Additional probes and alternative techniques for galaxy clustering

(Galaxy Clustering: Additional Probes work package)

Symposium Euclid France 2023

Florent Leclercq

www.florent-leclercq.eu

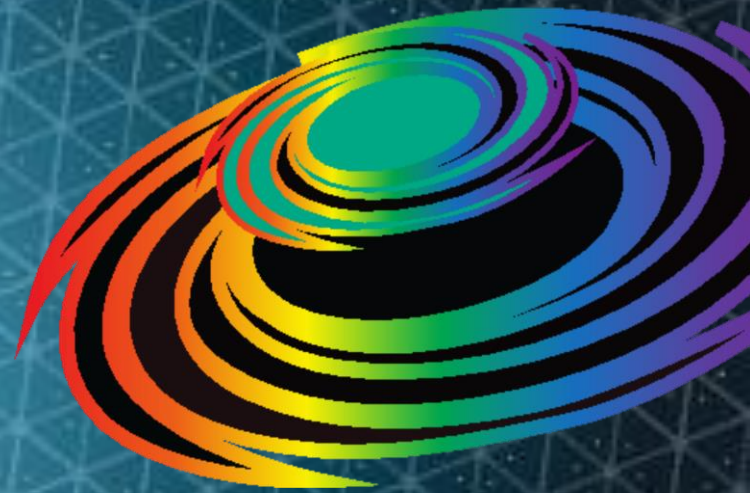
Institut d'Astrophysique de Paris
CNRS & Sorbonne Université

In collaboration with:
Adam Andrews (INAF Bologna), Deaglan Bartlett (IAP), Alan Heavens (Imperial College), Tristan Hoellinger (IAP), Jens Jasche (Stockholm U.), Guilhem Lavaux (IAP), James Prideaux-Ghee (Imperial College), Eleni Tsaprazi (Imperial College)

and the Aquila Consortium

www.aquila-consortium.org

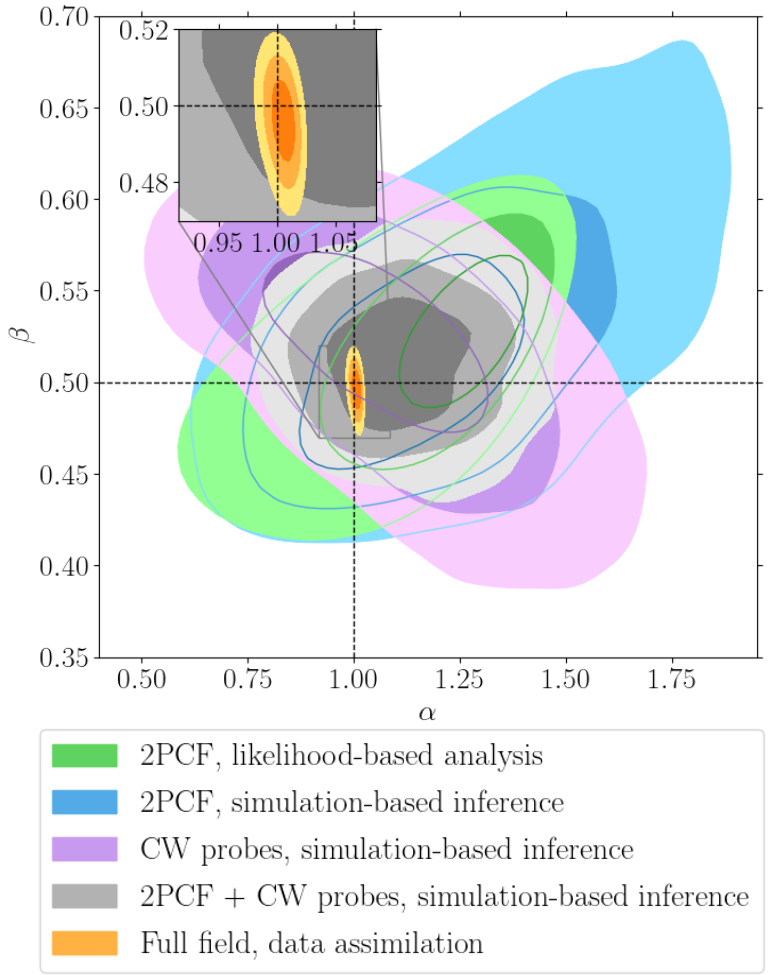
25 January 2024



Going beyond two-point correlations for galaxy clustering: the "implicit" and "explicit" approaches

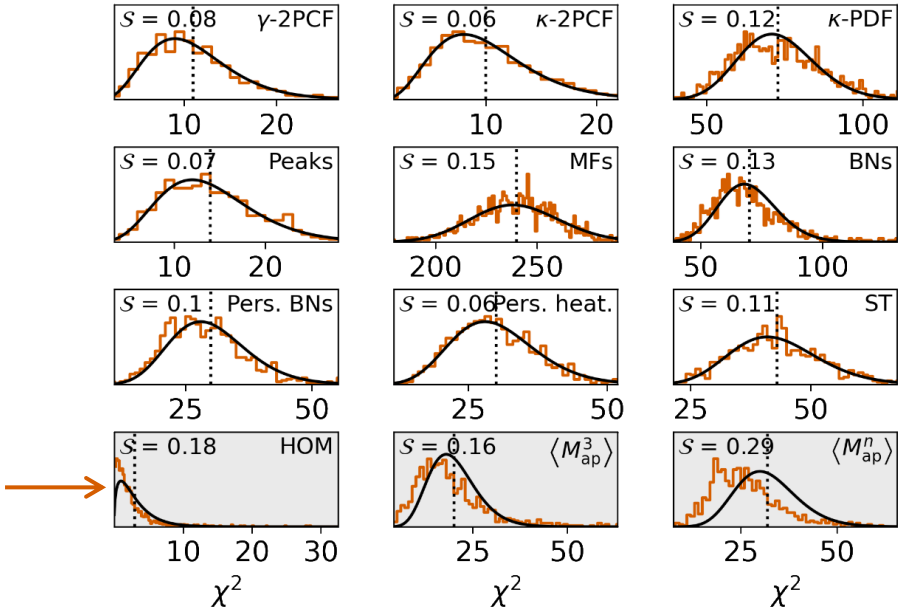
Note:
- likelihood-free inference (LFI) ≈ simulation-based inference (SBI) ≈ implicit likelihood inference (ILI)
- likelihood-based approach = explicit likelihood inference

- A question of accuracy: first, avoid biases.



FL & Heavens, 2103.04158

- Some weak lensing additional probes also have a non-Gaussian distribution.



- A question of precision: can numerical forward models be used to push further than $k \gtrsim 0.15\ h/\mathrm{Mpc}$? The full field contains much more information.

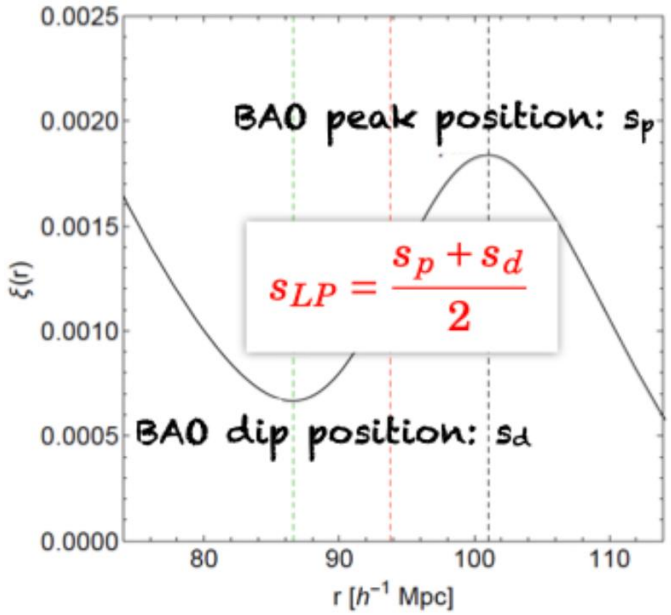
Euclid HOWLS-KP paper 1, Ajani *et al.*, 2301.12890

Additional probes: non-standard and informative summary statistics of the field

- Some examples of additional probes that we're looking at:

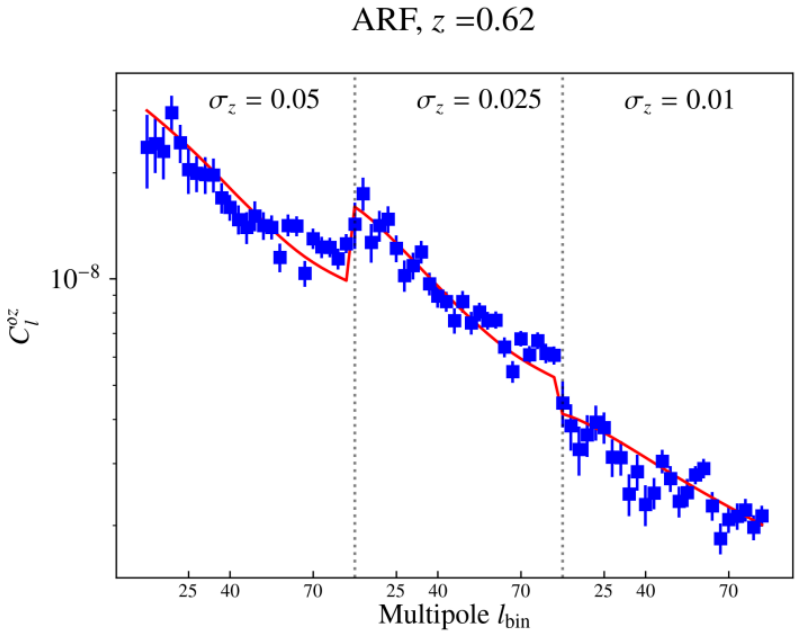
  Euclid definition: anything that is not (higher-order) correlation functions or voids

Linear point BAO



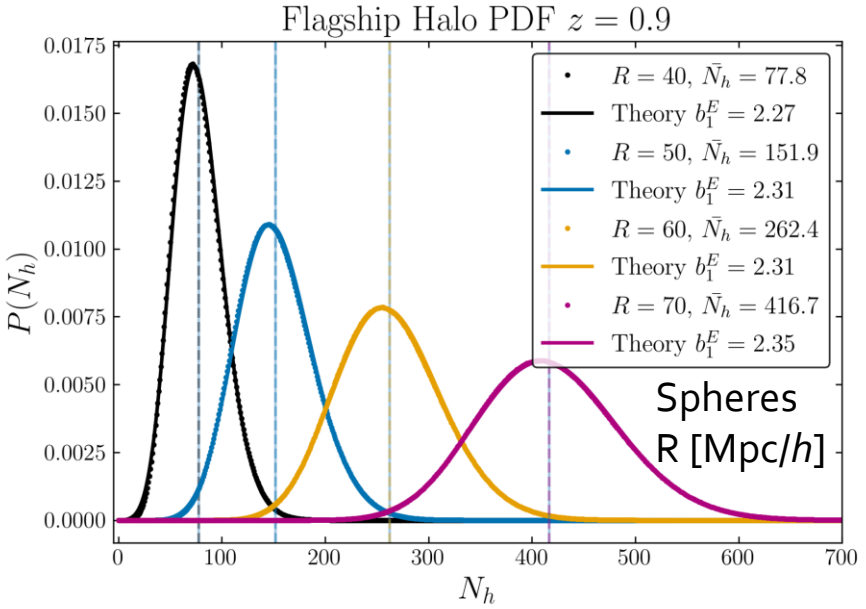
Anselmi, Oppizzi, Corasaniti, *et al.*

Angular redshift fluctuations



Hernández–Monteagudo, *et al.*

One-point statistics



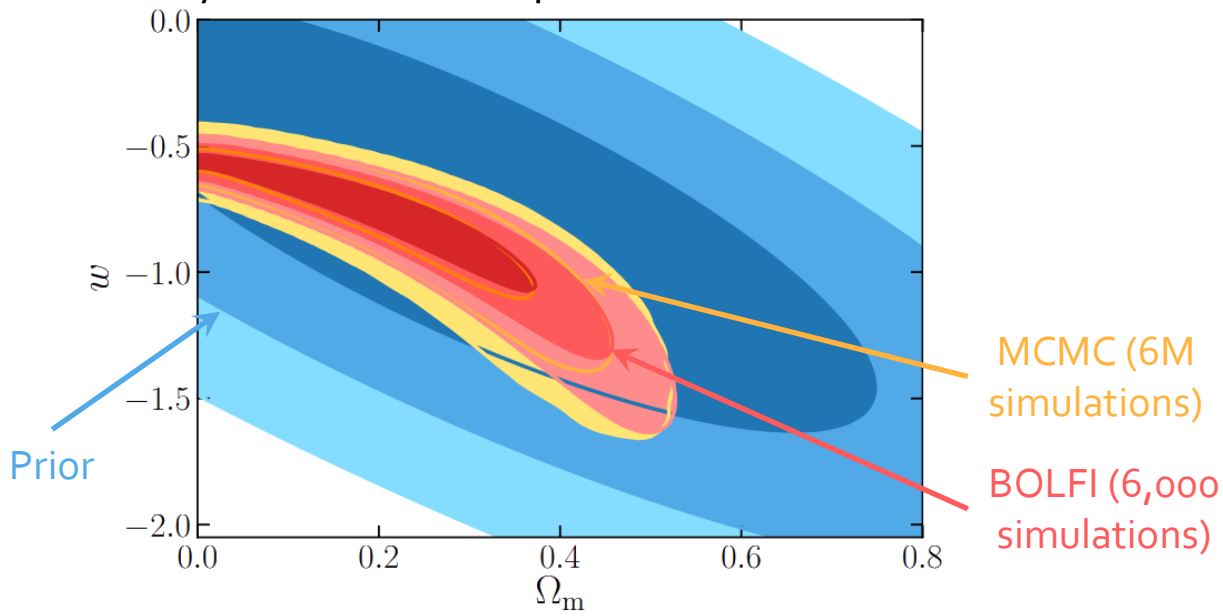
Uhlemann, Castiblanco Tolosa, Homer, Codis, *et al.*

- Typically, we look at the distribution of these probes in mock catalogues and fit a standard pdf.
- But all of these can also be used in an "implicit likelihood inference" approach…

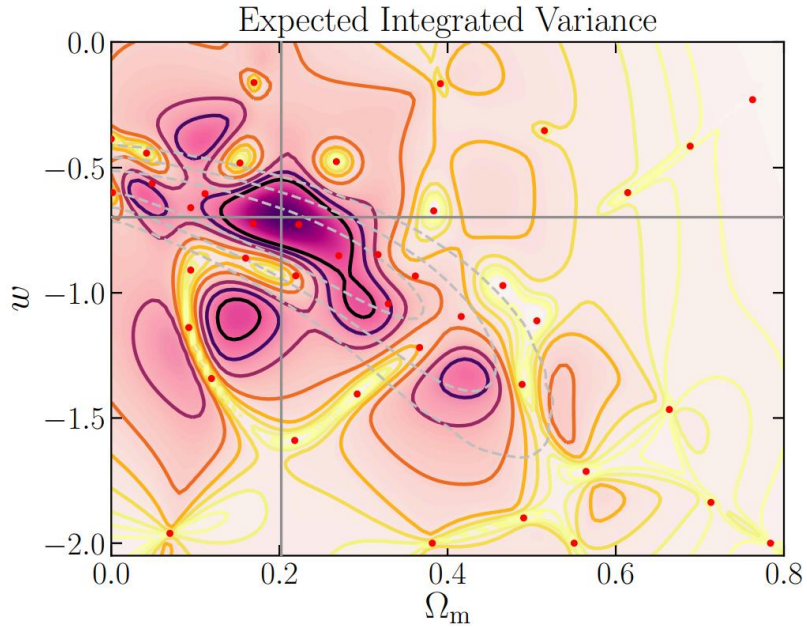
Dealing with expensive simulators in implicit likelihood inference (ILI) problems: The BOLFI algorithm (*Bayesian Optimisation for Likelihood-Free Inference*)

- The simulator will typically be extremely expensive (*N*-body simulation, halo finding, complex observational effects). We can typically afford O(10,000) evaluations.
- Emulation of the data model is not the only option.

Re-analysis of the JLA supernovae data:



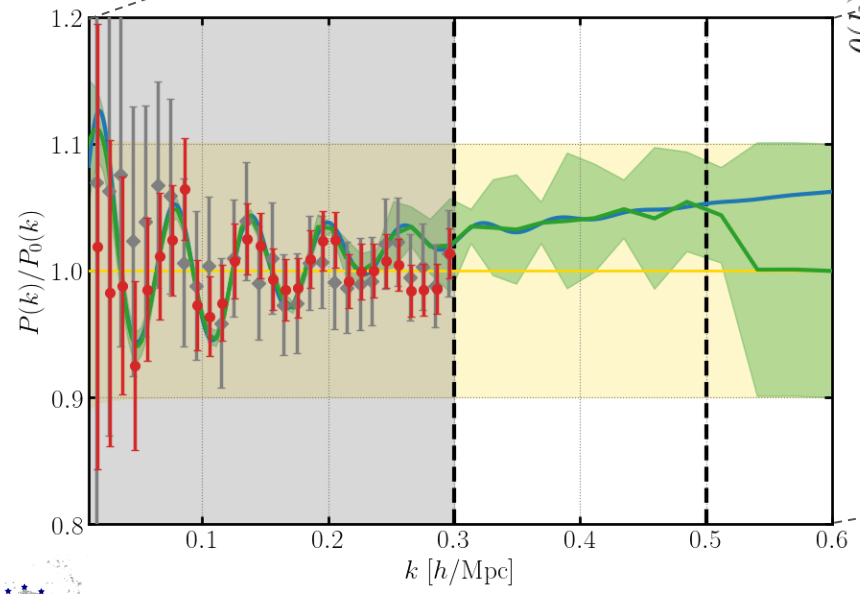
- BOLFI (*Bayesian Optimisation for Likelihood-Free Inference*) uses an acquisition function to place expensive simulations in the parameter space.
- The optimal acquisition function for implicit inference can be derived: the Expected Integrated Variance.

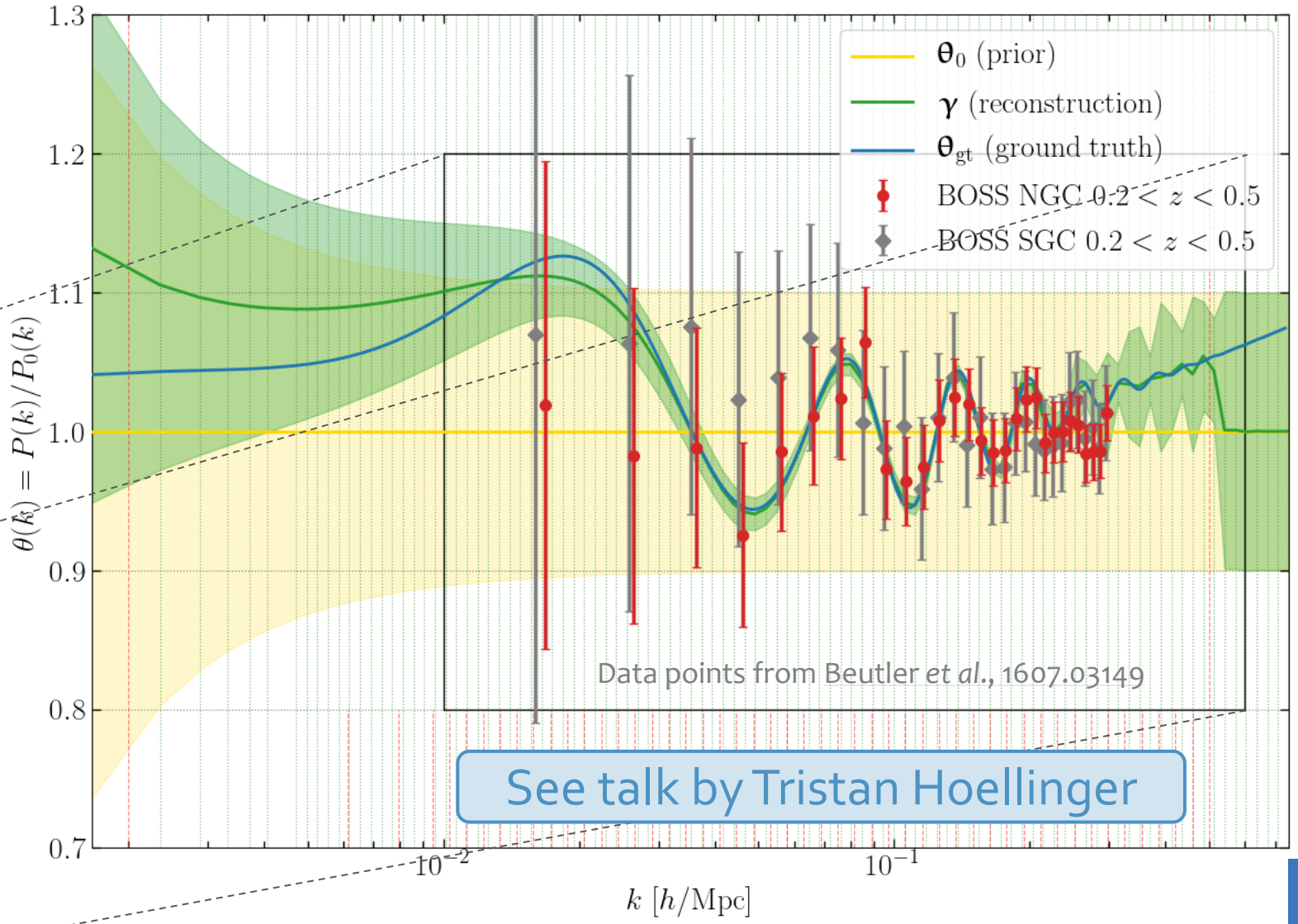


FL, 1805.07152

SELFI (*Simulator Expansion for Likelihood-Free Inference*): ILI of the initial power spectrum
Euclid forecast vs BOSS data

- Numerical data models allow using the galaxy power spectrum as summary statistics up to at least $k \gtrsim 0.5\ h/\mathrm{Mpc}$ safely
- $N_{\mathrm{modes}} \propto k^3$: 5 times more modes are used in the analysis.



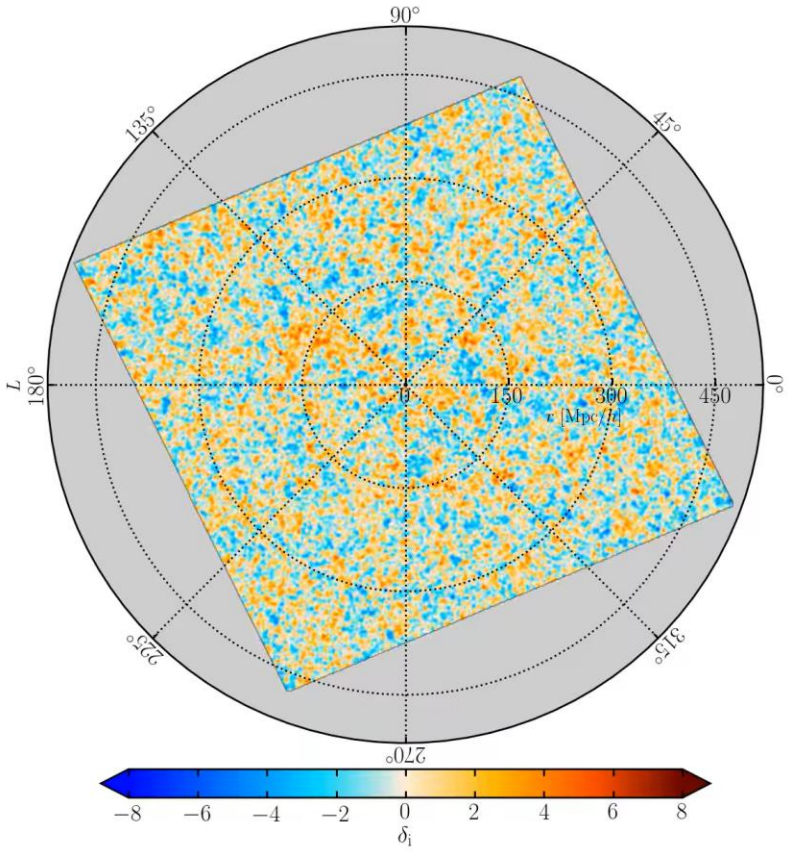


See talk by Tristan Hoellinger

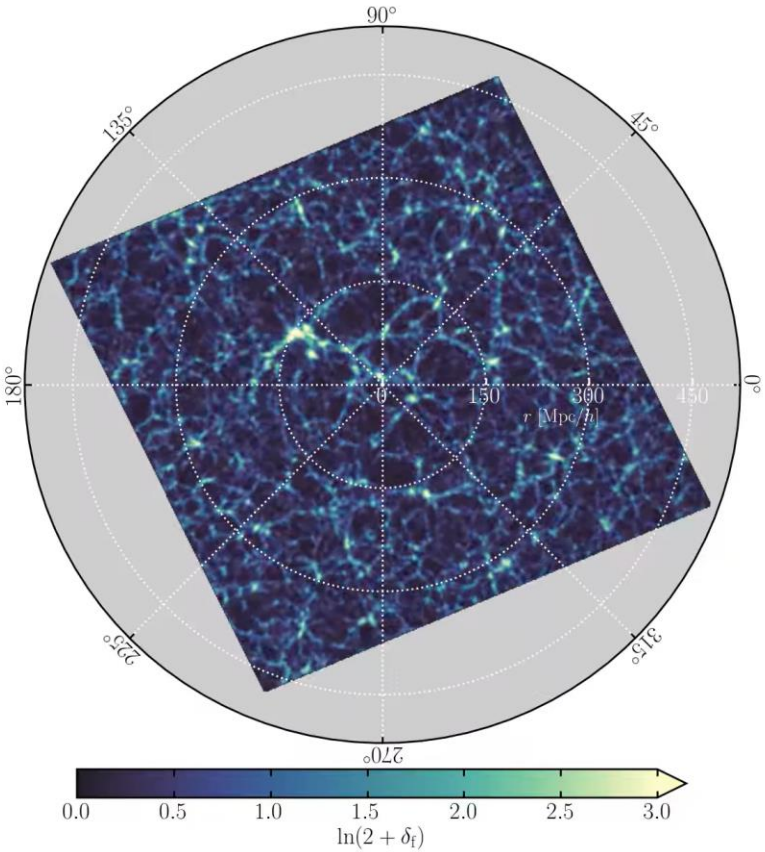
FL *et al.*, 1902.10149, FL, 2209.11057, Hoellinger & Leclercq, in prep.

Inference with an explicit field-level likelihood: The BORG algorithm (*Bayesian Origin Reconstruction from Galaxies*)

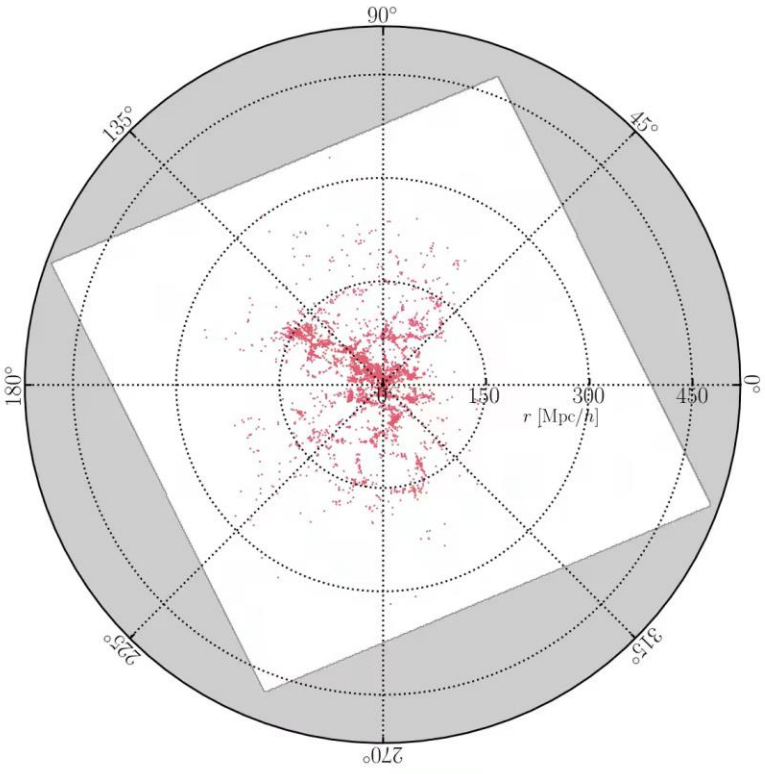
Initial conditions



Final conditions



Observations

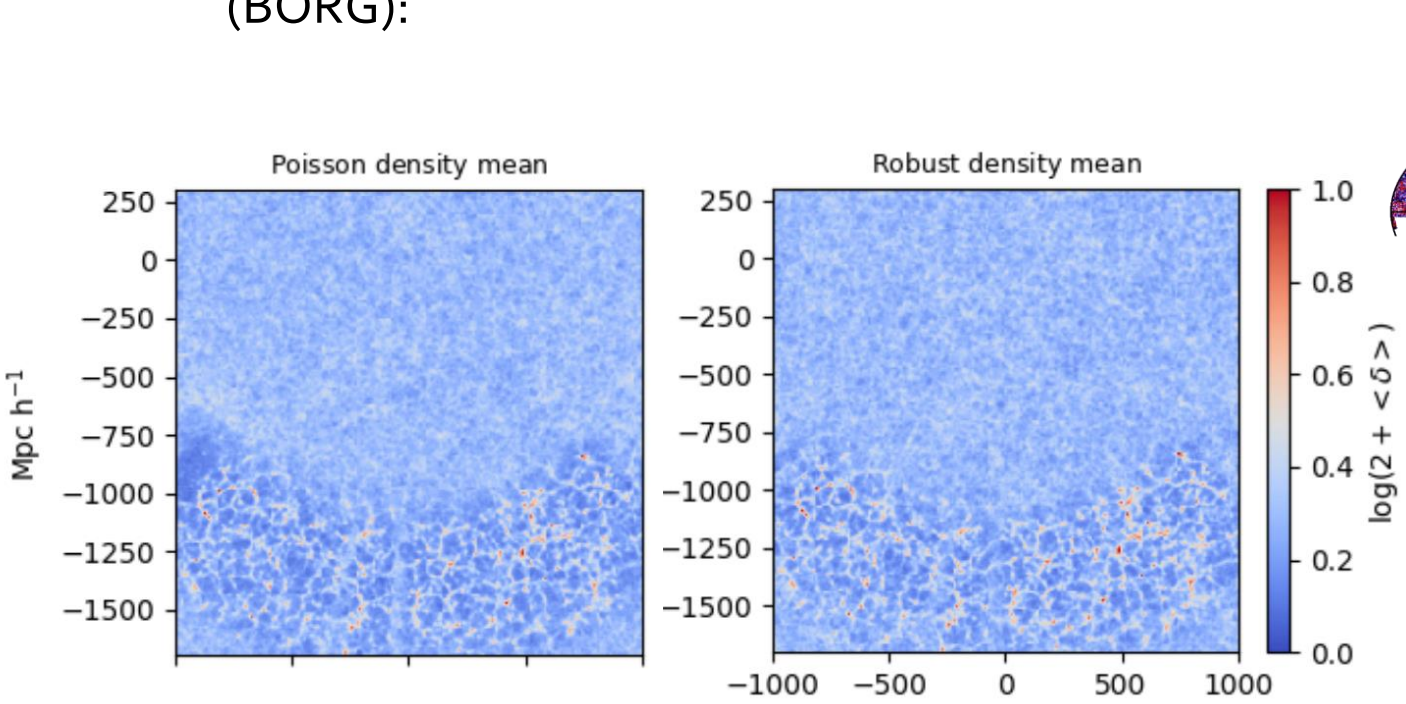


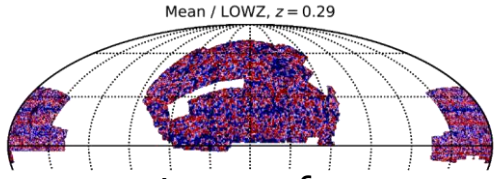
67,224 galaxies, ≈ 17 million parameters, 5 TB of primary data products, 10,000 samples, ≈ 500,000 forward and adjoint gradient data model evaluations, 1.5 million CPU-hours

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

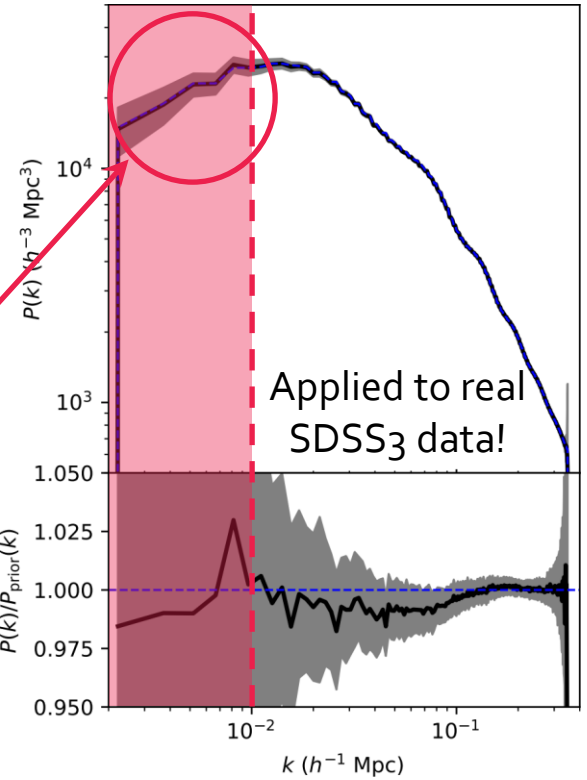
Model misspecification and unknown systematics with an explicit field-level likelihood

- Systematic effects are an issue of model misspecification: when the model differs from the actual data-generating process, posteriors tend to be biased and/or overly concentrated.
- In cosmology, we are sometimes unable to formulate ***any*** model that fits the data in some regimes.
- Machine-aided report of unknown systematic effects is possible with an explicit field-level likelihood (BORG):





A map of an unknown foreground



Applied to real SDSS3 data!

No apparent contamination, even well beyond the turn-over

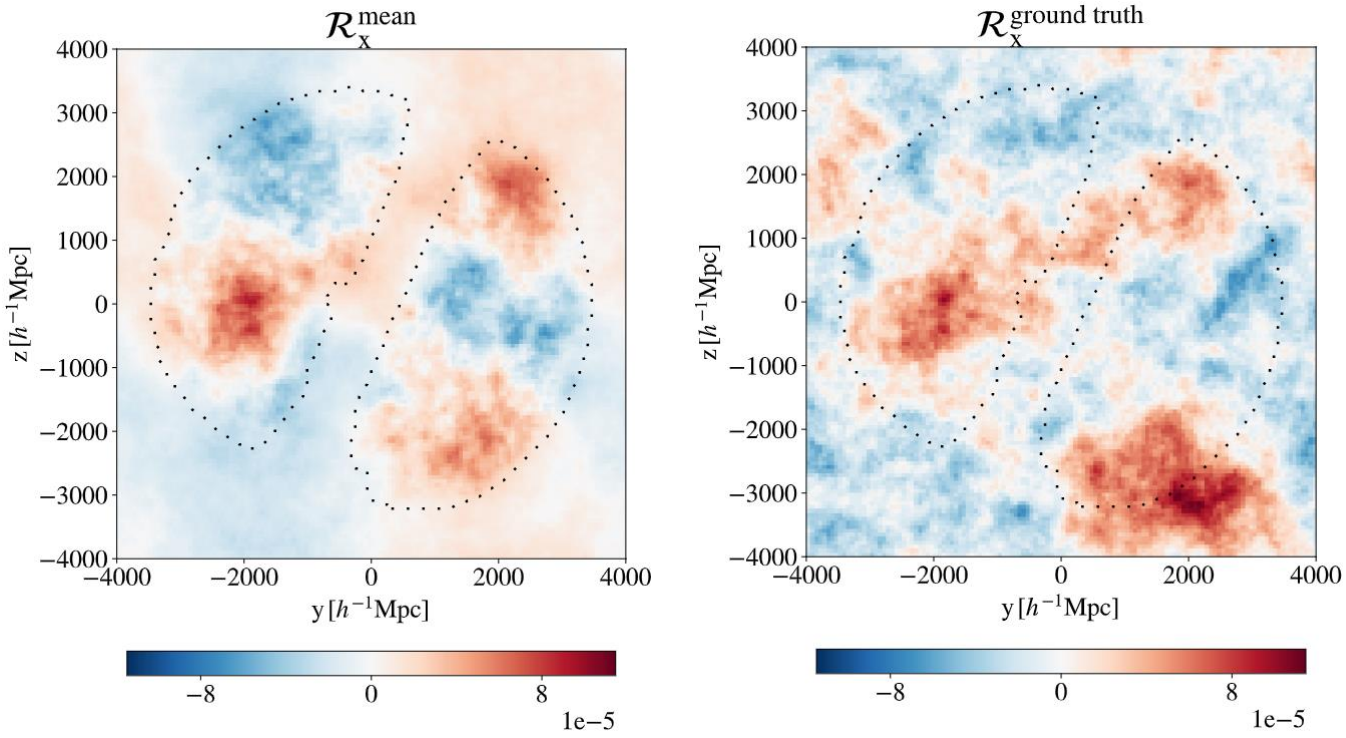
Porqueres *et al.*, 1812.05113

Lavaux, Jasche & FL, 1909.06396

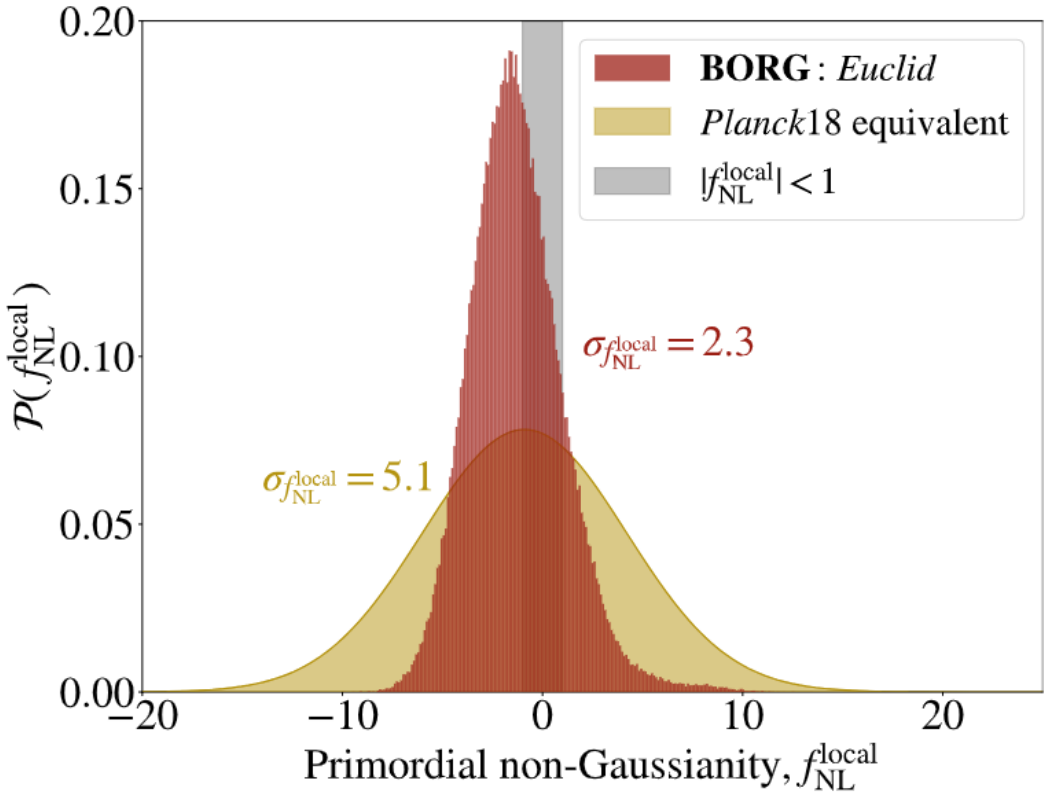
Field-based primordial physics: joint inference of primordial non-Gaussianity and initial conditions

- The physical model can be extended (as long as it is differentiable), e.g. with primordial physics:

Reconstruction of adiabatic curvature fluctuations



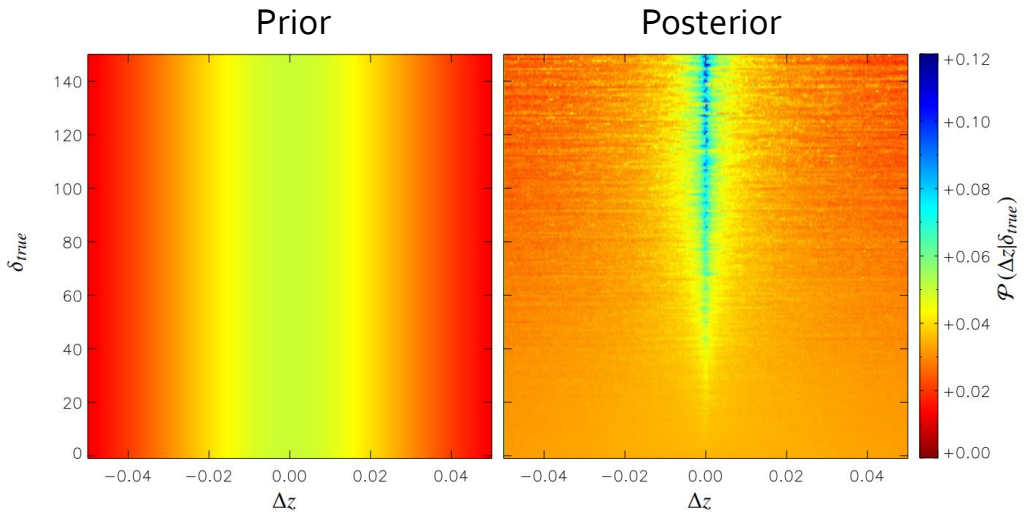
Posterior on $f_{\text{NL}}^{\text{local}}$

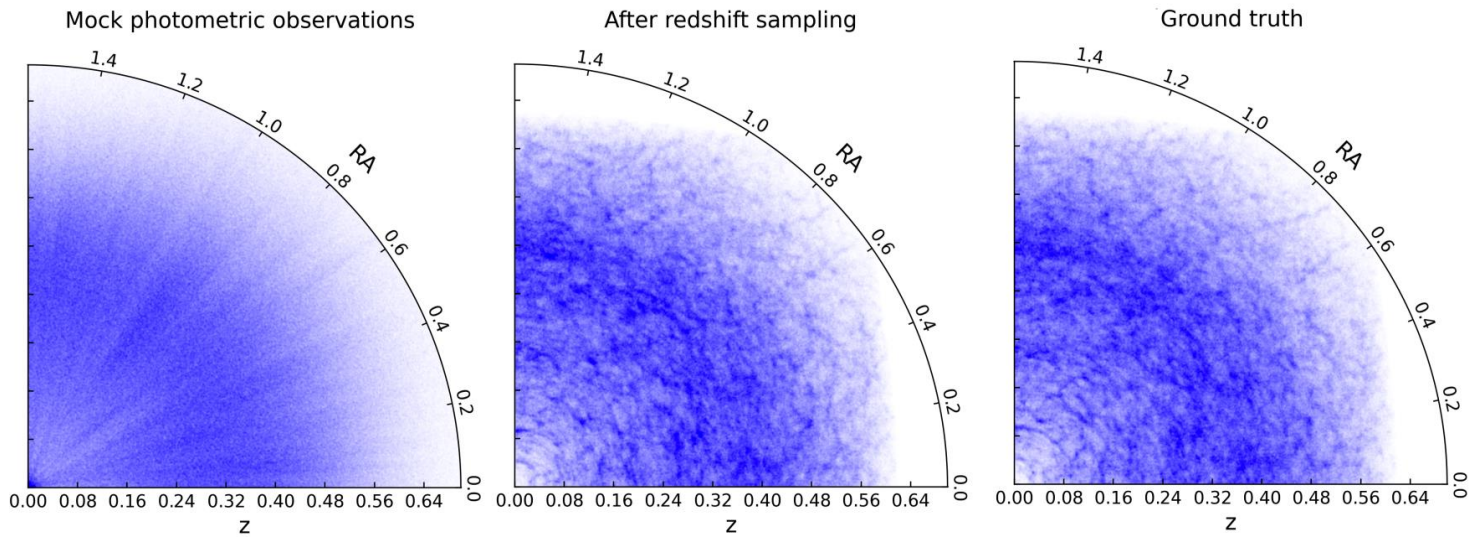


Andrews *et al.*, 2203.08838; Andrews *et al.*, in prep. (Euclid TWG WP4)

Field-based observational uncertainties: joint inference of photometric redshifts and density fields

- Sampling redshifts conditional on the density field sharpens the redshift pdfs…
- and propagates photometric galaxy clustering information to the density field reconstruction.

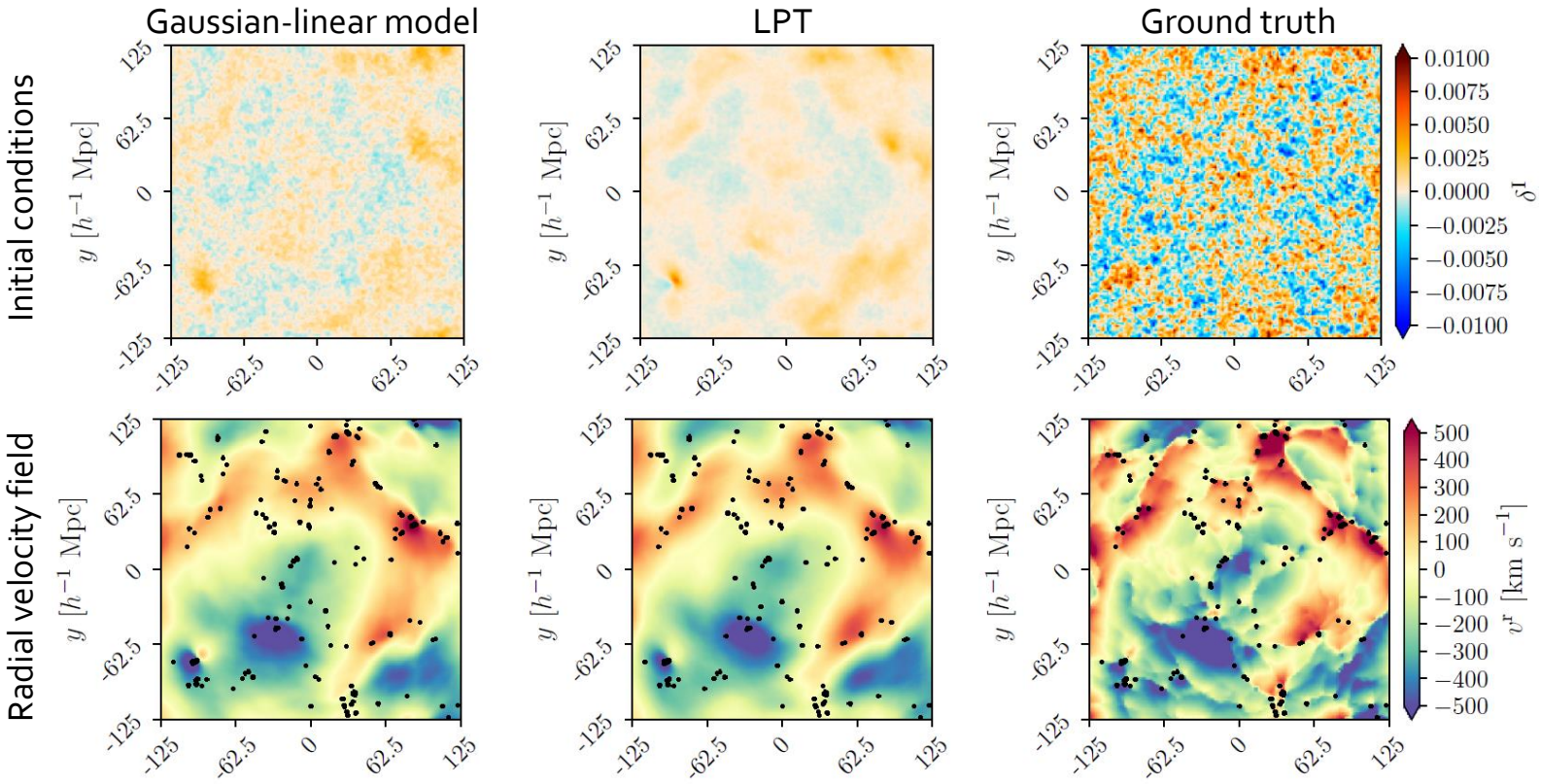




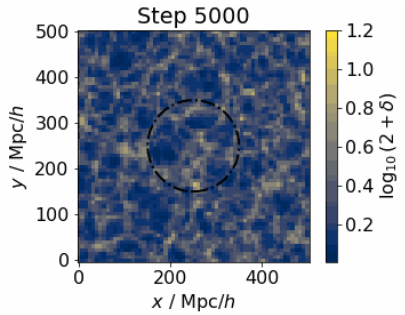
Jasche & Wandelt, 1106.2757; Tsaprazi, Jasche, Lavaux & FL, 2301.03581

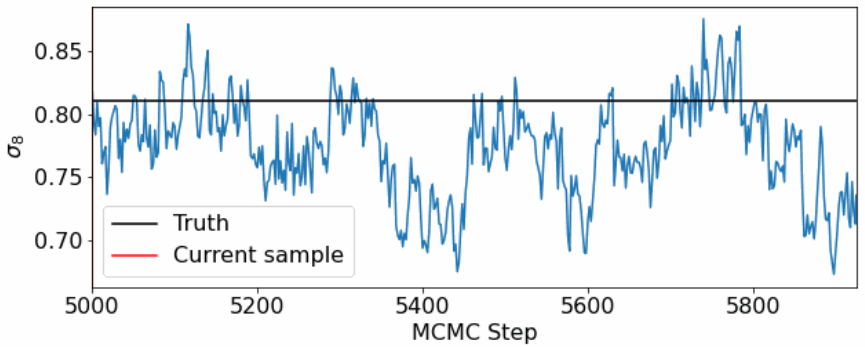
Field-level multi-tracer approach: joint initial conditions and velocity field reconstruction using distance tracers

- A field level approach naturally extends to muti-tracer / multi-wavelength / multi-messenger cosmology.



- New model in development:
  - Homogeneous and inhomogeneous Malmquist bias
  - Non-linear gravity (Lagrangian perturbation theory, ++)
  - Cosmological parameter sampling ($f\sigma_8$)





Lavaux, 1512.04534; Boruah, Lavaux & Hudson, 2111.15535; Prideaux-Ghee, FL, Lavaux, Heavens & Jasche, 2204.00023; Bartlett *et al.*, in prep.
See also work from Hoffman, Courtois, Sorce & CLUES team

Concluding thoughts

- Do not forget additional probes – they carry information needed to get the best science out of Euclid.
- Do not neglect alternative methods – they increase robustness to systematics and answer new questions.
- Bayesian analyses of galaxy surveys with fully non-linear numerical models is not an impossible task!
  - Implicit likelihood inference: algorithms for targeted questions, allowing the use of accurate simulators including all relevant physical and observational effects.
  - Analyses using an explicit field-level likelihood: general purpose inference of the initial conditions from cosmological observables (galaxy clustering, weak lensing, distance tracers), providing new measurements and predictions.