



## The Aemulus Project. II. Emulating the Halo Mass Function

Thomas McClintock<sup>1</sup>, Eduardo Rozo<sup>1</sup>, Matthew R. Becker<sup>2,3,4</sup>, Joseph DeRose<sup>2,3</sup>, Yao-Yuan Mao<sup>5</sup>, Sean McLaughlin<sup>2,3</sup>, Jeremy L. Tinker<sup>6</sup>, Risa H. Wechsler<sup>2,3</sup>, and Zhongxu Zhai<sup>6</sup>

<sup>1</sup>Department of Physics, University of Arizona, Tucson, AZ 85721, USA

<sup>2</sup>Kavli Institute for Particle Astrophysics and Cosmology and Department of Physics, Stanford University, Stanford, CA 94305, USA

<sup>3</sup>Department of Particle Physics and Astrophysics, SLAC National Accelerator Laboratory, Stanford, CA 94305, USA

<sup>4</sup>Civis Analytics, Chicago, IL 60607, USA

<sup>5</sup>Department of Physics and Astronomy and the Pittsburgh Particle Physics, Astrophysics and Cosmology Center (PITT PACC), University of Pittsburgh, Pittsburgh, PA 15260, USA

<sup>6</sup>Center for Cosmology and Particle Physics, Department of Physics, New York University, 4 Washington Place, New York, NY 10003, USA

Received 2018 April 16; revised 2018 November 19; accepted 2018 November 29; published 2019 February 8

### Abstract

Existing models for the dependence of the halo mass function on cosmological parameters will become a limiting source of systematic uncertainty for cluster cosmology in the near future. We present a halo mass function emulator and demonstrate improved accuracy relative to state-of-the-art analytic models. In this work, mass is defined using an overdensity criteria of 200 relative to the mean background density. Our emulator is constructed from the AEMULUS simulations, a suite of 40  $N$ -body simulations with snapshots from  $z = 3$  to  $z = 0$ . These simulations cover the flat  $\Lambda$ CDM parameter space allowed by recent cosmic microwave background, baryon acoustic oscillation and SNe Ia results, varying the parameters  $w$ ,  $\Omega_m$ ,  $\Omega_b$ ,  $\sigma_8$ ,  $N_{\text{eff}}$ ,  $n_s$ , and  $H_0$ . We validate our emulator using five realizations of seven different cosmologies, for a total of 35 test simulations. These test simulations were not used in constructing the emulator, and were run with fully independent initial conditions. We use our test simulations to characterize the modeling uncertainty of the emulator, and introduce a novel way of marginalizing over the associated systematic uncertainty. We confirm nonuniversality in our halo mass function emulator as a function of both cosmological parameters and redshift. Our emulator achieves better than 1% precision over much of the relevant parameter space, and we demonstrate that the systematic uncertainty in our emulator will remain a negligible source of error for cluster abundance studies through at least the LSST Year 1 data set.

*Key words:* large-scale structure of universe – methods: numerical – methods: statistical

### 1. Introduction

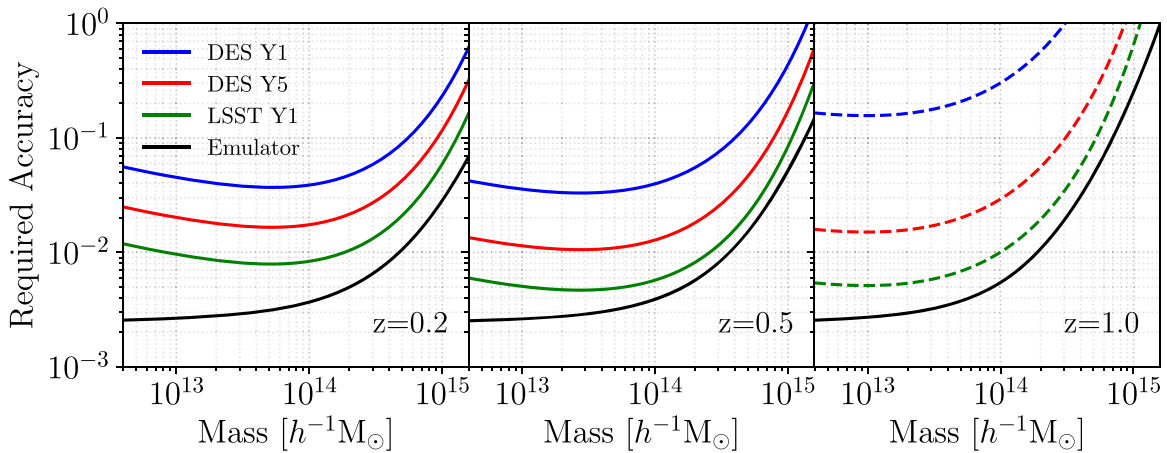
The dependence of the halo mass function—i.e., the comoving density of dark matter halos per unit mass—on cosmological parameters enables the abundance of galaxy clusters to be an exceptionally promising dark energy probe (Albrecht et al. 2006). Indeed, the recent Dodelson et al. (2016) report shows that galaxy clusters could provide the tightest constraints yet on cosmological parameters, provided the associated systematic uncertainties can be adequately controlled. Current optical surveys such as the Dark Energy Survey (DES Dark Energy Survey Collaboration 2005, 2016; Melchior et al. 2017), the Hyper Suprime-Cam Survey (Miyazaki et al. 2012; Aihara et al. 2018), and the Kilo-Degree Survey (de Jong et al. 2017; Radovich et al. 2017) can expect to observe  $\sim 100,000$  galaxy clusters, with this number growing by up to an order of magnitude in the LSST era (Weinberg et al. 2013). SZ and X-ray cluster catalogs (e.g., Piffaretti et al. 2011; Bleem et al. 2015) will reach similarly large samples of clusters. Here, we take an important step toward calibrating the cosmology dependence of the halo mass function at a precision that is sufficiently high to ensure this source of systematic uncertainty remains negligible in the LSST era.

The largest source of uncertainty in cluster abundance studies is generally cluster mass calibration. Observationally, cluster masses are estimated using a mass–observable relation. The observable can be the thermal SZ effect (Bleem et al. 2015; Dietrich et al. 2019; Stern et al. 2019), X-ray emission from hot gas (Piffaretti et al. 2011; Mehrrens et al. 2012), or optical richness (Rykoff et al. 2016; Melchior et al. 2017; Simet et al. 2017). The recent analysis

of McClintock et al. (2019) in the DES Year 1 data set achieved a 5% calibration of the mass–richness relation of the redMaPPer clusters. This is expected to improve to  $\sim 1\%$  for LSST and extend out to redshifts  $z \approx 1$ .

In addition to uncertainties in the mass–observable relation, the total error budget in a cluster abundance analysis must include a contribution from modeling uncertainty. One such example is the systematic uncertainty associated with modeling the halo mass function. Reed et al. (2007) present a review of halo mass function models published up to that point. A discussion on accuracy requirements for mass function models can be found in Reed et al. (2013). Assuming clusters map perfectly onto halos, the systematic uncertainty in the calibration of the halo mass function directly translates into a systematic uncertainty in the predicted cluster abundance. The Tinker et al. (2008) mass function reports an accuracy of  $\approx 5\%$  for  $10^{11} < M < 10^{15} h^{-1} M_\odot$  halos at  $z = 0$ . The calibration was obtained using a suite of flat  $\Lambda$ CDM simulations that varied the parameters  $\Omega_m$ ,  $\Omega_b$ ,  $\sigma_8$ ,  $n_s$ , and  $H_0$ .

Figure 1 presents the required accuracy for the systematic uncertainty in the halo mass function to be negligible (quantitatively defined below) in the DES Year 1, DES Year 5, and LSST Year 1 cluster abundance analyses. These requirements are set by rescaling the statistical error reported in McClintock et al. (2019) for the uncertainty in halo mass calibration. Specifically, we demand that the theoretical uncertainty in the halo mass function must not increase the total error budget of a cluster abundance analysis by more than 10% relative to the error from mass calibration alone. The details of this calculation, including assumptions regarding the



**Figure 1.** Required accuracy of the halo mass function for different surveys. These accuracies are calculated by demanding that the uncertainty in the halo mass function not increase the error budget of a cluster abundance measurement by more than 10% relative to the uncertainty from cluster mass calibration. Mass calibration uncertainties are estimated based on the DES Y1 measurements of DES Y1 cluster (blue; McClintock et al. 2019). DES Y5 (red) and LSST Y1 (green) requirements are projected based on expected improvements of the weak lensing masses. Dashed lines are extrapolations to  $z = 1$ . The black curve indicates the current accuracy of the emulator. For details of how the accuracy requirements are computed, see Appendix B.

lens source densities, survey areas, and depths, are provided in Appendix B. Importantly, Figure 1 shows that the claimed 5% accuracy at  $z = 0$  of the Tinker et al. (2008) mass function is nearly sufficient for the DES Y1 data set, but not for future data sets. Evidently, a new calibration of the halo mass functions is required to prevent systematic uncertainties in the halo mass function from contributing significantly to the cosmological error budget in cluster abundance studies.

Early halo mass function models were analytic estimates (Press & Schechter 1974). More recent analyses have calibrated fitting functions from simulations (Sheth & Tormen 1999; Sheth et al. 2001). Jenkins et al. (2001) and Evrard et al. (2002) attempted to find a universal fitting function accurate to  $\sim 10\%$ – $20\%$ , where the mass function did not explicitly depend on cosmological parameters. Later, Warren et al. (2006) calibrated a fitting function on a single cosmology at  $z = 0$  and achieved  $\sim 5\%$  accuracy. Tinker et al. (2008) calibrated the halo mass function to  $\approx 5\%$  at  $z = 0$  for virial masses identified using the spherical overdensity (SO) algorithm in the range  $10^{11} \leq M \leq 10^{15} h^{-1} M_{\odot}$ . Tinker et al. (2008) has been the standard for cluster abundance analyses since then (Zu et al. 2014; Mantz et al. 2016; Planck Collaboration et al. 2016). Similar high-precision calibrations of the halo mass function for friends-of-friends halos also exist (Bhattacharya et al. 2011).

More recently, Heitmann et al. (2016) built a high-dimensional interpolator using Gaussian Processes (Rasmussen & Williams 2005) to model the halo mass function for halos identified using a *friends-of-friends* (FOF) algorithm. In an eight-dimensional cosmological parameter space they were able to accurately model the mass function for their entire mass range at  $z = 0$ . The choice of halo finding algorithm impacts the recovered halo mass function. In practice, the appropriateness of different halo definitions depends on the specific science question under consideration. In this work, we have chosen to focus exclusively on SO halos. Note that the connection between FOF and SO halos is nontrivial, and exhibits strong asymmetric scatter (Tinker et al. 2008; More et al. 2011). Thus we caution that our results are only appropriate for the specific mass definition we have adopted.

The focus of this paper is to present a halo mass function emulator calibrated specifically for SO halos in numerical simulations. For the purposes of this work, halo mass is defined using an overdensity threshold  $\Delta = 200b$ , which is 200 times the mean matter density of the universe at the epoch at which the halo is identified. Our approach involves emulating parameters in a fitting function rather than  $n(M, z)$  directly. This approach has an advantage over directly emulating the mass function in that it requires significantly less training data. Moreover, because our fitting functions are expressed in terms of the peak height  $\nu$ , much of the cosmological dependence of the mass function is already removed. This way, the emulator need only characterize any remaining cosmological sensitivity.

The original mass function by Press & Schechter (1974) was universal when written in terms of the peak height  $\nu$ . That is, the authors proposed a form of the mass function that did not explicitly depend on cosmological parameters nor did it evolve over time. Recently, Despali et al. (2016) found that the mass function using the  $M_{\text{vir}}$  definition is universal at the few percent level, and that other definitions such as  $M_{200b}$  are less universal. For constant overdensities  $\Delta$  Tinker et al. (2008) demonstrated nonuniversality at the 20%–50% level between different epochs in their simulations. Our emulator incorporates nonuniversality by design, and we demonstrate the dependence of the emulated halo multiplicity function on redshift,  $\Omega_m$ , and  $\sigma_8$ . We note that different definitions of halo mass may lead to increased or decreased universality.

We quantitatively characterize the performance of our emulator a posteriori and provide a model for the residuals of the halo mass function emulator. This residual model accurately describes the nonconstant uncertainty over the masses and epochs in our simulations, allowing for the uncertainty in the mass function emulator to be properly propagated into abundance analyses. We find that our emulator achieves subpercent accuracy for certain mass ranges, and is accurate enough for the LSST Y1 analysis out to at least  $z = 1$ .

The layout of this paper is as follows. In Section 2, we briefly discuss the simulations used in this work, detail the process of identifying halos, and define the fitting function used. Section 3 describes our method for training the Gaussian

processes and constructing the halo mass function emulator. In Section 4 we show tests of the accuracy of the emulator, while in Section 5 we present a model for the residuals of the comparison between the emulator predictions and the measured halo mass functions. Section 6 contains a discussion of nonuniversality in the emulator for the halo mass function. Section 7 shows the emulator predictions for additional simulations that probe both lower and higher masses than those used to construct our emulator. Section 8 summarizes our results. In Appendix A we present tests of the residual model, while in Appendix B we derive the accuracy requirements for current and near-future cluster cosmology analyses. We make our simulations available at <https://AemulusProject.github.io>.

## 2. Simulations

Massive clusters live in the high-mass exponentially decaying tail of the halo mass function. Properly investigating this regime requires  $N$ -body simulations with enough resolution and volume to ensure sufficient statistics for high-mass halos. Collectively, all our simulations comprise the AEMULUS suite (DeRose et al. 2018; Zhai et al. 2018). The “training simulations” used to construct the emulator consist of 40  $N$ -body simulations. Each simulation has a length  $L = 1050 h^{-1}$  Mpc with periodic boundary conditions and  $1400^3$  particles. The cosmologies of these simulations span the allowed  $4\sigma$  parameter space spanned by a union of *Planck* and WMAP9 (Hinshaw et al. 2013; Planck Collaboration et al. 2014) plus BAO from BOSS (Anderson et al. 2014) and the Union 2.1 SNIa data (Suzuki et al. 2012). Initial conditions vary between each simulation, and are specified using CAMB. As described in DeRose et al. (2018), the cosmological sampling of the simulations was designed using an orthogonal Latin hypercube design (Heitmann et al. 2009) and they have an effective volume of  $42 (h^{-1} \text{Gpc})^3$ . Particle masses are given by  $M_{\text{part}} = 3.513 \times 10^{10} \left(\frac{\Omega_m}{0.3}\right) h^{-1} M_{\odot}$ .

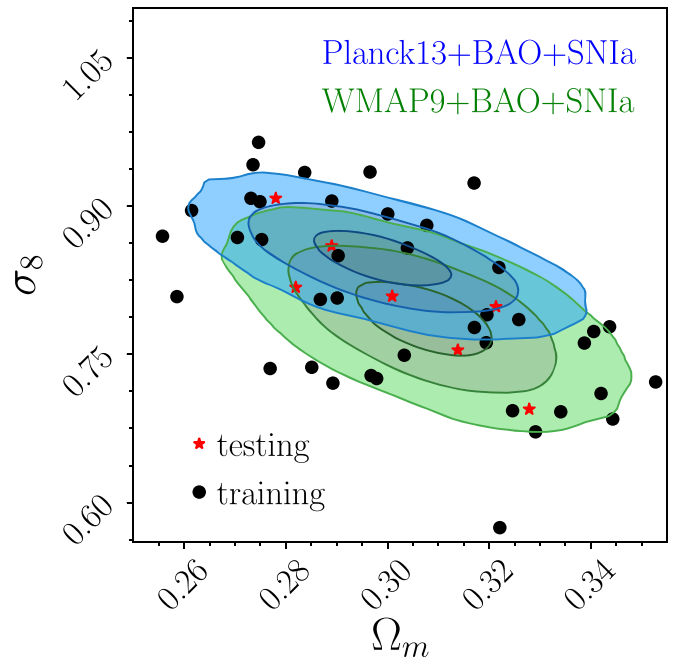
In order to assess the performance of the emulator we ran another set of 35 simulations, dubbed the “test simulations.” These simulations are comprised of seven different cosmologies with five different sets of initial conditions per cosmology. The five realizations are combined in order to reduce sample variance when validating the emulator performance. None of the test simulations were used in the construction of the emulator, and they span the range of cosmological parameter space used to define the training simulations.

### 2.1. Cosmological Models

Cluster abundance is most sensitive to the matter power spectrum normalization as specified by  $\sigma_8$  and to the matter density  $\Omega_M$ . The training simulations exist in the parameter space  $p \in [\Omega_b h^2, \Omega_c h^2, w, n_s, H_0, N_{\text{eff}}, \sigma_8]$ , where  $\Omega_b$  is the baryonic matter fraction,  $\Omega_c$  is the cold dark matter fraction,  $n_s$  is the power spectrum index,  $h = H_0(100 \text{ km s}^{-1} \text{ Mpc}^{-1})^{-1}$  is the Hubble constant, and  $N_{\text{eff}}$  is the effective number of relativistic species. Both the training and test simulations are shown as points in Figure 2 overlaid on top of the likelihood contours they are designed to span.

### 2.2. Halo Identification

Halos were identified using the ROCKSTAR halo finder (Behroozi et al. 2013), which identifies halos across simulation snapshots. We use the  $M_{200b}$  mass definition, where the halo is



**Figure 2.** CMB+BAO+SNIa allowed parameter space (contours) for  $\sigma_8$  and  $\Omega_M$ . This is a combination of BAO from BOSS DR11, the Union 2.1 SNIa catalog, and *Planck*/WMAP9. Contour levels are the  $1\sigma$ ,  $2\sigma$ , and  $3\sigma$  confidence contours. Points are the locations of the training simulations used to construct the emulator. The red stars mark the locations of the test simulations.

defined as a SO  $\Delta = 200$  times more dense than the background. We use the strict SO definition, where strict refers to the inclusion of unbound particles in the halo mass estimates. We conservatively only consider halos with 200 or more particles. The mass and abundance of the halos with the smallest particle numbers (below 500–1000 particles) were found to depend on the mass resolution of the simulations. To account for this systematic, we applied a correction to the recovered abundances as described in Section 4.2.5 of DeRose et al. (2018).

Halos in each snapshot were split into 20 mass bins beginning at the minimum halo mass resolved in each simulation. The maximum edge was fixed arbitrarily at  $10^{17} h^{-1} M_{\odot}$ ; however, the maximum halo mass in the simulations was  $7.06 \times 10^{15} h^{-1} M_{\odot}$ . This resulted in the number of bins with halos varying across snapshots, with  $\sim 5$  at  $z = 3$  and  $\sim 15$  at  $z = 0$ . Subhalos were ignored, where we used the default ROCKSTAR subhalo definition in which subhalos are identified when a halo is within the radius of a larger halo, defined by  $R_{200b}$ . Using  $8^3 = 512$  spatial jackknife subregions, we estimated the covariance matrix between bins in a given snapshot. We ignore correlations between mass bins across different snapshots when performing the fits described in Section 2.3.

### 2.3. Mass Function

Our emulators were not trained on the measured mass functions directly. Instead, we fit the mass function of each simulation snapshot with a modified version of the mass function presented in Appendix C of Tinker et al. (2008). Similar fitting functions were presented in Jenkins et al. (2001) and Warren et al. 2006. In Tinker et al. (2008), the cosmological parameters altered the mass function in two ways: (1) they alter the contribution of matter to the critical



density  $\Omega_m \rho_c$  and (2) they change the mapping from mass to the rms variance of the linear density field  $\sigma(M, z)$ . We extend this approach by allowing the fitting function parameters to have cosmological dependence as well, which is captured using the Gaussian processes that underpin the emulator.

This fitting function has the following form:

$$\frac{dn}{dM} = G(\sigma) \frac{\bar{\rho}_m}{M} \frac{d \ln \sigma^{-1}}{dM}, \quad (1)$$

where the halo multiplicity function  $G(\sigma)$  is given by

$$G(\sigma) = B \left[ \left( \frac{\sigma}{e} \right)^{-d} + \sigma^{-f} \right] \exp(-g/\sigma^2). \quad (2)$$

Here,  $\sigma^2$  is the rms variance of the linear density field

$$\sigma^2 = \int \frac{dk}{k} \frac{k^3 P(k, z)}{2\pi^2} |\hat{W}(kR)|^2 \quad (3)$$

evaluated at the Lagrangian scale of the halo, i.e.,  $R = (3M/4\pi\bar{\rho}_m)^{1/3}$ .  $P(k, z)$  is the linear matter power spectrum as a function of wavenumber  $k$  and redshift  $z$ , and  $\hat{W}$  is the Fourier transform of the real-space top-hat window function. Additionally, we require all dark matter to reside in halos, which implies

$$\int d \ln \sigma^{-1} G(\sigma) = 1. \quad (4)$$

The peak height of a halo is defined as  $\nu = \delta_c/\sigma(M)$ , where  $\delta_c = 1.686$  is the critical density for collapse and  $\sigma(M)$  is given by Equation (3).

The simulation is unable to sample arbitrarily large modes of the power spectrum due to the finite size of the box. We confirmed that our results are insensitive to a cut in  $k = 2\pi/R$  at the scale of our simulation  $R = 1.05 h^{-1} \text{ Gpc}$ . We use the publicly available CLASS<sup>7</sup> to calculate the power spectrum. We note that attempting to use analytic transfer functions resulted in a clear decrease in performance by the emulator. By solving Equation (4) we can write  $B$  in terms of the remaining free parameters in Equation (2):

$$B = 2 \left[ e^d g^{-d/2} \Gamma\left(\frac{d}{2}\right) + g^{-f/2} \Gamma\left(\frac{f}{2}\right) \right]^{-1}. \quad (5)$$

The fitting function parameters are  $d$ ,  $e$ ,  $f$ , and  $g$ . The parameter  $e$  sets the amplitude and  $d$  and  $f$  set the slope of low-mass power law;  $g$  determines the cutoff mass where the abundance of halos decreases exponentially. Empirically we find that for every cosmology tested, the redshift evolution of the best-fit parameters is well described as a linear function of the scale factor  $a$ . This was confirmed by fitting for the parameters individually in each snapshot and observing a linear trend with scale factor. Therefore, in our analysis we simultaneously fit all snapshots of each simulation simultaneously enforcing linear evolution of the fitting parameters. Specifically, we set:

$$p(\Omega, a) = p_0(\Omega) + (a - 0.5)p_1(\Omega). \quad (6)$$

In this equation,  $p_0$  is the value of the fitting function parameter at a scale factor of  $a = 0.5$  and  $p_1$  is the slope of the parameter as a function of the scale factor. We found that fixing  $d_0 = 2.4$

and  $f_1 = 0.12$  optimized the performance of our emulator while removing degeneracies among our full parameter set. We still allow  $d_1$  and  $f_0$  to vary as a function of cosmological parameters. Note that by construction the Gaussian processes only interpolated over cosmological parameter space and not over redshift and mass.

#### 2.4. Fitting the Training Simulations

After measuring the mass function in each simulation, we found the best-fit model using the fitting function in Equation (2). The best-fit parameters are later interpolated across cosmological parameter space using Gaussian processes in order to create the emulator. Our best-fit models are obtained by maximizing the likelihood

$$\ln \mathcal{L} = -\frac{1}{2} \sum_{i=0}^9 (N(z_i) - N_{\text{model}}(z_i))^T C_{N,i}^{-1} (N(z_i) - N_{\text{model}}(z_i)), \quad (7)$$

where  $i$  indexes the 10 snapshots at each of the  $z_i$  redshifts,  $C_{N,i}$  is the jackknife-estimated covariance matrix of the abundances at snapshot  $i$ , and  $N(z_i)$  is the mass function data at each mass bin in that snapshot. When inverting  $C_{N,i}$  we applied the correction from Hartlap et al. (2007). The measured quantity is the number of halos in a given ( $j$ th) mass bin  $N_j$ . The mass bins are not small, so they are modeled by

$$N_{\text{model},j} = V \int_{M_{\text{min},j}}^{M_{\text{max},j}} dM \frac{dn}{dM}, \quad (8)$$

where  $V = 1.05^3 (h^{-1} \text{ Gpc})^3$  is the volume of the simulation, and  $M_{\text{min},j}$  ( $M_{\text{max},j}$ ) is the minimum (maximum) mass in the  $j$ th bin. Choosing the number of bins meant striking a balance between obtaining as much resolution as possible while keeping the covariance matrix small enough to estimate by using a spatial jackknife. We used the MCMC code *emcee* (Foreman-Mackey et al. 2013) to obtain full posterior probability distributions for each of the parameters  $d_1$ ,  $e_0$ ,  $e_1$ ,  $f_0$ ,  $g_0$ , and  $g_1$  for each simulation, imposing flat priors on each of the parameters.

### 3. Emulator Design

The emulator uses Gaussian processes to perform regression on the parameters  $d_1$ ,  $e_0$ ,  $e_1$ ,  $f_0$ ,  $g_0$ , and  $g_1$  as a function of the cosmological parameters discussed in Section 2.1. In total the emulator is comprised of six independent Gaussian processes. Since the Gaussian processes are independent, we had to rotate the parameters into an orthogonal basis. After doing so each new parameter was labeled  $d'_1$ ,  $e'_0$ ,  $e'_1$ ,  $f'_0$ ,  $g'_0$ , and  $g'_1$ . To achieve this we took the MCMC chain from the central-most simulation, index 34, and computed the covariance matrix of the parameters in this chain. Then, by computing its eigenvectors we constructed a rotation matrix that diagonalizes this covariance matrix. We applied the rotation matrix to the MCMC chains for each simulation, so that the parameters of the new chains were all approximately orthogonal. We tested that the rotation matrix used depended only weakly on cosmology, and that the choice of which simulation was used to construct the rotation matrix did not influence the performance of the emulator.

<sup>7</sup> <http://class-code.net/>

Each Gaussian process uses the mean and variance of the posterior probability distributions of the orthogonal parameters to interpolate across cosmological parameter space. When the Gaussian process was used to predict parameters at a new cosmology, the predicted values were rotated back to the original basis.

### 3.1. Gaussian Process Kernel

Consider a set of  $n$  measured data  $y_i$   $i \in [1, \dots, n]$  with variance  $\sigma_i^2$ , located at  $\mathbf{x}_i$ . If the covariance between two measurements  $y_i$  and  $y_j$  depends only on  $\mathbf{x}_i$  and  $\mathbf{x}_j$ , then the measurements can be approximated by a Gaussian process  $y = \mathcal{GP}(\mu(\mathbf{x}), \mathbf{C})$ , where  $\mu$  is the ‘‘mean function’’ and  $\mathbf{C}$  is the covariance matrix. As described in Rasmussen & Williams (2005), modeling some data as a Gaussian process amounts to correctly modeling the covariance, since the mean of the observations  $y$  can be subtracted off such that  $\mu = 0$ . Once a set of data is successfully modeled in this fashion, the Gaussian process can be used for interpolation. Thus, our approach is to find an optimal covariance matrix describing the mass function parameters as a function of cosmology. This allows us to interpolate from the locations in parameter space corresponding to our simulations to arbitrary cosmologies. We use the Gaussian process implementation in the PYTHON package `george`<sup>8</sup> (Ambikasaran et al. 2015).

We model the covariance matrix as

$$\mathbf{C} = \mathbf{K} + \mathbf{S}, \quad (9)$$

where  $\mathbf{K}$  is the kernel matrix and  $\mathbf{S}$  is a diagonal matrix containing the variance of the training data  $\sigma_{x_i}^2$ . Each element of  $\mathbf{K}$  contains the covariance between two data points given by a kernel function. The Gaussian processes were constructed using a squared-exponential kernel:

$$k(\mathbf{x}, \mathbf{x}') = k_0 \exp\left(-\sum_i^{N_p} \frac{(x_i - x'_i)^2}{2L_i^2}\right), \quad (10)$$

where  $L_i$  is the length scale of the  $i$ th cosmological parameter,  $k_0$  is the kernel amplitude, and  $N_p$  is the number of cosmological parameters. We found fixing  $k_0 = 1$  to be an optimal configuration of the emulators in that residuals of the emulator prediction were minimized. In this approach the  $L_i$  are the hyperparameters of the kernel that we seek to optimize in order to model the covariance of the data. Their values are found by maximizing the likelihood

$$\ln \mathcal{L} = -\frac{1}{2} \Delta p^T \mathbf{C}^{-1} \Delta p - \frac{1}{2} \ln \det \mathbf{C}, \quad (11)$$

where  $\Delta p = p - p_{\text{emu}}$ . That is,  $\Delta p$  is the difference between the measured mass function parameters  $p \in \{d'_1, e'_0, e'_1, f'_0, g'_0, g'_1\}$  and the predictions from the emulator. Attempting to include  $k_0$  as a hyperparameter to optimize along with the  $L_i$  allowed the Gaussian processes too much flexibility despite maximizing the likelihood in Equation (11). This resulted in poor performance in the tests discussed in Section 4.

## 4. Emulator Accuracy

We tested the overall accuracy of our emulator using two sets of tests: leave-one-out tests and predictions of the mass function measured in the test simulations. In both of these tests, the quantity we calculated was the residual difference between the emulator prediction and the observed number of halos in a given mass bin. Specifically,

$$R = \frac{n - n_{\text{emu}}}{n_{\text{emu}}}. \quad (12)$$

In order to compare the emulator performance across simulations of different volumes, we have switched to halo number densities  $n$  rather than the actual number of halos  $N$ .

### 4.1. Leave-one-out Tests

The leave-one-out tests consist of retraining the Gaussian processes, leaving out the  $i$ th simulation, and then attempting to predict its mass function. The black points in the left panel of Figure 3 show the fractional difference between the measurement and emulator prediction from these tests. The inverse-variance weighted mean of the absolute value of the residuals is 0.57%. These tests are Poisson limited past the exponential cutoff at  $\nu \approx 3$ , where the residuals begin to fan out. The mass of the exponential cutoff depends on the redshift and cosmological parameters.

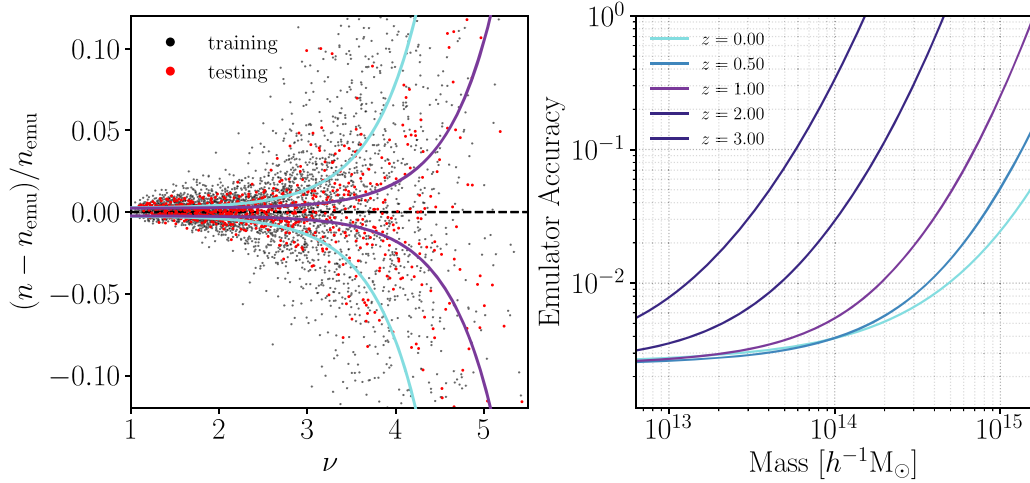
### 4.2. Validation on the Test Simulations

We test the performance of our emulator with the independent test simulations. We first calculated the average mass function for all simulations that used a single cosmology. Within a given snapshot we calculated the average of the five jackknife estimates of the covariance matrix between the mass bins, and divided this by a factor of five to arrive at the covariance matrix for the mean mass function.

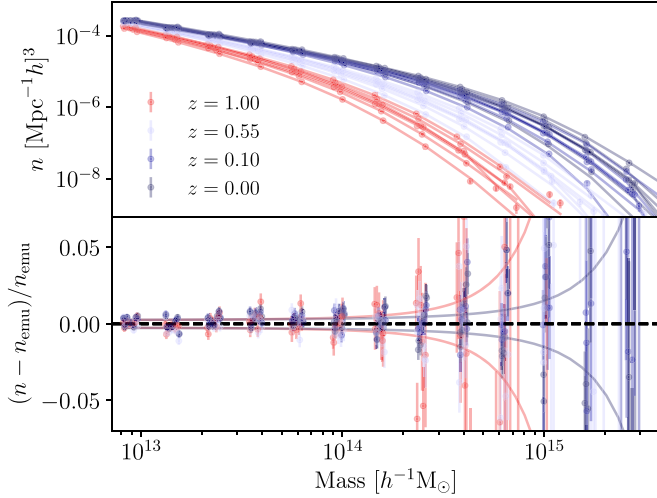
The red points in the left panel of Figure 3 show the fractional difference at all redshifts for all test simulations. The inverse-variance weighted mean of the absolute value of the residuals is 0.47%, lower than the corresponding value for the leave-one-out tests. Figure 4 shows the prediction and residuals for all the test simulations at four redshifts. The lower panel also shows the modeled emulator uncertainty described below in Section 5.

We also test the accuracy of the Tinker et al. (2008) fitting function for the test simulations and compare it to the emulator in Figure 5. The top panels show histograms of the residuals between the simulations and the Tinker et al. (2008) model, while the bottom panels show the emulator, with bars colored by redshift: higher redshifts in red and lower redshifts in blue. The inverse-variance weighted mean of the absolute value of the residuals of the Tinker et al. (2008) model is 3.4%, substantially larger than the 0.47% found for the emulator. The Tinker et al. (2008) model performs worse than the emulator at all redshifts, though the accuracy is consistent with  $\sim 5\%$  at  $z = 0$ . At high redshifts, their model deviates from the simulations significantly, with differences in excess of 100% for snapshots at  $z = [2, 3]$ . In contrast, the emulator is much more consistent with the simulation, achieving subpercent accuracy for all snapshots. Though not apparent from Figure 5, the accuracy of the Tinker et al. (2008) model depends on the cosmological parameters of the simulation, unlike the emulator. This is not unexpected because the Tinker et al. (2008) model

<sup>8</sup> <http://george.readthedocs.io/en/latest/>



**Figure 3.** Left: residual difference between the emulator prediction and the measured mass function for each mass bins in all snapshots of all of the training and test simulations. The light and dark lines indicate the modeled emulator uncertainty (Equation (14)) at  $z = [0, 1]$ , respectively. Note at fixed peak height, the emulator precision improves with increasing redshifts. Right: the modeled emulator accuracy at five redshifts as a function of mass. At fixed mass the emulator accuracy degrades with increasing redshift due to the decreasing halo density with increasing redshift. By contrast, when working at peak height, the opposite trend is seen: higher precision at increasing redshift. The difference reflects the redshift-dependent mapping between mass and peak height. For more details about fitting and verifying the model for the emulator accuracy, see Appendix A.

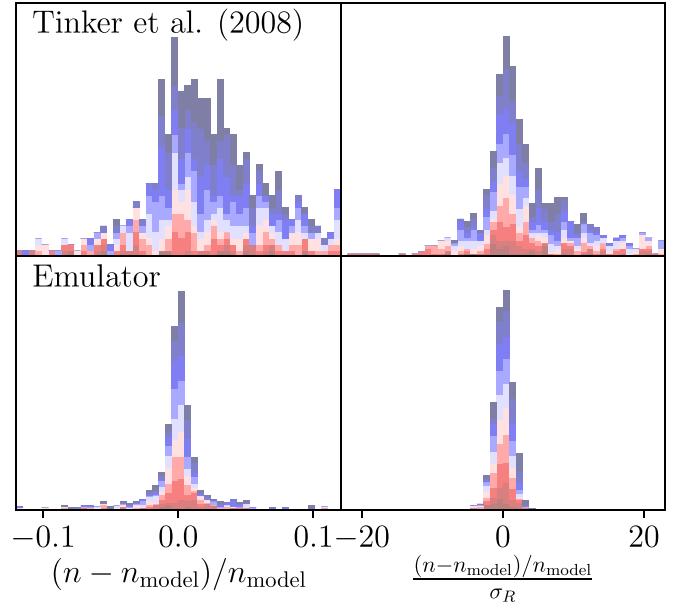


**Figure 4.** Mass function measurements and emulator predictions for the seven test simulations at four redshifts. These simulations were not used to train the Gaussian processes in the emulator, and span the entire range of cosmological design space. Points are placed at the mean halo mass of the corresponding bin, and can sometimes scatter left or right when halos are scarce at high masses. Lines in the lower panel show the modeled accuracy predictions from Equation (14) at  $z = [1, 0]$  in red and blue, respectively.

was calibrated using simulations in a lower-dimensional parameter space. We performed an identical analysis on the mass function to that presented in Tinker et al. (2010) and found comparable results.

### 5. Modeling the Residuals

The leave-one-out tests and comparisons to the test simulations demonstrate that the emulator accurately models the mass function. Additionally, Figure 3 shows that the level of accuracy is mass dependent. We model this mass dependence using the residuals from the test simulations as defined in Equation (12). The covariance between the residuals



**Figure 5.** Residuals from comparing the simulations to the emulator prediction (bottom) and the Tinker et al. (2008) model (top) colored by redshift, where red is high redshift and blue is low redshift. The left panels show the fractional differences, while the right panel show the distribution of  $\chi$ . The residual model has not been included in the denominator of the right panels, in comparison to Equation (17).

of a single mass bin in the  $i$ th snapshot in a given simulation is

$$C_{R,i} = \frac{C_{n,i}}{n_{\text{emu}}^2}, \quad (13)$$

where  $C_n = C_N/V^2$  is the covariance between the cluster counts used in Equation (7) divided by the square of the volume. As before, we ignore covariance between snapshots from the same simulations.

The left panel of Figure 3 demonstrates how the residuals depend on peak height and redshift. We model the mass-dependent emulator uncertainty, or residual model, using a

power-law function in  $\nu$  and scale factor  $a = 1/(1+z)$  given by

$$\sigma_{\text{model}}(\nu, z) = A_R + 10^{b_R + c_R \bar{\nu} + d_R \bar{z}}, \quad (14)$$

where  $\bar{\nu} = \nu - 3$  and  $\bar{z} = \frac{1}{1+z} - 0.5$ . The free parameters are  $A_R$ ,  $b_R$ ,  $c_R$ , and  $d_R$ . This residual model was added to the diagonal of the covariance matrix in Equation (13)

$$\tilde{\mathbf{C}}_{R,i} = \mathbf{C}_{R,i} + \mathbf{I} \sigma_{\text{model}}^2. \quad (15)$$

Additionally, we added in quadrature additional uncertainty along the diagonal equal to half the correction to the abundances described in Section 2.2. This only negligibly affected the recovered residual model. The free parameters of the residual model were found by maximizing the likelihood

$$\ln \mathcal{L}_R = -\frac{1}{2} \sum_i (\mathbf{R}_i^T \tilde{\mathbf{C}}_{R,i}^{-1} \mathbf{R}_i + \text{Indet} \tilde{\mathbf{C}}_{R,i}), \quad (16)$$

where  $\mathbf{R}_i$  contains the residuals of the  $i$ th snapshot of a simulation and the sum runs over all snapshot in all simulations. In this way, we model the residuals of individual snapshots as a Gaussian with zero mean and covariance given by Equation (15).

The residual model evaluated at five redshifts appears in the right panel of Figure 3. At fixed mass the accuracy gets worse with redshift, due to the scarcity of massive halos at high redshift. Notably, the trend of the residual model at fixed peak height as a function of redshift is opposite to the trend for mass. At fixed mass, the accuracy degrades at higher redshifts due to there being fewer halos. The halos found at high redshift reside in the highest peaks of the density field, meaning the model is accurate at high peak height at high redshift. This is due to the nonlinear dependence of  $\nu(M)$ . Additionally, because the Poisson noise is large at high redshift, the residual model described here disfavors a large uncertainty at these redshifts.

In order to assess the performance of the residual model, we computed the distribution of

$$\chi_R = \frac{(n - n_{\text{emu}})/n_{\text{emu}}}{\sqrt{\sigma_R^2 + \sigma_{\text{model}}^2}}, \quad (17)$$

and performed numerous tests detailed in Appendix A to confirm its validity. These included comparing the distribution of  $\chi_R$  for different cuts in mass and redshift, as well as computing the  $\chi^2$  of individual snapshots. The residual model passed all tests, and we found that six out of 668 mass bins with more than 20 halos were more than  $3\sigma$  outliers. Notably, four of the outlier mass bins occurred in the same cosmology.

Modeling the residuals allows us to make random realizations of our model uncertainty. Doing so allows one to propagate the mass-dependent modeling uncertainty forward in an abundance analysis. This is accomplished by drawing random residuals from the recovered model uncertainty. Critically, we do not expect this model uncertainty to oscillate wildly as we vary the peak height  $\nu$ . Rather, we expect the modeling uncertainties to correspond to large-scale fluctuations, meaning fluctuations in neighboring peak-height values and/or neighboring redshifts must be strongly correlated. Given two peak heights  $\nu_1$  and  $\nu_2$  at two different redshifts  $z_1$  and  $z_2$ , we model the correlation coefficient of the model uncertainties with a simple exponential. That is, we set the

covariance matrix of the model uncertainties to be:

$$C(\nu_1, z_1, \nu_2, z_2) = e^{-\Delta\nu^2 - \Delta z^2} \sigma_{\text{model}}(\nu_1, z_1) \sigma_{\text{model}}(\nu_2, z_2), \quad (18)$$

where  $\Delta\nu = \nu_1 - \nu_2$  and  $\Delta z = z_1 - z_2$  and we have assumed a correlation length of 1 for  $\nu$  and  $z$ . This choice was justified by computing the correlation length from the mass function covariance matrices  $\mathbf{C}_N$ , which was approximately unity.

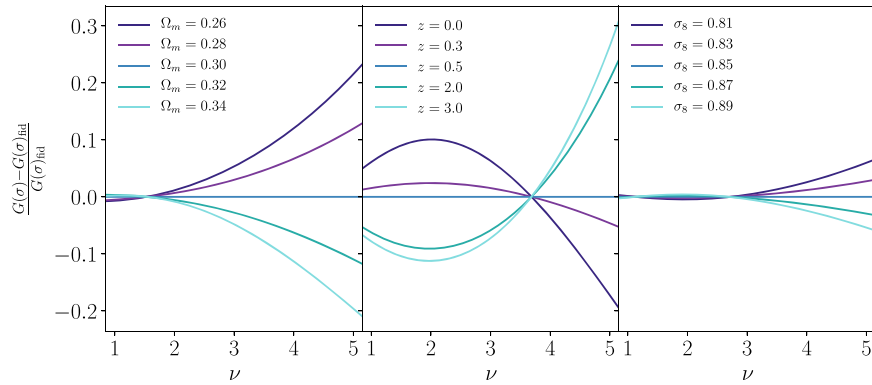
With the covariance matrix of Equation (18), we can model the residual systematics as a Gaussian random field. By defining a grid in  $\nu$  and  $z$ , we can compute the corresponding covariance matrix, and make random realizations along this grid that can be interpolated over in order to arrive at smooth residual functions that fall within the statistically allowed range based on our simulation tests. By performing these random draws as part of an MCMC algorithm, one can effectively marginalize over the theoretical uncertainty in the mass function from our emulator. Our implementation of the emulator includes this capability so that analyses can easily marginalize over the emulator uncertainty.

## 6. Mass Function Universality

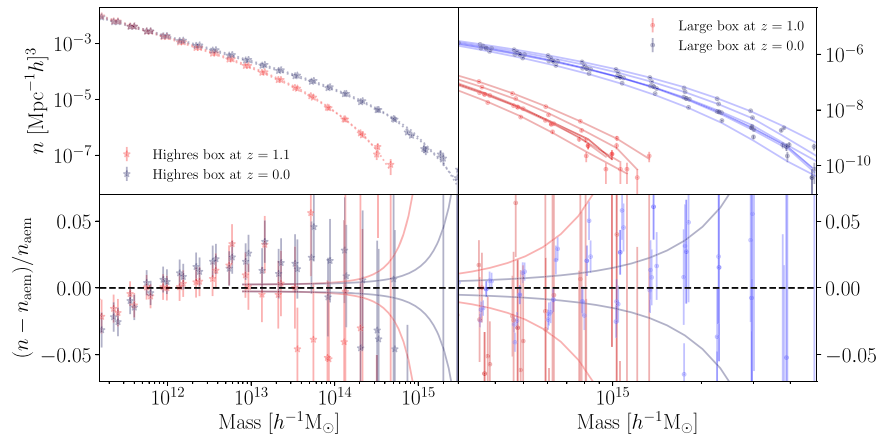
The halo mass function is universal if the halo multiplicity function  $G(\sigma)$  in Equation (2) does not depend explicitly on cosmological parameters or redshift (Jenkins et al. 2001; Evrard et al. 2002). Tinker et al. (2008) found that the mass function was not universal as a function of redshift, or correspondingly large changes in cosmological parameters. This was true whether they used  $M_{200b}$  halos or those defined by  $\Delta_{\text{vir}}$ . Recently, Despali et al. (2016) presented a fitting function to convert SO definitions to  $\Delta_{\text{vir}}$ , which they found to be a more universal definition than others, at least at  $z = 0$ . However, testing the universality of the mass function is limited by the precision of the simulations, and also may be influenced by baryonic effects (see Bocquet et al. 2016). In practice, whether the mass function is universal or not is not relevant for a cluster abundance analysis as long as a model exists that predicts the mass function adequately well. Our emulator explicitly contains nonuniversality by allowing the parameters in Equation (2) to vary with cosmology.

Figure 6 shows the variation of the multiplicity function in the emulator as a function of  $\Omega_m$ ,  $z$ , and  $\sigma_8$ .  $G(\sigma)$  depends sensitively on each of these as well as the other cosmological parameters varied in our simulations. The trend of  $G(\sigma)$  with  $\Omega_m$  are the most straightforward. Recall that, in this panel, all models are constrained to have the same  $\sigma_8$  at  $z = 0$ . Thus, at a given peak height—relative to the mean density—structure collapses earlier when the matter density is lower. Earlier collapse leads to higher halo concentration, and a higher mass within  $R_{\text{vir}}$  at  $z = 0$ . Thus the abundance of halos at fixed  $\nu$  increases with lower  $\Omega_m$ . Similar logic can mostly explain the other trends—halos at fixed  $M$  have higher concentrations at lower redshifts, thus higher  $G(\sigma)$ . And higher values of  $\sigma_8$  lead to more early forming, highly concentrated structure as well. This demonstrates that the halo multiplicity function in our emulator is not universal. This figure serves the same purpose as the sensitivity plots presented in Heitmann et al. (2009), where they demonstrated the dependence of their power spectrum emulator on cosmological parameters.





**Figure 6.** Fractional difference of the halo multiplicity function  $G(\sigma)$  as a function of cosmological parameters and redshift compared to a fiducial cosmology at  $z = 0.5$ . If  $G(\sigma)$  were universal it would not depend on cosmological parameters or redshift. For clarity the plots are shown as a function of peak height  $\nu = \delta_c/\sigma$ , where  $\delta_c = 1.686$ .



**Figure 7.** Comparison between the mass function predicted by the emulator and (A) two high resolution boxes with  $400 h^{-1} \text{Mpc}$  on a side (left), and (B) even simulations with  $3 h^{-1} \text{Gpc}$  on a side (right). Note that the emulator predictions (dotted lines) for the two high resolution simulations nearly overlap. Lines in the lower panels are the emulator modeling uncertainty from Equation (14) plotted in the mass range available to the simulations used to train the emulator. The emulator extrapolates well down to low masses. We conservatively claim an accuracy of 5% to low masses. At high masses the observed abundances are consistent with the emulator predictions.

## 7. Comparison to Other Simulations

To assess the performance at lower halo masses, we compared the emulator to a set of simulations run with  $2048^3$  particles in volumes  $400 h^{-1} \text{Mpc}$  per side resolving halos down to  $\sim 10^{11} h^{-1} M_\odot$  with about 200 particles. These are two of our high resolution suite of simulations. The full high resolution suite is still in progress; it will be comprised of 25 simulations run within the currently allowed parameter space shown in Figure 2. These are not incorporated into the training data of any of our emulators at present.

We also ran additional simulations with  $3 h^{-1} \text{Gpc}$  per side and  $2048^3$  particles in order to resolve high-mass halos of mass up to  $\sim 5 \times 10^{15} h^{-1} M_\odot$ . The cosmological parameters of the large boxes were identical to our test simulations to check for consistency across the simulations. We note that we have seen clear evidence that the finite particle correction described in Section 4.2.5 of DeRose et al. (2018) is resolution dependent, and therefore this correction should not be applied to our large volume simulations. For this reason, when comparing our large volume simulations to our emulator we restrict ourselves to halos with masses  $M \geq 3 \times 10^{14} h^{-1} M_\odot$ , corresponding to halos with  $\geq 1000$  particles.

The comparison to these simulations is shown in Figure 7. Our modeling uncertainty (Section 5) is only valid for the range

of masses used to construct the emulator and may not extrapolate well. However, as can be seen from Figure 7, our emulator extrapolates well down to  $\approx 10^{11} h^{-1} M_\odot$  at the  $\sim 2\%$  level. Due to the possibility of systematics from unconstrained cosmological dependencies and sample variance, we conservatively estimate a precision at these low masses at  $\approx 5\%$ . In the high resolution simulations it appears that the emulator underpredicts the mass function at  $M \gtrsim 10^{13} h^{-1} M_\odot$ . However, this is actually due to the correlation of mass bins below the limits of our training data with those within the limits. That is, the low-mass bins  $< 10^{13} h^{-1} M_\odot$  scattered high, causing the bins above this mass scale to also scatter high. This is reflected in the fact that we found an acceptable  $\chi^2$  value for mass bins within our mass limits, but a poor  $\chi^2$  value when including the low-mass bins. At high mass the emulator is consistent with the large simulations.

## 8. Conclusions

We present an emulator for the halo mass function constructed from halos identified with spherical overdensities 200 times more dense than the background, the  $M_{200b}$  mass definition. This emulator relies on Gaussian processes trained on parameters of a fitting function for the halo multiplicity. In



this way, nonuniversality is directly incorporated into the design of the emulator.

We construct the emulator from a suite of 40 simulations. An additional independent set of 35 simulations is used for validation. These test simulations are comprised of five independent simulations of seven different cosmologies, enabling us to reduce the sample variance for each of the seven test cosmologies in the test simulations by a factor of five.

The performance of the emulator is measured with two sets of tests: leave-one-out tests and comparisons to the test simulations. In both tests the emulator successfully predicted the mass function to better than 1% accuracy in the low-mass (i.e., power-law) regime for the redshifts considered. Validation of the accuracy at the high-mass end (i.e., the exponential tail) is limited by Poisson noise in the number of halos. We successfully modeled the overall uncertainty of the emulator across our entire mass range. Finally, we demonstrated in Figure 1 that the precision achieved by our emulator is sufficiently high for systematic uncertainties in the halo mass function from dark matter simulations to remain negligible for the DES Year 5 and LSST Year 1 cluster abundance analyses.

T.M. and E.R. are supported by DOE grant DE-SC0015975. E.R. acknowledges additional support by the Sloan Foundation, grant FG-2016-6443. J.L.T. and R.H.W. acknowledge support of NSF grant AST-1211889. J.D., R.H.W., S.M., and M.R.B. received support from the U.S. Department of Energy under contract number DE-AC02-76SF00515. Y.Y.M. is supported by the Samuel P. Langley PITT PACC Postdoctoral Fellowship. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

*Software:* Python, Matplotlib (Hunter 2007), NumPy (van der Walt et al. 2011), SciPy (Jones et al. 2001), emcee (Foreman-Mackey et al. 2013), george (Ambikasaran et al. 2015), CAMB (Lewis & Bridle 2002), CLASS (Lesgourgues 2011), ROCKSTAR (Behroozi et al. 2013), GADGET (Springel 2005), 2LPT (Crocco et al. 2006).

## Appendix A

### Assessing the Emulator Uncertainty Model

Residuals between the emulator and the simulations occur for three reasons: systematic differences between the emulator and the true mass function, Poisson noise in the number of halos, and sample variance due to the particular realization of density modes present in the simulation. Poisson noise and sample variance are estimated through the jackknife covariance matrix from each simulation snapshot. The accuracy of the emulator  $\sigma_{\text{model}}$  is calibrated as described in detail in Section 5. To validate the estimated accuracy of the emulator, we define

$$\chi_R = \frac{(n - n_{\text{emu}})/n_{\text{emu}}}{\sqrt{\sigma_R^2 + \sigma_{\text{model}}^2}}. \quad (19)$$

Figure 8 shows the distribution of  $\chi_R$  for two splits of the residuals, by mass and by redshift, for all snapshots for bins with more than 20 halos. The splits were chosen to have approximately the same number of residuals on either side. Outliers ( $>3\sigma$ ) occurred as follows: 1 out of 43 bins at  $z = 3$ , 3

out of 44 bins at  $z = 2$ , and 1 out of both 83 and 85 bins at  $z = 0.1$  and  $z = 0$ , respectively. This is 6 mass bins out of a total of 668 across all redshifts in all simulations. The Tinker et al. (2008) model was significantly more discrepant at all redshifts compared to the emulator.

While Figure 8 is useful for visualization, the fact that different bins are correlated implies that we cannot use the distribution of residuals  $\chi_R$  to establish goodness of fit. Instead, we compute the total  $\chi^2$  of the emulator across all simulations for each individual snapshot. For each individual snapshot of each simulation, we computed

$$\chi^2 = \mathbf{RC}^{-1}\mathbf{R}^T, \quad (20)$$

where  $\tilde{\mathbf{C}}$  is defined in Equation (15), and we incorporate the finite resolution correction discussed in Section 2.2 as additional uncertainty. We then summed over all our test simulations to arrive at a final  $\chi^2$ . We found that all snapshots exhibited an acceptable total  $\chi^2$  across all simulations for  $z \leq 2$ . At  $z = 3$ , the  $\chi^2$  of the emulator is slightly high, yielding  $\chi^2/\text{dof} = 47/28$ . The worst of the remaining  $\chi^2$  values is  $\chi^2/\text{dof} = 96/85$  at  $z = 0$ . These results demonstrate that the emulator accurately predicts the mass function up to at least  $z = 2$ . At  $z = 3$ , the emulator might be slightly over-optimistic, though the difference is clearly not strongly significant.

## Appendix B

### Accuracy Requirements

The required accuracy of the halo mass function is set by considering the uncertainty on cluster abundances in real surveys. In a survey, clusters are binned by some observable that is related to mass, for example, redMaPPer richness (Rykoff et al. 2014), X-ray luminosity or temperature, or Sunyaev–Zeldovich signal. The mean cluster mass of objects in these bins is measured using techniques such as weak lensing. If clusters are binned by mass, the number of clusters in a single bin is

$$\ln N = \Delta \ln M \frac{d \ln N}{d \ln M}, \quad (21)$$

where we assume a fiducial bin width of  $\Delta \ln M = 0.2 \ln 10$ , or five bins per decade. If the bin has a true mean mass  $\ln M_b$  measured with some uncertainty  $\sigma_{\ln M}$ , then error in the abundance is found via propagation of errors. One has

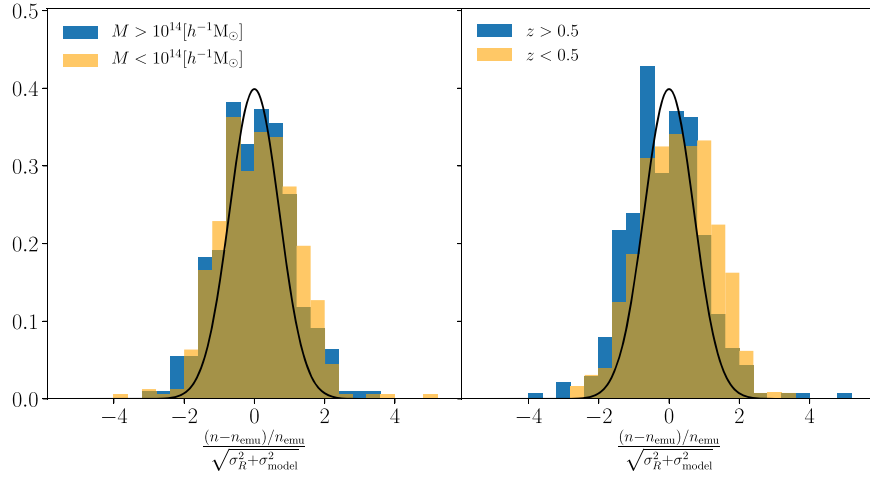
$$\ln N(\ln M_b + \sigma_{\ln M}) = \Delta \ln M \frac{d \ln N}{d \ln M} \Big|_{\ln M_b + \sigma_{\ln M}} \quad (22)$$

$$= \Delta \ln M \left( \frac{d \ln N}{d \ln M} \Big|_{\ln M_b} + \sigma_{\ln M} \frac{d^2 \ln N}{d \ln M^2} \Big|_{\ln M_b} \right) \quad (23)$$

$$= \ln N + \Delta \ln M \sigma_{\ln M} \frac{d^2 \ln N}{d \ln M^2} \Big|_{\ln M_b}. \quad (24)$$

Note that  $d^2 \ln N / d \ln M^2$  is the first derivative of the halo mass function. The uncertainty on the abundance due to uncertainty in the mass is

$$\sigma_{\ln N, M} = \Delta \ln M \sigma_{\ln M} \frac{d^2 \ln N}{d \ln M^2} \Big|_{\ln M_b}. \quad (25)$$



**Figure 8.** Histograms of the residuals of all snapshots for all testing simulations divided by the square root of their variances plus the emulator uncertainties added in quadrature, as compared to a unit normal distribution in black. Left: residuals split by mass at  $10^{14} h^{-1} M_{\odot}$ . Right: residuals split by redshift at  $z = 0.5$ . We caution that this figure cannot be used to gauge whether the calibration of the emulator uncertainty is correct via  $\chi$ -by-eye, as neighboring mass and redshift bins are correlated. A goodness-of-fit test that accounts for these covariances results in acceptable  $\chi^2$  values.

McClintock et al. (2019) found the statistical uncertainty on the mean masses of clusters in the DES Y1 data to be  $\sigma_{\ln M} \sim 8\%$ . Of course, this uncertainty will change between surveys and is sensitive to lens source density, survey area, and survey depth. We consider two contributions to the uncertainty in weak lensing mass calibration: shape noise (the dominant source of statistical uncertainty for the DES), and cluster sample variance

$$\sigma_{\ln M}^2 = \sigma_{\ln M}^{2,\text{SN}} + \sigma_{\ln M}^{2,\text{SV}}. \quad (26)$$

The sample variance is easily understood:

$$\sigma_{\ln M}^{2,\text{SV}} = \frac{\text{Var}(M|\lambda)}{N} = \frac{0.4^2}{N}, \quad (27)$$

where the scatter in mass at fixed richness is roughly 40%. This term is subdominant to the shape noise, which is more complicated. Melchior et al. (2017) found that  $M \propto \Delta\Sigma^{4/3}$ , where  $\Delta\Sigma$  is the weak lensing profile. This means that the uncertainty in the mass due to shape noise is given by

$$\sigma_M^{2,\text{SN}} = \sigma_{\Delta\Sigma}^2 M^{1/2} = AM^{1/2} (N_c n)^{-1}, \quad (28)$$

where  $N_c$  is the number of clusters,  $A$  is some constant, and  $n$  is the lensing source density. However, because clusters at higher redshift have fewer sources behind them, this term is redshift dependent. In detail this dependence is complicated, but it will generally resemble

$$n(z) \approx n_0 \exp\left(-\frac{1}{2} \frac{z^2}{z_*^2}\right), \quad (29)$$

where  $z_*$  is a characteristic redshift where the local source density peaks. For DES Y1,  $n_0$  is  $6.3 \text{ arcmin}^{-2}$ . We found that the parameters  $A = 0.7 \times 10^{22} [h^{-1} M_{\odot}]^{1.5} \text{ arcmin}^{-2}$  and  $z_* = 0.33$  recovered the uncertainties of the DES Y1 masses within 10%. In a generic mass bin at mass  $M$  for a given survey, we can use these values to compute the shape noise uncertainty using

$$\sigma_M^{2,\text{SN}} = \frac{AM^{1/2} 6.3 \text{ arcmin}^{-2}}{N n_0} \exp\left(-\frac{1}{2} \frac{z^2}{z_*^2}\right). \quad (30)$$

The  $z_*$  values for DES Y1, Y5, and LSST Y1 are 0.5, 0.6, and 8.5 respectively. The number of clusters in the bin  $N$  scales with the survey area, and is calculated within a redshift slice of  $\Delta z = 0.15$ , the width in redshift of the cluster bins in McClintock et al. (2019). The lens source densities for DES Y5 and LSST Y1 are estimated to be 8.4 and 10  $\text{arcmin}^{-2}$ . The survey areas for DES Y1, Y5, and LSST Y1 are 1514, 5000, and 18000  $\text{deg}^2$ . The shape noise uncertainty in  $\ln M$  is  $\sigma_{\ln M}^{\text{SN}} = \sigma_M^{\text{SN}} / M$ , which allows us to compute the total uncertainty on the mass bin  $\sigma_{\ln M}$ . From this we calculate the total uncertainty on the abundance as

$$\sigma_{\ln N}^2 = \sigma_{\ln N, \text{emu}}^2 + \sigma_{\ln N, M}^2, \quad (31)$$

where  $\sigma_{\ln N, \text{emu}}$  is the accuracy of the halo mass function emulator. We require that adding the emulator uncertainty increases the error budget for each survey by no more than 10%, so that  $\sigma_{\ln N}^2 = 1.1^2 \sigma_{\ln N, M}^2$ . We use this equation to derive the calibration requirements shown in Figure 1.

## ORCID iDs

Thomas McClintock <https://orcid.org/0000-0001-8764-5271>  
 Eduardo Roza <https://orcid.org/0000-0002-1666-6275>  
 Matthew R. Becker <https://orcid.org/0000-0001-7774-2246>  
 Joseph DeRose <https://orcid.org/0000-0002-0728-0960>  
 Yao-Yuan Mao <https://orcid.org/0000-0002-1200-0820>  
 Sean McLaughlin <https://orcid.org/0000-0002-7042-6424>  
 Risa H. Wechsler <https://orcid.org/0000-0003-2229-011X>  
 Zhongxu Zhai <https://orcid.org/0000-0001-7984-5476>

## References

- Aihara, H., Armstrong, R., Bickerton, S., et al. 2018, *PASJ*, 70, S8  
 Albrecht, A., Bernstein, G., Cahn, R., et al. 2006, arXiv:astro-ph/0609591  
 Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W., & O’Neil, M. 2015, *ITPAM*, 38, 2  
 Anderson, L., Aubourg, É., Bailey, S., et al. 2014, *MNRAS*, 441, 24  
 Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013, *ApJ*, 762, 109  
 Bhattacharya, S., Heitmann, K., White, M., et al. 2011, *ApJ*, 732, 122  
 Bleem, L. E., Stalder, B., de Haan, T., et al. 2015, *ApJS*, 216, 27  
 Bocquet, S., Saro, A., Dolag, K., & Mohr, J. J. 2016, *MNRAS*, 456, 2361

- Crocce, M., Pueblas, S., & Scoccimarro, R. 2006, *MNRAS*, 373, 369
- Dark Energy Survey Collaboration 2005, arXiv:astro-ph/0510346
- Dark Energy Survey Collaboration 2016, *MNRAS*, 460, 1270
- de Jong, J. T. A., Kleijn, G. A. V., Erben, T., et al. 2017, *A&A*, 604, A134
- DeRose, J., Wechsler, R. H., Tinker, J. L., et al. 2018, arXiv:1804.05865
- Despali, G., Giocoli, C., Angulo, R. E., et al. 2016, *MNRAS*, 456, 2486
- Dietrich, J. P., Bocquet, S., Schrabback, T., et al. 2019, *MNRAS*, 483, 2871
- Dodelson, S., Heitmann, K., Hirata, C., et al. 2016, arXiv:1604.07626
- Evrard, A. E., MacFarland, T. J., Couchman, H. M. P., et al. 2002, *ApJ*, 573, 7
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306
- Hartlap, J., Simon, P., & Schneider, P. 2007, *A&A*, 464, 399
- Heitmann, K., Bingham, D., Lawrence, E., et al. 2016, *ApJ*, 820, 108
- Heitmann, K., Higdon, D., White, M., et al. 2009, *ApJ*, 705, 156
- Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, *ApJS*, 208, 19
- Hunter, J. D. 2007, *CSE*, 9, 90
- Jenkins, A., Frenk, C. S., White, S. D. M., et al. 2001, *MNRAS*, 321, 372
- Jones, E., Oliphant, T., Peterson, P., et al. 2001, SciPy: Open source scientific tools for Python, <https://www.scipy.org/>
- Lesgourgues, J. 2011, arXiv:1104.2932
- Lewis, A., & Bridle, S. 2002, *PhRvD*, 66, 103511
- Mantz, A. B., Allen, S. W., Morris, R. G., et al. 2016, *MNRAS*, 463, 3582
- McClintock, T., Varga, T. N., Gruen, D., et al. 2019, *MNRAS*, 482, 1352
- Mehrtens, N., Romer, A. K., Hilton, M., et al. 2012, *MNRAS*, 423, 1024
- Melchior, P., Gruen, D., McClintock, T., et al. 2017, *MNRAS*, 469, 4899
- Miyazaki, S., Komiyama, Y., Nakaya, H., et al. 2012, *Proc. SPIE*, 8446, 84460Z
- More, S., Kravtsov, A. V., Dalal, N., & Gottlöber, S. 2011, *ApJS*, 195, 4
- Piffaretti, R., Arnaud, M., Pratt, G. W., Pointecouteau, E., & Melin, J.-B. 2011, *A&A*, 534, A109
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, *A&A*, 571, A16
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, *A&A*, 594, A24
- Press, W. H., & Schechter, P. 1974, *ApJ*, 187, 425
- Radovich, M., Puddu, E., Bellagamba, F., et al. 2017, *A&A*, 598, A107
- Rasmussen, C. E., & Williams, C. K. I. 2005, Gaussian Processes for Machine Learning (Cambridge, MA: MIT Press)
- Reed, D. S., Bower, R., Frenk, C. S., Jenkins, A., & Theuns, T. 2007, *MNRAS*, 374, 2
- Reed, D. S., Smith, R. E., Potter, D., et al. 2013, *MNRAS*, 431, 1866
- Rykoff, E. S., Rozo, E., Busha, M. T., et al. 2014, *ApJ*, 785, 104
- Rykoff, E. S., Rozo, E., Hollowood, D., et al. 2016, *ApJS*, 224, 1
- Sheth, R. K., Mo, H. J., & Tormen, G. 2001, *MNRAS*, 323, 1
- Sheth, R. K., & Tormen, G. 1999, *MNRAS*, 308, 119
- Simet, M., McClintock, T., Mandelbaum, R., et al. 2017, *MNRAS*, 466, 3103
- Springel, V. 2005, *MNRAS*, 364, 1105
- Stern, C., Dietrich, J. P., Bocquet, S., et al. 2019, *MNRAS*, in press (10.1093/mnras/stz234/5298897)
- Suzuki, N., Rubin, D., Lidman, C., et al. 2012, *ApJ*, 746, 85
- Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, *ApJ*, 688, 709
- Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, *ApJ*, 724, 878
- van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *CSE*, 13, 22
- Warren, M. S., Abazajian, K., Holz, D. E., & Teodoro, L. 2006, *ApJ*, 646, 881
- Weinberg, D. H., Mortonson, M. J., Eisenstein, D. J., et al. 2013, *PhR*, 530, 87
- Zhai, Z., Tinker, J. L., Becker, M. R., et al. 2018, arXiv:1804.05867
- Zu, Y., Weinberg, D. H., Rozo, E., et al. 2014, *MNRAS*, 439, 1628