CAMELS SAM

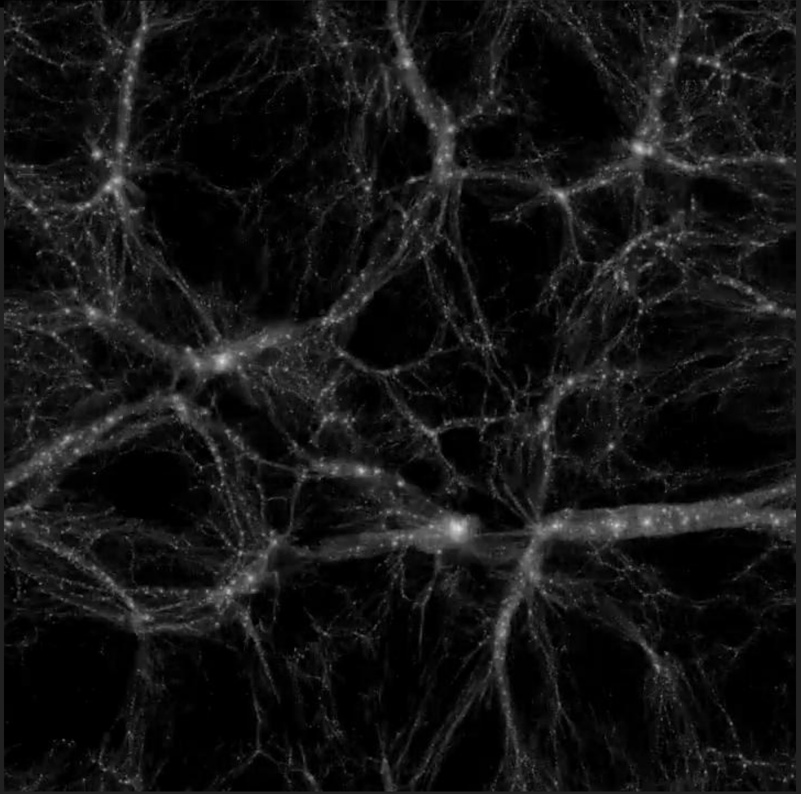
Lucia A. Perez (Princeton & CCA)
Shy Genel, Paco Villaescusa-Navarro, Rachel Somerville, Daniel Angles-Alcazar, Austen Gabrielpillai

New large-volume simulation 'hump' of CAMELS project

- 1000+ N-body simulations: $(100\ h^{-1}\ \mathrm{Mpc})^3$ large ; $N=640^3$ particles of ~1-6 x $10^8\ h^{-1}\ M_{sol}$; 100 snapshots between $0<z<27$
- Cosmological parameter space: Ω_m (fraction of energy density in DM+baryons) & σ_8 (~amplitude of density fluctuations)
- Run 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?

Proof-of-concept in Perez+2022: constraining power of galaxy clustering statistics (3D two-point correlation function, count-in-cells, Void Probability Function)



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

CAMELS SAM

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

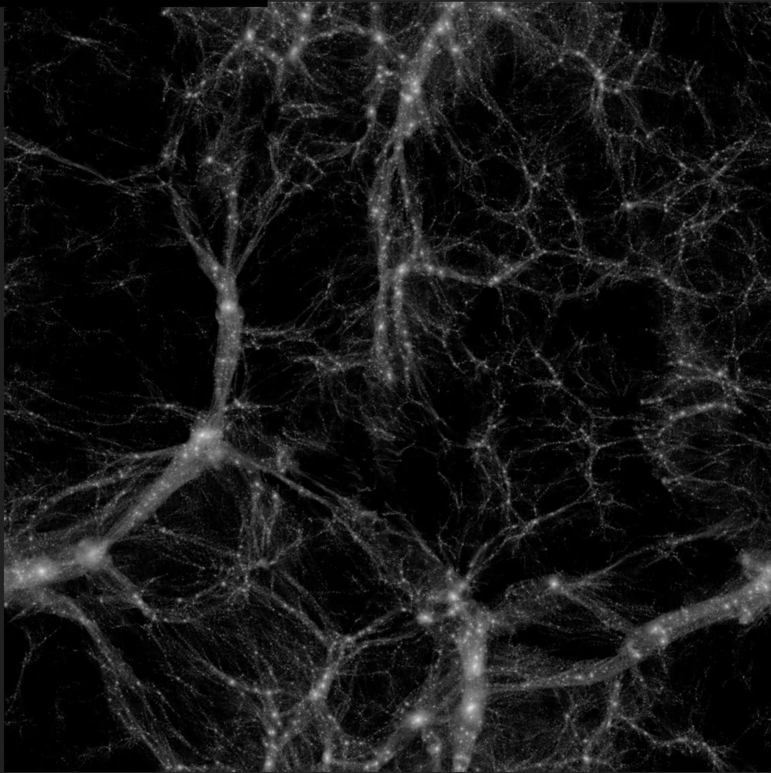
CAMELS-SAM public data | *camels-sam.readthedocs.io*

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: fiducial Ω_m=0.3, σ_8=0.8, default SC-SAM, unique random seeds
- **12 ~1P** galaxy catalogs: fiducial cosmology, min/max SC-SAM parameters for 2 unique random seeds



#643: Ω_m = 0.131 ; σ_8 = 0.986 Lucia A. Perez

The Santa Cruz SAM:

A_{SN1} x, +A_{SN2}

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

A_{AGN1} x

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

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

Unique/notable in the SC-SAM:

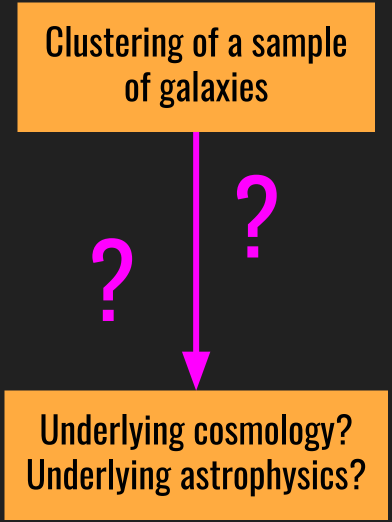
- Multiphase partitioning & tracking of the ISM
- How supermassive black holes form and grow, ‘black hole feedback’

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

Clustering!

Why clustering?



CAMELS SAM

Clustering!

Why clustering?

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

Clustering!

Why clustering?

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!

Underlying cosmology?
Input astrophysics?

Predictions only as good as training!
Interpretability?

Big CAMELS-SAM suite

Theory of clustering & statistics

Galaxy clustering as Lucia does it:

Two-point correlation function

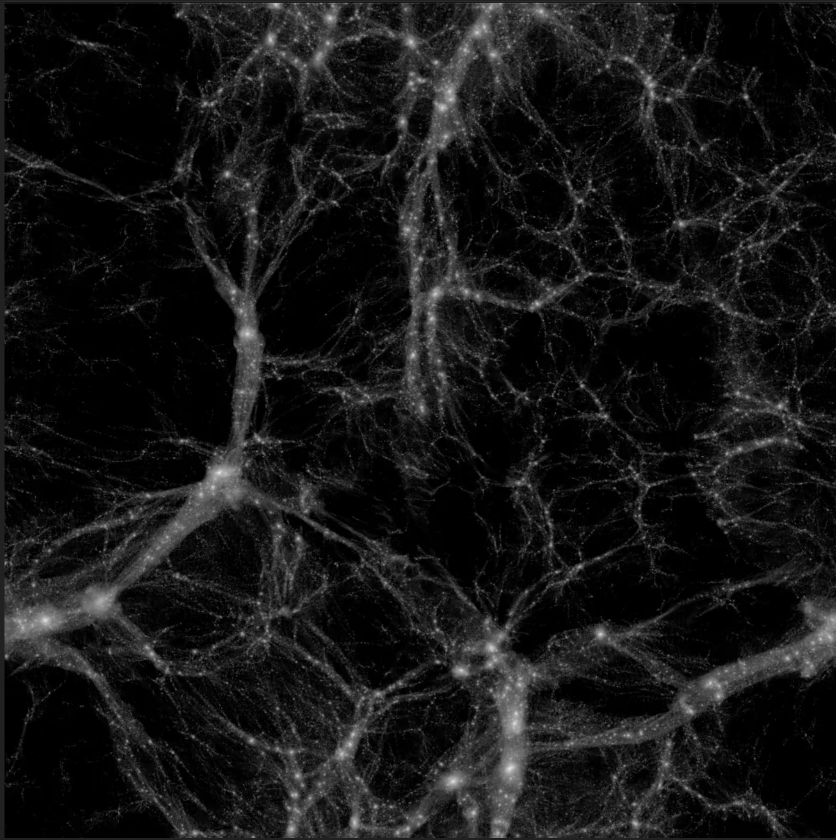
- Fourier Transform of power spectrum, common in observations
- Compare galaxies to a random distribution; pair counts
- Brief summary statistic: 1 R gives 1 ξ value

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

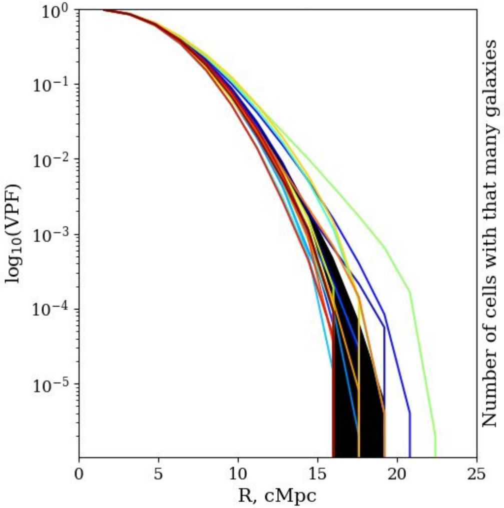
- Only *empty* test spheres–very cheap to calculate
- Influenced by higher order correlations
- Brief summary statistic: 1 R gives 1 VPF value



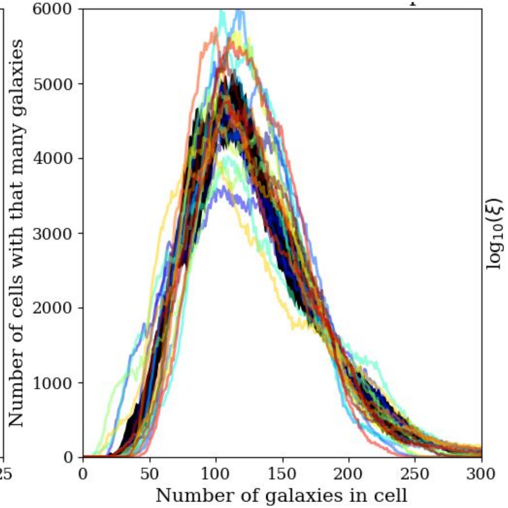
CAMELS SAM

Clustering!

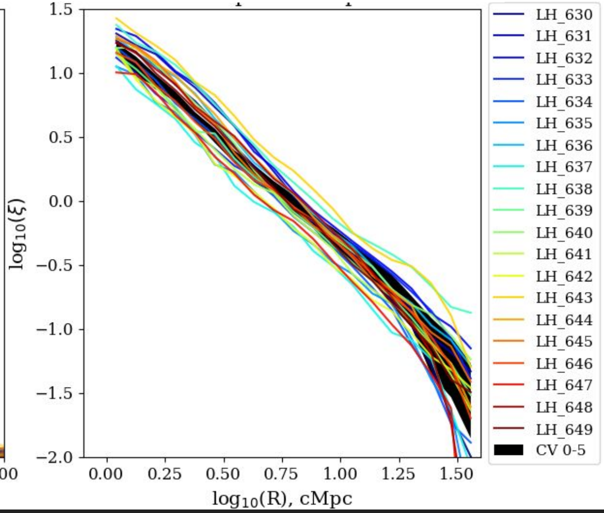
Void Probability Function



Count in Cells PDF (R=27.2 cMpc)



Real space 3D 2pt Correlation Func.



Best at Ω_M

Best at σ_8

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!

CAMELS

SAM

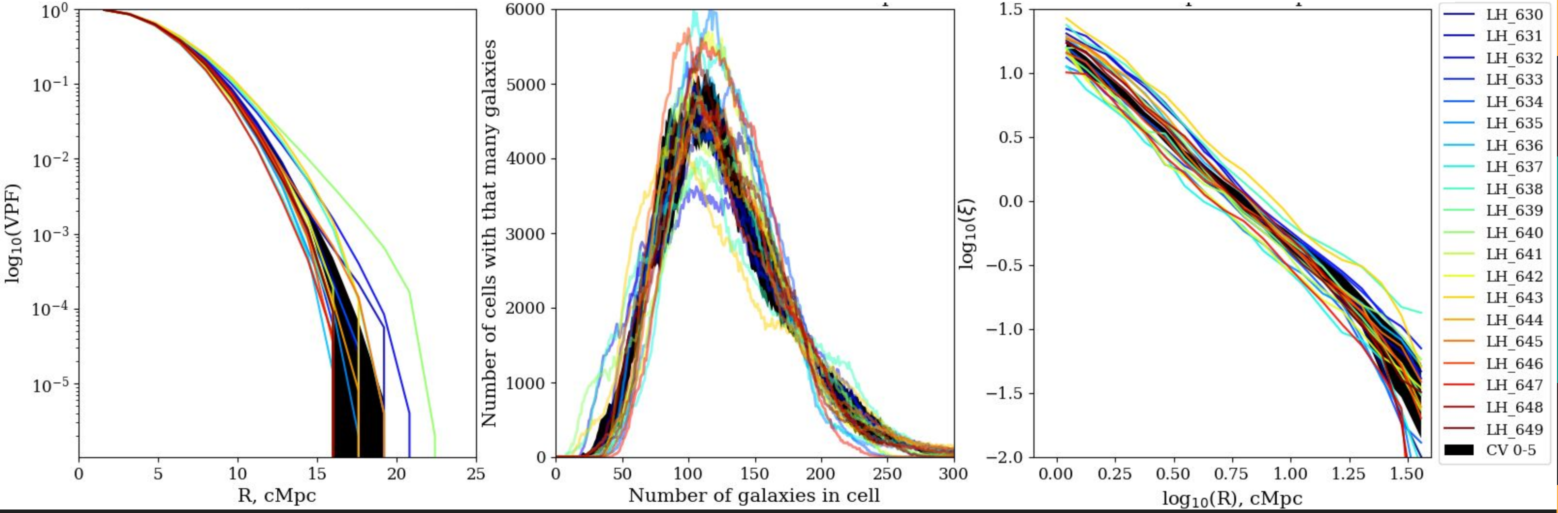
Clustering!

Corrected for number density!!!!

Void Probability Function

Count in Cells PDF (R=27.2 cMpc)

Real space 3D 2pt Correlation Func.



Best at Ω_M

Best at σ_8

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!

CAMELS SAM Clustering!

Basics of assessing neural network results:

Parameter regression + likelihood-free inference

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

→

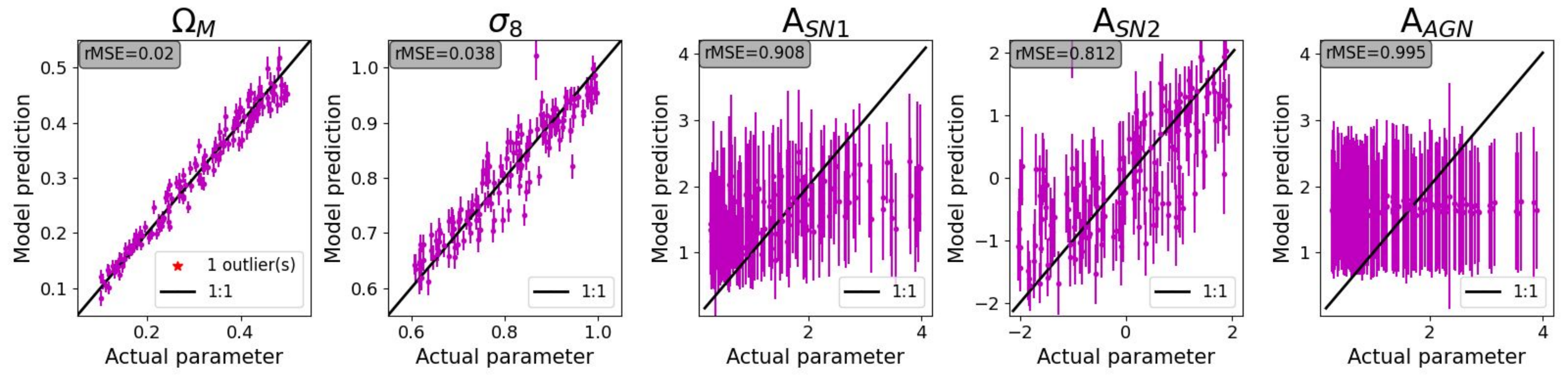
Neural network

750/150/150 split training/validation/test

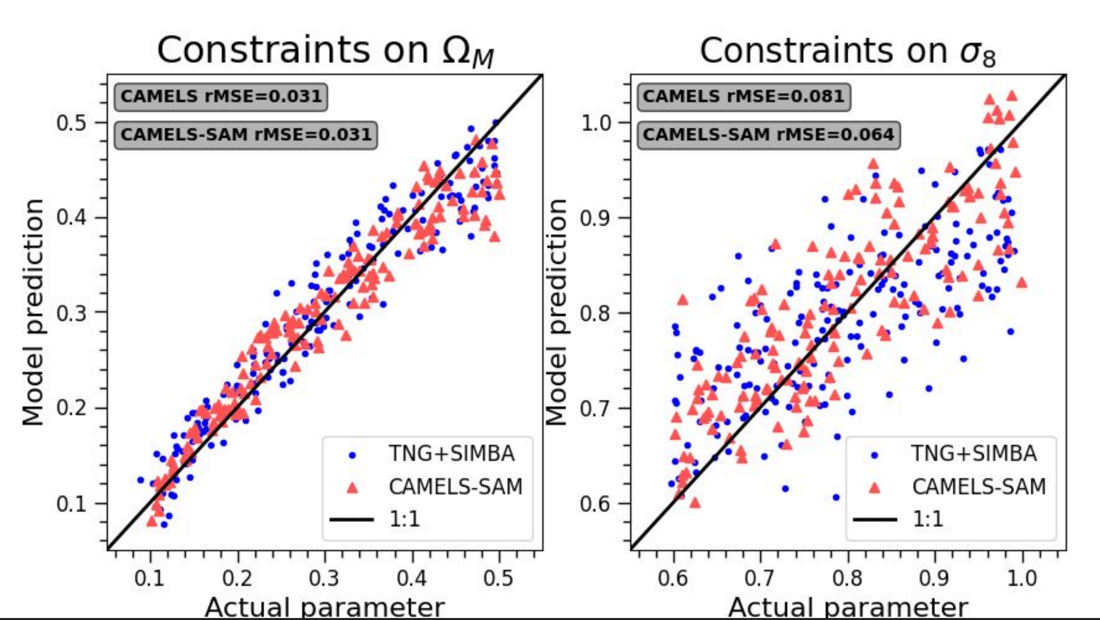
→

Predictions for the 5 input parameters!

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



OK, how rough are constraints with galaxy clustering in *CAMELS* ?



2ptCF, CiC, VPF at z={0, 0.1, 0.5, 1} for 1000 randomly selected halos with mass greater than:

- 2e10 M_{sol} h^{-1} CAMELS SIMBA+TNG (blue; within 10cMpc)
- 1.2e12 M_{sol} h^{-1} CAMELS-SAM (red; within 40 cMpc)

1. Cosmic variance worsens constraints
2. Small volume focuses to small scales
3. Almost *no* galaxy selection can be done across SIMBA *and* TNG!
   a. SIMBA makes way more galaxies
   b. Poisson noise starts mattering
4. Best case scenario–high density dark matter only selection–gets 10% errors on cosmology
5. CAMELS-SAM at much lower density meets or exceeds these easily, *and* allows for galaxy selections
6. Can we do it better? Rank-order galaxy selection! Stick to one CAMELS hump! Do this experiment properly

CAMELS SAM Clustering!

Takeaways from initial clustering work:

1. Use more than two-point statistics to improve constraints
2. SAM galaxy clustering measures cosmology well! Beginning to reach non-linear scales, too
3. Using a SAM, these clustering statistics do sense astrophysics!

How good are our constraints from clustering?

- Cosmology: 3-15% fractional error
  - Best-performing selections: low-threshold stellar mass
- Astrophysical SC-SAM feedback: >30% fractional error
  - Note: using clustering with low-hanging fruit ('pure' properties)

Galaxy clustering: more observable, loses info…

FULL 2D maps of CAMELS get 3-4% on Ω_M–what could CAMELS-SAM's do?!

CAMELS SAM

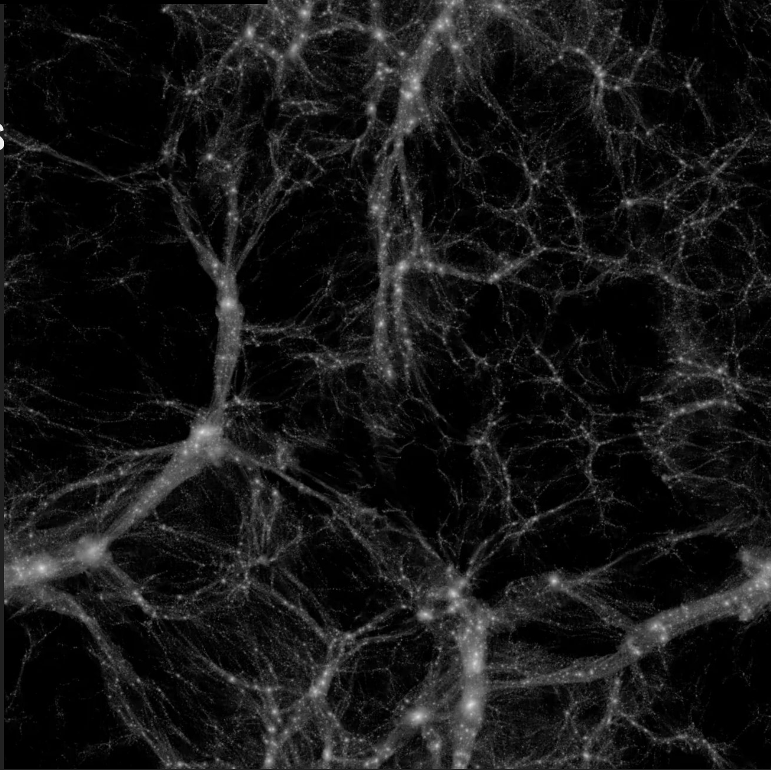
Lucia A. Perez
(Princeton & CCA)
Shy Genel, Paco Villaescusa-Navarro, Rachel Somerville, Daniel Angles-Alcazar, Austen Gabrielpillai

CAMELS-SAM work in progress:

- A new set of LH SC-SAM catalogs over *all* SC-SAM parameters (Z. Tremitiere)
- Influence of baryonic physics on PNG bias? (w/ A. Barreira)
- A new LH hump with the Galacticus SAM *(open to collaborators!)*

Work with me!

- Let’s run some HODs!
- More SAMs?
- Other work with the DMO volumes?



#643: Ω_m = 0.131 ; σ_8 = 0.986   Lucia A. Perez

Answering big questions with CAMELS SAM

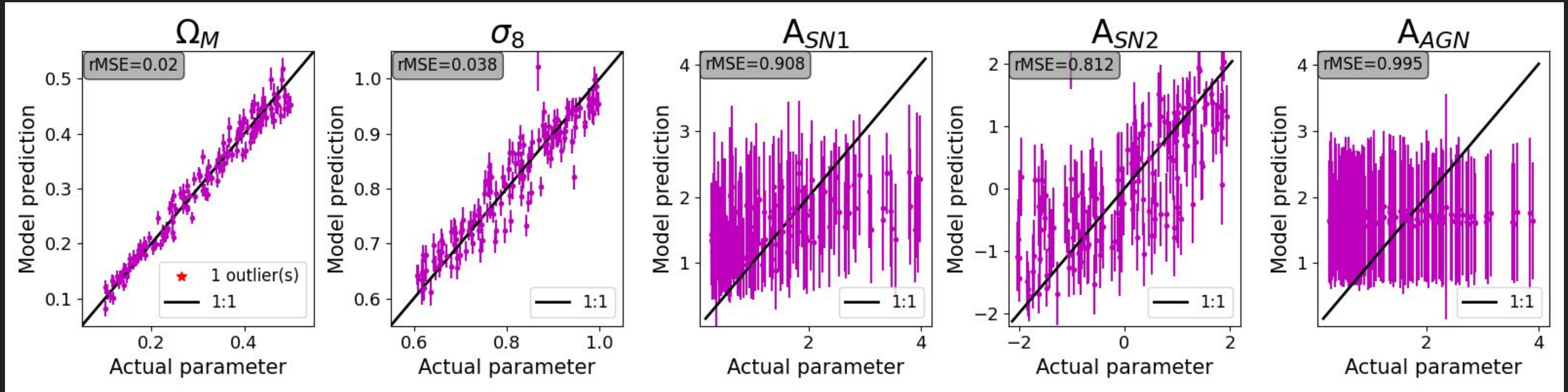
How much information is lost to astrophysics?

- Try running the most realistic fiducial SC-SAM over all the cosmologies:
  - Then, constraints from galaxy clustering are *at least as good as* DMO clustering!
  - Constraints slightly better for σ_8, without SC-SAM variation, especially if don't correct for number density
- The neural network maintains good accuracy on cosmology even when including astrophysics!
  - Cosmology constraints only a few percent worse (e.g. 9% vs. 5%)
  - Still learn about some of the SC-SAM parameters

Other big questions from CAMELS SAM

What happens if our ranges are just too big? Are we hurting our own efforts?

- Definitely want to avoid priors that bias the neural network
- Is a very broad astrophysical parameter range limiting the constraints on cosmology?
  - Note: predicting parameters one at a time does the same or a little worse than all 5
- How do you tell apart *'you need more training data'* vs. *'no information exists'*?



Other big questions from CAMELS SAM

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

- **Pros:**
  - Don’t have to identify a likelihood, or creating a covariance matrix, or create emulator for one given cosmology
  - Instead, create a large & representative enough training set
  - We’re probing non-linear scales with robust galaxies that theory can struggle with
- **Cons:**
  - Constraints could be much better, how do we improve them?
  - Is clustering an unoptimized application of CAMELS-SAM?

CAMELS SAM

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

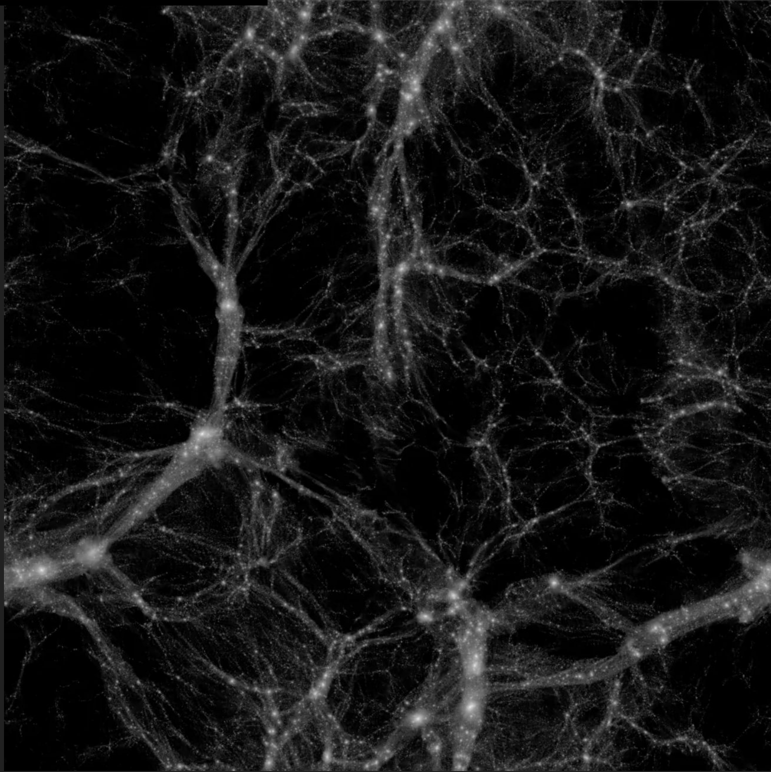
CAMELS-SAM public data | *camels-sam.readthedocs.io*

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
- *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: fiducial Ω_m=0.3, σ_8=0.8, default SC-SAM, unique random seeds
- **12 ~1P** galaxy catalogs: fiducial cosmology, min/max SC-SAM parameters for 2 unique random seeds



#643: Ω_m = 0.131 ; σ_8 = 0.986   Lucia A. Perez