

# 1 Introduction

We want to make a convincing mock. To do this, we will take an N-body sim (the 1Gpc MDPL) and put galaxies in it using the [Behroozi et al., 2010] prescription.

$$\log_{10}(M_{\text{halo}}) = \log_{10}(M_1) + \beta \log_{10}\left(\frac{M_*}{M_{*,0}}\right) + \frac{\left(\frac{M_*}{M_{*,0}}\right)^\delta}{1 + \left(\frac{M_*}{M_{*,0}}\right)^{-\gamma}} - \frac{1}{2} \quad (1)$$

We also add a mass varying scatter

$$\sigma = aM_{\text{halo}} + b \quad (2)$$

From [Leauthaud et al., 2012] we know the parameters for the functional form are approximately  $10^{12.52}, 10^{10.91}, 0.45, 0.6, 1.83$  and from my paper we know that the scatter is approximately 0.3 and 0.18 at  $M_{\text{halo}} = 10^{13}$  and  $10^{15}$ .

However, we want to make sure that the mock matches HSC as closely as possible. We will therefore optimize these parameters to fit the SMF and clustering, beginning from these starting points.

## 2 Optimizing model

### 2.1 $M_{\text{halo}}$ vs $V_{\text{max}}@M_{\text{peak}}$

Initially we modelled (as in the previous equation)  $M_*$  as a function of  $M_{\text{halo}}$  and  $M_{\text{halo,peak}}$ . However, per Peter we should instead be using  $V_{\text{max}}@M_{\text{peak}}$ . I think this is because  $M_{\text{halo}}$  begins to be stripped before a halo merger takes place, while  $V_{\text{max}}@M_{\text{peak}}$  continues to increase (or at least stays constant). This causes a problem as we now need a functional form to relate  $M_*(V_{\text{max}}@M_{\text{peak}})$

Figure 1 shows the relation between  $V_{\text{max}}@M_{\text{peak}}$  and  $M_{\text{halo}}$ . This is roughly linear in log space, so we can write:

$$\begin{aligned} \log_{10}(V_{\text{max}}@M_{\text{peak}}) &= m \log_{10}(M_{\text{halo}}) + c \\ \log_{10}(V_{\text{max}}@M_{\text{peak}}) &= m * (\log_{10}(M_1) + \beta \log_{10}\left(\frac{M_*}{M_{*,0}}\right) + \frac{\left(\frac{M_*}{M_{*,0}}\right)^\delta}{1 + \left(\frac{M_*}{M_{*,0}}\right)^{-\gamma}} - \frac{1}{2}) + c \end{aligned} \quad (3)$$

We now have two options: 1) Use this new function with the known  $m$  and  $c$  and the old starting parameters 2) Numerically find (guess) a new starting point for the five numbers that folds  $m$  and  $c$  into them. I did this.

This is not purely guessing. We can rationalise that  $M_1$  will decrease significantly (as  $V_{\text{max}}@M_{\text{peak}} < M_{\text{halo}}$ ),  $M_{*,0}$  will stay the same as the kink will be in about the same place. We can then guess at  $\beta, \delta, \gamma$ . We plot a couple of guesses at this and get the params of:  $10^{2.4}, 10^{10.91}, 0.45, 0.3, 0.2$ . We can use Figure 1 to map the scatter params. We parameterize this as a slope and a y intercept.

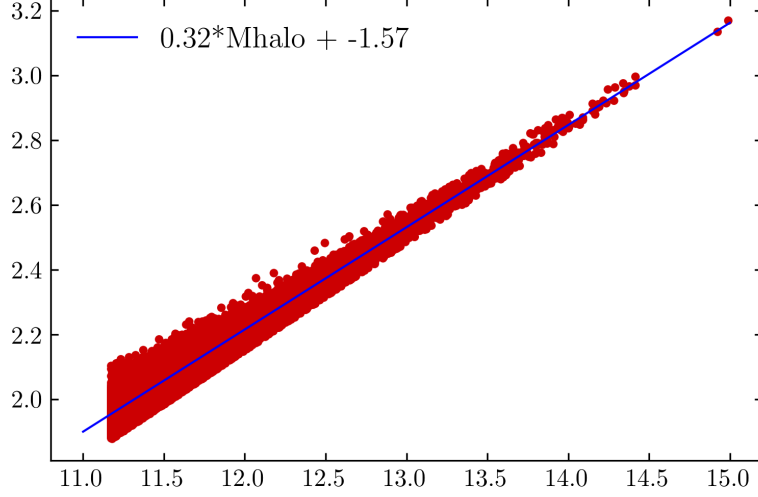


Figure 1:  $V_{\text{max}}@M_{\text{peak}} - M_{\text{halo}}$  relation.

In mass space this would be  $-0.06, 1.08$ . In velocity space it is approximately  $-0.1, 0.5$ .

We also note that Equation 1 relates  $M_{\text{halo}}(M_*)$ , or  $V_{\text{max}}@M_{\text{peak}}(M_*)$ . We numerically invert this to get the inverse.

However, this functional form is designed to fit the full SM-HM range. As we are only fitting the very high mass end it is absolute overkill. This is problematic for two reasons, 1) optimization/sampling will take longer because we have a 7 dimensional space, 2) Optimization and sampling will struggle because of the degeneracies/useless parameters.

At the high mass end, the SMHM is basically just a line, though with a bit of curvature. We therefore want to keep 3 parameters. We will fix the other two at their initial points.

$$\log_{10}(V_{\text{max}}@M_{\text{peak}}) = \log_{10}(M_1) + \beta \log_{10}\left(\frac{M_*}{M_{*,0}}\right) + \frac{\left(\frac{M_*}{M_{*,0}}\right)^\delta}{1 + \left(\frac{M_*}{M_{*,0}}\right)^{-\gamma}} - \frac{1}{2} \quad (4)$$

$$\log_{10}(V_{\text{max}}@M_{\text{peak}}) = \log_{10}(M_1) + \beta \log_{10}(c_1 M_*) + \frac{(c_1 M_*)^\delta}{c_2} - \frac{1}{2}$$

We can fit a line + some power

## 2.2 HaloTools?

Halotools is awesome and we should use it as much as possible. It also has a number of functions for creating a [Behroozi et al., 2010] model. This seems great!

However, I didn't know about this when I started working on it and the actual part of the my code that halotools would take the part of is tiny. It is just the mapping of  $V_{\text{max}}@M_{\text{peak}}$  to  $M_*$ . I have compared to ensure that my code and halotools are identical. The only difference is that my code assumes that all parameters (*e.g.*  $M_1$ ) have no little h. While in halotools they are assumed to have littleh of 0.7. Both will work fine they will just give different values for the same numerical values.

## 2.3 Comparisons to observations

We have an HSC galaxy catalog and an HSC SMF. Using the galaxy catalog we construct an estimate of clustering the 'counts in cylinders'. For this we define a cylinder 1Mpc in radius and 10Mpc in half length. We then pick a mass cut, *e.g.*  $10^{11}$ . Around each galaxy larger than the mass cut we count the number of galaxies between the mass cut and 0.1 dex below it. From this we construct a clustering signal using the Landay Szalay estimator using a random catalog.

We do this for a number of mass cuts.

## 2.4 Optimization

Given a halo catalog (MDPL) and a way to add galaxies to them (see previous 2 sections) we can now generate a mock. Each time we generate a mock we compute an SMF and an 'counts in cylinders' clustering signal analagous to what we have in data. We want the SMF and the clustering to be similar to the data.

To optimize this we assume that all points are uncorrelated and try to minimize error, scaled by the uncertainties on each point.

Note that we don't use MCMC to optimize as that is not what it is good at [Hogg and Foreman-Mackey, 2018]. It is slow and also a bad optimizer for high dimensional problems [Betancourt, ]. A downside of this though is that we don't get uncertainty estimates and so we don't know if some other parameter set is just a bit worse than the bestfit or is terrible. If we need that I can run an MCMC.

After running this, the best fit model has params:

2.379, 10.92, 0.356, 0.185, 0.278, -0.110, 0.519

This fits the SMF and clustering well (see Figure 2, Figure 3).

## 3 Tests

While we have shown that the functional form is flexible enough to fit the HSC SMF and clustering, this doesn't mean that the resulting mock is a good model.

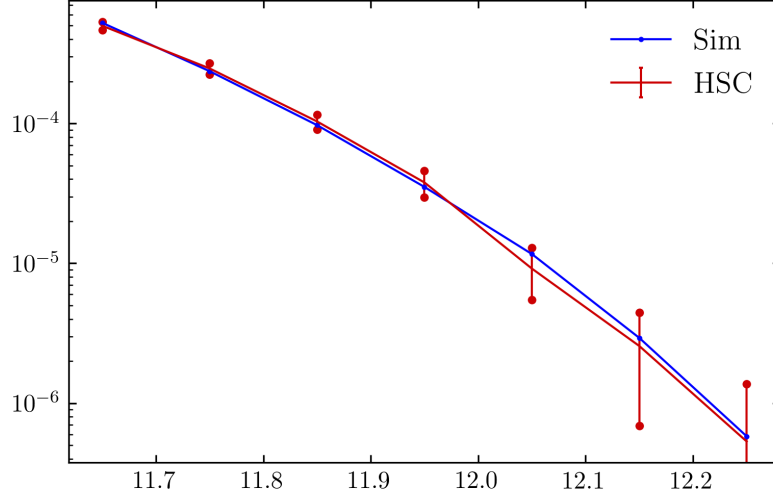


Figure 2: Fit to SMF - pretty good

It might have got a bunch of other important things that we didn't fit for wildly wrong. Let's check a couple of those things, and also just sanity check the mock.

### 3.1 Scatter

Based on how we parameterize scatter (linear decrease vs  $V_{\text{max}}@M_{\text{peak}}$ ), we expect for it to linearly decrease with increasing  $M_{\text{halo}}$ . We also expect it to roughly match the numbers we found in the UM. It does – see Figure 4.

### 3.2 Two point correlation function

Does our two point autocorrelation function look good? Compare to [Tinker et al., 2017] who used CMASS. See Figure 5.

### 3.3 Satellite Fraction

We want the fraction of galaxies at a given stellar mass that are satellites to match well with HSC. This is important because if we use this mock to test a cluster finder, an important thing a cluster finder does is identify centrals and satellites.

However, how does one define a central and a satellite? In a simulation it is easy – Rockstar does it for you. But you can't compare Rockstar satellites to anything in observations...

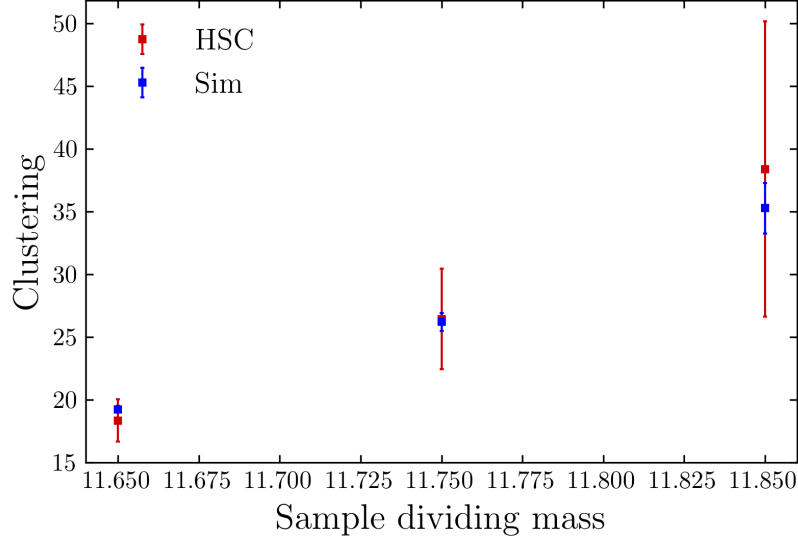


Figure 3: Fit to clustering - pretty good

Again we can go with a cylinder based approach. We define anything that is in a cylinder of a given size centered on a larger (by stellar mass) galaxy as a satellite. This can be done both in the mock and the data and so can be compared. This is shown in Figure 6.

This feels pretty tightly linked to clustering. I'm not sure how surprised we should be that this matches if we fit clustering.

### 3.4 Mag Gap

Related to the satellite fraction, we can look at the magnitude gap. I'm not really sure whether we can compare to obs because I don't know how we would see this?

## References

- [Behroozi et al., 2010] Behroozi, P. S., Conroy, C., and Wechsler, R. H. (2010). A Comprehensive Analysis of Uncertainties Affecting the Stellar Mass-Halo Mass Relation for  $0 < z < 4$ , 717:379–403.
- [Betancourt, ] Betancourt, M. A conceptual introduction to hamiltonian monte carlo.

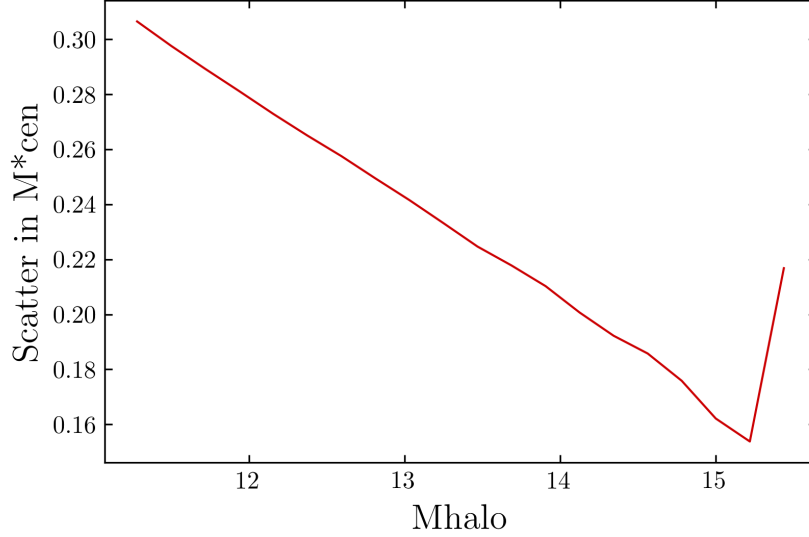


Figure 4: Looks like what we expect from the scatter params. Ignore the kick up at the end, I'm pretty sure the uncertainty on that point is huge.

[Hogg and Foreman-Mackey, 2018] Hogg, D. W. and Foreman-Mackey, D. (2018). Data analysis recipes: Using markov chain monte carlo. , 236:11.

[Leauthaud et al., 2012] Leauthaud, A., Tinker, J., Bundy, K., Behroozi, P. S., Massey, R., Rhodes, J., George, M. R., Kneib, J.-P., Benson, A., Wechsler, R. H., Busha, M. T., Capak, P., Cort  s, M., Ilbert, O., Koekemoer, A. M., Le F  vre, O., Lilly, S., McCracken, H. J., Salvato, M., Schrabback, T., Scoville, N., Smith, T., and Taylor, J. E. (2012). New Constraints on the Evolution of the Stellar-to-dark Matter Connection: A Combined Analysis of Galaxy-Galaxy Lensing, Clustering, and Stellar Mass Functions from  $z = 0.2$  to  $z = 1$ . , 744:159.

[Tinker et al., 2017] Tinker, J. L., Brownstein, J. R., Guo, H., Leauthaud, A., Maraston, C., Masters, K., Montero-Dorta, A. D., Thomas, D., Tojeiro, R., Weiner, B., Zehavi, I., and Olmstead, M. D. (2017). The correlation between halo mass and stellar mass for the most massive galaxies in the universe. , 839:121.

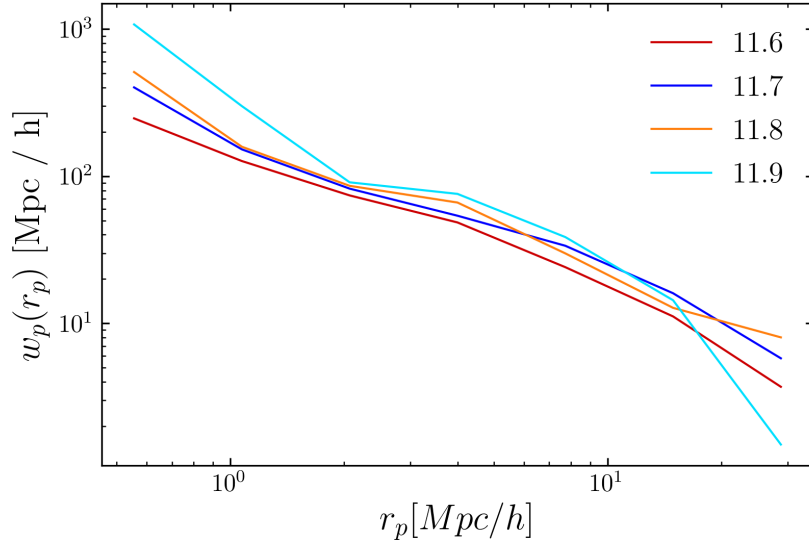


Figure 5: Satellite fraction with 2 methods of cluster finding. The first just uses Rockstar cen/sat. The second constructs cylinders 1MPC in radius and 10MPC half length. A galaxy is a satellite if it is in the cylinder of a larger galaxy. The mock matches HSC well. I'm not sure how surprised we should be about this as it seems to follow pretty directly from clustering?

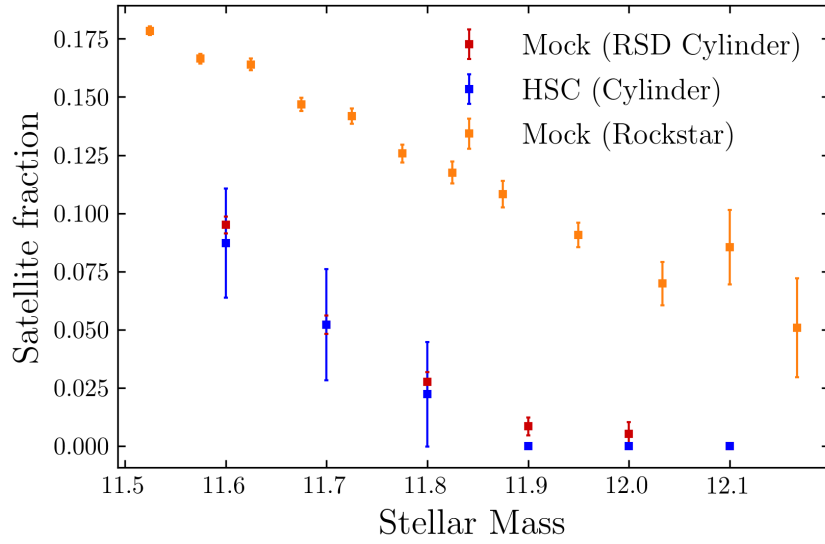


Figure 6: Satellite fraction with 2 methods of cluster finding. The first just uses Rockstar cen/sat. The second constructs cylinders 1MPC in radius and 10Mpc half length. A galaxy is a satellite if it is in the cylinder of a larger galaxy. The mock matches HSC well. I'm not sure how surprised we should be about this as it seems to follow pretty directly from clustering?



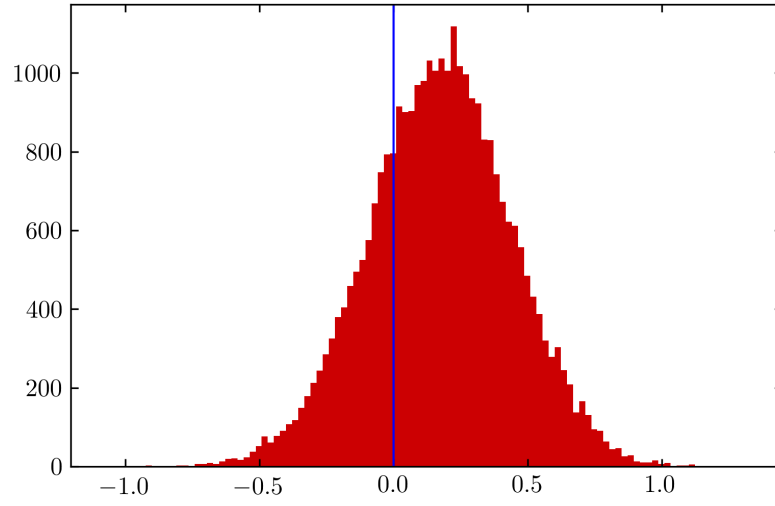


Figure 7: Difference in SM between the central and the largest satellite for halos  $> 14$

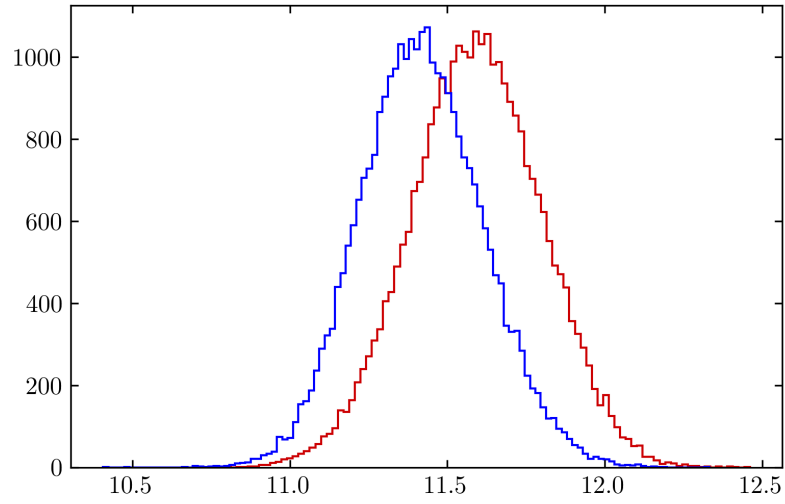


Figure 8: Distribution of SM for centrals and satellites for halos  $> 14$