






# The Aemulus Project. III. Emulation of the Galaxy Correlation Function

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

<sup>1</sup> Center for Cosmology and Particle Physics, Department of Physics, New York University, 4 Washington Place, New York, NY 10003, USA; [zz681@nyu.edu](mailto:zz681@nyu.edu)

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

Received 2018 April 16; revised 2019 February 26; accepted 2019 March 5; published 2019 March 27

## Abstract

Using the  $N$ -body simulations of the AEMULUS Project, we construct an emulator for the nonlinear clustering of galaxies in real and redshift space. We construct our model of galaxy bias using the halo occupation framework, accounting for possible velocity bias. The model includes 15 parameters, including both cosmological and galaxy bias parameters. We demonstrate that our emulator achieves  $\sim 1\%$  precision at the scales of interest,  $0.1 h^{-1} \text{ Mpc} < r < 10 h^{-1} \text{ Mpc}$ , and recovers the true cosmology when tested against independent simulations. Our primary parameters of interest are related to the growth rate of structure,  $f$ , and its degenerate combination,  $f\sigma_8$ . Using this emulator, we show that the constraining power on these parameters monotonically increases as smaller scales are included in the analysis, all the way down to  $0.1 h^{-1} \text{ Mpc}$ . For a BOSS-like survey, the constraints on  $f\sigma_8$  from  $r < 30 h^{-1} \text{ Mpc}$  scales alone are nearly a factor of two tighter than those from the fiducial BOSS analysis of redshift-space clustering using perturbation theory at larger scales. The combination of real- and redshift-space clustering allows us to break the degeneracy between  $f$  and  $\sigma_8$ , yielding an 11% constraint on  $f$  alone for a BOSS-like analysis. The current AEMULUS simulations limit this model to surveys of massive galaxies. Future simulations will allow this framework to be extended to all galaxy target types, including emission-line galaxies.

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

## 1. Introduction

The spatial distribution of luminous matter is one of the key windows to understanding the distribution and properties of the energy density of the universe. As galaxy clustering has emerged as an essential tool in our effort to understand the accelerated expansion of the universe, the amount of data from galaxy redshift surveys has increased by orders of magnitude over the past decade. Groundbreaking results from the first truly large-scale redshift surveys, the Sloan Digital Sky Survey (SDSS-I/II; York et al. 2000; Abazajian et al. 2009) and the Two Degree Field Galaxy Redshift Survey (2dFGRS; Colless et al. 2001; Cole et al. 2005), have spawned successor programs such as BOSS (Dawson et al. 2013), eBOSS (Dawson et al. 2016), and WiggleZ (Drinkwater et al. 2010), which are either complete or in progress. Near-term surveys such as the Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration et al. 2016), 4MOST (de Jong et al. 2016), and PFS (Takada et al. 2014) will represent another leap in our ability to create maps of the universe. With the current and near-future surveys, we expect to have taken tens of millions of spectra as a community, covering the last 10 billion yr of the history of the universe. In this paper, we propose a new method to make more efficient use of these spectra and increase the constraining power of these data sets. We will demonstrate that incorporating clustering information at nonlinear scales can more than double the power of these data to constrain the growth of structure.

Most of the applications of these data sets have focused on the retrieval of cosmological information from large scales. The appearance of the baryon acoustic peak at  $\sim 100 h^{-1} \text{ Mpc}$

allows galaxy clustering to be used as a standard ruler in geometric probes of the expansion history of the universe, but the detailed shape and amplitude of the measured correlation function also contain significant information. The amplification of clustering through galaxy peculiar velocities—an effect called redshift-space distortions (RSDs)—has become the primary method for measuring the growth of structure using spectroscopic surveys. Such measurements are complementary to geometric probes of the universe because they are less sensitive to variations in the equation of state of dark energy and more sensitive to possible variations in the underlying theory of gravity. Both theories have been proposed to explain the accelerated expansion, but as yet we do not have the data to determine which class of theories is correct.

The design of current and future surveys is built around the rule of thumb  $nP \approx 1$ , where  $n$  is the space density of the targets and  $P$  is the amplitude of the power spectrum in the region of interest (DESI Collaboration et al. 2016). Adhering to this rule creates a survey that is not shot noise-limited but maximizes volume by limiting the sampling of the density field with a fixed budget of observing time. However, in such survey designs, typical clustering measurements are most precise at the Mpc scale. This is far below the minimum scale considered for analyses of the shape of clustering measurements or the impact of RSD. These analyses are based on variations of higher-order perturbation theory and usually model clustering at scales down to 30–40 Mpc (e.g., Carlson et al. 2009, 2013).

Retrieving information from these scales has been a goal of modern cosmology, but it has also been a challenge (Kwan et al. 2015). Nonlinear dynamics of dark matter are captured with excellent precision in modern cosmological  $N$ -body

simulations (see, e.g., Klypin et al. 2011, 2016). The challenge of constraining cosmology with such simulations is twofold: (1) an accurate and flexible model of the galaxy bias is required, and (2) one needs to be able to properly sample cosmological parameter space, which becomes computationally intractable for a standard Monte Carlo Markov chain (MCMC) analysis. Because of these limitations, the amount of information that is extractable from small-scale galaxy clustering is simply unknown. The measurement precision of the data is orders of magnitude higher than at large scales, but the theoretical complexity increases significantly. How much information is recoverable after accounting for all possibilities in galaxy bias? In this paper, for the case of RSD, we will show that after marginalizing over numerous galaxy bias parameters and incorporating the theoretical uncertainty in the galaxy clustering model, it is still possible to extract more constraining power from the growth of structure measurements than what is achievable using perturbation theory on large scales.

To solve problem (1), we use the halo occupation distribution (HOD; Benson et al. 2000; Peacock & Smith 2000; Seljak 2000; White et al. 2001; Berlind & Weinberg 2002; Cooray & Sheth 2002). The HOD approaches galaxy bias by quantifying the statistical relationship between galaxies and dark matter halos. In its most basic form, the HOD is mostly determined by the probability distribution  $P(N|M)$ , the probability that a halo of mass  $M$  contains  $N$  galaxies of a given class. Once  $P(N|M)$  is combined with prescriptions for the spatial and velocity bias of galaxies within halos, this model offers a nearly complete description of the spatial distribution of galaxies for a given halo population. This simple approach of  $P(N|M)$ , however, ignores the possibility that  $N$  may depend on some secondary halo property. If this halo property is correlated with the spatial distribution of halos, this could create a “galaxy secondary bias” (also known as galaxy assembly bias) that would have to be incorporated in the probability distribution in order to create a fully descriptive HOD (Sheth & Tormen 2004; Gao et al. 2005; Harker et al. 2006; Wechsler et al. 2006). The optimal method for incorporating galaxy assembly bias into the HOD and tests against models that contain these effects is left to another paper (S. McLaughlin et al. 2018, in preparation). Our emphasis here is on determining the total constraining power and the scales from which these constraints come, under the assumption that the assumed HOD approach is sufficient for modeling galaxy bias.

To solve problem (2), we use a combination of novel space-filling algorithms and statistical techniques to create an *emulator* for galaxy clustering. Our approach follows from the work of Heitmann et al. (2009, 2010), Lawrence et al. (2010), and Heitmann et al. (2014), who created an emulator for the nonlinear matter power spectrum. Although full coverage of parameter space with simulations is infeasible, advancements in computing technology and force-calculation algorithms have pushed the field to a state where suites of simulations can be produced and analyzed in a tractable amount of time. If parameter space is properly sampled, novel interpolation schemes can be used to create high-accuracy estimates of statistics at any point within the space. The simulations and the parameter space spanned by the emulator in this paper are described in detail in DeRose et al. (2018). These simulations and their application constitute the AEMULUS Project. Use of these simulations to emulate the halo mass function

is presented in McClintock et al. (2019). The goals of the AEMULUS Project are not limited to these statistics; these papers are the first step toward a full accounting of the extractable information from small-scale galaxy clustering, including higher-order statistics, void statistics, galaxy–galaxy lensing, and numerous combinations of galaxies with clusters.

Our goal in this paper is to construct an emulator that achieves 1% accuracy in its prediction of the real-space correlation function and the monopole and quadrupole of the redshift-space correlation function over the scales  $1 h^{-1}\text{Mpc} \lesssim r \lesssim 10 h^{-1}\text{Mpc}$ . Motivated by both the mass resolution of the AEMULUS simulations and the massive amount of data on massive galaxies, we construct our emulator to model the class of galaxies known as luminous red galaxies (LRGs). Compiling the data from the completed SDSS-I/II and BOSS surveys, as well as the ongoing eBOSS survey, yields nearly two million spectra sampling a volume of nearly  $10 \text{ Gpc}^3$  (Dawson et al. 2016). This is the ideal target sample to begin the exploration of the constraining power of small-scale clustering. In a pilot study, Reid et al. (2014) used simulations of a single cosmology to analyze small-scale RSD measurements. Their analysis yielded a constraint on the growth of structure—through the parameter combination  $f\sigma_8$ —of 2.5%, a factor of four smaller than analyses using large-scale measurements combined with perturbation theory. In this paper, we take the next step forward in this analysis, expanding both the cosmological parameter space and the halo occupation parameter space to produce a robust model with which to analyze massive galaxy data sets.

Our paper is organized as follows. In Section 2, we introduce the method of the emulator and a trial application to the correlation function calculated with analytical methods in real space. Section 3 presents the emulator for the correlation function measured from  $N$ -body simulations and the performance. Section 4 shows that we can achieve unbiased recovery of cosmological parameters with the emulator and explores the constraining power of small-scale clustering. We discuss and list our conclusions in Section 5.

## 2. Constructing the Emulator

Here we discuss the observables, cosmological parameter space, HOD parameter space, and implementation of the Gaussian process (GP) emulator. We first test this process by building an emulator around an analytical model for one of our observables, the projected two-point galaxy correlation function  $w_p(r_p)$ . For this statistic, there is a robust analytic model for nonlinear galaxy clustering using halo occupation as the galaxy bias model. Here we use the specific implementation described in Tinker et al. (2005, 2012). This model is accurate to 5%–10%, but the accuracy of the model is less relevant than its role as a means by which to test the accuracy of our emulator given the sampling of parameter space, the expected error of our training sample simulations, and the GP itself.

### 2.1. Galaxy Clustering: The Observables

The clustering property of the galaxies can be characterized by the two-point correlation function  $\xi(r)$ , which is defined as a measure of the excess probability, relative to a Poisson distribution, of finding two galaxies at the volume elements  $dV_1$  and  $dV_2$  separated by a vector distance  $\mathbf{r}$  (Peebles 1980):

$$dP_{12} = n^2 [1 + \xi(r)] dV_1 dV_2, \quad (1)$$

**Table 1**  
The Parameters Used in Our Emulator, Their Physical Meaning, and the Range in the Parameter Space

	Parameter	Meaning	Range
Cosmology	$\Omega_m$	The matter energy density	[0.255, 0.353]
	$\Omega_b$	The baryon energy density	[0.039, 0.062]
	$\sigma_8$	The amplitude of matter fluctuations on $8 h^{-1}$ Mpc scales	[0.575, 0.964]
	$h$	The dimensionless Hubble constant	[0.612, 0.748]
	$n_s$	The spectral index of the primordial power spectrum	[0.928, 0.997]
	$w^a$	The dark energy equation of state	[-1.40, -0.57]
	$N_{\text{eff}}^a$	The number of relativistic species	[2.62, 4.28]
	$\gamma_f^a$	The amplitude of the halo velocity field relative to $w\text{CDM}+\text{GR}$	[0.5, 1.5]
HOD	$\log M_{\text{sat}}$	The typical mass scale for halos to host one satellite	[13.8, 14.5]
	$\alpha$	The power-law index for the mass dependence of the number of satellites	[0.2, 1.8]
	$\log M_{\text{cut}}$	The mass cutoff scale for the satellite occupation function	[10.0, 13.7]
	$\sigma_{\log M}$	The scatter of halo mass at fixed galaxy luminosity	[0.05, 0.6]
	$\eta_{\text{con}}^a$	The concentration of satellites relative the dark matter halo	[0.2, 2.0]
	$\eta_{\text{vc}}^a$	The velocity bias for central galaxies	[0.0, 0.7]
	$\eta_{\text{vs}}^a$	The velocity bias for satellite galaxies	[0.2, 2.0]

**Note.**

<sup>a</sup> This parameter is not used in the emulator of  $w_p$  with the analytic method.

where  $n$  is the mean number density over the whole sample volume. For a pair of galaxies with redshift-space positions  $s_1$  and  $s_2$ , the dependence of the correlation function is only through  $s = s_1 - s_2$  and the orientation of  $s$  relative to the line of sight.<sup>7</sup> In this case, we may calculate the correlation function of a two-dimensional grid of separations perpendicular ( $r_p$ ) and parallel ( $\pi$ ) to the line of sight  $\xi_Z(r_p, \pi)$  through

$$\pi = \frac{s \cdot l}{|l|}, \quad r_p = s \cdot s - \pi^2, \quad (2)$$

with  $l = (s_1 + s_2)/2$  (Davis & Peebles 1983; Fisher et al. 1994).

In order to mitigate the effect of RSDs and examine the real-space correlation function, we compute the projected correlation function from  $\xi(r_p, \pi)$  (Davis & Peebles 1983):

$$w_p(r_p) = 2 \int_0^\infty d\pi \xi_Z(r_p, \pi) = 2 \int_0^\infty d\pi \xi_R(r = \sqrt{r_p^2 + \pi^2}). \quad (3)$$

We truncate the integrand to  $\pi_{\text{max}} = 80 h^{-1}$  Mpc, which is large enough to include most of the correlated pairs. The projected correlation function eliminates the effect of RSD at all scales  $r_p < 10 h^{-1}$  Mpc. Coherent inflows can change the amplitude of the  $w_p$  at larger scales when  $\pi_{\text{max}}$  is finite, but in our analytic method, we account for this using the linear theory of Kaiser (1987). In subsequent calculation based on simulations, any impact of finite  $\pi_{\text{max}}$  is already built in to the same result.

Equivalently, we can also write  $\xi_Z(s_1, s_2)$  as  $\xi_Z(s, \mu)$ , where  $\mu = r_p/s$ . We can then expand the correlation function in harmonics of  $\mu$  for a given  $s$ , and the resulting Legendre multipoles  $\xi_\ell(s)$  are given by

$$\xi_\ell(s) = \frac{2\ell + 1}{2} \int_{-1}^1 L_\ell(\mu) \xi_Z(s, \mu) d\mu, \quad (4)$$

<sup>7</sup> We differentiate the correlation function in real and redshift space as  $\xi_R$  and  $\xi_Z$ , respectively.

where  $L_\ell$  is the Legendre polynomial of order  $\ell$ .

## 2.2. Cosmological Parameter Space

In our analytic calculation, we apply the spatially flat  $\Lambda\text{CDM}$  model with the following parameters: matter density  $\Omega_m$ , baryon density  $\Omega_b$ , Hubble constant  $h \equiv H_0(100 \text{ km s}^{-1} \text{ Mpc}^{-1})^{-1}$ , spectral index of the primordial perturbation  $n_s$ , and perturbation amplitude  $\sigma_8$ . In the simulation-based emulator, we consider a flat  $w\text{CDM}$  with the following additional parameters: the constant equation of state of dark energy  $w_0$  and the number of relativistic species  $N_{\text{eff}}$ . We note that these parameters have minimal impact on the small-scale spatial clustering of galaxies. The final parameter used in the simulation-based emulator is  $\gamma_f$ , the amplitude of the halo velocity field relative to the  $w\text{CDM} + \text{general relativity (GR)}$  prediction. The growth rate of the structure is the logarithmic derivative of the growth factor,  $D$ , with  $f = d \ln D / d \ln a$ . Thus,  $f$  determines the amplitude of the matter velocity field, which in turn is a product of the matter density and theory of gravity. Our parameter  $\gamma_f$  is defined as  $f/f_{w\text{CDM}}$ , and all halo velocities are rescaled by this value.<sup>8</sup> Therefore,  $\gamma_f$  is a test of gravity: within the cosmological parameter space allowed by the cosmic microwave background (CMB) and geometric constraints, can the data be fit with a model where  $\gamma_f$  is consistent with unity?

The emulator is based on an efficient parameter sampling strategy and an effective interpolation scheme. The former is realized by a Latin hypercube method as introduced in Heitmann et al. (2009), while the latter uses a GP. As mentioned in the Introduction, we first build the emulator with the analytic method, which can provide a thorough estimate of the error and its dependence on scale and location in parameter space. We calculate the matter power spectrum from Eisenstein & Hu (1998) and ignore the effect from the equation of state of dark energy and extra relativistic species. This results in a five-dimensional cosmological model with parameters  $\Omega_m$ ,  $\Omega_b$ ,  $n_s$ ,  $\sigma_8$ , and  $h$ . The 40 training cosmologies for the analytic emulator, as well as seven test cosmologies, are shown in

<sup>8</sup> In Reid et al. (2014),  $\gamma_f$  was labeled  $\gamma_{\text{HV}}$ .

Figure 3 of DeRose et al. (2018). The details of the design of the cosmologies can be found in their Table 1. The ranges of the HOD parameters used in this paper are summarized in Table 1.

### 2.3. Halo Occupation Parameter Space

In this paper, we apply the HOD framework, which approaches the problem of galaxy bias in a statistical way. In its most basic form, the HOD constructs a probability distribution  $P(N|M)$ : the probability that a halo of mass  $M$  contains  $N$  galaxies of a given class. Because the clustering, abundance, and interior structure of dark matter halos is well known from simulations, specifying  $P(N|M)$  provides a complete description of the spatial distribution of galaxies. This description is only complete if  $N$  depends only on halo mass  $M$ . If  $N$  depends on some secondary halo property, and this halo property depends on the spatial distribution of the halos, this could create a “galaxy assembly bias” that would have to be incorporated into the HOD model to be fully descriptive. We leave this to another paper (S. McLaughlin et al. 2018, in preparation).

For the HOD parameterization, it is necessary to separate the contribution of the central galaxies from that of the satellite galaxies with the mean occupancy of halos:

$$\langle N(M) \rangle = \langle N_{\text{gal}}(M) \rangle = \langle N_{\text{cen}}(M) \rangle + \langle N_{\text{sat}}(M) \rangle. \quad (5)$$

The mean number of central galaxies in each halo is modeled with a smooth transition between zero and one galaxy,

$$N_{\text{cen}}(M) = \frac{1}{2} \left[ 1 + \operatorname{erf} \left( \frac{\log_{10} M - \log_{10} M_{\text{min}}}{\sigma_{\log M}} \right) \right], \quad (6)$$

and the mean number of satellite galaxies is parameterized as

$$N_{\text{sat}}(M) = \left( \frac{M}{M_{\text{sat}}} \right)^\alpha \exp \left( -\frac{M_{\text{cut}}}{M} \right) N_{\text{cen}}(M). \quad (7)$$

The numbers of central and satellite galaxies in the population are drawn from Bernoulli and Poisson distributions, respectively. This HOD model has been applied widely in the study of galaxy clustering as first proposed in Zheng et al. (2005). Multiplying the central galaxy occupation function into the satellite occupation function guarantees that the satellite occupation terminates at a mass higher than the central occupation cutoff; i.e., halos cannot host satellites with no central galaxy. In this HOD model,  $M_{\text{min}}$ ,  $\sigma_{\log M}$ ,  $\alpha$ ,  $M_{\text{sat}}$ , and  $M_{\text{cut}}$  are the free parameters to be fit by observations, which include both  $w_p(r_p)$  and the observed number density of galaxies. Briefly,  $M_{\text{min}}$  is the mass at which half the halos have a central galaxy,  $\sigma_{\log M}$  physically relates to the scatter of halo mass at fixed galaxy luminosity,  $\alpha$  is the power-law index for the mass dependence of the number of satellites,  $M_{\text{sat}}$  is a typical mass for halos to host one satellite, and  $M_{\text{cut}}$  allows for the cutoff in the satellite occupation function to vary with halo mass.

In the simulation-based emulator, we also introduce three additional parameters related to the halo occupation.

1.  $\eta_{\text{con}}$ : the ratio between the concentration parameters of the satellites and dark matter halos  $\eta_{\text{con}} = c_{\text{sat}}/c_{\text{halo}}$ , where the dark matter halos are assumed to have a

Navarro–Frenk–White (NFW) profile (Navarro et al. 1996).

2.  $\eta_{\text{vs}}$ : the velocity bias parameter for the satellite galaxies,  $\sigma_{\text{sat}} = \eta_{\text{vs}} \sigma_{\text{halo}}$ , which rescales the velocity of the satellite galaxies relative to the host halos;  $\sigma_{\text{sat}}$  and  $\sigma_{\text{halo}}$  are the velocity dispersion of satellite galaxies and dark matter halos, respectively.
3.  $\eta_{\text{vc}}$ : the same as  $\eta_{\text{vs}}$  but for central galaxies.

The ranges of the HOD parameters used in this paper are summarized in Table 1. The application of this model within the range of the parameters can give consistent clustering measurements with various observational data, such as Zheng et al. (2007), White et al. (2011), Zehavi et al. (2011), Parejko et al. (2013), Reid et al. (2014), Zhai et al. (2017), and references therein. The HOD designs for the emulator are realized in the Latin hypercube design through the maximin distance design as detailed in the Appendix of Heitmann et al. (2009). Specifically, we initialize a random distribution of a number of points in a hypercube within the range (0, 1) and switch the coordinate in one direction of a randomly chosen pair of points. We iterate this process until the summation of the distance between all of the possible pairs maximizes under certain convergence conditions. The resulting hypercube is linearly mapped into the HOD parameter space that forms the basis for the training of the emulator.

We build a model of galaxy clustering for a sample at  $z = 0.57$  and a space density of  $4.2 \times 10^{-4} (h^{-1} \text{Mpc})^{-3}$ . These choices roughly approximate the BOSS CMASS galaxy sample (Reid et al. 2014). The next paper of this series will apply the model constructed in this work to the clustering measurements of CMASS. The techniques developed here are easily applicable to LRGs in current observations such as eBOSS and near-future observations from the DESI survey. The halo occupation of LRGs has been well studied for a decade; thus, we are on firm theoretical ground to build our clustering emulator. In this work, we fix the number density of the sample and calculate  $M_{\text{min}}$  once all other HOD parameters are known as

$$\bar{n} = \int \frac{dn}{dM} N(M), \quad (8)$$

where  $dn/dM$  is the halo mass function taken from Tinker et al. (2008a) for the analytic model. In future work, we will use the mass function emulator derived from these simulations. We note that in tests, the use of Tinker et al. (2008a) does not bias the results of this paper.

### 2.4. GP Parameter Space

A GP is a collection of random variables, any finite number of which have a joint Gaussian distribution. For a more detailed discussion of GPs and their features, see Rasmussen & Williams (2006). In the calculation throughout this work, we employ the Python code `george`<sup>9</sup> developed by Ambikasaran et al. (2015).

A GP can be written as

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \quad (9)$$

where  $m(\mathbf{x})$  and  $k(\mathbf{x}, \mathbf{x}')$  are the mean function and covariance function, respectively. For the sake of simplicity and with no

<sup>9</sup> <http://george.readthedocs.io/en/latest/>

loss of generality, we take the mean function to be zero in the following. Given the input training data  $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$  at coordinates  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$  with Gaussian noise  $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$ , we can write the joint distribution of the observation and the function values  $\mathbf{f}_*$  at the test locations  $\mathbf{x}_*$  as

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(\mathbf{x}, \mathbf{x}) + \sigma_n^2 I & K(\mathbf{x}, \mathbf{x}_*) \\ K(\mathbf{x}_*, \mathbf{x}) & K(\mathbf{x}_*, \mathbf{x}_*) \end{bmatrix}\right),$$

where  $K(\mathbf{x}, \mathbf{x}_*)$  denotes the covariance matrix of all of the pairs of training and test points, and the other entries have similar meanings. The conditional distribution of a predicted function value  $\mathbf{y}_*$  can then be calculated as

$$\mathbf{y}_* | \mathbf{x}, \mathbf{y}, \mathbf{x}_* \sim \mathcal{N}(\bar{\mathbf{y}}_*, \text{cov}(\mathbf{y}_*)), \quad (10)$$

where

$$\begin{aligned} \bar{\mathbf{y}}_* &= K(\mathbf{x}_*, \mathbf{x}) [K(\mathbf{x}, \mathbf{x}) + \sigma_n^2 I]^{-1} \mathbf{y}, \\ \text{cov}(\mathbf{y}_*) &= K(\mathbf{x}_*, \mathbf{x}_*) - K(\mathbf{x}_*, \mathbf{x}) \\ &\quad \times [K(\mathbf{x}, \mathbf{x}) + \sigma_n^2 I]^{-1} K(\mathbf{x}, \mathbf{x}_*). \end{aligned} \quad (11)$$

The remaining problem now is to model the covariance matrix for the training set and test points. This is implemented by choosing a kernel function to populate the elements in the covariance matrix. This is the crucial ingredient in a GP predictor, as it encodes our assumptions about the function we wish to learn (Rasmussen & Williams 2006). Due to the lack of knowledge about the correlation in the coordinates, the choice of the kernel function is not restrictive. The basic assumption is that points that are close in parameter space are more strongly correlated than points that are further separated, independent of their absolute location in parameter space. Explicitly, we assume the kernel function in our GP modeling to be a radial basis function that just depends on  $r = |\mathbf{x} - \mathbf{x}'|$ . Some examples of commonly used kernel functions in the literature include the squared exponential covariance function,

$$k_{\text{exp}}(r) = \exp\left(-\frac{r^2}{2l^2}\right), \quad (12)$$

where the hyperparameter  $l$  defines the characteristic length scale. Another example is the Matérn class,

$$k_{\text{Matérn}}(r) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}r}{l}\right)^\nu K_\nu\left(\frac{\sqrt{2\nu}r}{l}\right), \quad (13)$$

where  $\nu$  and  $l$  are the parameters, and  $K_\nu(r)$  is a modified Bessel function. A special case in machine learning is  $\nu = 3/2$ , which results in

$$k_{3/2} = \left(1 + \frac{\sqrt{3}r}{l}\right) \exp\left(-\frac{\sqrt{3}r}{l}\right). \quad (14)$$

In our calculation, we find that a combination of  $k_{\text{exp}}(r) + k_{3/2}(r)$  is already flexible enough to model the correlation in the parameter space; therefore, more complicated kernel functions are not used here. The hyperparameter  $l$  has a separate length scale for each dimension in the cosmology and HOD parameter space, corresponding to each component of the argument  $r$ . This kernel function can have extra scalar hyperparameters to scale individual terms or the overall

function, but it is not expected to have a significant impact. The next step is the training of the GP, which involves the selection of the hyperparameters of the kernel function. This is through the maximization of the log-likelihood of the training data:

$$\ln \mathcal{L} = -\frac{1}{2} \mathbf{y}^T (K + \sigma_n^2 I)^{-1} \mathbf{y} - \frac{1}{2} \log |K + \sigma_n^2 I| - \frac{n}{2} \log 2\pi. \quad (15)$$

After this process is completed and the hyperparameters are known, we can substitute the values into Equation (10) to make predictions for the test points.

### 2.5. Estimating the Error on the Training Sample

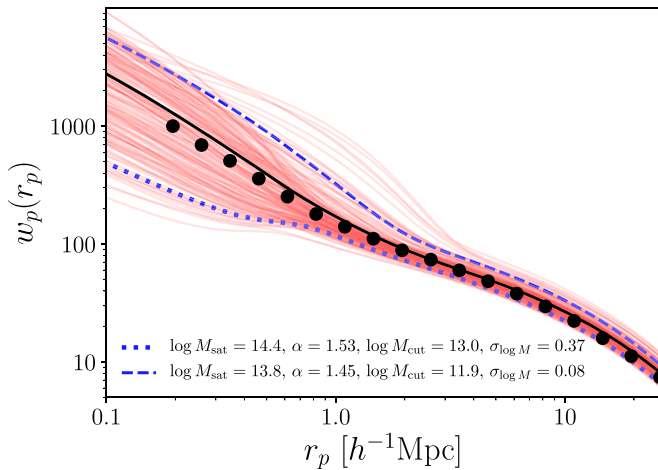
In order to investigate the performance of the GP in building the emulator, we first apply the above method to calculate  $w_p$  with the analytical method (Tinker et al. 2005, 2012). We also need to estimate the appropriate error that the emulator will take as input from the training sample. To estimate the error, we use the suite of test simulations. Each of the seven test cosmologies has five realizations (i.e., different initial conditions); all other simulation properties, including volume and mass resolution, are the same as the training sample. There are two sources of error in the numerical simulations: (1) sample variance in the cosmic structure and (2) shot noise from the finite number of galaxies per halo. To isolate the sample variance, we take each simulation and populate the halos 10 times with the same HOD but different random seeds. We then take the average value of the clustering over these 10 populations. For each cosmology, we obtain a mean value of the clustering and the five deviations from this mean. For the total of seven cosmologies, we have 35 ‘‘differences from the mean,’’ and the sample variance is taken to be the variance of these 35 numbers.

We estimate the shot noise from the variance of the 10 populations of the HOD for a single simulation. Thus, we have 35 estimates of the shot noise, from which we take the average. The total error is the quadrature sum of the shot noise and the sample variance, shown in Figure 2. Sample variance dominates at scales larger than  $\sim 1 h^{-1}$  Mpc. When expressed as a fractional error, we find that this error has little dependence on either the cosmology or the HOD parameters. Thus, when estimating the error from the training sample simulations, we apply the result obtained above as a fractional error to all simulations.

### 2.6. Implementing the Emulator Using the Analytic Clustering Model

As with the design of the cosmologies, we apply the Latin hypercube method and choose  $N_{\text{HOD}}$  designs to sample the HOD parameter space<sup>10</sup> Figure 1 shows a randomly chosen subsample of  $w_p$  calculated with the analytical method. To reduce memory consumption and CPU time, we select a

<sup>10</sup> For the analytic model, we adopt  $N_{\text{HOD}} = 400$  and therefore a sampling scheme with overlap; for a simulation-based emulator, we choose  $N_{\text{HOD}} = 2000$ , which results in a nonoverlap sampling scheme when the number of subsamples of HOD models is 50 (randomly chosen for each cosmology). In this case, for the analytic-based emulator, there is some subset of sampling points with identical HOD parameters but distinct cosmological parameters, while for the simulation-based emulator, each cosmology has a completely distinct HOD subset.



**Figure 1.** Randomly chosen subsample of  $w_p$  calculated with the analytic method for the input training set (red) and the mean (black). The dotted and dashed blue curves show two HOD models lower and higher than the mean  $w_p$ , respectively, with the same cosmology. For comparison, the measurements from BOSS DR11 (Reid et al. 2014) are shown as dots with error bars ignored for visualization purposes.

subsample of HOD models for each cosmology. This provides full coverage of the HOD parameter space without requiring  $400 \times 40$  training points. We construct independent emulators for each bin in  $r_p$  or  $s$  such that the emulator for a specific  $r_p$  or  $s$  bin has its own optimized hyperparameters. Although this ignores a correlation between  $r_p$  bins, we find that this approach is optimal considering the balance between speed and accuracy. Taking into account the correlation between different  $r_p$  bins can increase the training set and memory consumption significantly, which can affect the optimization of the GP parameters.

We first build the emulator under the ideal conditions, where the training data are the original  $w_p$  calculated with the analytical method and the error  $\sigma = 0$ . In order to evaluate the emulator performance, we first randomly generate 200 cosmology sets and 200 HOD sets within the input parameter space as Table 1, then calculate the  $w_p$  with the same analytical method at the same scales. We compare the “truth” with the prediction from the emulator and calculate the fractional error; the result is shown as the purple line in Figure 2. The error is estimated as the 68% and rms error (the latter is not shown in the figure for clarity) for all of the test points, and the results of these two estimates are consistent, implying no catastrophic outliers. This result shows that under the ideal conditions, the emulator can give accurate predictions of the correlation function, and that the GP modeling is robust. The error is relatively constant at all scales, with an average value of 0.3%.

Next, we build the emulator with the same input training coordinates, but we add a Gaussian random deviation with a width of  $1\sigma$  to each training data point, where  $\sigma$  is the error estimated from the  $N$ -body simulations as detailed in the previous section. The comparison with the test points is shown in Figure 2. This demonstrates that the accuracy is degraded as we add noise to the input training data, but the decrease is not significant relative to the  $1\sigma$  error level itself. Then we add Gaussian noise with a width of  $1\sigma$  to the test points as well (green line in Figure 2). The results show that when both the input training data and the test data have noise, the overall error of the emulator is dominated by the error from the test points. This implies that the GP-based emulator can provide estimates

of clustering with higher accuracy than the input training sample. We note that this test is constructed over somewhat idealized circumstances; the true error distribution may not be exactly Gaussian, and it assumes that different  $r_p$  bins are uncorrelated. But even with these caveats, Figure 2 shows the potential of the emulator to make high-accuracy predictions of clustering.

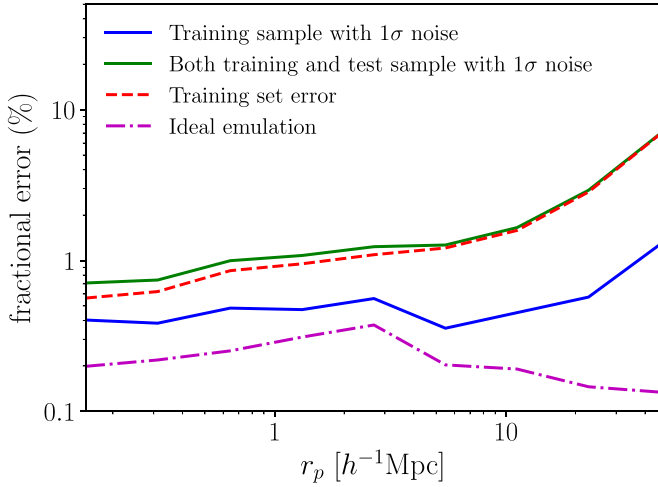
Using the analytic model for  $w_p$ , we can easily explore how the accuracy of the emulator varies with position within the cosmological parameter space. Figure 3 shows the 68% error on  $w_p$  as a contour plot on 2D projections of the cosmological parameter space. The emulator in this test has  $1\sigma$  uncertainty on the training sample, and the analytic model predictions are taken as truth for the test points (i.e., the blue curve in Figure 2). The error in this case purely comes from the understanding of the GP from the noisy training set. The result shows that the error of the emulator is more sensitive to the value of  $\sigma_8$  than other parameters, but the variations of the accuracy are small relative to the  $1\sigma$  error level on the training data. This implies that being near the edge in the parameter space does not degrade the emulator accuracy.

### 3. Building the Emulator with $N$ -body Simulations

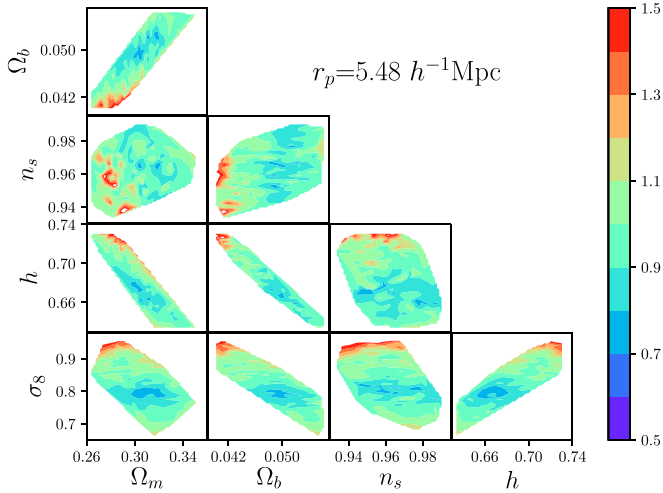
The previous section demonstrates that the GP is a powerful tool to emulate the galaxy correlation function. We now apply this methodology to the estimate of the correlation function directly from  $N$ -body simulations. The details of the simulations are presented in DeRose et al. (2018). Briefly, the simulation products are  $(1.05 h^{-1} \text{Gpc})^3$  boxes with  $1400^3$  particles, resulting in a mass resolution of  $3.51 \times 10^{10} \left(\frac{\Omega_m}{0.3}\right) h^{-1} M_\odot$ . Compared with the cosmology designs used for the analytic emulator, the  $N$ -body simulations have three extra parameters:  $w$ , the equation of state of dark energy;  $N_{\text{eff}}$ , the number of relativistic species; and  $\gamma_f = f/f_{\text{wCDM}}$ , the factor used to scale all halo velocities in the simulation. A fractional change in this parameter is proportional to the change in the linear growth rate at linear and nonlinear scales (Reid et al. 2014). In addition, we also add three HOD parameters,  $\eta_{\text{con}}$ ,  $\eta_{\text{vc}}$ , and  $\eta_{\text{vs}}$ , to incorporate the spatial and velocity bias of galaxies within halos.

Our emulator of the galaxy correlation function using  $N$ -body simulations has 15 parameters in the input parameter set. Because of this increase in dimensionality, we find that a design with 400 HODs is not sufficient to fully sample the space. We increase the number of HOD designs,  $N_{\text{HOD}}$ , to 2000 to obtain the training set. Note that the test cosmologies have five boxes each, so we can get a more accurate estimate of the correlation functions for these test points.

We build the emulator for the projected correlation function  $w_p$ , redshift-space monopole  $\xi_0$ , and quadrupole  $\xi_2$  estimated from  $N$ -body simulations using the same strategy as the previous section for GP modeling, including the kernel functions and scale binning. We estimate the training error for these statistics as detailed in the previous section and apply it as a constant fractional error. The accuracy of the emulator is obtained by comparing with the measurements from test boxes that contain seven cosmologies, randomly choosing 100 HOD sets in the same parameter space as the input training sample. This results in a test sample of 700 models. The three columns of Figure 4 present the performance of the emulator for  $w_p$ ,  $\xi_0$ , and  $\xi_2$ .



**Figure 2.** Performance of the emulator with the analytic method. The red dotted curve is the error level of the input training set ( $1\sigma$ ). The purple line in the bottom represents the emulator performance when the input training sample and test sample have no error. When the training sample is perturbed with  $1\sigma$  noise, we retrain the GP and generate emulator predictions to compare with the test sample, which has zero and  $1\sigma$  noise added. The performance is shown by the blue and green lines, respectively. The 68% and rms error (not shown for clarity) are nearly the same, implying that there are no catastrophic outliers. The blue curve represents the error purely due to emulation. Assuming that the errors from emulation and on the test points are independent, the final error estimate of the emulator (green) is the two added in quadrature.



**Figure 3.** Error of the emulator of  $w_p$  with the analytic method shown in cosmological parameter space, obtained from 200 cosmologies randomly sampled within the training parameter space. The error of each cosmology is estimated from the 68% error of 200 test HODs. The figure shows the error projected onto 2D parameter planes at scale:  $r_p = 5.48 h^{-1} \text{Mpc}$ . The results for other scales have similar patterns. The errors shown here are normalized by the mean of these errors at this scale (the blue solid line in Figure 2), and the fluctuations at various positions in the parameter space are mostly around 20%–30%. Compared with the input training error, which is a few times larger than the mean, the error from the emulator is fairly constant across the parameter space.

The top left panel of Figure 4 shows a few examples of  $w_p$  chosen to spread over the  $w_p$  amplitude calculated from the emulator and directly from  $N$ -body simulations, respectively. It shows that the emulator can generate high-accuracy predictions. The bottom left panel shows the performance of  $w_p$  from the emulator. The solid red line is the error of the input training set, which has a single population of dark matter halos for each position in the parameter space. The dashed red line shows the

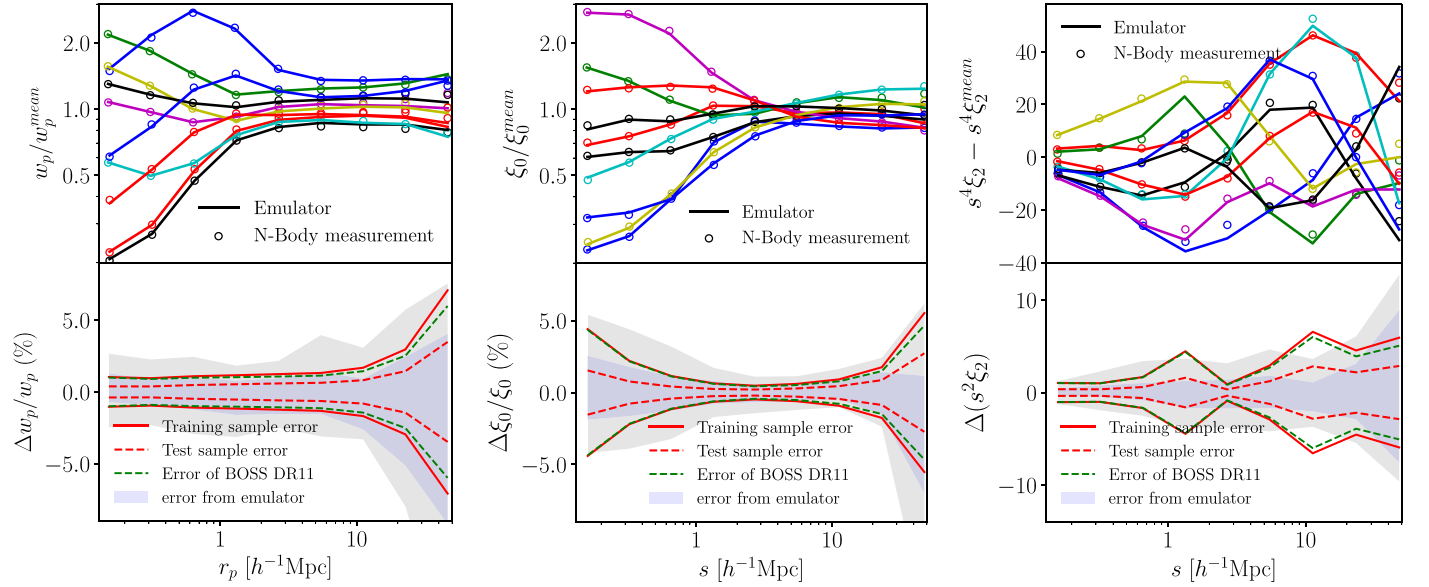
error for the test points, which is smaller than the training error, since the test boxes have larger volume ( $5 (1.05 h^{-1} \text{Gpc})^3$  boxes for each test cosmology) with multiple populations to suppress the shot noise and sample variance. The emulator performance is represented by the distribution of the shaded area ( $1\sigma$  and  $2\sigma$ , respectively). The overall error of the emulator is 0.9% in the  $1$ – $10 h^{-1} \text{Mpc}$  range. Note that the emulator prediction is directly compared with the measurements from the test cosmology and HOD model (the test sample). Thus, the accuracy (or residual) shown as the shaded area has two components: the intrinsic emulator uncertainty (the dot-dashed purple and solid blue lines in Figure 2) and the error of the test sample, which is nonzero due to the finite population and simulation volume.

The middle column of Figure 4 shows a similar result for the monopole  $\xi_0(s)$ . A similar conclusion can be drawn about this emulator. Note that the errors for the training sample and the emulator are larger than  $w_p$  at small scales due to the shot noise. The right column of Figure 4 represents the result for quadrupole  $\xi_2(s)$ . Because  $\xi_2(s)$  approaches zero and goes negative at some scale, the fractional error is not a useful statistic for the emulator performance; here the results from the emulator and the test boxes are shown as absolute errors. The overall  $1\sigma$  error of the emulator is smaller than 0.7 of the training error at the scales of interest.

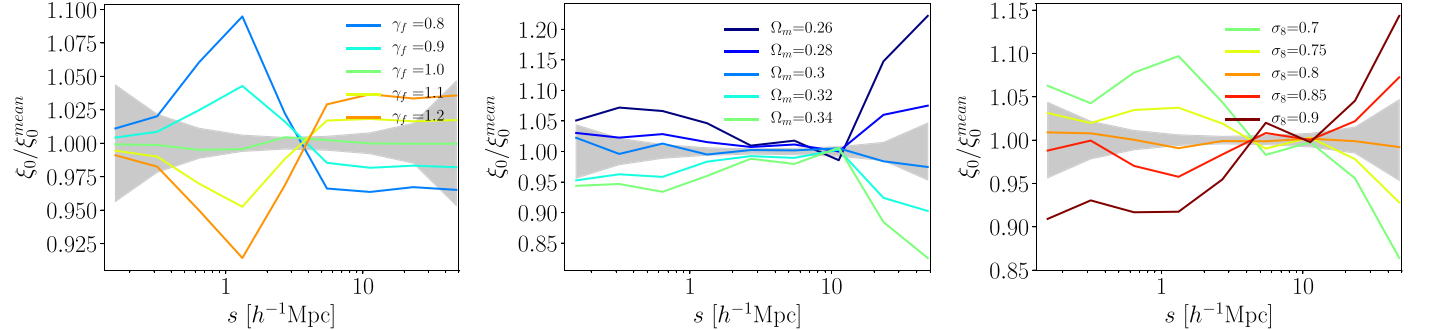
The error as estimated from the test points is smaller than the error on the training sample, implying that the emulator is performing better than the errors on its inputs. This demonstrates that as long as the error of the input training set is smaller than 1%, the emulator can generate predictions better than 1%. In the following calculation, we define the pure error from the emulator as the emulator uncertainty, which is assumed to be independent from the error of the test points. The addition of these error budgets in quadrature gives the total error of the emulation, represented by the envelope of the shaded area in the bottom panels of Figure 4. Compared with the  $w_p$  emulator using the analytic method in Section 2.6, the simulation-based emulator does not perform much better than the training error. This is partly due to the extended parameter space and better modeling of the noise in the analytic-based emulator, which is designed to be Gaussian-distributed. We note that there exists a correlation between different statistics and  $r_p$  or  $s$  bins. Properly taking this correlation into account is expected to improve the emulator performance. However, this also means increasing the training sample significantly for individual emulators and increasing memory consumption. This implies that other interpolating schemes beyond GP are worth investigating in future work.

#### 4. Recovery of Cosmological Parameters and Constraint from the Emulator

In linear theory, RSDs carry information about coherent flows of matter into overdense regions. This boosts the amplitudes of the RSD multipoles in a manner that is only dependent on the degenerate parameter combination  $f\sigma_8$ , where  $f = \gamma f_{w\text{CDM}}$ , and  $f_{w\text{CDM}}$  is the linear growth rate in the cold dark matter cosmology. Scaling this factor by the halo velocity parameter  $\gamma_f$  gives a direct estimate of the linear growth of structure (Reid et al. 2014). At translinear and nonlinear scales, the impact of peculiar velocities is more than simply a change in amplitude. Nonlinearities are most obviously present in the so-called finger-of-god effect (e.g., Jackson 1972), seen as an



**Figure 4.** Performance of the emulator for  $w_p$  (left),  $\xi_0$  (middle), and  $\xi_2$  (right) measured from  $N$ -body simulations. Top: the 10 randomly chosen models for  $w_p$ ,  $\xi_0$ , and  $\xi_2$  measured from simulations and the prediction from the emulator. They are shown as the ratio with respect to the mean measurement of the training sample for visualization. Bottom: the 68% and 95% error distributions of the emulator performance, represented by the shaded area. The training sample error is the sample variance on an individual simulation used in the training of the emulator, while the test sample error is from the mean of the five simulations, all of which have the same volume as the training sample. Thus, the test sample has an error (dashed red) smaller than the training sample (solid red) due to larger simulation volumes. The emulator performance (or residual) shown by the shaded area is directly obtained by comparing the emulator prediction with the measurements of the test cosmology and HOD model from the simulation boxes. Note that the results for  $\xi_2$  (right panels) are presented by absolute error instead of fractional error, and the top panel shows the performance of  $\xi_2 s^4$  instead of  $\xi_2 s^2$  for plotting purposes. The green dashed lines show the sample variance equivalent to the BOSS DR11 volume.



**Figure 5.** Cosmological parameter dependence of the redshift-space monopole given the real-space clustering amplitude. In the test, we fix all of the cosmological parameters and perform an HOD fit to the given  $w_p$ . From left to right, the cosmological parameters are the same except for  $\gamma_f$  (left),  $\Omega_m$  (middle), and  $\sigma_8$  (right), as labeled. The gray band shows the sample variance equivalent to the BOSS DR11 volume, which is calculated by scaling the error estimate from the test boxes by the volume ratio.

elongation of the contours of the correlation function along the line of sight (see, e.g., Peacock & Smith 2000). The amplitude of this effect is sensitive to the mass scale of the halos themselves, which is sensitive to the abundance of massive halos and the details of the galaxy bias model.

In Figure 5, we present a pedagogical test in which we vary a single cosmological parameter and show the impact on the RSD monopole. Each time a parameter is varied, the real-space correlation function,  $w_p(r_p)$ , is refit to determine new HOD parameters. Only HOD parameters that control the mean occupation function are allowed to vary. In the left panel, we vary the amplitude of the halo velocity field,  $\gamma_f$ , from 0.8 to 1.2. The real-space clustering of each model is nearly identical in each fit. At  $s > 5 h^{-1}$  Mpc, we have the expected linear behavior where increasing  $f$  increases the amplitude of  $\xi_0(s)$ . At smaller scales, however, increasing  $f$  suppresses the monopole. This is due to the increase in random motions of close halo pairs, spreading these pairs out along the line of sight. This

suppression goes away at the smallest scales, where the pairs are mostly within a single halo. The middle panel of Figure 5 shows variations in  $\Omega_m$ . Although increasing the matter density increases the amplitude of the velocity field akin to changing  $f$ , in this panel, we do not see the expected linear behavior at large scales. This is because of the impact  $\Omega_m$  has on the shape of the matter correlation function; increasing  $\Omega_m$  at fixed  $\sigma_8$  increases large-scale power at  $s > 10 h^{-1}$  Mpc for dark matter, which means lower galaxy bias to preserve the real-space clustering, and decreases the redshift-space monopole. In each of these models, the fit to  $w_p$  is consistent at  $r_p < 10 h^{-1}$  Mpc, but at larger scales, the galaxy correlation functions diverge, leading to the divergence seen in  $\xi_0(s)$ . At  $s < 10 h^{-1}$  Mpc, nonlinearities are already dominant over coherent velocity flows, producing a trend of lower  $\Omega_m$  yielding higher  $\xi_0(s)$ . Changing the matter density also changes the mass of dark matter halos; thus, galaxy pairs within halos have higher relative velocities. The right panel shows variations in  $\sigma_8$  at  $z = 0$ . Increasing  $\sigma_8$



increases the large-scale power in redshift space, as expected. But  $\sigma_8$  also influences the abundance of massive halos: larger  $\sigma_8$  results in more cluster-sized objects, which are the source of galaxy pairs within the same halo at  $r \approx 1 h^{-1}$  Mpc. Because the halos are massive, the relative velocities of galaxies within these halos are large, leading to increased suppression of redshift-space clustering at the transition between one-halo and two-halo clustering.

These parameters impact galaxy velocities at large and small scales, as well as the shape of real-space clustering. The combination of real- and redshift-space observables therefore provides enough leverage to break degeneracies and independently constrain each of these three quantities, without the need for a strong CMB prior. In this section, we demonstrate the ability of our clustering emulator to recover the input cosmologies both on the test  $N$ -body simulations and with even higher-resolution simulations that can track substructure within halos.

#### 4.1. Recovery Test on Simulation Boxes

With the above emulator for the galaxy correlation function, we can now fit a given measurement in cosmological and galaxy occupation parameter space. The result can also provide a direct measurement of  $f\sigma_8$  at  $z \sim 0.57$ . Moreover, the combination of real- and redshift-space clustering affords enough constraining power to break this degeneracy and constrain  $f$  directly. As a first test, we randomly choose a parameter set (including both cosmological and HOD parameters) from the test boxes and assume that the correlation function measurements are “observational data.” Then we use our emulator to generate predictions and construct a likelihood function. To estimate the likelihood of a given model, we need to estimate the covariance in the data. This includes not just the correlation between  $r$  bins of a given statistic but also the correlation between the statistics themselves. In order to estimate this correlation between  $w_p$ ,  $\xi_0$ , and  $\xi_2$  for the likelihood function, we use the Minerva simulations, a set of 100  $N$ -body simulations (Grieb et al. 2016). The parameters of the HOD model are chosen to be “CMASS-like” at  $z \sim 0.57$ , but in our test, we find that the errors depend little on the HOD model. We calculate the galaxy correlation function from these galaxy catalogs and estimate the correlation matrix (by normalizing the covariance matrix for these galaxy correlation functions).<sup>11</sup> As a conservative test, we first choose the error of the correlation function to be an addition in quadrature of the error in the test simulations (which corresponds to a simulation volume of about  $5 (h^{-1} \text{Gpc})^3$ ) and the emulator uncertainty as defined above. The resulting error thus corresponds to a simulation volume smaller than  $5 (h^{-1} \text{Gpc})^3$ . We use this error estimate as the diagonal elements combined with the

<sup>11</sup> We have done a singular value decomposition (SVD) analysis as in Gaztañaga & Scoccimarro (2005) and Sinha et al. (2018). We find that 23 of the 27 singular values meet the criteria to contain useful information due to the limited number of mocks that may impact the result. We further perform a denoising analysis for the covariance matrix by setting the eigenvalues of the covariance corresponding to the noisy eigenvectors to zero and use the remaining eigenvalues to get the inverse of the covariance matrix. The resulting cosmological constraint is degraded by a factor of 20%, which is a modest impact on our final result. We note that this method may lose some of the signal in the data but can serve as a conservative result (Sinha et al. 2018).

correlation matrix to populate the covariance matrix in our construction of the likelihood function.<sup>12</sup>

The resulting likelihood function can be written as

$$\ln \mathcal{L} = -\frac{1}{2}(\xi_{\text{emu}} - \xi_{\text{obs}})C^{-1}(\xi_{\text{emu}} - \xi_{\text{obs}}), \quad (16)$$

where  $\xi_{\text{emu}}$  and  $\xi_{\text{obs}}$  are the correlation function from the emulator and observation, respectively, and  $C$  is the covariance matrix as defined above. The exploration is obtained through an MCMC test with the Python package `emcee`<sup>13</sup> (Foreman-Mackey et al. 2013), which is based on an affine-invariant ensemble-sampling algorithm (Goodman & Weare 2010). Figure 6 shows constraints on the key cosmological parameters of interest, as well as key HOD parameters that could show degeneracies with cosmology. This result is obtained with flat priors on the parameters, which are uninformative. As expected, there is a strong degeneracy between  $\gamma_f$  and  $\sigma_8$ . The HOD parameter most degenerate with  $\gamma_f$  is  $M_{\text{sat}}$ ; lowering  $M_{\text{sat}}$  increases the bias of the sample, reducing the amplitude of the two-halo redshift-space clustering. To counterbalance this,  $\gamma_f$  increases the redshift distortions. This is the same reason for the degeneracy between  $\gamma_f$  and  $\sigma_8$ —lower  $\sigma_8$  requires a higher  $M_{\text{sat}}$  and bias to fit  $w_p$ . For reference, the true cosmology is indicated with the black cross.

The above constraints are obtained with no CMB priors other than flat priors that define the parameter space of our simulations. We also apply the constraint from Planck measurement as priors of the cosmological parameter. In particular, we choose the constraint on  $\Omega_m$ ,  $\Omega_b$ ,  $h$ ,  $\sigma_8$ , and  $n_s$  from the TT+lowP+lensing chain (Planck Collaboration et al. 2015) of the  $\Lambda$ CDM model. The constraints are shown as the red contours in Figure 6. Applying this Planck prior significantly strengthens the constraints on both cosmology and HOD parameters. This is primarily due to the strong prior on  $\sigma_8$  introduced by the Planck data.

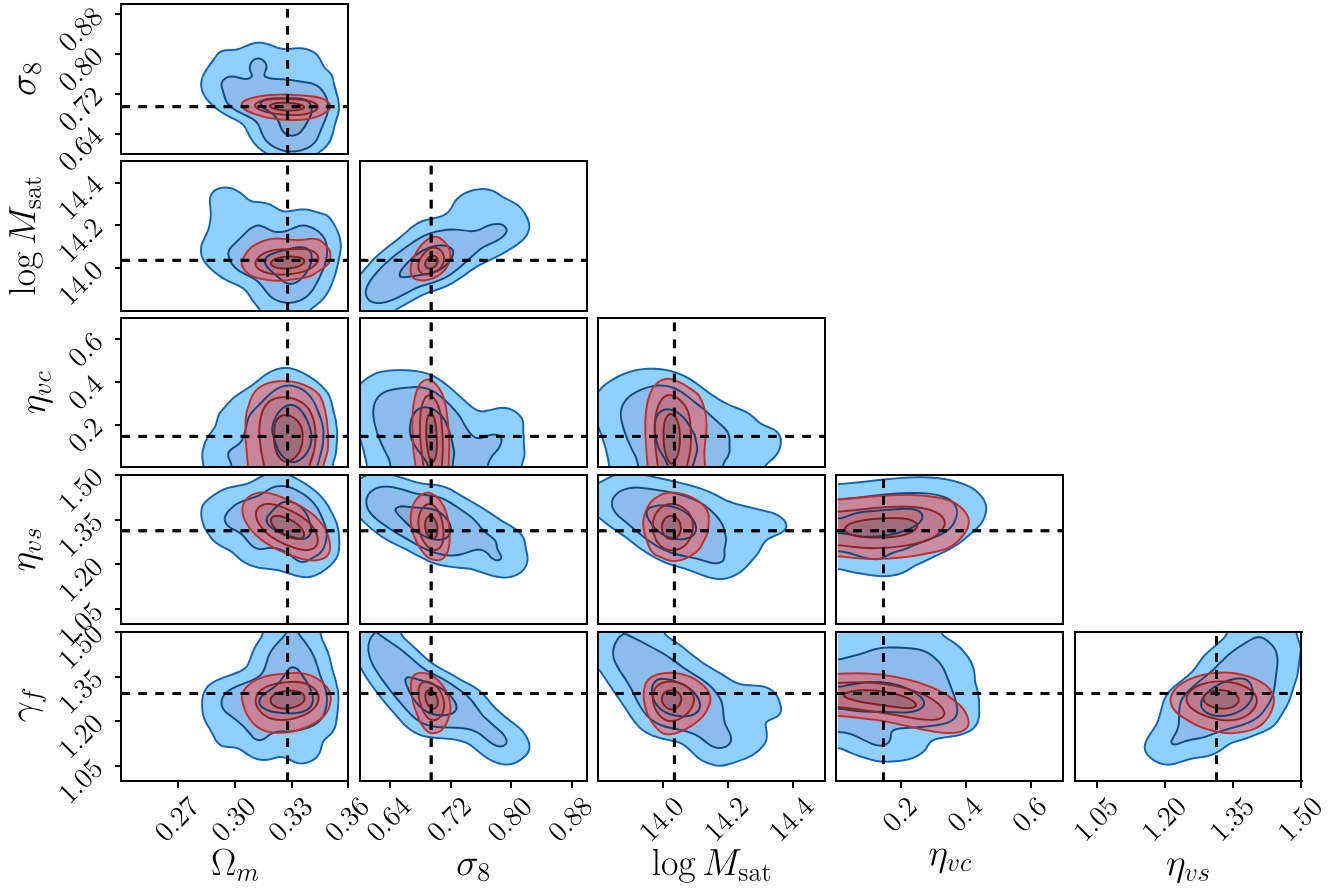
We repeat the above test for 10 randomly chosen test simulation boxes and HODs. The fractional errors on  $f$  and  $f\sigma_8$  are shown in Figure 7. The fractional error of  $f$  is mostly at 5% or higher and can be improved significantly by the Planck data to 2%–3%. However, for the product of  $f\sigma_8$ , the improvement from the Planck data is marginal compared with the correlation function only, which constrains  $f\sigma_8$  to the 2%–3% level.

#### 4.2. Scale Dependence of the Constraint on Structure Growth

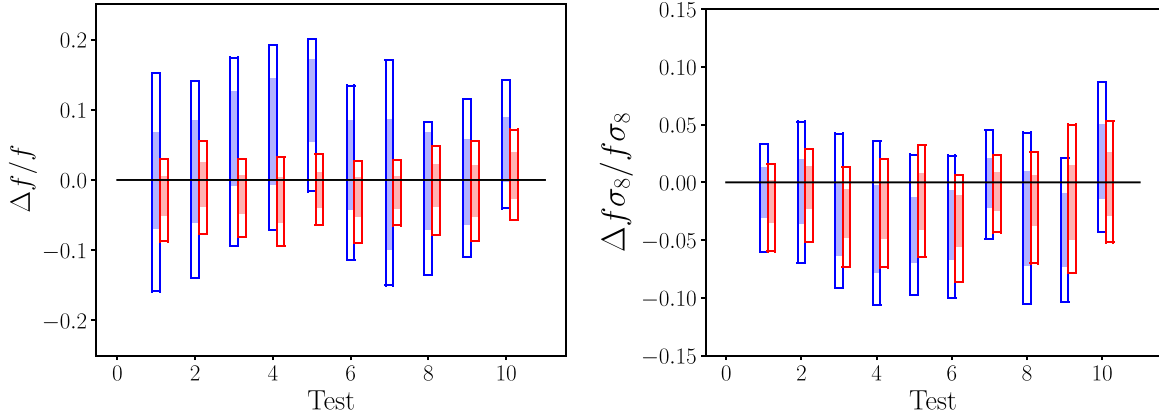
The emulator for small-scale clustering allows exploration of where the constraining power of structure growth comes from. In this test, we randomly choose a test cosmology and constrain the parameters as in Section 4.1, but now varying the minimum scale of the data from  $0.1$  to  $10 h^{-1}$  Mpc. We perform this scale variation while also adjusting the inputs of the emulator in two ways: (1) we compare the results of our standard (noisy) emulator to those where the emulator predictions are considered perfect (noiseless), and (2) we remove the quadrupole from the analysis to determine how much constraining power comes from the monopole only. The result

<sup>12</sup> This implicitly assumes that the emulator prediction has an identical correlation as the “observation.” This is not guaranteed in the application to real or mock data, but the final result should not be affected significantly with a more realistic model.

<sup>13</sup> <http://dfm.io/emcee/current/>



**Figure 6.** Recovery test with the emulator on a randomly chosen test cosmology. The blue contours show the  $1\sigma$  and  $2\sigma$  constraint on a subset of the cosmological and HOD parameters based on the correlation function measurements. The red contours indicate results with a Planck prior. The true (input of the test cosmology) parameters are marked as dashed lines.

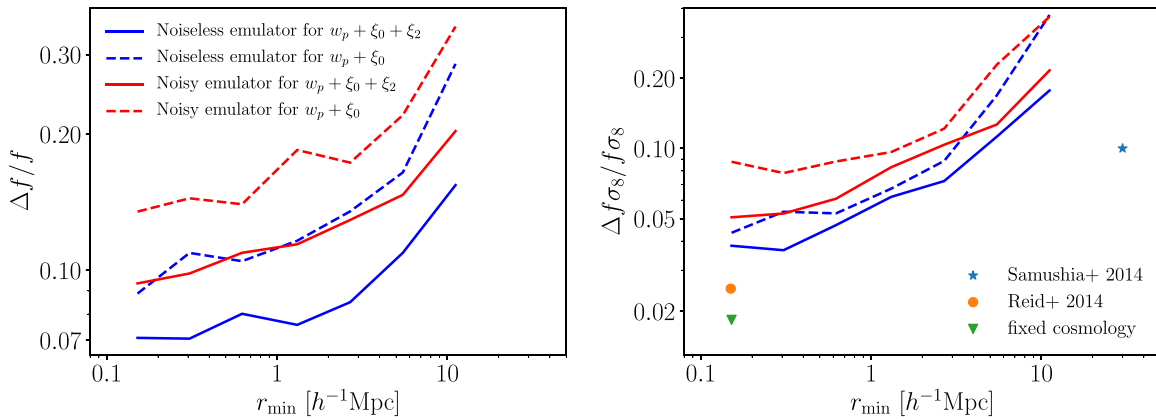


**Figure 7.** Constraints on  $f$  (left) and  $f\sigma_8$  (right) for 10 randomly chosen cosmology+HOD models, shown as fractional errors. The filled and open squares show  $1\sigma$  and  $2\sigma$  constraint, respectively. The blue and red bands show results without and with Planck priors. The average  $1\sigma$  ( $2\sigma$ ) constraint on  $f$  from all of these models is 5.3% (9.9%) without and 2.3% (4.7%) with Planck priors, respectively, while the corresponding results for  $f\sigma_8$  are 3.0% (6.0%) without and 2.2% (4.5%) with Planck priors.

is presented in Figure 8 for  $f$  and  $f\sigma_8$ . For comparison, the results from Reid et al. (2014) and Samushia et al. (2014) are also shown on the panel for  $f\sigma_8$ ; all emulator results have been scaled to the volume of BOSS DR11. It is clear that the cosmological information monotonically increases with decreasing minimum scale. The addition of the quadrupole tightens the constraints by a factor of 1.5 and 1.3 for  $f$  and  $f\sigma_8$ , respectively. At the minimum scale, our noiseless emulator shows that a 3.6% measurement of  $f\sigma_8$  is possible with the DR11 volume. For the noisy emulator, the accuracy of the

measurement is 5% compared with that measured from Reid et al. (2014) for the same survey volume. Reid et al. (2014) employed a single simulation box, which is equivalent to fixing the cosmological model. We perform a similar test with this  $\delta$ -function prior on the shape of the matter power spectrum, and the constraint is tightened by a factor of 3, as shown by the triangle in Figure 8.

The above estimation can be generalized to other current or future galaxy surveys. For a DESI-like LRG survey that covers  $14,000 \text{ deg}^2$  of the sky, the probed volume from  $z = 0.6$  to 1.0



**Figure 8.** Scale dependence of the constraint on  $f$  (left) and  $f\sigma_8$  (right) from galaxy clustering measurements, assuming an SDSS DR11–like galaxy sample. Solid and dashed lines correspond to constraints with and without quadrupole data, respectively. Blue lines assume that our emulator has no error in its prediction, i.e., the covariance matrix in Equation (16) only contains sample variance, while the red lines assume that the emulator has error that is added in quadrature with the sample variance. In the panel for  $f\sigma_8$ , two measurements from Reid et al. (2014) and Samushia et al. (2014) for the BOSS CMASS galaxy sample are marked by a circle and star. The triangle represents the noisy emulator test with fixed cosmological parameters. For the constraint with the noisy emulator (solid red), we find that the error for  $f$  and  $f\sigma_8$  scales roughly with the minimum scale as  $r_{\min}^{0.2}$  and  $r_{\min}^{0.4}$ , respectively. (Note that the “noisy” and “noiseless” emulators have the same prediction of the galaxy correlation function for a given cosmology and HOD model; the difference in the results is whether the emulator uncertainty is added to the observational uncertainty in the likelihood function.)

is about  $12 (h^{-1} \text{Gpc})^3$ , or twice the volume of one test cosmology in our simulation. This leads to a final constraint of 2% for  $f\sigma_8$  and 4.6% for  $f$ , respectively, in this redshift range.

## 5. Discussion and Conclusion

Using the simulations of the AEMULUS Project (DeRose et al. 2018), we have demonstrated the feasibility of constructing an emulator for the real- and redshift-space clustering of galaxies. Using simulations independent of the training sample, our design for the emulator is able to predict the clustering of galaxies to high accuracy over a wide range of both cosmological and galaxy bias parameter space. For the scales at which current surveys yield their most precise measurements,  $1 h^{-1} \text{Mpc} < r < 10 h^{-1} \text{Mpc}$ , our model predicts the galaxy clustering signal to better than 1% and yields predictions that are significantly better than the sample variance of the training sample simulations at smaller and larger scales.

The primary purpose of the galaxy clustering emulator presented here is to constrain cosmological parameters, with emphasis on the growth of dark matter structure, parameterized through  $f$  and the degenerate parameter combination  $f\sigma_8$ . We have shown that constraints on these parameters tighten monotonically as smaller scales are included in the analysis. For a BOSS-like survey, we estimate that we can achieve 5% accuracy on  $f\sigma_8$  and 9% accuracy on  $f$  itself without using CMB priors on any other cosmological parameters. This projection for  $f\sigma_8$  is two times larger than what is achievable through perturbation theory analysis of larger-scale information.

As we prepare this model for application to existing data, there are several additional steps required. A fully robust model requires incorporation of galaxy assembly bias into the halo occupation model, which we will leave to a future work (S. McLaughlin et al. 2018, in preparation). Expanding our parameter space may degrade our constraints; thus, incorporating additional observational measures may prove fruitful. Void statistics have been shown to constrain environmental dependence of halo occupation (Tinker et al. 2006, 2008b). Other statistics may also help with the degeneracies already seen in the parameter constraints in Section 4.1 and Figure 6 (Wibking

et al. 2019). Guo et al. (2015) showed that measurements of the small-scale three-point correlation function can significantly enhance constraints on the velocity bias of central and satellite galaxies, both of which show degeneracy with  $f$ . Among the galaxy bias parameters, the strongest degeneracy with  $f$  is with  $M_{\text{sat}}$ , the mass scale of satellite galaxies. This parameter can be measured directly through galaxy clusters; indeed, the  $M/N$  ratio within clusters itself contains significant cosmological information (Tinker et al. 2012; Reddick et al. 2014).

In addition to reducing the theoretical uncertainties of modeling clustering at nonlinear scales, the simulation-based approach used here is ideal for tackling observational systematics as well. The next version of this emulator will be applied to the existing LRG data sets, including CMASS, LOWZ, and the eBOSS LRG sample at higher redshift (Parejko et al. 2013; Zhai et al. 2017). The dominant observational systematic for these samples is fiber collisions: the constraint that two galaxies closer than  $62''$  cannot be observed at the same time. Nearly all previous attempts to account for this effect involve correcting the data (e.g., White et al. 2011; Zehavi et al. 2011). By using simulations directly, it is possible to forward model the impact of fiber collisions on observational measures of clustering and incorporate any uncertainties in the model itself.

Although we have focused on demonstrating the constraining power of small-scale clustering for a galaxy sample of a given redshift and number density, the ultimate goal of the AEMULUS galaxy clustering emulator is to build a robust tool to allow modeling of galaxies at any number density and redshift. This will significantly increase the parameter space and the dynamic range of clustering signals to be modeled. This may require numerical algorithms beyond the traditional GP (Ng & Deisenroth 2014). Our current results represent the first significant step on the path to that goal, which we expect to be a core technique in the analysis of next-generation galaxy surveys.

The authors thank Michael Blanton, Elisabeth Krause, Boris Leistedt, David Hogg, and David Weinberg for helpful comments and suggestions. This work received support from

the U.S. Department of Energy under contract No. DE-AC02-76SF00515. J.L.T. and R.H.W. acknowledge the support of NSF grant AST-1211889. Y.Y.M. is supported by the Samuel P. Langley PITT PACC Postdoctoral Fellowship. This research made use of computational resources at SLAC National Accelerator Laboratory, and the authors thank the SLAC computational team for support. 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), George (Ambikasaran et al. 2015), Emcee (Foreman-Mackey et al. 2013).

### ORCID iDs

Zhongxu Zhai  <https://orcid.org/0000-0001-7984-5476>  
 Yao-Yuan Mao  <https://orcid.org/0000-0002-1200-0820>  
 Risa H. Wechsler  <https://orcid.org/0000-0003-2229-011X>

### References

- Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, *ApJS*, **182**, 543
- Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W., & O’Neil, M. 2015, *ITPAM*, **38**, 252
- Benson, A. J., Cole, S., Frenk, C. S., Baugh, C. M., & Lacey, C. G. 2000, *MNRAS*, **311**, 793
- Berlind, A. A., & Weinberg, D. H. 2002, *ApJ*, **575**, 587
- Carlson, J., Reid, B., & White, M. 2013, *MNRAS*, **429**, 1674
- Carlson, J., White, M., & Padmanabhan, N. 2009, *PhRvD*, **80**, 043531
- Cole, S., Percival, W. J., Peacock, J. A., et al. 2005, *MNRAS*, **362**, 505
- Colless, M., Dalton, G., Maddox, S., et al. 2001, *MNRAS*, **328**, 1039
- Cooray, A., & Sheth, R. 2002, *PhR*, **372**, 1
- Davis, M., & Peebles, P. J. E. 1983, *ApJ*, **267**, 465
- Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, *AJ*, **151**, 44
- Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, *AJ*, **145**, 10
- de Jong, R. S., Barden, S. C., Bellido-Tirado, O., et al. 2016, *Proc. SPIE*, **9908**, 99081O
- DeRose, J., Wechsler, R. H., Tinker, J. L., et al. 2018, arXiv:1804.05865
- DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, arXiv:1611.00036
- Drinkwater, M. J., Jurek, R. J., Blake, C., et al. 2010, *MNRAS*, **401**, 1429
- Eisenstein, D. J., & Hu, W. 1998, *ApJ*, **496**, 605
- Fisher, K. B., Davis, M., Strauss, M. A., Yahil, A., & Huchra, J. 1994, *MNRAS*, **266**, 50
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, **125**, 306
- Gao, L., Springel, V., & White, S. D. M. 2005, *MNRAS*, **363**, L66
- Gaztañaga, E., & Scoccimarro, R. 2005, *MNRAS*, **361**, 824
- Goodman, J., & Weare, J. 2010, *Communications in Applied Mathematics and Computational Science*, **5**, 65
- Grieb, J. N., Sánchez, A. G., Salazar-Albornoz, S., & Dalla Vecchia, C. 2016, *MNRAS*, **457**, 1577
- Guo, H., Zheng, Z., Jing, Y. P., et al. 2015, *MNRAS*, **449**, L95
- Harker, G., Cole, S., Helly, J., Frenk, C., & Jenkins, A. 2006, *MNRAS*, **367**, 1039
- Heitmann, K., Higdon, D., White, M., et al. 2009, *ApJ*, **705**, 156
- Heitmann, K., Lawrence, E., Kwan, J., Habib, S., & Higdon, D. 2014, *ApJ*, **780**, 111
- Heitmann, K., White, M., Wagner, C., Habib, S., & Higdon, D. 2010, *ApJ*, **715**, 104
- Hunter, J. D. 2007, *CSE*, **9**, 90
- Jackson, J. C. 1972, *MNRAS*, **156**, 1P
- Jones, E., Oliphant, T., Peterson, P., et al. 2001, SciPy: Open Source Scientific Tools for Python, <http://www.scipy.org/>
- Kaiser, N. 1987, *MNRAS*, **227**, 1
- Klypin, A., Yepes, G., Gottlöber, S., Prada, F., & Heß, S. 2016, *MNRAS*, **457**, 4340
- Klypin, A. A., Trujillo-Gomez, S., & Primack, J. 2011, *ApJ*, **740**, 102
- Kwan, J., Heitmann, K., Habib, S., et al. 2015, *ApJ*, **810**, 35
- Lawrence, E., Heitmann, K., White, M., et al. 2010, *ApJ*, **713**, 1322
- McClintock, T., Rozo, E., Becker, M. R., et al. 2019, *ApJ*, **872**, 53
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, *ApJ*, **462**, 563
- Ng, J. W., & Deisenroth, M. P. 2014, arXiv:1412.3078
- Parejko, J. K., Sunayama, T., Padmanabhan, N., et al. 2013, *MNRAS*, **429**, 98
- Peacock, J. A., & Smith, R. E. 2000, *MNRAS*, **318**, 1144
- Peebles, P. J. E. 1980, *The Large-scale Structure of the Universe* (Princeton, NJ: Princeton Univ. Press)
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2015, *A&A*, **594**, A13
- Rasmussen, C., & Williams, C. 2006, *Gaussian Processes for Machine Learning* (Cambridge, MA: MIT Press)
- Reddick, R. M., Tinker, J. L., Wechsler, R. H., & Lu, Y. 2014, *ApJ*, **783**, 118
- Reid, B. A., Seo, H.-J., Leauthaud, A., Tinker, J. L., & White, M. 2014, *MNRAS*, **444**, 476
- Samushia, L., Reid, B. A., White, M., et al. 2014, *MNRAS*, **439**, 3504
- Seljak, U. 2000, *MNRAS*, **318**, 203
- Sheth, R. K., & Tormen, G. 2004, *MNRAS*, **350**, 1385
- Sinha, M., Berlind, A. A., McBride, C. K., et al. 2018, *MNRAS*, **478**, 1042
- Takada, M., Ellis, R. S., Chiba, M., et al. 2014, *PASJ*, **66**, R1
- Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008a, *ApJ*, **688**, 709
- Tinker, J. L., Conroy, C., Norberg, P., et al. 2008b, *ApJ*, **686**, 53
- Tinker, J. L., Sheldon, E. S., Wechsler, R. H., et al. 2012, *ApJ*, **745**, 16
- Tinker, J. L., Weinberg, D. H., & Warren, M. S. 2006, *ApJ*, **647**, 737
- Tinker, J. L., Weinberg, D. H., Zheng, Z., & Zehavi, I. 2005, *ApJ*, **631**, 41
- van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *CSE*, **13**, 22
- Wechsler, R. H., Zentner, A. R., Bullock, J. S., Kravtsov, A. V., & Allgood, B. 2006, *ApJ*, **652**, 71
- White, M., Blanton, M., Bolton, A., et al. 2011, *ApJ*, **728**, 126
- White, M., Hernquist, L., & Springel, V. 2001, *ApJL*, **550**, L129
- Wibking, B. D., Salcedo, A. N., Weinberg, D. H., et al. 2019, *MNRAS*, **484**, 989
- York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, *AJ*, **120**, 1579
- Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2011, *ApJ*, **736**, 59
- Zhai, Z., Tinker, J. L., Hahn, C., et al. 2017, *ApJ*, **848**, 76
- Zheng, Z., Berlind, A. A., Weinberg, D. H., et al. 2005, *ApJ*, **633**, 791
- Zheng, Z., Coil, A. L., & Zehavi, I. 2007, *ApJ*, **667**, 760