-contiguous fields to imitate the SDF and GOODS-North field. The second way to group the mock fields is called *random*, whereby all the 12 fields are selected in such a way as

¹ <http://www.multidark.org/MultiDark/>

to avoid that any two volumes are contiguous. In this paper we only report the results obtained by the `match` method since we do not find any significant difference with respect to the `random` method.

Figure ?? shows the spatial distribution for one mock survey constructed using the `match` method. Each field corresponds to one of the observational fields. The model parameters to build the mock are $M_{\min} = 10^{10.4} h^{-1} M_{\odot}$, $M_{\max} = 10^{10.5} h^{-1} M_{\odot}$ and $f_{\text{occ}} = 0.1$. The figure shows only one out of the 15 different mock surveys that are constructed for each model. We note that we only use $15 \times 12 = 180$ mock fields out of the total of 210 available sub-volumes. The reason is that the `match` method imposes constraints on the way the 7 fields mimicking the SSA22 can be distributed. This restriction makes unable some of the sub-volumes in the box. We decide to keep the number of mock surveys fixed to 15 also for the `random` method in order to allow a fair comparison between the two methods.

2.4. Exploring and selecting good models

We make a thorough exploration of the parameter space for the models \mathcal{M} where $\log_{10} M_{\min}$ takes 30 values from 10.0 up to 12.9 with an even spacing of 0.1 dex, $\log_{10} M_{\max}$ takes values in the same range as $\log_{10} M_{\min}$ with a displacement of 0.1 dex in the whole range. The occupation fraction f_{occ} takes 10 different values from 0.1 to 1 regularly spaced by 0.1. In total the number of different models \mathcal{M} that are explored is $30 \times 30 \times 10 = 9000$.

The lower limit for the parameter M_{\min} is set by the minimum occupation fraction we decide to consider. At $M_{\min} = 10^{10} h^{-1} M_{\odot}$ the halo number density around that mass range is ~ 10 times higher than the observational constraints for LAEs. This means that models in that mass range and an occupation fraction $f_{\text{occ}} = 0.1$ have the possibility to be compatible with observations. Lower values for M_{\min} require $f_{\text{occ}} < 0.1$, which are not considered in this paper. In turn exploring occupation fractions on the order of $f_{\text{occ}} = 0.01$ only makes sense for halo populations in the mass range of $10^9 h^{-1} M_{\odot}$ which are abundant enough to fit observations with a very low occupation fraction. In our case, this is a mass range unresolved by the Bolshoi simulation and therefore cannot be considered in the present study.

For each mock survey generated in a given model \mathcal{M} we compute the surface density in its 12 mock fields. We perform a Kolmogorov-Smirnov (KS) to compare these values against the 12 observational values. This tests gives us a value $0 < P < 1$ to reject the null hypothesis, namely that two data sets come from the same distribution. In this paper we consider that for values $P > 0.05$ the two distributions can be thought as coming from the same distribution.

We begin by considering that a model \mathcal{M} with at least one mock survey (out of 15) consistent with observations has viable parameters to host LAEs. From that we consider a stronger constraints to reduce the number of models by asking that all the 15 mocks to be consistent with observations and analyze again the properties of the resulting models. Finally we add the ACF as an additional constraint and consider all the models having their 15 mocks consistent with observations.

The ACF is computed using the Landy & Szalay estimator (Landy & Szalay 1993) on fields of size 1 deg^2 to

be compared against the results reported by Ouchi et al. (2010).

The observed and mock ACF are fit to a power-law function:

$$\omega(\theta) = \left(\frac{\theta}{\theta_0} \right)^{-\beta}, \quad (1)$$

where θ_0 and β are free parameters. The fit is done using a least square minimization procedure. For each mock field we obtain a covariance matrix that gives us the uncertainty in the parameters β and θ_0 . We consider that a mock field is consistent with observations if the two parameters β and θ_0 are equal within a $1\text{-}\sigma$ range. We perform this comparison only with the fields constructed using the `match` method.

3. RESULTS

The main purpose of this section is to show how the different observational constraints narrow down the parameters space of allowed models. Each sub-section presents the effect of adding a new piece of observational or statistical evidence.

3.1. Dark Matter Halo Number Density

The right panel in Figure 1 shows the integrated dark matter halo surface density as a function of minimum halo mass M_{\min} . Each line corresponds to one of the 210 sub-volumes in the Bolshoi simulation. The gray band indicates the surface density values for LAEs allowed by observations (Yamada et al. 2012). The dashed lines represent the average values in the fields observed by (Ouchi et al. 2008).

This plot allows us to understand why only a specific range of models \mathcal{M} can be expected to be consistent with observations. From Figure 1 we can read that models with a minimum mass $M_{\min} > 10^{12} h^{-1} M_{\odot}$ always have a surface number density lower than the observational constrain, making them incompatible with observations; there are simply too few halos compared to observed LAEs. The opposite is true in models with $M_{\min} < 10^{11.0} h^{-1} M_{\odot}$, which have a surface number density larger observations. In those cases the maximum mass M_{\max} and the occupation fraction $f_{\text{occ}} < 1.0$ can be tuned in order to lower the halo number density to match observations.

Figure 1 also illustrates the impact of cosmic variance. At fixed minimum mass there is a scatter of $0.3\text{--}0.6$ dex in the number density abundance, which is of the same order of magnitude as the scatter in the observational data. As a consequence, the variation in the number density in mocks for models with the same mass range and occupation fraction can be by factors of $\sim 2\text{--}5$. This scatter induced by cosmic variance is naturally included in the mock construction process.

3.2. Maximally Consistent Models

We impose two major constraints in our models. First, we require a models to have all its 15 mocks with a number density distribution consistent with observations. Second, we require that the average values among all mocks for the $\theta_0\text{--}\beta$ values in the correlation function fit to be consistent with observations (Ouchi et al. 2008).

The left panel in Figure 2 shows the results for the best estimates of $\theta_0\text{--}\beta$ used in the ACF parameterization

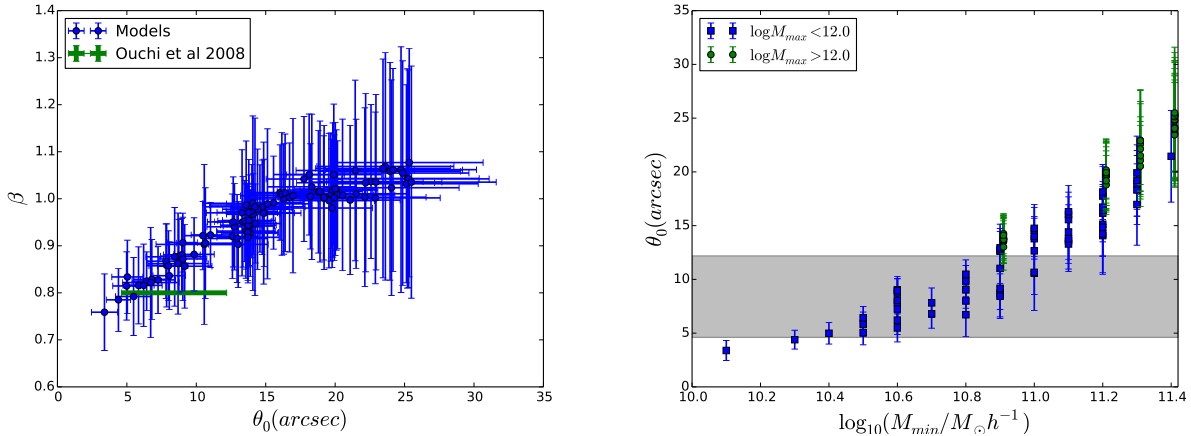


FIG. 2.— Left panel. Values for the free parameters θ_0 and β in the fitting formula (Eq. 1) for the angular correlation function. Blue dots correspond to simulations and the green line to observations by Ouchi et al. (2008, 2010). The error bars in the theoretical data correspond to the quadratic average of the fitting uncertainties for each mock survey. Right panel. Values of θ_0 and $\log_{10}(M_{\min}/M_{\odot}h^{-1})$. Blue and green point represents the models where $\log_{10}(M_{\max}/M_{\odot}h^{-1}) < 12.0$ and $\log_{10}(M_{\min}/M_{\odot}h^{-1}) > 12.0$ respectively. The horizontal gray band represents the observational constraints on θ_0 established by Ouchi et al. (2008, 2010). Green points has been displaced by 0.01 dex of its original value to avoid overlapping between blue and green points.

on the best models previously selected. Blue circles with error bars represent the results from the mocks and the horizontal line the observational results of Ouchi et al. (2008) over a field with average number density where the β parameter was fixed in the fit. From this Figure we observe that there are models with too large values of θ_0 and/or β that can be ruled out as inconsistent.

Figure 3 presents the consistent models in the planes M_{\min} - M_{\max} and M_{\min} - f_{occ} . A model is consistent with observations if there is a $1\text{-}\sigma$ overlap between both the correlation length θ_0 and the exponent β . We find that we end up with 40 models consistent with the ACF constraints. This is a reduction of a factor of ~ 2 with respect to the number of models consistent with the number density distribution.

4. DISCUSSION

Out of the initial set of 9000 models we end up with 40 that are consistent with the observational constraints. To facilitate the discussion of these models we define a new quantity, the halo mass range $\Delta M \equiv \log_{10} M_{\max} - \log_{10} M_{\min}$, which together with the occupation fraction, f_{occ} , and the minimum mass M_{\min} allows us to classify all the successful models into three families:

- (1) Low occupation fraction $f_{\text{occ}} \leq 0.2$ and narrow mass range $\Delta M \leq 1.0$ dex: 16 models.
- (2) High occupation fraction $f_{\text{occ}} > 0.2$ and narrow mass range $\Delta M \leq 1.0$: 17 models
- (3) Low occupation fraction $f_{\text{occ}} \leq 0.2$ and wide mass range $\Delta M > 1.0$: 7 models

A complete parameter list for the three families is presented in Tables 1, 2 and 3.

The models in the third family are barely consistent with the constraints from the ACF. These models they have another particular feature: their minimum mass is exactly $M_{\min} = 10^{10.9}h^{-1}M_{\odot}$ and $M_{\max} > 10^{12.0}h^{-1}M_{\odot}$. They have values of β close to 1.0 (instead of the

observational value of 0.8) and θ_0 on the range of $13'$ (the mean observational value is close to $8'$) but are considered consistent because their large error bars barely overlap with the observational expectation (see figure 2). Because these models are barely compatible with observations and are a minority of the consistent models, we exclude them from the discussion.

4.1. Two relevant features in the preferred models

There are two interesting features in the remaining two first families. First, the occupation fraction can take any value from 0.1 to 1.0. Second, the halos hosting LAEs all have a narrow mass range $\Delta M \leq 1.0$, half of them $\Delta M \leq 0.5$.

The existence of models with high occupation fractions $f_{\text{occ}} > 0.2$ is unexpected from previous clustering analysis that do not take account the effect of cosmic variance in the same way we propose in this paper. Nevertheless, values of an occupation fraction $f_{\text{occ}} = 0.1$ and the mass range predicted by all of our models is consistent with already published observational results (Gawiser et al. 2007b; Ouchi et al. 2010). This suggests that explicitly taking cosmic variance of the observed fields into account (the only point not included in the analysis already cited) hinders the possibility of constraining the LAEs occupation fraction.

Concerning the narrow mass range $\Delta M < 1.0$ the question arises of how can we give a physical interpretation for the existence of such a narrow mass range. We can start with a reasonable assumption. Namely, that star formation rate increases with halo's mass. Under this assumption the cut at the low mass end, M_{\min} , can be interpreted in terms of the minimal star formation rate required to produce a Ly α luminosity above observational detection thresholds. A cut at higher halo masses M_{\max} requires a different justification. There are two complementary physical scenarios that could provide it.

One scenario can be presented in terms of a decreasing escape fraction of Ly α radiation in massive systems. Detailed galaxy formation models support the idea that

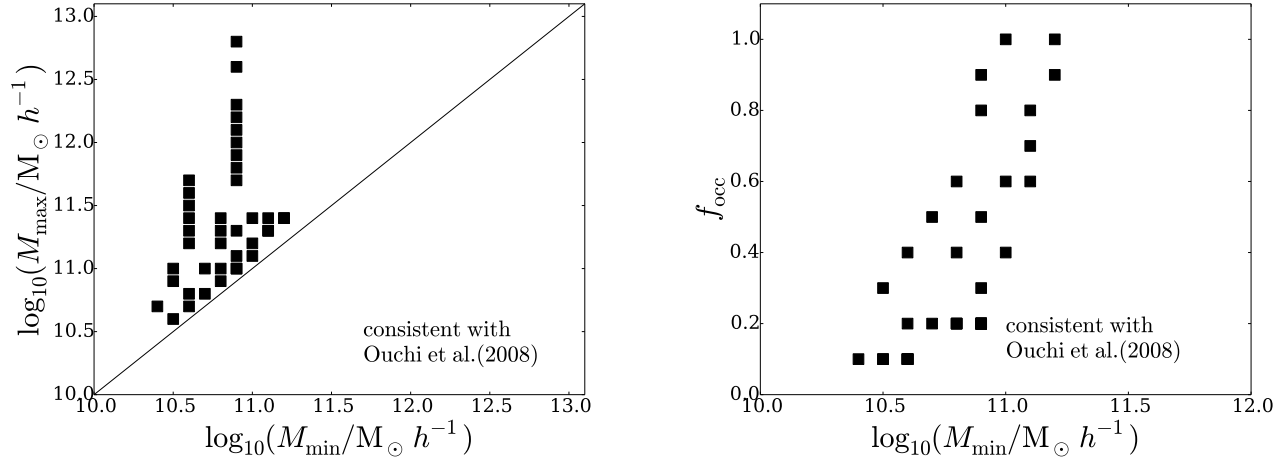


FIG. 3.— M_{\min} - M_{\max} (left) and M_{\min} - f_{occ} (right) planes showing the models fulfilling both constraints on the maximal number of consistent mocks and the angular correlation function from Ouchi et al. (2008, 2010).

massive galaxies with higher metallicities have larger dust contents and a less concentrated ISM than lower mass systems. Due to the resonant nature of the Ly α line the probability of absorption of Ly α photons increases in massive systems, producing high absorption of the Ly α line but not of UV continuum or other non-resonant lines (Laursen et al. 2009; Forero-Romero et al. 2011). In a second scenario larger systems have more extended gaseous envelopes which due to resonance effects of the Ly α line, induces a low surface brightness and a broader line, making these systems less observable in narrow band filter surveys (Laursen et al. 2009; Zheng et al. 2010).

4.2. Comparison to previous clustering estimates

Observational evidence based on the ACF inferred from photometric measurements in the Extended Chandra Deep Field South has shown that the median dark matter masses of halos hosting LAEs is $\log_{10} M_{\text{med}} = 10.9^{+0.5}_{-0.9} M_{\odot}$, with a corresponding occupation fraction of 1 – 10% (Gawiser et al. 2007b). Ouchi et al. (2010) presents analysis of LAE observations in the redshift interval $3.1 < z < 7.0$ and at $z = 3.1$ They quote an average mass for the host dark matter halos of $M_h = 2.9^{+24.0}_{-2.9} \times 10^{10} h^{-1} M_{\odot}$ with a corresponding duty cycle of 0.008 ± 0.03 .

Our results are in a general good agreement with those estimates for the host mass and the upper limit of the occupation fraction. This is not completely unexpected given that we have also required consistency with ACF measurements. These expectations are mostly matched by the first family of models, also summarized in Table 1. These models, which favor only the low ($\sim 10\%$) occupations fractions, are also consistent in that regard with the observational expectations by Gawiser et al. (2007b).

The novelty in our results is that we have a detailed estimate for host halo mass range together with the occupation fraction. This allows us to show that the halo mass range can, in some cases, be narrow $\Delta M < 0.3$ dex, something that cannot be inferred from ACF analysis alone. Furthermore, In contrast to Gawiser et al. (2007b) and Ouchi et al. (2010) we find that an ACF analysis is not enough to rule out models with a high

$\log_{10} M_{\min}$	$\log_{10} M_{\max}$	f_{occ}	ΔM
10.4	10.7	0.1	0.3
10.5	10.9	0.1	0.4
10.5	11.0	0.1	0.5
10.6	10.8	0.2	0.2
10.6	11.2	0.1	0.6
10.6	11.3	0.1	0.7
10.6	11.4	0.1	0.8
10.6	11.5	0.1	0.9
10.6	11.6	0.1	1
10.7	11.0	0.2	0.3
10.8	11.2	0.2	0.4
10.8	11.3	0.2	0.5
10.8	11.4	0.2	0.6
10.9	11.9	0.2	1.0
10.9	11.7	0.2	0.8
10.9	11.8	0.2	0.9

TABLE 1
LIST OF PARAMETERS FOR THE FIRST FAMILY OF MODELS. NARROW MASS RANGE $\Delta M \leq 1.0$ DEX AND LOW OCCUPATION FRACTION $f_{\text{occ}} \leq 0.3$.

occupation fraction $f_{\text{occ}} > 0.2$.

4.3. In the context of abundance matching models

The abundance matching methods are based on observational results for Lyman Break Galaxies (LBGs) (Behroozi et al. 2013b,a). In the case of Behroozi et al. (2013b) the minimum halo mass considered to be relevant in their analysis is $10^{11.4} h^{-1} M_{\odot}$. They report stellar mass around $(1.0 \pm 0.3) \times 10^{9.0} h^{-1} M_{\odot}$, while their star formation rate is in the range $0.6 \pm 0.2 M_{\odot} \text{yr}^{-1}$, which nevertheless are close to the lower bound of values inferred for LAEs at high redshift (Gawiser et al. 2007a; Nilsson et al. 2009; Pentericci et al. 2009).

In our results, all the preferred models have a halo mass range lower than the minimum of $M_{\min} < 10^{11.4} h^{-1} M_{\odot}$ considered in abundance matching at $z = 3$. Our results confirm the expectations that most of LAEs are to be found in less massive halos LBG hosts. A detailed analysis of the spectral and photometric properties of LAEs coupled to the kind of analysis performed in this paper can be a guide in the study of the properties of low

$\log_{10} M_{\min}$	$\log_{10} M_{\max}$	f_{occ}	ΔM
10.5	10.6	0.3	0.1
10.6	10.7	0.4	0.1
10.7	10.8	0.5	0.1
10.8	10.9	0.6	0.1
10.8	11.0	0.4	0.2
10.9	11.0	0.8	0.1
10.9	11.0	0.9	0.1
10.9	11.1	0.5	0.2
10.9	11.3	0.3	0.4
11.0	11.1	1.0	0.1
11.0	11.2	0.6	0.2
11.0	11.4	0.4	0.4
11.1	11.3	0.7	0.2
11.1	11.3	0.8	0.2
11.1	11.4	0.6	0.3
11.2	11.4	0.9	0.2
11.2	11.4	1.0	0.2

TABLE 2

LIST OF PARAMETERS FOR THE SECOND FAMILY OF MODELS. NARROW MASS RANGE $\Delta M \leq 1.0$ DEX AND HIGH OCCUPATION FRACTION $f_{\text{occ}} > 0.2$.

$\log_{10} M_{\min}$	$\log_{10} M_{\max}$	f_{occ}	ΔM
10.6	11.7	0.1	1.1
10.9	12.0	0.2	1.1
10.9	12.1	0.2	1.2
10.9	12.2	0.2	1.3
10.9	12.3	0.2	1.4
10.9	12.6	0.2	1.7
10.9	12.8	0.2	1.9

TABLE 3

LIST OF PARAMETERS FOR THE THIRD FAMILY OF MODELS. WIDE MASS RANGE $\Delta M > 1.0$ DEX AND LOW OCCUPATION FRACTION $f_{\text{occ}} \leq 0.2$. THESE MODELS ARE BARELY CONSISTENT WITH THE CONSTRAINTS FROM THE ANGULAR CORRELATION FUNCTION AND HAVE BEEN EXCLUDED FROM THE MAIN DISCUSSION.

mass dark matter halos at $z = 3.1$, extending the capabilities of abundance matching methods.

4.4. Caveats of our method

There are three minor caveats for the work presented here that are important to keep in mind.

The first caveat is the precise values for the mass intervals. These values are quoted from halos defined using a FOF halo finder. Different halo finders and definitions for the detection density threshold can yield different masses up to a factor ~ 2 (More et al. 2011). For instance a Friends-of-Friends algorithm with linking length $l = 0.20$ times the average inter-particle distance finds halos on average 1.4 times less massive than halos defined with an spherical overdensity algorithm halos (Klypin et al. 2011). Therefore, the mass values for M_{\min} and M_{\max} should not be considered exact within less than ~ 0.2 dex. However, this represents a systematic change in the mass values and the mass interval ΔM quoted here should remain unchanged.

The second caveat is the exact value that was used to turn filter values into cosmological volumes. Strictly speaking the volume probed is a function of the Ly α luminosity. Brighter LAEs probe a larger volume than fainter ones (Gronwall et al. 2007). A proper way to ac-

count for this effect would require to assign luminosities to each mock LAE and process them in a similar way as mock observations. However, we can estimate that under the considerations presented in (Gronwall et al. 2007), one could expect the survey depth to change at most by a factor of ~ 2 . Considering the dependence of the number density on halo mass (right panel of Figure 1), such change in volume could be compensated by different values for the favored model parameters M_{\min} and M_{\max} by about the same factor of ~ 2 or 0.3 dex.

A third caveat is the comparison of results for the number density statistics and the ACF from surveys with different EW cuts. In this sense we highlight again that the two filters and selection criteria used by (Ouchi et al. 2008) and (Yamada et al. 2012) are only slightly different. The median number density in the general fields by (Yamada et al. 2012) is 0.20 arcmin^{-2} , while in (Ouchi et al. 2008) is $0.099 \pm 0.005 \text{ arcmin}^{-2}$. This gives us confidence that the two different samples can give a self-consistent picture.

4.5. On the reproducibility of our results

All the software, raw and processed data to produce the results and plots in this paper are publicly available in a github repository². Most of the code to produce the plots can be found as an Ipython notebook (Pérez & Granger 2007) in the same repository.

5. CONCLUSIONS

In this paper we quantify the impact of cosmic variance in constraining the mass and occupation fraction of dark matter halos hosting Lyman α Emitters at redshift $z = 3.1$ in a Λ CDM cosmology. To this end we build a large number of mock catalogs matching observational geometries. The mocks are constructed from a N-body simulation by assigning a single LAE to a DM halo.

We then proceed with a thorough exploration of the space of free parameters M_{\min} , M_{\max} and f_{occ} describing the model. We look for consistency with two observational constraints: the surface number density distribution and the angular correlation function. Out of the initial 9000 combinations of parameters in the model we find 40 combination of parameters consistent with observations.

We find two interesting features in these models. First, the occupation fraction remains unconstrained; second, many models have a narrow mass range models with $\Delta M \leq 0.5$.

The wide range of occupation fraction values is unexpected from previous studies (i.e. Gawiser et al. 2007a; Ouchi et al. 2010) that unanimously favor low values on the range of 10^{-2} . Nevertheless, all the halo mass ranges deduced in previous analysis are a subset of the models we find in this paper, all of the best models support the notion that the most massive halos at $z = 3.1$ do not host detectable LAEs.

From the point of view of Dark Energy surveys such as the Hobby-Eberly Telescope Dark Energy Experiment (HETDEX) the narrow mass range for LAEs can be seen as an advantage because its simplicity in their modeling aiming at recovering cosmological parameters. We foresee that the observations with new instruments (such as

² <https://github.com/forero/CosmicVarianceLAEs>

the Multi Unit Spectroscopic Explorer (MUSE), Hyper SuprimeCam and HETDEX) covering larger fields and a wider range of luminosities will be key in reducing cosmic variance and imposing tighter constraints on the properties of dark matter halos hosting LAEs.

Finally, additional modeling for Ly α radiation transfer is needed to put tighter constraints on occupation fraction in high redshift halos, also paying attention to other physical phenomena, such as the stochasticity (Forero-Romero & Dijkstra 2013) in the star formation process, which might play a role in introducing detection biases in high redshift LAEs.

ACKNOWLEDGMENTS

REFERENCES

- Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013a, *ApJ*, 762, L31
- . 2013b, *ApJ*, 770, 57
- Bond, N. A., Gawiser, E., Gronwall, C., Ciardullo, R., Altmann, M., & Schawinski, K. 2009, *ApJ*, 705, 639
- Colberg, J. M., White, S. D. M., Yoshida, N., MacFarland, T. J., Jenkins, A., Frenk, C. S., Pearce, F. R., Evrard, A. E., Couchman, H. M. P., Efstathiou, G., Peacock, J. A., Thomas, P. A., & Virgo Consortium. 2000, *MNRAS*, 319, 209
- Dayal, P., Ferrara, A., Saro, A., Salvaterra, R., Borgani, S., & Tornatore, L. 2009, *MNRAS*, 400, 2000
- Dijkstra, M., & Kramer, R. 2012, *MNRAS*, 424, 1672
- Dijkstra, M., Mesinger, A., & Wyithe, J. S. B. 2011, *MNRAS*, 414, 2139
- Forero-Romero, J. E., & Dijkstra, M. 2013, *MNRAS*, 428, 2163
- Forero-Romero, J. E., Yepes, G., Gottlöber, S., Knollmann, S. R., Cuesta, A. J., & Prada, F. 2011, *MNRAS*, 415, 3666
- Forero-Romero, J. E., Yepes, G., Gottlöber, S., & Prada, F. 2012, *MNRAS*, 419, 952
- Garel, T., Blaizot, J., Guiderdoni, B., Schaerer, D., Verhamme, A., & Hayes, M. 2012, *MNRAS*, 422, 310
- Gawiser, E., Francke, H., Lai, K., Schawinski, K., Gronwall, C., Ciardullo, R., Quadri, R., Orsi, A., Barrientos, L. F., Blanc, G. A., Fazio, G., & Feldmeier, J. J. 2007a, *ApJ*, 671, 278
- Gawiser, E., Francke, H., Lai, K., Schawinski, K., Gronwall, C., Ciardullo, R., Quadri, R., Orsi, A., Barrientos, L. F., Blanc, G. A., Fazio, G., Feldmeier, J. J., Huang, J.-s., Infante, L., Lira, P., & Padilla, N. 2007b, *ApJ*, 671, 278
- Gronwall, C., Ciardullo, R., Hickey, T., Gawiser, E., Feldmeier, J. J., van Dokkum, P. G., Urry, C. M., Herrera, D., Lehmer, B. D., Infante, L., Orsi, A., Marchesini, D., Blanc, G. A., Francke, H., Lira, P., & Treister, E. 2007, *ApJ*, 667, 79
- Guaita, L., Francke, H., Gawiser, E., Bauer, F. E., Hayes, M., Östlin, G., & Padilla, N. 2013, *A&A*, 551, A93
- Hayashino, T., Matsuda, Y., Tamura, H., Yamauchi, R., Yamada, T., Ajiki, M., Fujita, S. S., Murayama, T., Nagao, T., Ohta, K., Okamura, S., Ouchi, M., Shimasaku, K., Shioya, Y., & Taniguchi, Y. 2004, *AJ*, 128, 2073
- Jarosik, N., Bennett, C. L., Dunkley, J., Gold, B., Greason, M. R., Halpern, M., Hill, R. S., Hinshaw, G., Kogut, A., Komatsu, E., Larson, D., & Limon, M. 2011, *ApJS*, 192, 14
- Jose, C., Srianand, R., & Subramanian, K. 2013, *ArXiv e-prints*
- Klypin, A. A., Trujillo-Gomez, S., & Primack, J. 2011, *ApJ*, 740, 102
- Koehler, R. S., Schuecker, P., & Gebhardt, K. 2007, *A&A*, 462, 7
- Komatsu, E., Dunkley, J., Nolte, M. R., Bennett, C. L., Gold, B., Hinshaw, G., Jarosik, N., Larson, D., Limon, M., Page, L., Spergel, D. N., & Halpern, M. 2009, *ApJS*, 180, 330
- J.E.F-R was supported by the FAPA grant by Vicerrectoría de Investigaciones at Universidad de los Andes.
- J.E.F-R thanks the hospitality of Changbom Park and the Korea Institute for Advanced Study where the first full draft of this paper was completed. The authors also thank Peter Laursen, Paulina Lira, Alvaro Orsi and Mark Dijkstra for helpful comments on the physical interpretation and presentation of our results.
- The MultiDark Database used in this paper and the web application providing online access to it were constructed as part of the activities of the German Astrophysical Virtual Observatory as result of a collaboration between the Leibniz-Institute for Astrophysics Potsdam (AIP) and the Spanish MultiDark Consolider Project CSD2009-00064. The Bolshoi and MultiDark simulations were run on the NASA's Pleiades supercomputer at the NASA Ames Research Center.
- Kudritzki, R.-P., Méndez, R. H., Feldmeier, J. J., Ciardullo, R., Jacoby, G. H., Freeman, K. C., Arnaboldi, M., Capaccioli, M., Gerhard, O., & Ford, H. C. 2000, *ApJ*, 536, 19
- Landy, S. D., & Szalay, A. S. 1993, *ApJ*, 412, 64
- Laursen, P., Duval, F., & Östlin, G. 2013, *ApJ*, 766, 124
- Laursen, P., Razoumov, A. O., & Sommer-Larsen, J. 2009, *ApJ*, 696, 853
- Laursen, P., & Sommer-Larsen, J. 2007, *ApJ*, 657, L69
- Matsuda, Y., Yamada, T., Hayashino, T., Tamura, H., Yamauchi, R., Murayama, T., Nagao, T., Ohta, K., Okamura, S., Ouchi, M., Shimasaku, K., Shioya, Y., & Taniguchi, Y. 2005, *ApJ*, 634, L125
- More, S., Kravtsov, A. V., Dalal, N., & Gottlöber, S. 2011, *ApJS*, 195, 4
- Neufeld, D. A. 1991, *ApJ*, 370, L85
- Nilsson, K. K., Möller, P., Möller, O., Fynbo, J. P. U., Michałowski, M. J., Watson, D., Ledoux, C., Rosati, P., Pedersen, K., & Grove, L. F. 2007, *A&A*, 471, 71
- Nilsson, K. K., Tapken, C., Möller, P., Freudling, W., Fynbo, J. P. U., Meisenheimer, K., Laursen, P., & Östlin, G. 2009, *A&A*, 498, 13
- Orsi, A., Lacey, C. G., & Baugh, C. M. 2012, *MNRAS*, 425, 87
- Ouchi, M., Shimasaku, K., Akiyama, M., Simpson, C., Saito, T., Ueda, Y., Furusawa, H., Sekiguchi, K., Yamada, T., Kodama, T., Kashikawa, N., Okamura, S., Iye, M., Takata, T., Yoshida, M., & Yoshida, M. 2008, *ApJS*, 176, 301
- Ouchi, M., Shimasaku, K., Furusawa, H., Saito, T., Yoshida, M., Akiyama, M., Ono, Y., Yamada, T., Ota, K., Kashikawa, N., Iye, M., Kodama, T., Okamura, S., Simpson, C., & Yoshida, M. 2010, *ApJ*, 723, 869
- Peebles, P. J. E. 1980, *The large-scale structure of the universe*
- Pentericci, L., Grazian, A., Fontana, A., Castellano, M., Giallongo, E., Salimbeni, S., & Santini, P. 2009, *A&A*, 494, 553
- Pérez, F., & Granger, B. E. 2007, *Comput. Sci. Eng.*, 9, 21
- Riebe, K., Partl, A. M., Enke, H., Forero-Romero, J., Gottlber, S., Klypin, A., Lemson, G., Prada, F., Primack, J. R., Steinmetz, M., & Turchaninov, V. 2013, *Astronomische Nachrichten*, 334, 691
- Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N., Gao, L., Navarro, J., Thacker, R., Croton, D., Helly, J., Peacock, J. A., Cole, S., Thomas, P., Couchman, H., Evrard, A., Colberg, J., & Pearce, F. 2005, *Nature*, 435, 629
- Verhamme, A., Schaerer, D., & Maselli, A. 2006, *A&A*, 460, 397
- Yajima, H., Choi, J.-H., & Nagamine, K. 2012, *MNRAS*, 427, 2889
- Yamada, T., Nakamura, Y., Matsuda, Y., Hayashino, T., Yamauchi, R., Morimoto, N., Kousai, K., & Umemura, M. 2012, *AJ*, 143, 79
- Zheng, Z., Cen, R., Trac, H., & Miralda-Escudé, J. 2010, *ApJ*, 716, 574