Using LasDamas: What more can we learn from the galaxy distribution?

Cameron K. McBride
Vanderbilt University

Berkeley Cosmology Seminar

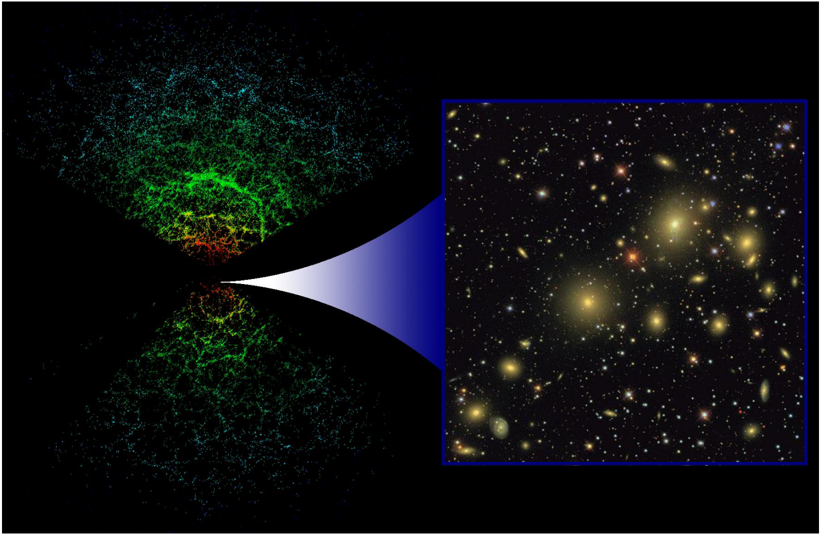
4 January 2011

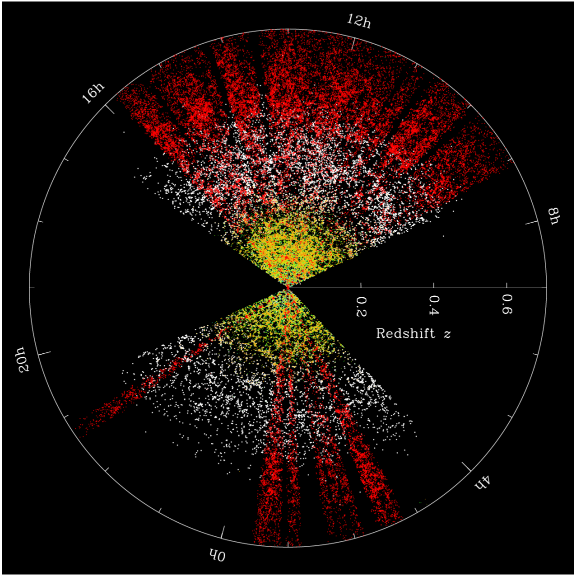
Sloan Digital Sky Survey

APO: Sloan Telescope

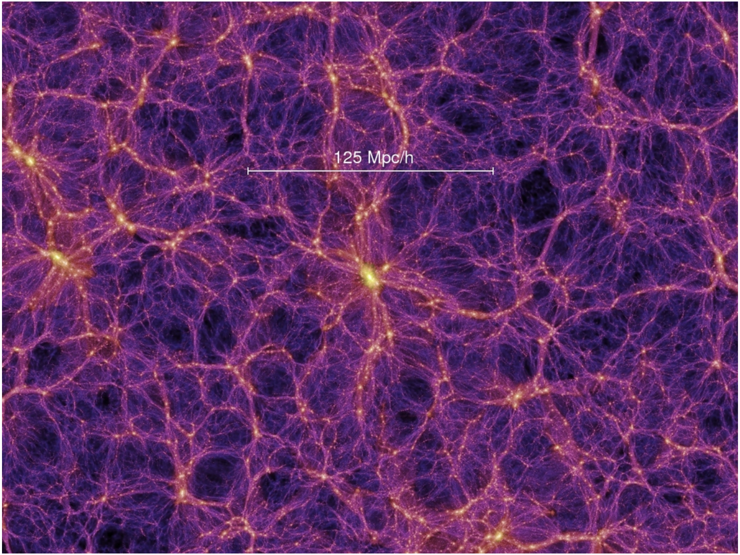
- 1.6 Million Spectra
- 8423 sq deg
- 929,555 DR7 galaxy spectra
- over 105,000 DR7 LRGs ($z < 0.47$)
- BOSS: Over 1 million luminous galaxies ($z < 0.7$)

Spectroscopic galaxy distribution





SDSS-II + BOSS Galaxies (image: Blanton)



Millennium Simulation: Springel, et al. 2005

Correlation Functions

- $\xi(r)$: connected spatial two point correlation function

$$\delta P = n^2 \delta V_1 \delta V_2 \left[1 + \xi(r_{12})\right]$$

- Defined by density field:

$$\begin{aligned} \delta(\vec{x}) &= \frac{\rho(\vec{x})}{\bar{\rho}} - 1 \\ \xi(r_{12}) &= \langle \delta(\vec{x}_1)\delta(\vec{x}_2) \rangle \\ \zeta(r_{12}, r_{23}, r_{31}) &= \langle \delta(\vec{x}_1)\delta(\vec{x}_2)\delta(\vec{x}_3) \rangle \end{aligned}$$

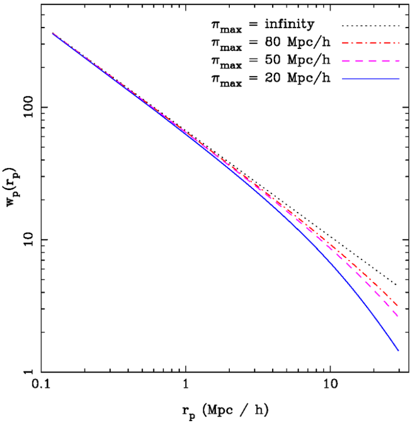
- $\xi(r)$ close to a power law for galaxies
- infinite series need to describe arbitrary point distribution (2PCF, 3PCF, etc)
- Gaussian distribution fully described by 2PCF

Two Point Correlation Function

- 3D clustering statistic
- redshift distortions
- with distortions: redshift space
- without distortions: real space

Projected 2PCF

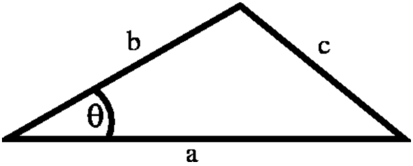
$$w_p(r_p) = 2\int_0^{\pi_{max}} \xi(r_p, \pi)\, d\pi$$

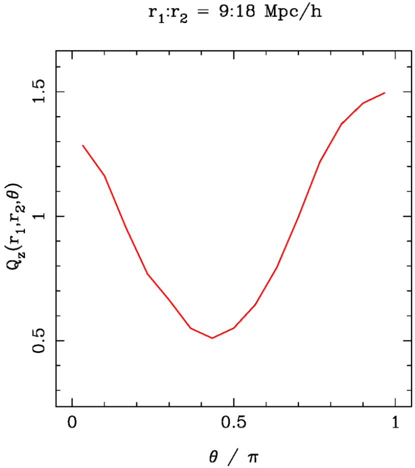


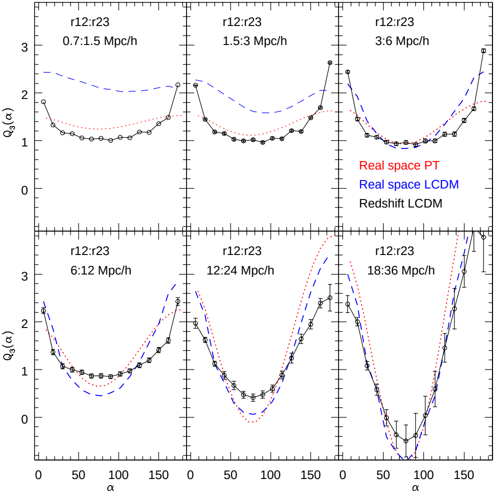
Three Point Correlation Function

Q: Reduced 3PCF

$$Q(r_a, r_b, \theta) = \frac{\zeta(r_a, r_b, r_c)}{\xi_a\xi_b + \xi_b\xi_c + \xi_c\xi_a}$$



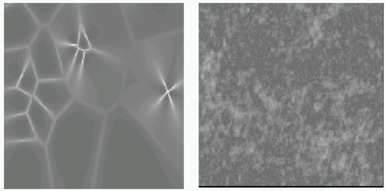



r12:r23
0.7:1.5 Mpc/h
r12:r23
1.5:3 Mpc/h
r12:r23
3:6 Mpc/h
Real space PT
Real space LCDM
Redshift LCDM
r12:r23
6:12 Mpc/h
r12:r23
12:24 Mpc/h
r12:r23
18:36 Mpc/h
$Q_3(\alpha)$
α
0
1
2
3
50
100
150

Why bother with 3PCF?

Higher order clustering is important!

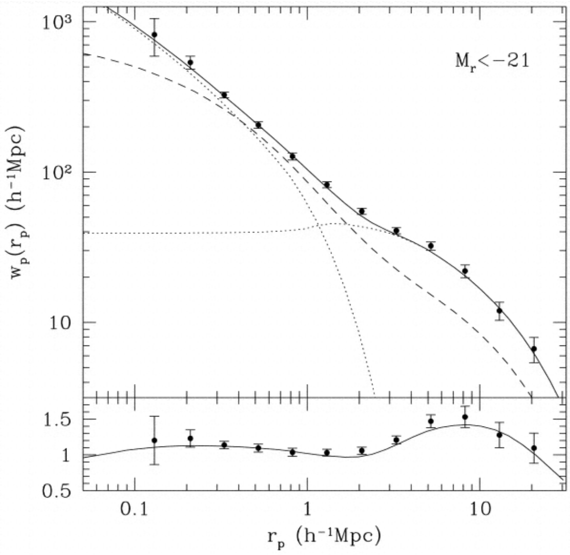
- more information than 2PCF analyses
- encoded shape information
- break degeneracies in models
    - halo model parameters
    - linear bias and σ_8
- data samples are large and growing



- phase information randomized for image on right
- same 2PCF but very different distributions

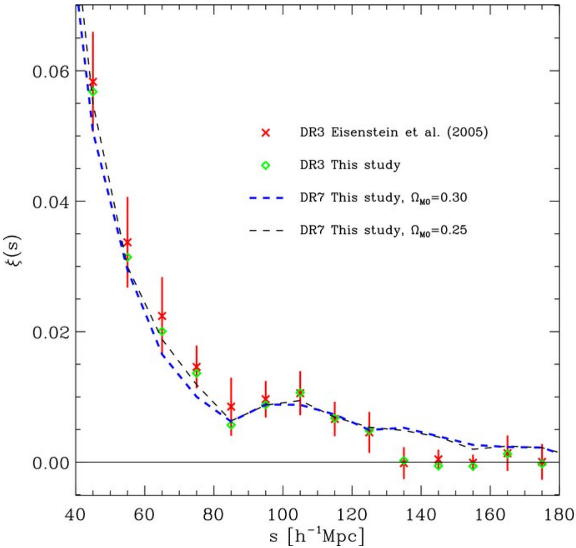
Some clustering results

Earlier SDSS: Power Law vs Halo Model

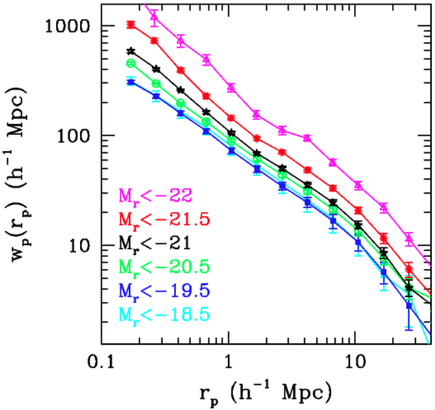


Zehavi, et al. 2004

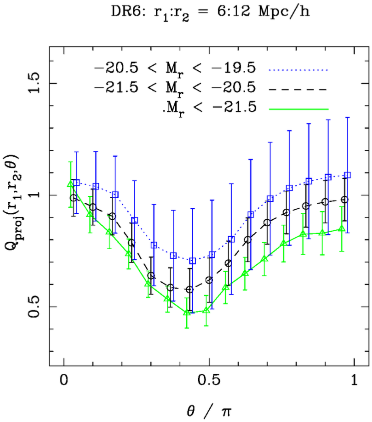
Some clustering results
SDSS: BAO Peak



Kazin, et al. 2010

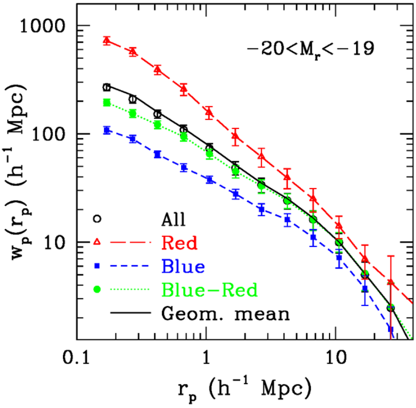


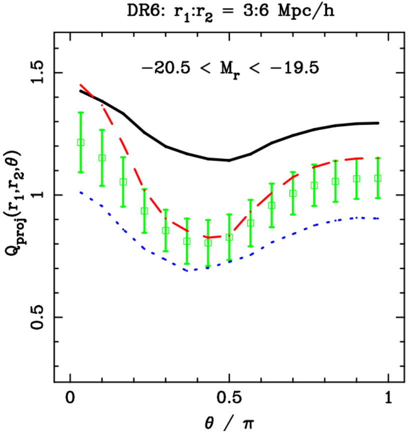
1000
100
10
0.1
1
10
$w_p(r_p)$ (h^{-1} Mpc)
r_p (h^{-1} Mpc)
$M_r<-22$
$M_r<-21.5$
$M_r<-21$
$M_r<-20.5$
$M_r<-19.5$
$M_r<-18.5$



DR6: r_1:r_2 = 6:12 Mpc/h
$-20.5 < M_r < -19.5$
$-21.5 < M_r < -20.5$
$M_r < -21.5$
$Q_{proj}(r_1,r_2,\theta)$
0.5
1
1.5
0
0.5
1
θ / π

Some clustering results
Color Dependence





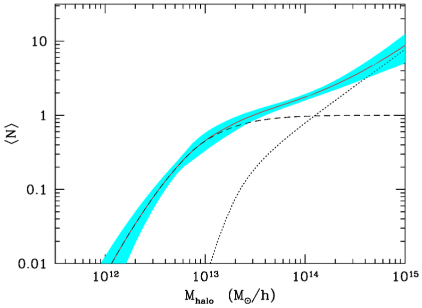
Modeling Galaxy Distribution

1. Our models
   - cosmological parameters / dark energy
   - early conditions (ICs)
   - general relativity
2. Characterize dark matter halos
3. Map Galaxies to Halos
4. Observe galaxy distribution

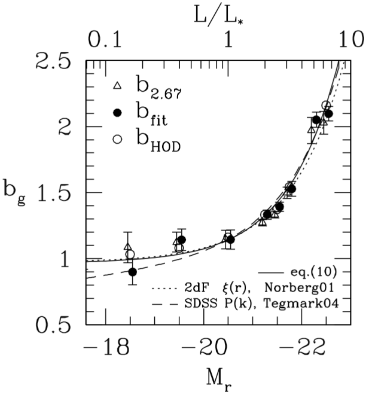
Understanding the difference in clustering:

- galaxies to mass: understand cosmology
- parameterization of bias: constraints on galaxy formation

Modeling: some results

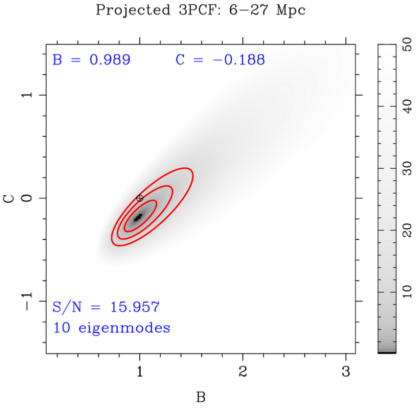


BOSS Galaxies; White, et al.



Zehavi, et al. 2010

Modeling: Galaxy-Mass Bias Constraints



local bias model:

$$\delta_g \approx b_1 \delta_{mass} + \frac{b_2}{2} \delta_{mass}^2$$

Bias in Projected 3pt

$$Q_g \approx \frac{1}{B} \left(Q_{mass} + C \right)$$

$B = b_1$ and $C = \frac{b_1}{b_2}$

Understanding Errors

To use observations for constraints, we need to understand the properties of our errors

Significant errors in:

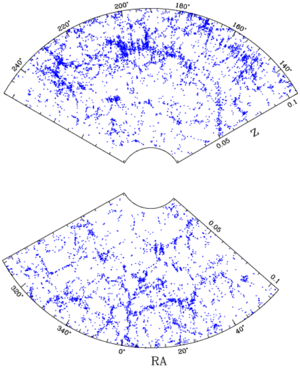
- Methods of analysis
- Observational Systematics
- Statistical

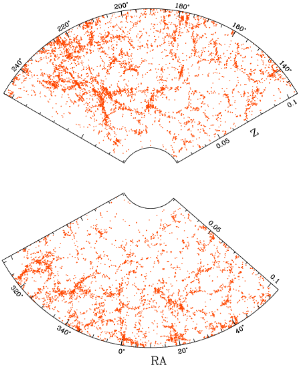
Ways of Estimating Errors:

1. theory
2. internal estimates on data (jackknife, bootstrap)
3. many realizations with signifcant assumptions
4. numerical modeling, typically few realizations (mocks)

Faking it: Mock Galaxy Catalogs

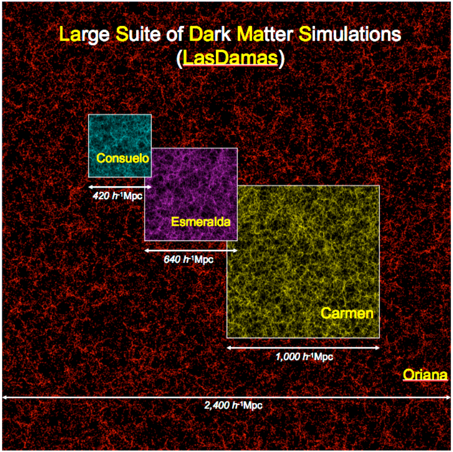
which is real?





Some goals:

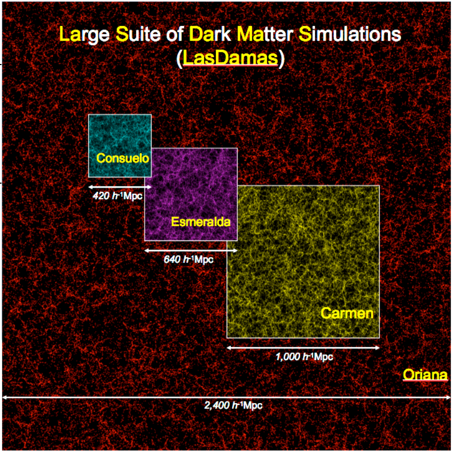
- target DR7 galaxy samples
- run many realizations
- focus on statistical studies
- sufficient resolution
- model observational limitations
- study halo properties
- make galaxy mocks public



CKM, A. Berlind (Vanderbilt); R. Scoccimarro (NYU); M. Manera (Portsmouth); R. Wechsler (Stanford); M. Busha (Zurich); F. van den Bosch (Yale)

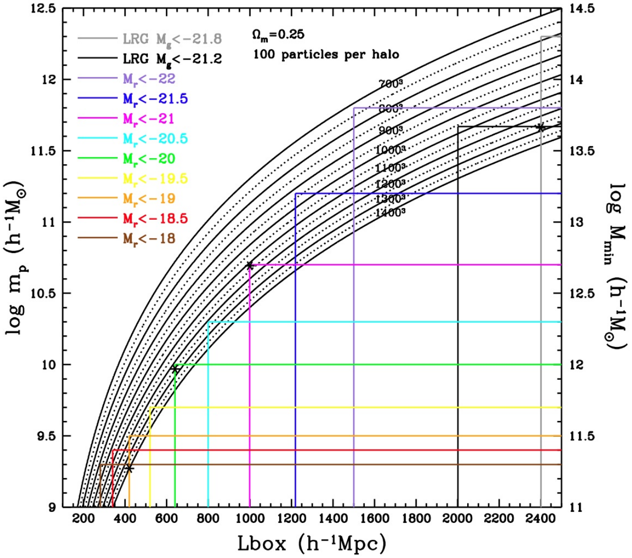
Name	Npart	Lbox(h^{-1}Mpc)
Oriana	1280^3	2400
Carmen	1120^3	1000
Esmeralda	1250^3	640
Consuelo	1400^3	420

- 50 realizations
- 200 simulations
- 10-12 million CPU hours
- 2LPT initial conditions
- over 750 $h^{-3}\mathrm{Gpc}^3$



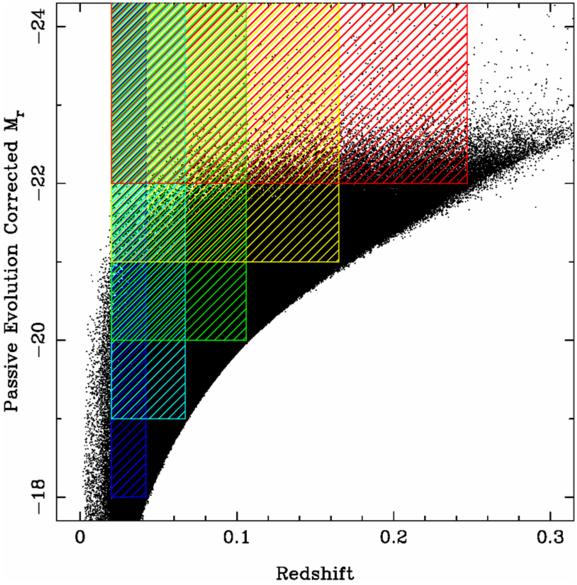
http://lss.phy.vanderbilt.edu/lasdamas

Simulations Parameters



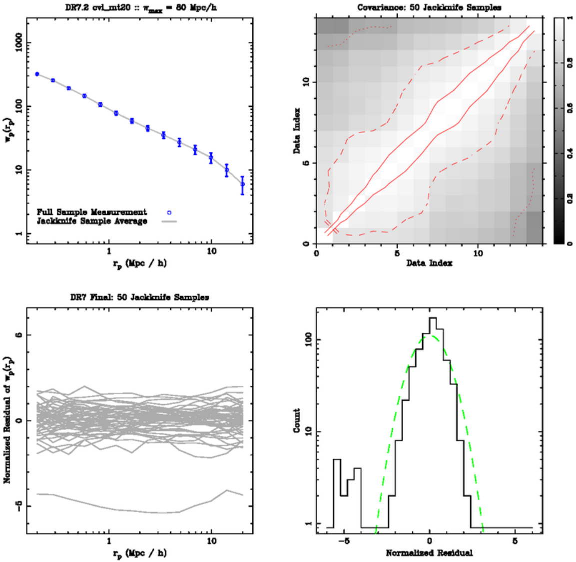
SDSS Galaxy Samples (DR7)

volume-limited selection



SDSS Galaxy Samples (DR7)

measurements

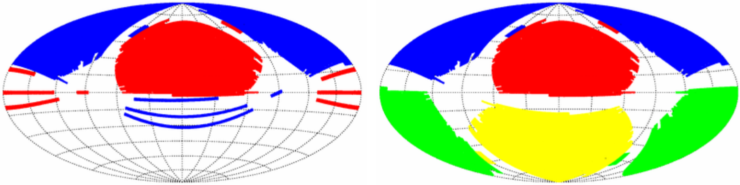


LasDamas Mock Fitting

Idea is to use clustering and galaxy number density as "truth"

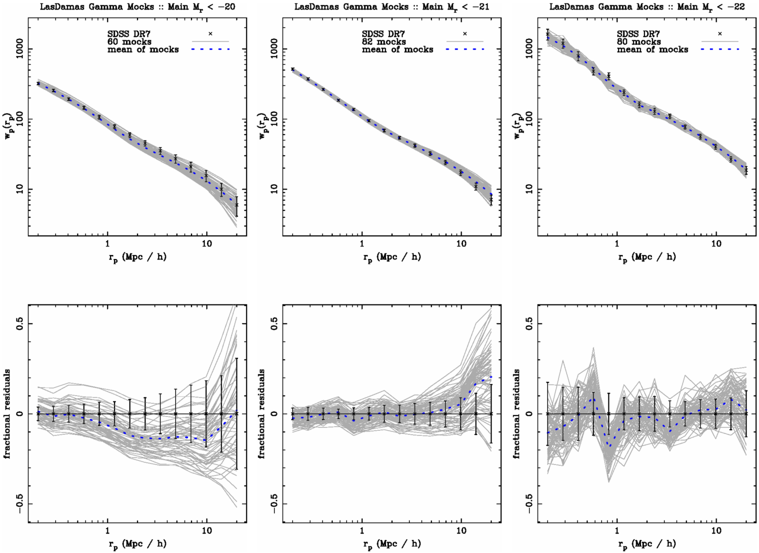
1. identify halos (currently FOF, soon SO)
2. make galaxy distribution (HOD)
3. apply redshift distortions
4. measure $w_p(r_p)$, evaluate fit to data
5. repeat (simple steepest descent)
6. done when "good enough"

LasDamas: SDSS Mocks
Apply SDSS Geometry

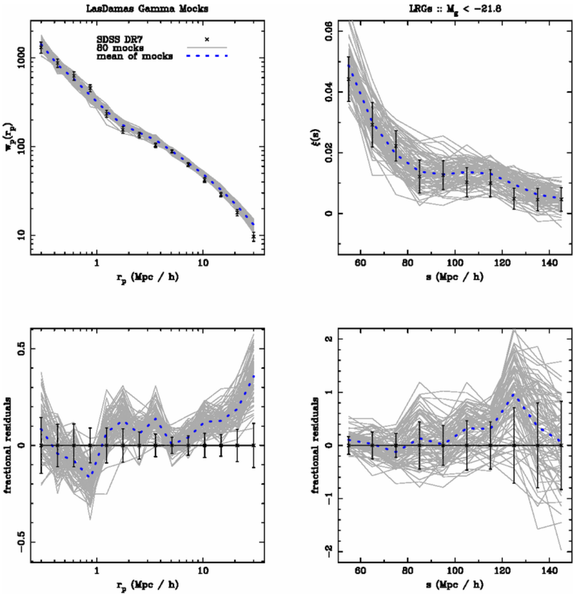


LasDamas: SDSS Mocks

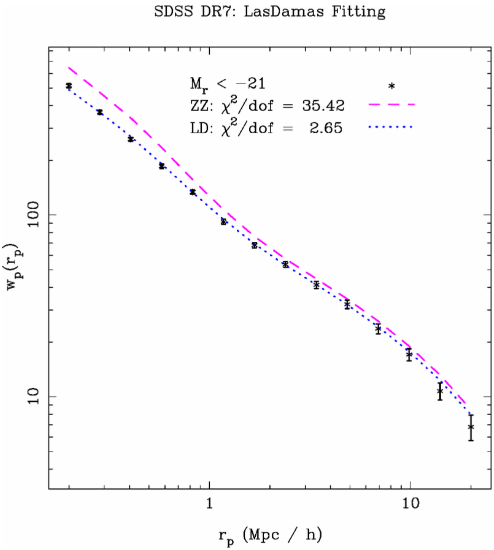
How well do they do?



LasDamas: SDSS Mocks
How well do they do?



LasDamas: SDSS Mocks
Why care so much?



LasDamas: SDSS Mocks
What's available?

"gamma" release

- 3 MAIN samples: main20, main21, main22
- 3 LRG samples: lrg21p2, lrg21p8, lrgFull*
- Publicly available in both real and redshift space.

version 1 : coming soon!

- 2 more MAIN samples: main18, main19
- more realizations!
- fiber collisions included
- adding spheres, as well as SDSS footprint

`http://lss.phy.vanderbilt.edu/lasdamas`

LasDamas: Abundance of High Mass Halos

Cluster counts used to constrain cosmology (ACT, SPT, Planck)

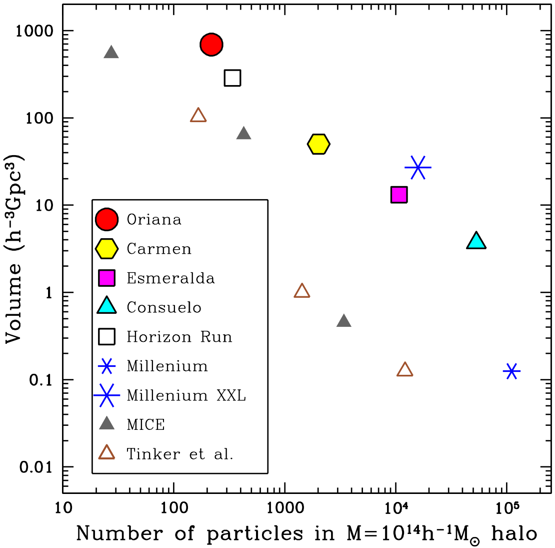
Cluster abundance

$$N(> O_{min}) = VOLUME \times \int P(> O_{min}|M, z) \frac{dn(z)}{dM} dM$$

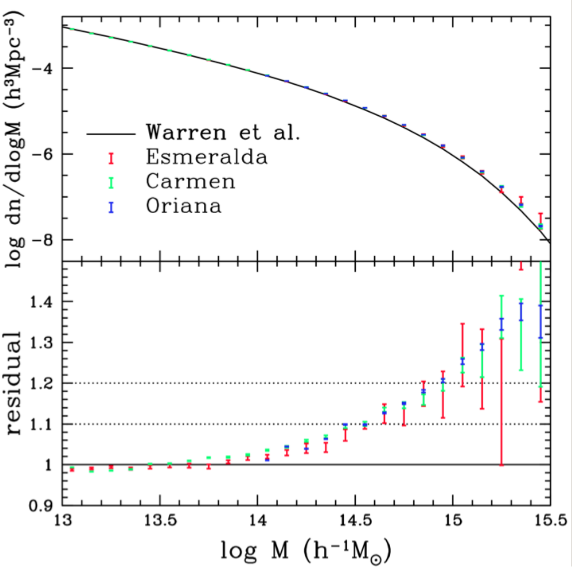
Our era of "precision cosmology" has high demands!

- halo mass function calibrated on simulations
- observable-mass relation hard
- systematic effect on abundance can affect constraints!
- affect of gas physics? (Stanek, et al.)

LasDamas: Volume and Resolution



LasDamas: halo mass function



To run an $n - body$ simulation, need input cosmology and initial conditions at a high redshift: z_{init}.

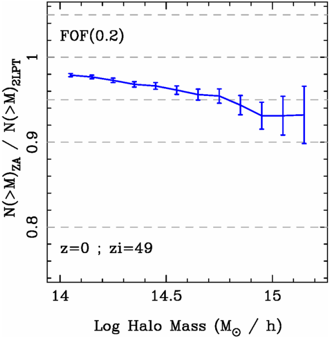
- $z >> z_{init}$ evolution handled by initial condition generation
- inaccuracies cause transients in evolution
- transients impede collapse

- ZA does not preserve angular momentum
- 2LPT allows curvature in early trajectories
- differences significant in IC velocities

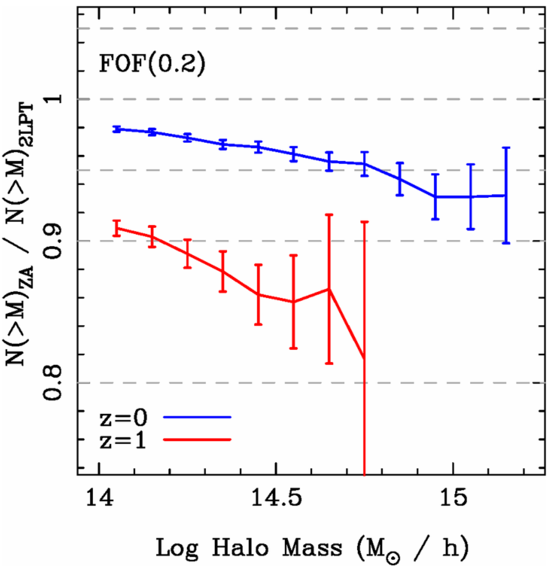
See Scoccimarro 1998, Crocce et al 2006, Crocce et al 2010

Test ICs with LasDamas
ZA vs 2LPT

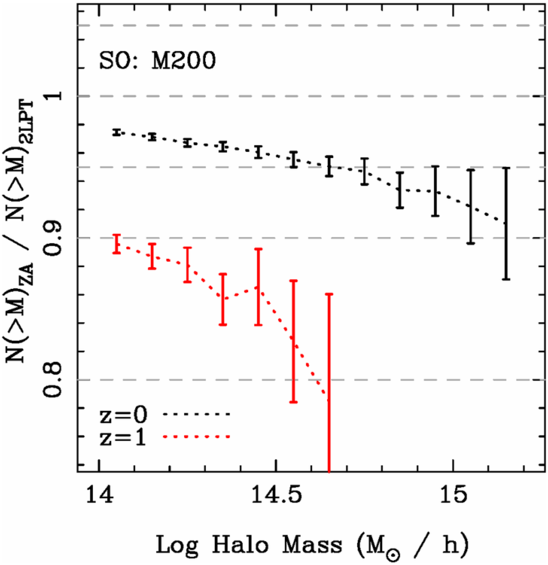
- Use Oriana, seed one with ZA
- preserve phases
- compare cumulative counts
- ACT/SPT: $4.9 \times 10^{14}\ h^{-1} M_{\odot}$



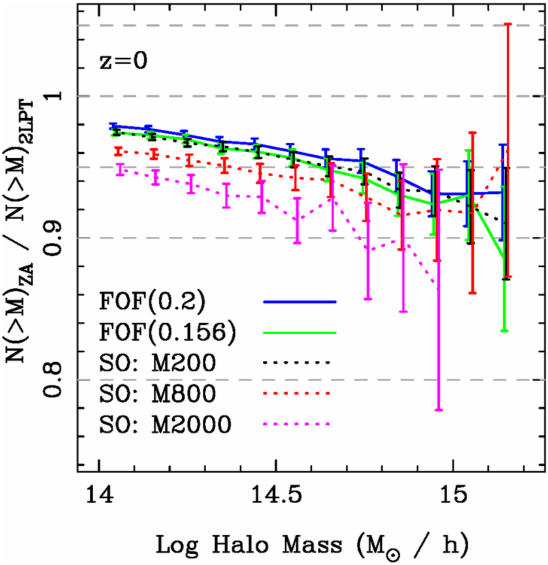
ZA vs 2LPT



ZA vs 2LPT

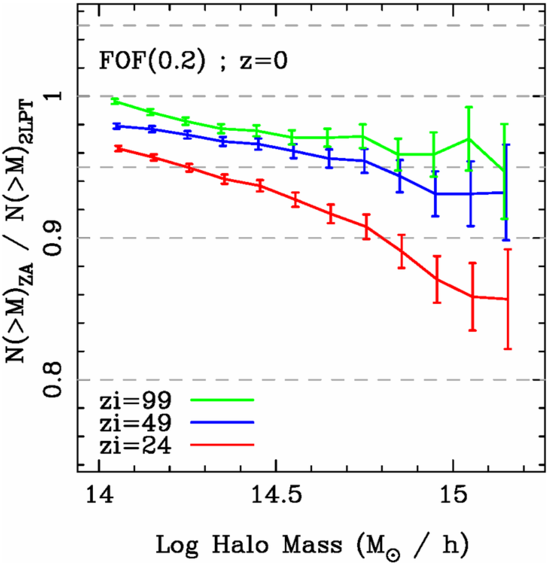


ZA vs 2LPT



z=0
FOF(0.2)
FOF(0.156)
SO: M200
SO: M800
SO: M2000
0.8
0.9
1
14
14.5
15
$N(>M)_{ZA}$ / $N(>M)_{2LPT}$
Log Halo Mass ($M_\odot$ / h)

ZA vs 2LPT

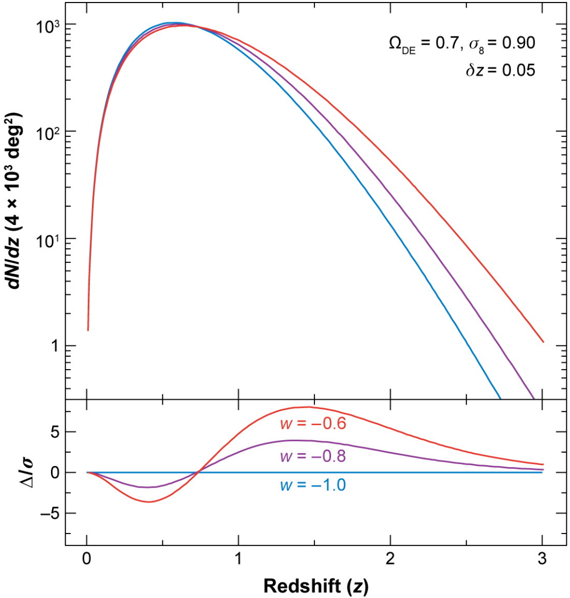

FOF(0.2) ; z=0
zi=99
zi=49
zi=24
N(>M)ZA / N(>M)2LPT
0.8
0.9
1
14
14.5
15
Log Halo Mass (M⊙ / h)

ZA vs 2LPT

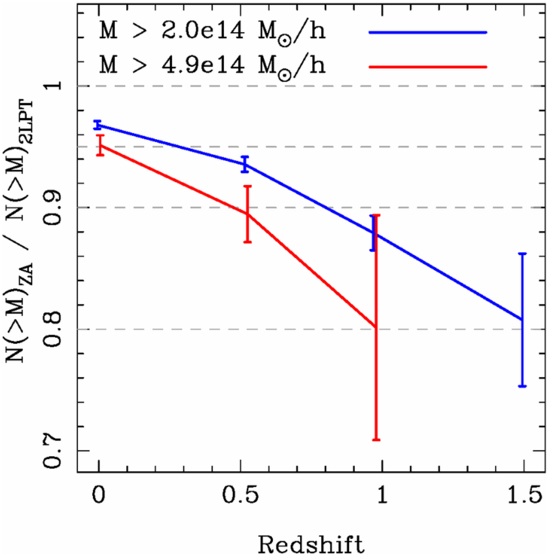
This seems to disprove some "common views":

- larger boxes can start at smaller z_{init} (Warren, et al)
- Higher z_{init} removes effect
- differences are insignificant, especially for cumulative functions

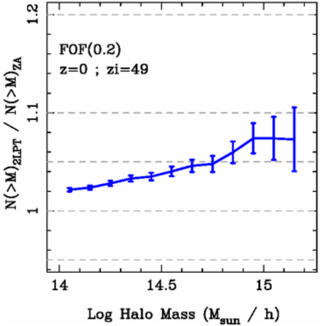
Effect with Redshift

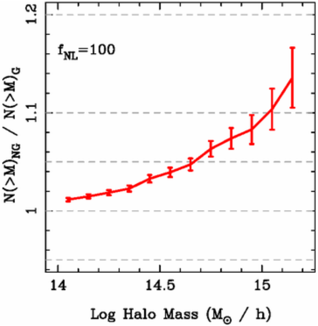


Degenerate with constraints on w



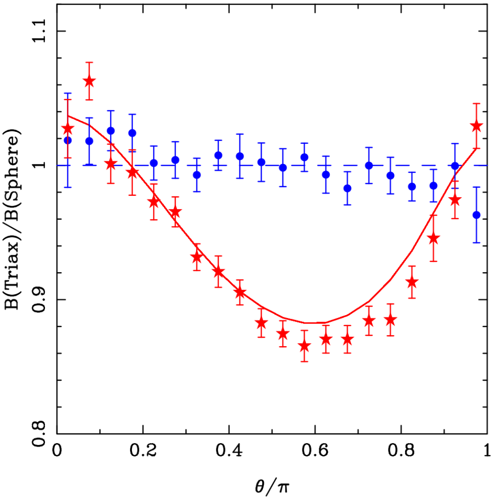
Can mimic primordial non-Gaussianity





LasDamas: in use!

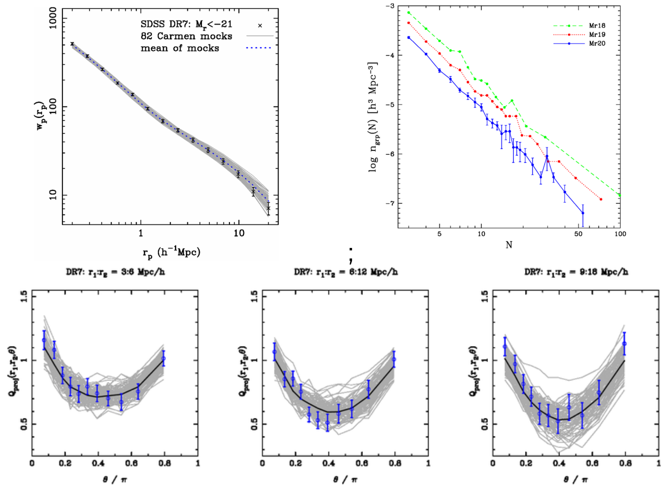
1. Constraining Halo Shape (McBride)
2. Joint Fitting: 2PCF, 3PCF, multiplicity (McBride)
3. primordial non-Gaussianity (Scoccimarro, Manera; Mao)
4. cosmology from redshift distortions (Kazin)
5. halo profiles: checking C(M) relation (Musher, AAS poster)
6. halo merger trees (Behroozi)
7. small scale galaxy clustering (Watson, Piscionere)
8. galaxy groups (Jackson)
9. ongoing BOSS analyses (Sanchez ; Tojeiro)



Smith, Watts & Sheth 2006

- B bispectrum: Fourier transform of the 3PCF

Project: Joint Fitting
self consistent fit over several statistics



THANKS!