Mon. Not. R. Astron. Soc. **000**, 000–000 (0000) Printed 10 April 2013 (MN LATEX style file v2.2)

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

Julian Mejía[2] and Jaime E. Forero-Romero[1]

[1]

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

10 April 2013

ABSTRACT

We use simulations and recent observations 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 cosmic variance between observed and mock fields. We build a model for LAEs on top of a large N-body cosmological simulation (Bolshoi) where the basic principle is that a dark matter halo can only host one LAE with a probability $f_{\rm occ}$ if its mass is found withing a certain range mass range delimited by two threshold values, $M_{\rm min}$ and $M_{\rm max}$. We generate hundreds of mock catalogs in a thourough exploration of the parameter space of this model. These catalogs are constructed in such a way as to reproduce the spatial correlation between the different observed fields. We find that the models that best match the observed cosmic variance statistics are those with halo masses in the range $M_{\rm min} =$ and $M_{\rm max} =$ with and occupation fraction that scales as $f_{\rm occ} =$ with the minimum mass. We find that the angular correlation function is in agreement with the observational constraints mostly due to the cosmic variance in the small angular fields where this statistics has been computed so far. We also present predictions for future measurements of the correlation function on larger continous fields. We make available the mock data for the best models in a public repository.

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

1 INTRODUCTION

Lyman-α emitting galaxies (LAEs) have become in the last decade a central topic in studies of structure formation in the Universe. The reason is the diverse range of fields where they are helpful, LAEs can be used as probes of reionization, tracers of large scale structure, signposts for low metallicity stellar populations and markers of the the galaxy formation process through cosmic history.

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 largely determines key properties of the galaxy such as its stellar mass and star formation rate. Processes that regulate the star formation cycle are also though 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 LCDM.

Some theoretical approaches to this problem have been based on a forward modelling. Starting from the DM halo population, the corresponding intrinsic star formation properties are infered 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 hidrodynamical simulations.

Added to the uncertainties in the astrophysical processeses describing star formation in galactic populations, a highly debated steps in this approach is the calculation of the fraction of Lyman-α photons that escape the galaxy to the observer. Given the resonance nature of the line, the radiative transfer of Lyman-α is sensitive to the density, temperature, topology and kinematics of the neutral Hydrogen in the interstellar medium (ISM).

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 ΛCDM 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 km s^{-1} Mpc^{-1}.

2 METHODOLOGY

In this paper we constrain the typical mass of dark matter halos hosting LAES at $z = 3.1$. Our model is based on the number density information obtained in the recent large scale survey presented by XXX where XXX LAEs are detected over 7 fields of $\sim 46 \times 35$Mpc$^2h^{-2}$ in area comoving in area corresponding to observed fields of XXX deg^2.

Both the spatial distribution and luminosities of the galaxies have, at least in a statistical sense, informatio to constrain theoretical model of LAEs. The most detailed theoretical models are also faced to diverse physical and astrophysical uncertainties in obtaining statistical prediction for the Lyman-alpha line. This uncertainties are largest impediment to consruct an ab-iitio model for LAEs.

In this paper we want to step back and reduce the complexity of our model, with the sole objective of reproducing the cosmic variance in the number density of LAEs. Afterwards we will interpret the implications of this result for physical models for Lyman-alpha emitting galaxies.

Our model is based on the predictions of a large volume high resolution N-body simulation describing the gravitational dynamics of dark matter. We do not have an strong bias towards the theoretical expectation of what the mass of the dark matter halo hosting the galaxy should be. Instead, we fully explore the parameter space of our simplified model. The only cut we impose is that observed LAEs do not reside in dark matter halos with masses less than 10^{10}hMsun [citation].

In the following subsections we describe the most relevant features of the observational data, the N-body simulation we use, our model and its parameters together with the method to compare its predictions against observations.

2.1 The Observational Constraints

Our observational reference are the recently published results of a panoramic survey of LAES at $z = 3.1$ by Yamada et al. (2012). This survey was conducted with the Subaru 8.2m telescope and the Subaru Prime Focus Camera, which has a field of view covering 34×27 arcmin, corresponding to a comoving scale of 46×35 Mpc h^{-1} at $z = 3.09$. The narrow band filter is centered at 4977 Åwith a 77Åwidth, corresponding to the redshift range $z = 3.062 - 3.125$ and 41 Mpc h^{-1} comoving scale for the detection of the Lyman-α line centered at $z = 3.09$.

The choice to have only one the data from Yamada et al 2012 as reference was made because their surveys is the largest in area with a set of homogenous conditions that define the LAE sample. Other surveys by XXX an XXX that cover similar regions, but they use different criteria on the equivalent width (EW) cuts to construct the LAE samples. Different cuts in the EW can change the number of LAEs to be included in the catalog. This cuts have an impact on the fainter LAEs which are more abundante than brighter ones. Different definitions of the EW cuts can yield number densities different by a factor of two [REF, I think Yamada has some numbers].

The survey covered four independent fields. The first is the SSA22 field of 1.38 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 deg^2 and 386 LAEs. The third and fourth fields are the Subaru Deep Field (SDF) with 0.22 deg^2 and 196 LAEs, and the fild arotund the Great Observatory Optical Deep Survey North (GOODS-N) with 0.24 deg^2 and 185 LAEs. In Table 1 we summarize the values we use in throughout this paper for the each field, covered area, measured surface LAE number density and inferred number volume density.

2.2 The Simulation and Halo Catalogs

The Bolshoi simulation was performed in a cubic volume of 250 h^{-1} Mpc on a side. It includes dark matter distribution is sampled using 2048^3 particles, which translates into a particle mass of $m_{\rm p} = 1.35 \times 10^8$ h^{-1} M$_\odot$. The cosmological parameters are consistent with a WMAP5 and WMAP7 data with a matter density $\Omega_{\rm m} = 0.27$, 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 $\sigma_8 = 0.82$ [REF].

We use halo catalogs constructed with a Friend-of-Friends (FOF) algorithm with a linking lenght of 0.17 times the interparticle distance. We have veryfied that the main results we present in this paper also hold if instead we use halo catalogs constructed from a the Bound Density Maxima (BDM) algorithm (Klypin et al. 1999) that are defined to have an density of 200 times the critical density. 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 [1] (Riebe et al. 2011).

2.3 Populating Halos with LAEs

The model that populates halos with LAEs is based on a one-to-one correspondence: each halo can only host a single LAE. There are three physical parameters in the model: the halo mass range $M_{\rm min} < M_{\rm halo} < M_{\rm max}$ where LAEs reside and the fraction $f_{\rm occ}$ of such halos that effectively host a LAE. In what follows we will describe by the letter $\mathcal{M}$ a model defined by these three parameters $M_{\rm min}$, $M_{\rm max}$ y $f_{\rm occ}$.

We stress that we do not intent to build a model for the luminosity of each LAE. Physically speaking we are primarely interested in constraining the halo mass above which there are detectable LAEs. under the conditions defined by Yamada et al.

For each mode $\mathcal{M}$ we create mock field from disjoint

[1] `http://www.multidark.org/MultiDark/`

volumes in the simulation with the same geometry probed by Suprime-CAM and the narrow band filter, namely $46 \times 35 \times 41\ h^{-3}\mathrm{Mpc}^3$ where the last dinemsion goes in the redshift direction, corresponding to a total area of 880 arcmin^2 for each mock field. There is a total $5 \times 7 \times 6 = 210$ of such sub-volumes in a snapshot of the Bolshoi simulation.

Next we group these 210 mock fields in three different ways to construct the LAEs number density distribution. The first way (match method) we follow the observational setup and constructs 15 different mock surveys, each one composed of 12 mock fields, out of which 7 correspond to contiguous sub-boxes in the simulation to mimick the whole SSA22, 3 are also contiguous between them but not to the first 7 fields to mimick the SXDS fields and finally 2 non-contiguous fields that correspond to the SDF and GOODS-North fields. This will produce 15 different distributions for the number density for a given model M. The second (random method) is similar to the first one. There aare 15 different mock surveys with 12 mock fields each, but this time each field corresponds to uncorrelated sub-boxes in the simulation. The third (full method) way in only has 1 mock survey containing all the 210 mock fields, in this setup there is only one predicted number density distribution for each model M

The advantage of these three sampling ways is that they allow us to explore the effects of both cosmic variance and the correlation between fields. Comparing the results of the first and second method will help us to quantify the effect of field correlation, while compating the first and the third method will serve us to gauge the impace of cosmic variance.

2.4 Model Sampling and Selection

We generate a series of models $\mathcal{M}$ with different input parameters$\{M_{\rm min}, M_{\rm max}, f_{\rm occ}\}$ as follows. $M_{\rm min}$ and $M_{\rm max}$ are allowed to take 30 different values evenly spaced by 0.1 dex, $M_{\rm min}$ ranges from $\log_{10} M_{\rm min} = 10.0$ up to $\log_{10} M_{\rm min} = 12.9$, while $M_{\rm max}$ range from $\log_{10} M_{\rm min} = 10.1$ up to $\log_{10} M_{\rm min} = 13.0$. The occupation fraction $f_{\rm occ}$ takes 100 different values from 0.01 to 1.00 regularly spaced by 0.01. In total the number of different sets of input parameters to be explored is $30 \times 30 \times 100 = 9 \times 10^4$.

For each model $\mathcal{M}$ we compute the LAE surface density distributions for the three different ways of grouping the mock fields, as described in the previous section. For each sub-volume we project the positions of the LAE hosting halos along the z direction and calculate its surface number density in units of sources per arcmin^2. For each number density distribution we perform a Kolmogorov-Smirnov againsg the 12 surface density observational values. From this test we obtain the value $0 < P < 1$ to reject the null hypothesis, namely that the two data sets come from the same distribution. In this paper we use values of $P > 0.1$ to consider that the simulated and observed number densities come from the same distribution.

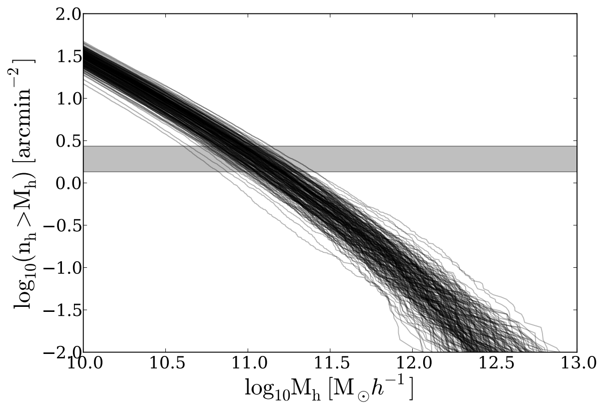


Figure 1. Cumulative mass function of dark matter haloes in the 210 sub-volumes of $46 \times 35 \times 41\ h^{-3}Mpc^3$. The variation in the total number of dark matter halos per sub-volume evidences the effect of cosmic variance at such sub-volume scale. It is also appreciable the low population $\lesssim 10^{-3}h^2Mpc^{-2}$ of halos with $log(M/M_\odot) > 12.0$

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 correspond 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 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 $M_{\rm min} > 3 \times 10^{11} h^{-1}\mathrm{M}_\odot$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 $M_{\rm min} < 3 \times 10^{10} h^{-1}\mathrm{M}_\odot$. Models with this minimum mass hava a chance for successfuly reproducing observations if the occupation fraction $f_{\rm occ} < 1$ 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 $M_{\rm min}$-$M_{\rm max}$ 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 with $P > 0.1$. The third panel shows the result for the method `Full`, in this case the color

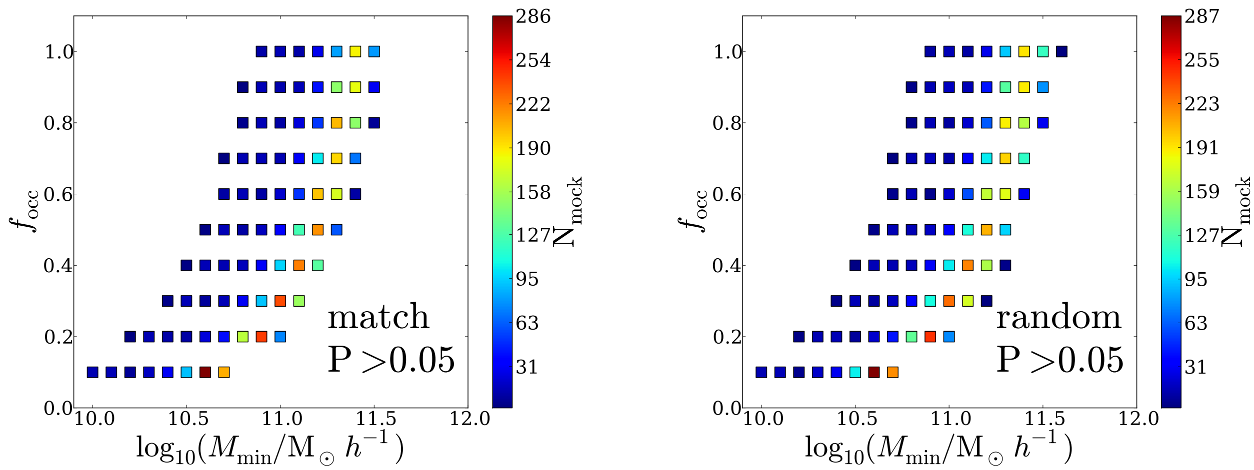


Figure 3. $M_{\rm min}$-$M_{\rm max}$ plane for all models with $P > 0.05$.

code correspond 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 $M_{\rm min} > 10^{11.5}h^{-1}{\rm M}_\odot$are not compatible with observations. There is a second regime for masses below $M_{\rm min} = 3 \times 10^{10}$ any values for $M_{\rm min}$ and $M_{\rm max}$ can be made compatible with observations, provided that $f_{\rm occ}$ is fine tuned to do it. In an intermediate mass regime, for minimum mass values $3\times10^{10}h^{-1}{\rm M}_\odot < M_{\rm min} < 3 \times 10^{11}h^{-1}{\rm M}_\odot$ only a limited range of models with $M_{\rm max}$ with occupation fraction $f_{\rm occ} \sim 1$ is able to reproduce observations.

In these three different mass regimes the occupation monotonically decreases as a function of the minimum halo mass $M_{\rm min}$. In Figure XXX we show this treend 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 $f_{\rm occ} \sim 1$, $0 < f_{\rm occ} < 1$ and $f_{\rm occ} \sim 0$, 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 sucessful model M present a high value $P > 0.1$, only a modest seems to be consistent with observations. This shows that the cosmic variance is still present on the physical scales probed by observations.

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.

3.3 Observational constraints on the occupation fraction

Hayes et al. (2010) constrained the value of $f_{\rm occ}$ at $z = 2.2$ to be $f_{\rm occ} = 0.10$. This estimation was based on blind surveys of the Hα and Lyman alpha line with the European Southern Observatorio (ESO) Very Large Telescope (VLT). Using corrections by extinction to obtain an estimate for the intrinsic Hα luminosity, and using values for the theoretical expectation of the ratio Lymanα/Hα they derive an bulk escape fraction for the Lymanα radiation of $f_{\rm esc} = (5.3 \pm 3.8)\%$ or $f_{\rm esc} = (10.7\pm2.8)$ if a different dust correction is used. 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 $f_{\rm occ} = 0.10$. Obervations are consistent with these numbers are consistent with being constant for every galaxy regardless of its luminosity.

For the cosmological parameters used in this paper the age of the universe between $z = 3.1$ and $z = 2.2$ has changed by ~ 1 Gyr, it is reasonable to assume that the physical conditions that determine the escape fraction $f_{\rm esc}$ and the occupation fraction $f_{\rm occ}$ remain constant. This assumption allows us to further pick models that have an occupation fraction of $f_{\rm occ} = 0.10$. Under this selection only 18 models can be selected. Considering an occupation fraction $f_{\rm occ} = 0.20$ another 39 models can be considered. The minimum and maximum mass are found in the appendix in Table XX.

3.4 Constraints from the Angular Correlation Function

We calculate the mean angular correlation function (ACF) for all models with $P > 0.05$ using the match and random methods. These correlation functions is calculated over the denseset subfield in all the 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 4 (match) and Figure 5 (random) present such comparison. The error bars in these figures represent the

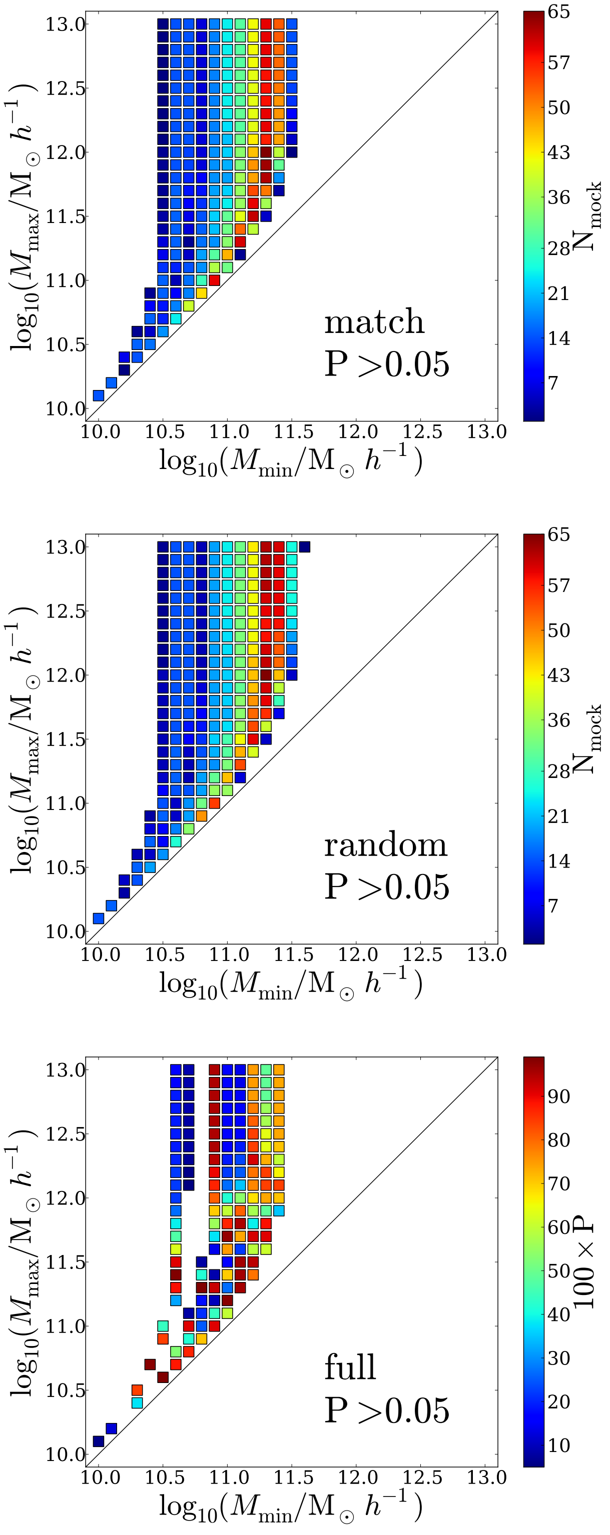


Figure 2. $M_{\rm min}$-$M_{\rm max}$ plane for all models with $P > 0.05$ in the three different of grouping the mock fields. In the case of the `match` and `random` methods the color code corresponds to the number of mock surveys that are found to be compatible with observations. For the `full` method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

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

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

where the sum is done over all the angle values θ_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 oserved in the SSA22 region by Yamada et el in 2012 it is expected that the predicted ACF in the SSA22 region should reproduce this observation. In the left panel of figure 4 we can see the predicted ACFs an their corresponding standard deviation over the seven fields that mock the SSA22 region. It can be seen that the model with $M_min = 10.6$ 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.

3.5 Prediction for the angular correlation function

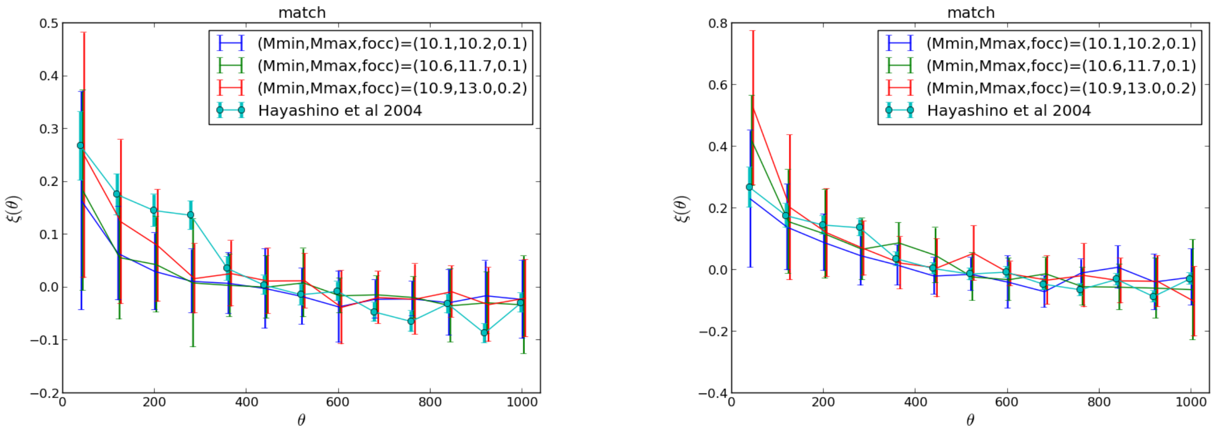
We also present the ACF in the case of the full SSA22 region which has been homogeneously observed by (Yamada et al. 2012). To this date the observational ACF has not been reported in the litereature, therefore our calculations can be considered as predictions.

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

4 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 from differente observed fields to mock fields based on a simplified model with three basic parameters. The halo masses that define the halos where LAEs can be found, $M_{\rm min} < M_{\rm h} < M_{\rm max}$, and the fraction of the halos in this range that are actully occupied, $f_{\rm occ}$. After a thorugh exploration of the paremeter space we are able to constrain the mass range of dark matter halos hosting LAEs to be in the range $< M_{\rm h} <$ and a corresponding occupation fraction that escales as $f_{\rm occ} = M_{\rm min}$.

In this letter, our estimates are based on a large volume, high resolution, dark matter only Nbody simulation. From the halo catalogs in the simulation a fraction of halos in a



f

Figure 4. mean ACFs and their correponding standard deviation (error bars) of some selected models in different mass ranges over the 7 subfields f the SSA22 field (left) and the entire 12 field sample (rigth) using thematch configuration.

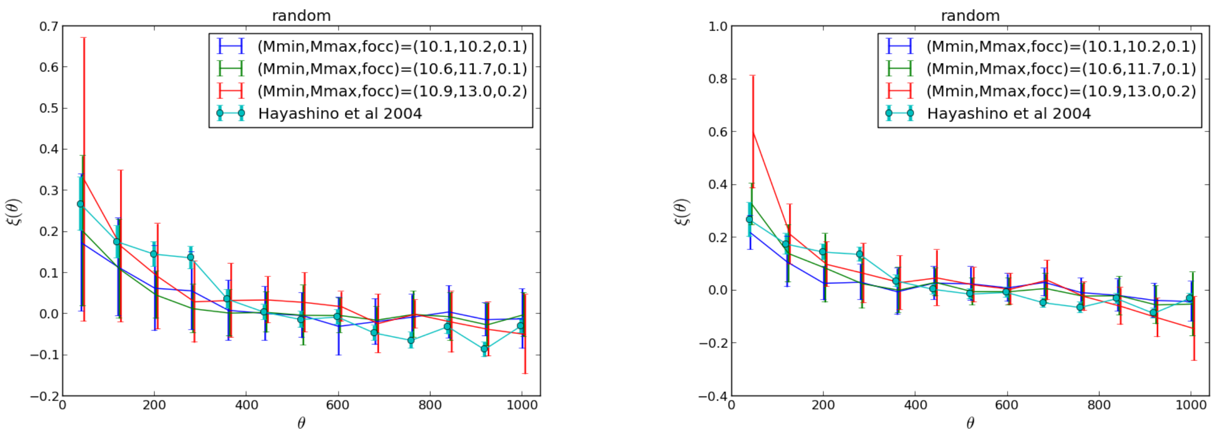


Figure 5. mean ACFs and their correponding standard deviation (error bars) of some selected models in different mass ranges over the 7 subfields of the SSA22 field (left) and the entire 12 field sample (rigth) using the random configuration.

given mass range to host LAEs. We use this information to construct mock observations to derive the statistics that can be compared against observations. The spatial information in this catalogs is also kept, which allows us to compute angular correlation functions and include these results in our analysis.

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 `random` method. We do not find a significant difference between the two first methods.

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 $< M_{\rm h} <$. Including the information from the angular correlation function (ACF) does not allows 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

ACKNOWLEDGMENTS

REFERENCES

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

Klypin A., Gottlöber S., Kravtsov A. V., Khokhlov A. M., 1999, ApJ, 516, 530

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

Riebe K., Partl A. M., Enke H., Forero-Romero J., Gottloeber S., Klypin A., Lemson G., Prada F., Primack J. R., Steinmetz M., Turchaninov V., 2011, ArXiv e-prints

Yamada T., Nakamura Y., Matsuda Y., Hayashino T., Yamauchi R., Morimoto N., Kousai K., Umemura M., 2012, AJ, 143, 79