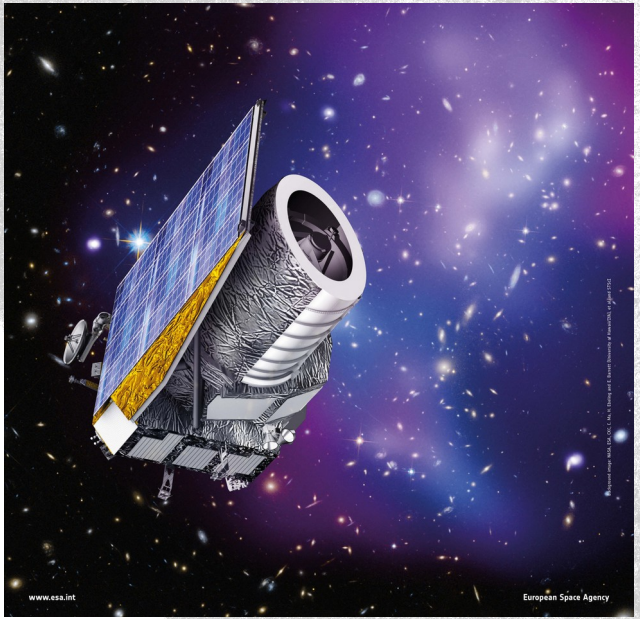
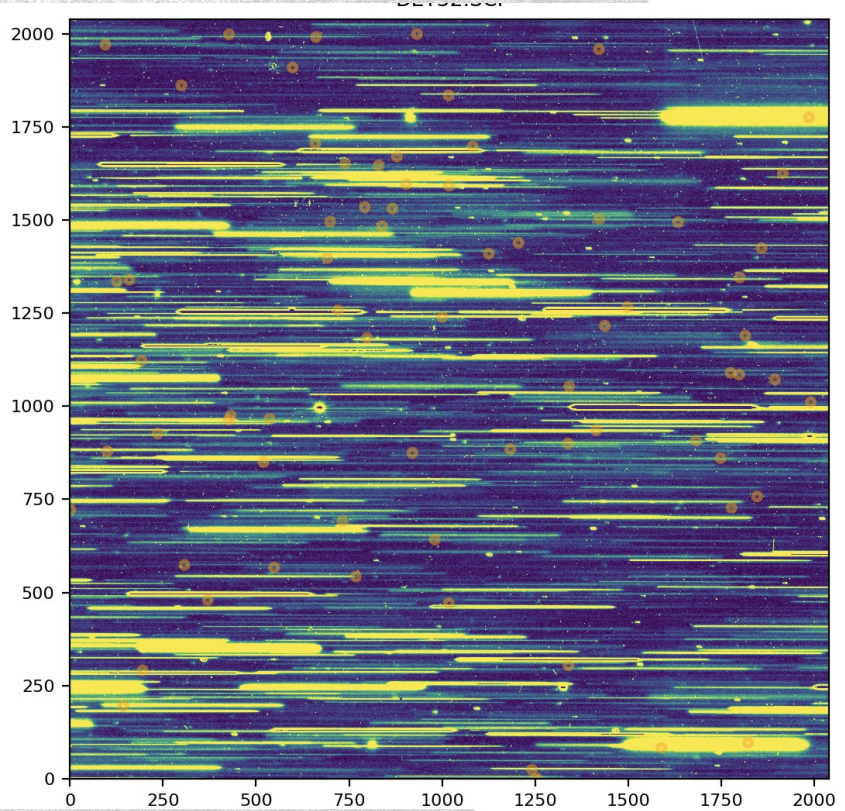
Mocking the Euclid spectroscopic survey

P. Monaco, Universita` di Trieste, INAF-OATs, INFN, IFPU

Role in Euclid:
co-lead of Observational Systematics WP, Galaxy Clustering SWG
co-lead of Simulations for Galaxy Clustering WP, CosmoSim SWG
lead of covariance Processing Functions for OU-LE3



credit: Ben Granett + OU-SIM

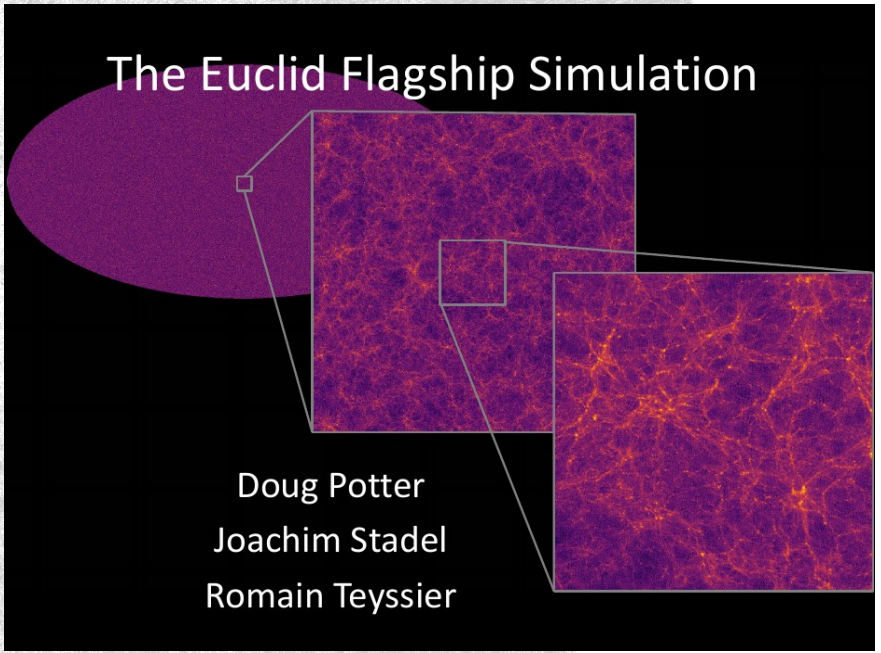


Mocking the Euclid spectroscopic survey

P. Monaco, Universita` di Trieste, INAF-OATs, INFN, IFPU

Role in Euclid:
co-lead of Observational Systematics WP, Galaxy Clustering SWG
co-lead of Simulations for Galaxy Clustering WP, CosmoSim SWG
lead of covariance Processing Functions for OU-LE3

Requirements for a simulation:
Sampling BAO scales with subpercent precision:
box size > 1 Gpc/h
Past light cone to z~2 with no (few) replications:
box size >~ 4 Gpc/h
Sampling the largest scales (PNGs, SSC etc.):
box size >> 1 Gpc/h
Sampling halos for Euclid spectroscopic sample:
M_{part} <~ 3e10 M_{sun}/h
Sampling halos for Euclid photometric sample:
M_{part} <~3e9 M_{sun}/h
-> at least ~6000^3 particles realizations



The number of needed mocks for the covariance matrix

Noise in the CM due to sampling with a finite number of mocks propagates to the parameter covariance (errorbars) as a function of:

- N_b lenght of the data vector (number of bins),
- N_p number of parameters
- N_s number of simulations (mocks)

$$\langle \Delta\theta_i \Delta\theta_j \rangle = f \langle \Delta\theta_i \Delta\theta_j \rangle_{\rm ideal}$$

$$f = 1 + B\,(N_{\rm b} - N_{\rm p})$$

$$B = \frac{(N_{\rm s} - N_{\rm b} - 2)}{(N_{\rm s} - N_{\rm b} - 1)(N_{\rm s} - N_{\rm b} - 4)}$$

$N_{\rm b}$ = Number of bins in **D**

$N_{\rm p}$ = number of parameters

(credits: A. Sanchez, L. Blot)

In Euclid we aim at **3500 simulations**, good for two cases:

- N_b = 600 with required 10% precision in the errorbar,
- N_b = 300 with required 5% precision in the errorbar.

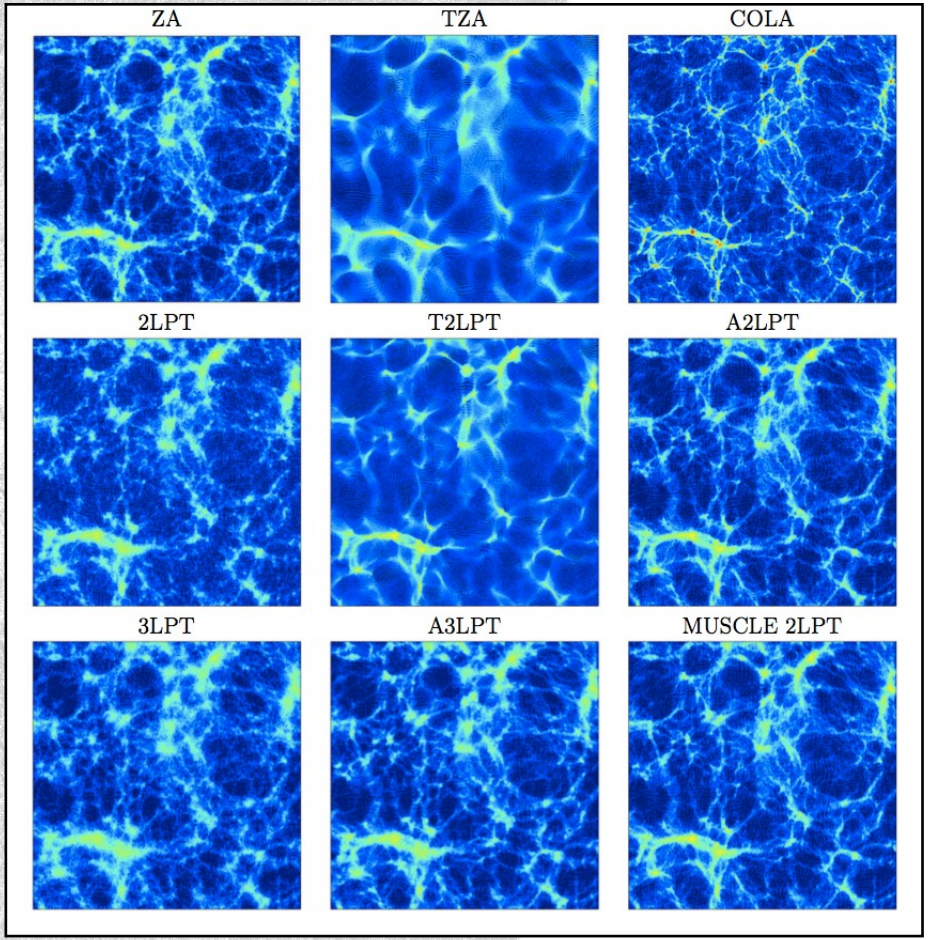
Advanced techniques (tapering, eigenvalue decomposition, shrinkage, fitting a model for the CM, denoising with ML...) will weaken this constraint,
higher orders or cross correlation will increase the data vector.

Approximate methods

Several methods have been proposed to generate catalogs of DM halos with approximate methods (or fast simulations). These can be divided into two broad classes:

- **Lagrangian**, or predictive (peak-patch, pinocchio, cola, fast-pm...)
- **Bias-based**, or calibrated (log-normal, patchy etc., ez-mocks, **BAM**...)

Criterion for validation:
the relative change of parameter errorbars induced by the use of approximate methods should be <10%



P.M., 2016, *Galaxies* 4, 53

PINOCCHIO

PINpointing Orbit Crossing-Collapsed HIerarchical Objects (Monaco+ 2002 ... Munari+2017)

Computing the "collapse" (OC) time of a fluid element

- Taylor expansion of the gravitational potential:

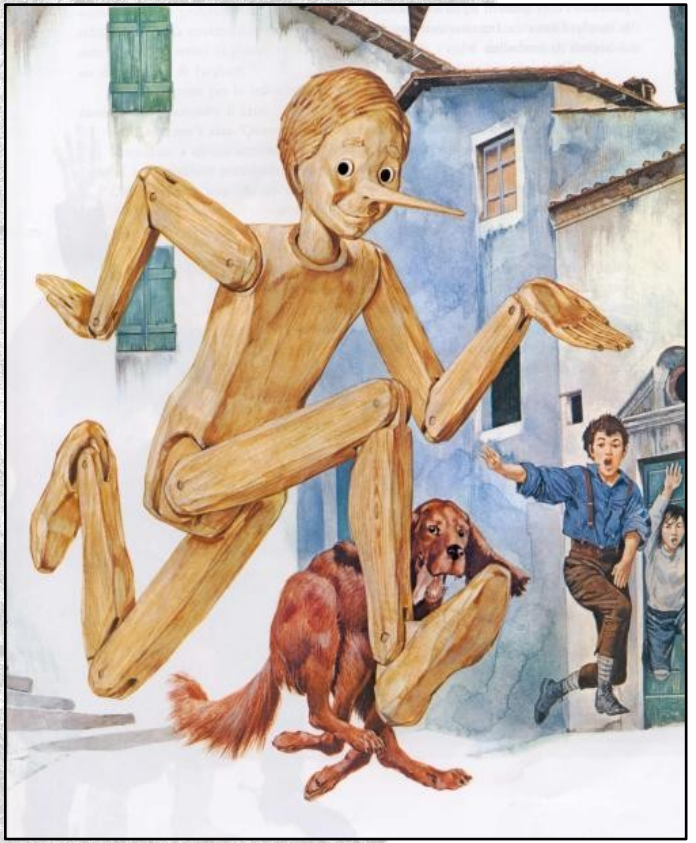
$$\Phi(\vec{q}_0) \simeq \cancel{\Phi_0} + \Phi_{,i}(\vec{q}_0)(\vec{q}-\vec{q}_0)_i + \Phi_{,ij}(\vec{q}_0)(\vec{q}-\vec{q}_0)_{ij}$$

bulk flow — second-order term

- => evolution of a homogeneous ellipsoid
  - numerical solution, or
  - solution with 3LPT up to **ORBIT CROSSING**

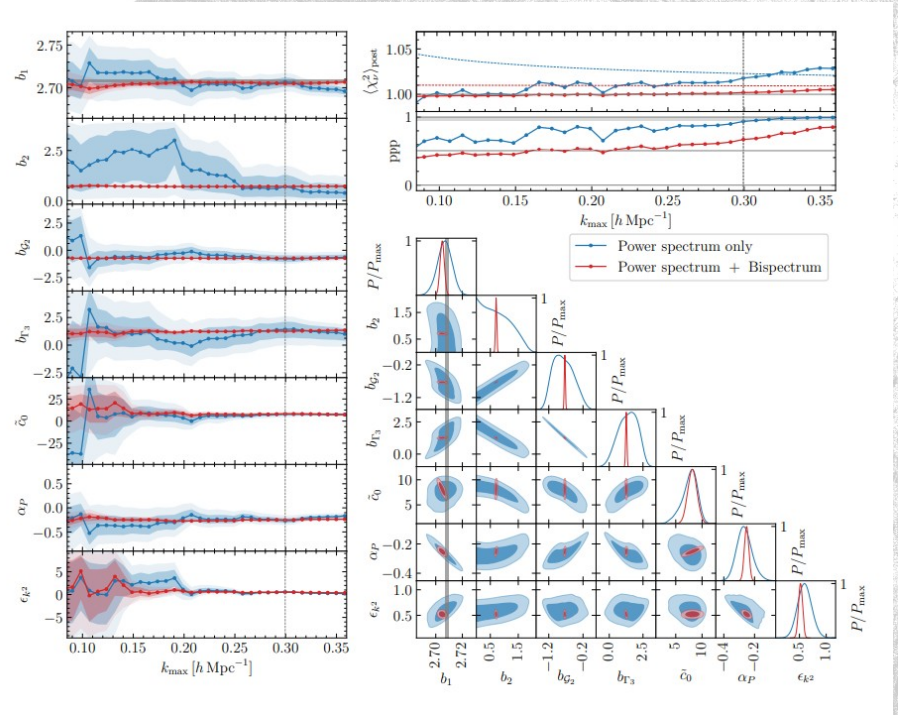
  + correction for quasi-spherical cases

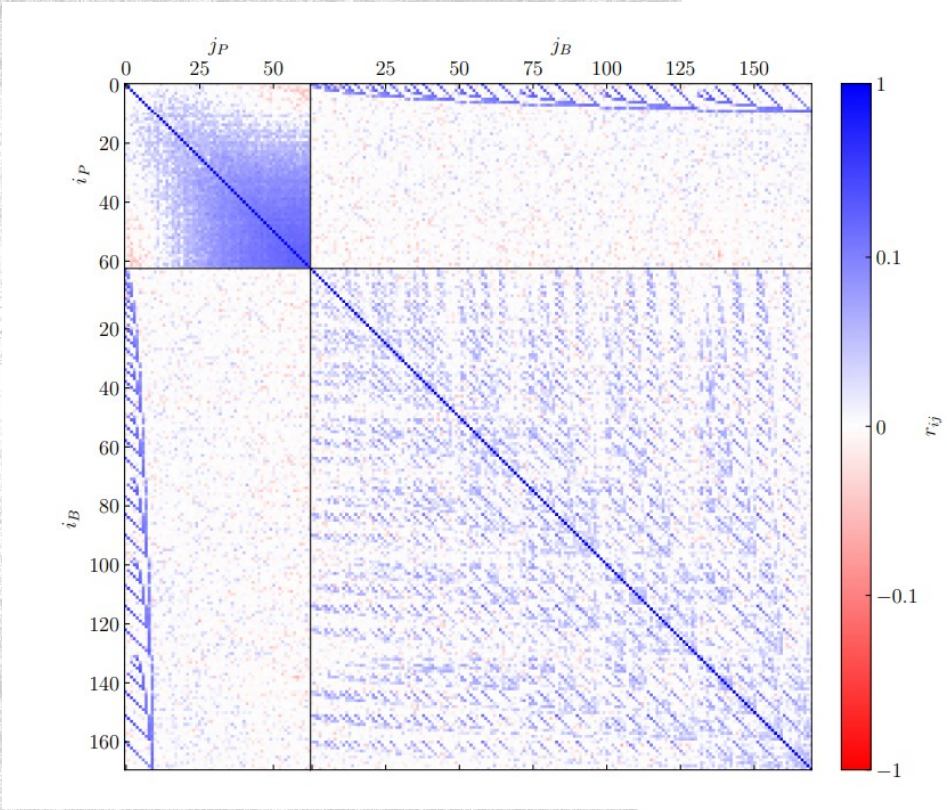
(P.M. 1995, 1997a)



Covariances for a P(k) + B(k) analysis

Oddo+2021 used 10000 Pinocchio simulations in periodic boxes to construct a CM for fitting P(k) and B(k) in real space, demonstrating the power of the joint analysis to constrain bias parameters.



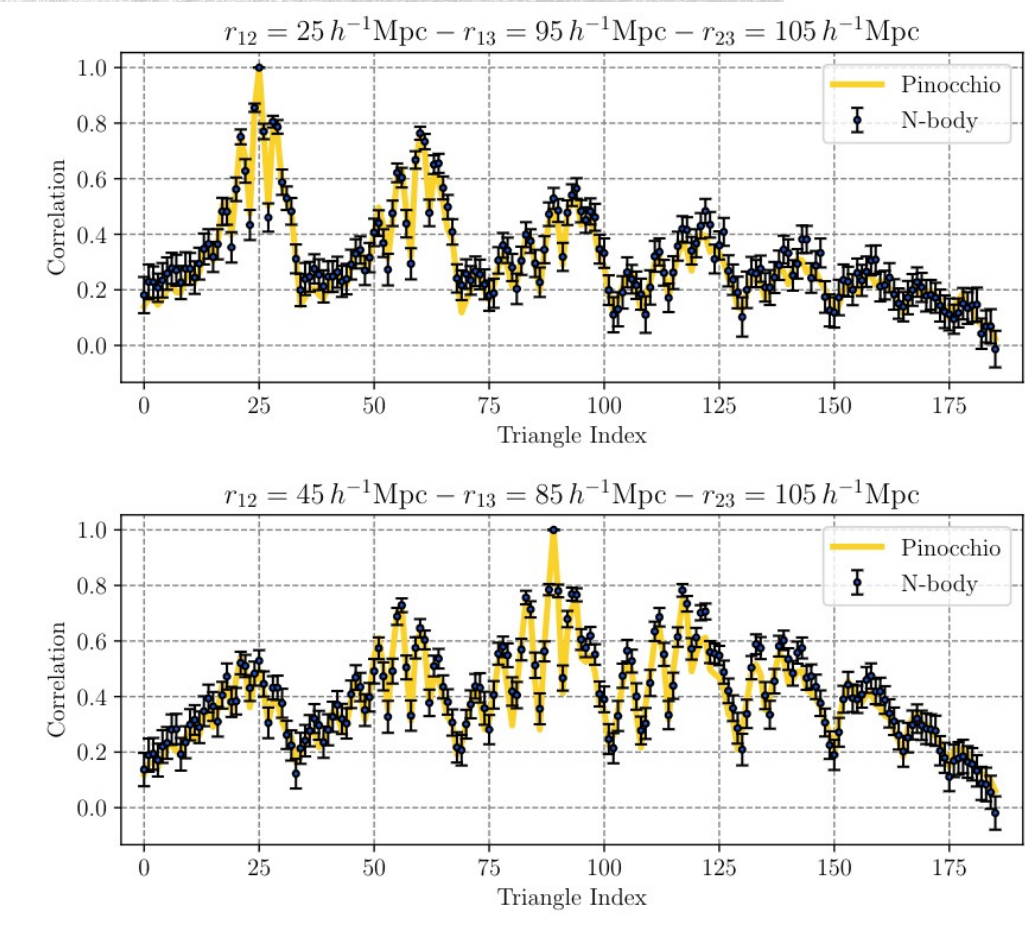


Covariances for a 3PCF analysis

Veropalumbo+2022 (in prep.) how the CM constructed from 298 N-body simulations and 298 pinocchio simulations run on the same initial conditions compare. Here the sampling noise is exactly the same for the two sets, so the comparison is expected to be better than the (jackknife) errorbars.

Reason of success:

- 3LPT displacements (that represent the 4-point function) applied to sets of particles in Lagrangian space that approximate well halo particles,
- i.e. it's better than 1-loop PT!

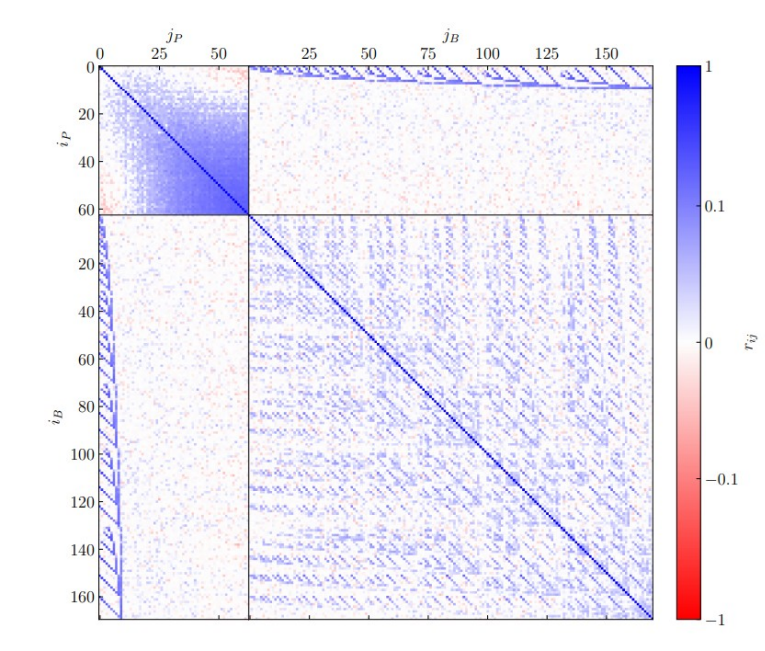


Available mocks

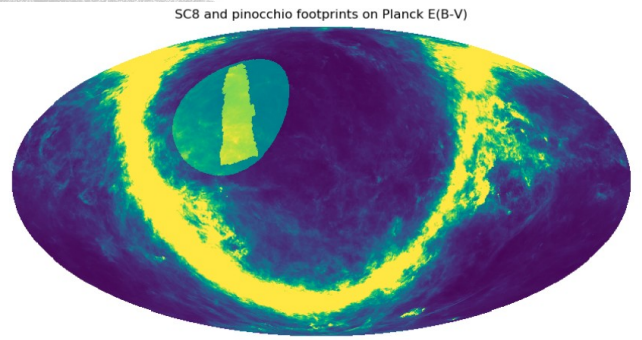
Minerva-like (ThirdEuclidMocks)

Simulation performed with PINOCCHIO v4.1.3:

- WMAP-like cosmology
- 1500 Mpc/h box sampled with 1000^3 particles
- M_p=3.8e11 M_{sun}/h, smallest halo M_h=1.15e13 M_{sun}/h (30 part.)
- **10000 realizations avaliable**
- no lightcones
- outputs at z=2, 1.8, 1.35, 1, 0.9, 0.57, 0,32 and 0
- storage: 8.3 TB
- used in many papers



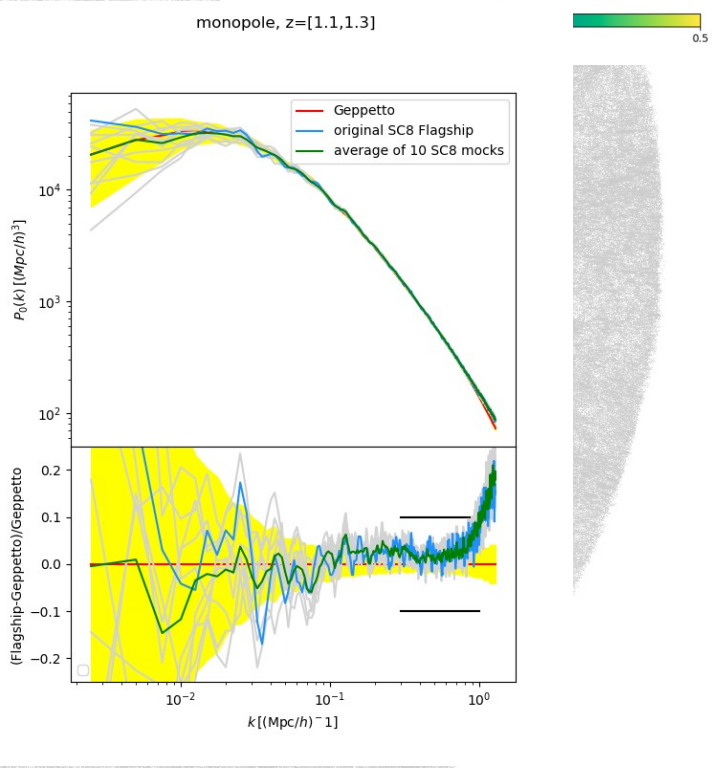
Available mocks



GeppettoFC

Simulation performed with PINOCCHIO v4.1.3:

- LambdaCDM cosmology similar to Flagship 1
- 1200 Mpc/h box sampled with 2160^3 particles
- M_p=1.5e10 M_{sun}/h, smallest halo M_h=1.5e11 M_{sun}/h (10 part.)
- **600 realizations available, 3500 planned**
- light-cone covering a circle of radius 30 deg, 2763 sq deg, starting at z=2 (containing SC8 wide field)
- outputs at z=2, 1.8, 1.35, 1, 0.9, 0.5 and 0 + lightcone + histories
- storage: 37 Gb each (17 Gb for the lightcone)
- used for SC8 mocks

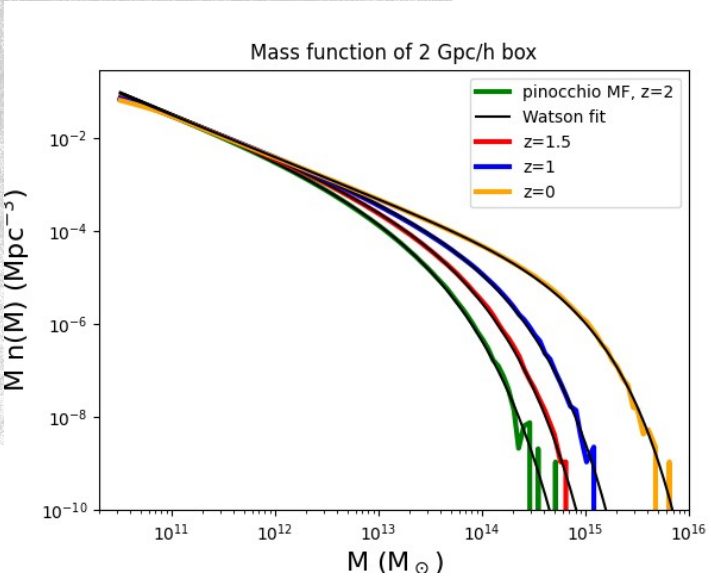


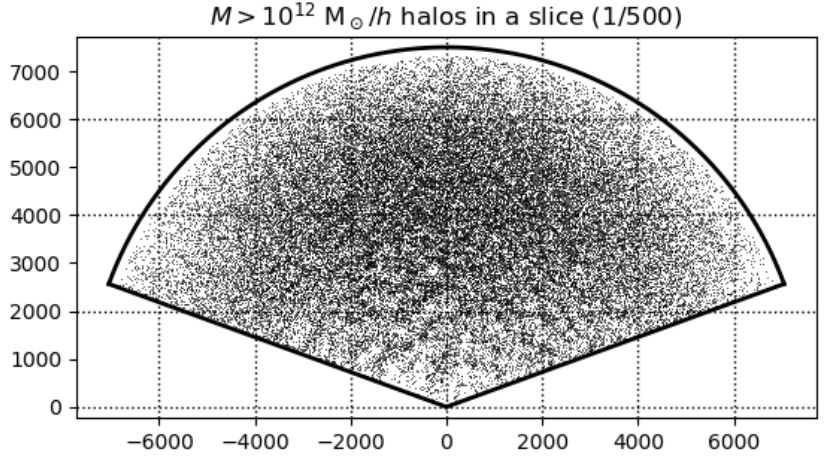
Available mocks

EuclidMocks2000

Simulation performed with PINOCCHIO v5.0:

- LambdaCDM cosmology similar to Flagship 1
- 2000 Mpc/h box sampled with 6144^3 particles
- M_p=3e9 M_{sun}/h, smallest halo M_h=3e10 M_{sun}/h (10 part.), same as the Flagship 1 simulation
- **50 realizations, target: 500**
- light-cone covering ~half sky starting at z=4
- central l.o.s. aligned with the z axis
- 2 outputs (z=1 and z=0) + lightcone + histories,
- storage: 573 Gb (450 Gb for the lightcone)
- issue with replications?





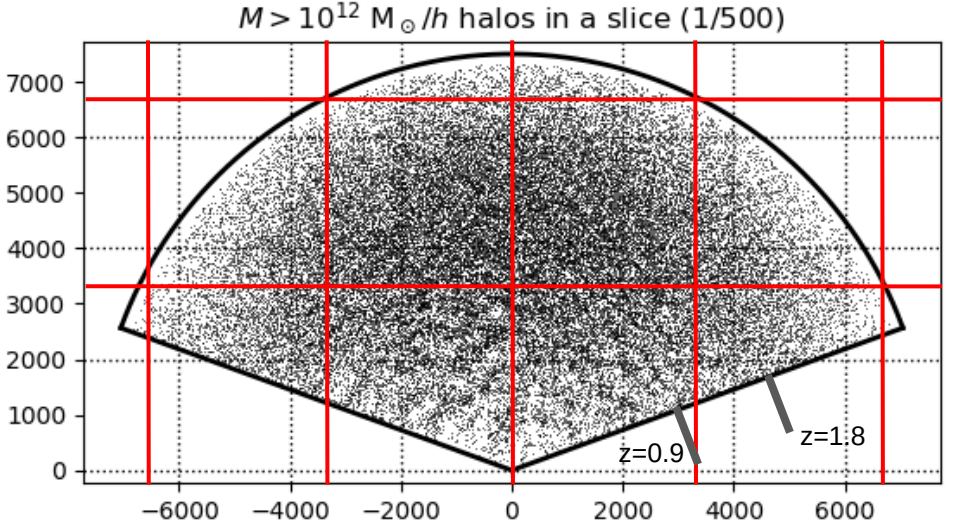
computer time provided by **INFN-Euclid**

Available mocks

EuclidMocks4000

Simulation performed with PINOCCHIO v5.0:

- LambdaCDM cosmology similar to Flagship 1
- 3380 Mpc/h box sampled with 6144^3 particles
- M_p=1.5e10 M_{sun}/h, smallest halo M_h=1.5e11 M_{sun}/h (10 part.)
- **217 realizations, target: 3500**
- light-cone covering ~half sky starting at z=4
- central l.o.s. aligned with the z axis **or** with the main diagonal **or** random
- 2 outputs (z=1 and z=0) + lightcone + histories,
- storage: 220 Gb (50 Gb for the lightcone)
- 52 TB total



computer time provided by **INFN-Euclid**

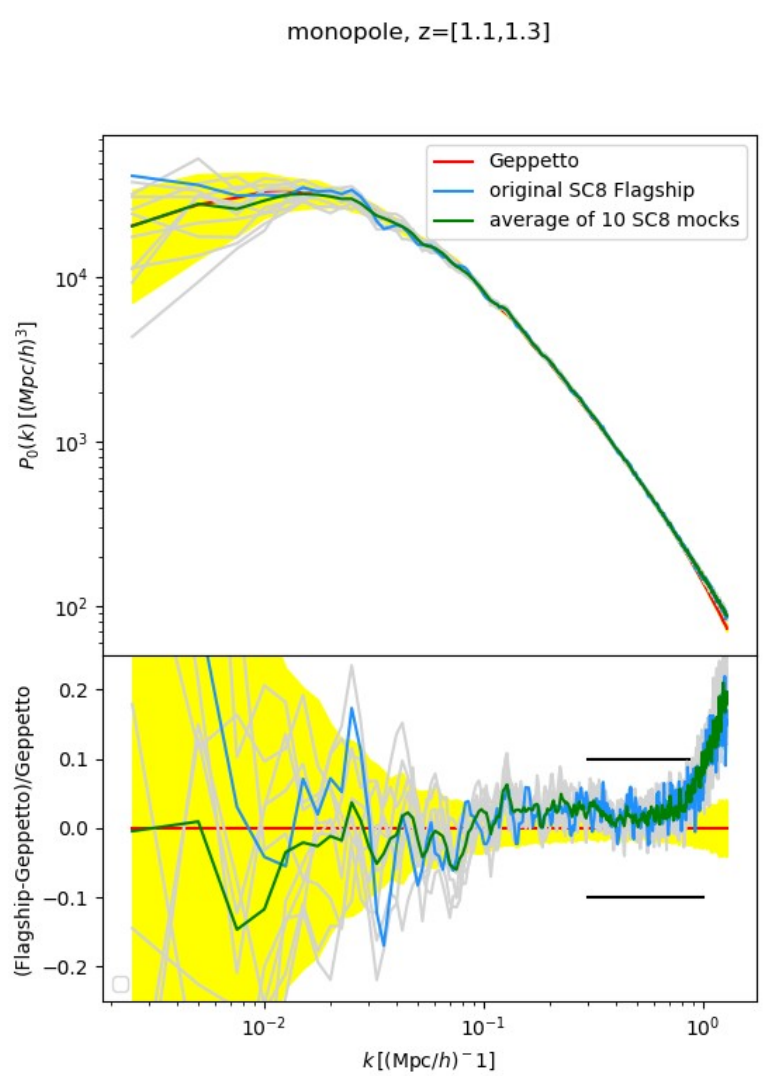
From halos to galaxies

Flagship mock galaxy catalog:

- match DM halos to r-band galaxies using an HOD,
- assign SFR, B/T ratio, bulge and disk sizes,
- sample galaxy SEDs according to their properties,
- add emission lines, including internal extinction,
- “predict” observed magnitudes and line fluxes and calibrate them against available observations

Mocks for the covariance:

- only produce position, redshift, line flux
- “measure” HOD from Flagship to populate halos
- process the Flaghsip mock with image simulations (**bypasses**)
- compute the probability of detecing a galaxy, already marginalized over galaxy properties, given image noise and exposure time
- use this probability to select galaxies



From halos to galaxies

Flagship mock galaxy catalog:

- match DM halos to r-band galaxies using an HOD,
- assign SFR, B/T ratio, bulge and disk sizes,
- sample galaxy SEDs according to their properties,
- add emission lines, including internal extinction,
- “predict” observed magnitudes and line fluxes and calibrate them against available observations

Mocks for the covariance:

- only produce position, redshift, line flux
- “measure” HOD from Flagship to populate halos
- process the Flaghsip mock with image simulations (**bypasses**)
- compute the probability of detecing a galaxy, already marginalized over galaxy properties, given image noise and exposure time
- use this probability to select galaxies

Parallel effort by Galaxy Evolution SWG (De Lucia):

- collect galaxy catalogs from hydro simulations and SAMs,
- project them on the past lightcone (limited areas)
- produce a consistent set of observed magnitudes

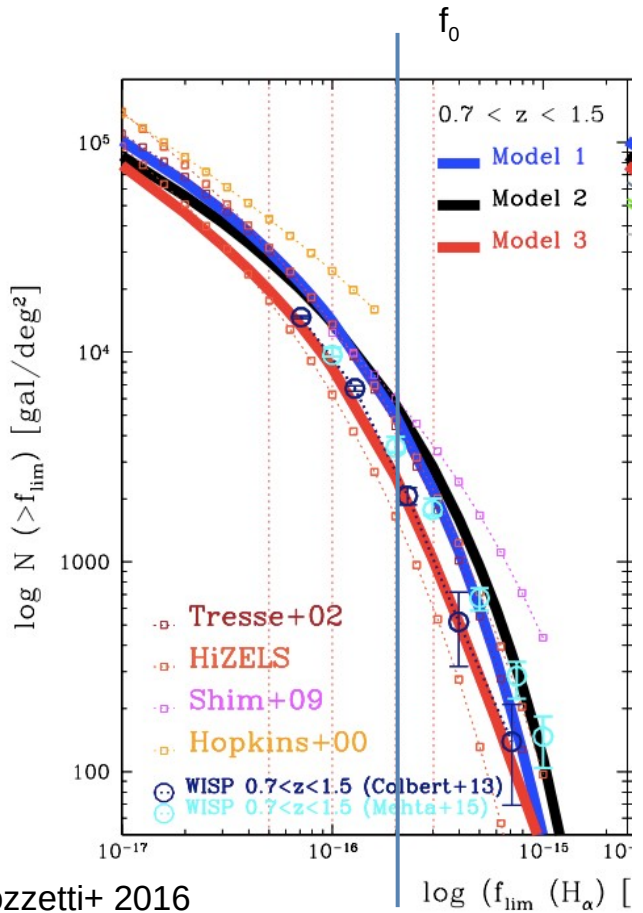
Galaxy Clustering systematics

Assumptions:
+ we restrict to linear galaxy bias
+ galaxy bias does not depend on luminosity
then the shape of the luminosity function is universal

We aim at measuring the galaxy number density:

$$n_{\rm g} = \int_{f_0}^{\infty} [1 + \delta_{\rm g}(\mathbf{x})]\, \Phi(f|z)\, df$$

f_0 is a fiducial Halpha flux limit (e.g. f_0 = 2e-16 erg/s/cm^{-2}).



Pozzetti+ 2016

Galaxy Clustering systematics

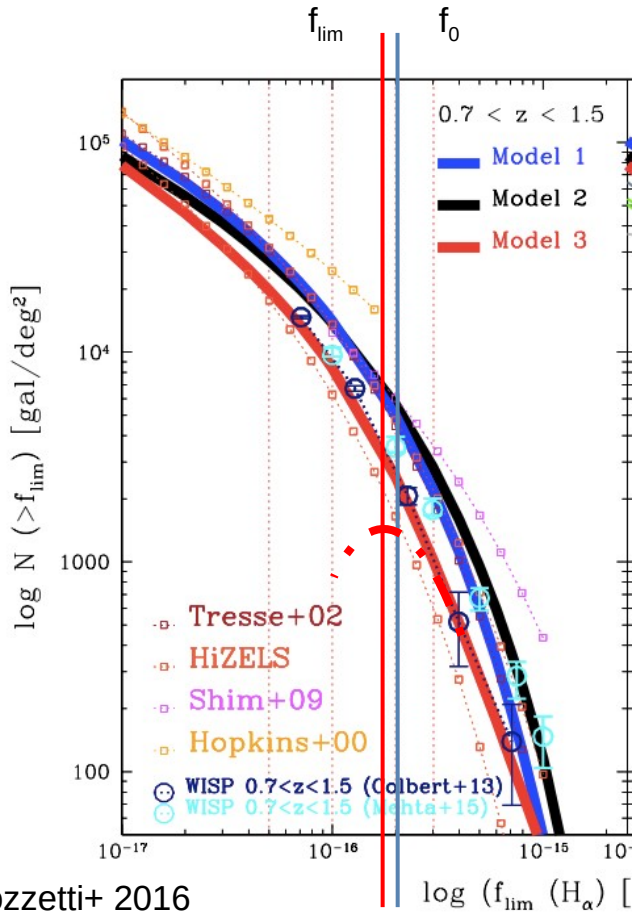
Our measurement involves a completeness function:

$$n_{\rm o} = \int_0^\infty \left[1 + \delta_{\rm g}(\mathbf{x})\right] \Phi(f|z)\, \mathcal{C}(f, z, \{N_i\})\, df$$

Completeness can be recast in terms of a flux limit f_{lim}:

$$n_{\rm o} = \int_{f_{\rm lim}(\boldsymbol{\vartheta}, z, \{N_i\})}^\infty \left[1 + \delta_{\rm g}(\mathbf{x})\right] \Phi(f|z)\, df$$

Here $\{N_i\}$ represent a generic set of noise terms that determine the probability of a galaxy being detected.



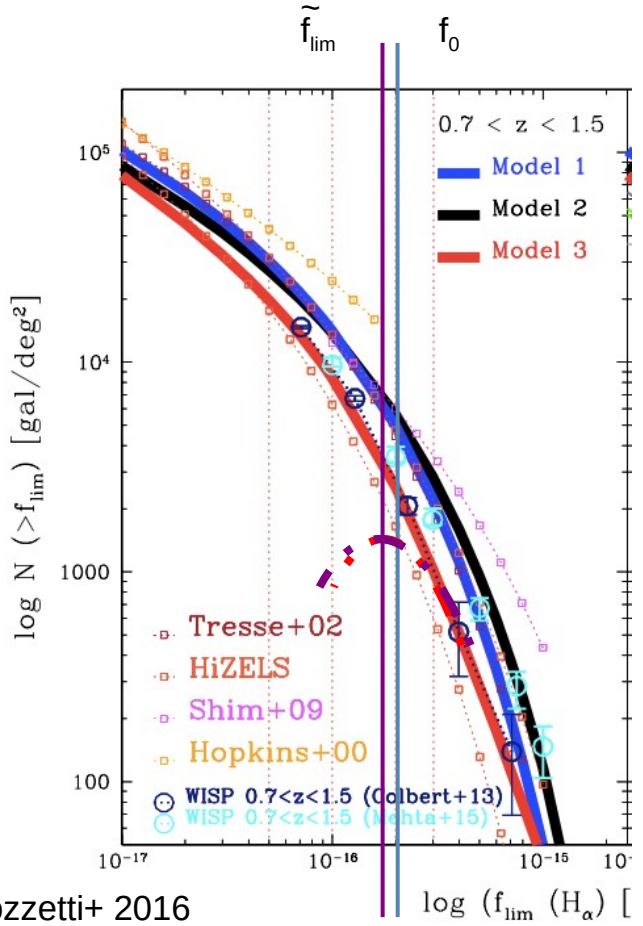
Pozzetti+ 2016

Galaxy Clustering systematics

Galaxy density is compared with that of a random catalog, created on the basis of an approximation of the completeness function (=visibility mask):

$$n_{\rm r} = N_{\rm r} \int_0^\infty \Phi(f|z)\,\tilde{\mathcal{C}}(f,z,\{N_i\})\,df$$

$$= \int_{\tilde{f}_{\rm lim}(\boldsymbol{\vartheta},z,\{N_i\})}^\infty \Phi(f|z)\,df$$



Pozzetti+ 2016

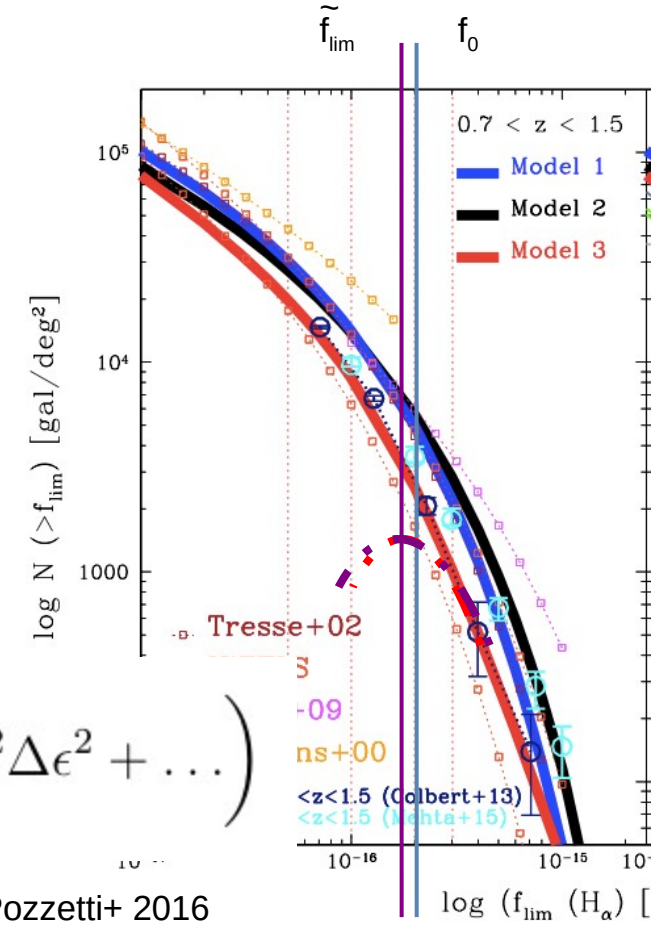
Galaxy Clustering systematics

The observed density contrast can then be expressed in terms of the galaxy (true) density contrast and the uncertainty in the flux limit ($\alpha = 1/N_r$):

$$f_{\rm lim}(\boldsymbol{\vartheta}, z, \{N_i\}) = f_0\left[1 + \epsilon(\boldsymbol{\vartheta}, z, \{N_i\})\right]$$

$$\Delta\epsilon = \tilde{\epsilon} - \epsilon$$

$$n_{\rm o} - \alpha n_{\rm r} = \delta_{\rm g}\alpha n_{\rm r} + (1+\delta_{\rm g})\left(1 + \Phi f_0 \Delta\epsilon + \frac{1}{2}\frac{d\Phi}{df} f_0^2 \Delta\epsilon^2 + \dots\right)$$



Pozzetti+ 2016

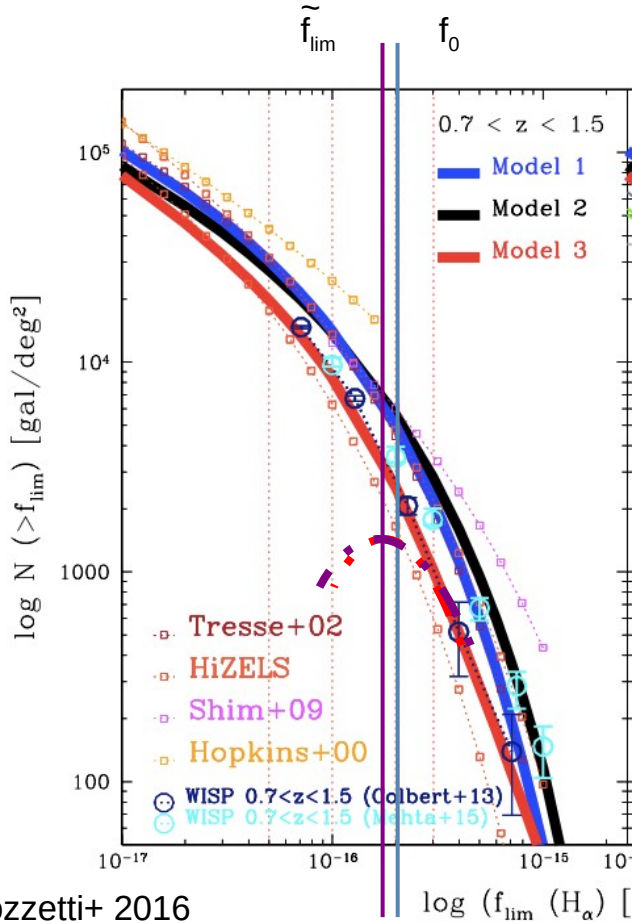
Galaxy Clustering systematics

This means that the measured signal **is not** a sum of cosmic signal + noise:

$$\delta_{\rm o} \neq \delta_{\rm g} + A$$

$$1 + \delta_{\rm o}(\mathbf{x}) = [1 + \delta_{\rm g}(\mathbf{x})]\,[1 + A(\boldsymbol{\vartheta}, z)]$$

$$A = \Phi f_0 \Delta\epsilon + \frac{1}{2}\frac{d\Phi}{df} f_0^2 \Delta\epsilon^2 + \dots$$



Pozzetti+ 2016

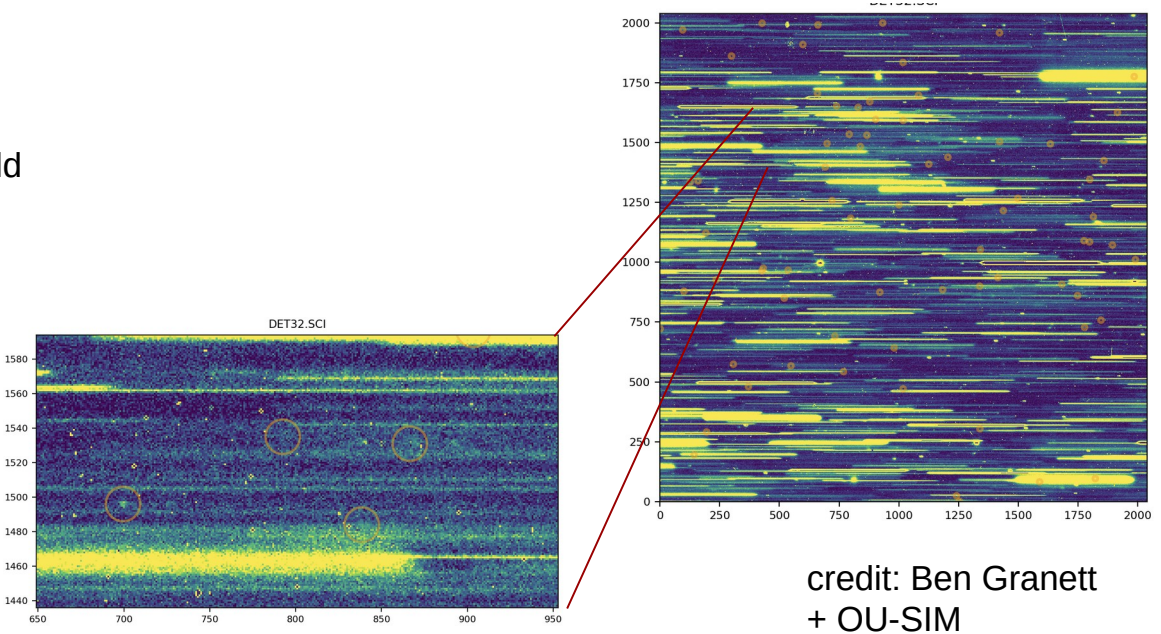
Mitigation through the random catalog

Construction of the random catalog

- place a galaxy in a random sky position
- measure the noise level at that position
- add galaxy properties drawn from the deep field
- use a bypass (pypelid) to estimate its SNR
- map SNR to detection probability
- decide if the galaxy gets into the random

Degrees of freedom:

- random error on noise
- calibration error
- assigning properties to the galaxy
- from detection probability to getting or not in the random

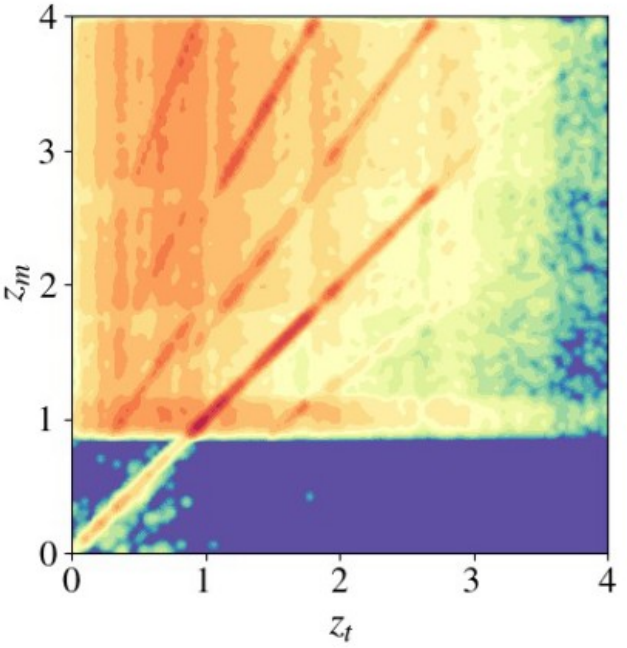


credit: Ben Granett
+ OU-SIM

Redshift errors

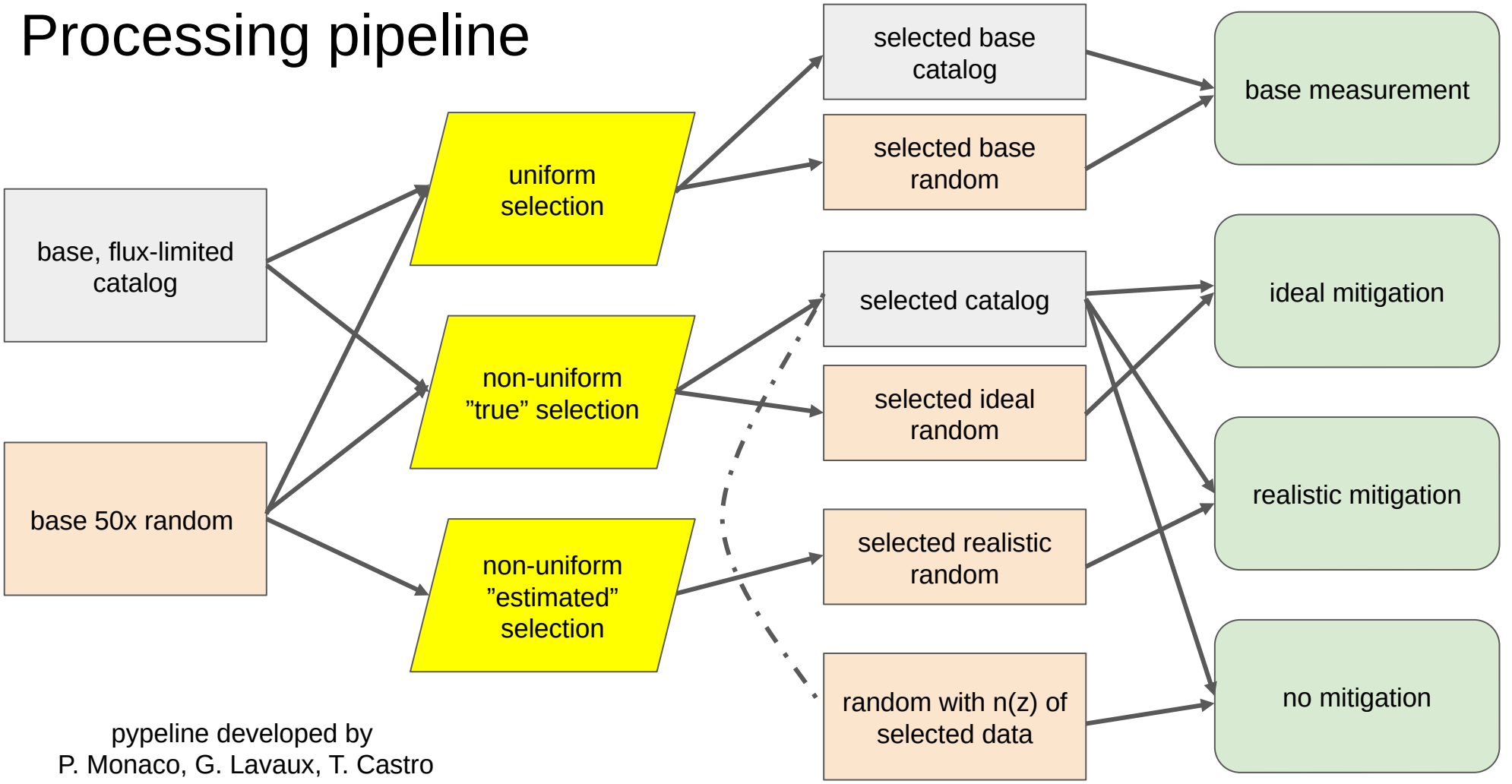
```
observed galaxies =

  good galaxies
+ line misidentifications   -> rescaled correlation
+ noise spikes              -> uncorrelated?
+ galactic objects          -> correlated like MW
```

Line misidentifications can be treated at the likelihood level.
Noise spikes and galactic objects must be represented in the random



credit: Sylvain De La Torre
+ OU-SPE

Processing pipeline



pypeline developed by
P. Monaco, G. Lavaux, T. Castro

Galaxy Clustering covariance

This propagates to the two-point CF (one more term) in and covariance, with **74 more terms** to add:

$$\xi_{\rm o} = \xi_{\rm g}(1+2\langle A\rangle) + \xi_A + \xi_{\rm g}\xi_A$$

$$\begin{aligned}
C_{\rm o} &= \langle(\xi^{(r)}_{\rm o12} - \xi_{\rm o12})(\xi^{(r)}_{\rm o34} - \xi_{\rm o34})\rangle \\
&= C_{\rm g} + C_A \\
&\quad + \xi_{\rm g13}\xi_{A34} + \xi_{\rm g14}\xi_{A23} + \xi_{\rm g23}\xi_{A14} + \xi_{\rm g24}\xi_{A13} \\
&\quad + \langle\delta_{\rm g1}\delta_{\rm g2}\delta_{\rm g3}\rangle[\langle A_4\rangle + \xi_{A14} + \langle A_1A_2A_4\rangle] + \text{perm.} && \text{28 terms} \\
&\quad + \langle\delta_{\rm g1}\delta_{\rm g2}\delta_{\rm g3}\delta_{\rm g4}\rangle[\langle A_1\rangle + \xi_{A12} + \langle A_1A_2A_3\rangle] + \langle A_1A_2A_3A_4\rangle] + \text{perm.} && \text{19 terms} \\
&\quad + \xi_{\rm g12}[\langle A_1A_3A_4\rangle + \langle A_1A_2A_3A_4\rangle] + \text{perm.} && \text{18 terms} \\
&\quad - \xi_{\rm g12}\xi_{\rm g34}\xi_{A12} - \xi_{\rm g12}\xi_{\rm g34}\xi_{A34} - \xi_{\rm g12}\xi_{A12}\xi_{A34} - \xi_{\rm g34}\xi_{A12}\xi_{A34} \\
&\quad - \xi_{\rm g12}\xi_{\rm g34}\xi_{A12}\xi_{A34}
\end{aligned}$$

Galaxy Clustering covariance

We worked out the expression for the power spectrum in Colavincenzo et al. (2017):

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

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

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

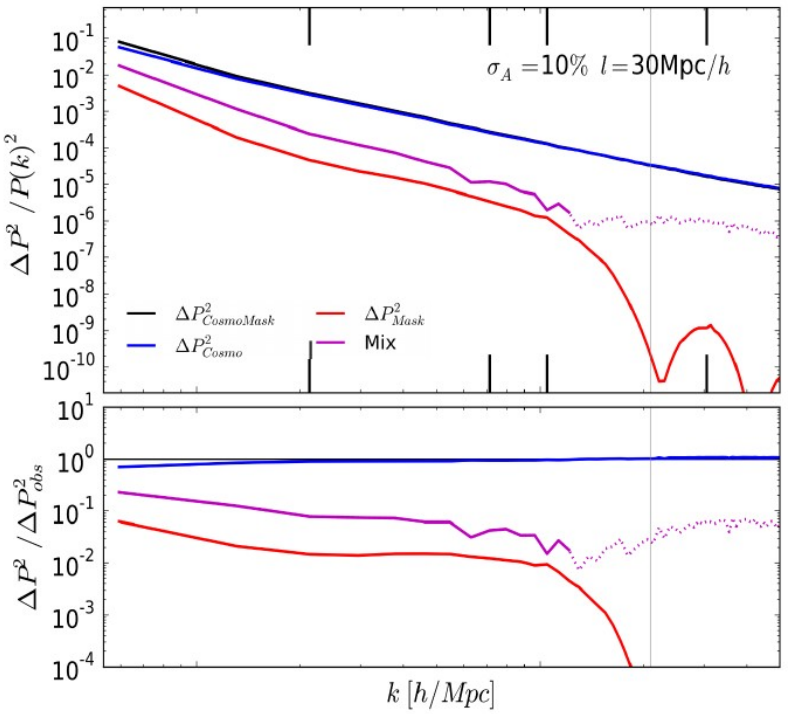
$$\hat{G} = 2\delta_{\mathbf{q}}\delta_{\mathrm{mask},\mathbf{q}} - \delta_{\mathbf{q}}\delta_{\mathrm{conv},\mathbf{q}} + \delta_{\mathrm{mask},\mathbf{q}}\delta_{\mathrm{conv},\mathbf{q}}$$

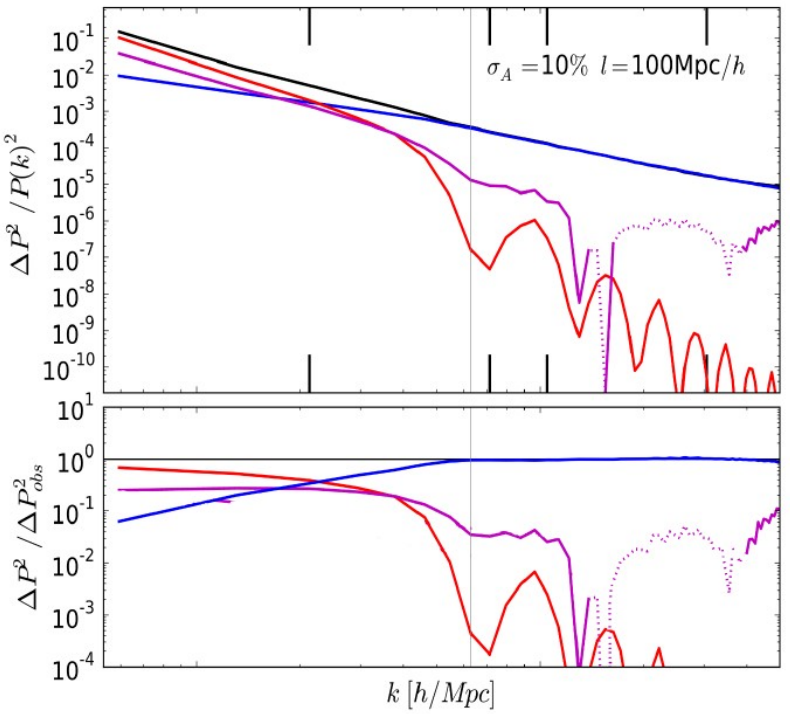
Galaxy Clustering covariance

Results: Variance

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

10 000 Pinocchio realizations



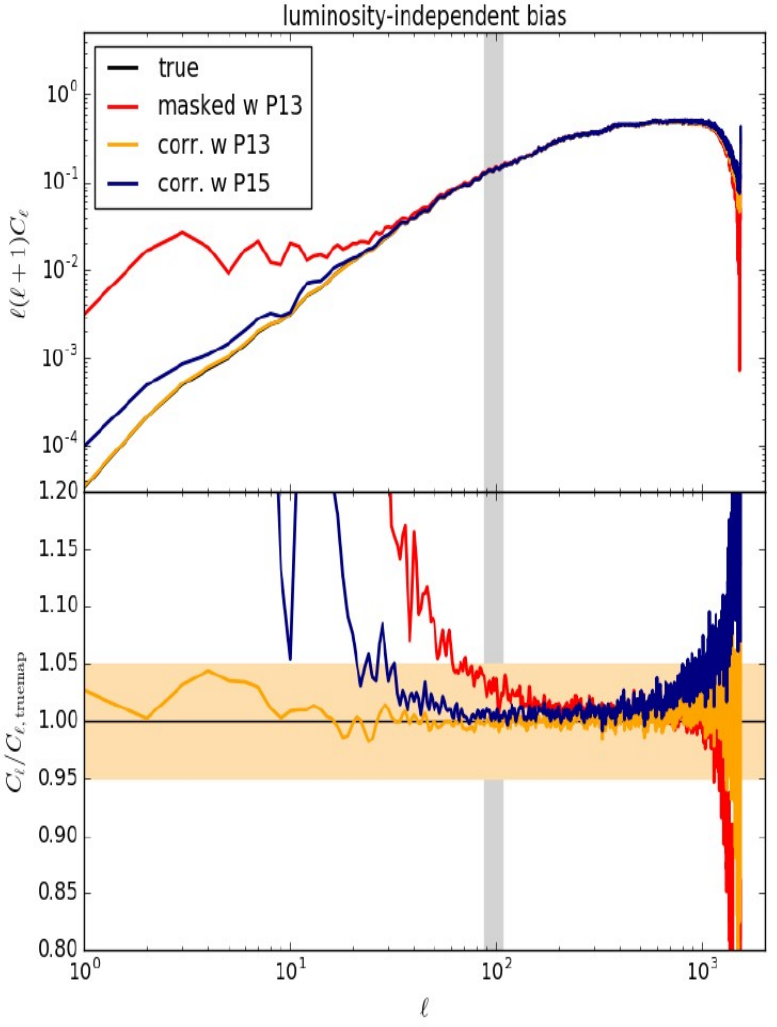


What is missing

Systematics may have error correlations that are not caught by this scheme.
Example: MW extinction.

$$1 + \delta_{\rm o}(\mathbf{x}) = [1 + \delta_{\rm g}(\mathbf{x})]\,[1 + A(\boldsymbol{\vartheta}, z)]$$

We need to identify and mitigate “unknown unknowns”.

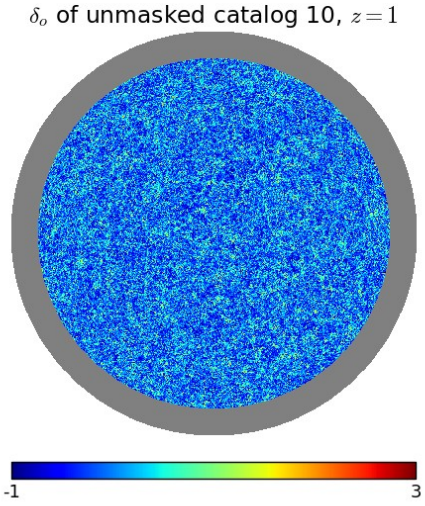


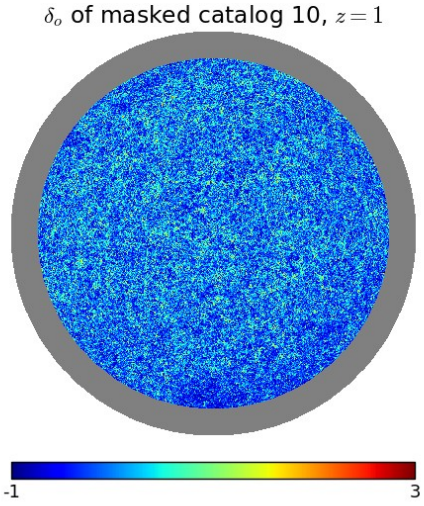
Journal of Cosmology and Astroparticle Physics
An IOP and SISSA journal

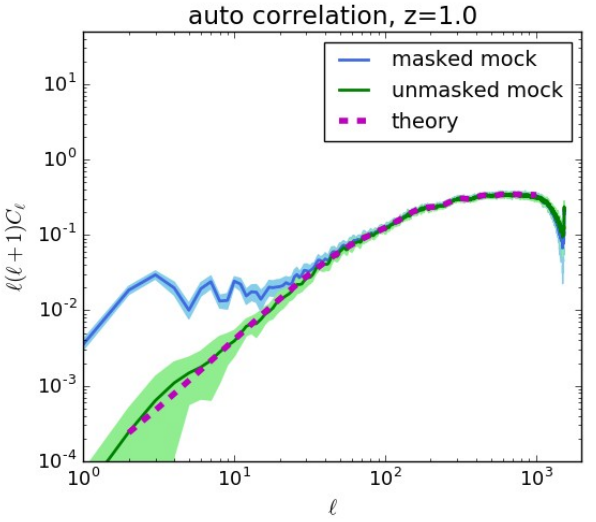
A blind method to recover the mask of a deep galaxy survey

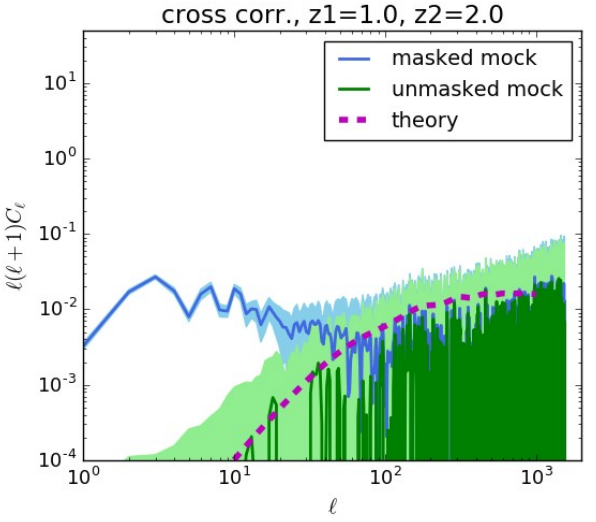
Pierluigi Monaco,[a,b,c] **Enea Di Dio**[d,e] **and Emiliano Sefusatti**[b,c]

Journal of Cosmology and Astroparticle Physics, 04, 023 (2019).







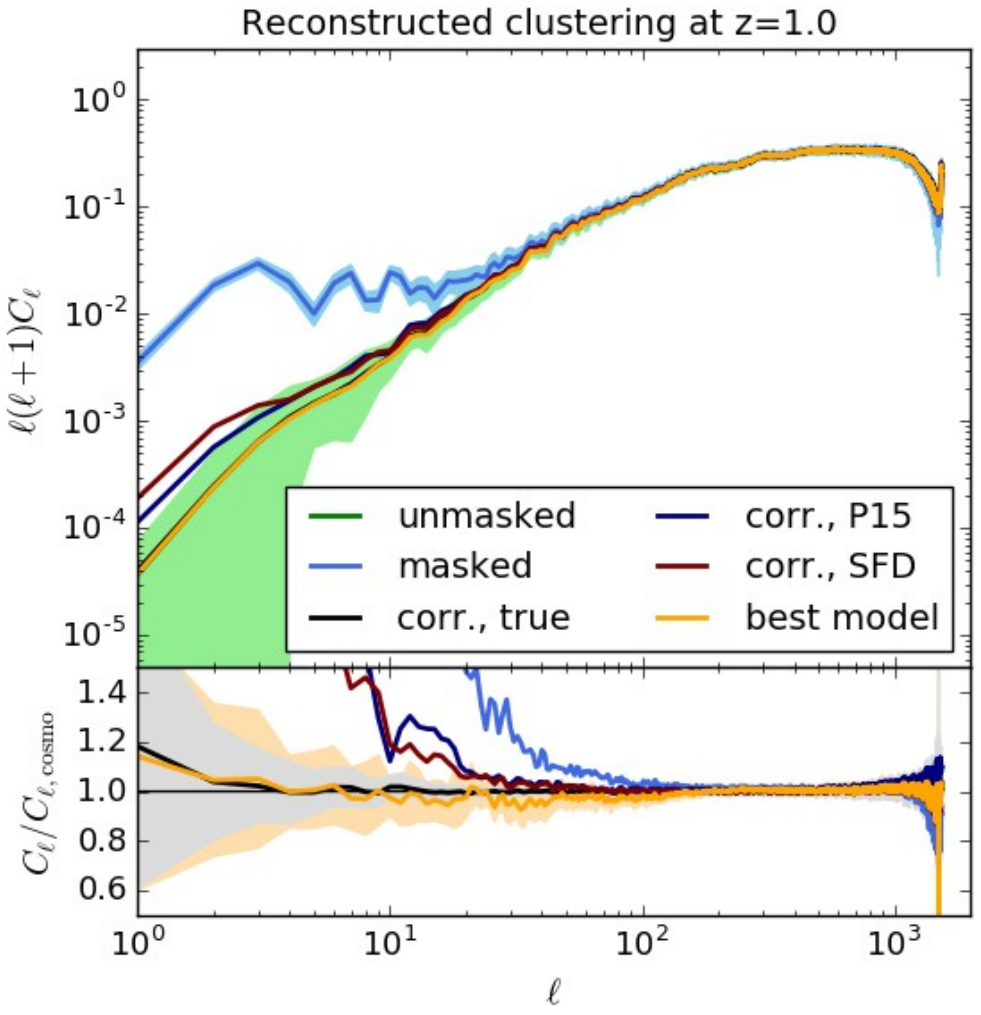


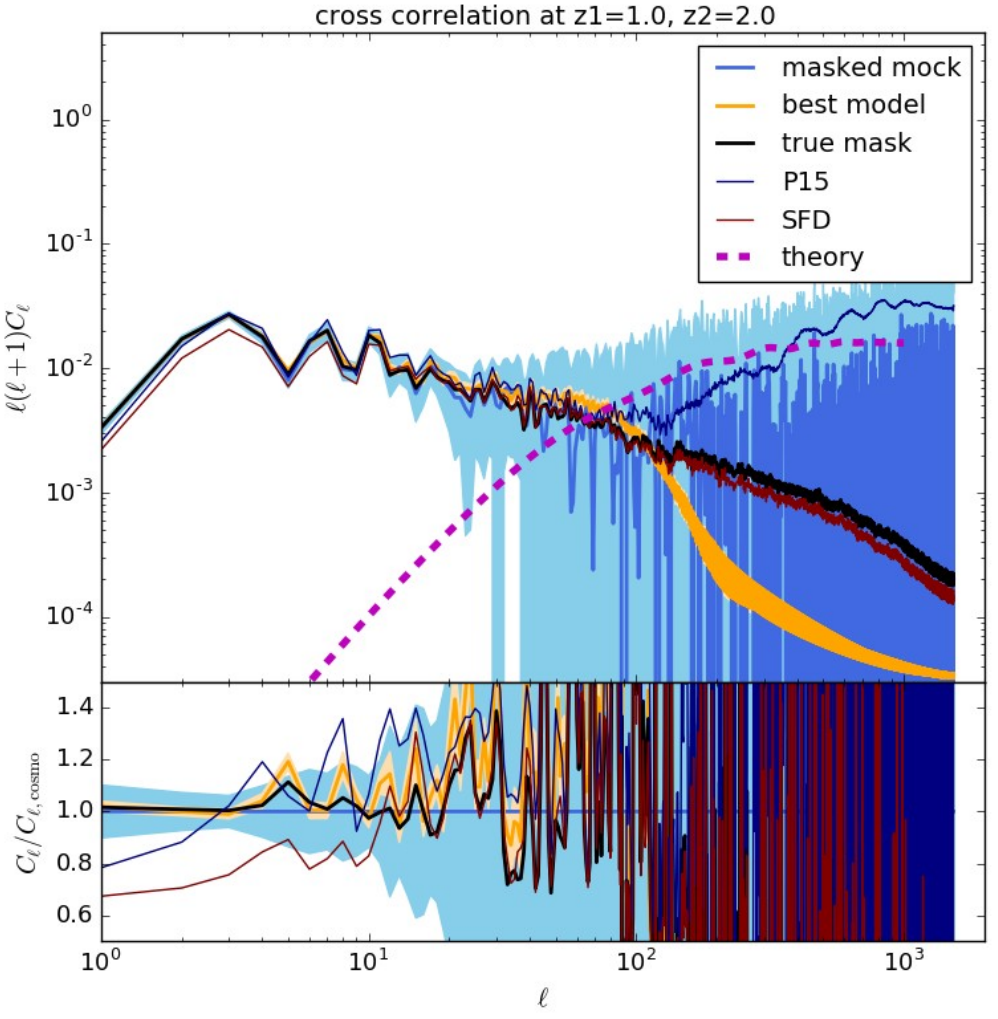
One can construct estimators of the visibility mask by averaging the density contrast over the line of sight, at the 1-point or 2-point level

$$E_{\rm av}(\boldsymbol{\theta}) \equiv \frac{1}{N_z}\sum_i \delta_{\rm r}([z_i,\boldsymbol{\theta}]) \simeq -\frac{1}{N_z}\sum_i \frac{S_{Ci}}{S_{Ai}}\mathcal{M} - \frac{1}{N_z}\sum_i \frac{S_{Bi}S_{Ci}^2}{S_{Ai}}\mathcal{M}^2 + O(\mathcal{M}^3)\,. \qquad (4.4)$$

$$E_{\rm sq}(\boldsymbol{\theta}) \equiv \frac{1}{N_p}\sum_i\sum_{j>i} \delta_{\rm r}([z_i,\boldsymbol{\theta}])\delta_{\rm r}([z_j,\boldsymbol{\theta}]) \qquad (4.5)$$

$$\simeq \frac{1}{N_p}\sum_i\sum_{j>i} \frac{S_{Ci}S_{Cj}}{S_{Ai}S_{Aj}}\mathcal{M}^2 + \frac{1}{N_p}\sum_i\sum_{j>i} \frac{S_{Ci}S_{Cj}(S_{Bi}S_{Ci}+S_{Bj}S_{Cj})}{S_{Ai}S_{Aj}}\mathcal{M}^3 + O(\mathcal{M}^4)\,.$$


Reconstructed clustering at z=1.0
$\ell(\ell+1)C_\ell$
unmasked
masked
corr., true
corr., P15
corr., SFD
best model
$C_\ell/C_{\ell,\mathrm{cosmo}}$
ℓ


cross correlation at z1=1.0, z2=2.0
masked mock
best model
true mask
P15
SFD
theory
$\ell(\ell+1)C_\ell$
$C_\ell/C_{\ell,\mathrm{cosmo}}$
ℓ

Possible synergies

...will we share the mocks?

but, what about mass resolution?
would it be convenient to add small halos with a BAM-like model (see De La Torre & Peacock 2012)?