

Building the ultimate Bayesian machine to interpret cosmological datasets

Guilhem Lavaux (IAP/CNRS)
for the Aquila Consortium

University of Oxford, May 7th 2019



From pictures... to physics of Universe at large



From pictures... to physics of Universe at large

In a handful of model parameters



Ultimately: we want to fit a model to this kind of picture, and the pixel by pixel spectrum



Ultimately: we want to fit a model to this kind of picture, and the pixel by pixel spectrum
That's very challenging, probably impossible → we reduce those datasets

Outline



Introduction



The chosen path: embrace the complexity



First results on 2M++ and SDSS3

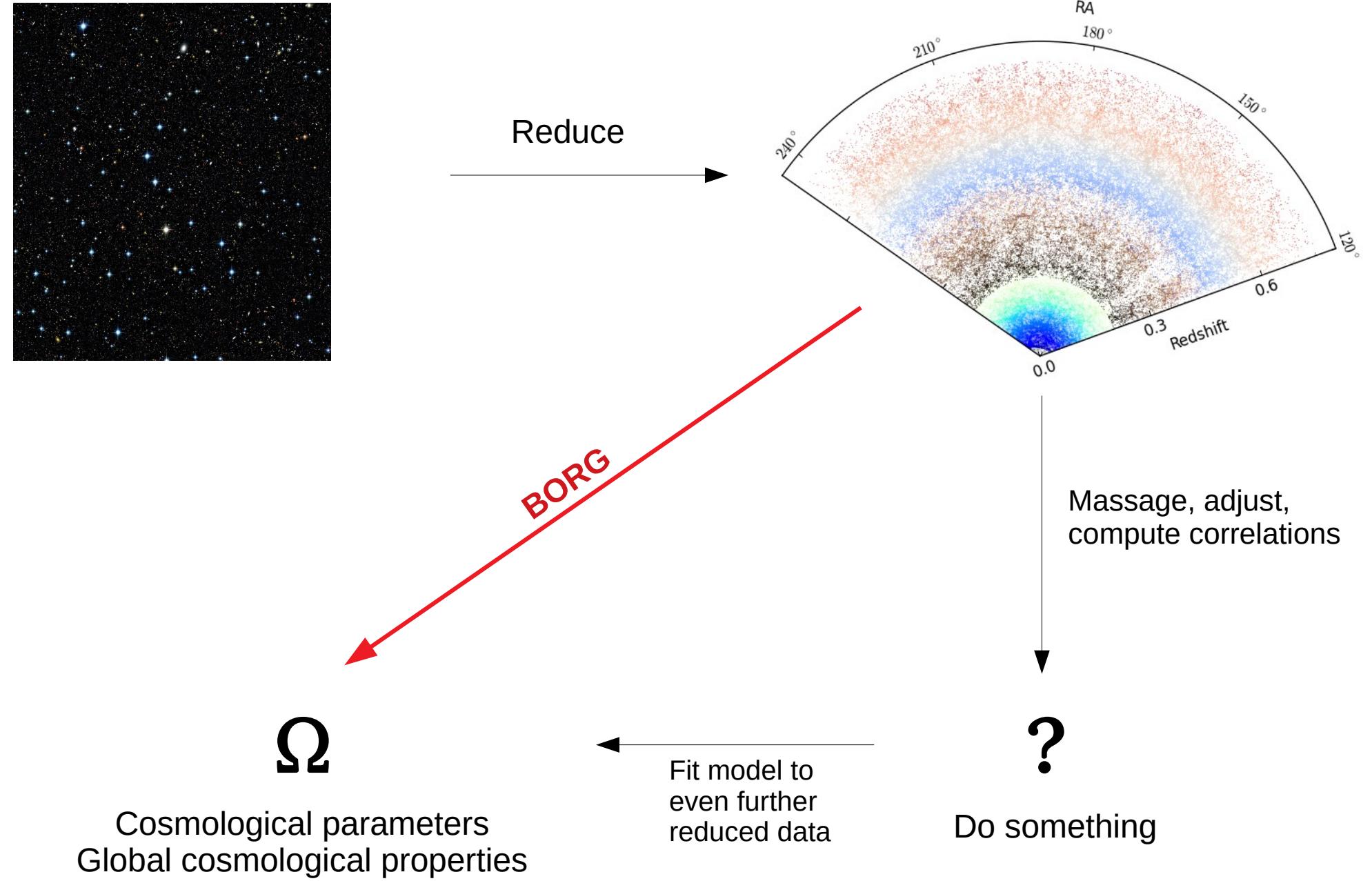


More models considerations:

- Altair (Alcock Pasczyński test)
- VIRBIUS2 (Flow inference with distance data)
- Lyman alpha
- Neural networks



The path forward / Conclusion



From theory to observations...

Model

- Perfect
- Complete description
- Full knowledge of physics
- Did I say perfect ?



Observations

- Great but messy
- We do not understand the physics
- Systematics not fully known
- Good attempt by observers to seemingly make our life easier end up bad

Various hacking to make sense of data



From theory to observations...

Model

- Perfect
- Complete description
- Full knowledge of physics
- Did I say perfect ?



Observations

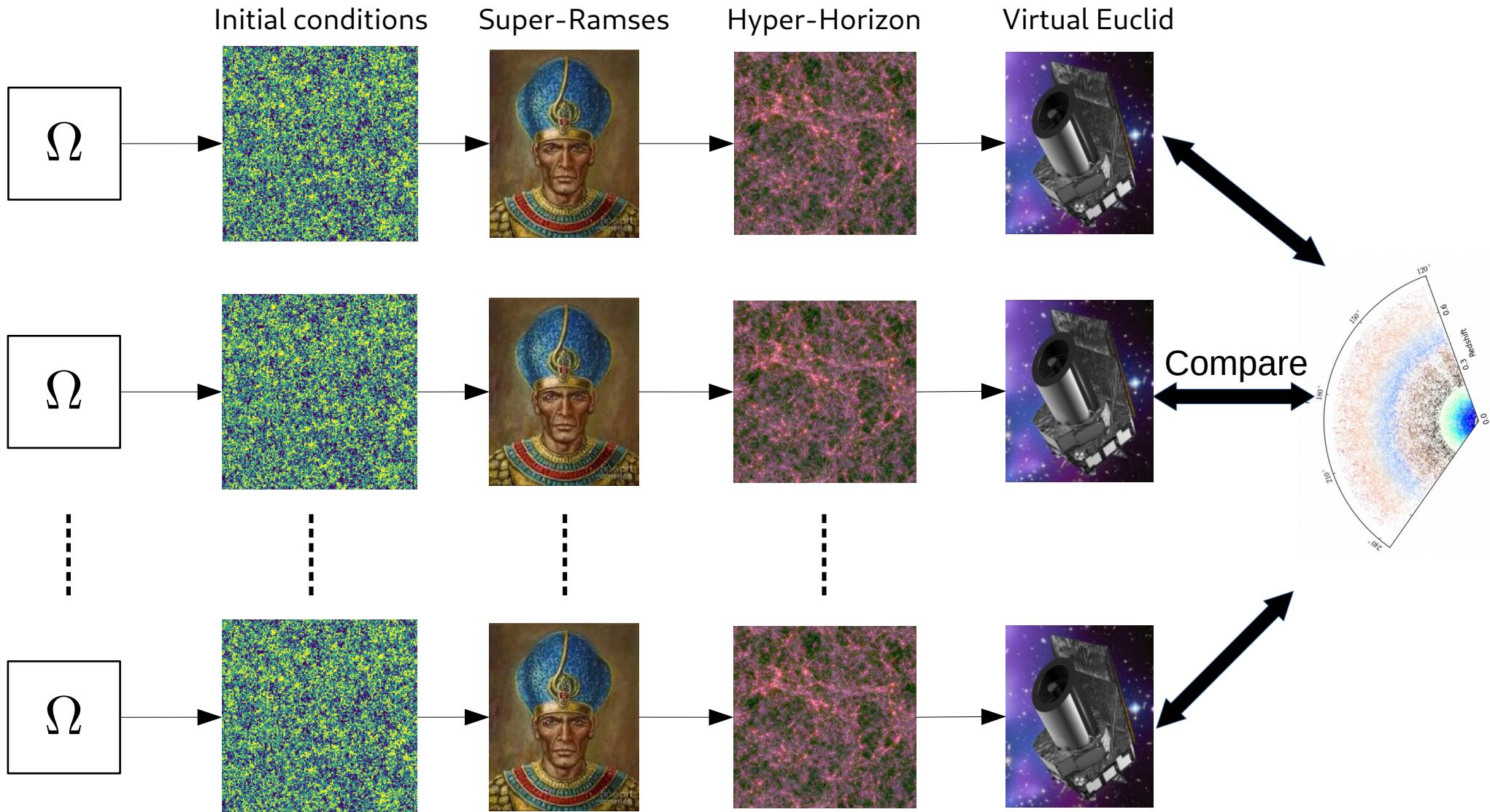
- Great but messy
- We do not understand the physics
- Systematics not fully known
- Good attempt by observers to seemingly make our life easier end up bad



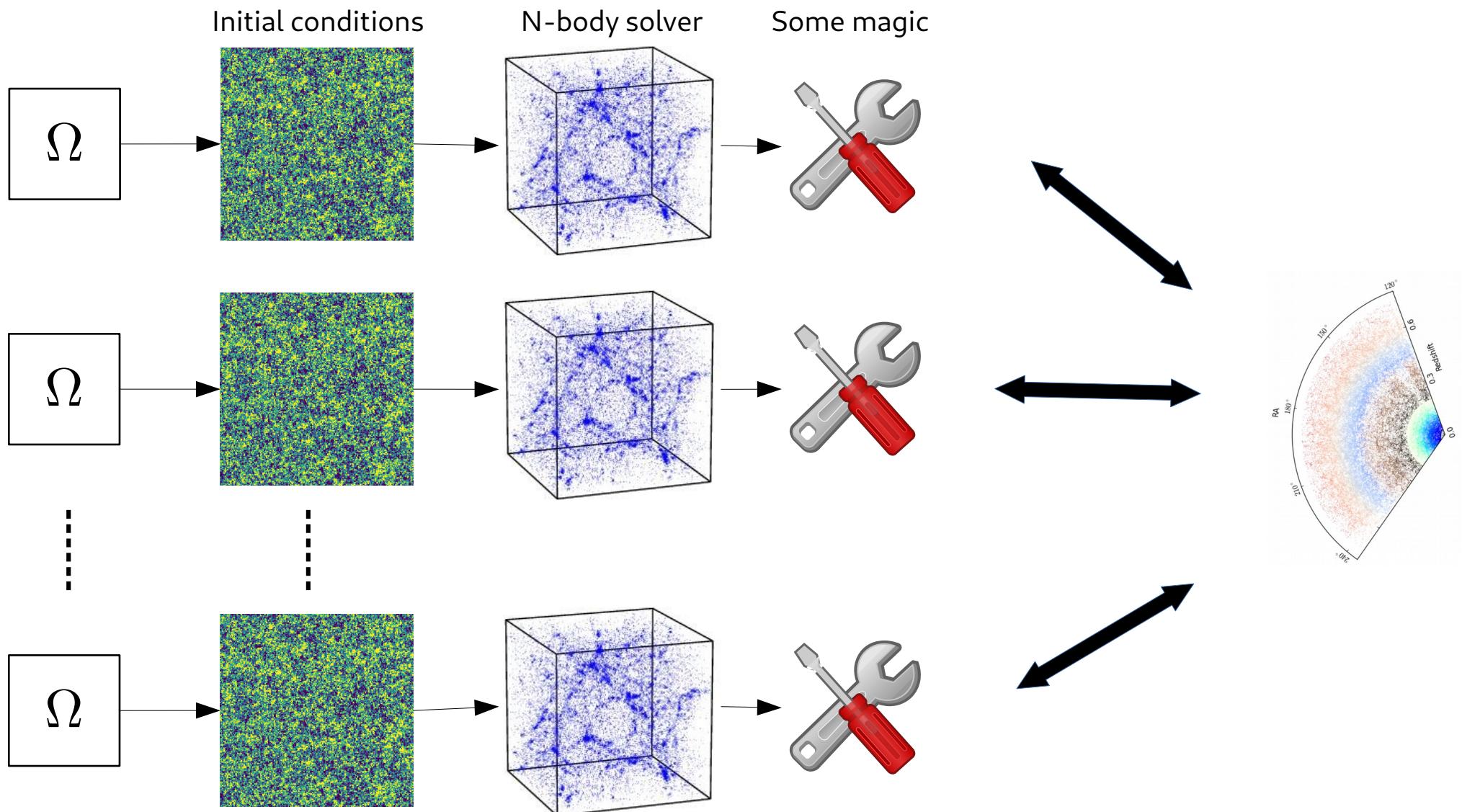
Still far too perfect though... (see later)



The ideal scheme



The more pragmatic scheme



In practice...



The BORG cube

In practice...

github.com/AlDanial/cloc v 1.72 T=0.26 s (1649.8 files/s, 264776.5 lines/s)				
Language	files	blank	comment	code
C++	191	7515	4704	23358
C/C++ Header	235	6407	4066	22438
Julia	4	92	64	366
SUM:	430	14014	8834	46162

Check ARES at https://bitbucket.org/bayesian_lss_team/

The BORG cube

The BORG3 inference framework

$$\pi(\hat{\delta}) \propto \exp\left(-\frac{1}{2} \sum_k |\hat{\delta}_k|^2 / P_k\right)$$

Initial conditions

The BORG3 inference framework

$$\pi(\hat{\delta}) \propto \exp\left(-\frac{1}{2} \sum_k |\hat{\delta}_k|^2 / P_k\right)$$

Initial conditions

Total evolved matter density $\rho_m = \mathcal{F}[\delta]$

The BORG3 inference framework

$$\pi(\hat{\delta}) \propto \exp\left(-\frac{1}{2} \sum_k |\hat{\delta}_k|^2 / P_k\right)$$

Initial conditions

Total evolved matter density $\rho_m = \mathcal{F}[\delta]$

Biased galaxy distribution $\rho_g \propto \rho_m^\alpha \exp\left(-(\rho_m/\rho_0)^{-\epsilon}\right)$

Selected/contaminated sample $\rho_g^s(\vec{x}) = S(\vec{x})\rho_g(\vec{x})$

**Random extraction
(i.e. observational metric)** $N_g^s \leftarrow \mathcal{P}(\rho_g^s)$ (Poisson, Negative binomial, ...)

The BORG3 inference framework

$$\pi(\hat{\delta}) \propto \exp\left(-\frac{1}{2} \sum_k |\hat{\delta}_k|^2 / P_k\right)$$

Initial conditions

Total evolved matter density

$$\rho_m = \mathcal{F}[\delta]$$

Biased galaxy distribution

$$\rho_g \propto \rho_m^\alpha \exp\left(-(\rho_m/\rho_0)^{-\epsilon}\right)$$

Selected/contaminated sample $\rho_g^s(\vec{x}) = S(\vec{x})\rho_g(\vec{x})$

**Random extraction
(i.e. observational metric)**

$$N_g^s \leftarrow \mathcal{P}(\rho_g^s) \quad (\text{Poisson, Negative binomial, ...})$$

Forward and adjoint model

The BORG3 inference framework

$$\pi(\hat{\delta}) \propto \exp\left(-\frac{1}{2} \sum_k |\hat{\delta}_k|^2 / P_k\right)$$

Initial conditions

Total evolved matter density

$$\rho_m = \mathcal{F}[\delta]$$

Biased galaxy distribution

$$\rho_g \propto \rho_m^\alpha \exp\left(-(\rho_m/\rho_0)^{-\epsilon}\right)$$

Selected/contaminated sample $\rho_g^s(\vec{x}) = S(\vec{x})\rho_g(\vec{x})$

**Random extraction
(i.e. observational metric)**

$$N_g^s \leftarrow \mathcal{P}(\rho_g^s) \quad (\text{Poisson, Negative binomial, ...})$$

Easily exchangeable to try
your favorite differentiable model

The BORG3 inference framework

$$\pi(\hat{\delta}) \propto \exp\left(-\frac{1}{2} \sum_k |\hat{\delta}_k|^2 / P_k\right)$$

Initial conditions

Total evolved matter density $\rho_m = \mathcal{F}[\delta]$

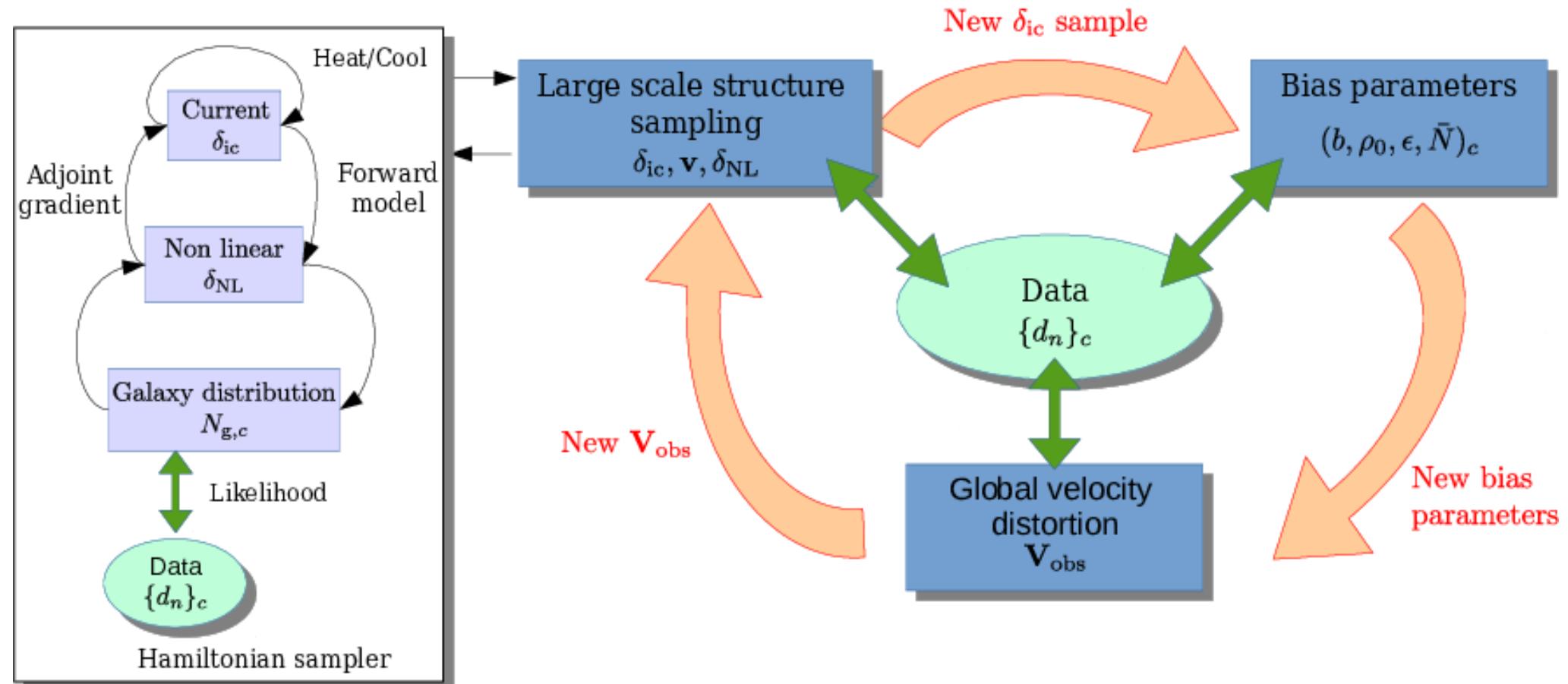
Biased galaxy distribution $\rho_g \propto \rho_m^\alpha \exp\left(-(\rho_m/\rho_0)^{-\epsilon}\right)$

Selected/contaminated sample $\rho_g^s(\vec{x}) = \mathbf{S}(\vec{x})\rho_g(\vec{x})$

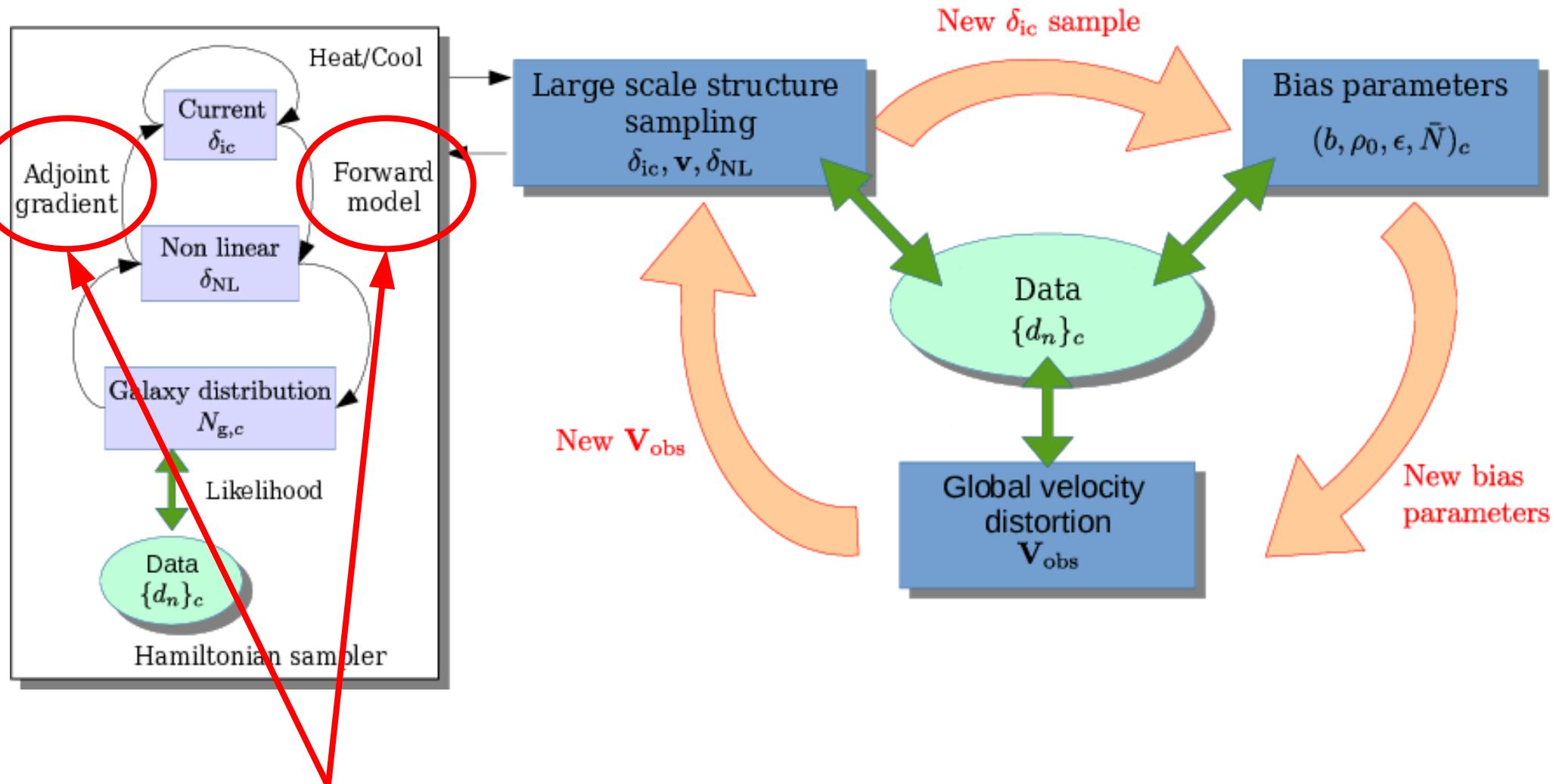
**Random extraction
(i.e. observational metric)** $N_g^s \leftarrow \mathcal{P}(\rho_g^s)$ (Poisson, Negative binomial, ...)

Encode survey systematic effects with expansions: $S(\hat{x}) = S_0(\hat{x}) \prod_{f=1}^N (1 + \alpha_f F_f(\hat{x}))$

The BORG3 machine

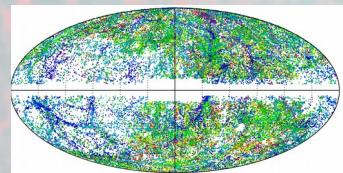


The BORG3 machine



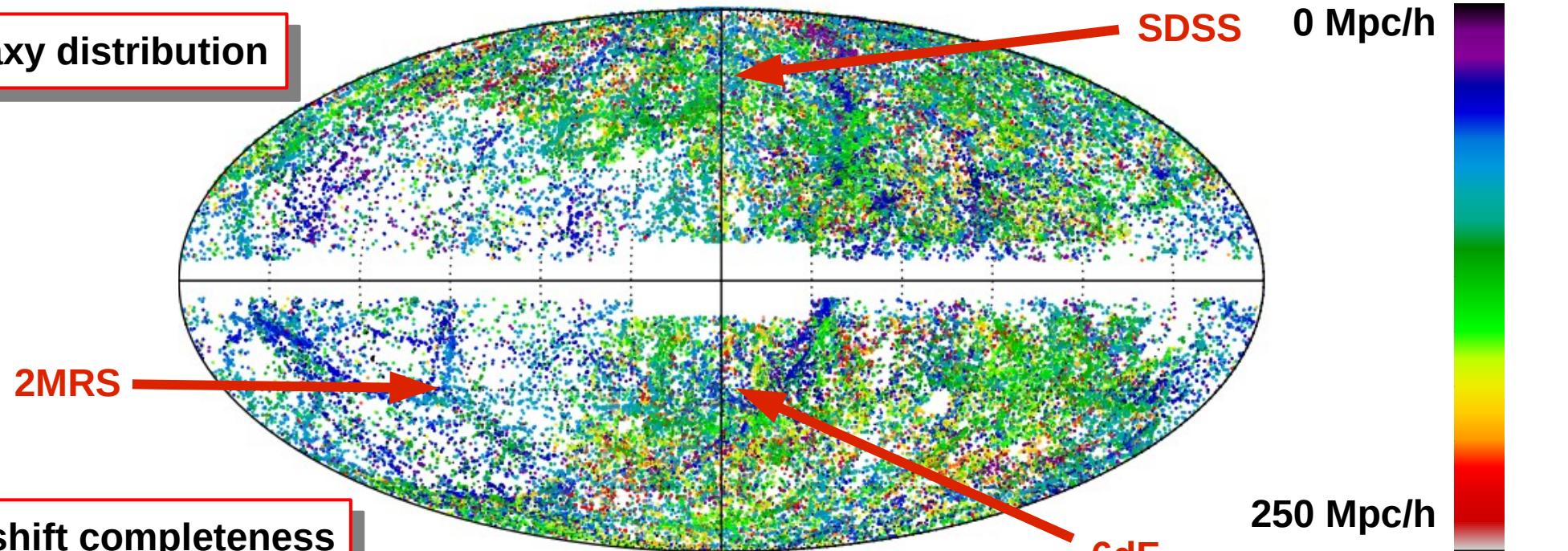
Gradient is only ~2x more expensive than the forward

Application to 2M++: Detailed dynamical modeling



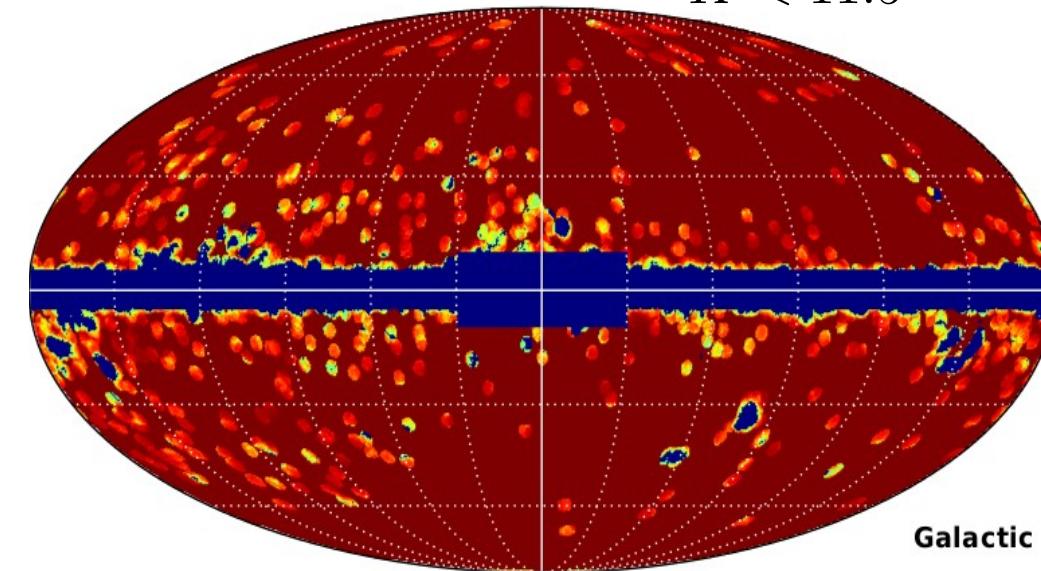
The 2M++ galaxy compilation

Galaxy distribution



Redshift completeness

$K < 11.5$



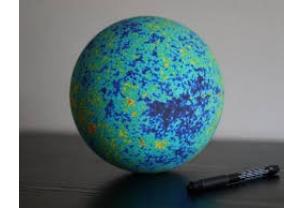
~70 000 galaxies

Galactic

Lavaux & Hudson (MNRAS, 2011)

The model

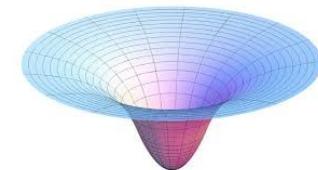
Λ CDM Universe with Planck+15 cosmological parameters



Box of $(677.7 \text{ Mpc}/h)^3$
 256^3 initial condition elements
 512^3 particles



Particle mesh solver
Redshift space distortions derived from particle simulations

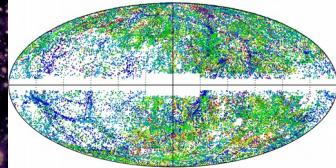


Bias model: $\rho_g \propto \rho_m^\alpha \exp\left(-(\rho_m/\rho_0)^{-\epsilon}\right)$

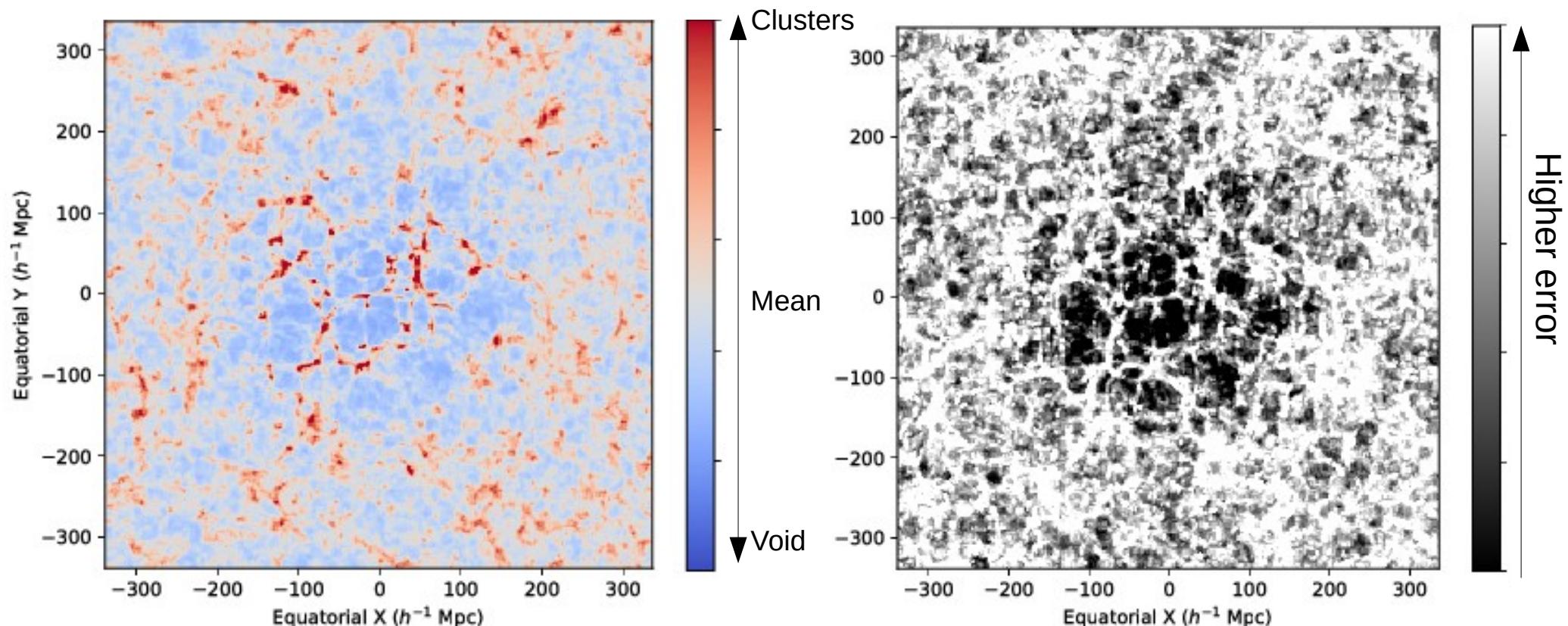
Selection derived from Schechter luminosity function



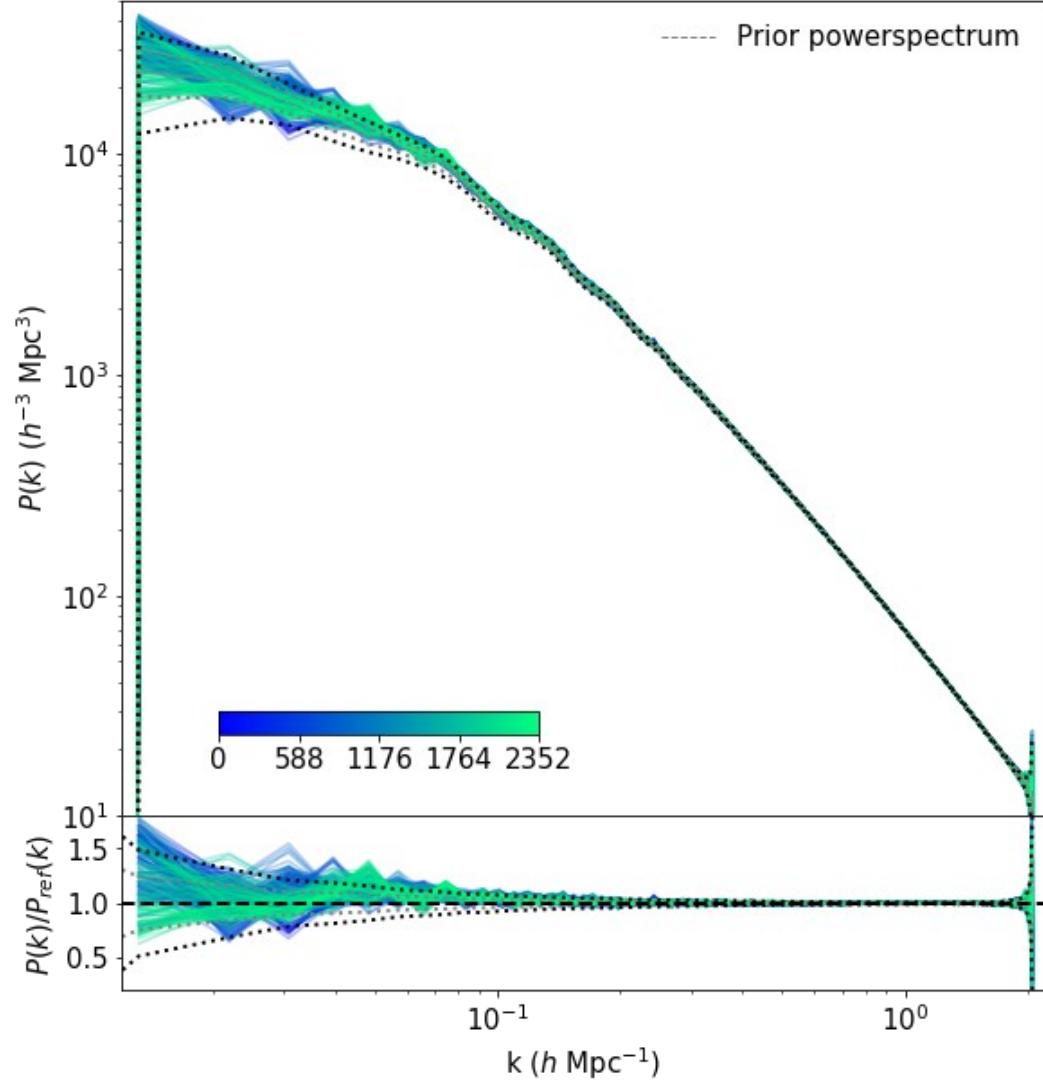
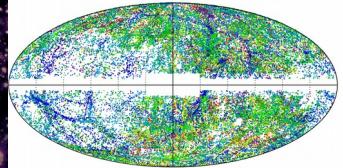
Inferred density fields



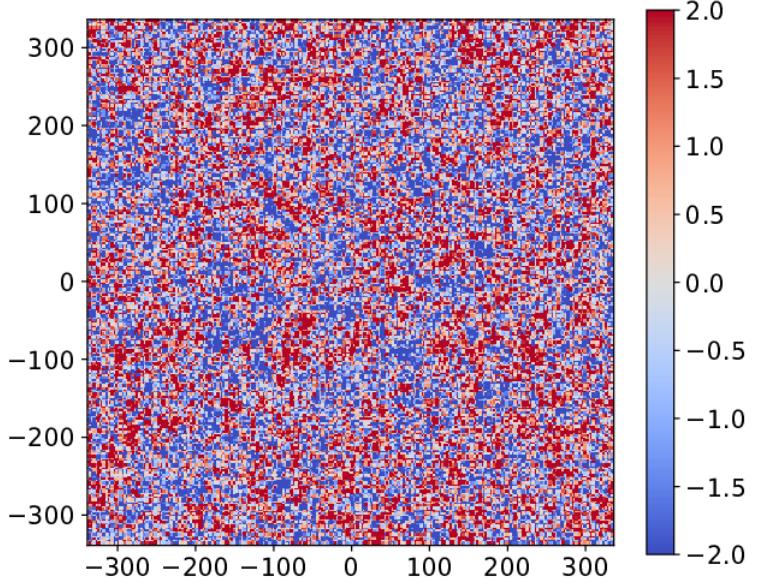
Ensemble average density fields at $z=0$



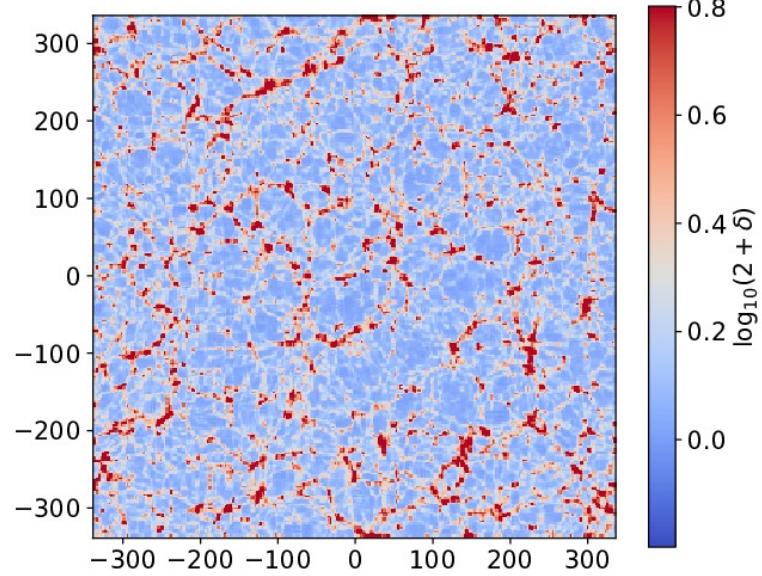
Initial condition powerspectrum



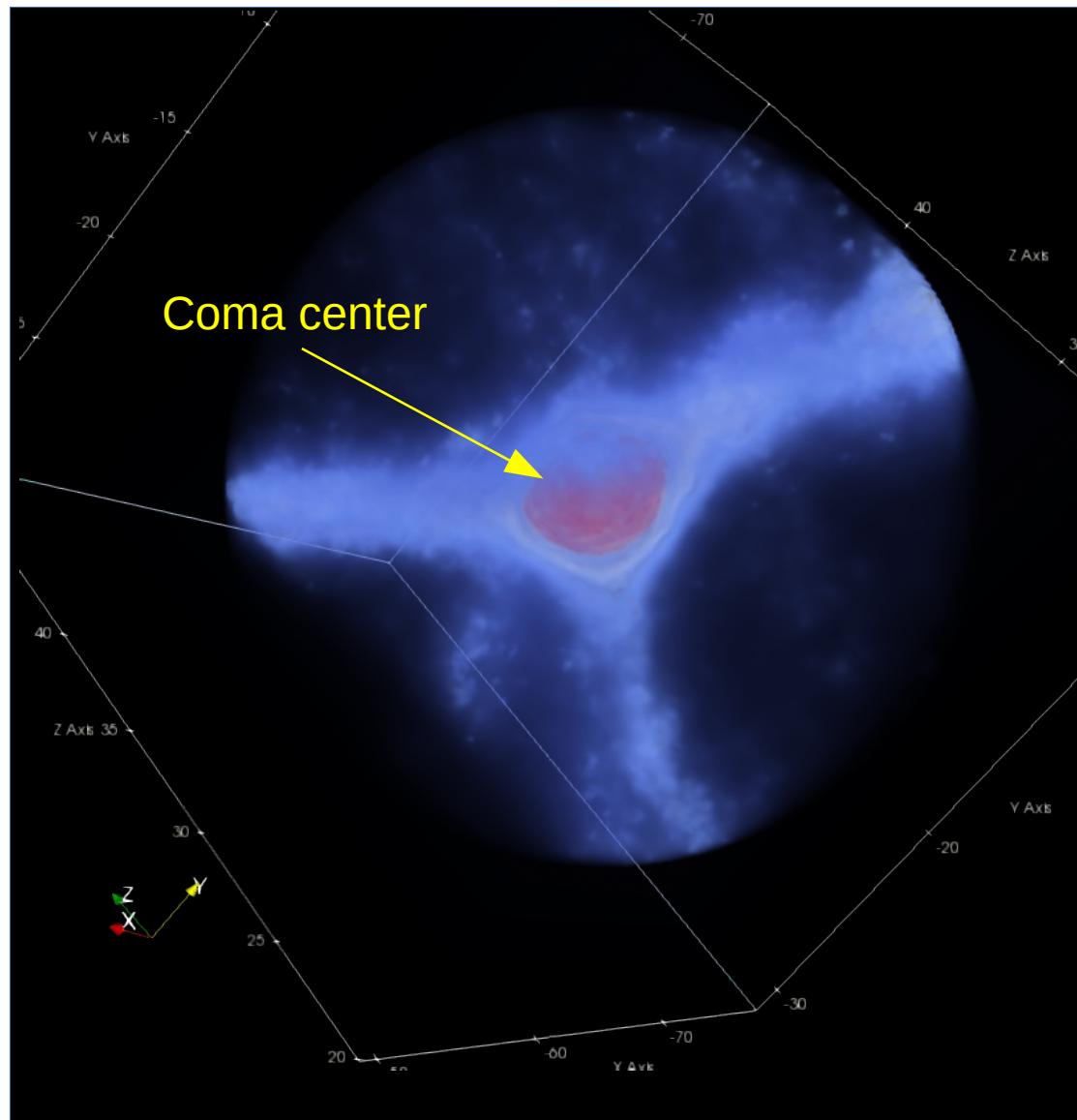
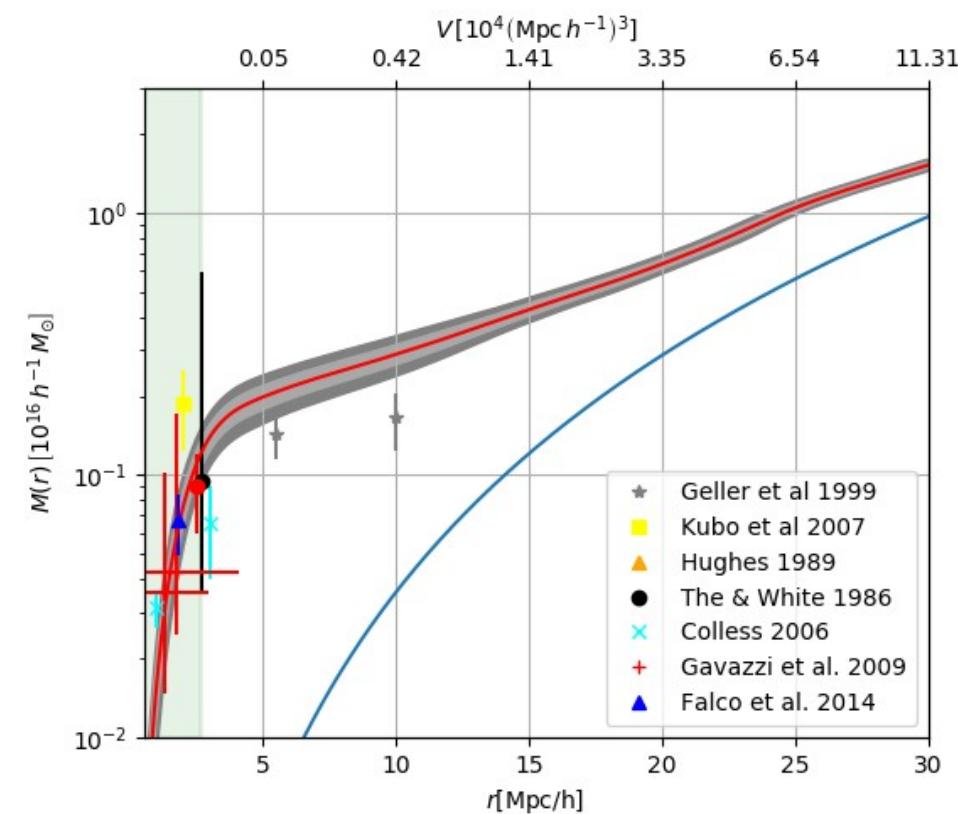
Initial conditions



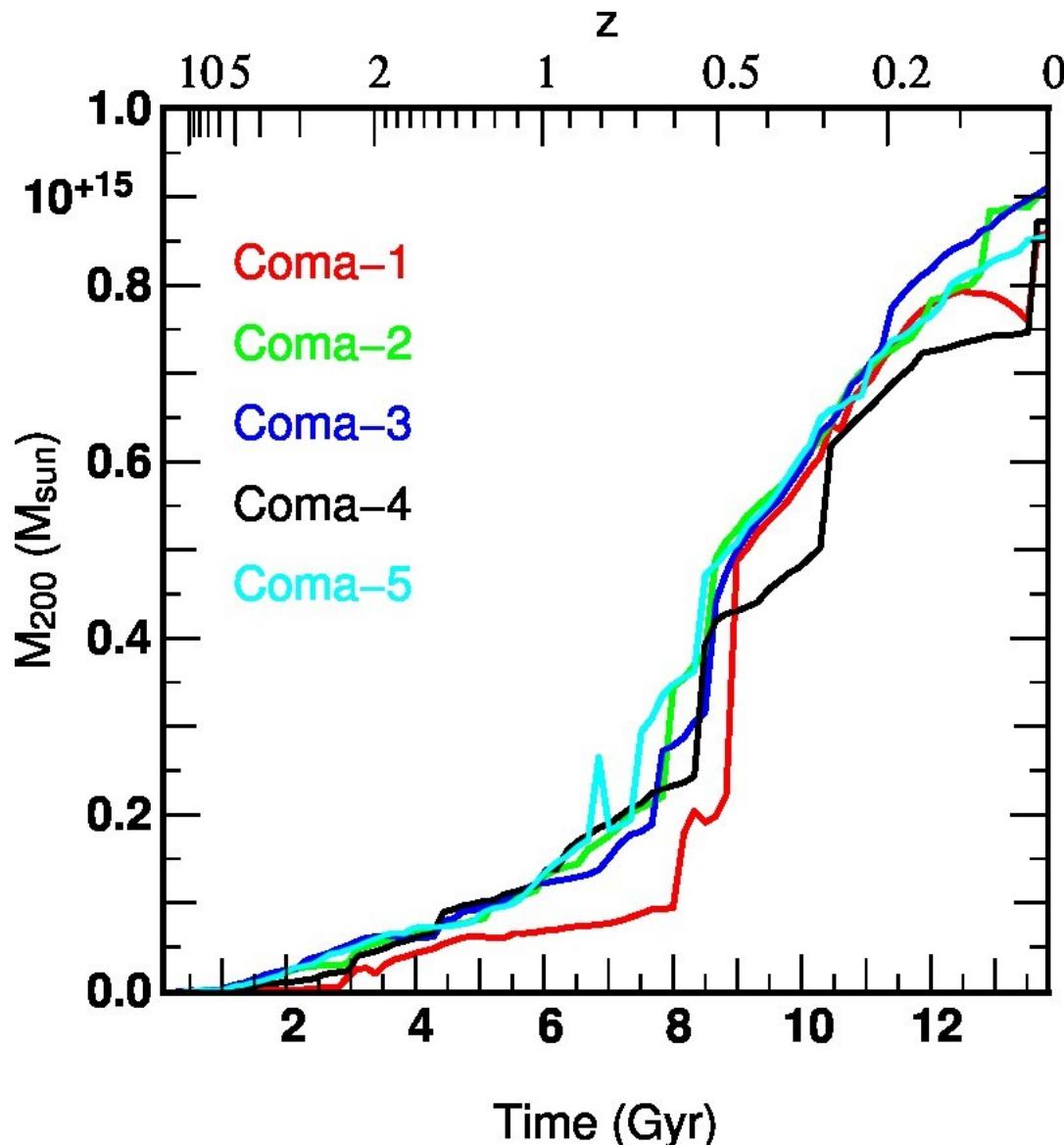
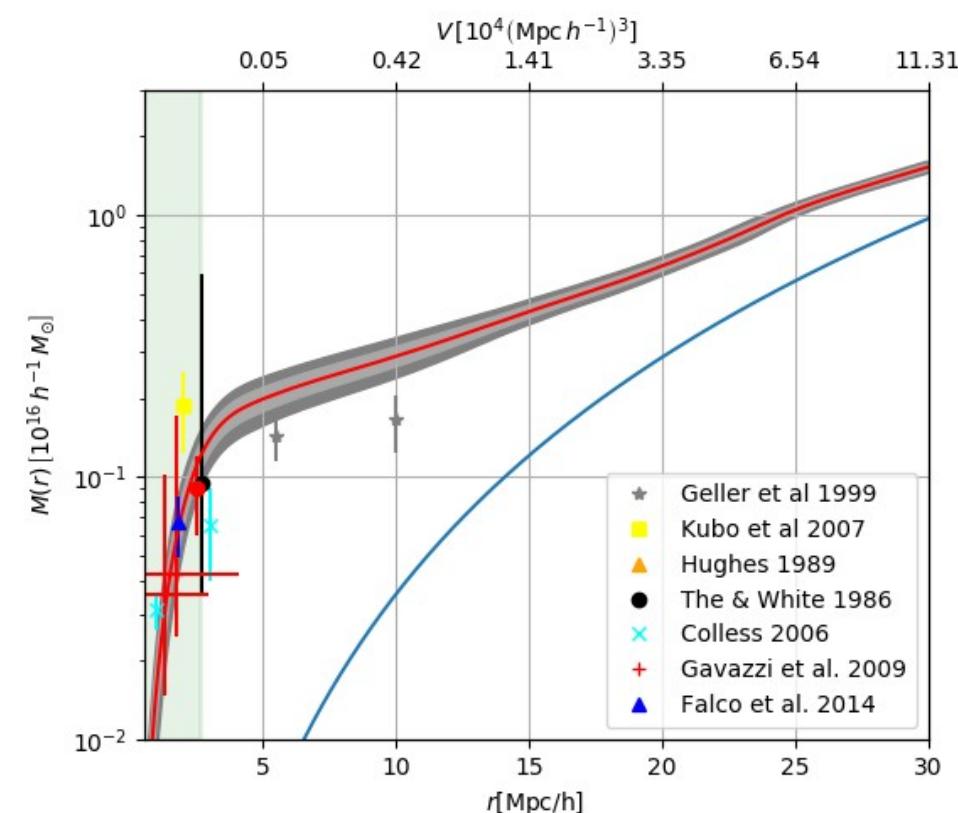
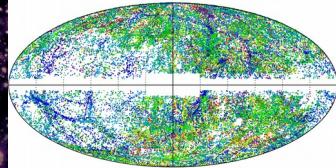
Post PM simulation



Coma dynamical properties



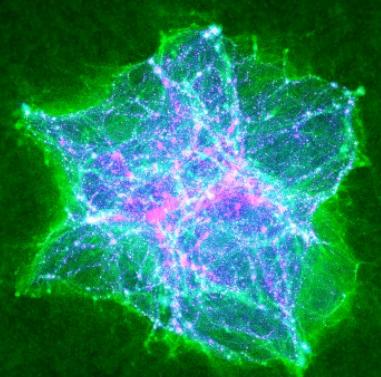
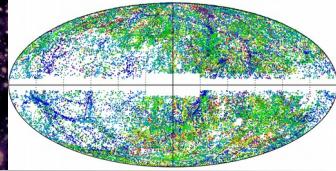
Coma dynamical properties



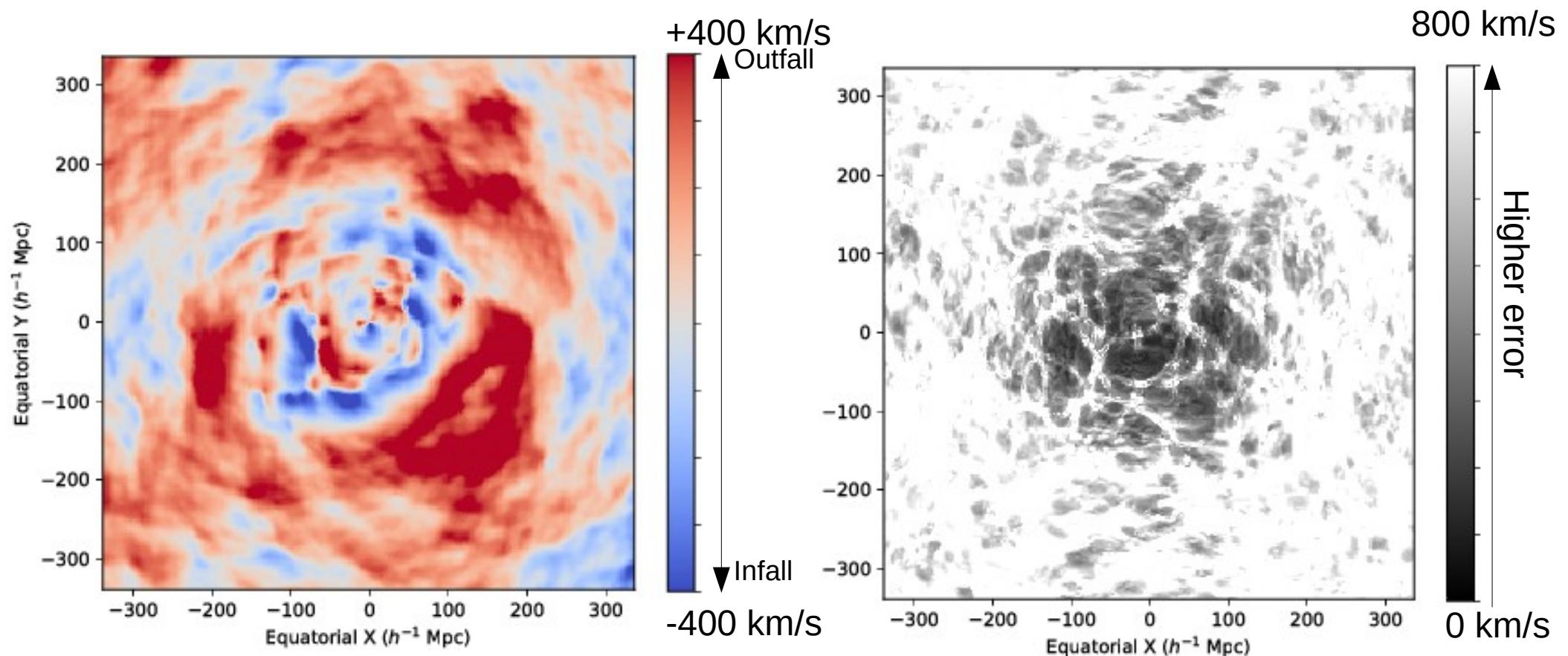
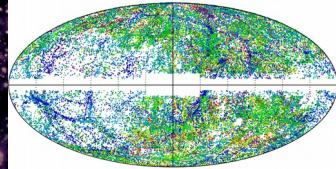
Zoom simulation on Coma
(~250 Mpart in zoom)

$4 \times 10^7 h^{-1} M_\odot/\text{part}$

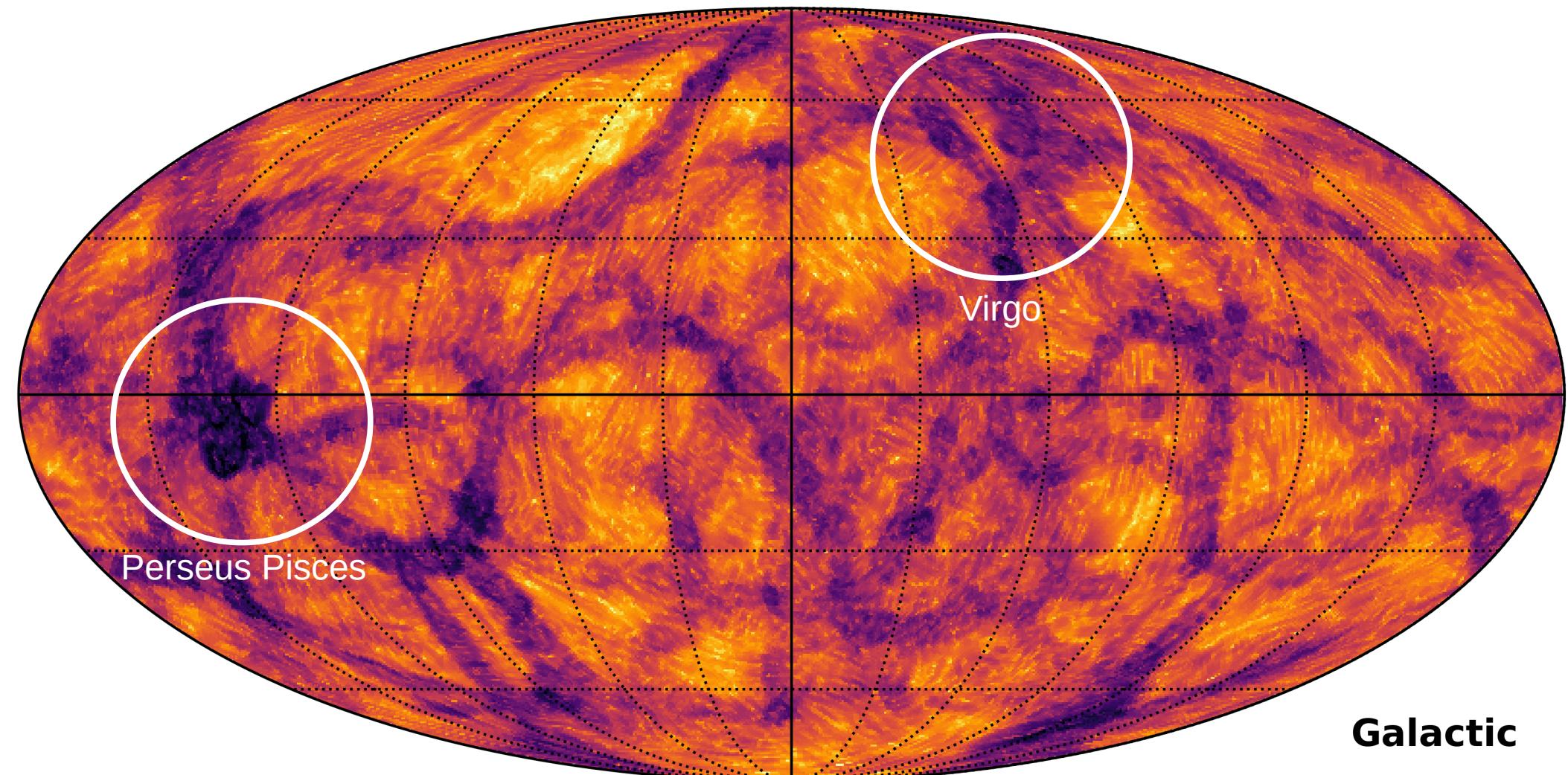
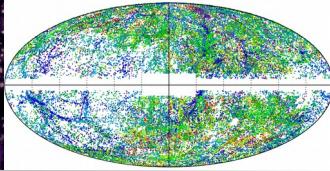
Zoomed coma



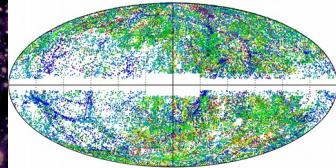
Inferred velocity fields



Peculiar velocity field vorticity

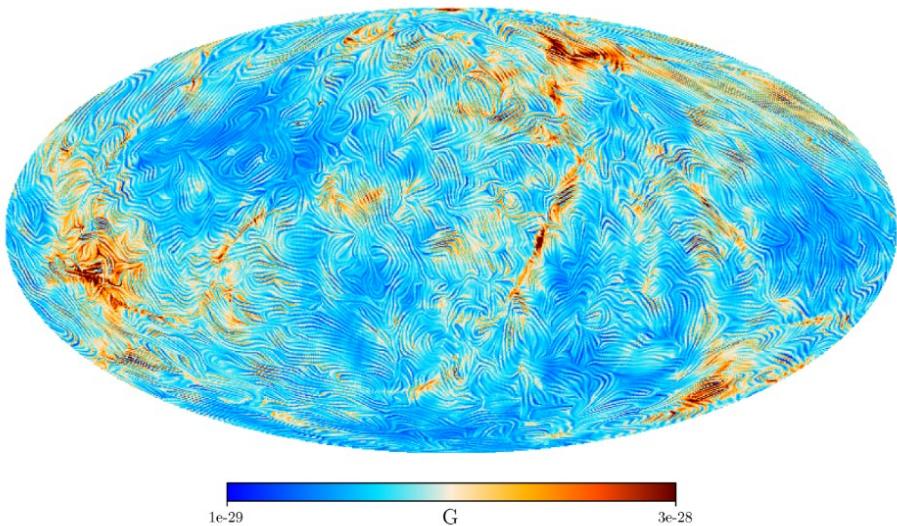


More applications



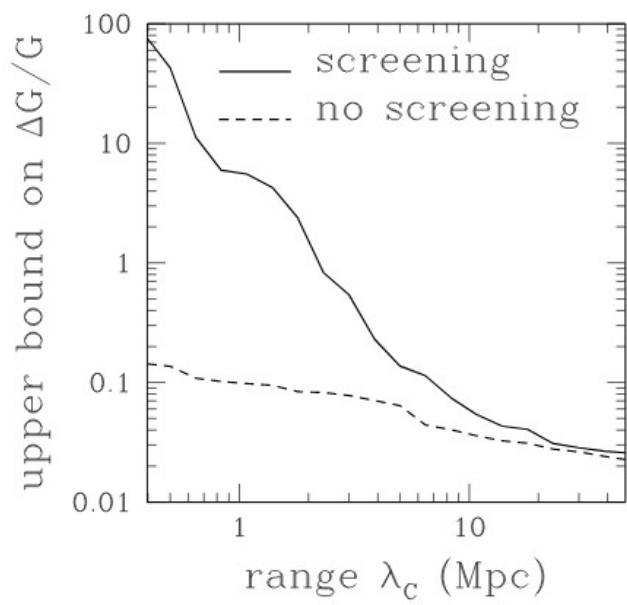
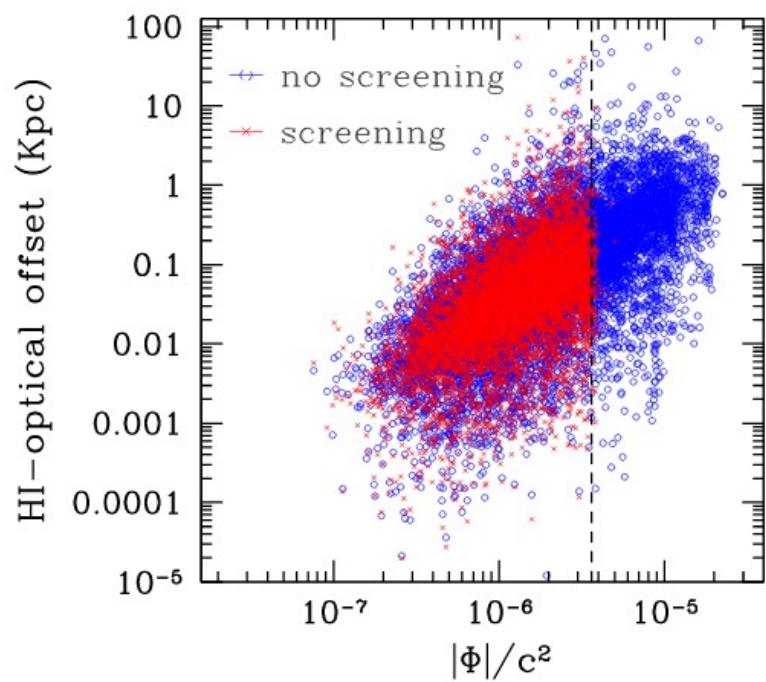
Magnetic field in our backyard generated by primordial mechanisms

Hutschenreuter et al. (2018, CQG)

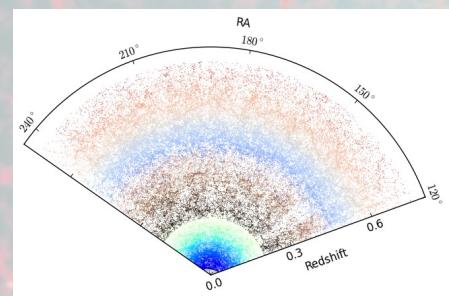


“Fifth-force” constraints

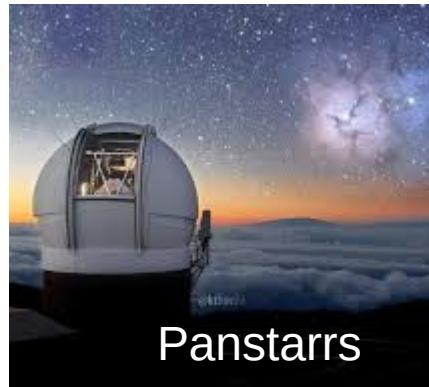
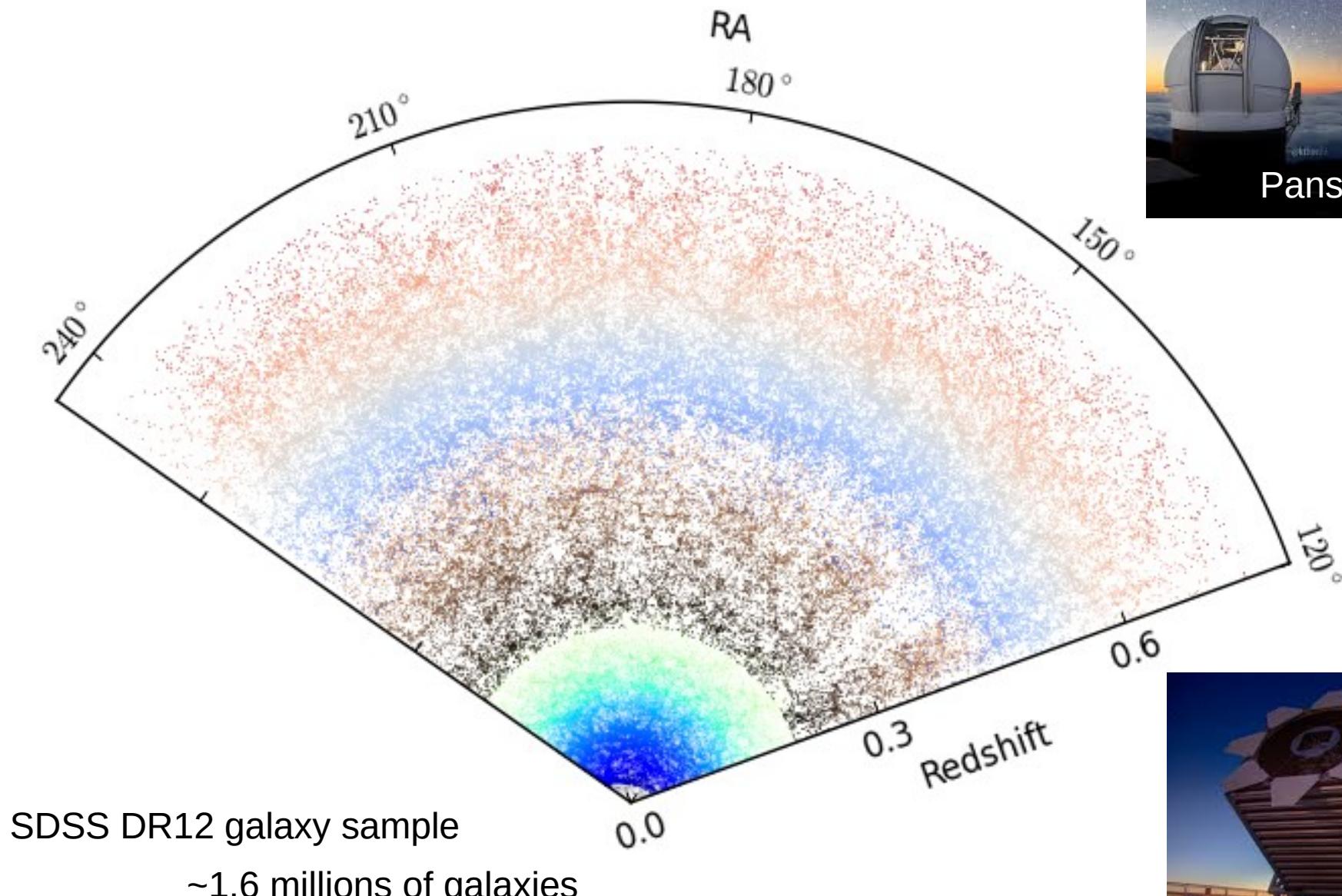
Desmond et al.
(2018abc, PRD, MNRASL)



Application to Sloan Digital Sky Survey III: Systematic cleanings

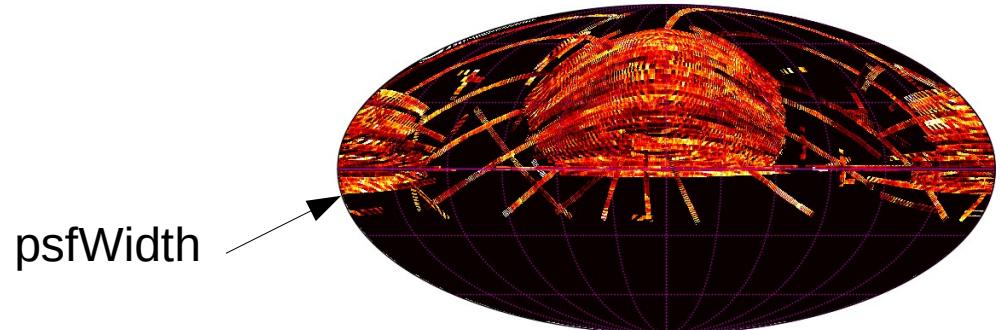
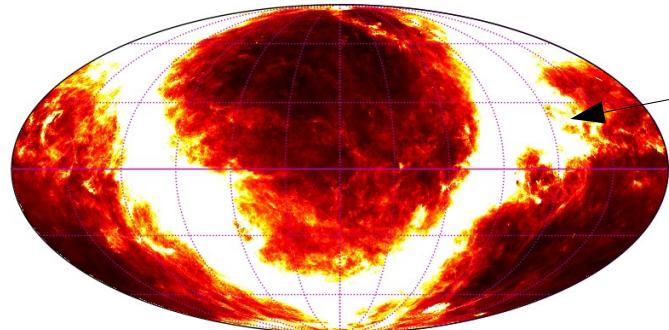


SDSS3 data

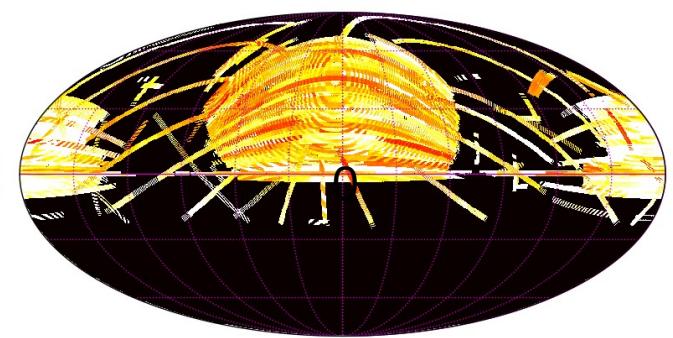
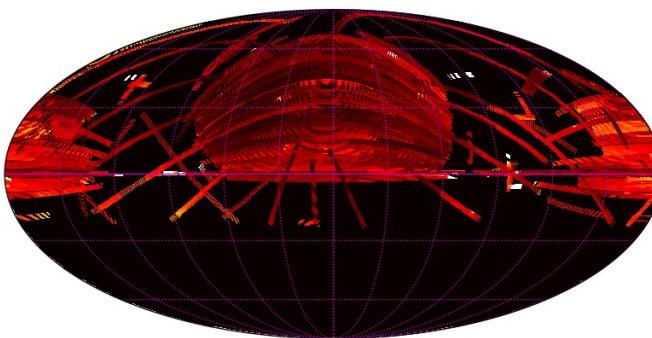
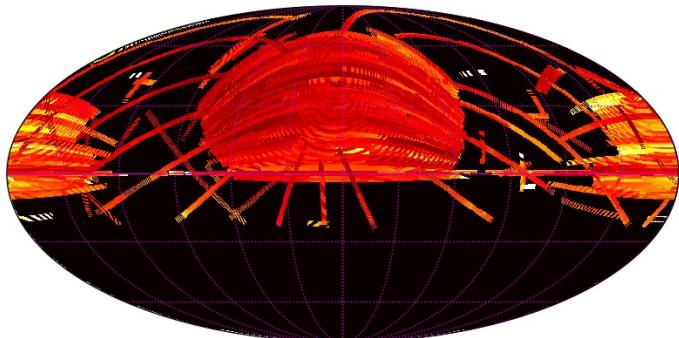


Non exhaustive list of contaminations

11 foregrounds (here only 8)... still much less than Leistedt & Peiris (2014) but improving

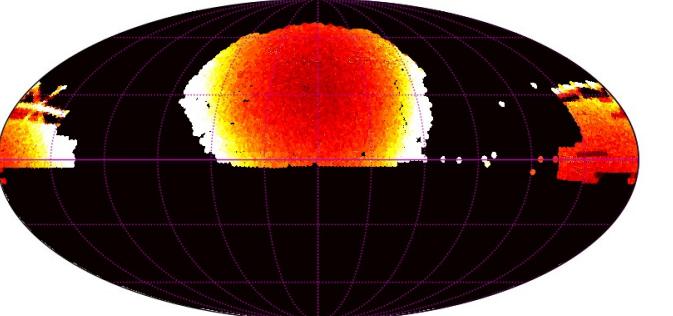
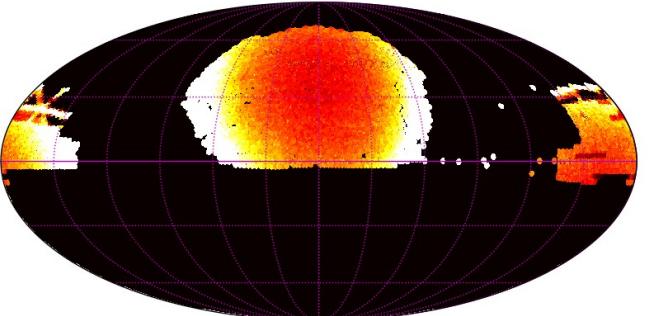
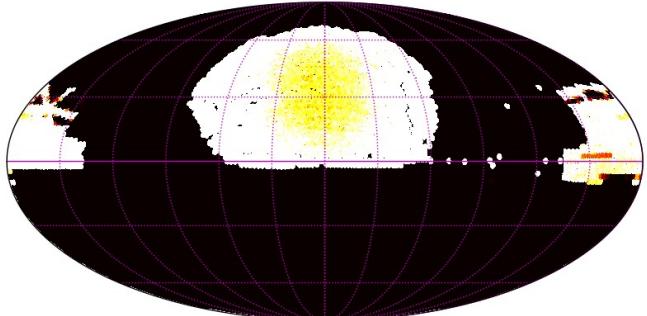


Sky fluxes



...

Star densities



...

Robust likelihood

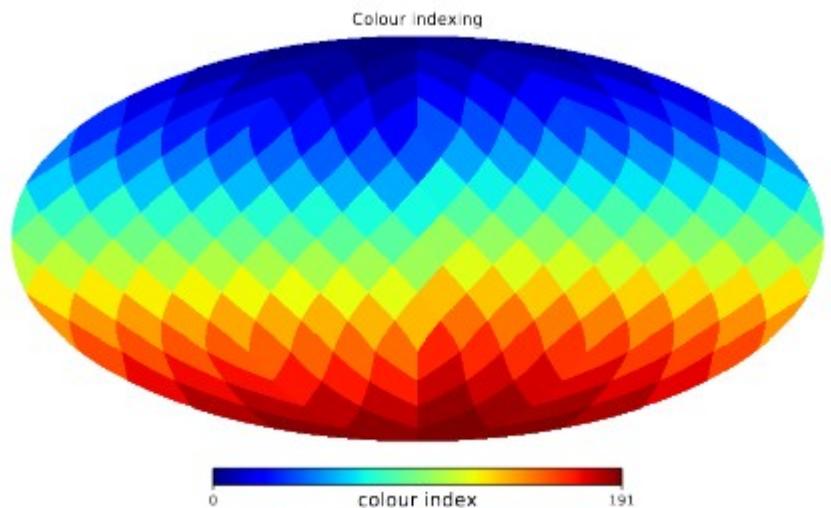


Each count in 3d patch ↪ Poisson probability

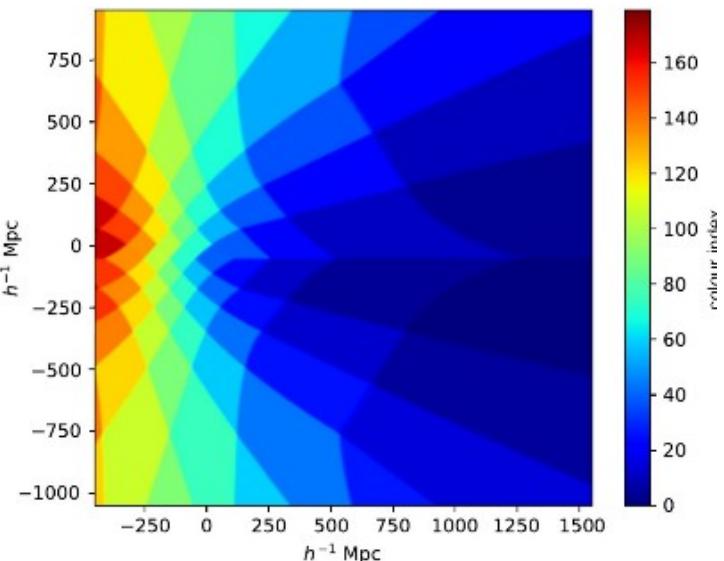
Yield a new effective likelihood

$$P(\{N\}|\{\lambda\}) \propto \prod_{\text{patch } i \in \text{patch}} \prod_{j \in \text{patch}} \left(\frac{\lambda_i}{\sum_{j \in \text{patch}} \lambda_j} \right)^{N_i}$$

Map of the patches on the sky



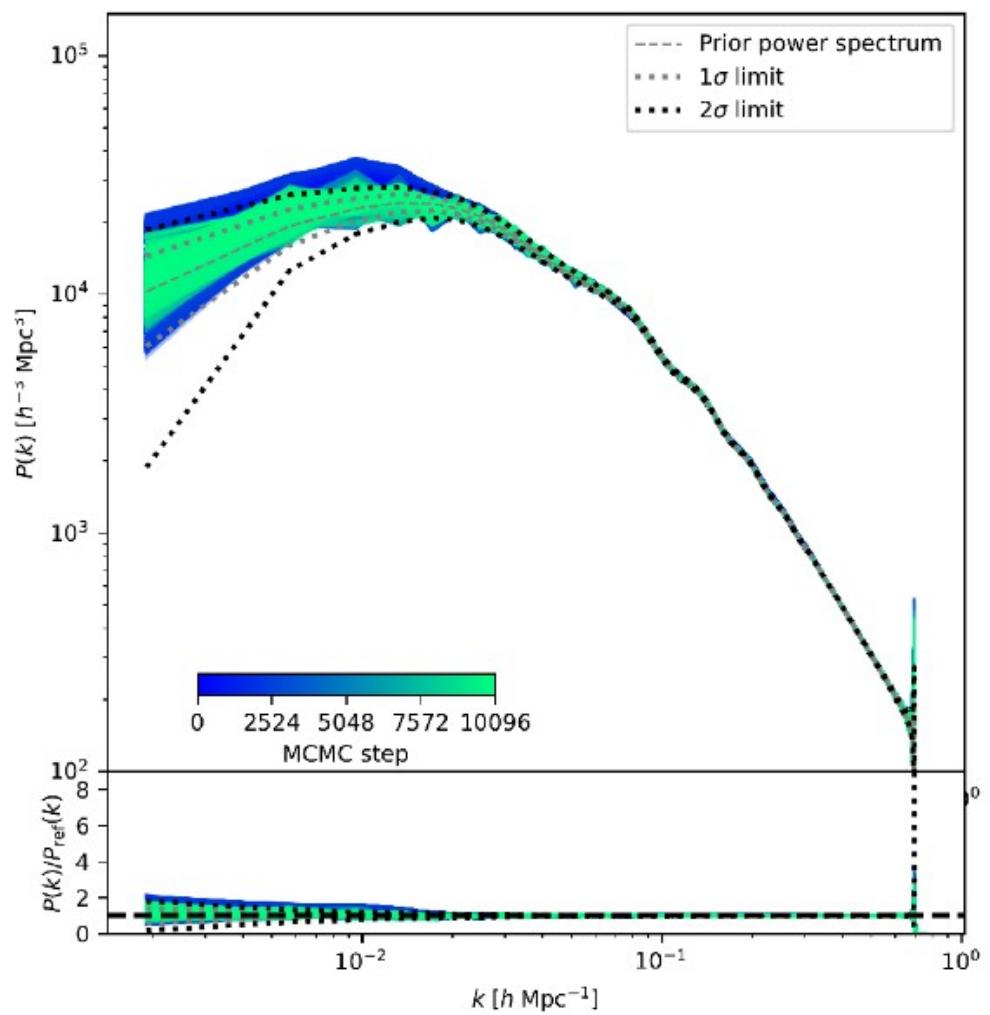
...Extruded in 3d



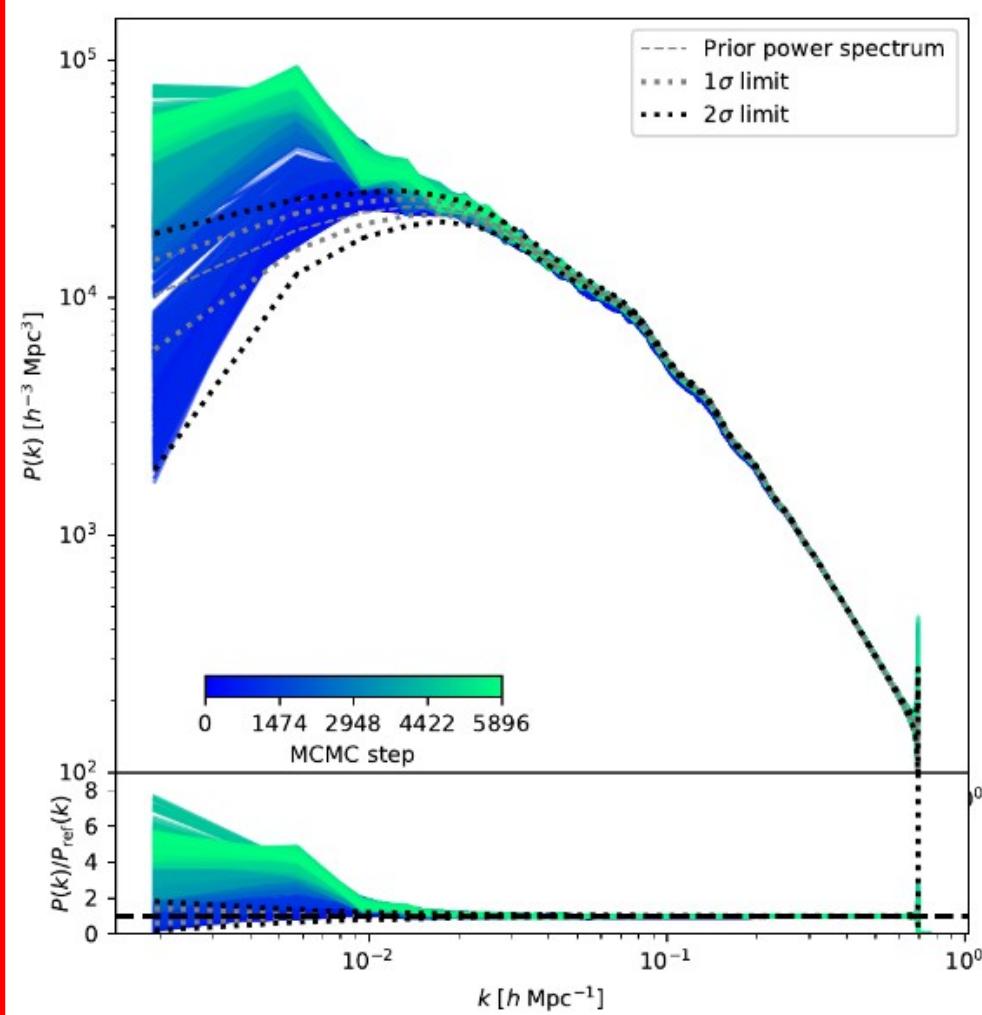
Results on mock SDSS3 data



Robust inference of contaminated data

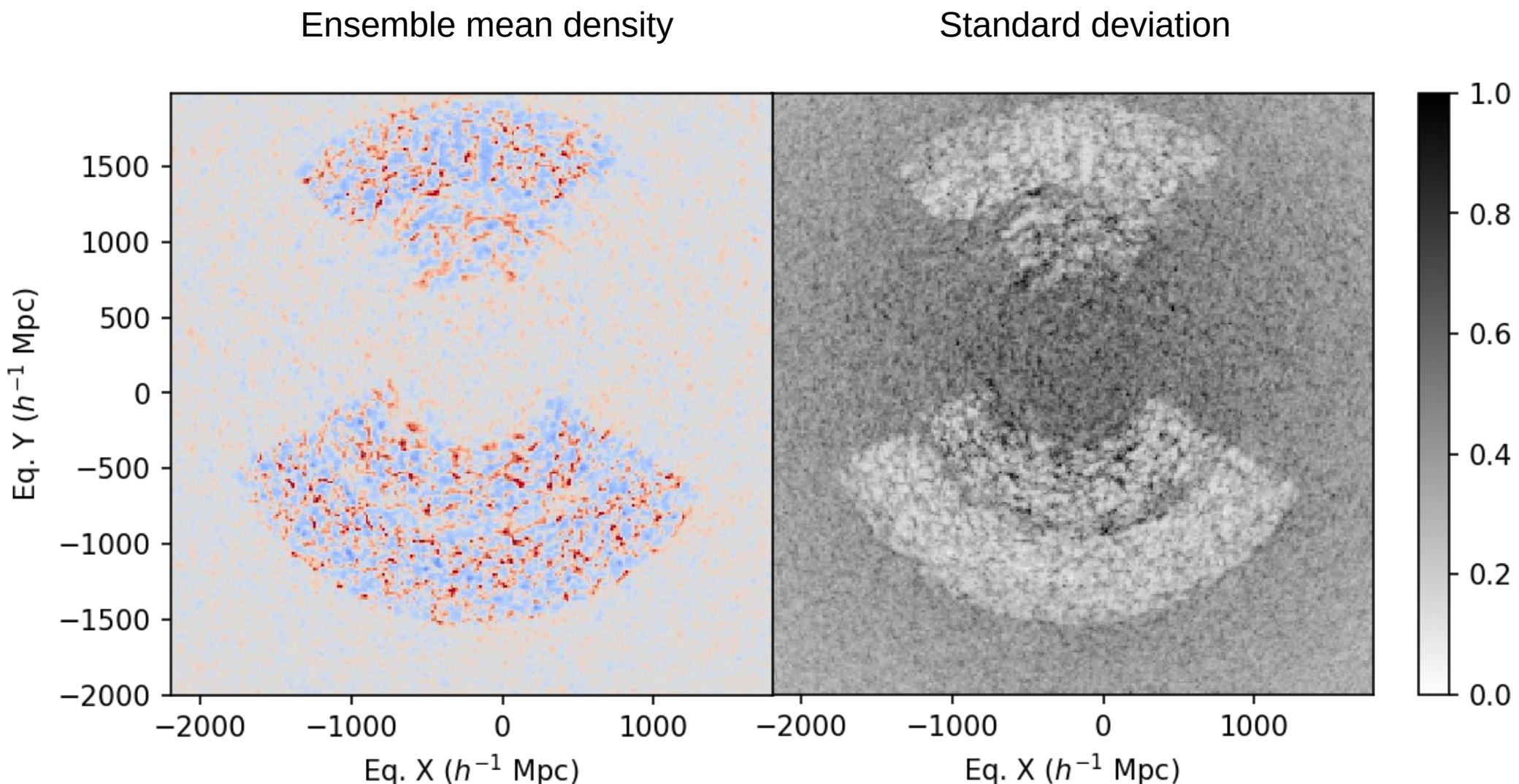


Inference of contaminated data without correction



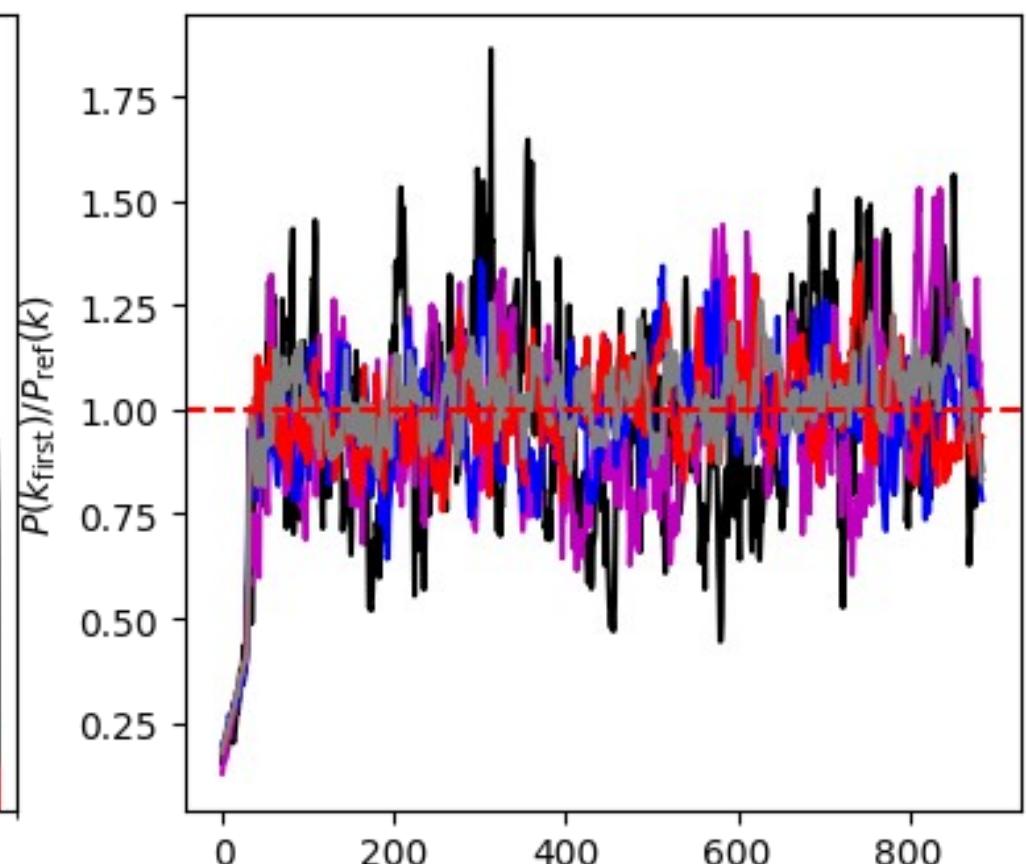
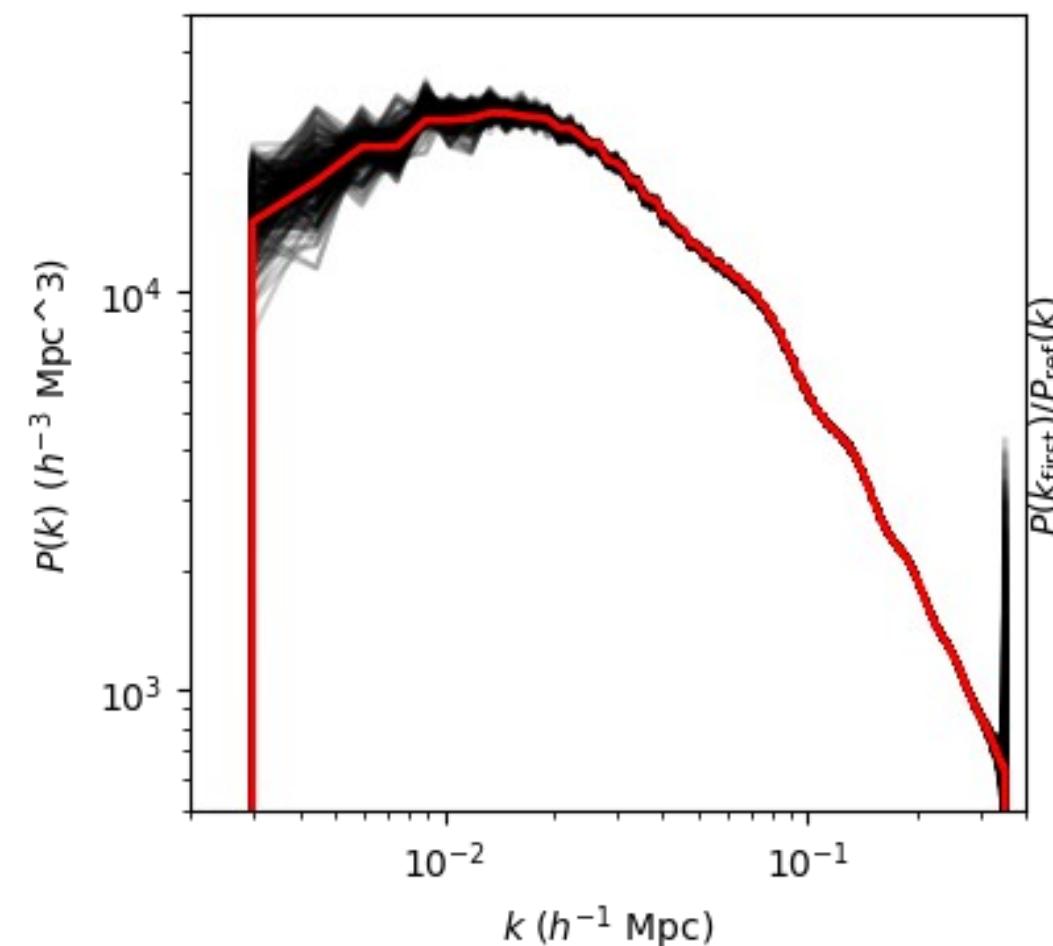
Inference results: density and P(k)

Preliminary



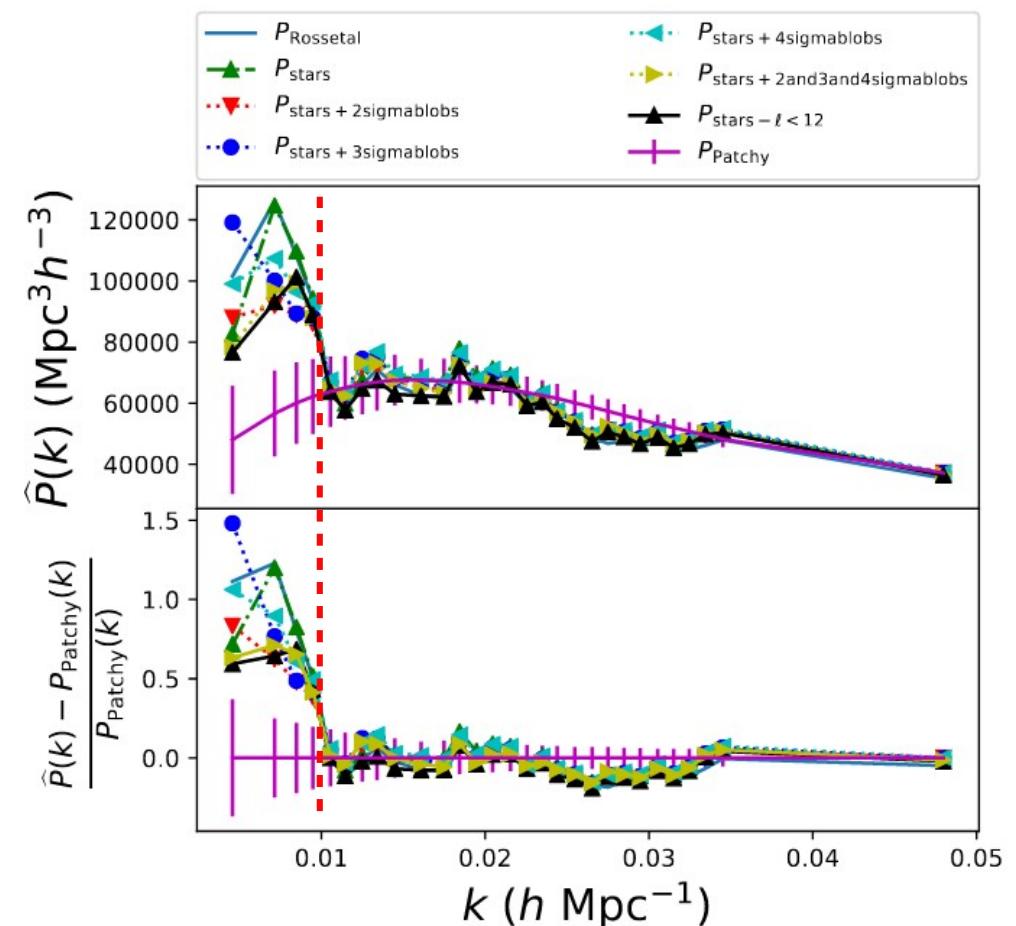
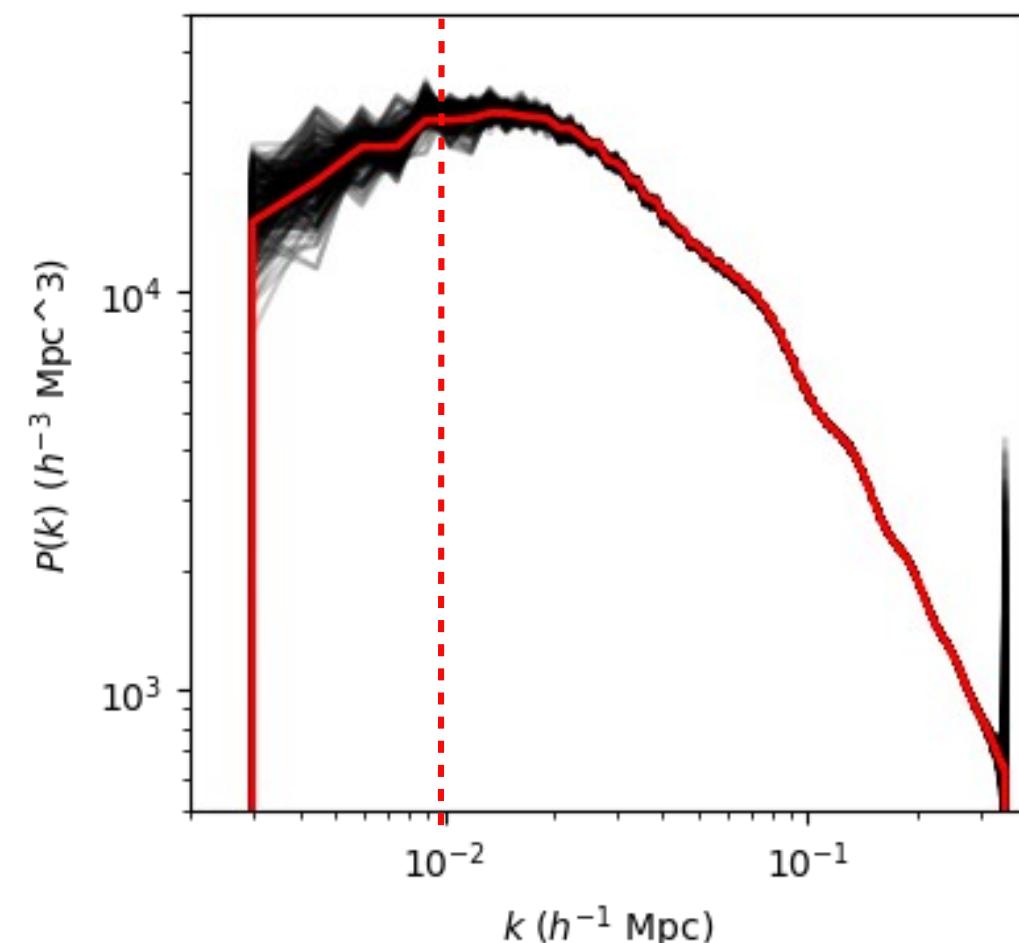
Inference results: density and $P(k)$

Preliminary



Inference results: density and $P(k)$

Preliminary



Application to CMB lensing

Preliminary

Estimated error from
MCMC

Convergence

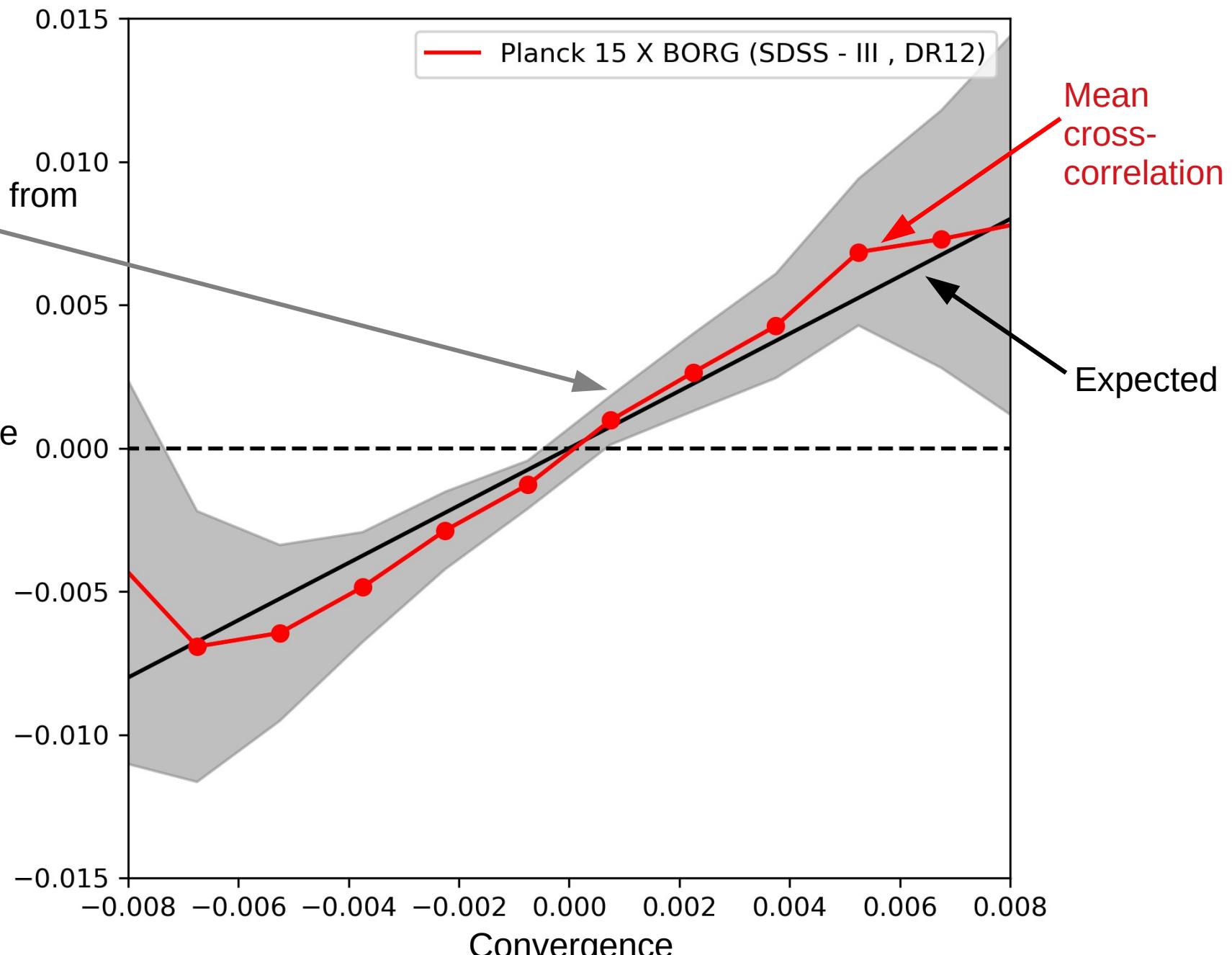
K_{Planck}

Planck 15 X BORG (SDSS - III , DR12)

Mean
cross-
correlation

Expected

$K_{\text{BORG / CMASS}}$



More modeling efforts



© AlltheSky.com

Highlights of the Aquila constellation

Aquila

Aquila

Universe expansion test

Bias model theory

Sunyaev-Zel'dovich

Statistical methods

Lyman- α forest modeling

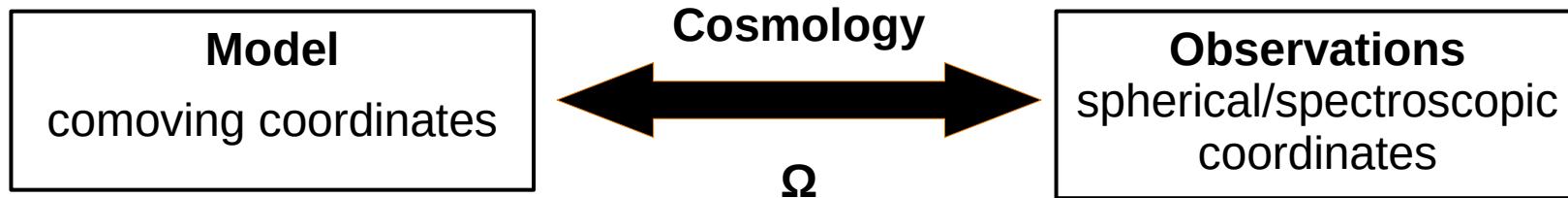
Bias model with Neural network

Fast simulation with Neural network

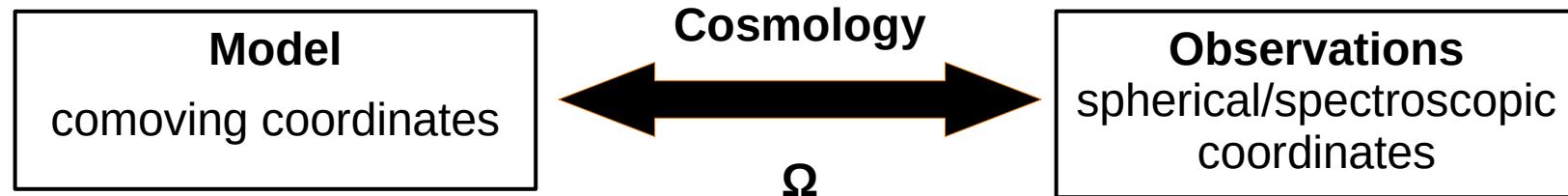
Distance data modeling/Velocity inference

Fifth force, Magnetic field, Gravitational lensing, photo-z...

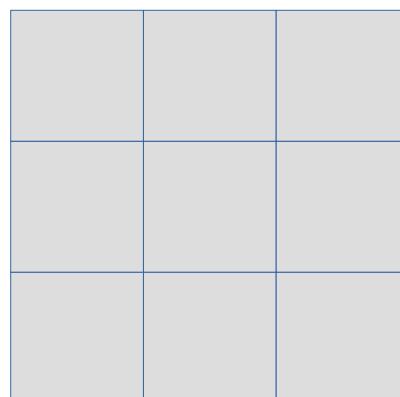
Use expansion of Universe (A/P)



Use expansion of Universe (A/P)

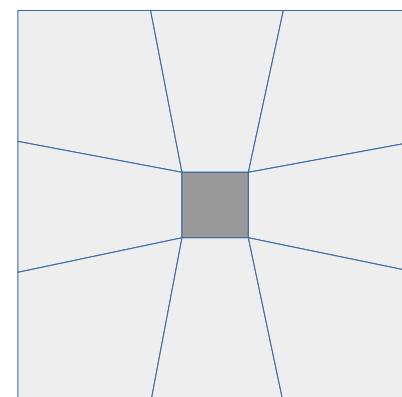


Added in BORG as density remapping:



Comoving
coordinates

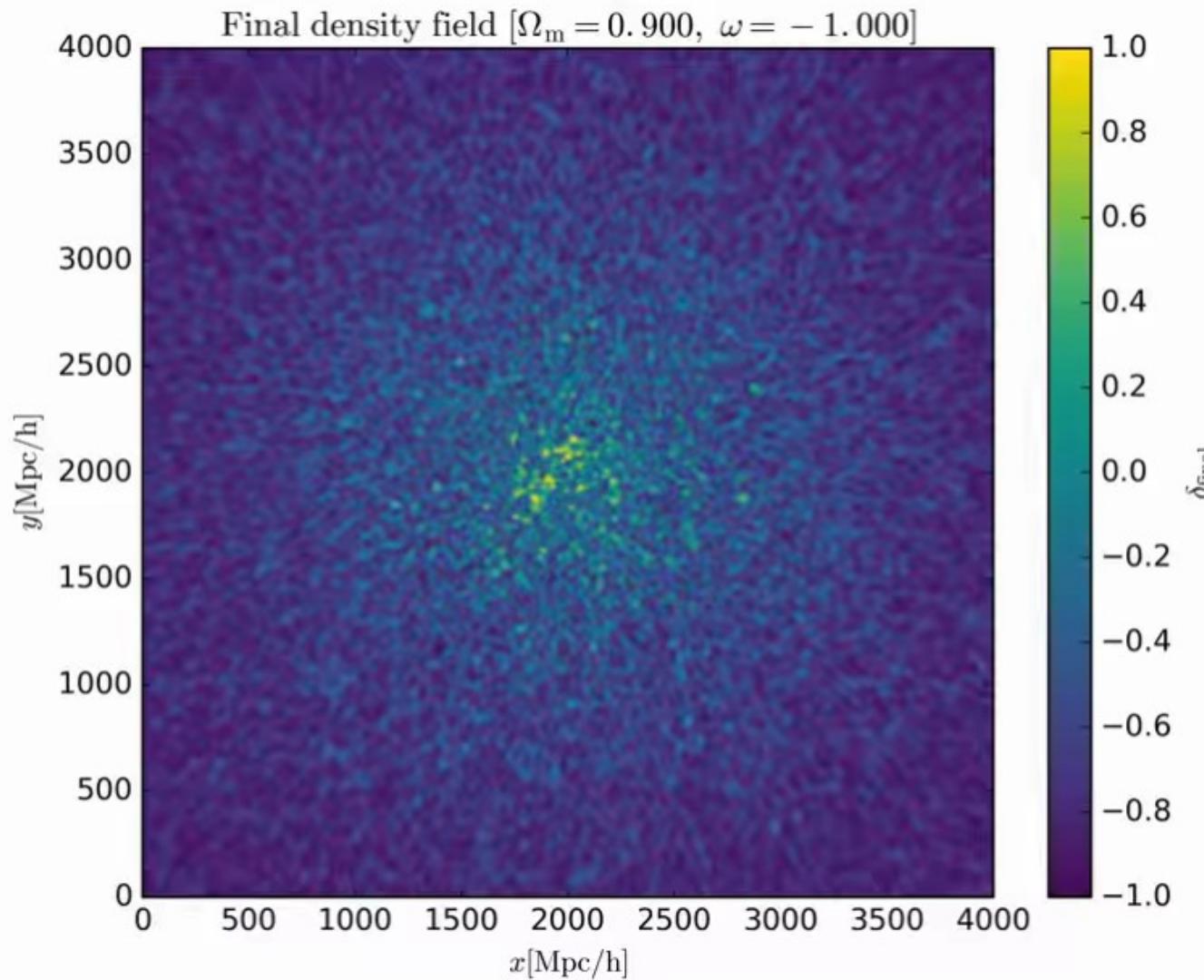
$$\vec{x}$$



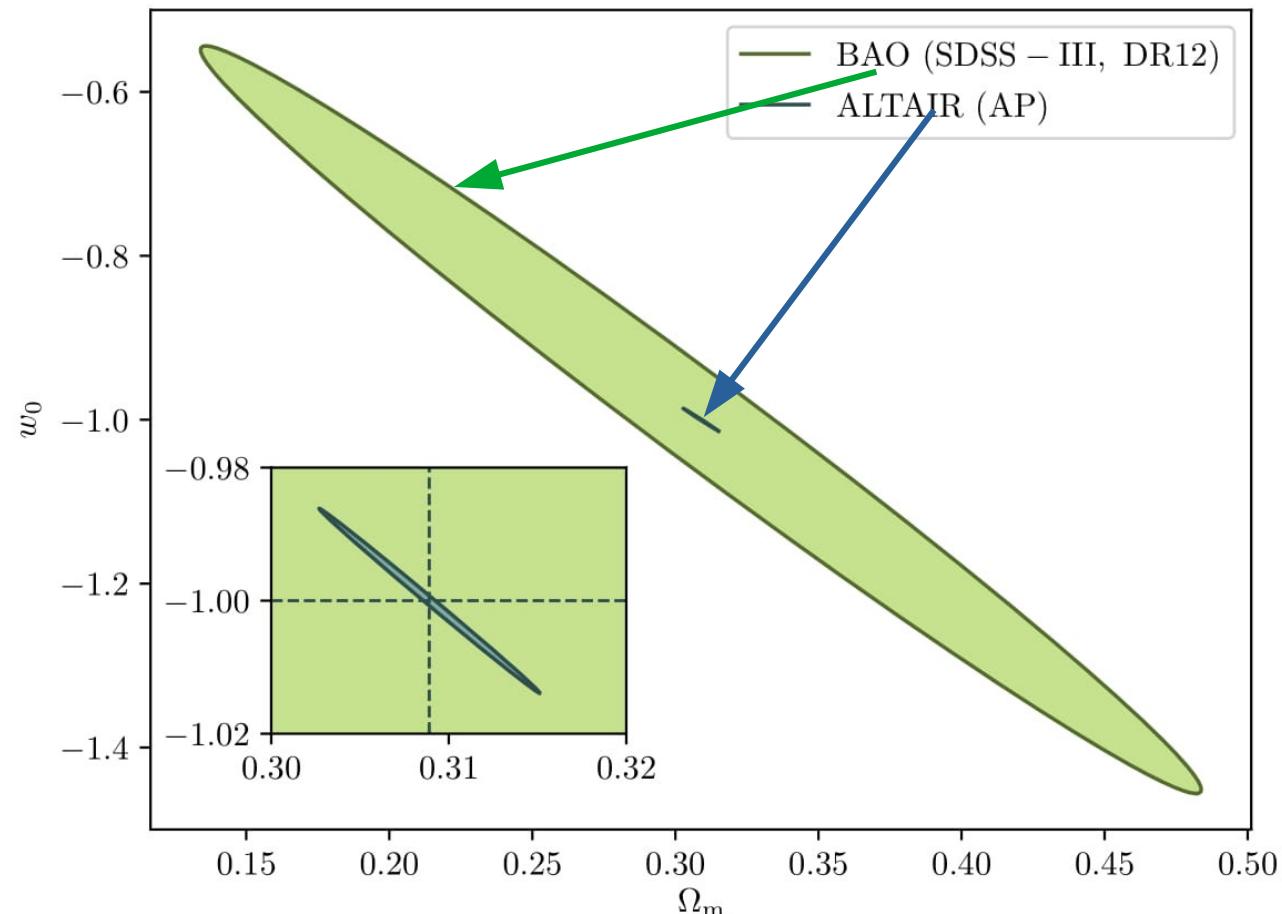
Scaled redshift
coordinates

$$\vec{z}_i = \frac{c}{H_0} z_i \hat{u}_i$$

A/P: dependence on cosmology



A/P: mock test



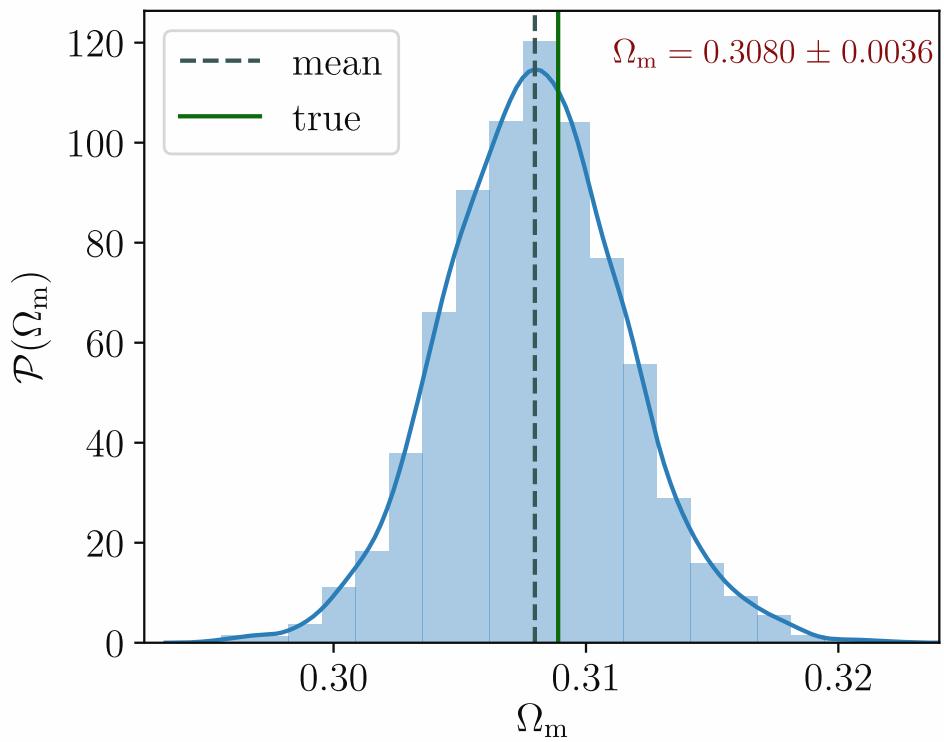
Source of additional information:
complete use of all the modes, and high order statistics

A/P: mock test

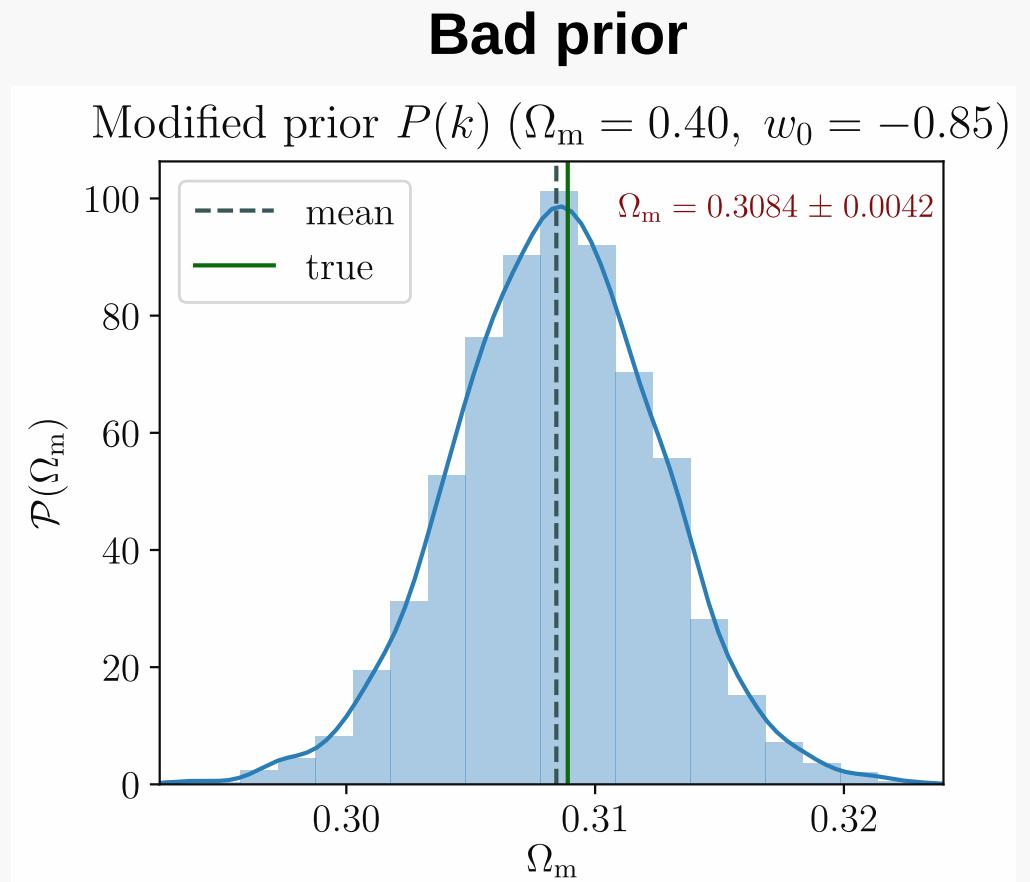


Constraints resilient to isotropic prior biases

Good prior



Bad prior



Highlights of the Aquila constellation

Sunyaev-Zel'dovich

Aquila

Universe expansion test

Bias model theory

Statistical methods

Lyman- α forest modeling

Bias model with Neural network

Fast simulation with Neural network

Distance data modeling/Velocity inference

Fifth force, Magnetic field, Gravitational lensing, photo-z...

Aquila



Distance data/velocity inference

Cosmic flows from observed distances

$$cz \simeq H d + v_r$$

Spectroscopic redshift

Supernovae Ia
Tully-Fisher
Fundamental plane
...



Distance data/velocity inference

Cosmic flows from observed distances

$$cz \simeq H d + v_r$$

Spectroscopic redshift

Supernovae Ia
Tully-Fisher
Fundamental plane
...

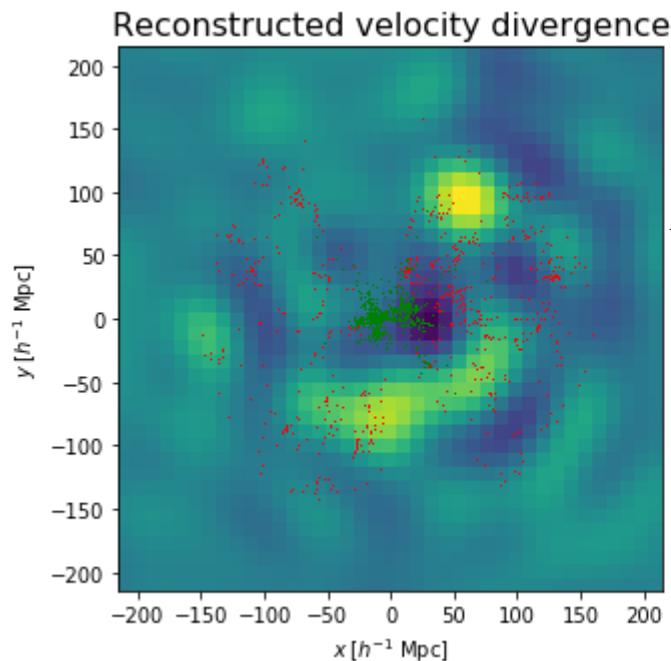
Model include:

- flexible distance prior
- redshift cut selection
- mixture error distribution
- Gaussian random field cosmic velocity field
- zero point calibration

Stochastic Velocity model

Application: 6dFv + Spitzer data (from CosmicFlows 3 database, Tully et al. 2016)

VIRBIUS2: inferred maps



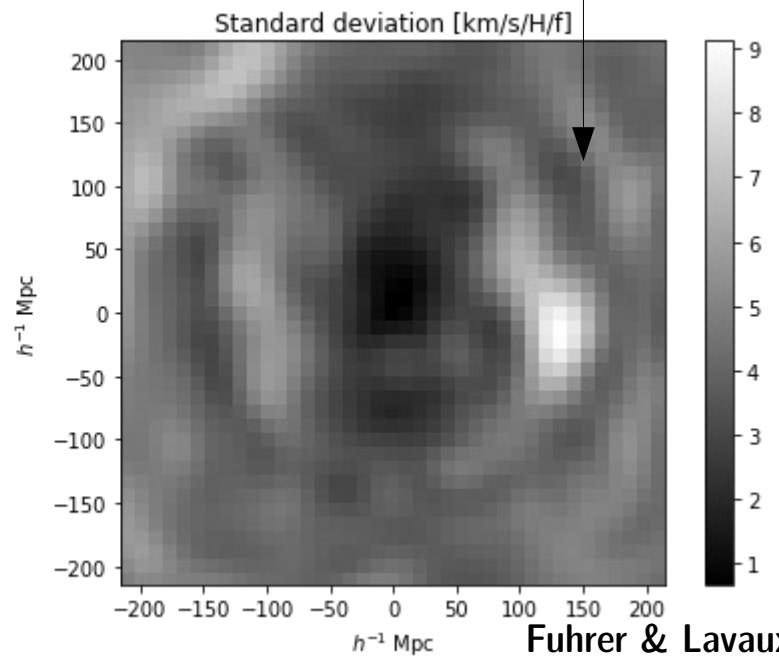
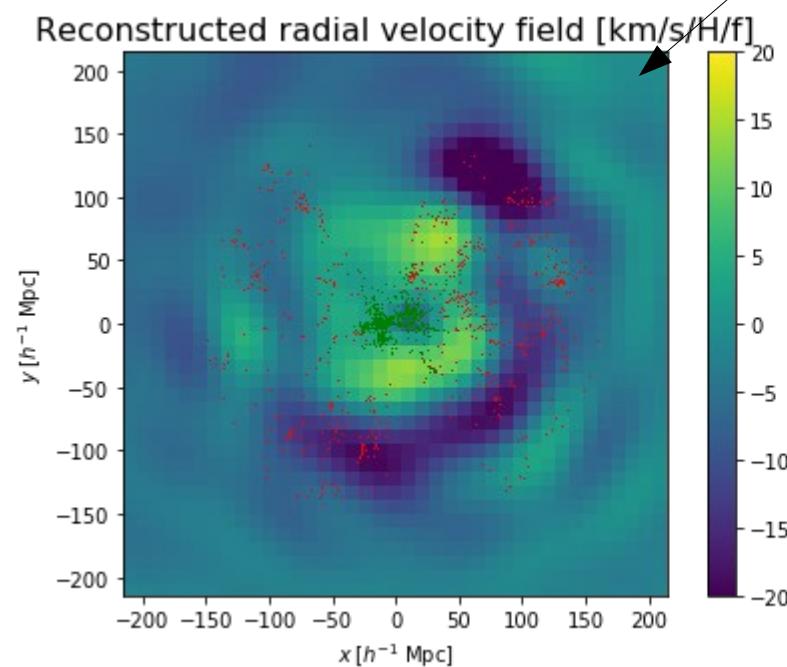
6dFv

Spitzer

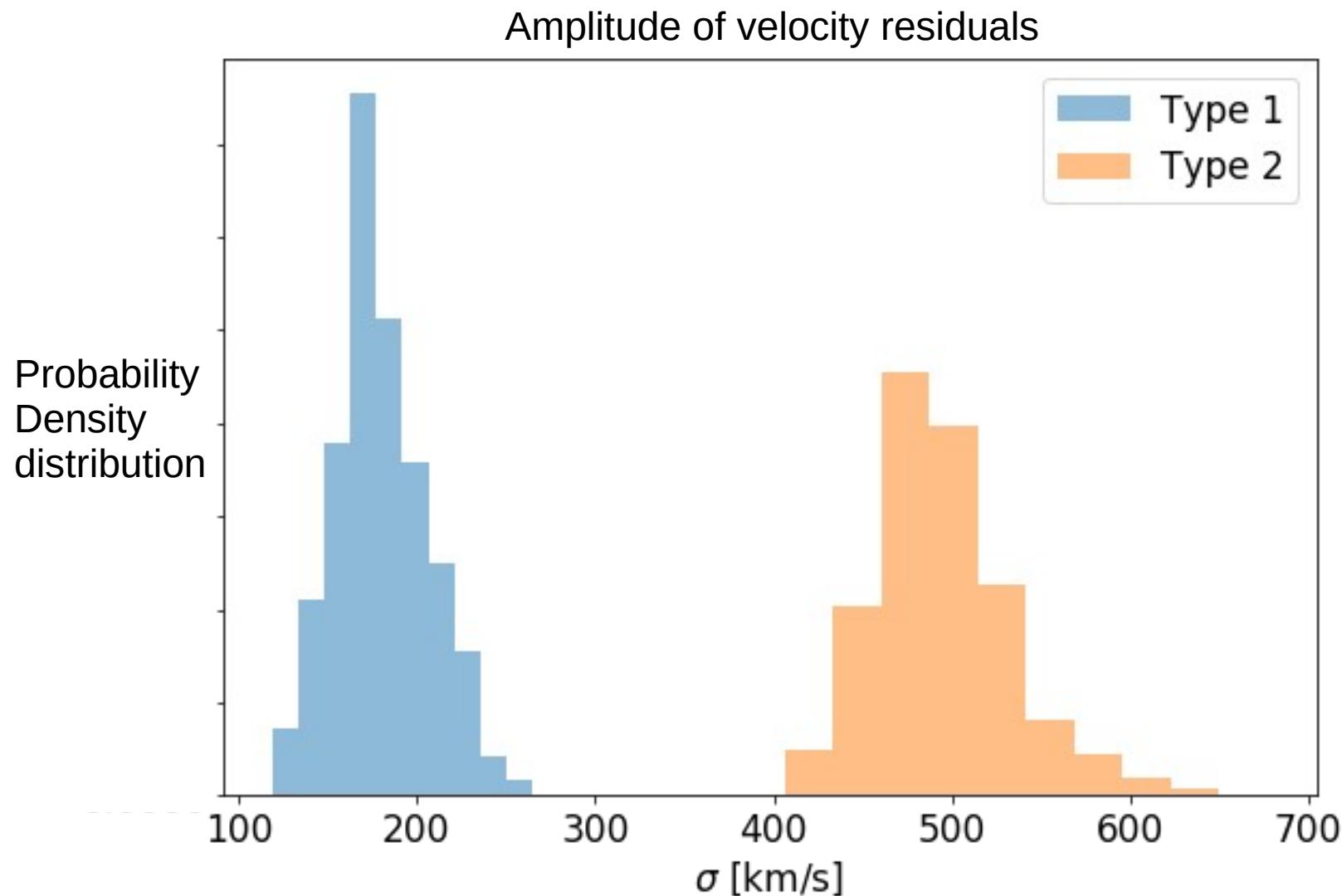
~Density

Mean radial velocity

Estimated error



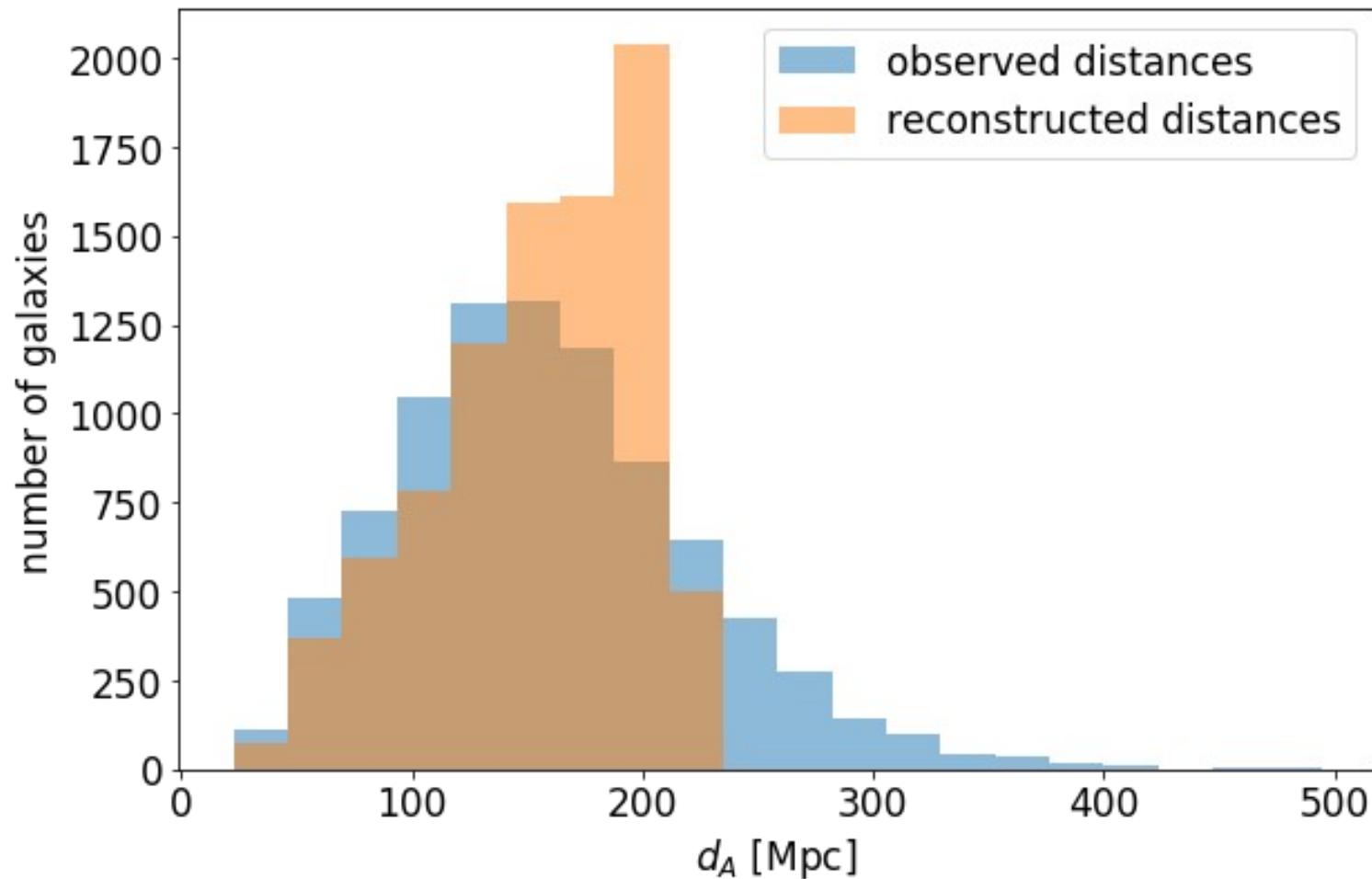
VIRBIUS2: mixture error distribution



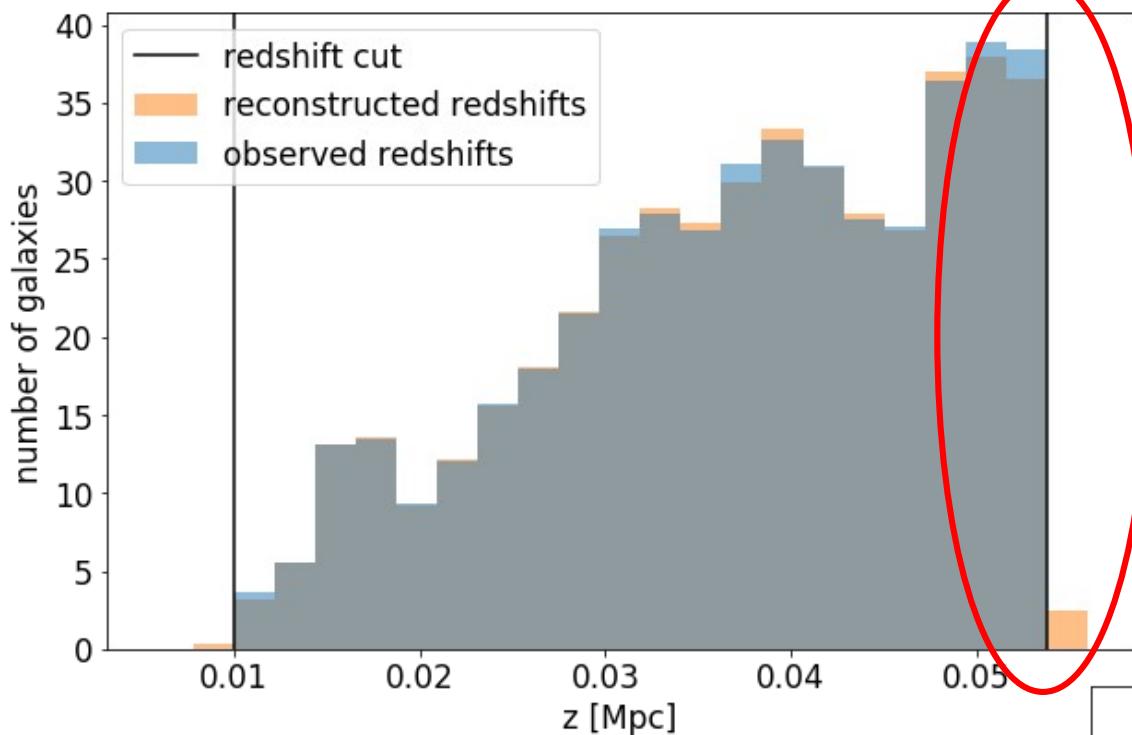
VIRBIUS2 distances



Inferred vs raw observed distances



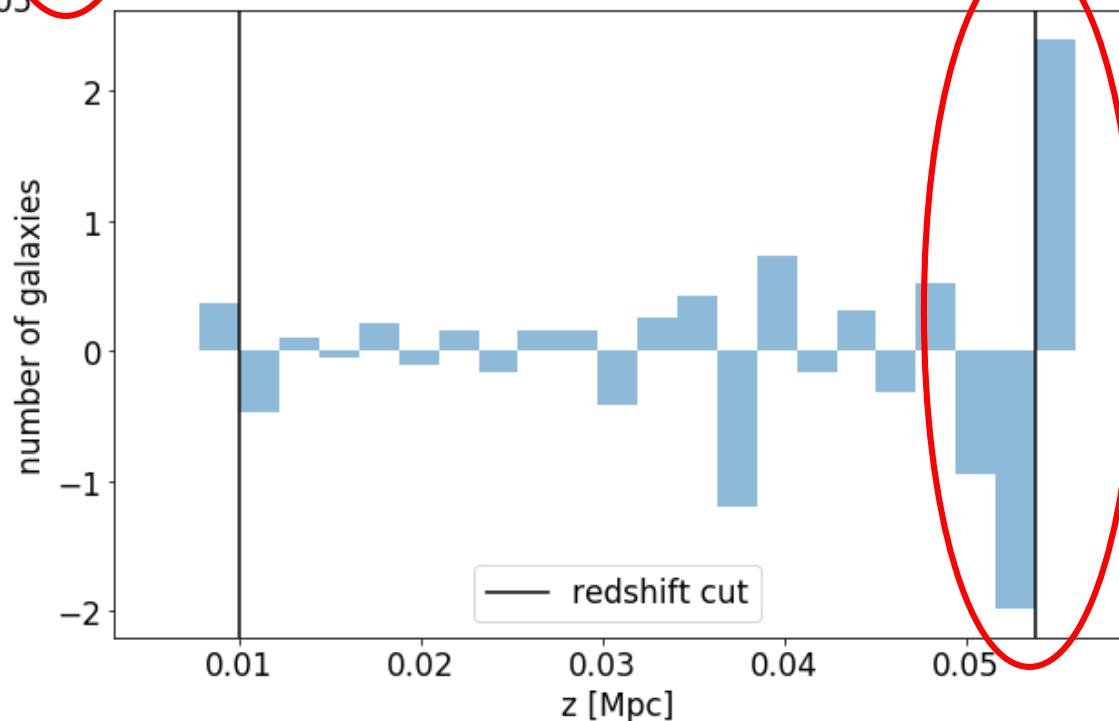
VIRBIUS2 redshifts



Full redshift distribution
before/after inference

Selection boundary effects

Distribution
differences



Highlights of the Aquila constellation

Sunyaev-Zel'dovich

Aquila

Universe expansion test

Bias model theory

Statistical methods

Lyman- α forest modeling

Bias model with Neural network

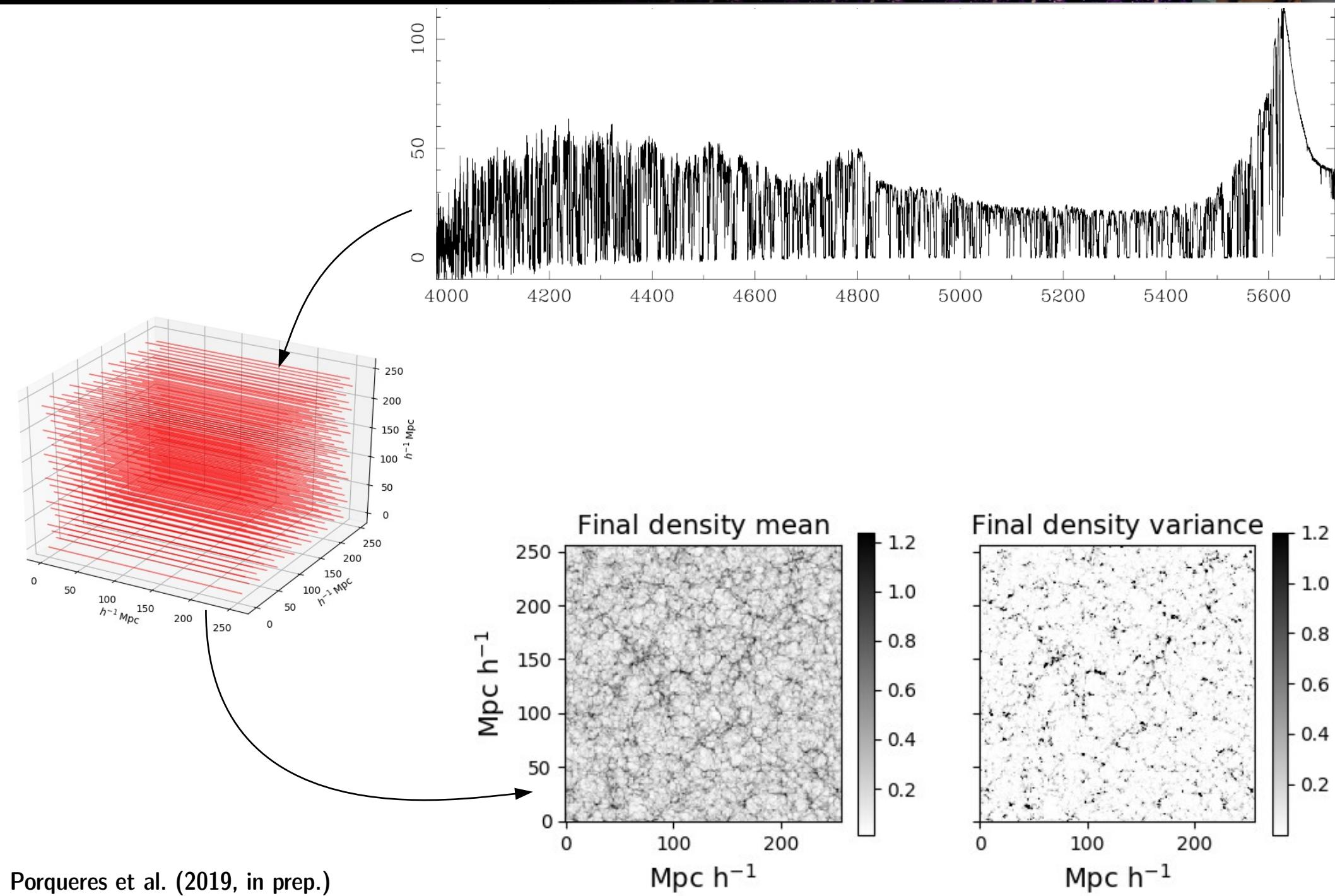
Fast simulation with Neural network

Distance data modeling/Velocity inference

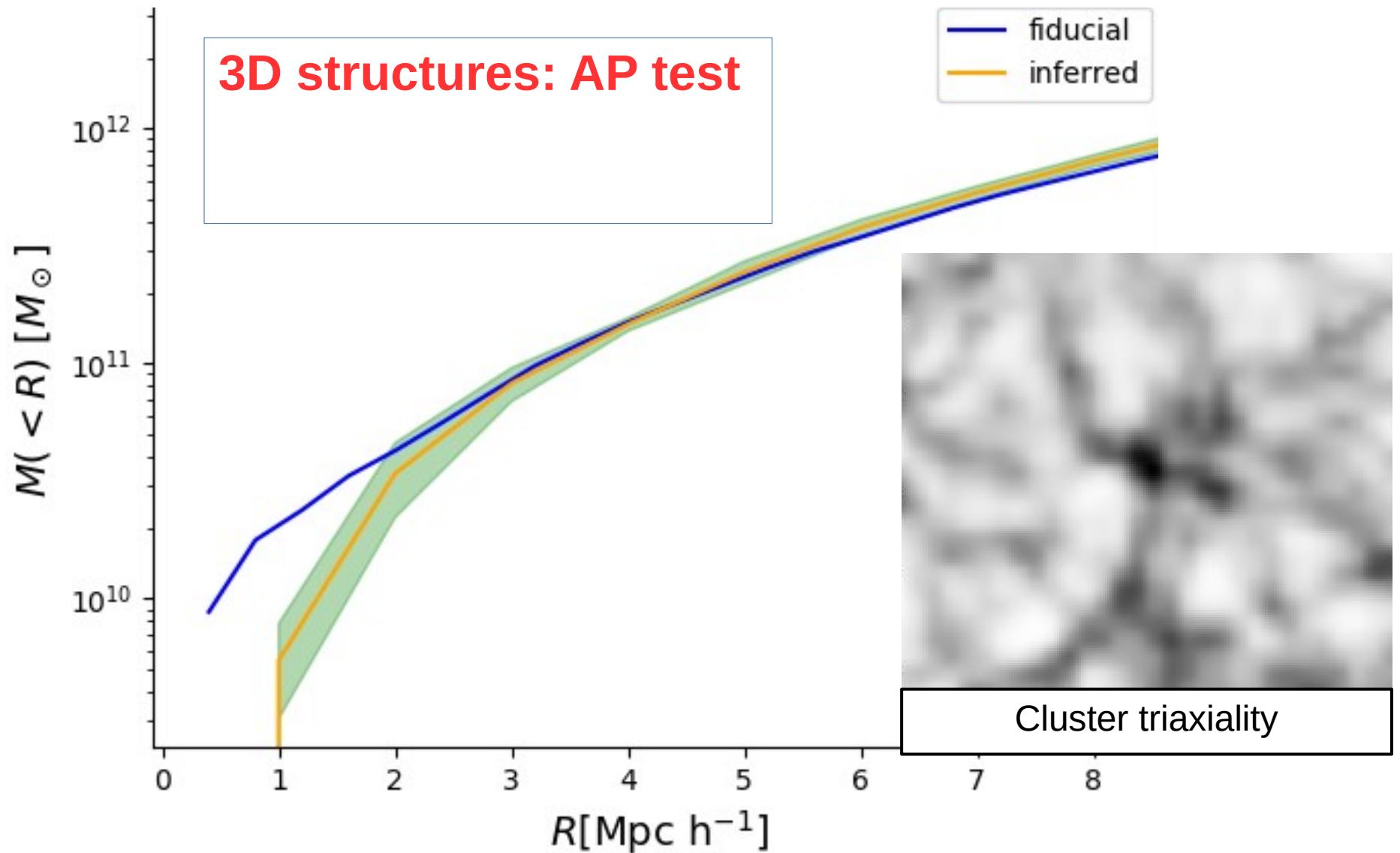
Fifth force, Magnetic field, Gravitational lensing, photo-z...

Aquila

Lyman- α forest: generic inference



Lyman- α forest: cluster mass profile



Highlights of the Aquila constellation

Sunyaev-Zel'dovich

Aquila

Universe expansion test

Bias model theory

Statistical methods

Lyman- α forest modeling

Bias model with Neural network

Fast simulation with Neural network

Distance data modeling/Velocity inference

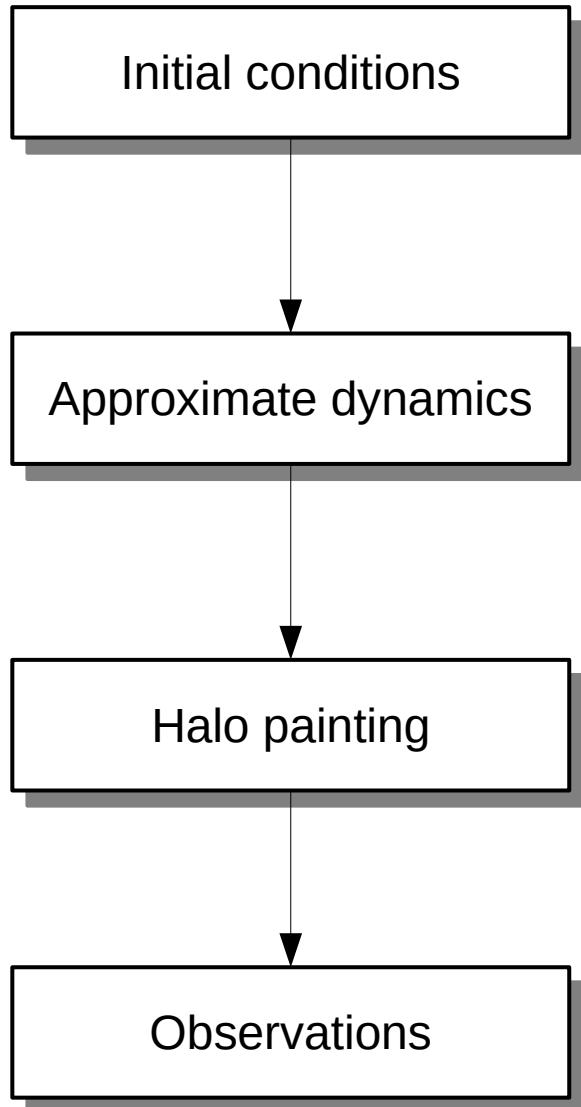
Fifth force, Magnetic field, Gravitational lensing, photo-z...

Aquila

Neural networks painting



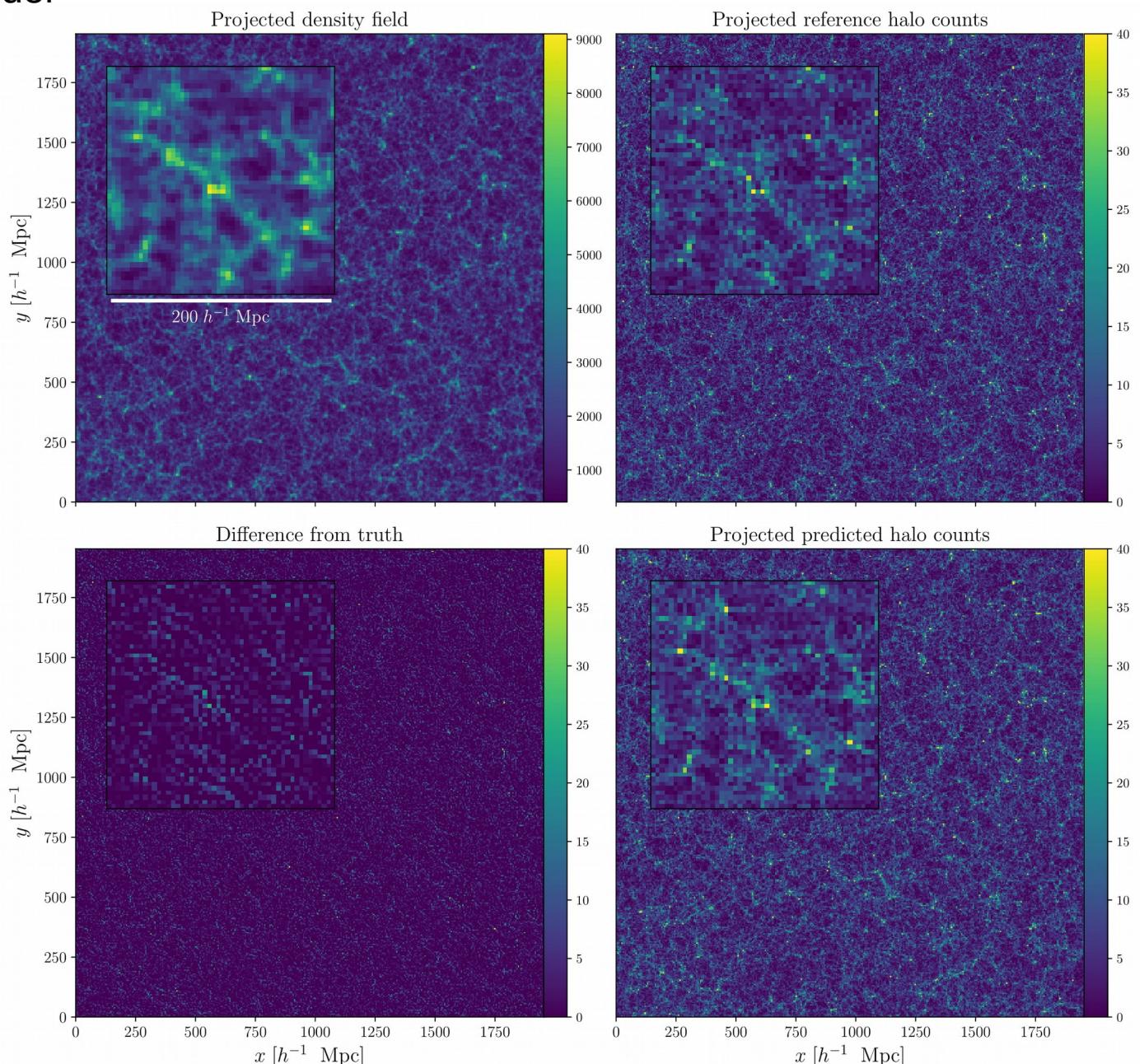
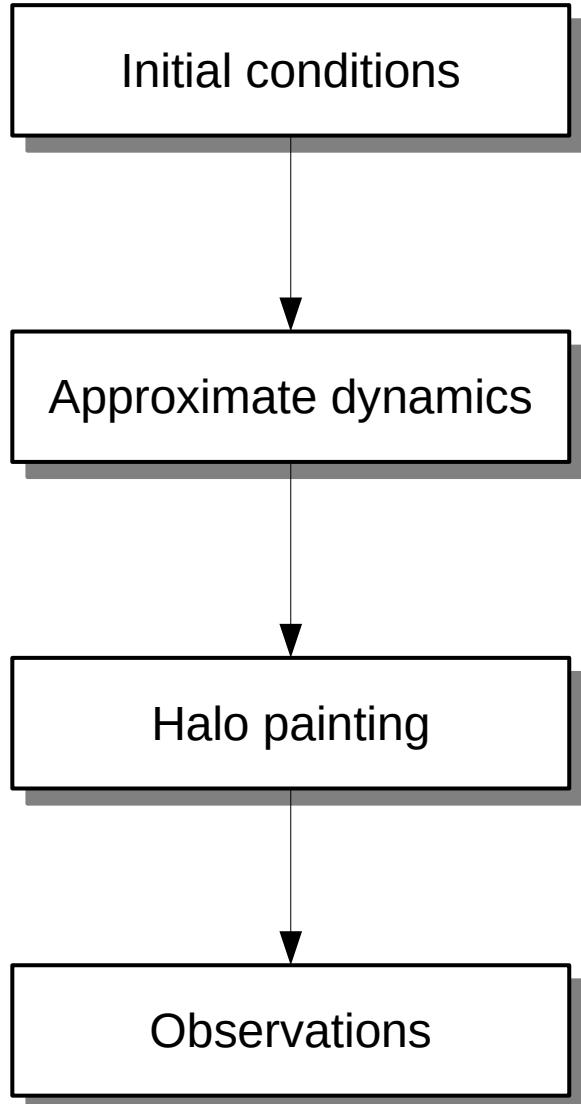
BORG large-scale-structure model



Neural networks painting



BORG large-scale-structure model



Highlights of the Aquila constellation

Sunyaev-Zel'dovich

Aquila

Universe expansion test

Bias model theory

Statistical methods

Lyman- α forest modeling

Bias model with Neural network

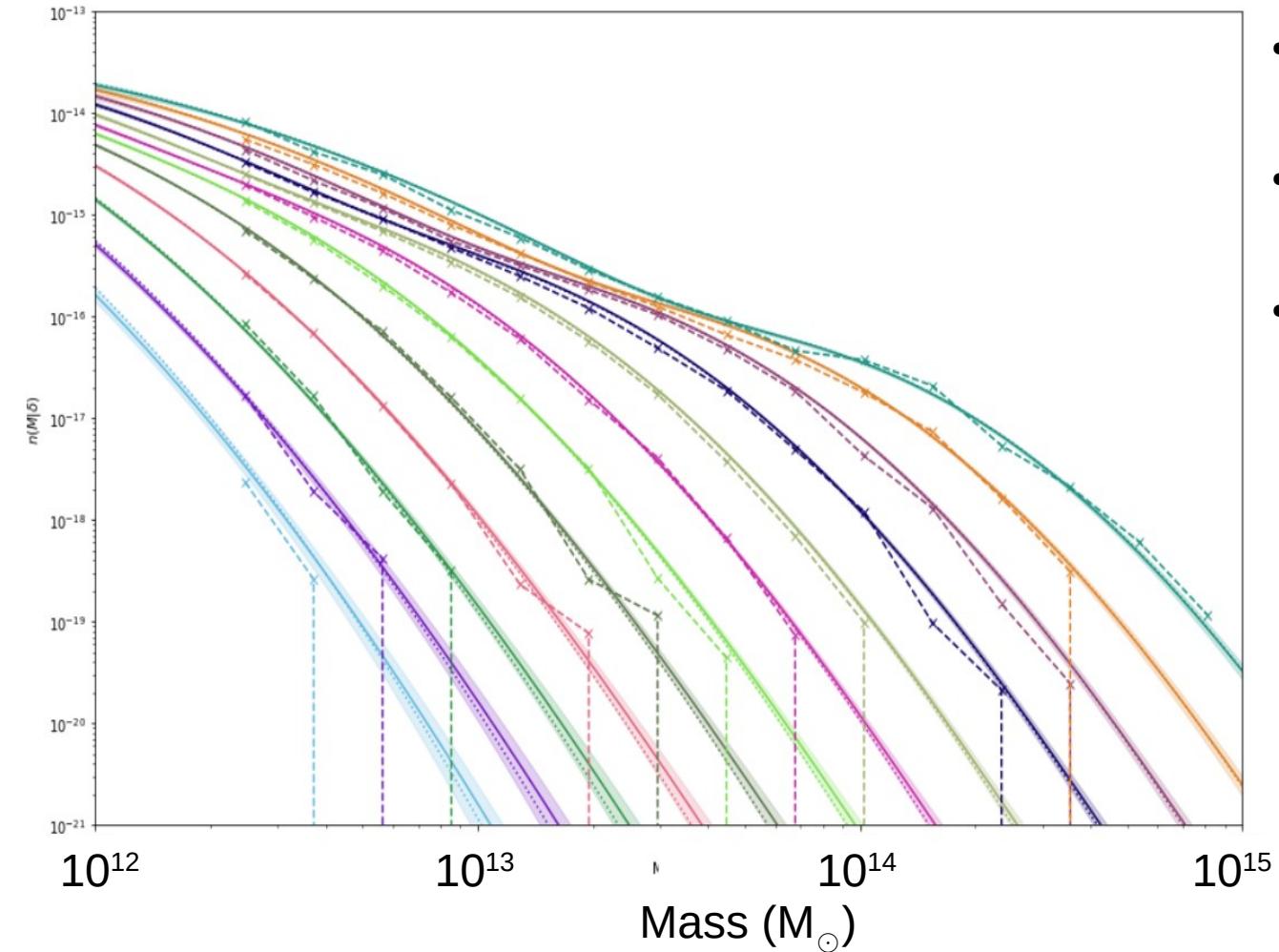
Fast simulation with Neural network

Distance data modeling/Velocity inference

Fifth force, Magnetic field, Gravitational lensing, photo-z...

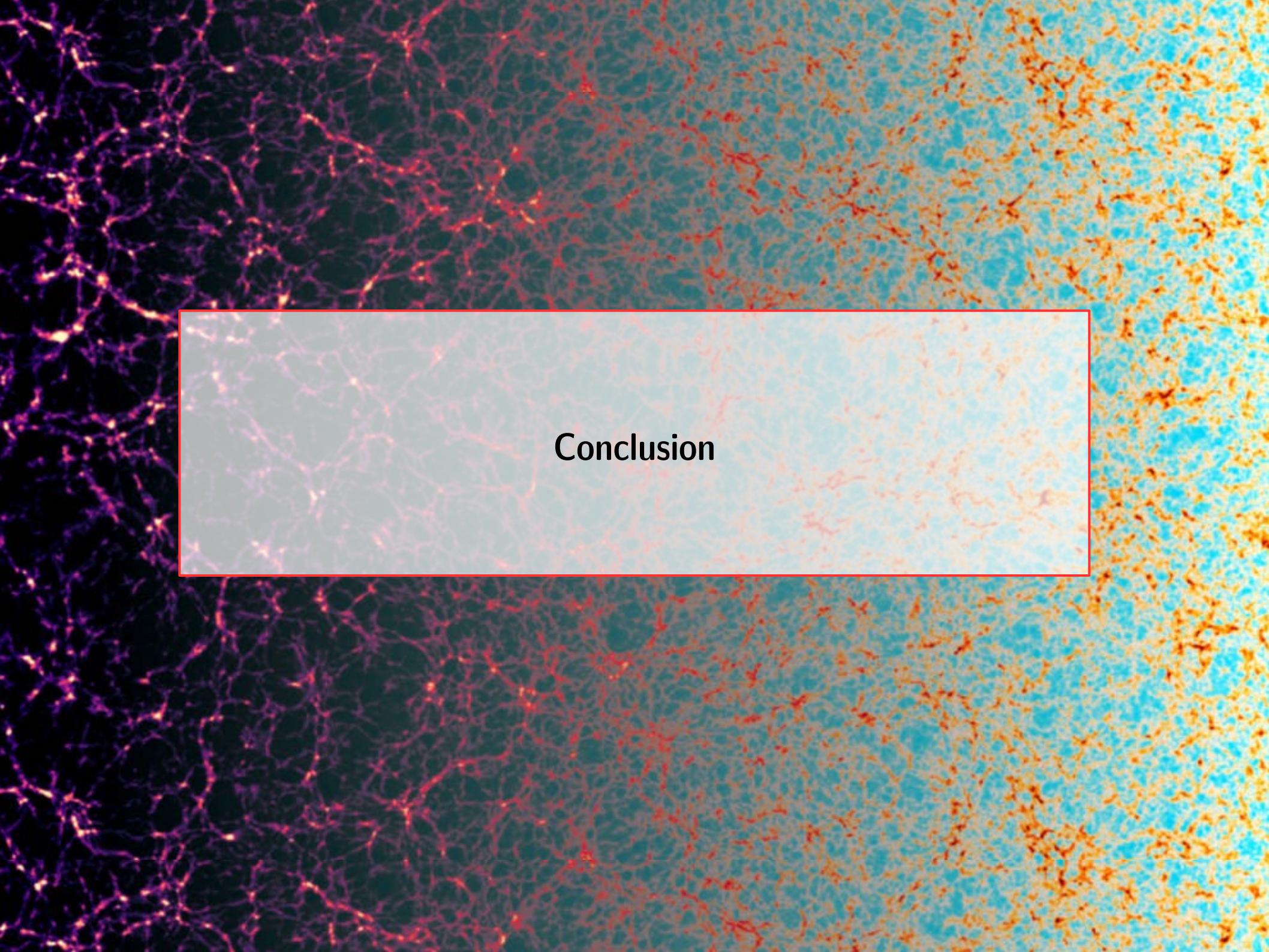
Aquila

Neural networks inference



- Exchange binning for distribution function
- Allow neural network to capture non-local Information
- Attached to the BORG machine

(lots of other work on IMNN, DELFI, N-body emulator...)



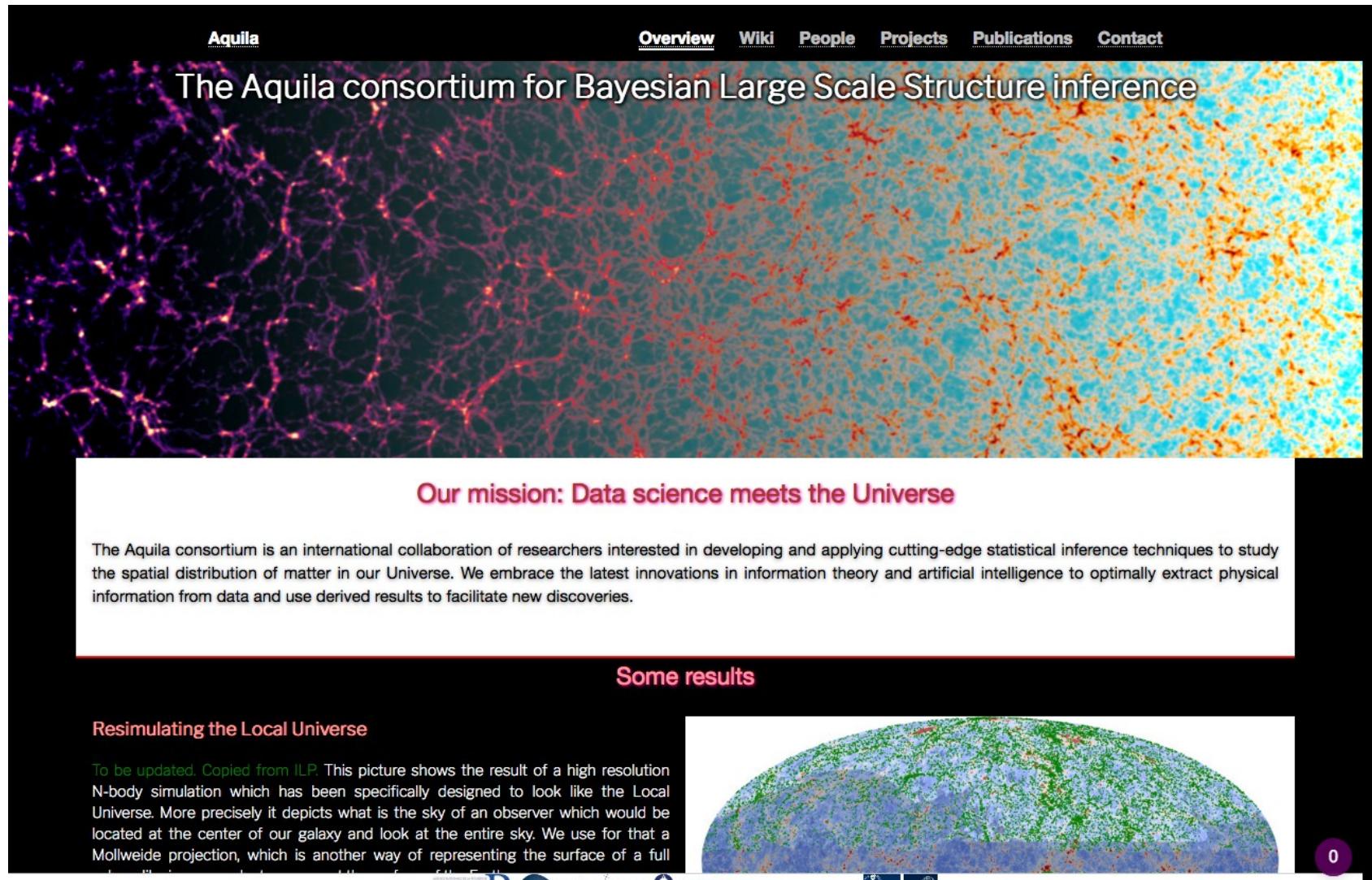
Conclusion

The Aquila consortium

Aquila

- Founded in 2016
- Gather people interested in working with each other on developing the Bayesian pipelines and run analysis on data.

<https://aquila-consortium.org/>



The screenshot shows the homepage of the Aquila consortium website. At the top, there is a navigation bar with links to Overview, Wiki, People, Projects, Publications, and Contact. Below the navigation bar is a large banner featuring a simulation of the large-scale structure of the universe, transitioning from purple to red to yellow. The text "The Aquila consortium for Bayesian Large Scale Structure inference" is displayed above the simulation. A central white box contains the text "Our mission: Data science meets the Universe". Below this box, a paragraph describes the consortium's mission: "The Aquila consortium is an international collaboration of researchers interested in developing and applying cutting-edge statistical inference techniques to study the spatial distribution of matter in our Universe. We embrace the latest innovations in information theory and artificial intelligence to optimally extract physical information from data and use derived results to facilitate new discoveries." Further down the page, there is a section titled "Some results" with a sub-section titled "Resimulating the Local Universe". A text box states "To be updated. Copied from ILP." followed by a detailed description of the simulation. To the right of this text is a Mollweide projection map of the sky, showing a distribution of points in green, blue, and red. At the bottom of the page, there are logos for ANR, CNRS, ARI, SORBONNE UNIVERSITÉ, Institut d'Astrophysique de Paris, Institut Élie Cartan de Lorraine, University of Oxford, and University of Birmingham. A small purple circle with the number "0" is located in the bottom right corner.

Aquila

Overview Wiki People Projects Publications Contact

The Aquila consortium for Bayesian Large Scale Structure inference

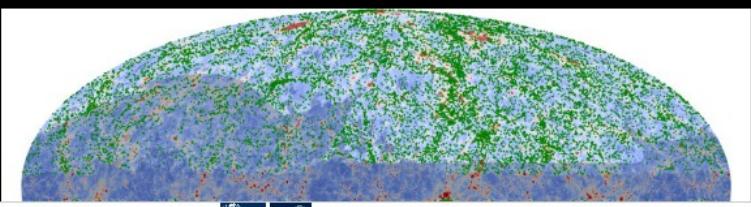
Our mission: Data science meets the Universe

The Aquila consortium is an international collaboration of researchers interested in developing and applying cutting-edge statistical inference techniques to study the spatial distribution of matter in our Universe. We embrace the latest innovations in information theory and artificial intelligence to optimally extract physical information from data and use derived results to facilitate new discoveries.

Some results

Resimulating the Local Universe

To be updated. Copied from ILP. This picture shows the result of a high resolution N-body simulation which has been specifically designed to look like the Local Universe. More precisely it depicts what is the sky of an observer which would be located at the center of our galaxy and look at the entire sky. We use for that a Mollweide projection, which is another way of representing the surface of a full sphere.



ANR CNRS ARI SORBONNE UNIVERSITÉ Institut d'Astrophysique de Paris Institut Élie Cartan de Lorraine University of Oxford University of Birmingham

0

The Aquila consortium

Aquila

- Founded in 2016
- Gather people interested in working with each other on developing the Bayesian pipelines and run analysis on data.

<https://aquila-consortium.org/>

A biased list of Aquilians... check the website!



Natalia Porqueres



Minh Nguyen



George Kyriacou



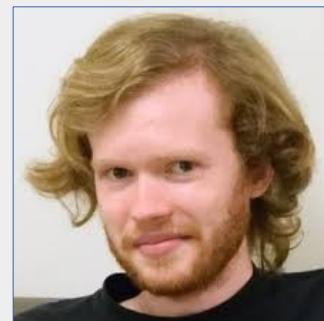
Adam Andrews



Doogesh Kodi Ramanah



Tom Charnock



Harry Desmond



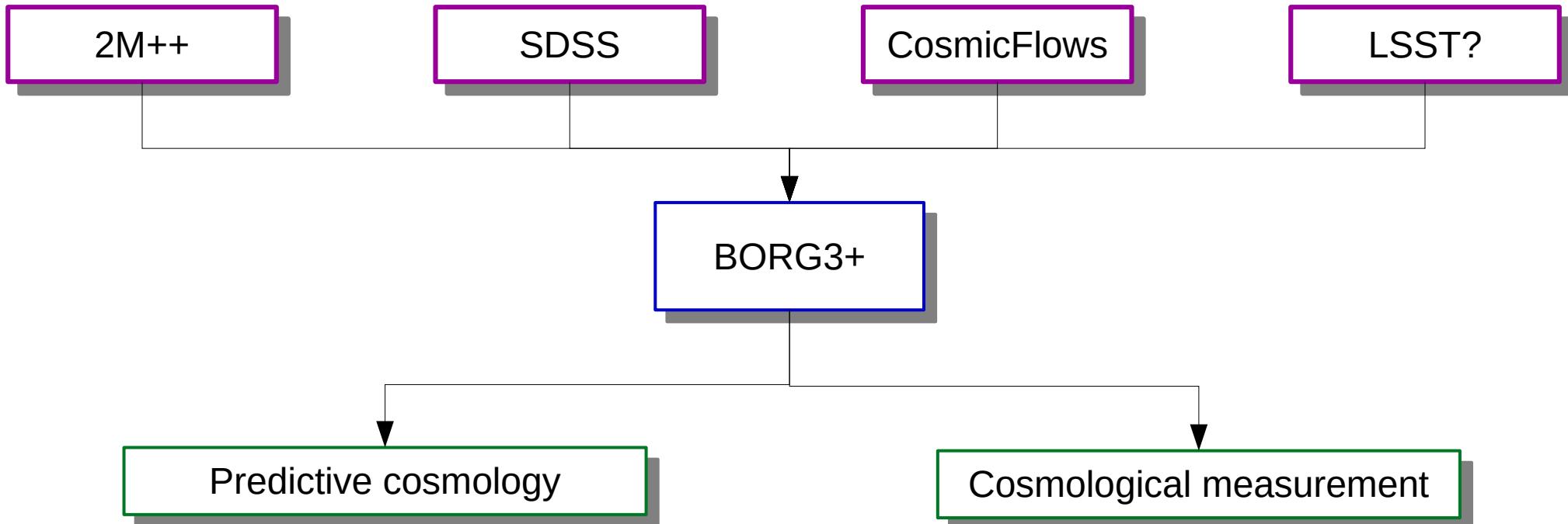
Florian Fuhrer



Florent Leclercq

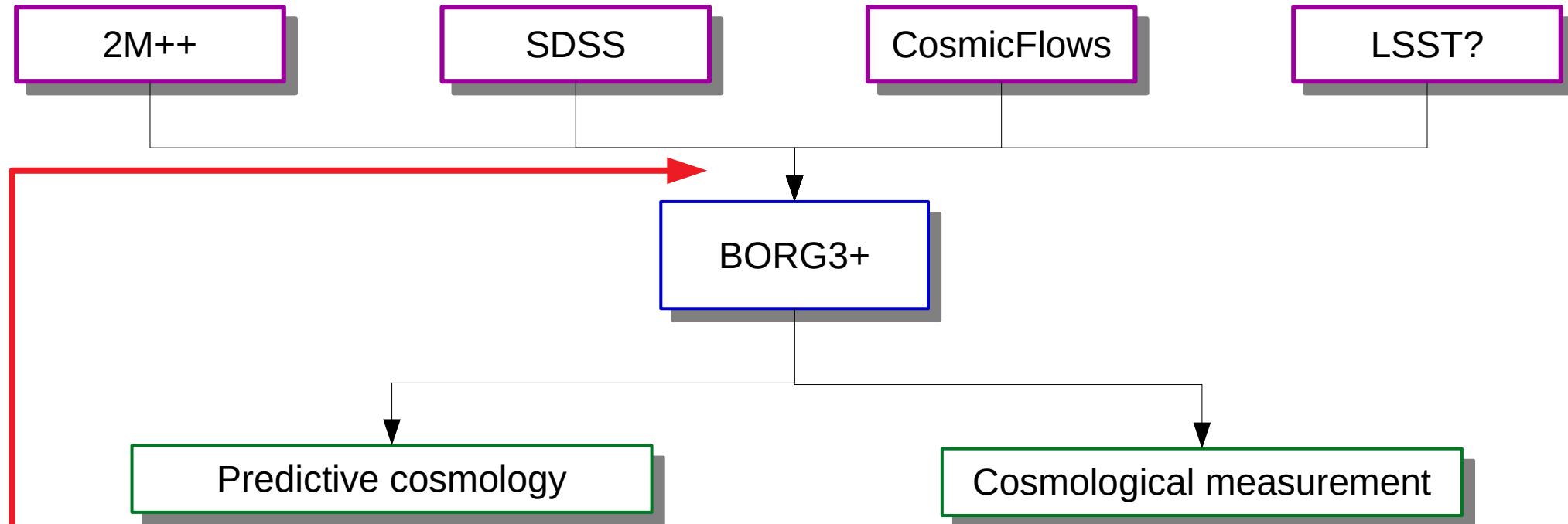


Conclusion: great future



- Velocity field
 - X-ray cluster emission
 - Kinetic Sunyaev Zel'dovich
 - Rees-Sciama
 - Dark matter ?
-
- Cosmic expansion
 - Power spectrum (and governing parameters)
 - Gaussianity tests of initial conditions
 - Direct probe of dynamics

Conclusion: great future and challenges



- Velocity field
- X-ray cluster emission
- Kinetic Sunyaev Zel'dovich
- Rees-Sciama
- Dark matter ?

- Cosmic expansion
- Power spectrum (and governing parameters)
- Gaussianity tests of initial conditions
- Direct probe of dynamics

Galaxy formation: bias and likelihood

Instrument modeling