Mon. Not. R. Astron. Soc. **000**, 000–000 (0000) Printed 10 February 2015 (MN LATEX style file v2.2)

Discovering the True Regulator of Stellar Mass

Andrew P. Hearin & Peter S. Behroozi

10 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 is the core assumption that permits us to use N-body simulations in cosmological parameter inference. In this paper we introduce the weighted rank-order technique (WRO), which allows *the choice of halo property itself* to be parameterized and continuously varied in a likelihood analysis. There are significant, systematic differences between empirical models based on different halo properties, and it has recently been shown that the associated systematic error known as *galaxy assembly bias* threatens the precision cosmology program. We describe how to use WRO to generalize abundance matching models as well as halo occupation distribution (HOD) models to make them more robust to assembly bias. We conclude by discussing the role that the WCA technique will play in future constraints on cosmology with large-scale structure.

1 INTRODUCTION

(i) The halo model is the premier tool to constrain dark energy with large-scale structure measurements.

(ii) How does that work? First you enumerate $\mathrm{d}n_{\rm h}/\mathrm{d}M_{\rm vir}$ and $b_{\rm h}(M_{\rm vir})$ through a painful calibration exercise.

Next you write down a parameterized form of an occupation model such as $\Phi(M_*|M_{\rm vir})$ or $\langle N_{\rm gal}\rangle(M_{\rm vir})$.

Finally, you use e.g., clustering and lensing predictions of the model to constrain both cosmological and galaxy-halo parameters.

(iii) However, if you base your model on different halo properties, your model predictions may be qualitatively distinct, because assembly bias is scale dependent. Halo mass $M_{\rm vir}$ is the most widespread, but its definition is based on spherical top-hat collapse, which is poorly motivated in detail. The impact of the choice of halo parameter has been poorly studied heretofore, because there has been no way to easily compare models based on different properties, and because precision demands have only recently reached this point.

(iv) This paper provides a new technique that allows us to construct and compare models that are identical in every way, but differ by the choice of halo property.

The technique also allows us to parameterize our choice of the halo property in a continuous fashion, so that we can marginalize over our uncertainty in the true halo property.

(v) The AM formulation of the generalization is in § 2, and the HOD/CLF formulation is in § 3.

2 ABUNDANCE MATCHING WITH A GENERALIZED HALO PROPERTY

2.1 Traditional Implementation

The basic principle of abundance matching is that there exists a single (sub)halo property which is in monotonic correspondence with the stellar mass of the galaxy residing at the center of the (sub)halo. Once the amount of scatter between this halo property and M_* is specified, the mapping between galaxies and dark matter halos is determined.

The standard implementation of this model begins by first selecting a particular halo property, such as peak historical mass $M_{\rm peak}$. Next, one computes the cumulative number density of halos and subhalos as a function of $M_{\rm peak}$, which we will write as $N_{\rm halo}(>M_{\rm peak})$. This calculation is extraordinarily simple. The (sub)halos are rank-ordered according to the value of $M_{\rm peak}$, and the cumulative number density of the $i^{\rm th}-$rank (sub)halo is given by

$$N_{\rm halo}(>M^{\rm i}_{\rm peak}) = \frac{i}{V_{\rm box}},$$

where $V_{\rm box}$ is the simulated volume.

The key insight behind abundance matching is that there exists a unique deterministic mapping $\bar{M}_*(M_{\rm peak})$ that preserves the observed stellar mass function. Once the value of $N_{\rm halo}(>M_{\rm peak})$ is known for every (sub)halo, this mapping can be numerically calculated by finding the zero of the function

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

where $N_{\rm gal}(>M_*)$ is the observed cumulative stellar mass function. For any particular choice of (pos-

sibly $M_{\rm peak}-$dependent) stochasticity, there also exists a unique map $\bar{M}_*(M_{\rm peak})$ that includes this scatter and preserves the observed $N_{\rm gal}(>M_*)$. Determining this map and introducing stochasticity between M_* and $M_{\rm peak}$ can be accomplished by the deconvolution method described in Section 3.3.1 of Behroozi et al. (2010) or in Appendix A of Kravtsov (2013).

Conventionally, the upshot of the above calculation is the determination of the mapping $\bar{M}_*(M_{\rm peak})$. This well-studied map is most commonly employed in the literature with convenient fitting functions such as those given in Behroozi et al. (2010); Moster et al. (2010); Kravtsov (2013).

2.2 Weighted Rank-Ordering: application to abundance matching

Of course, there is no a priori reason that the choice of $M_{\rm peak}$ is innocuous. In fact, it has been shown in Reddick et al. (2013) that different choices for the abundance matching parameter lead to models with different clustering predictions. This directly implies that existing SDSS data can discriminate between different choices for the primary halo property regulating M_*. We now describe how to adapt the methods outlined above in § 2.1 to build models based on a halo property that is "intermediate" between two particular, discrete choices.

For illustrative purposes, let's begin with a specific example. Suppose we want to conduct abundance matching 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:

$$N_{\rm halo}(>M^{\rm i}_{\rm peak}) = \frac{i_{M_{\rm peak}}}{V_{\rm box}}$$
$$N_{\rm halo}(>V^{\rm j}_{\rm peak}) = \frac{j_{V_{\rm peak}}}{V_{\rm box}},$$

Next we define the weighted logarithmic sum of the $M_{\rm peak}-$ and $V_{\rm peak}-$based ranks:

$$\mathcal{R}_x(w|M_{\rm peak},V_{\rm peak}) \equiv w\log i_{M_{\rm peak}} + (1-w)\log j_{V_{\rm peak}}$$

where w is any number between 0 and 1. This gives us a new quantity $\mathcal{R}_x$ for every (sub)halo in the simulation that we can rank-order. The resulting ranks $k_{\rm x}$ allow us to compute a new, $x-$based cumulative number density:

$$N_{\rm halo}(>x^{\rm k}) = \frac{k_{\rm x}}{V_{\rm box}}. \quad (2)$$

Finally 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 $M_{\rm peak}$ and $V_{\rm peak}$ 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).

Brief comment on how to implement scatter without ever knowing the value of x.

3 THE HOD WITH A GENERALIZED HALO PROPERTY

3.1 Traditional Implementation

3.2 Weighted Rank-Ordering: application to the HOD

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

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

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

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

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

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

This is accomplished by function composition:

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

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

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

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

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

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

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 $P(N_{\rm gal}|M_{\rm vir})$, but instead belong to a continuous family of halo properties spanned by $V_{\rm max}$ and $M_{\rm vir}$.

4 CONCLUSION

We provide a new technical tool to use in empirical modeling.

REFERENCES

Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ , 717, 379

Kravtsov A. V., 2013, ApJL , 764, L31

Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, ApJ , 710, 903

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

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