



FAIR
Universe

Weak Lensing ML Uncertainty Challenge

Join the competition today from [HERE!](#)

One of the **NeurIPS 2025 Competitions**



U.S. DEPARTMENT
of ENERGY

IAS | INSTITUTE FOR
ADVANCED STUDY

BERKELEY LAB

Berkeley
UNIVERSITY OF CALIFORNIA

CHA
LEARN

Stanford
University

université
PARIS-SACLAY

SLAC NATIONAL
ACCELERATOR
LABORATORY

W
UNIVERSITY of WASHINGTON

UCI University of
California, Irvine

Weak Lensing ML Uncertainty Challenge

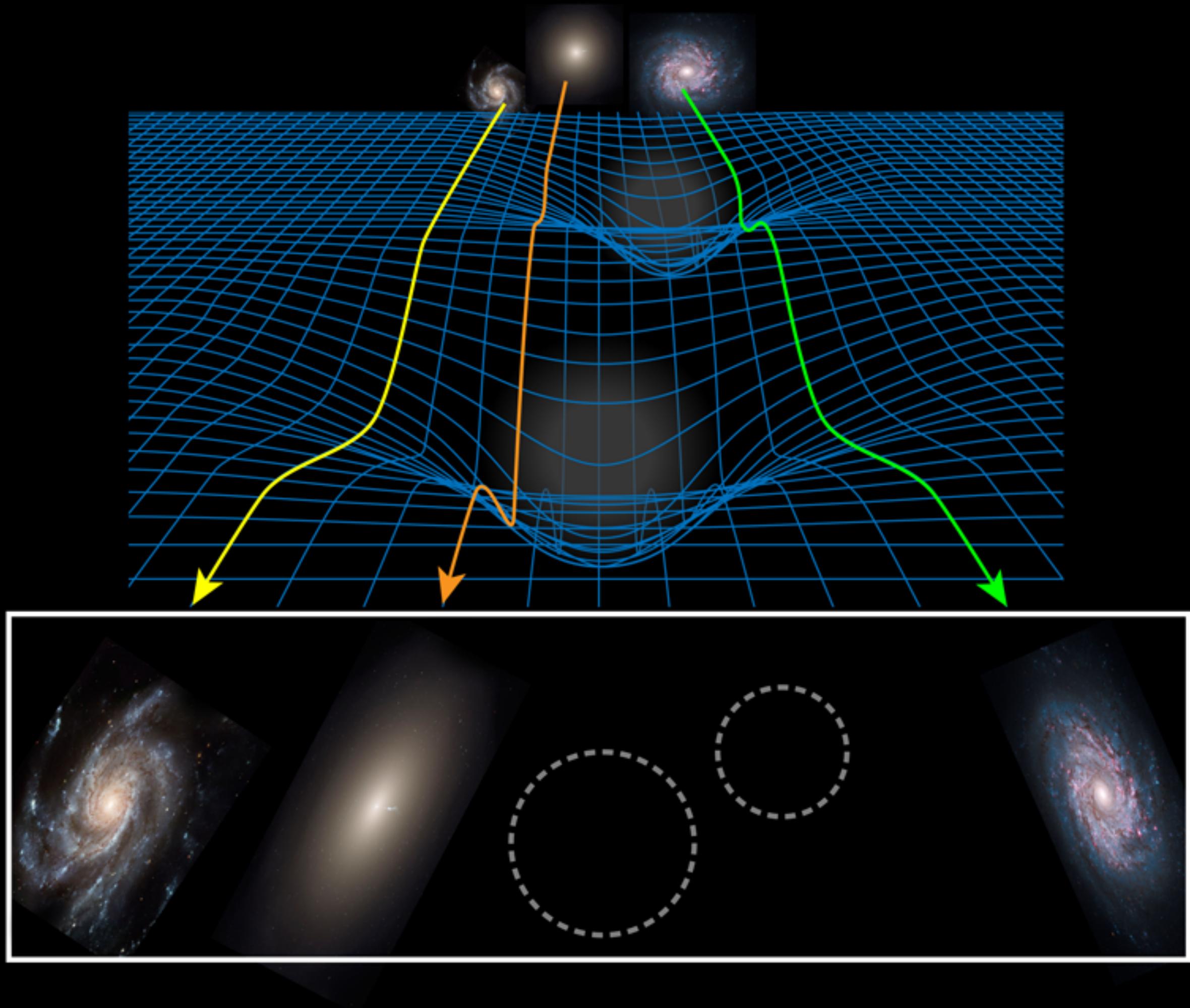


Overview

Weak gravitational lensing

- The gravity of matter warps the surrounding space-time and causes distortions in the observed shapes of the background galaxies.
- Powerful probe of the matter distribution in our universe from coherent patterns of galaxy shapes.
- Numerous current and upcoming WL surveys: DES, HSC, Euclid, Rubin LSST, Roman, etc.
- Traditional analysis based on two-point correlation functions can only capture limited amount of information from the weak lensing data.
- **AI/ML-based approaches could capture more information hidden in higher-order correlations!**

WE NEED YOU!



Weak Lensing ML Uncertainty Challenge



Competition tasks

The competition tasks will be structured into **2 phases**:

- **Phase 1: Cosmological Parameter Estimation**

Participants will develop models that:

- Accurately infer cosmological parameters ($\hat{\Omega}_m, \hat{S}_8$) from the weak lensing image data.
- Quantify uncertainties via the 68% confidence intervals of the parameters of interest. ($\hat{\sigma}_{\Omega_m}, \hat{\sigma}_{S_8}$).

- **Phase 2: Out-of-Distribution Detection**

Some test data will be generated with different physical models (OoD), leading to some distribution shifts with respect to the test data in Phase 1. Participants will develop models that:

- Identify test data samples inconsistent with the training distribution (OoD detection).
- Provide probability estimates indicating data conformity to training distributions.

Scoring metrics:

KL divergence between the true Gaussian-like posterior distribution and the Gaussian with the predicted mean and standard deviation:

$$\text{score}_{\text{inference}} = -\frac{1}{N_{\text{test}}} \sum_i^{N_{\text{test}}} \left\{ \frac{(\hat{\Omega}_{m,i} - \Omega_{m,i}^{\text{truth}})^2}{\hat{\sigma}_{\Omega_{m,i}}^2} + \frac{(\hat{S}_{8,i} - S_{8,i}^{\text{truth}})^2}{\hat{\sigma}_{S_{8,i}}^2} \right. \\ \left. + \log(\hat{\sigma}_{\Omega_{m,i}}^2) + \log(\hat{\sigma}_{S_{8,i}}^2) + \lambda \left[(\hat{\Omega}_{m,i} - \Omega_{m,i}^{\text{truth}})^2 + (\hat{S}_{8,i} - S_{8,i}^{\text{truth}})^2 \right] \right\}$$

$\lambda \equiv 10^3$: penalty factor for bad point estimates

Binary cross-entropy:

$$\text{score}_{\text{OoD}} = \frac{1}{N_{\text{test}}} \sum_i^{N_{\text{test}}} [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)]$$

where $y_i = 1$ if the dataset is InD; $y_i = 0$ if the dataset is OoD, and p_i is the probability estimates predicted by the model.

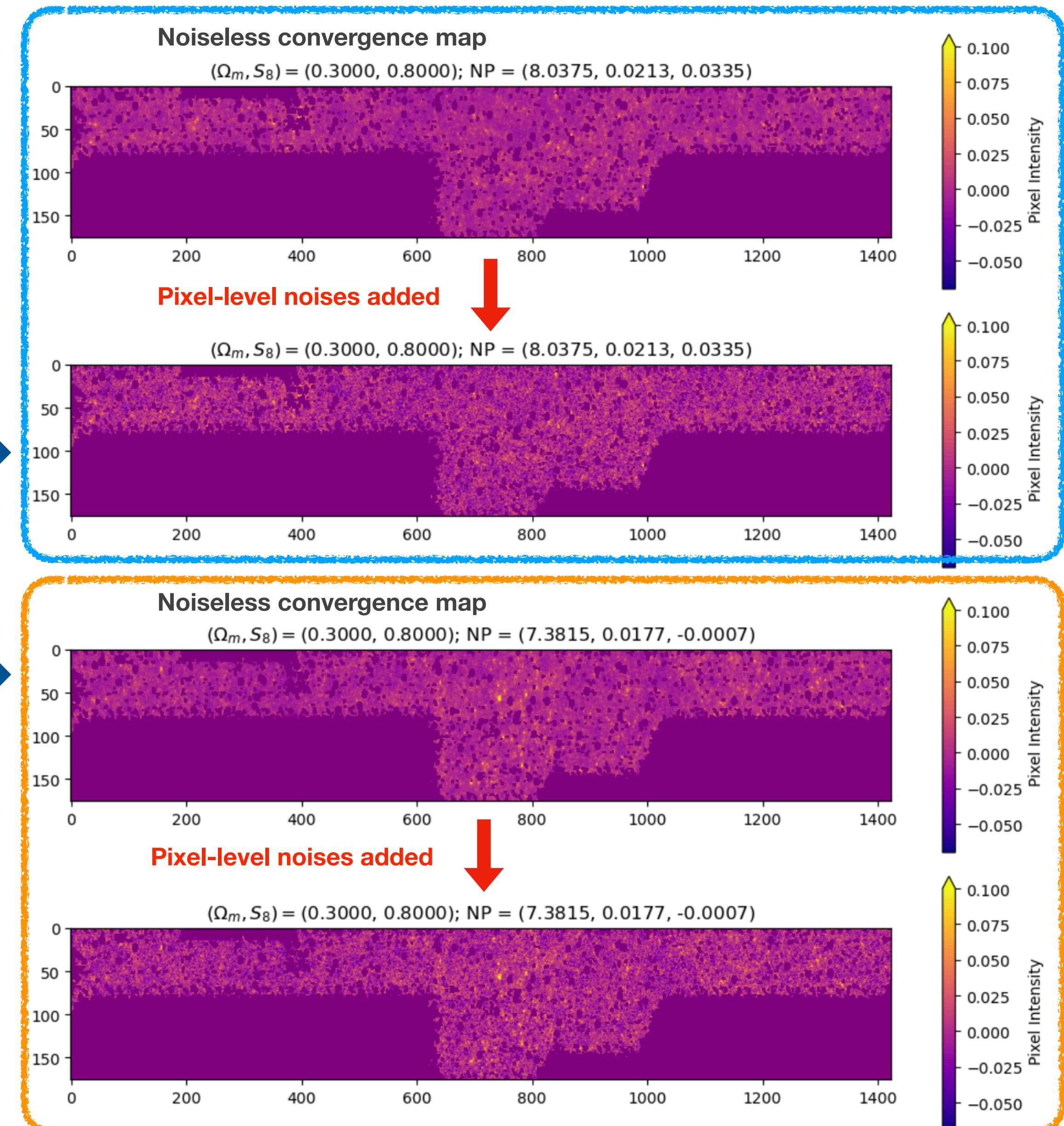
Weak Lensing ML Uncertainty Challenge



Dataset

- Mock galaxy catalogs predicted with N-body simulations and lens algorithms at 101 cosmological parameters (Ω_m, S_8)
- Pixelized 2D weak lensing images: convergence maps
- The model must take into account the systematic uncertainties from **3 realistic systematic effects**
 - 2 baryonic effect uncertainties**
 - 1 photometric redshift uncertainty**along with tunable **pixel-level noises**

Same cosmology,
different systematics



Weak Lensing ML Uncertainty Challenge

Dataset

The participants will be provided with:

- **Public training set:**

- Image data; shape = (101, 256, 1424, 176)
- Label shape = (101, 256, 5)

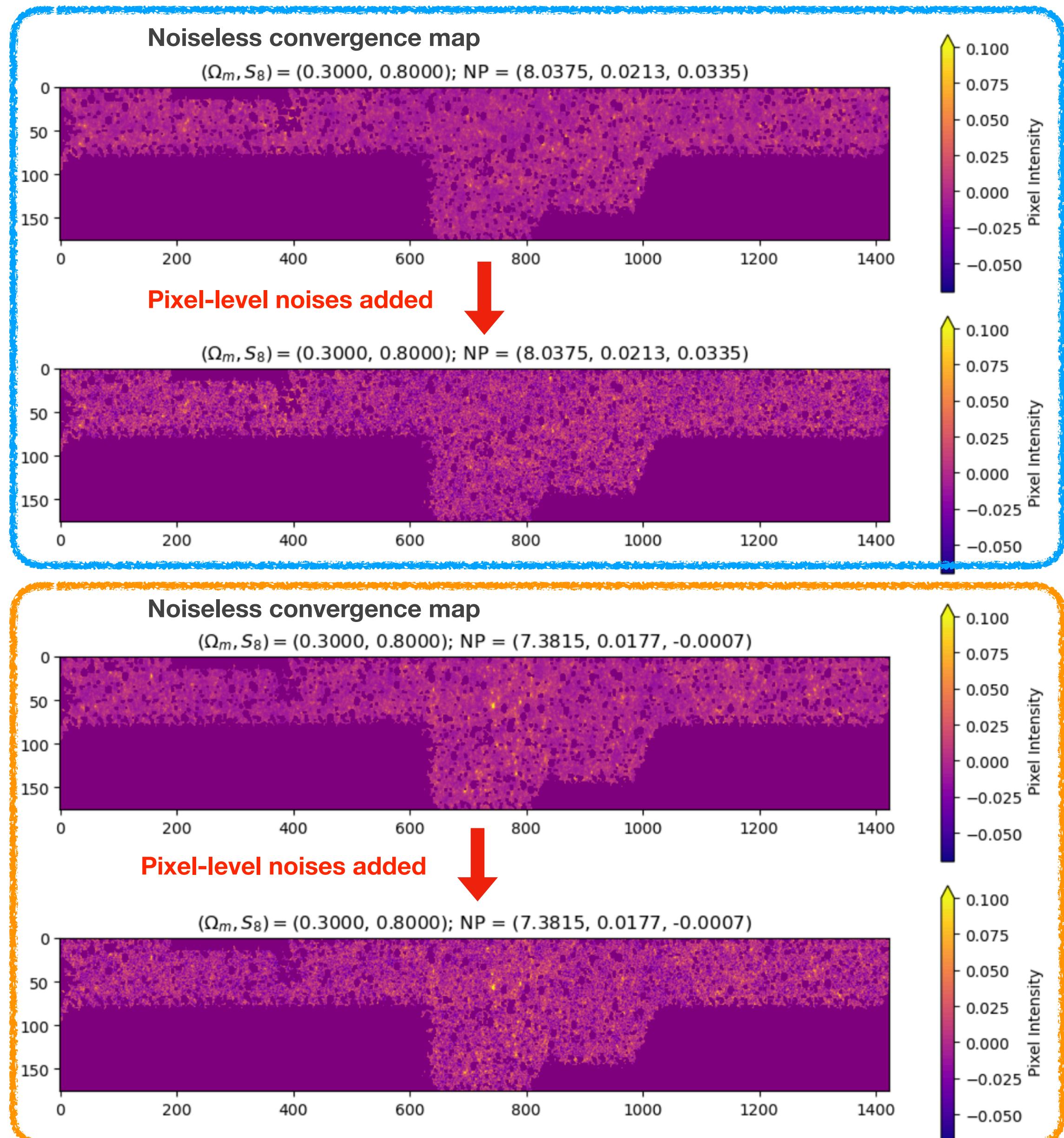
101 = Realizations of cosmological models; each characterized with 2 parameters of interest (Ω_m, S_8)

256 = Realizations of 3 nuisance parameters for systematics (1) and (2)

(1424, 176) = Image dimension

5 = 2 parameters of interest (Ω_m, S_8)
+ 3 nuisance parameters for systematics (1) and (2)

- **The provided training set is noiseless.** Participants can generate **pixel-level noise** to augment their training data using a simple `add_noise` function we provide



Weak Lensing ML Uncertainty Challenge



Phase 1 Dataset

The participants will be provided with:

- **Public test set:**
 - Image data; shape = (4000, 1424, 176)
 - 4000 = Number of test images
 - (1424, 176) = Image dimension
 - The test images are generated with random cosmological parameters, random nuisance parameters, and random pixel-level noises.

Phase 1 Evaluation

The true parameters ($\Omega_m^{\text{truth}}, S_8^{\text{truth}}$) of the public test set are unknown to the participants.

Participants will submit predictions of

- **Cosmological parameters** ($\hat{\Omega}_m, \hat{S}_8$)
- **Their uncertainties** ($\hat{\sigma}_{\Omega_m}, \hat{\sigma}_{S_8}$)

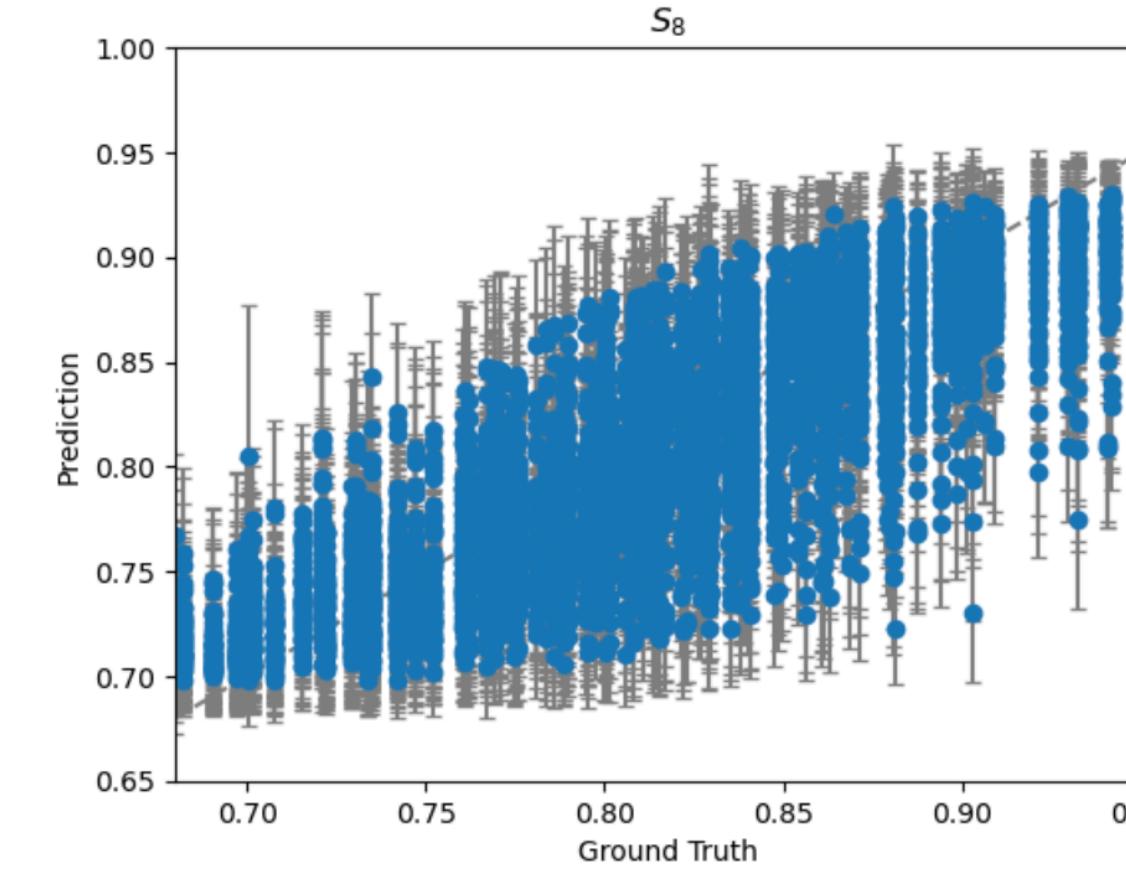
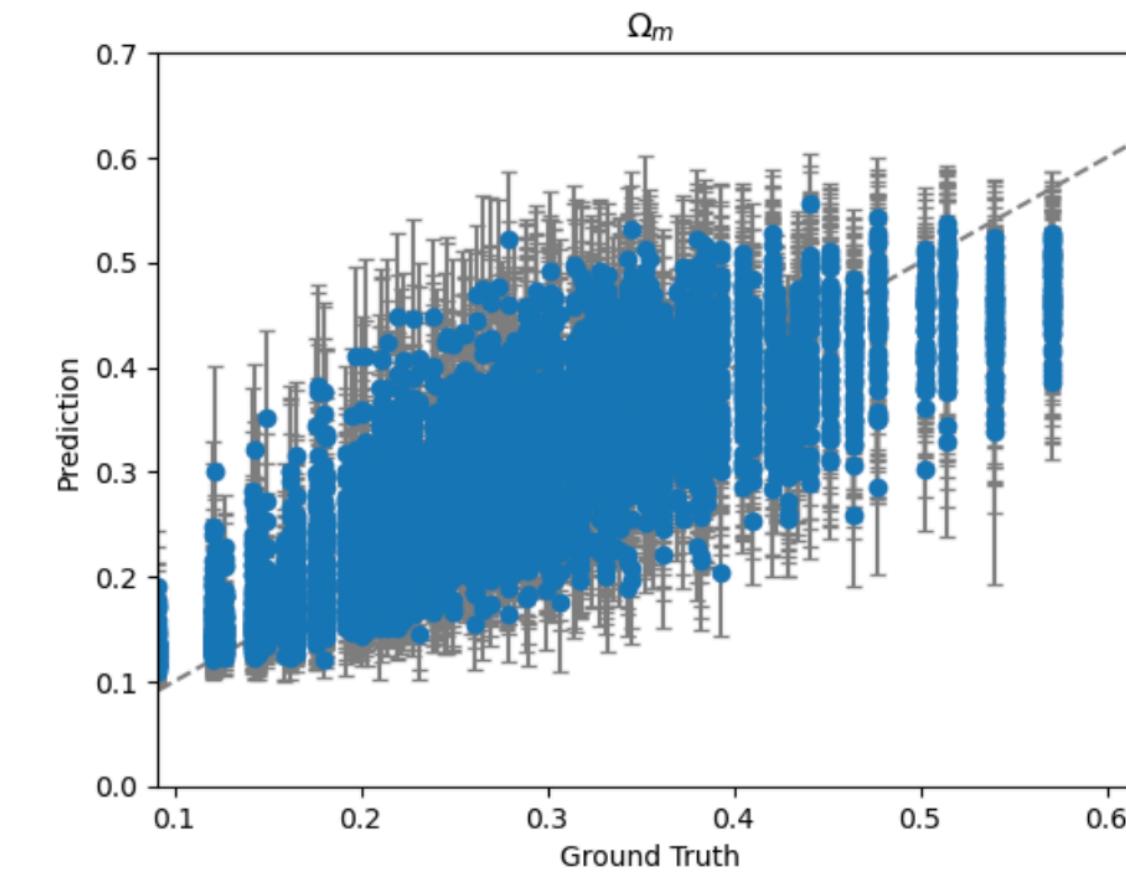
to [Codabench](#), our competition platform.

The model performance will then be evaluated with the hidden ground truth based on our scoring metrics.

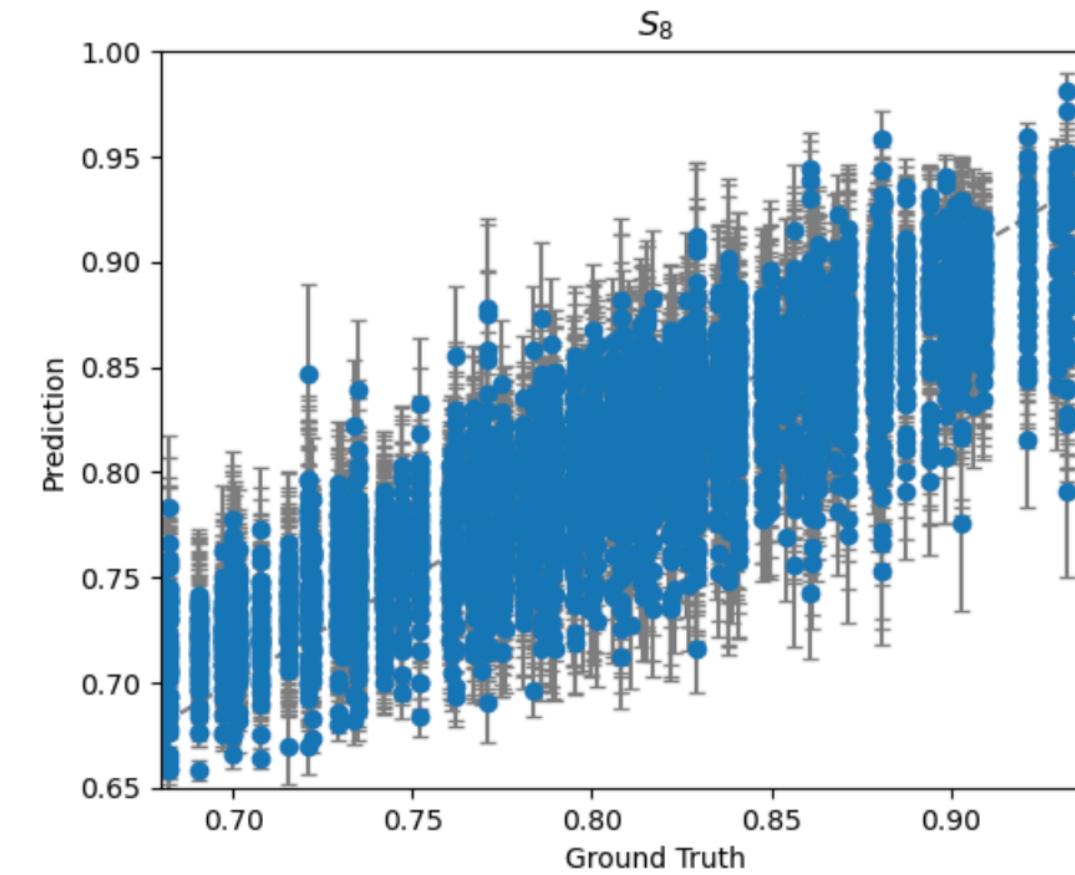
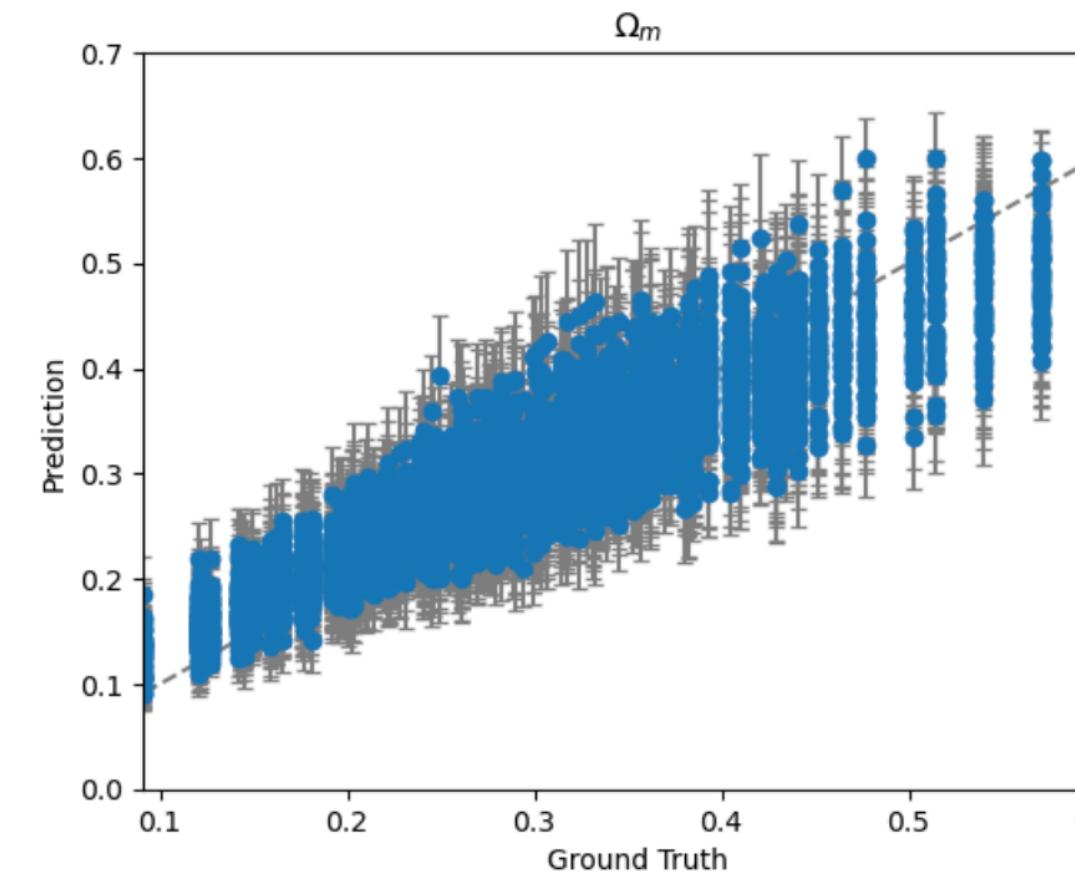
Weak Lensing ML Uncertainty Challenge

Phase 1 Baseline solutions (tested on validation data)

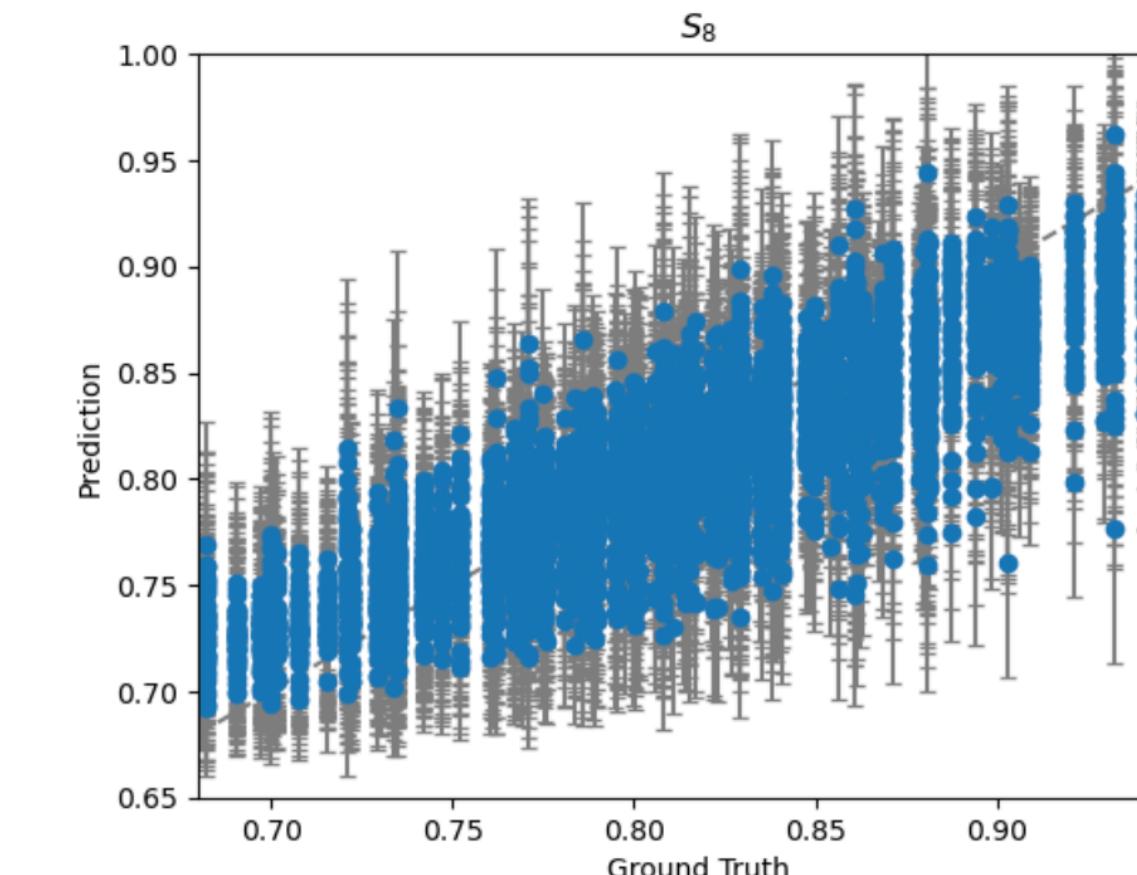
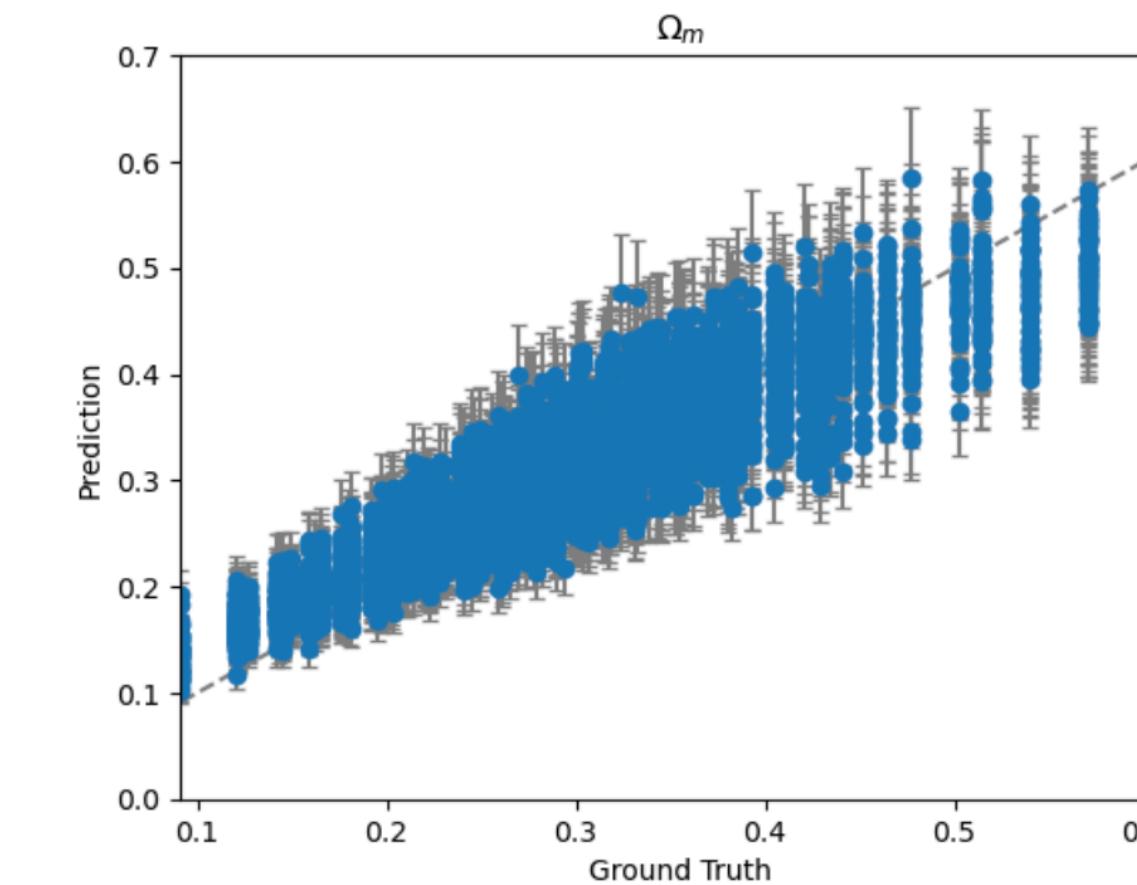
Power spectrum analysis
w/ MCMC



CNN predictions
w/ MCMC



CNN direct predictions



Weak Lensing ML Uncertainty Challenge



Phase 1 Baseline solutions (tested on validation data)

Phase 1 Competition leaderboard (in the final phase of testing)

#	Participant	Date	ID	Method Name	Score	MSE	R^2	Coverage
1	pwchang	2025-08-14 09:52	353450	Baseline_CNN_MCMC_it100k_006_new_data_score CNN predictions w/ MCMC	8.6995	0.1924	0.804	0.7278
2	pwchang	2025-08-14 09:52	353451	Baseline_CNN_direct_new_data_score CNN direct predictions	8.2667	0.1957	0.8006	0.734
3	pwchang	2025-08-14 10:34	353466	Baseline_PSA_MCMC_it100k_002_new_data_score Power spectrum analysis w/ MCMC	4.5766	0.349	0.6446	0.6726

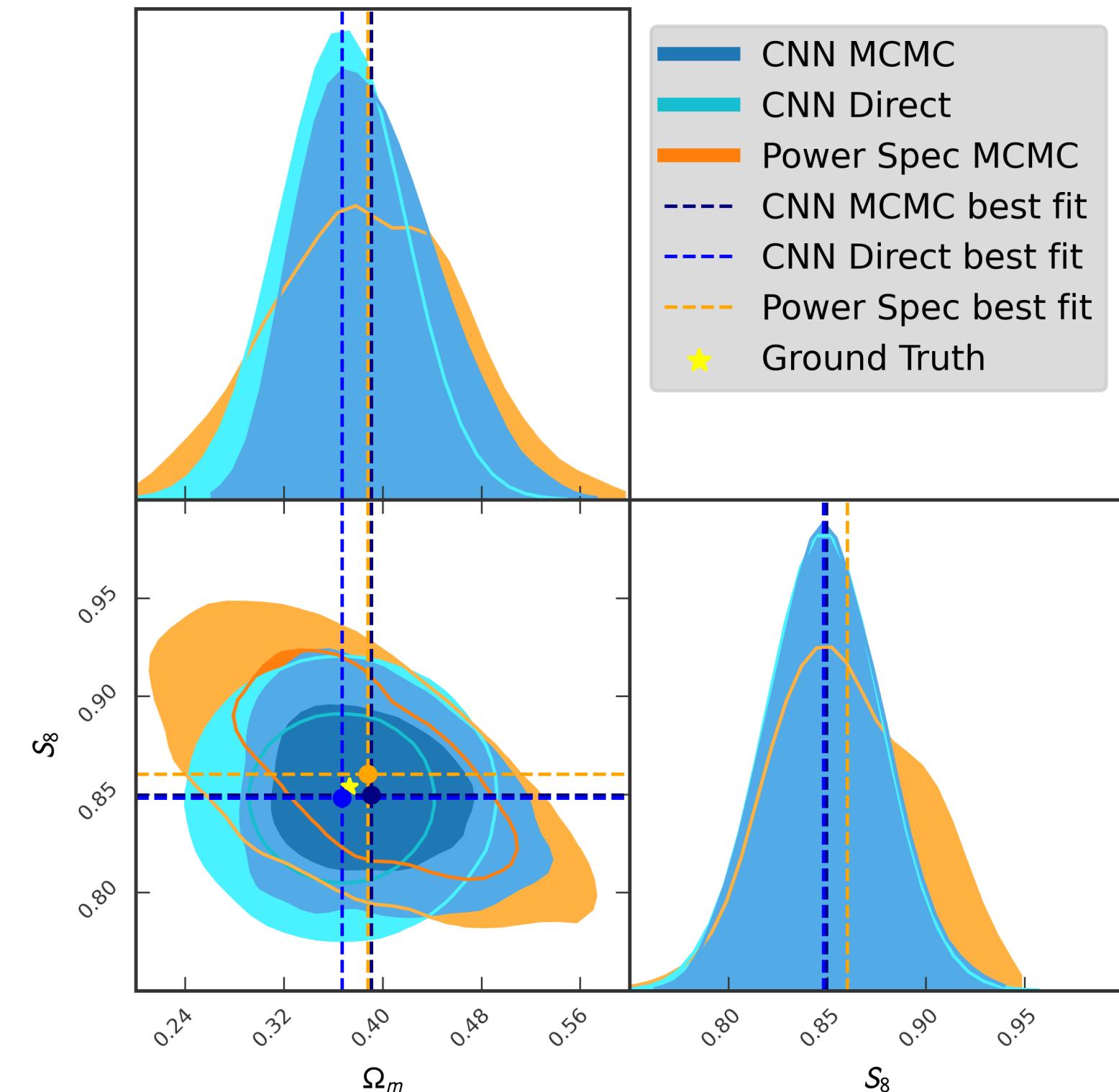
Weak Lensing ML Uncertainty Challenge



Phase 1 Baseline solutions (tested on validation data)

Phase 1 Competition leaderboard (in the final phase of testing)

#	Participant	Date	ID	Method Name	Score
1	pwchang	2025-08-14 09:52	353450	Baseline_CNN_MCMC_it100k_006_new_data_score CNN predictions w/ MCMC	8.6995
2	pwchang	2025-08-14 09:52	353451	Baseline_CNN_direct_new_data_score CNN direct predictions	8.2667
3	pwchang	2025-08-14 10:34	353466	Baseline_PSA_MCMC_it100k_002_new_data_score Power spectrum analysis w/ MCMC	4.5766



A **starting kit** of the power spectrum analysis baseline solution for the Phase 1 competition is now available on our [GitHub repository](#) or [Google Colab](#)

Weak Lensing ML Uncertainty Challenge



Prizes

The Phase 1 and Phase 2 of the competition will have separate leaderboards.

We will award the monetary prizes to the top performers in either phases of the competition.

For the **Phase 1**, the top three performers will be awarded the following prizes:

- 1 First Place: \$2,000
- 2 Second Place: \$1500
- 3 Third Place: \$500

We will announce the prize for the **Phase 2** soon.

In addition to monetary prizes, we will invite the participants who come up with top solutions to present at our competition workshop in NeurIPS 2025.

Weak Lensing ML Uncertainty Challenge



Status and timeline

- This challenge is one of the **NeurIPS 2025 competitions**.

Envisioned competition schedule (UTC)		
Competition Phase	Date	Description
Phase 1	Aug 27th 2025 – Nov 16th 2025	Open submissions
	Nov 17th 2025 – Nov 30th 2025	Evaluating top submissions on hidden dataset
	Dec 1st 2025	Announcement of winners
Phase 2	End Sep – Early Jan 2026	Open submissions
	Early Jan 2026 – Mid Jan 2026	Evaluating top submissions on hidden dataset
	Mid Jan 2026	Announcement of winners
NeurIPS 2025	Dec 7th 2025	Winners will be invited to our workshop

[Join the competition today from HERE!](#)



U.S. DEPARTMENT
of ENERGY

BERKELEY LAB



IAS | INSTITUTE FOR
ADVANCED STUDY

Berkeley
UNIVERSITY OF CALIFORNIA

Stanford
University

université
PARIS-SACLAY

SLAC NATIONAL
ACCELERATOR
LABORATORY

W
UNIVERSITY of WASHINGTON

UCI University of
California, Irvine

Weak Lensing ML Uncertainty Challenge



Longterm benchmark

UCI ML Datasets

A screenshot of the UC Irvine Machine Learning Repository website. The interface includes a navigation bar with links for Datasets, Contribute Dataset, and About Us, and a search bar labeled "Search datasets...". On the left, there is a "Filters" sidebar with sections for "Keywords", "Attributes", "Data Type" (which is selected), and "Task" (which is also selected). Under "Data Type", options like "Image" and "Multivariate" are shown with checkboxes. Under "Task", "Classification" and "Regression" are shown with checkboxes. The main area is titled "Browse Datasets" and features buttons for "SORT BY # VIEWS, DESC" and "EXPAND ALL". At the bottom right, there are pagination controls for "Rows per page" (set to 5) and "0 to 0 of 0".

This could be the first benchmark dataset for the image regression task for a well-established scientific problem with systematic effects