The CAMELS project

Francisco Villaescusa-Navarro

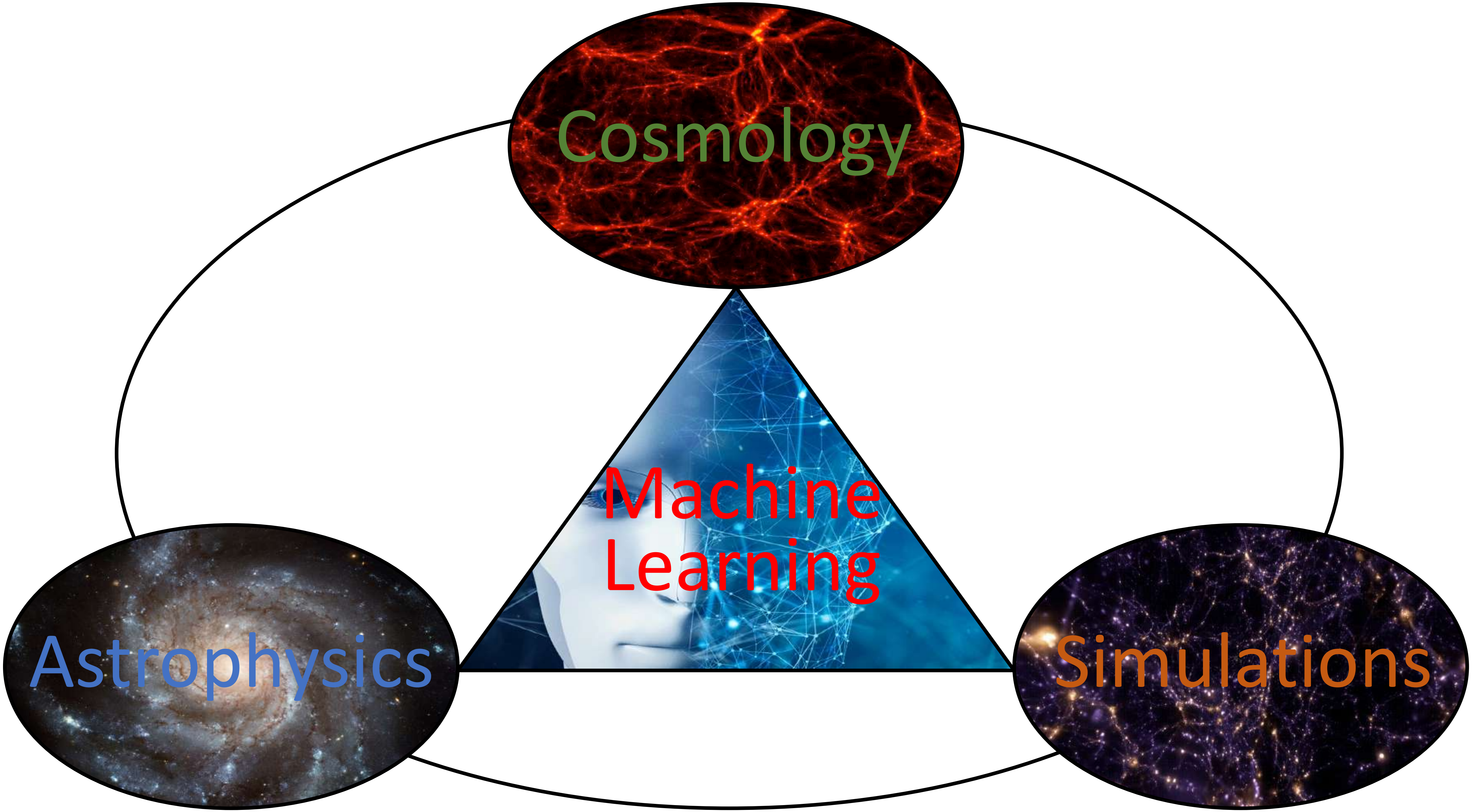
SIMONS FOUNDATION
FLATIRON INSTITUTE
PRINCETON UNIVERSITY

Berkeley

April 12th 2022

Outline

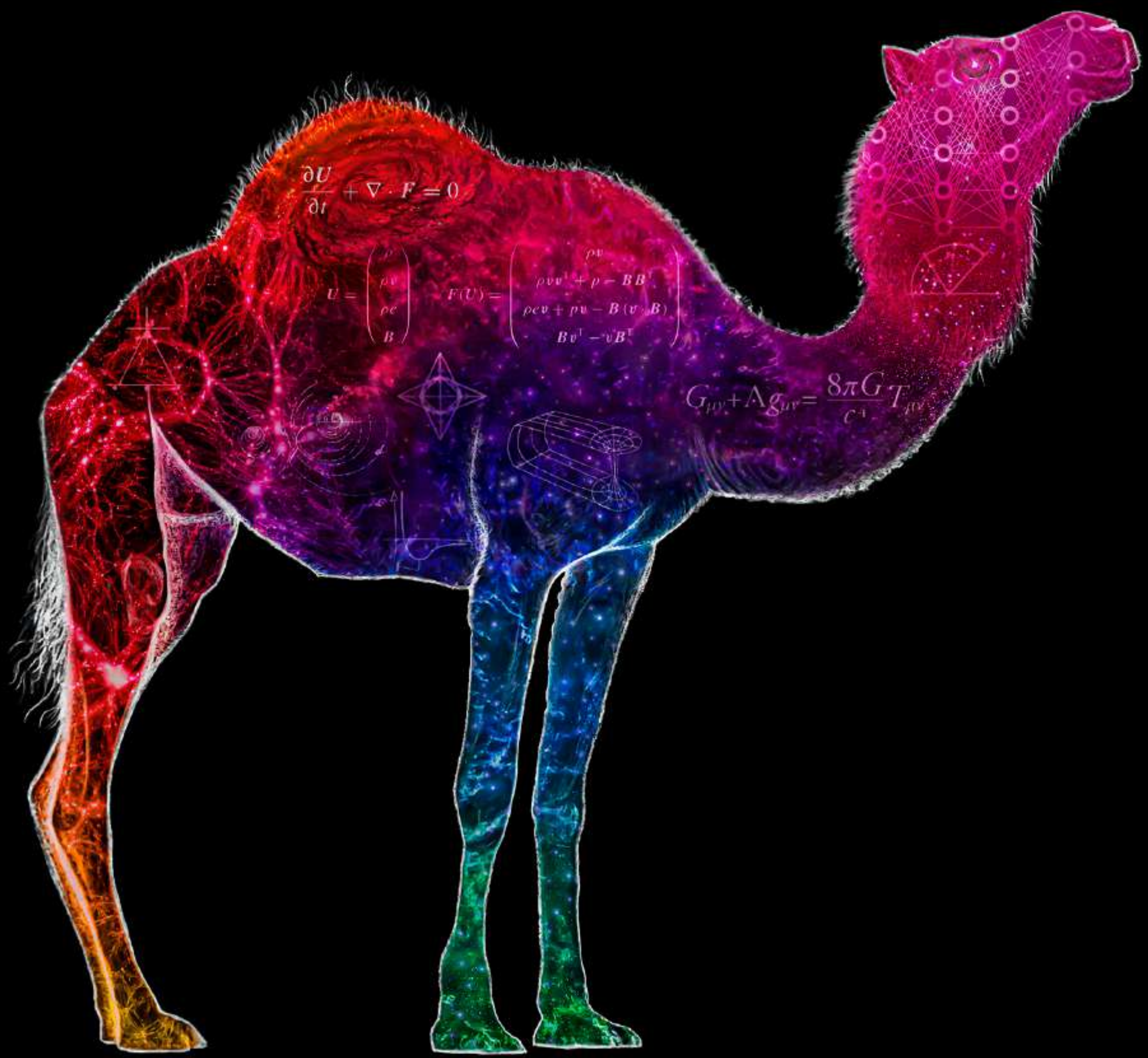
1. Motivation & Description
2. Results
3. Conclusions


Cosmology
Machine
Learning
Astrophysics
Simulations

CAMELS

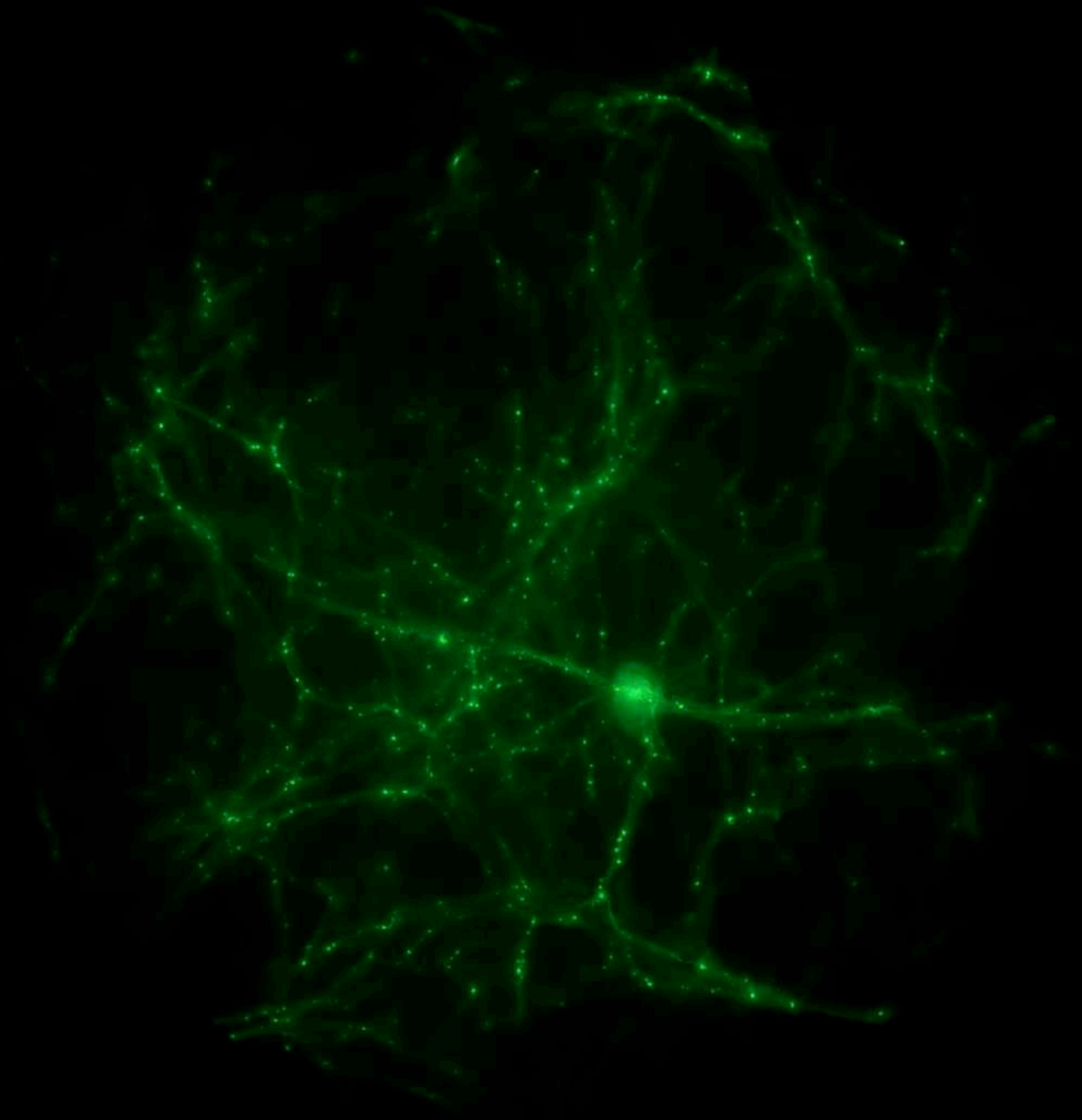
https://www.camel-simulations.org

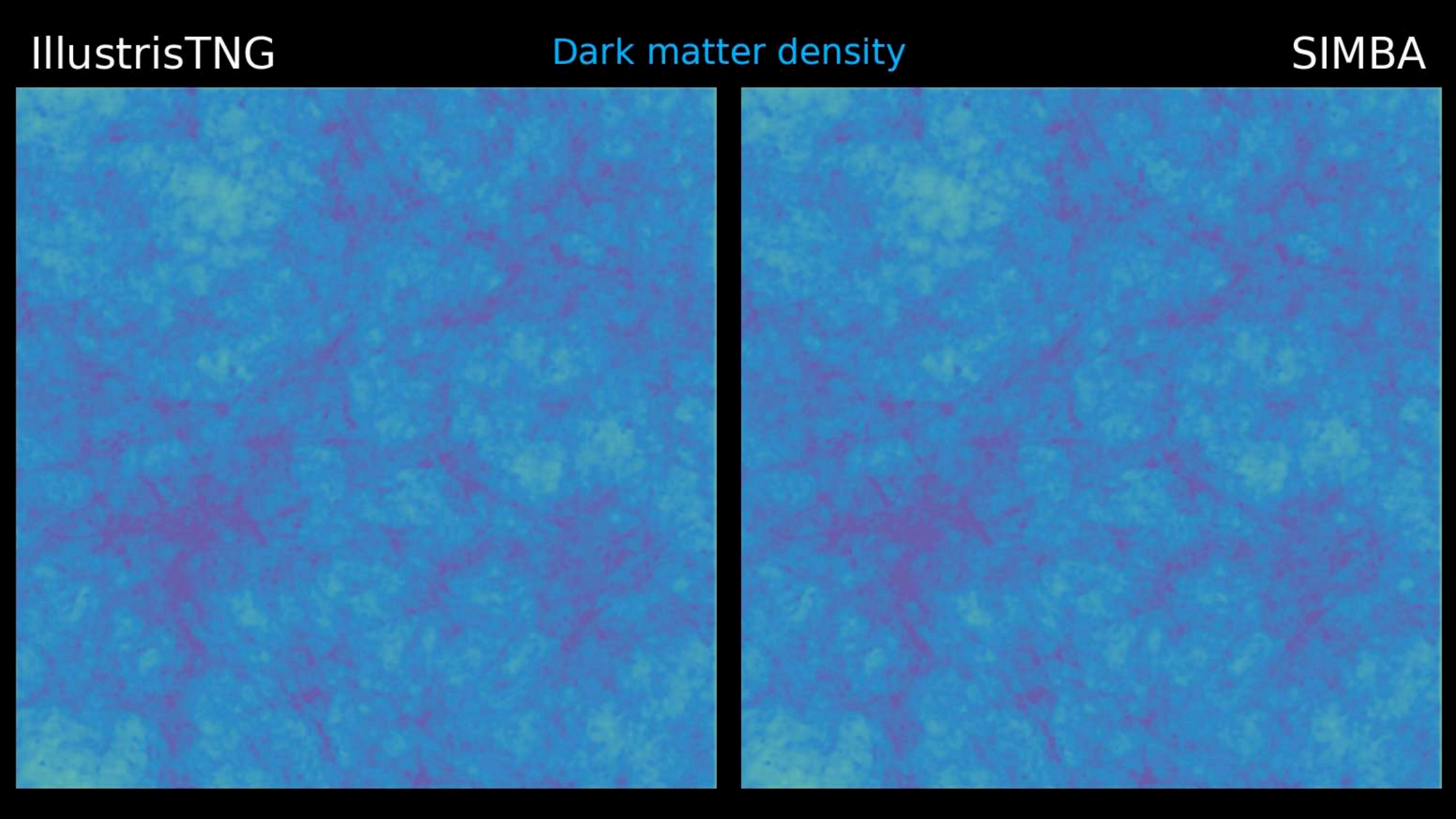
Cosmology and Astrophysics with MachinE Learning Simulations

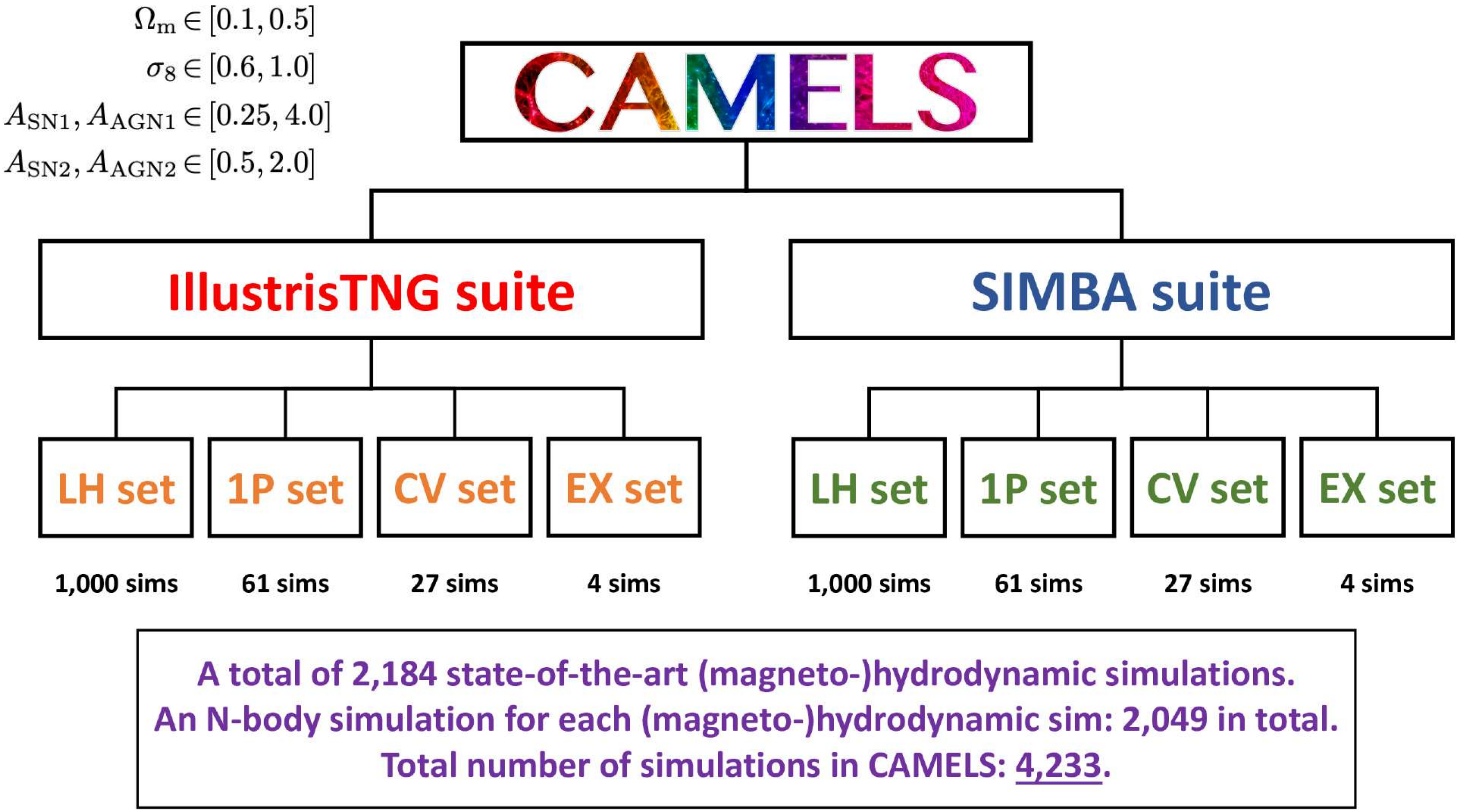


- A suite of 4,233 simulations
- 2,049 N-body; Gadget-III
- 2,184 state-of-the-art (magneto-)hydrodynamic sims
- AREPO/IllustrisTNG + GIZMO/SIMBA
- 6 parameters: $\{\Omega_m, \sigma_8, A_{\mathrm{SN1}}, A_{\mathrm{SN2}}, A_{\mathrm{AGN1}}, A_{\mathrm{AGN2}}\}$
- More than 2,000 cosmologies & astrophysics models; more than 140,000 snapshots
- Designed for machine learning applications
- Everything publicly available!!

Ω_m	0.1
σ_8	0.8
A_{SN1}	1.0
A_{AGN1}	1.0
A_{SN2}	1.0
A_{AGN2}	1.0




IllustrisTNG
Dark matter density
SIMBA



$\Omega_{\rm m} \in [0.1, 0.5]$
$\sigma_8 \in [0.6, 1.0]$
$A_{\rm SN1}, A_{\rm AGN1} \in [0.25, 4.0]$
$A_{\rm SN2}, A_{\rm AGN2} \in [0.5, 2.0]$
CAMELS
IllustrisTNG suite
SIMBA suite
LH set
1P set
CV set
EX set
LH set
1P set
CV set
EX set
1,000 sims
61 sims
27 sims
4 sims
1,000 sims
61 sims
27 sims
4 sims
A total of 2,184 state-of-the-art (magneto-)hydrodynamic simulations.
An N-body simulation for each (magneto-)hydrodynamic sim: 2,049 in total.
Total number of simulations in CAMELS: 4,233.

CAMELS

Francisco Villaescusa-Navarro, Daniel Angles-Alcazar, Shy Genel

Adrian Bayer
Alex Barreira
Ana Maria Delgado
Andrina Nicola
Alice Pisani
Benjamin Oppenheimer
Benjamin Wandelt
Blakesley Burkhart
ChangHoon Hahn
Colin Hill
Core Francisco Park
Daisuke Nagai
Desika Narayanan
David Spergel
Emily Moser
Erwin T. Lau
Faizan Mohammad
Gabrielle Parimbelli
Greg Bryan
Gabriella Contardo
Helen Shao
Jay Wadekar
Jingjing Shi
Joyce Caliendo
Lucia Perez
Lars Hernquist
Leander Thiele
Luis F. Machado Poletti
Matteo Viel
Matthew Gebhardt
Megan Tillman
Michael Eickenberg
Neerav Kaushal
Nicholas Battaglia
Oliver Philcox
Pablo Villanueva-Domingo
Rachel Somerville
Romeel Dave
Stephanie Tonnensen
Sultan Hassan
Romain Teyssier
Ulrich Steinwandel
Valentina La Torre
Vid Irsic
William Coulton
Yin Li
Yongseok Jo
Yueying Ni

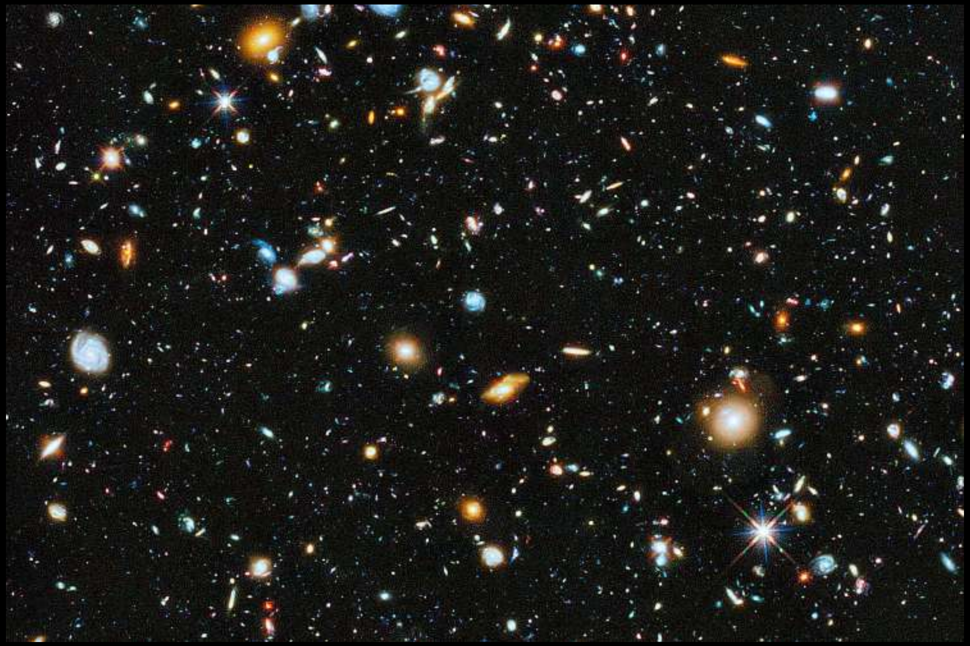
Summary

1. <u>Motivation & Description</u>

2. Results

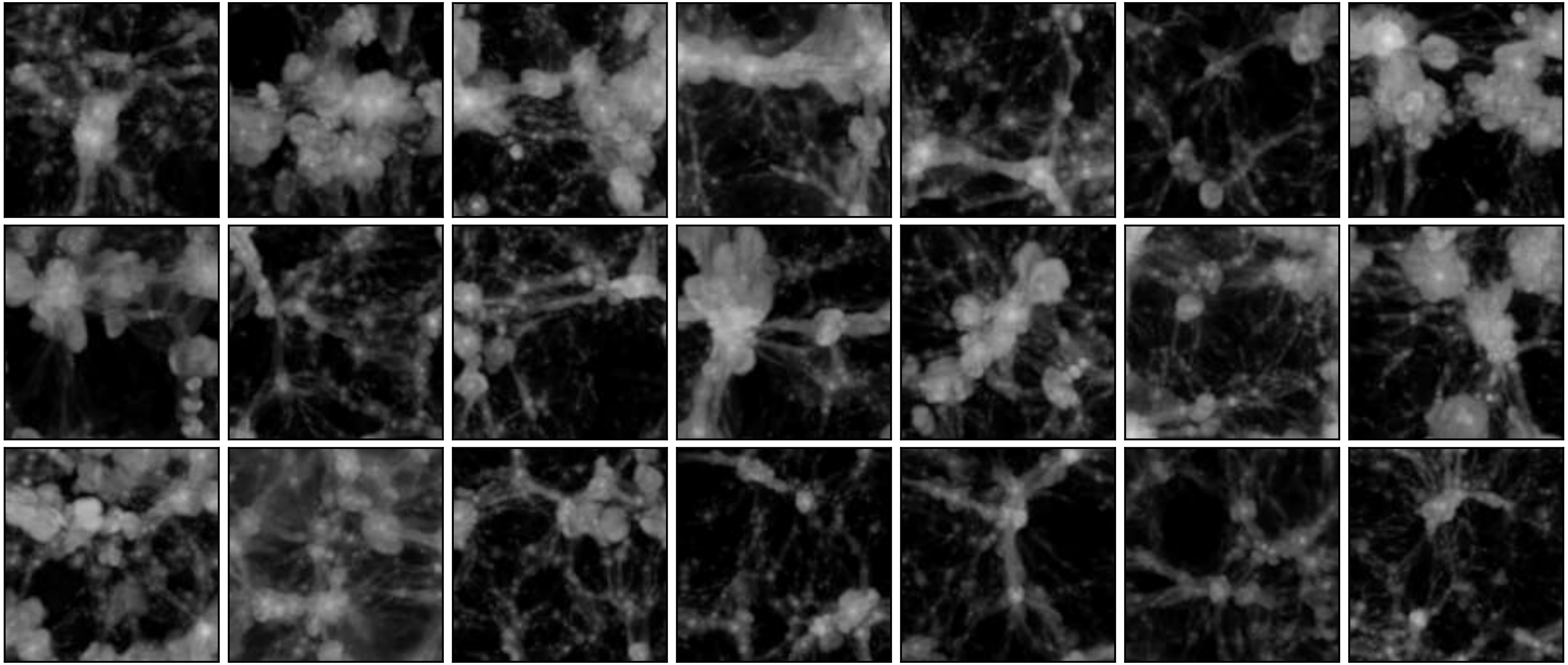
3. Conclusions

The problem



Goal: extract the maximum information from surveys

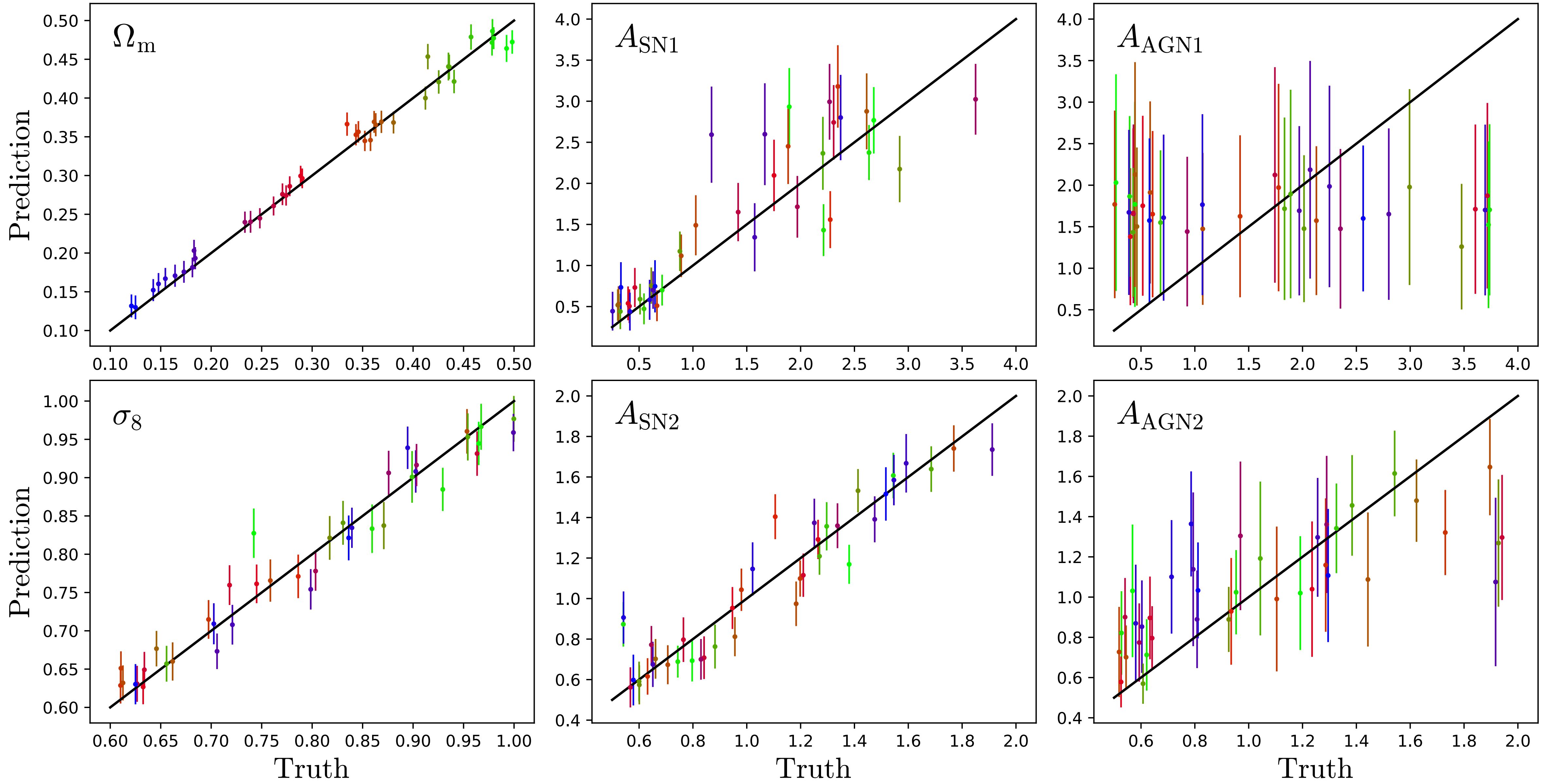
Example I: Gas temperature



Every map has 256×256 pixels, covers an area of 25×25 $(h^{-1}\mathrm{Mpc})^2$, and has a different cosmology & astrophysics. 15,000 images in total.

Likelihood-free inference: gas temperature

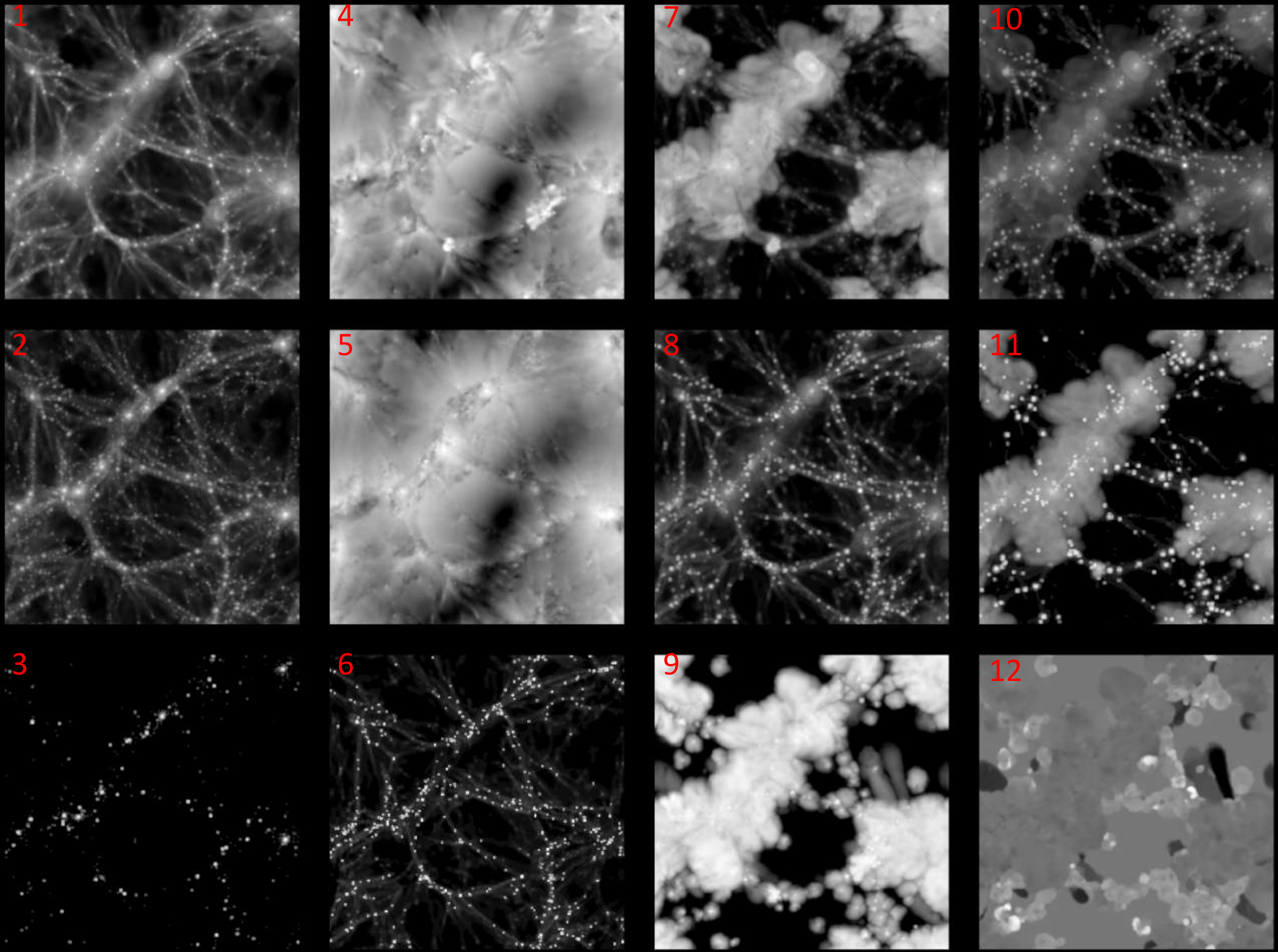
FVN et al. 2021a



15 different fields; 15,000 maps/field

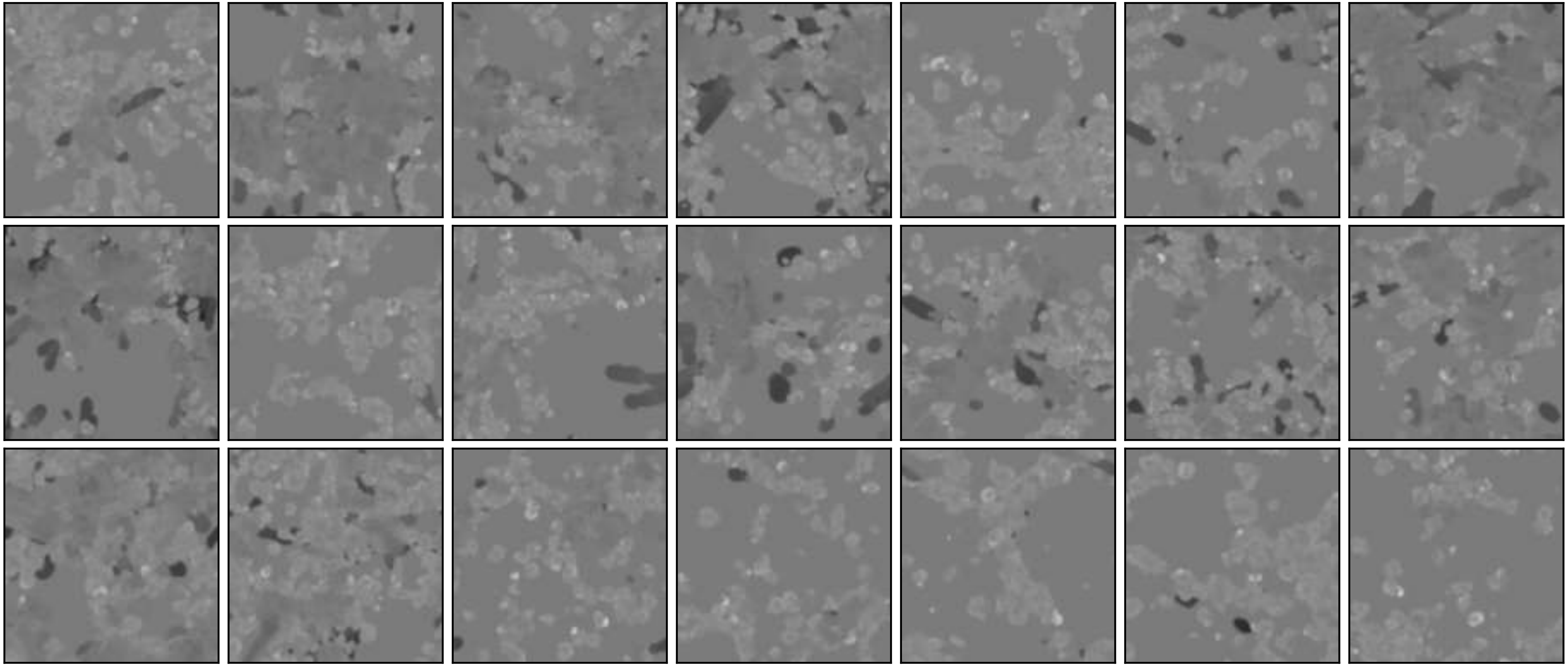
1. Gas mass
2. Dark matter mass
3. Stellar mass
4. Gas velocity
5. Dark matter velocity
6. Neutral hydrogen mass
7. Gas temperature
8. Electron density
9. Gas metallicity
10. Gas pressure
11. Magnetic fields
12. Mg/Fe
13. Total mass
14. Total mass
15. Multifield

Magneto-hydrodynamic simulations
N-body simulations

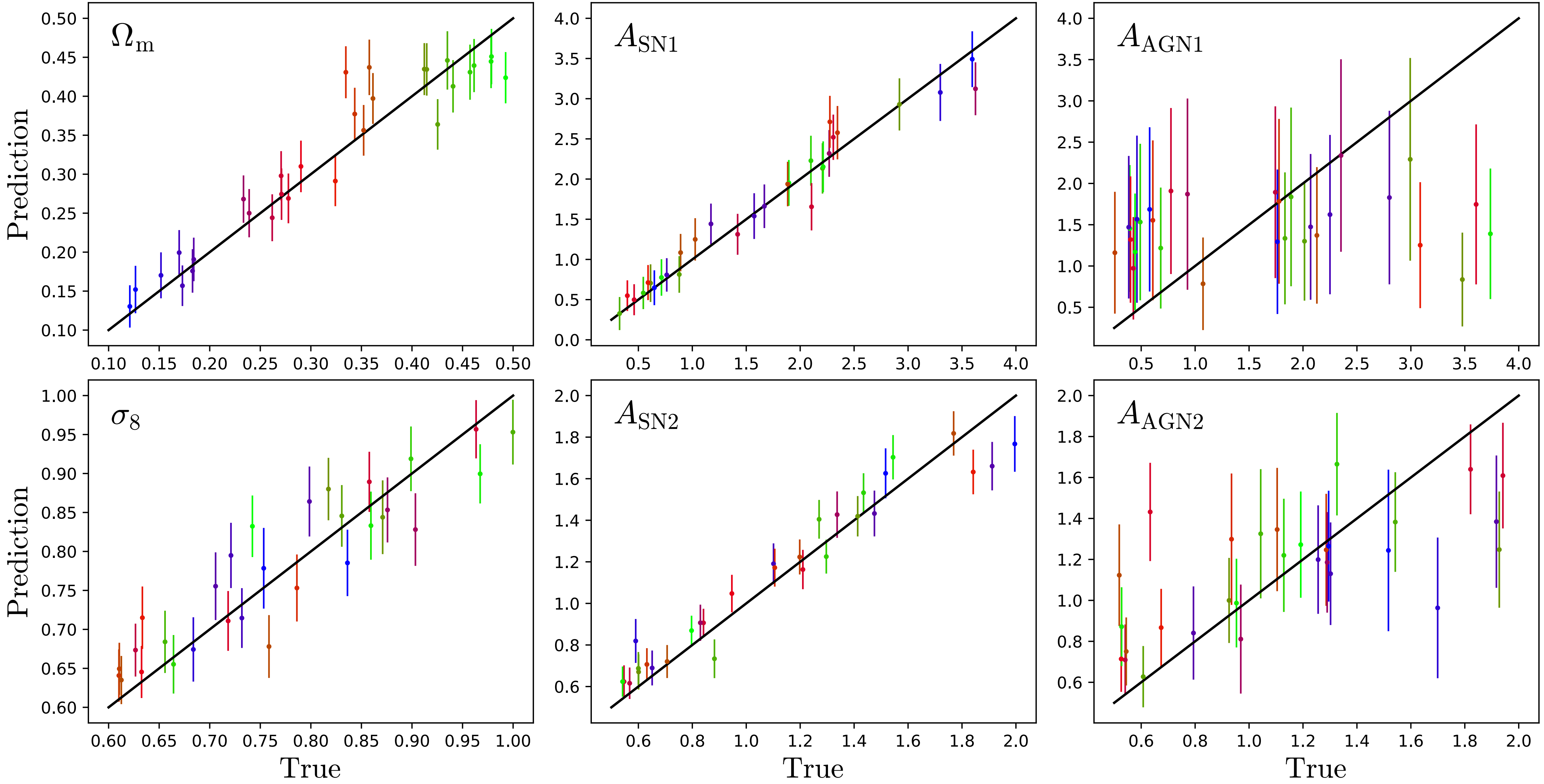


Example II: Magnesium over Iron (Mg/Fe)

FVN et al. 2021a

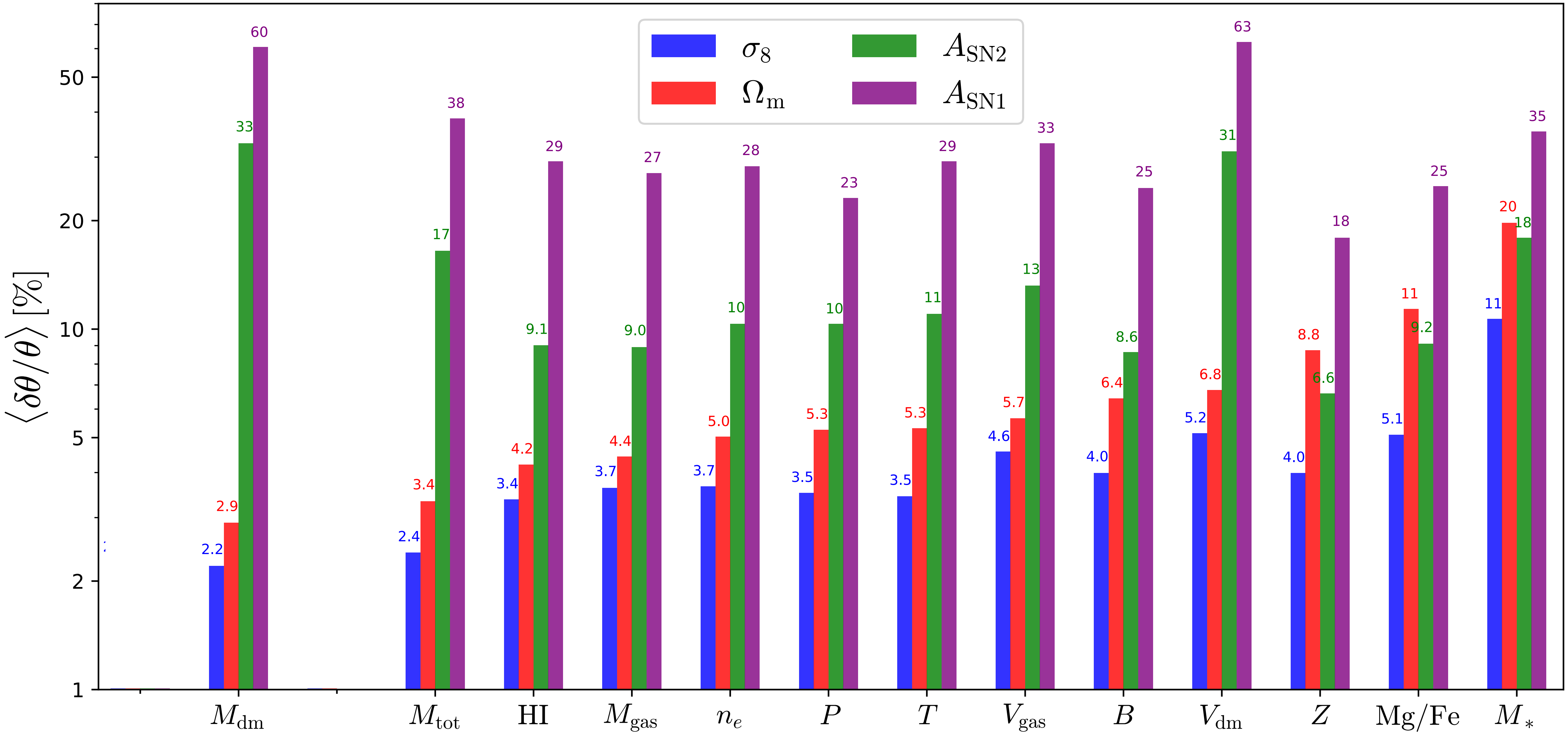


Likelihood-free inference: Mg/Fe



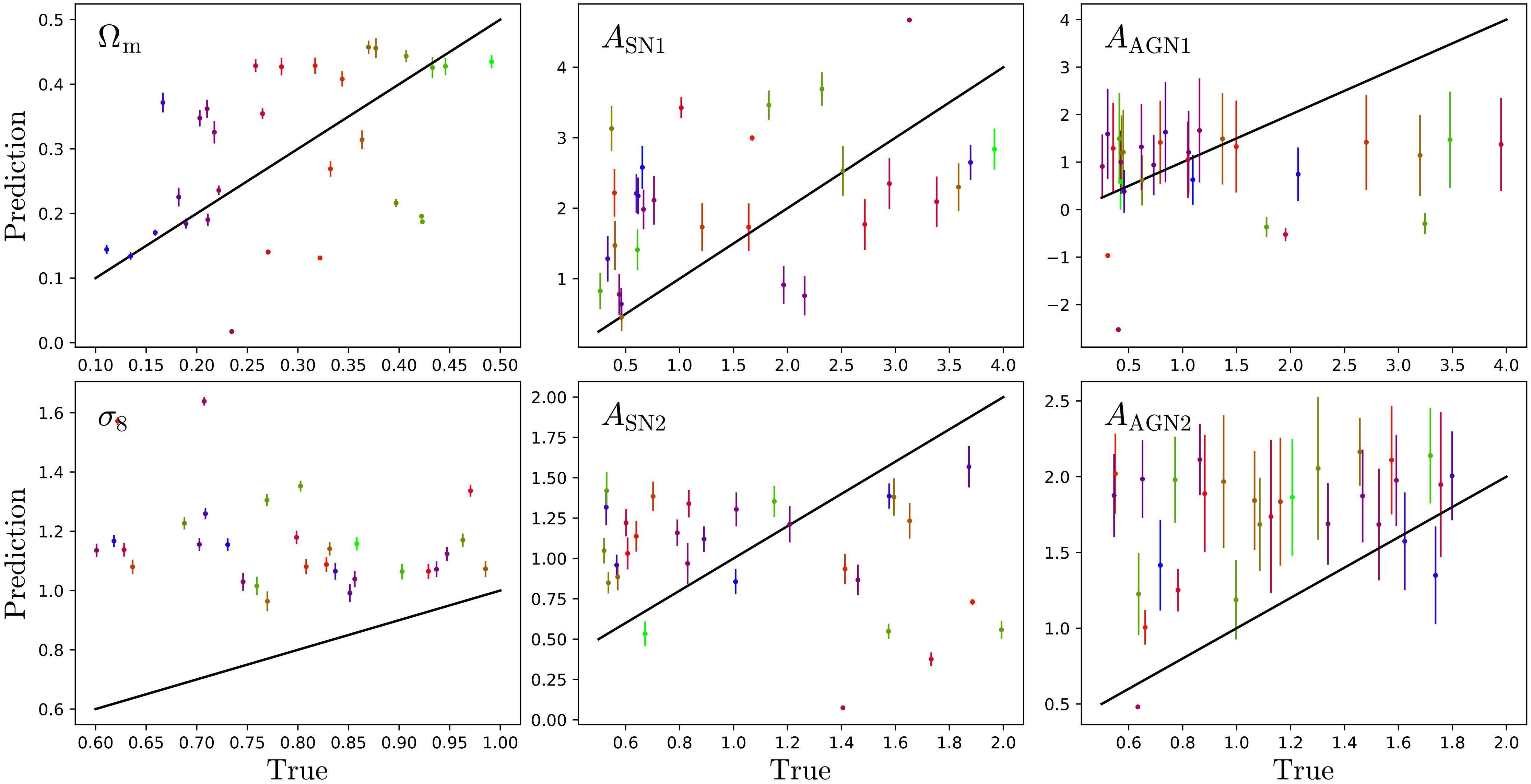
Fields comparison

FVN et al. 2021a



Robustness: gas temperature

Network trained on IllustrisTNG and tested on SIMBA

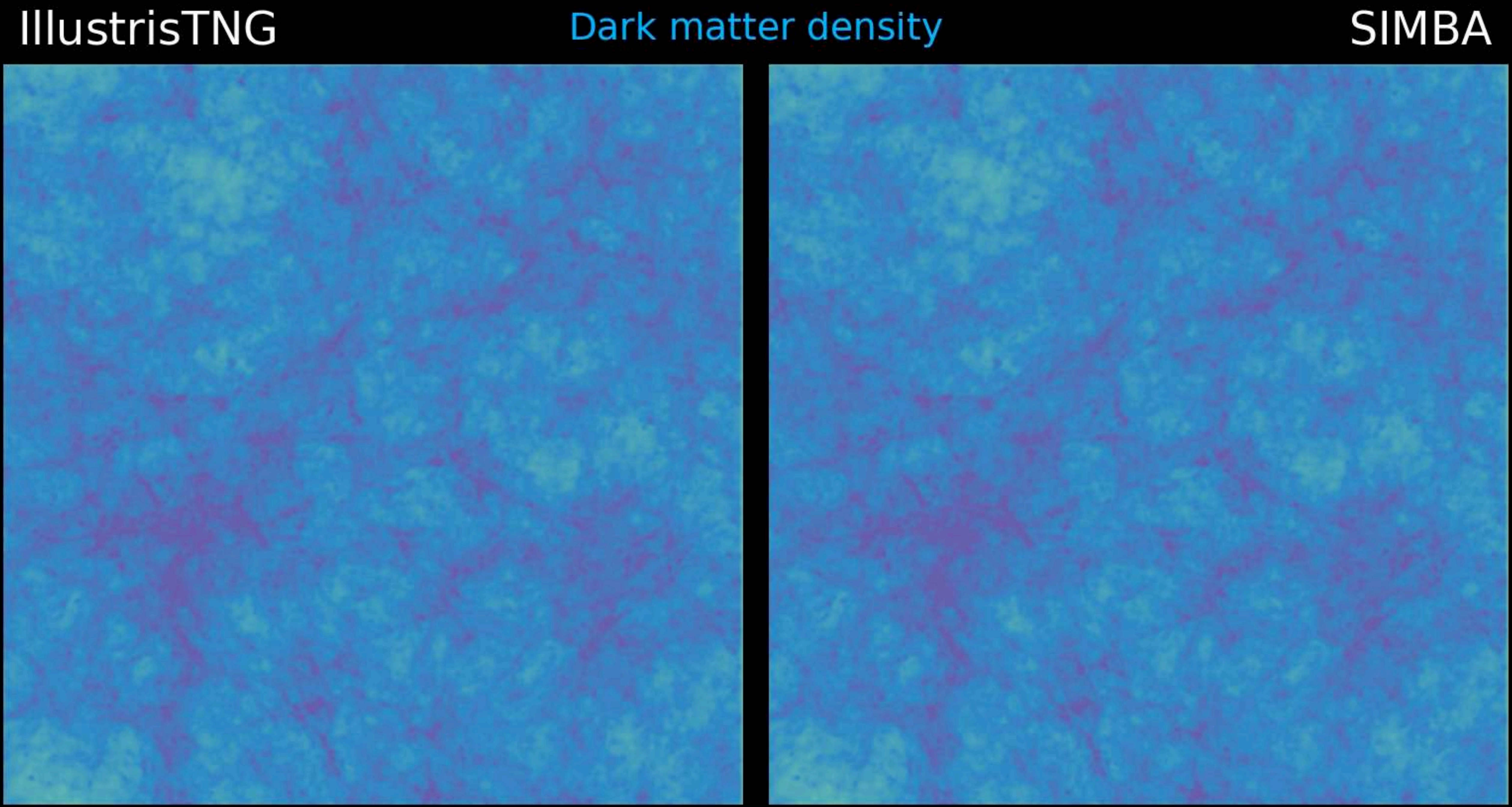


Robustness: gas temperature

Why?

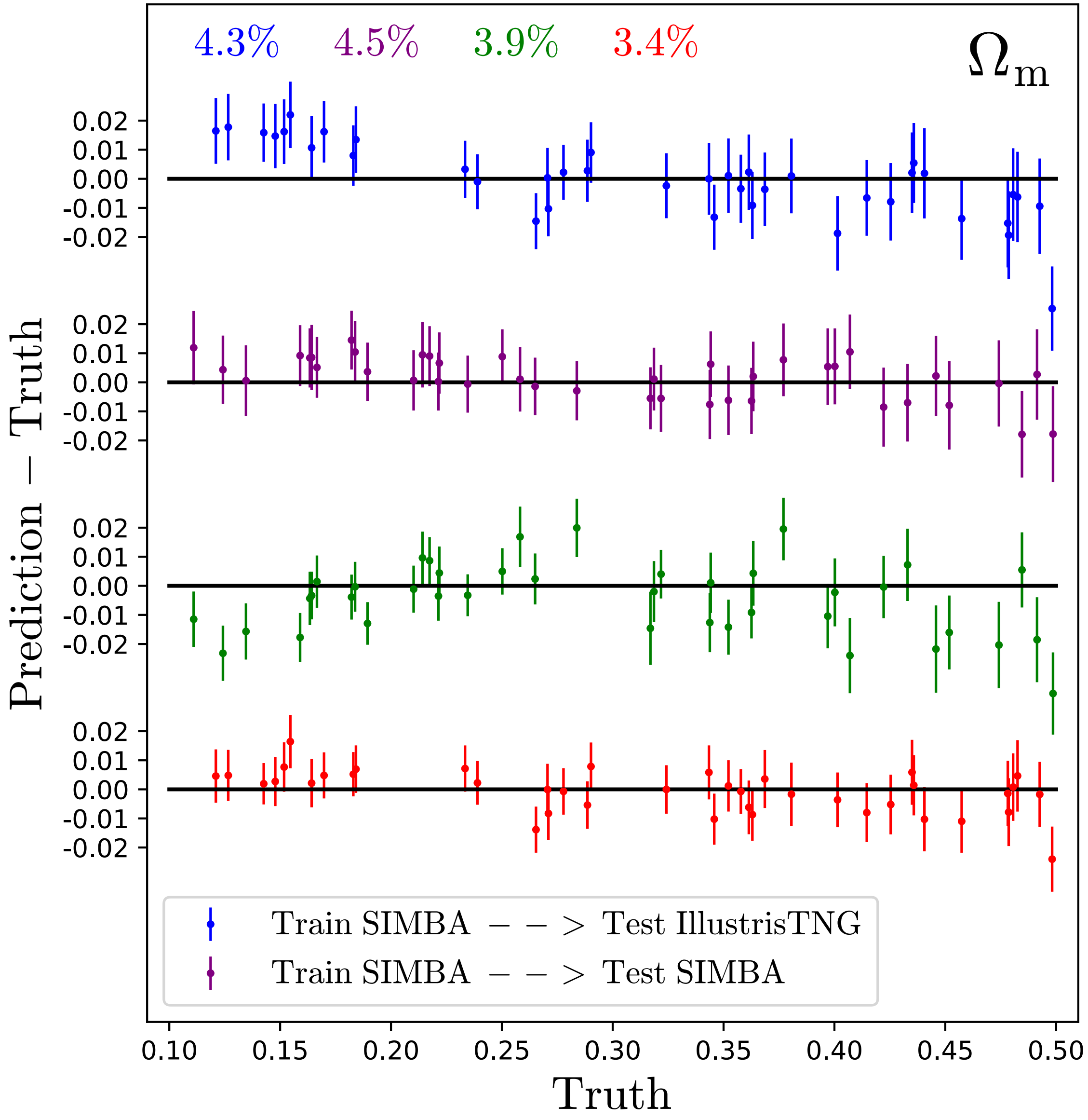
- Maybe simulations live in different regions in parameter space
- Maybe different representations of the data (e.g. resolution)
- Maybe the network is learning numerical artifacts from the simulations

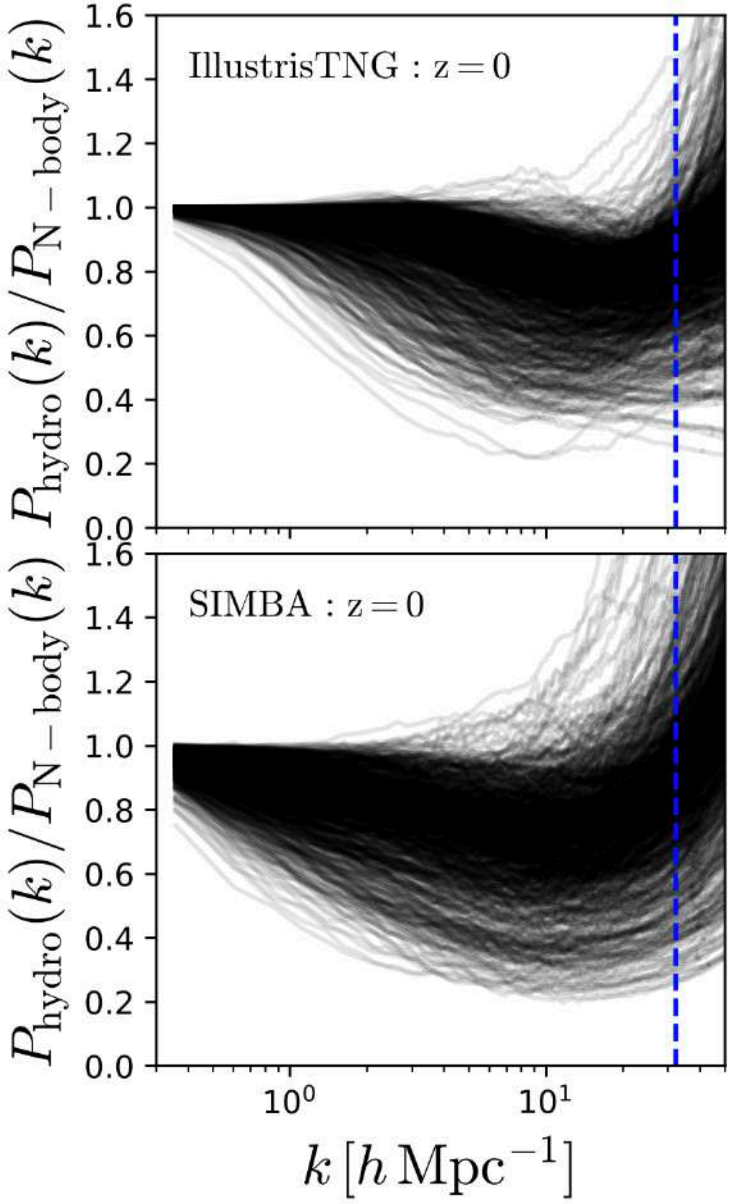
Dark matter density



Robustness: total matter

FVN et al. 2021b



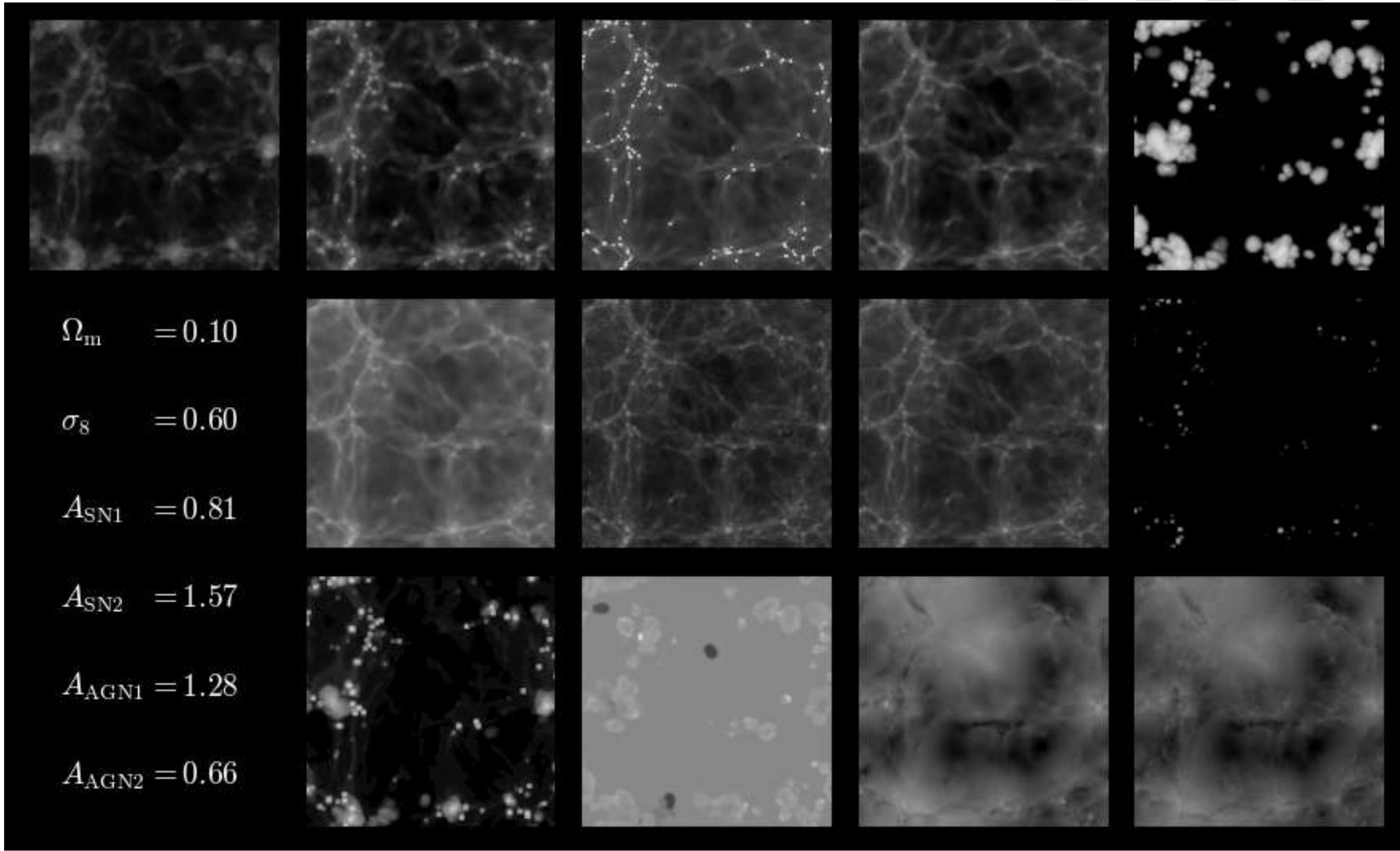


CAMELS Multifield Dataset

FVN et al. 2021c

https://camels-multifield-dataset.readthedocs.io

- Hundreds of thousands of labeled 2D maps and 3D grids
- Several redshifts: 0, 0.5, 1, 1.5, 2
- Three different resolutions
- 13 different fields:
  1. Gas density
  2. Gas temperature
  3. Gas metallicity
  4. Gas pressure
  5. Neutral hydrogen density
  6. Electron number density
  7. Dark matter density
  8. Total matter density
  9. Stellar mass density
  10. Gas velocity
  11. Dark matter velocity
  12. Magnetic fields
  13. Mg/Fe
- 70 Tb of data; Publicly available
- The MNIST of cosmology

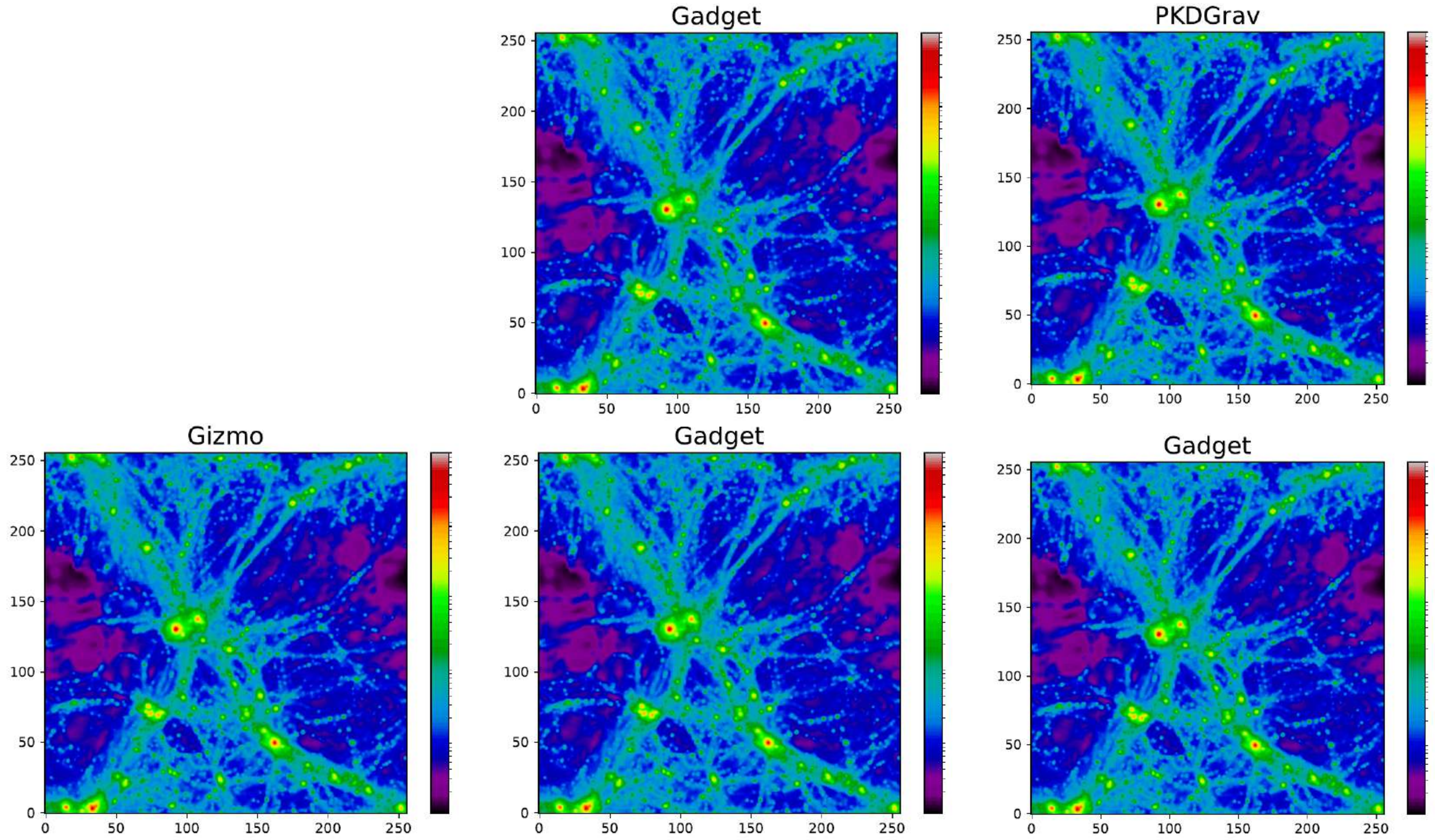


Field level comparison: Nbody

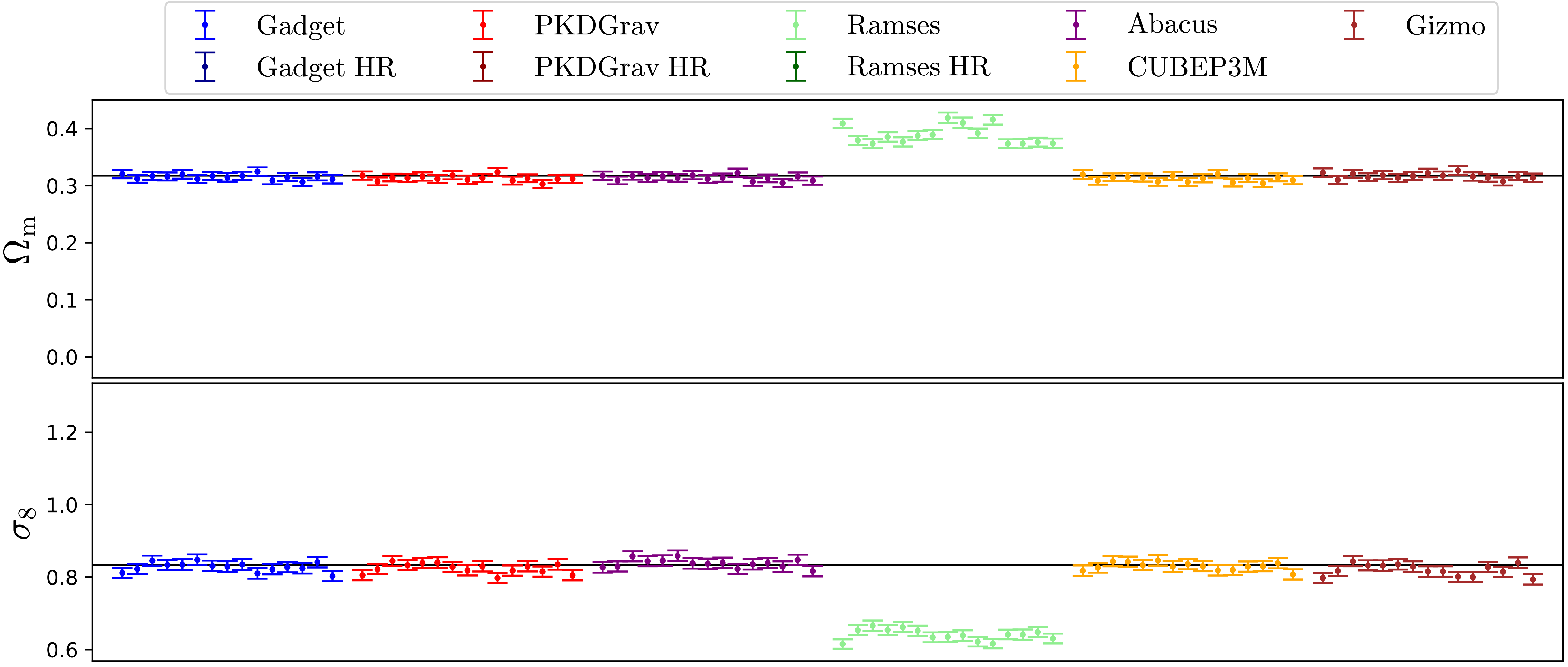
FVN et al. (in prep)

- Gadget
- Abacus
- PKDGrav
- Ramses
- CUBEP3M
- Gizmo

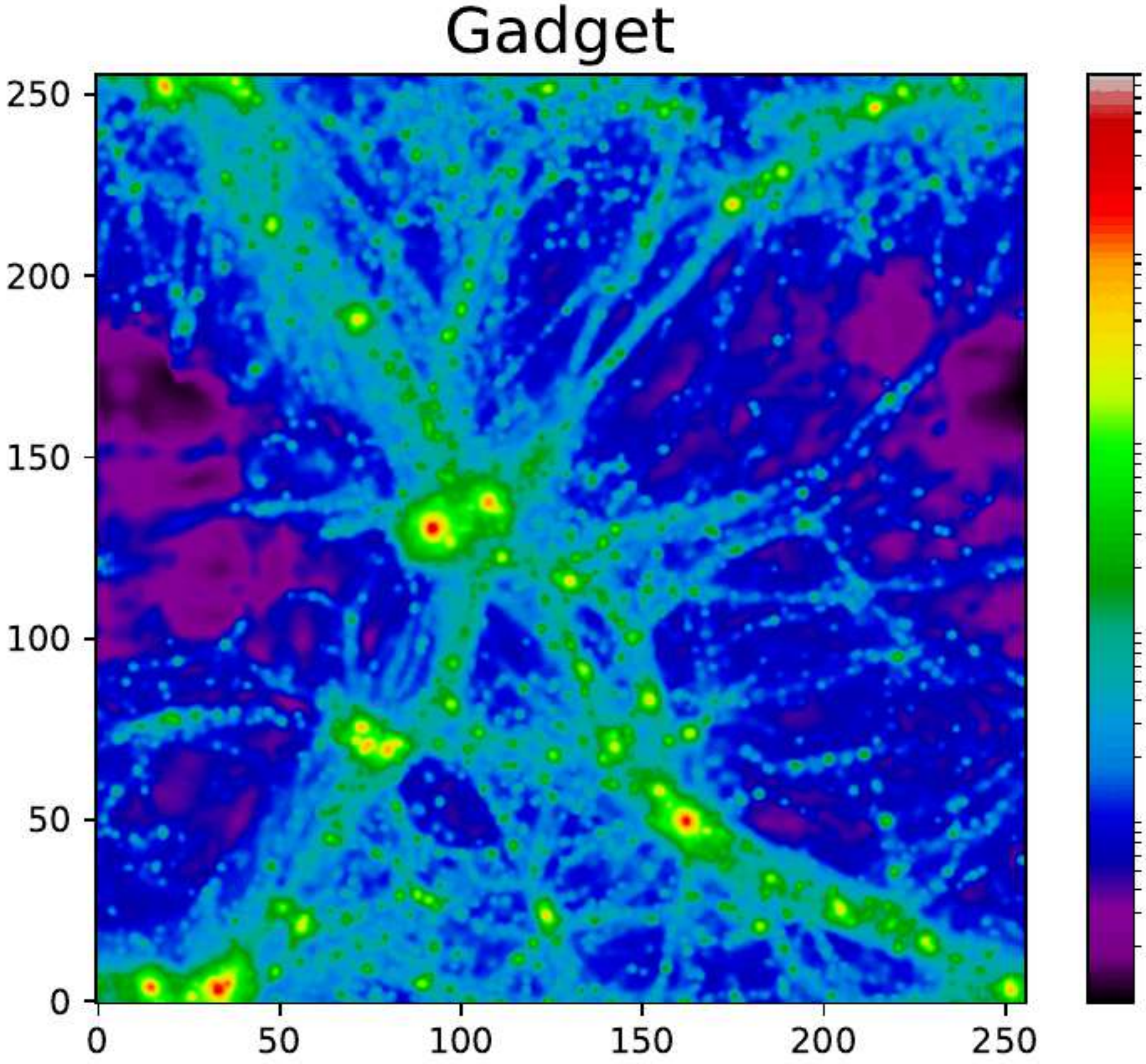
- COLA
- CAMELS



Field level comparison: Nbody



Field level comparison: Nbody



Field level comparison: Nbody

Training on Gadget + CAMELS does not work on Ramses
“Marginalizing” over resolution with Gadget does not work on Ramses

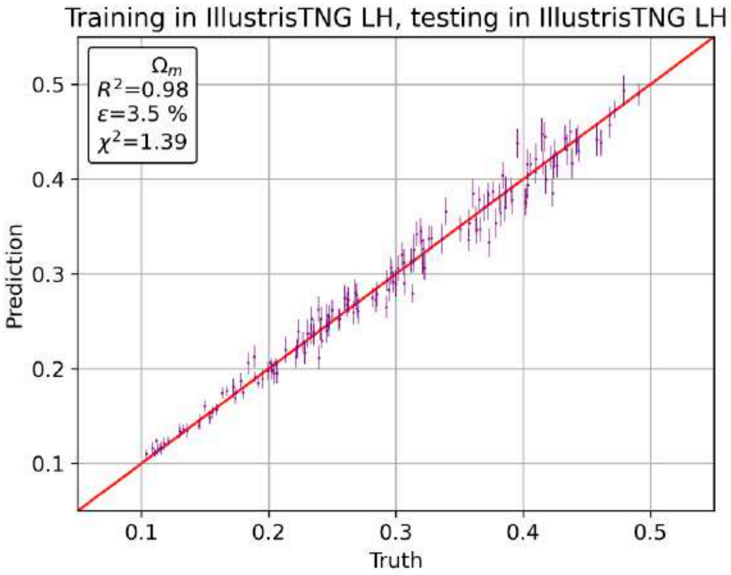
What guarantee do we have that simulations will overlap with reality?

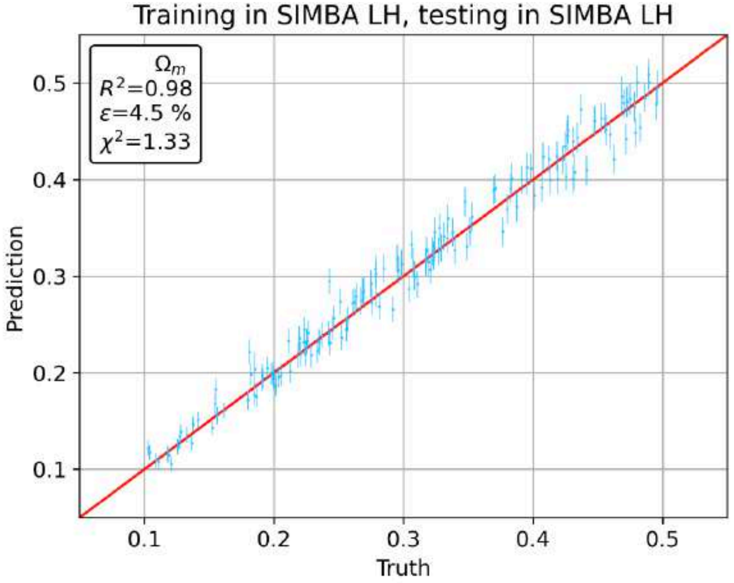
Galaxy field LFI

Galaxy field: Results

Villanueva-Domingo & FVN (in prep)

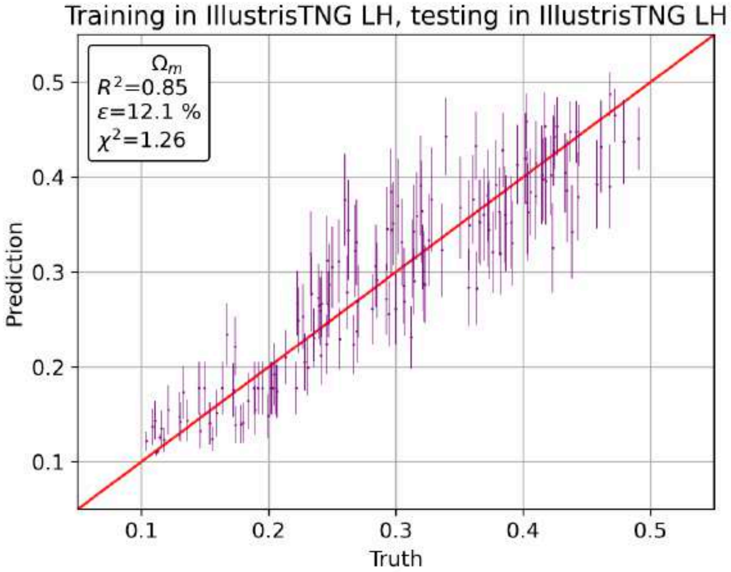
Pablo Villanueva-Domingo
(Valencia)

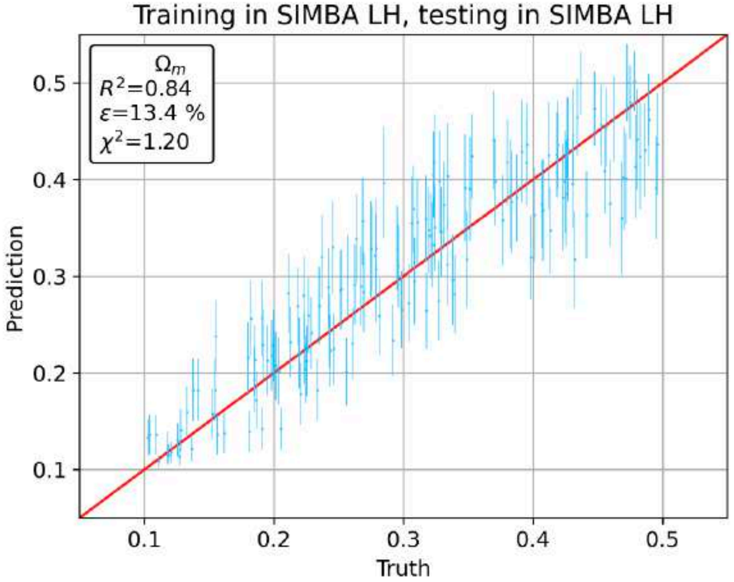




Translational and rotational invariance built-in.

It marginalizes over baryonic effects automatically

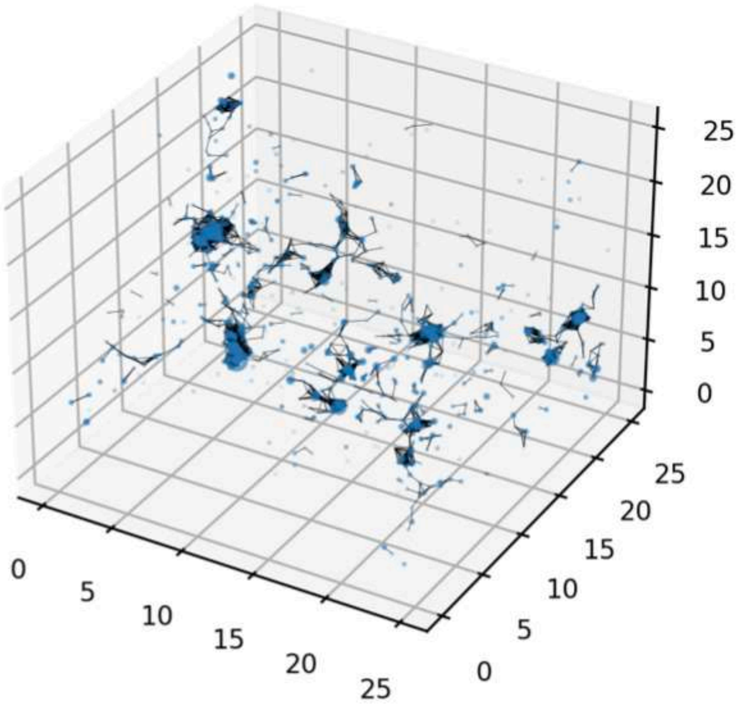


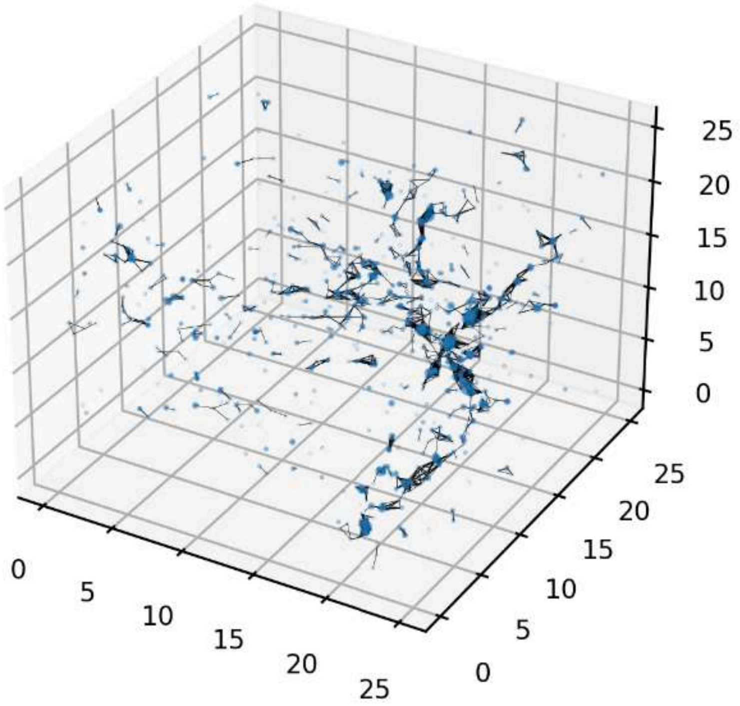


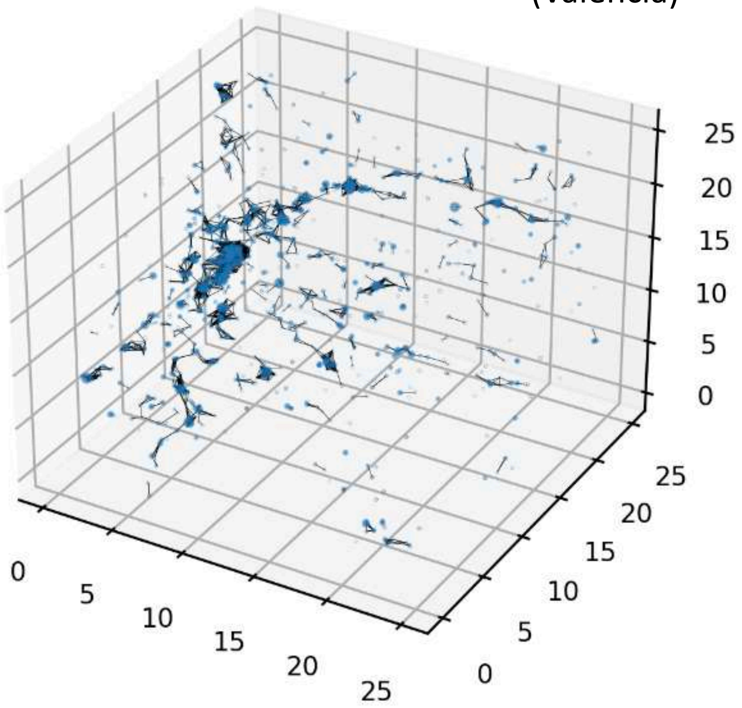
Galaxy field: graphs

Villanueva-Domingo & FVN (in prep)

Pablo Villanueva-Domingo
(Valencia)



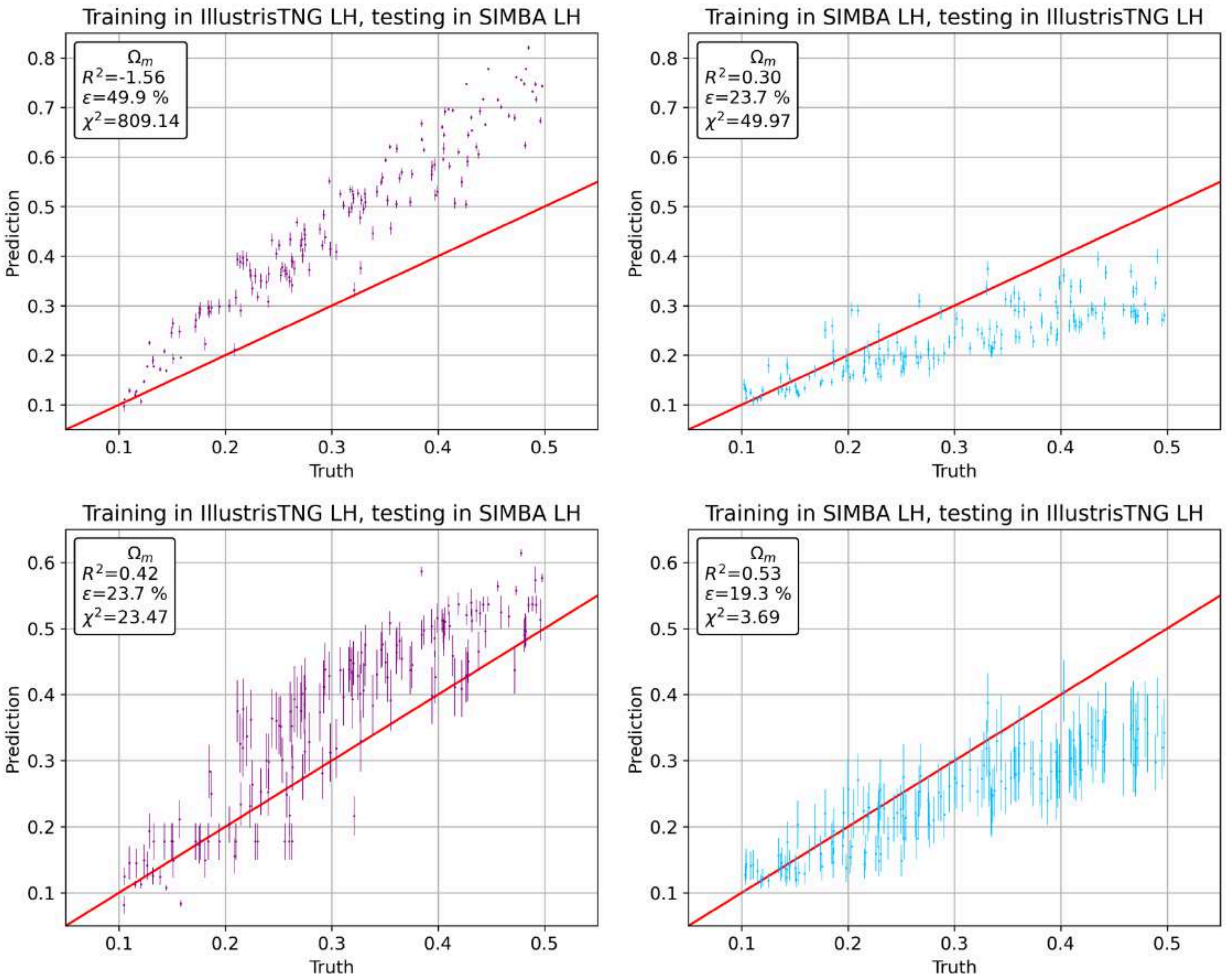




Galaxy field: Robustness

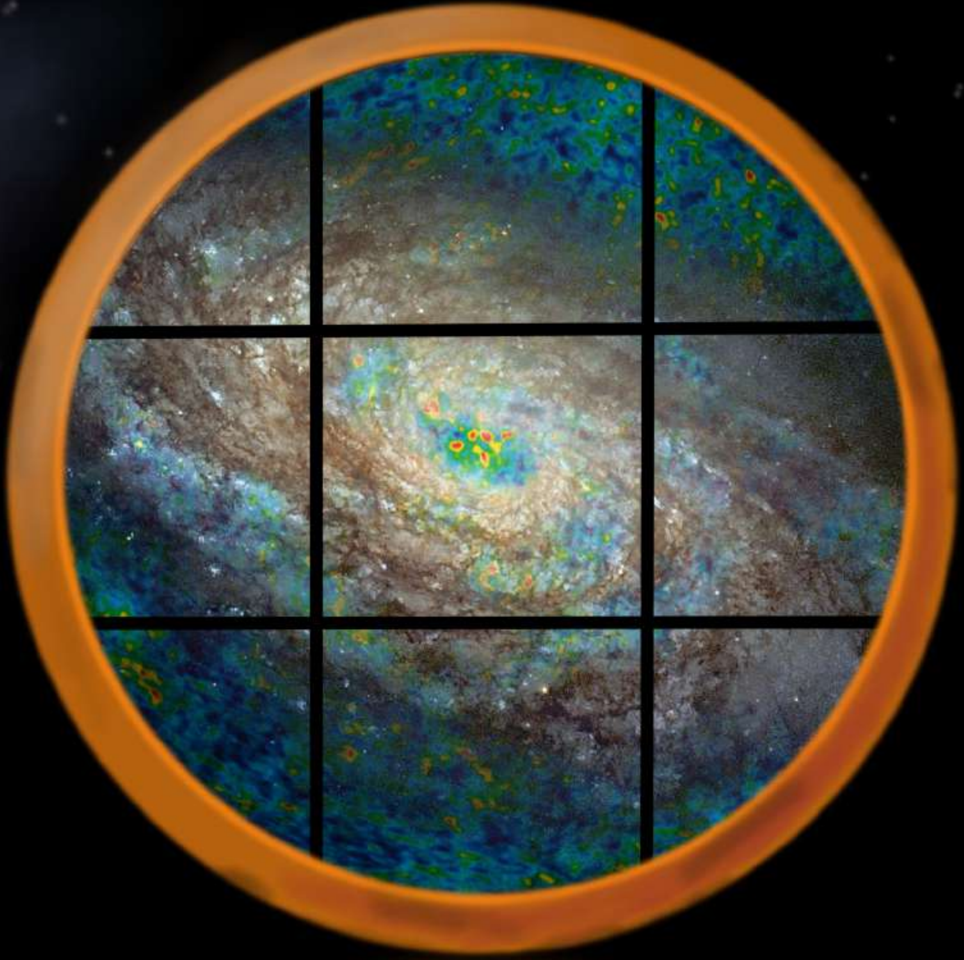
Villanueva-Domingo & FVN (in prep)

Pablo Villanueva-Domingo
(Valencia)

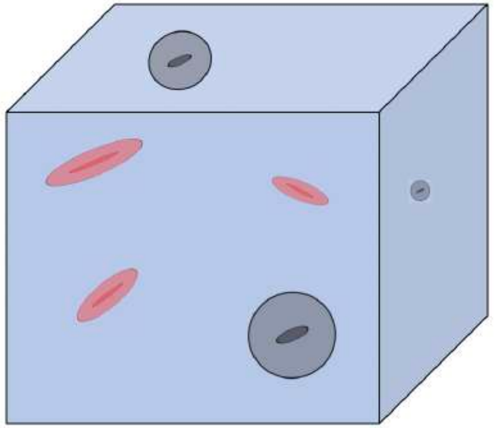


Cosmology with 1 galaxy?

FVN et al. 2022b



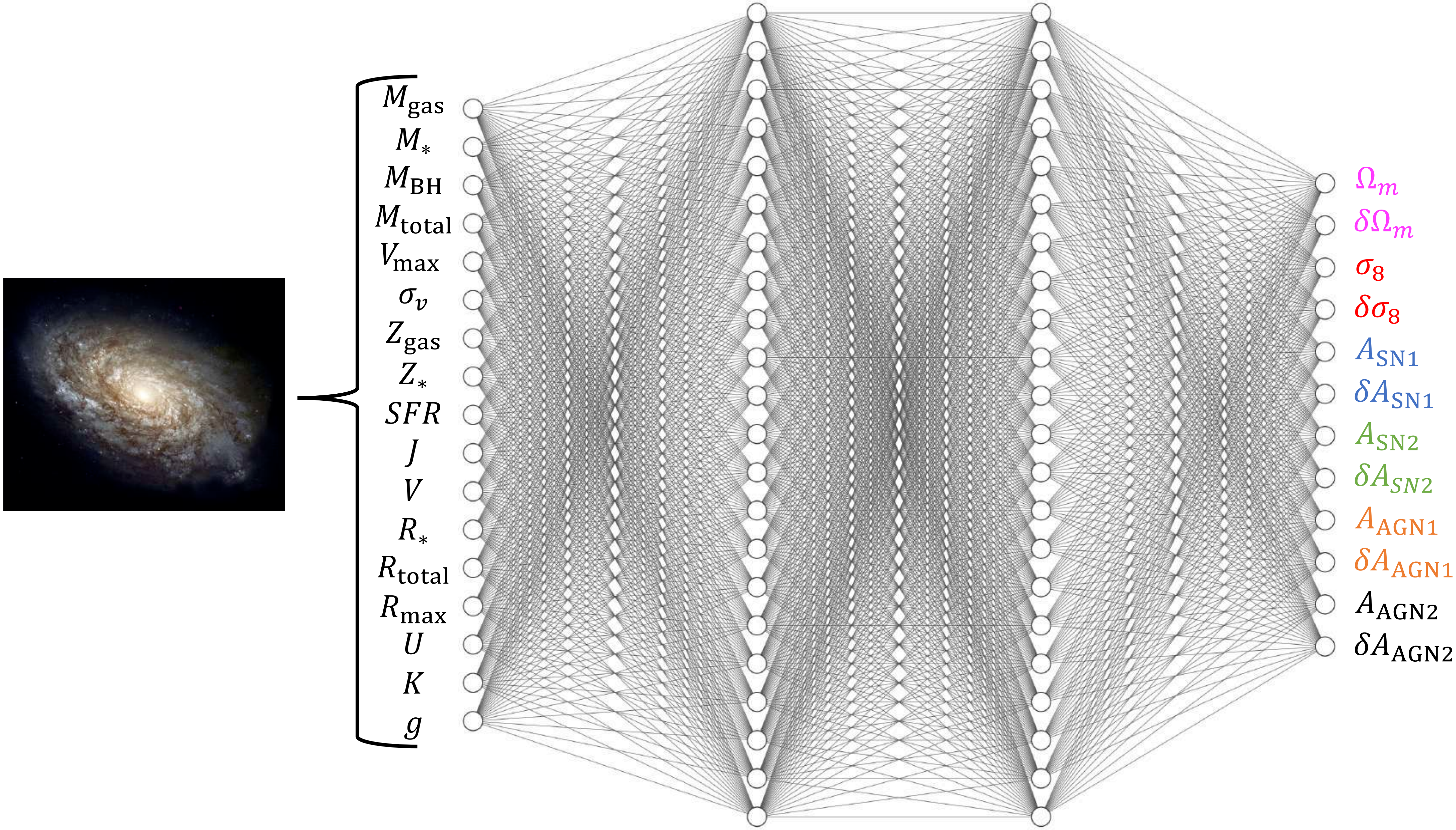
In collaboration with Jupiter Ding, Shy Genel, Stephanie Tonnesen, Valentina La Torre, David Spergel, Romain Teyssier, Yin Li, Caroline Heneka, Pablo Lesmos, Daniel Angles-Alcazar, Daisuke Nagai

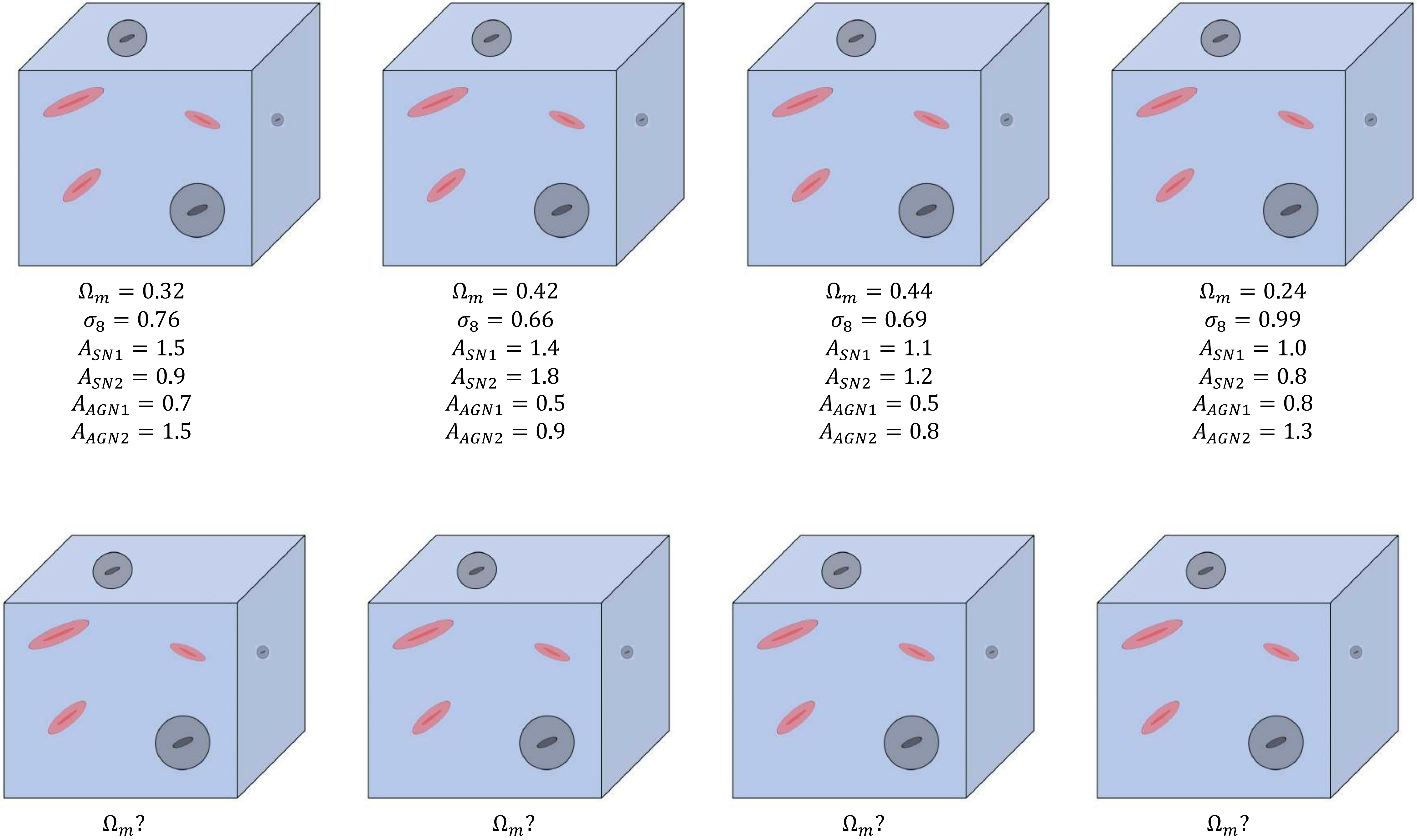


$\Omega_m = 0.32$
$\sigma_8 = 0.76$
$A_{SN1} = 1.5$
$A_{SN2} = 0.9$
$A_{AGN1} = 0.7$
$A_{AGN2} = 1.5$

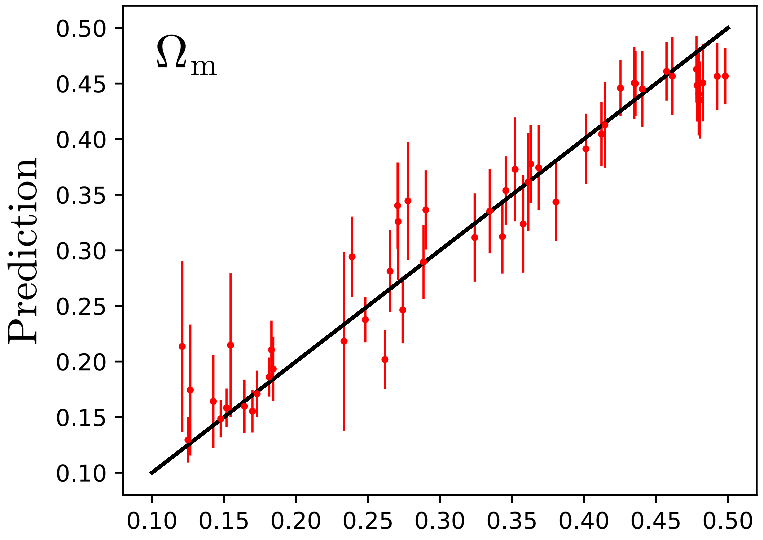
CAMELS

Cosmology with 1 galaxy

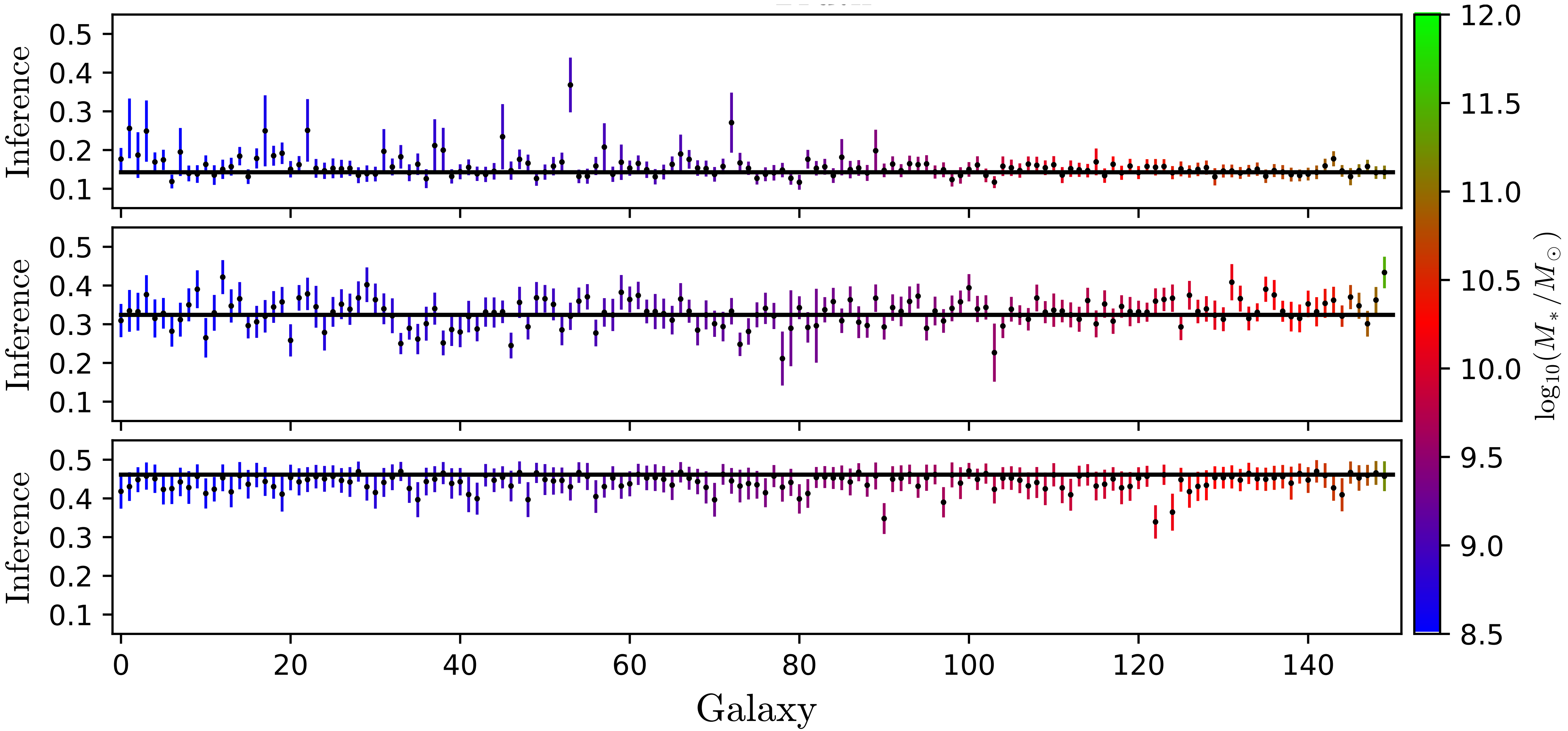



$\Omega_m = 0.32$
$\sigma_8 = 0.76$
$A_{SN1} = 1.5$
$A_{SN2} = 0.9$
$A_{AGN1} = 0.7$
$A_{AGN2} = 1.5$
$\Omega_m = 0.42$
$\sigma_8 = 0.66$
$A_{SN1} = 1.4$
$A_{SN2} = 1.8$
$A_{AGN1} = 0.5$
$A_{AGN2} = 0.9$
$\Omega_m = 0.44$
$\sigma_8 = 0.69$
$A_{SN1} = 1.1$
$A_{SN2} = 1.2$
$A_{AGN1} = 0.5$
$A_{AGN2} = 0.8$
$\Omega_m = 0.24$
$\sigma_8 = 0.99$
$A_{SN1} = 1.0$
$A_{SN2} = 0.8$
$A_{AGN1} = 0.8$
$A_{AGN2} = 1.3$
Ω_m?
Ω_m?
Ω_m?
Ω_m?

Cosmology with 1 galaxy

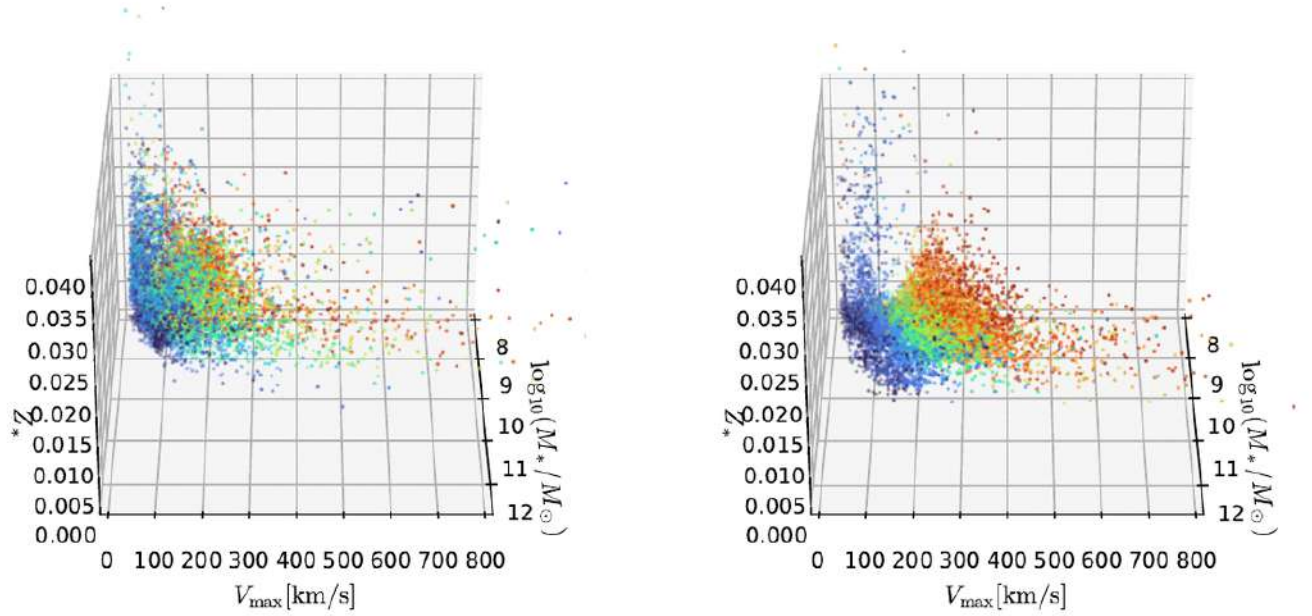


Cosmology with 1 galaxy

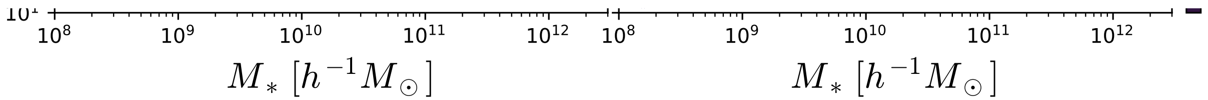


Cosmology with 1 galaxy

IllustrisTNG	SIMBA
$M_*, V_{\max}, Z_*, R_*, \mathrm{K}$	$M_*, V_{\max}, Z_*, R_*, R_{\max}$



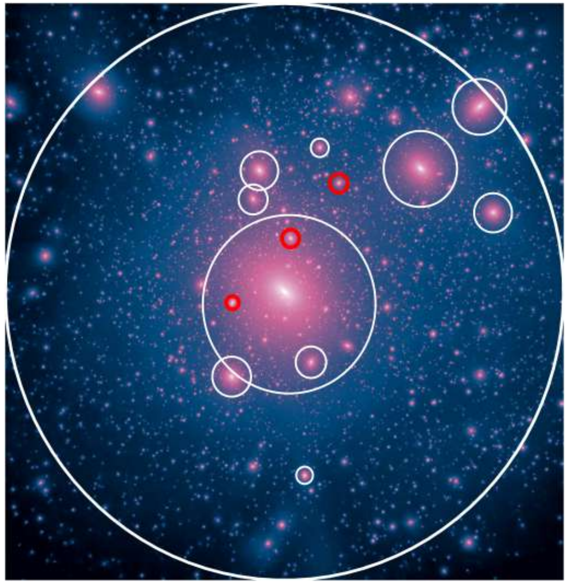
σ_8



Using AI to learn physics

Shao, FVN et al. 2021
(2109.04484)

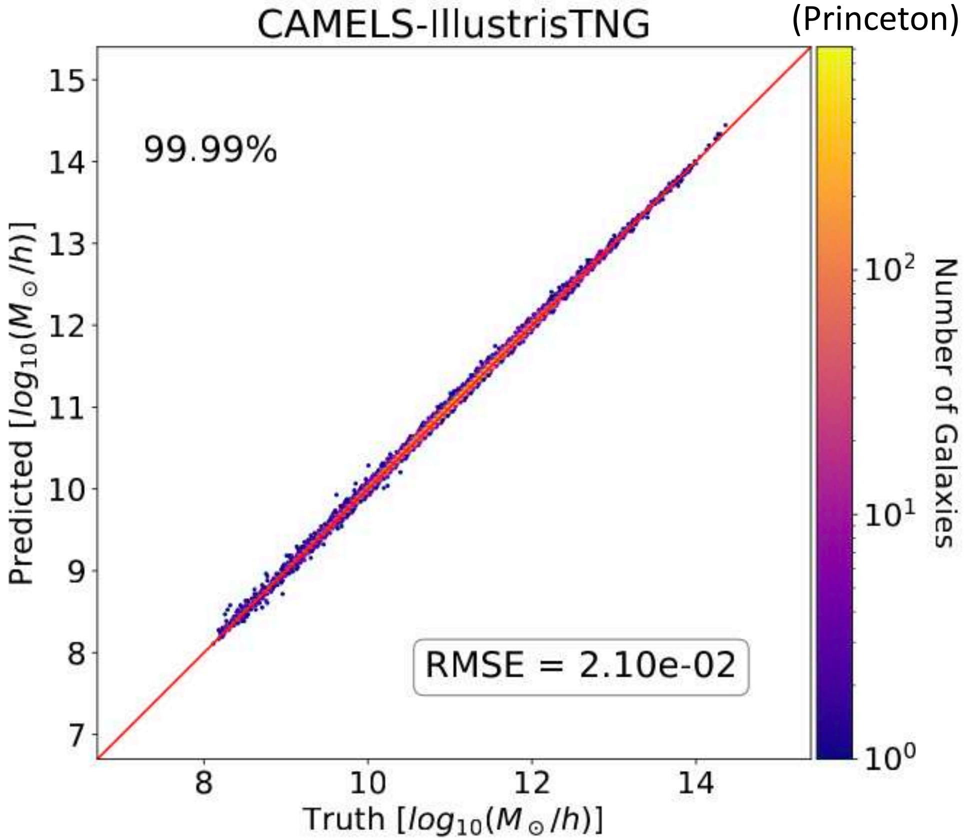
Helen Shao
(Princeton)



- Gas mass
- Stellar mass
- Black hole mass
- Gas metallicity
- Stellar metallicity
- Radius
- V_{max}
- Velocity dispersion
- Star-formation rate
- Spin

Total subhalo mass

$$M_{\rm tot} = A\sigma^{(\alpha_0+\alpha_1 \log\sigma)} R^{(\beta_0+\beta_1 \log R)} V_{\rm max}^{(\gamma_0+\gamma_1 \log V_{\rm max})}$$

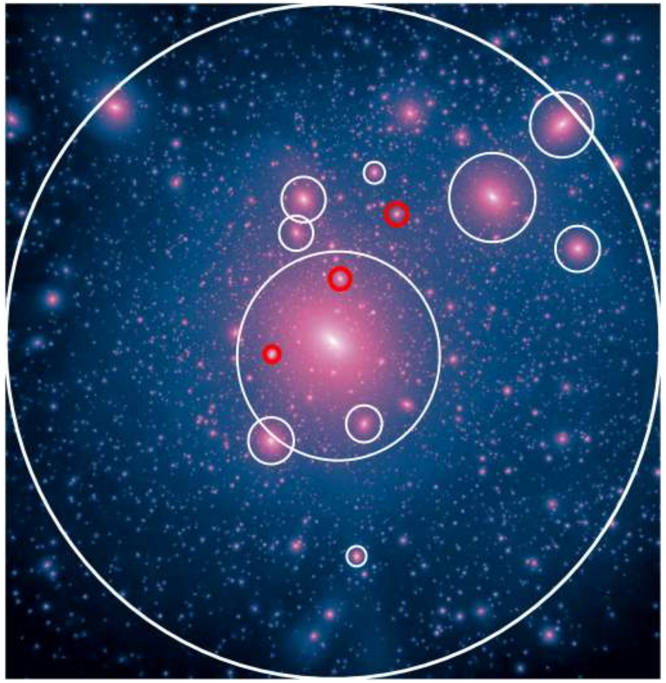


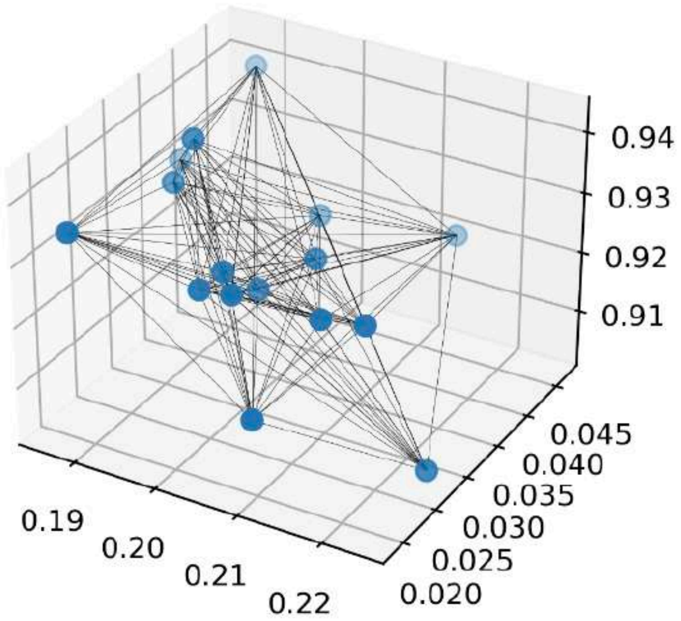
It works for any subhalo (central or satellite) containing any type of galaxy at any redshift from simulations with different cosmologies, different astrophysics, different subgrid physics, different resolutions, and different volumes. *Some version of the virial theorem*

Weighing halos with galaxies

Villanueva-Domingo, FVN et al. 2021a

Pablo Villanueva-Domingo
(Valencia)

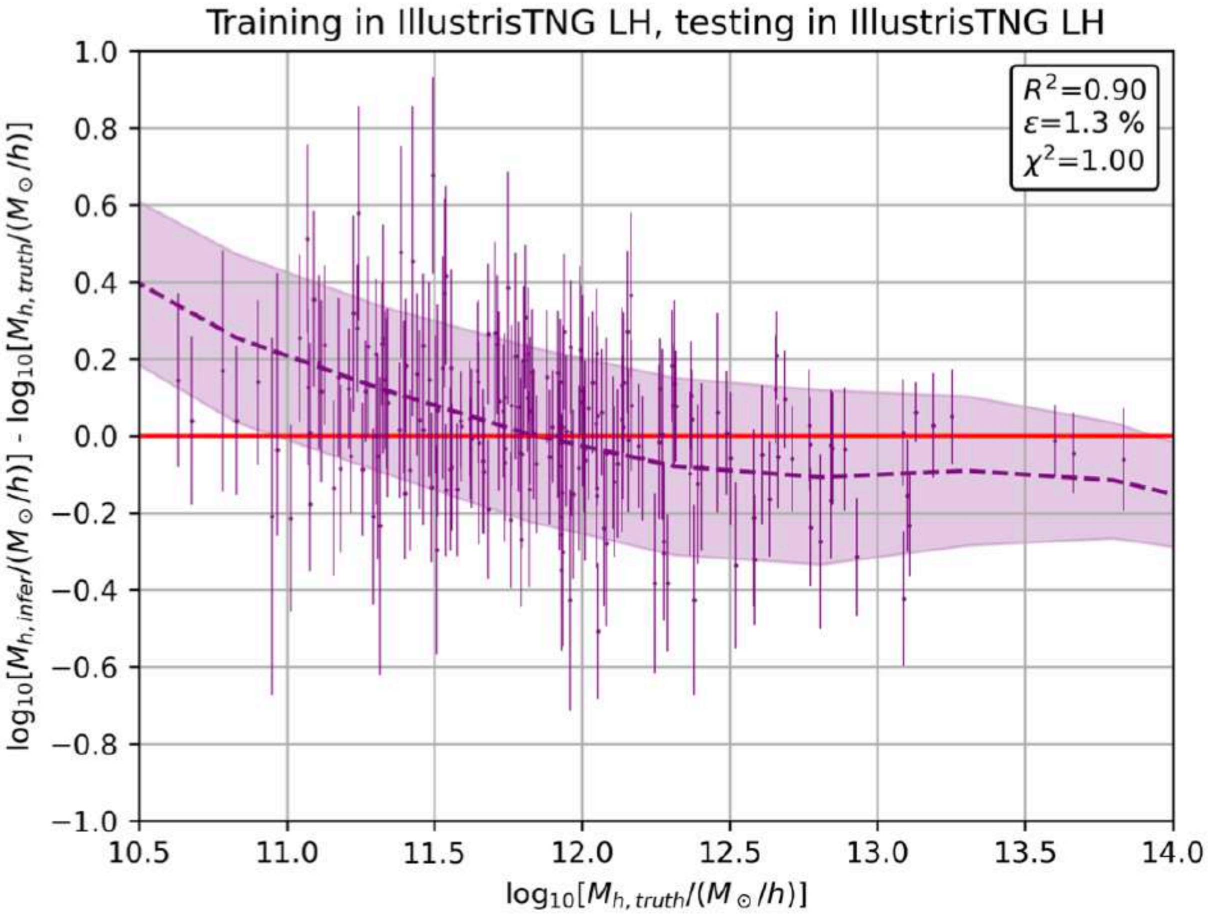


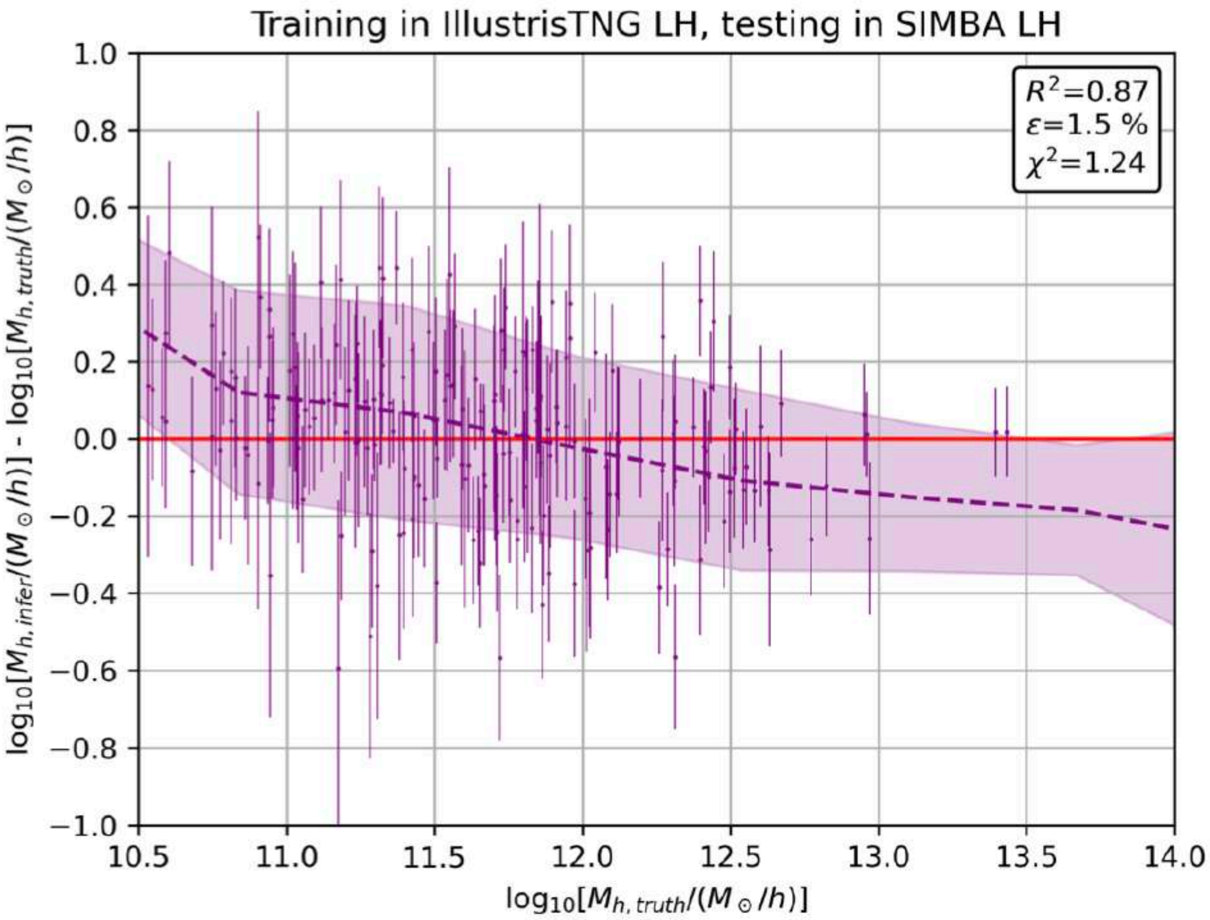


Weighing halos with galaxies

Villanueva-Domingo, FVN et al. 2021a

Pablo Villanueva-Domingo
(Valencia)

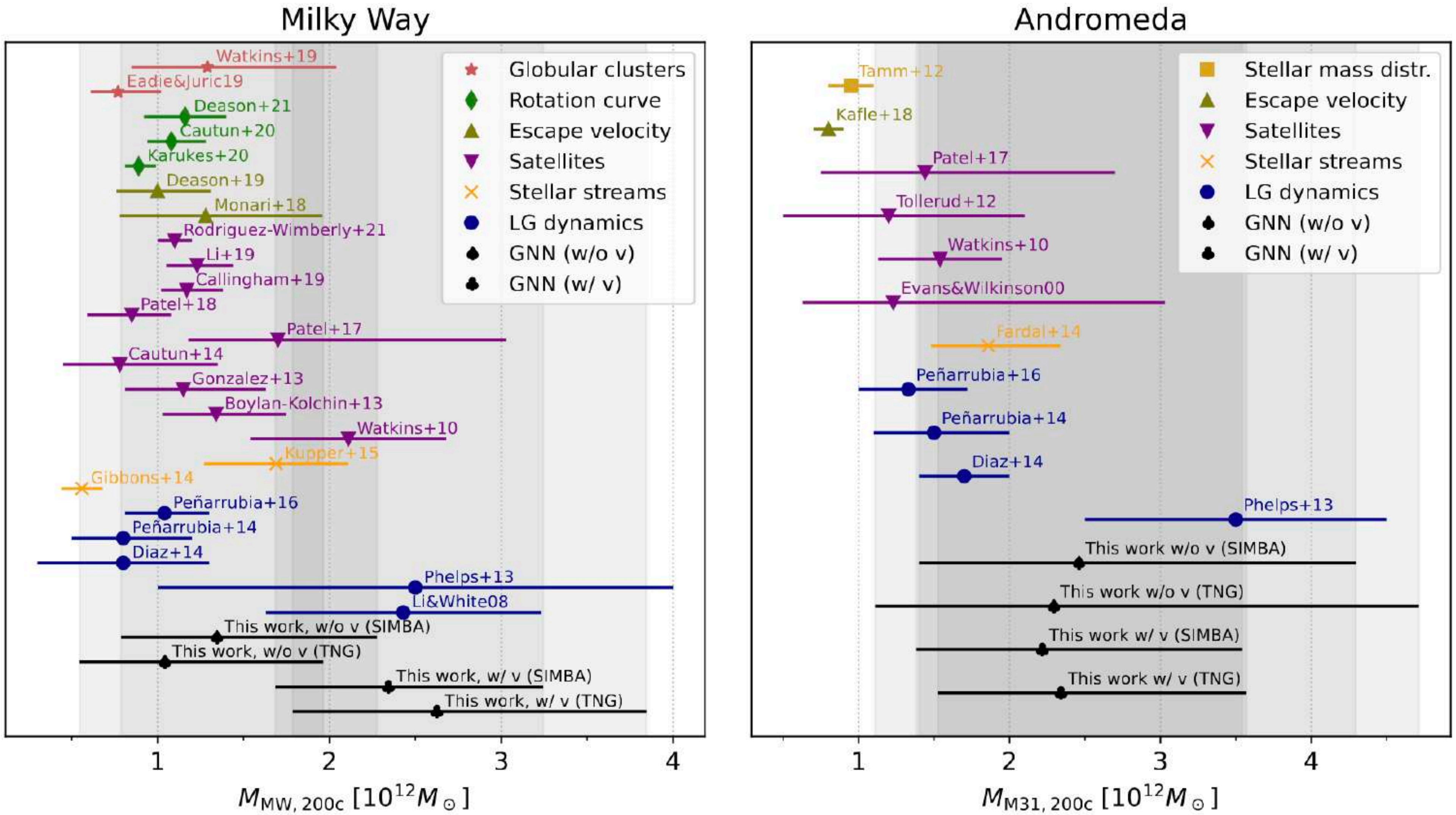




Weighing our own galaxy!

Villanueva-Domingo, FVN et al. 2021b

Pablo Villanueva-Domingo
(Valencia)



Other works

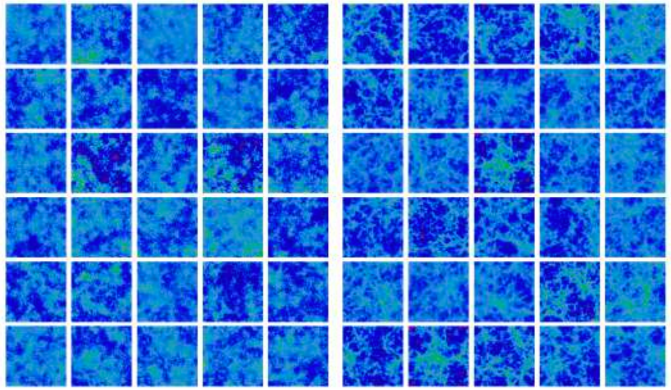
Andrina Nicola
(Princeton)

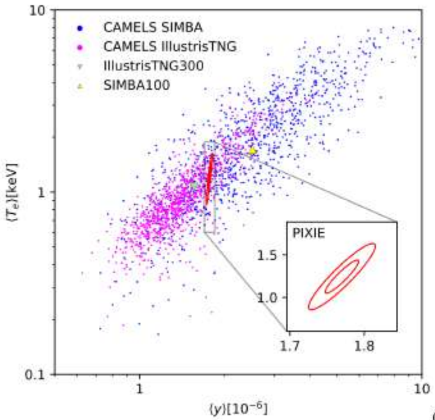
Sultan Hassan
(CCA)

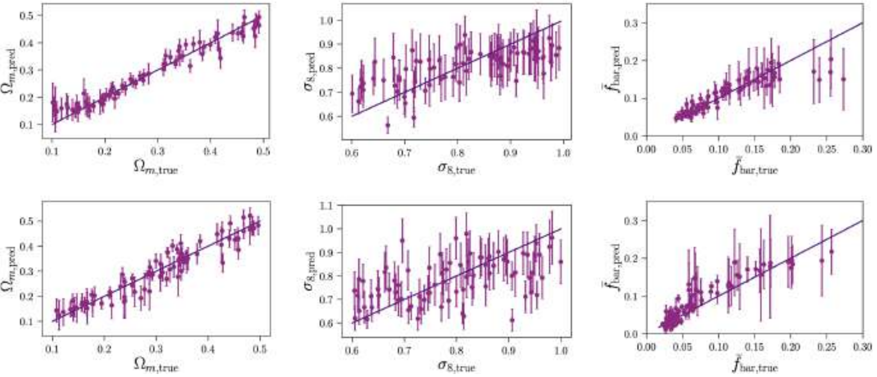
Emily Moser
(Cornell)

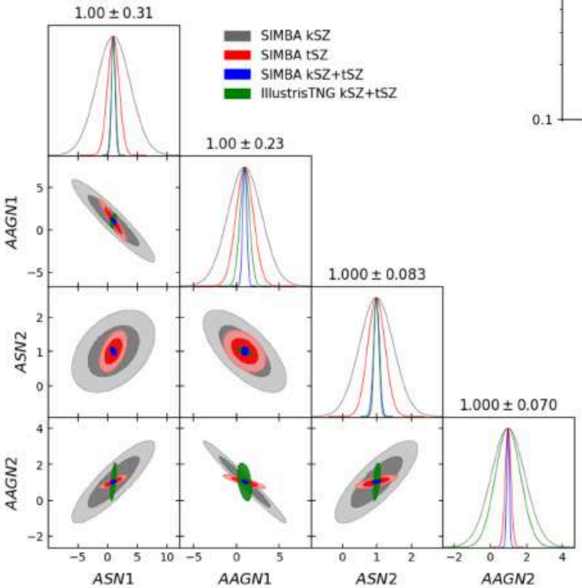
Leander Thiele
(Princeton)

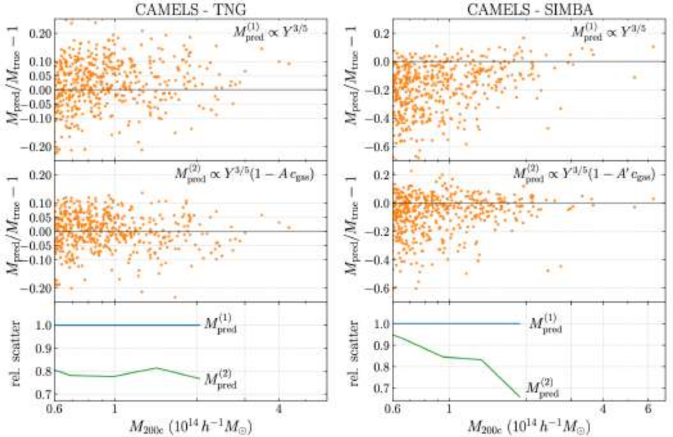
Jay Wadekar
(IAS)







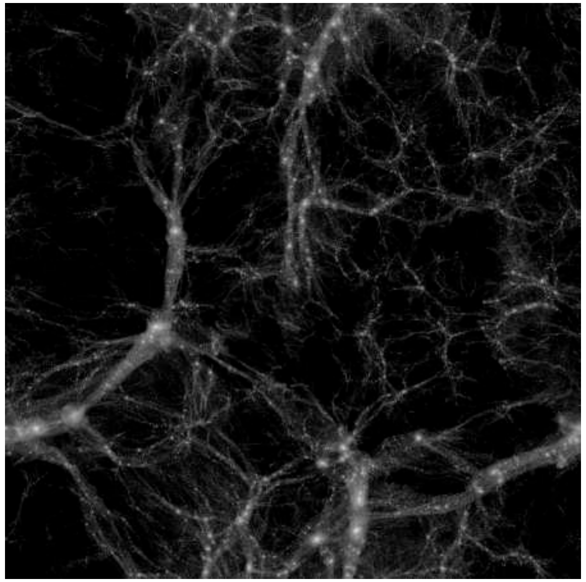




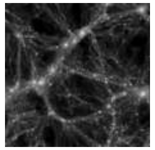
CAMELS-SAM

Lucia Perez
(Arizona/CCA)

Perez, Genel et al. 2022 (in prep)



25Mpc/h
CAMELS
hydro box



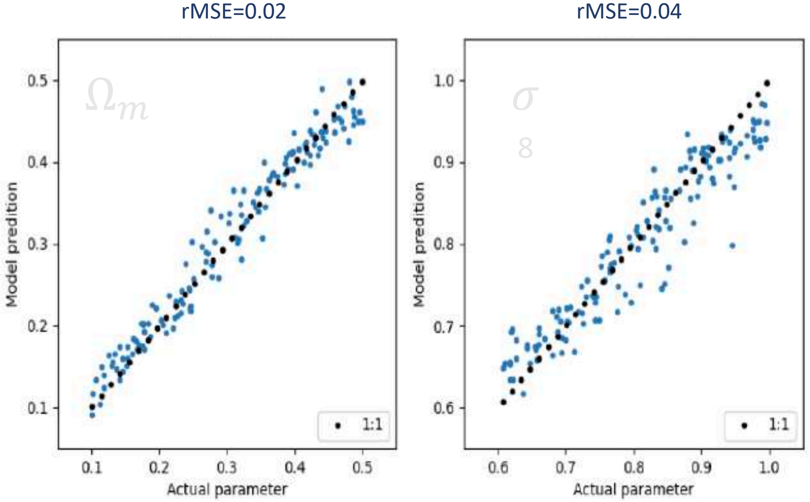
A suite of 1000 DM-only runs:

- 100 Mpc/h on a side
- 640^3 particles ($m_{DM} \approx 3 \cdot 10^8\, M_\odot$)
- $0.1 < \Omega_m < 0.5$; $0.6 < \sigma_8 < 1.0$

And on top of which 1000 runs of the Santa Cruz SAM (Somerville+ 2015) with variations of baryonic physics:

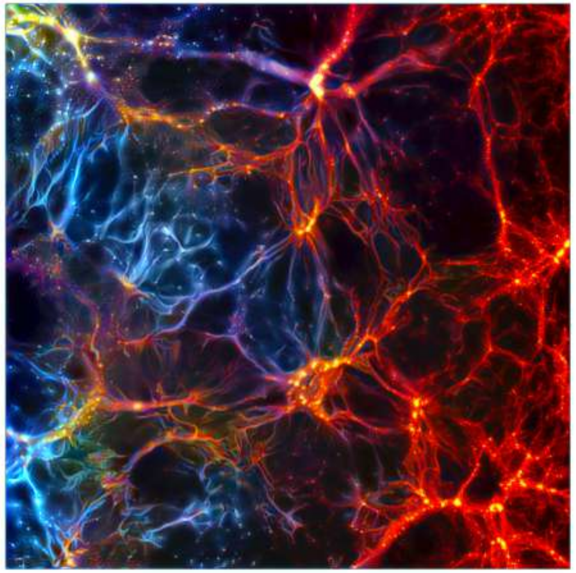
$$\dot{m}_{\rm out} = \overset{A_{\rm SN1}}{\boxed{\epsilon_{\rm SN}}} \left(\frac{V_0}{V_{\rm c}}\right)^{\overset{A_{\rm SN2}}{\boxed{\alpha_{\rm rh}}}} \dot{m}_*.$$

$$\dot{m}_{\rm radio} = \overset{A_{\rm AGN1}}{\boxed{\kappa_{\rm radio}}} \left[\frac{kT}{\Lambda(T, Z_{\rm h})}\right] \left(\frac{M_{\rm BH}}{10^8\,{\rm M_\odot}}\right)$$

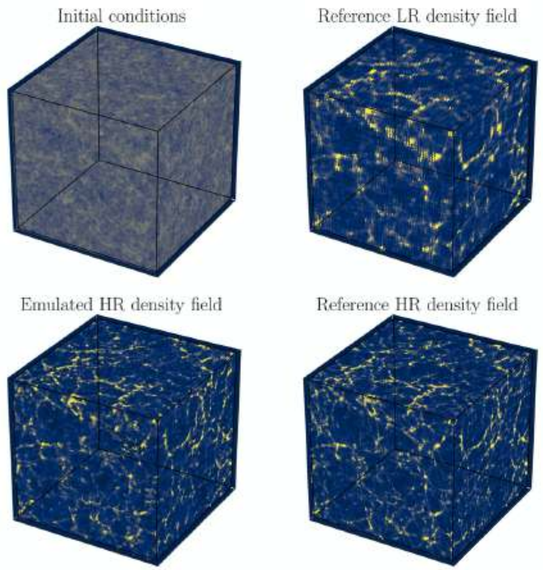


Steps towards field-level inference

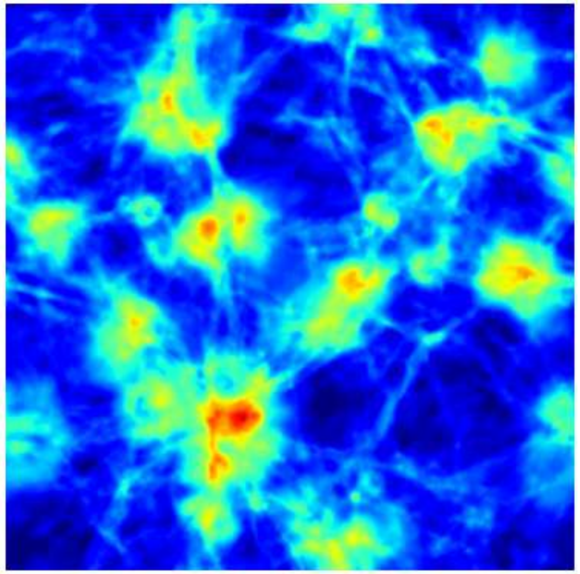
Quijote
Thousands of cosmologies



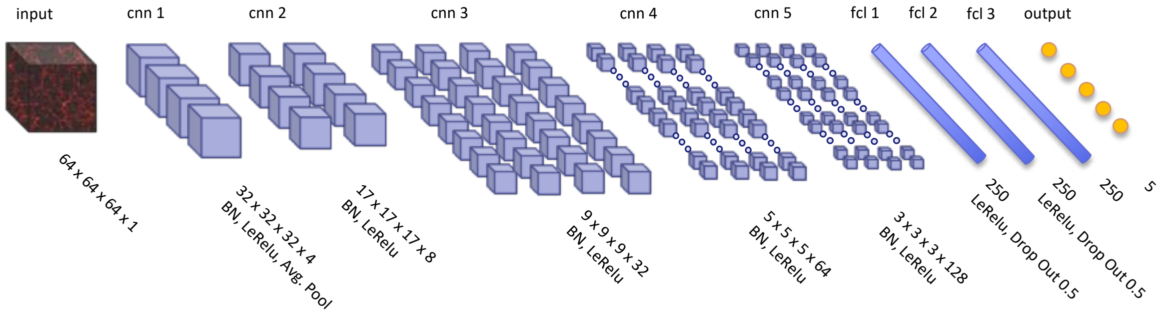
Super Resolution
$10^9\ \mathrm{h^{-1}M_\odot}$



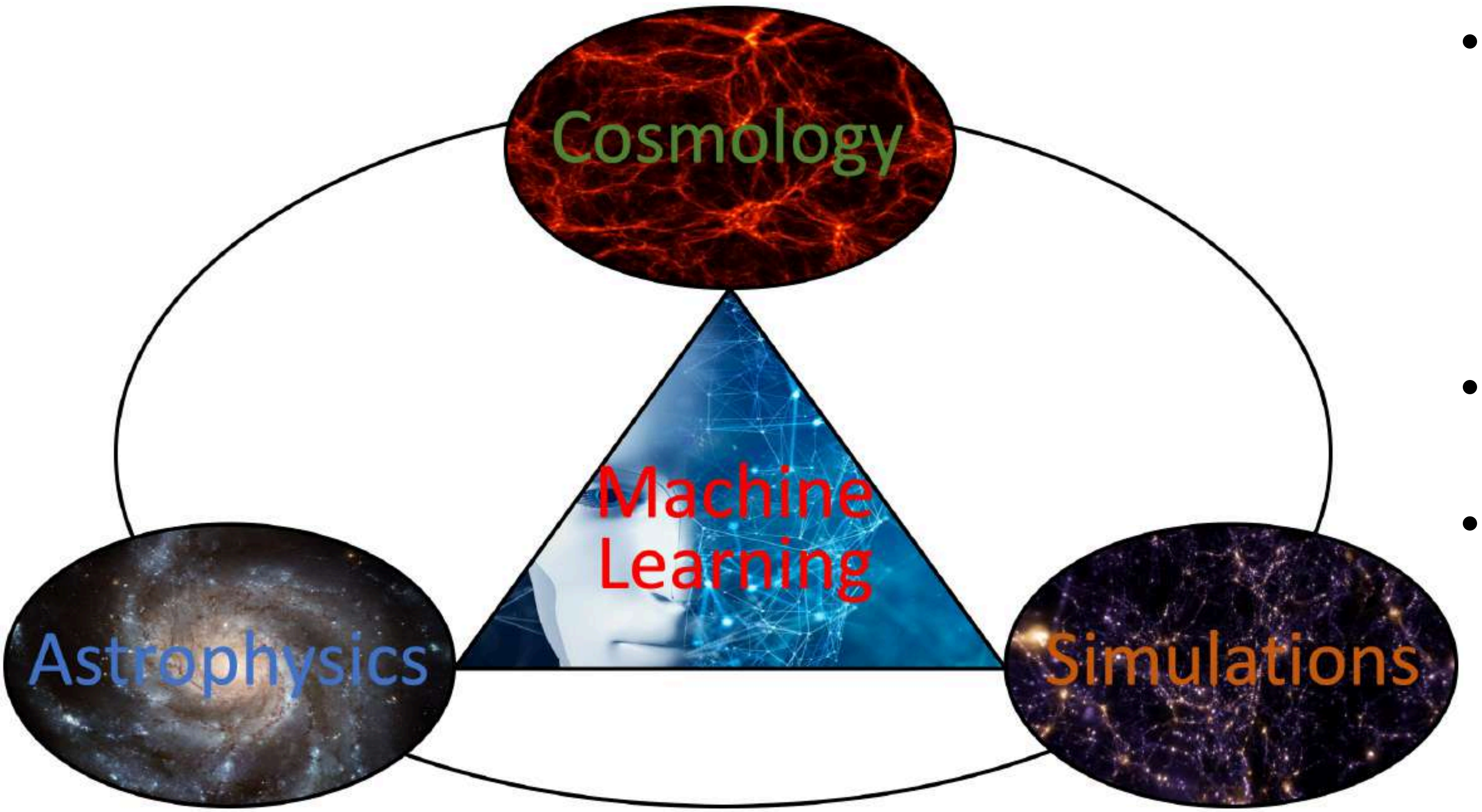
CAMELS
Thousands of astrophysics models



Likelihood-free inference
Extract all information. Marginalize over baryonic effects



Summary



- CAMELS: a large dataset to create connections between cosmology and galaxy formation through machine learning
- Hundreds of Tb publicly available!
- Many interesting results: field level LFI with CNNs, GNNs, first constrain on the mass of MW and M31, cosmology with one galaxy?
- Robustness is a major issue