Statistical Challenges in 21st Century Cosmology
Proceedings IAU Symposium No. 306, 2014
A. F. Heavens, J.-L. Starck & A. Krone-Martins, eds.
doi:10.1017/S1743921314010710

Modelling Galaxy Populations in the Era of Big Data

S. G. Murray[1,2]†, C. Power[1,2] and A. S. G. Robotham[1]

[1]ICRAR, University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009, Australia
[2]ARC Centre of Excellence for All-Sky Astrophysics (CAASTRO)

Abstract. The coming decade will witness a deluge of data from next generation galaxy surveys such as the Square Kilometre Array and Euclid. How can we optimally and robustly analyse these data to maximise scientific returns from these surveys? Here we discuss recent work in developing both the conceptual and software frameworks for carrying out such analyses and their application to the dark matter halo mass function. We summarise what we have learned about the HMF from the last 10 years of precision CMB data using the open-source `HMFcalc` framework, before discussing how this framework is being extended to the full Halo Model.

1. Introduction

The Halo Model has proven successful in predicting the large-scale clustering of dark matter halos and galaxies, and associated properties (Zheng(2004), Zehavi *et al.*(2011), Beutler *et al.*(2013)). With the emerging era of "big data" the Halo Model is set to be one of the key interpretive frameworks in the context of galaxy evolution and spatial statistics. It combines several components, and our aim in this short work is to further understanding of these components in a statistical sense. We also present a seamlessly integrated implementation of the various components, as a platform for next-generation modelling.

2. Uncertainty in the Halo Mass Function

The Halo Mass Function (HMF) is a standard tool in cosmology, used for consistency checking in simulations, fitting of cosmological parameters (especially σ_8) and as a component of other models. However, to properly apply the HMF requires a knowledge of its uncertainty distribution. We thus ask the question: "given our current 'best-bet' cosmology, with associated covariant uncertainty, what is the resulting uncertainty in the predicted mass function?" As a first-order application, this gives a rule-of-thumb estimate to assess whether a given measurement is consistent with current theory, but we show that such an analysis reveals more subtle insights as well (Murray *et al.*(2013a)).

To answer the question, we used the cosmological parameters, with covariant errors, from the past decade of WMAP and Planck results randomly drawing 5000 realisations of the four key parameters $\Omega_b h^2$, $\Omega_c h^2$, σ_8 and n_s from the MCMC chains of four releases. From these parameters we calculated 5000 HMF realisations for each of 12 fitting functions, calculating the resulting 1-σ uncertainties in both the amplitude and slope (shown in Fig. 1).

We find that over the course of the past decade of precision CMB measurements, the uncertainty in the HMF has accordingly decreased, and we give rule-of-thumb figures of 6% (20%) uncertainty at low (high) mass with current Planck results.

† steven.murray@uwa.edu.au

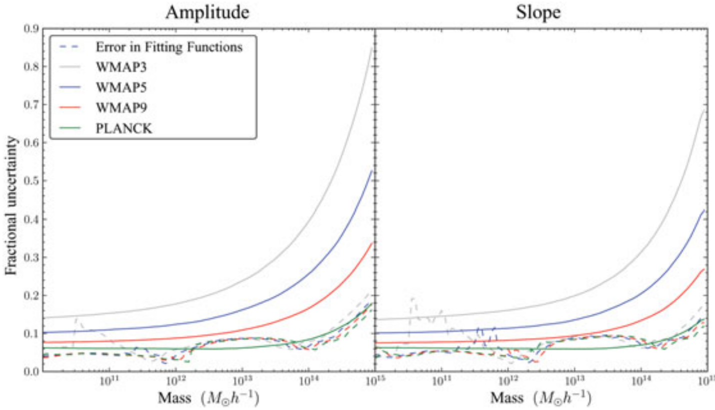

Figure 1. Fractional uncertainty in the amplitude (left) and slope (right) of the HMF for a range of parameter sets. Dashed lines indicate uncertainty due to 'arbitrary' choice of fitting function.

Consistent with previous studies (Vikhlinin *et al.*(2009)), we find that the HMF is most sensitive to the parameters Ω_m and σ_8 (at low-mass and high-mass respectively), which confirms the HMF as a useful tool for constraining these parameters.

Twelve fitting functions from the literature were used at each stage of the analysis which enabled some comparative statistics. Firstly, while the amplitude of each fit is different, the uncertainty behaves in the same way across all fits, allowing the use of one fit representatively. Secondly, we find that the scatter between fits, assuming arbitrary choice of fit, is now comparable to the uncertainty due to cosmology (see Fig. 1), highlighting the need for more robust fitting (cf. Bhattacharya *et al.*(2011)), and an educated choice of fit.

Finally, we calculated the uncertainty at which a viable Warm Dark Matter (WDM) model was more than 1-σ from its CDM counterpart, indicating the level of cosmological uncertainty necessary to detect WDM (see Fig. 2). Linearly extrapolating the trend in uncertainty reduction over the previous decade, we predict another 20 years before we can differentiate CDM and WDM on group mass scales ($\sim 10^{13} M_\odot h^{-1}$) using the HMF.

3. Halo Model Implementation

Though the Halo Model is set to be a key framework in the interpretation of the next decade of galaxy survey data, there does not exist an open-source implementation for its calculation. Our new code `halomod`† fills this gap, in the philosophy of our `hmf` code (Murray *et al.*(2013b)).

The `halomod` code was developed with primary goals of flexibility and ease-of-use. To this end, the user may submit a custom fitting function, or one of several other models, to the engine without modifying source code at all – a beneficial feature in a field where new models arise frequently. It also natively supports WDM models, has online documentation, includes most of the models already available in the literature, is simple to install and is algorithmically optimised for iteratively modifying model parameters.

Accompanying our `hmf` code is a web-application interface, `HMFcalc`‡, which exposes all appropriate functionality of `hmf` with a simple web-interface. We are preparing an extension of `HMFcalc` which extends the capabilities to the domain of `halomod`.

† Available at `github.com/steven-murray/halomod`
‡ Found at `hmf.icrar.org`

https://doi.org/10.1017/S1743921314010710 Published online by Cambridge University Press

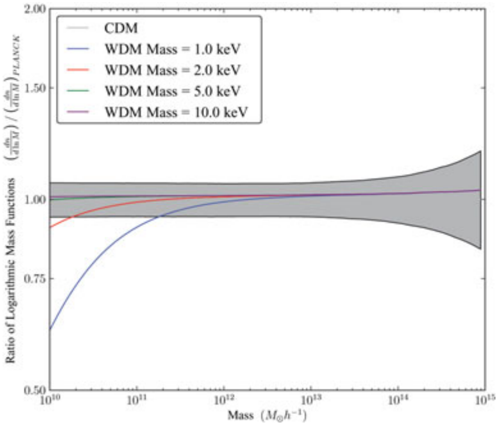

Figure 2. WDM HMF's for specified particle masses against the shaded uncertainty region in Planck CDM. "Boundary Crossing" represents the mass at which a WDM model is detectable for current cosmological uncertainty.

4. Halo Model Fitting

A common application of the Halo Model framework is to estimate the parameters of a Halo Occupation Distribution (HOD) model given an observed projected correlation function (Zehavi *et al.*(2011), Beutler *et al.*(2013)). The HOD parameters and their associated derived parameters are useful tracers of galaxy evolution properties.

Commonly this procedure is performed with a fiducial cosmology while the HOD parameters are fit via a Monte-Carlo routine. We investigate whether fixing the cosmology introduces a bias in the HOD measurement or a misestimation of its uncertainty. This requires leaving the cosmology to vary in the procedure in tandem with the HOD parameters. As a corollary, we determine whether this method can be used to fit certain cosmological parameters.

While this work is preliminary, early results show that several HOD-Cosmology parameter combinations exhibit strong correlations in their posterior distributions, suggesting that they are likely to be biased if the fixed cosmology is biased, and that their uncertainties will have been underestimated. Furthermore, while none of the cosmological parameters are constrained tightly, all marginalised posteriors are bound, suggesting that if appropriate priors are chosen this method could provide an independent and complimentary means of constraining cosmology.

5. Conclusion

We have discussed the application of our new codes `hmf` and `halomod` to the emerging problem of efficient modelling in the era of big data. Specifically, our implementations are simple to use and algorithmically efficient – enabling the seamless modelling of spatial galaxy statistics with current models. We have applied this software to diagnostic statistics for the HMF and two-point galaxy statistics to illustrate their benefit.

References

Bhattacharya, S., Heitmann, K., White, M., *et al.* 2011, *ApJ*, 732, 122
Zehavi, I., Zheng, Z., Weinberg, D. H., *et al.* 2011, *ApJ*, 736, 59
Zheng, Z. 2004, *ApJ*, 610, 61
Beutler, F., Blake, C., Colless, M., *et al.* 2013, *MNRAS*, 429, 3604
Murray, S. G., Power, C., & Robotham, A. S. G. 2013, *MNRAS*, 434, L61
Murray, S. G., Power, C., & Robotham, A. S. G. 2013, *Astronomy and Computing*, 3, 23
Vikhlinin A., *et al.* 2009, *ApJ*, 692, 1060

https://doi.org/10.1017/S1743921314010710 Published online by Cambridge University Press