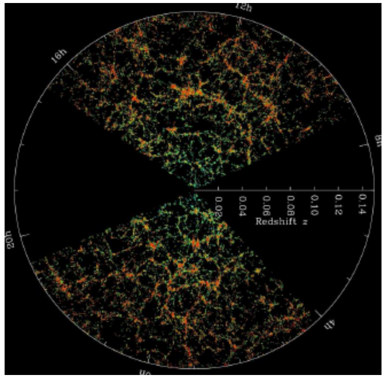
Uncertainty in correlation function estimators and BAO detection in current galaxy surveys

Antoine Labatie

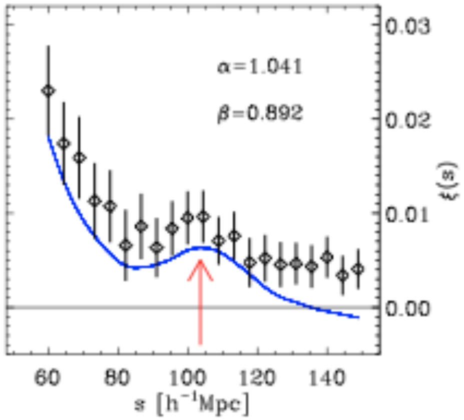
PhD Supervisors: Jean-Luc Starck and Marc Lachièze-Rey

Study of the galaxy distribution

- Galaxy surveys become very large
  - Testing cosmological model (ΛCDM)
  - Constraints on cosmic parameters (precision cosmology)



- Method:
  - Comparison with N-body simulations
  - Comparison with theoretical predictions



Fitting model correlation function to the data

Baryon Acoustic Oscillations (I)

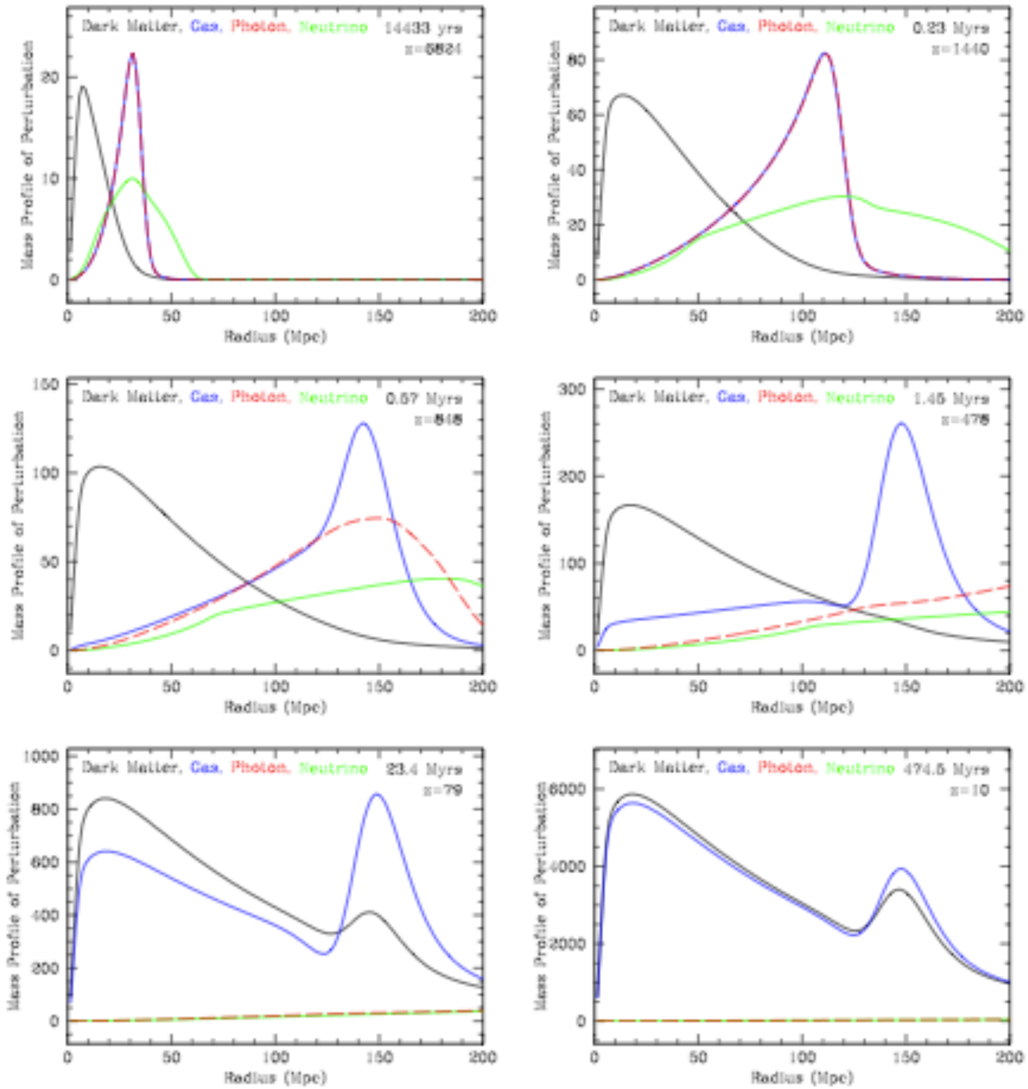
1- single pulse propagation with baryons and photons

2- drag of baryons and photons on dark matter

3- start of recombination: photons leak away

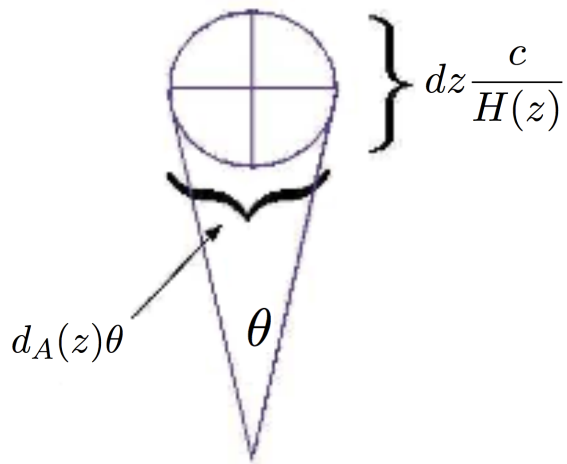
4- end of recombination: photons freely steam away from the 150Mpc over-density

5 & 6- dark matter drags baryon to the origin and baryons drag dark matter to the 150Mpc over-density



Baryon Acoustic Oscillations (II)

- BAOs are a feature imprinted in the galaxy distribution
- They create an excess of clustering at the sound horizon scale r_s
- They provide a statistical standard ruler for studying the geometry of the Universe
- If we can find this scale r_s in a redshif survey, we know the corresponding distance:



I) Uncertainty in the 2-point Correlation function & BAO study

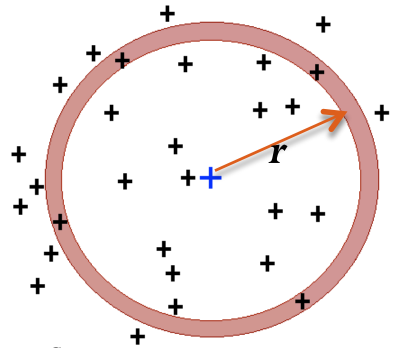
Standard tool for analysis: 2-point correlation function

- Probability to find a galaxy in volume δV:

$$\delta P = \bar{n}\delta V$$

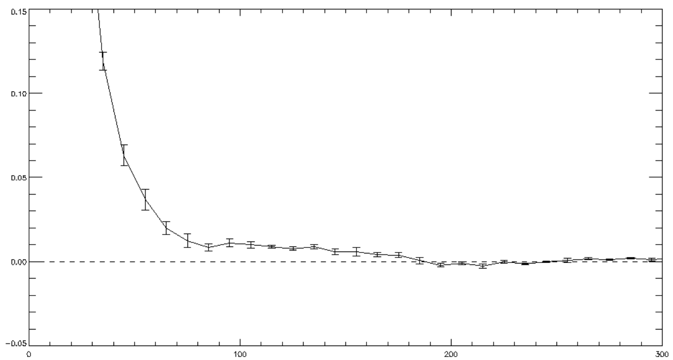
- Conditional to the existence of a galaxy, the probability to find another galaxy in volume δV at a distance ***r***:

$$(\delta P)_p(r) = \bar{n}(1+\xi(r))\delta V$$



*excess of galaxies at distance **r***

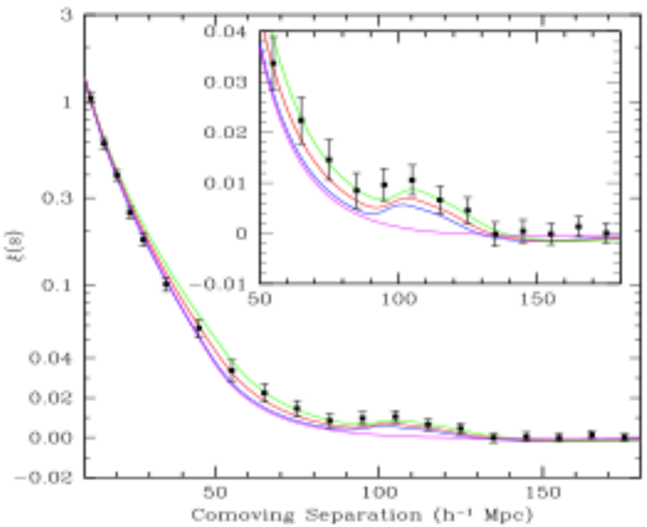
- $\xi(r)$ gives excess of probability to find pairs at distance ***r*** compared to a random distribution:



SDSS DR7 LRG ≈ 30000 galaxies

Correlation function in ΛCDM model

- Power law behavior at small scale: $\xi(r) \approx C\, r^{-\gamma}$
- *Baryon Acoustic Peak* at distance $\boldsymbol{r} \approx 105\ h^{-1}$ Mpc
- Correlation function can be computed for given cosmological parameters (assuming mass-luminosity bias)



Estimators of the Correlation function

- Estimated with ***uncertainty***:
  - only a finite volume (cosmic variance)
  - only finite number of galaxies (shot noise)

- Use random catalogues to calculate excess of pairs at distance ***r***

- Different estimators:
  - Peebles-Hauser (1974)

$$\tilde{\xi}_{PH}(r) = \frac{N_{RR}}{N_{DD}} \frac{DD(r)}{RR(r)} - 1$$

  - Davis-Peebles (1983)

$$\tilde{\xi}_{DP}(r) = \frac{N_{DR}}{N_{DD}} \frac{DD(r)}{DR(r)} - 1$$

  - Hamilton (1993)

$$\tilde{\xi}_{HAM}(r) = \frac{{N_{DR}}^2}{N_{DD}N_{RR}} \frac{DD(r)RR(r)}{[DR(r)]^2} - 1$$

  - Landy-Szalay (1993)

$$\tilde{\xi}_{LS}(r) = 1 + \frac{N_{RR}}{N_{DD}} \frac{DD(r)}{RR(r)} - 2\frac{N_{RR}}{N_{DR}} \frac{DR(r)}{RR(r)}$$

The integral constraint (I)

- By construction the estimators verify the integral constraint (because mean density is only estimated using sample)

estimated with sample density

$$(\delta P)_p(r) = \bar{n}(1+\xi(r))\delta V$$

$$\int \hat{\xi}(\mathbf{r})d^3\mathbf{r} \approx 0$$

- However the real correlation function $\xi(r)$ does ***not*** necessarily verify it

- Imposes a bias on the estimation

$$\mathbb{E}\left[\int \hat{\xi}(\mathbf{r})d^3\mathbf{r}\right] = \int \mathbb{E}[\hat{\xi}(\mathbf{r})]d^3\mathbf{r} \approx 0$$

$$\int \xi(\mathbf{r})d^3\mathbf{r} \neq 0$$

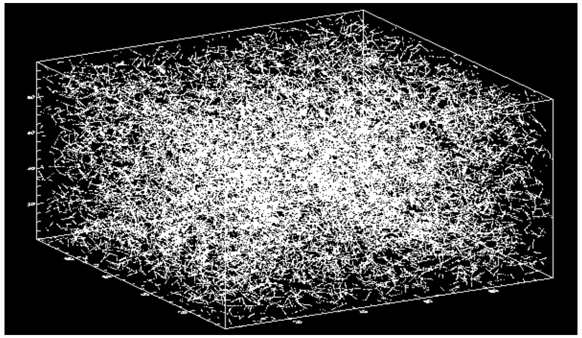
The integral constraint (II)

- The integral constraint has an effect for all survey sizes in a fractal Universe:
  - Idea supported by group of researchers (Labini et al.) claiming integral constraint makes estimation unreliable

- If we assume homogeneity at large scale, the integral constraint has an effect for too small survey sizes

Toy Model: Cox Segments (I)

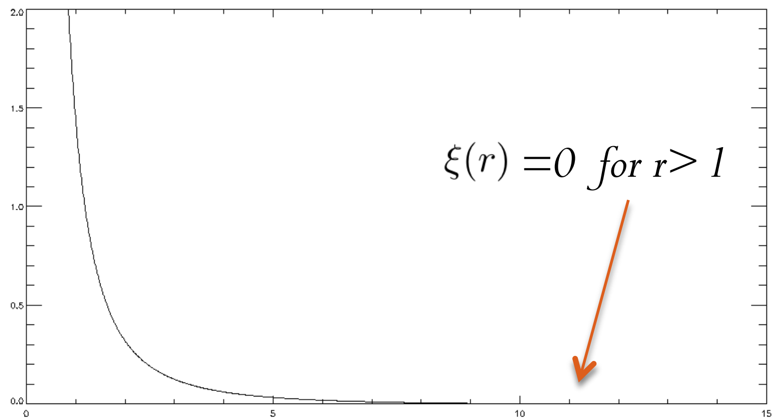
- Segments of length *l* put randomly in the volume



- Points put randomly on each segment

- Correlation function known analytically and ***always* ≥0**

$$\xi(r) = \begin{cases} \frac{1}{2\pi r^2 L_V} - \frac{1}{2\pi r l L_V} & \text{for} \quad r \leq l \\ 0 & \text{for} \quad r \geq l \end{cases}$$

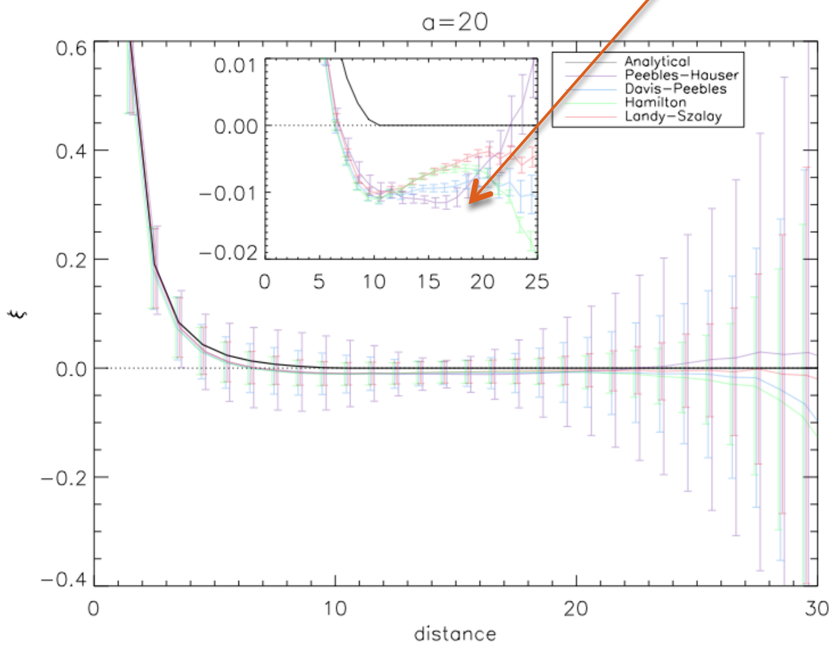


Toy Model: Cox Segments (II)

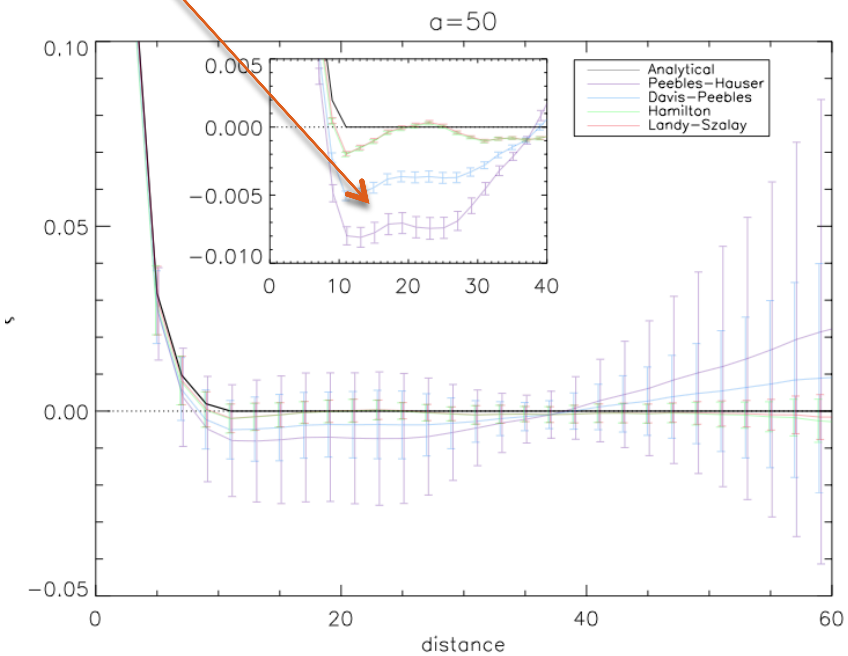
- Presence of a bias for cubic volumes of size a=20 and a=50 with l=10

Negative bias on the estimators:

$$\mathbb{E}[\xi(\hat{r})] \leq \xi(r)$$



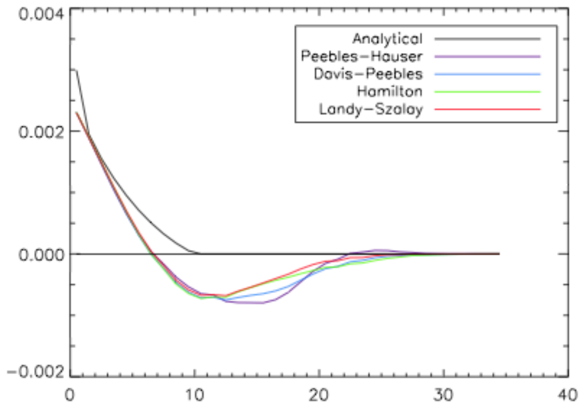
a=20



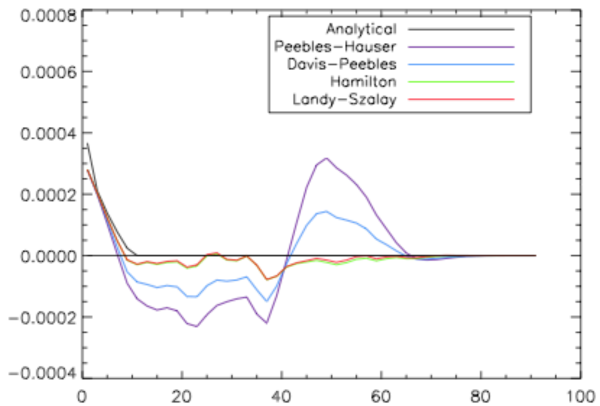
a=50

Toy Model: Cox Segments (III)

- Verification of the integral constraint: $$\int \hat{\xi}(\mathbf{r})d^3\mathbf{r} = \sum_i f(r_i)\xi(r_i) \approx 0 \quad ?$$
- The weighted correlation function sums up to 0 (effect is clear for a=20 but not for a=50)



a=20



a=50

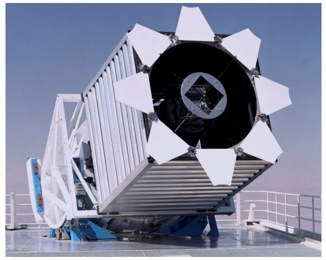
- Estimation is affected by integral constraint until volume gets large enough

Goals of our study

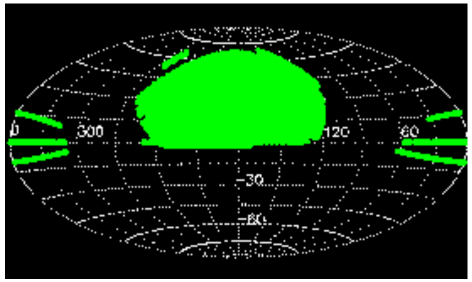
- We assume a ΛCDM model and generate lognormal realizations of galaxy surveys:
  - Compare the different estimators at large-scale for BAO study (bias and variance)
  - Evaluate the effect of the integral constraint: Is it causing a bias in the estimation for current galaxy surveys?
  - Is BAO detection reliable using correlation function ?

SDSS galaxy survey DR7

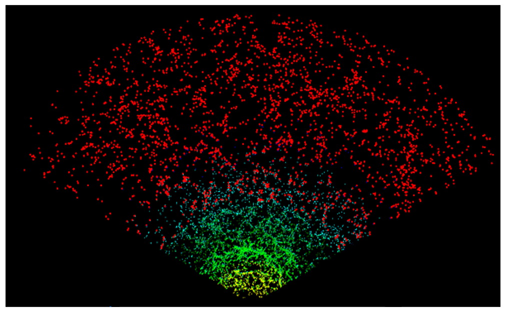
- 8 year program with 2.5m telescope at Apache Point (New Mexico)



  - Mapped 7500 square degree of the sky



  - Spectrum for 930 000 galaxies (largest galaxy survey up to date)
  - 1 magnitude-limited samples of galaxies (main) up to $D \approx 600\ h^{-1}$Mpc

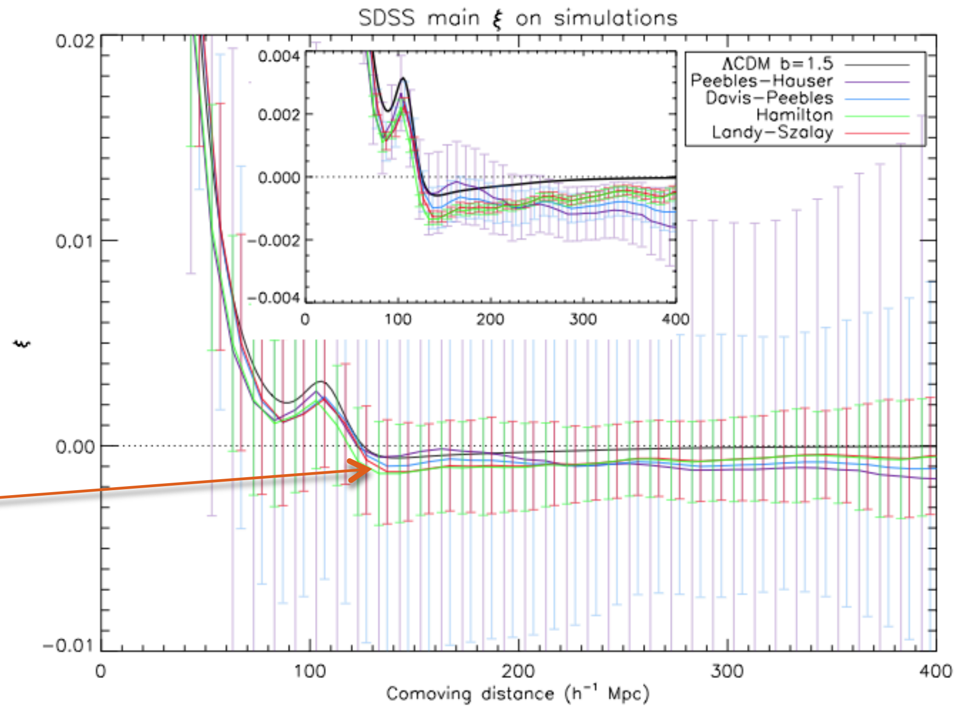


  - 1 approximately volume-limited of luminous red galaxies (LRG) up to $D \approx 1150\ h^{-1}$Mpc

Integral Constraint in SDSS (I)

- 200 lognormal simulations of SDSS main sample:
  - Superiority of Landy-Szalay and Hamilton estimators
  - Bias for all estimators ≈ half of the variance
  - Bias must be taken into account in ***uncertainty***

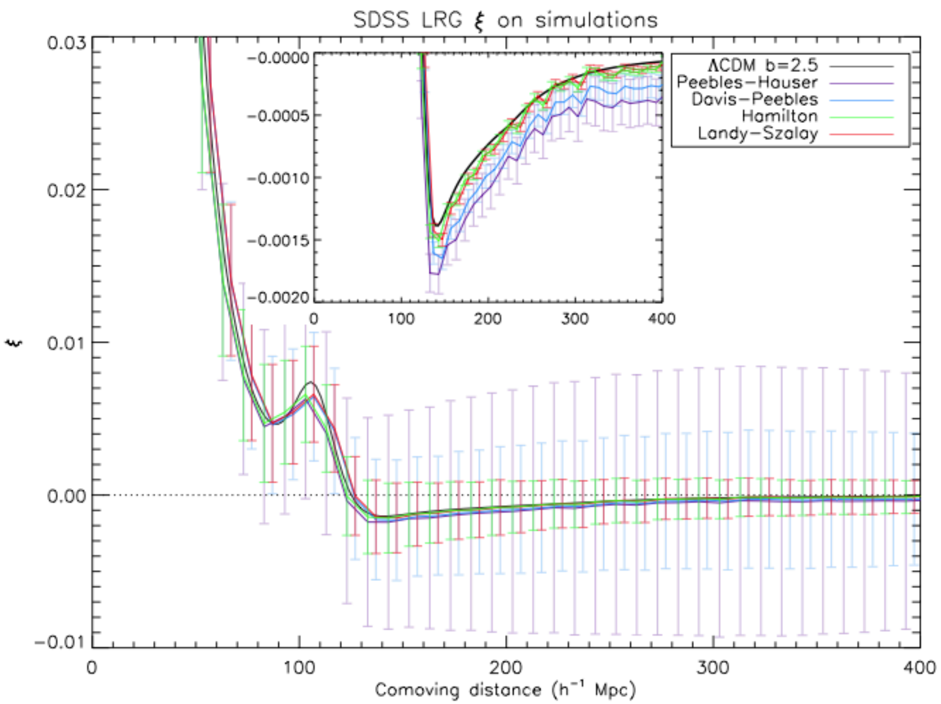
Negative bias of the estimators



Correlation function on SDSS Main simulations

Integral Constraint in SDSS (II)

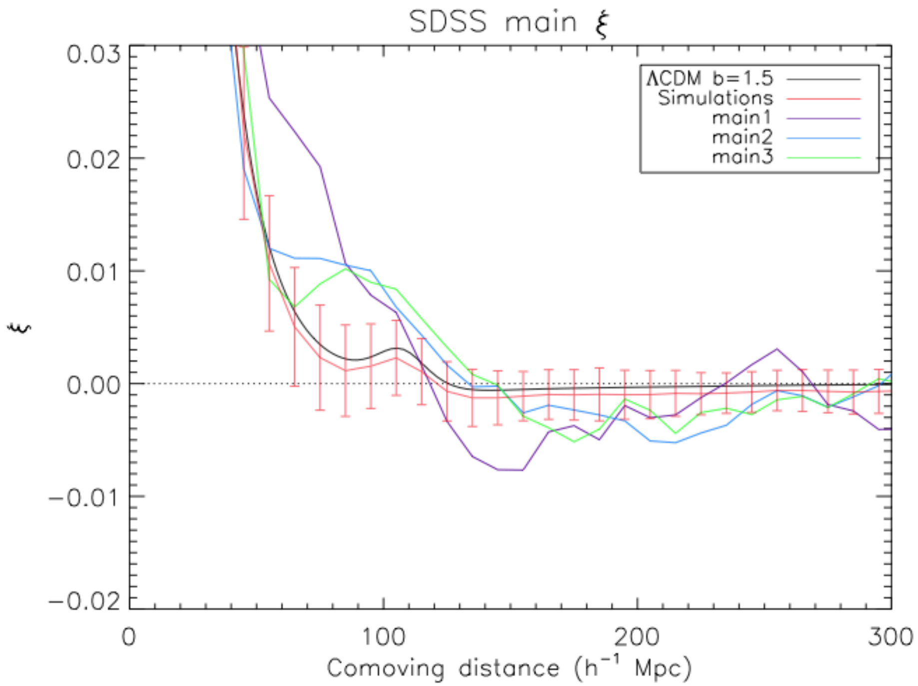
- 2000 lognormal simulations of SDSS LRG sample:
  - Superiority of Landy-Szalay and Hamilton estimators
  - Estimators are ***unbiased***



Correlation function on SDSS LRG simulations

BAO detection in SDSS DR7 (I)

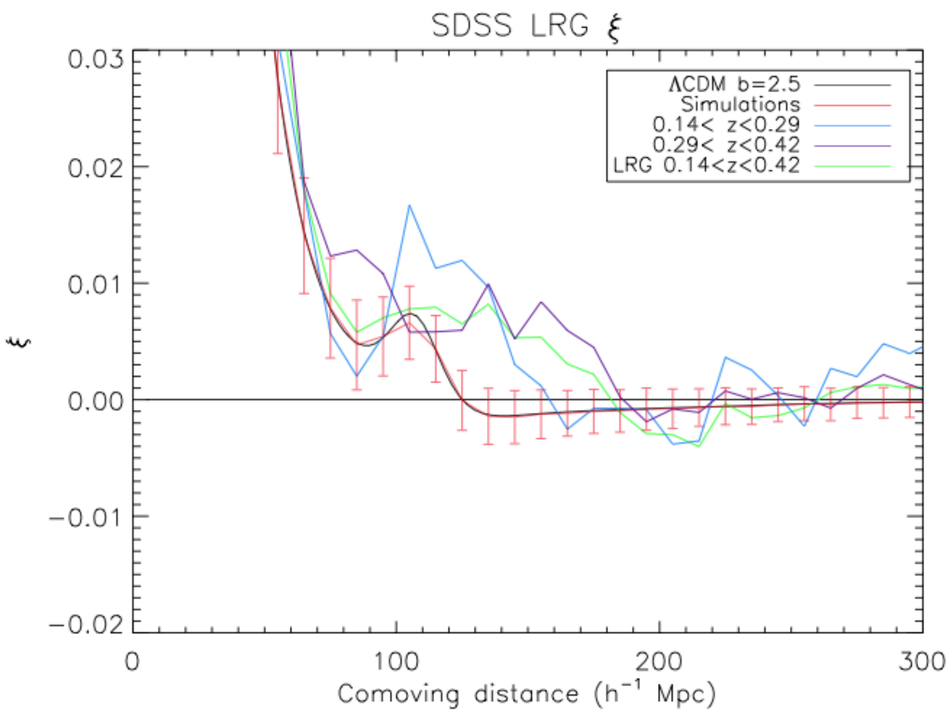
- Signal is too low in main Sample for BAO study



Correlation function on SDSS Main samples

BAO detection in SDSS DR7 (II)

- Reliable detection of the BAO in LRG sample *(without any bias in the estimation)*
- Problem: peak wider than expected *(χ^2 test: p-value ≈ 0.01)*



Correlation function on SDSS LRG samples

Conclusion

- We find agreement with previous studies: superiority of Landy-Szalay and Hamilton estimators
- We find that SDSS LRG sample used for BAO is not affected by the integral constraint (unbiased estimation)
- We find BAO detection is not possible in Main sample but is possible in LRG sample
- This agrees with BAO peak in SDSS LRG correlation but:
  - unexplained excess of clustering (peak wider than expected)