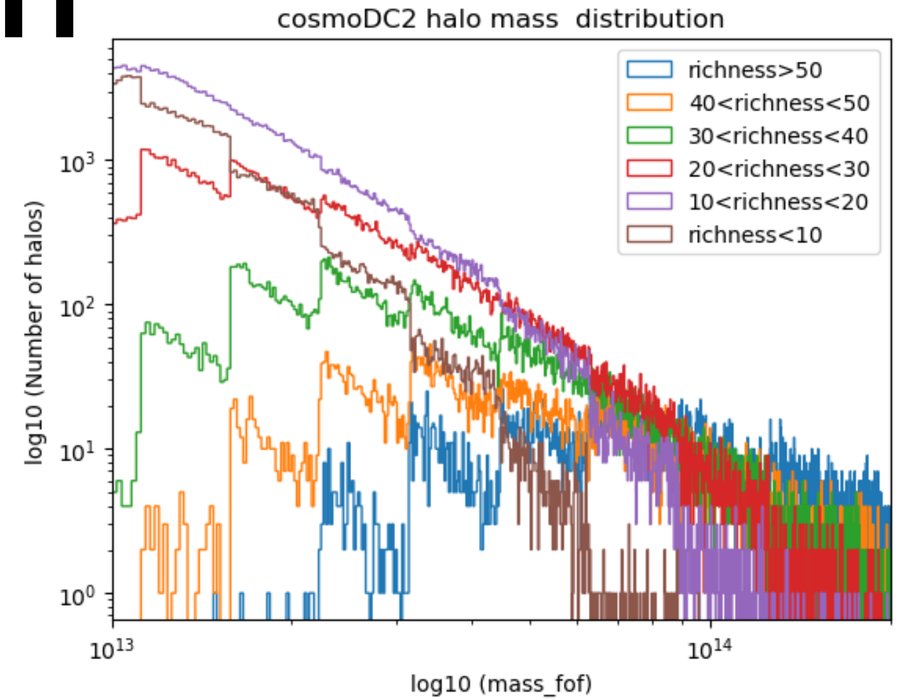
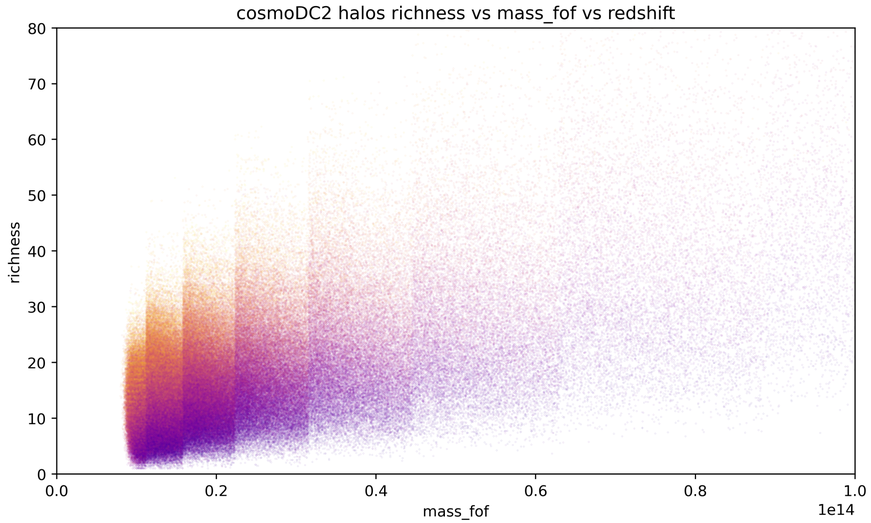
Comparison between different LSST catalogs

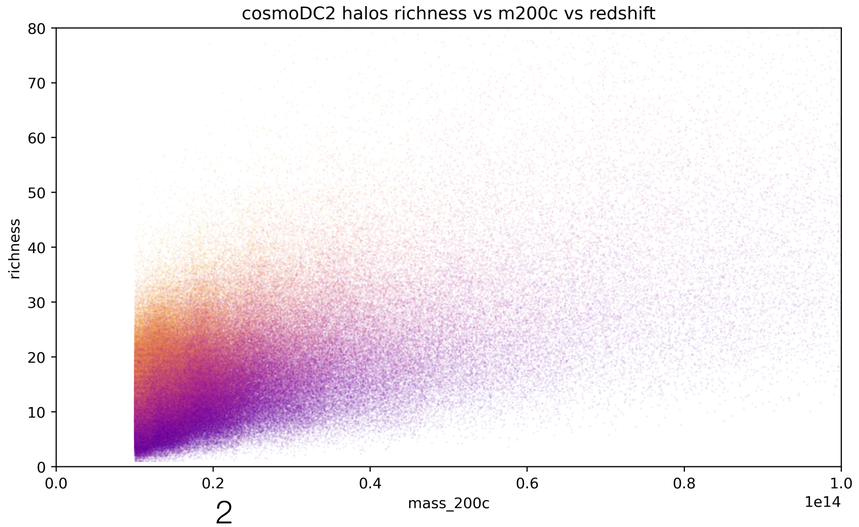
Narei Lorenzo Martinez
with help from Vincent (thanks !)

Introduction

- Last time, I found some discontinuities on richness vs mass_fof in the catalog i was studying
  - but not seen when using m200c instead
- Today: investigation on these discontinuities

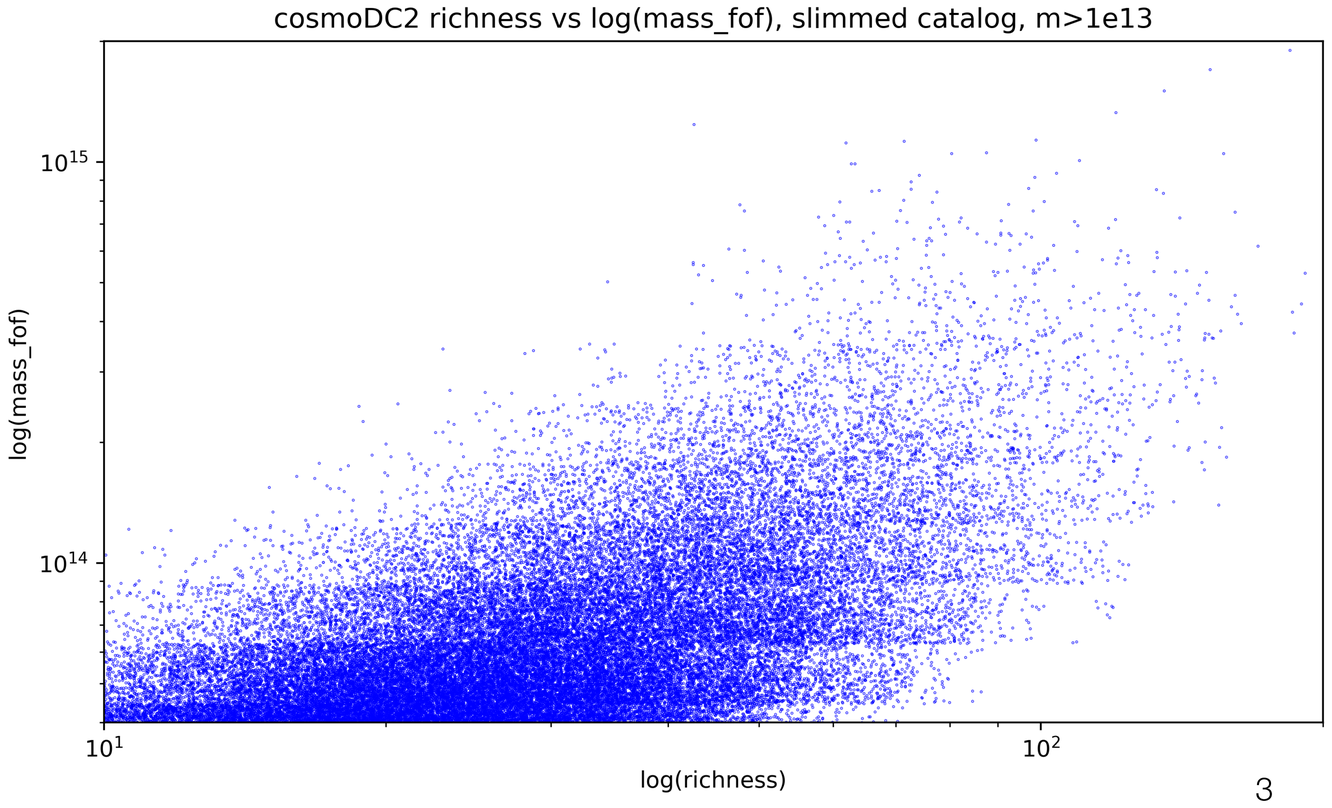


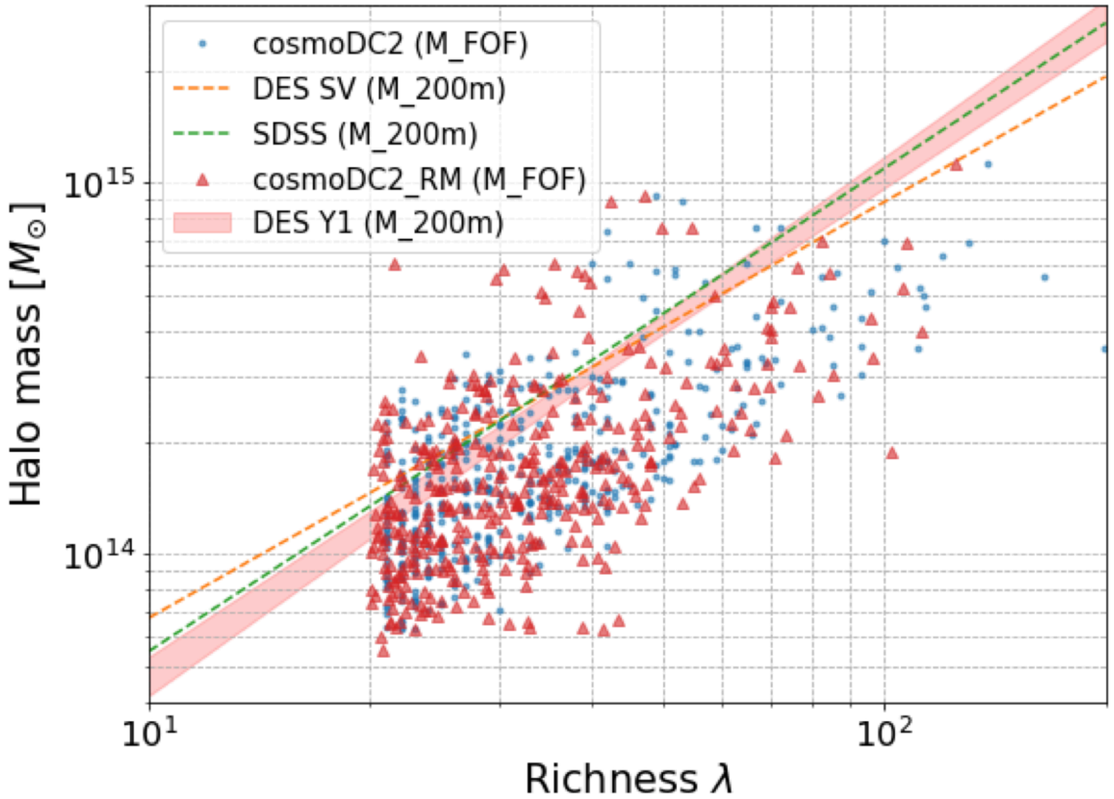




Comparison with validation paper

- Mimicking plot in validation paper : https://arxiv.org/pdf/2110.03769.pdf (figure 21)
  - difference is that galaxies in validation paper are matched to the ones found in RedMapper
  - granularity/stat shown in plot from paper does not allow to see steps in mass_fof ?

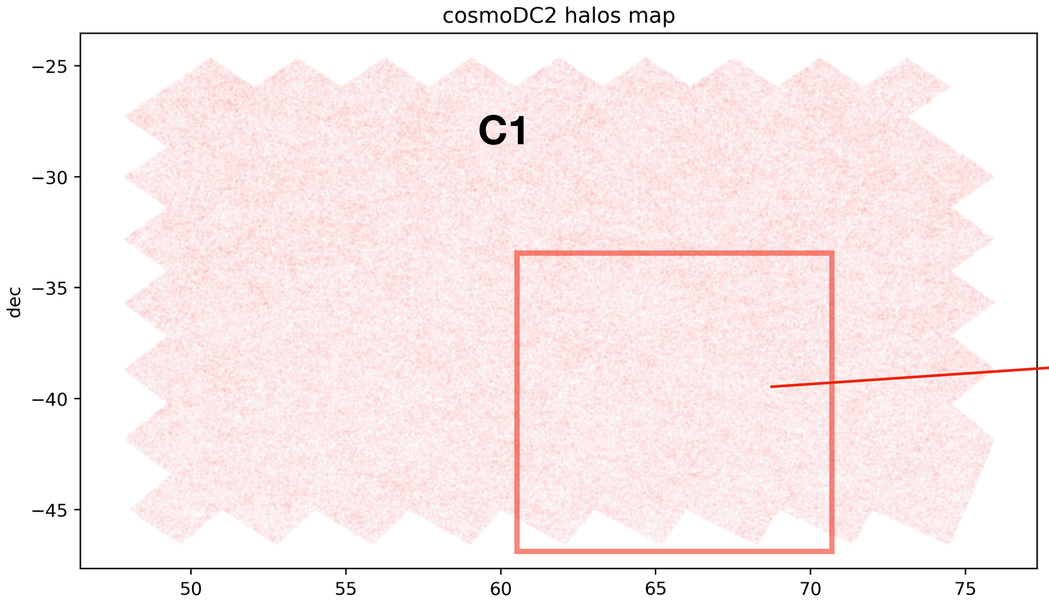


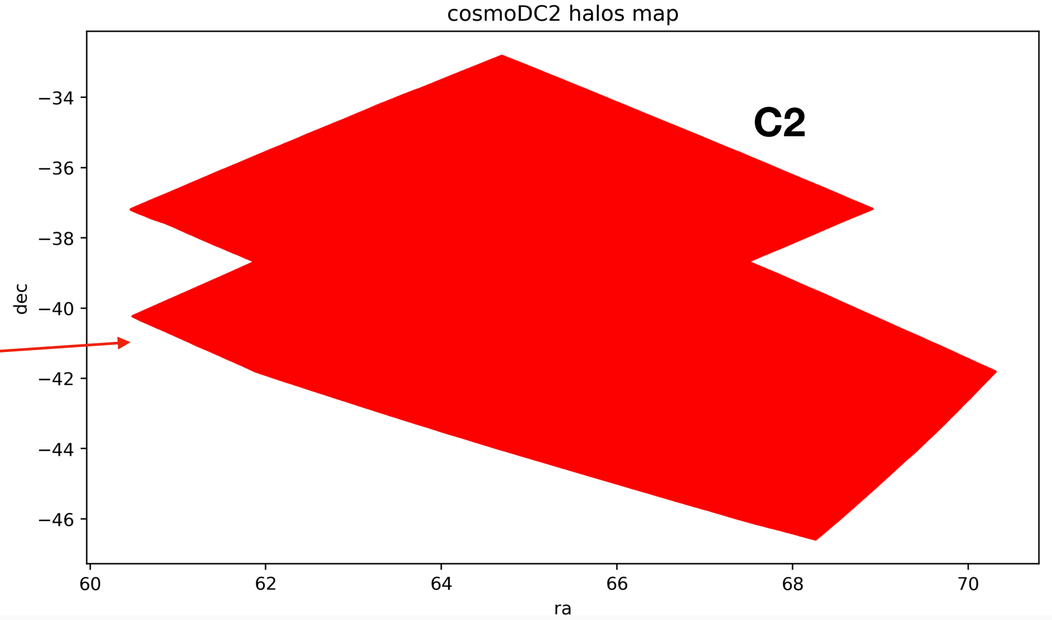


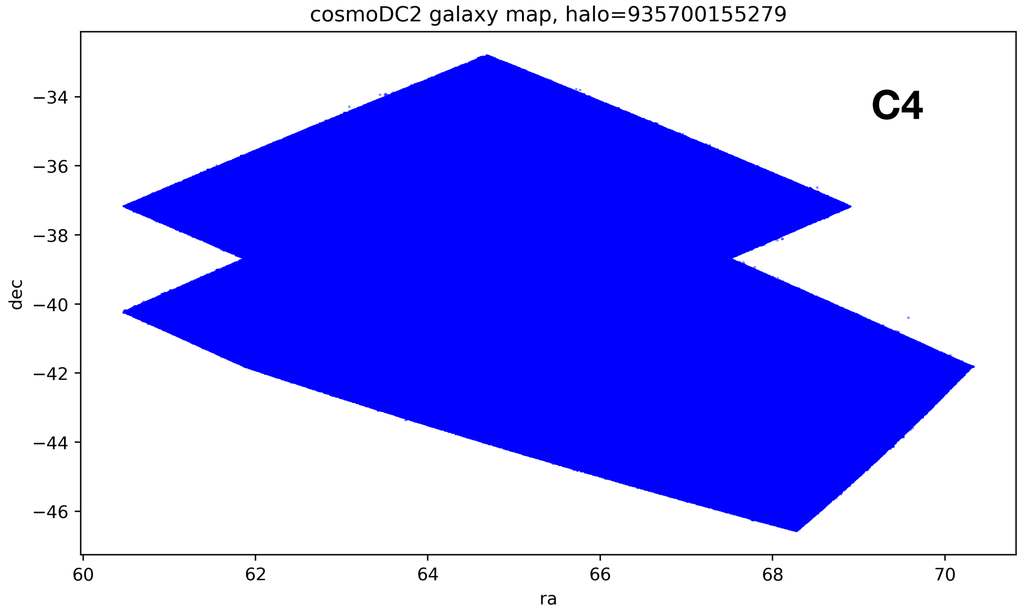
4 catalogs studied

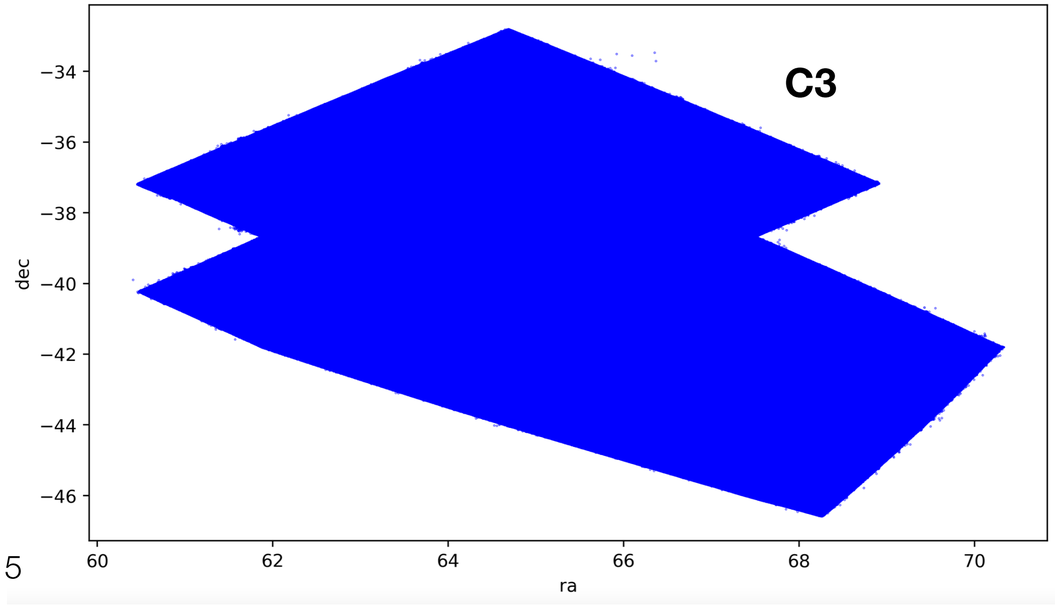
- **C1: CosmoDC2, 440 sq, mass>1e13, fit format** : /sps/lsst/groups/clusters/dc2/**cosmoDC2_v1.1.4/extragal/halos/halos_m200c_13.0.fits**
  - halo variables: ['halo_id', 'ra_true', 'dec_true', 'ra', 'dec', 'redshift_true', 'mass_fof', 'm200c', 'r200c', 'skysim_halo_id', 'NMEM', 'richness', 'NMEM_g_star2', 'NMEM_r_star2', 'NMEM_i_star2', 'NMEM_z_star2', 'NMEM_y_star2', 'richness_g_star2', 'richness_r_star2', 'richness_i_star2', 'richness_z_star2', 'richness_y_star2', 'ra_bary', 'dec_bary']
- **C2: CosmoDC2, Small part of the sky (~50 sq), original catalog (no cuts), fit format**: /sps/lsst/users/tguillem/debug/for_narei/Catalog.fits
  - halo variables: ['id', 'ra', 'dec', 'z', 'mass', 'log_mass']
- **C3: CosmoDC2, Small part of the sky (~50 sq), original catalog (no cuts), GCR (Generic Catalog Reader) format** : https://github.com/LSSTDESC/gcr-catalogs?tab=readme-ov-file
  - catalog : **cosmoDC2_v1.1.4_small**
  - variables : huge number of them, among them: ['ra', 'dec', 'halo_id', 'redshift', 'halo_mass', 'galaxy_id']
- **C4: Skysim5000, small part of the sky (~50 sq), original catalog (no cuts), GCR format**
  - catalog : **skysim5000_v1.2_small**
  - variables : huge number of them, among them: ['ra', 'dec', 'halo_id', 'redshift', 'halo_mass', 'galaxy_id']

Comparison of areas



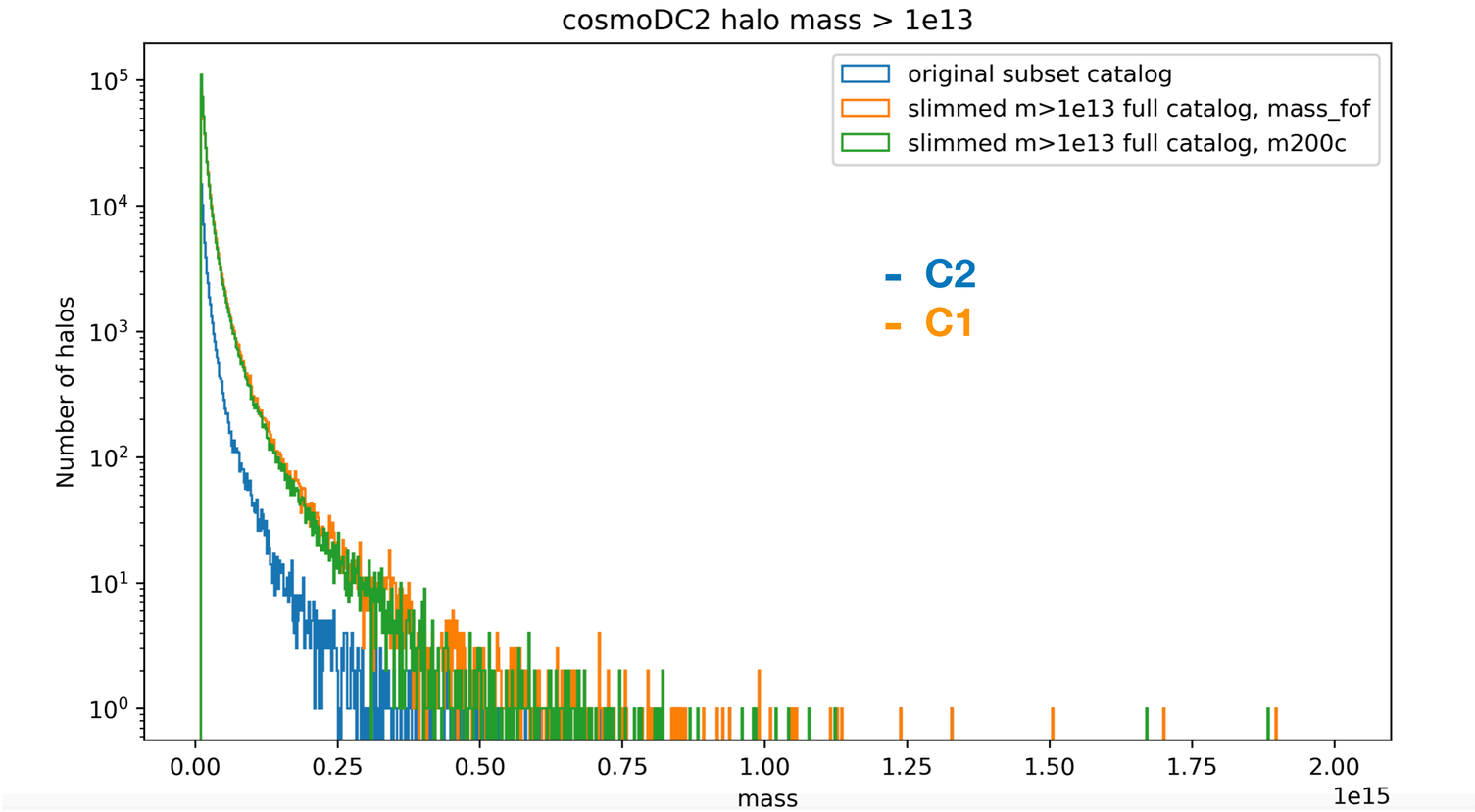






Comparison of C1 and C2
(to understand if discontinuities come from reduction of catalog)

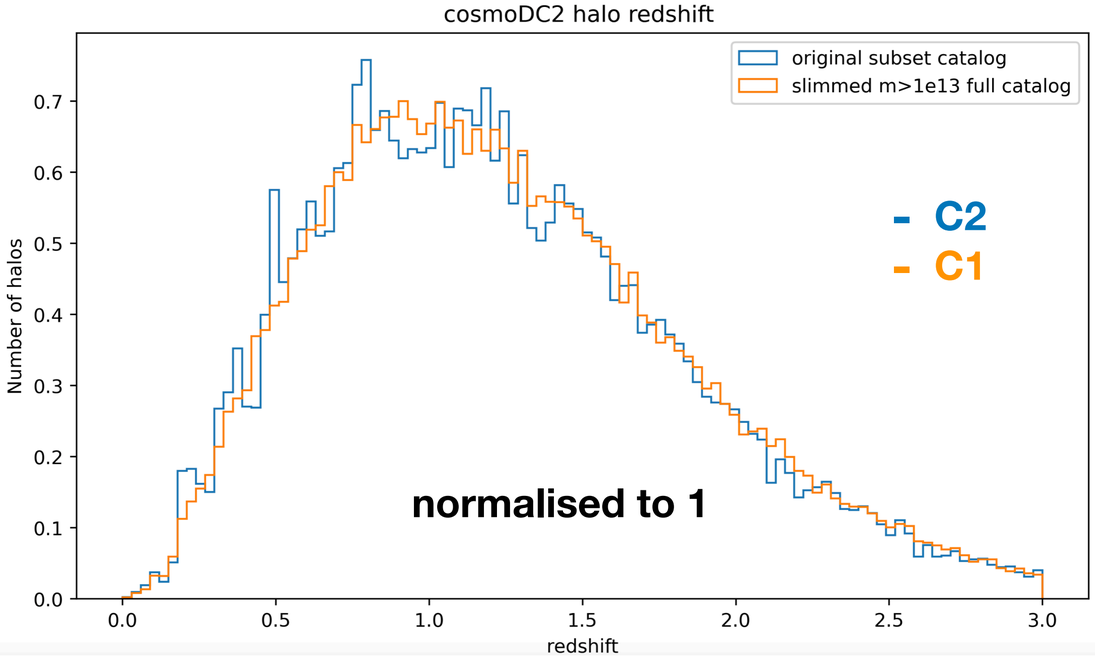
original catalogs, m>1e13

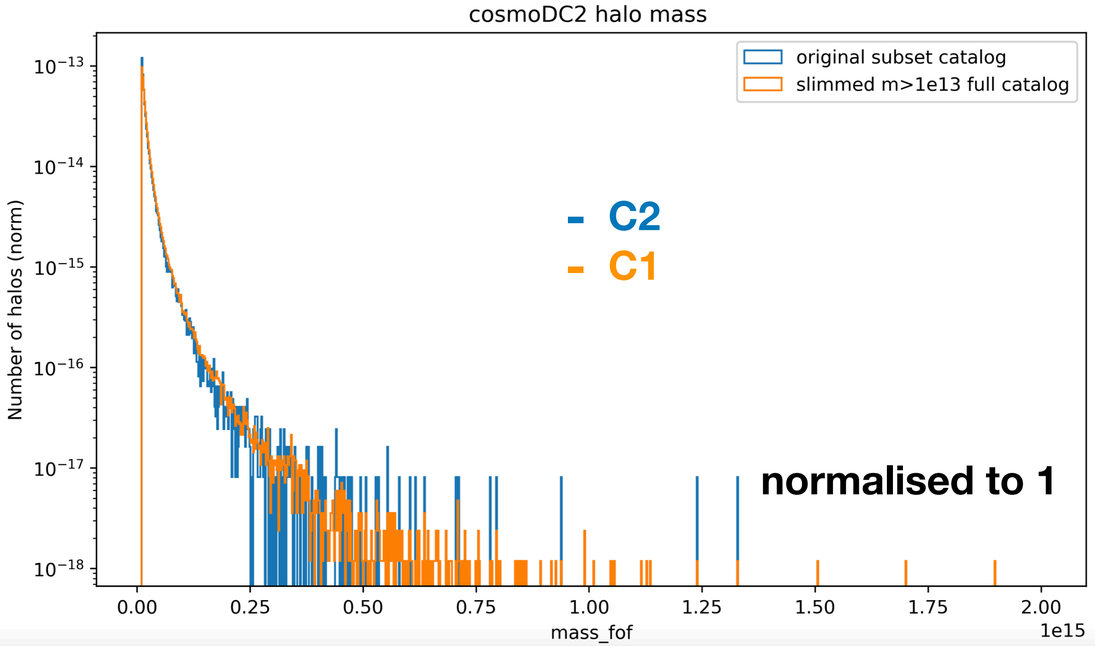


- In C2 (and C3), only mass_fof is available

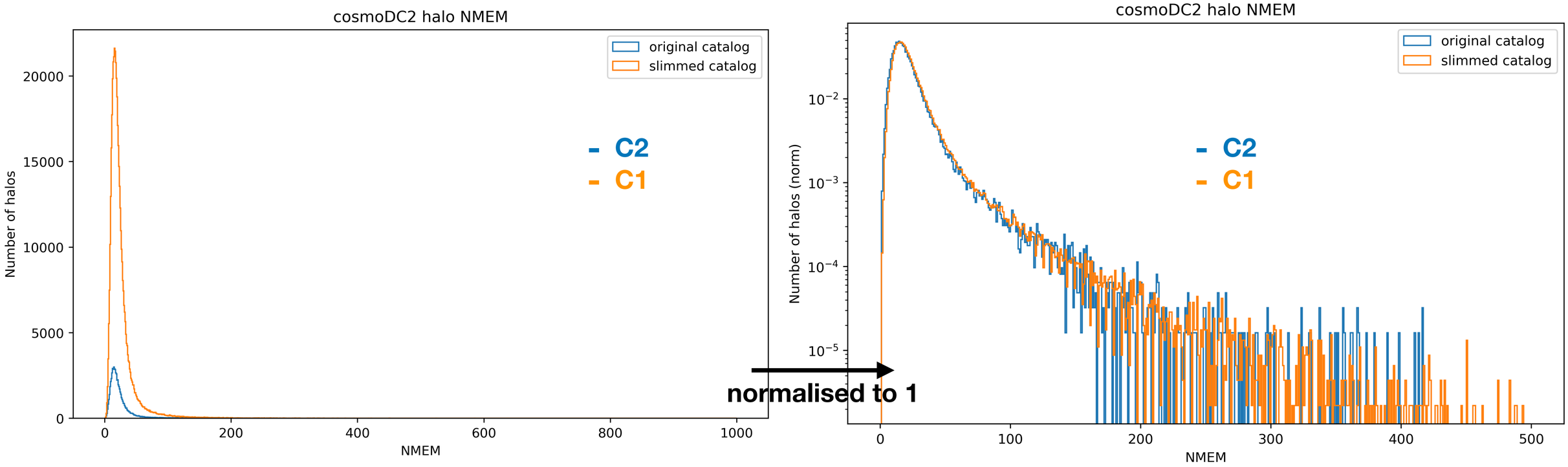
comparison of C1 and C2

- For halos with same ID, checked all quantities (mass_fof, ra, dec, redshift)
  - everything is exactly the same except for **redshift**
    - small difference -> why ?





comparison of C1 and C2



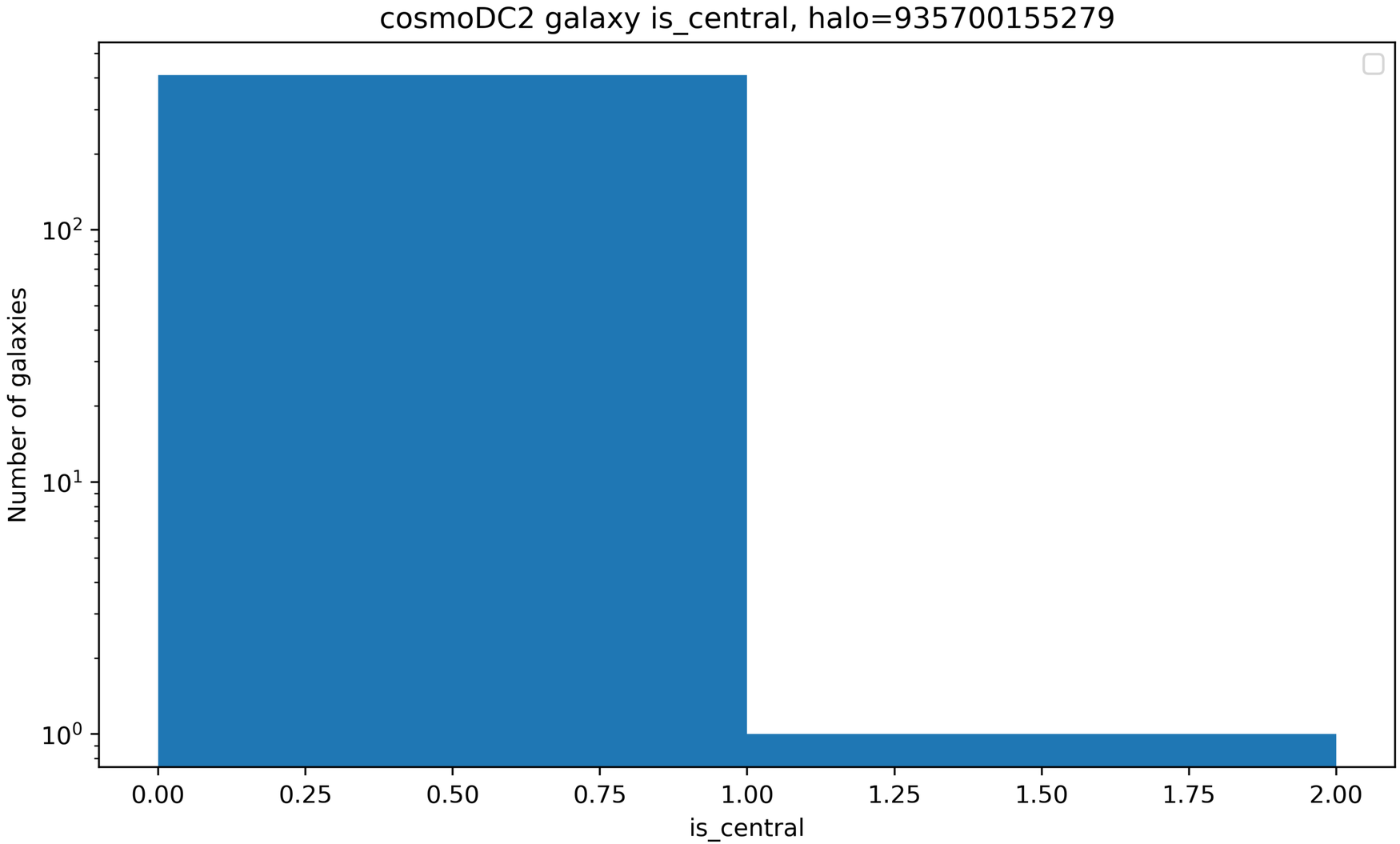
Example of halo= 935700155279

- In original catalog (C2) : 410 galaxy members
- mass (fof) =132834633216865 3.5

One galaxy difference between C1 and C2
Why ?

- In slimmed catalog (C1): 411 galaxy members
- NMEM (cluster) = 411
- richness (cluster) = 120.22
- mass (fof) = 132834633216865 3.5
- mass 200c= 314882894937419.75 (factor >2 diff with mass_fof !)

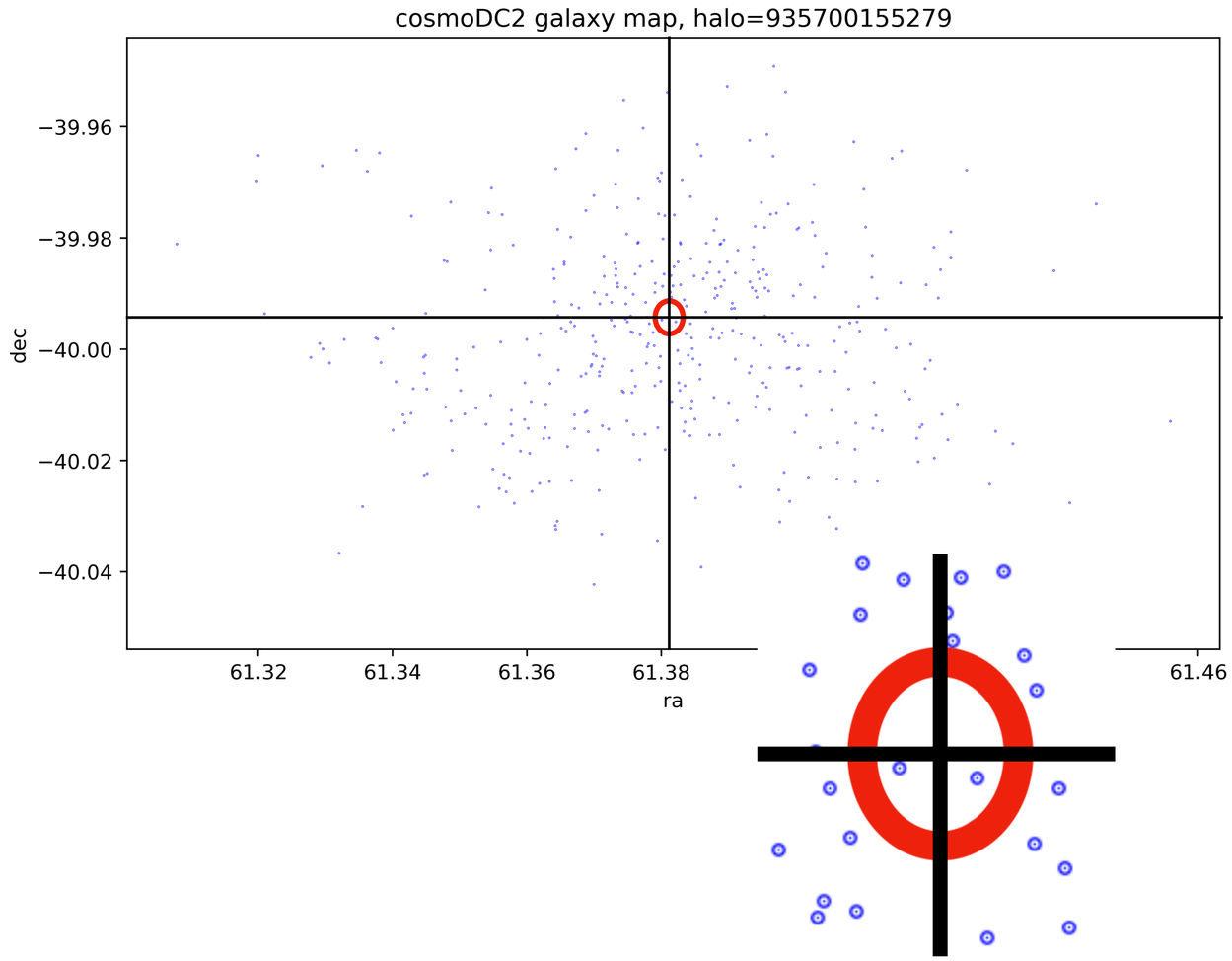
Central galaxy



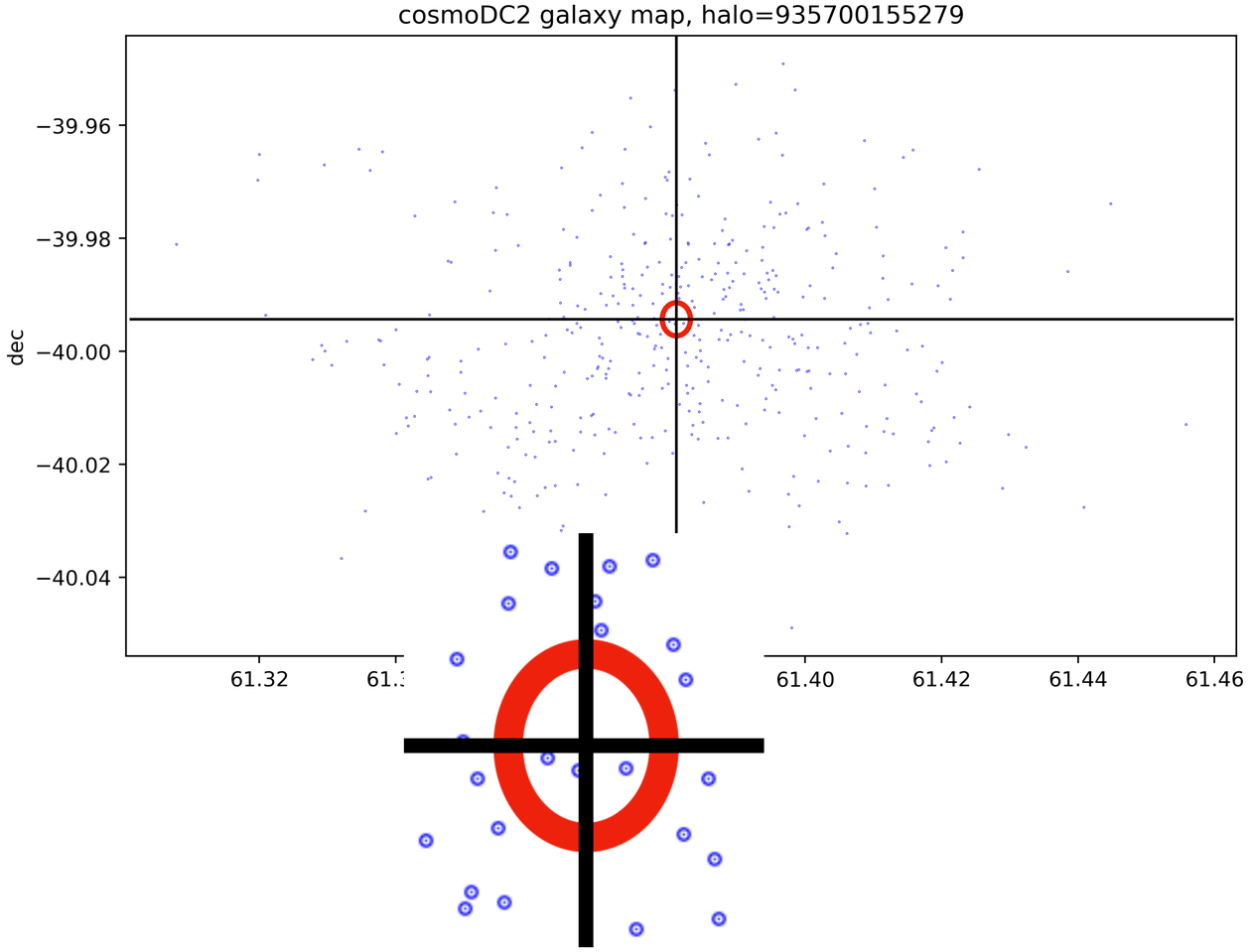
- In slimmed catalog (C1)
- galaxyID isCentral = 9688655685
- (ra,dec) = (61.381, -39.995)

Halo galaxies distribution

original catalog (C2)

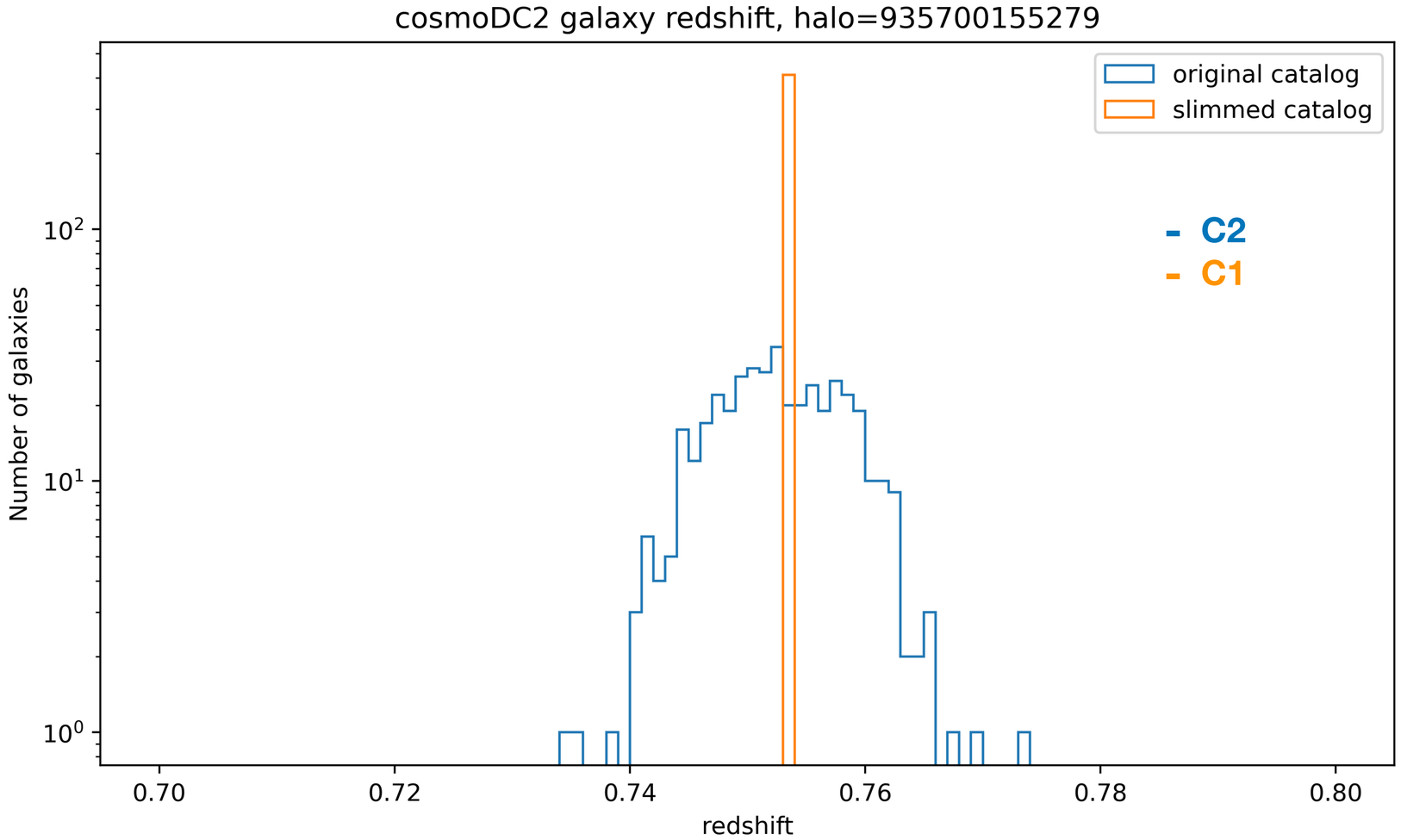


slimmed catalog (C1)



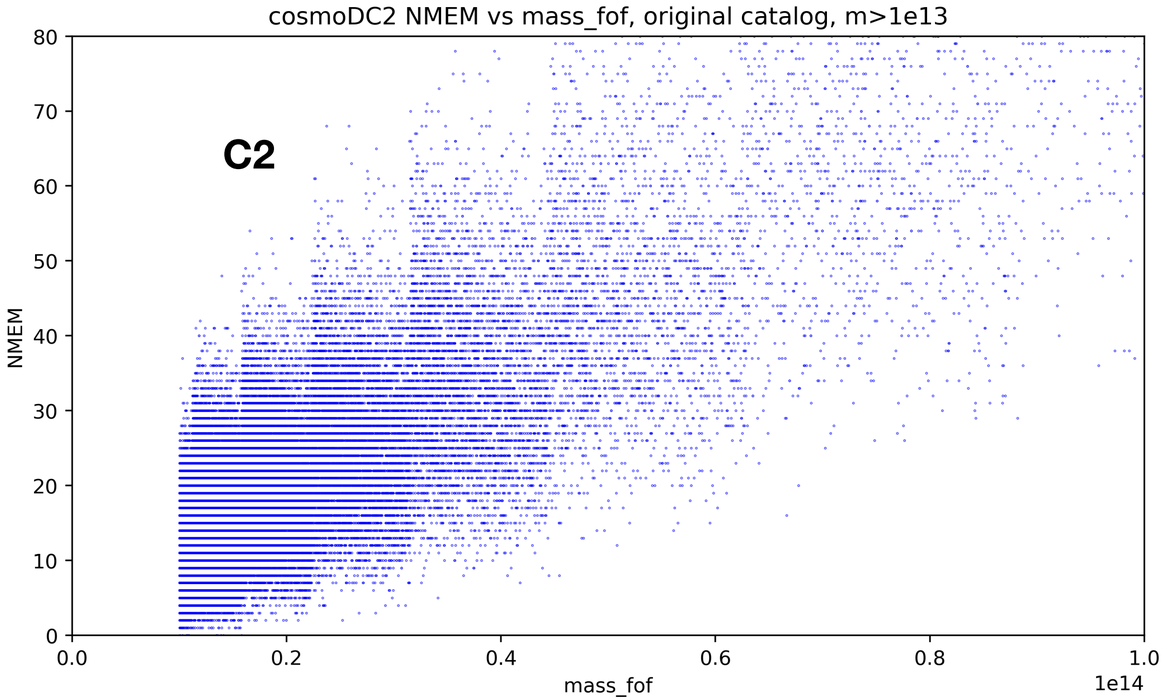
The missing galaxy in C2 is the central galaxy -> added by hand in slimmed catalog ?

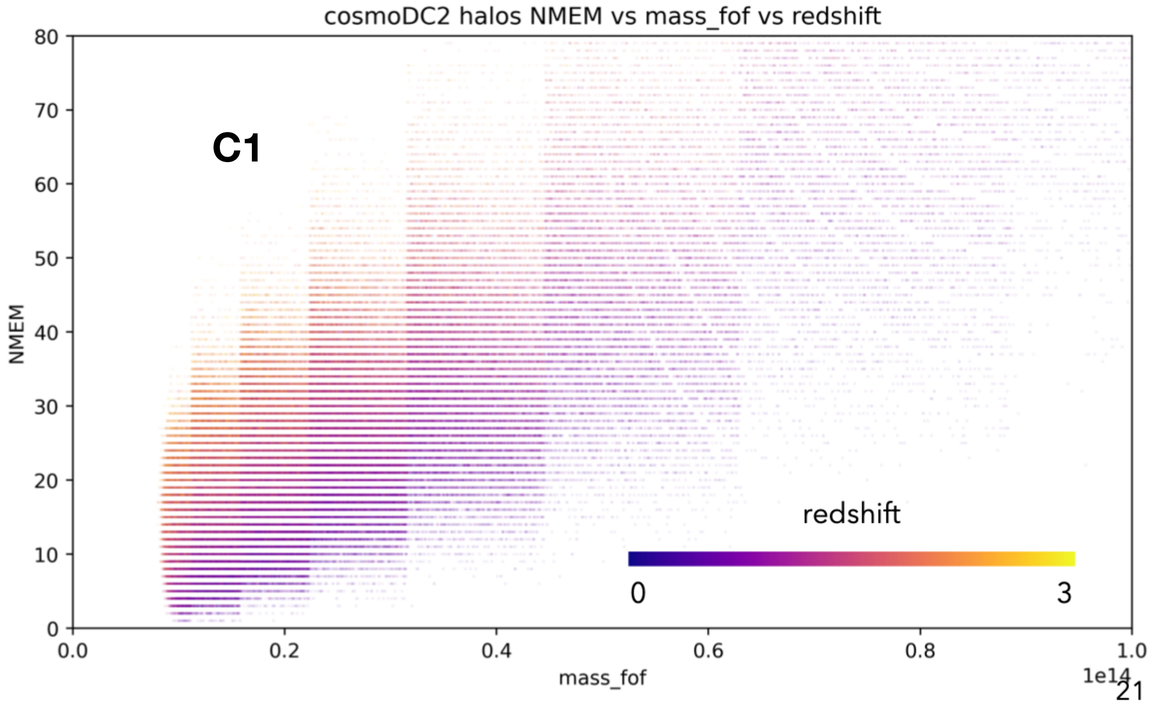
Redshift distribution for chosen halo



- in original catalog (C2), galaxies have different redshifts
- In C1, galaxies redshift re-attributed to be the halo's one ?
- how is redshift calculated for galaxies and cluster ?

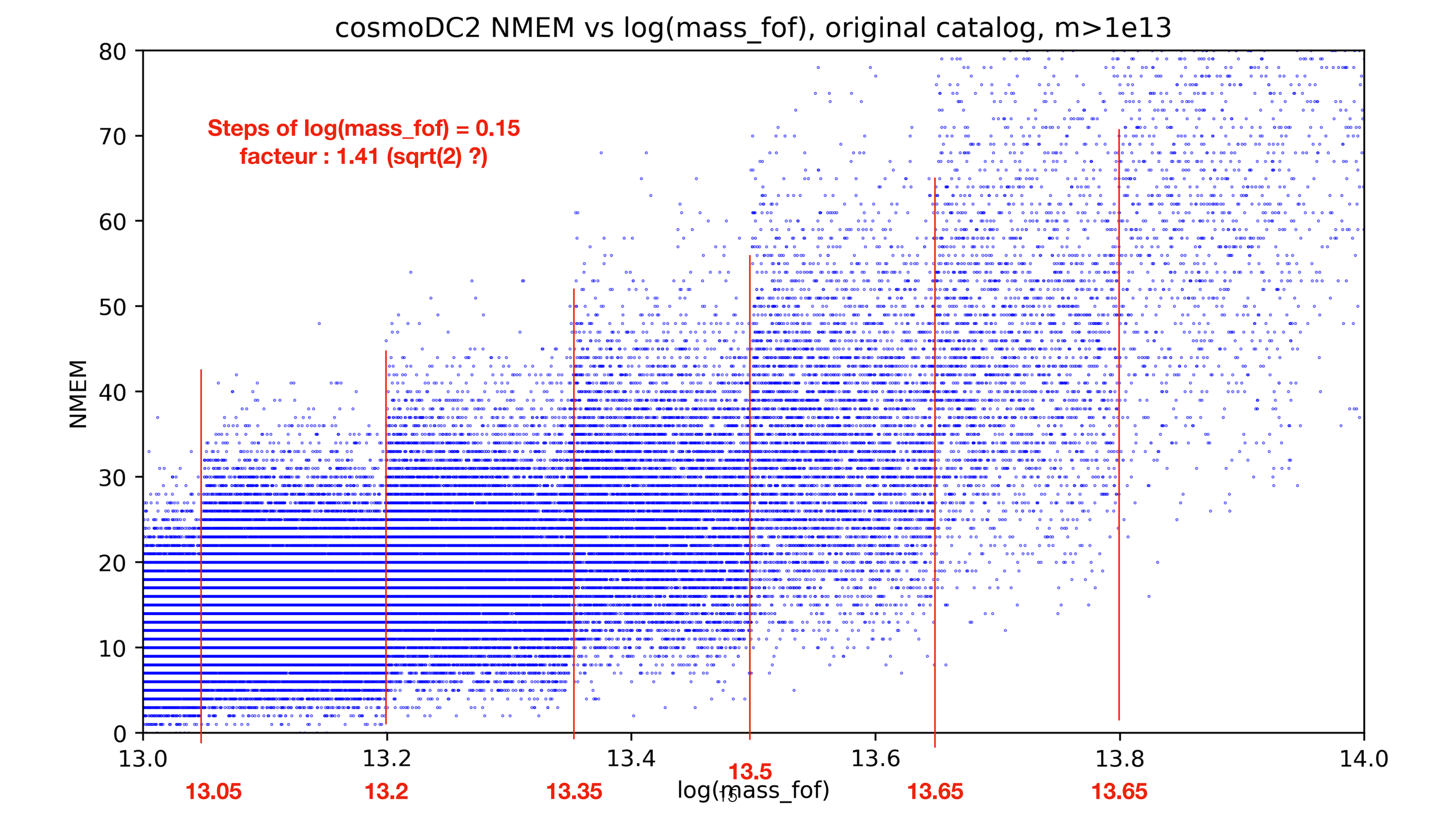
Comparison of NMEM vs mass_fof





21

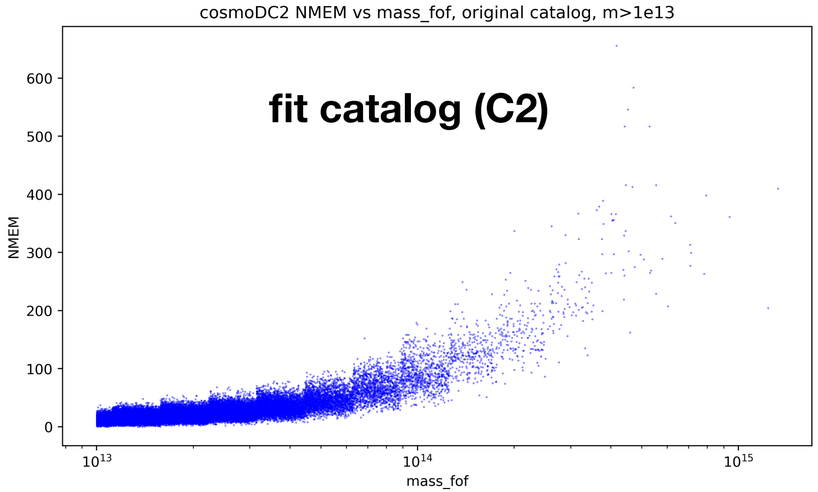
- Same patterns are observed

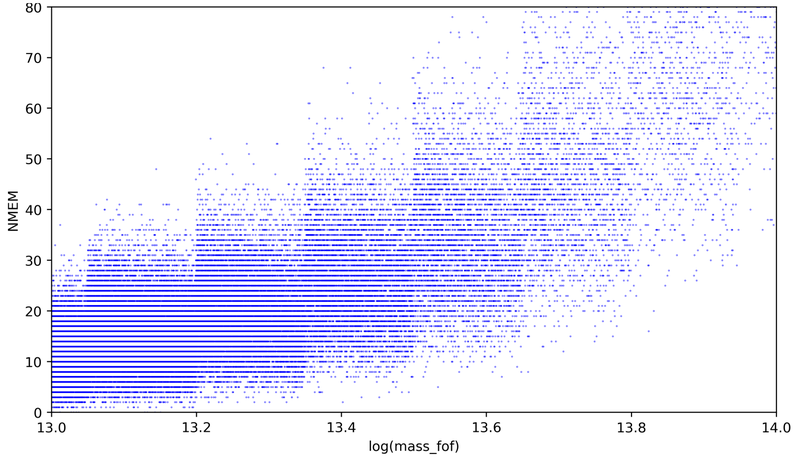

cosmoDC2 NMEM vs log(mass_fof), original catalog, m>1e13
Steps of log(mass_fof) = 0.15
facteur : 1.41 (sqrt(2) ?)
NMEM
log(mass_fof)
0
10
20
30
40
50
60
70
80
13.0
13.2
13.4
13.6
13.8
14.0
13.05
13.2
13.35
13.5
13.65
13.65

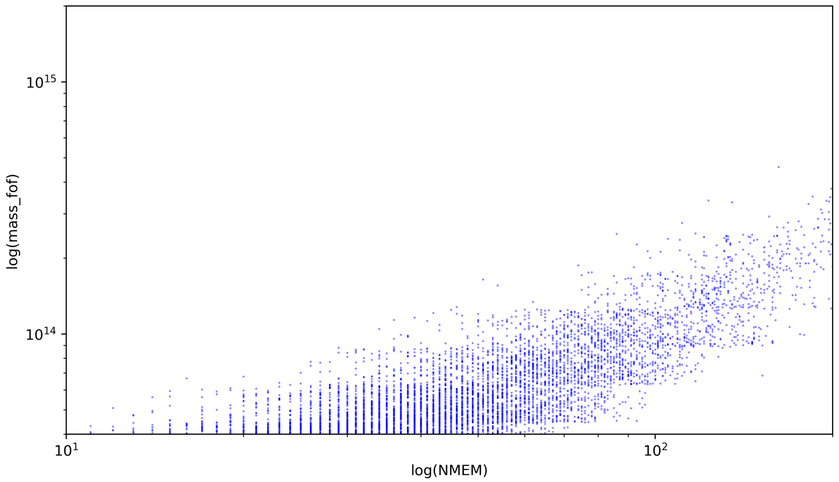
Comparison of C2 and C3
(to understand if discontinuities come from GCR -> fit conversion)

Check with GCR catalog

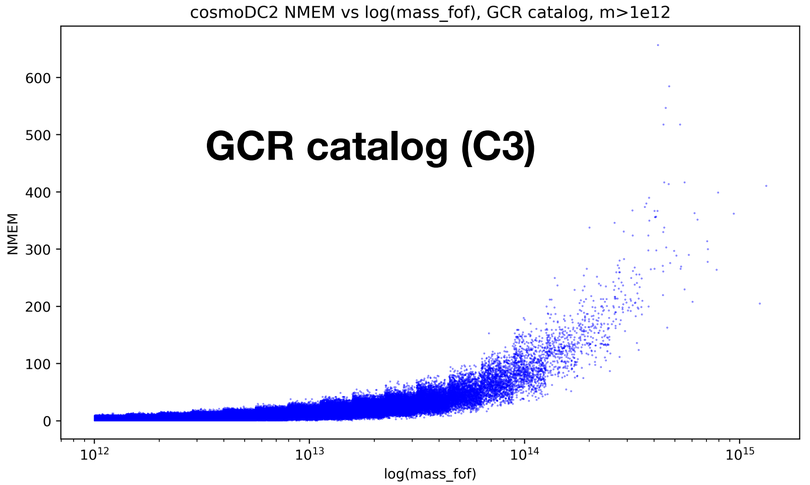
- After discussion with Eve Kovacs, she suspects that the bug was introduced during the process of conversion GCR->fit catalog
- She suggested that i look instead directly at GCR catalogs
- So i used the one presented at the beginning of these slides

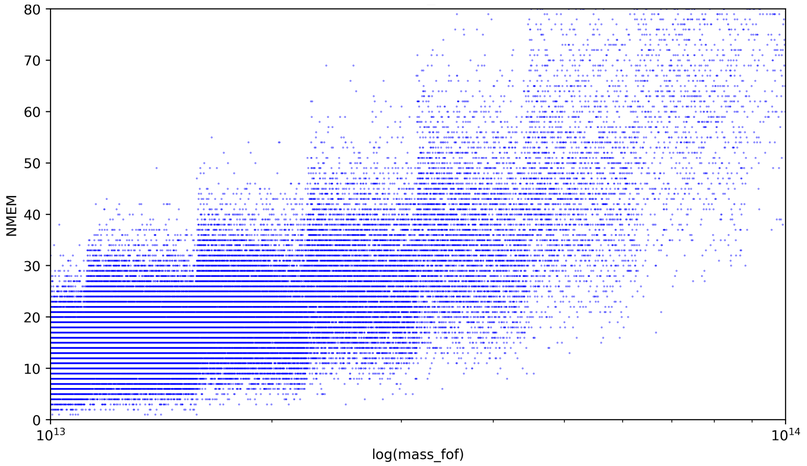


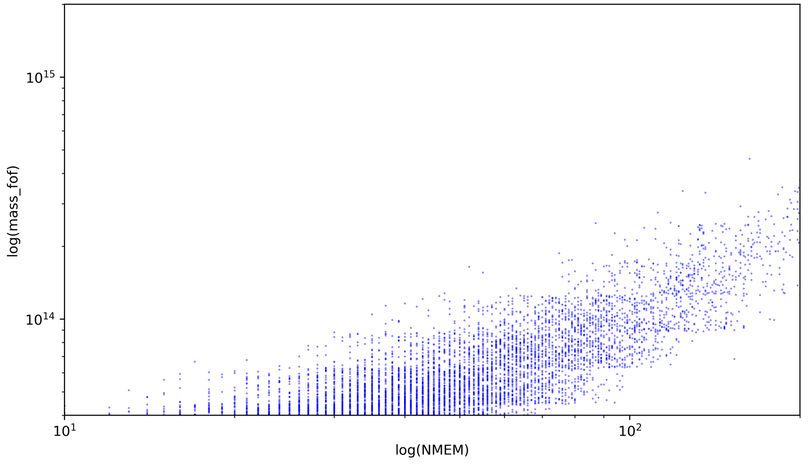




- Exactly same results
- So steps not introduced in fit catalog, already present in GCR one







Comparison of C3 and C4
(to understand if discontinuities come from cosmoDC2)

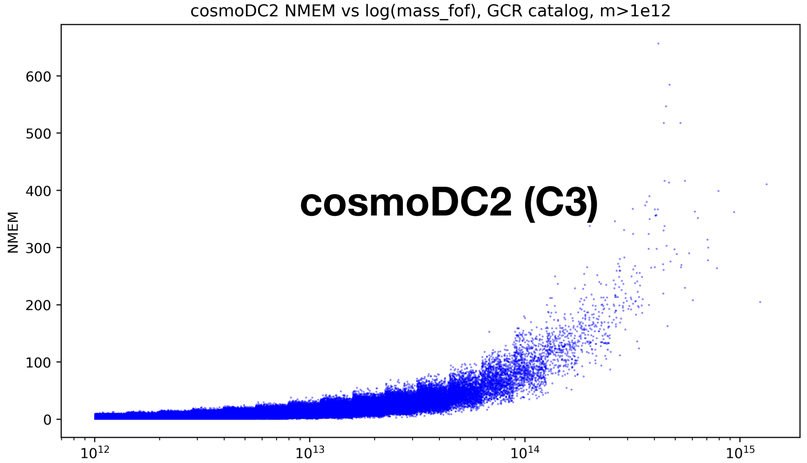
Check with skysim5000 catalog

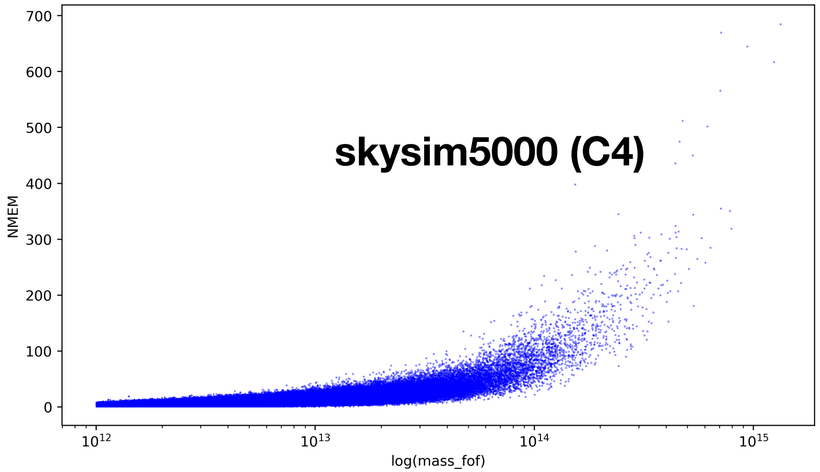
- Rediscussing with Eve Kovacs, she had an idea how this could arrive

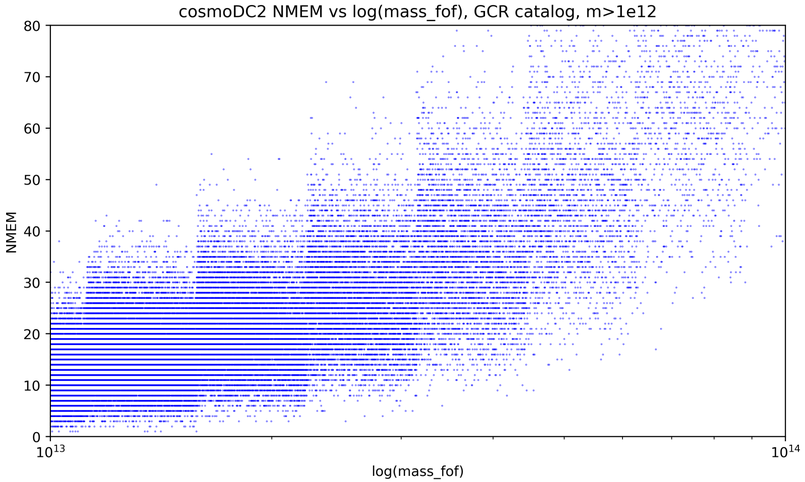
  - « In cosmoDC2 and skysim5000, we resampled galaxies from Universe Machine (UM) by matching halo masses across the simulations. For cosmoDC2, we did the matching by binning in the halo masses and choosing a matching halo randomly from the same mass bin in UM. Due to the steeply falling mass function, this resulted in a “pile-up” at low halo masses in the selected galaxies and introduced some discontinuities in galaxy properties as a function of halo mass. I think this is exactly what you are seeing. For skysim5000, we switched to using kdeTree matching, so this effect went away »

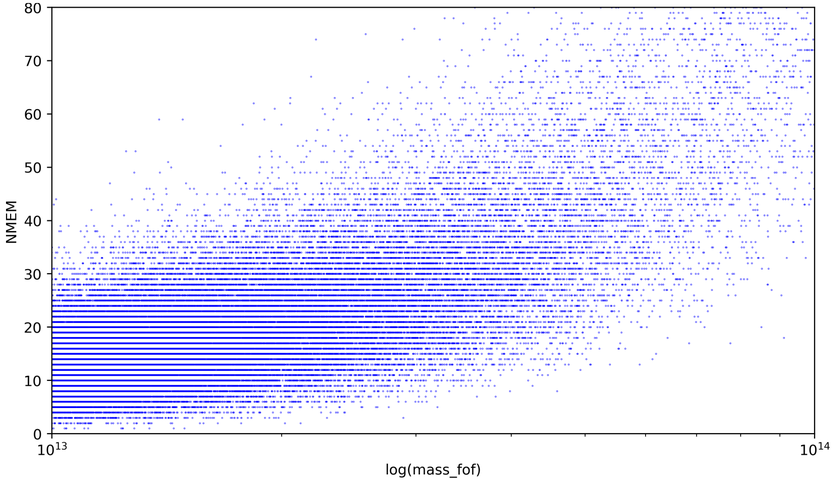
- When i asked the potential effect on cosmological parameters of using cosmoDC2 :

  - « I don’t know if the discontinuities will affect the extraction of cosmological parameters. The mass-richness relation has so much scatter anyway. You would have to ask someone in the CL WG about that.»

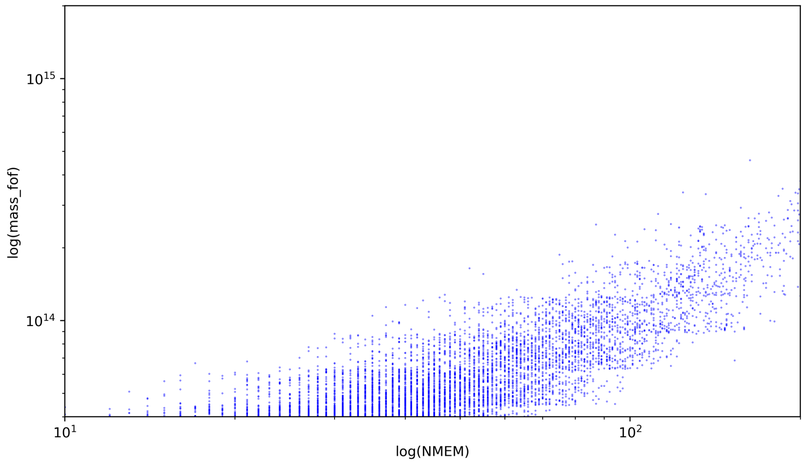


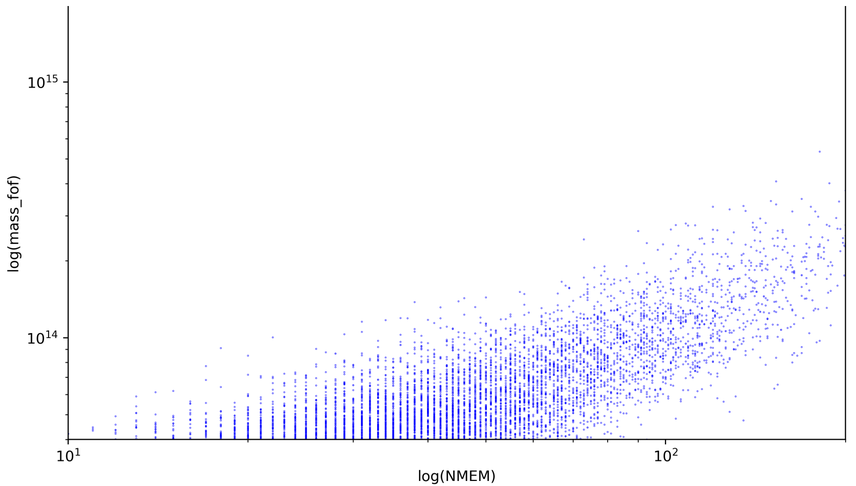




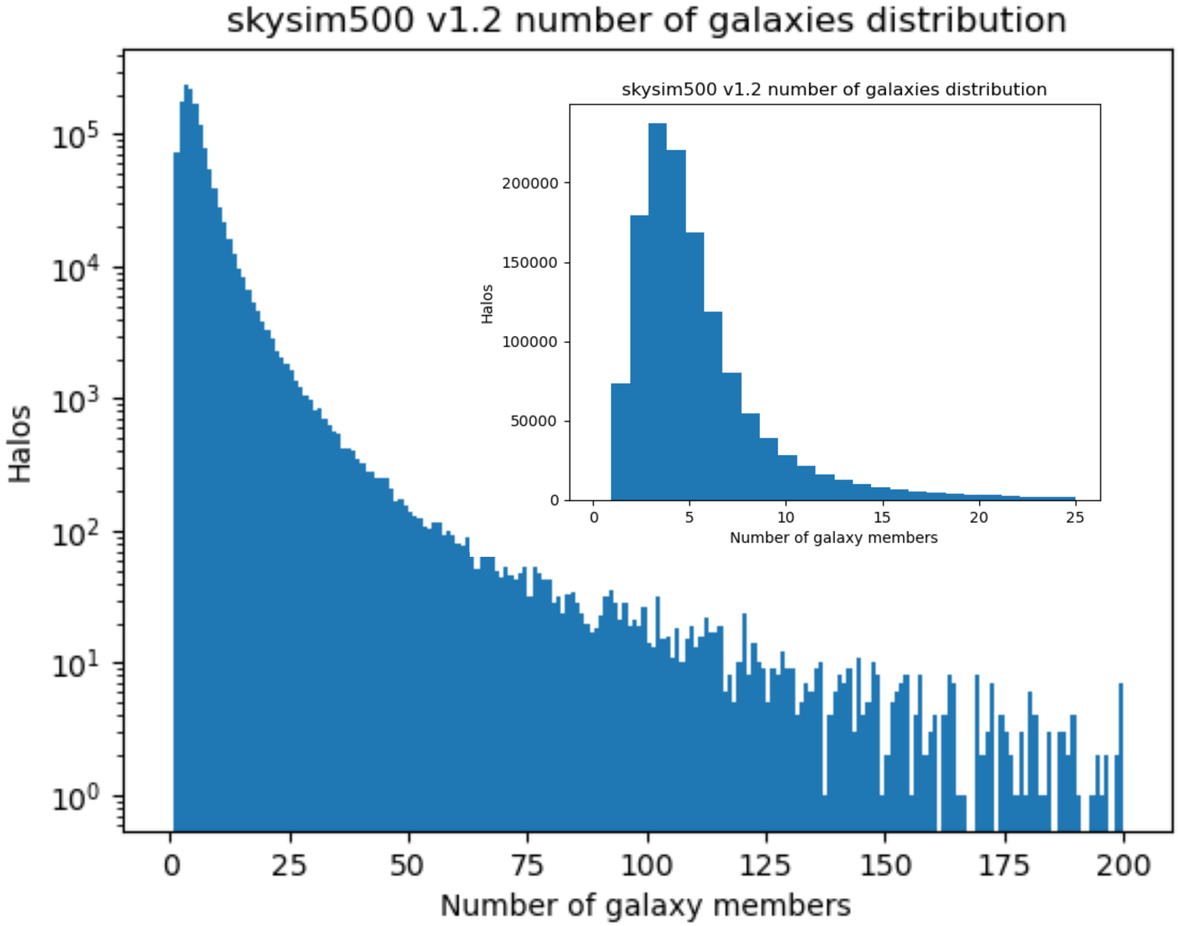


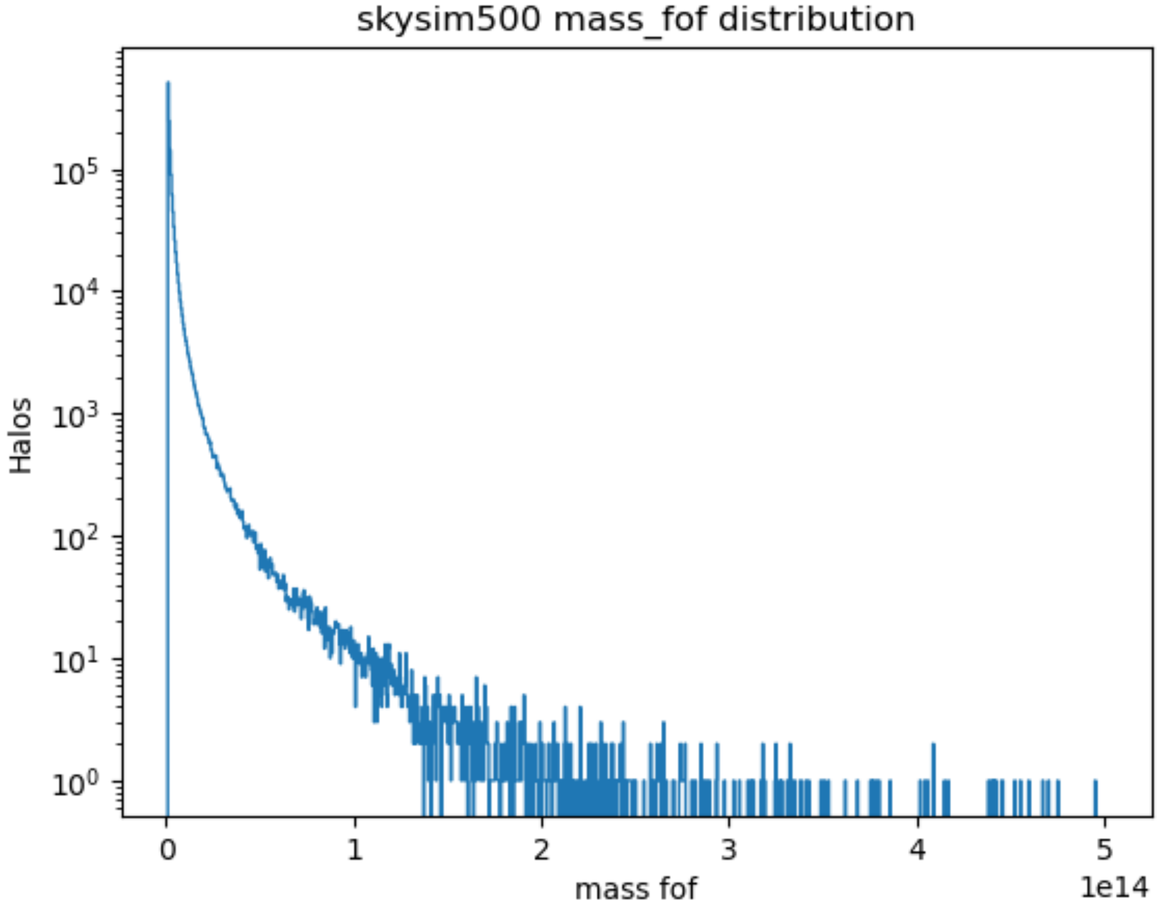
Now, the discontinuities indeed disappeared !



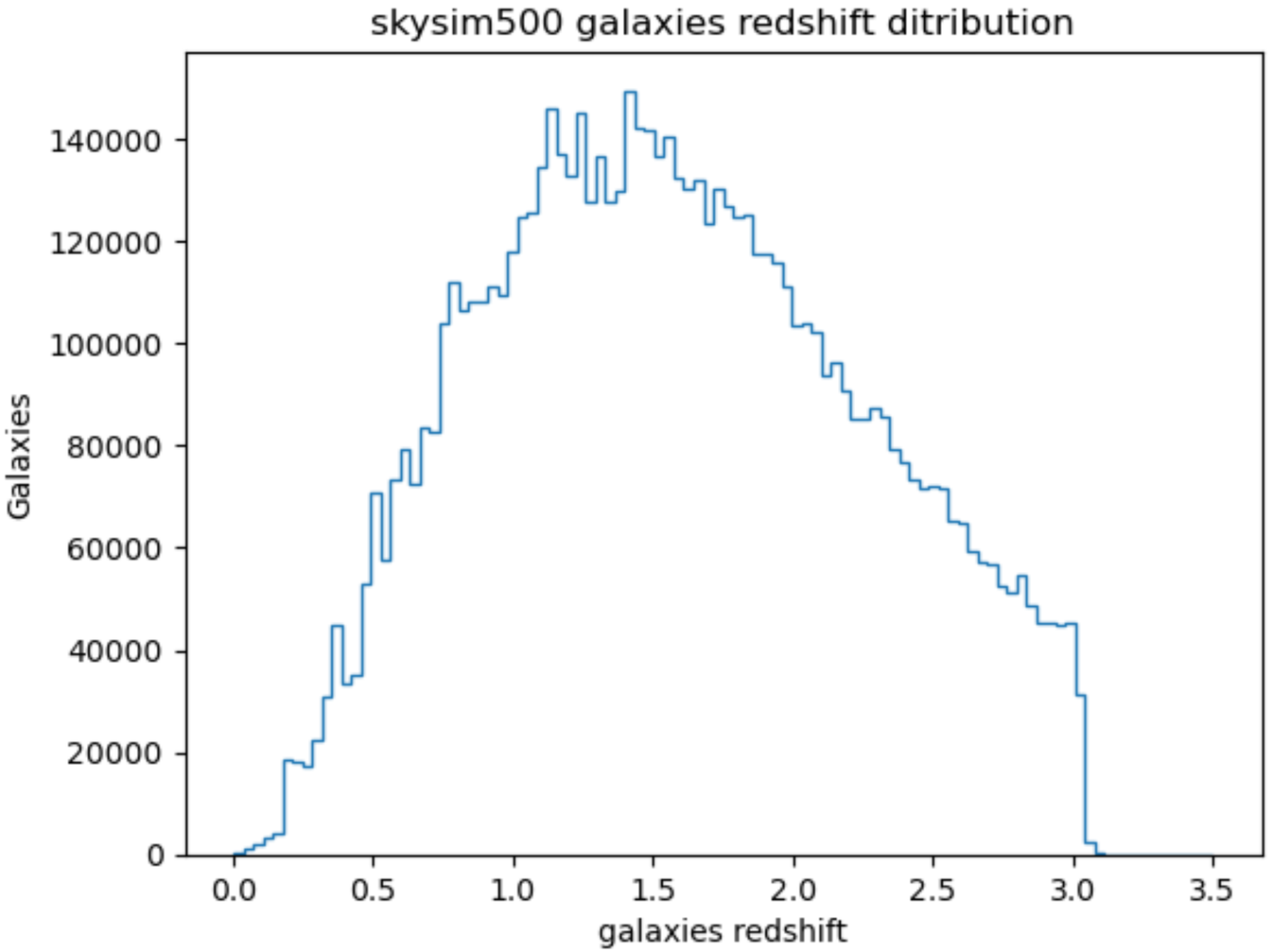


distributions for M>1e12

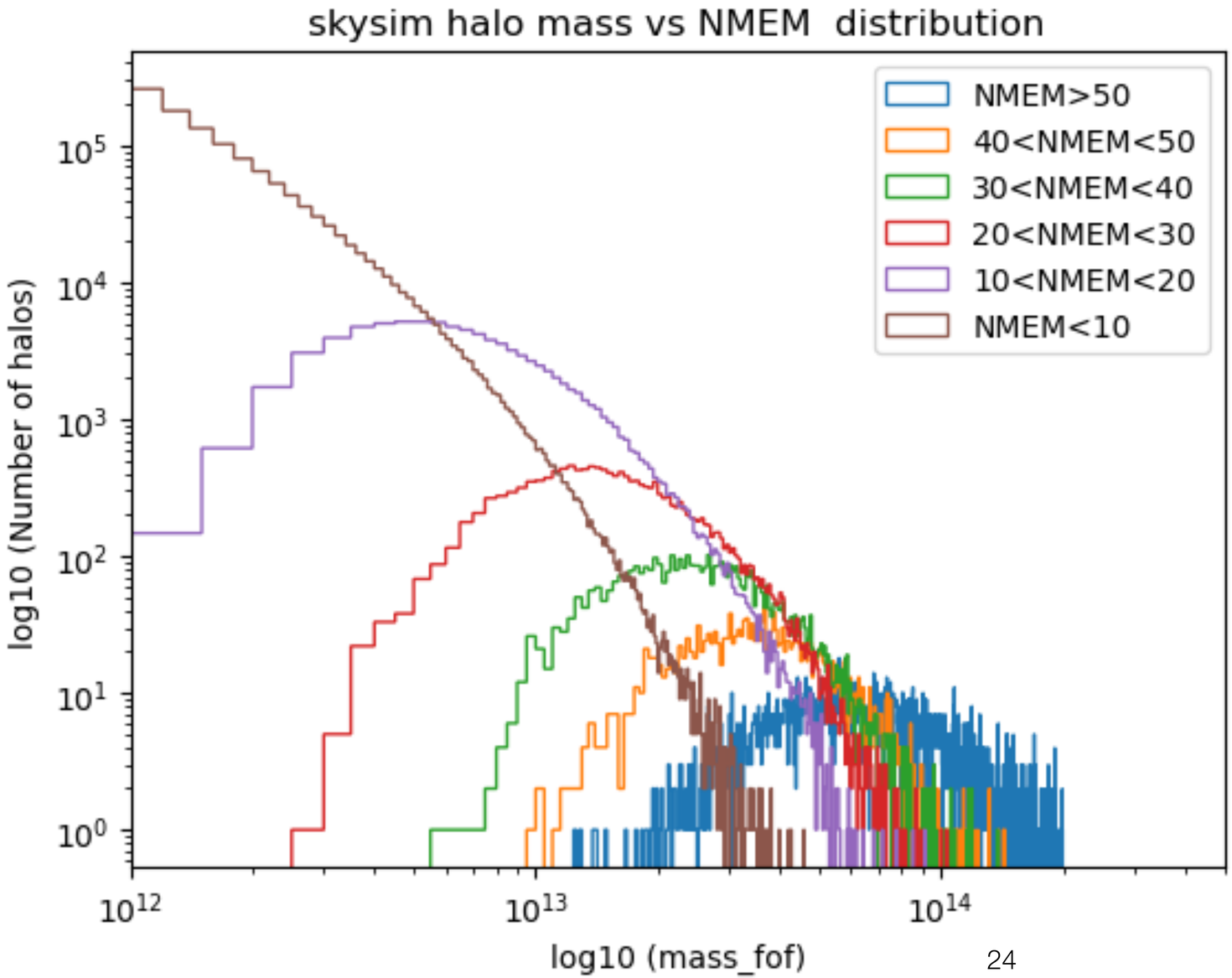




Galaxies redshift



Halo mass in bins of number of galaxies members

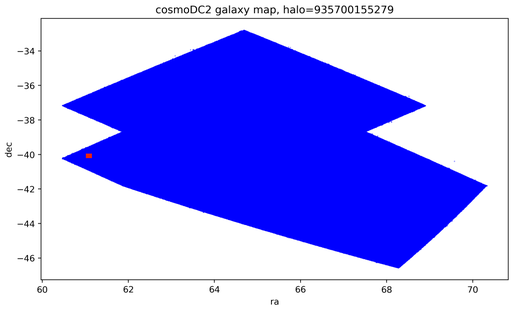


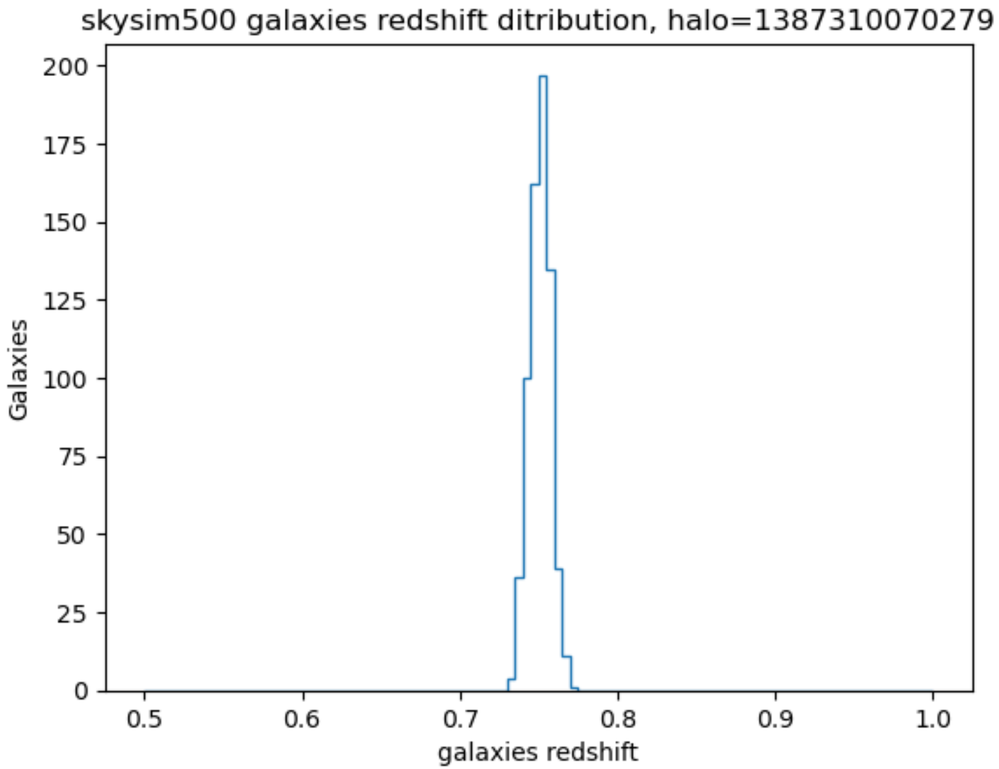
Large halos properties: m>5e14

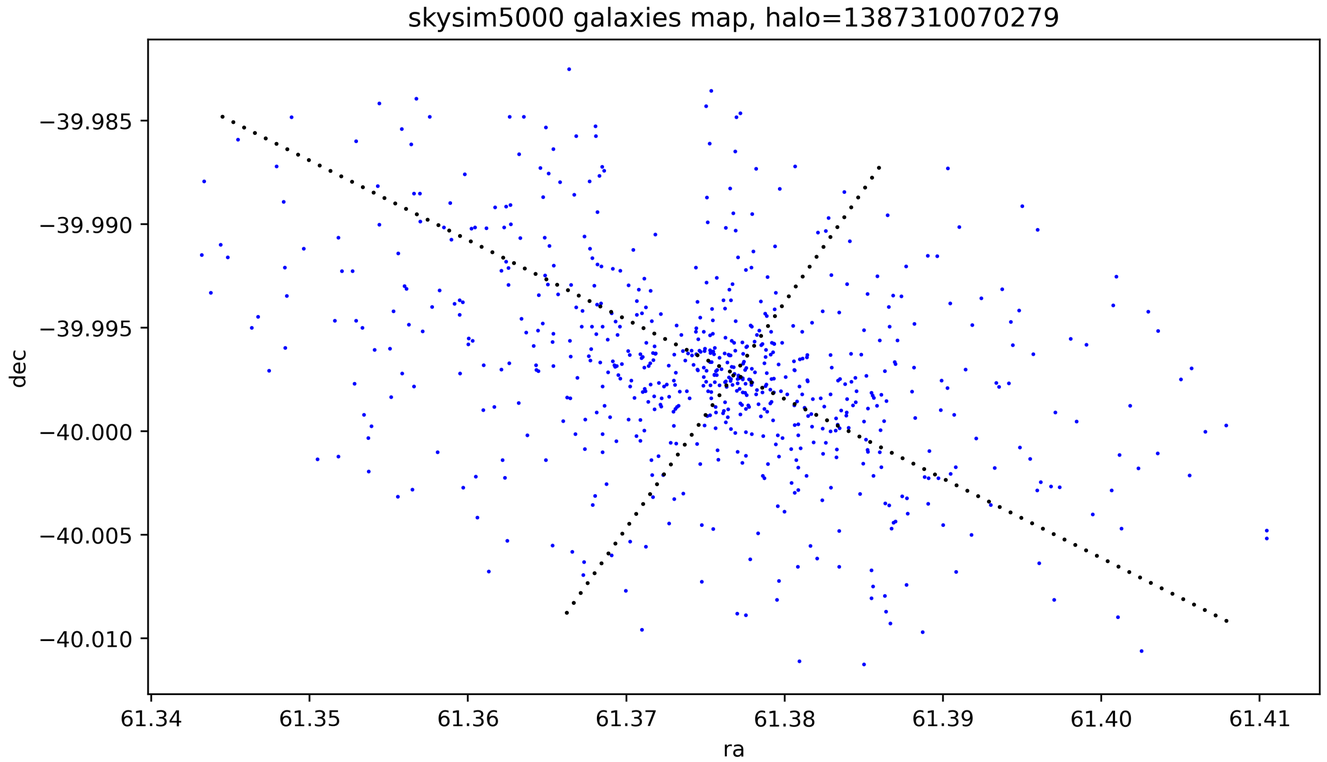
- A few halos with m>5e14

halo_id	NMEM	halo_mass
1387310070279	685	1.328346e+15
21709942331	670	7.123371e+14
364910326272	645	9.387429e+14
78609942421	617	1.239144e+15
482210072315	566	7.064627e+14
6410710199213	502	6.165222e+14
562109816259	450	5.283154e+14
111809816411	355	7.086487e+14
308410072401	351	7.817641e+14
149509814401	344	5.305248e+14
117610450411	319	7.940985e+14
6310327373	302	5.795759e+14
73109559373	285	6.356945e+14
144309942421	265	5.544304e+14
183209559373	258	6.032164e+14
555409814355	247	5.070324e+14
231010072392	181	5.339013e+14

halo = 1387310070279

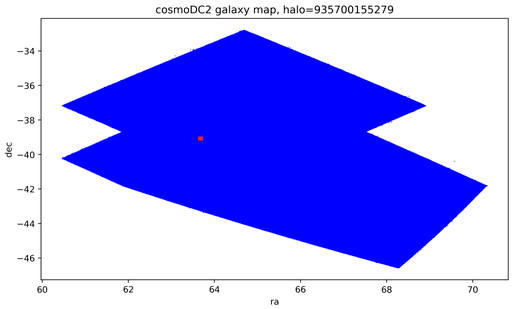


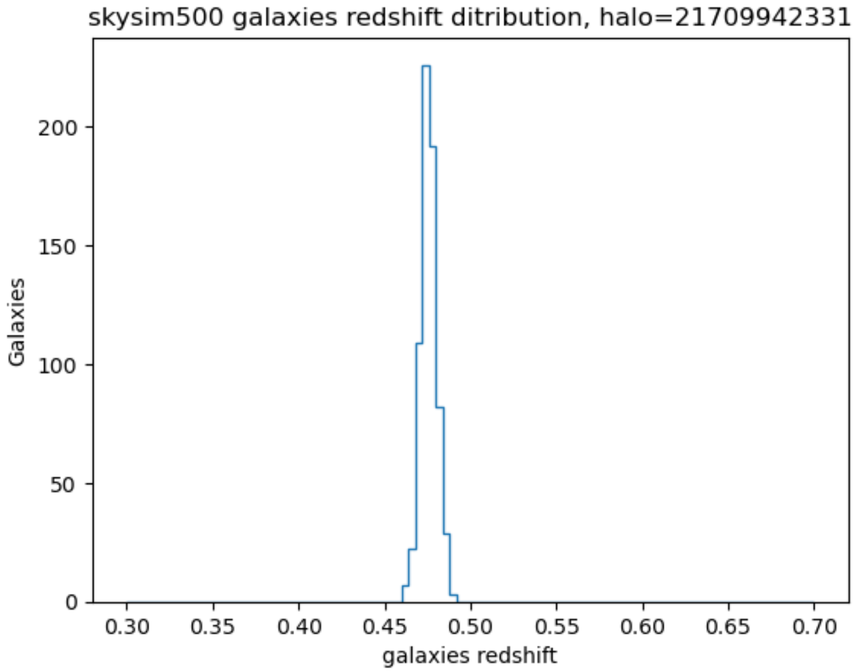


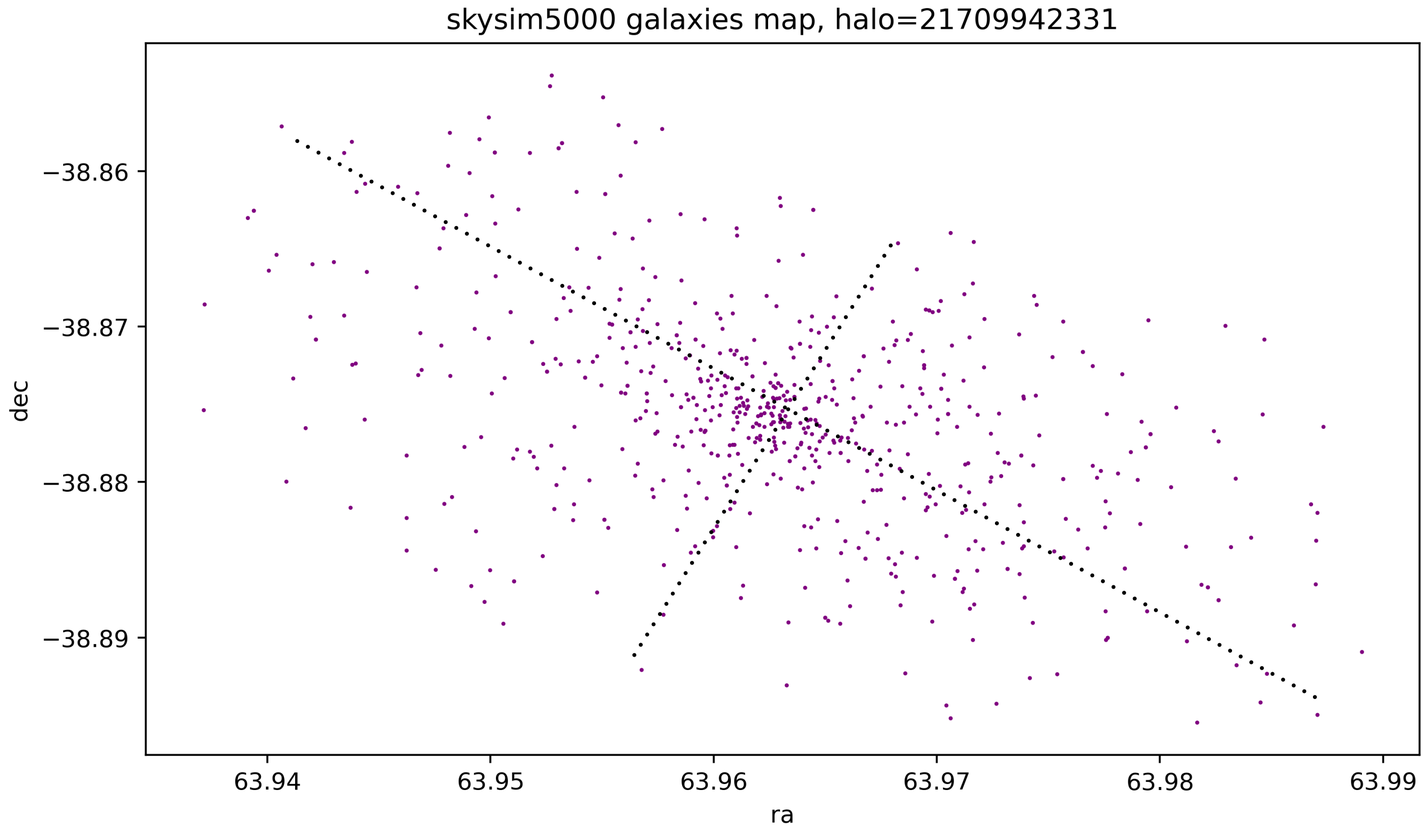


- m_fof = 1.3 e15 solar mass
- Far from being circular -> elliptical ?
- From validation paper: «The halo-mass definition used for the observed data is M_{200m}, which is approximately equivalent to M_{FoF}, the halo-mass definition used in the cosmoDC2 simulation »

halo=21709942331

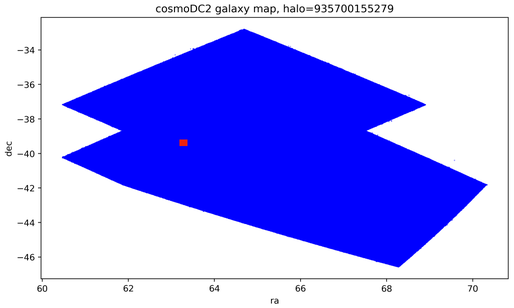


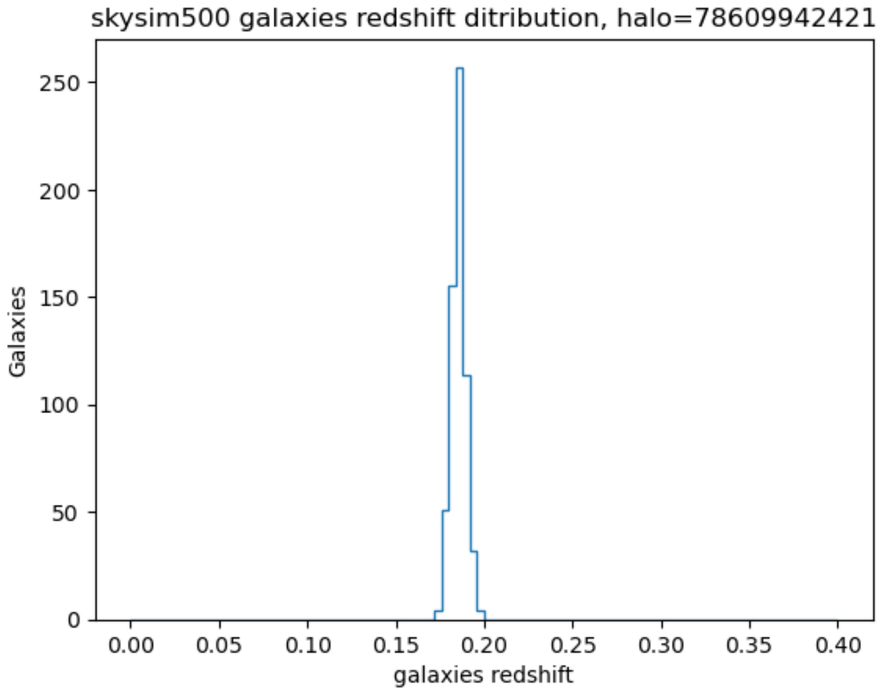


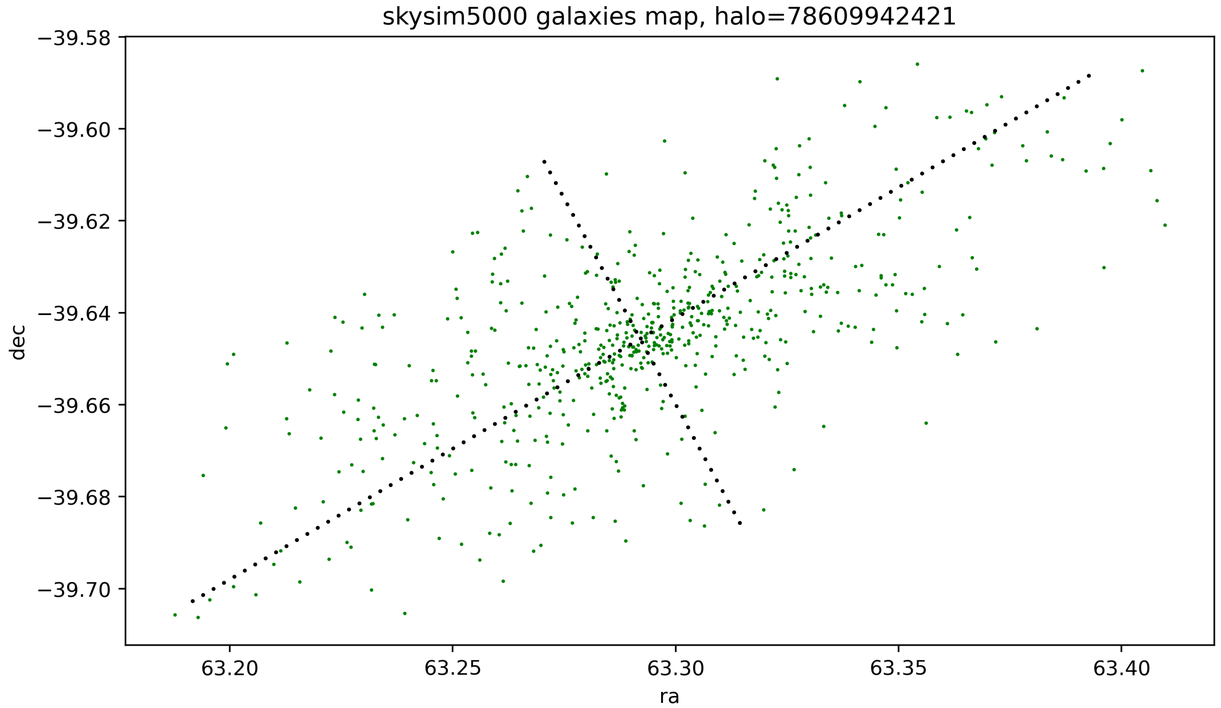


- m_fof = 7.1 e14 solar mass
- Far from being circular ! -> elliptical ?

halo=78609942421







- m_fof = 1.2 e15 solar mass
- Far from being circular ! -> elliptical ?

Conclusions/Next steps

- For cosmological constraints, seems more robust to use skysim5000
- Comparison of mass_fof and m200c (called baseDC2/sod_halo_mass in skysim5000 simulation) especially for large clusters
  - which variable is best to use to do cosmological constraints ?
- Since in skysim5000, only redshift of galaxy is provided, find a proxy for cluster redshift (average of galaxie redshifts ?)
  - to study mass/redshift relationship
- Use shape of large clusters (elliptical, axis direction, extension) rather than just mass ?
- Other ideas ?