

# Using cosmic variance to constrain the dark matter halo mass of Lyman-alpha emitting galaxies at $z=3.1$

Jaime E. Forero-Romero<sup>1</sup> and Julian Mejía<sup>2</sup>

<sup>1</sup> *Departamento de Física, Universidad de los Andes, Cra. 1 No. 18A-10, Edificio Ip, Bogotá, Colombia*

<sup>2</sup> –

27 May 2013

## ABSTRACT

We use cosmological N-body simulations to find the characteristic mass of dark matter halos hosting Lyman-Alpha Emitting (LAE) galaxies at a redshift of  $z = 3.1$ . The method is based on matching the statistics for the number density between mock and observed fields. The mock catalogs are constructed using a simple model where a dark matter halo can only host one LAE with a probability  $f_{\text{occ}}$  if its mass is found within a certain range mass range delimited by two threshold values,  $M_{\text{min}}$  and  $M_{\text{max}}$ . We find that the most of the models that are consistent the observed cosmic variance statistics are those with halo masses in the range  $10.5 < \log_{10} M_{\text{min}}/h\text{M}_{\text{sun}} < 11.5$  and  $\log_{10} M_{\text{max}}/h^{-1}\text{M}_{\odot} < 13.5$  with an occupation fraction that scales as  $f_{\text{occ}} =$ . We explore three additional constraints to narrow down these range: the number of mocks consistent with observations, observational constraints on the occupation fraction and the angular correlation function. The first two conditions narrow down the space parameter to  $M_{\text{min}} =$  and  $M_{\text{max}}, f_{\text{occ}}$ . The angular correlation function does not add a significant constraint due to the cosmic variance in the small angular fields where this statistics has been computed so far. We make available the mock data for the best models in a public repository. Implications for galaxy formation models?

**Key words:** galaxies: kinematics and dynamics, Local Group, methods:numerical

## 1 INTRODUCTION

Lyman- $\alpha$  emitting galaxies (LAEs) have become in the last decade a central topic in studies of structure formation in the Universe. They are helpful in a diverse range of fields. 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 and markers of the galaxy formation process through cosmic history (Forero-Romero et al. 2012).

At the same time, theoretical and observational developments have contributed to the emergence of a paradigm to describe structure formation in a cosmological context. In this context it is considered that dominant matter content of the Universe is to be found in dark matter, whereby each galaxy is hosted by larger dark matter structure known as a halo.

Most models of galaxy formation find that the mass of the halo can be used to predict properties of the galaxy such as its stellar mass and star formation rate (Behroozi et al. 2012). Processes that regulate the star formation cycle are also thought to be strongly dependent on its mass. Furthermore, the spatial clustering of galaxies on large scales is entirely dictated by the halo distribution. For the reasons

mentioned above, finding the typical dark matter halo mass hosting LAEs represents a significant step forward to understand the nature of this population in the context of Lambda Cold Dark Matter ( $\Lambda$ CDM) paradigm.

Some theoretical approaches to this problem have been based on a forward modeling. Starting from the DM halo population, the corresponding intrinsic star formation properties are inferred and statistics such as the luminosity function, the correlation function and the equivalent width distributions. Such modelling has been implemented from analytic considerations, semi-analytic models and full N-body hydrodynamical simulations (Dayal et al. 2009; Forero-Romero et al. 2011; Yajima et al. 2012; Forero-Romero et al. 2012).

Added to the uncertainties in the astrophysical processes describing star formation in galactic populations, a highly debated step in this approach is the calculation of the fraction of Lyman- $\alpha$  photons that escape the galaxy to the observer. Given the resonance nature of the line, the radiative transfer of Lyman- $\alpha$  is sensitive to the density, temperature, topology and kinematics of the neutral Hydrogen in the interstellar medium (ISM) (Neufeld 1991; Forero-Romero et al. 2011; Laursen et al. 2013).

This complexity makes the use of monte-carlo simulations for the radiative transfer a required tool to obtain physically sound results, although the degeneracy in the physical parameters involved in the problem makes it difficult to achieve a robust consensus on what is the theoretical expected value for the Lyman- $\alpha$  escape fraction in high redshift.

Throughout this paper we assume a  $\Lambda$ 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 to constrain the typical mass of a dark matter halos hosting LAEs at  $z = 3.1$  is based on the comparison of observational results on the surface number density and the predictions of a simple model that uses the outputs from cosmological N-body simulations.

In the next subsections we describe in detail the four key elements of this workflow. First, we present the observations we take as a benchmark. Second, the N-body simulation and the halo catalogs we use. Third, the simplified model that allows us to translated halo catalogs into mock LAE observations. Fourth, the statistics we use to compare observational results against our theoretical predictions.

### 2.1 Observational Constraints

The observational benchmark 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 is centered at  $4977 \text{ \AA}$  with a  $77 \text{ \AA}$  width, corresponding to the redshift range  $z = 3.062 - 3.125$  and  $41 \text{ Mpc } h^{-1}$  comoving scale for the detection of the Lyman- $\alpha$  line centered at  $z = 3.09$ . The authors report a total of 2161 LAEs with an observed equivalent width larger than  $190 \text{ \AA}$  over a total survey area of  $2.42 \text{ deg}^2$ , this corresponds to an average surface number density of  $0.20 \pm 0.01 \text{ arcmin}^{-2}$ .

The survey covered four independent fields. The first is the SSA22 field of  $1.38 \text{ deg}^2$  with 1394 detected LAEs, this field has been known to harbor a region with a large density excess of galaxies. The second observed region is composed by the fields Subaru/XMM-Newton Deep Survey (SXDS)-North, -Center and -South, with a total of  $0.58 \text{ deg}^2$  and 386 LAEs. The third and fourth fields are the Subaru Deep Field (SDF) with  $0.22 \text{ deg}^2$  and 196 LAEs, and the field around the Great Observatory Optical Deep Survey North (GOODS-N) with  $0.24 \text{ deg}^2$  and 185 LAEs.

There is abundant observational work done on LAEs at redshift  $z = 3.1$  (Kudritzki et al. 2000; Matsuda et al. 2005; Gawiser et al. 2007; 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 instrumental 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 due only to astrophysical reasons and not different observational conditions or criteria to construct the catalogs.

### 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. It includes dark matter distribution is sampled using  $2048^3$  particles, which translates into a particle mass of  $m_p = 1.35 \times 10^8 h^{-1} \text{ M}_\odot$ . The cosmological parameters are consistent with a WMAP5 and WMAP7 data with a density  $\Omega_m = 0.27$ , cosmological  $\Omega_\Lambda = 0.73$ , 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 interparticle distance. The minimum halo mass in the models we construct in this paper correspond to groups of  $\sim 75$  particles. The catalogs were obtained from the publicly available Multidark database<sup>1</sup> (Riebe et al. 2011). We focus our work on halos more massive than  $1 \times 10^{10} h^{-1} \text{ M}_\odot$  that are resolved with at least 70 particles.

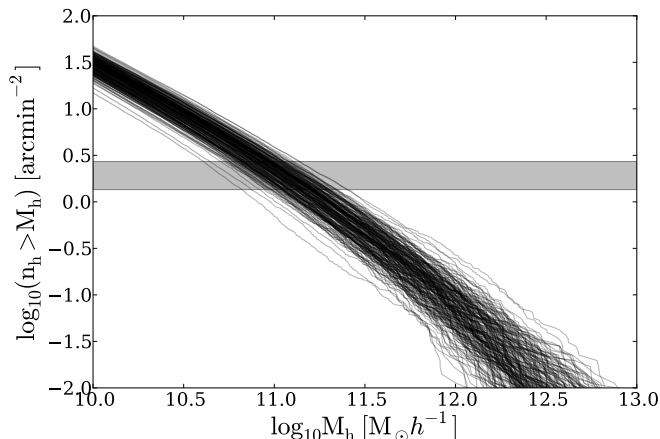
### 2.3 A Model to Populate Halos with LAEs

In our model a dark matter halo can only host one or zero LAE. There are three parameters in the model that decide whether a halo can host a galaxy or not: the lower and upper bounds for the mass range  $M_{\min} < M_h < M_{\max}$  where LAEs reside and the fraction  $f_{\text{occ}}$  of such halos that effectively host a LAE. We do not assign a luminosity to each LAE. We are primarily interested in constraining the halo mass range hosting detectable LAEs under the conditions defined by Yamada et al. (2012). In what follows will describe by the letter  $\mathcal{M}$  a model defined by an specific choice of the three scalar parameters  $M_{\min}$ ,  $M_{\max}$  y  $f_{\text{occ}}$ .

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  $46 \times 35 \times 41 h^{-1} \text{ Mpc}^3$  where the last dimension goes in the redshift direction. This corresponds to a total area of  $880 \text{ arcmin}^2$  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 11 mock fields out of the 210 available volumes. In total 15 mock surveys are constructed for each model  $\mathcal{M}$ . The grouping is done in two different ways. In the first way, called *match*, we follow the clustering of the observed fields. From the 11 mock fields, 7 are constructed from contiguous fields in the simulation to mimic the SSA22 region, 3 are also

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



**Figure 1.** Surface density of dark matter halos as a function of a minimum halo mass to count the total number of elements in a volume. Each line represents one of the 210 volumes of dimensions  $46 \times 35 \times 41 h^{-1} \text{Mpc}^3$  in the Bolshoi simulation. The horizontal grey band represents the range of surface densities observed for LAEs at  $z = 3.1$  as reported by (Yamada et al. 2012).

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. Our main goal with this selection is to test the impact on the final statistics of having 7 clustered fields. The second way to group the mock fields is called *random*, whereby all the 11 fields are selected in such a way as to avoid that any two volumes are contiguous.

## 2.4 Sampling and Selecting Models

We make a thorough exploration of the parameter space for the models  $\mathcal{M}$ .  $\log_{10} M_{\min}$  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  $\log_{10} M_{\min}$  only with a displacement of 0.1 dex in the whole range. The occupation fraction  $f_{\text{occ}}$  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  $30 \times 30 \times 10 = 9000$ .

For each mock survey generated in a given model  $\mathcal{M}$  we compute the surface density in the 12 mock fields. We perform a Kolmogorov-Smirnov (KS) to compare this mock data against the 12 observational values. From this test we obtain 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.

In this paper we consider that a model  $\mathcal{M}$  that has at least one (1) mock survey (out of 15) consistent with the observed distribution of LAE number densities has viable parameters that deserve to be considered for further analysis.

## 3 RESULTS

### 3.1 Dark Matter Halo Number Density

In Figure ?? we present the results for the integrated dark matter halo surface density as a function of halos mass. Each line corresponds to one of the 210 sub-volumes in the Bolshoi simulation. The shadowed area indicates the surface density values for LAEs allowed by the observations.

This result allows us to better understand the expected trends for the LAEs' preferred mass and the occupation fraction. From this Figure we can read which models do not have a chance to reproduce the observations. Regions in the plot where the halo surface density values are below the observational constraint correspond to high masses halo masses. For a LAE model  $\mathcal{M}$  with a minimum mass  $M_{\min} > 3 \times 10^{11} h^{-1} \text{M}_{\odot}$  located in that mass range, the surface density is too low compared with observations.

Conversely, there are regions in the plot where the halo surface density is always higher than the observational constraints correspond to models  $\mathcal{M}$  with a minimum mass below  $M_{\min} < 3 \times 10^{10} h^{-1} \text{M}_{\odot}$ . Models with this minimum mass have a chance for successfully reproducing observations if the occupation fraction  $f_{\text{occ}} < 1$  is tuned as to lower the halo number density down to the observed value.

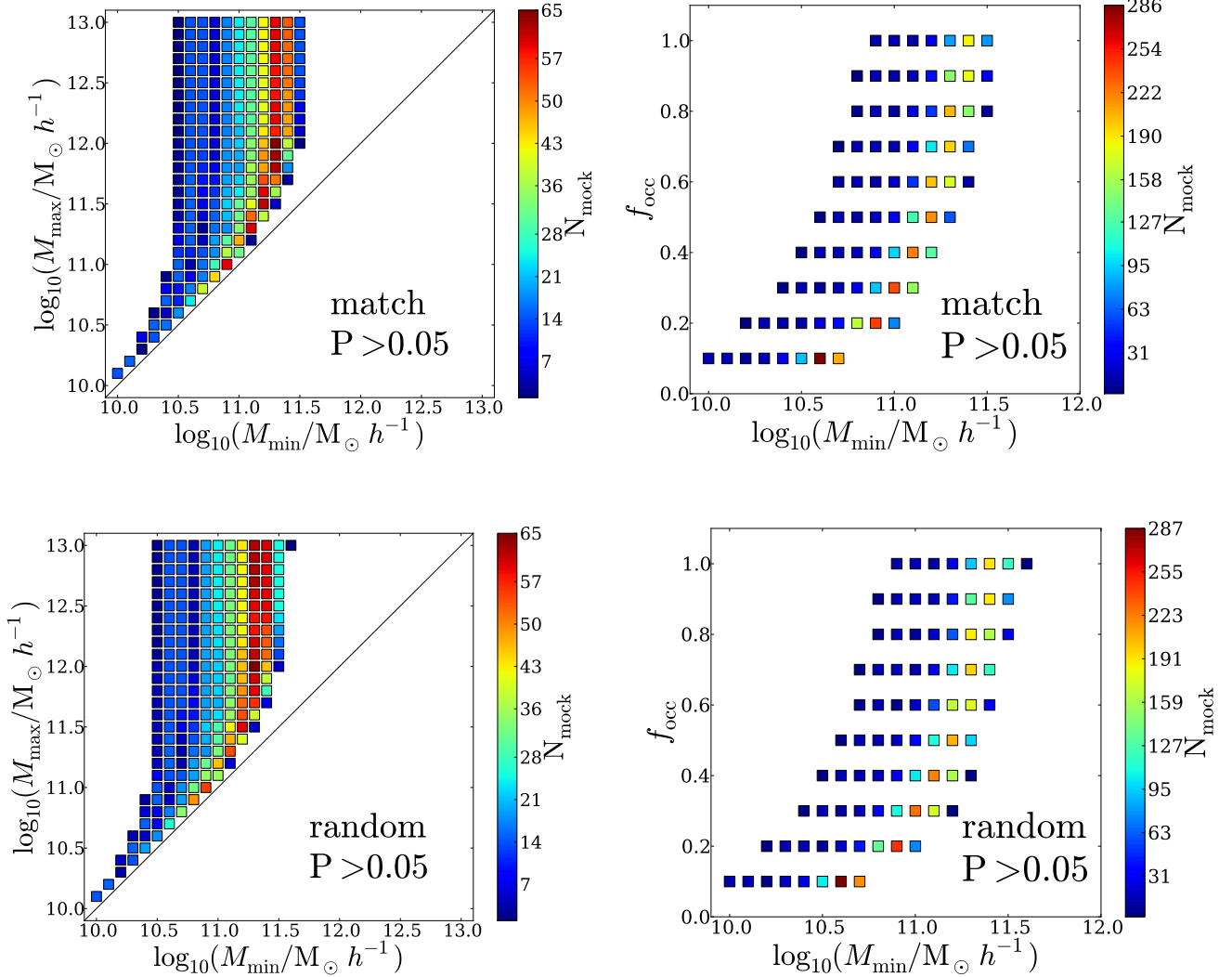
### 3.2 Kolmogorov-Smirnov Tests

Figure presents the regions in the parameter space  $M_{\min}$ - $M_{\max}$  where the KS test yields values of  $P > 0.05$ . We consider that for those models it is not possible to rule out the null hypothesis, namely that the number density in simulated data and the observations come from the same parent distribution. Each panel corresponds to the three different ways of grouping the mock fields. In the case of the methods *Match* and *Random* the color code indicates the fraction of these 15 mock surveys with  $P > 0.1$ . The third panel shows the result for the method *Full*, in this case the color code corresponds to the value maximum value of  $100 \times P$  for a model with those mass ranges.

These results clearly distinguish three mass regimes. In the first regime, at high mass values, we find that LAE models with minimum mass of  $M_{\min} > 10^{11.5} h^{-1} \text{M}_{\odot}$  are not compatible with observations. There is a second regime for masses below  $M_{\min} = 3 \times 10^{10}$  any values for  $M_{\min}$  and  $M_{\max}$  can be made compatible with observations, provided that  $f_{\text{occ}}$  is fine tuned to do it. In an intermediate mass regime, for minimum mass values  $3 \times 10^{10} h^{-1} \text{M}_{\odot} < M_{\min} < 3 \times 10^{11} h^{-1} \text{M}_{\odot}$  only a limited range of models with  $M_{\max}$  with occupation fraction  $f_{\text{occ}} \sim 1$  is able to reproduce observations.

In these three different mass regimes the occupation fraction monotonically decreases as a function of the minimum halo mass  $M_{\min}$ . In Figure XXX we show this trend in three panels following the same correspondence as Figure XXX. From these results we interpret that the different mass regimes that were identified correspond to best fit models with  $f_{\text{occ}} \sim 1$ ,  $0 < f_{\text{occ}} < 1$  and  $f_{\text{occ}} \sim 0$ , respectively.

The two *match* and *random* methods present the highest number of matching mock surveys in the medium mass regime. However, it is important to keep in mind that not all the mock surveys for a successful model  $\mathcal{M}$  present a high



**Figure 2.**  $M_{\min}$ - $M_{\max}$  plane for all models with  $P > 0.05$  in the three different of grouping the mock fields. In the case of the **match** and **random** methods the color code corresponds to the number of mock surveys that are found to be compatible with observations. For the **full** method the colour code corresponds to the results of 100 times the maximum  $P$  value resulting from the KS test. .

value  $P > 0.1$ , only a modest fraction seems to be consistent with observations. This shows that the cosmic variance is still present on the physical scales probed by observations. This will be considered in more detail in the discussion section.

To illustrate this point, in Figure ?? we present the results for two mock fields for the **Match** method for a model with the same parameters, but two extreme values for the KS test.

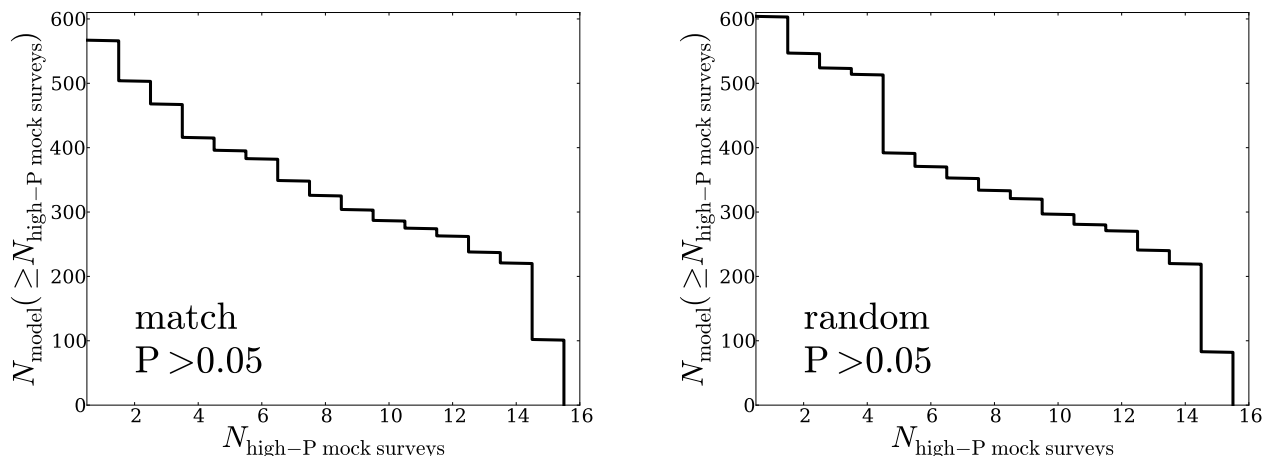
Conversely there are different models where the KS test yield values of  $P \sim 1$ . To illustrate the kind of success represented by these models, we have selected these best ones in the case of the method **MatchObs**. Figure XXX shows in the main panel the spatial distribution of the mock surveys, the smaller panel shows the corresponding surface density distribution and the observational constraint.

In what follows we will focus our discussion on the mocks constructed with the **Match** and **Random** methods.

#### 4 ADDITIONAL CONSTRAINTS

The raw results of our experiments yield a preferred range for halo masses hosting LAES bounded by a minimum mass  $10^{10} h^{-1} M_{\odot} < M_{\min} < 10^{11.5} h^{-1} M_{\odot}$  and without any limitation on the maximum mass. In this regime the average occupation fraction varies with an approximate dependence of the kind  $f_{\text{occ}} \approx 0.1 (M_{\min}/10^{10})^{X.X}$ . These results discard models that were nevertheless disfavored from the very beginning based on the results of the mass functions (see Figure 1.).

In this section we consider three different ways to impose tighter constraints on this mass range by making use of the results we have derived so far together with additional statistical and astrophysical constraints. In the first constraint we select the models where all the mock surveys present KS-test values consistent with observations. The second constraints uses recent observational results on the average occupation fraction for LAEs at high- $z$ . The third ex-



**Figure 3.** Number of models with a minimum number of mock survey realizations that are consistent with observations. .

exploits the information in the Angular Correlation Function (ACF).

#### 4.0.1 Models with the highest success rates

For each model there are 15 different mock survey realizations. In the previous section we presented the models that had at least one (1) mock survey realization with  $P > 0.05$ .

In Figure 3 we show the number of models that have at least  $n$  realizations with  $P > 0.05$  for the **match** and **random** methods. This shows that there are around 550 to 600 different models that have at least one mock survey realization consistent with observations. At the other extreme, there are 80 to 100 models with all the 15 realizations with  $P > 0.05$ . Here we focus on the latter models. The best models represent  $\sim 15\%$  of the number of initially considered good models.

In Figure XX we present the loci of these models in the parameter space  $M_{\min} - M_{\max}$  and  $M_{\min} - f_{\text{occ}}$ . In Figure XX we show the spatial distribution for two mock surveys corresponding to one of such models.

#### 4.0.2 Observational constraints on the occupation fraction

We now impose a different restriction using the observational results by Hayes et al. (2010). These authors constrained the value of  $f_{\text{occ}}$  at  $z = 2.2$  to be  $f_{\text{occ}} = 0.10$ . This estimation was based on blind surveys of the H $\alpha$  and Lyman alpha line with the European Southern Observatory (ESO) Very Large Telescope (VLT). Using corrections by extinction to obtain an estimate for the intrinsic H $\alpha$  luminosity, and using values for the theoretical expectation of the ratio Lyman $\alpha$ /H $\alpha$  they derive an bulk escape fraction for the Lyman $\alpha$  radiation of  $f_{\text{esc}} = (5.3 \pm 3.8)\%$  or  $f_{\text{esc}} = (10.7 \pm 2.8)\%$  if a different dust correction is used. The authors show that the luminosity function for LAEs at  $z = 2.2$  is consistent with the escape fraction being constant for every galaxy regardless of its luminosity. From this results they derive that almost 90% of the star forming galaxies emit insufficient Lyman-alpha to be detected, effectively setting the occupation fraction to be  $f_{\text{occ}} = 0.10$ .

For the cosmological parameters used in this paper the age of the universe between  $z = 3.1$  and  $z = 2.2$  has changed by  $\sim 1$  Gyr. We assume that the physical conditions that determine the escape fraction  $f_{\text{esc}}$  and the occupation fraction  $f_{\text{occ}}$  remain constant over that time scale. This assumption allows us to further pick models that have an occupation fraction of  $f_{\text{occ}} = 0.10$ . Under this selection only 18 models can be selected. Considering an occupation fraction  $f_{\text{occ}} = 0.20$  another 39 models can be considered. This constraint help us to select  $\sim 10\%$  of the original models that were considered as consistent with observations in the previous section.

Figure XX shows the preferred models in the planes  $M_{\min} - M_{\max}$  and  $M_{\min} - f_{\text{occ}}$  for the **match** and **random** methods. The list for the model parameters is found in the appendix in Table XX.

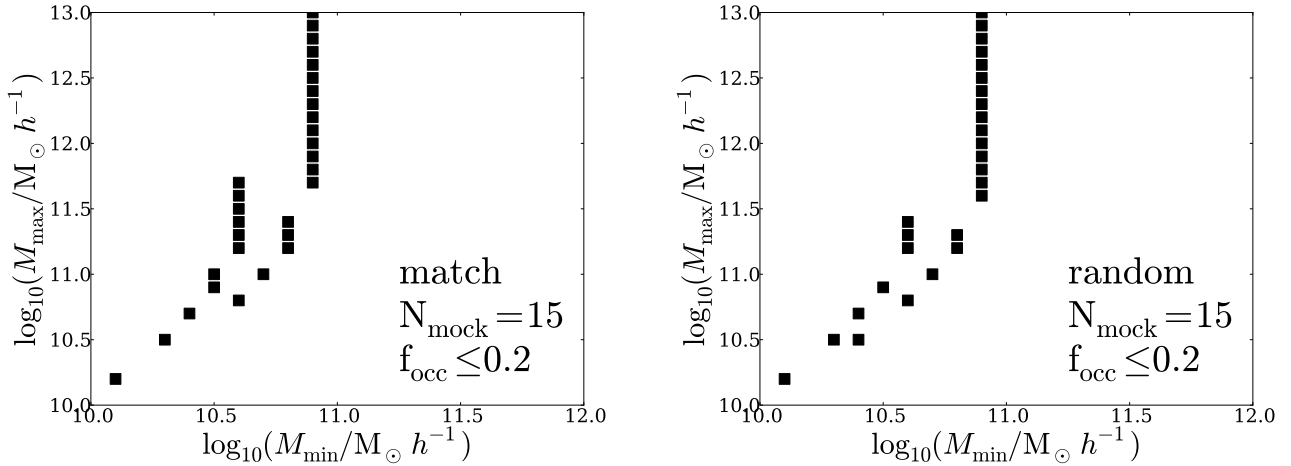
In terms of the constraints done in the previous subsection, we find that the models in this region of parameter space have an average number of XX mock catalogs consistent with observations.

#### 4.0.3 Constraints from the Angular Correlation Function

We calculate the mean angular correlation function (ACF) only for the models that showed that all of the 15 mock surveys are consistent at the  $P > 0.05$  level with observations. The ACF is computed only the densest subfield in all the 15 mock surveys corresponding to the SSA22 region. These results are compared against the observations reported by Hayashino et al in 2004 over the same region, which were also performed on the densest field.

Figure ?? (match) and Figure ?? (random) present such comparison. The error bars in these figures represent the standard deviation of the ACF over all the sub-fields. In general, we observe that the standard deviation of the computed ACF in the subfields increases with  $M_{\min}$  following the same trend as in Figure 1, as a direct consequence of cosmic variance.

The comparison between the simulated and observed ACFs is also done using a  $R^2$  statistic which includes the information on measurement uncertainties



**Figure 4.** Favored models when the constraints on the maximal number of consistent mocks and the observational constrain on the occupation fraction are included.

$$R^2 = \sum_{\theta_i} \frac{(\xi_{\text{obs}}(\theta_i) - \xi_{\text{sim}}(\theta_i))^2}{\sigma_{\text{obs}}^2(\theta_i) + \sigma_{\text{sim}}^2(\theta_i)}, \quad (1)$$

where the sum is done over all the angle values  $\theta_i$  where the ACF has been computed. In Figure XX we plot the integrated distributions for this  $R^2$  statistics

Given that the ACF reported by Hayashino et al en 2004 is taken over the densest field observed in the SSA22 region by Yamada et al in 2012 it is expected that the predicted ACF in the SSA22 region should reproduce this observation. In the left panel of figure ?? we can see the predicted ACFs and their corresponding standard deviation over the seven fields that mock the SSA22 region. It can be seen that the model with  $M_{\text{min}} = 10.6$  seems to better reproduce the Hayashino's ACF and that the corresponding field is in fact an overdense field in the SSA22 region covered by Yamada et al.

From these tests we conclude that the ACF on small fields does not provide additional constraints to further select models for halos hosting LAEs. The reason is that cosmic variance is large and the statistical uncertainties on the ACF render almost any model compatible with the observational constraints.

We also present the ACF in the case of the full SSA22 region which has been homogeneously observed by (Yamada et al. 2012). To this date the observational ACF has not been reported in the literature, therefore our calculations can be considered as predictions.

In Figure X we present the results for the models. The full list of these correlation functions can be found in the data repository for this paper in `github`.

We also present the results for the angular correlation function in terms of the correlation length obtained by fitting the following function

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma} \quad (2)$$

The results are shown in Figure in a  $r_0 - \gamma$  plane where

the average and standard deviation for each mock are shown in comparison with the result derived from observations.

## 5 DISCUSSION

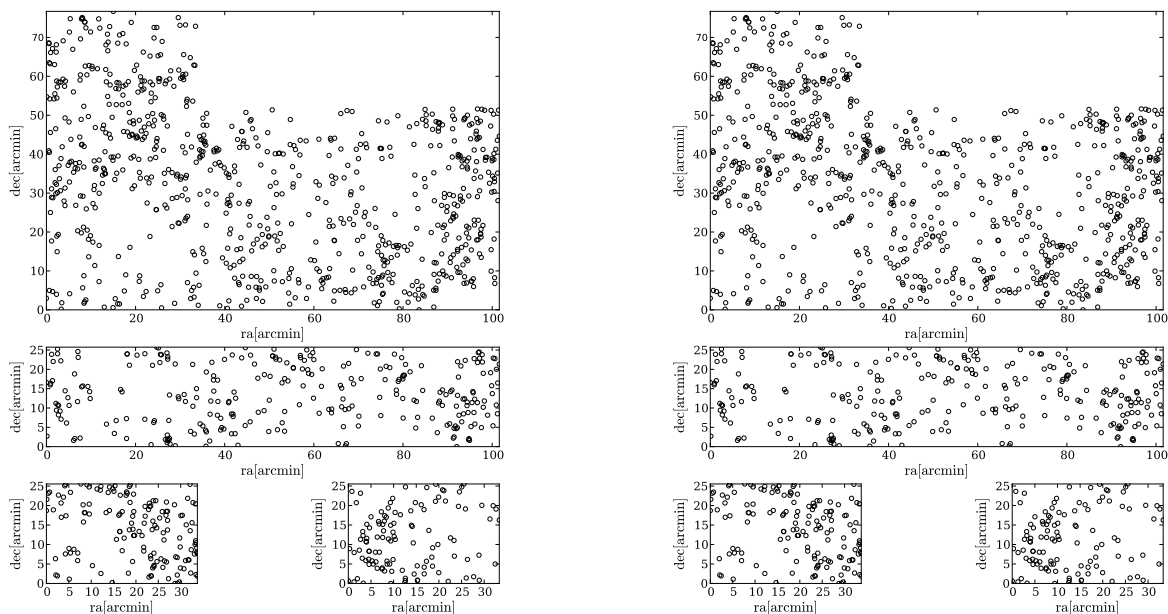
When we include the tightest constraints on the mock catalogs, we find that there are 30 set of parameters of our model, out of the original 90000 initial models, that are consistent with the observational constraints at redshift 3.1: the distribution of the number density, the inferred values for the average occupation fraction. The consistency with the angular correlation, in terms of the  $\chi^2$  statistics did not help to discard any additional models with a significant degree of confidence.

These 30 models can be classified into two families of the same size. The first, where the range  $M_{\text{min}} - M_{\text{max}}$  is narrow, typically of less than  $< 1.0$  dex. While in the second family the extent  $> 1.0$  dex. In the first case the minimum halo mass is found to be in a wide range  $10^{10} h^{-1} M_{\odot} < M_{\text{min}} < 10^{11.5} h^{-1} M_{\odot}$  while in the second case, only models with  $M_{\text{min}} \sim 10^{10.9} h^{-1} M_{\odot}$  are compatible with the observational constraints. In what follows we discuss the implications of the existence of these two families of models.

### 5.1 Implications for galaxy formation models

In the case of a narrow of masses to host LAEs the upper masses are bound to be  $M_{\text{max}} < 10^{11.5} h^{-1} M_{\odot}$  as it is shown in Figure 4. For halos more massive than this bound it is naturally expected that the galaxies can be observed as Lyman Break Galaxies (LBGs). This would imply that not all the bright LBGs can be detected as a LAE.

We have the opposite situation in the second family of models. If we have a wide range in halo masses, where the upper end of the halo masses can be considered as observed LAEs, one can expect that bright LBGs will have a correspondence as observed LAEs. The most interesting aspect is that there is a clear cut in the minimal mass that can be attained by observed LAEs  $M_{\text{min}} > 10^{11} h^{-1} M_{\odot}$ . This puts



**Figure 5.** Spatial distribution for two mocks corresponding to the model  $M_{\min} =$ ,  $M_{\max} =$  and  $f_{\text{occ}} =$ . All the 15 different mock surveys for this model in the match configuration are consistent with observations at the  $P > 0.05$  level. The full data for all the mocks can be found in the github repository for this paper. .

a tight constraint on the relationship between the minimum star formation rate required to be observed as a LAE and this minimal halo mass.

- ... Intrinsic emission and escape fraction.
- ... Star formation rate efficiency at this redshift.
- ... Mass dependence of the escape fraction.

## 5.2 Implications for large LAEs surveys

- ... The bias for the preferred halo mass.
- ... The scale at which cosmic variance drops.
- ... This can be observationally tested with HETDEX.

## 5.3 On the reproducibility of our results

- ... All the software to produce the results in this paper is publicly available.
- ... The raw catalogs can be obtained from the MultiDark database but can also be obtained in the repository of this paper on github.

## 6 CONCLUSIONS

In this paper we constrain the preferred mass for dark matter halos hosting Lyman Alpha Emitters at a redshift  $z = 3.1$ . We use a method that matches the cosmic variance in the surface density number of LAEs between mock and real observations. The mock catalogs are based on a simplified model with three basic parameters: the halo mass range where LAEs can be found,  $M_{\min} < M_h < M_{\max}$ , and the fraction of the halos in this range that are actually occupied,  $f_{\text{occ}}$ . After a thorough exploration of the parameter space we are able to constrain the mass range of dark matter halos

hosting LAEs to be in the range  $< M_h <$  and a corresponding occupation fraction that scales as  $f_{\text{occ}} = M_{\min}$ .

We use three additional constraints to reduce the allowed range of models. The first imposes a tighter criterion to consider a model successful, namely that all the mock surveys for a given model must be consistent with observations. This restriction narrows down the allowed range of models to be.

The second constraint is based on the observational results that high redshift LAEs have a bulk Lyman alpha escape fraction of  $XX$  which can be also interpreted as an average occupation fraction of  $XX$ .

Including additional observational constraints on the occupation fraction allows us to reduce the range of allowed halo masses to be in a narrower range of  $< M_h <$ . Including the information from the angular correlation function (ACF) does not allow us to impose further constraints. This is due to the scatter in the ACF due to the cosmic variance on the field observed by  $XXX$ .

We simulation allows us to extract 210 sub-boxes each of which has a comparable volume to the individual fields of view observed by Yamada et al. (2012). The comparison of the observed number density distribution against the results from our model is based on three different ways of constructing mock surveys. The first reproduces the spatial correlation between the 12 observational fields (*match*), the second breaks this spatial correlation while keeping the number of fields (*random*) and the third one simply includes all the 210 sub-boxes (*full*). We find that the methods *match* and *random* allow a larger set of models than the *full* method. We do not find a significant difference between the two first methods.

10.1	10.2	0.1
10.3	10.5	0.1
10.4	10.7	0.1
10.5	10.9	0.1
10.6	11.2	0.1
10.6	11.3	0.1
10.6	11.4	0.1
10.1	10.2	0.1
10.3	10.5	0.1
10.4	10.7	0.1
10.5	10.9	0.1
10.5	11.0	0.1
10.6	11.2	0.1
10.6	11.3	0.1
10.6	11.4	0.1
10.6	11.5	0.1
10.6	11.6	0.1
10.6	11.7	0.1
10.4	10.5	0.2
10.6	10.8	0.2
10.7	11.0	0.2
10.8	11.2	0.2
10.8	11.3	0.2
10.9	11.6	0.2
10.9	11.7	0.2
10.9	11.8	0.2
10.9	11.9	0.2
10.9	12.0	0.2
10.9	12.1	0.2
10.9	12.2	0.2
10.9	12.3	0.2
10.9	12.4	0.2
10.9	12.5	0.2
10.9	12.6	0.2
10.9	12.7	0.2
10.9	12.8	0.2
10.9	12.9	0.2
10.9	13.0	0.2
10.6	10.8	0.2
10.7	11.0	0.2
10.8	11.2	0.2
10.8	11.3	0.2
10.8	11.4	0.2
10.9	11.7	0.2
10.9	11.8	0.2
10.9	11.9	0.2
10.9	12.0	0.2
10.9	12.1	0.2
10.9	12.2	0.2
10.9	12.3	0.2
10.9	12.4	0.2
10.9	12.5	0.2
10.9	12.6	0.2
10.9	12.7	0.2
10.9	12.8	0.2
10.9	12.9	0.2
10.9	13.0	0.2

- Tornatore L., 2009, MNRAS, 400, 2000
- Dijkstra M., Mesinger A., Wyithe J. S. B., 2011, MNRAS, 414, 2139
- 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
- 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., 2007, ApJ, 671, 278
- Hayes M., Östlin G., Schaerer D., Mas-Hesse J. M., Leitherer C., Atek H., Kunth D., Verhamme A., de Barros S., Melinder J., 2010, Nature, 464, 562
- 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
- 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
- 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
- Laursen P., Duval F., Östlin G., 2013, ApJ, 766, 124
- 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, ApJL, 634, L125
- Neufeld D. A., 1991, ApJL, 370, L85
- Nilsson K. K., Møller 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
- 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
- Riebe K., Partl A. M., Enke H., Forero-Romero J., Gottlöber S., Klypin A., Lemson G., Prada F., Primack J. R., Steinmetz M., Turchaninov V., 2011, ArXiv e-prints
- 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

## ACKNOWLEDGMENTS

## REFERENCES

- Behroozi P. S., Wechsler R. H., Conroy C., 2012, ArXiv e-prints
- Dayal P., Ferrara A., Saro A., Salvaterra R., Borgani S.,