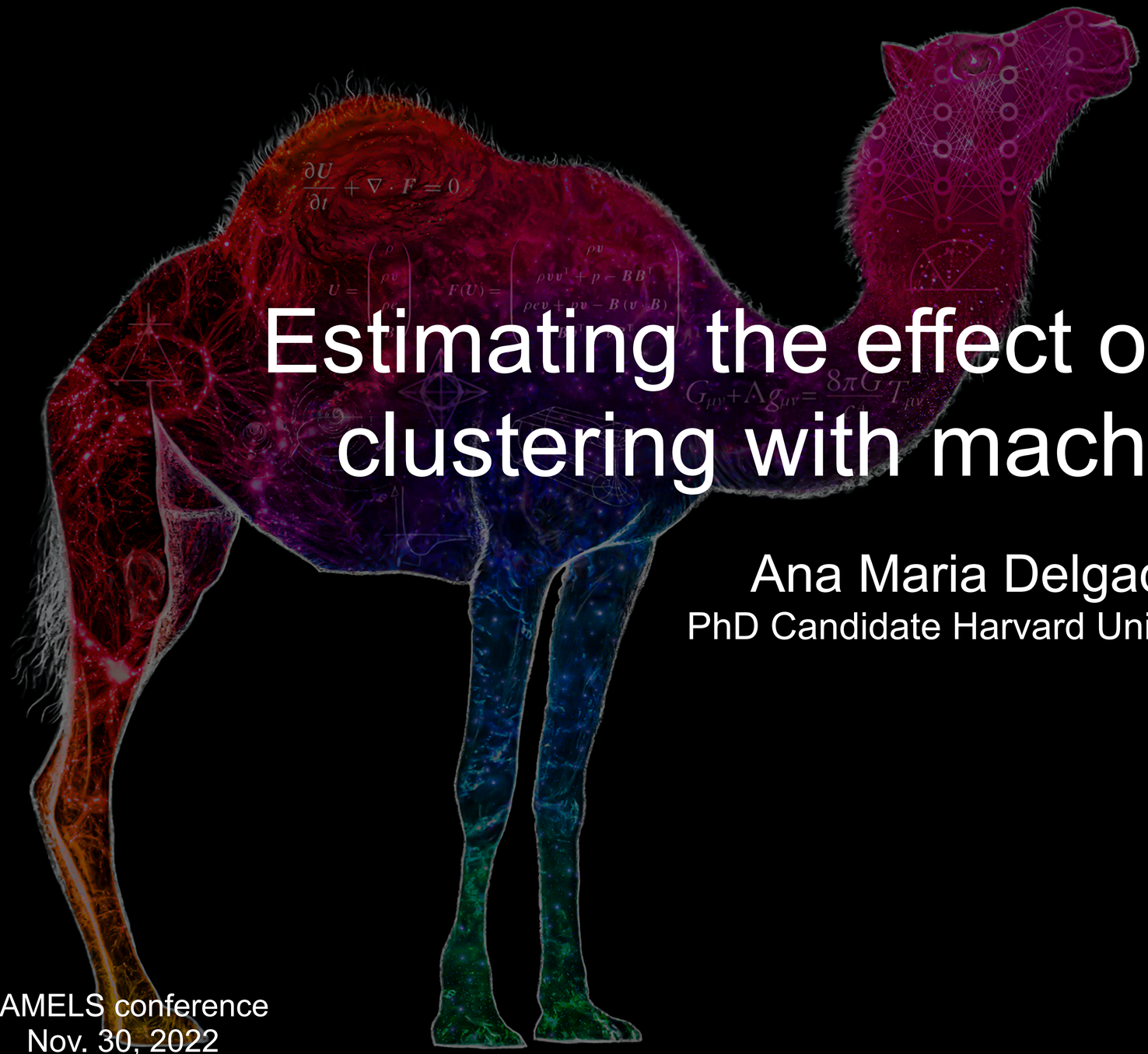


Estimating the effect of feedback on clustering with machine learning

Ana Maria Delgado
PhD Candidate Harvard University

CAMELS conference
Nov. 30, 2022

CENTER FOR ASTROPHYSICS
HARVARD & SMITHSONIAN

Motivation

Overarching Idea

In order to fully realize the statistical power of upcoming surveys, we need a precise understanding of the distribution of matter.

Motivation

Overarching Idea

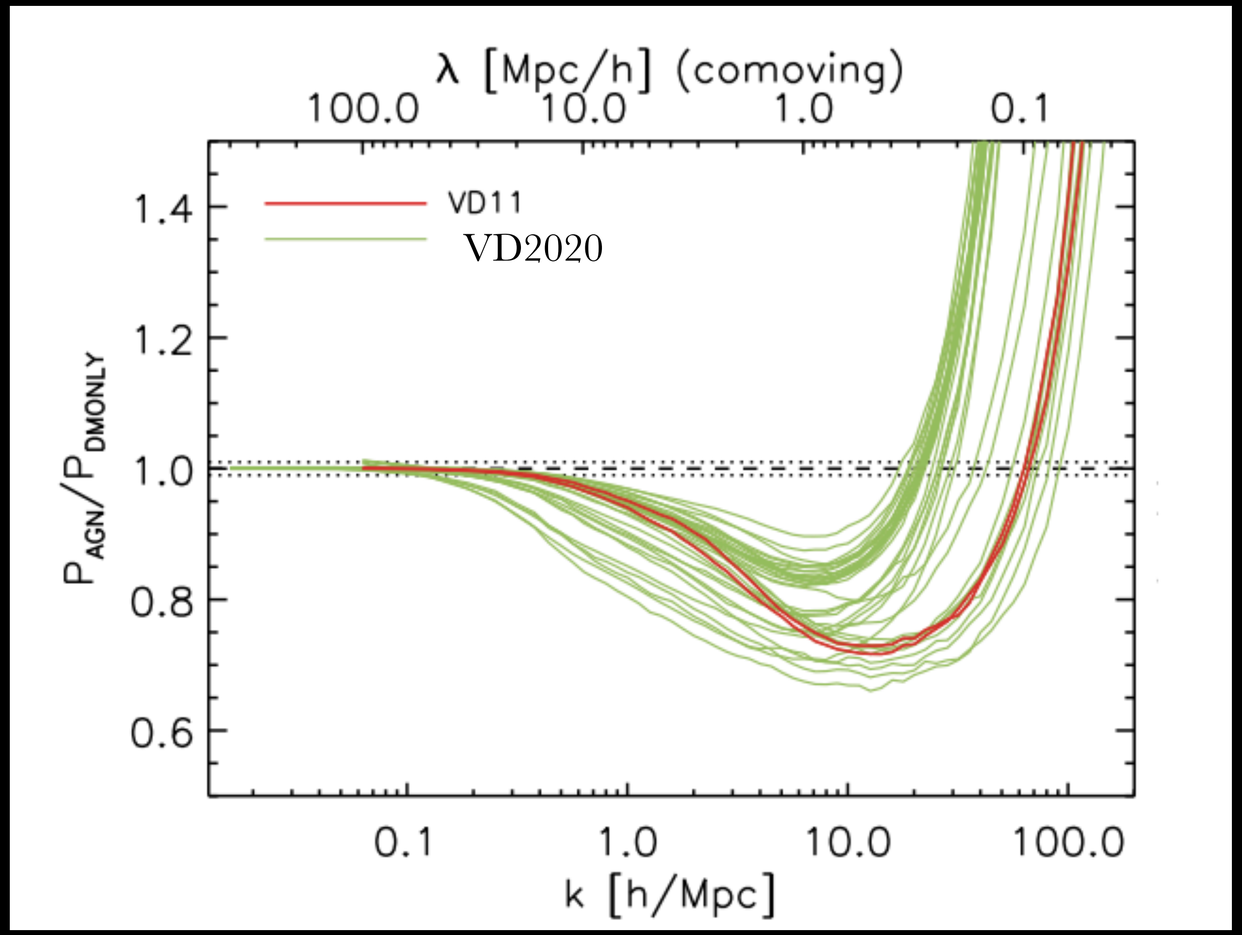
In order to fully realize the statistical power of upcoming surveys, we need a precise understanding of the distribution of matter.

total matter power spectrum

Motivation

Feedback processes make this challenging

Baryonic effects impact the matter power spectrum



van Daalen, McCarthy, Schaye 2020

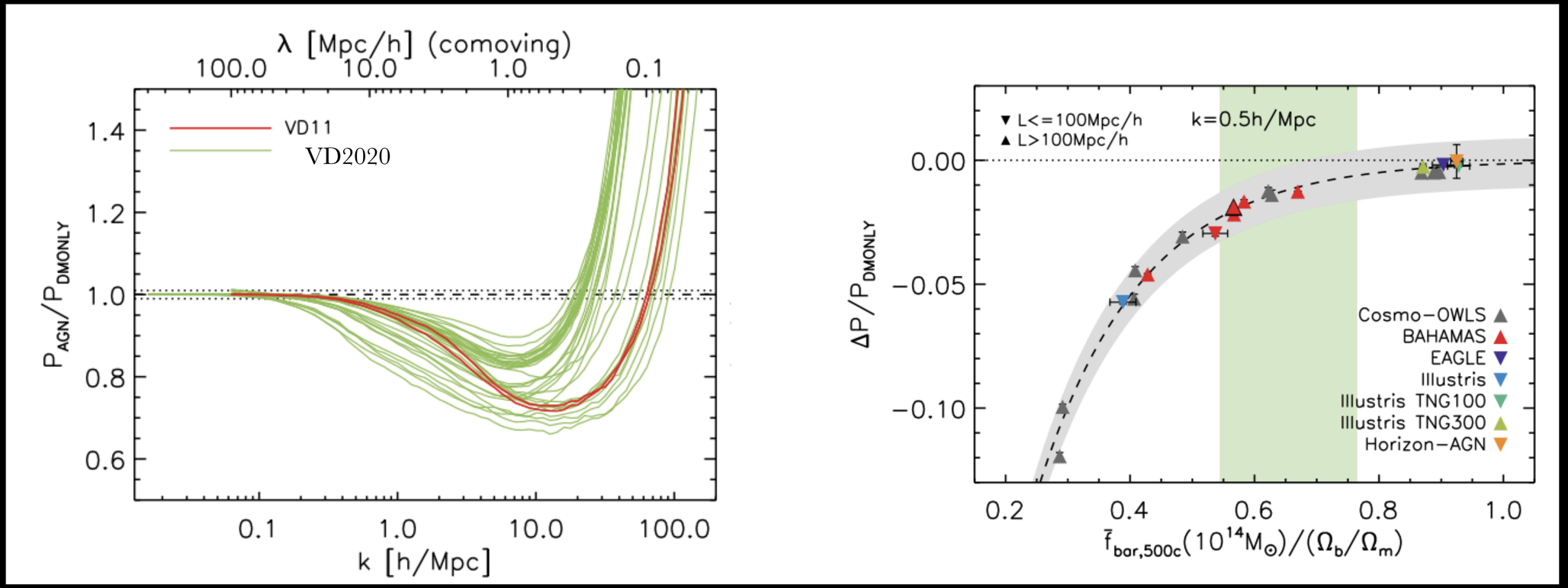
Motivation

$$\frac{\Delta P}{P_{\mathrm{DM}}} = \frac{P_{\mathrm{hydro}} - P_{\mathrm{DM}}}{P_{\mathrm{DM}}}$$

Relative difference between the total matter power spectrum of hydrodynamical and DM simulations

Motivation

$\Delta P/P_{\rm DM}$ as a function of mean baryon fraction of high-mass halos



van Daalen, McCarthy, Schaye 2020

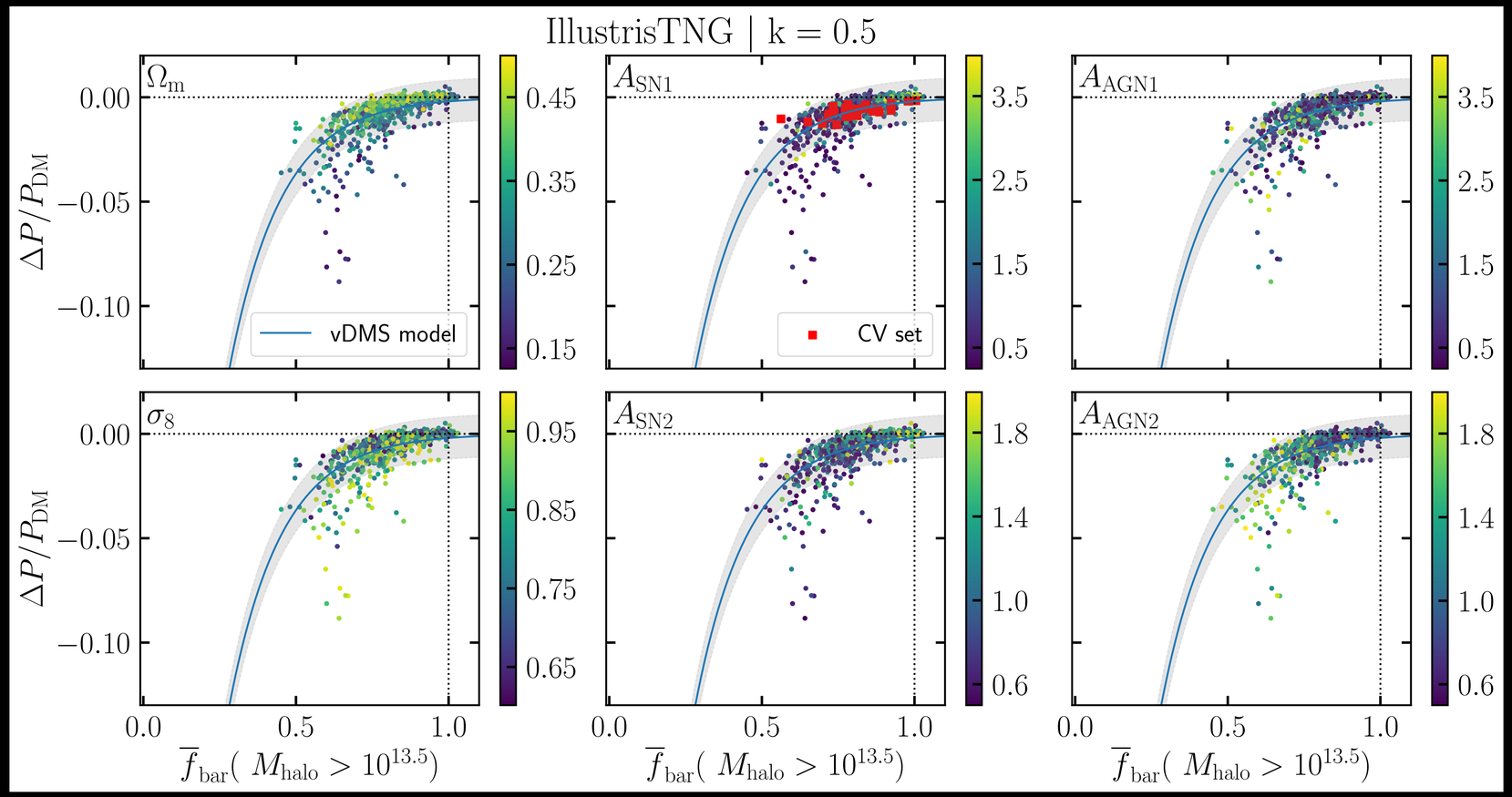
Motivation

CAN WE…

- Examine the robustness of this method over a broader range of cosmology and feedback parameters
- Extract information from a full range of halo masses
- Probe the non-linear regime

Motivation

CAMELS

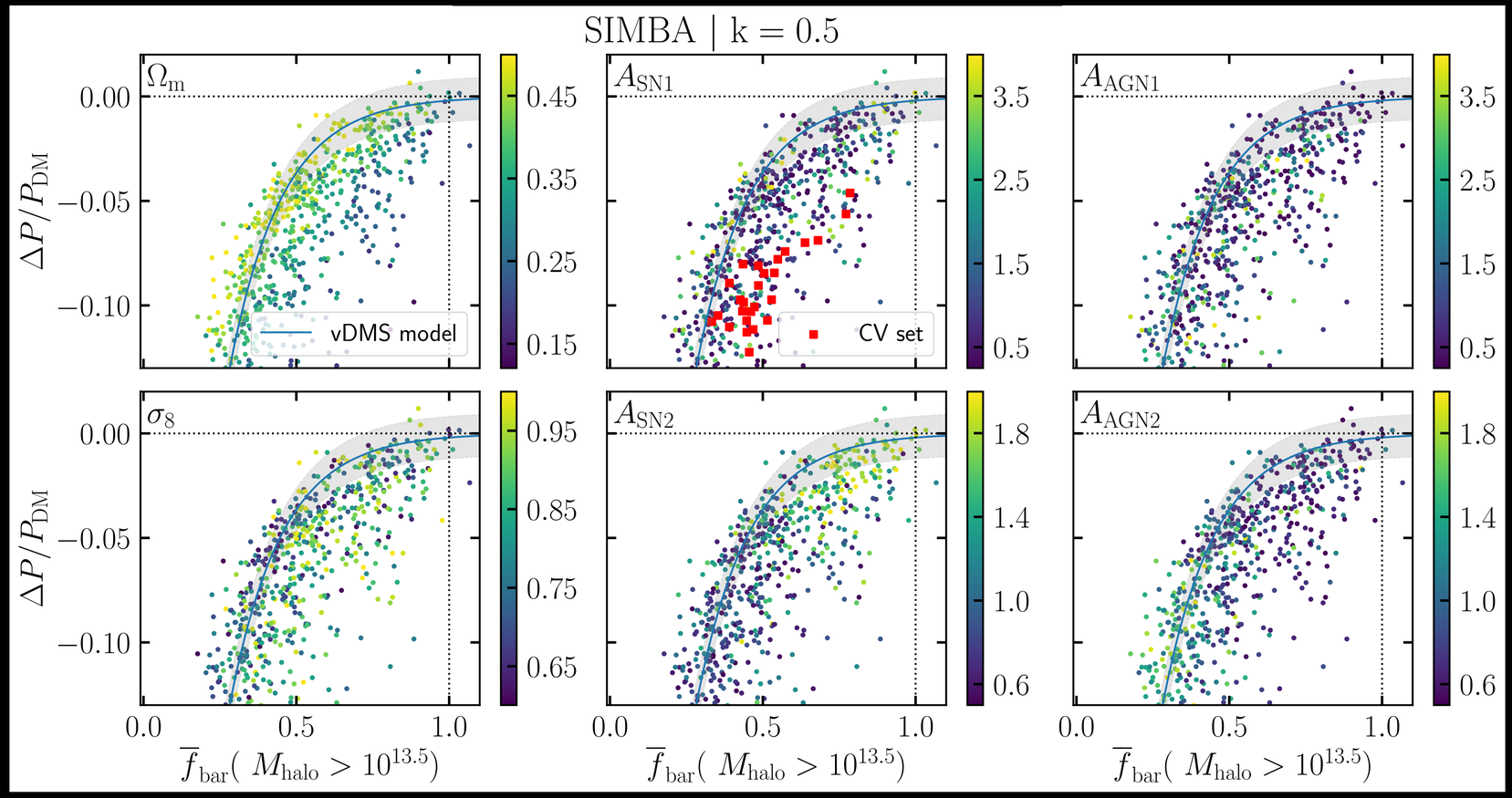


Delgado+ (in prep)

** all masses are in [h^{-1} $M_\odot$] ; all k modes are in [h Mpc^{-1}]

Motivation

CAMELS



Delgado+ (in prep)

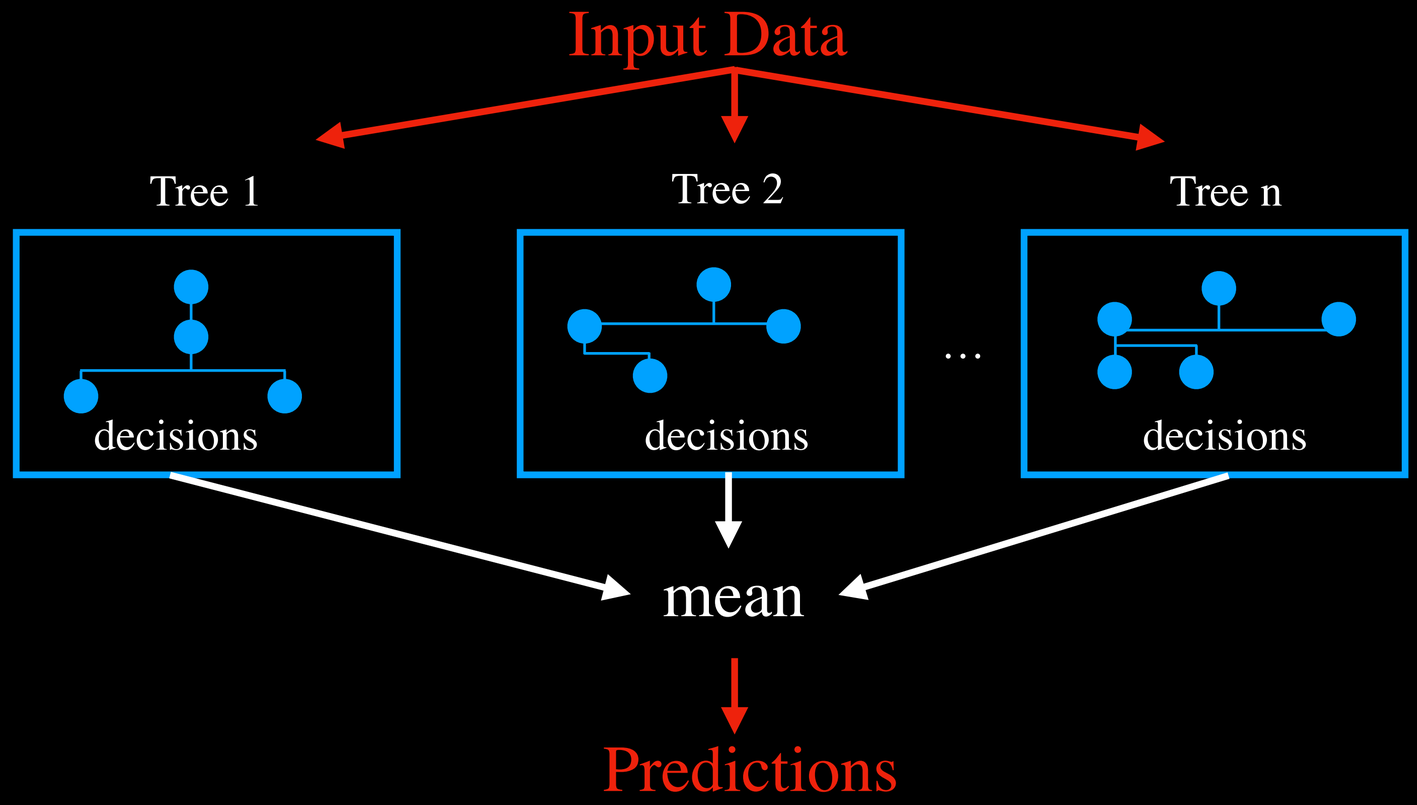
** all masses are in [h^{-1} $M_\odot$] ; all k modes are in [h Mpc^{-1}]

Methods

Random Forest Regressor (RF)

Advantages:

- Little hyperparameter tuning is needed
- Computationally efficient
- Ensemble characteristic lessens overfitting
- Some interpretability (feature importance)



Methods

Training Features

$\bar{f}_{\mathrm{bar}}(M_{\mathrm{halo}} > 13.5)$ ⟶ mean baryon fraction of high-mass halos

N_{halo} ⟶ number of halos

$\bar{f}_{\mathrm{bar}}(M^{j}_{\mathrm{halo}})$ ⟶ mean baryon fraction for a range of halo masses

N^{j}_{halo} ⟶ number of halos per halo mass bin

Methods

Training Features

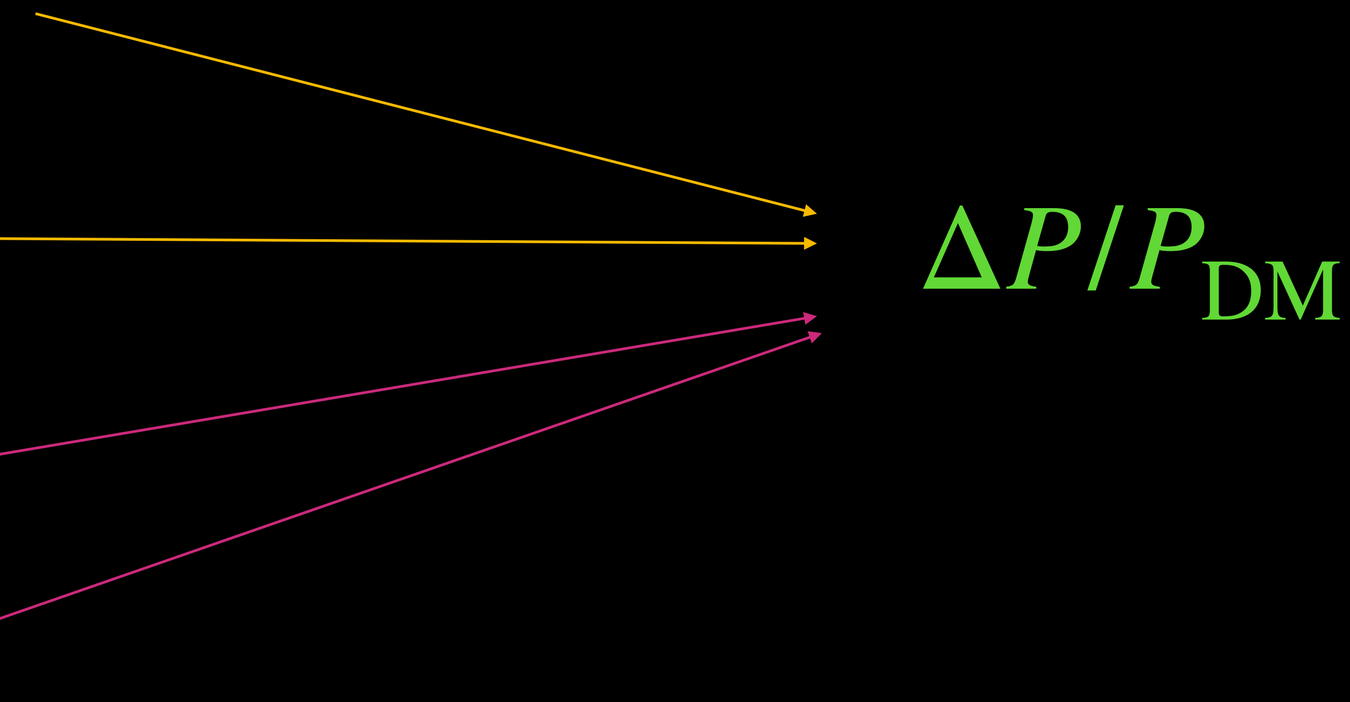
$\bar{f}_{\rm bar}(M_{\rm halo} > 13.5)$

$N_{\rm halo}$

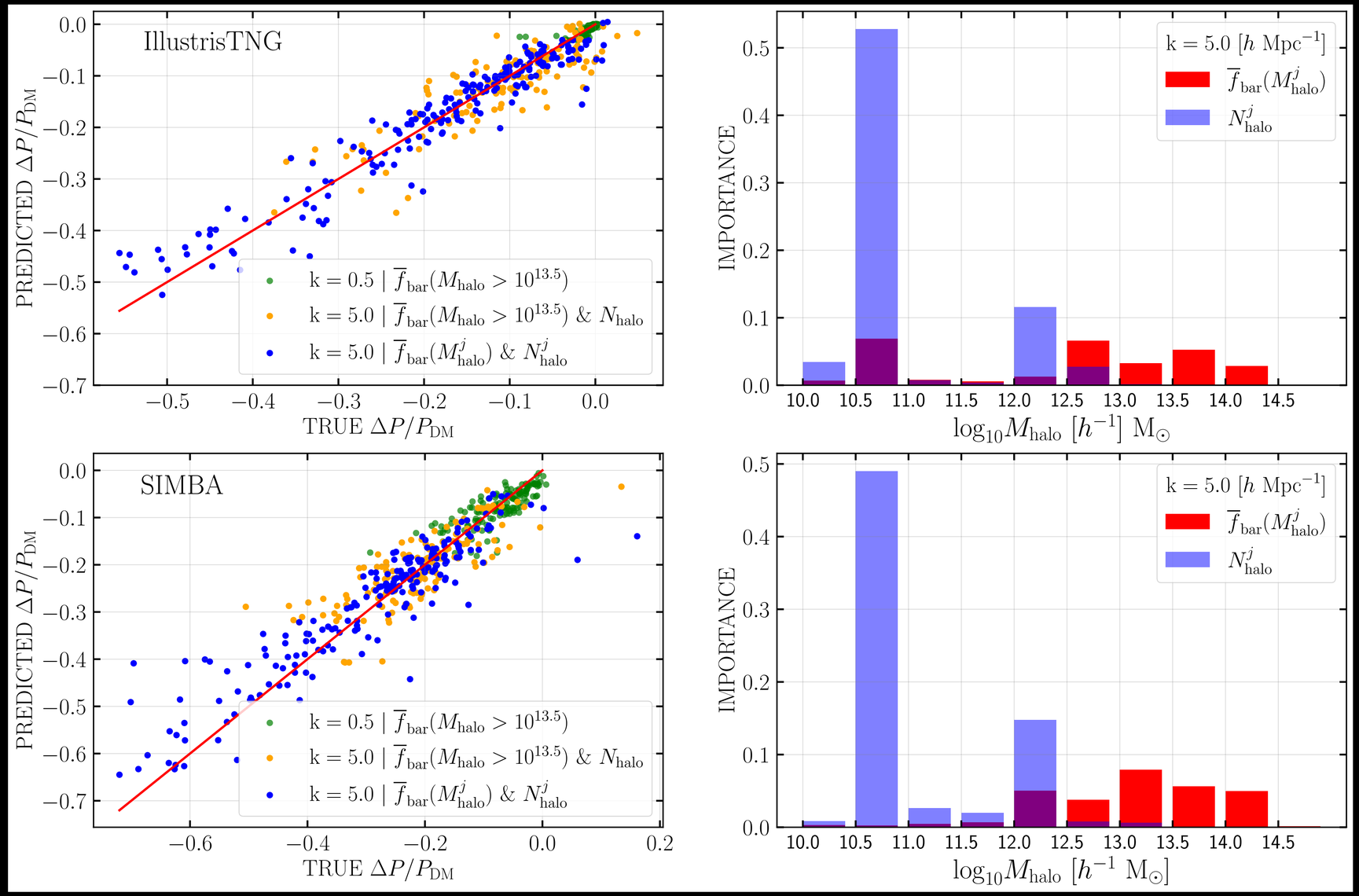
$\bar{f}_{\rm bar}(M^j_{\rm halo})$

$N^j_{\rm halo}$

Target



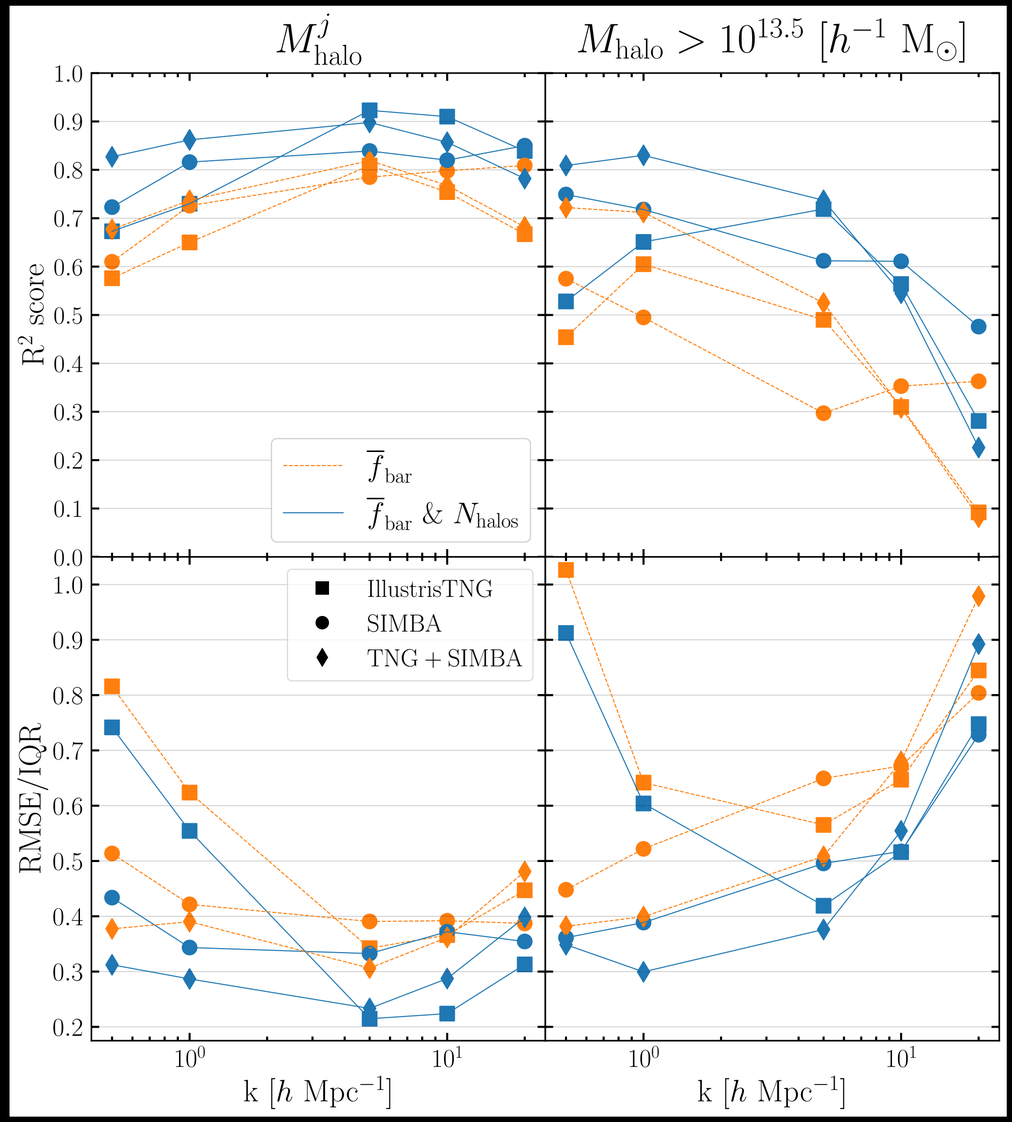
Results



Delgado+ (in prep)

Results

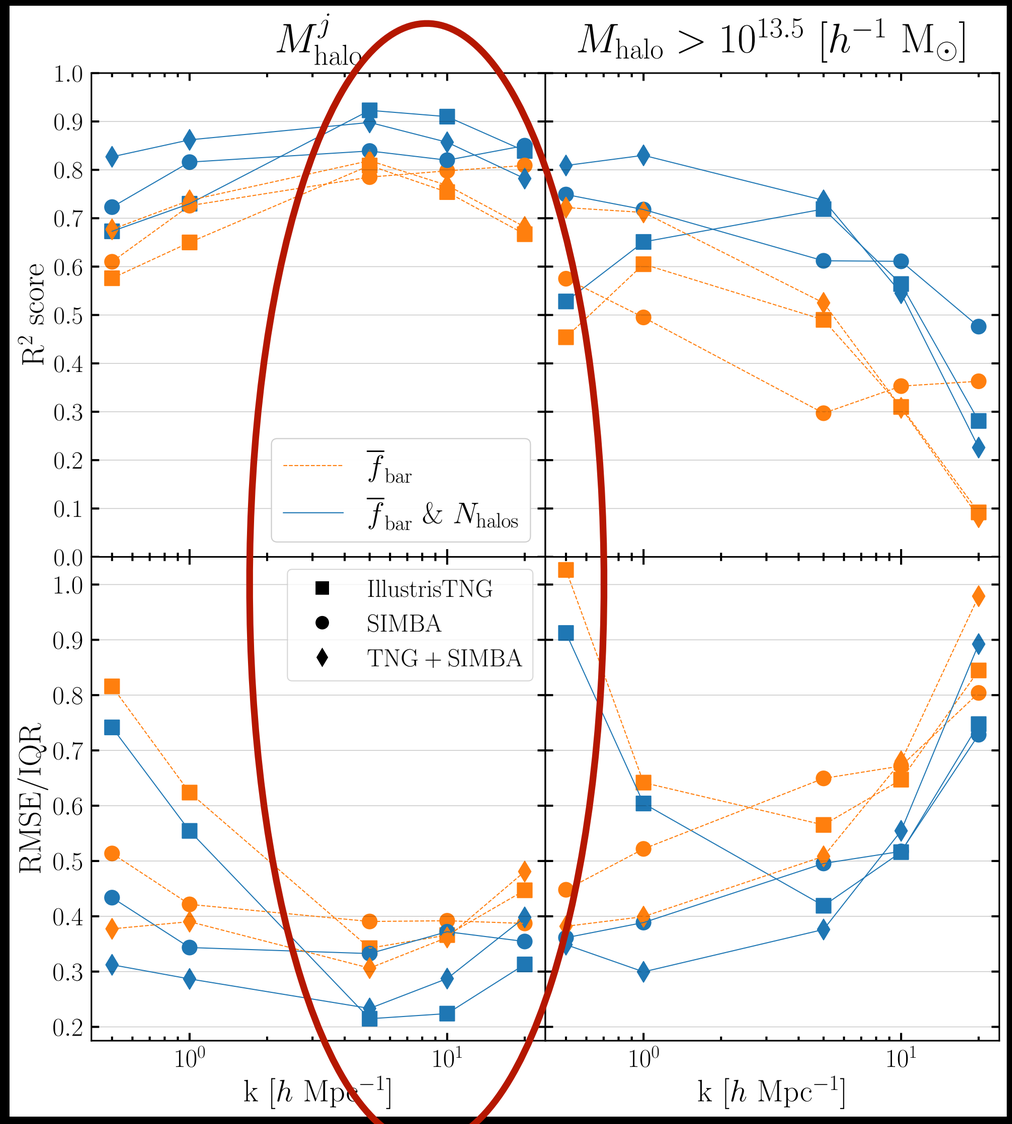
We are able to extract information from a range of halo masses in the non-linear regime to $k = 20.0\ h\ \mathrm{Mpc}^{-1}$



Delgado+ (in prep)

Results

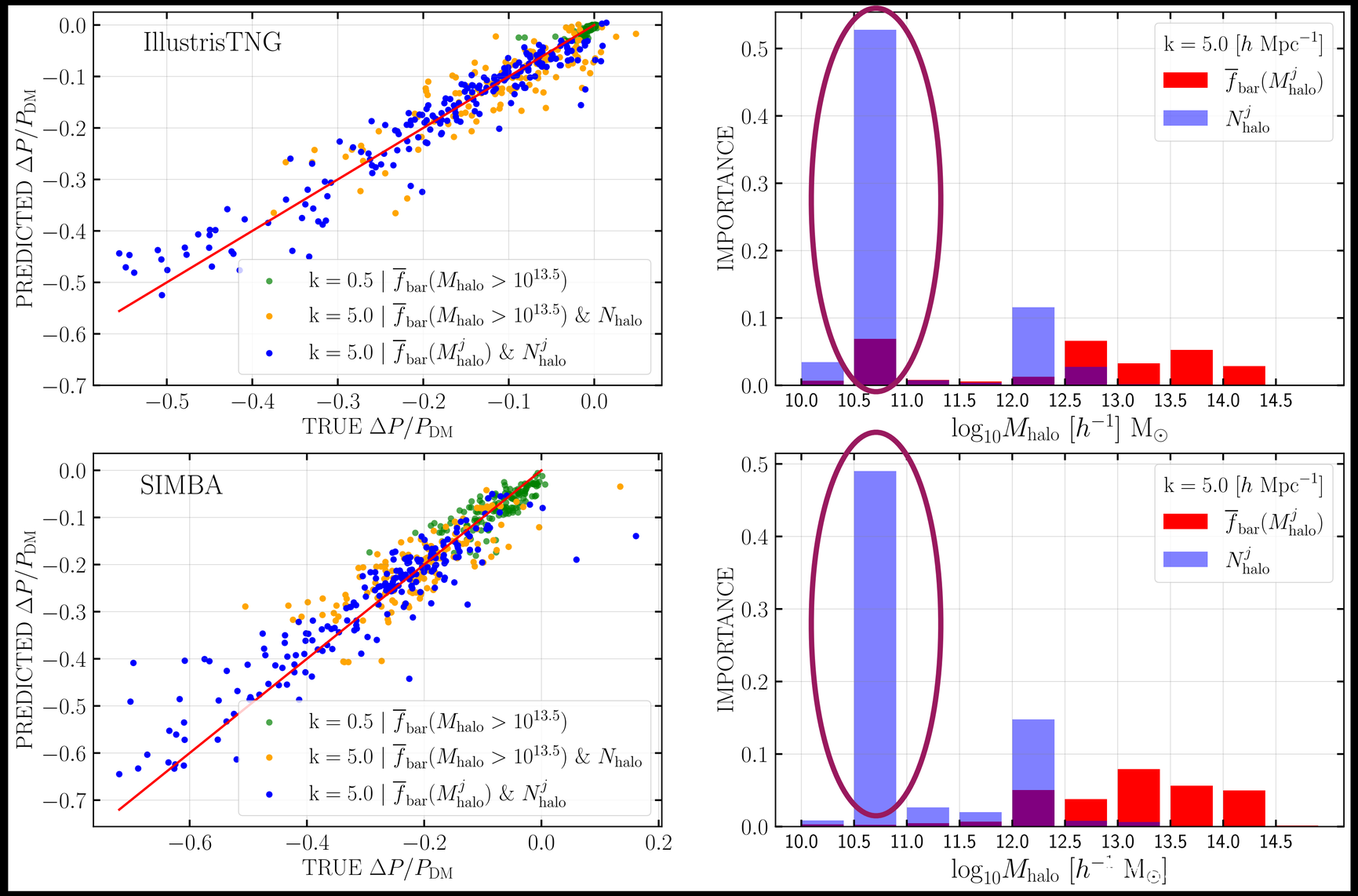
We are able to extract information from a range of halo masses in the non-linear regime to $k = 20.0\ h\ \mathrm{Mpc}^{-1}$



Delgado+ (in prep)

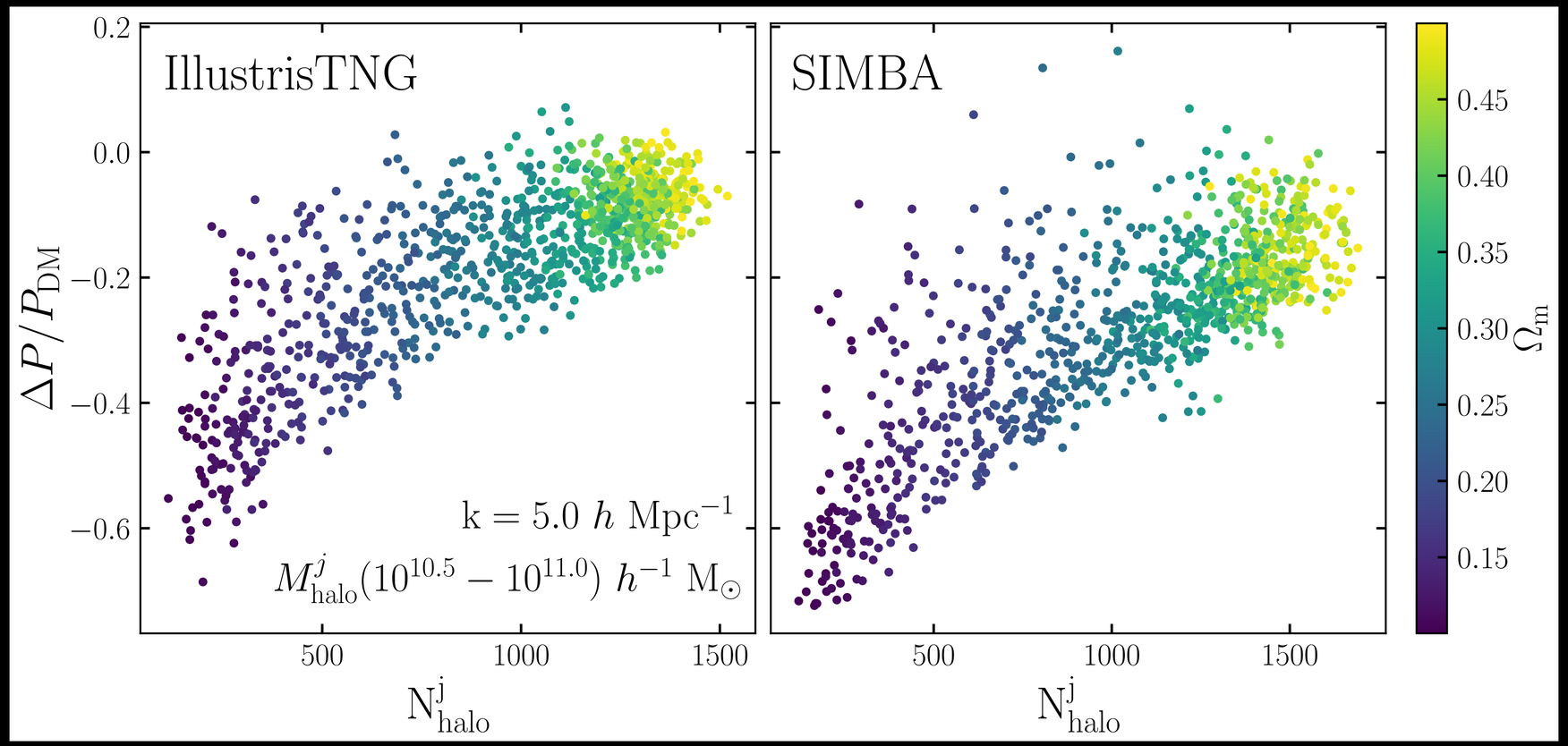
Interpretation

RF reported $N^j_{\rm halo}$ at $10^{10.5} \leq M_{\rm halo} < 10^{11.0}$ as highest ranked feature



Interpretation

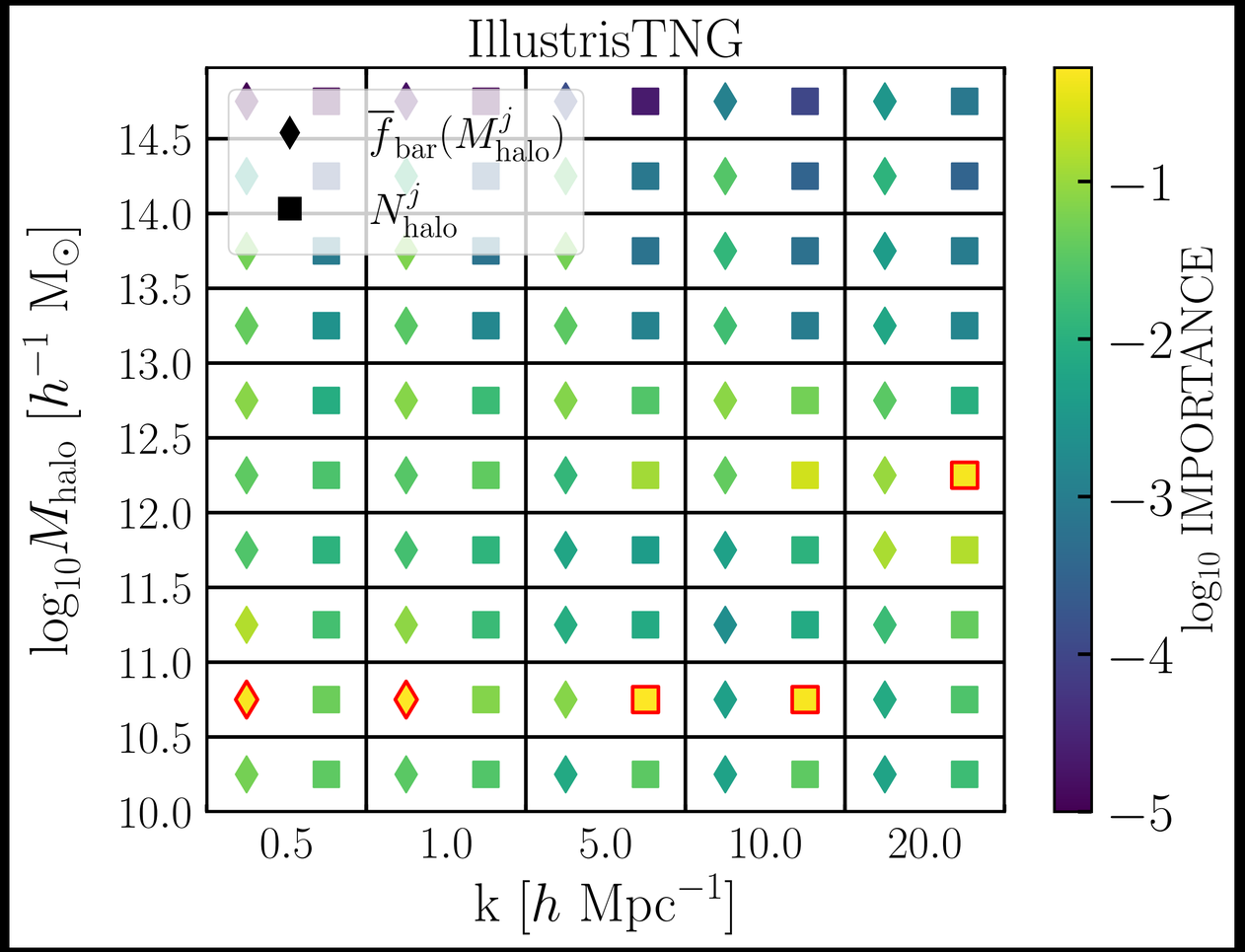
The highest ranked features by the RF are strong tracers of Ω_m



Delgado+ (in prep)

Interpretation

Examining the feature importance reported by the RF per k-mode, there is no clear trend for highest ranked feature.

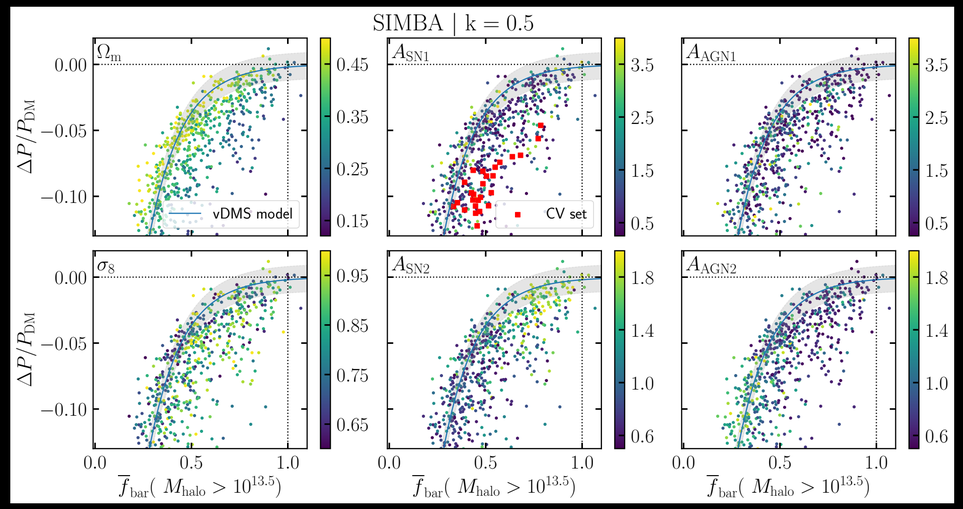


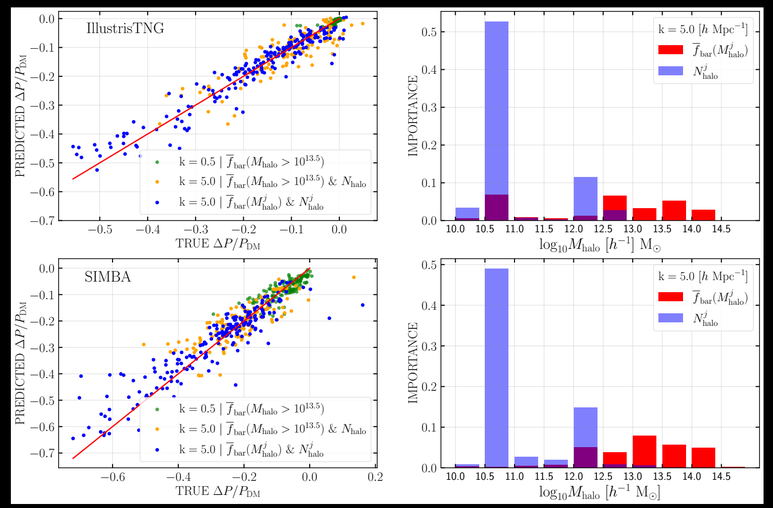
Delgado+ (in prep)

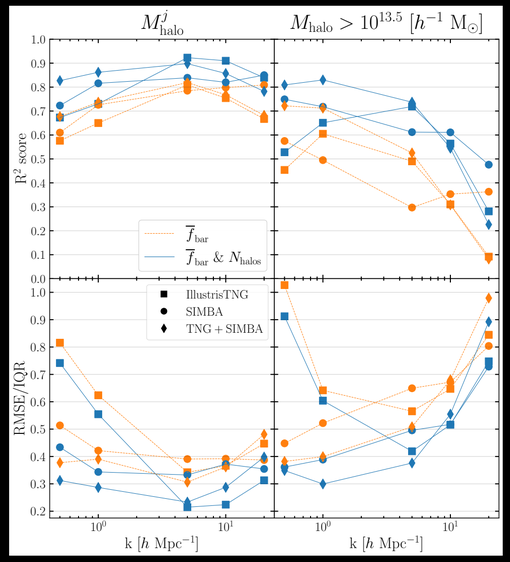
Many Thanks.

- CAMELS generally agree with previous works: $\Delta P/P_{\rm DM}$ can be modeled as a function of mean baryon fraction.
    - models that can generalize to broader set of feedback parameters are needed

- A RF can extract information about $\Delta P/P_{\rm DM}$ from broad range of halo masses and in the non-linear regimes to estimate a more general model

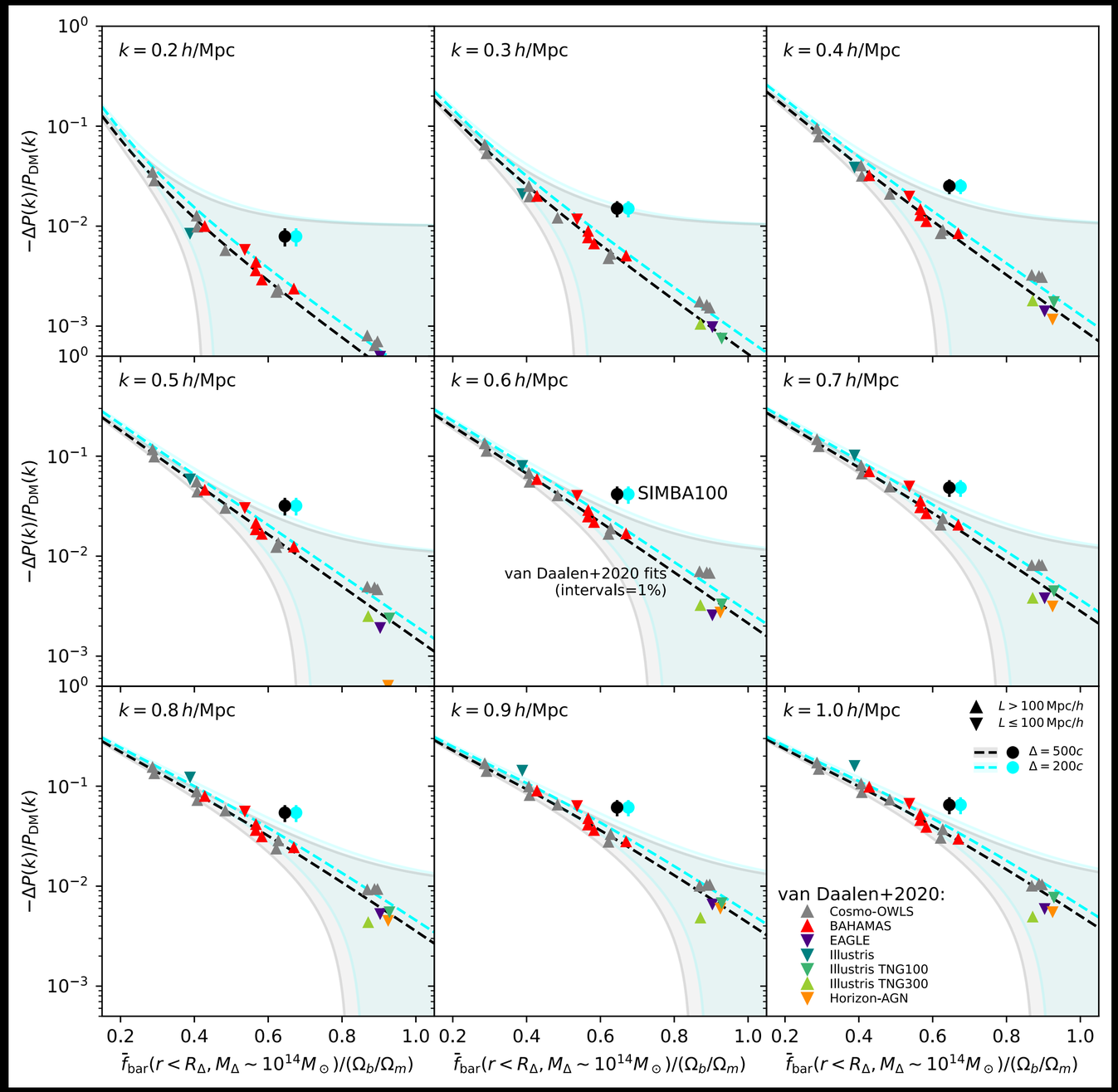






Supplementary

Leander Thiele



Delgado+ (in prep)