

IMPACT OF COSMIC VARIANCE ON THE GALAXY–HALO CONNECTION FOR $\text{Ly}\alpha$ EMITTERSJULIÁN E. MEJÍA-RESTREPO^{1,2} AND JAIME E. FORERO-ROMERO³¹ Departamento de Astronomía, Universidad de Chile, Camino el Observatorio 1515, Santiago, Chile; jemejia@das.uchile.cl² FAcCom-Instituto de Física-FCEN, Universidad de Antioquia, Calle 70 No. 52-21, Medellín, Colombia³ Departamento de Física, Universidad de los Andes, Cra. 1 No. 18A-10, Edificio Ip, Bogotá, Colombia; je.forero@uniandes.edu.co

Received 2016 April 26; revised 2016 June 21; accepted 2016 June 22; published 2016 August 23

ABSTRACT

In this paper we study the impact of cosmic variance and observational uncertainties in constraining the mass and occupation fraction, f_{occ} , of dark matter (DM) halos hosting $\text{Ly}\alpha$ -emitting galaxies (LAEs) at high redshift. To this end, we construct mock catalogs from an N -body simulation to match the typical size of observed fields at $z = 3.1$ ($\sim 1 \text{ deg}^2$). In our model a DM halo with mass in the range $M_{\text{min}} < M_{\text{h}} < M_{\text{max}}$ can only host one detectable LAE at most. We proceed to explore the parameter space determined by M_{min} , M_{max} , and f_{occ} with a Markov Chain Monte Carlo algorithm using the angular correlation function and the LAEs' number density as observational constraints. We find that the preferred minimum and maximum masses in our model span a wide range $10^{10.0} h^{-1} M_{\odot} \leq M_{\text{min}} \leq 10^{11.1} h^{-1} M_{\odot}$, $10^{11.0} h^{-1} M_{\odot} \leq M_{\text{max}} \leq 10^{13.0} h^{-1} M_{\odot}$, followed by a wide range in the occupation fraction $0.02 \leq f_{\text{occ}} \leq 0.30$. As a consequence, the median mass, M_{50} , of all the consistent models has a large uncertainty $M_{50} = 3.16_{-2.37}^{+9.34} \times 10^{10} h^{-1} M_{\odot}$. However, we find that the same individual models have a relatively tight 1σ scatter around the median mass $\Delta M_{1\sigma} = 0.55_{-0.31}^{+0.11}$ dex. We are also able to show that f_{occ} is uniquely determined by M_{min} , regardless of M_{max} . We argue that upcoming large surveys covering at least 25 deg^2 should be able to put tighter constraints on M_{min} and f_{occ} through the LAE number density distribution width constructed over several fields of $\sim 1 \text{ deg}^2$.

Key words: dark matter – galaxies: halos – galaxies: high-redshift – galaxies: statistics – methods: numerical

1. INTRODUCTION

$\text{Ly}\alpha$ -emitting galaxies (LAEs) are central to a wide range of subjects in extragalactic astronomy. LAEs can be used as probes of reionization (for a recent review see Dijkstra 2014, and references therein), tracers of large-scale structure (Koeehler et al. 2007), signposts for low-metallicity stellar populations (for a recent review see Hayes 2015, and references therein), markers of the galaxy formation process at high redshift (Partridge & Peebles 1967; Rhoads et al. 2000; Blanc et al. 2011), and tracers of active star formation.

In most of those cases, capitalizing on the observations requires understanding how LAEs are formed within an explicit cosmological context. Under the current structure formation paradigm, the dominant matter content of the universe is dark matter (DM). Each galaxy is thought to be hosted by a larger DM structure known as a halo (Peebles 1980; Springel et al. 2005). Understanding the cosmological context of LAEs thus implies studying the galaxy–halo connection. Galaxy formation models suggest that the physical processes that regulate the star formation cycle are dependent on halo mass (e.g., Behroozi et al. 2013). Therefore, the mass becomes the most important element in the halo–galaxy connection.

The goal becomes finding the typical DM halo mass of halos hosting LAEs. In the case of LAEs there are different ways to find this mass range. One approach is theoretical, using general astrophysical principles to find the relationship between halo mass, intrinsic $\text{Ly}\alpha$ luminosities, and observed $\text{Ly}\alpha$ luminosities. This approach is usually implemented through semianalytic 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).

The downside of these calculations is the uncertainty in the estimation of the escape fraction of $\text{Ly}\alpha$ photons. Given the

resonant nature of the $\text{Ly}\alpha$ line, the escape fraction is sensitive to the dust contents, density, temperature, topology, and kinematics of the neutral hydrogen in the interstellar medium (ISM). The process of finding a consensus on the expected value for the $\text{Ly}\alpha$ escape fraction in high-redshift galaxies is still a matter of open debate (Neufeld 1991; Verhamme et al. 2006; Dijkstra & Kramer 2012; Forero-Romero et al. 2012; Orsi et al. 2012; Laursen et al. 2013; Yajima et al. 2014).

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 (ACF) 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. 2007; Nilsson et al. 2007; Ouchi et al. 2010; Bielby et al. 2016). In these studies the observations are done on fields of $\sim 1 \text{ deg}^2$ and the conclusions derived on the halo host mass do not elaborate deeply on the uncertainty resulting from the cosmic variance on these fields. For instance, Guaita et al. (2010) attempted to analytically estimate the uncertainty from cosmic variance following Somerville et al. (2004) and Peebles (1980). Their prescription assumed that the correlation function can be represented by a power law $\xi(r) = (r_0/r)^\gamma$. Their estimation of cosmic variance strongly depends on r_0 and γ . However, γ is only poorly constrained with current observations, which in turn indicates that their cosmic variance estimation could not be precise.

In this paper we investigate the impact of cosmic variance in constraining the mass and occupation fraction of halos hosting LAEs at $z = 3$. We build mock surveys from a cosmological N -body simulation to compare them against the observations of

Bielby et al. (2016) using the ACF. We use a simple model to populate a halo in the simulation with an LAE assuming a minimum mass, M_{\min} , and maximum mass, M_{\max} , for the DM halos hosting LAEs. We do not assume an underlying relation between the Ly α luminosity and the DM halo mass. This approach bypasses all the physical uncertainties associated with star formation and radiative transfer. We then use the Markov Chain Monte Carlo technique to obtain the likelihood of the parameters given the observational constraints. This approach allows us to estimate cosmic variance directly from simulations without making any assumption on the correlation function behavior.

Throughout this paper we assume a Λ CDM cosmology with the following values for the cosmological parameters: $\Omega_m = 0.30711$, $\Omega_\Lambda = 0.69289$, and $h = 0.70$, corresponding to the matter density, vacuum density, and the Hubble constant in units of $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$. The values are consistent with Planck Collaboration et al. (2014) results.

2. METHODOLOGY

The base of our method is the comparison between observations and mock catalogs. This approach allows us to take explicitly into account cosmic variance. The comparison has four key elements. First, we take the observations as a benchmark. Second, we use the N -body simulation and the halo catalogs to build the mocks. Third, we use the parameters describing our model to assign an LAE to a halo. Fourth, we adopt the statistical method to compare observations and simulations. We describe in detail these four elements in the following subsections.

2.1. Observational Constraints

Bielby et al. (2016) used narrowband imaging to detect 643 LAE candidates at $z \sim 3$ with Ly α rest-frame equivalent widths $\gtrsim 65 \text{ \AA}$ and an Ly α flux limit of $2 \times 10^{17} \text{ erg cm}^{-2} \text{ s}^{-1}$ ($L \sim 7 \times 10^{42} \text{ erg s}^{-1}$). Using follow-up spectroscopy, they found a 22% contamination fraction f_c . Their observations cover five (out of nine) independent and adjacent fields of the VLT LBG Redshift Survey. The total observed area corresponds to 1.07 deg^2 , which translates to $\sim 80^2 h^{-1} \text{ Mpc}^2$ in a comoving scale. Bielby et al. (2016) used the NB497 narrowband filter, whose 77 \AA FWHM and 154 \AA full width tenth maximum correspond to a total observational comoving depth of 44 and $82 h^{-1} \text{ Mpc}$, respectively.

2.2. Simulation and Halo Catalog

We use results from the BolshoiP simulation (Klypin et al. 2011, 2016) performed in a cubic comoving volume of $250 h^{-1} \text{ Mpc}$ on a side. The DM distribution was sampled using 2048^3 particles.

We chose this simulation based on three requirements. First, the simulation has cosmological parameters consistent with the most recent constraints from cosmic microwave background experiments and provides public access to the data. Second, the halo mass function from the simulation is robust down to $10^{10} h^{-1} M_\odot$, which is the lower halo mass bound suggested by previous observational studies (Hayashino et al. 2004; Gawiser et al. 2007; Ouchi et al. 2010; Bielby et al. 2016). Third, the total volume of the simulation is at least 10 times larger than the equivalent volume of a single LAE observational field. These conditions are met as we describe below.

The cosmological parameters are consistent with *Planck* results (Planck Collaboration et al. 2014) with a matter density $\Omega_m = 0.307$, cosmological constant $\Omega_\Lambda = 0.693$, dimensionless Hubble constant $h = 0.678$, slope of the power spectrum $n = 0.96$, and normalization of the power spectrum $\sigma_8 = 0.823$. This translates into a particle mass of $m_p = 1.5 \times 10^8 h^{-1} M_\odot$. Data are available to the public through an online interface⁴ (Riebe et al. 2013).

We use halo catalogs constructed with a bound density maxima (BDM) algorithm. For each halo in the box we extract its comoving position and mass. We focus our work on halos more massive than $6.0 \times 10^9 h^{-1} M_\odot$ resolved with at least 40 particles to guarantee statistical significance and a well-behaved halo mass function. We do not take into account subhalos.

We split the simulation volume at $z \sim 3$ into 27 smaller mock volumes mimicking the area and depth reported in Bielby et al. (2016) and described in Section 2.1.

2.3. A Simple LAE Model

We build the simplest possible model to assign an LAE to each DM halo without trying to compute an LAE luminosity.

We first assume that a DM halo can host one detectable LAE at most. This assumption 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). Then we say that a halo will host an LAE with probability f_{occ} if and only if the halo mass is in the range $M_{\min} < M_h < M_{\max}$. We also use the variable $\Delta \log M$ to represent the mass range width, $\Delta \log M \equiv \log M_{\max} - \log M_{\min}$.

An astrophysical interpretation of f_{occ} convolves at least four phenomena: the actual presence of a star-forming galaxy in a halo, a duty cycle in the star formation rate, the escape fraction of Ly α radiation, and its detectability as an LAE with Ly α equivalent widths of $\gtrsim 65 \text{ \AA}$. We do not try to disentangle these effects and do not assume an underlying relation between the α luminosity and the DM halo mass. We instead opt for a purely arithmetic interpretation by setting f_{occ} as the ratio between the observed number of LAEs and the number of halos within the considered mass range, that is, $f_{\text{occ}} \equiv N_{\text{LAEs}}/N_{\text{halos}}$.

For each mock catalog we also randomly remove a fraction $f_c = 0.22$ of the mock LAEs and replace them with randomly distributed points to mimic the effect of interloper contamination in the Bielby et al. (2016) observations. On top of that we apply rejection sampling to our LAE selection along the radial direction, taking the transmission function of the NB479 filter used in the Bielby et al. (2016) observations.

Figure 1 shows the cumulative halo number density for all 27 subvolumes in the simulation, with a normalization by the median number density among fields, $\langle n \rangle$. Each line represents a different model \mathcal{M} with fixed $f_{\text{occ}} = 1$ and $\Delta \log M = 1.0$ and varying M_{\min} . This figure shows that the halo number density varies across subvolumes, as an expression of cosmic variance. As a consequence, the f_{occ} also varies across the mock fields by the same factor.

In what follows we will describe by the letter \mathcal{M} a model defined by a particular choice of the two parameters M_{\min} and M_{\max} . For each model \mathcal{M} we define \tilde{f}_{occ} as the median occupation fraction within the mock fields.

⁴ <http://www.multidark.org/MultiDark/>

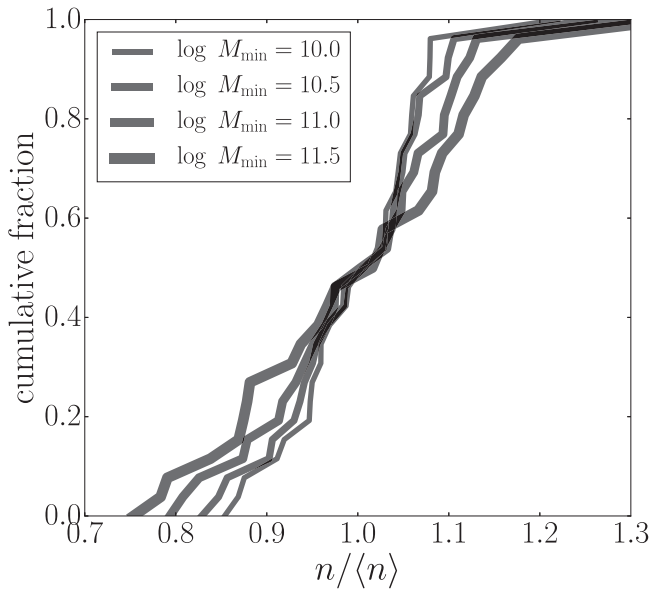


Figure 1. Cumulative halo number density distribution function over 27 mock fields. Each line corresponds to a different model with increasing values of M_{\min} . Different models produce different number density distributions. The width of the distribution increases with M_{\min} .

2.4. Model Selection

We explore the parameter space of the models \mathcal{M} by means of the affine-invariant Markov Chain Monte Carlo technique using the EMCEE python package (Foreman-Mackey et al. 2013, and references therein).

The MCMC exploration is done using a total of 24 seeds and 400 iterations (9600 models) to sample the posterior probability distribution function, $P(\mathcal{M}|\text{observations})$, based on the ACF. We put a flat prior on $\log M_{\min}$ and $\log M_{\max}$ between 9.8 and 13.4, corresponding to the halo mass range of the simulation at $z = 3$. We restrict the selection to models that give a minimal number density $N_{\text{halos}} > N_{\text{LAE}}/3$. This means that it is possible to have $N_{\text{halos}} < N_{\text{LAE}}$ and hence $f_{\text{occ}} \equiv N_{\text{LAE}}/N_{\text{halos}} > 1$. We include the 1/3 factor to account for the uncertainty in the number density of LAEs due to cosmic variance, expecting it to be on the same order as the DM halo cosmic variance shown in Figure 1. Our likelihood is taken proportional to $\exp(-\chi_{\mathcal{M}}^2/2)$, where

$$\chi_{\mathcal{M}}^2 = \sum_{\theta} \left[\frac{(\text{ACF}_{\mathcal{M}}(\theta) - \text{ACF}_{\text{obs}}(\theta))^2}{\sigma_{\mathcal{M}}^2(\theta) + \sigma_{\text{obs}}^2(\theta)} \right]. \quad (1)$$

Here $\text{ACF}_{\mathcal{M}}$ and ACF_{obs} are the ACF of the explored model \mathcal{M} and the observational ACF reported by Bielby et al. (2016), respectively. $\sigma_{\mathcal{M}}$ is the associated 1σ scatter of the $\text{ACF}_{\mathcal{M}}$ as a product of cosmic variance, and σ_{obs} is the observational error associated to ACF_{obs} .

We do not include the covariance matrix due to the practical impossibility of estimating it in a robust way. The small size of the fields and galaxies makes the application of jackknife and bootstrap techniques unfeasible (Norberg et al. 2009). An estimate from N -body simulation faces the same problem, with the additional limitation that we want to solve in the first place, namely, not knowing the range of halo masses that produces the observed clustering signal. Tests that we performed in the simulation show that, due to the small number of points to estimate the clustering, the covariance

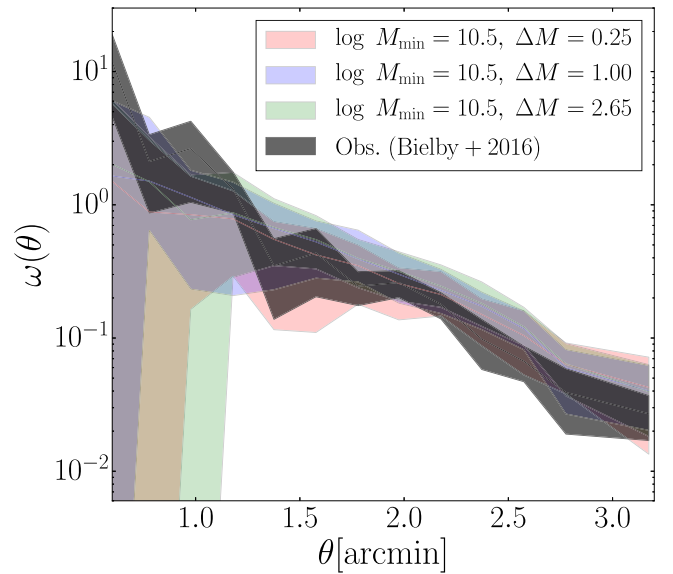


Figure 2. Angular correlation functions for $\log M_{\min} [M_{\odot} h^{-1}] = 10.5$ and different values of $\Delta \log M$. The shaded region in the models represents the 1σ variation due to cosmic variance. Radically different models in $\Delta \log M$ are consistent with observations once cosmic variance is modeled in detail.

matrix is not stable, even for similar models. Using the full error covariance matrix, when not noise dominated, versus only diagonal elements usually has a small effect on clustering analyses (Zehavi et al. 2011).

We compute the $\text{ACF}_{\mathcal{M}}$ using the Landy & Szalay estimator (Landy & Szalay 1993). After this, we correct the computed ACF from the contamination fraction of interlopers f_{c} by multiplying the ACF by a factor of $1/(1 - f_{\text{c}})^2$ following the procedure described in Bielby et al. (2016).

Figure 2 shows the observational ACF by Bielby et al. (2016) compared to the ACF in three different models with a wide range in $\Delta \log M$. This already shows that radically different models can be compatible with observations once cosmic variance uncertainties are modeled in detail.

3. RESULTS AND DISCUSSION

3.1. Constraints on Model Parameters

Figure 3 shows the one- and two-dimensional projections of the posterior probability distributions of the parameters in our LAE model. This figure represents our main result: M_{\min} , M_{\max} , and f_{occ} cannot be tightly constrained from the available observations.

The preferred 1σ range for the masses is $10.0 < \log M_{\min} < 11.2$ and $11.0 < \log M_{\max} < 13.0$. f_{occ} is completely determined by M_{\min} from $f_{\text{occ}} = 0.004$ when $\log M_{\min} = 10.0$ to $f_{\text{occ}} = 0.38$ when $\log M_{\min} = 11.15$. We compute the power-law dependence between f_{occ} and M_{\min} to be

$$f_{\text{occ}} = 0.05 \left(\frac{M_{\min}}{10^{10} h^{-1} M_{\odot}} \right)^{0.77}. \quad (2)$$

We remind the reader that models with $\log f_{\text{occ}} > 0.00$ ($f_{\text{occ}} > 1$) correspond to cases where the halo number density is smaller than the number of observed LAEs but are still considered consistent because of the expected uncertainty in

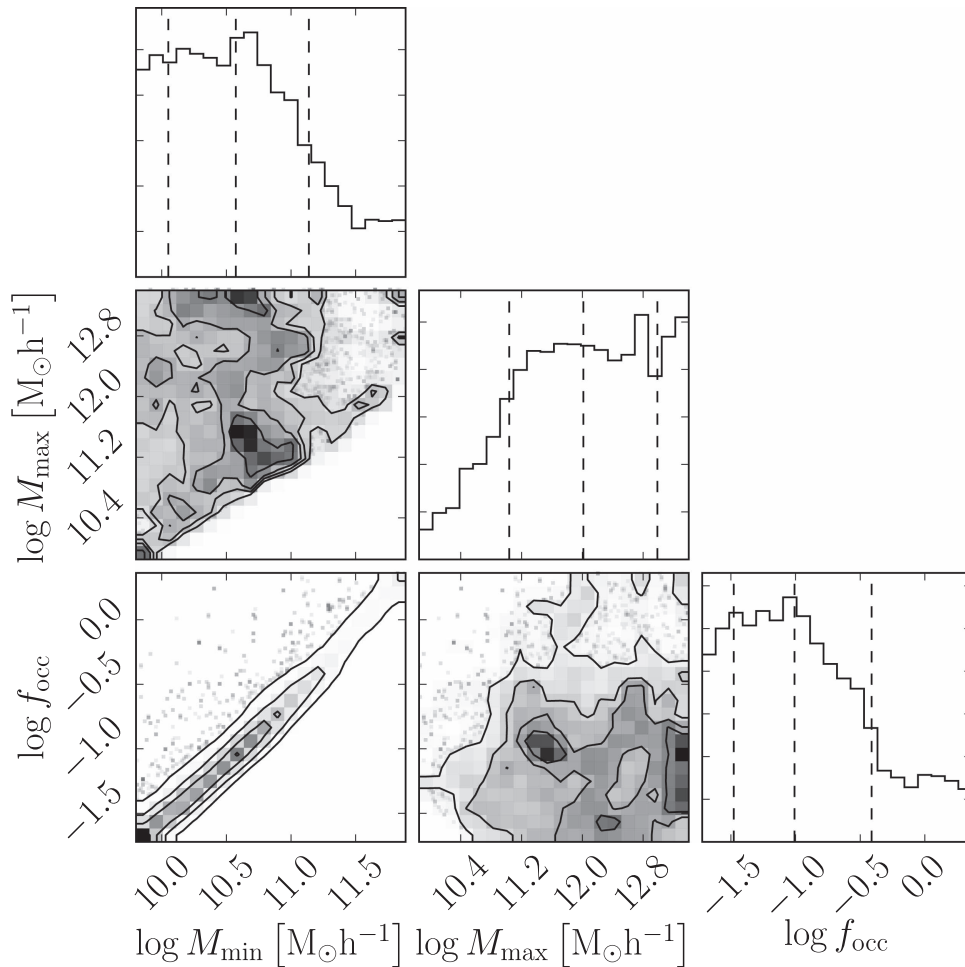


Figure 3. One- and two-dimensional projections of the posterior probability distributions of M_{\min} , M_{\max} , and f_{occ} . Vertical lines in the one-dimensional projections correspond to the percentiles 14, 50, and 84 of the distributions. Thick lines in the two-dimensional planes mark the 1σ , 2σ , and 3σ contours of the corresponding joint distributions. The models with $\log f_{\text{occ}} > 0.00$ ($f_{\text{occ}} > 1$) correspond to models where the number density of halos is smaller than the number of observed LAEs, but we consider them as consistent because of the uncertainty in the median number density due to cosmic variance. See Figure 1 and Section 2.4 for details.

the median number density of LAEs due to cosmic variance (see Figure 1 and Section 2.4).

The result in Equation (2) can be qualitatively understood as follows. A choice of $f_{\text{occ}} < 1$ means that the halo mass function is scaled down. If M_{\min} decreases, the total number of halos increases because less massive halos are more abundant. This requires smaller values of f_{occ} to keep the total number of halos hosting LAEs equal to the number of LAEs in observations.

3.2. Median Halo Mass and Halo Mass Width within Models

We now compute the median mass, $\log M_{50}$, and the 1σ halo mass width, $\Delta \log M_{1\sigma} \equiv \log M_{84} - \log M_{16}$, of the LAE hosting halos for each model \mathcal{M} (here M_p represents the p percentile of the mass distribution); $\log M_{50}$ and $M_{1\sigma}$ allow a direct comparison between our results and previous results in the literature (e.g., Hayashino et al. 2004; Gawiser et al. 2007; Ouchi et al. 2010; Bielby et al. 2016) that used a more simplified semianalytical approach.

In Figure 4 we show the $\log M_{50} - \Delta \log M_{1\sigma}$ plane for the models selected to have $\chi^2_{\mathcal{M}}/2 < 1$, which roughly correspond to the 1σ region in the posterior distribution for the model

parameters. The color encodes the $\chi^2_{\mathcal{M}}/2$ value. The median mass has a wide distribution spanning two orders of magnitude.

From this distribution we obtain $\log M_{50} = 10.8 \pm 0.6$ or equivalently $M_{50} = 6.3^{+18.8}_{-4.7} \times 10^{10} h^{-1} M_{\odot}$. The 1σ uncertainty for this median value is estimated from the 16th and 84th percentile values in the $\log M_{50}$ distribution.

Our result is consistent within the statistical uncertainties with previous estimates reported by Bielby et al. (2016) ($M_{50} = 10^{11.0 \pm 0.6} h^{-1} M_{\odot}$), Gawiser et al. (2007) ($M_{50} = 10^{10.9 \pm 0.9} h^{-1} M_{\odot}$), and Ouchi et al. (2010) ($M_{50} = 10^{10.8^{+1.8}_{-0.8}} h^{-1} M_{\odot}$) using semianalytical approaches. Our result is slightly below the Bielby et al. (2016) median mass estimation. This could probably be attributed to the fact that Bielby et al. (2016) only represent a particular field of the universe while our $\log M_{50}$ represents the median mass of the universe after taking into account cosmic variance that allows lower-mass halos to be also consistent with observations.

Figure 4 also shows something that semianalytical approaches were not able to predict. The mass width, $\Delta \log M_{1\sigma}$, which is the width of the mass distribution for a given model with fixed M_{\min} and M_{\max} , has a median value and 1σ uncertainty of $\Delta \log M_{1\sigma} = 0.51^{+0.10}_{-0.30}$ dex. This means that the mass range for halos hosting LAEs is very narrow. There is only a factor of 2–4 between the lower and upper mass

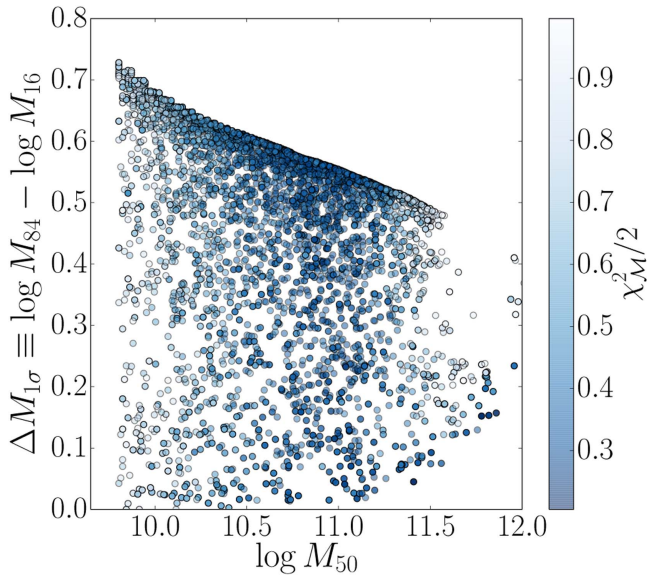


Figure 4. Median mass, $\log M_{50}$, and mass width $\Delta \log M_{1\sigma} = \log M_{84} - \log M_{16}$, for all models with low $\chi^2_{\mathcal{M}}/2 < 1.0$. The median mass spans two orders of magnitude with a mean value around $10^{10.5} h^{-1} M_{\odot}$. The mass width $\Delta \log M_{1\sigma}$ has a median of 0.55 dex (a factor of ~ 3 in mass) with a maximum value of 0.7 dex (a factor of 5). This means that all models consistent with observations have a very narrow range in halo mass.

boundary of the central 68th percentile of the mass distribution. We emphasize that M_{\max} can still be very large while the width is small due to the asymmetry in the DM halo mass function.

Summarizing, the median mass could be anything in the range from $10^{10.2} h^{-1} M_{\odot}$ to $10^{11.4} h^{-1} M_{\odot}$ (a 2.0 dex range), but the 1σ width of the mass distribution is highly constrained to be between 0.2 and 0.6 dex.

In Figure 2 we show the computed $\text{ACF}_{\mathcal{M}}$ of models with $\log M_{\min} = 10.5$ and different values of $\Delta \log M$. We can see that the clustering gets slightly stronger for larger values of $\Delta \log M$. Nevertheless, due to the large impact of cosmic variance at the volume of the current observations, all the models are basically consistent within errors. The last result, together with the large Poissonian observational error in the ACF, explains the current difficulty in putting tighter constraints on $\log M_{\max}$ in our model.

3.3. Constraining Dark Matter Halo Mass with Cosmic Variance

Figure 1 shows the halo number distribution (HND) in the mock fields of the simulation for different models \mathcal{M} . By simple inspection one can infer that the distribution width increases with M_{\min} .

In Figure 5 we confirm this trend by plotting the HND 1σ width for good models ($\chi^2_{\mathcal{M}}/2 < 1$) as a function of $\log M_{\min}$. We find that considering all 27 mock fields, the 1σ width, $W_{1\sigma}$, increases with M_{\min} following

$$W_{1\sigma} = (0.138 \pm 0.002) \times \left(\frac{M_{\min}}{10^{11} h^{-1} M_{\odot}} \right)^{0.177 \pm 0.009}. \quad (3)$$

This result opens the possibility of constraining the $\log M_{\min}$ (as well as the median mass) of LAEs by simply measuring the width of the distribution of observed LAE along several observational fields. This idea has been already explored for $z > 6$ galaxies (Robertson 2010).

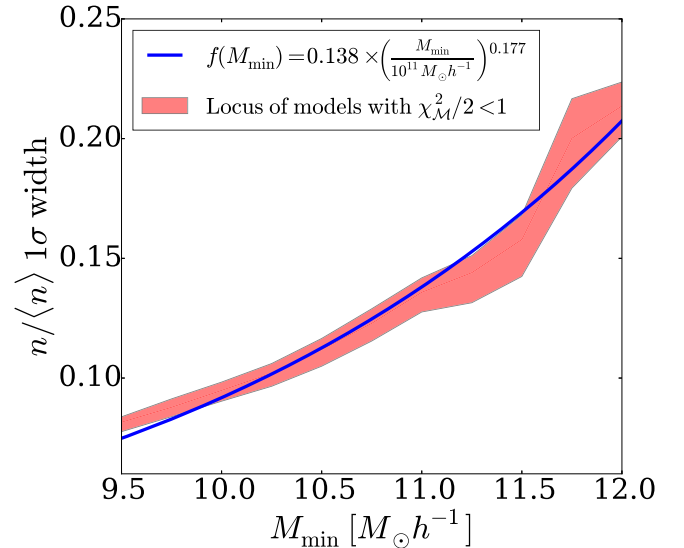


Figure 5. 1σ width of the number density distribution as a function of M_{\min} . Only the models that have a good match with observations ($\chi^2_{\mathcal{M}}/2 < 1$) are included. The strong dependence of $\langle n \rangle$ with M_{\min} was parameterized as a power law with the parameters shown in the legend.

To keep the validity of Equation (3), there should at least be 27 fields (to follow the same numbers we use here) of size $\sim 1 \text{ deg}^2$ with the same observational conditions (filters, equivalent width cuts). However, repeating the same kind of cosmic variance study for different survey strategies should help to provide a constraint of the same kind.

To keep the validity of Equation (3), there should be of the order of 27 fields or even a larger number. Each of these fields should be $\sim 1 \text{ deg}^2$ in size and reproduce the same observational conditions of Bielby et al. (2016) in terms of filters and equivalent width cuts. However, repeating the same kind of cosmic variance study for different survey strategies should help to provide an equivalent constraint.

3.4. Hints of Larger Uncertainties

Figure 3 shows an interesting feature for M_{\min} . The histogram (upper left) showing the probability for this parameter given the observational constraints does not considerably decrease for lower masses. This raises the following question: down to which values of M_{\min} does the probability significantly decrease?

Lower values of M_{\min} would imply a lower f_{occ} and larger uncertainties on the parameters M_{\min} , f_{occ} , and M_{50} . In other words, the limitations in our priors for M_{\min} (which set the limitations for the kind of numerical simulation we use) might be underestimating the uncertainties and slightly biasing our results for M_{\min} , f_{occ} , and M_{50} .

Although Figure 4 suggests that models with lower values of M_{50} , i.e., lower values of M_{\min} , have a higher $\chi^2_{\mathcal{M}}$ and could thus be discarded, we suggest that a proper test of this hypothesis requires performing new cosmological simulations to be able to probe M_{\min} masses below our limit of $6 \times 10^9 h^{-1} M_{\odot}$.

This highlights the main thesis of this paper, namely, the high impact that cosmic variance can have on constraining the parameters of our model. With the current limitations in the cosmological simulations available to the public, we cannot

significantly decrease our prior on M_{\min} . We consider such tests beyond the scope of this paper.

This extension to larger priors cannot change other central results of our paper, such as the relationship between f_{occ} and M_{\min} , the narrow mass range for individual models, and the possibility of constraining the halo mass with cosmic variance on the number density counts. However, we suggest that future work building on the technique presented in this paper should count with simulations that extend at least one order of magnitude below our limit of $6 \times 10^9 h^{-1} M_{\odot}$.

4. CONCLUSIONS

In this paper we studied the impact of cosmic variance and observational uncertainties in constraining the mass range and occupation fraction of DM halos hosting LAEs. To this end we used the BolshoiP N -body simulation to construct 27 mock fields with the same typical size of observed fields at $z = 3.1$ ($\sim 1 \text{ deg}^2$). In our model a DM halo with mass in the range $M_{\min} < M_h < M_{\max}$ can only host one detectable LAE at most. We explored the parameter space determined by M_{\min} and M_{\max} using affine-invariant Monte Carlo Markov Chain minimization to match the observed ACF and mean number density of LAEs. It is the first time that such a thorough exploration of the cosmic variance impact on LAE statistics is presented in the literature.

We find that once cosmic variance is taken into account, the observational data can only put weak constraints on M_{\min} and f_{occ} . When we translate these loose constraints into the median mass, we find a general consistency with previous works (Hayashino et al. 2004; Gawiser et al. 2007; Ouchi et al. 2008), especially with the most recent one (Bielby et al. 2016).

This work also highlights the need to explore a wider range on the prior of M_{\min} if one wants to better estimate the uncertainty on M_{\min} , f_{occ} , and M_{50} . In this paper we use priors given by previous observations, but our results suggest that once cosmic variance is explicitly taken into account, even lower values of M_{\min} might be, in fact, consistent with observations. We recommend that future work making use of the technique presented in this paper should extend the prior for M_{\min} to masses at least one order of magnitude below $6 \times 10^9 h^{-1} M_{\odot}$.

Nevertheless, our analysis allowed us to draw two results that can be used to put tighter constraints on M_{\min} and f_{occ} once upcoming large LAE surveys, such as the HETDEX project (Adams et al. 2011) and the HSC ultra deep survey, are available:

1. f_{occ} is uniquely determined by M_{\min} , regardless of M_{\max} . A precise determination of M_{\min} will thus fix f_{occ} . As a consequence of this, for a given M_{\min} the median mass is also fixed.
2. The width of the LAE number distribution function obtained over several fields of $\approx 1 \text{ deg}^2$ is tightly correlated with M_{\min} . That measurement with next-generation surveys will be able to constrain M_{\min} within a factor of ~ 2 .

We also find that the 1σ width of the mass distribution is highly constrained to be between 0.2 and 0.6 dex. This result can be used to test different models for LAE formation in a cosmological context to better understand why observable LAEs seem to be constrained into a narrow halo mass range.

The authors acknowledge the referee, Nelson Padilla, for useful remarks and questions that helped us to improve the quality and clarity of our paper. The authors also acknowledge Valentino Gonzalez, Eric Gawiser, and Mark Dijkstra for the fruitful discussion and suggestions that also contributed to improving the quality of this work.

J.E.M.-R. acknowledges “CONICYT-PCHA/doctorado Nacional para extranjeros/2013-63130316” for their PhD scholarship support.

J.E.F.-R. acknowledges financial support from Vicerrectoría de Investigaciones at Uniandes from an FAPA project.

The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.Gauss-centre.eu) and the Partnership for Advanced Supercomputing in Europe (PRACE; www.prace-ri.eu) for funding the MultiDark simulation project by providing computing time on the GCS Supercomputer SuperMUC at Leibniz Supercomputing Centre (LRZ; www.lrz.de). The Bolshoi simulations have been performed within the Bolshoi project of the University of California High-Performance AstroComputing Center (UC-HiPACC) and were run at the NASA Ames Research Center.

REFERENCES

- Adams, J. J., Blanc, G. A., Hill, G. J., et al. 2011, *ApJS*, 192, 5
- Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013, *ApJ*, 770, 57
- Bielby, R. M., Tummungpak, P., Shanks, T., et al. 2016, *MNRAS*, 456, 4061
- Blanc, G. A., Adams, J. J., Gebhardt, K., et al. 2011, *ApJ*, 736, 31
- Bond, N. A., Gawiser, E., Gronwall, C., et al. 2009, *ApJ*, 705, 639
- Colberg, J. M., White, S. D. M., Yoshida, N., et al. 2000, *MNRAS*, 319, 209
- Dayal, P., Ferrara, A., Saro, A., et al. 2009, *MNRAS*, 400, 2000
- Dijkstra, M. 2014, *PASA*, 31, e040
- Dijkstra, M., & Kramer, R. 2012, *MNRAS*, 424, 1672
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306
- Forero-Romero, J. E., Yepes, G., Gottlöber, S., et al. 2011, *MNRAS*, 415, 3666
- Forero-Romero, J. E., Yepes, G., Gottlöber, S., & Prada, F. 2012, *MNRAS*, 419, 952
- Garel, T., Blaizot, J., Guiderdoni, B., et al. 2012, *MNRAS*, 422, 310
- Gawiser, E., Francke, H., Lai, K., et al. 2007, *ApJ*, 671, 278
- Guaita, L., Gawiser, E., Padilla, N., et al. 2010, *ApJ*, 714, 255
- Hayashino, T., Matsuda, Y., Tamura, H., et al. 2004, *AJ*, 128, 2073
- Hayes, M. 2015, *PASA*, 32, e027
- Jose, C., Srianand, R., & Subramanian, K. 2013, *MNRAS*, 435, 368
- Klypin, A., Yepes, G., Gottlöber, S., Prada, F., & Heß, S. 2016, *MNRAS*, 457, 4340
- Klypin, A. A., Trujillo-Gomez, S., & Primack, J. 2011, *ApJ*, 740, 102
- Koehler, R. S., Schuecker, P., & Gebhardt, K. 2007, *A&A*, 462, 7
- Landy, S. D., & Szalay, A. S. 1993, *ApJ*, 412, 64
- Laursen, P., Duval, F., & Östlin, G. 2013, *ApJ*, 766, 124
- Laursen, P., & Sommer-Larsen, J. 2007, *ApJL*, 657, L69
- Neufeld, D. A. 1991, *ApJL*, 370, L85
- Nilsson, K. K., Møller, P., Möller, O., et al. 2007, *A&A*, 471, 71
- Norberg, P., Baugh, C. M., Gaztañaga, E., & Croton, D. J. 2009, *MNRAS*, 396, 19
- Orsi, A., Lacey, C. G., & Baugh, C. M. 2012, *MNRAS*, 425, 87
- Ouchi, M., Shimasaku, K., Akiyama, M., et al. 2008, *ApJS*, 176, 301
- Ouchi, M., Shimasaku, K., Furusawa, H., et al. 2010, *ApJ*, 723, 869
- Partridge, R. B., & Peebles, P. J. E. 1967, *ApJ*, 147, 868
- Peebles, P. J. E. 1980, The Large-Scale Structure of the Universe
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, *A&A*, 571, A16
- Rhoads, J. E., Malhotra, S., Dey, A., et al. 2000, *ApJL*, 545, L85
- Riebe, K., Partl, A. M., Enke, H., et al. 2013, *AN*, 334, 691
- Robertson, B. E. 2010, *ApJL*, 716, L229
- Somerville, R. S., Lee, K., Ferguson, H. C., et al. 2004, *ApJL*, 600, L171
- Springel, V., White, S. D. M., Jenkins, A., et al. 2005, *Natur*, 435, 629
- Verhamme, A., Schaerer, D., & Maselli, A. 2006, *A&A*, 460, 397
- Yajima, H., Choi, J.-H., & Nagamine, K. 2012, *MNRAS*, 427, 2889
- Yajima, H., Li, Y., Zhu, Q., et al. 2014, *MNRAS*, 440, 776
- Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2011, *ApJ*, 736, 59