Astro@TS

Covariance matrix estimation for the statistics of galaxy clustering

Manuel Colavincenzo

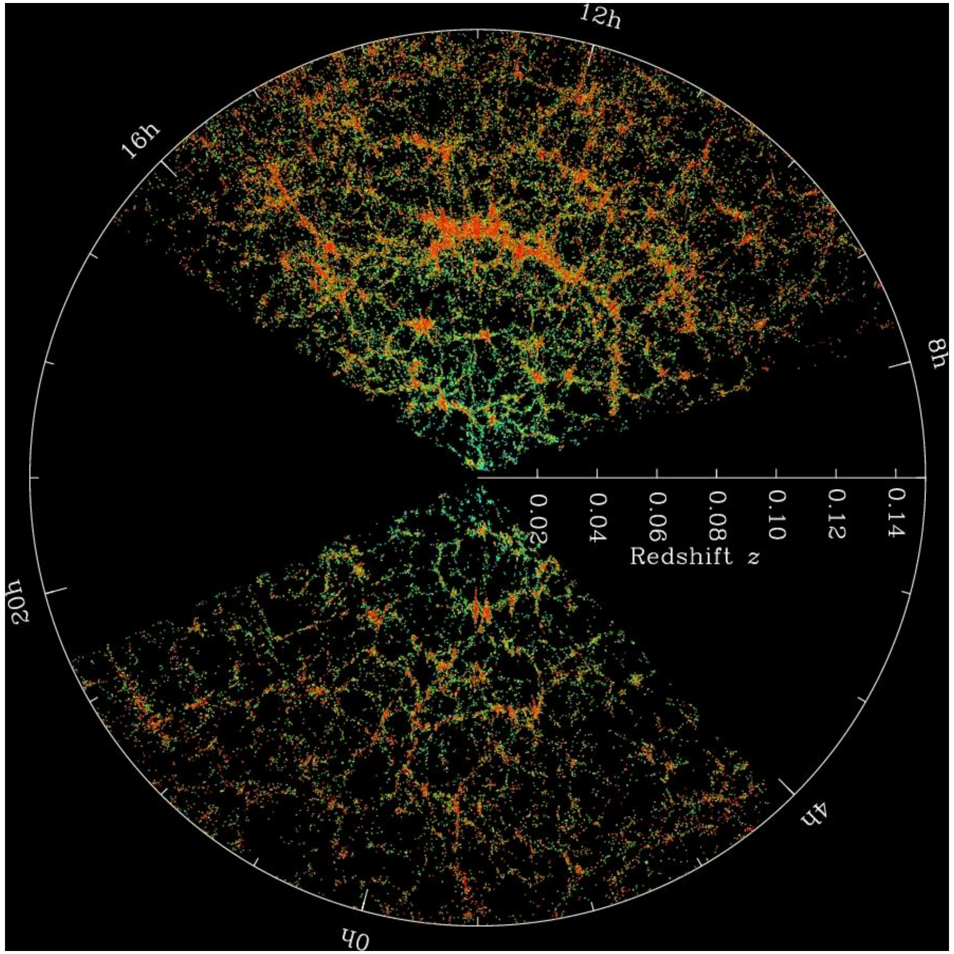
Supervisor: Pierluigi Monaco
Co-Supervisors: Stefano Borgani, Emiliano Sefusatti

UNIVERSITÀ
DEGLI STUDI DI TRIESTE

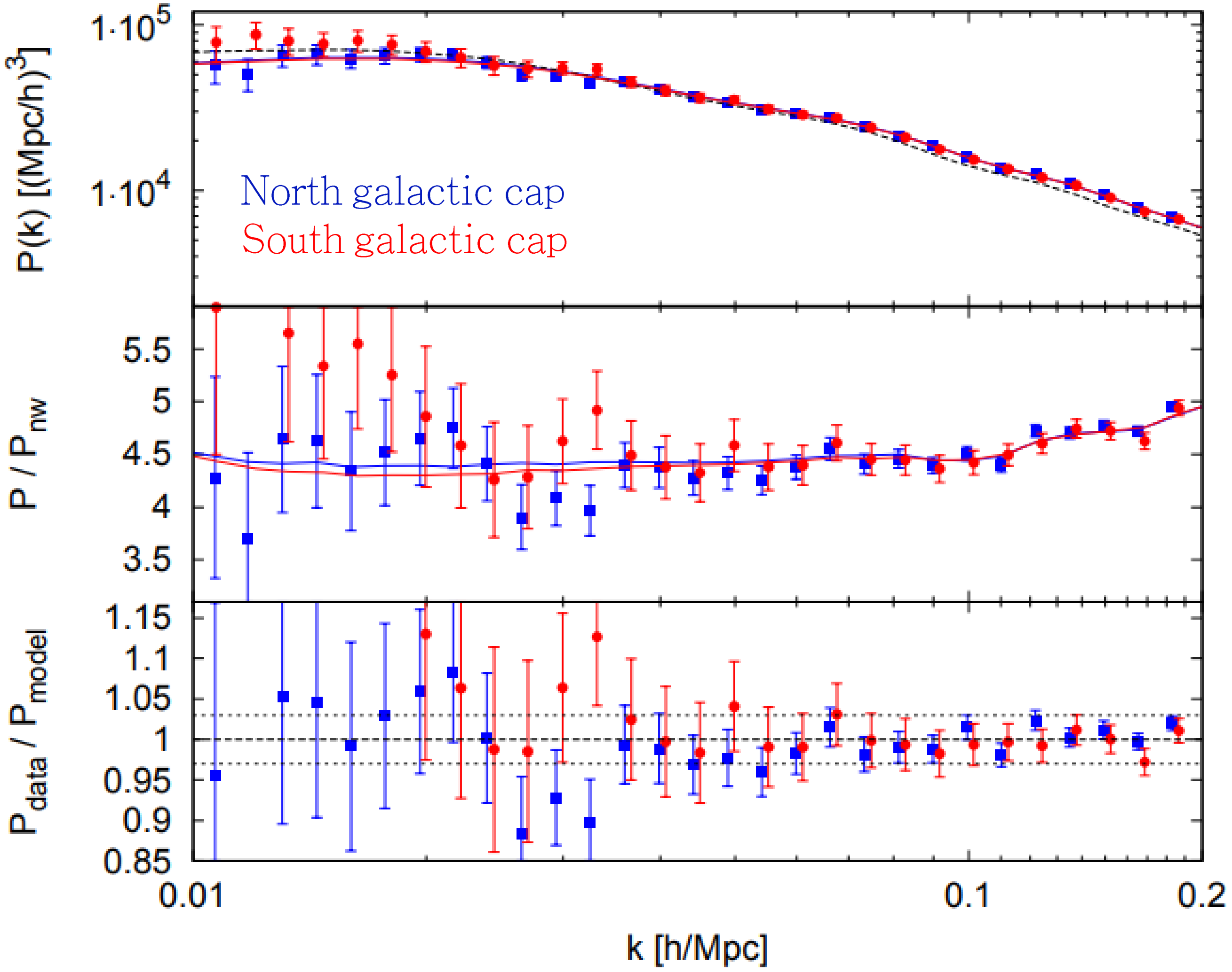
Trieste, 26.09.2017

General introduction

Gil-Marin et al. 2015
(SDSS-BOSS paper)

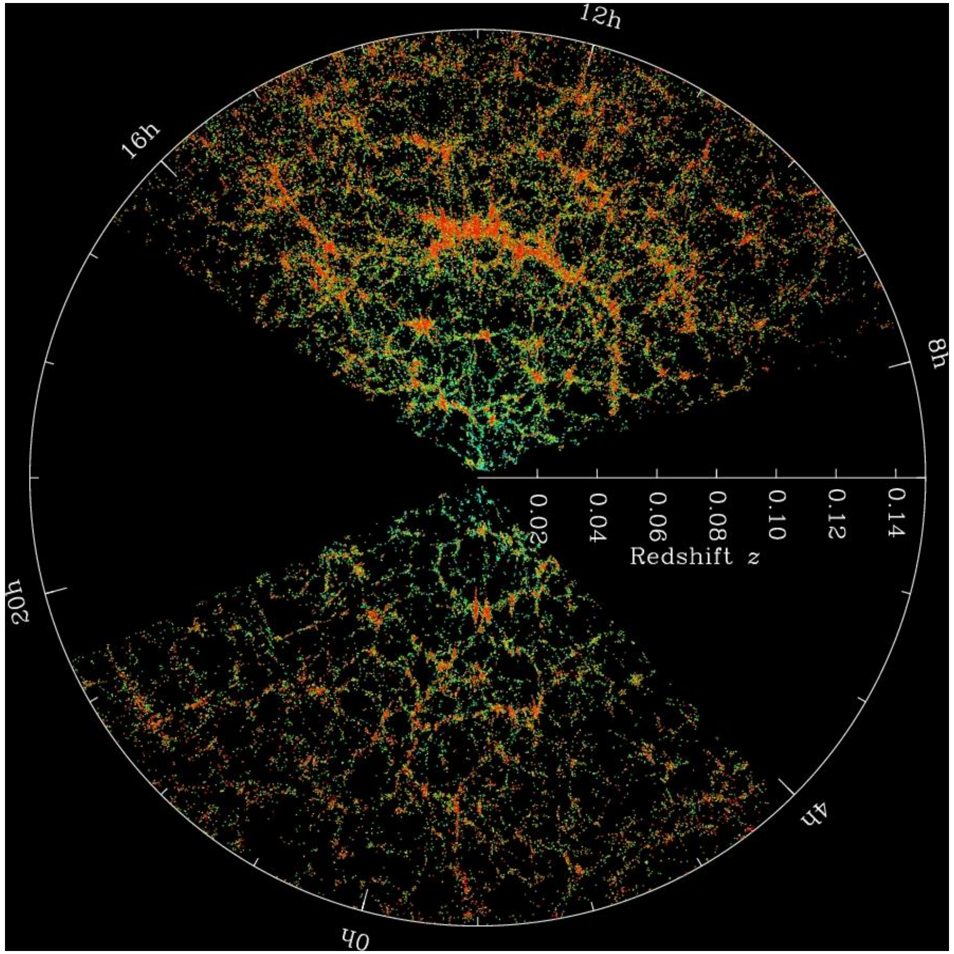


Credits: SDSS

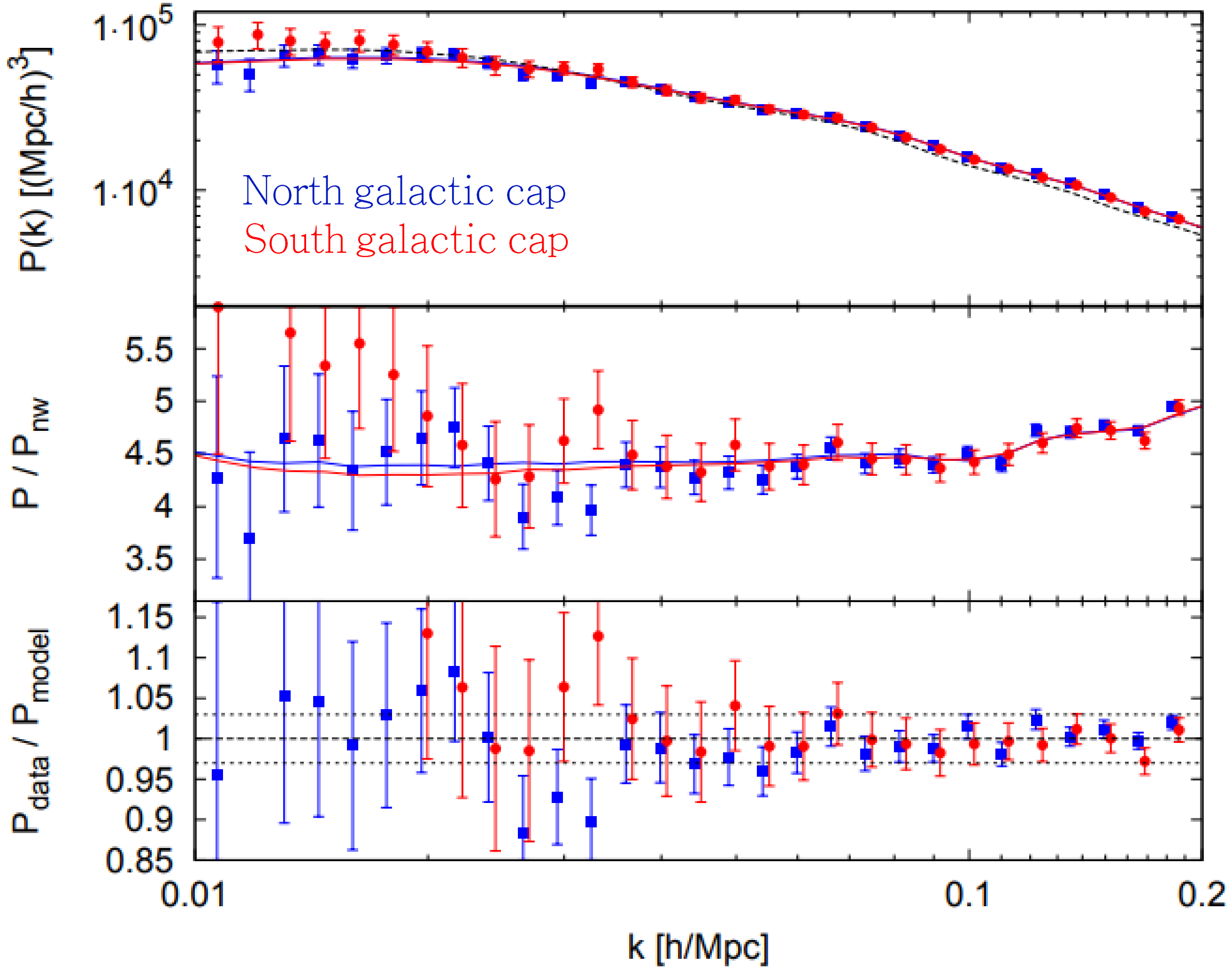


General introduction

Gil-Marin et al. 2015
(SDSS-BOSS paper)



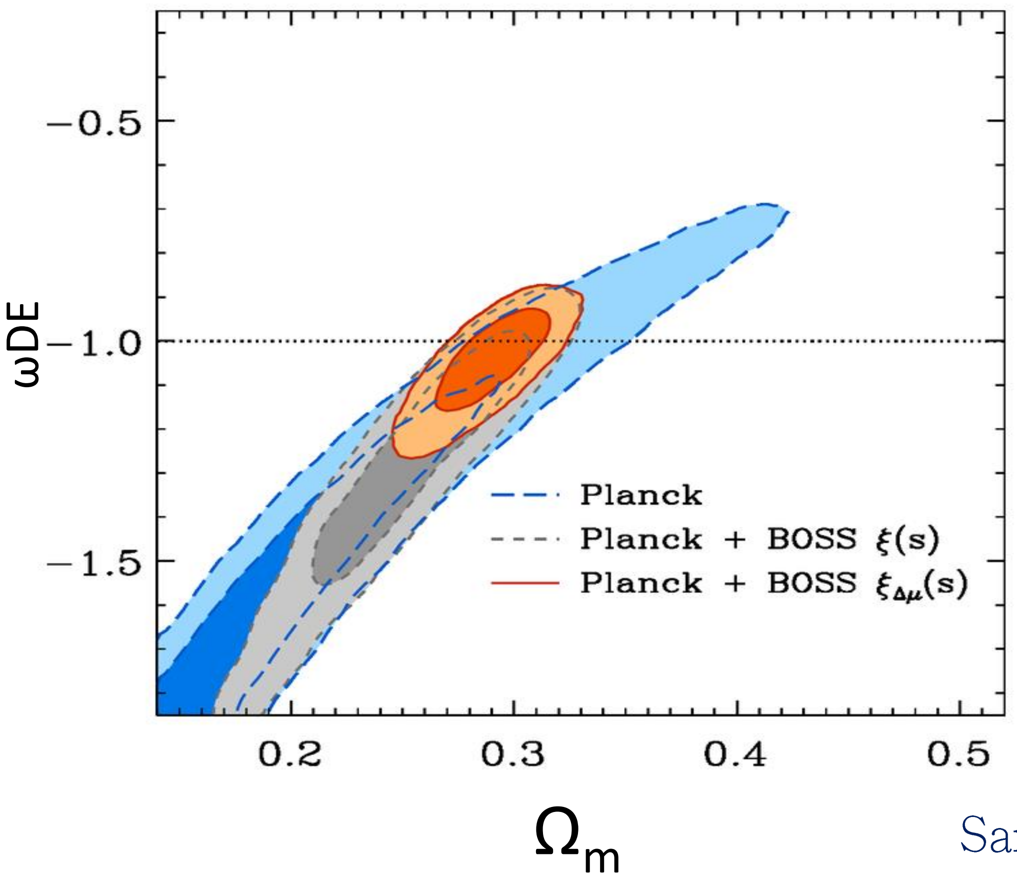
Credits: SDSS



Error bars: dispersion among 2048 simulated galaxy catalogues

General introduction

An accurate modeling of the power spectrum covariance matrix is needed to constrain **cosmological parameters**



Sanchez et al. 2013

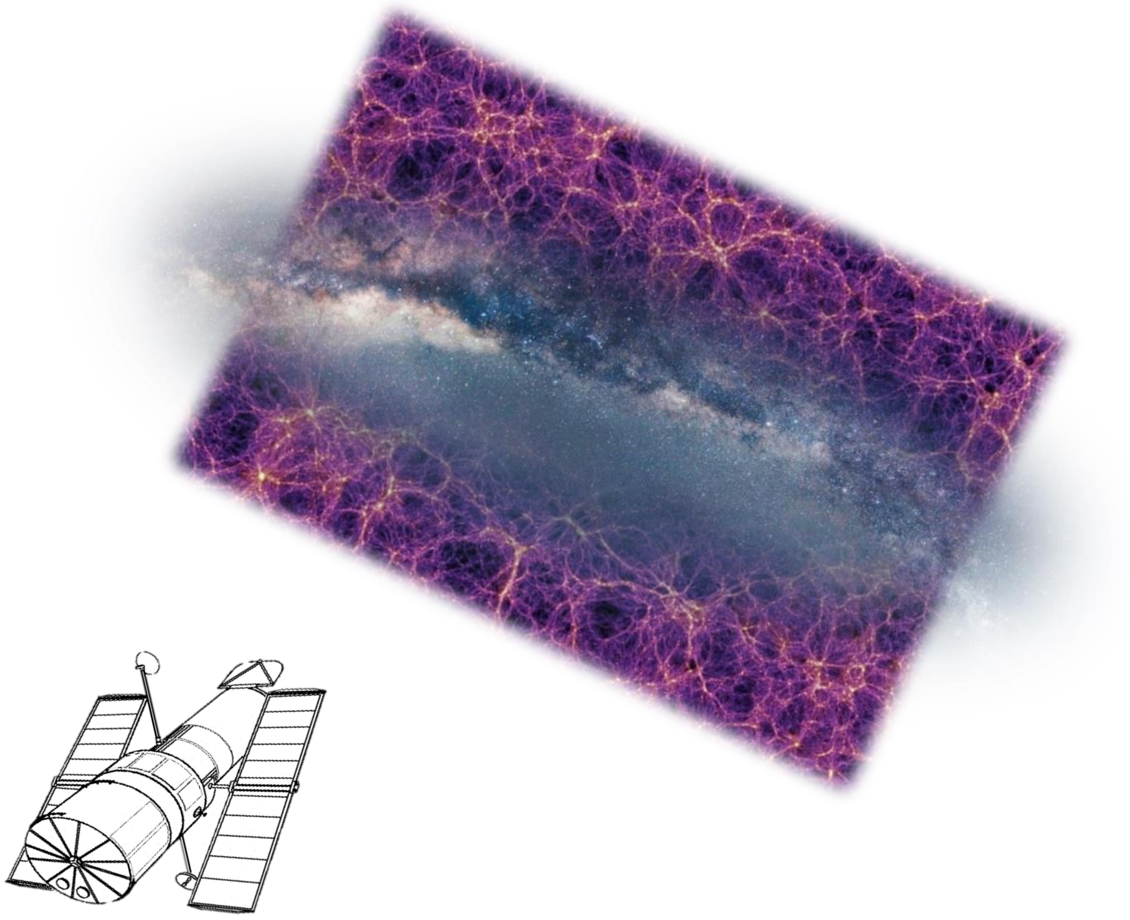
Uncertainty in the visibility mask of a survey and its effects on the clustering of biased tracers

M. Colavincenzo,[a] P. Monaco,[a,b,d] E. Sefusatti,[c,e] S. Borgani[a,b,d]

Published 27 March 2017 •
Journal of Cosmology and Astroparticle Physics, Volume 2017, March 2017

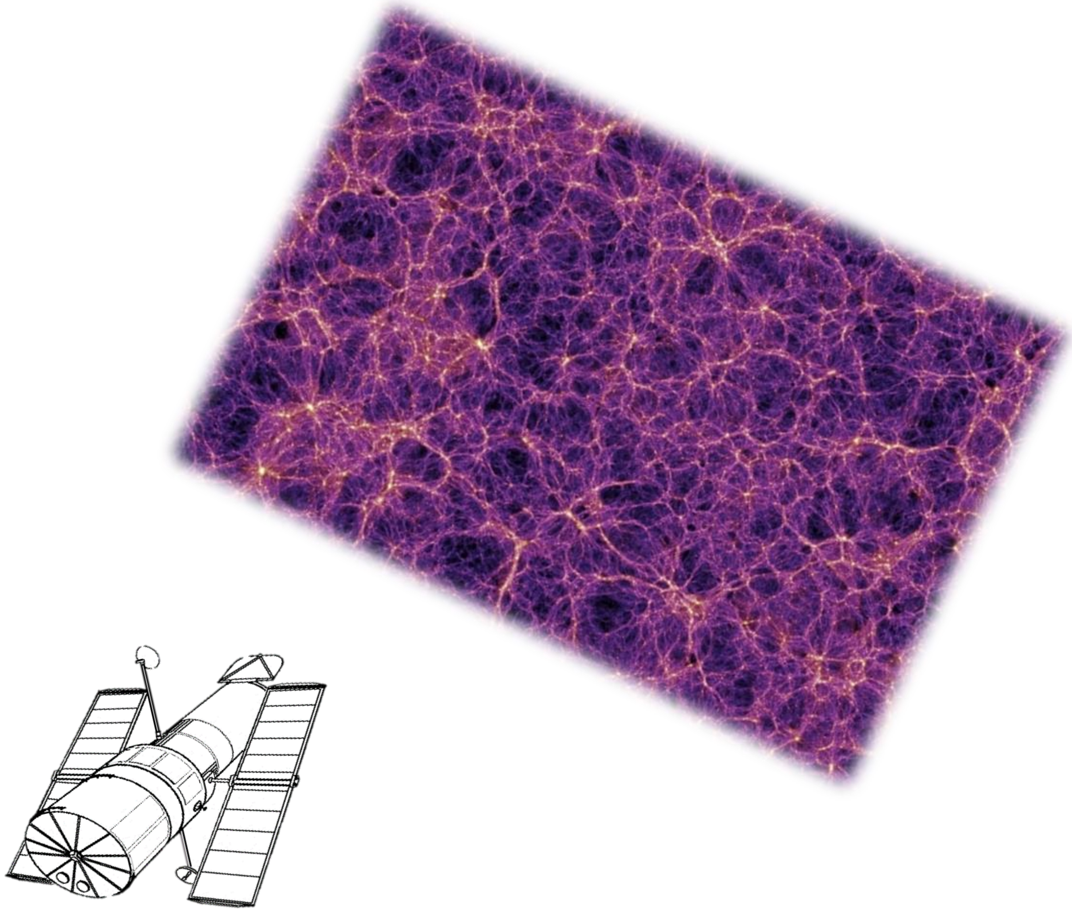
The foreground problem

The limiting magnitude of a galaxy sample is **modulated by foregrounds**



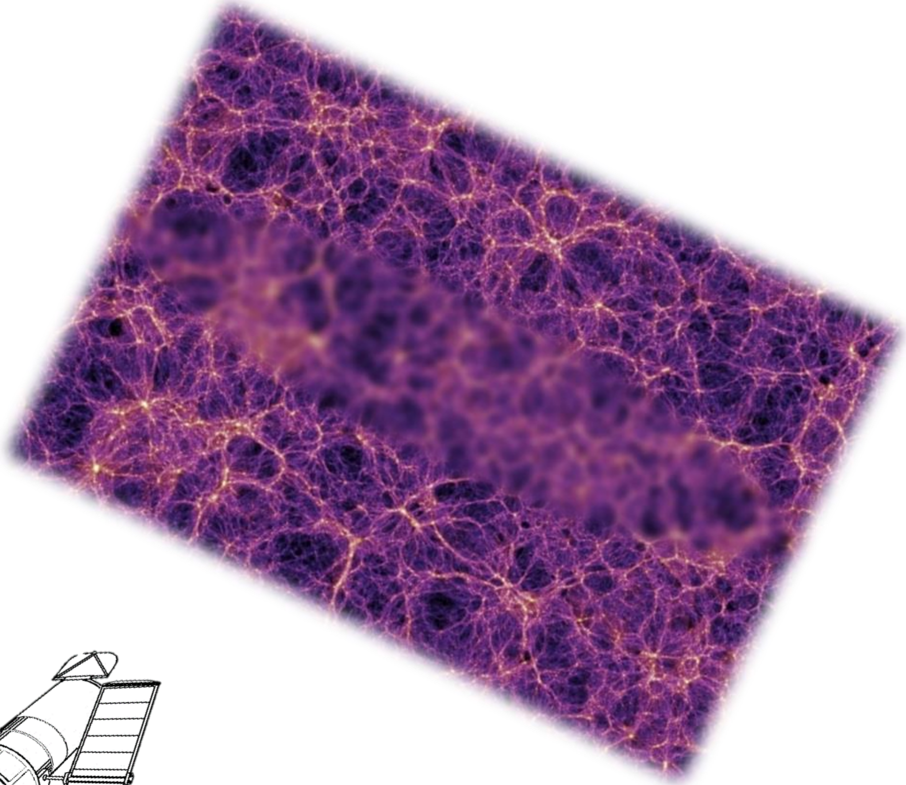
The foreground problem

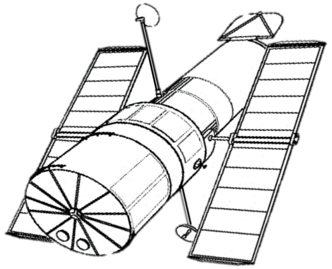
Cosmological information can be extracted after foreground subtraction



The foreground problem

This subtraction is subject to an uncertainty and the residual foreground will be correlated on the sky and will create structure on large scales





The foreground problem

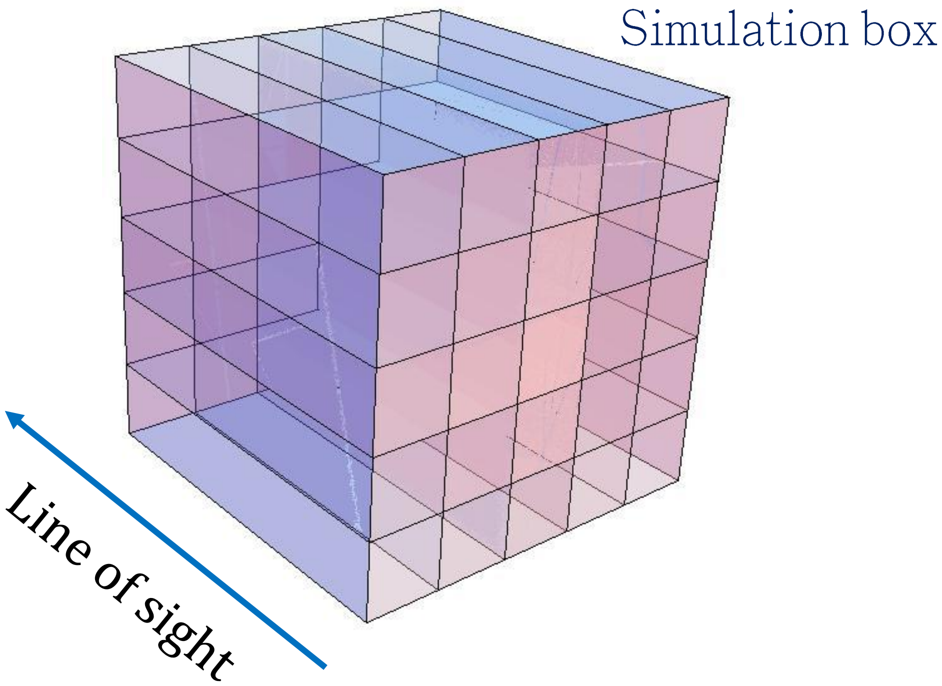
Why is it fundamental the foreground analysis?

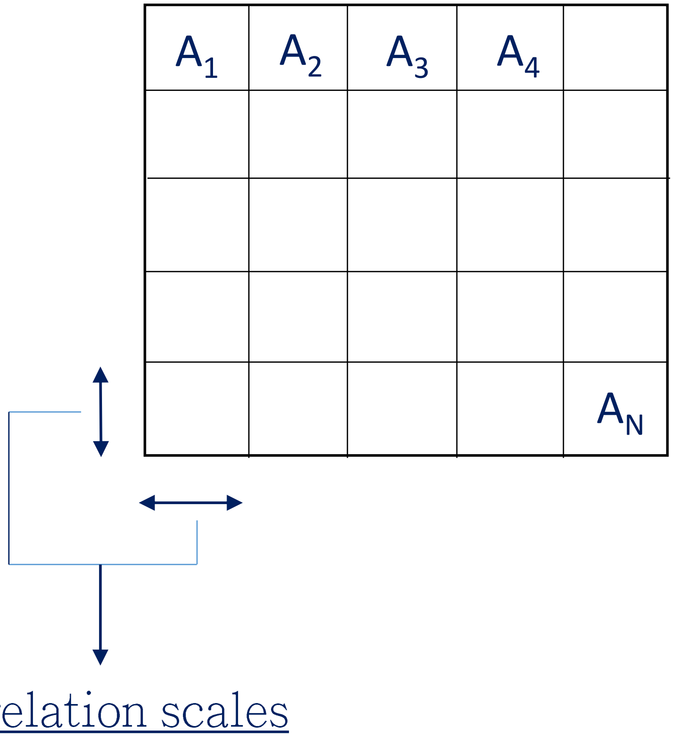
Future galaxy surveys will provide very high statistics -----> the error budget will be dominated by **systematics**

- Milky way extinction
- Instrumental and survey features
- …

The error on foreground subtraction must be properly propagated to correctly assess the error on cosmological parameters estimation.

The model for the "mask"





$$< A > = \frac{1}{N}\sum_{i=1}^{N} A_i = 0$$

$$< A^2 > = 0.01, 0.05, 0.1, 0.2$$

$$A(\boldsymbol{\theta}) \equiv \frac{\delta M(\boldsymbol{\theta})}{M_0} \longrightarrow \delta_{\rm mask}(\mathbf{x}) = -\frac{M_0\,\bar{\Phi}(M_0)}{\bar{n}} A(\boldsymbol{\theta}) + \mathcal{O}(A^2)$$

$$\sigma^2_{\rm mask} \simeq \frac{M_0^2 \bar{\Phi}^2(M_0)}{\bar{n}^2(M_0)} \sigma^2_A$$

Prediction for the PS covariance matrix

$$C_{ij} \equiv \mathrm{cov}[\hat{P}(k_i),\ \hat{P}(k_j)] = \langle \delta\hat{P}(k_i)\delta\hat{P}(k_j)\rangle$$

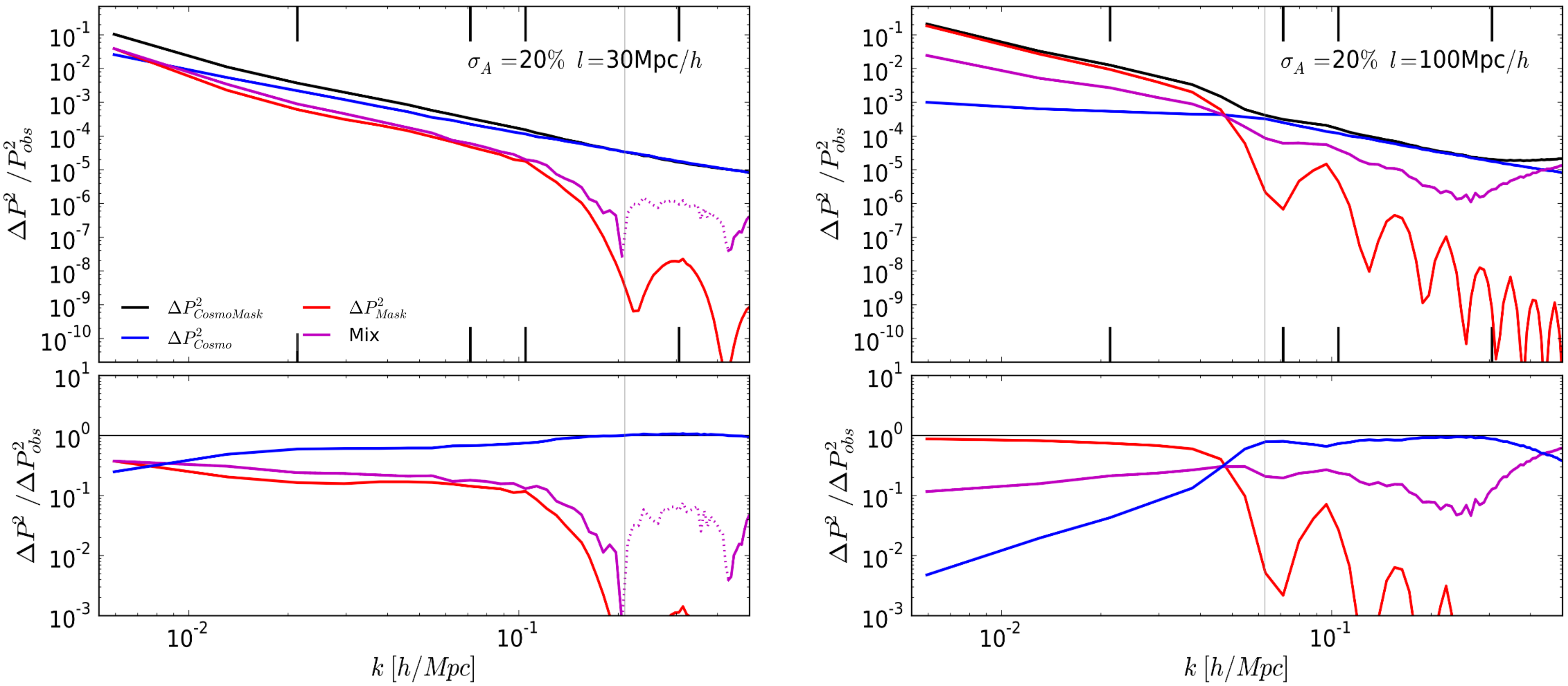
$$C_{ij}^{\mathrm{obs}} = C_{ij}^{cosm} + C_{ij}^{mask} + C_{ij}^{mixed}$$

$$\begin{aligned} C_{ij}^{mixed} = \ & \langle \hat{P}_{\mathrm{conv}}(k_i)\hat{P}_{\mathrm{conv}}(k_j)\rangle - \langle \hat{P}_{\mathrm{conv}}(k_i)\rangle\langle \hat{P}_{\mathrm{conv}}(k_j)\rangle + \\ & \langle \hat{P}_{\mathrm{cosmo}}(k_i)\hat{P}_{\mathrm{conv}}(k_j)\rangle - \langle \hat{P}_{\mathrm{cosmo}}(k_i)\rangle\langle \hat{P}_{\mathrm{conv}}(k_j)\rangle + \\ & \langle \hat{P}_{\mathrm{mask}}(k_i)\hat{P}_{\mathrm{conv}}(k_j)\rangle - \langle \hat{P}_{\mathrm{mask}}(k_i)\rangle\langle \hat{P}_{\mathrm{conv}}(k_j)\rangle + \\ & \langle \hat{P}_{\mathrm{cosmo}}(k_i)\hat{G}(k_j)\rangle + \langle \hat{P}_{\mathrm{mask}}(k_i)\hat{G}(k_j)\rangle + \\ & \langle \hat{P}_{\mathrm{conv}}(k_i)\hat{G}(k_j)\rangle + \langle \hat{G}(k_i)\hat{G}(k_j)\rangle \end{aligned}$$

$$\hat{G} = 2\delta_{\mathbf{q}}\delta_{\mathrm{mask},\mathbf{q}} - \delta_{\mathbf{q}}\delta_{\mathrm{conv},\mathbf{q}} + \delta_{\mathrm{mask},\mathbf{q}}\delta_{\mathrm{conv},\mathbf{q}}$$

Results: Variance

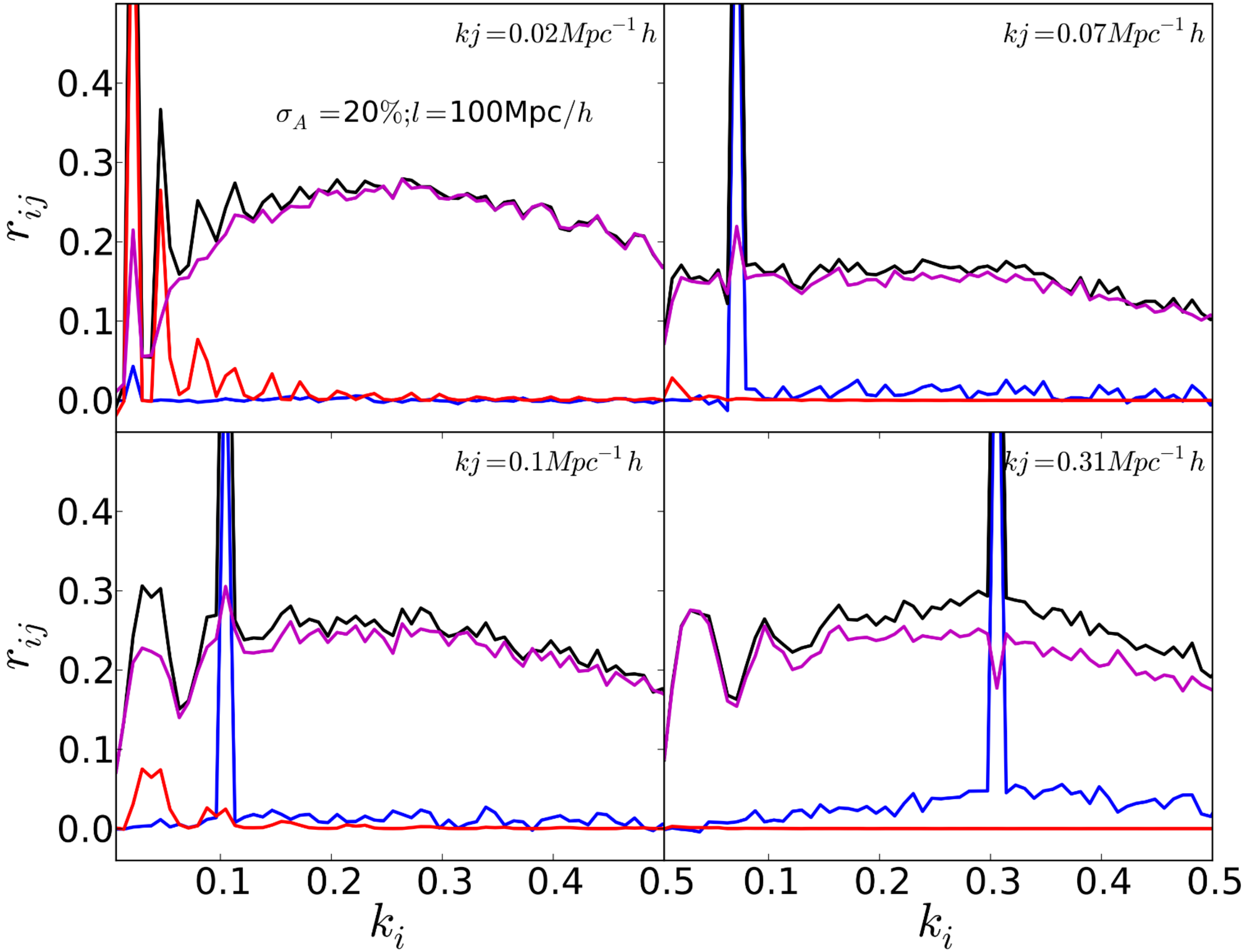
10 000 simulated galaxy catalogs



Results: Covariance

10 000 simulated galaxy catalogs

$$\frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}} = \frac{\mathrm{cov}_{ij}}{\sqrt{\mathrm{var}_i \mathrm{var}_j}}$$



Results

The error in the subtraction of the foreground is relevant and may be not negligible.

The convolution term, in the power spectrum, enters in the covariance mixing the modes and transferring power from large to small scales.

Non negligible off-diagonal terms could be enhanced by the coupling of the mask and the cosmological power spectrum

Results

The error in the subtraction of the foreground is relevant and may be not negligible.

The convolution term, in the power spectrum, enters in the covariance mixing the modes and transferring power from large to small scales.

Non negligible off-diagonal terms could be enhanced by the coupling of the mask and the cosmological power spectrum

Next steps

Study the effects using more realistic foreground (extinction maps, zodiacal light maps, ...)

Covariance matrix comparison

M. Colavincenzo & L. Blot, M. Crocce, P.Monaco, A. Sanchez ,E. Sefusatti, et al. (in preparation)

The approximate methods

Data

Covariance matrix

Model

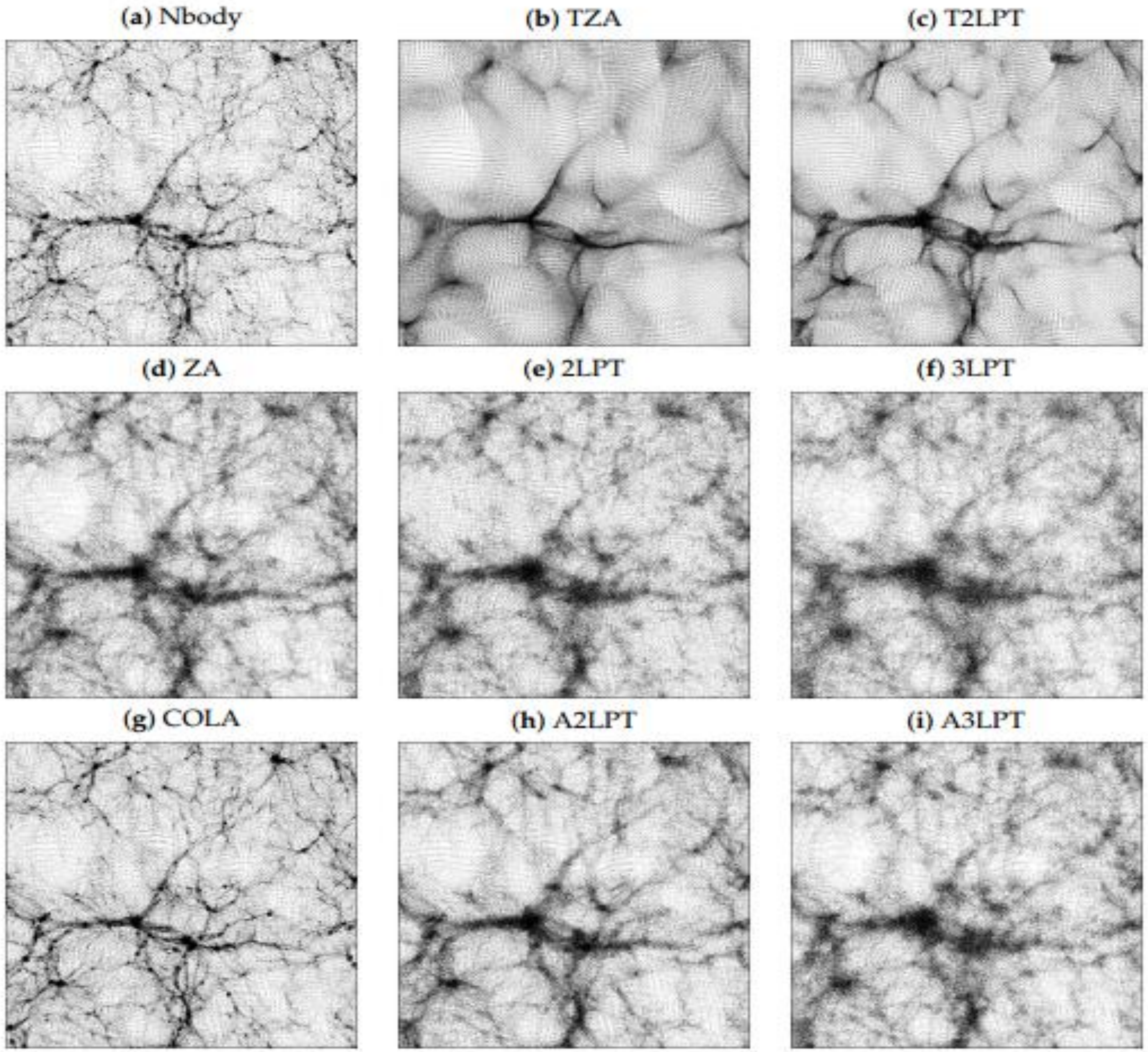
$$\mathcal{L}(\boldsymbol{D}|\boldsymbol{\theta},\boldsymbol{C}) \propto |\mathrm{C}|^{\frac{1}{2}} \exp\left[-\frac{1}{2}\sum_{i,j}(D_i - T_i(\boldsymbol{\theta}) C_{ij}^{-1}(D_j - T_j(\boldsymbol{\theta}))\right]$$

Set of cosmological parameters

Inverse of the covariance matrix → Precision matrix

In most analyses of clustering measurements, the **precision matrix is estimated from an ensemble of simulated catalogues** → All estimates of the precision matrix based on a finite number of catalogues are affected by **noise**. → A **large number of simulated catalogues** might be necessary in order to keep the **uncertainty under control**

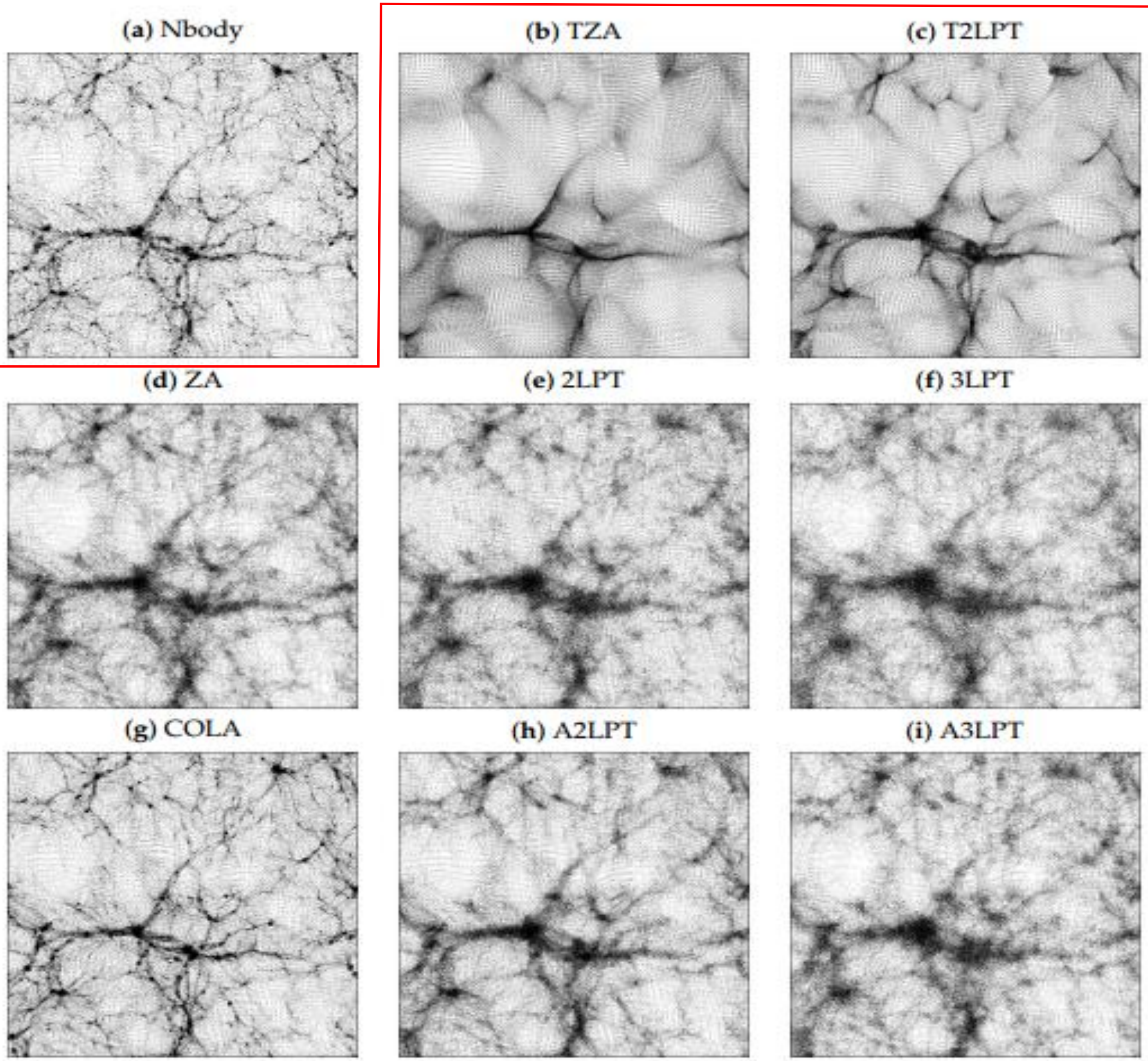
The approximate methods



N-Body simulations: calculate the **non-linear growth** of structures in the Universe by following the trajectories of N particles interacting between each others through gravity

Approximate methods: take advantage of **analytic approximations** to produce a large number of galaxy catalogues reducing the computing time of a factor ~ 1000, paying a loss in accuracy of the small scales ($k \sim 0.5\ h^{-1}$Mpc)

The approximate methods



N-Body simulations: calculate the **non-linear growth** of structures in the Universe by following the trajectories of N particles interacting between each others through gravity

Approximate methods: take advantage of **analytic approximations** to produce a large number of galaxy catalogues reducing the computing time of a factor ~ 1000, paying a loss in accuracy of the small scales ($k \sim 0.5\ h^{-1}$Mpc)

The approximate methods

Main questions:

- Can we use approximate methods to estimate the covariance matrices of clustering measures?
- Do approximate methods reproduce the error on cosmological parameters?

Tools we will use to answer:

- Halo power spectrum (L. Blot et al.)
- **Bispectrum (M. Colavincenzo et al.)**
- Two-point correlation function (A. Sanchez et al.)

Approximate methods:

- Pinocchio
- Cola
- Halogen
- PeakPatch
- Patchy, …

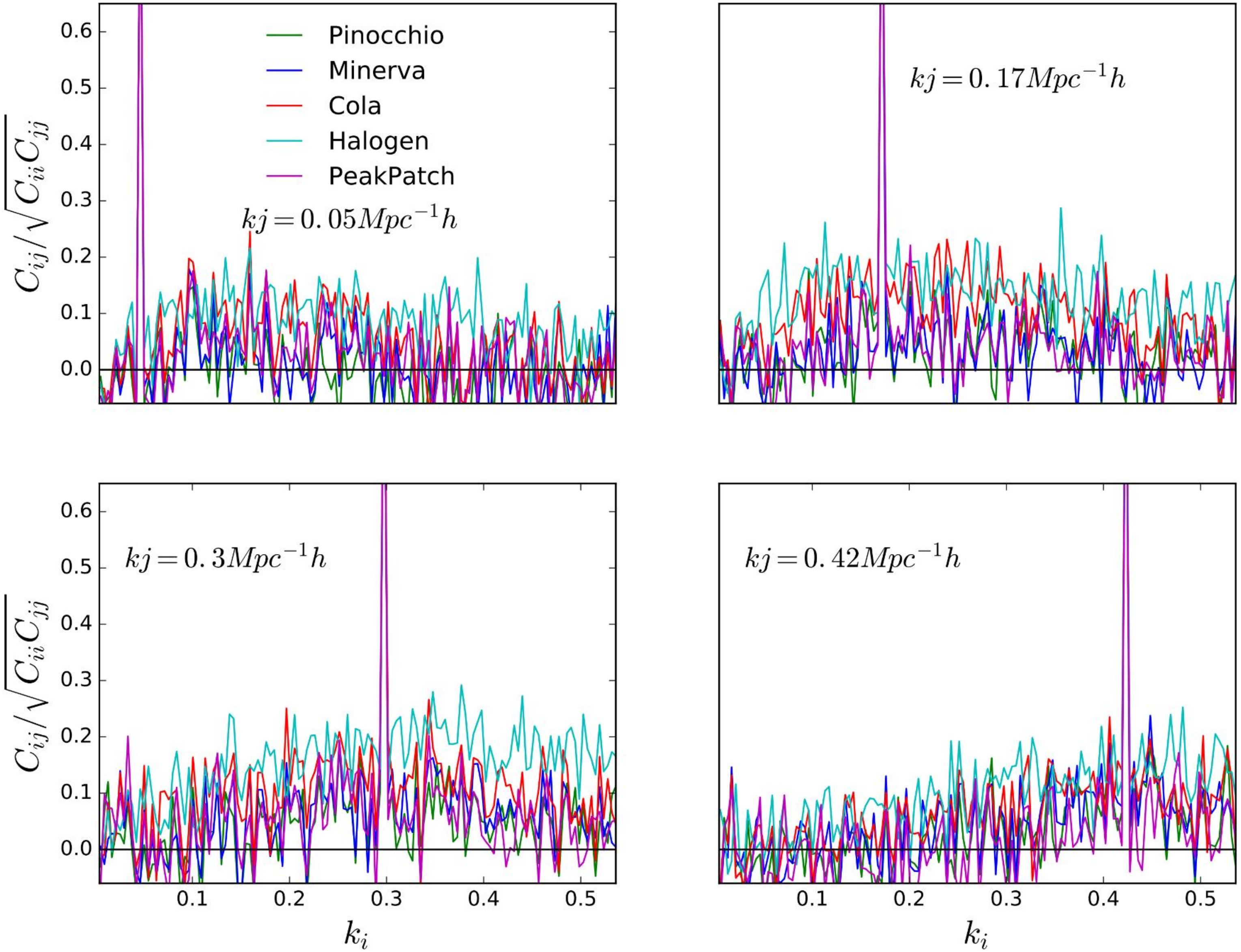
Reference simulations:

- N-body (Minerva) [Gadget-2]

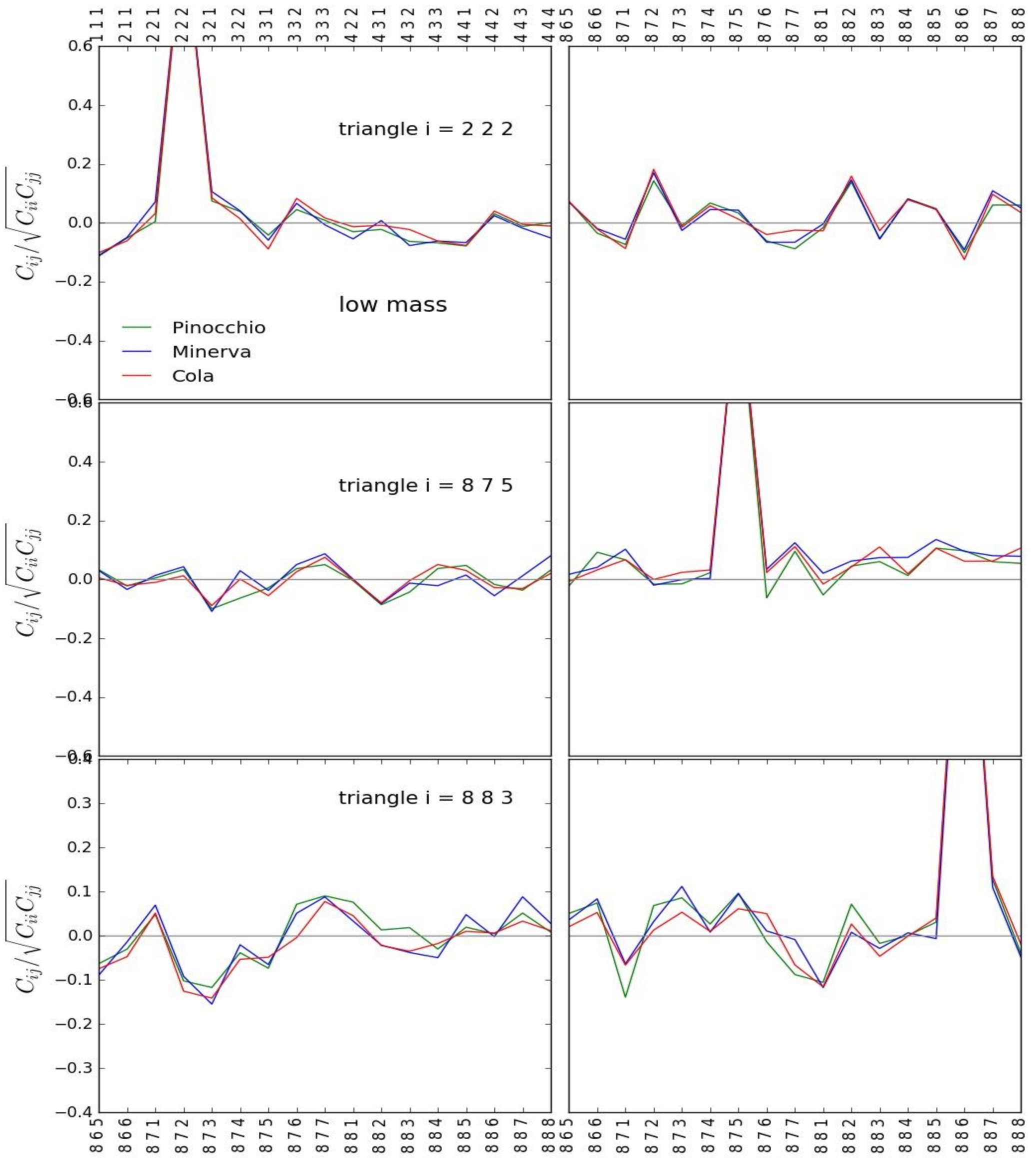
Simulation infos:

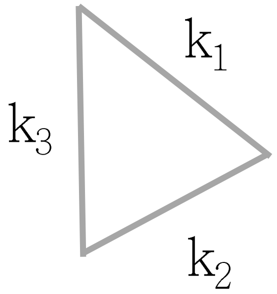
- Cubic box: 1500 Mpc/h
- $z = 1$
- Real and Redshift space
- **300 realizations**

Power spectrum analysis



Bispectrum analysis





Results

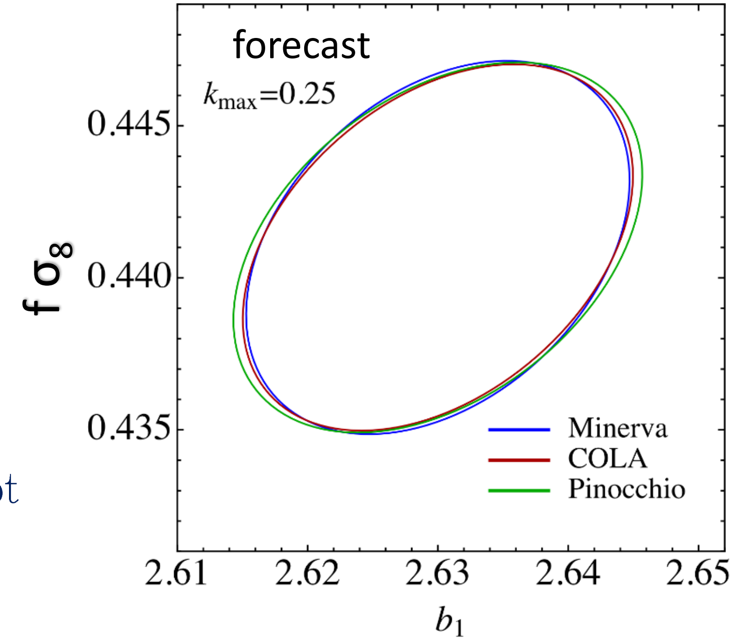
- Different approximate methods allow us to obtain the **same correlated noise** in the power spectrum and bispectrum covariance matrix

- We expect the physical signal is the same

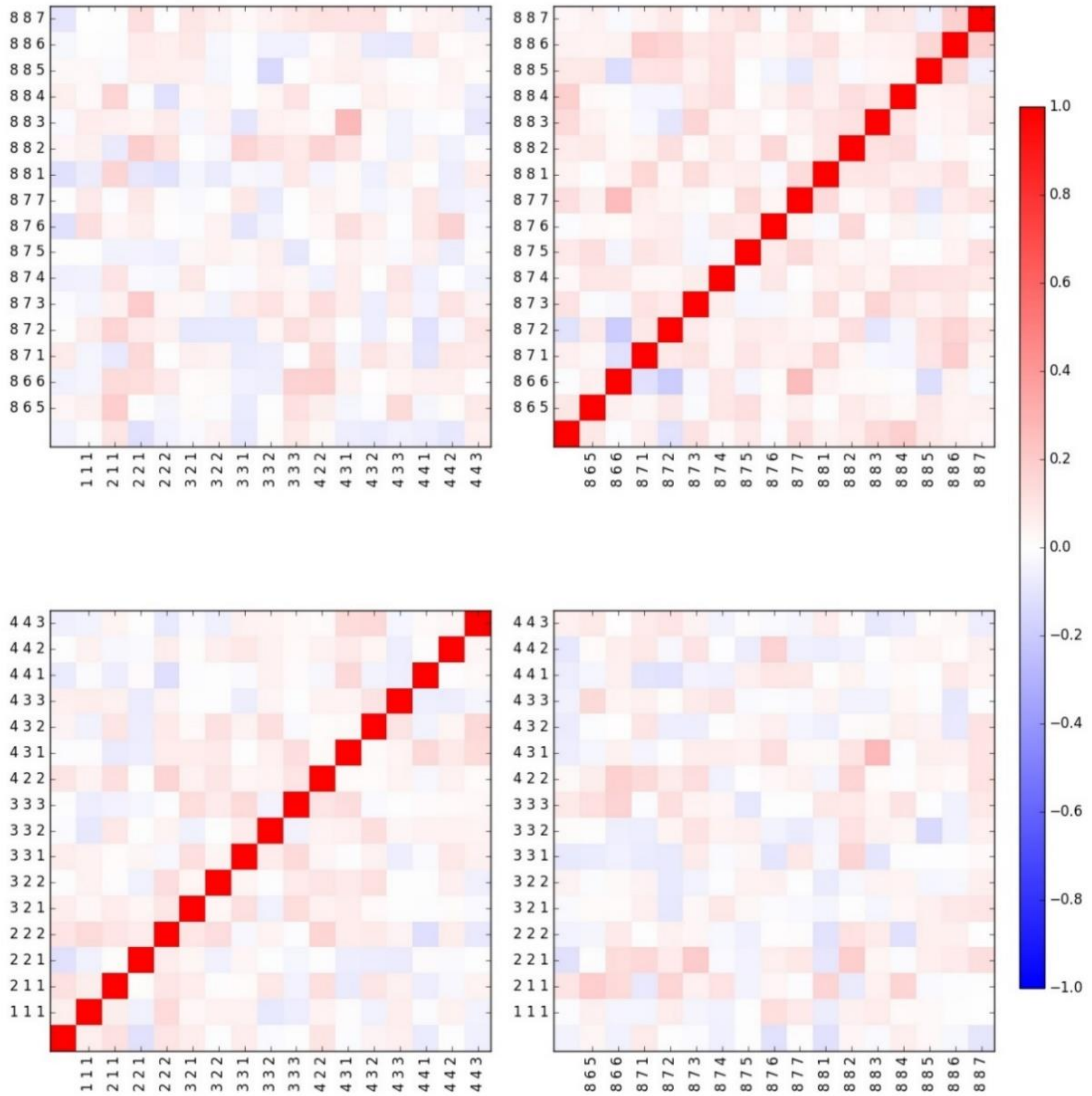
Preliminary

MCMC results - 3 multipoles - $k_{\mathrm{Max}} = 0.17\, h\, \mathrm{Mpc}^{-1}$					
	$f \times \sigma_8$	b_1	b_2	γ_3	a_{vir}
Minerva	0.444 ± 0.022	2.625 ± 0.015	2.669 ± 0.4	1.83 ± 0.134	1.14 ± 0.76
COLA	0.446 ± 0.023	2.627 ± 0.013	2.467 ± 0.348	1.86 ± 0.097	0.80 ± 0.27
Pinocchio	0.451 ± 0.019	2.627 ± 0.012	2.326 ± 0.254	1.87 ± 0.086	0.63 ± 0.35

Credits: Linda Blot

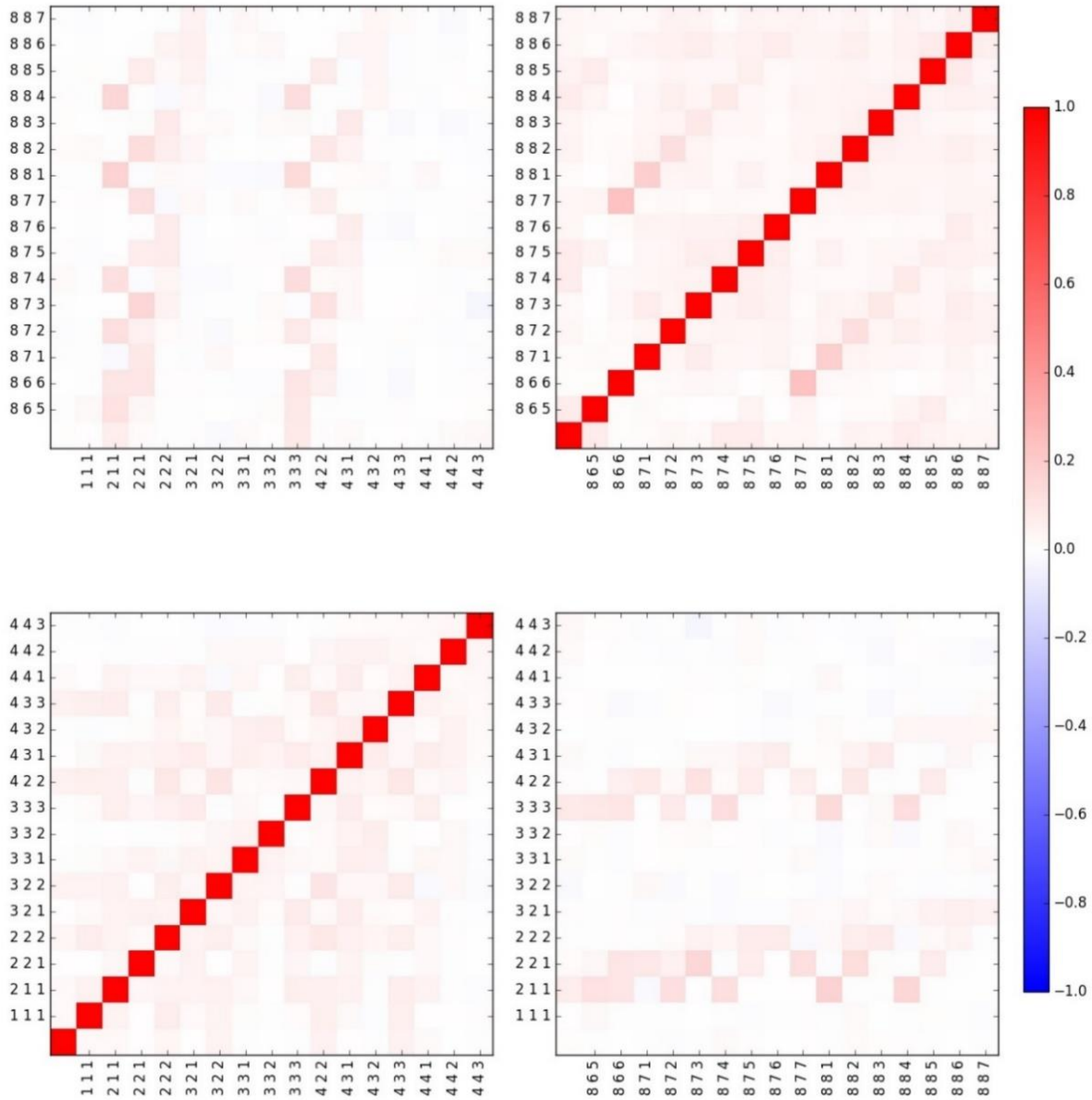


$$\frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}} = \frac{\mathrm{cov}_{ij}}{\sqrt{\mathrm{var}_i \mathrm{var}_j}}$$



300 N-body simulations

$$\frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}} = \frac{\mathrm{cov}_{ij}}{\sqrt{\mathrm{var}_i \mathrm{var}_j}}$$



10 000

PINOCCHIO

realizations