27 January 2025 · CC-BY 4.0 Qeios

Peer Review

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

Andrea DeMarco[1]

1. Institute of Space Sciences and Astronomy, University of Malta, Malta

The manuscript introduces a novel approach combining Convolutional Neural Networks (CNNs) and Variational Diffusion Models (VDMs) to generate high-fidelity galaxy fields from dark matter simulations. This innovative method attempts to address the limitations of traditional methods like the Halo Occupation Distribution (HOD) and CNN-based frameworks by introducing a probabilistic, uncertainty-aware model. While the approach is promising, a thorough assessment highlights strengths, limitations, and areas for improvement.

Novelty: The use of VDMs for galaxy field generation is innovative, extending diffusion models from computer vision to astrophysics. The probabilistic framework effectively captures the stochastic nature of galaxy formation, surpassing traditional deterministic models like HOD and CNNs.

Performance: The results show significant improvements over benchmarks: a 60% reduction in mean squared error (MSE) compared to HOD; a 20% improvement over CNNs, particularly at capturing intermediate and small-scale galaxy clustering patterns. The inclusion of power spectrum and cross-power spectrum analyses provides strong evidence of the model's accuracy.

If I had to comment on possible areas for improvement for this submission:

1. Model generalization: how does the model generalize to data outside the CAMELS dataset, especially simulations with significantly different astrophysical models or observational noise?
2. Parameter selection: a sensitivity analysis on the impact of VDM hyperparameters (e.g., noise schedule or UNet depth) would strengthen the methodology section.
3. Cosmological constraints: the paper mentions implications for constraining cosmological parameters but provides no quantitative results. Including examples of parameter inference based on VDM outputs would substantiate these claims.

 doi.org/10.32388/TK5QN2

4. Improve figures, particularly Figure 2, which could benefit from improved annotation and resolution to make differences between models more discernible.
5. Reproducibility: the authors reference their use of GPUs; they should include implementation details (e.g., training configurations, computational requirements) to enhance reproducibility.

Declarations

Potential competing interests: No potential competing interests to declare.