Discovering the True Regulator of Stellar Mass

Andrew P. Hearin

9 February 2015

ABSTRACT

Empirical models of the galaxy-halo connection are built upon the assumption that some host halo property serves as the statistical regulator of the stellar mass M_* content of the halo. This assumption provides the key ingredient permitting us to use N-body simulations in cosmological parameter inference. While the true statistical regular of M_* is not known, the most widespread choice is halo mass $M_{\rm vir}$, although $V_{\rm max}$ is also common. There are significant, systematic differences between $V_{\rm max}-$ vs. $M_{\rm vir}-$based models, and it has recently been shown that the associated systematic error known as *galaxy assembly bias* threatens the precision cosmology program. In these notes I introduce the weighted cumulative abundance (WCA) technique, which allows *the choice of halo property itself* to be parameterized and continuously varied in a likelihood analysis. I describe how to use WCA to generalize abundance matching models as well as halo occupation distribution (HOD) models to make them more robust to assembly bias. I conclude by discussing the role that the WCA technique will play in future constraints on cosmology with large-scale structure.

1 INTRODUCTION

The first step in all traditional formulations of empirical models is to choose a particular halo property upon which the model is based. In AM-type models, this property is commonly something like $V_{\rm peak}$ or $M_{\rm peak}$ of both hosts and subs, while in HOD/CLF-type models this property is usually host halo $M_{\rm vir}$.

Due to assembly bias, and particularly its scale-dependence, models formulated on one halo property need not necessarily map onto models based on a different property. So it may not be possible to get a good χ^2 fit to a model based on one halo property if the model in your likelihood analysis is based on a different halo property. This is the bird's-eye-view upshot of the Phantom Menace exercise.

The problem is that we don't know what the true halo property in Nature is, and this lack of knowledge poses a systematic for both galaxy evolution and cosmology applications. These notes show how to build models based on continuous variations between any finite number of different halo properties. The AM formulation of the generalization is in § 2, and the HOD/CLF formulation is in § 3.

2 APPLICATION TO ABUNDANCE MATCHING

Suppose we want to build a model based on a some property x, which we want to be in some sense intermediate between two halo properties we already know and love. For simplicity and definiteness, let's figure out how to do this using $V_{\rm peak}$ and $M_{\rm peak}$ in the absence of scatter.

In the standard approach, we first compute the cumulative number density of the halos, and calculate our mapping $\bar{M}_*(x)$ in the normal way, by finding the zero of the function

$$\mathcal{F}(M_*|x) \equiv N_{\rm halo}(> x) - N_{\rm gal}(> M_*), \qquad (1)$$

where $N_{\rm halo}(> x)$ is the cumulative abundance of (sub)halos as a function of x, and $N_{\rm gal}(> M_*)$ is the observed cumulative stellar mass function,

$$N_{\rm halo}(> x) \equiv \int_x^\infty {\rm d}x'\, {\rm d}n_{\rm h}/{\rm d}x', \qquad (2)$$

$$N_{\rm gal}(> M_*) \equiv \int_{M_*}^\infty {\rm d}M_*'\, \Phi_{\rm gal}(M_*'). \qquad (3)$$

Ok, so the core quantity that we need to do abundance matching with, say, $V_{\rm peak}$, is just the value of $N_{\rm halo}(> V_{\rm peak})$ for every (sub)halo in our box. Computing this quantity is trivial,

$$N_{\rm halo}(> V_{\rm peak}^{\rm i}) = {\rm i}/L_{\rm box}^3,$$

where i is the $V_{\rm peak}-$based rank-order of the (sub)halos.

That's all the standard jam. But now suppose we want to conduct AM using a property x that is "between" $V_{\rm peak}$ and $M_{\rm peak}$. We begin in the normal way by computing the numerical value of $N_{\rm halo}(> M_{\rm peak})$ and $N_{\rm halo}(> V_{\rm peak})$ for every (sub)halo in the box, exactly as above. Next we define the following *weighted cumulative abundance*:

$$N_{\rm halo}(> x) \equiv w N_{\rm halo}(> V_{\rm peak}) + (1-w)N_{\rm halo}(> M_{\rm peak}), \qquad (4)$$

where w is any number between 0 and 1. This gives us a numerical value of $N_{\rm halo}(> x)$ for every (sub)halo in

the box. Now we just use Eq. 1 to assign M_* to every (sub)halo using AM based on our newly defined property x.

Notice that we never even assigned values of x to the (sub)halos! So this trick entirely circumvents the problem that M_{peak} and V_{peak} do not even have the same units. **Upshot:** we can allow w to be a continuously varying parameter in an MCMC, allowing us to marginalize over our uncertainty about the "true" halo property in the AM-prediction of M_*. This provides a natural extension to the abundance matching parameter scan explored in Reddick et al. (2013).

3   APPLICATION TO THE HOD/CLF

Now let's build new families of HOD models $P(N_{\mathrm{gal}}|x)$ based on a generalized host halo property x that's "between" M_{vir} and V_{max}.

(i) Write down your favorite parameterization for $P(N_{\mathrm{gal}}|M_{\mathrm{vir}})$.

(ii) Compute $N_{\mathrm{halo}}(> M_{\mathrm{vir}})$ for every halo.[1]

The outcome of this step is to provide a bijective mapping between $N_{\mathrm{halo}}(> M_{\mathrm{vir}}) \leftrightarrow M_{\mathrm{vir}}$.

(iii) Use the mapping provided by Step (2) to recast

$$P(N_{\mathrm{gal}}|M_{\mathrm{vir}}) \Rightarrow P(N_{\mathrm{gal}}|N_{\mathrm{halo}}(> M_{\mathrm{vir}})).$$

This is accomplished by function composition:

$$\langle N_{\mathrm{gal}}\rangle(M_{\mathrm{vir}}) \Rightarrow \langle N_{\mathrm{gal}}\rangle(N_{\mathrm{halo}}(> M_{\mathrm{vir}})).$$

This doesn't change our original model in the slightest; we're now just using a new label $N_{\mathrm{halo}}(> M_{\mathrm{vir}})$ in our equations.

(iv) Compute $N_{\mathrm{halo}}(> V_{\mathrm{max}})$ for every halo, giving us $N_{\mathrm{halo}}(> V_{\mathrm{max}}) \leftrightarrow V_{\mathrm{max}}$.

(v) By replacing all appearances of $N_{\mathrm{halo}}(> M_{\mathrm{vir}})$ with $N_{\mathrm{halo}}(> V_{\mathrm{max}})$, we construct a new HOD model $P(N_{\mathrm{gal}}|V_{\mathrm{max}})$ that is the "$V_{\mathrm{max}}-$equivalent" of our original model $P(N_{\mathrm{gal}}|M_{\mathrm{vir}})$.

(vi) More than this, we can also construct entirely new models using

$$N_{\mathrm{halo}}(> x) \equiv w N_{\mathrm{halo}}(> V_{\mathrm{max}}) + (1 - w) N_{\mathrm{halo}}(> M_{\mathrm{vir}}).$$

Those models will be the "$x-$equivalent" HOD model to our original. Again, at no point do we ever need to compute the value of x.

This gives us a recipe to explore entirely new models that are in every way equivalent to our original $P(N_{\mathrm{gal}}|M_{\mathrm{vir}})$, but instead belong to a continuous family of halo properties spanned by V_{max} and M_{vir}.

REFERENCES

Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, ApJ , 771, 30

---

[1] Again, this is computed as i/L_{box}^3, except this time neglecting subhalos from all aspects of the calculation, as is usual for this type of model.