

Peer Review

Review of: "Probabilistic Galaxy Field Generation with Diffusion Models"

Gabriele Parimbelli¹

1. Institute of Space Sciences, Barcelona, Spain

The paper discussed an innovative method to investigate the galaxy-halo connection through machine learning instead of relying on approximate schemes such as HOD or semi-analytical models. The results are promising; overall, the paper is well written, and it was a pleasure to read.

I have these main comments:

1. I believe most of the abstract can be moved to the introduction in order to save some space to report some of the significant results later exposed in the main text.
2. Among all the possible approximate schemes for assessing the galaxy-halo connection, HOD is the only one that is mentioned here. While I agree that it is a good benchmark to be used, the authors should at least mention in the text other possible methods, i.e., SHAM and semi-analytical models.
3. I have some doubts about the section describing the implementation of the HOD. The authors first say that satellites are randomly distributed around the halo center. This is way too simplistic, as we know that satellite galaxies do not follow a random profile. I expect this assumption to have a non-negligible impact on the one-halo term, and in fact, the differences with respect to the target model arise for $k > 1 \text{ h/Mpc}$ (fig. 3). To be clear, I understand that we want to “move away” from the HOD, and I strongly agree, but this seems to me a little bit “unfair” in respect to the HOD model. Could the authors explain and motivate their choices?
4. Moreover, the authors mention that they tune the 3 HOD parameters in order to minimize the MSE of the power spectrum. I believe also the total number density of the HOD should be the same as that of the target. Do the authors take that into account?
5. In the bottom left panel of fig. 3, I would use a linear y-axis; I believe the units of the $P(k)$ should be $(\text{Mpc}/h)^3$.

6. I am not sure why the amplitudes of the cross spectra should be different (up to an order of magnitude), given that the auto power spectra have the same amplitude. Could the authors clarify?

Declarations

Potential competing interests: No potential competing interests to declare.