

Galaxy-(sub) halo relation in the eye of machine learning

Xiaoju Xu

Shanghai Jiao Tong University

collaborators: Xiaohu Yang (SJTU), Haojie Xu (SHAO), Youcai Zhang (SHAO)

The 2nd Shanghai Assembly on Cosmology and Structure Formation, 11.01

Establishing galaxy-halo connection

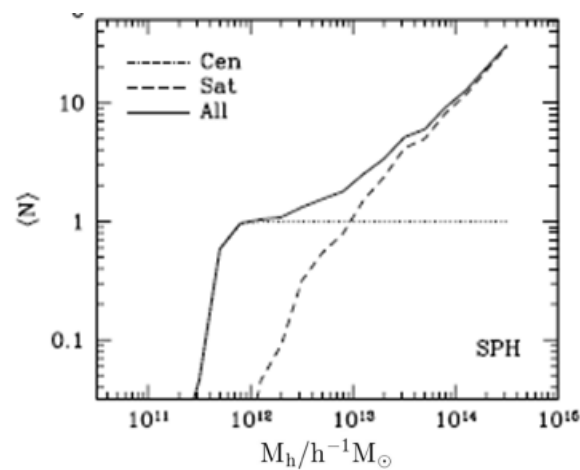
- Empirical models
- SAM and hydrodynamic simulation
- Machine learning methods

Predicting galaxy with (sub) halo properties

- Predicting halo occupation
- Predicting galaxy properties
- Predicting galaxy properties in observation

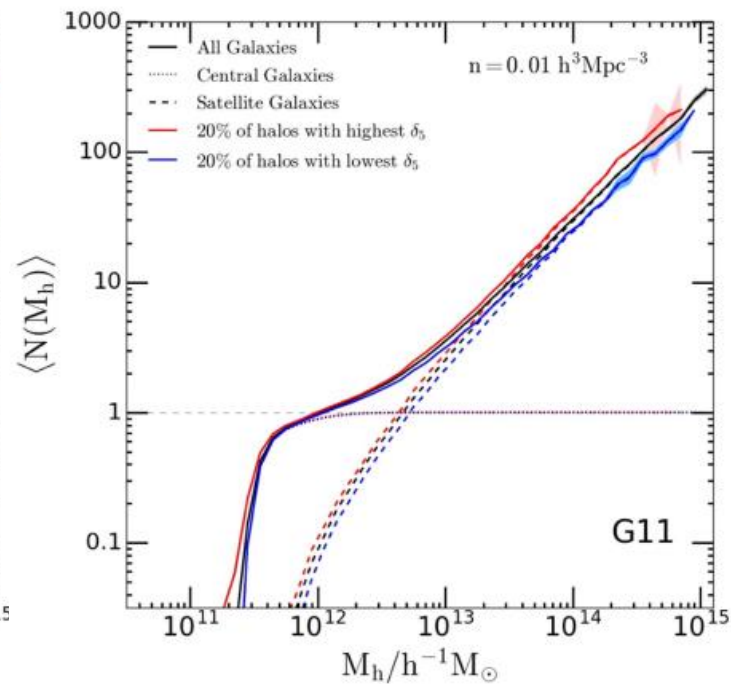
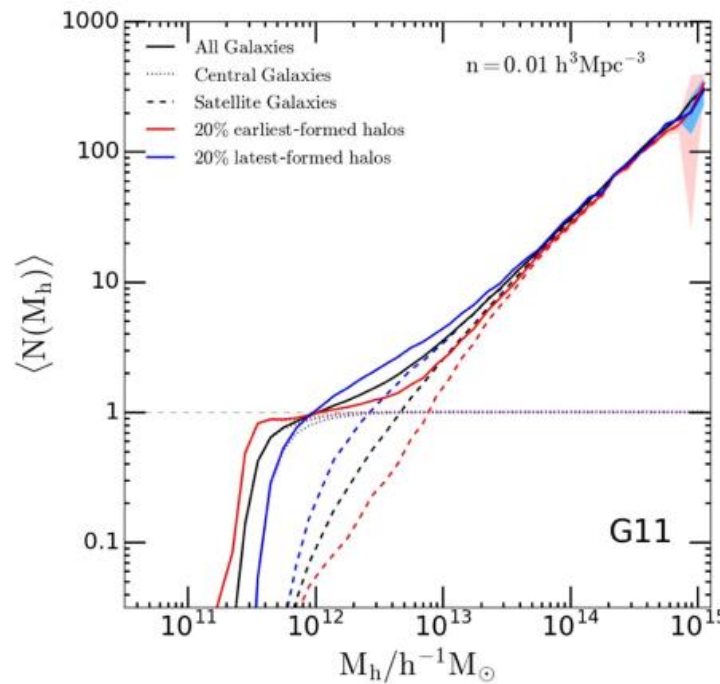
Predicting galaxy with (sub) halo properties

- Predicting halo occupation



Zheng et al. 2005

traditional HOD

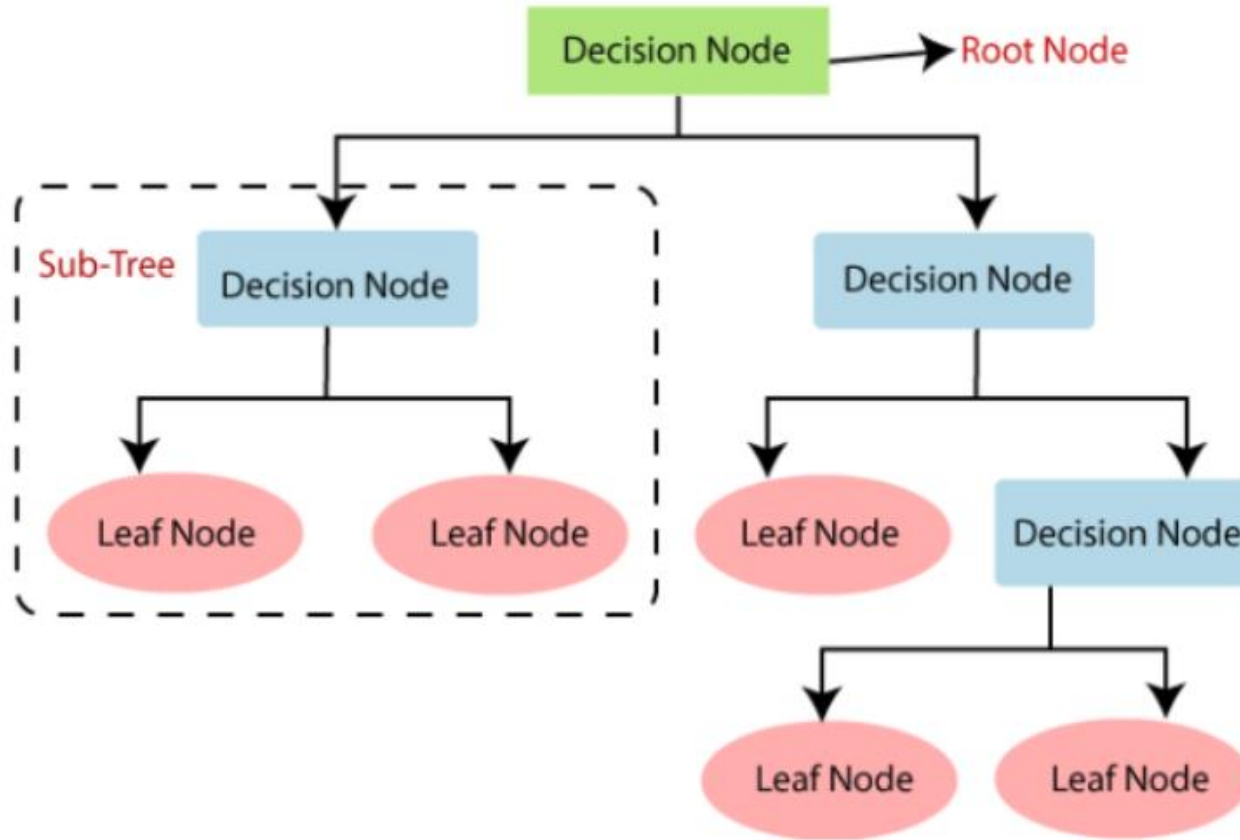


Zehavi et al. 2018

HOD in SAM

Predicting halo occupation in SAM

Random Forest



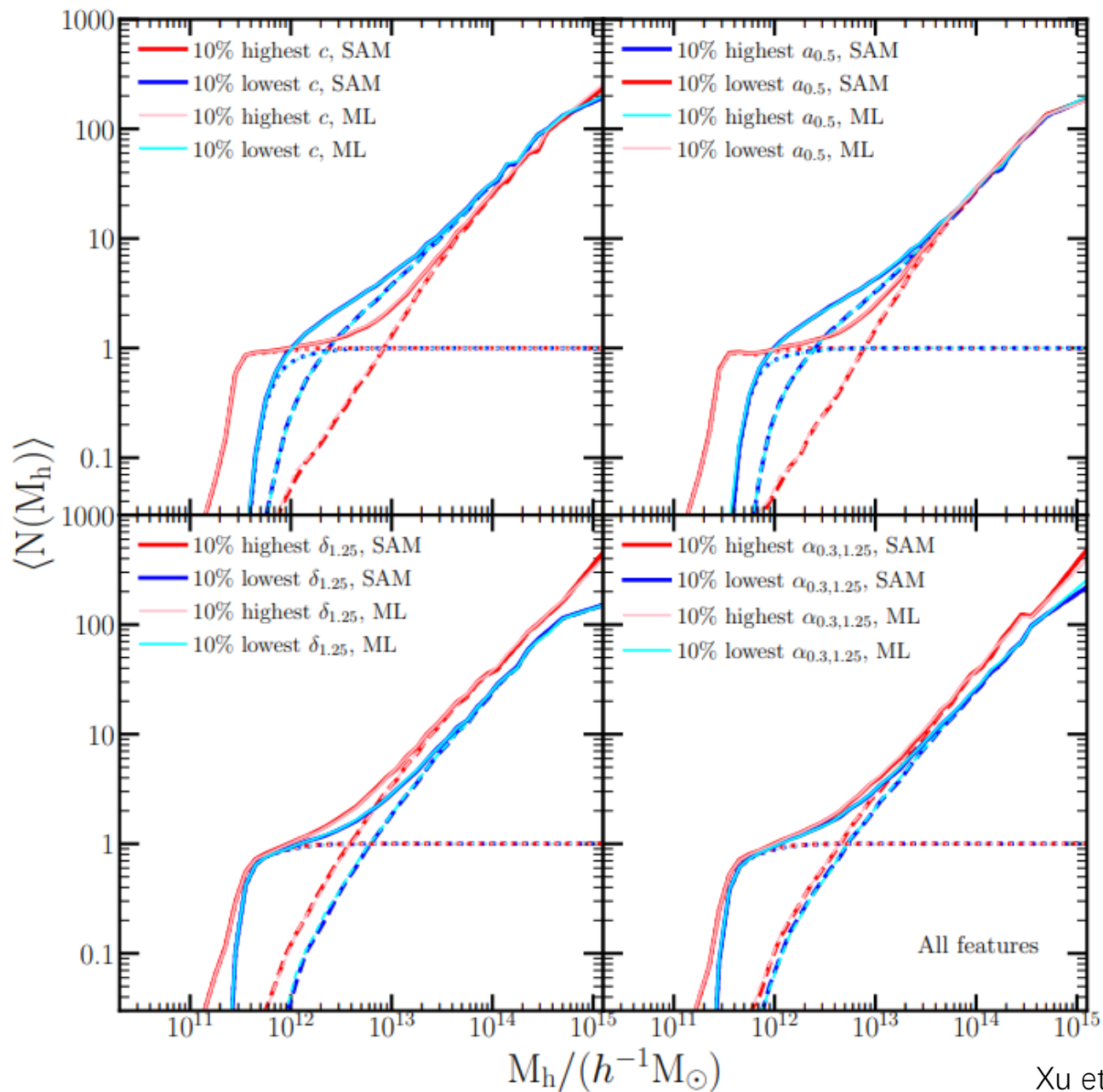
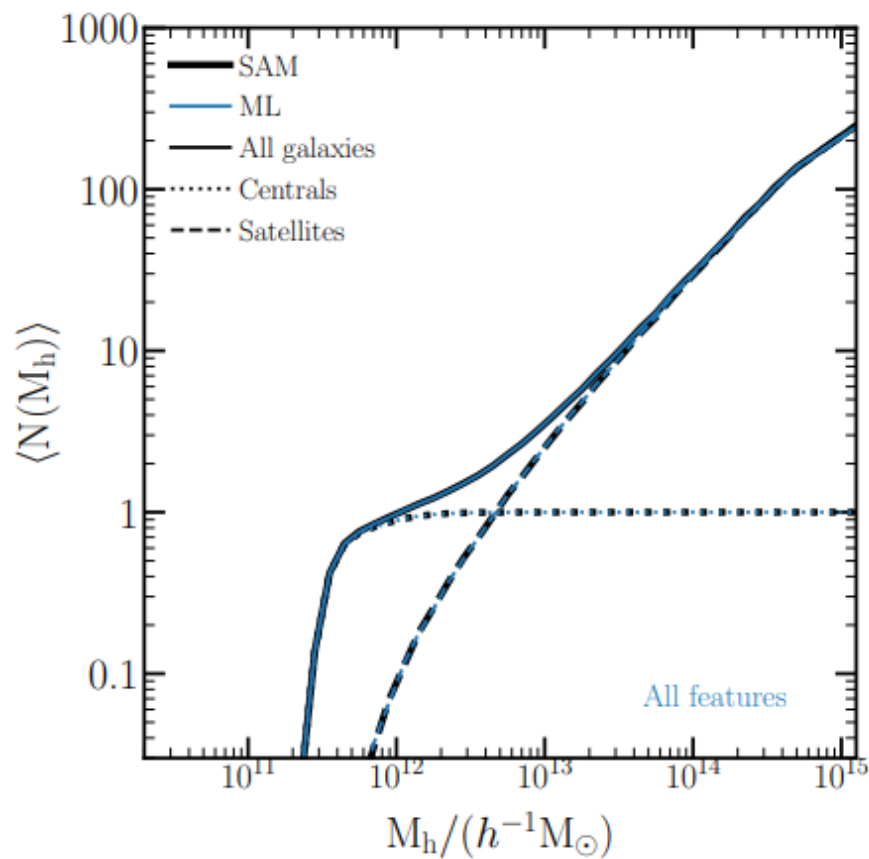
input: halo properties

M_{vir} , V_{peak} , $a_{0.5}$, concentration,
 a_{first} , a_{last} , n_{merge} ,
and environment (density, anisotropy)

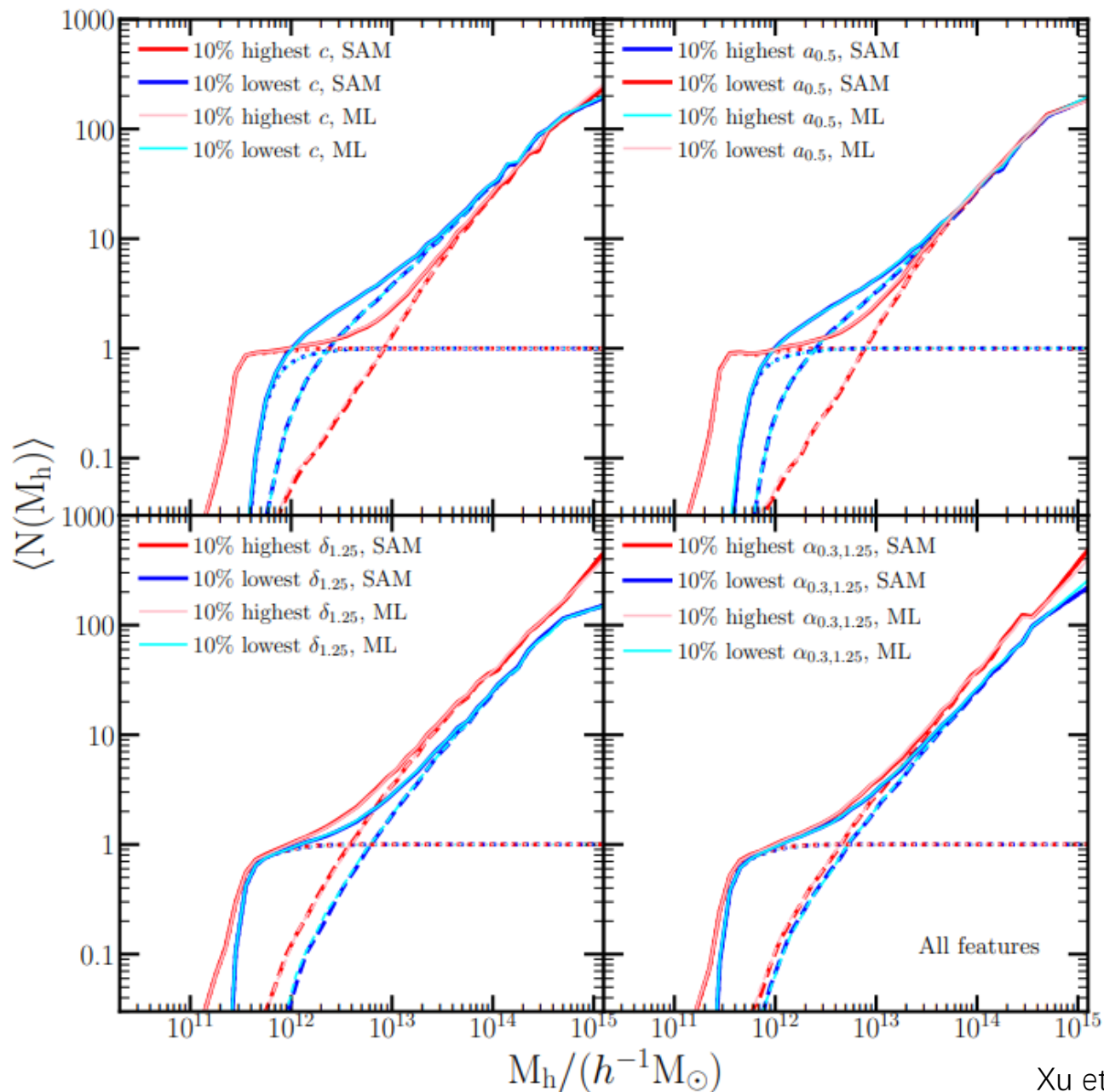
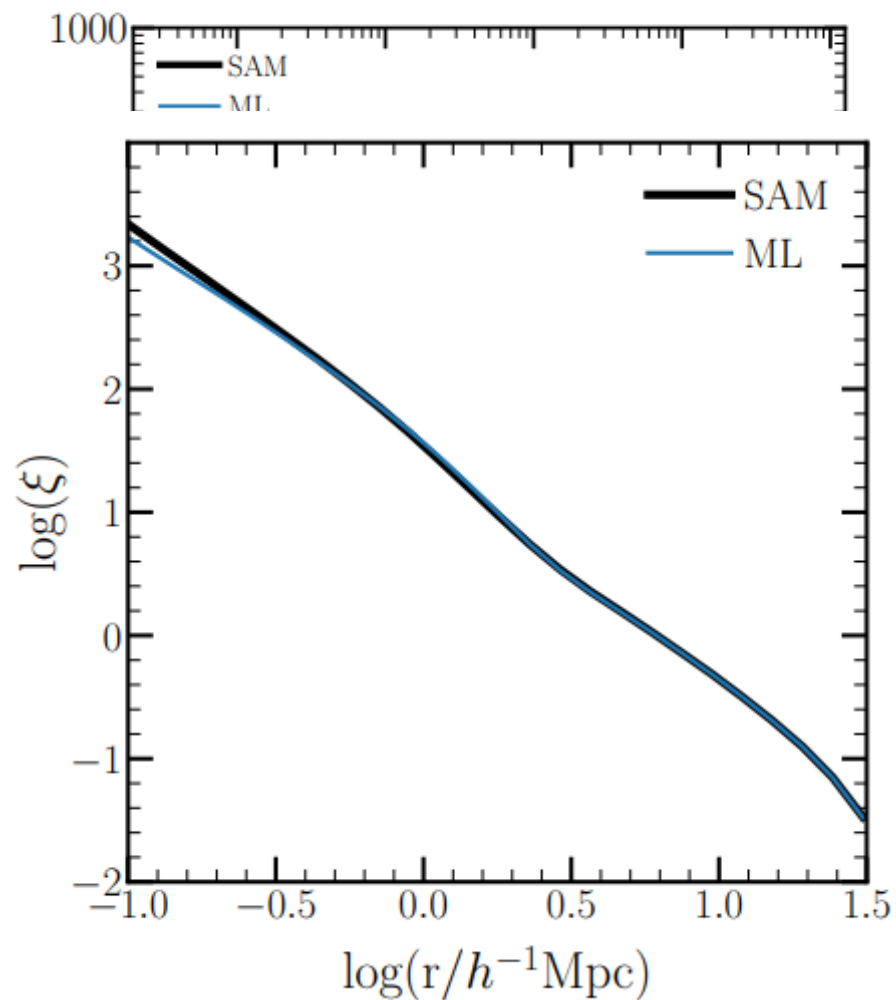
output: halo occupation

N_{cen} , N_{sat}

Predicting halo occupation in SAM

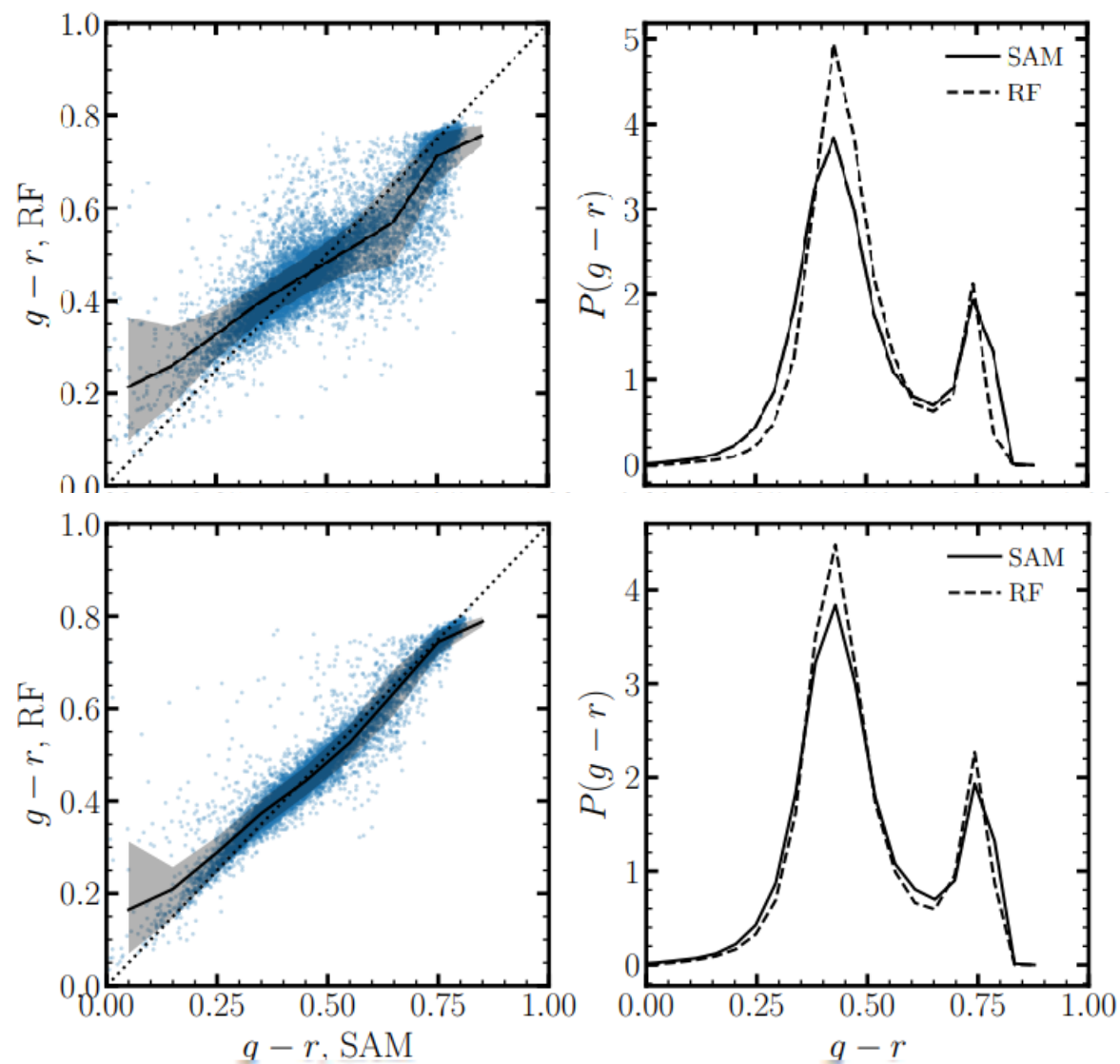


Predicting halo occupation in SAM



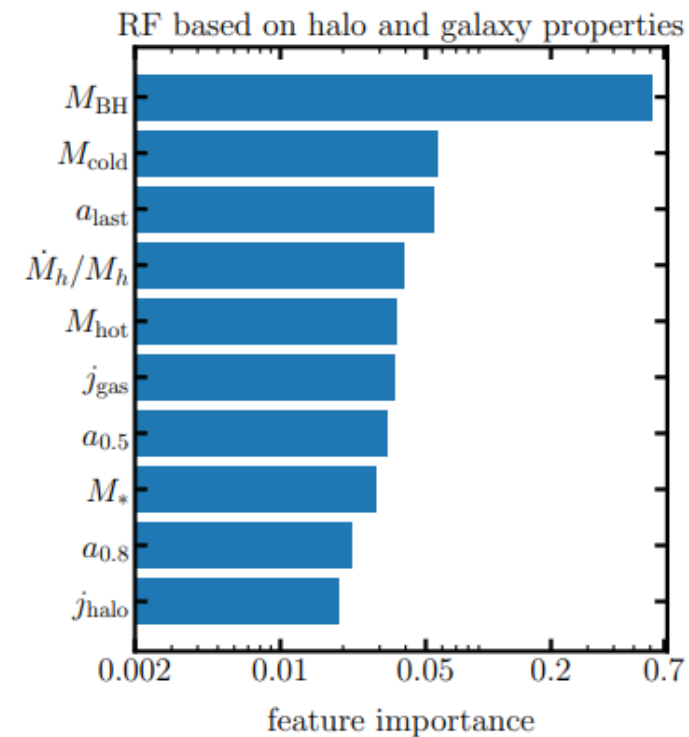
Predicting central galaxy property in SAM

Halo properties only



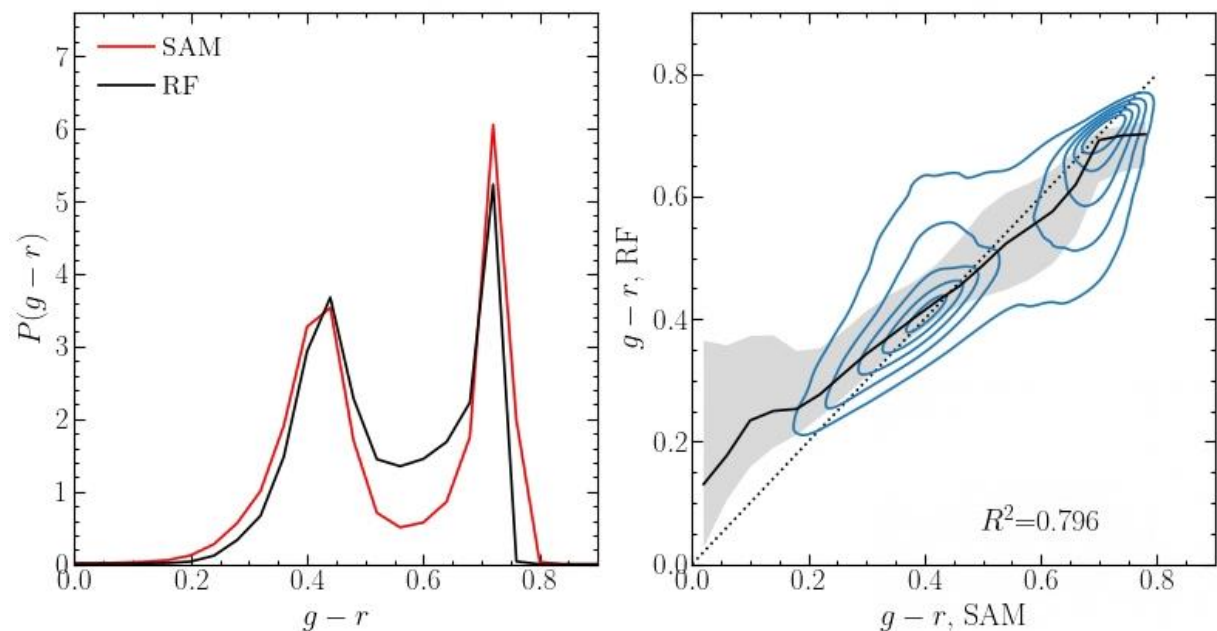
halo+galaxy properties

mass of central black hole, mass of cold gas, ...

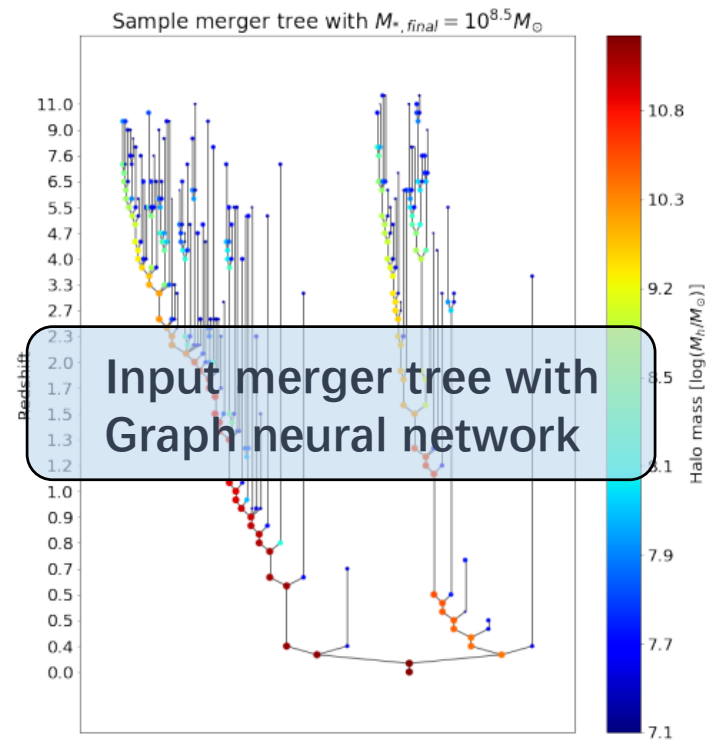
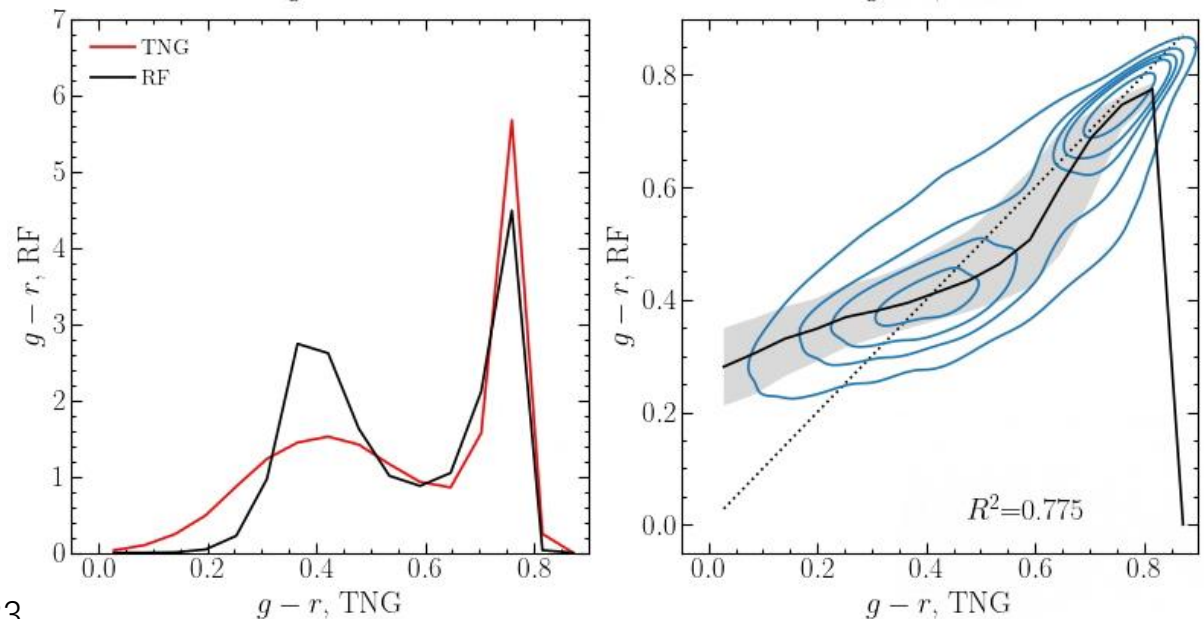


Predicting cen+sat property with subhalo

SAM

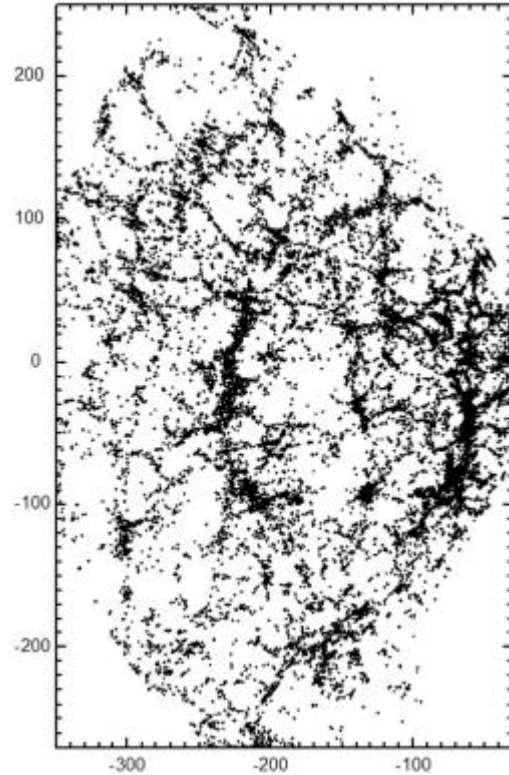


TNG



Target	Paper/Method	R^2 - score
<i>SFR</i>	Mangrove	0.876
	Final halo only	0.847
	A18	0.555
	JK19	0.760
	dS22	0.094
	L22	-
	SAM probabilistic limit	0.947

Can we predict galaxy properties in observation?

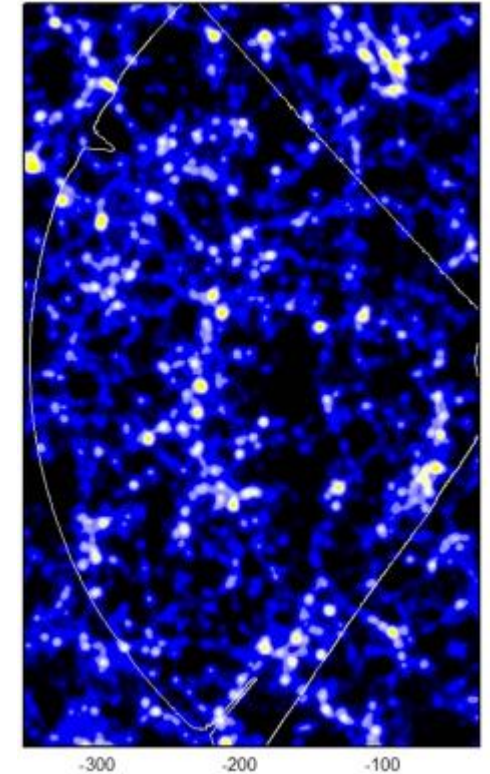


SDSS galaxies

ELUCID simulation

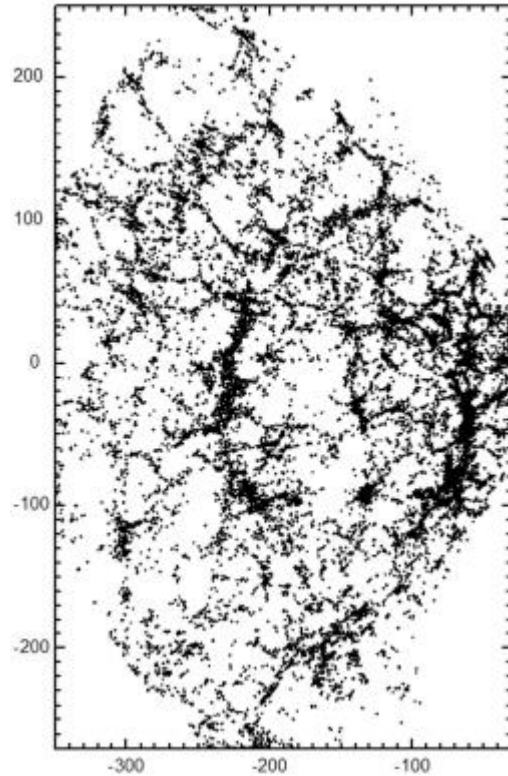
- Constrained N-body simulation constructed based on SDSS
- Recover large scale structures in SDSS

(Wang et al, 2014, 2016)



ELUCID subhalos

Can we predict galaxy properties in observation?

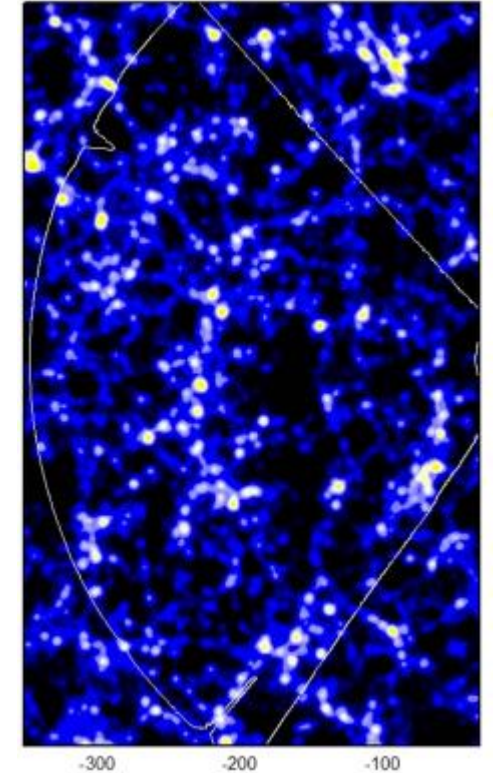


SDSS galaxies

neighborhood abundance matching (Yang2018)

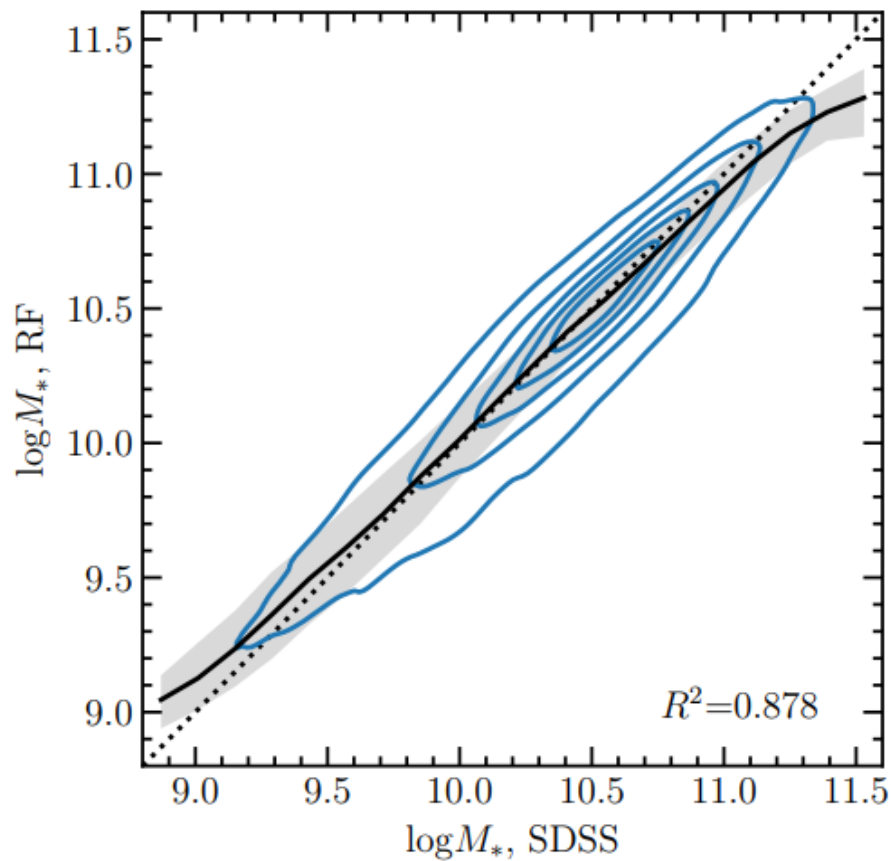
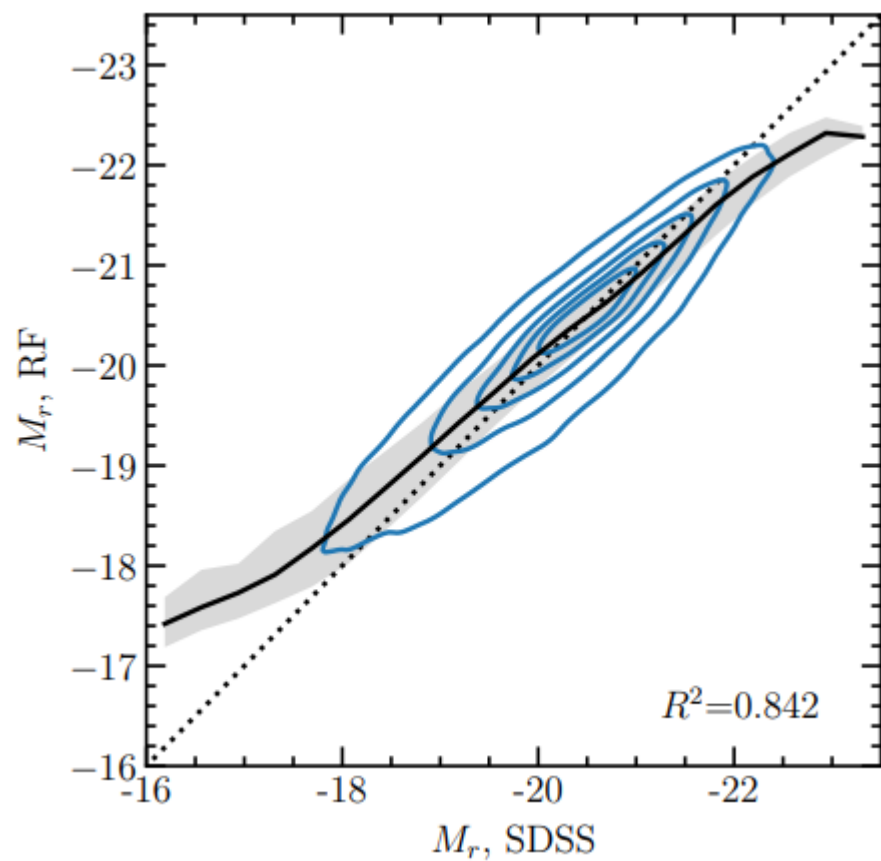


assigning SDSS galaxies to ELUCID subhalos
according to position and subhalo mass

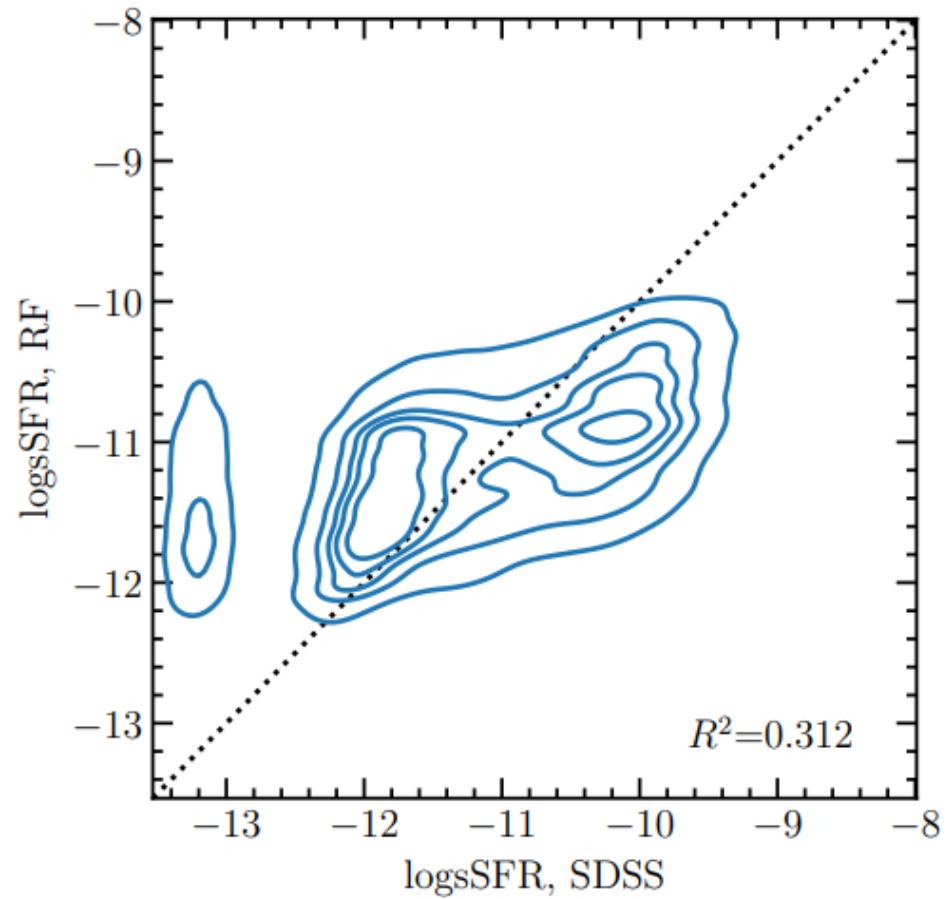
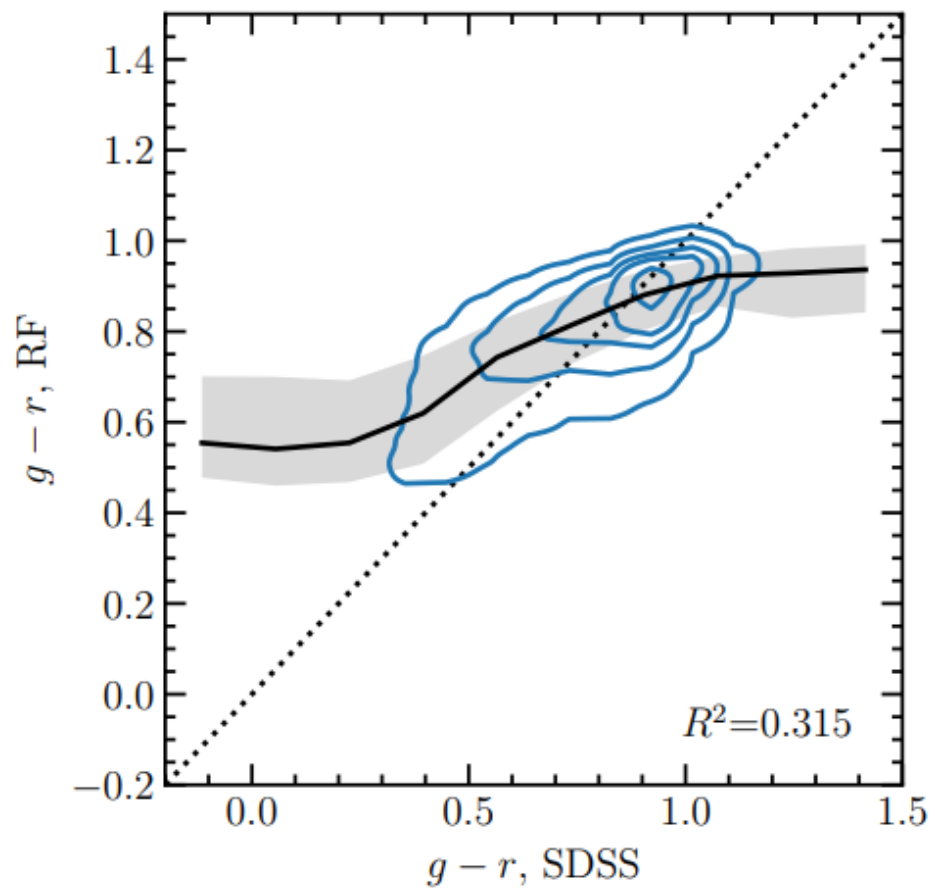


ELUCID subhalos

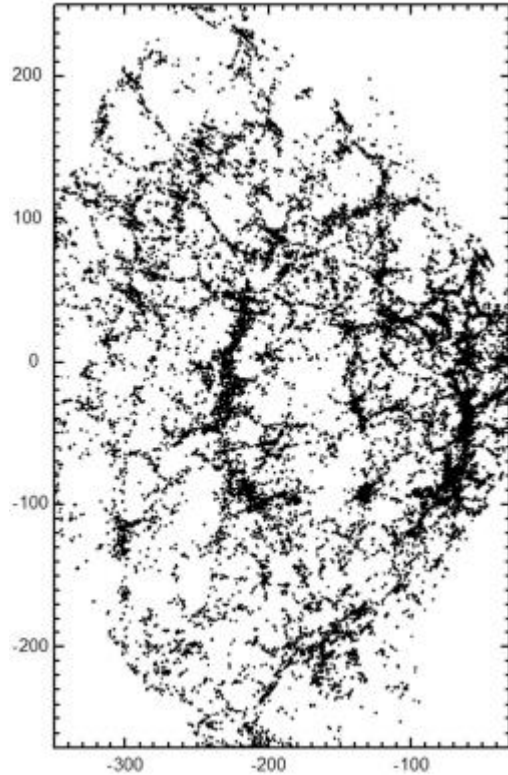
Predict galaxy properties in observation



Predict galaxy properties in observation



Predict galaxy properties in observation

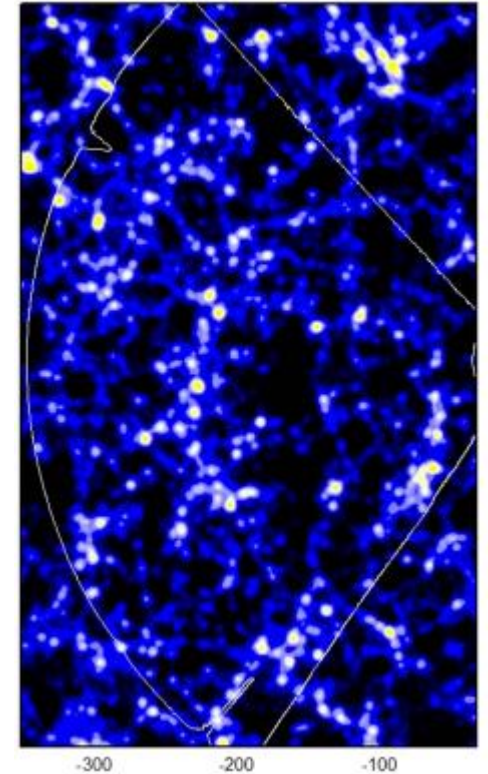


SDSS galaxies

neighborhood abundance matching (Yang2018)

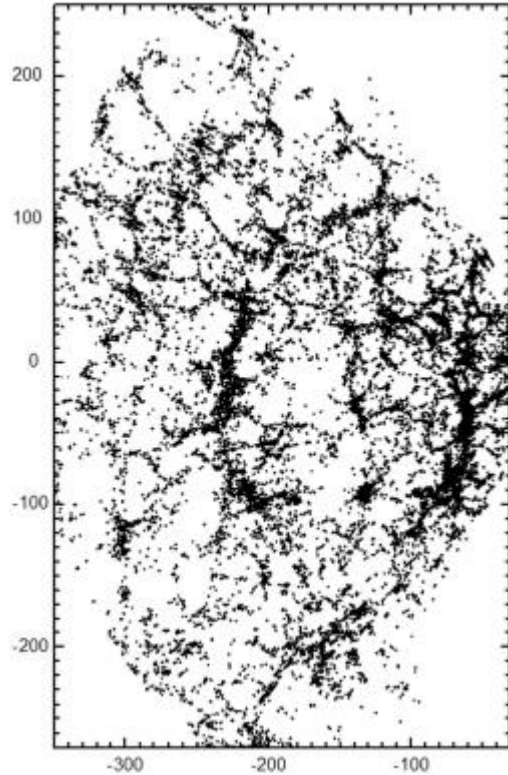


assigning SDSS galaxies to ELUCID subhalos
according to position and subhalo mass



ELUCID subhalos

Possible Problems

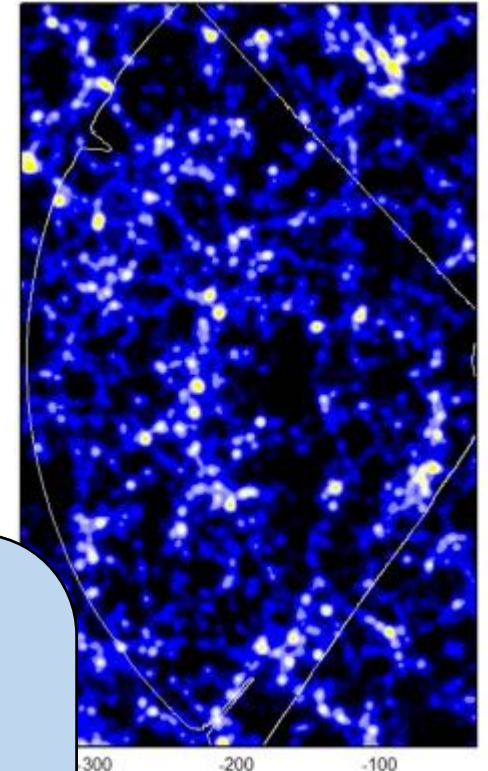


SDSS galaxies

neighborhood abundance matching (Yang2018)



assigning SDSS galaxies to ELUCID subhalos according to position and subhalo mass

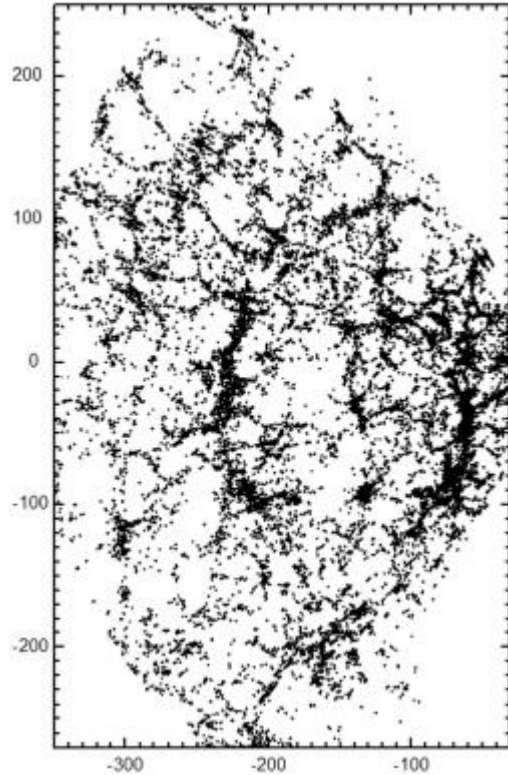


ELUCID subhalos

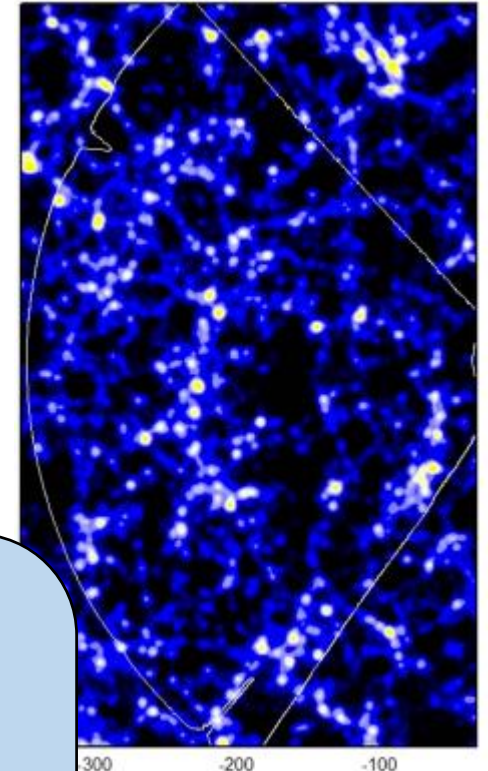
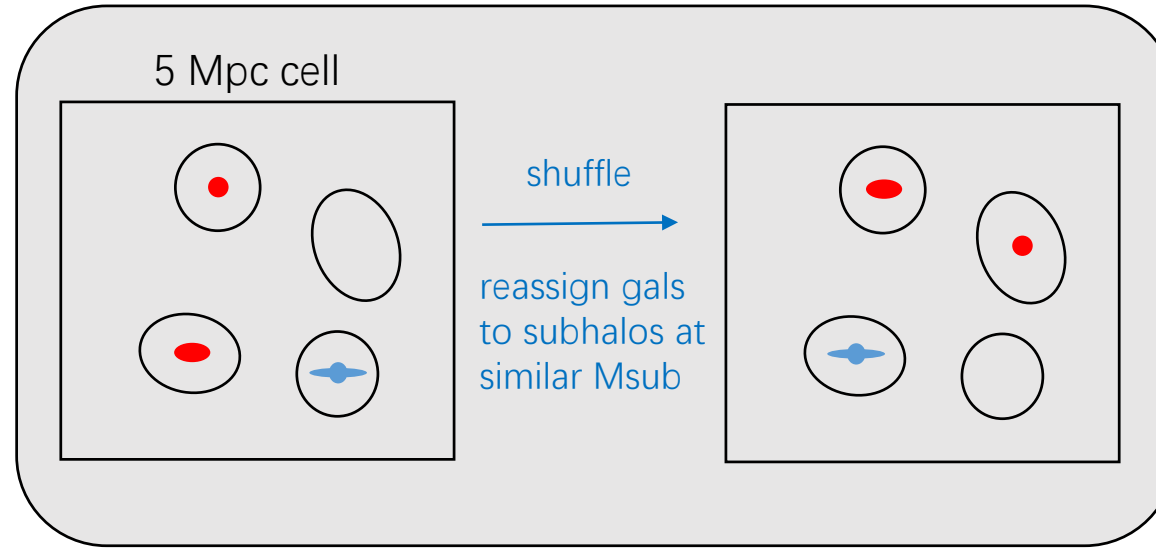
Possible problems

- ELUCID didn't recover the correct subhalos
- Galaxies are matched to wrong halos
- galaxy-halo connections may be wrong on some level
- noises in ML training sample

Possible Problems



SDSS galaxies

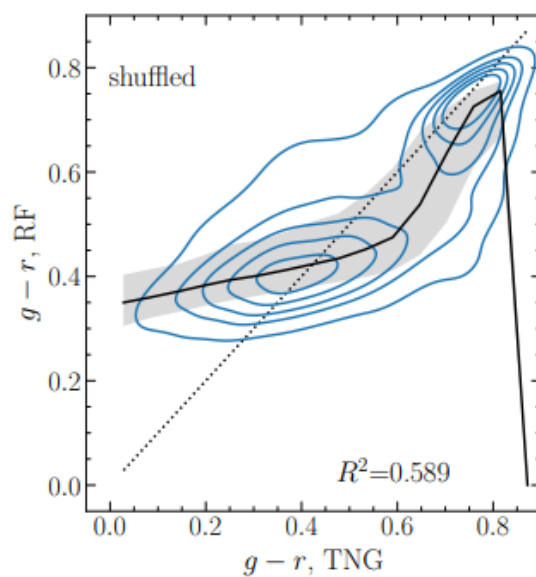
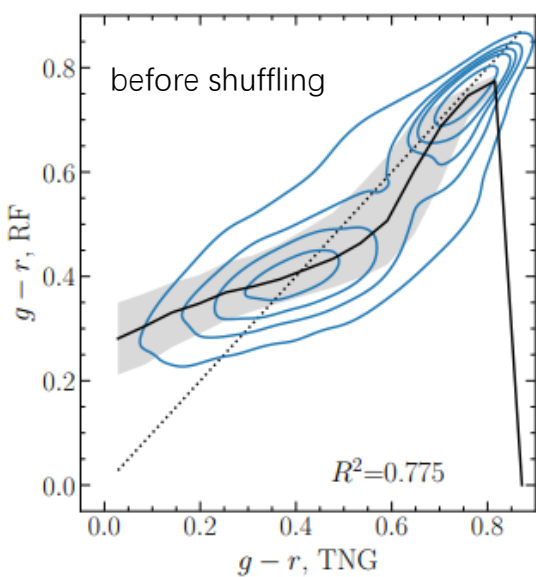
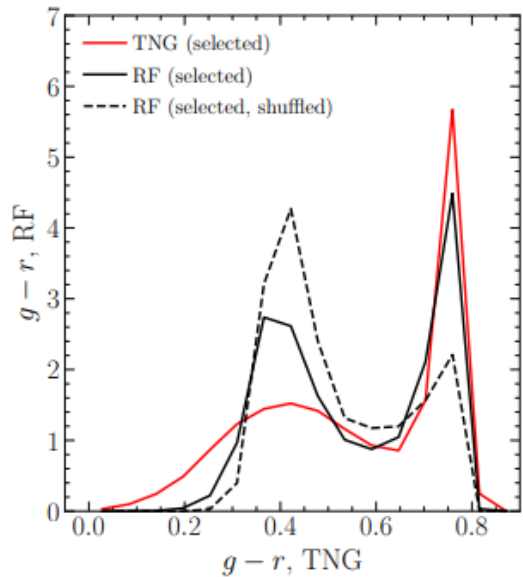
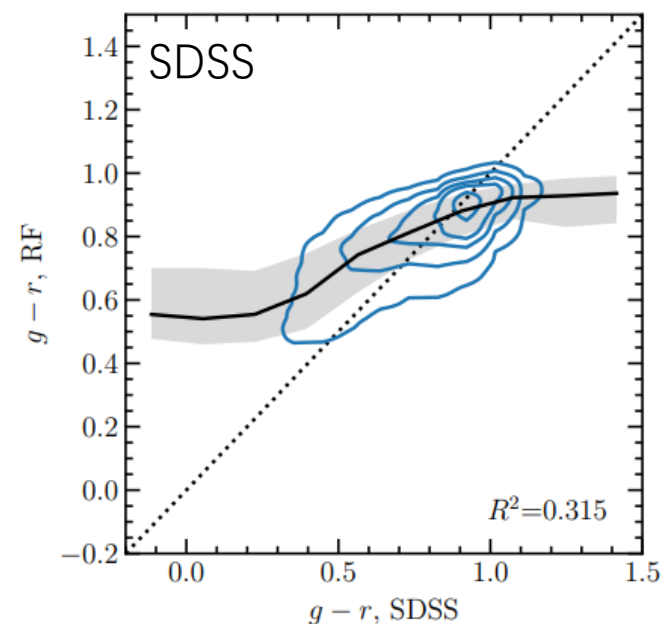
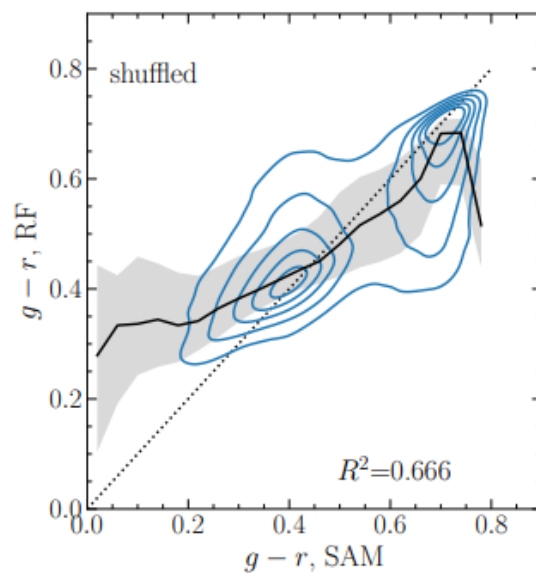
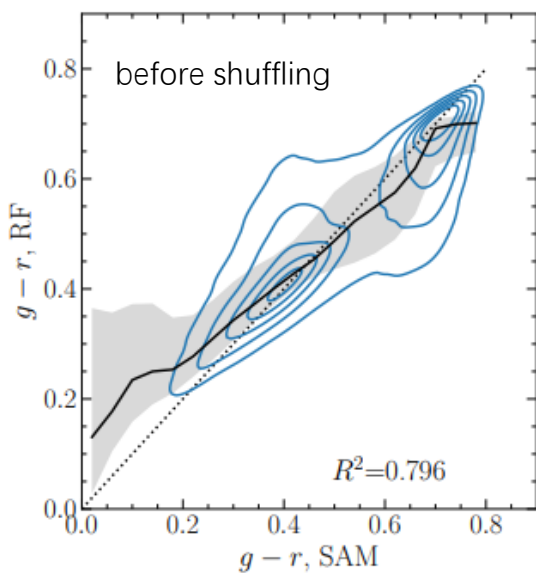
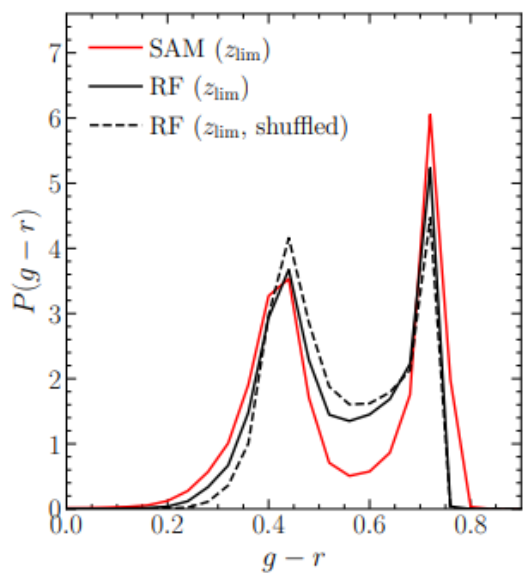


ELUCID subhalos

Possible problems

- ELUCID didn't recover the correct subhalos
- Galaxies are matched to wrong halos
- galaxy-halo connections may be wrong on some level
- noises in ML training sample

Predicting galaxy properties with mismatch



Prediction performance is lower



Color-subhalo connection is lower

Summary

- in SAM and hydrodynamic simulation:
 - RF can predict halo occupation or magnitude with high accuracy
 - RF can predict color with relatively high accuracy
- in SDSS
 - RF can predict magnitude or stellar mass with relatively high accuracy
 - Prediction accuracy of color is low
- The color-subhalo connection in SDSS may be much lower than those in simulated galaxies