Universidad de La Laguna

Modelling and inferring the relationship between the dark matter and the galaxy distribution

Francisco-Shu Kitaura

ESAC Madrid 13/03/2019

The golden era of galaxy surveys

Large number of large-scale galaxy surveys are being prepared after the success of SDSS I/II/III (BOSS), 2dF, 2MASS, Vipers, WiggleZ, Wise

Ongoing SDSS IV (eBOSS), DES

New generation of galaxy surveys:

DESI, EUCLID, J-PAS, 4MOST, LSST, Subaru PFS, Taipan, WFirst, ...

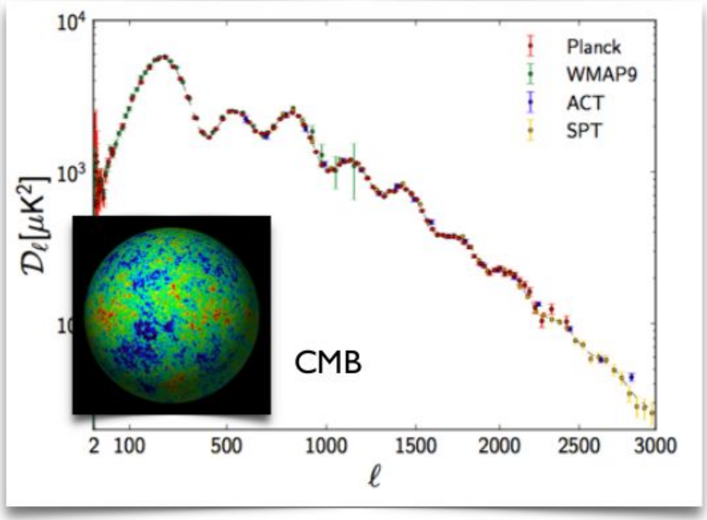
But what are they looking for?

understanding galaxy formation and evolution, structure formation …. solving the big mystery of the accelerated expansion of the Universe

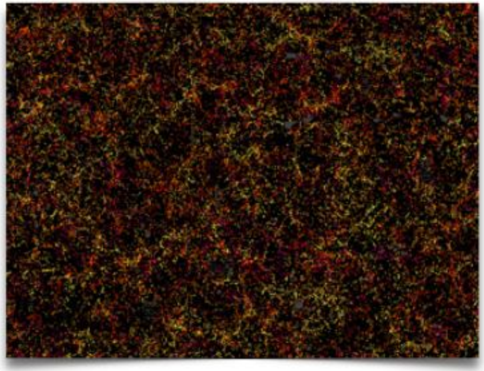
STANDARD RULERS

BARYON ACOUSTIC OSCILLATIONS

Planck collaboration 2013



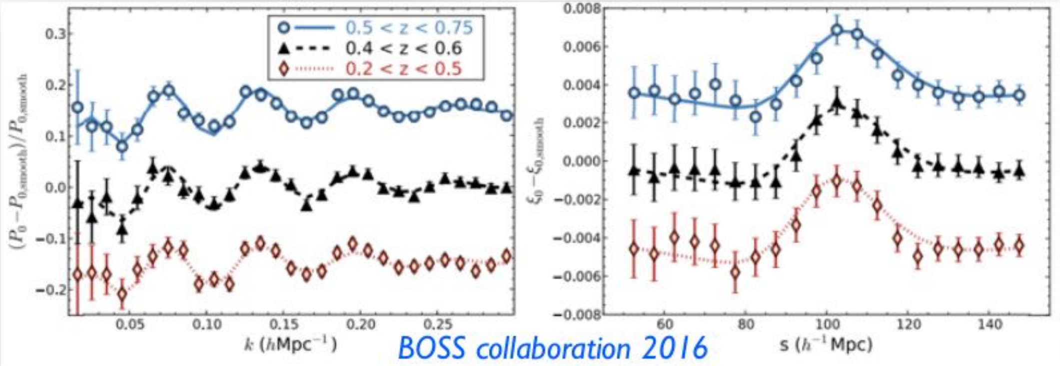
photon pressure in each DM&baryon overdensity causes spherical sound waves with 57% light speed



CMASS galaxies BOSS DR12

galaxies more likely to appear in spheres of ~490M lys radius

these sharp BAOs are reconstructed! and how do we get error bars?

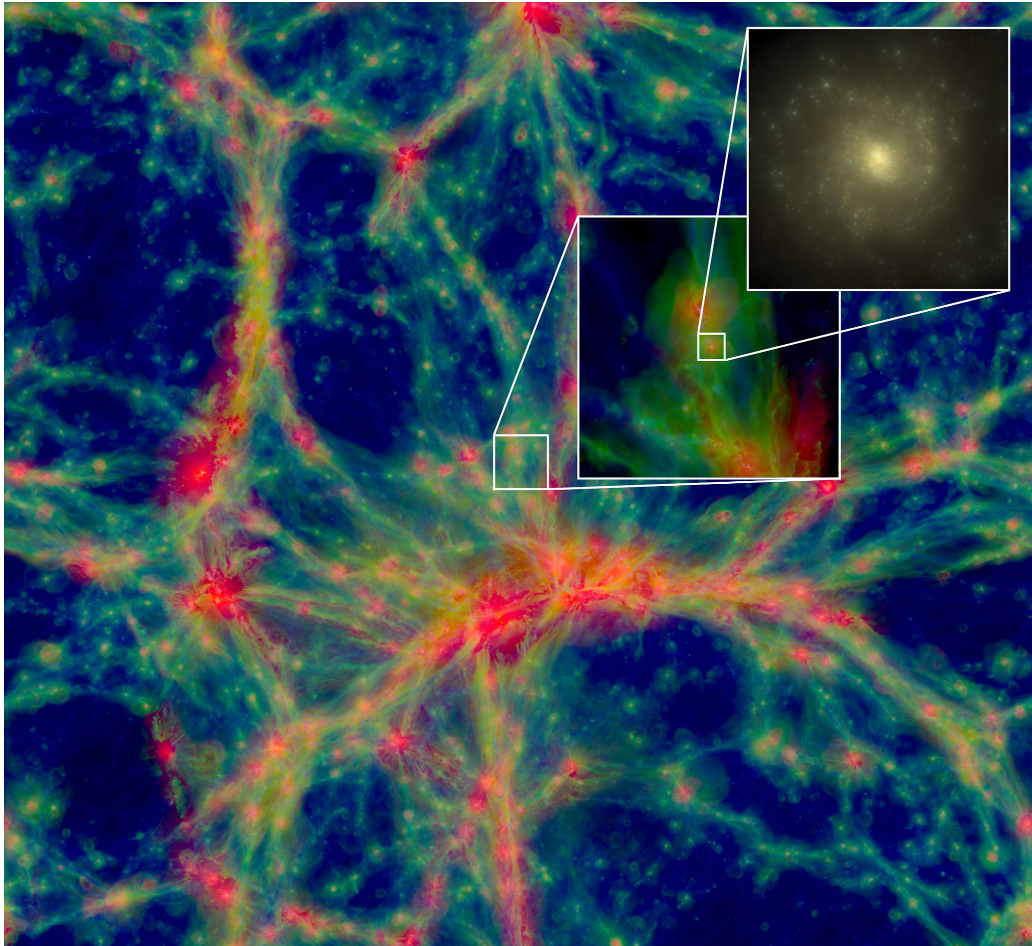


BOSS collaboration 2016

Why not directly simulate the galaxy distribution with all the involved physics?

Current best simulations cover volumes of about (200 Mpc/h) cube

Illustris, Eagle simulations ...



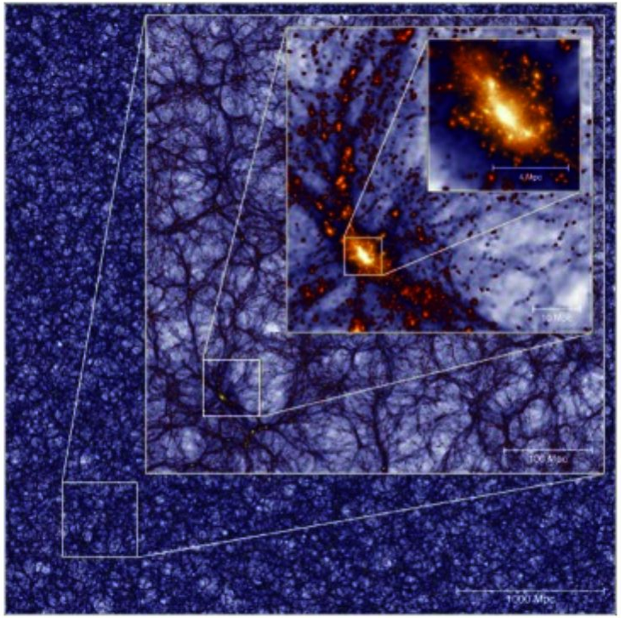
Can we simulate the whole Universe?

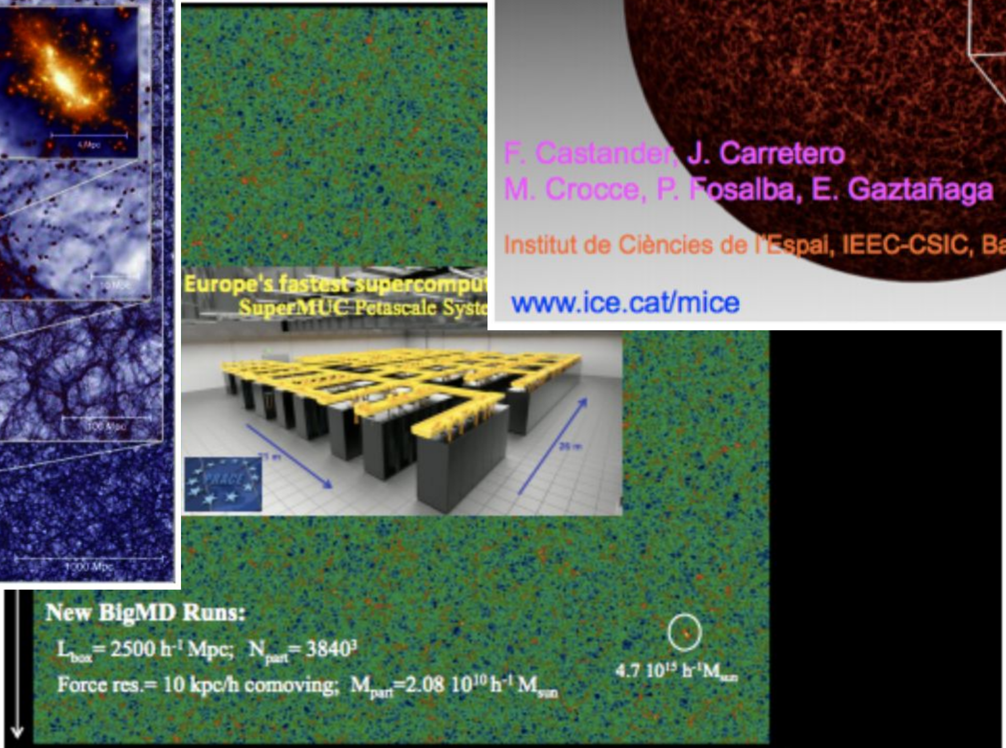
Let us look at the numbers (Anatoly Klypin):

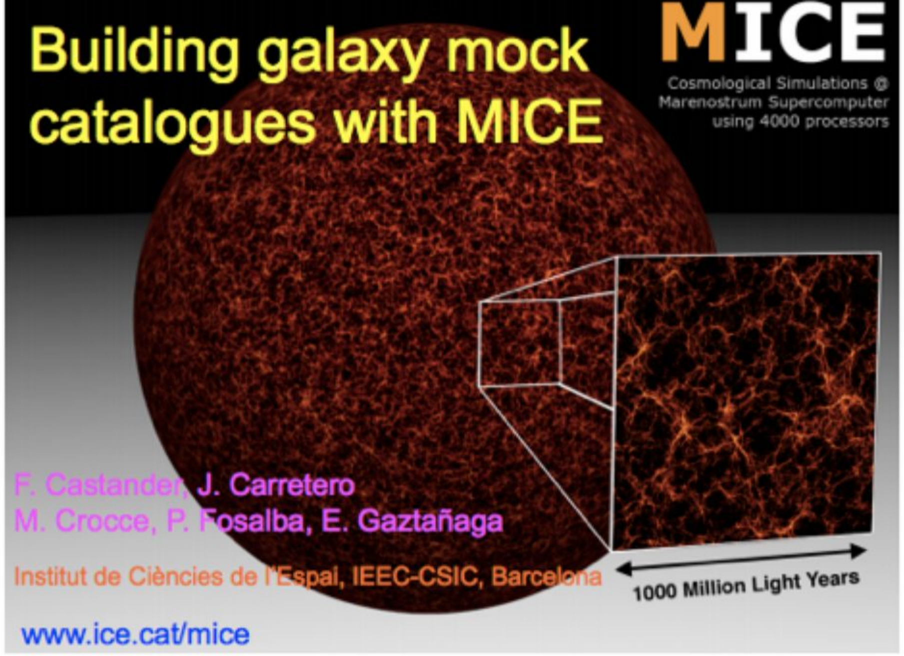
- Galaxies: the number of MW-type galaxies in 1Gpc cube volume is about 10,000,000
- Typical number of required realisations of Universes: 1000
- Total number of galaxies: 10^{10}
- One simulated galaxy with the resolution of 10pc to 1Mpc requires 500,000 CPU hrs
- Total required computational time is 10^{14} CPU hrs
- Total available computational time in the world 10^{11} CPU hrs

Dark Matter only simulations!

MXXL, MICE, BigMD, HORIZON, DEUS,
DARK SKY, FLAGSHIP simulation
~10-50M CPU HRS







A few large simulations

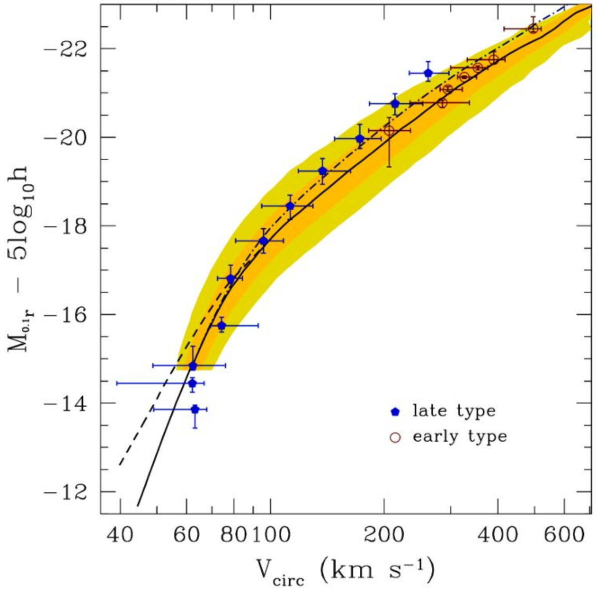
They even have names!

How can we connect the dark matter to the galaxy distribution?

- Semi-analytic models (e.g., Cole 00; Haton 03; Croton 06; Bower 06; Monaco 2007; Benson 10; deLucia & Blaizot 07; Baugh 06)
- HOD (e.g., Berlind & Weinberg 02; Zheng 05; Zehavi et al 11)
- HAM or SHAM (e.g., Kratsov 04; Tasitsiomi 04; Vale & Ostriker04; Conroy 06; Behroozi et al 10; Trujillo-Gomez et al 11; Nuza et al 12; Rodriguez-Torres et al 16; Favole et al 16)

There is a tight relation between circular velocity in galaxies and dark matter in central regions of galaxies

(e.g., Trujillo-Gomez et al 11)



Are there faster ways of producing mock galaxy catalogs? Approximate methods

Analytic halo formation models:

- Peak-Patch Method (Bond et al 96)
- Pinocchio (Monaco et al 02, 13)
- PThalos (Scoccimarro & Sheth 02; Manera et al 11, 13)

many equations
many parameters

Fast gravity solvers:

- COLA,L-PICOLA (Tassev et al 13; Koda et al 16; Izard et al 16)
- FastPM (Feng et al 16)
- PPMGLAM (Klypin et al 17)

requires same number of particles as a full N-body calculation to resolve distinct halos (substructures cannot be modelled)

Combination of fast gravity solvers with halo or galaxy bias prescriptions

- Patchy (FSK et al 14,15,16; Vakili, FSK et al 17)
- QPM (White et al 14)
- EZmocks (Chuang, FSK et al 15)
- Halogen (Avila et al 15)

methods calibrated on full calculations
accuracy depends on bias prescription and gravity treatment

PATCHY approach

1. Low resolution approximate gravity solver → large-scale dark matter field

   Trust the approximate gravity solver
   only on a few Mpc scales!

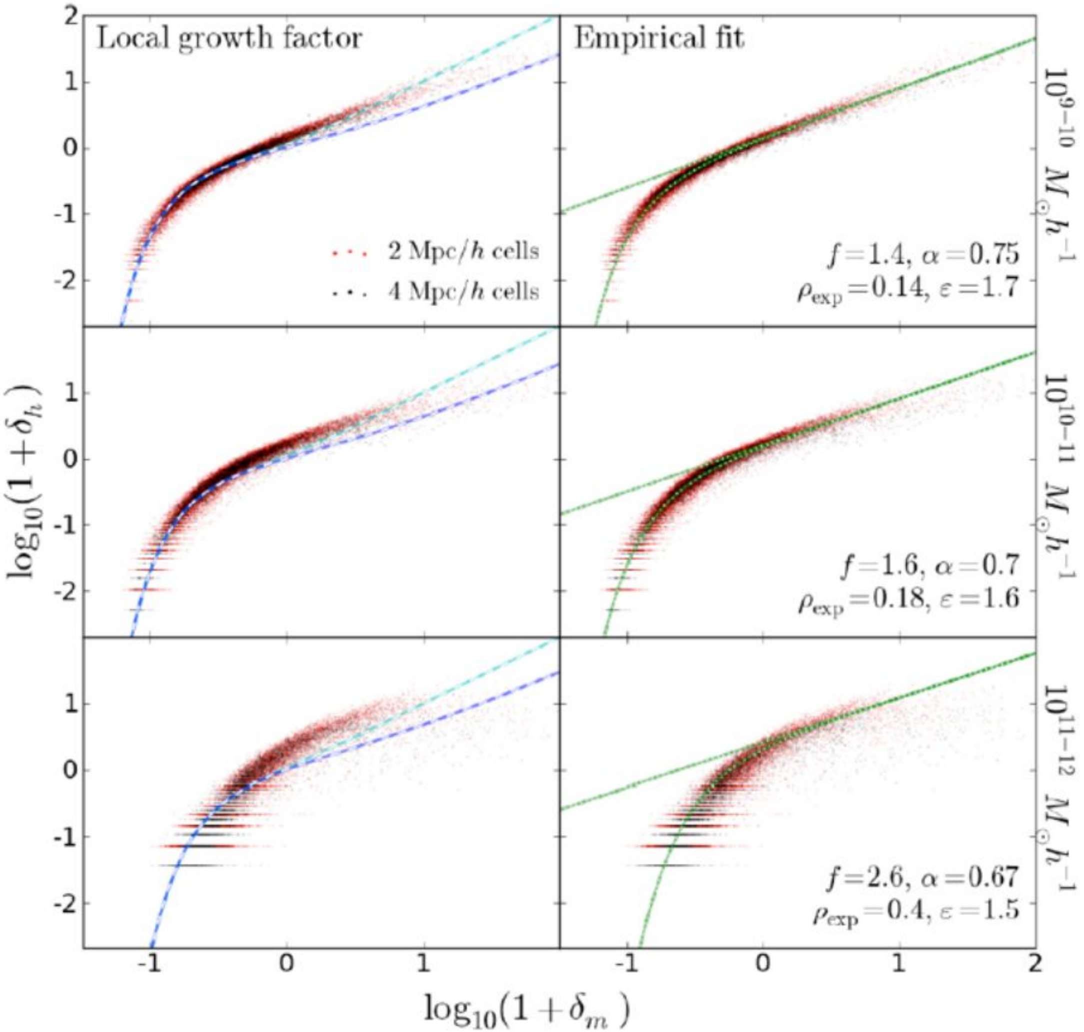
2. Bias model calibrated on a reference simulation → halo or galaxy catalog

   Find the effective bias model calibrated on
   simulations or observations

FSK & Hess 13	fast Lagrangian PT method
FSK, Yepes & Prada 14	deterministic and stochastic bias
FSK, Gil-Marin, Scoccola, Chuang et al 15	constraining higher order statistics
Zhao, FSK, Chuang et al 15	assigning masses
FSK , Rodriguez-Torres, Chuang et al 16	fitting mock galaxies, BOSS mocks

Effective bias
From N-body sims

Neyrinck et al 14
Aragón-Calvo et al 14



Parametric bias model

Nonlinear deterministic bias

Kaiser 84
Fry & Gaztanaga 93
de la Torre & Peacock 13
FSK et al 14, 15, 16
Neyrinck et al 14

Stochastic component

Peebles 80
Sheth 95
Lahav & Lemson 99
FSK et al 14, 15, 16
Neyrinck et al 14

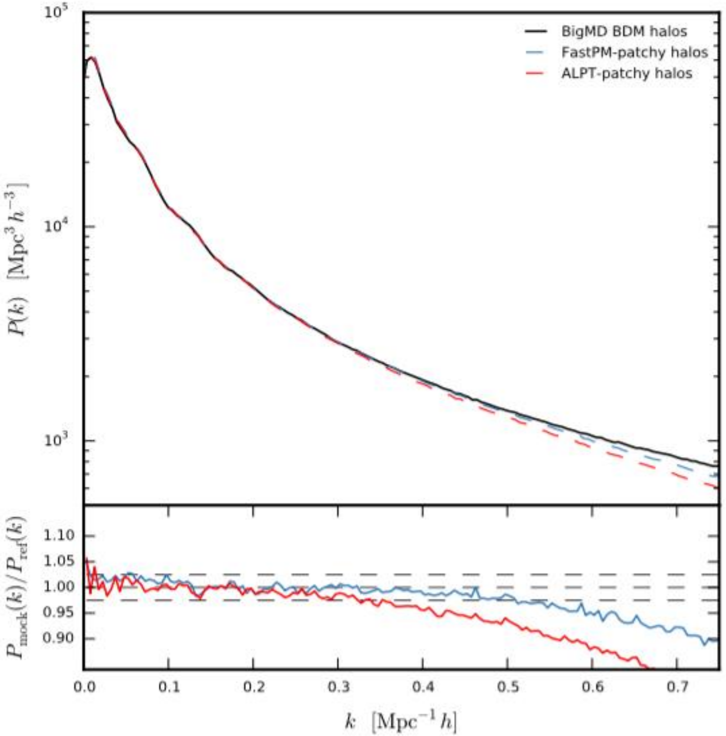
$$\langle \rho_{\rm h} \rangle_{{\rm d}V} = f_{\rm h}\, B(\rho_{\rm h}|\rho_{\rm m}),$$

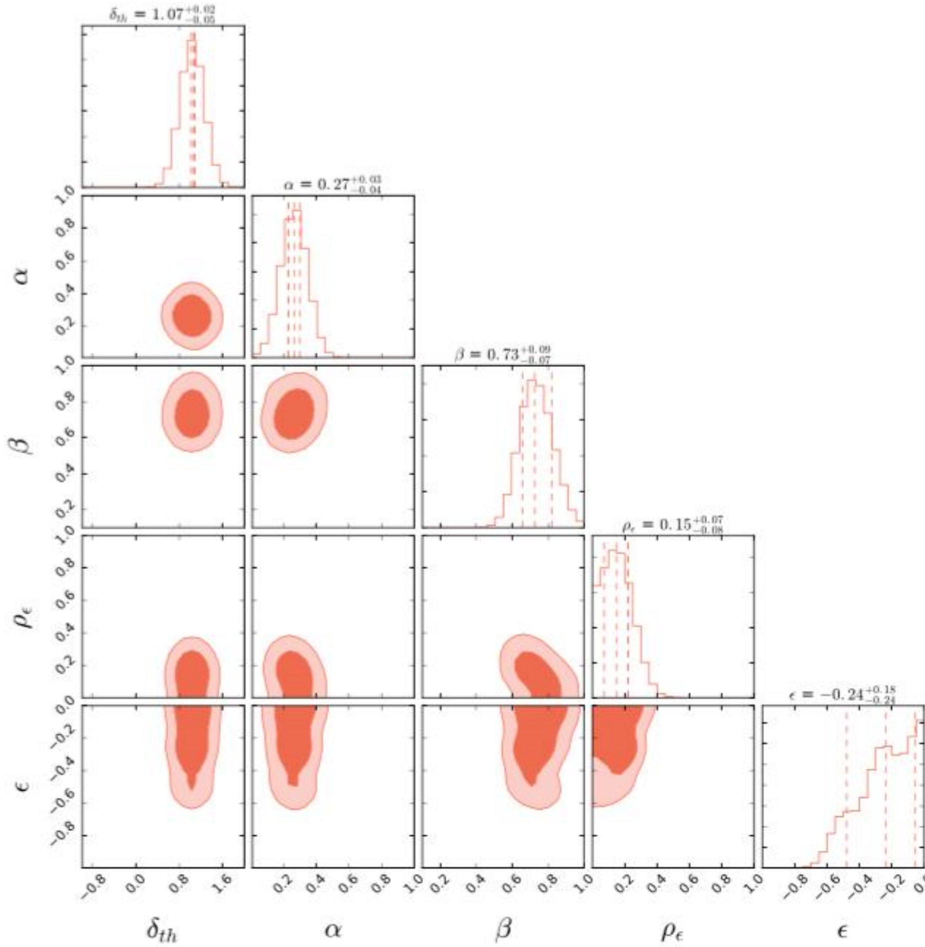
$$f_{\rm h} = \frac{n_{\rm h}}{\langle B(\rho_{\rm h}|\rho_{\rm m})\rangle_V},$$

$$B(\rho_{\rm h}|\rho_{\rm m}) = \underbrace{\rho_{\rm m}^{\alpha}}_{\text{nonlinear bias}} \times \underbrace{\theta(\rho_{\rm m} - \rho_{\rm th})}_{\text{threshold bias}} \times \underbrace{\exp\left(-(\rho_{\rm m}/\rho_\epsilon)^\epsilon\right)}_{\text{exponential cutoff}},$$

$$P(N_{\rm h}|\lambda_{\rm h},\beta) = \underbrace{\frac{\lambda_{\rm h}^{N_{\rm h}}}{N_{\rm h}!}e^{-\lambda_{\rm h}}}_{\text{Poisson distribution}} \times \underbrace{\frac{\Gamma(\beta+N_{\rm h})}{\Gamma(\beta)(\beta+\lambda_{\rm h})^{N_{\rm h}}} \times \frac{e^{\lambda_{\rm h}}}{(1+\lambda_{\rm h}/\beta)^{\beta}}}_{\text{Deviation from Poissonity}}.$$

PATCHY approach





Mohammadjavad Vakili, FSK et al 17 fastPM, MCMC automatic bias calibration

Comparison project within EUCLID

special Credit to: Linda Blot (et al 19), Martha Lippich (et al 19), Manuel Colavincenzo (et al 19)
Ariel Sánchez, Martín Crocce, Emiliano Sefusatti, Pierluigi Monaco, Claudio dalla Vecchia

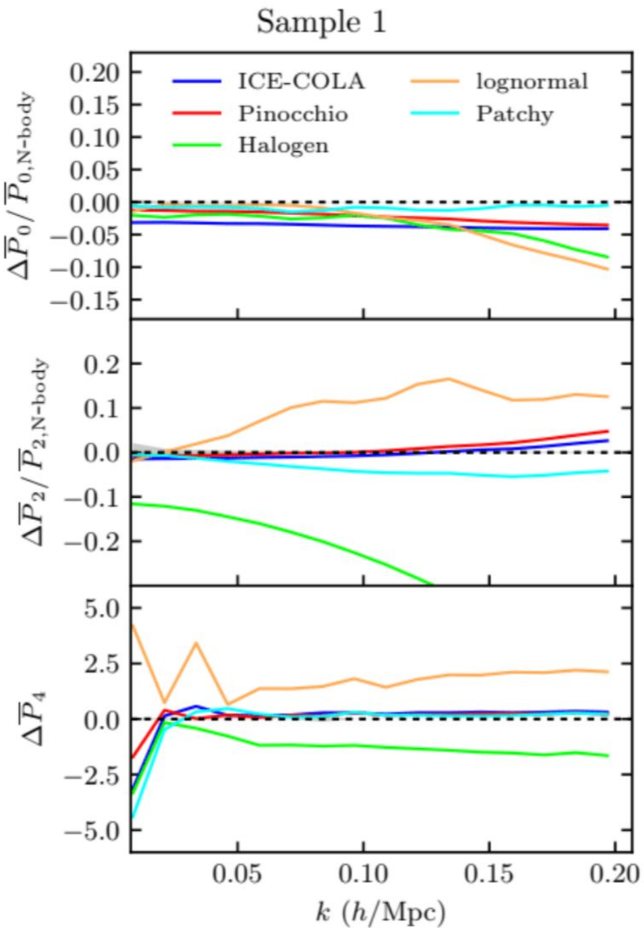
Method	Algorithm	Computational Requirements	Reference
Minerva	**N-body** Gadget-2 Halos : SubFind	CPU Time: 4500 hours Memory allocation: 660 Gb	Grieb et al. (2016) *https://wwwmpa.mpa-garching.mpg.de/gadget/*
ICE-COLA	Predictive 2LPT + PM solver Halos : FoF(0.2)	CPU Time: 66 hours Memory allocation: 340 Gb	Izard, Crocce & Fosalba (2016) Modified version of: *https://github.com/junkoda/cola_halo*
PINOCCHIO	Predictive 3LPT + ellipsoidal collapse Halos : ellipsoidal collapse	CPU Time: 6.4 hours Memory allocation: 265 Gb	Monaco et al. (2013); Munari et al. (2017b) *https://github.com/pigimonaco/Pinocchio*
PEAKPATCH	Predictive 2LPT + ellipsoidal collapse Halos : Spherical patches over initial overdensities	CPU Time: 1.72 hours* Memory allocation: 75 Gb*	Bond & Myers (1996a,b,c) Not public
HALOGEN	Calibrated 2LPT + biasing scheme Halos : exponential bias	CPU Time: 0.6 hours Memory allocation: 44 Gb Input: $\bar{n}$, 2-pt correlation function halo masses and velocity field	Avila et al. (2015). *https://github.com/savila/halogen*
PATCHY	Calibrated ALPT + biasing scheme Halos : non-linear, stochastic and scale-dependent bias	CPU Time: 0.2 hours Memory allocation: 15 Gb Input: $\bar{n}$, halo masses and environment	Kitaura, Yepes & Prada (2014) Not Public

Reference N-body simulations
4500 CPU hrs
660 Gb

approximate analytic solvers
2 orders of magnitude less CPU hrs
Same order of magnitude memory requirements

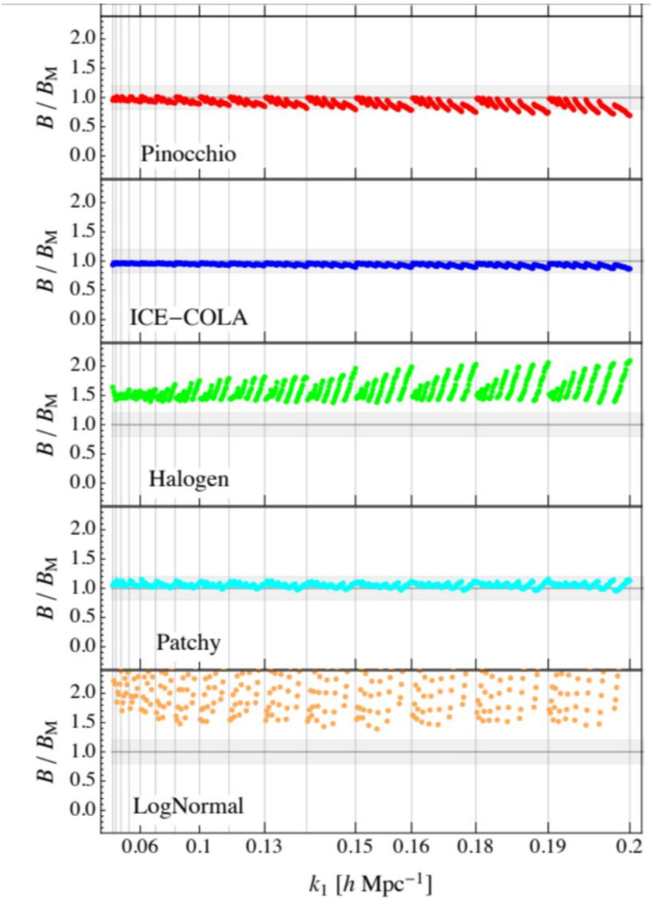
PATCHY code
4 orders of magnitude less CPU hrs
2 orders of magnitude less memory requirements

Comparison project within EUCLID

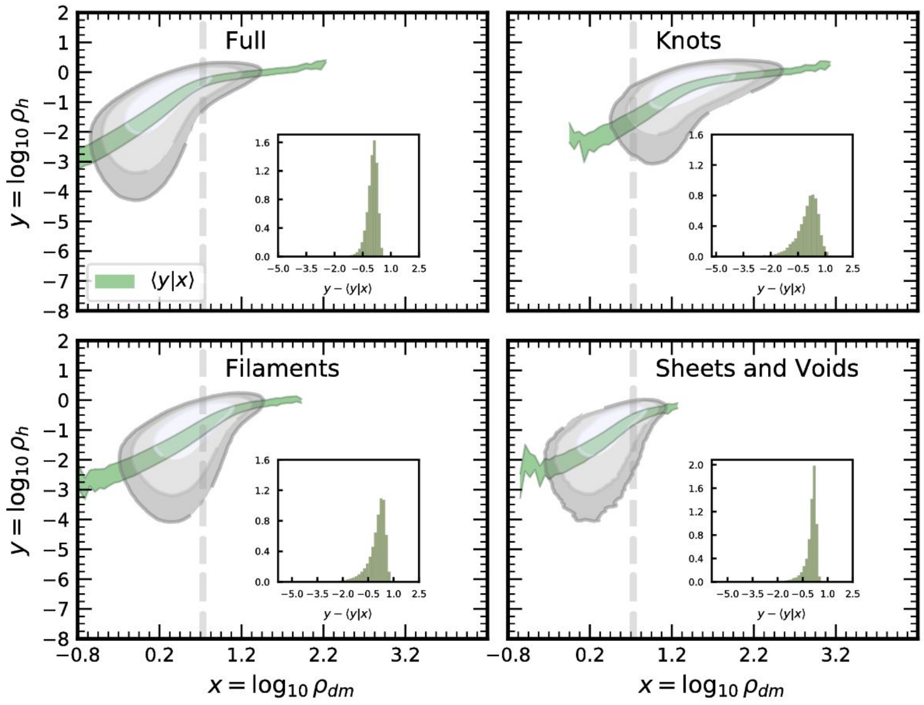


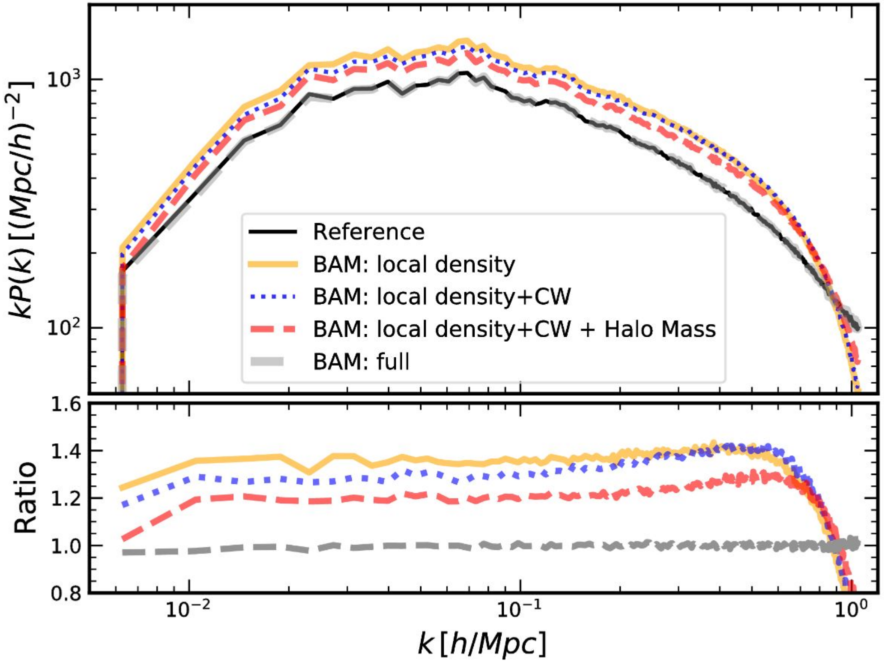
Most accurate method in 2-point statistics

At the level of N-body solvers in 3-point statistics



Why not directly map the bias relation from accurate calculations?
BAM percentage accuracy up to the Nyquist frequency

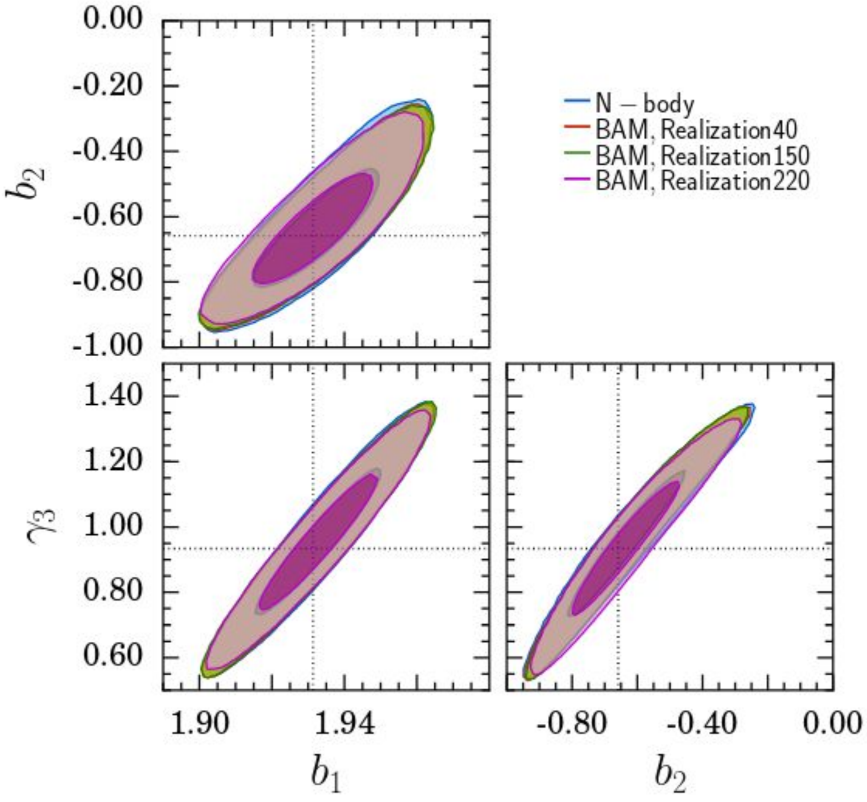


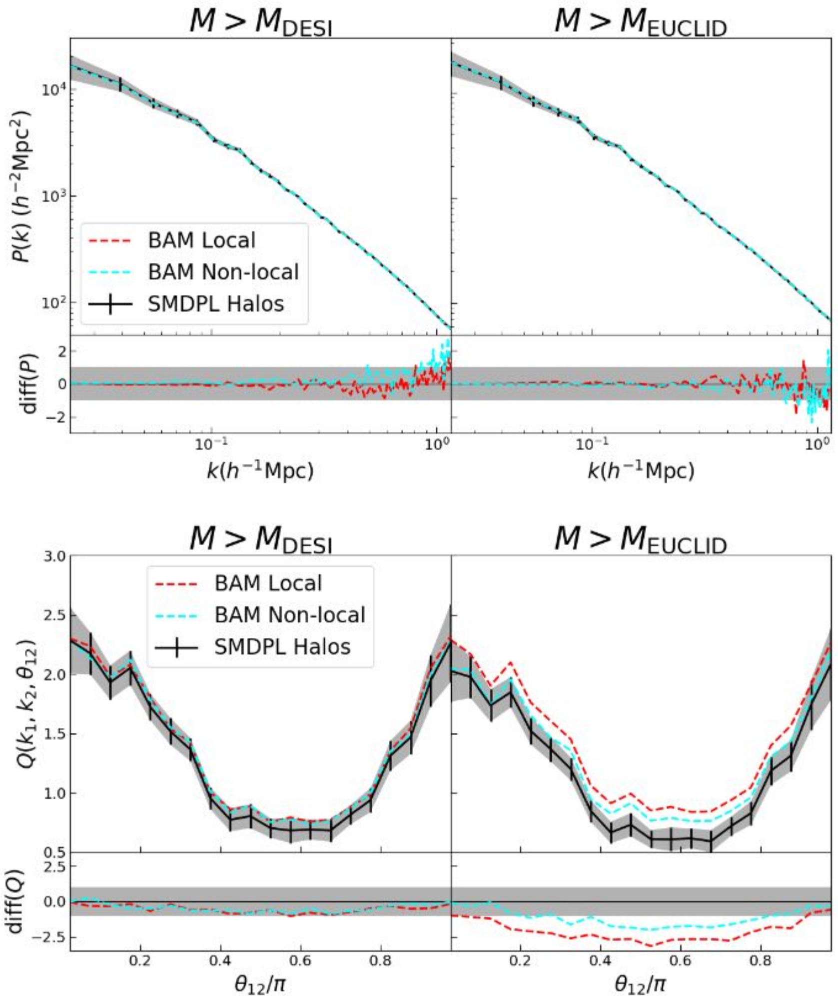


Learn the effective bias relation from N-body simulations including local and nonlocal dependencies

Andres Balaguera-Antolinez, FSK, Pellejero-Ibanez, Zhao, & Abel 19 arXiv:1806.05870

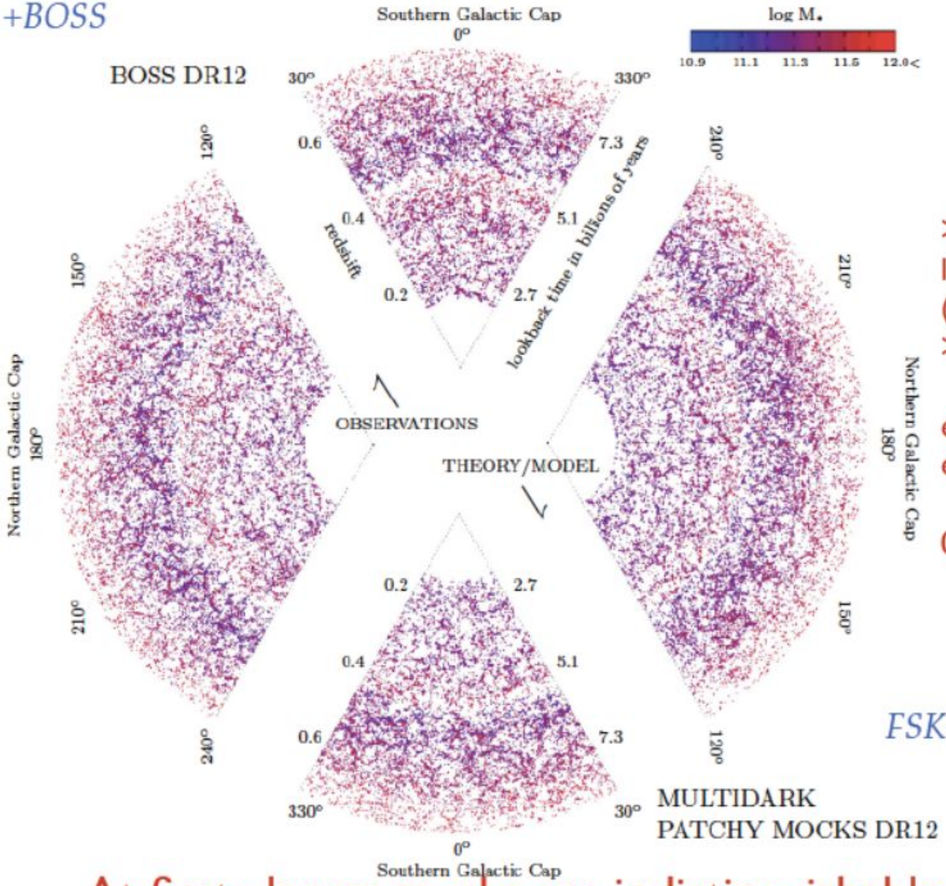
BAM 2-, 3-, and 4- point statistics.
Power, Bispectra and Covariance matrices





PATCHY MULTIDARK BOSS DR11/DR12

collaborators Chia-Hsun Chuang, Sergio Rodriguez-Torres, Cheng Zhao, F. Prada, G. Yepes, A. Klypin +BOSS



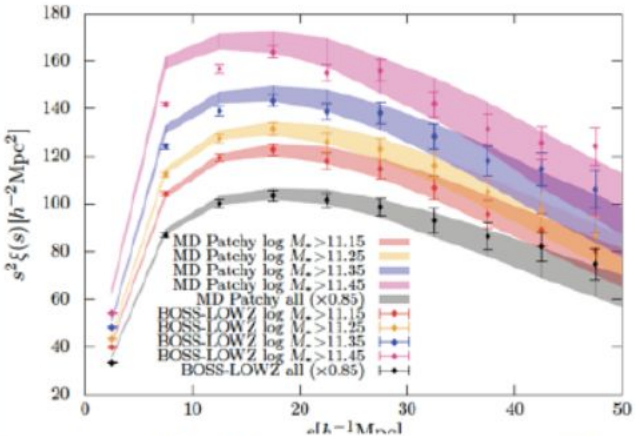
>12000 mocks including light cone and evolution effects (>4000 LOWZ, >4000 CMASS, >4000 COMBINED SAMPLE)

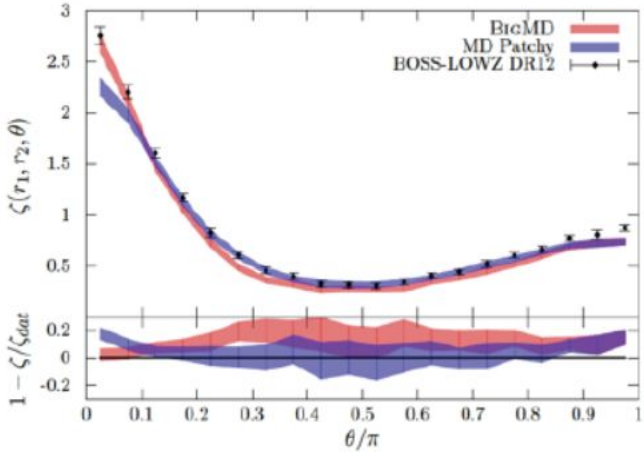
effective V ~ 192,000 [Gpc/h]^3
effective # p ~(61,440)^3

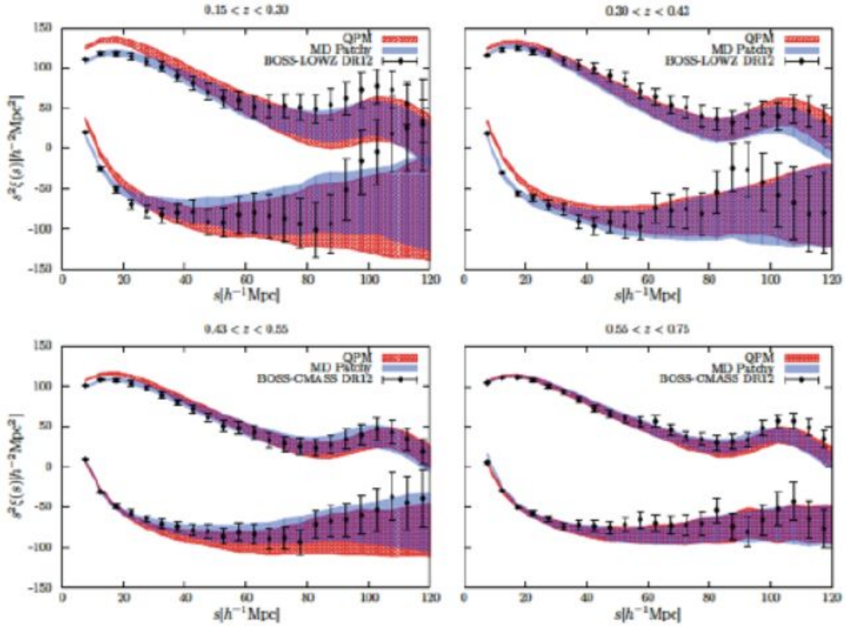
0.5M CPU hr vs 9billion CPU hr

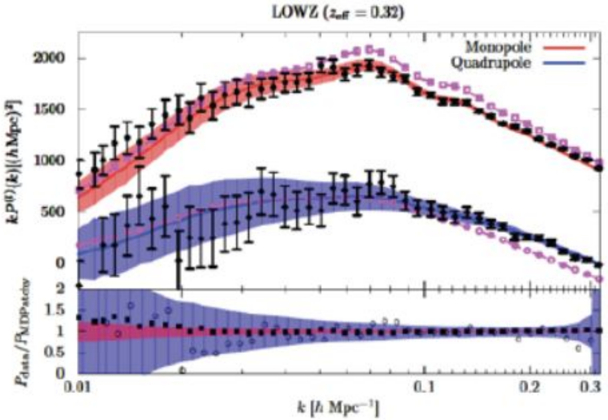
FSK+BOSS arXiv:1509.06400v1

At first glance: mocks are indistinguishable from observations!









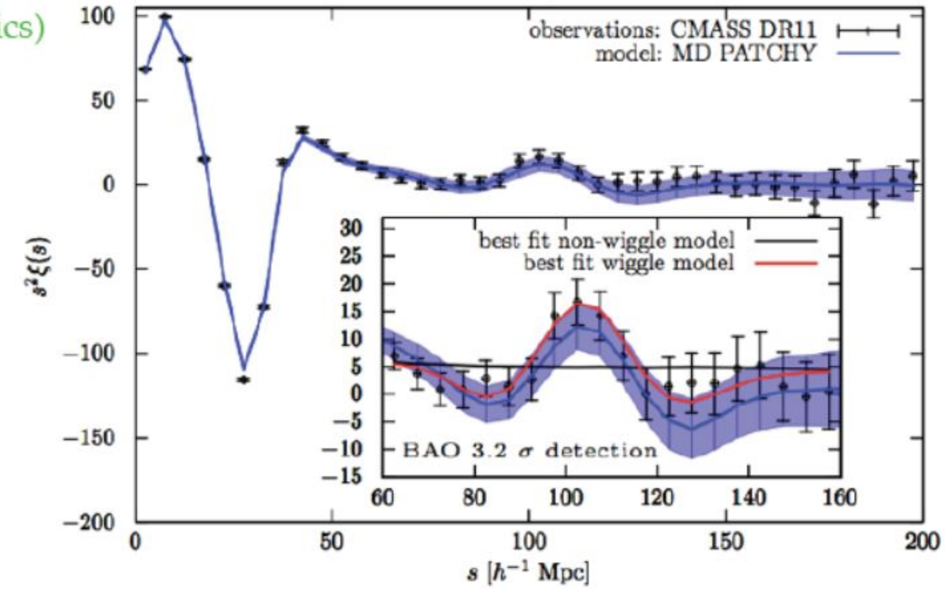
Accurate up to k~0.3, can we go beyond that?

FSK+BOSS arXiv:1509.06400v1

Application using PATCHY mocks:
First BAO from troughs in voids using BOSS data

collaborators Chia-Hsun Chuang, Yu Liang, Cheng Zhao, Charling Tao, Kitaura + BOSS

excellent agreement between observations and mocks (voids are constructed from the higher order statistics)



FSK, Chuang, Liang, Zhao, Charling+BOSS 15 arxiv:1511.04405

DIVE void finder code: *Zhao+15 arxiv:1511.04299*

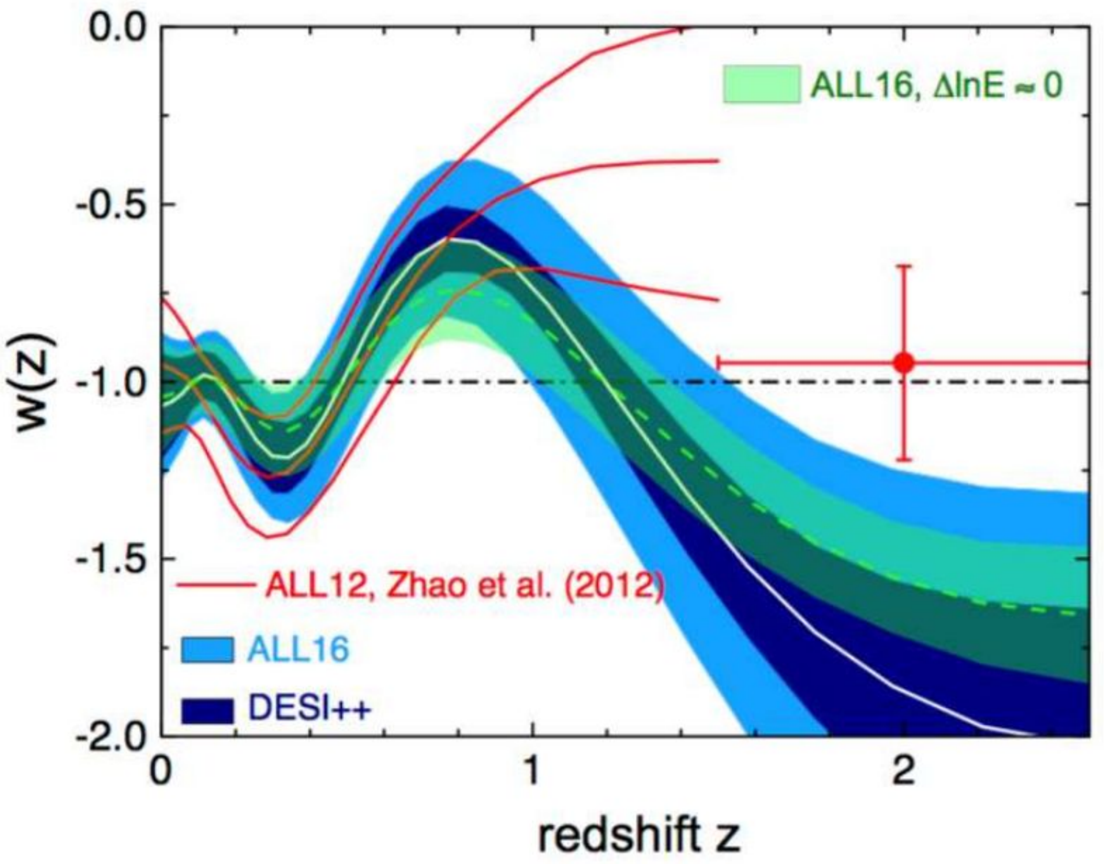
Optimal BAO detection from voids *Liang+15 arxiv:1511.04391*

Cheng Zhao's and Yu Liang's PhD thesis

Tsinghua University

Dynamical dark energy

Some evidence on evolving dark energy in the equation of state



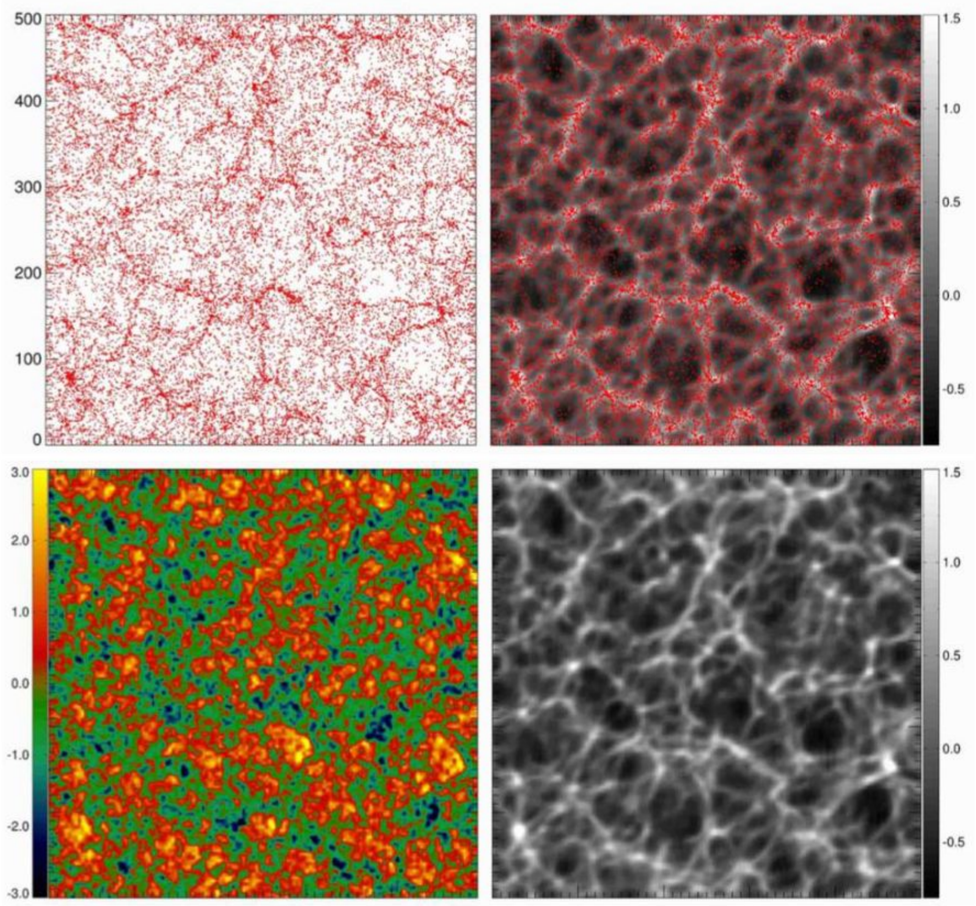
Gong-Bo Zhao et al 17 Nature Astronomy

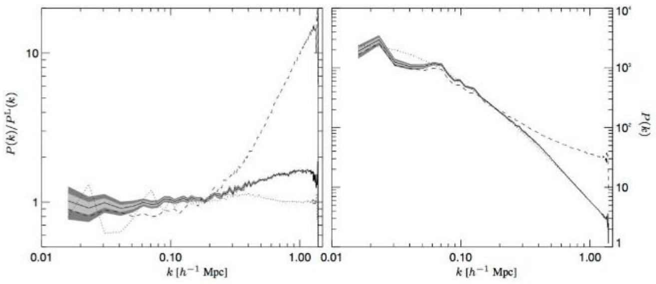
Can we infer the dark matter field from the galaxy distribution?

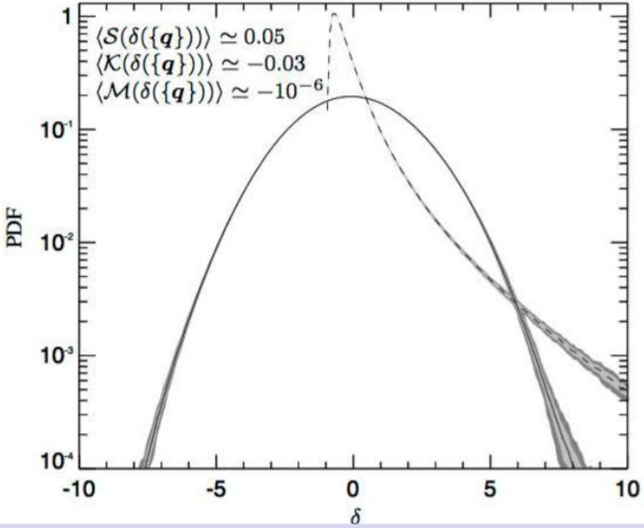
Forward modelling with Bayesian approaches taking advantage of the simple statistics (Gaussian) of the primordial Universe.

A forward modelling code KIGEN

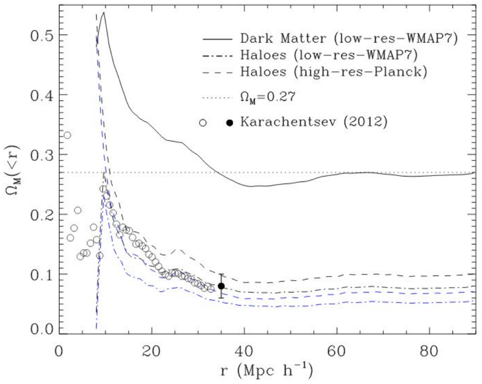
FSK 12

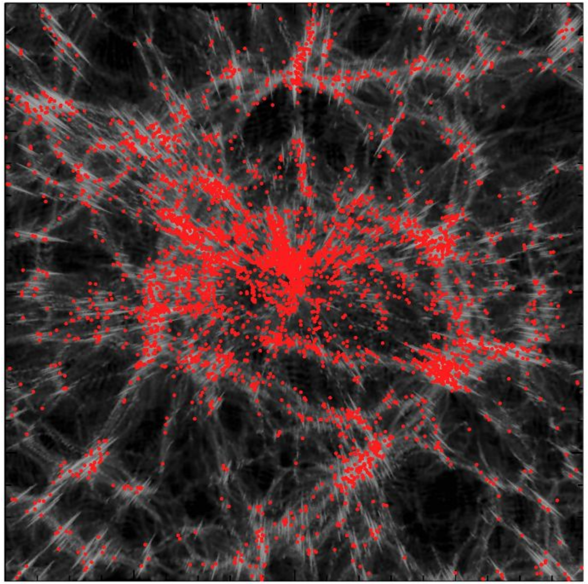


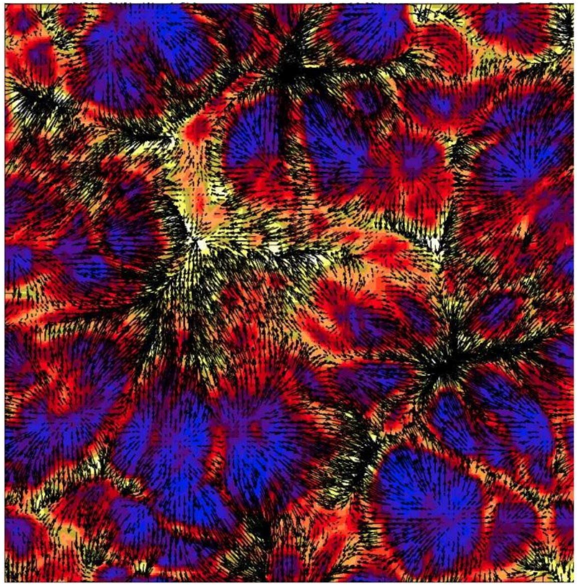


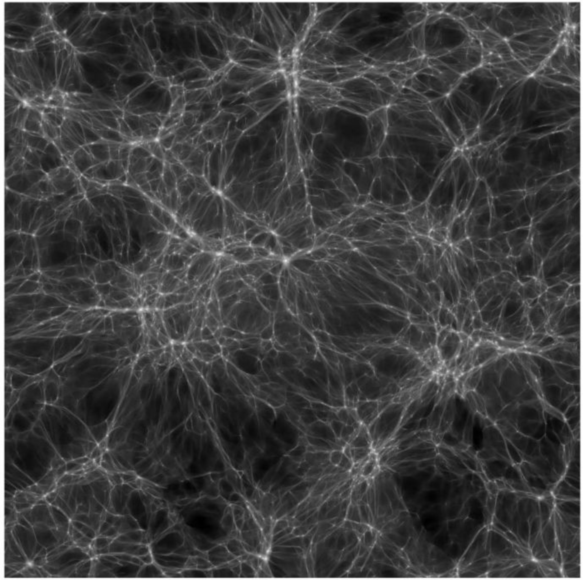


Application to the Local Universe 2MRS



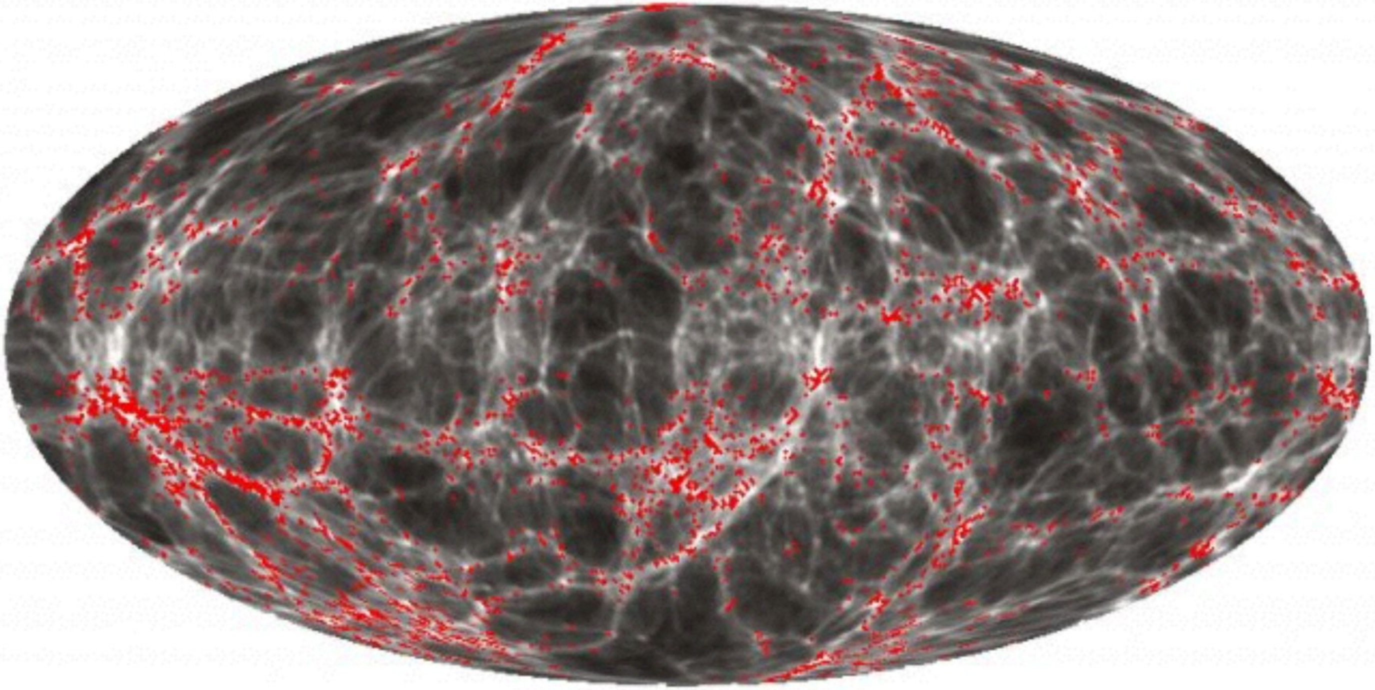






FSK et al 12; Hess, FSK et al 13; Nuza, FSK et al 14; Hess & FSK 16

Application of KIGEN to the Local Universe (FSK et al 2012)



Problems:
1) Bias including internal variables are chosen to fit an input power spectrum
2) Survey mask is ignored
Gaps are filled with random mock galaxies
3) Lightcone effects are ignored

FSK et al 12; Hess, FSK et al 13; Nuza, FSK et al 14; Hess & FSK 16

Getting ready for DESI, EUCLID ... : Cosmic BIRTH code

FSK in prep

mix between

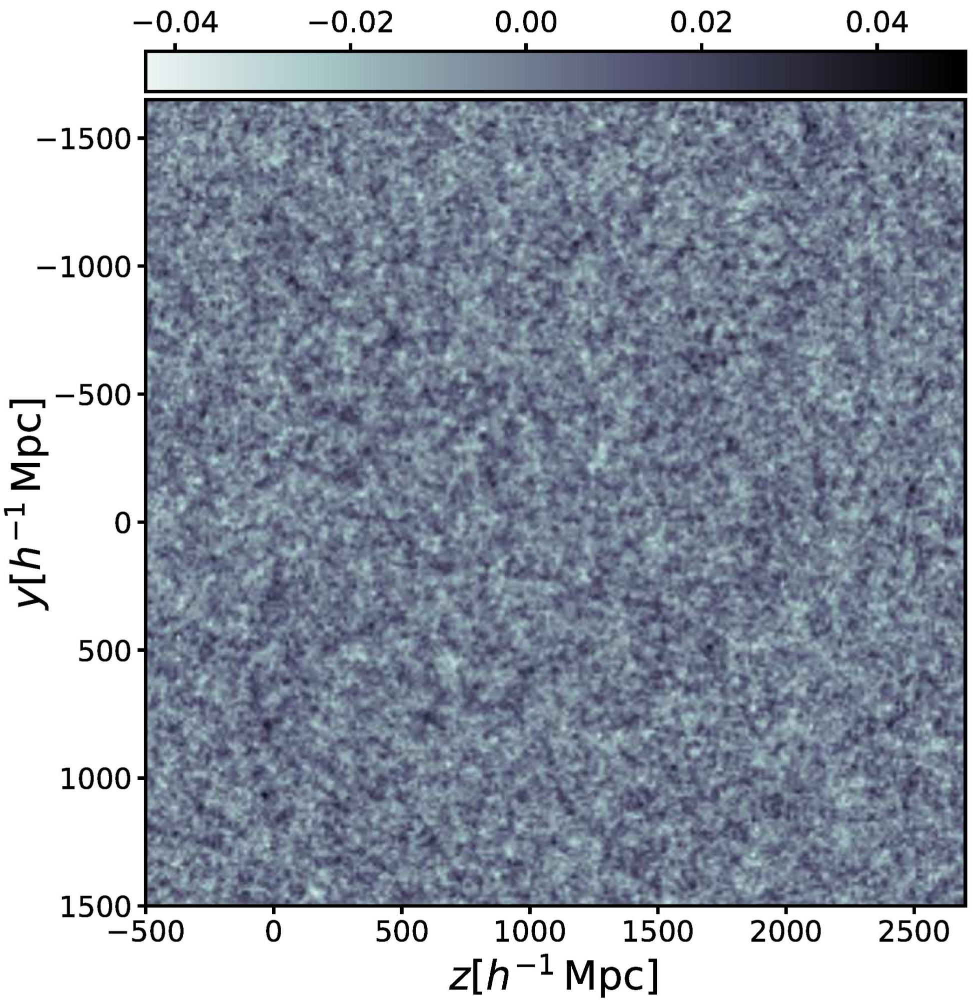
- ARGO (FSK et al 2008, 2016)
- KIGEN (FSK 2013)

What is new?

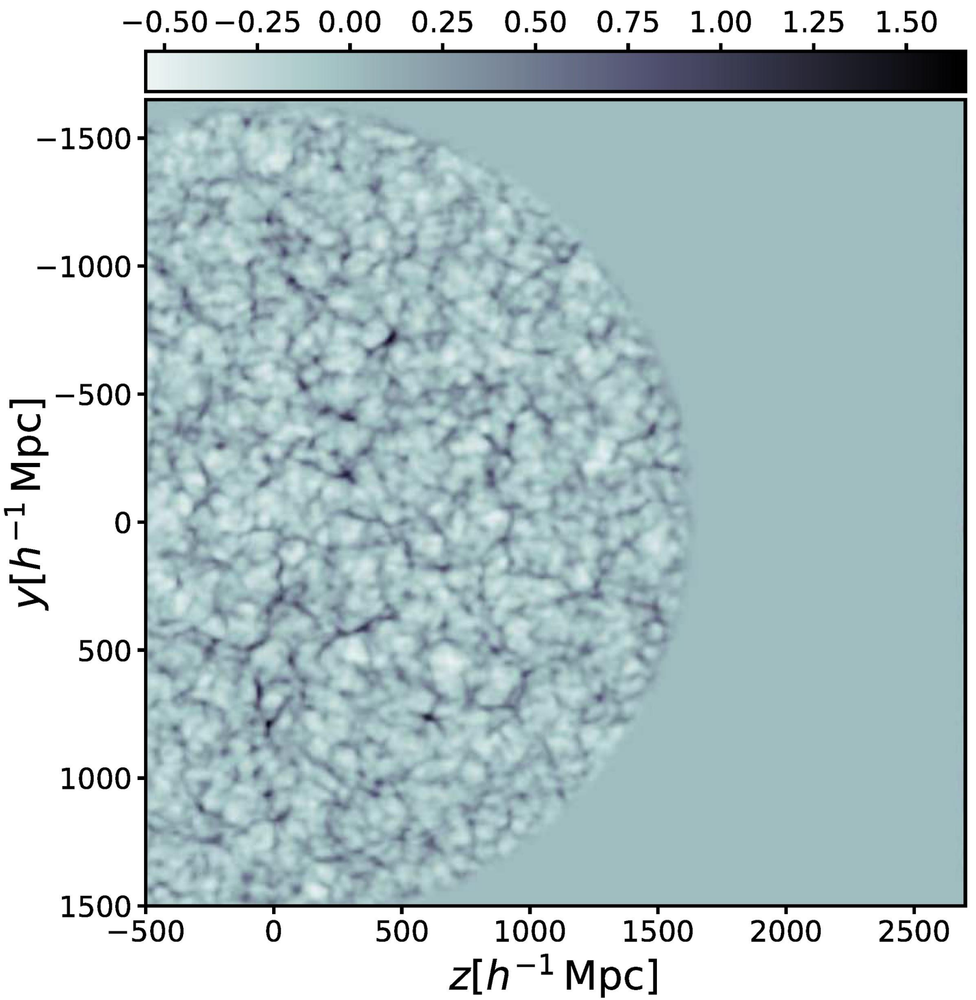
- Automatic nonlinear bias sampling
- Reconstruction of the completeness at early cosmic times (survey geometry, selection function, mixing)
- Light-cone reconstruction of bias, displacements, velocity fields, completeness

We have produced a new mock galaxy catalog fitting CMASS galaxies as a function of redshift of CMASS galaxies based on the BigMD simulation. and the dark matter lightcone to compare with the reconstruction.

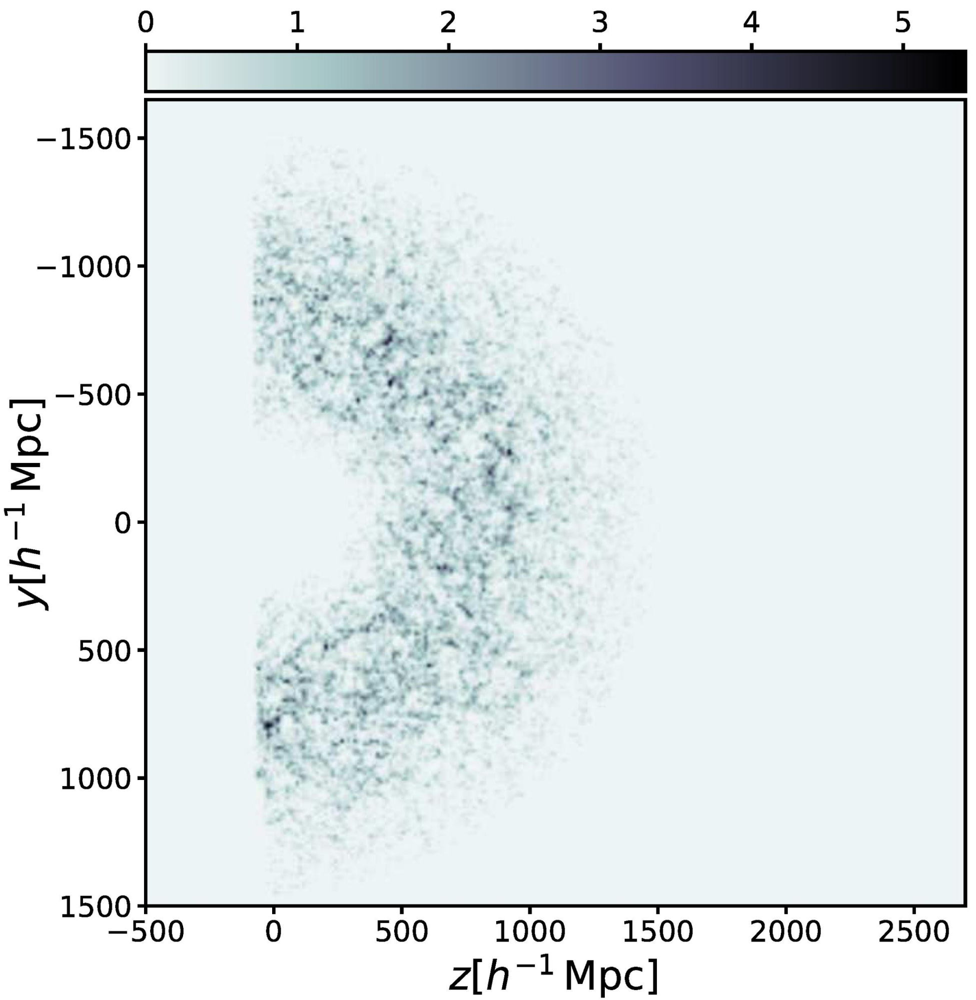
(credit Sergio Rodriguez and Gustavo Yepes)



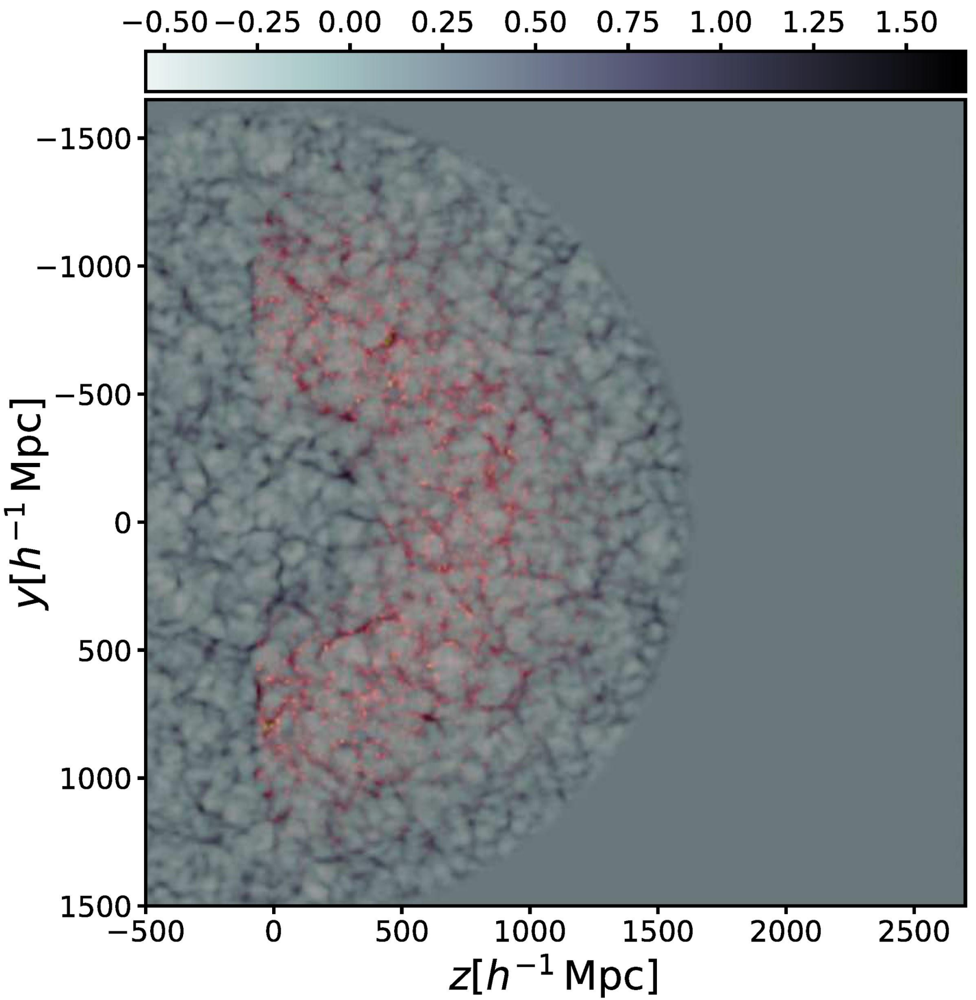
−0.04
−0.02
0.00
0.02
0.04
−1500
−1000
−500
0
500
1000
1500
$y[h^{-1}\,\mathrm{Mpc}]$
−500
0
500
1000
1500
2000
2500
$z[h^{-1}\,\mathrm{Mpc}]$



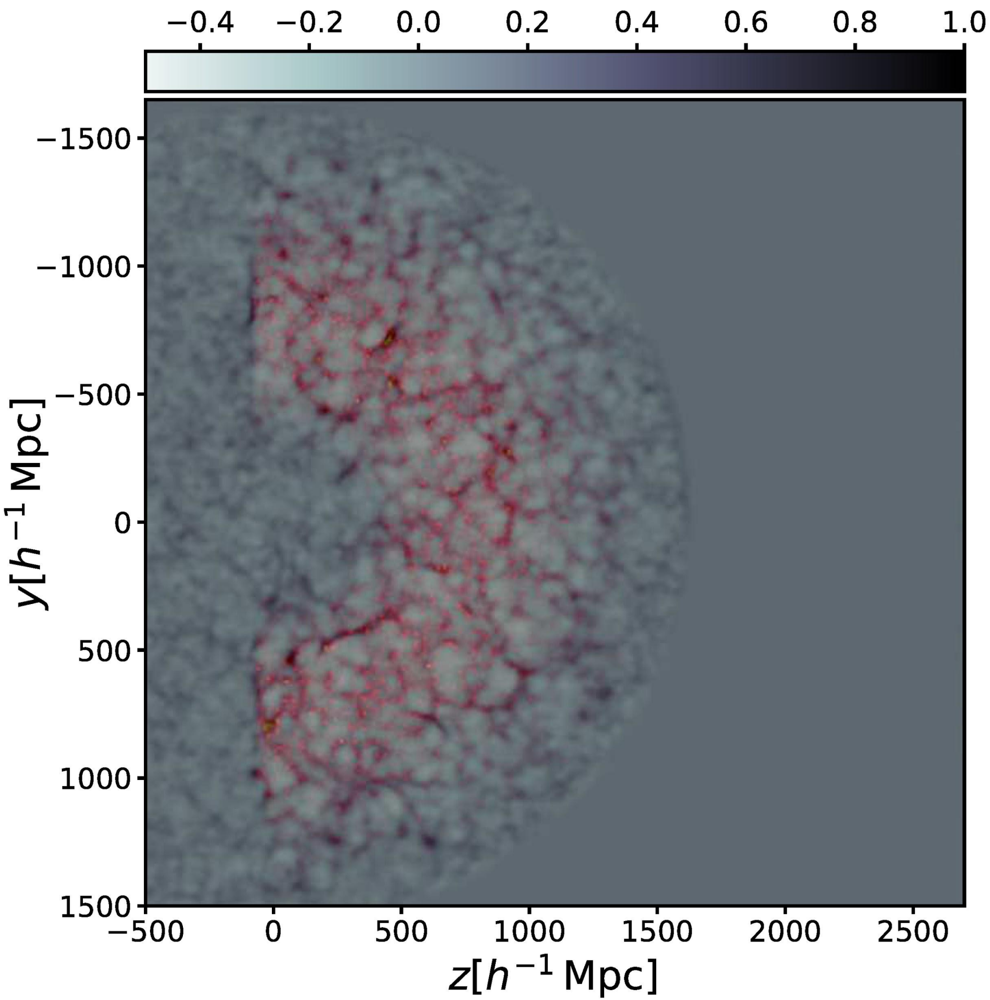
−0.50 −0.25 0.00 0.25 0.50 0.75 1.00 1.25 1.50
−1500
−1000
−500
0
500
1000
1500
$y[h^{-1}\,\mathrm{Mpc}]$
−500 0 500 1000 1500 2000 2500
$z[h^{-1}\,\mathrm{Mpc}]$

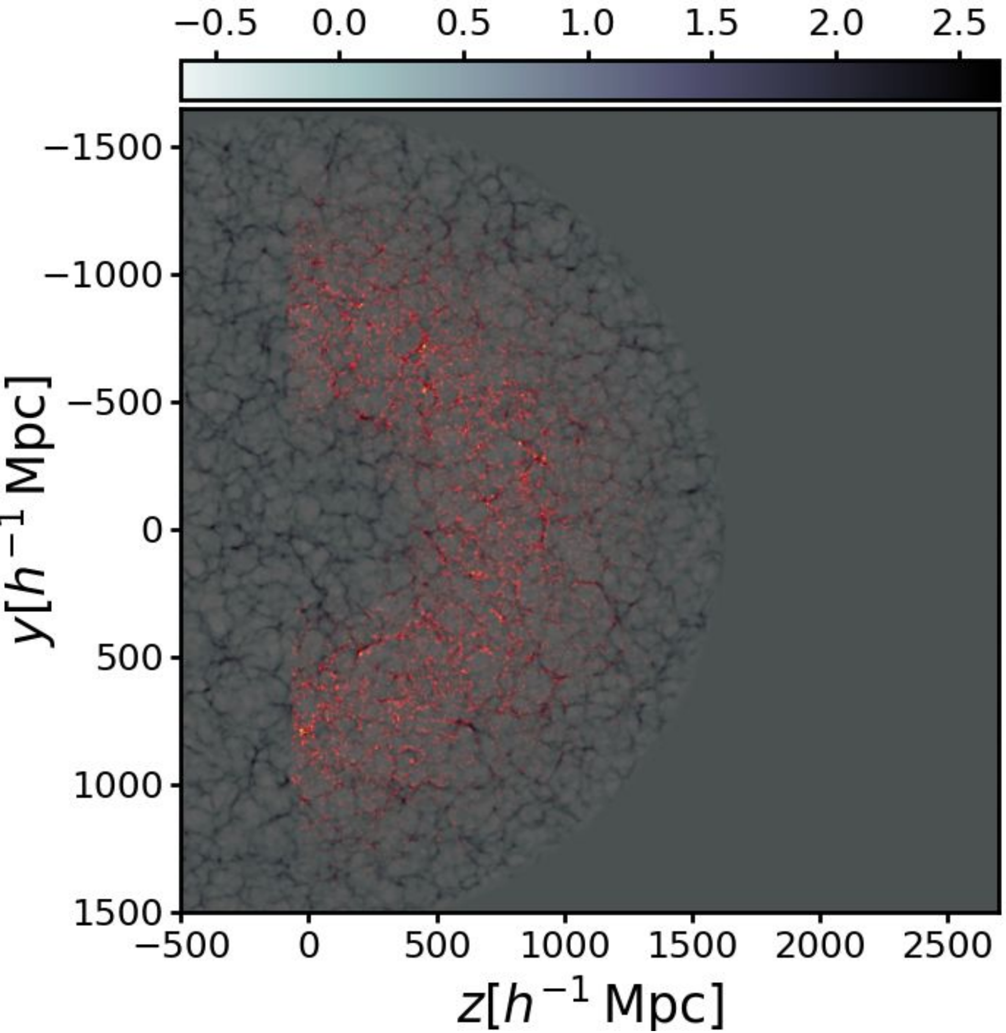


0
1
2
3
4
5
−1500
−1000
−500
0
500
1000
1500
$y[h^{-1}\,\mathrm{Mpc}]$
−500
0
500
1000
1500
2000
2500
$z[h^{-1}\,\mathrm{Mpc}]$



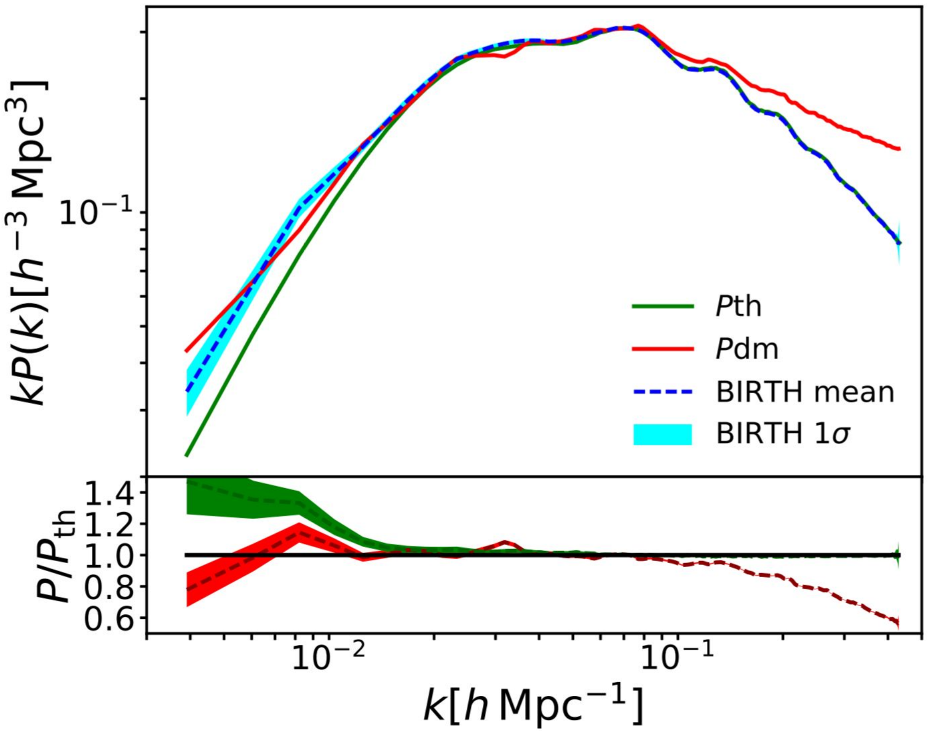
−0.50 −0.25 0.00 0.25 0.50 0.75 1.00 1.25 1.50
−1500
−1000
−500
0
500
1000
1500
−500 0 500 1000 1500 2000 2500
$y[h^{-1}\,\mathrm{Mpc}]$
$z[h^{-1}\,\mathrm{Mpc}]$


−0.4
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
−1500
−1000
−500
0
500
1000
1500
−500
0
500
1000
1500
2000
2500
$y[h^{-1}\,\mathrm{Mpc}]$
$z[h^{-1}\,\mathrm{Mpc}]$



−0.5
0.0
0.5
1.0
1.5
2.0
2.5
−1500
−1000
−500
0
500
1000
1500
$y[h^{-1}\,\mathrm{Mpc}]$
−500
0
500
1000
1500
2000
2500
$z[h^{-1}\,\mathrm{Mpc}]$

Unbiased power spectra up to the Nyquist frequency



CONCLUSIONS

- Modelling the galaxy distribution is possible with dark matter only simulations resolving the halos hosting those galaxies, but only about 1 simulation can be done for a survey
- One can learn from full calculation simulations and train algorithms to reproduce those results on coarse grids
- Effective bias models are required
- We can achieve high accuracy for the next generation of galaxy surveys in the 2-, 3-, and 4- point statistics mapping the full bias relation
- We need many mock galaxy catalogs to robustly extract cosmological information from galaxy surveys
- Bayesian forward modelling approaches help us to reconstruct the dark matter field from the galaxy distribution