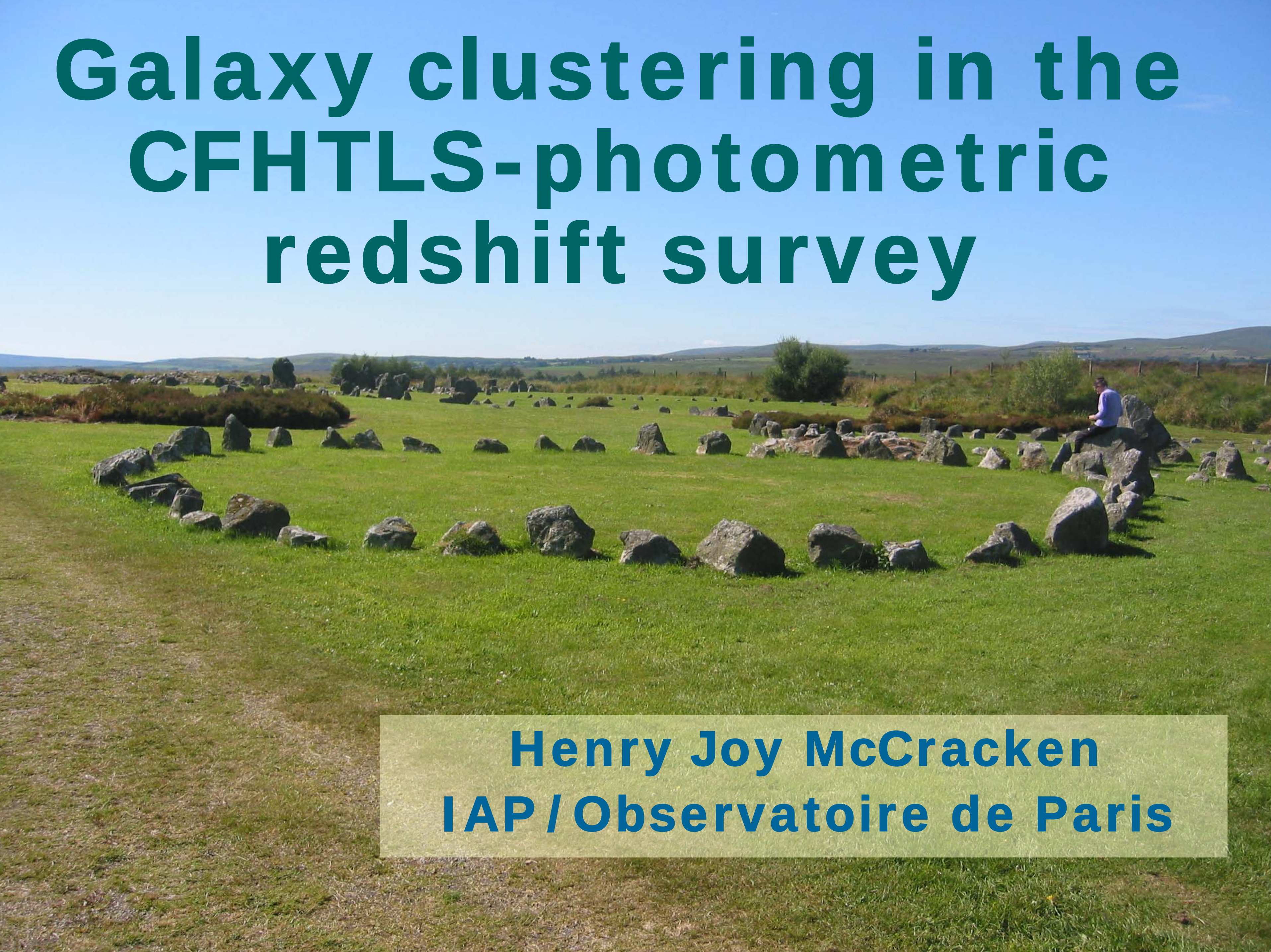
Galaxy clustering in the CFHTLS-photometric redshift survey



Henry Joy McCracken
IAP/Observatoire de Paris

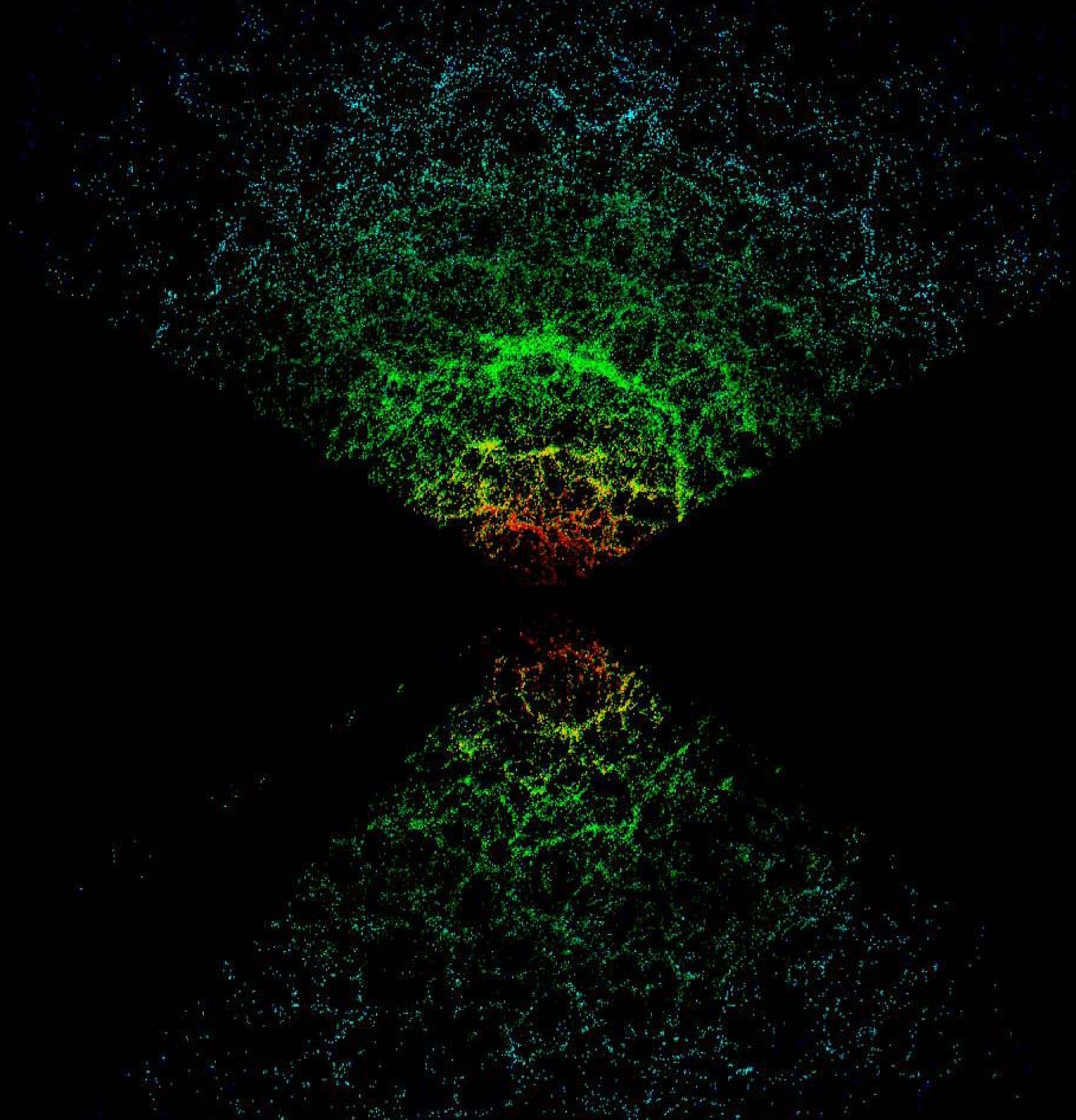
Here's the summary!

- We present in detail a new, very large *photometric* redshift survey of the distant universe, comprising 250,000 galaxies extracted from the four Canada-France-Legacy Survey Fields
- We present initial results from the clustering of galaxies as a function of intrinsic luminosity and type out to $z \sim 1$

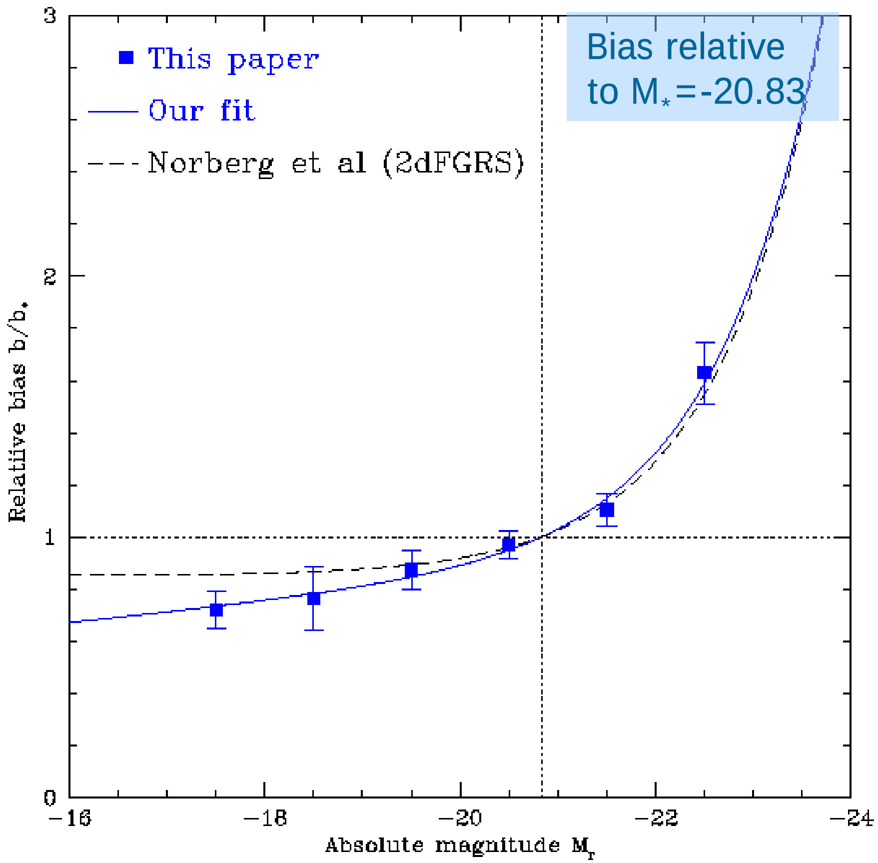
Growth of structure in the Universe

- Structures in the Universe grow from tiny fluctuations in the CMB under the influence of gravitational instabilities
- The **mass correlation function** depends on our choice of cosmology (now known!) and the power spectrum of the initial fluctuations
- We can trace the hierarchical evolution of the dark matter component on large scales using large numerical simulations or analytic theory
- **But what about the galaxies?**
- 'b' is our 'ignorance' parameter...

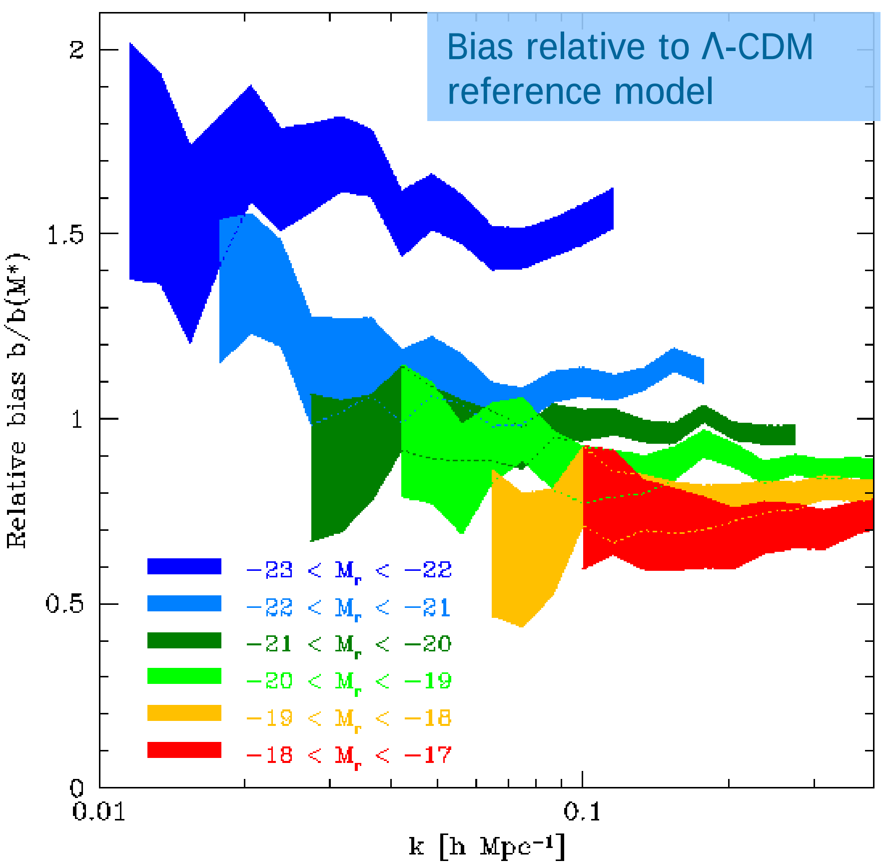
$$\xi_2^{gal} = b^2 \xi^{DM}$$



Galaxy biasing at z~0.1

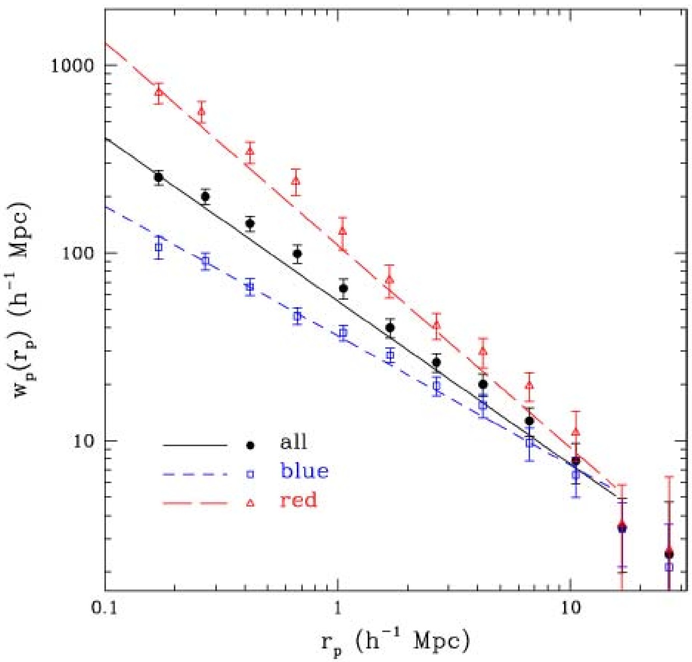


Tegmark et al 2003

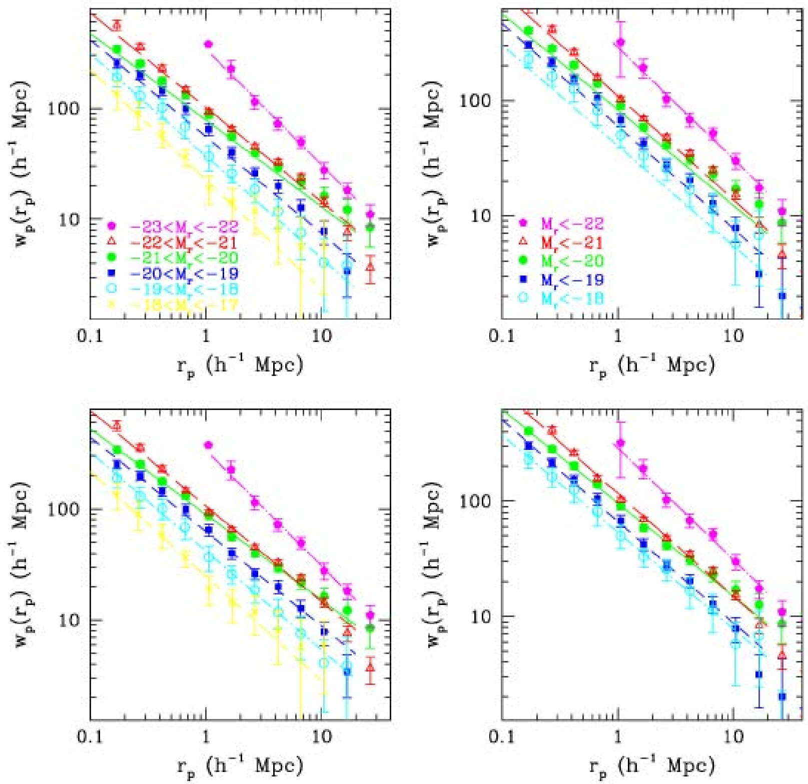


Tegmark et al 2003

Type-dependent clustering

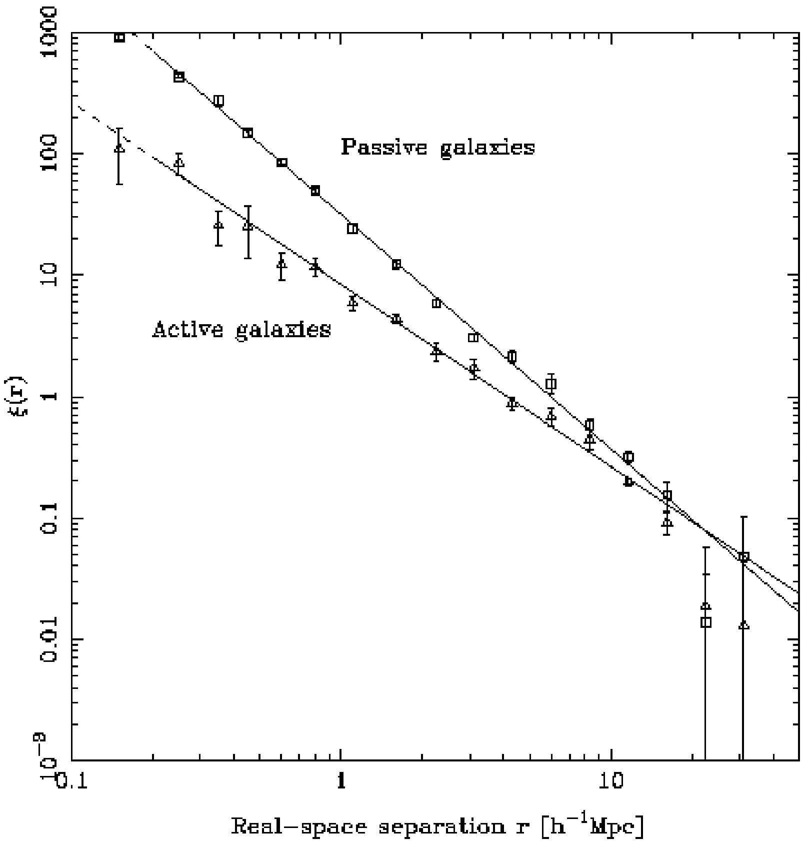


Zehavi et al. 2004

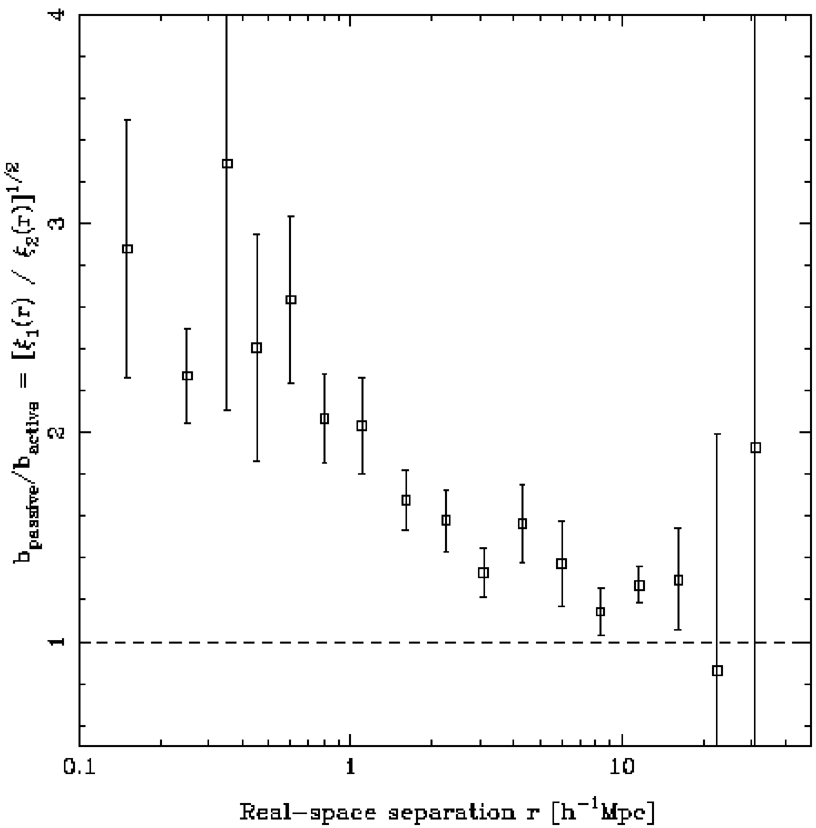


Zehavi et al 2004

...and from the 2df



Madgwick et al 2003



Madgwick et al 2003

Summary of low redshift measurements

- At low redshifts, the clustering amplitude depends on type, luminosity and colour
- For faint galaxies ($L<L^*$), the dependency is weak
- There is some evidence that the correlation function slope for red galaxies is different than for blue galaxies
- Clustering amplitudes follow very well the hierarchical scaling relation

Some outstanding questions:

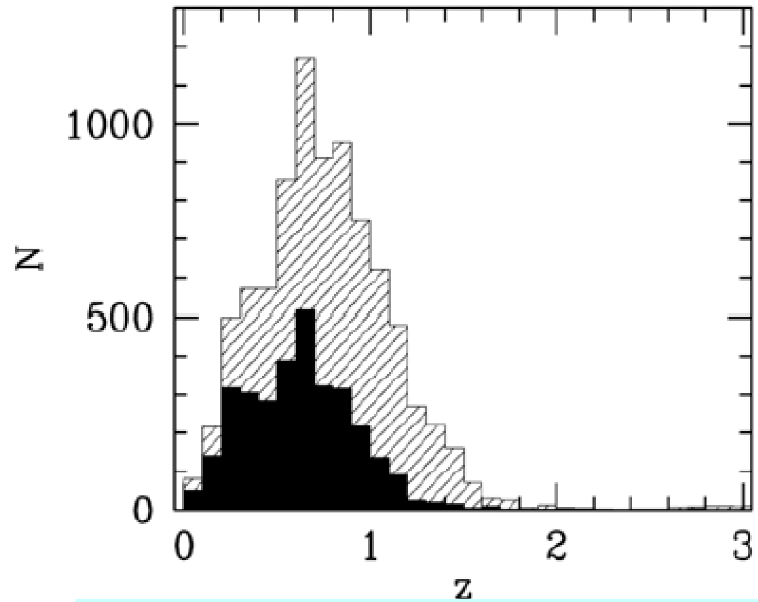
- How are luminous objects distributed at intermediate to high redshift?
- How is the clustering of luminous matter related to the underlying dark matter density field?
- What is the occupation function for dark matter haloes? Or: how many galaxies per dark-matter halo?

Finding out about the distant Universe

- We would like to survey a volume of the universe large enough to be representative, **free from cosmic variance effects**
- We would like to have enough objects so that galaxy properties can be investigated in terms of environment and type

How to survey the distant Universe?

- Can make a pre-selection in colour-colour space: this can very complicated – and you only get only what you look for!
- Can just make a magnitude limited sample: very time-consuming but provides the most unbiased sample
- Best solution is to couple spectroscopic redshifts with photometric ones!
- Compute z_phot for all objects and calibrate them using a large base of high-quality spectra
- With large-format array detectors like MEGACAM it's possible

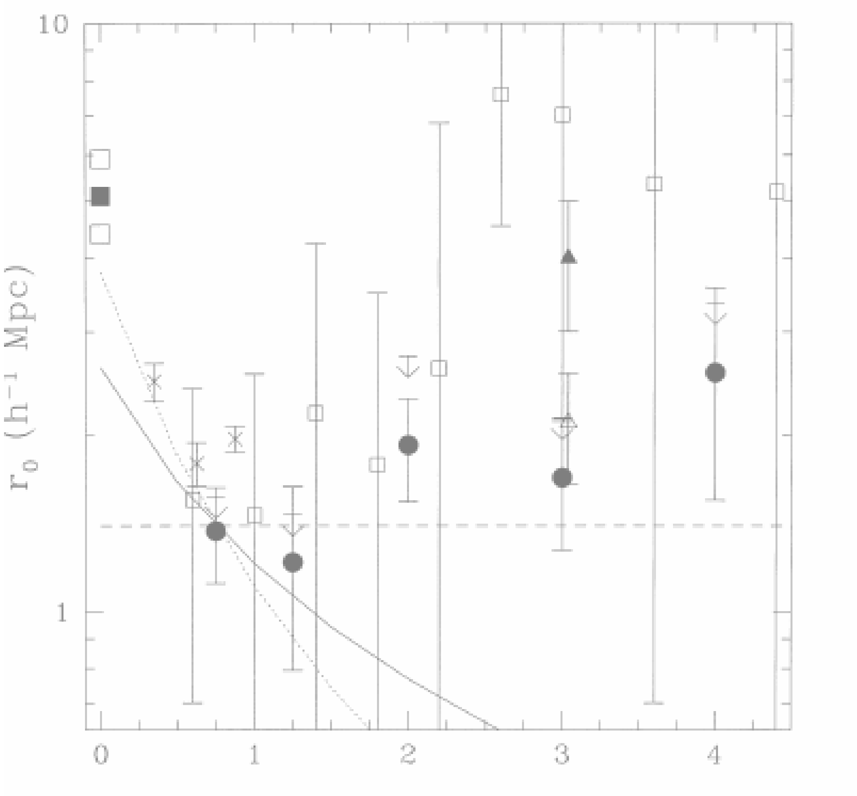


VLT-VIRMOS deep survey, Le Fevre et al 2005

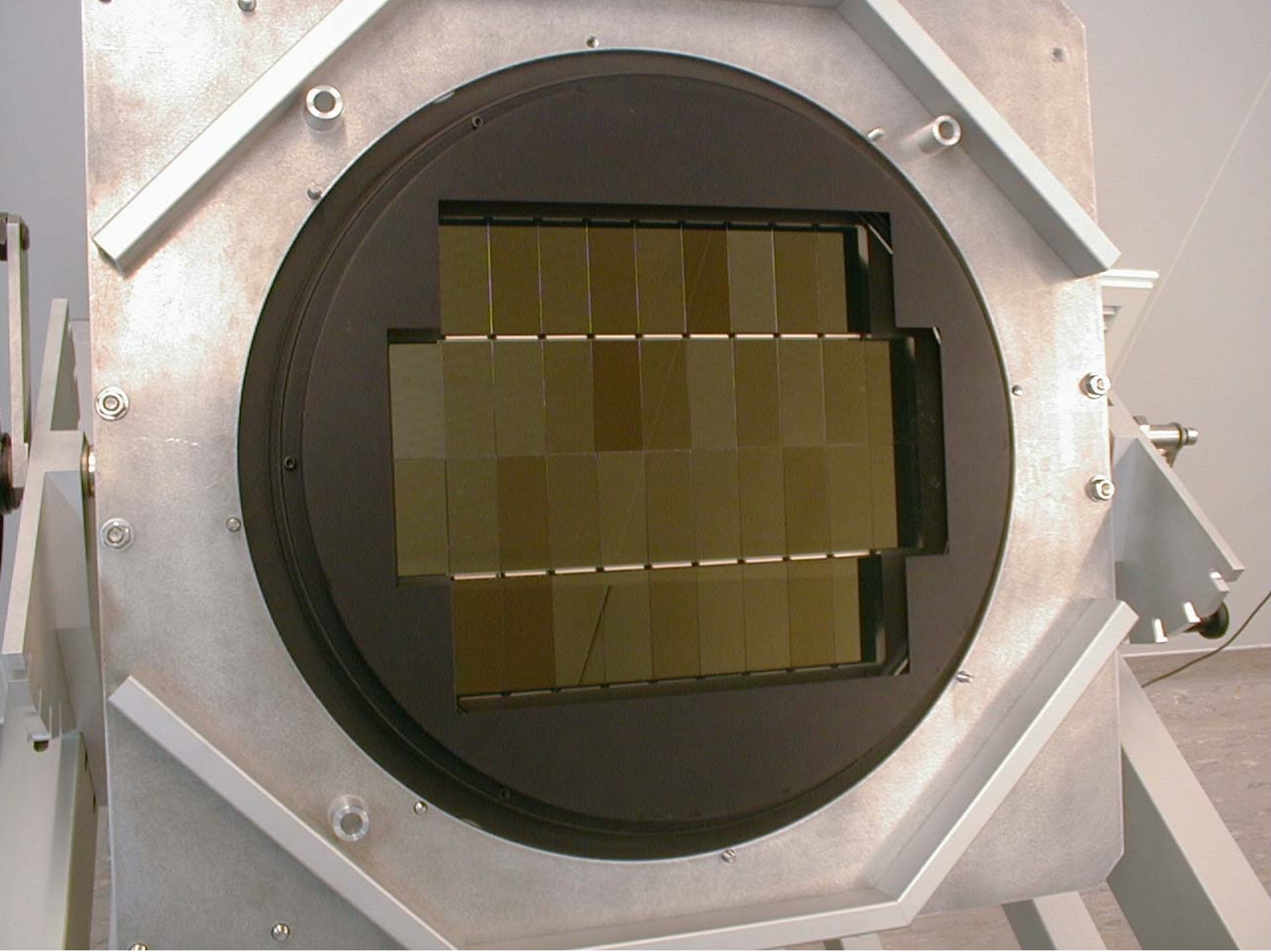
$$U_n - G < G - \mathcal{R} + 0.2.$$

Deep photometric surveys

- Previous “pencil beam” deep surveys have suffered badly from cosmic variance and shot noise effects
- Very small volumes probed at low redshifts meant was difficult to connect high-redshift studies to local analogues like the SDSS
- Low numbers of galaxies make it impossible to make divisions by type or luminosity
- Large-area ground based surveys were plagued by calibration issues or insufficient depth



Arnouts et al 1999, HDF-N, 4 arcmin^2

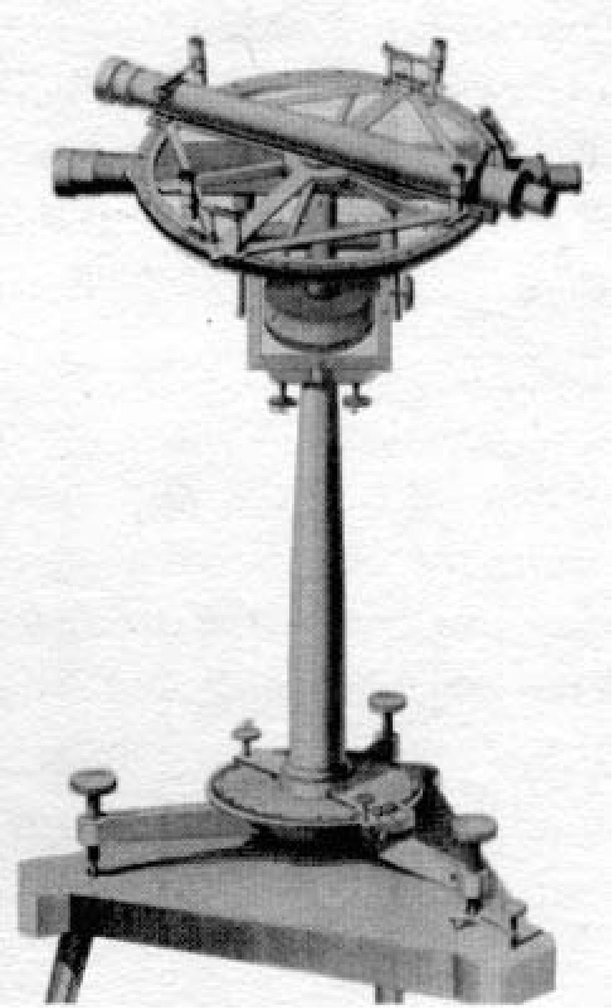


The Canada-France Legacy Survey

- **500 nights @CFHT: SDSS for the high-redshift Universe**
- Four deep (each comprising a megacam pointings) and three wide fields each in five filters
- **Median redshift of cfhtls galaxies is around one!**
- Several of the deep fields contain regions where there are or will be a very large number of spectroscopic redshifts and large ancillary datasets (COSMOS, VIRMOS-F02)
- Only excellent quality images are accepted
- Data processing and stacking carried out at TERAPIX, a special data centre designed to cope with the massive flow of data from MEGACAM

We would like to preserve the quality of “artisinale” (*fatto in casa*!) reductions whilst duplicating them on a very large scale...

It’s important to understand the difference between systematic and random errors…



… as systematic errors will limit the successful scientific exploitation of any large survey, where they are the dominant type of errors!

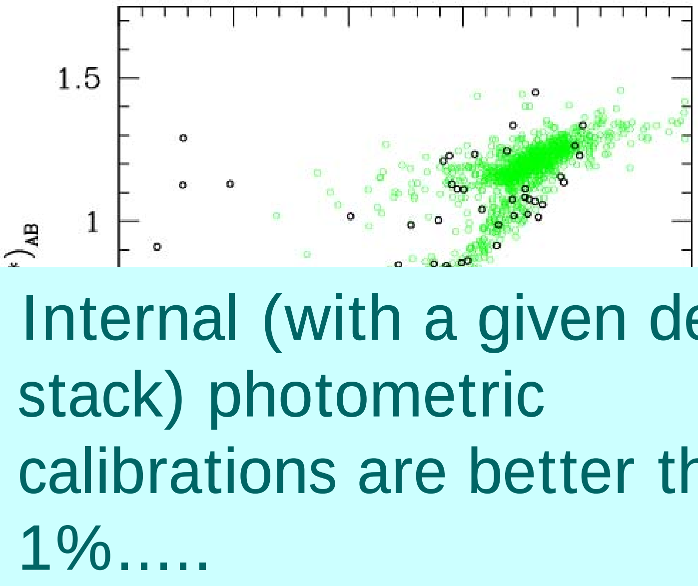
TERAPIX and CFHTLS

- TERAPIX is a data processing centre at the IAP dedicated to handling data from large-format array cameras
- For the moment almost all the data at TERAPIX comes from MEGACAM, although WIRCAM data is arriving
- TERAPIX produces catalogues and data products derived for the CFHTLS
- TERAPIX handles the total reduction chain from flat-fielded images to catalogue generation
- TERAPIX receives flat-fielded images from CFHT with a chip-to-chip photometric accuracy of 1%

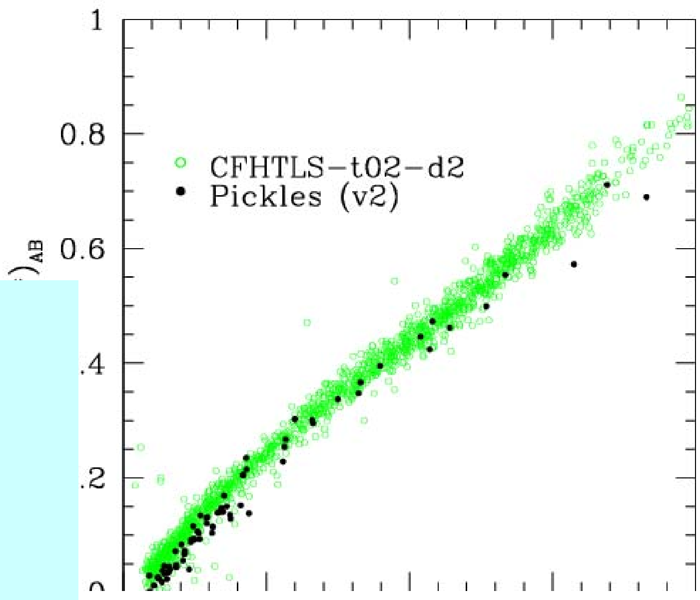
- Very large amount of available computational power (more than 1000 Gflops) and disk (100TB) means that any aspect of the reduction can be quickly re-done, based on any problems encountered in the scientific analysis.
- Data processing and scientific analysis are closely coupled

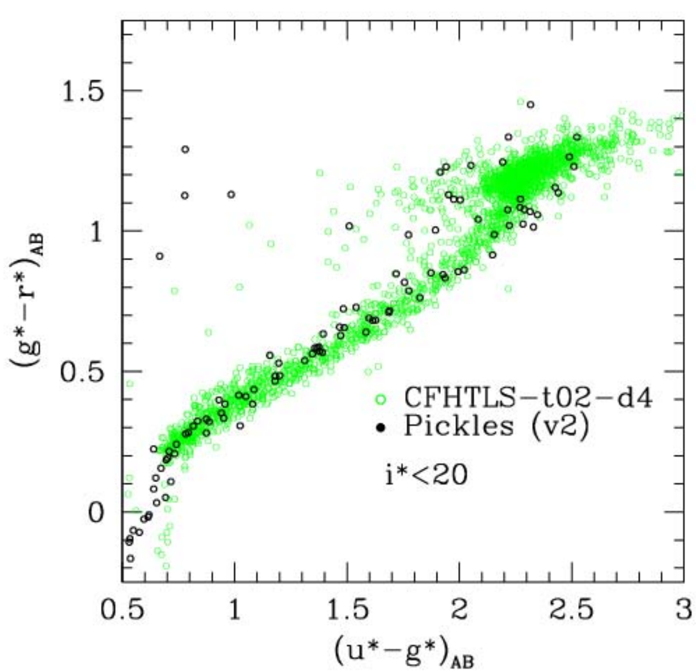
The CFHTLS-t02 deep stacks

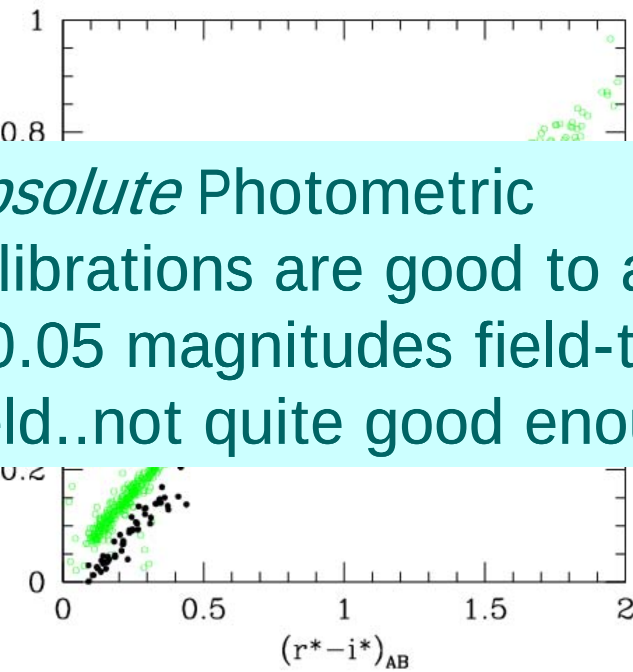
- All data between taken between June 2003 and December 2004
- Only images with seeing better than 1.1''
- Four independent fields each of which has an effective area of 0.8 deg^2 after masking
- Coverage in five broad band filters (*ugriz*), reaching approximately AB~26 in all bands
- Data released publicly to the French and Canadian communities – see the CADC/TERAPIX web sites



Internal (with a given deep stack) photometric calibrations are better than 1%.....



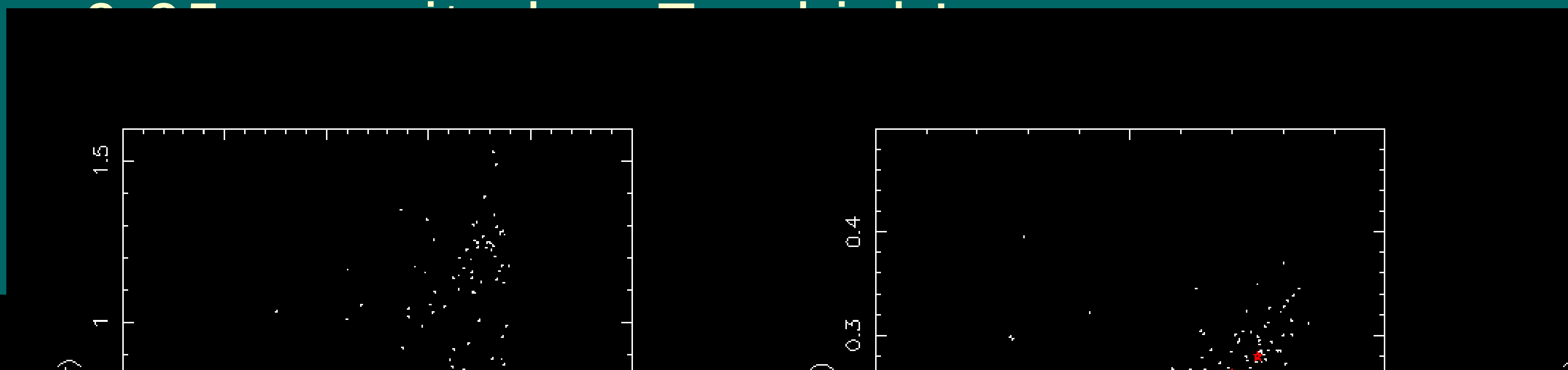




Absolute Photometric calibrations are good to around ~0.05 magnitudes field-to-field..not quite good enough!

Photometric (re)calibrations

- Photometric calibrations “out of the box” have a systematic field-to-field dispersion of around



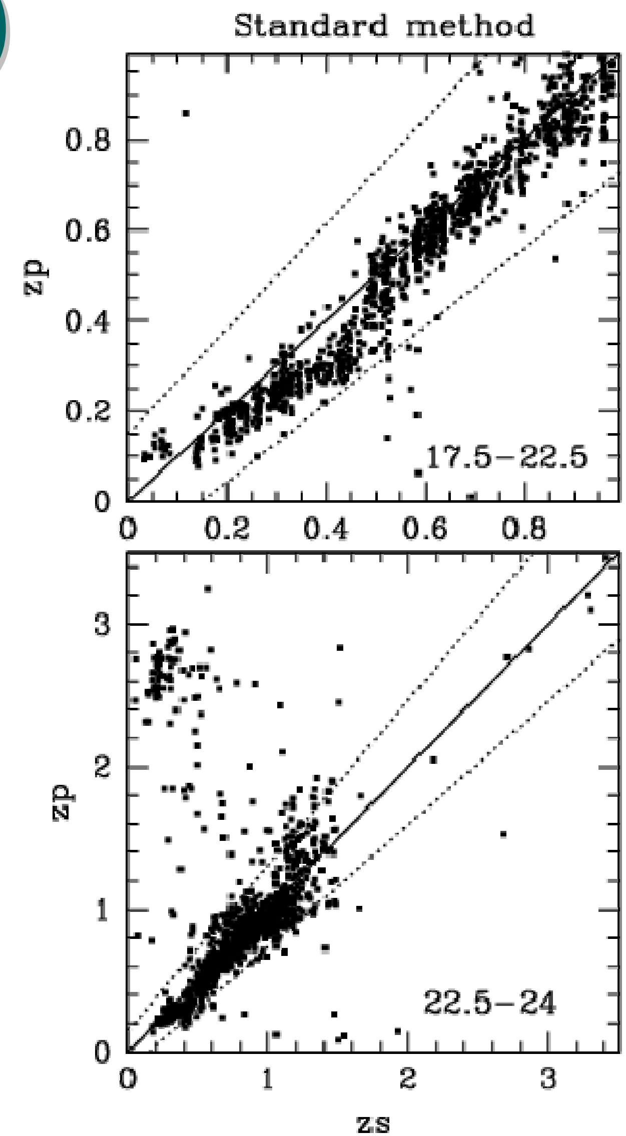
- Minimisations repeated over all four fields taking one field as the reference
- These offsets offsets applied and catalogues re-extracted
- Final catalogues have absolute calibrations at the ~0.025-0.01 magnitude level

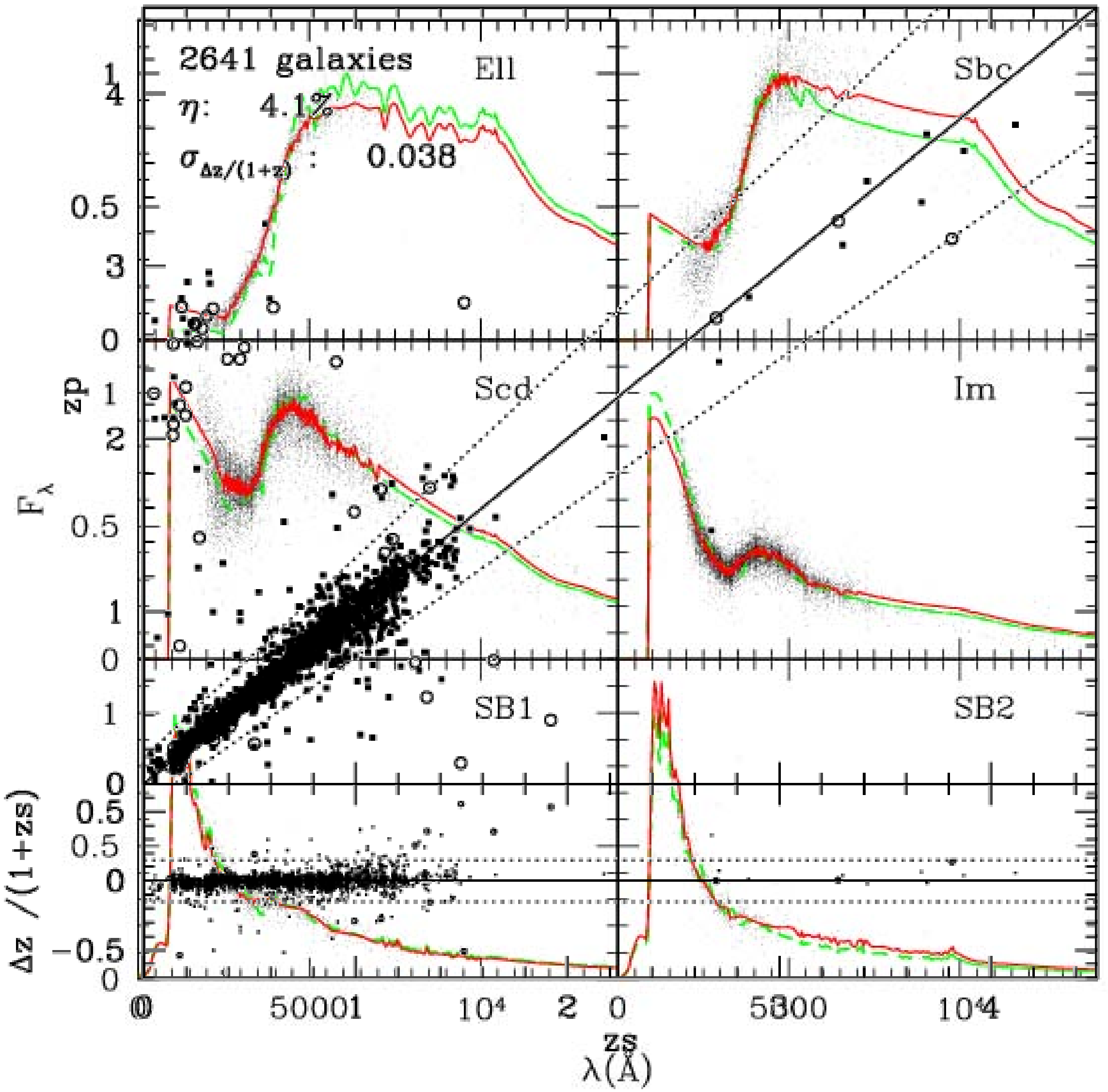
The CFHTLS/VVDS field

- Old cfh12k photometry re-swarped into the MEGACAM astrometric frame
- New catalogue constructed containing BVRI cfh12k photometry for all cfhtls-d1 objects and also matched with spectroscopic redshifts for 10,000 objects.
- Photometric redshifts computed using this combined catalogue
- This is the key field which is used to calibrate the photometric redshifts

Photo-zeds: "Le Phare" (Ilbert/Arnouts)

- Chi-squared fitting technique with the standard interpolated Coleman, Wu and Weedman templates (+ starburst type)
- Nasty systematics at low redshift!
- Many catastrophic errors!
- Photometric redshifts demand *precise knowledge* of the instrumental response function – we need to re-calibrate our templates

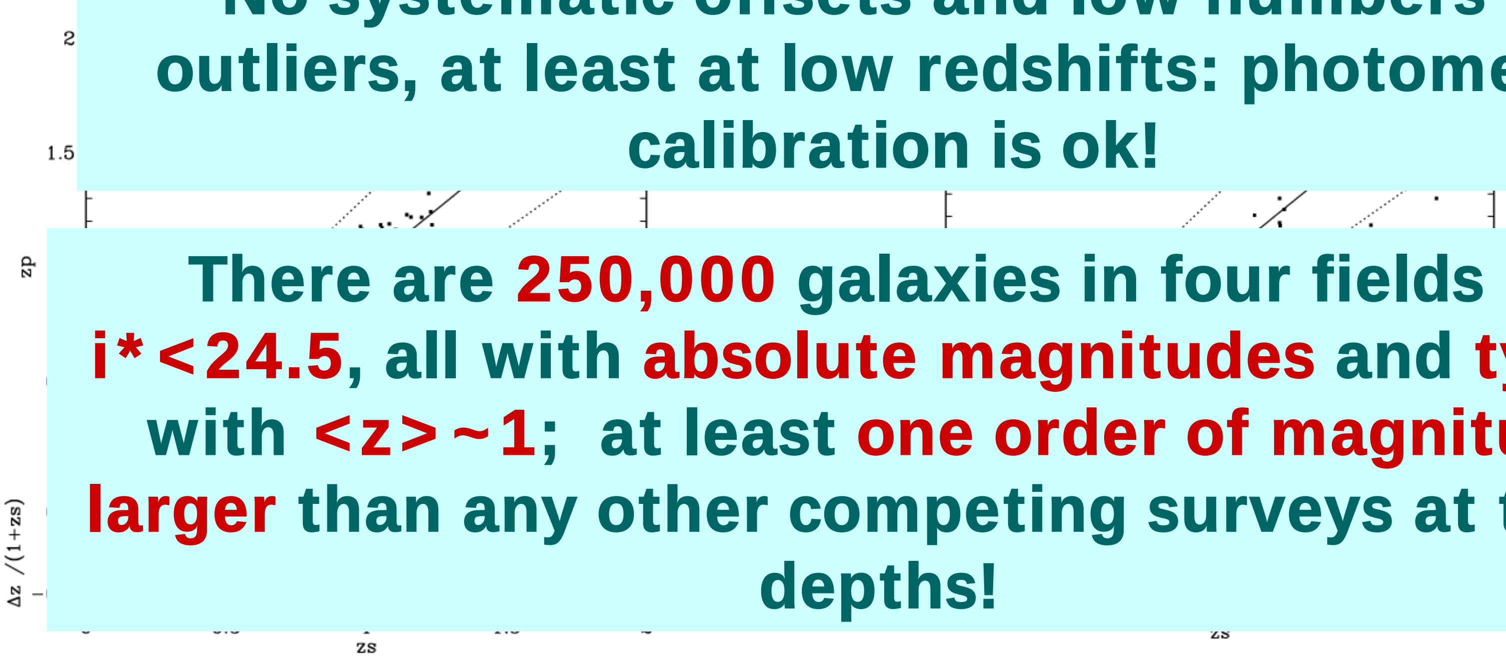




- Control sample of 468 galaxies with i*<21.5 and spectroscopic redshifts are used to produced “corrected” templates.
- These corrected templates produce much better photmetric redshifts with *no* systematic effects
- And also with a much smaller number of catastrophic outliers

Computing phot-zeds in the other CFHTLS deep fields

- In the d3 and d4 fields there are a small number of spectroscopic redshifts at lower redshifts from other surveys (SDSS, CFRS) which allow us to validate the templates derived from the cfhtls-d1 field



No systematic offsets and low numbers of outliers, at least at low redshifts: photometric calibration is ok!

There are 250,000 galaxies in four fields to i*<24.5, all with absolute magnitudes and types, with <z>~1; at least one order of magnitude larger than any other competing surveys at these depths!

Characterising the galaxy distribution

Method	Angular correlation function	Two point correlation function	Counts in cells
What it tells us	Projected excess of pairs with respect to a random distribution	Excess of pairs with respect to a random distribution	Moments of the galaxy distribution (variance, skewness...)
Advantages	Works with just positions: don't need redshifts	Easier to compare with theory	Very fast
Disadvantages	Need to know the source redshift distribution	Have to measure spectroscopic redshifts!	Can be difficult to interpret: higher-order moments depend sensitively on photometric errors

Computing the comoving correlation length-I

$$\omega(\theta) = \frac{H_0 H_\gamma}{c}\,\theta^{1-\gamma}\,\frac{\int_0^\infty N^2(z)\,r_0^\gamma(z)\,[x(z)]^{1-\gamma}\,E(z)\,F(z)\,dz}{[\int_0^\infty N(z)\,dz]^2}$$

Relativistic limber equation

$$\omega(\theta) = A_\omega \theta^{-\delta}$$

Assuming w(θ) is a power law...

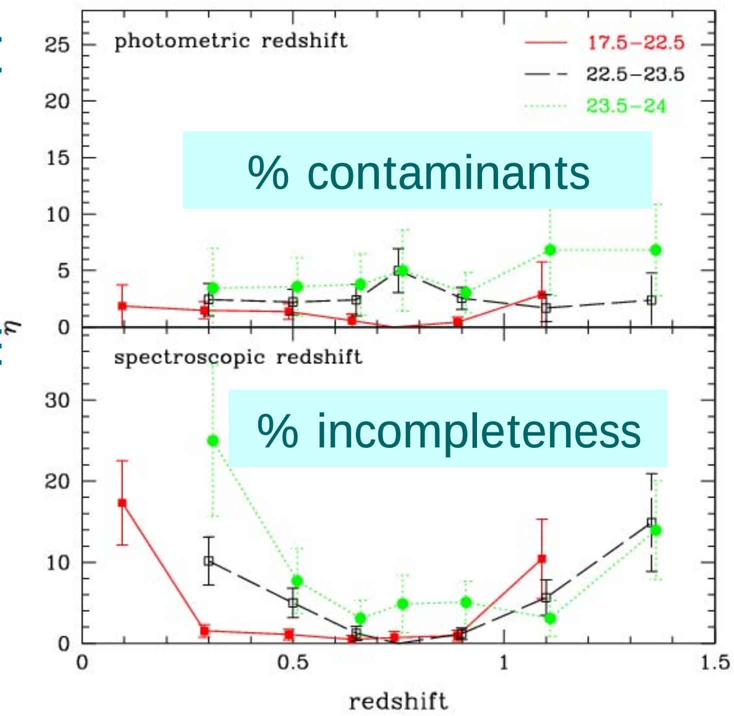
$$r_0^\gamma(z_{\rm eff}) = A_\omega \left[\frac{H_0 H_\gamma}{c}\,\frac{\int_{z_1}^{z_2} N^2(z)\,[x(z)]^{1-\gamma}\,E(z)\,dz}{[\int_{z_1}^{z_2} N(z)\,dz]^2}\right]^{-1}$$

$$\omega(\theta) = \frac{DD - 2DR + RR}{RR},$$

Which you get from computing pair counts on your catalogue....

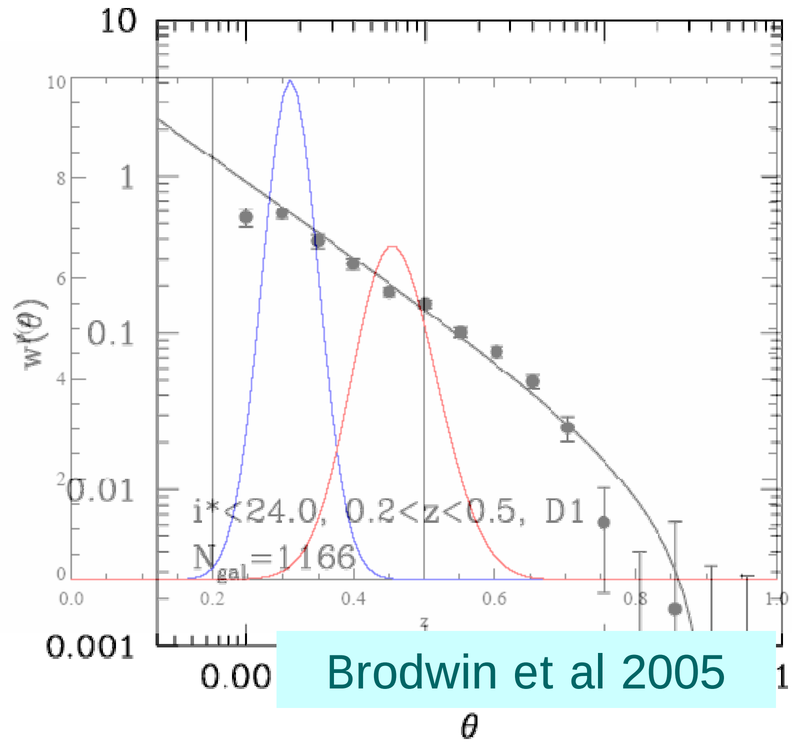
Computing comoving correlation lengths-II

- We compute the **projected correlation function w(θ)** fc each field and for each magnitude slice.
- We select galaxies in redshift slices corresponding the range where our photometric redshifts have the highest accuracy (lowest numbers of catastrophic outliers)
- For the moment, we consider galaxies with **0.2<z<1.2**



Computing the comoving correlation length-III

- For each galaxy in each redshift slice we compute the area under that galaxy's probability distribution function
- These areas are used as weights in the correlation function measurement
- This ensures that **all** information about the reliability of each photometric redshift is used
- The resulting measurements are then fitted with a power law with the appropriate finite-volume correction.

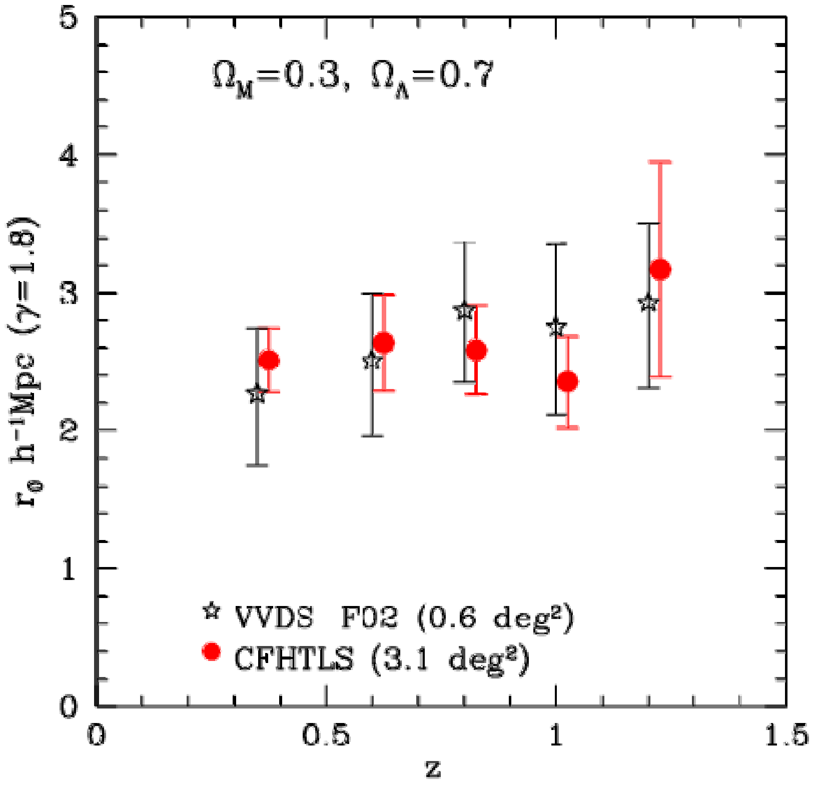


Brodwin et al 2005

$$\omega(\theta) = \frac{DD - 2DR + RR}{RR},$$

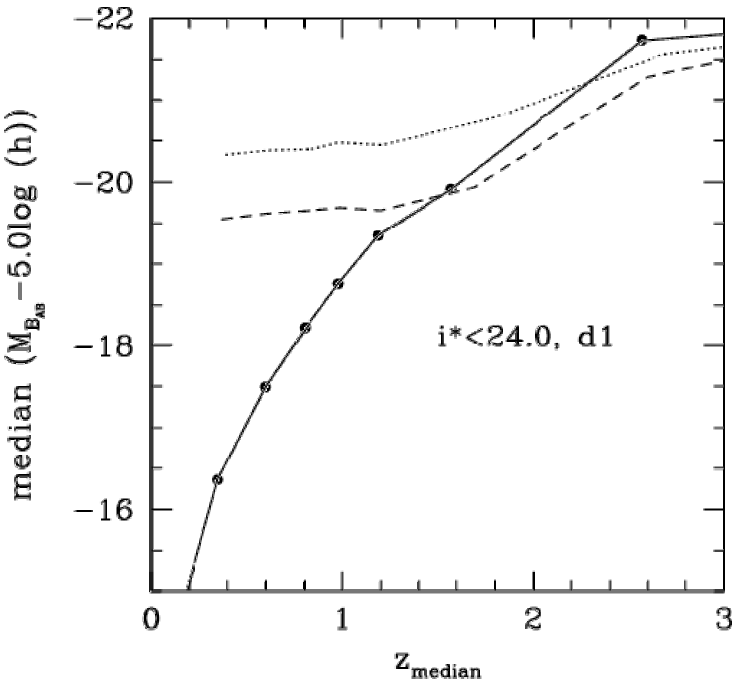
Comoving correlation length as a function of redshift

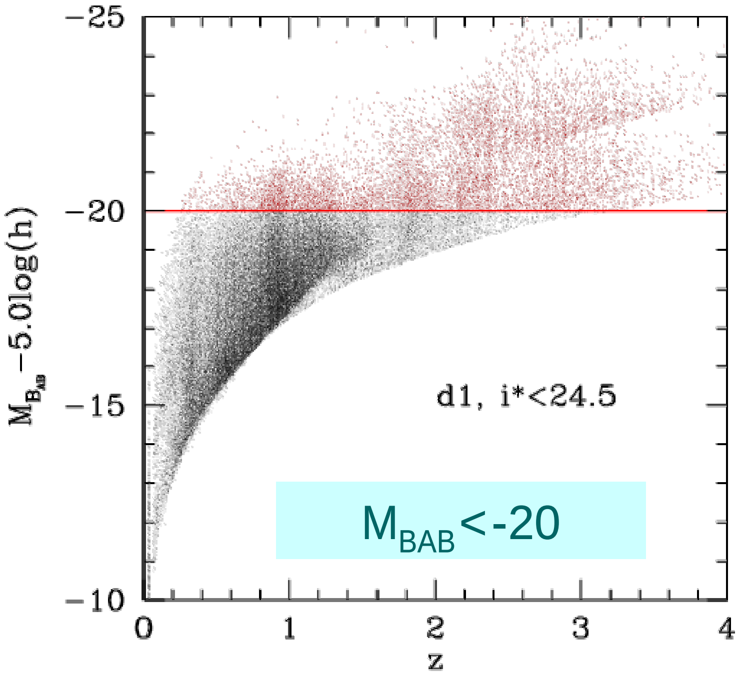
- We compute r_0 as a function for z for all four fields of the CFHTLS
- Error bars computed from the field-to-field variance – they are true "cosmic" error bars
- Remarkably good agreement with the VVDS spectroscopic survey measurements (which enclose one of the cfhtls survey fields)

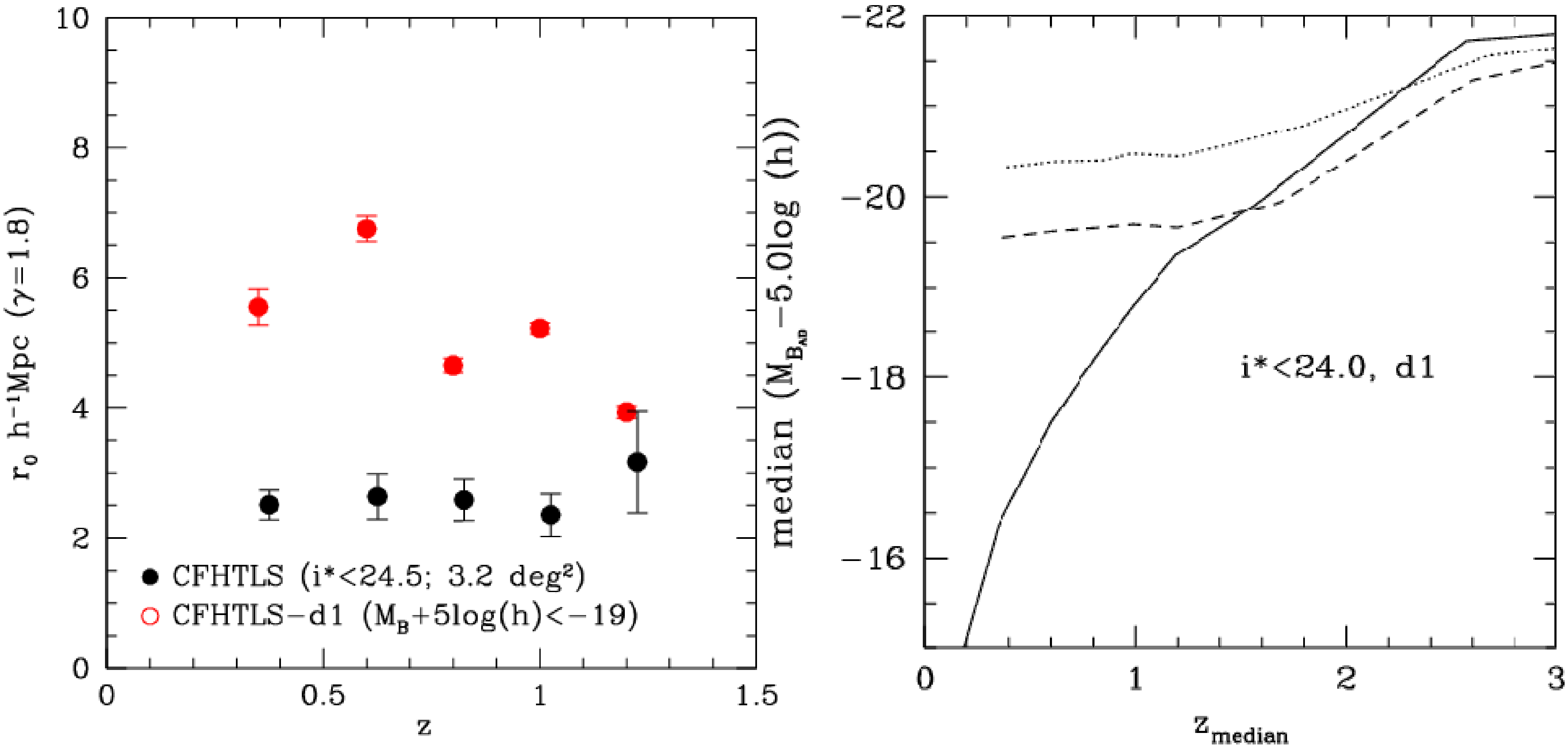


Luminosity limited samples

- Median luminosity in redshift slices is a strong function of redshift…
- Making luminosity-limited samples creates volume-limited samples



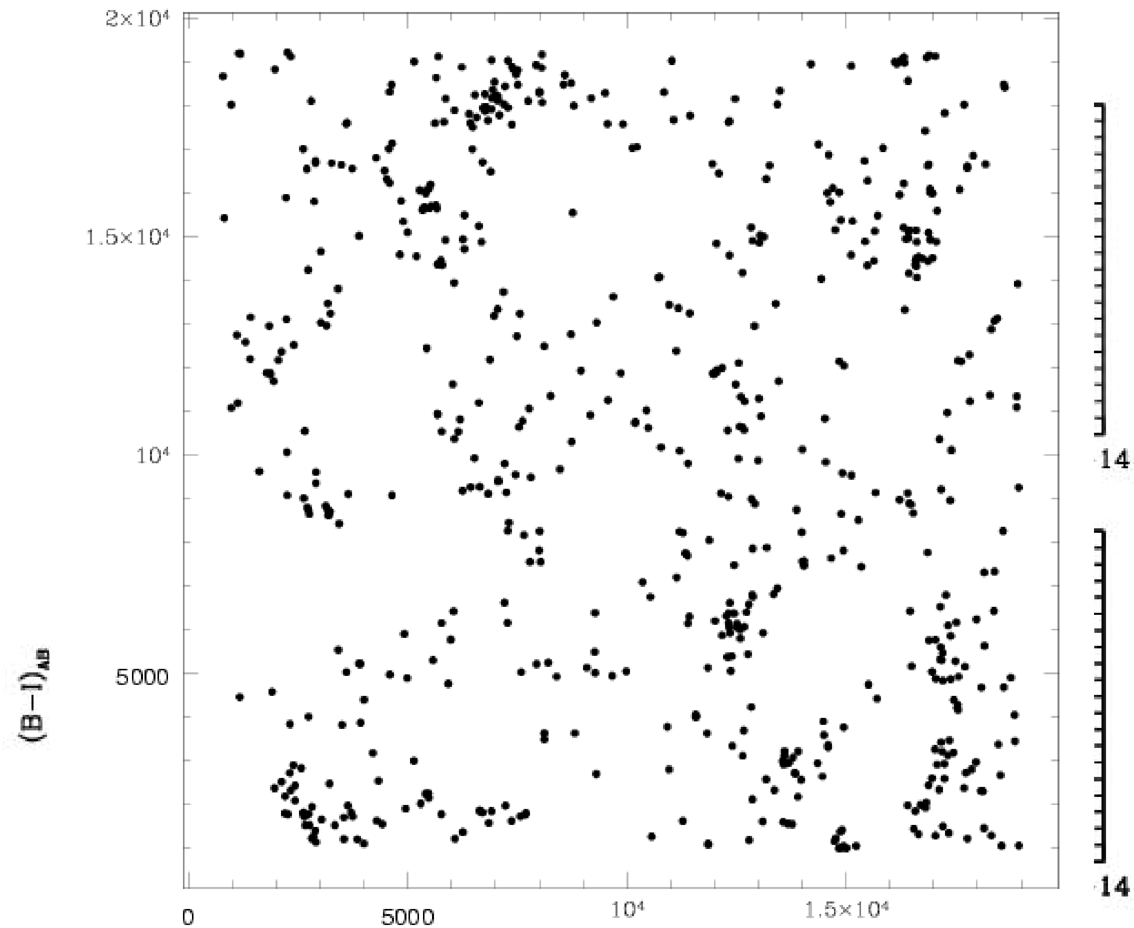




- Clustering amplitudes much higher than the magnitude limited sample, because the mean absolute magnitude is higher; bias depends on luminosity
- Does r0 decrease for these galaxies? (you might expect this if they were weakly biased…

Clustering by type to z~1

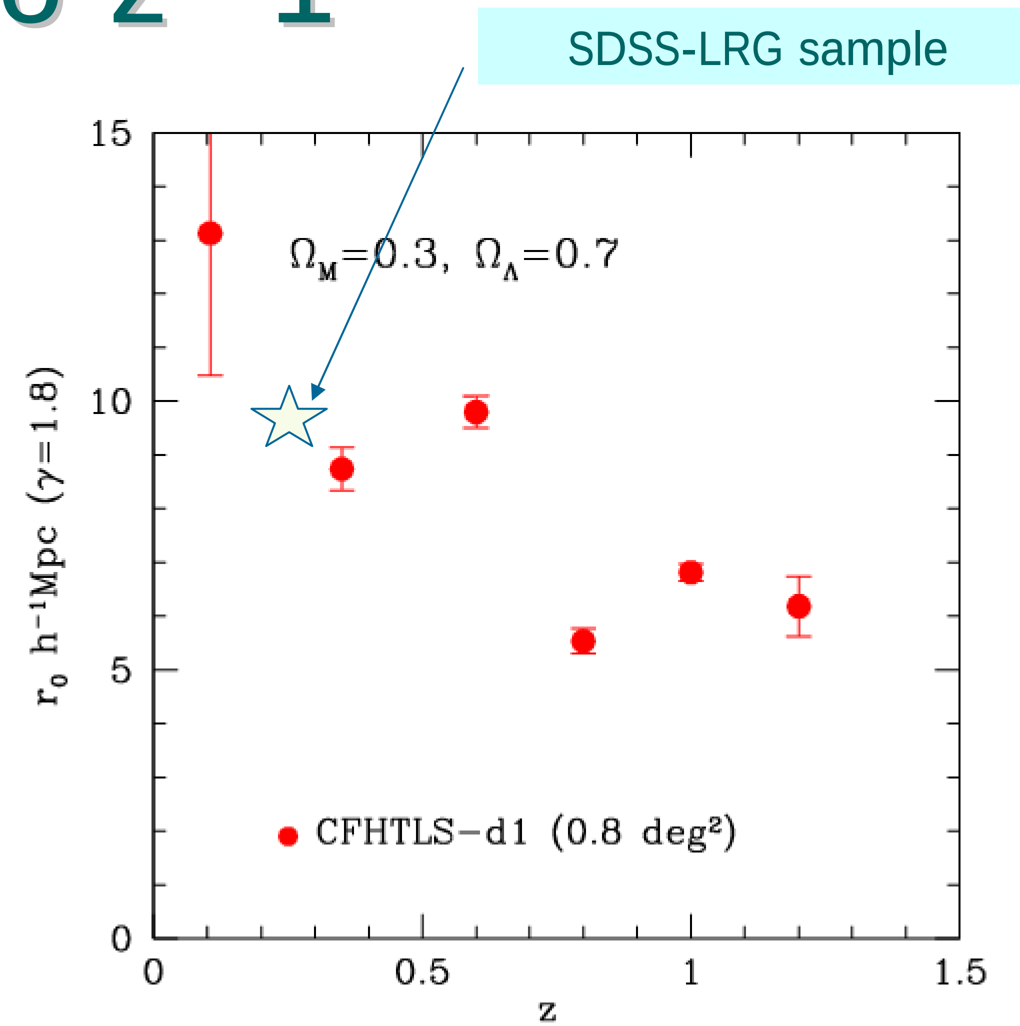
- What about the colour and type evolution of galaxies?
- Photometric redshift code provides *types* of best fitting templates
- These objects are **even more strongly clustered**



Elliptical galaxies in the cfhtls-d1 0.2<z<0.5 redshift slice

Clustering of early types to z~1

- Clustering of early types at z~1 is even higher than the luminosity limited samples at the same redshifts



What does it all mean?

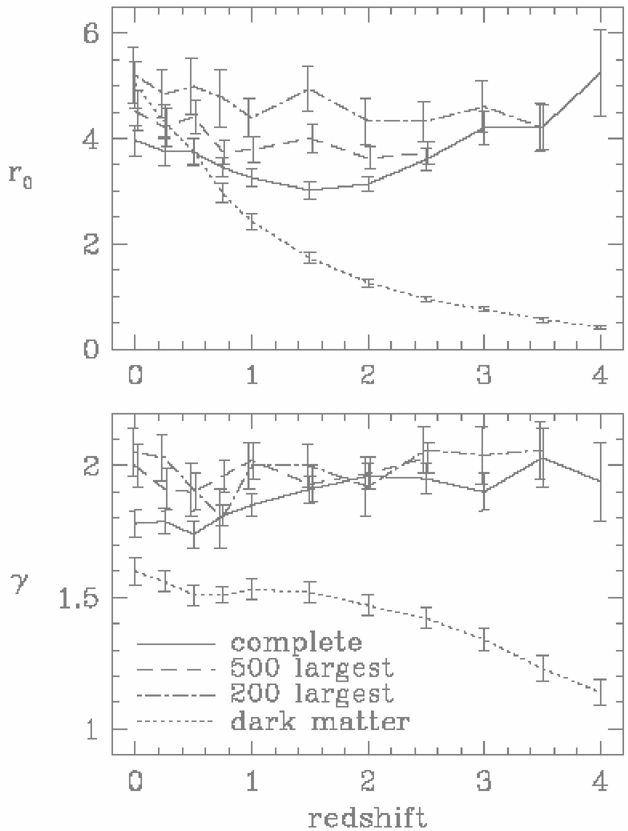


FIG. 5.—Evolution of the correlation length r_0 (in comoving h^{-1} Mpc) and power-law index γ, for all galaxies (*solid line*), the 500 most massive galaxies (*dashed line*), the 200 most massive galaxies (*dot-dashed line*), and the dark matter (*dotted line*). Error bars are obtained from the power-law fits, using the jackknife errors on $\xi(r)$. Lines for the 500 largest galaxies stop at $z = 2.5$, since the complete sample contains fewer than 500 galaxies at higher redshift; likewise, lines for the 200 largest galaxies stop at $z = 3.5$.

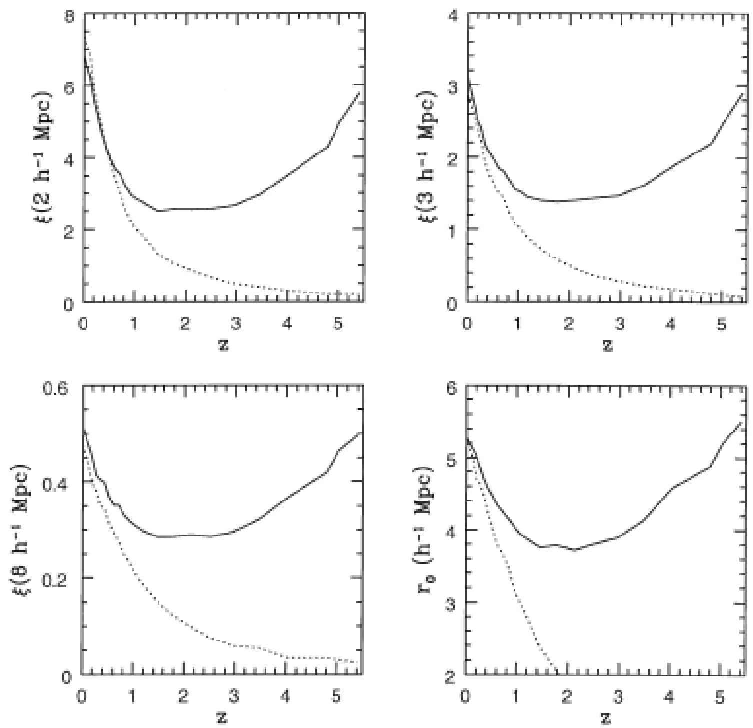
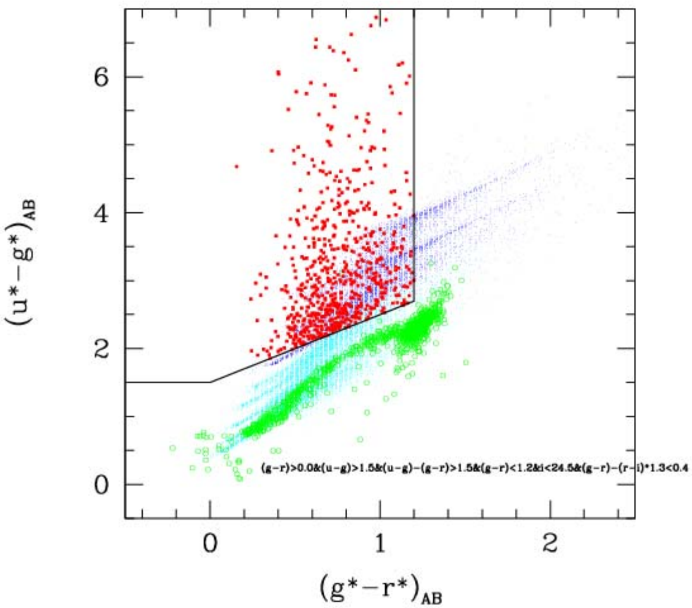


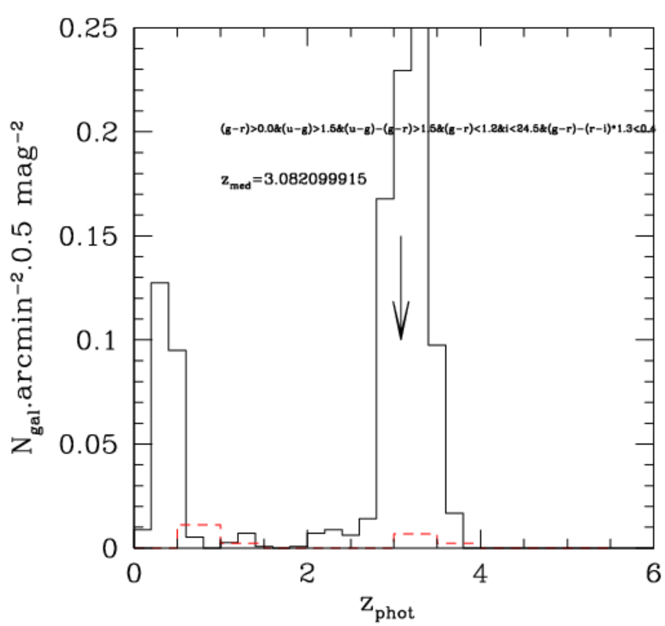
Figure 4. Evolution of clustering the the ΛCDM model. In the first three panels, the clustering amplitude is plotted against redshift for galaxies with rest frame *B*-band magnitudes brighter than $-19 + 5\log h$ (solid lines) and for the dark matter (dotted lines). Results are shown for $\xi(r)$ evaluated at $r = 2$, 3 and $8\,h^{-1}\,\mathrm{Mpc}^{-1}$. In the fourth panel, the comoving correlation length r_0 is plotted against redshift both for the galaxies and for the dark matter.

Kaufmann et al 1999

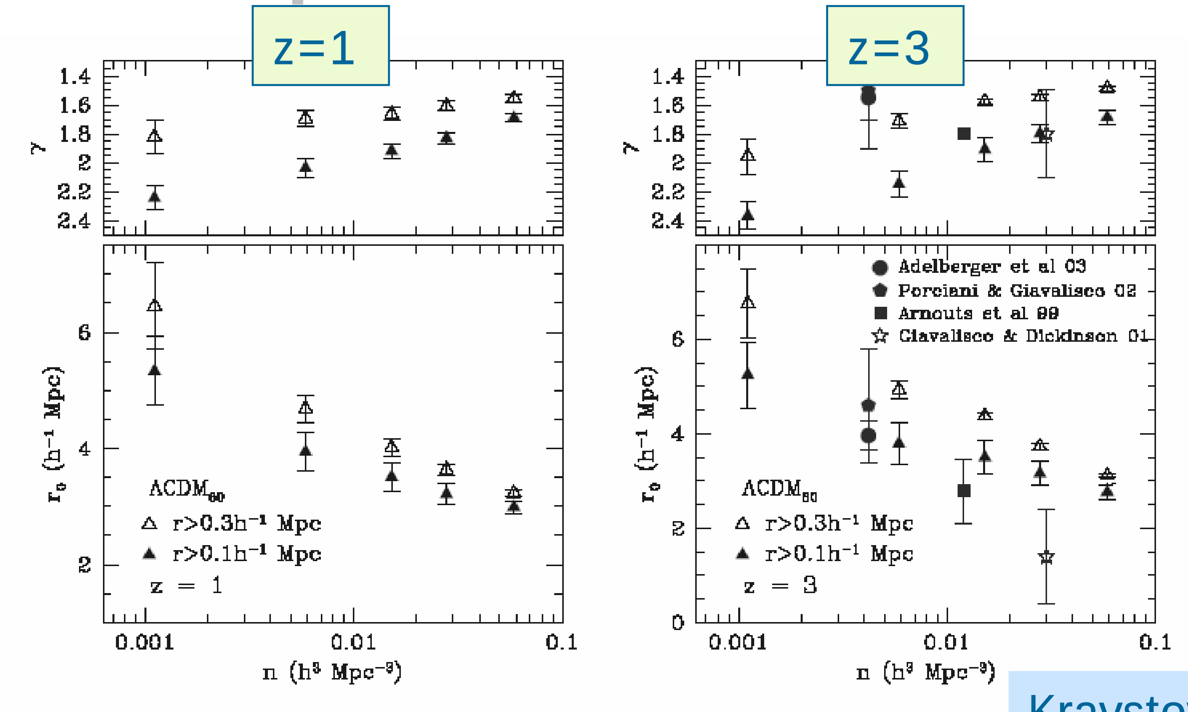
What's next: lyman-break galaxy samples

- There are several thousand z~4 and z~3 lyman-break galaxies in the CFHTLS survey fields...
- Megacam is very efficient in u*





Measuring the halo occupation function



Kravstov et al 2003

- Modelling the occupation function of dark matter haloes perhaps can provide some insight into how galaxies cluster at small separations where traditionally predictions are very difficult
- Does this explain the deviation from the power law behaviour seen for objects at z~4?
- We should be able to make a direct measurement of this quantity with the CFHTLS-zphot survey