Bayesian large-scale structure inference

A new approach toward joint BAO and RSD analysis

Florent Leclercq

Institute of Cosmology and Gravitation, University of Portsmouth

http://icg.port.ac.uk/~leclercq/

July 8th, 2016

In collaboration with:

Jens Jasche (ExC Universe, Garching), Guilhem Lavaux (IAP),
Will Percival (ICG), Benjamin Wandelt (IAP/U. Illinois)

A disclaimer

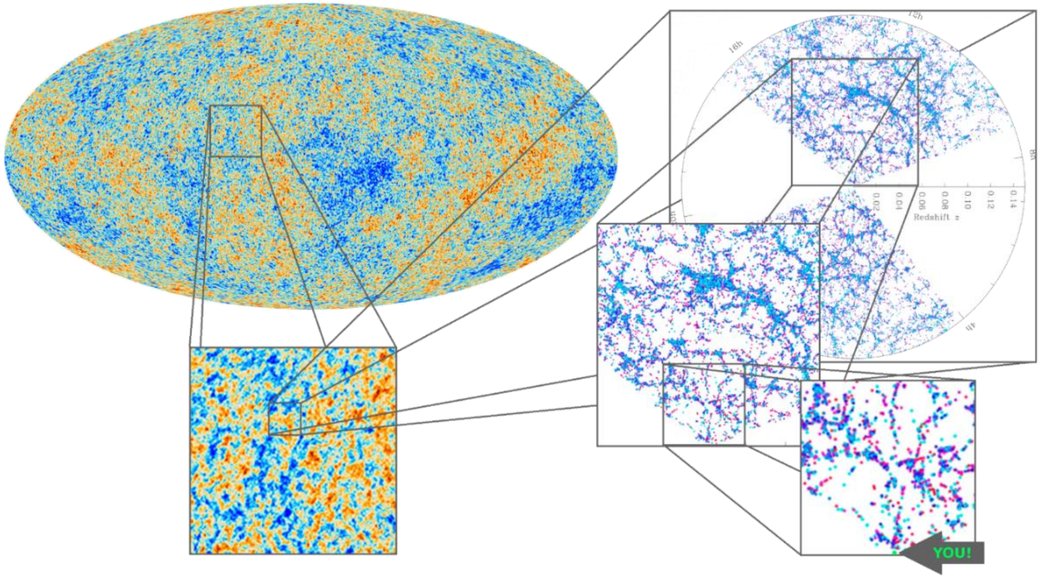
E. T. Jaynes (1922-1998)

"A previous acquaintance with probability and statistics is not necessary; indeed, a certain amount of innocence in this area may be desirable, because there will be less to unlearn."
— Edwin Thompson Jaynes (2003), *Probability Theory: The Logic of Science*

- For the purpose of this talk, please forget the following concepts:
  - "**measurements**" of power spectra / correlation functions, etc.
  - "**mock**" catalogs
  - "**weights**"
  - inverse modeling for "**BAO reconstruction**"
  - "**corrections**" to the data
  - estimators / χ^2 / maximum likelihood, etc.

What we want to know from the LSS

1. How do we **test** our understanding of structure formation?
   Precise tests of ΛCDM require many modes…
   Fitting of average power spectra will hit a fundamental limit due to the absence of phase information.
2. How did structure appear in **our** Universe?



Can we just **fit the entire survey**?

The CMB problem

- The "genius idea": **complicate to simplify**!

$$\mathcal{P}(C_\ell|d) \quad D \approx 10^3 \quad \Longrightarrow \quad \mathcal{P}(s, C_\ell|d) \quad D \approx 10^7$$

- Original Gibbs sampling algorithm:

$$C_\ell \curvearrowleft \mathcal{P}(C_\ell|d, s)$$ (power spectrum sampling from inverse-Γ distribution)

$$s \curvearrowleft \mathcal{P}(s|d, C_\ell)$$ (sampling from the Wiener filter posterior)

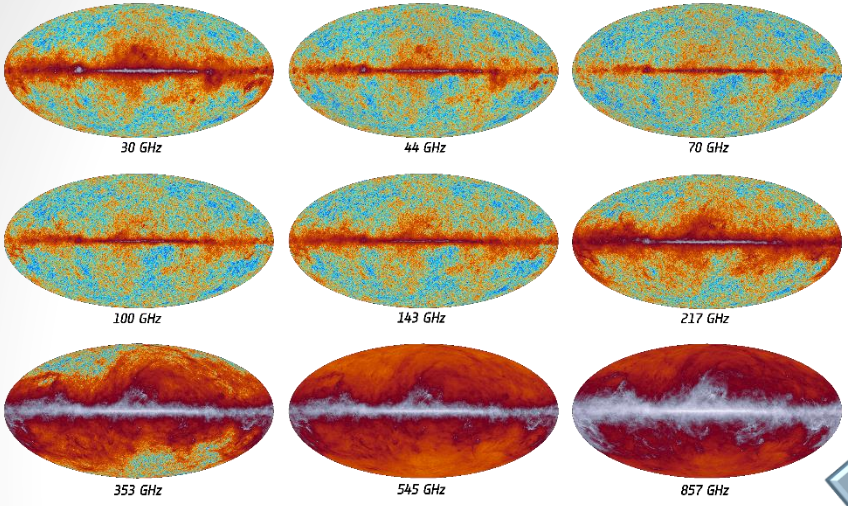
Generates samples of $\mathcal{P}(s, C_\ell|d)$!

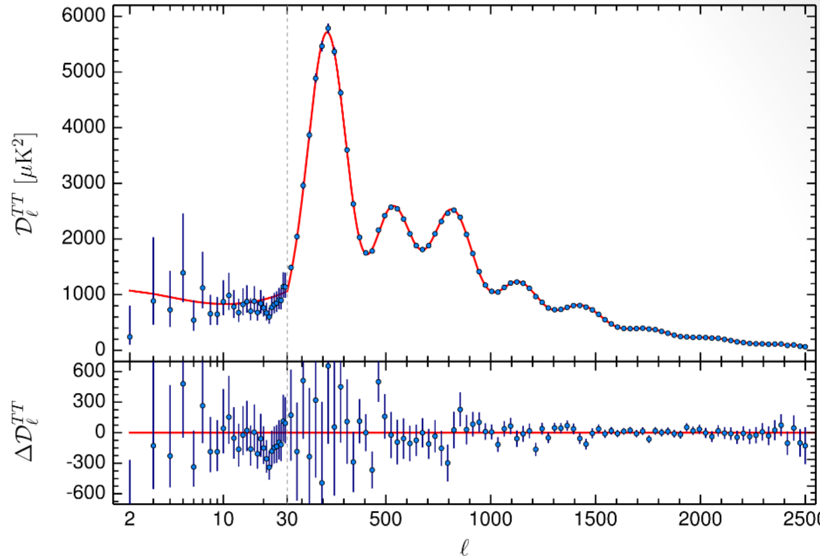
Wandelt, Larson & Lakshminarayanan 2004, arXiv:astro-ph/0310080

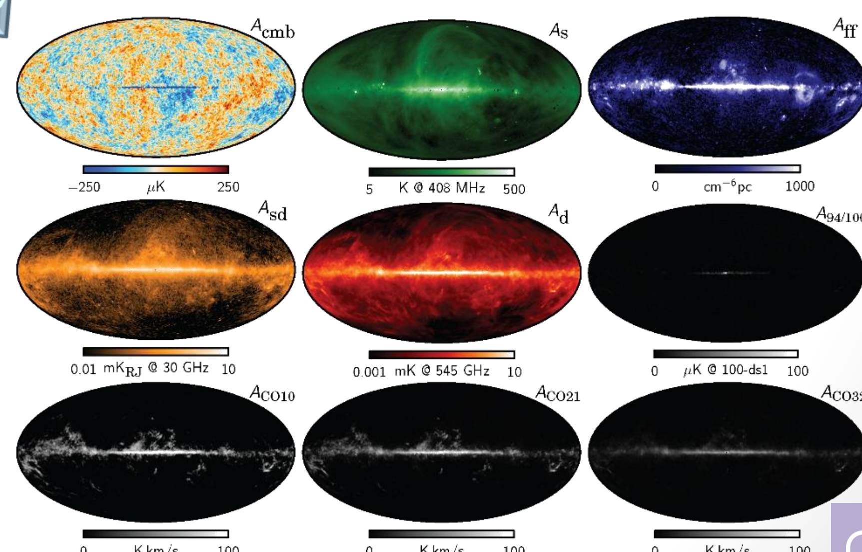
- But it is also possible with more parameters:

$$\begin{aligned} C_\ell &\curvearrowleft \mathcal{P}(C_\ell|s, a_\mathrm{fg}, \beta_\mathrm{fg}, d) \\ s &\curvearrowleft \mathcal{P}(s|C_\ell, a_\mathrm{fg}, \beta_\mathrm{fg}, d) \\ a_\mathrm{fg} &\curvearrowleft \mathcal{P}(a_\mathrm{fg}|s, C_\ell, \beta_\mathrm{fg}, d) \\ \beta_\mathrm{fg} &\curvearrowleft \mathcal{P}(\beta_\mathrm{fg}|s, C_\ell, a_\mathrm{fg}, d) \end{aligned} \quad \Longrightarrow \quad \mathcal{P}(s, C_\ell, a_\mathrm{fg}, \beta_\mathrm{fg}|d)$$

The CMB problem







Planck collaboration (2013-2015)

Bayesian Large-Scale Structure Inference Codes

Bayesian large-scale structure inference codes

- Density field inference:

Physical modeling →

ARES, HADES, BORG, BORGPM

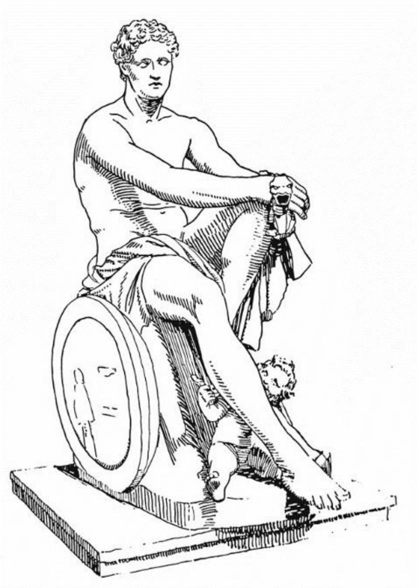
← Data aspects

- Velocity field / distance inference:

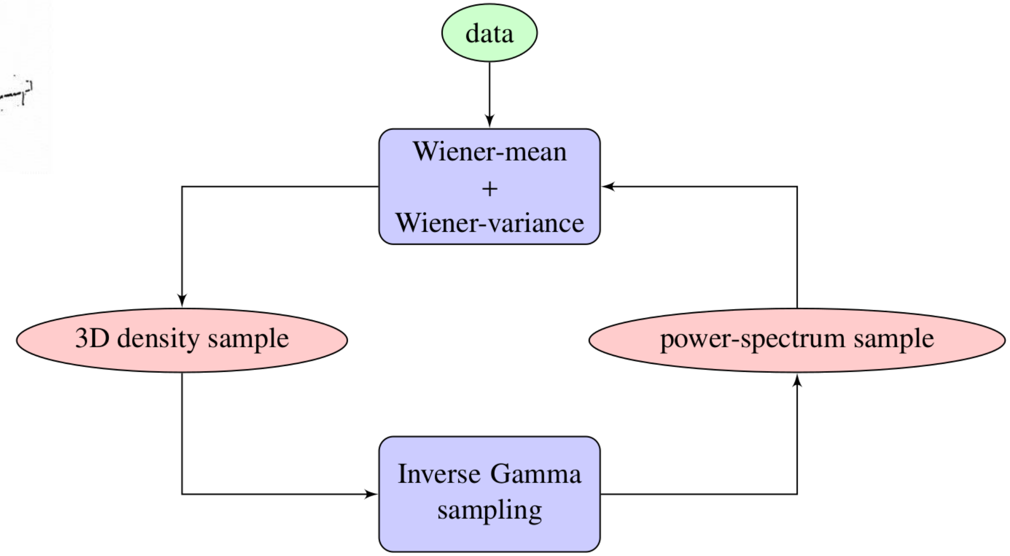
VIRBIUS

- All codes give **samples** of the respective posterior distributions.
  ⇒ **Only one run** of the code gives uncertainties!

ARES: *Algorithm for REconstruction and Sampling*

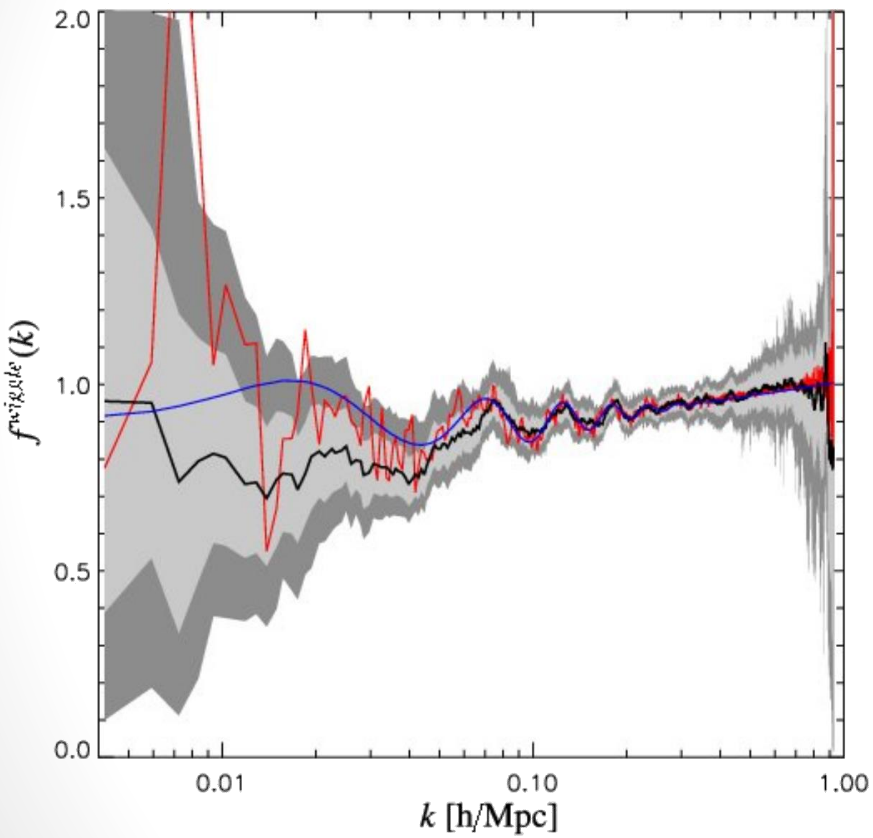


- **Data model**:
  - Gaussian random field for the density (Wiener filtering)
  - Inverse-Γ distribution for the power spectrum
- **Sampler**: Gibbs sampling

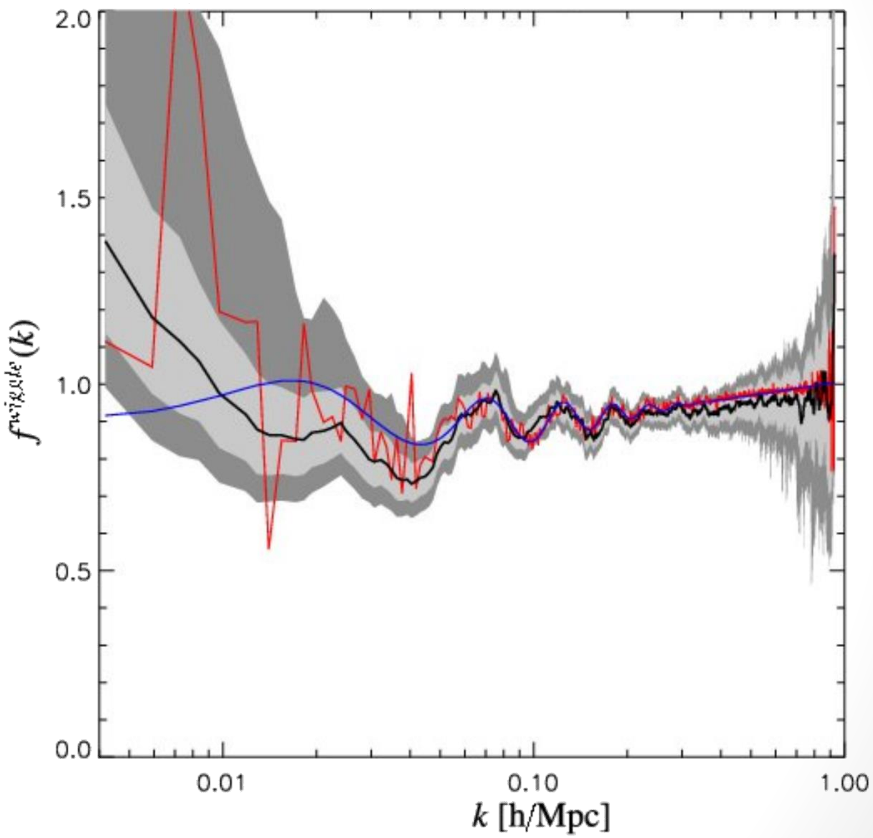


Jasche, Kitaura, Wandelt & Enßlin 2010, arXiv:0911.2493

BAO inference by ARES



Jeffrey's prior

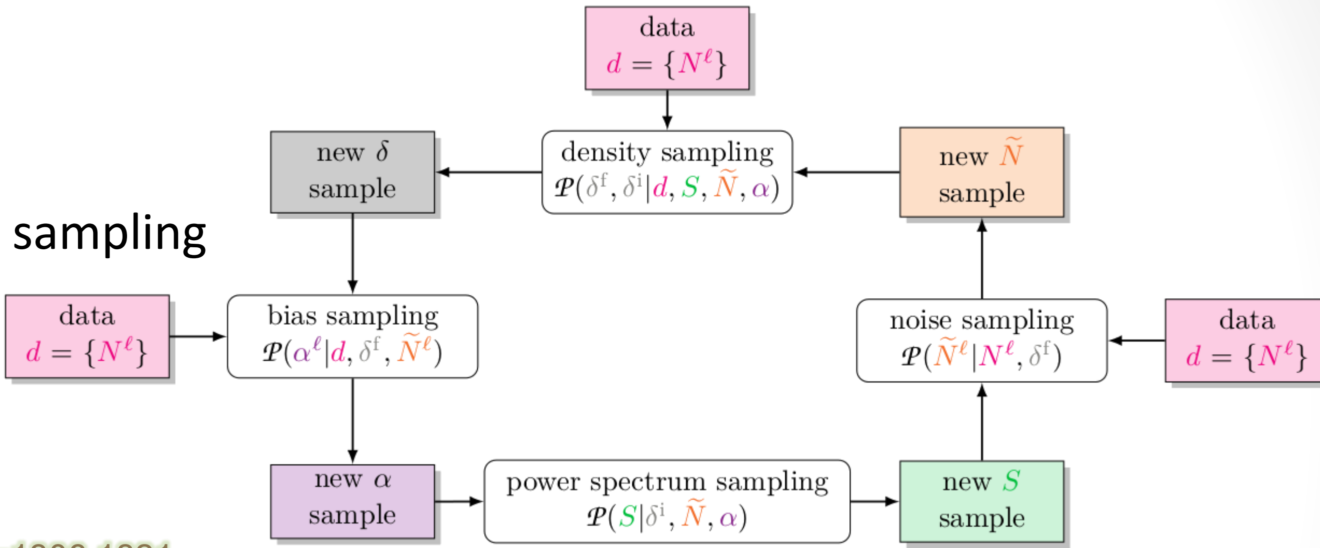


Inverse-Γ prior

Jasche, Kitaura, Wandelt & Enßlin 2010, arXiv:0911.2493

ARES: *Algorithm for REconstruction and Sampling*

- ARES2
  - multi-survey
  - bias and noise sampling



Jasche & Wandelt 2013, arXiv:1306.1821

- ARES3
  - complete rewriting of the code
  - includes **redshift-space distortions**
  - samples **foregrounds** (galactic dust, stars, etc.)

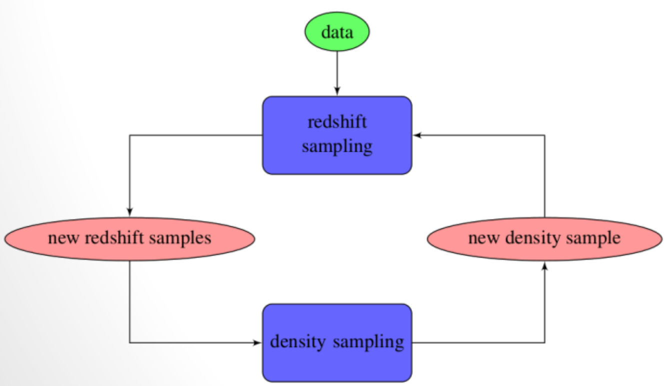
Jasche & Lavaux, in prep.
Lavaux & Jasche, in prep.

HADES: *HAmiltonian Density Estimation and Sampling*

- **Data model**:
  - Log-normal density field
  - Poisson likelihood
- **Sampler**: Hamiltonian Monte Carlo algorithm

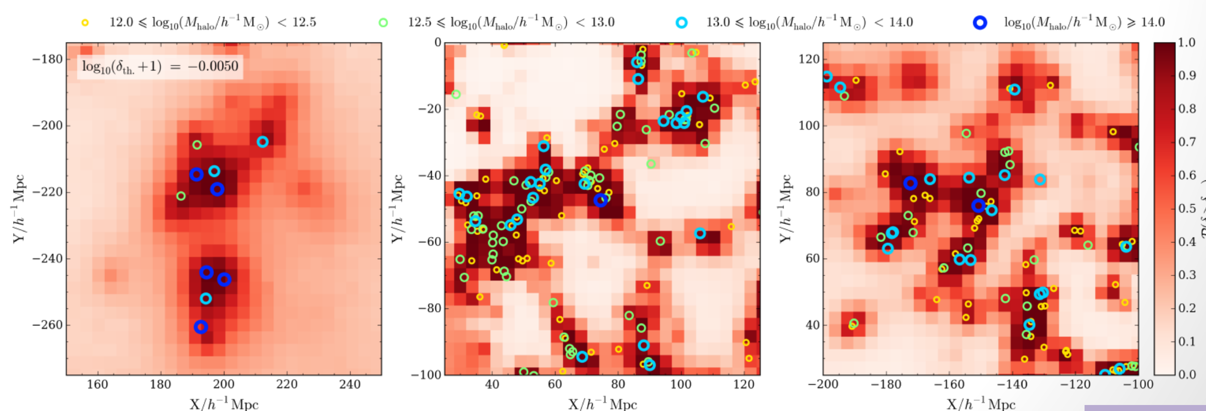
Jasche & Kitaura 2010, arXiv:0911.2496

Photometric redshift sampling



Jasche & Wandelt 2013, arXiv:1203.3639

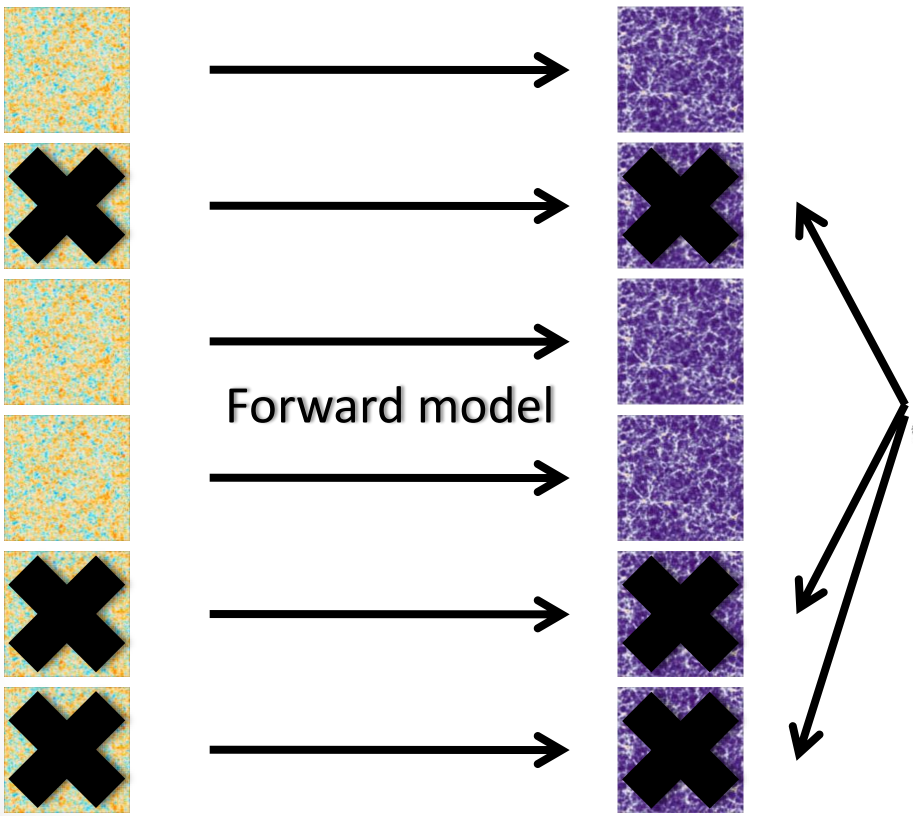
Halo detection using HADES

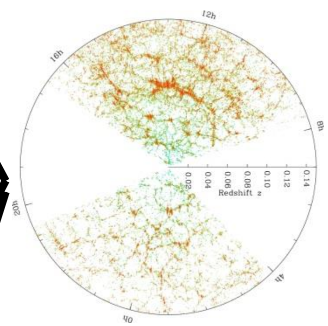


Merson *et al.* 2015, arXiv:1505.03528

Bayesian forward modeling: the ideal scenario

Forward model = N-body simulation + Halo occupation + Galaxy formation + Feedback + …



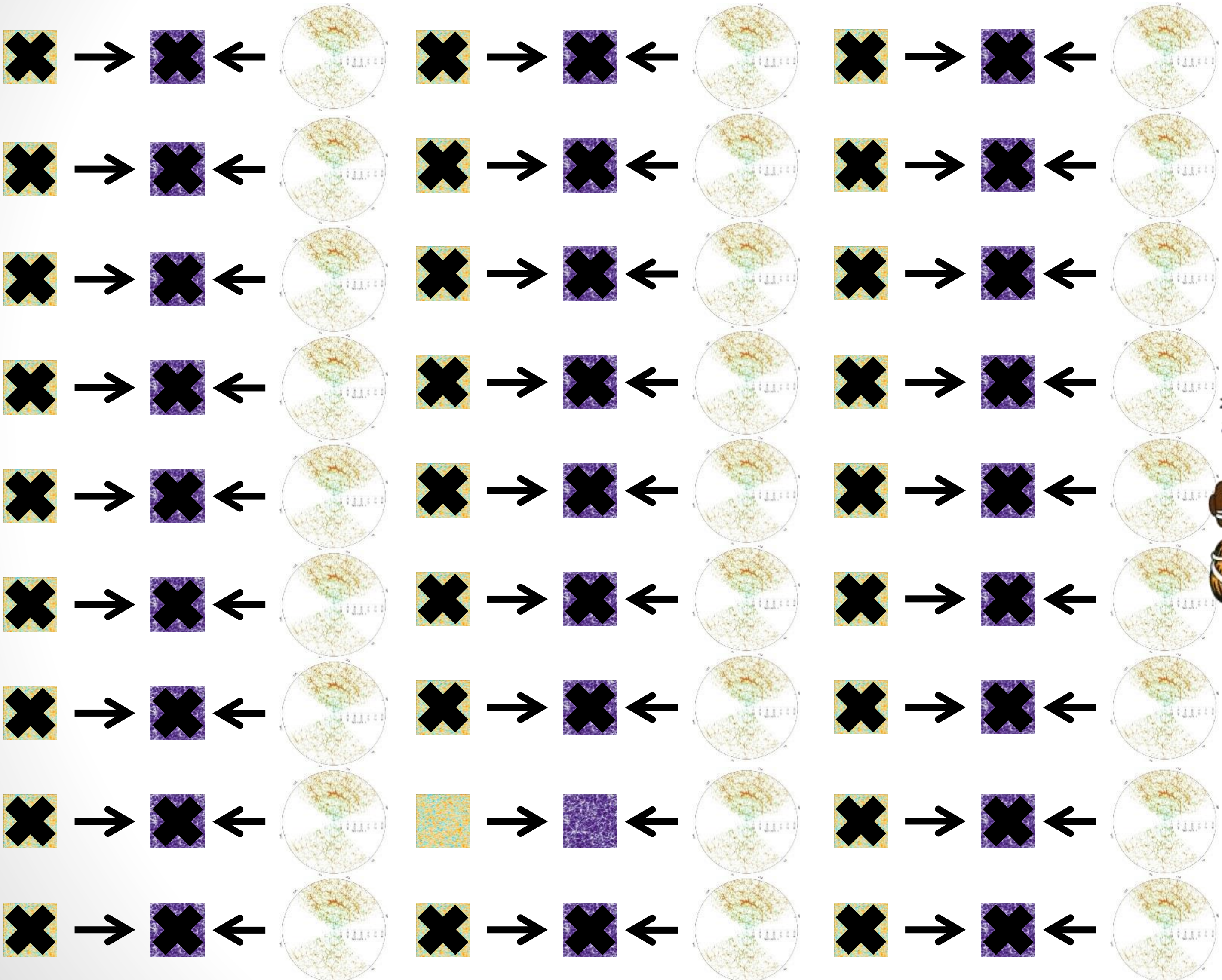


Observations

All possible ICs

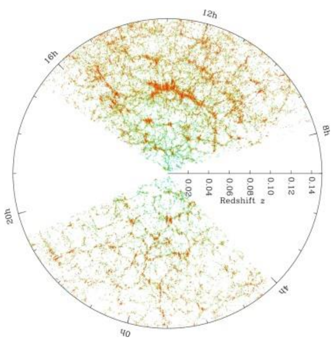
All possible FCs

Bayesian forward modeling: the ideal scenario



BORG: *Bayesian Origin Reconstruction from Galaxies*

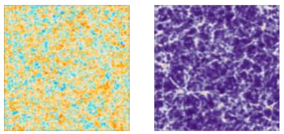
- **Sampler**: Hamiltonian Monte Carlo
- **Data model**:
  - Gaussian prior for the initial conditions
  - Second-order Lagrangian perturbation theory (2LPT)
  - Poisson likelihood



Observations

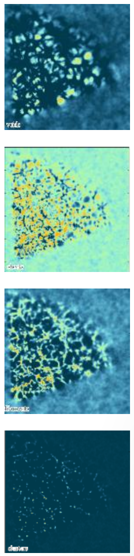
(galaxy catalog + meta-data: selection functions, completeness…)

→ BORG →



Inferred dark matter density

→



Cosmic web analysis

see also:
Kitaura 2013, arXiv:1203.4184
Wang, Mo, Yang & van den Bosch 2013, arXiv:1301.1348

Jasche & Wandelt 2013, arXiv:1203.3639

BORG: *Bayesian Origin Reconstruction from Galaxies*

- BORG2
  - luminosity-dependent galaxy **bias**
  - automatic calibration of **noise** levels

Jasche, FL & Wandelt 2015, arXiv:1409.6308

- BORG3
  - entire code rewriting, MPI+OpenMP parallel
  - improved **bias** model
  - includes **redshift-space distortions**

Lavaux & Jasche, in prep.
Jasche & Lavaux, in prep.

- BORGPM
  - includes a full **particle-mesh code**

see also Wang *et al.* 2014, arXiv:1407.3451

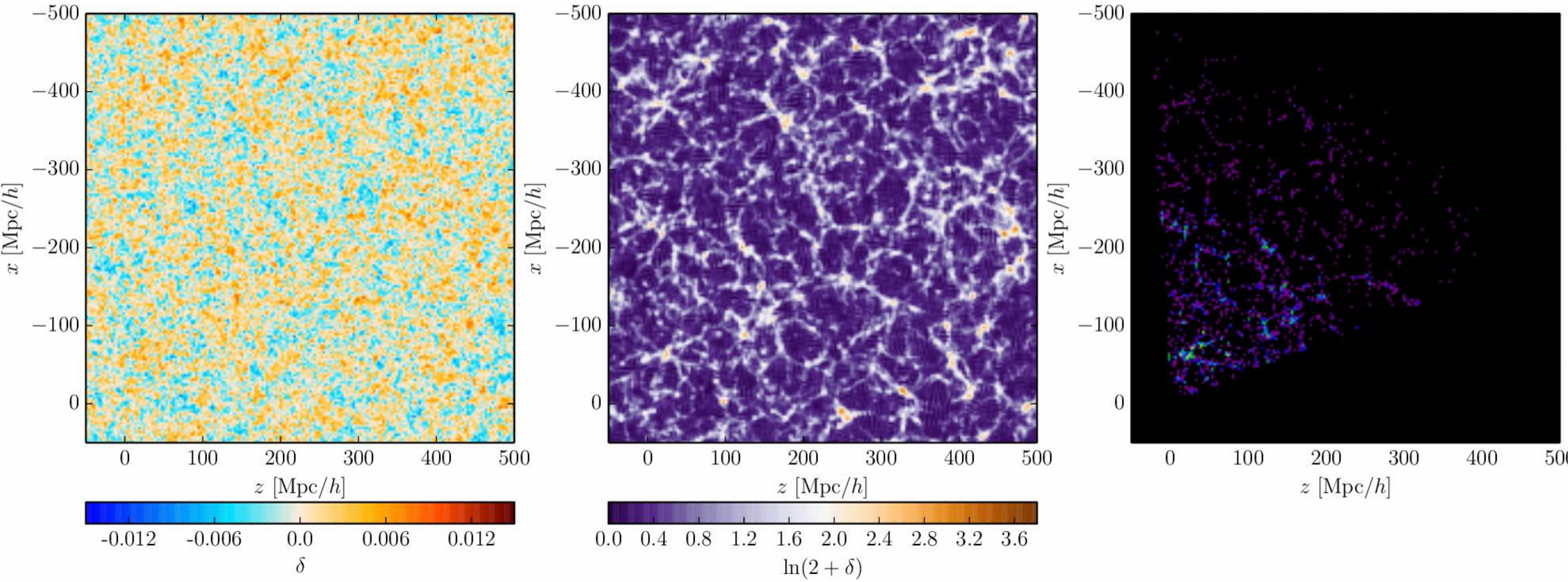
VIRBIUS: *VelocIty Reconstruction using Bayesian Inference Software*

- **velocity**/density field inference
- true **distances** of galaxies

Lavaux 2016, arXiv:1512.04534

INFERENCE RESULTS

BORG at work: SDSS chrono-cosmography



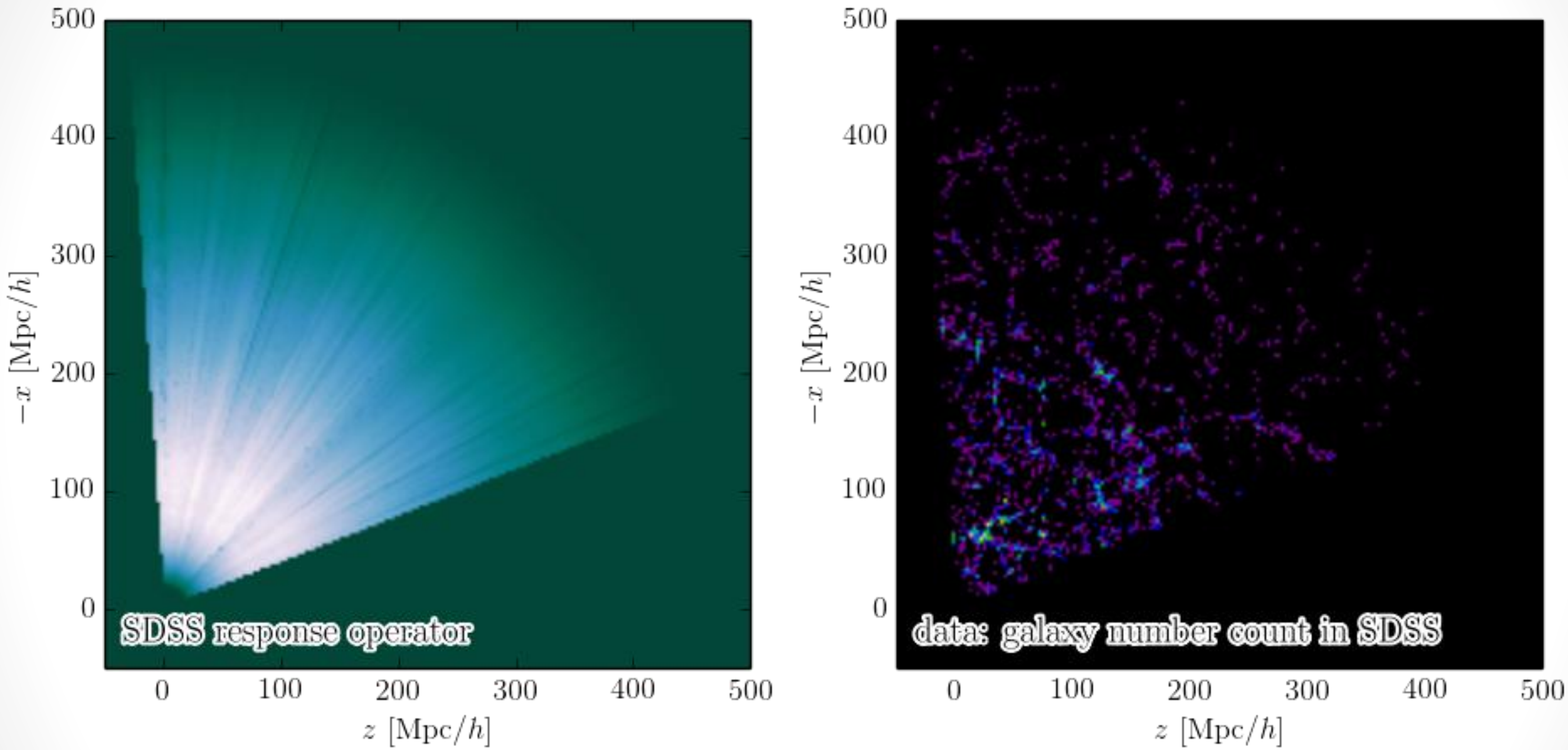
Initial conditions | Final conditions | Observations

The BORG SDSS run:

334,074 galaxies, ≈ 17 millions parameters, 3 TB of primary data products,
12,000 samples, ≈ 250,000 data model evaluations, 10 months on 32 cores

Jasche, FL & Wandelt 2015, arXiv:1409.6308

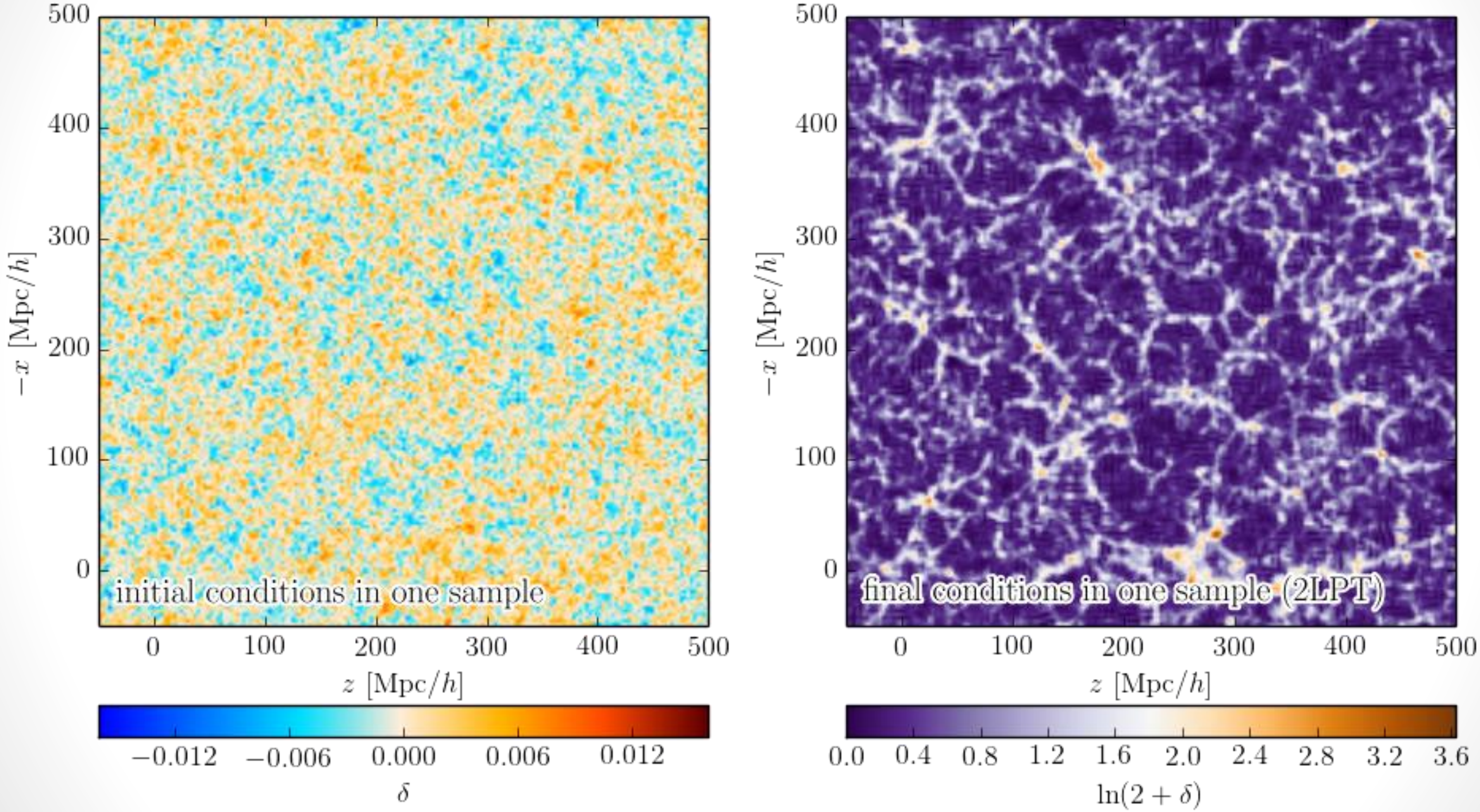
Bayesian chrono-cosmography from SDSS DR7



Data

Jasche, FL & Wandelt 2015, arXiv:1409.6308

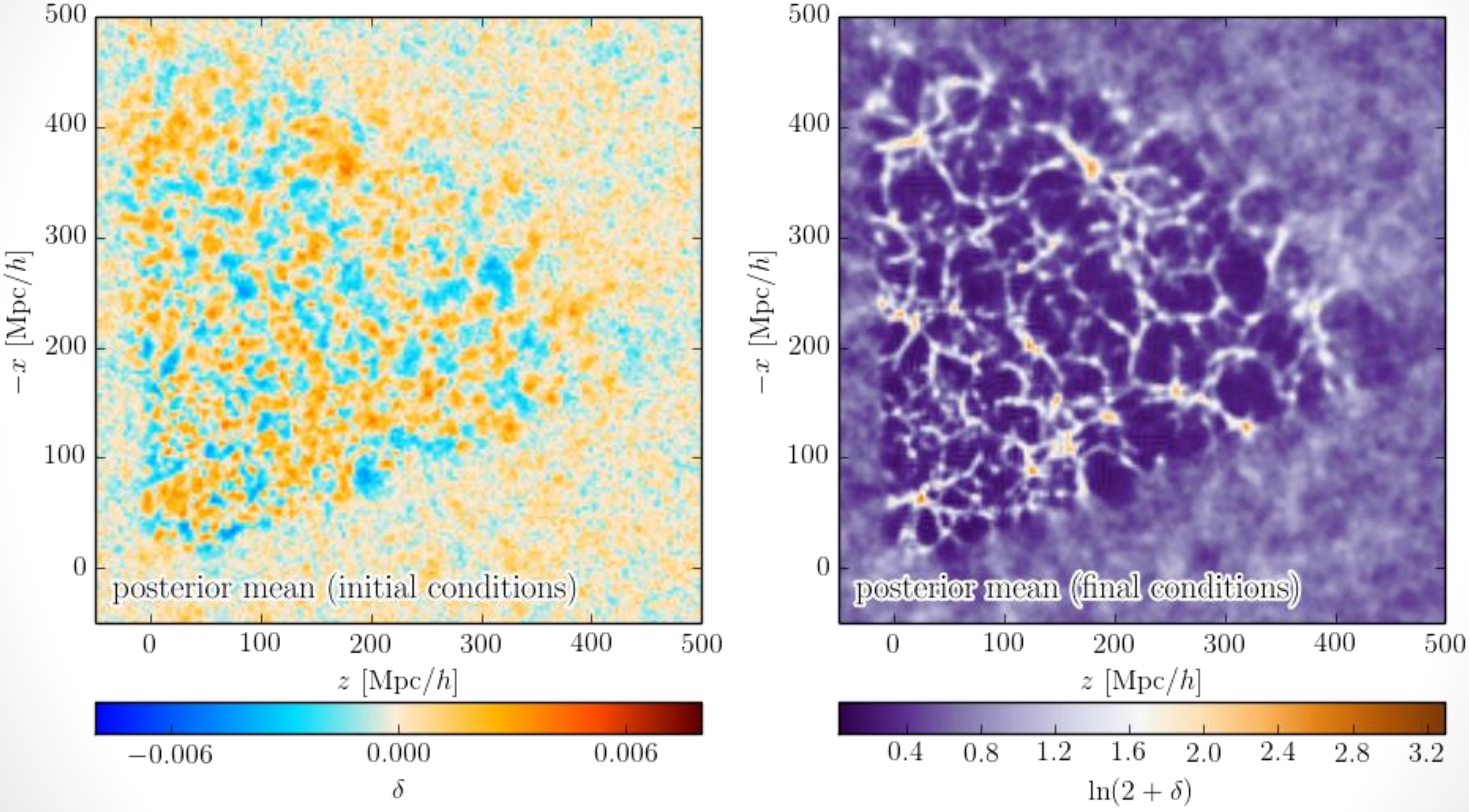
Bayesian chrono-cosmography from SDSS DR7



One sample

Jasche, FL & Wandelt 2015, arXiv:1409.6308

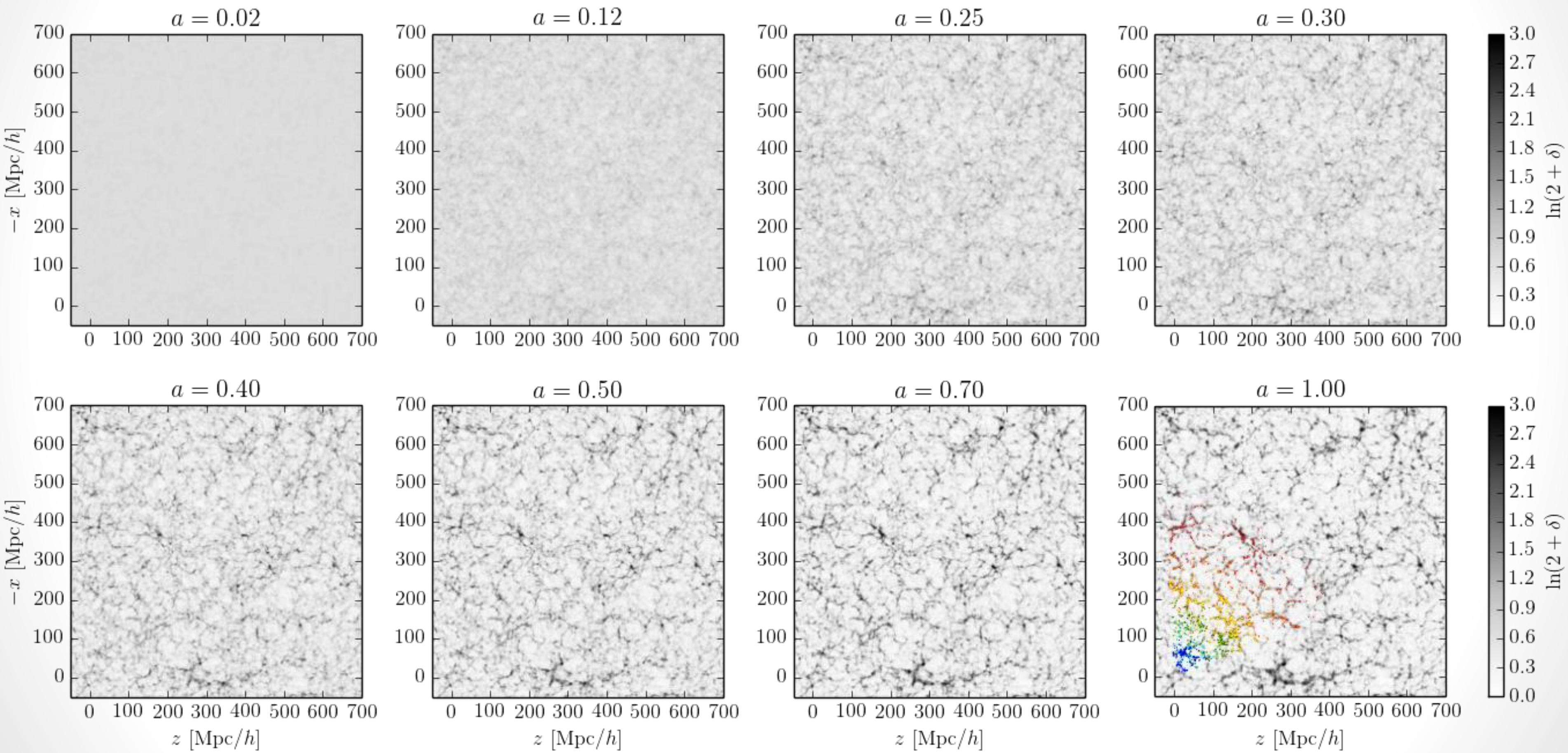
Bayesian chrono-cosmography from SDSS DR7



Posterior mean

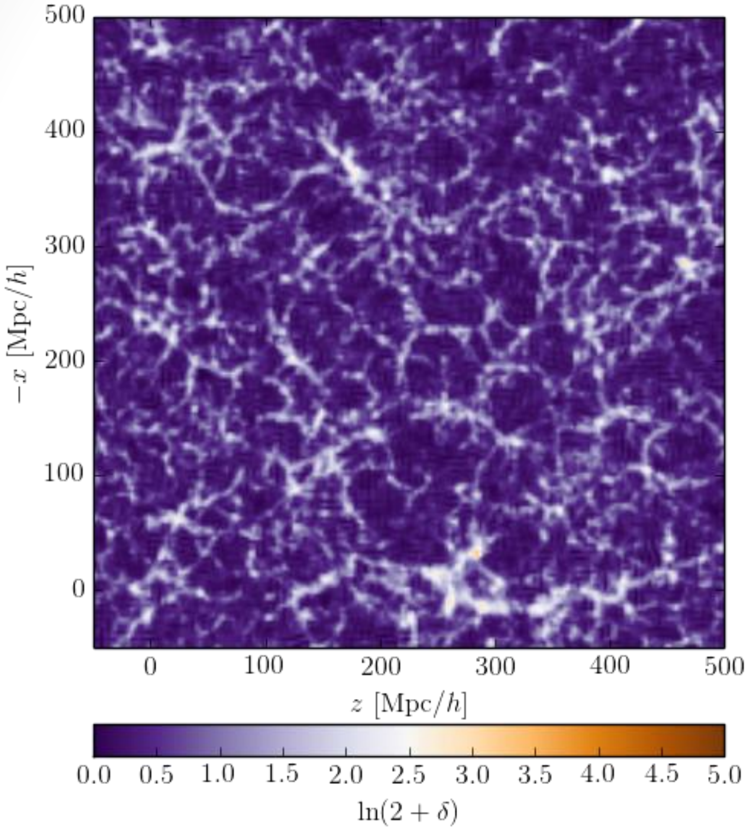
Jasche, FL & Wandelt 2015, arXiv:1409.6308

Evolution of cosmic structure



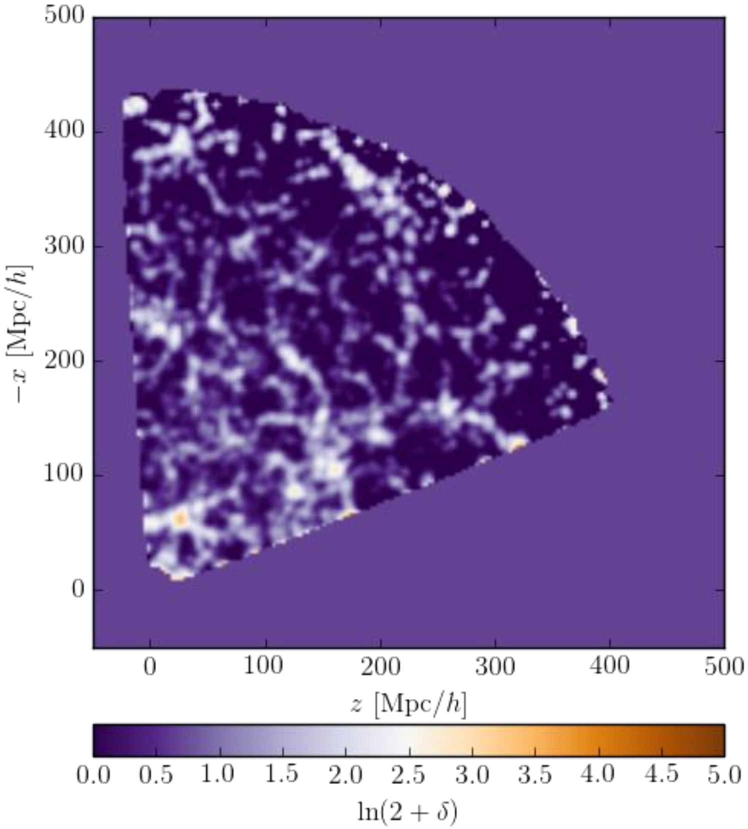
Jasche, FL & Wandelt 2015, arXiv:1409.6308

"Inference" vs "Reconstruction"



One sample of the BORG SDSS run

Jasche, FL & Wandelt 2015, arXiv:1409.6308

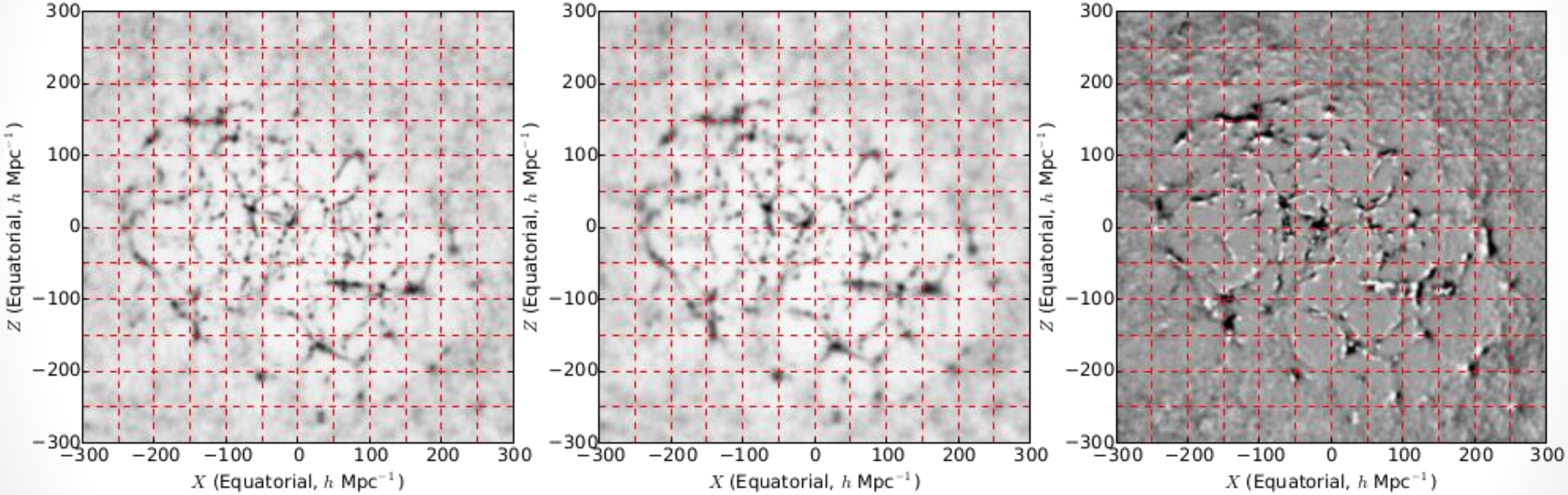


Field used for reconstruction of the SDSS (code from W. Percival)

Burden, Percival & Howlett 2015, arXiv:1504.02591

The differences: Gaussian prior – treatment of observational effects – galaxy bias – uncertainty quantification – free parameters

Impact of redshift-space distortions



Real space

Redshift space

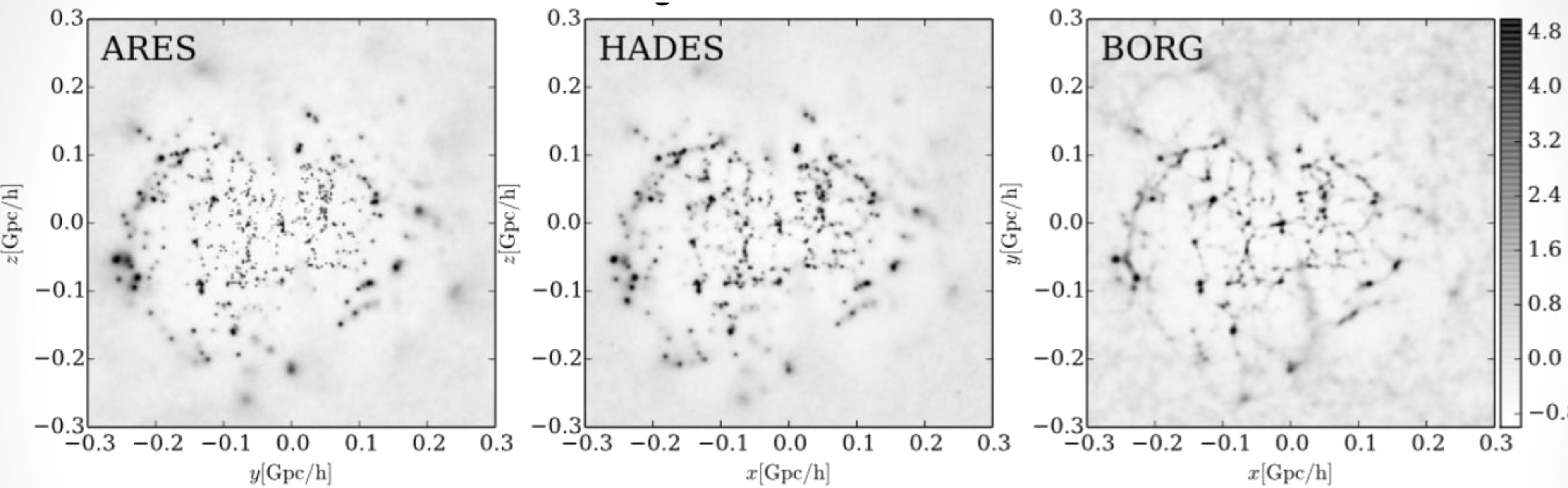
Difference

2M++, mean final matter density field

Lavaux & Jasche 2015, arXiv:1509.05040

Comparing BLSS methods

Gaussian (a.k.a. Wiener filter) | Lognormal – Poisson | 2LPT – Poisson



Jasche *et al.* 2010, arXiv:0911.2493
Jasche & Wandelt 2013, arXiv:1306.1821

Jasche & Kitaura 2010, arXiv:0911.2496

Jasche & Wandelt 2013, arXiv:1203.3639

- Which scheme performs best? Ask the data!

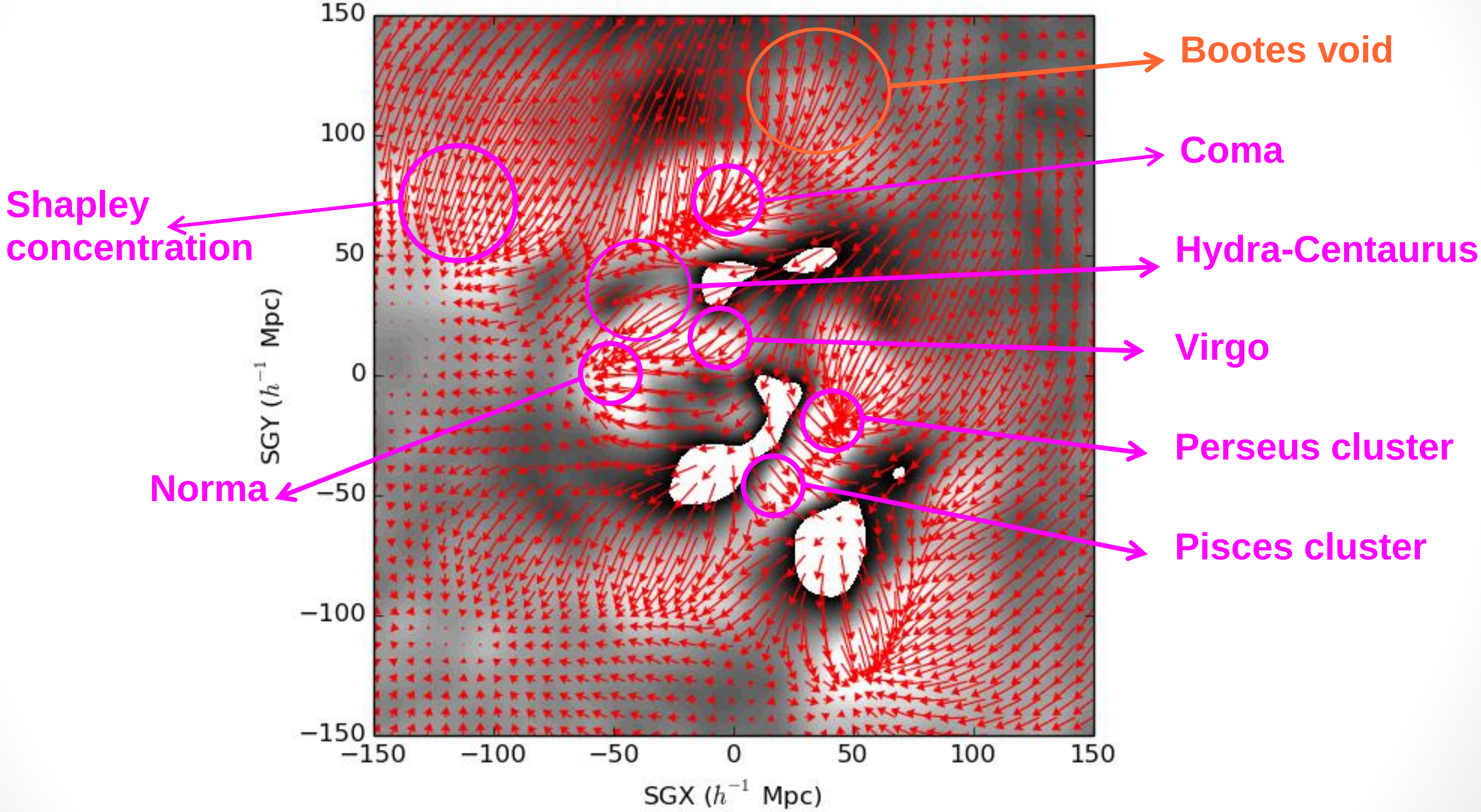
$$A_{ij} = \ln(\mathcal{P}(d|\delta_i)) - \ln(\mathcal{P}(d|\delta_j))$$

	ARES	HADES	BORG
ARES	0	-219580.31	-383482.25
HADES	219580.31	0	-163901.94
BORG	383482.25	163901.94	0.

VIRBIUS density and velocity fields

PRELIMINARY

Supergalactic plane



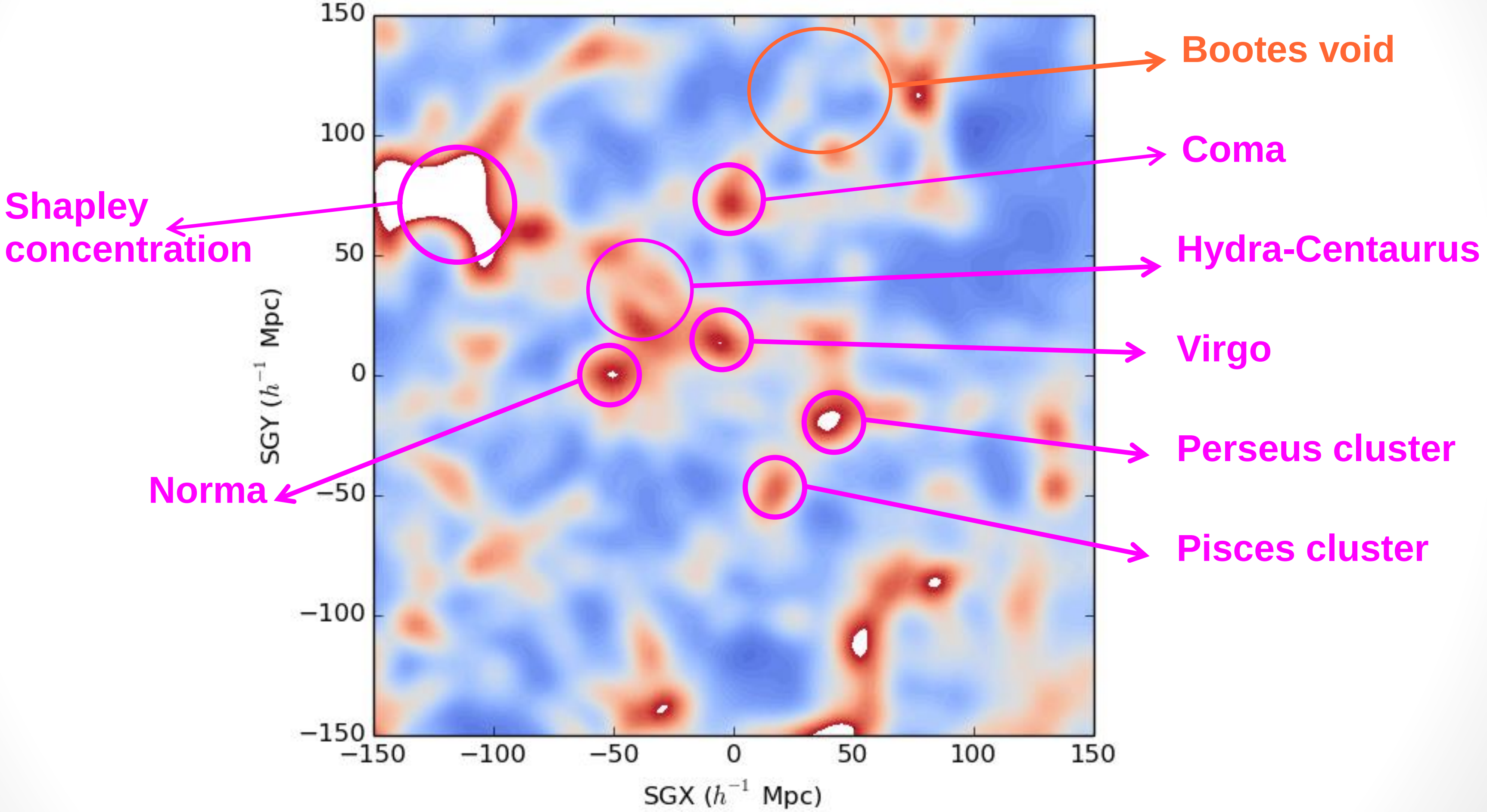
CosmicFlows 2.1, mean density and velocity field given data

Lavaux & Jasche, in prep.

BORG3 density field

PRELIMINARY

Supergalactic plane, final density field smoothed to 5 Mpc/h (Gaussian)



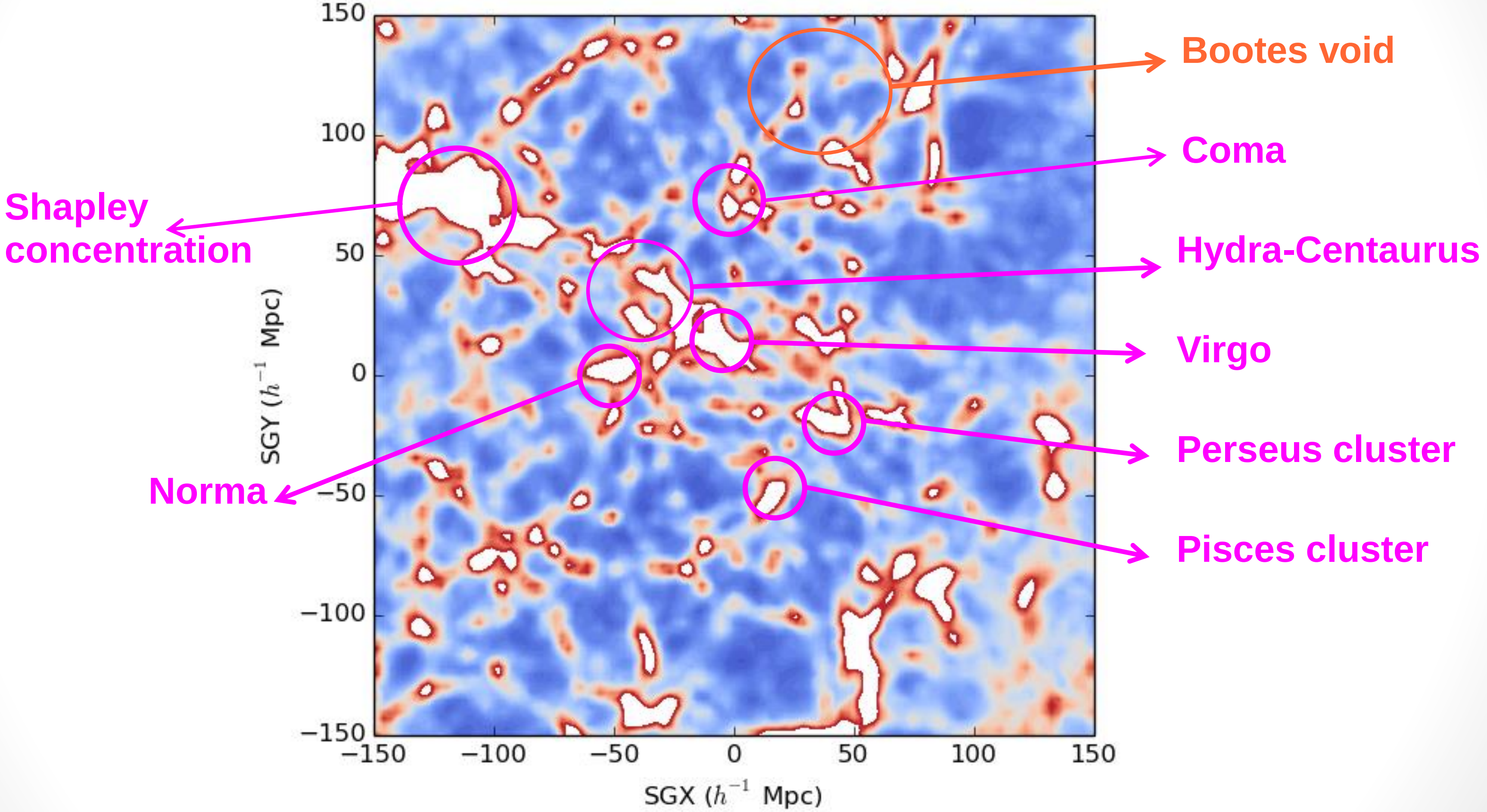
2M++, mean final matter density field

Lavaux & Jasche, in prep.

BORG3 density field

PRELIMINARY

Supergalactic plane, no smoothing



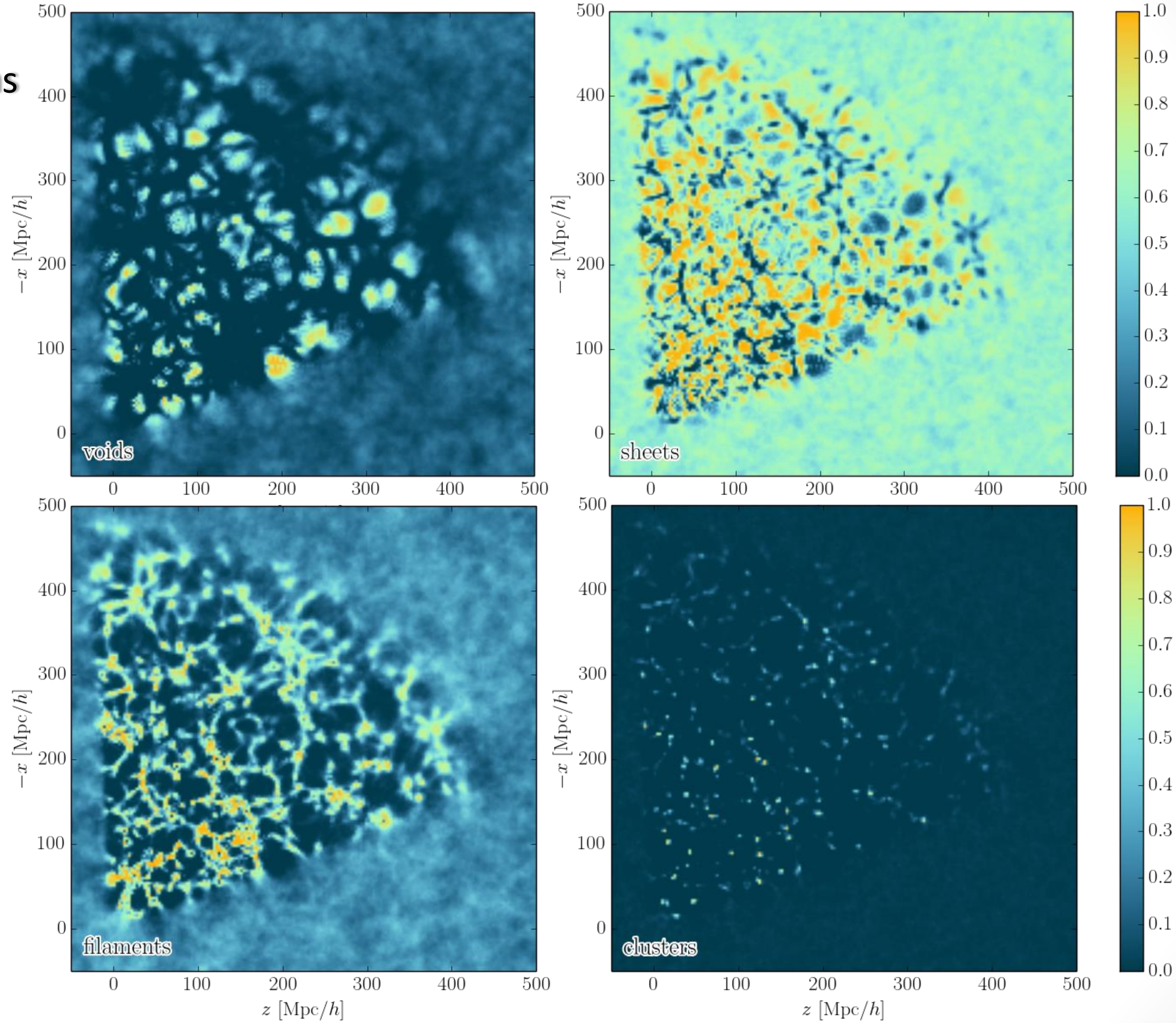
2M++, mean final matter density field

Lavaux & Jasche, in prep.

Cosmic Web Analysis

T-web structures inferred by BORG

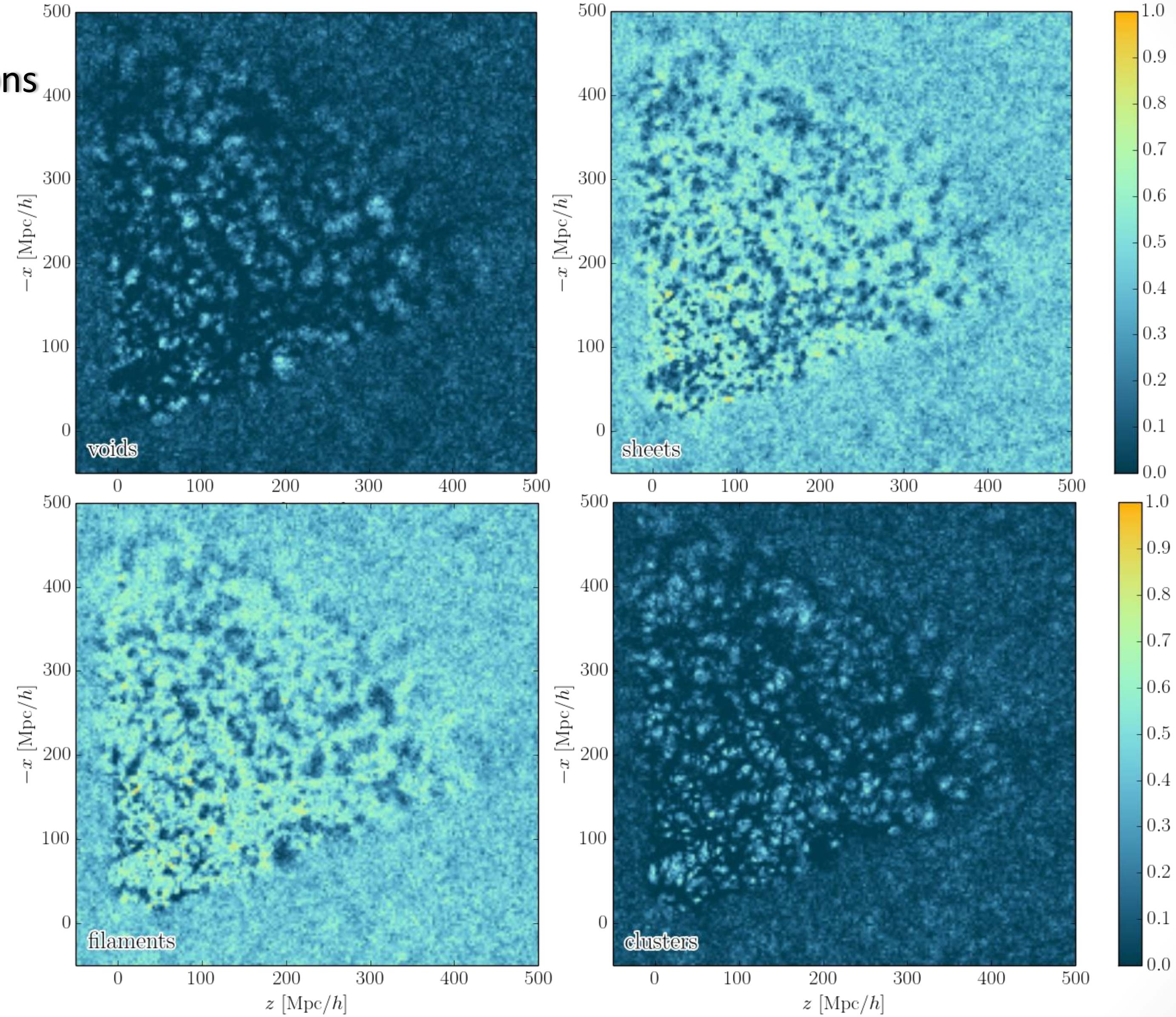
Final conditions



FL, Jasche & Wandelt 2015, arXiv:1502.02690

T-web structures inferred by BORG

Initial conditions



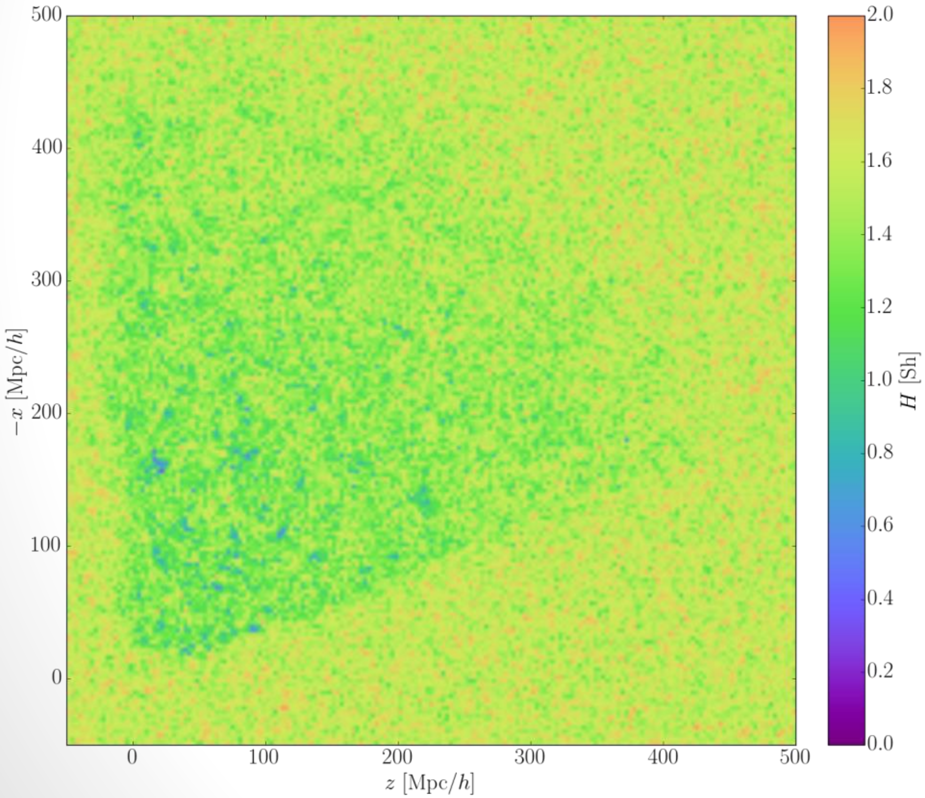
FL, Jasche & Wandelt 2015, arXiv:1502.02690

What is the information content of these maps?

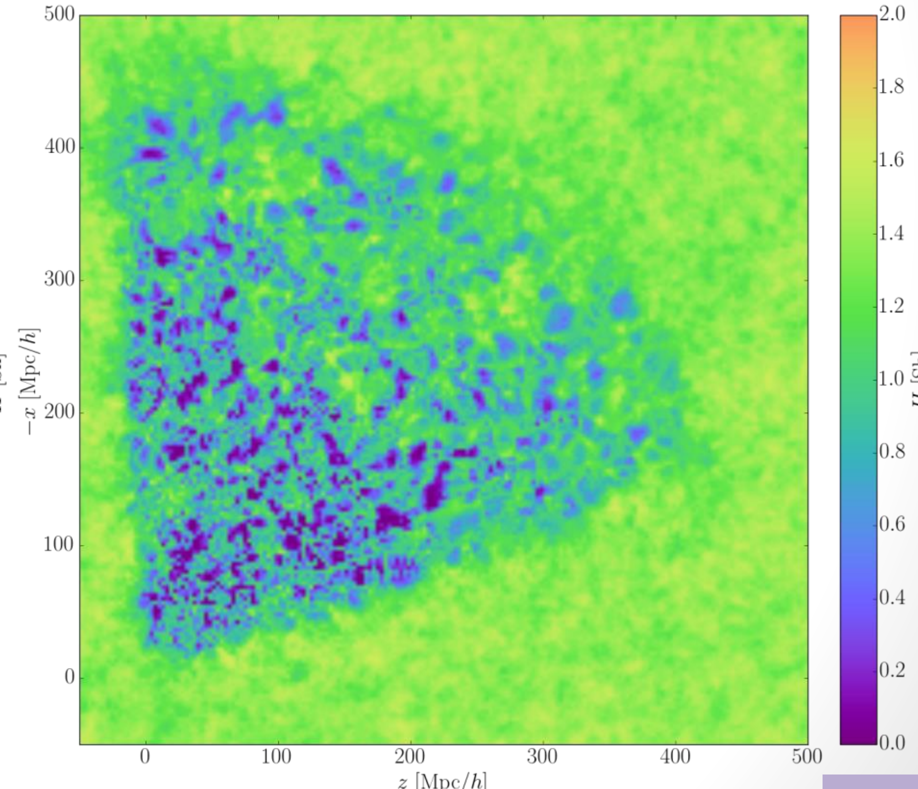
Shannon entropy

$$H\left[\mathcal{P}(\mathrm{T}(\vec{x}_k)|d)\right] \equiv -\sum_{i=0}^{3} \mathcal{P}(\mathrm{T}_i(\vec{x}_k)|d) \log_2(\mathcal{P}(\mathrm{T}_i(\vec{x}_k)|d))$$ in shannons (Sh)

Initial conditions



Final conditions

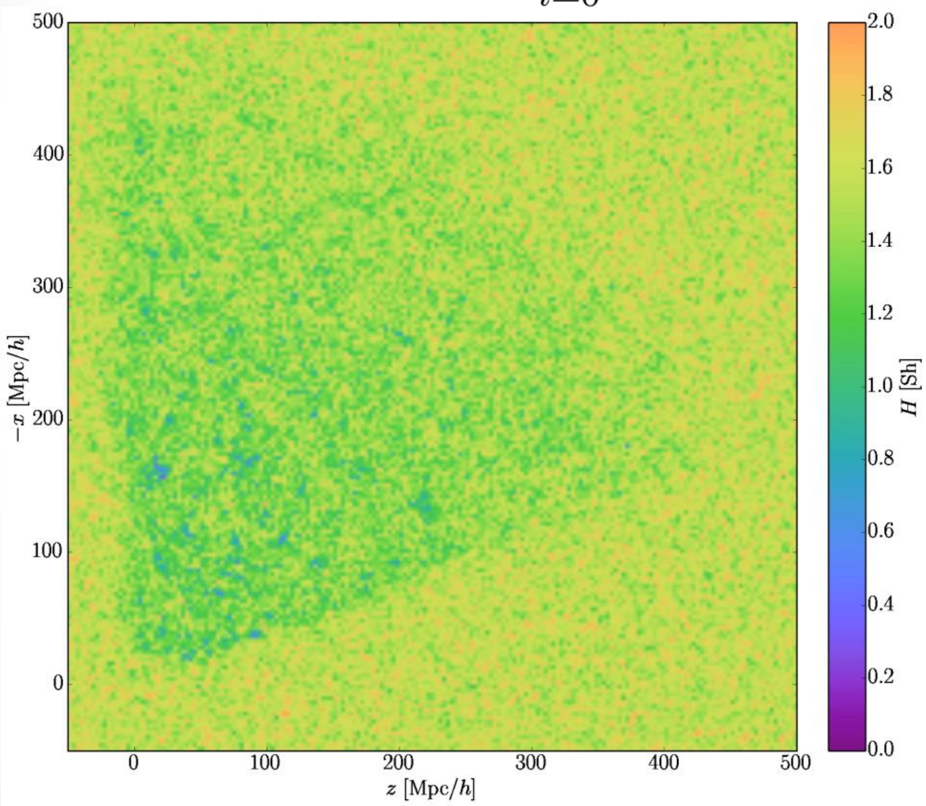


FL, Jasche & Wandelt 2015, arXiv:1502.02690

How is information propagated?

Shannon entropy

$$H\left[\mathcal{P}(\mathrm{T}(\vec{x}_k)|d)\right] \equiv -\sum_{i=0}^{3} \mathcal{P}(\mathrm{T}_i(\vec{x}_k)|d)\log_2(\mathcal{P}(\mathrm{T}_i(\vec{x}_k)|d))$$ in shannons (Sh)



More about cosmic web analysis:

FL, Jasche & Wandelt 2015, arXiv:1502.02690
(T-web, entropy, relative entropy)
FL, Jasche & Wandelt 2015, arXiv:1503.00730
(decision theory for structure classification)
FL, Jasche, Lavaux & Wandelt 2016, arXiv:1601.00093
(classifications with DIVA & ORIGAMI)
FL, Lavaux, Jasche & Wandelt 2016, arXiv:1606.06758
(mutual information, classifier utilities)

FL, Jasche & Wandelt 2015, arXiv:1502.02690

Concluding thoughts

- **Bayesian large-scale structure inference** is not an impossible problem!
- A new method for principled analysis of galaxy surveys:
  - Uncertainty quantification (noise, survey geometry, selection effects and biases)
  - Various degrees of physical modeling
  - Non-linear and non-Gaussian inference, with improving techniques
- Bayesian large-scale structure inference has moved beyond the proof-of-concept stage to **routine applications to real data**.
- These techniques allow **full and statistically accurate** inference from galaxy surveys, which paves the path toward **joint BAO/RSD** analyses.

Epilogue: uncertainty quantification in LSS maps