An information theory based search for the scale of cosmic homogeneity

Biswajit Pandey

Department of Physics, Visva-Bharati University

13th January, 2016

TIFR, Mumbai

Large-scale Cosmic Homogeneity

- Cosmological principle: Universe is homogeneous and isotropic

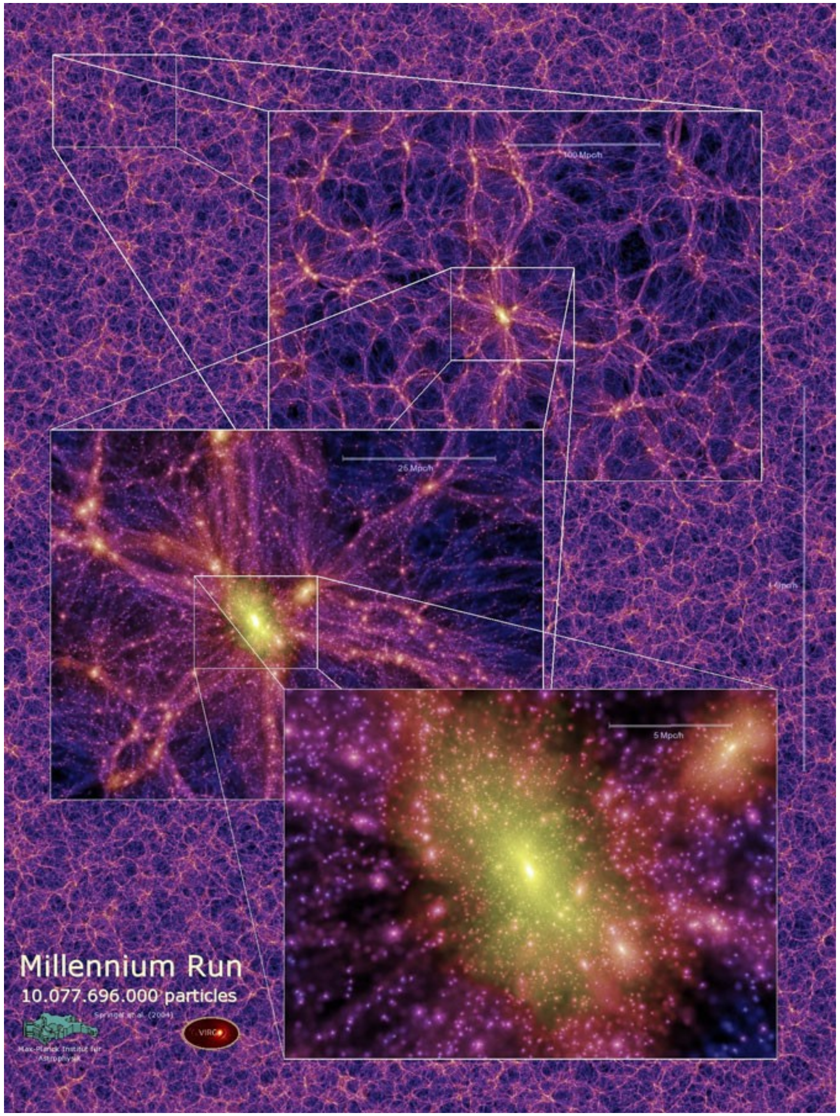
− Homogeneous: different regions of the Universe have the same properties

− Isotropic: looks the same in all directions

- Allows use of Friedmann-Robertson-Walker (FRW) spacetime metric

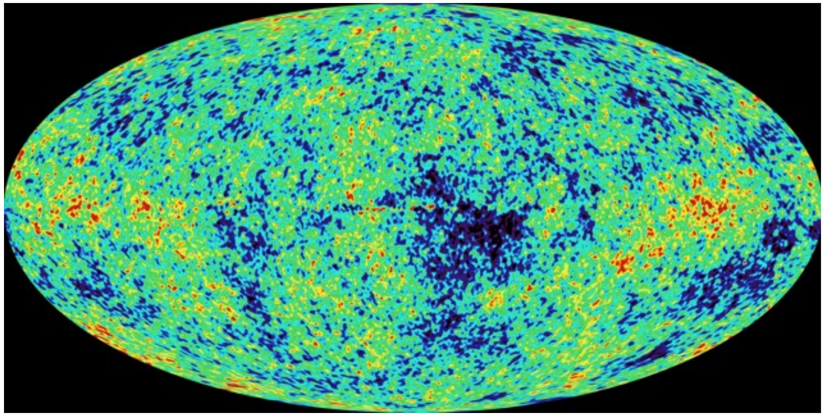
- Need FRW to convert redshifts to distances, via Friedmann eqn:

$$\frac{H^2}{H_0^2} = \Omega_R a^{-4} + \Omega_M a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda .$$



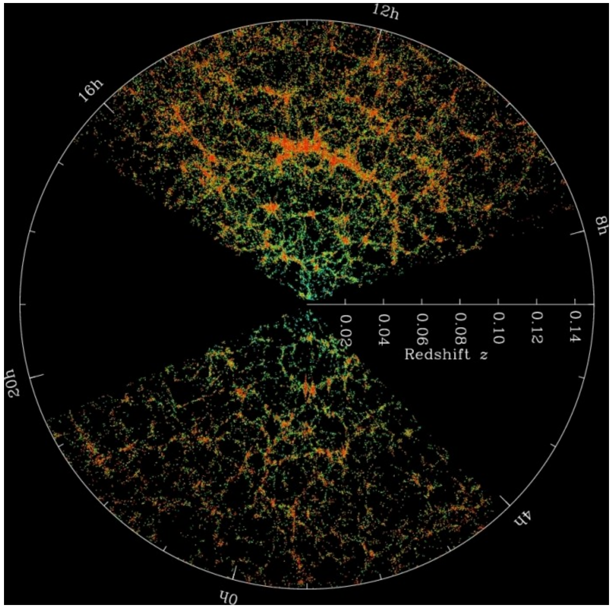
Inhomogeneities in the Universe

- Quantum fluctuations during inflation produce inhomogeneities.
- CMB fluctuations ~ 1 in 100,000.
- Galaxies in the present Universe are clumped.



Current paradigm:

- Gravity amplifies fluctuations.
- Before Recombination - competing effects of gravity and pressure.
- Laws of physics predict what we see today - mainly gravity and electromagnetism.



Inhomogeneity: An alternative to Dark Energy

- Is ‘perturbed FRW’ a valid description?
- Large inhomogeneities → breakdown of FRW

− Light paths distorted: distances inferred from redshifts are wrong (e.g. Wiltshire 2010)

− “Averaging problem” and backreaction: different-density regions evolve differently, can have global accelerated expansion without Dark Energy (e.g. Buchert 2007, Li & Schwarz 2009)

− Void models, e.g. Lemaître-Tolman-Bondi model

Quantifying inhomogeneity

Multifractal Analysis

The generalised correlation integral is defined as

$$C_q(r) = \frac{1}{MN} \sum_{i=1}^{M} [n_i(< r)]^{q-1}$$

where M is the number of centers, N is the total number of galaxies and $q - 1$ refers to a particular moment of the galaxy counts.

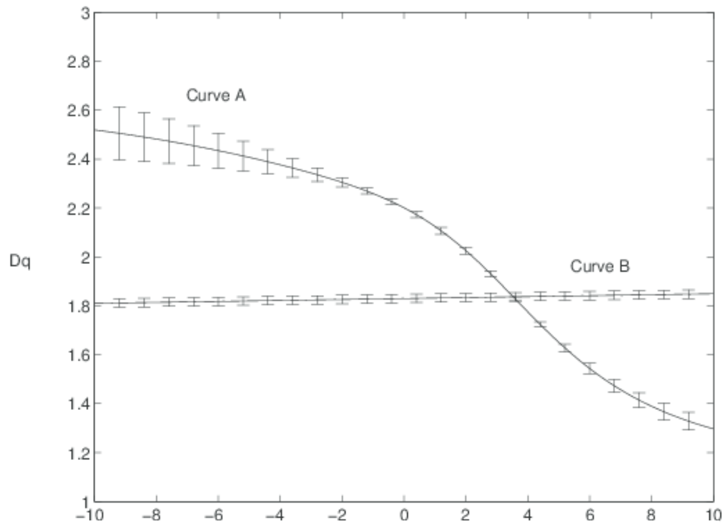
The generalised Minkowski-Bouligand dimension D_q is given by

$$D_q(r) = \frac{1}{q-1} \frac{d \log C_q(r)}{d \log r}$$

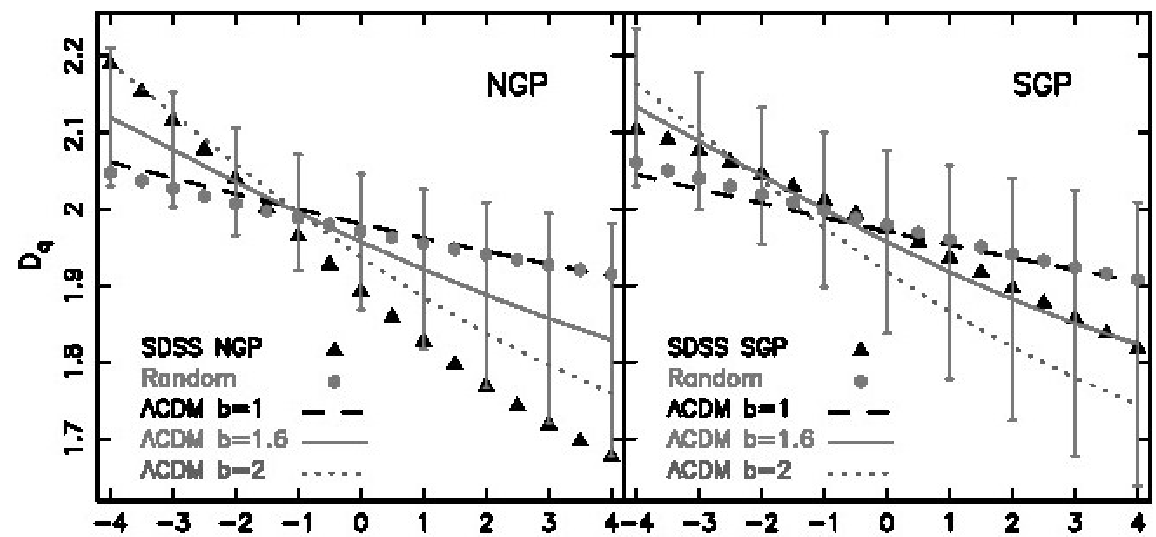
$$C_2(r) \propto r^{D_2}$$

$D_2 = 3$ for a homogeneous distribution.

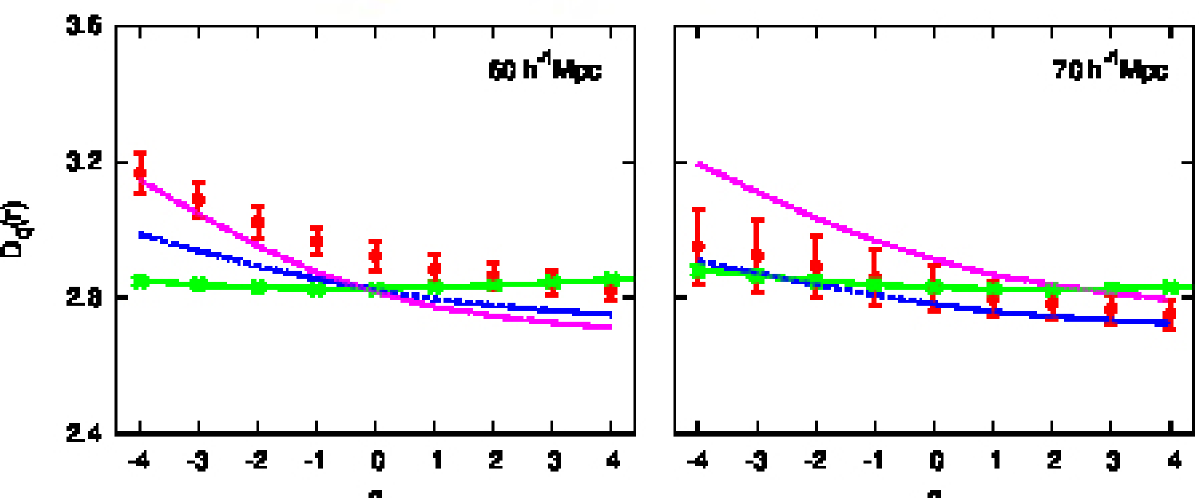
Some results.....



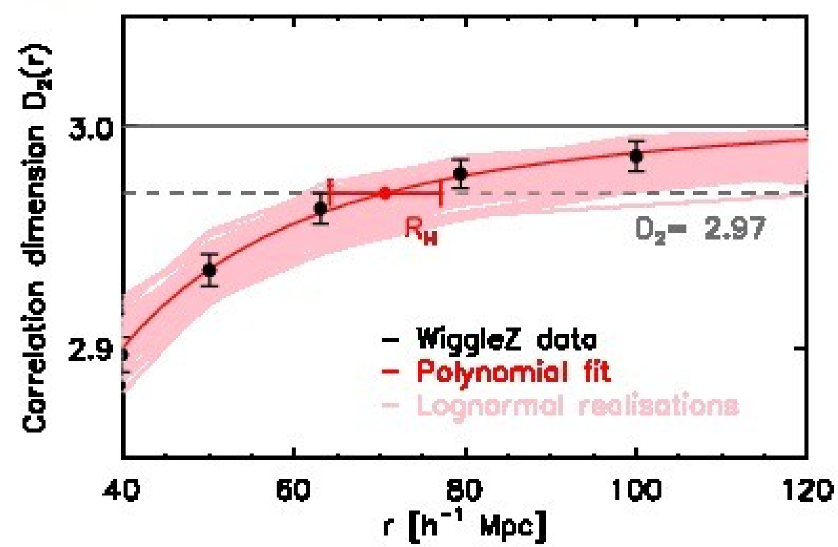
Bharadwaj et al., 1999



Yadav et al., 2005



Sarkar et al., 2009



Scrimgeour et al., 2012

So far the results are conflicting!

Homogeneity at $\sim 70-80h^{-1}$Mpc

LCRS - Bharadwaj et al. (1999)

SDSS DR1 - Yadav et al. (2005)

SDSS LRGs - Hogg et al. (2005)

SDSS DR6 - Sarkar et al. (2009)

WiggleZ - Scrimgeour et al. (2012)

But several other works find fractal structures upto the scale of the survey indicating no transition to homogeneity.

Coleman & Petronero (1992)

Amendola & Palladino (1999)

Sylos Labini et al. (2007,2009)

Sylos Labini (2011)

Information Entropy

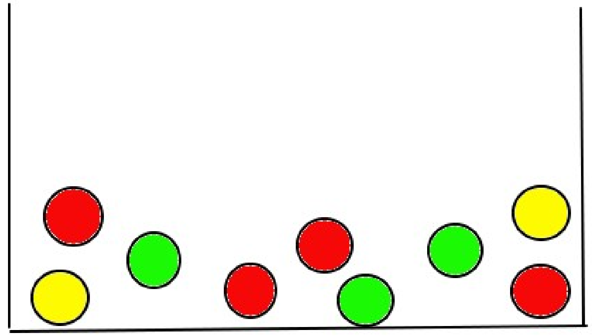
Shannon (1948)

The Information entropy for a discrete random variable X with n outcomes $\{x_i : i = 1, ...n\}$ is a measure of uncertainty denoted by $H(X)$ defined as,

$H(X) = -\sum_{i=1}^{n} p(x_i) \, log \, p(x_i)$

The information is $I(p) = -\log_b(p)$, where b is the base and p is the probablity of the event happening. So the total information from N occurences $I = -\sum_{i=1}^{n} N \times p(x_i) \, log \, p(x_i)$ and therefore entropy $H(x)$ is the expected average amount of information in a certain event.

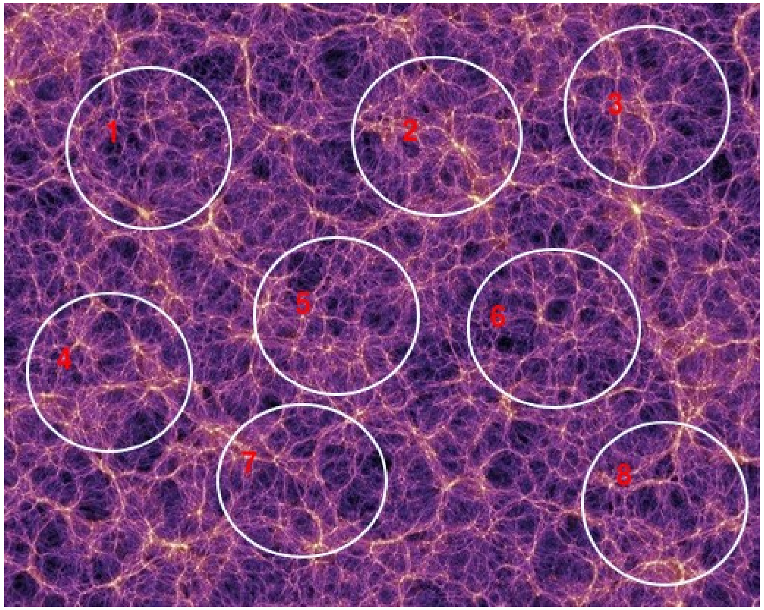
A simple example



- Pick a ball randomly from the bucket.
- The ball can be either red, green or yellow, total $n = 3$ outcomes.
- Entropy$=(-\frac{4}{9}\log\frac{4}{9})+(-\frac{3}{9}\log\frac{3}{9})+(-\frac{2}{9}\log\frac{2}{9}) = 1.5304755$ implies that you are expected to get 1.5304755 bits information each time you choose a ball from the bucket.
- Minimum entrpy $= 0$ occurs when one of the probabilities is 1 and rest are 0.
- Maximum entropy $= \log(n)$ occurs when all the probabilties have equal values of $\frac{1}{n}$.

Shannon entropy as a measure of homogeneity

Pandey(2013)



- Place a number of spheres with radius r and labell them. At each radius r only a certain number say $M(r)$ spheres can be placed depending on the volume and geometry of the region.
- Go to each galaxy and ask which particular sphere does it belong to ? There are $M(r)$ likely outcomes for each of them. We define a random variable X_r for each radius r which has $M(r)$ possible outcomes each given by,

$$f_{i,r} = \frac{n_i(<r)}{\sum_{i=1}^{M(r)} n_i(<r)}$$

with the constraint $\sum_{i=1}^{M(r)} f_{i,r} = 1$.

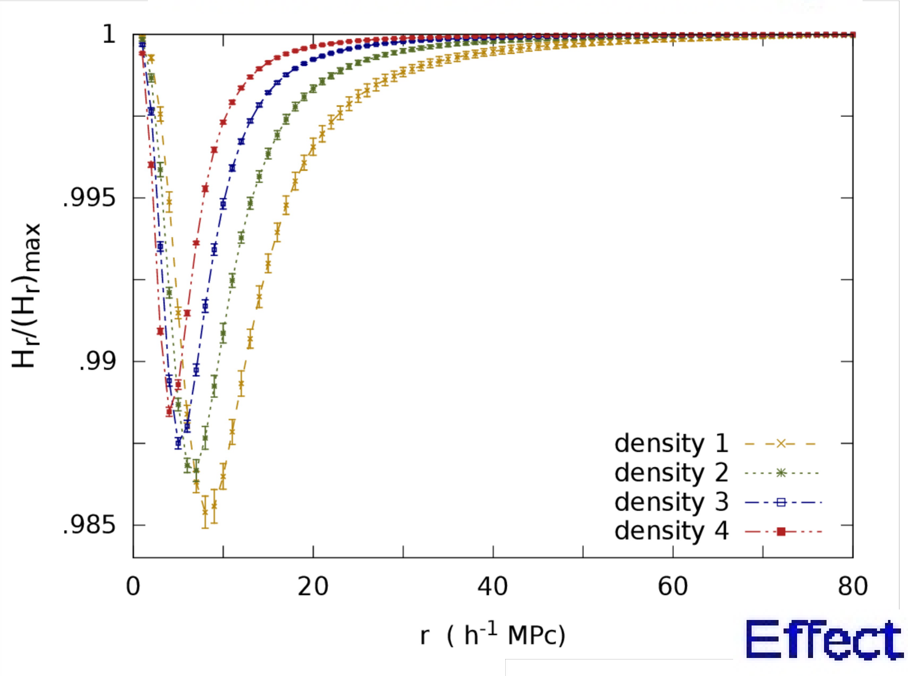
- The Shannon entropy associated with the random variable X_r can be written as

$$H_r = -\sum_{i=1}^{M(r)} f_{i,r} \log f_{i,r}$$
$$= \log(\sum_{i=1}^{M(r)} n_i(< r)) - \frac{\sum_{i=1}^{M(r)} n_i(<r) \log(n_i(<r))}{\sum_{i=1}^{M(r)} n_i(<r)},$$

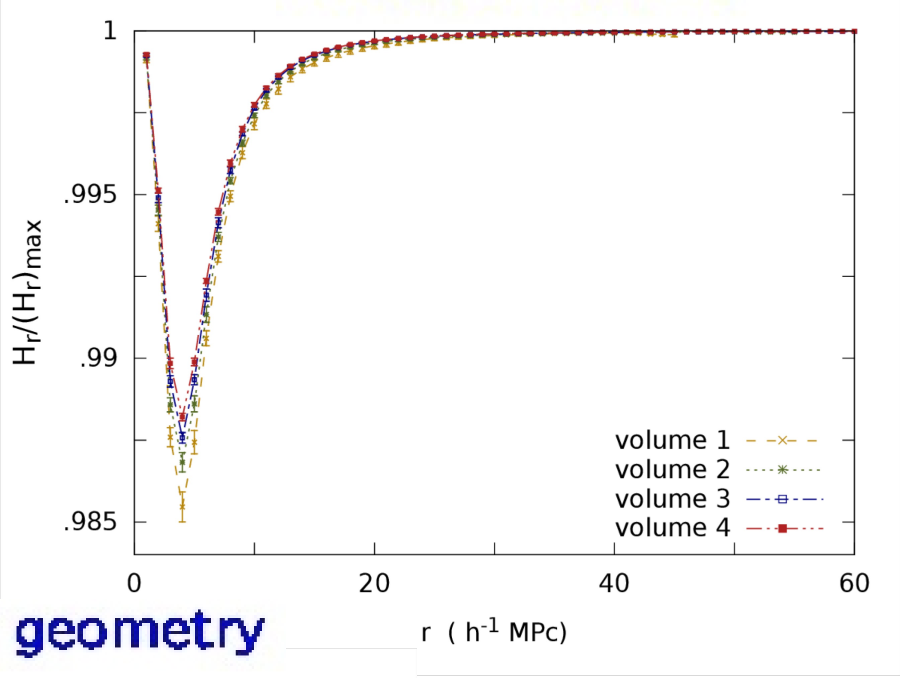
- $f_{i,r}$ will have the same value $\frac{1}{M(r)}$ for all the centers when $n_i(< r)$ is same for all of them which maximizes the Shannon entropy to $(H_r)_{max} = \log M(r)$ for radius r.
- As the max entropy depends on the number of possible outcomes we use the relative Shannon entropy $\frac{H_r}{(H_r)_{max}}$. This quantifies the degree of uncertainty in the knowledge of the random variable X_r at any r. $1 - \frac{H_r}{(H_r)_{max}}$ quantify the information available in X_r at any r.
- The distribution becomes completely uniform when $\frac{H_r}{(H_r)_{max}} = 1$ is reached.

Some tests on 3D Poisson distributions

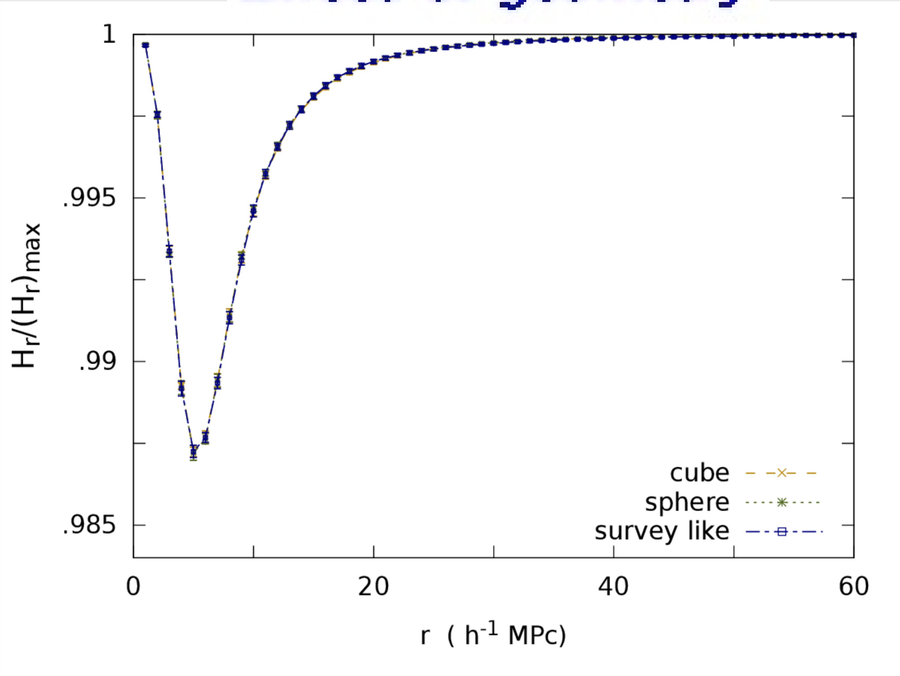
Effect of number density



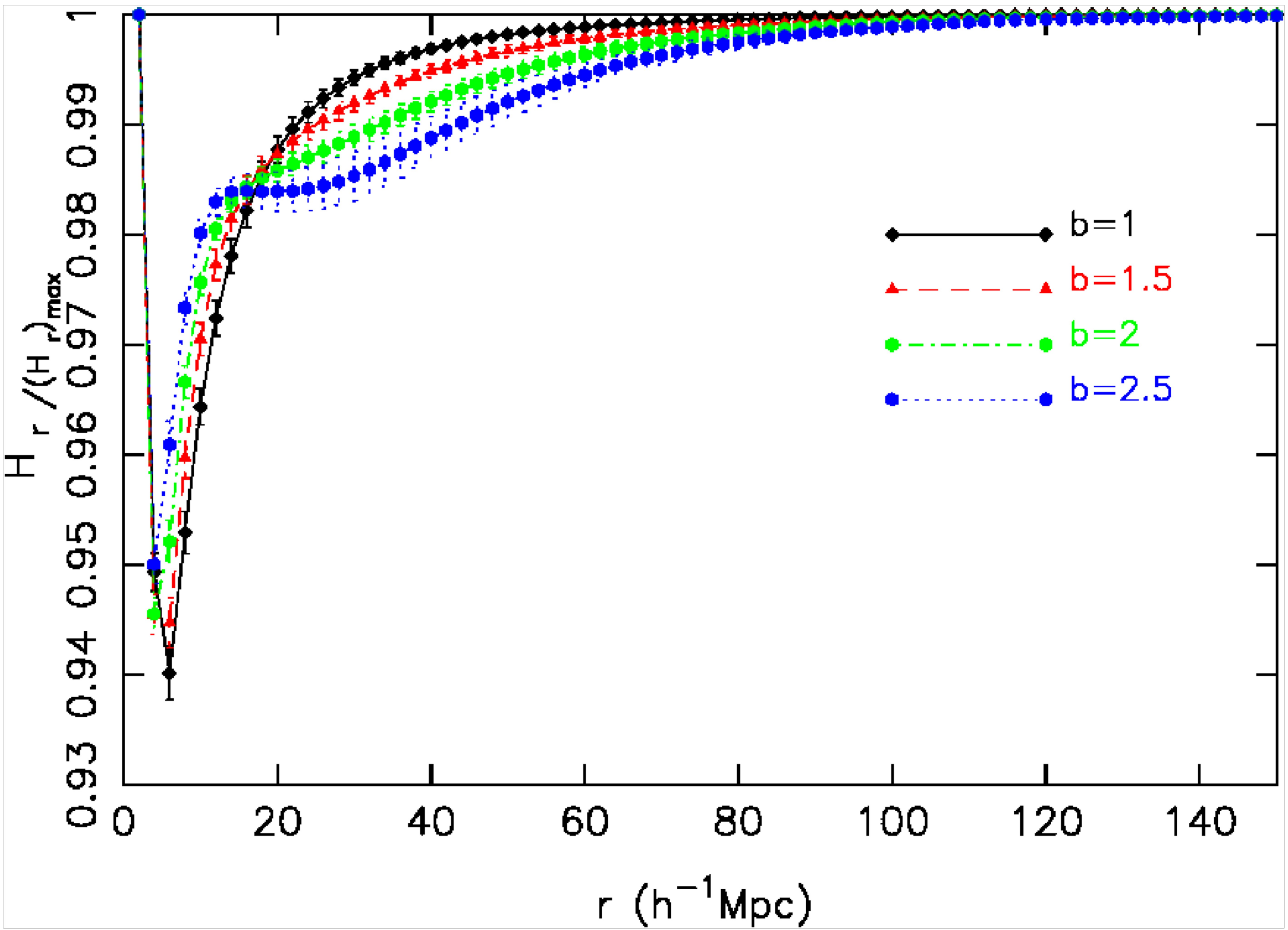
Effect of volume



Effect of geometry



Results for Λ CDM N-body simulations

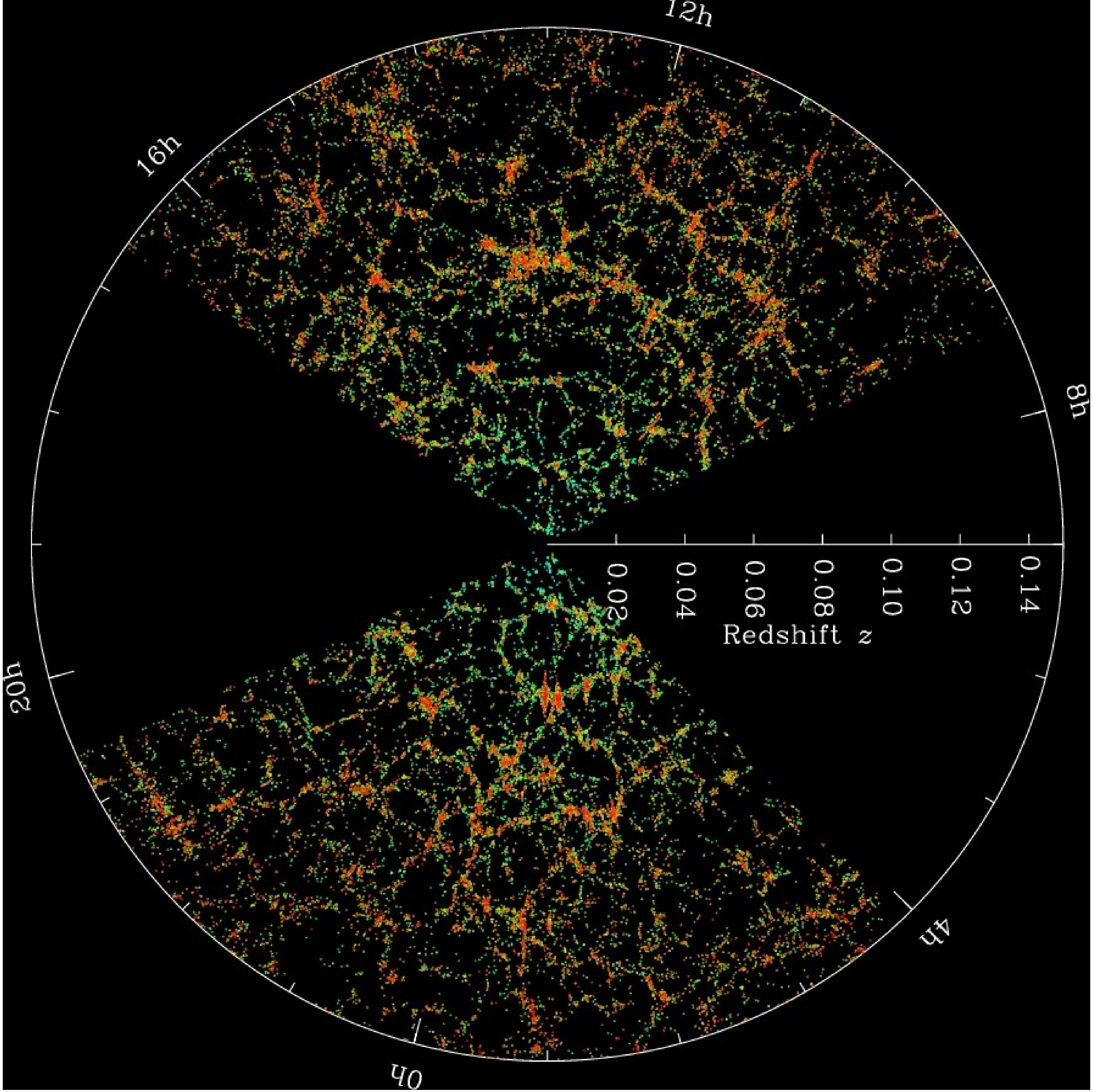


Sloan Digital Sky Survey (SDSS)

- SDSS is the largest galaxy survey todate.

- SDSS is a multi-filter (u,g,r,i,z) imaging and spectroscopic redshift survey using a dedicated 2.5-m wide-angle optical telescope at Apache Point Observatory in New Mexico, United States.

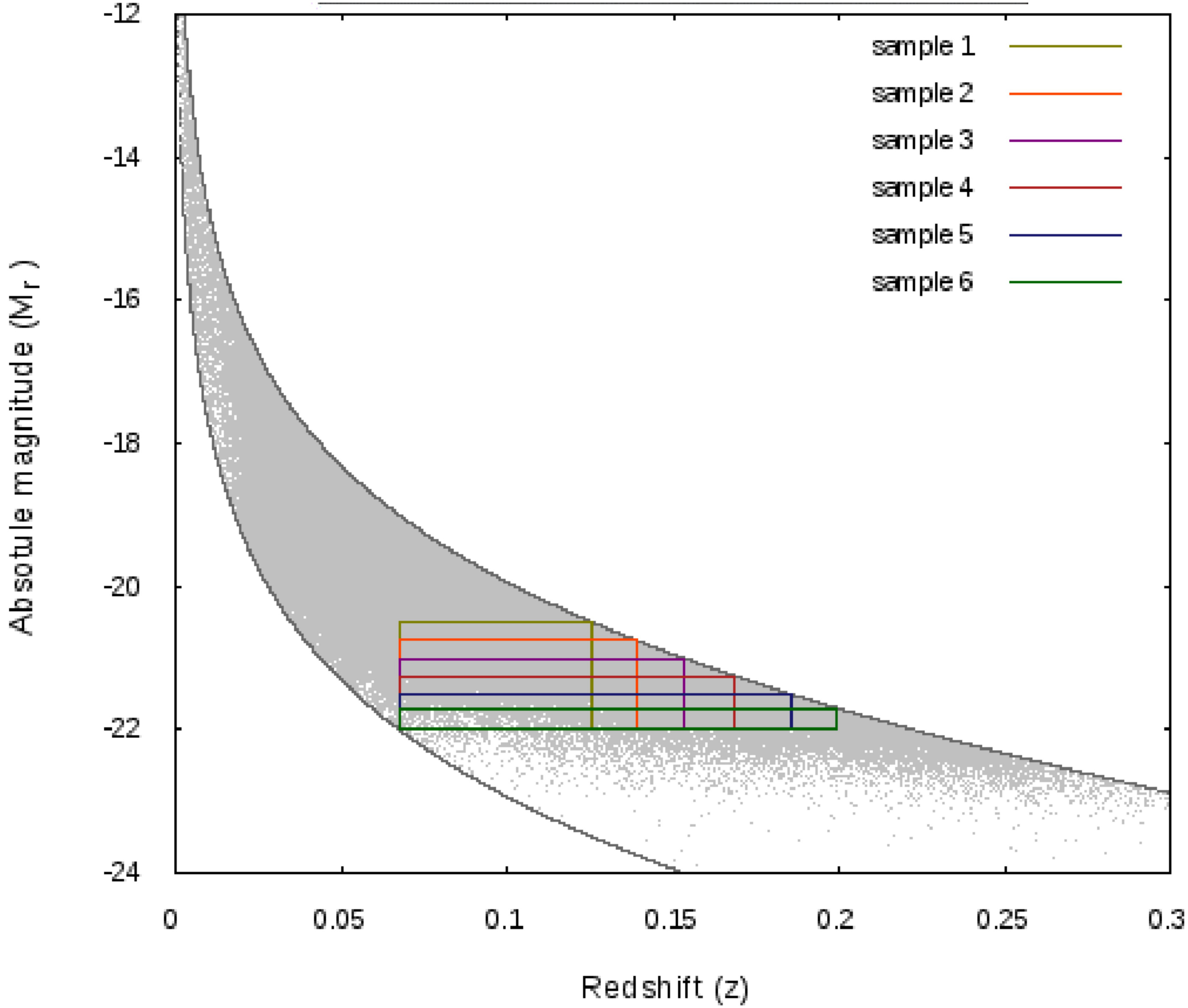
- We use the final data release SDSS DR12 which contains all data taken by all phases of the SDSS through July 14, 2014. This contains altogether optical spectroscopy of 2401952 galaxies and 477161 quasars.



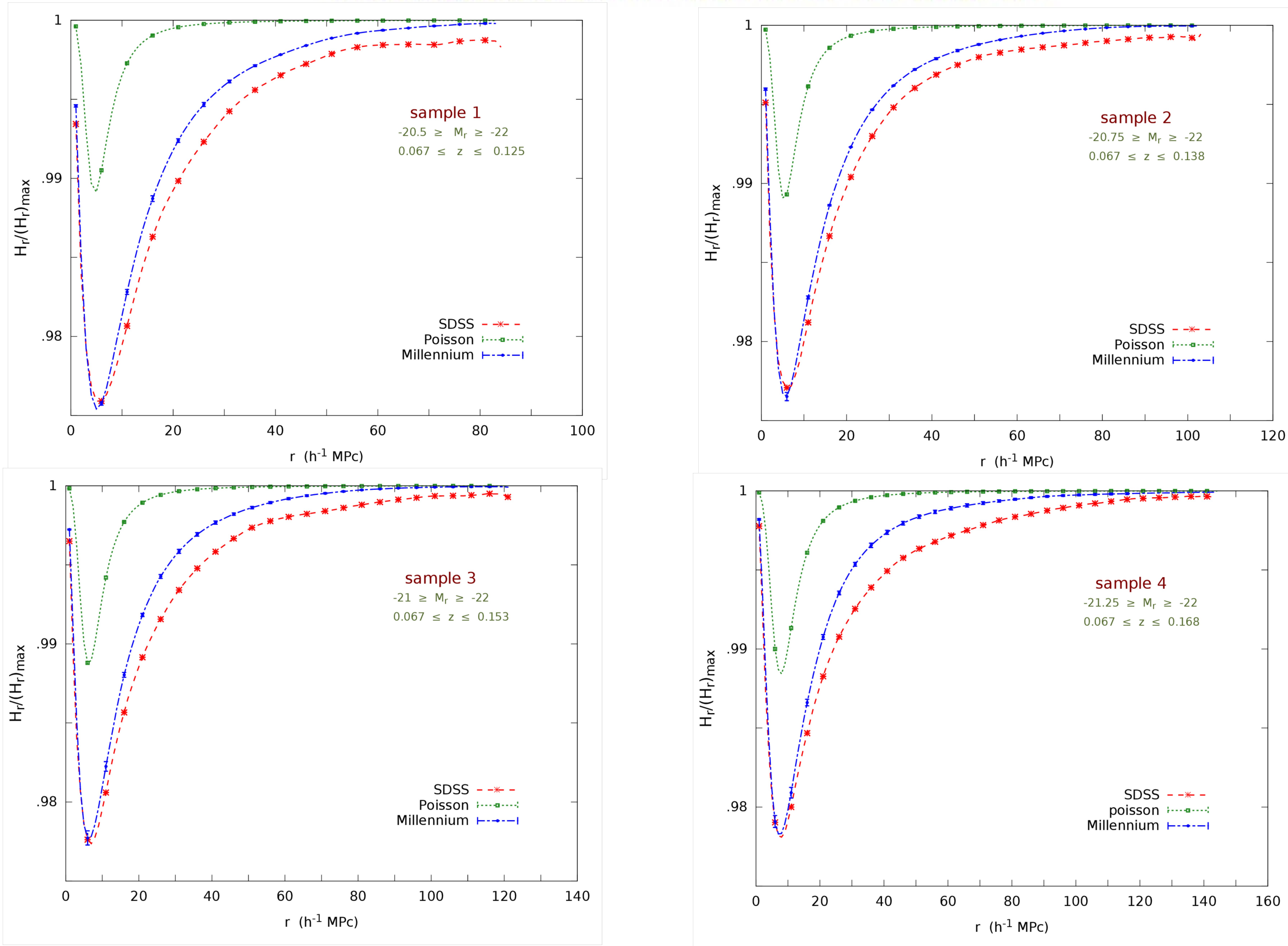
12h
16h
8h
20h
4h
0h
0.02
0.04
0.06
0.08
0.10
0.12
0.14
Redshift z

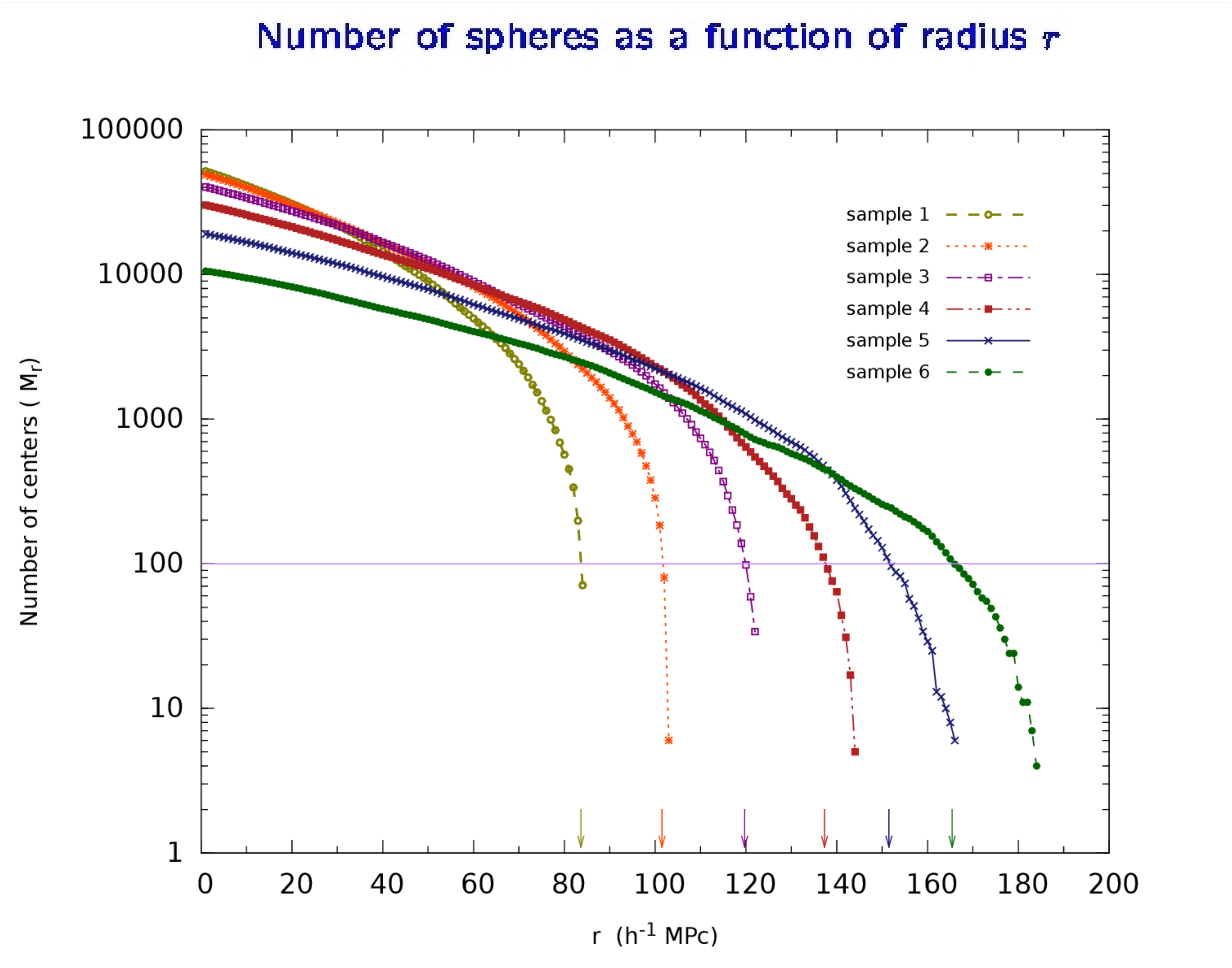
SDSS volume limited samples

Galaxy sample	Absolute magnitude range	Redshift range	Number of galaxies (N)	Volume of the region $(h^{-1}\,\mathrm{Mpc})^3$	Number density $(h^{-1}\,\mathrm{Mpc})^{-3}$	Mean separation $(h^{-1}\,\mathrm{Mpc})$	Largest sphere $(h^{-1}\,\mathrm{Mpc})$
Sample 1	$-20.5 \geq M_r \geq -22$	$0.067 \leq z \leq 0.125$	51 804	1.86×10^7	2.77×10^{-3}	7.1	83
Sample 2	$-20.75 \geq M_r \geq -22$	$0.067 \leq z \leq 0.138$	48 992	2.61×10^7	1.87×10^{-3}	8.1	103
Sample 3	$-21 \geq M_r \geq -22$	$0.067 \leq z \leq 0.153$	40 263	3.58×10^7	1.12×10^{-3}	9.6	122
Sample 4	$-21.25 \geq M_r \geq -22$	$0.067 \leq z \leq 0.168$	30 196	4.84×10^7	6.23×10^{-4}	11.7	144
Sample 5	$-21.5 \geq M_r \geq -22$	$0.067 \leq z \leq 0.185$	19 063	6.45×10^7	2.95×10^{-4}	15	167
Sample 6	$-21.7 \geq M_r \geq -22$	$0.067 \leq z \leq 0.199$	10 567	8.05×10^7	1.31×10^{-4}	19.6	185



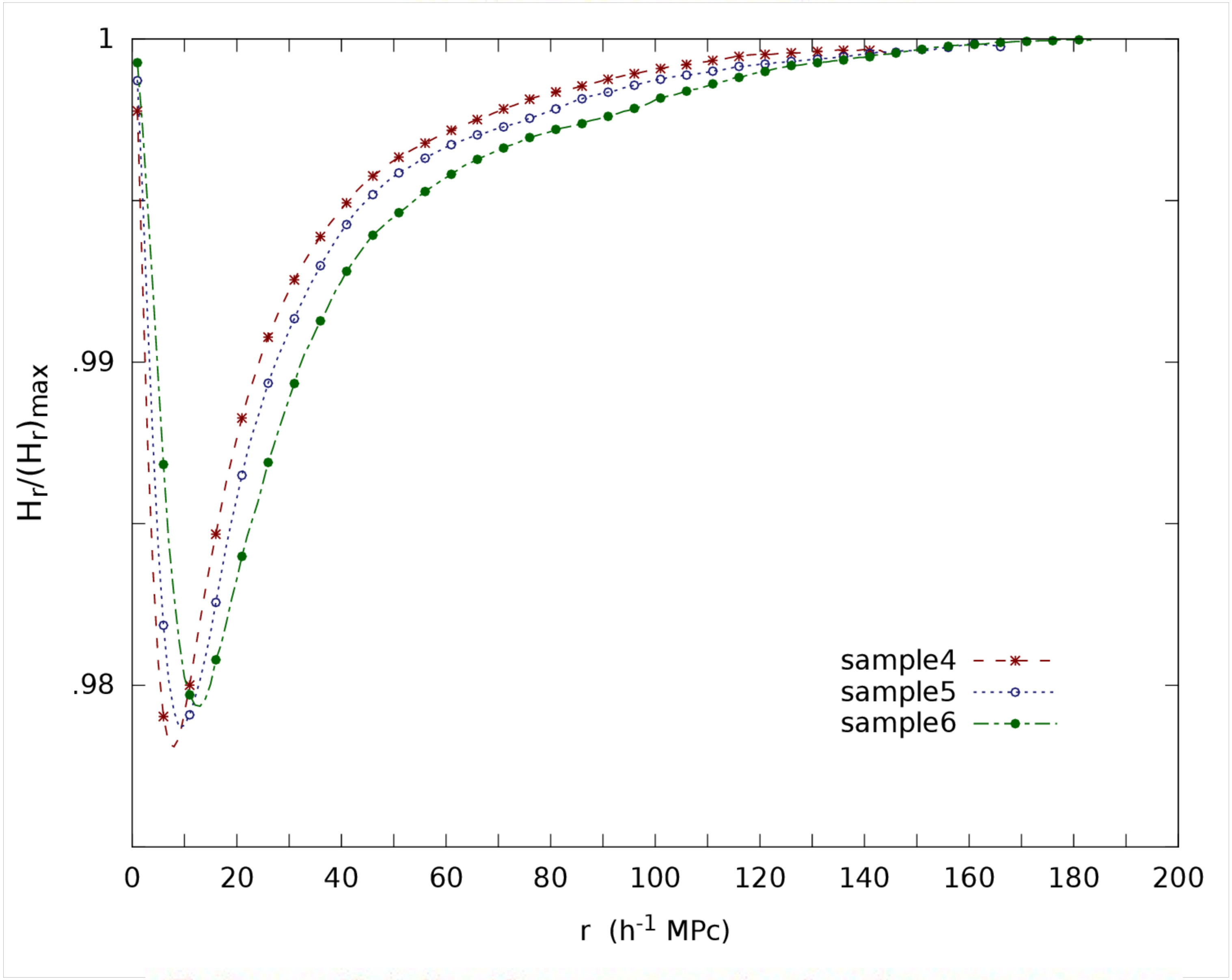
Results for the SDSS samples and their mock catalogues from Millennium simulation





Number of spheres as a function of radius r
100000
10000
1000
100
10
1
0
20
40
60
80
100
120
140
160
180
200
Number of centers (M_r)
r (h-1 MPc)
sample 1
sample 2
sample 3
sample 4
sample 5
sample 6

Results for the sample 6 (largest) along with sample 5 and sample 4



Galaxy distribution seems to be homogeneous above $140h^{-1}$Mpc.

Pandey and Sarkar (2015)

The results should be taken with a grain of salt!.

Confinement bias and Overlapping bias are matters of real concern.

Effects of overlapping and confinement

- At each r we use a finite set of spheres $S_r = \{s_{1,r}, s_{2,r}, s_{3,r}, \ldots, s_{M(r),r}\}$.
- The probability that a randomly drawn point would appear somewhere in the sample is 1. If the spheres are disjoint then at any r there are $M(r)+1$ outcome of this experiment i.e the point would lie either in any one of the $M(r)$ spheres or somewhere in the sample outside the spheres. But given the fact that the spheres overlap the point could also appear at the intersections of multiple spheres.

Effects of overlapping and confinement

- Given a finite sample A the total probability can be written as,

$$P(A) = P(\cup_{i=1}^{M(r)} s_{i,r}) + P((\cup_{i=1}^{M(r)} s_{i,r})^c)$$

$$= \sum_{i=1}^{M(r)} P(s_{i,r})$$

$$- \sum_{i \neq j,1}^{M(r)} P(s_{i,r} \cap s_{j,r})$$

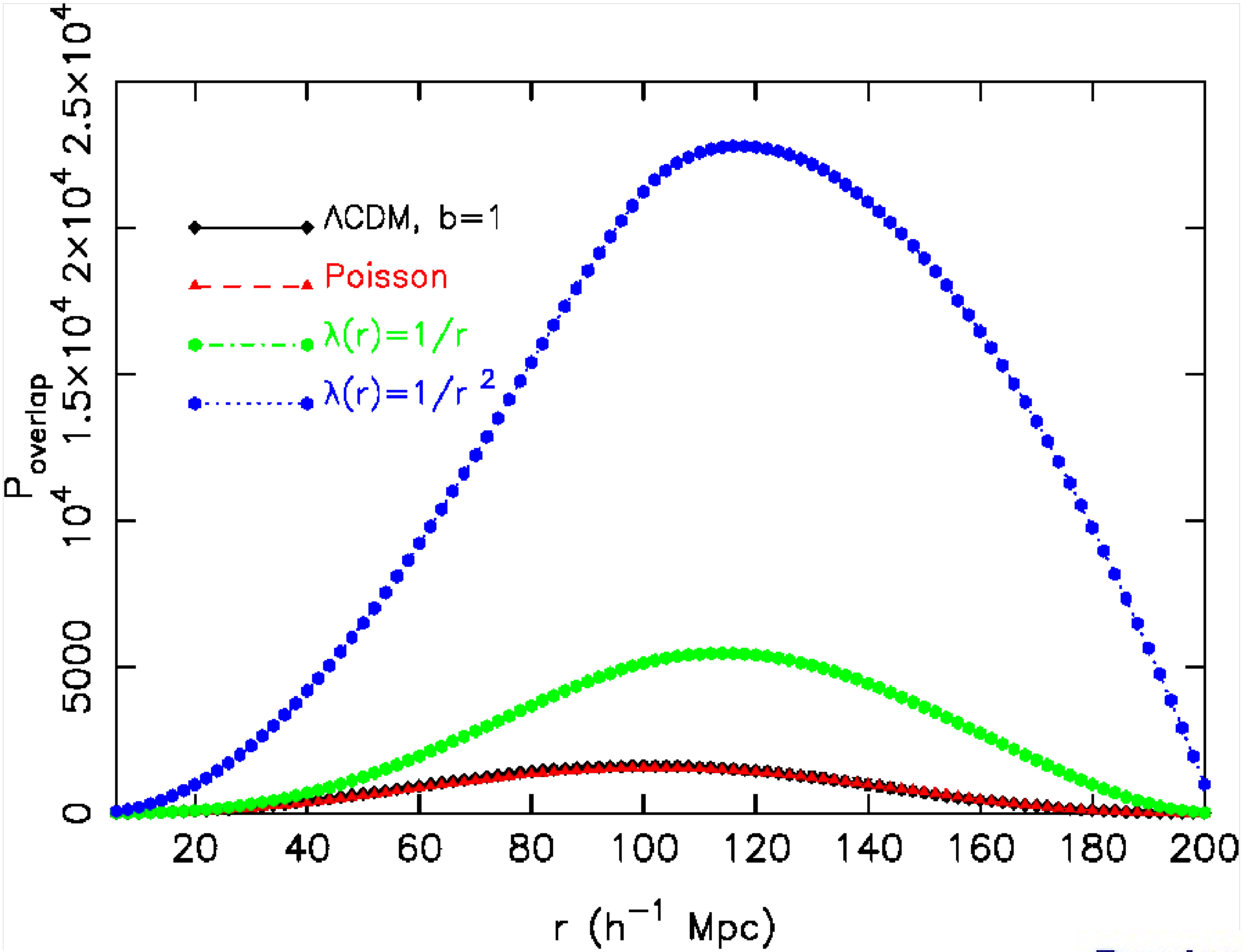
$$+ \sum_{i \neq j \neq k,1}^{M(r)} P(s_{i,r} \cap s_{j,r} \cap s_{k,r})$$

$$- \ldots + (-1)^{M(r)-1} P(\cap_{i=1}^{M(r)} s_{i,r})$$

$$+ P((\cup_{i=1}^{M(r)} s_{i,r})^c)$$

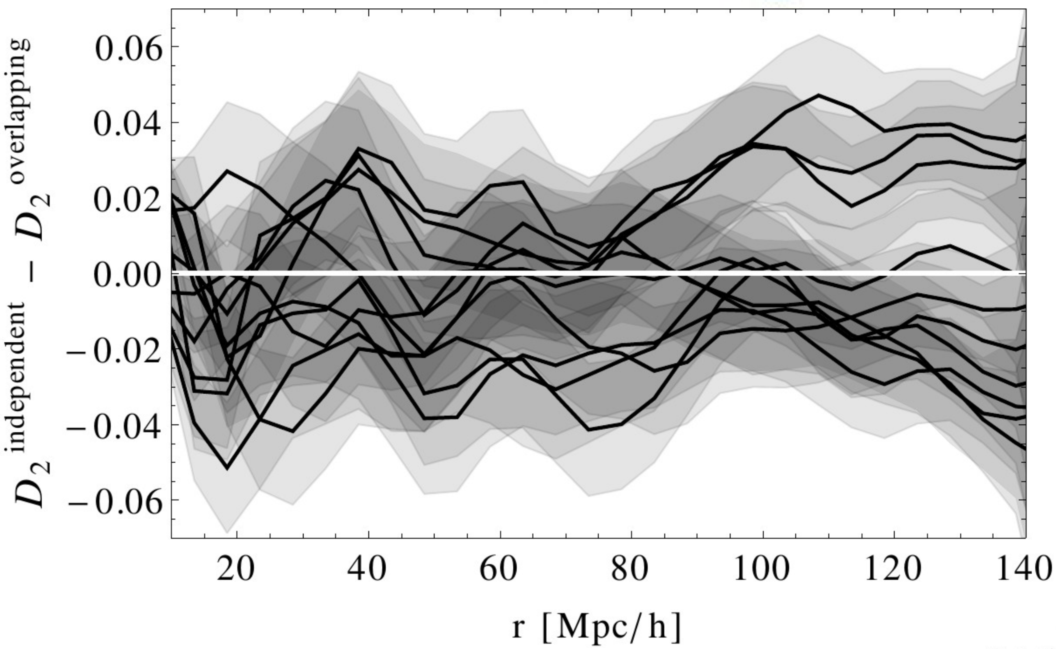
$$= \sum_{i=1}^{M(r)} P(s_{i,r}) - P_{overlap} + P((\cup_{i=1}^{M(r)} s_{i,r})^c) = 1$$

Effects of overlapping and confinement

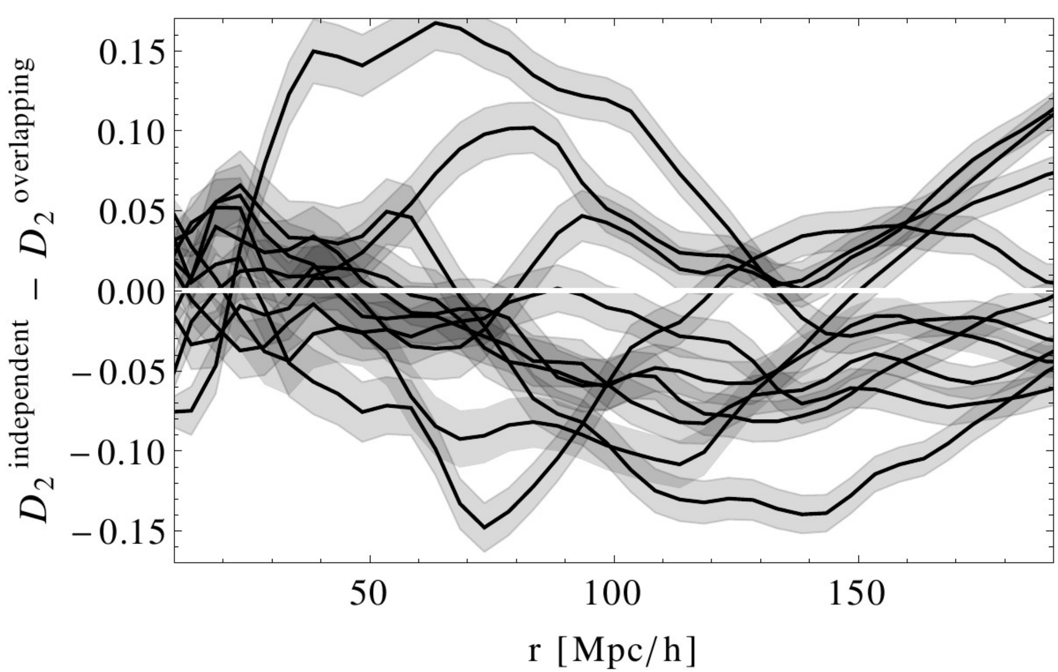


Pandey (2013)

Some recent results on the effects of overlapping and confinement bias on the correlation dimension D_2



Kraljic (2015)



The Millennium Run simulation volume is too small and prone to bias to reliably identify the onset of homogeneity.

The confinement and overlapping biases can be avoided if we have access to 3D galaxy samples covering much larger volumes. The Luminous Red Galaxy (LRG) distribution extends to a much deeper region of the Universe as compared to the SDSS Main galaxy sample.

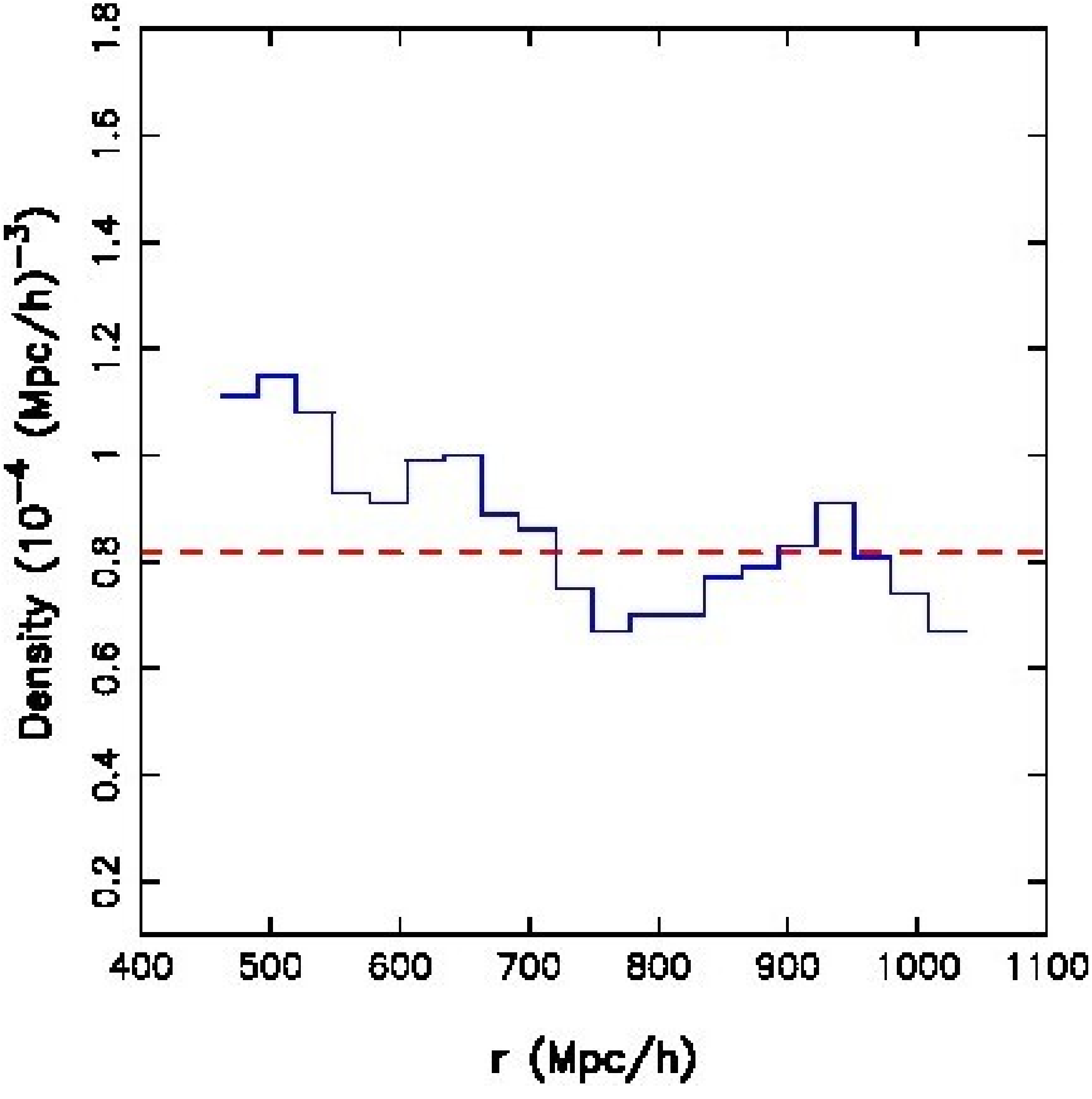
- They can be observed to greater distances as compared to normal L_* galaxies for a given magnitude limit.
- Their stable colors make them relatively easy to pick out from the rest of the galaxies using SDSS multi-band photometry.
- Not too rare, number density $\sim 10^{-4}(Mpc/h)^{-3}$.
- Trace matter well, linear bias $b \sim 2$.

The properties of our Luminous Red Galaxy (LRG) sample from SDSS DR7

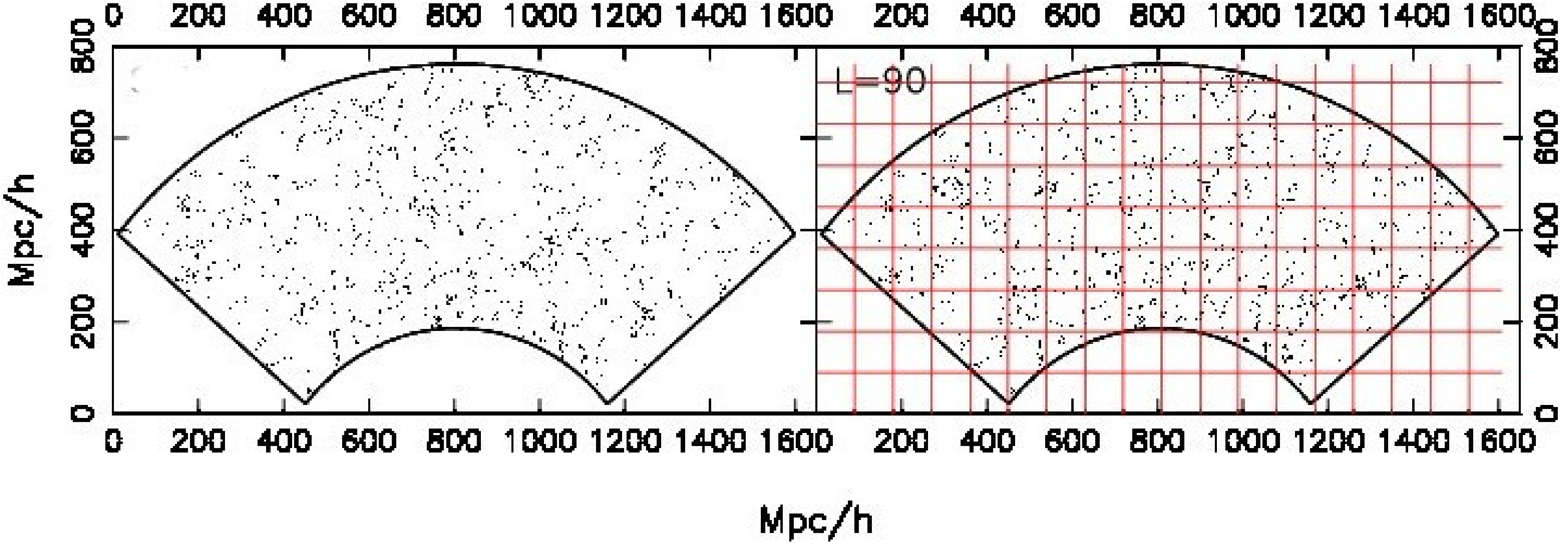
- Absolute magnitude range $-23 < M_g < -21$.
- Redshift range $0.16 < z < 0.38$
- $-50 < \lambda < 50$ and $-33.5 < \eta < 29.5$ where λ and η are the survey co-ordinates.

Combining these cuts provides us a LRG sample which radially extends from $462\,h^{-1}$Mpc to $1037\,h^{-1}$Mpc and consists of 50513 luminous red galaxies.

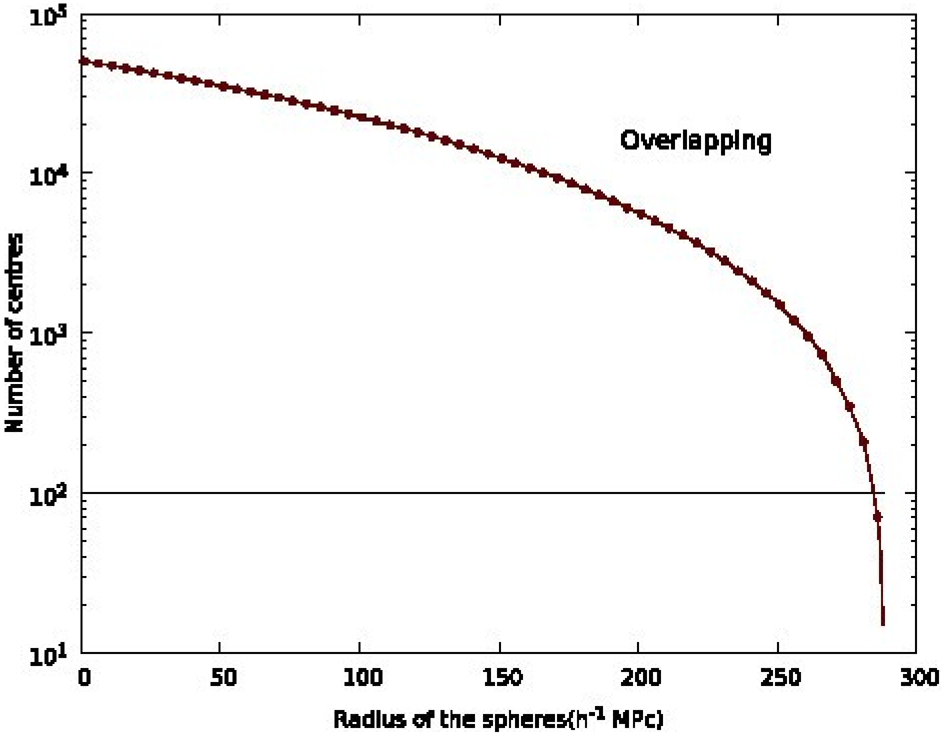
Comoving number density of LRGs

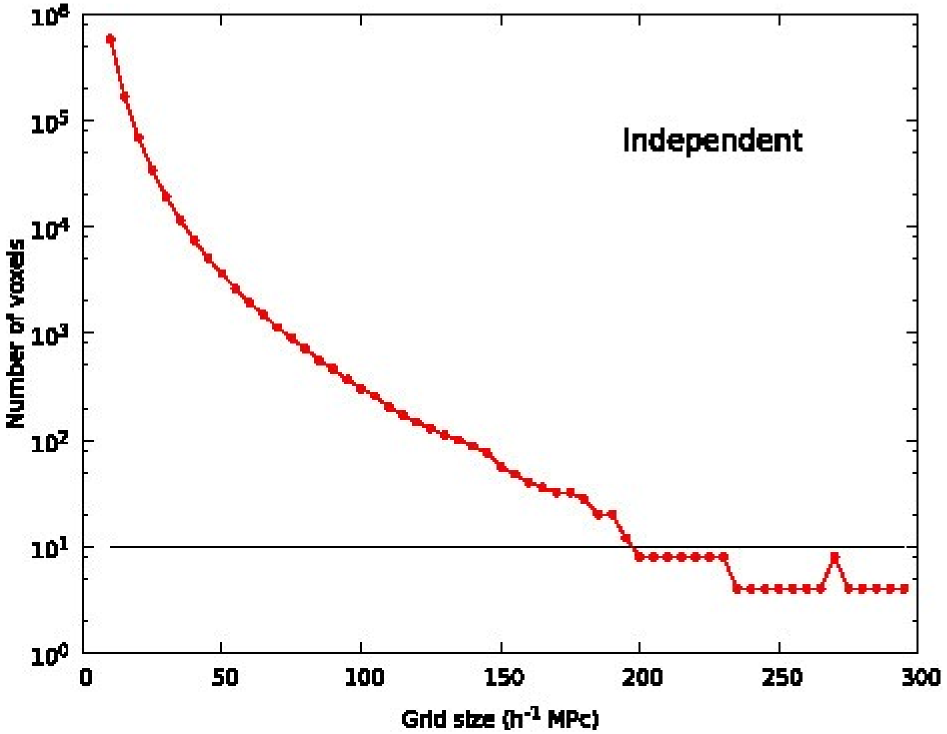


Analysis with non-overlapping regions

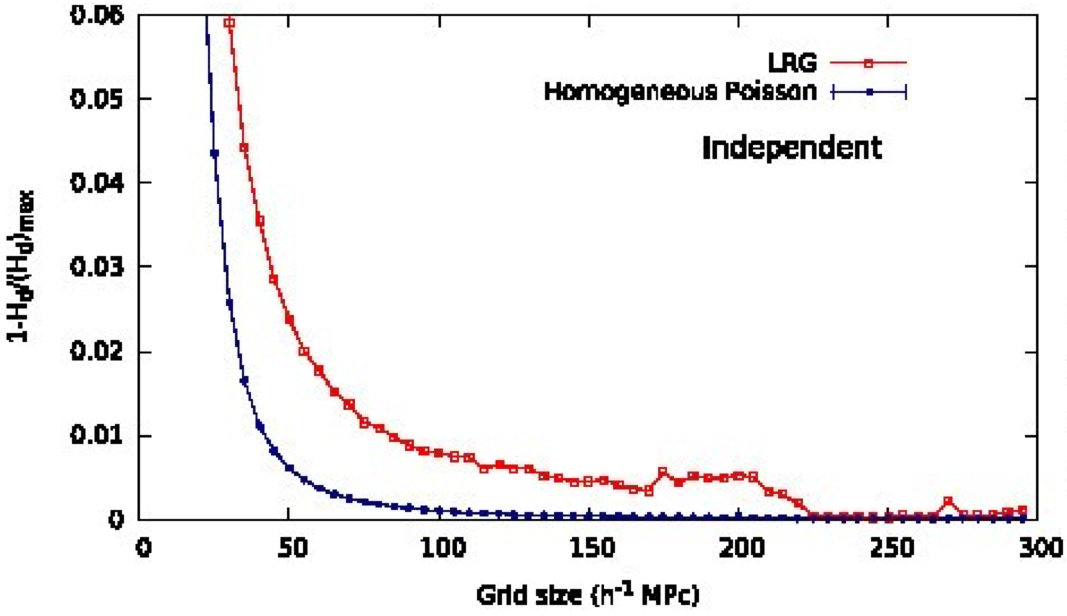


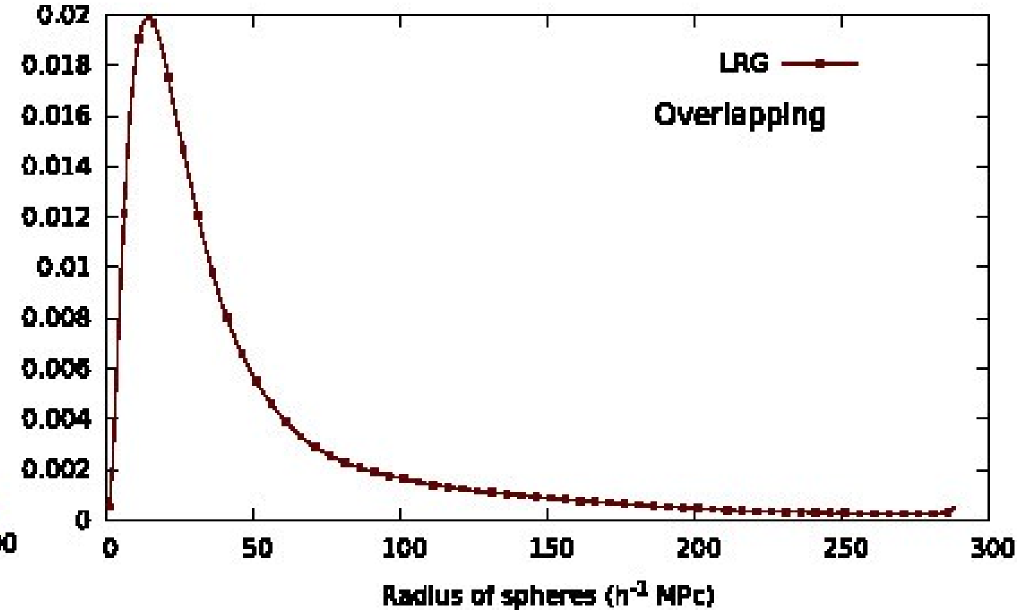
Number of measuring voulmes as a function of length scale

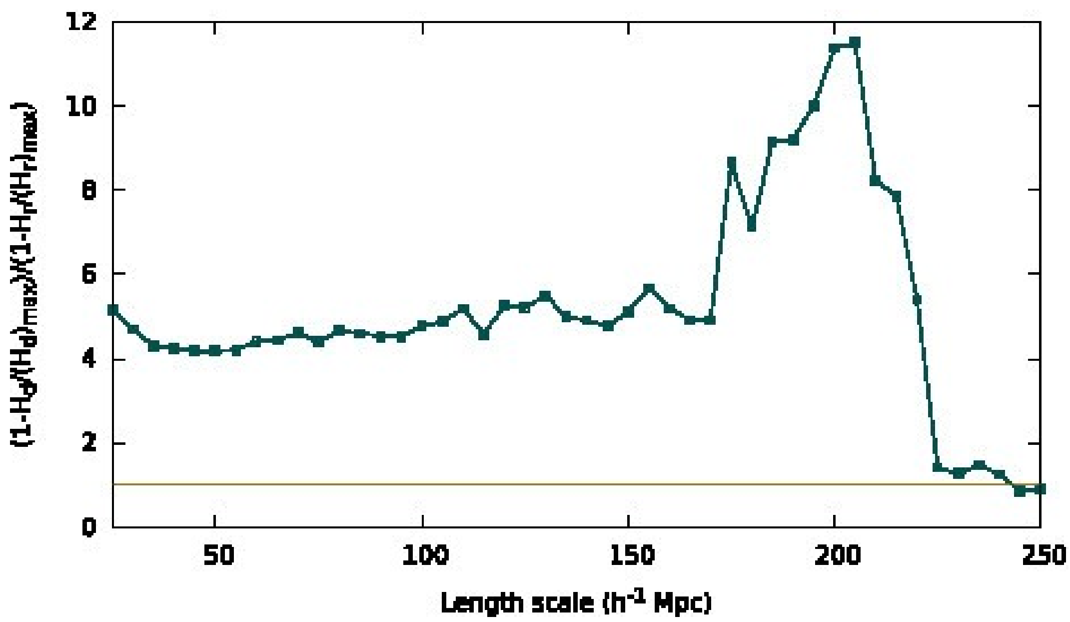




Results for the LRG distribution





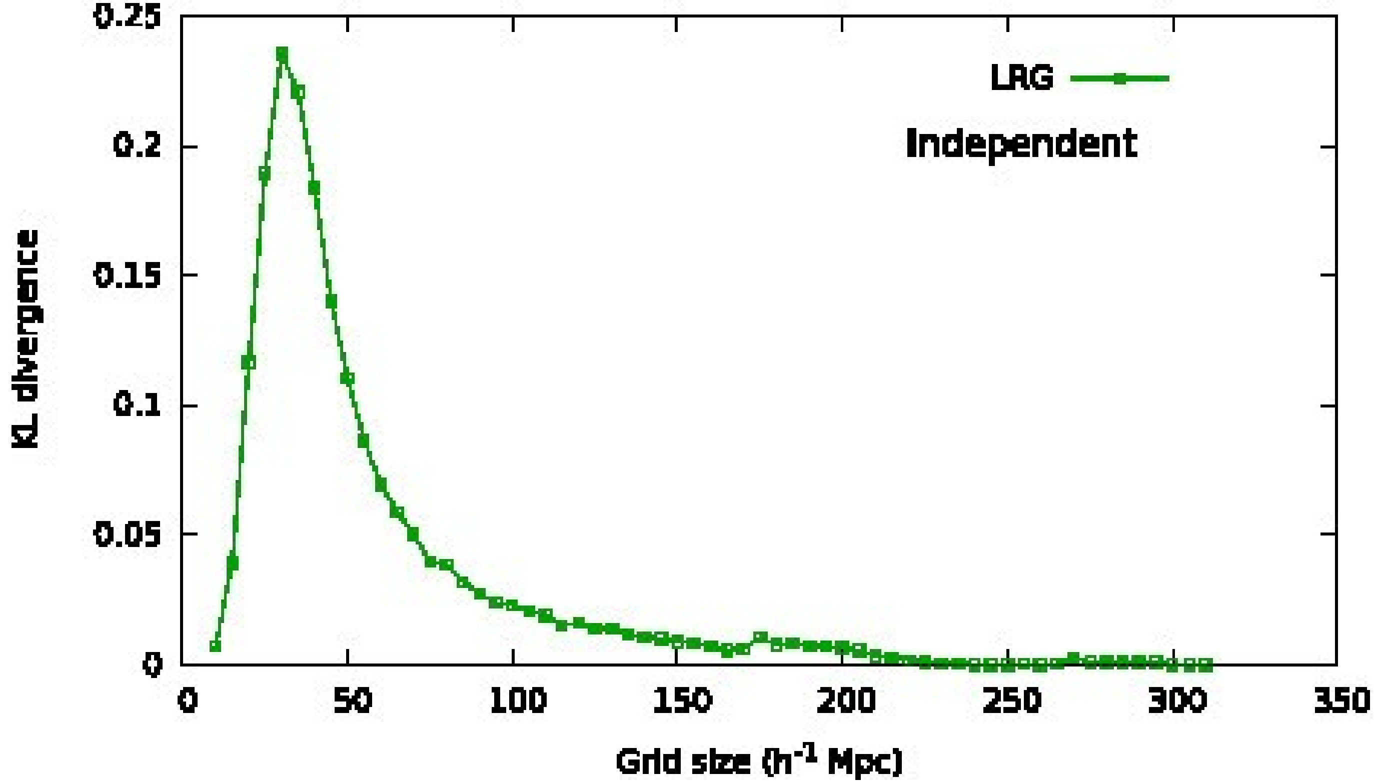


• Kullback-Leibler divergence: An alternative measure of inhomogeneity

In information theory KL divergence is used to measure the difference between two probability distributions $p(x)$ and $q(x)$.
$D_{KL}(p|q) = \sum_i p(x_i) \log \frac{p(x_i)}{q(x_i)}$

The KL divergence between the actual and homogeneous random data is then given by,
$D_{KL}(D|R) = \frac{(\sum_i n_{Di} \log n_{Di} - \sum_i n_{Di} \log n_{Ri})}{\sum_i n_{Di}} - \log \frac{\sum_i n_{Di}}{\sum_i n_{Ri}}$
where n_{Di} and n_{Ri} are the counts in the i^{th} voxel for actual data and random data respectively.

KL divergence measure as a function of length scale for the LRG distribution



- Our studies with the SDSS Main galaxy sample and the Luminous Red Galaxy (LRG) sample suggest that the Universe is homogeneous beyond a scale of $\sim 150\, h^{-1}$Mpc.

THANKS FOR YOUR KIND ATTENTION