Mon. Not. R. Astron. Soc. **000**, 000–000 (0000) Printed 8 March 2016 (MN LATEX style file v2.2)

# Uncertainty on the galaxy-halo connection for Lyman-$\alpha$ emitters at $z = 3$

J. Mejia-Restrepo ⋆[1], Jaime E. Forero-Romero †[2]

[1],[3] *Departamento de Astronomía, Universidad de Chile, Camino el Observatorio 1515, Santiago, Chile*
[2] *Departamento de Física, Universidad de los Andes, Cra. 1 No. 18A-10, Edificio Ip, Bogotá, Colombia*

8 March 2016

**ABSTRACT**
We present

**Key words:** Cosmology: theory - large-scale structure of Universe - Methods: data analysis - numerical - N-body simulations

## 1 INTRODUCTION

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

In most of those cases, capitalizing 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 dark matter 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 (Behroozi et al. 2013). 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 Ly$\alpha$ luminosities and observed Ly$\alpha$ luminosities. This approach is usually implemented through semi-analytic models (Garel et al. 2012; Orsi et al. 2012) and full N-body hydrodynamical simulations (Laursen & Sommer-Larsen 2007; Dayal et al. 2009; Forero-Romero et al. 2011; Yajima et al. 2012).

The downside of these calculations is the uncertainty in the estimation of the escape fraction of Ly$\alpha$ photons. Given the resonant nature of the 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 Ly$\alpha$ escape fraction in high redshift galaxies is still matter of open debate (Neufeld 1991; Verhamme et al. 2006; Forero-Romero et al. 2011; Dijkstra & Kramer 2012; Laursen et al. 2013; Orsi et al. 2012).

A different approach to infer the typical mass of halos hosting LAEs is based on the spatial clustering information. This approach uses the fact that in CDM cosmologies the spatial clustering of galaxies on large scales is entirely dictated by the halo distribution (Colberg et al. 2000), which in turn has a strong dependence on halo mass. Using measurements of the angular correlation function of LAEs, observers have put constraints on the typical mass and occupation fraction of the putative halos hosting these galaxies (Hayashino et al. 2004; Gawiser et al. 2007; Nilsson et al. 2007; Ouchi et al. 2010; Bielby et al. 2016). In these studies the observations are done on fields of $\sim 1$ deg$^2$ and the conclusions derived on the halo host mass do not elaborate on the uncertainty resulting from the cosmic variance on these fields.

In this letter 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 to compare them against the observations in Bielby et al. (2016) using the angular correlation function. We use a simple model to populate a halo in the simulation with a LAE assuming a minimum ($M_{\rm min}$) and maximum mass ($M_{\rm max}$) for the dark matter halos hosting LAEs without predicting a Ly$\alpha$ luminosity. This approach bypasses all the physical uncertainties associated to star formation and radiative transfer. We use the Markov Chain Monte Carlo technique to obtain the likelihood of the parameters given the observational constraints.

Throughout this letter we assume a $\Lambda$CDM cosmology

⋆ jemejia@das.uchile.cl
† je.forero@uniandes.edu.co

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

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, the observations we take as a benchmark. Second, the N-body simulation and the halo catalogs we use to build the mocks. Third, the parameters describing our model to assign a LAE to a halo. Fourth, the statistical method we adopt 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 narrow band imaging to detect 643 LAE candidates at $z \sim 3$ with equivalent widths of $\gtrsim$65Åand a flux limit of $2 \times 10^17$erg/cm$^2$/s ($L \sim 7 \times 10^{42}$erg/s). Using spectroscopy they found a 22% contamination fraction. Their observations cover 5 (out of 9) independent and cospatial fields of the VLT LBG Redshift Survey (VLRS). The total observed area corresponds to 1.07deg$^2$ that translates to $\sim 80^2 h^{-1}$Mpc$^2$ in a comoving scale. Bielby et al. (2016) used the NB497 narrow-band filter whose 77ÅFWHM and 154ÅFWTM correspond to a total observational depth of $44h^{-1}$Mpcand $82h^{-1}$Mpccomoving, respectively.

### 2.2 Simulation and halo catalog

We use results from the ●Bolshoi simulation (Klypin et al. 2011)(●REFS)which was performed in a cubic volume of 250 $h^{-1}$ Mpc comoving on a side. The dark matter distribution was sampled using ●$2048^3$ particles. The cosmological parameters are consistent with a Wilkinson Microwave Anisotropy Probe (WMAP) ninth year data ●PLANK? ((●REFS)) with a matter density $\Omega_\mathrm{m} = 0.307$, cosmological constant $\Omega_\Lambda = 0.693$, dimensionless Hubble constant $h = 0.70$, slope of the power spectrum ●$n = 0.95$ and normalization of the power spectrum ●$\sigma_8 = 0.82$ ●. This translates into a particle mass of ●$m_\mathrm{p} = 1.54 \times 10^8\ h^{-1}$ M$_\odot$.

We use halo catalogs constructed with a Bound-Density-Maxima (BDM) algorithm. The catalogs were obtained from the publicly available Multidark database ●[1] (Riebe et al. 2013). For each halo in the box we extract its comoving position and mass. We focus our work on halos more massive than $1.54 \times 10^9 h^{-1}$M$_\odot$resolved with at least 10 particles.

We divide the total volume of the snapshot of the simulation at z$\sim$3 into 27 smaller mock volumes mimicking the comoving area and depth reported in Bielby et al. (2016) and described in §2.1. This allows us to take explicitly into account the effects of cosmic variance.

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

### 2.3 A model to populate halos with LAEs

We build a model to assign LAEs to a DM halo. We are not concerned on the exact LAE luminosity, we only care about a yes/no answer to the following question. Does this halo host a LAE?

We first assume that a dark matter 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 a LAE with probability $f_\mathrm{occ}$ if and only if the halo mass is in the range $M_\mathrm{min} < M_\mathrm{h} < M_\mathrm{max}$. The probability $f_\mathrm{occ}$ can be thought as the occupation fraction of halos which can be automatically set as the ration of the observed number of LAEs to the number of halos within the considered mass range, that is $f_\mathrm{occ} \equiv N_\mathrm{LAE}/N_\mathrm{halos}$.

The occupation fraction varies along the mock fields tracing the cosmic variance changing in a factor $\sim 2$ ($\sim 0.3$, see left panel in Fig. 3). The interpretation of the occupation fraction $f_\mathrm{occ}$ involves two phenomena: the actual presence of a star forming galaxy in a halo and its detectability as a LAE.

We do not explore any physical model to disentangle these two effects. This means that our model does not assign a luminosity or escape fraction to each LAE. We avoid this in order to maintain theoretical uncertainties to a minimum. This flexibility allows us to explore a wide range of possible masses for the host halos without any strong theoretical prejudice regarding the details of star formation and Ly$\alpha$ escape fraction.

We also artificially contaminate our mock catalogs with 22% randomly distributed data points to mimic the fraction of interlopers in observations Bielby et al. (2016) . On top of that we apply rejection sampling to our LAE selection taking the transmission function of the NB479 filter used in their observations as a radial selection factor.

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

### 2.4 Exploring and selecting consistent models

We use the angular correlation function to compare the observations with our mock catalogs. This is done trough a thorough exploration of the parameter space of the models $\mathcal{M}$ by meansof a Monte Carlo Markov Chain minimization using the EMCEE python package (Foreman-Mackey et al. 2013). We put a flat prior on $\log M_\mathrm{min}$ and $\log M_\mathrm{max}$ to vary between 9.2 up to 13.4, corresponding to the halo mass range of the simulation at $z = 3$. Given that the typical scatter in $N_\mathrm{halos}$ is about 0.3 dex (i.e. about a factor of 2), our parameter space is restricted to models where the median number of dark matter halos is larger than $N_\mathrm{LAE}/3$. The MCMC exploration is done using a total of 24 seeds and 400 iterations (9600 models) to sample the posterior PDF, $P(M_\mathrm{min}, M_\mathrm{max}, f_\mathrm{occ}|\mathcal{M}) \propto \exp(-\chi^2_\mathcal{M}/2)$, with

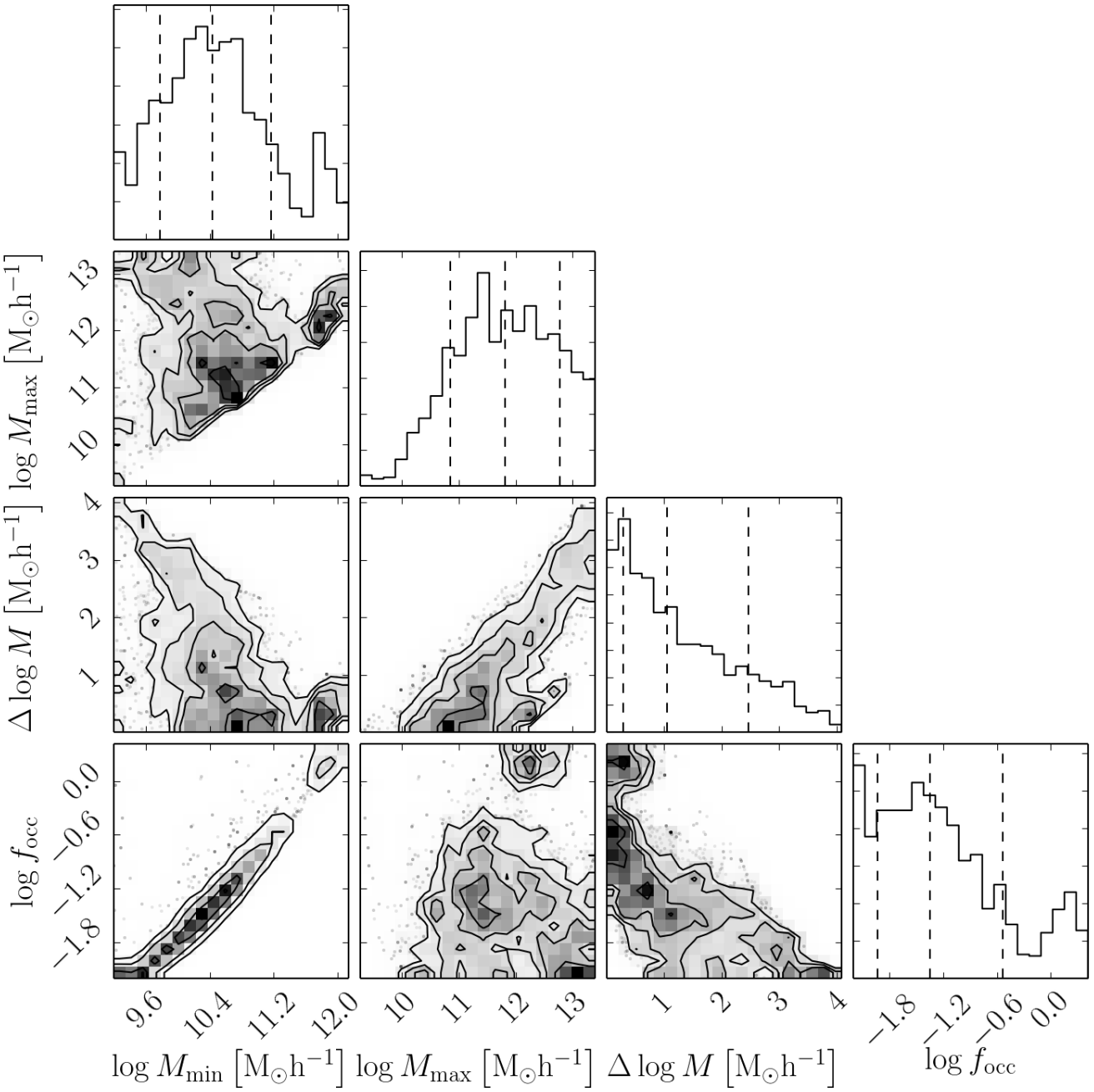


**Figure 1.** Likelihood

$$\chi^2_{\mathcal{M}} = \sum_{\theta} \left[ \frac{\left(\mathrm{ACF}_{\mathcal{M}}(\theta) - \mathrm{ACF}_{\rm obs}(\theta)\right)^2}{\sigma^2_{\mathcal{M}}(\theta) + \sigma^2_{\rm obs}(\theta)} \right] \qquad (1)$$

where $\mathrm{ACF}_{\mathcal{M}}$ and $\mathrm{ACF}_{\rm obs}$ 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 1-$\sigma$ scatter of the $\mathrm{ACF}_{\mathcal{M}}$ as a product of cosmic variance and $\sigma_{\rm obs}$ is the observational error associated to $\mathrm{ACF}_{\rm obs}$. The $\mathrm{ACF}_{\mathcal{M}}$s are computed using the Landy & Szalay estimator (Landy & Szalay 1993).

## 3 RESULTS AND DISCUSSION

### 3.1 Constraining $M_{\rm min}$, $M_{\rm max}$and $f_{\rm occ}$in DHM hosting LAEs

In Fig. 1 the posterior probability distributions in the $M_{\rm min}$-$M_{\rm max}$, $M_{\rm min}$-$f_{\rm occ}$, $M_{\rm min}$-$\Delta M$, $M_{\rm max}$-$f_{\rm occ}$, $M_{\rm max}$-$\Delta M$and $f_{\rm occ}$-$\Delta M$ parameter spaces. We find that the $M_{\rm min}$and $M_{\rm max}$parameter space cannot be well constrained mainly because of cosmic variance and the large observational Poissonnian uncertainty in the ACF. We however found that 10.0< log $M_{\rm min}$ <11.1 and 11.0< log $M_{\rm max}$ <12.8. We interestingly find that $f_{\rm occ}$ is basically completely determine by $M_{\rm min}$going from $f_{\rm occ}$=0.015 when log $M_{\rm min}$ =10.0 to

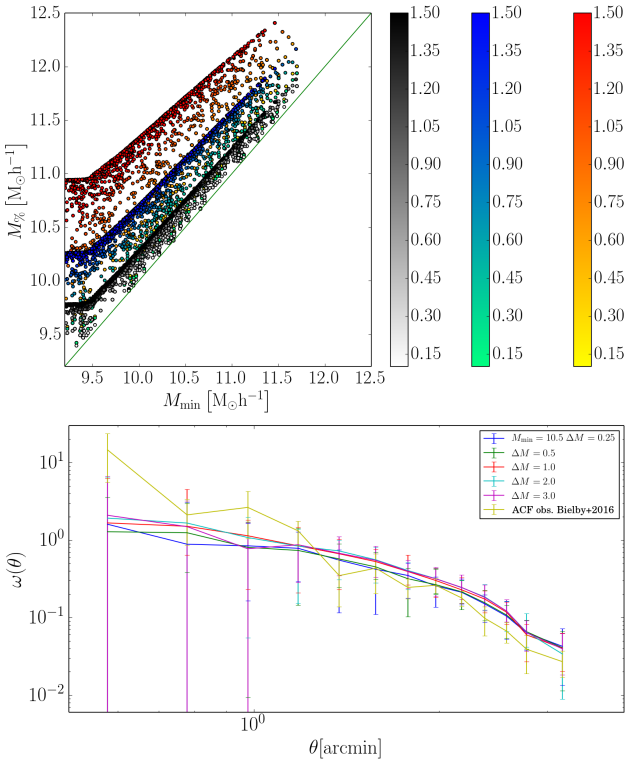


**Figure 2.** Left: $\log M_{\rm min}$ vs $\log M_{50\%}$ (black), $\log M_{16\%}$-$\log M_{84\%}$ (blue) and $\log M_{2.5\%}$-$\log M_{97.5\%}$(red) for different values of $\log M_{\rm max}$. The points are color coded by $\Delta \log M$ [$\rm M_\odot h^{-1}$]]. THe green line accounts for the 1:1 relation. Right: Angular Correlation functions for $\log M_{\rm min}$[$\rm M_\odot h^{-1}$] = 10.5 and different values of $\Delta M$. The error bars represent the 1 sigma deviation due to cosmic variance.

$f_{\rm occ}$=0.43 when $\log M_{\rm min}$ =11.1. We remark that those models where $\log f_{\rm occ} > 0.0$ ($f_{\rm occ} > 1$) correspond to models where the number density of halos is smaller than the number of observed LAEs but are still consistent because of the uncertainty in the median number of LAEs in the universe due to cosmic variance.

### 3.2 Halo mass distribution within models

In the top panel of Fig. 2 we show the 50 ($\log M_{50}$, black dots), the 84 ($\log M_{84}$,blue diamonds) and 95 ($\log M_{95}$, red squares) percentiles of the LAE halo mass as a function of $\log M_{\rm min}$ for each of the models that we run in our MCMC simulation. The points are color coded according to their $\Delta \log M$ associated value. We can see that the median mass, $M_{50}$, and $M_{84}$ are not very sensitive to $\log M_{\rm max} = \log M_{\rm min} + \Delta \log M$ specially when $\Delta \log M \gtrsim$ 1.0dex. We particularly found that $\log M_{\rm min} \lesssim \log M_{50} \lesssim \log M_{\rm min} + 0.2$dex and that $\log M_{\rm min} + 0.1$dex $\lesssim \log \rm M_{50} \lesssim \log M_{\rm min} + 0.5$dex regardless of the value of $\Delta \log M$. The latter is a consequence of the very steepen distribution of the dark matter mass function and is at the same time the reason for the almost perfect one to one relation between $M_{\rm min}$ and $f_{\rm occ}$. However, it can also be seen in Fig. 2 that $\log M_{84}$ is very sensitive to $\log M_{\rm max}$($\log M_{\rm min} + 0.2$dex $\lesssim \log \rm M_{50} \lesssim \log M_{\rm min} + 1.5$dex). Thereby, any differences in the clustering strength of models sharing the same $M_{\rm min}$but different $M_{\rm max}$should be mainly driven by the $\sim 16\%$ most massive halos of each model $\mathcal{M}$.

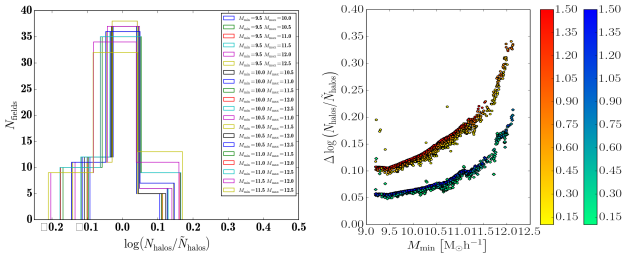


**Figure 3.** Left: Halo number distribution over the 54 mock field of the simulation for different models of $M_{\rm min}$-$M_{\rm max}$.Right:Width of the halo number distribution over the 54 mock field of the simulation ($\Delta \log\left(N_{\rm halos}/\tilde{N}_{\rm halos}\right)$) of the central 68 (blue diamonds) and 95(red circles) percentiles vs $M_{\rm min}$. Points are color coded according to their associated value of $\Delta M \equiv M_{\rm min} - M_{\rm min}$.

In order to estimate the effect of most massive halos in $ACF_{\mathcal{M}}$ in the bottom panel of Fig. 2 we show the computed $ACF_{\mathcal{M}}$ of models with $\log M_{\rm min} = 0.5$ and different values of $\Delta \log M$. We can see that the clustering gets 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 explain the current difficulty to put tighter constrains in $\log M_{\rm max}$ in our model.

### 3.3 Constraining DMH mass with cosmic variance

In the left panel of Fig. 3 we show the number halo distribution (NHD) in the mock fields of the simulation for different models $\mathcal{M}$. By simple inspection one can infer an increase in the distribution with as $\log M_{\rm min}$increase. This trend is confirmed in the right panel of Fig. 3 where we plot the central 1-$\sigma$ (blue diamonds) and 2-$\sigma$ (red dots) widths of the NDH as a function of $\log M_{\rm min}$. We particularly found that when we consider survey fields of $\sim 1\rm deg^2$, the central 1-$\sigma$ (2-$\sigma$) width of the NHD increases monotonically from 0.05dex (0.10dex) when $\log M_{\rm min} = 9.5$ to 0.20dex (0.35dex) when $\log M_{\rm min} = 12.0$. The latter result opens the possibility to constrain the $\log M_{\rm min}$(as well as the median mass) of LAEs by simply measure the width of the distribution of observed LAE along several observational fields mapping $\sim 1\rm deg^2$ in area and the observational depth determined by the NB479 filter.

## REFERENCES

Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57

Bielby R. M., Tummuangpak P., Shanks T., Francke H., Crighton N. H. M., Bañados E., González-López J., Infante L., Orsi A., 2016, MNRAS, 456, 4061

Bond N. A., Gawiser E., Gronwall C., Ciardullo R., Altmann M., Schawinski K., 2009, ApJ, 705, 639

Colberg J. M., White S. D. M., Yoshida N., MacFarland T. J., Jenkins A., Frenk C. S., Pearce F. R., Evrard A. E., Couchman H. M. P., Efstathiou G., Peacock J. A., Thomas P. A., Virgo Consortium 2000, MNRAS, 319, 209

Dayal P., Ferrara A., Saro A., Salvaterra R., Borgani S., Tornatore L., 2009, MNRAS, 400, 2000
Dijkstra M., Kramer R., 2012, MNRAS, 424, 1672
Dijkstra M., Mesinger A., Wyithe J. S. B., 2011, MNRAS, 414, 2139
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Forero-Romero J. E., Yepes G., Gottlöber S., Knollmann S. R., Cuesta A. J., Prada F., 2011, MNRAS, 415, 3666
Forero-Romero J. E., Yepes G., Gottlöber S., Prada F., 2012, MNRAS, 419, 952
Garel T., Blaizot J., Guiderdoni B., Schaerer D., Verhamme A., Hayes M., 2012, MNRAS, 422, 310
Gawiser E., Francke H., Lai K., Schawinski K., Gronwall C., Ciardullo R., Quadri R., Orsi A., Barrientos L. F., Blanc G. A., Fazio G., Feldmeier J. J., Huang J.-s., Infante L., Lira P., Padilla N., 2007, ApJ, 671, 278
Guaita L., Francke H., Gawiser E., Bauer F. E., Hayes M., Östlin G., Padilla N., 2013, A&A, 551, A93
Hayashino T., Matsuda Y., Tamura H., Yamauchi R., Yamada T., Ajiki M., Fujita S. S., Murayama T., Nagao T., Ohta K., Okamura S., Ouchi M., Shimasaku K., Shioya Y., Taniguchi Y., 2004, AJ, 128, 2073
Jose C., Srianand R., Subramanian K., 2013, ArXiv e-prints
Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ, 740, 102
Koehler R. S., Schuecker P., Gebhardt K., 2007, A&A, 462, 7
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., Fynbo J. P. U., Michałowski M. J., Watson D., Ledoux C., Rosati P., Pedersen K., Grove L. F., 2007, A&A, 471, 71
Orsi A., Lacey C. G., Baugh C. M., 2012, MNRAS, 425, 87
Ouchi M., Shimasaku K., Furusawa H., Saito T., Yoshida M., Akiyama M., Ono Y., Yamada T., Ota K., Kashikawa N., Iye M., Kodama T., Okamura S., Simpson C., Yoshida M., 2010, ApJ, 723, 869
Peebles P. J. E., 1980, The large-scale structure of the universe
Riebe K., Partl A. M., Enke H., Forero-Romero J., Gottlber S., Klypin A., Lemson G., Prada F., Primack J. R., Steinmetz M., Turchaninov V., 2013, Astronomische Nachrichten, 334, 691
Springel V., White S. D. M., Jenkins A., Frenk C. S., Yoshida N., Gao L., Navarro J., Thacker R., Croton D., Helly J., Peacock J. A., Cole S., Thomas P., Couchman H., Evrard A., Colberg J., Pearce F., 2005, Nature, 435, 629
Verhamme A., Schaerer D., Maselli A., 2006, A&A, 460, 397
Yajima H., Choi J.-H., Nagamine K., 2012, MNRAS, 427, 2889