The Aquila Consortium: Building the ultimate Bayesian machine to interpret cosmological dataset

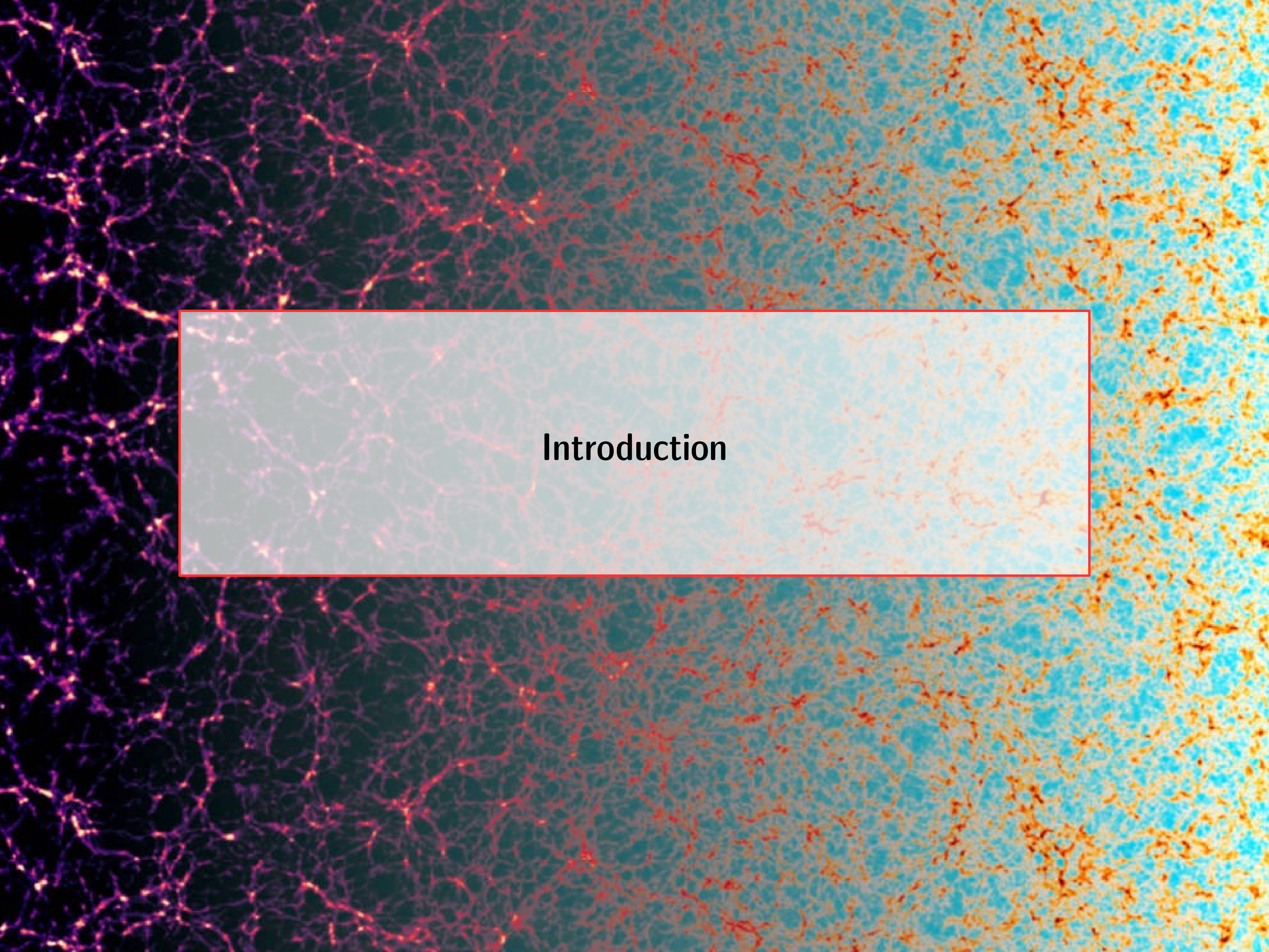
Guilhem Lavaux (IAP/CNRS)
and Aquila Consortium

IHP Trimester - “Statistical inference workshop”

Outline

- Introduction
- The chosen path: embrace the complexity
- Two specific sub-models:
  - Altair (Alcock Pasczyǹ ski test)
  - VIRBIUS2 (Flow inference with distance data)
- The path forward / Conclusion

Introduction



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

The ultimate Bayesian machine ?

The ideal scheme

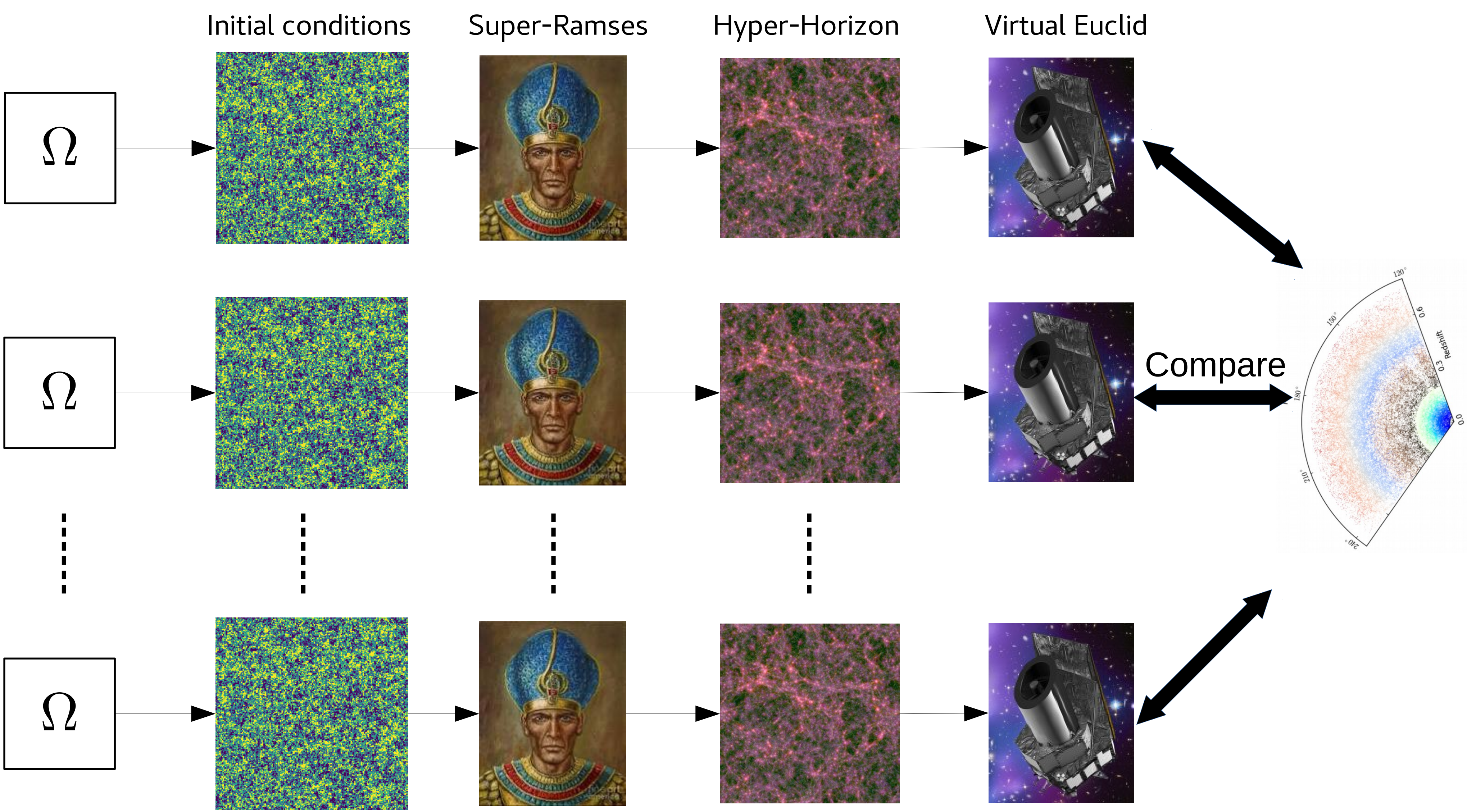
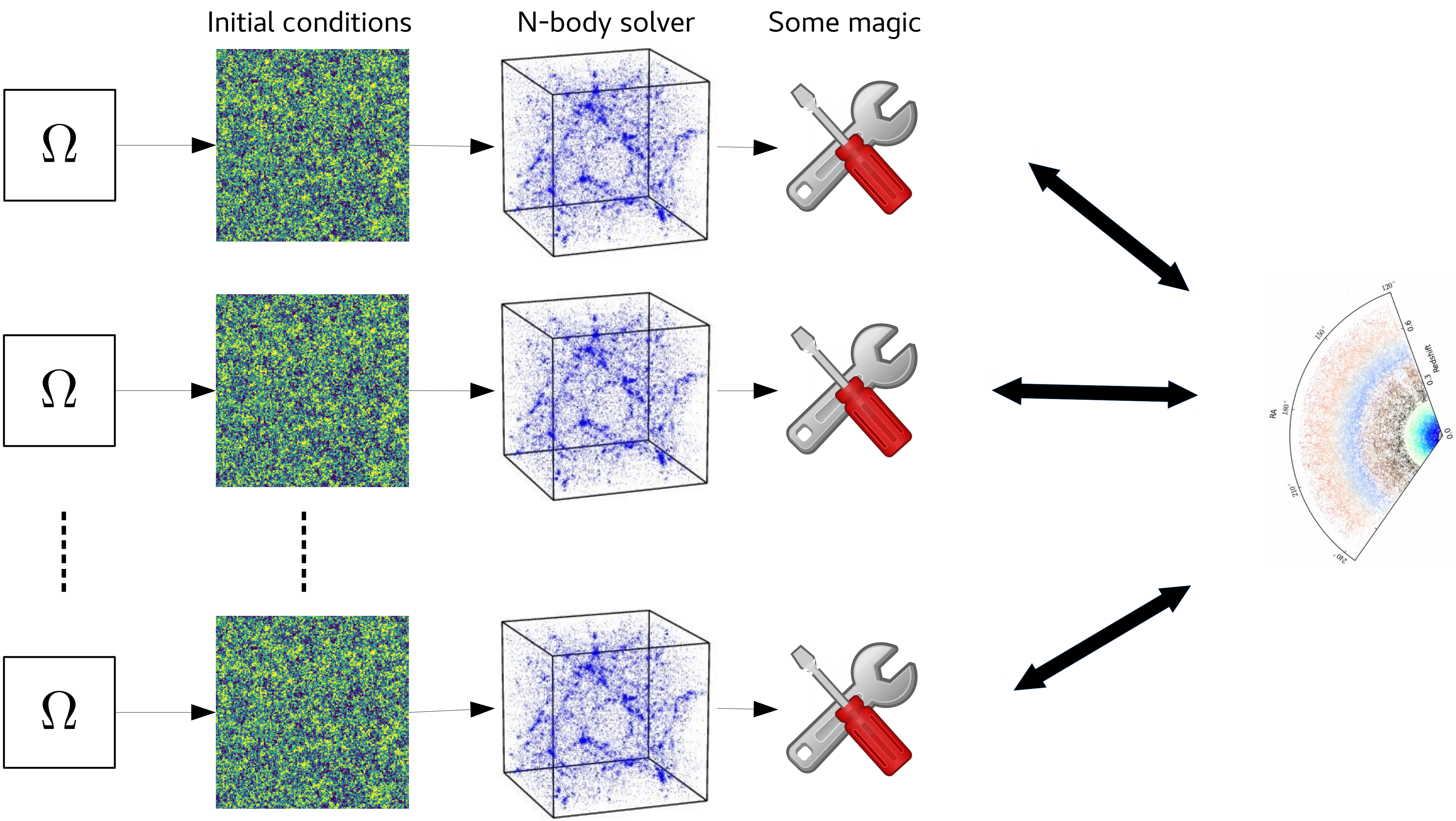


Image credit: Arturas Slapsys, Horizon AGN collaboration

The more pragmatic scheme



Enter...

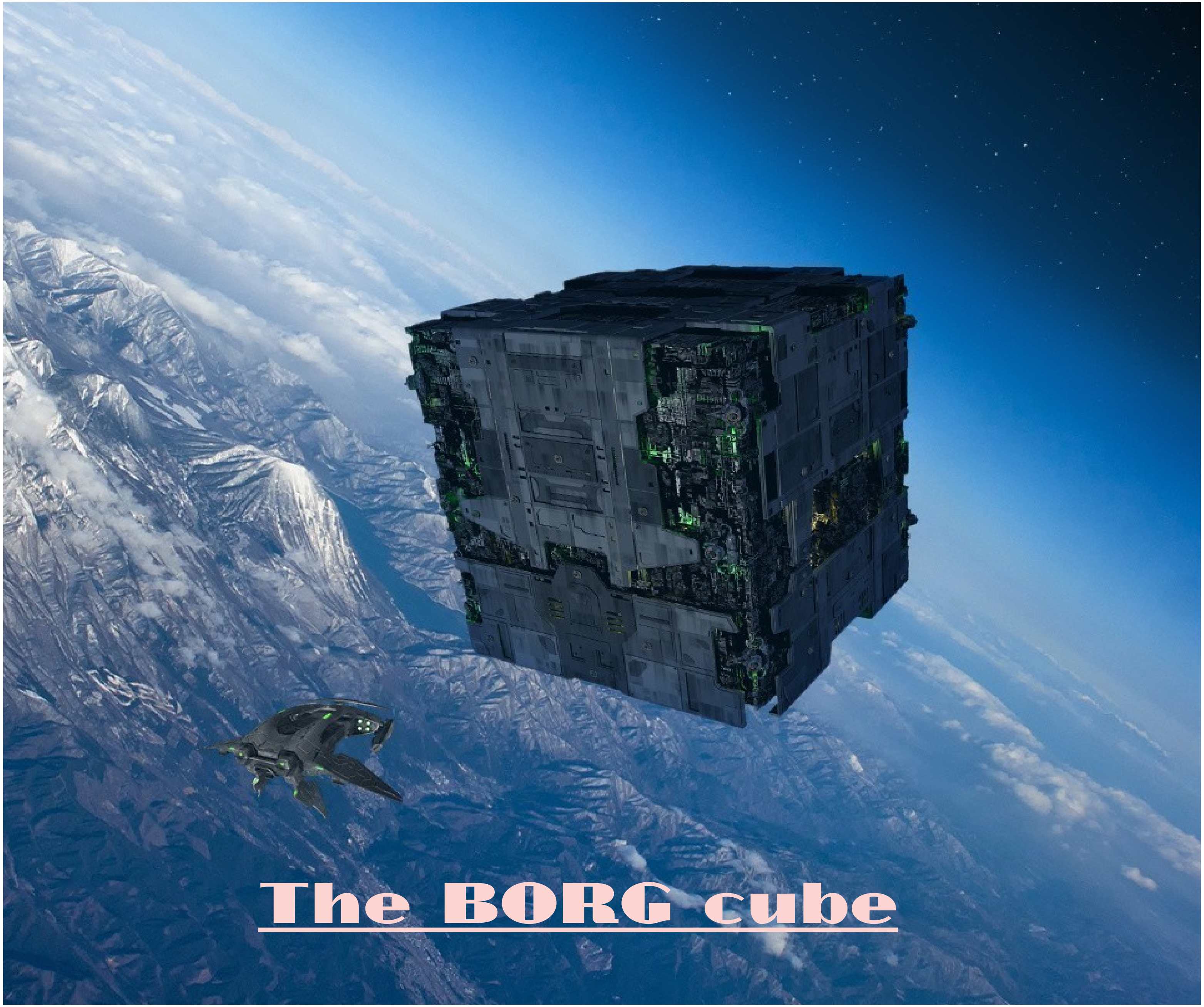
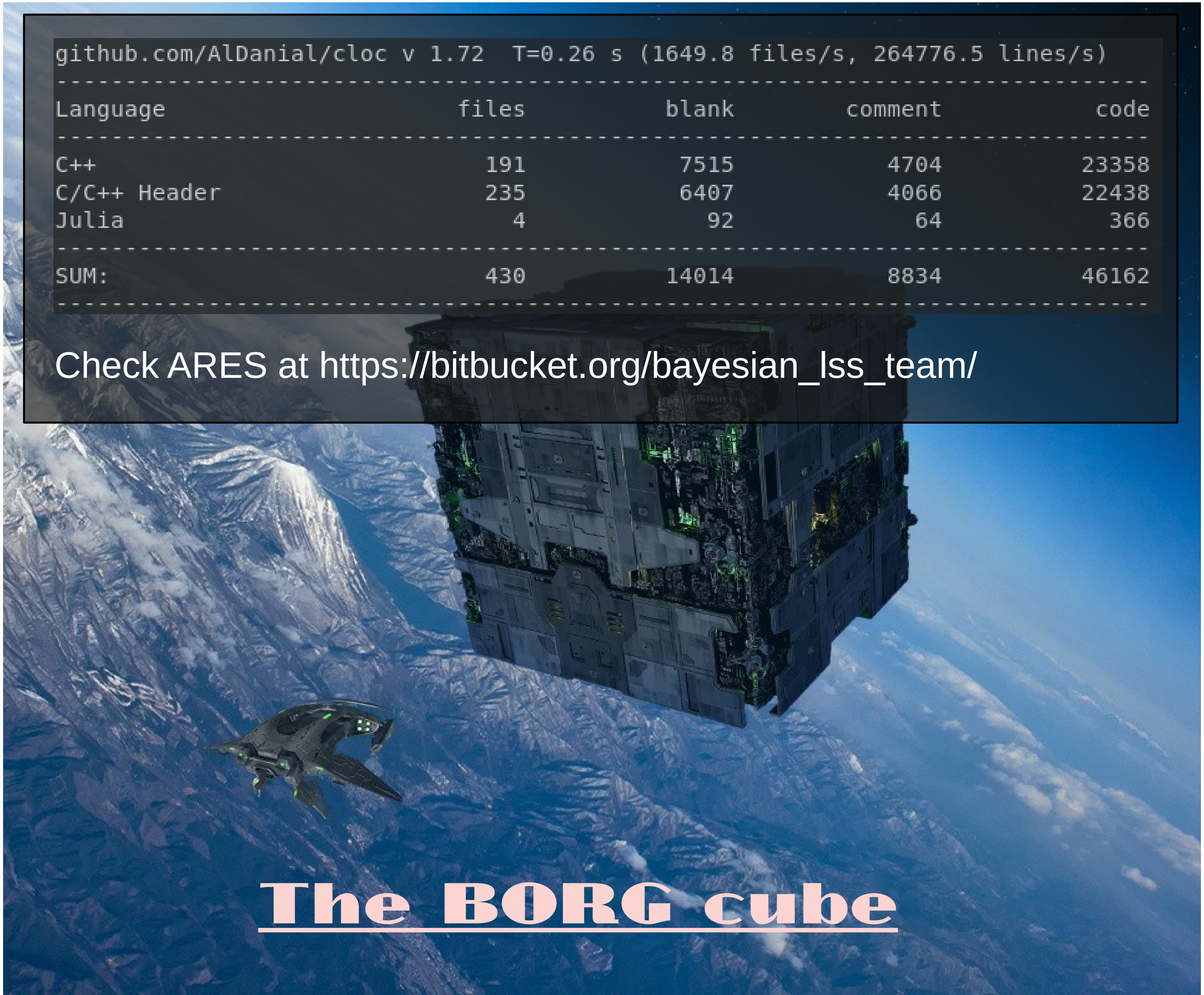


Image credits: Paramount Pictures

Enter...



```
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/

Image credits: Paramount Pictures

The BORG3 inference framework

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

Initial conditions

Jasche et al. (2010), Jasche & Wandelt (2013), Lavaux & Jasche (2016), Jasche & Lavaux (2017, 2018)

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

Jasche et al. (2010), Jasche & Wandelt (2013), Lavaux & Jasche (2016), Jasche & Lavaux (2017, 2018)

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 $\rho_g = \mathcal{B}(\rho_m)$

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

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

Jasche et al. (2010), Jasche & Wandelt (2013), Lavaux & Jasche (2016), Jasche & Lavaux (2017, 2018)

The BORG3 inference framework

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

Initial conditions

Forward and adjoint model

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

Biased galaxy distribution $\rho_g = \mathcal{B}(\rho_m)$

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

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

Jasche et al. (2010), Jasche & Wandelt (2013), Lavaux & Jasche (2016), Jasche & Lavaux (2017, 2018)

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 $\rho_g = \mathcal{B}(\rho_m)$

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

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

Easily exchangeable to try
your favorite differentiable model

Jasche et al. (2010), Jasche & Wandelt (2013), Lavaux & Jasche (2016), Jasche & Lavaux (2017, 2018)

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 $\rho_g = \mathcal{B}(\rho_m)$

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

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

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

Jasche et al. (2010), Jasche & Wandelt (2013), Lavaux & Jasche (2016), Jasche & Lavaux (2017, 2018)

Some BORG3 bias models

Voxel predictibility → freedom in bias model choice $\mathcal{B}$

Purely local models

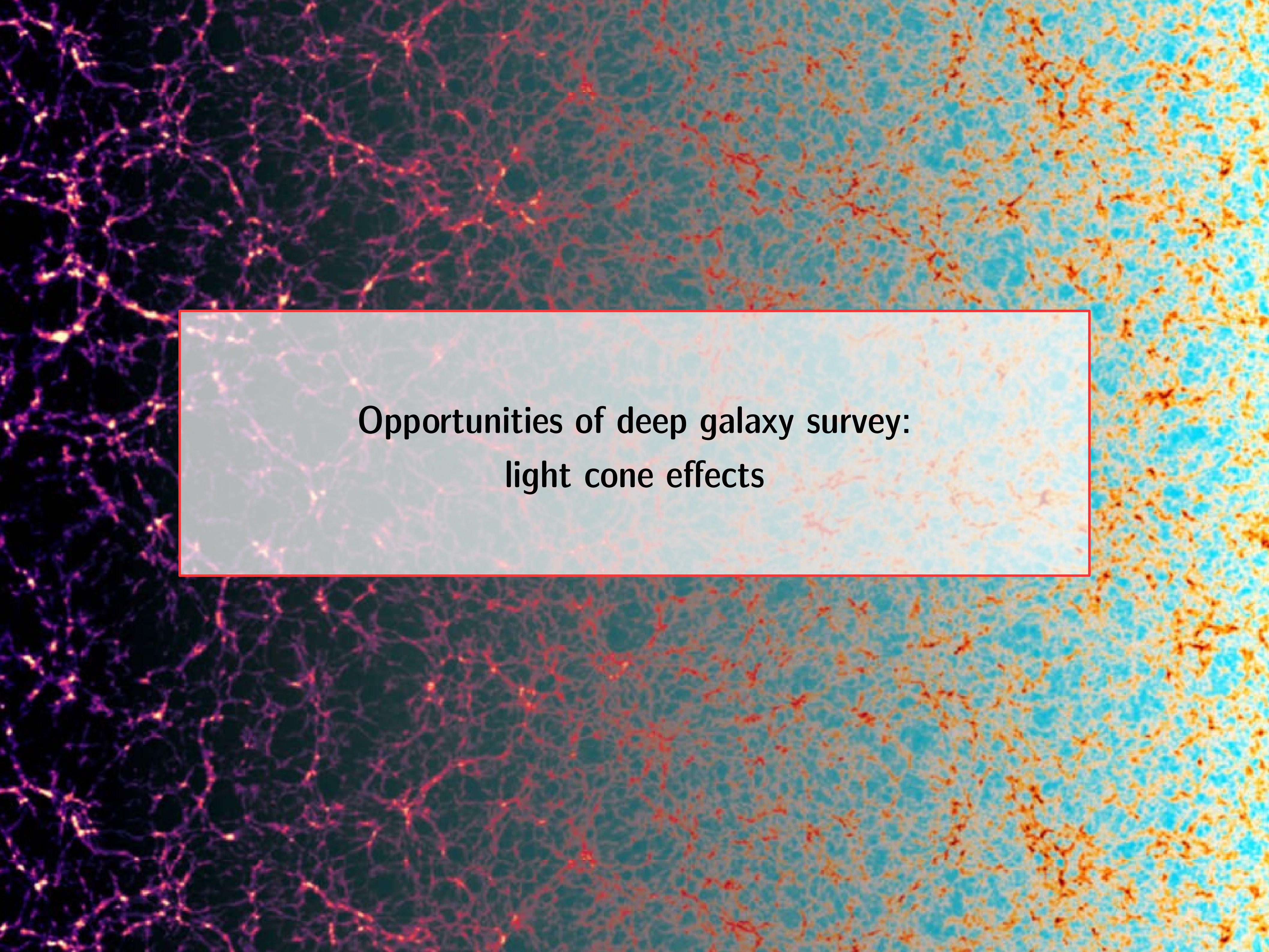
- Linear $\delta_g = b\delta_m$
- Power law $\rho_g \propto \rho_m^\alpha$
- Double power law $\rho_g \propto \dfrac{\rho_m^\alpha}{1+(\rho_m/\rho_0)^\beta}$
- 1-pt halo empirical halo $\rho_g \propto \rho_m^\alpha \exp\left(-(\rho_m/\rho_0)^{-\epsilon}\right)$
- Full 1-pt Halo distribution $P(M|\vec{a}) \propto n(M,\vec{a})$

Non-local models

- "EFT" / Second order
- Oct-tree

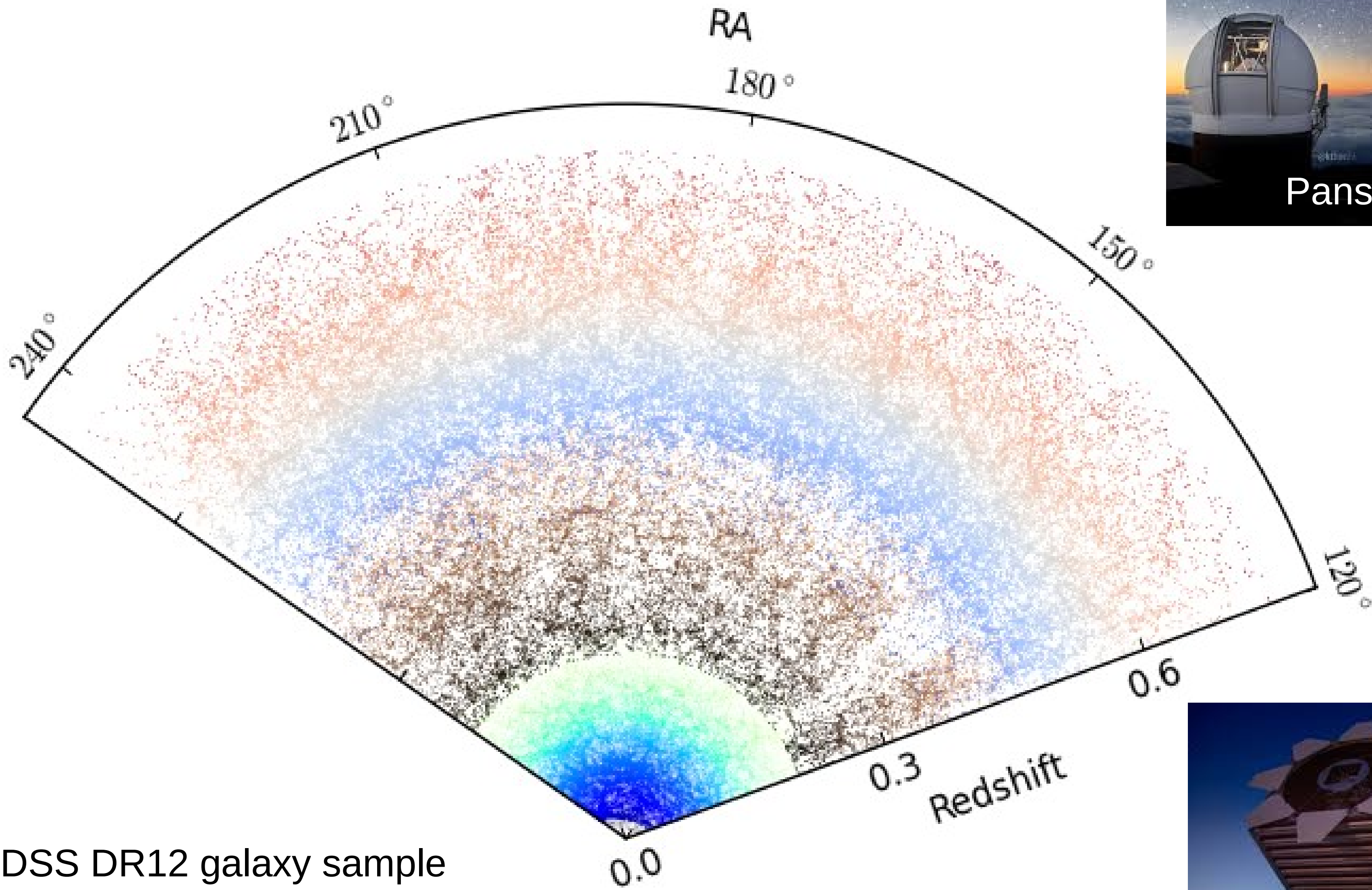
Applications to data

See Jens' talk for many results on data



Opportunities of deep galaxy survey: light cone effects

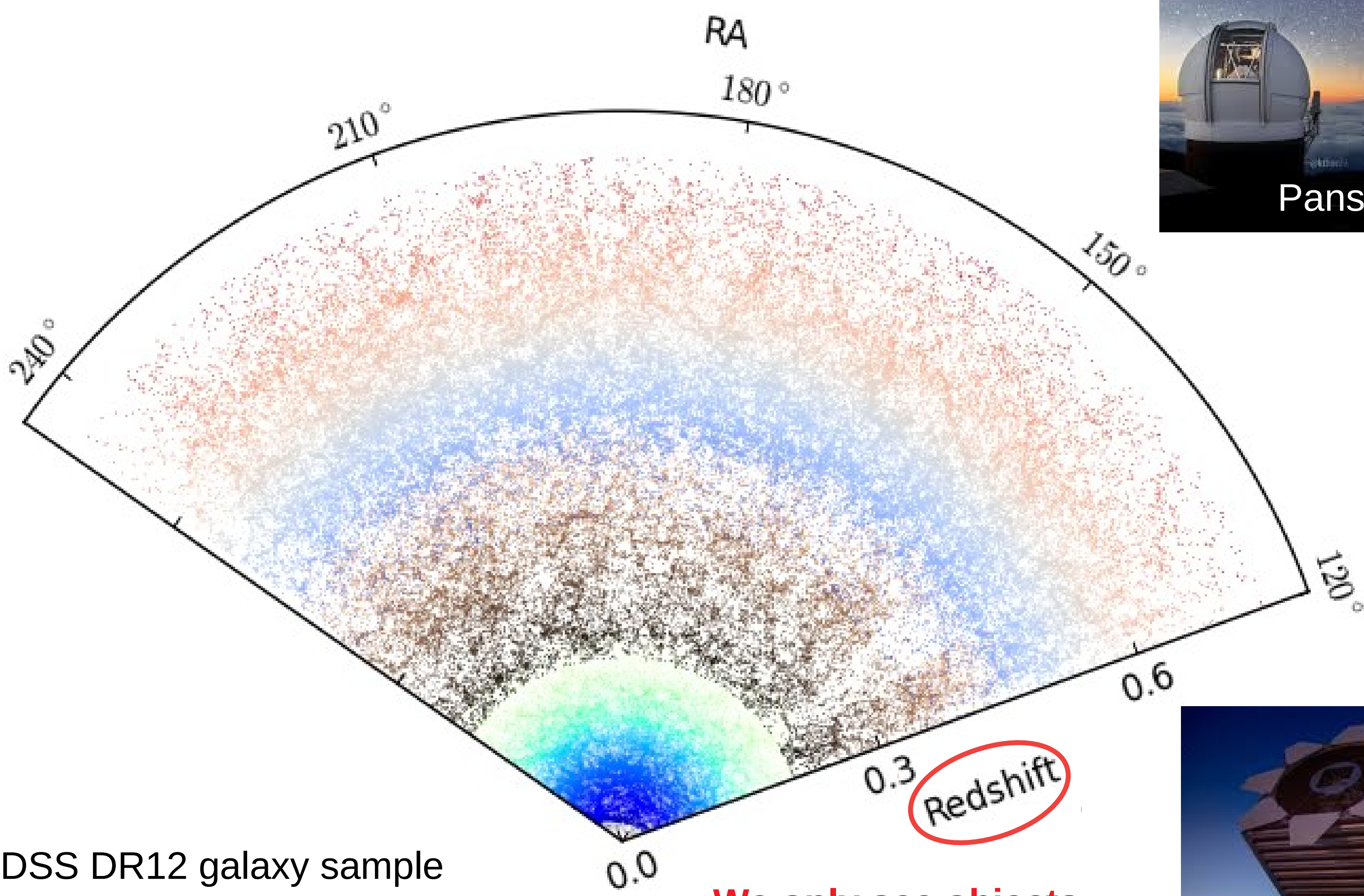
SDSS3 data



SDSS DR12 galaxy sample

~1.6 millions of galaxies

SDSS3 data



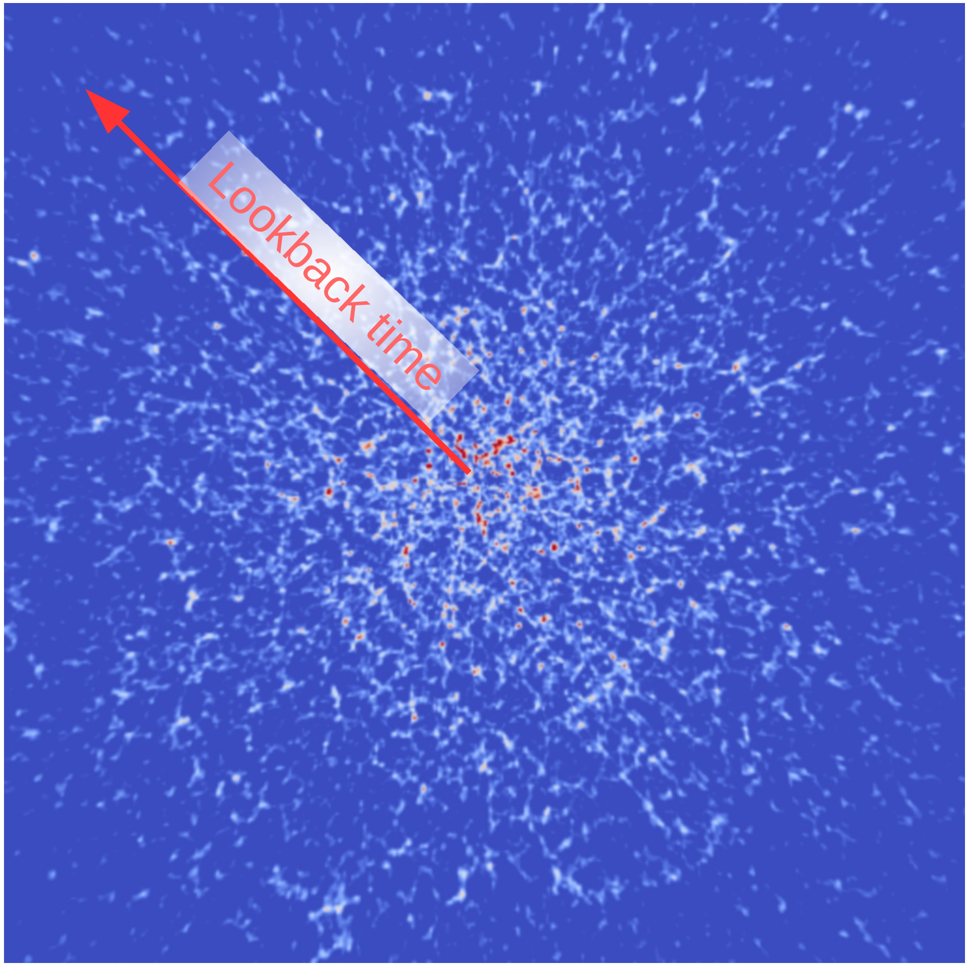
SDSS DR12 galaxy sample

~1.6 millions of galaxies

We only see objects as they appear on our lightcone

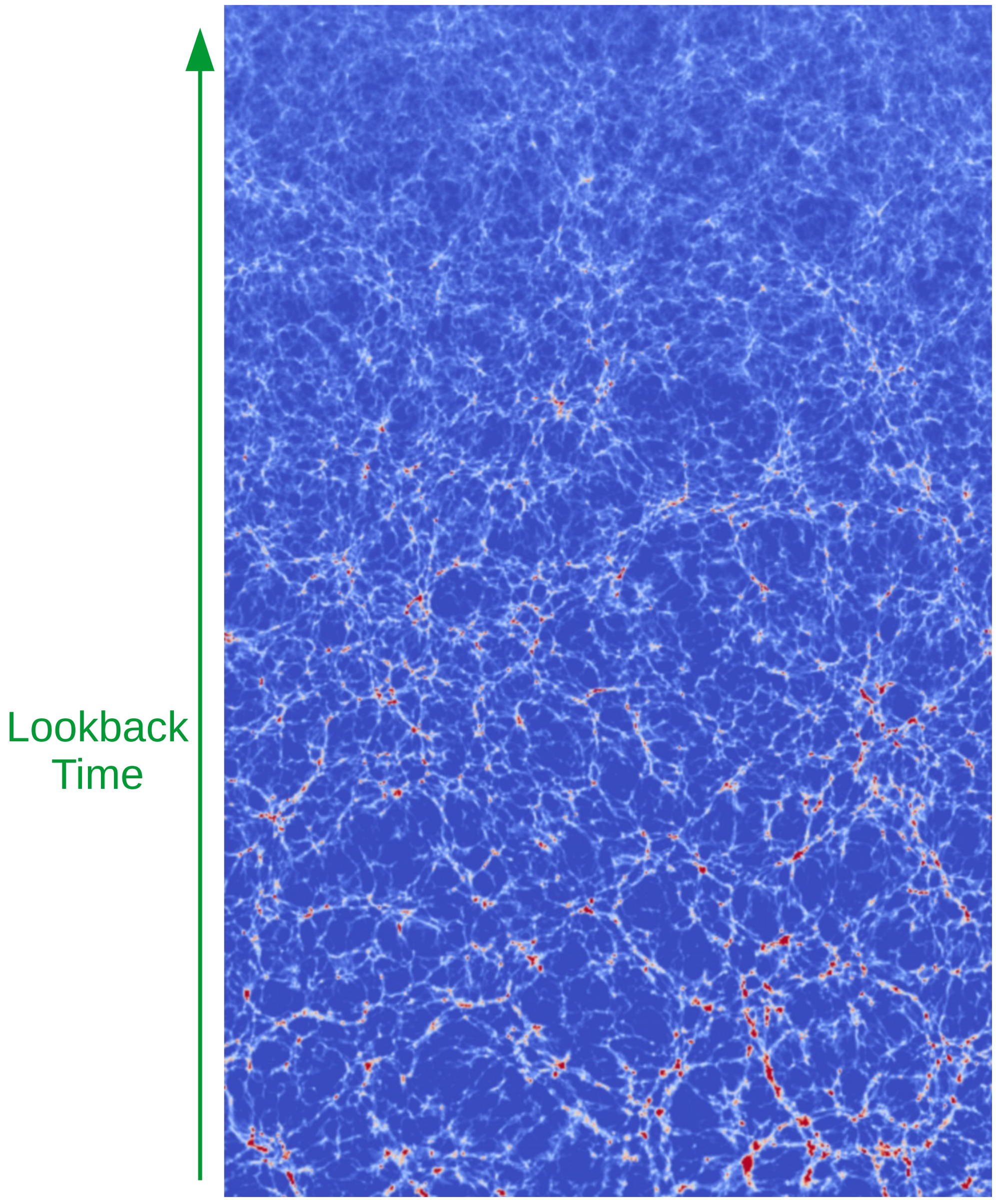
Two components of light cone effect

Cosmic expansion



Ramanah et al. (2018, submitted)

Cosmic growth of structures



The BORG/ALTAIR model

Bayesian **O**rigins **R**econstruction from **G**alaxies

ALcock-Paczyński consTr**AI**ned **R**econstruction

Initial conditions

↓

Dynamical model (LPT, 2LPT, PM)

↓

Cosmic expansion

↓

Bias transformation

↓

Selection function

↓

Likelihood

↓

Data

Doogesh K. Ramanah

Image credits: Paramount Pictures, Ubisoft

Ramanah et al. (2018, submitted)

The BORG/ALTAIR inference

Bayesian Origins Reconstruction from Galaxies

ALcock-Paczyński consTrAIned Reconstruction

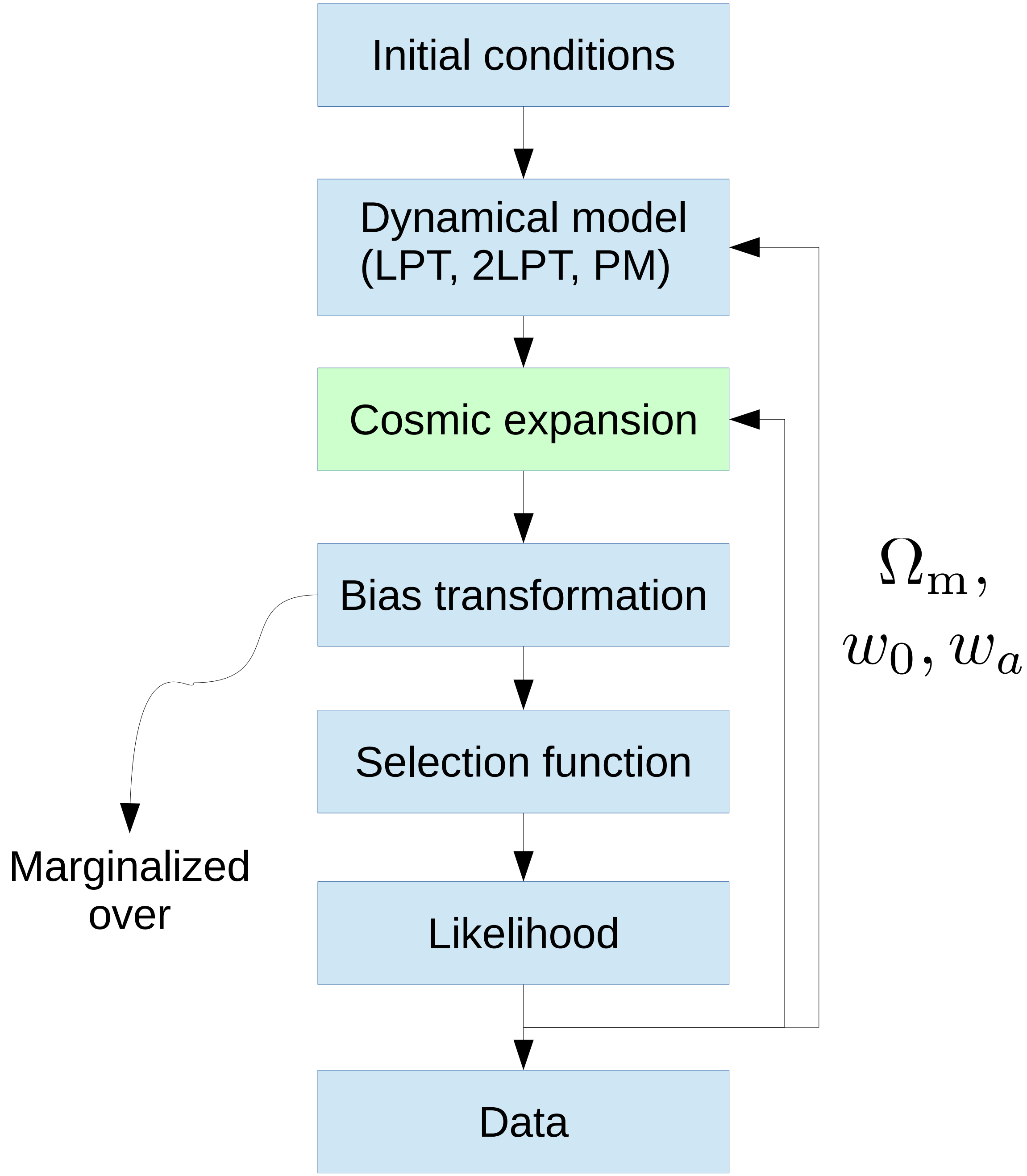
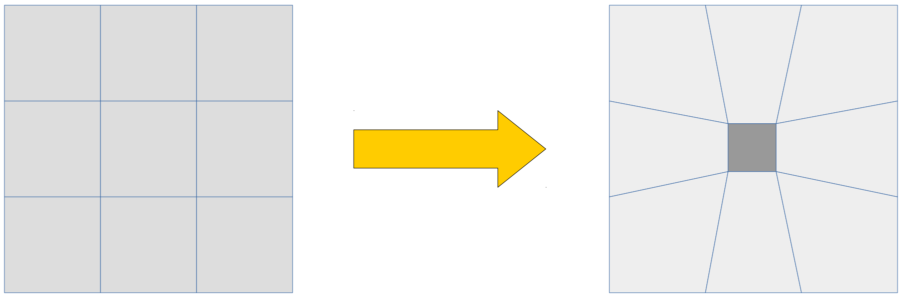


Image credits: Paramount Pictures, Ubisoft

Ramanah et al. (2018, submitted)

Cosmic expansion in practice...

Non-linear density remapping:



Comoving coordinates

Scaled redshift coordinates

$$\vec{x}$$

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

- Implemented with an p^{th} order linear polynomial interpolation

$$\rho^{(z)} = J(\vec{x}) \sum_i \left(\prod x_{a_i}^{q_i} x_{b_i}^{r_i} \ldots \right) \rho_i^{(x)}$$

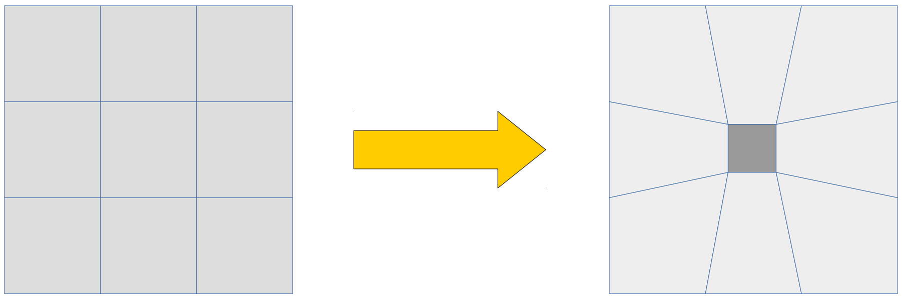
$$\Omega$$

- Required to avoid strong grid-on-grid interpolation aliasing
(5th order ok in practice)

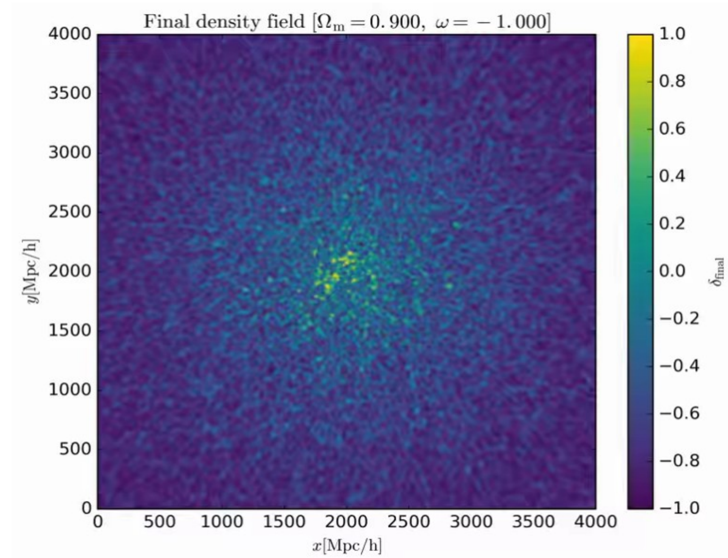
Ramanah et al. (2018, submitted)

Cosmic expansion in practice...

Non-linear density remapping:



Comoving coordinates	Scaled redshift coordinates
$\vec{x}$	$\vec{\mathfrak{z}}_i = \dfrac{c}{H_0} z_i \hat{u}_i$



- Implemented with an p^th order linear polynomial interpolation

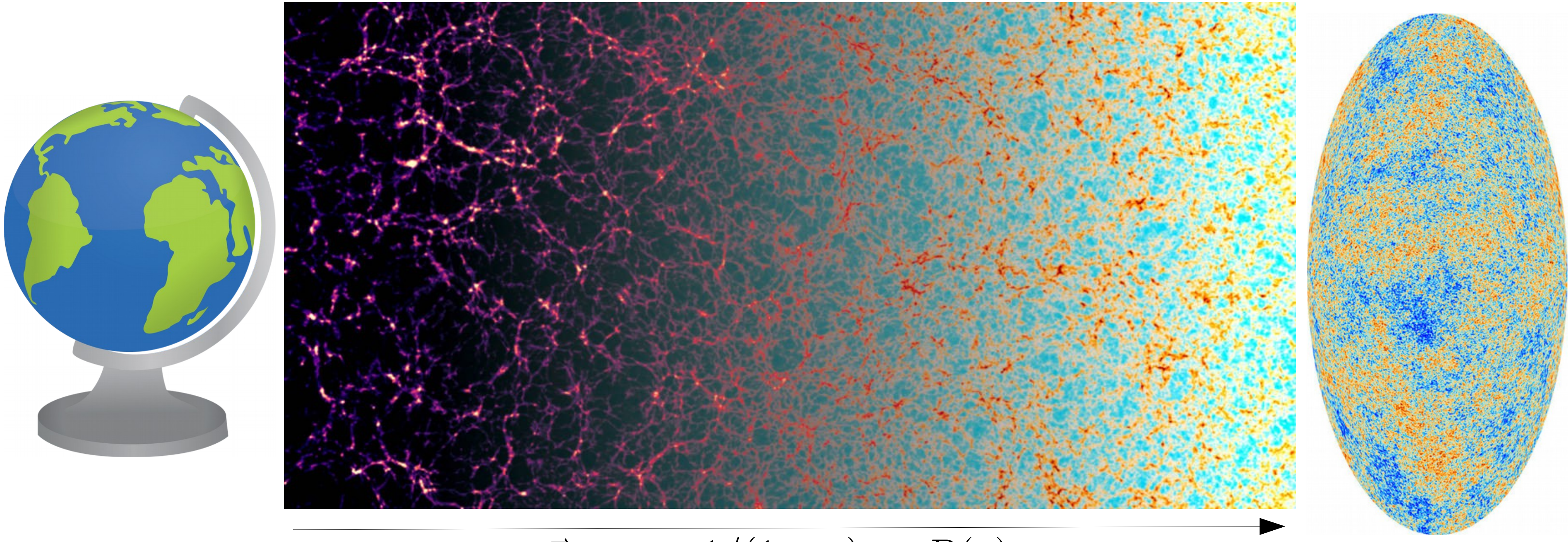
$$\rho^{(z)} = J(\vec{x}) \sum_i \left(\prod x_{a_i}^{q_i} x_{b_i}^{r_i} \ldots \right) \rho_i^{(x)}$$

$$\Omega$$

- Required to avoid strong grid-on-grid interpolation aliasing (5^th order ok in practice)

Movie Credit: Doogesh K. Ramanah   Ramanah et al. (2018, submitted)

Lightcone in 1/2LPT

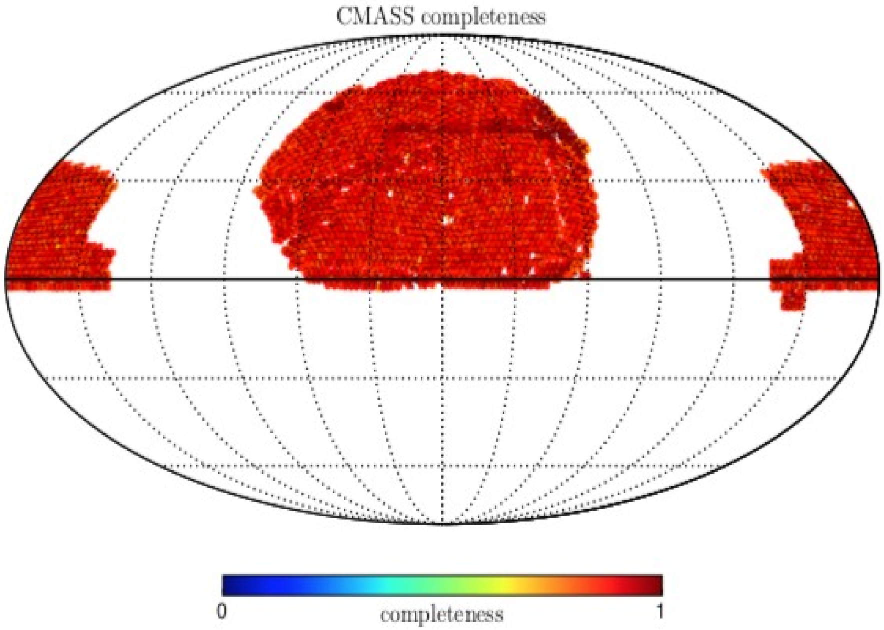


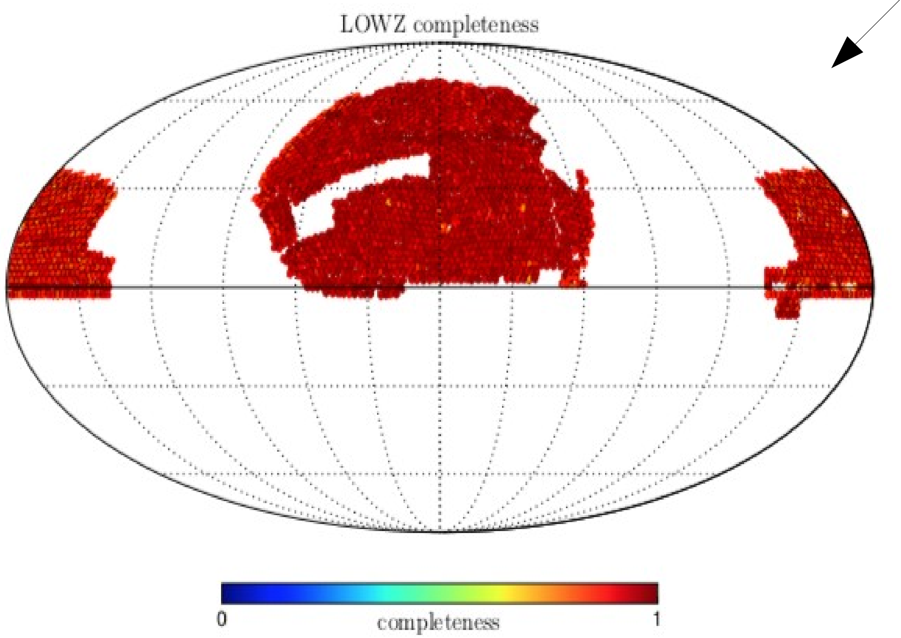
$$\vec{z} \rightarrow a = 1/(1+z) \rightarrow D(a)$$

$$\vec{x}(\vec{q}) = \vec{q} + \Psi(\vec{q}) \underset{LPT}{\simeq} \vec{q} + D(a(||\vec{q}||))\vec{\Psi}(\vec{q})$$

- ✓ Can be generalized to PM
- ✓ Is differentiable
- ✓ Misses some higher order effect, but exact solution can be implemented also

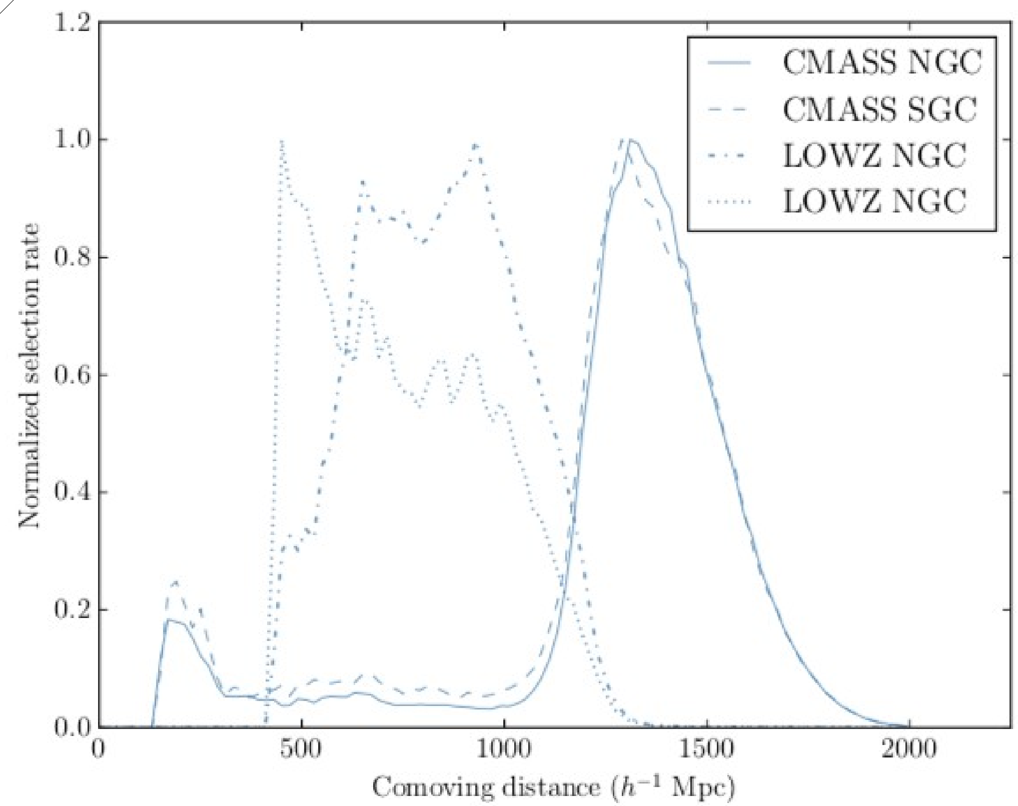
Mock test setup: SDSS3 template





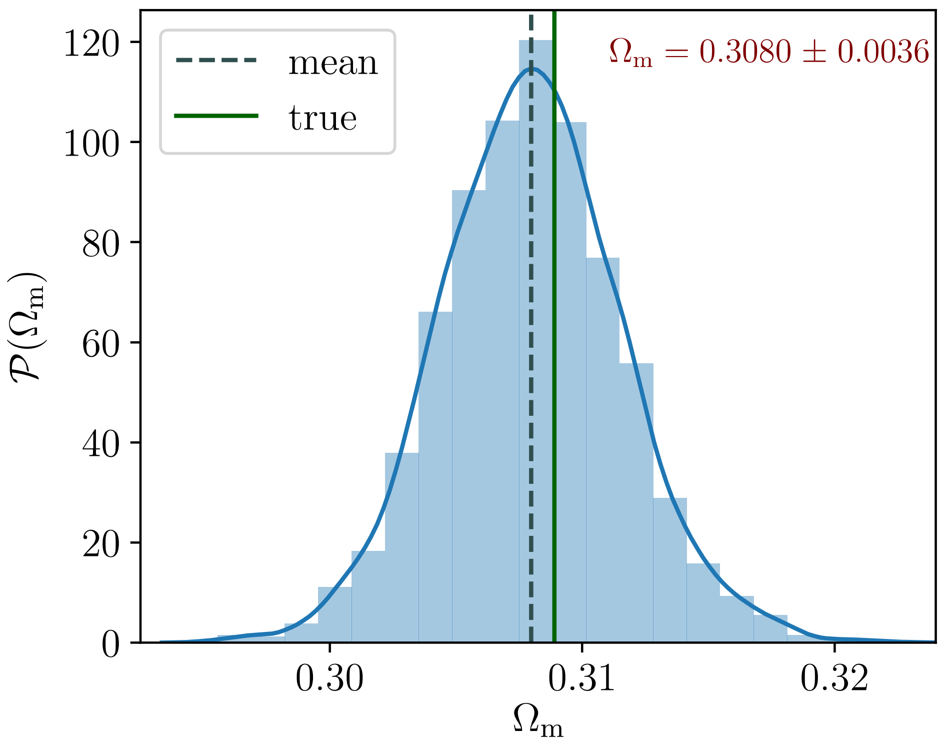
Sky completeness

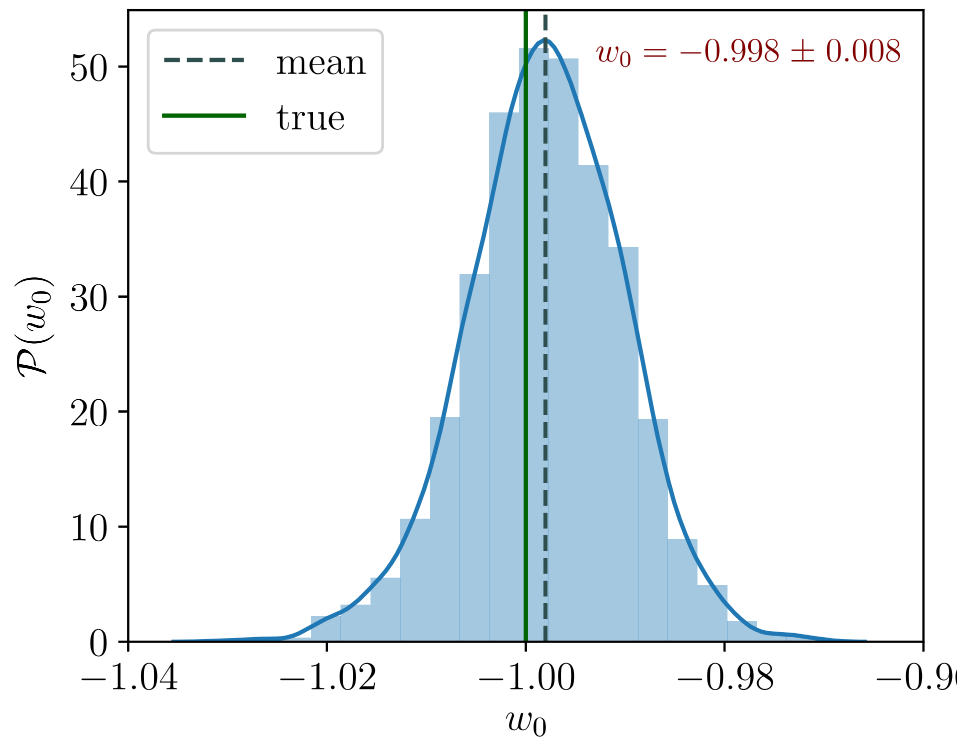
Redshift completeness



Ramanah et al. (2018, submitted)

Mock test setup: results





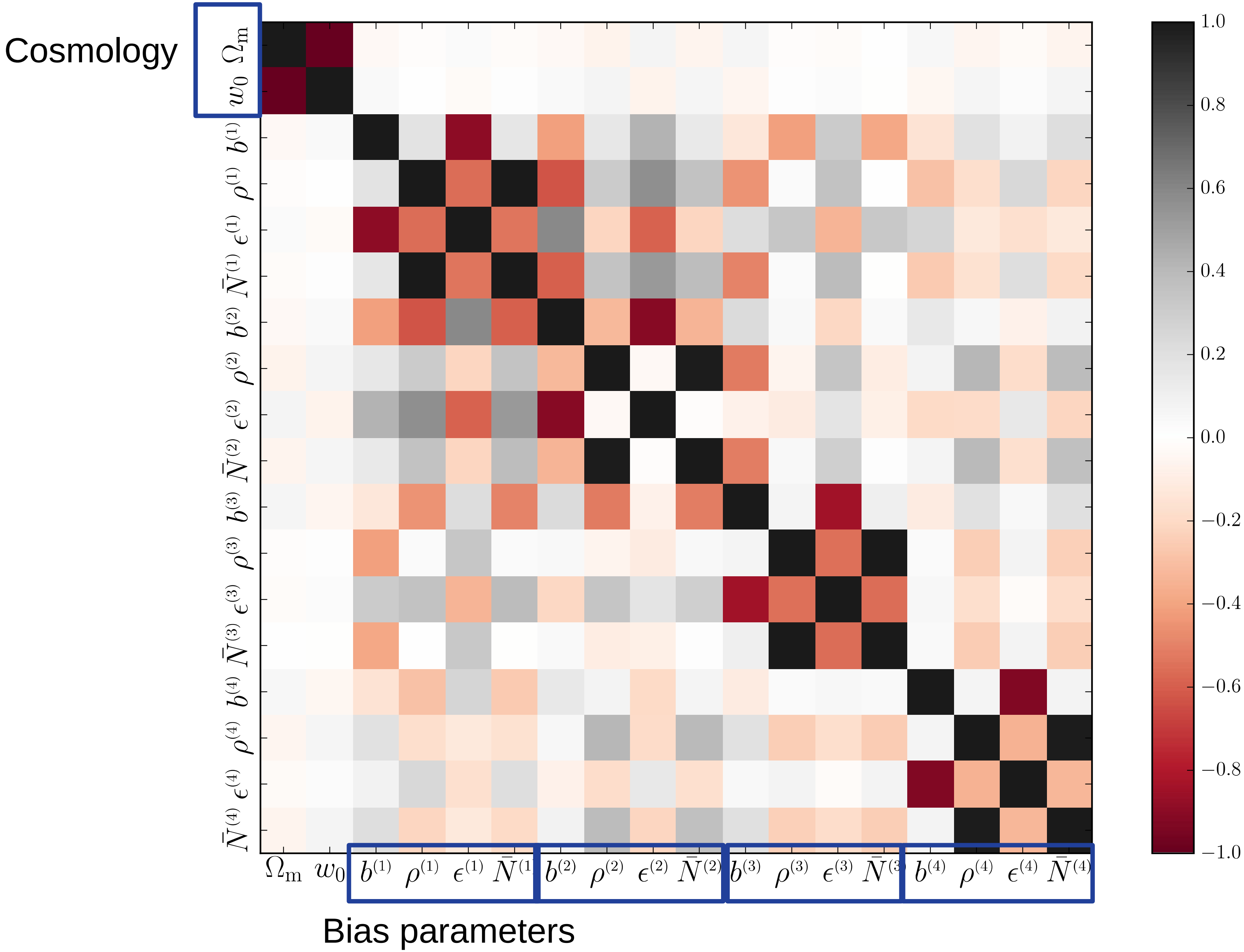
Constraints resilient to isotropic biases

Most of the information comes from geometry

Another interesting POV is Li et al. (ApJ, 2014, 2016)

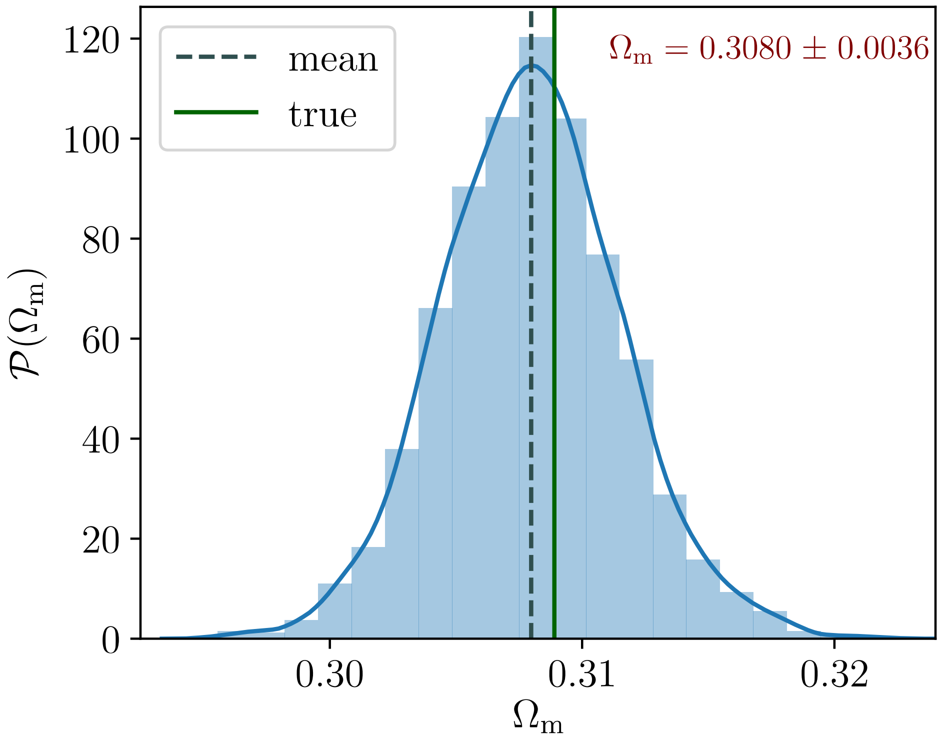
Ramanah et al. (2018, submitted)

Robustness assessment



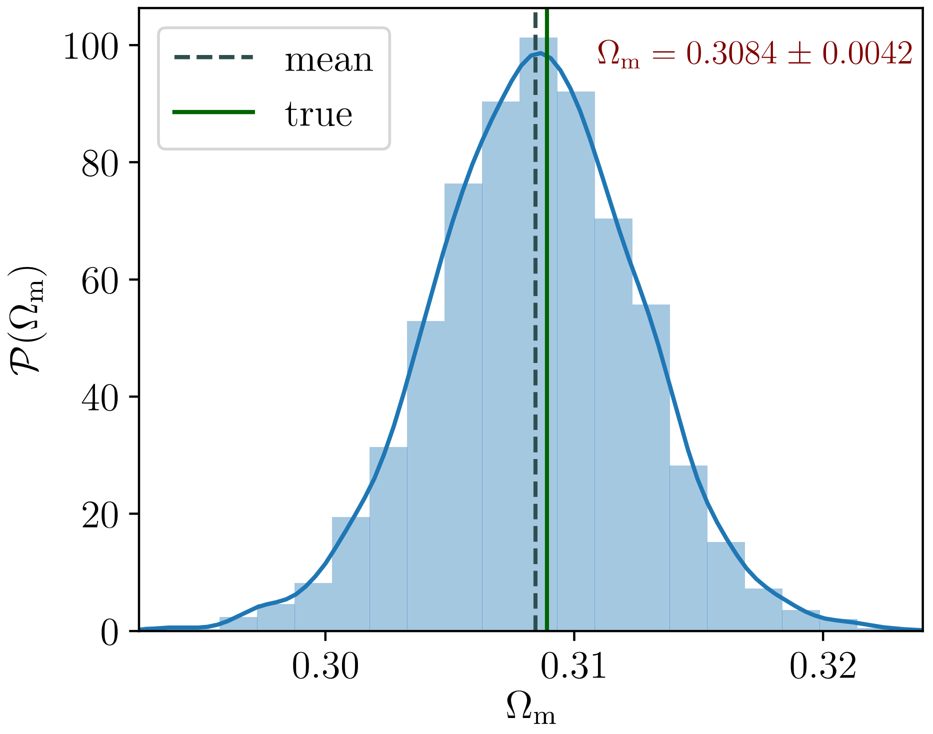
Robustness to improper prior

Good prior



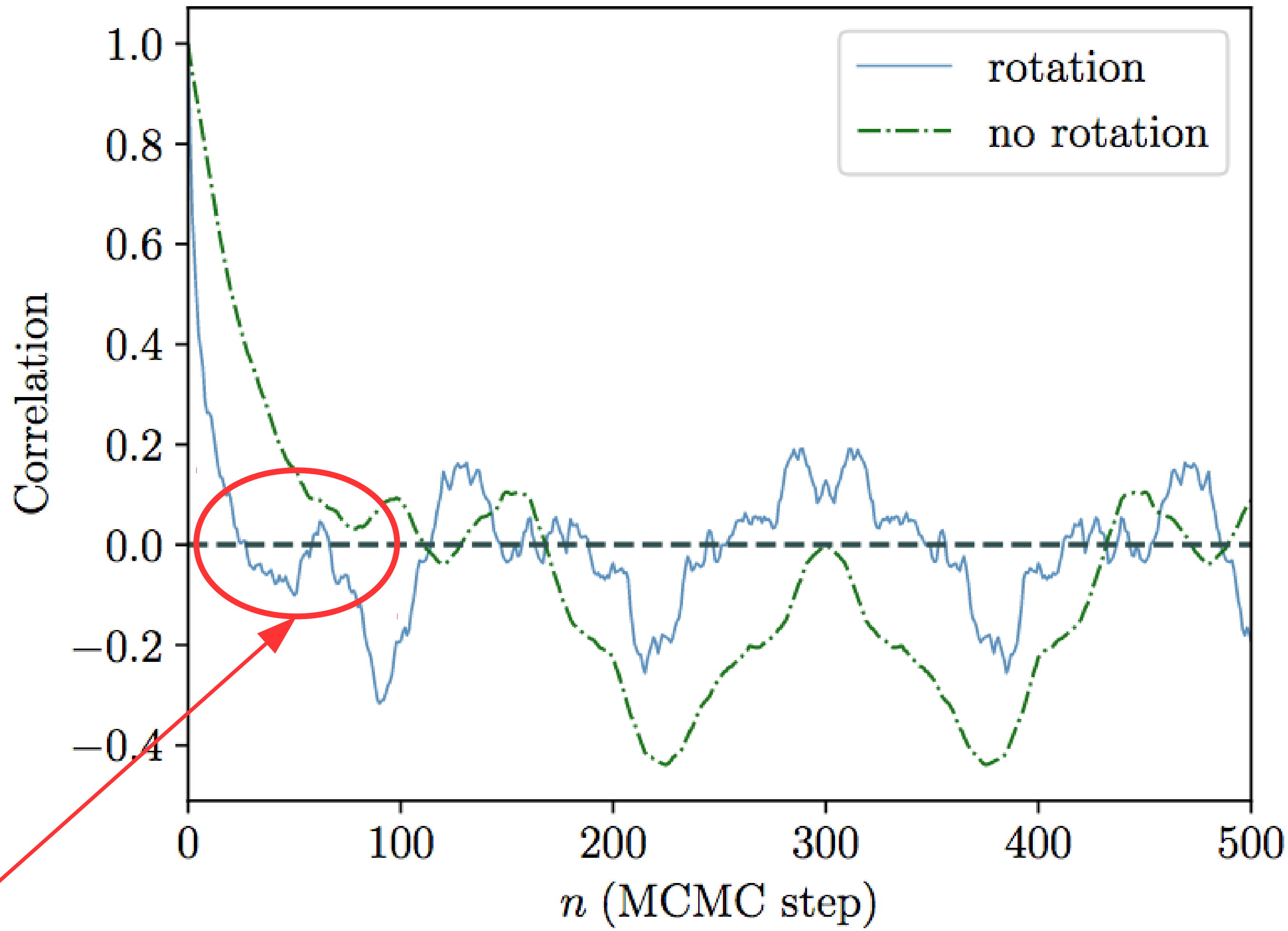
Bad prior

Modified prior $P(k)$ ($\Omega_{\mathrm{m}} = 0.40,\ w_0 = -0.85$)



Constraints resilient to isotropic prior biases

MCMC efficiency



Decorrelation

Summary

- The BORG3 machine is starting to become a powerful machine to exploit spectroscopic surveys, with flexible internals and parallelization.
- The ALTAIR extension is showing promising results for cosmological information extraction, relying on the detailed anisotropic distortion of the density field.
- Data application of BORG/ALTAIR are coming.

Velocity field inference from noisy distance data

... and some application to data

Why cosmic flows are interesting

Direct observation of gravitational field is difficult

- Lensing
- Integrated Sachs Wolf effect
- Cosmic velocities
- Galaxy distribution

Cosmic velocities have several advantages:

- Do not need background sources
- High S/N nearby
- Less assumptions

Some big disadvantages:

- Prone to high systematics
- Very noisy at high distances
- Unclear calibration of distance indicators (Fundamental plane, Tully-Fisher, ...)

The Cosmic Flows inference problem

It is old/recent (Late 1970s, Jim Peebles action method, linear gravity solver)

Got mostly abandoned in 1999 (except for pure RSD)

Now solvable with Bayesian Hierarchical Modeling (BHM)

Will be needed for Taipan, ZTF, LSST

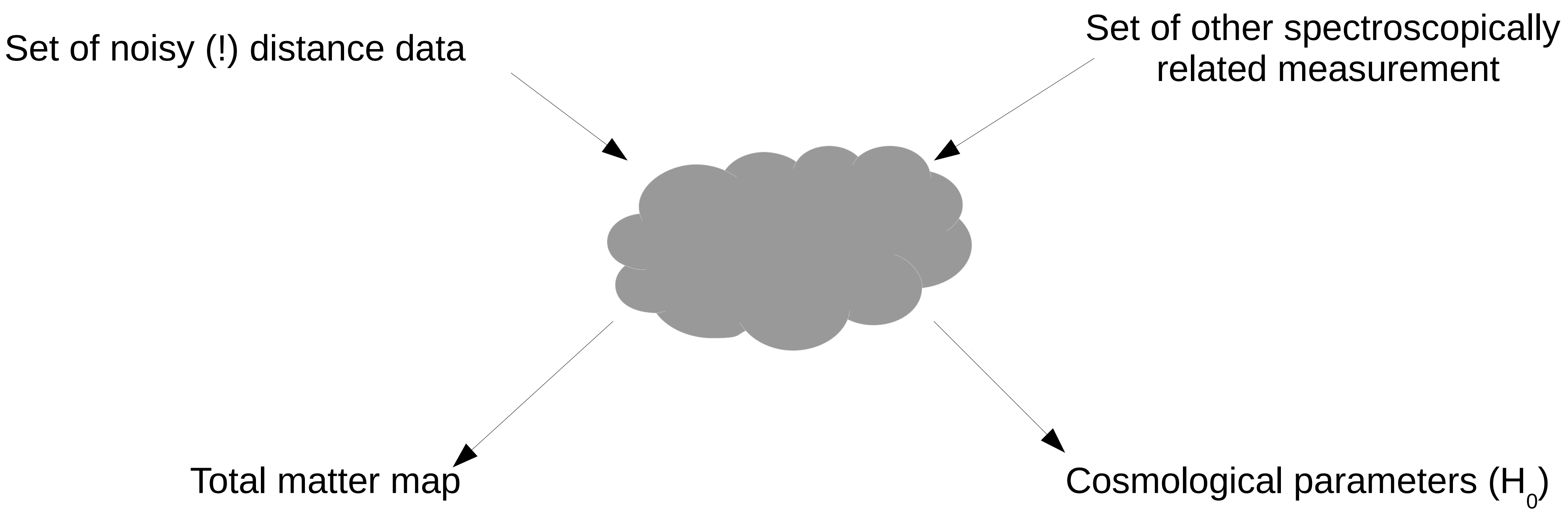
The Cosmic Flows inference problem

It is old/recent   (Late 1970s, Jim Peebles action method, linear gravity solver)

Got mostly abandoned in 1999 (except for pure RSD)

Now solvable with Bayesian Hierarchical Modeling (BHM)

Will be needed for Taipan, ZTF, LSST

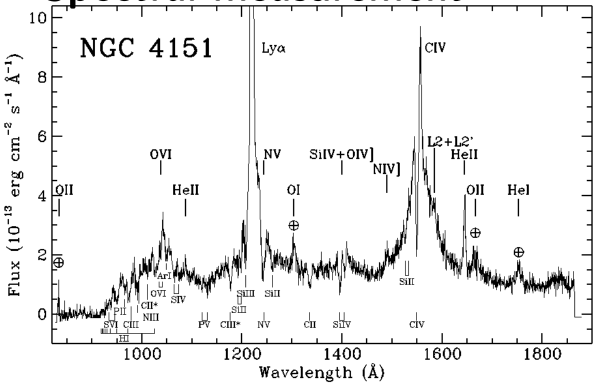


VIRBIUS model: forward

Distance estimate

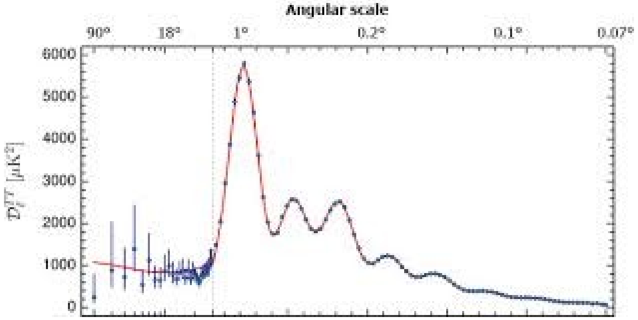
$$\mu_i = \mu_i^{\mathrm{obs}} + \epsilon_i^{\mu}$$

Spectral measurement



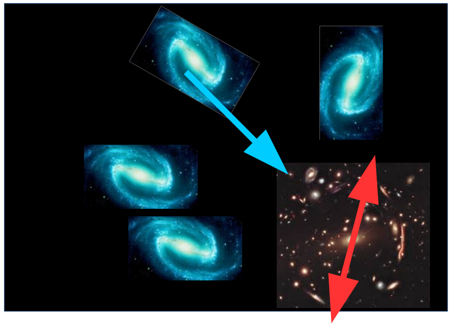
$$cz_i^{\mathrm{obs}} = c\bar{z}_i + (1+\bar{z}_i)\left(v(\vec{x}_i(\bar{z}_i)) + \epsilon_i^{v}\right) + \epsilon_i^{z}$$

Cosmology



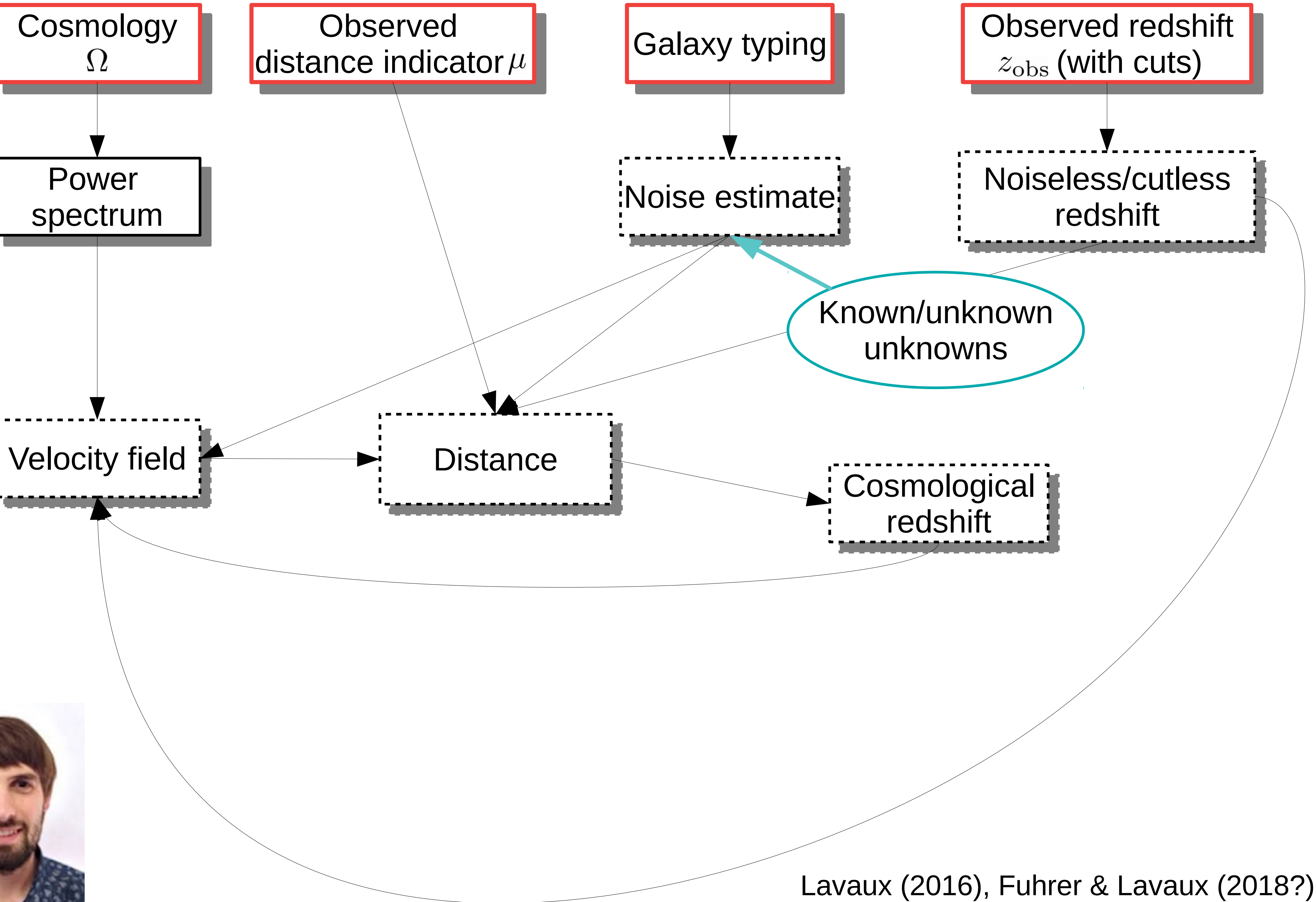
$$\langle v(\vec{x})v(\vec{y})\rangle = \xi(|\vec{x}-\vec{y}|)$$

Classifier / supplemental error



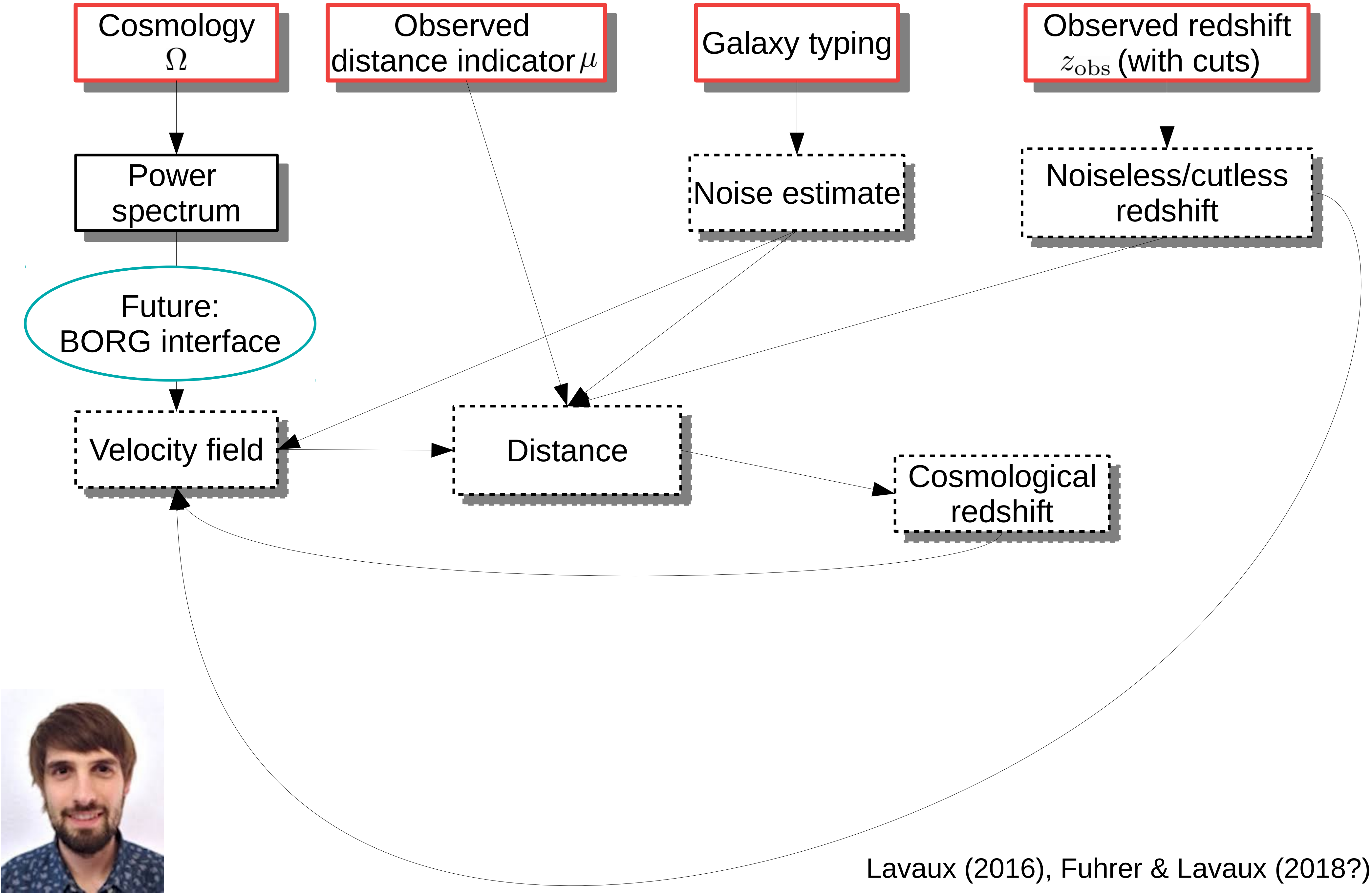
$$\langle(\epsilon_i^{v})^2\rangle = \sigma_{p(i)}^2$$

The VIRBIUS model (Diagram attempt)



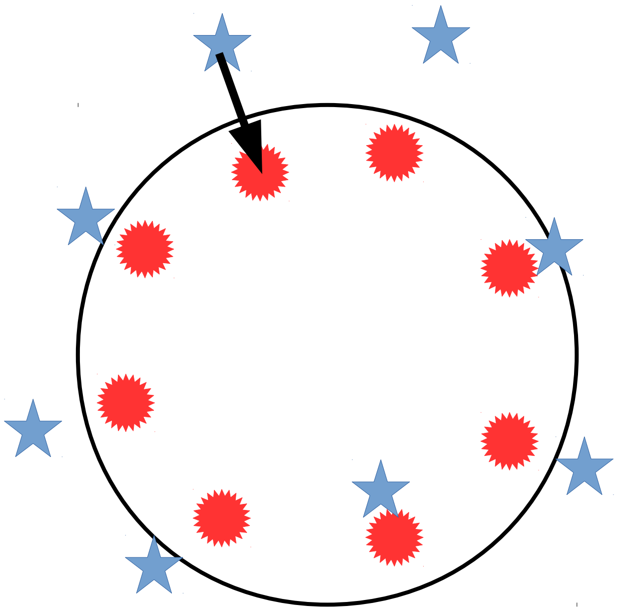
Lavaux (2016), Fuhrer & Lavaux (2018?)

The VIRBIUS model (Diagram attempt)

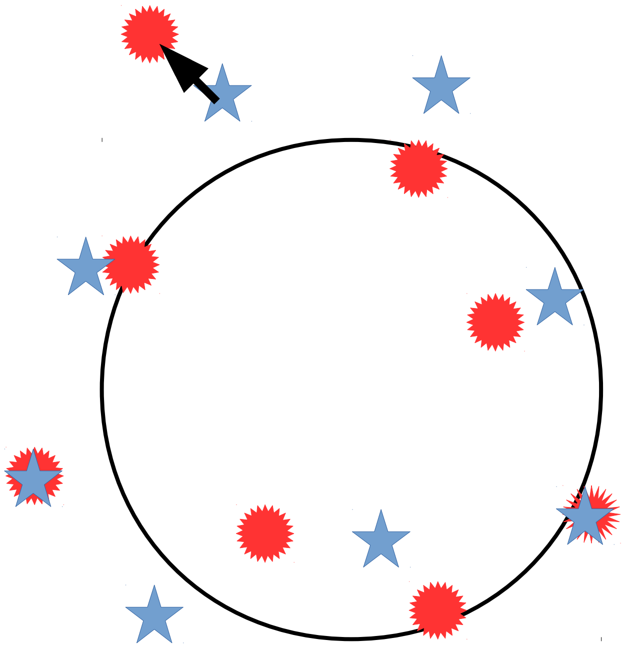


Lavaux (2016), Fuhrer & Lavaux (2018?)

Redshift cut biases

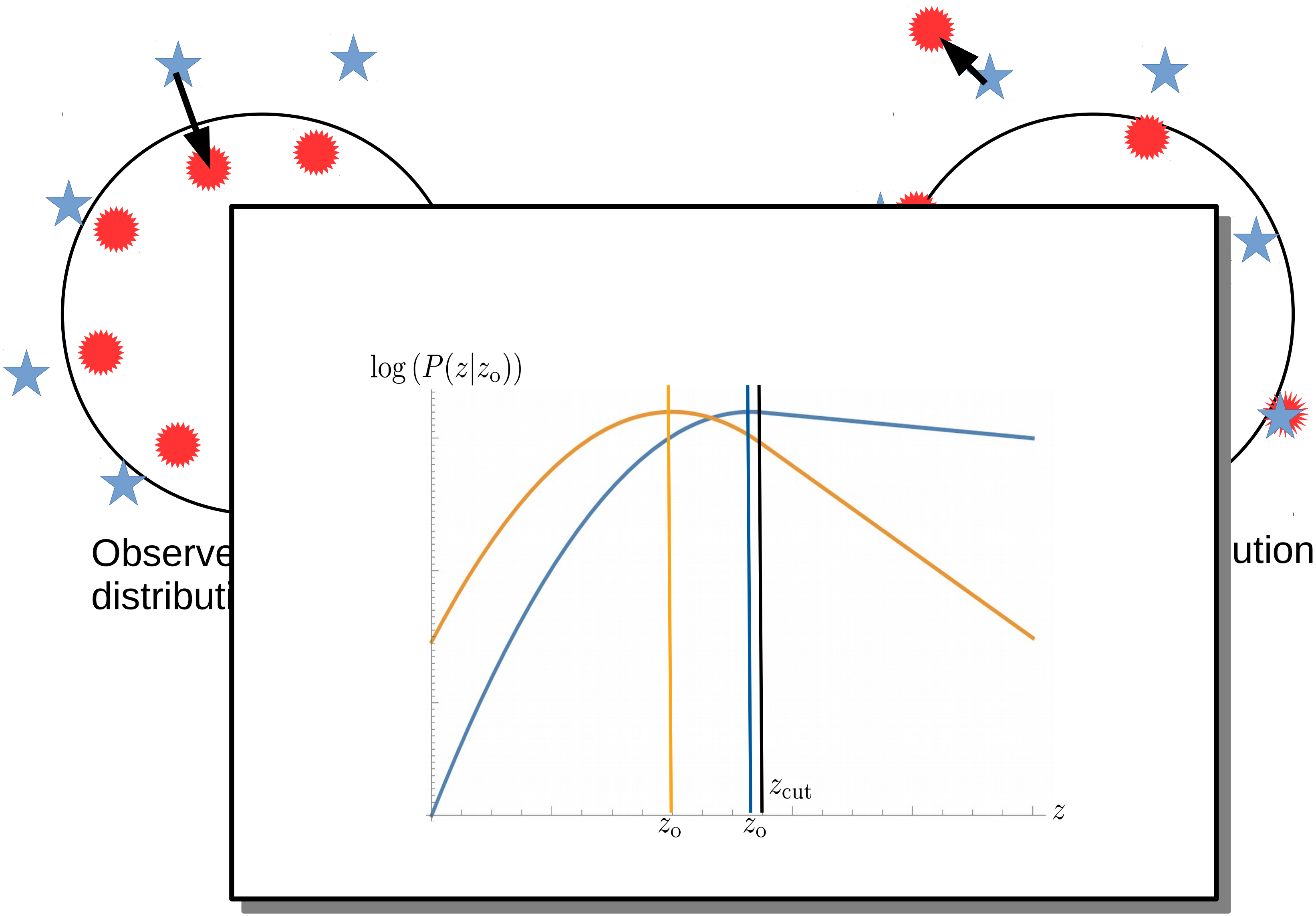


Observed redshift distribution with cut

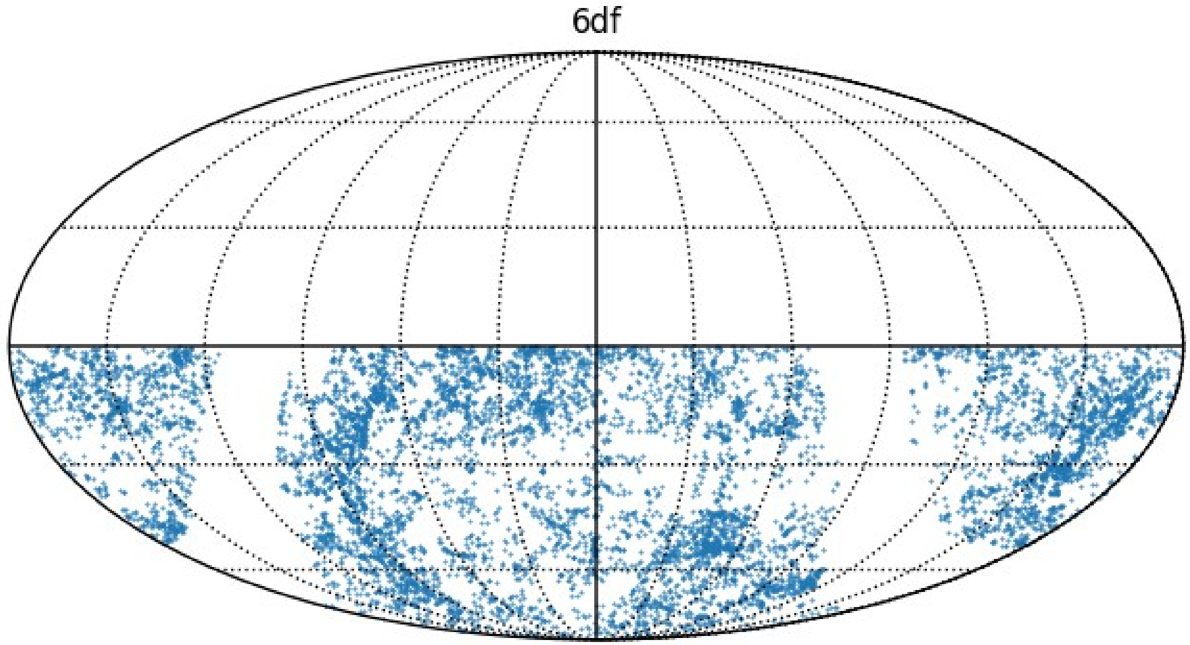


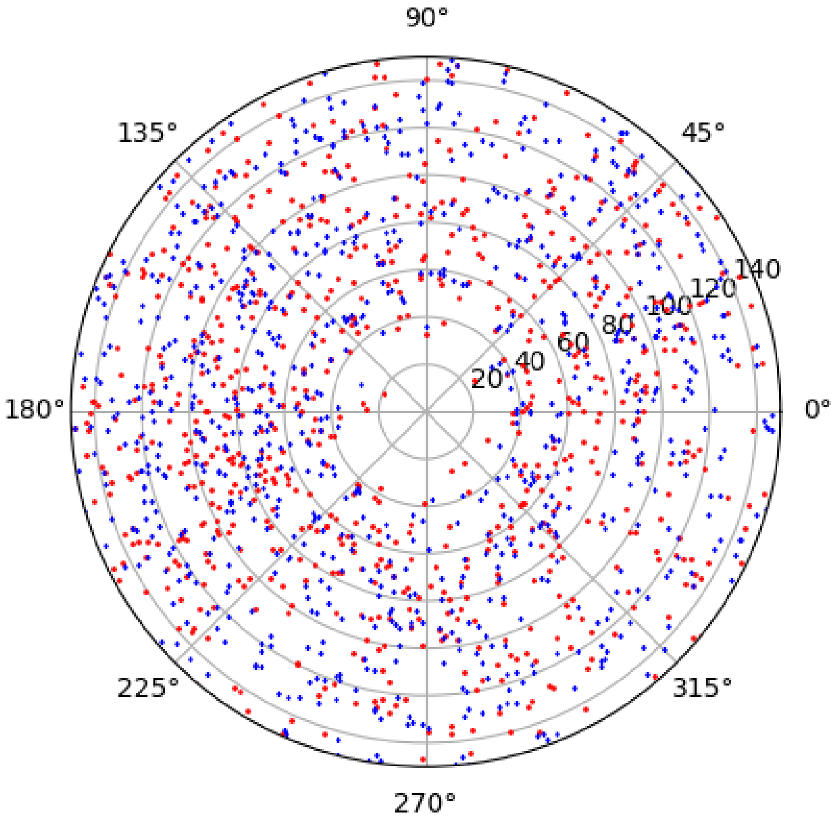
Actual redshift distribution with cut

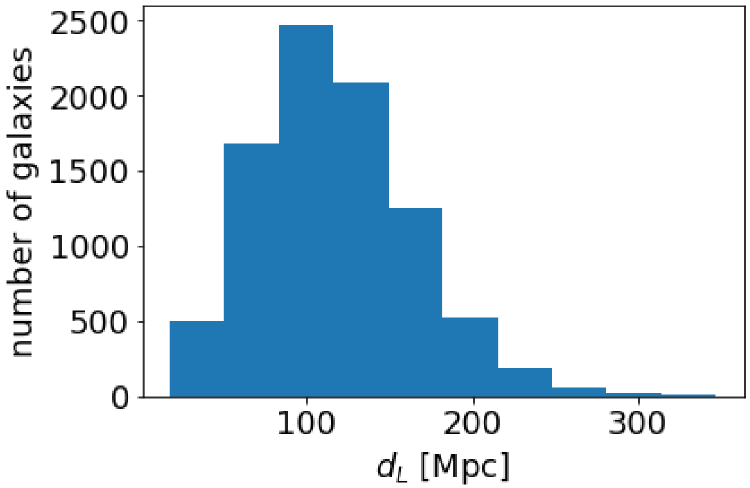
Redshift cut biases

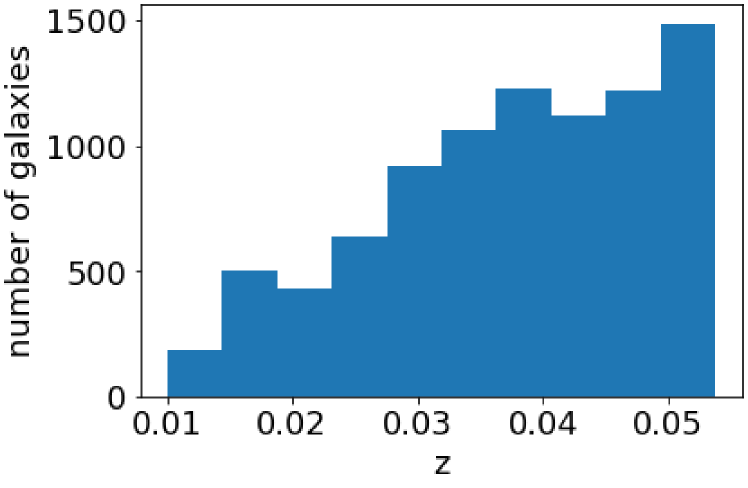


6dF dataset







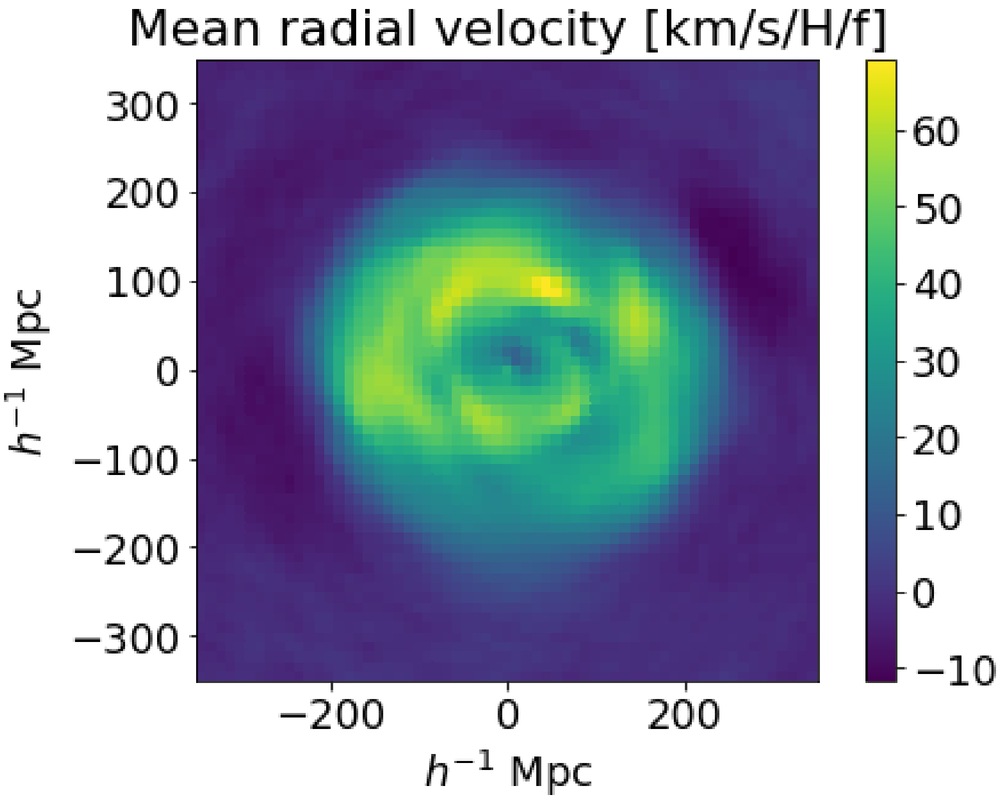


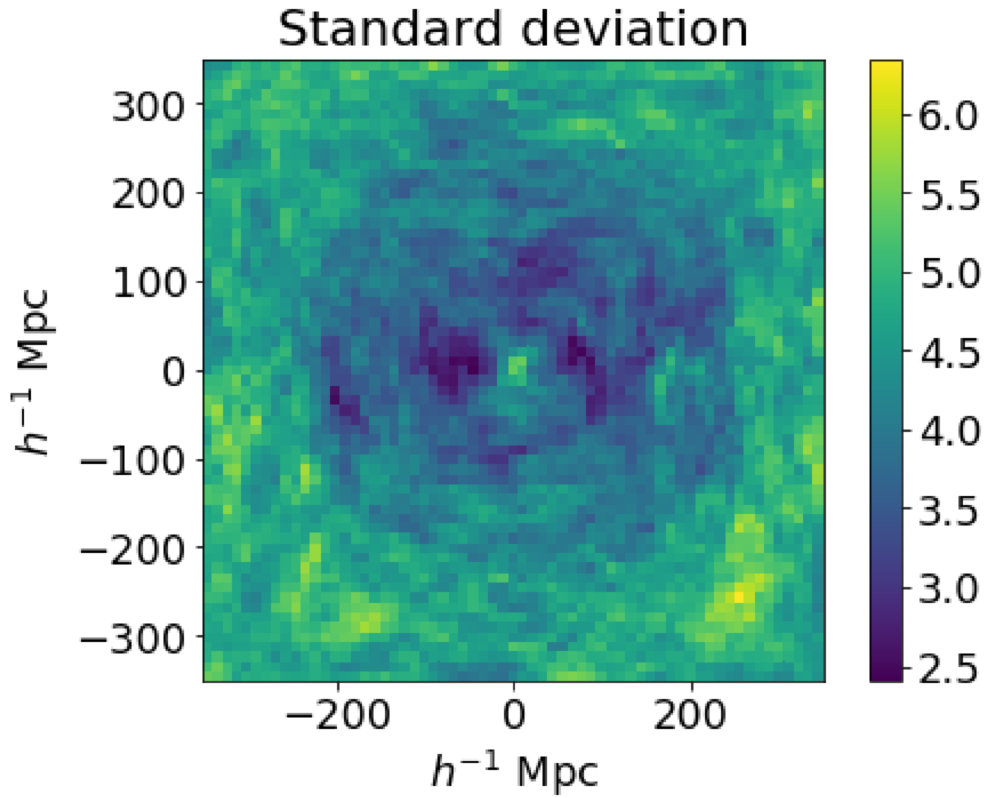
Red = redshift

Blue = Estimated distance

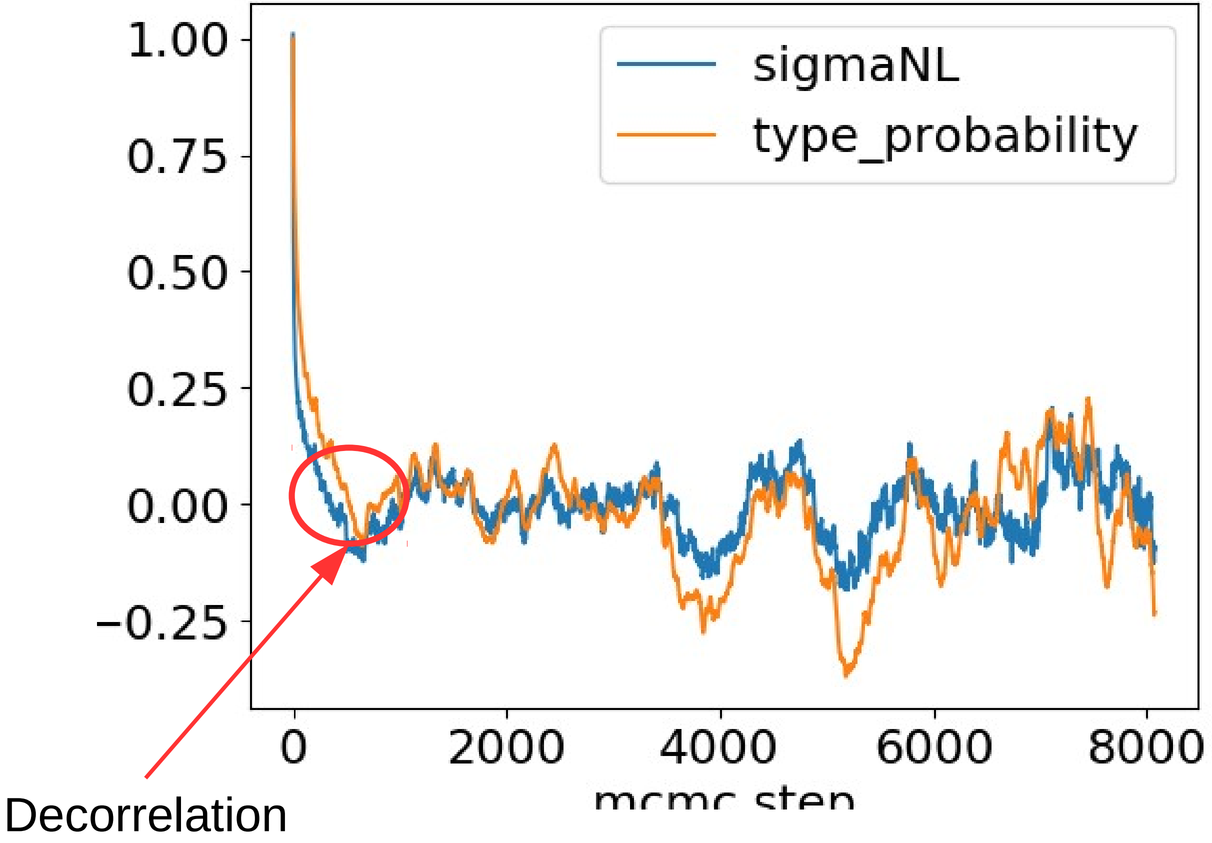
~9000 galaxies

Results on mock data



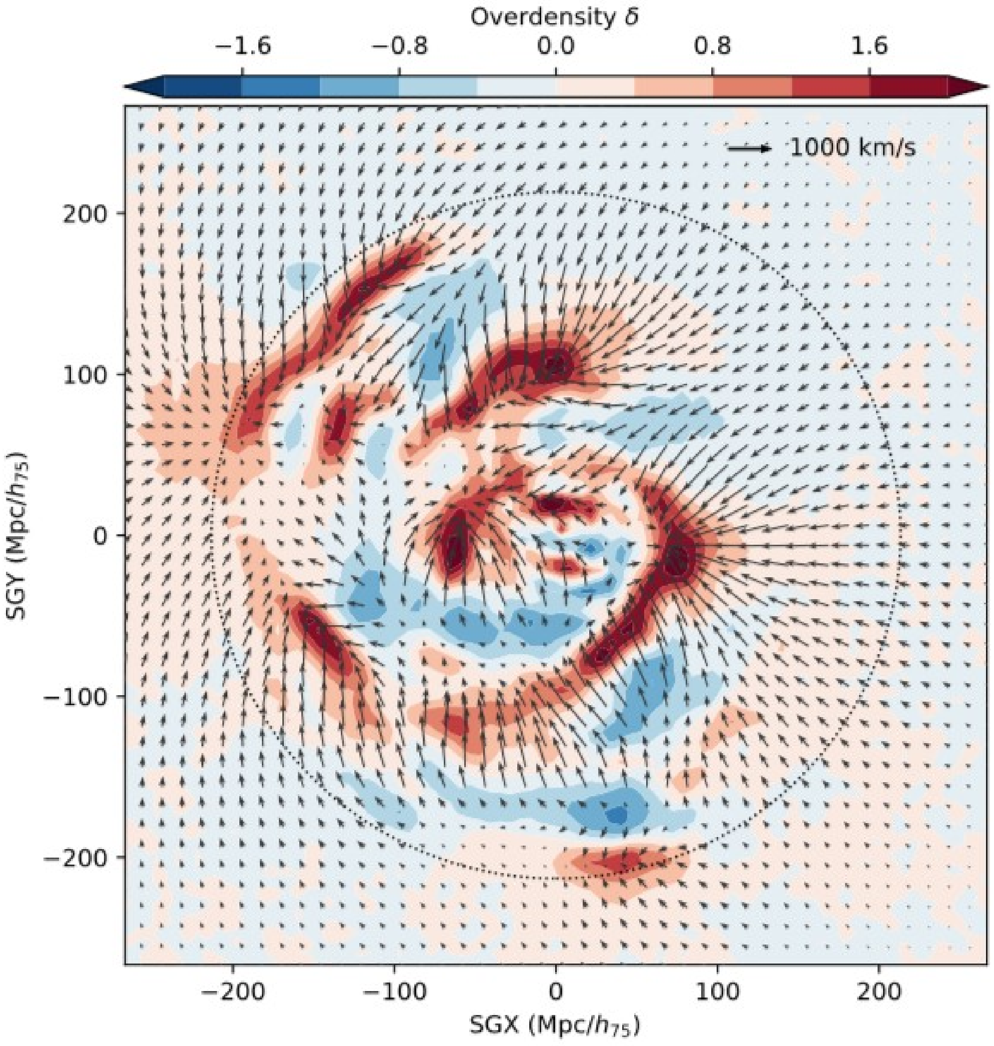


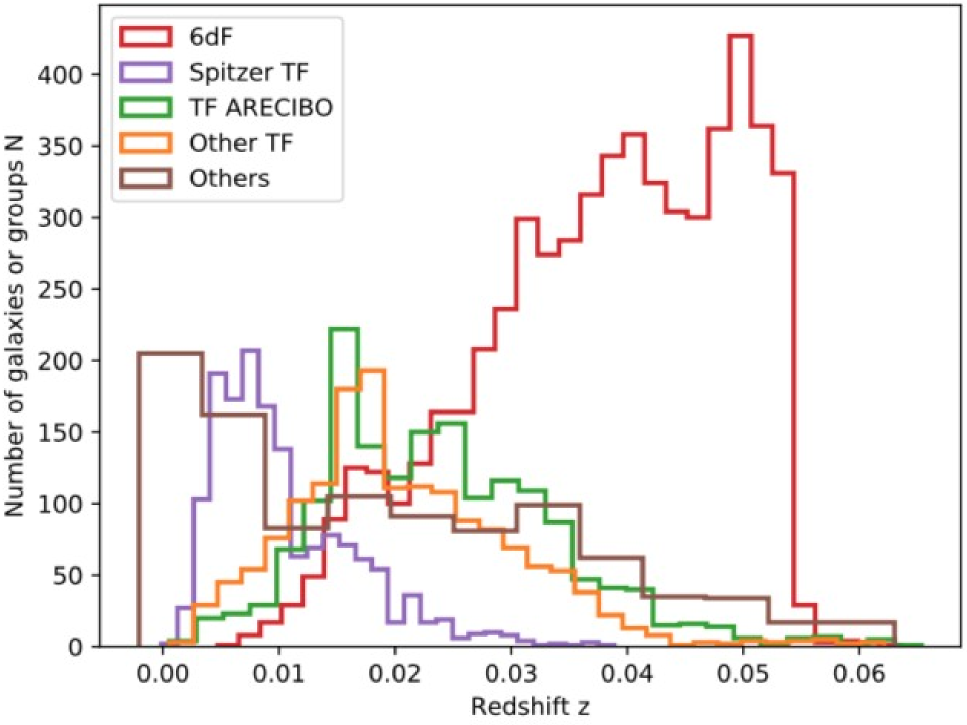
MCMC efficiency



...Some results on data with VIRBIUS1

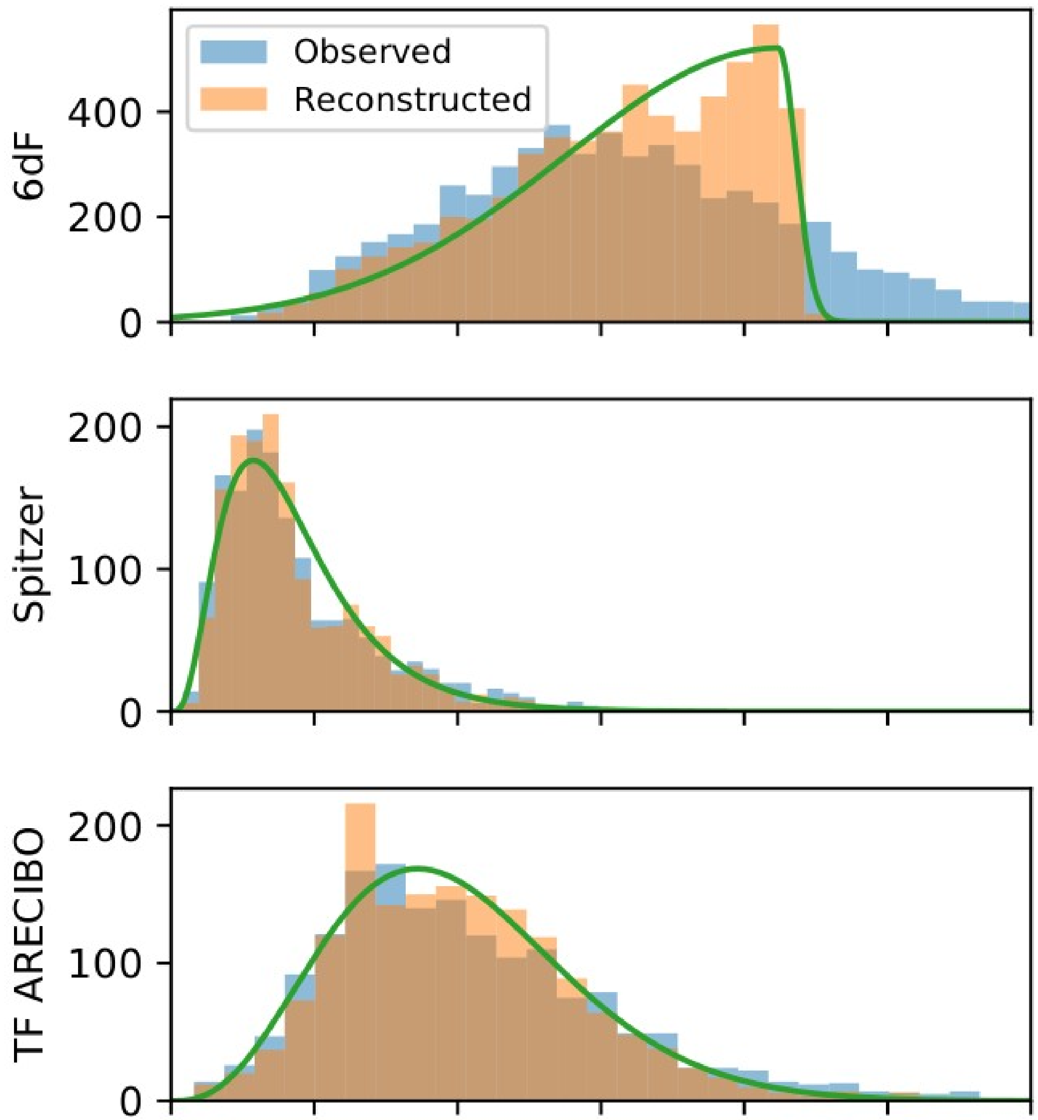
Other work: VIRBIUS1 on CF3





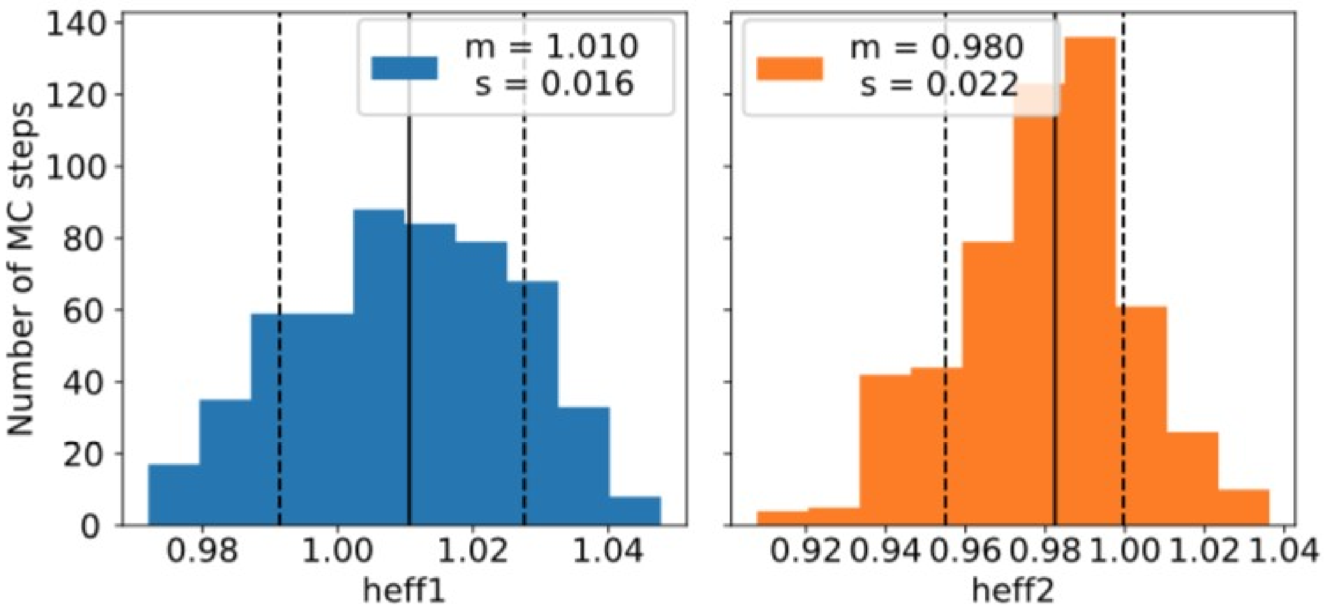
Graziani, Courtois, Lavaux et al. (2018 submitted)

CF3: Example of distance inference



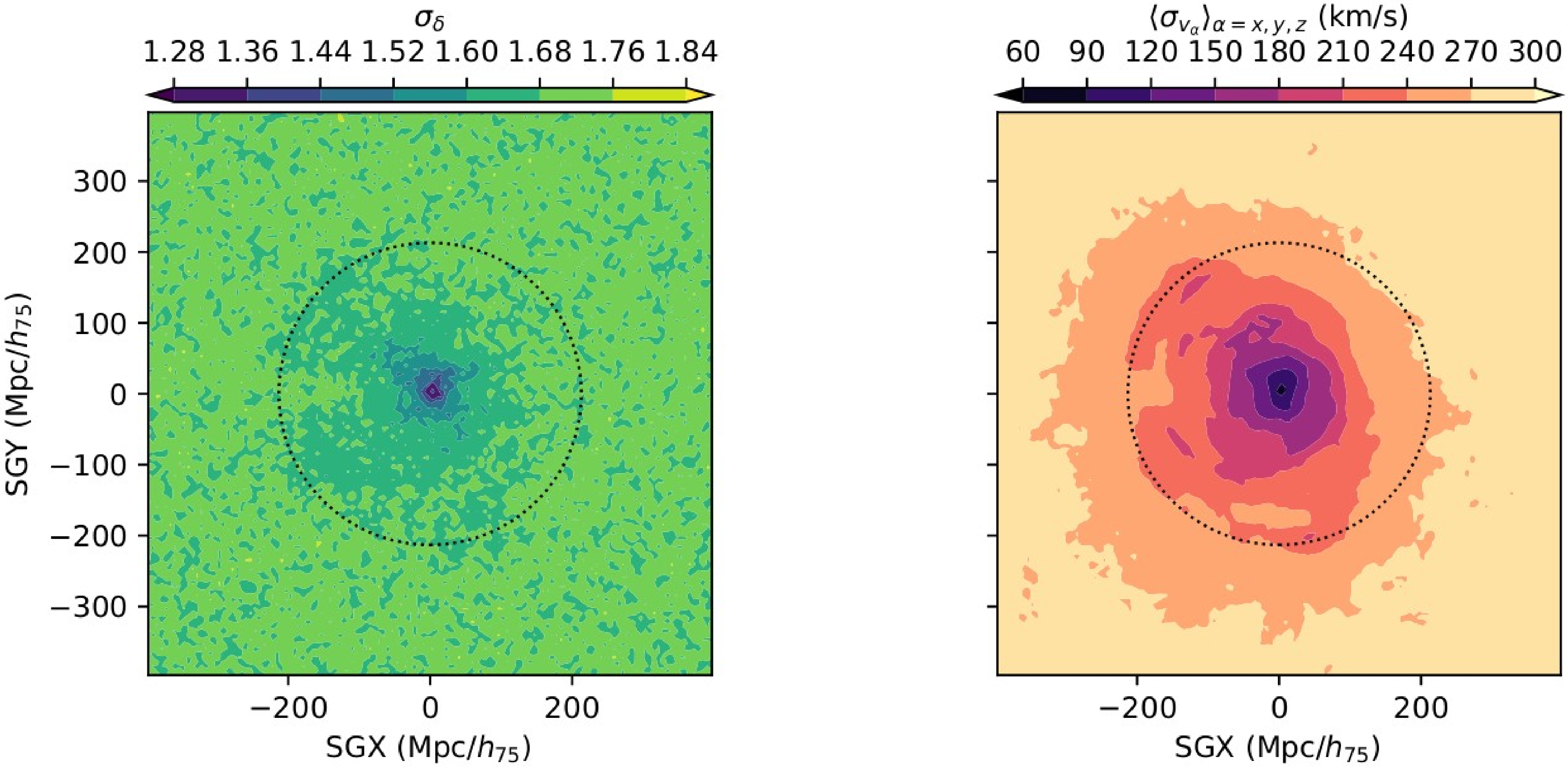
Graziani, Courtois, Lavaux et al. (2018 submitted)

CF3: Calibration checks



Graziani, Courtois, Lavaux et al. (2018 submitted)

Error on density / velocity



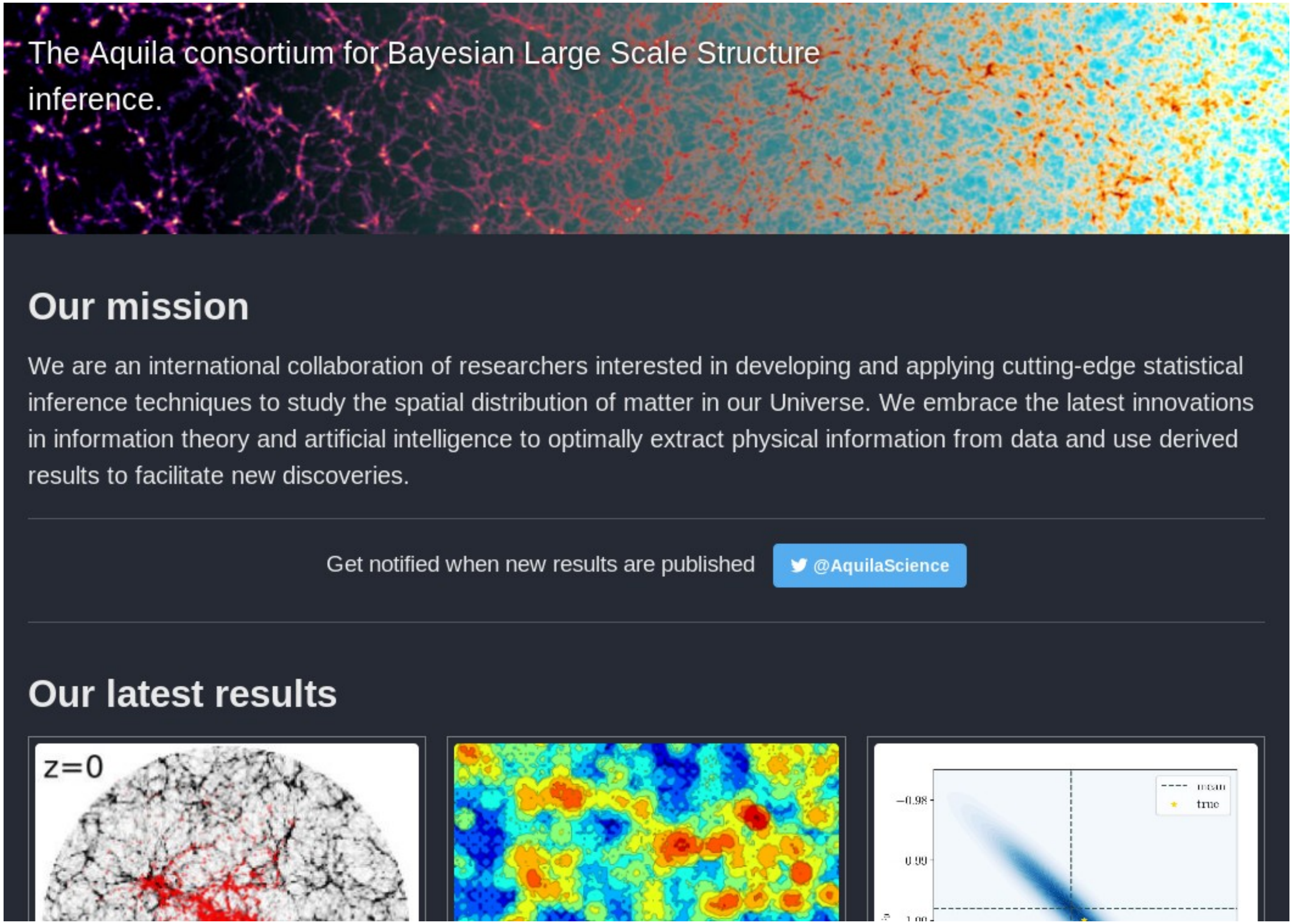
Graziani, Courtois, Lavaux et al. (2018 submitted)

The Path forward / Conclusion

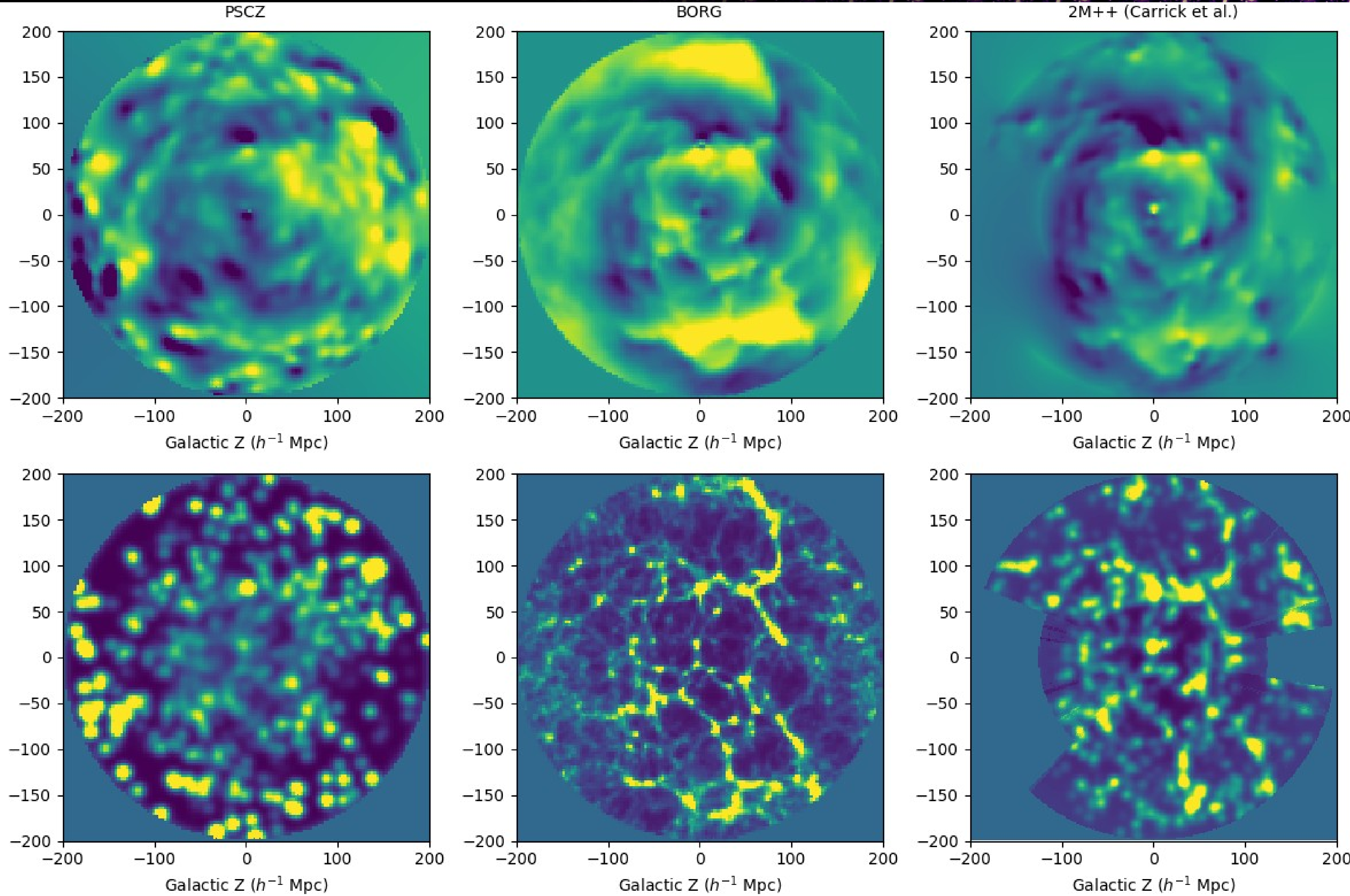
The Aquila consortium

- 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/

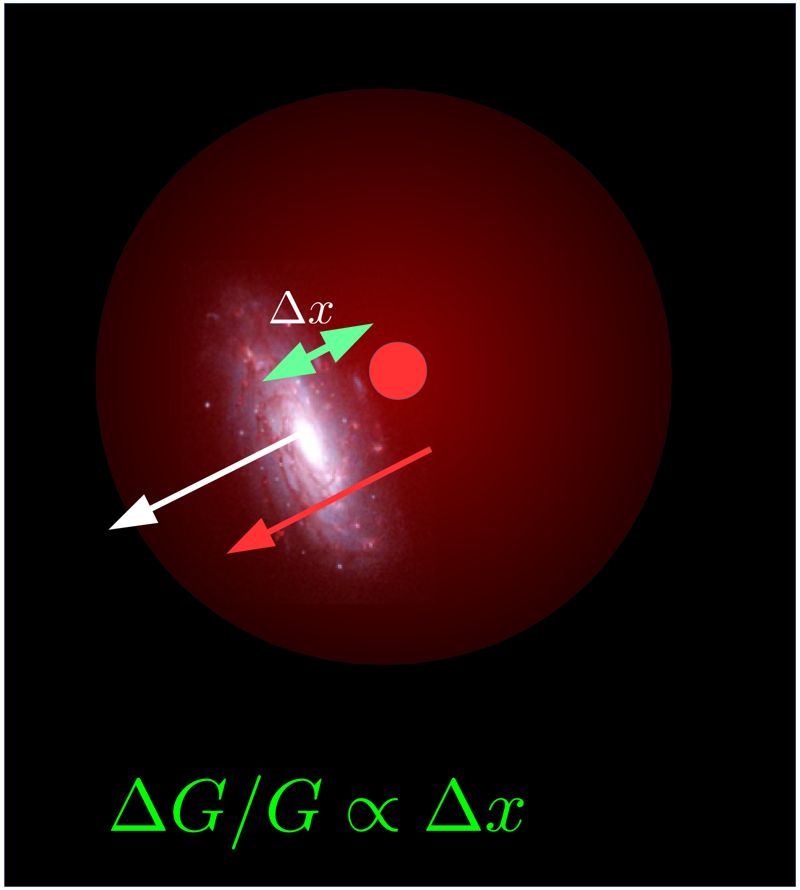


Predictive cosmology: cosmic flows



Lavaux & Jasche (2018, in prep.)

Predictive cosmology: fifth force



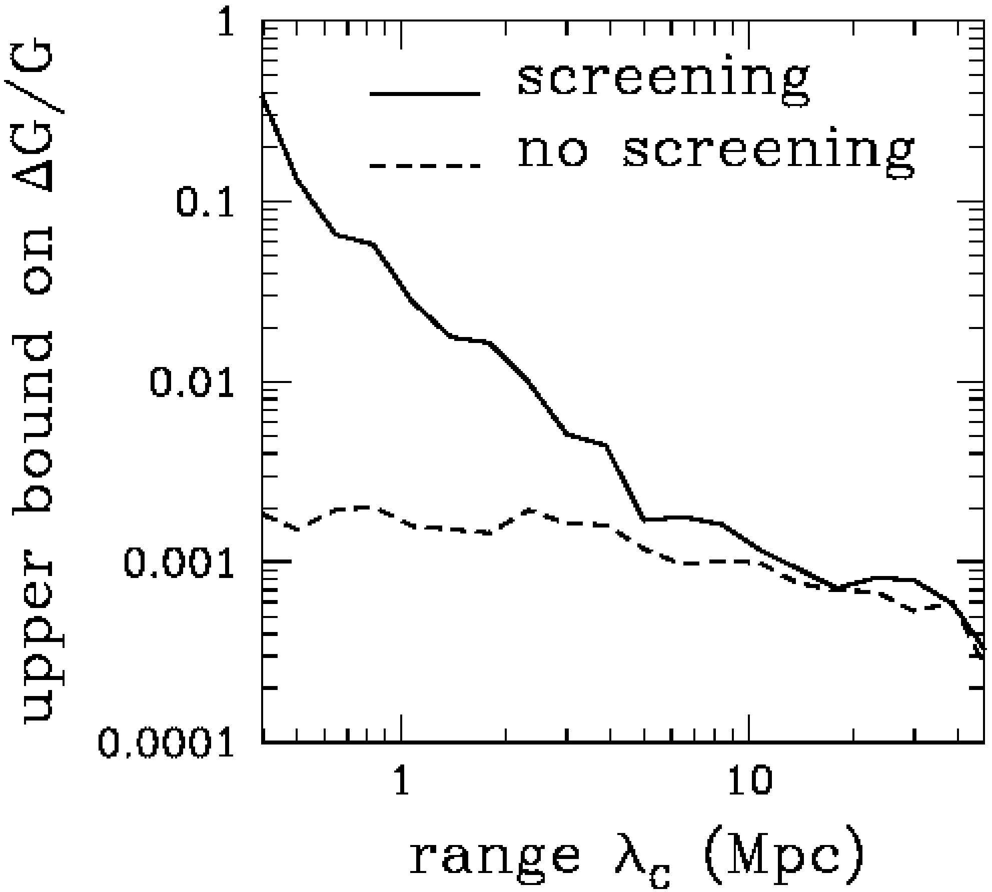
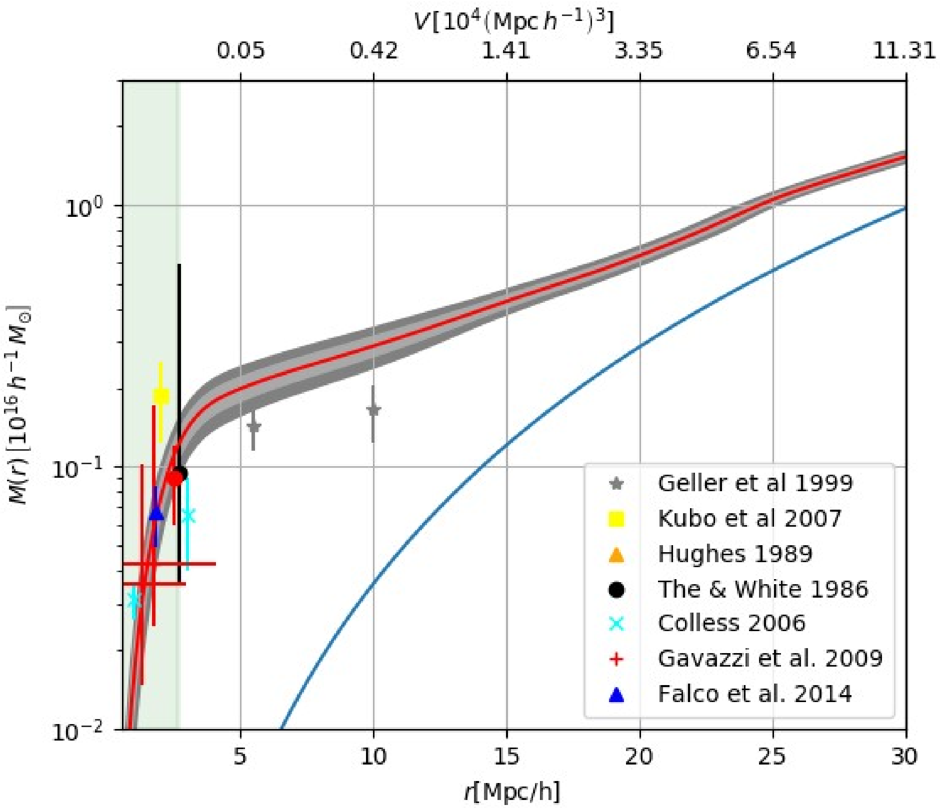
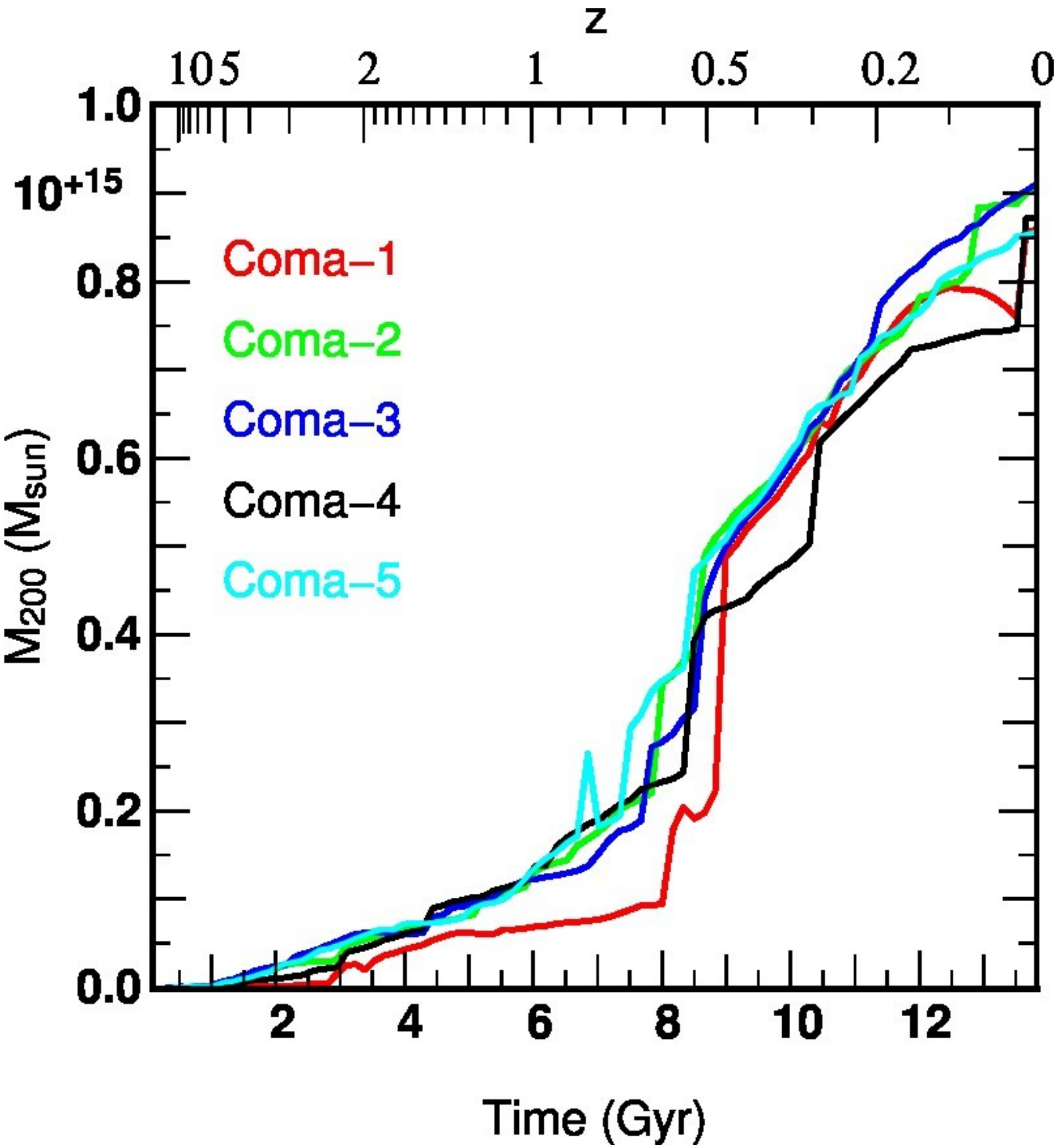


Image credit: Sukyong Yi

Desmond et al. (2017, 2018a,b,c)

Predictive cosmology: Coma dynamics



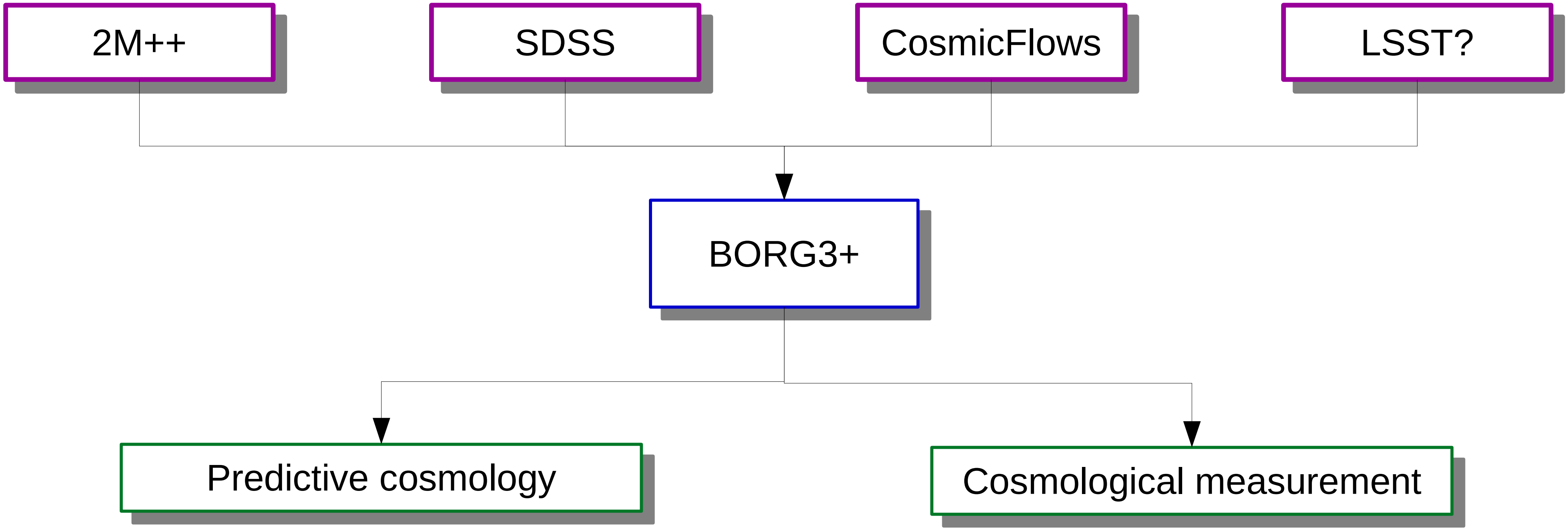


Zoom simulation on Coma
(~250 Mpart in zoom)

$$4 \times 10^7 h^{-1} \mathrm{M}_\odot/\mathrm{part}$$

Jasche & Lavaux; Lavaux & Jasche; Peirani, Lavaux & Jasche (2018, in prep.)

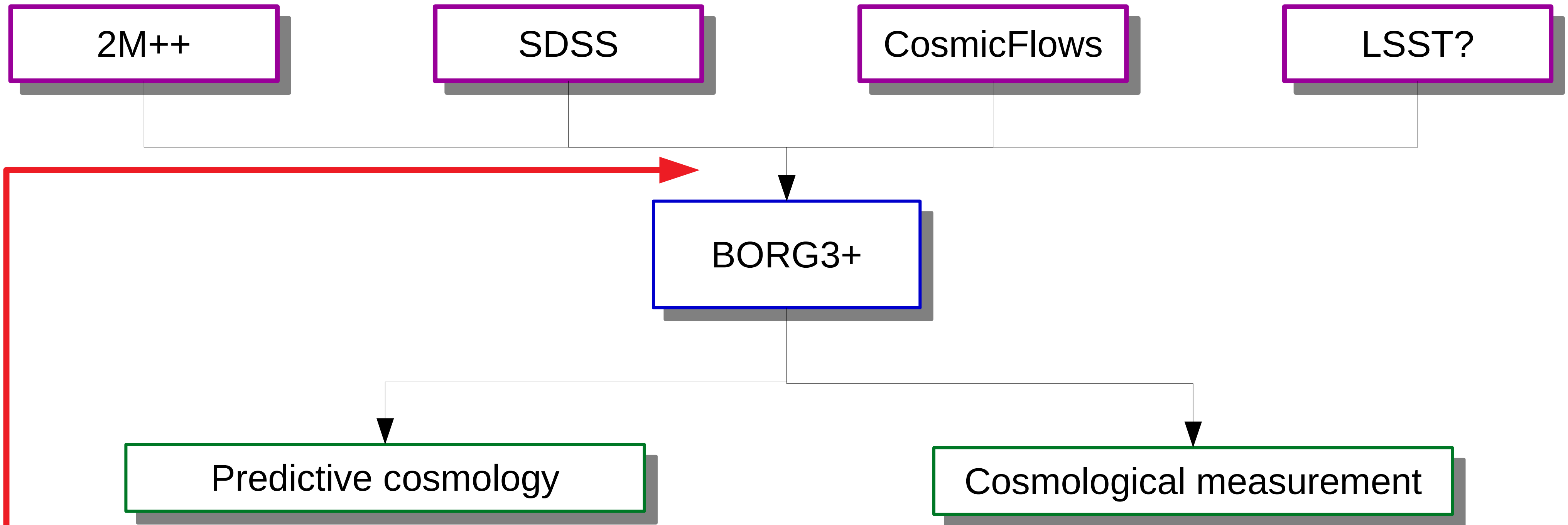
Conclusion: great future



- Velocity field (also VIRBIUS with F. Fuhrer)
- 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