Field-level inference lecture 4: State-of-the-art

Cosmology Beyond the Analytic Lamppost course (2025)

Florent Leclercq

www.florent-leclercq.eu

Institut d'Astrophysique de Paris
CNRS & Sorbonne Université

17 June 2025

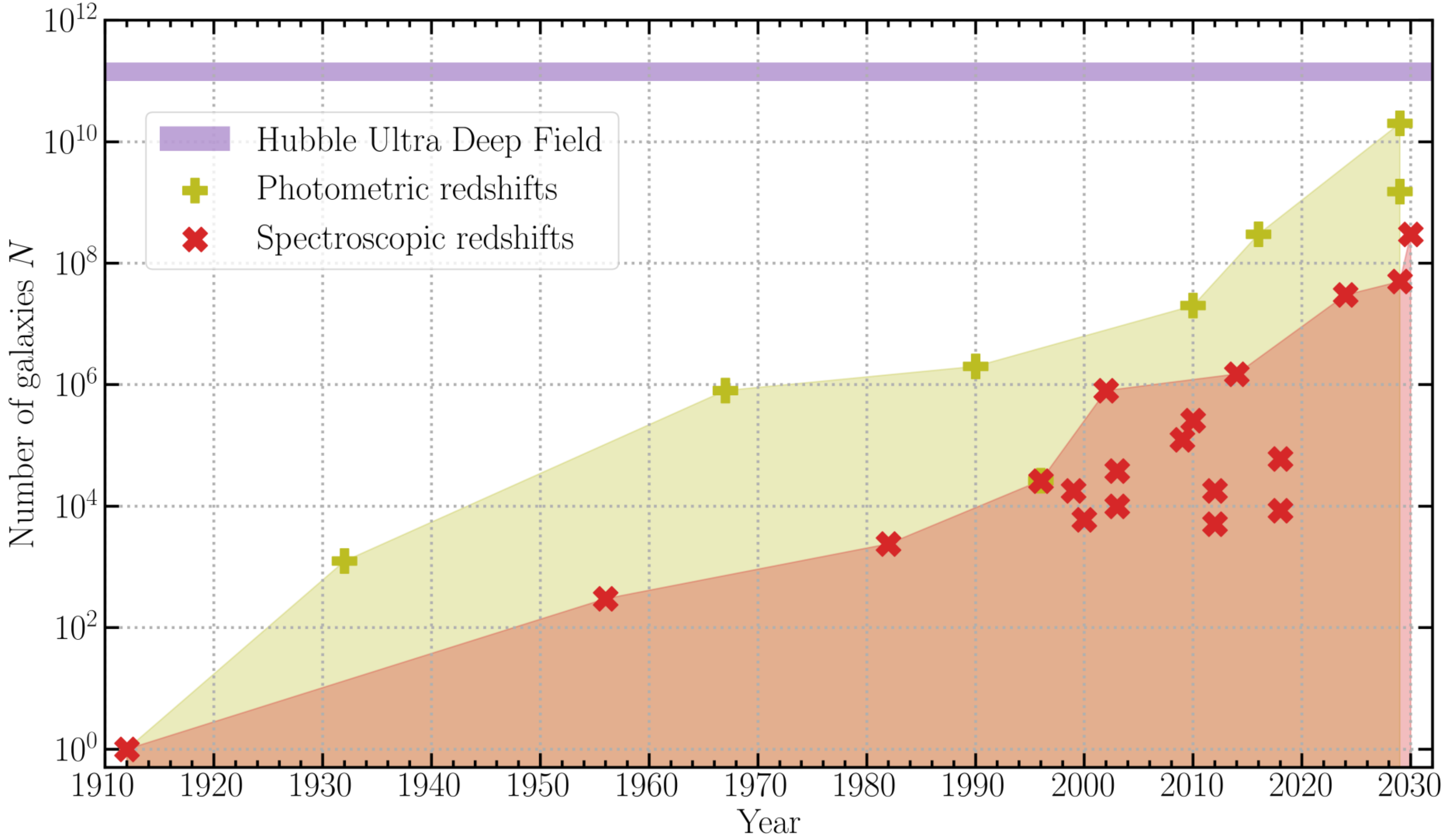
Turnagain Arm, Alaska

04 STATE-OF-THE-ART

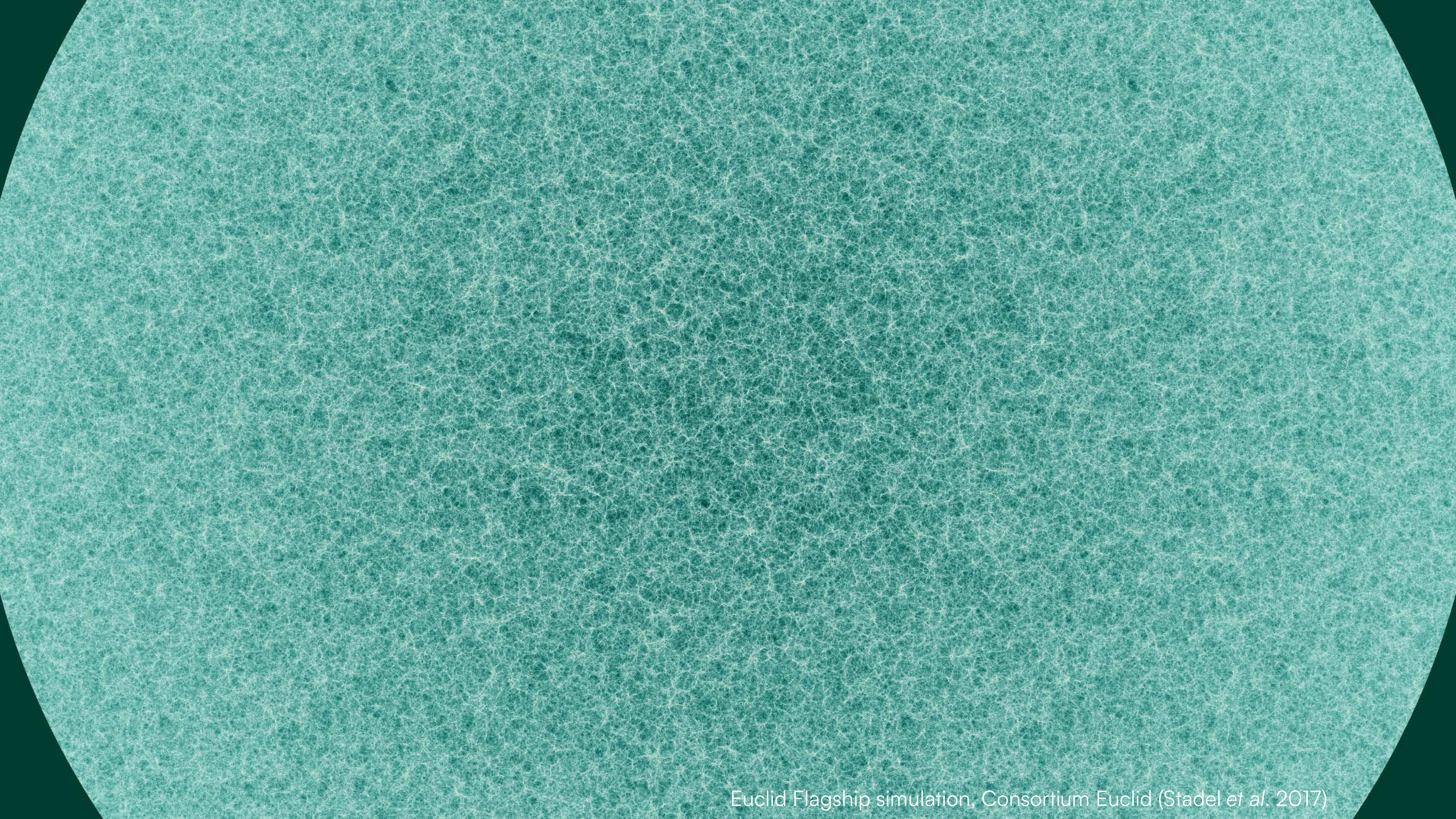
Euclid's view of the Perseus cluster of galaxies, ESA, 07/11/2023

The growth of data

- The number of observed galaxies has grown exponentially since 1910.



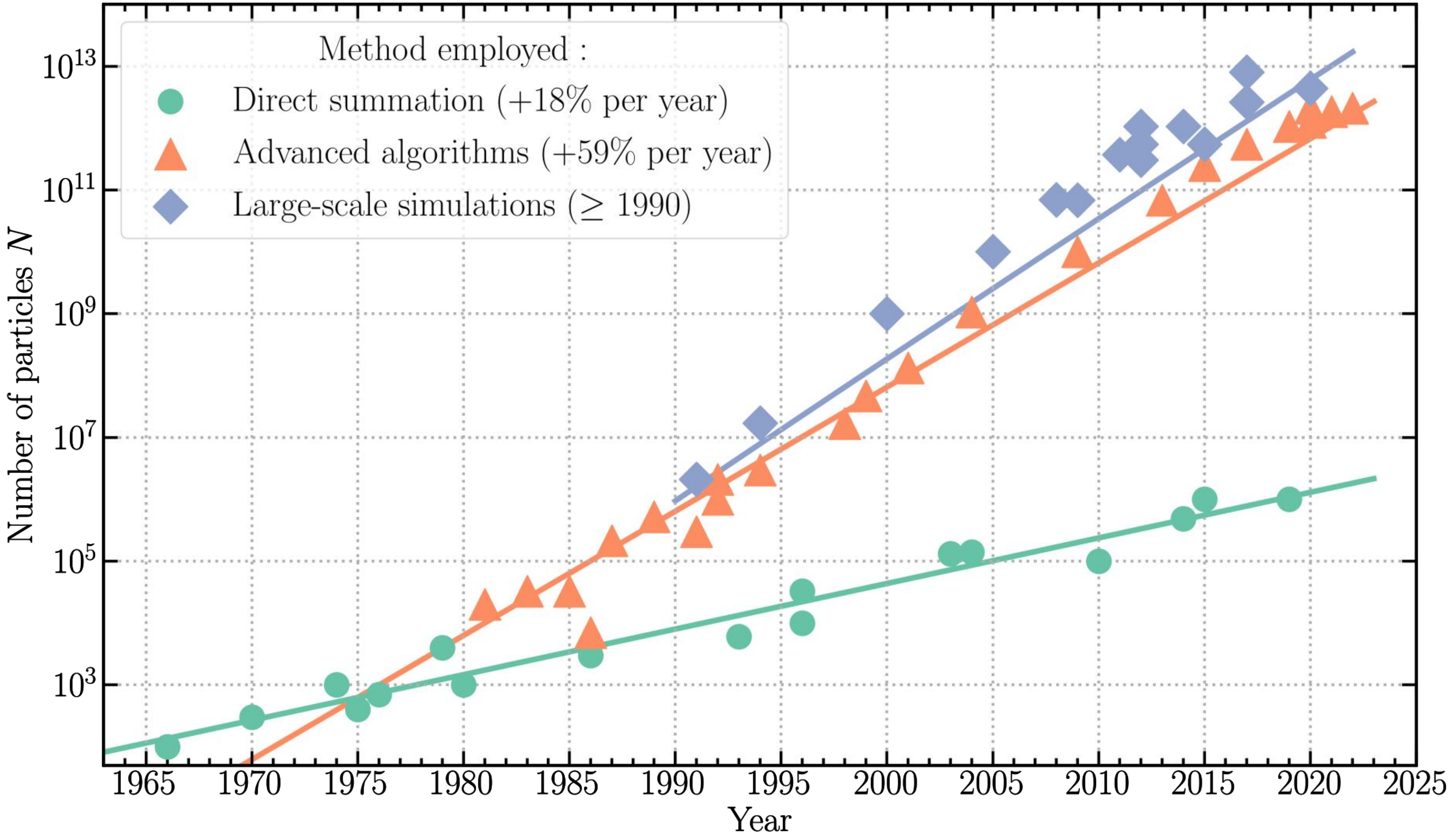
Galaxy surveys: figure inspired by J. Peacock, data collected by J. Jasche



Euclid Flagship simulation, Consortium Euclid (Stadel *et al*. 2017)

The growth of models

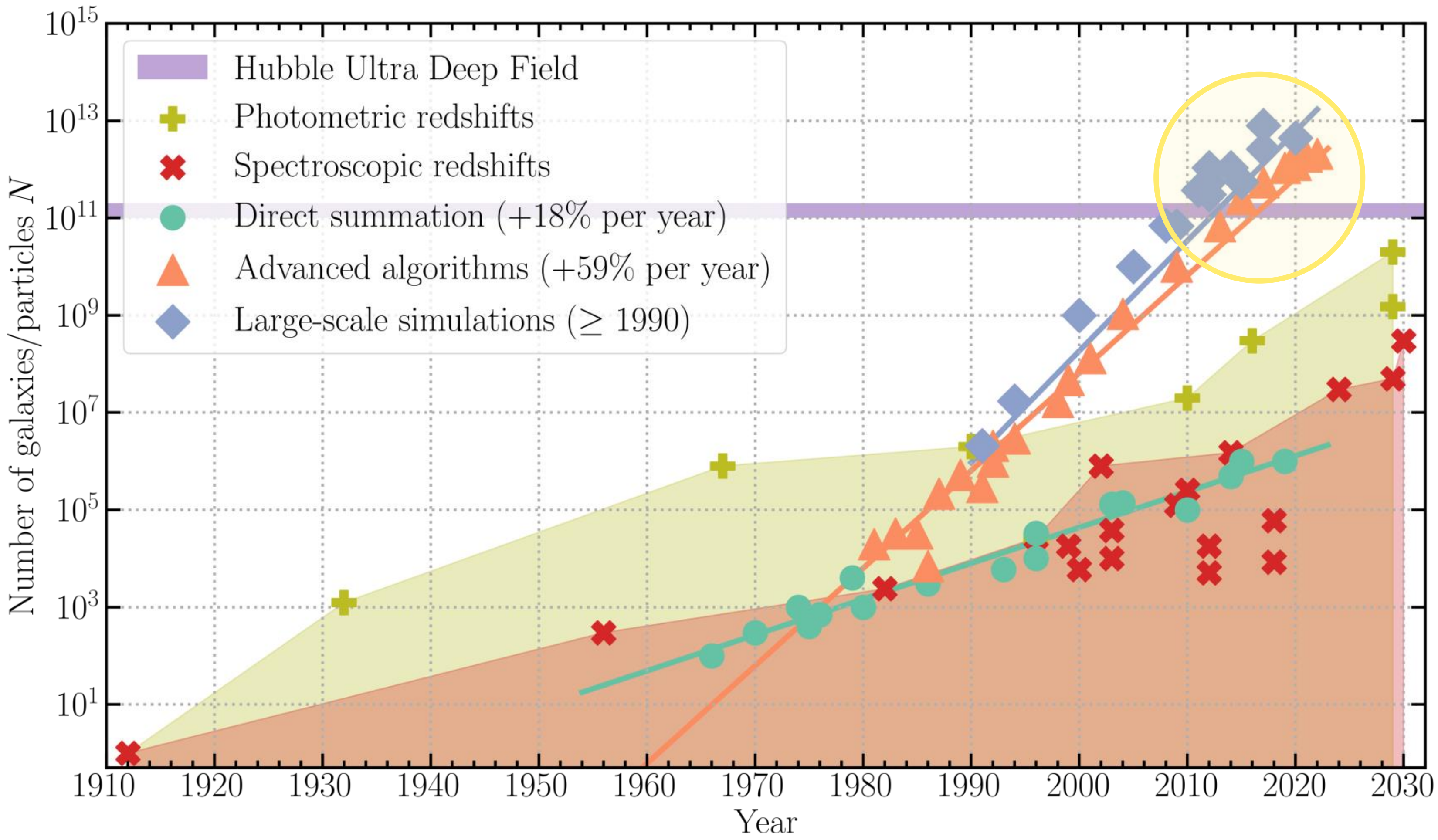
- Numerical simulations are the new way to express theoretical models.



Cosmological simulations: Github:florent-leclercq/Moore_law_cosmosims

Comparative growth of data and models

- We are already using more particles in simulations than there are galaxies in the observable Universe!



Galaxy surveys: figure inspired by J. Peacock, data collected by J. Jasche
Cosmological simulations: Github:florent-leclercq/Moore_law_cosmosims

INTRODUCTION TO STATE-OF-THE-ART FIELD-LEVEL INFERENCE

What is forward modelling?

- Data analysis is the art of having the two ends meet…

Forward modelling

Any step here is expensive and possibly complicated

"Backward modelling" assumptions, corrections, etc.

Any step here is lossy and potentially biasing

Cosmological parameters → Initial conditions → Final conditions → Mock catalogues → Power spectrum ↔ Power spectrum ← Observed catalogues ← Images / spectra ← Photons on instruments

- Ideally…

Forward modelling → Photons on instruments

- Less ideally, but still unrealistic:

Forward modelling → Images / spectra ← Photons on instruments

- Current developments in cosmology focus on:

Forward modelling

(…) → Mock catalogues → Any summary ↔ Any summary ← Observed catalogues ← (…)

IMPLICIT LIKELIHOOD INFERENCE

OR

Forward modelling

(…) → Mock catalogues ↔ Observed catalogues ← (…)

FIELD-LEVEL INFERENCE

What is field-level inference?

- What it is (a personal point of view):
  - Any Bayesian inference method where the likelihood lives **at the level of the field**, **without compression** of the data of any kind.
  - It reconstructs **maps** as a natural by-product.
- What it is not (also a personal point of view):
  - Any method that includes a compression of the map-level information:
    - Usual summary statistics (correlation functions, cluster counts, void properties, etc.)
    - Statistical compression like MOPED or score compression
    - Any other compression, be it “optimal” in any sense, e.g. information-maximising neural networks (IMNN), neural compression…
  - Because if compressing the data is allowed, any cosmological analysis could qualify as “field-level”, since any summary statistics (e.g. the power spectrum) comes from a field at some point.
- For these reasons:
  - The likelihood in field-level inference is (almost always) explicit, at the level of the (data) field.
  - Field-level inference problems are very high dimensional problems, sampling from their posterior (almost always) requires gradients of the data model.

Explicit and implicit likelihood inference

- Simulated data are nothing other than draws from the likelihood of your problem!
- You may not be able to score variables with your likelihood (an "intractable likelihood"), but you are always able to sample it (i.e. draw samples = simulations).
  - Because if you can't even simulate the data, you're in big trouble… The only way forward is to make simplifying assumptions about the data-generating process.

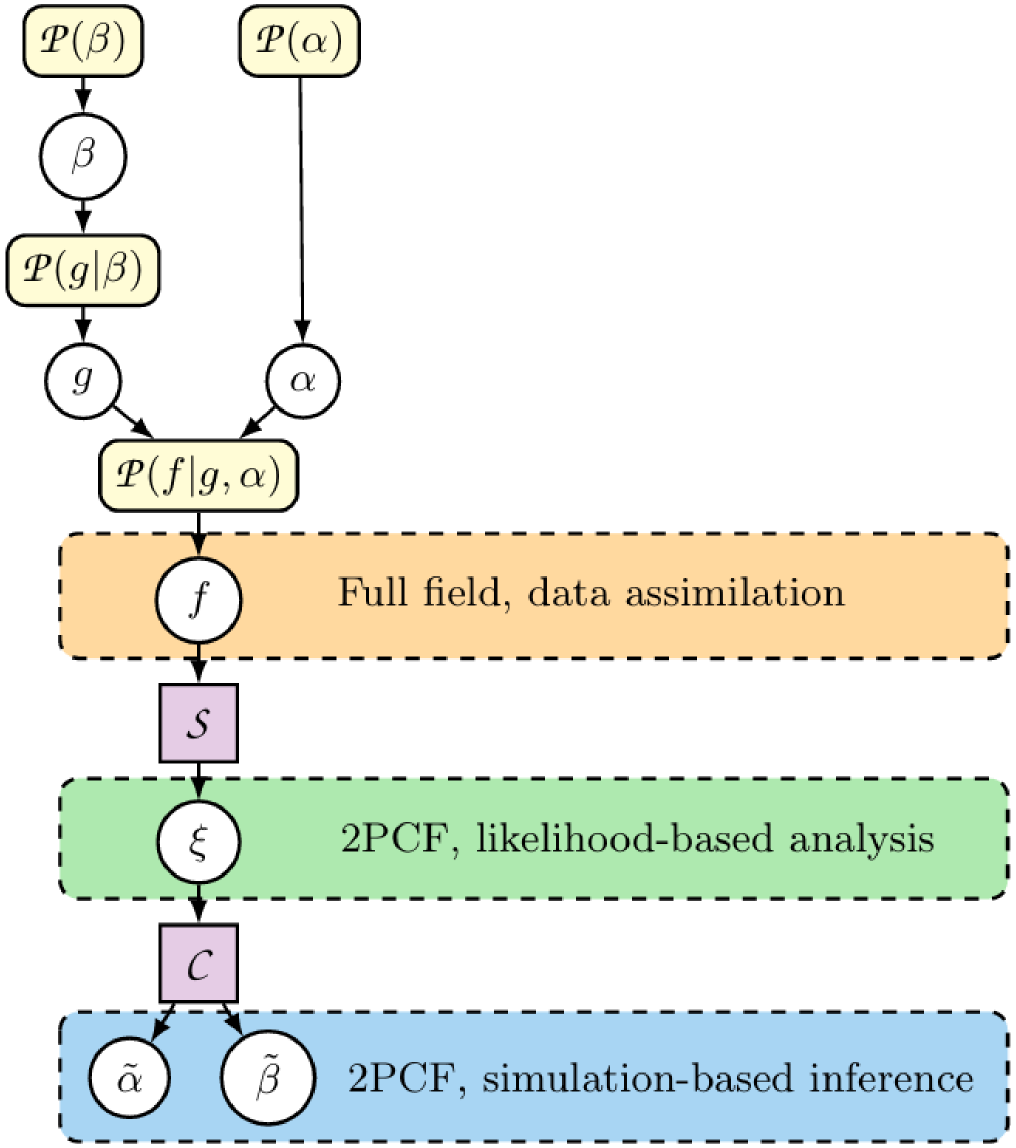
> What I cannot create, I do not understand.
> R. Feynman

- The Bayesian problem of inferring parameters, using Bayes' theorem, when the some of the pdfs (usually the likelihood) are implicitly defined by a simulator is called implicit (likelihood) inference (a.k.a. simulation-based inference or likelihood-free inference)

	Explicit inference	Implicit inference
What can I do with the likelihood?	**Score** and **sample**	**Sample** only
With compression (summary statistics level)	*Traditional cosmological data analysis:* correlation function, clusters, voids…	ABC, ABC-PMC, DELFI, NPE, NRE, BOLFI, SELFI…
No compression (field level)	BORG, Almanac, LEFTField…	*Just becoming conceivable:* LULO

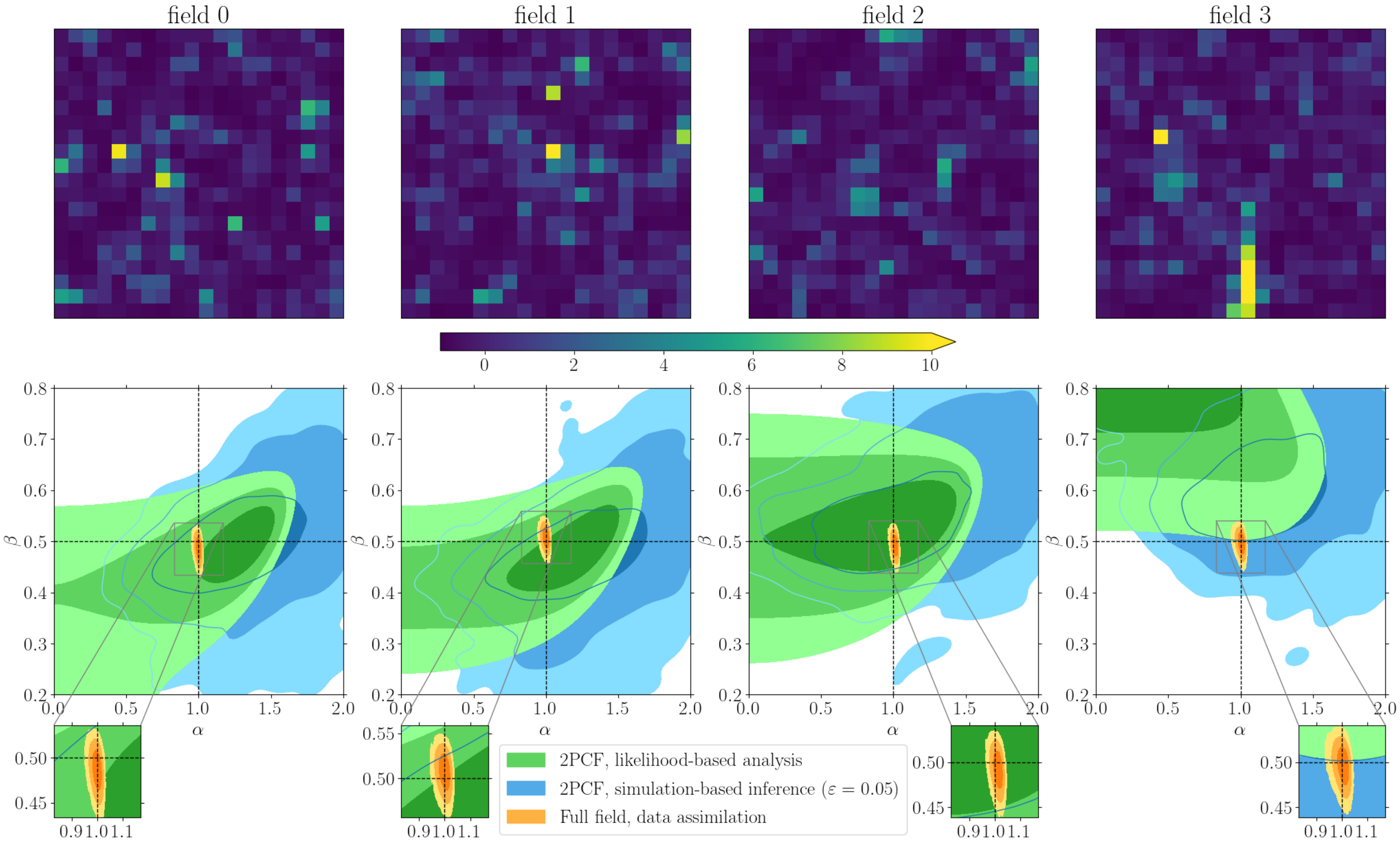
Summary statistics vs field-level inference in cosmology

- How to optimally extract information from a field?
- Compare three approaches:
  - Field-level data assimilation (DA), using Bayesian forward modelling and Hamiltonian Monte Carlo (HMC)
  - Standard likelihood-based analysis (LBA) of the two-point correlation function (2PCF), assuming a Gaussian distribution and a fixed covariance matrix
  - Simulation-based inference (SBI) (a.k.a. implicit likelihood inference, ILI), using Approximate Bayesian Computation (ABC), based on the 2PCF



FL & Heavens, 2103.04158

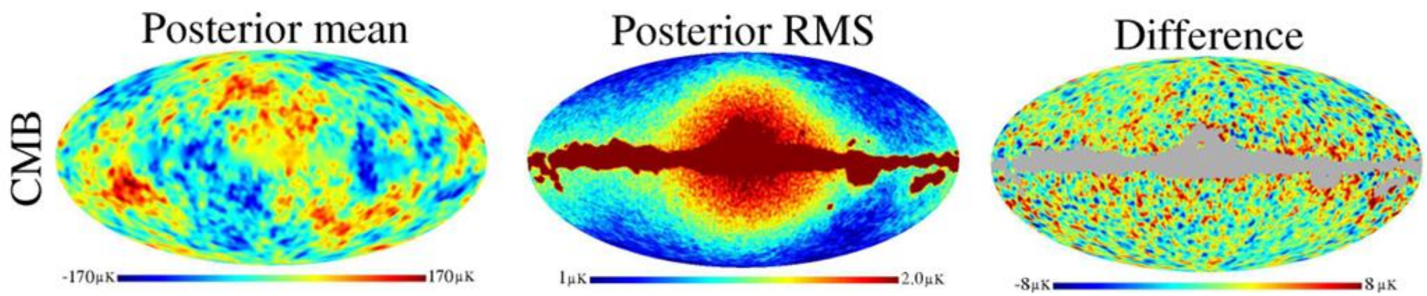
Summary statistics vs field-level inference in cosmology



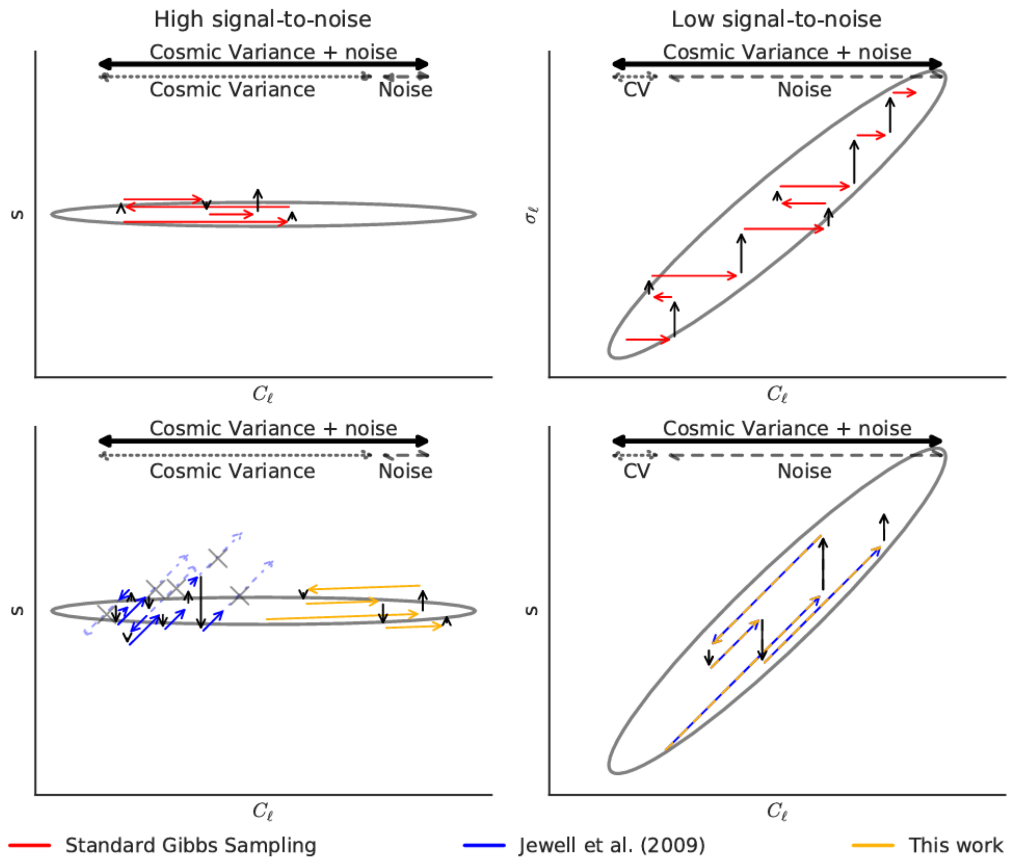
FL & Heavens, 2103.04158

FIELD-LEVEL INFERENCE FOR THE COSMIC MICROWAVE BACKGROUND

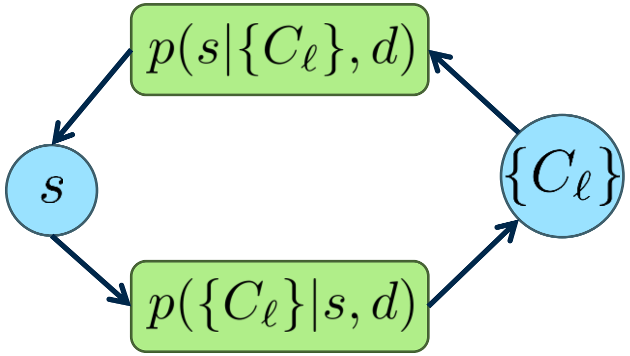
Cosmic Microwave Background



- CMB was the first field-level inference problem solved in cosmology.
- Assuming that the CMB map is a Gaussian random field (⇒ Wiener filtering), there remains important questions to solve:
  - Efficient joint sampling of the field and the power spectrum/cosmological parameters (Gibbs sampling)
  - Component separation (SMICA, BICA)



To sample $p(s, \{C_\ell\}|d)$:



Wandelt *et al.*, astro-ph/0310080; Eriksen *et al.*, 0709.1058; Jewell *et al.*, 0807.0624; Racine *et al.*, 1512.06619; …

FIELD-LEVEL INFERENCE FOR GALAXY CLUSTERING

Bayesian forward modelling cosmic structure surveys

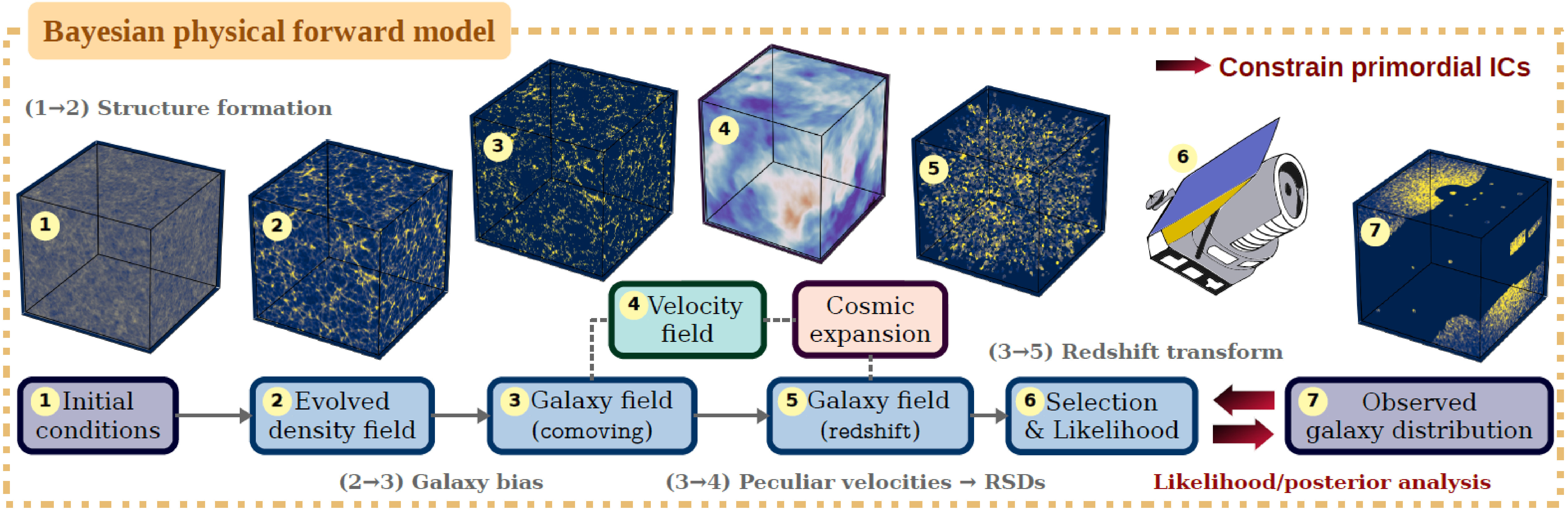
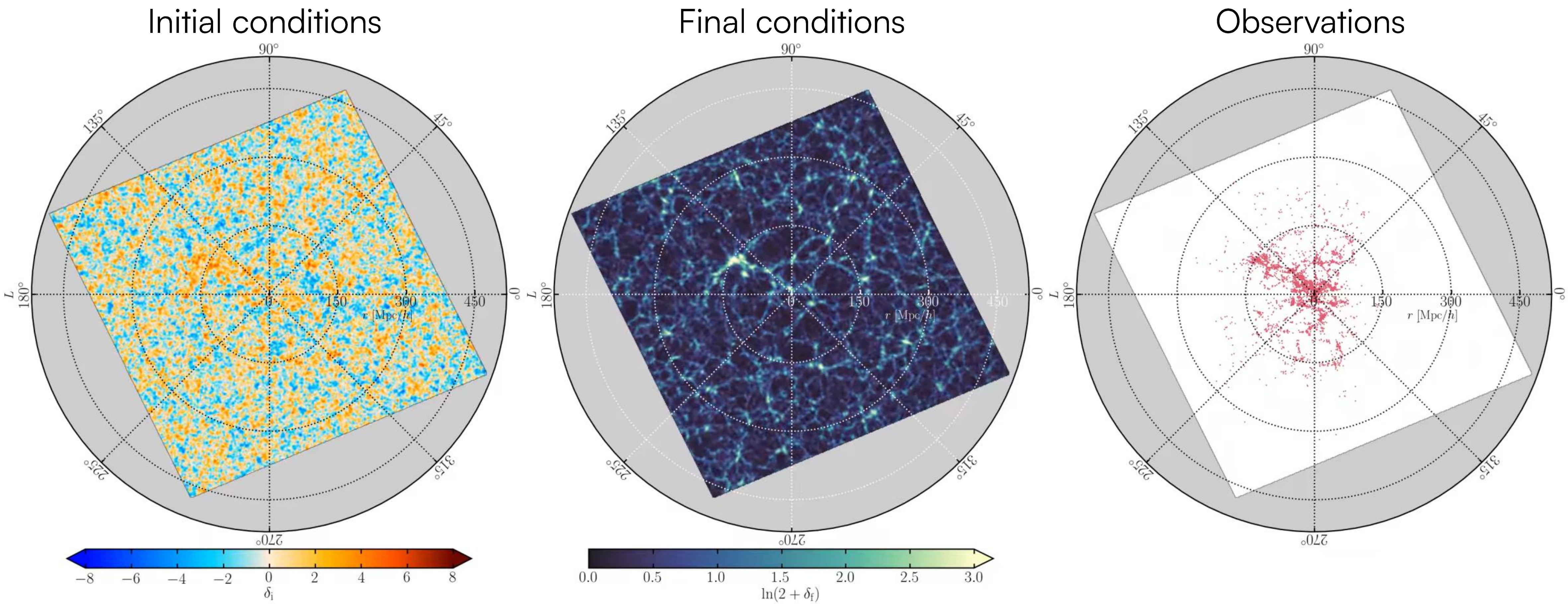


Figure from Kostić *et al.*, 2107.00657

Bayes at work in cosmology:
The BORG algorithm (*Bayesian Origin Reconstruction from Galaxies*)

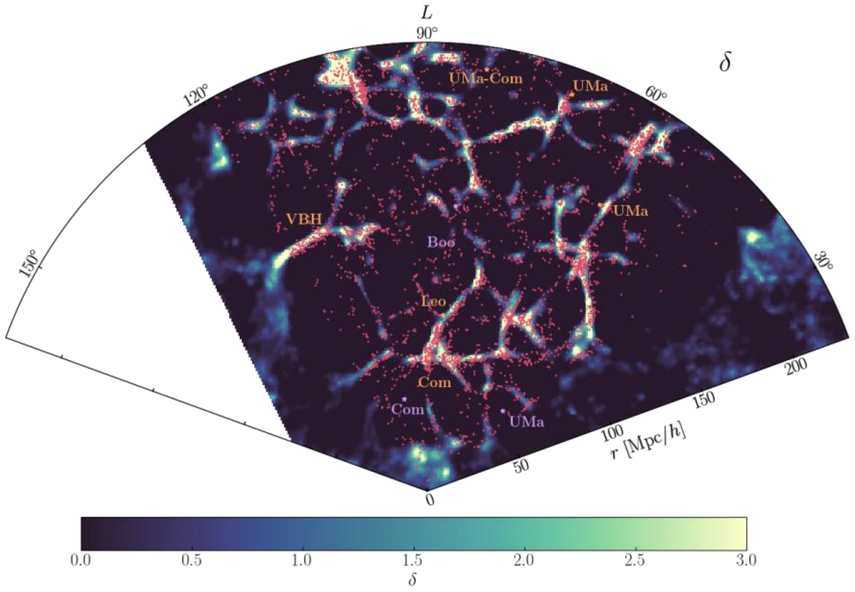


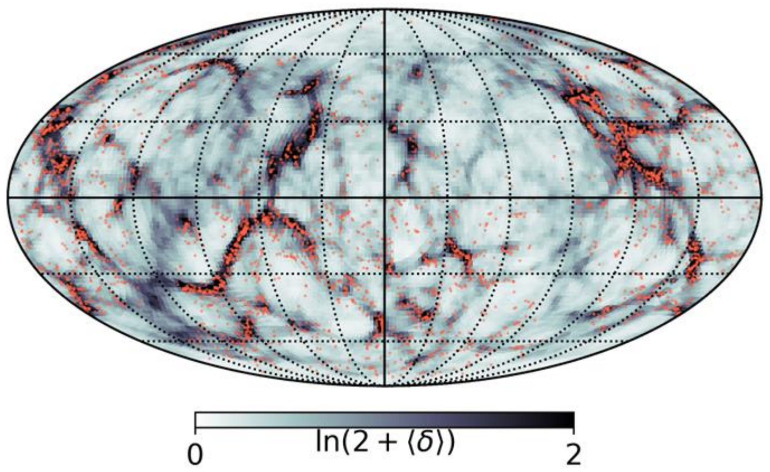
67,224 galaxies, ≈ 17 million parameters, 5 TB of primary data products, 10,000 samples,
≈ 500,000 forward and adjoint gradient data model evaluations, 1.5 million CPU-hours

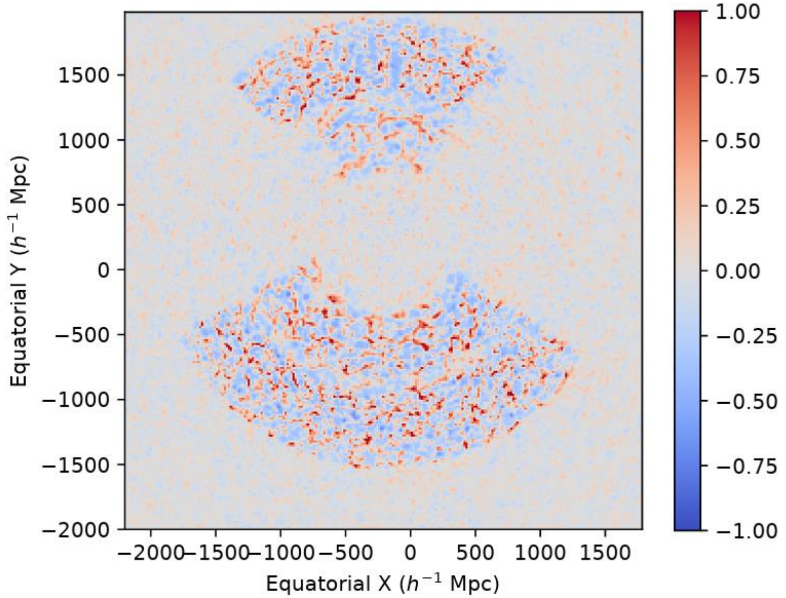
Jasche & Wandelt, 1203.3639; Jasche, FL & Wandelt, 1409.6308; Jasche & Lavaux, 1806.11117; Lavaux, Jasche & FL, 1909.06396

Applications to real data

- SDSS main galaxy sample
- 2M++
- SDSS3-BOSS





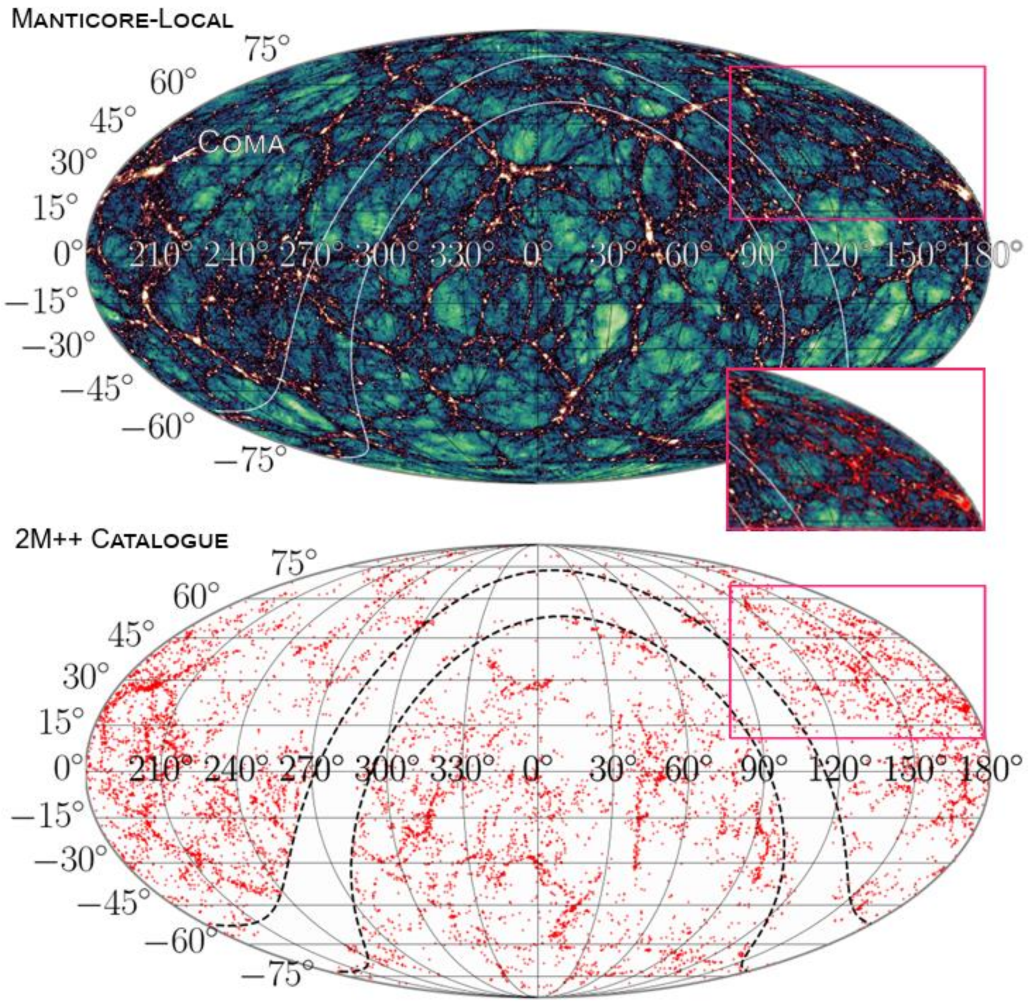


Jasche, FL & Wandelt, 1409.6308

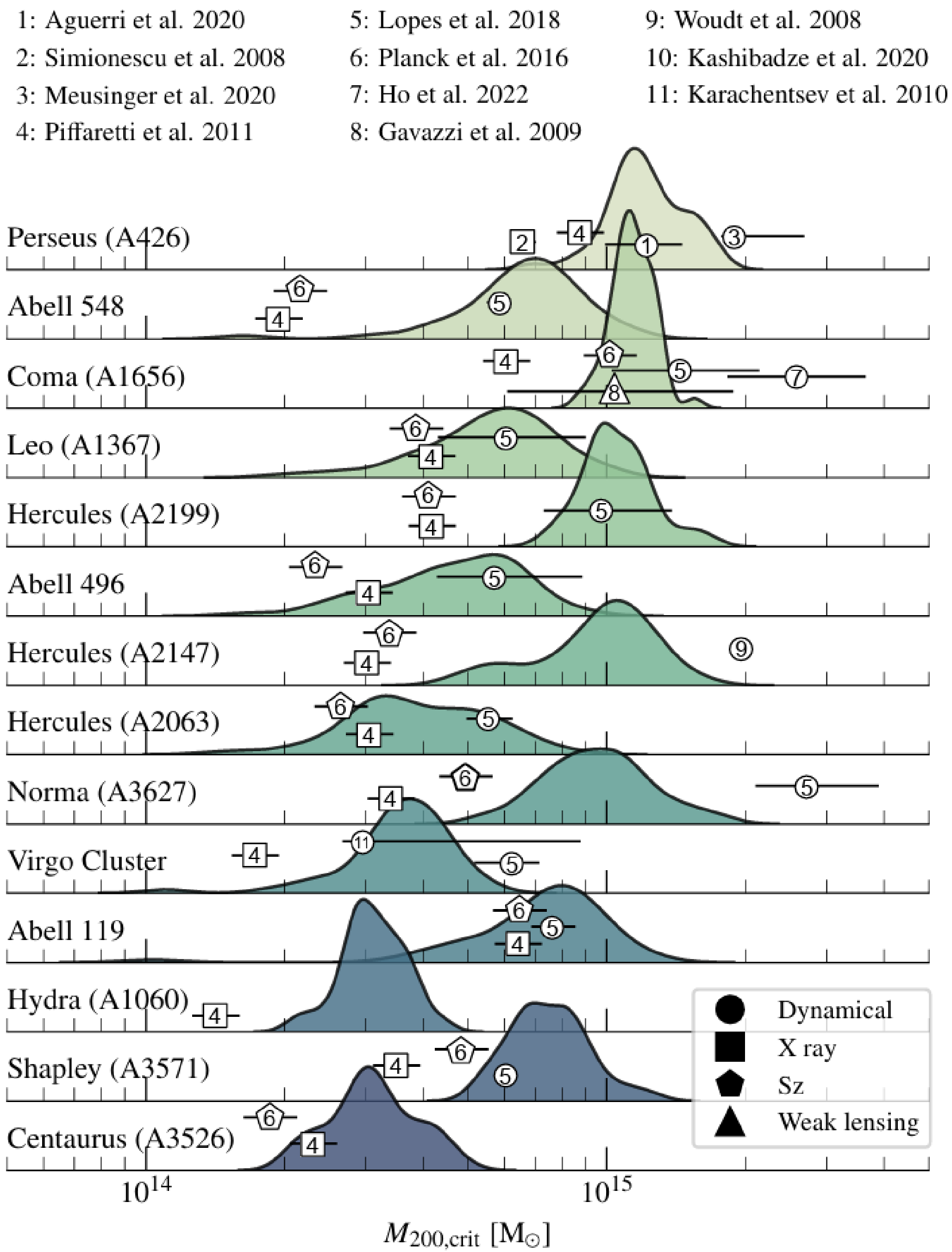
Lavaux & Jasche, 1509.05040;
Jasche & Lavaux, 1806.11117

Lavaux, Jasche & FL, 1909.06396

The Manticore project



McAlpine *et al.*, 2505.10682



Gravity models

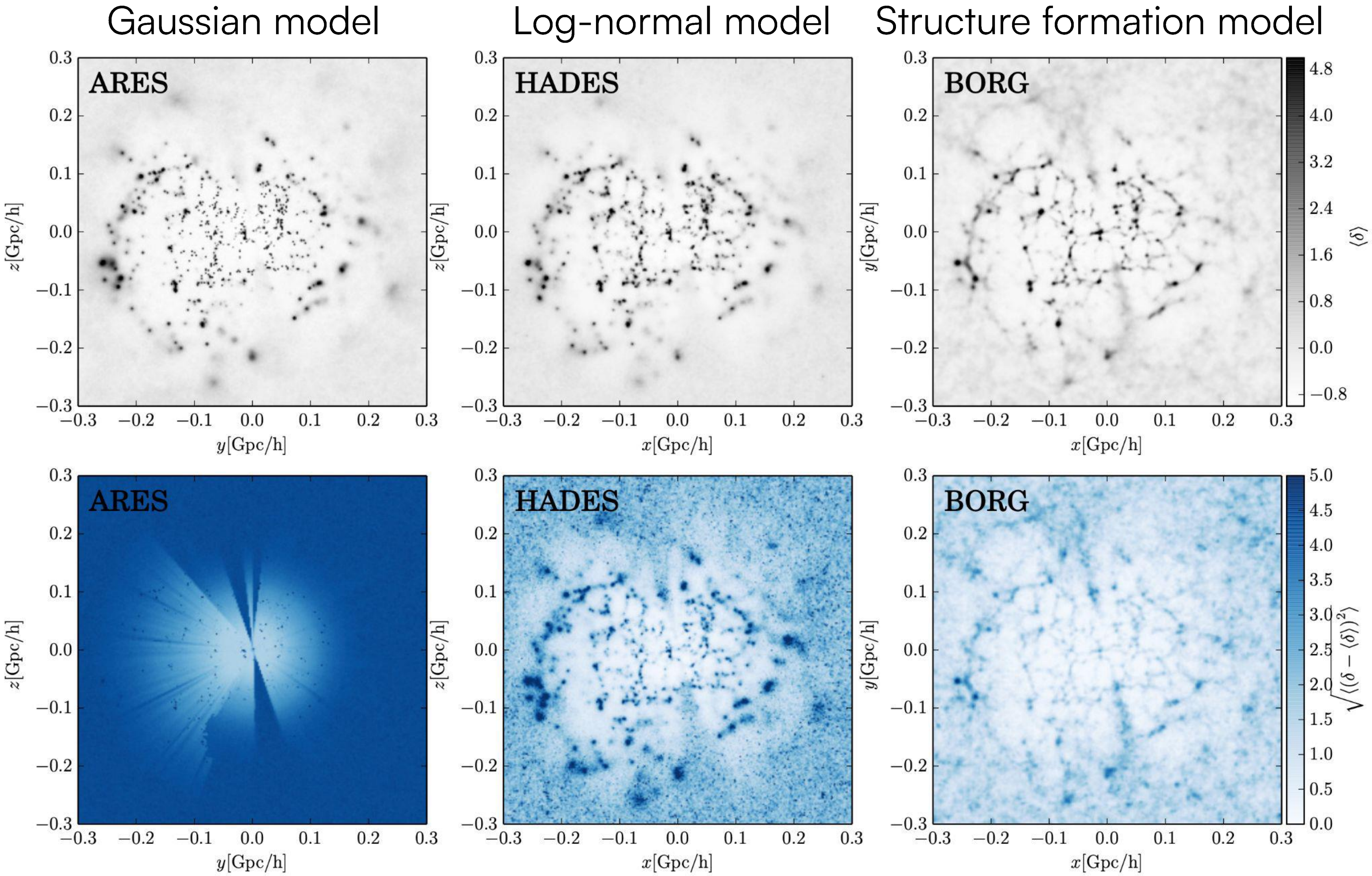
- Phenomenological density field models:
  - Gaussian-linear (Wiener filter)
  - Log-normal
- Structure formation models:
  - Lagrangian Perturbation Theory (LPT)
  - Quantum LPT/Propagator PT (QLPT/PPT)
  - Particle-Mesh (PM)

Wang *et al.* 1407.3451; Jasche & Lavaux, 1806.11117

  - COmoving Lagrangian Acceleration (COLA)

Tassev *et al.* 1301.0322; Stopyra *et al.* 2304.09193

- Others:
  - Neural network emulator (BORGEmu)
    Doeser *et al.*, 2312.09271
  - Several PM grids (ZoomPM)
    Wempe *et al.*, 2406.02228



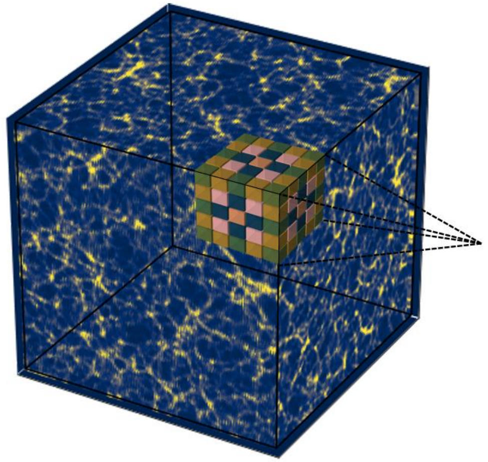
e.g. Zaroubi, 0010561; Erdoğdu *et al.*, 0312546; Kitaura *et al.* 0906.3978; Jasche *et al.* 0911.2493; Granett *et al.* 1112.0008

e.g. Kitaura *et al.*, 0911.1407; Jasche & Kitaura, 0911.2496; Ata *et al.*, 1408.2566; Böhm *et al.* 1701.01886

e.g. Jasche & Wandelt, 1203.3639; Kitaura, 1203.4184; Wang *et al.*, 1301.1348; Ata *et al.*, 2206.01115; Modi *et al.* 2104.12864

Galaxy bias model

- Galaxy bias (relation between dark matter density and galaxy number counts):
  - Field-level inference requires a voxel-based bias model (existing bias models for fitting correlation functions are inadequate for field-level inference). Constraints are to ensure small-scale accuracy and expected galaxy number count ≥ 0 (for a Poisson likelihood).
  - Galaxy bias for field-level inference is usually luminosity-dependent, non-local and deterministic.
    - Power-law (Jasche, FL & Wandelt, 1409.6308) $N_{\mathrm{g}} \propto (1+\delta_{\mathrm{m}})^{\alpha}$
    - Truncated power law (Neyrinck *et al.*, 1309.6641; used in Jasche & Lavaux, 1806.11117, Stopyra *et al.*, 2304.09193)
      $N_{\mathrm{g}} \propto (1+\delta_{\mathrm{m}})^{\alpha} \exp\left(-\rho_{\mathrm{m},0}(1+\delta_{\mathrm{m}})^{-\epsilon}\right)$
    - Quadratic form in $1, \delta_{\mathrm{m}}, \delta_{\mathrm{m}}^2, \delta_{\mathrm{m}}^{(2)}, (\delta_{\mathrm{m}}^{(2)})^2$ (fields and smoothed fields) (Lavaux, Jasche & FL, 1909.06396)
    - Scale-dependent second-order bias for models with f_{NL} (Euclid Collaboration (Andrews *et al.*), 2412.11945)
      $N_{\mathrm{g}} \propto b_1 \delta_{\mathrm{m}} + \frac{b_2}{2}\delta_{\mathrm{m}}^2 + b_K K^2 + b_\phi f_{\mathrm{NL}} \phi_{\mathrm{g}} + b_{\phi,\delta} f_{\mathrm{NL}} \delta_{\mathrm{m}} \phi_{\mathrm{g}}$
    - Sigmoid-Truncated Double Power-Law (STDP) (McAlpine *et al.*, 2505.10682)
      $$N_{\mathrm{g}} \propto (1+\delta_{\mathrm{m}})^{\alpha} \times \left[1 + \left(\frac{\rho_{\mathrm{m}}}{\rho_{\mathrm{m},0}}\right)^{\beta}\right]^{-\beta} \times \frac{1}{1+\left(\frac{\rho_{\mathrm{m}}}{\rho_{\mathrm{m},1}}\right)^{-\gamma}}$$
    - A small, interpretable neural network without pre-training: PineTree (Ding, Lavaux & Jasche, 2407.01391)

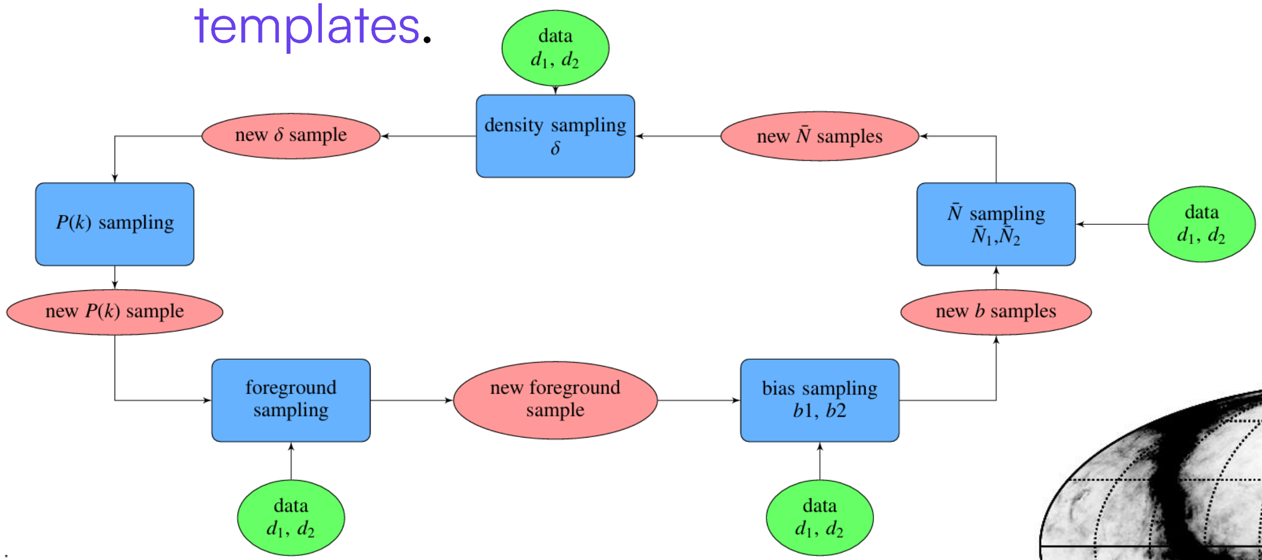


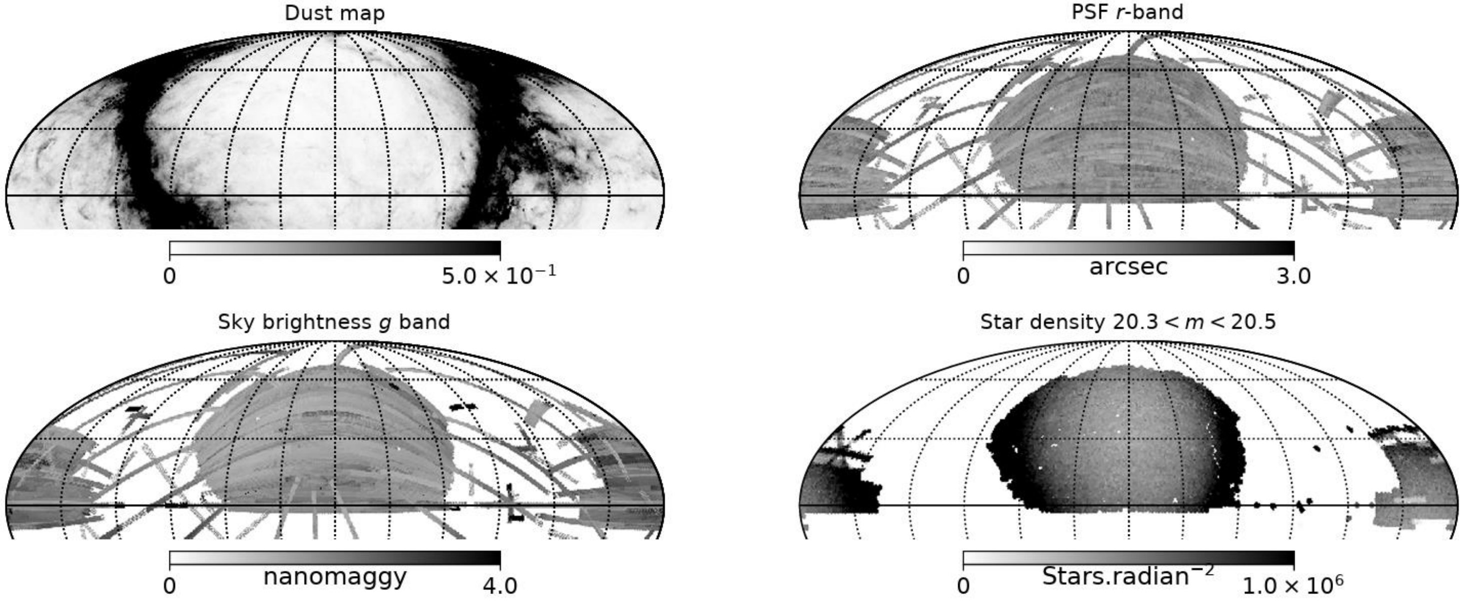
Field-level likelihood function

- Likelihood (probability distribution to be used for $p(\{N_{\mathrm{g}}\}|s, \{\Omega\})$):
  - The likelihood accounts for the stochasticity in the galaxy formation process. Usually we want it to predict an integer number of galaxy per voxel (a probability *mass* function).
  - Several studies have observed super-Poissonian behaviour in the scatter of galaxy counts, particularly in high-density regions, where the variance exceeds the mean.
    - Gaussian distribution (ARES, HADES codes)
    - Poisson distribution, with mean equal to variance (original BORG code)
      $$\ln p(n = N|\lambda) = N \ln \lambda - \lambda + \text{const.}$$
    - Negative-binomial distribution (Ata *et al.*, 1408.2566)
      $$\ln p(n = N|r, p) = r \ln p + N \ln(1 - p) + \text{const.}$$
    - Generalised Poisson distribution (McAlpine *et al.*, 2505.10682)
      $$\ln p(n = N|\lambda, b) = (N - 1) \ln [\lambda(1 - b) + Nb] - \lambda(1 - b) - Nb + \text{const.}$$
    - Effective Field Theory (EFT): Gaussian likelihood in Fourier space (all of the Fourier grid) rather than in voxel space — the LEFTField code (Stadler, Schmidt & Reinecke, 2303.09876; Babić, Schmidt & Tucci, 2407.01524)

Foregrounds and systematics

- For ***known*** foregrounds, it is possible to sample the amplitude of each contamination based on templates.

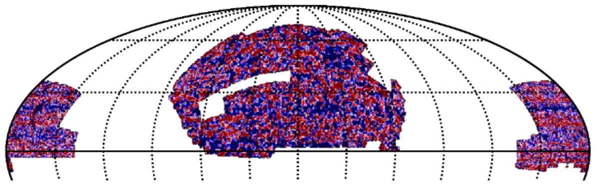




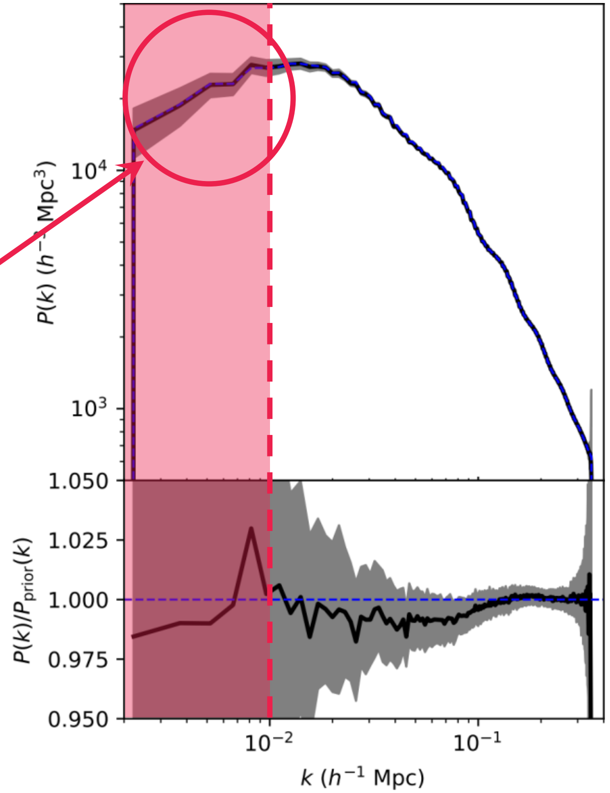
Jasche & Lavaux, 1706.08971; Lavaux, Jasche & FL, 1909.06396

Machine-aided report of unknown data contaminations
Application to SDSS-III/BOSS (LOWZ+CMASS)

Map of ***unknown*** foreground contaminant

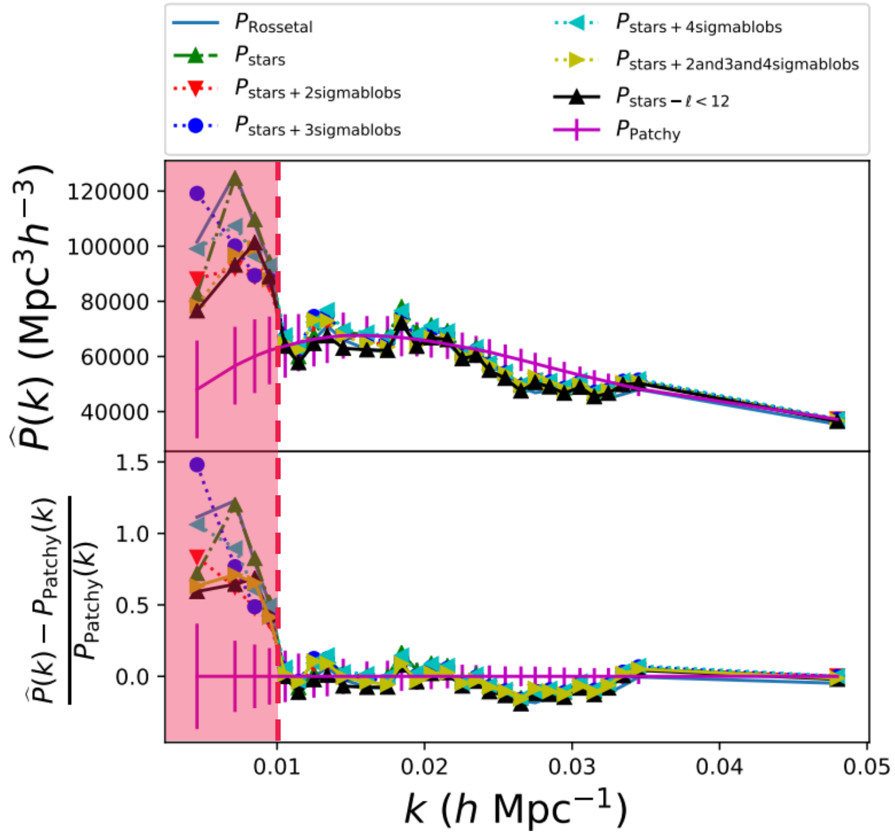


BORG *a posteriori* power spectrum



No apparent contamination, even well beyond the turn-over

State-of-the-art with backward-modelling technique (mode subtraction)

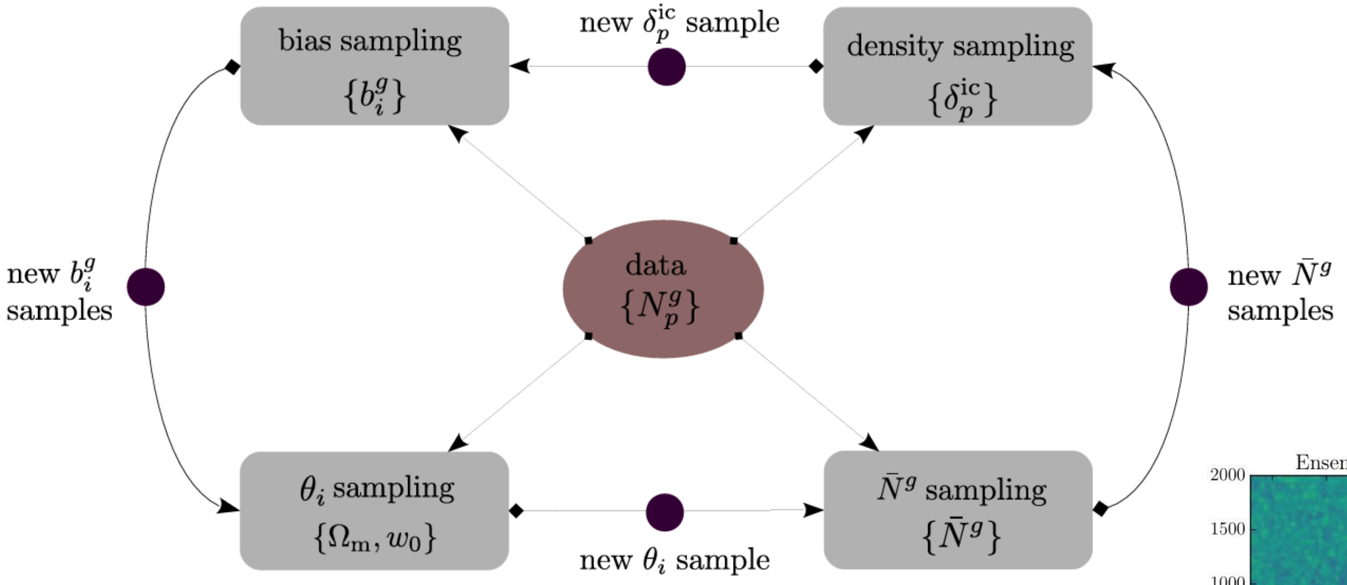


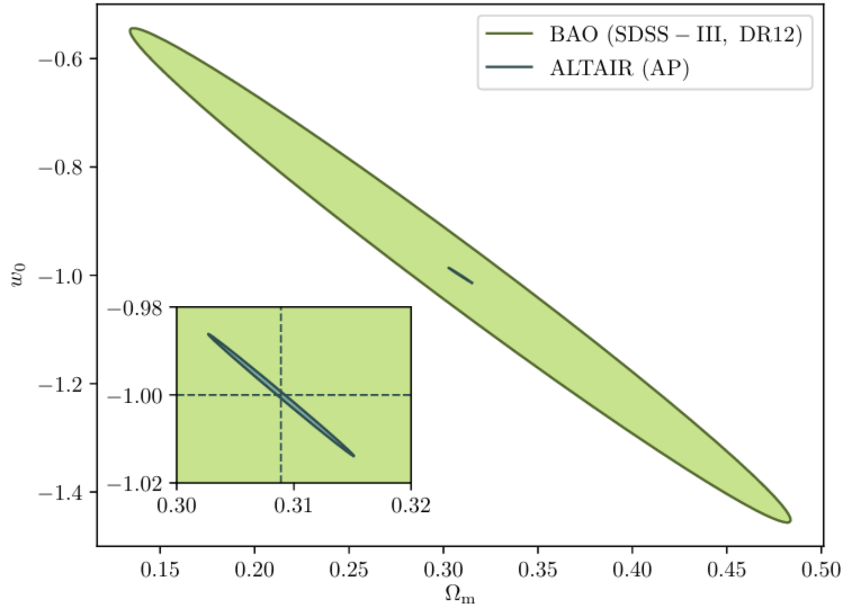
Porqueres, Kodi Ramanah, Jasche & Lavaux, 1812.05113
Lavaux, Jasche & FL, 1909.06396

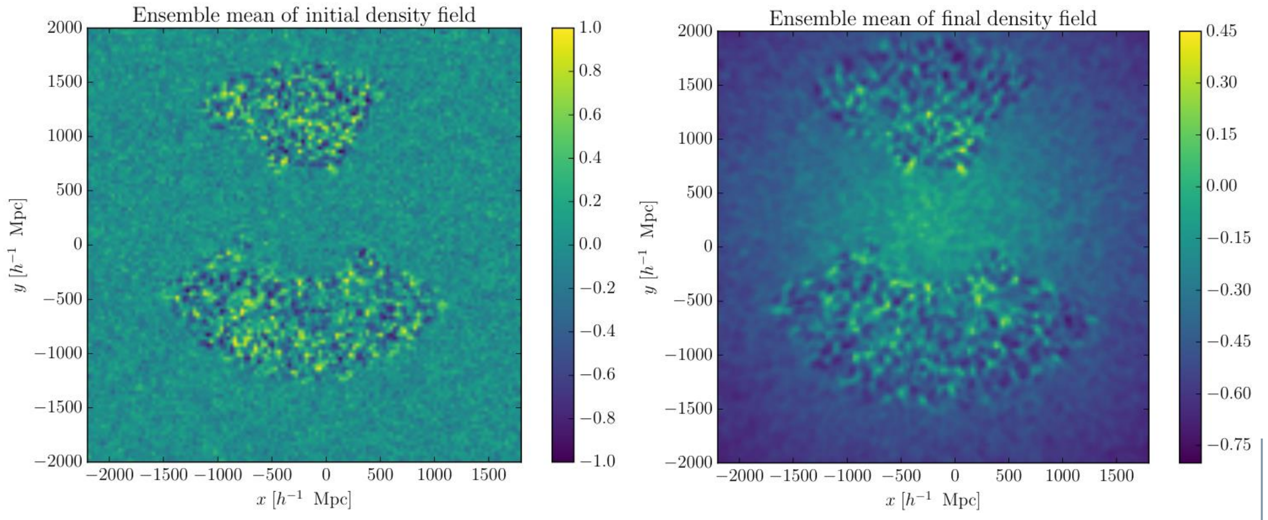
Kalus, Percival *et al.*, 1806.02789

Inferring cosmology: Alcock-Paczyński expansion test

- Field-level expansion test through the Alcock-Paczyński (AP) effect jointly constrains fields and cosmological parameters ($\Omega_{\rm m}, w$).



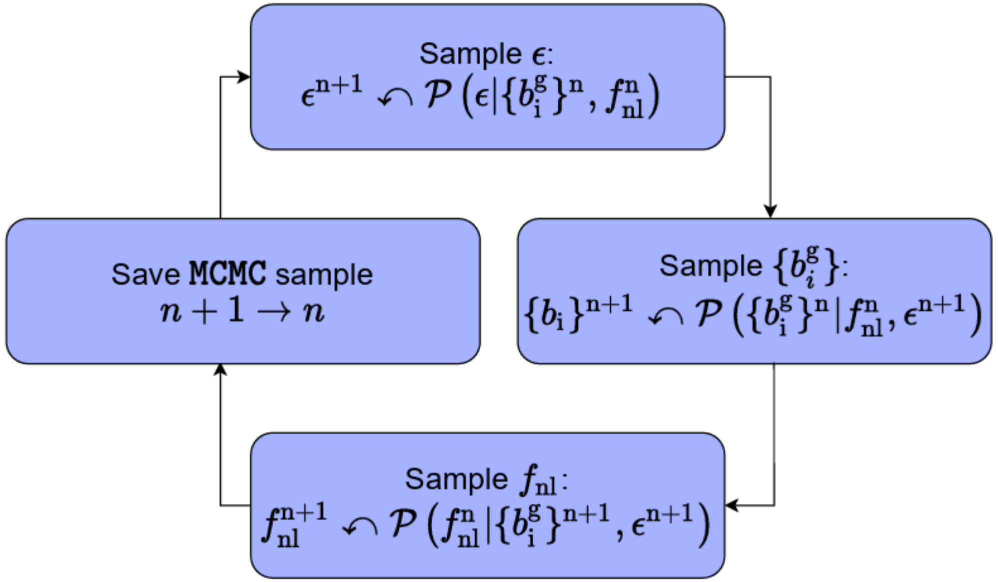


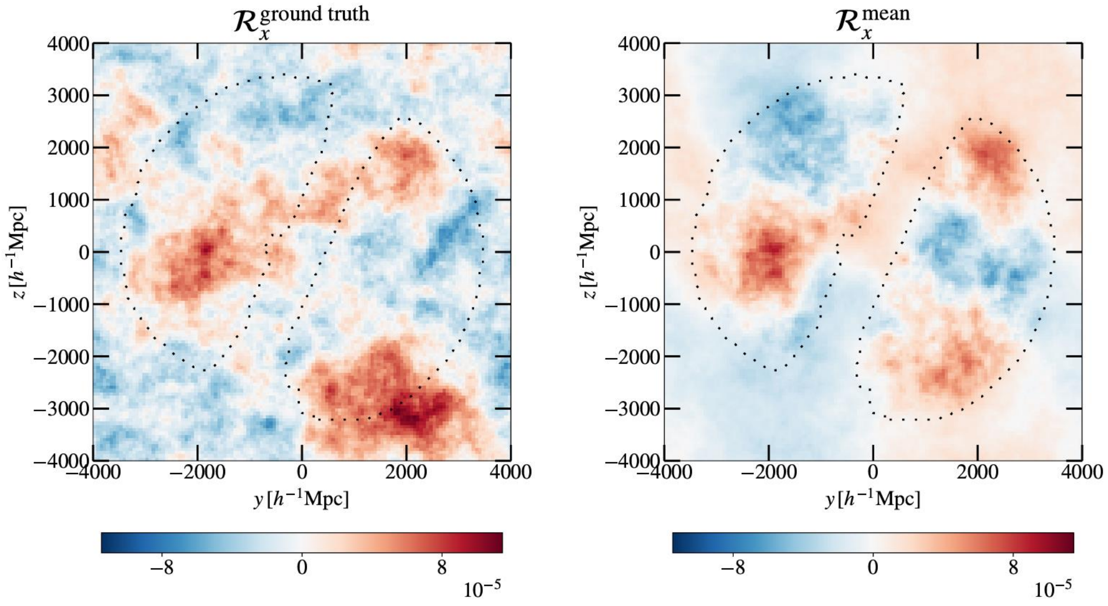


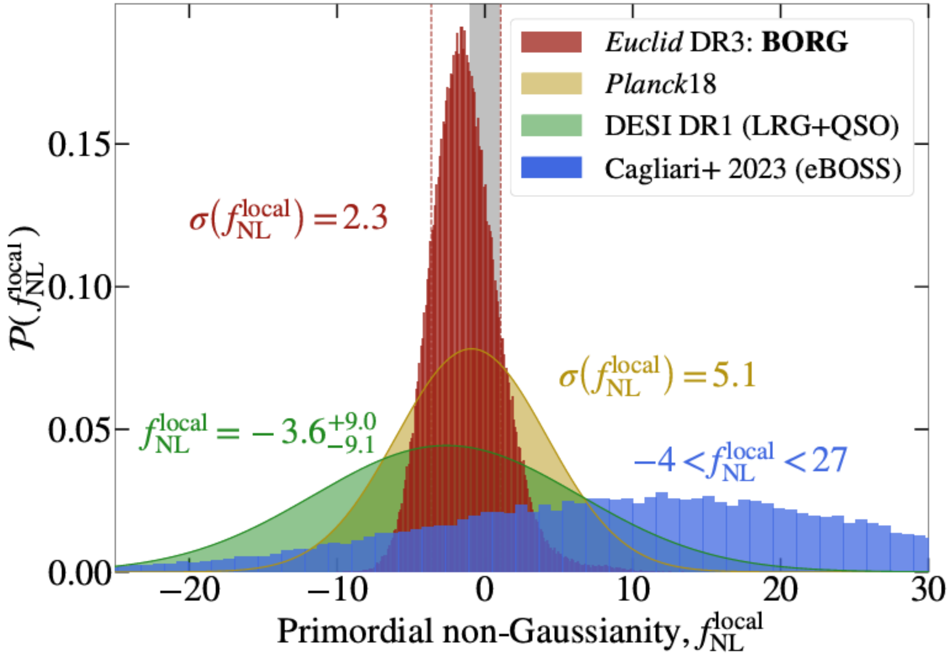
Kodi Ramanah *et al.*, 1808.07496

Inferring cosmology: primordial non-Gaussianity

- Joint inference of the primordial curvature fluctuation field and local primordial non-Gaussianity parameter f_{NL}.



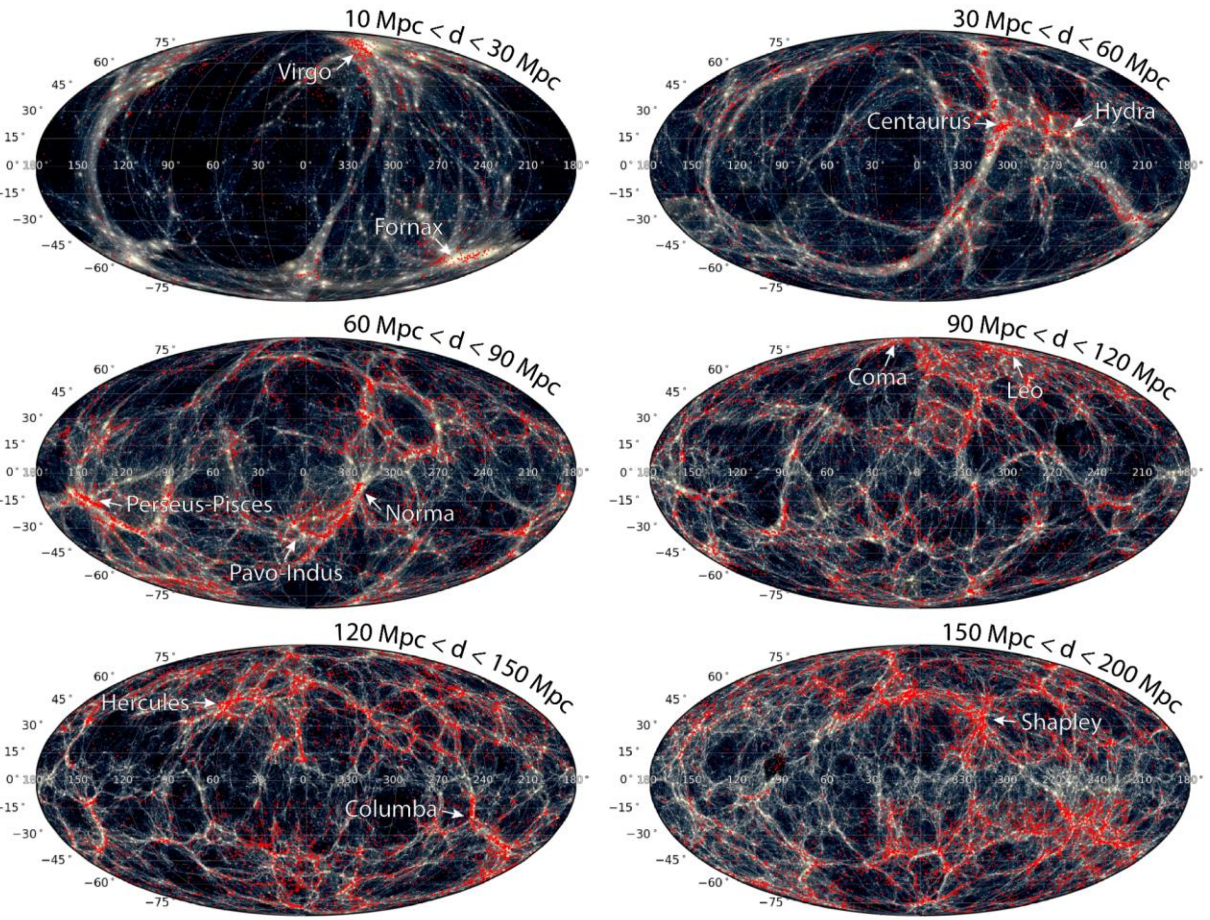


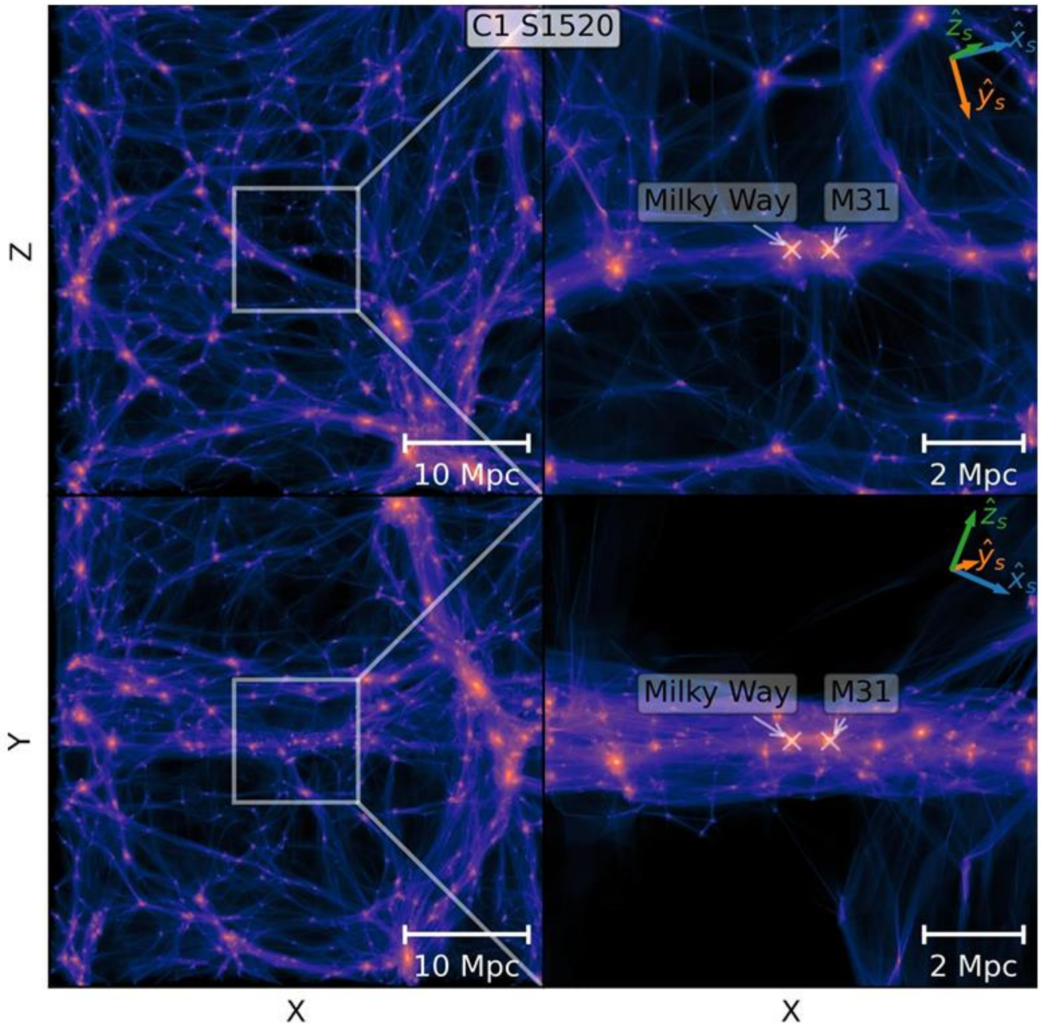


Andrews *et al.*, 2203.08838; Euclid Collaboration (Andrews *et al.*), 2412.11945

Small scales: constrained simulations and zoom-in reconstructions

- SIBELIUS & SIBELIUS-DARK simulations



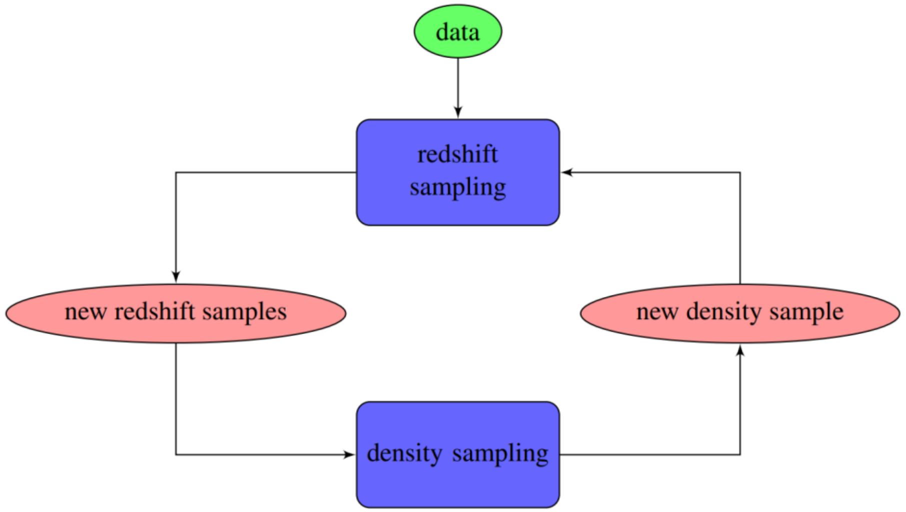


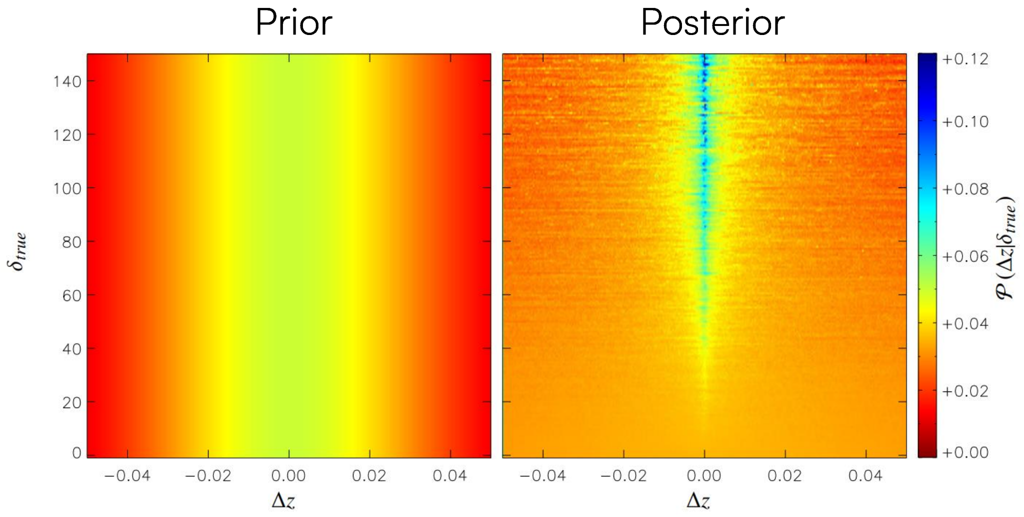
Sawala *et al.*, 2103.12073; McAlpine *et al.*, 2202.04099

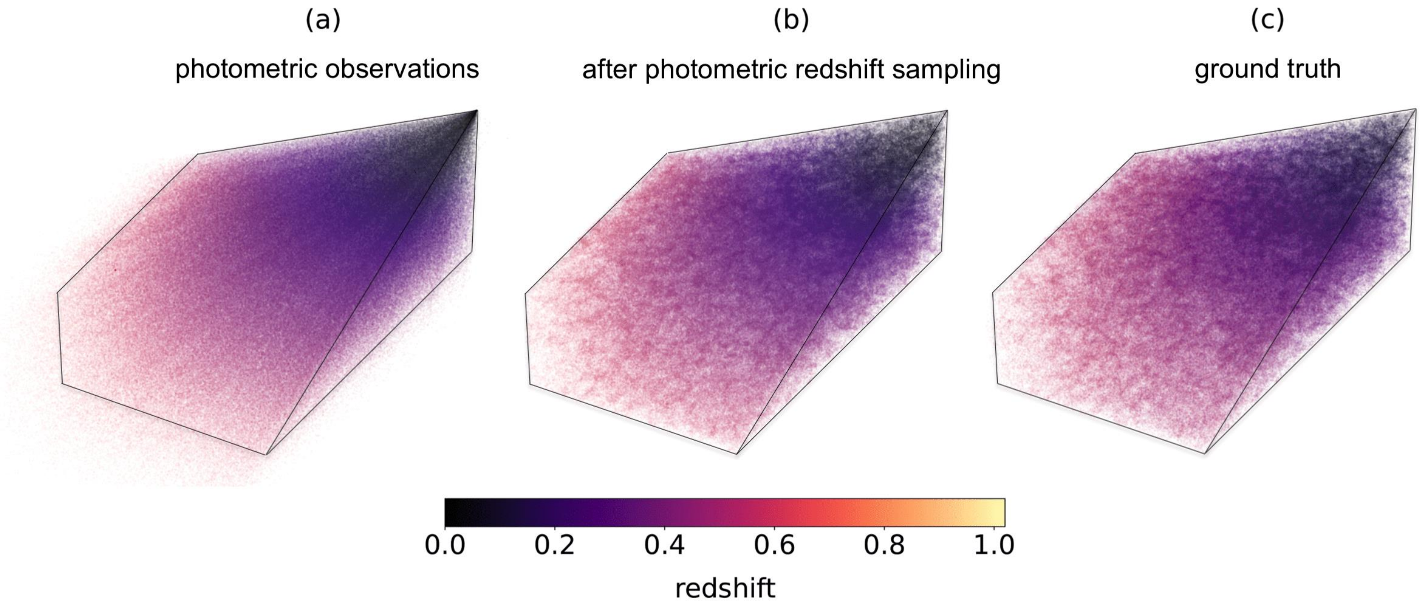
Wempe *et al.*, 2406.02228

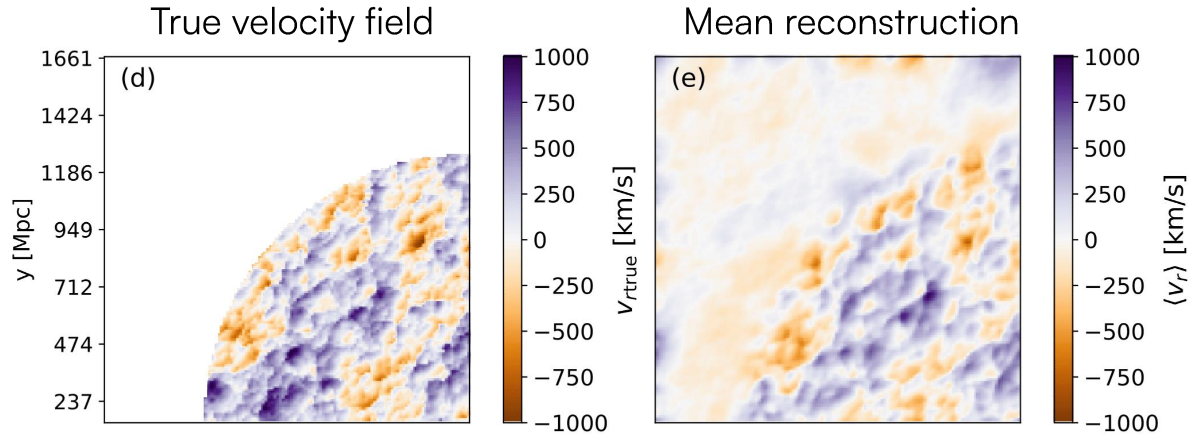
FIELD-LEVEL INFERENCE FOR OTHER/JOINT COSMOLOGICAL PROBES

Photometric galaxy clustering





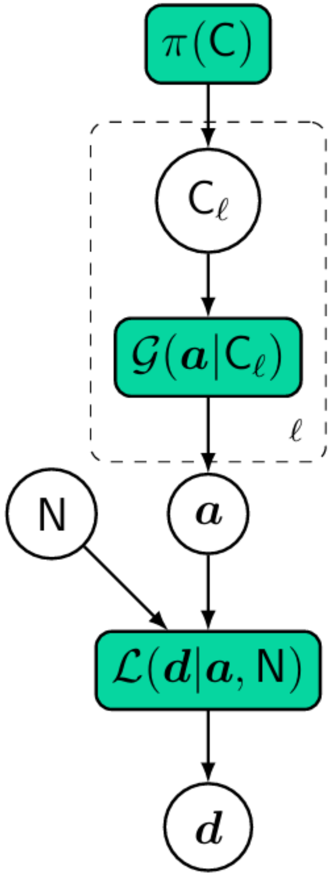


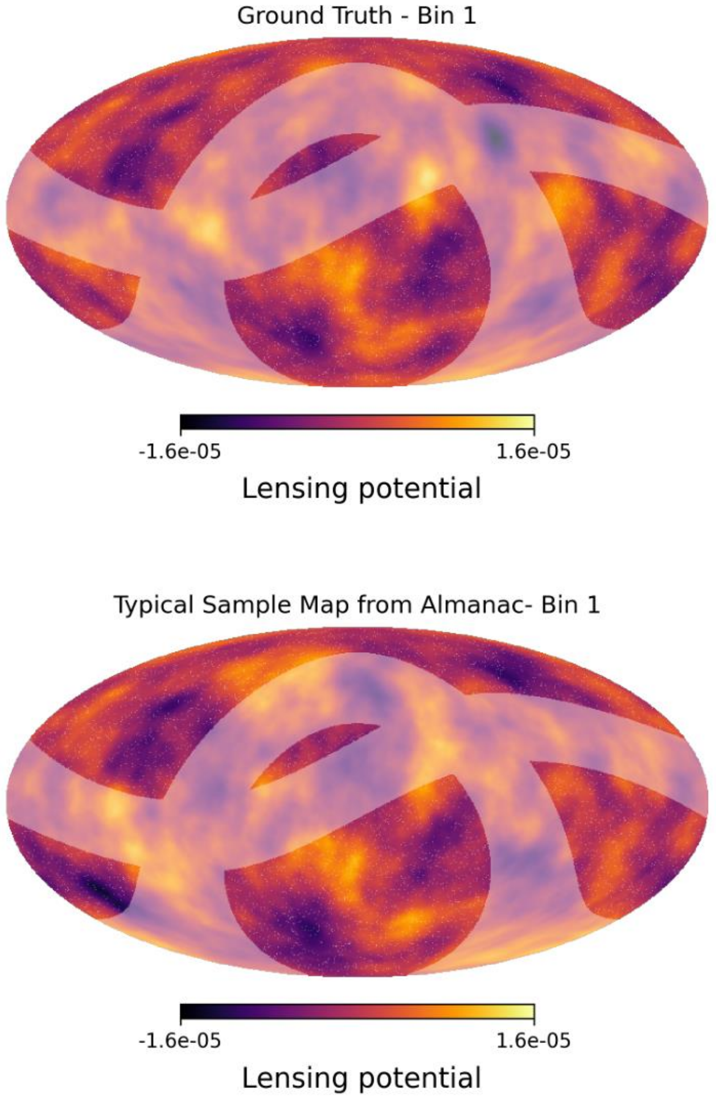


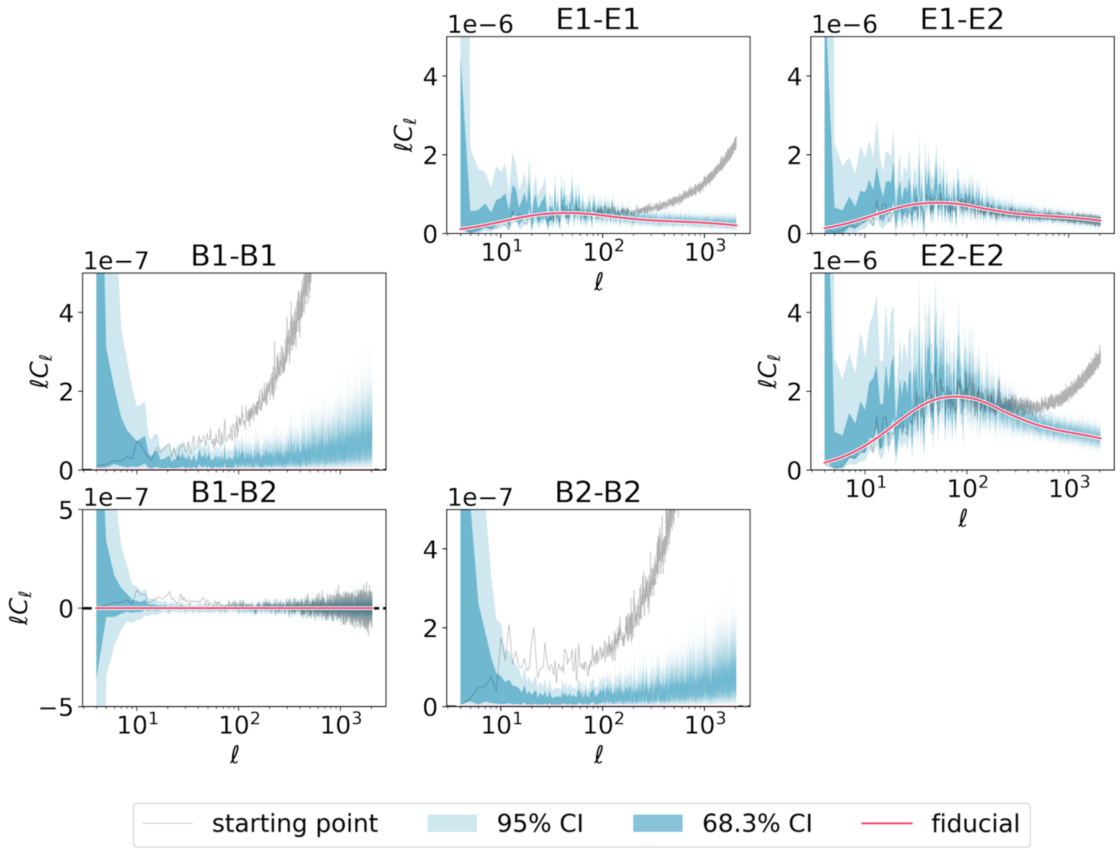
Jasche & Wandelt, 1106.2757; Tsaprazi *et al.*, 2301.03581

Weak lensing: map and power spectrum inference on the sphere

- Almanac algorithm



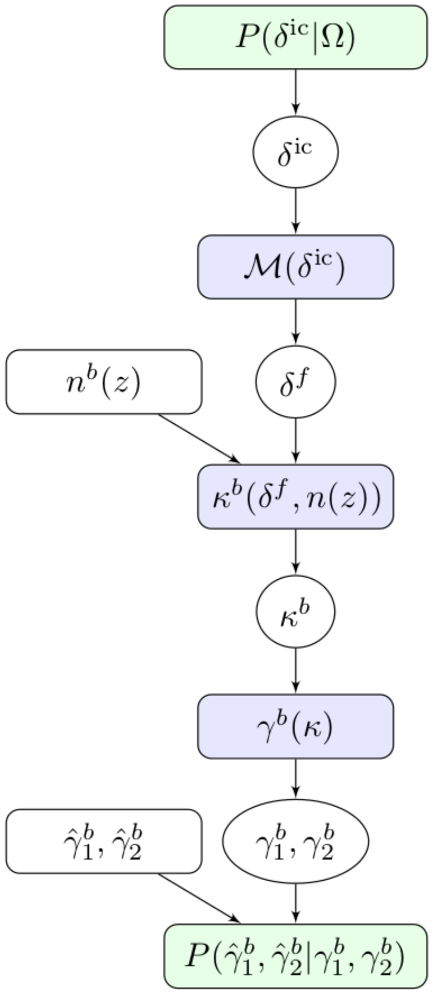


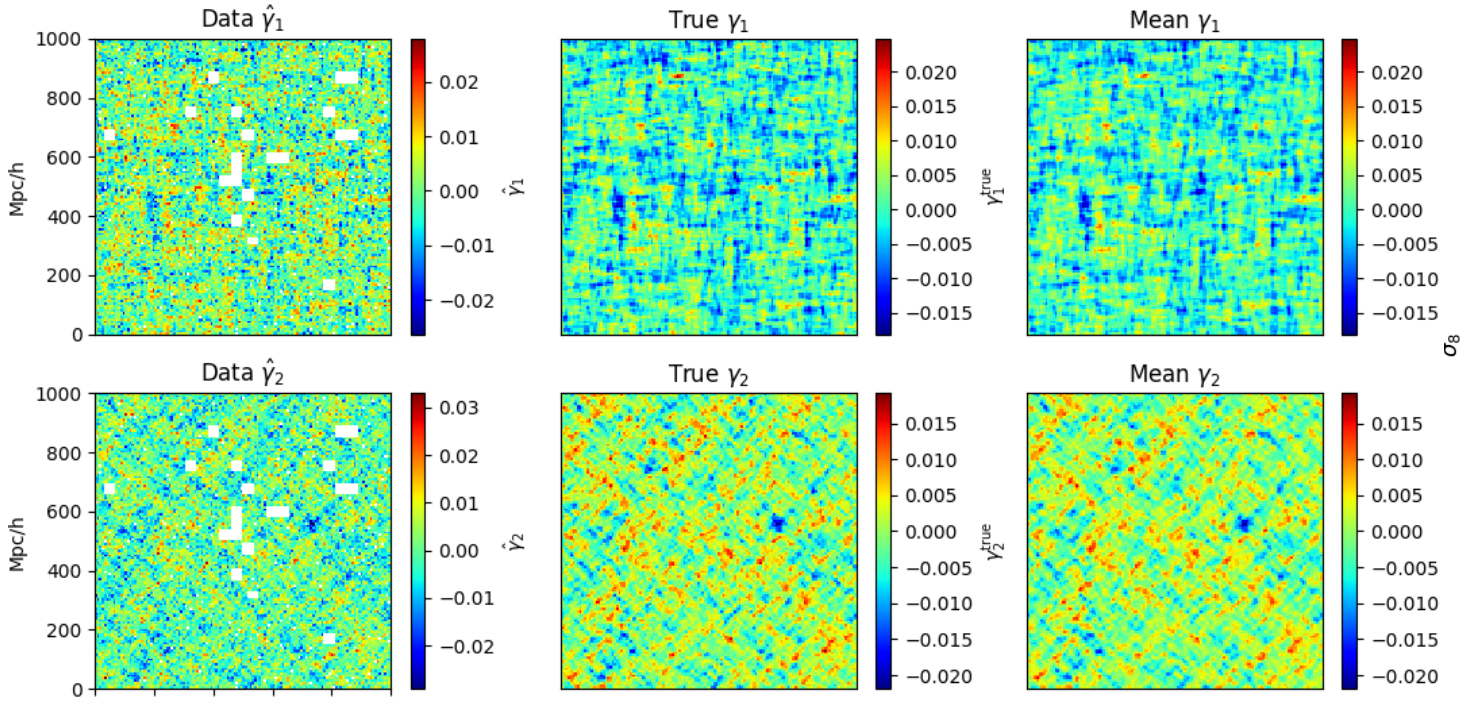


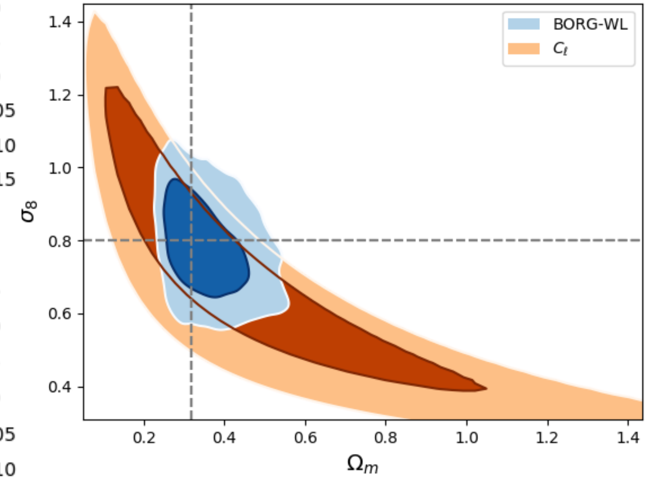
Loureiro *et al.*, 2210.13260; Sellentin *et al.*, 2305.16134

Weak lensing: three-dimensional map and cosmological parameter inference

- BORGWL algorithm



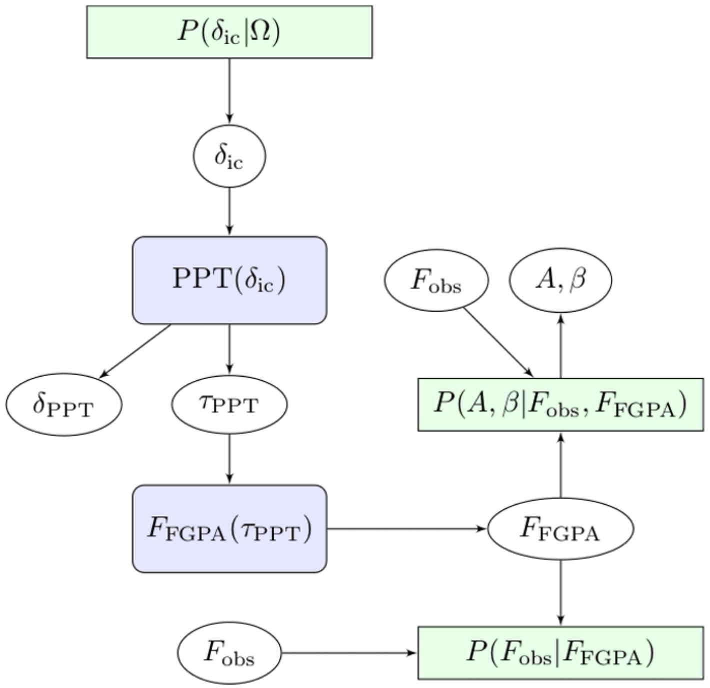


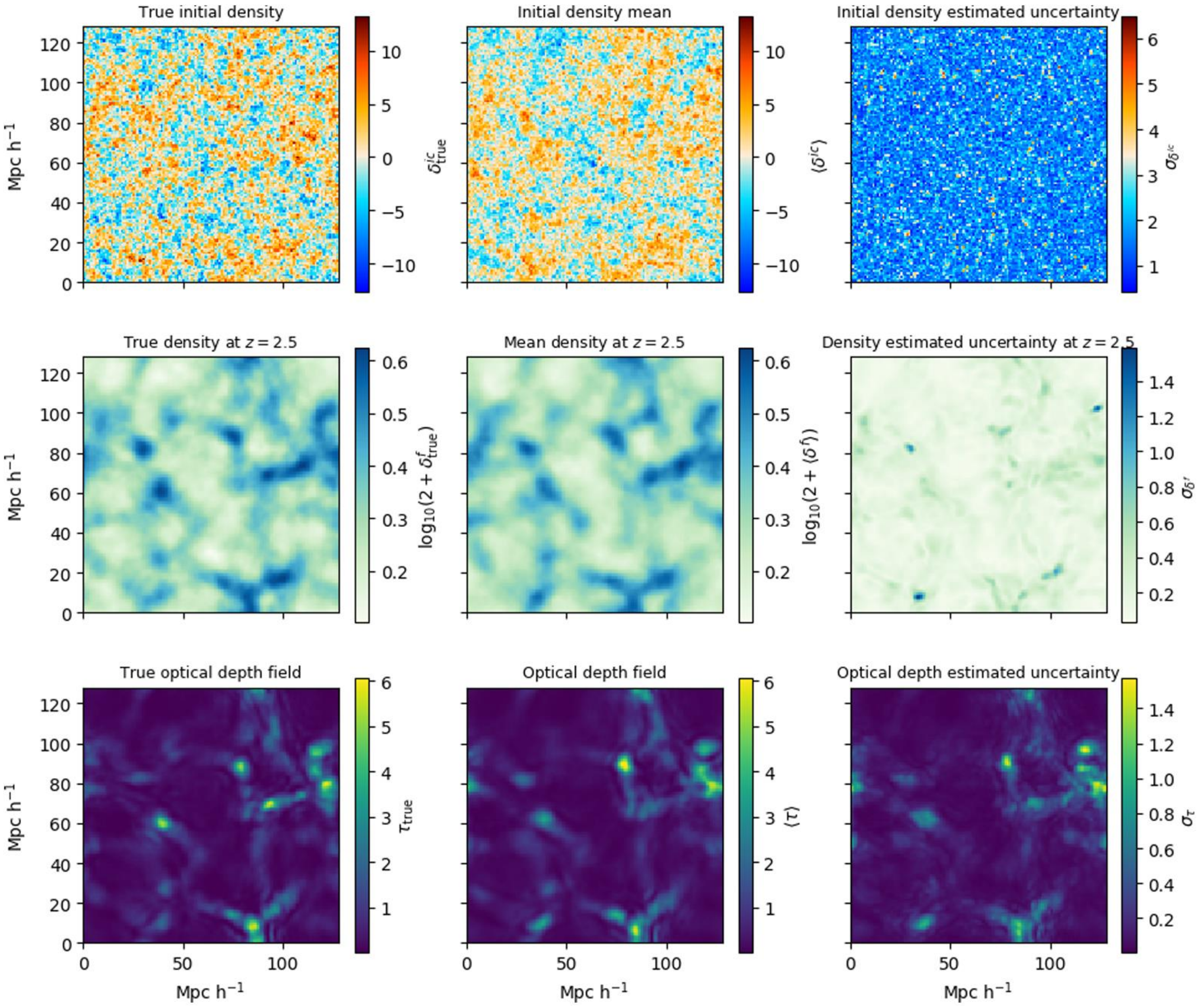


Porqueres, Heavens, Mortlock & Lavaux, 2011.07722; Porqueres, Heavens, Mortlock & Lavaux, 2108.04825

Lyman-alpha forest

- Data model for quasar spectra based on propagator perturbation theory (PPT) and the fluctuating Gunn-Peterson approximation (FGPA)

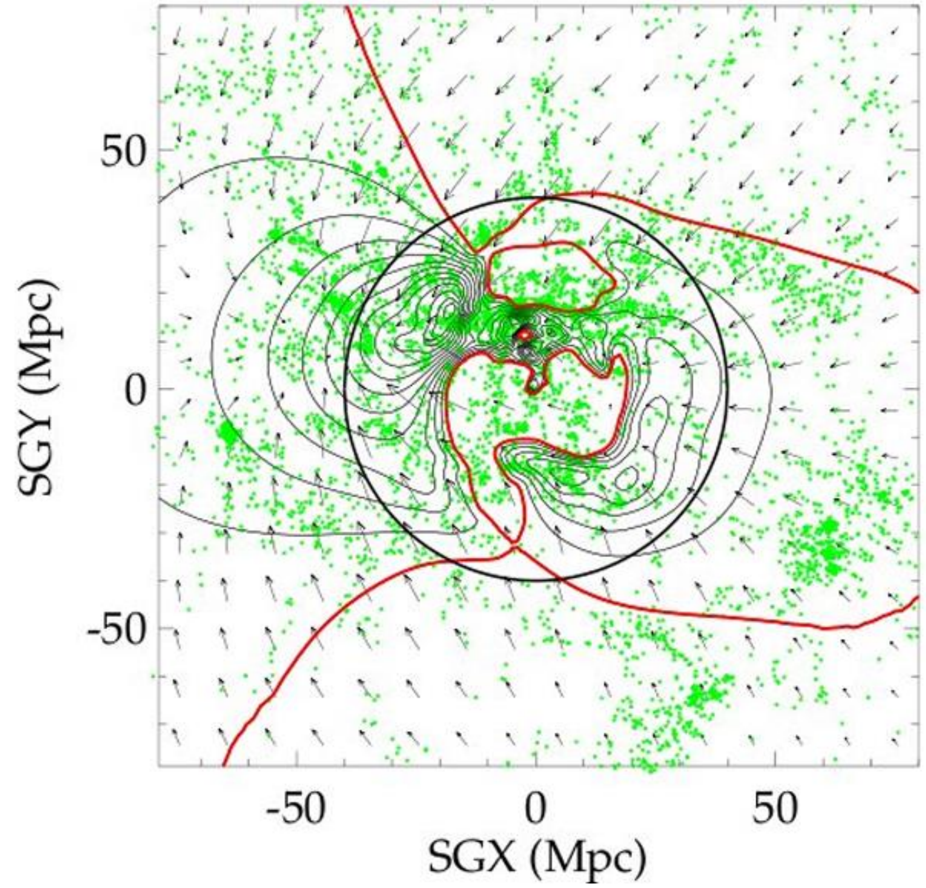


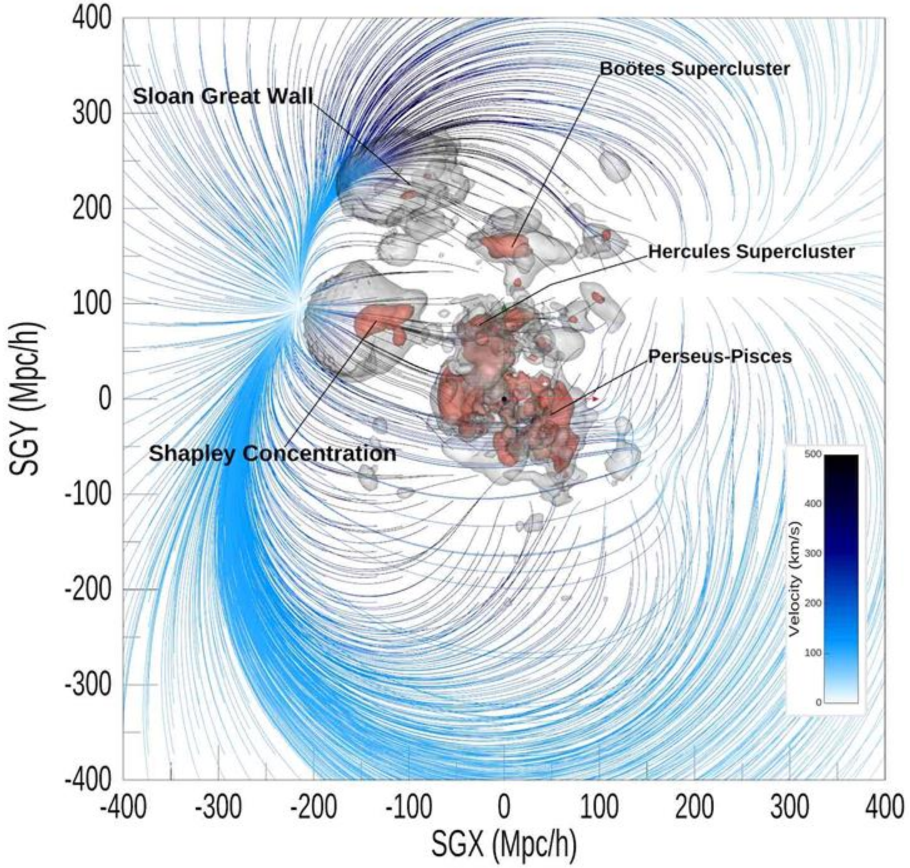


Porqueres, Hahn, Jasche & Lavaux, 2005.12928; Boonkongkird *et al.* 2303.17939

Velocity field inference from peculiar velocity tracers

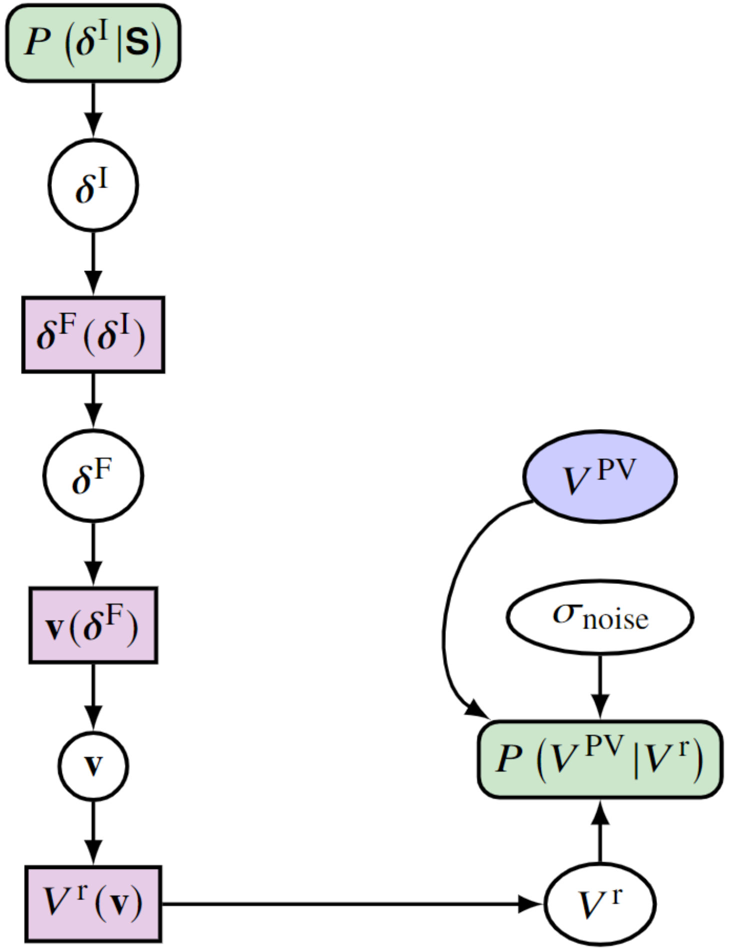
- Peculiar velocity data are combined in CosmicFlows-1 to 4 data
- Reconstruction of the field is traditionally done via Wiener filtering

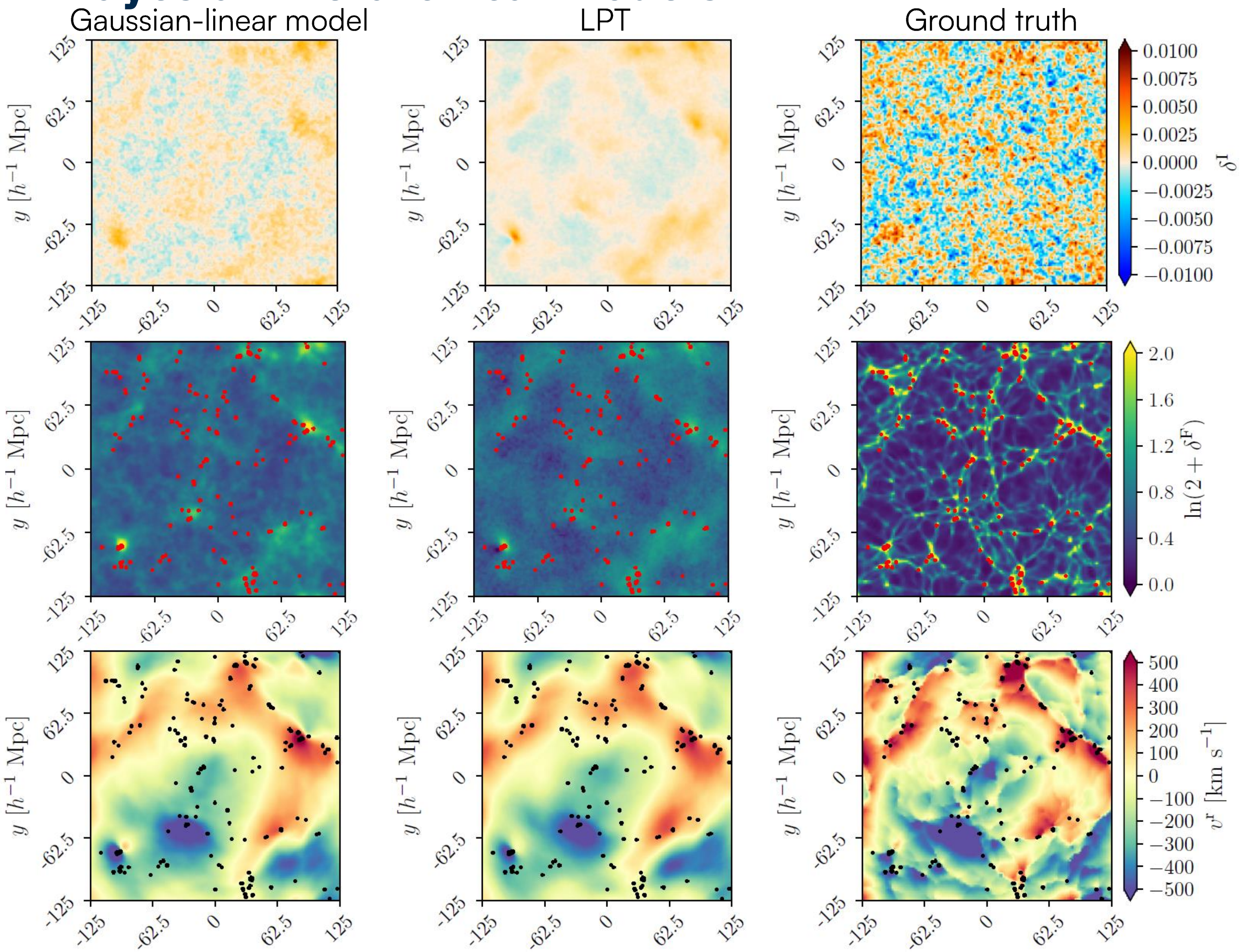




Zaroubi *et al.*, astro-ph/9410080; Courtois *et al.*, 1109.3856; Hoffman *et al.*, 1503.05422; Hoffman *et al.*, 2311.01340; CLUES collaboration

Velocity field inference with Bayesian hierarchical models

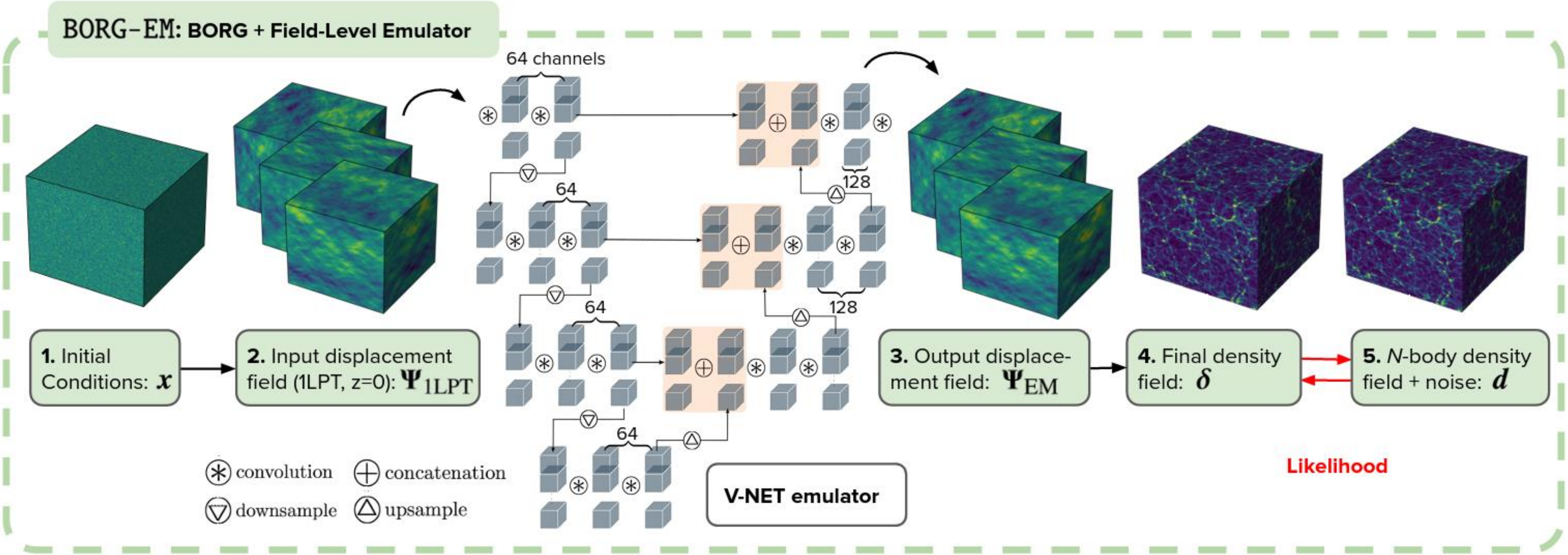


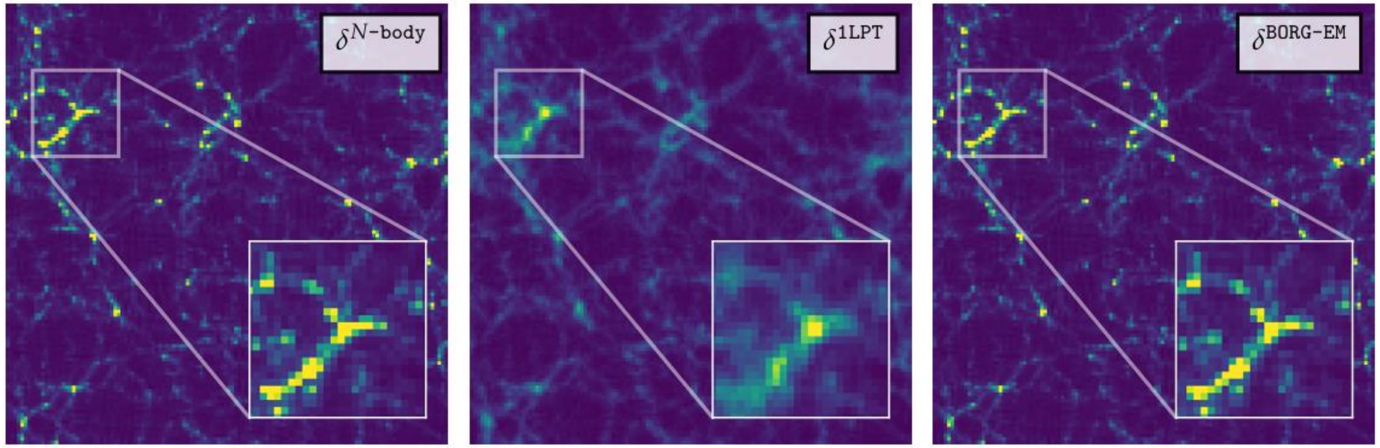


Lavaux, 1512.04534; Boruah, Lavaux & Hudson, 2111.15535; Prideaux-Ghee, FL, Lavaux, Heavens & Jasche, 2204.00023

MACHINE LEARNING FOR FIELD-LEVEL INFERENCE

Emulators: accelerated forward models at non-linear scales





Doeser *et al.*, 2312.09271

Approximate posterior sampling using diffusion models

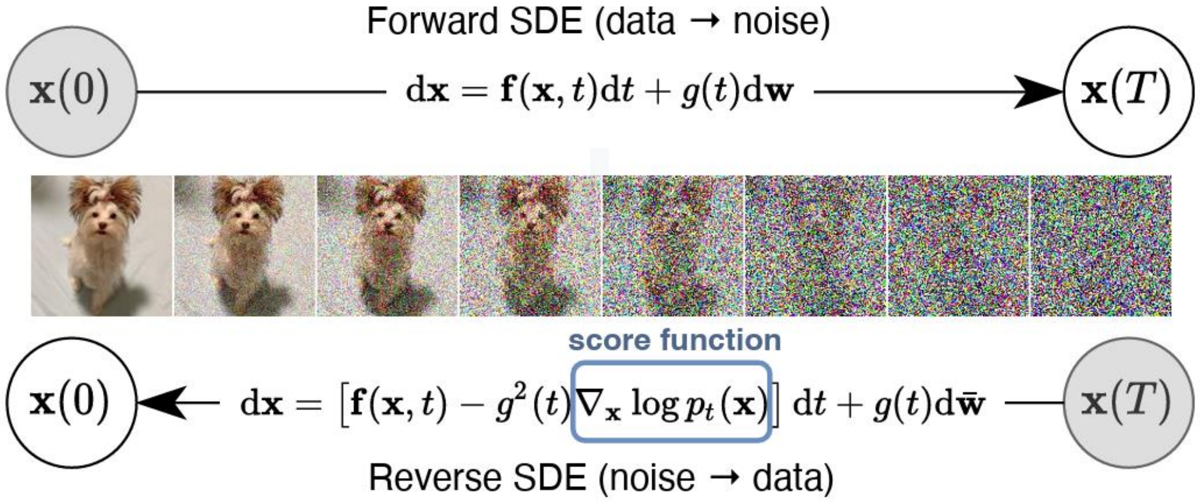
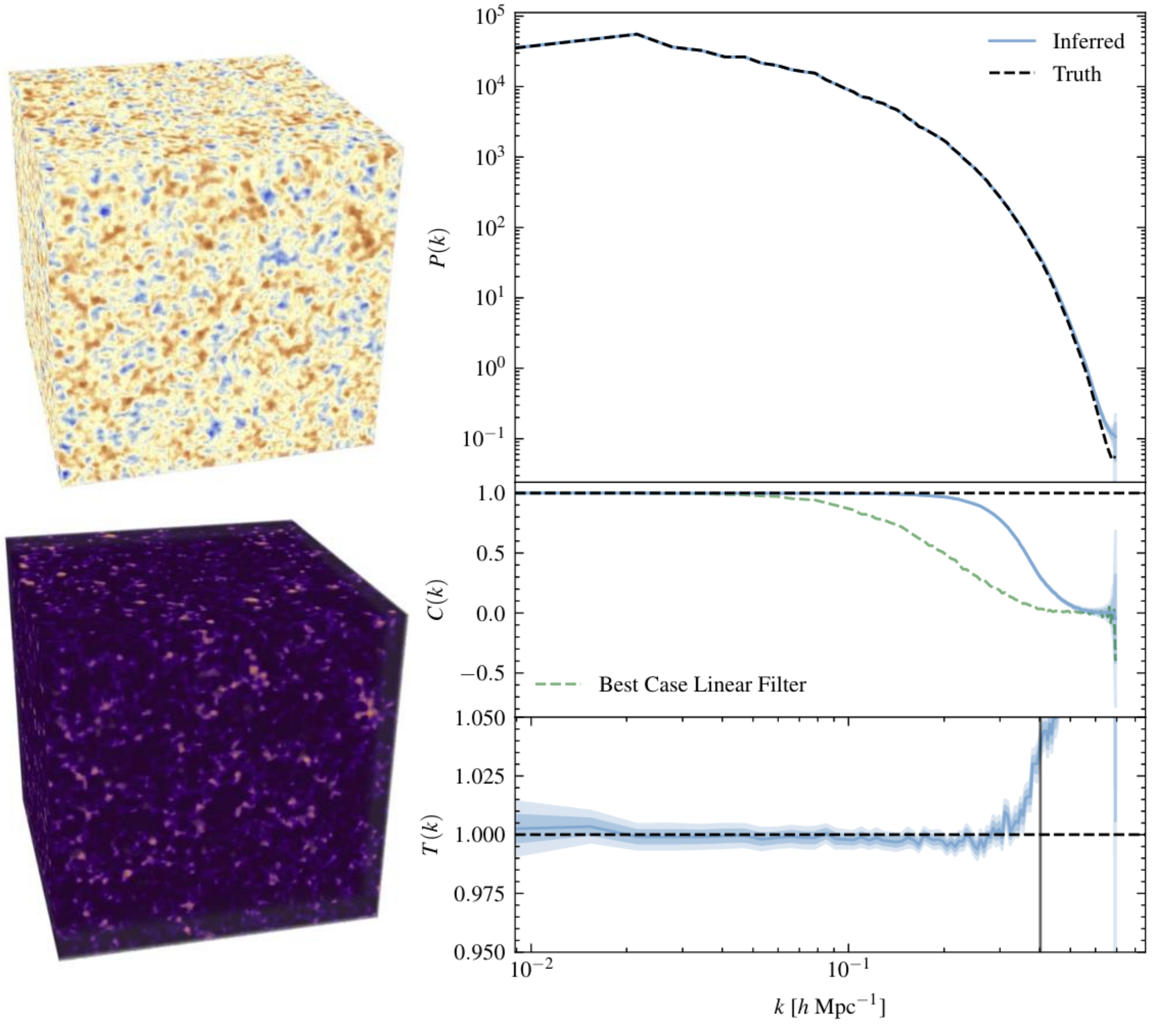


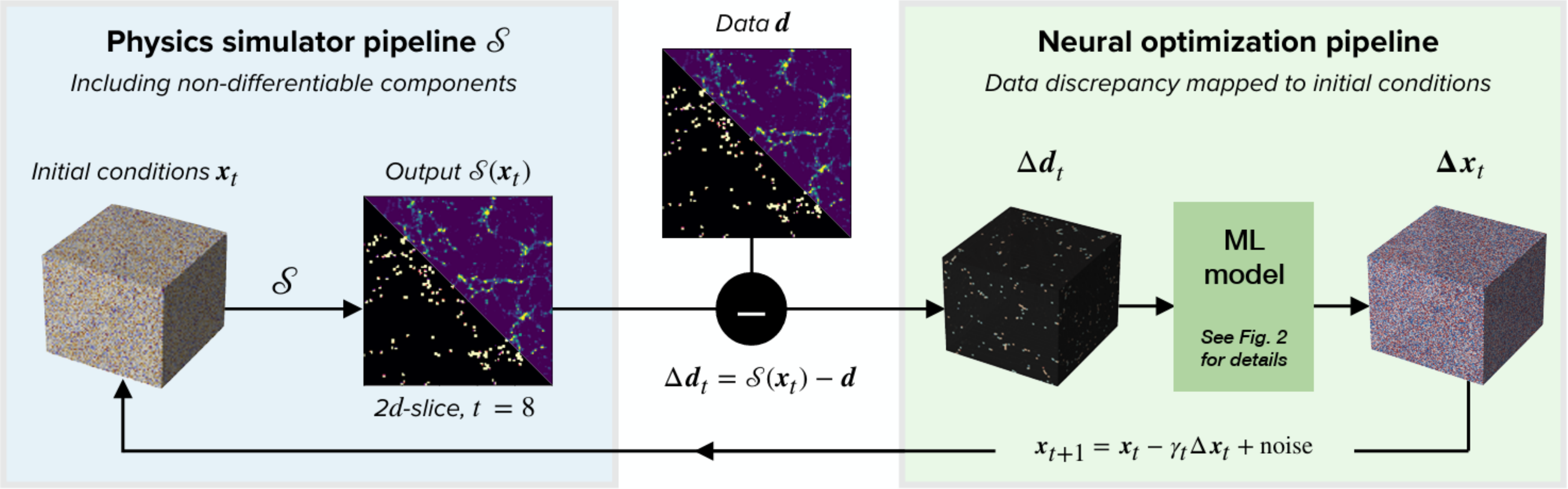
Figure from Song *et al.*, 2011.13456

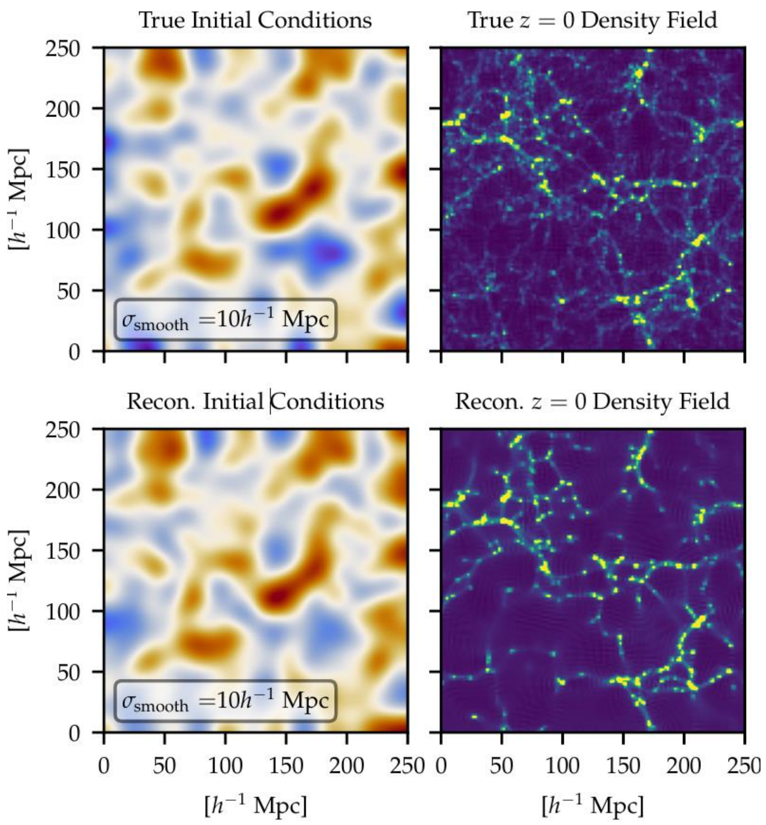


Legin *et al.*, 2304.03788

LULO: initial condition search process with a neural optimiser

- Learning the Universe: Learning to Optimise (LULO): enables the use of **non-differentiable** models (i.e. field-level inference becomes close to implicit inference).





[Doeser, Ata & Jasche, 2502.13243](https://arxiv.org/abs/2502.13243)

Thanks for listening!

https://florent-leclercq.eu/teaching.php