LSS Parameter Inference with Machine Learning - Recent Results and Challenges

CIPNAP 2025, University of Wisconsin-Madison

June 11th 2025

Moritz Münchmeyer, UW Madison

Funded by the U.S. National Science Foundation and the Simons Foundation

Talk outline

- First part of the talk:
    - Short review **of the general status of Machine Learning based parameter inference in cosmology**.
- Second part of the talk:
    - Some of our **recent work on a special case**: **Machine Learning in the squeezed limit.**

Goals and Challenges for AI in cosmology

Grand-Unified Cosmology Analysis

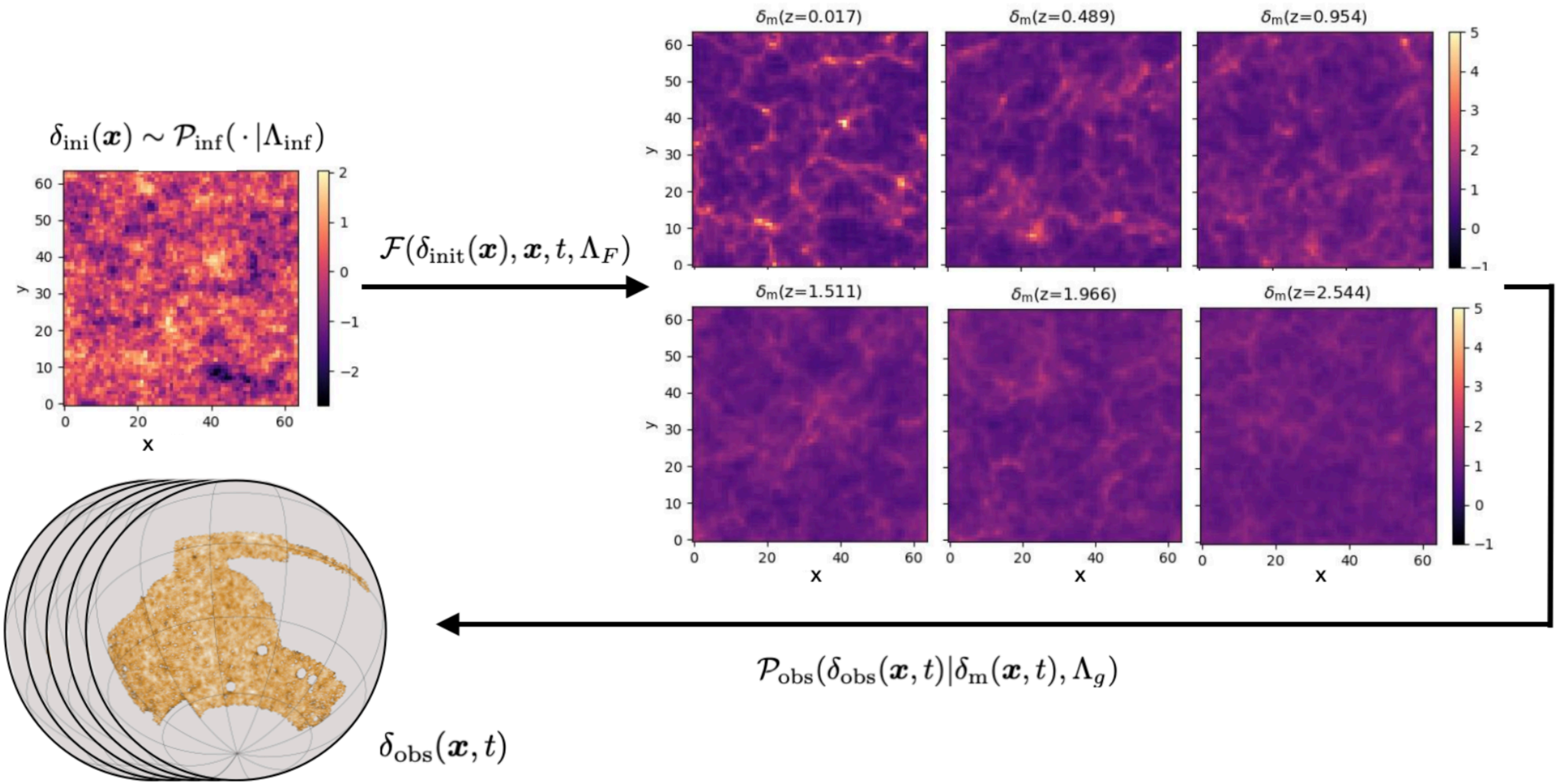


Figure credit: Chiway Chang, SkAI.

Main parameters we'd like to improve

- **Expansion**: H, w_0, w_a
  - Obviously exciting currently, IF gains are possible.
  - BAO dominated.
- Ω_m **and** σ_8.
  - Main parameters many AI in LSS methods focus on.
  - Interesting if there are tensions with other measurements (CMB).
- **Primordial power spectrum, Primordial non-Gaussianity**
  - Strong CMB data.
- I'll ignore scale-dependent, time-dependent, anisotropic, localized etc. models, where improvements can be larger.

General approaches

- There are **three directions people look at to improve constraints from LSS with these methods:**

    1. **Learning bias priors** for the PT based PS+BS large-scale analysis.

    2. Doing a **PT field level analysis** of the same scales to extract information beyond PS+BS (and break bias degeneracies).

    3. Going to **more non-linear scales scales** than PT allows by replacing it with **simulations**. Either SBI (implicit) or FLI (explicit).

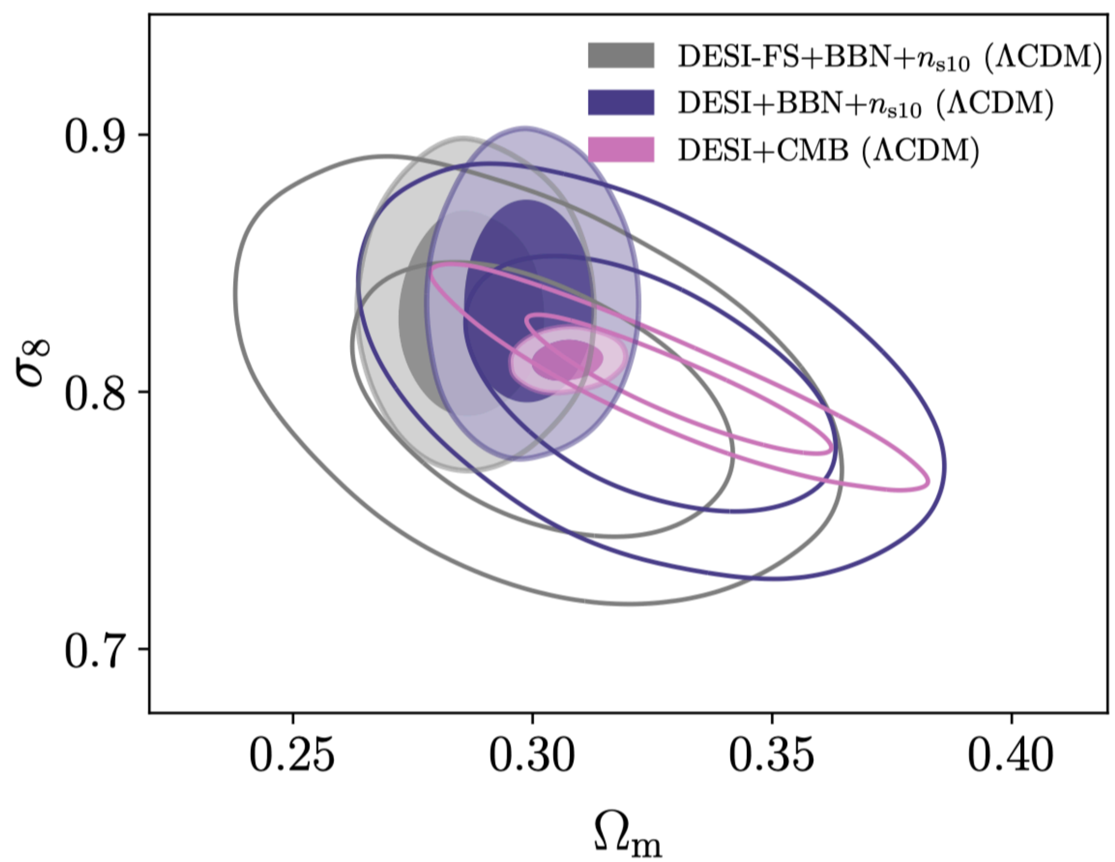
- I want to focus on results and problems for the third case.

Methods

- Many combinations of the following methods have been explored:
  - **Field level vs Summary Statistics:** E.g. use a NN vs wavelet scattering transform
  - **Model of the physics:** Perturbation theory forward model, N-body simulation, Neural Networks, combinations thereof.
  - **Re-learn vs augment**: Replace PT or N-body completely with a NN, or augment the classical method.
  - **Architectures for NN**: CNN, GNN, Normalizing Flow, Diffusion Models, Transformers, combinations thereof.
  - **Training techniques / objectives**: supervised MSE, contrastive loss, pre-training (foundation models), etc.
  - **Inference / Sampling methods**: MCMC, “likelihood free inference”, HMC, Variational Inference, Diffusion Score Matching, combinations thereof.

Problem 1: Measurements are already tight

- Measurements from the **linear and weakly non-linear regime using the power spectrum** from large volume surveys are **already strong**.



DESI + CMB:

$$\Omega_{\rm m} = 0.3072 \pm 0.0049$$
$$\sigma_8 = 0.8123 \pm 0.0052$$

Planck:

$$\Omega_m \approx 0.315 \pm 0.007$$

$$\sigma_8 \approx 0.811 \pm 0.006$$

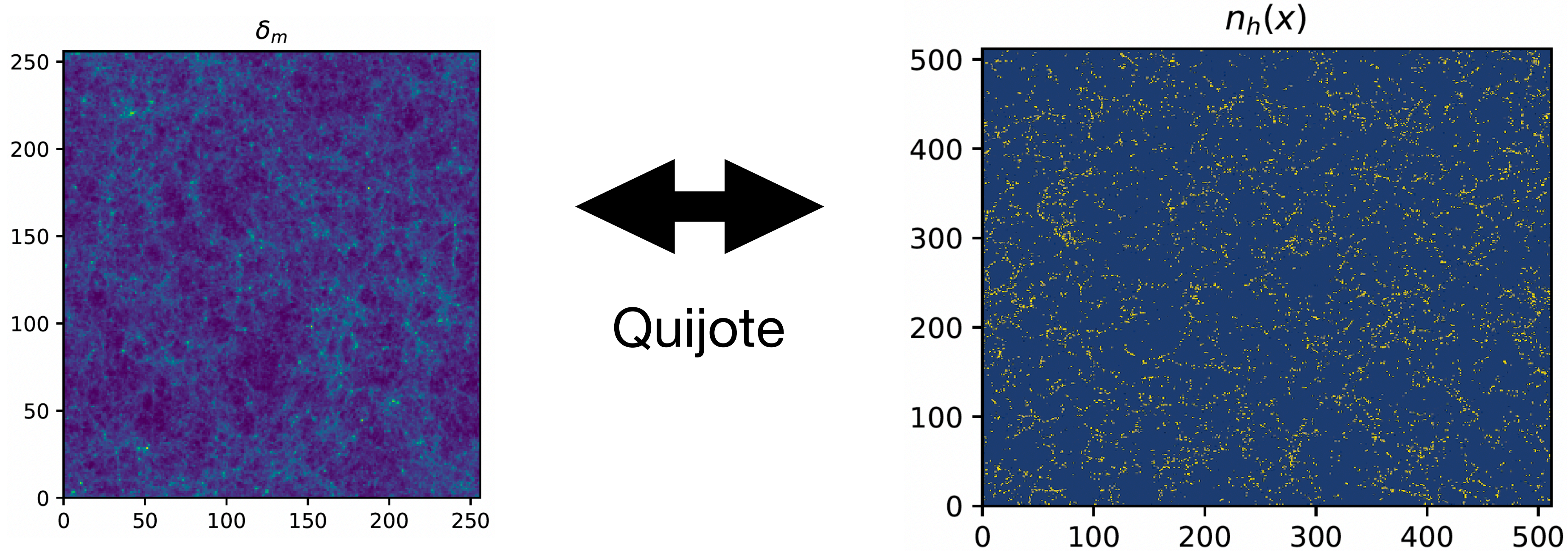
From: https://arxiv.org/abs/2504.10407

Enhancing DESI DR1 Full-Shape analyses using HOD-informed priors

- Roughly speaking, to beat large-volume surveys by accessing more non-linear scales, **we need to be able to control systematic simulation uncertainties on smaller scales at higher accuracy than the constraint from larger scales.**

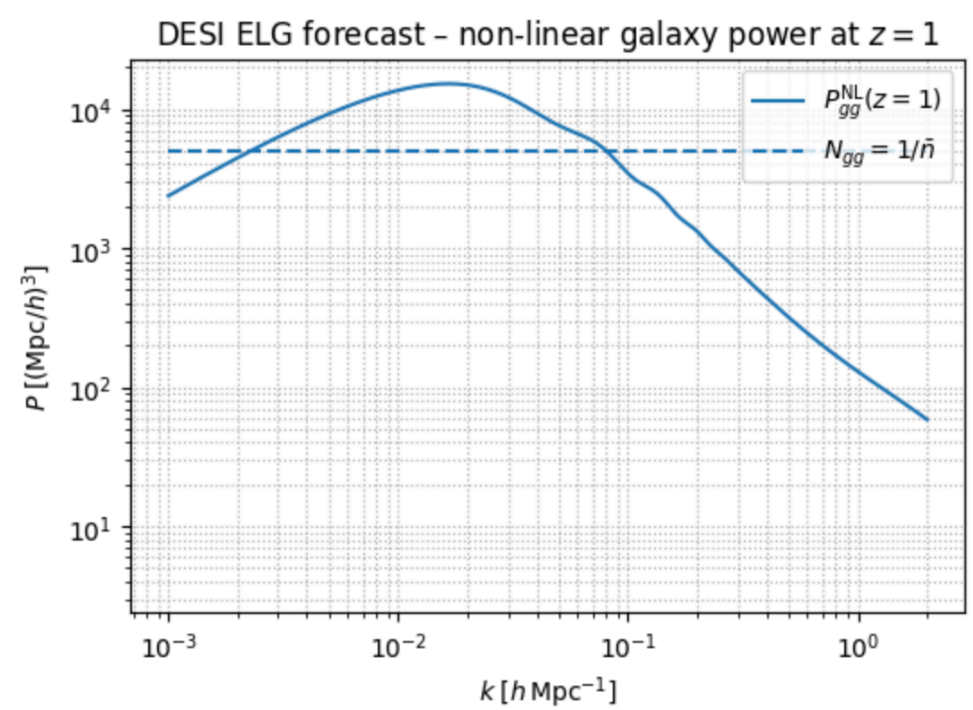
Problem 2: The shot noise is large

- A lot of works (including my own) have at first been developed at the level of the simulated dark matter distribution.



- Often the improvements achieved at matter level nearly go away at halo level.

- The perturbative range and the shot noise limited range are not so far away from another.



Problem 3: Likelihoods tend to be Gaussian

- A lot of work has gone into **learning likelihoods or posteriors of summary statistics**. However, likelihoods for summary statistics, like PS or WST, tend to be Gaussian to good approximation.
- It is always more sample efficient to put in the inductive bias of Gaussianity, when it applies.
- SBI methods also often do not work very well with realistic simulation budget, in my experience.

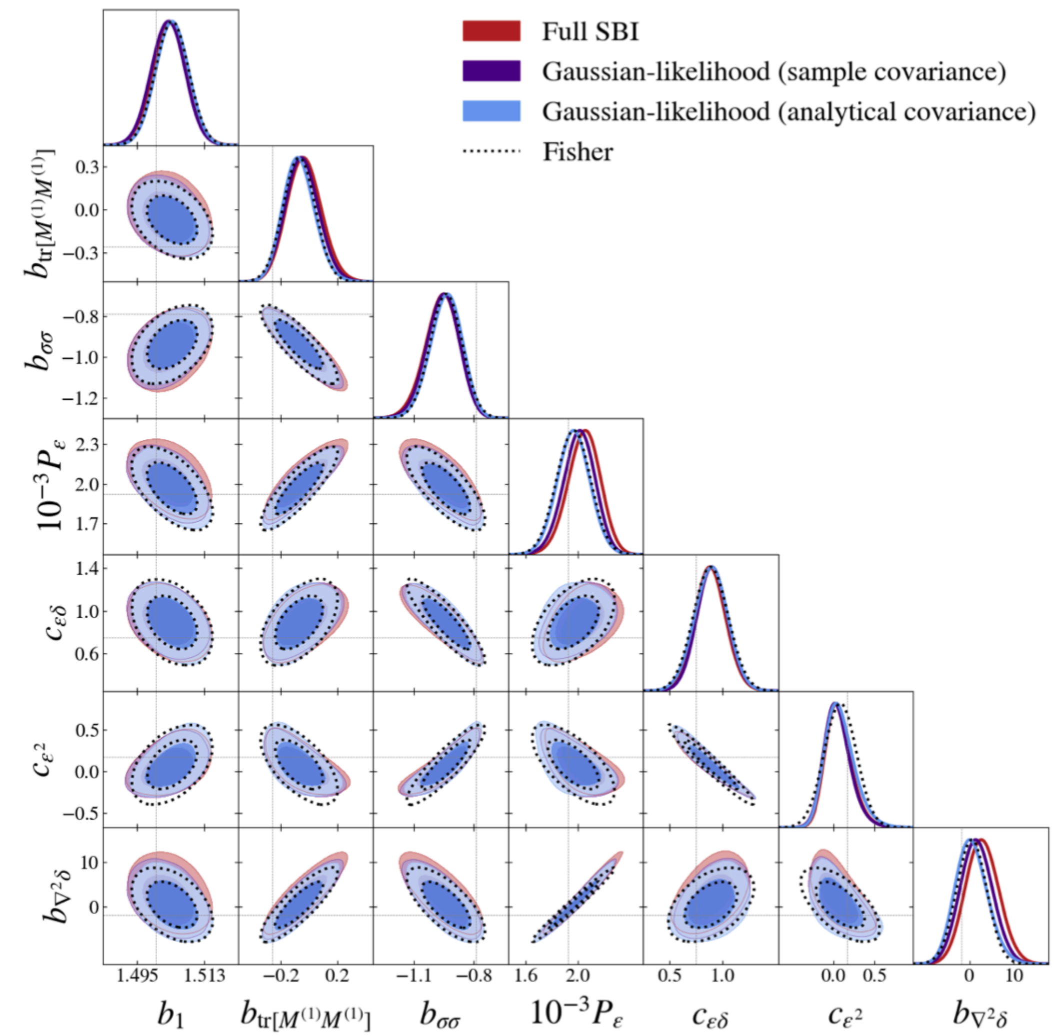


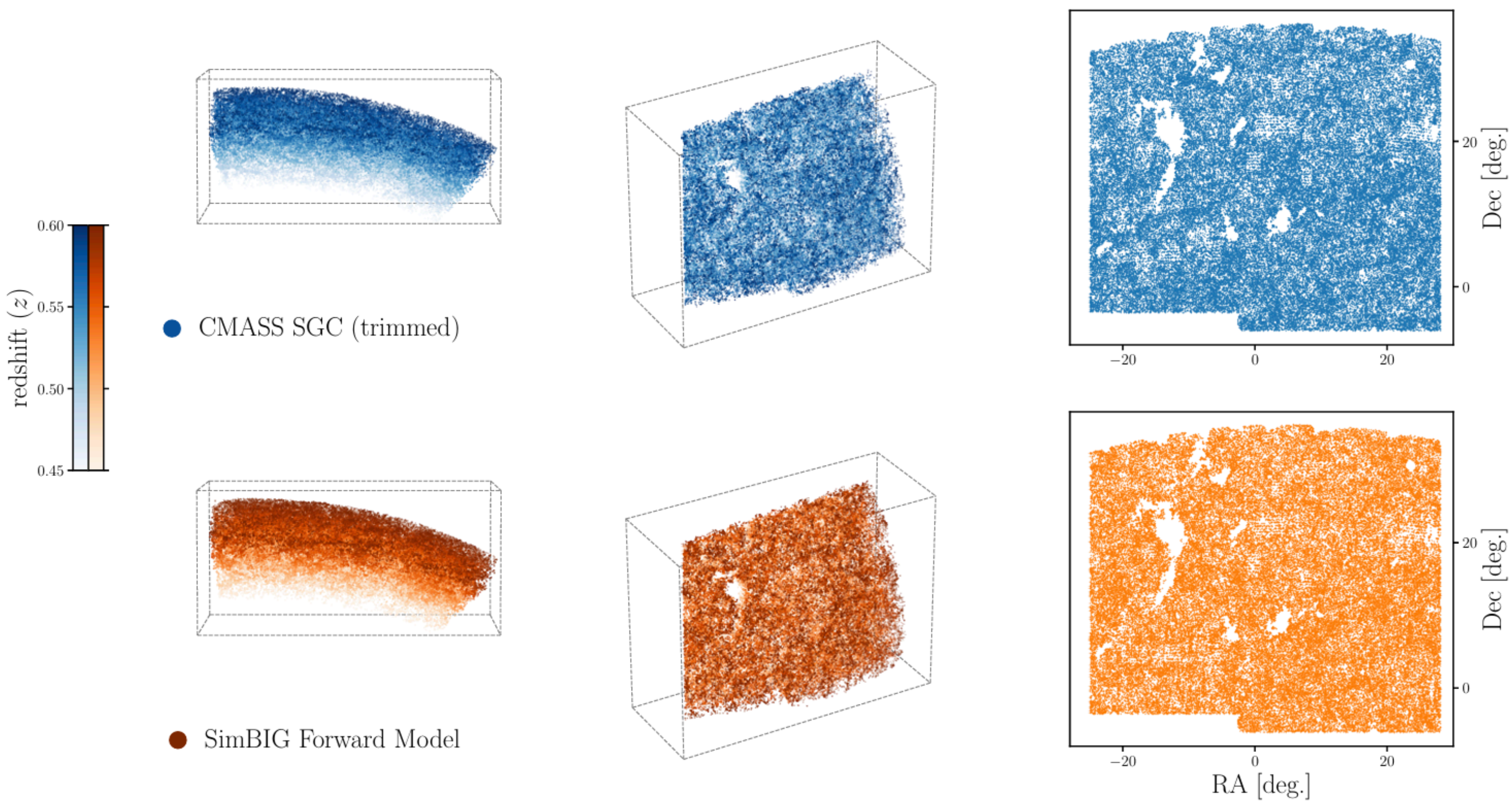
Figure 6. Parameter posterior for the case where the cosmological parameter α is fixed. Blue and purple contours correspond to the models where the data vector is sampled from a Gaussian likelihood with analytical and sample covariance, respectively, where $N_{\mathrm{cov}} = 10^4$. The red contours show the full SBI results, with no Gaussian assumption. Dotted lines indicate the Fisher prediction with sample covariance for reference. For all cases, the method NPE was used from a simulation budget of $N_{\mathrm{sim}} = 10^5$, scale cut of $k_{\mathrm{max}} = \Lambda = 0.1h\mathrm{Mpc}^{-1}$ and data vector dimension $D = 33$.

https://arxiv.org/abs/2310.03741 Tucci, Schmidt: EFTofLSS meets simulation-based inference: Sigma8 from biased tracers

Used PS+BS as summary statistics for SBI.

Example for SBI: SimBIG project

- SimBIG Forward model, based on Quijote simulations + HODs.



https://arxiv.org/abs/2211.00660 Hahn et. al. SIMBIG: Mock Challenge for a Forward Modeling Approach to Galaxy Clustering

Status of SBI methods: SIMBIG results

- https://arxiv.org/abs/2310.15246 Hahn et. al. **SIMBIG: The First Cosmological Constraints from Non-Gaussian and Non-Linear Galaxy Clustering**

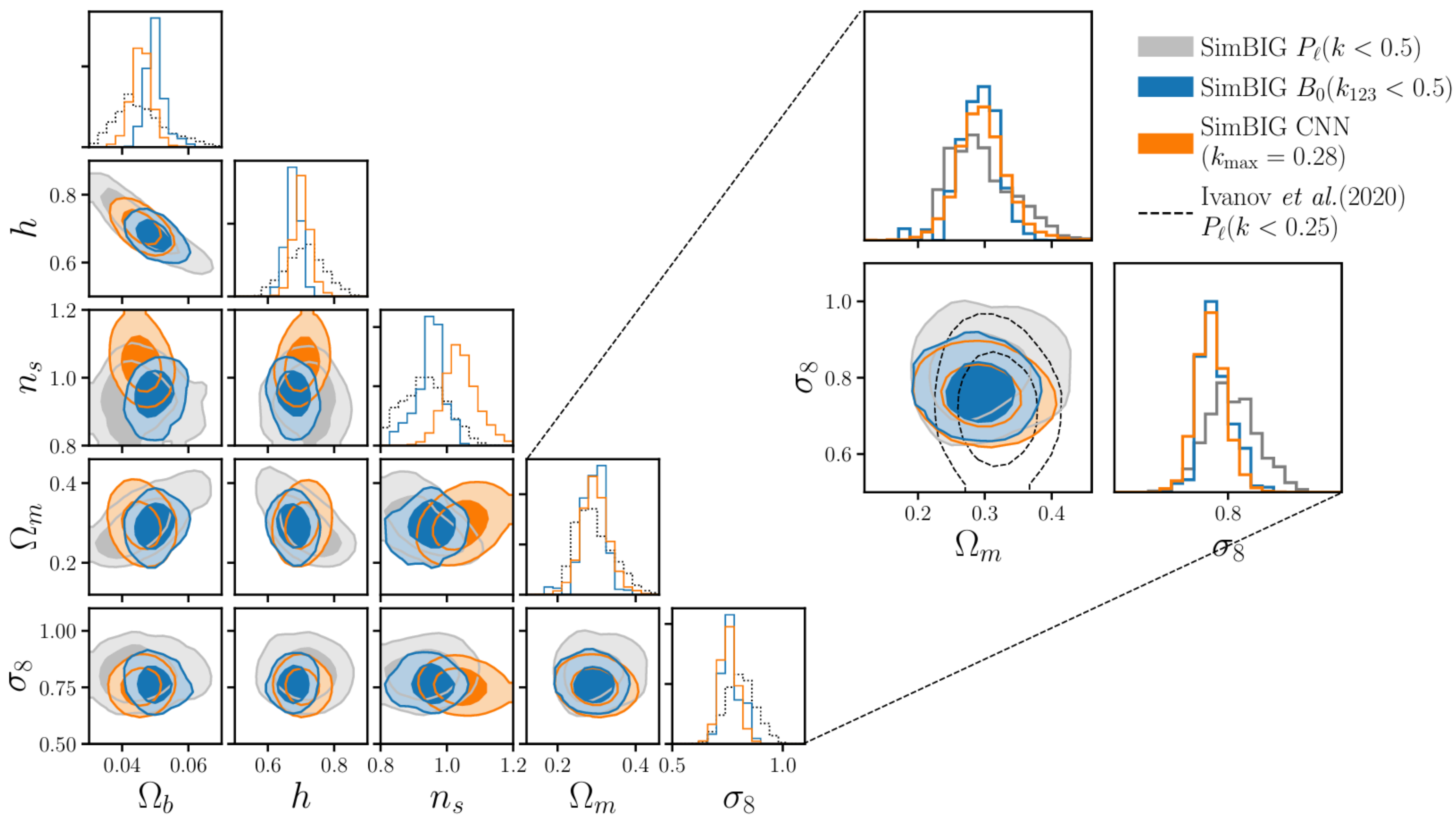


Figure 1. *Left*: Posteriors of cosmological parameters inferred from B_0 (blue) and CNN (orange) using SimBIG. All posteriors include a ω_b prior from BBN studies. The contours mark the 68 and 95 percentiles. For comparison, we include the posterior from the SimBIG P_ℓ analysis (gray). *Right*: We focus on the posteriors of Ω_m and σ_8, the parameters that can be most significantly constrained by galaxy clustering. We include the posterior from the Ivanov et al. (2020) PT-based $P_\ell(k < 0.25\,h/\mathrm{Mpc})$ analysis for reference (black dashed). The SimBIG B_0 and CNN constraints are significantly tighter yet consistent with P_ℓ constraints.

Despite only using 10% of the BOSS volume, we derive S_8 and H_0 constraints that are competitive with other cosmological probes and galaxy clustering analyses of the full BOSS volume. In short, our forward-modeling approach improves S_8 and H_0 constraints by $\sim$2.0 and 1.5× over the standard PT P_ℓ analysis. The improvement on S_8 is equivalent to applying the standard PT P_ℓ analysis to a galaxy sample *four times* the cosmic volume.

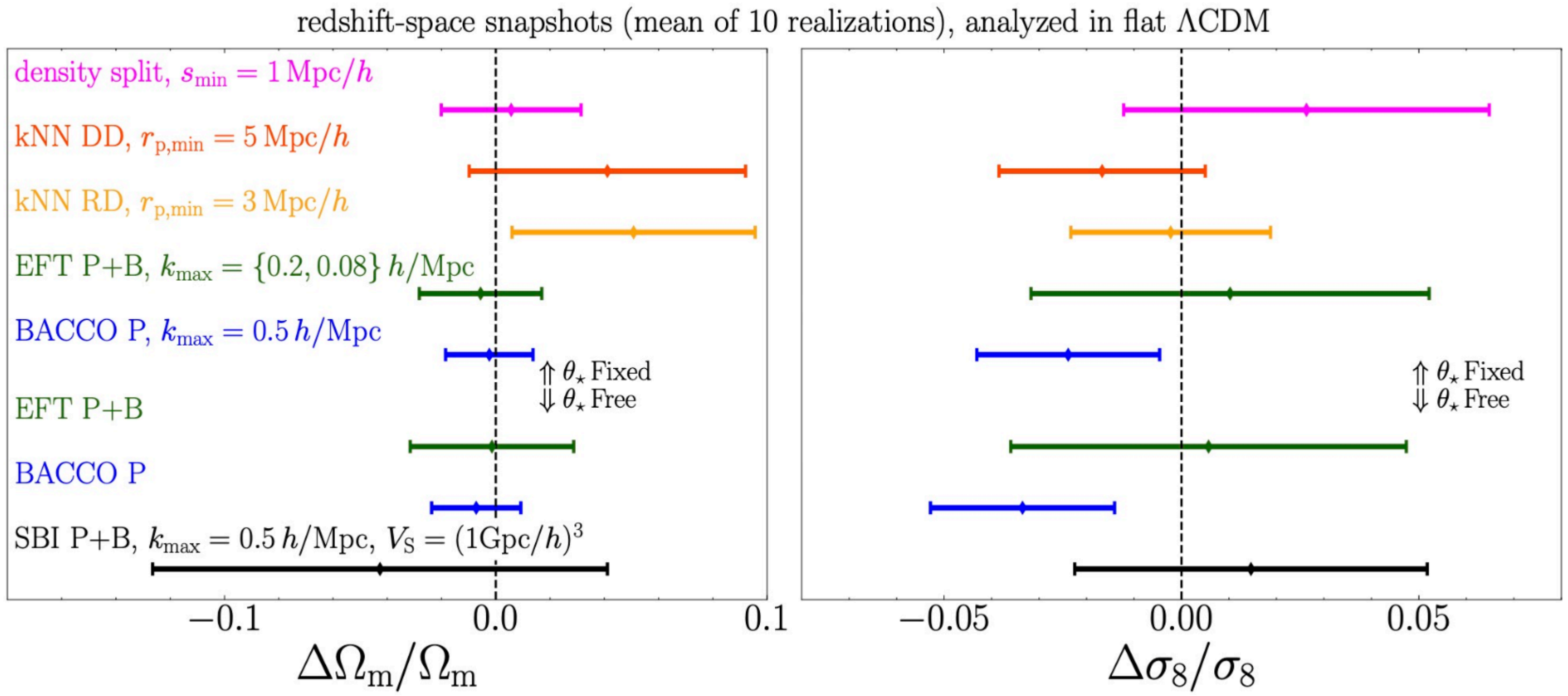
Tighter constraints on **smaller volume**, but also uses **higher k**.

Robustness depends on **forward model / HOD accuracy** and **convergence of SBI / NF** (both of which were investigated by the authors of course).

Related: https://arxiv.org/abs/2309.15071 Sensitivity Analysis of Simulation-Based Inference for Galaxy Clustering

Beyond 2pt mock challenge

https://arxiv.org/abs/2405.02252 Krause et. al. **A Parameter-Masked Mock Data Challenge for Beyond-Two-Point Galaxy Clustering Statistics**



- The most systematic comparison of methods so far.
- Bacco power spectrum emulator was leading (going to high k).
- SBI result only used a sub-volume (1/8th). Illustrates the difficulty of scaling up volume in this approach.
- FBI only applied to a simpler setup in this first round.

Squeezed-Limit Machine Learning for local NG

Yurii Kvasiuk,
Grad student at UW Madison

Kendrick Smith,
Perimeter Institute

Utkarsh Giri,
Postdoc at Caltech

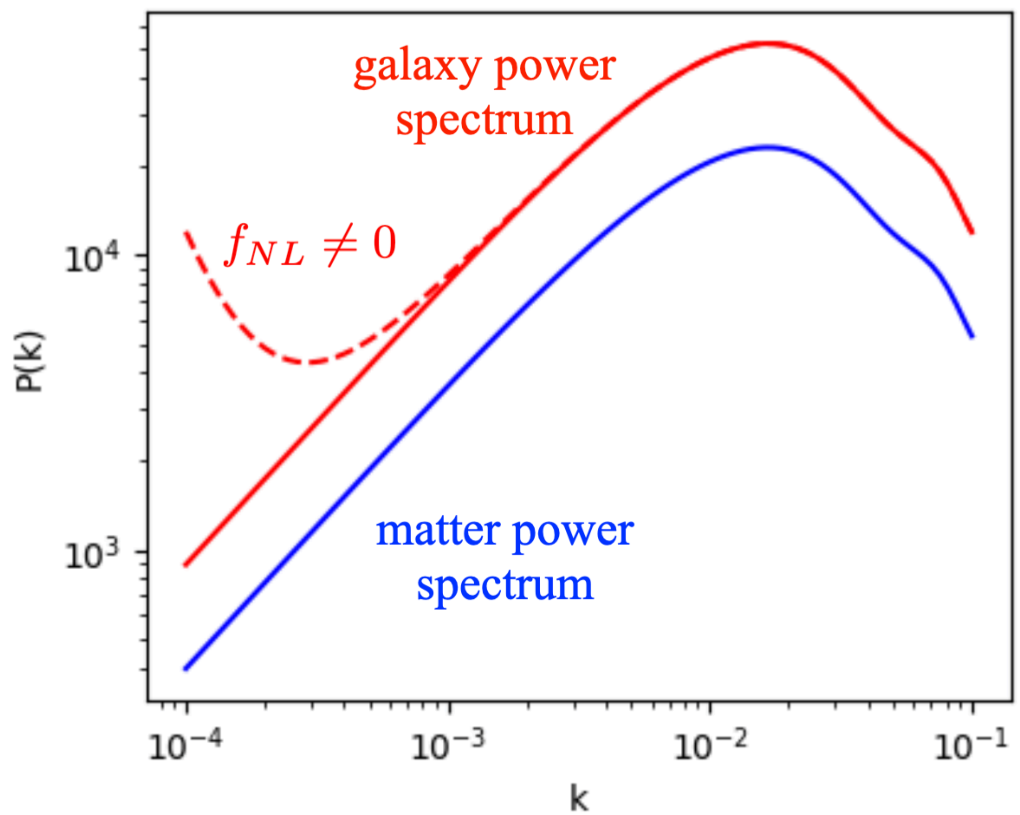
arXiv:2410.01007 A Tale of Two Fields: Neural Network-Enhanced non-Gaussianity Search with Halos
https://arxiv.org/abs/2205.12964 Robust Neural Network-Enhanced Estimation of Local Primordial Non-Gaussianity

General Idea

- There are situation in cosmology where we **use very non-linear small-scale structure to infer large-scale modes from them**.
- In that case, **uncertainty over small-scale physics can be parametrized in terms of one or more bias parameters on large scales**.
- If we use machine learning in this way, we **can create robust methods**.
- Of course this **only works for specific parameters**, and is not in competition with the methods I discussed previously.

Scale-dependent bias and f_{NL}

Local non-Gaussianity f_{NL} generates an excess clustering on large scales.This effect is called **scale-dependent bias** of the halo field (0710.4560).



$$\delta_h(\mathbf{k}_L) = \left(b_G + b_{NG} \frac{f_{NL}}{k_L^2} \right) \delta_m(\mathbf{k}_L)$$

Gaussian bias

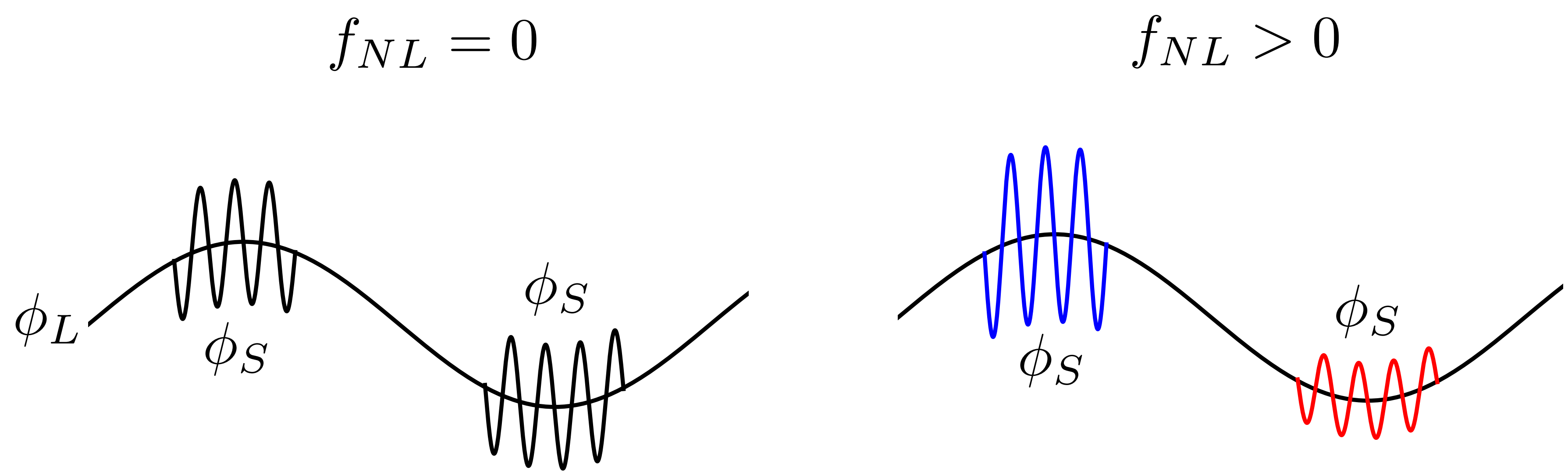
non-Gaussian bias, measures f_{NL}

The “kink” in the power-spectrum cannot be introduced by non-linear astrophysics. This **robustness is ultimately a consequence of Einstein’s Equivalence Principle**.

We have a symmetry protected observable. Can we enhance its SNR with a NN without spoiling the robustness?

The physics of local non-Gaussianity

- Local non-Gaussianity is **a large-scale modulation of small-scale power**.



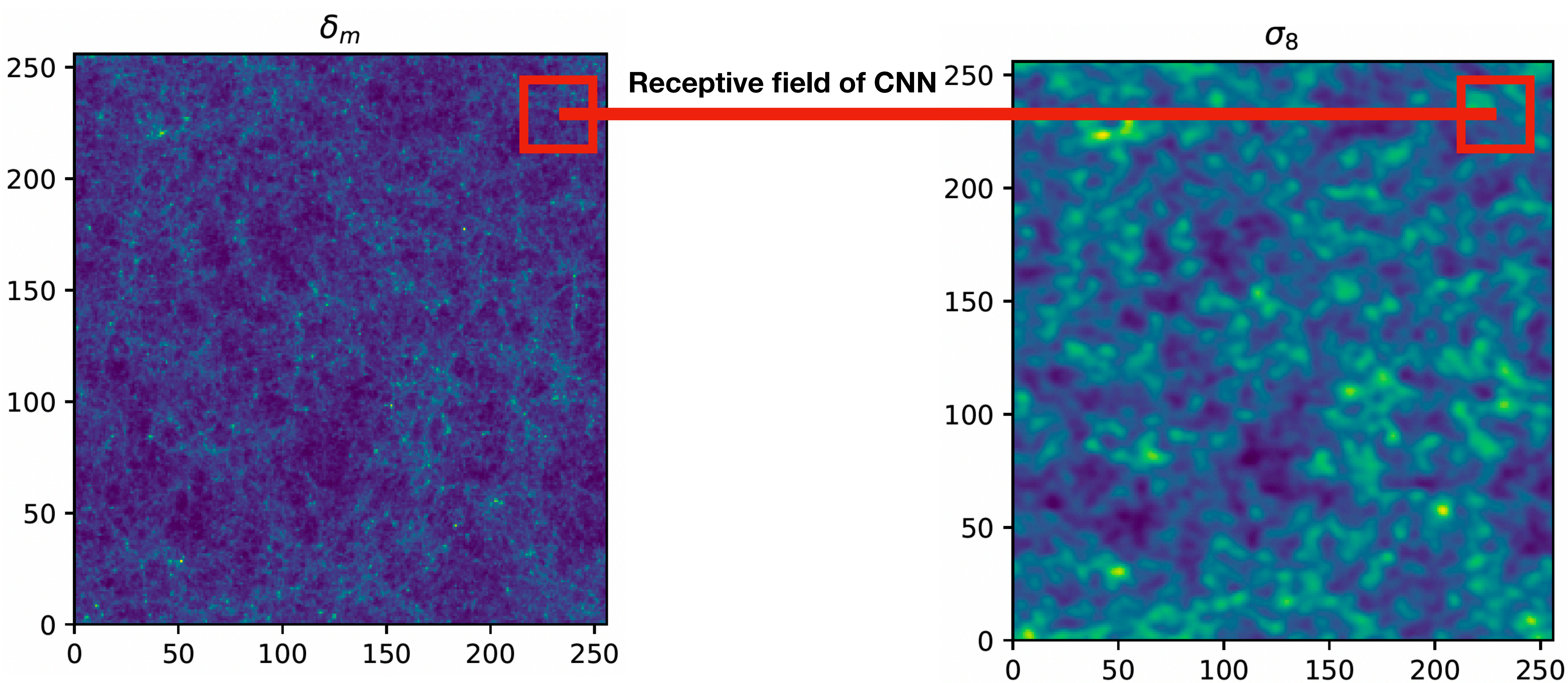
$$\sigma_8^{\mathrm{loc}}(\mathbf{x}) = \big(1 + 2 f_{NL} \Phi_l(\mathbf{x})\big)\, \bar{\sigma}_8$$

- The traditional scale-dependent halo bias works because the **local halo density is a sensitive measurement of** $\sigma_8^{\mathrm{loc}}(\mathbf{x})$.

- **Intuitively**: If we measure $\sigma_8^{\mathrm{loc}}(\mathbf{x})$, the primordial power spectrum amplitude, as well as possible, and we also measure the modulating long range mode as well as possible, then we should be able to make an optimal estimate of f_{NL}.

CNN for local σ_8 measurements

Idea: Local non-Gaussianity f_{NL} is **a large-scale modulation of local power**. Therefore if we have a **neural network that optimally probes local power (i.e. local σ_8)**, it would be the ideal field to base an f_{NL} estimate on.



This local **NN output field will also have a scale-dependent f_{NL} bias.**

Scale Scale-dependent bias in the π-field formalism

- Any field π(x) that is sensitive to $\sigma_8^{\mathrm{loc}}(\mathbf{x})$ can be modeled on large scales in the same way as the familiar halo field:

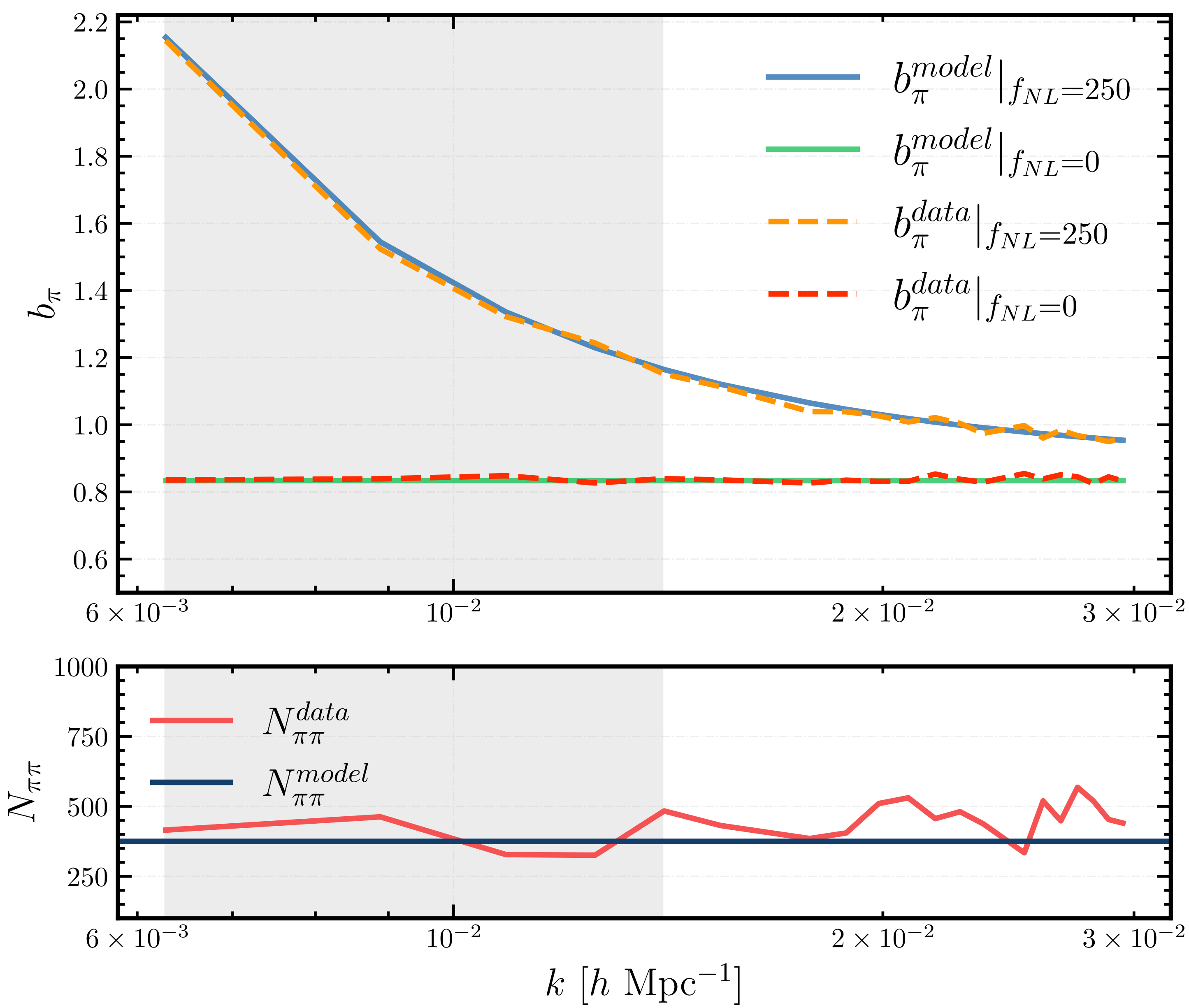
$$\delta_\pi(\mathbf{k}_L) = \left(\underbrace{b_\pi^G}_{\text{Gaussian bias}} + \underbrace{b_\pi^{NG} \frac{f_{NL}}{k_L^2}}_{\text{non-Gaussian bias}} \right) \delta_m(\mathbf{k}_L)$$

$$b_\pi^{NG} = \frac{\partial \bar{\pi}}{\partial \log \sigma_8}$$

- Examples of π(x) fields:
    - The **halo density** δ_h in some mass bin (traditional method).
    - The **locally measured (position-dependent) halo power spectrum.**
    - A **neural network trained to measure the local primordial power spectrum amplitude**.

The analytic model matches the NN output

"Halo bias" of the neural network output field π

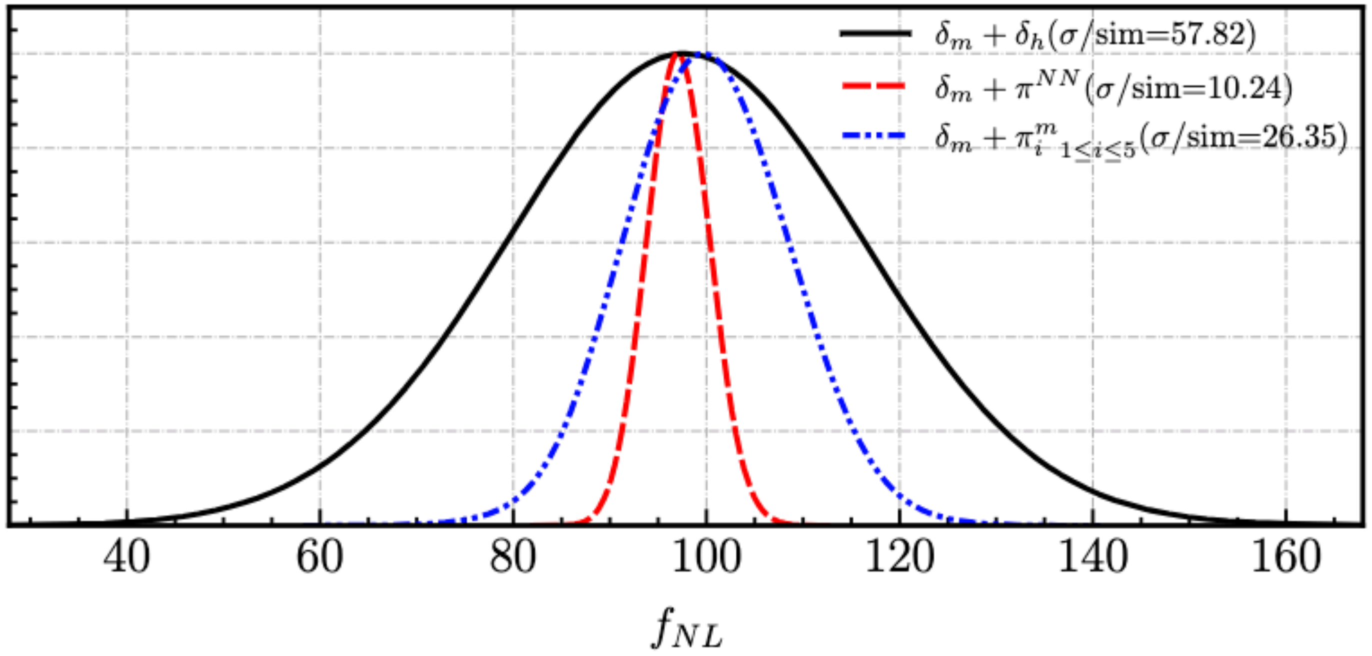


The signal power spectrum and noise power spectrum of the NN output behave exactly as we predict analytically.

We also have some a proof that our method is optimal under certain circumstances.

Result: Strong improvement in sensitivity to f_{NL}

Under simulation conditions, **the new method is several times more sensitive than the “old” analytic method.**



Black: analytic method without NN

Red: new NN method

Using Quijote-PNG simulations with f_{NL} = 100.

Caveat: In this analysis, the neural network gets to see the matter field, which is not directly observable.

Equivalence between f_{NL} and σ_8 measurement

Recall that non-Gaussian bias is defined as

$$b_\pi^{NG} = \frac{\partial \bar{\pi}}{\partial \log \sigma_8}$$

Thus any field π with $b_{ng} \neq 0$ will be sensitive to σ_8. We can construct a statistic that we will use to constrain σ_8 as follows

$$\bar{\pi} = \frac{1}{V} \int_x \pi(x)$$

The statistical error on σ_8 from this statistic can be calculated to be

$$\Delta\sigma_8 = \frac{\sigma_8}{V} \left(\frac{N_\pi^{1/2}}{b_\pi^{NG}} \right)$$

Thus **the Fisher information on σ_8 and on f_{NL} both scale in the same way:**

$$\Delta_{f_{NL}} \propto \Delta_{\sigma_8} \propto \frac{(N')^{1/2}}{b'_{ng}}$$

Optimality of the Loss function

The previous argument showed that **a field that is maximally good at measuring σ_8 will also be maximally good at measuring f_{NL}** through scale-dependent bias.

Thus we **can train a neural network to measure σ_8, on simulations without f_{NL}, and then use it to measure f_{NL}**.

An optimal loss function is given by

$$
\begin{aligned}
J &= \left\langle (\bar{\pi} - \sigma_8^{true})^2 \right\rangle_{simulations} \\
&= \left\langle \left(\frac{1}{V} \int \pi(\mathbf{x}) dV - \sigma_8^{true} \right)^2 \right\rangle_{simulations}
\end{aligned}
$$

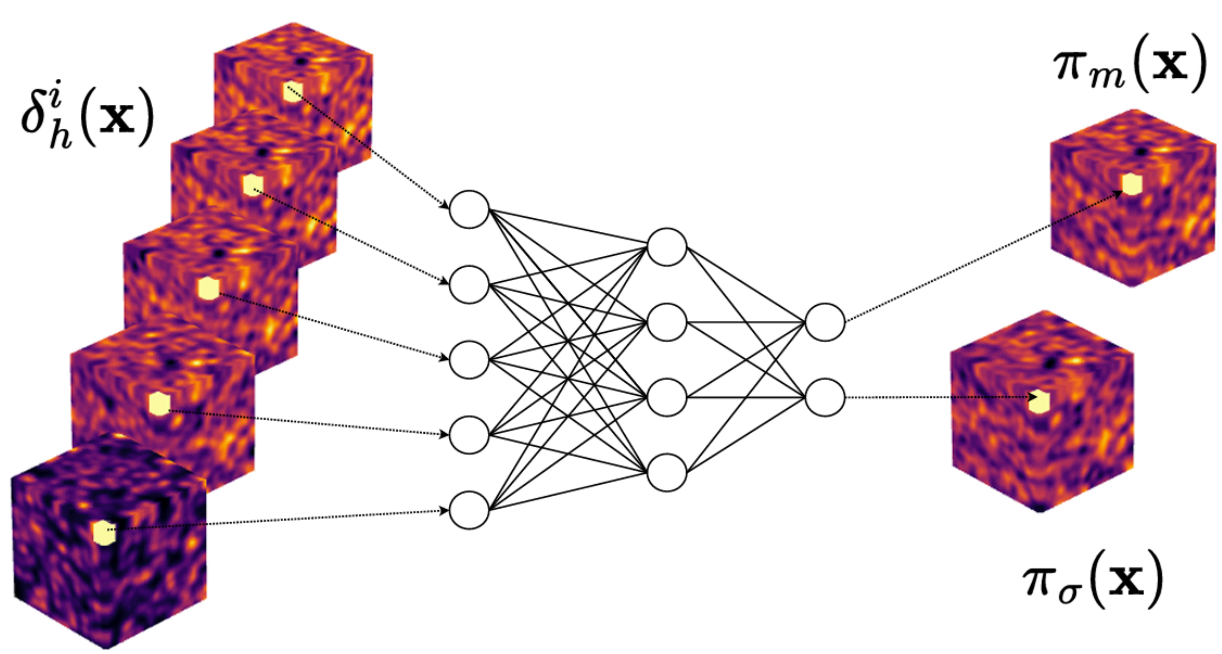
In the paper we have a more formal proof of optimality.

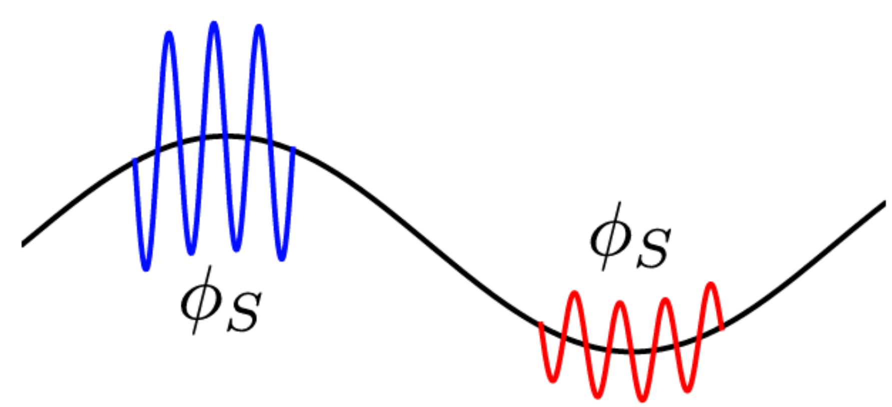
Halo analysis: Learning two optimal fields

If the matter field is not known (but rather we see only the halos), we can prove mathematically that it is possible to construct an optimal method by learning two fields:

1. a reconstruction of the local $\sigma_8^{\mathrm{loc}}(\mathbf{x})$ field (as in the last section)

2. a reconstruction of the matter field δ_m (i.e. learn to remove as much shot noise as possible)

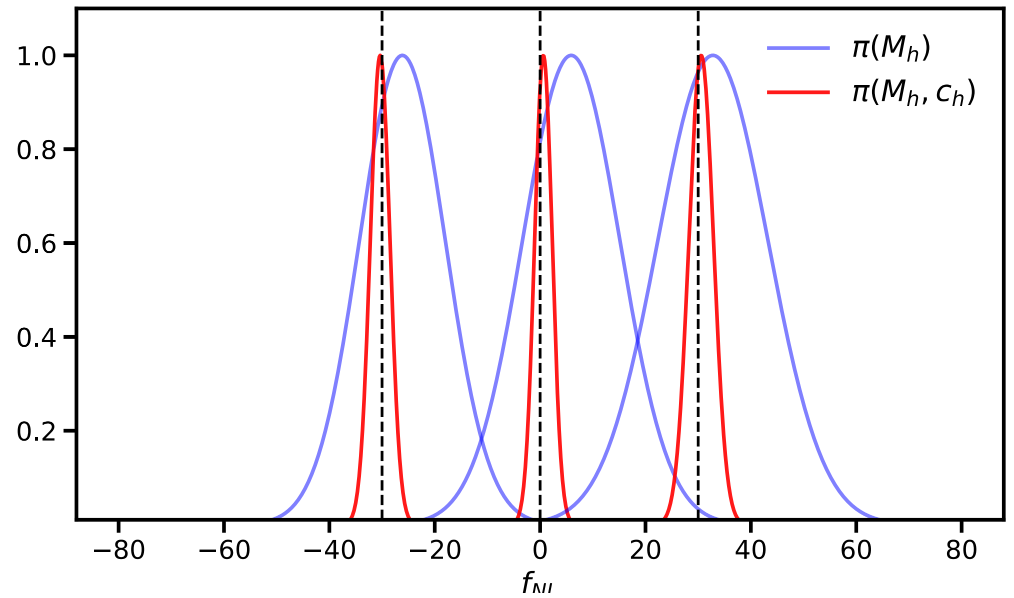


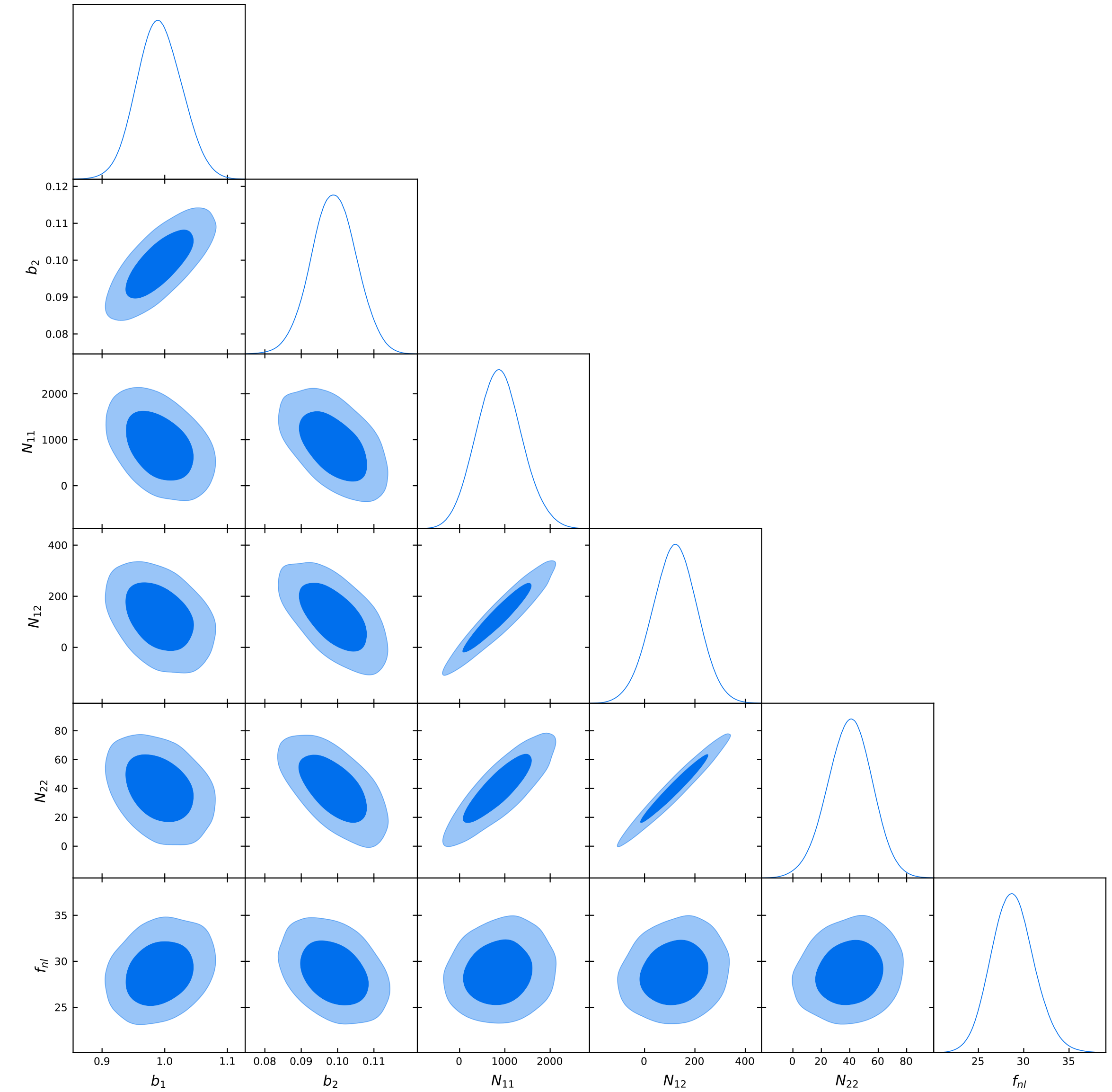


AbacusPNG with two CNN fields π_σ^{NN} and π_m^{NN}

- After training two CNNs, one to reconstruct the σ_8 field and one to reconstruct the δ_m field, we then **apply these networks to AbacusPNG test data**.
- We run the same MCMC analysis as for ordinary scale-dependent bias, but now using the two novel fields.
- We find **strong unbiased constraints on f_{NL} (in particular when adding halo concentrations)** and unbiased estimates.

	M		M,c	
f_{NL}	$\hat{f}_{NL}$	$\sigma_{\hat{f}_{NL}}$	$\hat{f}_{NL}$	$\sigma_{\hat{f}_{NL}}$
-30	-26.1	8.1	-30.4	1.9
0	5.9	9.2	0.6	1.7
+30	32.8	10.4	30.6	2.3

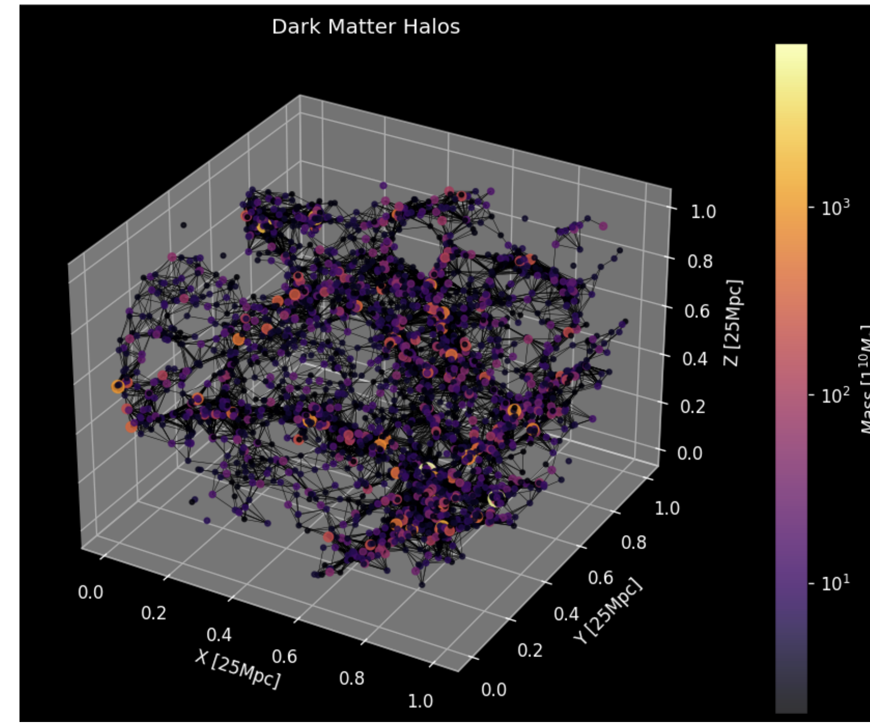




Upcoming: From CNN to GNN

- **Individual halo positions (and other individual properties) should contain extra information on σ_8.**
- Recent work (e.g. 2204.13713) has shown that **Graph Neural Networks** are very suitable to this task.
- We compared on **CAMELS whether the GNN beats halo counting for σ_8.**
- **Recall the** proportionality:

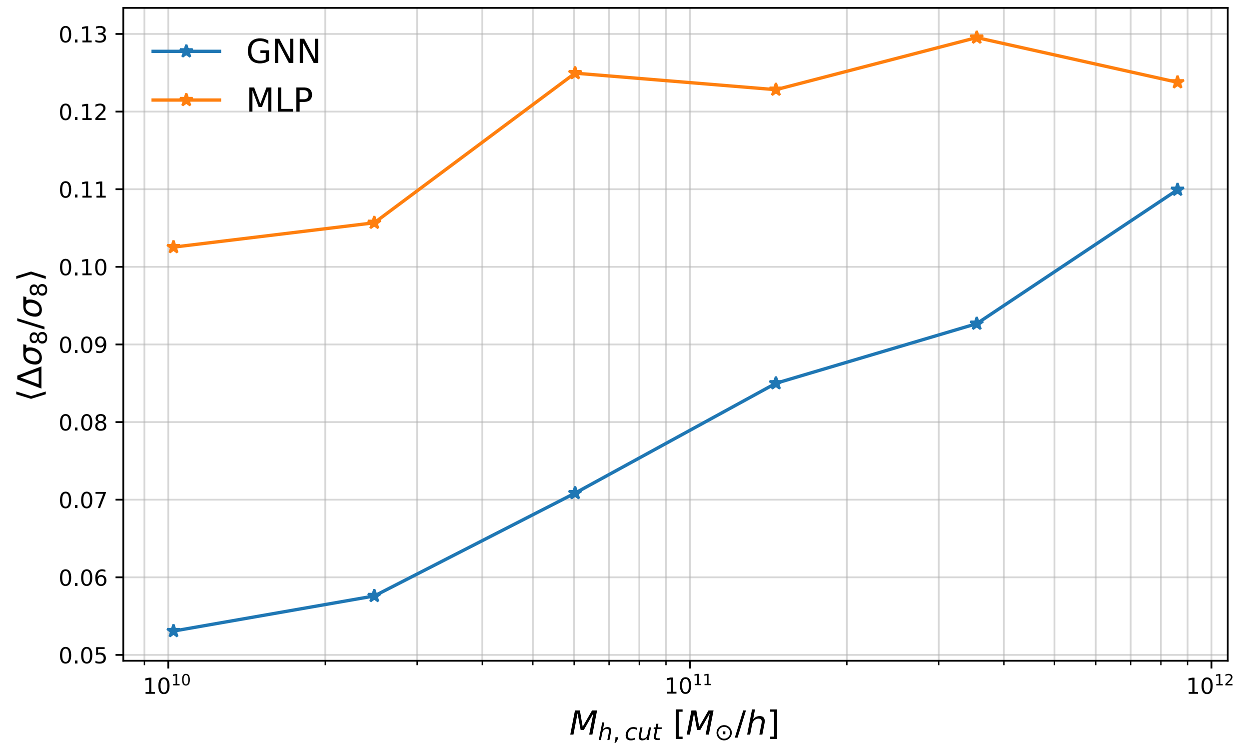
$$\Delta_{f_{NL}} \propto \Delta_{\sigma_8}$$



Graph example for (25MPc)^3
CAMELS Illustris LH hydro sims suite

Measuring σ_8 on CAMELS

- **Do halo positions add information on σ_8 over halo masses? YES.**
- For CAMELS, we find that the GNN (input: individual masses+positions) does outperform an MLP/CNN (input: local halo mass function) in predicting σ_8.



CAMELS Illustris Hydro. Using sub-halos as a proxy for galaxies.

So can we tighten f_{NL} significantly in practice?

- Even though our method has convenient properties, **this question is STILL hard to answer**.
  - On CAMELS, it appears that our method works but **CAMELS does not have large halos**, which contribute a lot of signal in larger volumes.
  - Results for halo or galaxy-wise measurements also depend on **what halo/galaxy properties can be measured** how well. E.g. halo concentration.
  - On **Abacus halos**, we are somewhat limited by the lack of continuous training data. Leads to convergence problems with the GNN.
- **Upcoming work**:
  - More detailed analysis for more realistic survey.
  - **Extend our method to weak lensing.**

Aside: Reconstruct the local electron and dark matter density

2411.02496 Reconstruction of Continuous Cosmological Fields from Discrete Tracers with Graph Neural Networks. And upcoming followup.

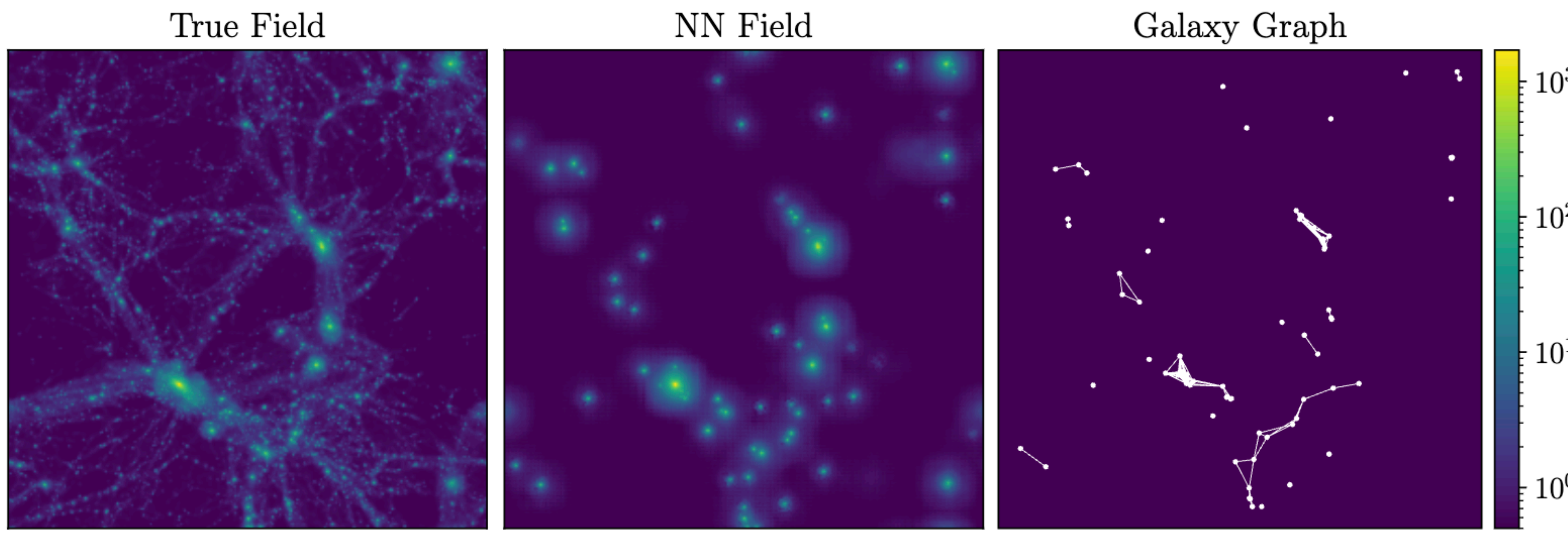


Figure 6: True (left) and predicted with GNN-CNN setup (middle) dark matter density in a $5 \times 25 \times 25$ $(\mathtt{Mpc/h})^3$ volume, averaged over the x axis. The rightmost column shows the visualization of the input - galaxy cloud in the same region as a graph

- If we use the reconstructed electron template for the quadratic estimator in kSZ velocity reconstruction, then this is **also a "squeezed limit machine learning method."**
    - Again, we can marginalize on large scales about a bias factor that takes into account simulation uncertainty.

Thanks!