Bayesian analyses of galaxy surveys

Florent Leclercq
www.florent-leclercq.eu

Imperial Centre for Inference and Cosmology
Imperial College London

Wolfgang Enzi, Alan Heavens, Andrew Jaffe,
Jens Jasche, George Kyriacou, Guilhem Lavaux,
Arrykrishna Mootoovaloo, Will Percival, Benjamin Wandelt

and the Aquila Consortium
www.aquila-consortium.org

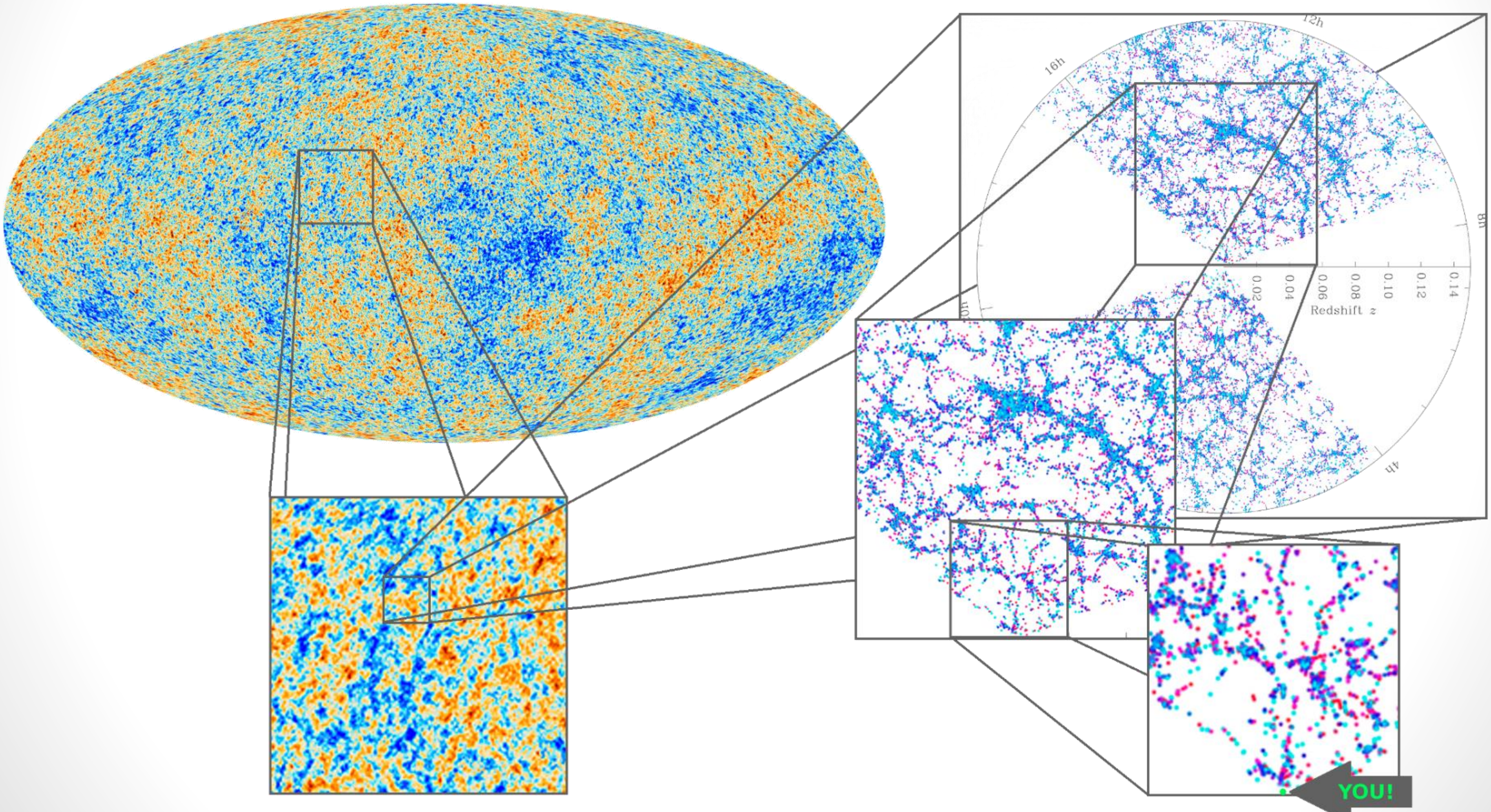
July 8th, 2019

ICIC Imperial Centre for Inference & Cosmology

Imperial College London

The big picture: the Universe is highly structured

You are here. Make the best of it…



Planck collaboration (2013-2015)

M. Blanton and the Sloan Digital Sky Survey (2010-2013)

What we want to know from the large-scale structure

The LSS is a vast source of knowledge:

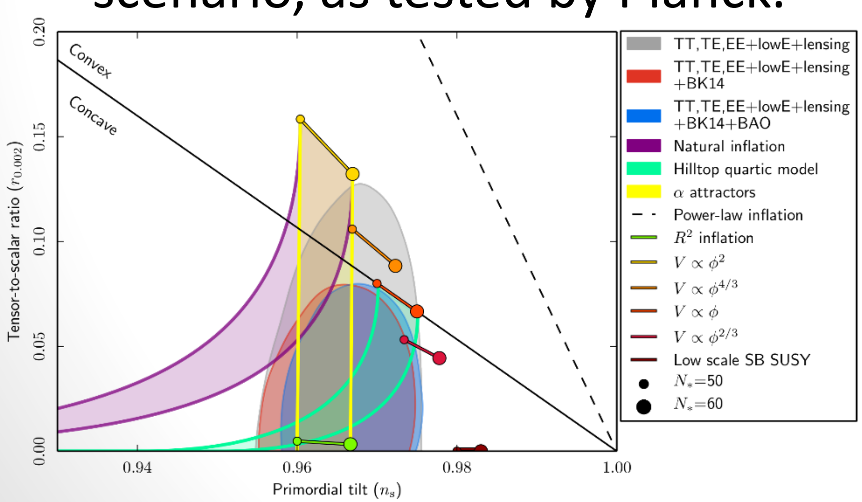
- Cosmology:
  - ΛCDM : cosmological parameters and tests against alternatives,
  - Physical nature of the dark components,
  - Neutrinos : number and masses,
  - Geometry of the Universe,
  - Tests of General Relativity,
  - Initial conditions and link to high energy physics
- Astrophysics: galaxy formation and evolution as a function of their environment
  - Galaxy properties (colours, chemical composition, shapes),
  - Intrinsic alignments, intrinsic size-magnitude correlations

We have theoretical and computer models…

- **Initial conditions:**
  a Gaussian random field

$$\mathcal{P}(\delta^{\mathrm{i}}|S) = \frac{1}{\sqrt{|2\pi S|}} \exp\left(-\frac{1}{2}\sum_{x,x'} \delta^{\mathrm{i}}_x S^{-1}_{xx'} \delta^{\mathrm{i}}_{x'}\right)$$

Everything seems consistent with the simplest inflationary scenario, as tested by Planck.

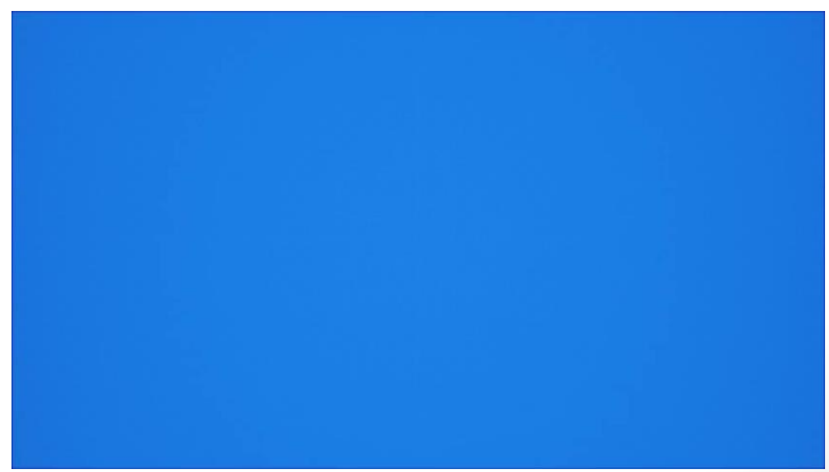


Planck 2018 X, 1807.06211

- **Structure formation:**
  numerical solution of the Vlasov-Poisson system for dark matter dynamics

$$\frac{\partial f}{\partial \tau} + \frac{\mathbf{p}}{ma} \cdot \nabla f - ma\nabla\Phi \cdot \frac{\partial f}{\partial \mathbf{p}} = 0$$

$$\Delta\Phi = 4\pi \mathrm{G} a^2 \bar{\rho}\delta$$



Y. Dubois & S. Colombi (IAP)

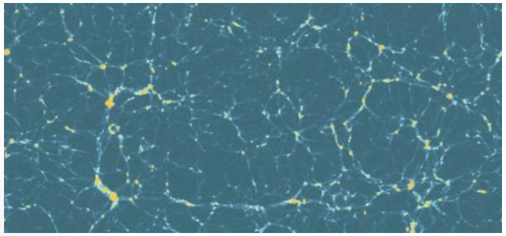
… how do we test these models against survey data?

J. Cham – PhD comics

Redshift range	Volume (Gpc^3)	k_{max} $(Mpc/h)^{-1}$	N_{modes}
0-1	50	0.15	10^7
1-2	140	0.5	$5x10^8$
2-3	160	1.3	10^{10}

M. Zaldarriaga

- Precise tests require many modes.
- In 3D galaxy surveys, the number of modes usable scales as k_{max}^3.



- The challenge: non-linear evolution at small scales and late times.
- The strategy:
  - Pushing down the smallest scale usable for cosmological analysis
  - Using a numerical model linking initial and final conditions

In other words: go beyond the linear and static analysis of the LSS.

Why Bayesian inference?

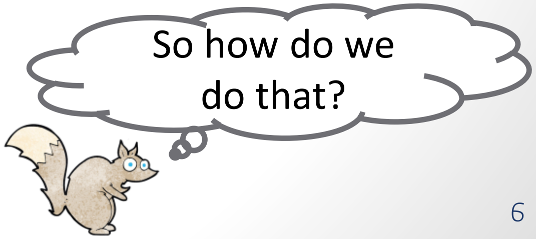
- Inference of signals = ill-posed problem
  - Incomplete observations: finite resolution, survey geometry, selection effects
  - Noise, biases, systematic effects
  - Cosmic variance

➡ No unique recovery is possible!

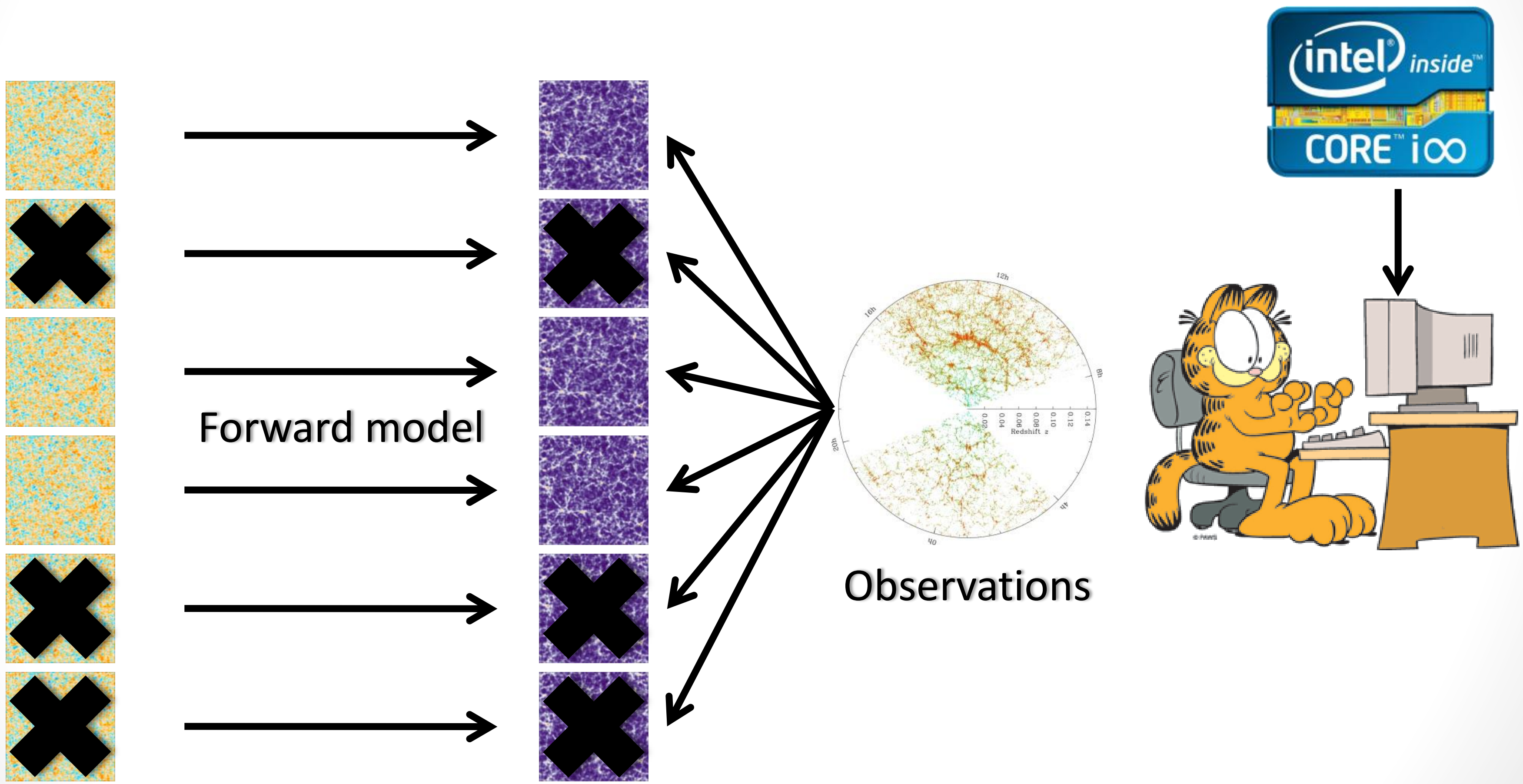
"What is the formation history of the Universe?" ➡ "What is the probability distribution of possible formation histories (signals) compatible with the observations?"

Bayes' theorem: $$\mathcal{P}(s|d)\mathcal{P}(d) = \mathcal{P}(d|s)\mathcal{P}(s)$$

- Cox-Jaynes theorem: Any system to manipulate "*plausibilities*", consistent with Cox's desiderata, is isomorphic to (Bayesian) probability theory



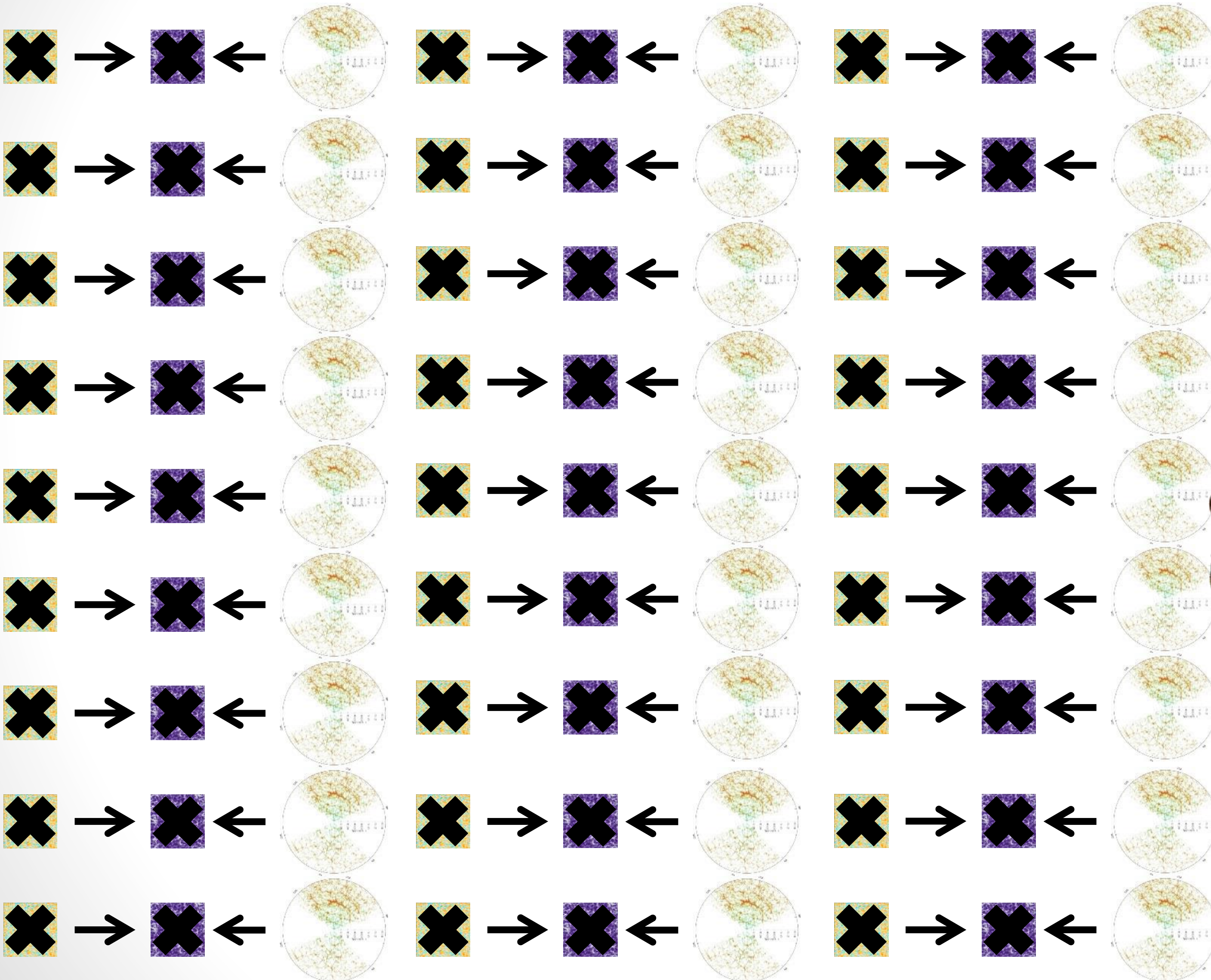
Bayesian forward modelling: the ideal scenario



All possible ICs

All possible FCs

Bayesian forward modelling: the challenge

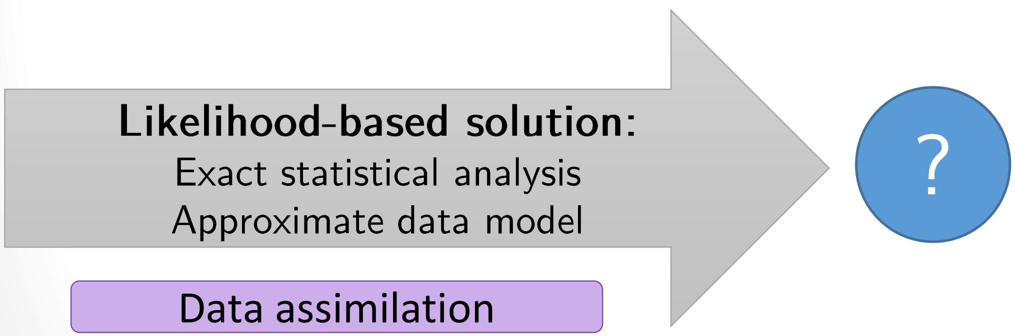


The (true) likelihood lives in

$d \approx 10^7$!

Likelihood-based solution: BORG

Bayesian Origin Reconstruction from Galaxies



Hamiltonian (Hybrid) Monte Carlo

- Use classical mechanics to solve statistical problems!
  - The potential: $\psi(\mathbf{x}) \equiv -\ln p(\mathbf{x})$
  - The Hamiltonian: $H(\mathbf{x},\mathbf{p}) \equiv \frac{1}{2}\mathbf{p}^\intercal \mathbf{M}^{-1}\mathbf{p} + \psi(\mathbf{x})$

$$(\mathbf{x},\mathbf{p}) \Rightarrow \left\{ \begin{aligned} \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} &= \frac{\partial H}{\partial \mathbf{p}} = \mathbf{M}^{-1}\mathbf{p} \\ \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} &= -\frac{\partial H}{\partial \mathbf{x}} = -\frac{\mathrm{d}\psi(\mathbf{x})}{\mathrm{d}\mathbf{x}} \end{aligned} \right\} \Rightarrow (\mathbf{x}',\mathbf{p}')$$

gradients of the pdf

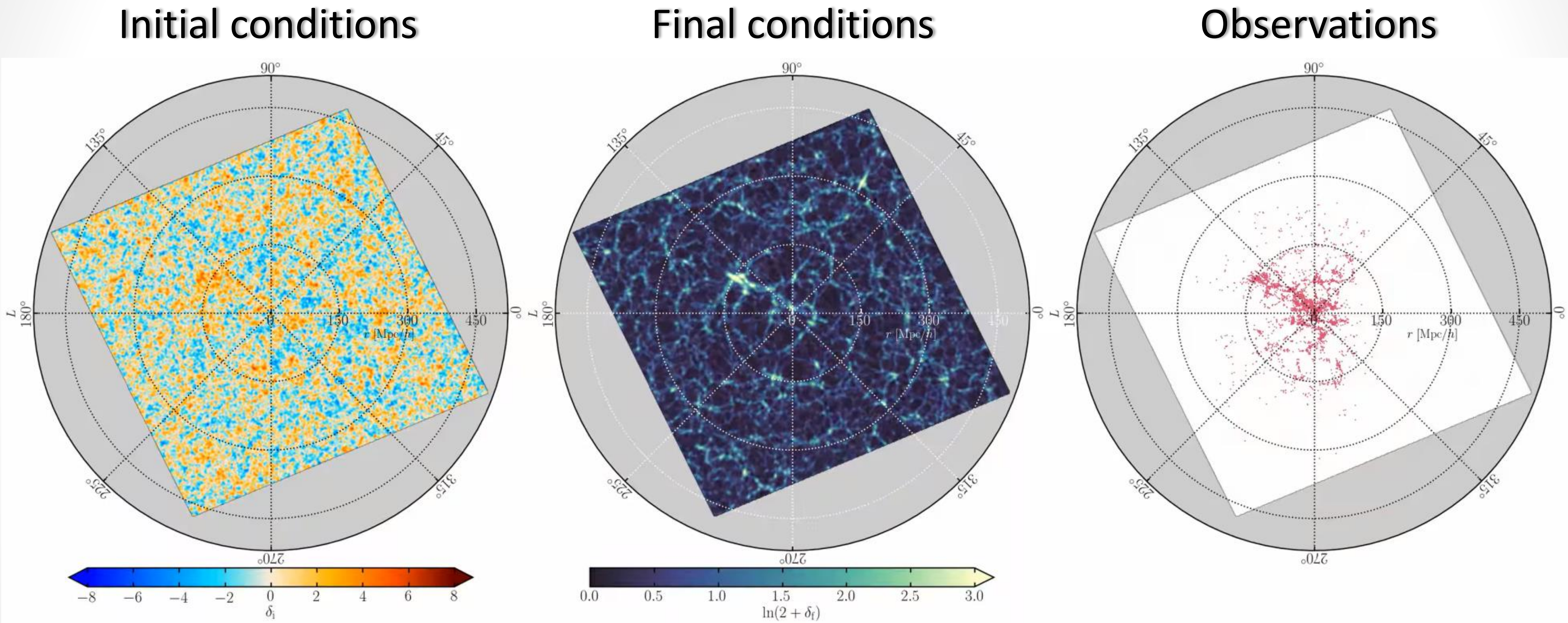
$$a(\mathbf{x}',\mathbf{x}) = \mathrm{e}^{-(H'-H)} = 1$$

acceptance ratio unity

- HMC beats the curse of dimensionality by:
  - Exploiting gradients
  - Using conservation of the Hamiltonian

Duane et al. 1987, Phys. Lett. B 195, 2

BORG at work: Bayesian chrono-cosmography

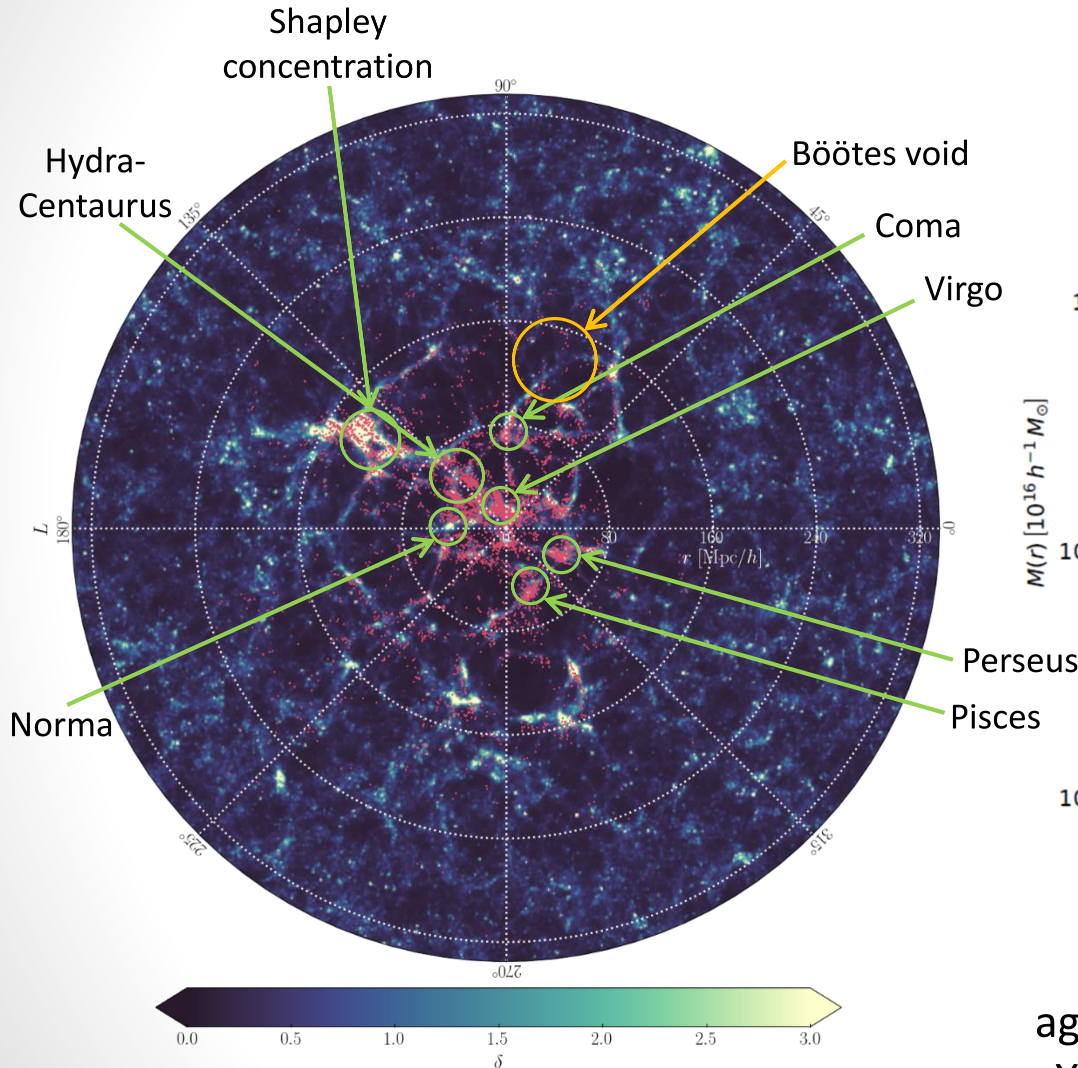


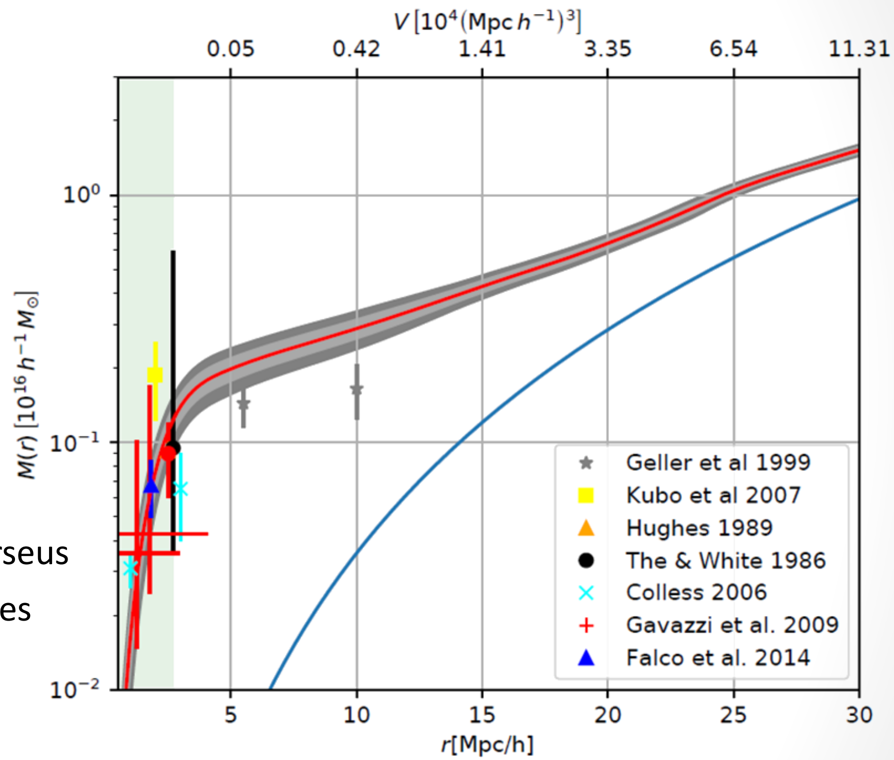
Supergalactic plane

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 & Lavaux 2019, 1806.11117 – FL, Lavaux & Jasche, in prep.

BORGPM density field: full non-linear dynamics



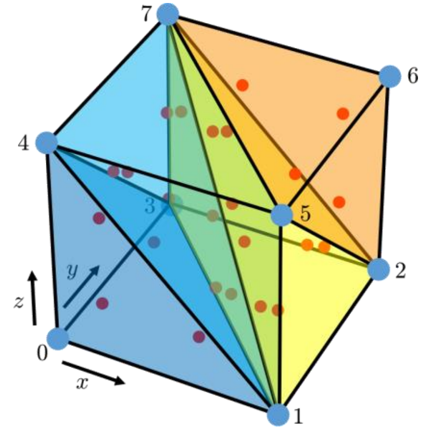


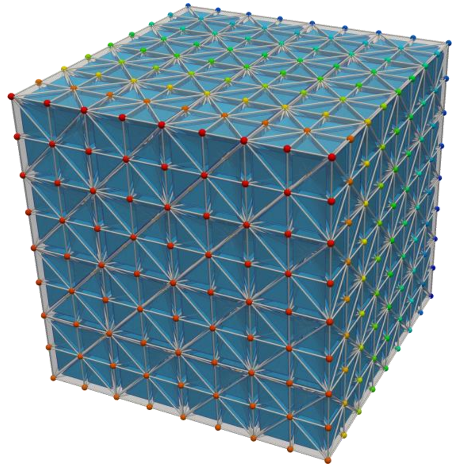
Mass profile of the Coma cluster, in agreement with gravitational lensing and X-ray observations down to a few Mpc.

Jasche & Lavaux 2019, 1806.11117 – FL, Lavaux & Jasche, in prep.

The phase-space structure of dark matter: tools

Delaunay tessellation of elementary Lagrangian cubes



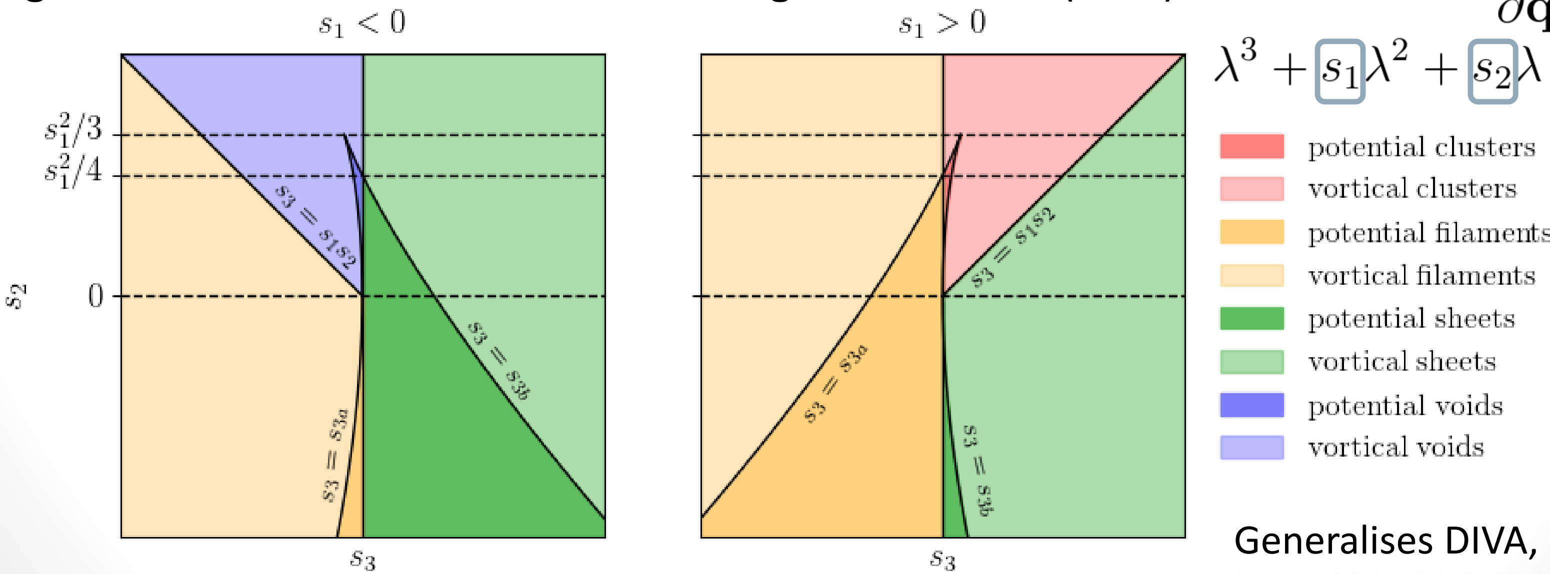


Abel, Hahn & Kaehler 2012, 1111.3944
Shandarin, Habib & Heitmann 2012, 1111.2366
Hahn, Abel & Kaehler 2013, 1210.6652
Hahn & Angulo 2016, 1501.01959
Sousbie & Colombi 2016, 1509.07720

Lagrangian Invariants Classification of Heterogeneous flows (LICH)

$$\mathcal{R}_{\ell m} \equiv \frac{\partial \Psi_\ell}{\partial \mathbf{q}_m}$$

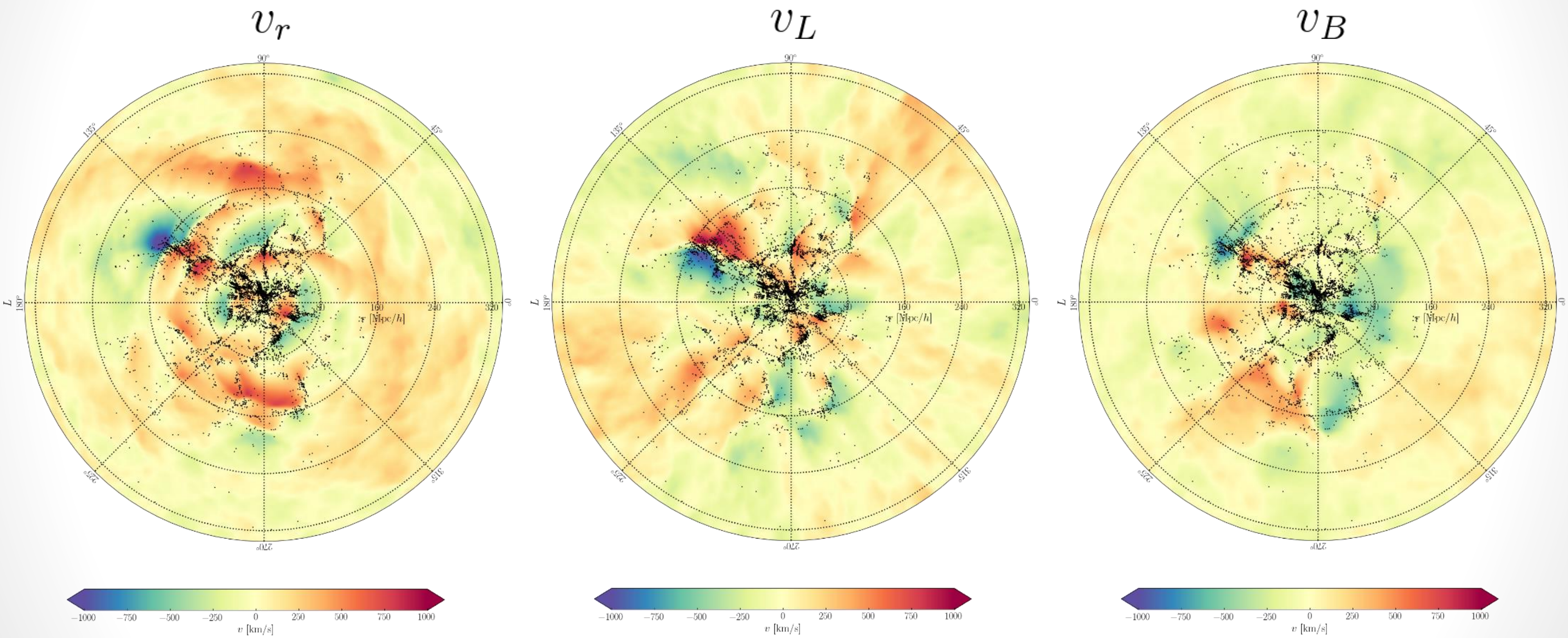
$$\lambda^3 + s_1 \lambda^2 + s_2 \lambda + s_3 = 0$$



FL, Jasche, Lavaux, Wandelt & Percival 2017, 1601.00093

Generalises DIVA,
Lavaux & Wandelt 2010, 0906.4101

Velocity field in the supergalactic plane

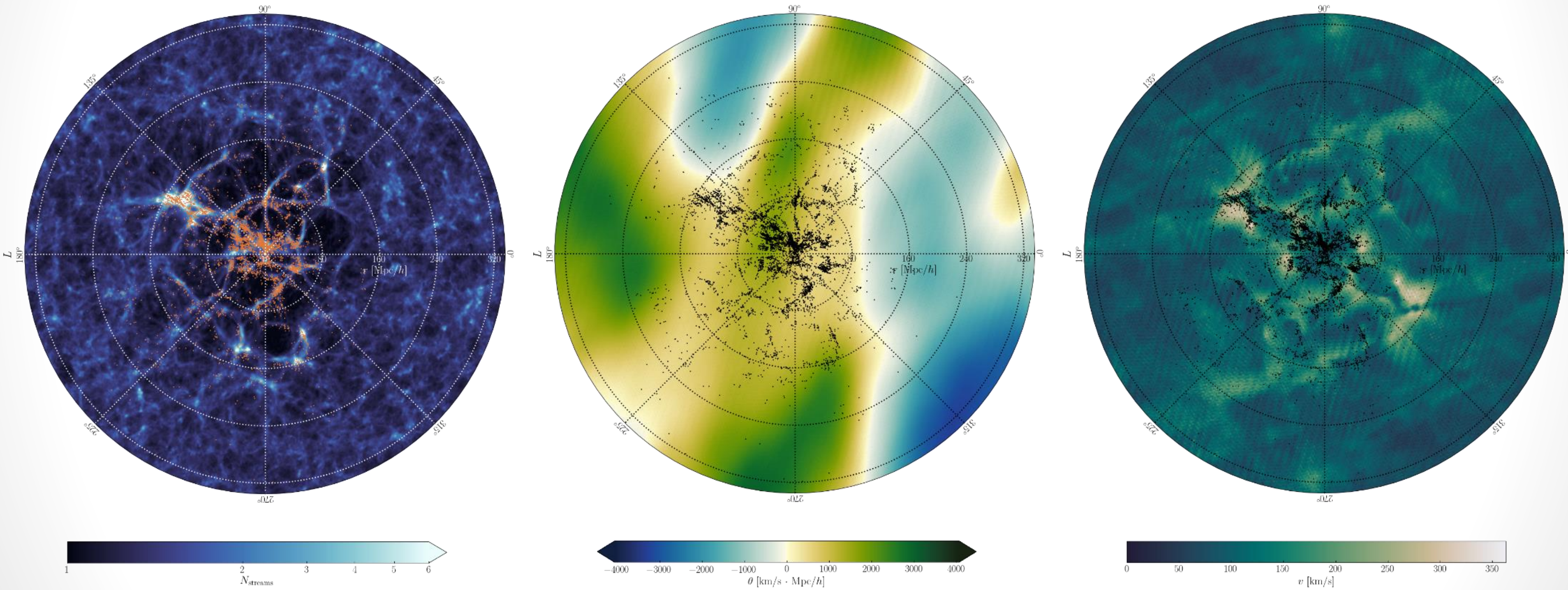


The **gravitational infall** of known structures can be observed.

FL, Lavaux & Jasche, in prep.

Number of streams and vorticity

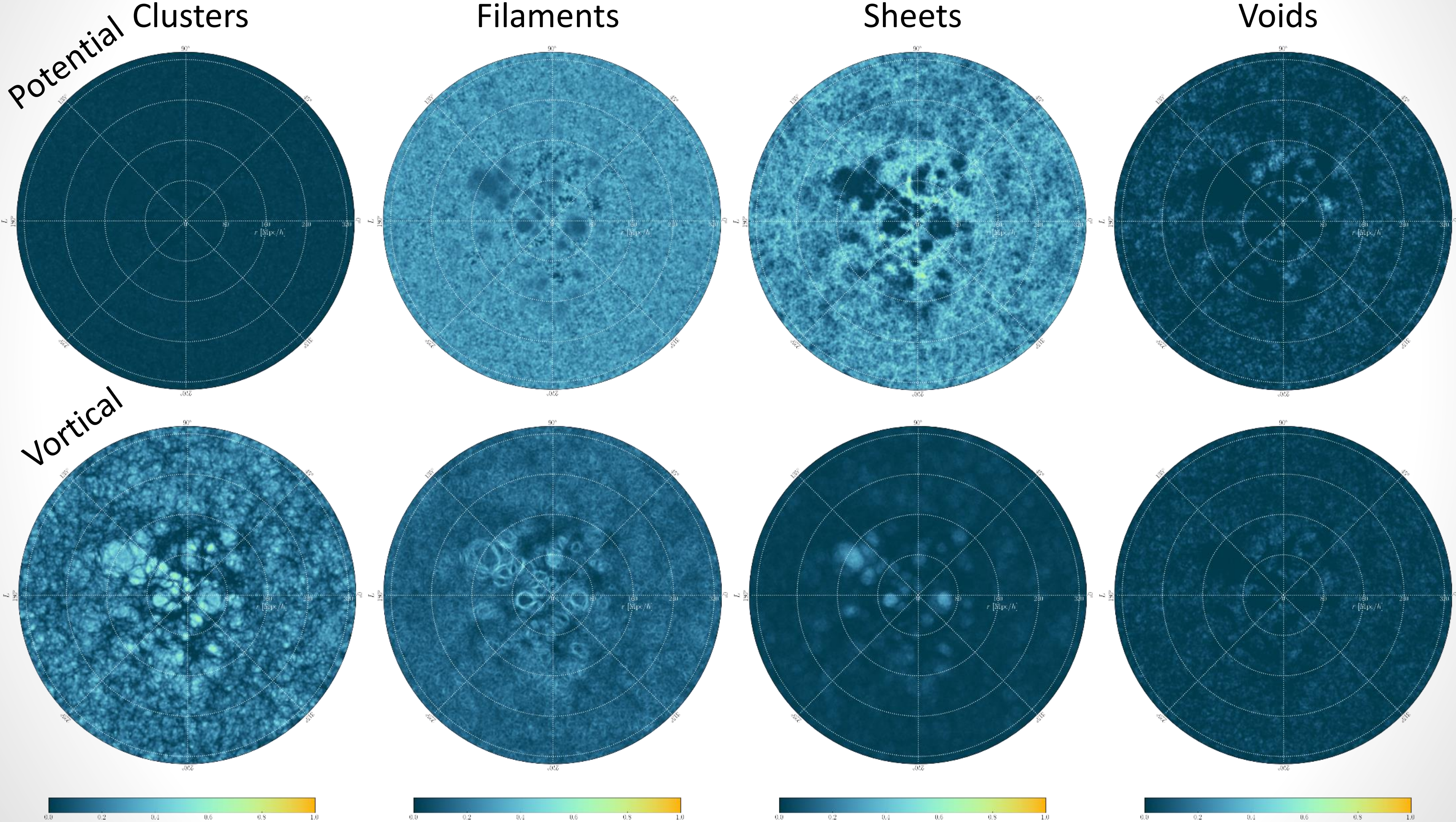
Number of streams Velocity potential Norm of vorticity



In earlier work (Leclercq, Jasche, Lavaux, Wandelt & Percival 2017, 1601.00093), these were postdictions. Thanks to BORGPM (full non-linear dynamics), we have now actual measurements - with uncertainties.

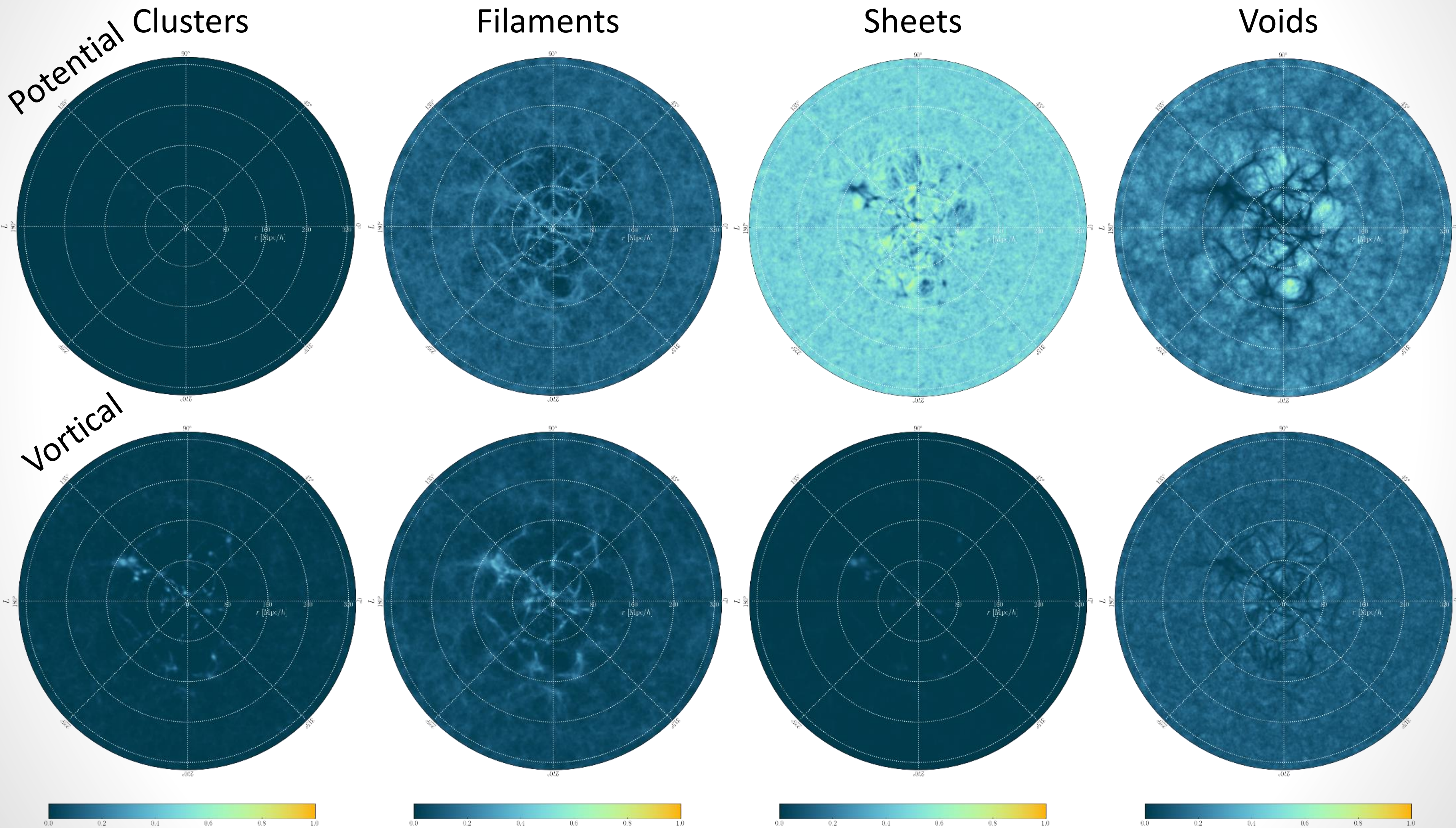
FL, Lavaux & Jasche, in prep.

LICH initial structures inferred by BORG



FL, Lavaux & Jasche, in prep.

LICH final structures inferred by BORG



FL, Lavaux & Jasche, in prep.

Mapping the Universe: epilogue?

J. Cham – PhD comics

Likelihood-free solution: SELFI

Simulator Expansion for Likelihood-Free Inference

Likelihood-based solution:
Exact statistical analysis
Approximate data model

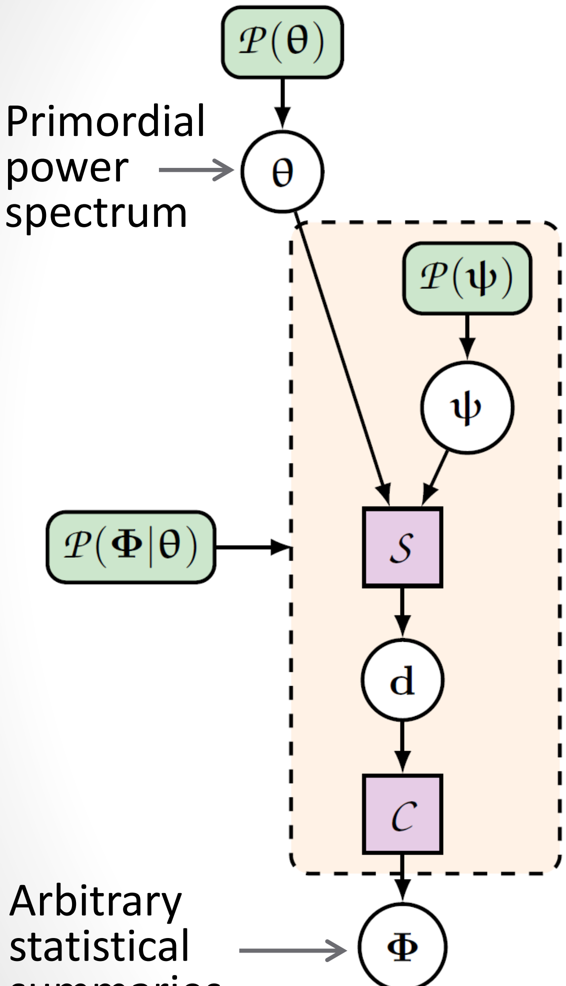
Data assimilation

?

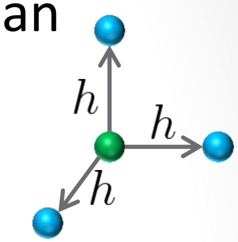
Likelihood-free solution:
Approximate statistical analysis
Arbitrary data model

Generative inference

SELFI: Method



- Gaussian prior + Gaussian effective likelihood
- Linearisation of the black-box around an expansion point + finite differences:



$$\hat{\Phi}_\theta \approx \mathbf{f}_0 + \nabla \mathbf{f}_0 \cdot (\theta - \theta_0)$$

➡ The posterior is Gaussian and analogous to a Wiener filter:

expansion point → θ_0; observed summaries → Φ_O

$$\gamma \equiv \theta_0 + \Gamma\,(\nabla \mathbf{f}_0)^\intercal\, \mathbf{C}_0^{-1}(\Phi_O - \mathbf{f}_0)$$

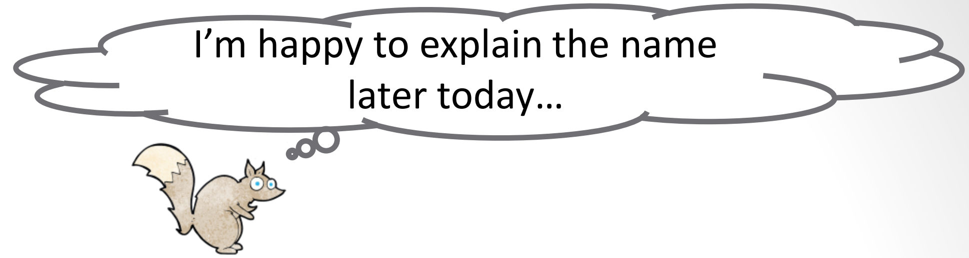
$$\Gamma \equiv \left[(\nabla \mathbf{f}_0)^\intercal\, \mathbf{C}_0^{-1} \nabla \mathbf{f}_0 + \mathbf{S}^{-1}\right]^{-1}$$

covariance of summaries → $\mathbf{C}_0$; gradient of the black-box → $\nabla \mathbf{f}_0$; prior covariance → $\mathbf{S}$

$\mathbf{f}_0, \mathbf{C}_0$ and $\nabla \mathbf{f}_0$ can be evaluated through simulations only.
The number of required simulations is fixed *a priori*.

FL, Enzi, Jasche & Heavens 2019, 1902.10149

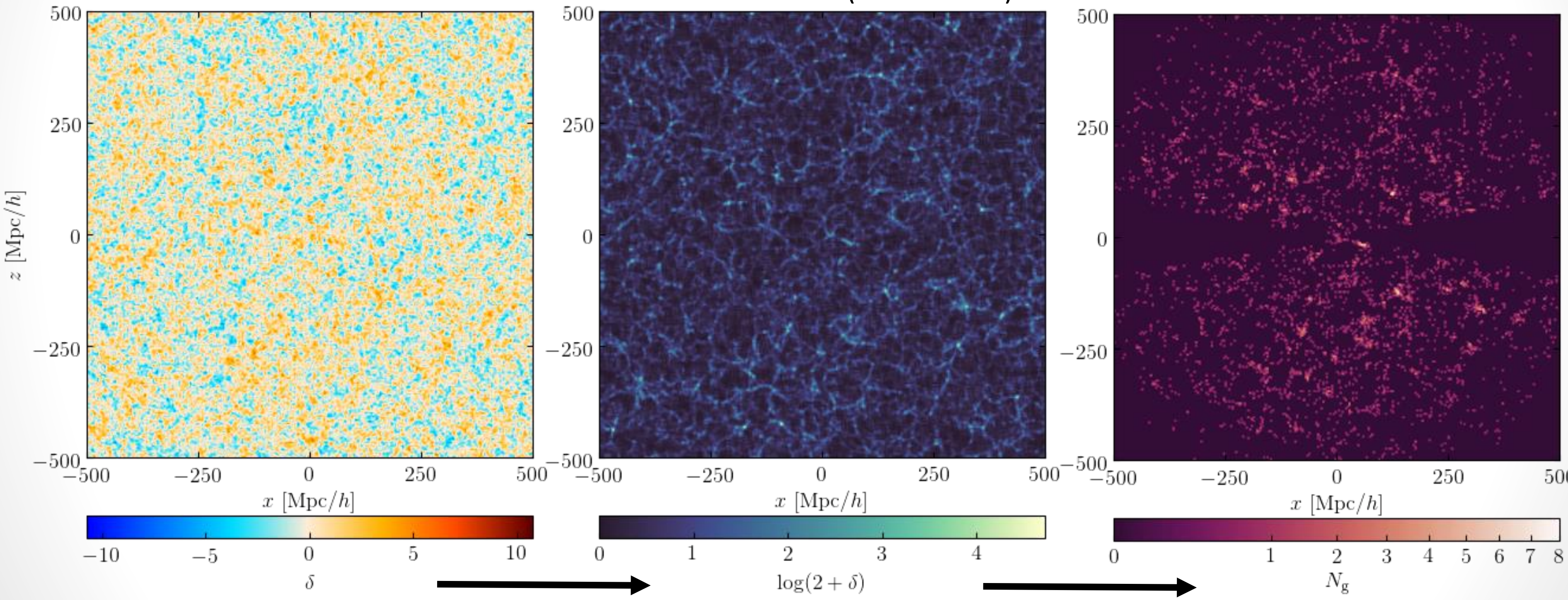
A black-box: Simbelmynë



Publicly available code:

https://bitbucket.org/florent-leclercq/simbelmyne/

Initial conditions | Final conditions (dark matter) | Galaxies

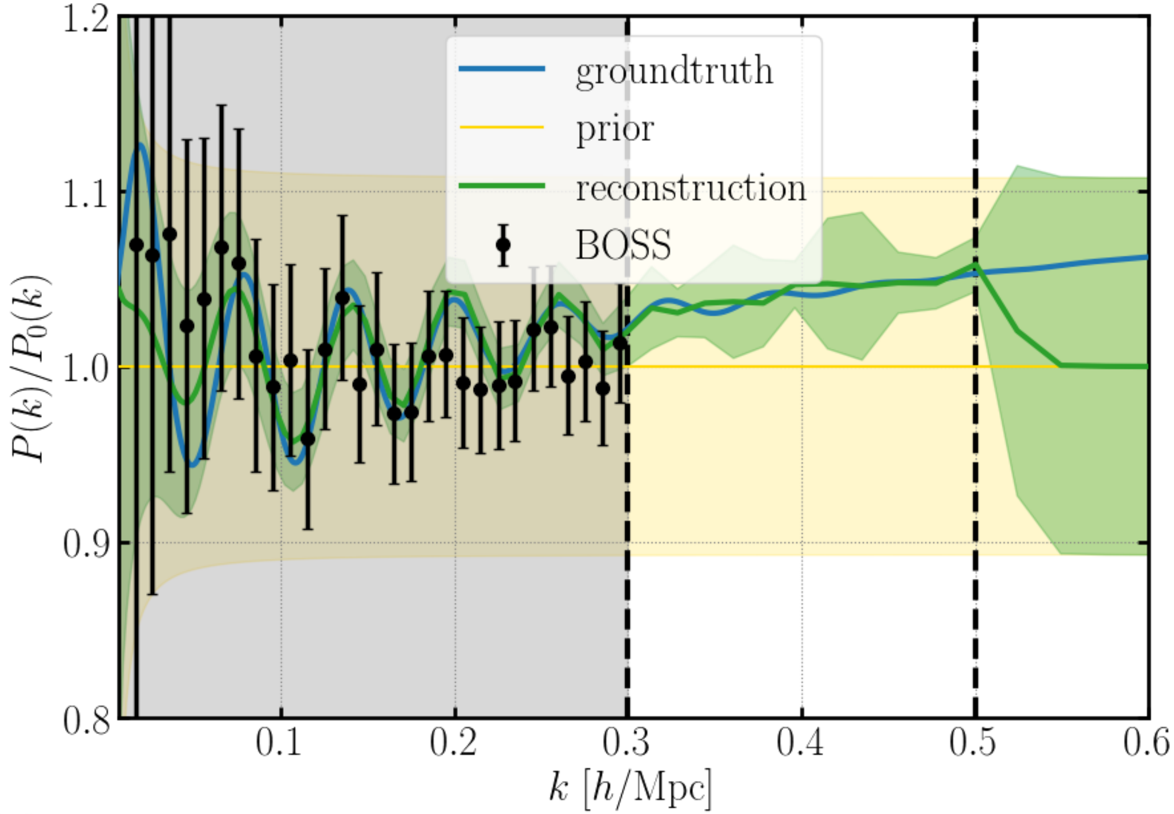


Dark matter simulation with COLA

Tassev, Zaldarriaga & Eisenstein 2013, 1301.0322

Survey simulation:
Redshift-space distortions, galaxy bias, selection effects, survey geometry, instrumental noise

SELFI + Simbelmynë: Proof-of-concept



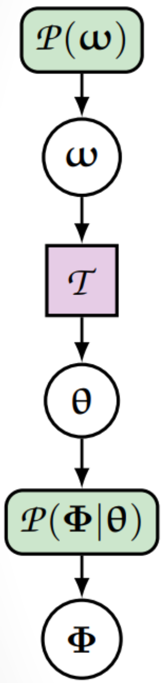
100 parameters are simultaneously inferred from a black-box data model

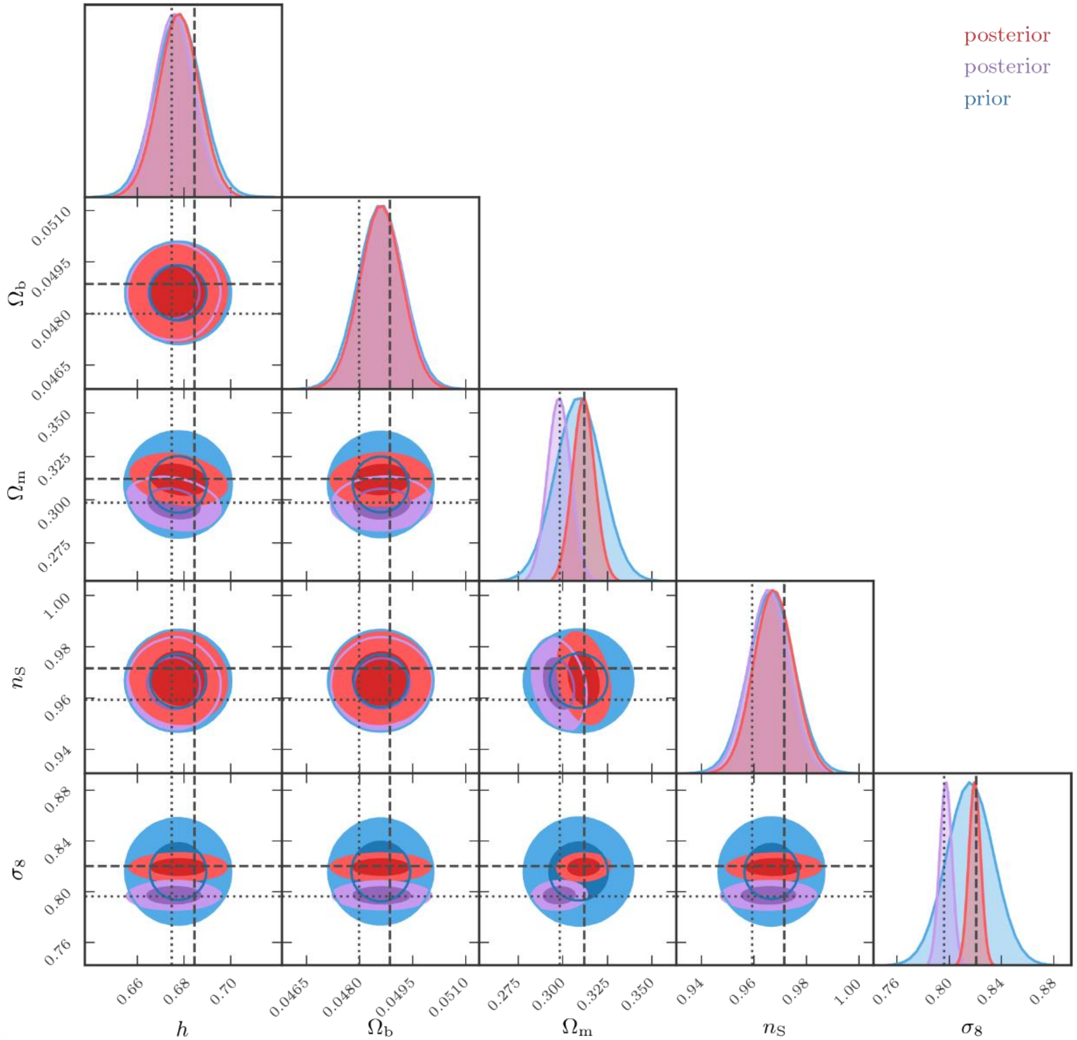
$N_{\mathrm{modes}} \propto k^3$: **5** times more modes are used in the analysis

1 (Gpc/h)^3 only! Much more potential for Euclid data…

FL, Enzi, Jasche & Heavens 2019, 1902.10149

SELFI + Simbelmynë: Proof-of-concept

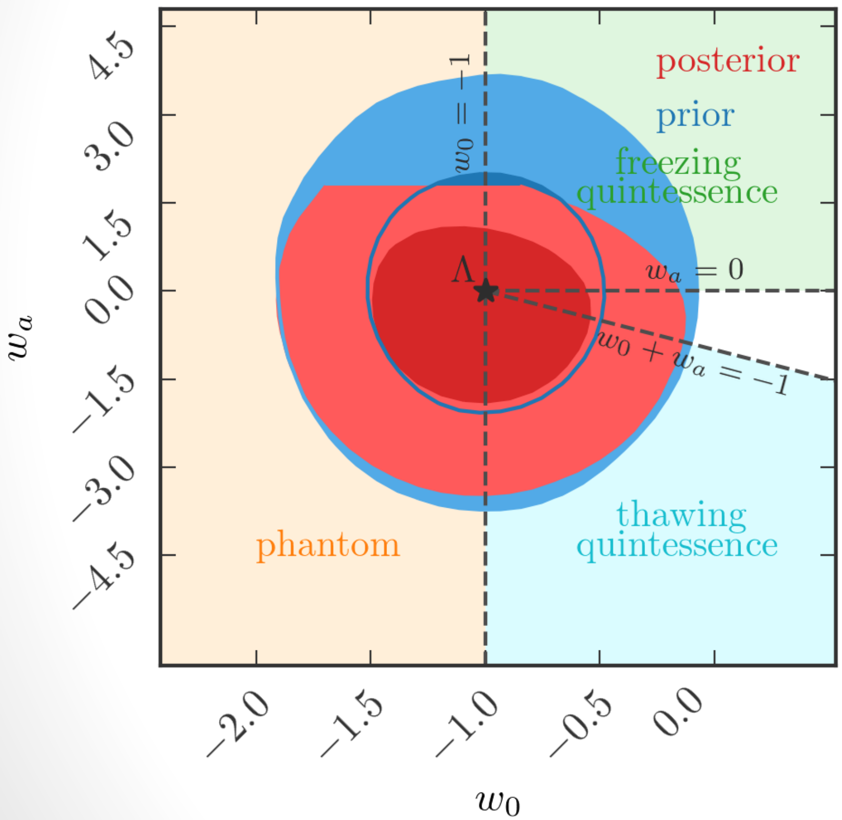


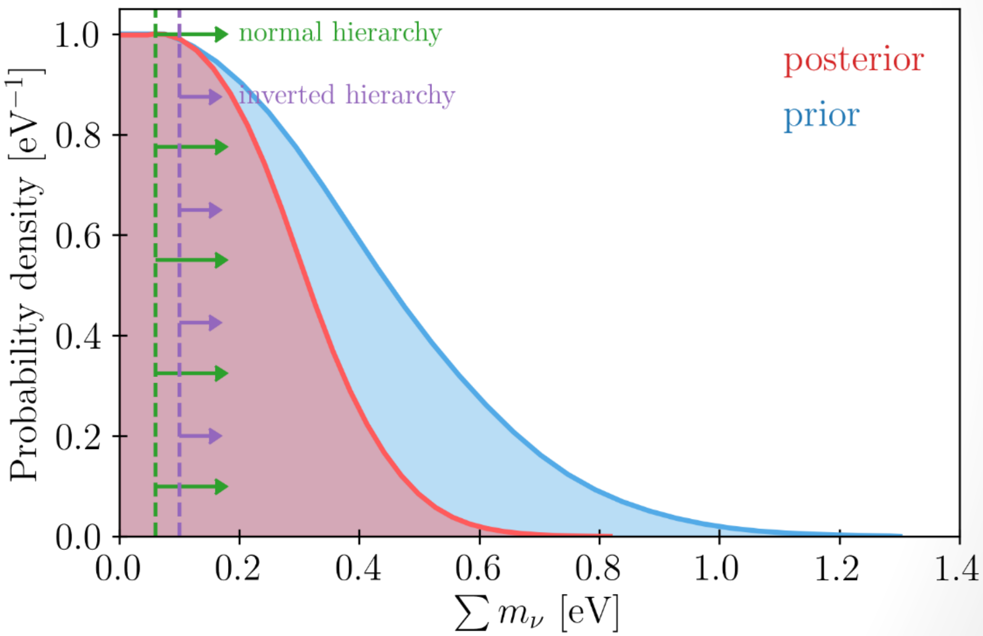


- Robust inference of cosmological parameters can be easily performed *a posteriori* once the linearised data model is learnt

FL, Enzi, Jasche & Heavens 2019, 1902.10149

Dark energy and neutrino masses with SELFI



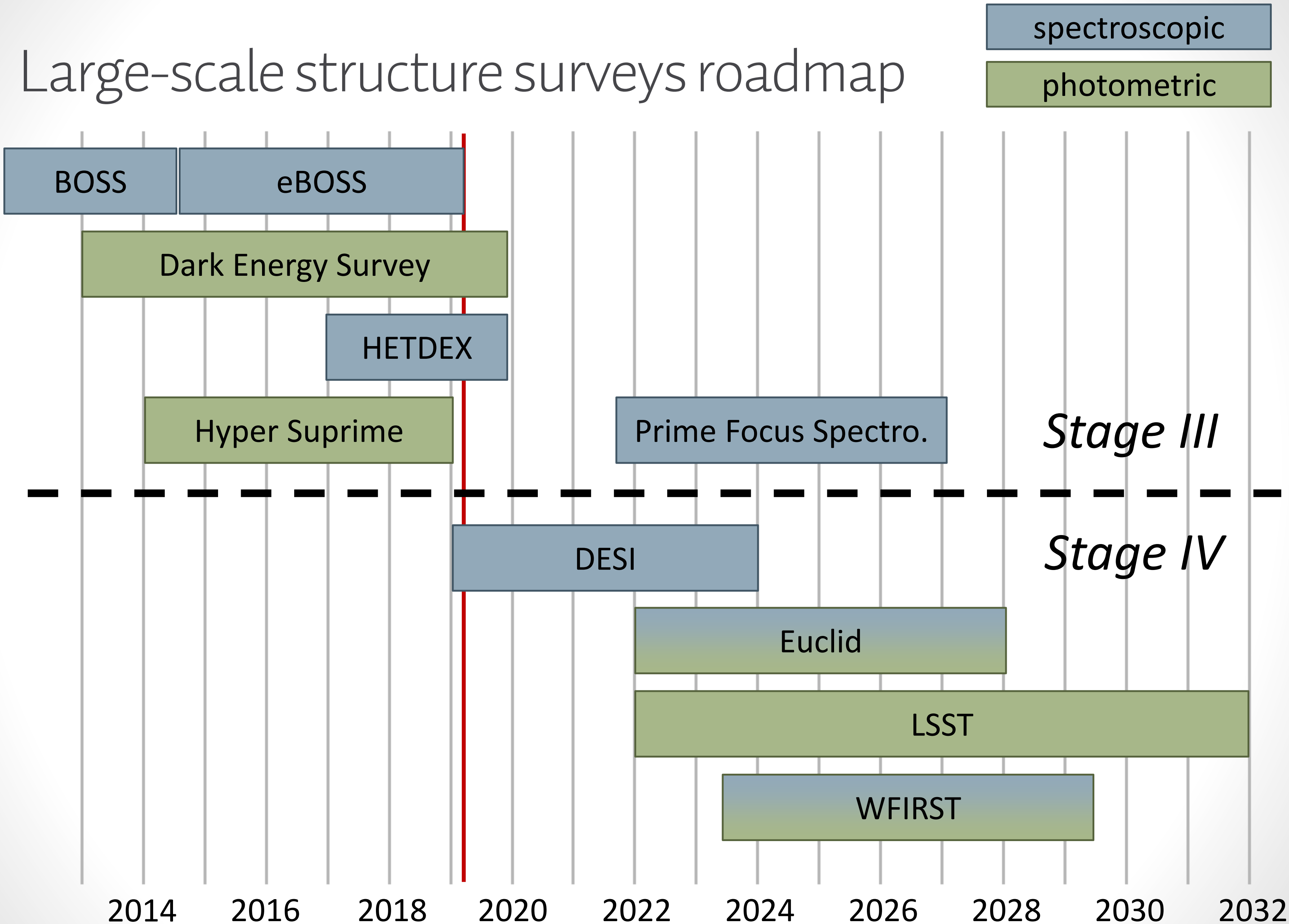


pyselfi will be made publicly available soon.

The Future: Opportunities & Challenges

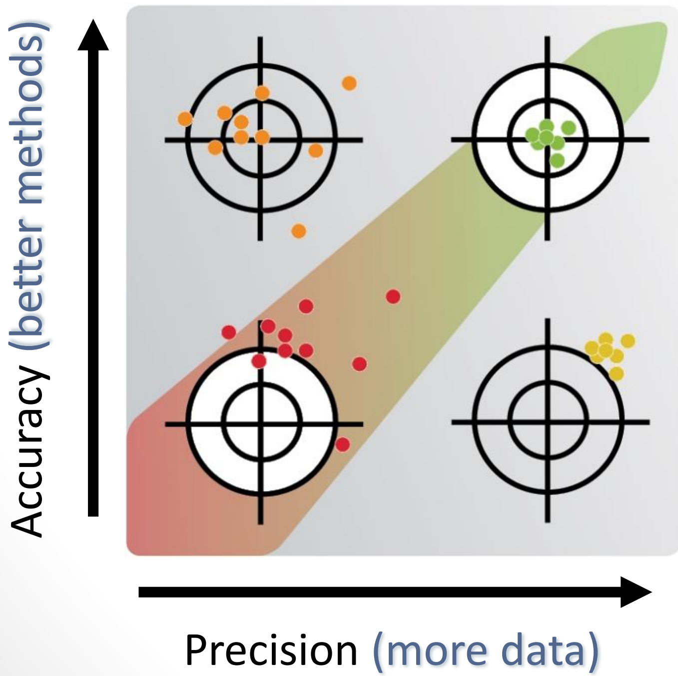
DESI, Euclid, LSST, WFIRST, and more…

Large-scale structure surveys roadmap



Data-intensive scientific discovery from galaxy surveys

- Next-generation surveys will be dominated by systematics
- 80% of the total signal will come from non-linear structures

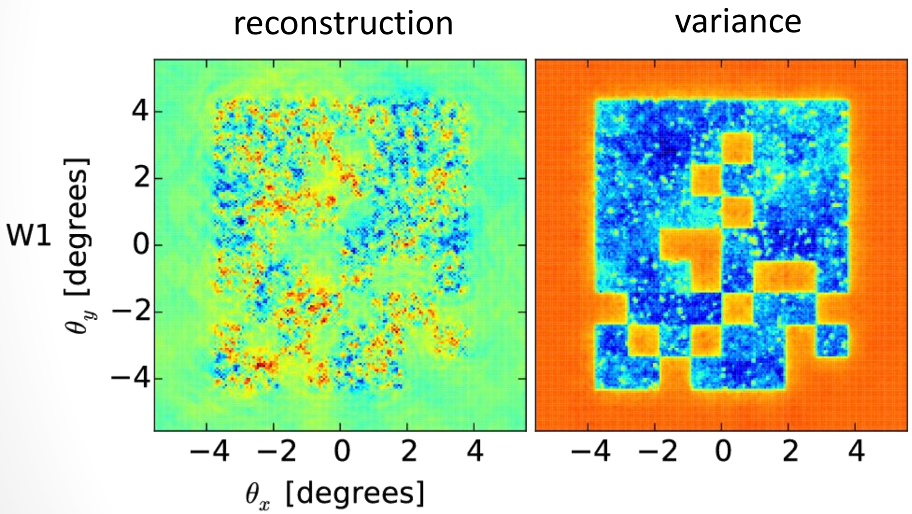


- Can data analysts keep pace?

The Imperial weak lensing inference framework

with George Kyriacou, Arrykrishna Mootoovaloo, Alan Heavens & Andrew Jaffe

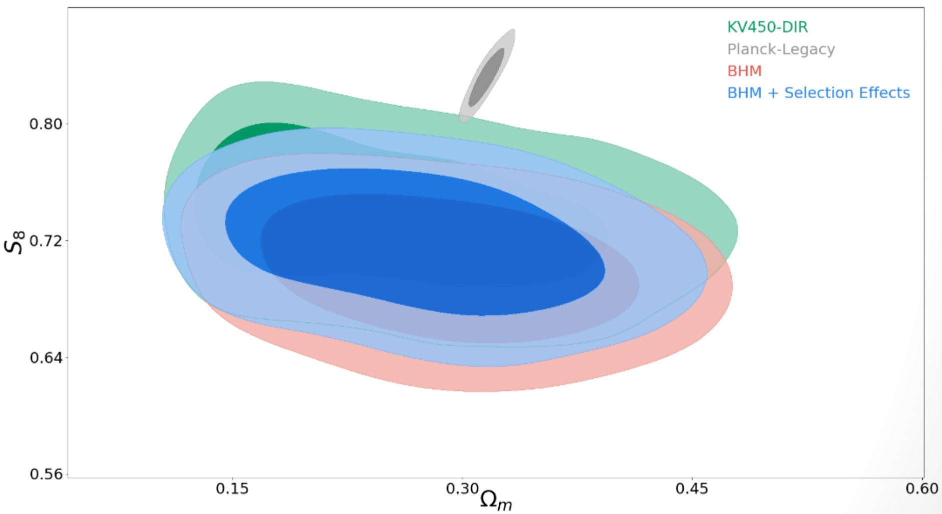
Joint inference of cosmic shear maps and power spectra/cosmology from CFHTLenS



$$\sum m_\nu < 4.6\ \mathrm{eV}(95\%)$$ from lensing data alone

Alsing, Heavens & Jaffe 2016, 1607.00008

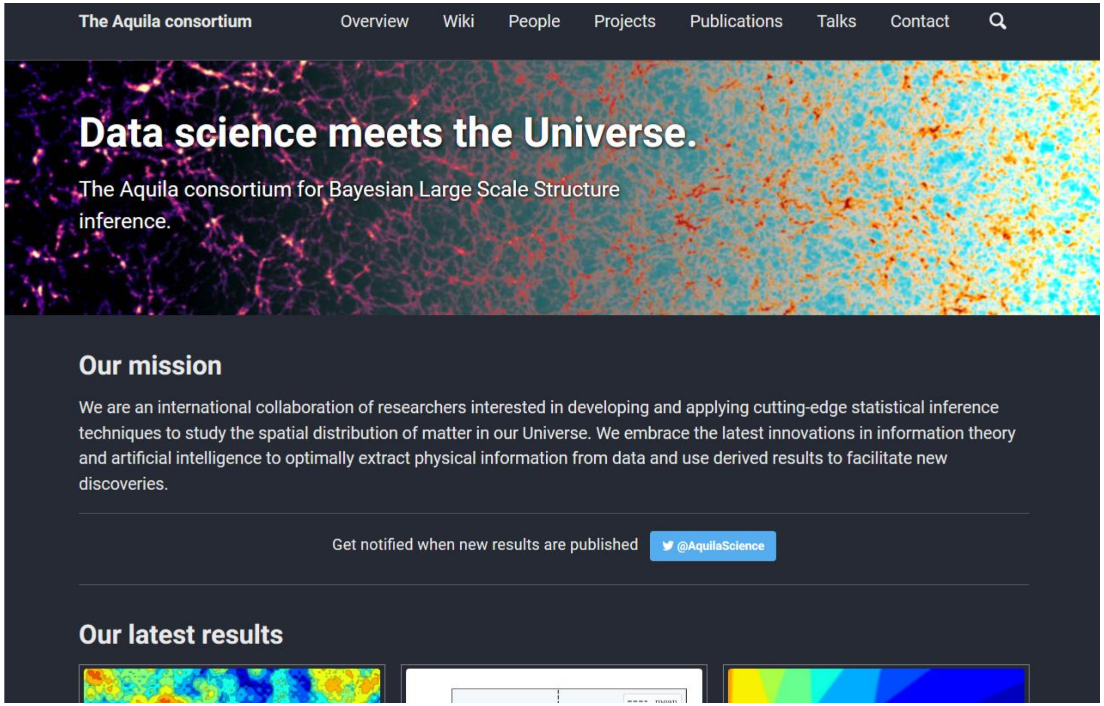
Bayesian hierarchical inference of galaxy redshift distributions $n(z)$



Kyriacou, FL, Heavens & Jaffe, in prep.

The Aquila Consortium

- Created in 2016. Members from the UK, France, Germany & Sweden.
- Gathers people interested in developing the Bayesian pipelines and running analyses on cosmological data.



www.aquila-consortium.org

Concluding thoughts

Likelihood-based solution:
Exact statistical analysis
Approximate data model

Data assimilation

?

Likelihood-free solution:
Approximate statistical analysis
Arbitrary data model

Generative inference

- Bayesian analyses of galaxy surveys with fully non-linear numerical models is not an impossible task!
- A likelihood-based solution (BORG): general purpose reconstruction of dark matter from galaxy clustering, providing new measurements and predictions
- A likelihood-free solution (SELFI): algorithm for targeted questions, allowing the use of simulators including all relevant physical and observational effects

Concluding thoughts

- The future: great science and challenges

Galaxy formation: bias model & likelihood
Large volume, photometric redshifts
Instruments modelling

DESI, Euclid, LSST, WFIRST

Bayesian large-scale structure analyses

Cosmological measurements:

- Cosmic expansion
- Power spectrum (and governing parameters)
- Gaussianity tests of the initial conditions
- Dark energy from the growth of structure

Predictive cosmology:

- Velocity field
- X-ray cluster emission
- Gravitational lensing
- CMB secondary effects
- Dark matter?