Pixel-level analysis of weak lensing maps:

an optimal inference approach

Natalia Porqueres

with Alan Heavens, Lucas Makinen, Guilhem Lavaux, Seung-gyu Hwang, Josh Williamson, Daniel Mortlock, Anya Paopiamsap, Jordy Ram

CIEMAT – 3/9/2026

Outline

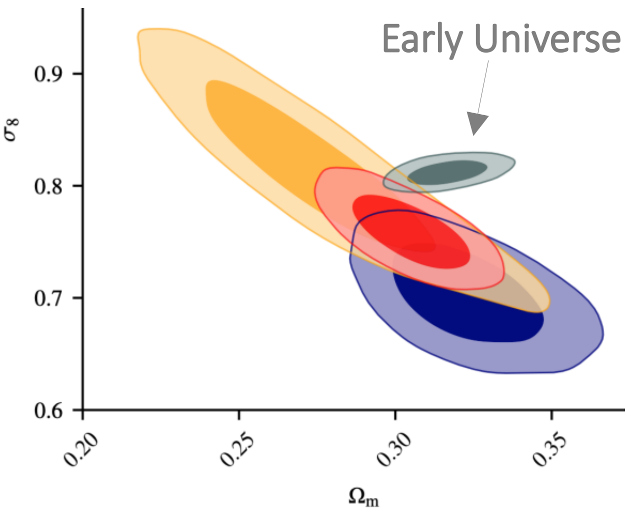
The method — A field-based approach of weak lensing

The proof of concept — Results with synthetic data

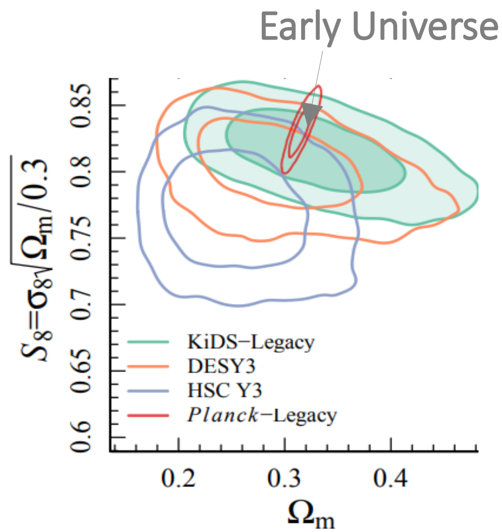
The challenge — Getting ready for real data

Cosmology now

Cosmology tension seems resolved

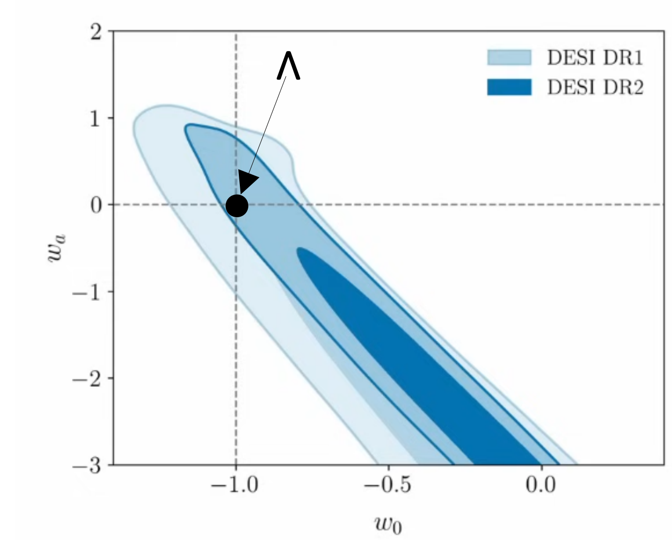


Heymans et al. (2020)



Wright et al. (2025)

What is dark energy?



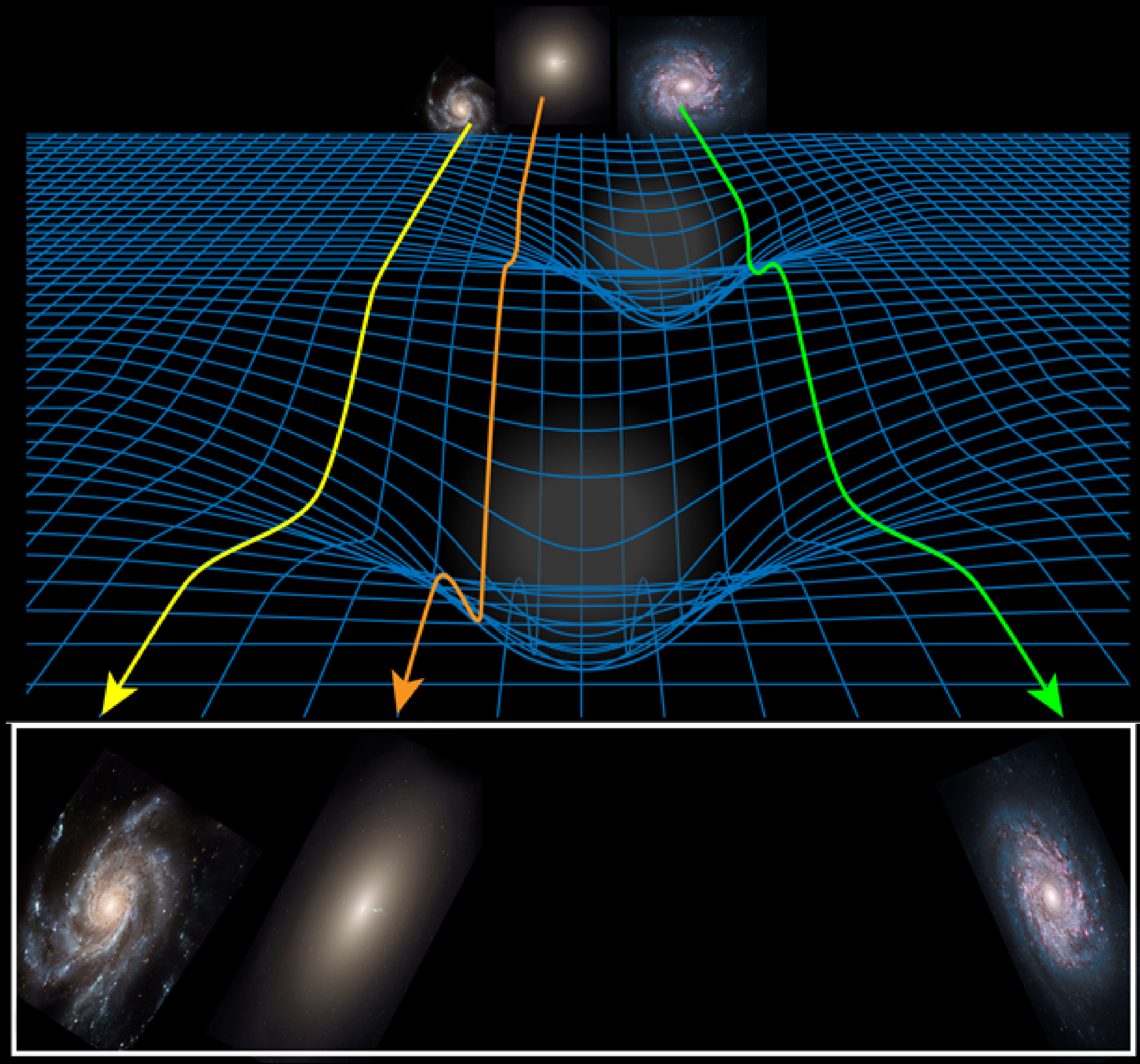
DESI collaboration (2025)

Upcoming datasets with unprecedented precision

Euclid consortium

VERA C. RUBIN OBSERVATORY

Weak gravitational lensing

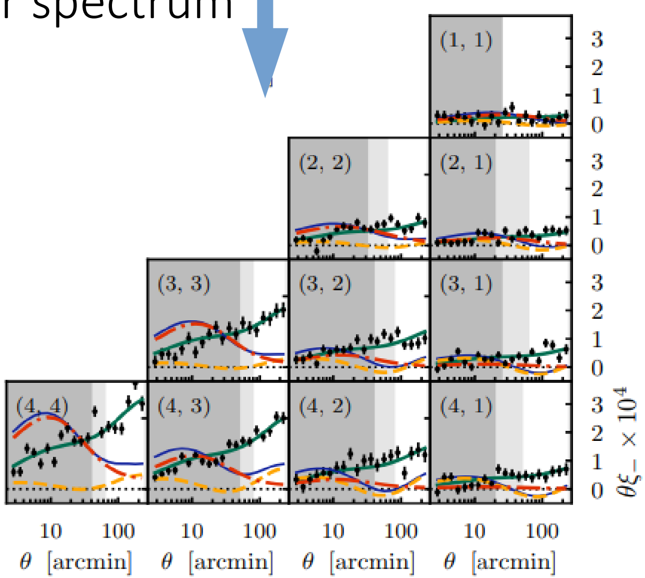


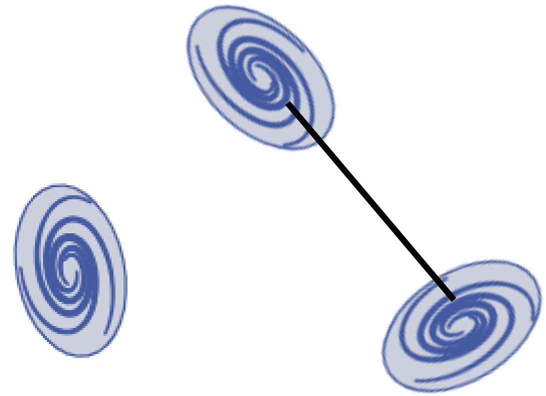
Credit: Stonebraker

Natalia Porqueres

The standard analysis of weak lensing

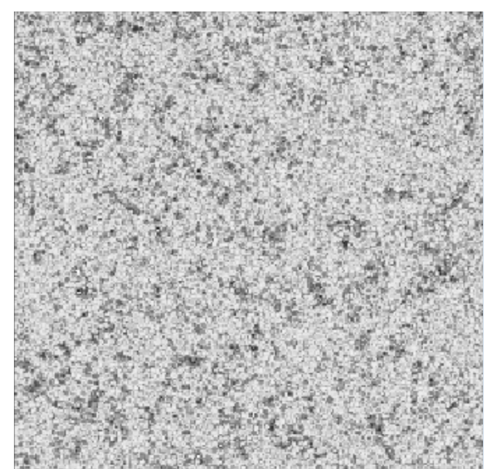
Compression to power spectrum





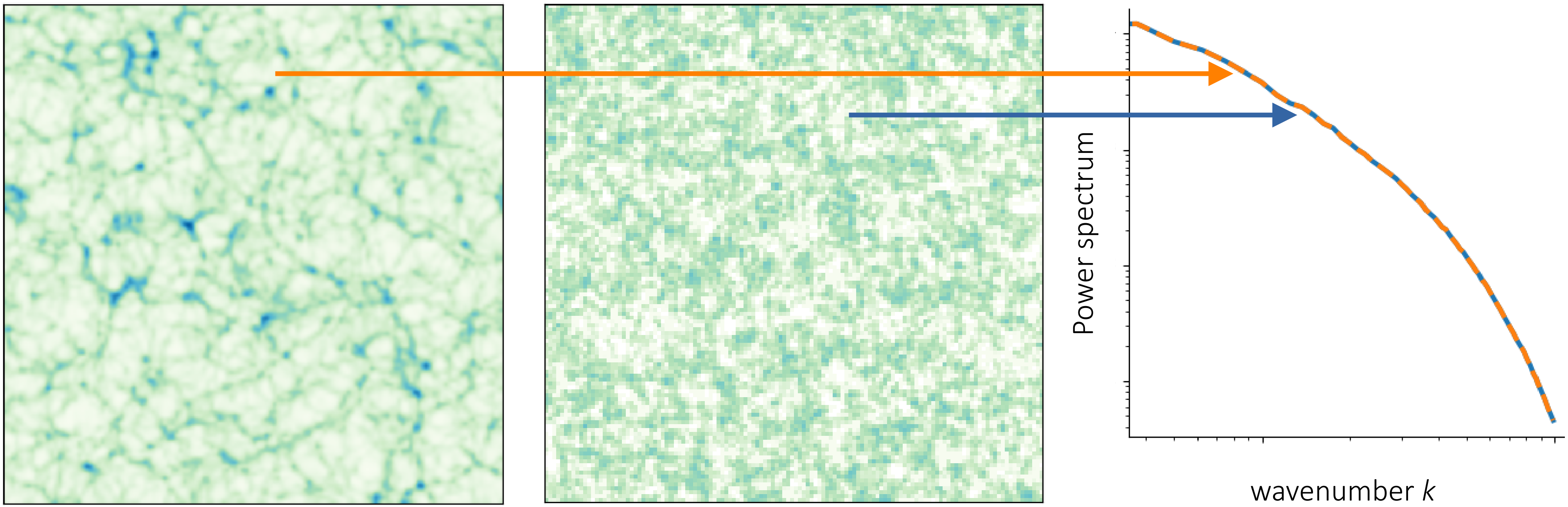
Correlations between pairs of galaxy shapes

Power spectrum is enough for Gaussian data



Natalia Porqueres

Is the standard approach enough?



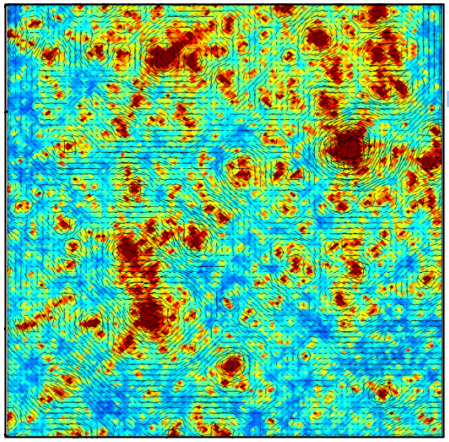
Our Universe

Gaussian random field

The standard approach misses the non-Gaussian information

We need new techniques to analyse the data

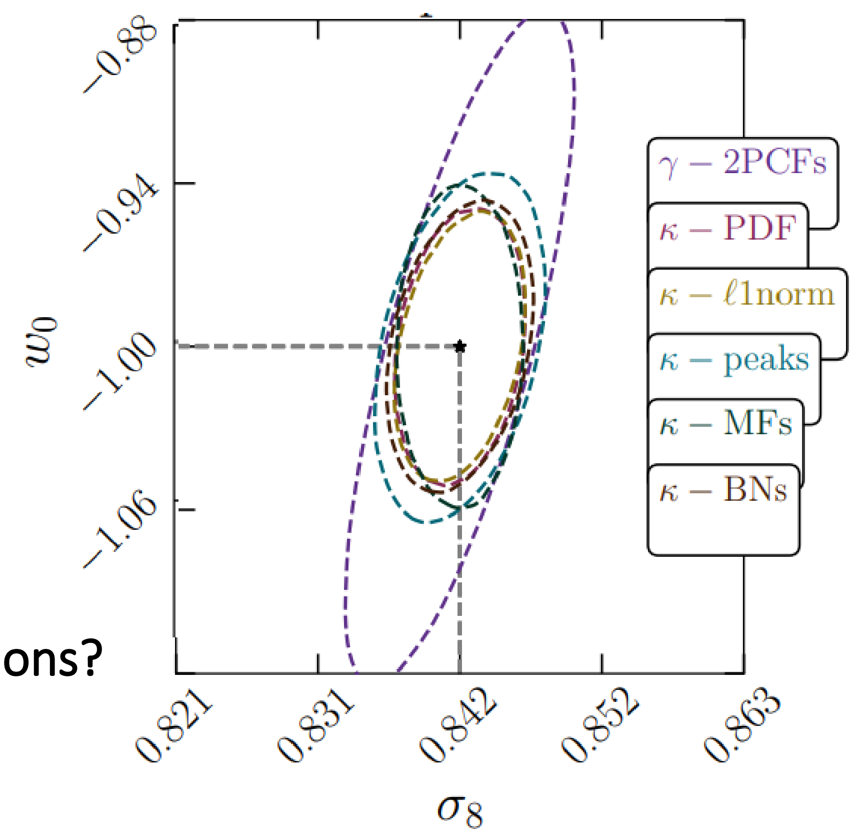
Extracting more information from the data



Data compression

High-order statistics (*Euclid* HOWLS team)

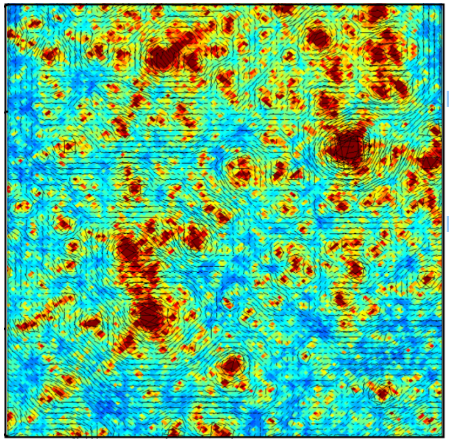
Assume distributions



Challenge: sampling distributions?

Euclid Collaboration, Vinciguerra et al. (2025)

Extracting more information from the data



Data compression

High-order statistics (*Euclid* HOWLS team)

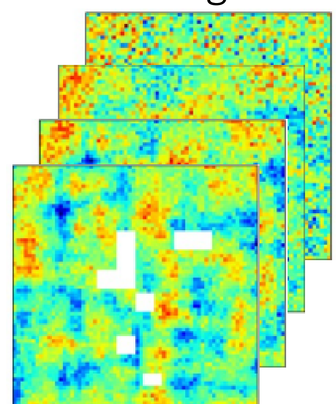
Assume distributions

Data compression

Simulation-based inference (neural network stats)

Learn distributions

Predicted lensing fields



Forward model (simulations)

Cosmological parameters

data

Data compression

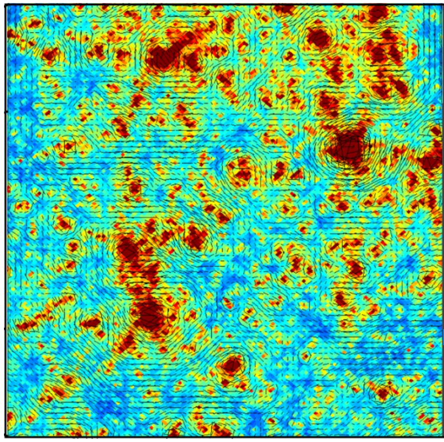
Data compression

Summary statistics

Summary statistics

Challenge: Optimal (extreme) data compression?

Extracting more information from the data

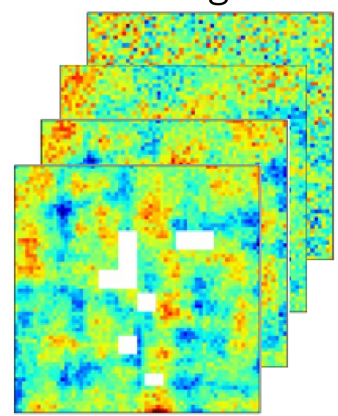


Data compression → High-order statistics (*Euclid* HOWLS team)
Assume distributions

Data compression → Simulation-based inference (neural network stats)
Learn distributions

→ **Field-level inference** (with a likelihood)
Write in terms of known distributions (BHM)

Cosmological parameters → Forward model (simulations) → Predicted lensing fields

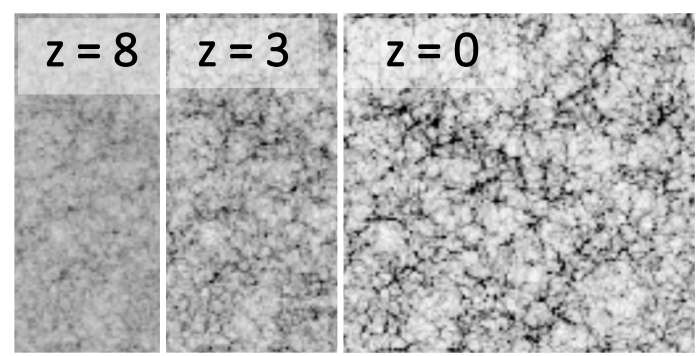


Likelihood at field level ⇄ data

Captures all the information

Easier to separate cosmology and systematics

Provides a digital copy of the Universe

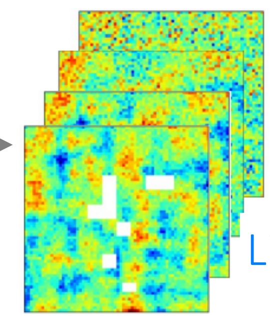


Field-level inference

Cosmological parameters

gravity simulations

Predicted observations

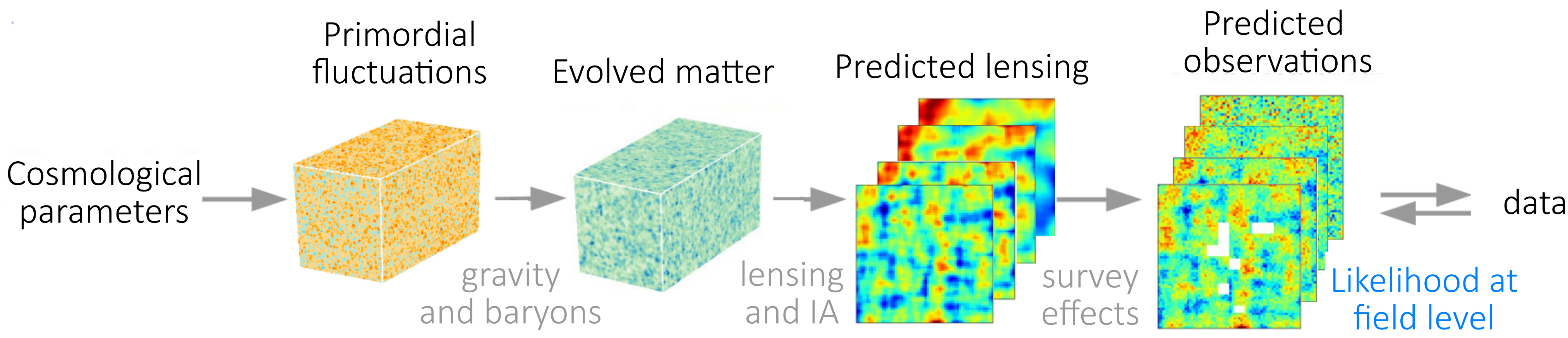


data

Likelihood at field level

Natalia Porqueres

Field-level inference: BORG-WL



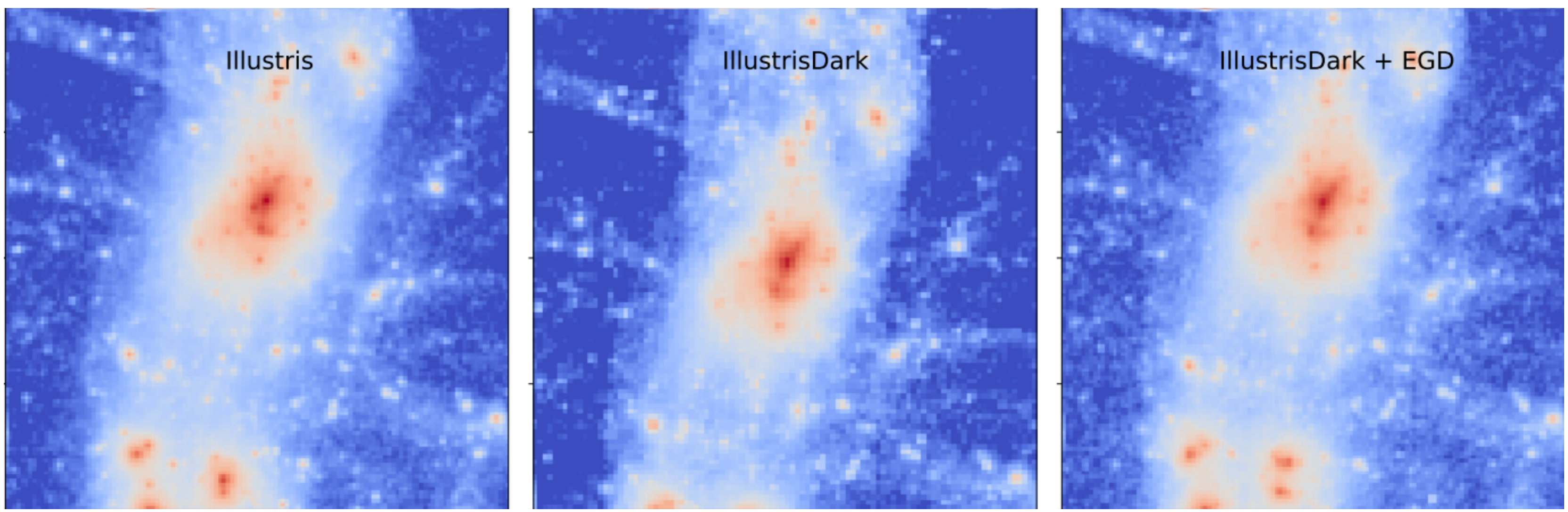
Natalia Porqueres

Baryon feedback: Enthalpy gradient descent model

Displaces dark matter particles with

$$\mathbf{S}_{\mathrm{baryons}} = -\frac{\beta}{H_0^2}\frac{k_B T_0}{\mu}\frac{\gamma}{\gamma-1}\nabla\left[\mathbf{O}_{\mathrm{J}}(1+\delta)\right]^{\gamma-1}$$

Assumes power law equation of state of baryons $T(\delta) = T_0(1+\delta_b)^{\gamma-1}$



Dai & Seljiak, 2018

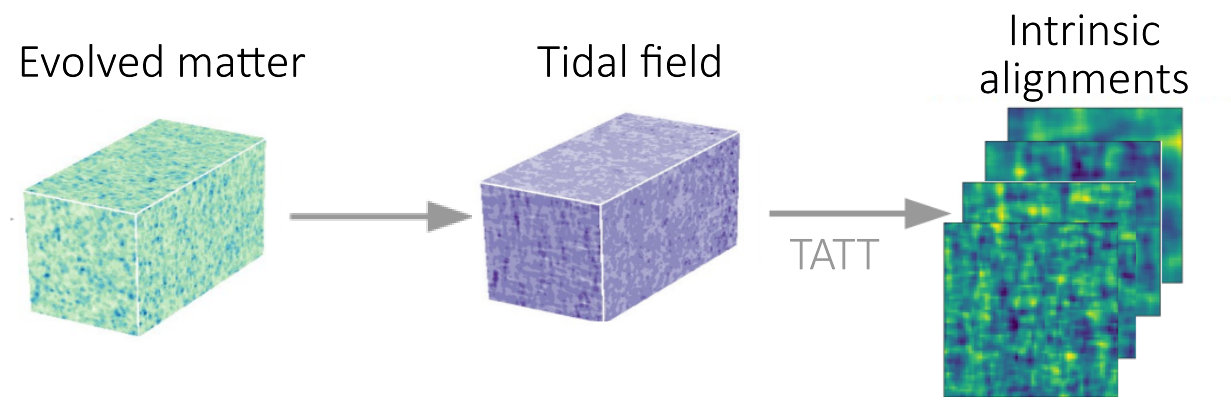
Self calibration of model parameters.

Displaces particles independently of whether they are in a halo.

Fully differentiable.

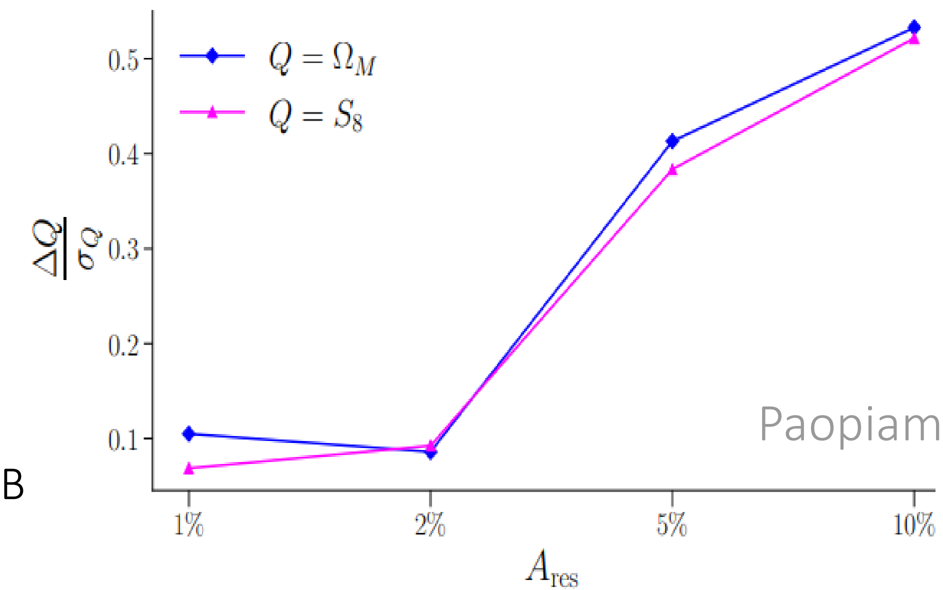
Intrinsic alignments: Tidal alignment and tidal torquing

$$\gamma_1^{\mathrm{IA}}(\boldsymbol{r},\boldsymbol{\vartheta}) = (C_1 + C_{1\delta}\delta)(s_{xx} - syy) + C_2(s_{xk}s_{xk} - s_{yk}s_{yk})$$
$$\gamma_2^{\mathrm{IA}}(\boldsymbol{r},\boldsymbol{\vartheta}) = 2(C_1 + C_{1\delta}\delta)s_{xy} + 2C_2 s_{xk}s_{yk},$$



Accuracy requirements StageIV: miss-modelling **10%** would lead to 0.5 sigma bias

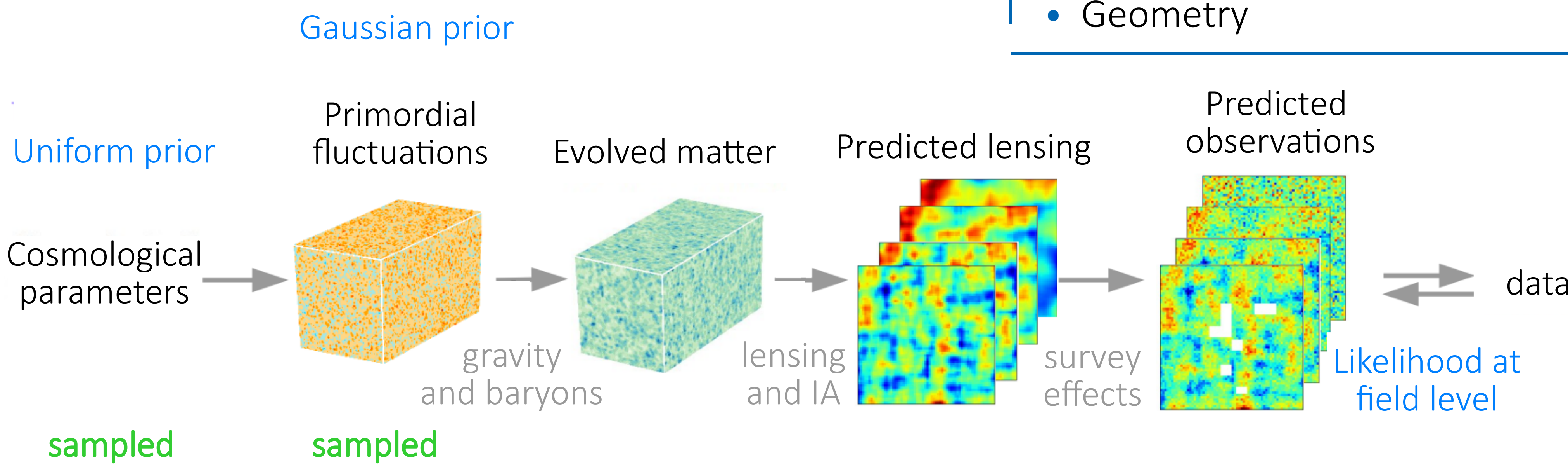
Anya Paopiamsap
MPhys Oxford → PhD at UB



Paopiamsap, NP+ 2024

BORG-WL: field-level inference of weak lensing

Sensitivity to cosmology:
- Initial matter power spectrum
- Growth of structures
- Geometry



Likelihood: Assuming Gaussian noise in pixelised shear

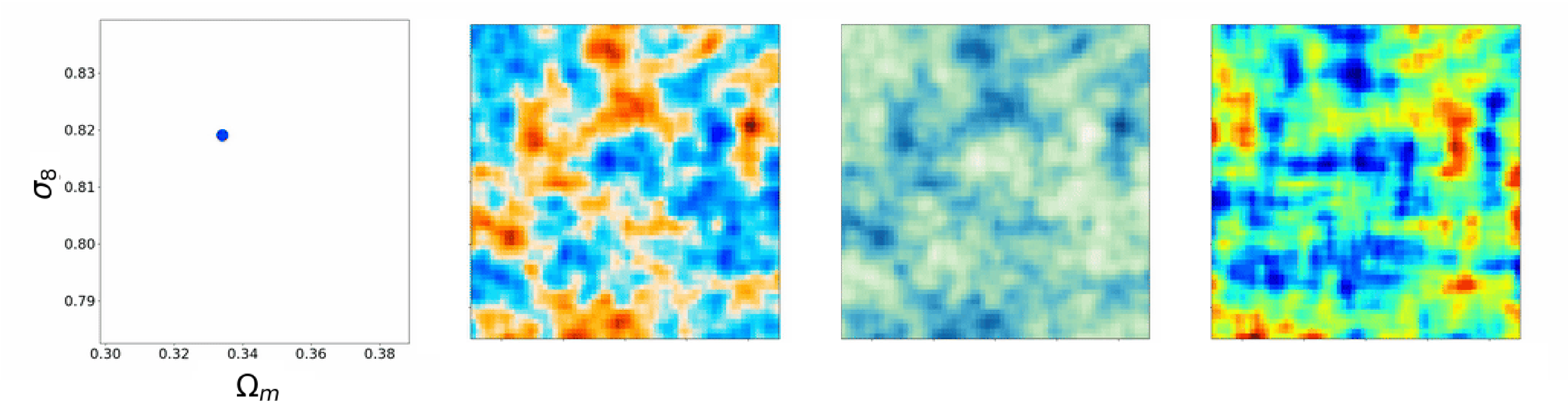
$$\log \mathcal{L} = -\frac{1}{2} \sum_{b} \sum_{mn} \frac{(\epsilon^{b}_{1,mn} - \hat{\epsilon}^{b}_{1,mn})^2 + (\epsilon^{b}_{2,mn} - \hat{\epsilon}^{b}_{2,mn})^2}{\sigma_b^2}$$

Doesn't need covariance matrix $\sigma_b = \sigma_\epsilon / \sqrt{N_b}$

Natalia Porqueres

The samples

Cosmological parameters | Primordial fluctuations | Evolved matter | Predicted lensing



sampled | sampled

10^7 parameters!

$$P(\theta, \delta^{\mathrm{ic}} | \hat{\epsilon}) \propto P(\hat{\epsilon}|\epsilon) P(\epsilon|\delta^{\mathrm{ic}}) P(\delta^{\mathrm{ic}}|\theta) P(\theta)$$

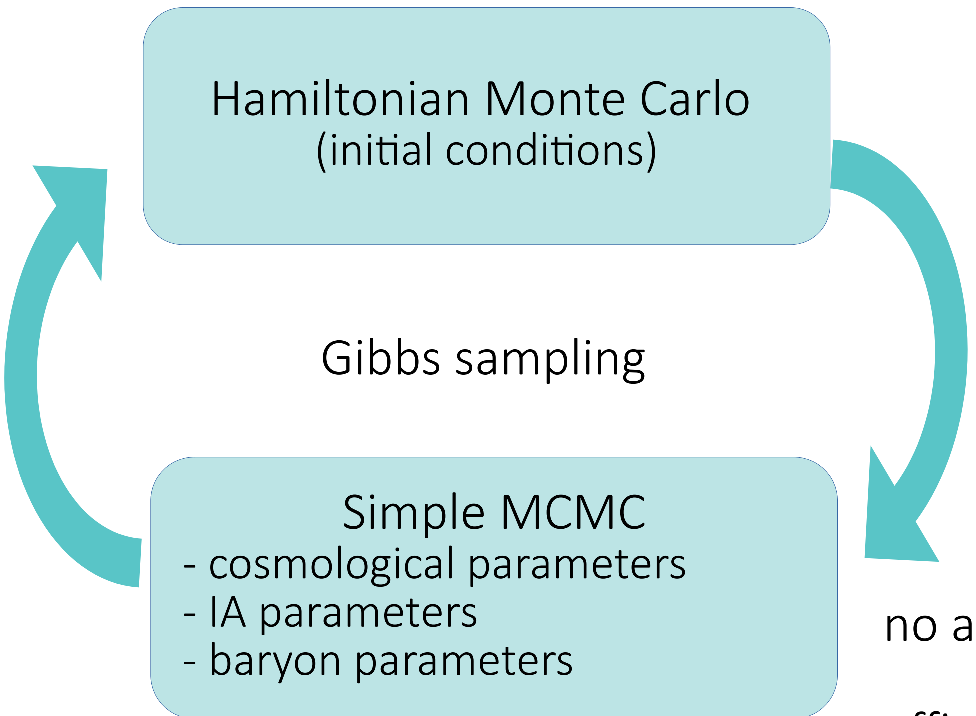
posterior　likelihood　prior

$$\delta^D \left[\epsilon - \mathcal{F}(\delta^{\mathrm{ic}})\right] \; G(0, S(\theta))$$

forward model

Natalia Porqueres

Sampling in a high-dimensional space



Addresses curse of dimensionality by:

1. exploiting information in gradients
2. using conserved quantities

no analytical derivatives

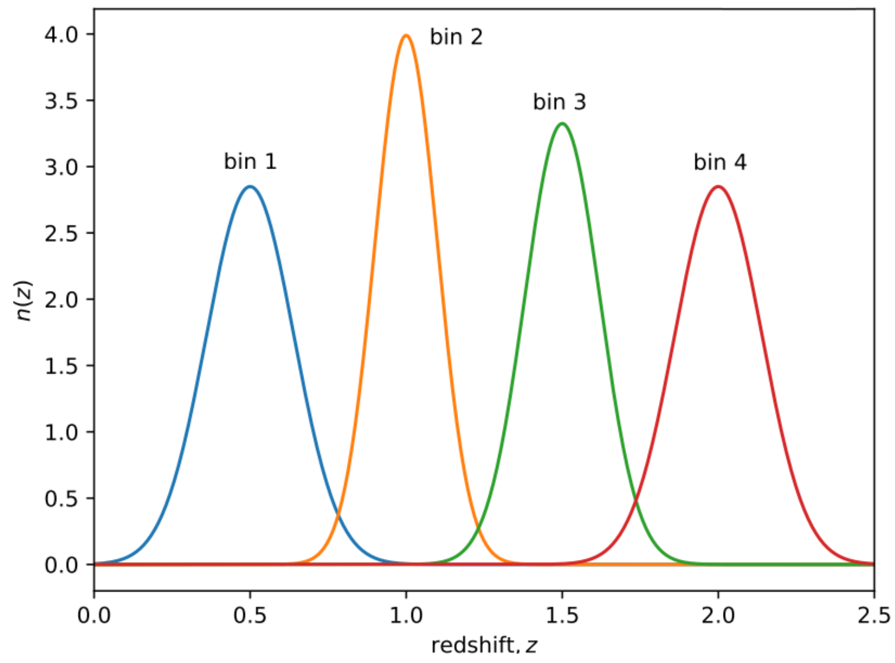
efficient mass matrix?

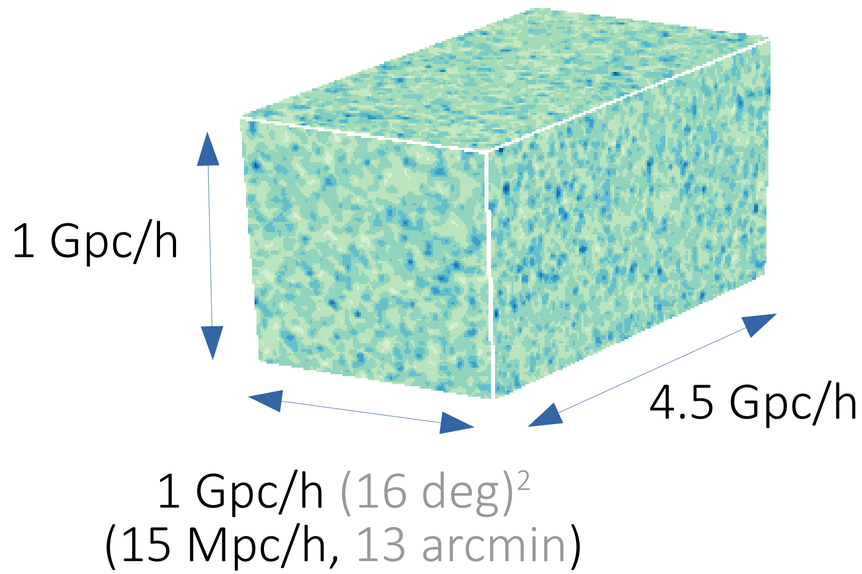
The proof of concept Results with synthetic data

How much do we get from going to the field level?

Simulated data

Tomographic bins





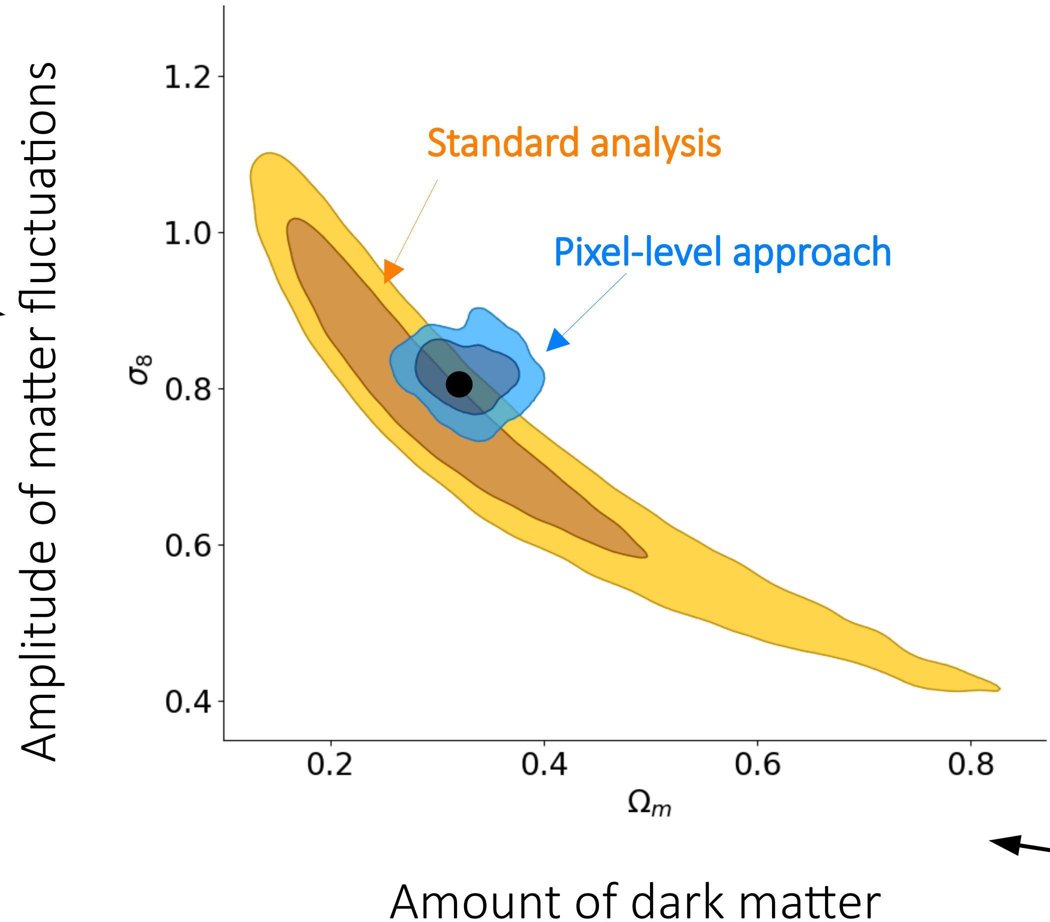
Gravity model: LPT.

Gaussian pixel-noise: 30 sources per square arcmin.

Intrinsic alignment amplitudes: DES values (0.18, 0.8, 0.1)

Comparison of cosmology constraints

marginal error
3 times smaller



(1 Gpc/h)^2, 4.5 Gpc/h
(16 deg)^2, 13 arcmin
LPT

marginal error
5 times smaller

Field-level approach **lifts degeneracy** by extracting more information from the data

NP, Heavens +2022, 2023

Where is the information coming from?

Comparison with high-order stats

Is there a summary statistic that captures as much information?

Extreme data compression that preserves all information

MOPED for Gaussian data, can we do it for non-Gaussian data?

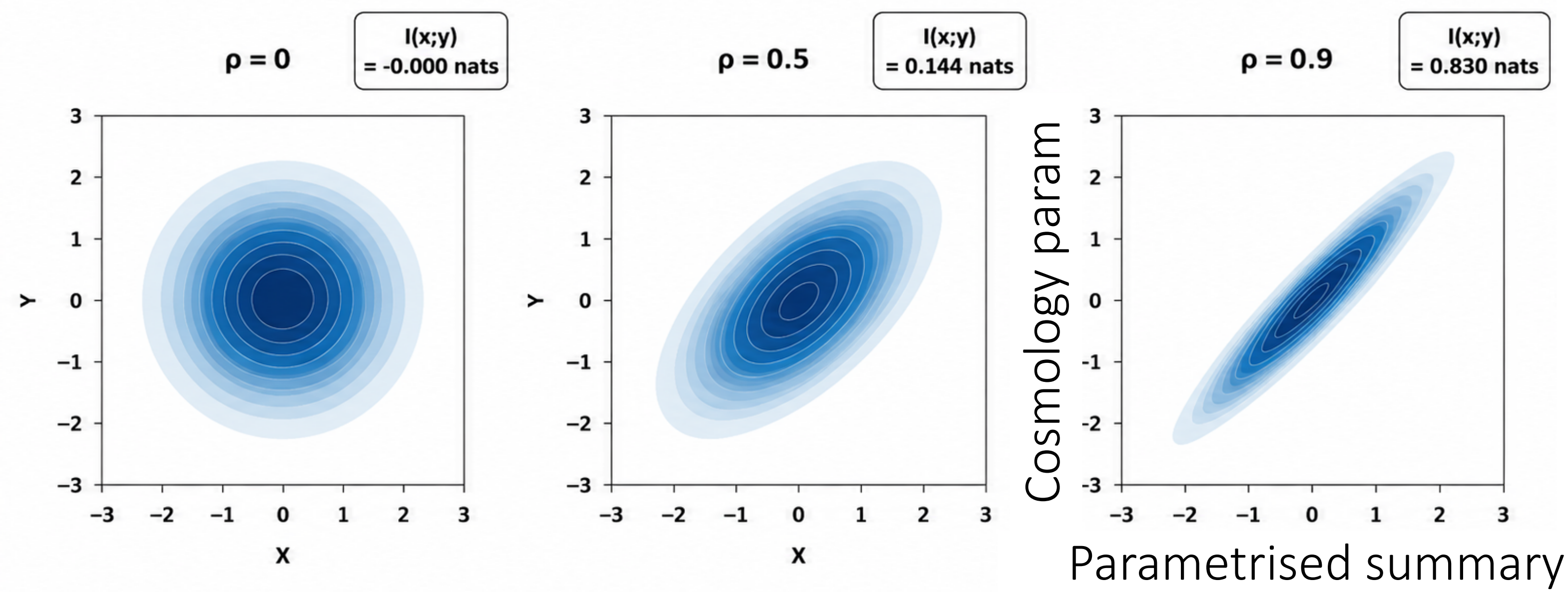
Mutual information

Lucas Makinen
Imperial → Cambridge

Find a few summary statistics that capture as much info as possible

Maximise mutual information:

$$I(A;B) = \iint p(a,b) \ln\left(\frac{p(a,b)}{p(a)p(b)}\right) \mathrm{d}a\, \mathrm{d}b$$



Find a set of (parametrised) summaries that maximises the MI with cosmological params

Maximise mutual information is a non-linear function of the distribution (correlation)

Using NN to compress data to parametrised summaries

Natalia Porqueres

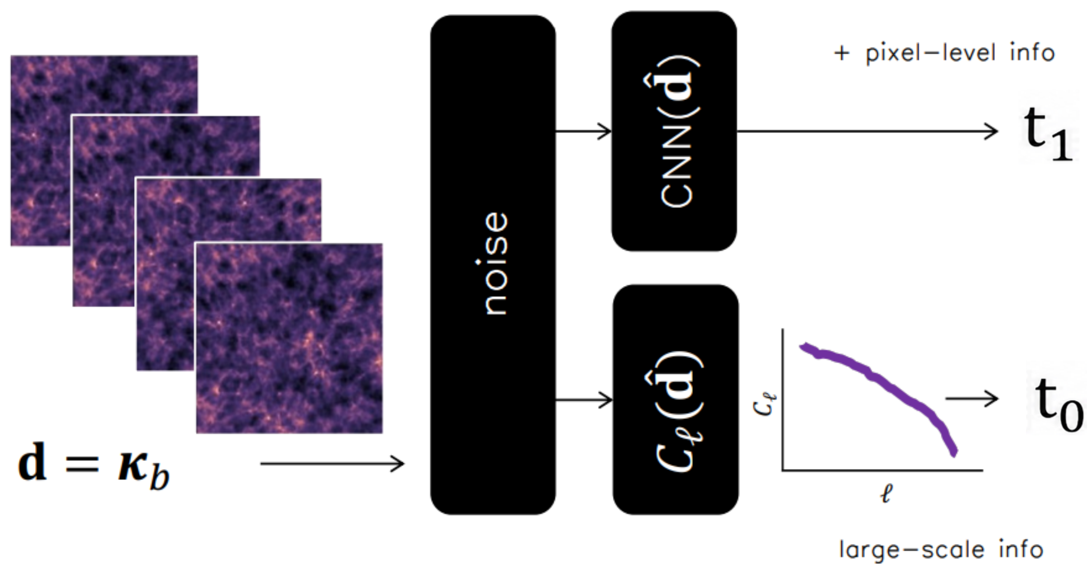
Hybrid summary statistics

Lucas Makinen
Imperial → Cambridge

Use Physics to assist network and make it scalable:

- On large scales, Cls are informative.
- On small scales, there is more information

Ask neural network to capture ***additional*** information
→ Maximise conditional mutual information



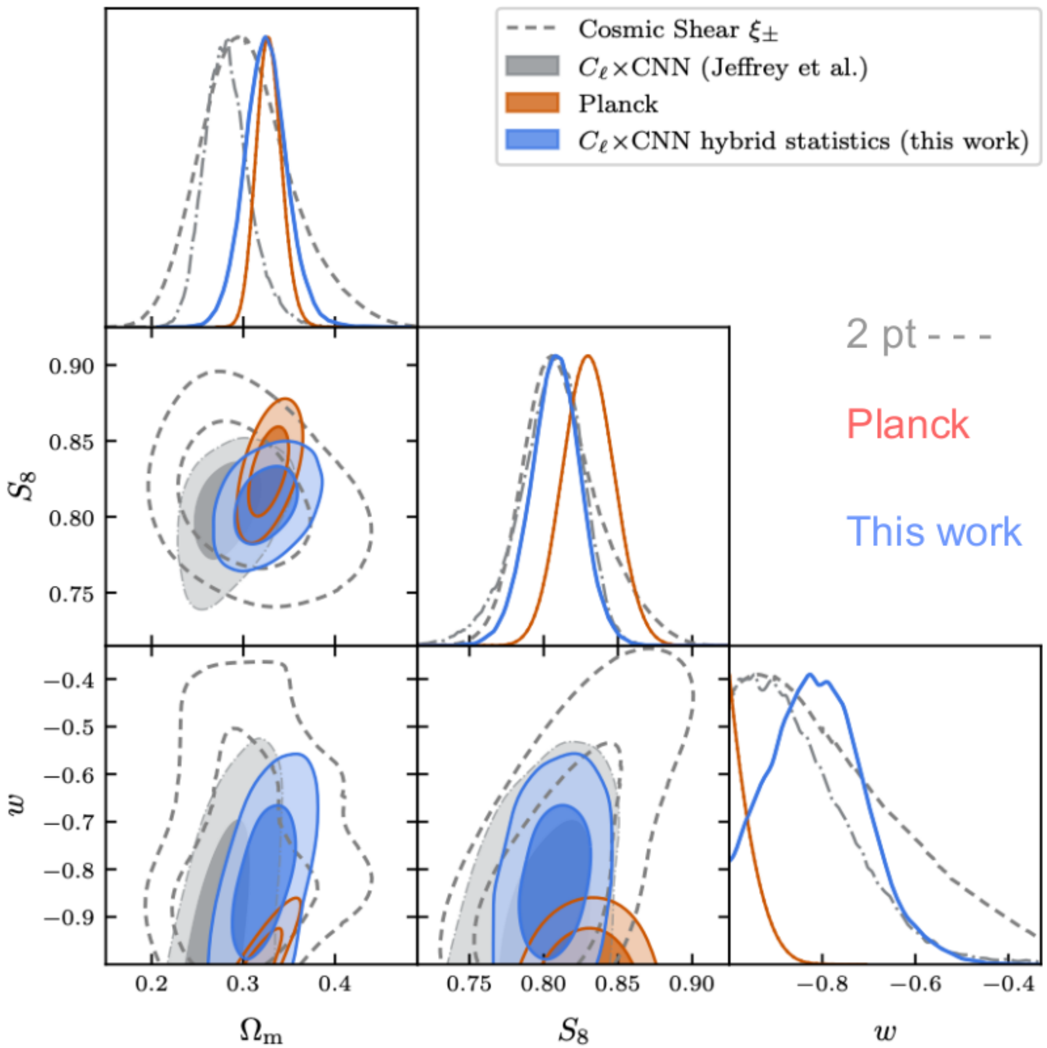
$$I(\Theta;\, T_1|T_0) \;=\; \int p(\theta, t_1, t_0)\, \ln\left[\frac{p(\theta, t_1|t_0)}{p(\theta|t_0)\, p(t_1|t_0)}\right] d\theta\; dt_1\; dt_0$$

Makinen+ 2024

Application to DES Y3: most precise constraints from WL

Gower Street simulations (Jeffrey & Gatti, 2021), robust to baryons/gravity model

Most precise constraints from WL alone



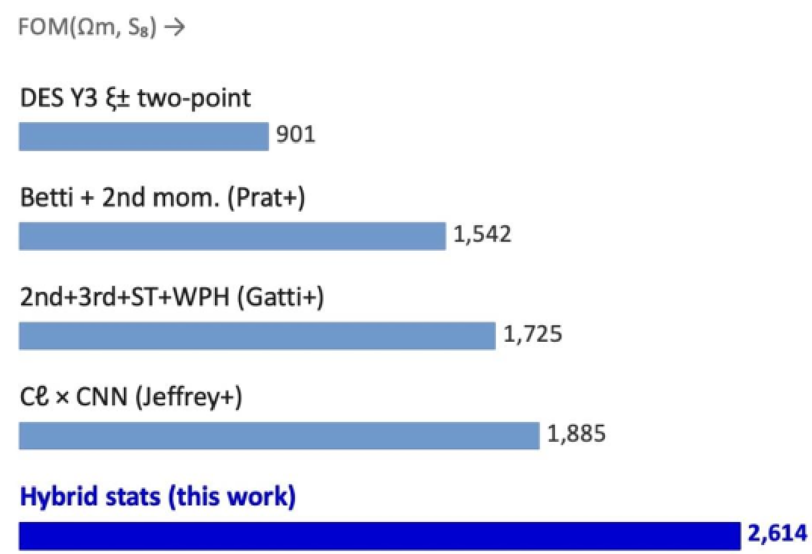
2 pt - - -

Planck

This work

Lucas Makinen
Imperial → Cambridge

Josh Williamson
UCL

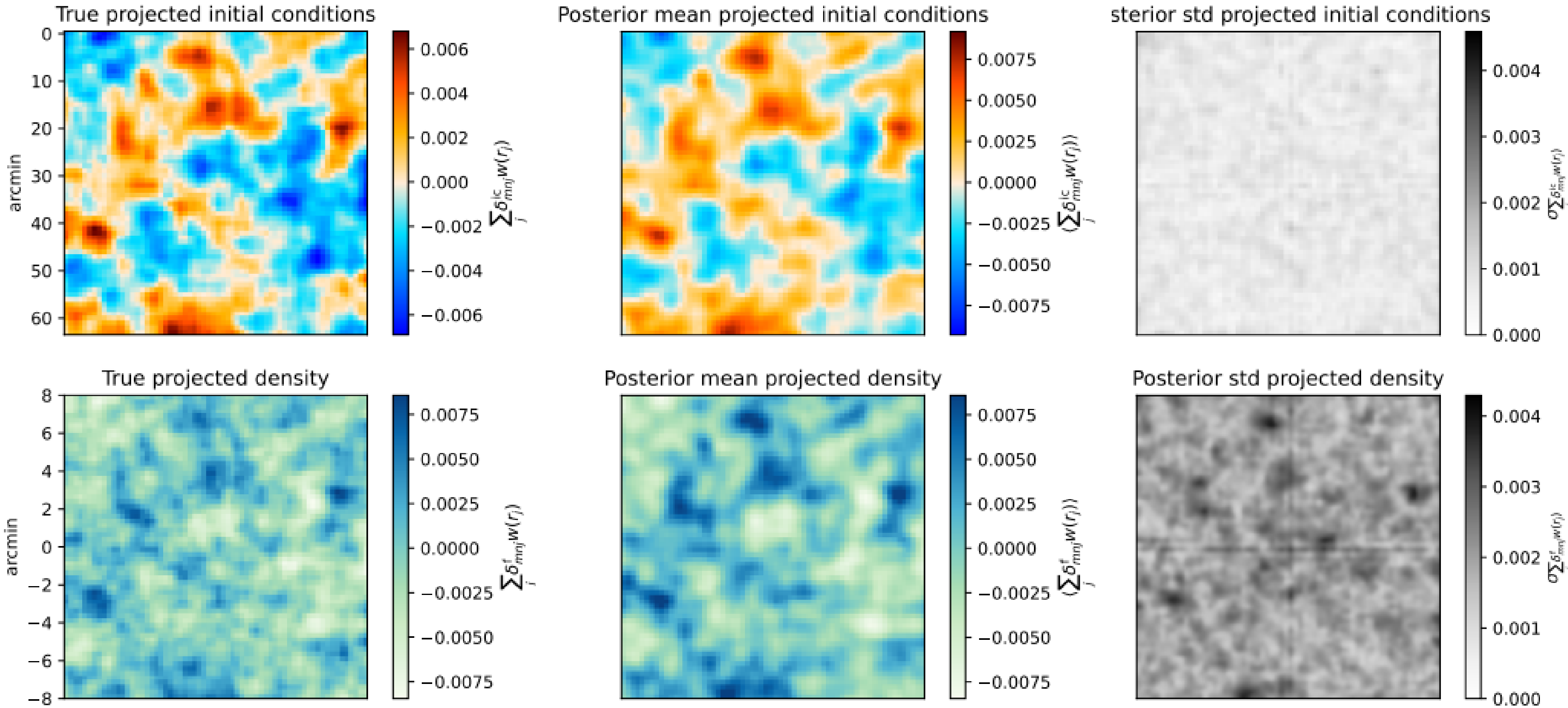


Williamson & Makinen, **NP**+ 2026

Going back to the field-level framework....

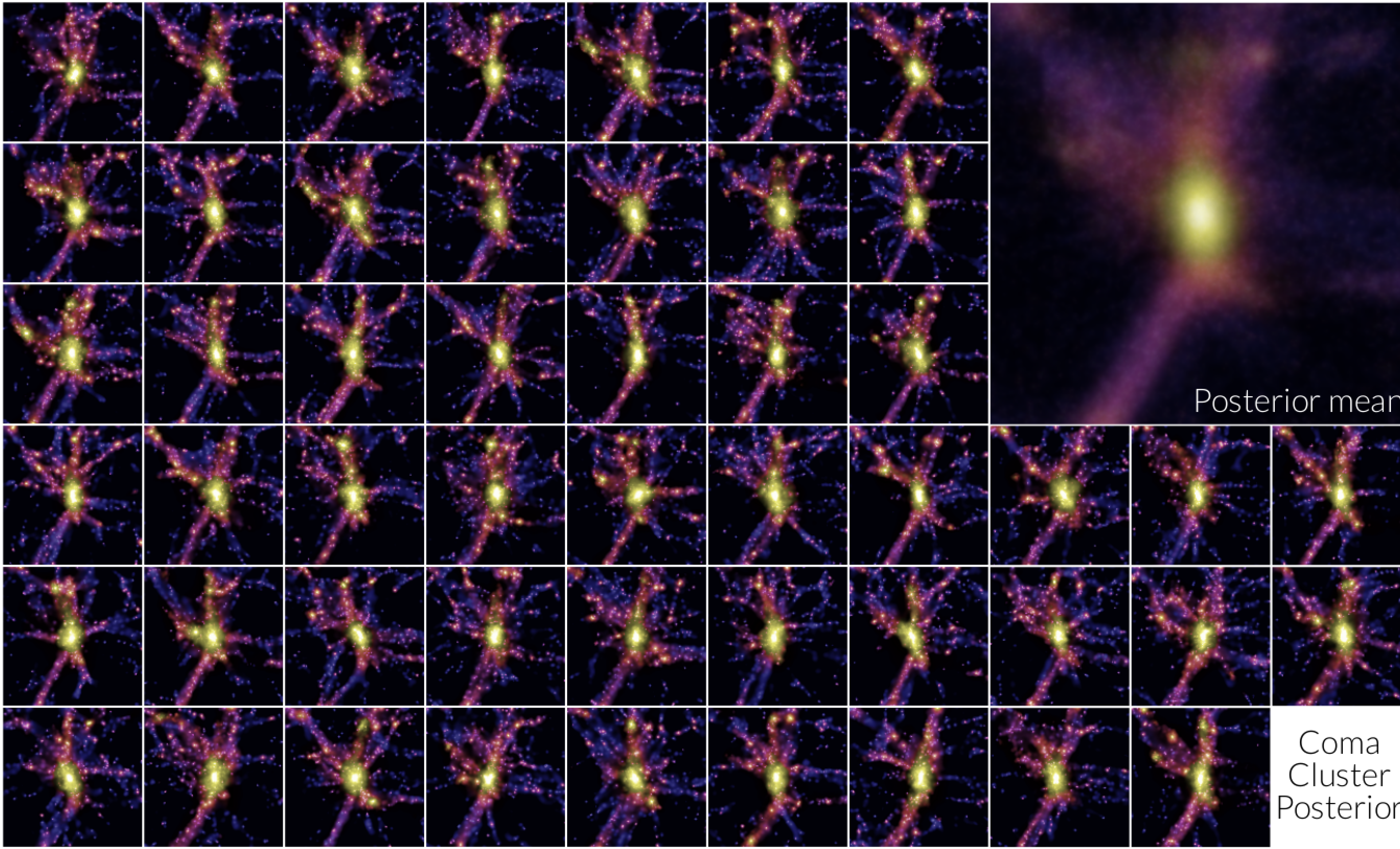
A digital twin of the Universe

Inferred density fields



NP+2022, 2023

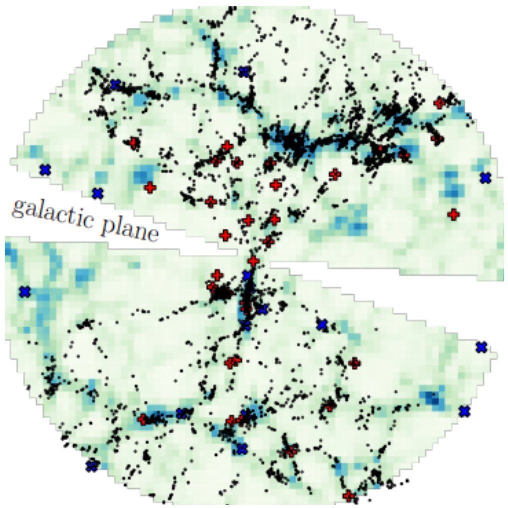
Posterior distribution



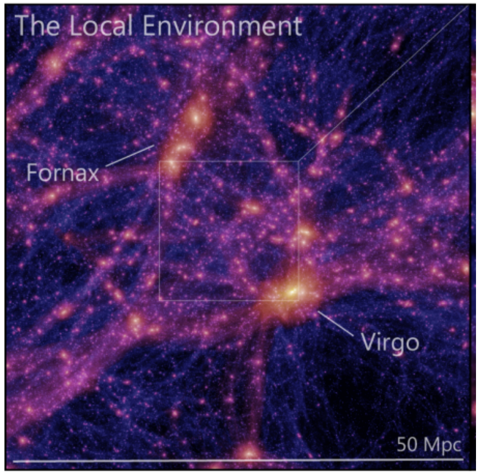
McAlpine+2025

What physics can we do with the density fields?

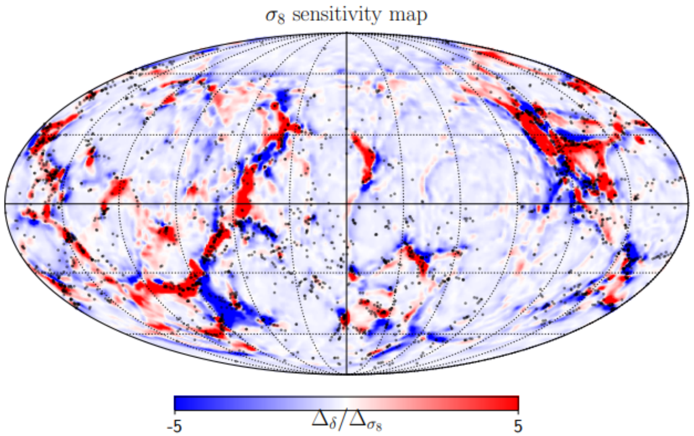
Galaxy environment
(**NP**+ 2017)



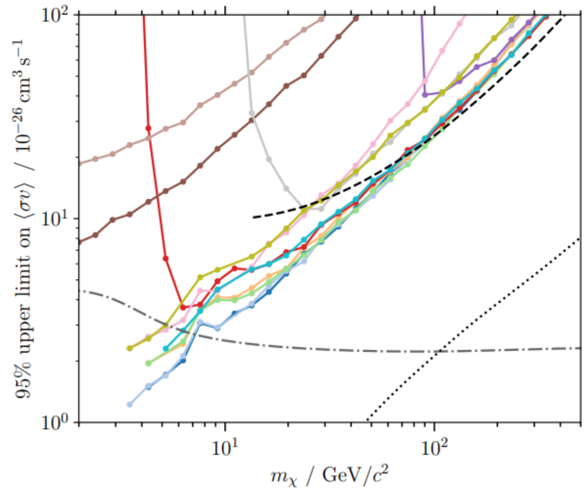
Constrained simulations
(McAlpine+ 2022)



Targeted surveys
(Kostic+ 2022)



Constrain DM annihilation
(Bartlett+ 2022)

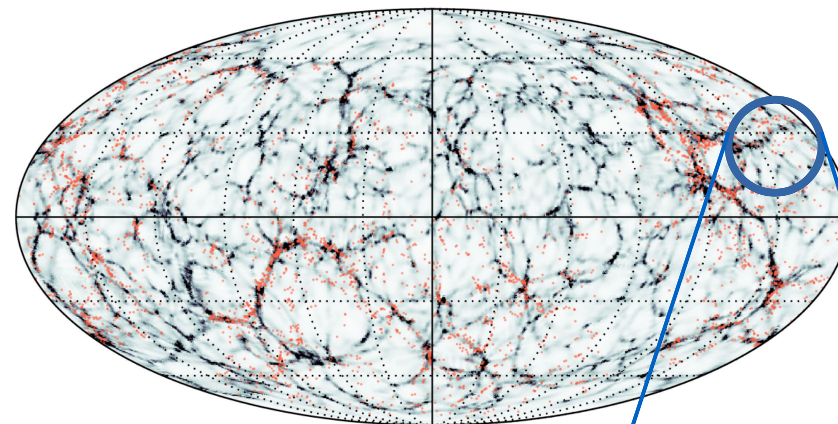


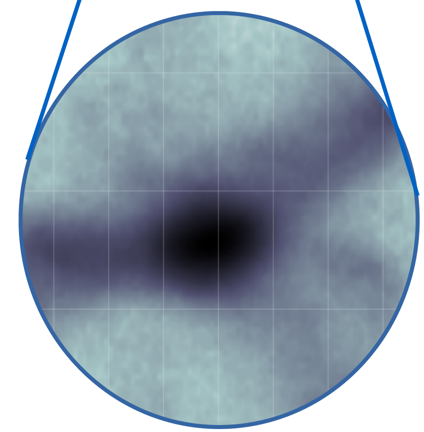
Natalia Porqueres

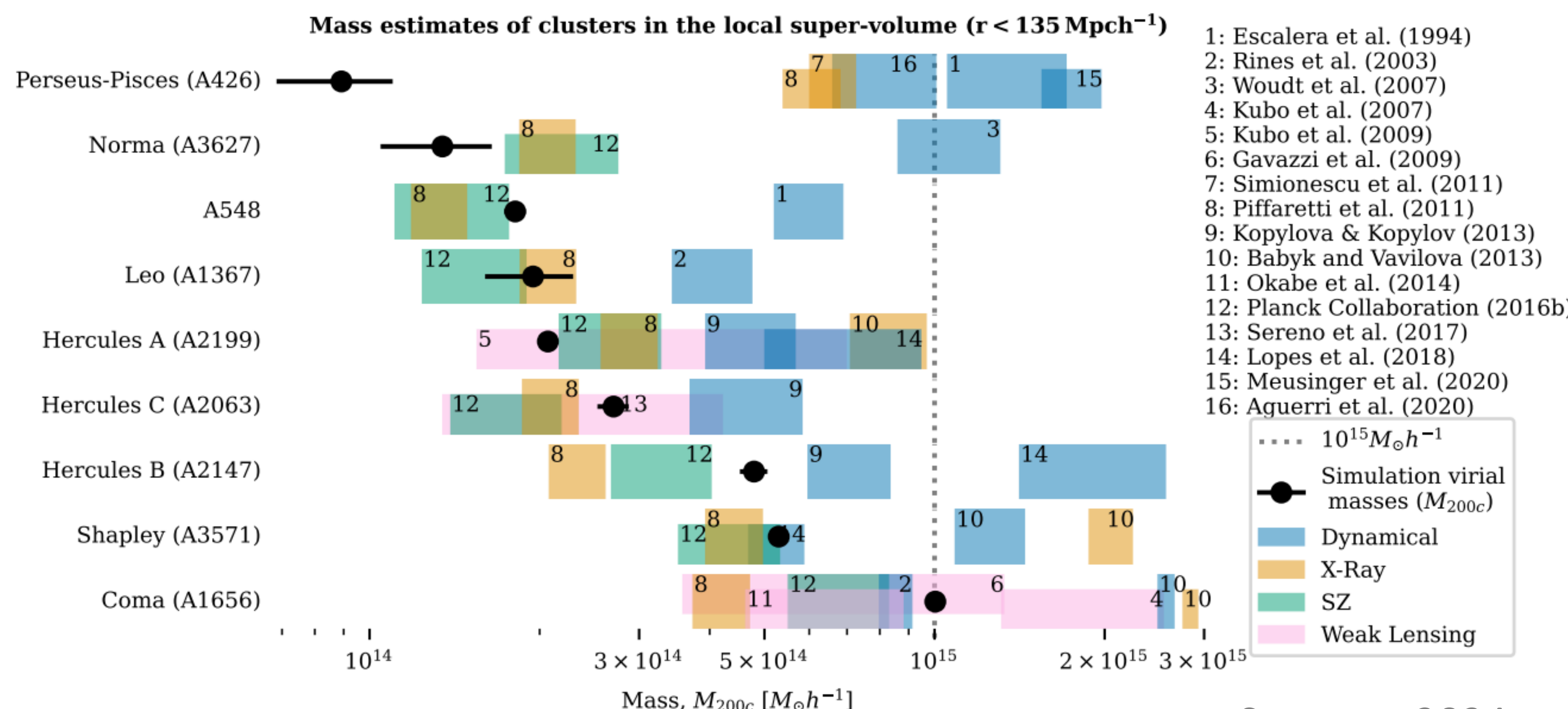
Density fields as a validation tool

Posterior predictive tests and cross-validation with independent measurements

Inferred mass profile agree with external measurements?







Stopyra+ 2024

Natalia Porqueres

The challenge Getting ready for Euclid data

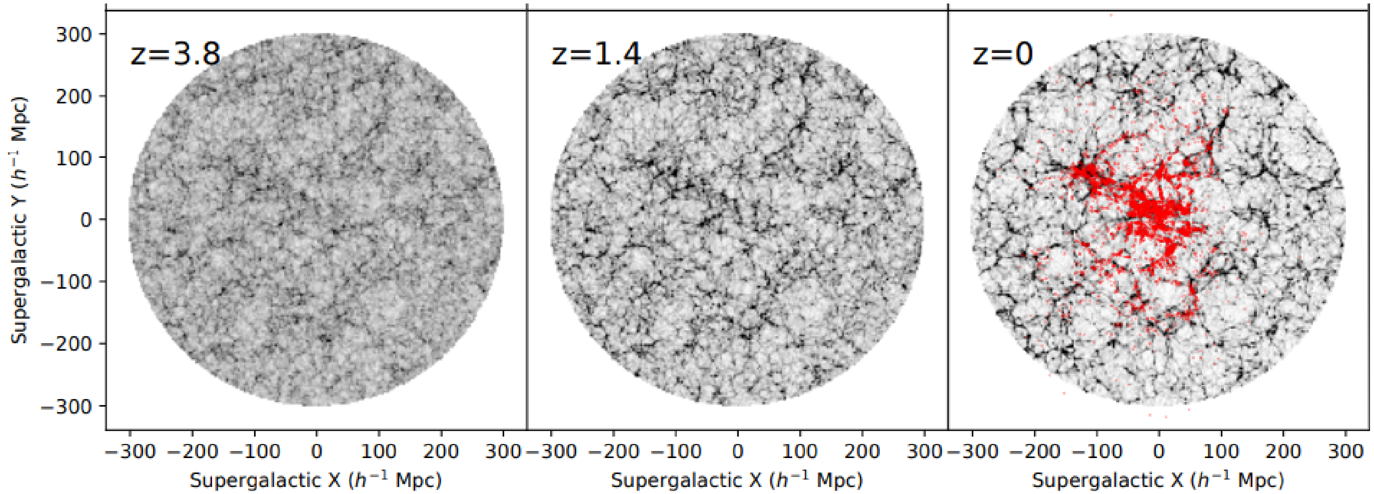
More realistic gravity model

Systematics (redshifts, model accuracy)

Scalability to Euclid data

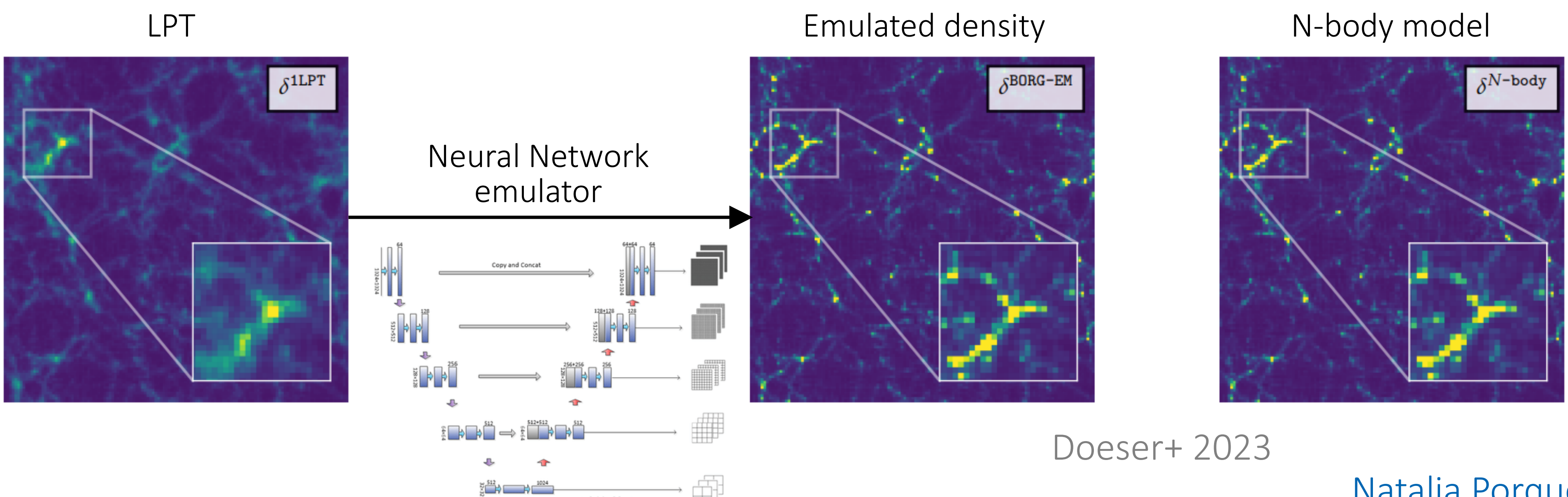
More realistic gravity model in BORG

Particle-mesh is **ready** and has been **used in real data** BORG analysis (2M++)



Jasche & Lavaux 2019

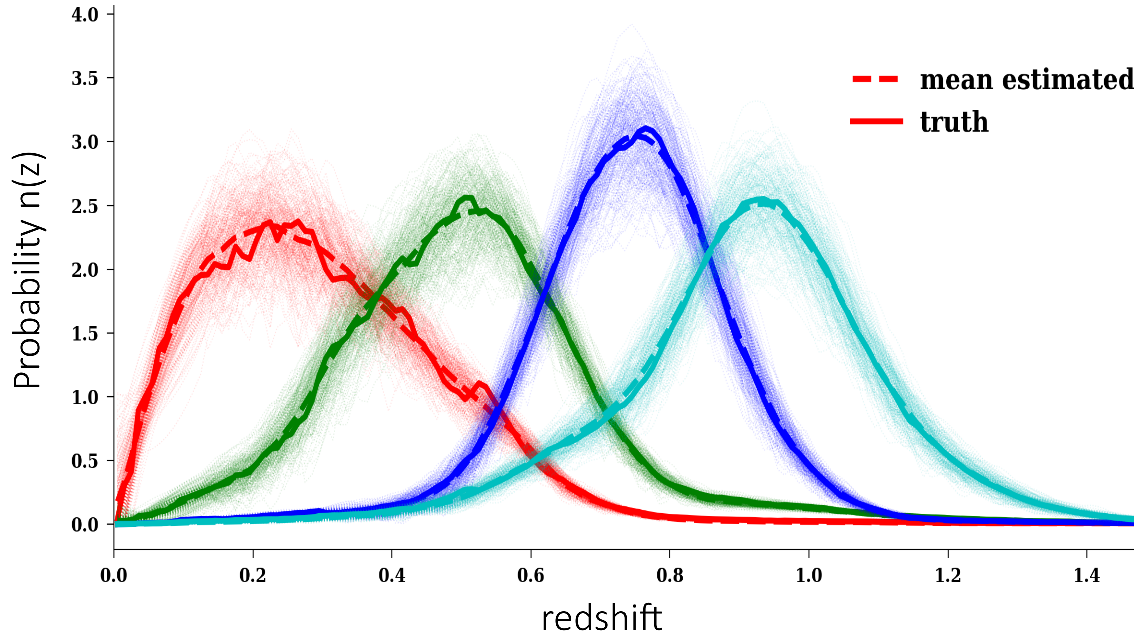
N-body emulator: 1% agreement (4x faster than PM, 1.6s/evaluation, 1Mpc resolution)
CNN (V-net) trained LPT & N-body pairs, loss preserves Eulerian and Lagrangian properties



Doeser+ 2023

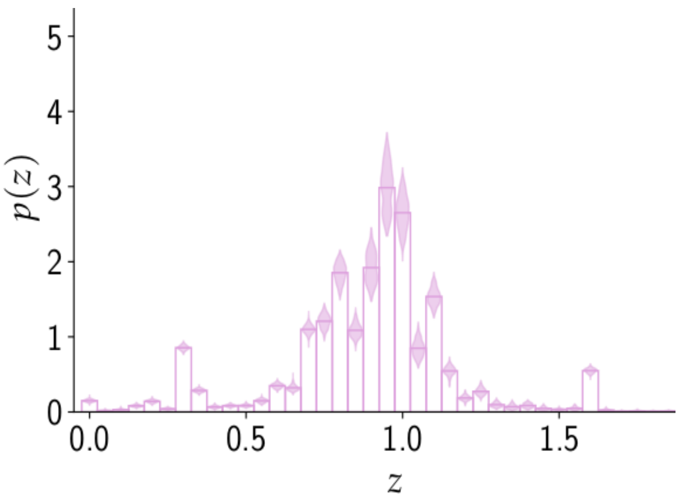
Natalia Porqueres

Systematics: Redshift distribution



Myles+ 2023

Jordy Ram
CEA (soon)



Leistedt+ 2016;
Kyriacou, Heavens+ in prep

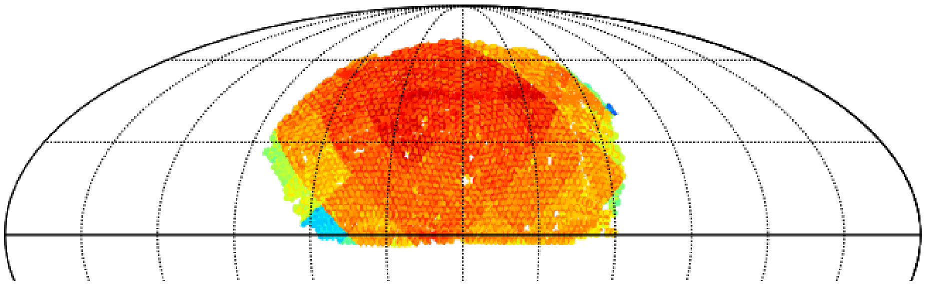
Hybrid physics-informed and machine-learning sampler

Sample n(z) as parametrised histograms

Marginalise over n(z) parameters

Systematics: Unknown contaminations

Unknown contaminations → Physical models do not exist



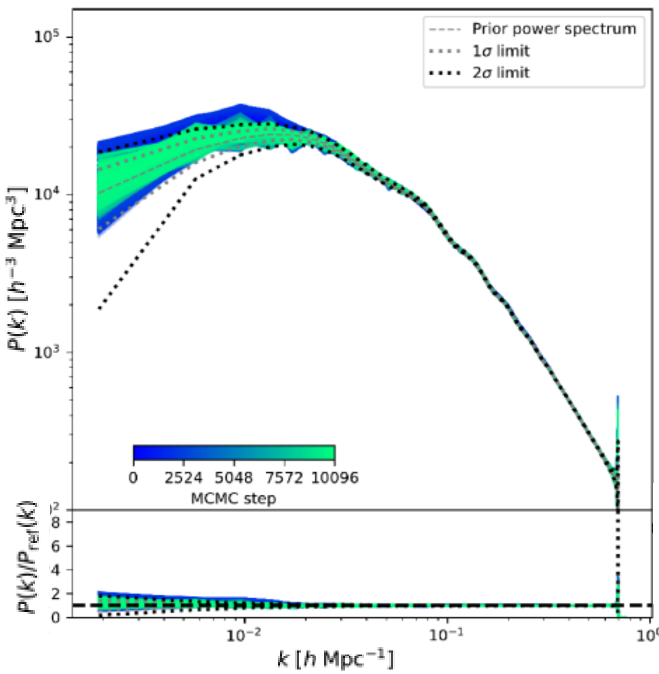
$$\lambda_c = A_c \bar{\lambda}_c$$

$$\downarrow$$

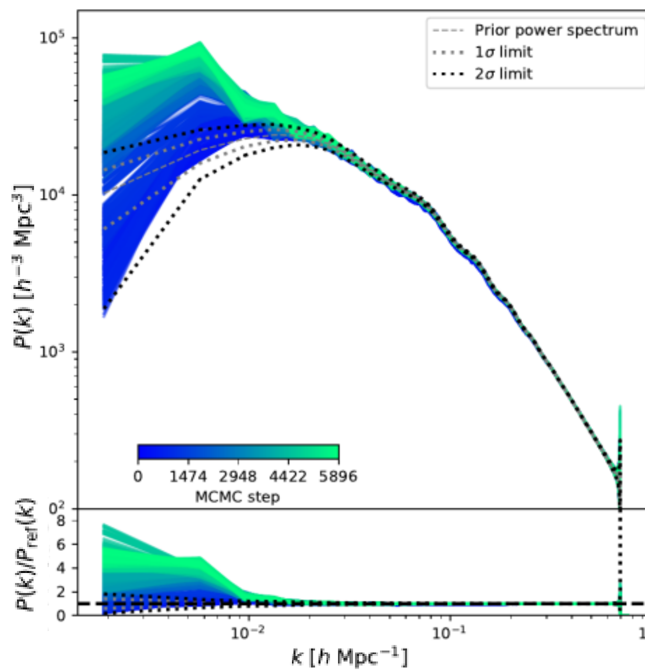
$$P(N|\lambda) = \prod_i \frac{e^{-\lambda_i}(\lambda_i)^{N_i}}{N_i!}$$

NP+ 2018

With unknown contaminations model



Without the model

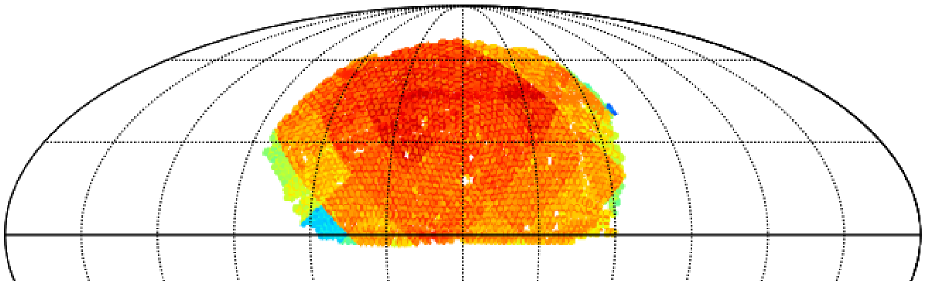


NP+ 2018

Natalia Porqueres

Systematics: Unknown contaminations

Unknown contaminations → Physical models do not exist



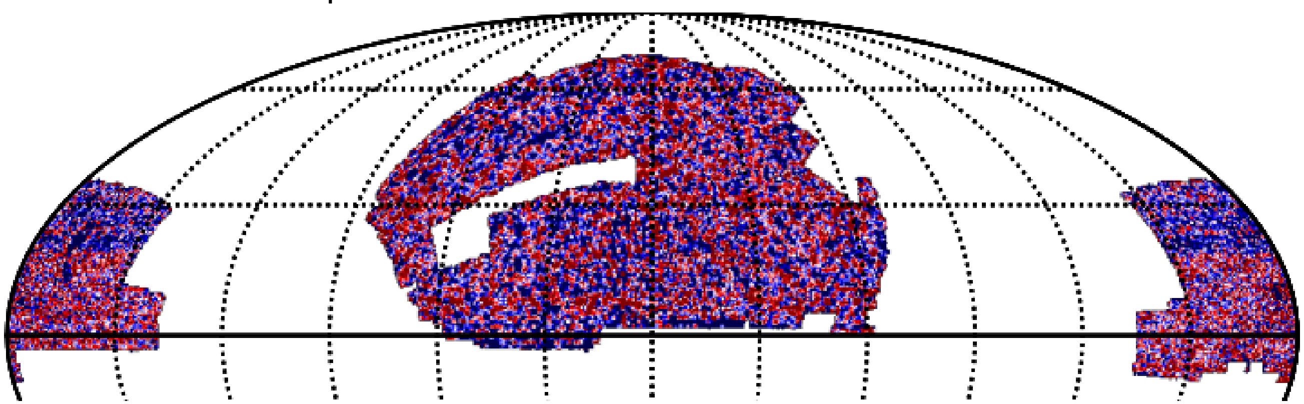
$$\lambda_c = A_c \bar{\lambda}_c$$

$$\downarrow$$

$$P(N|\lambda) = \prod_i \frac{e^{-\lambda_i}(\lambda_i)^{N_i}}{N_i!}$$

NP+ 2018

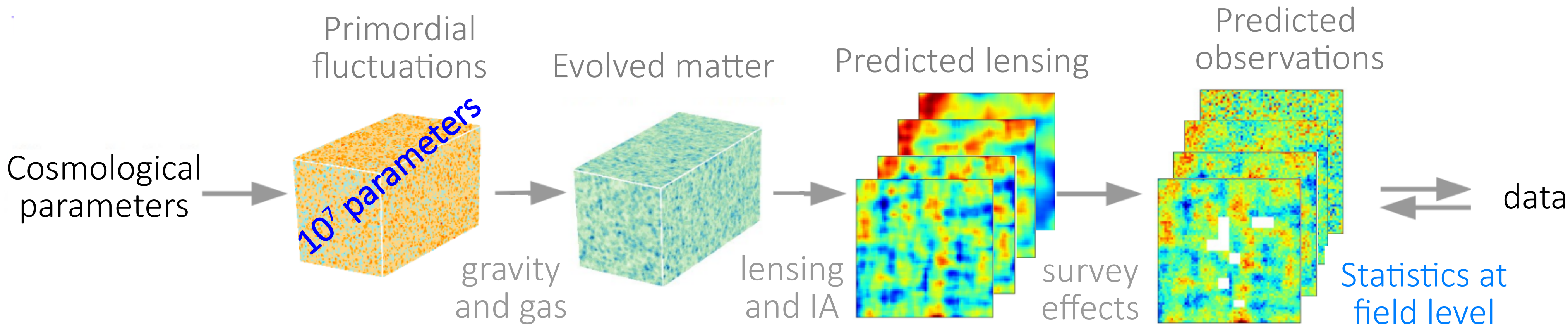
Maps of the unknown contamination



Lavaux+ 2019

Field-level inference to detect unknown contaminations

Scaling to the Euclid volume

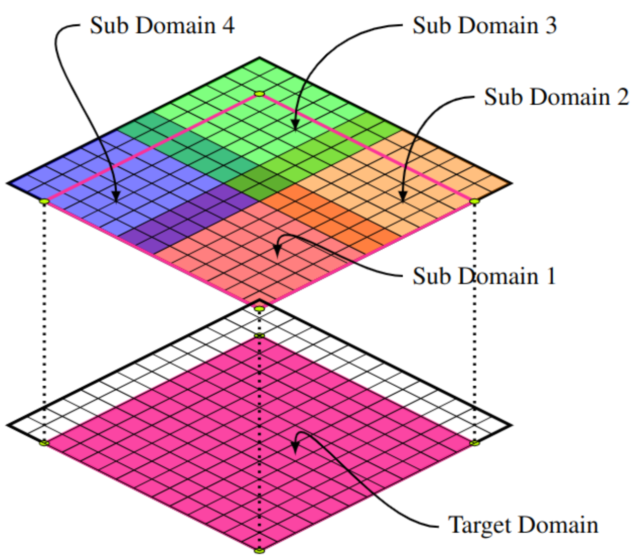


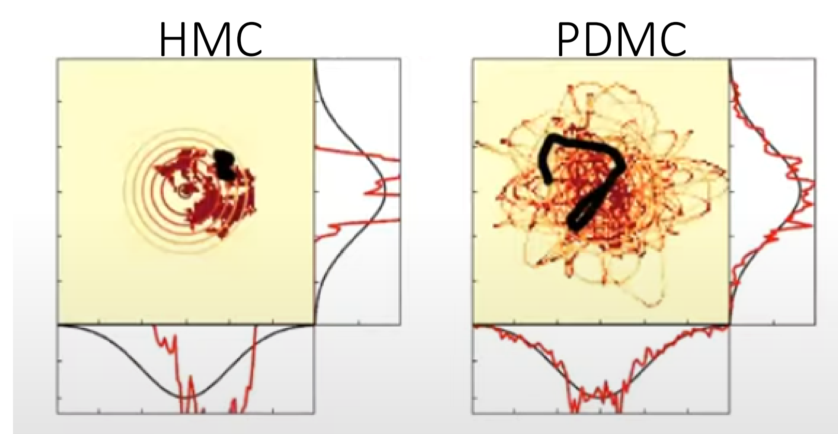
1. Gravity emulator

2. Efficient sampler

Bierkens+2023

3. Sub-volume analysis





Michel et al. (2020)

Natalia Porqueres

Summary

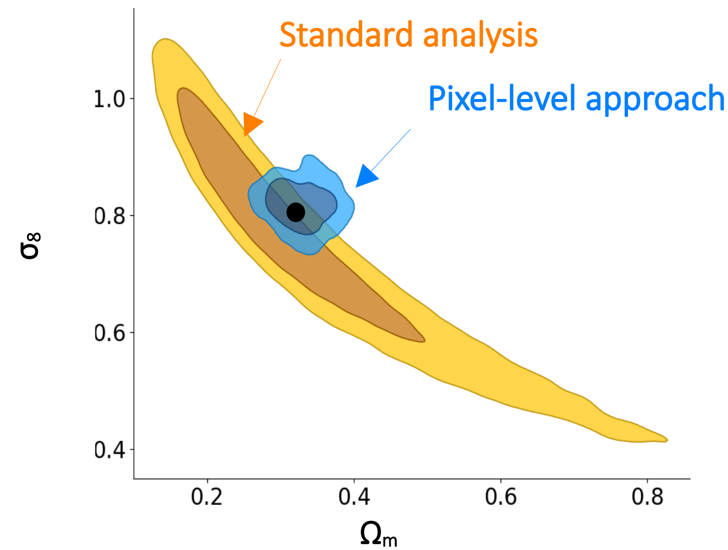
- There is more information in the data that the 2-point summary statistics do not capture.

- Field-based approach reduces marginal error up to a factor of 5.

- What's next?

  Testing with Euclid's Flagship simulations

  First data application!



A digital twin of the Universe