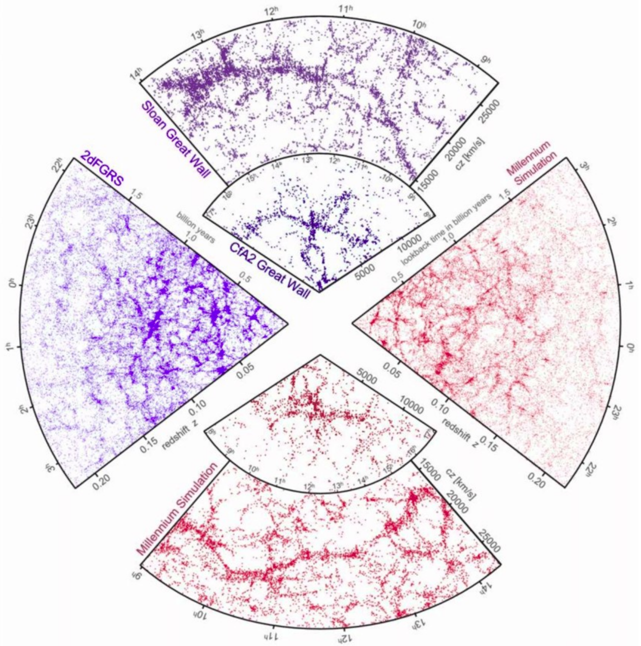
Expected measurements of the two-point correlation function in the first phase of DESI

D. Sierra-Porta and Jaime Forero-Romero
Universidad de los Andes
d.sierrap@uniandes.edu.co

The two-point correlation function.

Comparison between Sky Survey programs (Sloan Digital Sky Survey SDSS, Twodegree Field Galaxy Redshift Survey 2dFGRS) and Numerical Simulations (Millennium) from (Springel et al.,2006) figure 1. (© Nature Publishing Group)



- Most structures in the universe are thought to have in fact been possible due to fluctuations in the field of density of matter. A common measure for this is expressed as

$$\delta(x)=\frac{\rho(x)-\bar{\rho}}{\bar{\rho}}$$

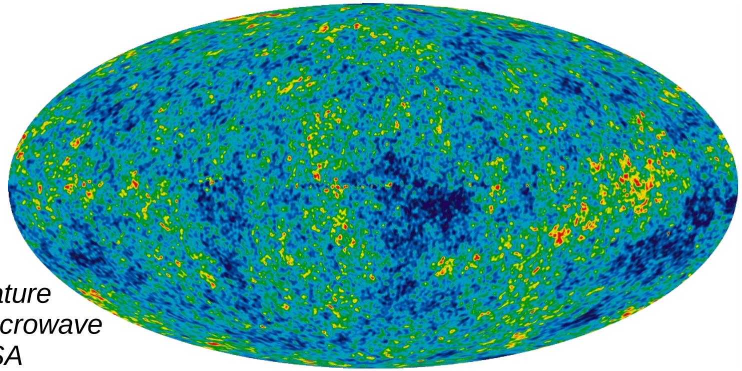
- where *δ(x)* measures these fluctuations at a point *x* in the universe and *ρ(x)* is the mass density at that point.

- It is convenient to consider the construction of a general field of densities by superposition of many ways.

$$\delta(x)=\frac{1}{(2\pi)^3}\int d^3k\, e^{ikx}\,\delta(k)$$

$$\delta(k)=\int d^3k\, e^{-ikx}\,\delta(x)$$

$$k=\frac{2\pi}{\lambda}$$



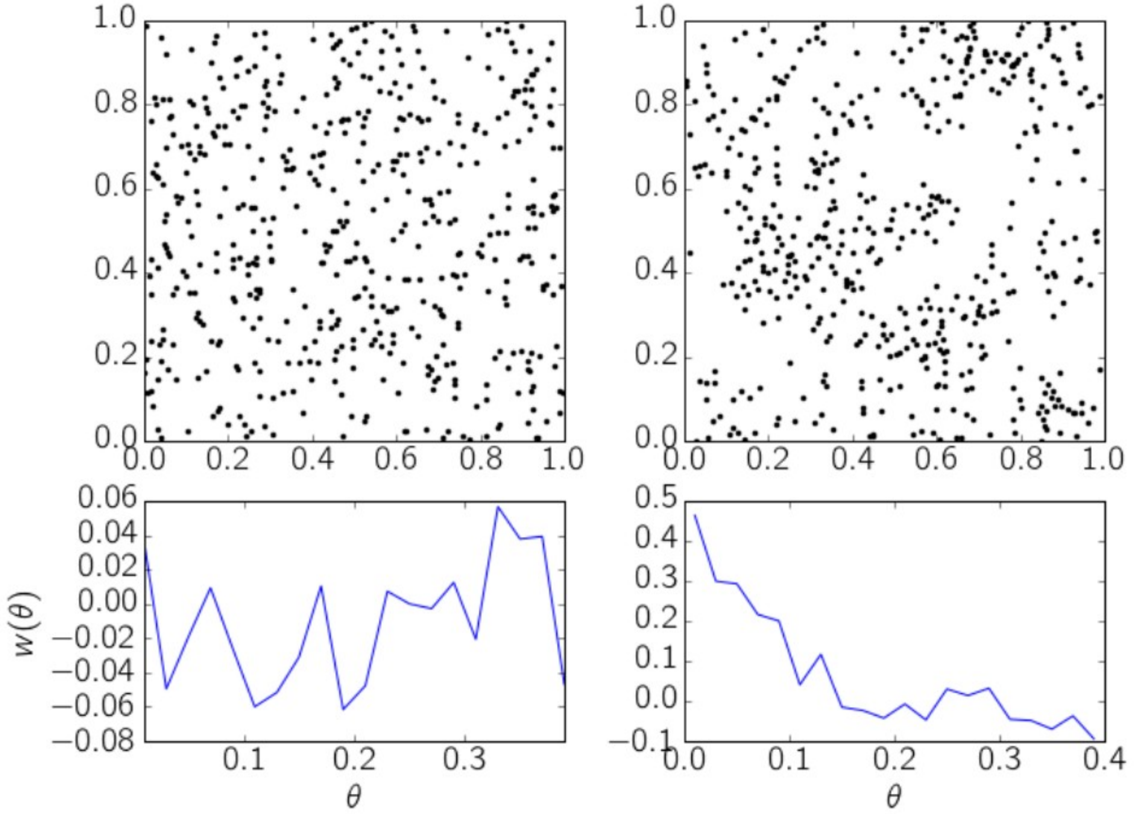
***The full-sky image** of temperature fluctuations in the cosmic microwave background. Credits: NASA*

- Simple way of describing the clustering of data points. This is an important characteristic of a dataset, which not not captured by statistics such as the mean, median standard deviation, etc.

$$\delta P = \rho^2 \left[1 + w(\theta) \right] \delta\Omega_1 \, \delta\Omega_2$$

- For example, the data points in Figure are drawn from distributions with equal densities, but different clustering properties. Using the two-point correlation function we can characterize the difference between the distributions.

- The two-point angular correlation function is convenient to measure, easy to interpret,and naturally accommodates complicated survey areas (typical in astronomical surveys).



“Is the probability, in excess of the random one, of finding a galaxy at a fixed distance from a random neighbor. Its Fourier transform is the power spectrum. "

Different estimators:

With the above preparation the pairwise estimators used in what follows are the natural estimator ξ_N, and the estimators due to Davis and Peebles (1983) ξ_{DP}, Hewett (1982) ξ_{He}, Hamilton (1993) ξ_{Ha}, and Landy and Szalay (hereafter LS, 1993) ξ_{LS}:

$$\xi_N = \frac{DD}{RR} - 1 \qquad \xi_{DP} = \frac{DD}{DR} - 1 \qquad \xi_{He} = \frac{DD - DR}{RR} \qquad \xi_{Ha} = \frac{DD \cdot DR}{(DR)^2} \qquad \xi_{LS} = \frac{DD - 2\,DR + RR}{RR}$$

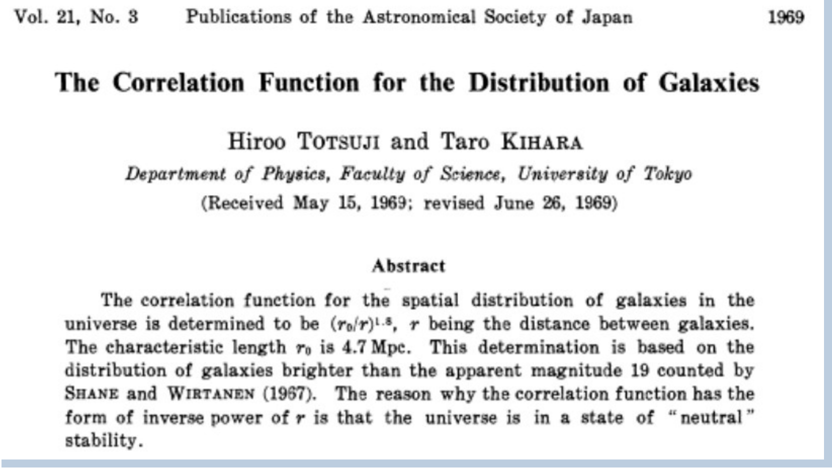
The Galaxy Correlation Function

- First measured by Totsuji and Kihara, then Peebles et al
- Mostly angular correlations in the beginning
- Later more and more redshift space
- Power law is a good approximation

$$\xi(r)=\left(\frac{r}{r_0}\right)^{-\gamma}$$

- Correlation length r_0=5.4 Mpc/h
- Exponent is around γ =1.8
- Corresponding angular correlations

$$w(r)=\left(\frac{\theta}{\theta_0}\right)^{1-\gamma}$$



Vol. 21, No. 3 Publications of the Astronomical Society of Japan 1969

The Correlation Function for the Distribution of Galaxies

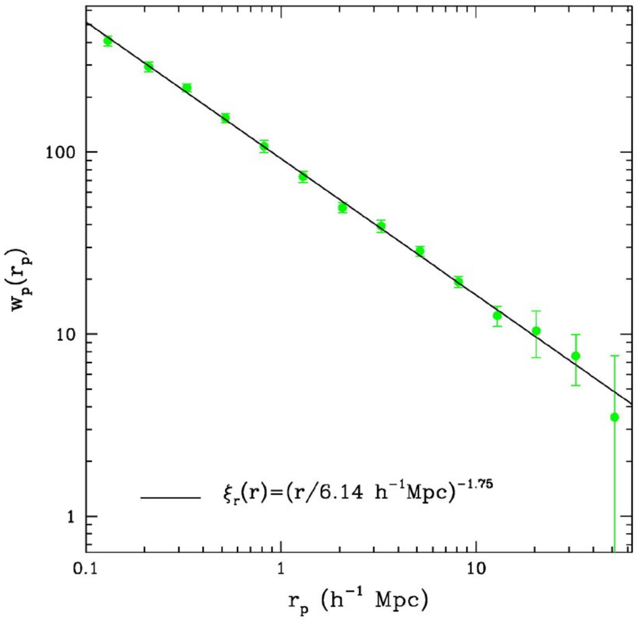
Hiroo TOTSUJI and Taro KIHARA

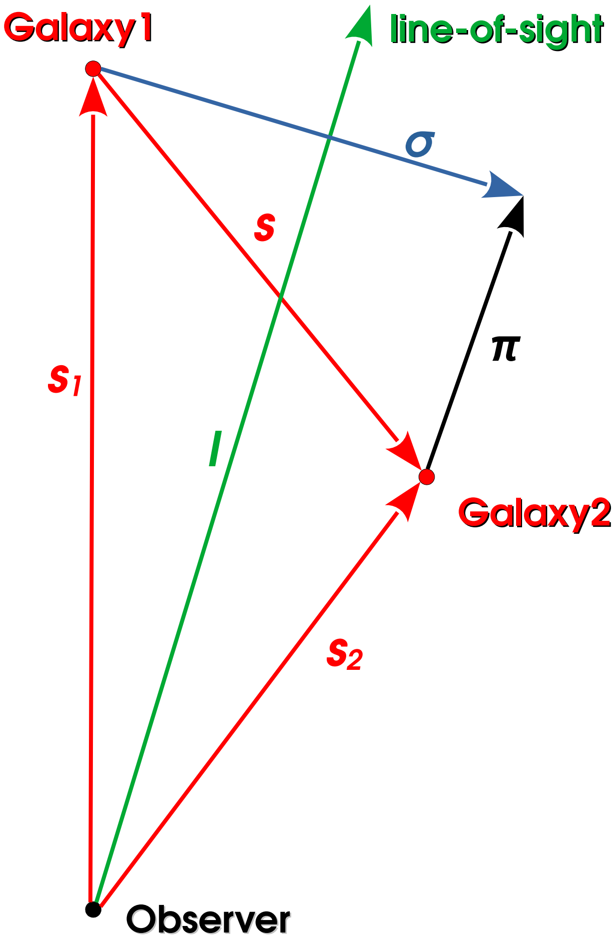
Department of Physics, Faculty of Science, University of Tokyo

(Received May 15, 1969; revised June 26, 1969)

Abstract

The correlation function for the spatial distribution of galaxies in the universe is determined to be $(r_0/r)^{1.8}$, r being the distance between galaxies. The characteristic length r_0 is 4.7 Mpc. This determination is based on the distribution of galaxies brighter than the apparent magnitude 19 counted by SHANE and WIRTANEN (1967). The reason why the correlation function has the form of inverse power of r is that the universe is in a state of "neutral" stability.





Some representations:

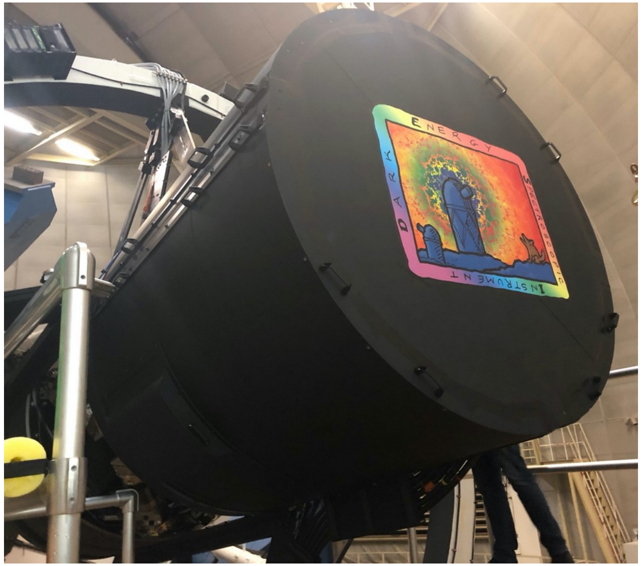
- Decomposition for the distances between pairs of galaxies in transverse coordinates (***σ***) and along the line of sight (***π***).

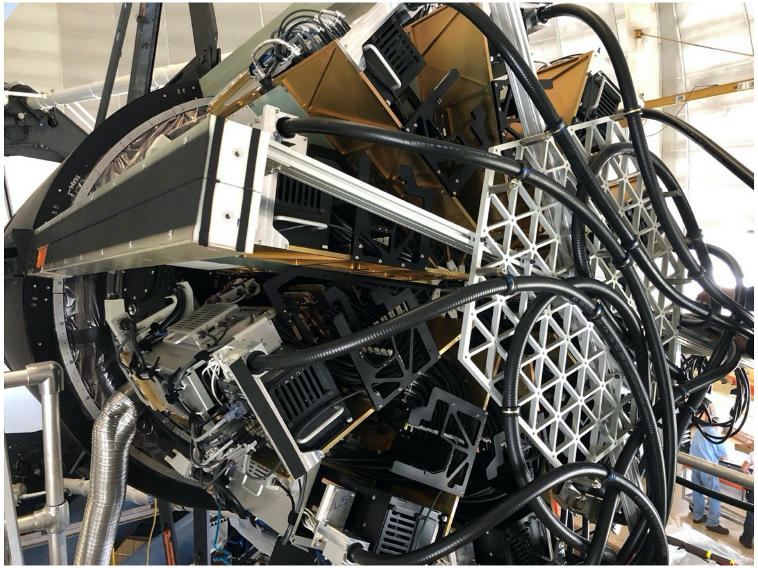
$$\pi \equiv \frac{|s \cdot l|}{|l|} \qquad \sigma \equiv \sqrt{s \cdot s - \pi^2}$$

Then the two-point function is expressed ***ξ(σ,π)*** → counting pairs in intervals of ***σ***, ***π*** and use one estimator: Landy-Szalay estimator

The instrument *installed on the 4-m Mayall Telescope at KittPeak, Arizona.*

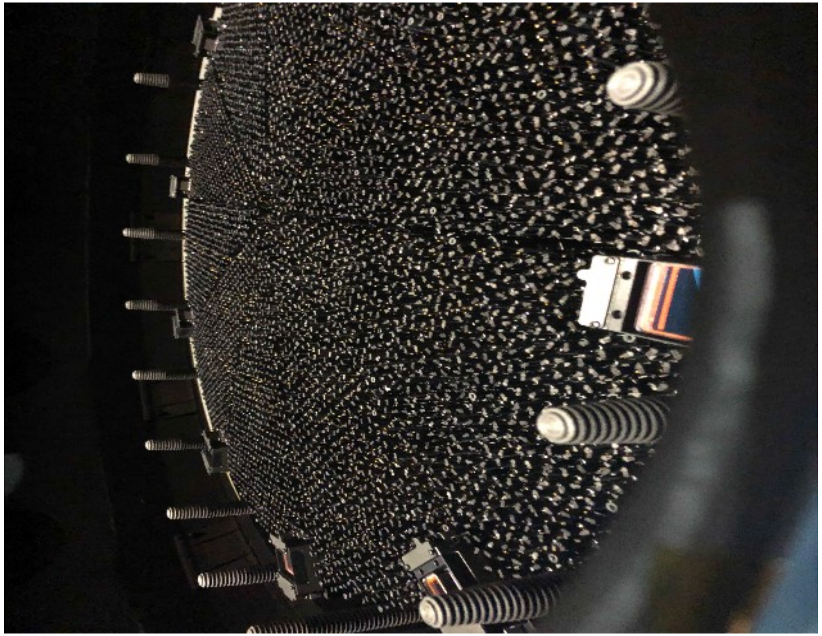
- DESI will consist of dark-time and bright-time programs.
  - The dark-time survey will measure spectra of 4 million luminous red galaxies (LRGs) ($0.4 < z < 1.0$), 17 million emission line galaxies (ELGs) ($0.6 < z < 1.6$), 1.7 million quasars ($z < 2.1$) and 0.7 million high redshift quasars ($2.1 < z < 3.5$) to probe the Ly-α forest.
  - The bright-time survey will consist of the bright galaxy survey (BGS), a low redshift, flux limited survey of $\sim$10 million galaxies with a median redshift $z_{med} \sim 0.2$, and a survey of Milky Way stars.





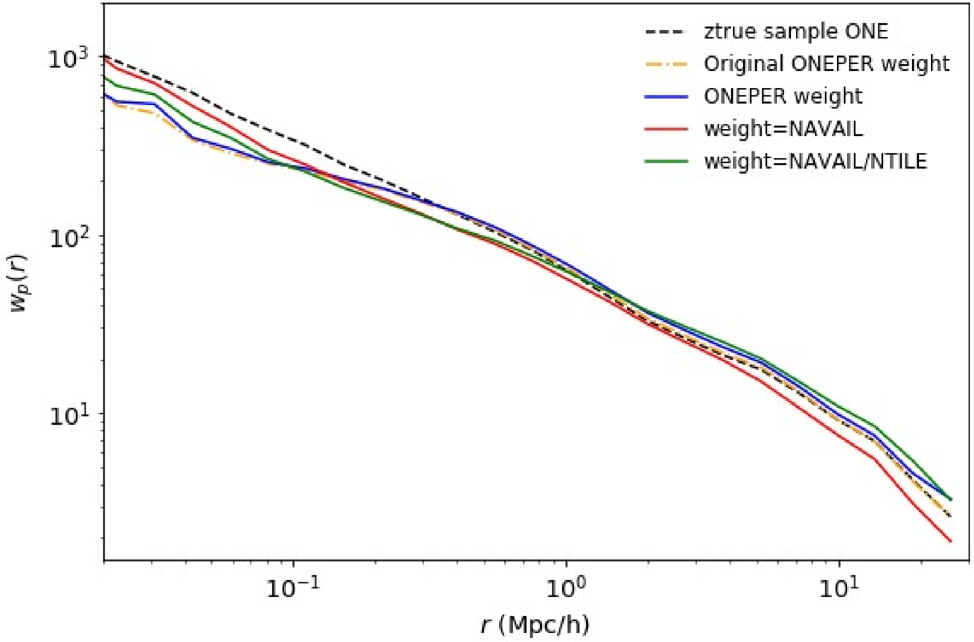
The Dark Energy Spectroscopic Instrument (DESI)

Installation of the focal plane instrument was completed in August, 2019. The picture below shows the fiber ends of the 5000 robotic positioners on the focal plane.

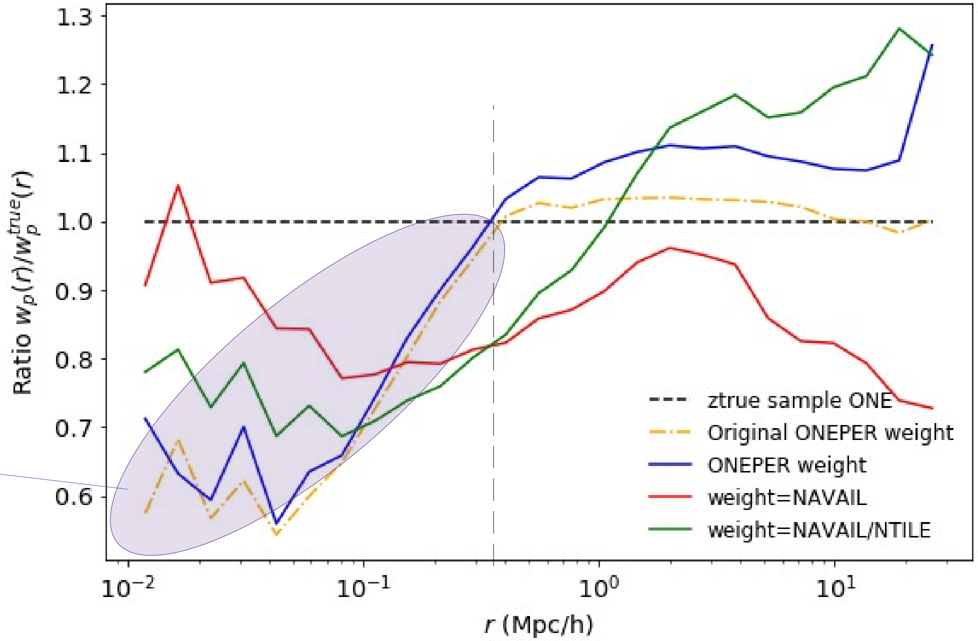


- The light from each target galaxy is collected by fibres located at the focal plane of the telescope, and taken to one of 10 spectrographs, where the spectrum is measured and a redshift determined.
- However, it is not possible to place a fibre on every single potential target, and even if it is, a redshift measurement can fail due to low surface brightness or weak spectral features.
- Other complications, such as observing conditions, also affect the redshift completeness in the final galaxy catalogue.
- To make precise measurements of galaxy clustering in order to reach the primary science aims of the survey, it is essential to correct for the effects of incompleteness.

A major systematic in galaxy clustering measurements is from the effect of fibre collisions, which occur because fibres cannot be placed arbitrarily close together. Since it is not possible to place a fibre on both objects in a close pair, that pair will be missing in the final catalogue, biasing the pair counts, particularly at small scales, which can bias galaxy clustering measurements...

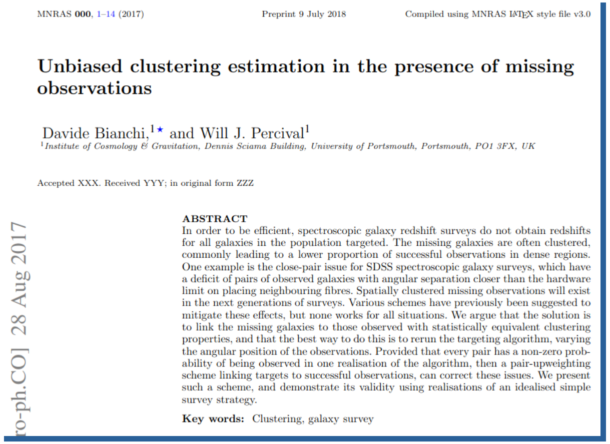


- The correlation function measured for different weights assigned to each galaxy
- The expected result is contrasted with the observed data



- *For r<0.35 Mpc/h there are errors in the measurement of the correlation function of up to 40% while for r>0.35 Mpc/h the results agree with what was expected very closely.*

The algorithm of (Bianchi & Percival 2017)


MNRAS 000, 1–14 (2017) Preprint 9 July 2018 Compiled using MNRAS LaTeX style file v3.0

Unbiased clustering estimation in the presence of missing observations

Davide Bianchi,[1★] and Will J. Percival[1]
[1]Institute of Cosmology & Gravitation, Dennis Sciama Building, University of Portsmouth, Portsmouth, PO1 3FX, UK

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT
In order to be efficient, spectroscopic galaxy redshift surveys do not obtain redshifts for all galaxies in the population targeted. The missing galaxies are often clustered, commonly leading to a lower proportion of successful observations in dense regions. One example is the close-pair issue for SDSS spectroscopic galaxy surveys, which have a deficit of pairs of observed galaxies with angular separation closer than the hardware limit on placing neighbouring fibres. Spatially clustered missing observations will exist in the next generations of surveys. Various schemes have previously been suggested to mitigate these effects, but none works for all situations. We argue that the solution is to link the missing galaxies to those observed with statistically equivalent clustering properties, and that the best way to do this is to rerun the targeting algorithm, varying the angular position of the observations. Provided that every pair has a non-zero probability of being observed in one realisation of the algorithm, then a pair-upweighting scheme linking targets to successful observations, can correct these issues. We present such a scheme, and demonstrate its validity using realisations of an idealised simple survey strategy.

Key words: Clustering, galaxy survey

ro-ph.CO] 28 Aug 2017

- Each pair of galaxies is weighted by but $w_{ij}=1/p_{ij}$, where p_{ij} is the probability that the pair is the observed one.
- Fiberassign code is runing N=128 times.
- For a galaxy i, a vector w_i of length 128 is stored, containing a 1
- if a fiber is assigned to the galaxy and a 0 otherwise.

$$w_{ij}=\frac{N_{real}}{w_i \cdot w_j}=\frac{N_{real}}{popcount(\vec{w}_i \wedge \vec{w}_j)}$$

Pair Inverse Probability (PIP)

- The corrected DD counts are calculated by adding the weights of the galaxy pairs at intervals of separation s

$$DD_w(\vec{s})=\sum_{s_i-s_j\approx s_{ij}} w_{ij}\frac{DD^{(p)}(\theta_{ij})}{DD_w(\theta_{ij})}$$

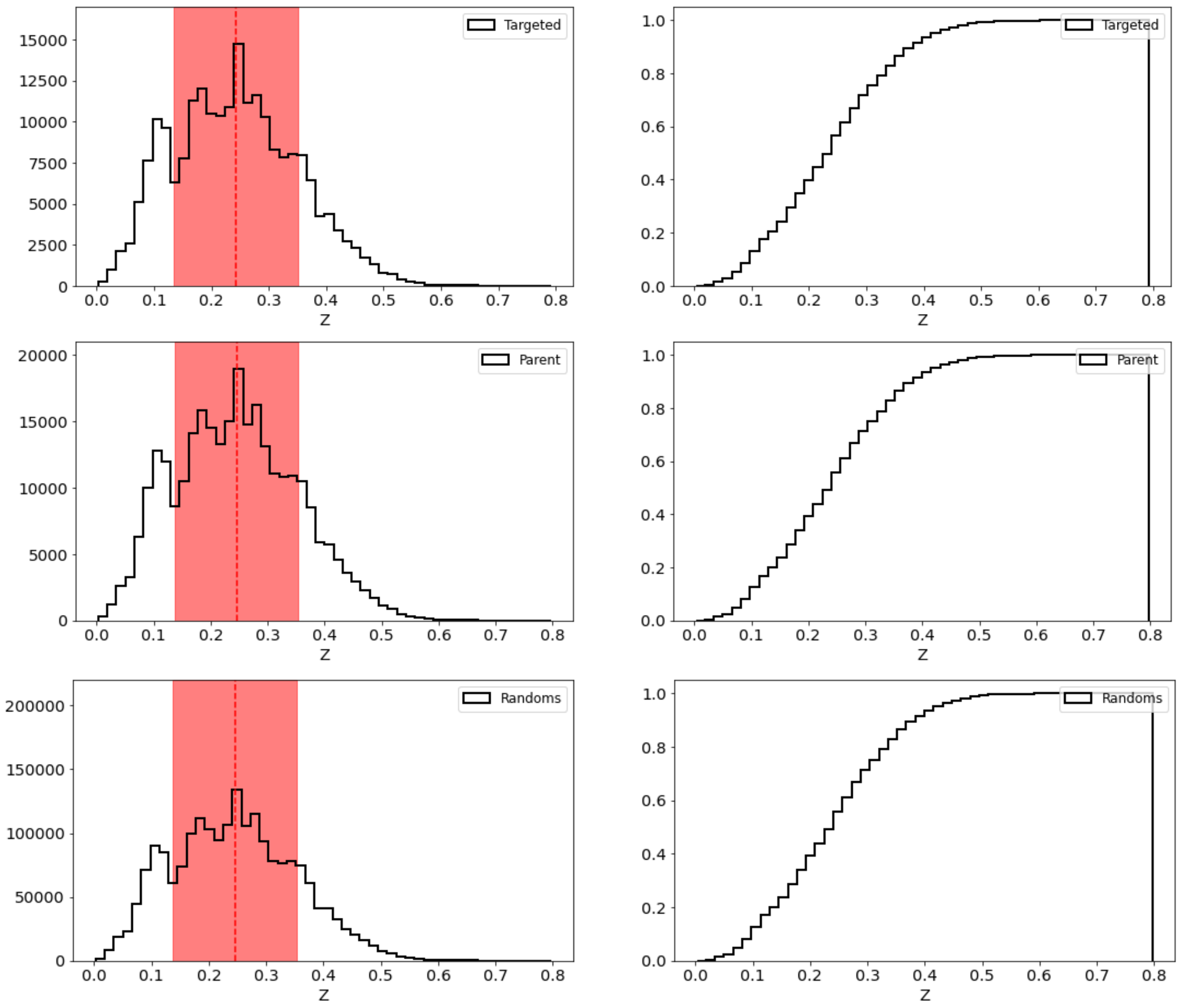
- Each galaxy is assigned an individual weight, which is the inverse of the probability that the galaxy is the target, $w_i=1/p_i$. This can be estimated from the same bitwise vectors used to estimate the probabilities of the pairs,

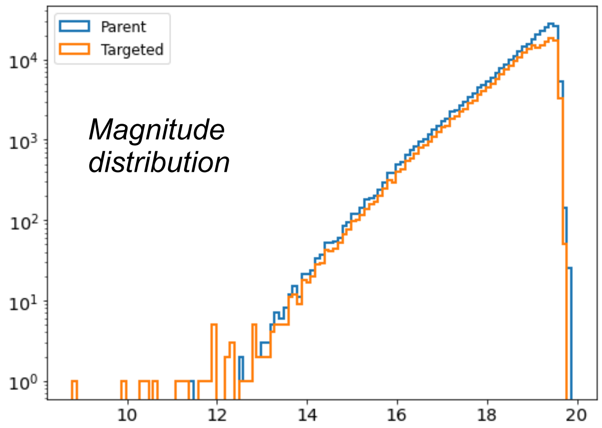
$$w_i=\frac{N_{real}}{popcount(\vec{w}_i)}$$

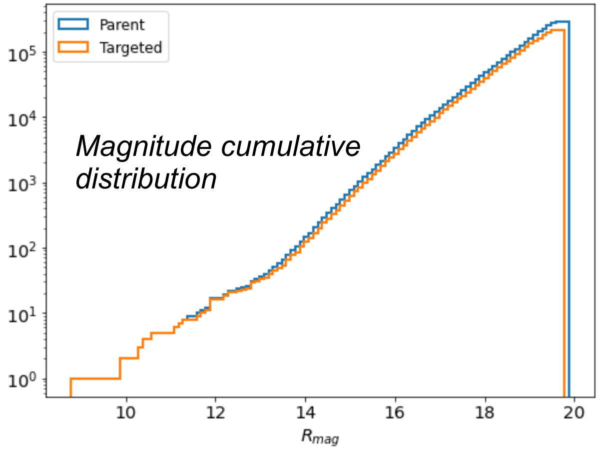
Individual Inverse Probability (IIP)

Redshift distribution

Redshift cumulative distribution





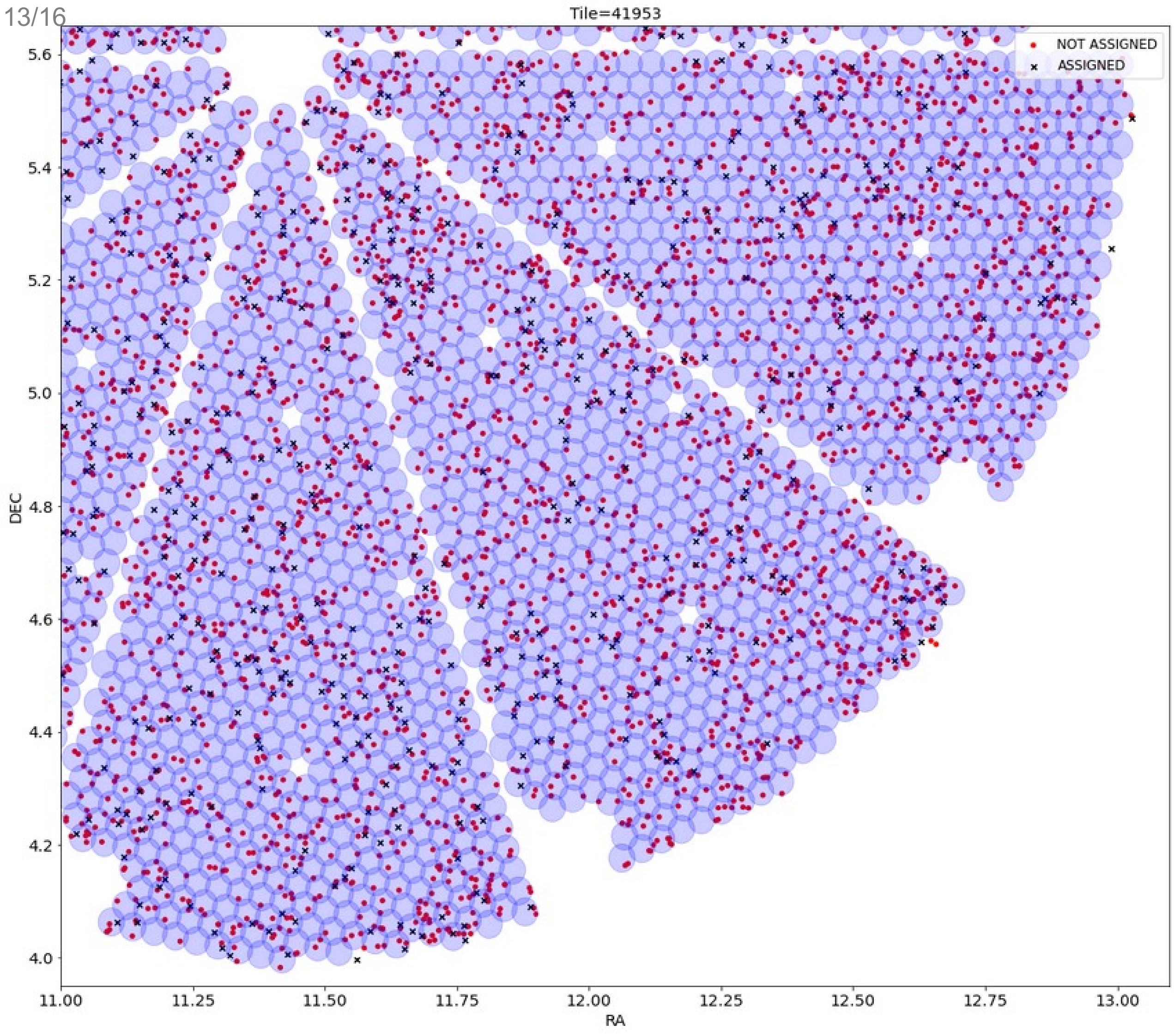


Selecting data:

(targets, randoms) ~ (2.1×10^5, 2.1×10^6)

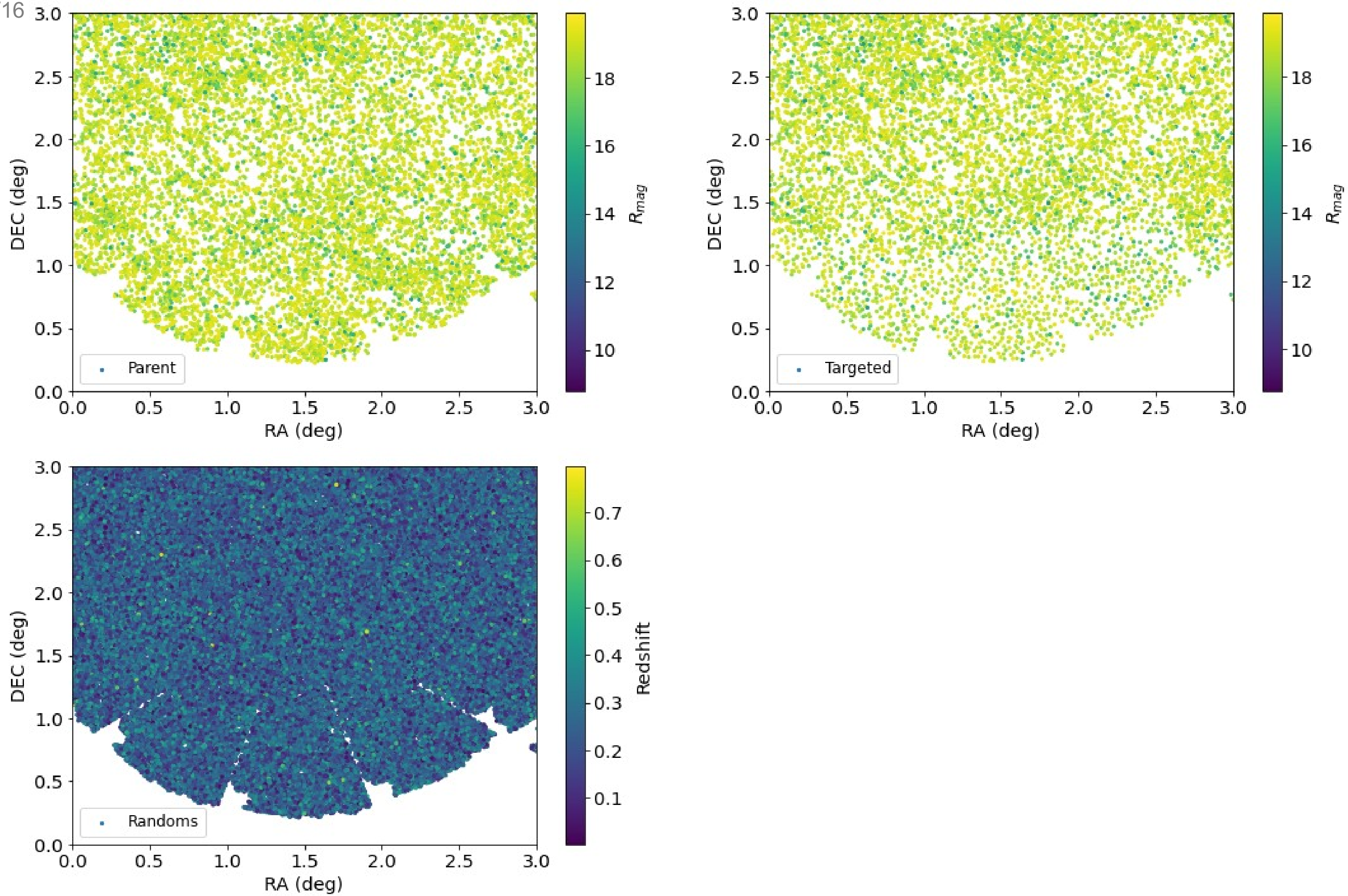
Redshift ~ [0.0024330453,0.79168683]

z_{mean}, σ_z ~ [0.24232799,0.10786242]

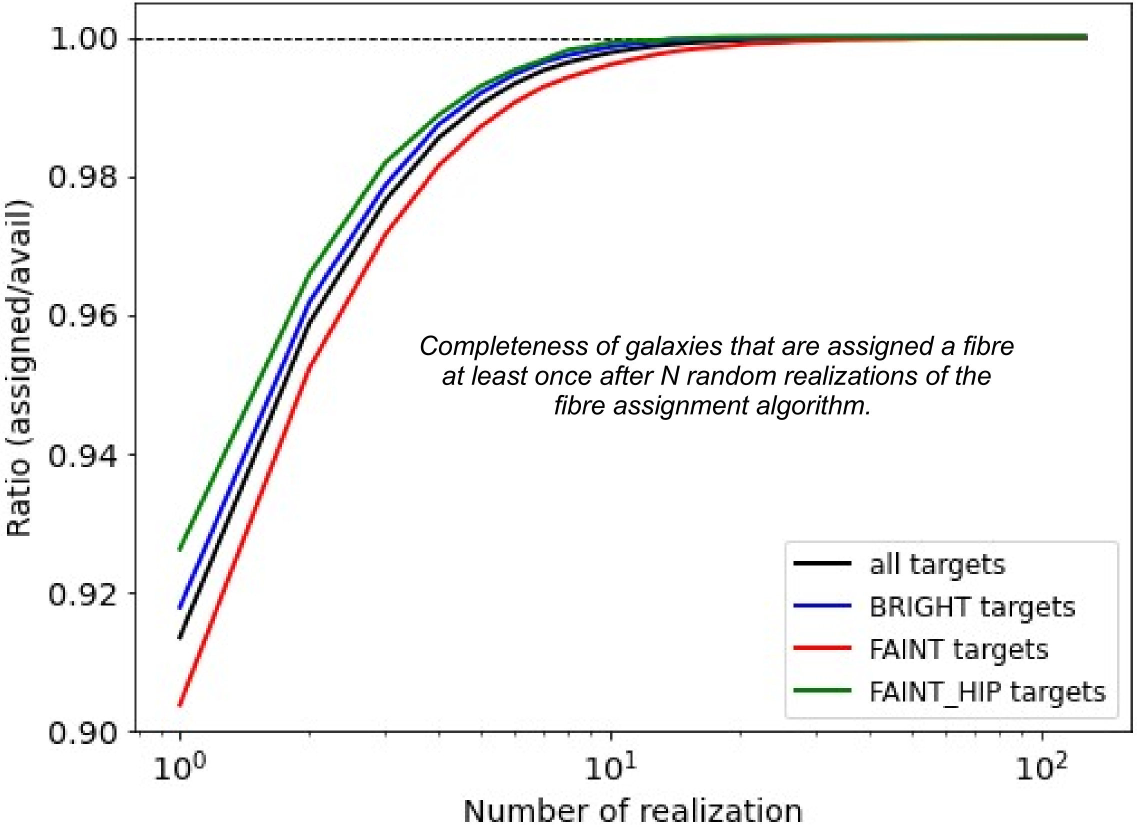


Observed and unobserved positions of galaxies in the BGS mock in fiber regions

- A close-up on a small section around the edge of the mosaic in a survey simulation, showing the positions of the BGS galaxies in relation to the fiber regions.
- The shaded circles indicate the region of each fiber, for a given mosaic.
- White regions cannot be reached with a fiber.
- The crosses indicate galaxies that were successfully assigned a fiber, while the circles show galaxies that could not be assigned to any fiber.

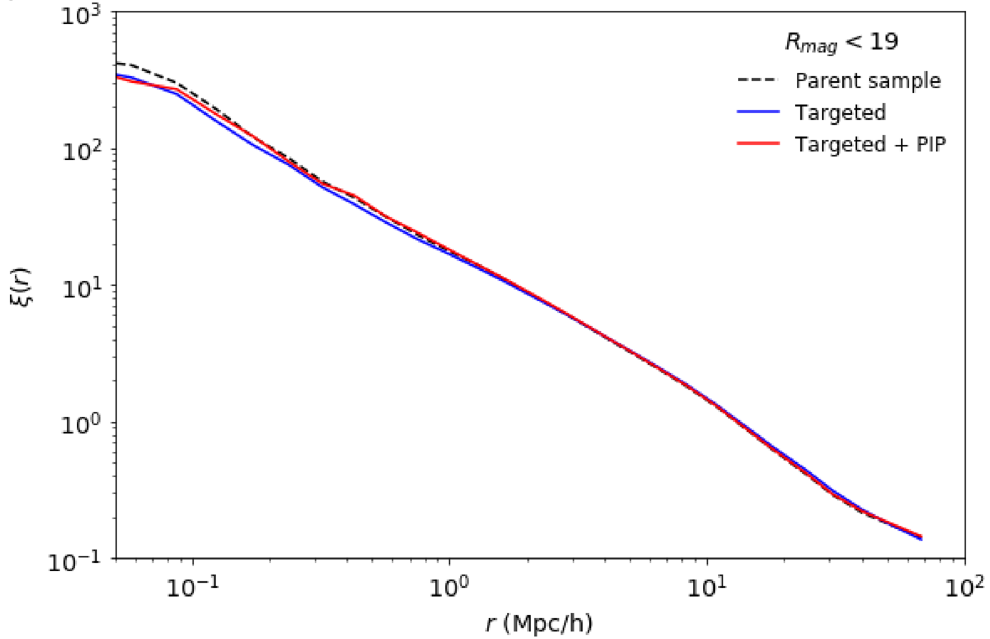

Parent
Targeted
Randoms
RA (deg)
DEC (deg)
R_{mag}
Redshift

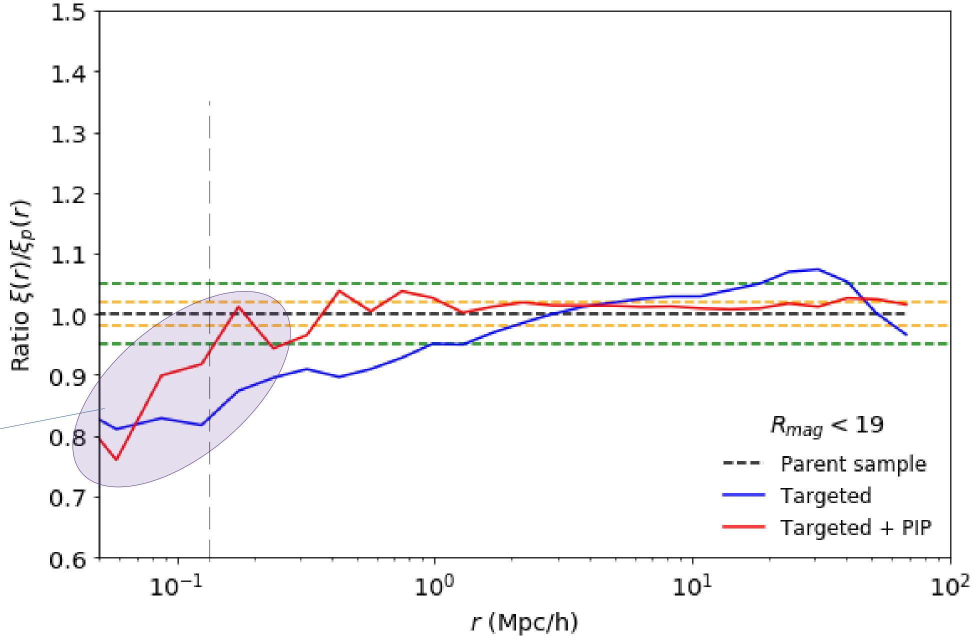
Completeness fiber assignations in all realizations



- The fraction of galaxies that are assigned at least once after 128 runs of the fiberassign algorithm for the targets and randoms data set.
- To achieve 99.99% completeness on the targets with three passes, only 30 realizations are necessary, while for the random set many more than the 128 performed are required.
- There are ∼2 galaxies that are never assigned to any fiber in the 128 realizations.
- This number is very small and can be considered as an effect that does not affect the measurement in the grouping measures.

PIP Implementation





- *For r<0.1 Mpc/h there are errors in the measurement of the correlation function of up to 20% while for r>0.35 Mpc/h the results agree with what was expected very closely in a +/- 2% of error.*