Cosmological constraints from the HSC survey's 1st yr data with deep learning

Lu, Haiman and Li 2023, MNRAS, 521, 2050

Tianhuan (Brian) Lu   Zoltán Haiman

Columbia University

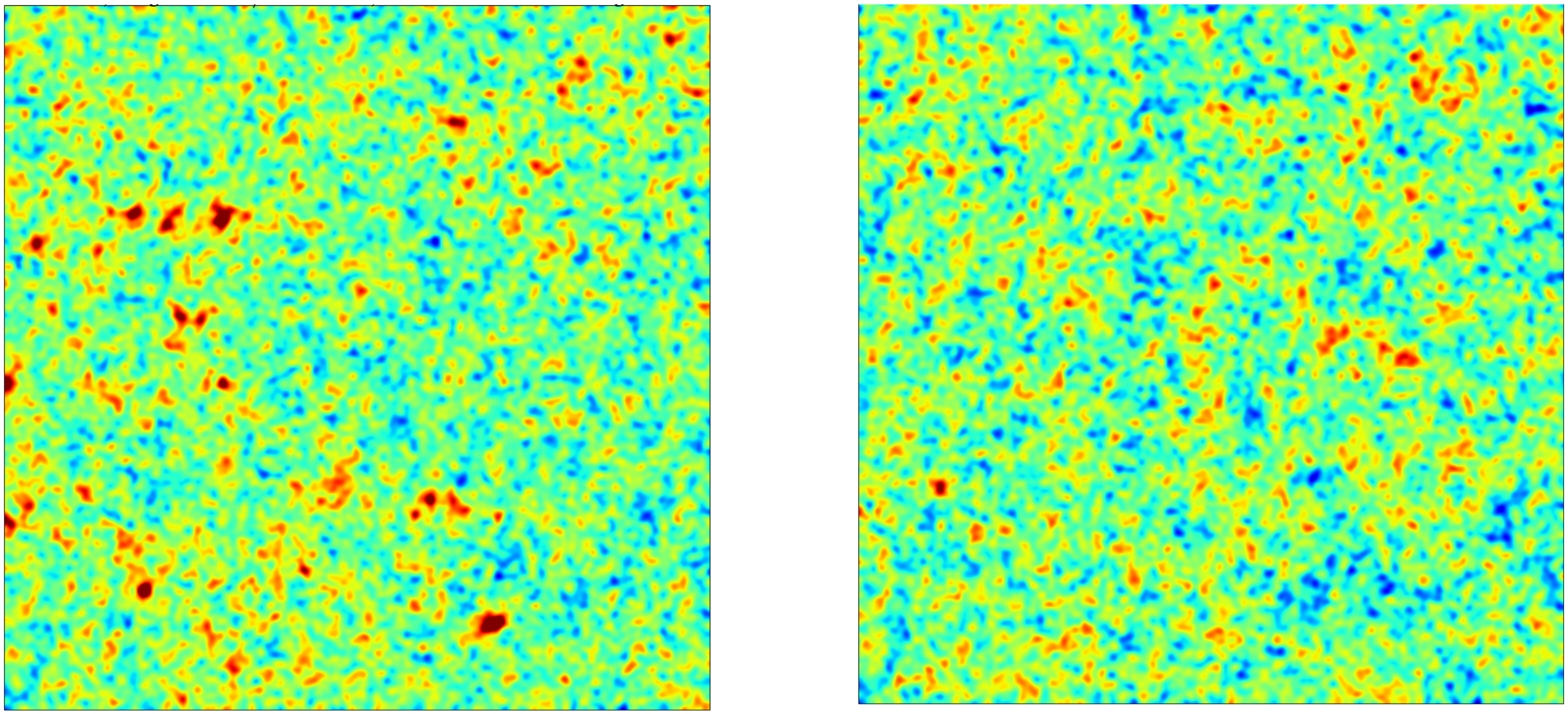
questions:

- How much cosmological information is contained, in principle, in a (perfect) weak lensing map?
- How well can we constrain background cosmology, in practice, from observed lensing data?

Cosmic Shear is Not Gaussian

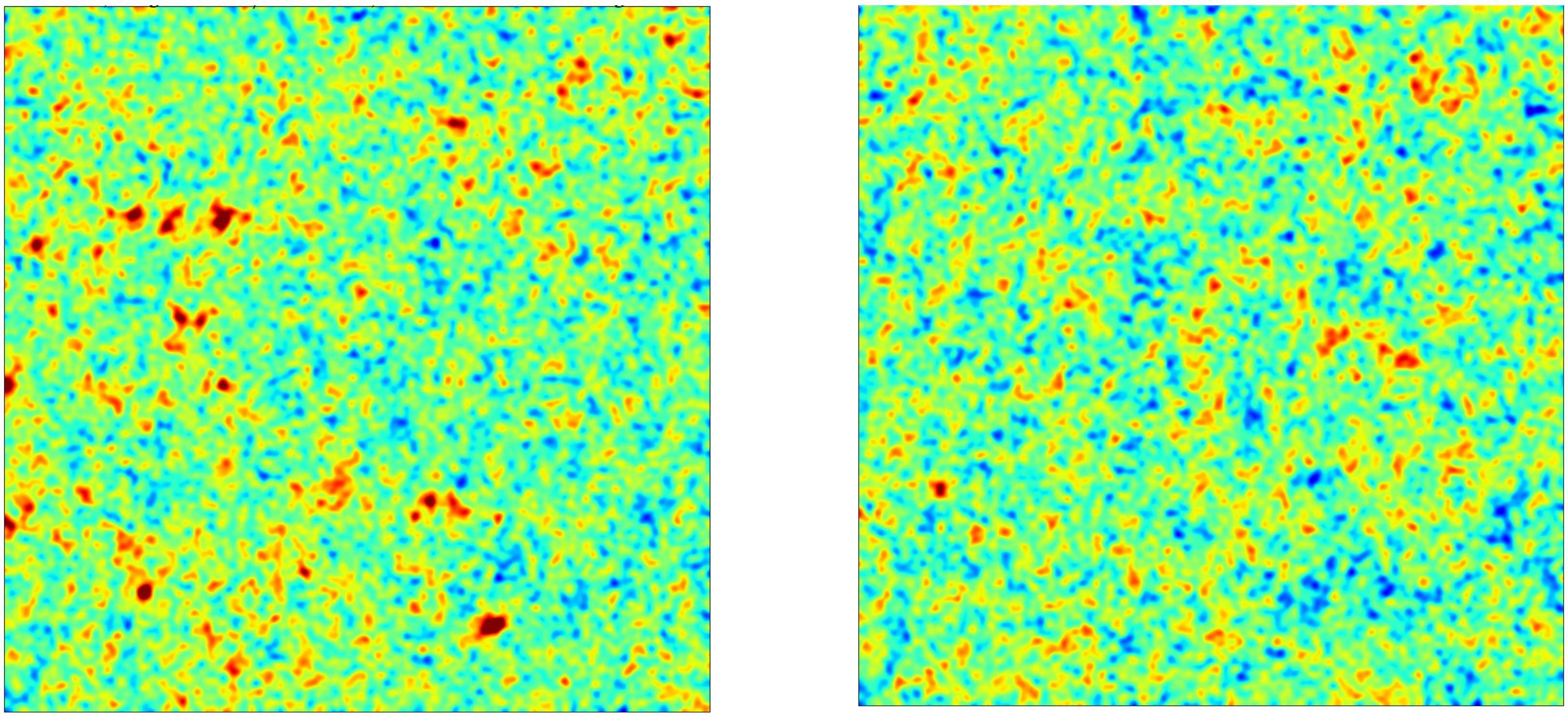
Millennium simulation – Volker Springel, MPA

Cosmic Shear is Not Gaussian



Millennium simulation – Volker Springel, MPA

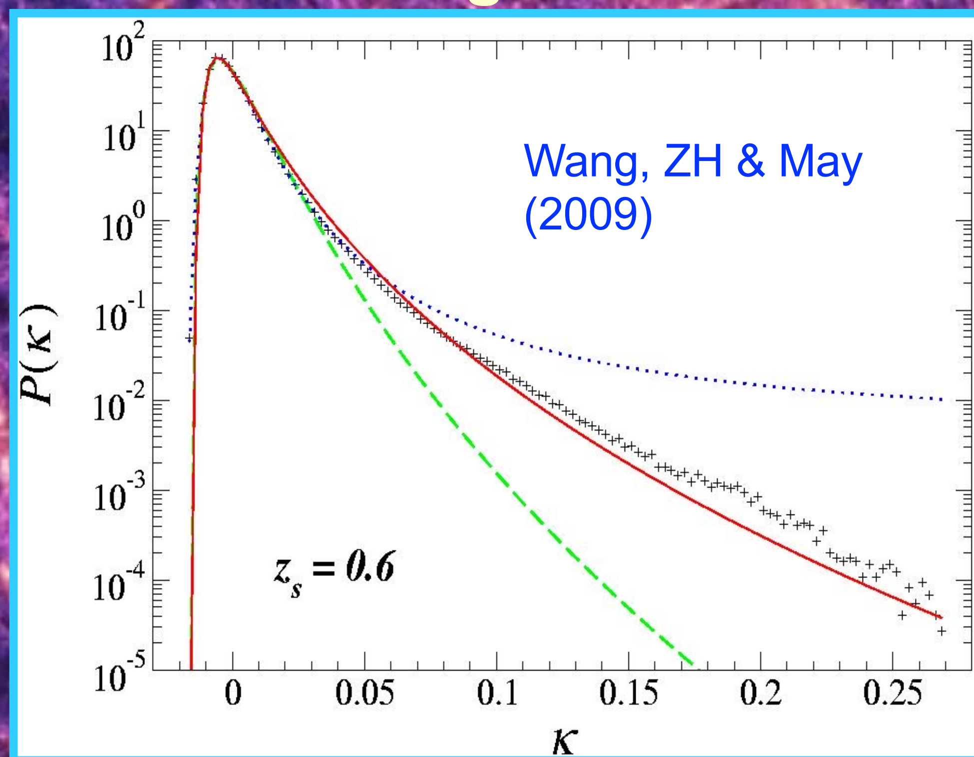
Cosmic Shear is Not Gaussian



Millennium simulation – Volker Springel, MPA

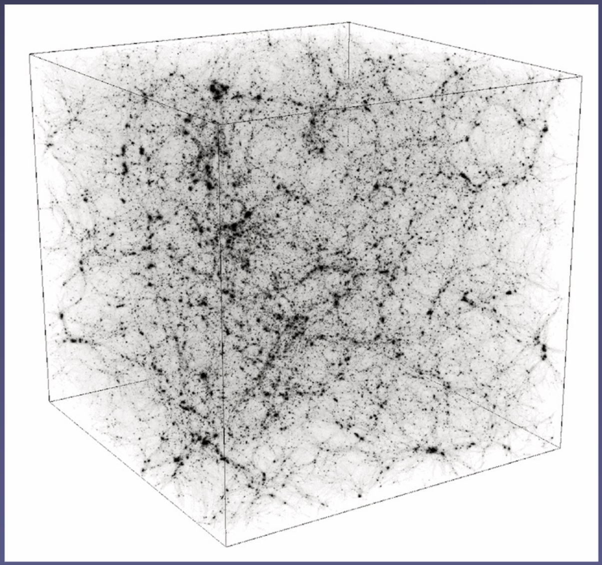
Cosmic Shear is Not Gaussian

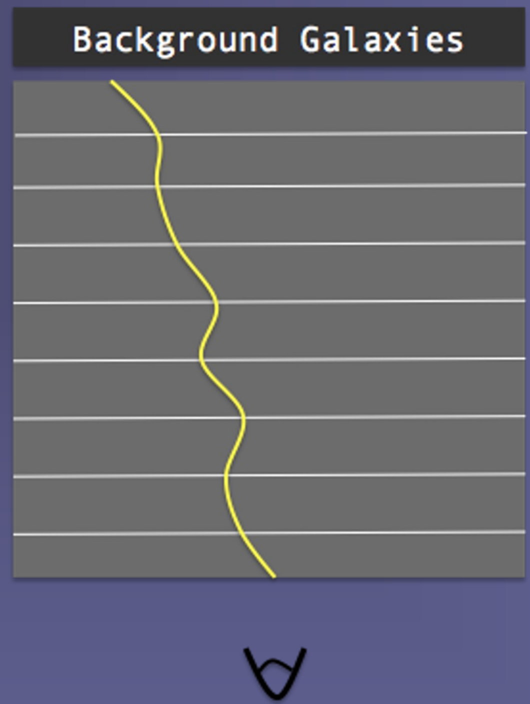
PDF of convergence:

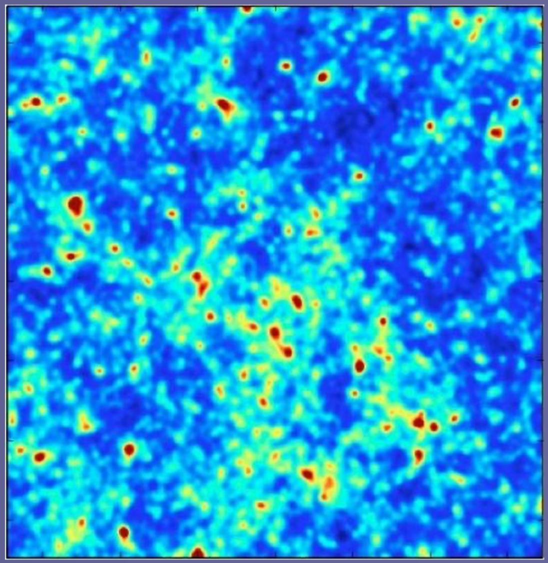


Millennium simulation – Volker Springel, MPA

Raytracing Simulation Pipeline







(1) N-body sims

- Gadget-2
- 512^3 (240 Mpc)^3
- ~100 $[\Omega_m, \sigma_8]$'s

(2) Ray-tracing

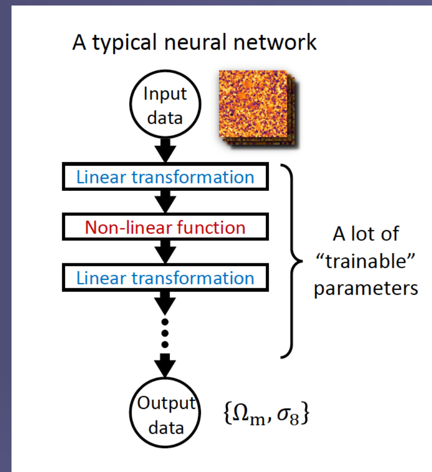
- 2048x2048
- 3x3 deg^2
- z_s=1, 1.5, 2

(3) Convergence maps

- smooth ~amin
- shape noise
- 1000/model

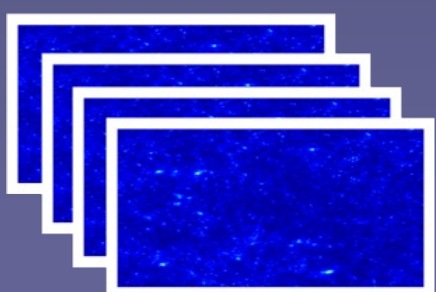
LensTools public on github

Deep convolutional neural network



Training

Training set of maps (50-70%)



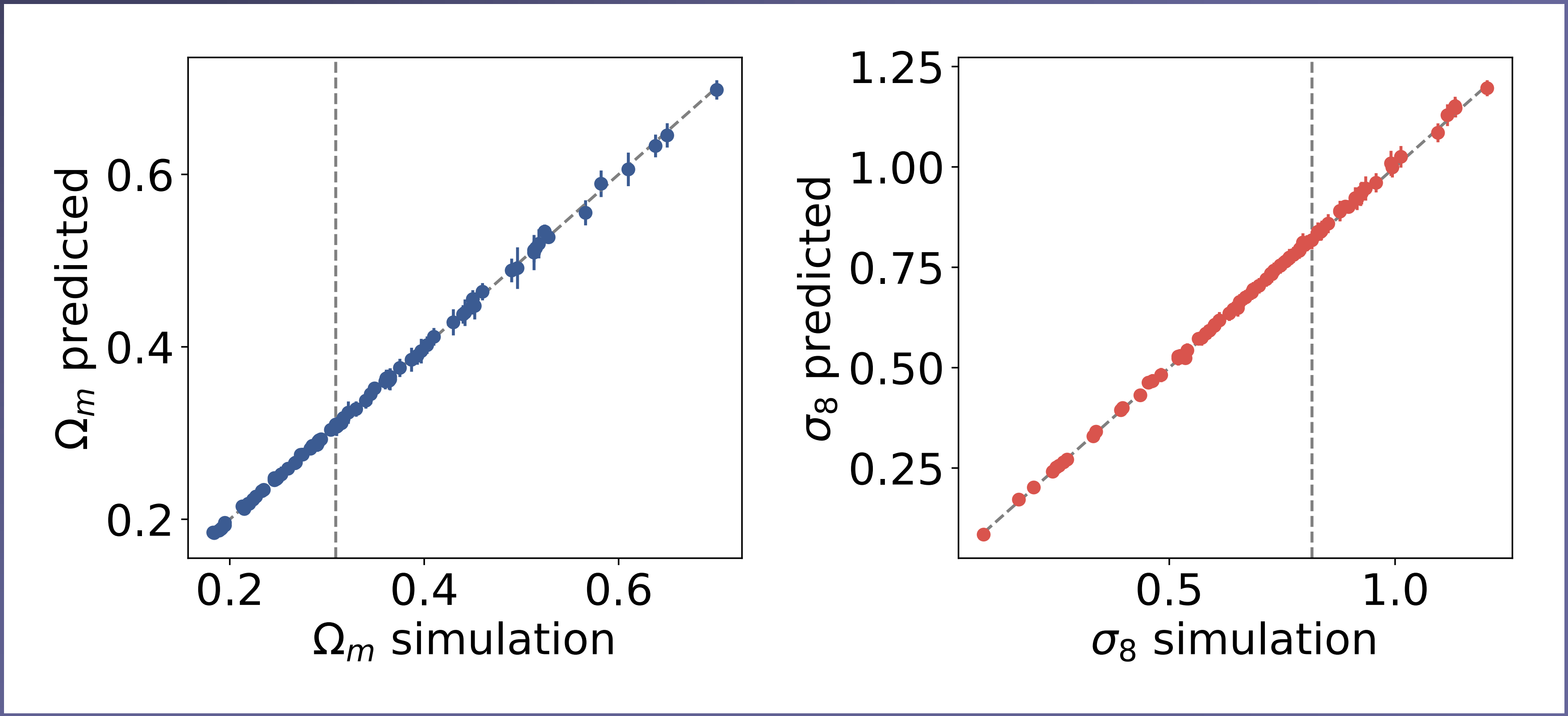
CNN (black box)

Loss function (to be minimized)

$$\sum_{\text{map}\in\text{batch}} W_{\text{cosmo}}\left(|\sigma_8^{\text{pred}} - \sigma_8^{\text{true}}| + |\Omega_m^{\text{pred}} - \Omega_m^{\text{true}}|\right).$$

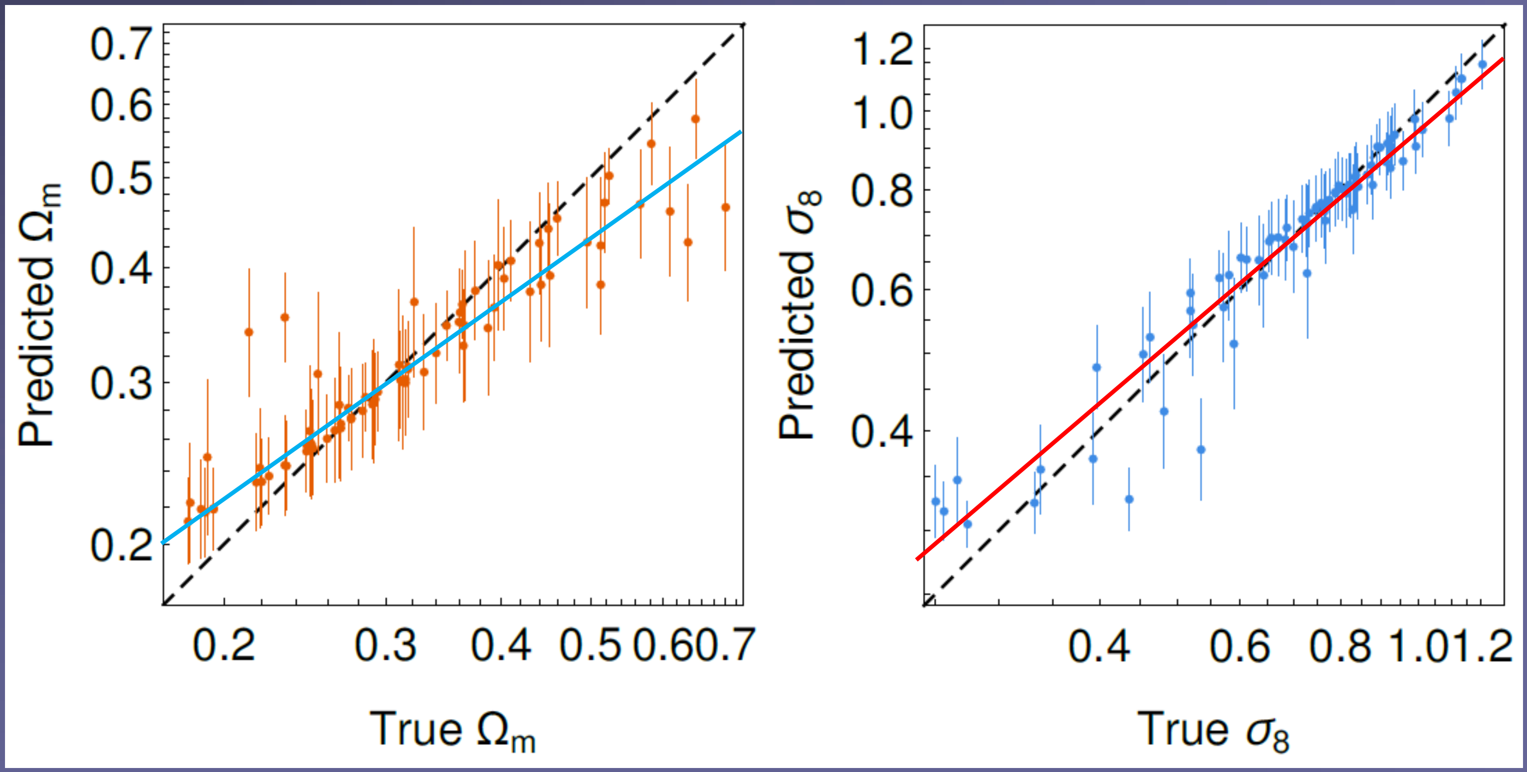
Update weights
(e.g. Adams optimizer,
stochastic gradient descent)

Parameter estimates from CNN



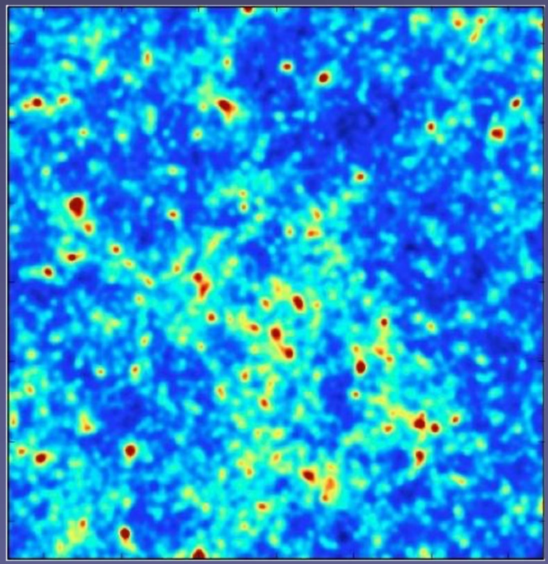
raw noise-free maps

Parameter estimates from CNN



maps with noise + 2 arcmin smoothing

Parameter inference



Power spectrum → $C_{\ell_1}, C_{\ell_2}, C_{\ell_3}, \cdots, C_{\ell_n}$

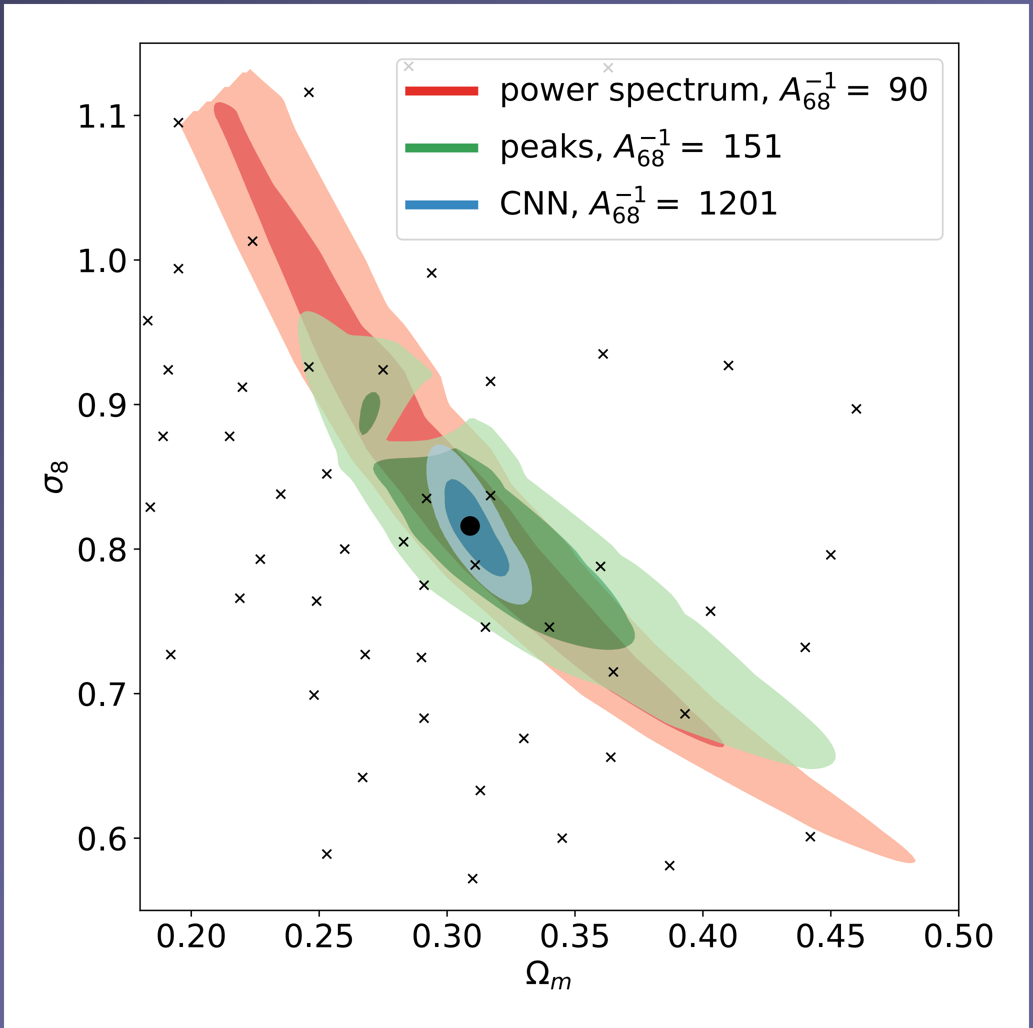
Deep learning → ~~Ω_m~~, ~~σ_8~~, …

y_1 y_2

treat CNN output y_1 the same way as C_{ℓ_1}

→ define likelihood, obtain posteriors …

Noiseless maps



Gupta+2018
Ribli+2019

Constraints improve by factor of 13(!)
[passes Gaussian-field null-test]

Unsolved problems

- How much cosmological information is contained, in principle, in a (perfect) weak lensing map?

- How well can we constrain background cosmology, in practice, from observed lensing data?

Unsolved problems

- How much cosmological information is contained, in principle, in a (perfect) lensing map?

→ **At least an order of magnitude more than in power spectrum**

- How well can we constrain background cosmology, in practice, from observed lensing data?

Unsolved problems

- How much cosmological information is contained, in principle, in a (perfect) lensing map?

→ **At least an order of magnitude more than in power spectrum**

- How well can we constrain background cosmology, in practice, from observed lensing data?

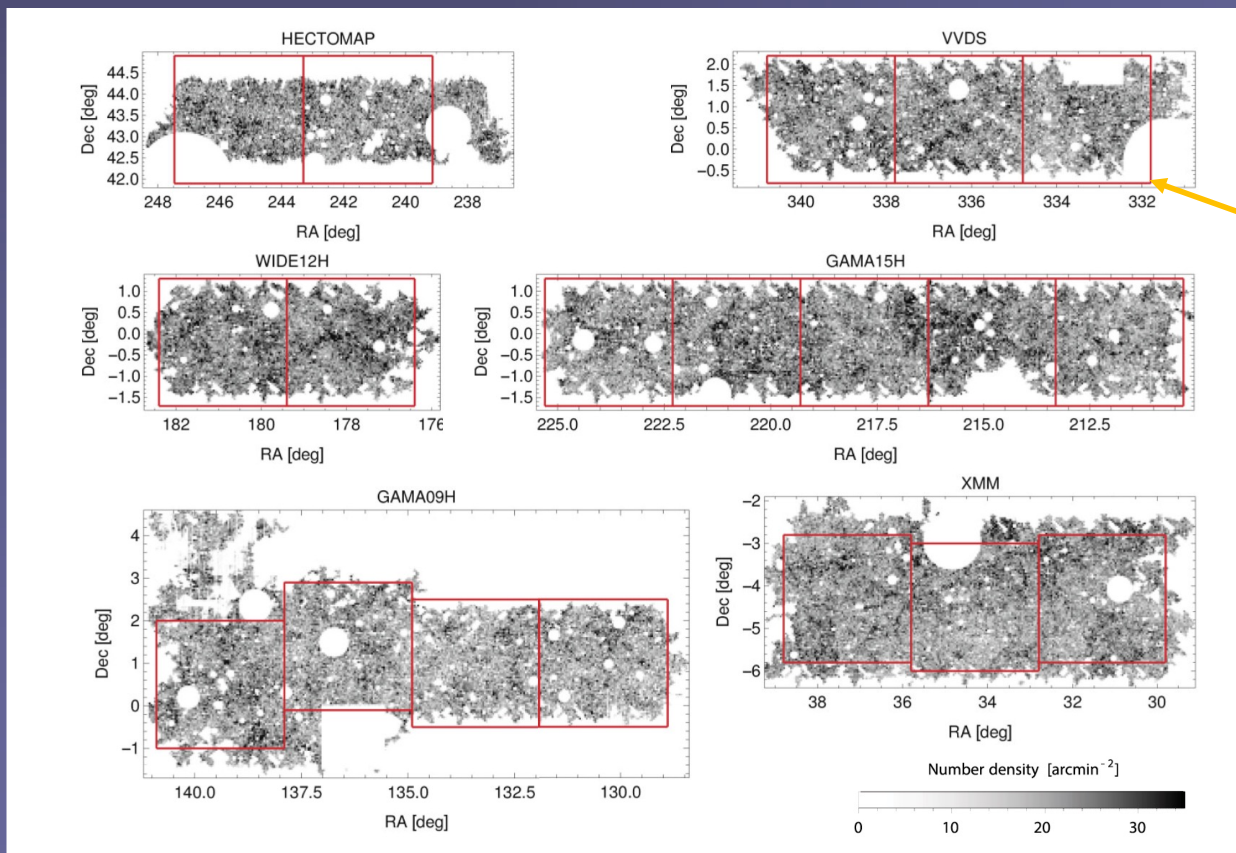
Surveys

Survey	**Area** (deg^2)	**Depth** (gal/arcmin^2)	**Current State**
CFHTLenS	150	14	Completed
KiDS	1350	9	Ongoing
DES	5000	6	Ongoing
HSC	1500	17	Ongoing
Euclid Space Mission	15000	~30	Starting
LSST	18000	~30	Starting
Roman Space Telescope	2200	~50	Planned

approach: forward-model survey, train CNN on sims, apply to data

HSC Yr-1 catalogs (public)

- 19 3×3 deg^2 subfields cropped and used for forward modeling
- four redshift bins ($0.3 < z < 1.5$): 11.9 → 8.5 million galaxies
- catalog has sky position, redshift, shear, weight for each galaxy

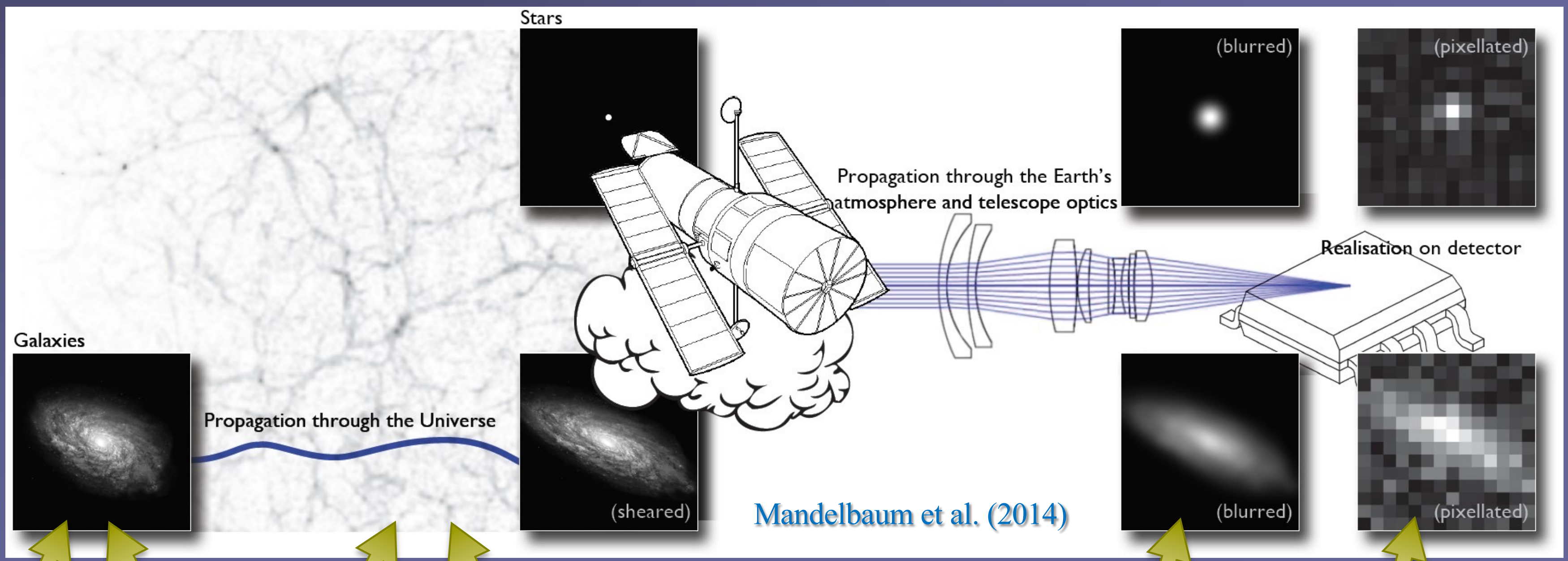


each subfield is forward-modeled

Subaru telescope

Forward-model suite with systematics

- *79 cosmologies*: 1600 realizations each in 2D (Ω_m, σ_8)
- *baryons*: Sobol sequence in 4D space, 1600 combos (2.5×10^6 maps total)
- *CNNs*: 50% training, 50% inference



Shape noise

Photo-*z*

Baryonic effects

+4 parameters

Intrinsic alignments

+1 parameter

PSF correction

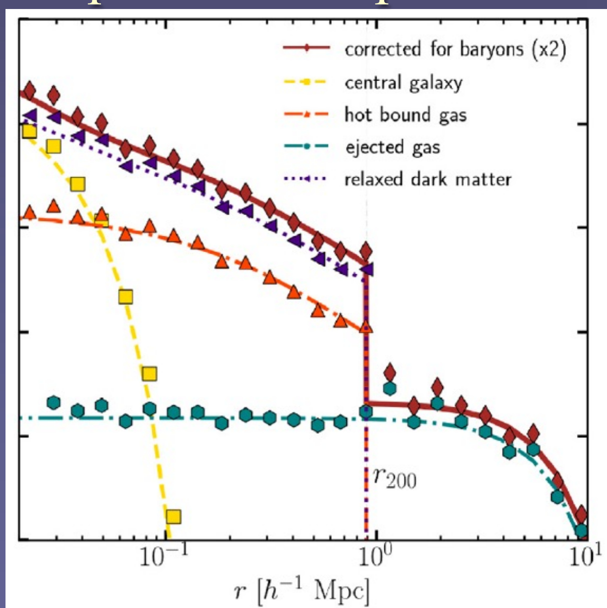
Shape measurement

Baryons

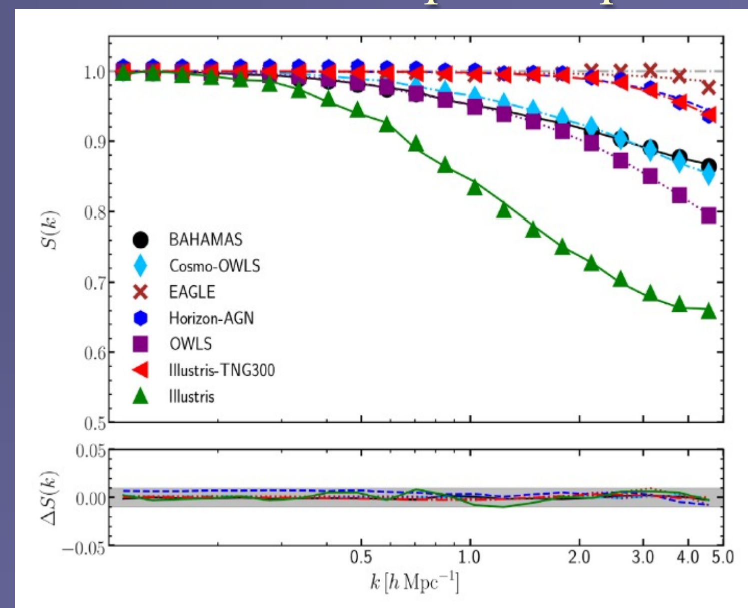
- Halo profiles based on hydro sims (painted on lens planes)
- Baryon correction models (BCM)

Schneider & Teyssier (2015)
Arico+ (2020, 2021)

Impact on halo profile



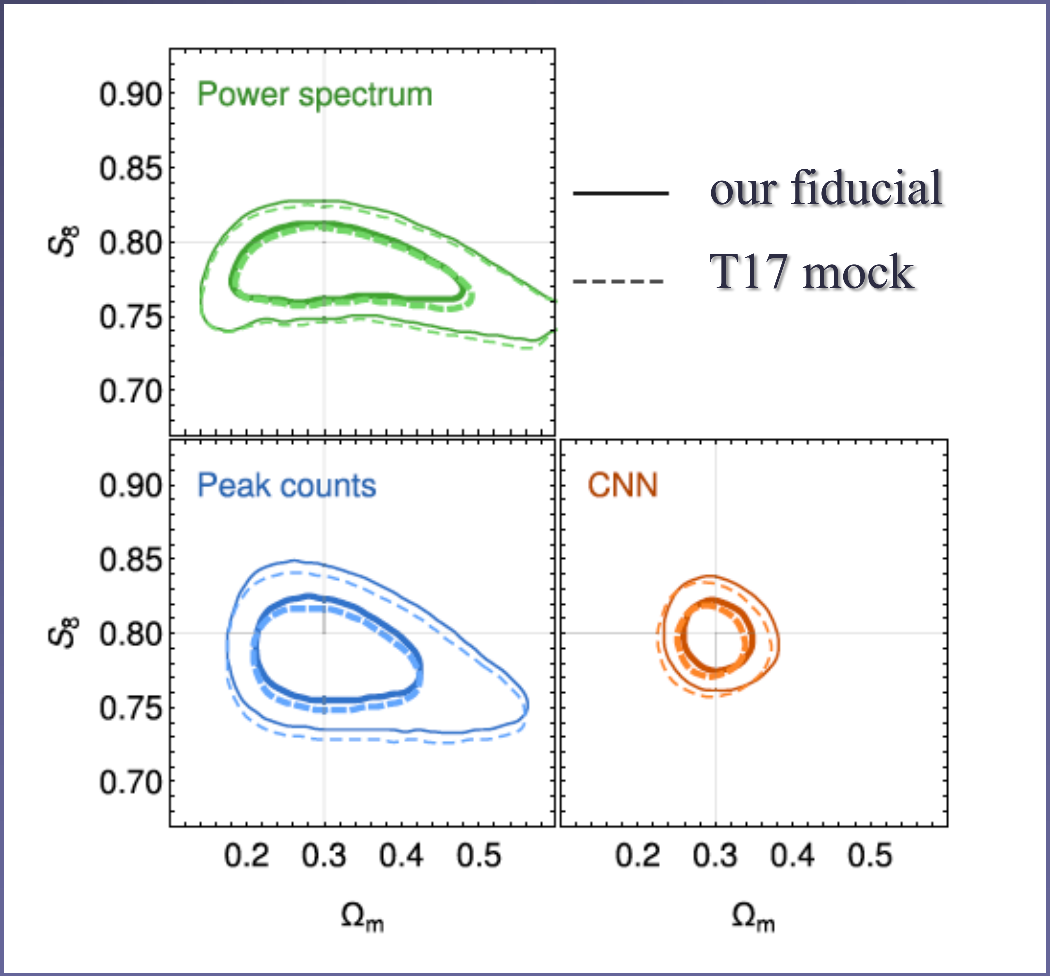
Can fit 3d matter power spectra



Parameter	Description	Fiducial Value ($z = 0$)
M_c	Halo mass scale for retaining half of the total gas	$3.3 \times 10^{13}\ h^{-1}\ \mathrm{M}_\odot$
M_1	Characteristic halo mass for a galaxy mass fraction $\epsilon = 0.023$	$8.63 \times 10^{11}\ h^{-1}\ \mathrm{M}_\odot$
η	Maximum distance of gas ejection in terms of the halo escape radius	0.54
β	Slope of the gas fraction as a function of halo mass	0.12

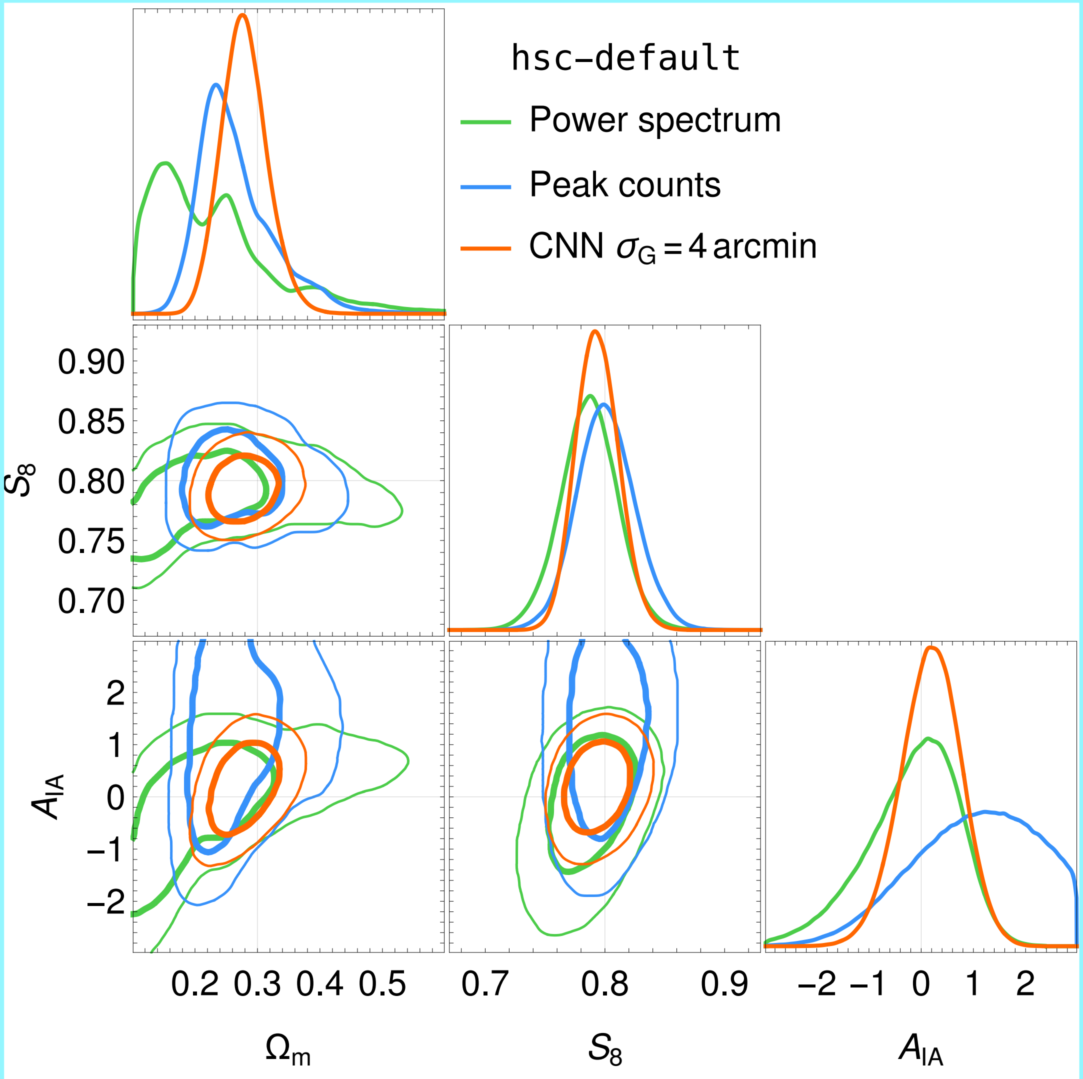
Validation

apply inference to Takahashi et al. (2017) mock catalog



Results – no baryons

Lu, ZH & Li (2023)

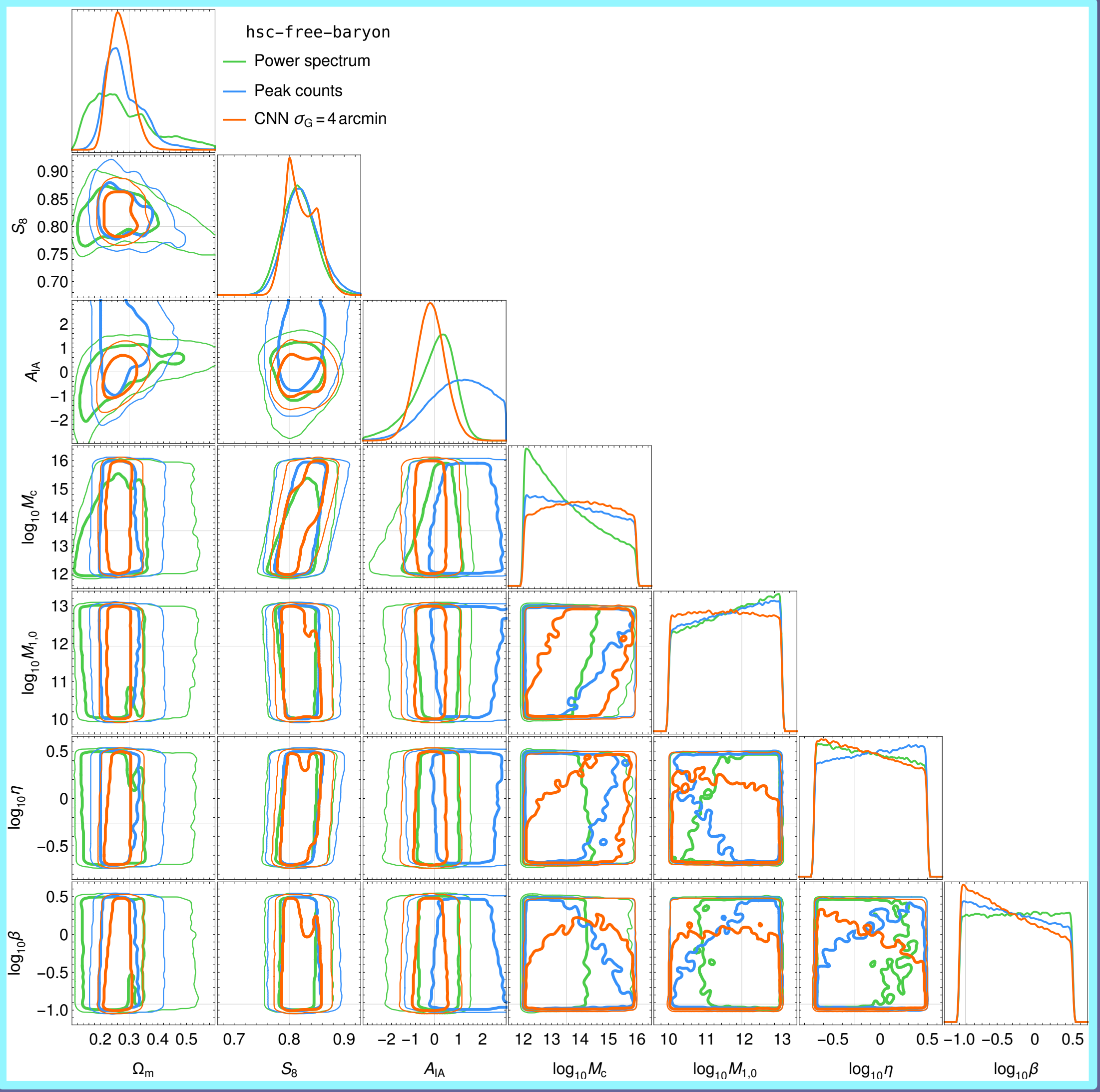


CNNs tighten Ω_m, S_8, A_{IA} constraints by up to a factor of 2.5

similar results if baryon parameters fixed at fiducial values

Results - free baryons

Lu, ZH & Li (2023)



- baryons increase Ω_m,S_8 uncertainty
- CNNs improve Ω_m but S_8 constraint only marginally better
- baryon parameters unconstrained (external priors will help)

Results - summary

Lu, ZH & Li (2023)

Model	Summary statistic	Ω_m	S_8	A_{IA}
Default (no baryon)	Power spectrum	$0.211^{+0.105}_{-0.072}$ (1.00)	$0.787^{+0.023}_{-0.023}$ (1.00)	$-0.07^{+0.75}_{-0.99}$ (1.00)
	Peak counts	$0.254^{+0.072}_{-0.042}$ (0.64)	$0.800^{+0.025}_{-0.023}$ (1.03)	$1.10^{+1.15}_{-1.31}$ (1.41)
	CNN	$0.278^{+0.037}_{-0.035}$ (0.41)	$0.793^{+0.018}_{-0.017}$ (0.76)	$0.20^{+0.55}_{-0.58}$ (0.65)
No IA ($A_{IA} = 0$)	Power spectrum	$0.235^{+0.111}_{-0.063}$ (1.00)	$0.794^{+0.022}_{-0.022}$ (1.00)	–
	Peak counts	$0.218^{+0.049}_{-0.039}$ (0.51)	$0.803^{+0.025}_{-0.024}$ (1.12)	–
	CNN	$0.281^{+0.031}_{-0.029}$ (0.35)	$0.793^{+0.018}_{-0.017}$ (0.80)	–
Fiducial baryonic model	Power spectrum	$0.290^{+0.115}_{-0.087}$ (1.00)	$0.820^{+0.025}_{-0.026}$ (1.00)	$0.18^{+0.61}_{-0.82}$ (1.00)
	Peak counts	$0.278^{+0.063}_{-0.051}$ (0.57)	$0.819^{+0.029}_{-0.026}$ (1.08)	$1.11^{+1.14}_{-1.28}$ (1.70)
	CNN	$0.250^{+0.036}_{-0.031}$ (0.33)	$0.826^{+0.020}_{-0.019}$ (0.77)	$-0.04^{+0.58}_{-0.61}$ (0.83)
Free baryonic model	Power spectrum	$0.259^{+0.139}_{-0.083}$ (1.00)	$0.816^{+0.031}_{-0.030}$ (1.00)	$0.08^{+0.68}_{-0.98}$ (1.00)
	Peak counts	$0.268^{+0.082}_{-0.047}$ (0.58)	$0.822^{+0.034}_{-0.029}$ (1.04)	$1.07^{+1.19}_{-1.32}$ (1.51)
	CNN	$0.268^{+0.040}_{-0.036}$ (0.35)	$0.819^{+0.034}_{-0.024}$ (0.95)	$-0.16^{+0.59}_{-0.58}$ (0.71)

Results - summary

Lu, ZH & Li (2023)

Model	Summary statistic	Ω_m	S_8	A_{IA}
Default (no baryon)	Power spectrum	$0.211^{+0.105}_{-0.072}$ (1.00)	$0.787^{+0.023}_{-0.023}$ (1.00)	$-0.07^{+0.75}_{-0.99}$ (1.00)
	Peak counts	$0.254^{+0.072}_{-0.042}$ (0.64)	$0.800^{+0.025}_{-0.023}$ (1.03)	$1.10^{+1.15}_{-1.31}$ (1.41)
	CNN	$0.278^{+0.037}_{-0.035}$ (0.41)	$0.793^{+0.018}_{-0.017}$ (0.76)	$0.20^{+0.55}_{-0.58}$ (0.65)
No IA ($A_{IA} = 0$)	Power spectrum	$0.235^{+0.111}_{-0.063}$ (1.00)	$0.794^{+0.022}_{-0.022}$ (1.00)	–
	Peak counts	$0.218^{+0.049}_{-0.039}$ (0.51)	$0.803^{+0.025}_{-0.024}$ (1.12)	–
	CNN	$0.281^{+0.031}_{-0.029}$ (0.35)	$0.793^{+0.018}_{-0.017}$ (0.80)	–
Fiducial baryonic model	Power spectrum	$0.290^{+0.115}_{-0.087}$ (1.00)	$0.820^{+0.025}_{-0.026}$ (1.00)	$0.18^{+0.61}_{-0.82}$ (1.00)
	Peak counts	$0.278^{+0.063}_{-0.051}$ (0.57)	$0.819^{+0.029}_{-0.026}$ (1.08)	$1.11^{+1.14}_{-1.28}$ (1.70)
	CNN	$0.250^{+0.036}_{-0.031}$ (0.33)	$0.826^{+0.020}_{-0.019}$ (0.77)	$-0.04^{+0.58}_{-0.61}$ (0.83)
Free baryonic model	Power spectrum	$0.259^{+0.139}_{-0.083}$ (1.00)	$0.816^{+0.031}_{-0.030}$ (1.00)	$0.08^{+0.68}_{-0.98}$ (1.00)
	Peak counts	$0.268^{+0.082}_{-0.047}$ (0.58)	$0.822^{+0.034}_{-0.029}$ (1.04)	$1.07^{+1.19}_{-1.32}$ (1.51)
	CNN	$0.268^{+0.040}_{-0.036}$ (0.35)	$0.819^{+0.034}_{-0.024}$ (0.95)	$-0.16^{+0.59}_{-0.58}$ (0.71)

- Extra info found by CNN: factor of ~3 improvement for Ω_m

  mostly from degeneracy breaking (S_8 not improved)

Results - summary

Lu, ZH & Li (2023)

Model	Summary statistic	Ω_m	S_8	A_{IA}
Default (no baryon)	Power spectrum	$0.211^{+0.105}_{-0.072}$ (1.00)	$0.787^{+0.023}_{-0.023}$ (1.00)	$-0.07^{+0.75}_{-0.99}$ (1.00)
	Peak counts	$0.254^{+0.072}_{-0.042}$ (0.64)	$0.800^{+0.025}_{-0.023}$ (1.03)	$1.10^{+1.15}_{-1.31}$ (1.41)
	CNN	$0.278^{+0.037}_{-0.035}$ (0.41)	$0.793^{+0.018}_{-0.017}$ (0.76)	$0.20^{+0.55}_{-0.58}$ (0.65)
No IA ($A_{IA}=0$)	Power spectrum	$0.235^{+0.111}_{-0.063}$ (1.00)	$0.794^{+0.022}_{-0.022}$ (1.00)	–
	Peak counts	$0.218^{+0.049}_{-0.039}$ (0.51)	$0.803^{+0.025}_{-0.024}$ (1.12)	–
	CNN	$0.281^{+0.031}_{-0.029}$ (0.35)	$0.793^{+0.018}_{-0.017}$ (0.80)	–
Fiducial baryonic model	Power spectrum	$0.290^{+0.115}_{-0.087}$ (1.00)	$0.820^{+0.025}_{-0.026}$ (1.00)	$0.18^{+0.61}_{-0.82}$ (1.00)
	Peak counts	$0.278^{+0.063}_{-0.051}$ (0.57)	$0.819^{+0.029}_{-0.026}$ (1.08)	$1.11^{+1.14}_{-1.28}$ (1.70)
	CNN	$0.250^{+0.036}_{-0.031}$ (0.33)	$0.826^{+0.020}_{-0.019}$ (0.77)	$-0.04^{+0.58}_{-0.61}$ (0.83)
Free baryonic model	Power spectrum	$0.259^{+0.139}_{-0.083}$ (1.00)	$0.816^{+0.031}_{-0.030}$ (1.00)	$0.08^{+0.68}_{-0.98}$ (1.00)
	Peak counts	$0.268^{+0.082}_{-0.047}$ (0.58)	$0.822^{+0.034}_{-0.029}$ (1.04)	$1.07^{+1.19}_{-1.32}$ (1.51)
	CNN	$0.268^{+0.040}_{-0.036}$ (0.35)	$0.819^{+0.034}_{-0.024}$ (0.95)	$-0.16^{+0.59}_{-0.58}$ (0.71)

- Extra info found by CNN: factor of ~3 improvement for Ω_m mostly from degeneracy breaking (S_8 not improved)
- S_8 tension: reduced from 2.2σ to 0.4σ by baryons (vs. Planck)

Results - summary

Lu, ZH & Li (2023)

Model	Summary statistic	Ω_m	S_8	A_{IA}
Default (no baryon)	Power spectrum	$0.211^{+0.105}_{-0.072}$ (1.00)	$0.787^{+0.023}_{-0.023}$ (1.00)	$-0.07^{+0.75}_{-0.99}$ (1.00)
	Peak counts	$0.254^{+0.072}_{-0.042}$ (0.64)	$0.800^{+0.025}_{-0.023}$ (1.03)	$1.10^{+1.15}_{-1.31}$ (1.41)
	CNN	$0.278^{+0.037}_{-0.035}$ (0.41)	$0.793^{+0.018}_{-0.017}$ (0.76)	$0.20^{+0.55}_{-0.58}$ (0.65)
No IA ($A_{IA} = 0$)	Power spectrum	$0.235^{+0.111}_{-0.063}$ (1.00)	$0.794^{+0.022}_{-0.022}$ (1.00)	–
	Peak counts	$0.218^{+0.049}_{-0.039}$ (0.51)	$0.803^{+0.025}_{-0.024}$ (1.12)	–
	CNN	$0.281^{+0.031}_{-0.029}$ (0.35)	$0.793^{+0.018}_{-0.017}$ (0.80)	–
Fiducial baryonic model	Power spectrum	$0.290^{+0.115}_{-0.087}$ (1.00)	$0.820^{+0.025}_{-0.026}$ (1.00)	$0.18^{+0.61}_{-0.82}$ (1.00)
	Peak counts	$0.278^{+0.063}_{-0.051}$ (0.57)	$0.819^{+0.029}_{-0.026}$ (1.08)	$1.11^{+1.14}_{-1.28}$ (1.70)
	CNN	$0.250^{+0.036}_{-0.031}$ (0.33)	$0.826^{+0.020}_{-0.019}$ (0.77)	$-0.04^{+0.58}_{-0.61}$ (0.83)
Free baryonic model	Power spectrum	$0.259^{+0.139}_{-0.083}$ (1.00)	$0.816^{+0.031}_{-0.030}$ (1.00)	$0.08^{+0.68}_{-0.98}$ (1.00)
	Peak counts	$0.268^{+0.082}_{-0.047}$ (0.58)	$0.822^{+0.034}_{-0.029}$ (1.04)	$1.07^{+1.19}_{-1.32}$ (1.51)
	CNN	$0.268^{+0.040}_{-0.036}$ (0.35)	$0.819^{+0.034}_{-0.024}$ (0.95)	$-0.16^{+0.59}_{-0.58}$ (0.71)

- Extra info found by CNN: factor of ~3 improvement for Ω_m mostly from degeneracy breaking (S_8 not improved)
- S_8 tension: reduced from 2.2σ to 0.4σ by baryons (vs. Planck)
- baryons: Ω_m modestly affected, S_8 degraded by factor of ~1.8

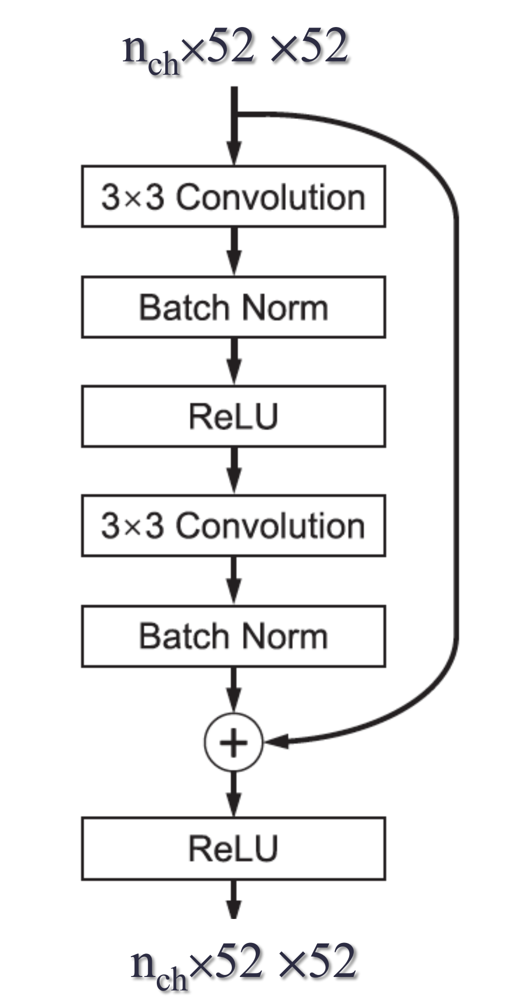
Conclusions

- CNNs: in the presence of baryon uncertainties, constraints improve by up to a factor of ~3 for Ω_m but only 5% for S_8, compared to the power spectrum
- Baryons: marginalizing over baryons degrade constraints by x 1.8 for S_8 , only modest effect on Ω_m . Baryon model unconstrained.
- S_8 tension: baryons smooth fluctuations on few arcmin scales → adding them increases S_8 and eliminates tension with CMB
- The future: 10^{7-8} → few x 10^9 gals with LSST, Euclid, Roman → can go to smaller scales; advantage of CNN should increase

The End

Network architecture

Layer/ block	Kernel size	Stride	Output dimension
(Input)			$4 \times 104 \times 104$
Convolution	7×7	2	$n_{ch} \times 52 \times 52$
Residual block 1	–	–	$n_{ch} \times 52 \times 52$
⋮			
Residual block n_{block}	–	–	$n_{ch} \times 52 \times 52$
Pooling	2×2	2	$n_{ch} \times 26 \times 26$
Convolution	3×3	1	$(2\,n_{ch}) \times 24 \times 24$
Pooling	2×2	2	$(2\,n_{ch}) \times 12 \times 12$
Convolution	3×3	1	$(4\,n_{ch}) \times 10 \times 10$
Pooling	2×2	2	$(4\,n_{ch}) \times 5 \times 5$
Convolution	3×3	1	$(8\,n_{ch}) \times 3 \times 3$
Pooling	3×3	2	$(8\,n_{ch}) \times 1 \times 1$
Linear	–	–	256
ReLU	–	–	256
Linear	–	–	N_θ

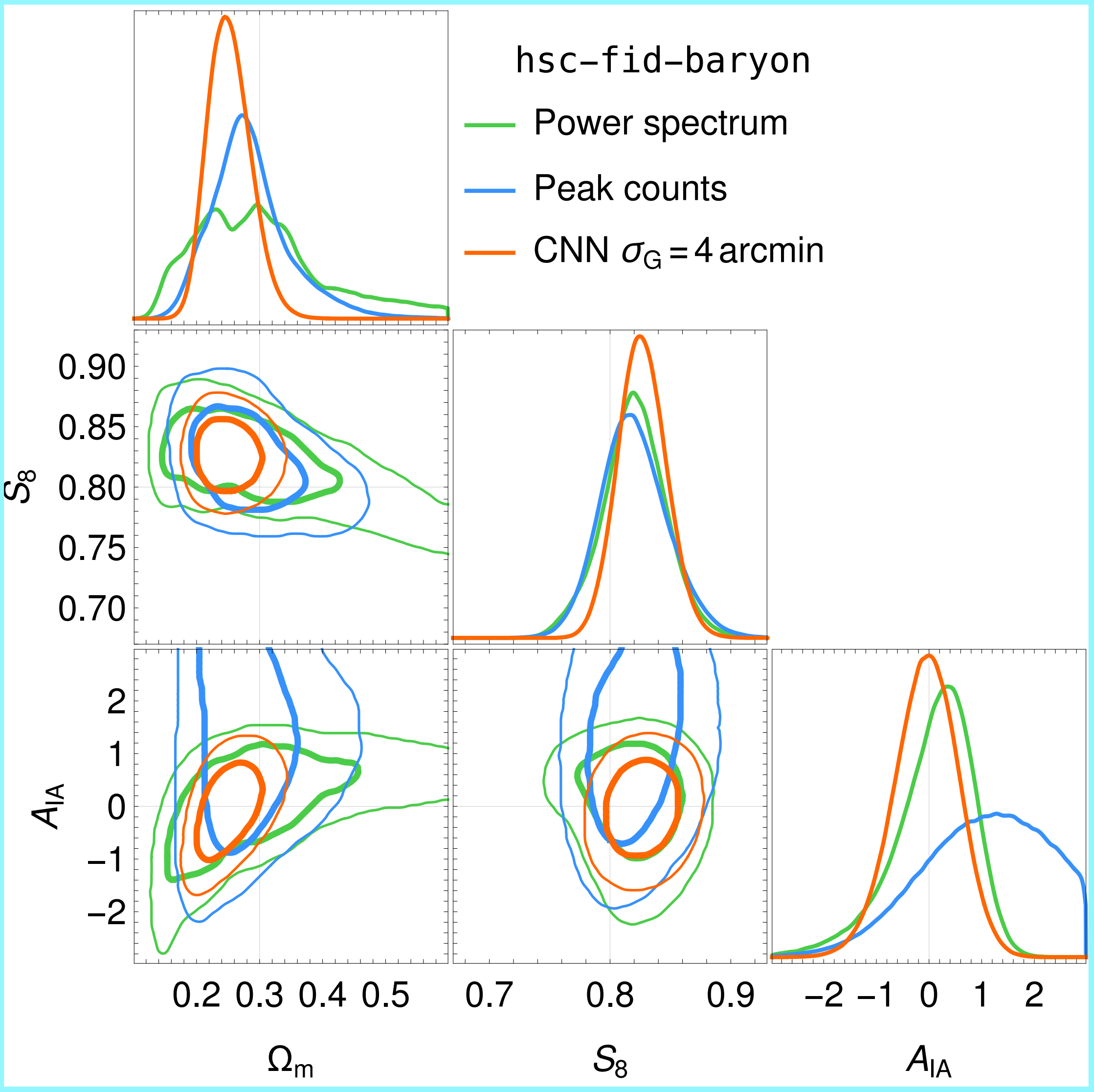


Alternative architectures

n_{block}	n_{ch}	σ_G	Statistical error		
			Ω_m	S_8	A_{IA}
5	32	4 arcmin	0.0778	0.0351	1.15
10	32	4 arcmin	0.0769	0.0351	1.15
15	32	4 arcmin	0.0782	0.0354	1.15
5	64	4 arcmin	0.0776	0.0354	1.14
10	**64**	**4 arcmin**	**0.0767**	**0.0355**	**1.16**
15	64	4 arcmin	0.0756	0.0351	1.16
5	96	4 arcmin	0.0779	0.0355	1.15
10	96	4 arcmin	0.0761	0.0349	1.15
15	96	4 arcmin	0.0765	0.0353	1.13
10	64	8 arcmin	0.0918	0.0389	1.20
10	**64**	**4 arcmin**	**0.0767**	**0.0355**	**1.16**
10	64	2 arcmin	0.0696	0.0338	1.13
10	64	1 arcmin	0.0696	0.0338	1.13

Results - fixed baryons

Lu, ZH & Li (2023)

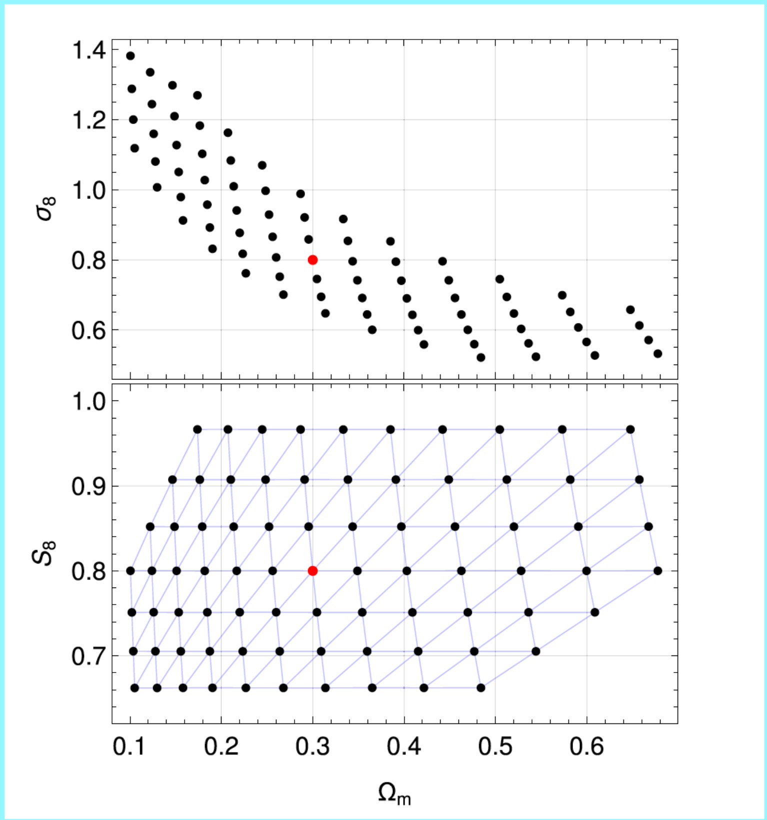


similar results

forward-modeling suite

Table 2. The prior distributions of all parameters.

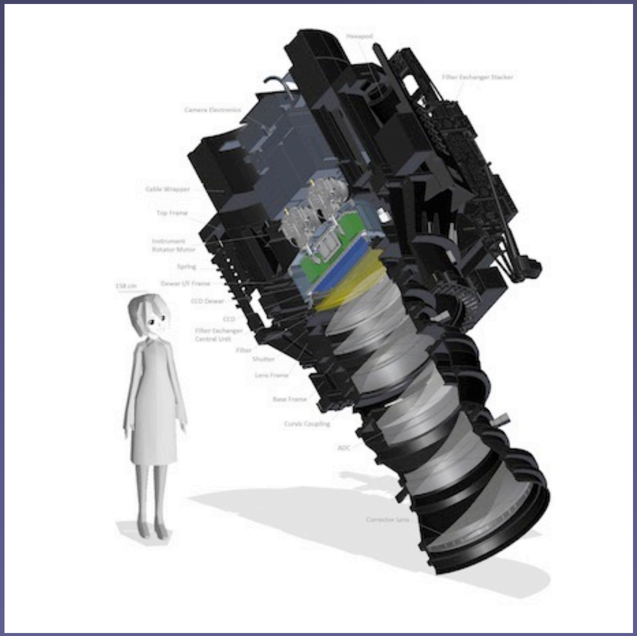
Parameter	Fiducial value	Prior
Cosmological and baryonic parameters		
Ω_m	0.3	flat within the boundary of simulated cosmologies
σ_8	0.8	flat within the boundary of simulated cosmologies
A_{IA}	0	flat $[-3, 3]$
M_c	$3.3 \times 10^{13} h^{-1} M_\odot$	log-uniform $(10^{12} h^{-1} M_\odot, 10^{16} h^{-1} M_\odot)$
$M_{1,0}$	$8.63 \times 10^{11} h^{-1} M_\odot$	log-uniform $(10^{10} h^{-1} M_\odot, 10^{13} h^{-1} M_\odot)$
η	0.54	log-uniform $(10^{-0.7}, 10^{0.5})$
β	0.12	log-uniform $(10^{-1.0}, 10^{0.5})$
Nuisance parameters		
Δz_1	–	Gaussian $\mathcal{N}(0, 0.0285)$
Δz_2	–	Gaussian $\mathcal{N}(0, 0.0135)$
Δz_3	–	Gaussian $\mathcal{N}(0, 0.0383)$
Δz_4	–	Gaussian $\mathcal{N}(0, 0.0376)$
Δm	–	Gaussian $\mathcal{N}(0, 0.01)$
α_{psf}	–	Gaussian $\mathcal{N}(0.030, 0.015)$
β_{psf}	–	Gaussian $\mathcal{N}(-0.89, 0.70)$

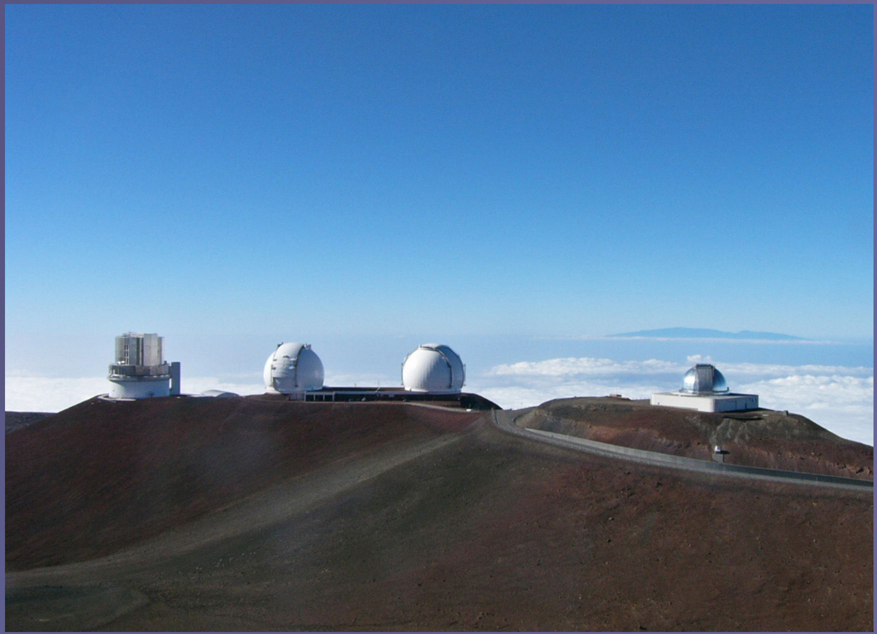


- *79 cosmologies*: 1600 realizations each
- *With baryons*: Sobol sequence in 4D baryon parameter space, 1600 combos (2.4×10^6 maps in total)
- *CNNs*: 50% training, 50% inference

Hyper Supreme-Cam (HSC)

"gigantic" 870 Mega-pixel camera on 8.2m Subaru telescope





- survey data acquired from 2014-2021
- covers 1500 deg^2 (3.5% of the sky)
- depth: 20 galaxies per $arcmin^2$
- Yr-1 public (134 deg^2 , 11.9 million galaxies)
- Yr-3 recently released

Systematic effects

- Shape noise: randomly rotate galaxies
- Photo-z: marginalize over uncertain Δz~0.01-0.04
- PSF: marginalize over residuals in PSF shape
- Shear measurement: include multiplicative and additive bias
- Intrinsic alignment: adopt "NLA" model on map level:

$$\kappa_{\rm IA}^{(b)}(\boldsymbol{x}) = \int {\rm d}z\, F(z) n_{\rm g}^{(b)}(z) \delta(\boldsymbol{x}, z),$$

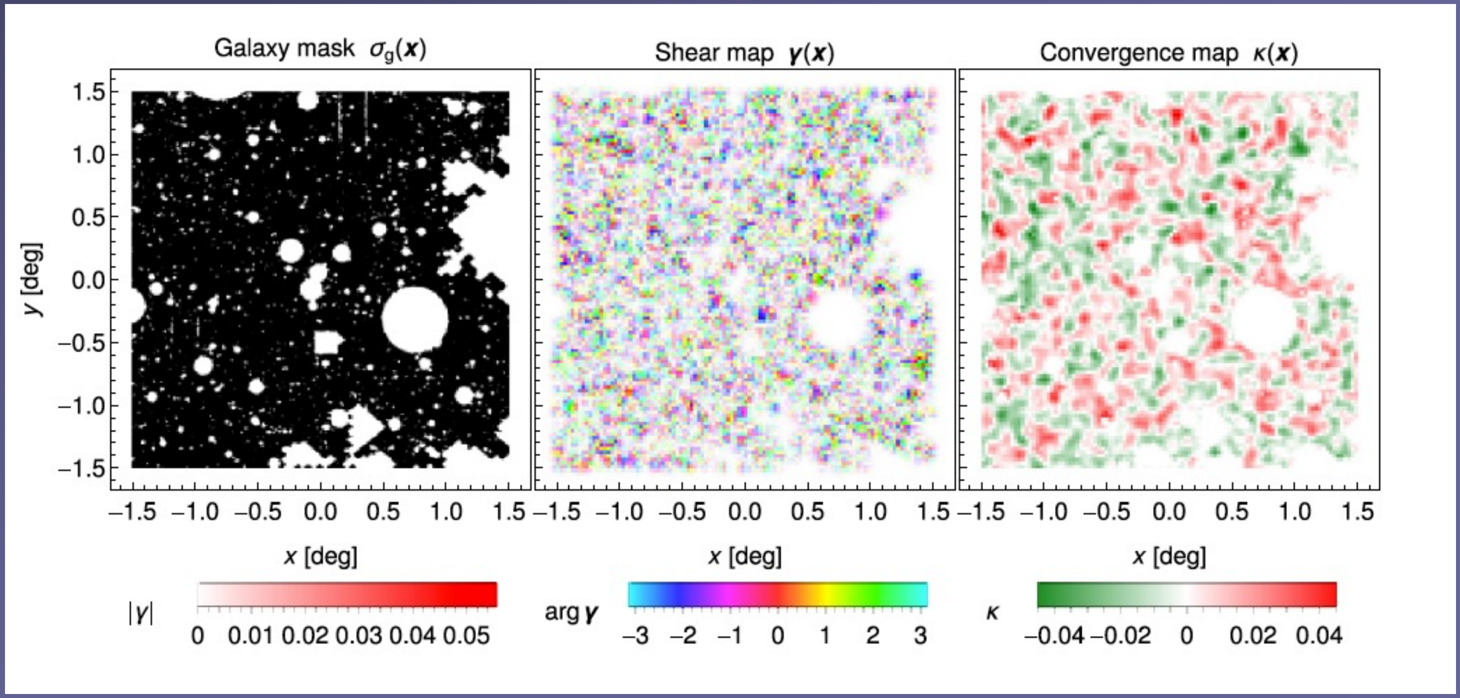
$$F(z) = -A_{\rm IA} C_1 \bar{\rho}(z) \frac{D(z)}{D(0)} \left(\frac{1+z}{1+z_0} \right)^{\eta} \left(\frac{\bar{L}}{L_0} \right)^{\beta}.$$

A_{IA} a free parameter

- Baryonic effects: paint baryons onto N-body sims with 4-parameter baryon-correction model (BCM): M_c M_1 η β

Example subfield

$3{\times}3$ deg^2 convergence (κ) map



Looking for beyond-Gaussian info

Approaches:

1 perturbative expansions:

higher-order moments (skewness, kurtosis …)

higher-order correlation functions (3pt, 4pt ….)

Fourier counterparts (bispectra, trispectra ….)

2 Other morphological ”features”:

peaks, Minkowski functionals, shapelets …

3 ”Gaussianization”: transform lensing field locally

4 machine learning: *can be cast as 2D image classification*

Questions:

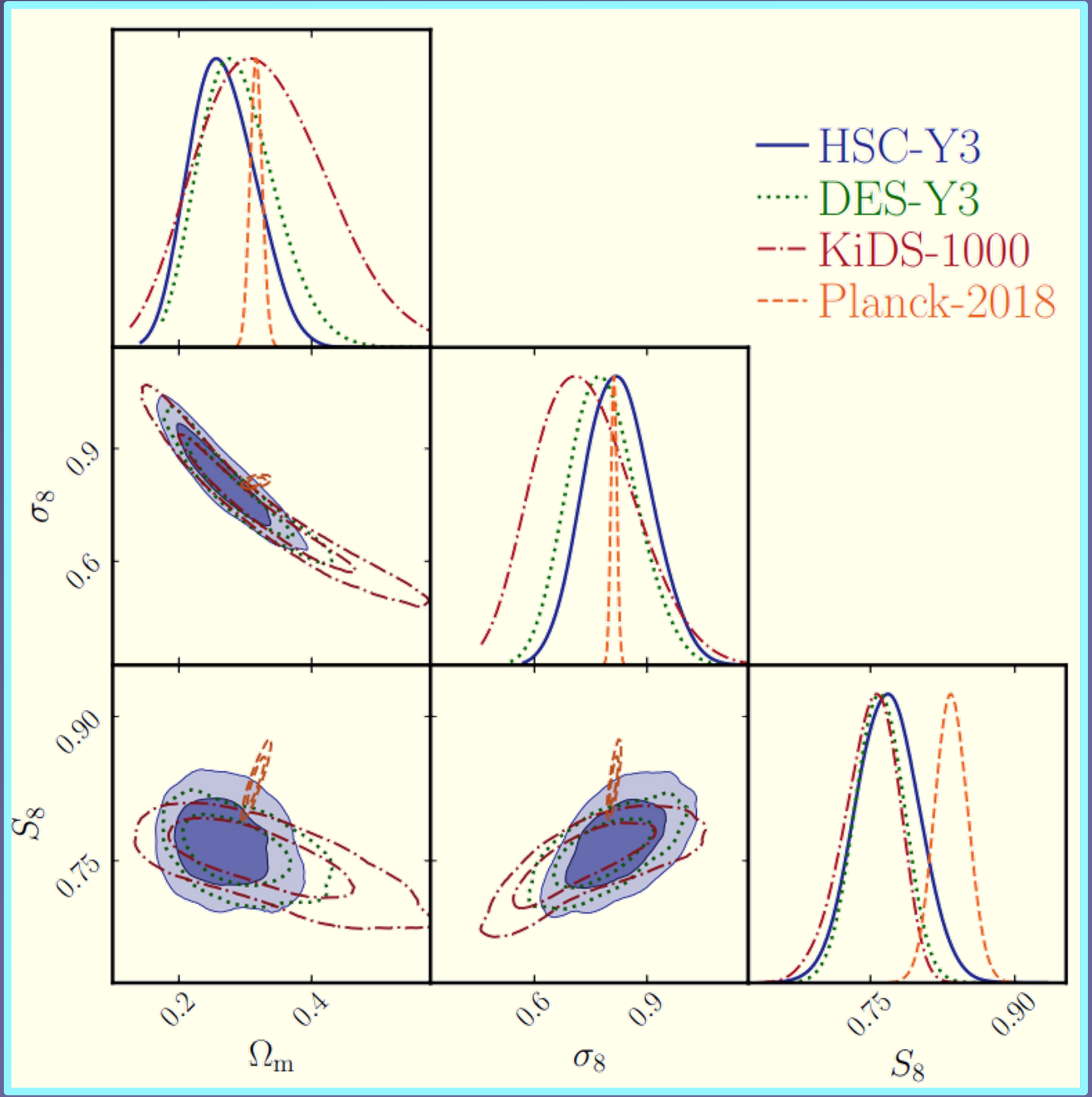
- how do these respond to cosmology *vs* systematics
- extra info is from small, nonlinear scales - modeling
- how do you tell whether most info has been found?

Cosmology Inference

HSC Y3 constraints using two-point correlation function

best-constrained combination

$$S_8 \equiv \sigma_8 \sqrt{\Omega_\mathrm{m}/0.3}$$

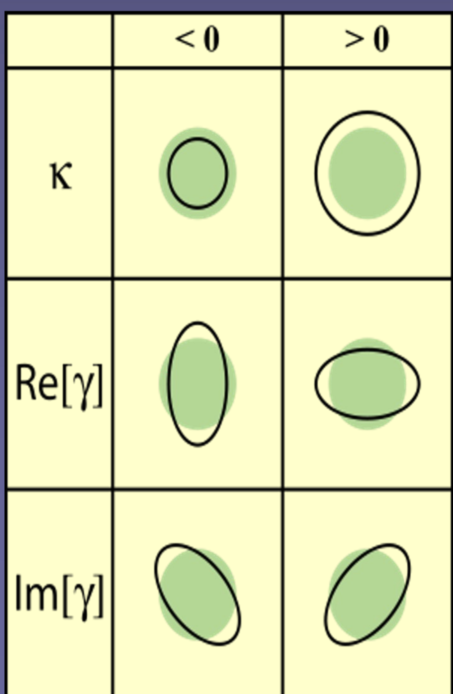


Gravitational Lensing

Unlensed position(θ_j)
Observed position (θ_i)

$$f_{obs}(\theta_i) = f_s(A_{ij}\theta_j)$$

$$A_{ij} = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix}$$



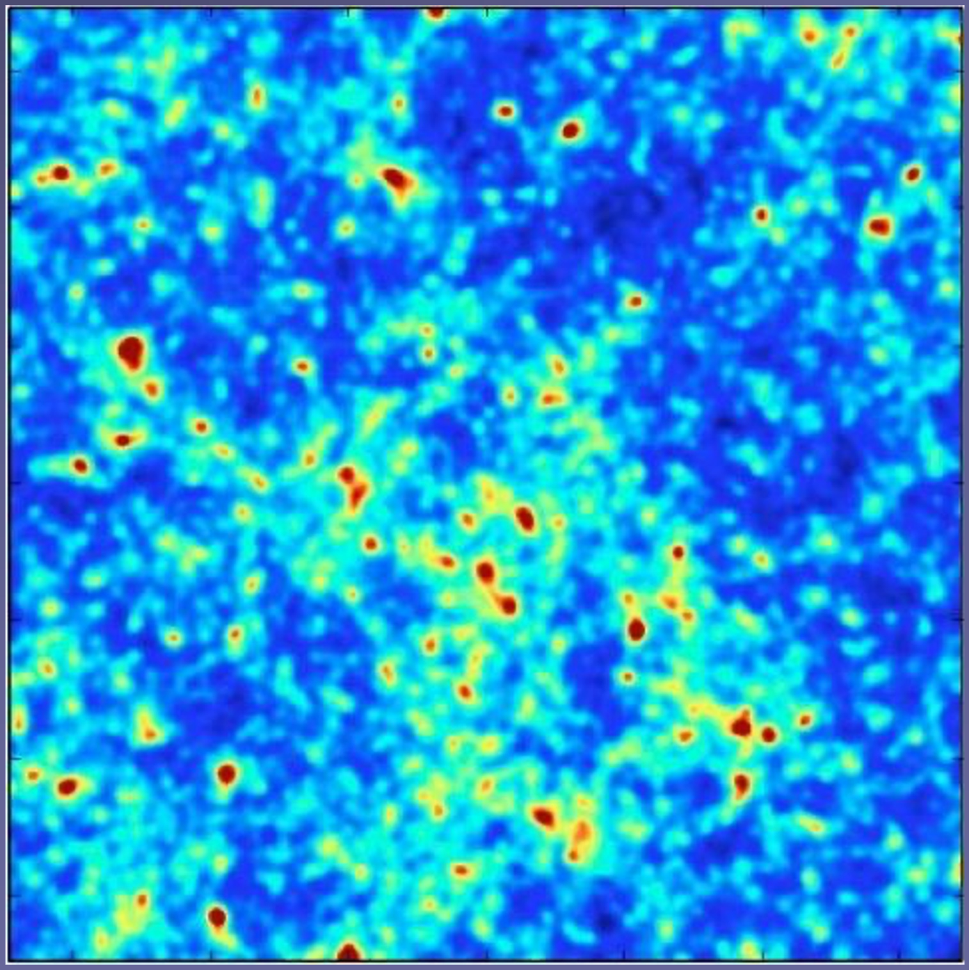
First deep field of JWST
(centered on a galaxy cluster)

Weak lensing: convergence map

- Measure ellipticities of galaxies
- Convert to convergence (=magnification)
- Smooth over few arcmin^2 patches (~100 gal)

$$\hat{\kappa}(\mathbf{s}) = \frac{1}{2}\left(\frac{k_1^2 - k_2^2}{k_1^2 + k_2^2}\right)\hat{\gamma}_1(\mathbf{s}) + \frac{k_1 k_2}{k_1^2 + k_2^2}\hat{\gamma}_2(\mathbf{s})$$

Kaiser & Squires (1993)



Workhorse: 2-point functions

- **Real-space: two-point correlation functions**

  $\xi(\vec{\Theta}) = \langle \kappa(\vec{\theta})\, \kappa(\vec{\theta}+\vec{\Theta})\rangle$

  $\langle \vec{\gamma}\,\vec{\gamma} \rangle$ in principle, same information

- **Fourier space: convergence power spectrum**

  $\langle \kappa(\vec{l})\, \kappa^*(\vec{l}-\vec{l}')\rangle = 2\pi\delta(\vec{l}-\vec{l}')\, P(l)$

$$P_\kappa(l) = \frac{9}{4}\Omega_m^2 \frac{H_0^4}{c^4}\int_0^\infty dz \left[\frac{d\chi(z)}{dz}\right] \frac{\xi^2[\chi(z)]}{a^2(z)} P_{3\mathrm{D}}\left(\frac{l}{\chi(z)};z\right),$$

$$\xi(\chi) = \int_z^\infty dz'\, n_{\mathrm{gal}}(z')\, \frac{\chi(z')-\chi(z)}{\chi(z')}.$$

Workhorse: 2-point functions

- **Real-space: two-point correlation functions**

  $\xi(\Theta) = \langle \kappa(\vec{\theta})\, \kappa(\vec{\theta}+\vec{\Theta}) \rangle$

  $\langle \vec{\gamma}\,\vec{\gamma} \rangle$ **in principle, same information**

- **Fourier space: convergence power spectrum**

  $\langle \kappa(\vec{l})\, \kappa^*(\vec{l}-\vec{l}') \rangle = 2\pi\delta(\vec{l}-\vec{l}')\, P(l)$

$$P_\kappa(l) = \frac{9}{4}\Omega_m^2 \frac{H_0^4}{c^4} \int_0^\infty dz \left[\frac{d\chi(z)}{dz}\right] \frac{\xi^2[\chi(z)]}{a^2(z)} P_{3D}\left(\frac{l}{\chi(z)}; z\right),$$

$$\xi(\chi) = \int_z^\infty dz'\, n_{gal}(z') \frac{\chi(z') - \chi(z)}{\chi(z')}.$$

geometry

Workhorse: 2-point functions

- **Real-space: two-point correlation functions**

  $\xi(\Theta) = \langle \kappa(\vec{\theta})\, \kappa(\vec{\theta}+\vec{\Theta}) \rangle$

  $\langle \vec{\gamma}\,\vec{\gamma} \rangle$ **in principle, same information**

- **Fourier space: convergence power spectrum**

  $\langle \kappa(\vec{l})\, \kappa^*(\vec{l}-\vec{l}') \rangle = 2\pi\delta(\vec{l}-\vec{l}')\, P(l)$

$$P_\kappa(l) = \frac{9}{4}\Omega_m^2 \frac{H_0^4}{c^4}\int_0^\infty dz \left[\frac{d\chi(z)}{dz}\right] \frac{\xi^2[\chi(z)]}{a^2(z)} P_{3D}\left(\frac{l}{\chi(z)};z\right),$$

$$\xi(\chi) = \int_z^\infty dz'\, n_{\rm gal}(z') \frac{\chi(z')-\chi(z)}{\chi(z')}.$$

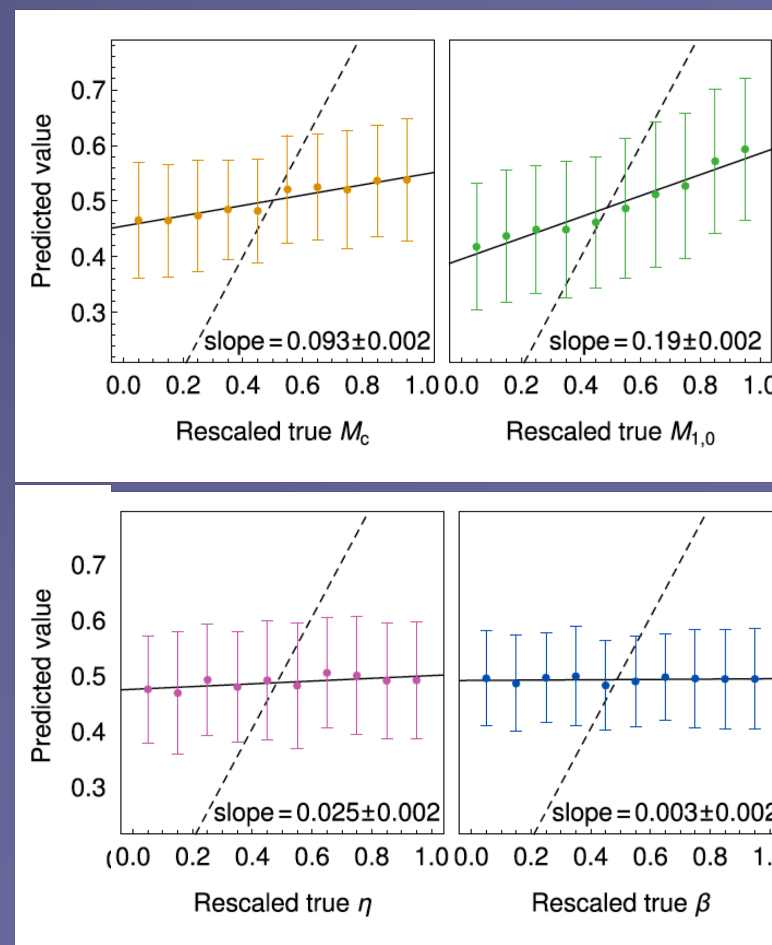
geometry
growth

Jointly fit cosmology & baryons

Lu, ZH & Zorrilla 2022

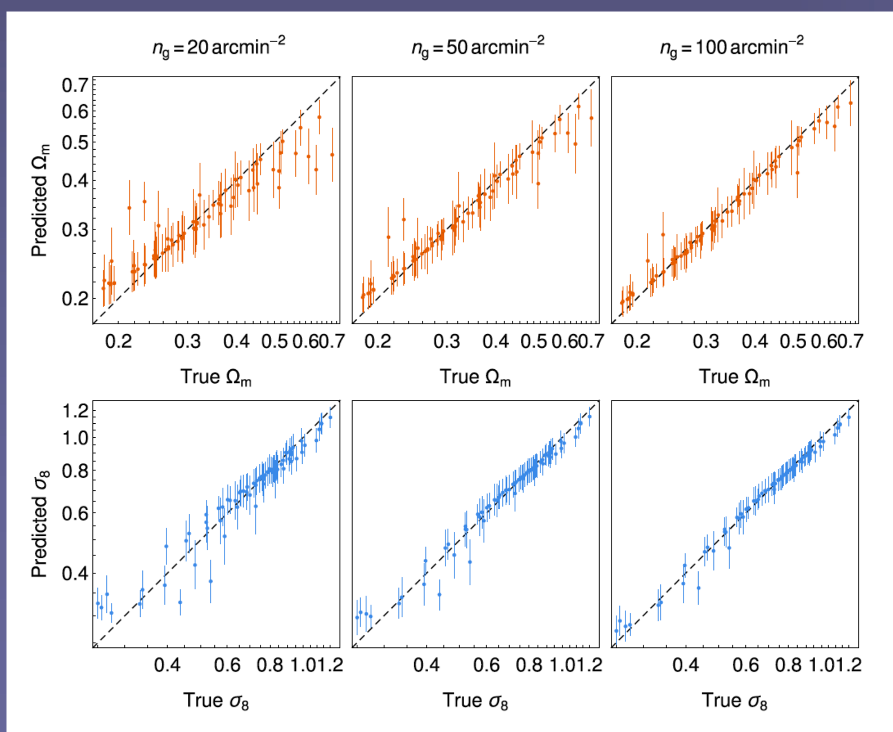
Baryons

- Network can learn M_c + $M_{1,0}$
- but not β or η



Cosmology

- Can predict parameters,
- tilt/bias (corrected in likelihood)



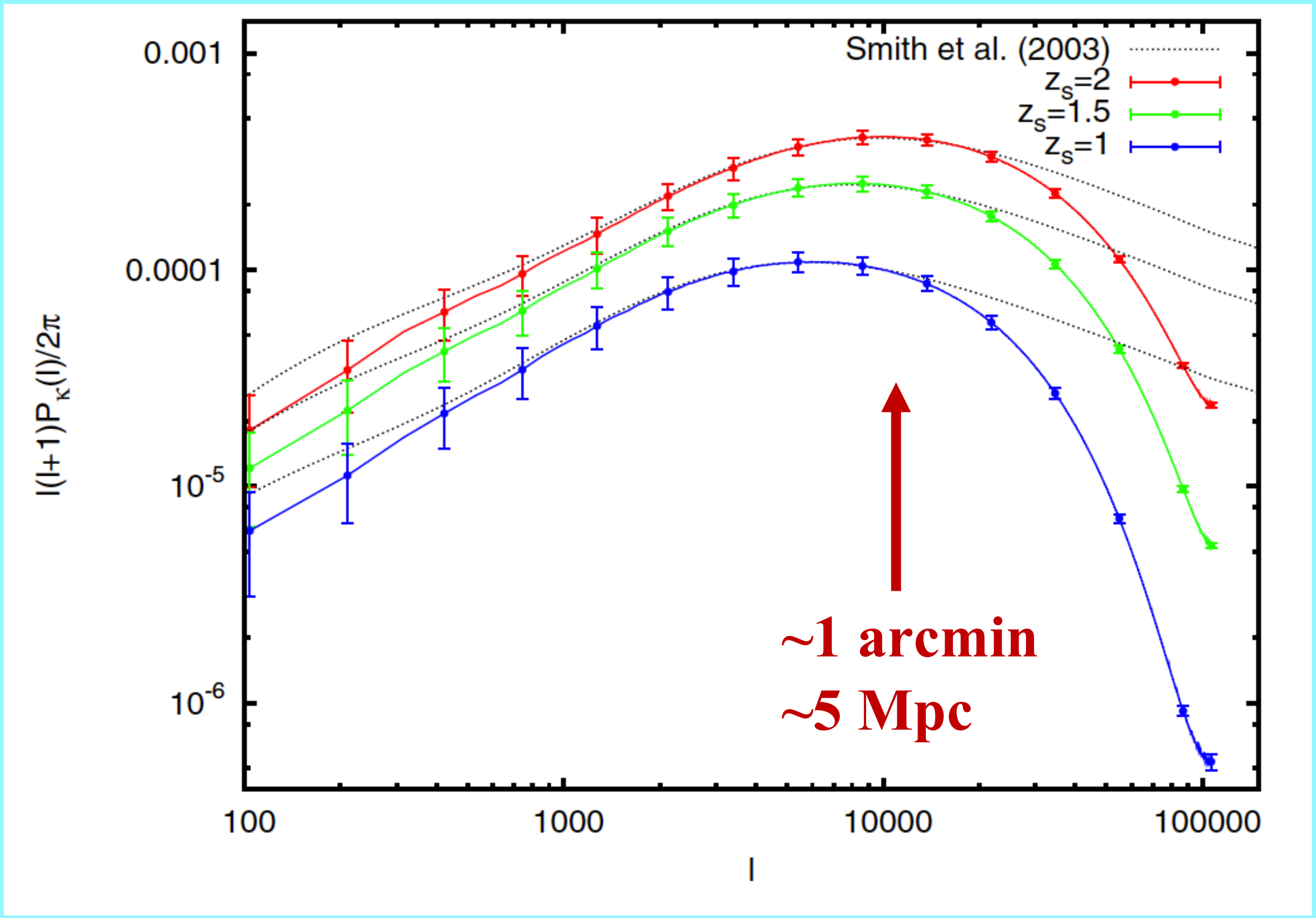
Baryons with machine learning

Lu, ZH & Zorrilla 2022

Methods	$\Omega_m - \sigma_8$			$M_{1,0} - \eta$		
	S_{full} ($\times 10^{-4}$)	S_{fid} ($\times 10^{-4}$)	S_{full}/S_{fid}	S_{full} ($\times 10^{-2}$)	S_{fid} ($\times 10^{-2}$)	S_{full}/S_{fid}
Power spectrum	3.45	0.93	3.71	10.4	3.6	2.88
Peak counts	5.89	0.94	6.28	30.6	7.3	4.16
CNN	2.08	0.44	4.70	13.0	3.7	3.48
CNN + power spectrum (L)	1.27	0.44	2.91	7.1	2.6	2.69
CNN + power spectrum (M)	1.11	0.42	2.61	6.9	2.8	2.41
CNN + power spectrum (S)	1.74	0.41	4.23	9.7	3.0	3.26
CNN + power spectrum (L, M)	1.01	0.42	2.39	5.2	2.3	2.24
CNN + power spectrum (full)	0.96	0.40	2.41	4.6	2.1	2.24

- CNN improves over peaks/power spectrum by factor of ~1.8.
- With baryons, peaks degrade the most
- CNN was unable to learn the medium and large-scale power spectrum – so their combination mitigates degradation
- For baryon parameters, CNN comparable to power spectrum but independent

Convergence power spectrum



Kratochvil et al. 2012

Cosmology results

signal is weak (~1%), must average over many galaxies:

(0.3/√900) → 900 galaxies for S/N=1 detection of a systematic γ ~0.01

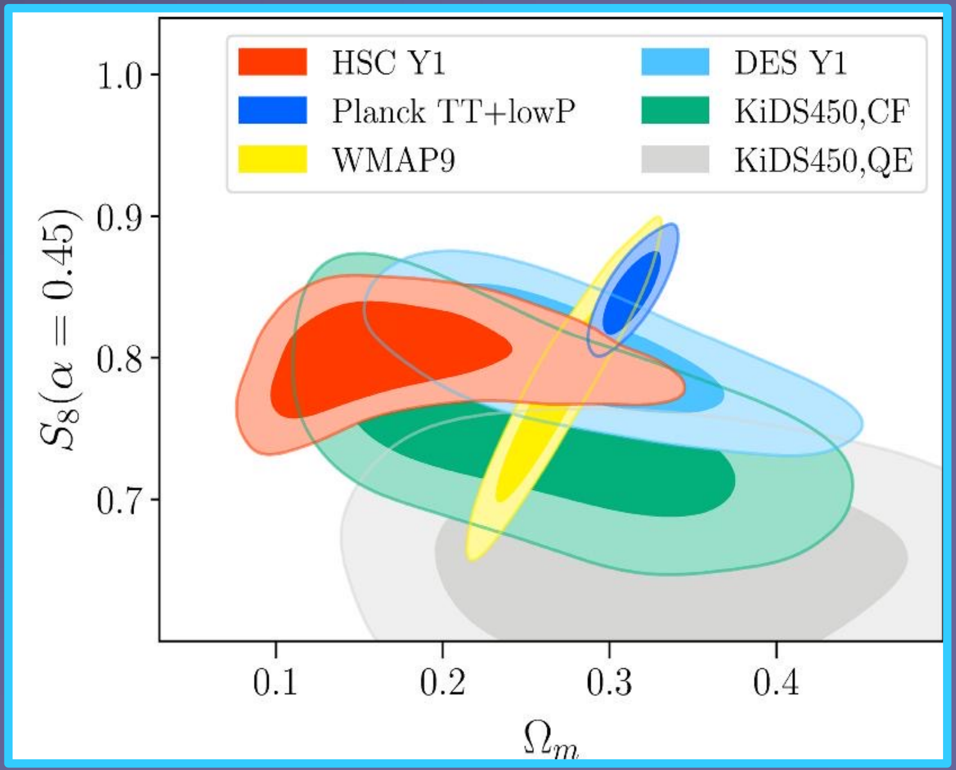
→ 900×10^4 ~ 10^7 galaxies for ~1% error on γ → need ~100 deg^2

Canada-France-Hawaii Telescope (CFHTLenS)
154 deg^2 imaging (6×10^6 gals)
Kilbinger et al. (2013)

Kilo Degree Survey (KiDS-1000)
1006 deg^2 imaging in 4 bands (25×10^6 gals)
Heymans et al. (2020)

Dark Energy Survey (DES; Year 3)
4143 deg^2 imaging in 5 bands (100×10^6 gals)
Amon et al. (2021)

Subaru Hyper Suprime-Cam (HSC; Year 1)
137 deg^2 imaging in 5 bands (9×10^6 gals)
Hikage et al. (2019)



The Future: Full HSC, Euclid, LSST, Roman 10^7 → 10^8 → 10^9+ gals

CNN on noisy maps

Confidence range ratios around two input cosmologies
$(\Omega_m, \sigma_8) = (0.26, 0.8) - (0.309, 0.816)$

Table 2. The table lists the relative sizes of the 68 percent credible contour areas of the power spectrum and peak counts compared to the CNN. The CNN achieves smaller 68 percent credible contour areas than the power spectrum for any noise level, and also outperforms the peak counts when the galaxy density is at least 30 arcmin^{-2}.

A_{68} ratio	Noiseless	100 gal arcmin^{-2}	75 gal arcmin^{-2}	50 gal arcmin^{-2}	30 gal arcmin^{-2}	10 gal arcmin^{-2}
Power spectrum / CNN	13	3.7–4.6	3.5–4.1	3–3.6	2.4–2.8	1.4–1.5
Peak counts / CNN	8	1.5–2.1	1.4–1.9	1.2–1.7	1.05–1.42	0.9–1.1