

Modelling and inferring the cosmological Large Scale Structure



Leibniz Institute for Astrophysics in Potsdam

Francisco-Shu Kitaura

Karl-Schwarzschild fellow

MPA Large Scale Structure

Modelling the cosmological Large Scale Structure

How can we generate accurate mock galaxy catalogs for present large-scale structure surveys? (BOSS, eBOSS, J-PAS, DESI, EUCLID, 4MOST, WEAVE, DES, LSST, ...)

4MOST: 4m VISTA 2021-2025
spectroscopic survey
~14k sq deg
15-20 M LRGs+eLGs+Quasars
 $z < 1.4$



JPAS: JST/T250 2016-2020
56 colors
1.2 Giga pixel camera
~8k sq deg
~200 M objects
 $z < 1.4$

Perform large N-body simulations:
MillenniumXXL, BigMultidark, MICE, DEUS, HORIZON
needed to produce reference catalogs.

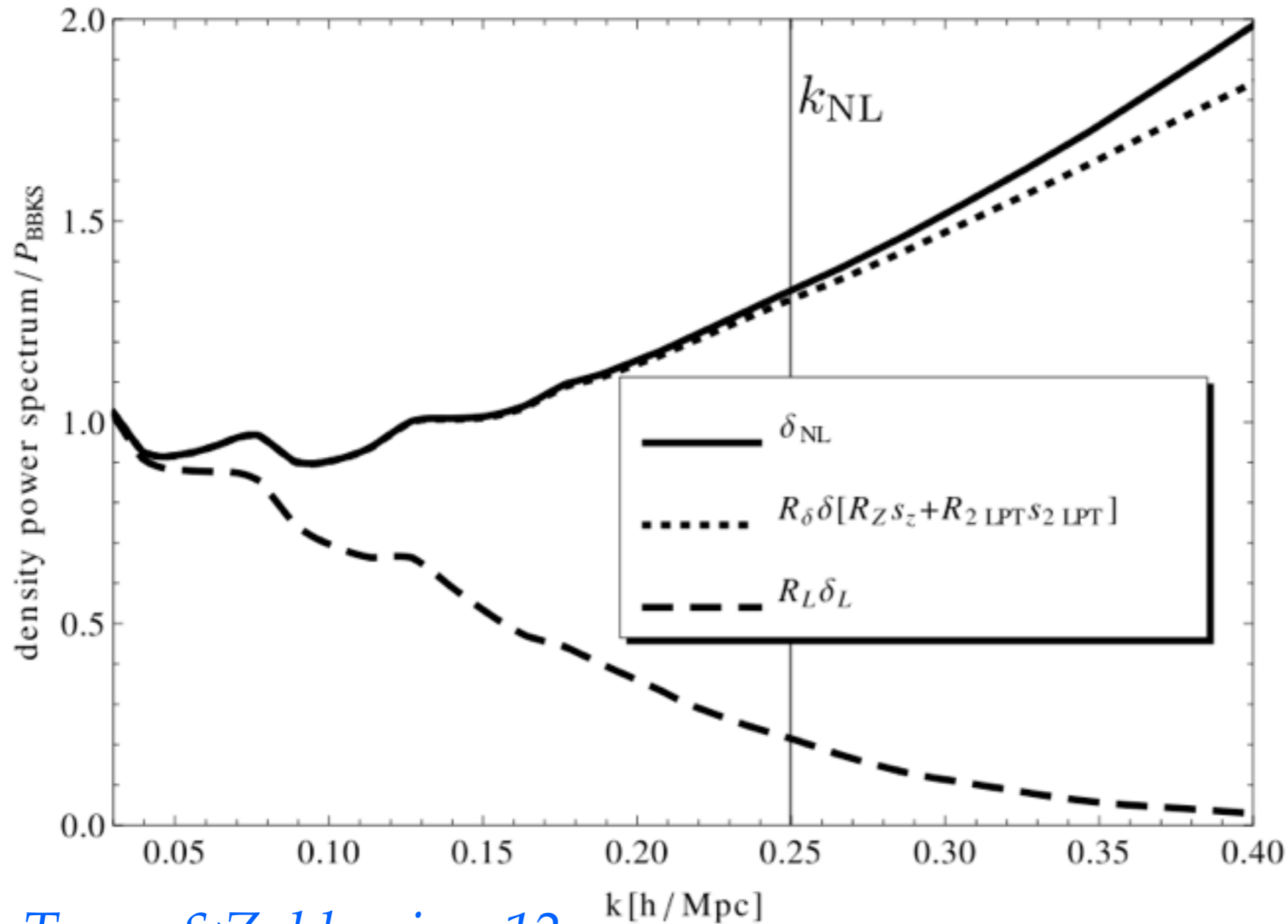
However,

**this effort can only be done a couple of times,
but not thousands of times!**

What is the bottle-neck of the computations?

1. the gravity solver

Perturbative approaches to model BAOs



$$\delta(\mathbf{k}) = R_{\delta}(k) \delta_1(\mathbf{k})$$

$$R_{\delta}(k, \eta) = \exp(0.58d)$$

missing nonlinear power bump
no one-halo term!

Tassev & Zaldarriaga 12

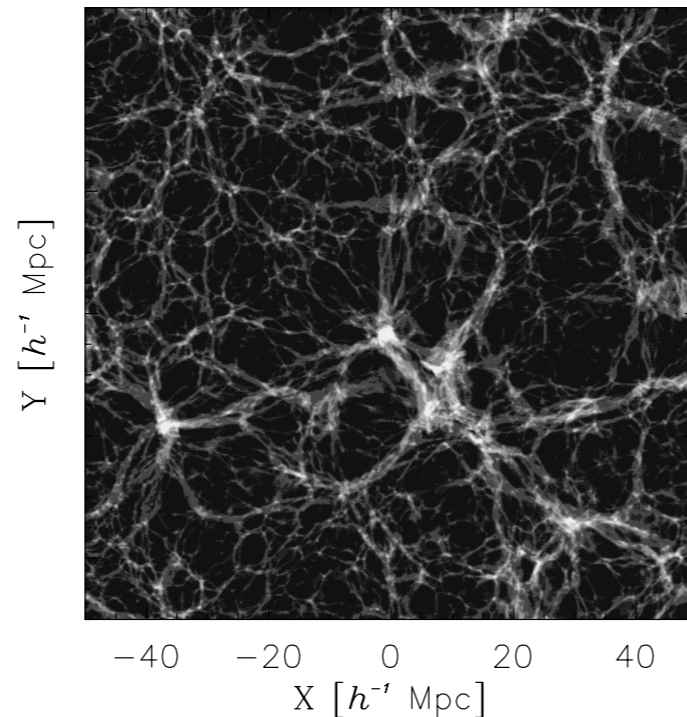
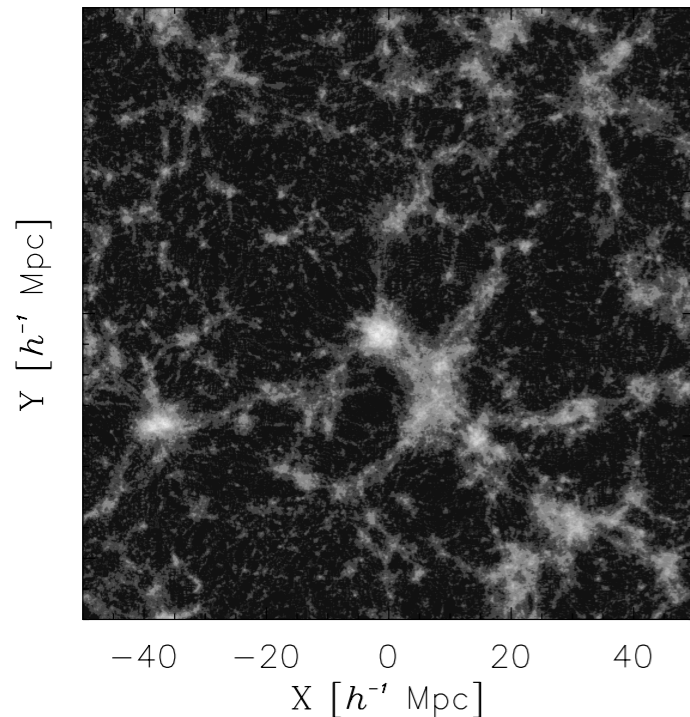
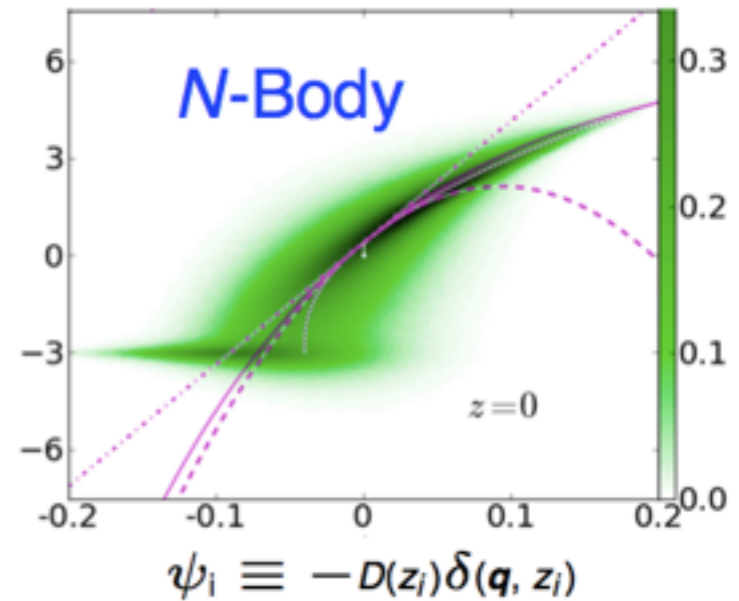
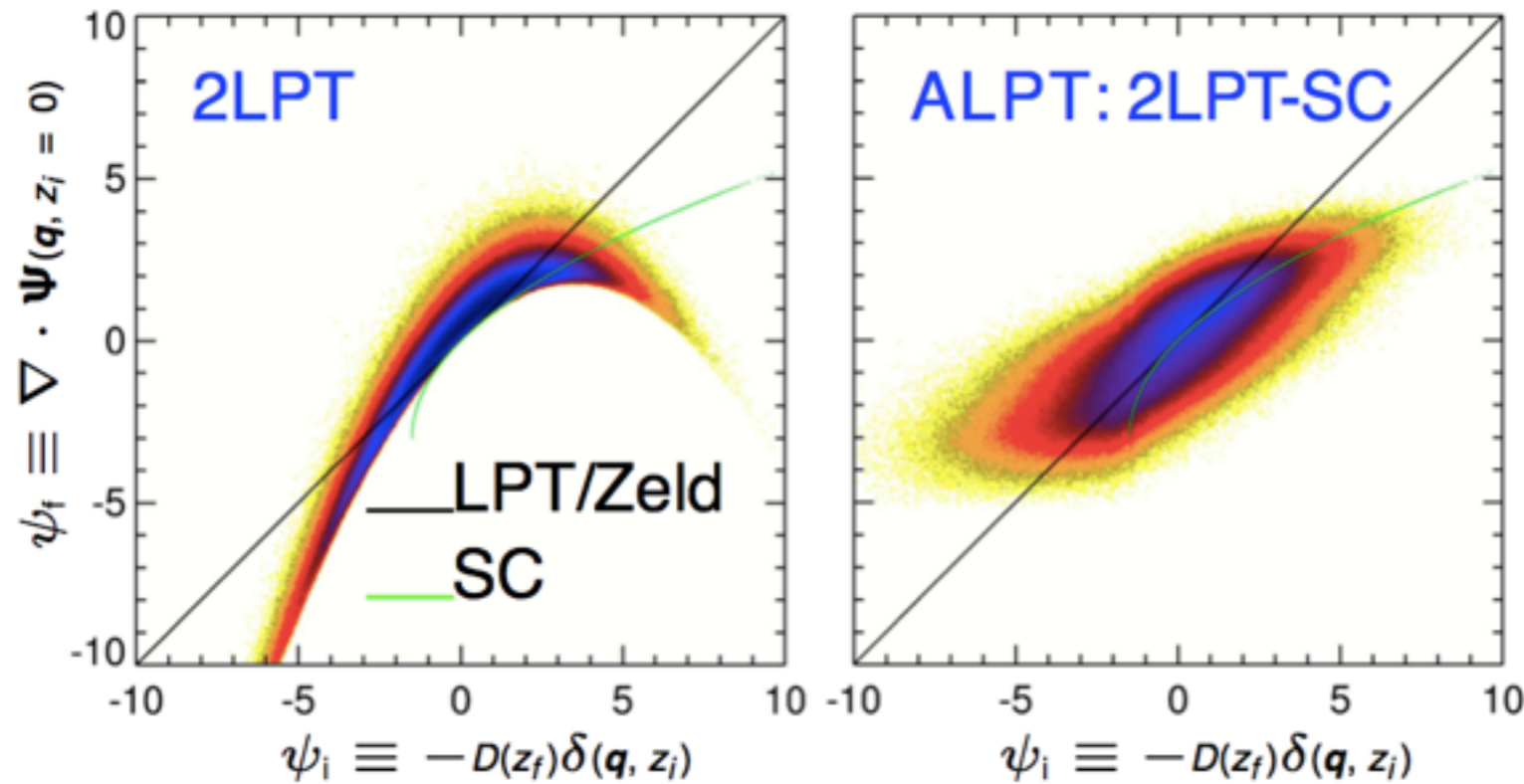
pioneering galaxy mock generation codes

Scoccimarro & Sheth 02 (PThalos)

Monaco + 02 (Pinocchio)

ALPT: Augmented Lagrangian Perturbation Theory

FSK & Heß 13, arxiv:1212.3514



*for N-body relation see Neyrinck13
 for the spherical collapse relation see also
 Bernardeau94; Mohayee06
 improvement for $k > 1$ found in Neyrinck15
 speed up N-body codes with LPT:
 COLA Tassev+13*

**ALPT is a fast one-step solver
 75% more correlated with the full N-body solution than 2LPT at $k \sim 2$ h/Mpc**

down to which scales can we trust these PT based approaches?

What is the bottle-neck of the computations?

2. the resolution

Let us ask:
What is the relation between
the halo/galaxy distribution
and
the underlying large-scale dark matter field
(on a few Mpc scales)?

Deterministic biasing

we need to know the deterministic bias!

$$\rho_h = f_h B(\rho_h | \rho_M)$$

but this implies knowing all the higher order correlation functions!

$$B(\rho_h | \rho_M) = B(\rho_h | \rho_M, \xi_1^h, \xi_2^h, \xi_3^h, \xi_4^h, \dots)$$

Deterministic biasing parametrization

Fry&Gaztañaga93

$$\rho_h = f_h^a \sum_i a_i \delta_M^i$$

Cen&Ostriker93; de la Torre&Peacock13

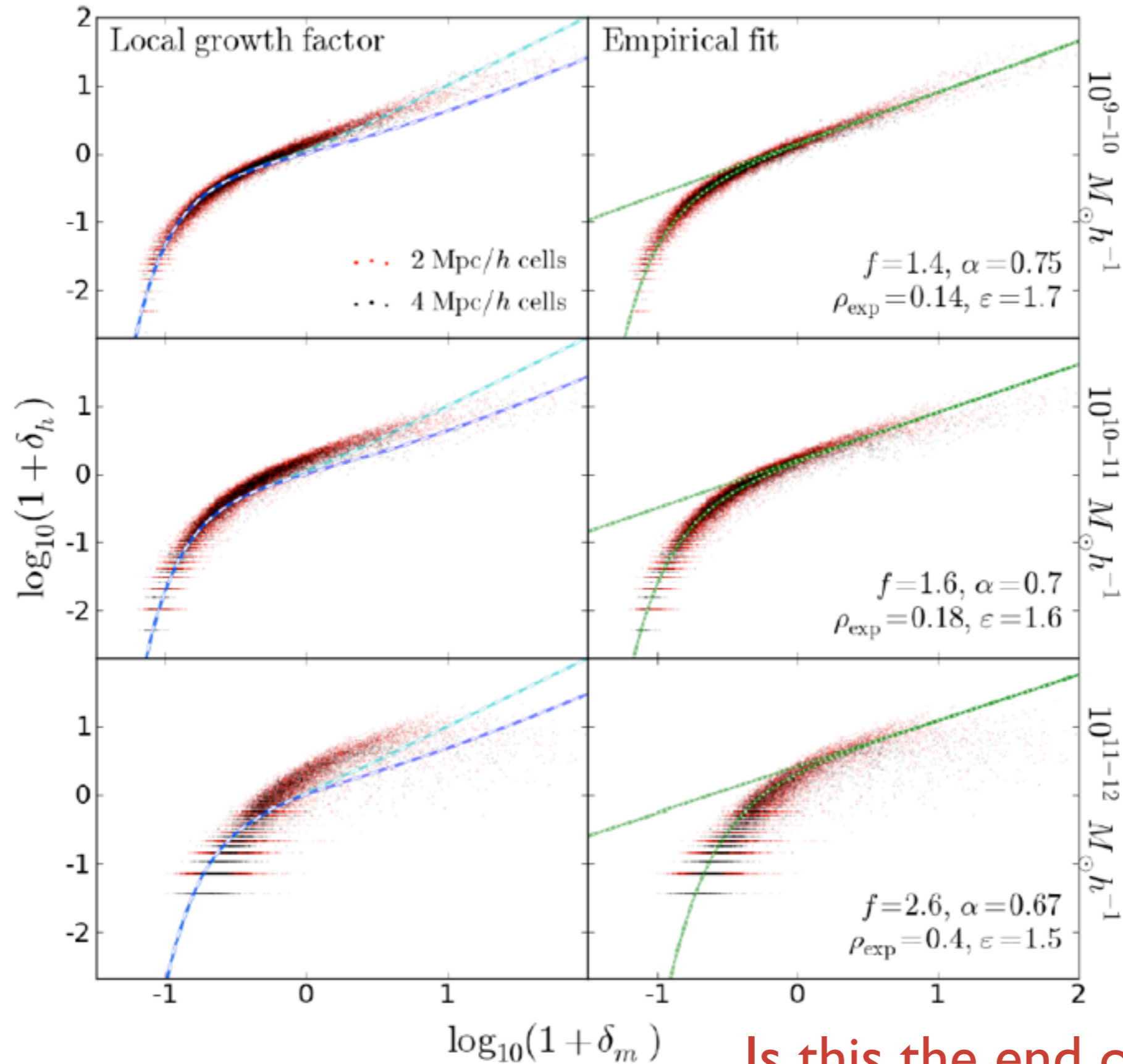
$$\rho_h = f_h^b \exp \left[\sum_i b_i \log (1 + \delta_M)^i \right]$$

FSK, Yepes & Prada14 + Neyrinck+14

$$\rho_h = f_h \theta(\rho_M - \rho_{\text{th}}) \rho_M^\alpha \exp \left[- \left(\frac{\rho_M}{\rho_\epsilon} \right)^\epsilon \right]$$

Neyrinck+13; Aragon-Calvo13

These relations are well known in the literature see eg Mo&White02



Is this the end of the story?

Stochastic biasing

caution! we still need to know the deviation from Poissonity!

$$N_h \curvearrowright \mathcal{P}(N_h | \rho_h)$$

$$N_h \curvearrowright \mathcal{P}(N_h | B(\rho_h | \rho_M, \xi_1^h, \xi_2^h, \xi_3^h, \xi_4^h, \dots))$$

non-Poissonian PDFs:

Saslaw&Hamilton84

Sheth95

stochastic bias:

Dekel&Lahav99

Sheth&Lemson99

Somerville+01

Casas-Miranda+02

PATCHY-code

Over-dispersion 10% effect at BAO scales in P_k
for LRGs, stronger effect 20-30% for eLGs

FSK, Yepes & Prada 14 arxiv:1307.3285

Finding the deterministic and stochastic parameters

FSK+14 arxiv:1407.1236

$$\bar{N}_h = \langle \rho_h \rangle \leftarrow \xi_1^h$$

$$P_h(k) \leftarrow \xi_2^h$$

$$\mathcal{P}^1(\rho) = \int_{-\sqrt{-1}\infty}^{\sqrt{-1}\infty} \frac{dt}{2\pi\sqrt{-1}} \exp(t\rho + \mathcal{C}(t))$$

$$\mathcal{P}_h^1(B(\rho_h | \rho_M)) \leftarrow \{\xi_1^h, \xi_2^h, \xi_3^h, \xi_4^h, \dots\}$$

$$N_h \curvearrowright \mathcal{P}(N_h | B(\rho_h | \rho_M, \bar{N}_h, P_h(k), \mathcal{P}_h^1))$$

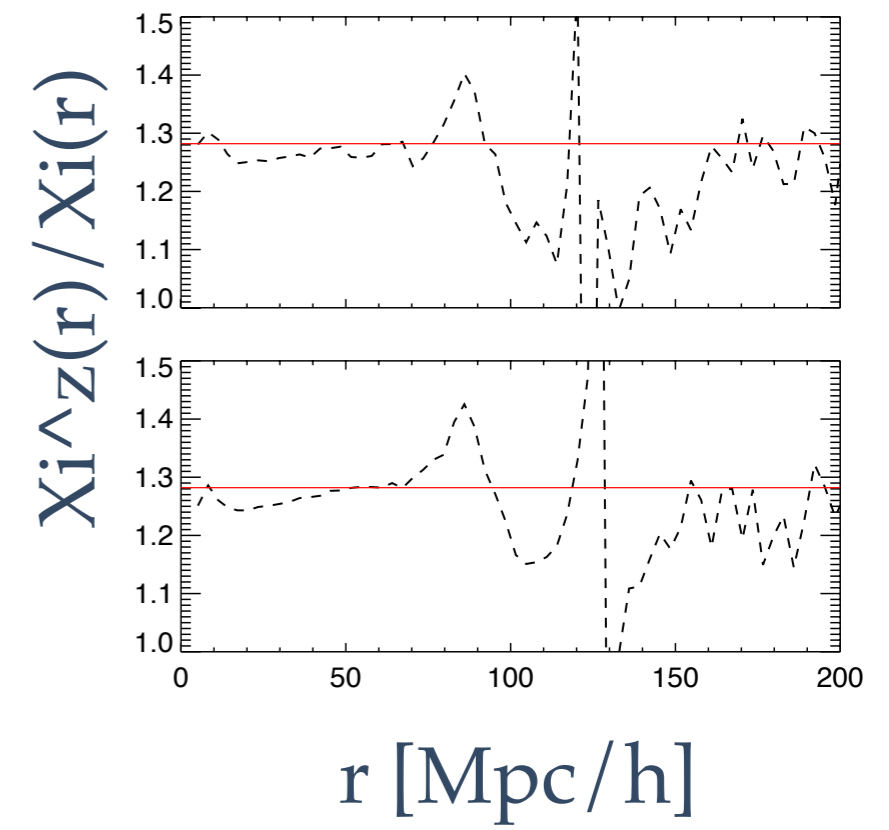
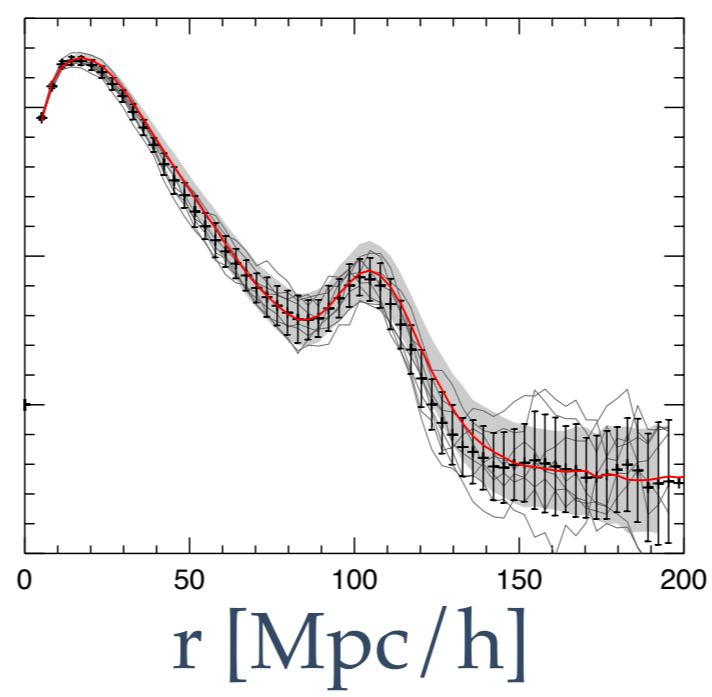
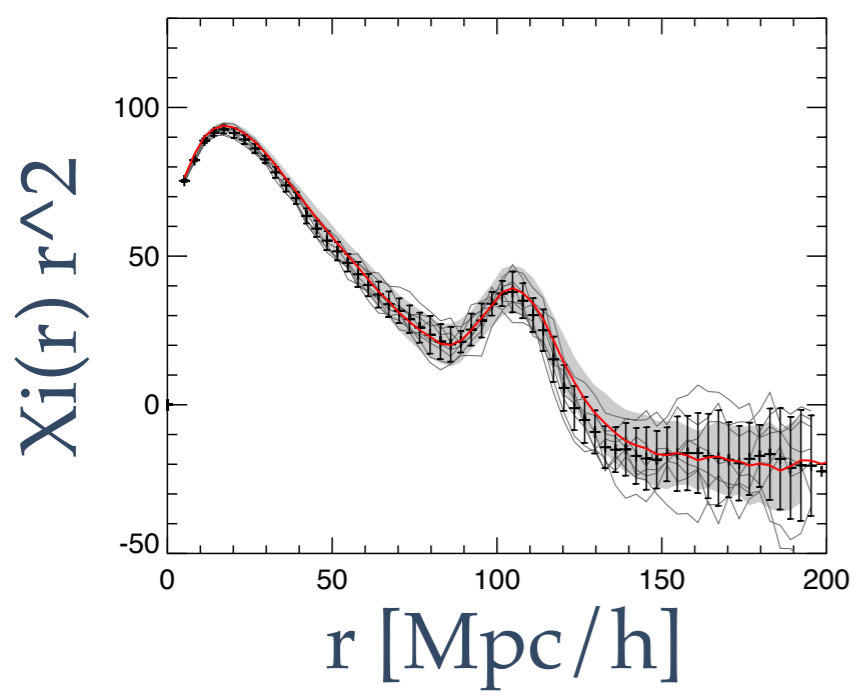
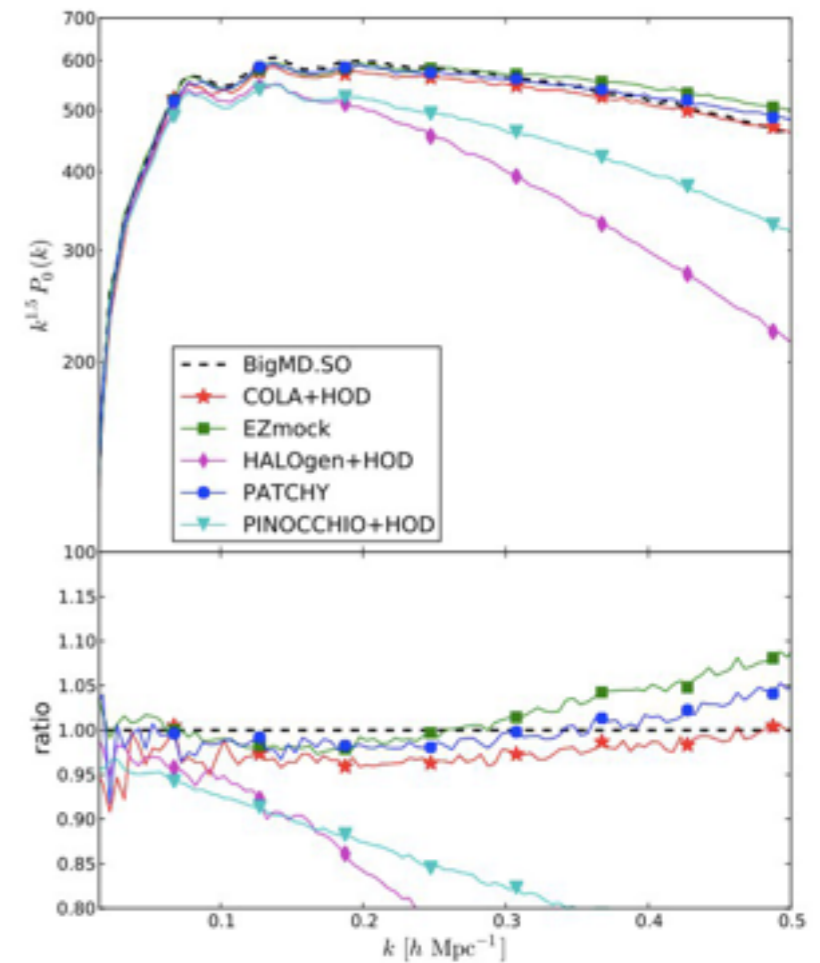
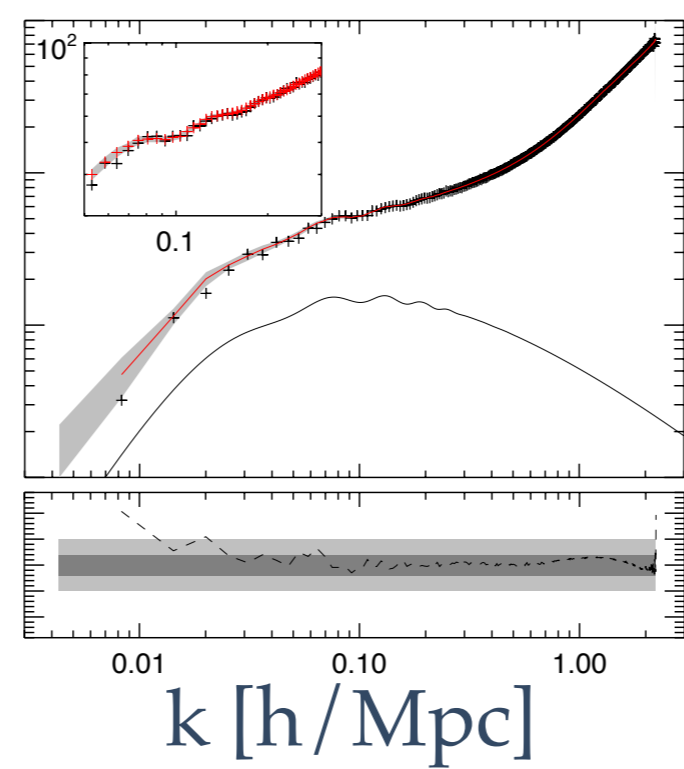
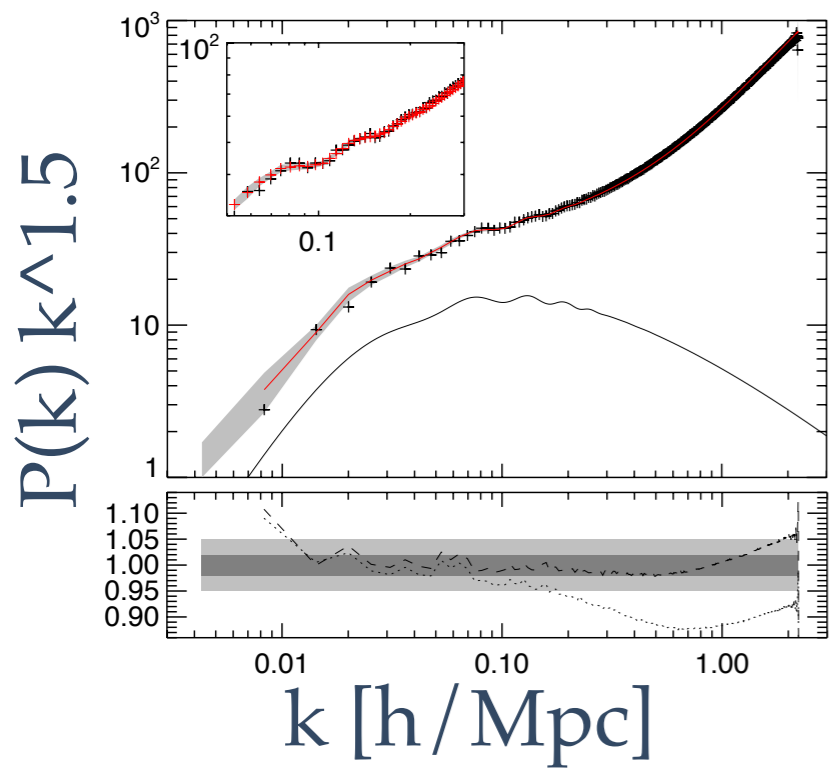
PATCHY-CODE

1pt + 2pt statistics

FK, Yepes & Prada 14, arxiv:1307.3285

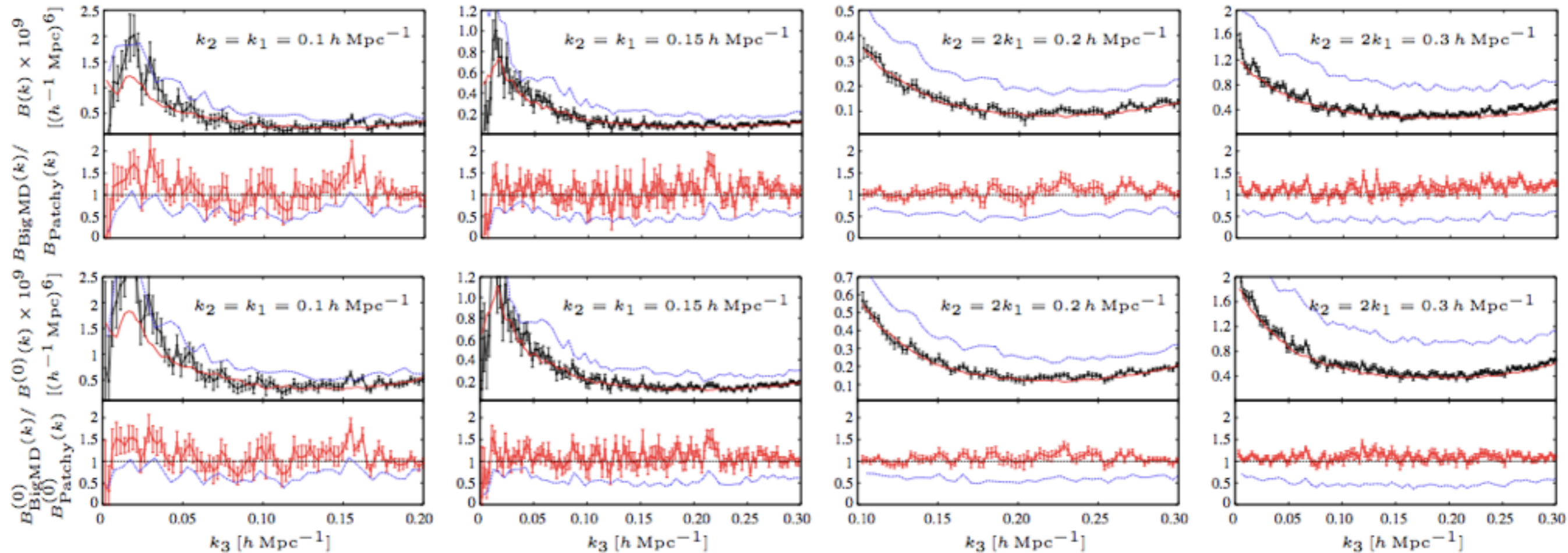
real-space

redshift-space



PATCHY CODE: 3pt statistics bispectrum

FSK, Hector Gil-Marín+14 arxiv:1407.1236



threshold bias is an indispensable ingredient!

FSK+14 arxiv:1407.1236

Kaiser84

Is this the end of the story?

Nonlocal biasing (as part of deterministic biasing)

What is the impact of non-local bias?

see eg McDonald&Roy09

Sherwin&Zadarriga12

Chan,Sheth&Schoccimarro12

Sheth,Chan&Schoccimarro13

Baldauf+12,13;

Saito+14

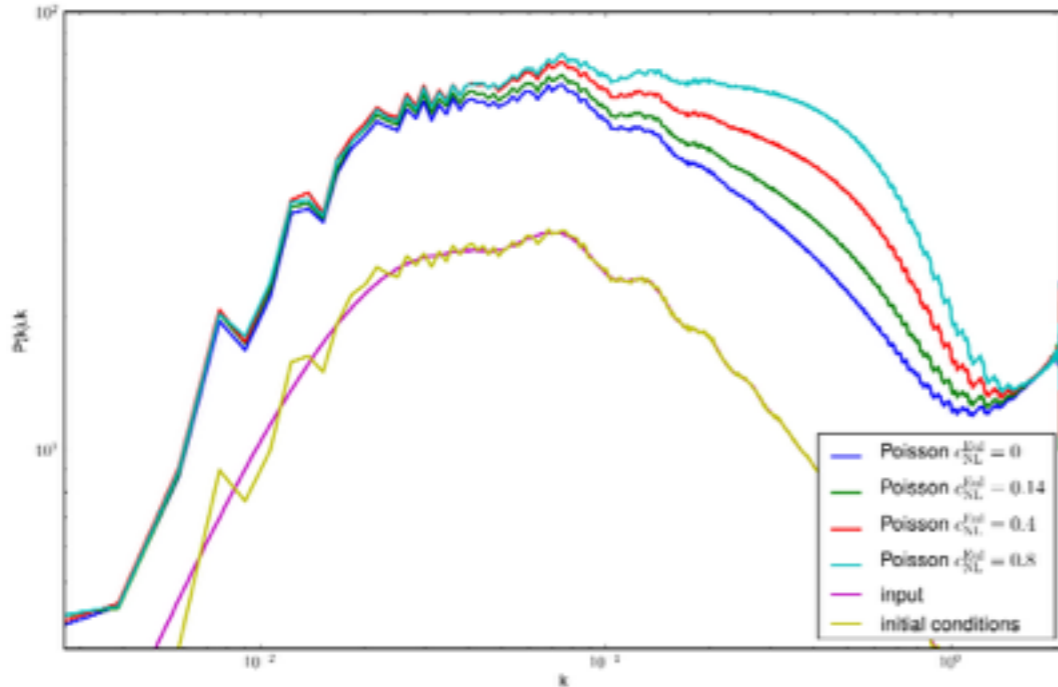
Can it be effectively modelled as part of the
stochastic bias?

Nonlocal biasing

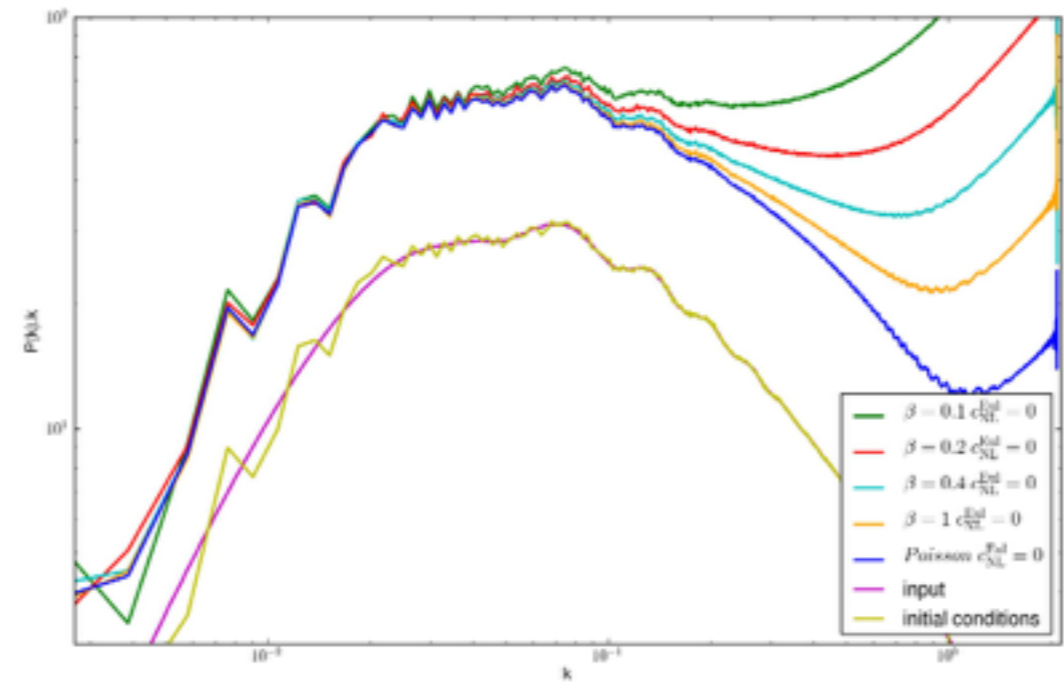
We include in PATCHY second order nonlocal bias:

$$B(\rho_h|\rho_M) = B_L(\rho_h|\rho_M) + B_{NL}(\rho_h|\rho_M)$$

Nonlocal bias



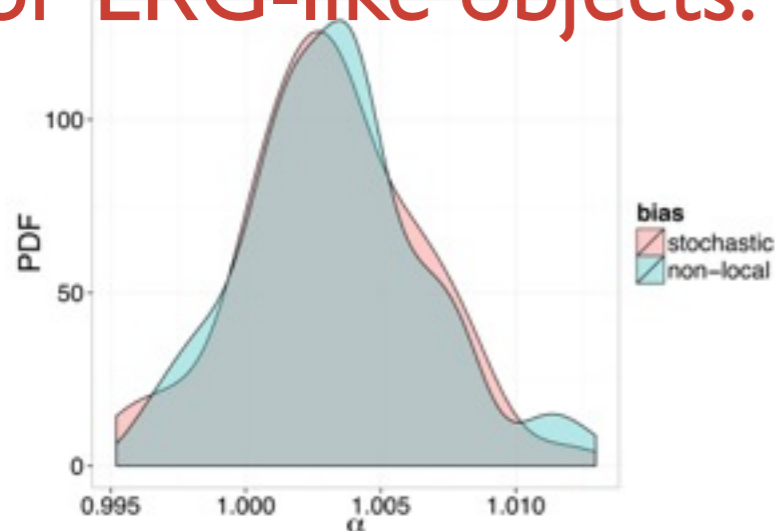
Stochastic bias



nonlocal bias is partially degenerate with stochastic bias!

not very relevant

for LRG-like objects:



Mathieu Autefage, FSK, Christian Wagner & Raul Angulo in prep

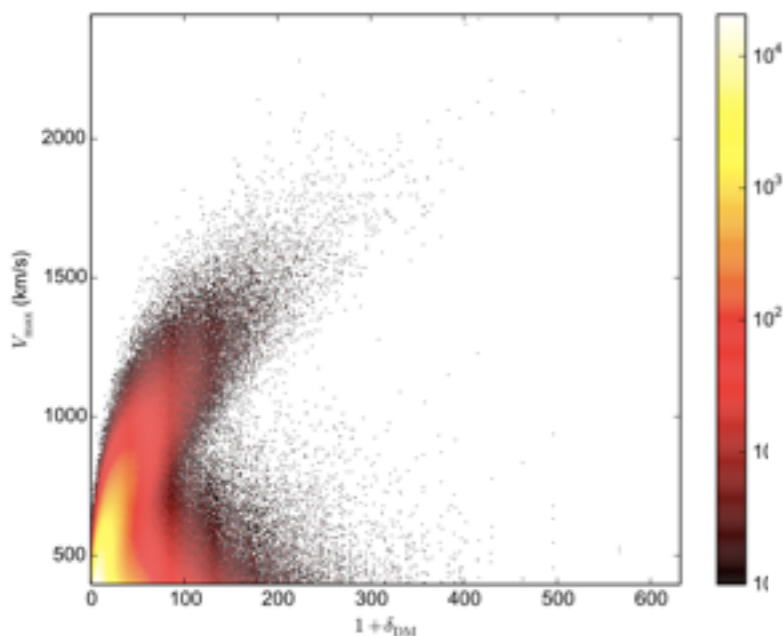
*Mathieu Autefage's master thesis
see his poster!*

Is this the end of the story?

How can we assign masses to the objects?

Cheng Zhao's master thesis

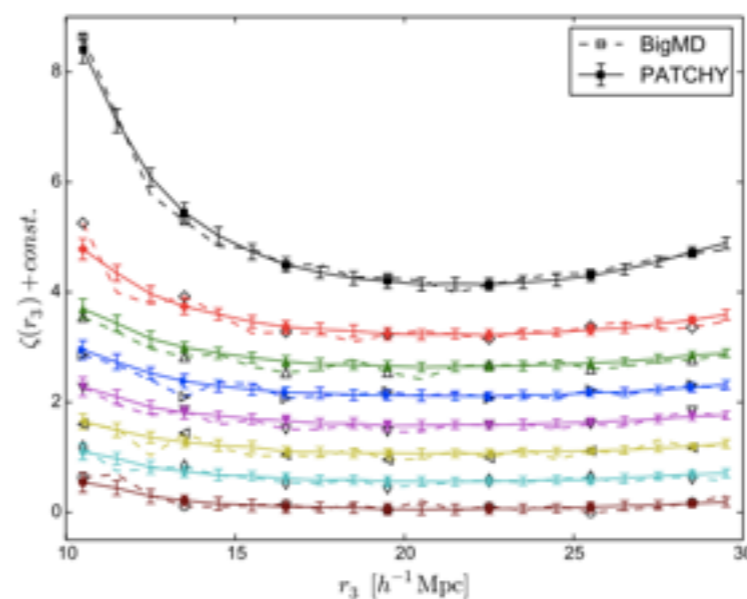
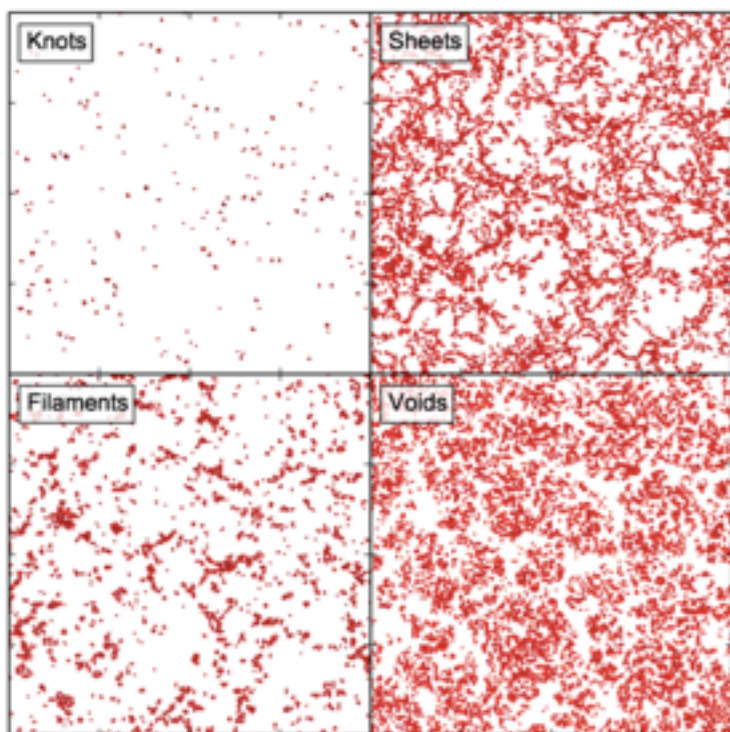
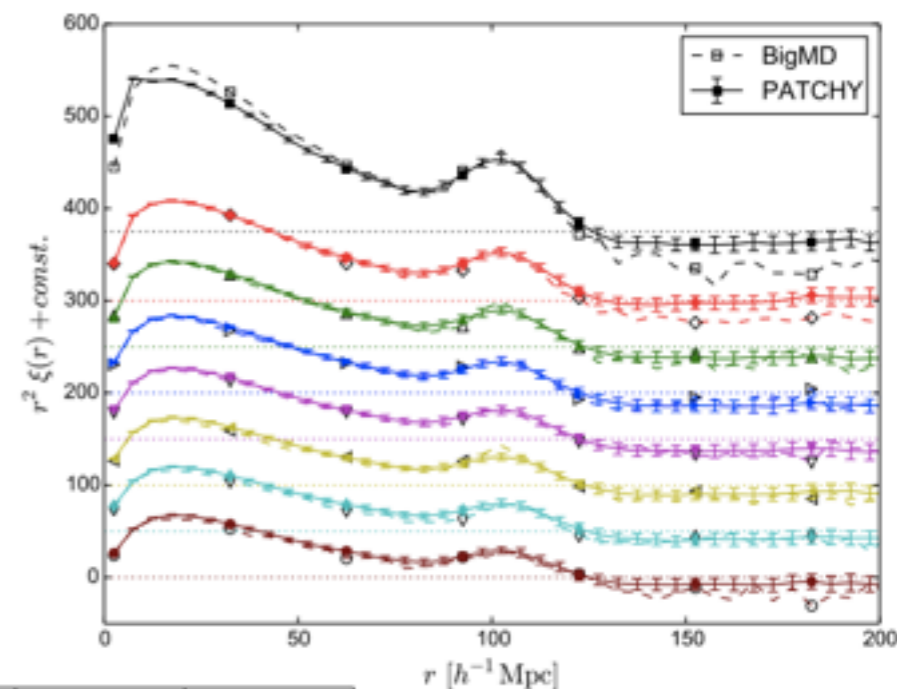
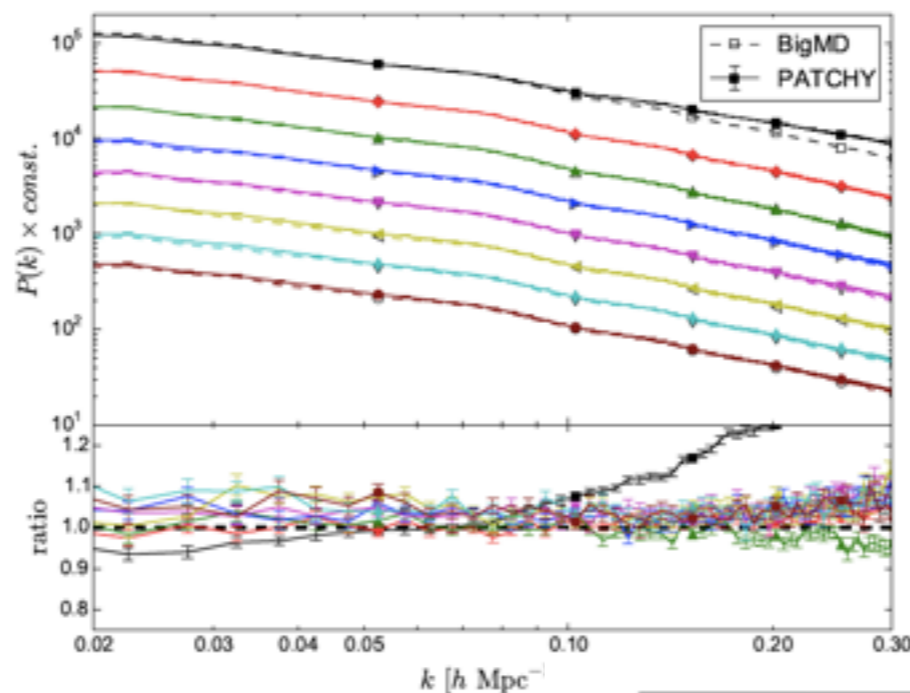
mass (V_{\max}) - density relation



HADRON code: Cheng Zhao, FSK+15 arXiv:1501.05520

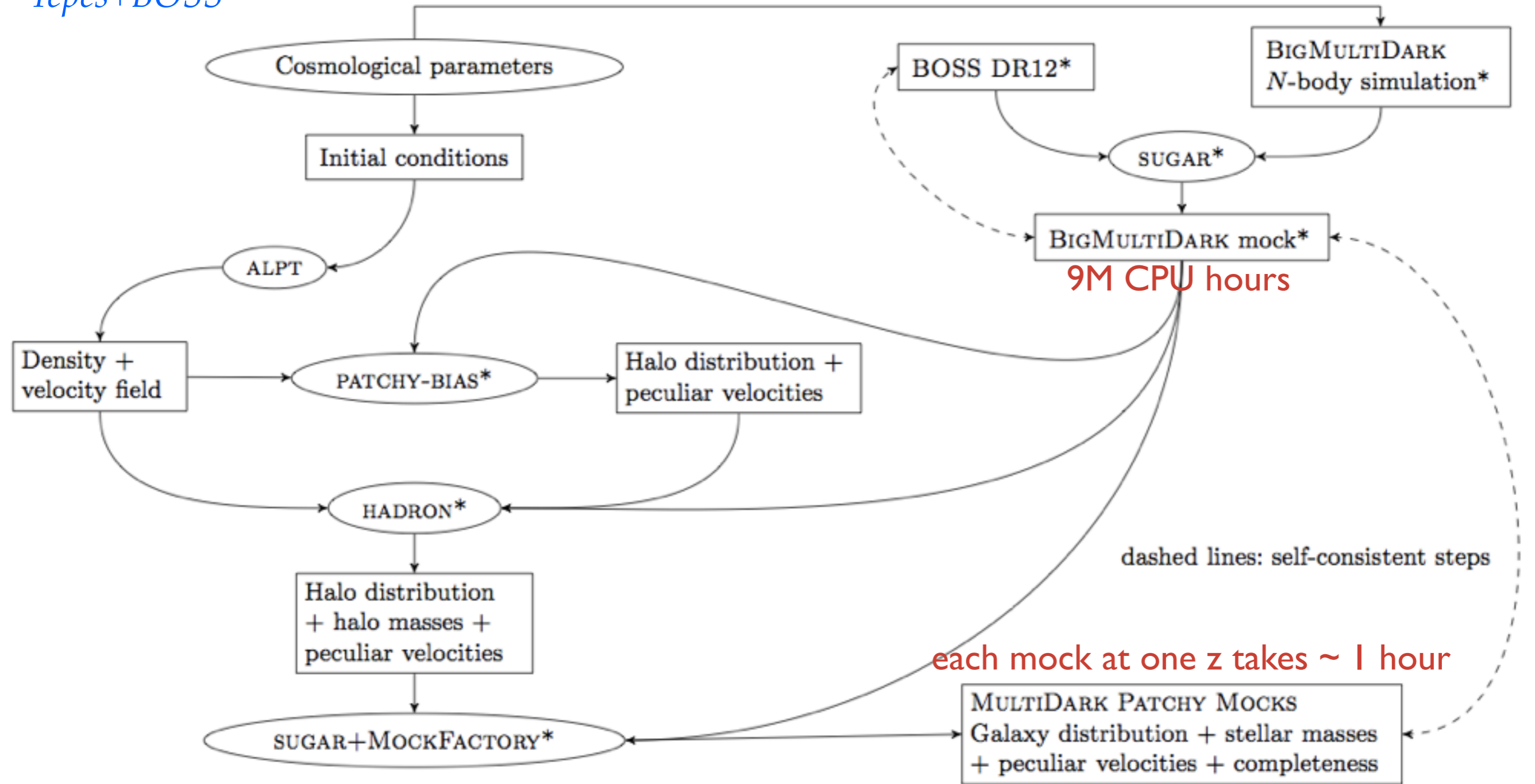
dependence on mass function, local density and cosmic web environment

$$P(M_h^i | \{r_h\}, \rho_M, T, \Delta r_{\min}^M, \{p_c\}, z)$$



PATCHY MULTIDARK BOSS DR11/DR12

collaborators Chia-Hsun Chuang, Sergio Rodriguez-Torres, Cheng Zhao, Francisco Prada, Gustavo Yepes+BOSS

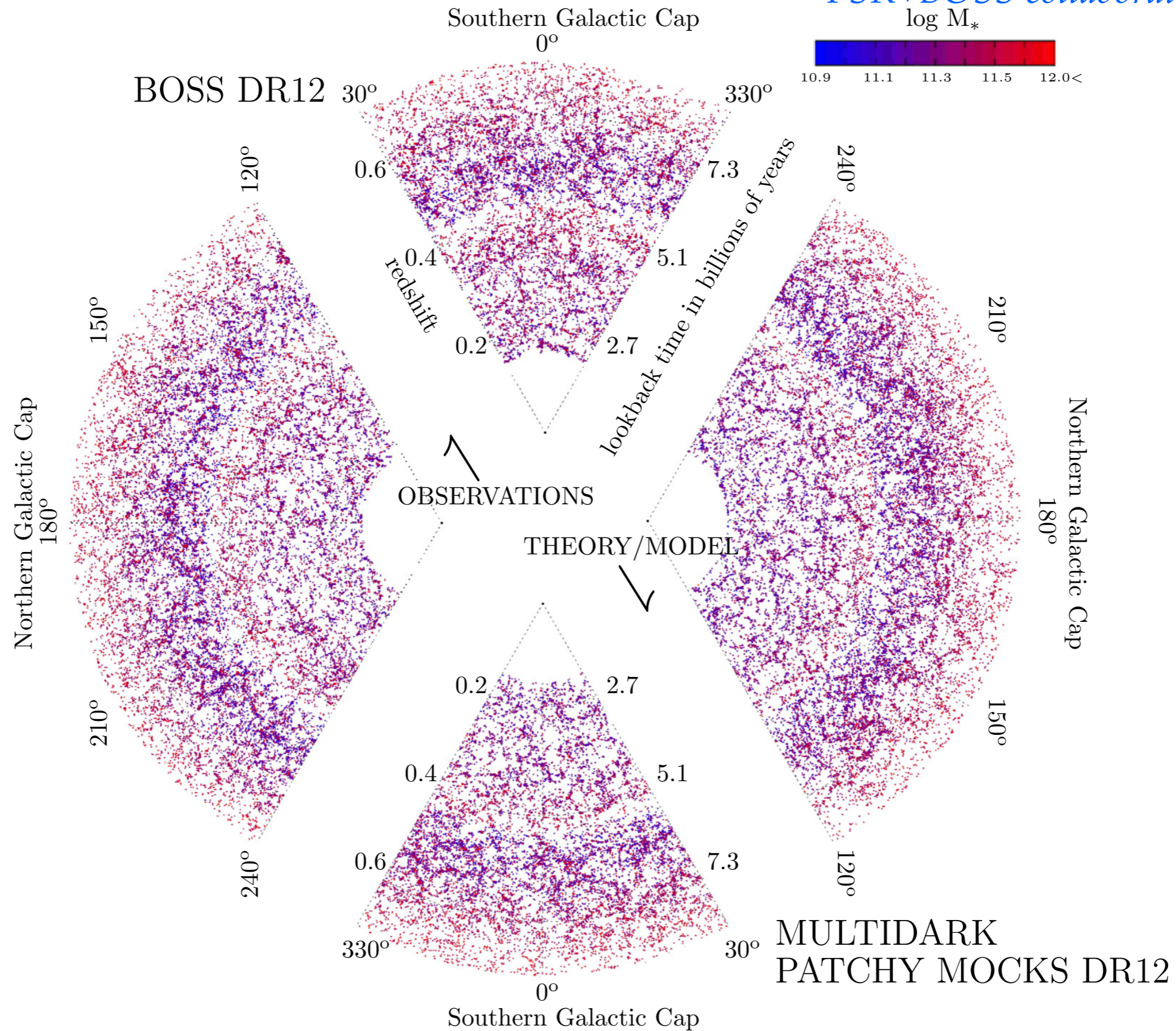


*: the asterisc indicates the steps in which calibration with observations and simulations is required

FSK+BOSS collaboration in prep

BigMultidark reference catalog Rodriguez-Torres+BOSS in prep

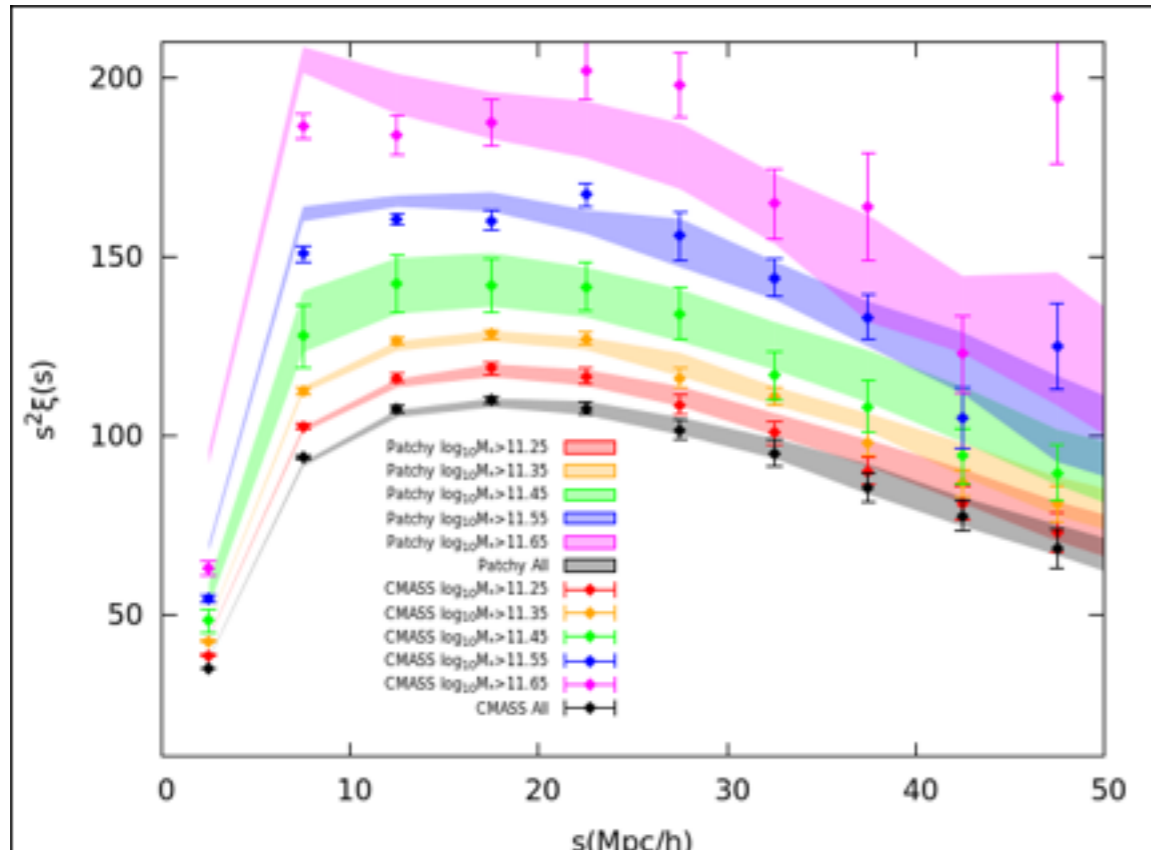
More than 12000 mocks including light cone and evolution effects
(>4000 LOWZ, >4000 CMASS, >4000 COMBINED SAMPLE) !



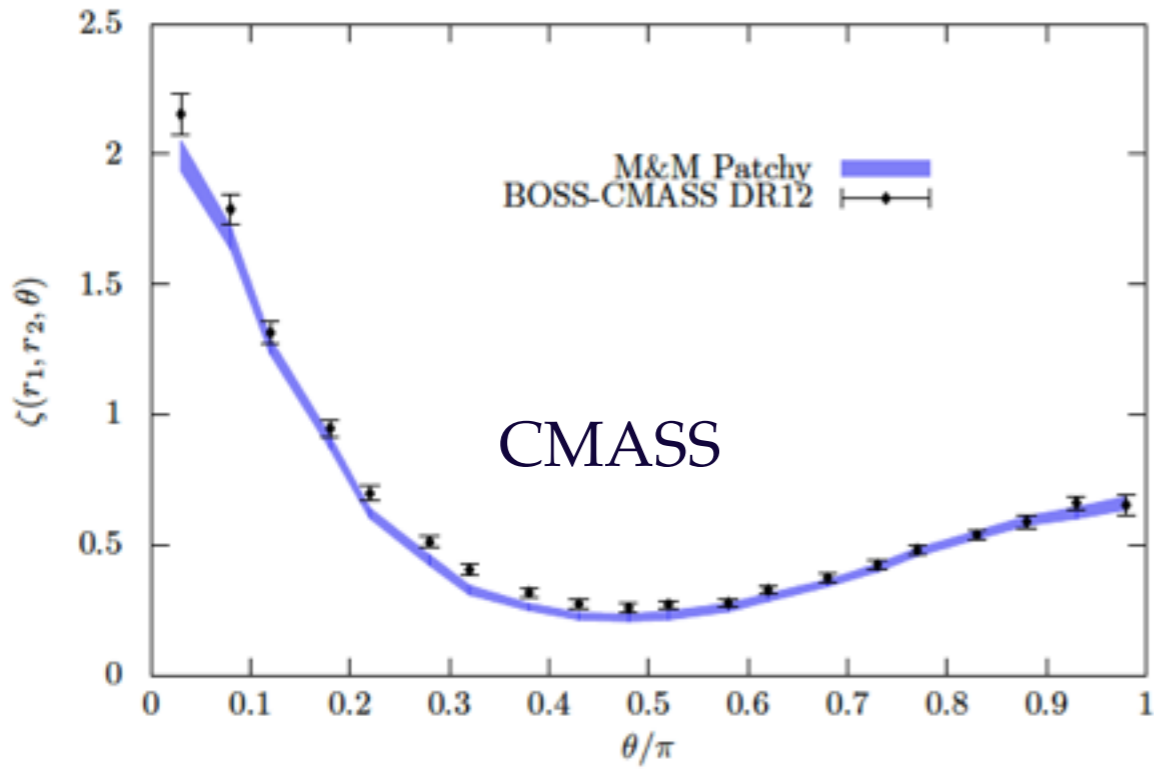
At first glance: mocks are indistinguishable from observations!

Compatible 2pt and 3pt statistics for arbitrary stellar mass cuts!

CMASS

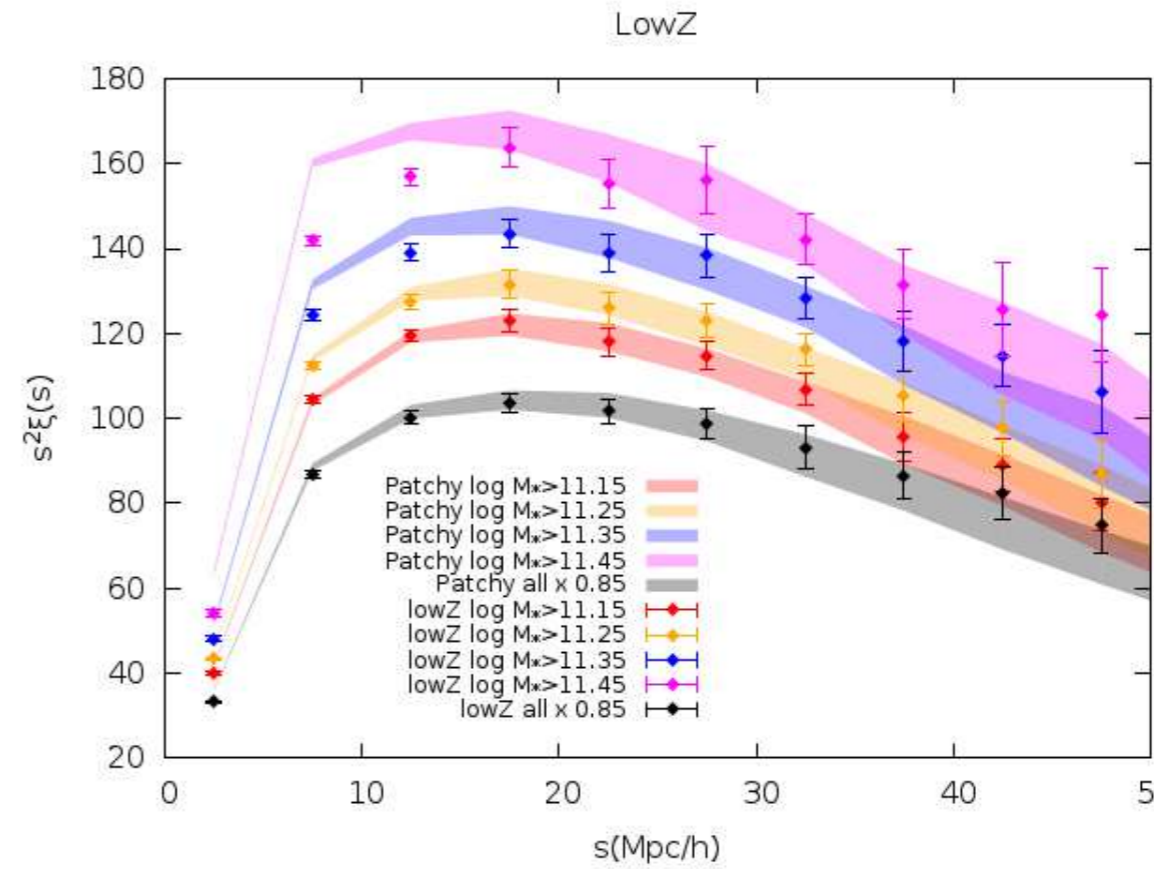


$0.43 < z < 0.6$

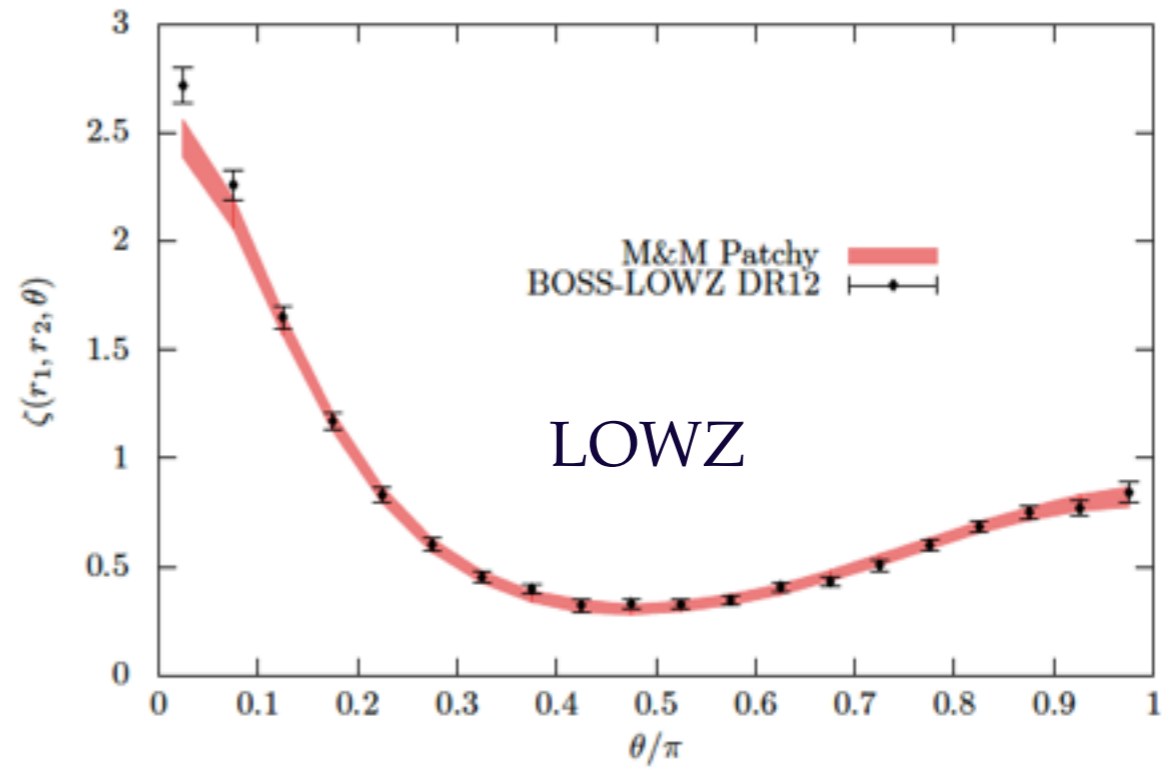


CMASS

LOWZ



$0.15 < z < 0.43$



LOWZ

Can we use our bias (and structure formation) knowledge
to make direct comparison between our models and
observations?

Inference of the cosmological Large Scale Structure

Let us reconstruct the particular realisation of our Universe!

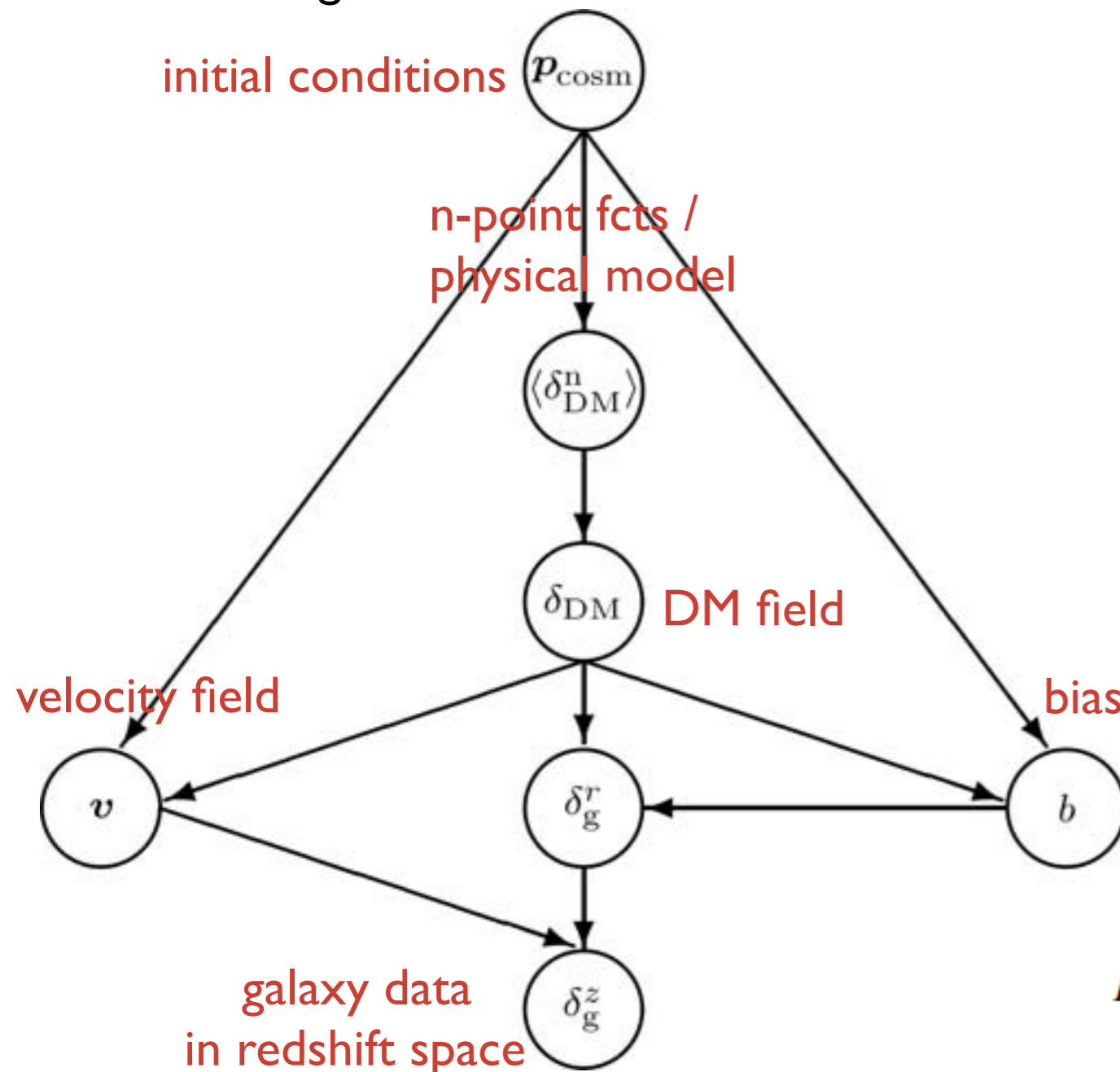
Bayesian reconstruction of the cosmological large-scale structure

F. S. Kitaura^{*} and T. A. Enßlin

[arXiv:0705.0429](https://arxiv.org/abs/0705.0429)

Max-Planck Institut für Astrophysik, D-85748 Garching, Germany

Abstract [...] These fast inverse methods ultimately will enable the application of sampling techniques to **explore complex joint posterior distributions**. We outline how the space of the **dark matter density field, the peculiar velocity field and the power spectrum** can jointly be investigated by a Gibbs-sampling process. Such a method can be applied for the **redshift distortions correction** of the observed galaxies and for time-reversal **reconstructions of the initial density field**.



$$s^{(j+1)} \sim P(s | v^{(j)}, \mathbf{S}^{(j)}, d),$$

$$\mathbf{S}^{(j+1)} \sim P(\mathbf{S} | s^{(j+1)}).$$

$$v^{(j+1)} \sim P(v | s^{(j+1)}).$$

coherent flows

$$r^{(j+1)} = z - v_r^{(j+1)}.$$

sampling dispersion term

$$P(v | s^{(j)}) \propto G[v - \langle v \rangle_M(s^{(j)}), \sigma_v^2(s^{(j)})]$$

Bayesian LSS reconstruction methods

Gaussian linear Wiener filter: [Zaroubi+95](#)

[Gibbs-sampling for 2D Gaussian linear CMB maps and power spectrum
first proposed by: [Jewell+04](#) ; [Wandelt+04](#)]

Gibbs-sampling for 3D density, peculiar coherent and dispersed velocity
fields and power spectra: first proposed by: [FSK & Enßlin 08](#)

Gibbs-sampling for Gaussian linear case without RSDs:

[Jasche, FSK, Wandelt & Enßlin 10](#) with linear bias: [Jasche&Wandelt13](#)

Lognormal-Poisson model:

[FSK, Jasche & Metcalf 10](#)

Hamiltonian sampling with Lognormal-Poisson model:

[Jasche & FSK 10](#)

Gibbs-and Hamiltonian sampling for Lognormal-Poisson case with RSD:

[FSK, Gallerani & Ferrara 12](#)

Hamiltonian sampling with Lognormal-Poisson model with photozs:

[Jasche&Wandelt12](#)

with Lognormal-Negative Binomial model: [Ata, FSK & Müller 15](#)

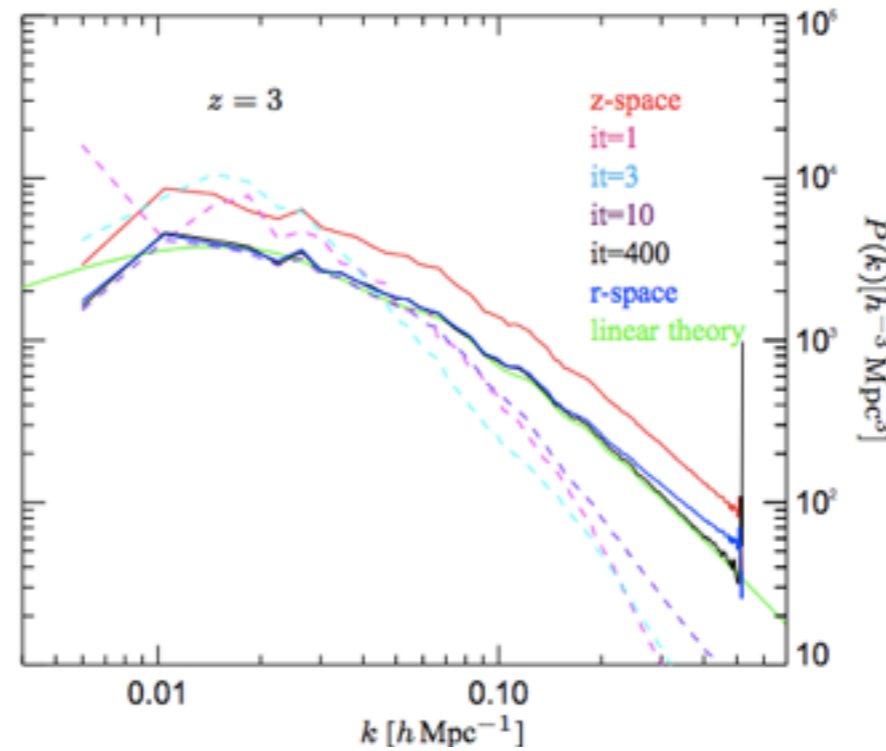
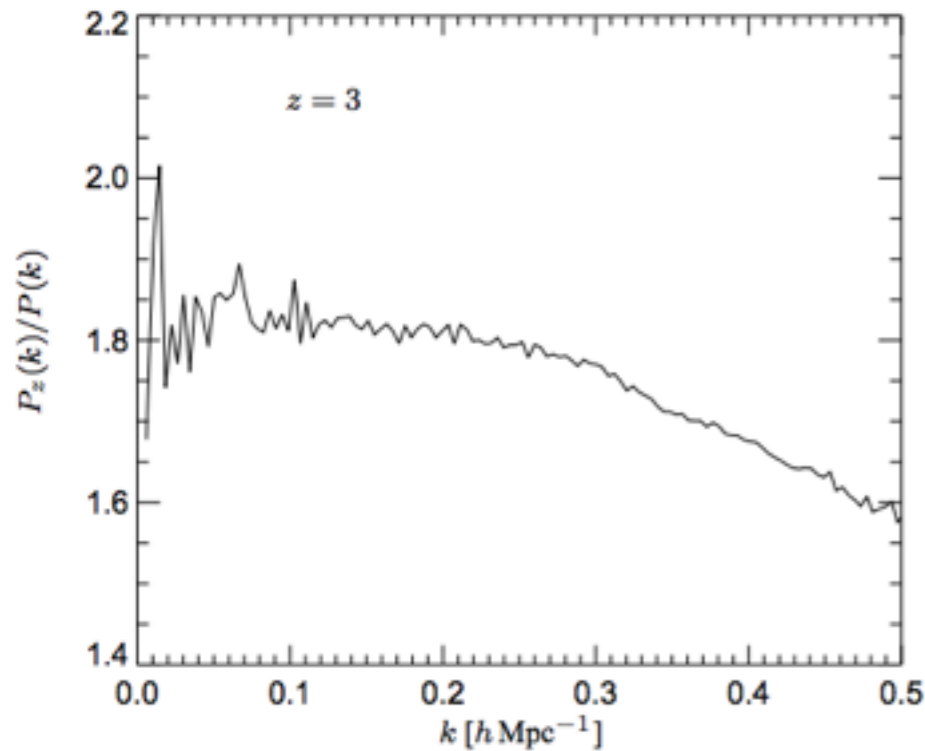
Gibbs-sampling for Gaussian linear case with 2D anisotropic P_k and
growth factor, linear bias, and selection function: [Granett+VIPERS 15](#)

including Lagrangian perturbation theory in a couple of slides

Gibbs-Hamiltonian sampling of densities,

velocity fields and power spectra

Test on simulations (*L-basic, Angulo+08*), context Lyman-a forest



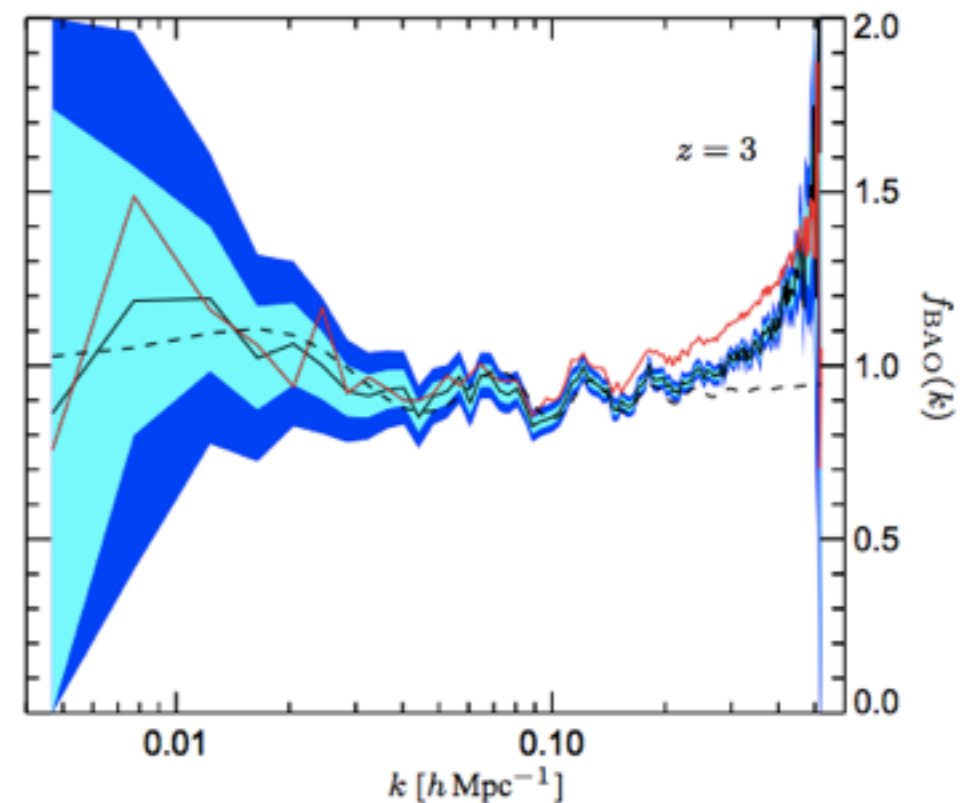
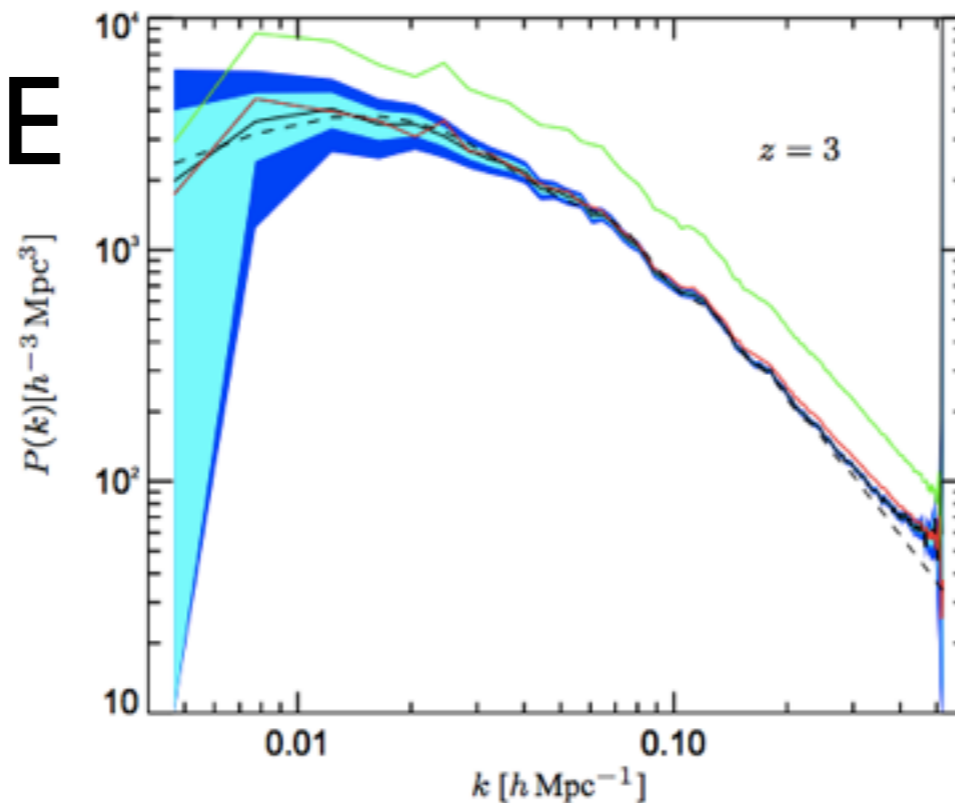
Idea: sample the Gaussian component and make the likelihood comparison with the nonlinear transformed one

$z=3$

ARGO-CODE

see also Neyrinck
+09,11,12

Carron&Szapudi13

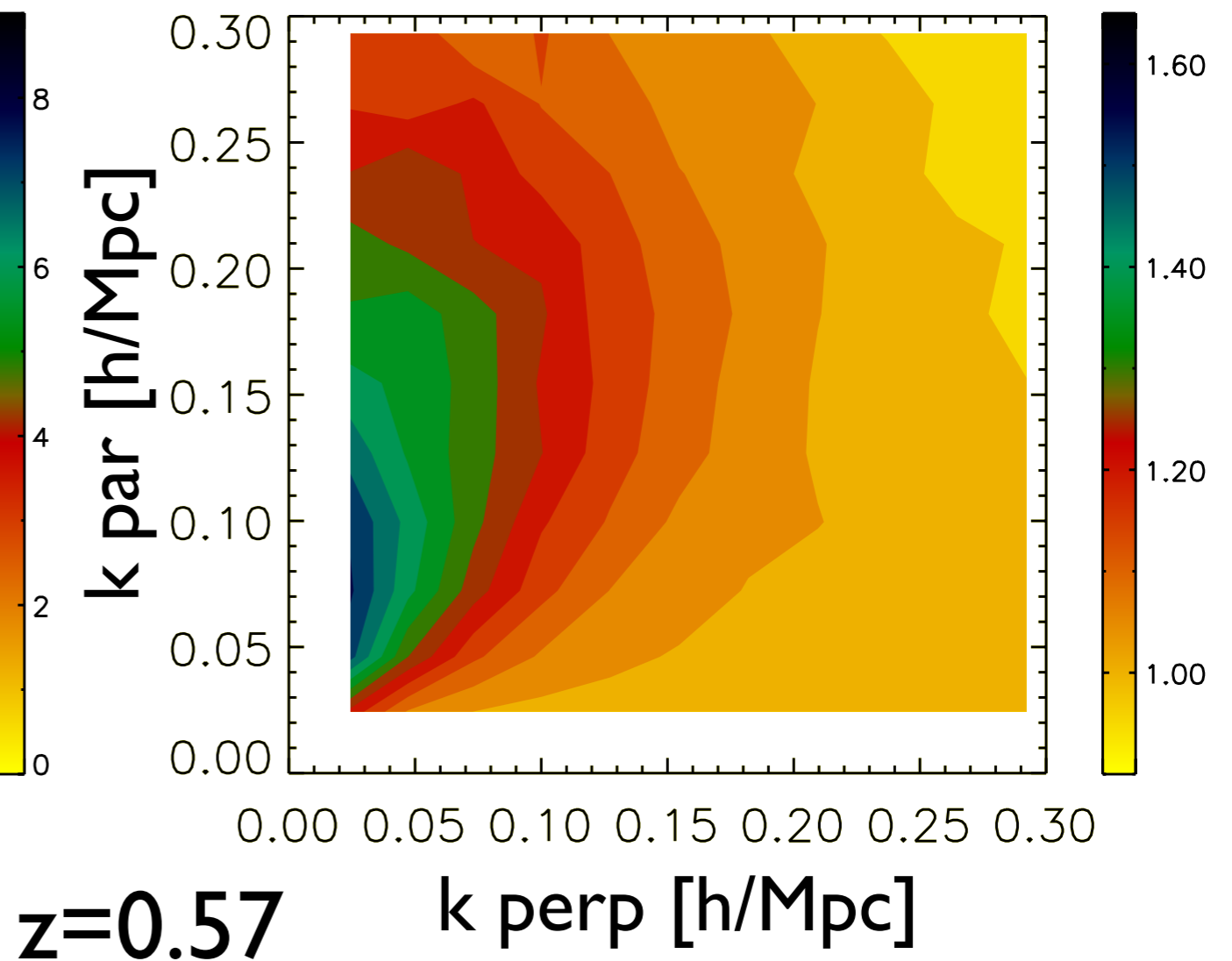
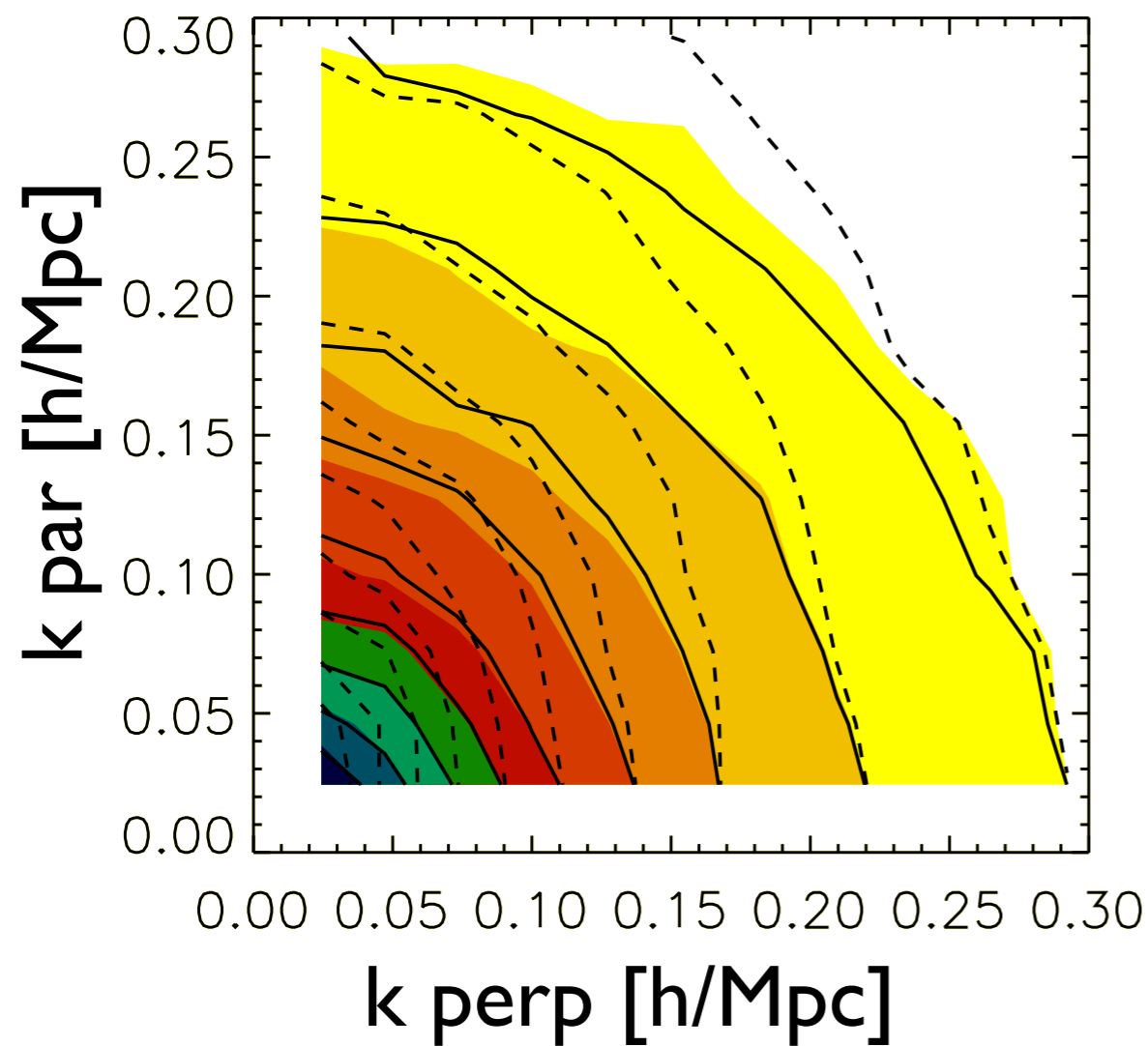


FSK, Gallerani & Ferrara 12, arXiv:1011.6233

Gibbs-Hamiltonian sampling of density and velocity fields (RSD corrections)

Test on BOSS DR12 mocks, no selection effects

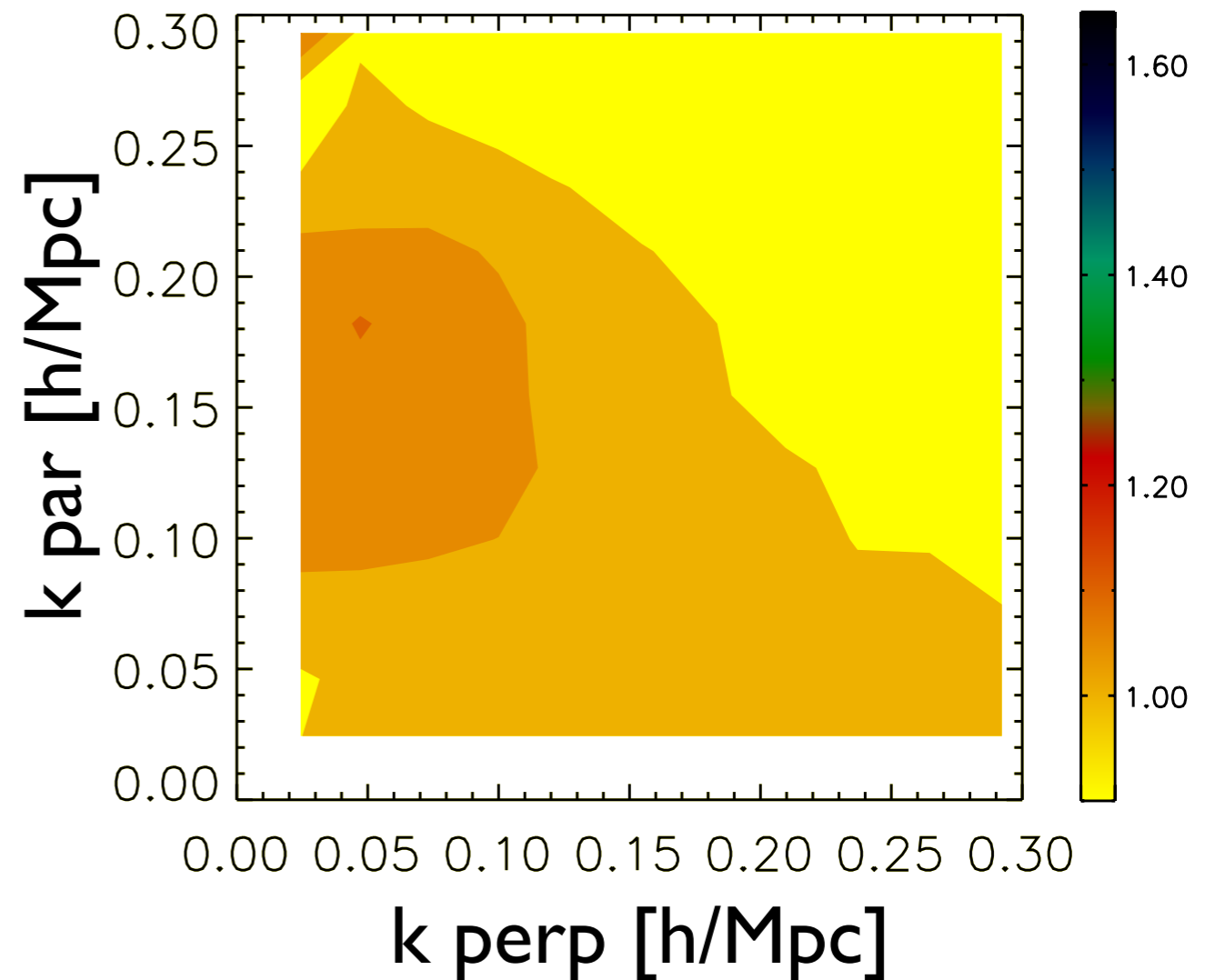
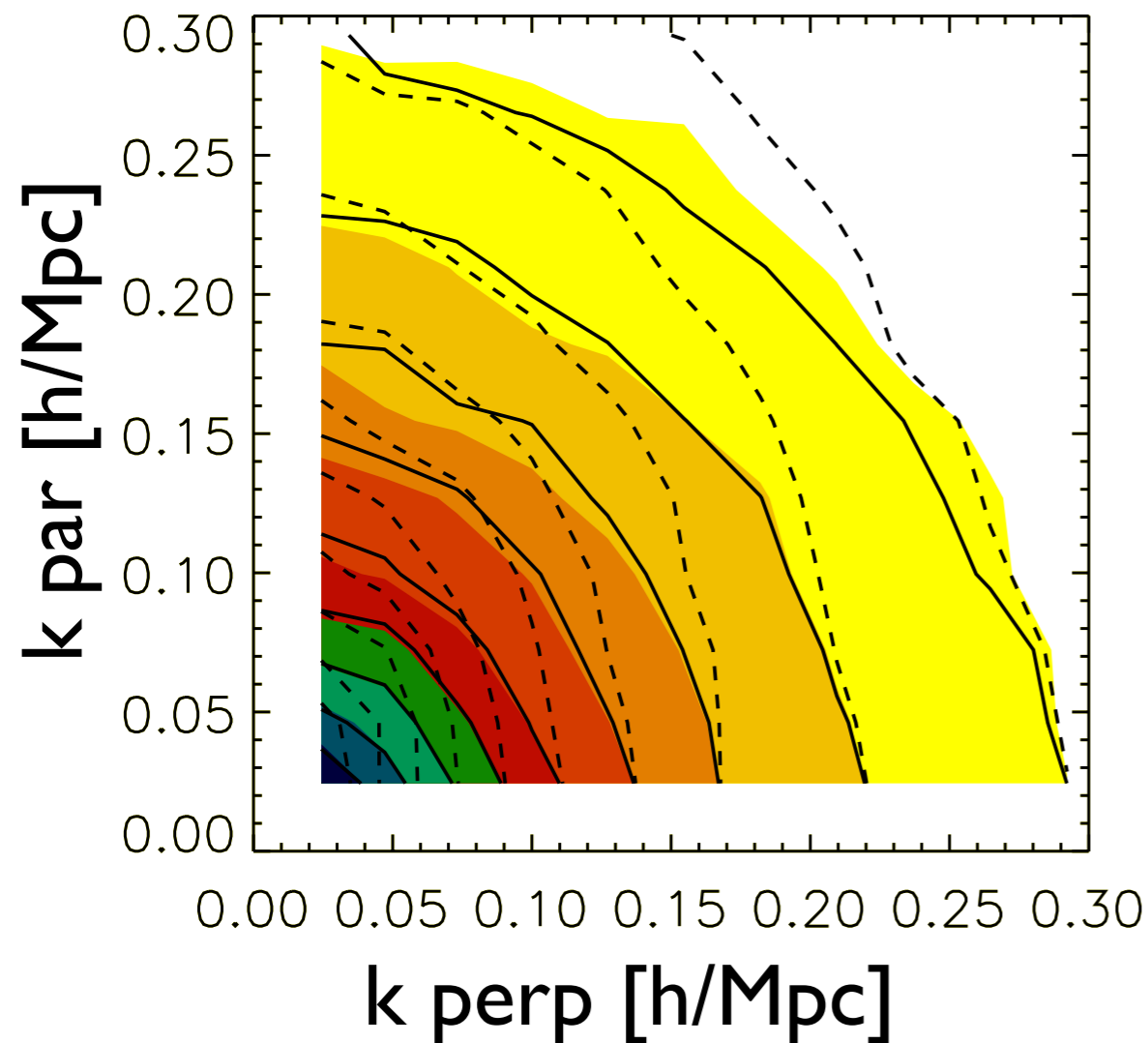
FSK, Metin Ata, Raul Angulo+ in prep



Gibbs-Hamiltonian sampling of density and velocity fields (RSD corrections)

Test on BOSS DR12 mocks, no selection effects

FSK, Metin Ata, Raul Angulo+ in prep



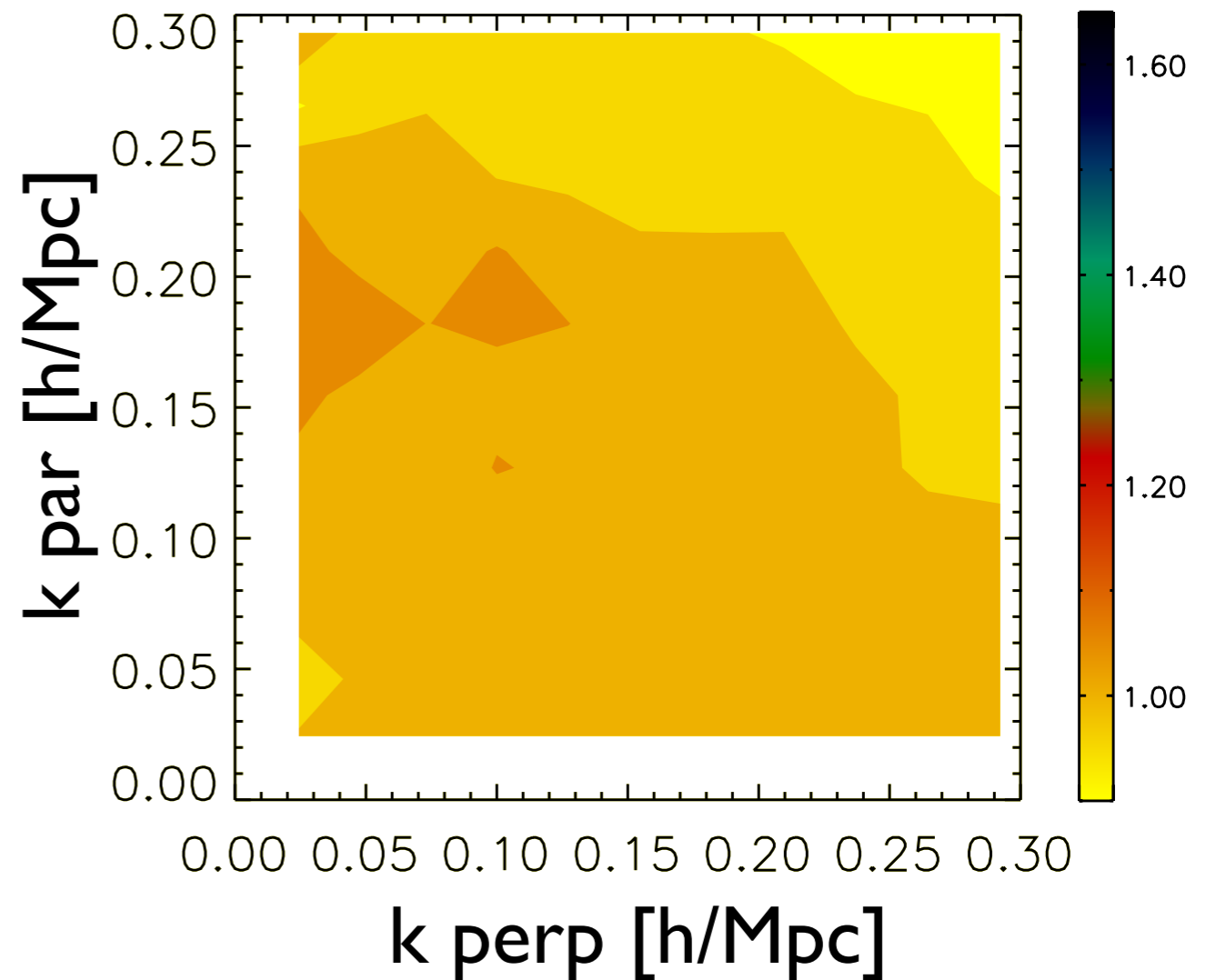
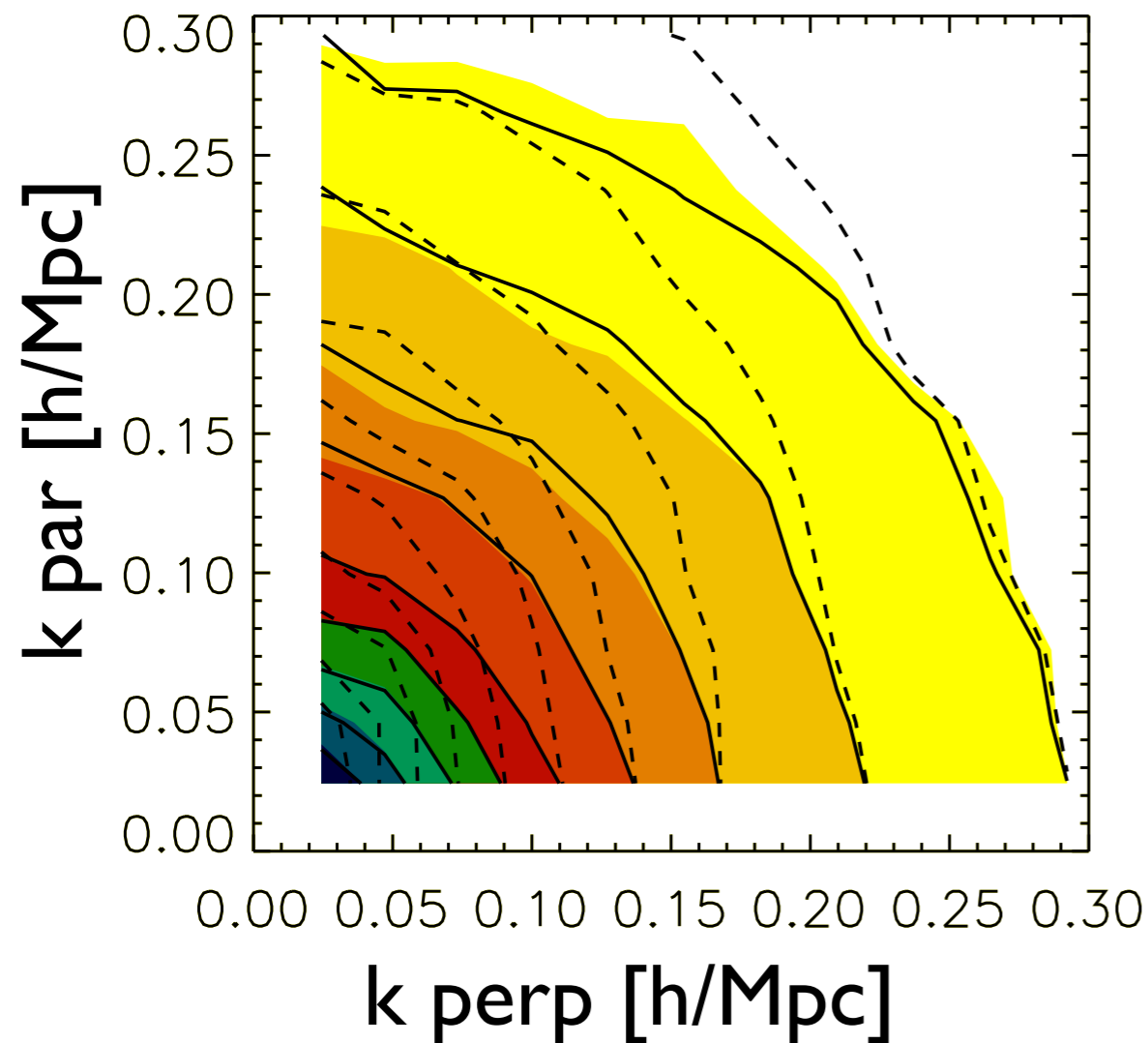
$$\mathbf{v}_{\text{bulk}} = f_{\Omega} H a \nabla \nabla^{-2} \theta (\delta)$$

$$\mathbf{r}_{\text{bulk}} = \mathbf{s}^{\text{obs}} - [\mathbf{v}_{\text{bulk}}(\mathbf{r}) \cdot \hat{\mathbf{r}}] \hat{\mathbf{r}}$$

Gibbs-Hamiltonian sampling of density and velocity fields (RSD corrections)

Test on BOSS DR12 mocks, no selection effects

FSK, Metin Ata, Raul Angulo+ in prep

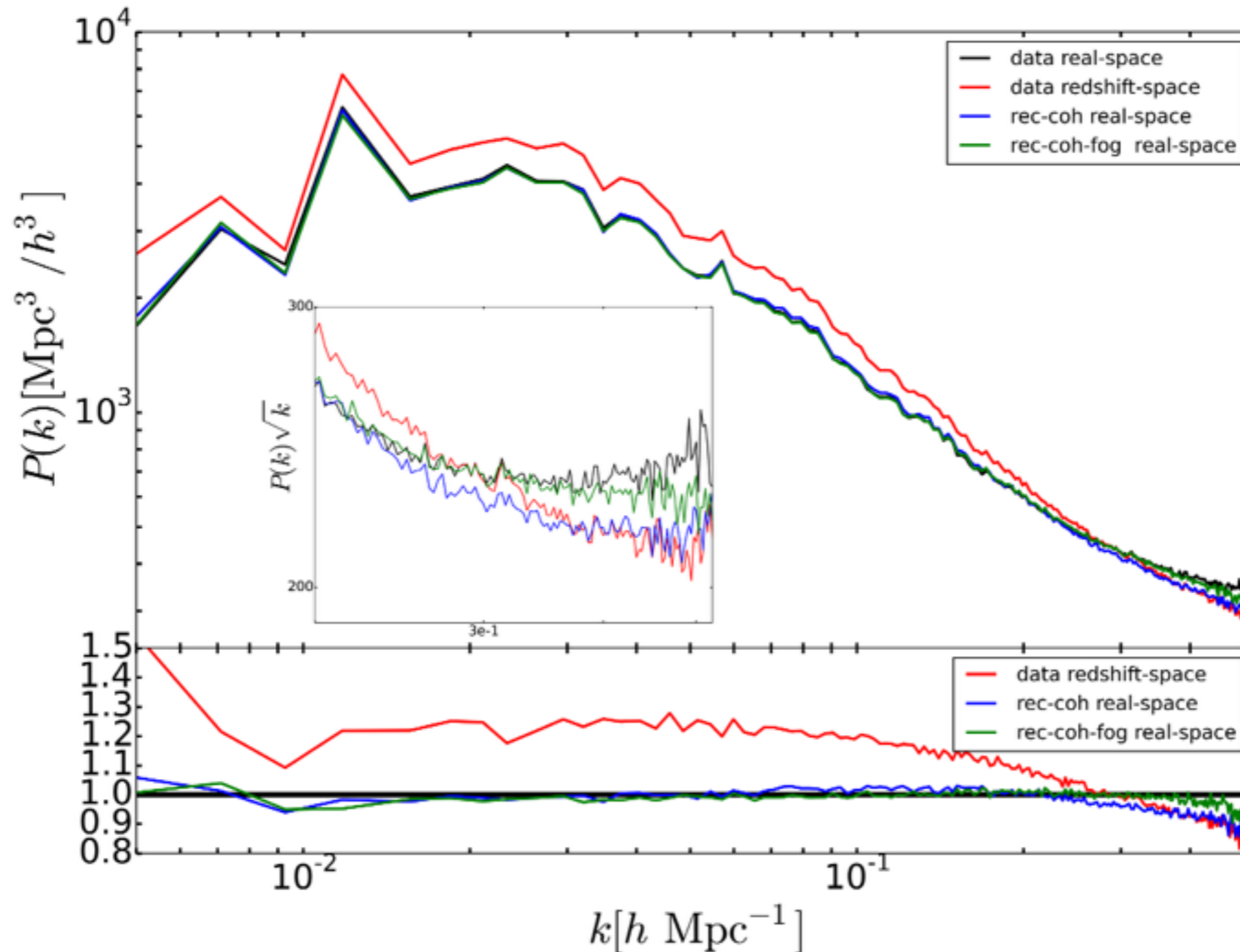


$$\begin{aligned}
 \mathbf{v}_r^\sigma(\mathbf{r}^k) &\leftarrow \mathcal{P}(\mathbf{v}_r^\sigma(\mathbf{r}^k) \mid \sigma(\delta(\mathbf{r}^k)), \forall \lambda(\mathbf{H}(\delta)) > 0) \\
 \mathbf{r}_{\text{bulk}}^k &= \mathbf{r}^k + \mathbf{v}_r^\sigma(\mathbf{r}^k) \quad \min(\mathbf{r}_{\text{bulk}} - \mathbf{r}_{\text{bulk}}^k (= \mathbf{r}^k + \mathbf{v}_r^\sigma(\mathbf{r}^k))) \text{ \& } \delta(\mathbf{r})_i \geq \delta(\mathbf{r})_{i-1}
 \end{aligned}$$

Gibbs-Hamiltonian sampling of density and velocity fields (RSD corrections)

Test on BOSS DR12 mocks, no selection effects

FSK, Metin Ata, Raul Angulo+ in prep



Gibbs-Hamiltonian sampling of density and velocity fields, and Pk sampling

Test on BOSS DR12 mocks, no selection effects

Ata, FSK+BOSS 15

collaborators Chia-Hsun Chuang, Sergio Rodriguez-Torres, Francisco Prada, Mark Neyrinck+BOSS

$$\begin{aligned} \delta &\curvearrowright \mathcal{P}(\delta \mid N(\{\mathbf{r}\}), w, \mathbf{C}, \{b_p\}), \\ \{\mathbf{r}\} &\curvearrowright \mathcal{P}(\{\mathbf{r}\} \mid \{\mathbf{s}^{\text{obs}}\}, \{\mathbf{v}(\delta, \mathbf{H}(\delta), f_\Omega)\}), \\ w &\curvearrowright \mathcal{P}(w \mid \{\mathbf{r}\}, m(\alpha, \delta)), \\ \mathbf{C} &\curvearrowright \mathcal{P}(\mathbf{C} \mid \Phi(\delta)). \end{aligned}$$

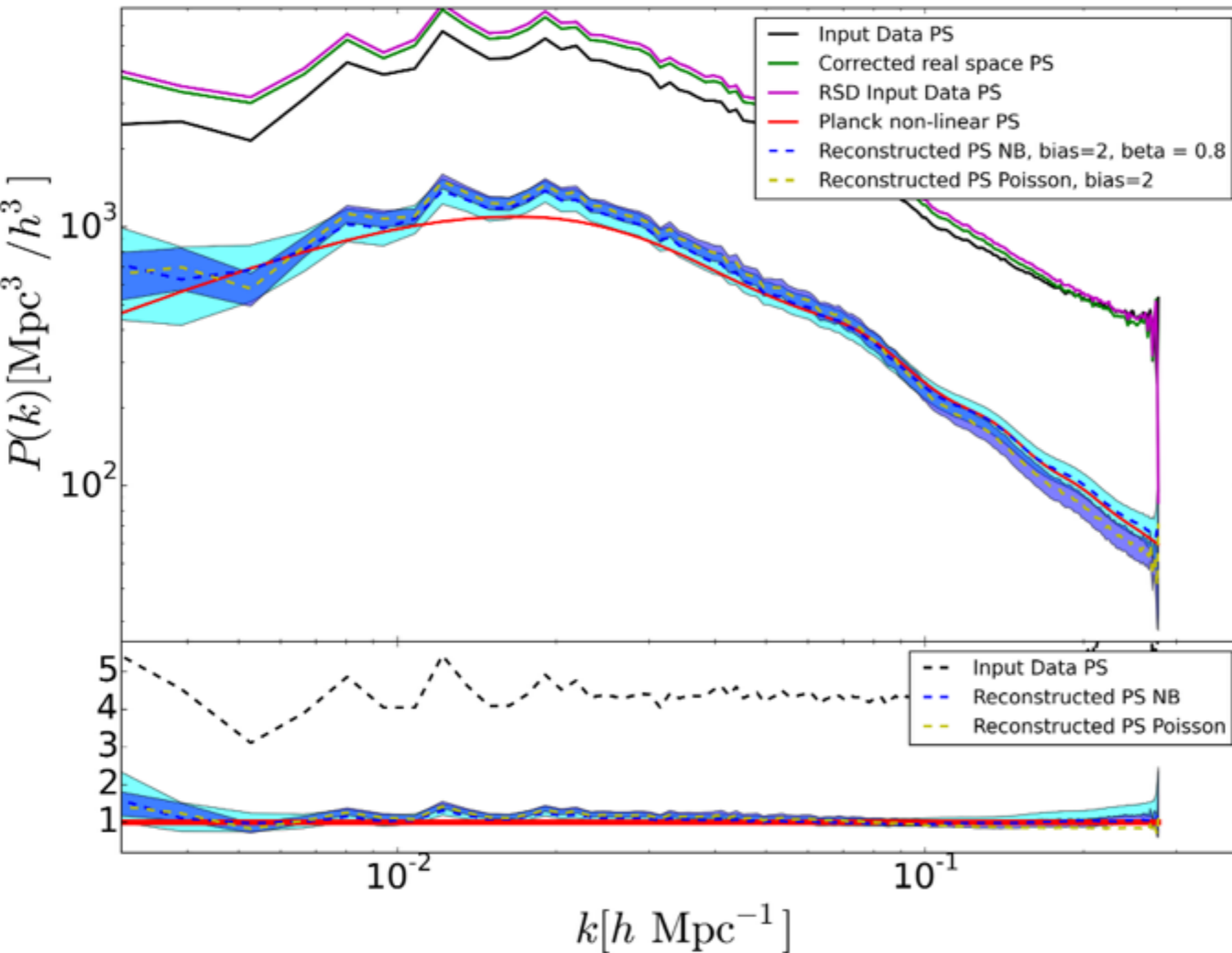
Metin Ata's PhD thesis

see Metin Ata's poster!

Gibbs-Hamiltonian sampling of density, velocity fields, and Pk sampling

Test on BOSS DR12 mocks, no selection effects

Metin Ata, FSK+BOSS 15

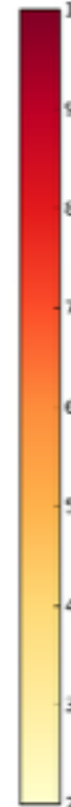
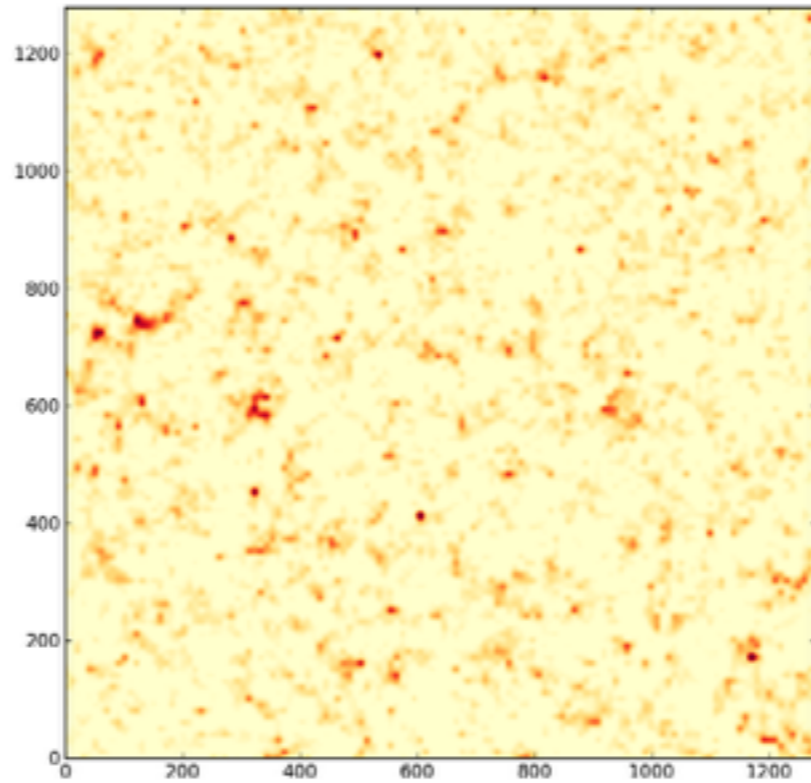
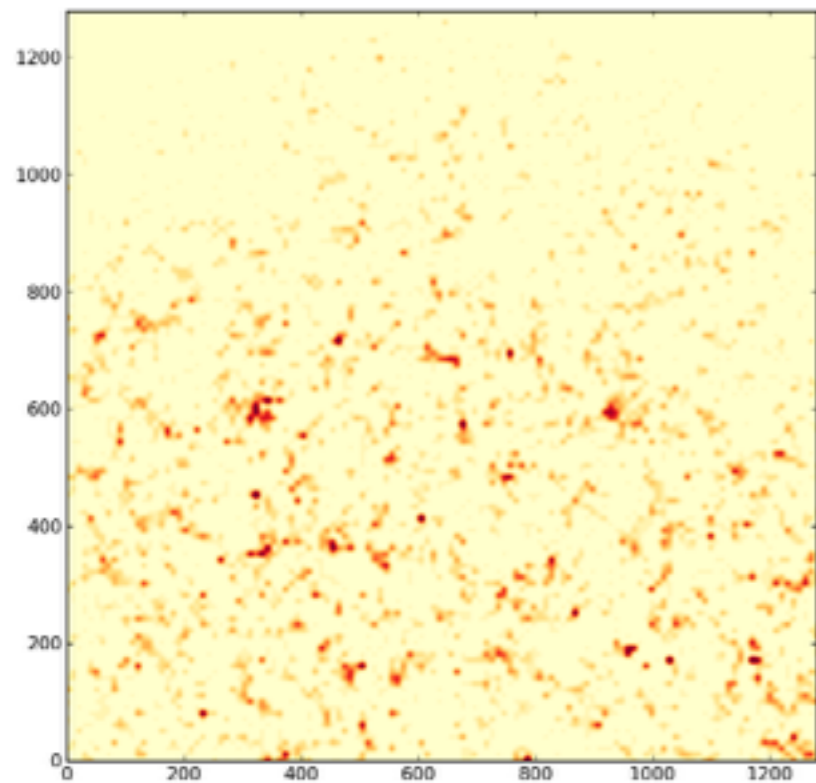
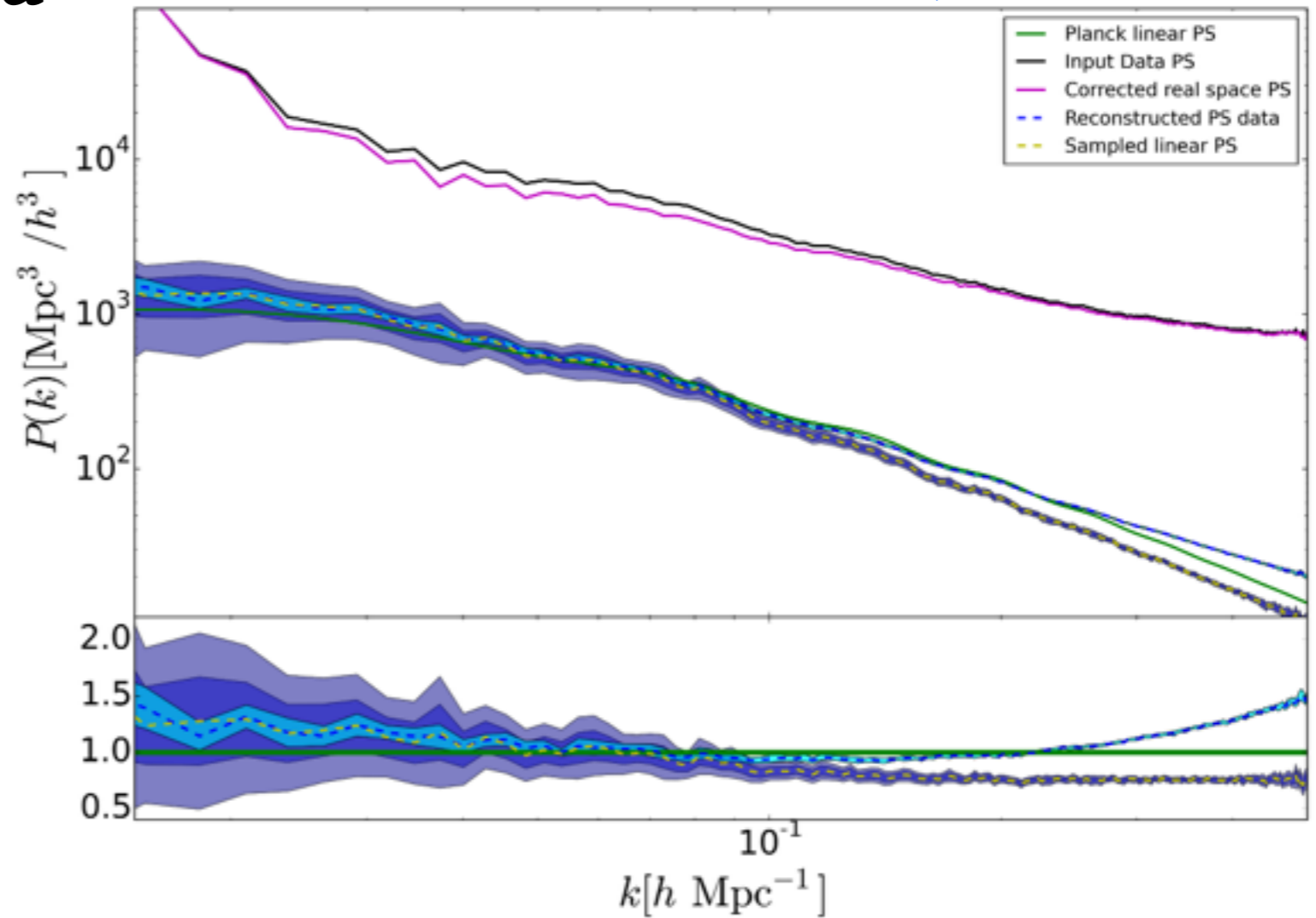
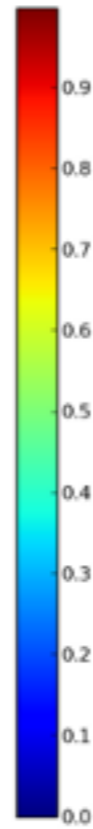
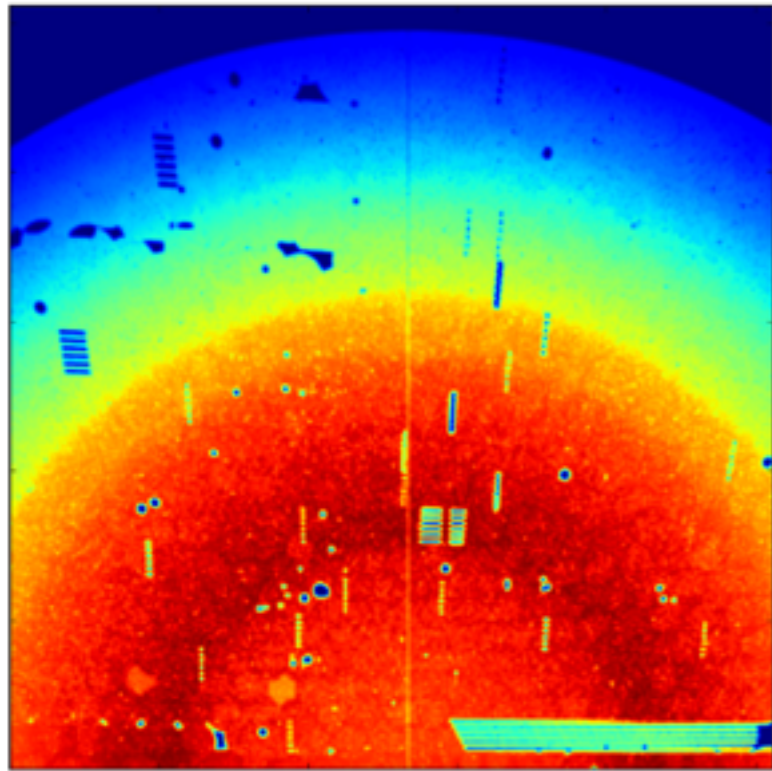


Including nonlinear stochastic bias: $\rho_G = \gamma \Theta (\rho_M - \rho_{th}) \rho_M^\alpha$

Metin Ata, FSK & Müller 14 arXiv:1408.2566

On BOSS DR12 data

Metin Ata, FSK+BOSS 15



Chia-Hsun Chuang is investigating the reconstructed linearised BAO peak. And the obtained fields are used for BAO reconstruction.

Can we include nonlocal bias in the Bayesian reconstruction method?

see Mathieu Autefage's poster!

$$s^2 = -2\mu^{(2)}[\Phi_M] + \frac{2}{3}\delta_M^2 \quad \rho_h = f_h \theta(\delta_M - \delta_M^{\text{th}}) \left[(1 + \delta_M)^\alpha + c_{\text{NL}}\mu^{(2)}(\nabla^{-2}\delta_M) \right]$$

$$2\mu^{(2)}(\Phi_M) \equiv - \sum_{ij} \Phi_{M,ij} \Phi_{M,ji} + \delta_M^2$$

necessary gradient for the Hamiltonian sampler trivially implemented in ARGO

FSK+ in prep

$$\begin{aligned} \sum_{\mathbf{r}} \frac{\partial \sum_{ij} \Phi_{M,ij}^2(\mathbf{r})}{\partial \delta_M(\mathbf{r}')} &= \sum_{\mathbf{r}} \sum_{ij} 2\Phi_{M,ij}(\mathbf{r}) \frac{\partial \Phi_{M,ij}(\mathbf{r})}{\partial \delta_M(\mathbf{r}')} \\ &= \sum_{\mathbf{r}} \sum_{ij} 2\Phi_{M,ij}(\mathbf{r}) \frac{1}{N} \sum_{\mathbf{k}} e^{-i\mathbf{k}\mathbf{r}} \frac{k_i k_j}{k^2} \sum_{\mathbf{r}''} e^{i\mathbf{k}\mathbf{r}''} \delta_{\mathbf{r}'', \mathbf{r}'} \\ &= \sum_{\mathbf{r}} \sum_{ij} 2\Phi_{M,ij}(\mathbf{r}) \frac{1}{N} \sum_{\mathbf{k}} e^{-i\mathbf{k}\mathbf{r}} \frac{k_i k_j}{k^2} e^{i\mathbf{k}\mathbf{r}'} \\ &= \frac{2}{N} \sum_{\mathbf{k}} e^{i\mathbf{k}\mathbf{r}'} \sum_{ij} \frac{k_i k_j}{k^2} \sum_{\mathbf{r}} e^{-i\mathbf{k}\mathbf{r}} \Phi_{M,ij}(\mathbf{r}) \\ &= \frac{2}{N} \sum_{\mathbf{k}} e^{i\mathbf{k}\mathbf{r}'} \sum_{ij} \frac{k_i k_j}{k^2} \bar{\Phi}_{M,ij}(\mathbf{k}) \\ &= \frac{2}{N} \sum_{\mathbf{k}} e^{i\mathbf{k}\mathbf{r}'} \sum_{ij} \frac{k_i^2 k_j^2}{k^4} \bar{\delta}_M(\mathbf{k}) \end{aligned}$$

Can we improve the prior with a physical model?

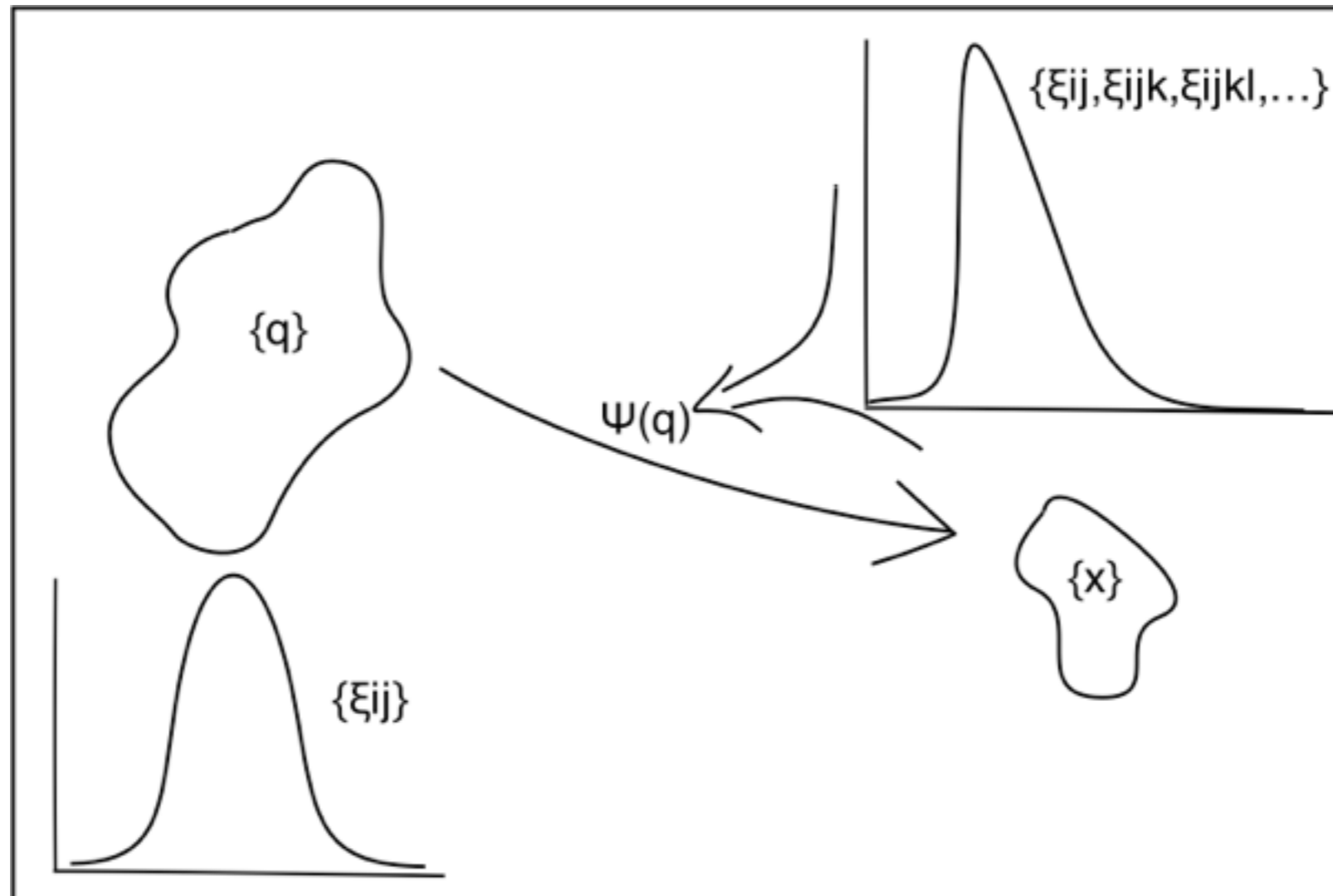
including Lagrangian perturbation theory in Bayesian reconstruction methods

first independently proposed methods by: *Jasche&Wandelt13 FSK13*
FSK+12

alternative later method by: *Wang+13* with approx PM: *Wang+14*

application to data: *FSK+12; Heß, FSK+13; Nuza, FSK+14*
Jasche+15 Leclerque, Jasche+15

The reconstruction problem: Lagrangian to Eulerian problem



primordial density fluctuations can be fully characterized by the 2-point correlation function (neglecting non-Gaussianities)

the action of gravity can be summarised by the displacement field

KIGEN-CODE

instead of making gradients of complex structure formation models:
radical simplification of the problem through Gibbs-sampling splitting
approach:

Bayesian Networks Machine Learning (artificial intelligence)

KIGEN-code (Kinetic GENeration of the initial conditions, japanese origin)

$$\begin{aligned}\delta(\mathbf{q}) &\curvearrowright P(\delta(\mathbf{q})|\{\mathbf{q}^\circ\}, \{p_c\}) \\ \{\mathbf{q}^\circ\} &\curvearrowright P(\{\mathbf{q}^\circ\}|\{\mathbf{s}^\circ\}, \delta(\mathbf{q}), w(\mathbf{s}), \mathcal{M}_\Psi : \mathbf{q} \rightarrow \mathbf{s}, \{p_c\})\end{aligned}$$

1) constrained realisation

Bertschinger 87; Hoffman & Ribak 91;

van de Weygaert & Bertschinger 96; Jasche & FSK 10, FSK+12

2) constrained simulation

arbitrary structure formation model!

new KIGEN-CODE *FSK in prep*

instead of making gradients of complex structure formation models:
radical simplification of the problem through Gibbs-sampling splitting
approach:

Bayesian Networks Machine Learning (artificial intelligence)

KIGEN-code (Kinetic GENeration of the initial conditions, japanese origin)

$$\delta(\mathbf{q}) \quad \curvearrowright \quad P(\delta(\mathbf{q}) | \Psi_{\text{L}}^{\circ}(\mathbf{q}, z_i), \{p_c\})$$

$$\Psi_{\text{L}}^{\circ}(\mathbf{q}, z_i) \quad \curvearrowright \quad P(\Psi_{\text{L}}^{\circ}(\mathbf{q}, z_i) | \{\mathbf{s}^{\circ}\}, \delta(\mathbf{q}), w(\mathbf{s}), \mathcal{M}_{\Psi} : \mathbf{q} \rightarrow \mathbf{s}, \{p_c\})$$

1) constrained realisation

Bertschinger 87; Hoffman & Ribak 91;

van de Weygaert & Bertschinger 96; Jasche & FSK 10, FSK+12

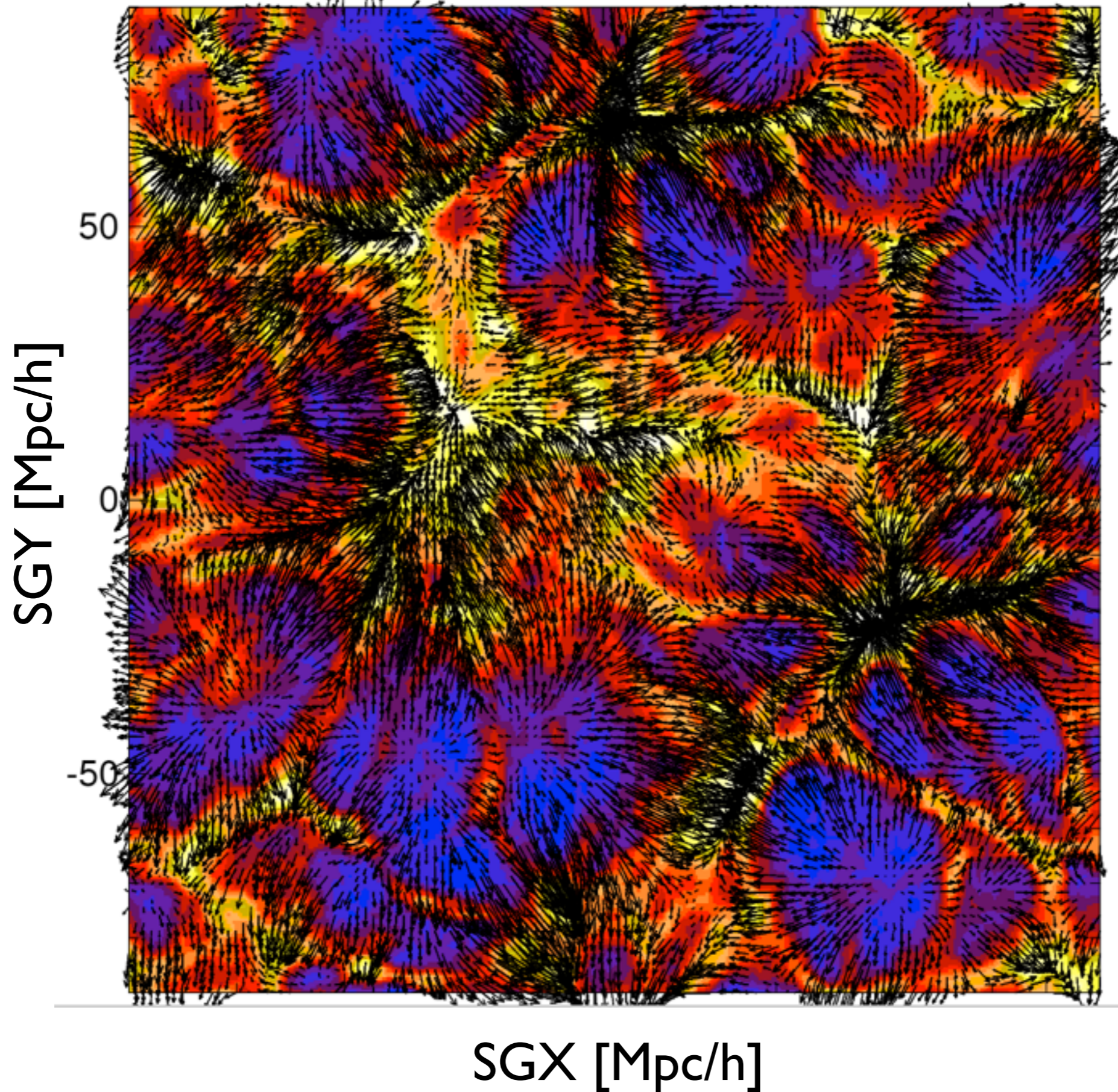
2) constrained simulation

arbitrary structure formation model!

**Applications to the Local Volume
with the old version of KIGEN**

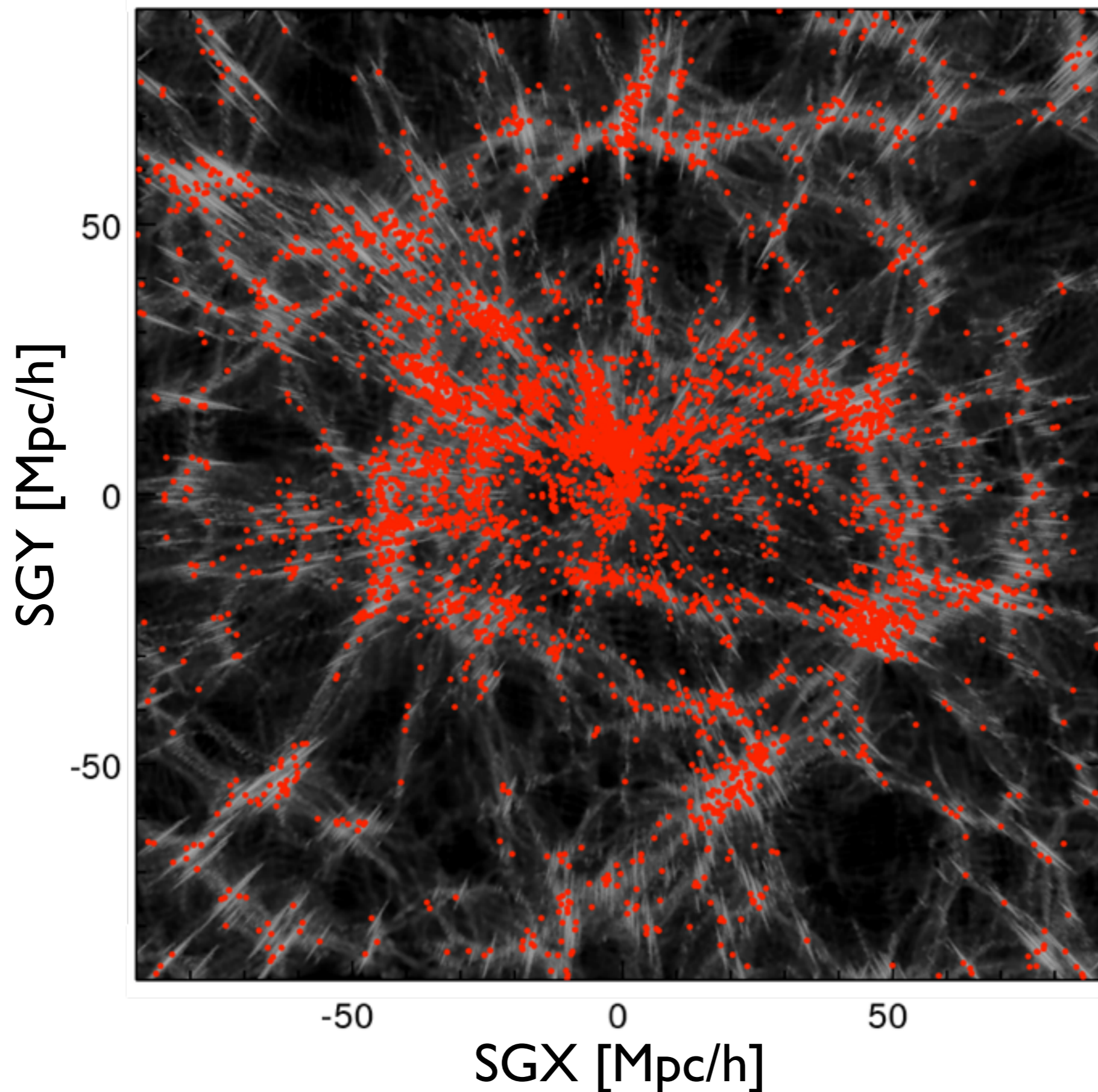
reconstructed peculiar velocity field from

FSK+12 arXiv:1205.5560



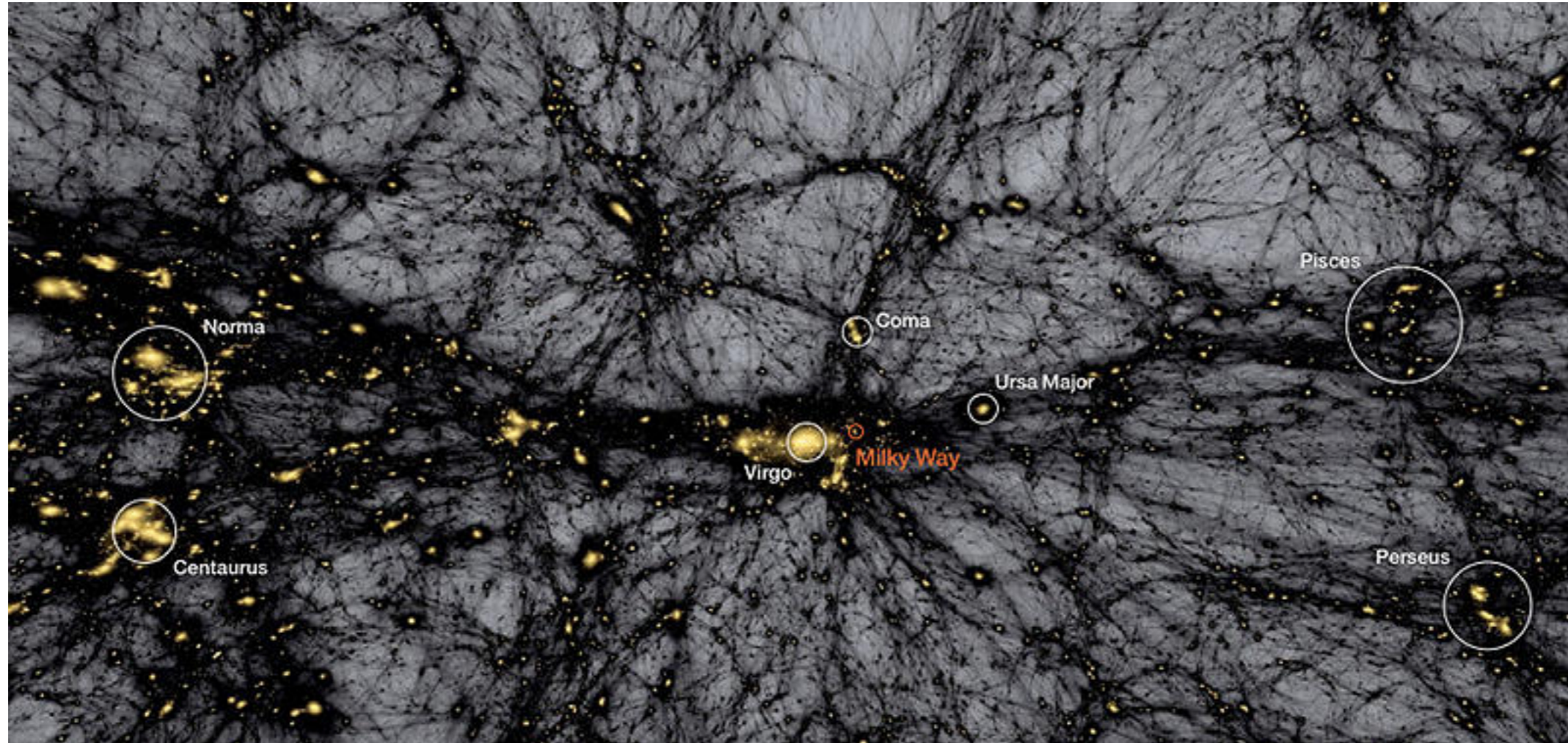
2013: 1st constrained simulation based on a self-consistent phase-space reconstruction code

Steffen Hess, FK & Gottlöber 13, arXiv:1304.6565



Red dots represent galaxies from the 2MRS survey (2% are randomly augmented in the galactic plane)

Underlying countour represents the DM constrained simulation



SIMULATION AND RECONSTRUCTION: STEFFEN HESS AND FRANCISCO-SHU KITaura, LEIBNIZ INSTITUTE FOR ASTROPHYSICS POTSDAM.

VISUALIZATION: TOM ABEL AND RALF KAEHLER, STANFORD KAVLI INSTITUTE FOR PARTICLE ASTROPHYSICS AND COSMOLOGY

Conclusions

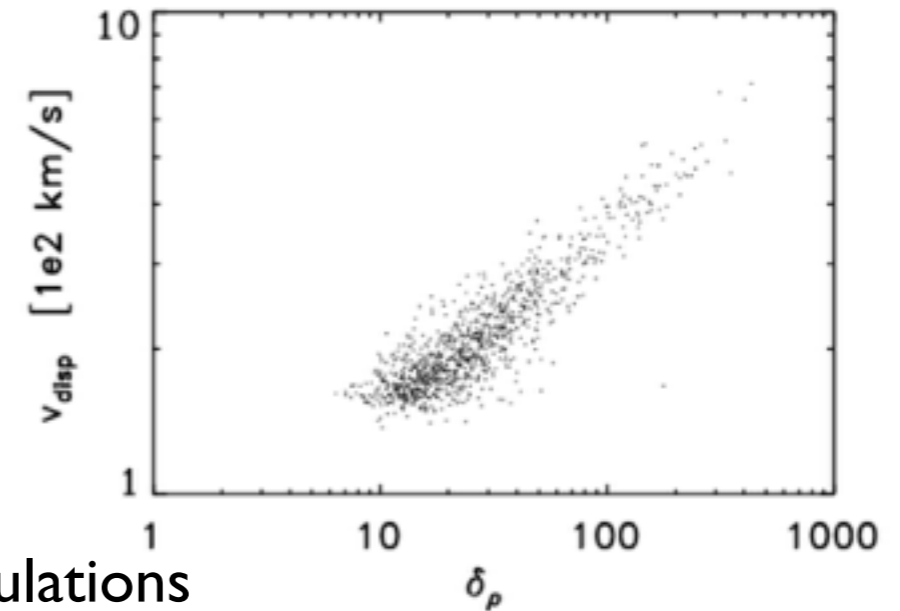
- * We have presented a number of bias components and their impact on halo/galaxy clustering (nonlinear, stochastic and nonlocal), and have shown how these can be incorporated in a Bayesian framework to make a complete self-consistent inference of the cosmological large-scale structure.
- * **ALPT**: Augmented Lagrangian Perturbation Theory, most accurate one-step gravity solver
- * **PATCHY**: ALPT + nonlinear stochastic and nonlocal halo / galaxy biasing accurate mocks for galaxy redshift surveys BOSS DR11 / DR12
- * **ARGO**: statistical based reconstruction code of density fields power spectra and RSDs, BAO, application to BOSS DR12
- * **KIGEN**: Nonlinear phase-space reconstructions are possible including redshift space uncertainties and bias!
- * The range of applications is very wide: CMB-dipole, Bulk flows, Hubble constant, WHIM, kSZ, ISW, BAO reconstruction, RSDs, cosmic web classification, environmental studies

Modelling redshift space distortions

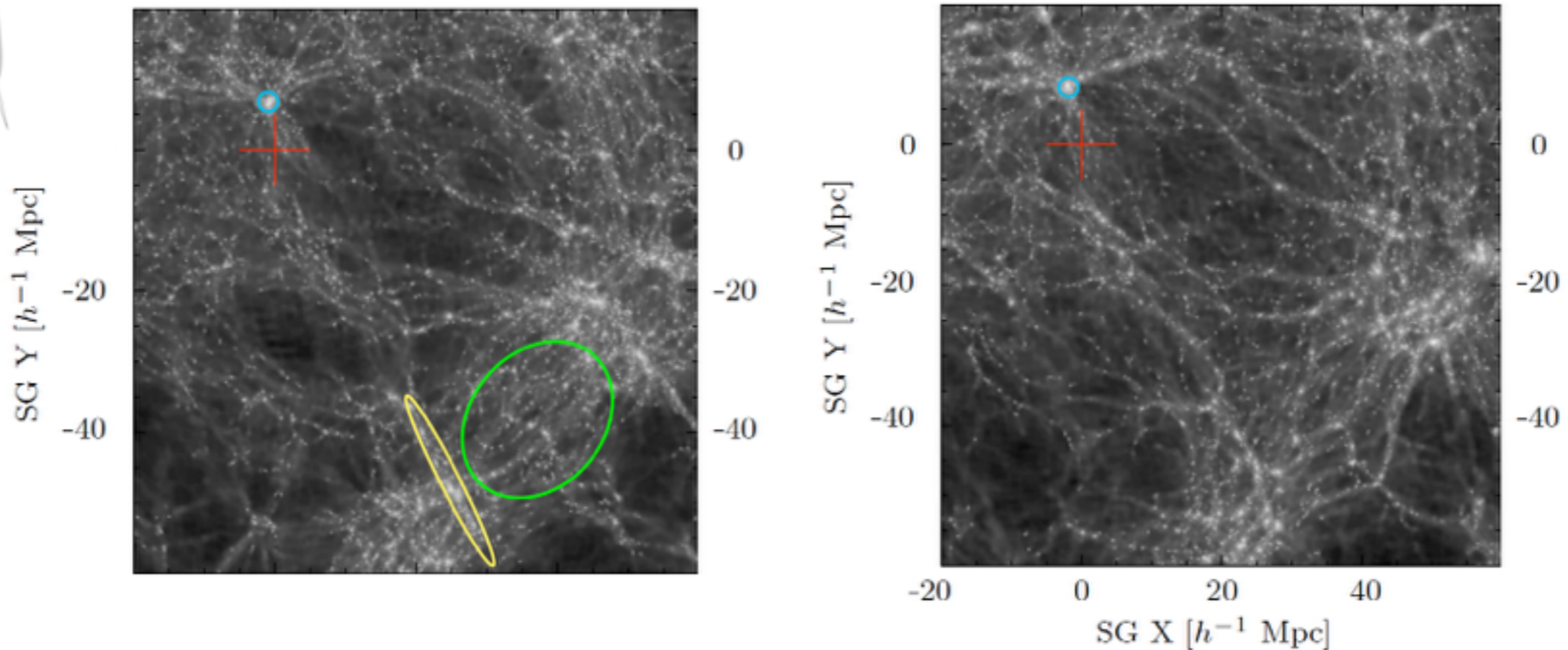
FSK PhD 2007 from MR sim
FSK+14

$$\mathbf{v} = \mathbf{v}^{\text{coh}} + \mathbf{v}^{\sigma}$$

$$\mathbf{v}_r^{\sigma} \equiv (\mathbf{v}^{\sigma} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} / (Ha) = \mathcal{G} \left(g \times \left(1 + b^{\text{ALPT}} \delta^{\text{ALPT}}(\mathbf{x}) \right)^{\gamma} \right) \hat{\mathbf{r}}$$



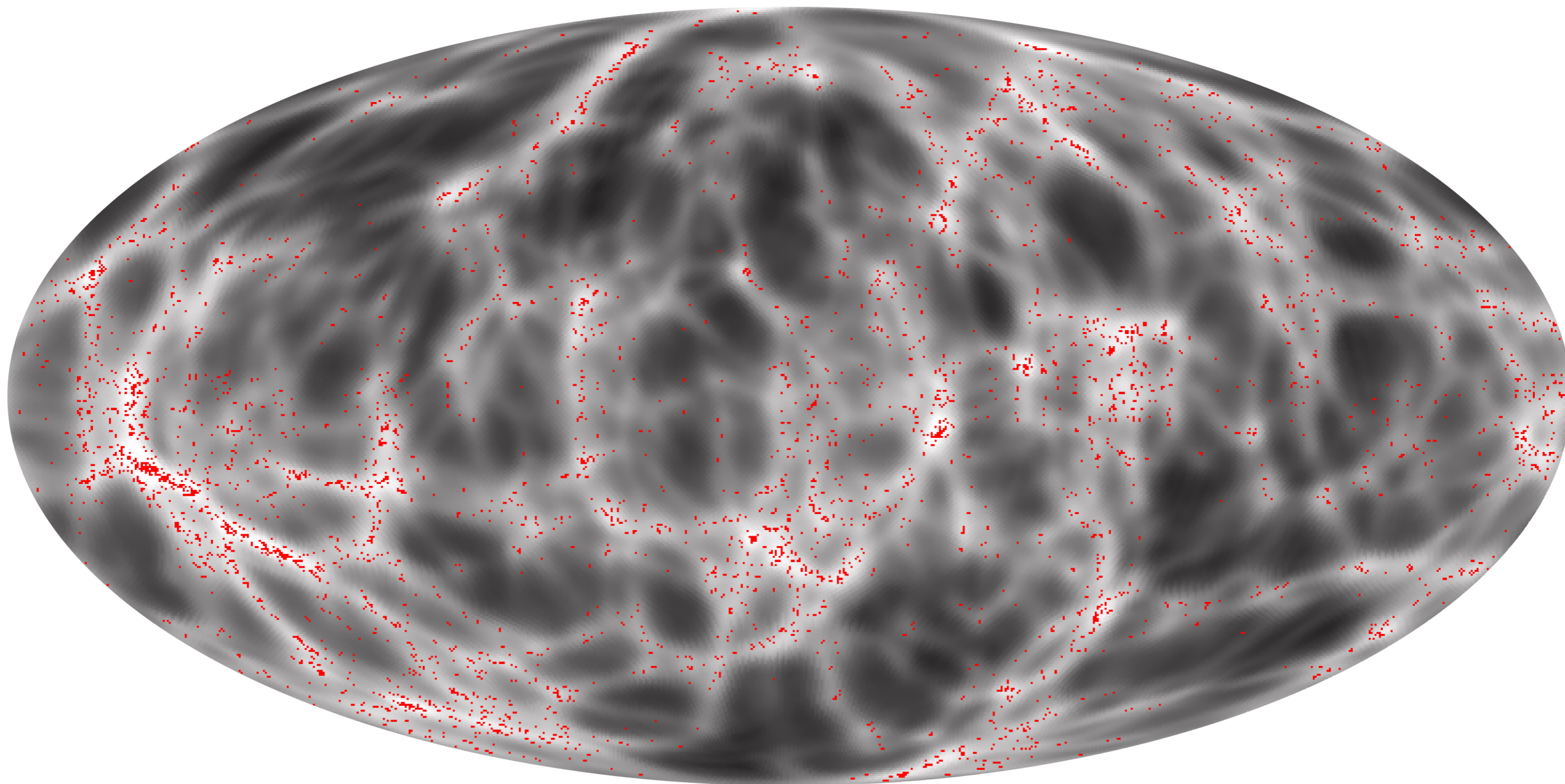
Application of KIGEN to 2MRS: constrained simulations



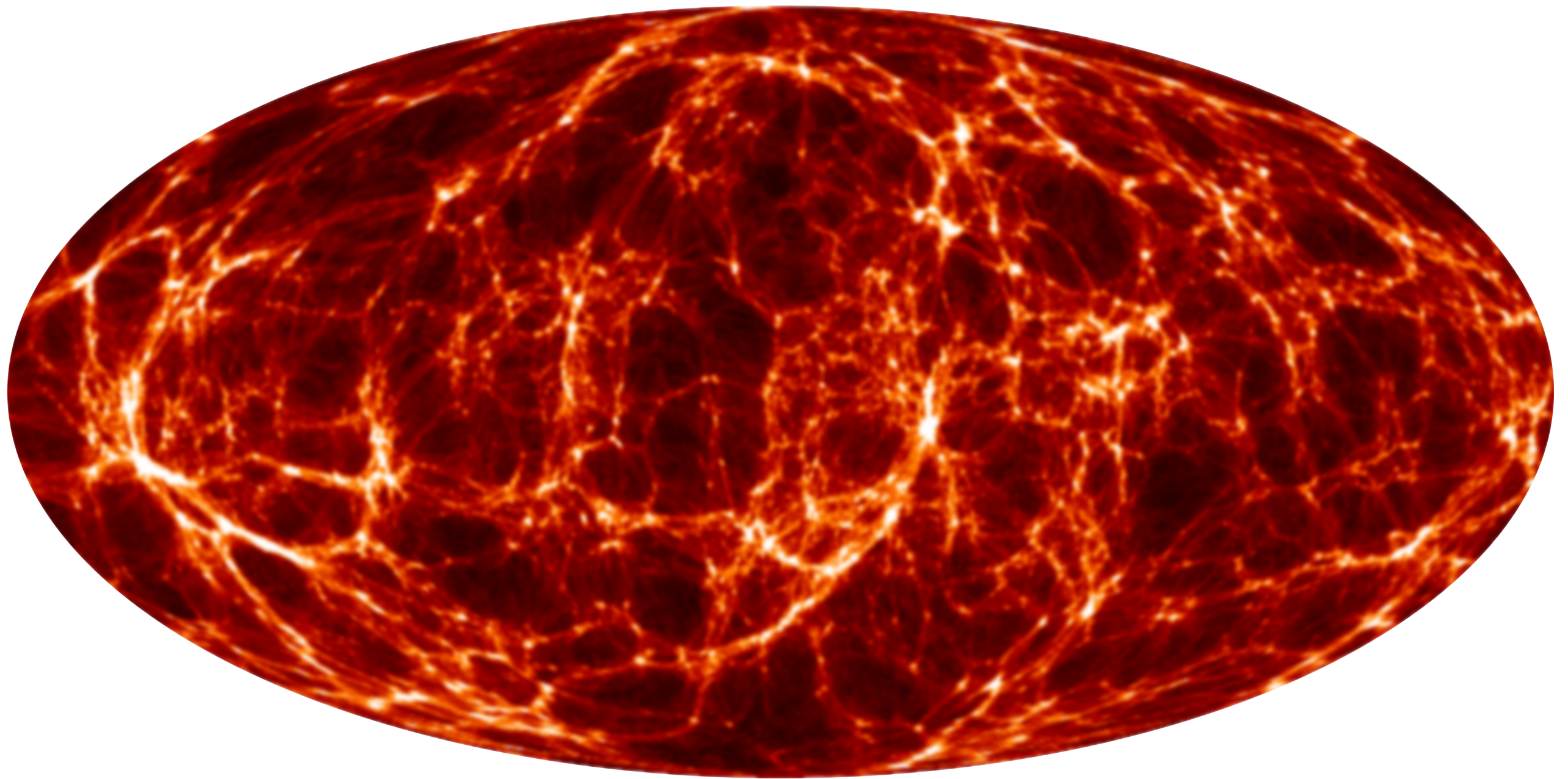
Left: constrained simulation based on 2LPT and compressed fogs

Right: constrained simulation based on ALPT modelling fogs

Hess Steffen, FSK. & Gottlöber 13 arXiv:1304.6565



mean over an ensemble of ALPT reconstruction!



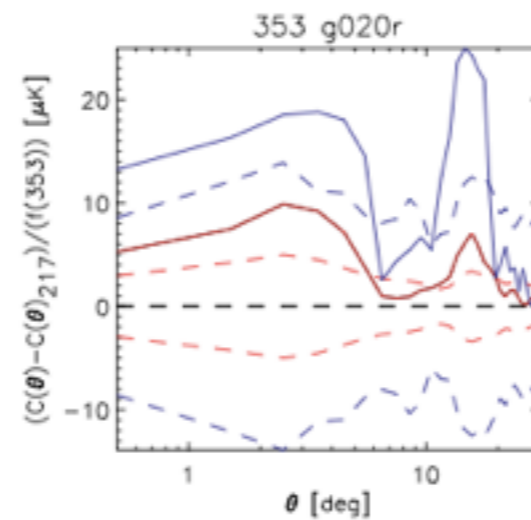
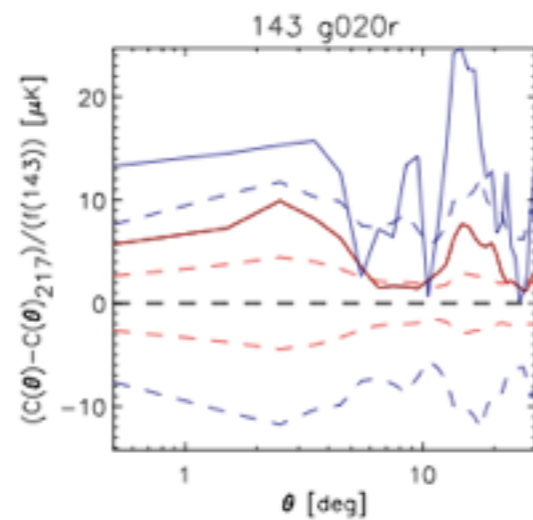
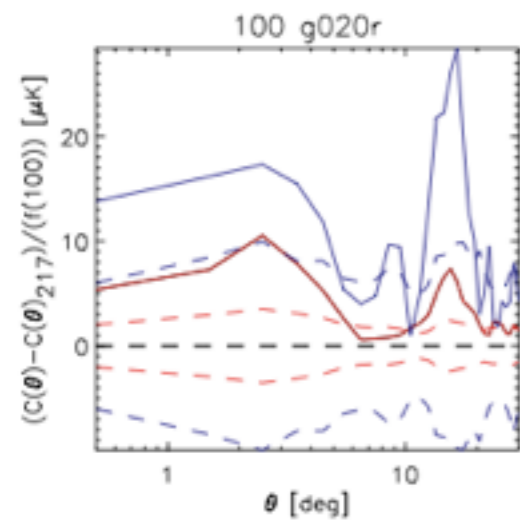
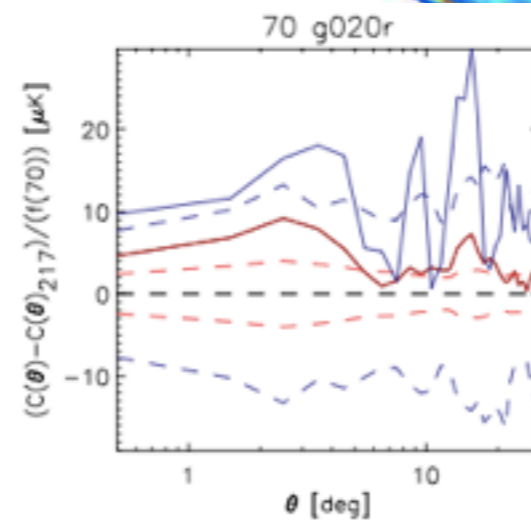
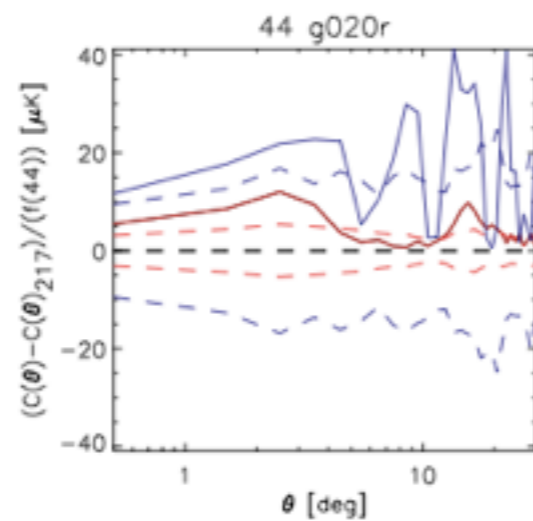
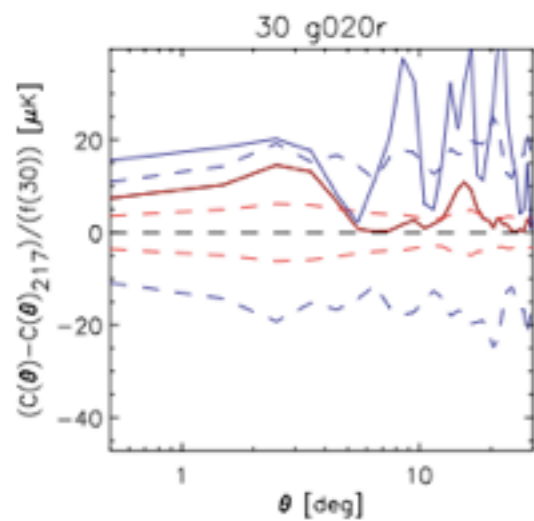
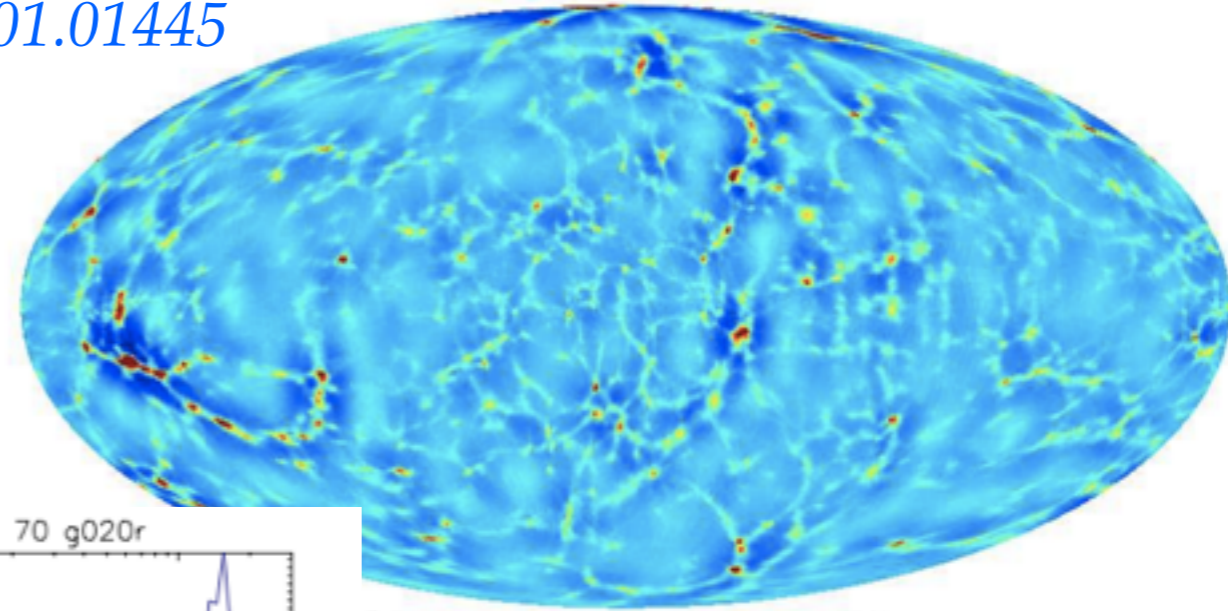
high resolution constrained N-body simulation!

Searching missing baryons from public PLANCK data!

Suarez-Velásquez, FSK, Atrio-Barandela, Mückel 13, arXiv:1303.5623

results from PLANCK

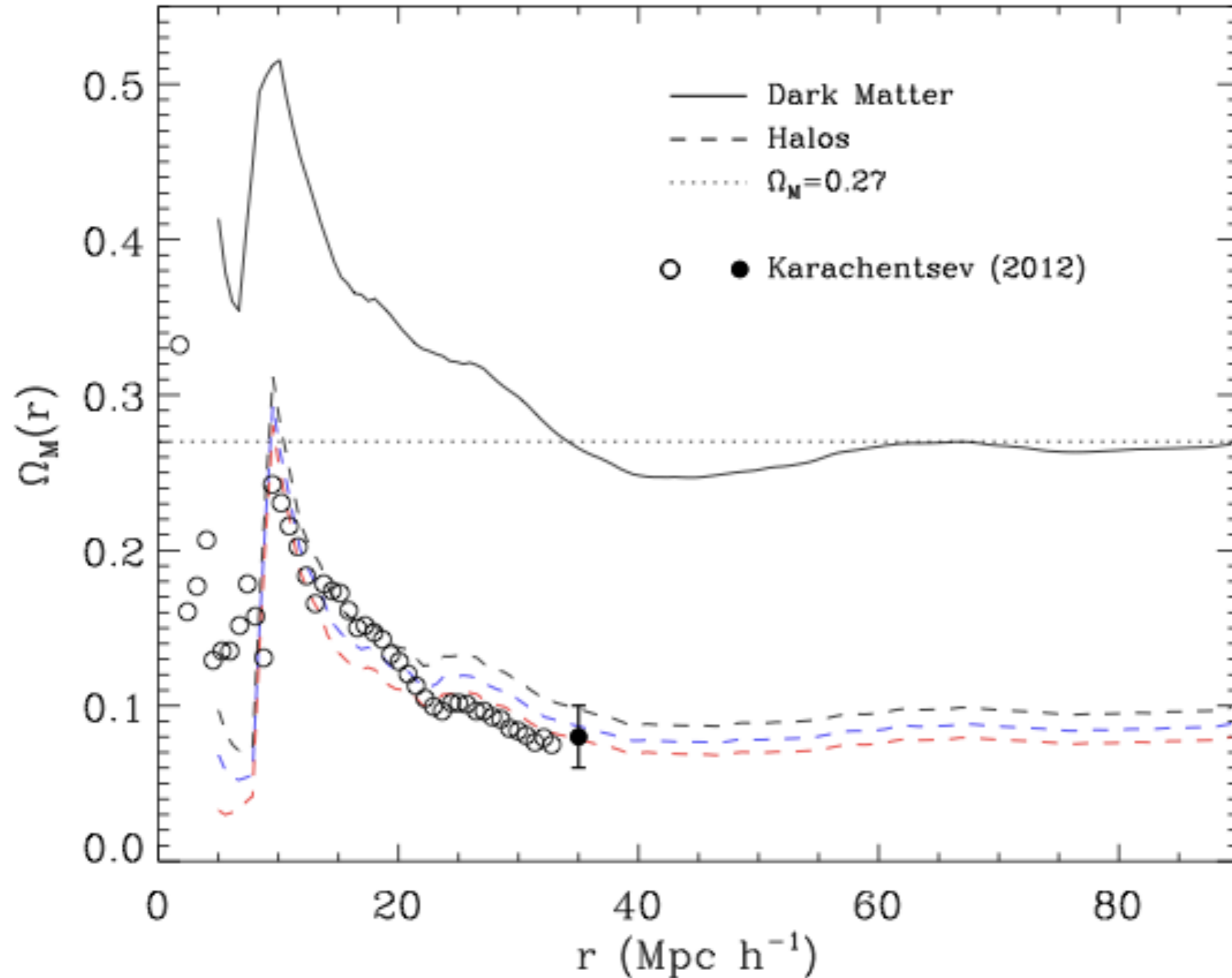
Genova-Santos, Atrio-Barandela, FSK, Mückel arxiv:1501.01445



but we have the full cosmic web...

Missing mass in the Local Universe?

Sebastian Nuza, FK, Heß, Libeskind & Müller 14, arXiv:1406.1004



Galaxy morphology-environmental study

Mercedes Filho, Sanchez-Almeida, Muñoz-Torron, Nuza, FSK, Heß 15 arXiv:1501.06709

