Qeios

Peer Review

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

Luz Ángela García Peñaloza[1]

1. Escuela Colombiana de Carreras Industriales, Colombia

The document “Probabilistic Galaxy Field Generation with Diffusion Models” presents an alternative to the traditional (and computationally expensive) method of HOD through two machine learning methods trained with Illustris TNG in the CAMELS dataset. The approach presented is quite insightful and enormously reduces the computing time taken to run any hydrodynamical simulations (such as Illustris TNG).

I have these main comments:

- The architecture of both the CNN and VDM is barely discussed throughout the paper. I would suggest including a diagram showing the final (detailed) architecture used to train the models, to guide the reader and ensure that the algorithms are reproduced in a similar fashion in future works. Taking into account that this is the main advantage of the manuscript, I consider it important to explain in extended detail the steps of the applied models.
- Also, it would be quite insightful to discuss why these models outperformed the HOD, besides the improvement in the errors of the power spectra shown in Table 1 and Figure 3. Is it a result of the large set of simulations used to train their models? Or instead, what feature of the HOD currently used makes its performance significantly decrease for large k's? Why aren't both the CNN and VDM affected at these scales? I think this is key for the community that wants to enhance their simulated datasets with the advent of large galaxy surveys.
- Could you provide the reader with the values (or ranges) of the hyper-parameters used to recover your best fit models? This information is missing in this version of the paper, and it is quite relevant for astronomers working in the field.
- At no point in the paper is there a discussion on the validity of these results with redshift. With the release of Euclid, DESI, and other surveys that are pushing the boundaries of the redshifts

 doi.org/10.32388/NAQZWE

measured for galaxies, as well as their census, it is imperative to know if these results extend to higher redshifts. The authors start section 2.2 mentioning that the models were tested at z=0, but thereafter, there is no comment on the redshift evolution. Please clarify this point.

Declarations

Potential competing interests: No potential competing interests to declare.