Qeios

Peer Review

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

Karthik Prabhu[1]

1. Physics and Astronomy, University of California, Davis, United States

Dear Authors,

I have reviewed the manuscript titled "Probabilistic Galaxy Field Generation with Diffusion Models" and am impressed by your innovative approach in addressing the challenges of galaxy formation modeling. The combination of Convolutional Neural Networks (CNNs) and Variational Diffusion Models (VDMs) is a significant contribution to the field, offering improved accuracy over traditional methods like HOD while enhancing computational efficiency. The results demonstrate promising improvements in capturing galaxy clustering patterns at various scales.

However, there are a few points that could benefit from further clarification or extension:

1. **Model Architecture:** Providing detailed diagrams or descriptions of both CNN and VDM architectures would enhance understanding and facilitate reproducibility. Including values for hyperparameters used in training best-fit models would enhance reproducibility for future research applications.

2. **Performance Comparison:** Including a discussion on why the VDM model outperforms HOD would strengthen the argument, as would discussing specific features or properties that are being addressed.
    1. The VDM exhibits a systematic overestimation of galaxy density at large scales (voids) and underestimation at small scales (halos), as seen in the Z-score map (Fig. 2) and power spectrum residuals (Fig. 3). What is the loss function being used in the VDM? Is it MSE loss? Have the authors explored alternative loss functions (e.g., perceptual loss, multi-scale

 doi.org/10.32388/14J7VV

SSIM) to better capture nonlinear clustering?

2. The overcorrelation in cross-power spectra (Fig. 3) is attributed to "modeling galaxy-halo connections" but lacks mechanistic insight. Is this due to missing baryonic physics in the training data or the model architecture? Further analysis is needed.
3. **Redshift Evolution:** It would be interesting to check if these results extend to higher redshifts.

Overall, this work presents a significant advancement in galaxy catalog generation, leveraging diffusion models to achieve state-of-the-art accuracy and uncertainty quantification. By addressing the above concerns, you can improve the clarity, quality, and applicability of your work. The paper is technically sound and merits publication post-revision, with the potential to influence both cosmological simulations and machine learning applications in astrophysics.

Declarations

Potential competing interests: No potential competing interests to declare.