Dear Editor,

We wish to thank the referee for a detailed report on our paper.

The referee expressed 5 major points of concern. Some of them were due to unclearly written sections, other made reference to issues with the observational data and choices to perform the comparison against the simulated data.

In this new version we respond to all these concerns. We have clarified the text where it was needed and performed new tests on the data. Therefore we resubmit this draft to be considered for publication in the MNRAS.

We wish to emphasize that our paper uses a new method in the context of the study of LAE galaxies. Correspondingly, the results are novel with respect to the already published literature.

In what follows we reply to each one of the major points raised by the referee. The comments by the referee are in boldface. Our reply follows to each of these comments.

Best regards,

Jaime E. Forero-Romero and Julian Mejia-Restrepo

Major points

1. **As the authors are aware, each observational survey for LAEs has different selection effects, and some use very different thresholds in equivalent width. Hence you need to use the number density and autocorrelation function (ACF) of LAEs from the *same* survey for an investigation like this. You could try to get Yamada's LAE positions to calculate the ACF, since you appear to prefer that survey for number density estimation. Or you could pursue joint analysis of number density and ACF for smaller surveys e.g., Ouchi's.**

   We agree with the referee that the two statistical tests should be performed on the same data set. Unfortunately neither the galaxy catalog nor the ACF from Yamada et al. 2012 are available. We contacted Dr. Yamada asking for information about the ACF, but we didn't get any reply.

   The currrent version of the draft features a comparison with the ACF from a survey with an average number density from Hayashino et al 2004.

   On the other hand, Using a small field as Ouchi's for our analysis is not feasible, given that our method is based on the analysis of the surface density *distribution among different* fields. Having the distribution data is central to discard a large fraction of models (Section 3.3).

2. **The major result claimed that all models have a narrow mass range of halos hosting LAEs is guaranteed by the choice of a minimum value of $f_{\rm occ}$ of 0.1. This should be obvious to the authors, as there is a degeneracy between $f_{\rm occ}$ and the mass range of halos that you have hidden by forcing $f_{\rm occ}$ to values larger than those previously claimed in the literature.**

   As the referee highlights, we have found a new set of acceptable parameters for the escape fraction than those that have been previously claimed by the literature.

   However, we do not find a clear evidence for a degeneracy between the occupation fraction $f_{\rm occ}$ and the mass range ΔM. There is, however, a clear degeneracy between the minium halo mass $M_{\rm min}$ and the occupation fraction.

   The reason for that trend is easy to understand. In LCDM the halo number density n is verysensitive to the mass. The dependency is roughly $n \propto M_h^{-1}$. Therefore, at a fixed minimum halo mass, increasing the mass range by 1.0 dex barely raises the number density by 0.10. Increasing the mass range by 2.0 dex raises the number density by 0.11. This explains that we cannot find any clear dependence between ΔM and the $f_{\rm occ}$. On the other hand, decreasing the minium mass by 1.0 dex would increas the number density by a factor of 10! Requiring that models with small values for the minium mass to have a small occupation fraction.

   This trend is clearly seen on the right panel of Figure 3,4 and 6. This implies that exploring very low escape fractions values of 0.01 would require the exploration of halo masses aroud the range of 10^9 Msun or below, which is lower than the minimum halo masss of $10^{9.5}$ Msun that is confidently resolved in the simulation.

3. **The authors cite similar work by Walker Soler et al. 2012 but fail to mention in the introduction how the current paper has the same goals as that earlier work. If applied more carefully, their approach should be able to improve on the earlier work, but it is important to be clear about how similar the papers are and how the current one hopes to improve on the previous one.**

   The work of Walker-Soler 2012 actually does not pursue a similar goal as our paper. They take a fixed set of results of a semi-analytical model to infer the mass of the *descendants* from LAEs at $z = 3.1$. We have removed the reference to that work that might show that we are improving work they have been done, because that's not the case.

4. **Comparing the most heavily clustered of 15 mocks against SSA22 is unfair. SSA22 is a 5 sigma overdensity, which is much rarer than 1/15.**

We agree with the referee that a comparison of a ~ 5 overdensity with an average field would not be fair. Therefore we have removed that constraint from the results in the new version of the paper. Now we only include the ACF comparison against the results of an average density field.

Related to this point we have added the left panel in Figure 1 and subsection 2.2 to discuss the adequacy of including SSA22 as an observational benchmark in the number density tests.

5. **There seems to be significant confusion between halo occupation fraction and Lyman alpha escape fraction. Only in an incredibly simplified model where 100% of Ly alpha photons escape from LAEs and 0% escape from non-LAEs would the comparison in section 4.1 make sense. In a more realistic model, it is not clear how this investigation could hope to comment on the escape fraction of Ly alpha photons.**

   We agree with the referee in this point.

   We have corrected all the mistakes where we meant occupation fraction but wrote escape fraction. We have alsoremoved the discussion comparing our results to escape fraction estimates, in spite of the effort of other authors to link these two different quantities.

Additional suggestions:

- **Only allowing one LAE per dark matter halo is a simplistic model. You could try more complicated HOD models where $f_occ(M)$ is not constant to see if they make a significant difference in the results. But arguing against this by citing the Jose et al. 2013 paper seems reasonable. You could also note the lack of close-proximity LAE pairs noted in papers led by Nick Bond as further evidence against multiple LAEs occurring in the same halo.**

  We would like to add a reference to the work of Nick Bond, but failed to fine what exact paper makes reference to the lack of close-proximity LAE pairs.

- **You could explore a wide range of f_occ down to properly tiny values such as 0.001 without increasing the run time by indexing it logarithmically.**

  As we mentioned before in the Reply to Major Point 2, tiny values of 0.001 can only produce reasonable models for very low mass halos. This is due to the contraint on the number density. An occupation fraction of 0.001 requires models with a minimum mass on the order of $M_{\min} = 10^9$ solar masses, which is well below the mass resolution of the simulation we use. This comment has been added to the paper.

- **Be careful when turning filters into cosmological volumes - you should not simply use the FWHM. See Gronwall et al. 2007 for a detailed discussion of this.**

- **Be careful when comparing results for halo occupation fraction for LAEs from a survey with a 190A observed EW threshold to the more typical 20 A EW rest-frame, which equals 80A observed at z=3.**

- **SSA22 is known to be an extraordinary overdensity, as you note. The overdense region, and possibly the entire field, will bias the measured number density of z=3.1 LAEs upwards. So one should be very cautious including this field in a study like this one.**

  We have included subsection 2.2 to comment on this point.

- **When using a constant value for $f_{\rm occ}$, you could use just $M_{\rm min}$, $M_{\rm max}$ as parameters, choose the value of $f_{\rm occ}$ that fits the observed number density (propagating uncertainties) and then test the resulting ACF versus data. This would be accurate because a random sub-sample of halos has the same clustering as the full sample.**

  We agree that this is a feasible approach. The emphasis we meant in our approach was the effect of cosmic variance on the size of observed fields. This is why we include the tests on all sub-samples and not just one.