Hierarchical Galaxy Clustering in the 2dFGRS

Peder Norberg (ETHZ)
in collaboration with
Carlton Baugh, Darren Crotton,
Enrique Gaztanaga
&
the 2dFGRS Team

Moriond, March 2004

Contents: Counts-in-Cells Analysis of 2dFGRS

- Overview of the 2dFGRS
- Counts-in-Cells & hierarchical scaling model
- Methodology of our analysis
  - volume limited samples
  - incompleteness corrections
  - error estimation & effects of superstructures
- Hierarchical clustering in 2dFGRS:
  - 2^{nd} order to 6^{th} order
  - luminosity segregation for 2^{nd} to 4^{th} order
- Conclusions

2dFGRS in a Nutshell

- The 2dF galaxy redshift survey (2dFGRS) input catalogue is selected from *UKST photographic plates*: SGP & NGP (main strips) + 100 random fields (spread over full APM survey).
- Magnitude limited survey: $14.0 \leq b_J \leq \sim 19.4$, with $\sigma(b_J) \sim 0.12$ mag.; galaxy completeness ~91%; stellar contamination ~6%; (B-R) from SCOS.
- 225’000 unique galaxy redshifts over ~1500 sq. deg., with average spectroscopic completeness ~85%; median redshift ~0.11
- 2dFGRS is since June’03 in the public domain!

Counts-in-Cells (CiC: I)

What does CiC consist in? Just counting the number of elements in a set of cells of volume V (for spheres: radius R) and create the associated count probability distribution function (CPDF):

$$P_N(R) = \frac{N_N}{N_T}$$

where N_N is the number of cells containing N galaxies out of a total number of cells thrown down N_T.

Counts-in-Cells (CiC: II)

- The moments of the CPDF are given by:

$$m_p(R) = \langle (N - \bar{N})^p \rangle = \sum_{N=0}^{\infty} P_N(R)(N - \bar{N})^p$$

where N is the mean number of galaxies obtained from the CPDF:

$$\bar{N} = \sum_{N=0}^{\infty} N P_N$$

- The logarithm of the moment generating function is equal to the cumulant generating function. The cumulant, μ_p, is simply the volume averaged reduced p-point correlation function:

$$< \delta^p >_c = \mu_p = \bar{\xi}_p(R)$$

The hierarchical scaling model

If gravitational instability is important in shaping large scale structure, then one expects the following scaling relation:

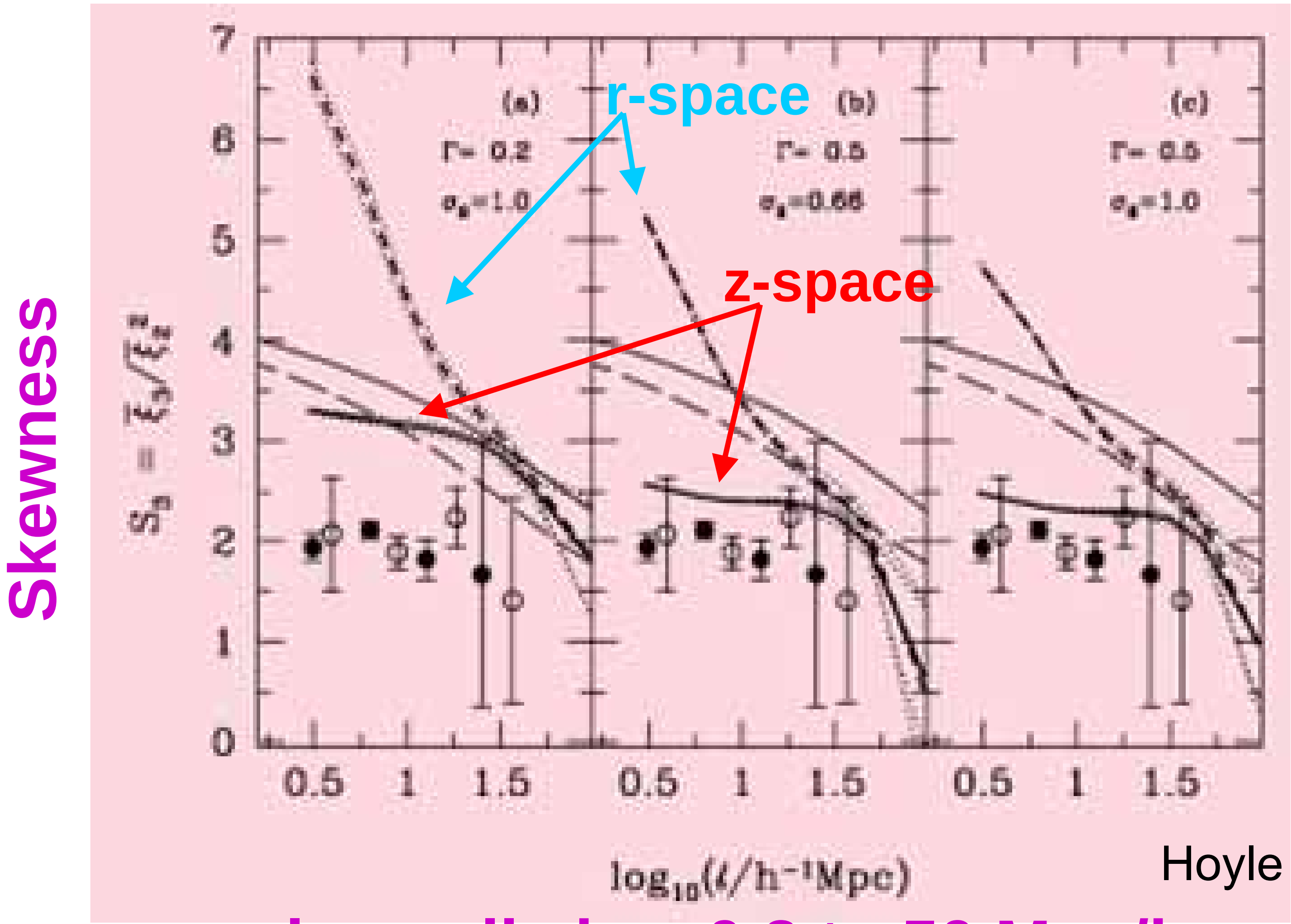
$$\bar{\xi}_p = S_p\, \bar{\xi}_2^{\,p-1}$$

where the volume averaged reduced p-point correlation function is:

$$\bar{\xi}_p(V) = \frac{1}{V^p}\int_V \mathrm{d}^3 r_1 \,\ldots\, \mathrm{d}^3 r_p\, \xi_p(\mathbf{r}_1, \,\ldots\, , \mathbf{r}_p)$$

NB: the volume average p-point correlation function is independent of position with the assumption of *statistical homogeneity* and *isotropy* of the cosmic density field: *not true on small scales in redshift space*! S_3 is usually called skewness, whereas S_4 is referred as kurtosis.

Real-Redshift Space & Small-Large Scales



Hoyle etal (2000)

log cell size: 0.3 to 50 Mpc/h

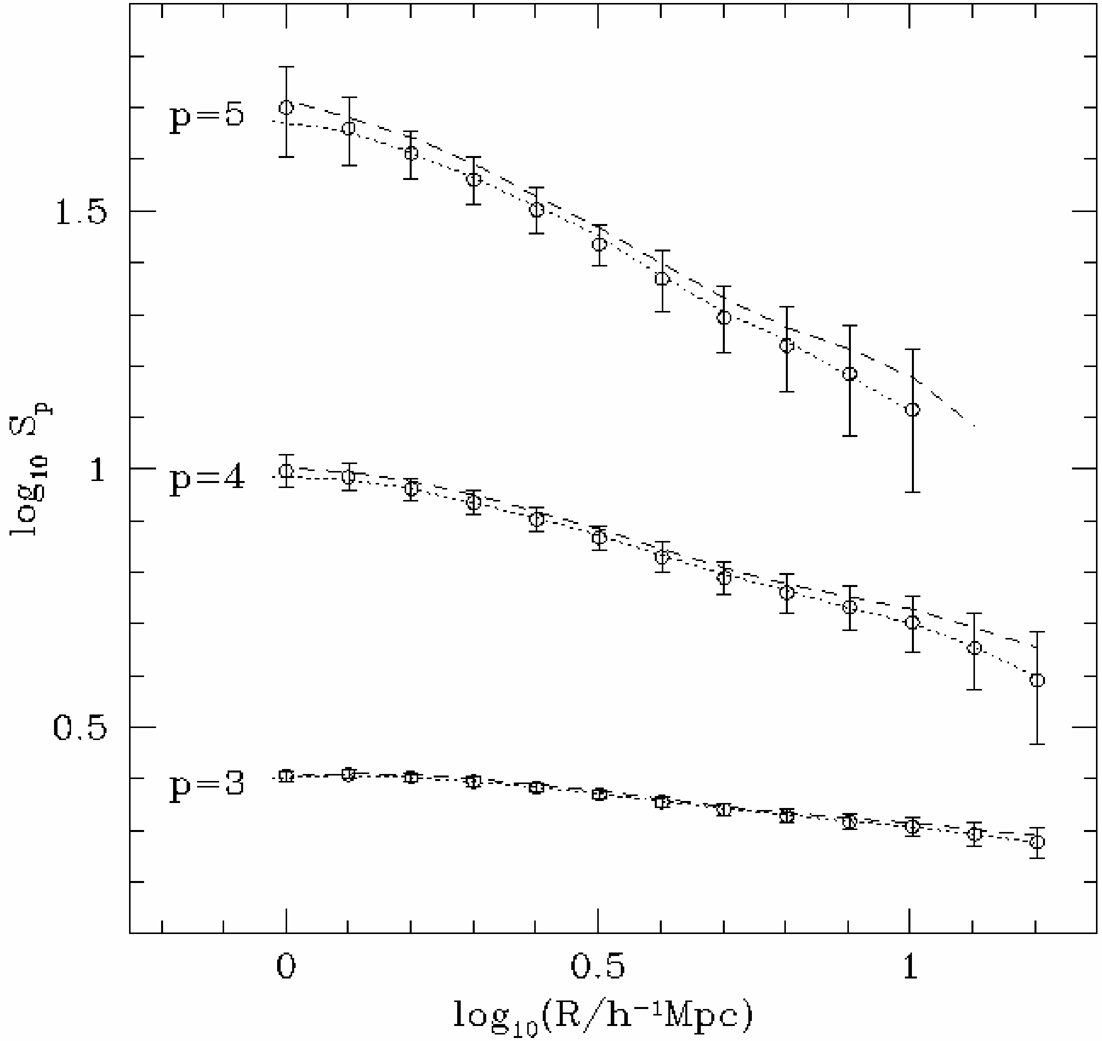
Counts-in-Cells (CiC: III)

- We opt for massively oversampling the survey volume, by randomly throwing down $25\ 10^6$ spheres of radius going from R=0.5 Mpc/h to 30 Mpc/h

- Issues:
  - selection function for the spheres: we use volume limited samples with constant radial mean density.
  - survey geometry (drill holes, edge effects) and spectroscopic completeness are accounted for: we apply a volume correction (instead of a number correction) and sphere which are reduced by more than 50% are discarded; hence spheres of a given size can by construction only contribute to their own bin!
  - Extremely correlated errors: principal component analysis shows that the first 2 or 3 eigenvectors are responsible for ~90% of the variance => PCA essential for the fits!

Incompleteness corrections

Log of Hierarchical Amplitudes: p=3,4,5



Log of sphere radius: 1.0 to 30 Mpc/h

Counts-in-Cells (CiC: III)

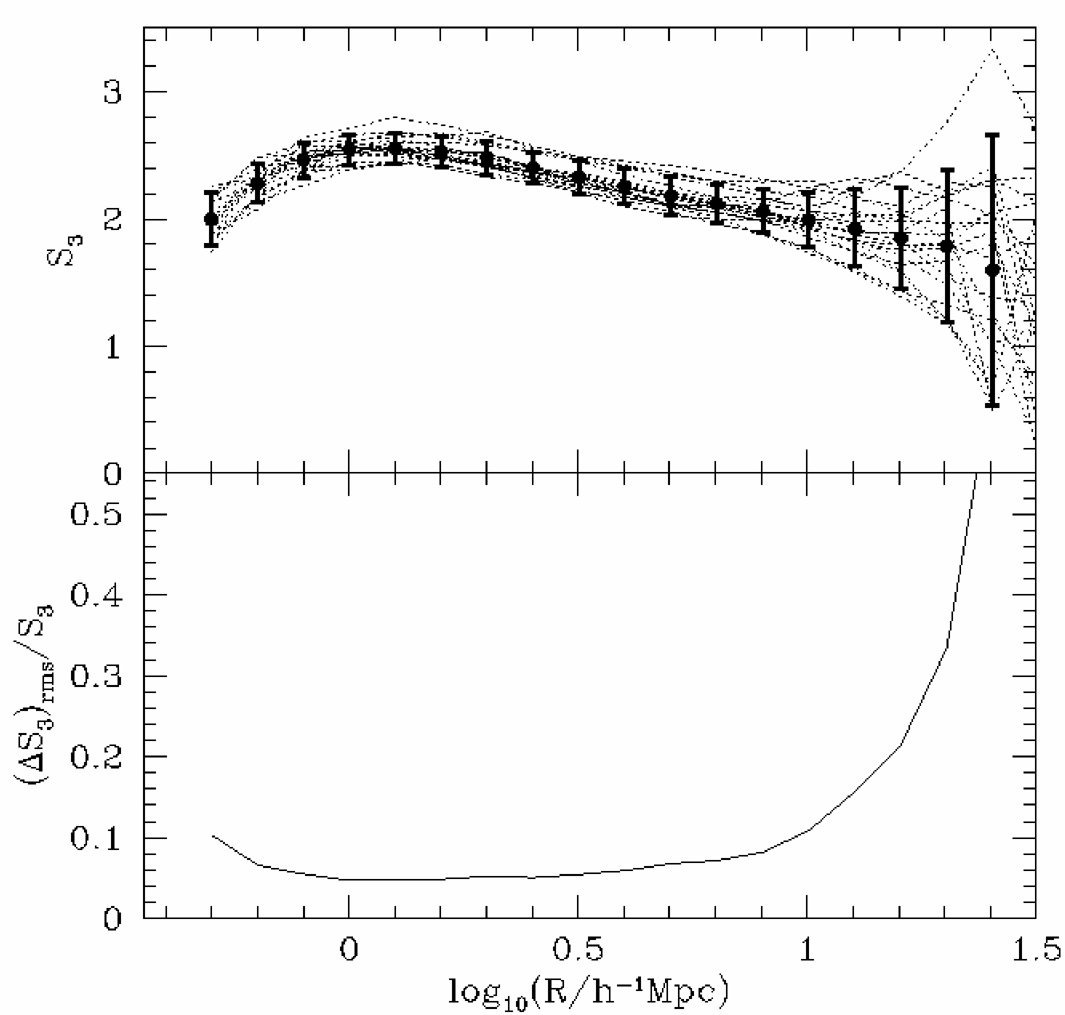
- We opt for massively oversampling the survey volume, by randomly throwing down $25\ 10^6$ spheres of radius going from R=0.5 Mpc/h to 30 Mpc/h

- Issues:
  - selection function for the spheres: we use volume limited samples with constant radial mean density.
  - survey geometry (drill holes, edge effects) and spectroscopic completeness are accounted for: we apply a volume correction (instead of a number correction) and sphere which are reduced by more than 50% are discarded; hence spheres of a given size can by construction only contribute to their own bin!
  - Extremely correlated errors: principal component analysis shows that the first 2 or 3 eigenvectors are responsible for ~90% of the variance => PCA essential for the fits!

Errors on the skewness estimated from 22 Hubble Volume mock catalogues

Skewness

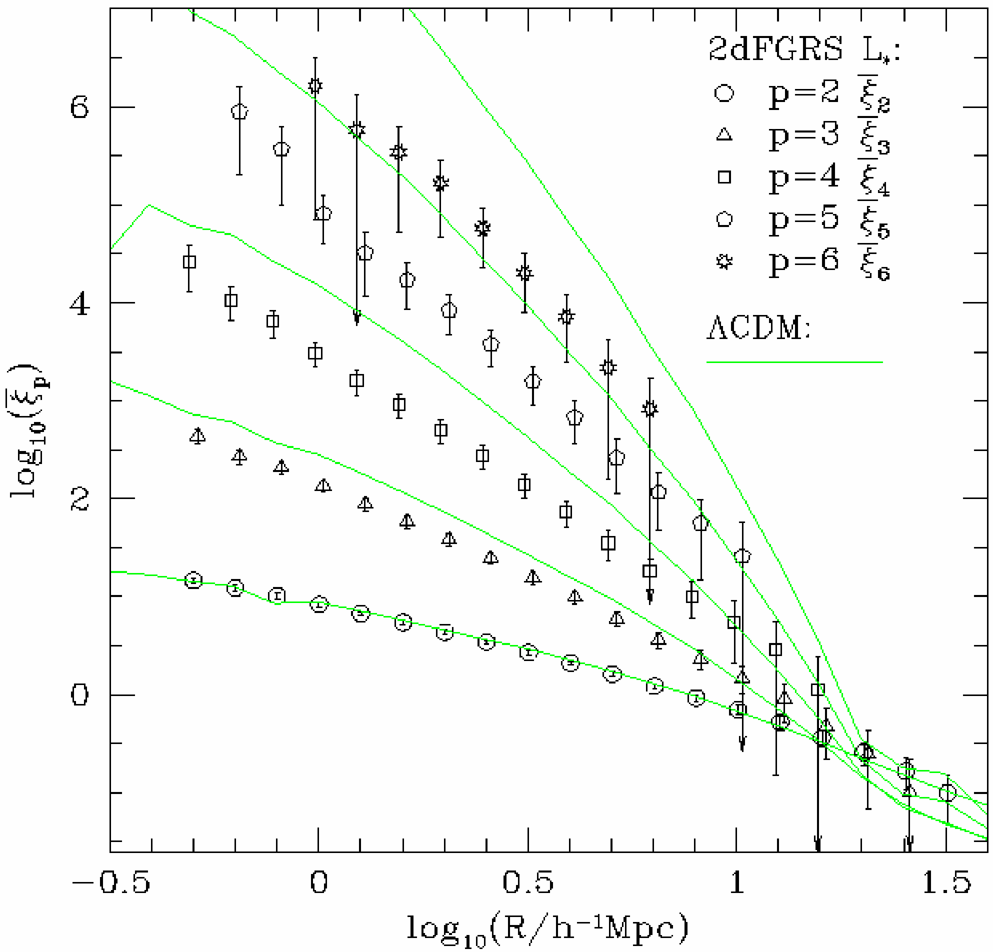
Fractional rms error



Log of sphere radius: 1.0 to 30 Mpc/h

Comparison between ΛCDM and 2dFGRS: L* sample in z-space

Log of ξ_p: p=2,3,4,5,6

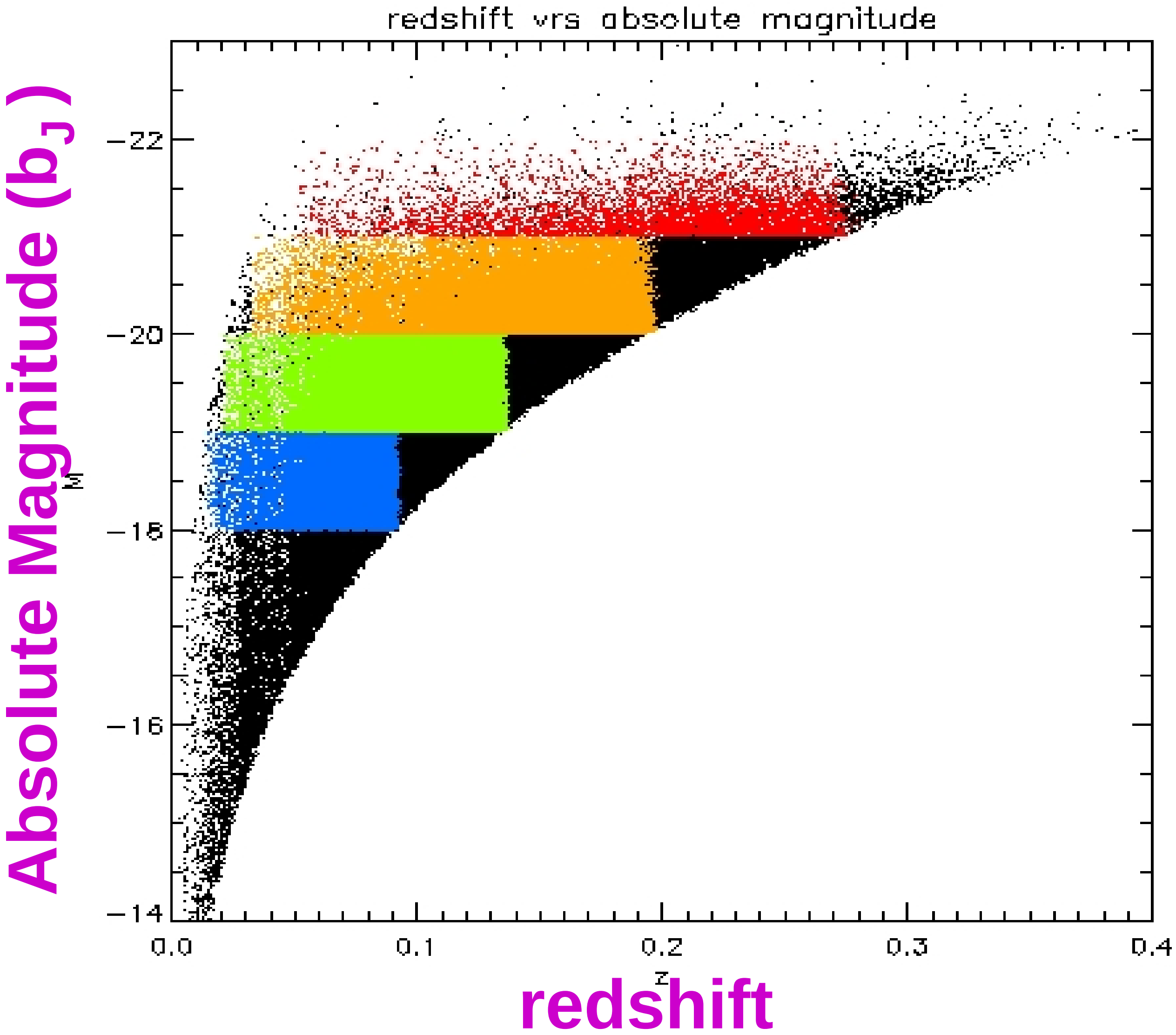


Log of sphere radius: 0.3 to 30 Mpc/h

Methodology: sample selection

- High completeness sectors in SGP & NGP: effective area of 1140 sq. deg., containing >190’000 galaxy redshifts.
- 5 volume limited samples (two below L*, one L* sample, two above L*):
  - each one magnitude wide, defined through a bright and faint absolute magnitude.
  - not fully independent volumes: the brighter the characteristic luminosity, the more independent the sample is with respect to the fainter ones.

Volume Limited Samples



Sample Properties

Table 1. Properties of the combined 2dFGRS SGP and NGP volume-limited catalogues (VLCs). Column 1 gives the numerical label of the sample. Columns 2 and 3 give the faint and bright absolute magnitude limits that define the sample. The fourth column gives the median luminosity of each volume limited sample in units of L_*, computed using the Schechter function parameters quoted by Norberg et al. (2002b). Columns 5, 6 and 7 give the number of galaxies, the mean number density and the mean inter-galaxy separation for each VLC, respectively. Columns 8 and 9 state the redshift boundaries of each sample for the nominal apparent magnitude limits of the survey; columns 10 and 11 give the corresponding comoving distances. Finally, column 12 gives the combined SGP and NGP volume. All distances are comoving and are calculated assuming standard cosmological parameters ($\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$).

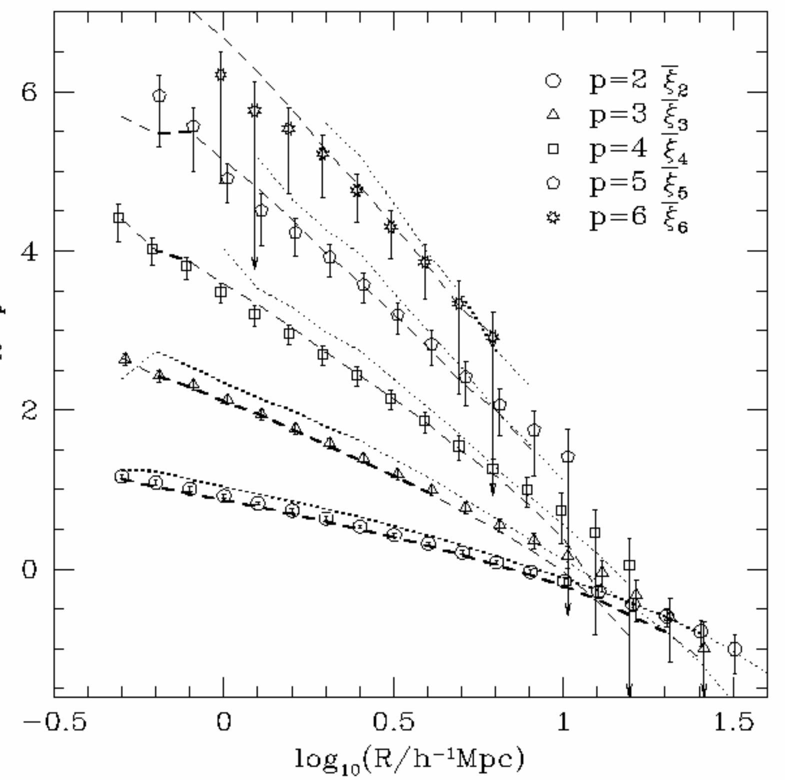
VLC ID	Mag. range $M_{b_J} - 5\log_{10} h$		Median lum. L/L_*	N_G	ρ_{ave} $10^{-3}h^3\mathrm{Mpc}^{-3}$	d_{mean} $h^{-1}\mathrm{Mpc}$	z_{min}	z_{max}	D_{min} $h^{-1}\mathrm{Mpc}$	D_{max} $h^{-1}\mathrm{Mpc}$	Volume $10^6 h^{-3}\mathrm{Mpc}^3$
1	−17.0	−18.0	0.13	8038	10.9	4.51	0.009	0.058	24.8	169.9	0.74
2	−18.0	−19.0	0.33	23290	9.26	4.76	0.014	0.088	39.0	255.6	2.52
3	−19.0	−20.0	0.78	44931	5.64	5.62	0.021	0.130	61.1	375.6	7.97
4	−20.0	−21.0	1.78	33997	1.46	8.82	0.033	0.188	95.1	537.2	23.3
5	−21.0	−22.0	3.98	6895	0.110	20.9	0.050	0.266	146.4	747.9	62.8

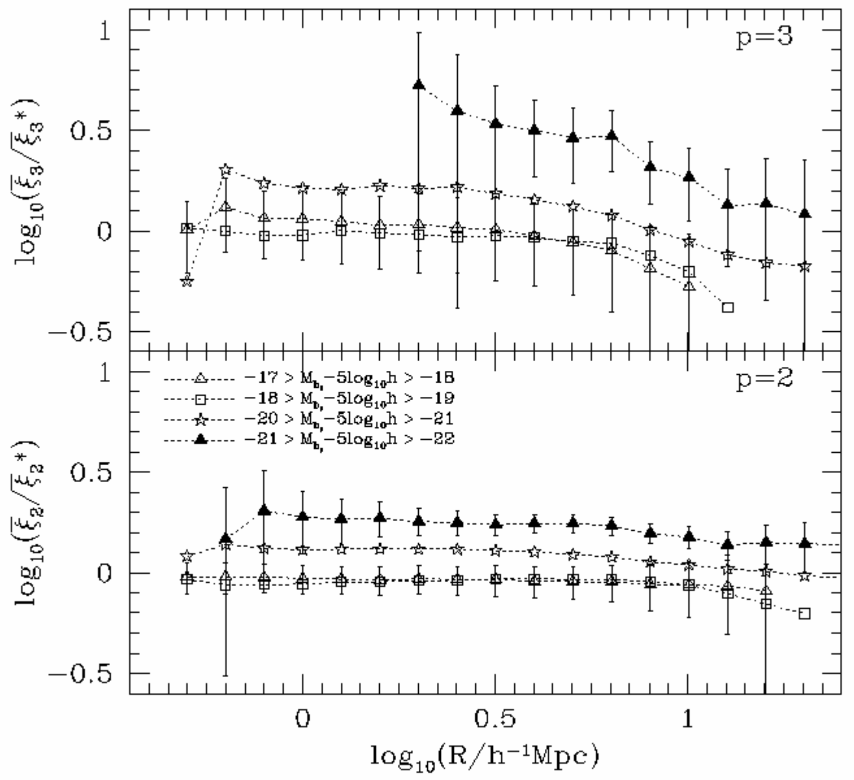
Table 2. The best fit values and 2−σ error ($\Delta\chi^2 = 4$) for S_p (columns 4 to 7). The range of scales used in the fits is given in columns 2 and 3. The number in brackets after each error gives the reduced χ^2 value for the fit, using the number of degrees of freedom derived from the principal component analysis. The last two columns give the relative linear bias, b_r (defined by Eq. 14) and the second order bias term, c_2' (defined by Eq. 17). Errors are 2−σ. The reference sample is sample number 3. These values are obtained for the full volume limited samples. A blank entry indicates that a reliable measurement of the particular hierarchical amplitude was not possible for the sample in question.

VLC ID	R_{min} $h^{-1}\mathrm{Mpc}$	R_{max} $h^{-1}\mathrm{Mpc}$	S_3	S_4	S_5	S_6	b_r	c_2'
1	0.71	7.1	$2.58 \pm 0.37\,(0.1)$	$9.3 \pm 4.0\,(0.1)$	$34 \pm 32\,(0.1)$	− − −	$0.96 \pm 0.16\,(0.1)$	$0.17 \pm 0.25\,(0.1)$
2	0.71	7.1	$2.38 \pm 0.25\,(0.1)$	$8.2 \pm 2.3\,(0.9)$	$36 \pm 20\,(0.4)$	$185 \pm 170(0.1)$	$0.96 \pm 0.08\,(0.3)$	$0.11 \pm 0.13\,(0.1)$
3	0.71	7.1	$1.95 \pm 0.18\,(6.1)$	$5.5 \pm 1.4\,(2.3)$	$18 \pm 11\,(1.9)$	$46 \pm 50(1.1)$	1	0
4	0.80	8.9	$2.01 \pm 0.17\,(1.2)$	$6.0 \pm 1.5\,(0.6)$	$22 \pm 12\,(0.4)$	$71 \pm 80(0.3)$	$1.13 \pm 0.06\,(2.8)$	$0.10 \pm 0.08\,(0.3)$
5	2.2	11.2	$2.39 \pm 0.63\,(0.5)$	$6.8 \pm 7.0\,(0.4)$	− − −	− − −	$1.30 \pm 0.14\,(0.9)$	$0.33 \pm 0.31\,(0.5)$

Higher Order Clustering: p=2 to 6

Log of ξ_p: p=2,3,4,5,6

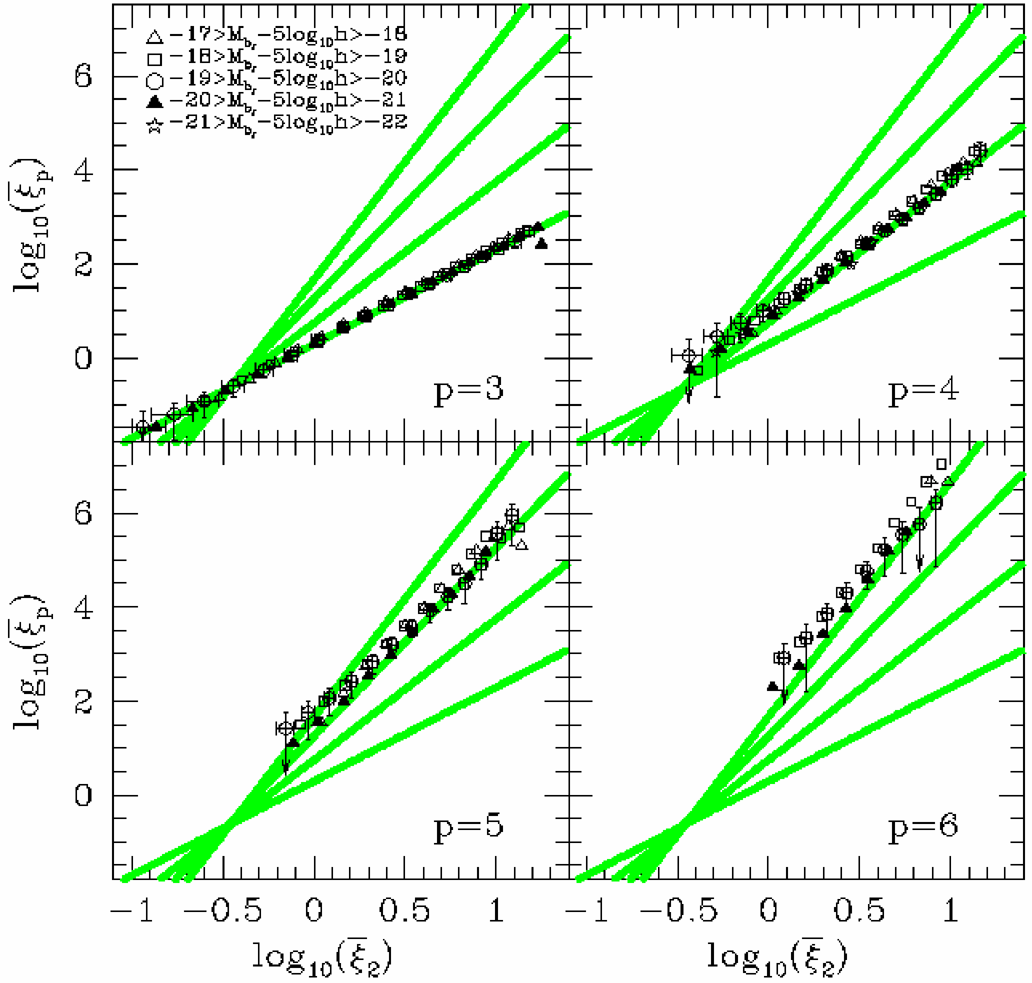




Log of sphere radius: 0.3 to 30 Mpc/h

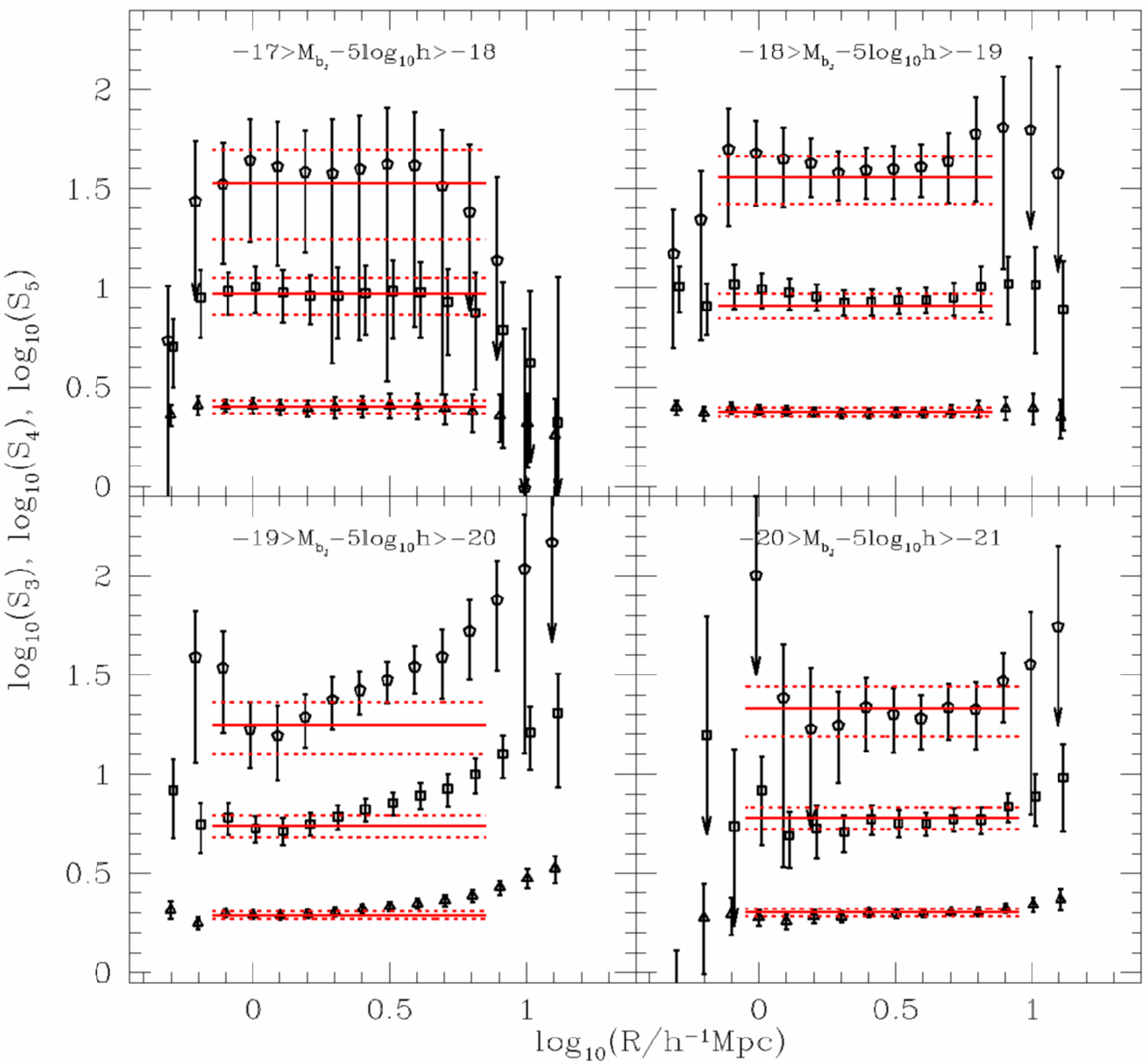
Hierarchical Scaling: p=3,4,5,6

Log of ξ_p: p=3,4,5,6



Log of variance => 30 to 0.3 Mpc/h

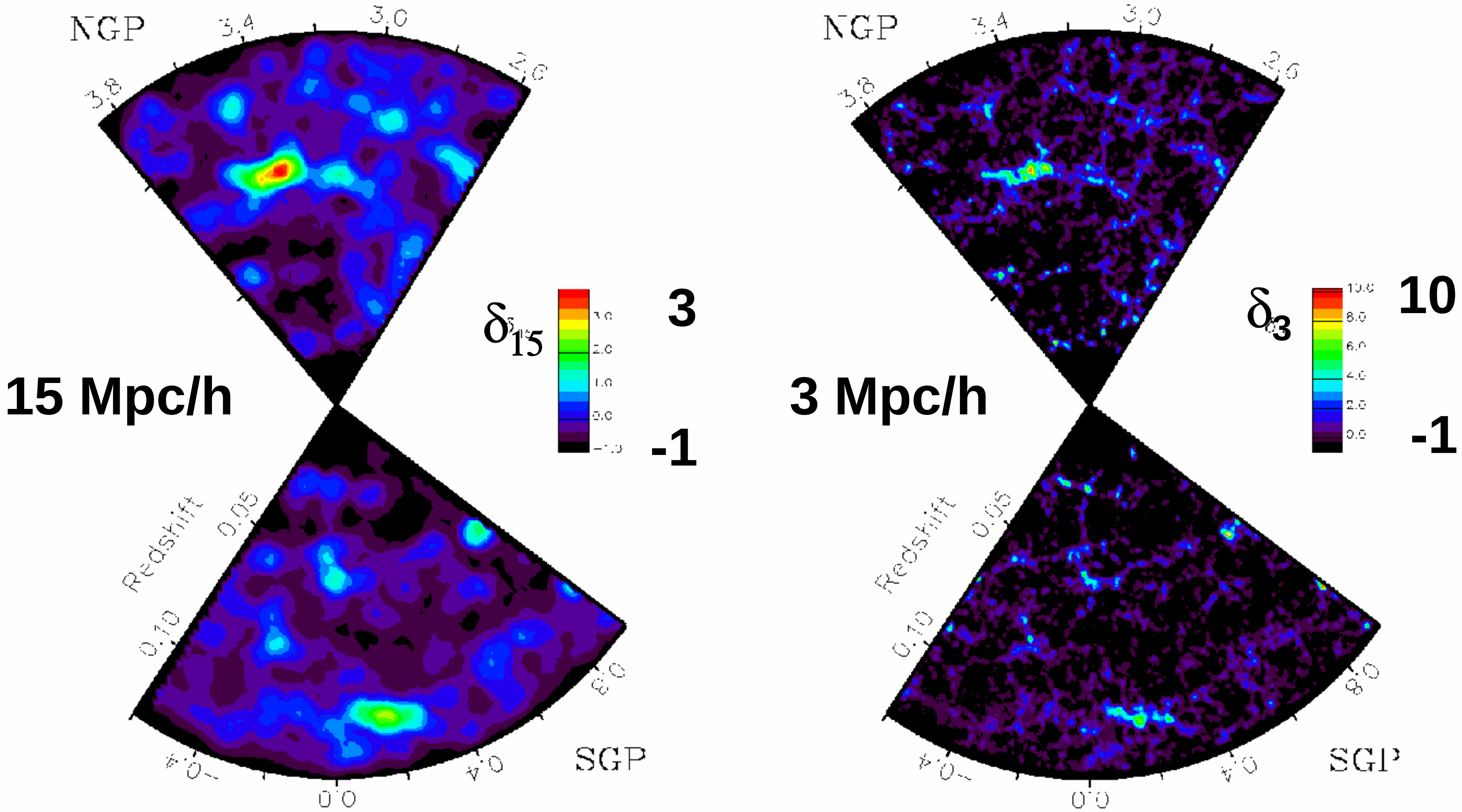
Hierarchical Amplitudes: S_3 to S_5



Log of Hierarchical Amplitudes: p=3,4,5

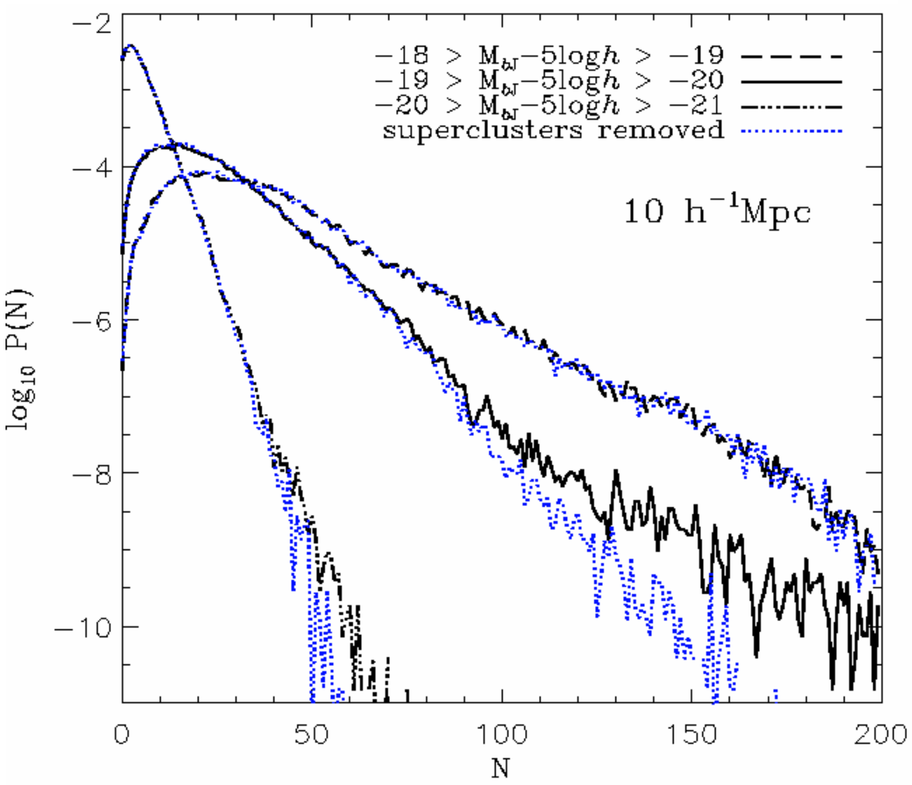
Log of sphere radius: fit between 0.8 & 8 Mpc/h

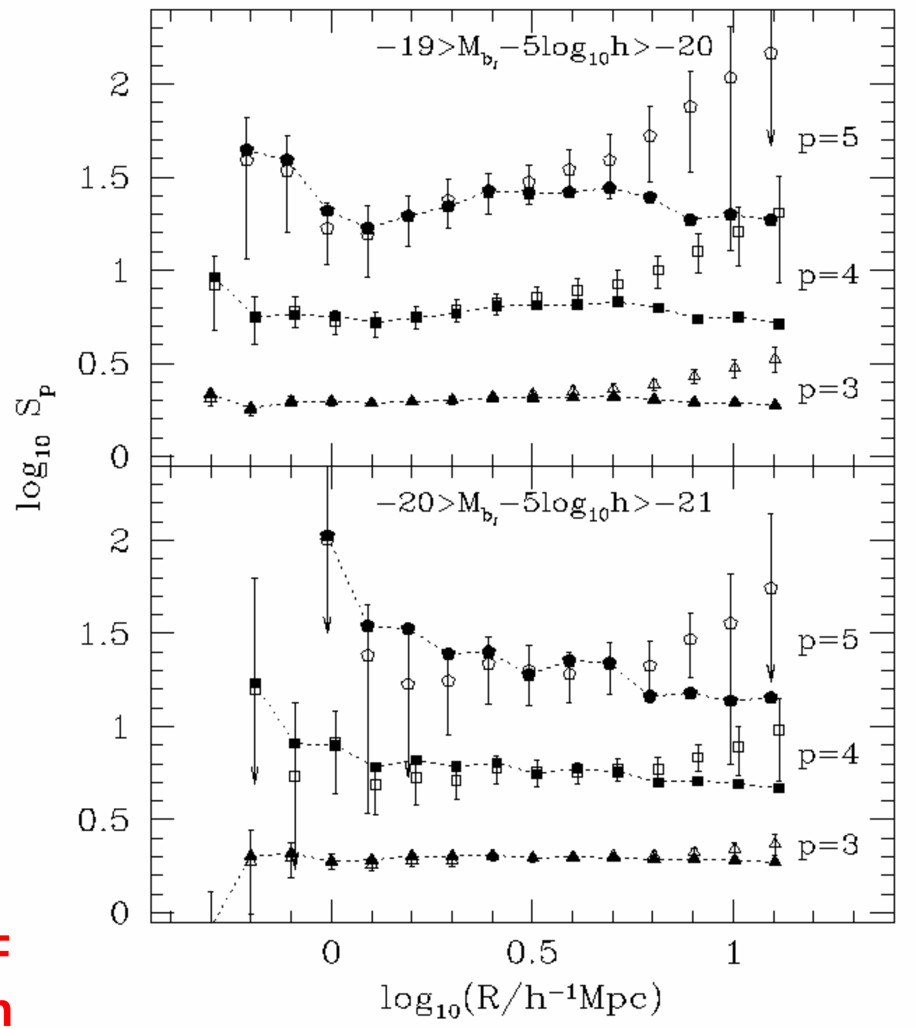
Projected galaxy density in L* volume limited sample



Effects of the superstructures on the hierarchical amplitudes

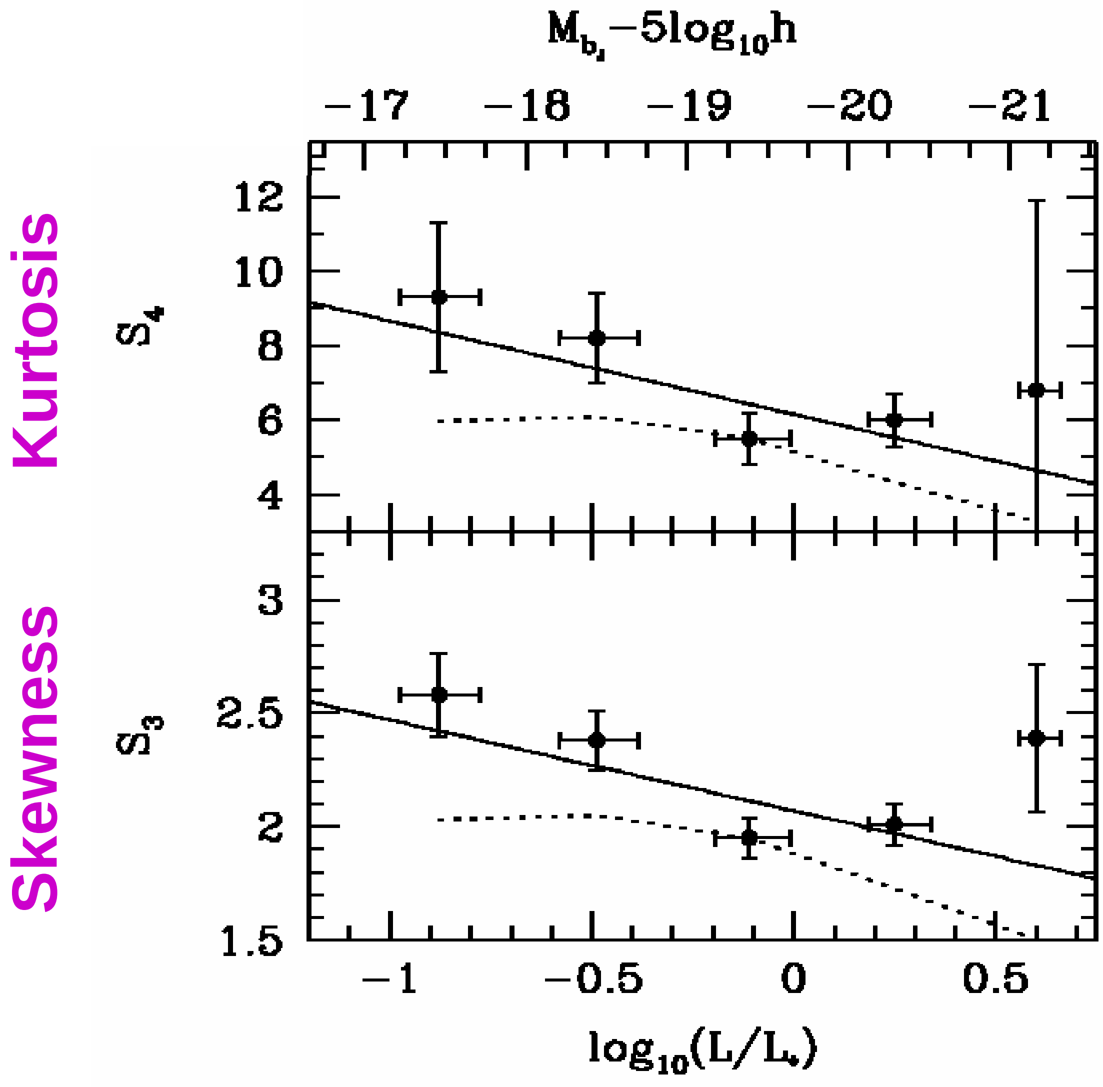
CPDF for R=10 Mpc/h





NB: the large difference in the CPDF for the 3 VLS is due to the change in the mean density only!

Some ‘evidence’ for luminosity dependent higher order clustering



Conclusions

- 2dFGRS galaxies follow a hierarchical clustering model up to 6th order in redshift space (in each luminosity bin).
- Hierarchical amplitudes are *approximately* independent of cell radius used to smooth the galaxy distribution on intermediate scales (~2 < R/Mpc/h < ~7).
- Presence of rare and extreme superstructures in the galaxy distribution can **strongly influence** results on larger scales (R > ~7 Mpc/h).
- Skewness (S_3) and Kurtosis (S_4) show both a *weak linear* dependence on log luminosity.
- Further details in astro-ph/0401405 (Baugh etal) and in astro-ph/0401434 (Croton etal).

Summary of Results

Table 1. Properties of the combined 2dFGRS SGP and NGP volume-limited catalogues (VLCs). Column 1 gives the numerical label of the sample. Columns 2 and 3 give the faint and bright absolute magnitude limits that define the sample. The fourth column gives the median luminosity of each volume limited sample in units of L_*, computed using the Schechter function parameters quoted by Norberg et al. (2002b). Columns 5, 6 and 7 give the number of galaxies, the mean number density and the mean inter-galaxy separation for each VLC, respectively. Columns 8 and 9 state the redshift boundaries of each sample for the nominal apparent magnitude limits of the survey; columns 10 and 11 give the corresponding comoving distances. Finally, column 12 gives the combined SGP and NGP volume. All distances are comoving and are calculated assuming standard cosmological parameters ($\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$).

VLC ID	Mag. range $M_{b_J} - 5\log_{10} h$		Median lum. L/L_*	N_G	ρ_{ave} $10^{-3}h^3\mathrm{Mpc}^{-3}$	d_{mean} $h^{-1}\mathrm{Mpc}$	z_{min}	z_{max}	D_{min} $h^{-1}\mathrm{Mpc}$	D_{max} $h^{-1}\mathrm{Mpc}$	Volume $10^6 h^{-3}\mathrm{Mpc}^3$
1	−17.0	−18.0	0.13	8038	10.9	4.51	0.009	0.058	24.8	169.9	0.74
2	−18.0	−19.0	0.33	23290	9.26	4.76	0.014	0.088	39.0	255.6	2.52
3	−19.0	−20.0	0.78	44931	5.64	5.62	0.021	0.130	61.1	375.6	7.97
4	−20.0	−21.0	1.78	33997	1.46	8.82	0.033	0.188	95.1	537.2	23.3
5	−21.0	−22.0	3.98	6895	0.110	20.9	0.050	0.266	146.4	747.9	62.8

Table 2. The best fit values and 2−σ error ($\Delta\chi^2 = 4$) for S_p (columns 4 to 7). The range of scales used in the fits is given in columns 2 and 3. The number in brackets after each error gives the reduced χ^2 value for the fit, using the number of degrees of freedom derived from the principal component analysis. The last two columns give the relative linear bias, b_r (defined by Eq. 14) and the second order bias term, c'_2 (defined by Eq. 17). Errors are 2−σ. The reference sample is sample number 3. These values are obtained for the full volume limited samples. A blank entry indicates that a reliable measurement of the particular hierarchical amplitude was not possible for the sample in question.

VLC ID	R_{min} $h^{-1}\mathrm{Mpc}$	R_{max} $h^{-1}\mathrm{Mpc}$	S_3	S_4	S_5	S_6	b_r	c'_2
1	0.71	7.1	2.58 ± 0.37 (0.1)	9.3 ± 4.0 (0.1)	34 ± 32 (0.1)	− − −	0.96 ± 0.16 (0.1)	0.17 ± 0.25 (0.1)
2	0.71	7.1	2.38 ± 0.25 (0.1)	8.2 ± 2.3 (0.9)	36 ± 20 (0.4)	185 ± 170 (0.1)	0.96 ± 0.08 (0.3)	0.11 ± 0.13 (0.1)
3	0.71	7.1	1.95 ± 0.18 (6.1)	5.5 ± 1.4 (2.3)	18 ± 11 (1.9)	46 ± 50 (1.1)	1	0
4	0.80	8.9	2.01 ± 0.17 (1.2)	6.0 ± 1.5 (0.6)	22 ± 12 (0.4)	71 ± 80 (0.3)	1.13 ± 0.06 (2.8)	0.10 ± 0.08 (0.3)
5	2.2	11.2	2.39 ± 0.63 (0.5)	6.8 ± 7.0 (0.4)	− − −	− − −	1.30 ± 0.14 (0.9)	0.33 ± 0.31 (0.5)

Content: Theoretical Background

- Dynamics of Gravitational Instability
- Random cosmic fields & their statistical description
- From dynamics to statistics:
  - N-point results
  - local cosmic fields
- Estimators and Errors:
  - Correlation functions: 2-point to N-point
  - Counts-in-Cells (CiC)

Statistical Homogeneity & Isotropy

- Definitions:
  - Random field is statistically homogeneous if all multipoint probability distribution functions p(\d(x_1),..,\d(x_N)) or its moments remain the same under translation of the coordinates (x_1,..x_N) in space. => probabilities depend only on the relative positions.
  - Random field is statistically isotropic if p(\d(x_1),..,\d(x_N)) is invariant under spatial rotations.
  - We assume that cosmic fields are statistically homogeneous and isotropic. This is a valid assumption for most cosmological theories. However, redshift space distortions in galaxy redshift surveys introduce significant deviations from statistical isotropy and homogeneity in the redshift space density field.

Non-zero higher order moments

For a Gaussian distribution of density fluctuations, the volume averaged reduced p-point correlation functions, x_p , are identically zero for p>2; the density field is completely described by the variance (p=2). The evolution of an initially Gaussian density field due to gravitational instability generates non-zero x_p . A basic test of the gravitational origin of the higher order moments is to determine their relation to the variance of the distribution.

Incompleteness corrections

- Geometric incompleteness (ie. fraction of cell volume sampling a region outside the 2dFGRS survey volume)
- Spectroscopic incompleteness (ie. fraction of galaxies actually observed in the line of sight of the cell volume)