# Breaking Galaxy-Halo Degeneracies with Anisotropic Small-Scale 3-Point Clustering

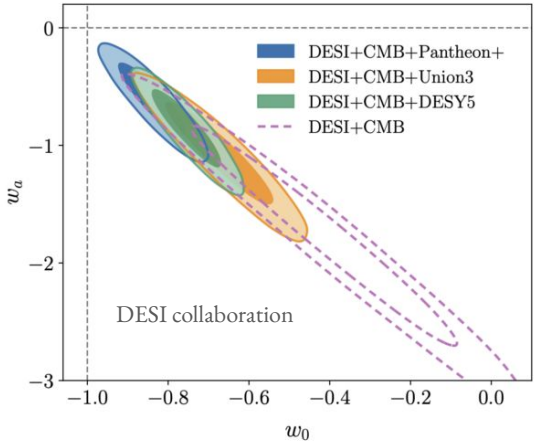
**Lado Samushia, Kansas State University**

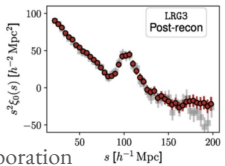
**GGI: Exploring New Frontiers in Cosmology, 23rd July 2026**

**In collaboration with: Zachary Brown, Nick Magnelli**

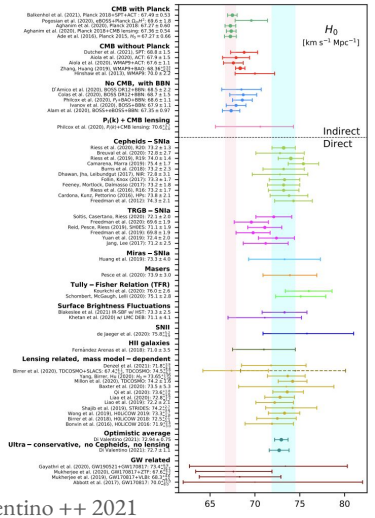
# Curse of the Finite Observable Universe

- Our flagship results, including the evidence for phantom dark energy rely on the analysis of 2-point correlation function on linear scales.
- The constraining power will increase with DR2 -> DR3 but not dramatically.
- The amount of information (the number of linear modes) in the visible universe is fundamentally limited (cosmic variance).
- To continue hunting the elephants we have to find new weapons.
- The two obvious frontieres are the higher order correlation patterns and nonlinear scales



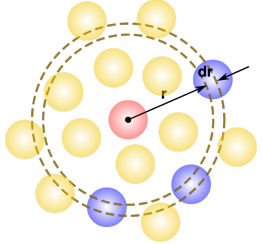



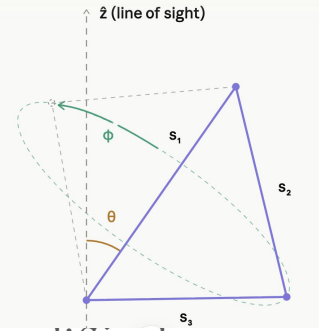

DESI collaboration




Di Valentino ++ 2021

# What kind of 3pt function?

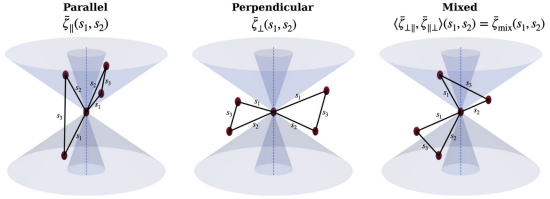


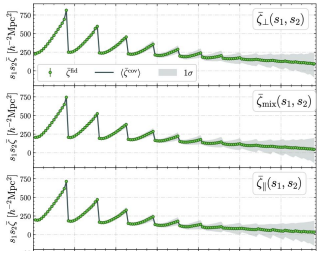

- The next obvious statistics to look at is the 3-point correlation function.
- Its analysis is associated with a number of inconveniences: it is slow to compute, difficult to model, and results in a very long data vector (2pt CF a 2D function, 3pt CF a 5D function).
- Historically the main bottleneck for npt function computation has been the cost of computation. Many ingenious algorithms have been developed to accelerate the computation.
- The fast algorithms sometimes “force” you into a specific basis.
- If CPU time was not an issue, what would you like to compute?
    - Clearly differentiate  line-of-sight and flat triangles. LOS differential clustering carries a lot of information through RSD
    - Coarse binning in triangle sides since there is no feature to resolve only slow variance of amplitude with shape
    - Coarse binning in LOS angle (for 2pt correlation function 2 multipoles are usually enough)
    - Make modeling easier by excluding small scales in all three sides
    - Make modeling easier by excluding nearly LOS sides

# Read our paper (Brown & Samushia) 2606.19518 ---->

- A new algorithm -- TriCo.
- Trico Measures: three classes of triangles (par,par); (per,per); (par,per) + (per,par); plus a coarse binning in the triangle sides $s_1$, $s_2$, $s_3$, enforcing $s_i > s_{min}$
- The algorithm uses some obvious accelerators (e.g. divides volume into subvolumes and only traverses neighboring subvolumes), but is otherwise slow.
- But we don't care. We submit the jobs to our Z8 workstation and check back in two weeks to make plots and write the paper.
- ~ 400-long vector for triangles between 5 and 80 Mpc/h separation.

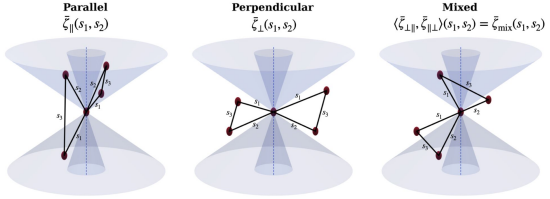



https://github.com/ladosamushia/trico

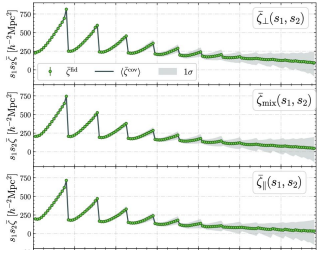


# Read our paper (Brown & Samushia) 2606.19518 ---->

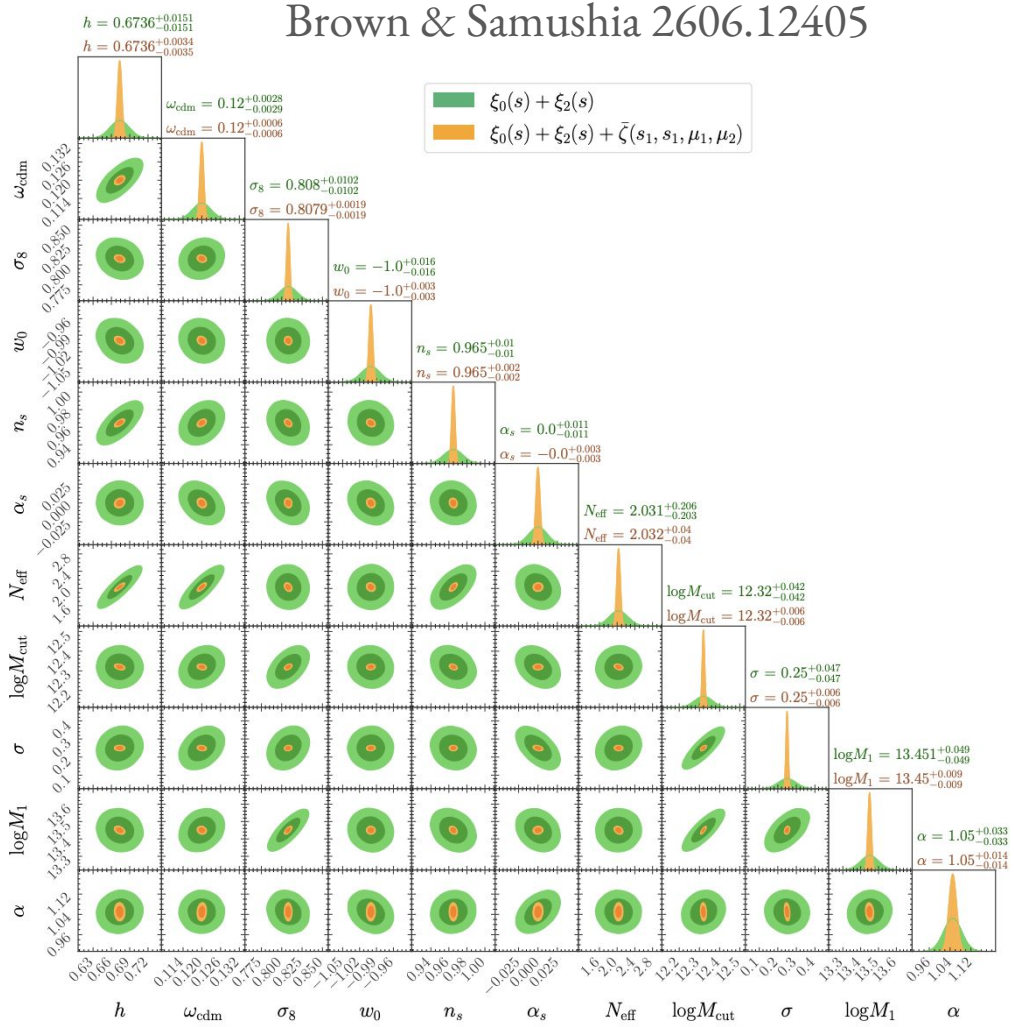
- A new algorithm -- TriCo.
- Trico Measures: three classes of triangles (par,par); (per,per); (par,per) + (per,par); plus a coarse binning in the triangle sides $s_1$, $s_2$, $s_3$, enforcing $s_i > s_{min}$
- The algorithm uses some obvious accelerators (e.g. divides volume into subvolumes and only traverses neighboring subvolumes), but is otherwise slow.
- But we don't care. We submit the jobs to our Z8 workstation and check back in two weeks to make plots and write the paper.
- ~ 400-long vector for triangles between 5 and 80 Mpc/h separation.
- Unsolicited advice to early career scientists: Don't optimize your code (unless you are writing something like CAMB or CLASS or GADGET).
- Write a clear code that does exactly what you want. Run it long. Enjoy life while it runs.

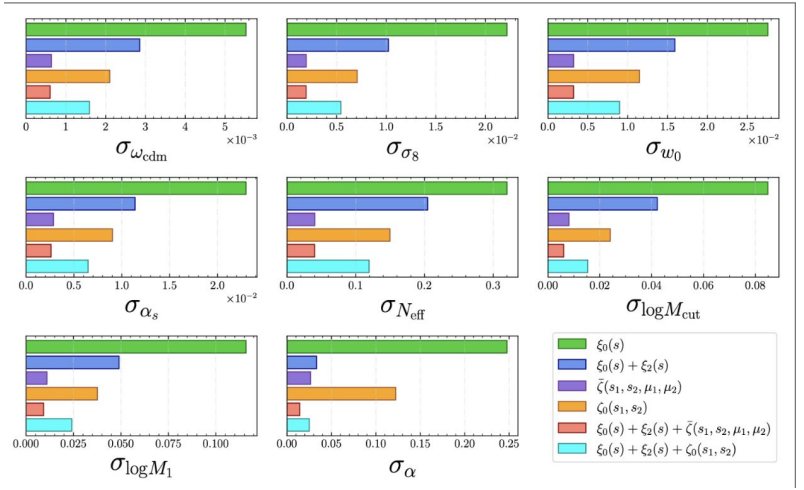



https://github.com/ladosamushia/trico



# Brown & Samushia 2606.12405

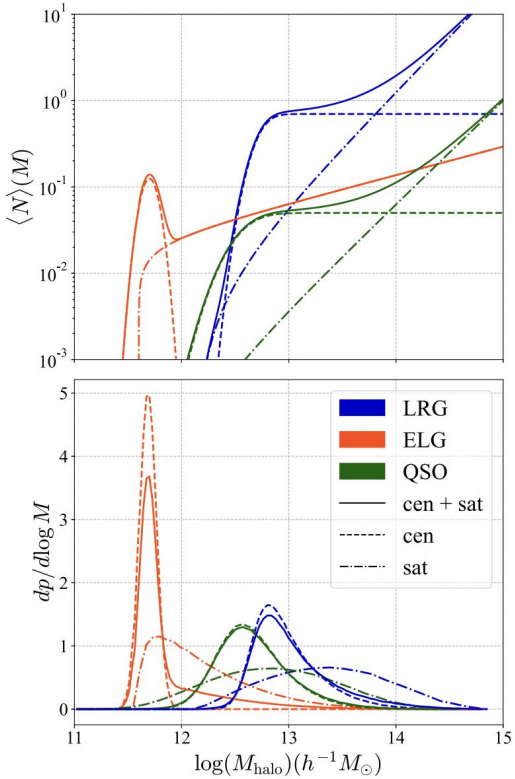



- Significant increase in the constraining power on all cosmological parameters and HOD parameters.
- The line-of-sight separation of triangles significantly enhances the constraints
- The useful information is not concentrated in any specific triangular shape or configuration. It is rather almost uniformly distributed between all triangles.

# Small Scale Clustering Analysis: Challenges

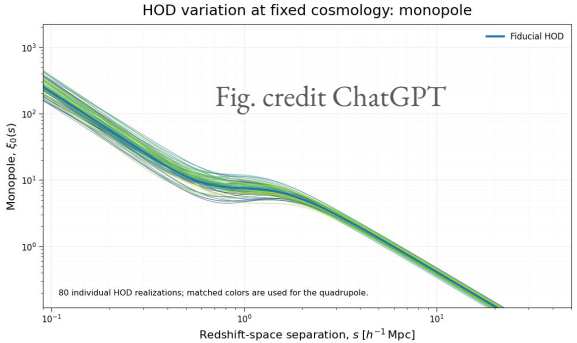
- Needs to be simulation based (Perturbation theory is not going to work).
  - Associated problems: volume of simulations, fidelity of simulations, coverage of cosmological parameter space, etc.
- Needs to somehow account for the connection between halos and galaxies (N-body simulations produce halos, we observe galaxies).
- A possible solution is to have hydro-simulations that actually produce galaxies but for hydro-simulations all of the above problems (volume, fidelity, coverage) become more severe.
- Current most popular approach: N-body simulations + HOD++ (Only mass dependence in the simplest implementation but can be made arbitrarily complex)

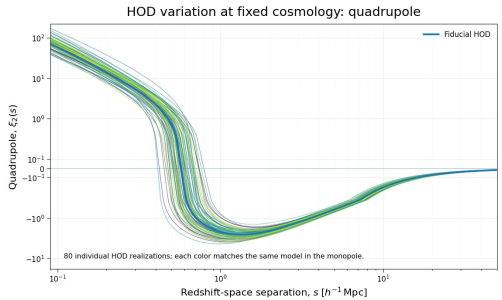



Zhang, Samushia, et al. 2203.17214

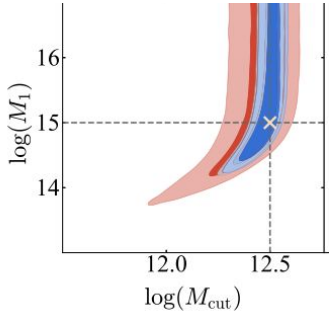
# Skeptic's view of HOD modelling (Levels of confidence)

- The fundamental problem: multiple correlation functions are possible within the same cosmology (not a unique prediction) and unlike the matter correlation function they can not be computed from cosmological parameters alone.
- The small-scale correlation function measurements are insanely precise. Can we ever hope to understand the galaxy formation physics to trust predictions on the correlation function to a <1% precision no matter how complicated the HOD model or the simulation is?
- Main assumption: galaxies reside in halos ← This is most certainly true. But ...
- When you run MCMC chains over HOD parameters not only do you trust the model predictions but you also trust that they properly weight the prior space of possible correlation functions in a given cosmology (related to the prior volume effects). This could be scary for a skeptic.
- If you are in the correct cosmology, one of the HOD parameter sets will give you close enough correlation function to the true galaxy correlation function in that cosmology. This is something even a skeptic will probably agree to.




Zhang, Samushia, et al. 2203.17214



# Read our paper (Magnelli, Brown & Samushia) 2606.12405 --->

- If we magically new HOD parameter values we could extract all cosmological information from small-scale 2pt CF <- The “ceiling”.
- If you are maximally but reasonably skeptical about HOD modelling you want to exclude smaller scales (< 5 Mpc/h?) and do profile likelihood instead of Bayesian marginalization <- The “floor”.
- The “ceiling” is the theoretical (unobtainable) ceiling.
- The “floor” is the most conservative and cautious approach (perhaps overly conservative).
- Based on AbacusSummit simulation suit that covers a large cosmological model space we found that there are several orders of magnitude in constraining power between the ceiling and the floor.
- Most cosmologies (except for very nearby ones) would be ruled out for the ceiling
- Most cosmologies (except for a few extremes) are consistent with Planck LCDM for the floor. Because you can always find some HOD values that bring you close enough to it from a different cosmology.
- Two ways of interpreting this:
  - pessimist: small-scale clustering can not be used in cosmology
  - optimist: It is crucial to understand galaxy-halo connection so that we can do amazing co