Colloque Action Nationale Dark Energy - November 3rd 2026



Beyond standard galaxy clustering methods with DESI

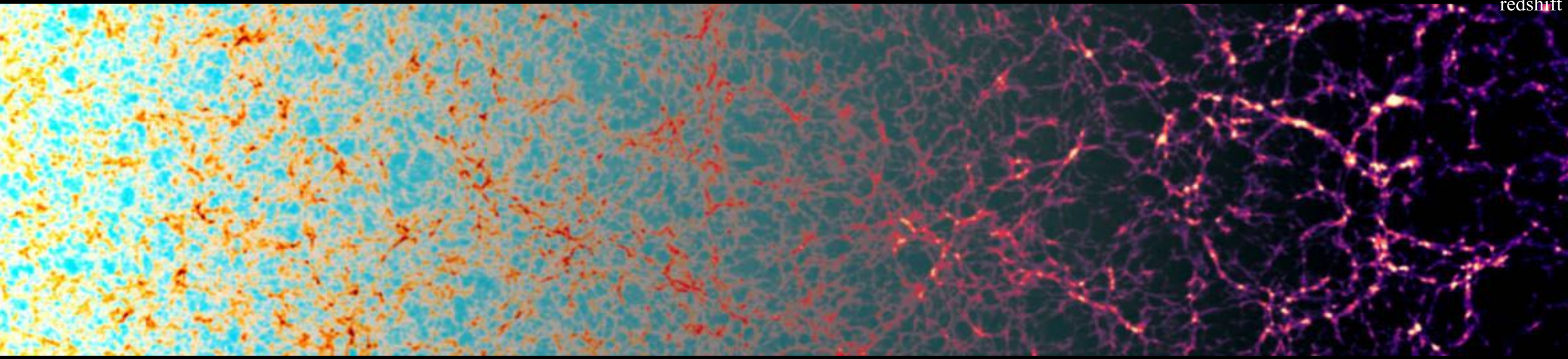
Pauline Zarrouk (CNRS-LPNHE, Sorbonne Université)

How to extract cosmological information?

Sensitivity to cosmology

Inflation and primordial features | Dark matter and neutrinos | Dark energy

redshift



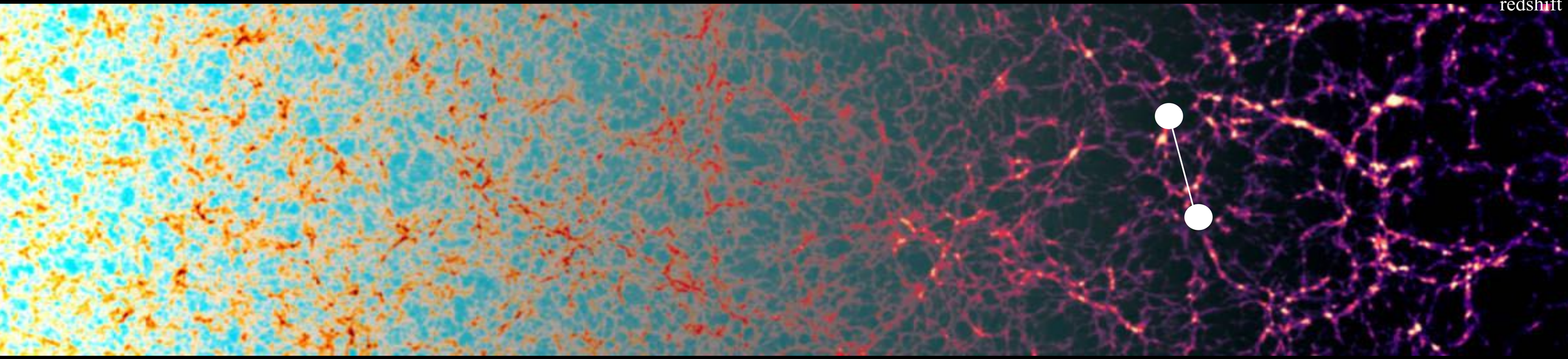
Initial Gaussian and linear density field

How to extract cosmological information?

Sensitivity to cosmology

Inflation and primordial features — Dark matter and neutrinos — Dark energy

redshift



Initial Gaussian and linear density field

← more robust

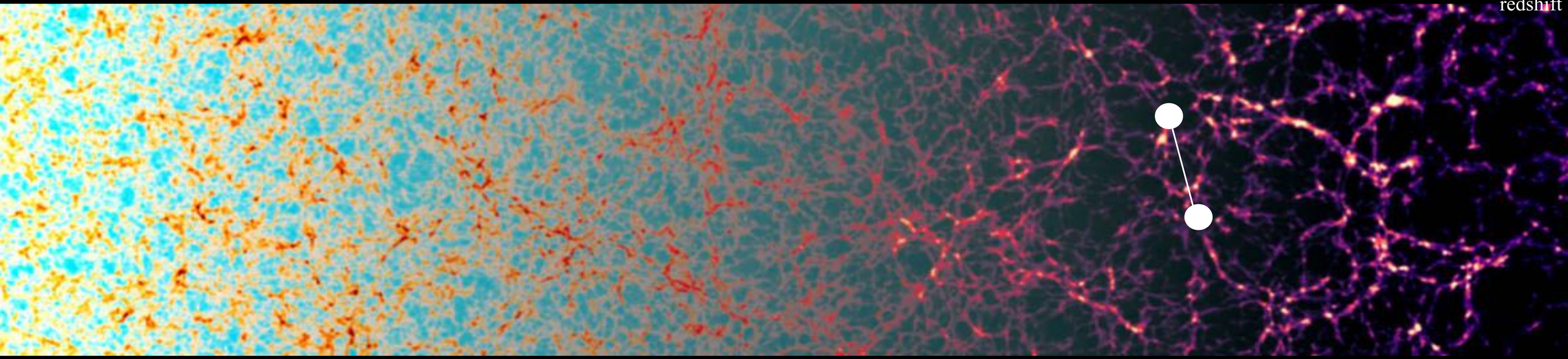
Two-point statistics (2pt)

- Most mature and robust technique
- Analytical model available up to quasi-linear scales (k_{max}=0.2-0.3 h/Mpc)
- Computationally cheap
- Only Gaussian and linear information

How to extract cosmological information?

Sensitivity to cosmology

Inflation and primordial features — Dark matter and neutrinos — Dark energy → redshift



Initial Gaussian and linear density field

← more robust

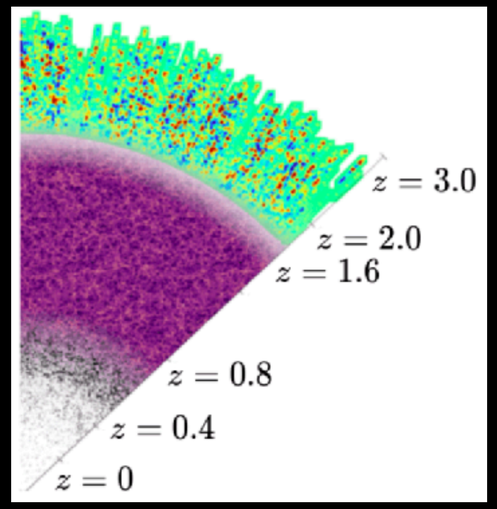
Two-point statistics (2pt)

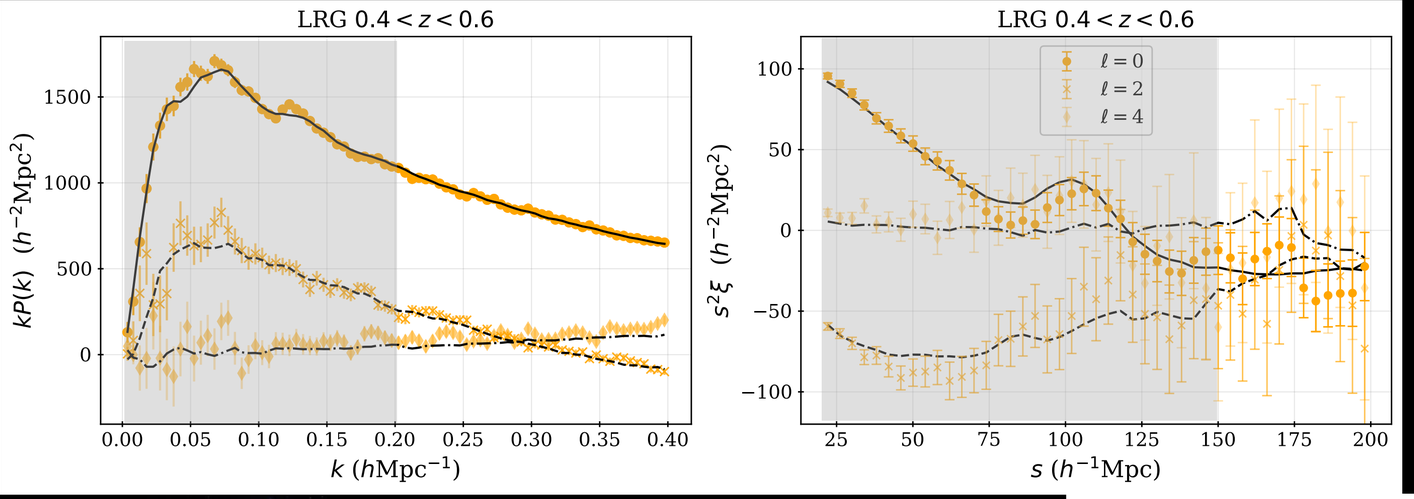
- Most mature and robust technique
- Analytical model available up to quasi-linear scales (k_{max}=0.2-0.3 h/Mpc)
- Computationally cheap
- Only Gaussian and linear information

DESI DR1 Galaxy Full-Shape, DESI collaboration
(arXiv:2411.12021 (co-lead), arXiv:2411.12022)
—> 2pt statistics / EFT

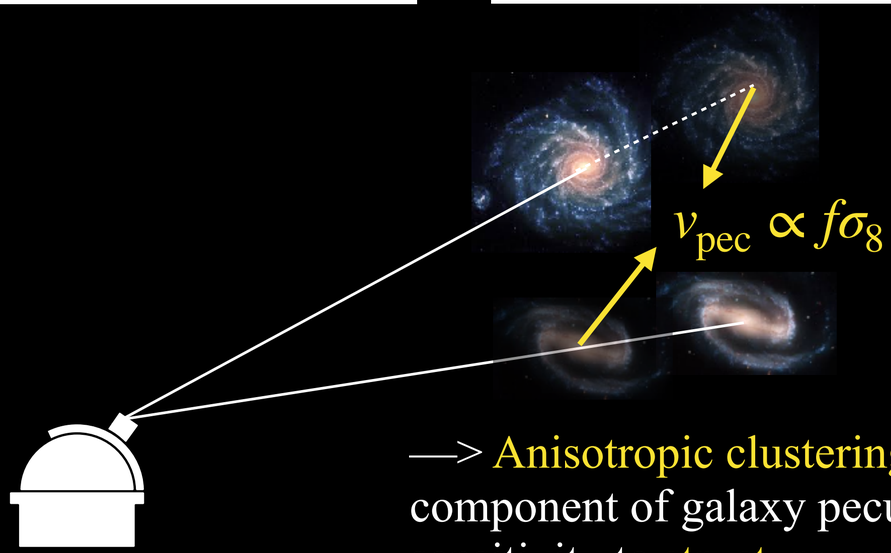
DESI DR1 Galaxy Full-Shape

3D galaxy catalogue (RA, Dec, z) ⟶ Power spectrum $P(k)$ or 2-point correlation function $\xi(s)$ ⟶ Cosmological constraints

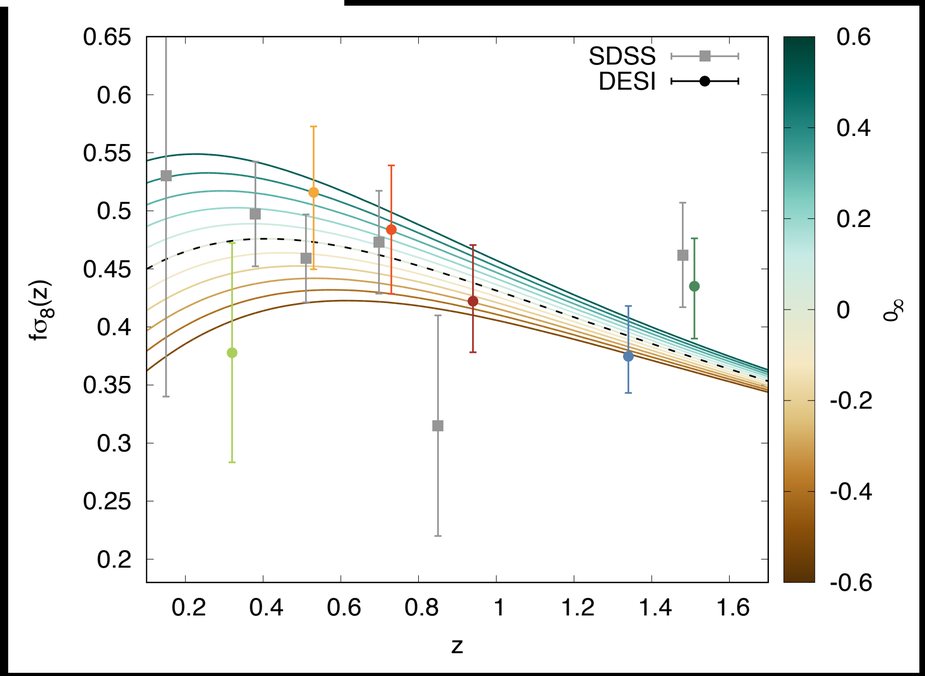




arXiv:2411.12021



—> Anisotropic clustering due to line-of-sight component of galaxy peculiar velocities: sensitivity to structure growth under gravity

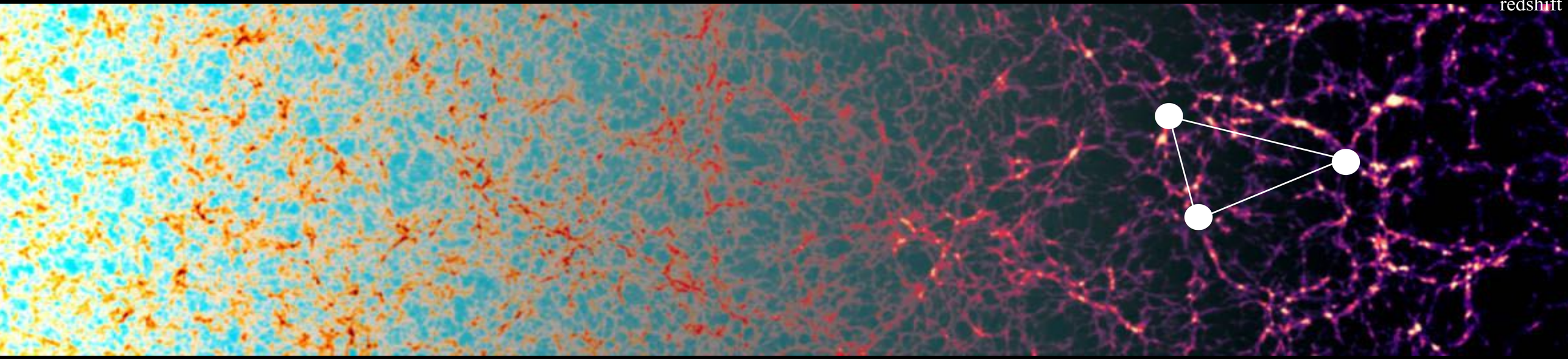


How to extract cosmological information?

Sensitivity to cosmology

Inflation and primordial features — Dark matter and neutrinos — Dark energy

redshift



Initial Gaussian and linear density field

*Today **non-Gaussian** and **non-linear** density field*

← more robust

more optimal →

Two-point statistics (2pt)

- Most mature and robust technique
- Analytical model available up to quasi-linear scales (k_{max}=0.2-0.3 h/Mpc)
- Computationally cheap
- Only Gaussian and linear information

Higher-order statistics (HOS)

- Capture some non-Gaussian information
- Mostly simulation-based model up to k_{max}=0.8 h/Mpc
- Computationally cheaper than FLI but still requires thousands of cosmological simulations
- Factor 2 improvement w.r.t. 2pt*

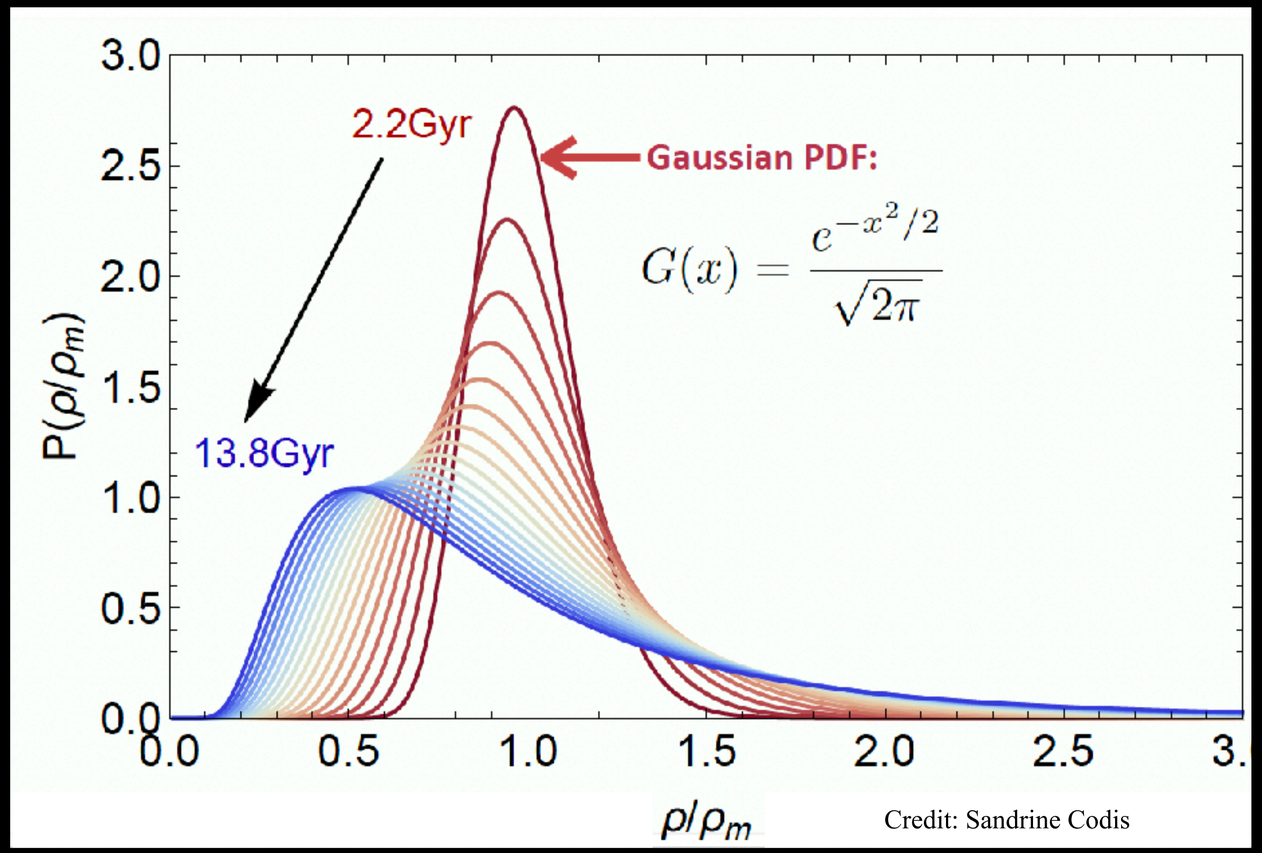
*Euclid collaboration V 2023

DESI DR2 Full-Shape (papers soon)
—> 2pt+3pt statistics / EFT

See Arnaud's talk

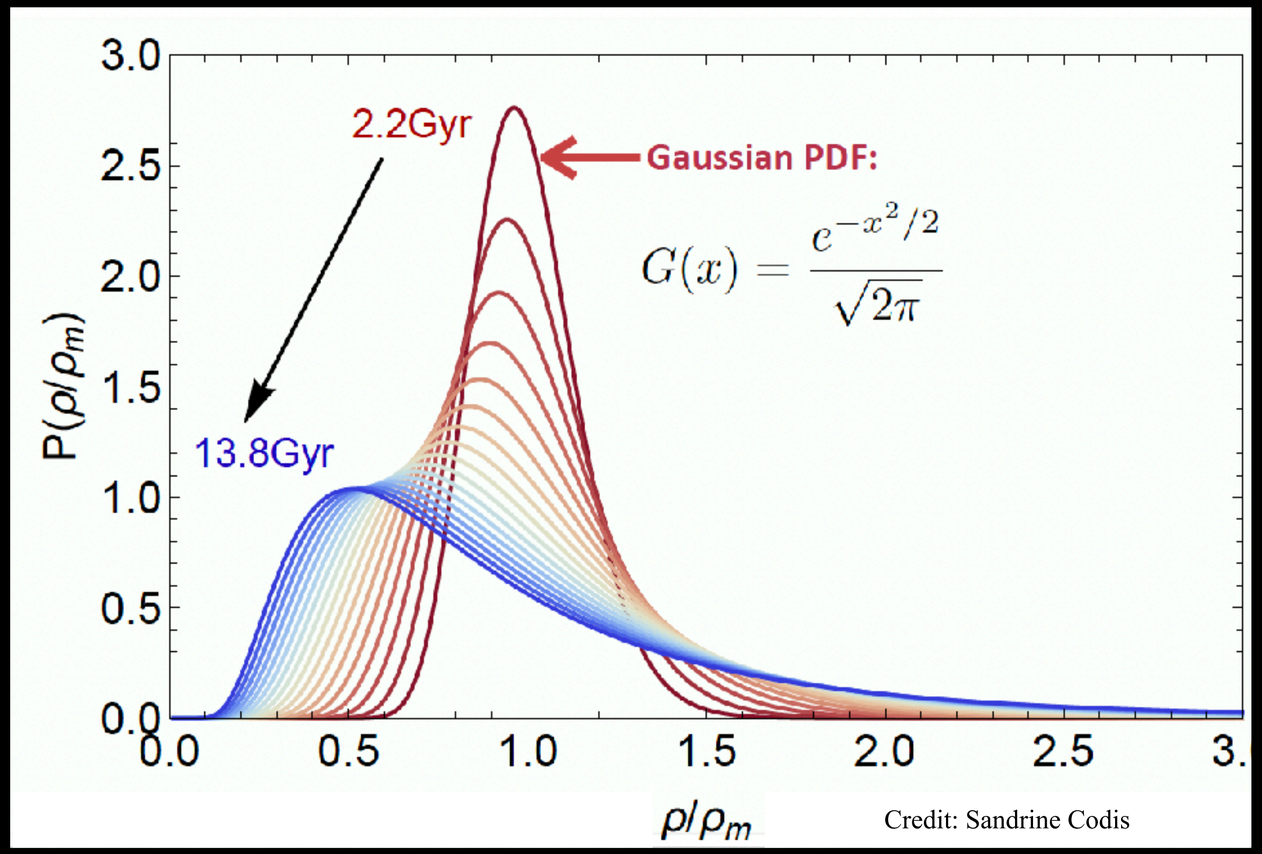
How to go beyond and extract more information?

How to go beyond and extract more information?

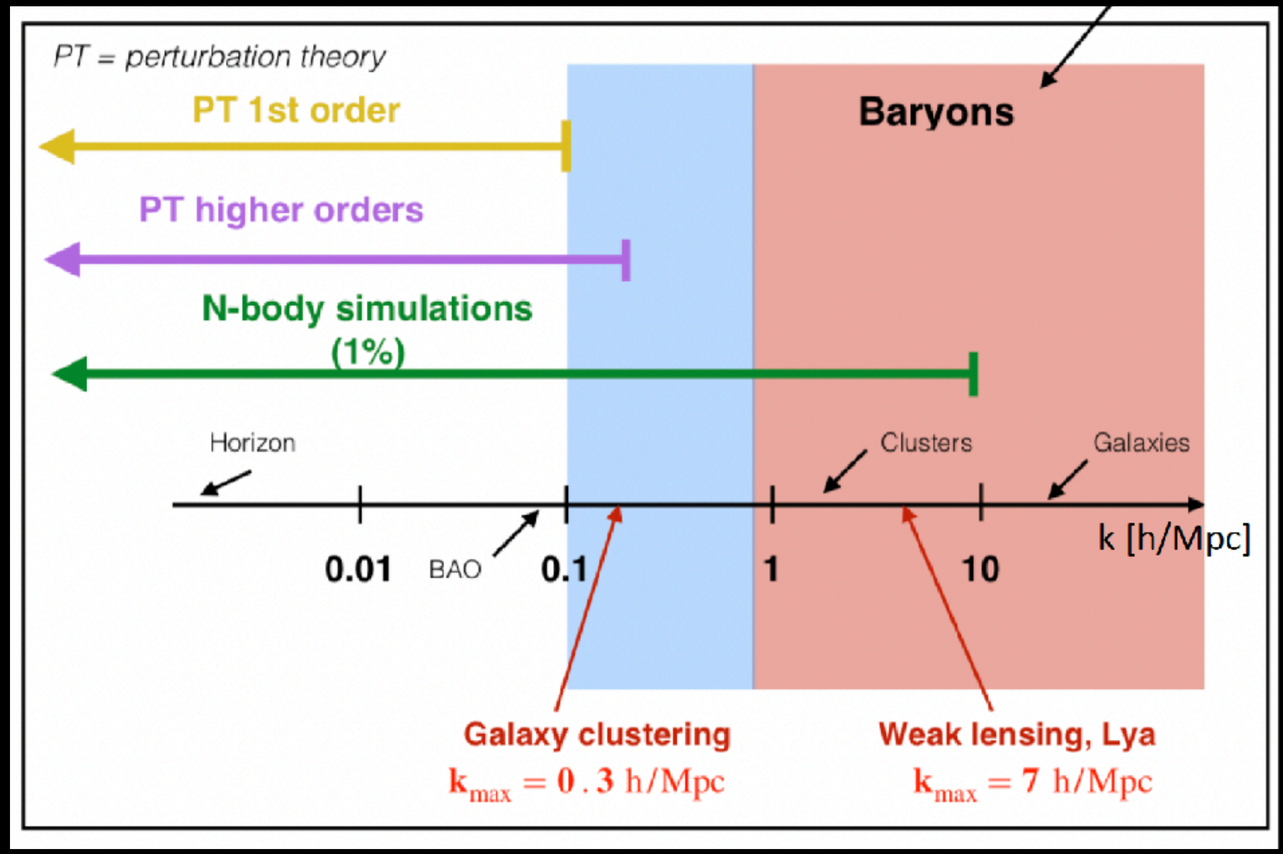


1) Develop other beyond-2pt summary statistics that can capture the non-Gaussian information of the density field

How to go beyond and extract more information?



Credit: Sandrine Codis

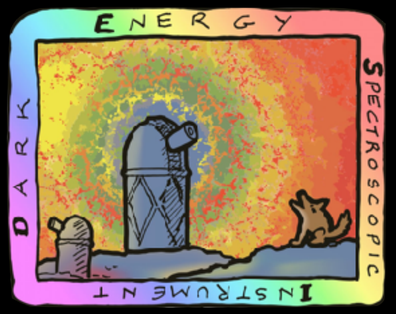


1) Develop other beyond-2pt summary statistics that can capture the non-Gaussian information of the density field

2) Develop simulation-based models that can capture the non-linear information of the density field at smaller scales

Alternative Clustering Methods (ACM) within DESI

ACM group: Gillian Beltz-Mohrmann, Davide Bianchi, Simon Bouchard, Alena Casella, Jean-Marc Colley, Carolina Cuesta-Lazaro, Zhao Chen, Daniel Forero-Sanchez, Nathan Findlay, Katayoon Ghaemi, Elisabeth Krause, Wei Liu, Sesh Nadathur, Krishna Naidoo, Enrique Paillas, Mathilde Pinon, Alice Pisani, Hernan Rincon, Nico Schuster, Georgios Valogiannis, Yu Yu and Pauline Zarrouk.





Alternative Clustering Methods within the DESI



ACM wishing list



One single pipeline for all statistics



One single type of simulation-based model
trained on the same suite of N-body simulations



Standardised pipeline for plug-and-play applications
(unit tests, functional tests, automated integration)
—> help from software engineer is a real asset

Alternative Clustering Methods within the DESI



ACM wishing list

- One single pipeline for all statistics
- One single type of simulation-based model trained on the same suite of N-body simulations
- Standardised pipeline for plug-and-play applications (unit tests, functional tests, automated integration) —> help from software engineer is a real asset



External constraints

- Mock-based covariance
- Explicit likelihood inference (assuming Gaussianity)

Alternative Clustering Methods within the DESI



ACM wishing list



One single pipeline for all statistics



One single type of simulation-based model
trained on the same suite of N-body simulations

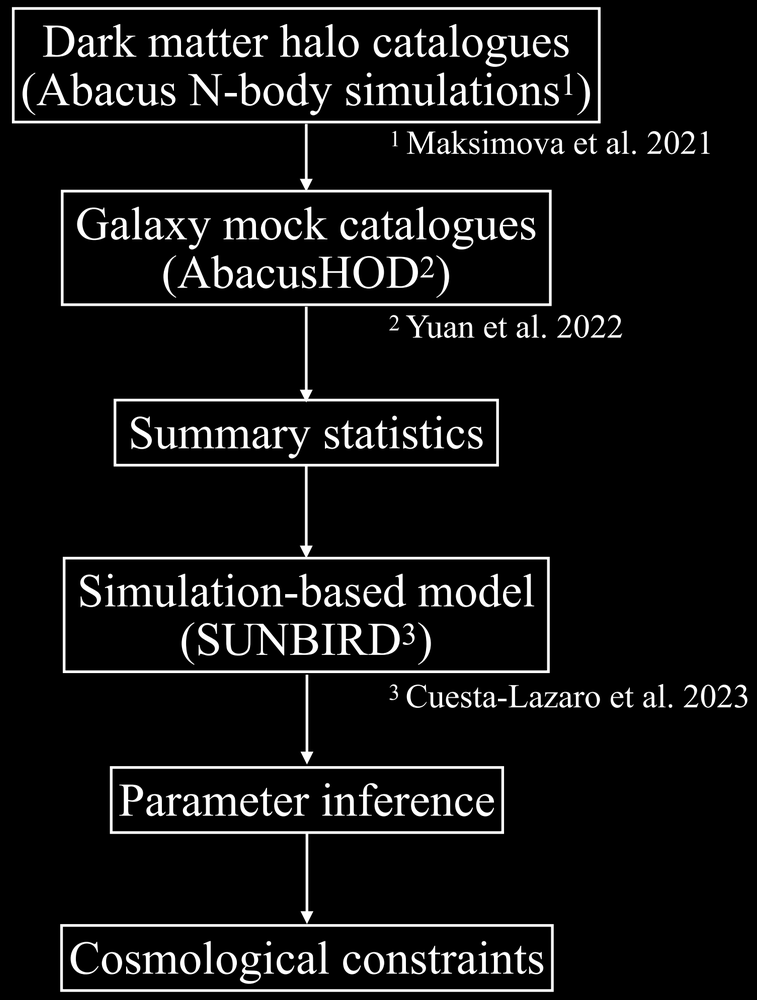


Standardised pipeline for plug-and-play applications
(unit tests, functional tests, automated integration)
—> help from software engineer is a real asset

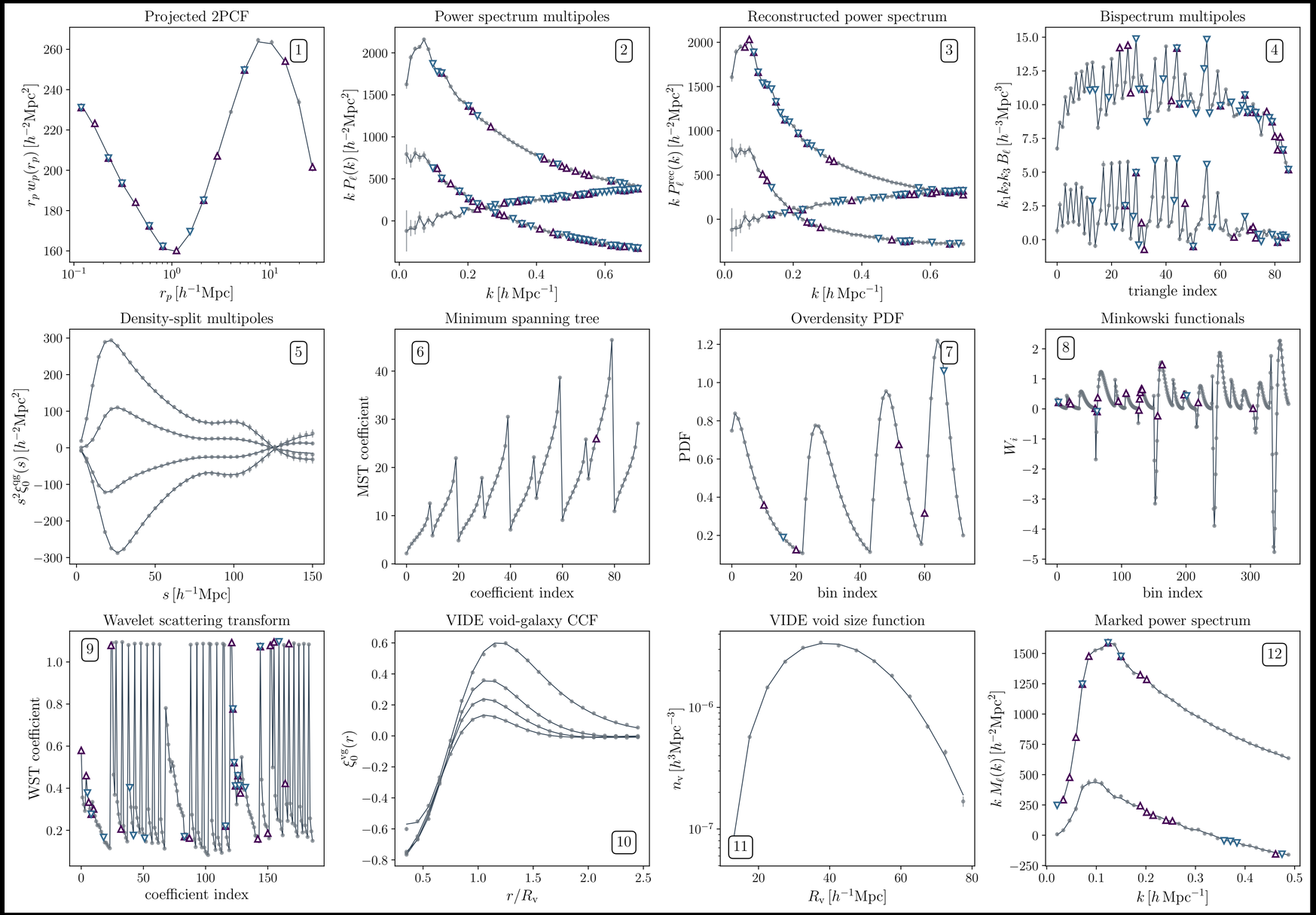


External constraints

- Mock-based covariance
- Explicit likelihood inference (assuming Gaussianity)



Alternative Clustering Methods within the DESI



Dark matter halo catalogues (Abacus N-body simulations[1])

[1] Maksimova et al. 2021

Galaxy mock catalogues (AbacusHOD[2])

[2] Yuan et al. 2022

Summary statistics

Simulation-based model (SUNBIRD[3])

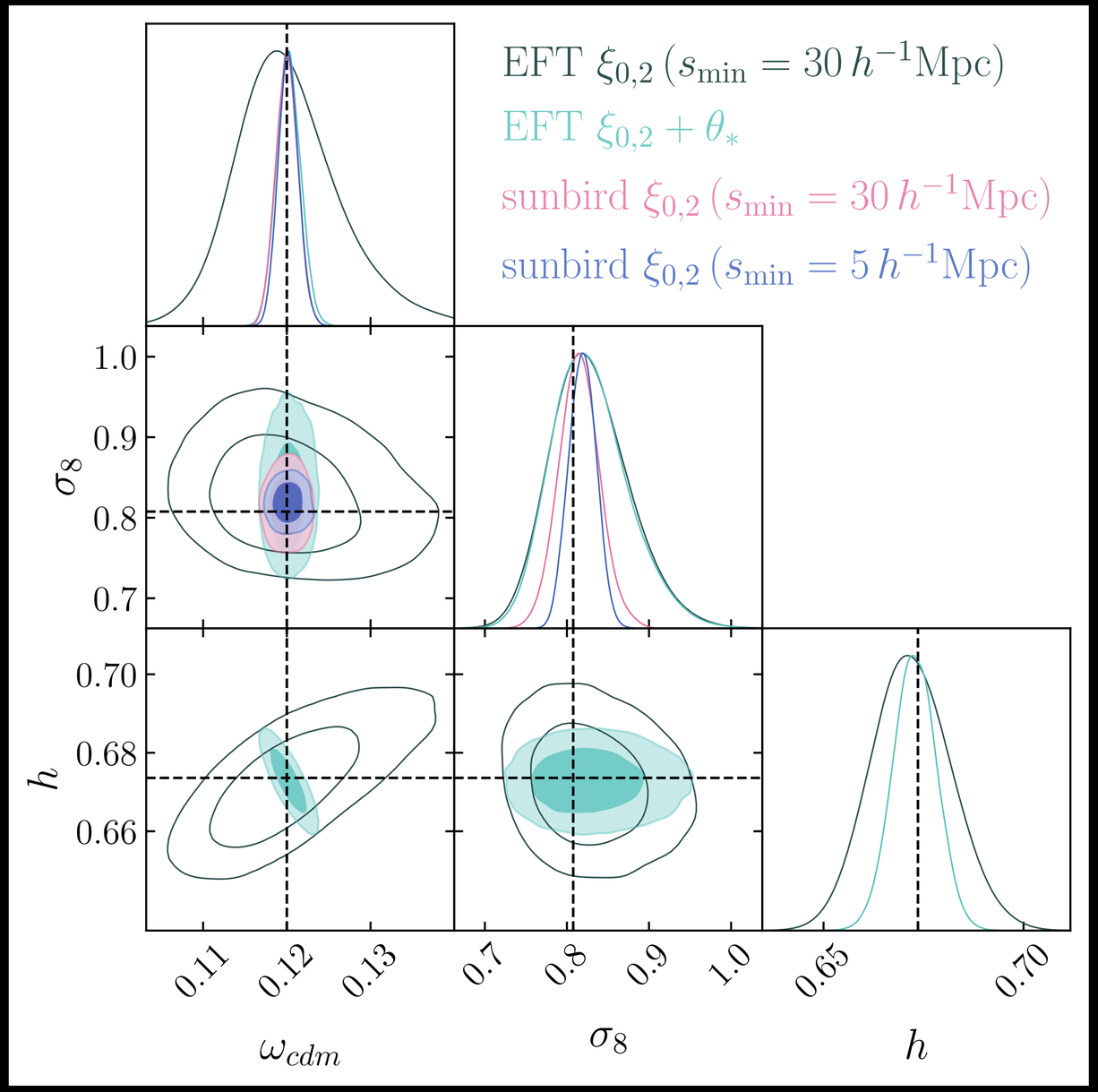
[3] Cuesta-Lazaro et al. 2023

Parameter inference

Cosmological constraints

DESI ACM pipeline, *in prep*

Alternative Clustering Methods within the DESI



DESI ACM pipeline, *in prep*

Dark matter halo catalogues (Abacus N-body simulations[1])

[1] Maksimova et al. 2021

Galaxy mock catalogues (AbacusHOD[2])

[2] Yuan et al. 2022

Summary statistics

Simulation-based model (SUNBIRD[3])

[3] Cuesta-Lazaro et al. 2023

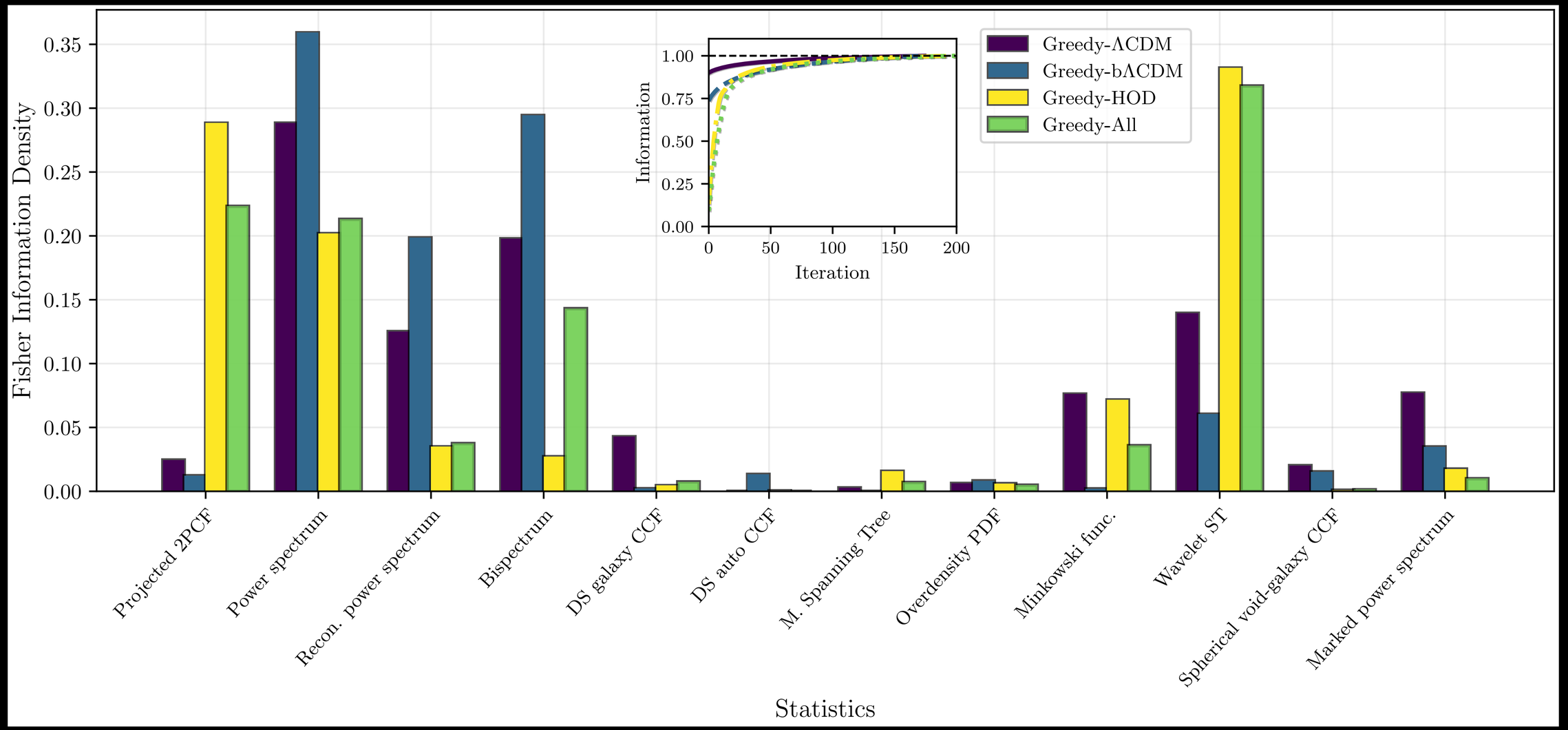
Parameter inference

Cosmological constraints

Alternative Clustering Methods within the DESI

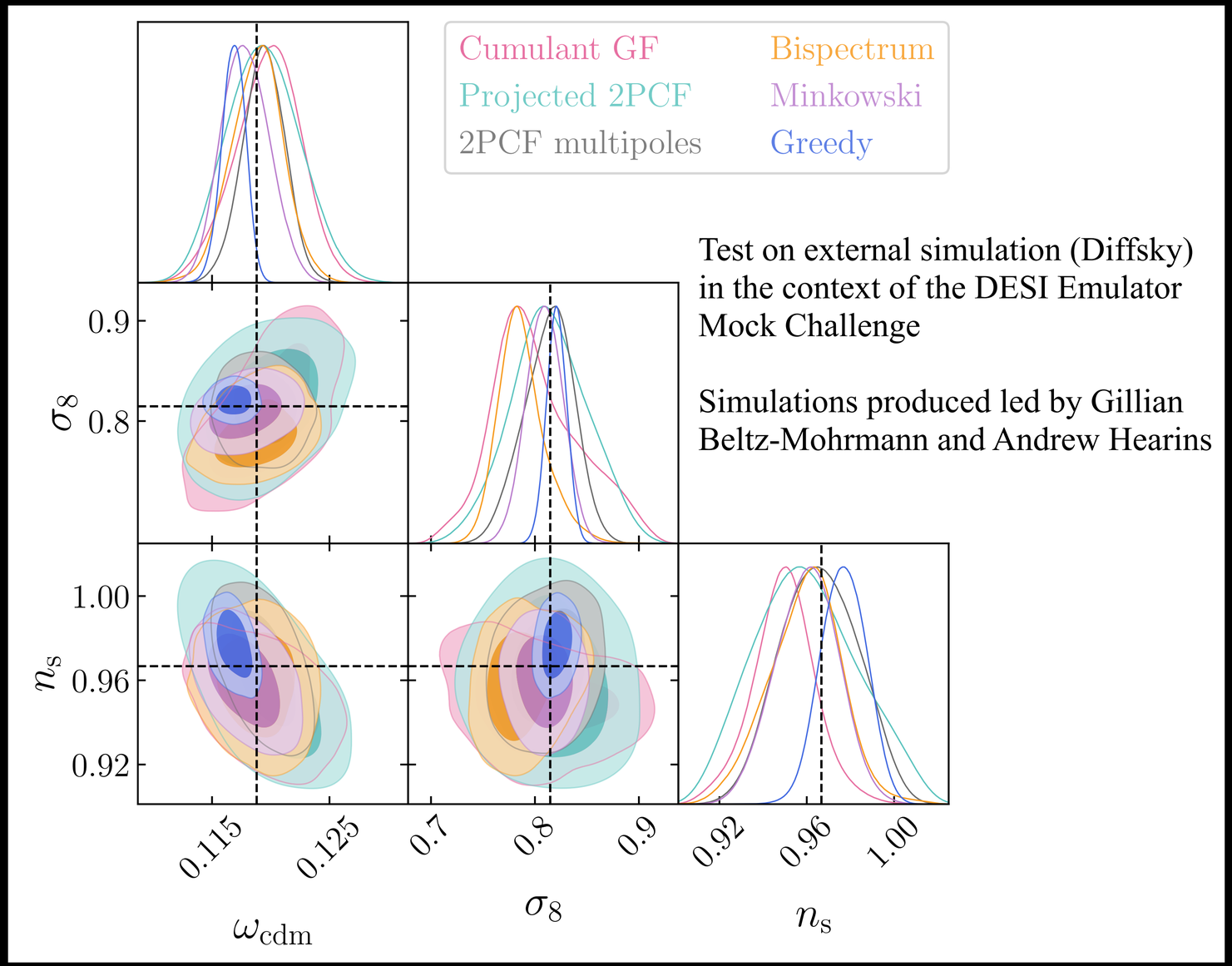
Which set of summary statistics maximises the Fisher information?

Constraint: Keep a data vector of length 200 for mock-based covariance matrix



DESI ACM pipeline, *in prep*

Alternative Clustering Methods within the DESI



DESI Emulator Mock Challenge, Beltz-Mohrmann et al., *under review*

Dark matter halo catalogues (Abacus N-body simulations[1])

[1] Maksimova et al. 2021

Galaxy mock catalogues (AbacusHOD[2])

[2] Yuan et al. 2022

Summary statistics

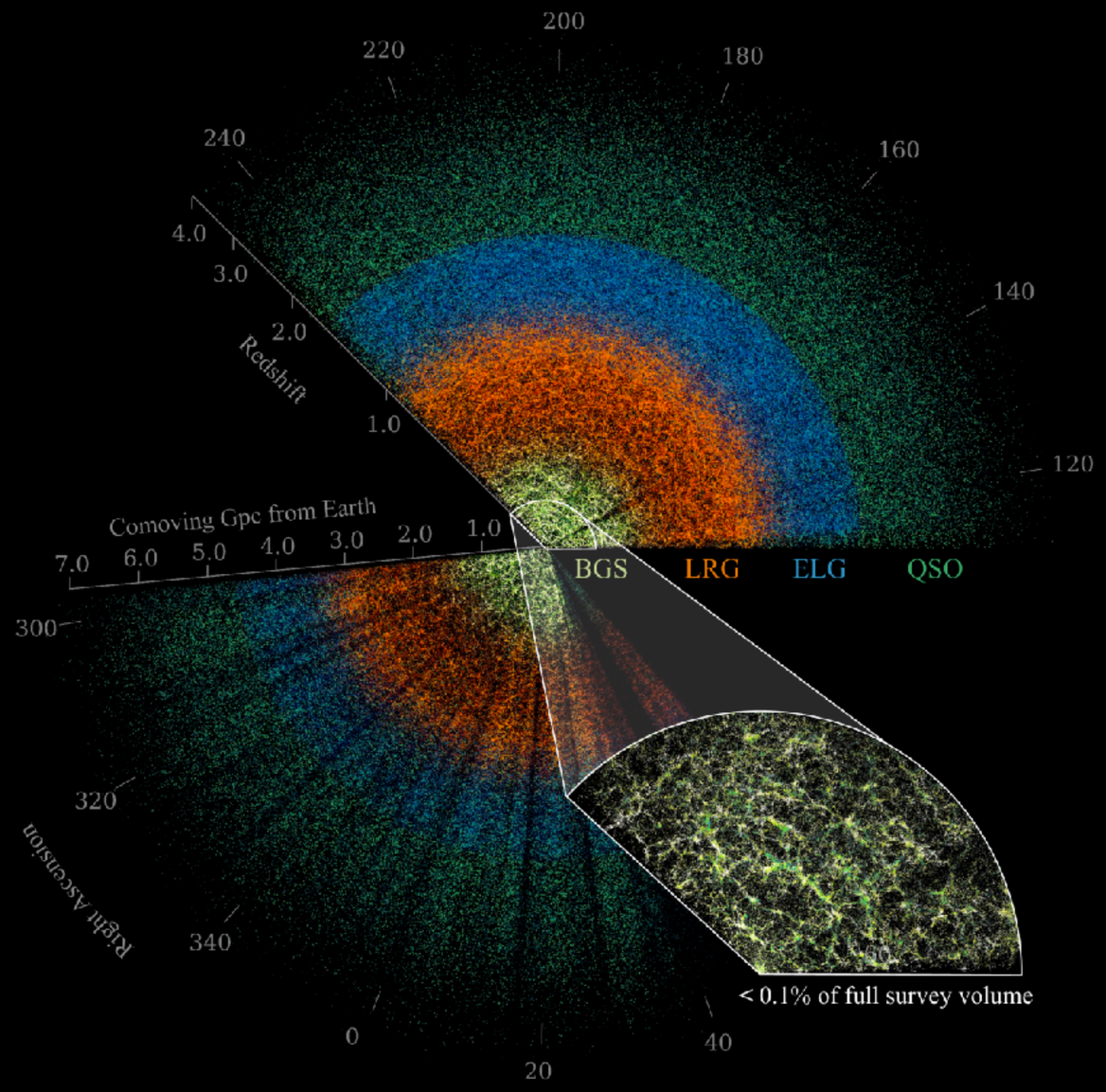
Simulation-based model (SUNBIRD[3])

[3] Cuesta-Lazaro et al. 2023

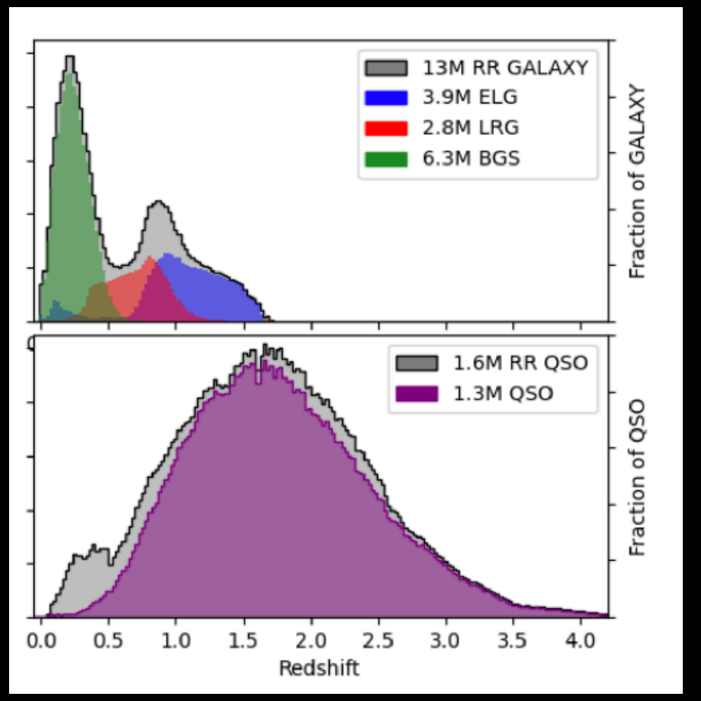
Parameter inference

Cosmological constraints

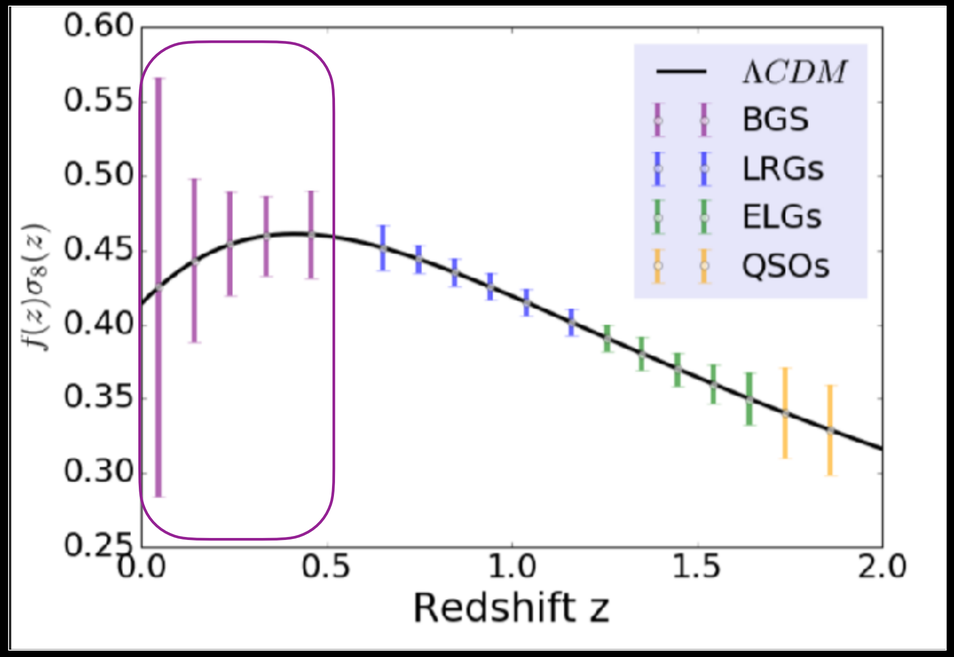
Special effort for the DESI BGS



DESI DR1, DESI collaboration 2025



DESI DR1, DESI collaboration 2025



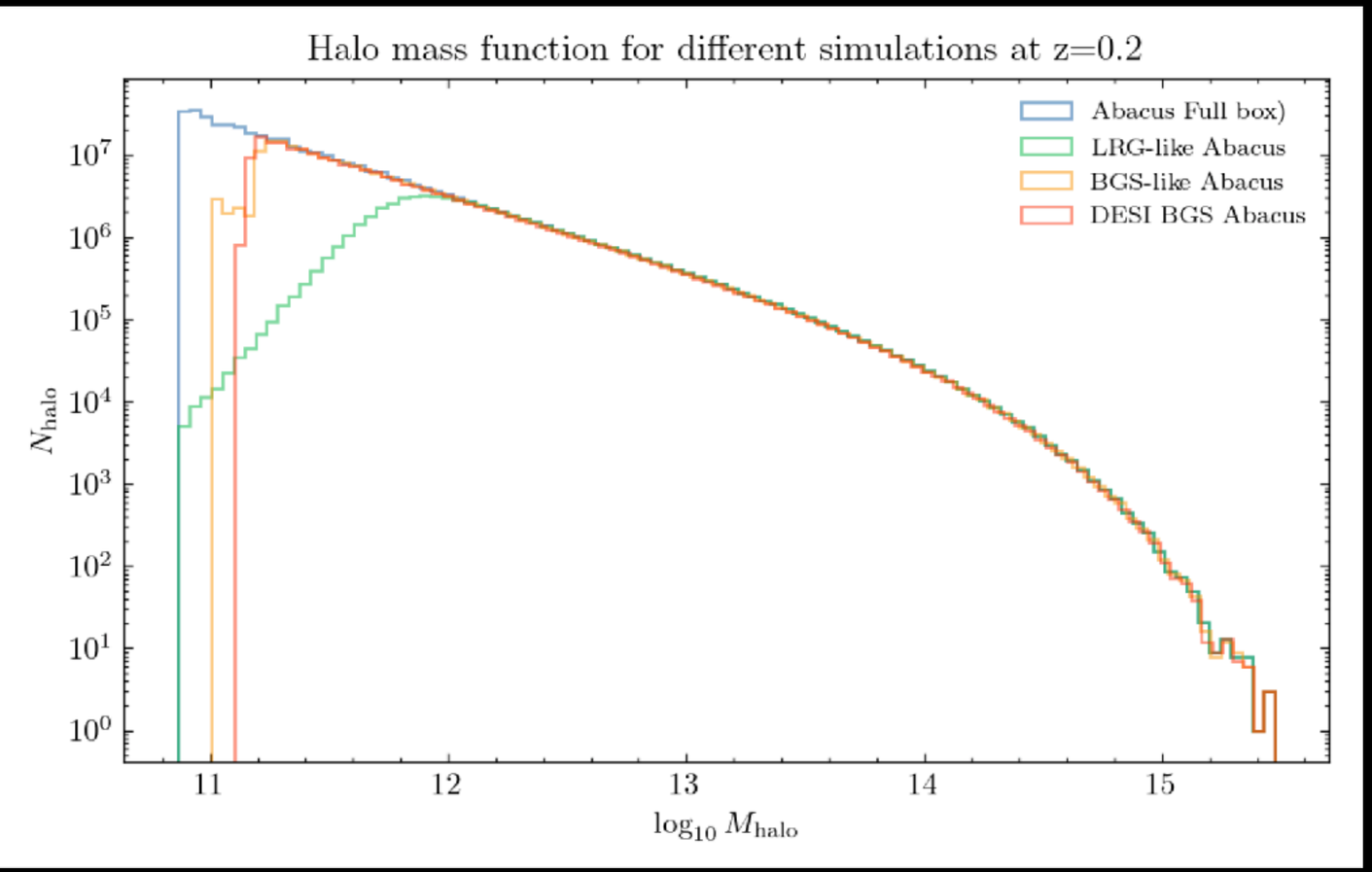
DESI collaboration 2016a

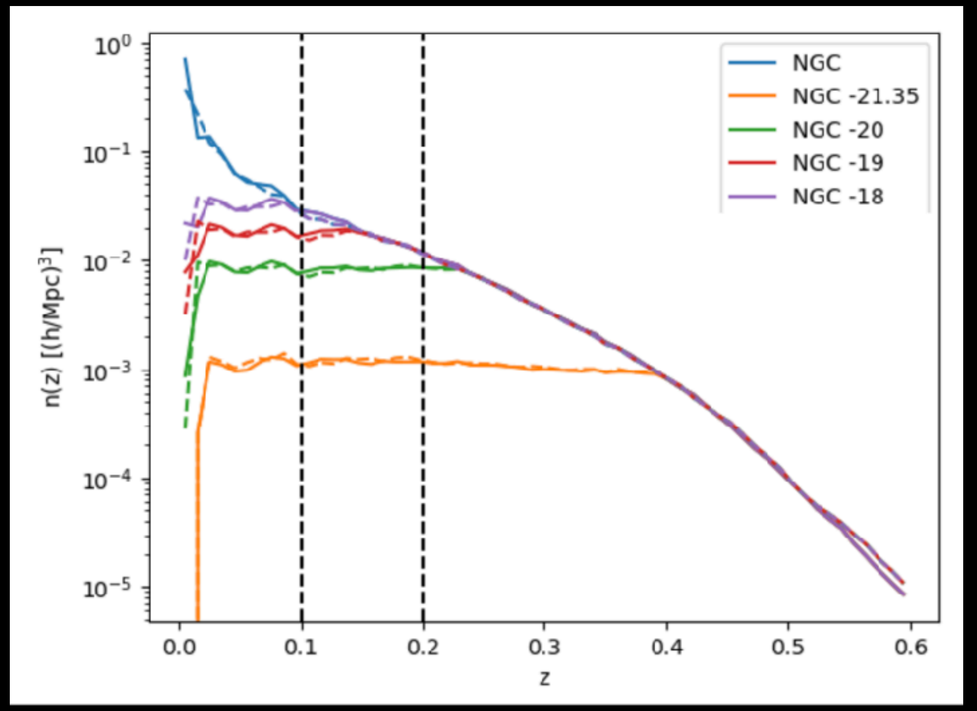
Alternative Clustering Methods with the DESI BGS

—> Specific needs for the **Bright Galaxy Survey**

Dark matter halo catalogues
(Abacus N-body simulations[1])

[1] Maksimova et al. 2021



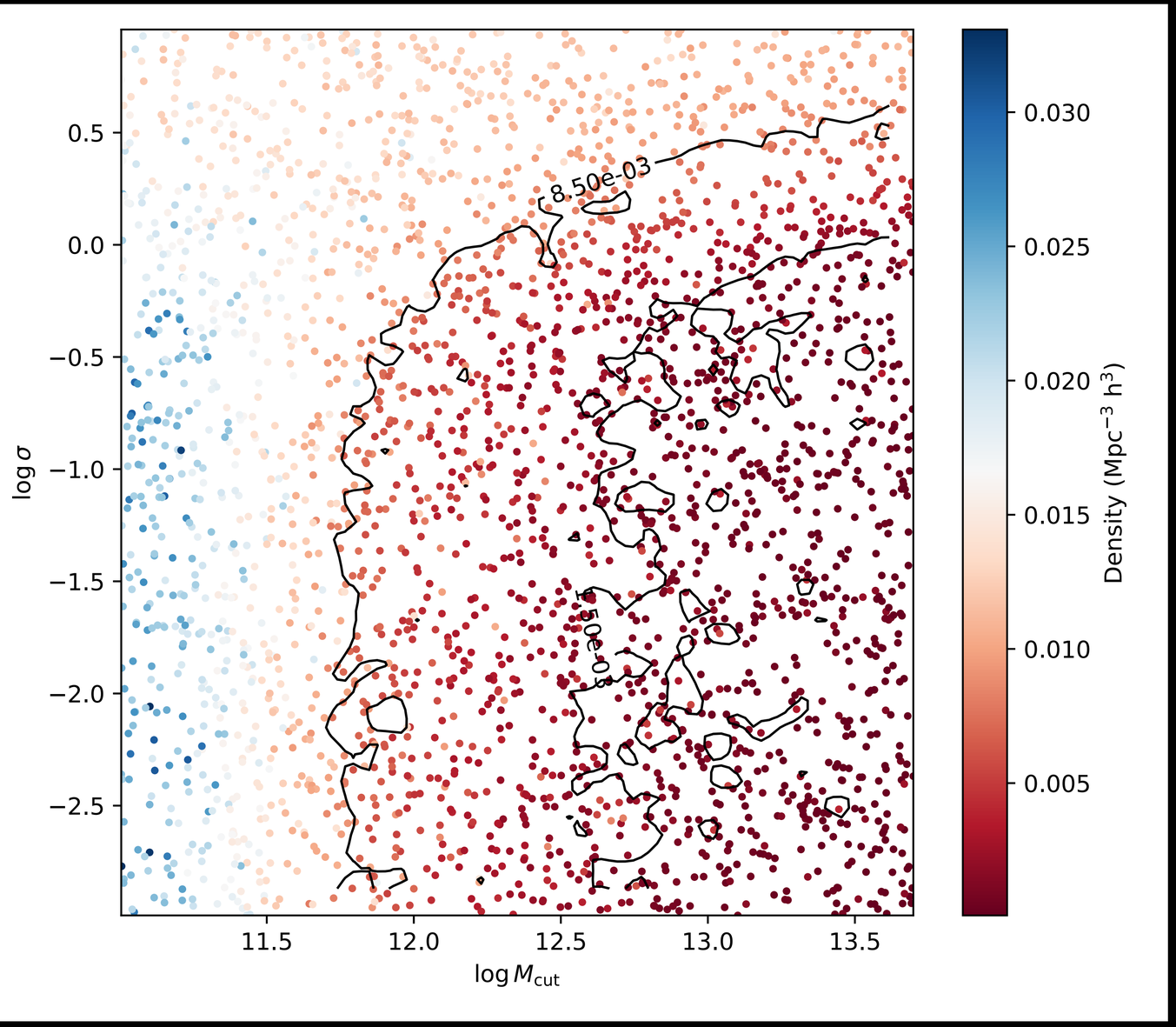




Simon Bouchard (2024-2027)
Optimal summary statistics for BGS

Bouchard, Zarrouk et al. *in prep*

Alternative Clustering Methods with the DESI BGS



Bouchard, Zarrouk et al. *in prep*

Dark matter halo catalogues
(Abacus N-body simulations[1])

[1] Maksimova et al. 2021

Galaxy mock catalogues
(AbacusHOD[2])

[2] Yuan et al. 2022

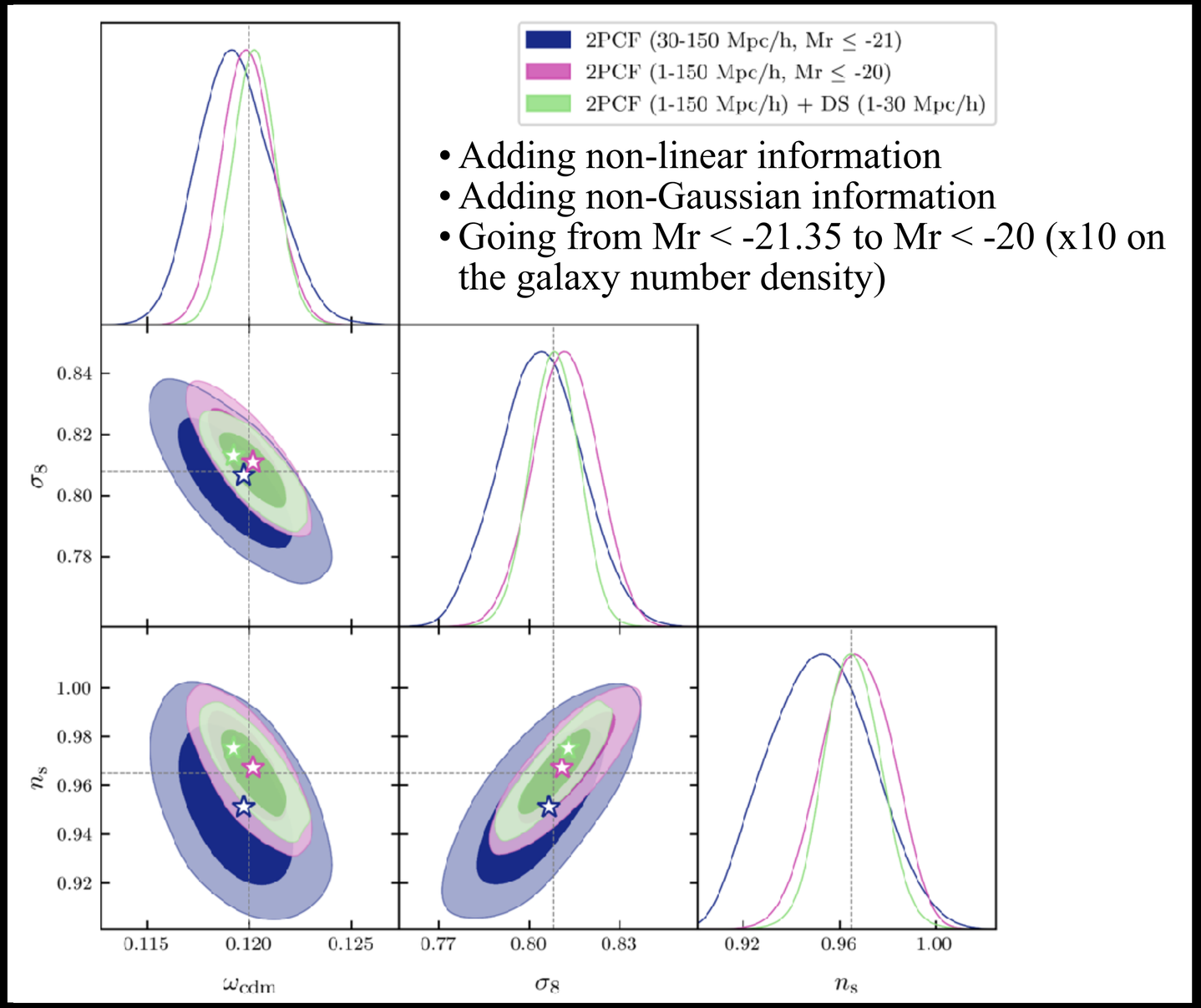
BGS storage
- dark matter catalogues: 8.5 Tb
- galaxy catalogues: ∼50 Tb —> not saved, analysis on-fly
- Statistics measurement ∼1 Tb

BGS computational resources
- 800 CPU hours for producing the dark matter halos catalogues
- 1200 GPU hours for generating the training set and computing the statistics



Simon Bouchard (2024-2027)
Optimal summary statistics for BGS

Alternative Clustering Methods with the DESI BGS



Bouchard, Zarrouk et al. *in prep*

Dark matter halo catalogues (Abacus N-body simulations[1])

[1] Maksimova et al. 2021

Galaxy mock catalogues (AbacusHOD[2])

[2] Yuan et al. 2022

Summary statistics

Simulation-based model (SUNBIRD[3])

[3] Cuesta-Lazaro et al. 2023

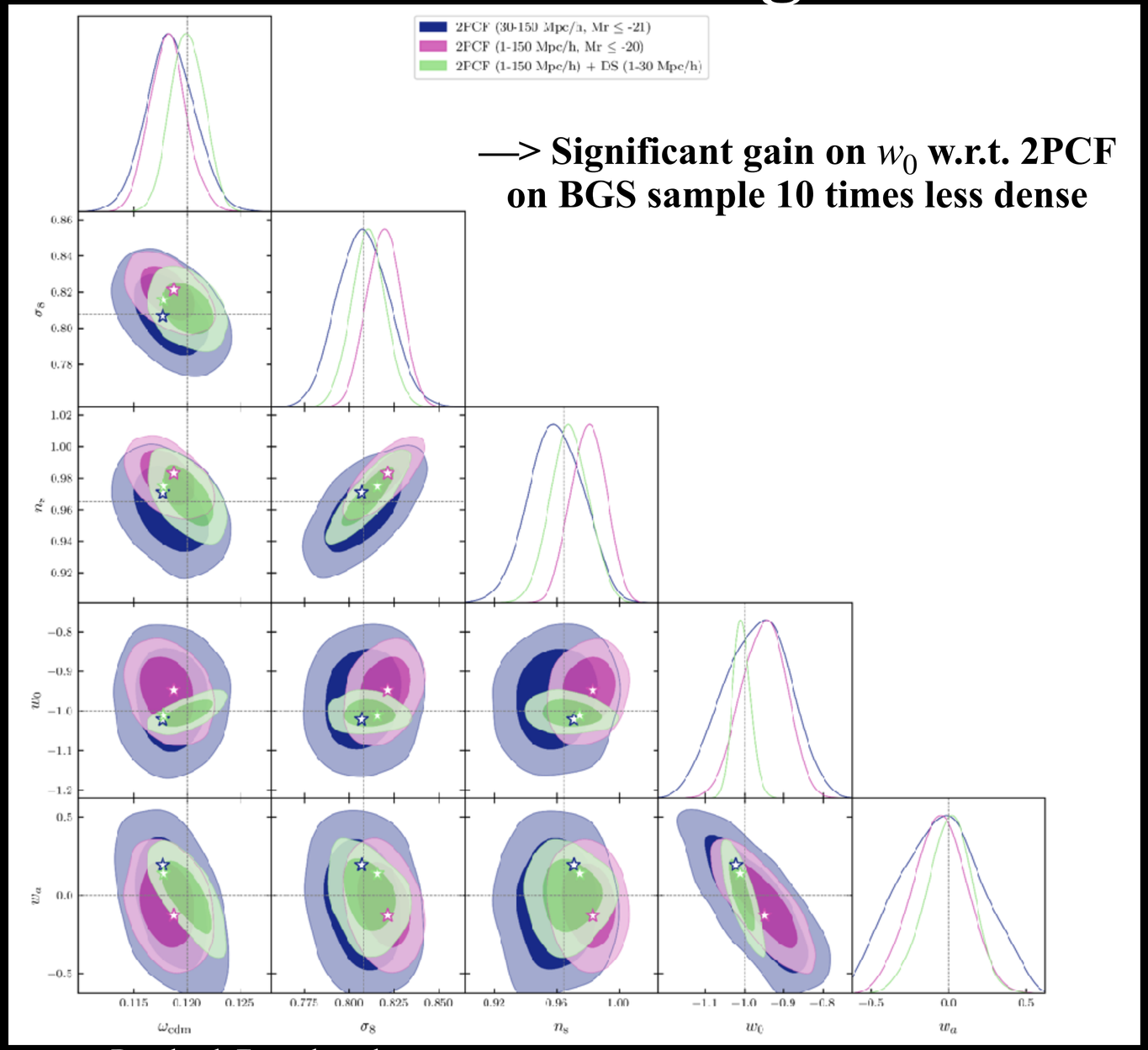
Parameter inference

Cosmological constraints ΛCDM



Simon Bouchard (2024-2027)
Optimal summary statistics for BGS

Alternative Clustering Methods with the DESI BGS



—> Significant gain on w_0 w.r.t. 2PCF on BGS sample 10 times less dense

Bouchard, Zarrouk et al. *in prep*

Dark matter halo catalogues (Abacus N-body simulations[1])

[1] Maksimova et al. 2021

↓

Galaxy mock catalogues (AbacusHOD[2])

[2] Yuan et al. 2022

↓

Summary statistics

↓

Simulation-based model (SUNBIRD[3])

[3] Cuesta-Lazaro et al. 2023

↓

Parameter inference

↓

Cosmological constraints $w_o w_a$CDM



Simon Bouchard (2024-2027)
Optimal summary statistics for BGS

Alternative Clustering Methods within DESI

Where are we today?



ACM working (cubic-box) pipeline for 12 summary statistics
—> two papers under review: ACM pipeline, emulator mock challenge



Special effort for BGS: methodology paper in preparation (Bouchard, Zarrouk et al.)



Standardisation and robustification of the entire ACM pipeline
(unit tests, functional tests, automated integration)

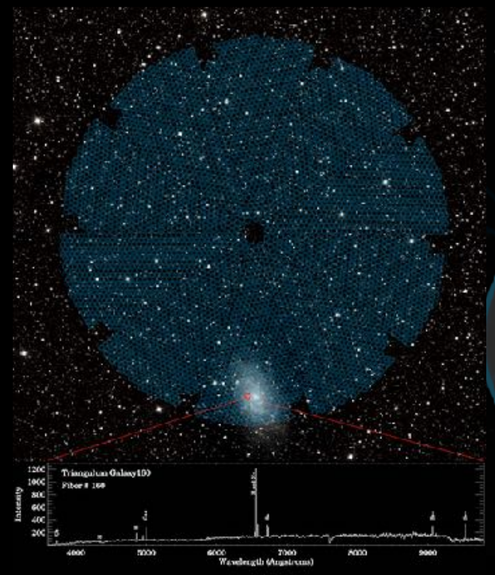
Alternative Clustering Methods within DESI

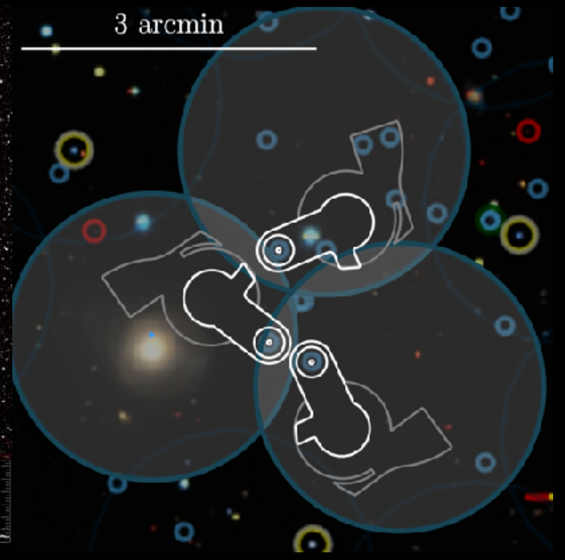
Where are we today?

- ACM working (cubic-box) pipeline for 12 summary statistics
  —> two papers under review: ACM pipeline, emulator mock challenge
- Special effort for BGS: methodology paper in preparation (Bouchard, Zarrouk et al.)
- Standardisation and robustification of the entire ACM pipeline
  (unit tests, functional tests, automated integration)

Towards DESI DR2 application

- Include DESI survey geometry
- Include DESI fibre-assignment





Fast emulation of DESI fibre assignment

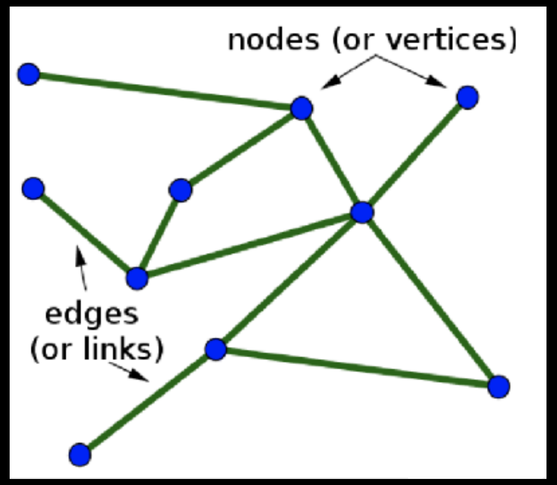
—> Use auto-regressive graph neural network to predict which targets will be observed and which won't



With Alena Casella and Elisabeth Krause (University of Arizona), Carolina Cuesta-Lazaro (CCA Flatiron)

Fast emulation of DESI fibre assignment

—> Use auto-regressive graph neural network to predict which targets will be observed and which won't



Nodes: galaxy target that contains input info (position, number of close neighbours, observation state, etc.)

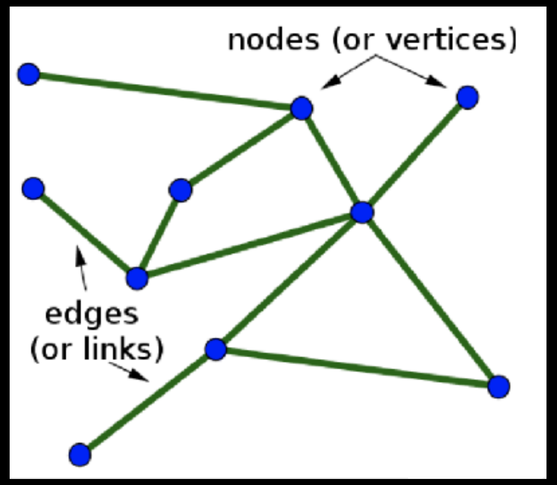
Edges: connects pairs in the patch; separation: dx, dy, distance
—> how the local environnement affects the fibre assignment



With Alena Casella and Elisabeth Krause (University of Arizona), Carolina Cuesta-Lazaro (CCA Flatiron)

Fast emulation of DESI fibre assignment

—> Use auto-regressive graph neural network to predict which targets will be observed and which won't



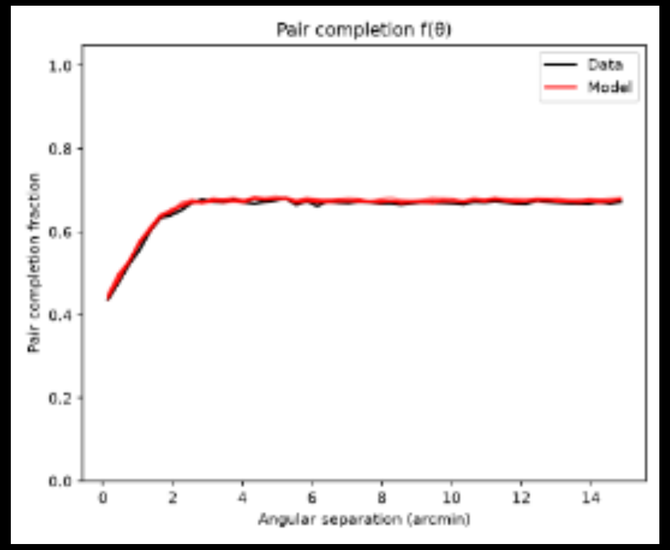
Nodes: galaxy target that contains input info (position, number of close neighbours, observation state, etc.)

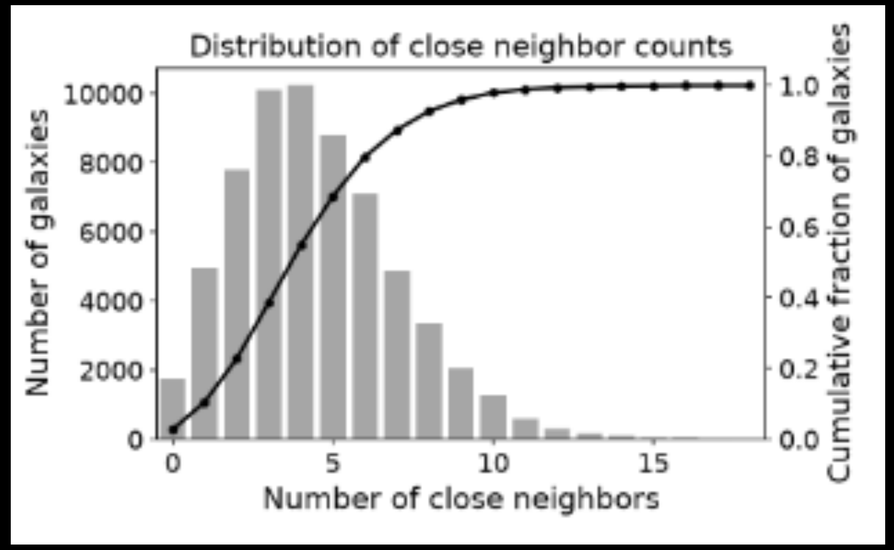
Edges: connects pairs in the patch; separation: dx, dy, distance
—> how the local environnement affects the fibre assignment



With Alena Casella and Elisabeth Krause (University of Arizona), Carolina Cuesta-Lazaro (CCA Flatiron)

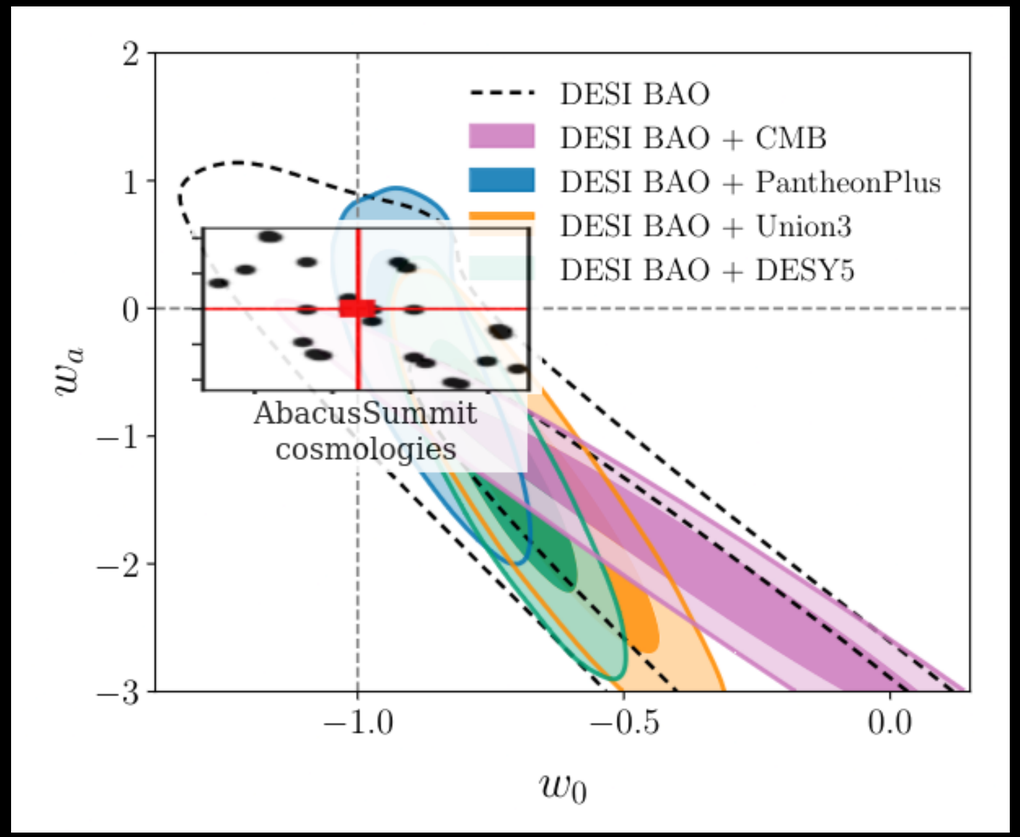






Alternative Clustering Methods within DESI

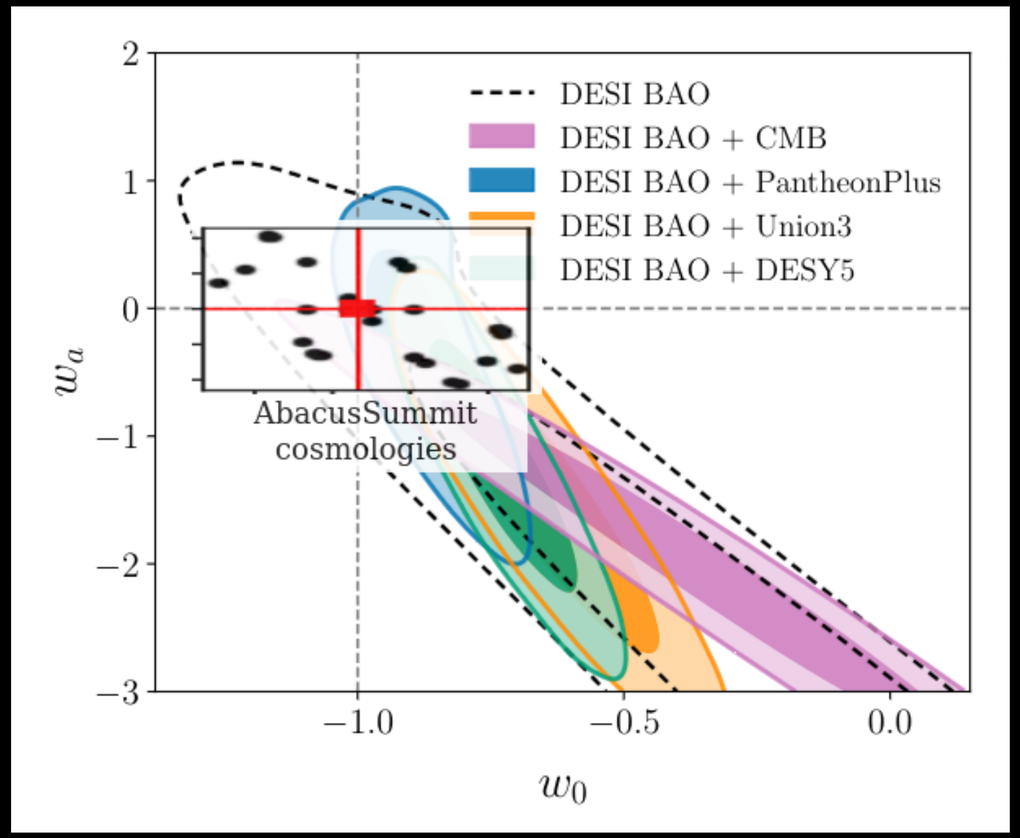
Constraining $w_0 w_a$CDM



—> Limited $w_o - w_a$ parameter space due to the input suite of simulations (AbacusSummit range)

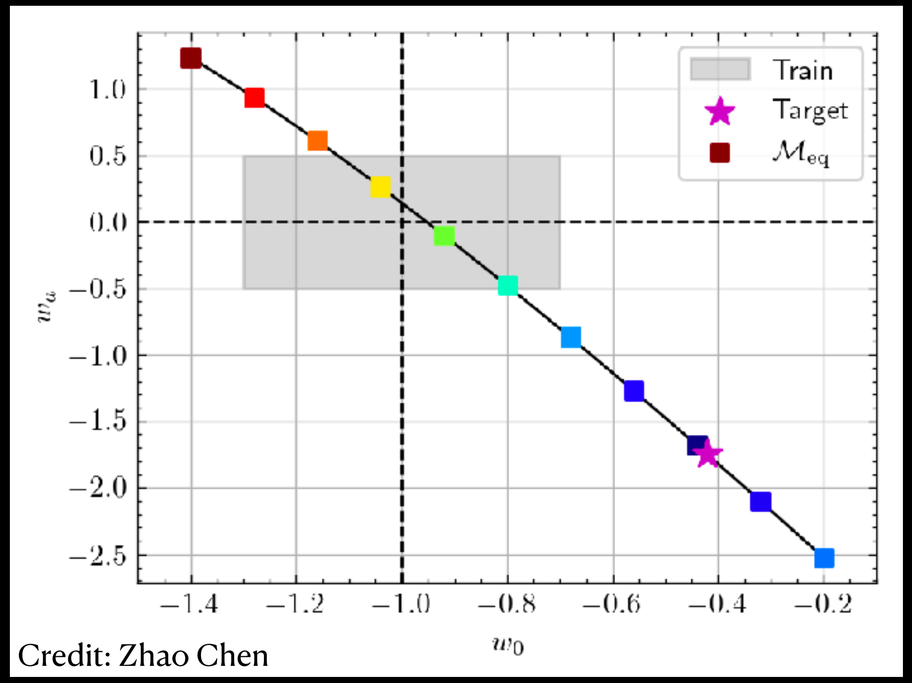
Alternative Clustering Methods within DESI

Constraining $w_0 w_a$CDM



—> Limited $w_o - w_a$ parameter space due to the input suite of simulations (AbacusSummit range)

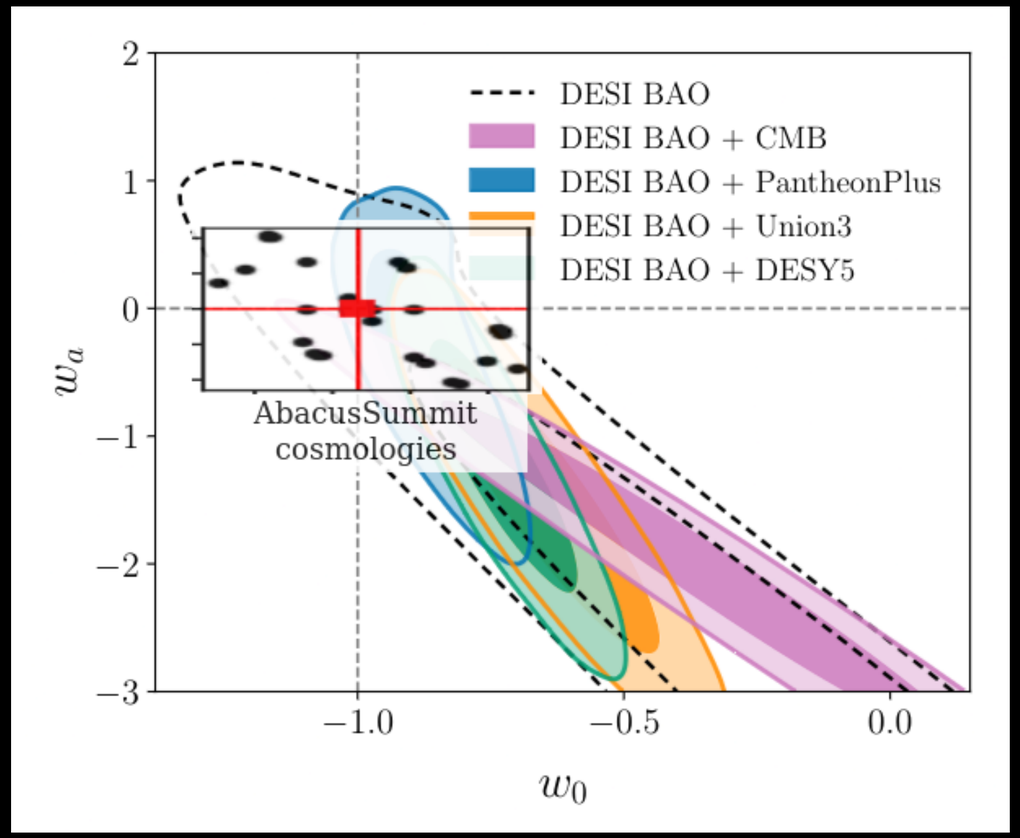
- One avenue to extend SUNBIRD $w_o - w_a$ parameter space: Spectral equivalence method (Zhao Chen & Yu Yu 2025)



Credit: Zhao Chen

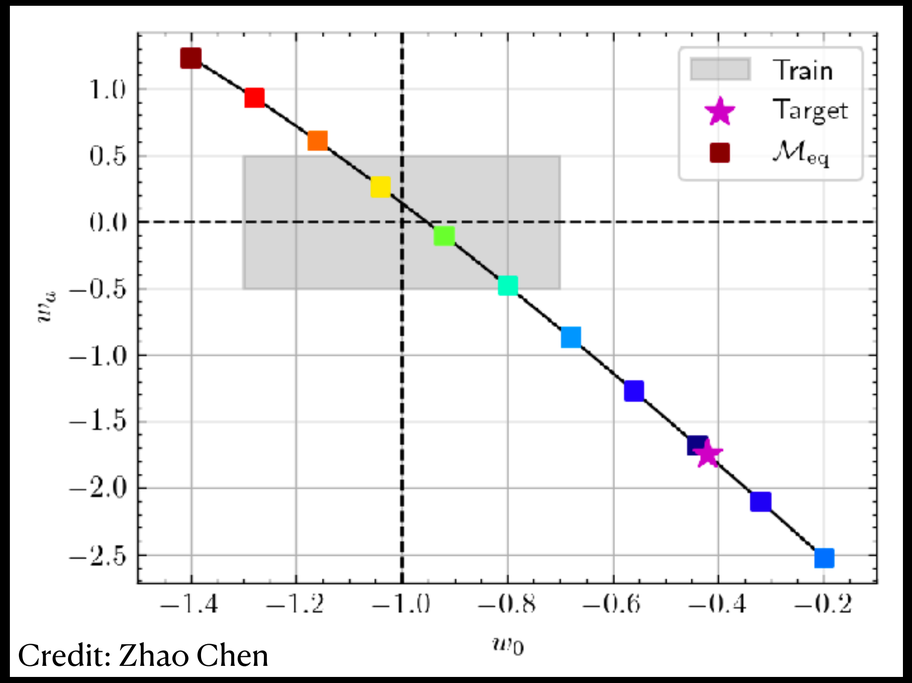
Alternative Clustering Methods within DESI

Constraining $w_0 w_a$CDM



—> Limited $w_o - w_a$ parameter space due to the input suite of simulations (AbacusSummit range)

- One avenue to extend SUNBIRD $w_o - w_a$ parameter space: Spectral equivalence method (Zhao Chen & Yu Yu 2025)



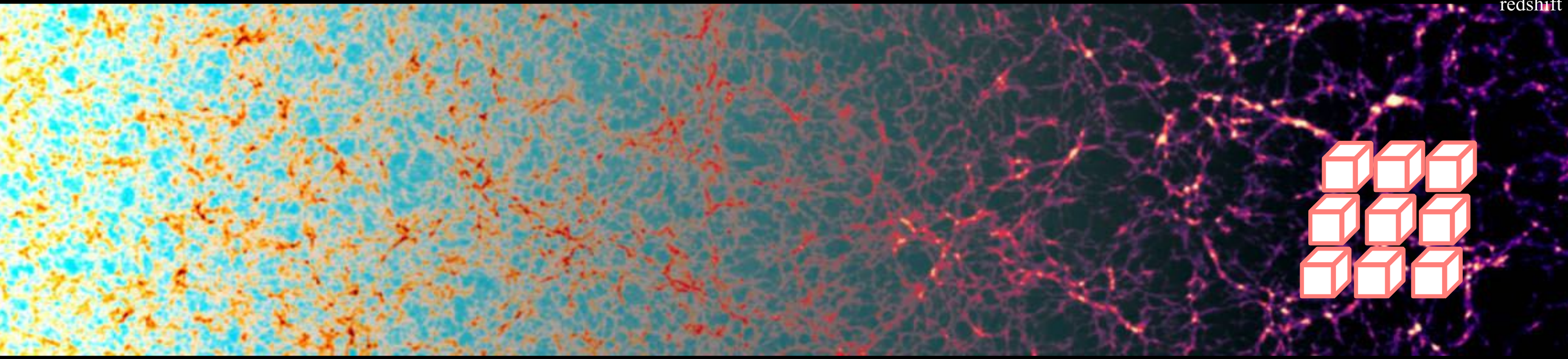
- Another long-term avenue (for DESI DR3): new suite of Abacus simulations should be available next year

How to extract cosmological information?

Sensitivity to cosmology

Inflation and primordial features — Dark matter and neutrinos — Dark energy

redshift



Initial Gaussian and linear density field

*Today **non-Gaussian** and **non-linear** density field*

← more robust

more optimal →

Two-point statistics (2pt)

- Most mature and robust technique
- Analytical model available up to quasi-linear scales (k_{max}=0.2 h/Mpc)
- Computationally cheap
- Only Gaussian and linear information

Higher-order statistics (HOS)

- Capture some non-Gaussian information
- Mostly simulation-based model up to k_{max}=0.8 h/Mpc
- Computationally cheaper than FLI but still requires thousands of cosmological simulations
- Factor 2 improvement w.r.t. 2pt*

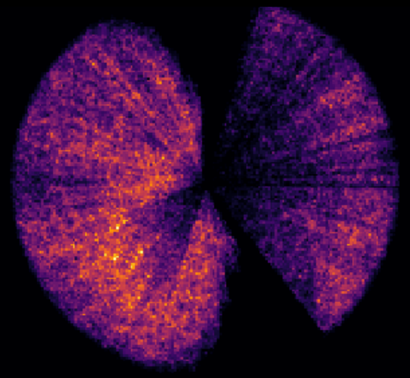
*Euclid collaboration V 2023

Field-level inference (FLI)

- Retrieve the maximal information & deliver constrained simulations
- Grid resolution up to k_{max}=0.8 h/Mpc
- Computationally very expensive
- Factor 2-5 improvement w.r.t. HOS**?

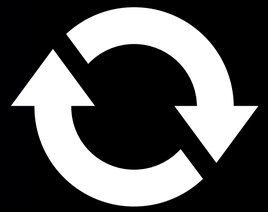
** Nguyen et al. 2025, Porqueres et al. 2023

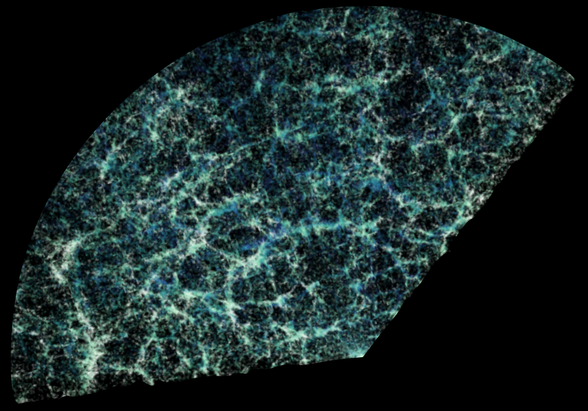
Bayesian field-level inference: how does it work?



Cosmic variance-aligned simulated
DESI DR1 Universe

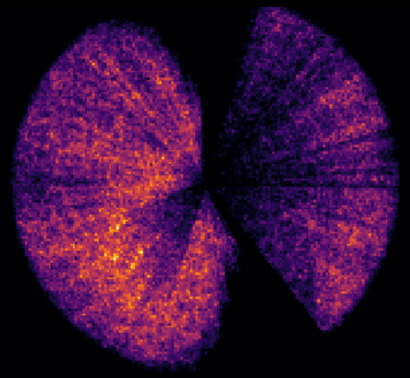
Voxel-by-voxel
comparison





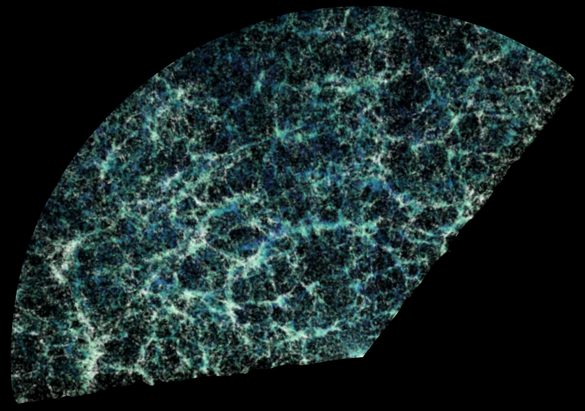
Real DESI DR1 Universe

Bayesian field-level inference: how does it work?



Cosmic variance-aligned simulated
DESI DR1 Universe

Voxel-by-voxel
comparison



Real DESI DR1 Universe

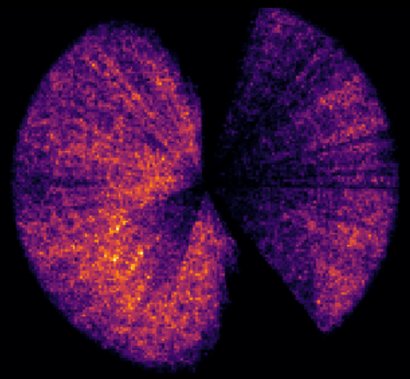
Three main ingredients:

1. Forward model (or the simulator)
2. (explicit) likelihood
3. Gradient-based sampling

Three main challenges:

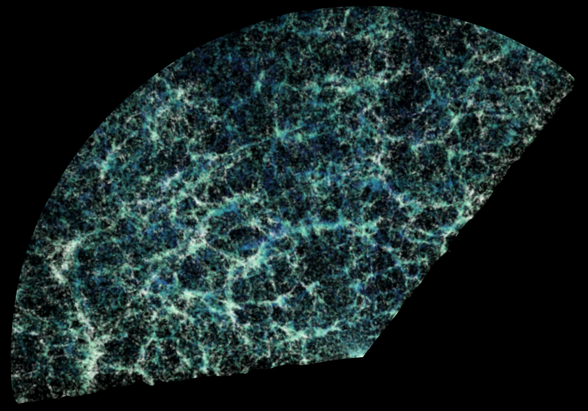
1. Model-mispecification
2. Scalability with larger and denser datasets
3. Multi-million dimensional parameter space

Bayesian field-level inference: how does it work?



Cosmic variance-aligned simulated DESI DR1 Universe

Voxel-by-voxel comparison



Real DESI DR1 Universe

Three main ingredients:

1. Forward model (or the simulator)
2. (explicit) likelihood
3. Gradient-based sampling

Three main challenges:

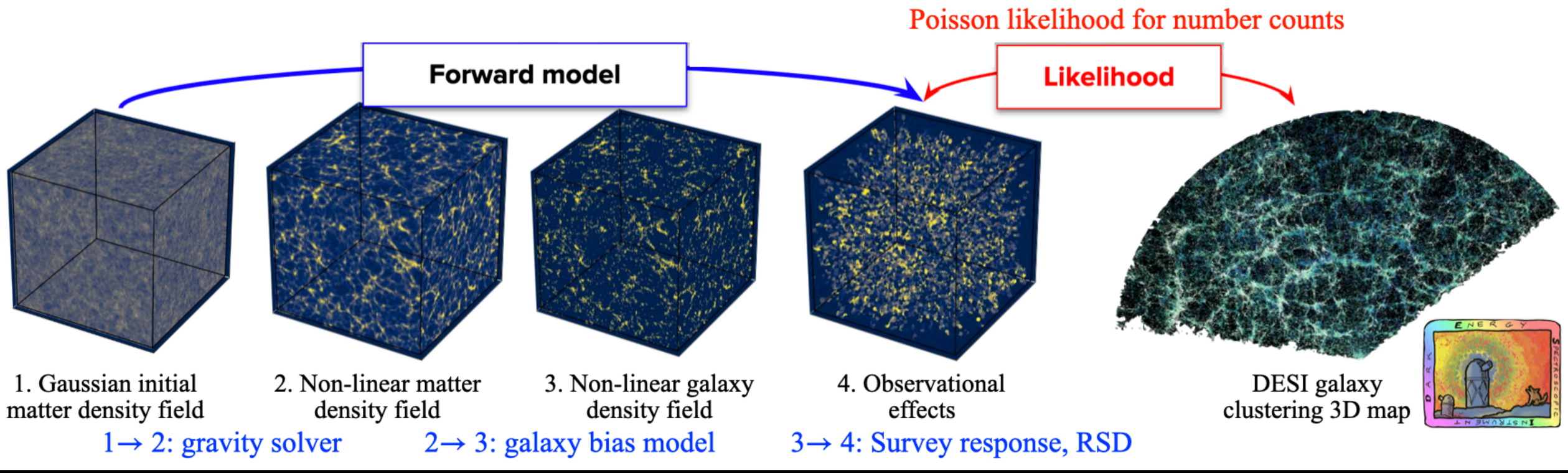
1. Model-mispecification
2. Scalability with larger and denser datasets
3. Multi-million dimensional parameter space

Digital twins of the Universe (cosmic variance-aligned / constrained simulations) offer a unique way of confronting theory against data

- Detailed probabilistic maps of the matter distribution and its history
- Access all latent variables (velocity field, tidal field, …)
- Posterior predictive tests and cross-validation checks

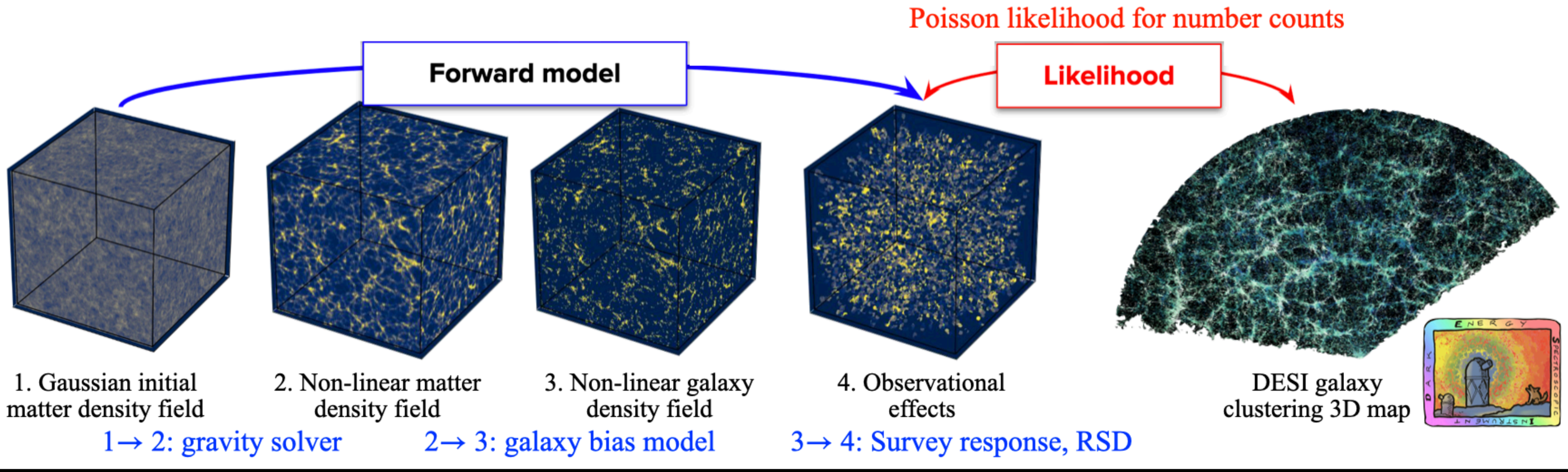
Field-level inference with galaxy clustering using BORG

Jasche & Wandelt (2013), Jasche et al. (2015), Jasche & Lavaux (2019)



Field-level inference with galaxy clustering using BORG

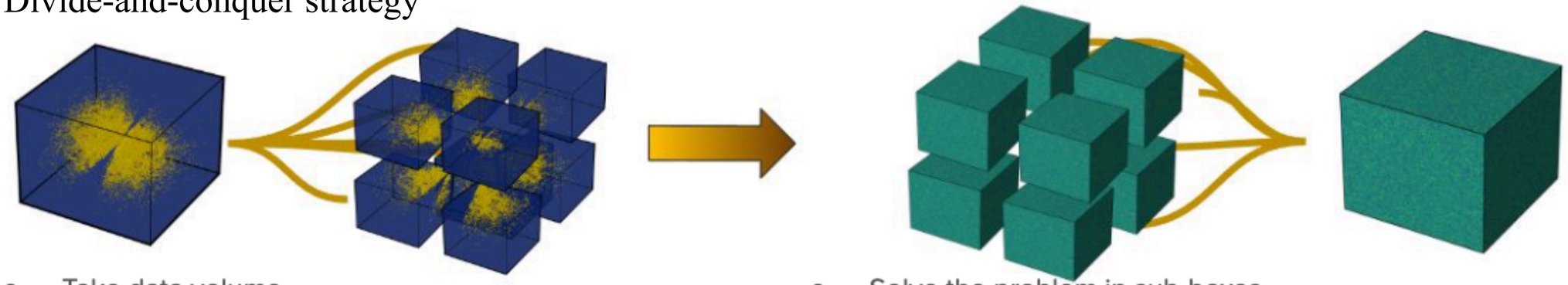
Jasche & Wandelt (2013), Jasche et al. (2015), Jasche & Lavaux (2019)



- Successful application to 2M++ —> Manticore-local (McAlpine et al. 2025)
- Successful application to SDSS up tp $z < 0.7$ —> Manticore-Deep (McAlpine et al. 2026)
- Ongoing application to DESI: proof-of-concept at low resolution (256^3), move to Manticore scalability strategy

Manticore-Deep: scalability strategy

Divide-and-conquer strategy

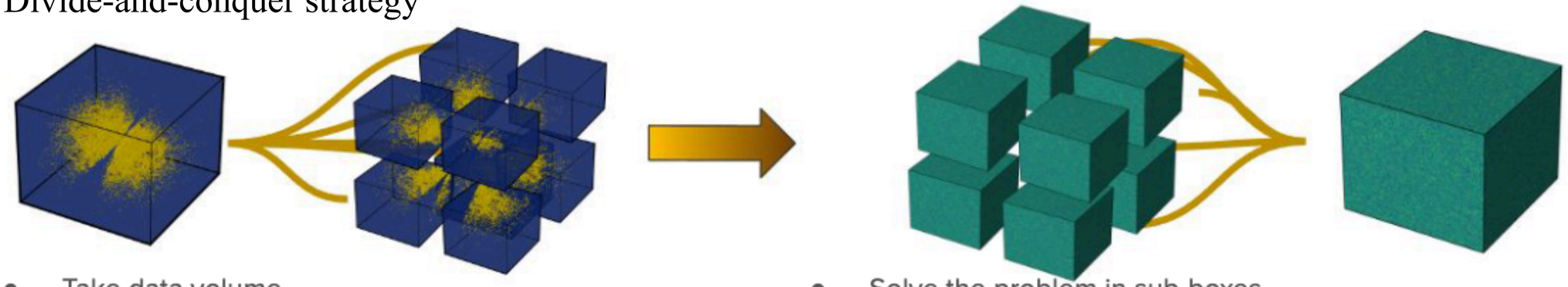


- Take data volume
- Split it into separate inverse problem

- Solve the problem in sub-boxes
- Recombine

Manticore-Deep: scalability strategy

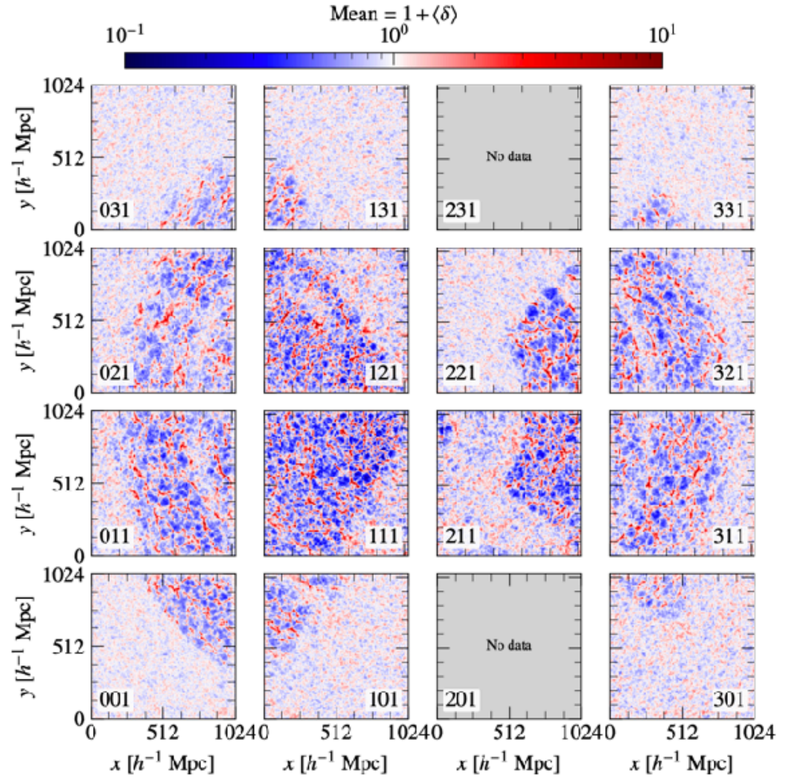
Divide-and-conquer strategy



- Take data volume
- Split it into separate inverse problem
- Solve the problem in sub-boxes
- Recombine

McAlpine et al. 2026

Manticore-Local	Manticore-Deep
L = 700 Mpc/h	L = 4096 Mpc/h
N = 256^3	N = 1024^3
—> 2.7 Mpc/h spatial resolution	—> 4 Mpc/h spatial resolution
	L_{subvol} = 1024 Mpc/h
	N = 256^3
	—> 4 Mpc/h spatial resolution



Summary and open questions

DESI DR2 Galaxy Full-Shape



2pt+3pt statistics using EFT-based model

See Arnaud's talk

Summary and open questions

DESI DR2 Galaxy Full-Shape



2pt+3pt statistics using EFT-based model

Alternative Clustering Methods within DESI



Working pipeline for 12 summary statistics validated against non-HOD simulations

Enough to validate the galaxy-halo connection assumptions?



Including observational systematics: fibre-assignment

Is the method and gain robust against systematics?



Extend the emulator $w_o - w_a$ parameter space

Do we have the required suites of simulations for simulation-based model?

Summary and open questions

DESI DR2 Galaxy Full-Shape



2pt+3pt statistics using EFT-based model

Alternative Clustering Methods within DESI



Working pipeline for 12 summary statistics validated against non-HOD simulations

Enough to validate the galaxy-halo connection assumptions?



Including observational systematics: fibre-assignment

Is the method and gain robust against systematics?



Extend the emulator $w_o - w_a$ parameter space

Do we have the required suites of simulations for simulation-based model?

Field-level inference within DESI



Validated framework on previous-generation data

Model-mispecification and scalability with new generation data?

Is the method worth it for parameter constraints?



Ongoing applications to DESI data

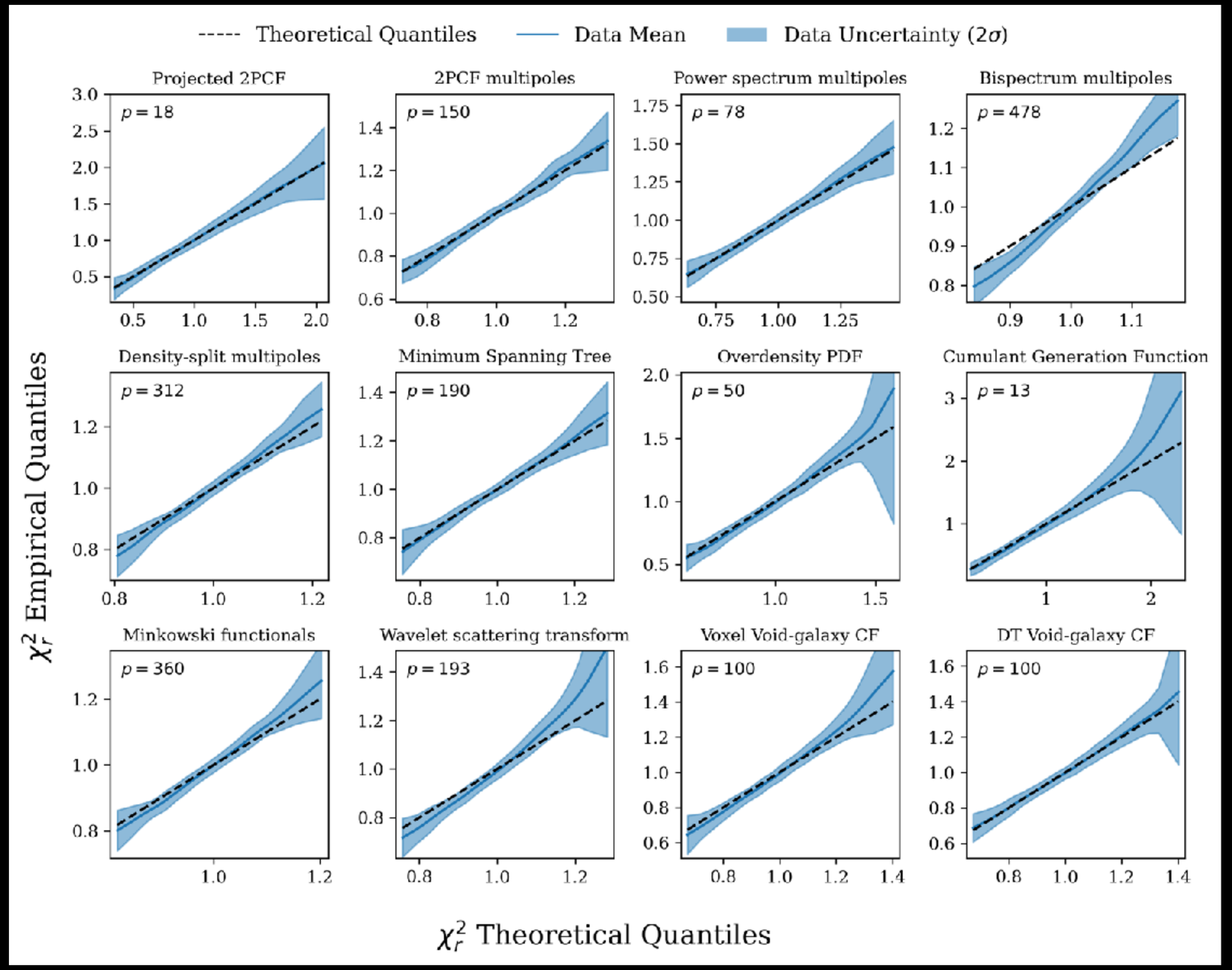


Thank you!

Pauline Zarrouk (CNRS-LPNHE, Sorbonne University)

Back up

Alternative Clustering Methods within the DESI



DESI ACM pipeline, *in prep*

Dark matter halo catalogues (Abacus N-body simulations[1])

[1] Maksimova et al. 2021

Galaxy mock catalogues AbacusHOD[2]

[2] Yuan et al. 2022

Summary statistics

Simulation-based model (SUNBIRD[3])

[3] Cuesta-Lazaro et al. 2023

Parameter inference

Explicit likelihood (Gaussian assumption)

Cosmological constraints

Bayesian field-level inference: how does it work?

Initial phases — World model

$$\pi(x, \Omega, M \mid d) = \pi(M \mid d)\pi(\Omega \mid d, M)\pi(x, d, M, \Omega)$$

Model parameters

$\pi(M \mid d)$

Model evidence
(here ΛCDM)

$\pi(\Omega \mid d, M)$

Parameter estimation
$\Omega_m, \sigma_8, n_s, h, \ldots$ + astro + nuisance

Ensemble view: laws and parameters, statistically consistent

$\pi(x \mid d, M, \Omega)$

Specific realisation
Initial phases / modes

Realization view: our Universe

Output of the re-simulations at the inferred initial conditions: digital twins of the Universe

The ultimate goal is to access $\pi(x, \Omega \mid M, d)$.

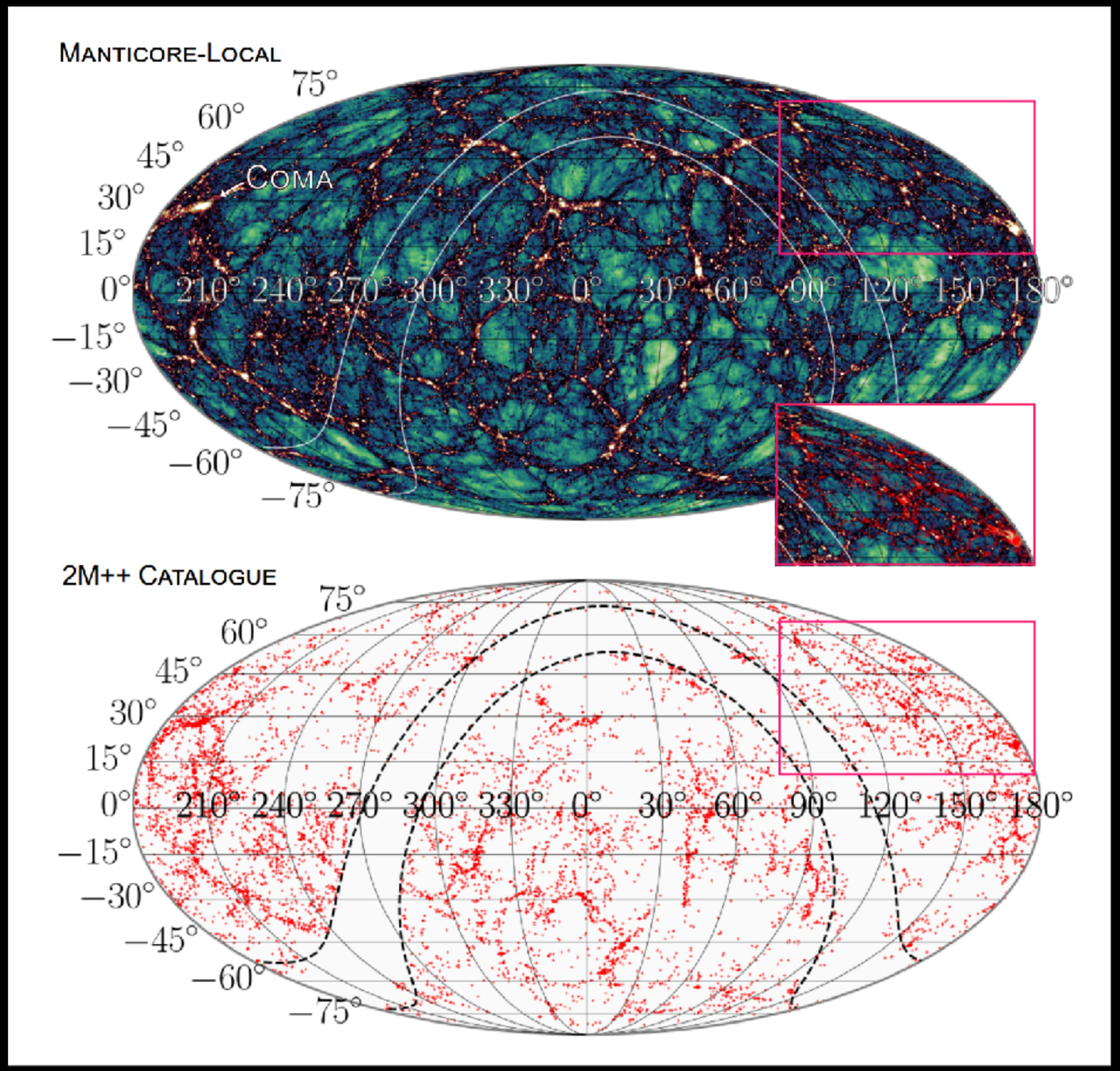
But it requires:

1. Fast gradient-based sampler (HMC) for the very high dimensional parameter space: $\sim 10^{8-9}$ parameters (mesh resolution)
2. Requires an accurate (differentiable) model that links the initial Gaussian random field to a given observable
3. Precise knowledge of the selection functions and mitigation of systematic effects at the field level

How? One way to do it is with BORG

Jasche & Wandelt (2013), Jasche et al. (2015), Jasche & Lavaux (2019)

Successful proof-of-concept: Manticore-Local



McAlpine et al. 2025

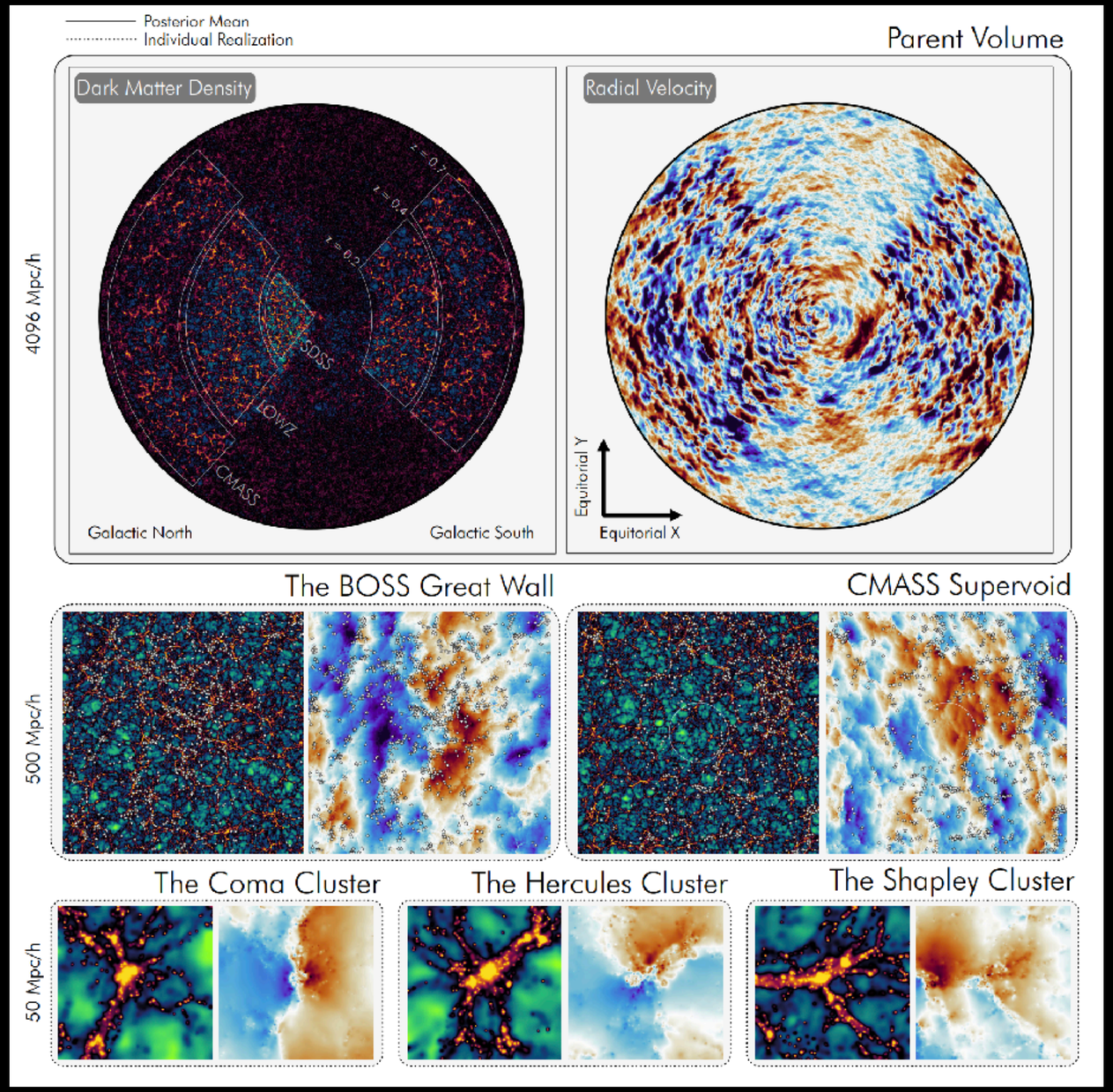
Posterior predictive tests

- ✓ Initial power spectrum (white noise field)
- ✓ Power spectrum and bispectrum of the matter field at z=0
- ✓ Halo mass function
- ✓ Velocity field
- ✓ Galaxy clusters detection and properties

Which additional tests to showcase the limitations of the galaxy bias model?

The limitations of the generalized Poisson likelihood?

Successful proof-of-concept: Manticore-Deep



McAlpine et al. 2026

Posterior predictive tests

- Initial power spectrum (white noise field)
- Power spectrum and bispectrum of the matter field at z=0
- Halo mass function
- Convergence field and cross-correlation with Planck PR3 CMB map at 7.4σ
- Velocity field and detection of tSZ effect at 3.5σ

Which additional tests to showcase the limitations of the galaxy bias model?

The limitations of the generalized Poisson likelihood?