MAKE NEW STANDARDS.
東海国立大学機構

NAGOYA UNIVERSITY

The 10th East Asian Numerical Astrophysics Meeting @Jeju, South Korea
2025/9/15

Environmental Dependence of Galaxy Properties: Machine Learning Analysis of IllustrisTNG300

Based on Uchida et al. (2025)
arXiv:2505.06815

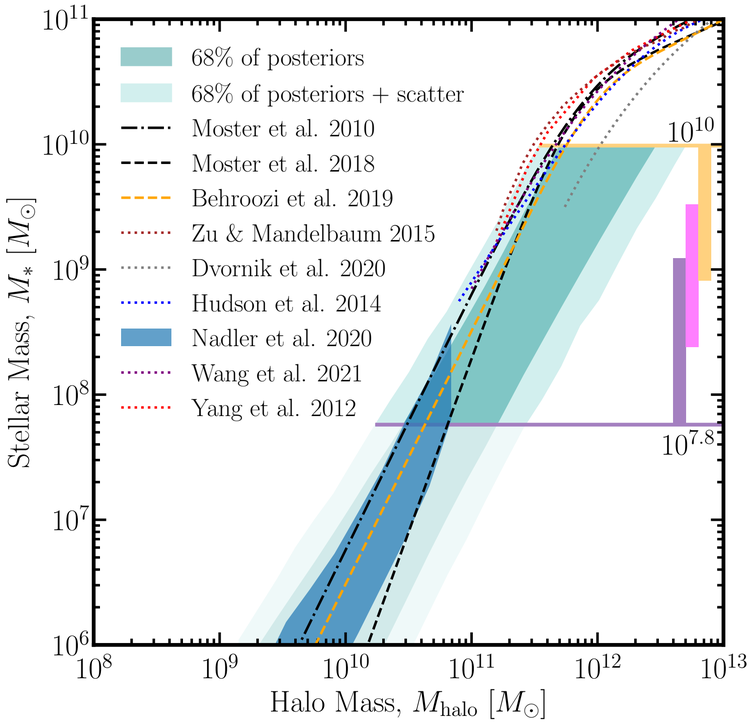
MNRAS accepted!

Introduction: galaxy-halo connection

Galaxies form and evolve within the dark matter halo.

Properties of **DM halo** (subhalo) are related to hosted galaxy properties.

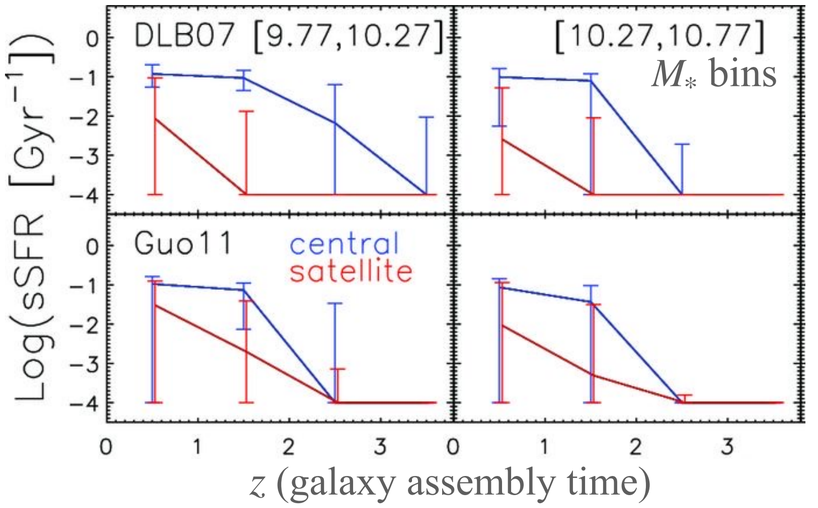
(Wechsler et al. 2018, ARA&A, 56. 435-487)



(Thornton et al. 2024, MNRAS, 535. 1-20)

- The galaxy properties depend on $M_{\rm halo}$ (stellar-to-halo mass relation)
  - ➡ e.g. HOD method, Abundance Matching method

- Assembly Bias (secondary halo bias):
  Galaxy properties depend not only on their mass but also on their assembly history.
  - ➡ e.g. Low-M_* galaxies with lower sSFR tend to assemble earlier.

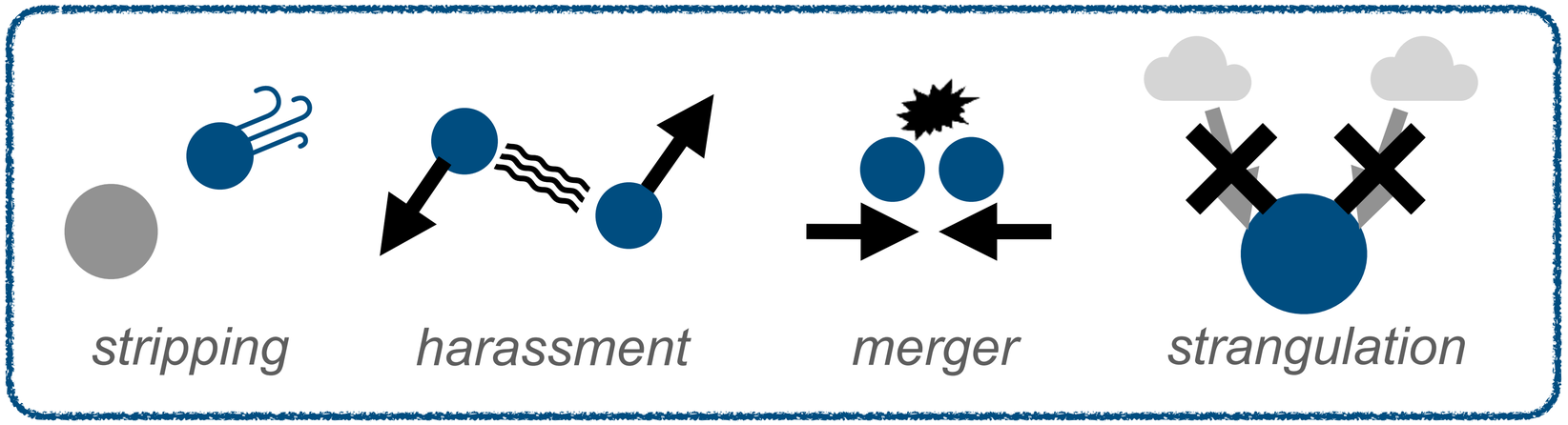


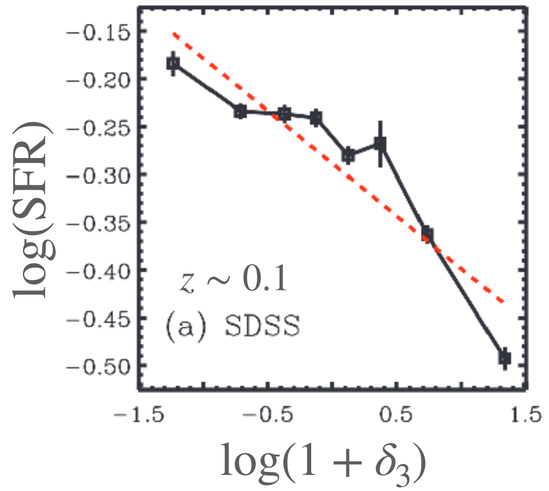
(Wang et al. 2013, MNRAS, 433, 512-520)

Introduction: environmental dependence of galaxy properties

- The galaxy-halo connection can also be considered through a **galaxy's environment**.
  The environment surrounding a galaxy influences its properties and evolution.

  ▸ On the local environment scales (within a central galaxy's virial radius),
  galaxies are affected by their environment, which modifies their gas content.





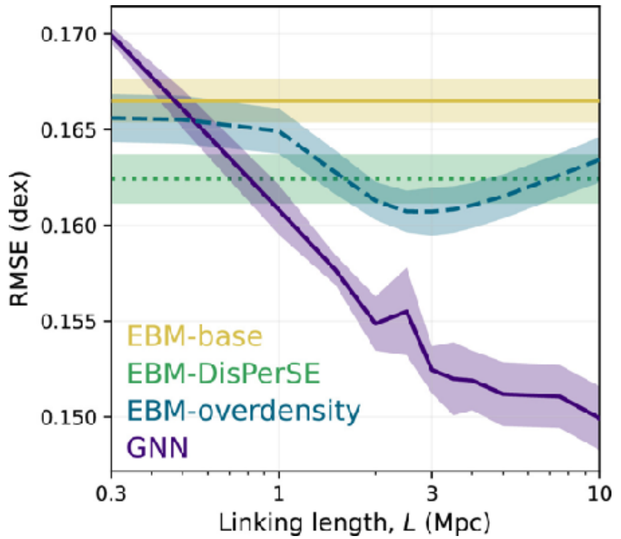
(Cooper et al. 2008, MNRAS, 383, 1058-1078)

  ▸ Environmental effects are also seen **between galaxies belonging to different host halos** (two-halo conformity)

  ➡ The star-formation indices correlate at 2 ~ 4 Mpc scales ($z < 0.03$). (Kauffmann et al. 2013)

Introduction: recent work & our motivation

- Various approaches to define environmental parameters
  e.g., fixed-radius overdensities, Nth-nearest neighbors, topological feature
- Machine learning analysis suggests **the optimal scale for quantifying the environment is 3 Mpc**, with some additional information up to 10 Mpc (Wu et al. 2024)
- However, quantifying environmental influence on different galaxy properties as a function of galaxy types remains challenging



(Wu et al. 2024, ApJ 976 37)

How much do galaxy properties get influenced by their environment?

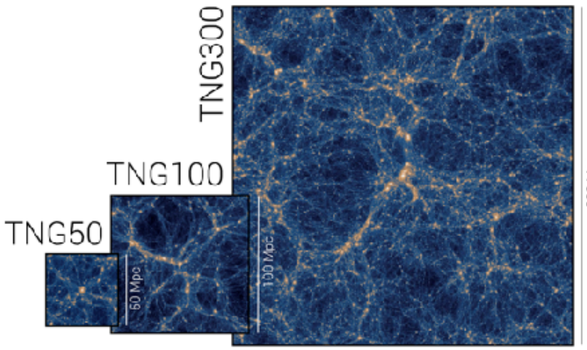
How does this influence vary across different galaxy types?

Which environmental parameters most strongly influence different galaxy populations?

Methodology & Data

- We constructed an interpretable neural network framework to characterize the surrounding environment of galaxies

  ➡predict M_* and SFR of the simulated galaxy from **the own and surrounding subhalos properties**



(Nelson et al. 2021)

- Our framework is trained on IllustrisTNG300-1 simulation result (subhalo catalog $z = 0$)

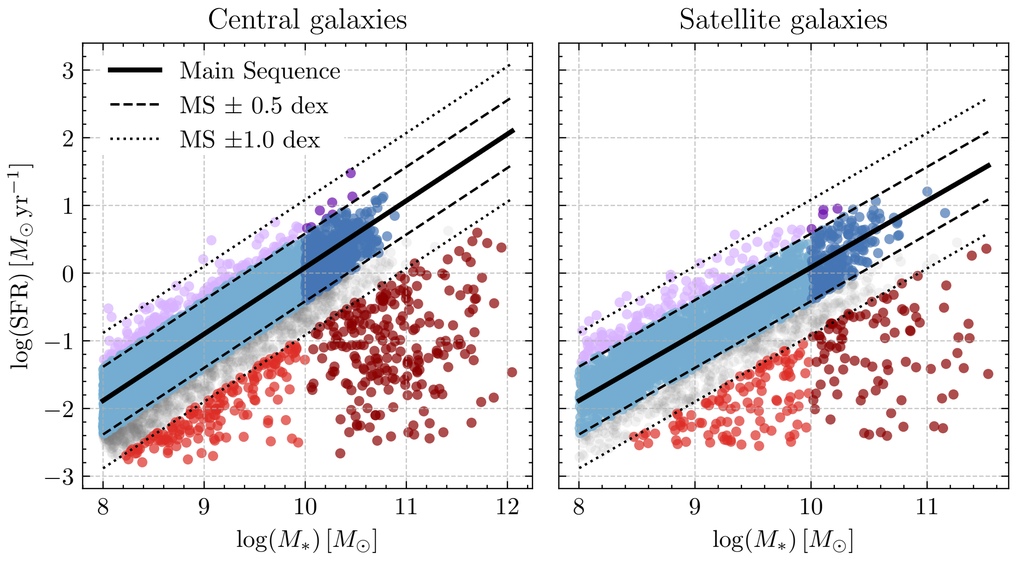
  - Galaxies meeting the following criteria were used

$$\log(\text{halo mass}) > 10 \ M_\odot$$

$$\log(\text{stellar mass}) > 8 \ M_\odot$$

$$\text{SFR} > 0 \ M_\odot/\text{yr}$$

  - We classified the test data (10,000 galaxies) into 6 groups separately for centrals and satellites, according to M_* and SFR



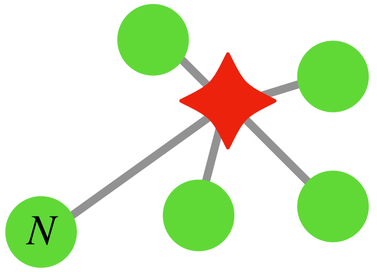
Methodology: neural network architecture

Model input features

- **Target galaxy's subhalo properties;**
  (x, y, z), (V_x, V_y, V_z), $M_{\rm halo}$, r_{50} (half $M_{\rm halo}$ radius),
  $V_{\rm max}$ (maximum circular velocity)

+

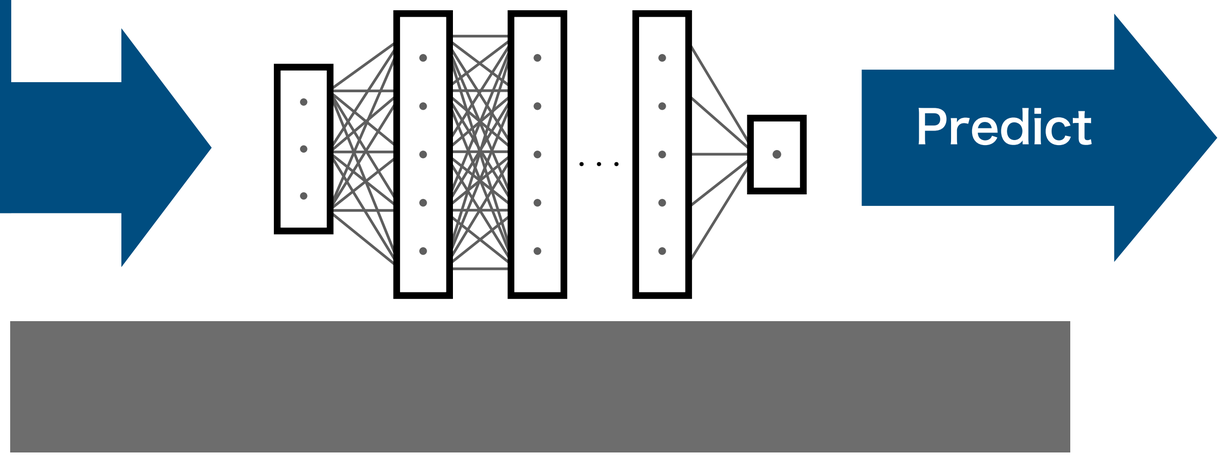
- **(Neighboring galaxy's subhalo properties) × *N*;**
  (D, θ, ϕ), (V_D, V_θ, V_ϕ), $M_{\rm halo}$, r_{50}, $V_{\rm max}$



✦ Target galaxy

● Surrounding galaxies:
N nearest neighbors ($N = 0 \sim 10, 15, 20, 25, 30$)

Training data:
300,000 galaxies



Predict

Model output

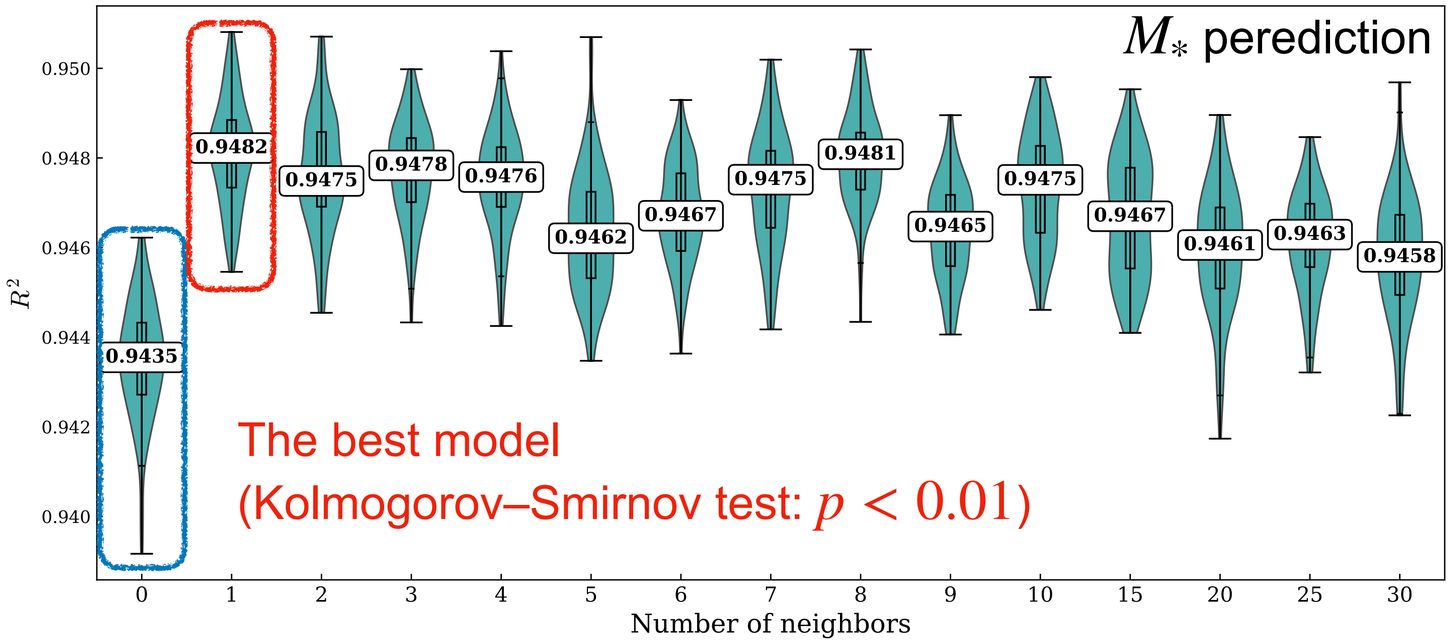
M_* or SFR

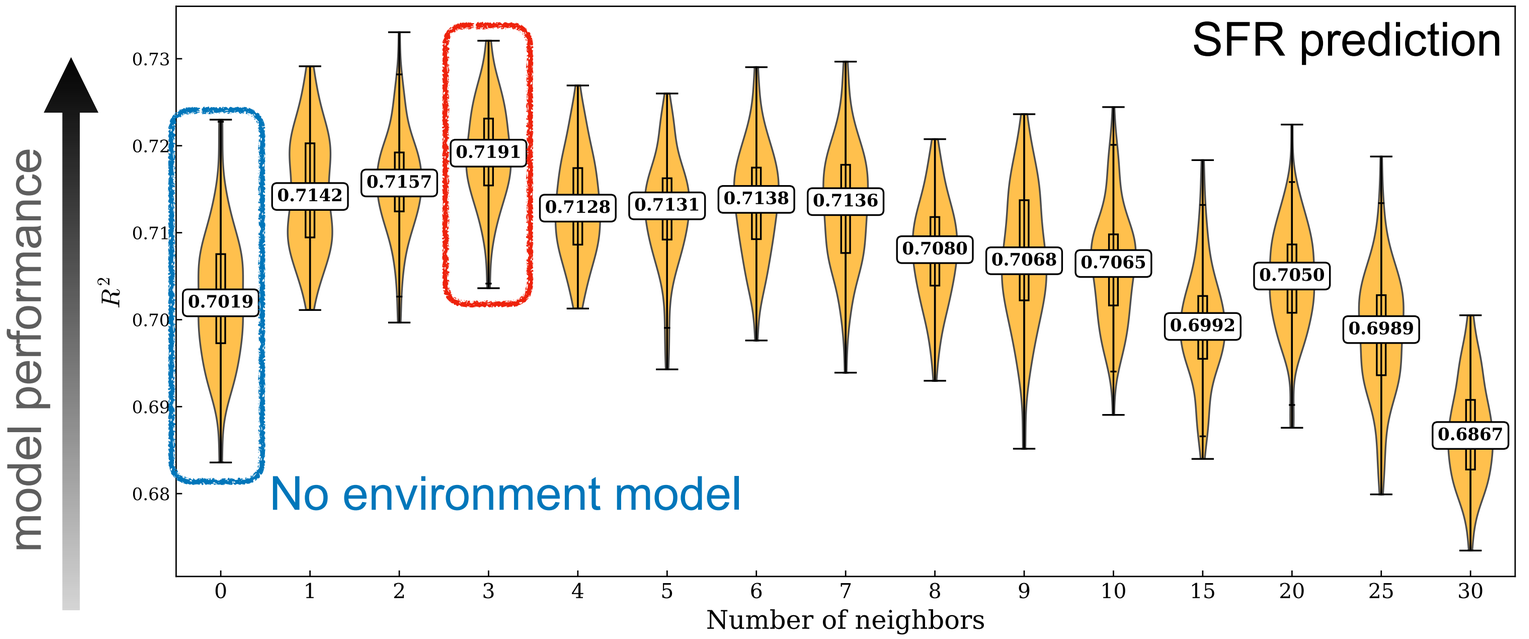
➡ Analyze feature contributions to the predictions

Result: model prediction (100 times bootstrap tests)

Comparison of different N nearest neighbors models by R^2

- **Best M_* prediction model:** 1-neighbor model.
- **Best SFR prediction model:** 3-neighbors model.
- Possible implications of **different scale dependencies of M_* and SFR.**
- We use these best models for subsequent analyses.

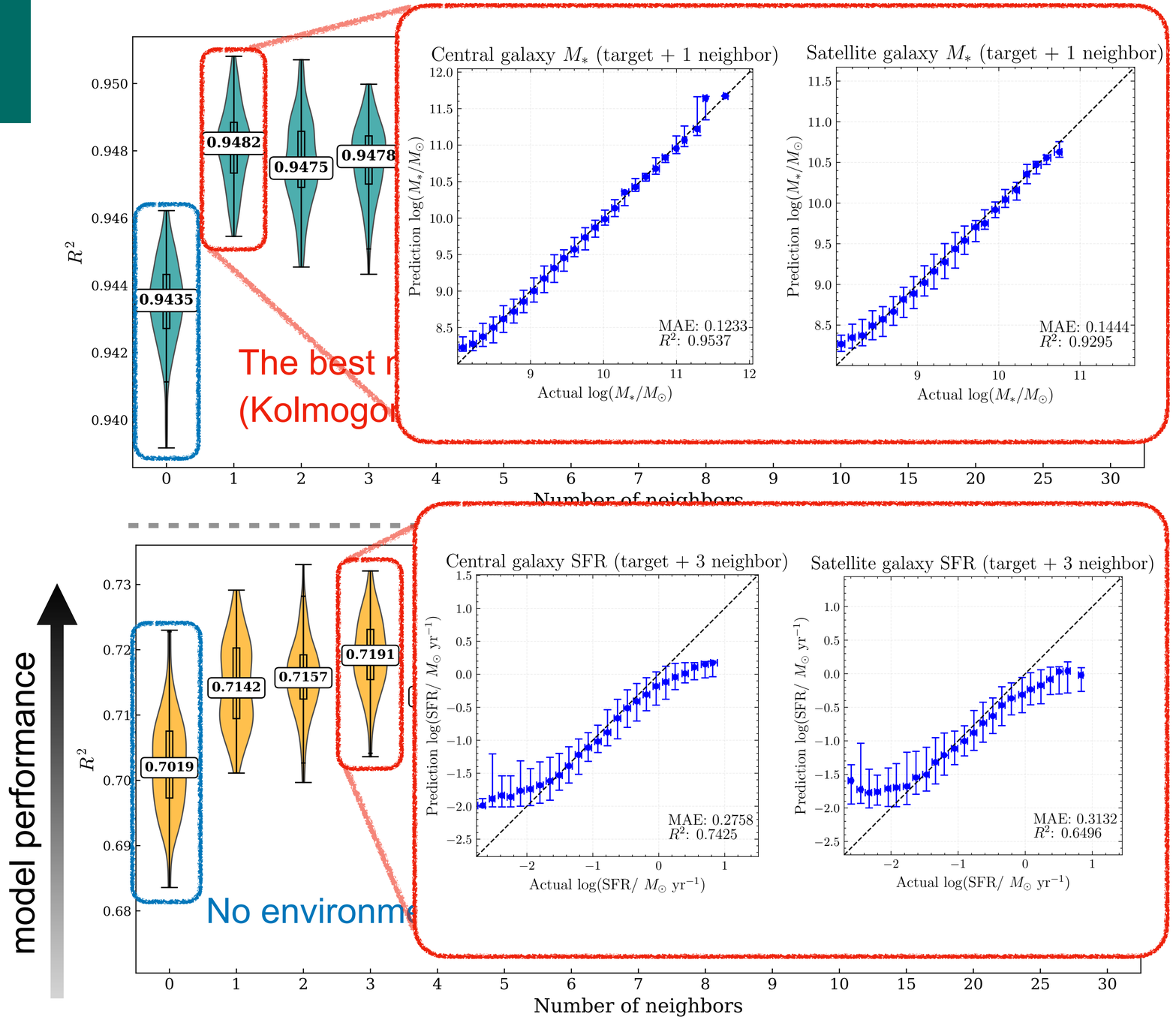




Result: model prediction (100 times bootstrap tests)

Comparison of different N nearest neighbors models by R^2

- **Best M_* prediction model:** 1-neighbor model.
- **Best SFR prediction model:** 3-neighbors model.
- Possible implications of **different scale dependencies of M_* and SFR.**
- We use these best models for subsequent analyses.

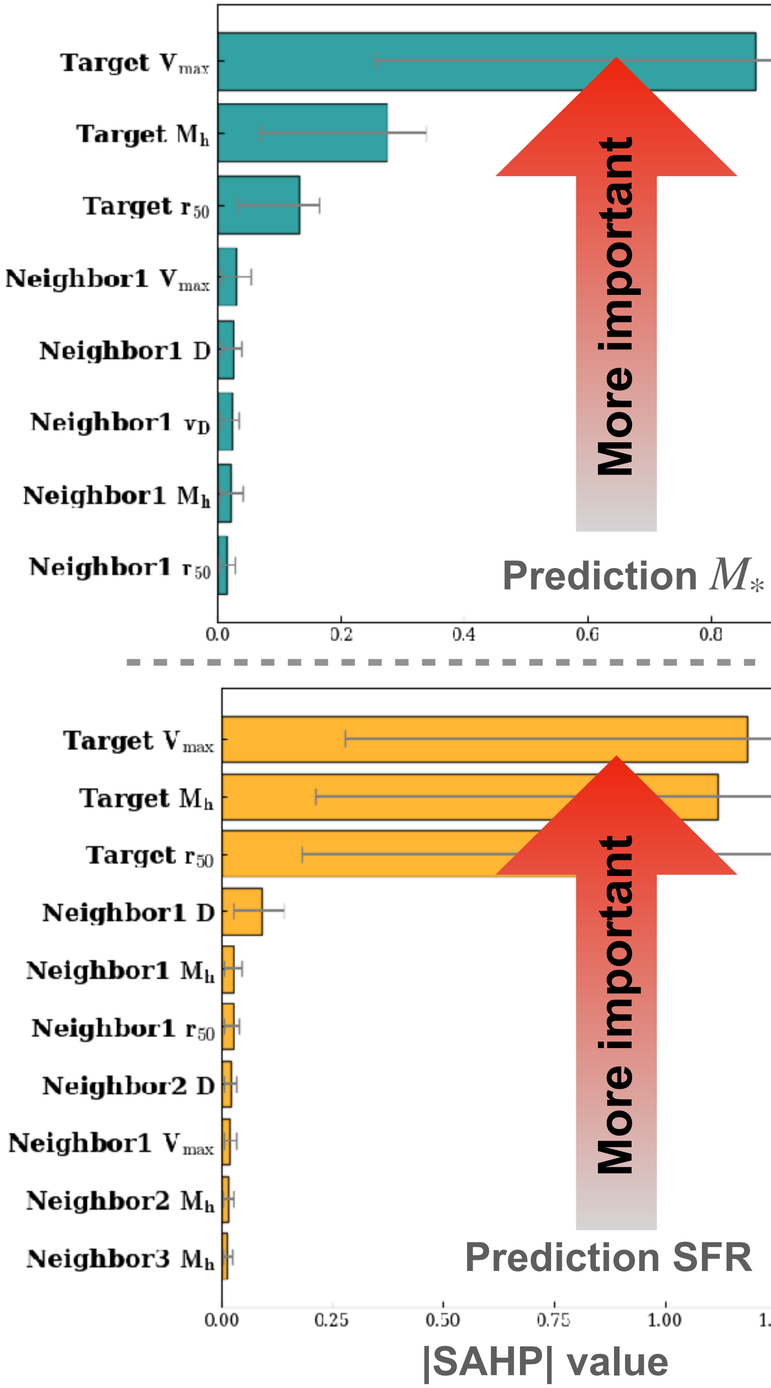


Influence from the environment

Interpreting Model Output: XAI (Explainable AI)

- **SHAP** (SHapley Additive exPlanations) value:
  Measures each input feature's contribution to predictions
  (correlation, not causation)

➡ The main contribution to prediction is the host subhalo properties.
(e.g., own $V_{\rm max}$, M_h , r_{50})



Influence from the environment

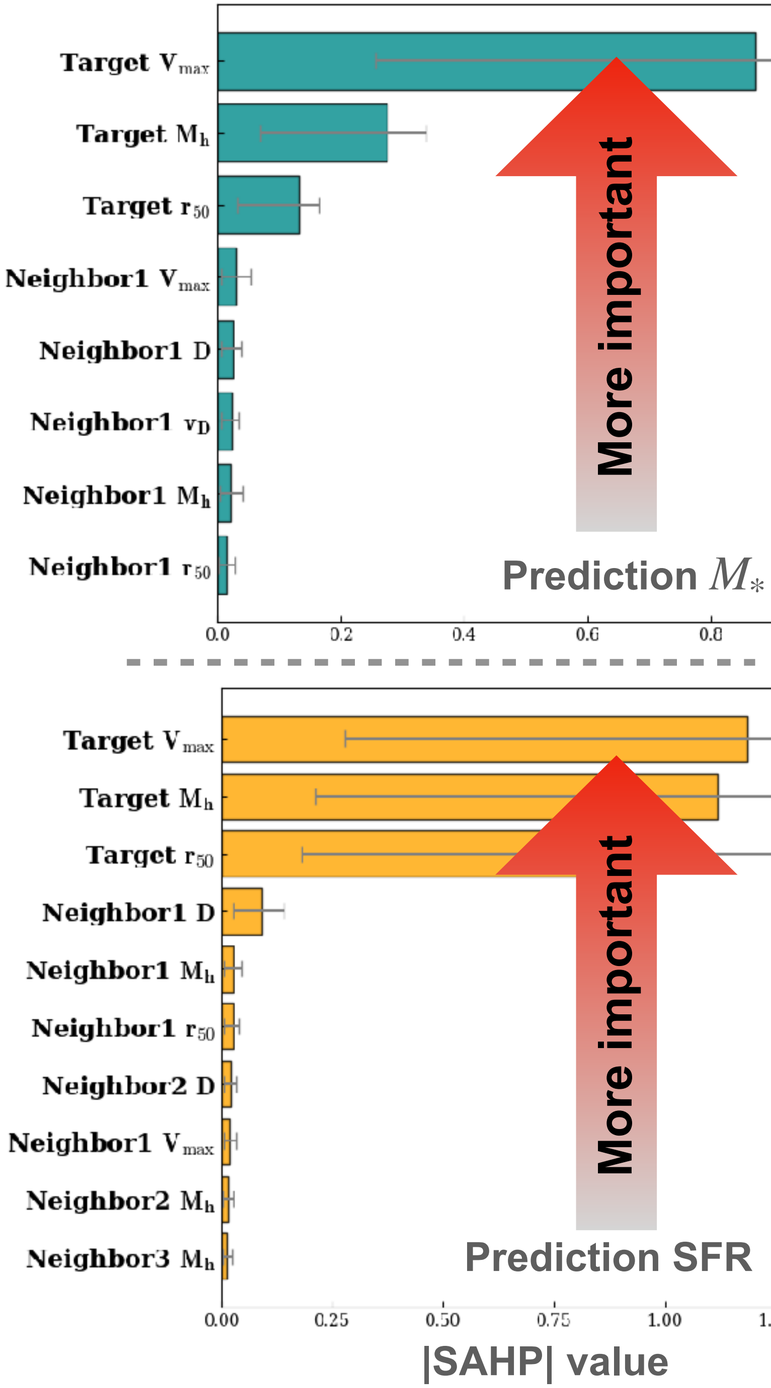
Interpreting Model Output: XAI (Explainable AI)

- **SHAP** (SHapley Additive exPlanations) value:
  Measures each input feature's contribution to predictions
  (correlation, not causation)

➡ The main contribution to prediction is the host subhalo properties.
(e.g., own $V_{\rm max}$, M_h , r_{50})

- To assess the influence of environmental effects,
  we excluded $V_{\rm max}$, and r_{50} from inputs in the SHAP calculation.
  ➡ The models rely as much as possible on the environment for predictions.

- Mutual correlations among M_h, $V_{\rm max}$, and r_{50}
  ➡ Complicates SHAP interpretation
- $V_{\rm max}$ and r_{50} reflect both environmental effects and **assembly history**
  ➡ Difficult to isolate the environmental effect

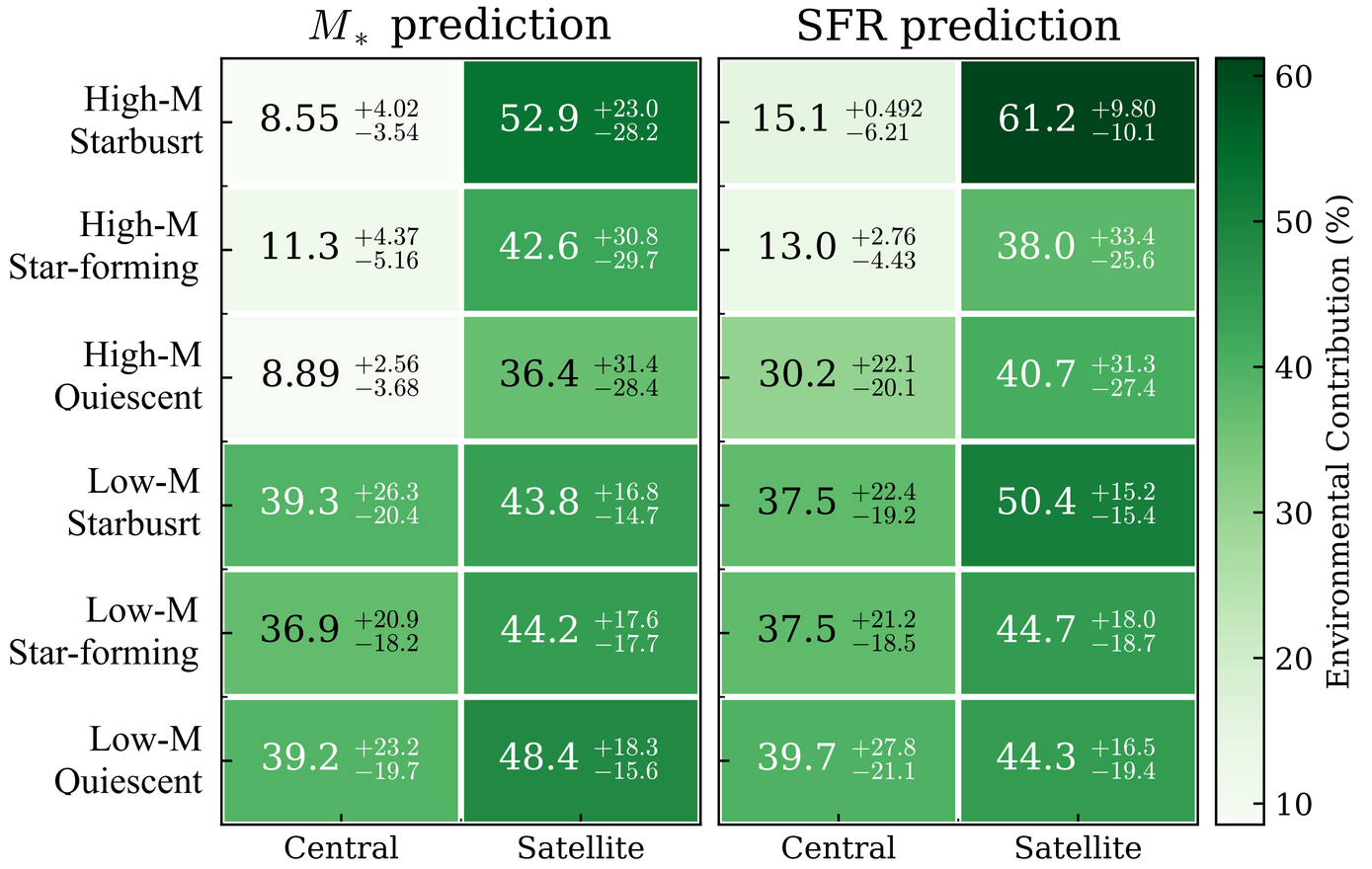


Result & Discussion: isolating the influence of environment

The effect of the environment in our model is about these ratios r of SHAP values S **at the most**.

$$r = \frac{\sum_j \sum_{i \in \text{neighbours input}} |S_i^j|}{\sum_j \sum_{i \in \text{all input}} |S_i^j|} \quad (S_i^j : \text{for } i\text{-th input for } j\text{-th galaxy})$$

- Environmental effects are not uniform and vary based on galaxy type.
- Low-mass and satellite galaxies received greater contributions from the surrounding environment
- Exception: Satellite Starburst Galaxies
  - High-mass satellite galaxies are more affected by the environment than low-mass ones.



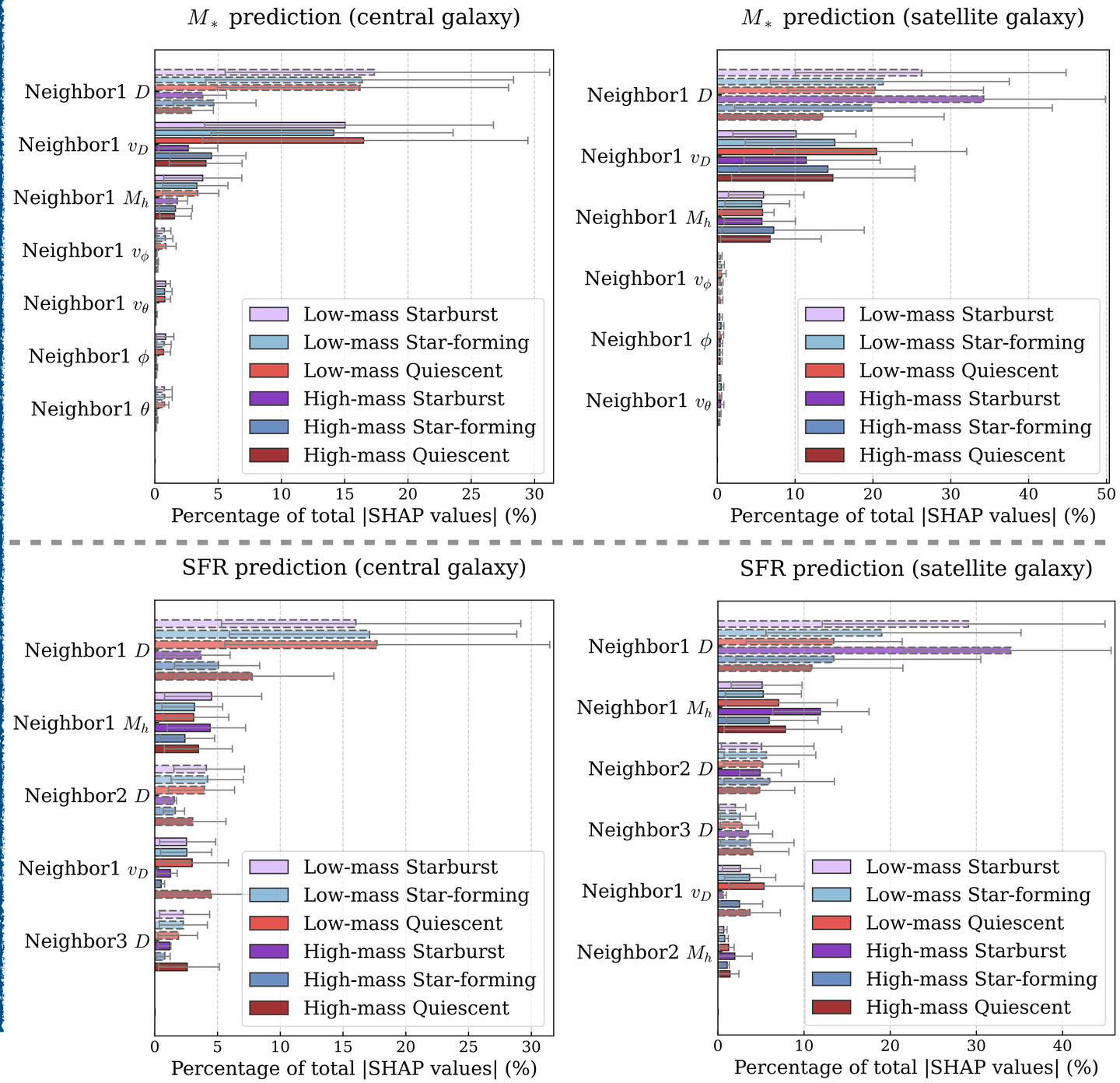
Result & Discussion: Contribution from the environment features

- Key environmental features:
  - Distance (D) to the surrounding galaxy (nearest galaxy)
  - M_h of the nearest galaxy

  ➡Suggests strong influence of close interactions

  - v_D of the nearest galaxy
    - Especially with central's M_* predictions
    - Indicates dynamic interactions

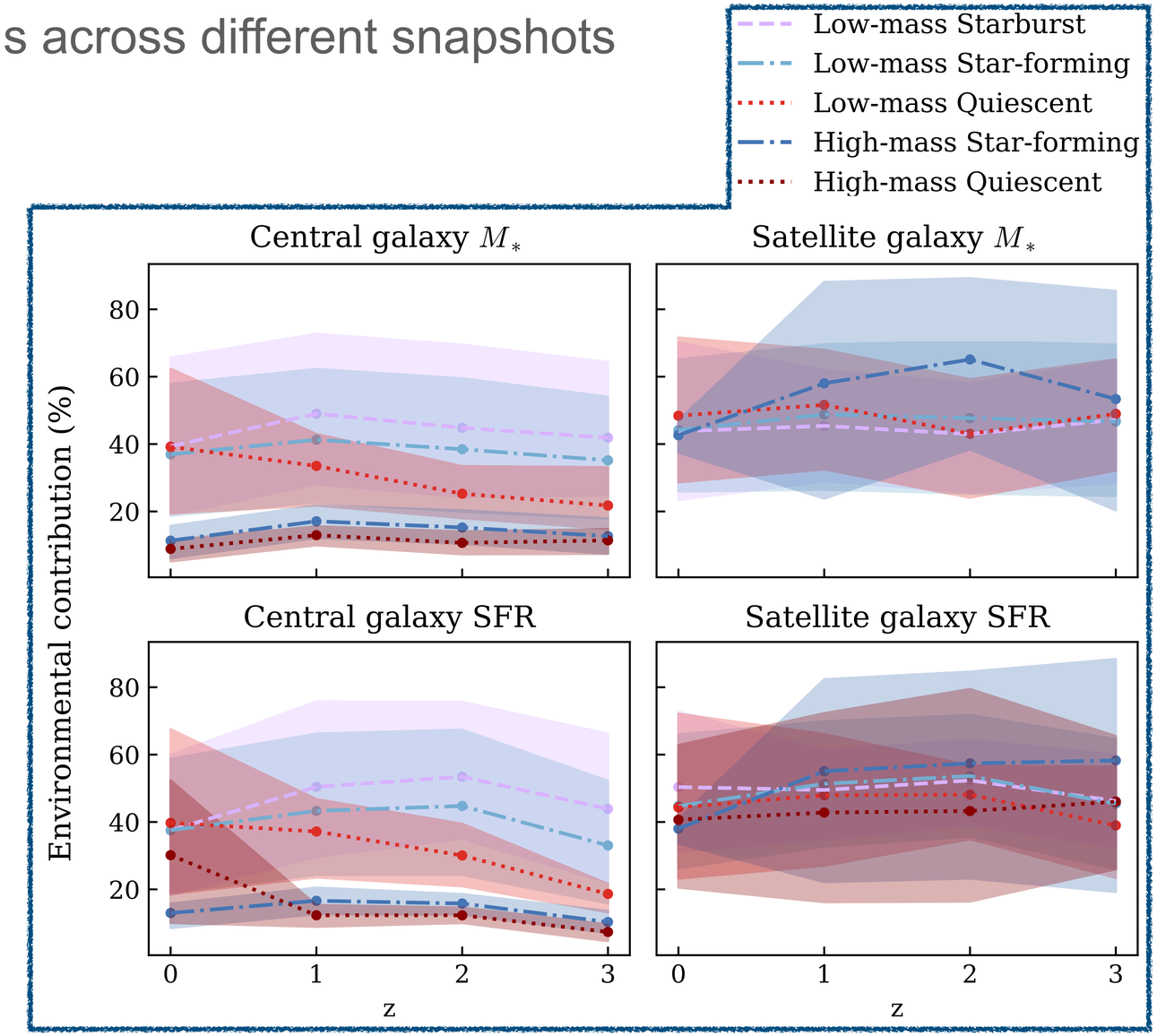
- M_* reflects long-term star formation history and aligns with SFR trends.



Result & Discussion:
Isolating the environmental influence across $z = 0 \sim 3$

We compare the relative environmental contributions across different snapshots ($z = 0,1,2,3$).

- For central galaxies:
  - The larger contribution of low-mass systems (compared to high-mass) persists up to $z = 3$.
  - At $z = 0$, no dependence on star-formation.
  - At $z = 1 \sim 3$, low-mass galaxies with higher star-formation activity show larger environmental contributions.
- For satellite galaxies:
  - No significant difference across redshift.



Summary

How do environmental factors influence galaxy properties (M_*, SFR) across different galaxy populations?

- **Method**
  Developed an interpretable neural network framework trained on **IllustrisTNG300-1** to predict M_* and SFR using itself and the surrounding DM subhalos properties.

- **Results & Key findings** (at $z = 0$)
  - M_* is best predicted using only **the nearest neighboring galaxy** as an environmental factor.
  - SFR prediction improves with up to **three nearest neighbors.**
  - The **distance to the nearest neighboring galaxy** is the most influential factor for predicting.
  - Low-mass galaxies and satellite galaxies are more strongly influenced by environmental factors, in star-forming and quiescent populations.
  - Satellite high-mass starburst galaxies experience the strongest environmental effects among all groups classified in this study.

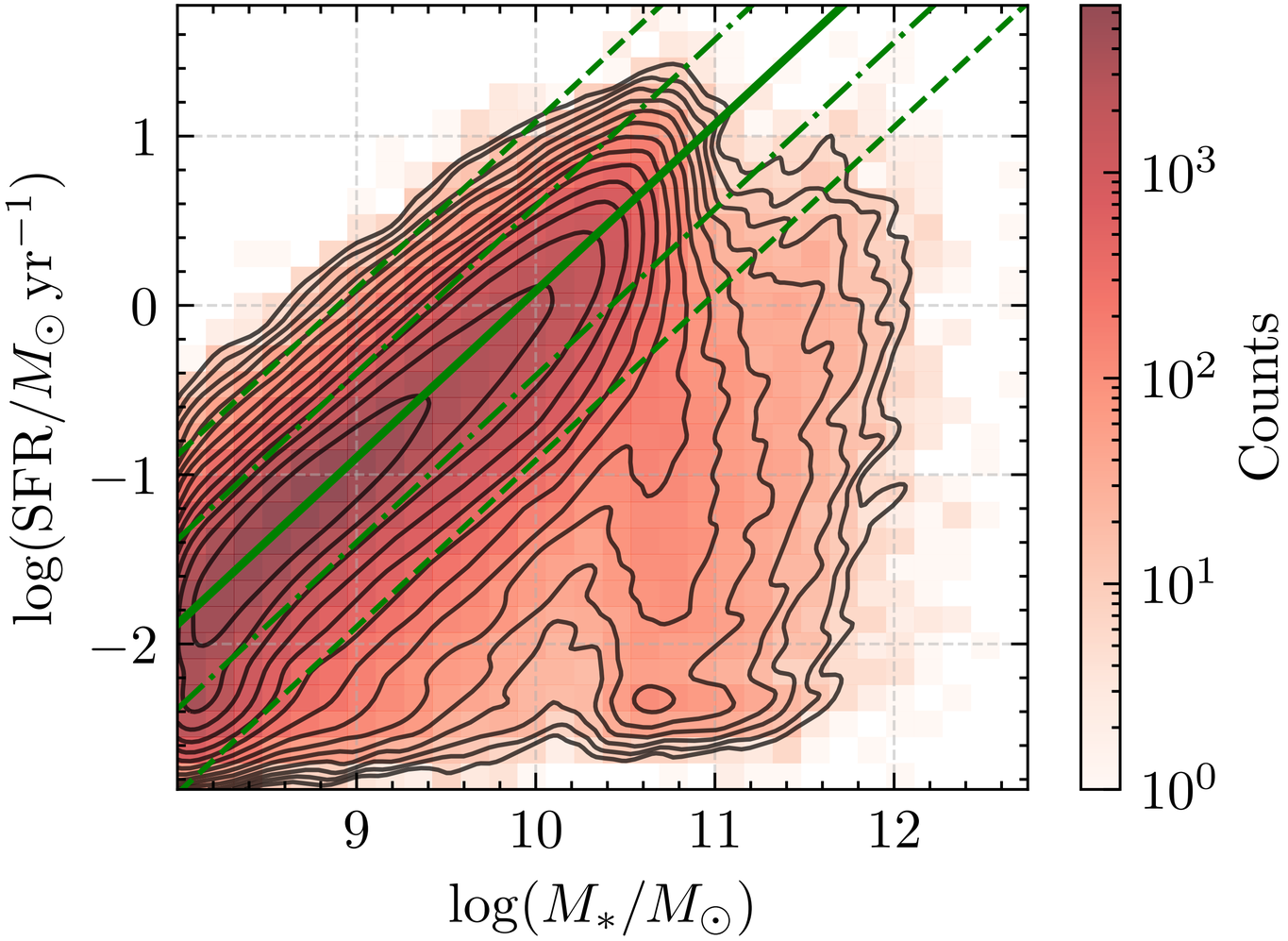
➡ **These results offer guidance for constructing more sophisticated empirical and semi-analytic models of galaxy formation that explicitly include environmental dependence as a function of galaxy type and mass.**

Appendix

IllusitirisTNG300 data

All datasets in this study

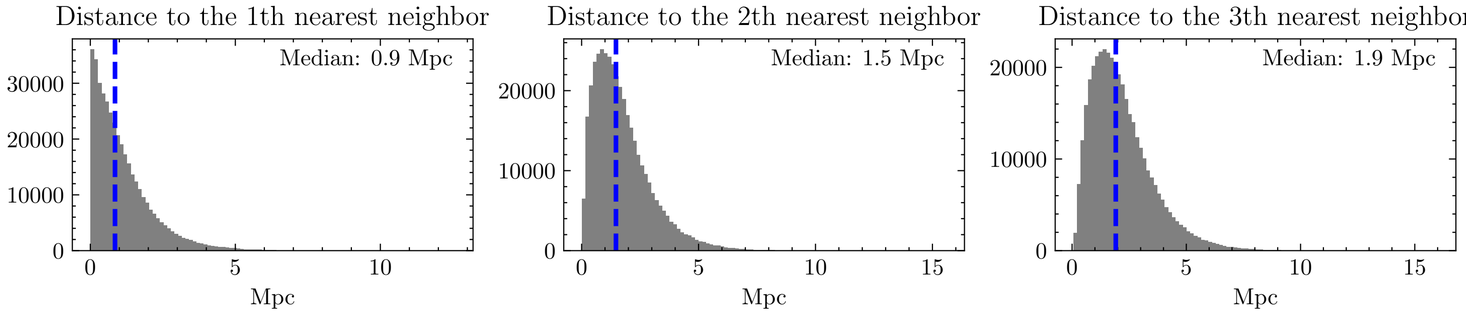
Central (291,885 objects) + Satellite (101,846 objects)

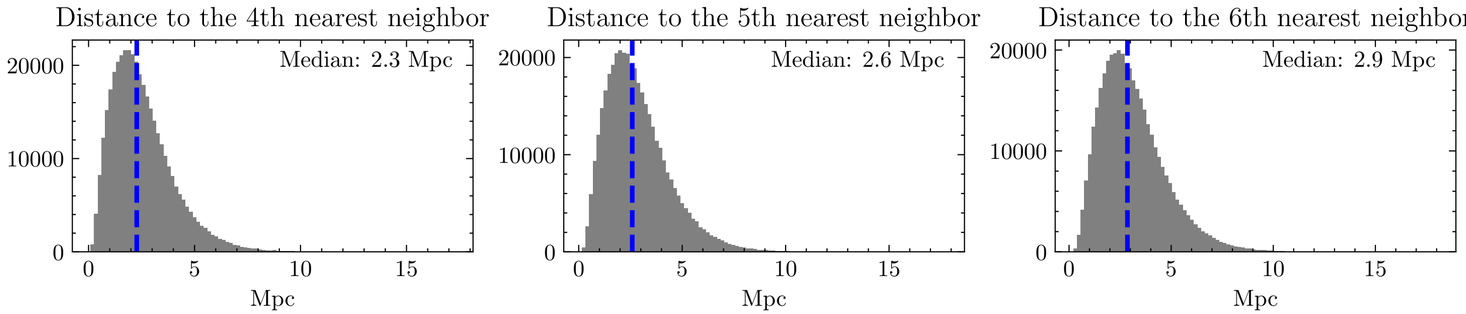


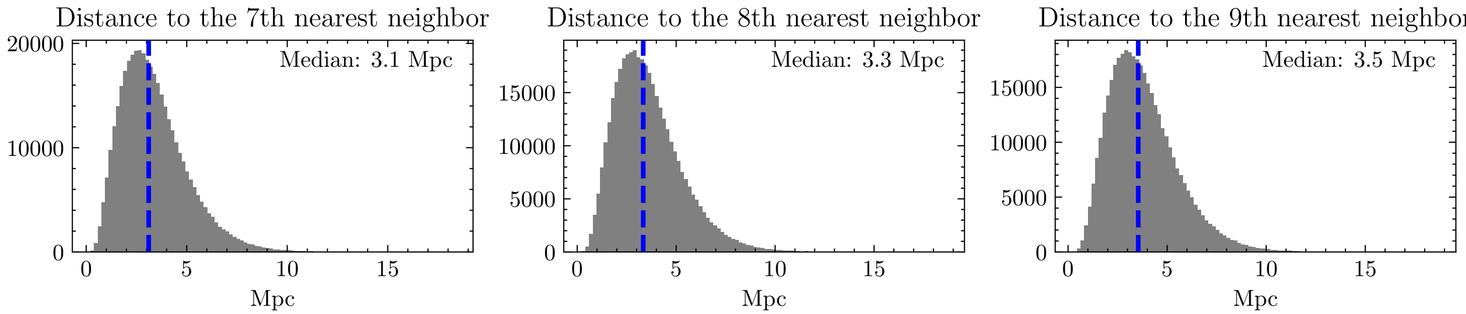
We applied PCA to high-density areas in the top 1% of KDE density

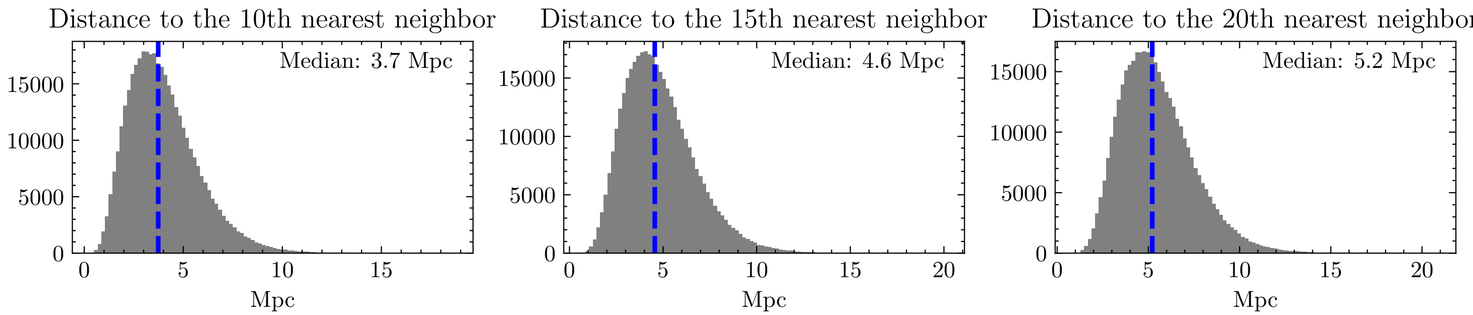
➡ To get MS

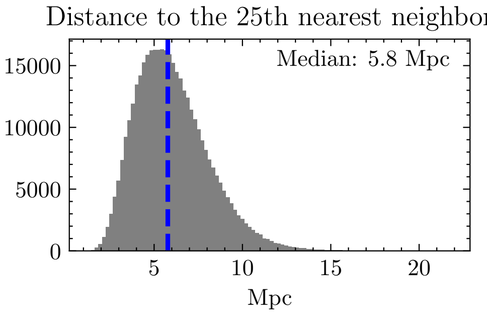
TNG300-1 (z=0): Distribution of distance to the neighbor galaxy

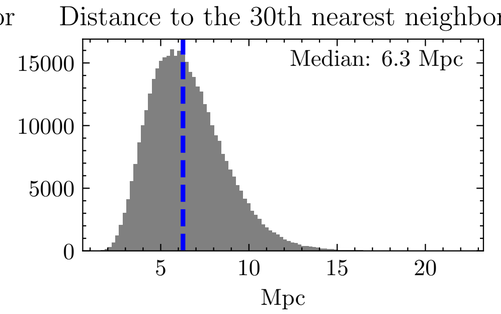








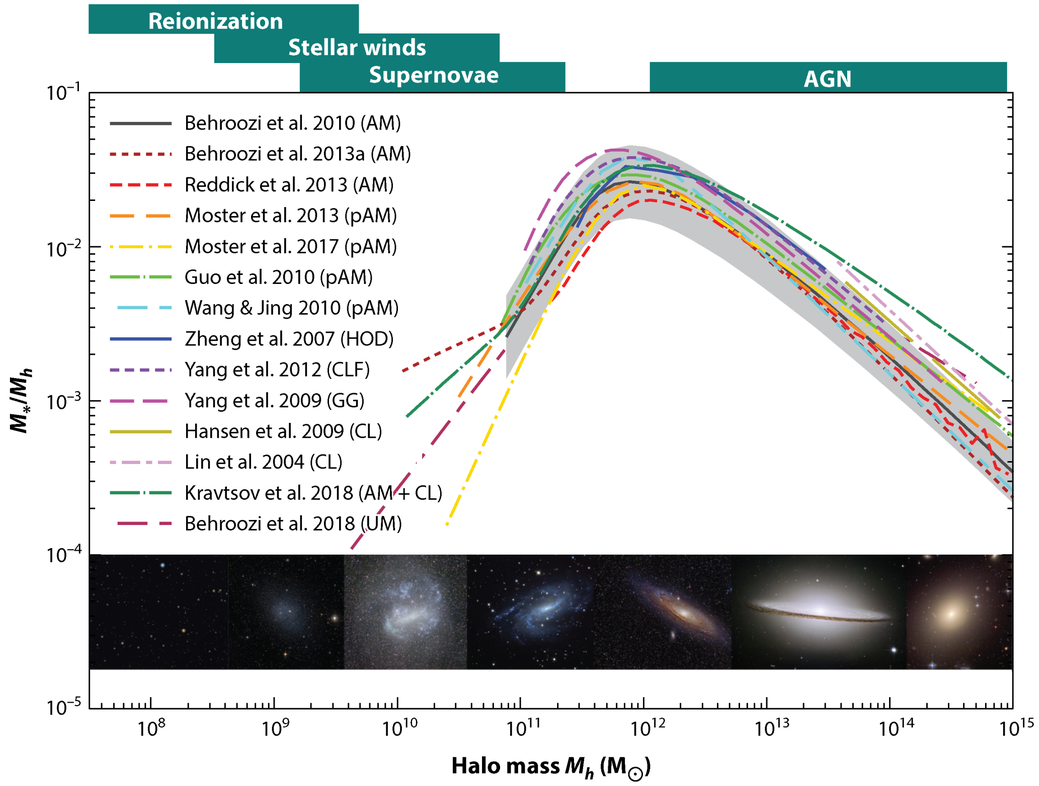




Number of neighboring galaxies for M_*, SFR prediction

SFR prediction accuracy declines beyond the 3rd nearest neighbor—Why?

- Possible difficulty in learning the model due to increased input dimensions.
- Possible implications of **different scale dependencies of M_* and SFR.**
  - Since M_* is strongly correlated with M_h, interactions with the nearest neighbor galaxy, which has strong gravitational ties, dominate.
  - **SFR is influenced on a broader scale**, not only by gas inflow and outflow due to gravitational interactions but also by various mechanisms.
    (e.g. stellar feedback, AGN feedback)

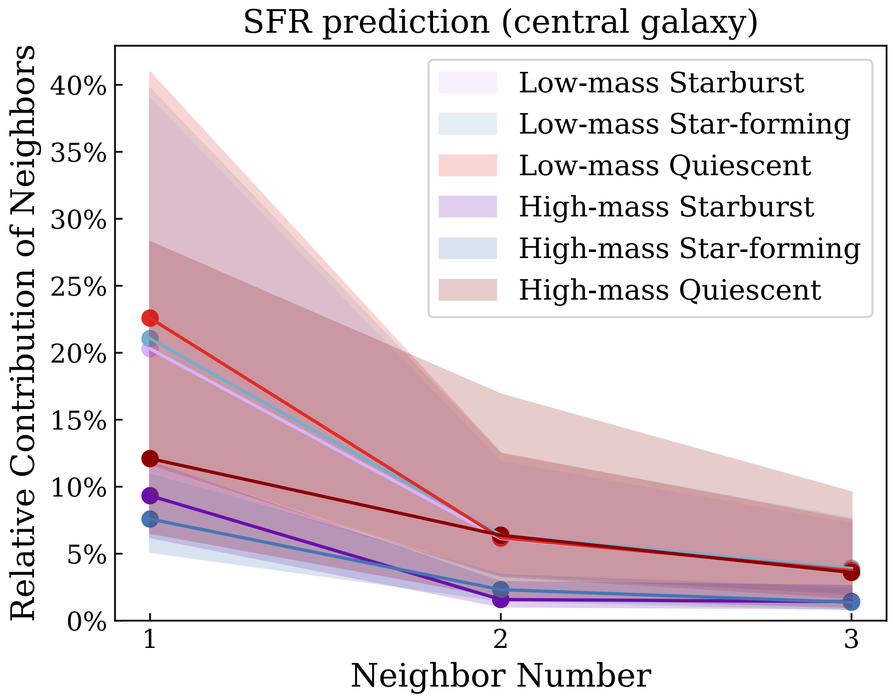


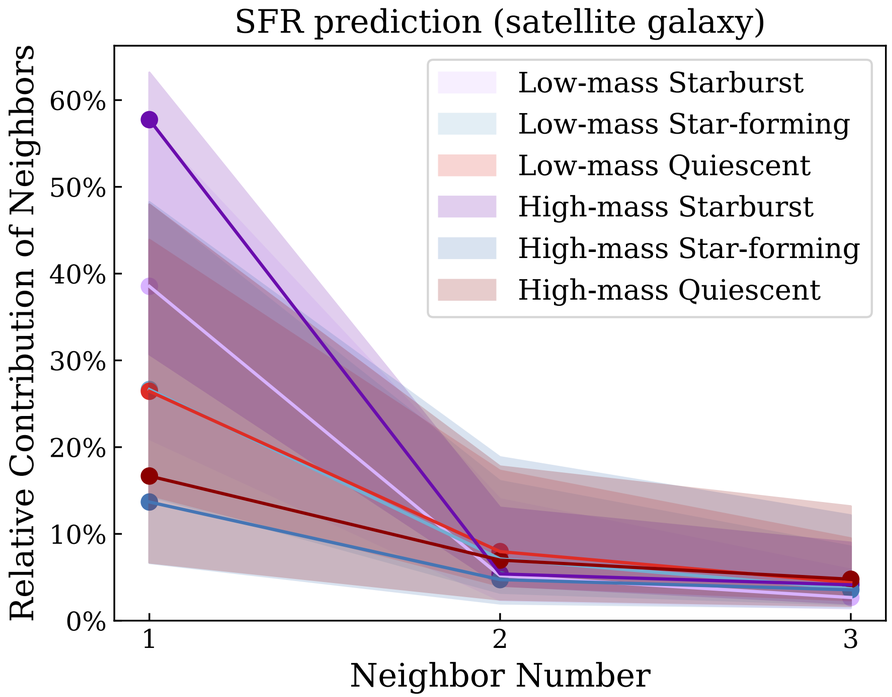
(Wechsler et al. 2018, ARA&A, 56, 435-487)

3-2. Result & Discussion: The relative contributions of individual neighboring galaxies

- The nearest neighbor providing the dominant contribution to SFR predictions
  ($\sim 8-23\,\%$ for centrals and $\sim 14-58\,\%$ for satellites)

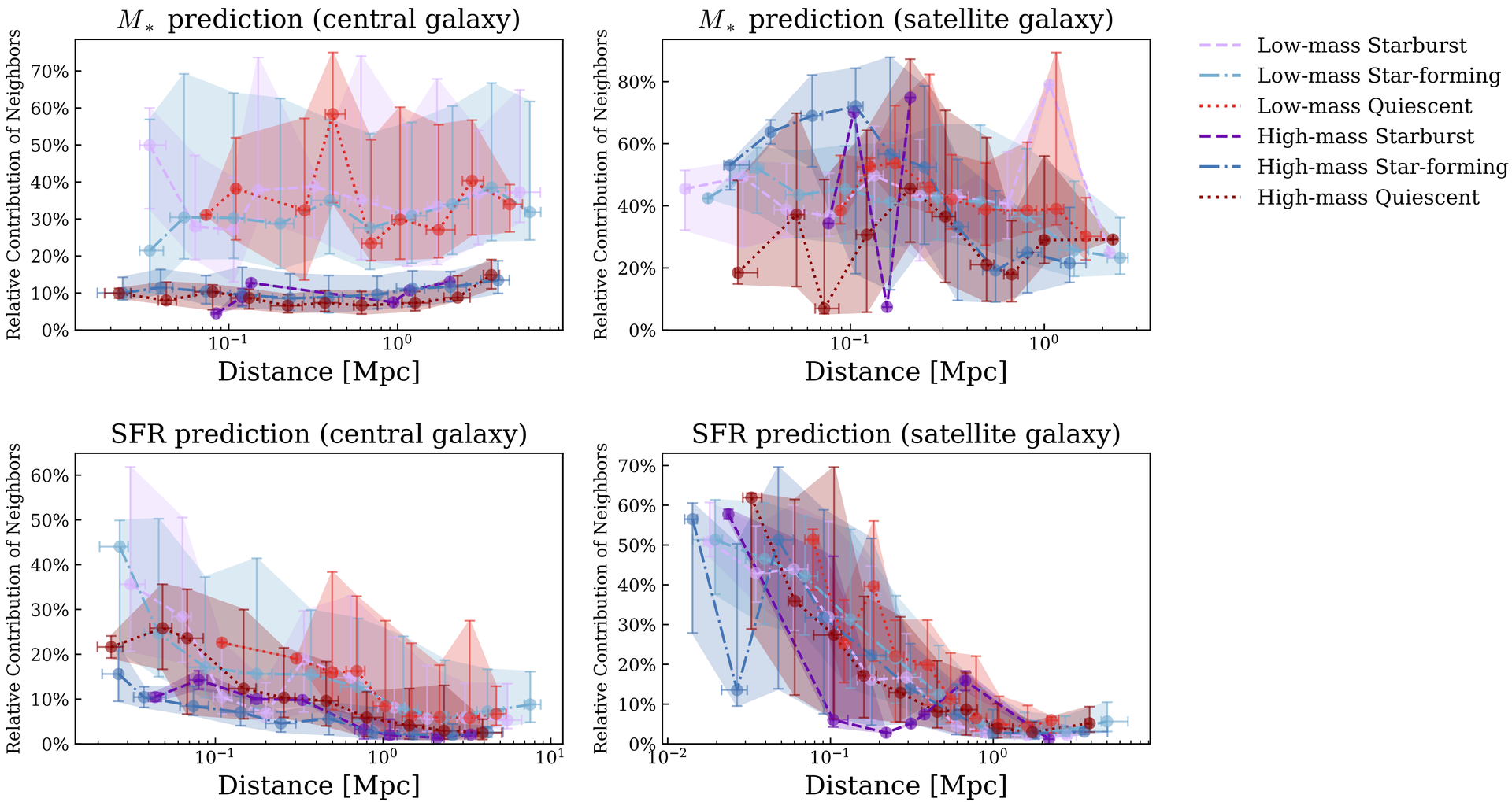
  ➡ While the overall magnitude of environmental dependence varies with galaxy types, the relative contribution of neighbors follows a distance-based hierarchy.





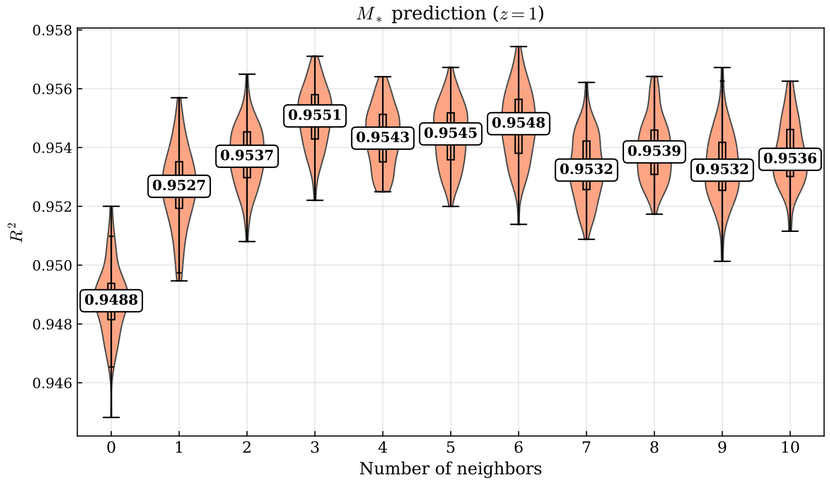
The dependence of the SHAP values of environmental features on distance

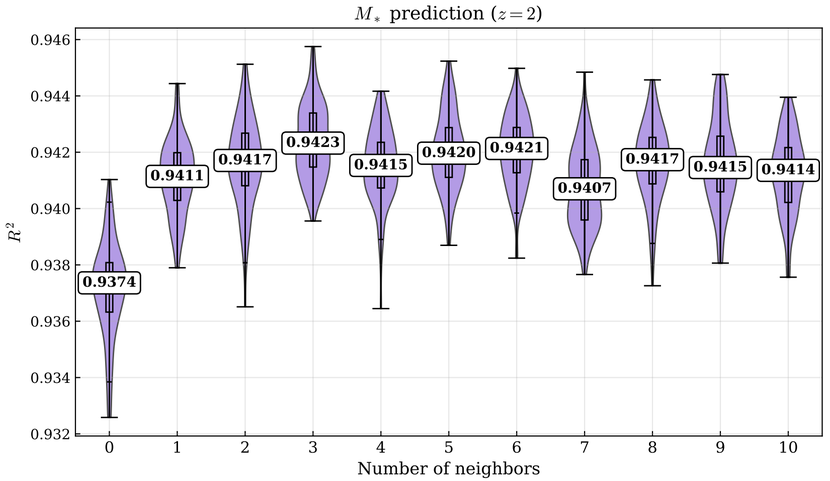
- Nearby galaxies into uniform bins on a log-scale

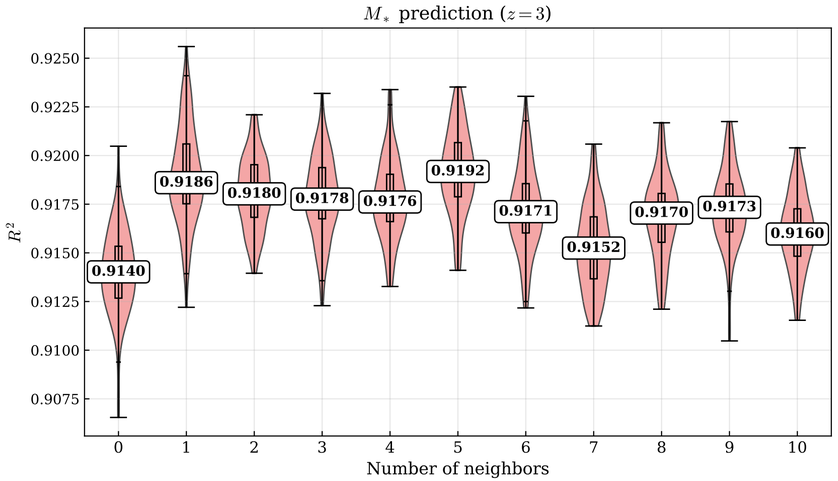


R^2 distribution at $z = 1,2,3$

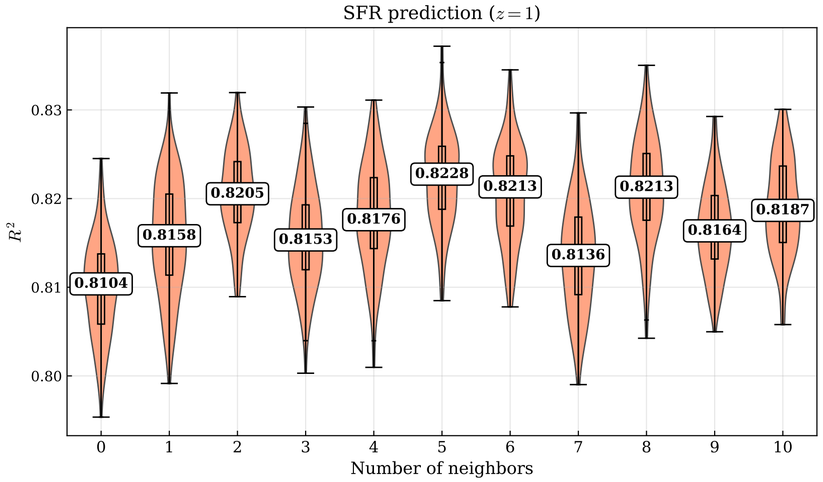
M_* prediction

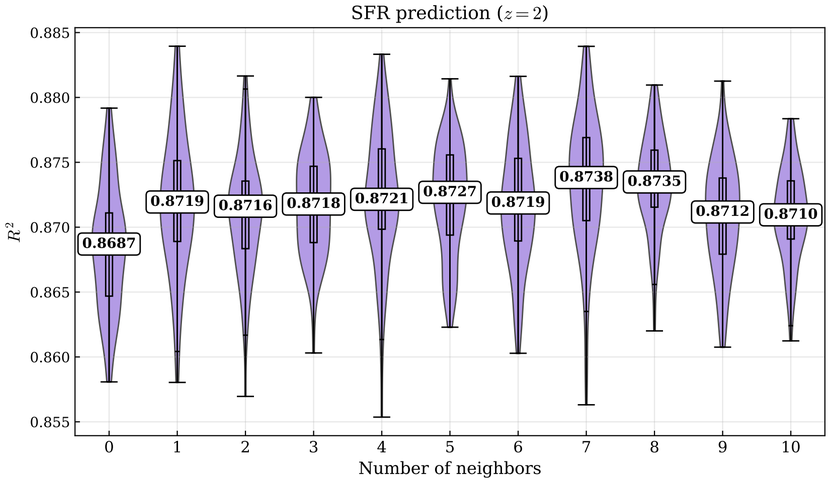


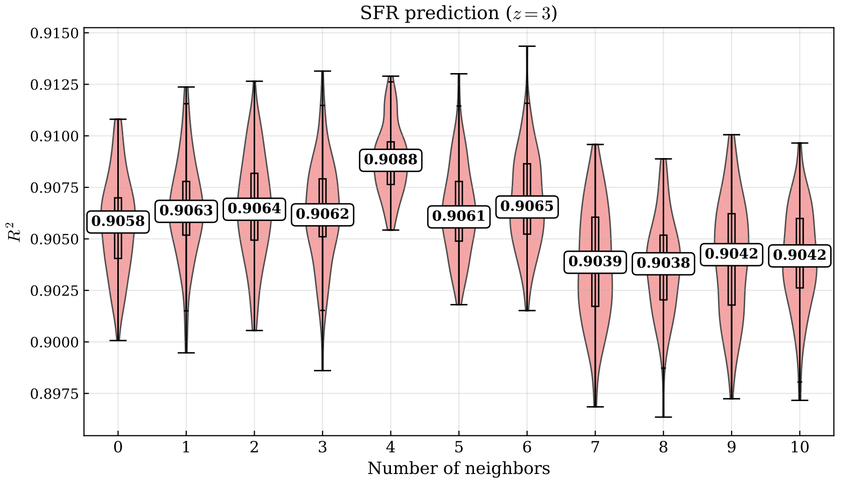


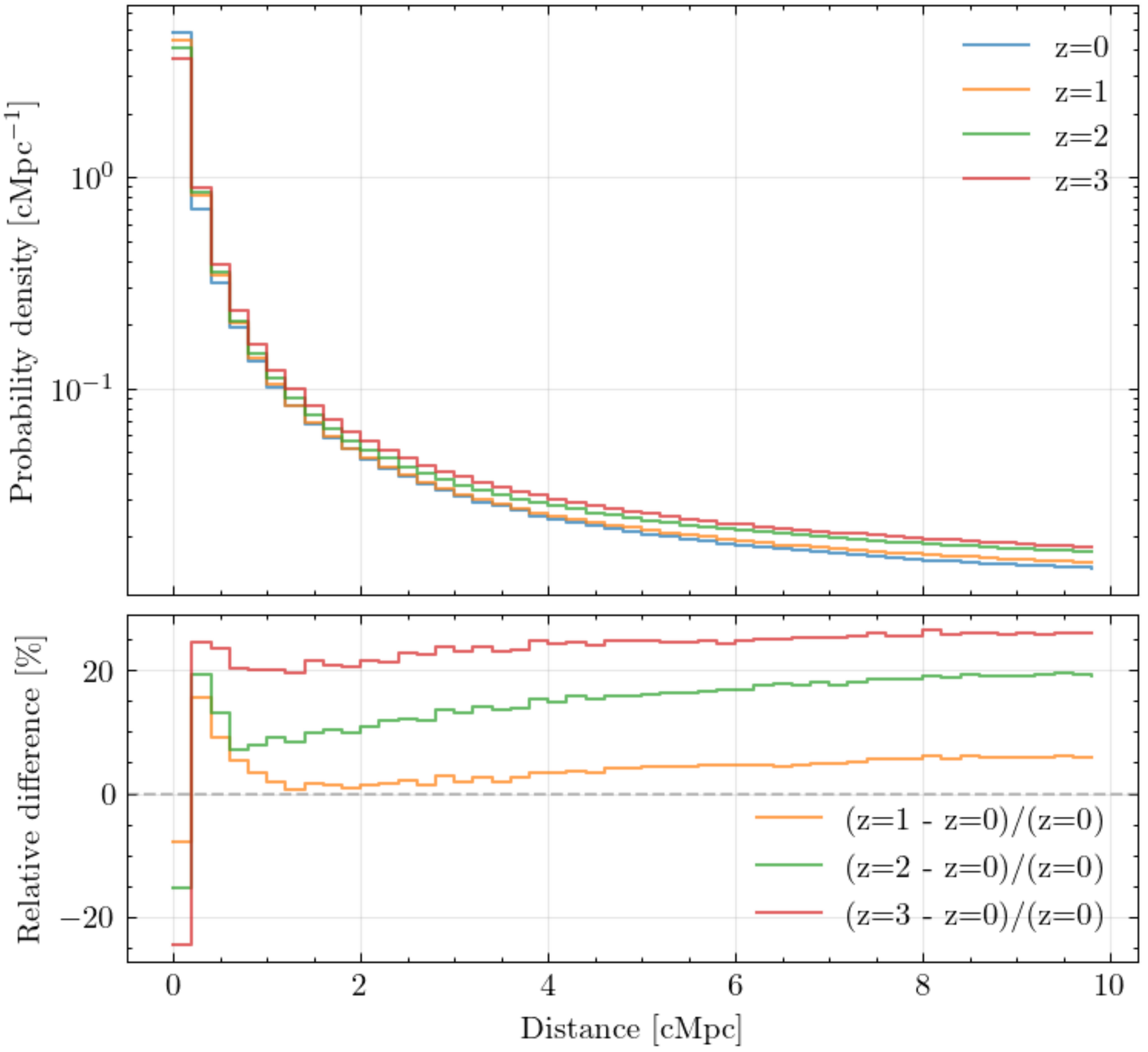


SFR prediction








TNG300-1: Distance distribution to neighboring galaxies within 10 cMpc
Probability density [cMpc^{-1}]
10^0
10^{-1}
z=0
z=1
z=2
z=3
Relative difference [%]
20
0
−20
(z=1 - z=0)/(z=0)
(z=2 - z=0)/(z=0)
(z=3 - z=0)/(z=0)
0
2
4
6
8
10
Distance [cMpc]