Bayesian field-level inference of primordial non-Gaussianity using next-generation galaxy surveys

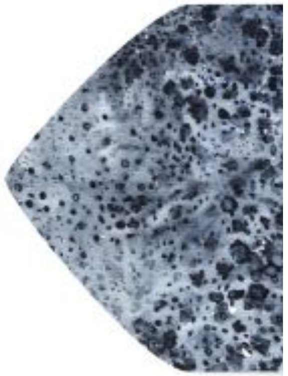
Guilhem Lavaux (IAP/CNRS)

A. Andrews (INAF Bologna)
J. Jasche (Stockholm U.)

with Aquila consortium, Euclid Collaboration & Learning the Universe collaboration

GDR CoPhy, IP2I, Lyon (May 21-23, 2024)

Cosmological context: current paradigm

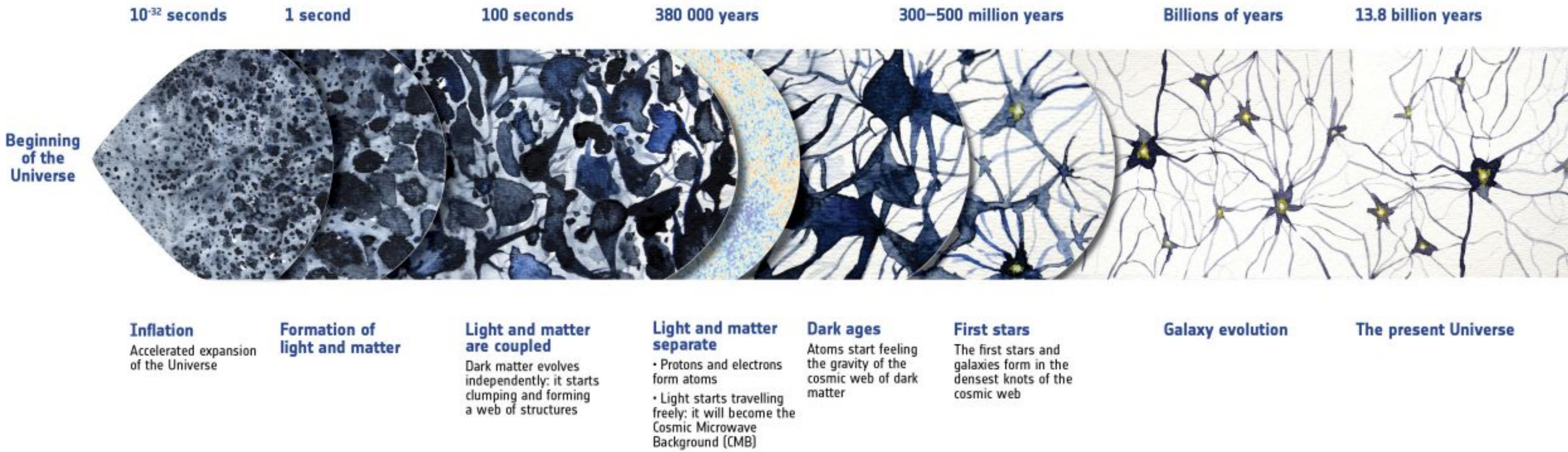


We want physics here

We observe here

Image credit: Planck collaboration

Cosmological context: current paradigm



Dynamical evolution of the universe from first instant to present time

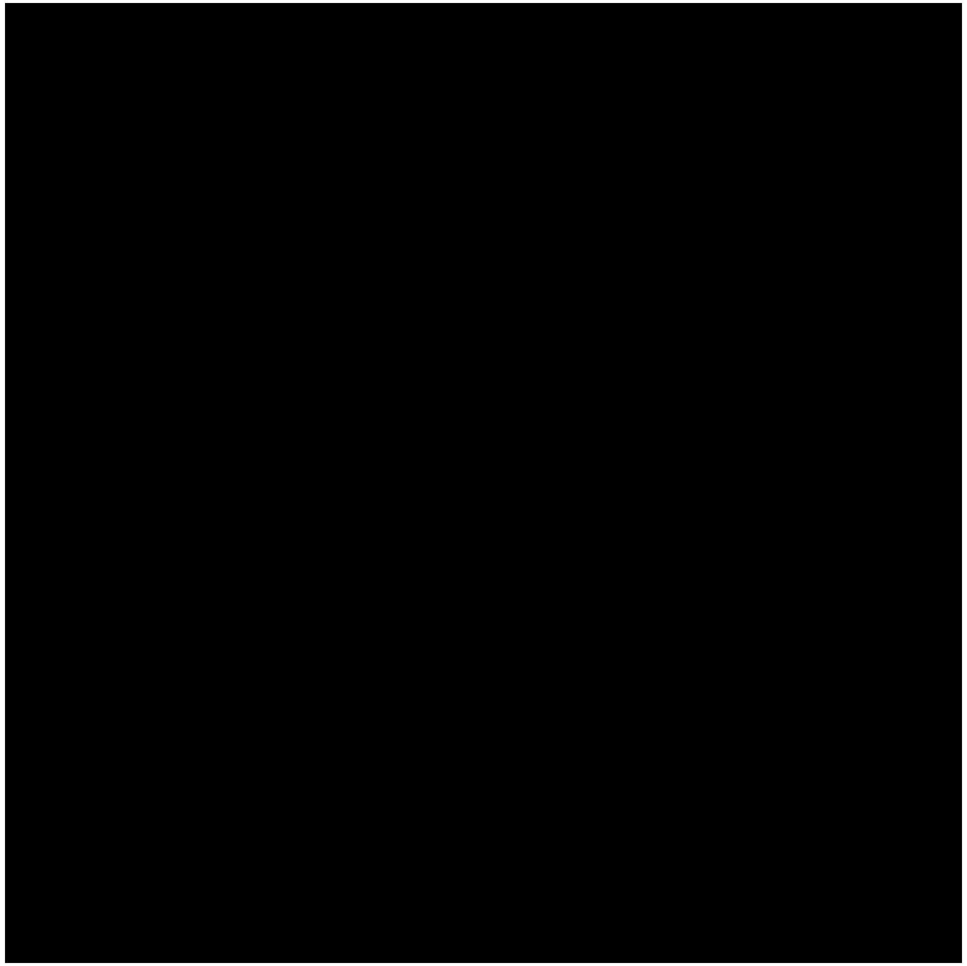
Causality

Image credit: Planck collaboration

Motivation: A complete characterization of cosmic structure

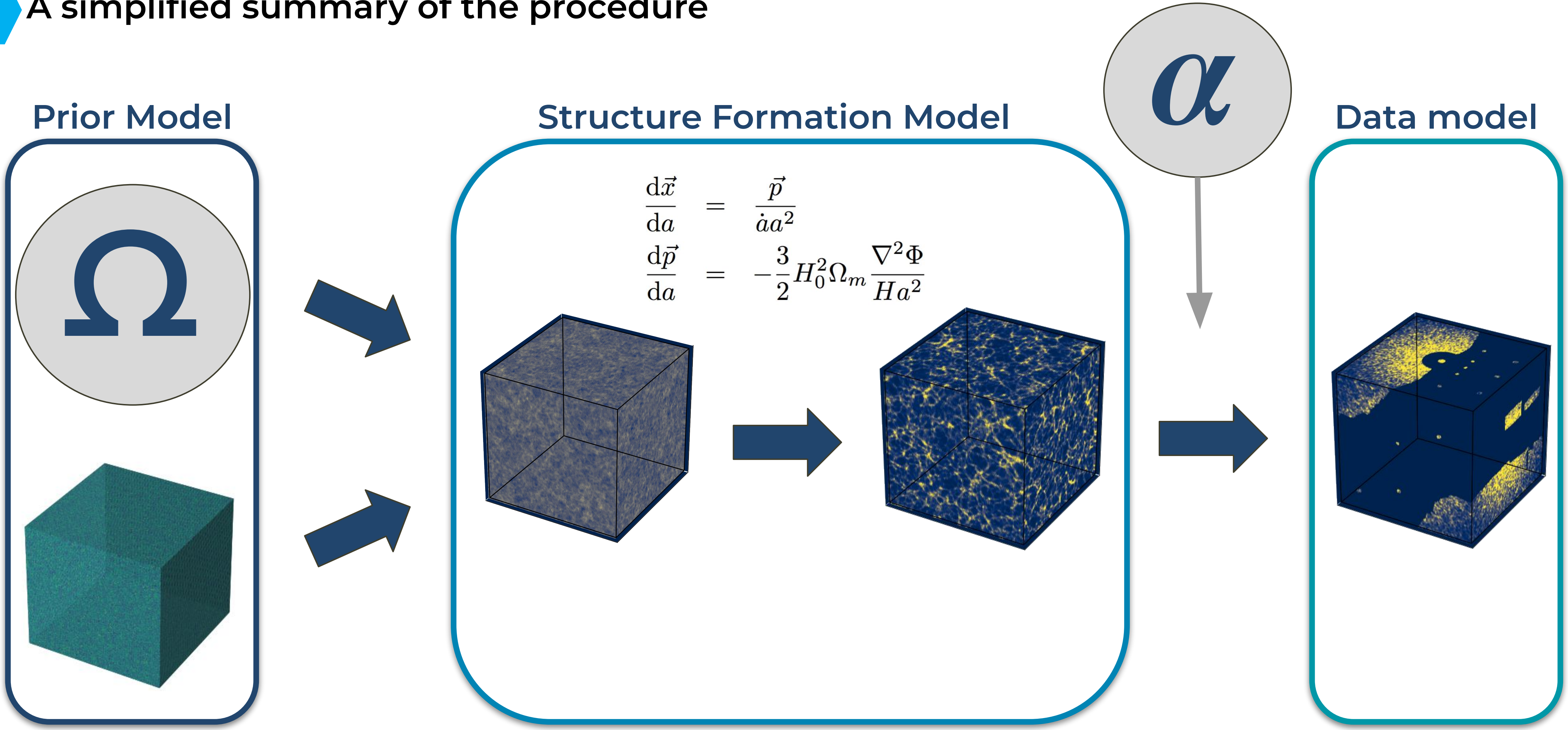
Bayesian Physical forward modeling

- **Field-level inference**
  - Beyond summary statistics
  - Beyond random realizations
- **Causal inference**
  - Beyond associative analyses
  - Easier to incorporate systematic effects than on summaries
  - Harder to separate model misspecifications
- **Non-linear and dynamical inference**
  - Beyond linear structure growth
  - Redshift Distortions
  - Light-Cone effects



Jasche & Lavaux 2019

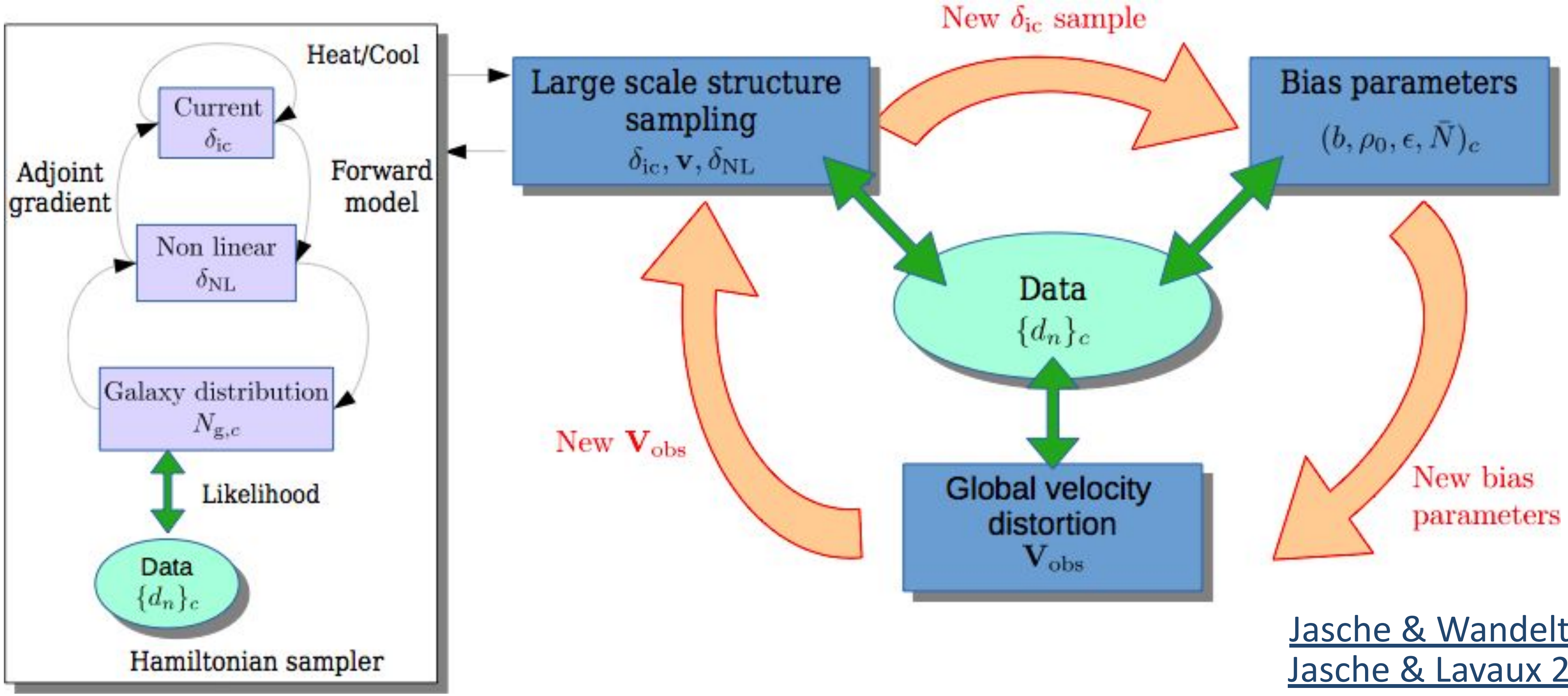
A simplified summary of the procedure



$$\pi\left(\mathbf{x}, \boldsymbol{\Omega}\right) \qquad \pi\left(\rho_{\mathbf{m}} \middle| \mathbf{x}, \boldsymbol{\Omega}\right) \qquad \pi\left(\mathbf{N_g} \middle| \rho_{\mathbf{m}}, \alpha, \boldsymbol{\Omega}\right)$$

BORG: A large scale MCMC framework

- **BORG's MCMC framework allows building flexible data models**
  - Hierarchical Bayes and block sampling
  - Efficient Hamiltonian Monte Carlo **(HMC)** technique
  - **Fully differentiable physics forward model**



Jasche & Wandelt 2014
Jasche & Lavaux 2019

1

Inferring f_{NL} with field level inference

$$\Phi_{\mathrm{primord}} = \phi_g + f_{\mathrm{NL}} \left(\phi_g^2 - \langle \phi_g^2 \rangle \right)$$

Other generators in Scoccimarro et al. (2012)

The dynamical forward model

Different choice possible:

- Log transform
- nLPT (1LPT, 2LPT)
- Quantum LPT **(Uhlemann et al. 2019)**
- PM-COLA **(Tassev et al. 2013)**
- LPT+Emulator (BORG-EMU, **Jamieson et al. 2023, Doeser et al. 2024**)
- Zoom-PM **(Wempe et al. 2024 in prep)**

For this work, we use 2LPT

Choice of dark matter / galaxy relation (1)

Scale dependent bias

$$\rho_g = \bar{N}\left[1 + b_1\delta_m + b_\phi f_{\rm NL}\phi_g^q\right]$$
$$\rho_g = \bar{N}\left[1 + b_1\delta_m + b_\phi f_{\rm NL}(Q * \delta_m)\right]$$

$$Q(k) \propto \frac{3\Omega_m H_0^2}{2k^2 T(k)}$$

Dalal et al. (2008), Slosar et al. (2008)

Advantages:

Reproduce phenomenology
Easy to implement
Relates to work on perturbation theory

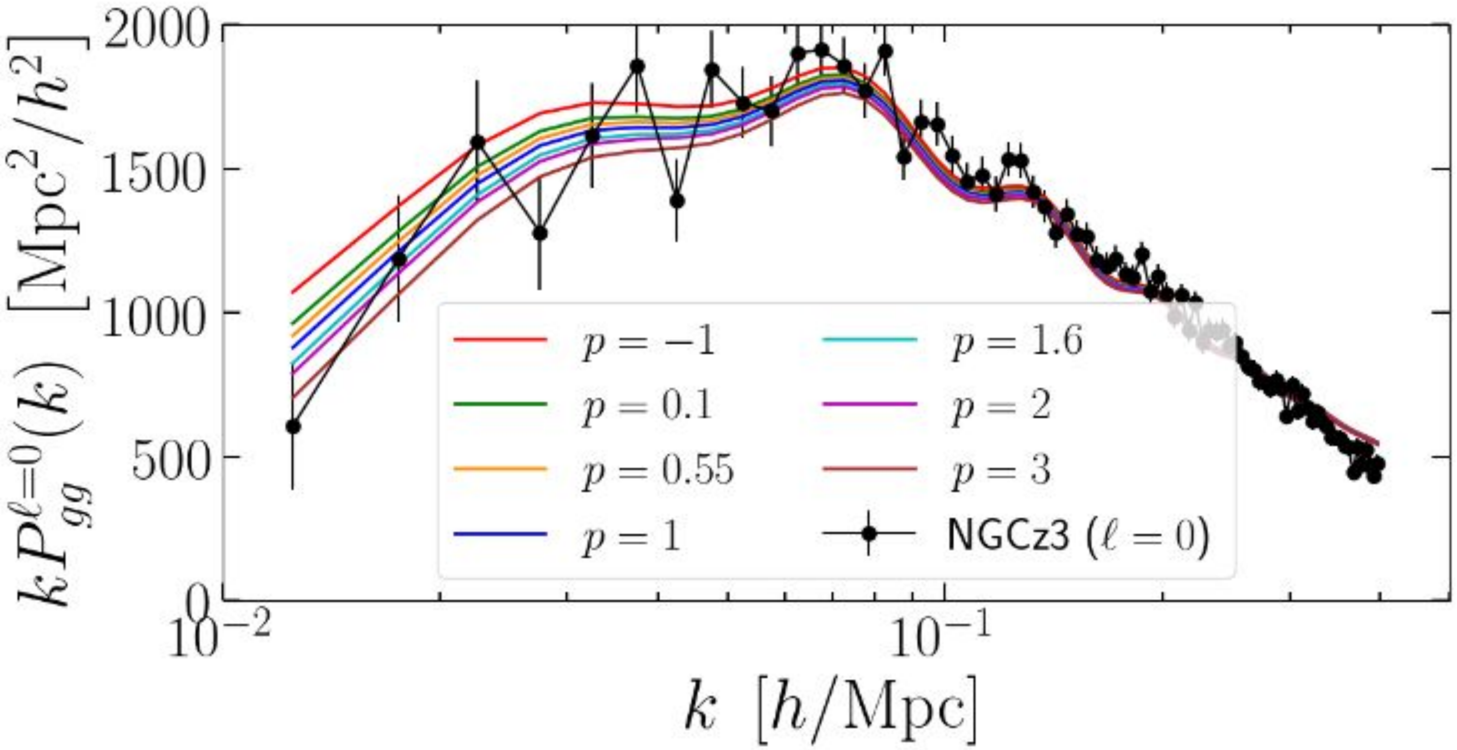
Disadvantages:

Harder to work on real data
Limited spatial resolution of the model

Choice of dark matter / galaxy relation (2) : impact of choices of bΦ / b1 relation

Scale dependent bias

$$\rho_g = \bar{N}\left[1 + b_1\delta_m + b_\phi f_{\rm NL}\phi_g^q\right]$$



f_{NL}=50, b_1=2.15

Barreira (2022)

Choice of dark matter / galaxy relation (3): further improvements

Scale dependent bias

$$\rho_g = \bar{N}\left[1 + b_1\delta_m + b_\phi f_{\rm NL}\phi_g^q\right]$$
$$\rho_g = \bar{N}\left[1 + b_1\delta_m + b_\phi f_{\rm NL}(Q * \delta_m)\right]$$

$$b_\phi = 2\delta_c(b_1 - p)$$
$$Q(k) \propto \frac{3\Omega_m H_0^2}{2k^2 T(k)}$$

Dalal et al. (2008), Slosar et al. (2008)

Scale dependent bias (higher-order, Andrews et al in prep.)

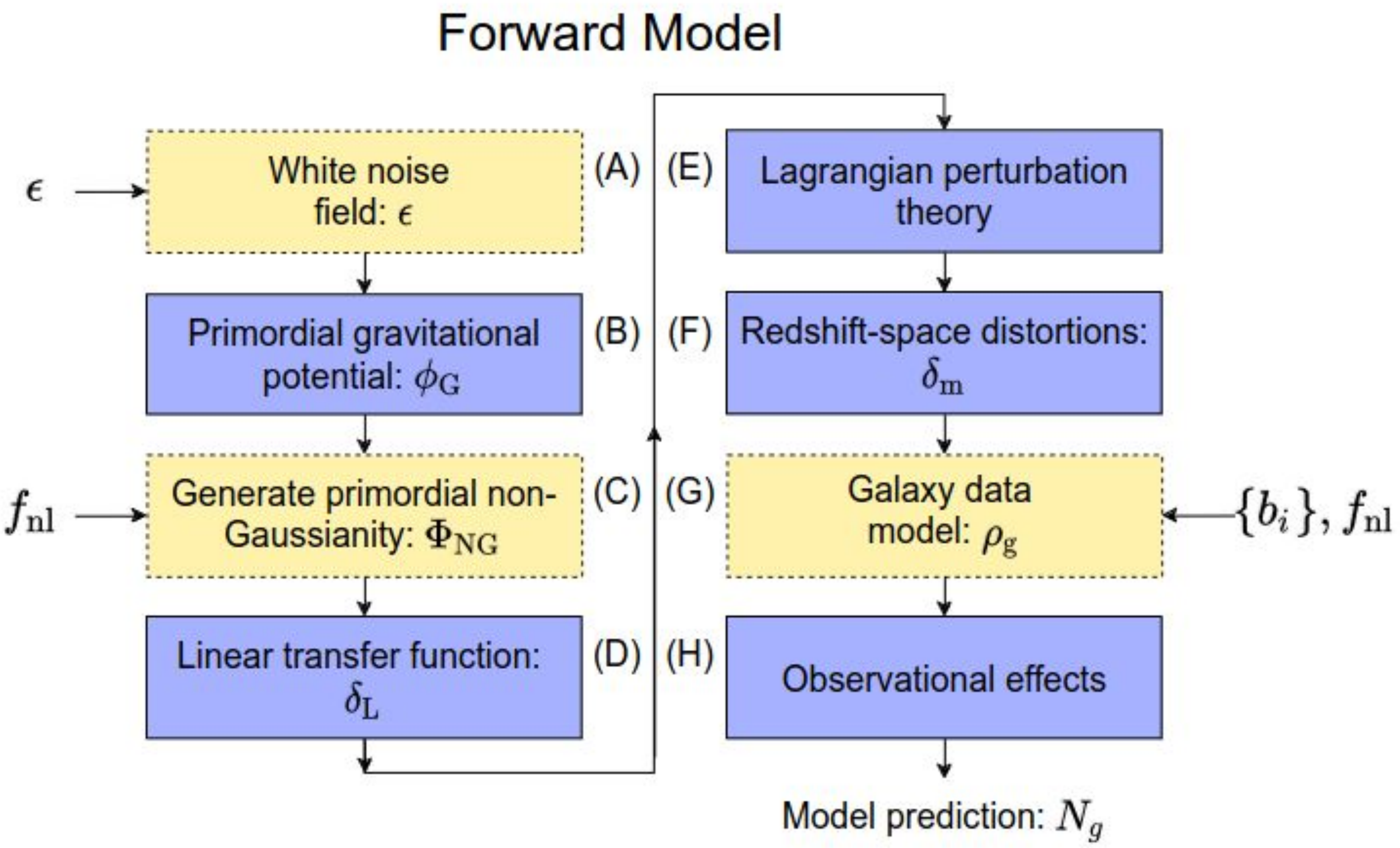
$$\rho_g = \bar{N}\left[1 + b_1\delta_m + \frac{b_2}{2}\delta_m^2 + b_K K^2 + b_\phi f_{\rm NL}\phi_g^q + b_{\phi,\delta} f_{\rm NL}\delta_m\phi_g^q\right]$$

Lazeyras et al. (2021), Barreira et al. (2021)

PineTree model (Ding et al. in prep)

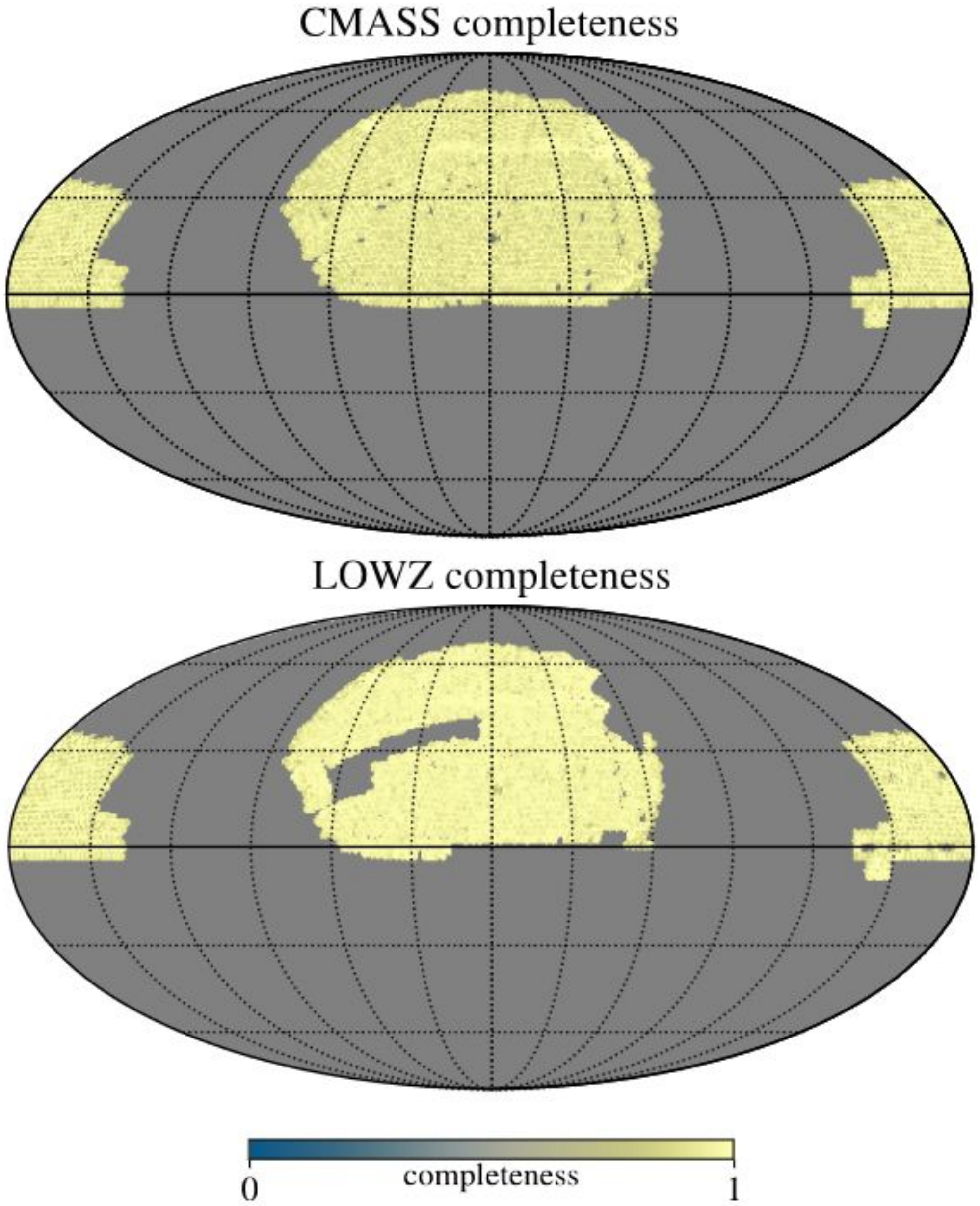
see later

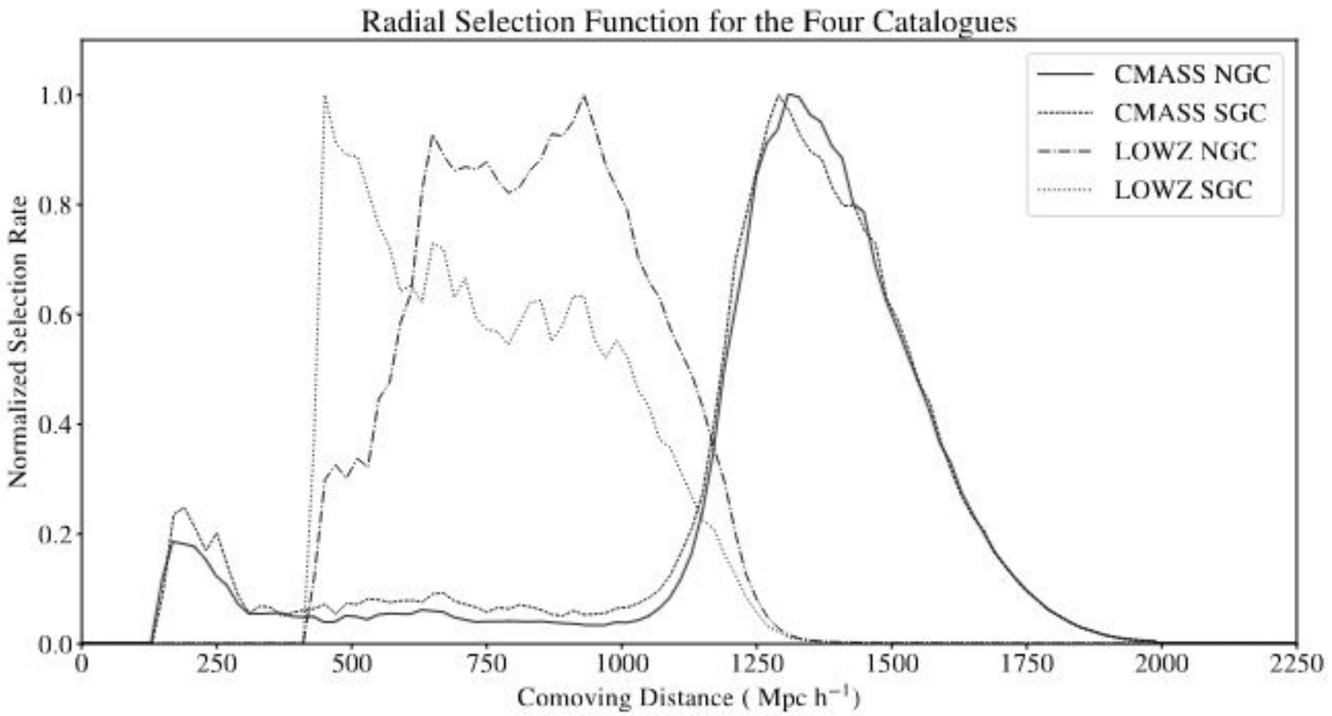
Full data model



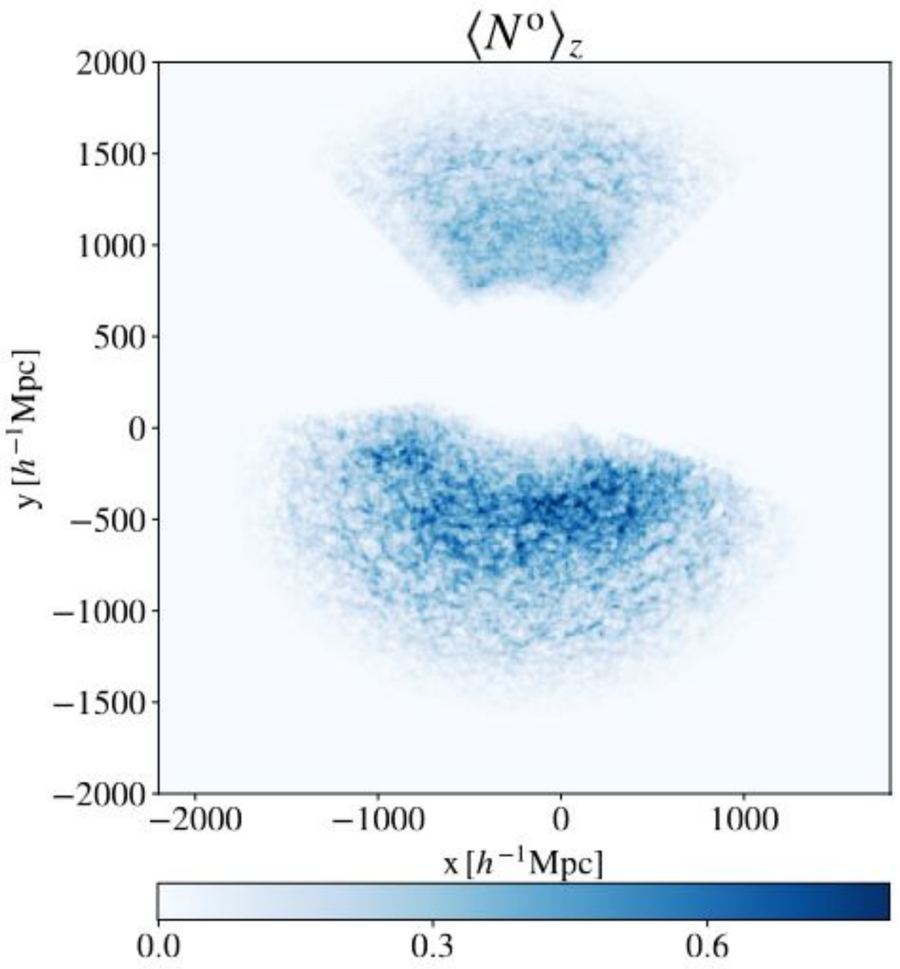
Andrews et al. (2023, 2024 in prep)

Test setup (mock data): visibility mask & radial selection

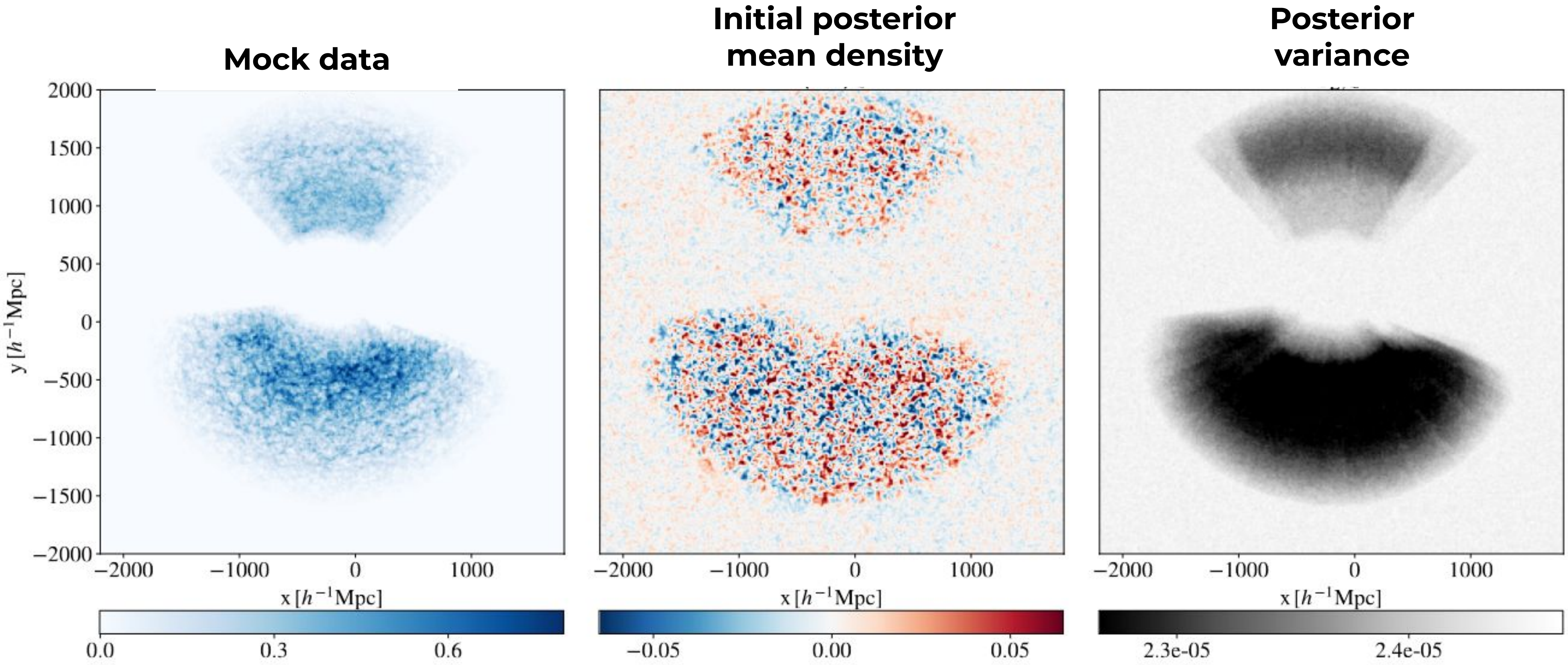




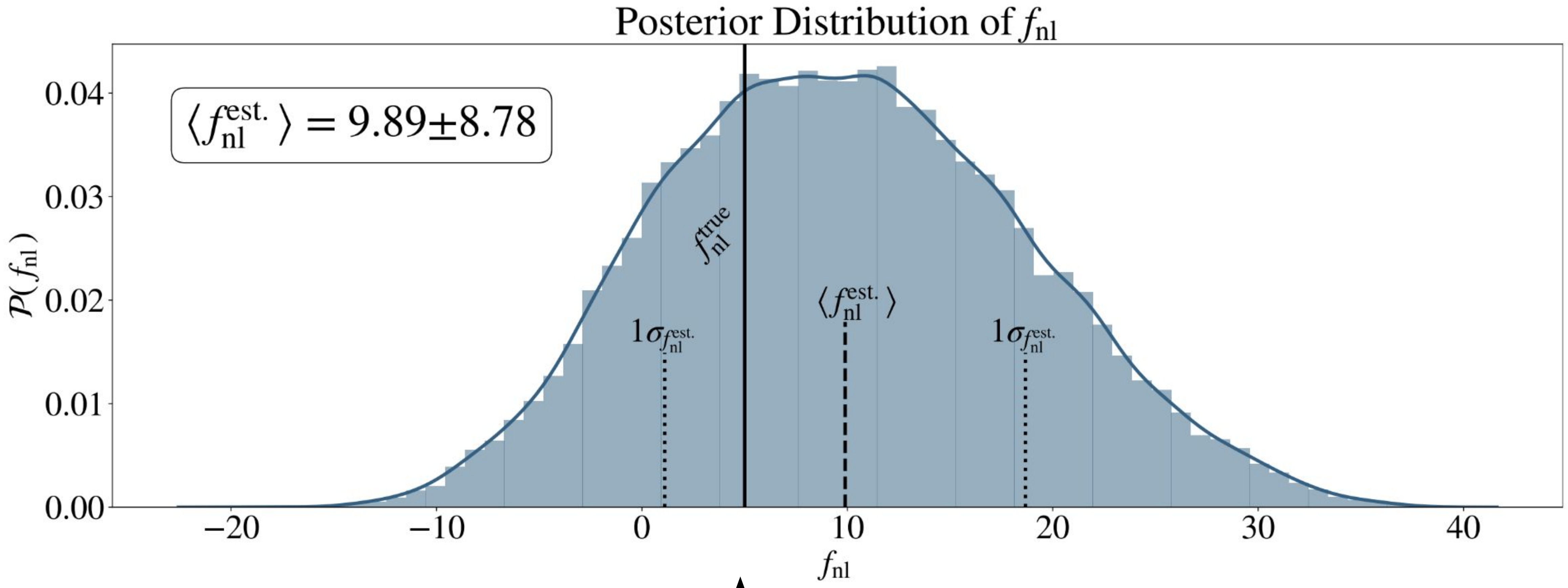
Self consistent mock data



Inferred maps with mock setup



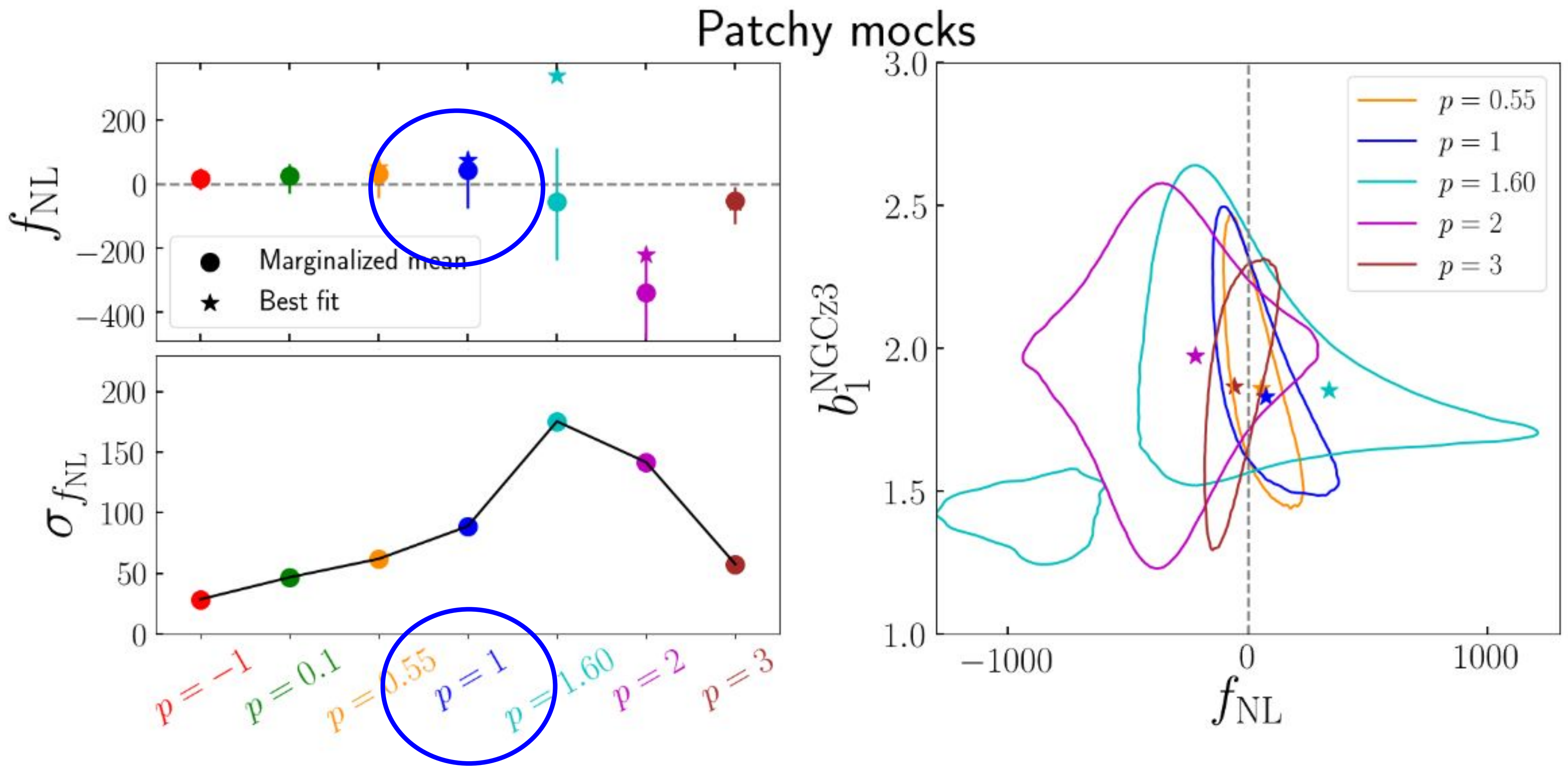
Inferred f_{NL} for reference run (mock-BOSS survey)



"True" f_{NL}

p=1 fixed, k_max = 0.2 h/Mpc

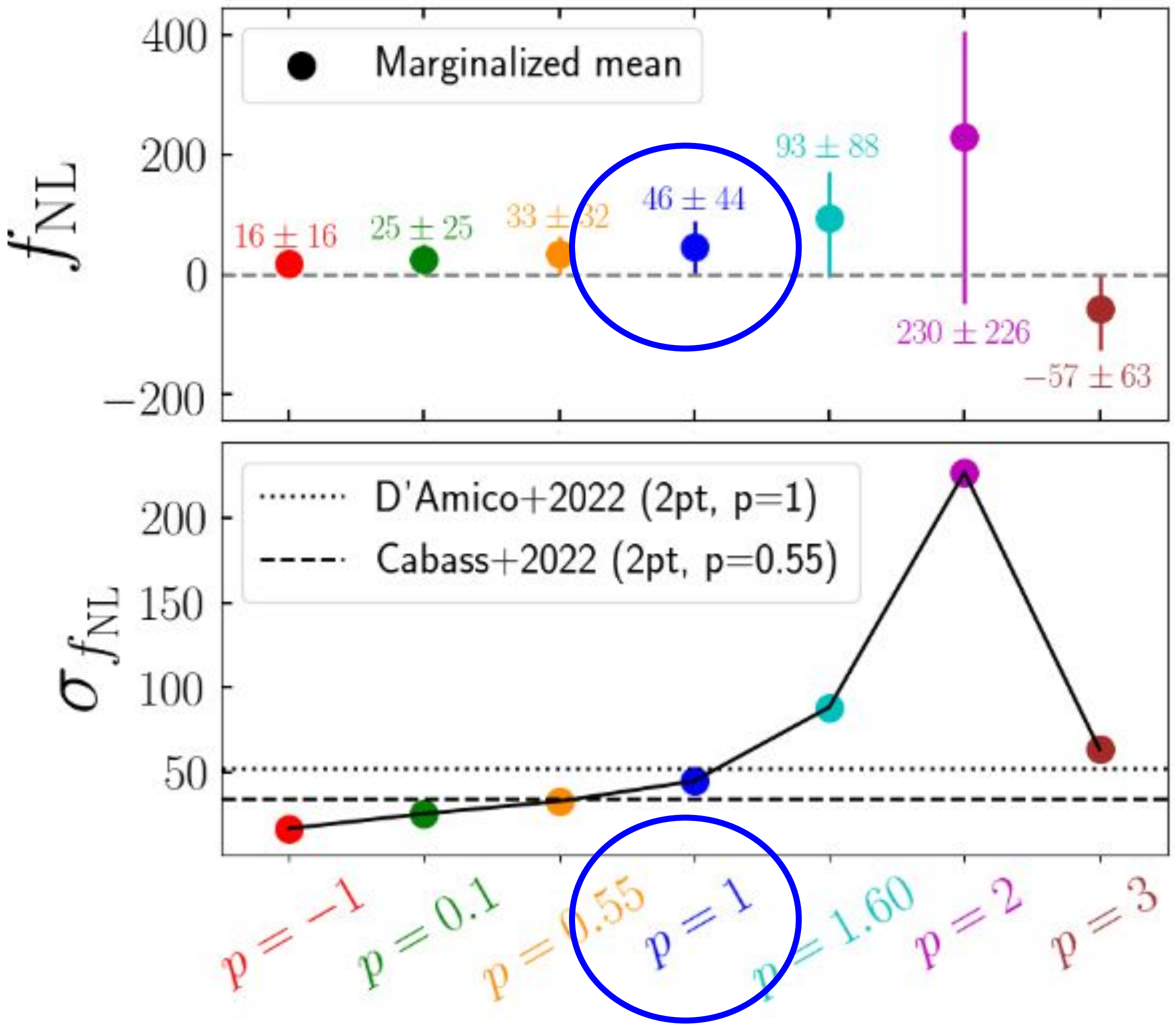
Reported forecast for BOSS and BOSS-like data (p=1)



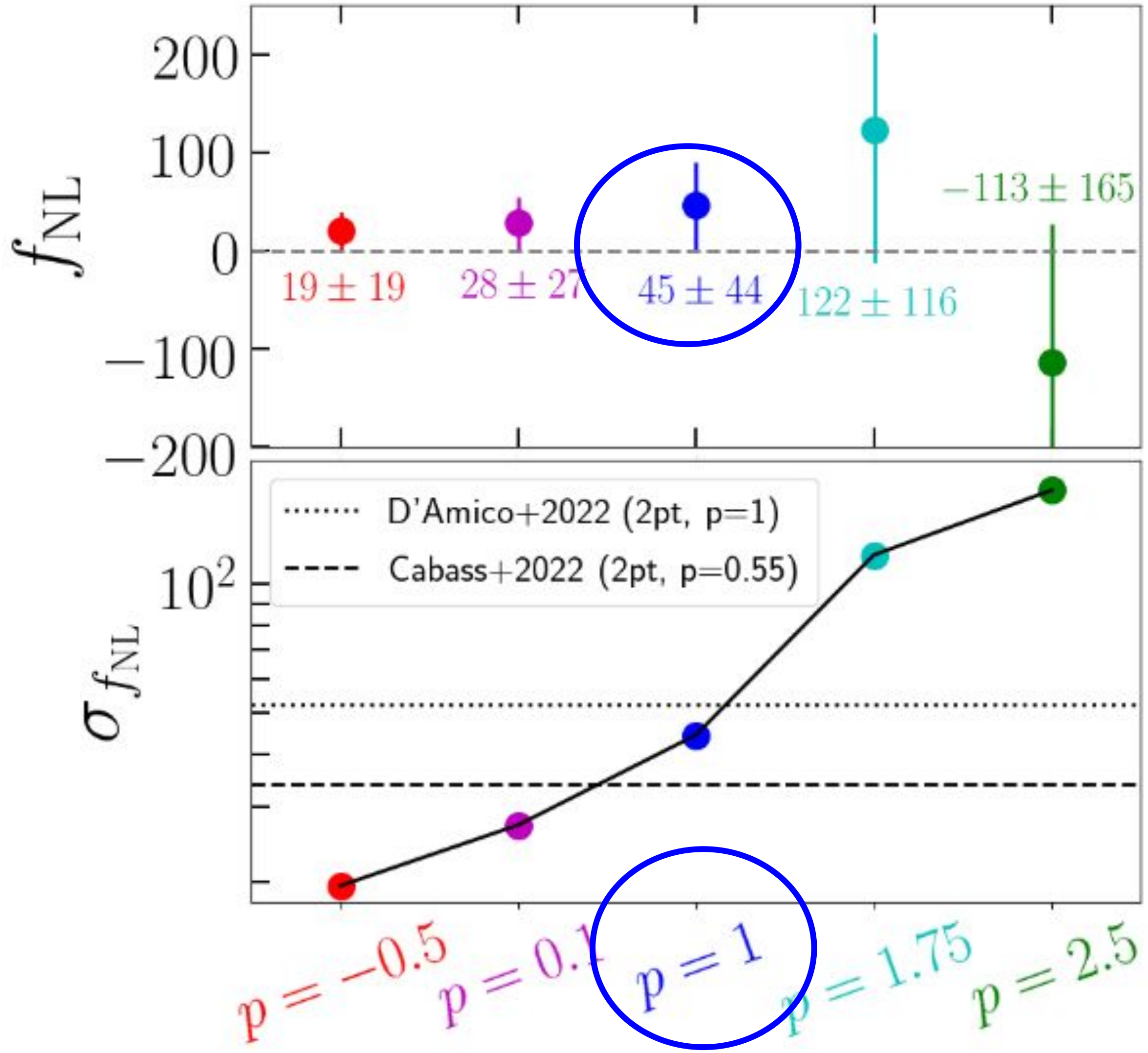
k_{max} = 0.05 h/Mpc

Lazeyras et al. (2023)

Reported constraints on f_{NL} in BOSS data (p=1)



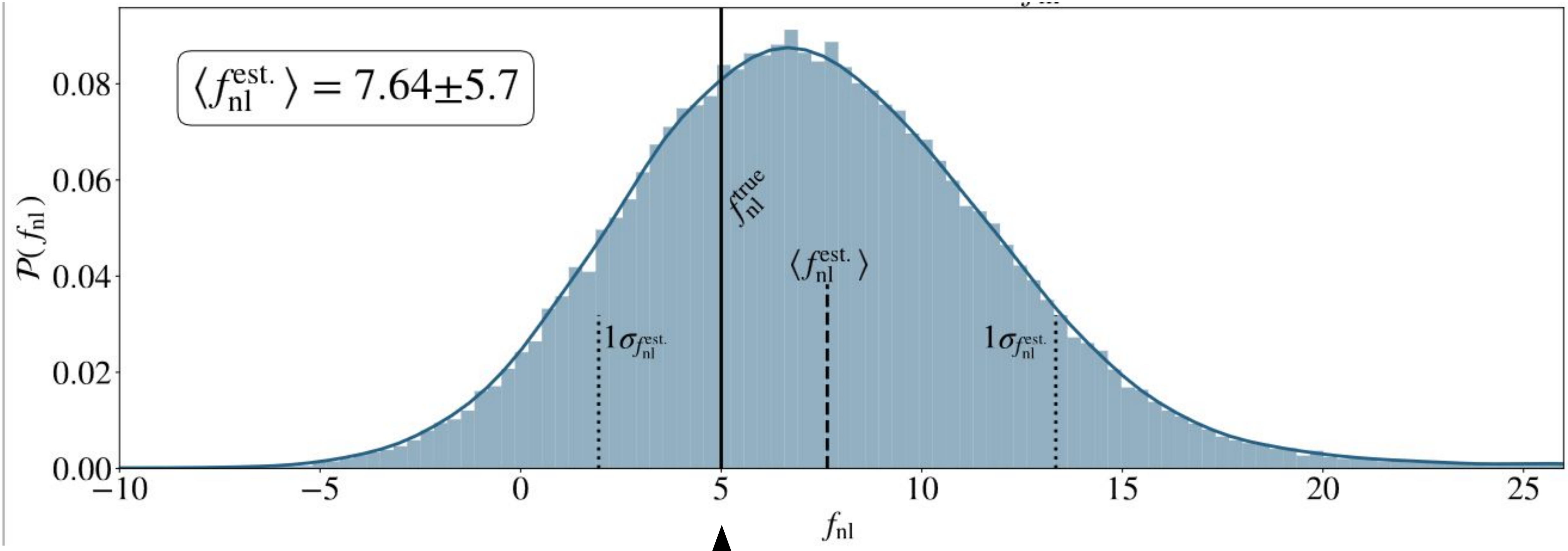
Barreira (2022)



Lazeyras et al. (2023)

k_{max} = 0.05 h/Mpc

Inferred f_{NL} for reference run (Stage-IV survey)

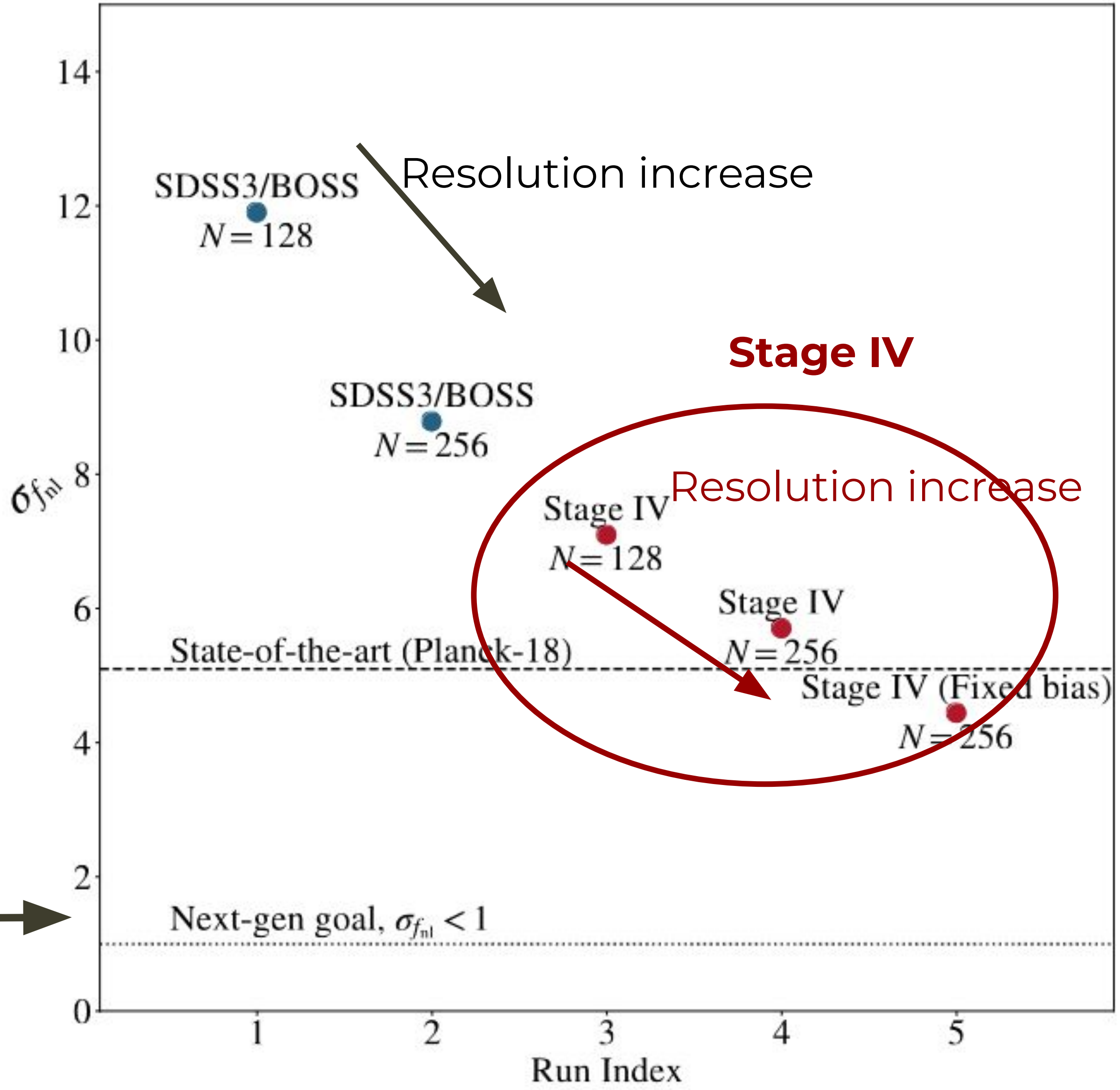


"True" f_{NL}

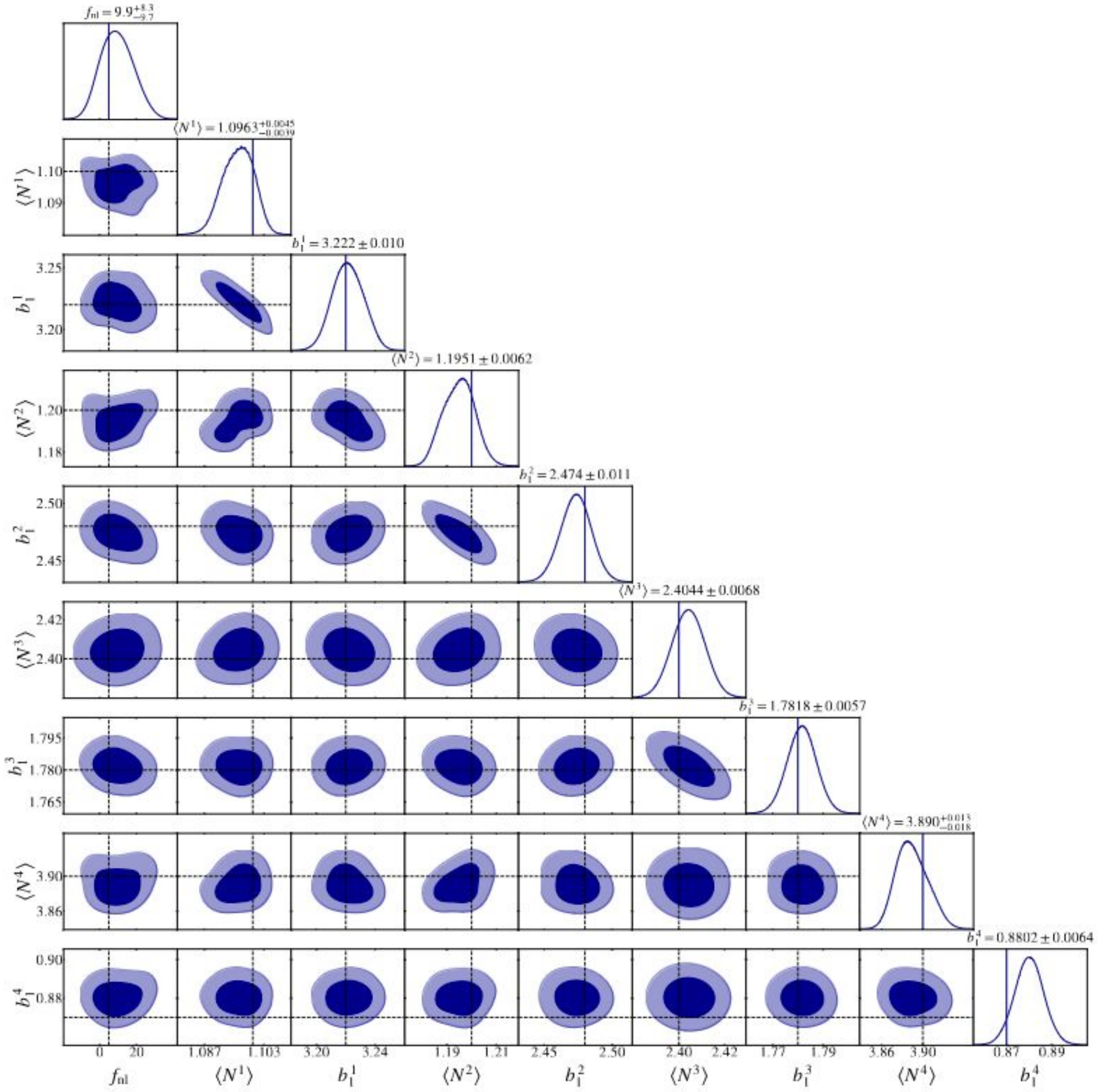
p=1 fixed, k_{max} = 0.2 h/Mpc

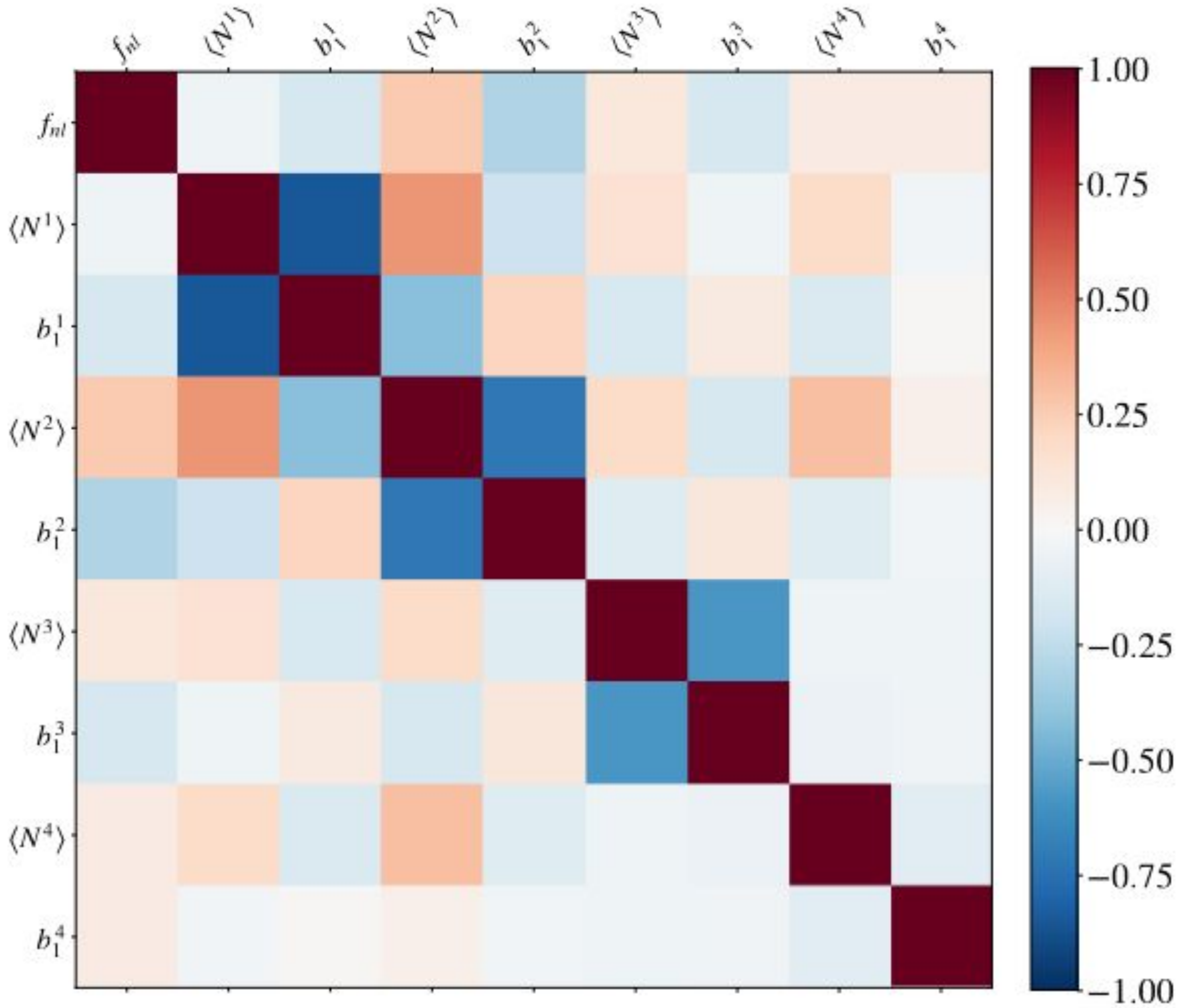
Forecast error to Stage-4 experiment

Goal $|f_{NL}|<1$

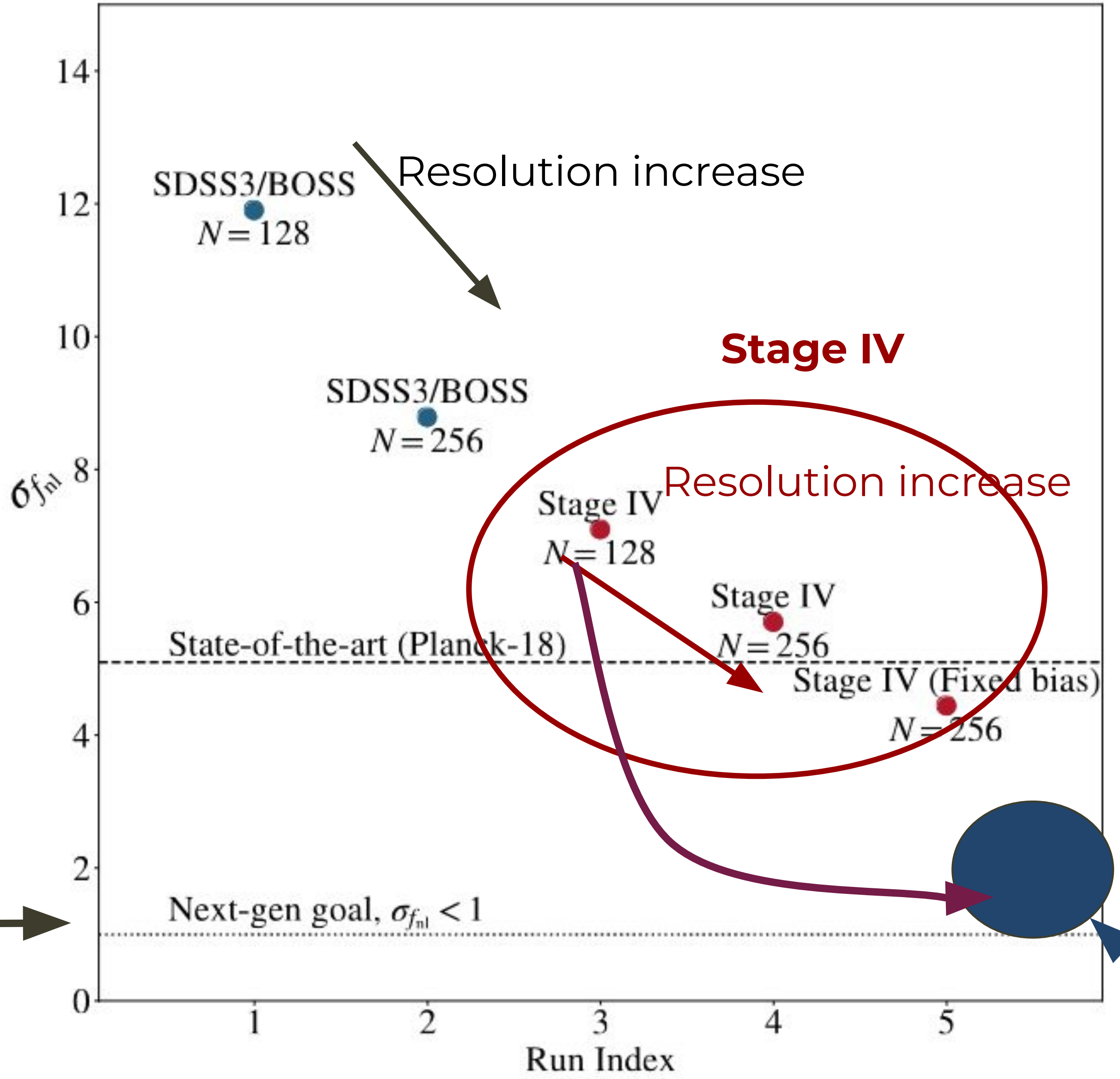


Relation to galaxy bias parameters





Forecast error for Euclid-like experiment



Goal $|f_{NL}|<1$

b_ϕ marginalized (prior p = 0.55 +/- 0.4, like Barreira 2022)

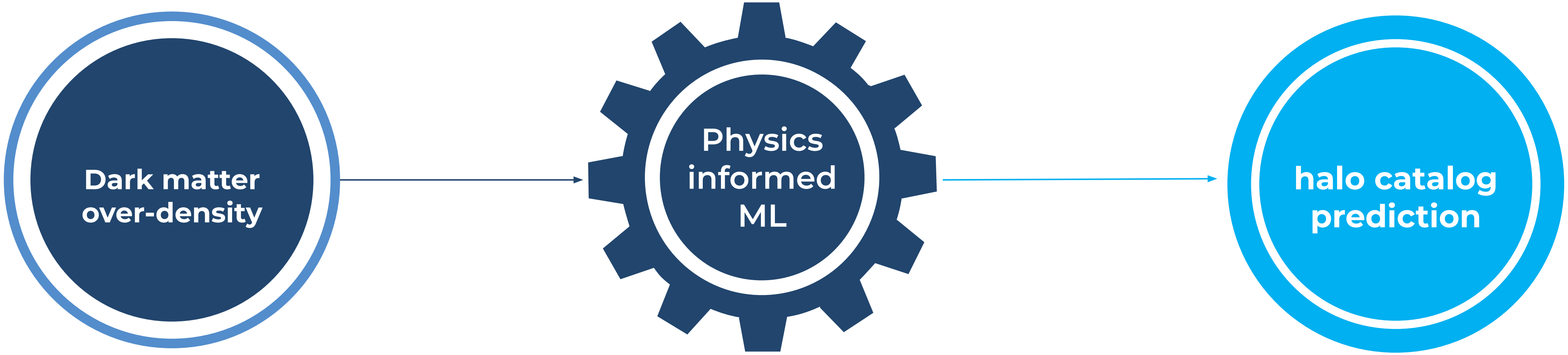
k_{max} = 0.05 - 0.1 h/Mpc + several forward model improvement w.r.t Andrews et al (2023)

Andrews et al. (2024 in prep)

2

Improving Galaxy/Dark matter connection

PineTree & CHARM (ex-NPE = Neural Physical Engine)
(S. Ding, S. Pandey, T. Charnock)



From approximate simulators
(e.g. 2LPT)

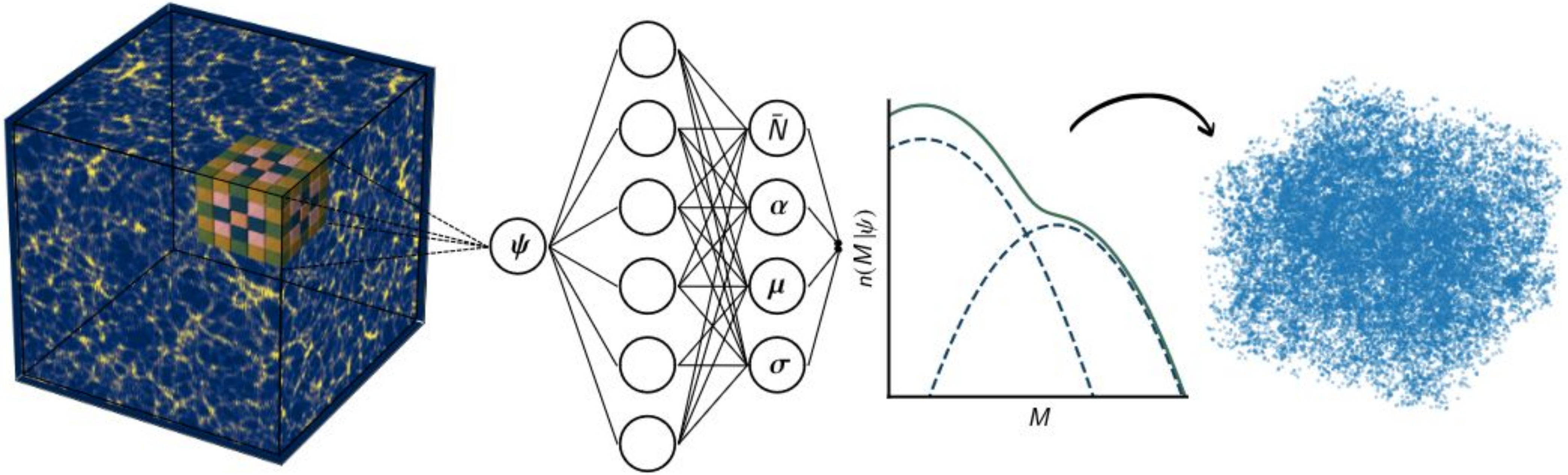
- Fast & Differentiable
- Stochastic
- Explainable
- 17-32 parameters

Validation:
- 1pt
- 2pt
- field-level

Ding et al (in prep), Pandey et al. (2024), Charnock et al 2020

PineTree: Physical and Interpretable NEtworks for TRacer Estimation/Emulation

Preliminary



overdensity field

PineTree forward model

halo catalogue

Ding et al. (in prep. 2024)

Training / validation procedure

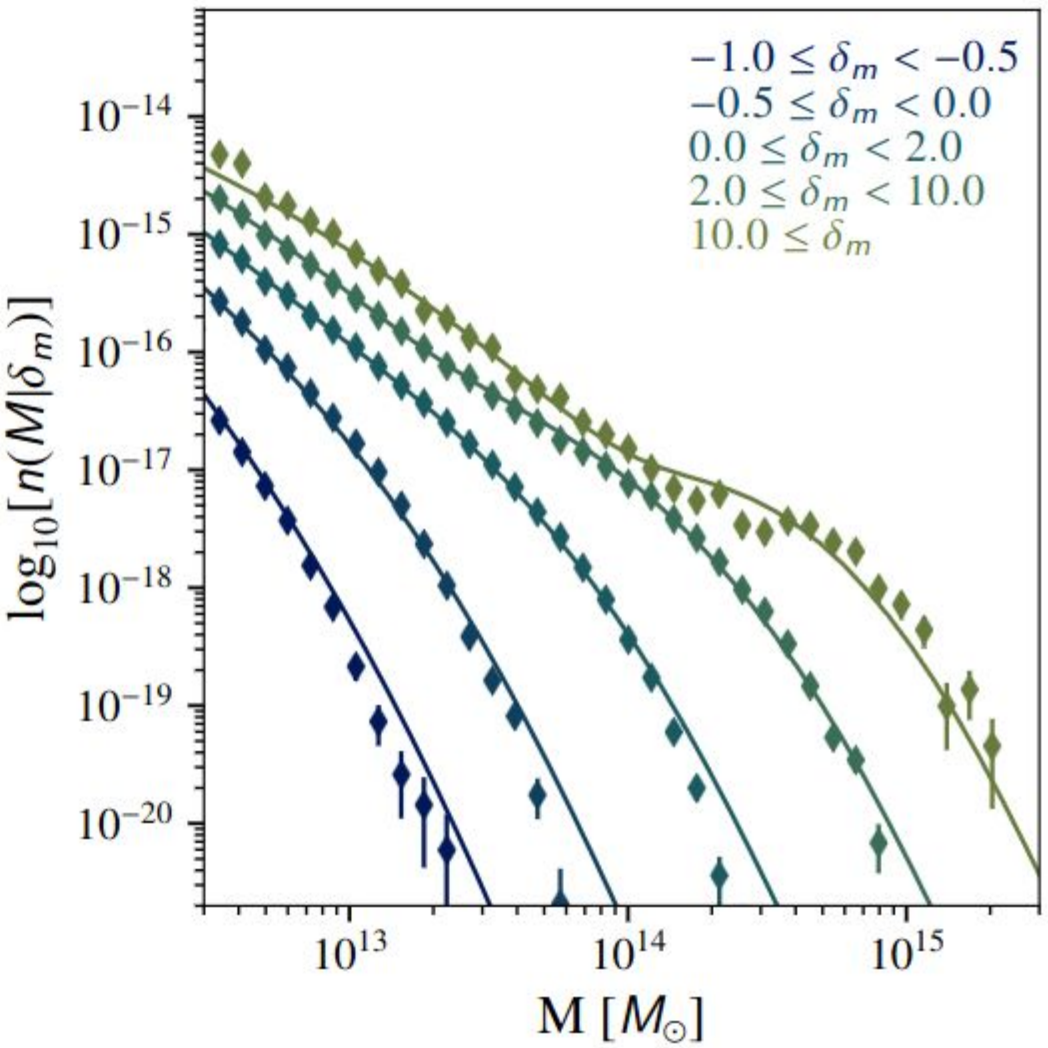
Preliminary

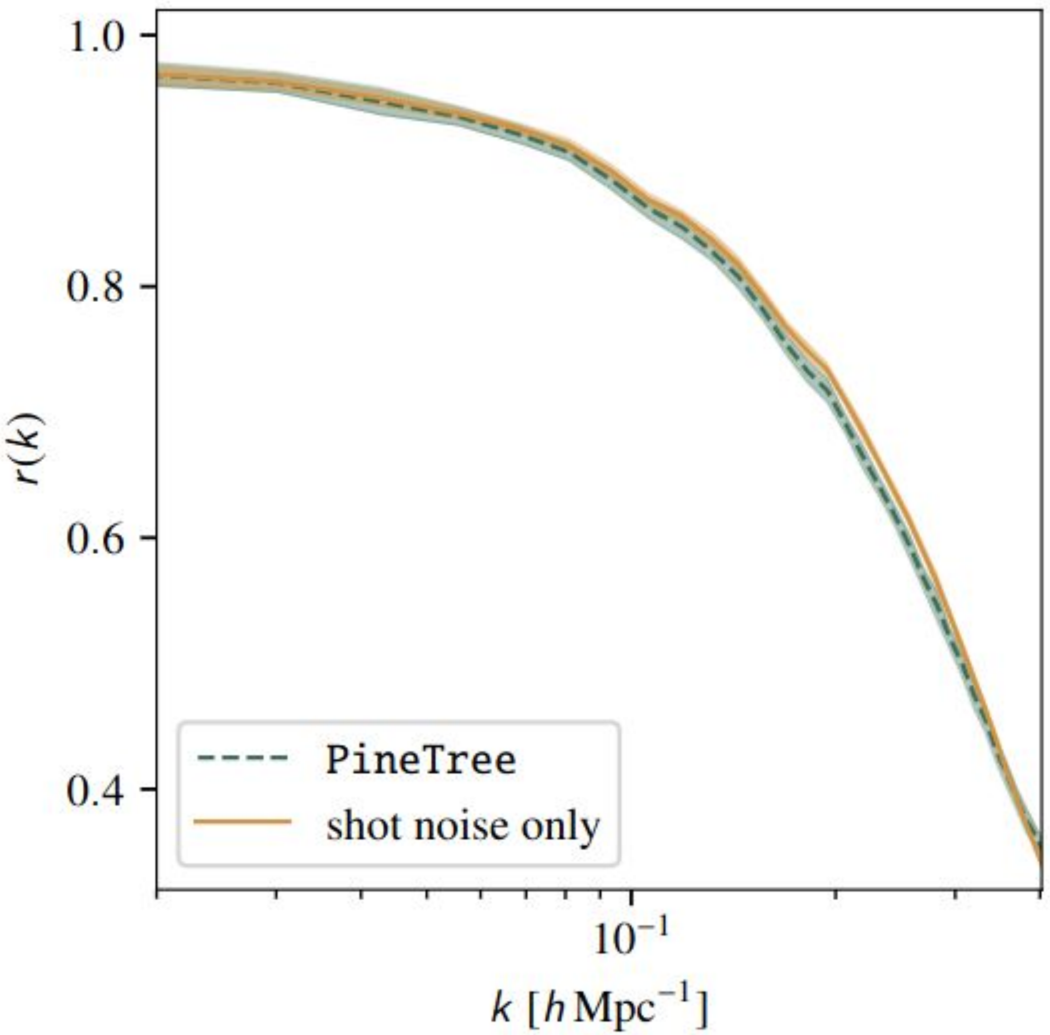
- Computed 40 N-body simulations
    - 500 Mpc/h, 512^3 particles
    - $m_p = 3 \times 10^{12}\ M_\odot$
- Training on:
    - baseline: one simulation
    - extended: 10 for training and 30 for validation
- Ideally: no training at all!

Ding et al. (in prep. 2024)

First look: mass function and halo field correlation

Preliminary

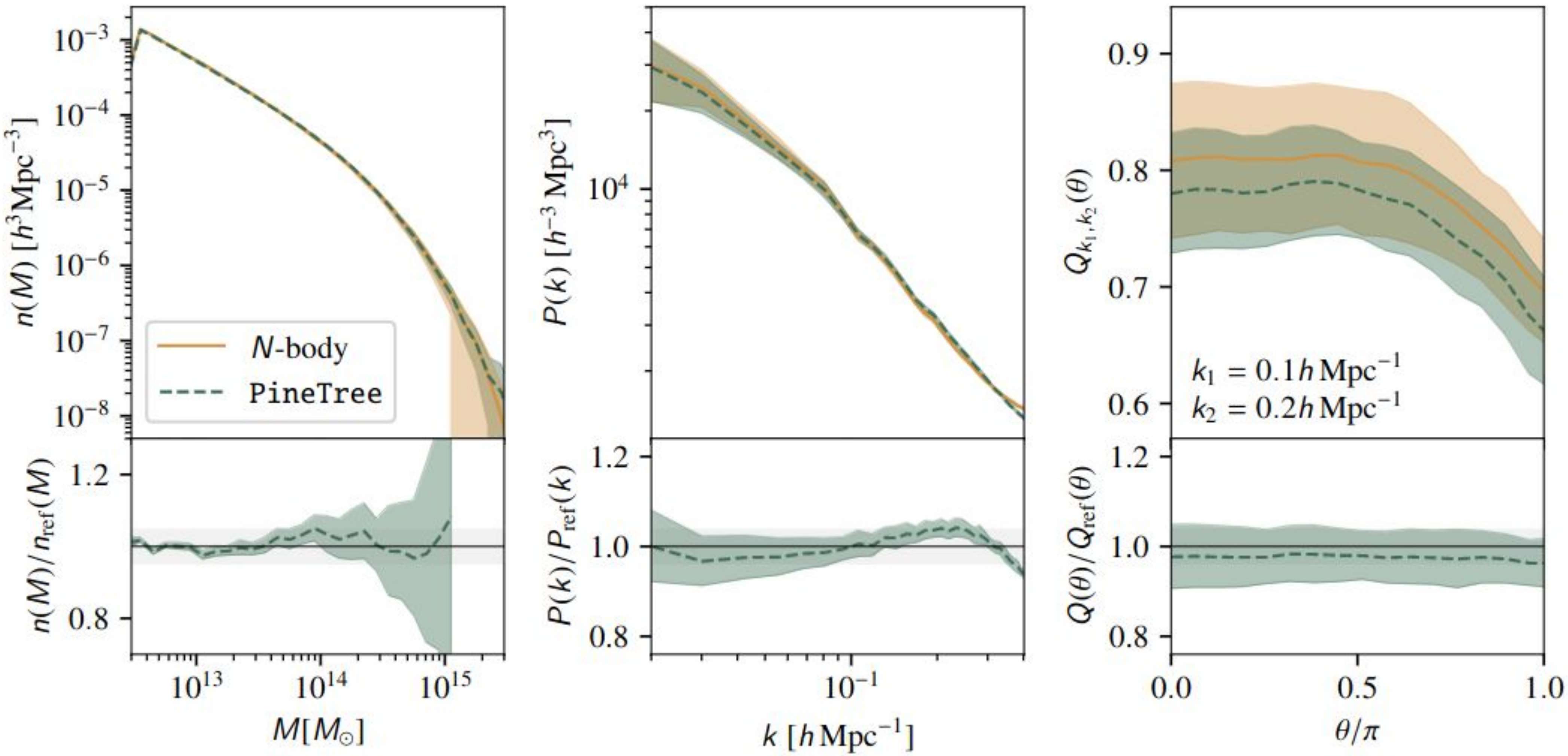




Ding et al. (in prep. 2024)

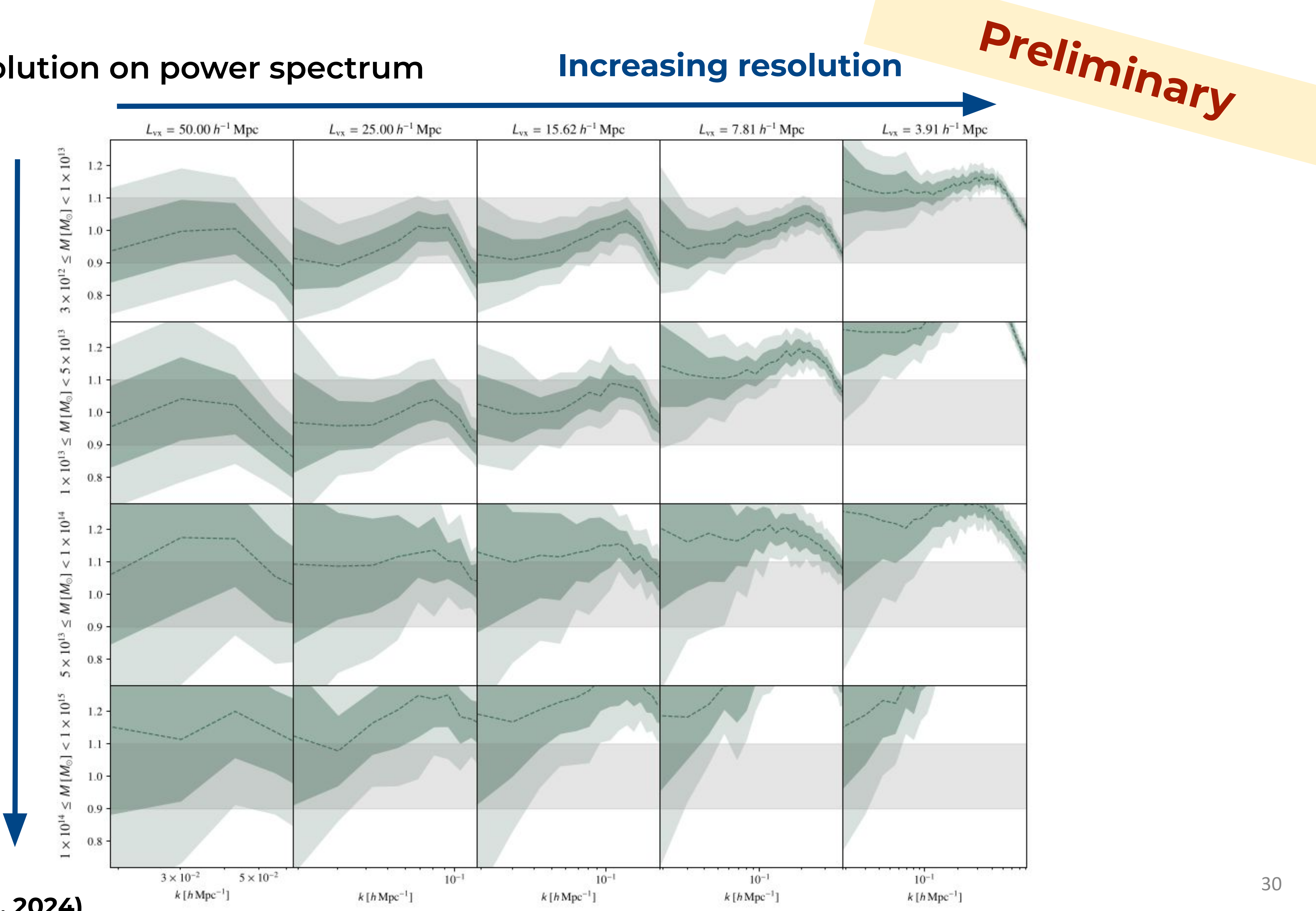
Second look: n(M), power spectra, bi-spectra

Preliminary



Ding et al. (in prep. 2024)

Effect of resolution on power spectrum



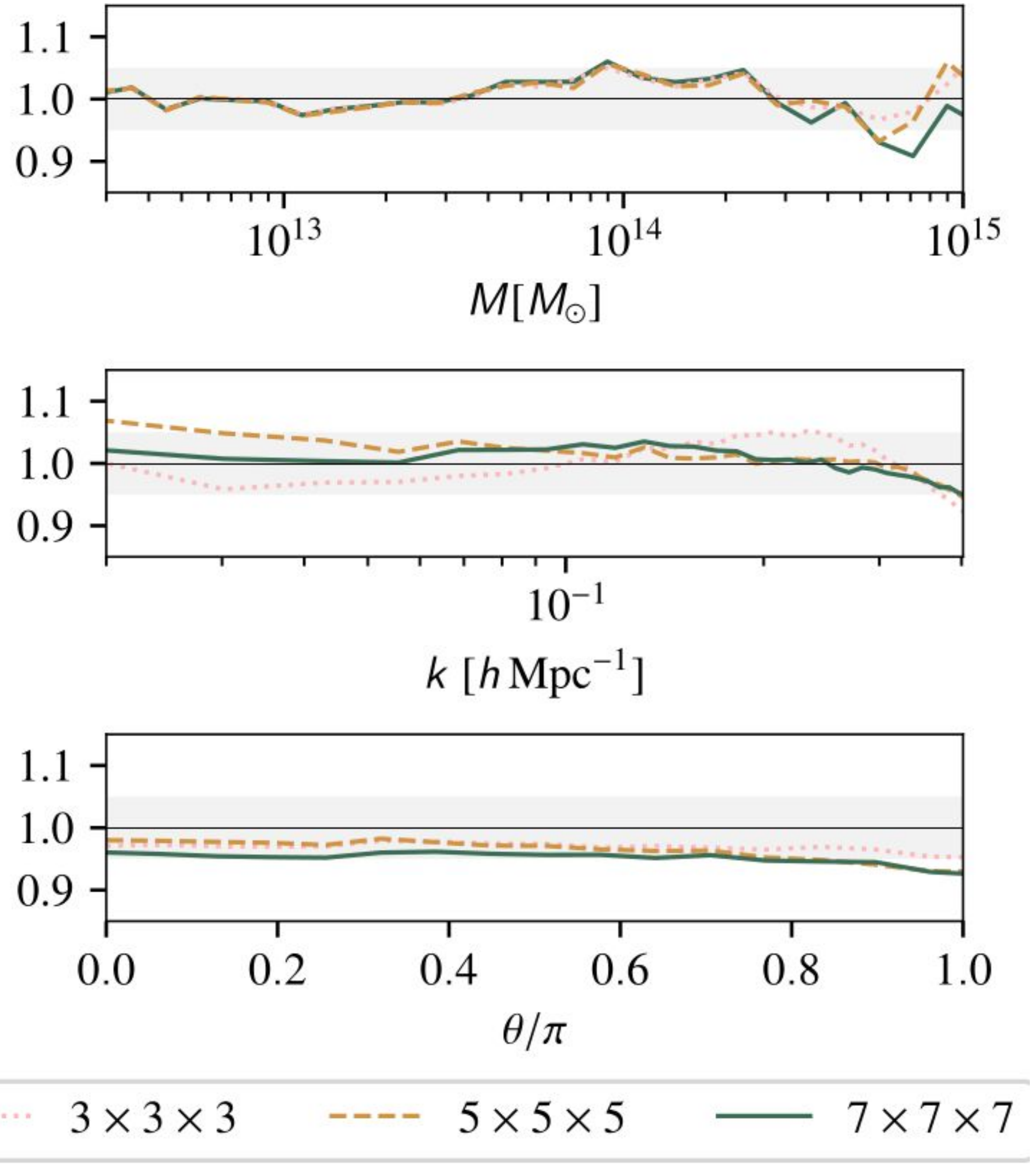
Ding et al. (in prep. 2024)

Impact of kernel size on summaries / Information content

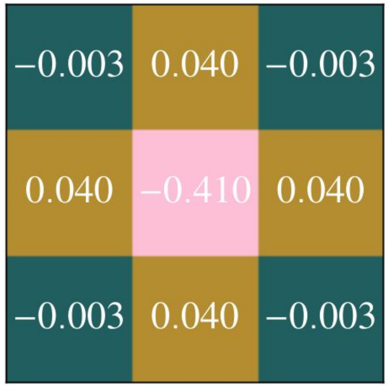
Relative mass function

Relative power-spectrum

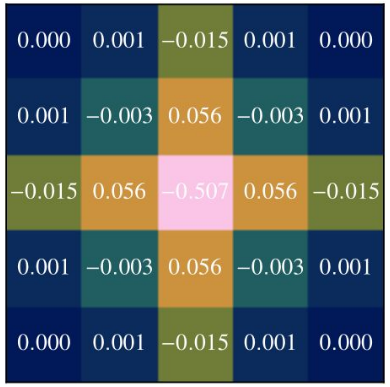
Relative bi-spectrum



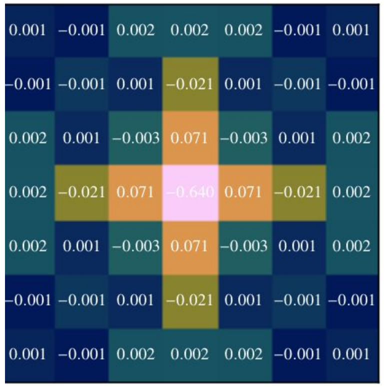
Kernel 3x3x3



Kernel 5x5x5



Kernel 7x7x7



3

Summary & Outlook

Field level inference for primordial non-gaussianities

The road so far:

- BORG: an automated machine for cosmology
- Models:
    - Different dynamical models
    - Galaxy scale dependent bias
    - PineTree (demo-ed for mock halo catalog generation)
- Improvements through unlocking more modes
- Immediately achievable constraints with Stage IV: $|f_{NL}| < 5$

Field level inference for primordial non-gaussianities

The road so far:

- BORG: an automated machine for cosmology
- Models:
    - Different dynamical models
    - Galaxy scale dependent bias
    - PineTree (demo-ed for mock halo catalog generation)
- Improvements through unlocking more modes
- Immediately achievable constraints with Stage IV: $|f_{NL}| < 5$

In preparation:

- Detailed Euclid forecasts
- Test of inference on Quijote halo mocks
- PineTree

Field level inference for primordial non-gaussianities

The road so far:

- BORG: an automated machine for cosmology
- Models:
  - Different dynamical models
  - Galaxy scale dependent bias
  - PineTree (demo-ed for mock halo catalog generation)
- Improvements through unlocking more modes
- Immediately achievable constraints with Stage IV: $|f_{NL}| < 5$

In preparation:

- Detailed Euclid forecasts
- Test of inference on Quijote halo mocks
- PineTree

Future:

- Inference on large galaxy catalogs (notably Euclid)