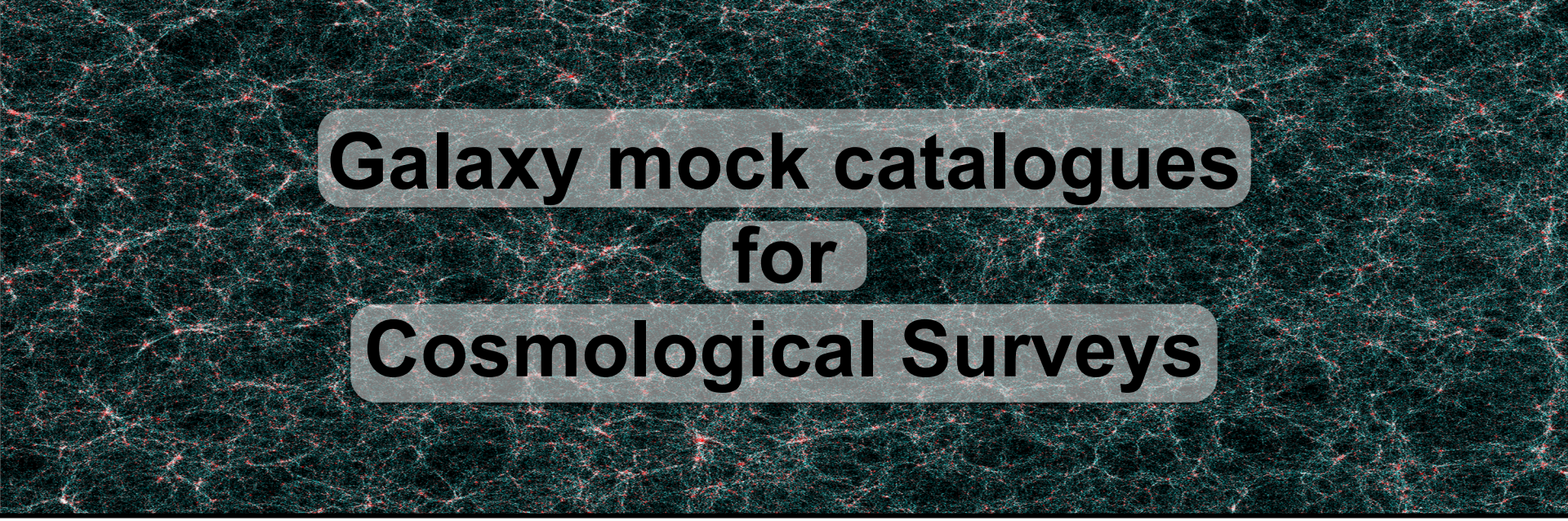


# Galaxy mock catalogues for Cosmological Surveys


Ismael Ferrero

- Postdoc at ITA -

Oslo, Norway



Iceland Liechtenstein Norway grants

**The unbearable lightness of the Universe**
Praha, 16-18 Sep 2021



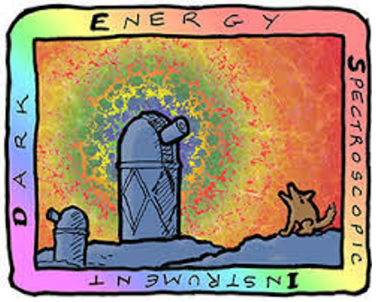

UiO : Institute of Theoretical Astrophysics
The Faculty of Mathematics and Natural Sciences

# Motivation

The picture can't be displayed.

Present and planned galaxy surveys like

DARK ENERGY SURVEY




euclid consortium

LSST Large Synoptic Survey Telescope

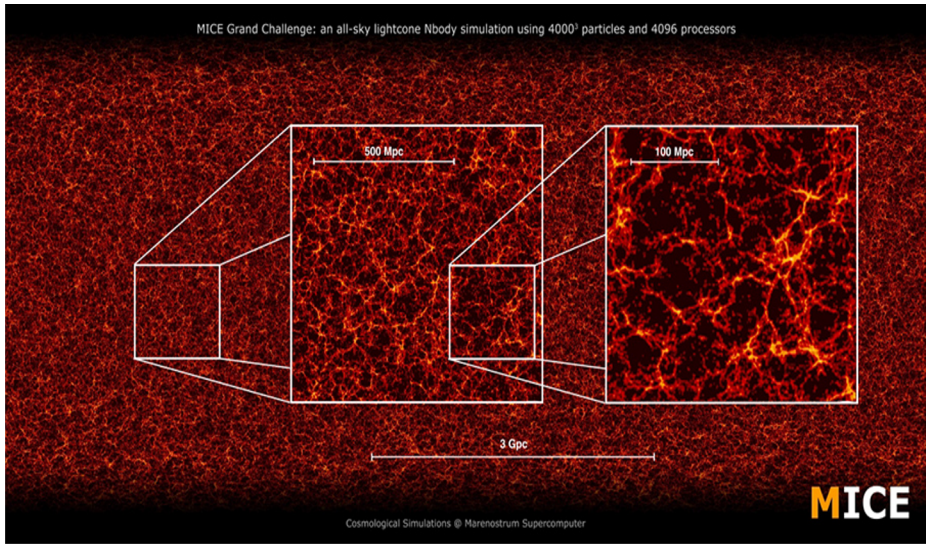
Will generate a wealth of high-quality data that will allow to test the nature of dark-energy and constrain possible deviations from the standard cosmological model.

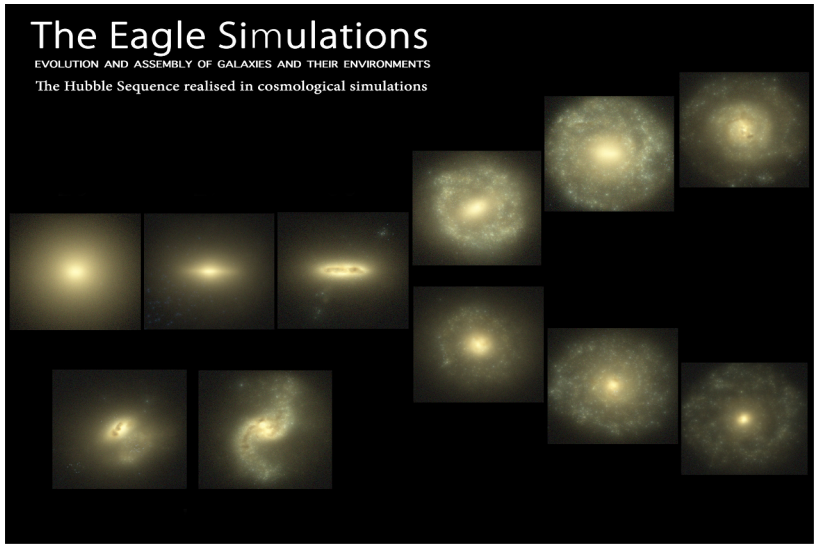
An optimal extraction of cosmological parameters from those very large and complex datasets will ultimately rely on our ability to model cosmological observables and their covariances with high accuracy

**This entails the development of synthetic observations**

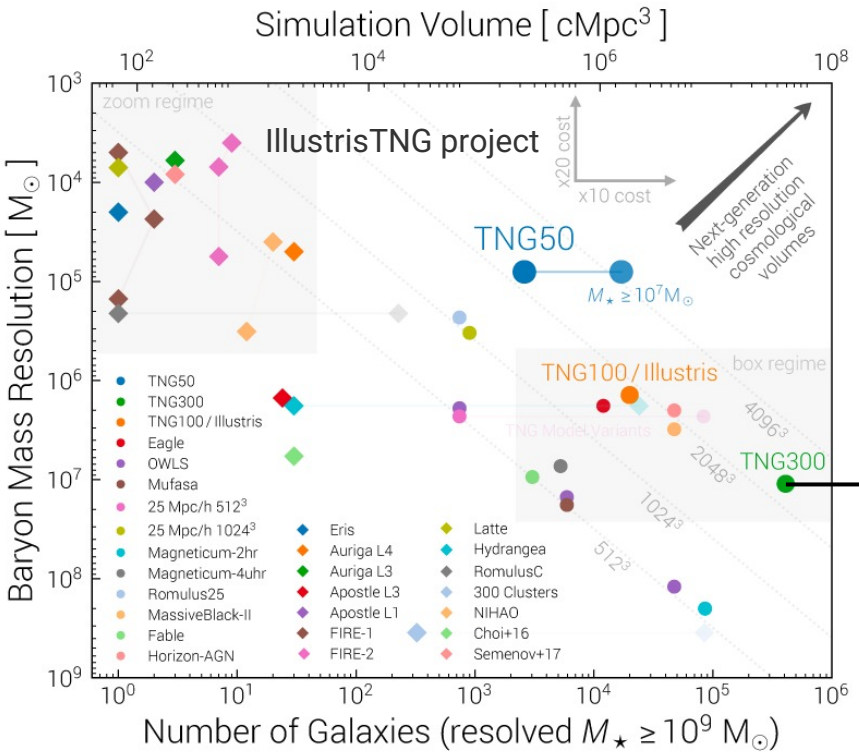
# Computational problem

The requirement of sampling large cosmological volumes while still resolving small scales is a big challenge to current N-body simulation codes.

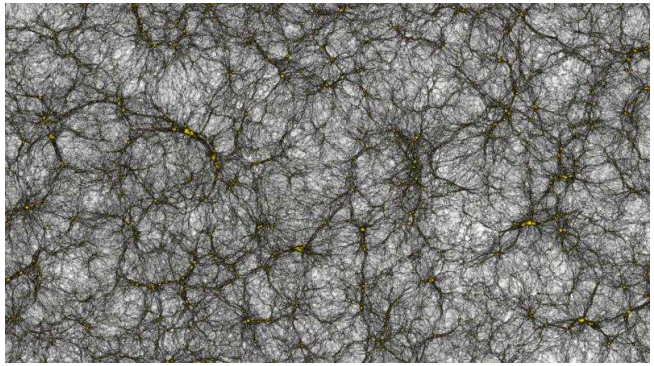


# Computational problem




Flagship: official mock for Euclid mission



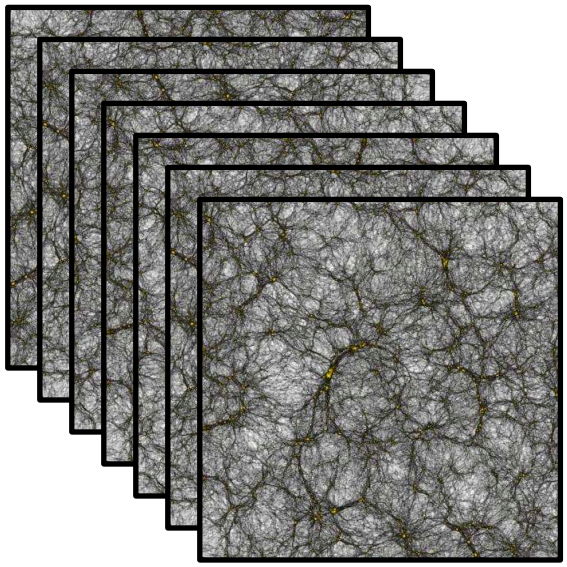
2000 times bigger

**Problem** To have galaxy catalogues of cosmological volumes

**Solution** Populate, by hand, cosmological dark matter-only simulations with galaxies

# Computational problem

Moreover, hundreds or thousands of realizations are needed for robustly estimating covariance matrices or for propagating errors in complex and non-linear analysis .



**Problem** Producing massive ensembles of N-body mocks is computationally prohibitive

**Solution** To use fast simulation

## What we need to create a galaxy catalogue given a halo catalogue?

**HOD**

number of galaxies per halo

**Halo mass profile**

satellites position

**AM & SHAM**

mass or luminosity

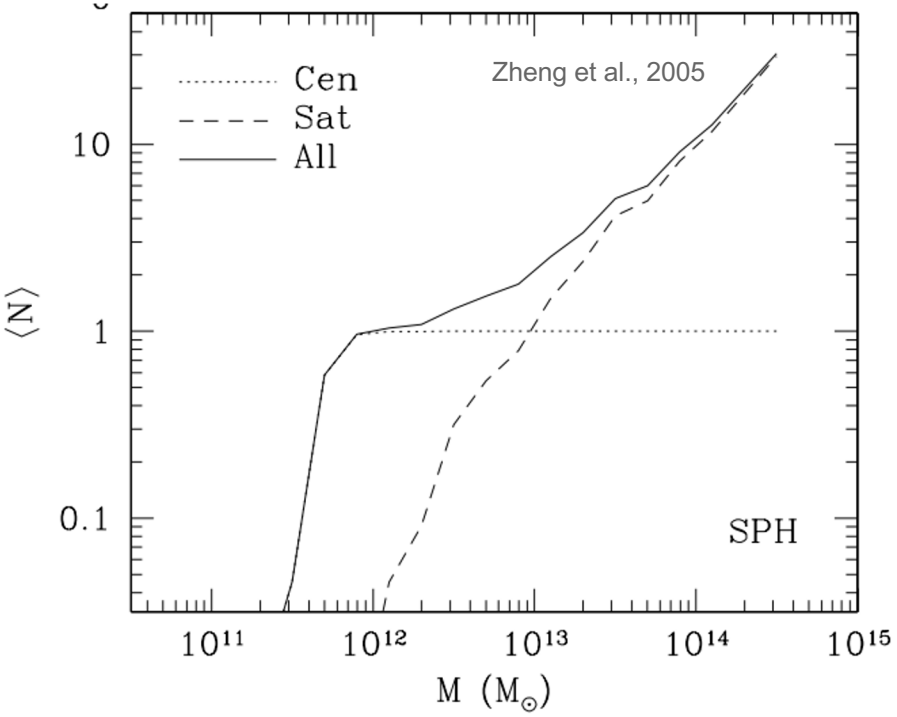
**Velocity distribution**

satellites velocity

# HOD

## Halo Occupation Distribution, for all kind of galaxies

Provides a view of how galactic matter is distributed within each of the dark matter clumps




$$< N_c >_M = \frac{1}{2}\left[1 + erf\left(\frac{logM - logM_{min}}{\sigma_{logM}}\right)\right]$$

$$< N_s >_M = \left(\frac{M}{M_1}\right)^{\alpha} \longrightarrow M > M_{cut}$$

$$< N_s >_M = 0 \longrightarrow M < M_{cut}$$

$M_{min}$ = characteristic minimum mass of haloes that host centrals above luminosity threshold

$\sigma_{\log M}$ = characteristic transition width due to scatter in L-M relation of centrals

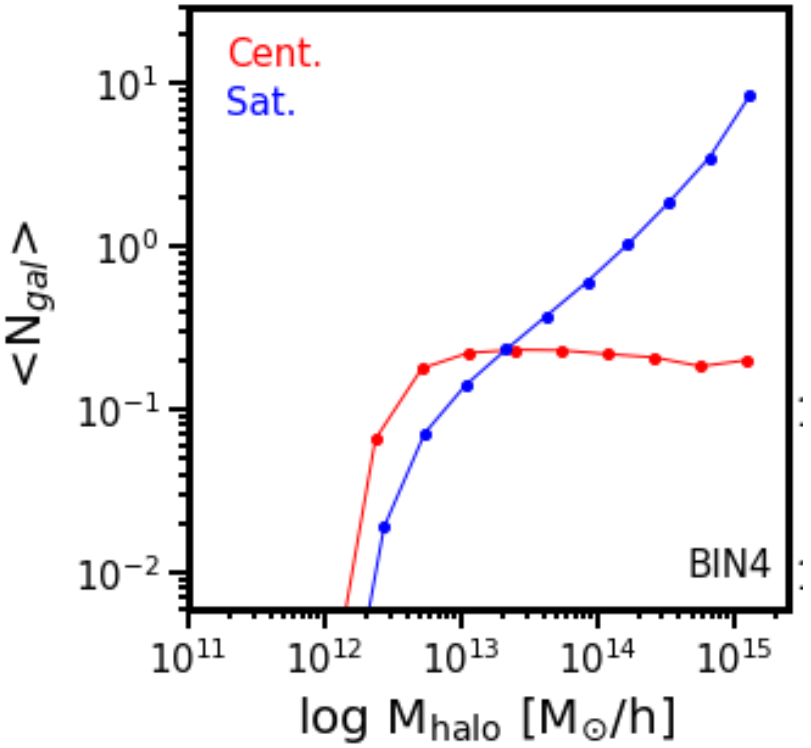
$M_{cut}$ = cut-off mass below which you have zero satellites above luminosity threshold

$M_1$ = normalization of satellite occupation numbers

$\alpha$ = slope of satellite occupation numbers

# HOD

## Halo Occupation Distribution, for a kind of galaxies




Sample of Emitter Line galaxies (ELG)

$$< N_{cen} >= f_{cen}^{max} \times \frac{1}{2}\left[1 + erf\left(\frac{logM_h - logM_{min}}{\sigma_{logM}}\right)\right] \times \left[1 - \frac{1 - f_{cen}^{min}/f_{cen}^{max}}{1 + 10^{\frac{2}{k}(logM_h - logM_{drop})}}\right]$$

$$< N_{sat} >= f_{cen}^{max} \times \frac{1}{2}\left[1 + erf\left(\frac{logM_h - logM_{min}}{\sigma_{logM}}\right)\right] \times \left(\frac{M_h}{M_1}\right)^{\alpha}$$

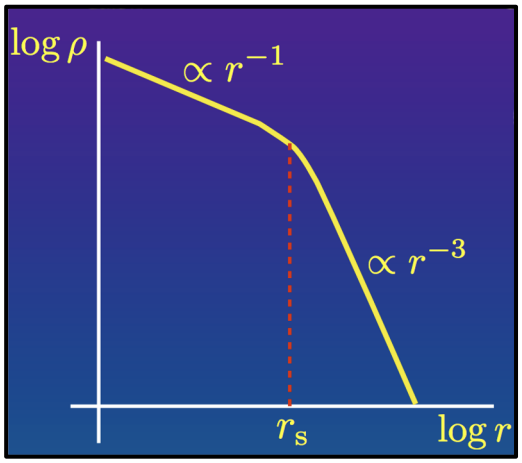
# Halo mass profile

## The NFW Profile

Where to put the satellite galaxies ?


In 1997, Navarro, Frenk & White wrote a seminal paper in which they showed that CDM haloes in Nbody simulations have a universal density profile, well fit by a double power-law...

$$\rho(r) = \rho_{crit} \frac{\delta_{char}}{(r/r_s)(1+r/r_s)^2}$$




It is completely characterized by the mass $M_{vir}$ and the concentration parameter$c = r_{vir}/r_s$ , which is related to the characteristic overdensity according to:

Since more massive haloes assemble later (on average) they are expected to be less concentrated, giving rise to an inverted concentration-mass relation. Strong correlation between $c$ and $M_{vir}$ parameters.

*"Mass distribution into a halo can be approximately modeled knowing only the halo total mass"*

# How we do to create catalogos

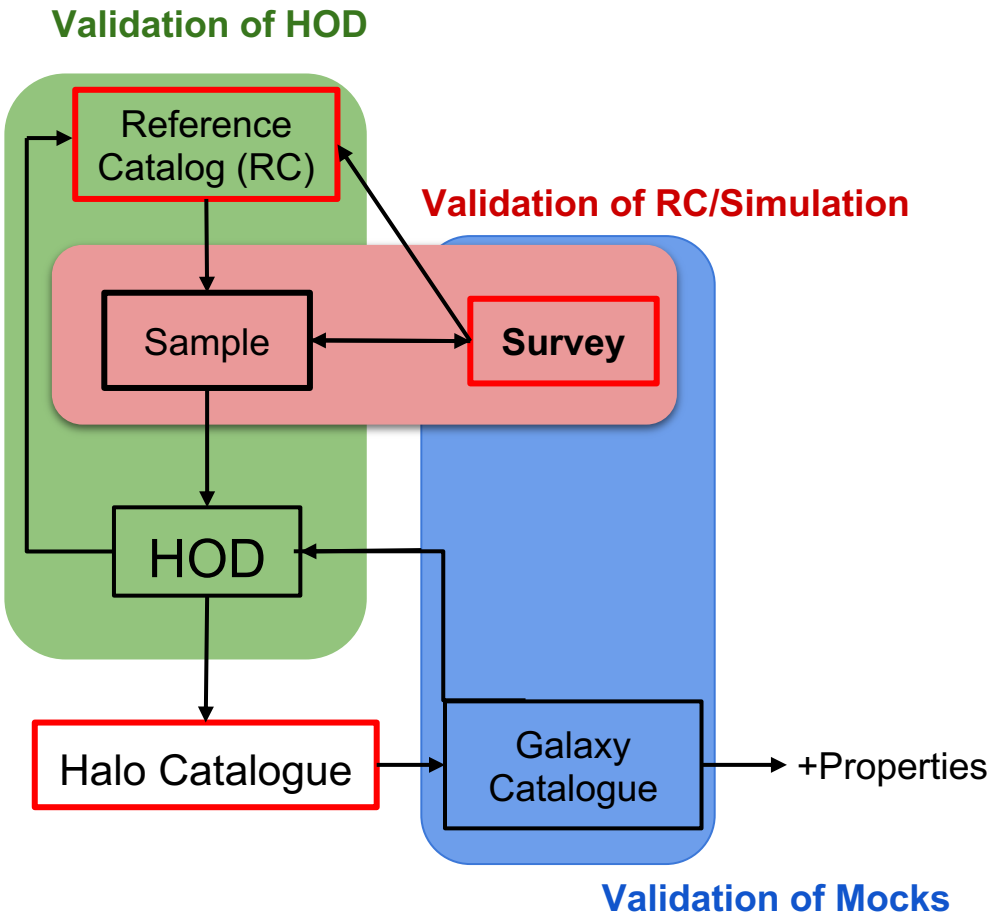
1. **Select a sample from the RC**

2. **Read off the HOD of the sample**
   spline, parameterized function, nº bins,..
3. **Apply that HOD back to the RC**
   triaxiality, concentration, …
4. **Test of HOD and implementation**
   n(z), clustering

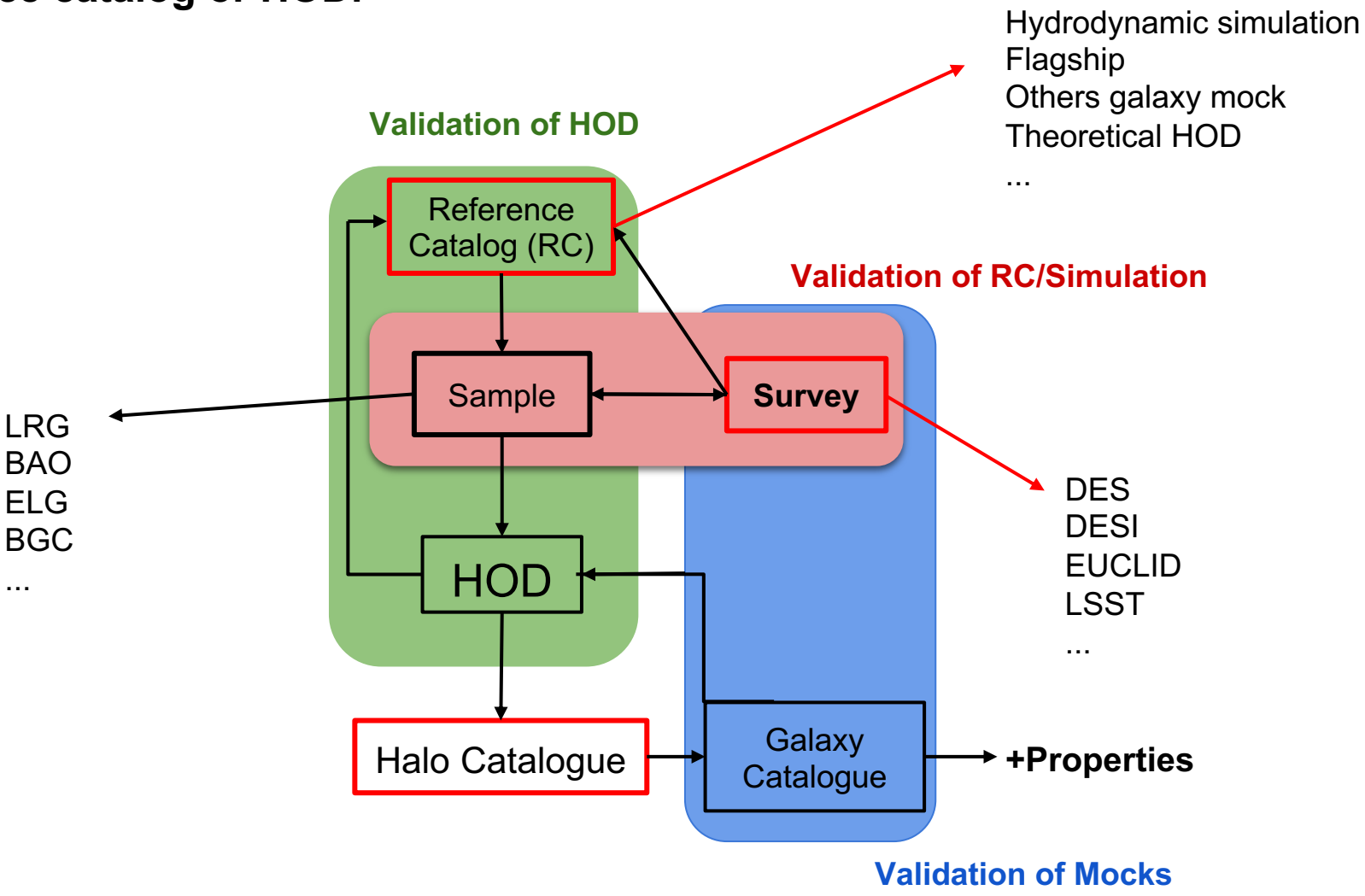
5. **Test against Survey data**
   n(z), clustering
6. **Calibrate HOD**
   parameters, implementation

7. **Apply resulting HOD to halo catalogue**
   mask
8. **Test against Survey data**
   n(z), clustering

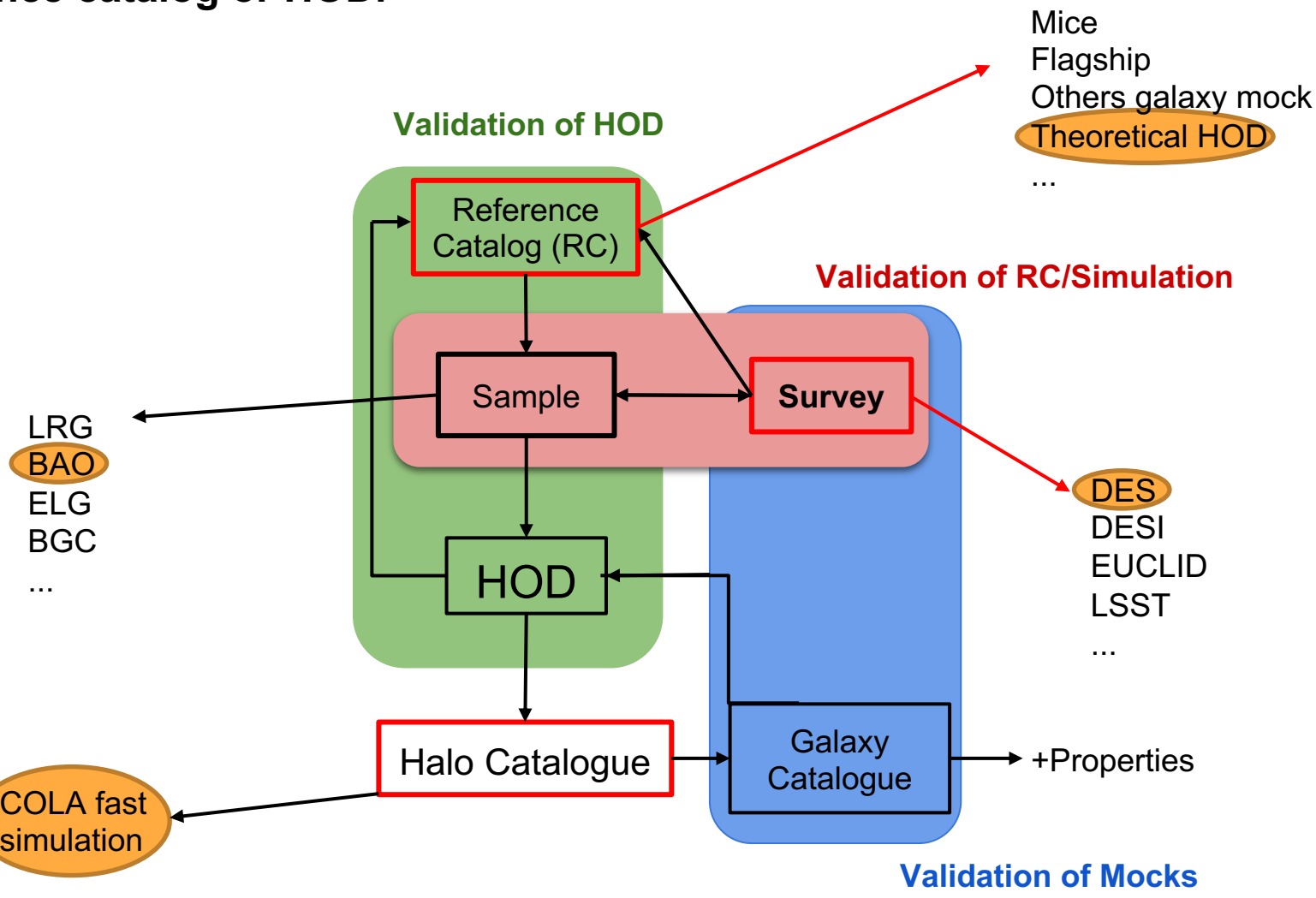
9. **Assign properties to galaxies**
   SHAM, colors, luminosities,...

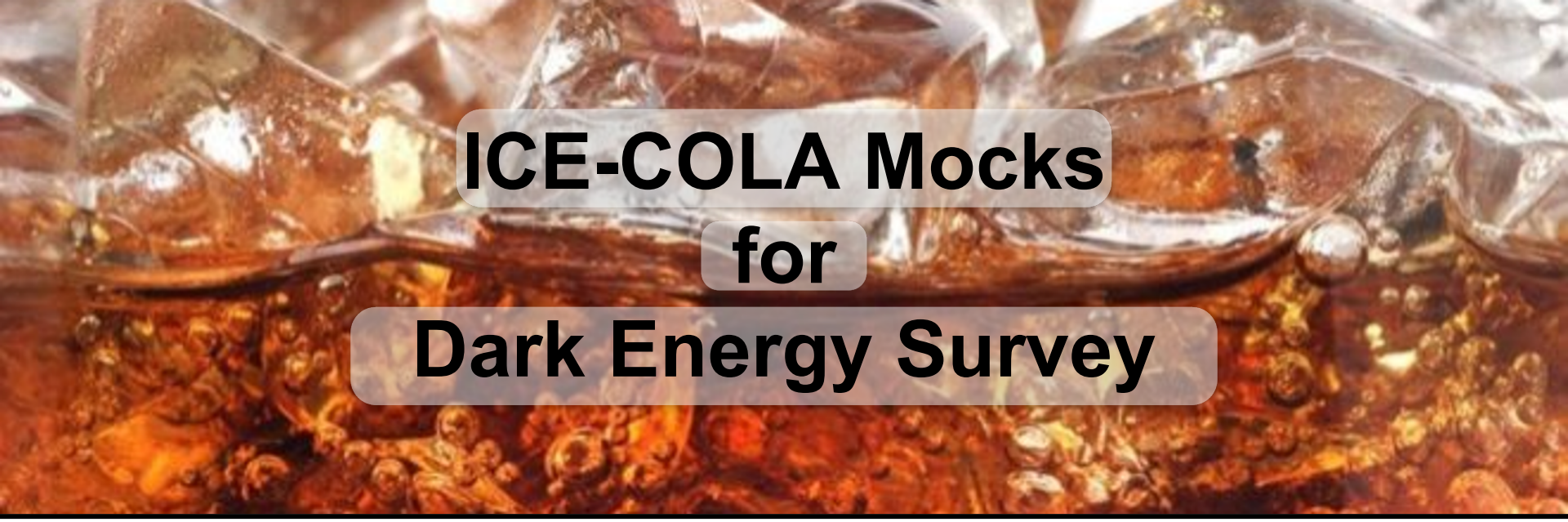


**Pipeline can be used for any survey, for any kind of sample and using any reference catalog or HOD.**



**Pipeline can be used for any survey, for any kind of sample and using any reference catalog or HOD.**





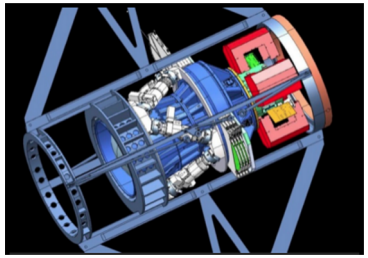
# ICE-COLA Mocks for Dark Energy Survey

**Mimic the BAO* sample of the DES survey and produce thousand of mock using COLA fast simulation**

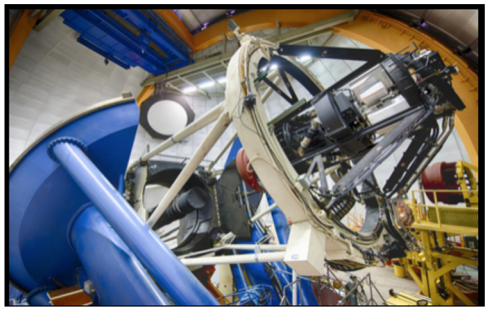
* A sample defined with color-magnitude cuts, dominated by red galaxies with a good compromise of photo-z accuracy and number density, optimal for the BAO measurement.

5 tomographics bins in the range Photo-z = [0.6 , 1.1]

# THE DARK ENERGY SURVEY



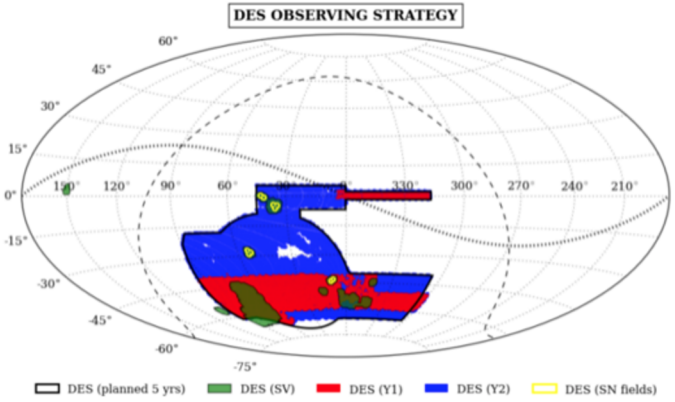
Dark Energy Camera



Victor M. Blanco Telescope

Cerro Tololo Inter-American Observatory

**DES has observed thousands of supernovae and hundreds of millions of galaxies to measure or constrain changes in dark energy over cosmic time.**

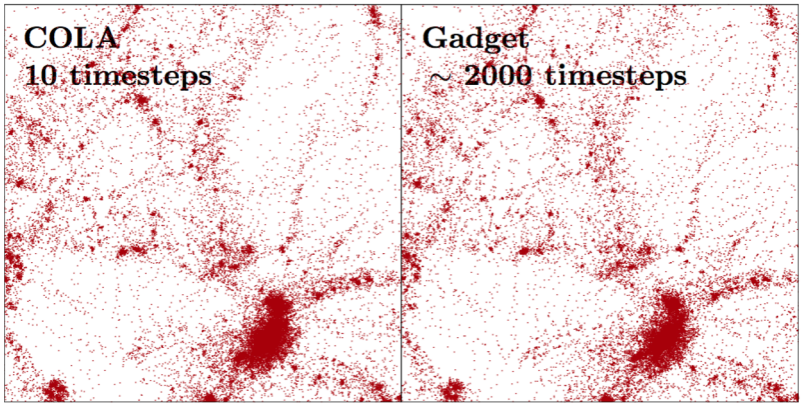



The *survey* covers a large, contiguous area of sky (5000 square degrees), each part of which is observed ten times in each filter over the course of the survey.

**COLA**
**CO**moving
**L**agrangian
**A**ccelerator

An N-body method for solving for Large Scale Structure (LSS) in a frame that is comoving with observers following trajectories calculated in Lagrangian Perturbation Theory (LPT)


Evolve particles according to the 2LPT trajectories + residual displacement evaluated by the N-body solver (Particle-Mesh)

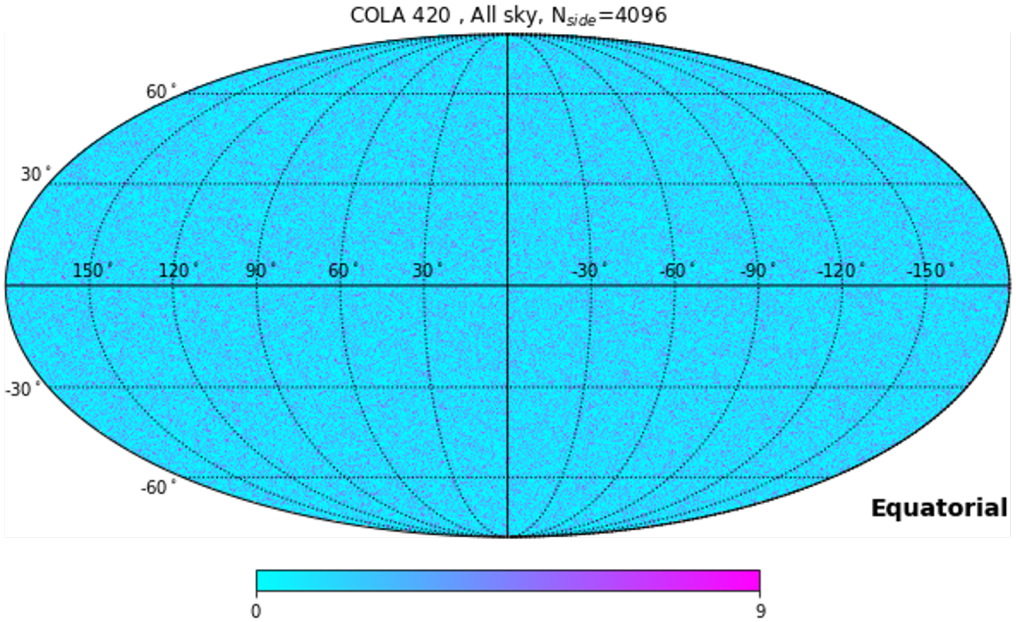



Tassev, Zaldarriaga, Eisenstein, 2013

- Orders of magnitude faster than a high-force accuracy N-body run
- With COLA one can straightforwardly trade accuracy at small-scales in order to gain computational speed, without sacrificing the accuracy at large scales

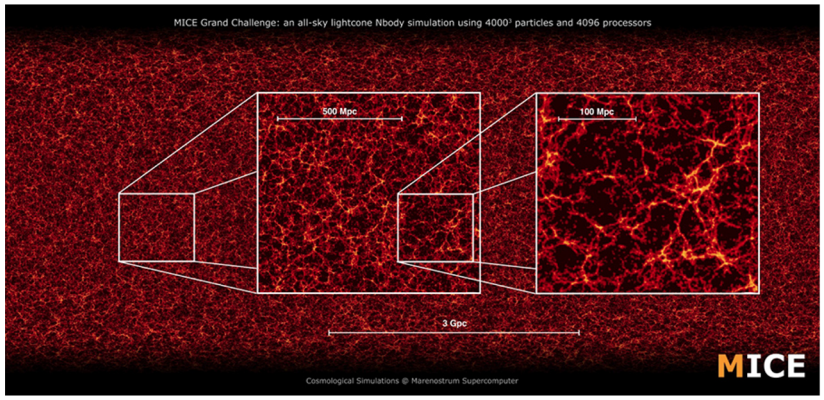
# What do we have?

- 488 full sky runs.
- Lbox=1536, npart=$2048^3$, LC on the fly from z=1.42, MICE cosmology.
- Mass particle = 2,927 x $10^{10}$ ---> Min. halo mass = 5,84 x $10^{11}$
- FoF lightcone: nº of particles, positions and velocities
- Healpix maps: particle counts, shear and kappa



# MICE grand challenge lightcone simulation

Marenostrum Institut de Ciències de l'Espai Simulations

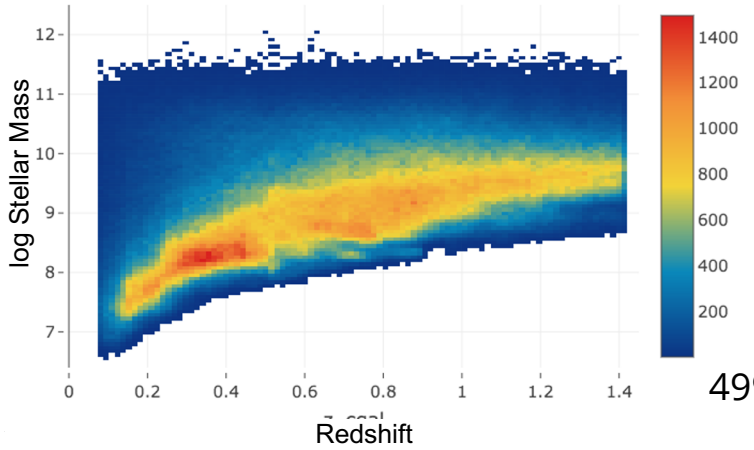



DARK ENERGY SURVEY

Uses MICE Grand Challenge simulation:

- 70 billion particles,
- 3 Gpc/h box,
- mp=3x10^10 M

Full-sky Lightcone without repetition to z=1.4



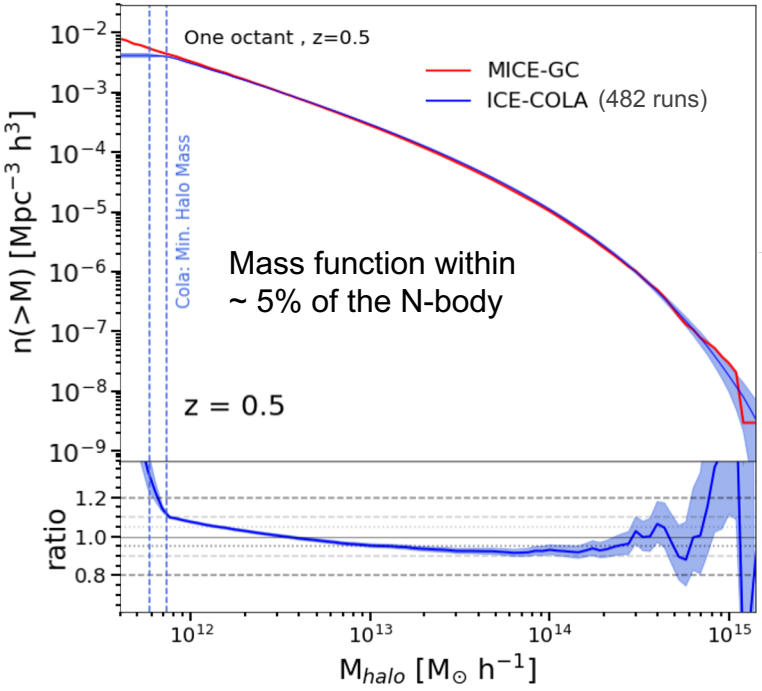

One Octant (~5000 deg2) filled with galaxies:

- positions, velocities
- luminosities
- colours
- galaxy properties
- lensing

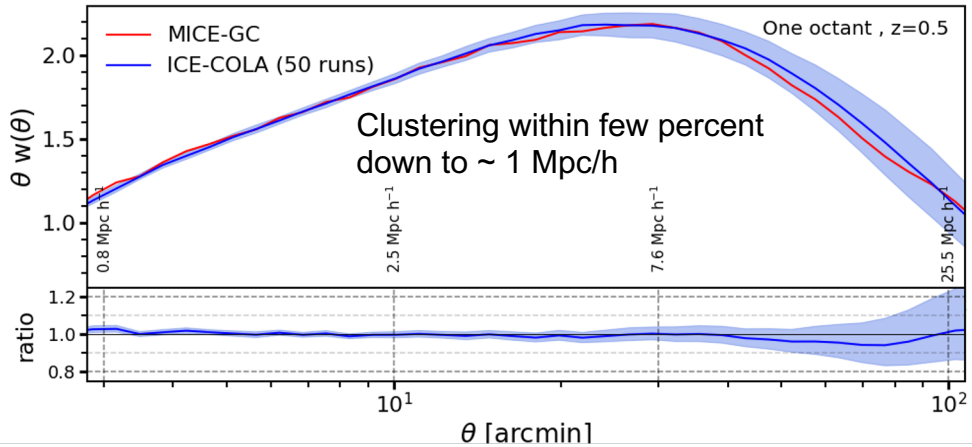
499.61 Millions of galaxies

# Cola validation

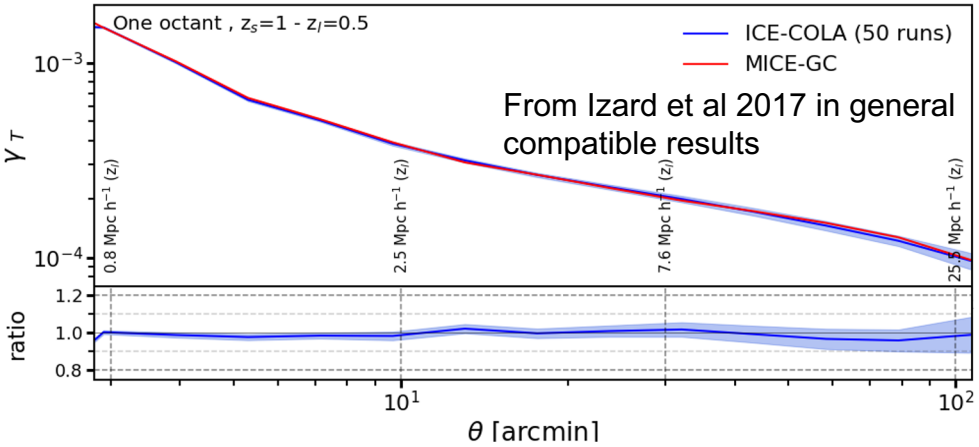
## Halo Mass function at z=0.5




## Pos-pos 2pcf at z=0.5

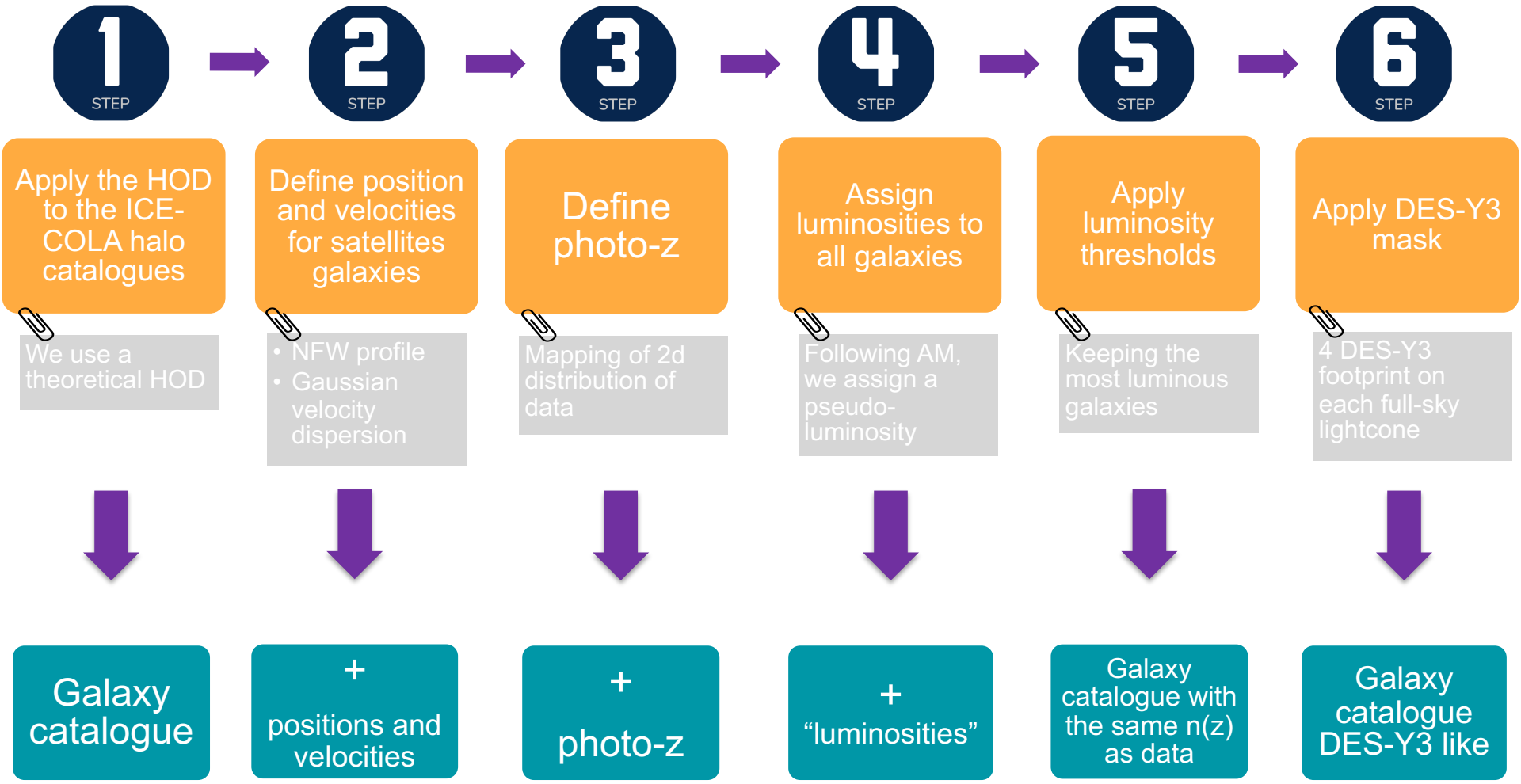



## Pos-shear 2pcf at $z_l$=0.5

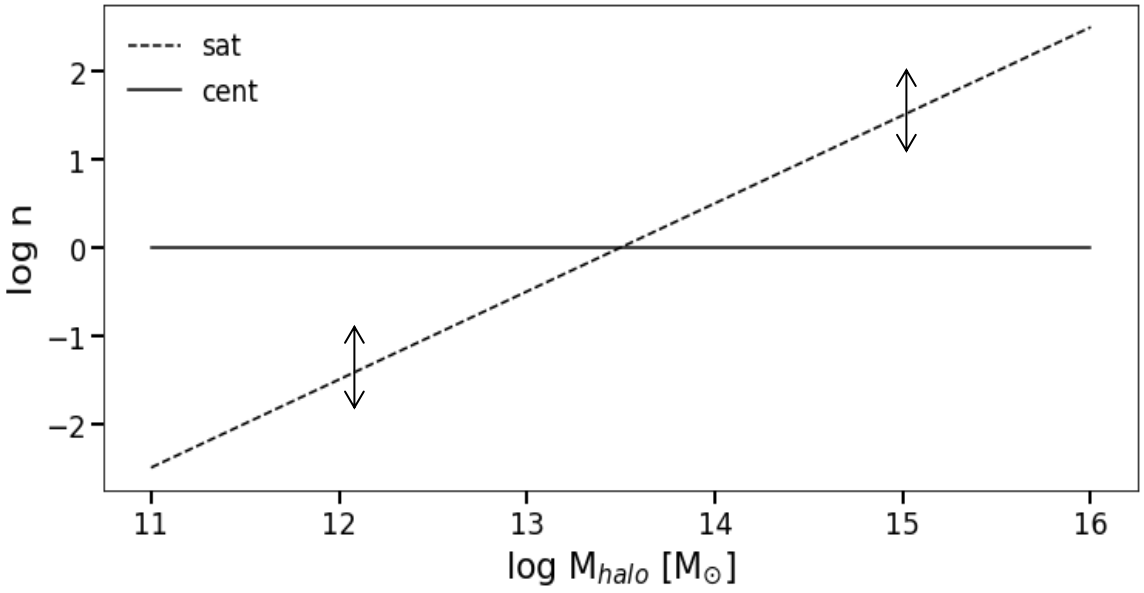



+ shear-shear
+ cls
+ also z=1

# Steps to make the catalogues



# Theoretical HOD for BAO sample




Simple HOD. Only 1 parameter

$$N_{\rm cent} = 1 \ , \qquad \langle N_{\rm sat} \rangle = \frac{M_h}{M_1} \ ,$$

**This can be improved by selection the BAO sample from MICE simulation and measuring the HOD**

# Satellites galaxies in place

## Positions

- Following a NFW profile, assuming triaxiality
- Input: Halo position and mass, z
- Output: (x,y,z)

## Velocities

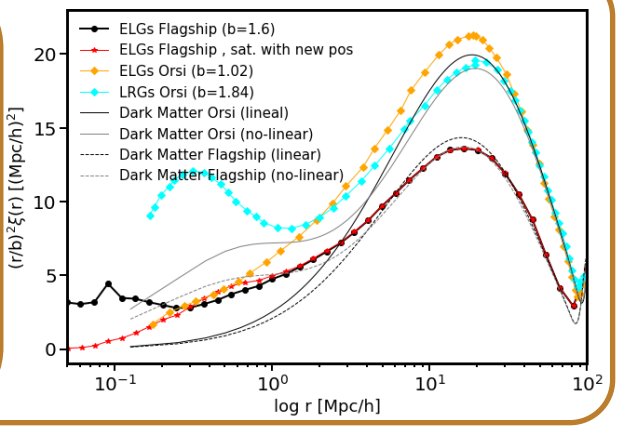
- Assuming a gaussian distribution
- Input: Halo velocity and mass, z
- Output: (Vx,Vy,Vz)

EUCLID project

**Modeling LRGs and ELGs in galaxy mocks**
Needed to provide a more realistic description of these samples

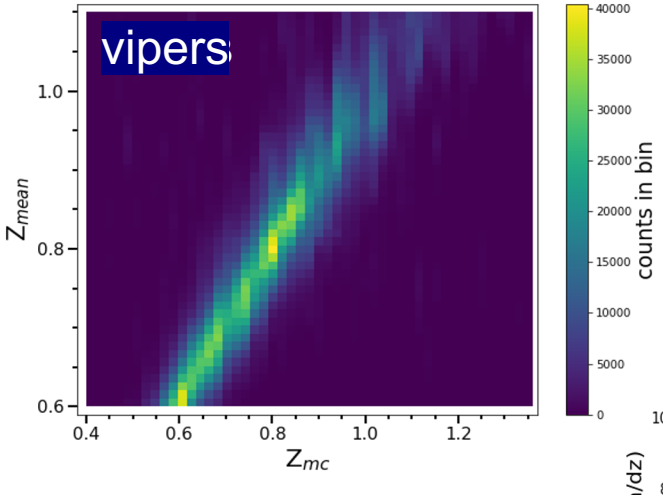
Analysing the way in how we place satellites and how this impact on real/redshift space and on cosmological parameters prediction.

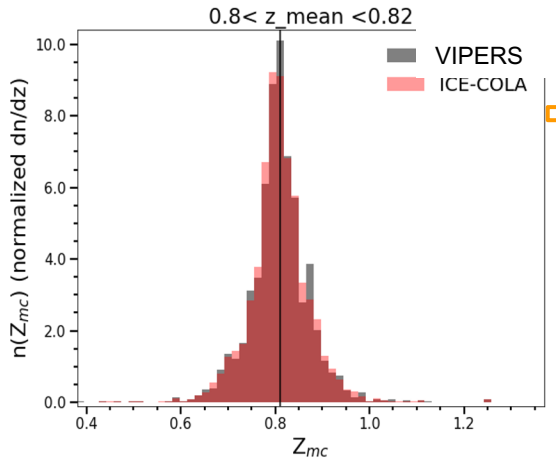
Radial distribution: NFW, triaxiality, concentration, ….
Velocity structure: single Gaussian distribution, fit to line-of-sight velocity distribution, a two-Gaussian fit,...
Dependency of HOD with environment

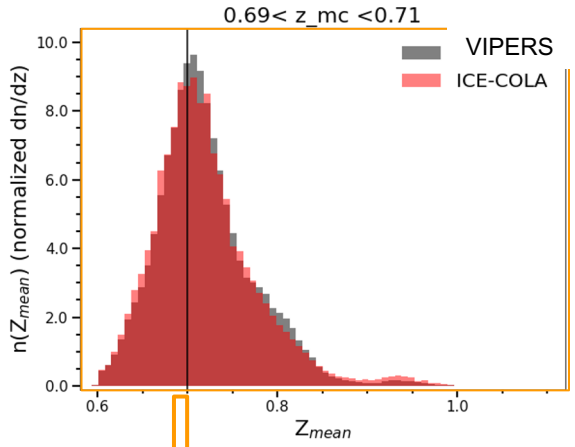


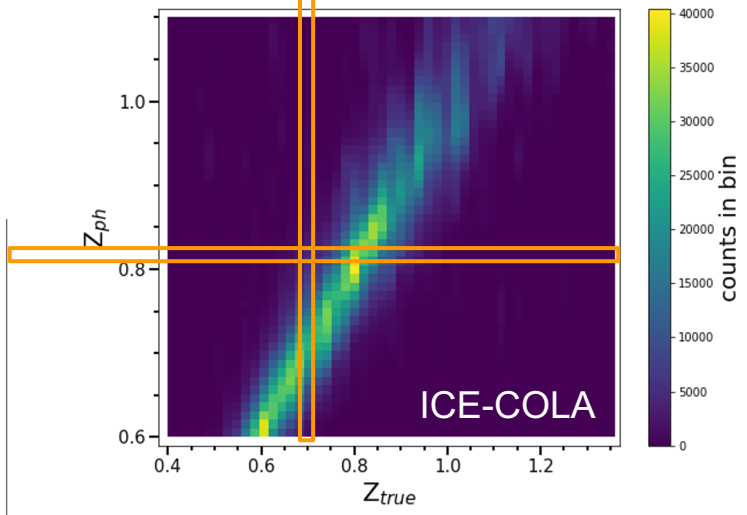
# Modeling of Photo_z (mapping of 2d distribution)

For ICE-COLA we have only the true redshift (z_true) and we need to assign photo_z, to do it we follow the 2d distribution of data (Zmc,Zmean):



# 4 STEP Assign luminosities

Following the concepts of Halo Abundance Matching, we assign a pseudo-luminosity lp to the galaxies. This lp is not intended to represent a realistic luminosity.

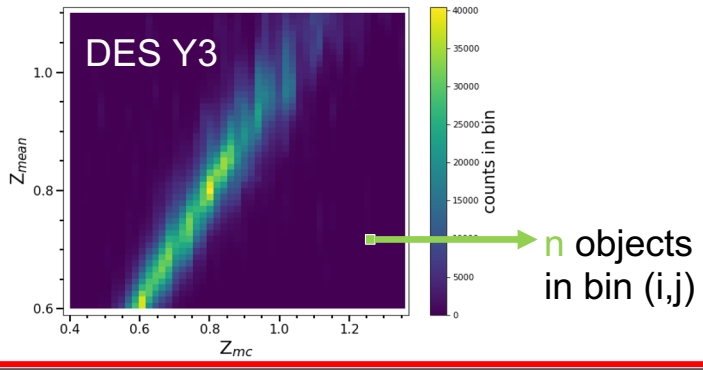
It is used to set the thresholds to match the abundance of the data in a simple way.

$$\log_{10}(l_p) = \log_{10}(M_h) + \Delta_{LM} \cdot R^{\text{gauss}}_{\mu=0\ \sigma=1}$$

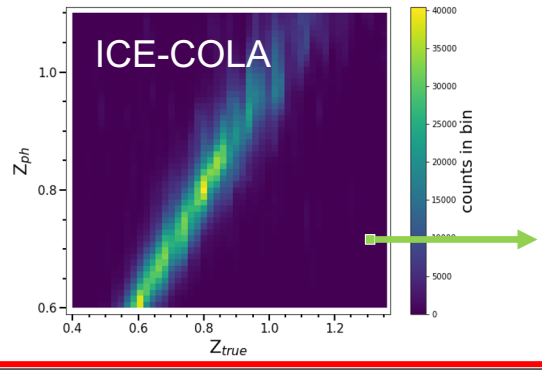
We model lp (in arbitrary scales) with a Gaussian scatter around the halo mass Mh in logarithmic scales, where ΔLM is a free parameter of the HAM model that controls the amount of scatter.

# 5 STEP Apply luminosity thresholds

As the number of galaxies in the simulation are higher than in the survey, the abundance is then fixed by setting luminosity thresholds




n objects in bin (i,j)

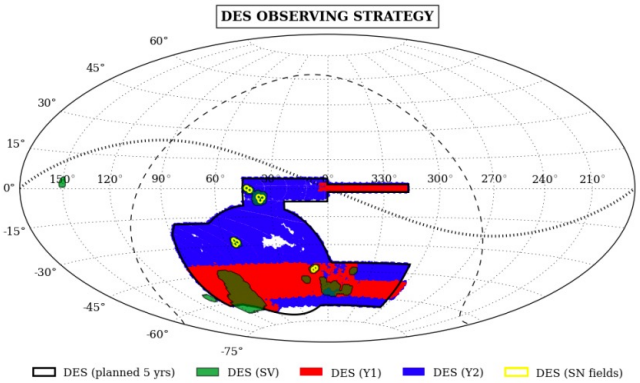


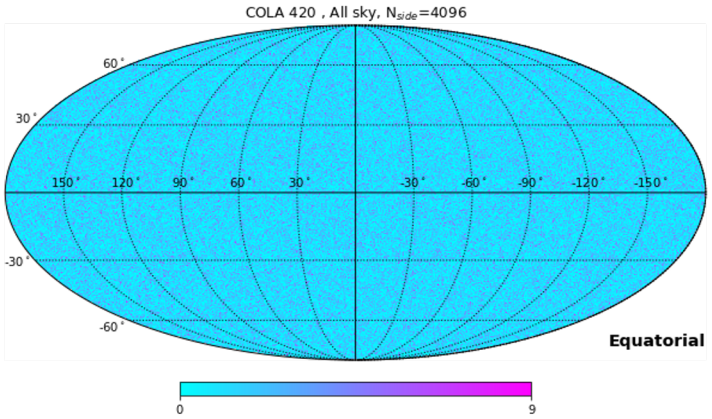

Then I keep the n objects most luminous of the bin (i,j)

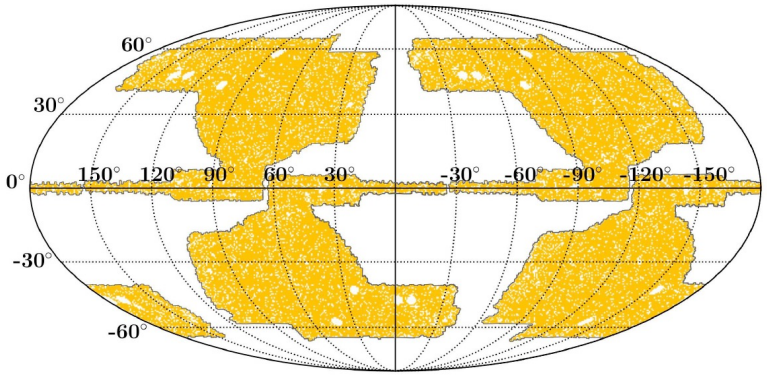
# Apply DES-Y3 mask

DES-Y3 footprint

BAO Area: 4148 deg$^2$



It allows us to have 4 mock catalogues for each ICE-COLA full sky lightcone

488 ICE-COLA seeds x 4 → **1952** DES-Y3 like galaxy mocks

# Theoretical HOD for BAO sample: Calibration (M1, ΔLM) ?

**Problem:**

- HOD parametrization need to be calibrated to match observed galaxy properties
- Hard to do “manually” when we have many parameters (2 x 5 z-bins = 10 parameters)

**Goal: Do the calibration in an automatized way**

A. Read array with the values of HOD parameters

B. Generate galaxy mock from halo catalog with given parametrization

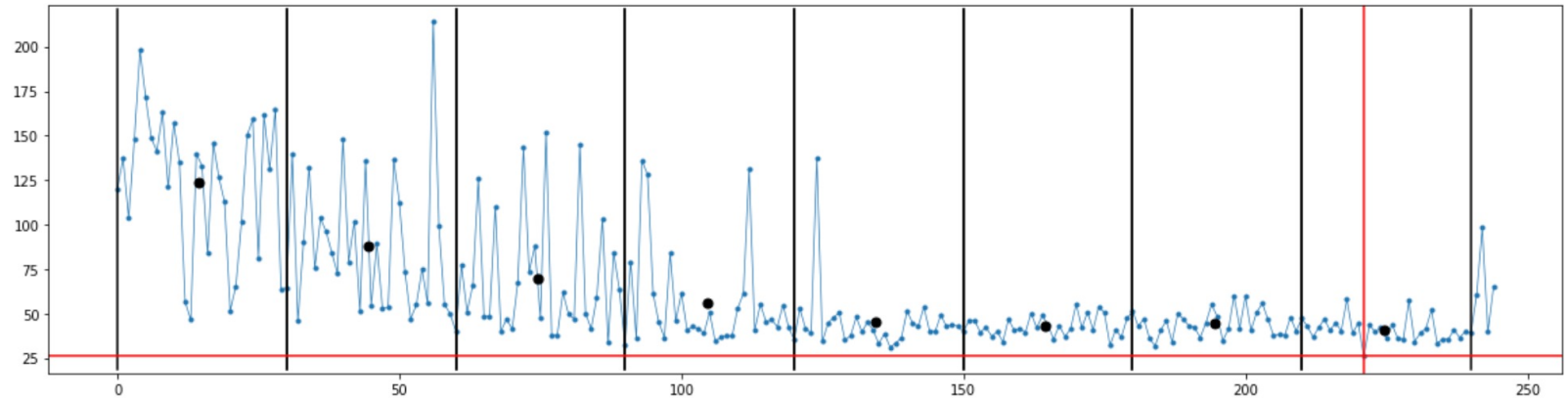

C. Compute clustering (or any other observable)

D. Compare to observational data vector (chi2)

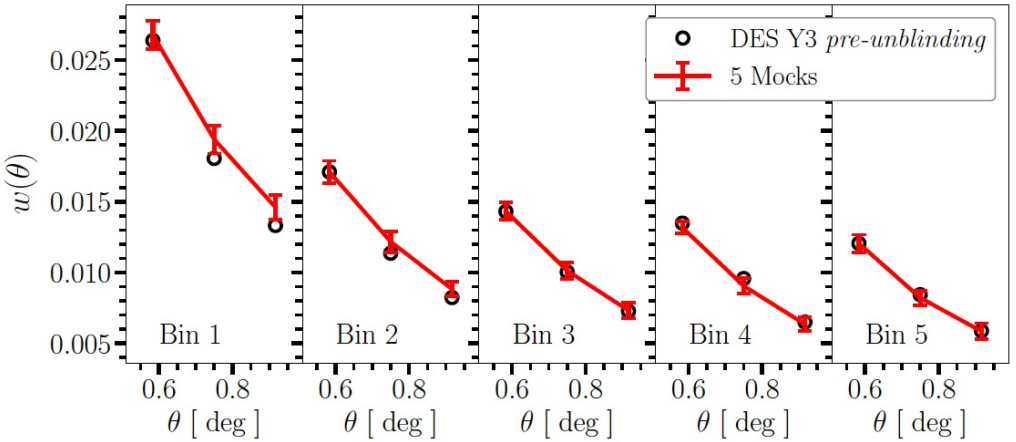
E. Select a different point in the parameter space and re-iterate until best-fit values are found

# Theoretical HOD for BAO sample: Calibration

Evolution of chi2 with iterations

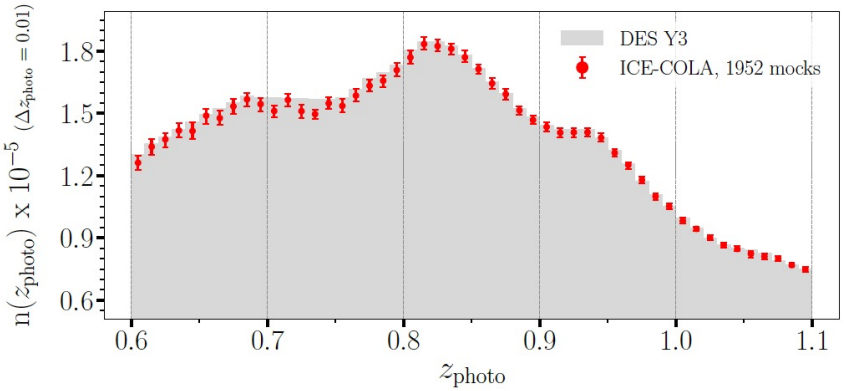


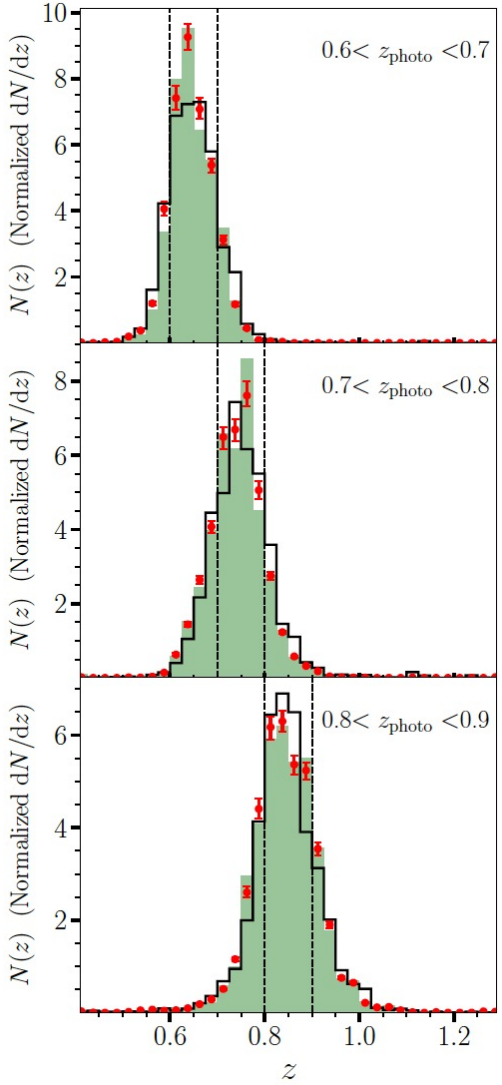

The input for calibration are the measurements of the angular two point correlation function for 3 different Θ for each tomographic bin.

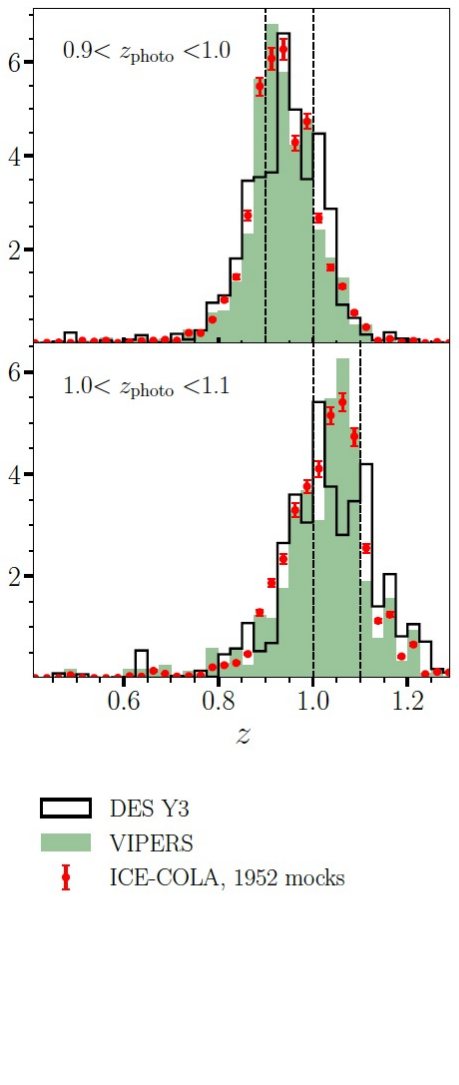


# Results

Redshift Distribution of mocks compared with data

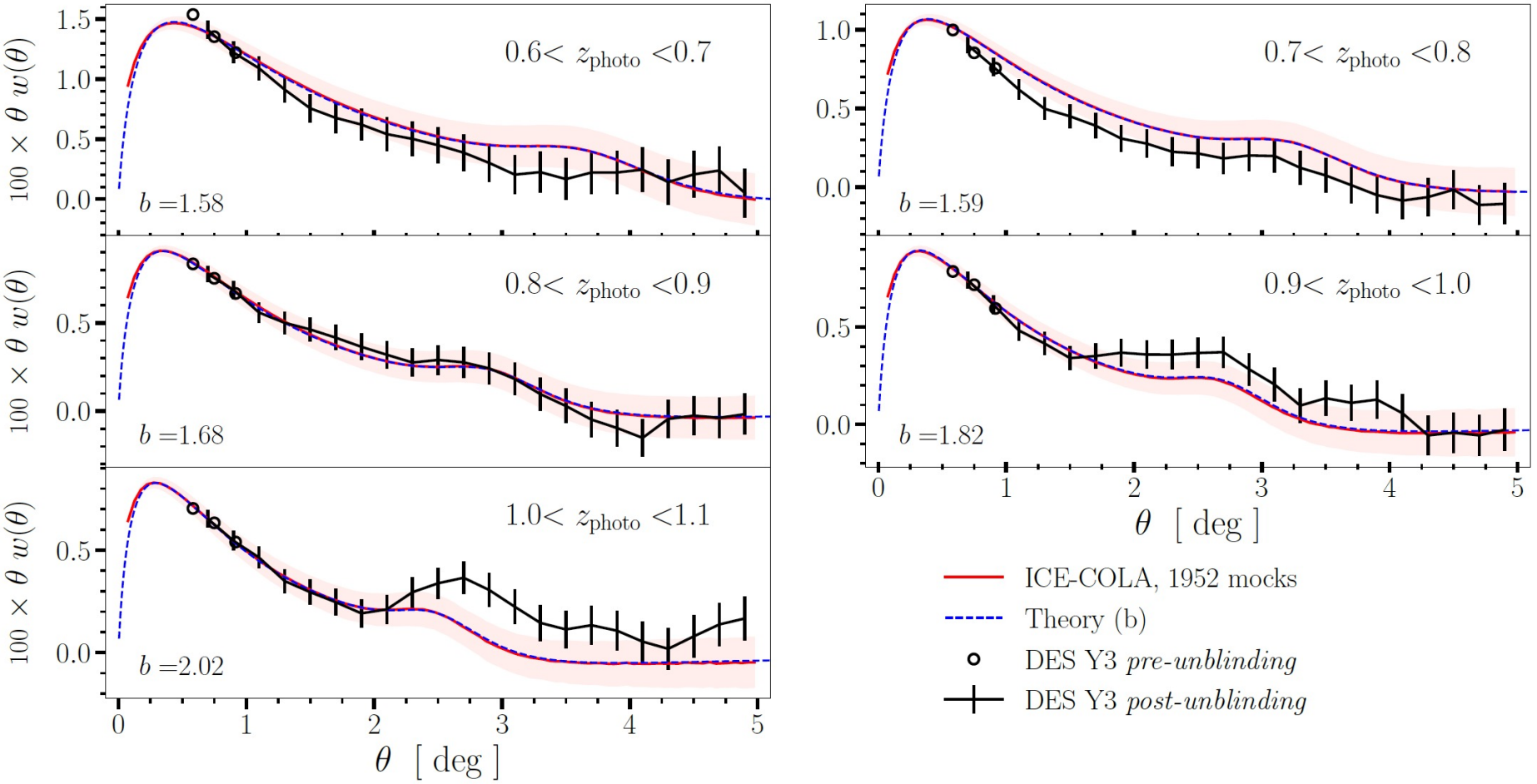


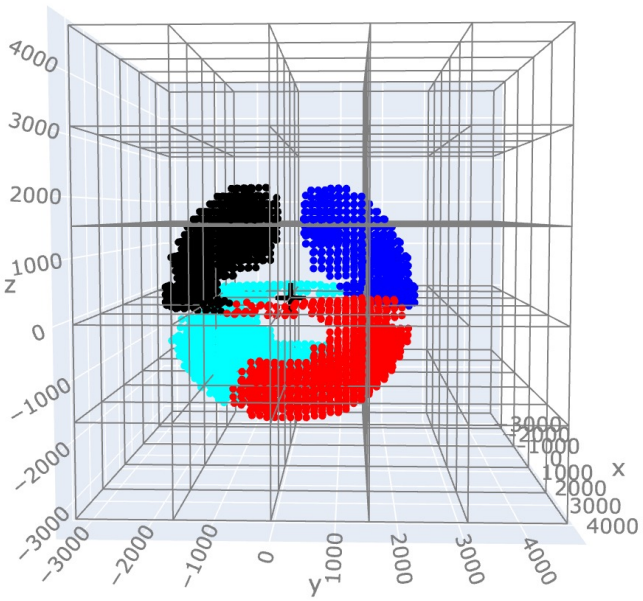



# Results

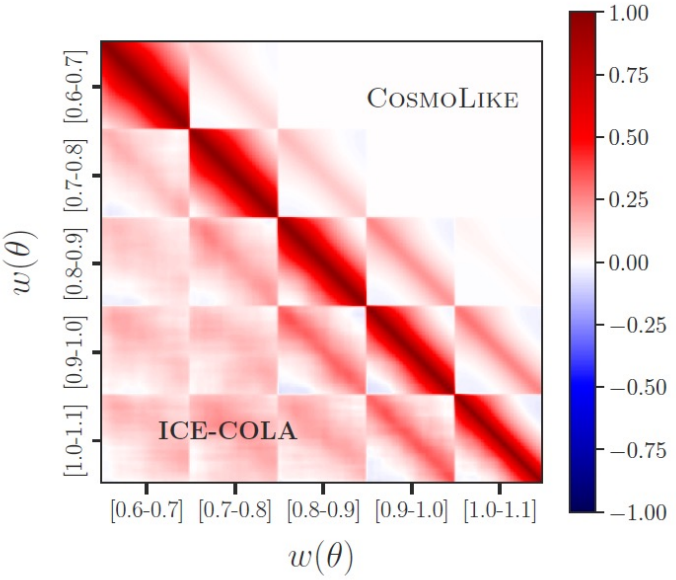
Angular 2pfc compared with calibration data and theoretical prediction

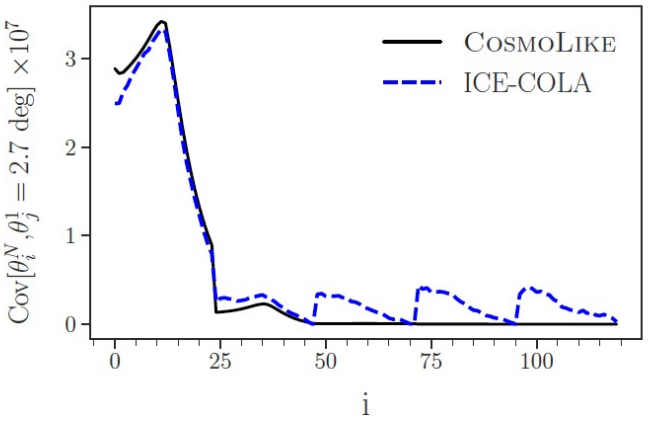


# Replication problems




| BIN | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 3.9 (22.6) | 9.7 (30.6) | 11.9 (30.2) | 15.9 (30.3) | 17.8 (30.8) |
| 2 | - | 5.9 (27.9) | 13.9 (28.7) | 15.7 (29.1) | 17.7 (28.9) |
| 3 | - | - | 8.1 (28.7) | 16.5 (29.2) | 18.5 (28.1) |
| 4 | - | - | - | 8.9 (30.7) | 18.2 (27.4) |
| 5 | - | - | - | - | 10.1 (29.9) |



# Summary

- Galaxy mock catalogues are an indispensable resource for cosmological data analysis, those are a “cheap” way to simulate samples of observed galaxies.
- Using fast simulation to create masive numbers of halo catalogues allow for robustly estimating covariance matrices or for propagating errors in complex and non-linear analysis.
- This pipeline can be used for any survey, for any kind of sample and using any reference catalog or HOD. It also allows to add extra modules to calculate galaxy properties.
- The automatic calibration is the key to get mock catalogues, whose match the observables.
- The shown ICE-COLA mocks for BAO (DES) is a clear example of the robustness of the method.
- It is crucial to make an exhaustive analisis of the impact of structures replications on covariance.