Simulation-based inference for Cosmology

Francisco Villaescusa-Navarro

SIMONS FOUNDATION
FLATIRON INSTITUTE
PRINCETON UNIVERSITY

IAU-IAA seminar

February 14th 2023

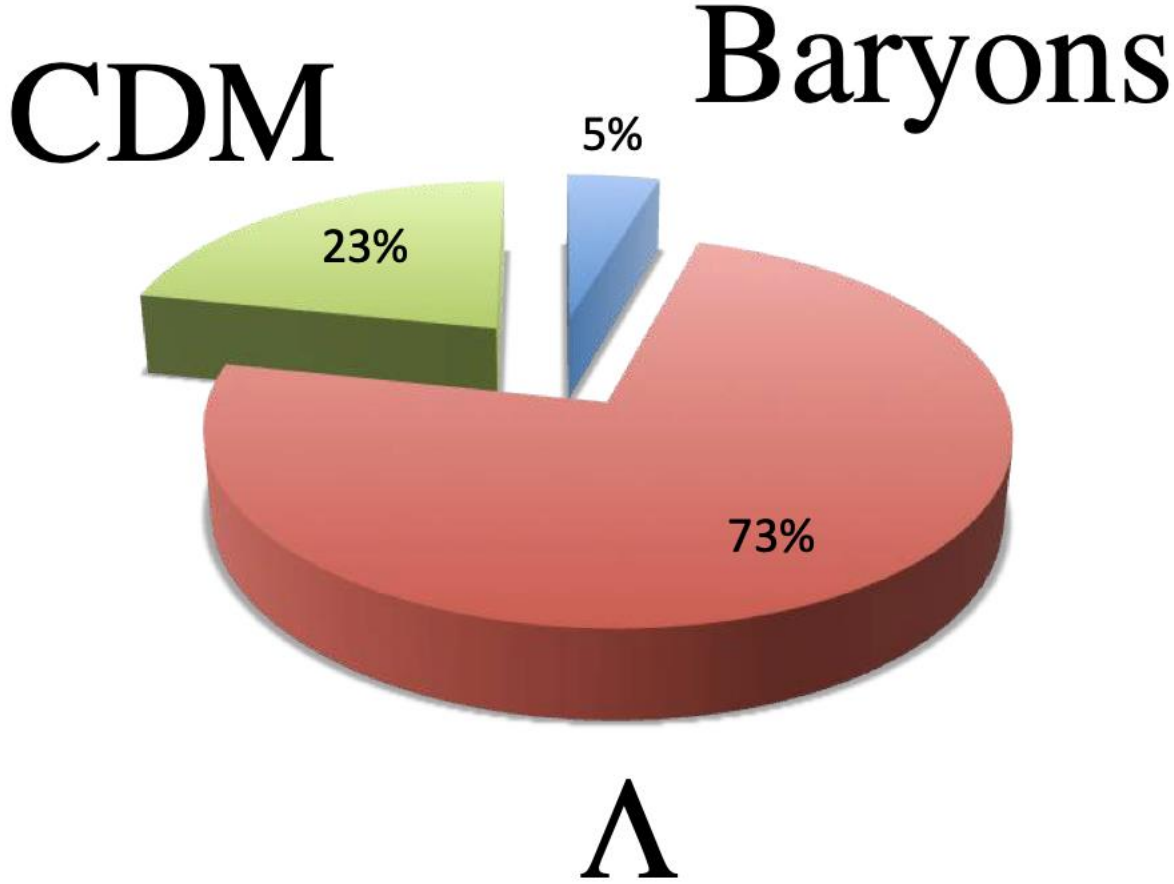
Outline

1. The ΛCDM model

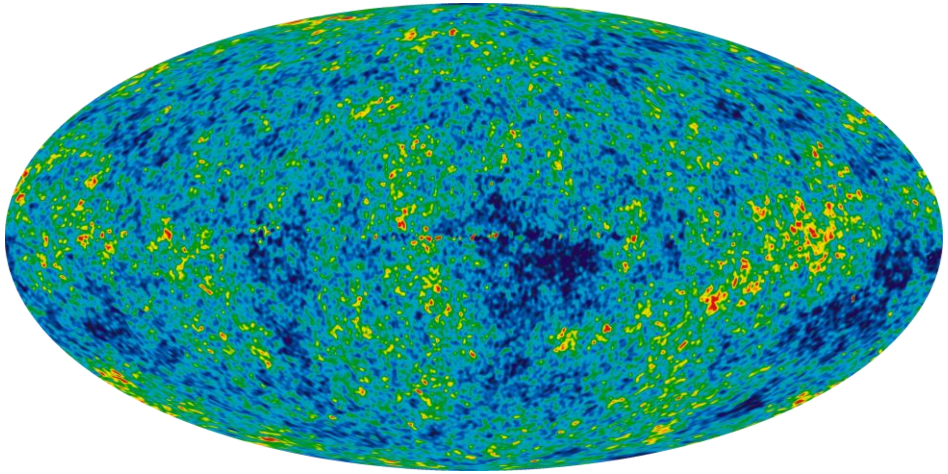
2. Parameter inference

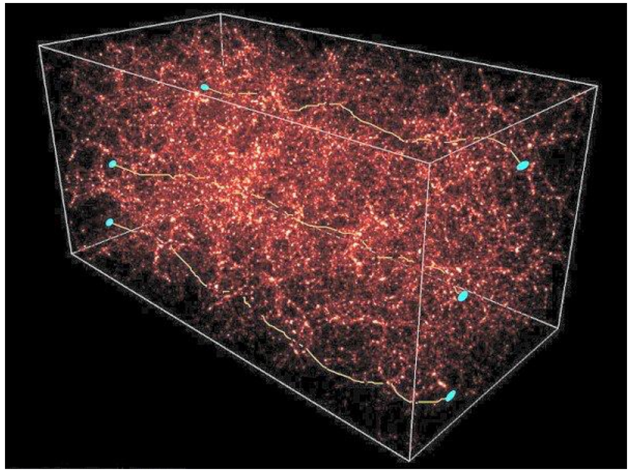
3. The role of AI

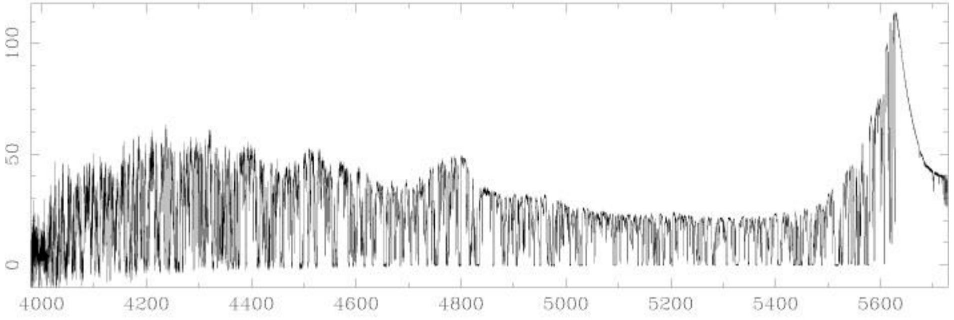
The ΛCDM model



The ΛCDM model







$$\Omega_m \pm \delta\Omega_m$$

$$\Omega_b \pm \delta\Omega_b$$

$$h \pm \delta h$$

$$n_s \pm \delta n_s$$

$$\sigma_8 \pm \delta\sigma_8$$

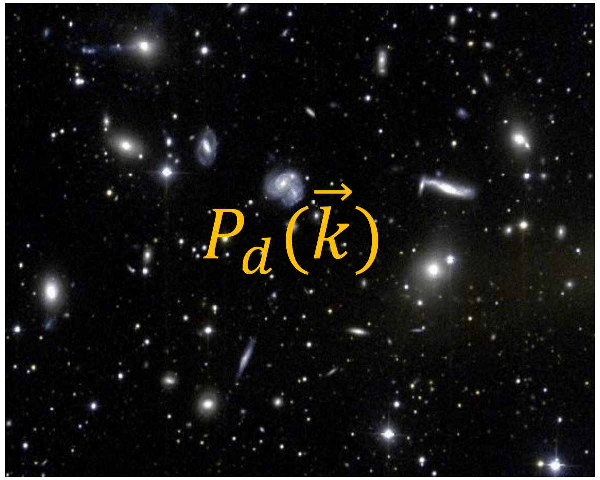
$$M_\nu \pm \delta M_\nu$$

$$w_0 \pm \delta w_0$$

$$w_a \pm \delta w_a$$

$$N_{\mathrm{eff}} \pm \delta N_{\mathrm{eff}}$$

Parameter inference

Observations	Theory
	$P_t(\vec{k}\|\vec{\theta})$

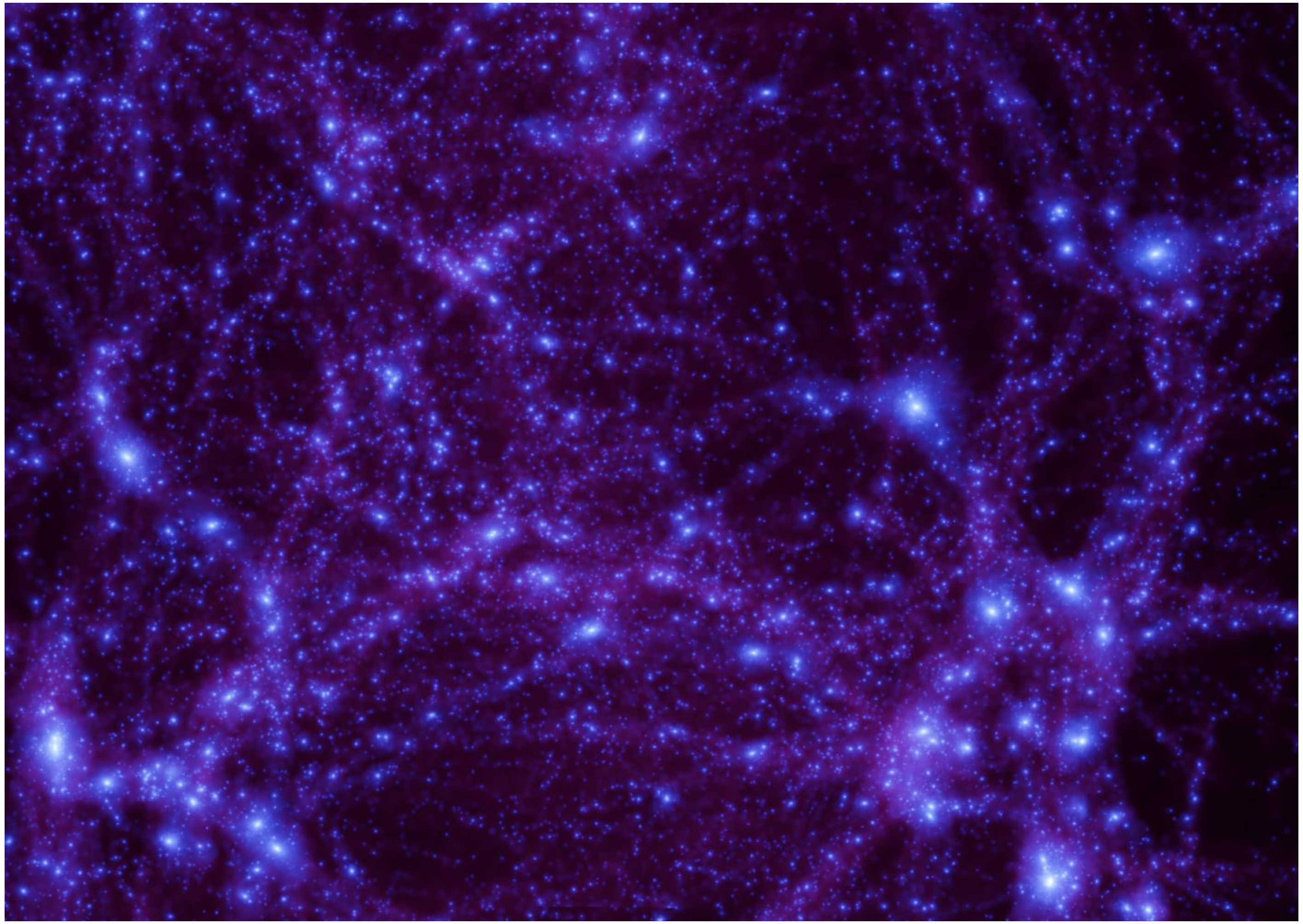
What summary statistics shall we use to determine $\vec{\theta}$ with the smallest error?

Power spectrum

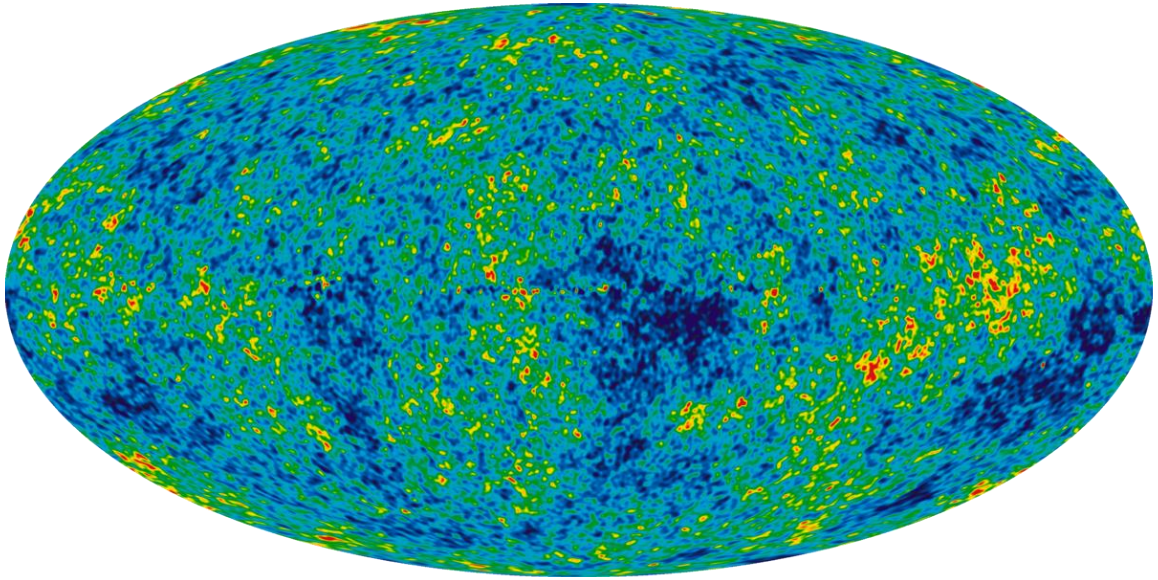
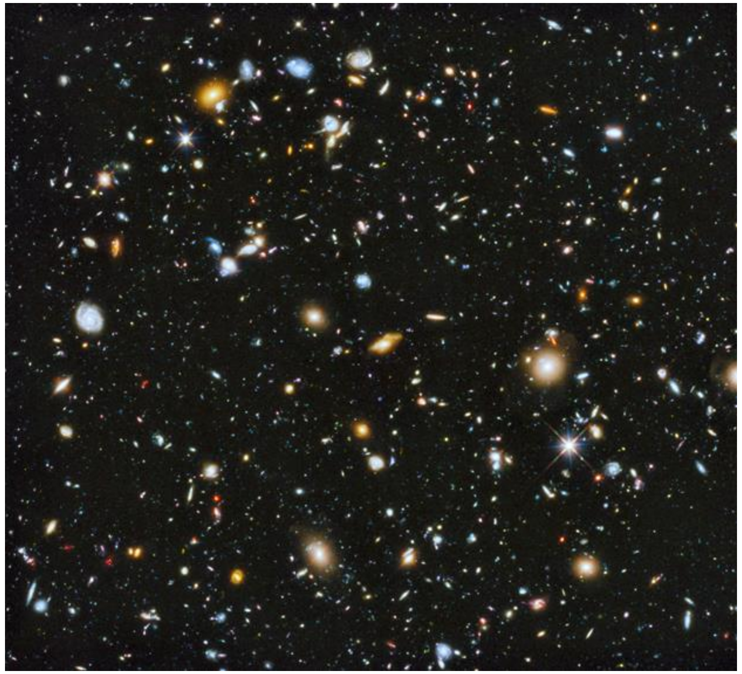
$$\delta(\vec{x}) = \frac{\rho(\vec{x})}{\bar{\rho}}$$

$$\delta(\vec{k}) = \int \delta(\vec{x}) e^{-i\vec{k}\,\vec{x}}$$

$$P(k) = \frac{1}{N_k} \sum_{k \in [k_i + \Delta k]} \delta(\vec{k}) \delta^*(\vec{k})$$

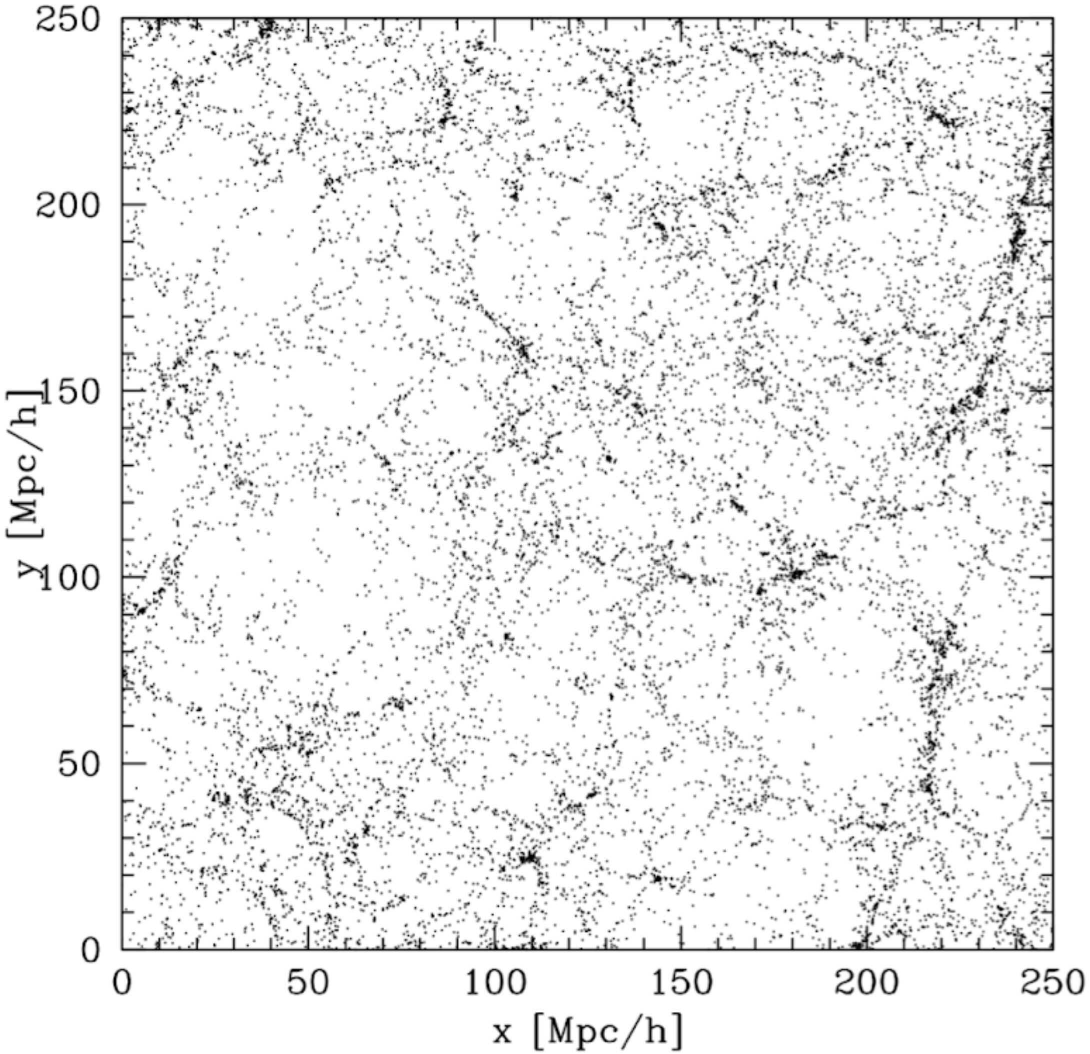


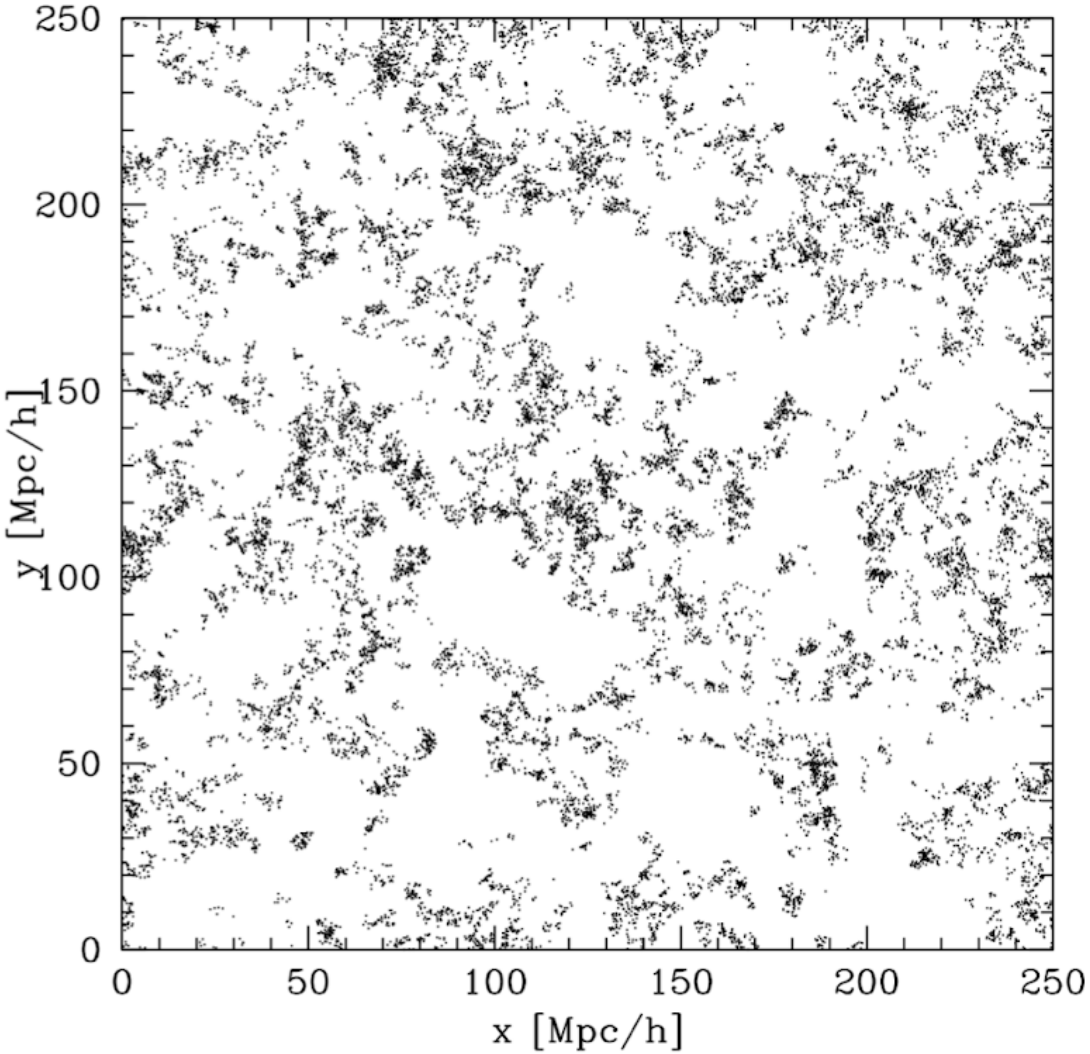
Parameter inference: summary statistics

Gaussian density field	**Non-Gaussian density field**
	
Optimal statistic: power spectrum	Optimal statistic: ???

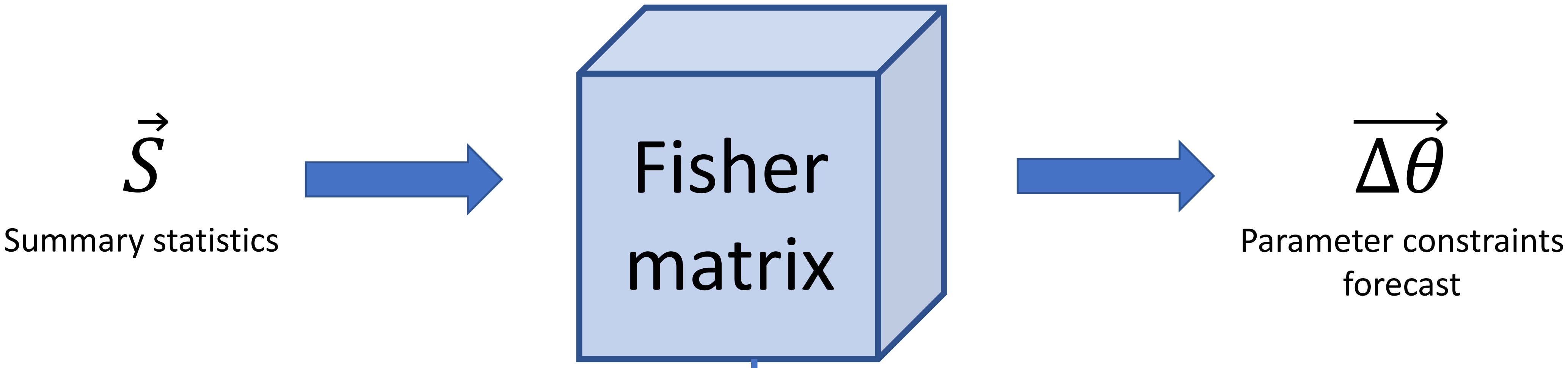
The standard method: power spectrum

Sefusatti & Scoccimarro 2004





Parameter inference: information content



$$F_{\alpha\beta} = \frac{\partial \vec{S}}{\partial \theta_\alpha} C^{-1} \frac{\partial \vec{S}}{\partial \theta_\beta}$$

$$\delta\theta_\alpha \geq (F^{-1})_{\alpha\alpha}$$

The Quijote Simulations

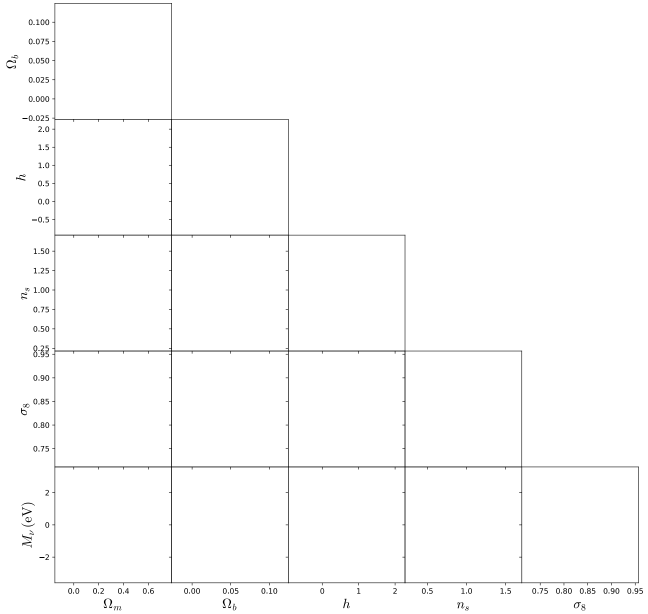
(https://quijote-simulations.readthedocs.io)

- A set of 47,500 full N-body simulations
- More than 9,000 cosmologies in $\{\Omega_m, \Omega_b, h, n_s, \sigma_8, M_\nu, w, f_{NL}\}$ hyperplane
- More than 10 trillion particles over a volume larger than entire observable Universe
- Catalogues with billions of halos, voids and galaxies: Molino and Gigantes datasets
- 35 Million CPU hours; 1 Petabyte of data
- Everything publicly available

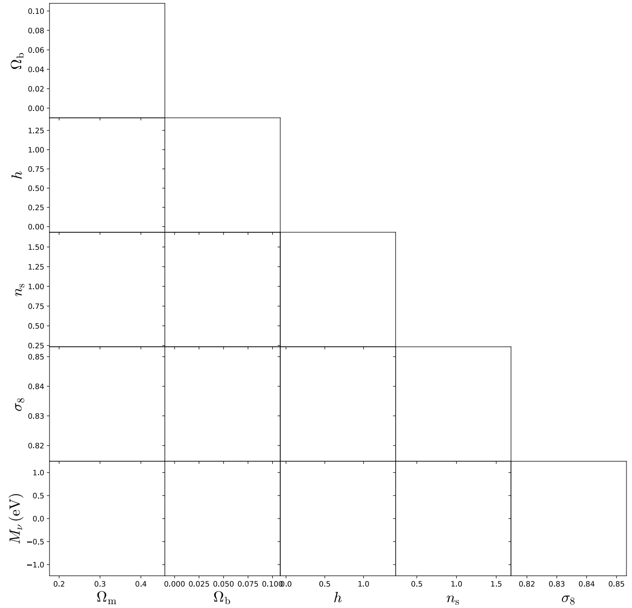
Information content

See https://quijote-simulations.readthedocs.io/en/latest/publications.html for more works

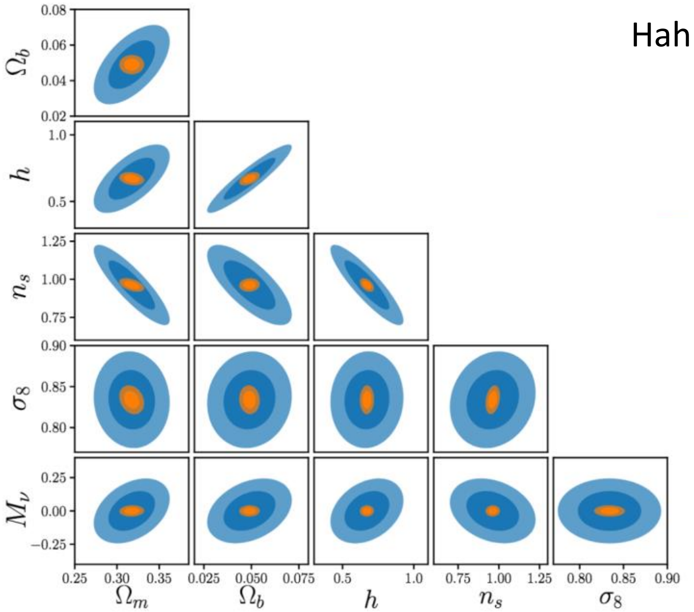
Bayer, FVN et al. 2021



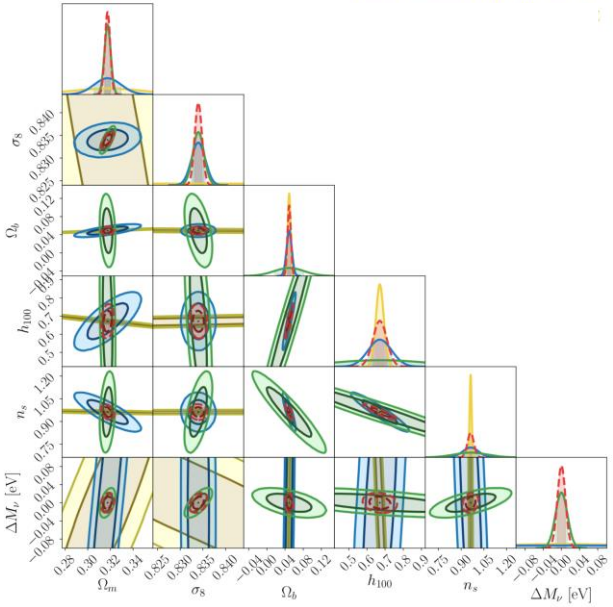
Massara, FVN et al 2020



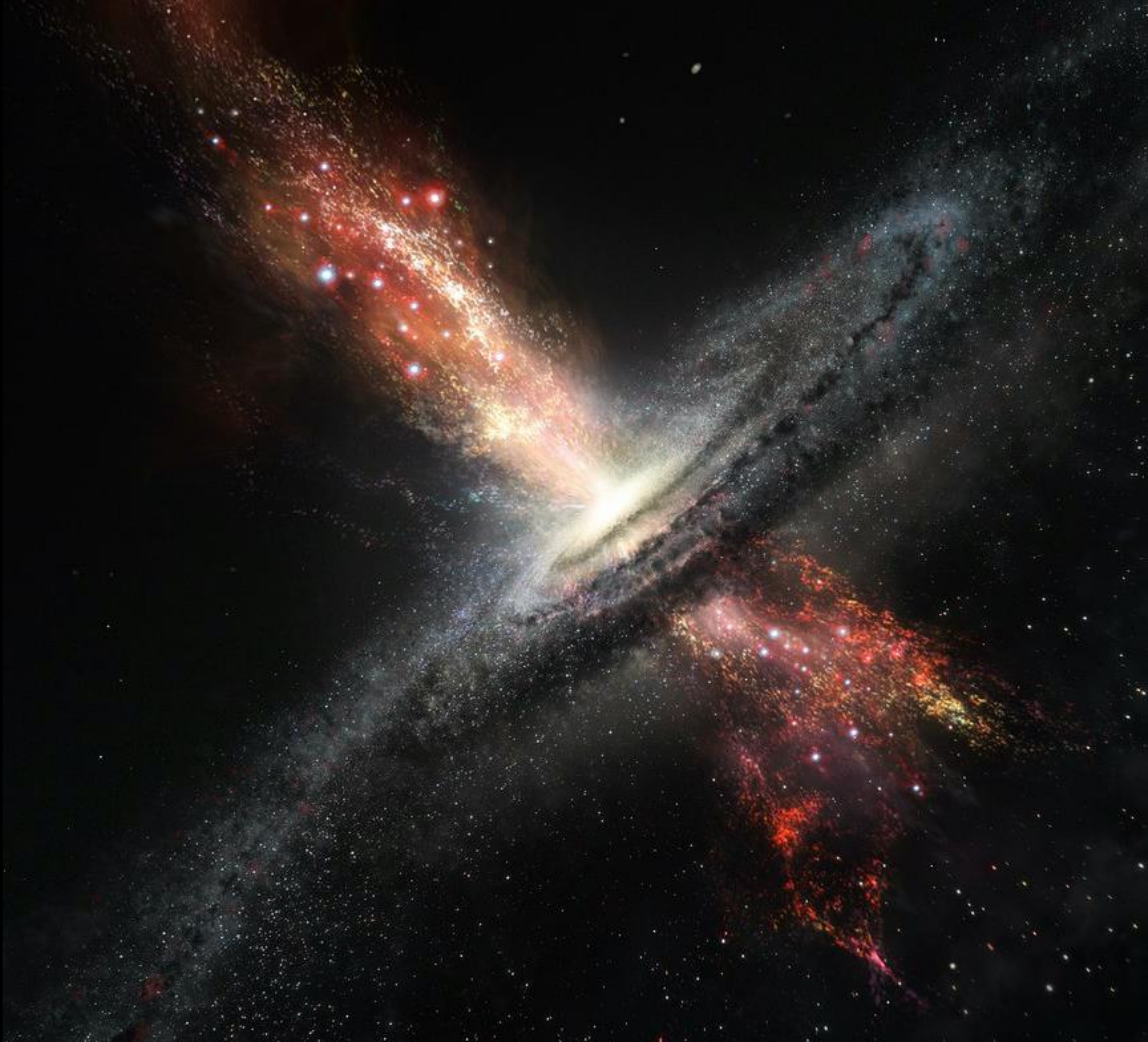
Hahn, FVN, Castorina, Scoccimarro 2019



Uhlemann, Friedrick, FVN, et al. 2019



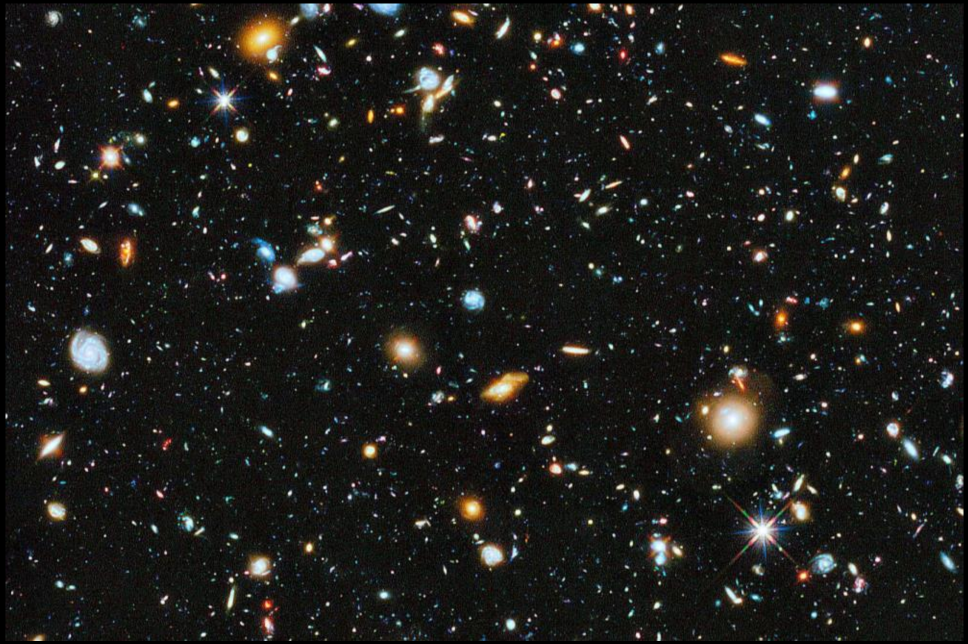
Generic conclusion:
Lots of information on small scales beyond P(k)



Benefits: Lots of information

Problems: Non-linearities & baryonic effects

The problem



Goal: extract the maximum information from surveys

Summary

1. <u>The ΛCDM model</u>

2. <u>Parameter inference</u>

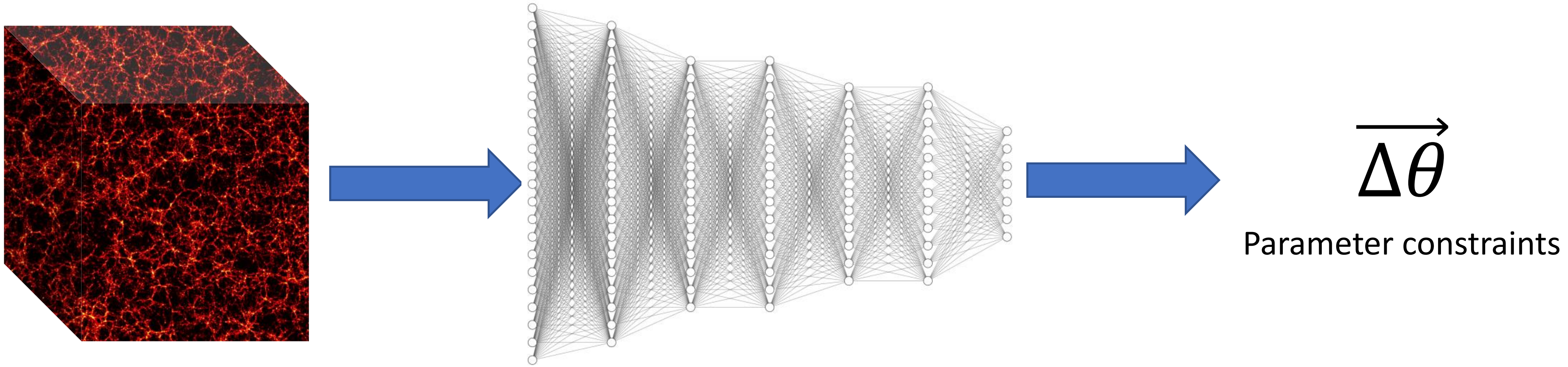
3. The role of AI

Neural networks

$$\mathrm{Output} = f(\mathrm{input})$$

Universal approximation theorem

The role of Artificial Intelligence



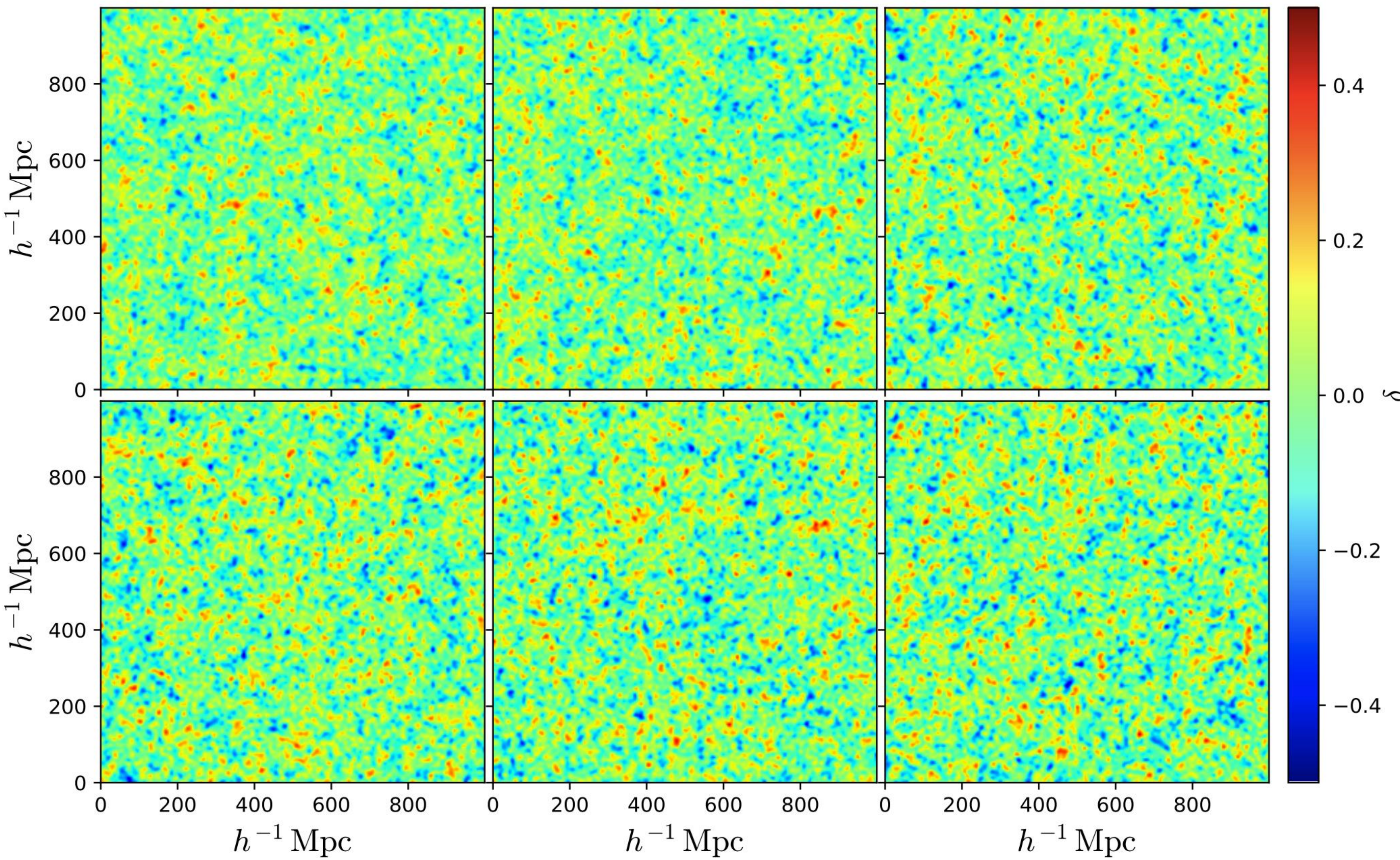
Toy Example: Gaussian density fields

FVN et al. 2020

$$P(k) = A/\sqrt{k}$$

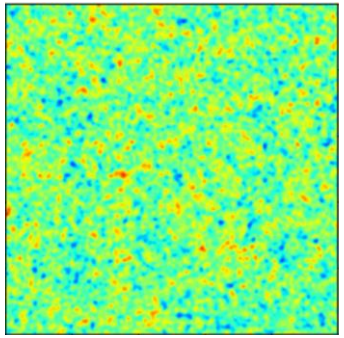
A is the cosmological parameter

Goal: infer value of A from the map itself



Toy Example: Gaussian density fields

FVN et al. 2020

 → Neural Network → A

Input 100,000 maps with A=1

Error on A calculated
with Fisher matrix

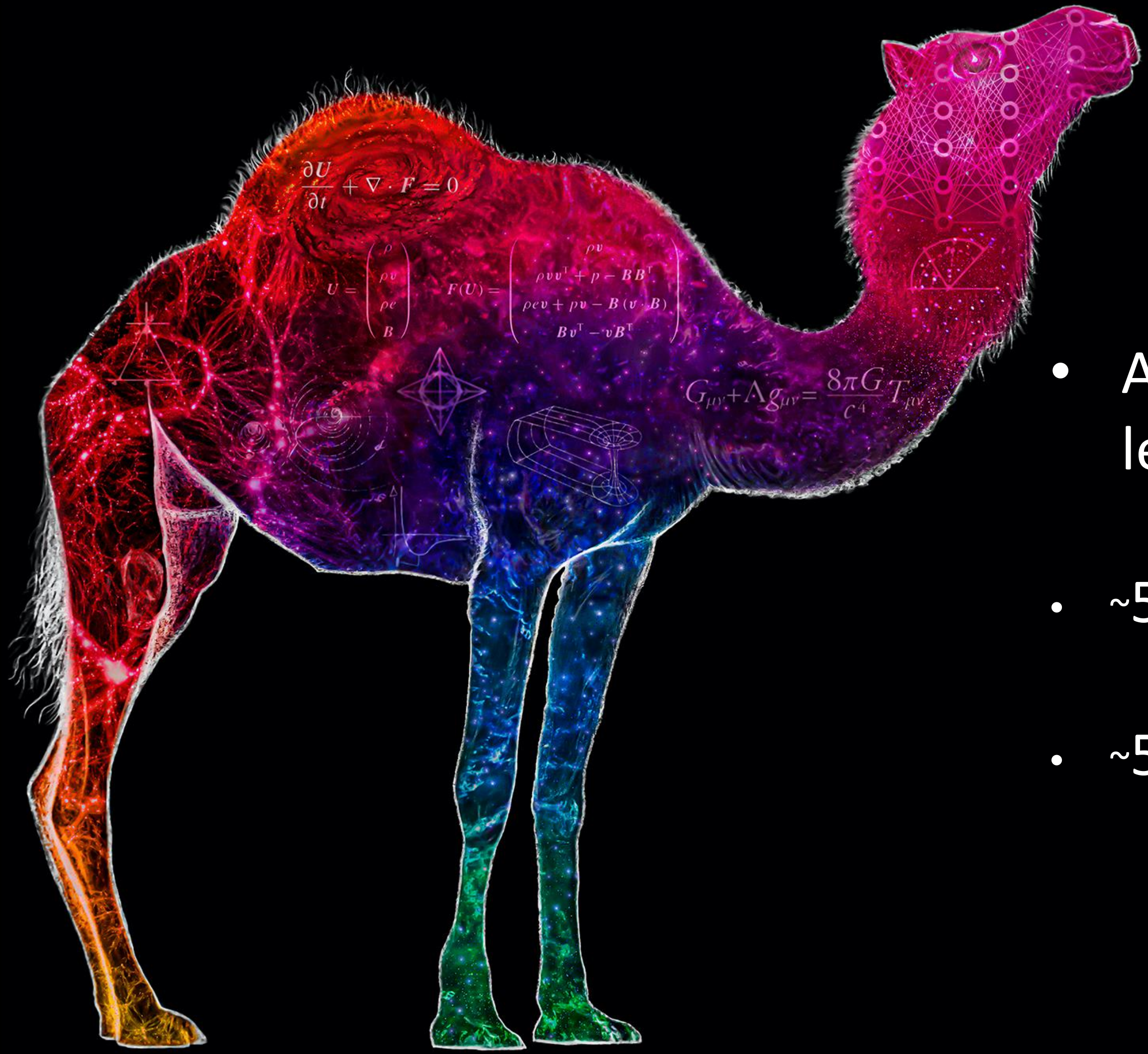
$$\sigma_A = A\sqrt{\frac{2}{N_{\text{pixels}}}}$$

$$A = \Bigg\{$$

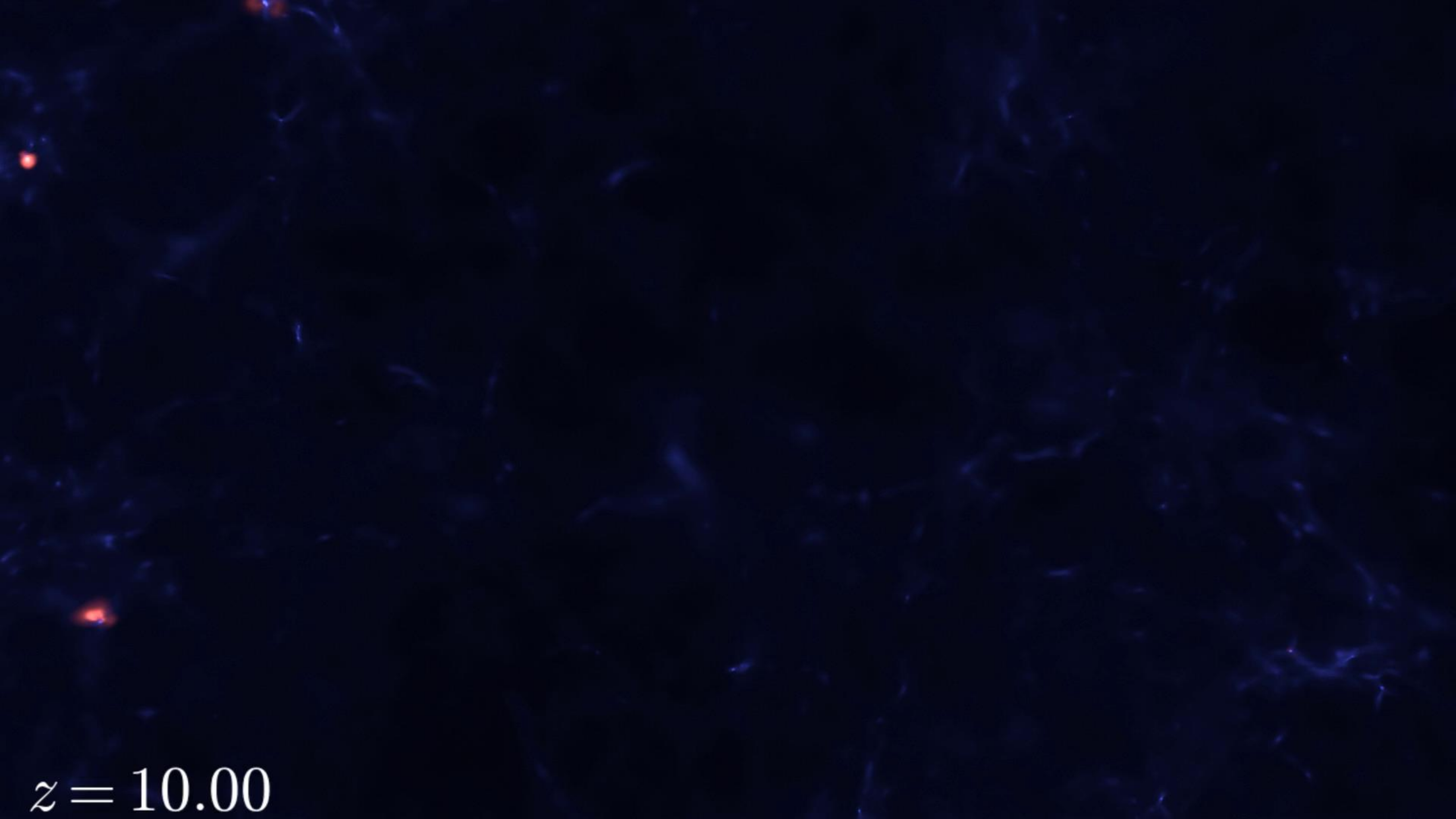
CAMELS

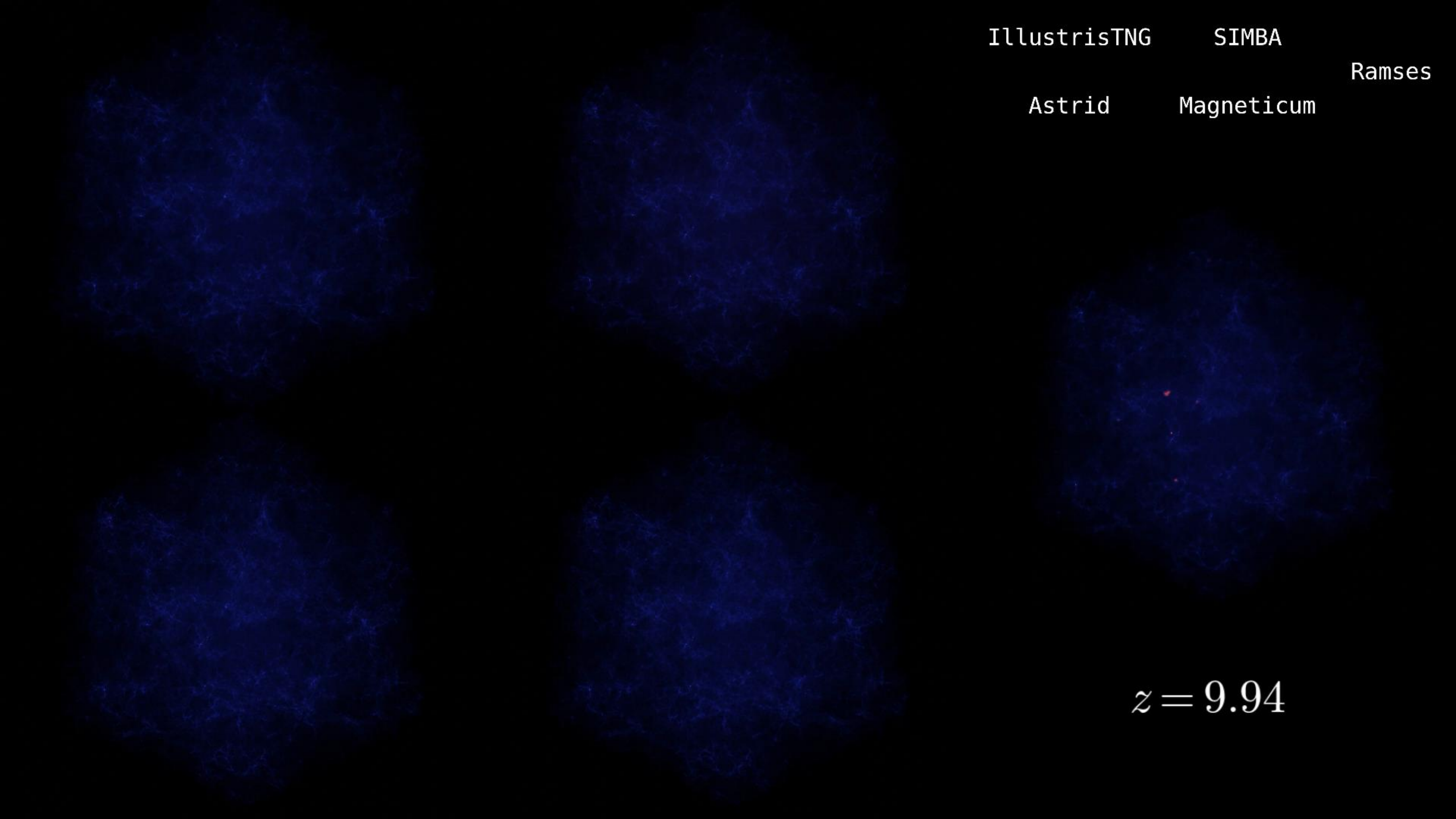
https://www.camel-simulations.org

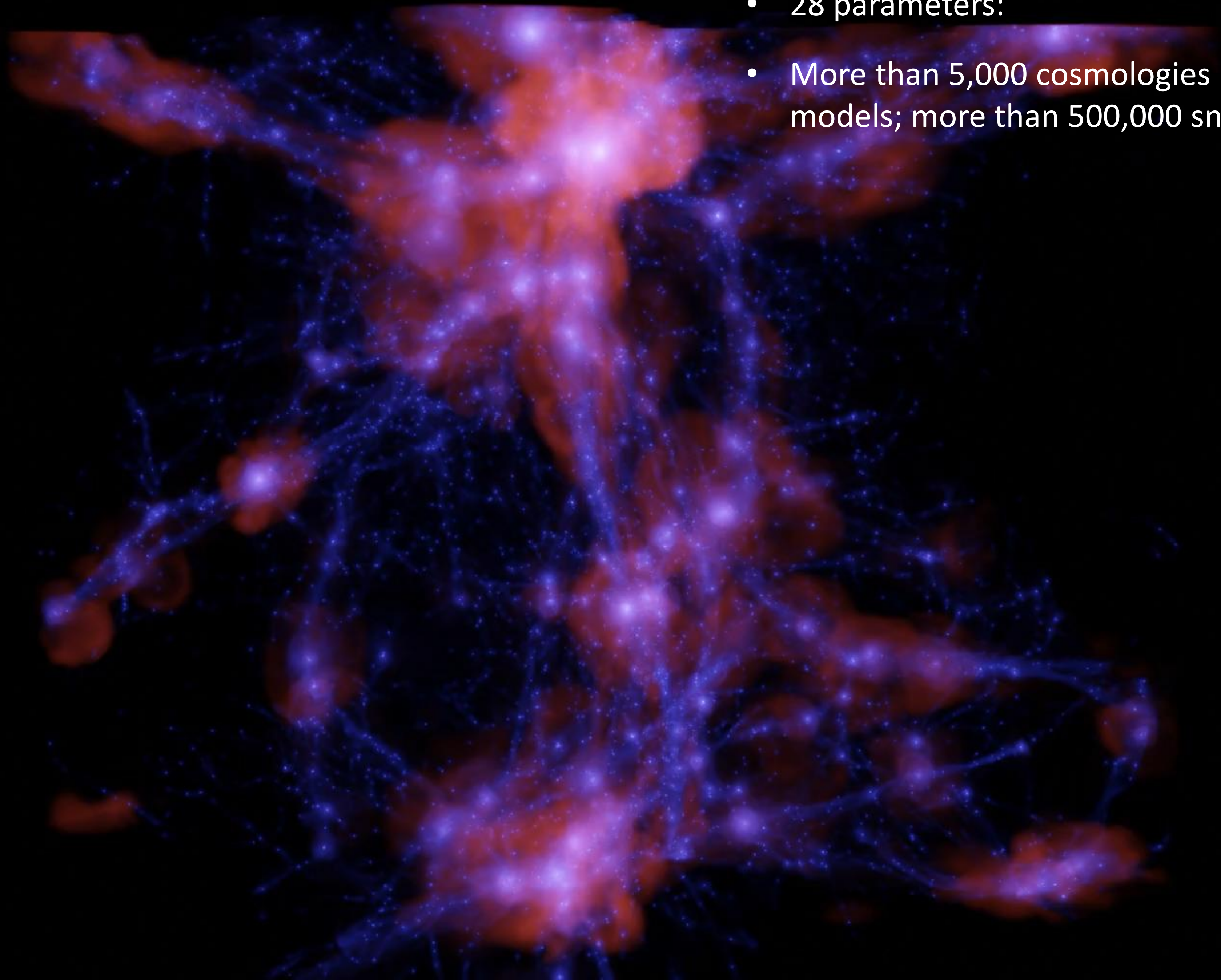
Cosmology and Astrophysics with MachinE Learning Simulations



- A suite of ~10,000 simulations designed for machine learning applications
- ~5,000 N-body; Gadget-III
- ~5,000 state-of-the-art (magneto-)hydrodynamic sims

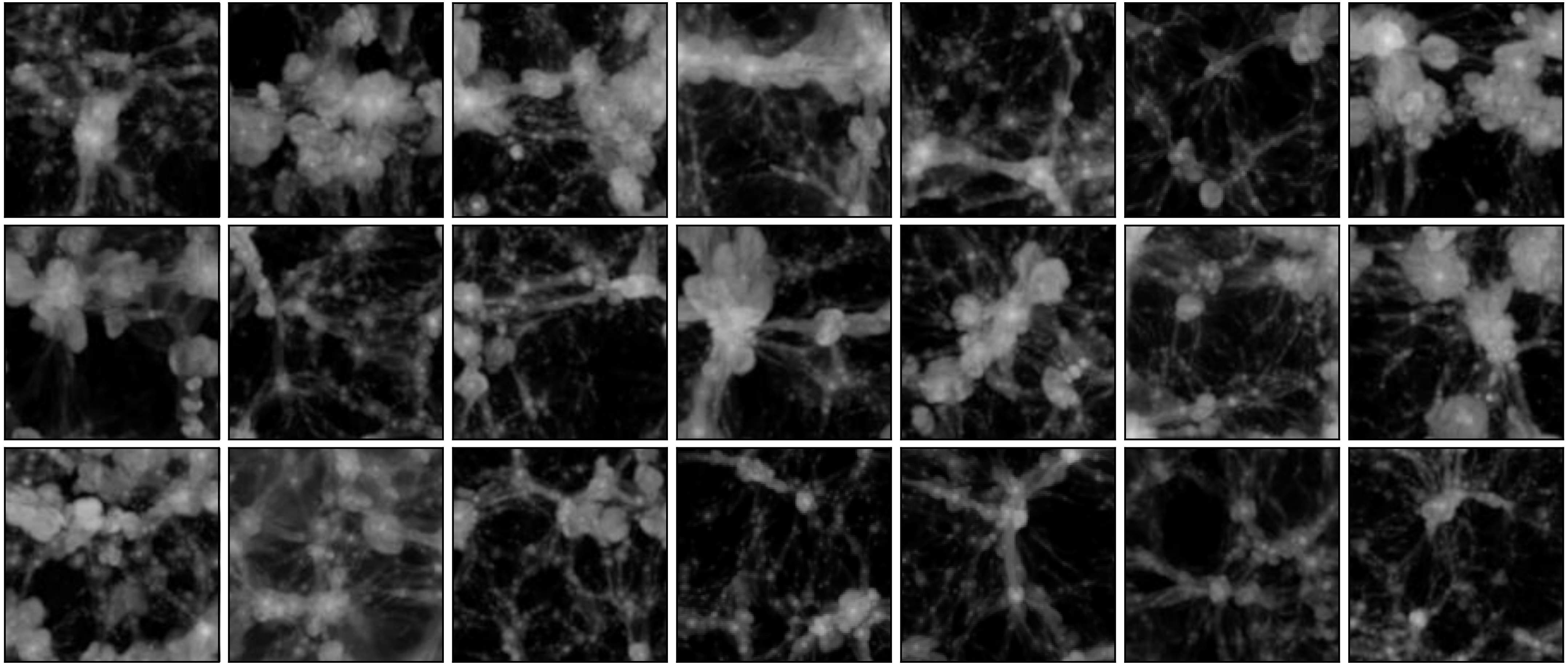

$z = 10.00$


IllustrisTNG
SIMBA
Ramses
Astrid
Magneticum
$z = 9.94$



- 6 parameters: $\{\Omega_m, \sigma_8, A_{\mathrm{SN1}}, A_{\mathrm{SN2}}, A_{\mathrm{AGN1}}, A_{\mathrm{AGN2}}\}$
- 28 parameters:
- More than 5,000 cosmologies & astrophysics models; more than 500,000 snapshots

Example I: Gas temperature



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

Inference with AI

Moments network
Jeffrey & Wandelt 2020
(2011.05991)

$$\mathcal{L} = \sum_{i=1} \log \left(\sum_{j \in \mathrm{batch}} (\theta_{i,j} - \mu_{i,j})^2 \right) + \sum_{i=1} \log \left(\sum_{j \in \mathrm{batch}} \left((\theta_{i,j} - \mu_{i,j})^2 - \sigma_{i,j}^2 \right)^2 \right)$$

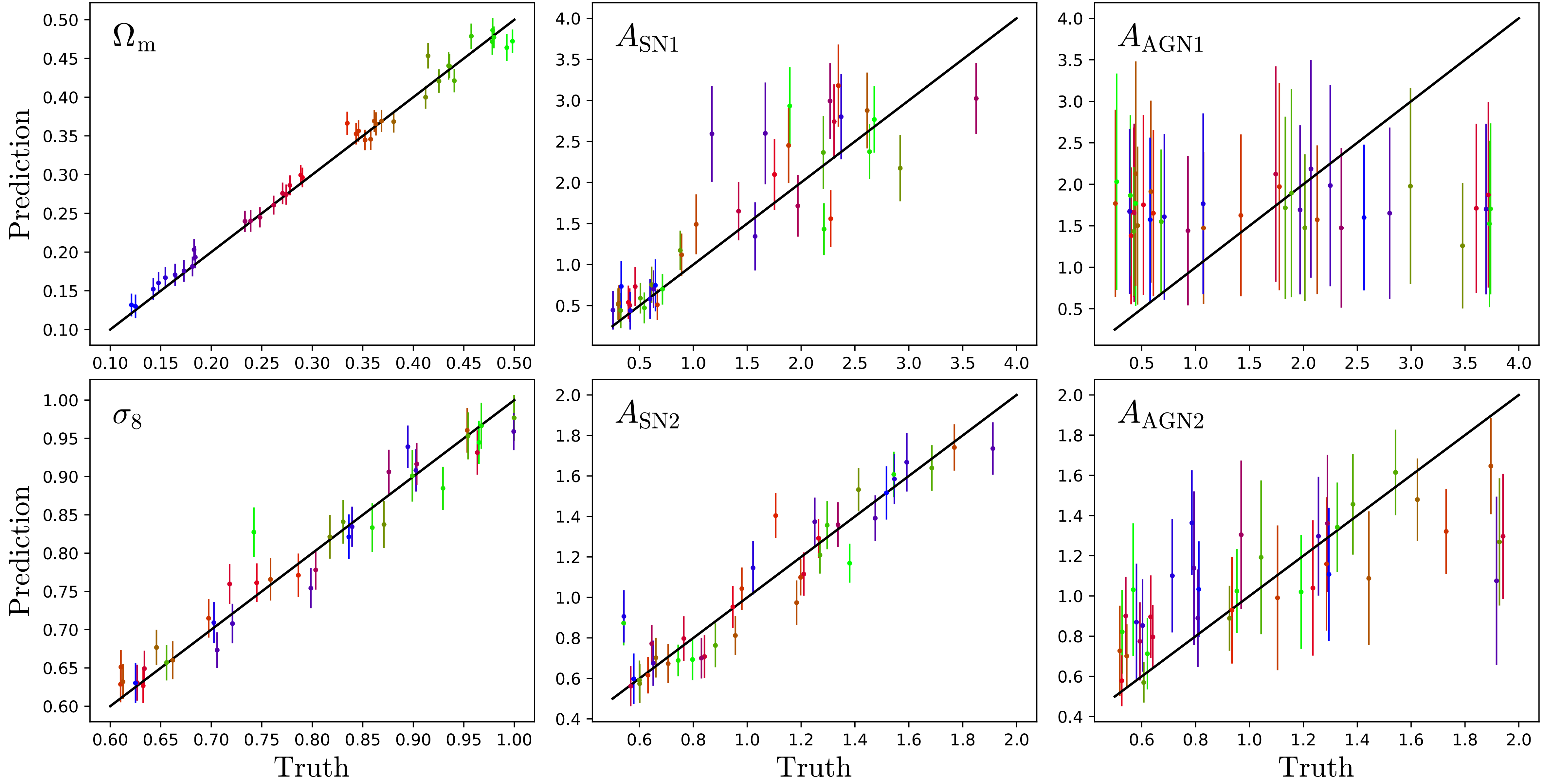
$$\mu_i(\mathbf{X}) = \int_{\theta_i} p(\theta_i|\mathbf{X})\theta_i d\theta_i \ ,$$

$$\sigma_i(\mathbf{X}) = \int_{\theta_i} p(\theta_i|\mathbf{X})(\theta_i - \mu_i)^2 d\theta_i \ ,$$

$$p(\theta_i|\mathbf{X}) = \int_{\theta} p(\theta_1, \theta_2, ...\theta_n|\mathbf{X}) d\theta_1 ... d\theta_{i-1} d\theta_{i+1} ... d\theta_n \ .$$

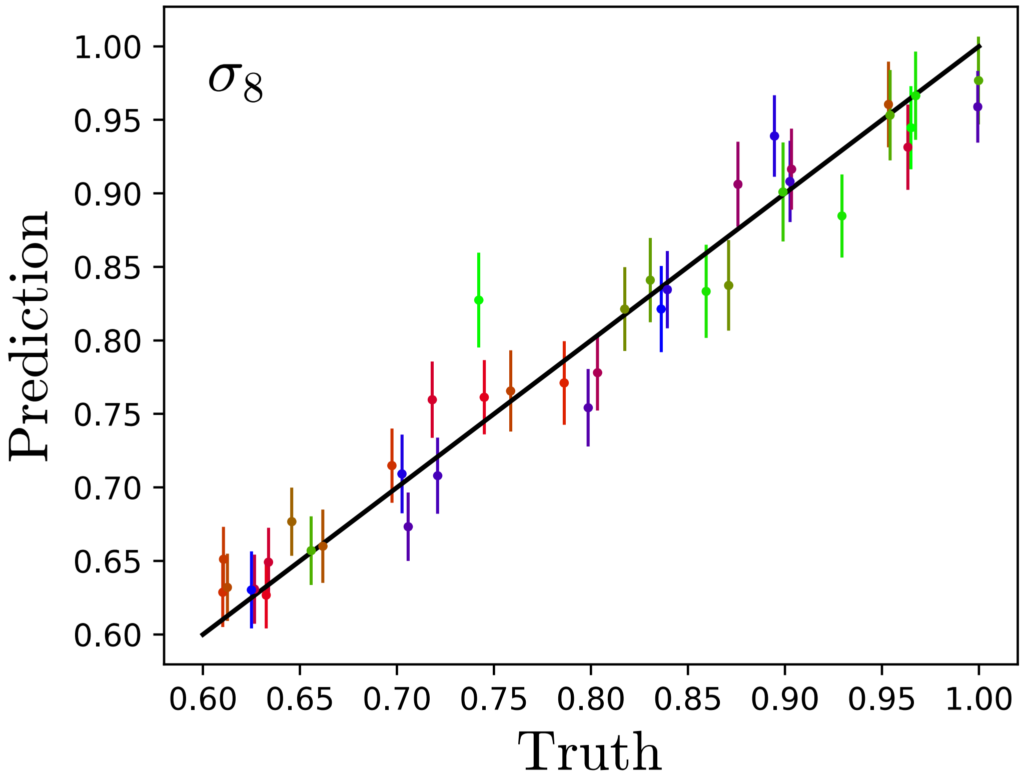
Likelihood-free inference: gas temperature

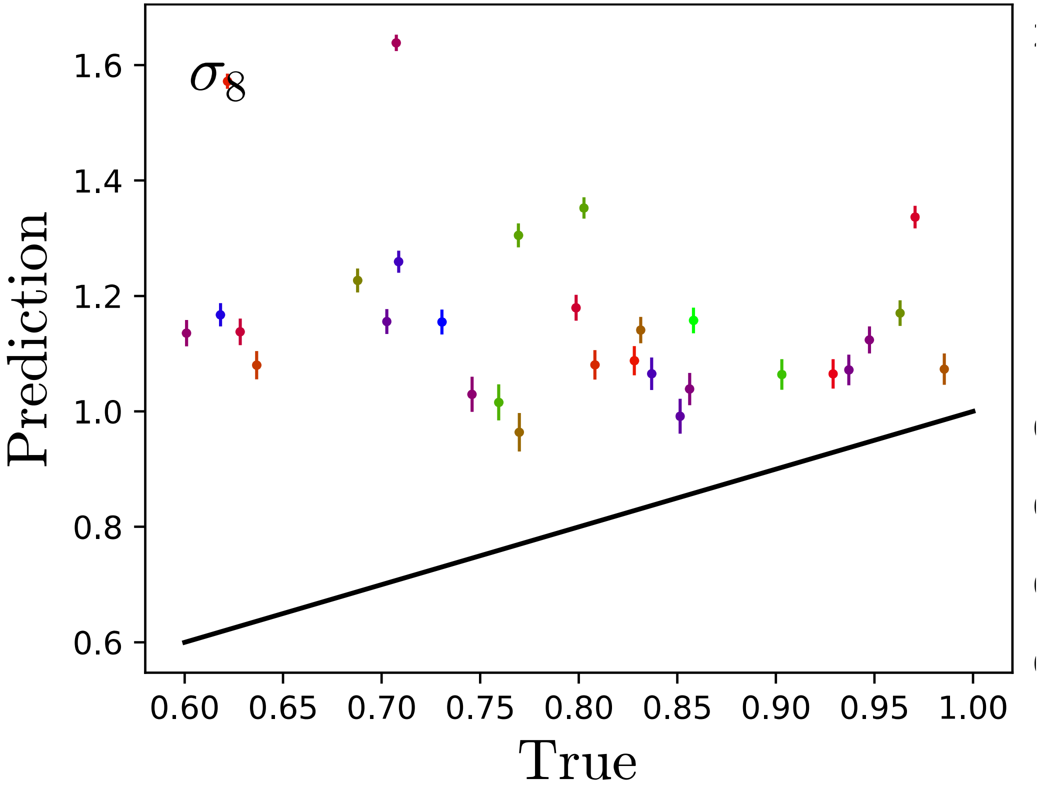
FVN et al. 2021a



Robustness: the AI evidence?

FVN et al. 2021a



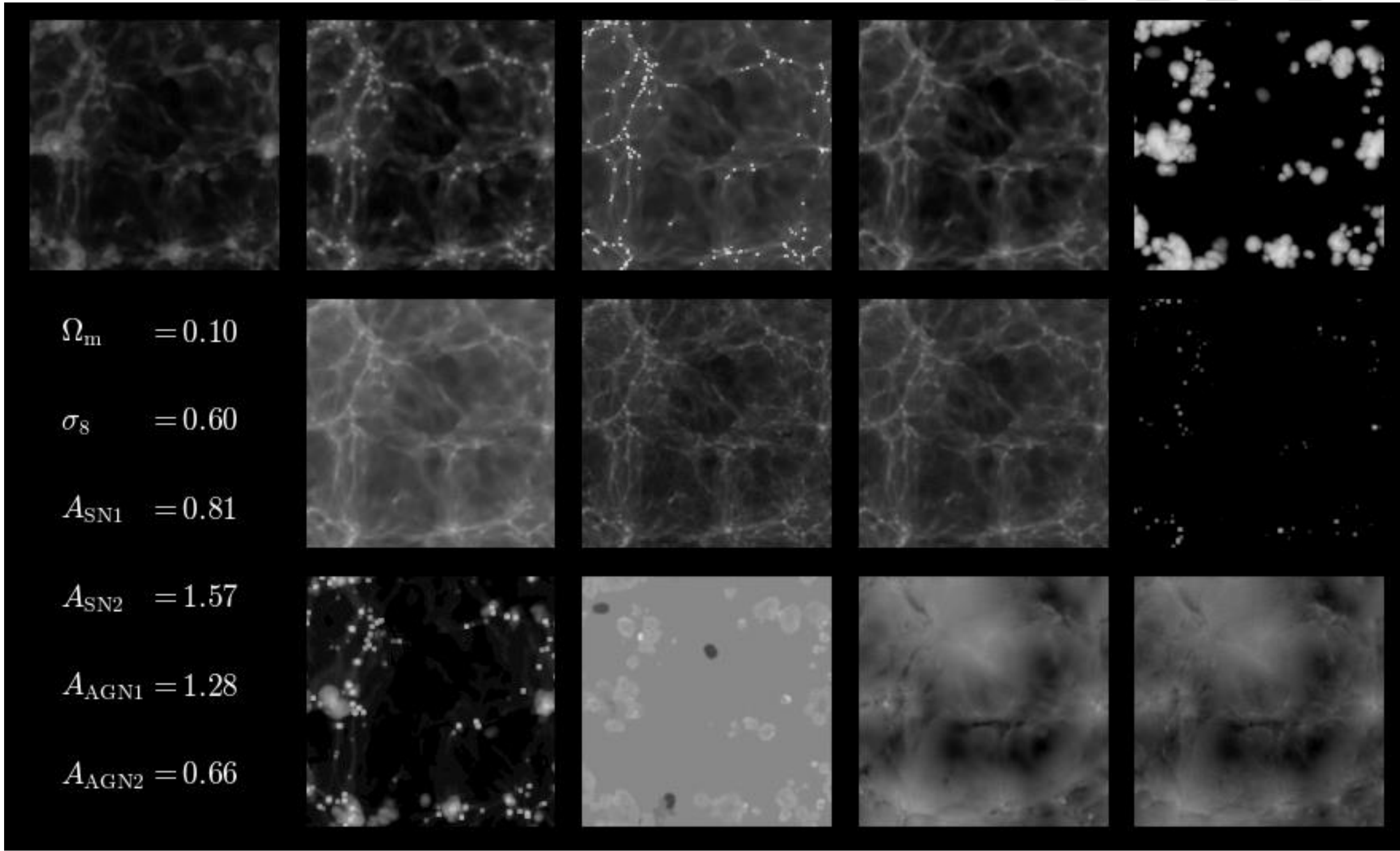


CAMELS Multifield Dataset

FVN et al. 2021c

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

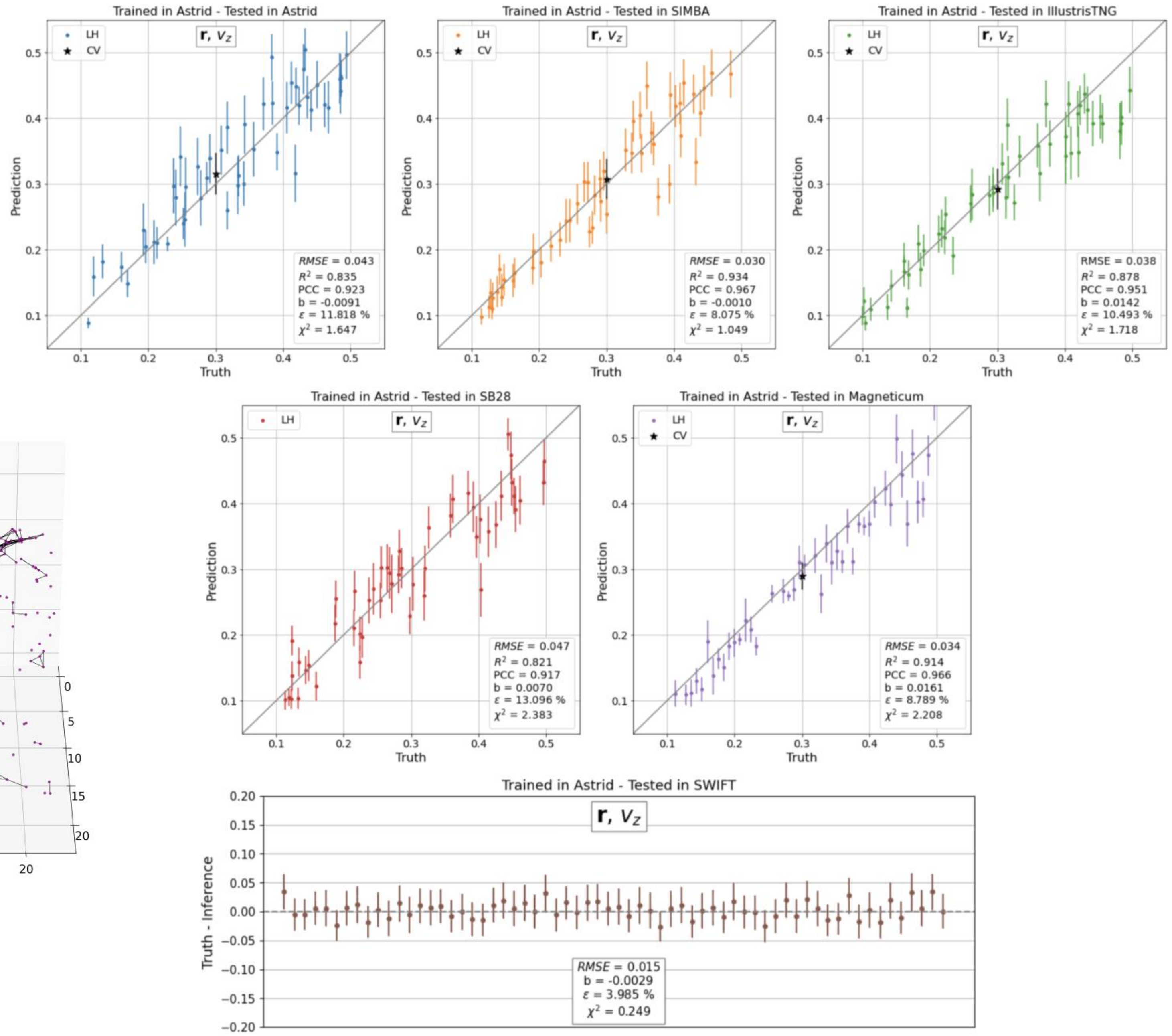
- Hundreds of thousands of labeled 2D maps and 3D grids
- Several redshifts: 0, 0.5, 1, 1.5, 2
- Three different resolutions
- 13 different fields:
  1. Gas density
  2. Gas temperature
  3. Gas metallicity
  4. Gas pressure
  5. Neutral hydrogen density
  6. Electron number density
  7. Dark matter 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

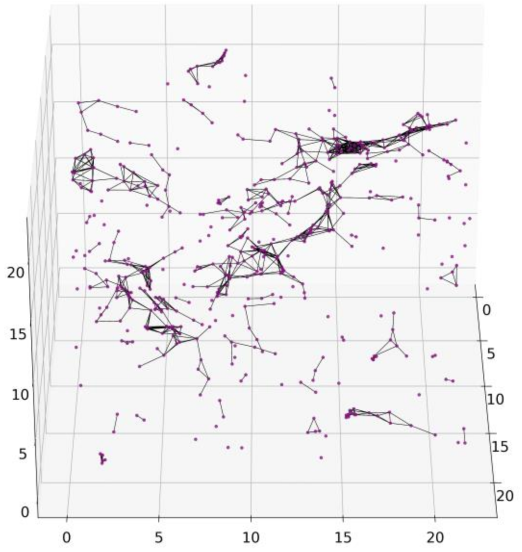


Cosmic graphs

Robust field-level likelihood-free inference with galaxy catalogues

De Santi et al. (in prep)
Shao et al. (in prep)





Natali de Santi
(Sao Paolo / Flatiron)

Helen Shao
(Princeton)

Robust field-level likelihood-free inference with galaxy catalogues

Derived a new universal equation to predict Ω_m from halo and galaxy catalogues.

Helen Shao
(Princeton)

Natali de Santi
(Sao Paolo / Flatiron)

$$1.32|v_i - v_j + 0.21| + 0.12(v_i - v_j) - 0.12(\gamma_{ij} + \beta_{ij} - 1.73)$$

$$|1.62(v_i - v_j) + 0.45| + 1.98(v_i - v_j) + 0.55$$

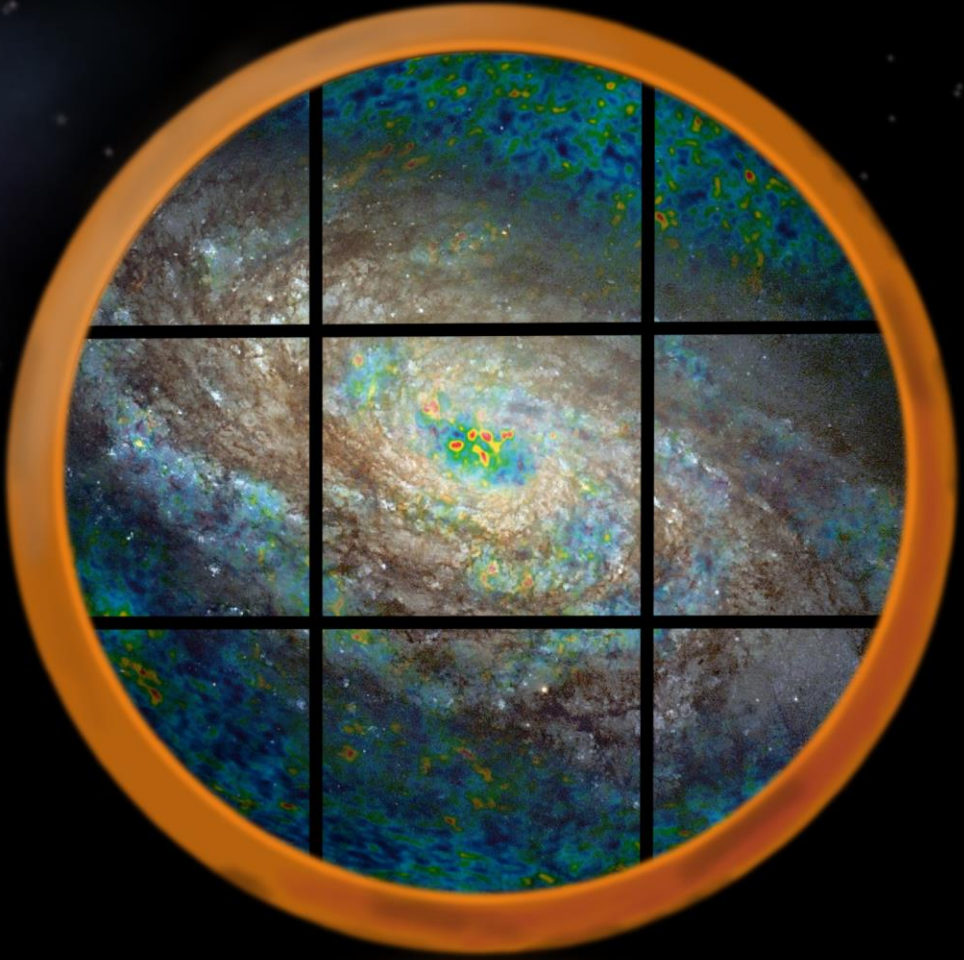
$$1.21^{v_i}(0.77^{3.29\sum_{j\in\mathcal{N}_j} e_1^{(1)} + \sum_{j\in\mathcal{N}_j} e_2^{(1)}}) + 0.12$$

$$0.78 - \sqrt{\log(0.16^{\sum_{j\in\mathcal{N}_j} e_2 + \sum_{j\in\mathcal{N}_j} e_1 - 0.41 v_i - 1.05})} + 1.45$$

$$a \cdot (-5.5\sum_{i\in\mathcal{G}} v_2^{(1)} + 2.21\sum_{i\in\mathcal{G}} v_1^{(1)} + |0.96\sum_{i\in\mathcal{G}} v_2^{(1)} + 0.82\sum_{i\in\mathcal{G}} v_1^{(1)}|) + b$$

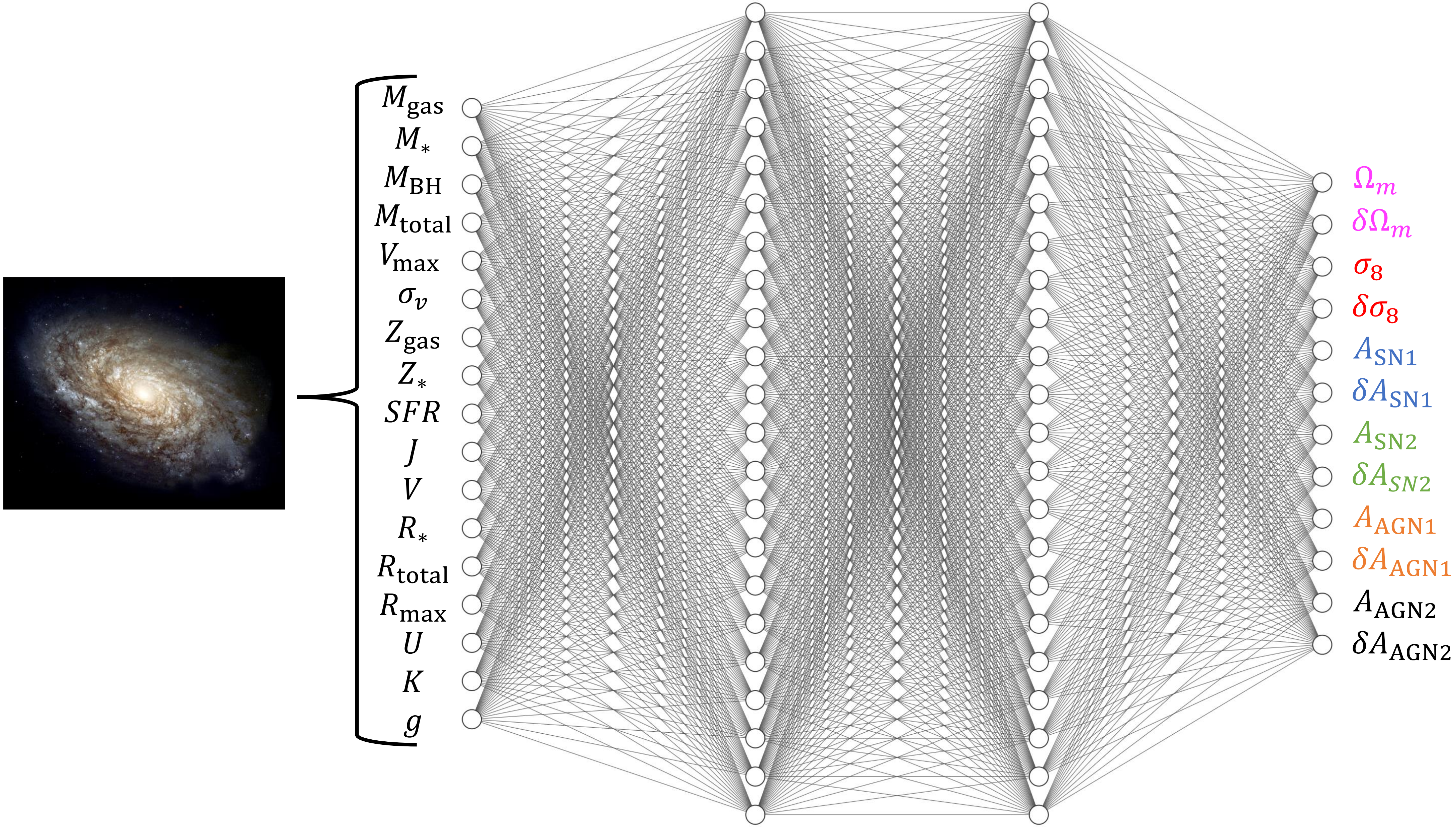
Cosmology with 1 galaxy?

FVN et al. 2022b

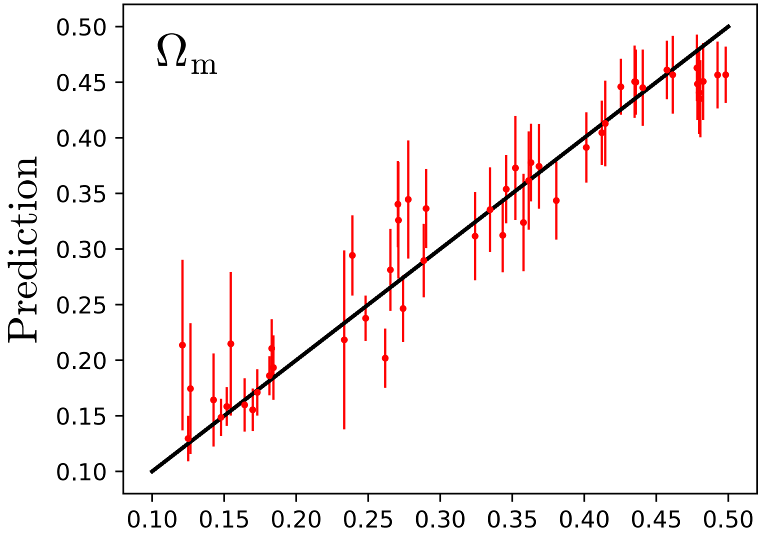


In collaboration with Jupiter Ding, Shy Genel, Stephanie Tonnesen, Valentina La Torre, David Spergel, Romain Teyssier, Yin Li, Caroline Heneka, Pablo Lesmos, Daniel Angles-Alcazar, Daisuke Nagai

Cosmology with 1 galaxy

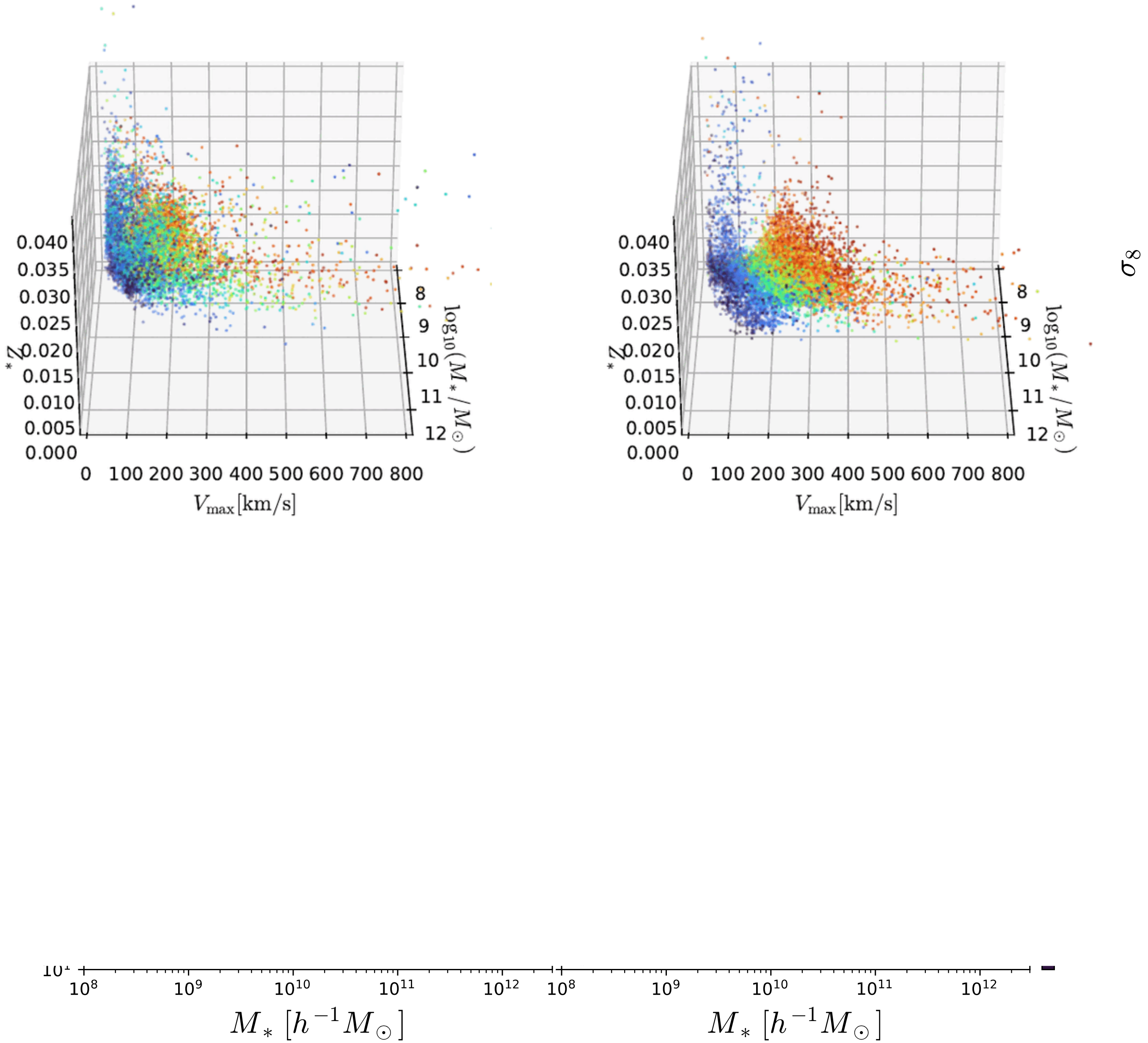


Cosmology with 1 galaxy



Cosmology with 1 galaxy

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



Conclusions

- We need to read ALL the information to decipher the mysteries of the Universe
- We don’t know how to read all information
- Neural networks can approximate generic functions
- We can learn about our own Universe by generating many fake ones

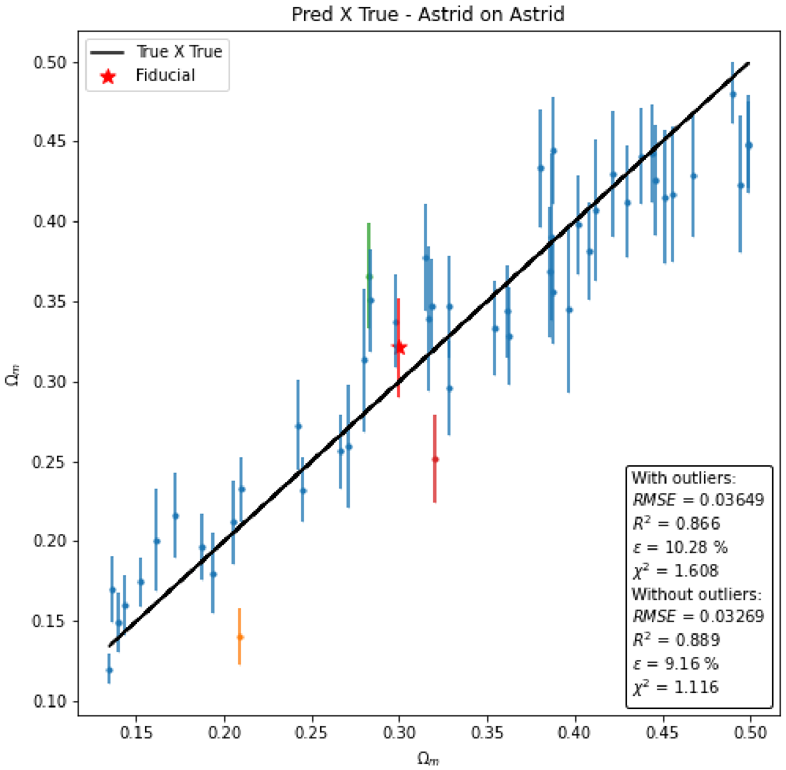
Robust field-level inference: galaxies

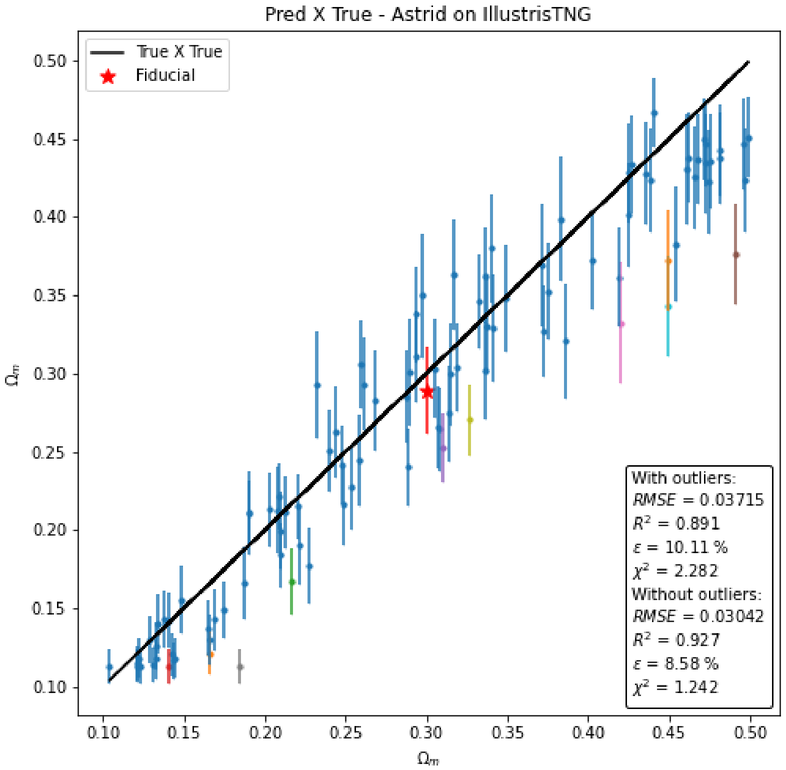
Villanueva-Domingo & FVN 2022
Shao, FVN et al. 2022
De Natali, FVN et al. 2022 (in prep)
See also Makinen et al 2022

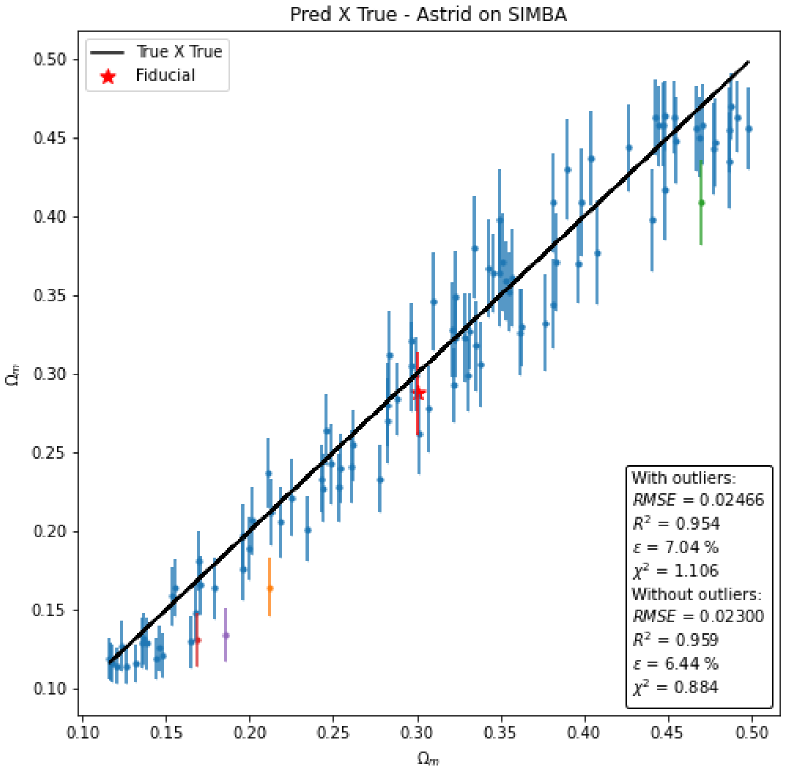
Pablo Villanueva-Domingo
Valencia

Helen Shao
Princeton

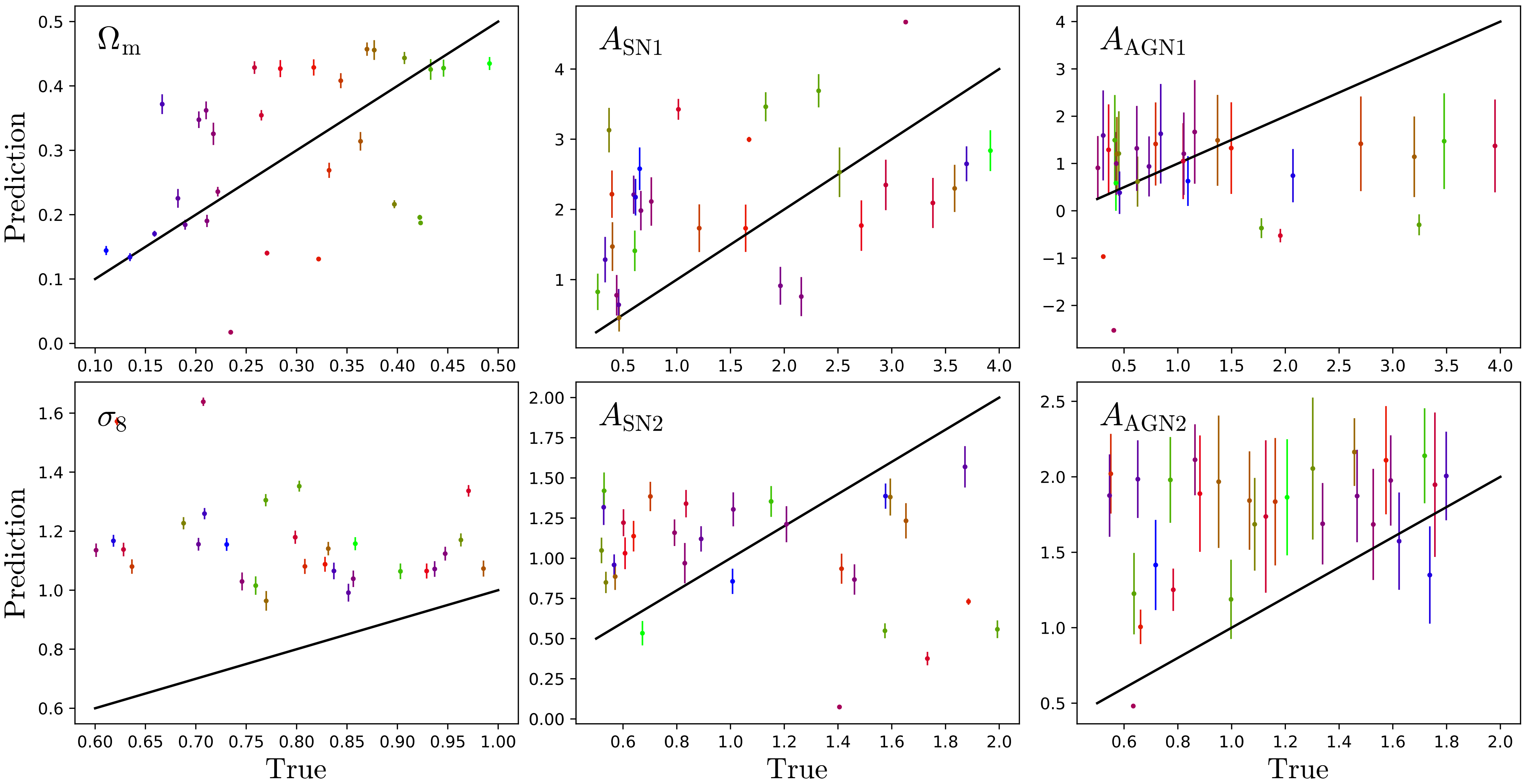
Natali de Santi
Sao Paolo / Flatiron





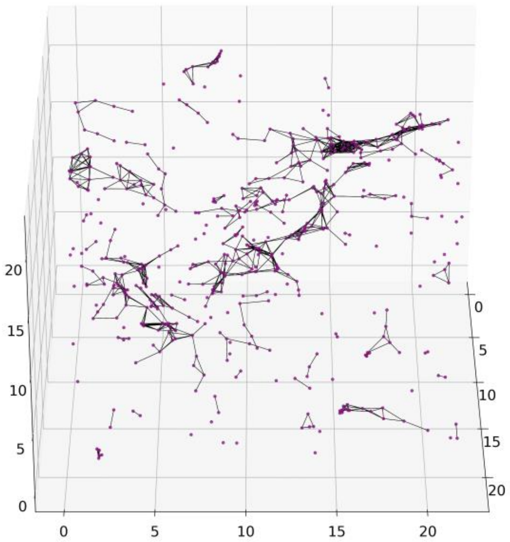


Network trained on IllustrisTNG and tested on SIMBA

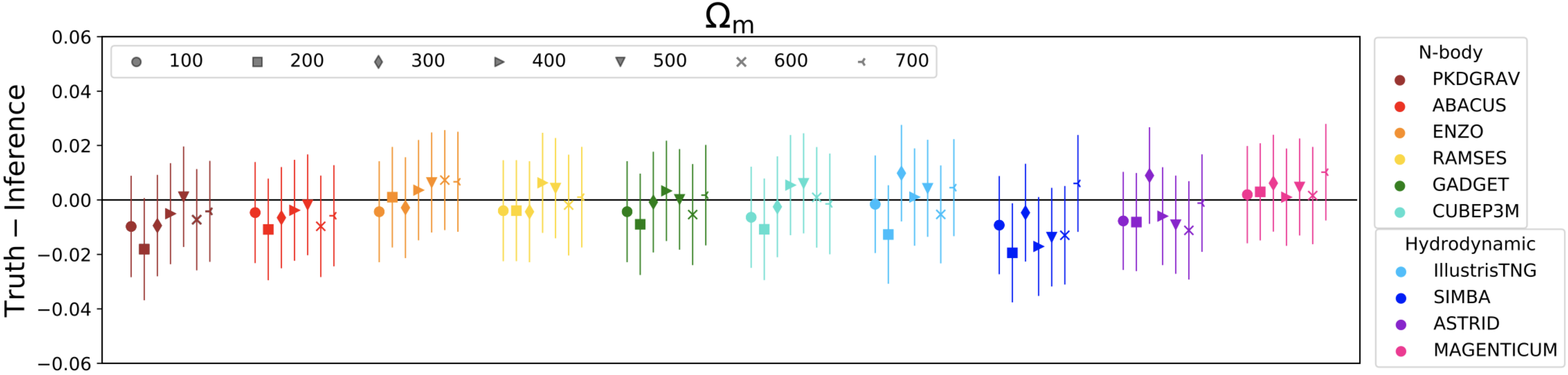


Robust field-level inference: halos

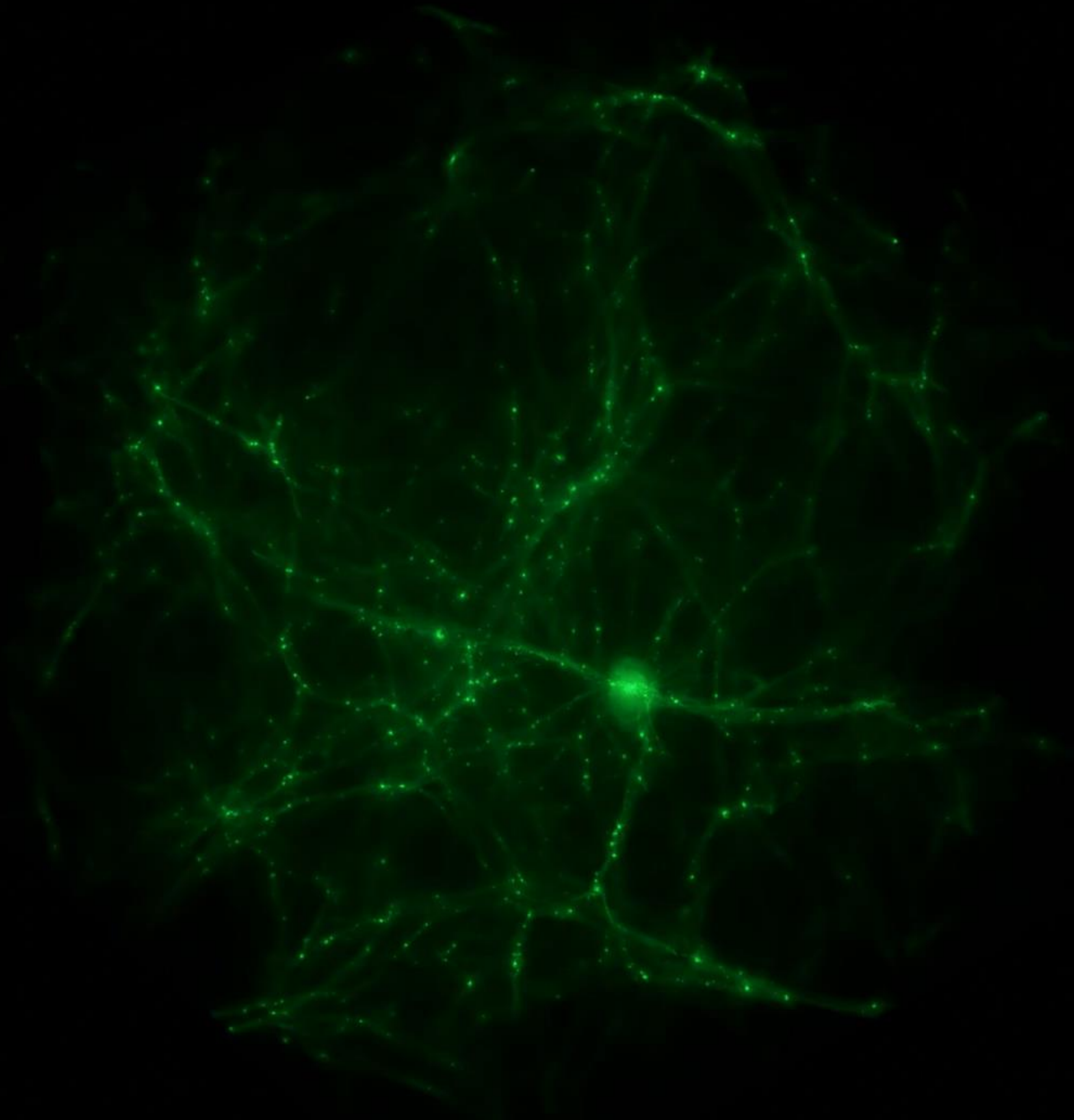
Helen Shao
Princeton



Careful about what variables to use
Mass, Vmax, concentration are not robust

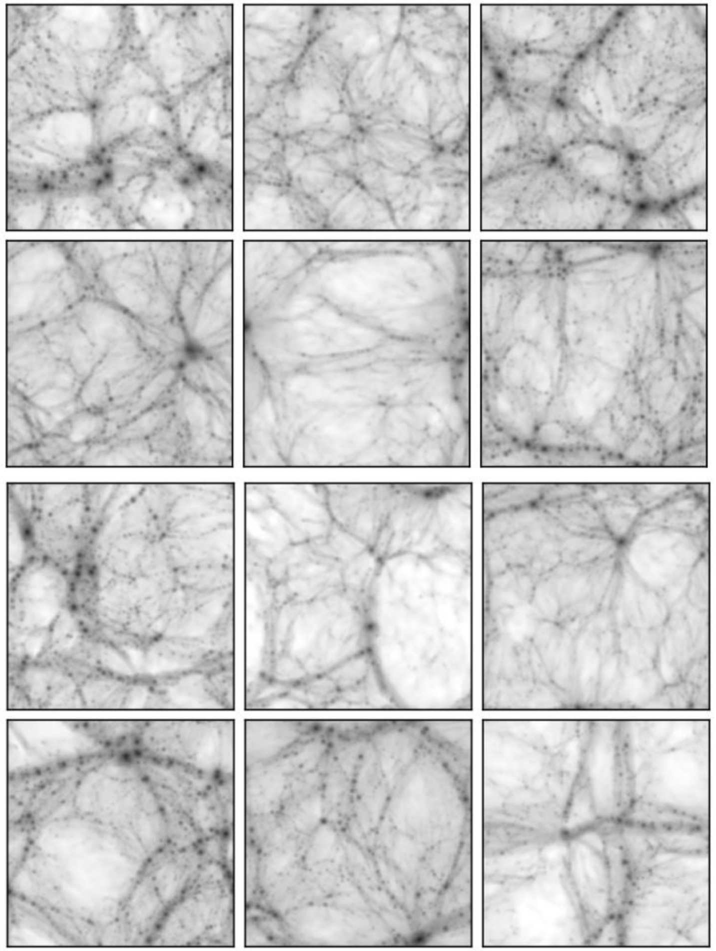


Ω_m	0.1
σ_8	0.8
A_{SN1}	1.0
A_{AGN1}	1.0
A_{SN2}	1.0
A_{AGN2}	1.0

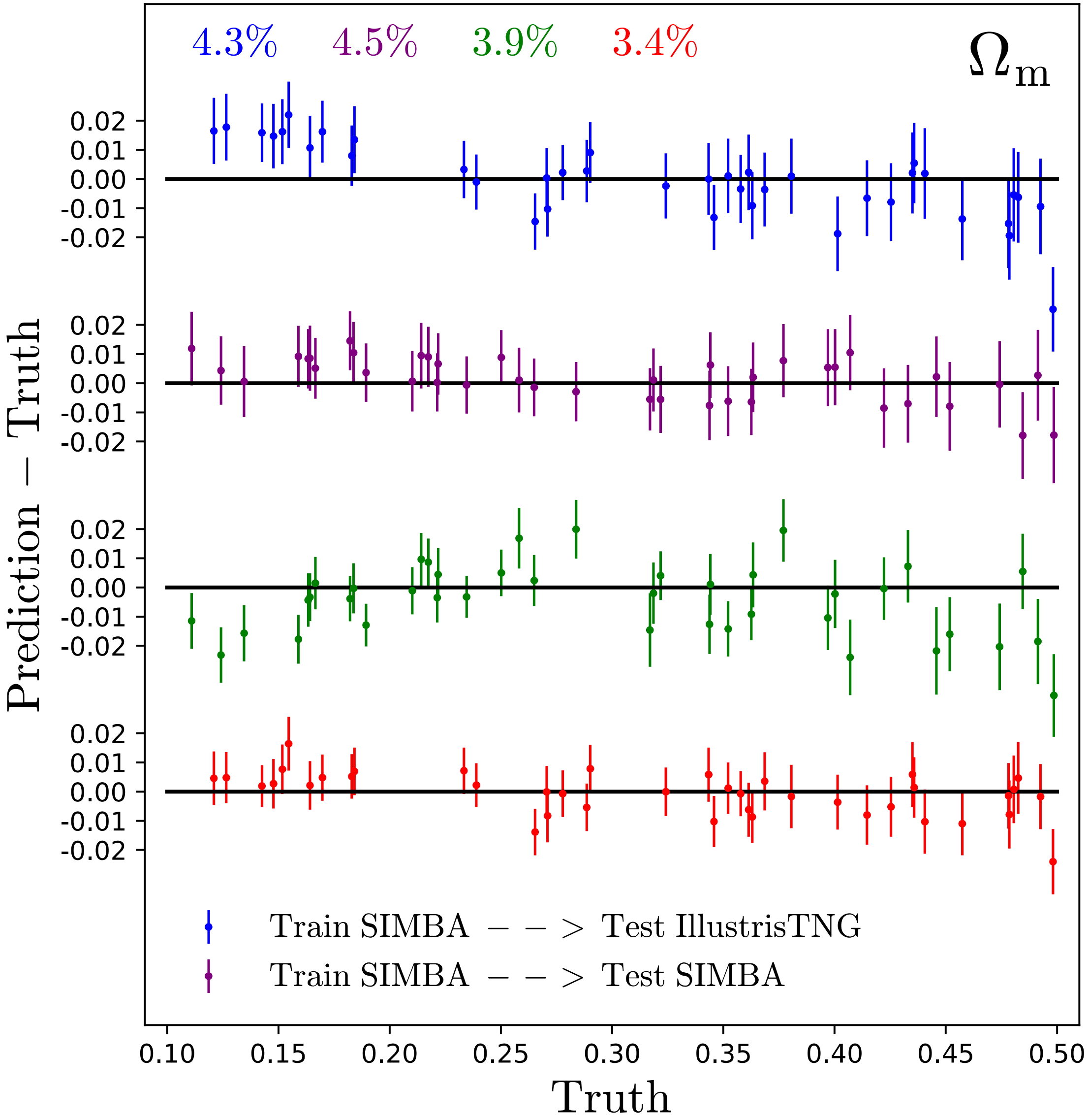


Cosmology at the field level

FVN et al. 2021a (2109.09747)
FVN et al. 2021b (2109.10360)

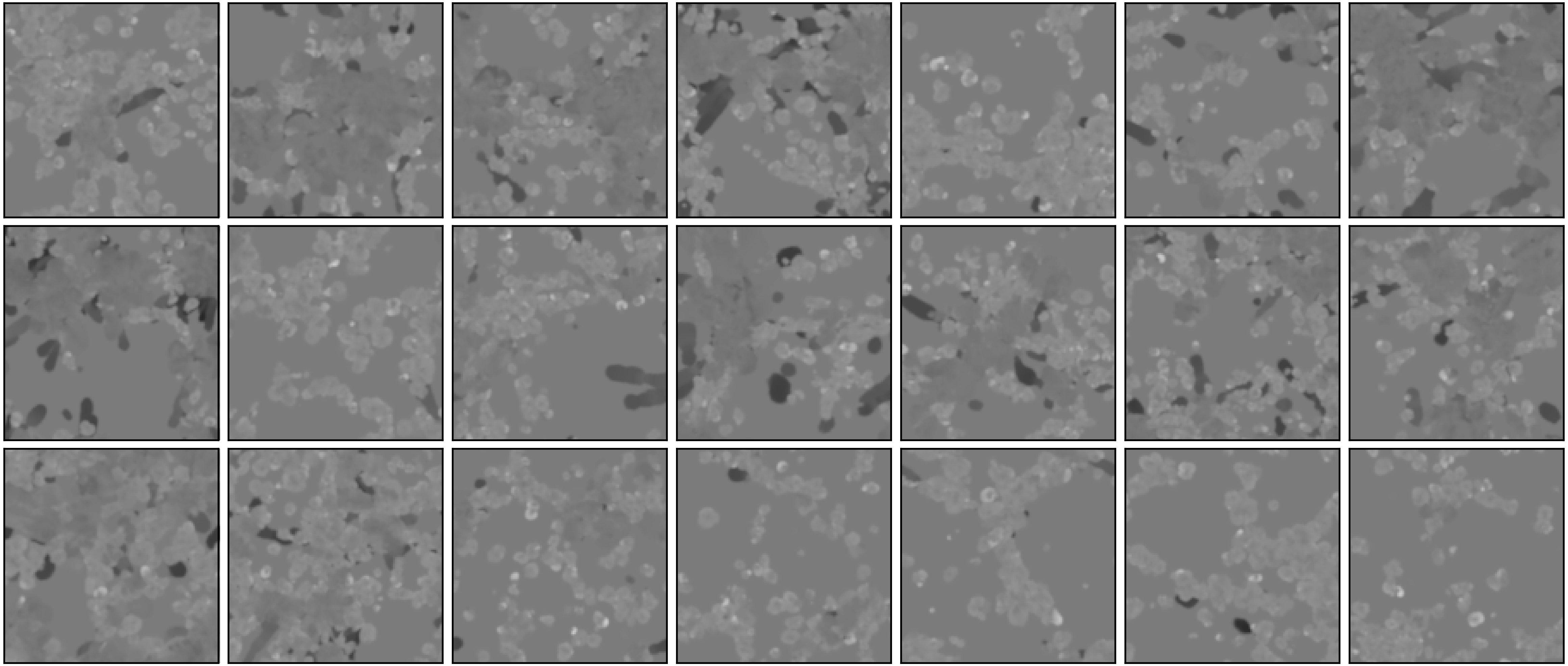


Multifield Cosmology!!!

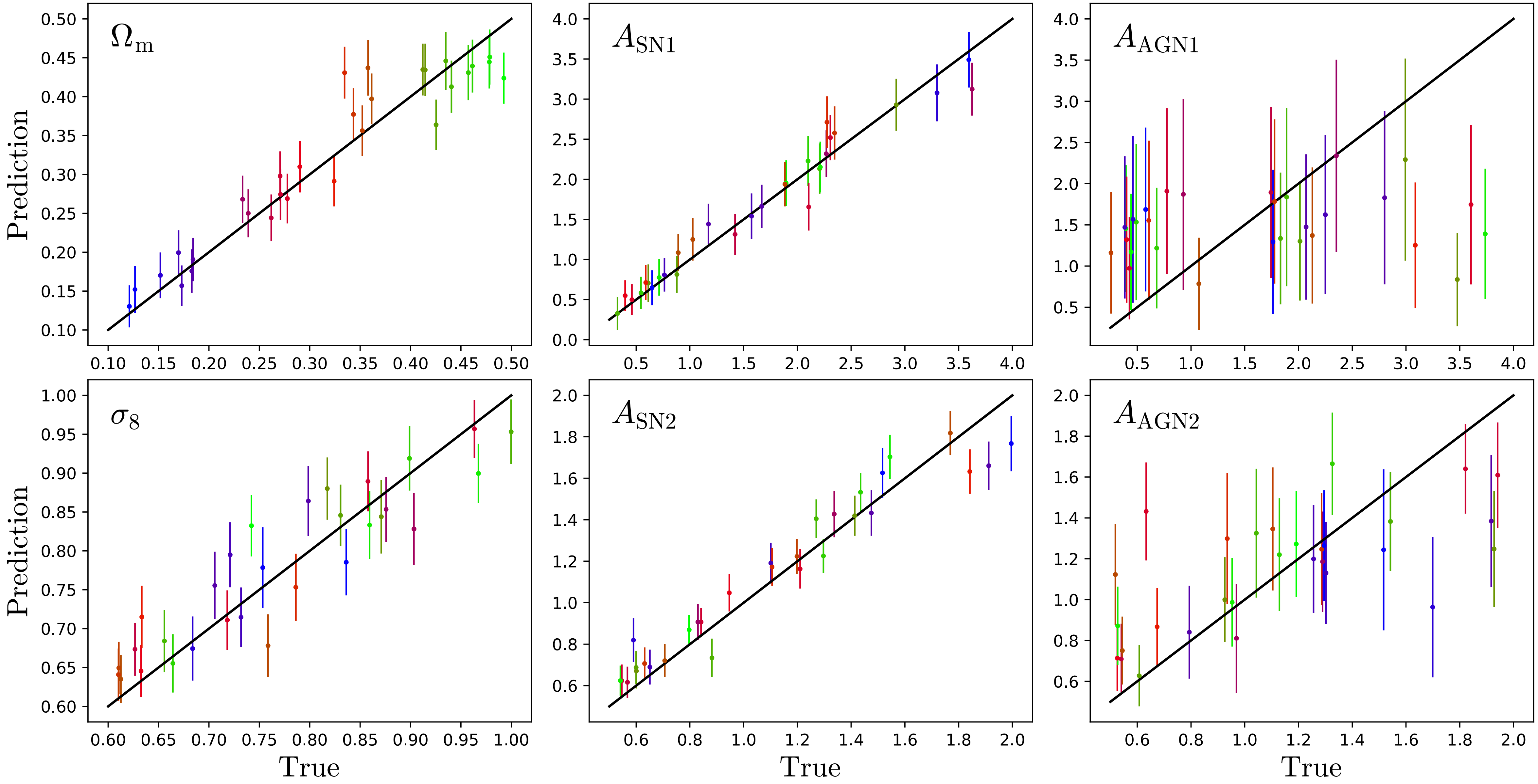


Example II: Magnesium over Iron (Mg/Fe)

FVN et al. 2021a



Likelihood-free inference: Mg/Fe



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