Simulations and symmetries (arXiv:1910.07097)

Martin White w/ Chirag Modi & Stephen Chen.

A simple (practical) question:

Can we get a “model” for the real-space power spectrum of cosmological objects for use in fitting upcoming photometric/lensing surveys?

… that leads in some very interesting directions …

The problem (opportunity)

- Lots of reasons to want (real space) P(k,z) for biased tracers.
  - My interest arose out of trying to model galaxy clustering in photometric surveys in combination with CMB lensing, but the problem is more general.
- Quite a few approaches developed over the years.
- Two approaches dominate today:
  - Simulations + HOD models (or SAMs or hydro, ...)
    - Computationally expensive, hard to assess convergence, many parameters, but can (in principle) predict many things.
    - Connection with “physical” models of galaxy formation is relatively direct (galaxies live in halos).
  - Perturbation theory
    - Reasonably “cheap”, generally applicable, often gives insight but has limited dynamic range and statistics it can explain.

Can we combine the advantages of both approaches?

What can we simulate?

- The non-linearity of the dark matter field does not itself pose insurmountable difficulties.
  - The evolution of DM under gravity is a well posed numerical problem.
  - %-level accuracy on low order statistics can be obtained.
  - Emulators can be constructed.
- By contrast the behavior of the baryonic component, including hydrodynamics, star and BH formation and feedback, remains a challenge.
  - Despite decades of progress in algorithms, codes and computers, quantitative understanding eludes us.
  - We still need parameterized, phenomenological models!
  - Convergence is … an issue.

Symmetries

On sufficiently large scales all of these complexities can be parameterized by a series of numbers, the bias expansion, ...

- While the process that form and shape galaxies and other objects are complex, all such objects arise from simple initial conditions acted upon by physical laws which obey well-known symmetries.
- For non-relativistic tracers these are the
  - equivalence principle and translational, rotational and Galilean invariance.
- This highly restricts the kinds of terms that can arise, no matter how complex the history.

This kind of “symmetries based bias expansion” is commonplace in PT ...

Lagrangian bias expansion

- We will use a Lagrangian (rather than Eulerian) bias expansion
- Assume the probability that a particle in an N-body simulation will end up in a halo (or galaxy, or QSO, or ...) is a function of the initial density and gravitational potential field.
  - Simplest example: galaxies form from peaks in initial field.
  - Press-Schechter, BBKS, peaks bias, ...
- We assume this function is local, over some radius R ~ 1Mpc.
- Further assume we can Taylor expand this function for small $\boldsymbol{\delta}$.
- Working to 2nd order our variables are:

$$\delta^i = \{1, \delta_L, \delta_L^2, s_L^2, \nabla^2 \delta_L\}$$

Lagrangian bias model in practice

Take a simulation and its ICs. For each particle initially at **q** determine a weight which is a linear combination of our $\boldsymbol{\delta}^i$:

$$w(q) = 1 + b_1 \delta_L(\mathbf{q}) + b_2 \left(\delta_L^2(\mathbf{q}) - \langle \delta_L^2(\mathbf{q}) \rangle\right) + b_s \left(s^2(\mathbf{q}) - \langle s^2(\mathbf{q}) \rangle\right) + b_\nabla \nabla^2 \delta_L(\mathbf{q}).$$

Now move these particles to their final positions, **x** (using N-body dynamics), and “paint” them onto a grid with these weights.
This is the model for the density field of the biased tracer (e.g. galaxies).

This can *also* be written as:

$$\delta^b(\mathbf{x}) = \delta_{[1]}(\mathbf{x}) + b_1 \delta_{[\delta_L]}(\mathbf{x}) + b_2 \delta_{[\delta_L^2]}(\mathbf{x}) + b_s \delta_{[s^2]}(\mathbf{x}) + b_{\nabla^2} \delta_{[\nabla^2 \delta]}(\mathbf{x})$$

Power spectra

- It follows immediately that the power spectrum of any biased tracer, or the cross-spectrum between two such tracers, can be written as a sum over component spectra.
- Each component, P_{ij}, can be computed directly from the N-body simulation as the cross-spectrum of two fields (each on a grid: FFT), e.g.

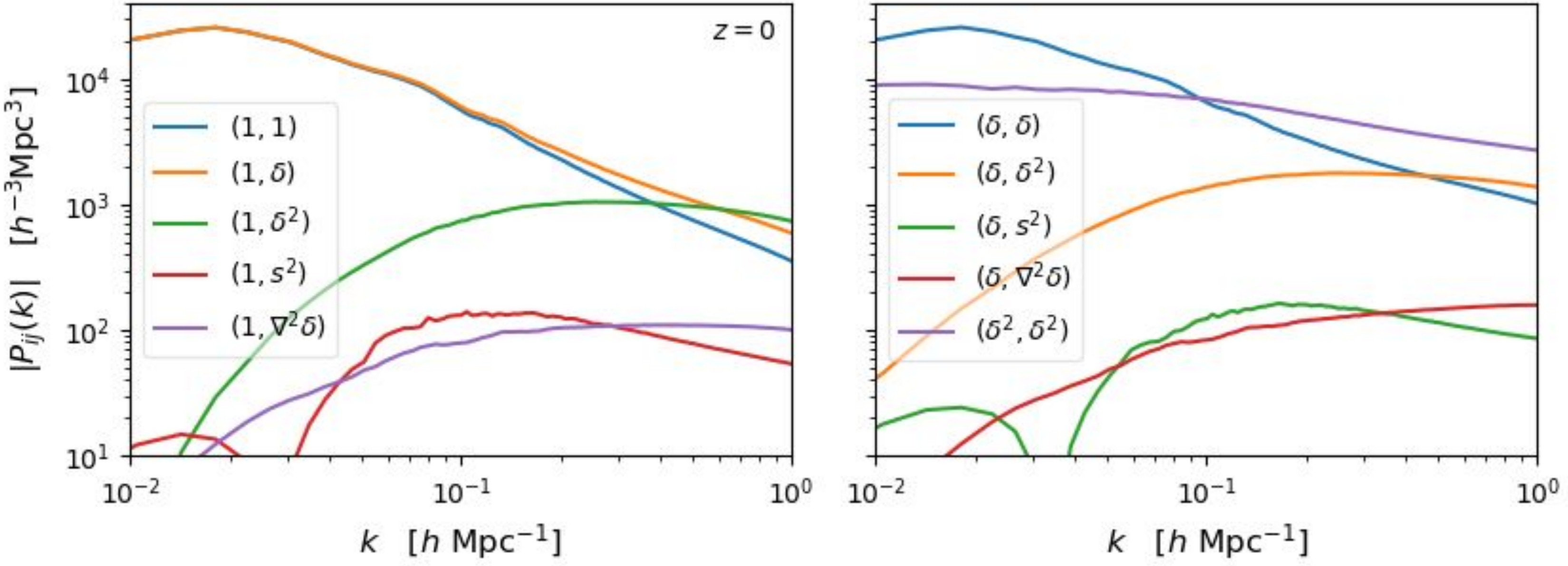
$$P_{1,\delta} = \left\langle \delta_{[1]}(\mathbf{k}) \delta^{\star}_{[\delta_L]}(\mathbf{k}) \right\rangle$$

-
- In general the cross spectrum between biased tracers *a* and *b* is:

$$P^{ab}(k) = \sum_{i,j} F^i_a F^j_b P_{ij}(k) + P_{SN}$$

$$F^i \in (1, b_1, b_2, b_s, b_{\nabla^2})$$

Component spectra



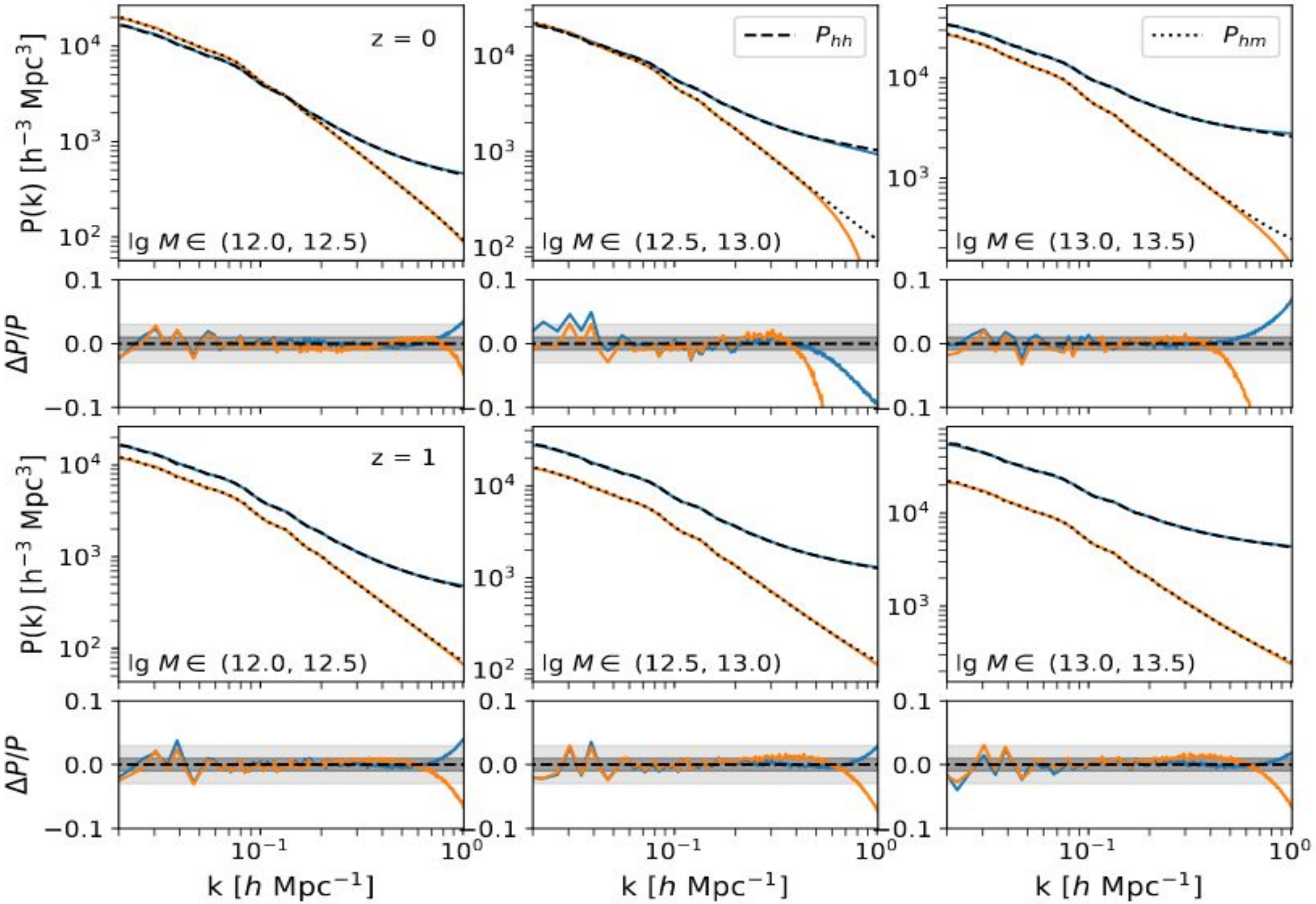
Any auto- or cross-spectrum is a linear combination of the “component spectra” ...

How well does it work?

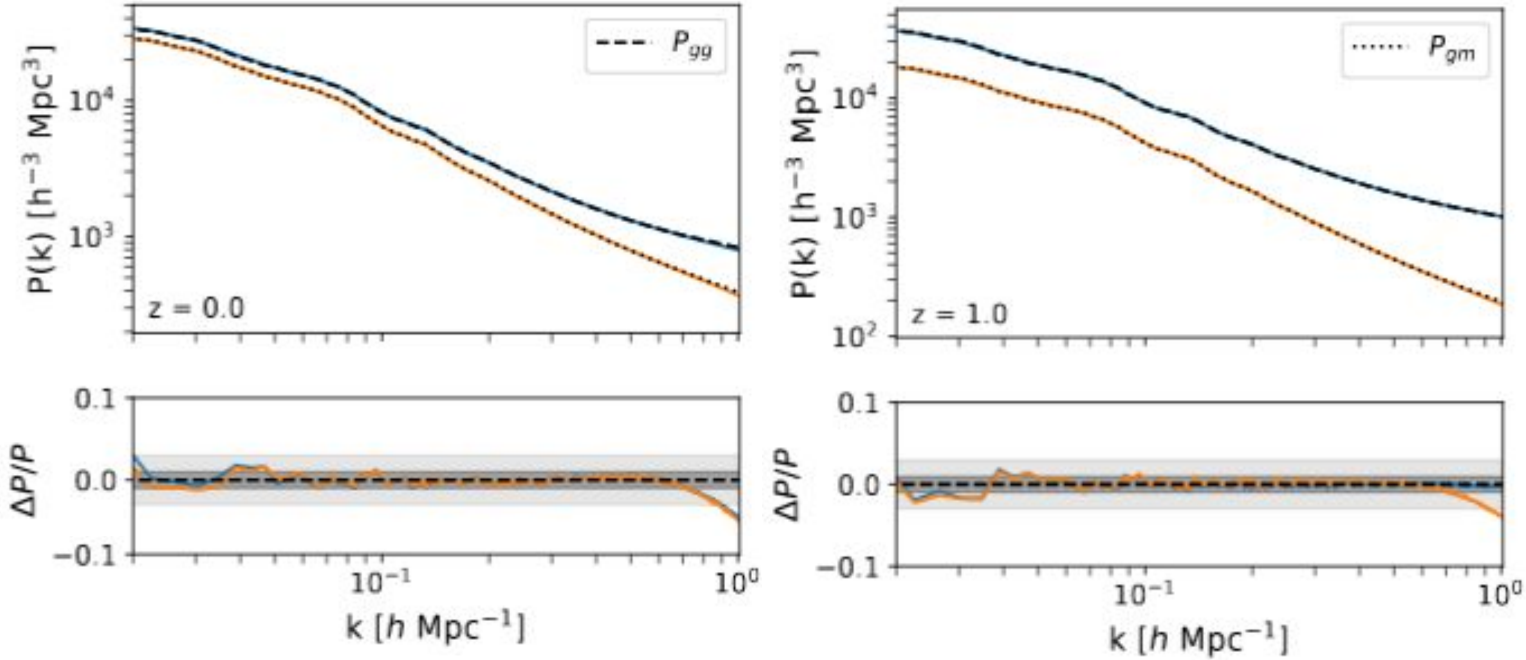
- The bias expansion can't work to infinitely small scales!
- How well does it work for "typical" halos and mock galaxies if we truncate to quadratic order?
- Start with a test on halos ... higher bias is harder ...

	$z = 0$	
$\log_{10} M$	$\bar{n}$	b
(12.0,12.5)	24.3	0.80
(12.5,13.0)	9.5	0.89
(13.0,13.5)	3.6	1.10

	$z = 1$	
$\log_{10} M$	$\bar{n}$	b
(12.0,12.5)	23.7	1.30
(12.5,13.0)	7.9	1.69
(13.0,13.5)	2.2	2.36



z = 0
z = 1
P_{hh}
P_{hm}
P(k) [h^{-3} Mpc^3]
ΔP/P
lg M ∈ (12.0, 12.5)
lg M ∈ (12.5, 13.0)
lg M ∈ (13.0, 13.5)
k [h Mpc^{-1}]



$$\langle N_{\rm cen}\rangle(M_h) = \frac{1}{2}\left\{1+\mathrm{erf}\left[\frac{\lg M/M_{\rm min}}{\sigma}\right]\right\}$$

$$M_{\rm min} = 10^{12.5}\, h^{-1} M_\odot$$

$$\sigma = 0.2\ dex$$

$$\langle N_{\rm sat}\rangle(M_h) = \Theta(M_h - M_{\rm min})\left(\frac{M_h - M_{\rm min}}{M_1}\right)^{\alpha}$$

$$M_1 = 20\, M_{\rm min}$$

$$\alpha = 0.9$$

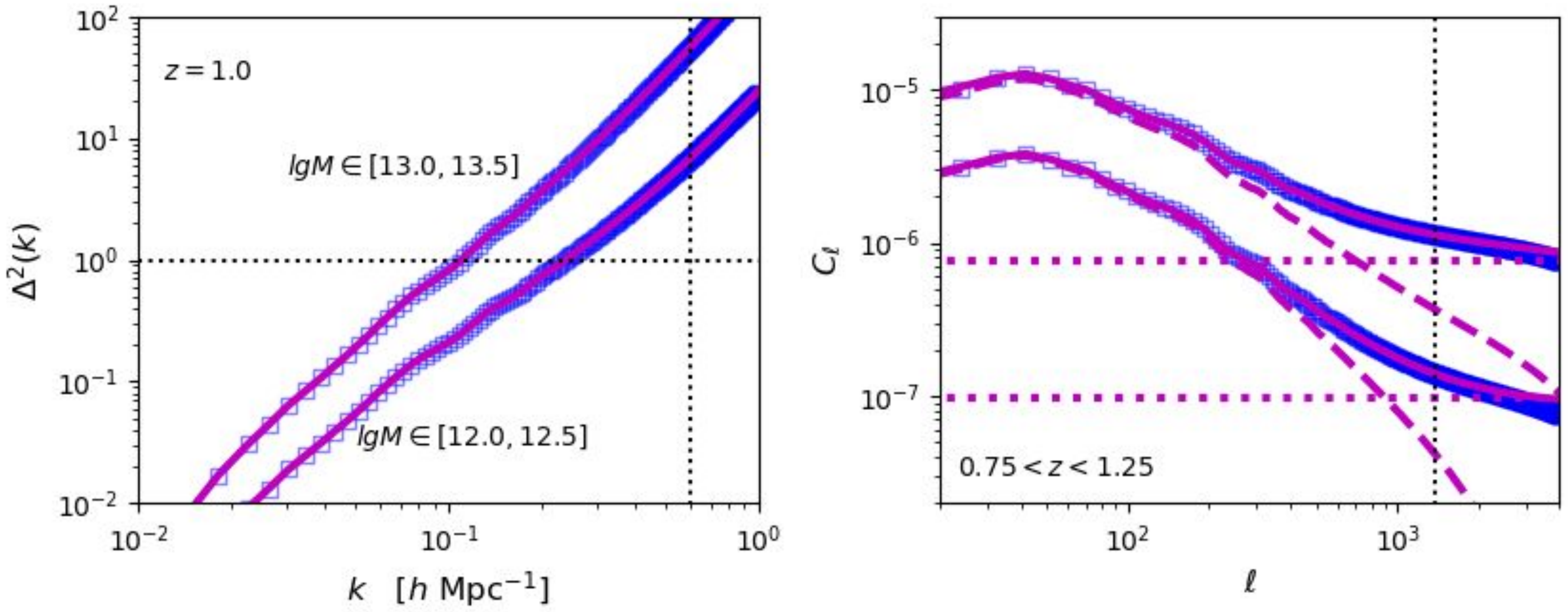
18% and 10% satellite fraction at z=0 and 1 respectively

Good enough?

- This is probably good enough for most “cosmology” applications.
- At higher k the field is highly non-linear
  - Modes become correlated with each other.
  - Modes become uncorrelated with their linear counterparts.
- For most samples, these scales are shot-noise dominated.
- At higher *k* modeling becomes increasingly complex.
  - Up to k~0.8h/Mpc can use k^2P approximation for baryons to ~1%.
  - At 2nd order we have 4 bias parameters at 3rd order double that!
- Anecdotally pushing to higher *k* mostly fixes nuisance params.

z	0.25	0.50	0.75	1.0
Lmax	400	800	1100	1400

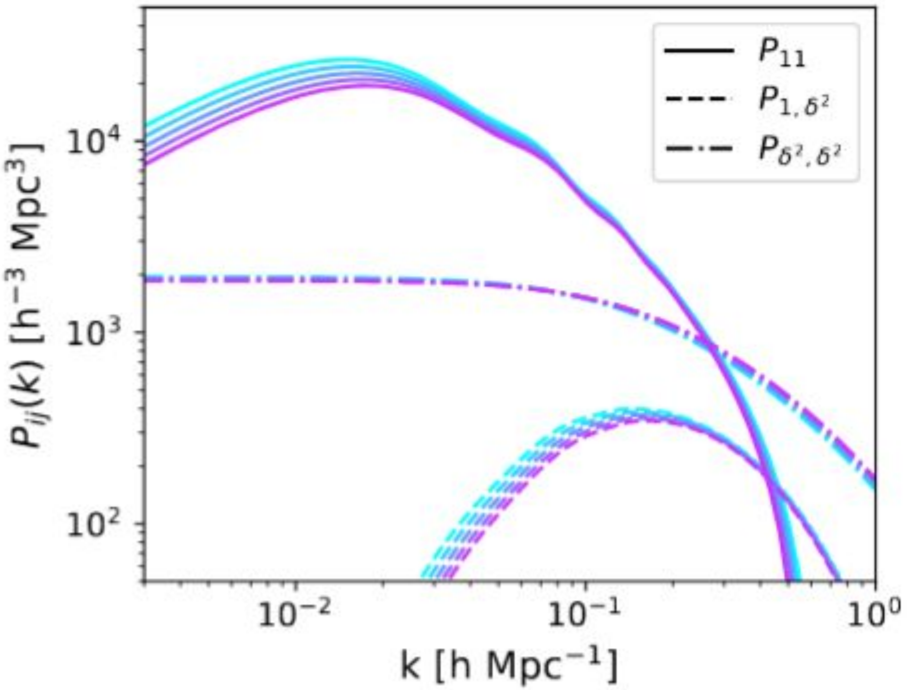
Good enough?



(Looks pretty similar at z<1)

But ... the emulator?

- So far this is all for 1 cosmology.
- How do we emulate?
- Same way as for P_m!
- The bias dependence is analytic, so only need cosmology parameters.
- Cosmology dependence of components similar to matter P(k) -- which we've done.
- Simulation requirements much relaxed ...

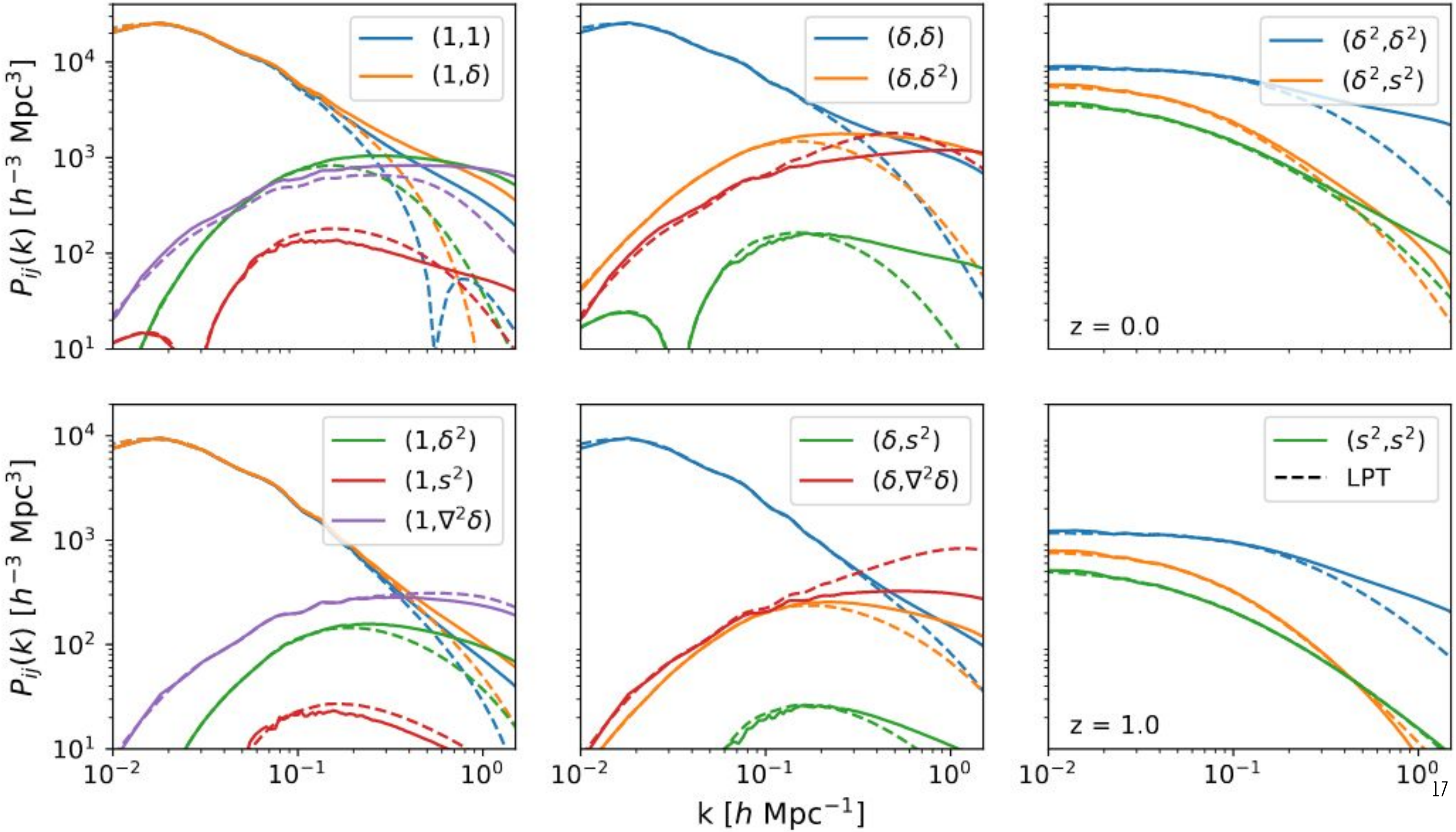


Variation in component spectra with ± 10% $\boldsymbol{\Omega_m}$

Conclusions

- The best of both worlds!
  - Accurate DM particle displacements from N-body.
  - Flexible and efficient bias expansion from PT.
  - Both are controlled approximations!
- For low z, and modest bias, doubles reach of PT and works well into the non-linear regime.
- Drastically reduces demands on the emulator framework.
  - Far fewer parameters, easier simulations (don't need halo shapes, no need to track subhalos, merger trees, etc.)
- Bias model actually works "at the field level", so other statistics should be emulatable as well (and mocks, samplers, …).

Thank You

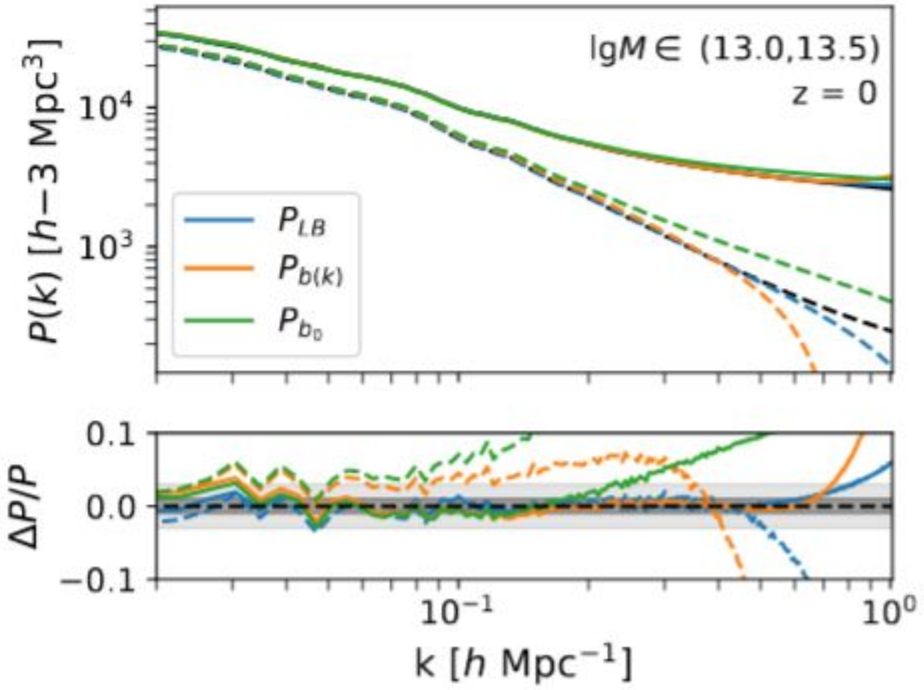


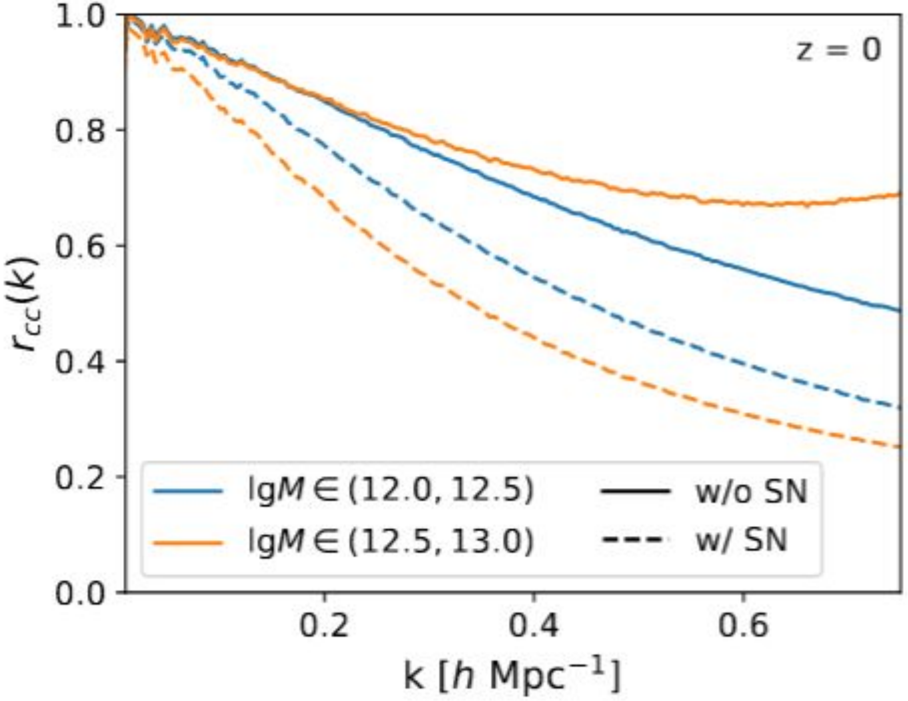
(1,1)
(1,δ)
(δ,δ)
(δ,δ²)
(δ²,δ²)
(δ²,s²)
z = 0.0
(1,δ²)
(1,s²)
(1,∇²δ)
(δ,s²)
(δ,∇²δ)
(s²,s²)
LPT
z = 1.0
$P_{ij}(k)$ [h^{-3} Mpc^3]
k [h Mpc^{-1}]

Contrast this with a "HaloFit model"

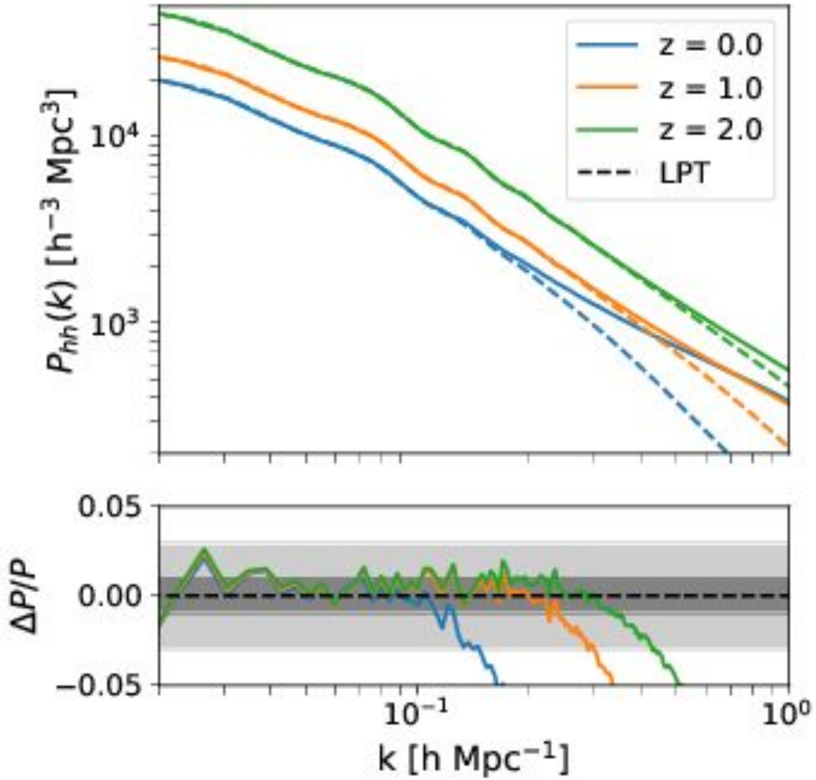
$$P_{hm} = \left[b'_0 + b'_1 k + b'_2 k^2\right] \ P_m(k)$$

$$P_{hh} = \left[b'_0 + b'_1 k + b'_2 k^2\right]^2 P_m(k) + P_{SN}$$





Comparison to (1-loop) LPT



- Can do a direct, theory-to-theory comparison of N-body derived P_{ij} and LPT derived P_{ij}.
- At low *k* they match very well.
- For low *z*, perturbative dynamics is more limiting than the bias expansion.
- At higher *z*, this is no longer the case.
- Emulator (roughly) doubles the range of the bias expansion over PT.

Simulations & Emulators

- This approach runs a "grid" of simulations
  - Find halos and subhalos (this is hard!)
- For (another) "grid" of HOD models (or <u>in principle</u> SAMs or hydro sims) populate sim with objects.
- Compute the statistics of interest.
- Use some form of interpolation to provide smooth predictions over the parameter space in an efficient manner.

- The number of samples, and the location of the samples, must be carefully chosen with an interpolation scheme in mind
  - High accuracy requires "many" simulations.
  - More parameters means more simulations.

Quite a few parameters ...

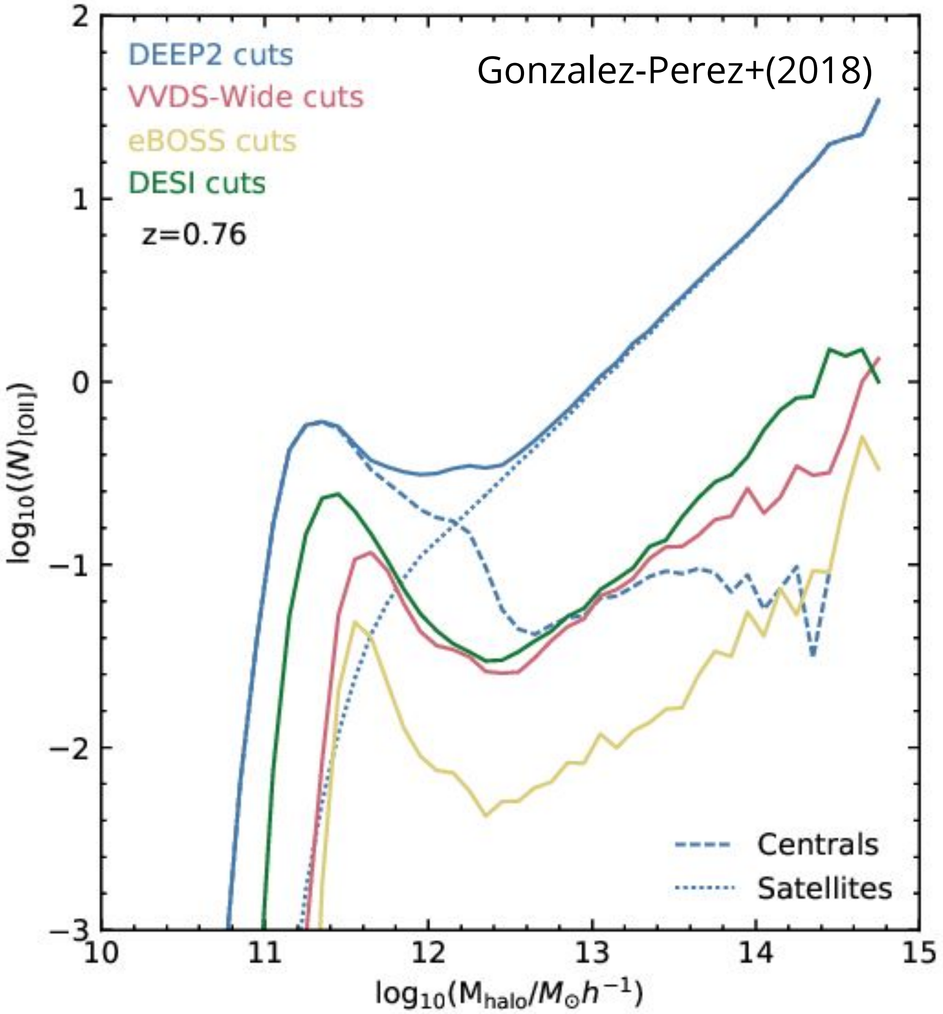
AEMULUS III: GALAXY CLUSTERING

Zhai+(2019)

	Parameter	Meaning	Range
Cosmology	Ω_m	The matter energy density	[0.255, 0.353]
	Ω_b	The baryon energy density	[0.039, 0.062]
	σ_8	The amplitude of matter fluctuations on 8 h^{-1}Mpc scales.	[0.575, 0.964]
	h	The dimensionless Hubble constant	[0.612, 0.748]
	n_s	The spectral index of the primordial power spectrum	[0.928, 0.997]
	$w^{\dagger}$	The dark energy equation of state	[-1.40, -0.57]
	$N_{\rm eff}{}^{\dagger}$	The number of relativistic species	[2.62, 4.28]
	$\gamma_f{}^{\dagger}$	The amplitude of halo velocity field relative to wCDM+GR	[0.5, 1.5]
HOD	$\log M_{\rm sat}$	The typical mass scale for halos to host one satellite	[13.8, 14.5]
	α	The power-law index for the mass dependence of the number of satellites	[0.2, 1.8]
	$\log M_{\rm cut}$	The mass cut-off scale for the satellite occupatioin function	[10.0, 13.7]
	$\sigma_{\log M}$	The scatter of halo mass at fixed galaxy luminosity	[0.05, 0.6]
	$\eta_{\rm con}{}^{\dagger}$	The concentration of satellites relative the dark matter halo	[0.2, 2.0]
	$\eta_{\rm vc}{}^{\dagger}$	The velocity bias for central galaxies	[0.0, 0.7]
	$\eta_{\rm vs}{}^{\dagger}$	The velocity bias for satellite galaxies	[0.2, 2.0]

Complex HODs

- Selections weighted by shape, color (star formation rate, emission line strength), photo-z, etc. often have quite "complex" HODs.
- HODs can vary strongly across photo-z bins.
- These require even more parameters to model.
- Add assembly bias, halo orientation, baryonic effects, ...



Gets hard ...

- This gets hard, and expensive, quite quickly.
- Validation and uncertainty quantification also become increasingly difficult.
- Due to the many ingredients being put in by hand, validation must be performed on each and every statistic, redshift, sample, etc., independently.

Nobody has successfully used an emulator for cosmic inference on a "modern" galaxy survey yet ...

Simulations and symmetries (arXiv:1910.07097)

Can we get a "model" for the real-space power spectrum of cosmological objects for use in fitting upcoming photometric /lensing surveys?

Our approach - Symmetries based bias model + N-body dynamics

- Marry two common approaches ...
- Adopt quadratic bias scheme that will work up to (about) the halo radius
- Advect bias-weighted distribution using "exact" N-body displacement

Simulations and symmetries (arXiv:1910.07097)

There are 2 types of people in cosmology:

1. **Theorists -** Symmetries based bias model + perturbative approaches for dynamics
2. **Simulators -** Halo based models + N-body dynamics (from simulations)
3. **Others -** people who disprove the first assertion.

We want to take the best of both "worlds", and marry them ...

Can we get a "model" for the real-space power spectrum of cosmological objects for use in fitting upcoming photometric surveys?

Our approach - Symmetries based bias model + N-body dynamics

- No need for complicated halo models
- Adopt quadratic bias scheme that will work up to about the halo radius
- Advect bias-weighted distribution using "exact" N-body displacement