Bayesian inference of dark matter voids in galaxy surveys

Florent Leclercq, Jens Jasche, Paul M. Sutter,
Nico Hamaus, Benjamin D. Wandelt
florent.leclercq@polytechnique.org, jasche@iap.fr, sutter@iap.fr,
hamaus@iap.fr, wandelt@iap.fr

Summary

Establishing a link between cosmological observations and theories describing the early Universe is important because it can further our knowledge of fundamental physics on a wide range of energy and distance scales. In particular, the analysis of large-scale structure in the mildly non-linear regime remains one of the biggest cosmological challenges.

We apply the BORG algorithm [1] to Sloan Digital Sky Survey Data Release 7. The method results in the physical inference of the four-dimensional initial density field at redshift $z = 1000$, evolving gravitationally to the observed density field at redshift $z = 0$, and provides an accurate quantification of corresponding uncertainties.

Building upon these results, we generate a set of constrained reconstructions of the present large-scale dark matter distribution.

As a physical illustration, we apply a void identification algorithm to them. In this fashion, we access voids defined by the inferred dark matter field, not by galaxies, greatly alleviating the bias problem. In addition, the use of full-scale physical density fields yields a drastic reduction of statistical uncertainty in void catalogs. This new catalog is an enhanced data set for cross-correlation with other cosmological probes such as the cosmic microwave background, to study the integrated Sachs-Wolfe effect, or gravitational lensing shear maps.

Materials and Methods

SDSS DR7 data. We employ the `Sample dr72` of the New York University Value Added Catalogue (NYU-VAGC), based on the final data release (DR7) of the Sloan Digital Sky Survey (SDSS).

BORG. The algorithm explores numerically the posterior distribution, via efficient Hamiltonian Markov Chain Monte Carlo (HMC) sampling. The approximate physical model for cosmic structure formation is second-order Lagrangian perturbation theory (2LPT).

VIDE. The void finder is a modified version of ZOBOV that uses Voronoi tessellations of the tracer particles to estimate the density field and a watershed algorithm to group Voronoi cells into voids.

References

[1] J. Jasche, B. D. Wandelt, *MNRAS* **432**, 894 (2013).

[2] J. Jasche, F. Leclercq, B. D. Wandelt, *in prep.*

[3] J. Jasche, F. Leclercq, E. Romano-Díaz, B. D. Wandelt, *in prep.*

[4] F. Leclercq, J. Jasche, P. M. Sutter, N. Hamaus, B. D. Wandelt, *in prep.*

[5] P. M. Sutter, G. Lavaux, B. D. Wandelt, D. H. Weinberg, M. S. Warren, *MNRAS* **438**, 3177 (2014).

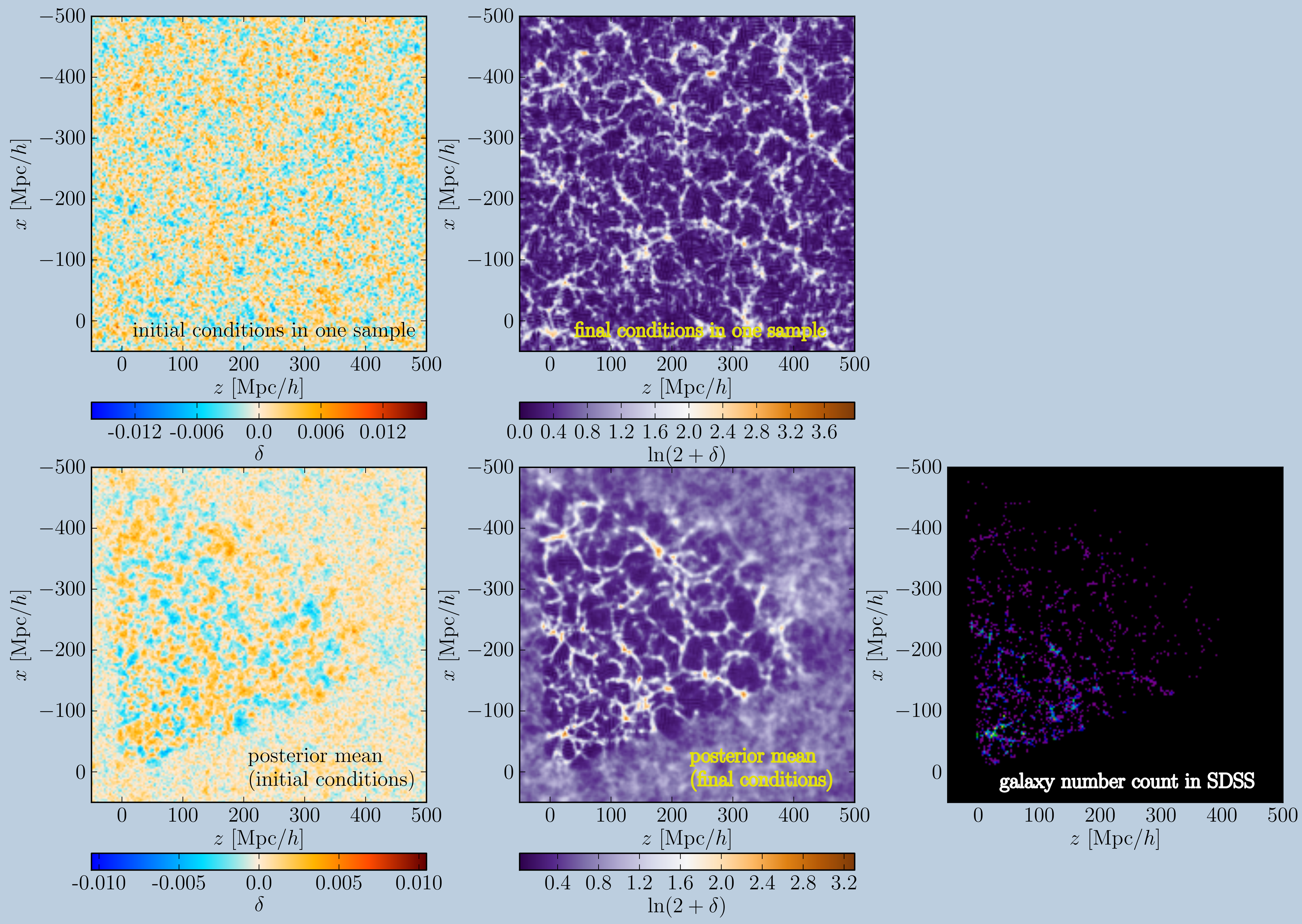
Acknowledgements

FL acknowledges funding from an AMX grant (École polytechnique ParisTech). JJ is supported by a Feodor Lynen Fellowship by the Alexander von Humboldt foundation and by BDW's senior Excellence Chair of the Agence Nationale de la Recherche. PMS and BDW acknowledge support from NSF grant AST-0908902. BDW acknowledges support from the Agence Nationale de la Recherche (ANR-10-CEXC-004-01), the UPMC Chaire Internationale in Theoretical Cosmology, NSF grants AST-0708849, NSF AST-0908693 ARRA and NASA subcontract JPL 1236748. This work made in the ILP LABEX (ANR-10-LABX-63) was supported by French state funds managed by the ANR within the Investissements d'Avenir programme (ANR-11-IDEX-0004-02).

Bayesian physical inference of the initial conditions [2]

Application of BORG (Bayesian Origin Reconstruction from Galaxies) [1] to the SDSS DR7 galaxies:

- 2LPT: **physical model of gravitational dynamics** in the linear and mildly non-linear regime, linking initial density fields ($z = 1000$) to the presently observed large-scale structure.
- HMC: **full-scale Bayesian inference framework**, exploring the posterior by sampling the joint distribution of all parameters involved.



Each sample is a "possible version of the truth" in the form of a full physical realization of dark matter particles, tracing both the density and the velocity fields. The variation between samples quantifies joint and correlated uncertainties (survey geometry, selection effects, galaxy bias, redshift distortions, noise, cosmic variance) inherent to any cosmological observation.

Data-constrained realizations of the Universe [3]

Generation of a set of data-constrained realizations of the present large-scale structure:

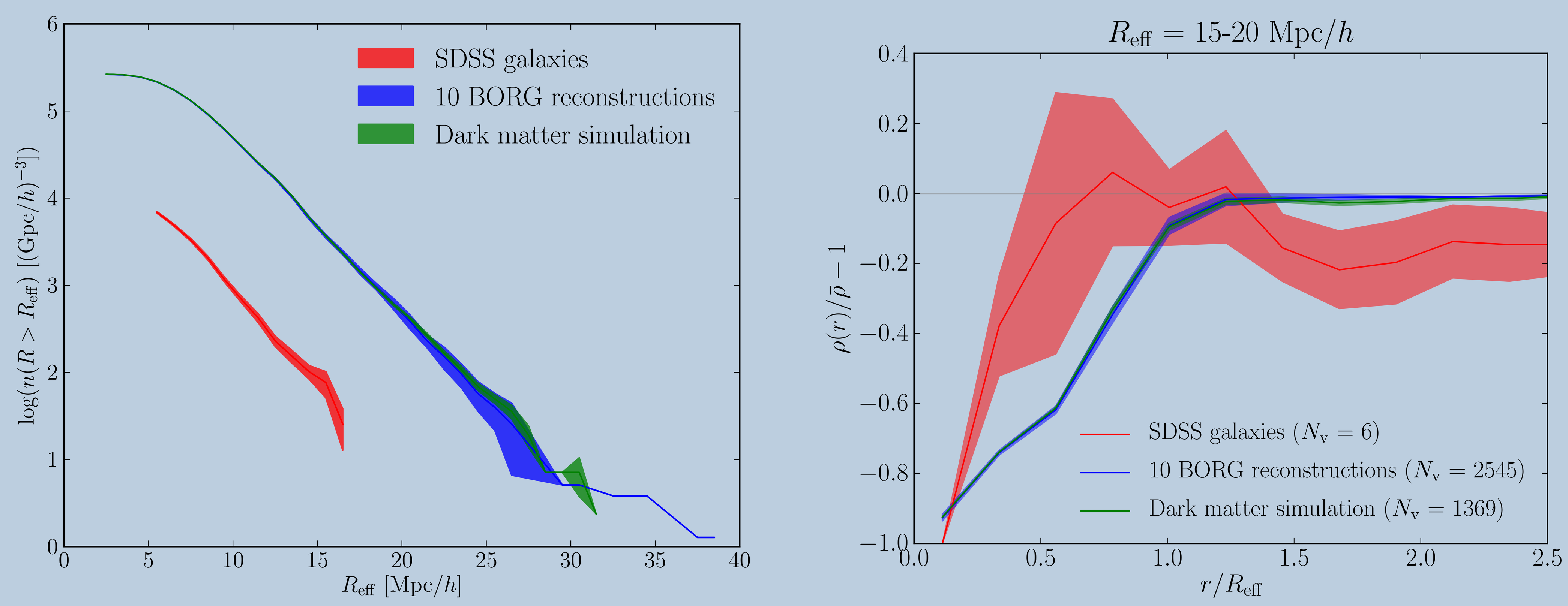
- Some samples of inferred primordial conditions are evolved with 2LPT to $z = 69$,
- then with a fully non-linear cosmological simulation (using GADGET-2) from $z = 69$ to $z = 0$.

A dynamic, non-linear physical model naturally introduces some correlations between the constrained and unconstrained parts, which yields **reliable extrapolations** for certain aspects of the model that have not yet been constrained by the data.

Dark matter voids in the SDSS galaxy survey [4]

Application of VIDE (the Void IDentification and Examination pipeline) to the constrained parts of these realizations:

- We find physical cosmic voids in the field traced by the dark matter particles, probing a level deeper in the mass distribution hierarchy than galaxies [5], and greatly **alleviating the bias problem** for cosmological interpretation of final results.
- Due to the high density of tracers, we find an order of magnitude more voids at small scales than the voids directly traced by the SDSS galaxies, which sample the underlying mass distribution only sparsely. Our inference framework therefore yields a drastic **reduction of statistical uncertainty** in voids catalogs.



For usual voids statistics such as density profiles of stacked voids, the results we obtain are consistent with N-body simulations prepared with the same setup.