Lucia Perez (postdoc @ Princeton & Flatiron Institute's CCA)

CAMELS SAM

untangling the galaxy-halo connection with machine learning and galaxy clustering

How can we generate large galaxy catalogs that vary in cosmology *and* astrophysics, to better study the galaxy-halo connection?

Start with dark matter & cosmology, then choose how to make galaxies…

Most realistic	**Hydrodynamic simulations**	**Most expensive**
⬇️	**Semi-analytic models**	⬇️
Most analytical	**Empirical models** of galaxy formation	**Least expensive**

Volume vs. Resolution vs. Sophistication

How can we generate large galaxy catalogs that vary in cosmology *and* astrophysics, to better study the galaxy-halo connection?

Start with dark matter & cosmology, then choose...

Most realistic	**Hydrodynamic simulations**	• Explicitly solve equations of gravity, (magneto)hydrodynamics, and thermodynamics for dark matter, gas, stars, black holes particles • Star & black hole formation & feedback must use subgrid recipes, still fundamentally not well understood	**Most** expensive
⬇️	**Semi-analytic models**	• Built on merger trees of dark matter halos (theoretical for fastest, or from N-body simulations for spatial information) • Gas accretion ~ dark matter accretion • Recipes model baryonic processes & feedback: like a "flow model"	⬇️
Most analytical	**Empirical models** of galaxy formation	• Parametrize relationship between halos & galaxies, fine tune with observations, *very* fast to run at large scales • **Halo Occupation Distribution** & subhalo abundance matching: how are galaxies distributed within a given dark matter halo? • e.g. GalaxyNet (Moster et al. 2020), uses machine learning	**Least** expensive

Volume vs. Resolution vs. Sophistication

How can we generate large galaxy catalogs that vary in cosmology *and* astrophysics, to better study the galaxy-halo connection?

Start with dark matter & cosmology, then choose...

Most realistic	**Hydrodynamic simulations**	• Explicitly solve equations of ity, (hy ics, and thermodynam da cles • Star & bl still sfunda	**Most** expensive
⬇️	**Semi-analytic models**	• Bu • • R	⬇️
Most analytical	**Empirical models** of galaxy formation	• P une wi • **Halo Occ** matching: how are ga en dark matter halo? • e.g. GalaxyNet (Moster et es machine learning	**Least** expensive

What if you want to study the dependence on cosmology? Or on your prescription for galaxy physics?

Volume vs. Resolution vs. Sophistication

The core CAMELS project

Volumes of $(25\ h^{-1}\ \mathrm{cMpc})^3$ with 256^3 dark matter + 256^3 gas particles

IllustrisTNG suite

SIMBA suite

Astrid suite

N-body only partners for each!

Each hydro- suite has…

- 1,000 simulations EACH across a *latin hypercube (LH)* of Ω_m, σ_8, 2 supernova, & 2 AGN feedback parameters
- Dozens of *one-parameter (1P)* simulations, varying astrophysical parameters one at a time
- Dozens of *Cosmic variance (CV)* simulations varying only the random seed
- BONUS: 1024 “SB” simulations of TNG over all 20+ astro parameters!
- On the way: Enzo, Magneticum, Ramses, & SWIFT-Eagle suites!

THIS, and more, is PUBLIC!!!

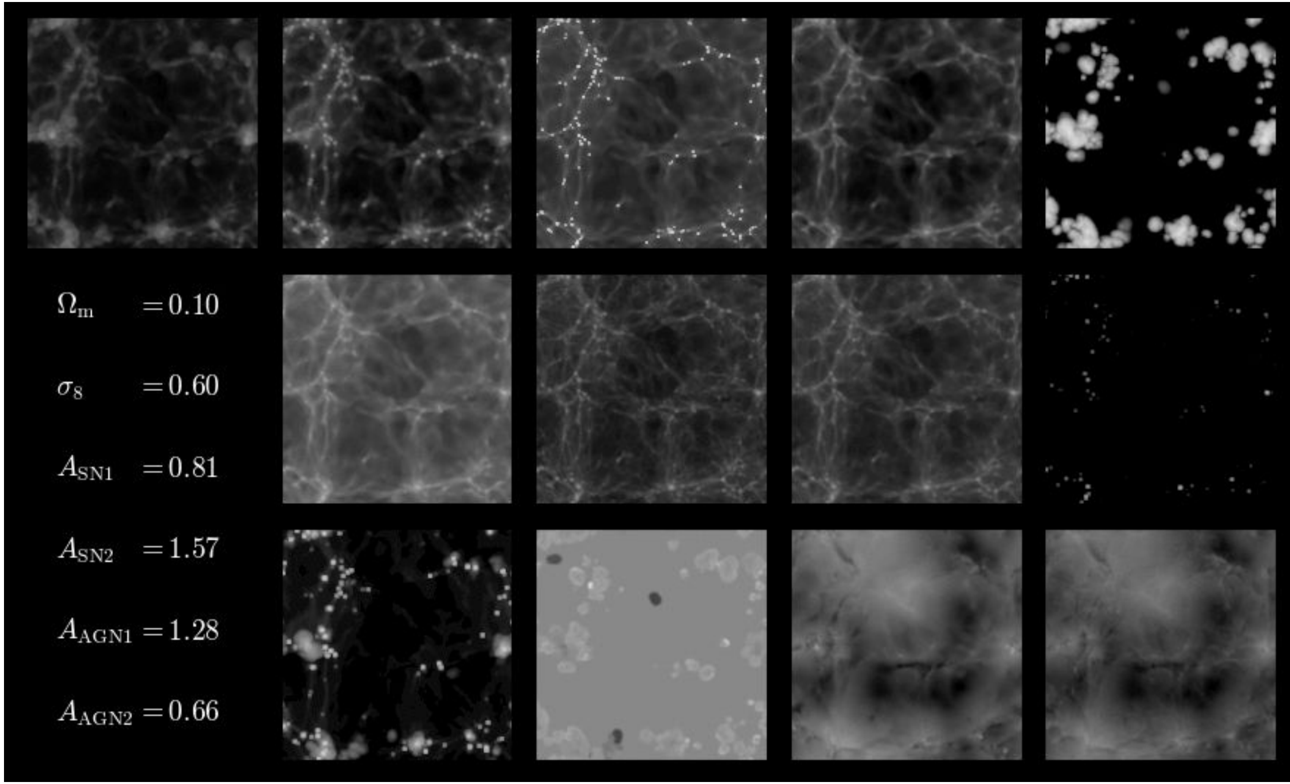
camels.readthedocs.io

CAMELS Multifield Dataset

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

Villaescusa-Navarro et al. 2021c (2109.10915)

- 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:
  - Gas density
  - Gas temperature
  - Gas metallicity
  - Gas pressure
  - Neutral hydrogen density
  - Electron number density
  - Dark matter density
  - Total matter density
  - Stellar mass density
  - Gas velocity
  - Dark matter velocity
  - Magnetic fields
  - Mg/Fe
- All data publicly available (70 Tb)
- The MNIST of cosmology



CAMELS projects on ADS, part 1

2023arXiv230406084E 2023/04
Cosmology with one galaxy? -- The ASTRID model and robustness
Echeverri, Nicolas; Villaescusa-Navarro, Francisco; Chawak, Chaitanya *and 7 more*

2023arXiv230402096N 2023/04 cited: 1
The CAMELS project: Expanding the galaxy formation model space with new ASTRID and 28-parameter TNG and SIMBA suites
Ni, Yueying; Genel, Shy; Anglés-Alcázar, Daniel *and 12 more*

2023ApJS..265...54V 2023/04 cited: 29
The CAMELS Project: Public Data Release
Villaescusa-Navarro, Francisco; Genel, Shy; Anglés-Alcázar, Daniel *and 45 more*

2023ascl.soft03020V 2023/03
HaloGraphNet: Predict halo masses from simulations
Villanueva-Domingo, Pablo; Villaescusa-Navarro, Francisco; Anglés-Alcázar, Daniel *and 7 more*

2023arXiv230307473A 2023/03
Invertible mapping between fields in CAMELS
Andrianomena, Sambatra; Hassan, Sultan; Villaescusa-Navarro, Francisco

2023arXiv230214591S 2023/02 cited: 2
A universal equation to predict Ω_m from halo and galaxy catalogues
Shao, Helen; de Santi, Natalí S. M; Villaescusa-Navarro, Francisco *and 13 more*

2023arXiv230214101D 2023/02 cited: 1
Robust field-level likelihood-free inference with galaxies
de Santi, Natalí S. M.; Shao, Helen; Villaescusa-Navarro, Francisco *and 12 more*

2023arXiv230201363O 2023/02 cited: 1
Topological data analysis reveals differences between simulated galaxies and dark matter haloes
Ouellette, Aaron; Holder, Gilbert; Kerman, Ely

2023MNRAS.519.2251B 2023/02 cited: 1
X-ray absorption lines in the warm-hot intergalactic medium: probing Chandra observations with the CAMEL simulations
Butler Contreras, Amanda; Lau, Erwin T.; Oppenheimer, Benjamin D. *and 5 more*

2023ApJ...944...67J 2023/02
Calibrating Cosmological Simulations with Implicit Likelihood Inference Using Galaxy Growth Observables
Jo, Yongseok; Genel, Shy; Wandelt, Benjamin *and 7 more*

2023ApJ...944...27S 2023/02 cited: 1
Robust Field-level Inference of Cosmological Parameters with Dark Matter Halos
Shao, Helen; Villaescusa-Navarro, Francisco; Villanueva-Domingo, Pablo *and 13 more*

2023arXiv230102231D 2023/01 cited: 1
Predicting the impact of feedback on matter clustering with machine learning in CAMELS
Delgado, Ana Maria; Angles-Alcazar, Daniel; Thiele, Leander *and 6 more*

2023arXiv230102186P 2023/01 cited: 2
Inferring the impact of feedback on the matter distribution using the Sunyaev Zel'dovich effect: Insights from CAMELS simulations and ACT+DES data
Pandey, Shivam; Lehman, Kai; Baxter, Eric J. *and 6 more*

2023AAS...24140604M 2023/01
Probing the Physics of Circum-Galactic Medium using Fast Radio Bursts: Insights from CAMELS simulations
Medlock, Isabel; Nagai, Daisuke

2022arXiv221205964S 2022/12 cited: 1
Baryonic Imprints on DM Halos: The concentration-mass relation in the CAMELS simulations
Shao, Mufan; Anbajagane, Dhayaa; Chang, Chihway

2022ApJ...941..132M 2022/12 cited: 2
Inpainting Hydrodynamical Maps with Deep Learning
Mohammad, Faizan G.; Villaescusa-Navarro, Francisco; Genel, Shy *and 2 more*

2022arXiv221112724F 2022/11
HIGlow: Conditional Normalizing Flows for High-Fidelity HI Map Modeling
Friedman, Roy; Hassan, Sultan

CAMELS projects on ADS, part 2

2022arXiv221105000A 2022/11
Emulating cosmological multifields with generative adversarial networks
Andrianomena, Sambatra; Villaescusa-Navarro, Francisco; Hassan, Sultan

2022ApJ...937..115V 2022/10 cited: 13
Learning Cosmology and Clustering with Cosmic Graphs
Villanueva-Domingo, Pablo; Villaescusa-Navarro, Francisco

2022ApJ...937...83H 2022/10 cited: 11
HIFLOW: Generating Diverse HI Maps and Inferring Cosmology while Marginalizing over Astrophysic Using Normalizing Flows
Hassan, Sultan; Villaescusa-Navarro, Francisco; Wandelt, Benjamin *and 11 more*

2022arXiv220902075W 2022/09 cited: 7
The SZ flux-mass (Y-M) relation at low halo masses: improvements with symbolic regression and strong constraints on baryonic feedback
Wadekar, Digvijay; Thiele, Leander; Hill, J. Colin *and 8 more*

2022arXiv220900657P 2022/09 cited: 1
Studying the Warm Hot Intergalactic Medium in emission: a reprise
Parimbelli, G.; Branchini, E.; Viel, M. *and 2 more*

2022arXiv220808927A 2022/08
Predictive uncertainty on improved astrophysics recovery from multifield cosmology
Andrianomena, Sambatra; Hassan, Sultan

2022ApJ...933..133M 2022/07 cited: 11
The Circumgalactic Medium from the CAMELS Simulations: Forecasting Constraints on Feedback Processes from Future Sunyaev-Zeldovich Observations
Moser, Emily; Battaglia, Nicholas; Nagai, Daisuke *and 9 more*

2022maff.confE..30L 2022/06
Constraining Cluster Astrophysics and Cosmology with X-ray Power Spectrum
Lau, Erwin

2022arXiv220402408P 2022/04 cited: 7
Constraining cosmology with machine learning and galaxy clustering: the CAMELS-SAM suite
Perez, Lucia A.; Genel, Shy; Villaescusa-Navarro, Francisco *and 5 more*

2022PhRvD.105h3505T 2022/04 cited: 7
Percent-level constraints on baryonic feedback with spectral distortion measurements
Thiele, Leander; Wadekar, Digvijay; Hill, J. Colin *and 7 more*

2022JCAP...04..046N 2022/04 cited: 6
Breaking baryon-cosmology degeneracy with the electron density power spectrum
Nicola, Andrina; Villaescusa-Navarro, Francisco; Spergel, David N. *and 8 more*

2022ApJS..259...61V 2022/04 cited: 31
The CAMELS Multifield Data Set: Learning the Universe's Fundamental Parameters with Artificial Intelligence
Villaescusa-Navarro, Francisco; Genel, Shy; Anglés-Alcázar, Daniel *and 25 more*

2022ApJ...929..132V 2022/04 cited: 13
Cosmology with One Galaxy?
Villaescusa-Navarro, Francisco; Ding, Jupiter; Genel, Shy *and 10 more*

2021arXiv211114874V 2021/11 cited: 13
Weighing the Milky Way and Andromeda with Artificial Intelligence
Villanueva-Domingo, Pablo; Villaescusa-Navarro, Francisco; Genel, Shy *and 6 more*

2021arXiv210910360V 2021/09 cited: 29
Robust marginalization of baryonic effects for cosmological inference at the field level
Villaescusa-Navarro, Francisco; Genel, Shy; Angles-Alcazar, Daniel *and 11 more*

2021arXiv210909747V 2021/09 cited: 32
Multifield Cosmology with Artificial Intelligence
Villaescusa-Navarro, Francisco; Anglés-Alcázar, Daniel; Genel, Shy *and 10 more*

2021ApJ...915...71V 2021/07 cited: 98
The CAMELS Project: Cosmology and Astrophysics with Machine-learning Simulations

In prep: Megan Tillman et al. : AGN feedback effects on the low-z Lyman-α forest, and the interplay between AGN and stellar feedback (mtt74@rutgers.edu)

How can we generate *large* galaxy catalogs that vary in cosmology *and* astrophysics, to better study the galaxy-halo connection?

CAMELS-SAM – moderate volumes, resolution, and sophistication… but flexible and physics-based!

Most realistic	Hydrodynamic simulations	Most expensive
⬇️	Semi-analytic models	⬇️
Most analytical	Empirical models of galaxy formation	Least expensive

Core CAMELS – small volumes, moderate resolution, highly sophisticated & varied

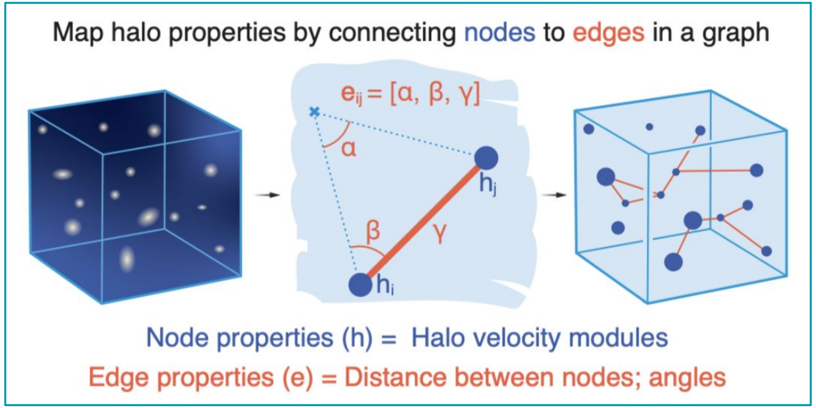
e.g. **SIMBIG** (Hahn et al. 2023) : giant low-res Quijote N-body volumes + decorated HOD

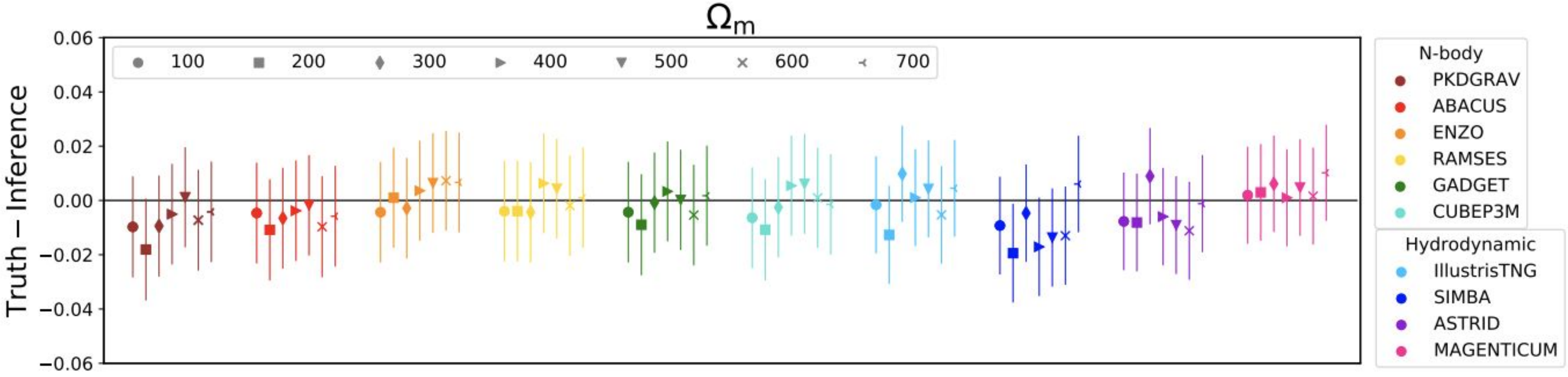
Volume vs. Resolution vs. Sophistication

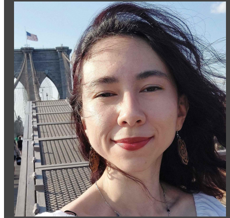
Robust field level inference of cosmological parameters with dark matter halos

Helen Shao
Princeton University
hshao@princeton.edu

- Halo field → **graph** → GNN → Ω_m & σ_8
- Train: Gadget N-Body halos
- Test: **5 N-Body** codes + hydrodynamic codes of **4 different** subgrid physics models
- Robust when using halo **velocities & positions**







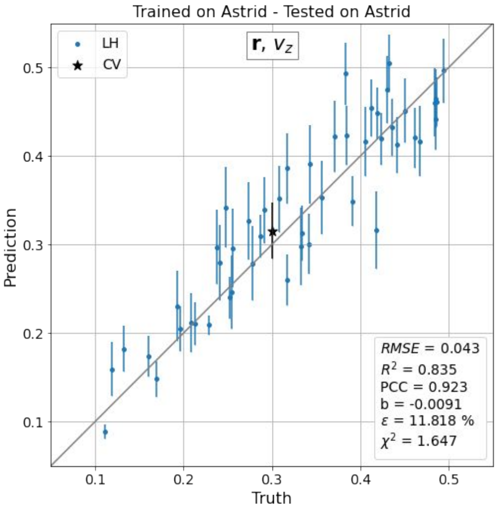
Natalí de Santi

Flatiron Institute
University of São Paulo

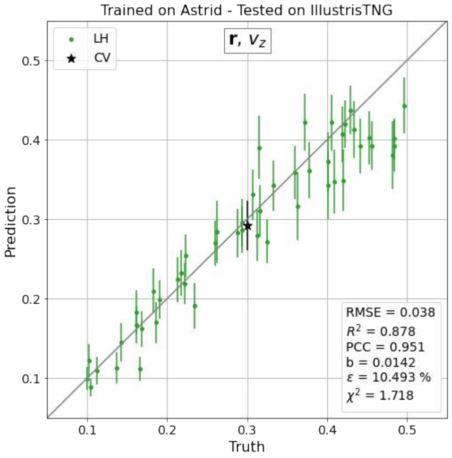
natalidesanti@gmail.com

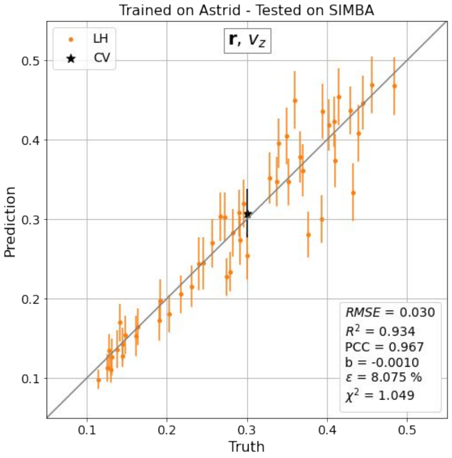
Robust field-level likelihood-free inference with galaxies

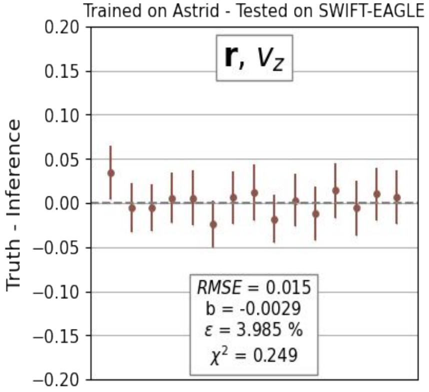
Dataset: Galaxies from Astrid
Machine Learning Method:
Graph Neural Networks
Objective: Ω_m inference

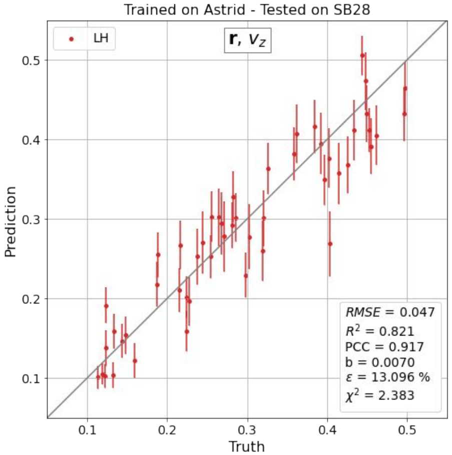


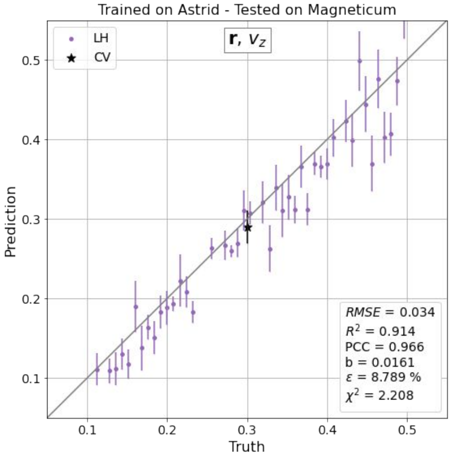
arXiv: 2302.14101











- Information came from galaxy positions and velocities;
- The broader variation in Astrid allowed a robust model across 5 different sub-grid physics sets;
- First steps to apply this machinery on real data.

A universal equation to predict Ω_m from halo and galaxy catalogues

Gadget Halo Catalogues

↓

Halo Graphs

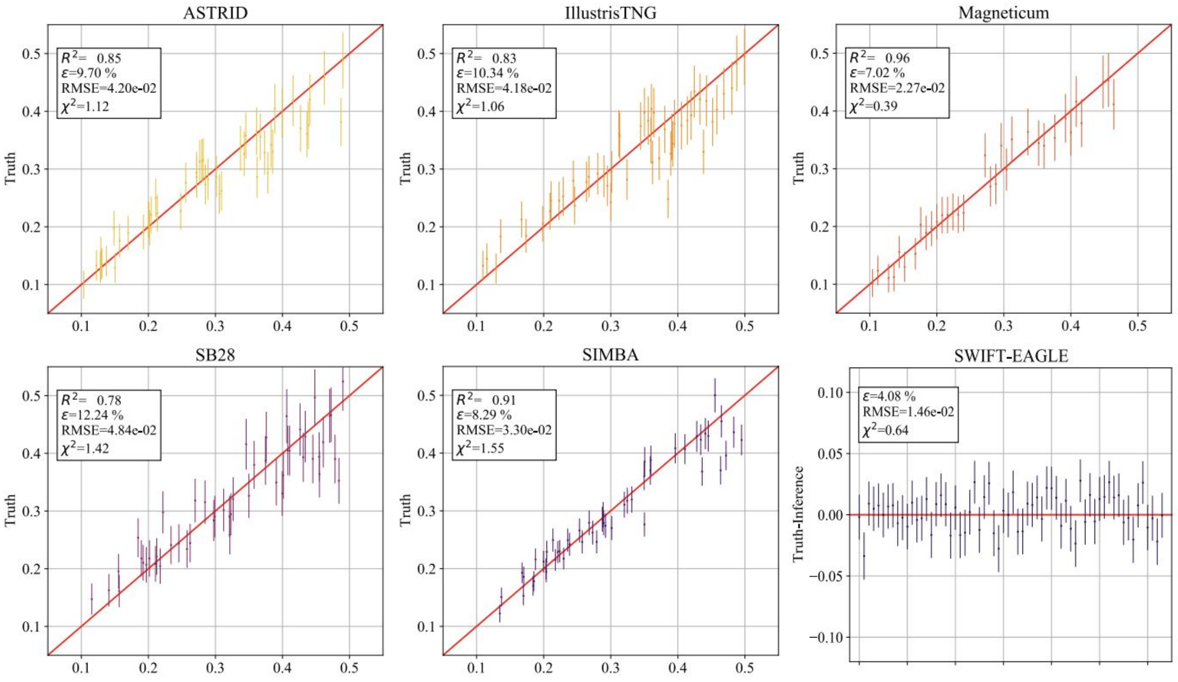
↓

Graph Neural Network

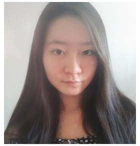
↓

Symbolic Regression

- Using **velocity modulus** and 3D-**positions**
- Train on **Gadget** N-body halos → test on halos & galaxies from:
  - **6** hydrodynamical suites
  - **6** N-body suites



GNN Component	Formula
Edge Model: $e_1^{(1)}$	$1.32\lvert v_i - v_j + 0.21\rvert + 0.12(v_i - v_j) - 0.12(\gamma_{ij} + \beta_{ij} - 1.73)$
Edge Model: $e_2^{(1)}$	$\lvert 1.62(v_i - v_j) + 0.45\rvert + 1.98(v_i - v_j) + 0.55$
Node Model: $v_1^{(1)}$	$1.21^{v_i}(0.77^{3.29\sum_{j\in\mathcal{N}_j} e_1^{(1)} + \sum_{j\in\mathcal{N}_j} e_2^{(1)}}) + 0.12$
Node Model: $v_1^{(1)} + v_2^{(1)}$	$0.78 - \sqrt{\log(0.16^{\sum_{j\in\mathcal{N}_j} e_2 + \sum_{j\in\mathcal{N}_j} e_1 - 0.41 v_i - 1.05})} + 1.45$
Final MLP: μ_{Ω_m}	$4\times10^{-4}\cdot(-5.5\sum_{i\in\mathcal{G}} v_2^{(1)} + 2.21\sum_{i\in\mathcal{G}} v_1^{(1)} + \lvert 0.96\sum_{i\in\mathcal{G}} v_2^{(1)} + 0.82\sum_{i\in\mathcal{G}} v_1^{(1)}\rvert) - 0.103$



Helen Shao
Princeton University
hshao@princeton.edu

https://arxiv.org/abs/2302.14591

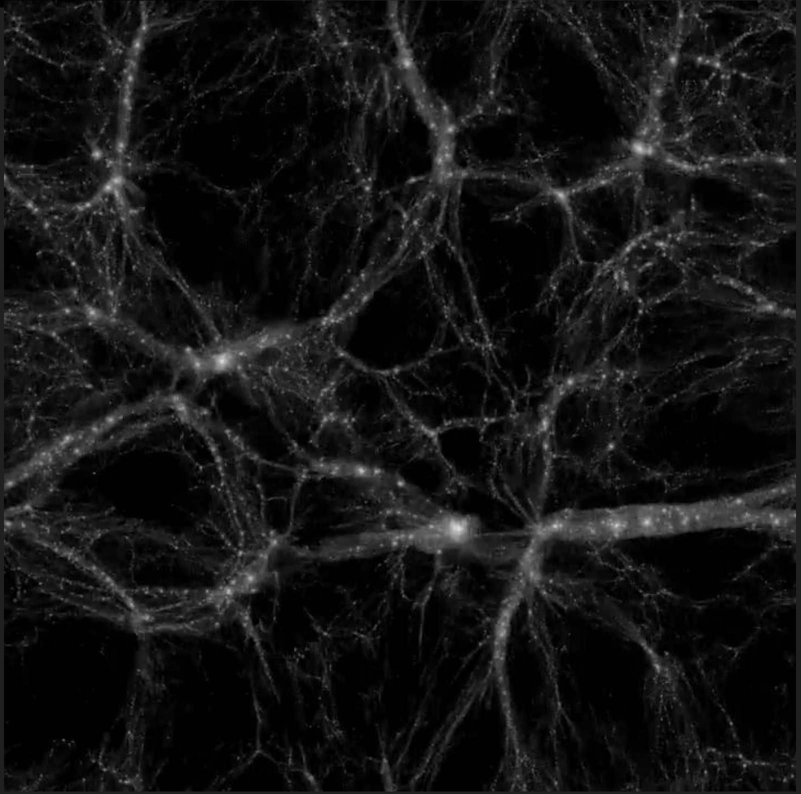
CAMELS SAM

Lucia A. Perez (Princeton & CCA)
Shy Genel, Paco Villaescusa-Navarro, Rachel Somerville, Daniel Angles-Alcazar, Austen Gabrielpillai

New large-volume simulation 'hump' of CAMELS project

- 1000+ N-body simulations: $(100\ h^{-1}\ \mathrm{Mpc})^3$ large ; $N=640^3$ particles of ~$1\text{-}6 \times 10^8\ h^{-1}\ M_{sol}$; 100 snapshots between $0<z<27$
- Cosmological parameter space: $\boldsymbol{\Omega_m}$ (fraction of energy density in DM+baryons) & $\boldsymbol{\sigma_8}$ (~amplitude of density fluctuations)
- Run each through the Santa Cruz Semi-Analytic Model:

  "A_{SN}": mass outflow + reheating rates of cold gas due to SNe + stars
  "A_{AGN}": AGN feedback, how much mass ejected in radio jets?



LH_643: $\Omega_m = 0.131$; $\sigma_8 = 0.986$

Perez+2023 (in press) | arXiv:2204.02408

camels-sam.readthedocs.io

The Santa Cruz Semi-Analytic Model for galaxy formation

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

⬆️ mass outflow rate due to SN & massive stars

⬇️ mass ejected by an AGN in radio jets

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

Typical to most SAMs are physically-motivated prescriptions for:

- How gas cools & accretes onto halos/galaxies
- How stars form from cooled gas in ISM
- How mass/metals return to the ISM

...using the information in halo *merger trees*!

Unique/notable in the SC-SAM:

- Multiphase partitioning & tracking of the ISM
- How supermassive black holes form and grow, 'black hole feedback'

Somerville et al. (2008, 2015, 2021) + Porter et al. (2014) + Gabrielpillai et al. (2022)

Cool example: mocks for JWST and Roman by Yung et al. 2019-2022!

CAMELS SAM

Lucia A. Perez
(Princeton & CCA)
Shy Genel, Paco Villaescusa-Navarro, Rachel Somerville, Daniel Angles-Alcazar, Austen Gabrielpillai

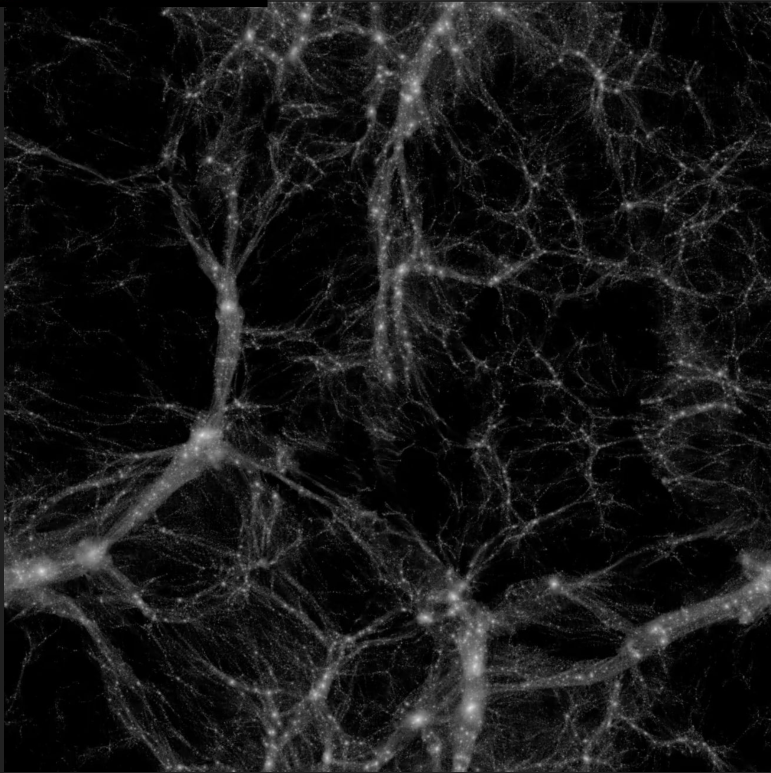
CAMELS-SAM public data | *camels-sam.readthedocs.io*

From 1000+ simulations with 100 snapshots between 20 < z < 0:

- *ROCKSTAR* halo catalogs
- *ConsistentTrees* merger trees
- Santa Cruz SAM galaxy catalogs
- *Full snapshots are on tape–reach out if you really want them!*

Data product flavors:

- **1000 LH** simulations over Ω_m, σ_8, A_{SN1}, A_{SN2}, A_{AGN}
- **5 CV** simulations: Ω_m=0.3, σ_8=0.8, default SC-SAM, unique random seeds
- **12** galaxy catalogs with fiducial cosmology, min/max SC-SAM parameters, for 2 unique random seeds



#643: Ω_m = 0.131 ; σ_8 = 0.986

Lucia Perez (postdoc @ Princeton & Flatiron Institute's CCA)

CAMELS SAM

Older draft – arXiv:2204.02408 ; final form in press

- Using clustering & neural networks, how much information about cosmology is lost when varying astrophysics?
- Can our neural networks marginalize over astrophysics to constrain cosmology with galaxy clustering?

Clustering of galaxies

Neural network

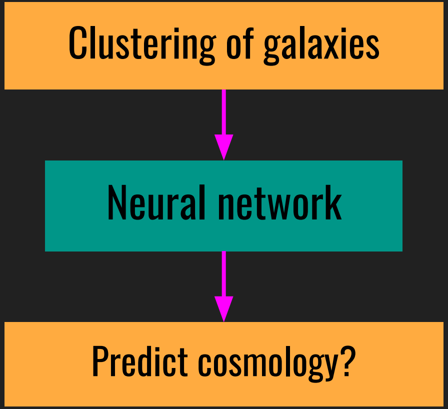
Predict cosmology?

Lucia Perez (postdoc @ Princeton & Flatiron Institute's CCA)

CAMELS SAM

Older draft – arXiv:2204.02408 ; final form in press

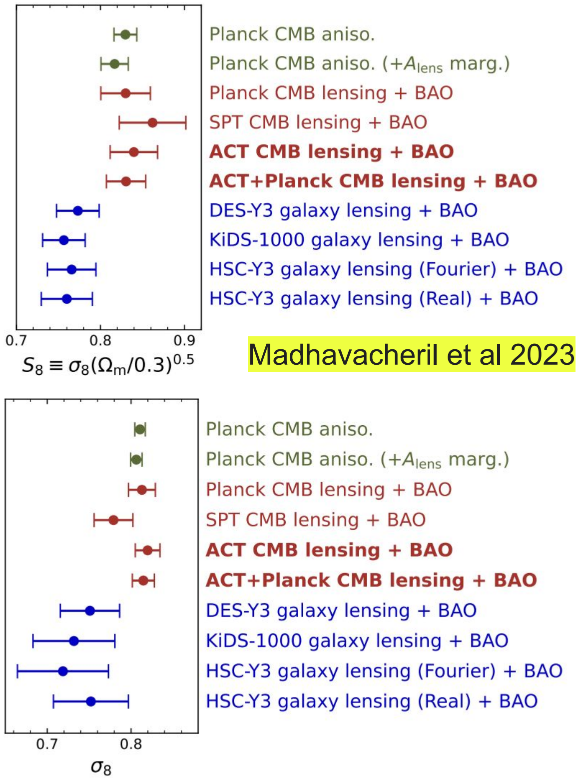
- Using clustering & neural networks, how much information about cosmology is lost when varying astrophysics?
  - very little, and you learn something about astrophysics!
- Can our neural networks marginalize over astrophysics to constrain cosmology with galaxy clustering?
  - yes, and while learning about astrophysics!



Why bother? Why only σ_8 & Ω_M?

Galaxy surveys constrain $S_8 = \sigma_8(\Omega_M/0.3)^{0.5}$

& find tension with CMB constraints for S_8!



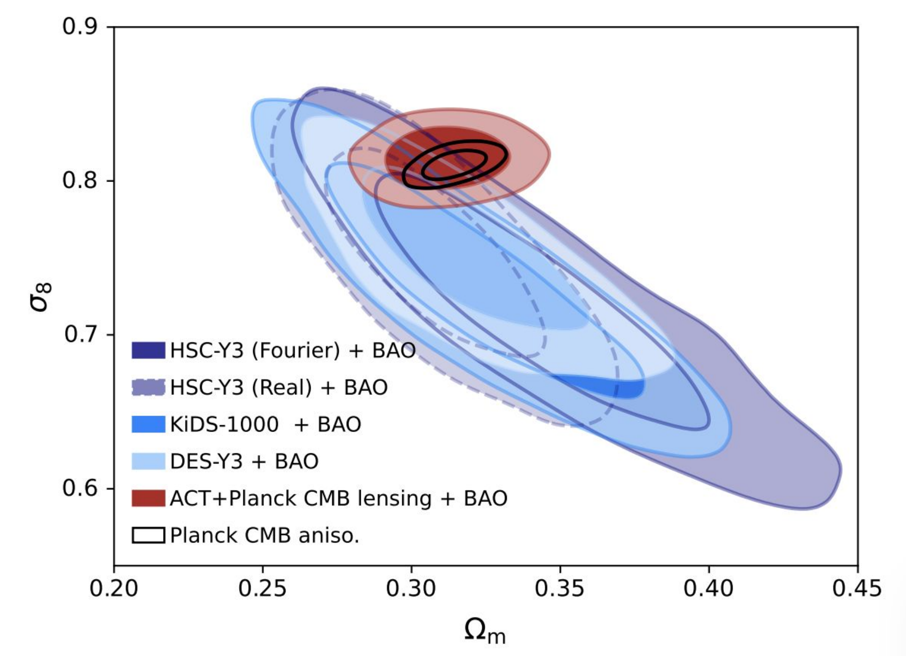
Madhavacheril et al 2023

Figure 7. Marginalized posteriors for various combinations of parameters measuring the amplitude of matter fluctuations. The top panel shows $S_8 \equiv \sigma_8(\Omega_m/0.3)^{0.5}$ which is best constrained by galaxy lensing, and the bottom panel shows σ_8. All lensing measurements shown here include BAO data. The *Planck* CMB anisotropy measurements are shown both without and with marginalization over late-time information; while the former is mostly an early-universe extrapolation, the latter is more fully so.

Don't forget:

the baryonic astrophysics of galaxies affects their large scale structure, too!

e.g. red vs. blue galaxies



CAMELS SAM

Learning the galaxy-halo connection? Classically…

Summary statistics to good start–full galaxy info HARD to learn, CAMELS is innovating a lot there!

Clustering of a sample of galaxies

? ?

Underlying cosmology?
Underlying astrophysics?

Clustering of a sample of galaxies

Theory!
e.g. bias models, halo occupation

Hard beyond linear large scales!

Underlying cosmology!
An approximate model for astrophysics?

Limited connection to astrophysics

CAMELS SAM

Learning the galaxy-halo connection?

Clustering of a sample of galaxies

?

?

Underlying cosmology?
Underlying astrophysics?

Clustering of a sample of galaxies

Theory!
e.g. bias models, halo occupation

Underlying cosmology!
An approximate model for astrophysics?

Clustering of a sample of galaxies

Neural network!

Predictions only as good as training!
Interpretability?

Big CAMELS-SAM suite

Theory of clustering & statistics

Underlying cosmology?
Input astrophysics?

CAMELS SAM

Thoughts as we went into this clustering proof-of-concept…

- Will SAM galaxies' clustering *actually* carry information about the astrophysics that the NN will pick up?
- How well will the neural networks marginalize over astrophysics?
- How well will the galaxy clustering do vs. dark matter clustering?
- *How will the different clustering statistics do?*

Galaxy clustering as Lucia does it:

Two-point correlation function

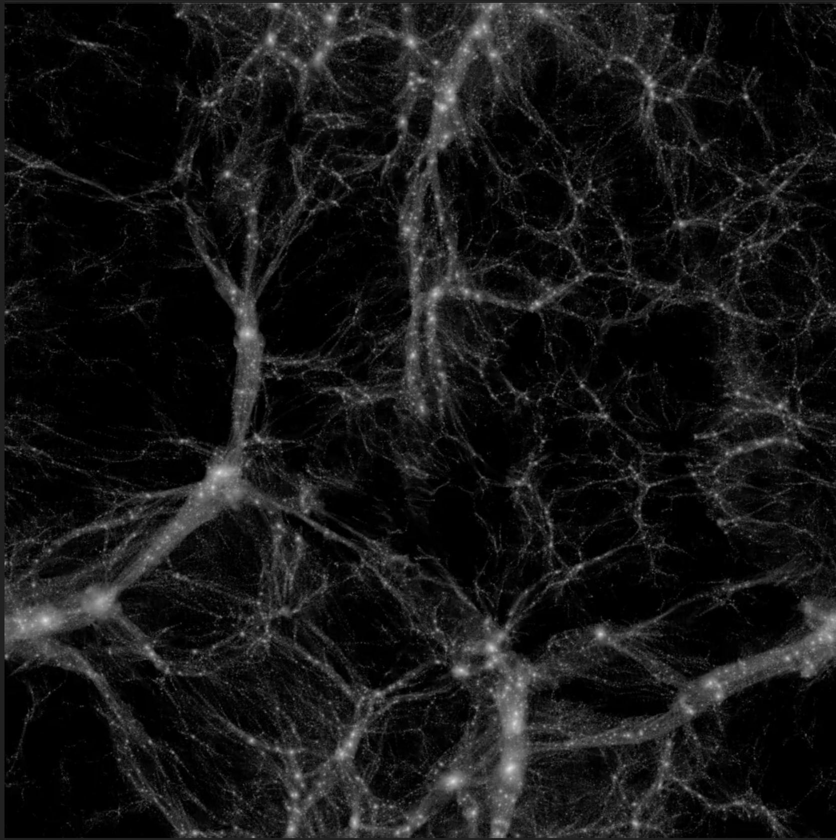
- Fourier Transform of power spectrum, common in observations
- Compare galaxies to a random distribution; pair counts
- Brief summary statistic: 1 R gives 1 ξ value

Count-in-cells

- Drop test spheres of a given size
- How many galaxies are in each test spheres?
- Volume averaged measurements
- Contains all higher order correlations!
- Computationally expensive + dense: 1 R can give 100's of points

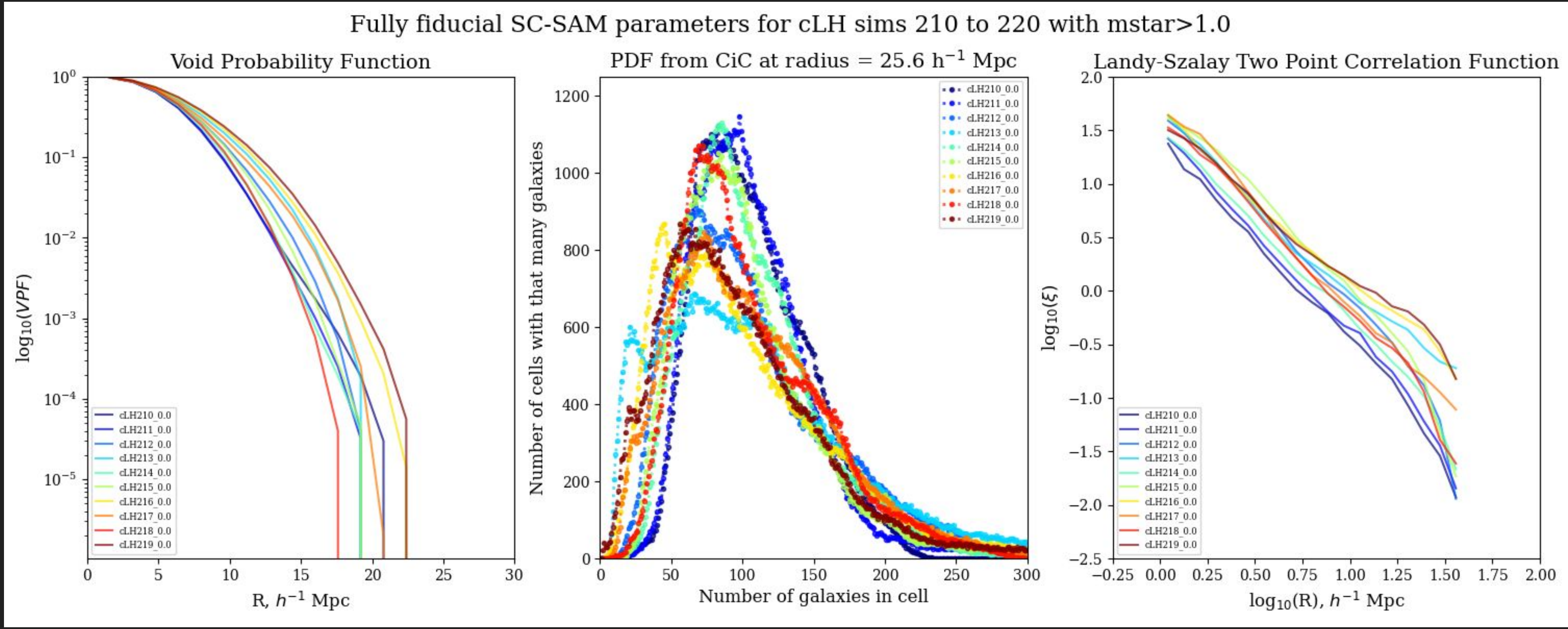
Void Probability Function

- Only *empty* test spheres–very cheap to calculate
- Influenced by higher order correlations
- Brief summary statistic: 1 R gives 1 VPF value



How much information about cosmology is lost when varying astrophysics?

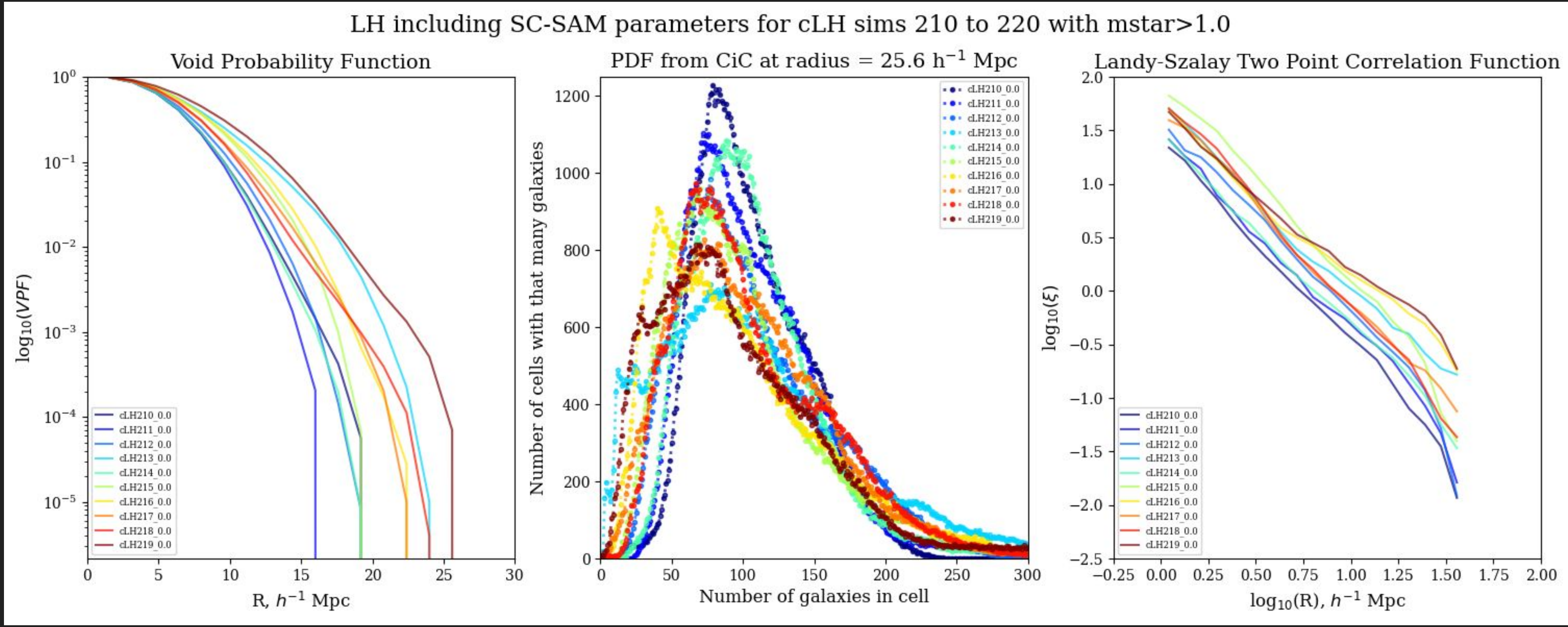
Some, but sufficiently varying over astrophysics gets most of it back



ONLY cosmology varies–SC-SAM kept to values fitting z=0 observations

How much information about cosmology is lost when varying astrophysics?

Some, but sufficiently varying over astrophysics gets most of it back



Also vary SC-SAM alongside cosmology — clustering looks different, can the NN learn?

CAMELS SAM Clustering!

Neural network setup & assessing its results:

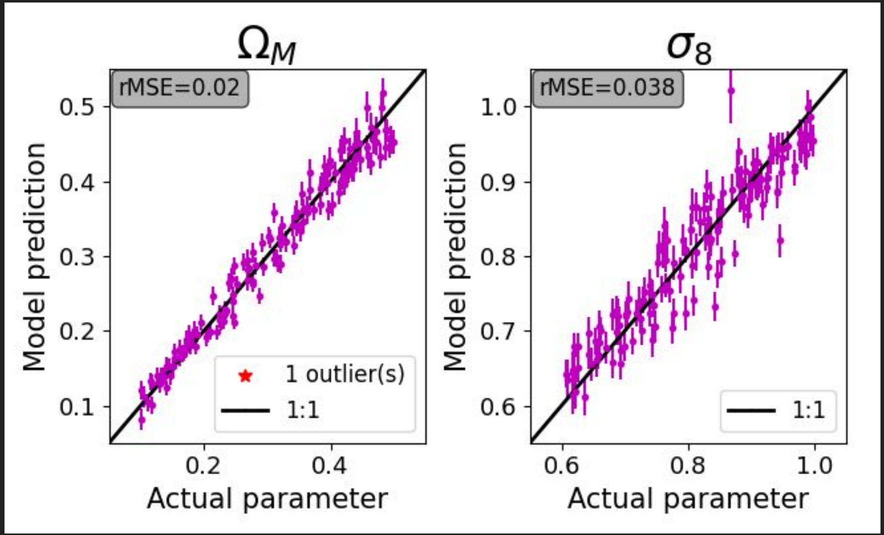
Clustering values for a set of galaxies, labeled w/ parameters
All clustering of 5000 randomly sampled SAM galaxies with stellar mass > 10^9 h^{-1} M_{solar}

→ **Neural network**
750/150/150 split training/validation/test

→ **Predictions for the 5 input parameters!**

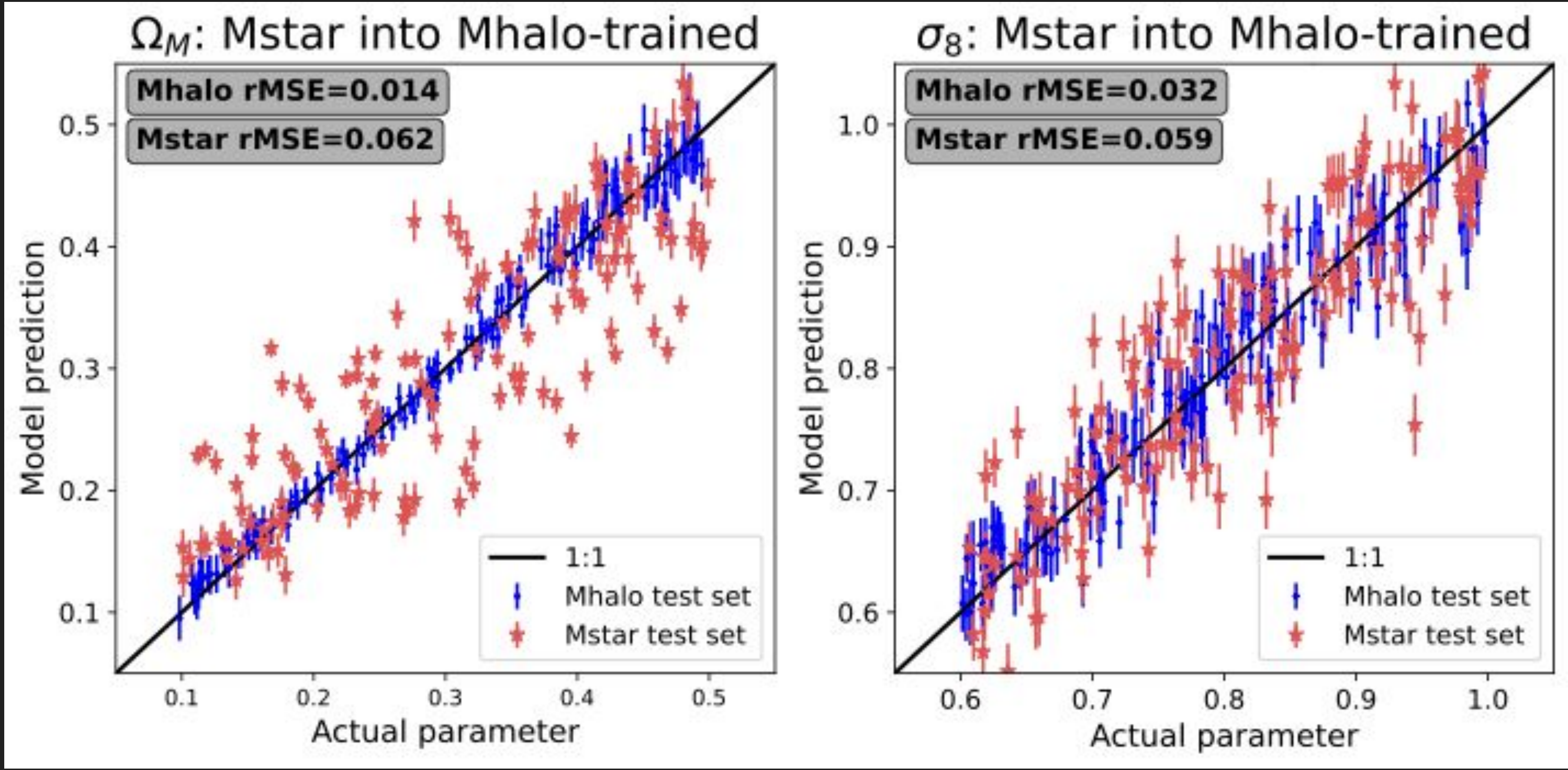
Parameter regression + likelihood-free inference
Moment networks, Jeffrey & Wandelt 2020: predict mean *and* standard deviation of marginal posterior. Great for cases with strong & weak parameters!

$$\mathcal{L}_{\mathrm{LFI}} = \sum_{i=1}^{5} \log\left(\sum_{j\in\mathrm{batch}} (\theta_{i,j\ \mathrm{true}} - \mu_{i,j})^2\right) + \sum_{i=1}^{5} \log\left(\sum_{j\in\mathrm{batch}} \left((\theta_{i,j\ \mathrm{true}} - \mu_{i,j})^2 - \sigma_{i,j}^2\right)^2\right)$$



SANITY CHECK & why we have to care about astrophysics

→ Train a neural network on DMO clustering, give it stellar mass clustering?



How much information about cosmology is lost when varying astrophysics?

Some, but sufficiently varying over astrophysics gets most of it back

Clustering values for a set of galaxies
All clustering of 5000 randomly sampled SAM galaxies with stellar mass > $10^9\ h^{-1}\ M_{solar}$

↓

Neural network

750/150/150 split training/validation/test

↓

Predictions for the 5 input parameters!

Parameter & constraints likelihood-free inference / rMSE	Train & Test on dark matter only clustering (similar scale of mass cut)	Train & Test on galaxy catalogs with the same fiducial SC-SAM	Train & Test on galaxy catalogs that also varied SC-SAM parameters
Ω_m	0.014 → 4.7% 0.014 → 4.7%	0.014 → 4.7% 0.016 → 5.3%	0.014 → 4.7% 0.02 → 6.7%
σ_8	0.024 → 3% 0.032 → 4%	0.018 → 2.25% 0.028 → 3.5%	0.021 → 2.6% 0.038 → 4.75%

How much information about cosmology is lost when varying astrophysics?

Some, but sufficiently varying over astrophysics gets most of it back

- Well-trained NNs with good architecture find nearly all information!
- Using a single model of galaxy physics reinforces cosmological information
- Not including astrophysics may be giving you *overly optimistic constraints!*

Parameter & constraints likelihood-free inference / rMSE	Train & Test on dark matter only clustering (similar scale of mass cut)	Train & Test on galaxy catalogs with the same fiducial SC-SAM	Train & Test on galaxy catalogs that also varied SC-SAM parameters
Ω_m	0.014 → 4.7% 0.014 → 4.7%	0.014 → 4.7% 0.016 → 5.3%	0.014 → 4.7% 0.02 → 6.7%
σ_8	0.024 → 3% 0.032 → 4%	0.018 → 2.25% 0.028 → 3.5%	0.021 → 2.6% 0.038 → 4.75%

Can we marginalize over astrophysics to constrain cosmology with galaxy clustering? **YES!**

CAMELS SAM

Clustering!

Basics of assessing neural network results:

Parameter regression + likelihood-free inference

$$\mathcal{L}_{\rm LFI} = \sum_{i=1}^{5} \log\left(\sum_{j\in{\rm batch}} (\theta_{i,j\ {\rm true}} - \mu_{i,j})^2 \right) + \sum_{i=1}^{5} \log\left(\sum_{j\in{\rm batch}} \left((\theta_{i,j\ {\rm true}} - \mu_{i,j})^2 - \sigma_{i,j}^2 \right)^2 \right)$$

Clustering values for a set of galaxies
All clustering of 5000 randomly sampled SAM galaxies with stellar mass > $10^9\ h^{-1}\ M_{\rm solar}$

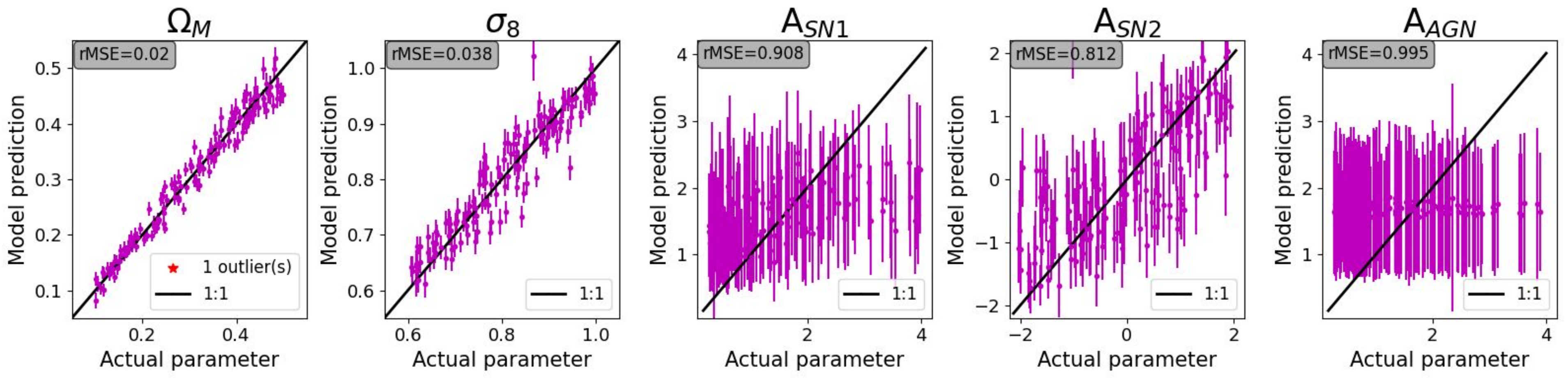
→

Neural network

750/150/150 split training/validation/test

→

Predictions for the 5 input parameters!



Can we marginalize over astrophysics to constrain cosmology with galaxy clustering?

YES! And, we can learn a little bit about the astrophysics itself!

- Well-trained NNs with good architecture find all the information!
- Using a single model of galaxy physics reinforces cosmological information
- Not including astrophysics likely gives you overly optimistic constraints!

Best of both worlds: get cosmology and some astro!!!

Parameter & constraints likelihood-free inference / rMSE	Train & Test on dark matter only clustering (similar scale of mass cut)	Train & Test on galaxy catalogs with the same fiducial SC-SAM	Train & Test on galaxy catalogs that also varied SC-SAM parameters
Ω_m	0.014 → 4.7% 0.014 → 4.7%	0.014 → 4.7% 0.016 → 5.3%	0.014 → 4.7% 0.02 → 6.7%
σ_8	0.024 → 3% 0.032 → 4%	0.018 → 2.25% 0.028 → 3.5%	0.021 → 2.6% 0.038 → 4.75%
A_{SN}	N/A	N/A	**40% on A_{SN1}** **30% on A_{SN2}**

Reflecting on this proof-of-concept

CAMELS SAM

Is this approach better to use than the ‘traditional’ method of constraining cosmology with galaxy clustering?

- **Pros:**
  - Don’t have to identify a likelihood or create a covariance matrix/emulator for one cosmology
  - Probing non-linear scales & non-Gaussian statistics with more realistic galaxies
- **Cons:**
  - There’s a reason observational cosmologists love their full posterior distributions
  - Constraints could be much better, and *must* be–how do we improve them?
  - Clustering is likely an unoptimized application of CAMELS-SAM

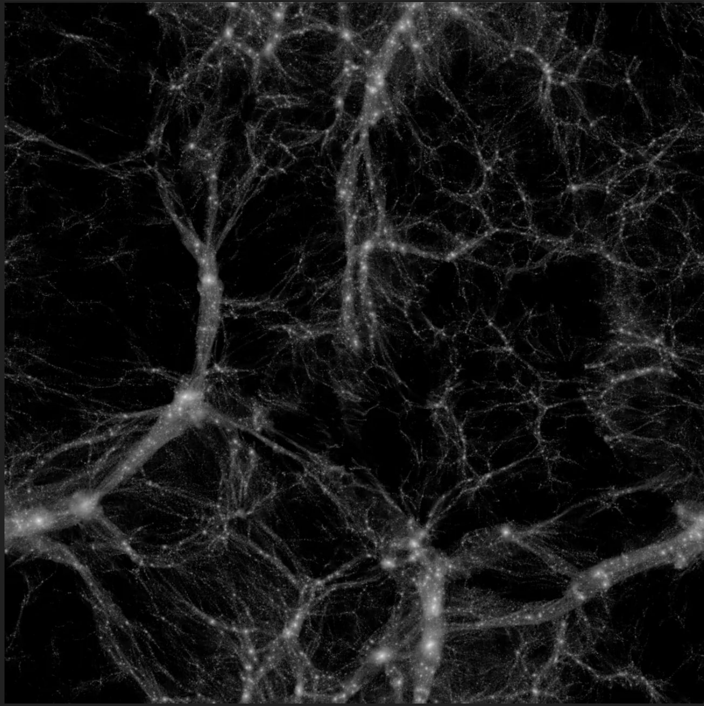
CAMELS SAM

Lucia A. Perez
(postdoc @ Princeton & CCA)
lucia.perez@princeton.edu

Let's collaborate
and do cool science!

Key results from Perez, Genel, et al. (2023):

- NNs *can* marginalize over astrophysics to constrain cosmology with clustering!
- Use more than two-point statistics to improve constraints–VPF and CiC!
- SAM galaxy clustering measures cosmology well, near DM constraints!
- Using a SAM, clustering *does* sense astrophysics! >30% fractional error



CAMELS-SAM public data | *camels-sam.readthedocs.io*

1000+ simulations with 100 snapshots between 27 < z < 0:

- $(100\ h^{-1}\ \mathrm{Mpc})^3$ large ; $N=640^3$ particles of ~$1\text{-}6 \times 10^8\ h^{-1}\ M_{sol}$
- *ROCKSTAR* halo catalogs
- *ConsistentTrees* merger trees
- Santa Cruz SAM galaxy catalogs

… all publicly available! More stuff on request