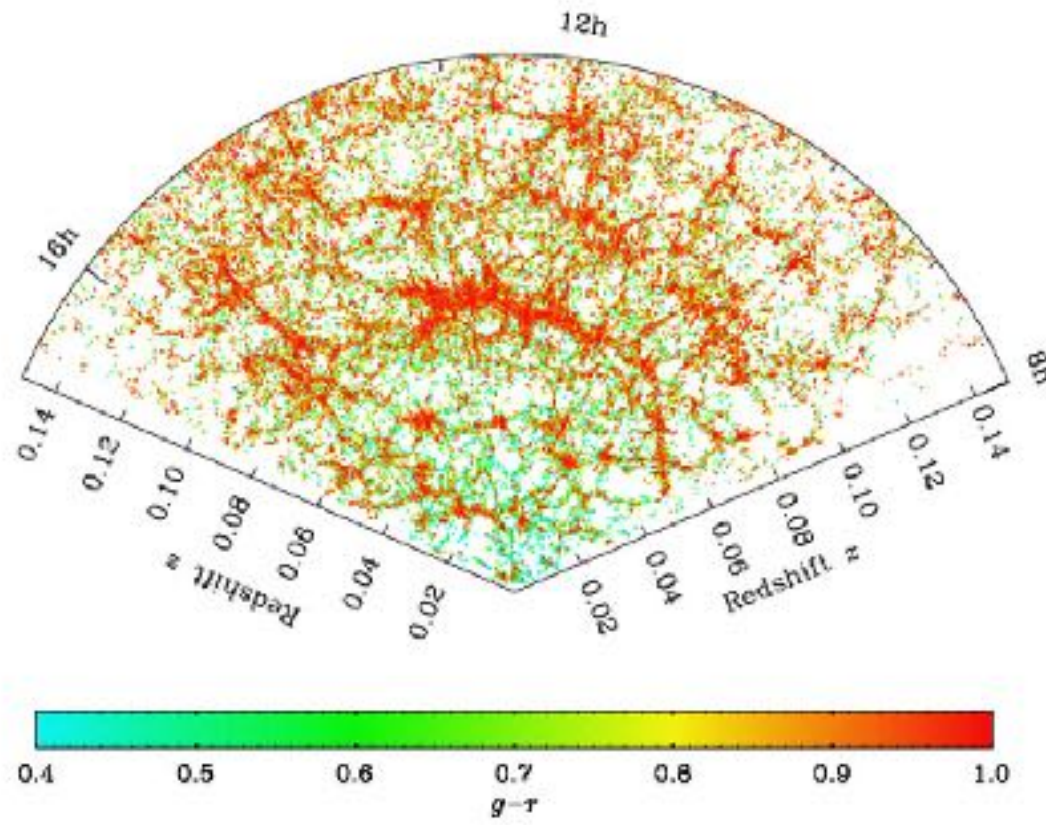
The large-scale structure of the Universe

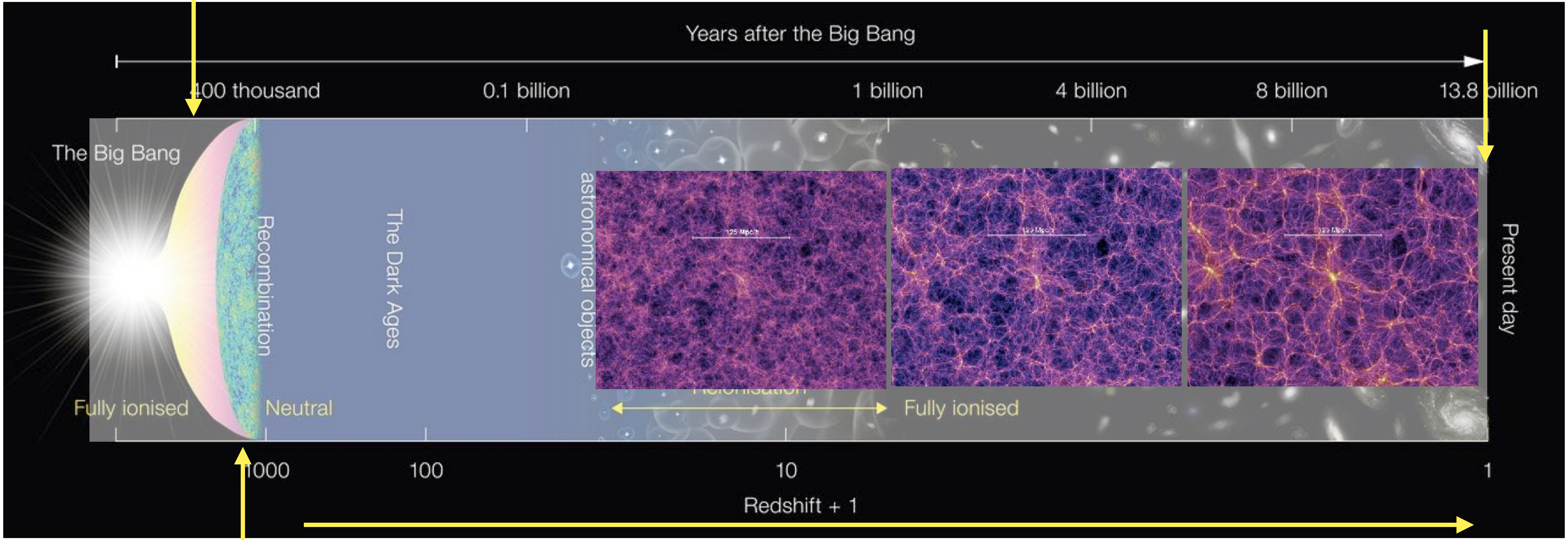
Cristina García Vergara

AstroTwinCoLo (Virtual) School 2020

Previous class: Structure formation and evolution



Inflation

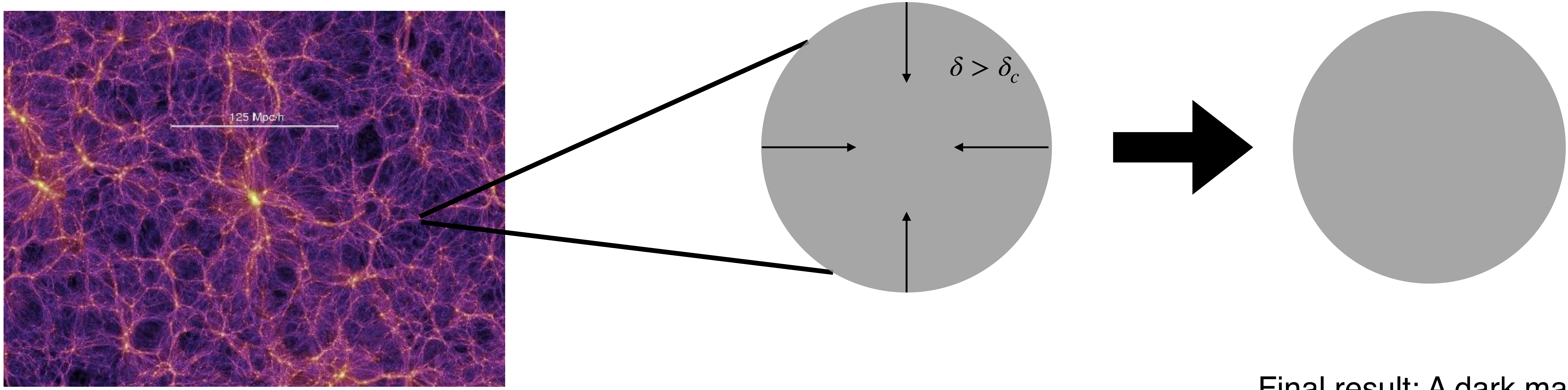


CMB

Universe is expanding and Density fluctuations evolve into structures we observe: galaxies, clusters, super-clusters.

Previous class: Structure formation and evolution

Overdensities will not growth their radius infinitely, they growth linearly, but because as their mass is also growing, at some point (density threshold) this will re-collapse to form a dark matter halo (spherical collapse model).



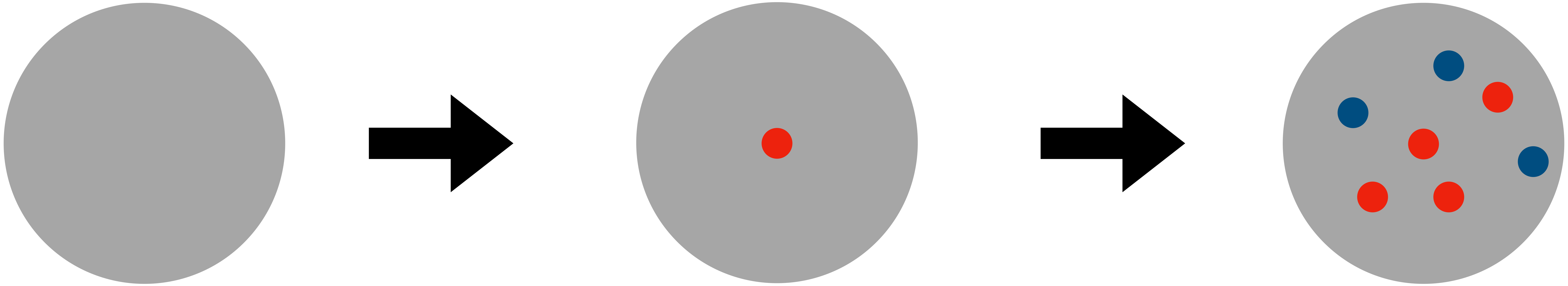
Final result: A dark matter halo in equilibrium.

If the density contrast δ is greater than a certain threshold over a certain volume, then the overdensity will start to collapse (local gravity strongly dominate over the universe expansion)

Dark matter halos are collapsed overdensities in the dark matter distribution with density ~200 times the men density in the universe, inside of which all mass is gravitationally bound.

Previous class: Structure formation and evolution

- Galaxies form in the center of the larger dark matter halos.
- To understand how galaxies form we need to understand galaxy formation models (not covered in this course).



- Almost all the dark matter halos, contain a galaxy in their center (central galaxy).
- Dark matter halos can also contain more than one galaxy (satellite galaxies).
- The most massive dark matter halos can contain a lot of satellite galaxies (for example cluster of galaxies).

Previous class: Structure formation and evolution

Why is important to trace and quantify the LSS?

Visually we can see and detect structure, but we need a mathematical formalism to quantify the density fluctuations and the level in which the matter is grouped.

Details of LSS depends on:

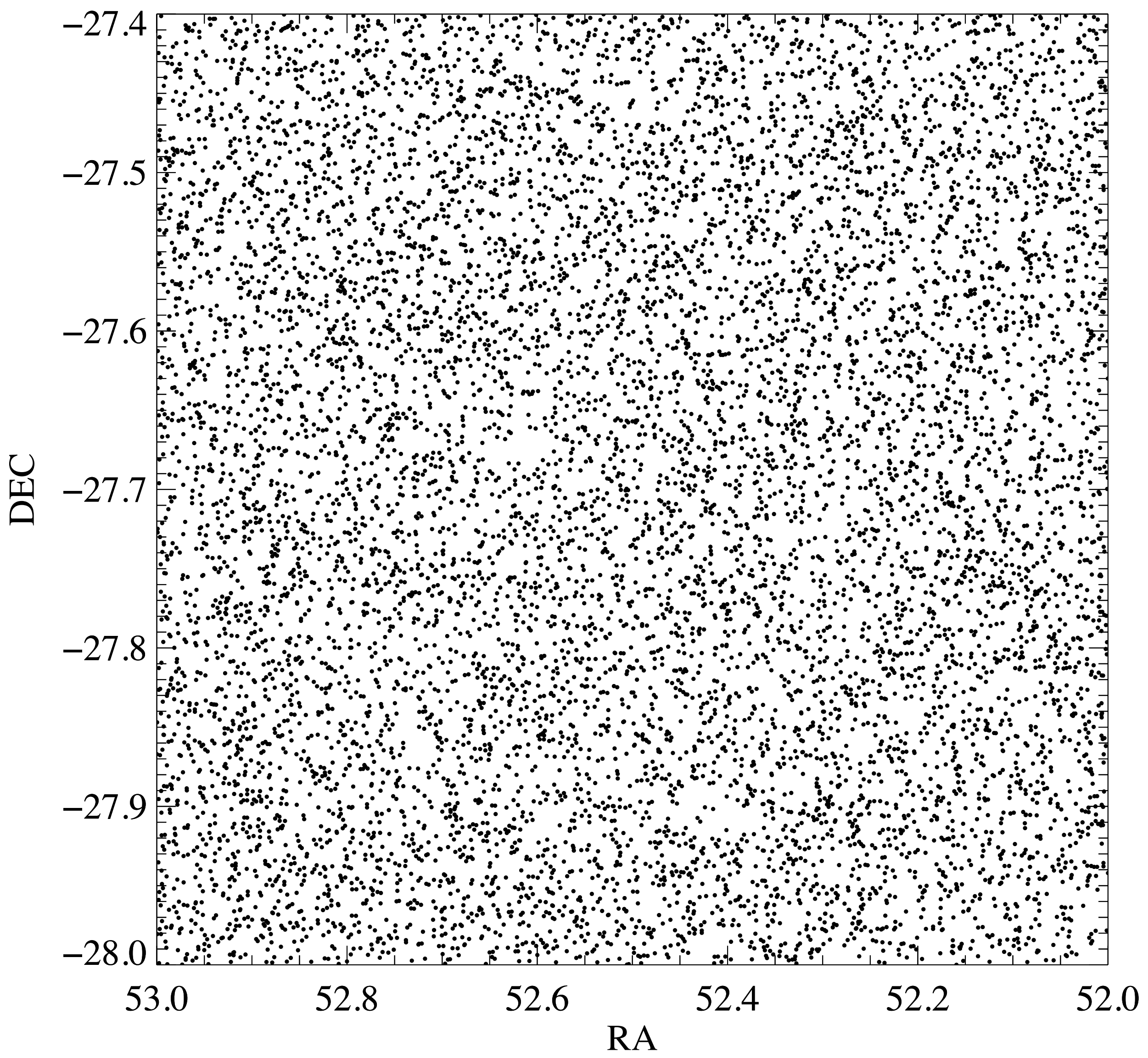
- Initial conditions (characteristics of the initial density field): CMB
- Cosmological parameters (matter density at each epoch, dark energy, etc).
- Formation and evolution of structure.
- Physical processes involved in the growth and evolution of individual galaxies.

Measurements of the large scale structure in combination of theoretical models allow us to constrain both cosmology and physics of galaxy evolution.

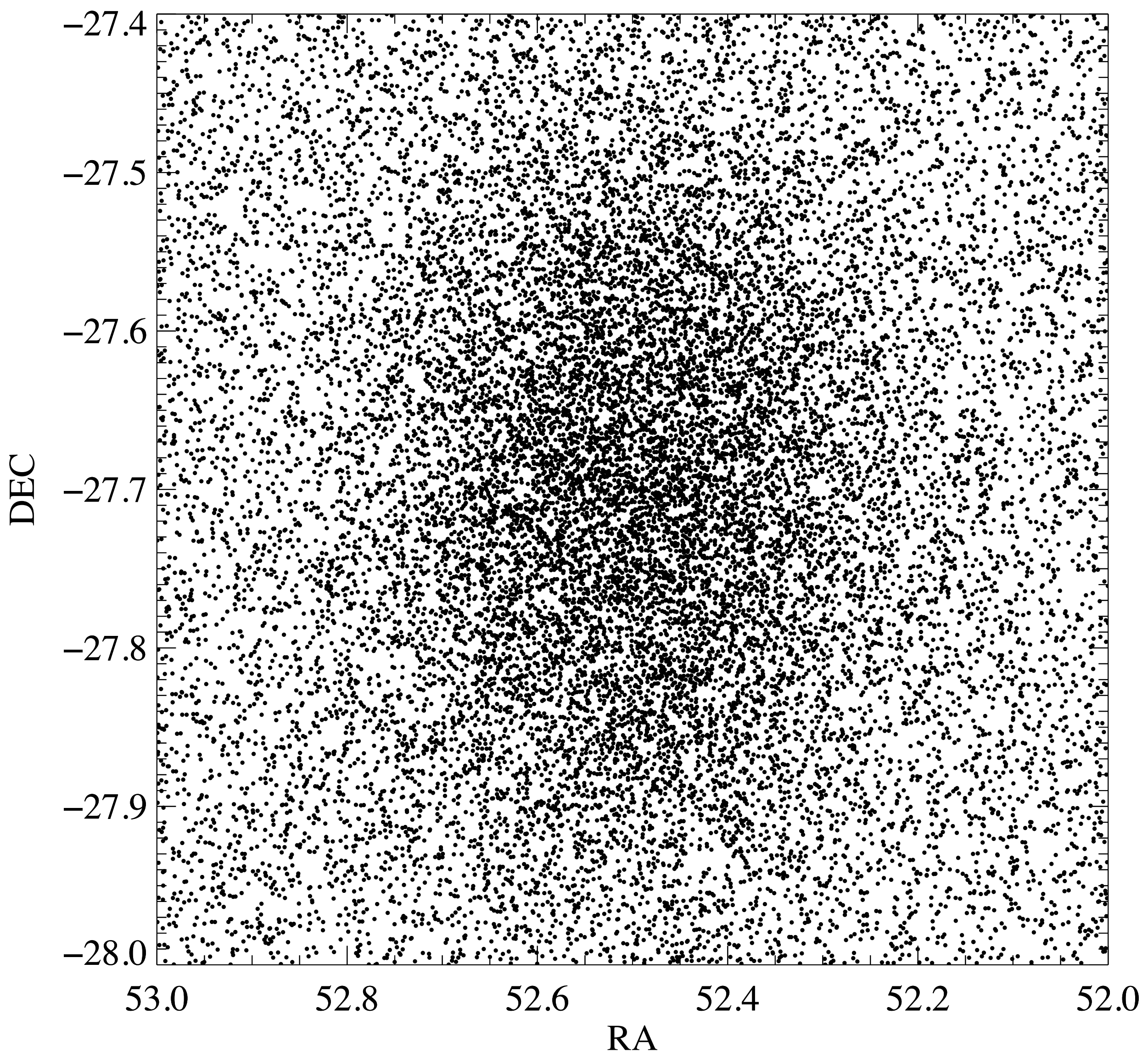
Statistical characterization of the LSS are needed to test models of structure formation and evolution, and models are needed to interpret LSS observations.

Clustering basics

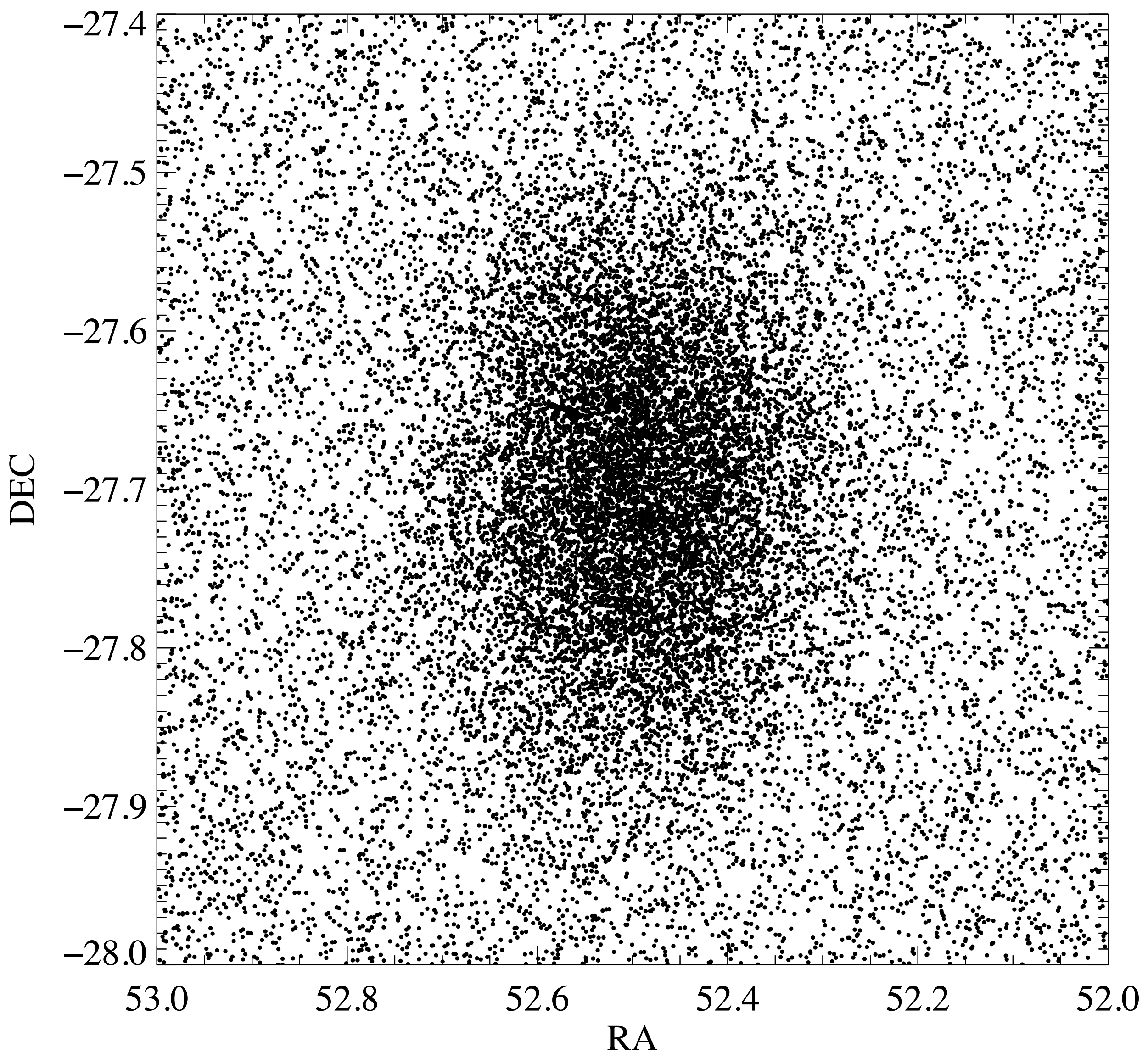
Clustering quantify how grouped the objects are: the more grouped objects are, the stronger the clustering is.



Objects randomly distributed
They are not clustered



Objects are clustered to each other at small scales



Objects are strongly clustered to each other at small scales

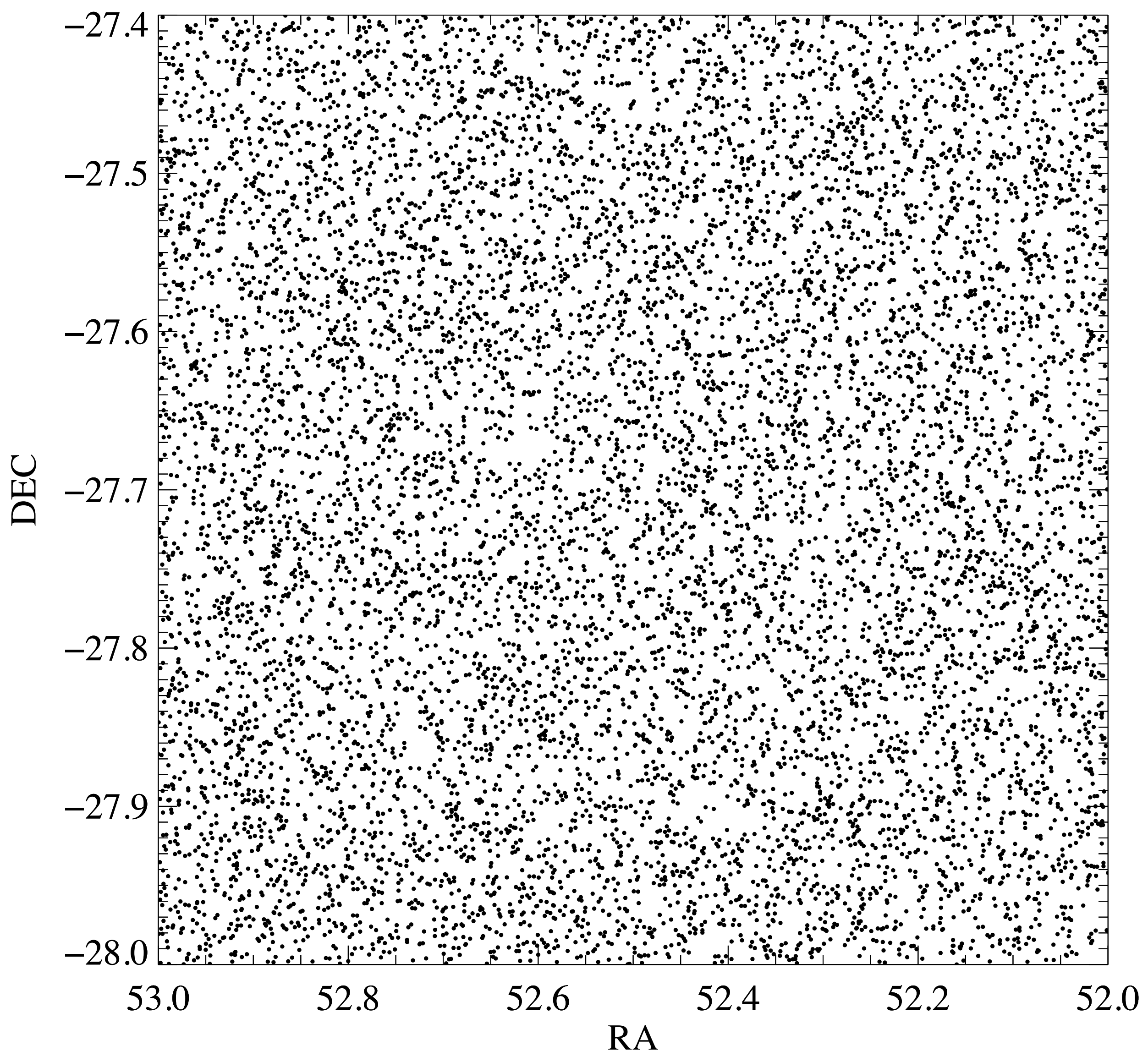
The mathematical formalism to describe the level of clustering is the two-point correlation function

Two-point correlation function

The two-point correlation function ξ(r) is defined as a measure of the excess probability dP, over a random occurrence of finding a galaxy in a volume element dV at a separation r from another galaxy.

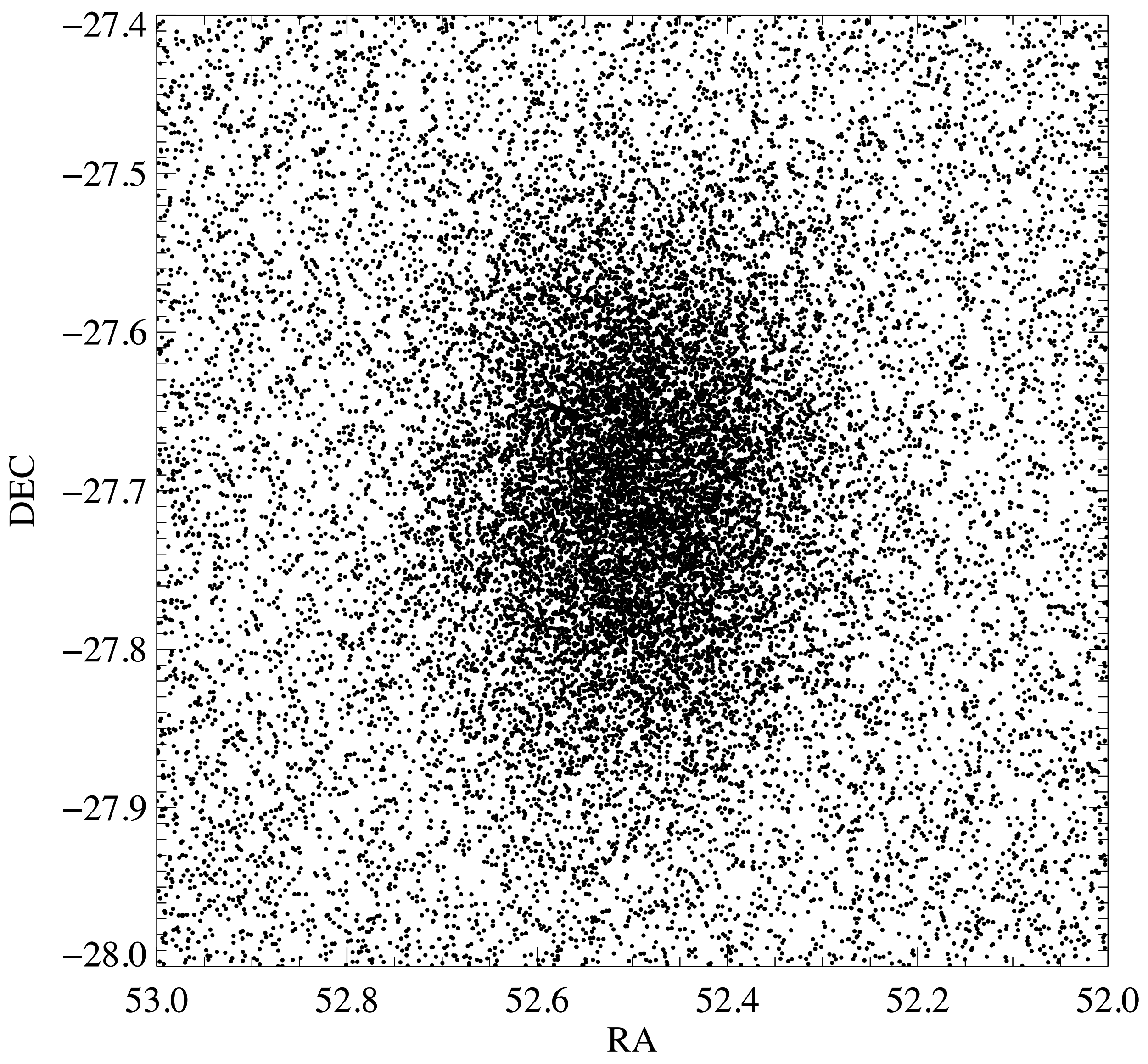
$$\boxed{dP = \bar{n}[1 + \xi(r)]\mathrm{dV}}$$

where n is the mean number density of the galaxy sample in question.



In a random distribution, the probability to find a galaxy in one place or another, is independent. Their positions are not correlated.

$$dP = \bar{n}\mathrm{dV}$$



In a strongly clustered population, if you find a galaxy it is highly probable that you find another galaxy close to it.

$$dP = \bar{n}[1 + \xi(r)]\mathrm{dV}$$

- The two-point correlation function traces the amplitude of clustering as a function of physical scale (clustering depends on scale!)

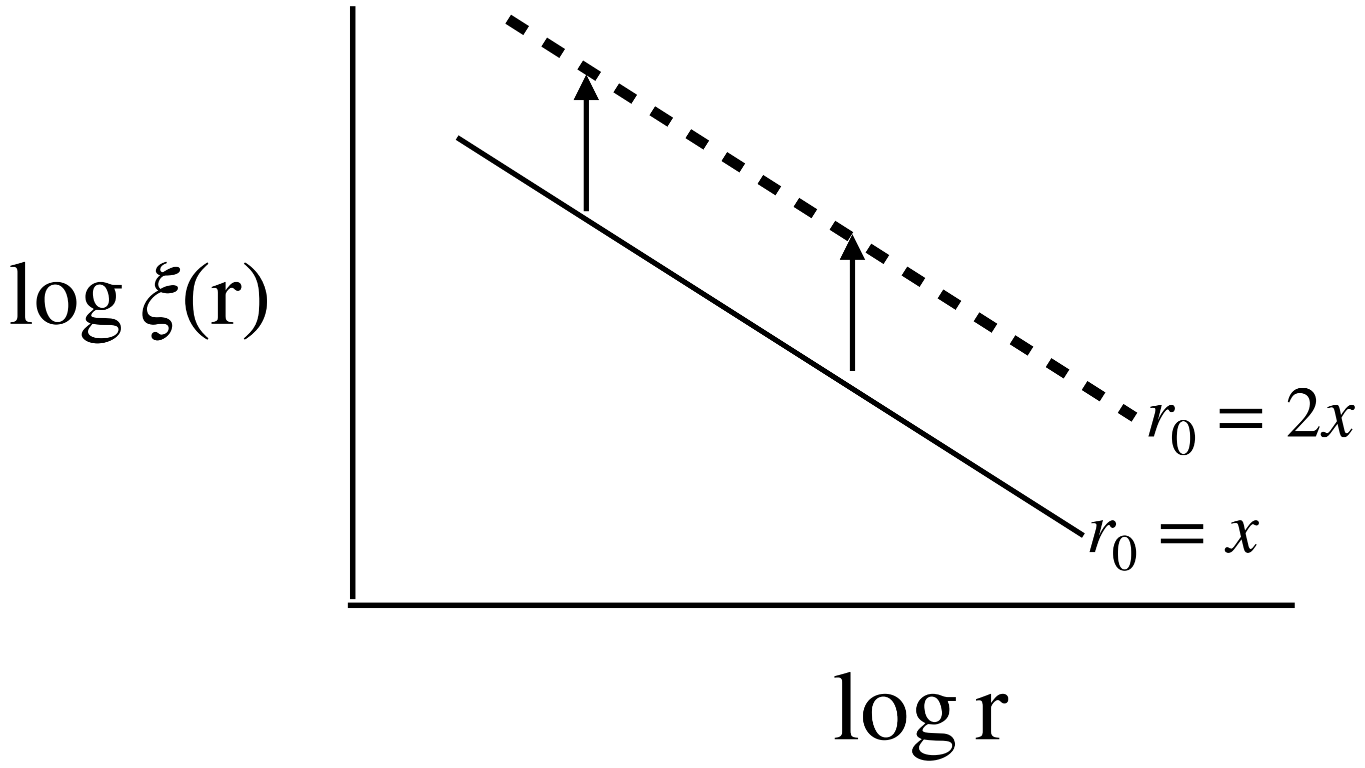
Two-point correlation function

Observations indicate that ξ(r) is well described by a power-law: $\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}$

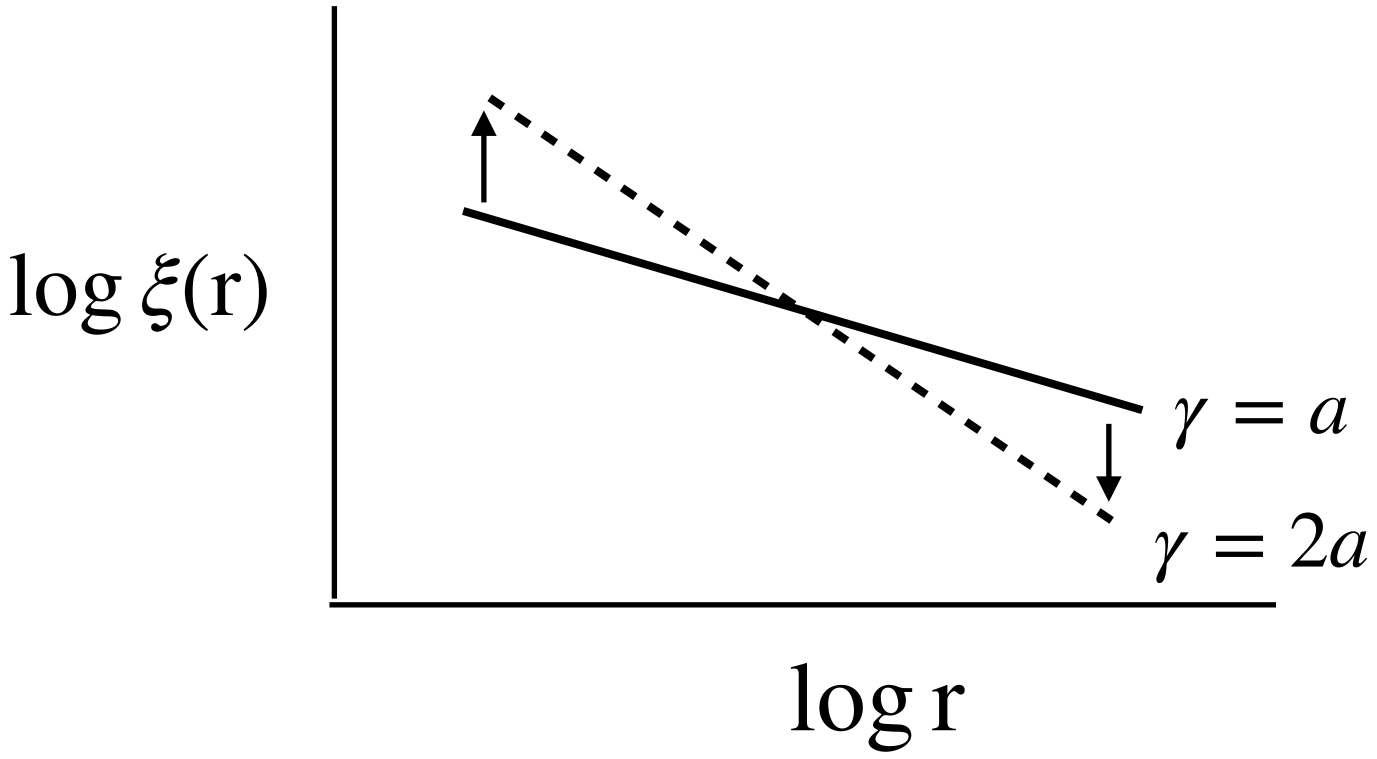
r_0 correlation length
γ slope (typically 1.8)

→ Strong clustering at small scales and weak clustering at large scales.

Effect of the correlation length:



Effect of the slope:



Higher clustering implies a higher ξ(r) and therefore a higher r_0

How is ξ(r) measured?

Estimators:

$$\xi = \frac{n_R^2}{n_D^2}\frac{DD}{RR} - 1$$

Natural estimator

One counts pairs of galaxies as a function of separation and divides by what is expected for an unclustered distribution (random distribution of points). The construction of a so-called “random catalog” is then required.

DD (data-data): number of pairs of galaxies in bins of separation.

RR (random-random): number of pairs of random points in bins of separation.

n_d and n_R are the number density of galaxies and random points in the catalogs.

Requirements for the random catalog:

- Sources need to be randomly distributed over the sky.
- Same selection function as the data (in 3D).
- Large enough such that it don’t introduce poisson errors in the estimation (RR and DD scale with the square of number of pairs).

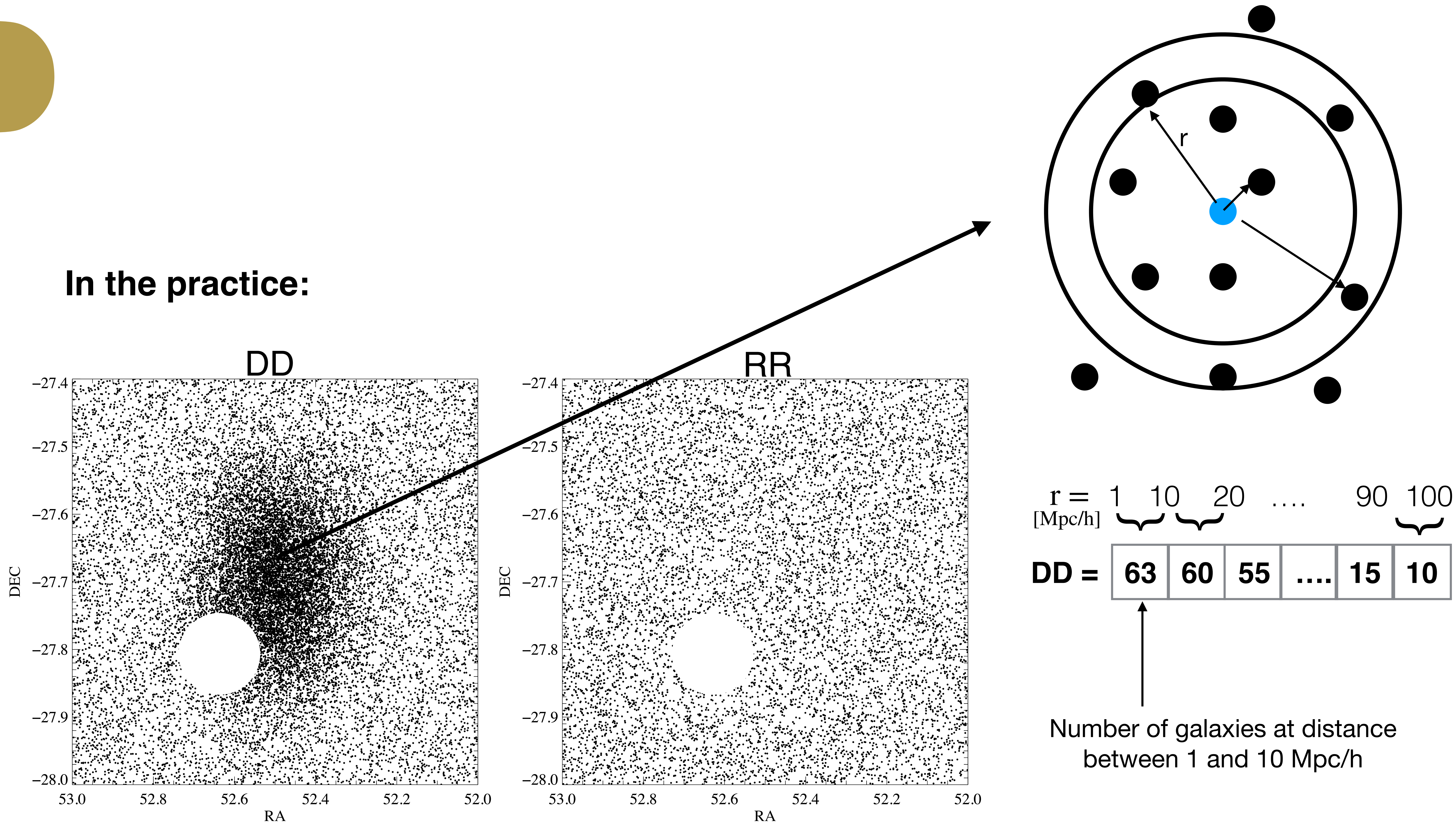
How is ξ(r) measured?

Estimators:

$$\xi = \frac{n_R^2}{n_D^2}\frac{DD}{RR} - 1$$

Natural estimator

In the practice:



r = 1 10 20 .... 90 100
[Mpc/h]

DD =	63	60	55	....	15	10

Number of galaxies at distance between 1 and 10 Mpc/h

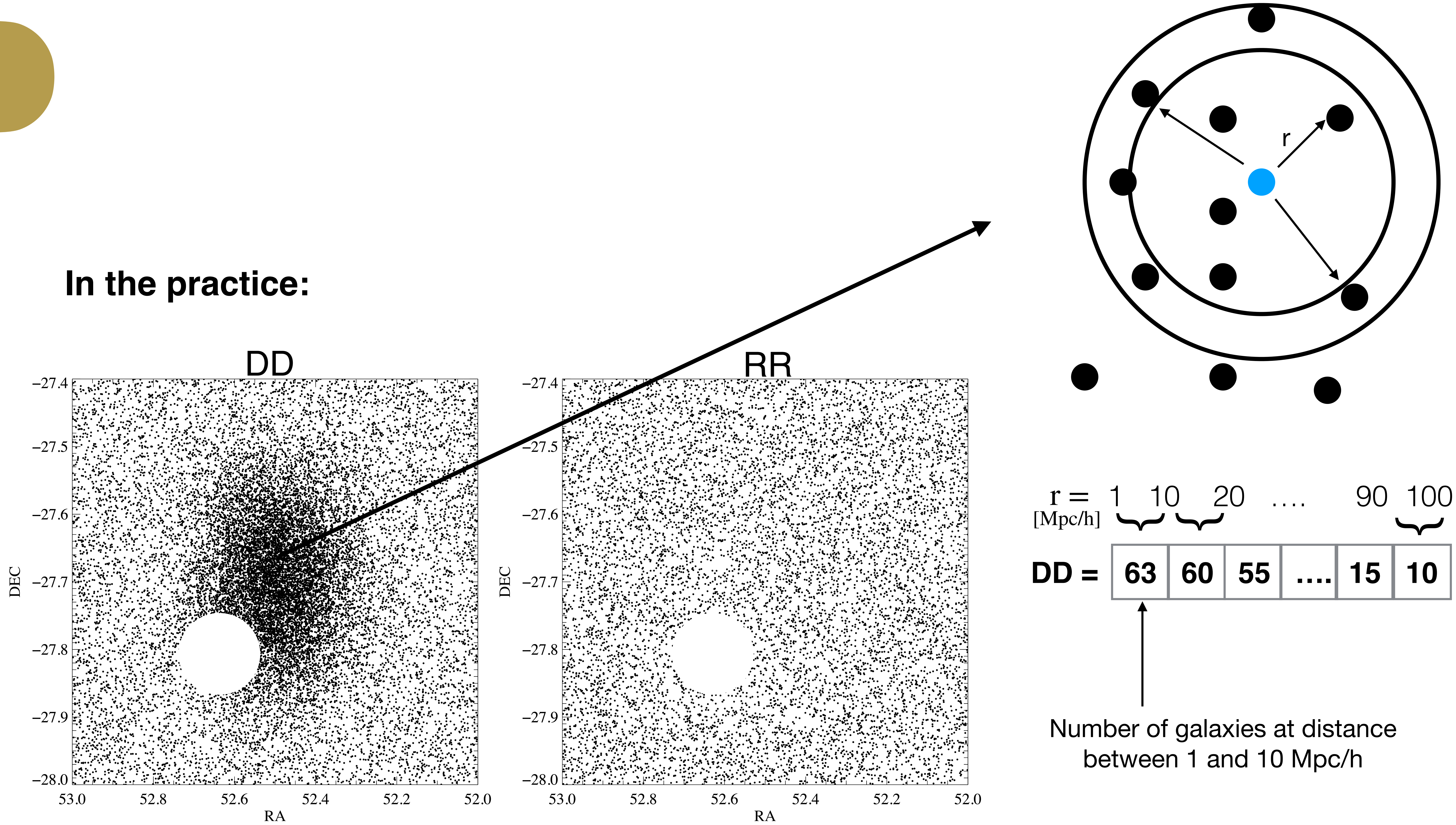
How is ξ(r) measured?

Estimators:

$$\xi = \frac{n_R^2}{n_D^2}\frac{DD}{RR} - 1$$

Natural estimator

In the practice:



$r =$ 1 10 20 .... 90 100
[Mpc/h]

DD =	**63**	**60**	**55**	**....**	**15**	**10**

Number of galaxies at distance between 1 and 10 Mpc/h

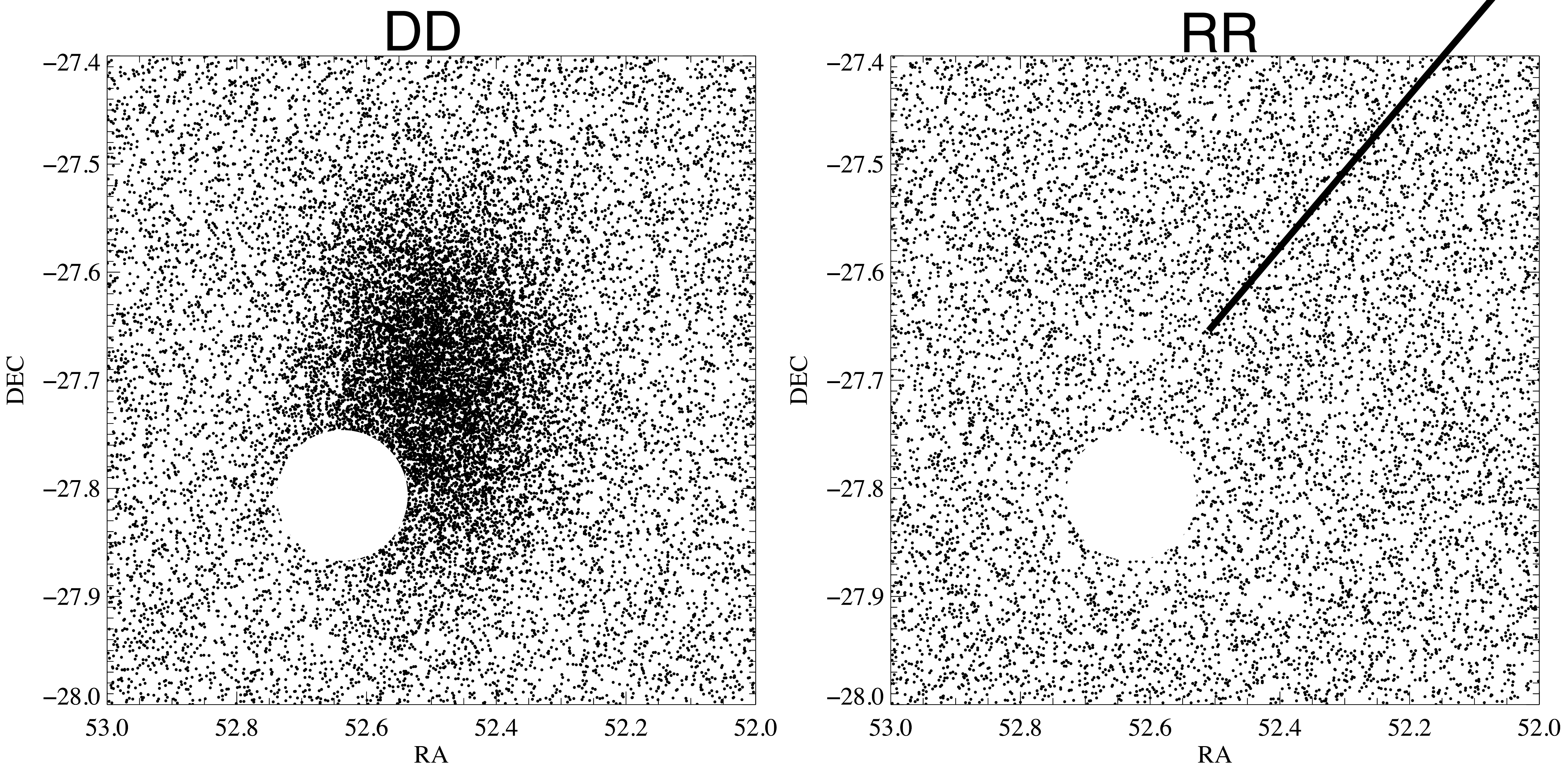
How is ξ(r) measured?

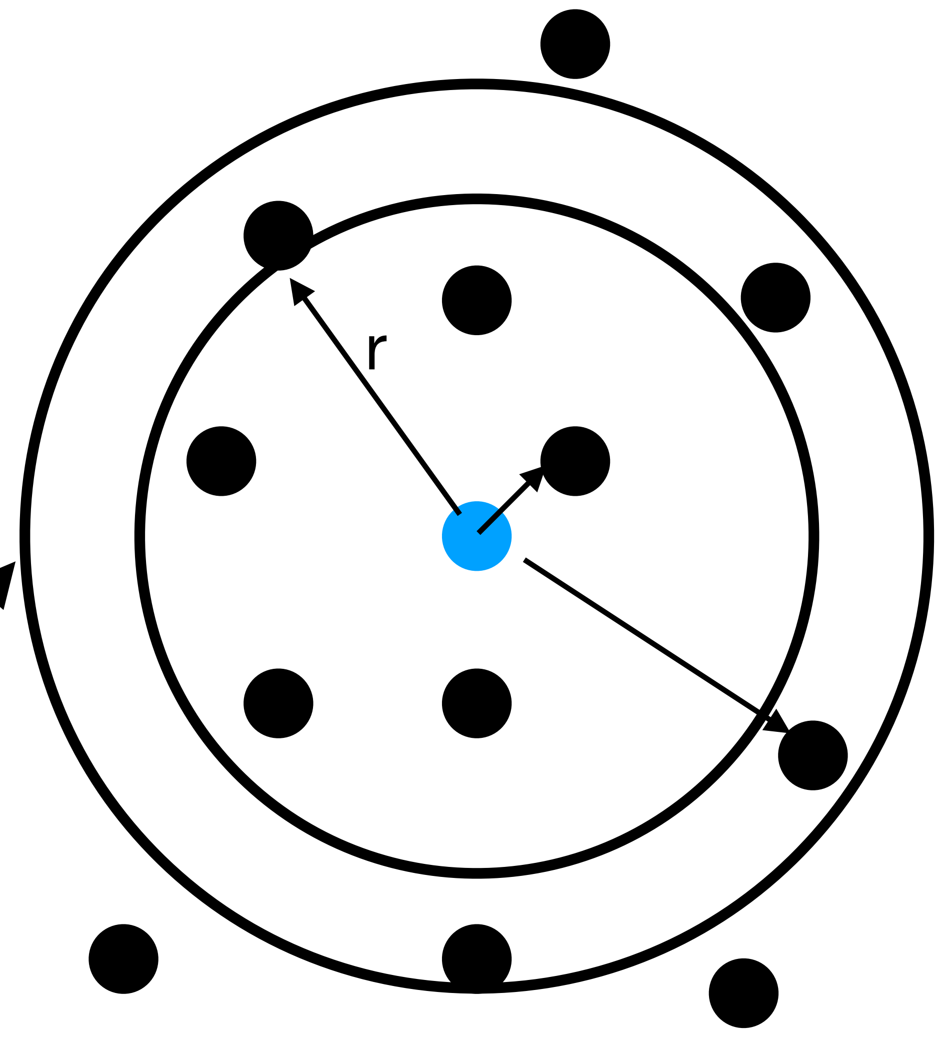
Estimators:

$$\boxed{\xi = \frac{n_R^2}{n_D^2}\frac{DD}{RR} - 1}$$

Natural estimator

In the practice:





r = 1 10 20 .... 90 100
[Mpc/h]

DD =	**63**	**60**	**55**	**....**	**15**	**10**

r = 1 10 20 .... 90 100
[Mpc/h]

RR =	**12**	**9**	**15**	**....**	**10**	**13**

How is ξ(r) measured?

One counts pairs of galaxies as a function of separation and divides by what is expected for an unclustered distribution (random distribution of points). The construction of a so-called “random catalog” is then required.

Estimators:

$$\xi = \frac{n_R^2}{n_D^2}\frac{DD}{RR} - 1$$

Natural estimator

$$\xi = \frac{n_R}{n_D}\frac{DD}{DR} - 1,$$

Davis & Peebles (1983)

$$\xi = \frac{DD\,RR}{(DR)^2} - 1$$

Hamilton (1993)

$$\xi = \frac{1}{RR}\left[DD\left(\frac{n_R}{n_D}\right)^2 - 2DR\left(\frac{n_R}{n_D}\right) + RR\right]$$

Landy & Szalay (1993)

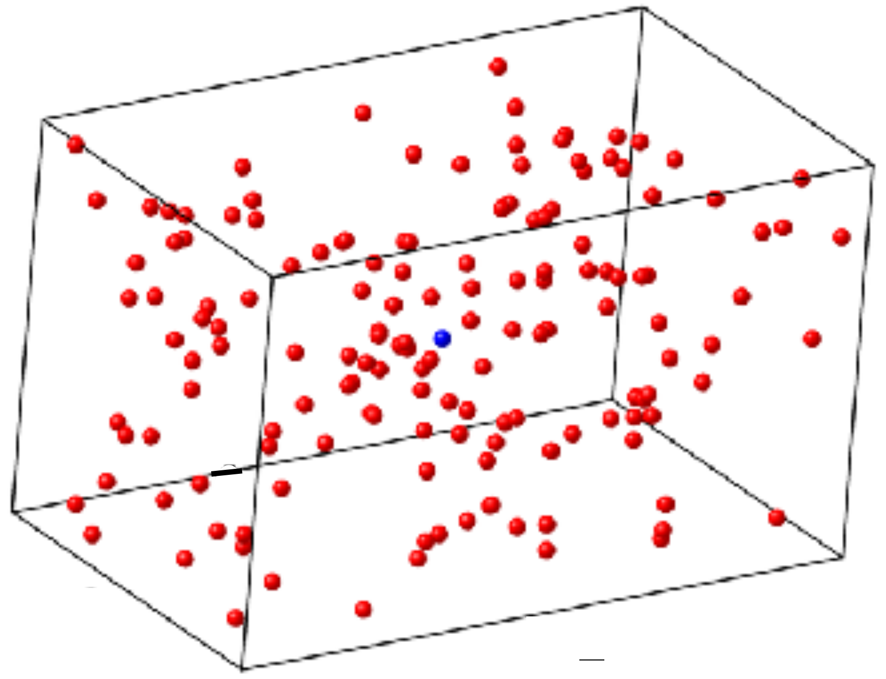
→ This is normally the preferred choice

DD (data-data): number of pairs of galaxies in bins of separation.
RR (random-random): number of pairs of random points in bins of separation.
n_d and n_R are the number density of galaxies and random points in the catalogs.

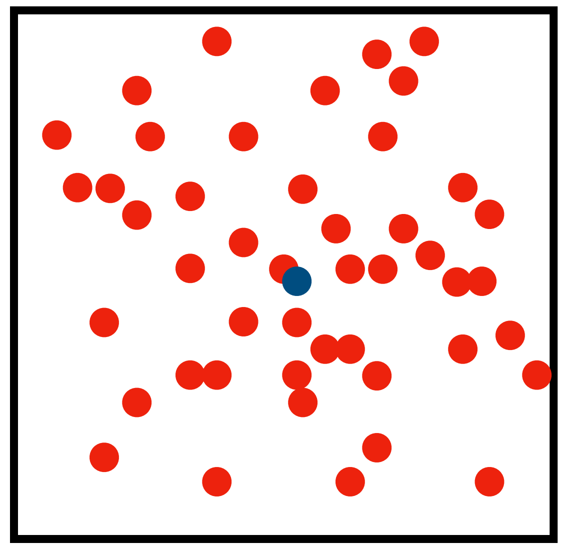
How can we do this in the practice?

- The exact distance r between objects is never possible to measure accurately (even if we have z information).
- Depending on the data that we have, we will be able to measure the:

Projected correlation function: if 3D information is available.



Projected angular correlation function: if only 2D information is available.



Projected angular correlation function

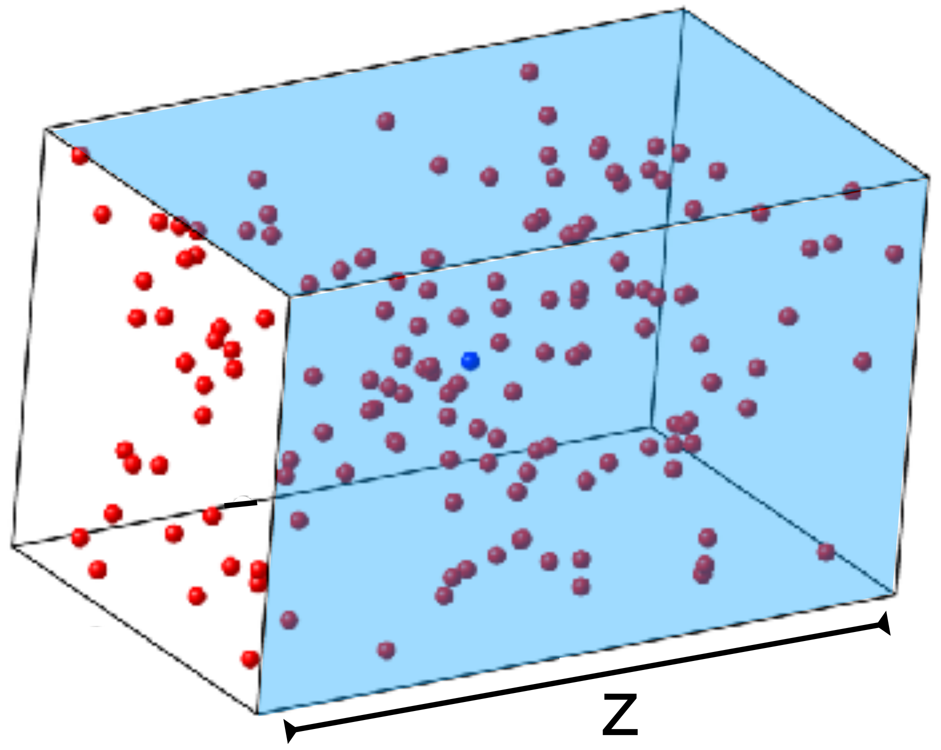
- We have RA and Dec for all the sources, but not z information.
- We can only measure θ (angular distance), and then we only can measure the angular correlation function ω(θ):

$$dP = n[1 + \omega(\theta)]d\Omega$$

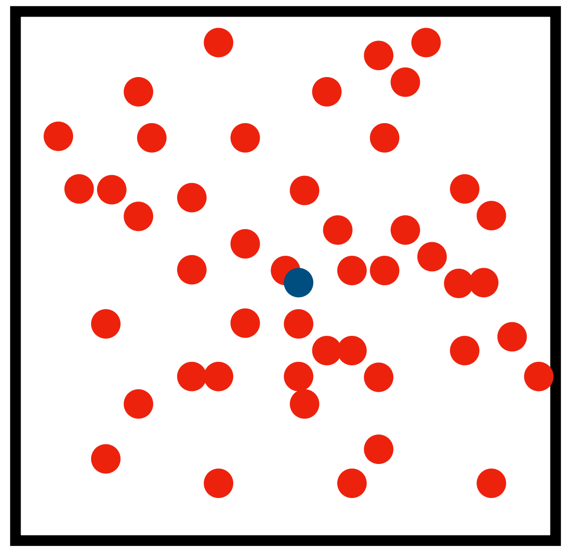
- The angular correlation function is normally well represented by a power law:

$$\omega(\theta) = A\theta^{\beta}$$

with A the clustering amplitude and β the slope



We see all the galaxies in the same plane, we see the information collapsed over the line-of-sight



Projected angular correlation function

- If we know the redshift distribution of the sources, then we can infer ξ(r) from ω(θ).

- The relation between ξ(r) and ω(θ) can be obtained integrating ξ(r) over the line-of-sight (Limber equation):

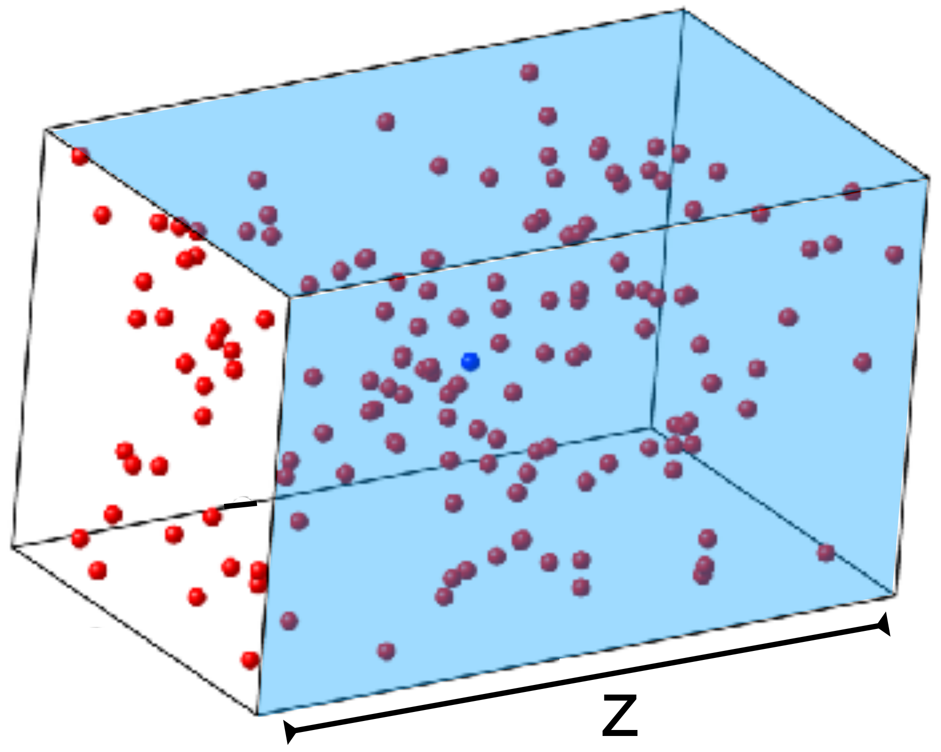
$$A = \frac{\int_0^\infty r_0^\gamma(z) g(z) \boxed{\left(\frac{dn}{dz}\right)^2} dz}{\left[\int_0^\infty \boxed{\left(\frac{dn}{dz}\right)^2} dz\right]^2}$$

with $g(z) = \left(\frac{dz}{dr}\right) r^{(1-\gamma)} F(r)$

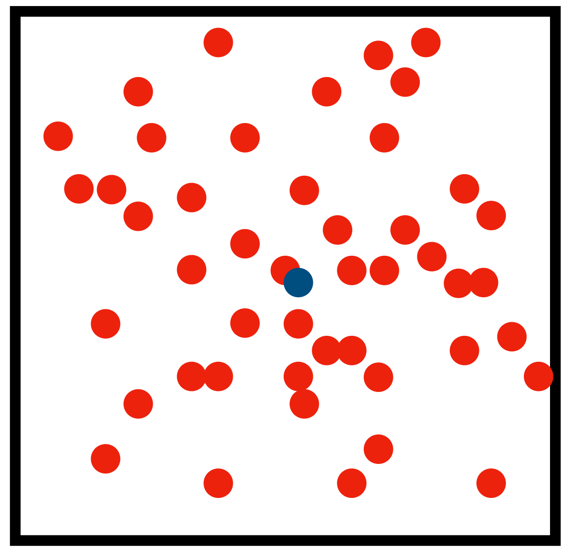
This depends on the cosmological model

Redshift distribution of the sample (number of objects as a function of redshift)

$$\beta = 1 - \gamma$$



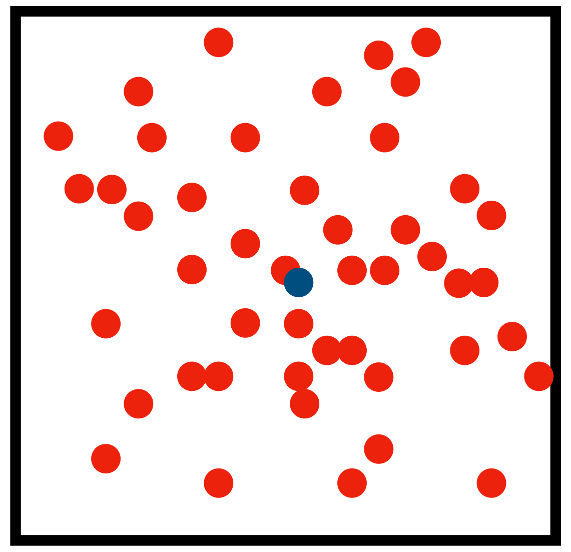
We see all the galaxies in the same plane, we see the information collapsed over the line-of-sight

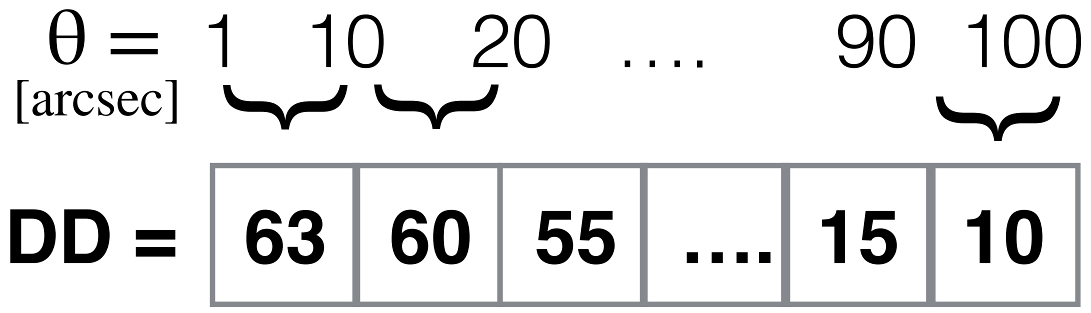


Projected angular correlation function

Cookbook:

1)





$\theta =$ 1 10 20 .... 90 100

[arcsec]

DD =

63	**60**	**55**	**....**	**15**	**10**

2) Repeat for a random catalog and obtain RR.

3) Use an estimator to compute $\omega(\theta)$.

4) Fit the measurement with a power law to obtain the parameters A and beta.

5) Assume a redshift distribution and use the Limber equation to obtain r_0, γ.

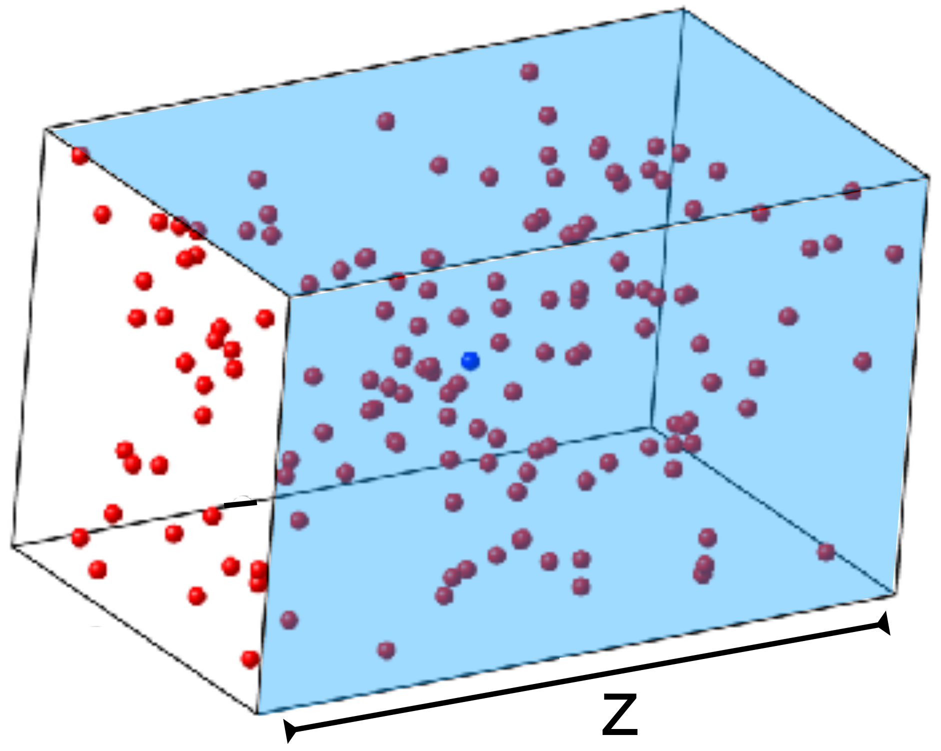
Projected angular correlation function

- Two important things about the angular correlation function:

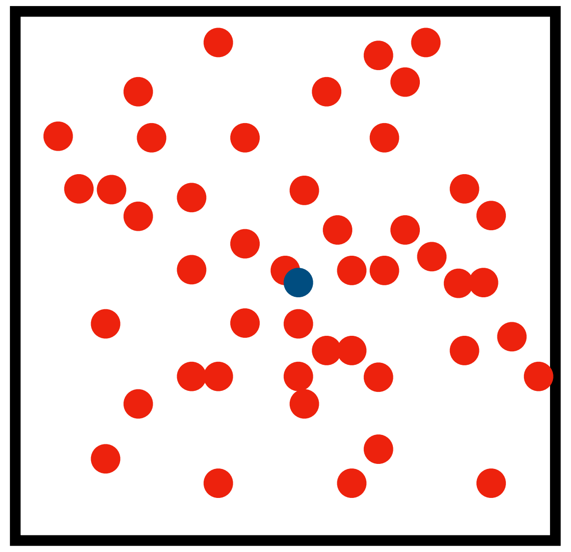
1) A can be dominated by errors given the lack of knowledge of the redshift distribution of the sample.

2) We are collapsing the information over large volumes, then the clustering signal can be diluted (projection effects over the line of sight).

Projection effects: Even if a population is strongly clustered in 3D, when we integrate over a long line-of-sight the signal may be washed out and the angular correlation function would be weak.



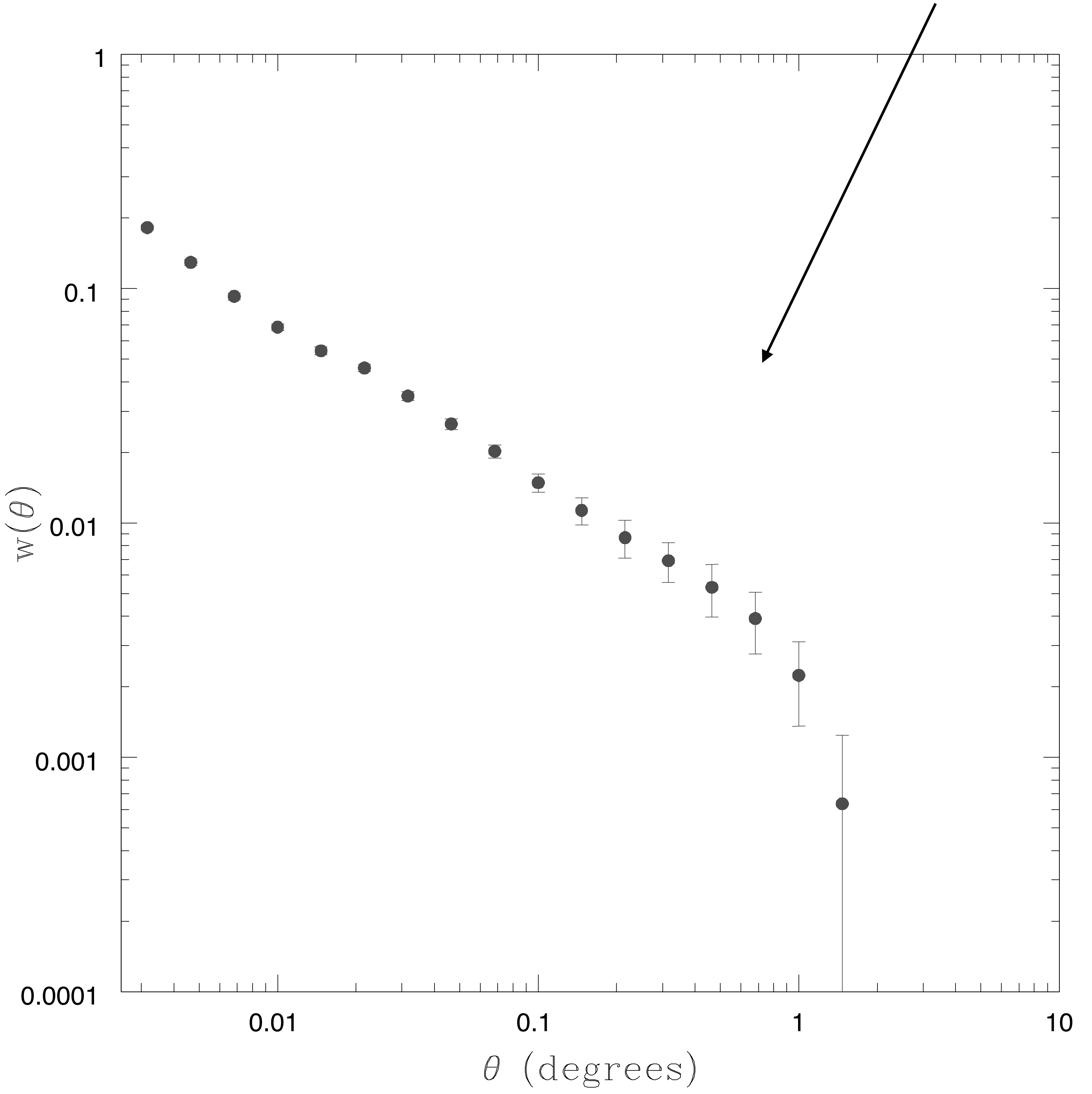
We see all the galaxies in the same plane, we see the information collapsed over the line-of-sight



Projected angular correlation function

One example: Angular two-point correlation function of galaxies from the SDSS.

Note that we usually choose bin logarithmically spaced bins

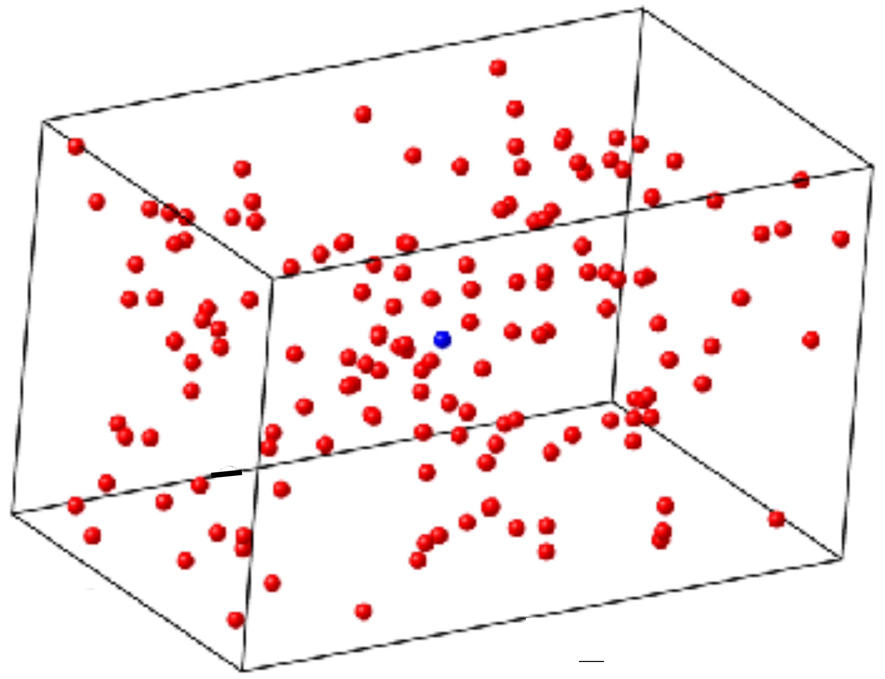


(Connolly et al. 2002)

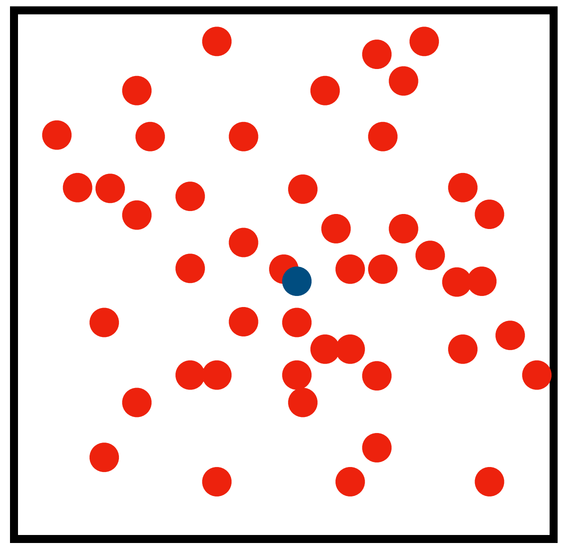
How can we do this in the practice?

- The exact distance r between objects is never possible to measure accurately (even if we have z information).
- Depending on the data that we have, we will be able to measure the:

Projected correlation function: if 3D information is available.



Projected angular correlation function: if only 2D information is available.



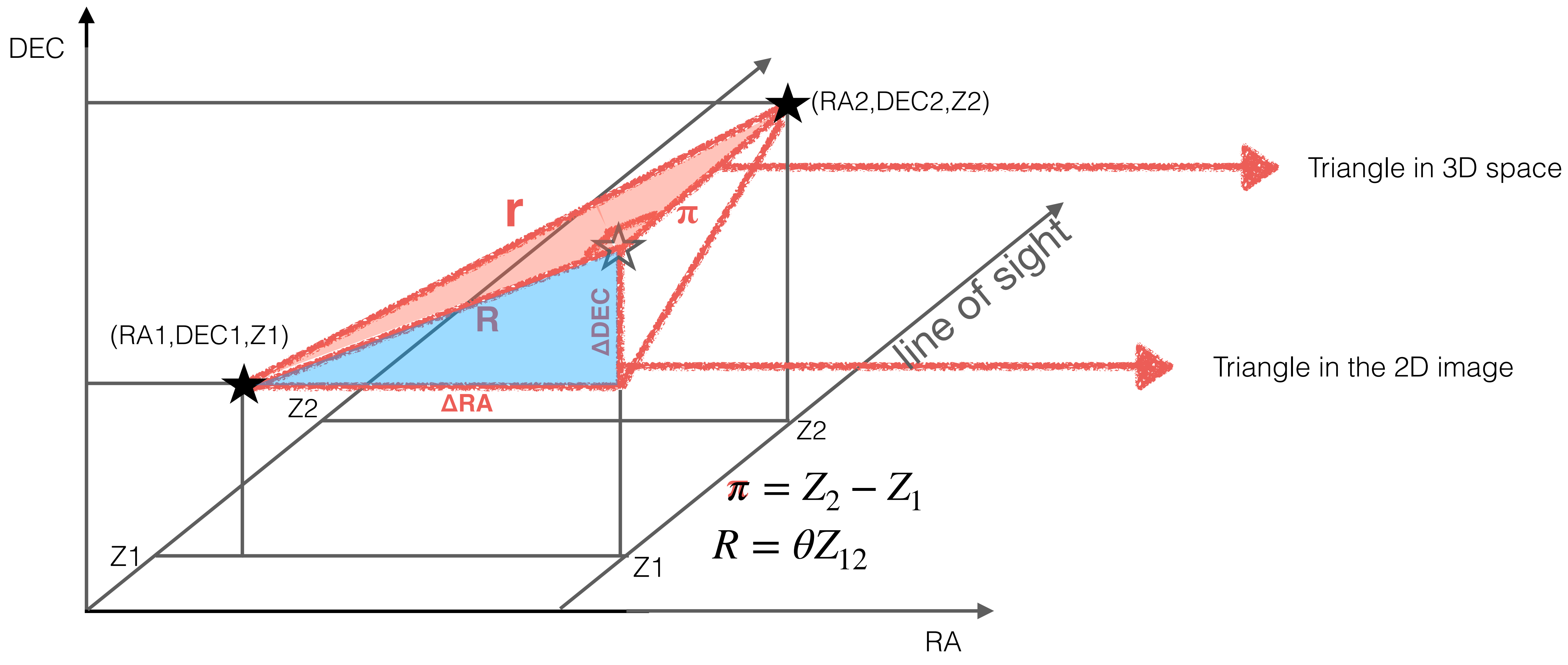
Projected correlation function

- We have ra, dec, z for all the sources.
- z is the redshift, and can be converted into a comoving distance in redshift space (Z [Mpc/h]) but:

1) there is a dependence of the used cosmology (z ⟶ Z).

2) it is affected by peculiar velocities.

Then we can never measure the distance r.



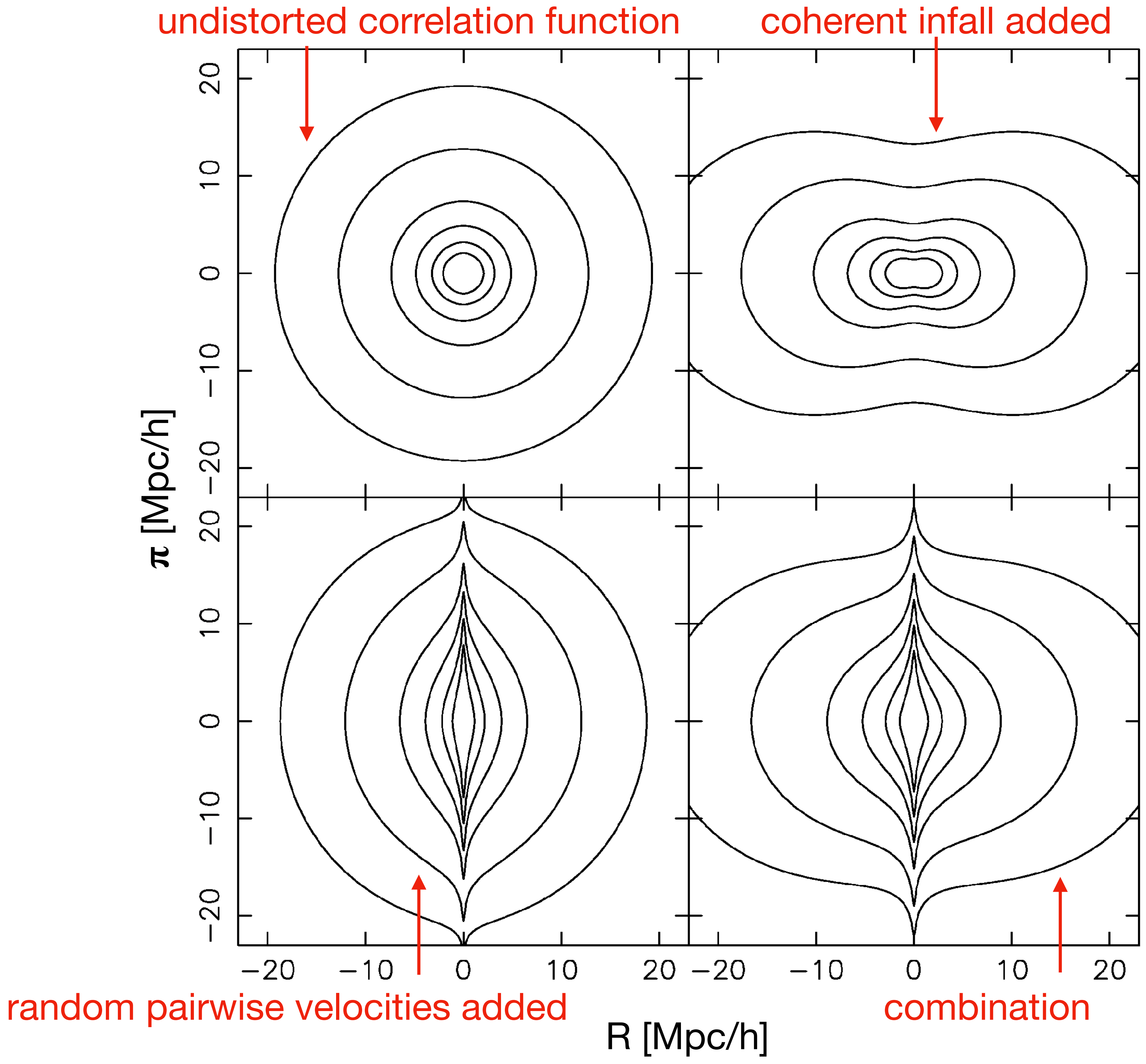
distance between objects expressed in two components: perpendicular (R) and parallel (**π**) to the line-of-sight

$$r^2 = R^2 + \pi^2$$

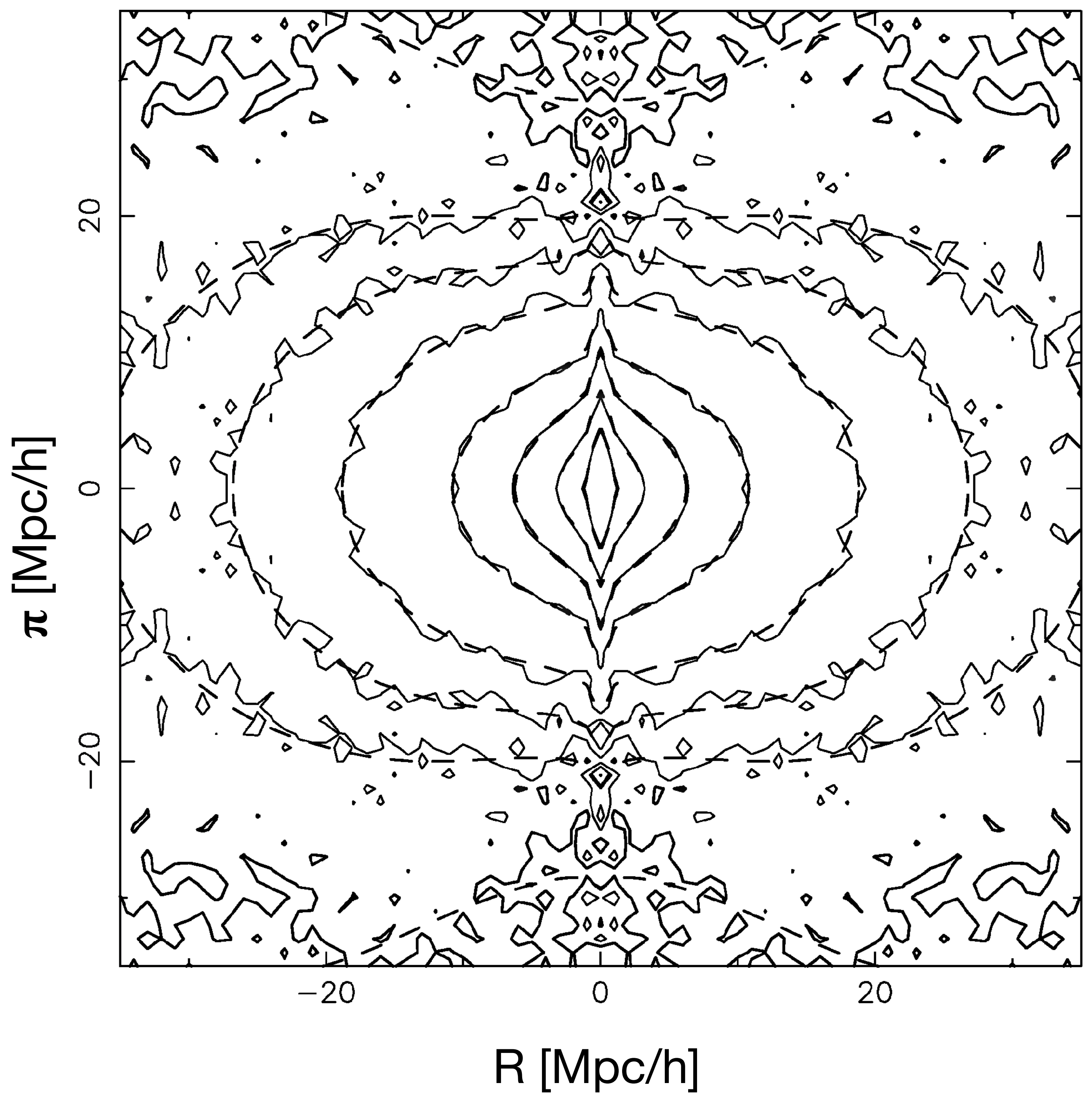
Projected correlation function

Redshift space distortions

Modeled contours of $\xi(R,\pi)$ (with same r0, gamma) with different added distortions.

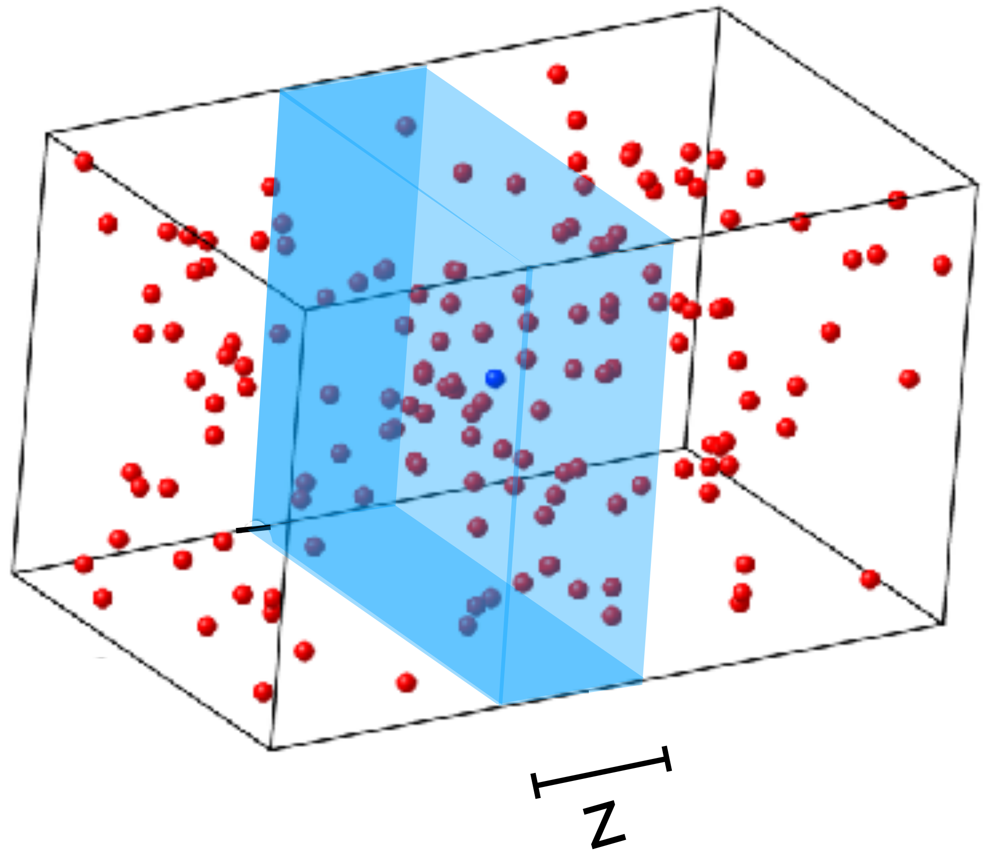


Observed contours of $\xi(R,\pi)$ from the 2dF data



We need to express the 3-dimensional two-point correlation function $\xi(r)$, as a 2-dimensional two-point correlation function $\xi(R,\pi)$

Projected correlation function

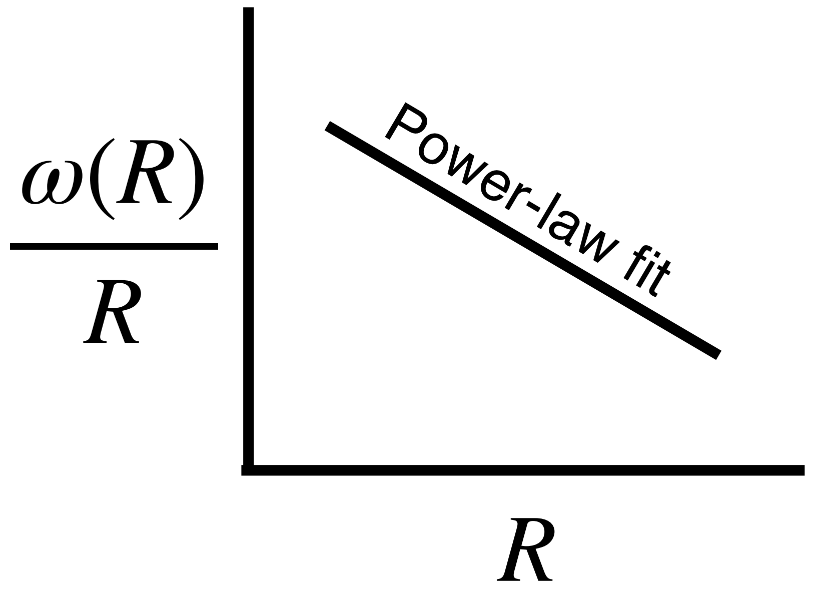


Real 3-d correlation function

Projected correlation function: $$\omega(R) = \int_{-\infty}^{\infty} \xi(R,\pi)\,d\pi$$

If we assume a power-law: $$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma} \longrightarrow \xi(R,\pi) = \left(\frac{\sqrt{R^2+\pi^2}}{r_0}\right)^{-\gamma}$$

Projected correlation function: $$\omega(R) = R\left(\frac{r_0}{R}\right)^{\gamma} \frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\gamma-1}{2}\right)}{\Gamma\left(\frac{\gamma}{2}\right)} \longrightarrow$$



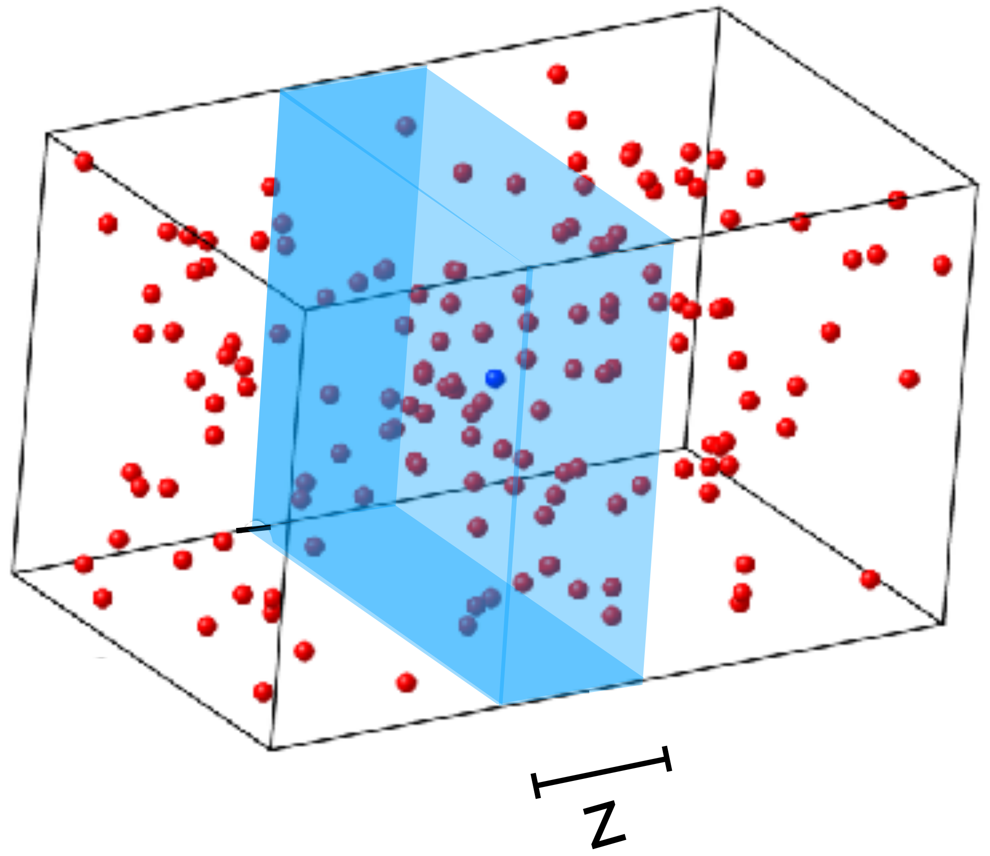
$$\longrightarrow r_0, \gamma$$

- In practice we don't integrate until infinity but until certain number (usually ~100 Mpc/h) for which the contribution is significant.

Projected correlation function

Cookbook:

1)



R = 1 10 20 .... 90 100
[Mpc/h]

[Mpc/h]
π=
1
10
20
....
90
100

DD =

63	**60**	**55**	**....**	**15**	**10**
58	**53**	**49**	**....**	**12**	**7**

Only go until certain number for which the contribution is significant

2) Repeat for a random catalog and obtain RR.

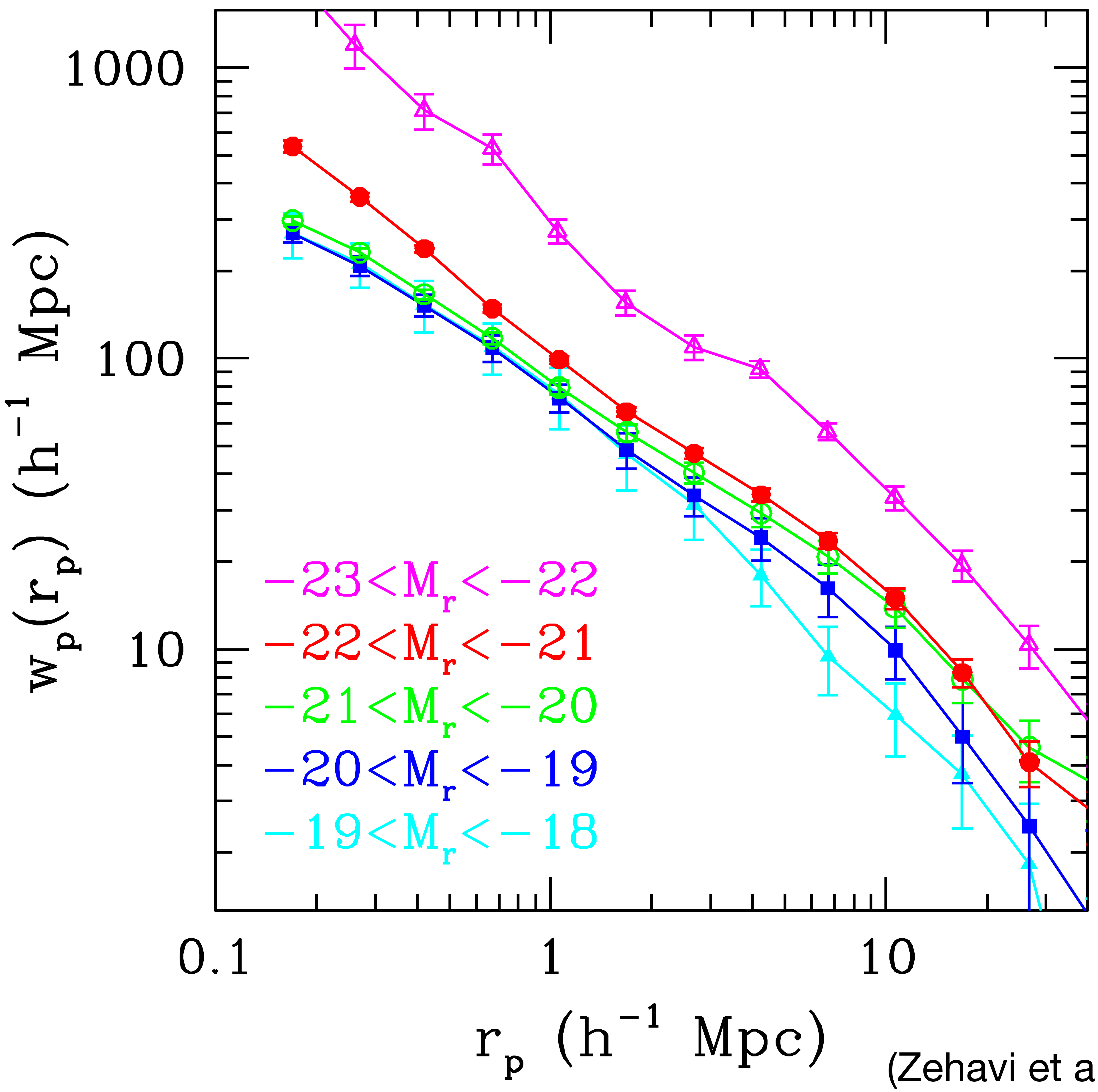
3) Use an estimator to compute $\xi(R,\pi)$.

4) Integrate the values of the grid over the $\boldsymbol{\pi}$ direction to obtain ω(R).

5) Fit the measurement with a power law to obtain the parameters r_0, $\boldsymbol{\gamma}$.

Projected correlation function

One example: Projected two-point correlation function for galaxies in SDSS of different luminosities.



(Zehavi et al. 2011)

Summary

If you have
2D positions

↓

Angular correlation
function $\omega(\theta)$

(Integrated over all the redshift space)

↓

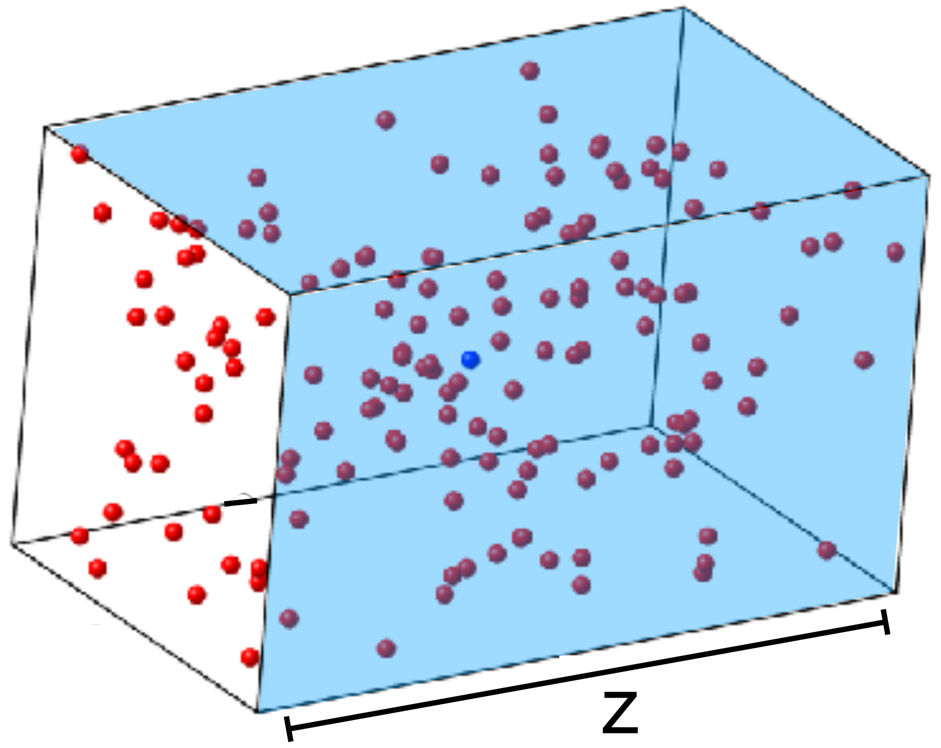
Fit the measurement
to get A and β

↓

Assumptions about
the z distribution

↓

Get r_0, γ



If you have
3D positions

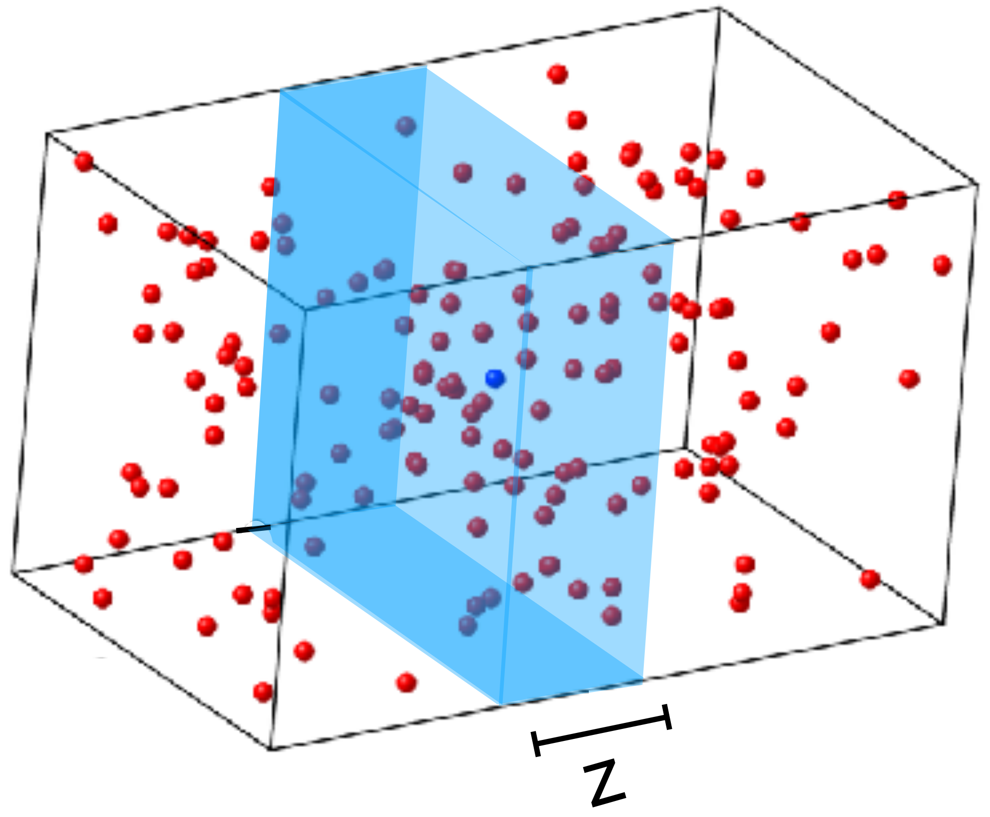
↓

Projected correlation
function $\omega(R)$

(Integrated over a narrow redshift space)

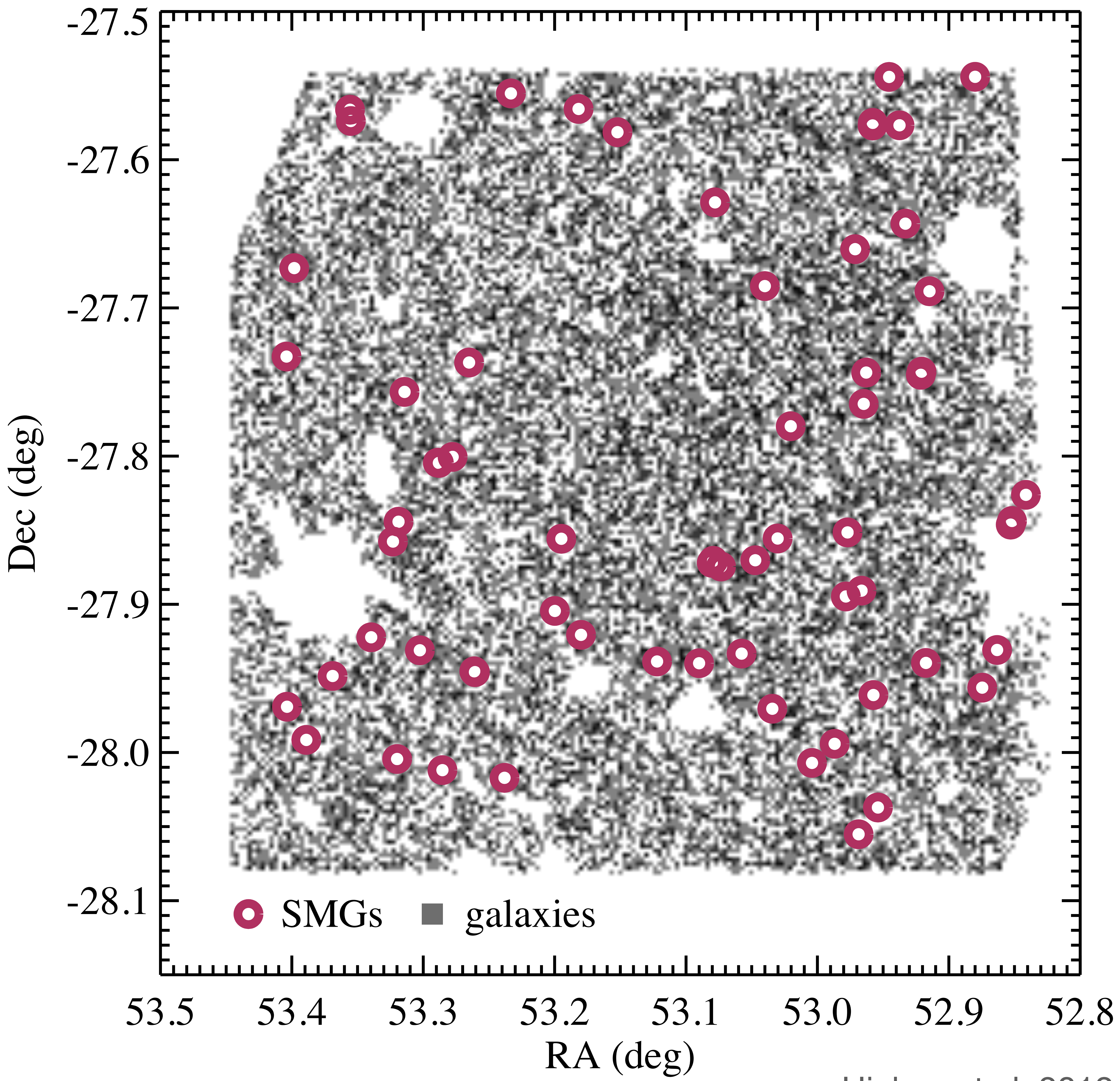
↓

Fit the measurement to
get r_0, γ



Auto-correlation vs Cross-correlation function

Cross-correlation function is when we compute the correlation function between different populations



Hickox et al. 2012

- If we compute the correlation function between galaxies with themselves this is an auto-correlation function.
- If we compute the correlation function between SMGs and galaxies, this is a cross-correlation function.
- For the cross-correlation function everything is the same, but the DD term is now computed using one catalog of one population and the other catalog of the other population:

$$\boxed{\xi = \frac{n_R}{n_D}\frac{DD}{DR} - 1}$$

Auto-correlation

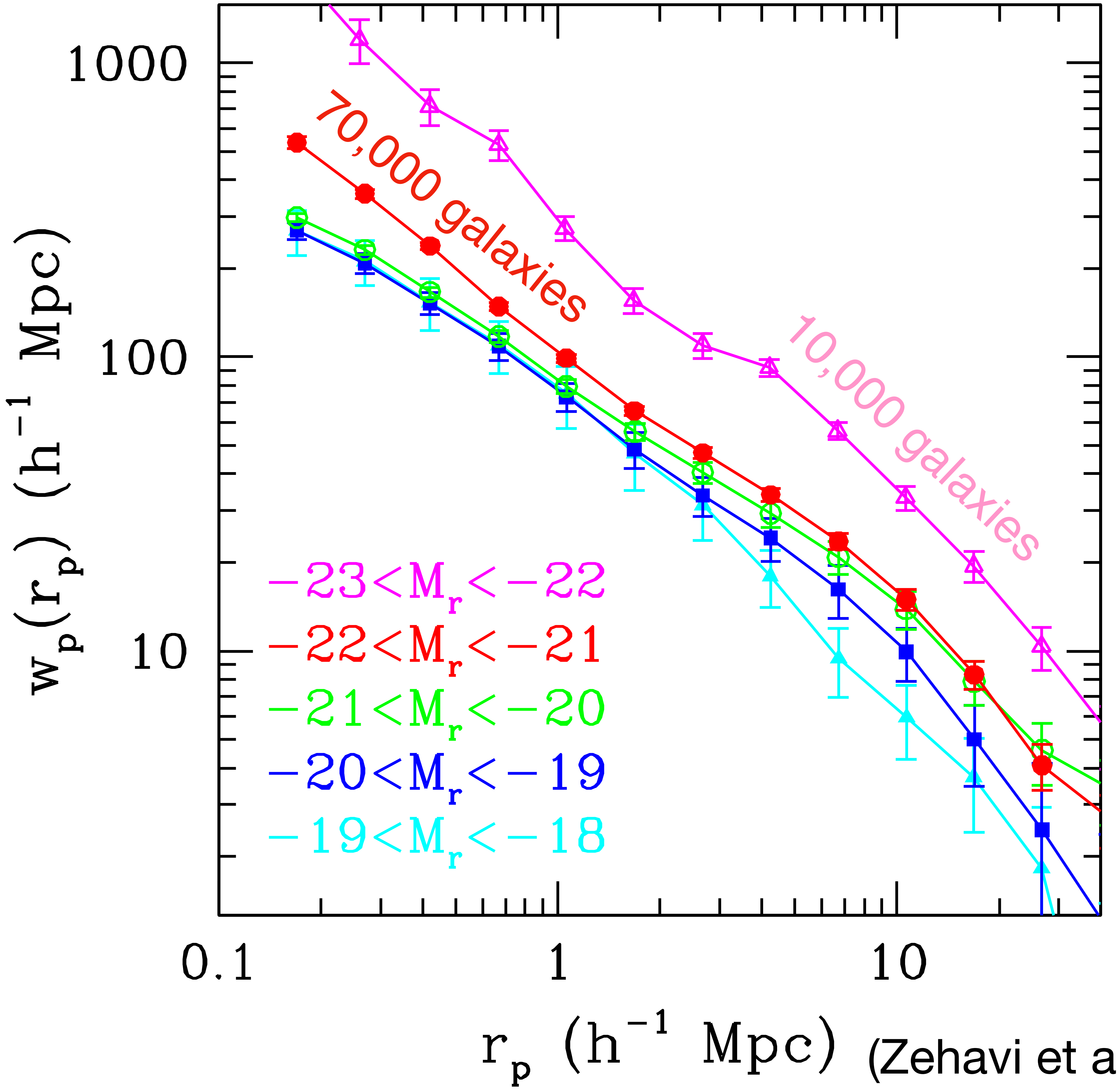
$$\boxed{\xi = \frac{n_{R2}}{n_{D2}}\frac{D_1D_2}{D_1R_2} - 1}$$

Cross-correlation

Uncertainties in clustering measurements

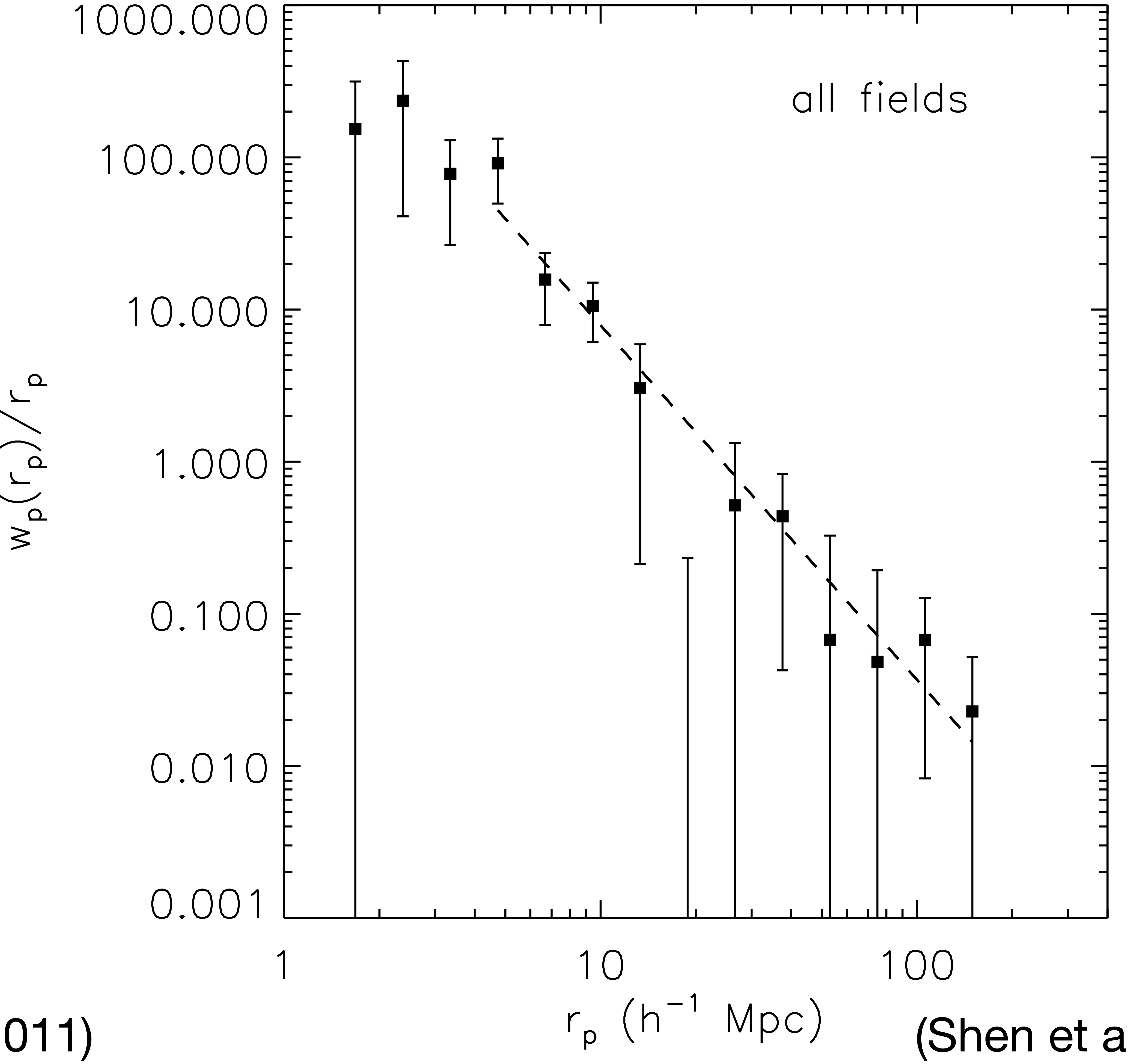
The clustering signal increases as the square of the number of galaxies in the sample

700,000 local galaxies in SDSS (over ~8000 deg²)



(Zehavi et al. 2011)

~4,400 High-z Quasars in SDSS (over ~4000 deg²)



(Shen et al. 2007)

Uncertainties in clustering measurements

Poisson errors

Since clustering is based on a pairs counting process, Poisson errors typically dominate the measurement.

Larger samples provide much better signal of clustering measurement.

Larger surveys → More sources → More pairs → Less uncertainties

In general terms (this also depends on how strong is the intrinsic clustering of the population).

$$\omega(R) = \frac{DD}{RR} - 1$$

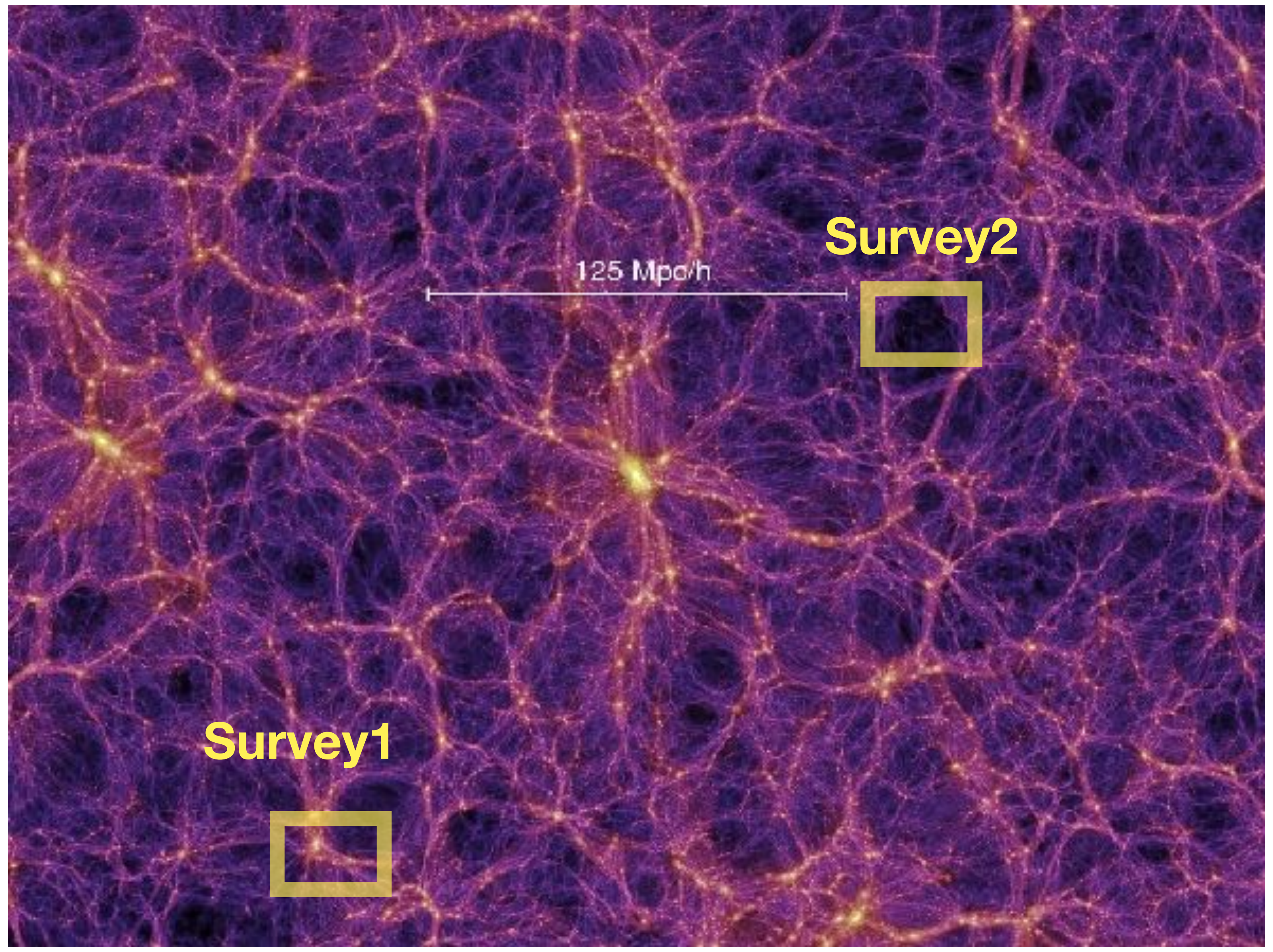
$$\Delta\omega(R) = \frac{\sqrt{DD}}{RR}$$

Uncertainties in clustering measurements

Cosmic variance

Beside Poisson errors, clustering measurements are associated with errors associated to cosmic variance effects, due to the fact that we are observing only one specific region of the universe.

These type of errors can be taken into account using statistical methods to compute error bars such a the jackknife method or the bootstrap method.



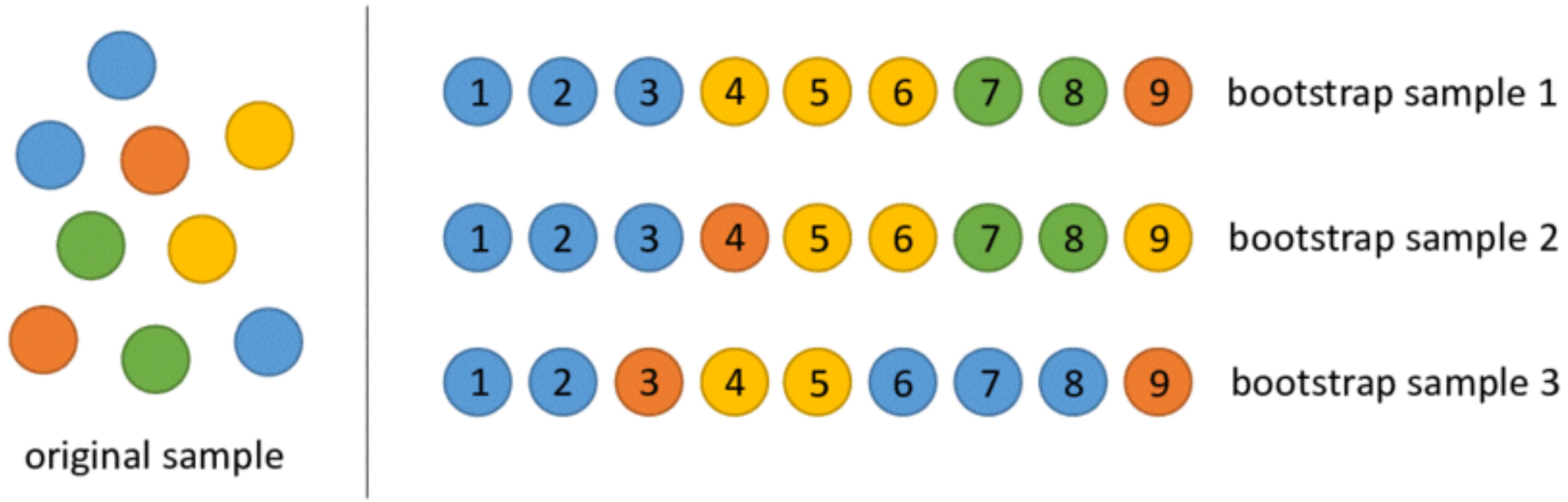
- The dominant error depends on the sample used to measure clustering. If the sample is small, Poisson errors will probably dominate the error budget.
- The jackknife or the bootstrap methods include both Poisson uncertainties and cosmic variance effects, then they are normally the preferred.

Uncertainties in clustering measurements

Bootstrap: Resampling method to estimate statistics on a population by sampling a dataset with replacement.

N Galaxies on the sky

Randomly select N galaxies from the sample



→ Compute w(R) or w(theta)

→ Compute w(R) or w(theta)

→ Compute w(R) or w(theta)

.....

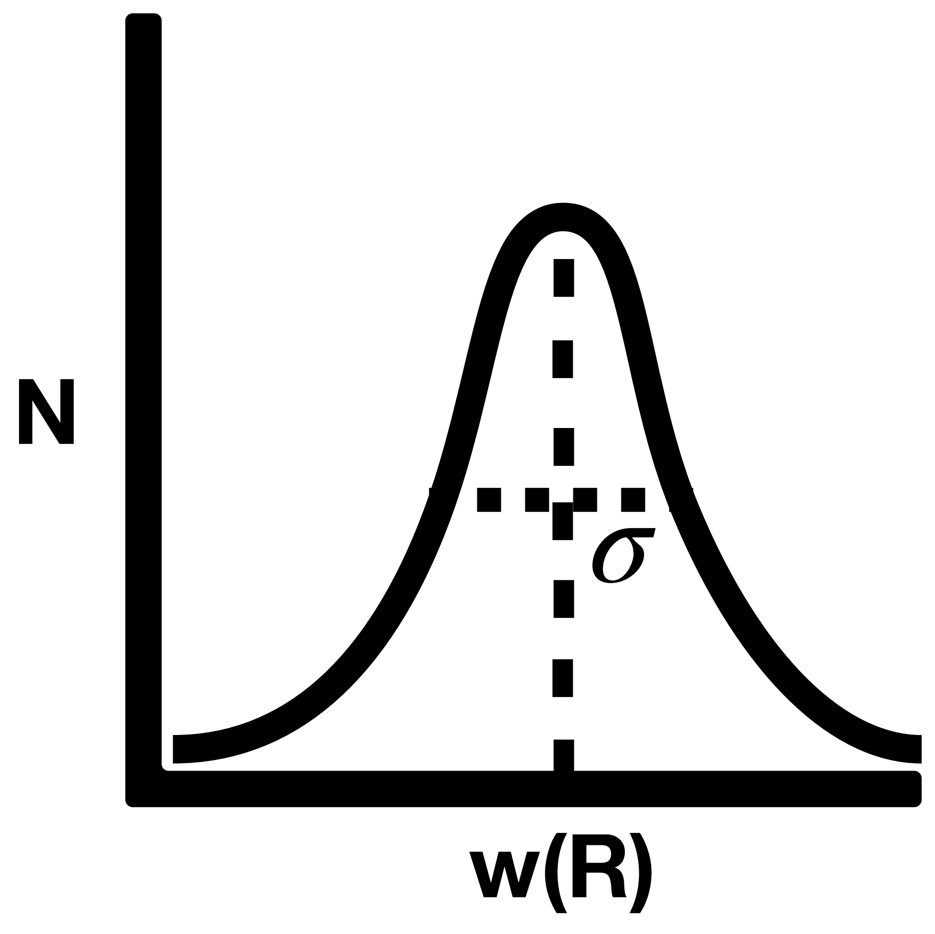
.....

bootstrap sample 10,000 → Compute w(R) or w(theta)

↓

Compute w(R) or w(theta)

Distribution of 10,000 w(R) or w(theta) values



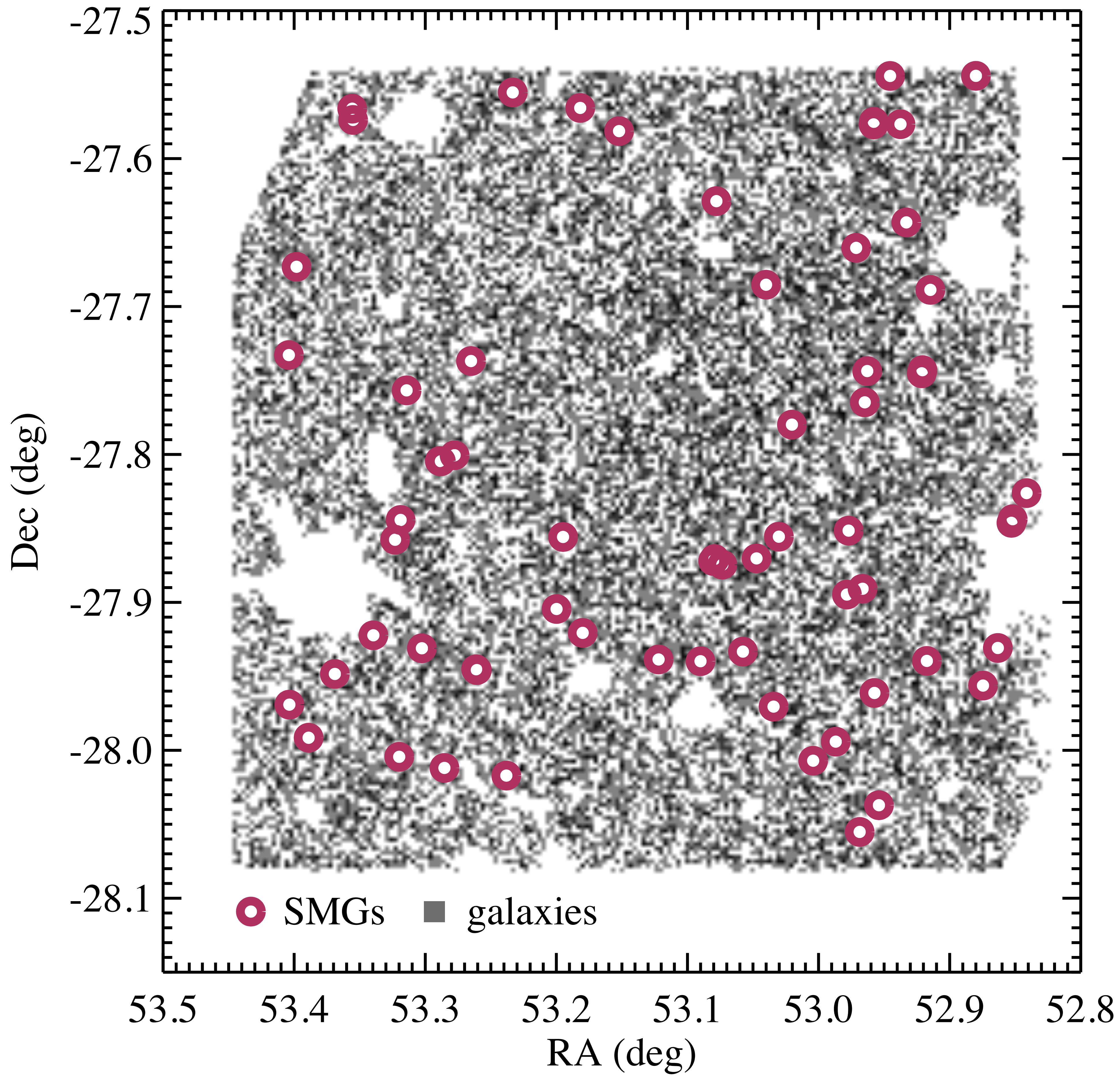
Uncertainty in w(R) is **σ**

(Adapted from Paola Galdi+2018)

Uncertainties in clustering measurements

Alternative techniques to reduce the uncertainties when the sample is small:

- Cross-correlations with a large population.



- We want to measure the clustering of SMGs, but we only have ~50 SMGs here, so Poisson errors will be huge.
- Fortunately, we also have a catalog of 10,000 normal galaxies over the same area of the sky.
- We measure the cross-correlation between SMGs and normal galaxies. Poisson errors will be small.
- I can also measure the auto-correlation of normal galaxies. Poisson errors will be small.
- Using the cross-correlation between SMG and normal galaxies and the auto-correlation of normal galaxies, I can infer what is the autocorrelation of SMGs.
- All the measurements have small Poisson errors.

Why is so important to compute the r_0 and γ?

- If we want to compare clustering between different populations (at the same epoch), comparing their r_0 and γ should be informative enough. This is useful to understand differences in the physical processes affecting each different population.

(Lecture 3 and 4)

- Additionally, r_0 and γ allow us to estimate the halo mass in which galaxies inhabit. In general terms, a higher clustered population reside in more massive halos. This provide insights to understand how galaxies populate the cosmic web, how different populations can be related (evolutionary link), and to constraint cosmological models.

(Lecture 4)

$$r_0, \gamma \longrightarrow M_{\rm halo}$$

Take home message

- The mathematical formalism to describe the level of clustering is the two-point correlation function $\xi(r)$

$$\boxed{dP = \bar{n}[1 + \xi(r)]\mathrm{dV}}$$

- To measure it we counts pairs of galaxies as a function of separation and divides by what is expected for an unclustered distribution.

- If we have 3D information (RA, Dec, z) we can measure the Projected correlation function. If we have 2D information (RA, Dec) we can measure the Projected angular correlation function.

- Larger samples provide much better signal of clustering measurement.