ETH zürich
Institute for Particle Physics
and Astrophysics

Multi-Probe Large-Scale Structure Cosmology with Simulation-Based Inference and Deep Learning

Jozef Bucko
Institute for Particle Physics and Astrophysics
ETH Zurich


Arne Thomsen

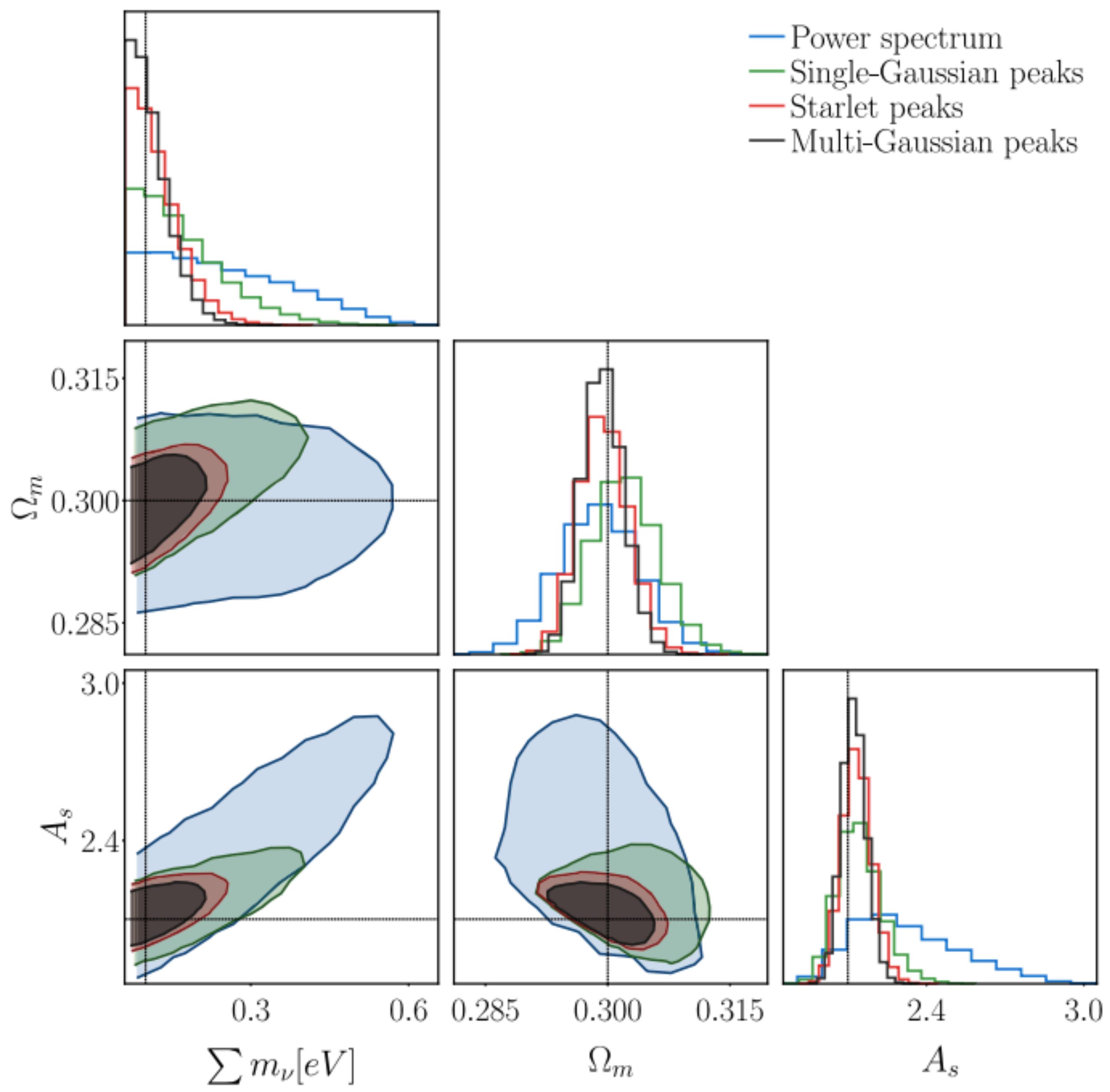

Tomasz Kacprzak


Virginia Ajani

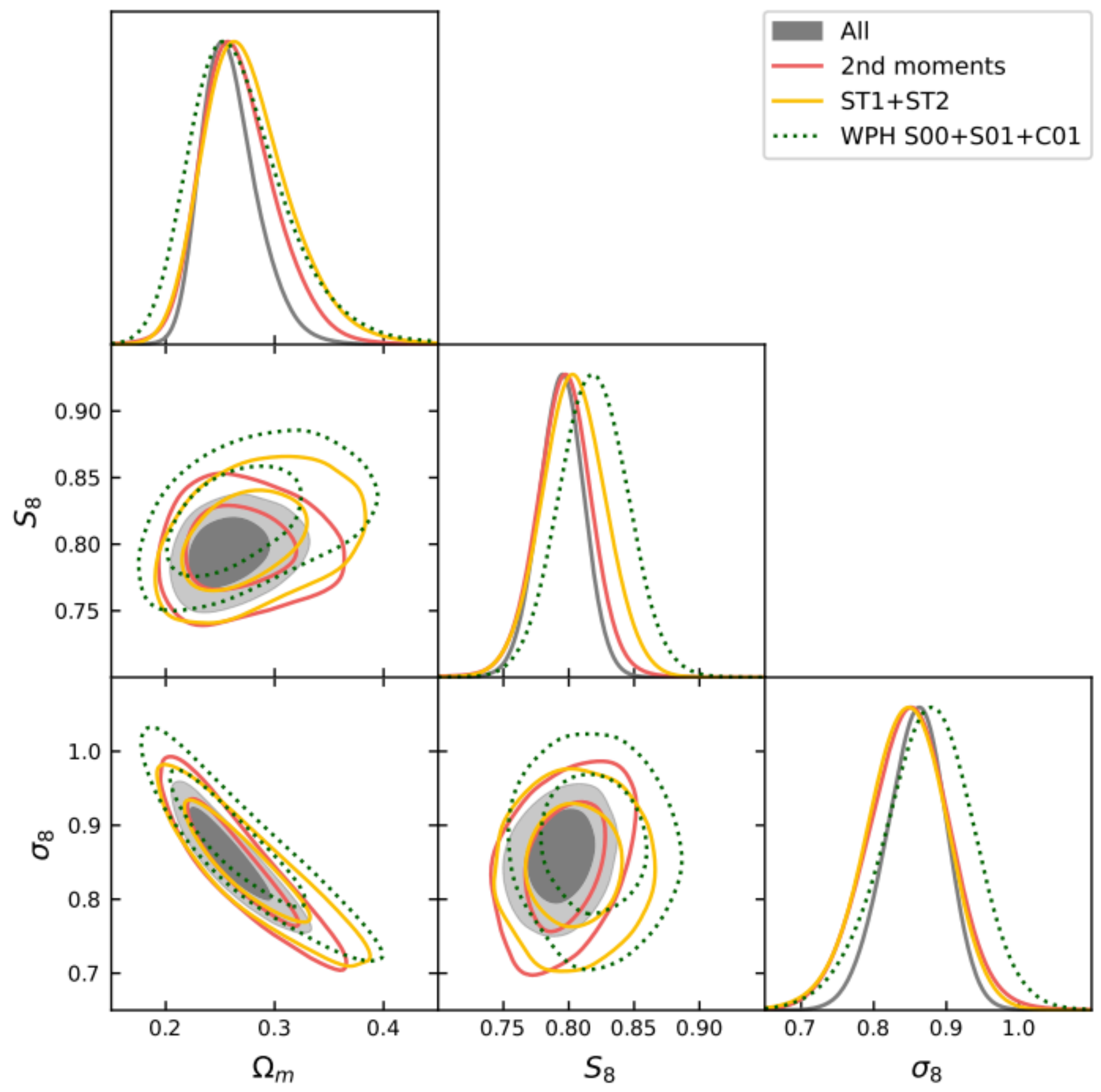

Alexandre Refregier

Motivation

- Non-Gaussian statistics is a powerful probe of LSS (Ajani et al. 2020, Kratochvil et al. 2012,...)
- Theoretical modeling is very challenging, extracting from simulations much less so
  - ➡ but one needs a solid simulation base
  - ➡ CosmoGrid simulation suite (Kacprzak et al. 2022a)



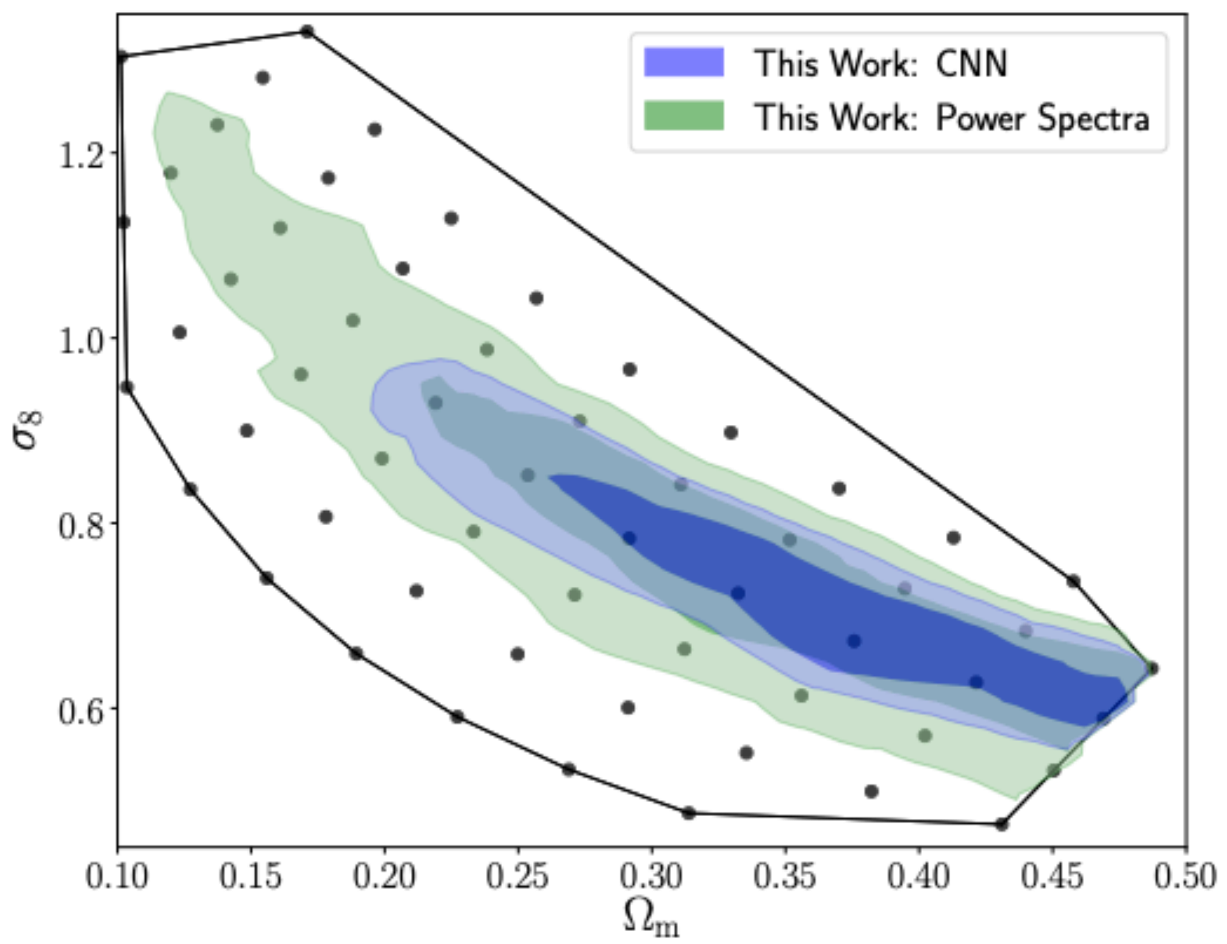
Ajani et al. 2020, stage IV-like, weak lensing



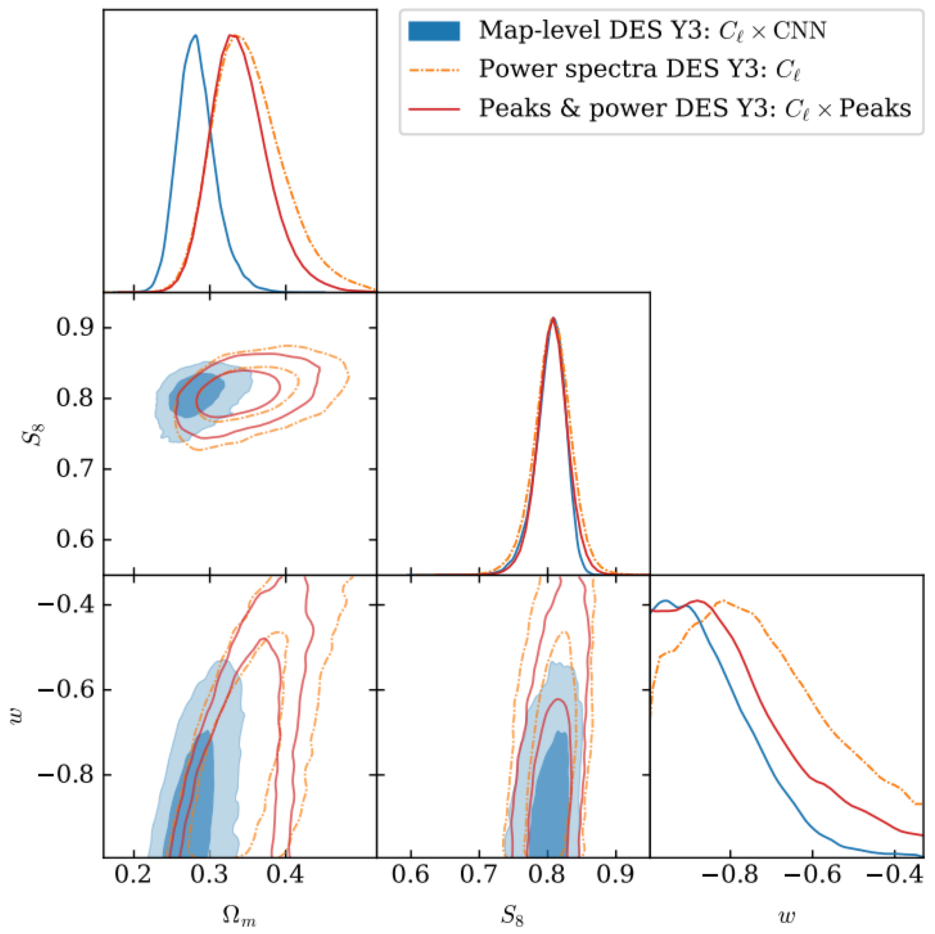
DES; Gatti et al. 2024, DES Y3, weak lensing

Motivation

➡ Non-Gaussian information can also be extracted using Deep Learning, aiming for minimal information loss



Fluri et al. 2019, KiDS-1000, weak lensing

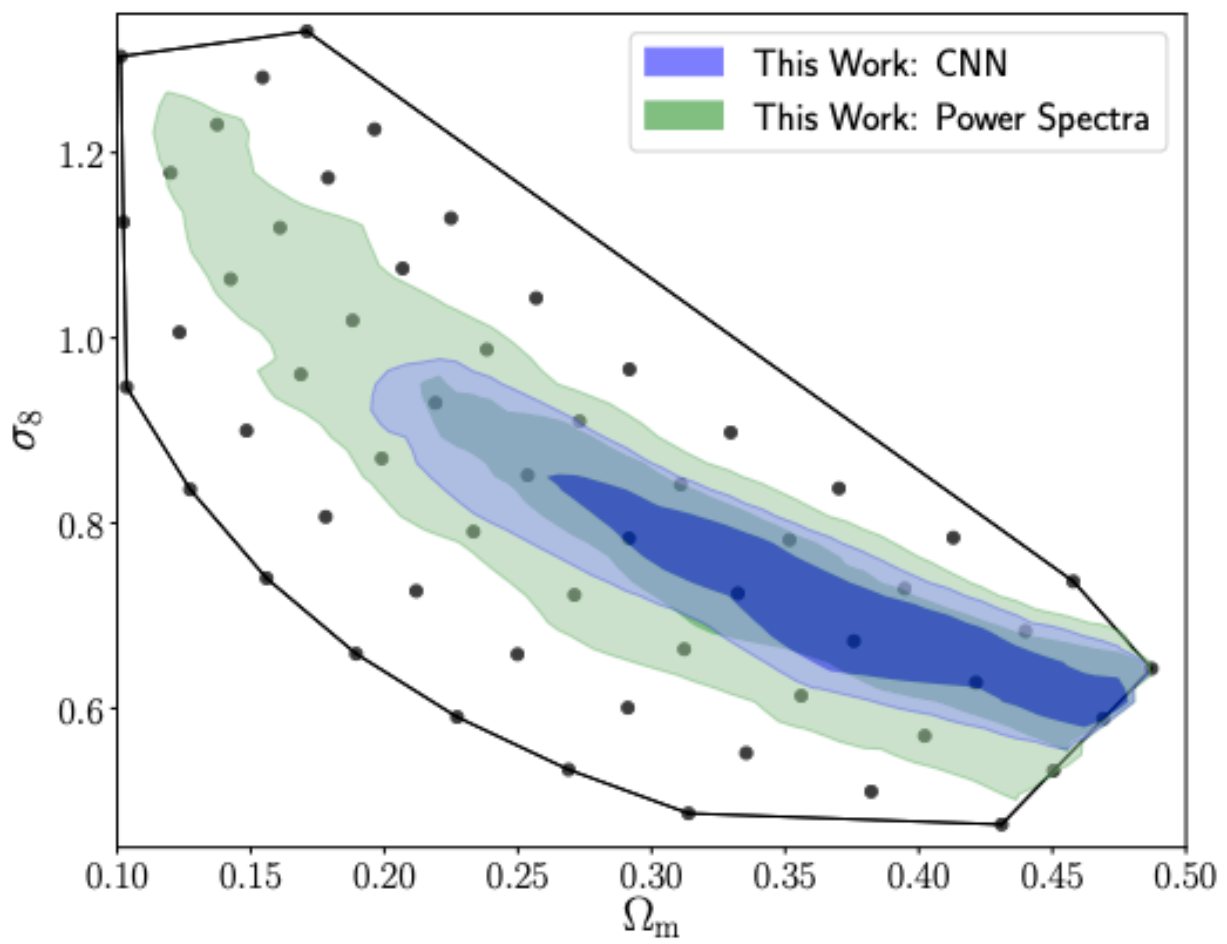


Jeffrey et al. 2024, DES Y3, weak lensing

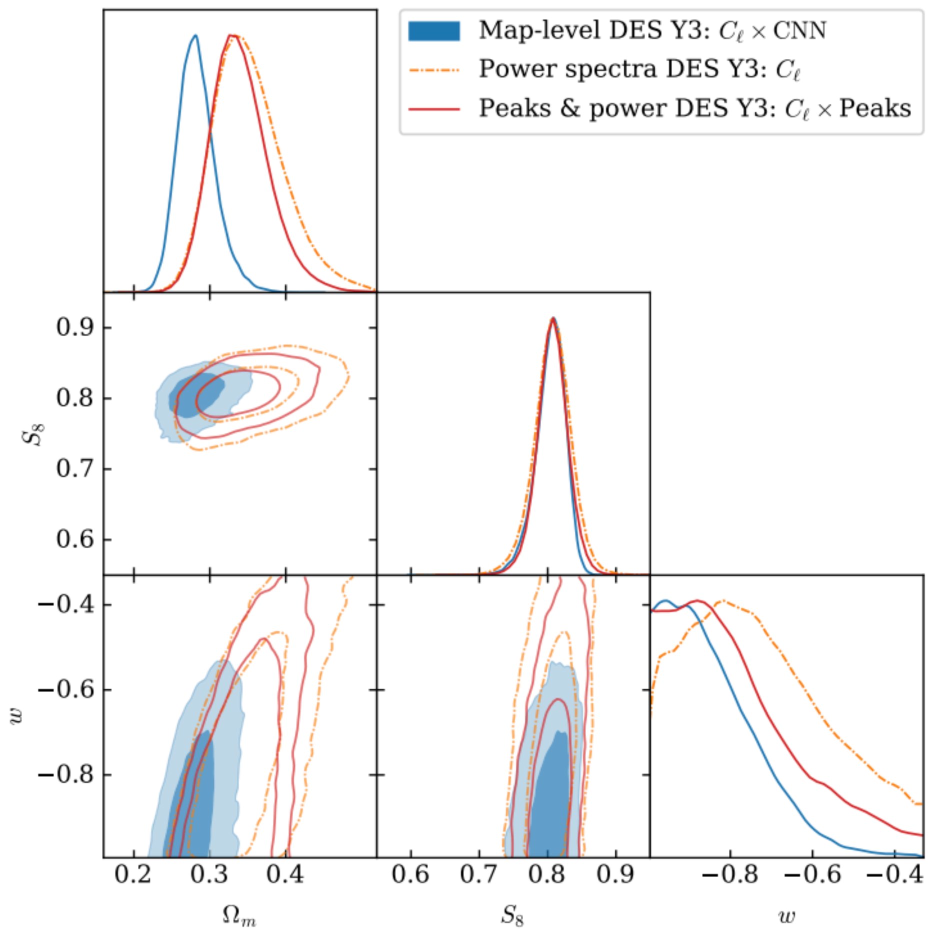
➡ Combine different probes to get more information

Motivation

➡ Non-Gaussian information can also be extracted using Deep Learning, aiming for minimal information loss

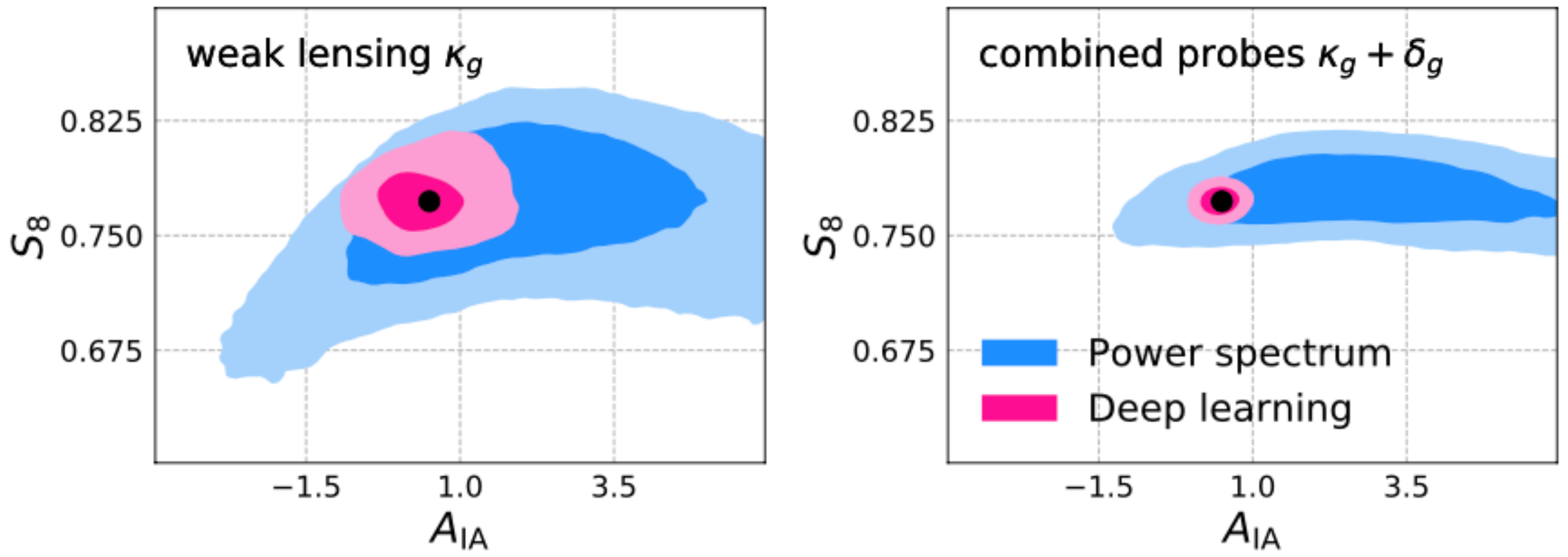


Fluri et al. 2019, KiDS-1000, weak lensing



Jeffrey et al. 2024, DES Y3, weak lensing

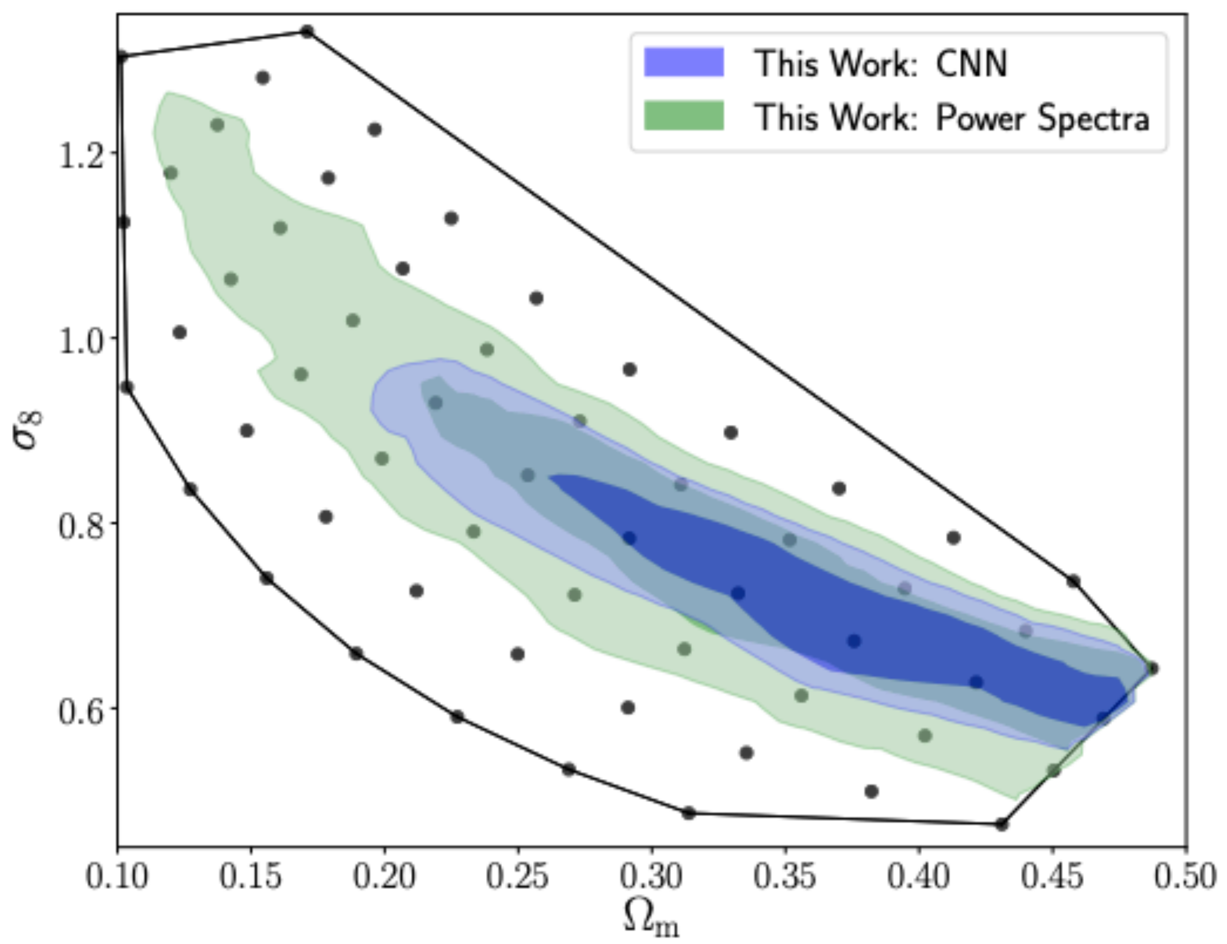
➡ Combine different probes to get more information



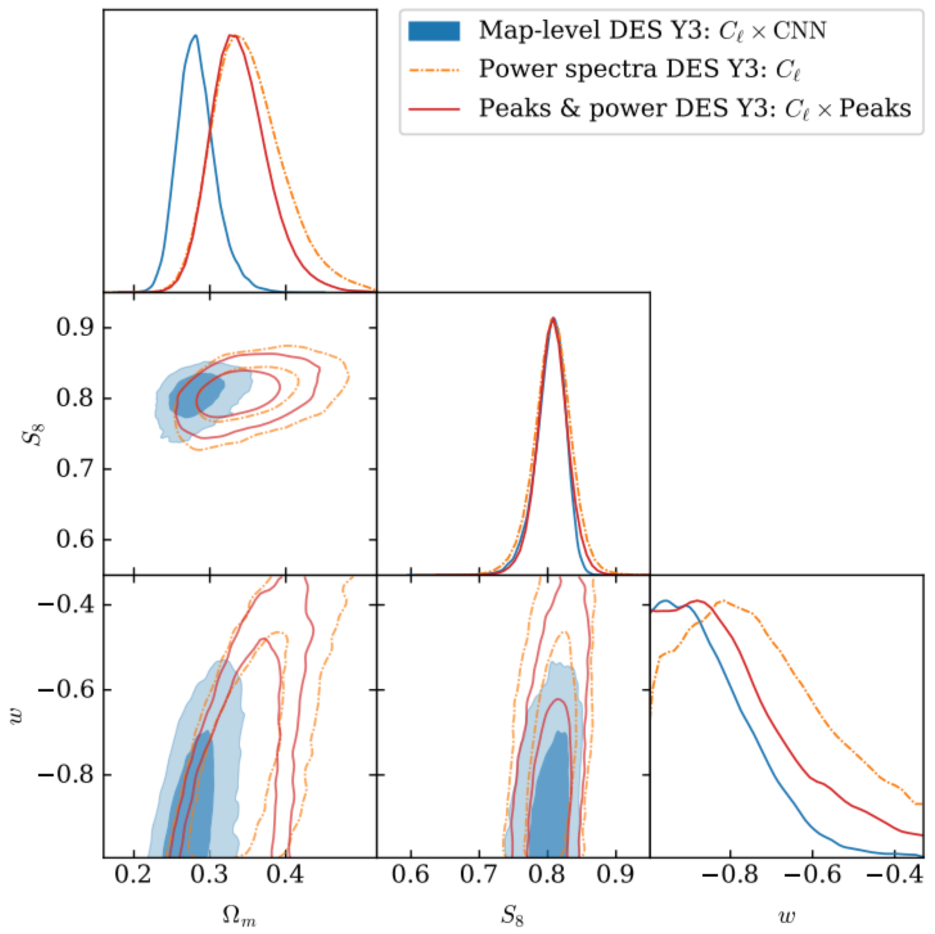
Kacprzak et al. 2022b, stage III-like, weak lensing+galaxy clustering

Motivation

➡ Non-Gaussian information can also be extracted using Deep Learning, aiming for minimal information loss

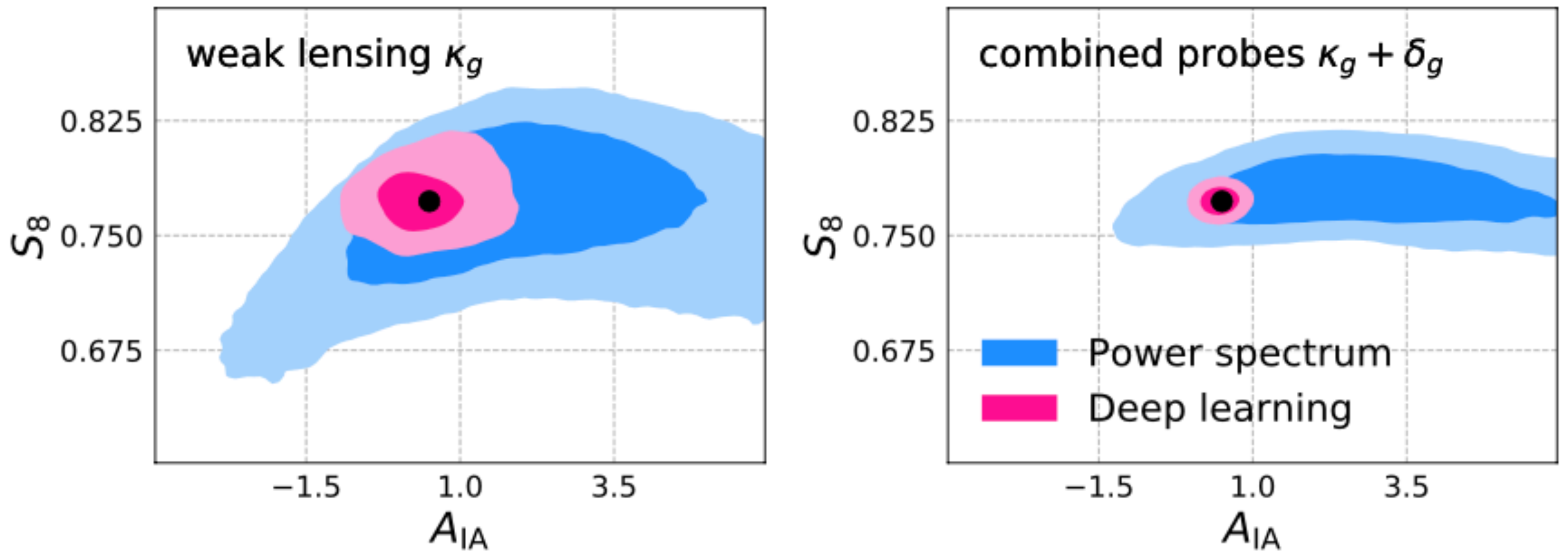


Fluri et al. 2019, KiDS-1000, weak lensing



Jeffrey et al. 2024, DES Y3, weak lensing

➡ Combine different probes to get more information



Kacprzak et al. 2022b, stage III-like, weak lensing+galaxy clustering

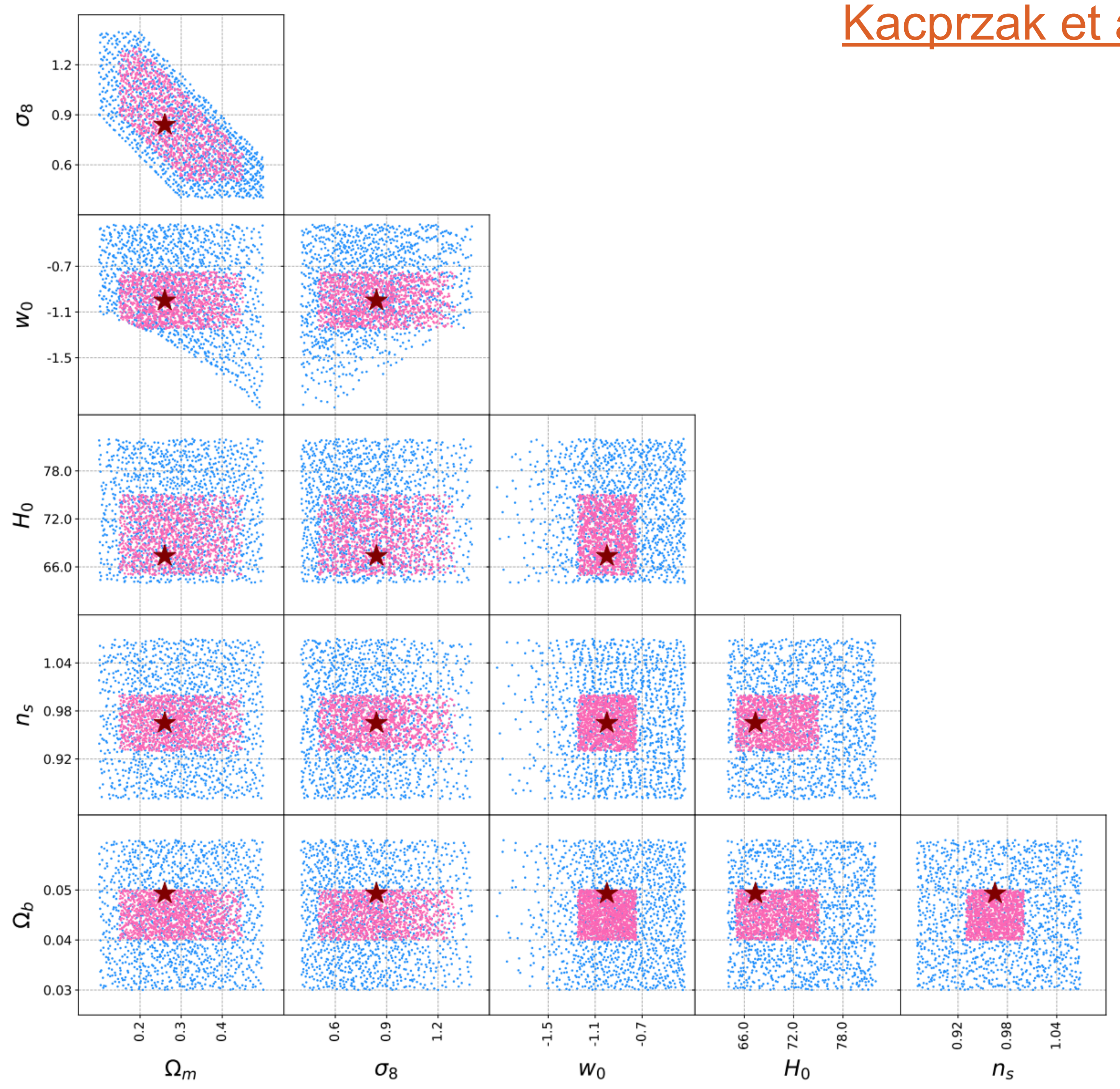
weak lensing + galaxy clustering SBI with peak statistics and deep learning summaries
(DES projects)

CosmoGrid

The CosmoGridV1 dataset consists of a total of 20128 simulations divided into three main parts:

- **grid**: a set of 2500 cosmologies, each with 7 simulations from unique initial conditions (a total of 17500 N-body runs),
- **fiducial**: simulations and the fiducial cosmology and its $\pm\Delta$ derivatives, with 200 unique initial conditions (2600 runs),
- **benchmark**: simulation benchmarks used for systematics testing of features chosen for parameter inference (28 runs).

Kacprzak et al. 2022a

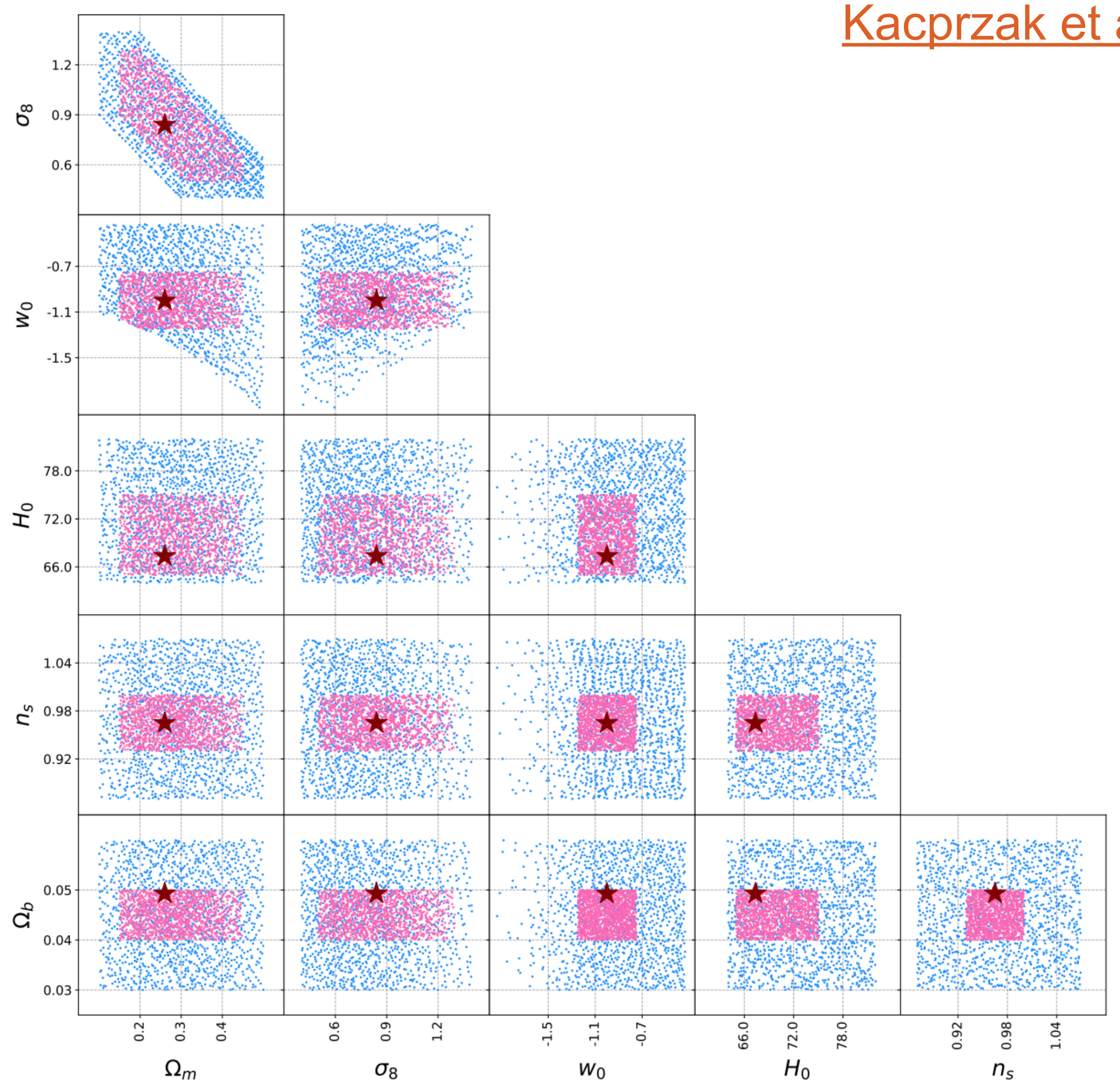


CosmoGrid

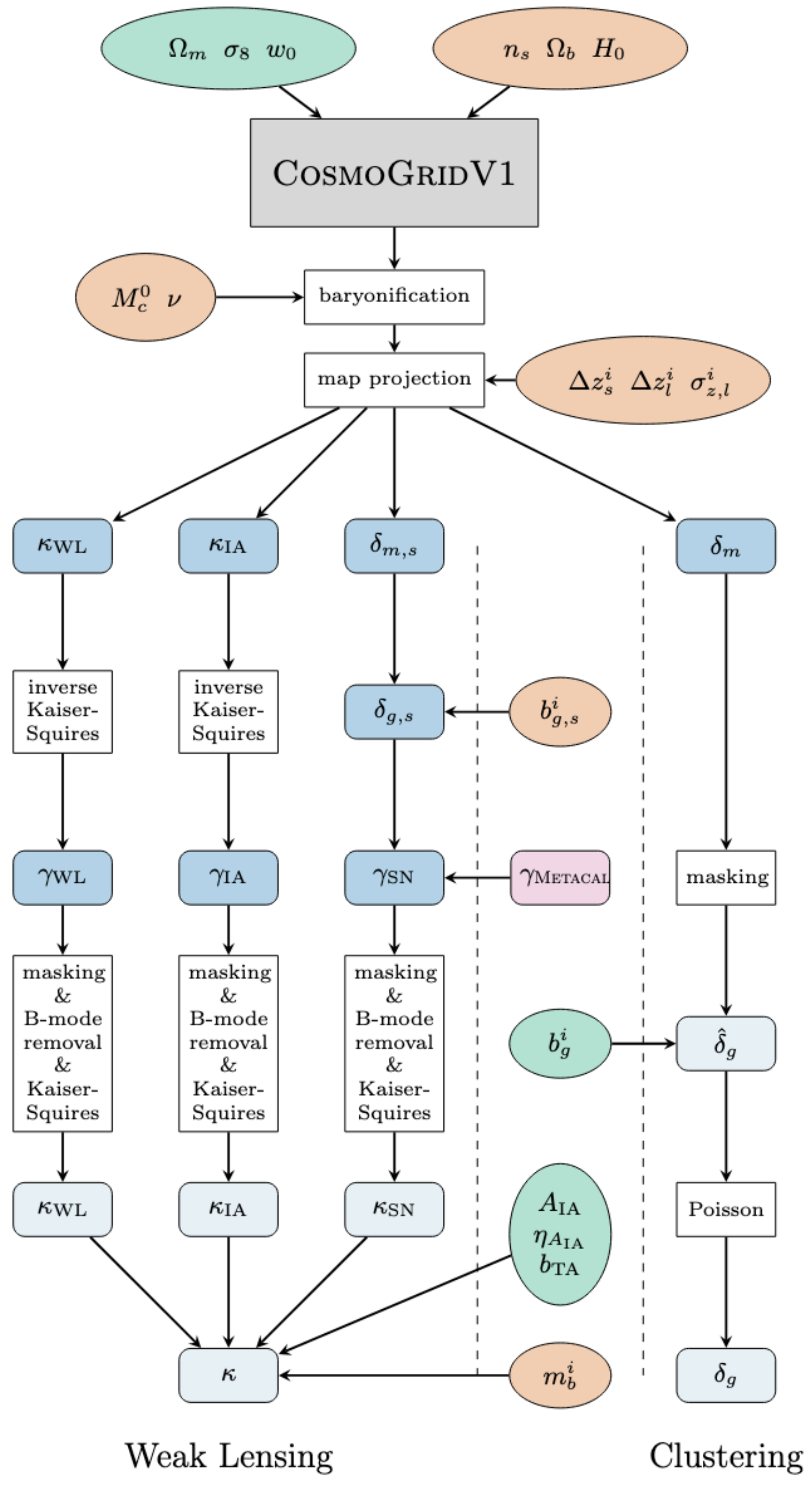
The CosmoGridV1 dataset consists of a total of 20128 simulations divided into three main parts:

- **grid**: a set of 2500 cosmologies, each with 7 simulations from unique initial conditions (a total of 17500 N-body runs),
- **fiducial**: simulations and the fiducial cosmology and its $\pm\Delta$ derivatives, with 200 unique initial conditions (2600 runs),
- **benchmark**: simulation benchmarks used for systematics testing of features chosen for parameter inference (28 runs).

Kacprzak et al. 2022a



Forward model



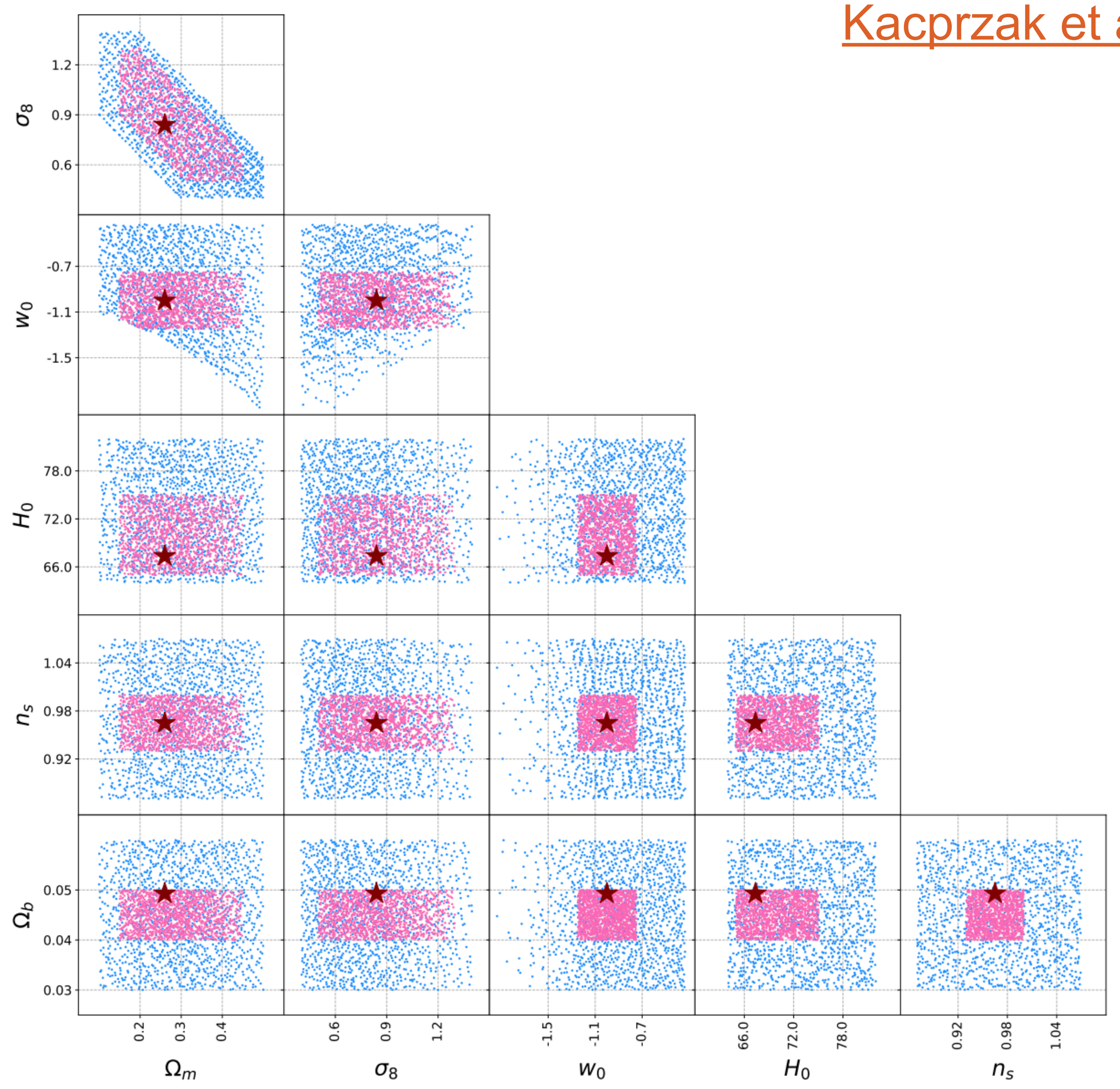
Thomsen, JB et al. (in prep)

CosmoGrid

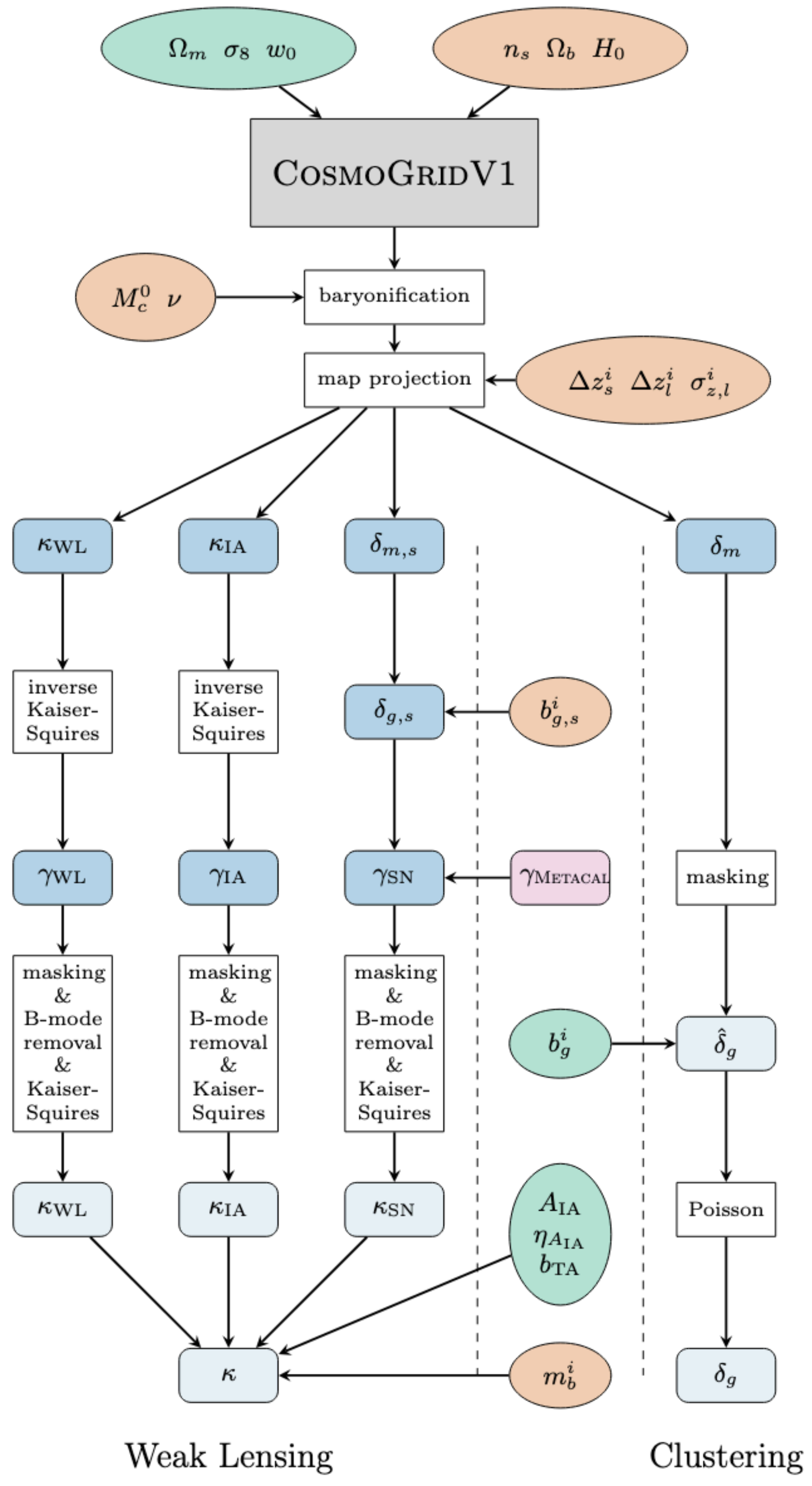
The CosmoGridV1 dataset consists of a total of 20128 simulations divided into three main parts:

- **grid**: a set of 2500 cosmologies, each with 7 simulations from unique initial conditions (a total of 17500 N-body runs),
- **fiducial**: simulations and the fiducial cosmology and its $\pm\Delta$ derivatives, with 200 unique initial conditions (2600 runs),
- **benchmark**: simulation benchmarks used for systematics testing of features chosen for parameter inference (28 runs).

Kacprzak et al. 2022a

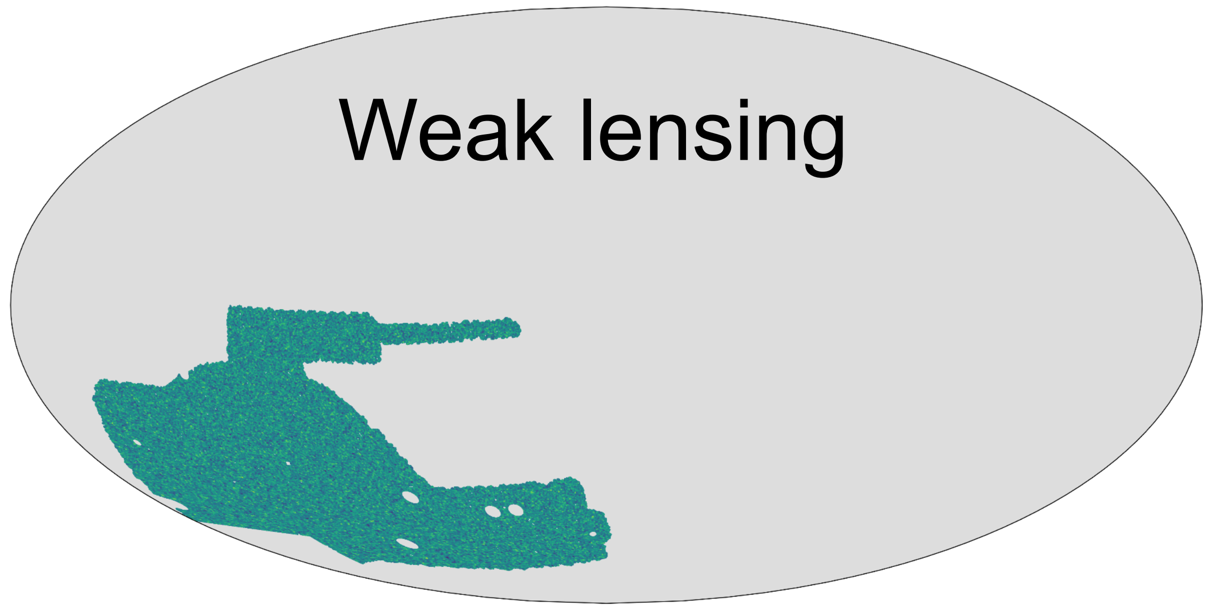


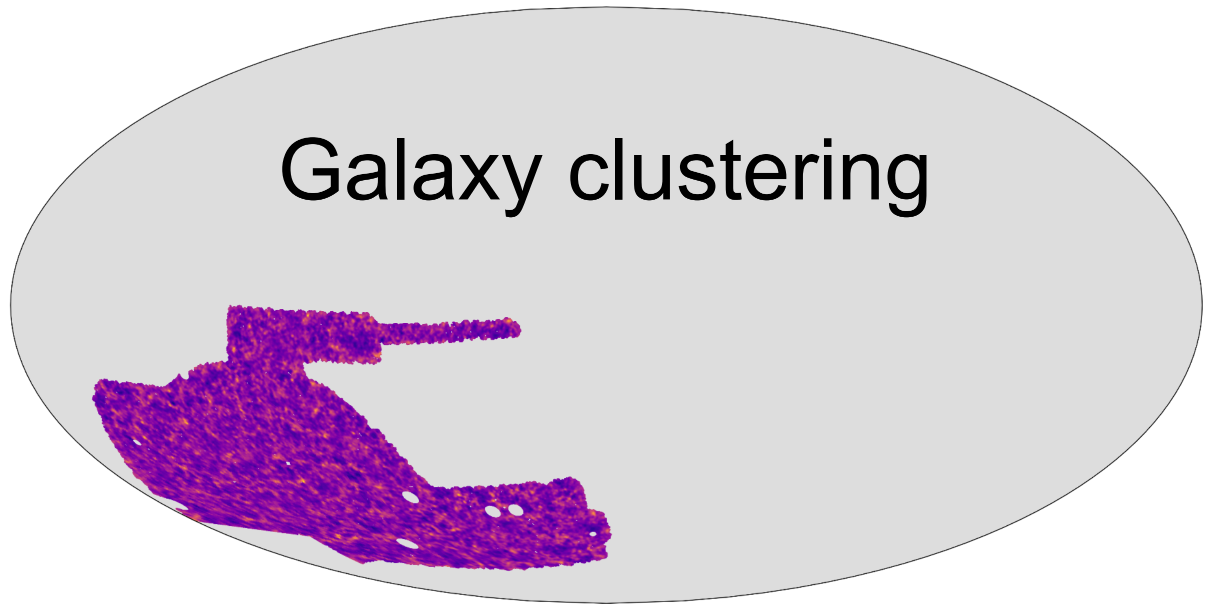
Forward model



Thomsen, JB et al. (in prep)

Synthetic DES Y3 data





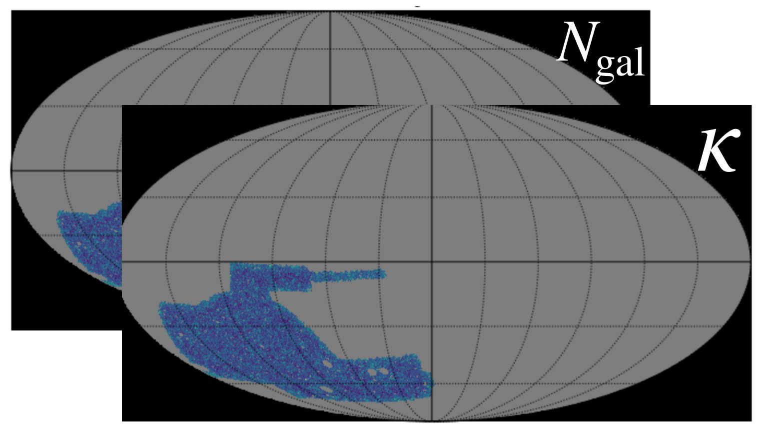
→ 1’000’000 independent survey maps

Pipeline

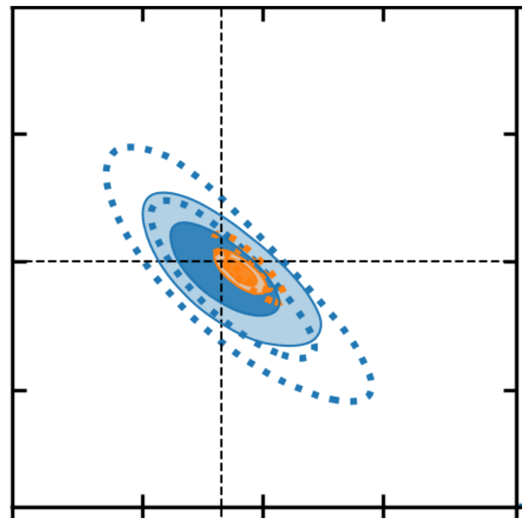
Synthetic DES Y3 data

Forward model

CosmoGrid → Simulated maps

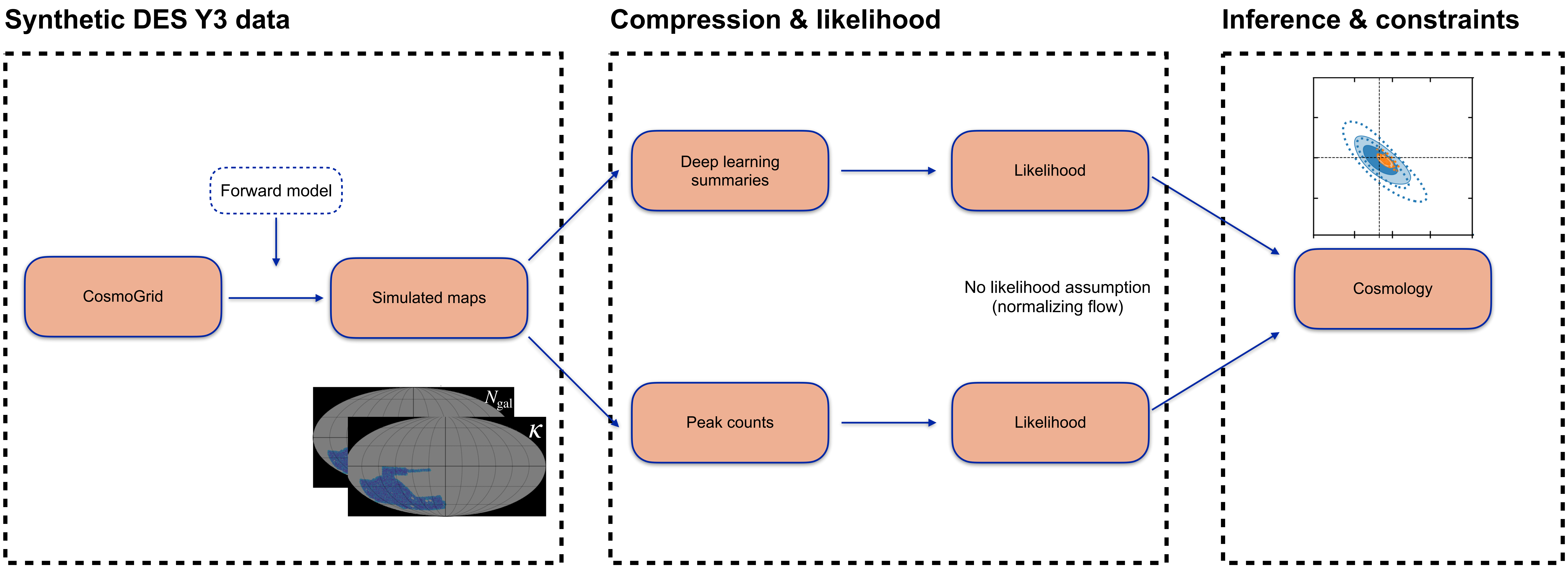


Inference & constraints



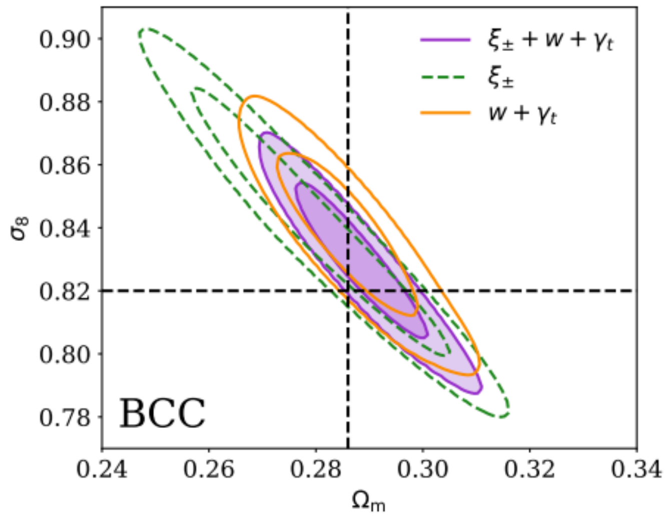
Cosmology

Pipeline


Synthetic DES Y3 data
Compression & likelihood
Inference & constraints
Forward model
CosmoGrid
Simulated maps
$N_{\rm gal}$
κ
Deep learning summaries
Peak counts
Likelihood
Likelihood
No likelihood assumption
(normalizing flow)
Cosmology

Towards real data

DES Collaboration: MacCrann et al. 2018



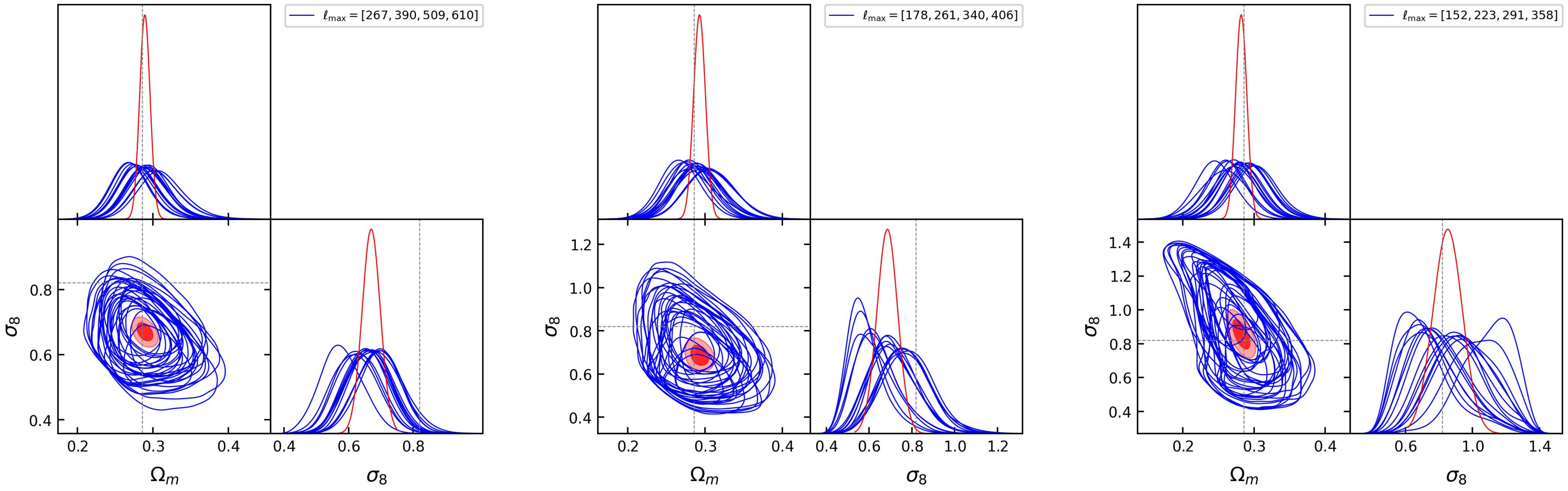
Validation:

1. Internal data (consistency): cosmology from CosmoGrid
   → no model mismatch

2. Mock DESY3-like data: cosmology from external simulations
   a. Buzzard (multiple realizations, De Rose et al. 2019)
   b. Cardinal (To et al. 2023)
   c. MICE (Fosalba et al. 2013)

3. tests of other systematics (source clustering, resolution,…)

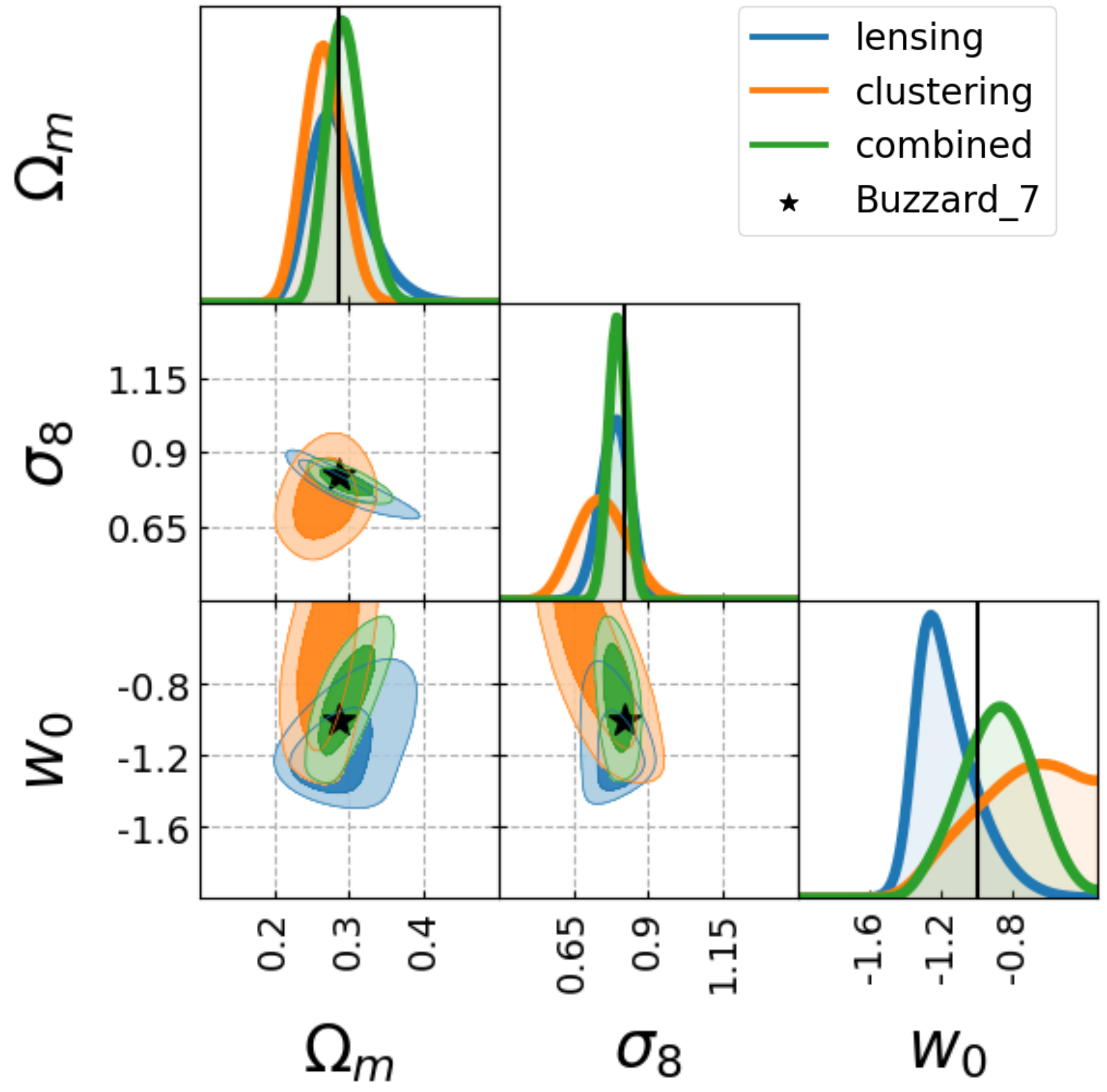
Smoother maps = less small-scale information



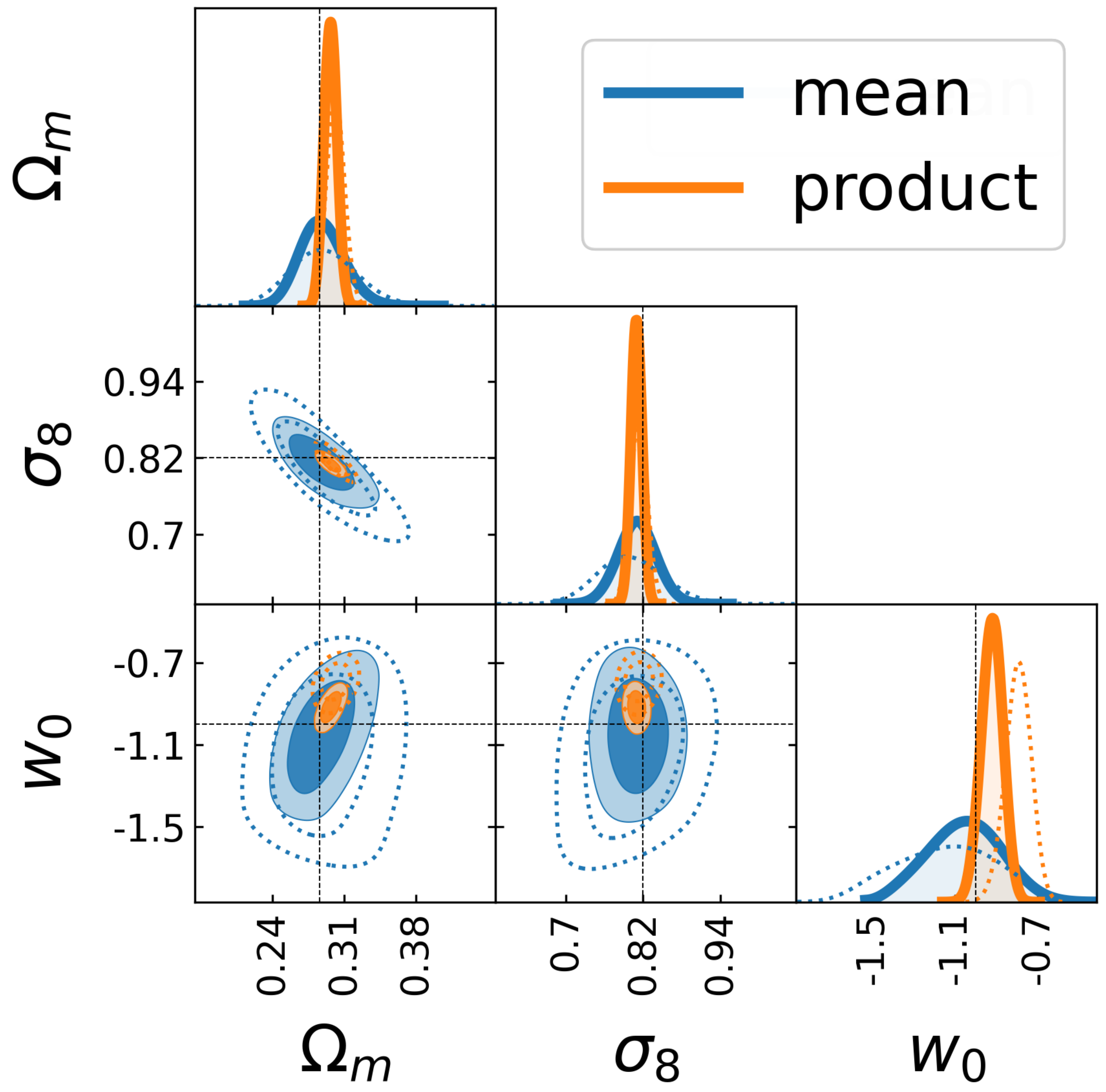
Bucko et al. (in prep)

Towards real data

single-probe vs. probe combination



2-pt vs. deep learning (probe combination)



Conclusion

- Simulation-based inference of Cosmology
  - CosmoGridV1 N-body suite
  - Forward modeling of DES Y3 mocks
  - No likelihood assumption

- From multi-probe maps
  - Weak lensing
  - Galaxy clustering

- Non-Gaussian information
  - Peaks
  - Learned summary statistics

- Using deep learning
  - Learned summary statistics
  - Likelihood as normalizing flow