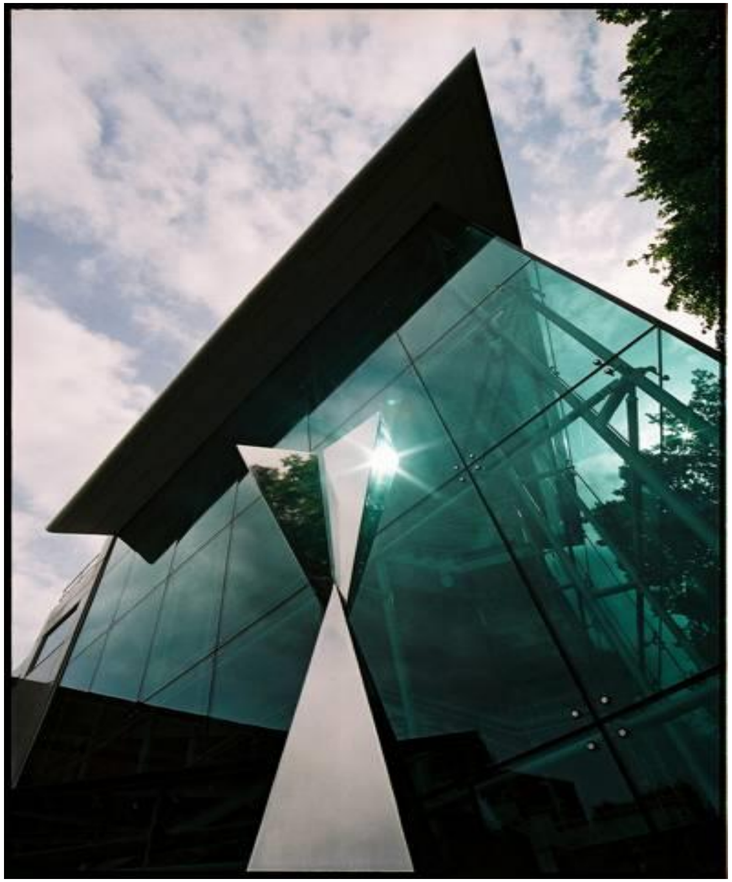
# The SHAM approach to modelling galaxy bias

Shaun Cole
(ICC, Durham University, UK)

Collaborators:
Vimal Simha (ICC, Durham)
Michael Schneider (LLNL)

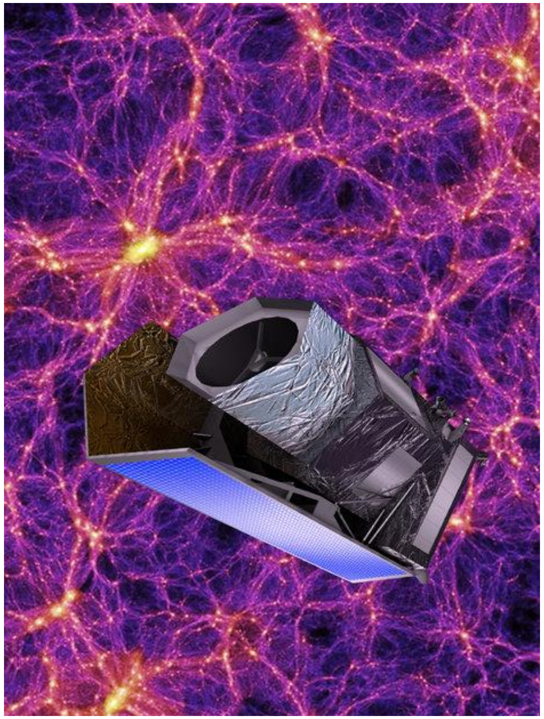
ICTP, 9th October 2013



# Outline

- Introduction
- What is SHAM
- Testing SHAM with SDSS
- Towards larger scales
- Summary

# Introduction



Euclid

- Goal: Accurate predictions of galaxy clustering as a function of cosmological parameters.
- Ideally we need to predict not only clustering, but also the expected covariance of an estimate from a given survey. At present this is computationally demanding even just for the mass distribution (brief aside).
- Here the focus is bias, the relationship between galaxy and mass clustering.

# Covariance (aside)

Schneider et al (2011, ApJ 737,11)

Running ensembles of "Euclid" scale N-body simulations to estimates covariance s such as $\langle \Delta P(k_1) \Delta P(k_2) \rangle$ is computationally expensive.
A single (or a few) large boxes are sufficient to fully sample the covariance of the small scale modes, but not the large scales modes.

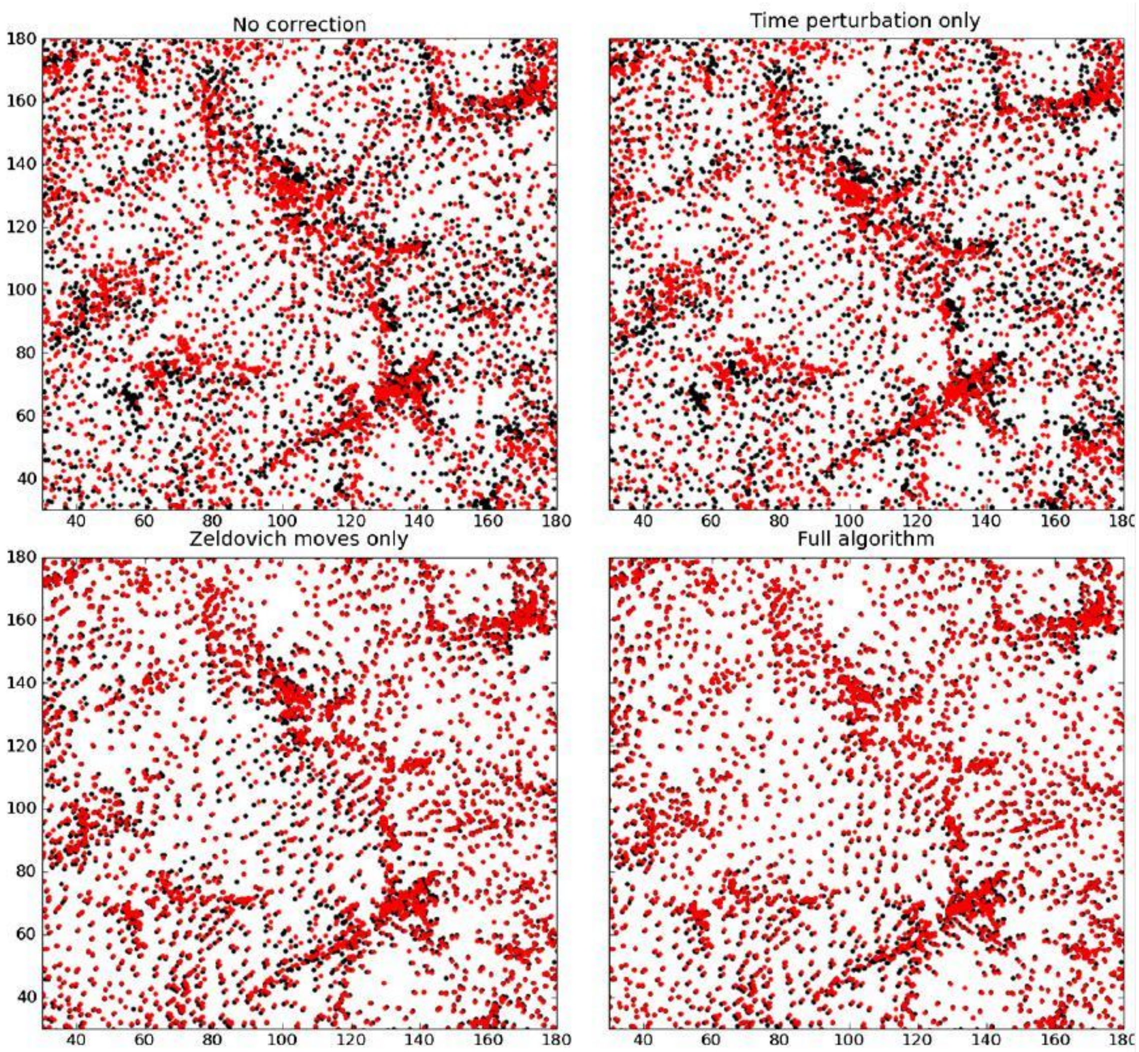
However:

i) Large scale modes are linear
ii) The non-linear coupling of the small scale modes to the large scale modes can be approximated as

$$\delta(\mathbf{x}, a) \approx \delta_L(\mathbf{x}, a) + \delta_S(\mathbf{x}', a) + \delta_L(\mathbf{x}, a) \frac{\partial \delta_S(\mathbf{x}', a)}{\partial \ln a} \left( \frac{d \ln D}{d \ln \Omega} \Big/ \frac{d \ln D}{d \ln a} \right).$$

(e.g. Peak-Background split, Kaiser 1984)

# Covariance (aside)



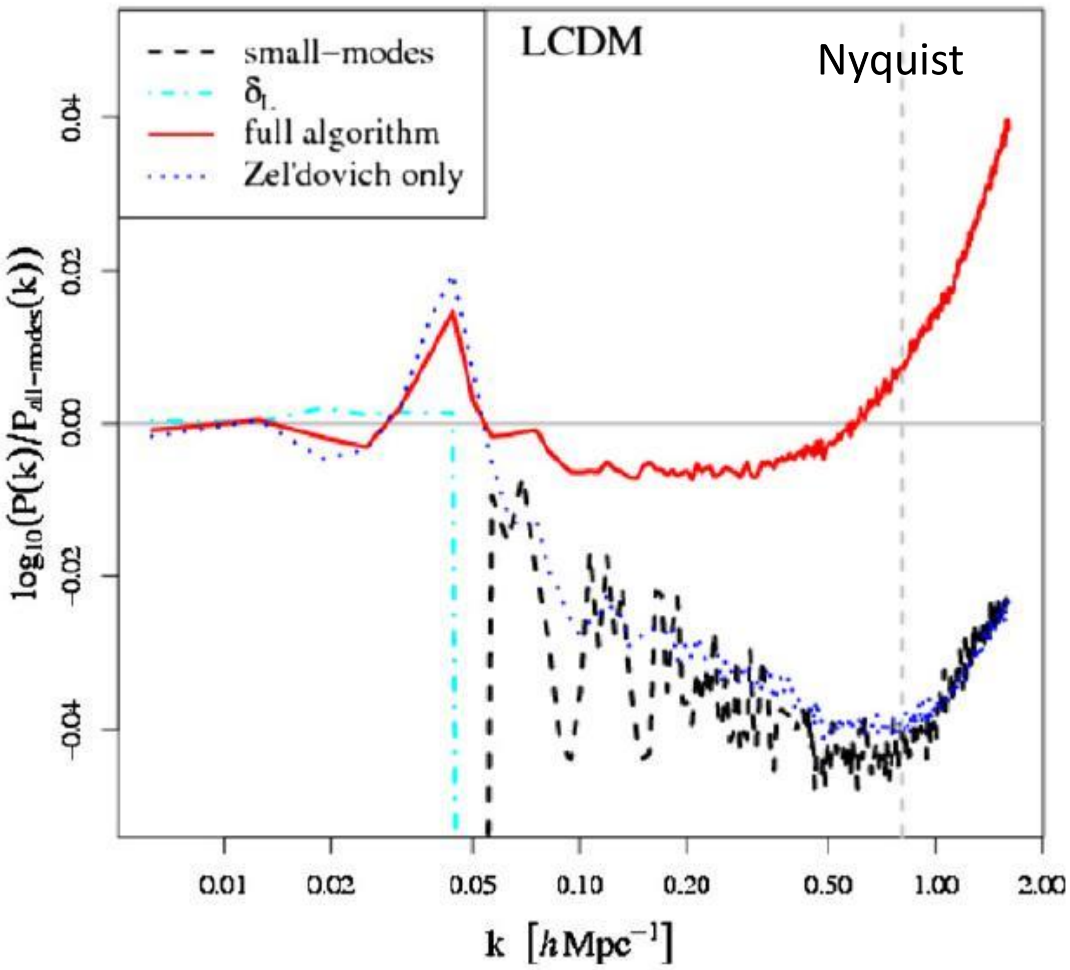

The algorithm can be implemented as combination of adding large scale modes using Zel'dovich shifts, together with a temporal shift of the small scale evolution to match the background perturbation's effect on the growth rate.

A good approximation!

Black points : full/normal simulation

Red points : long wavelength modes added after running the simulation

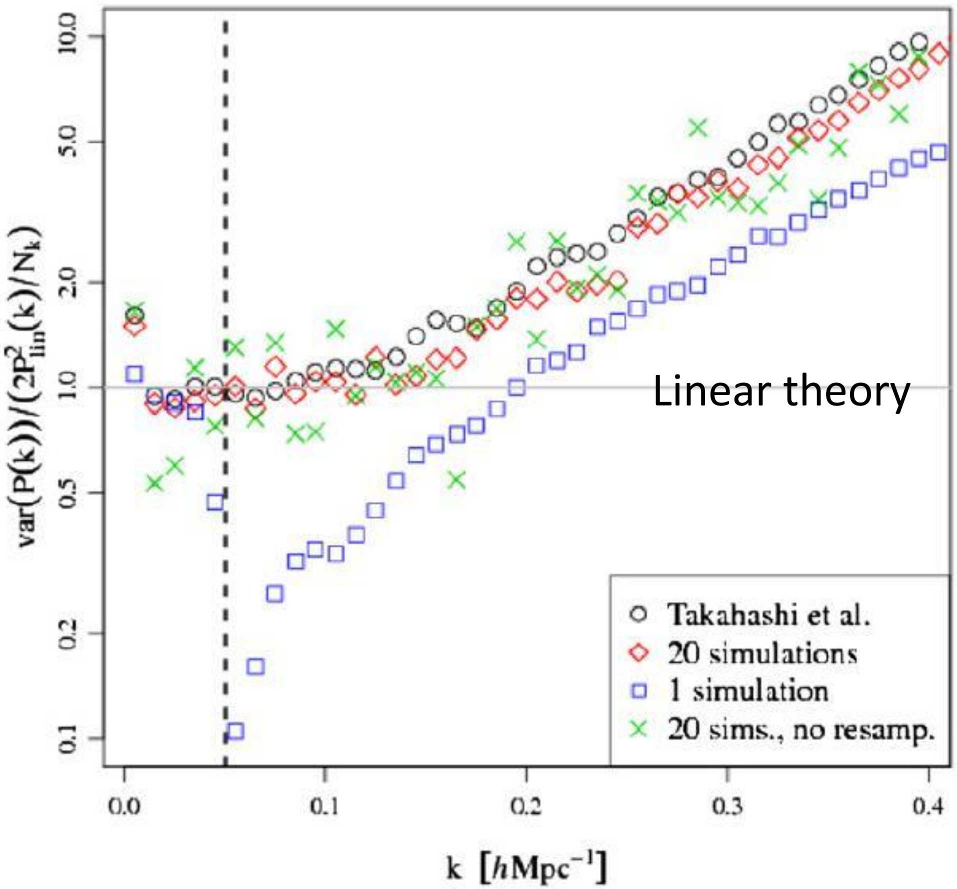
# Covariance (aside)




The effect on the mass power spectrum

Full algorithm recovers non-linear P(k) to sub 1% accuracy for k less than the Nyquist frequency of our grid.

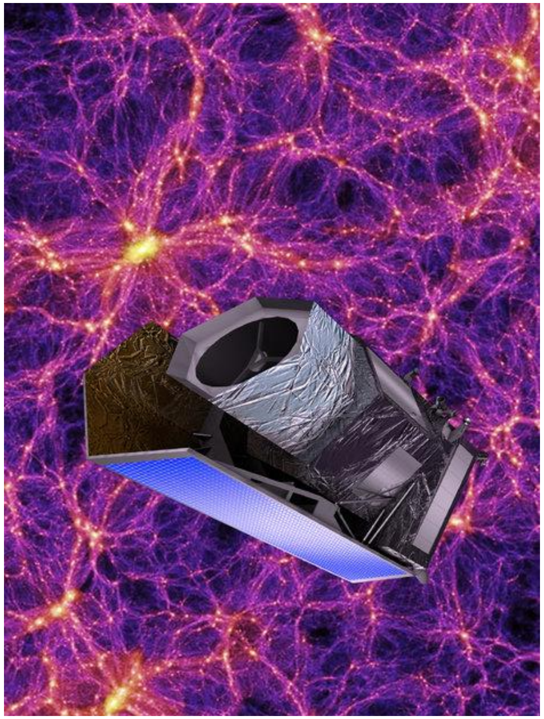
# Covariance (aside)




Can estimate the mass power spectrum covariance as accurately as Takahashi et al (2009), who used 500 simulations with just 20 simulations.

Schneider et al (2011, ApJ 737,11)

# Introduction



Euclid

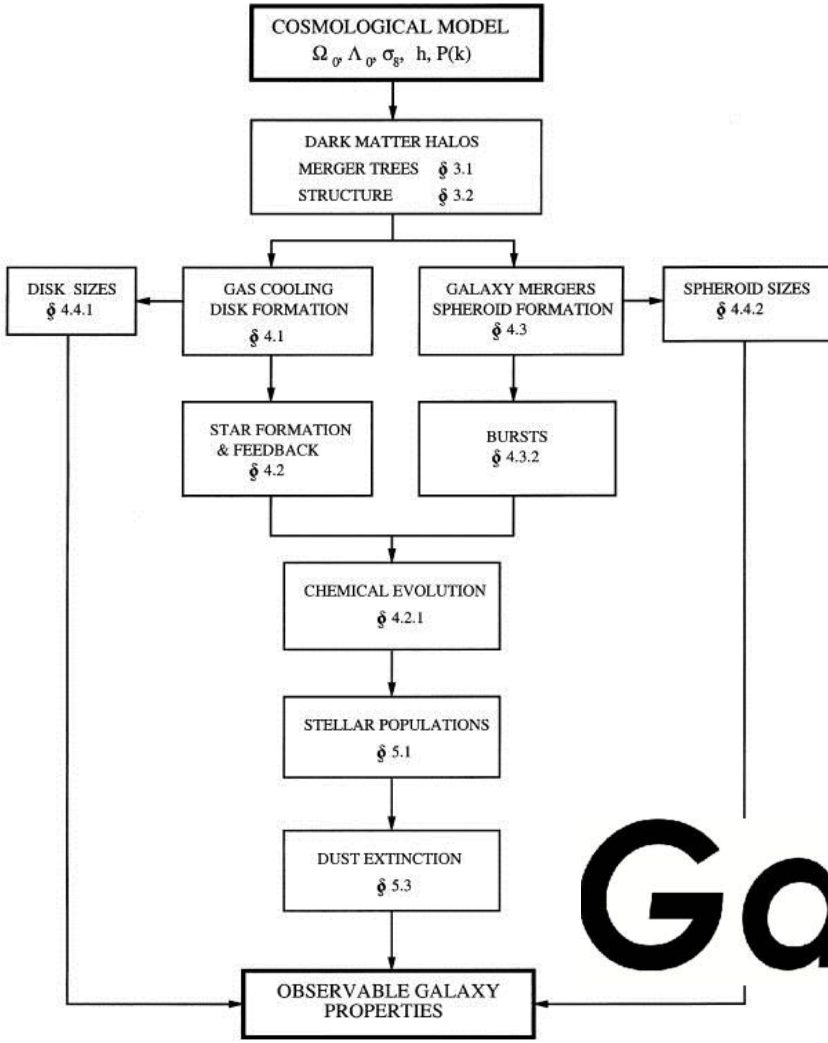
- Goal: Accurate predictions of galaxy clustering as a function of cosmological parameters.

- Ideally we need to predict not only clustering, but also the expected covariance of an estimate from a given survey. At present this is computationally demanding even just for the mass distribution (brief aside).

- Here the focus is bias, the relationship between galaxy and mass clustering.

# SHAM

Replace complicated astrophysics of galaxy formation by a simple ansatz.




Galform

Main Development:
Kravtsov et al 2004
Vale & Ostriker 2004
Conroy et al 2006
Moster et al 2010

See also
Reddick et al 2013

Aquarius, Springel et al (2008)

ICC

DU

# SHAM

Replace complicated astrophysics of galaxy formation by a simple ansatz.

Biggest galaxies form in the biggest (sub)haloes.

→ Galaxy stellar mass monotonically related to halo mass.

Note:

- There exist multiple galaxies per halo and so the mapping is between galaxies and subhaloes rather than haloes.
- The current mass of a subhalo is not the relevant quantity as it will be tidally stripped long before its host galaxy. Hence label subhaloes by their mass at infall

$$n_{\rm gal}(>M_{\rm stars}) = n_{\rm subhalos}(>M_{subhalo})$$
$$\Rightarrow M_{\rm stars}(M_{\rm subhalo})$$

Main Development:
Kravtsov et al 2004
Vale & Ostriker 2004
Conroy et al 2006
Moster et al 2010

See also
Reddick et al 2013

Aquarius, Springel et al (2008)

# SHAM

Replace complicated astrophysics of galaxy formation by a simple ansatz.

Biggest galaxies form in the biggest (sub)haloes.

→ Galaxy stellar mass monotonically related to halo mass.

Note:

- There exist multiple galaxies per halo and so the mapping is between galaxies and subhaloes rather than haloes.
- The current mass of a subhalo is not the relevant quaniity as it will be tidally stripped long before its host galaxy. Hence label subhaloes by their mass at infall

Main Development:
Kravtsov et al 2004
Vale & Ostriker 2004
Conroy et al 2006
Moster et al 2010

See also
Reddick et al 2013

Aquarius, Springel et al (2008)

Galaxy luminosity sometimes used instead of stellar mass.
Scatter can be introduced into the relation .
Other proxies such as $V_{max}$ exist..

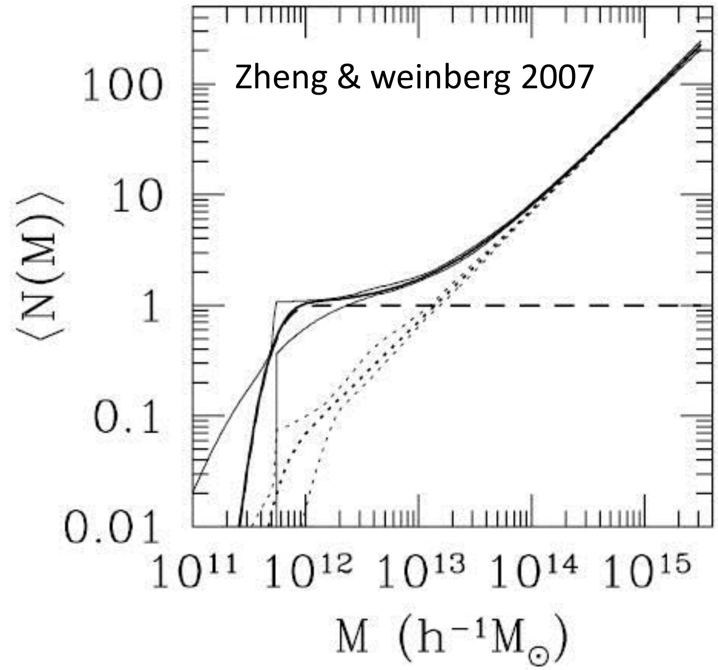
# SHAM vs HOD

Advantages of SHAM over HOD

$$M_{\mathrm{stars}}(M_{\mathrm{subhalo}})\ \mathrm{v.s.}\ \mathrm{P}(N, M_{\mathrm{halo}})$$

1. Fewer or even no free parameters
2. Does not ignore the subhalo content of haloes
   1. Makes use of merger history
   2. Includes halo formation bias
3. Predicts galaxy bias

HOD model:
Seljak 2000
Peacock & Smith 2000
Benson et al 2000
Berlind & Weinberg 2002

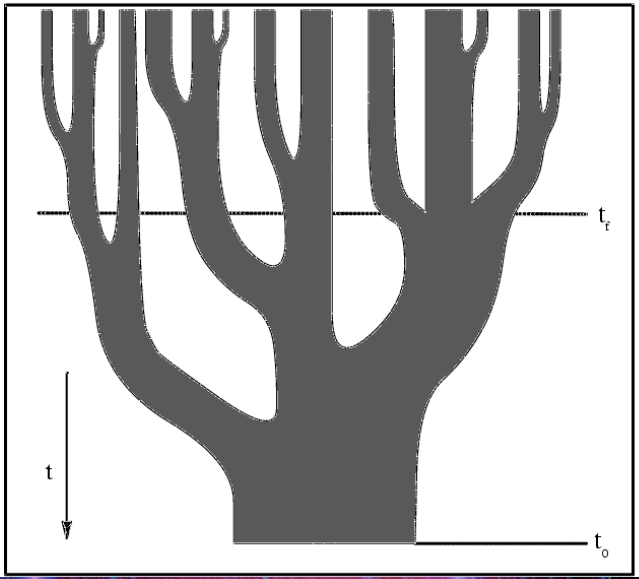


# SHAM vs HOD

Advantages of SHAM over HOD

$$M_{\rm stars}(M_{\rm subhalo})\ \text{v.s.}\ \mathrm{P}(N, M_{\rm halo})$$

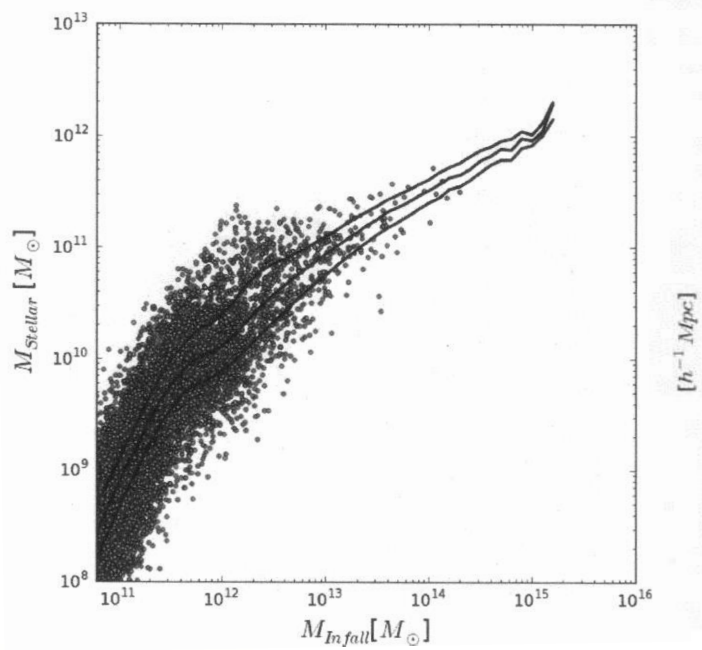
1. Fewer or even no free parameters
2. Does not ignore the subhalo content of haloes
   1. Makes use of merger history
   2. Includes halo formation bias
3. Predicts galaxy bias



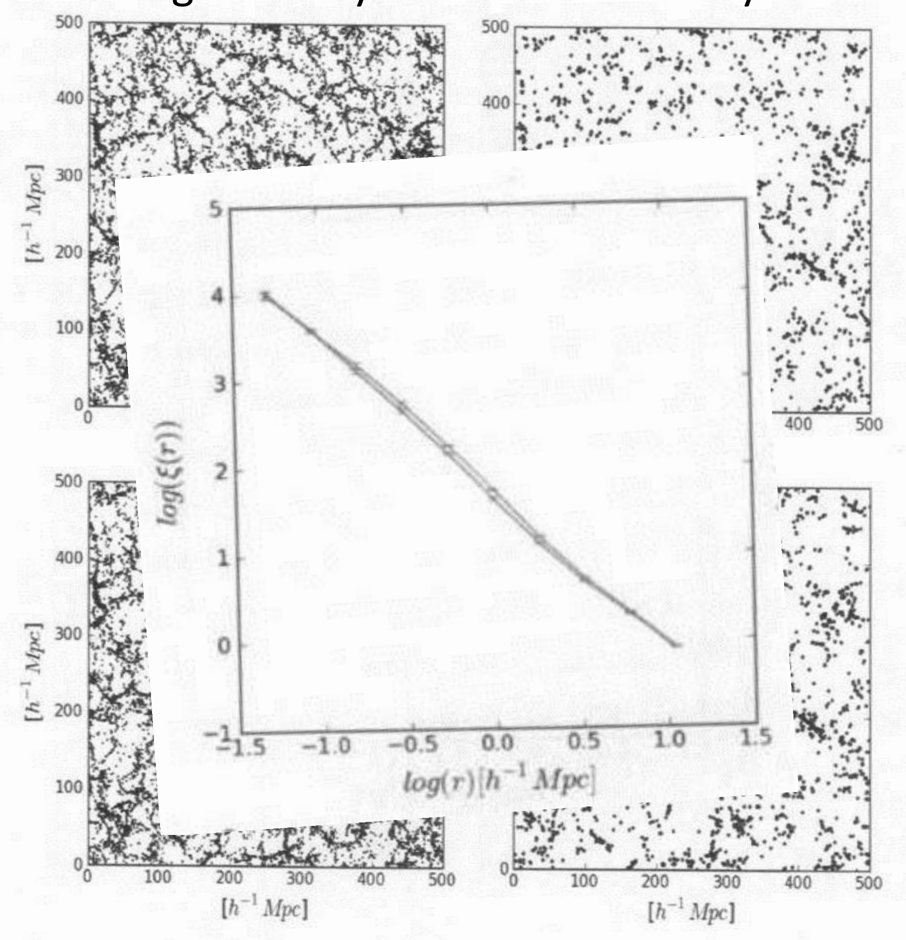

Could Mpc scale galaxy clustering be used to constrain cosmological parameters or is SHAM flawed?

# SHAM vs GALFORM

Test SHAM assumptions using the GALFORM semi-analytic model (Bower et al 2006)




High density | low density




Galaxies selected by stellar mass

Halos selected by infall mass

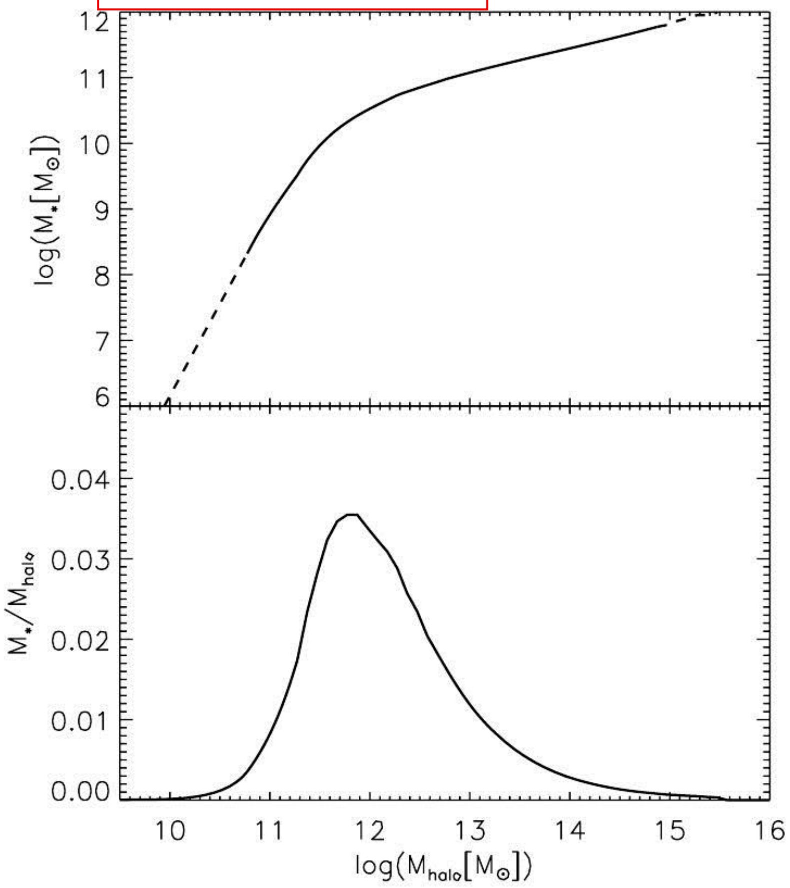
Ben Pike 2012, Durham undergraduate project

# SHAM: Millennium vs SDSS

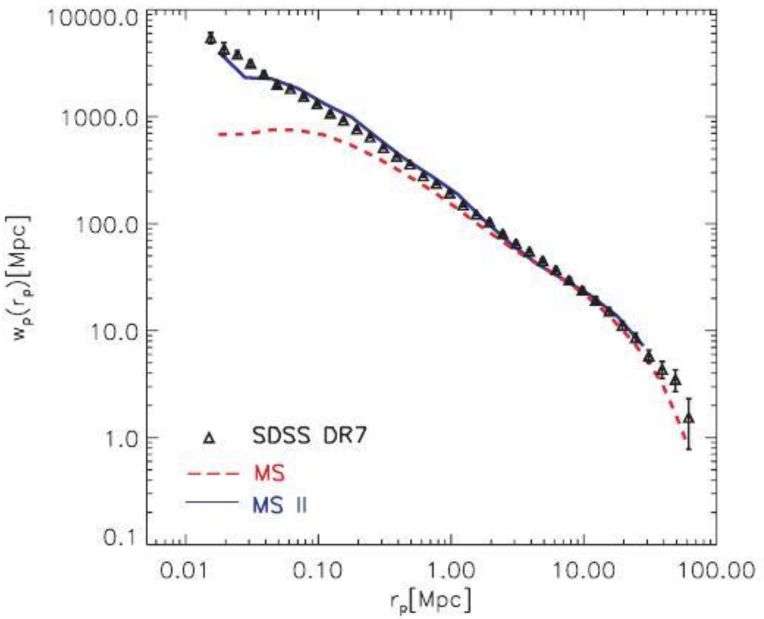
**Guo et al (2010)**

Millennium (WMAP1 cosmology) vs SDSS

$\Rightarrow M_{stars}(M_{subhalo})$



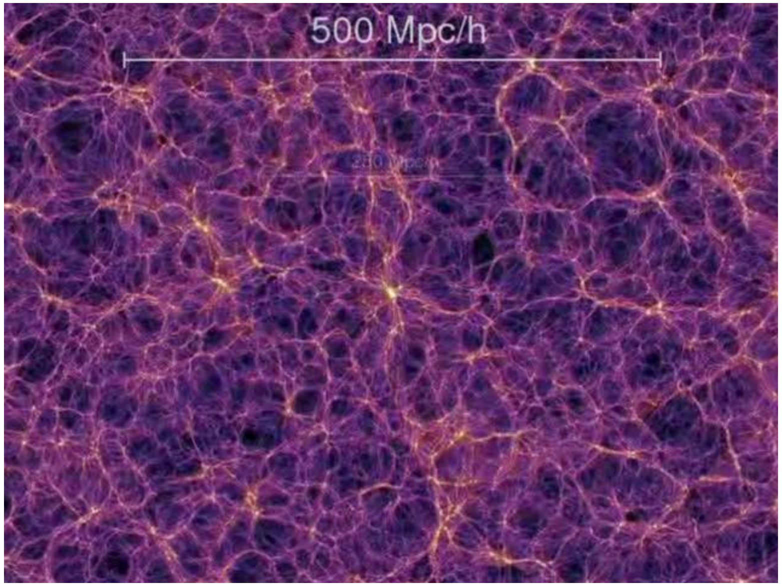

Then one can predict the clustering (here the stellar mass weighted projected correlation function) and compare to observations.

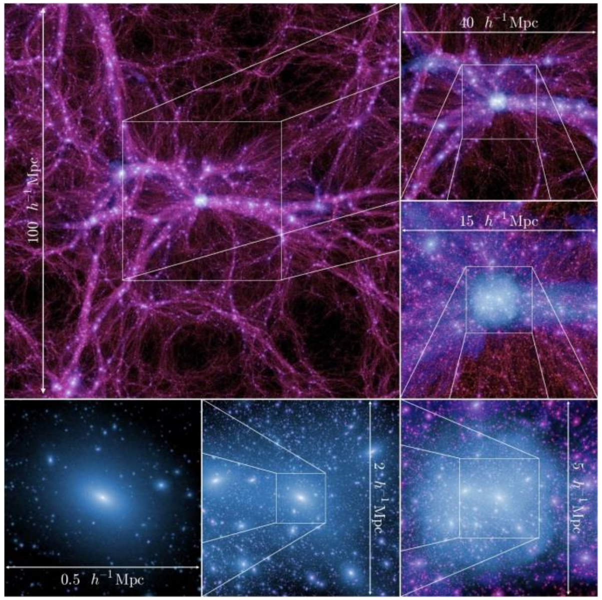



With the higher resolution MII simulation the match is pretty good but this is with WMAP1 not current WMAP/Planck cosmology

# Millennium I and II Simulations

Both have WMAP1 Cosmology: $\Omega_m = 0.25,\ \Omega_\Lambda = 0.75,\ n = 1,\ \sigma_8 = 0.9$




Millennium Simulation:
500 Mpc/h box
$2160^3$ particles
Mp=$8\times10^8$ Msol/h

Springel et al 2005

Millennium Simulation II:
100 Mpc/h box
$2160^3$ particles
Mp=$6\times10^6$ Msol/h

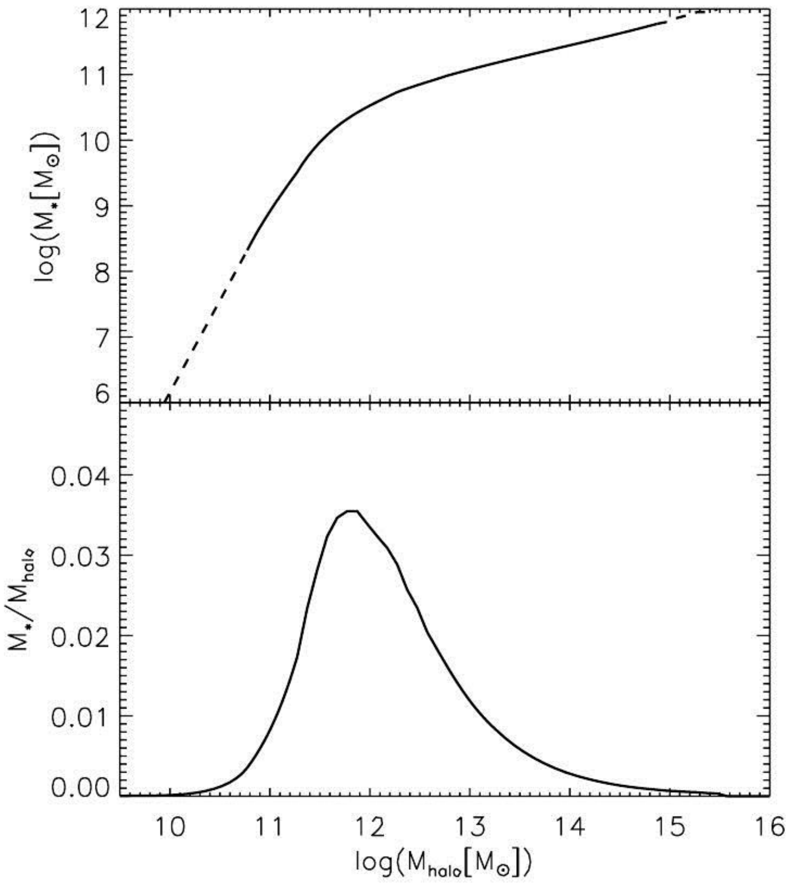
Boylan-Kolchin et al 2009

# SHAM: Millennium vs SDSS

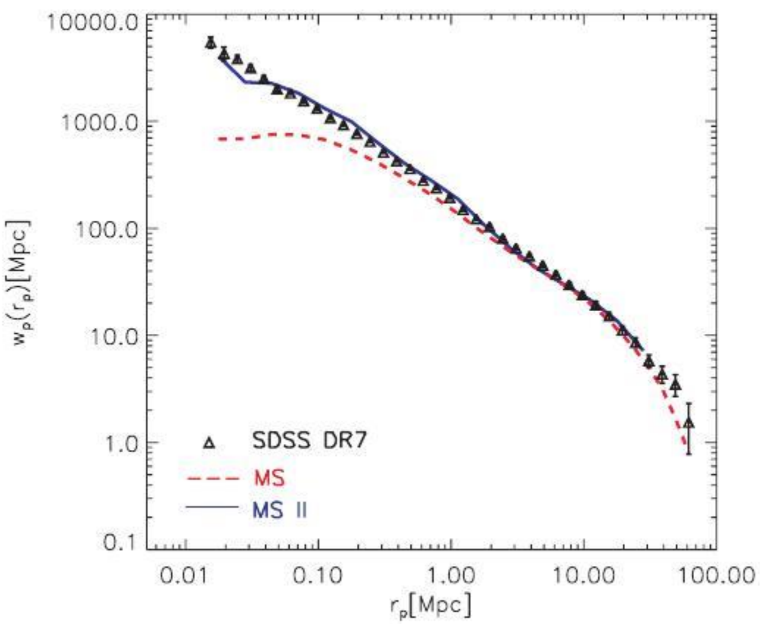
**Guo et al (2010)**

Millennium (WMAP1 cosmology) vs SDSS

$\Rightarrow M_{\text{stars}}(M_{\text{subhalo}})$



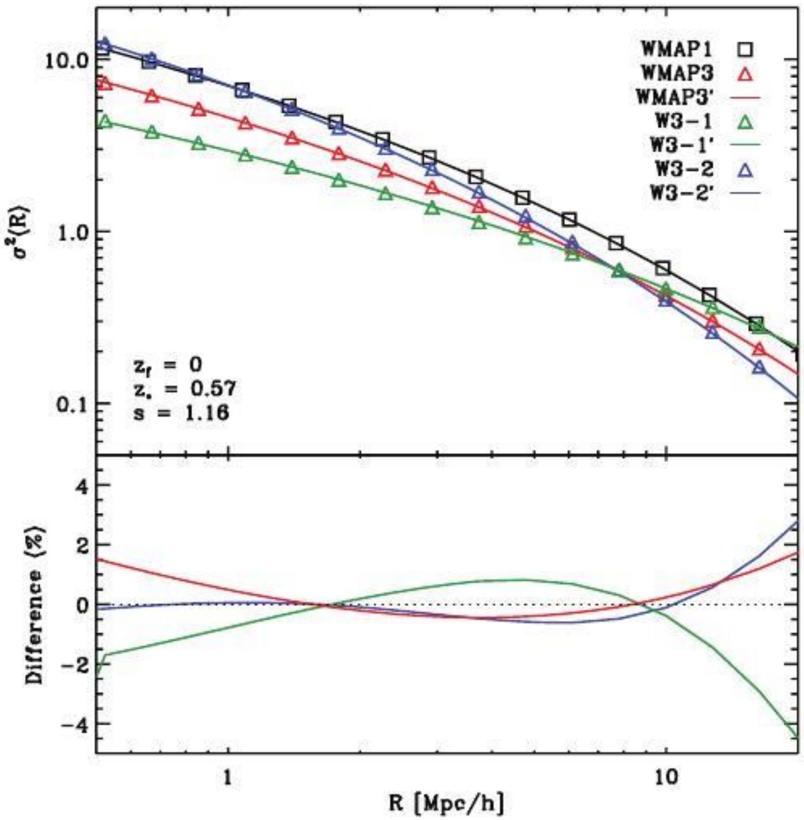

Then one can predict the clustering (here the stellar mass weighted projected correlation function) and compare to observations.

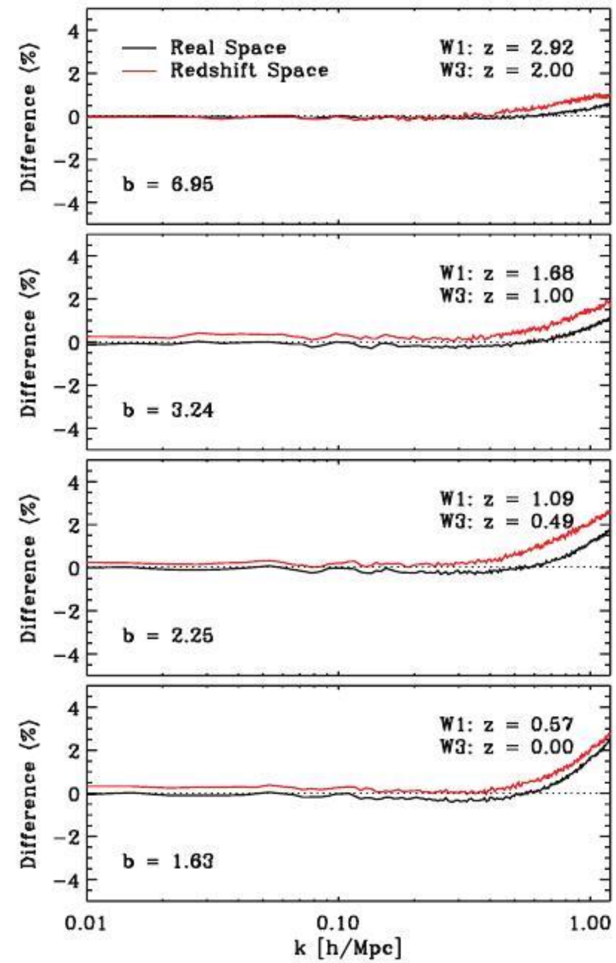



With the higher resolution MII simulation the match is pretty good but this is with WMAP1 not current WMAP/Planck cosmology

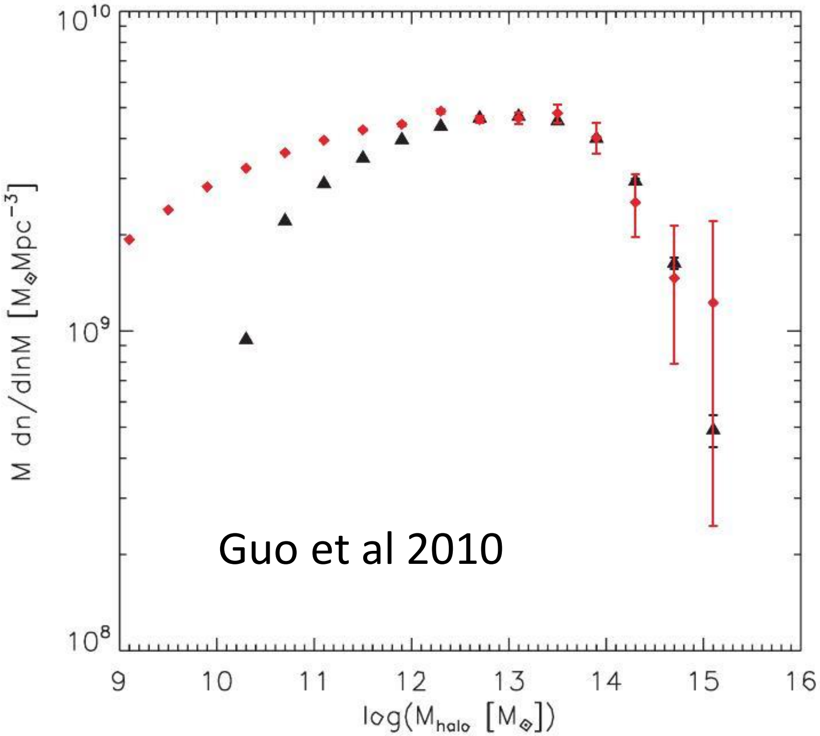
# SHAM: Varying Cosmology

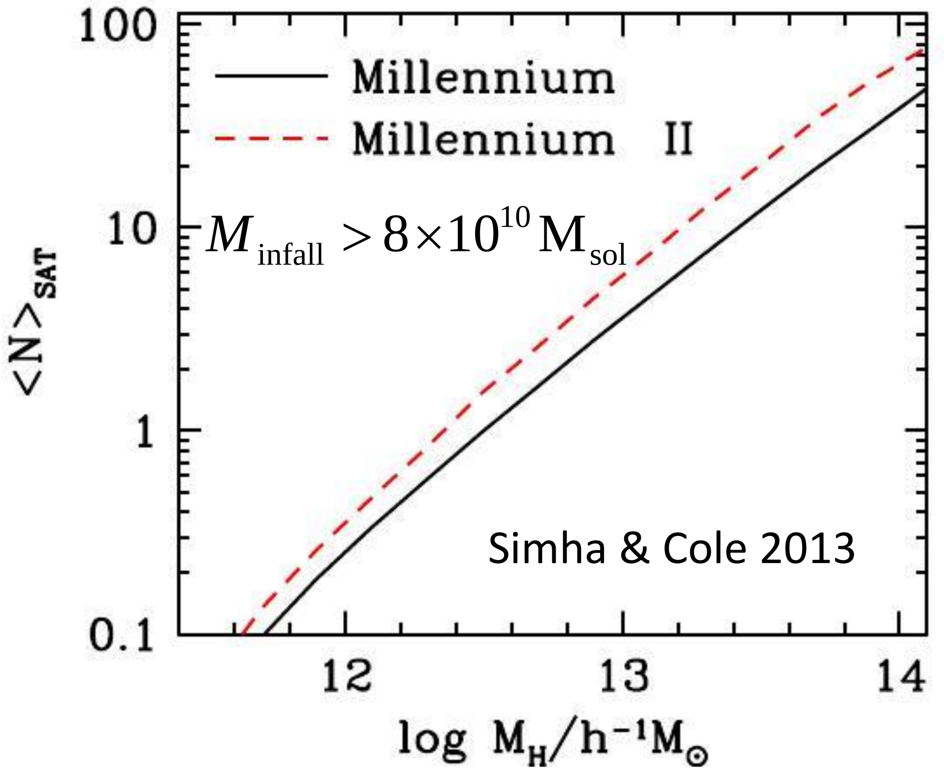
We don't have Millennium simulations  with a variety of cosmologies, but we can use the rescaling techniques of Angulo & White (2010).



# SHAM: Millennium+

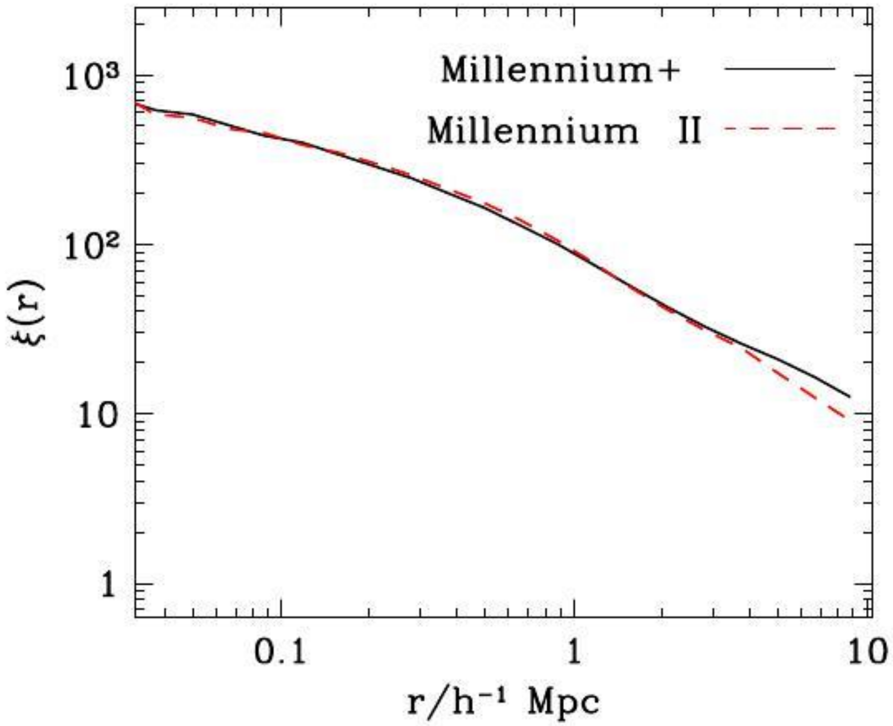



For standard SHAM to work the resolution must be such that no relevant subhaloes have been destroyed by tidal stripping. In practice this means subhalo catalogues are typically incomplete for $M_{infall}$<1000 particles

We want the volume of MI, but the resolution of MII.
We can almost achieve this by resurrecting some of the disrupted subhaloes to statistically match the satellite fraction seen in MII as a function of halo mass.

# SHAM: Millennium+




Correlation function of subhaloes with $M_{infall}$>8x$10^{10}$$M_{sol}$

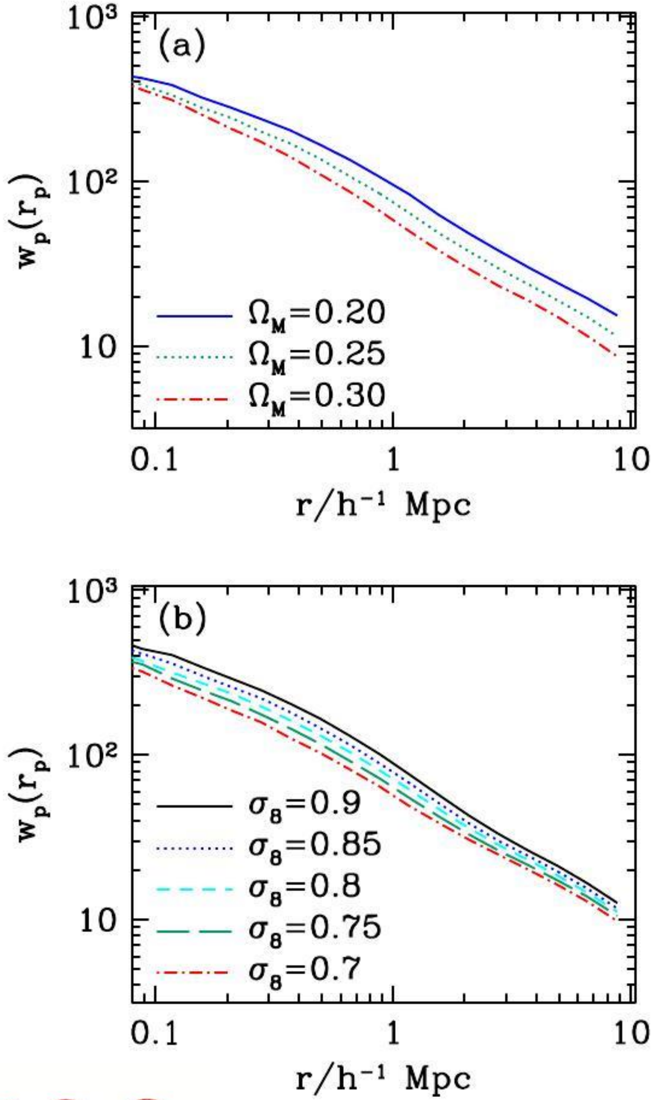
For MII this is for all such subhaloes that exist at z=0

For M+ this is for all such subhaloes that exist in MI at z=0 plus additional satellite subhaloes to match the satellite fraction (as a function o halo mass) in MII

Expect MII to be suppressed on largest scales due to its smaller box size.

→ Very good agreement..

# SHAM: predictions




This gives us all the ingredients to predict how the clustering of a particular galaxy sample should depend on cosmological parameters.

- Choose cosmological parameters
- Rescale Millennium+ so that input P(k) matches the linear theory (CAMB) expectation.
- Match abundances (SHAM)
- Populate the simulation
- Measure the clustering

Here we see that both decreasing $\Omega_m$ and increasing $\sigma_8$ boosts the clustering.

# SHAM: fitted to SDSS

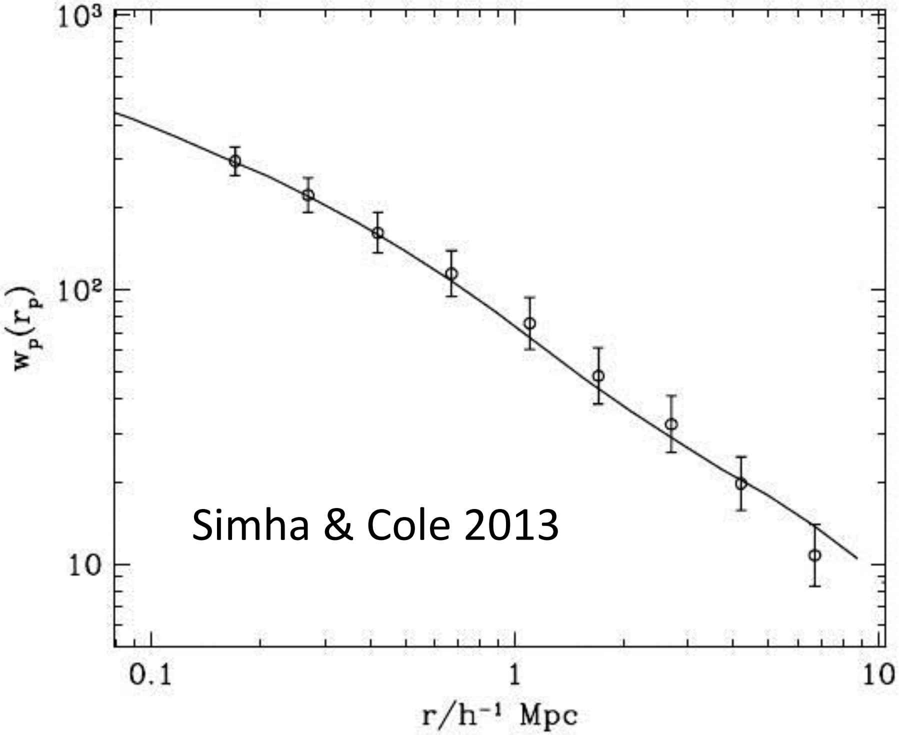
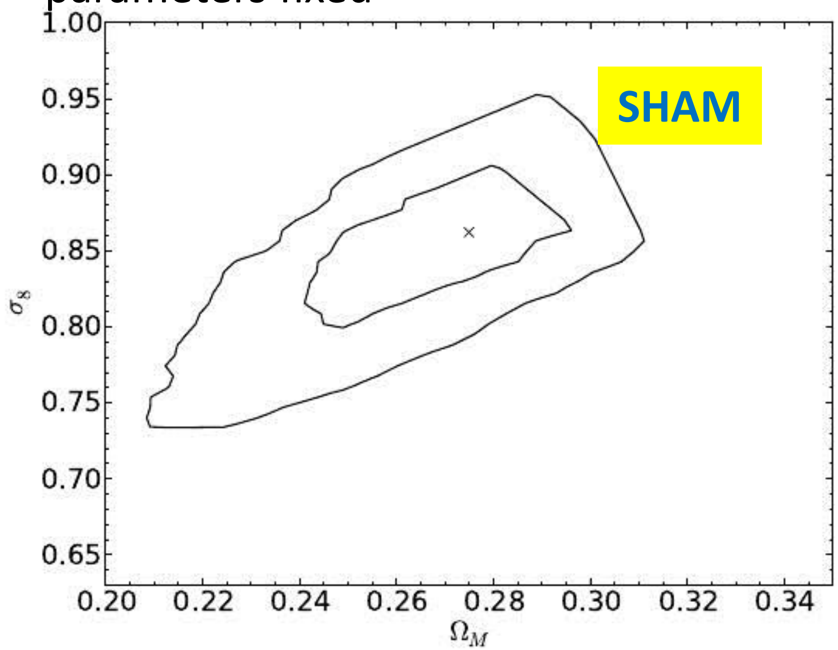



**Figure 4.** The solid curve is the galaxy two-point correlation function of our best-fit model with $\Omega_M = 0.275$ and $\sigma_8 = 0.86$. The points with error bars are the SDSS observed galaxy two-point correlation function from a volume limited sample of galaxies with $M_r \leq -18.0$.

Take r-band volume limited SDSS sample (Zehavi et al 2011) and fit SHAM by varying

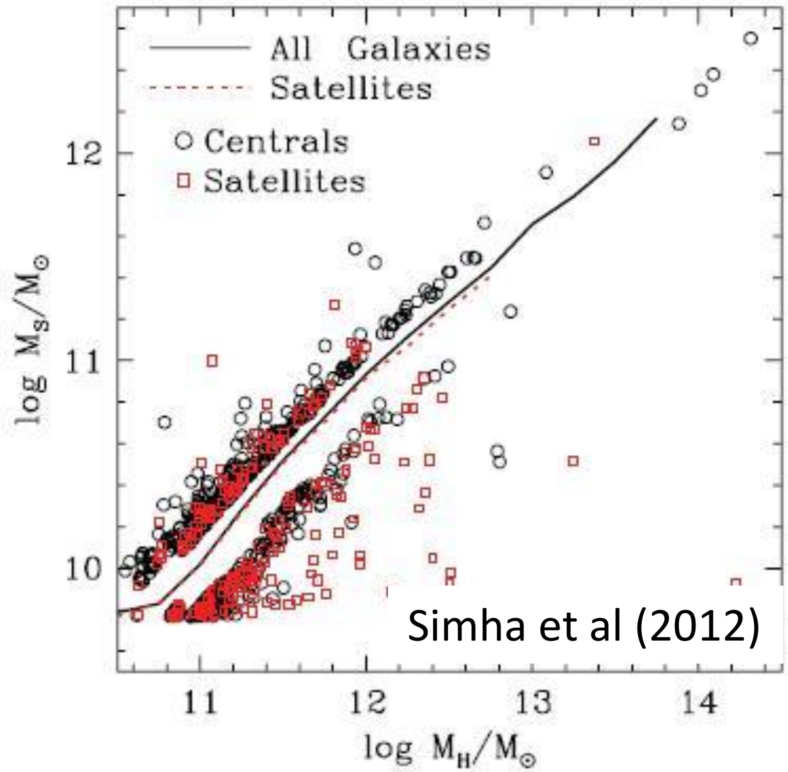
$$\Omega_{\rm m} - \sigma_8$$

and keeping all other cosmological parameters fixed

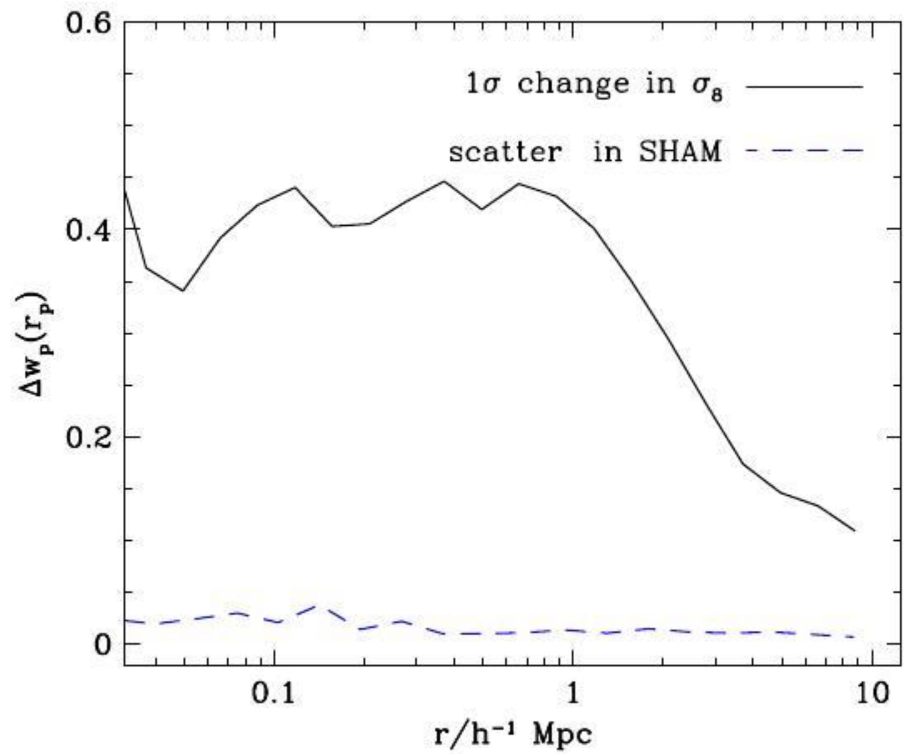


# SHAM: Systematics?

SHAM scatter in SPH simulation




Effect on projected correlation function




Effect of realistic scatter in the SHAM relation is small compared to the statistical errors

# SHAM: Systematics?

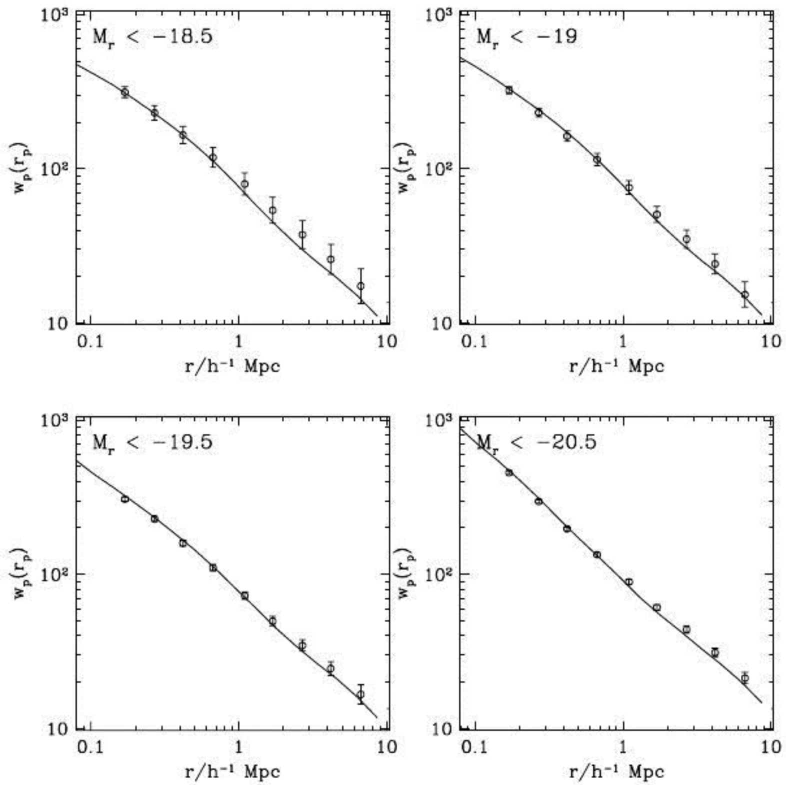



**Figure 9.** In each panel, the points with error bars are the SDSS observed galaxy two-point correlation function in a volume limited sample of galaxies brighter than $M_r$ = -18.5, -19, -19.5 and -20.5. The solid curve in each panel is the galaxy two-point correlation function predicted by our best-fit model with $\Omega_M$ = 0.275 and $\sigma_8$ = 0.86 for the corresponding galaxy sample.

Choice of luminosity threshold for the observational sample

All but the brightest are formally good fits.

The volume of the largest sample is greater than the Millennium volume and so statistical error on the model prediction become important.

Could combine samples to get stronger constraints, but would need inter-sample covariance.

# SHAM vs CMB

Compare WMAP7 CMB results and SHAM in the $\sigma_8$ $\Omega_m$ plane

Only very weak tension compared to the independent WMAP7 results

Doesn't fail!

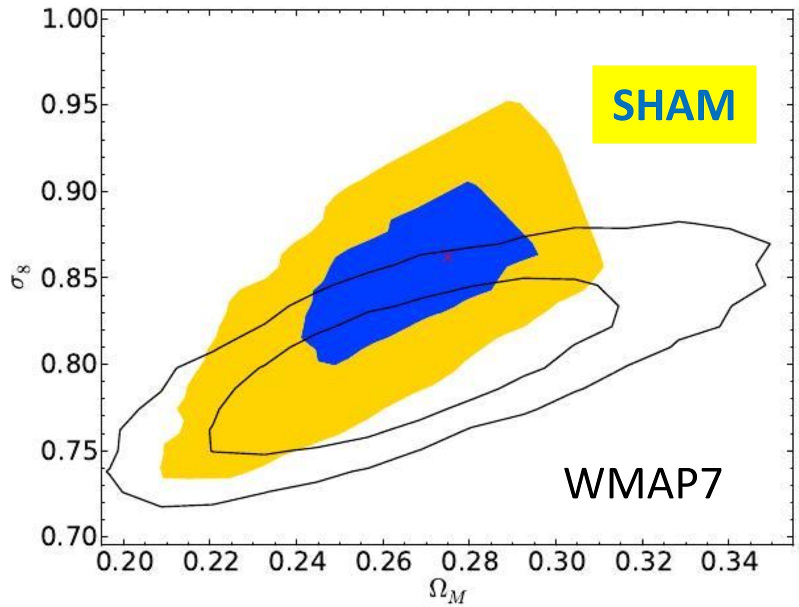



**Figure 11.** Joint constraint in the $\sigma_8$ - $\Omega_{\rm M}$ plane. The inner contour shows the boundary of the 68% confidence region and the outer contour shows the 95% confidence region. The filled contour is the result from this work while the black solid open contours are from WMAP7 (Komatsu et al. 2011).

# Towards Larger Scales

To apply the SHAM approach on larger scales requires being able to apply it to larger volume, lower resolution N-body simulations .

Two approaches:

1. Resolve (sub)haloes at infall, but use analytic calculations/approximations to determine whether they survive or are tidally disrupted and merge with the main halo.

   --Dynamical friction and tidal stripping.

2. Transplant halo merger histories and internal structure from higher resolution simulations.

   -- Neglect halo assembly bias

# Summary

1. SHAM predicts the amplitude of galaxy bias as well as the scale dependence (and higher order effects).
2. If accurate, then in combination with cosmological rescaling techniques moderate scale galaxy clustering it can be used to constrain cosmological parameters.
3. Passes the test when applied to SDSS main survey correlation functions of r-band selected samples.
4. Very large scale SHAM mocks will require modelling substructure survival or bootstrapping from smaller high resolution simulations .