The Astrophysical Journal, 961:208 (20pp), 2024 February 1 https://doi.org/10.3847/1538-4357/ad0ce8

The Aemulus Project. VI. Emulation of Beyond-standard Galaxy Clustering Statistics to Improve Cosmological Constraints

Kate Storey-Fisher[1], Jeremy L. Tinker[1], Zhongxu Zhai[2,3,4,5], Joseph DeRose[6], Risa H. Wechsler[7,8,9], and Arka Banerjee[10]

[1] Center for Cosmology and Particle Physics, Department of Physics, New York University, 726 Broadway, New York, NY 10003, USA; k.sf@nyu.edu
[2] Department of Astronomy, School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China
[3] Shanghai Key Laboratory for Particle Physics and Cosmology, Shanghai 200240, People's Republic of China
[4] Waterloo Center for Astrophysics, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[5] Department of Physics and Astronomy, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[6] Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
[7] Kavli Institute for Particle Astrophysics and Cosmology and Department of Physics, Stanford University, Stanford, CA 94305, USA
[8] SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA
[9] Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA
[10] Department of Physics, Indian Institute of Science Education and Research, Homi Bhabha Road, Pashan, Pune 411008, India

Received 2022 October 6; revised 2023 October 11; accepted 2023 November 14; published 2024 January 26

Abstract

There is untapped cosmological information in galaxy redshift surveys in the nonlinear regime. In this work, we use the AEMULUS suite of cosmological N-body simulations to construct Gaussian process emulators of galaxy clustering statistics at small scales (0.1–50 h^{-1} Mpc) in order to constrain cosmological and galaxy bias parameters. In addition to standard statistics—the projected correlation function $w_{\rm p}(r_{\rm p})$, the redshift-space monopole of the correlation function $\xi_0(s)$, and the quadrupole $\xi_2(s)$—we emulate statistics that include information about the local environment, namely the underdensity probability function $P_{\rm U}(s)$ and the density-marked correlation function $M(s)$. This extends the model of AEMULUS III for redshift-space distortions by including new statistics sensitive to galaxy assembly bias. In recovery tests, we find that the beyond-standard statistics significantly increase the constraining power on cosmological parameters of interest: including $P_{\rm U}(s)$ and $M(s)$ improves the precision of our constraints on $\Omega_{\rm m}$ by 27%, σ_8 by 19%, and the growth of structure parameter, $f\sigma_8$, by 12% compared to standard statistics. We additionally find that scales below $\sim$6 h^{-1} Mpc contain as much information as larger scales. The density-sensitive statistics also contribute to constraining halo occupation distribution parameters and a flexible environment-dependent assembly bias model, which is important for extracting the small-scale cosmological information as well as understanding the galaxy–halo connection. This analysis demonstrates the potential of emulating beyond-standard clustering statistics at small scales to constrain the growth of structure as a test of cosmic acceleration.

Unified Astronomy Thesaurus concepts: Large-scale structure of the universe (902); Cosmological parameters (339); Computational methods (1965); Astrostatistics (1882)

1. Introduction

Galaxy redshift surveys contain a wealth of information about the cosmological model. Galaxies trace the underlying matter distribution, and their clustering gives us detailed insight into the growth history of the universe. Recent spectroscopic surveys, including SDSS (York et al. 2000) and its extensions BOSS (Dawson et al. 2013) and eBOSS (Dawson et al. 2015), have provided impressive constraints on cosmology using galaxy clustering. Upcoming surveys such as DESI (Aghamousa et al. 2016), the Subaru Prime Focus Spectrograph (Takada et al. 2014), and eventually Euclid (Laureijs 2011) and the Nancy Grace Roman Space Telescope (Green et al. 2012), will measure tens of millions of spectroscopic redshifts, allowing for unprecedented cosmological measurements.

Most of the current state-of-the-art constraints from these data sets are based on galaxy clustering at large scales. One of the main probes used to measure the growth of structure in spectroscopic analyses is redshift-space distortions (RSDs), anisotropies in clustering induced by galaxy peculiar velocities. For the scales over which the RSD effect is typically analyzed, around $\sim$40–150 h^{-1} Mpc, the evolution of matter is close to linear and can be modeled with linear perturbation theory (e.g., Alam et al. 2017). While this approach has been very successful, current and future surveys will be most precise at much smaller scales, given their requirements on galaxy number density. It is not currently known how much additional information exists at these small scales, but recent work suggests that it is significant and may even exceed the information content at large scales (Zhai et al. 2019). Extracting this information requires accurately modeling the nonlinear dynamics of dark matter down to these scales. Cosmological N-body simulation have been remarkably successful at this (e.g., Klypin et al. 2011); however, they are very expensive to run, and including hydrodynamics is intractable for complete cosmological inference purposes.

In order to use N-body simulations for cosmological analysis, we require a galaxy bias model to populate the dark matter distribution with galaxies (Seljak 2000; Berlind & Weinberg 2002; Cooray & Sheth 2002; Zheng et al. 2005), which probabilistically describes the occupation number of galaxies in dark matter halos as a function of halo mass. The simple HOD model reconstructs galaxy clustering to a reasonable degree of accuracy; however, it has been shown that occupation has a small but non-negligible dependence on secondary halo properties, known as galaxy assembly bias (see, e.g., Wechsler et al. 2006; Croton et al. 2007; Zentner et al. 2014; Wechsler & Tinker 2018). Modeling assembly bias is critical for obtaining the most accurate cosmological constraints, as well as understanding the galaxy–halo connection.

Late-time galaxy clustering analyses have put increasingly strong constraints on the growth of structure parameter $f\sigma_8$. While some of these agree with results from the cosmic microwave background as measured by Planck (eBOSS Collaboration et al. 2021; Zhang et al. 2022b), others are in 1σ–4σ tension (e.g., Macaulay et al. 2013; Sánchez et al. 2014; De Mattia et al. 2021). A series of recent studies focusing on small scales have also found a few-sigma tension (Chapman et al. 2021; Lange et al. 2022; Yuan et al. 2022; Zhai et al. 2023), and these agree with the results of weak-lensing studies (e.g., MacCrann et al. 2015; Leauthaud et al. 2017; Joudaki et al. 2020). Improving the constraining power from clustering analyses is important for determining if the tension still holds; one avenue for doing this is expanding beyond RSD to include other clustering statistics.

Current cosmological analyses focus on a small set of two-point statistics of galaxy clustering that are well-understood theoretically. While these statistics are highly informative, it has been shown that there is significant additional information in other nonstandard observables. For instance, Tinker et al. (2006, 2008) demonstrated that the void probability function and underdensity probability function contribute complementary information to two-point statistics, due to their sensitivity to the environmental dependence of halo occupation. Other work has demonstrated the constraining power in these and other related counts-in-cells statistics (Walsh & Tinker 2019; Wang et al. 2019; Beltz-Mohrmann et al. 2020).

The marked correlation function (Sheth & Tormen 2004) has also been shown to contain information complementary to that in standard statistics. White & Padmanabhan (2009) demonstrated that when using a local density-based mark, the statistic is useful in constraining the cosmological parameter σ_8 by breaking degeneracies in HOD modeling; White (2016) found that it is sensitive to modifications to general relativity. Recently, Szewciw et al. (2022) aimed to optimally constrain the galaxy–halo connection, and confirmed that including the marked correlation function, as well as counts-in-cells statistics and others including the group multiplicity function and group velocity dispersion, significantly improve constraints on halo model parameters at fixed cosmology.

In this work, we combine the use of beyond-standard clustering statistics with the emulation approach. Emulation has recently been explored as a method for making highly accurate predictions for cosmology at nonlinear scales while minimizing requirements on cosmological simulations (Heitmann et al. 2009, 2010; Lawrence et al. 2010). The idea is to first construct a sparse training set of high-resolution N-body simulations that span the allowable parameter space. Then a model can be trained to make fast predictions of the output of the simulations, or summary statistics of the output, given the input parameters. This can finally be used in inference to fully explore the parameter space, essentially interpolating in high dimensions over the regions between input simulations. Machine-learning models are often used for this purpose, due to the need to model such a high-dimensional space and produce quick predictions.

Cosmological emulators typically aim to predict summary statistics of the matter and galaxy distributions. Two-point statistics, namely the power spectrum and its real-space counterpart the correlation function, are the key observables used to constrain cosmological models. There has been significant work emulating the matter power spectrum (Heitmann et al. 2009; Lawrence et al. 2017; Giblin et al. 2019; Ho et al. 2022). Recent work has extended and improved upon this approach, such as the incorporation of dynamical dark energy and massive neutrinos into emulators (Angulo et al. 2021), and the development of fully differentiable power spectrum emulators (Spurio Mancini et al. 2021; DeRose et al. 2022). Other emulators predict the galaxy power spectrum (Kwan et al. 2015; Pellejero-Ibañez et al. 2020; Kokron et al. 2021), and Wibking et al. (2019) recently emulated the galaxy correlation function along with galaxy–galaxy lensing.

Simulation-based emulators have been used to improve precision on cosmological parameter constraints from recent surveys: Miyatake et al. (2021) constrain S_8 from the HSC-Y1 and SDSS data using the DARKEMULATOR (Nishimichi et al. 2019). Neveux et al. (2020) apply a Gaussian process emulator to the BOSS galaxy and eBOSS quasar samples, obtaining constraints similar to those of SDSS using half the amount of data. Euclid Collaboration et al. (2019) constructed the EUCLIDEMULATOR to predict the nonlinear correction of the matter power spectrum in preparation for the upcoming Euclid survey; the improved version (Knabenhans et al. 2021) achieves 1% accuracy or better for $0.01\,h\,\mathrm{Mpc}^{-1} \leqslant k \leqslant 10\,h\,\mathrm{Mpc}^{-1}$.

This work is part of the AEMULUS Project, which uses a suite of high-resolution N-body simulations expressly designed for emulation at small scales to improve cosmological constraints. The previous papers in the project introduce the simulation suite (DeRose et al. 2019) and construct emulators of the halo mass function (McClintock et al. 2019a), the galaxy correlation function (Zhai et al. 2019), and halo bias (McClintock et al. 2019b). The AEMULUS emulator has been used to constrain the growth rate of structure in the BOSS-LOWZ sample (Lange et al. 2022) and the eBOSS LRG sample (Chapman et al. 2021), both obtaining nearly a factor-of-two increase in precision on $f\sigma_8$ compared to standard measurements at linear scales. Most recently, the AEMULUS project constructed two-point function emulators that include models of assembly bias and deviations from general relativity (GR) to provide improved precision on the growth rate of structure parameter from the BOSS survey (Zhai et al. 2023).

In this paper, we extend the work of AEMULUS III (Zhai et al. 2019) to include emulation of two beyond-standard observables: The underdensity probability function $P_{\mathrm{U}}(s)$, defined as the probability that a randomly placed sphere has a galaxy density less than some threshold (e.g., Hoyle & Vogeley 2004), and the marked correlation function $M(s)$, the two-point correlation function with galaxy pairs weighted by their properties (Sheth & Tormen 2004). We extend the HOD model of AEMULUS III to include a model of assembly bias, based on the local density. We also incorporate several more HOD parameters for increased flexibility, as well as a parameter that scales the velocity field to model deviations from GR, following AEMULUS V.

This paper is organized as follows: in Section 2, we describe the N-body simulations and halo occupation distribution model used, and in Section 3, we outline the five clustering statistics we use for inference. We detail our emulation and inference methods in Section 4, and show the results of recovery tests on both AEMULUS mocks and an external mock catalog in Section 5. In Section 6, we discuss the implications of these results and our conclusions.

2. Simulations and Galaxy Bias Model

In this section, we detail the AEMULUS N-body simulations that are used as the basis for our emulation (Section 2.1), and the halo occupation distribution model used to model the galaxy–halo connection and populate the simulations to construct mock galaxy catalogs (Section 2.2).

2.1. The Aemulus Simulations

We use the AEMULUS simulations, a suite of 75 high-resolution N-body simulations (DeRose et al. 2019). They have a box size $L = 1.05\ h^{-1}$ Gpc with 1400^3 dark matter particles, and a mass resolution of $\sim 3.5 \times 10^{10} h^{-1} M_\odot$ (depending on the cosmology). The training set consists of 40 different wCDM cosmologies, selected using a Latin hypercube to span the parameter space. The ranges of the cosmological parameters are shown in Table 3 of AEMULUS V (Zhai et al. 2023). The test set is comprised of seven different cosmologies, with five realizations with different initial conditions for each cosmology, totaling 35 test boxes (with a slightly reduced parameter space compared to the training simulations). We use the redshift $z = 0.55$ snapshot for this work. We use the training set to train our emulator, and the test set to verify its performance as well as to estimate the sample variance.

Our cosmological model consists of seven parameters: the matter energy density $\Omega_{\rm m}$, the baryon energy density Ω_b, the amplitude of matter fluctuations σ_8, the dimensionless Hubble constant h, the spectral index of the primordial power spectrum n_s, the dark energy equation of state parameter w, and the number of relativistic species $N_{\rm eff}$. These simulations are based on GR, so we include a scaling parameter γ_f to capture non-GR effects; it is defined as the amplitude of the halo velocity field relative to the wCDM+GR prediction. The parameters of interest for this work are $\Omega_{\rm m}$, σ_8, and γ_f; we do not expect our approach to be particularly sensitive to the other parameters (Zhai et al. 2019), and these are marginalized over. Most importantly, we are interested in the growth of structure parameter $f\sigma_8$, and we parameterize it to be independent of GR by including the velocity field scaling parameter γ_f (Reid et al. 2014). We henceforth compute and refer to the growth of structure parameter as $\gamma_f f\sigma_8$.

2.2. Halo Occupation Distribution Model

To create mock galaxy catalogs from these simulations, we use the halo occupation distribution to model the galaxy–halo connection. The HOD framework starts from the assumption that the number of galaxies N in a given dark matter halo depends only on the mass of the host halo M, and gives a probability distribution for N given M: $P(N|M)$. We base our HOD model on those of Zheng et al. (2005) and Reddick et al. (2013), which separate the contribution of central and satellite galaxies, $\langle N(M)\rangle = \langle N_{\rm cen}(M)\rangle + \langle N_{\rm sat}(M)\rangle$. The central galaxy occupation function is modeled as a Bernoulli distribution with a mean of

$$\langle N_{\rm cen}(M)\rangle = \frac{f_{\max}}{2}\left[1 + {\rm erf}\left(\frac{\log_{10} M - \log_{10} M_{\min}}{\sigma_{\log M}}\right)\right], \quad (1)$$

where `erf()` is the error function. The number of satellite galaxies is drawn from a Poisson distribution with a mean of

$$\langle N_{\rm sat}(M)\rangle = \left(\frac{M}{M_{\rm sat}}\right)^{\alpha} \exp\left(-\frac{M_{\rm cut}}{M}\right) N_{\rm cen}(M)\,. \quad (2)$$

The parameters are defined as follows: $M_{\min}$ is the mass at which half of the halos host a central galaxy, $\sigma_{\log M}$ controls the scatter of halo mass at fixed galaxy luminosity, α is the power-law index for the mass dependence of the number of satellites, $M_{\rm sat}$ is a typical mass for halos to host one satellite, $M_{\rm cut}$ varies the cutoff mass in the satellite occupation function, and $f_{\max}$ is the central occupation fraction of the highest-mass halos. When the $f_{\max}$ parameter equals unity, in the high halo mass limit, all halos host galaxies; setting $f_{\max} < 1$ adjusts this fraction. This accounts for bright galaxies missed in target selection—for example, due to color and magnitude effects, as is the case in the BOSS-LOWZ sample (Leauthaud et al. 2016). A similar parameterization has been used in other analyses (Hoshino et al. 2015; Guo et al. 2018; Chapman et al. 2021; Zhai et al. 2023).

We fix the number density to $\bar{n} = 2 \times 10^{-4}\ (h^{-1}\ {\rm Mpc})^{-3}$ by computing $M_{\min}$ to satisfy this number density after varying the other HOD parameters. This value is somewhat lower than the peak BOSS number density, but similar to that of a luminous red galaxy (LRG) sample, and it is designed to produce a sample closer to being volume limited; it is the number density used in the AEMULUS V analysis of BOSS-LOWZ+CMASS (Zhai et al. 2023). We note that the amplitude of density fluctuations is degenerate with the galaxy bias to linear order, so fixing the number density risks artificially breaking this degeneracy and biasing the results. However, our inclusion of the $f_{\max}$ parameter effectively allows for flexibility in the galaxy bias, as it sets a ceiling for the central galaxy occupation of the halo field. This has been shown by Chapman et al. (2021; Section 4.2): with a fixed number density emulator, they demonstrate that fixing $f_{\max} = 1$ results in a bias in the recovered halo velocity field rescaling parameter γ_f, while freeing $f_{\max}$ eliminates this bias. Chapman et al. (2021) also perform a test of the Alcock–Paczynski scaling effect (Alcock & Paczyński 1979; Section 3.6), which impacts the number density, and find that this change has a negligible effect on final constraints. Additionally, in our target sample, BOSS CMASS and LOWZ, the number density is well measured: we estimate the variation in number density using the BOSS QPM mocks, and find that the fractional uncertainty is 0.43%. To test that this small uncertainty does not affect our results, we construct mocks with 1% greater and lower number density than that at which the emulator is constructed; we find that the recovered parameters shift by only $\sim 0.25\sigma$, and none more than 1σ. We thus conclude that fixing the number density while having a free $f_{\max}$ in our emulator should allow for unbiased inference of cosmology.

We include three additional parameters in our HOD model related to halo occupation, following Zhai et al. (2019). In addition to the parameter γ_f described in Section 2.1 that rescales all halo velocities, we include velocity bias parameters for galaxies relative to the virial velocity of their DM halo $\sigma_{\rm halo}$.

We define $v_{\rm bc}$ as the velocity bias of central galaxies, which rescales the velocity of centrals $\sigma_{\rm cen}$ relative to that of host halos as $\sigma_{\rm cen} = v_{\rm bc}\,\sigma_{\rm halo}$. The velocity bias of satellite galaxies $v_{\rm bs}$ is defined in the same way as $v_{\rm bc}$. We also include a concentration parameter relating satellite and halo concentrations, where the concentration c is defined as the ratio between the halo outer radius and the scale radius (which depends on the halo density profile). We define the concentration ratio $c_{\rm vir}$ as the ratio between the concentration of satellites and DM halos, $c_{\rm vir} = c_{\rm sat}/c_{\rm halo}$.

We extend this standard HOD model to take into account the dependence on properties other than just the host halo mass; this secondary dependence is known as assembly bias. Here, we use the three-parameter assembly bias model of Walsh & Tinker (2019), which includes a dependence on the local dark matter density around a halo, because we might expect the external environment of halos to play a role in galaxy formation. Specifically, we define δ as the relative density in a sphere of radius 10 h^{-1} Mpc around a halo center. The assembly bias model adjusts the minimum halo mass needed to host a central galaxy, $M_{\rm min}$, to a threshold $M'_{\rm min}$ based on the local density. It is defined as

$$M'_{\rm min} = M_{\rm min}\left[1 + f_{\rm env}\ {\rm erf}\left(\frac{\delta - \delta_{\rm env}}{\sigma_{\rm env}}\right)\right], \tag{3}$$

where $f_{\rm env}$ controls the strength of the environmental dependence, $\delta_{\rm env}$ is the density threshold at which to move around satellites, and $\sigma_{\rm env}$ controls the sharpness of the transition between overdense and underdense regions. A value of $f_{\rm env} > 0$ means that a halo in a higher-density environment requires a higher mass to host a central galaxy, and a halo in a lower-density environment needs a lower mass, effectively moving galaxies from high- to low-density regions. Conversely, $f_{\rm env} < 0$ moves galaxies from low- to high-density regions. Setting $f_{\rm env} = 0$ turns off assembly bias.

After using the HOD to populate the simulation boxes with galaxies, we input redshift-space distortions. We do this by projecting the real-space positions along one of the axes $x_{\rm r}$ into redshift-space positions $x_{\rm s}$:

$$x_{\rm s} = x_{\rm r} + (1 + z)\frac{v}{H(z)}, \tag{4}$$

where z is the redshift of the simulation, v is the velocity of the galaxy along axis $x_{\rm r}$, and $H(z)$ is the Hubble parameter at that redshift for the given cosmology.

The HOD parameter ranges we use are the same as in Aemulus V (Zhai et al. 2023), shown in Table 3 of that work. We populate each of the 40 training boxes with 100 unique HOD models, chosen using the Latin hypercube method (Heitmann et al. 2009) with a total of 4000 samples. We populate the test boxes with another independent set of 100 HOD models, from a separate 100 sample draw from a Latin hypercube (it should be noted that, for the test set, we use the same 100 models to populate each of the 35 boxes, while for the training set every model is different). This results in a training set of 4000 catalogs and a test set of 3500 catalogs for the emulator. (We found that two of the 4000 training mocks resulted in unphysical values of clustering statistics and discarded these from our training set.) For the recovery tests, we use a subset of this test set consisting of 70 catalogs, with 10 unique HOD models per cosmology (complete recovery tests on all 3500 models would be both expensive and repetitive, but we do use the additional models for select tests as well as for sample variance estimation). Our complete model has seven cosmology parameters plus γ_f, seven HOD parameters, and three assembly bias parameters, for a total of 18 free parameters. These are the parameters that will be the inputs to our emulators and that we will later infer through Markov Chain Monte Carlo, based on the measured observables.

3. Observables

The goal of this work is to investigate the information in small-scale clustering using both standard statistics and other, beyond-standard observables that may contain important information. (It should be noted that we use the words “observables” and “statistics” interchangeably in this work.) The standard observables we use are:

1. The projected correlation function, $w_{\rm p}(r_{\rm p})$ (Section 3.1);
2. The monopole of the two-point correlation function, $\xi_0(s)$ (Section 3.2); and
3. The quadrupole of the two-point correlation function, $\xi_2(s)$ (Section 3.2).

The beyond-standard observables we include are:

1. The underdensity probability function, $P_{\rm U}(s)$ (Section 3.3); and
2. The marked correlation function, $M(s)$ (Section 3.4).

We discuss the covariances between these statistics in Section 4.2. The statistics measured in the given bins are shown in Figure 2 (circles in top panel), for the 3500 test set models.

3.1. The Projected Correlation Function, $w_p r_p$

The two-point correlation function is defined as the excess probability above a Poisson random distribution that two galaxies are separated by a given distance r. In practice, we work in redshift space with vector distance s, defining $s = s_2 - s_1$ and $l = (s_1 + s_2)/2$. We measure the two-dimensional correlation function $\xi_{\rm Z}(r_{\rm p}, \pi)$ on a grid, where the subscript Z denotes redshift-space, π is the transverse separation, and $r_{\rm p}$ is the line-of-sight separation, defined as

$$\pi = \frac{\boldsymbol{s}\cdot\boldsymbol{l}}{|\boldsymbol{l}|}, \qquad r_p^2 = \boldsymbol{s}\cdot\boldsymbol{s} - \pi^2\,. \tag{5}$$

Then, the projected correlation function is

$$w_{\rm p}(r_{\rm p}) = 2\int_0^\infty d\pi\ \xi_{\rm Z}(r_{\rm p}, \pi)\,. \tag{6}$$

In practice, we cut off the integral at a scale of $\pi_{\rm max} =$ 40 h^{-1} Mpc. This choice of a somewhat low $\pi_{\rm max}$ leaves $w_{\rm p}(r_{\rm p})$ sensitive to RSDs in the two-halo term. However, this preserves some cosmological information, and in any case, it is consistent in the constructed emulator, so it will not lead to a bias in parameter recovery.

We must use an estimator to measure the correlation function in data. We use the natural estimator (Peebles & Hauser 1974),

$$\xi(r_{\rm p}, \pi) = \frac{DD}{RR} - 1, \tag{7}$$

where DD is the number of data–data pairs in an $(r_{\rm p}, \pi)$ bin, and RR is the number of random–random pairs in a uniform random catalog of the same size as the data, each normalized by the total number of galaxy pairs in the respective catalog pair. Because we are working with periodic simulation boxes in this analysis, we can analytically compute the random–random (RR) term and only have to numerically compute the DD term. We note that the DR term cannot be analytically computed, so to avoid needing a random catalog, we do not use a lower-variance estimator such as the standard Landy & Szalay (1993) estimator. As there is no complex window function to introduce biases or additional noise, and we are doing inference using simulations rather than model comparison, the natural estimator should be sufficient.

We measure $w_p(r_p)$ in nine logarithmically spaced bins between $r_p = 0.1$ and $50\,h^{-1}$ Mpc. We use the software package `corrfunc` (Sinha & Garrison 2019, 2020) to compute this observable.

3.2. The Two-point Correlation Function Multipoles, and $\xi_0(s)$ and $\xi_2(s)$

We also measure the multipoles of the redshift-space correlation, now defining the coordinates $s = |\mathrm{s}|$ and $\mu = r_p/s$:

$$\xi_\ell(s) = \frac{2\ell + 1}{2} \int_{-1}^{1} L_\ell(\mu)\, \xi_Z(s, \mu)\, d\mu, \tag{8}$$

where L_ℓ is the Legendre polynomial of order ℓ (and ℓ indexes the multipole). Most of the information is contained in the few lowest-order multipoles, so for this analysis, we use only the monopole $\xi_0(s)$ and the quadrupole $\xi_2(s)$. We use the Peebles & Hauser (1974) estimator as in the previous section to measure the correlation functions in practice.

For $\xi_0(s)$ and $\xi_2(s)$, we use the same nine bins as we did for $w_p(r_p)$, between $s = 0.1$ and $50\,h^{-1}$ Mpc, and we use 15 μ bins. We use `corrfunc` (Sinha & Garrison 2019, 2020) and `halotools` (Hearin et al. 2017) to compute these statistics.

3.3. The Underdensity Probability Function, $P_U(s)$

The first beyond-standard statistic we use in our analysis is the underdensity probability function, $P_U(s)$ (e.g., Hoyle & Vogeley 2004). $P_U(s)$ is defined as the fraction of randomly placed spheres that are underdense compared to some threshold density. This is a more robust metric to measure than the void probability function, which uses a threshold of zero and is more sensitive to issues such as the angular selection function, shot noise, and fiber collisions. We can write $P_U(s)$ as

$$P_U(s) = \frac{1}{N} \sum_{i}^{N} \mathbb{1}(n_i(s) < n_{\mathrm{thresh}}), \tag{9}$$

where i indexes the N spheres, $n_i(s)$ is the number density of galaxies in sphere i with radius s, 1() is an indicator function that is 1 if its argument is true and 0 otherwise, and n_{thresh} is the threshold number density. We choose $N = 10^6$ and $n_{\mathrm{thresh}} = 0.2\bar{n}$, where $\bar{n}$ is the mean number density of the mock; this is the same value chosen by Hoyle & Vogeley (2004), which is slightly denser than the mean underdensity of large voids in the 2dF Galaxy Redshift Survey (Colless et al. 2003).

The $P_U(s)$ does not vary significantly at small scales ($s \lesssim 5h^{-1}$ Mpc) across different cosmology and HOD models (see the sample variance at small scales in Figure 2), so these scales are not as useful for parameter inference. Thus, we use nine linearly spaced radii between $s = 5$ and $45\,h^{-1}$ Mpc. To compute the statistic, we modify a standard k-d tree code[11] to work on a periodic box.[12]

3.4. The Marked Correlation Function, $M(s)$

The other beyond-standard statistic we investigate is the marked correlation function, $M(s)$ (Sheth & Tormen 2004). $M(s)$ is a generalization of the two-point correlation function with each galaxy weighted by some mark m. It is defined as

$$M(s) = \frac{1}{N_p(s)\bar{m}^2} \sum_{ij} m_i m_j, \tag{10}$$

where the sum is over all pairs with separation $s = s_{ij}$, N_p is the number of galaxy pairs at s, and $\bar{m}$ is the mean of the marks. Following White & Padmanabhan (2009), we choose the marks to be a function of the galaxy number density ρ_i around galaxy i, computed within a sphere of radius $10\,h^{-1}$ Mpc. Specifically, we use a mark of $m_i = [\rho_* + \bar{\rho}/(\rho_* + \rho_i)]^n$, where $\bar{\rho}$ is the mean density, following White (2016) and Satpathy et al. (2019). This mark tends to unity for $\rho \sim \bar{\rho}$, is less than unity for $\rho > \bar{\rho}$, and is greater than unity for $\rho < \bar{\rho}$, serving to upweight underdense regions and downweight overdense regions. The parameters ρ_* and n control the sharpness of the transition. We test a grid of ρ_* and n values and choose the values that balance two criteria. We first select three unique cosmology+HOD models that have a minimal distance between their measured $w_p(r_p)$ values. We then measure $M(s)$ for these catalogs on a grid of varying ρ_* and n values, and see which values maximize the distance between their $M(s)$ values, compared to the variance of the entire test set. The idea is that we want $M(s)$ to discriminate between models that are indistinguishable with just $w_p(r_p)$. We also want to maximize the variance of the $M(s)$ values overall, so that the predictions can be better distinguished. These criteria prefer different directions along the ρ_* and n axes, and we choose the values that optimally balance both of them: $n = 1$ and $\rho_* = 8\,\bar{\rho}$.

We measure $M(s)$ with the same binning we did $w_p(r_p)$, $\xi_0(s)$, and $\xi_2(s)$, from $s = 0.1$ to $50\,h^{-1}$ Mpc. We compute the marks using our modified k-d tree code, and use `corrfunc` (Sinha & Garrison 2019, 2020) to compute the $M(s)$.

4. Methods

To perform our analysis, we first construct a Gaussian process emulator for each observable, as explained in Section 4.1. Our inference will require the covariances between the observables and bins; we describe this computation in Section 4.2. We finally perform the inference using our emulator in combination with Markov Chain Monte Carlo, discussed in Section 4.3.

4.1. Gaussian Process Emulation

We use a Gaussian process to emulate the function relating the input cosmological, HOD, and assembly bias parameters to the observables. A Gaussian process is a collection of random variables for which any finite subsample is Gaussian

[11] https://github.com/jtsiomb/kdtree

[12] https://github.com/kstoreyf/clust

distributed. It can be described as a multivariate normal distribution generalized to infinite dimensions. Here, we follow the notation of Rasmussen & Williams (2006); a full discussion of GPs can be found in that text.

Given a training set with $N_{\rm train}$ inputs, each with $N_{\rm param}$ features $\boldsymbol{x}$ and a scalar output y, we can construct a design matrix X of shape ($N_{\rm param}$, $N_{\rm train}$) and a target vector y of length $N_{\rm train}$. We also have a test set with $N_{\rm test}$ inputs $\boldsymbol{x}_*$ from which we can similarly construct a design matrix X_* and a target vector y_*. We assume that these observations can be described by a function f, such that $y = f(\boldsymbol{x}) + \epsilon$, where ϵ is a noise model given by $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$.

The Gaussian process is a function $f(\boldsymbol{x})$ relating the input parameters to the output targets. We take it to have zero mean without loss of generality, and a covariance of $k(\boldsymbol{x}, \boldsymbol{x}')$, described by a kernel function k. Extending this to our full design matrices for the training set and including the noise model, the covariance on the targets becomes $\mathrm{cov}(\boldsymbol{y}) = K(X, X) + \sigma_n^2 I$. We can define the joint distribution of the training target values $\boldsymbol{y}$ and the function evaluated at the test inputs $\boldsymbol{f}_*$ as

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{f}_* \end{bmatrix} \sim \mathcal{N}\left(\boldsymbol{0}, \begin{bmatrix} K(X, X) + \sigma_n^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix}\right), \tag{11}$$

where $K(X, X_*)$ is the covariance matrix of the training and test set inputs, and the other covariances are defined similarly.

Then we can define the predictive function $\boldsymbol{f}_*$ as

$$\boldsymbol{f}_* | X, \boldsymbol{y}, X_* \sim \mathcal{N}(\bar{\boldsymbol{f}}_*, \mathrm{cov}(\boldsymbol{f})), \tag{12}$$

where the mean $\bar{\boldsymbol{f}}_*$ is defined as

$$\bar{\boldsymbol{f}}_* = K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1} \boldsymbol{y} \tag{13}$$

and the covariance $\mathrm{cov}(\boldsymbol{f}_*)$ as

$$\mathrm{cov}(\boldsymbol{f}) = K(X_*, X_*) \\ - K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1} K(X, X_*). \tag{14}$$

Next, we must choose our kernel function, which describes the expected properties of the function we are trying to learn. We assume the kernel to have only a dependence on the distance between the inputs in parameter space, $r = |\boldsymbol{x} - \boldsymbol{x}'|$ (i.e., a "stationary" kernel). We test common kernels and combinations, and choose the one that performs the best on our test set:

$$k(r) = k_{\exp}(r) k_{\rm const}(r) + k_{\rm M3/2}(r), \tag{15}$$

where $k_{\exp}(r)$ is the exponential squared kernel,

$$k_{\exp}(r) = \exp\left(-\frac{r^2}{2l^2}\right), \tag{16}$$

$k_{\rm const}$ is a constant kernel,

$$k_{\rm const}(r) = c, \tag{17}$$

and $k_{\rm M3/2}$ is a special case of the general Matérn kernel with $\nu = \frac{3}{2}$,

$$k_{\rm M3/2}(r) = \left(1 + \frac{\sqrt{3}\, r}{l}\right) \exp\left(-\frac{\sqrt{3}\, r}{l}\right), \tag{18}$$

where l is a characteristic length scale, and c is a constant.

We train the GP on our set of training catalogs to determine the $2\,N_{\rm param} + 1$ kernel parameters (the length scale l for each input parameter for each of the $k_{\exp}$ and $k_{\rm M3/2}(r)$ kernels, and the constant c for the $k_{\rm const}$ kernel) that result in the maximization of the log marginal likelihood:

$$\log p\left(\boldsymbol{y} | X\right) = -\frac{1}{2} \boldsymbol{y}^\top \left(K + \sigma_n^2 I\right)^{-1} \boldsymbol{y} \\ - \frac{1}{2} \log |K + \sigma_n^2 I| - \frac{n}{2} \log 2\pi. \tag{19}$$

We perform this optimization using the L-BFGS-B solver (Fletcher 1987) through `scipy`. We can then use these optimized parameters to evaluate the kernels in Equation (12), and use it to predict the target value for our test set inputs.

We train a separate GP model for each bin of each observable. To perform the Gaussian process computations, we use the `george` code (Ambikasaran et al. 2016), which is optimized for large data sets.

4.2. Covariance Matrix Construction

To perform inference using our emulator, we require a covariance matrix describing the correlations between the observables, as well as between the bins of a single observable. This covariance includes both the uncertainties introduced by the emulator, contained in $C_{\rm emu}$, and the sample variance of the data on which we are performing parameter recovery, $C_{\rm data}$. We combine these into the total covariance $C_{\mathcal{L}}$ that we will use in our likelihood function (see Section 4.3),

$$C_{\mathcal{L}} = C_{\rm emu} + C_{\rm data}. \tag{20}$$

We define the overall emulator performance covariance $C_{\rm perf}$ as the combination of both the intrinsic emulator prediction error ($C_{\rm emu}$) and the covariance of the data on which the emulator is tested, $C_{\rm test}$, so to obtain $C_{\rm emu}$ we must subtract off $C_{\rm test}$:

$$C_{\rm emu} = C_{\rm perf} - C_{\rm test}. \tag{21}$$

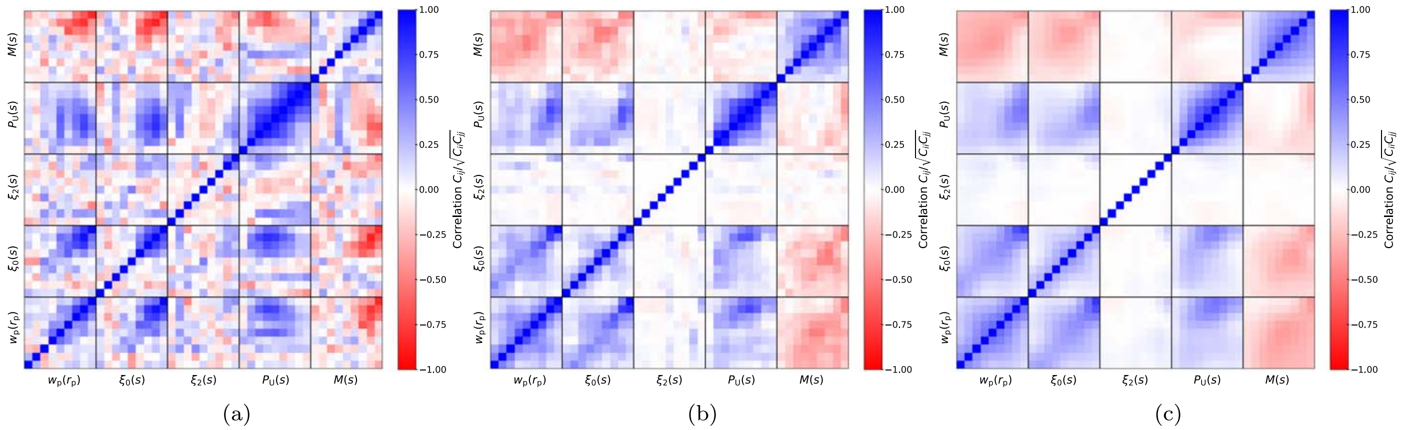
We obtain $C_{\rm perf}$ by computing the covariance of the fractional error between the emulator predictions and the measurements on the data (and then smoothing this matrix to handle noise from our limited number of simulations, as described below). The performance covariance on our test set with $N_{\rm test} = 3500$ observations indexed by n is then

$$C_{\rm perf} = \frac{1}{N_{\rm test} - 1} \sum_{n}^{N_{\rm test}} \boldsymbol{f}_n \cdot \boldsymbol{f}_n^\top, \tag{22}$$

$$\boldsymbol{f}_n = \frac{\boldsymbol{y}_{n,\rm pred} - \boldsymbol{y}_{n,\rm test}}{\boldsymbol{y}_{n,\rm test}}, \tag{23}$$

where $\boldsymbol{y}$ is a vector of the measured observables (which can be a concatenation of multiple observable vectors). It is worth noing that we know the expectation value of these fractional errors should be zero, so we assume $\bar{\boldsymbol{f}}_n = 0$ when computing the covariance. The computed $C_{\rm perf}$ is visualized in Figure 1(b), for all five observables.

We compute $C_{\rm test}$ using the AEMULUS test set, which has $N_{\rm cosmos} = 7$ different cosmologies c, and $N_{\rm box} = 5$ boxes (realizations) b for each cosmology. These are each populated with $H = 100$ HOD models h. We utilize the fact that we have multiple boxes per cosmology to estimate the sample variance. We choose a single HOD model in the middle of the parameter space, and for each cosmology populated with this HOD, we



(a)

(b)

(c)

Figure 1. Correlation matrices for visualizing the covariance matrices used in the analysis, for all five observables. The panels show correlation matrices constructed from (a) the AEMULUS sample covariance C_{aemulus}, (b) the emulator performance covariance C_{perf}, and (c) the performance covariance with a Gaussian smoothing $C_{\text{perf,smooth}}$. The color bar shows the correlation quantity $C_{ij}\big/\sqrt{C_{ii}C_{jj}}$, where C_{ij} are elements of the correlation matrix.

compute the mean value of the observable $\bar{\boldsymbol{y}}_c$ of the N_{box} boxes:

$$\bar{\boldsymbol{y}}_c = \frac{1}{N_{\text{box}}}\sum_b^{N_{\text{box}}} \boldsymbol{y}_{b,c}. \tag{24}$$

We compute the fractional deviation from this mean $\boldsymbol{d}_{b,c}$ for each of box of a given cosmology:

$$\boldsymbol{d}_{b,c} = \frac{\boldsymbol{y}_{b,c} - \bar{\boldsymbol{y}}_c}{\bar{\boldsymbol{y}}_c}. \tag{25}$$

We finally compute the covariance of these deviations from the mean:

$$C_{\text{aemulus}} = \frac{1}{N_{\text{box}}N_{\text{cosmos}} - 1}\sum_b^{N_{\text{box}}}\sum_c^{C} \boldsymbol{d}_{b,c}\cdot \boldsymbol{d}_{b,c}^{\top}. \tag{26}$$

The computed C_{aemulus} is shown in Figure 1(a).

When we compute Equation (22) used in C_{perf}, the observable values $\boldsymbol{y}_{n,\text{test}}$ we use are the mean value of the observable over the N_{box} test boxes for each cosmology. This essentially increases the volume of the test set by a factor of N_{box}, and uncertainty scales in inverse proportion to volume (Klypin & Prada 2018). Thus, in order to combine C_{perf} and C_{test}, we need to scale the latter to match the effective volume of the former:

$$C_{\text{test}} = \frac{1}{N_{\text{box}}}C_{\text{aemulus}}. \tag{27}$$

We can now use C_{test} to construct C_{emu}, and combine it with C_{data} to obtain the total covariance. For our tests, we are performing parameter recovery on the AEMULUS test simulations themselves, so we have $C_{\text{data}} = C_{\text{test}}$, and we get simply $C_{\mathcal{L}} = C_{\text{perf}}$. In future applications to real data, we will need to include both C_{data} and C_{test} in the covariance matrix construction.

We do use the AEMULUS covariance C_{test} as input to the Gaussian process emulator. The GP requires an estimation of the uncertainty on the training set. As the training and test sets are from the same simulation suite, but the test set contains multiple realizations of the same cosmology, we use the test set to estimate the training set uncertainty. We use the diagonal elements of C_{test} as the variances σ_n^2 in Equation (12).

We perform a smoothing on the total covariance matrix, here C_{perf}, to avoid inference issues due to the initially noisy matrix. Our procedure follows that of Lange et al. (2022), and it has been shown by Mandelbaum et al. (2013) to give essentially the same results as applying the Hartlap correction to unbias the inverse covariance matrix (Hartlap et al. 2007). We first compute the correlation matrices, with elements given by $C_{ij}\big/\sqrt{C_{ii}C_{jj}}$, where C_{ij} are the elements of the covariance matrix. The diagonal elements of the correlation matrix must be equal to 1, as each element is perfectly correlated with itself, and the surrounding elements are typically much smaller, so we start by replacing the diagonal elements with the mean of its four neighbors. We then apply a basic Gaussian kernel with width one, to smooth the matrix. Finally, we replace back the diagonal elements. The smoothed total covariance matrix, $C_{\text{perf,smooth}}$, is shown in Figure 1(c). A comparison between using the smoothed and original covariance matrices for parameter inference is shown in Appendix A.

4.3. Inference with Emulator+MCMC

We use Markov Chain Monte Carlo (MCMC) to infer the parameters of the mock catalog given the measured statistics, using the trained Gaussian process emulator to predict the statistic at each set of parameters. For the MCMC process, we use the package `dynesty` (Speagle 2020), which implements dynamic nested sampling. Nested sampling is a method for both obtaining posterior values from a likelihood function and estimating the Bayesian evidence (Skilling 2006); dynamic nested sampling improves upon this by varying the number of live points used in the computation (Higson et al. 2019). While we do not directly make use of the evidence in this work, dynamic nested sampling is faster and more robust than other standard MCMC approaches. We use an evidence threshold of `dlogz = 0.1`, and check that our chains are converged with respect to this threshold. We also perform extensive consistency and convergence tests for other MCMC hyperparameters.

For the HOD and assembly bias parameters, as well as γ_f, we use a uniform prior given by the training set parameter range, with an additional constraint on M_{cut} to be above $10^{11.5}\,M_\odot$. For the cosmological parameters, we use a multidimensional Gaussian prior defined by the mean and covariance of the cosmology training set parameter space (see Figure 3 in DeRose et al. 2019). We also try a flat prior and a high-dimensional ellipsoid for the cosmological parameters, and find no change in the results; we choose to use the multidimensional

Gaussian to improve the stability and speed of the MCMC runs.

We use a likelihood $\mathcal{L}$ of

$$\ln \mathcal{L} = -\frac{1}{2}\left(\frac{y_{\mathrm{pred}} - y_{\mathrm{test}}}{y_{\mathrm{test}}}\right)^{\top} C_{\mathcal{L}}^{-1}\left(\frac{y_{\mathrm{pred}} - y_{\mathrm{test}}}{y_{\mathrm{test}}}\right), \quad (28)$$

where $C_{\mathcal{L}}$ is the covariance matrix described in Section 4.2, and $\boldsymbol{y}$ is a vector containing the concatenated observables. Here, $\boldsymbol{y}_{\mathrm{test}}$ are the statistics measured directly on the test set mock catalog on which we are performing parameter recovery, averaged over the $N_{\mathrm{box}} = 5$ boxes per cosmology and HOD model, and $\boldsymbol{y}_{\mathrm{pred}}$ are the emulator predictions for the observables at the given point in parameter space.

5. Results

In this section, we present the results of our emulation and inference on the AEMULUS test suite. We show the emulator performance (Section 5.1), the results of recovery tests on a single test model (Section 5.2) and a larger test sample (Section 5.3), and an analysis of the scale dependence of our results (Section 5.4).

5.1. Emulator Performance

The performance of the emulators is shown in Figure 2, for each of the observables for all 700 test models. For each test cosmology, we compute the statistic for each of the $N_{\mathrm{box}} = 5$ realizations, and take the measured statistic to be the mean of these. We compute the fractional error between the predicted and measured statistic, and define the error as the symmetrized inner 68% error. We compare this error to the sample variance, the square root of the diagonal of C_{aemulus} for the given observable, as well as this uncertainty scaled by $\sqrt{N_{\mathrm{box}}}$. This scaled uncertainty takes into account the increased precision provided by comparing to the mean over multiple boxes; the covariance matrix scales as the inverse volume, as explained in Section 4.2, and averaging over multiple boxes effectively increases the volume, so we obtain this factor of $\sqrt{N_{\mathrm{box}}}$ (the result is equivalent to taking the square root of the diagonal of C_{test}).

Our emulators achieve very good accuracy across most observables and scales. For $w_{\mathrm{p}}(r_{\mathrm{p}})$, we obtain ∼1%–4%, with a minimum at intermediate scales. For $\xi_0(s)$, we achieve ∼1%–2% error on scales between 1 and 10 h^{-1} Mpc, and up to 13% for the smallest-scale bin. $\xi_2(s)$ has the lowest performance, due to high noise levels, with errors ranging from ∼5% to order unity depending on the scale. For $P_{\mathrm{U}}(s)$, we see extremely small errors of <1% below 20 h^{-1} Mpc scales, due to the low variation of the statistic there; up to 35 h^{-1} Mpc, we achieve ∼1%–7% error, with the error increasing even more for the highest-scale bins. Finally, for $M(s)$, we achieve 1%–3% error on scales below 1 h^{-1} Mpc, and <1% error at larger scales.

At most scales, we see that our emulator error is comparable to the raw sample variance of the AEMULUS simulations adjusted for the effective volume. The exception is $w_{\mathrm{p}}(r_{\mathrm{p}})$, whose error remains a bit larger than this level at all scales; however, this is not entirely unexpected, as the emulation performance error includes both the sample variance and the emulator prediction error.

5.2. Parameter Inference Recovery Tests on a Single Mock

We apply our approach with our GP emulator and MCMC to obtain the posterior distributions of the 18 parameters for a given cosmology+HOD model. As we have five realizations of each test cosmology, we populate all of these with the same HOD and measure the desired statistics on each of them, and then take the mean of these to obtain the measured statistic. These are the values that we compare to the emulator prediction at each step of the MCMC chain. The AEMULUS test volume summed over the five boxes is $N_{\mathrm{box}} \times (1.05 h^{-1}\,\mathrm{Gpc})^3 = 5.79\ (h^{-1}\,\mathrm{Gpc})^3$. This is significantly larger than the volume of the highest-redshift shell used in AEMULUS V: 1.63 $(h^{-1}\,\mathrm{Gpc})^3$, based on the redshift range $0.48 < z < 0.62$ and the CMASS +LOWZ area of 8447 deg^2. For that analysis, the CMASS data was subsampled to a number density of $2 \times 10^{-4} (h^{-1}\,\mathrm{Mpc})^{-3}$, the same as used here, and thus we can make a direct comparison of the volumes. The larger volume of the AEMULUS test boxes by a factor of a few suggests that these are a meaningful test of the precision we will achieve when we apply the approach to data.

We start by performing the inference based on each of the five observables alone. In Figure 3, we show the results on a single cosmology+HOD model; Figure 3(a) shows key cosmological parameters, and Figure 3(b) shows key HOD and assembly bias parameters. We have chosen the latter set of parameters because they are particularly degenerate with cosmological parameters. We see that the different observables have varying effectiveness at constraining the parameters. For instance, for this example mock, $\xi_0(s)$ provides strong constraints on its own on the cosmological parameters, while the other statistics constrain them more weakly, though $\xi_2(s)$ provides surprisingly strong constraining power on γ_f. For the HOD and assembly bias parameters, $P_{\mathrm{U}}(s)$ and $M(s)$ provide constraining power on f_{env}, though $w_{\mathrm{p}}(r_{\mathrm{p}})$ constrains it even more strongly, and $\xi_0(s)$ constrains v_{bs} well. We also note that, because our test set HOD parameter space has the same ranges as our training space, some of the test mocks will have some parameters near the edge of the parameter space. While this may slightly affect the robustness of the MCMC chains in those regions of parameter space, it applies to a small fraction of the parameters across the mocks and should not affect our overall results. Additionally, we note that we have applied optimized smoothing of the posteriors (the default of the `getdist` plotting software), and this occasionally leads to posterior tails that go slightly beyond the edge of the prior space; we check that the unsmoothed posteriors are entirely within the priors.

Next, we explore the constraining power of combining the observables when running the MCMC chains. We start with just $w_{\mathrm{p}}(r_{\mathrm{p}})$, and then one at a time add in $\xi_0(s)$, $\xi_2(s)$, $P_{\mathrm{U}}(s)$, and $M(s)$. The results are shown in Figure 4 for the same model and parameters as Figure 3. As additional observables are added, we obtain tighter and tighter constraints on the parameters. In particular, we can compare the constraints with the three standard observables to those when including the two beyond-standard statistics. For the parameters Ω_{m}, σ_8, M_{sat}, and f_{env}, we see a clear increase in both precision and accuracy when including these new statistics. This indicates promise for the power of the beyond-standard statistics to add additional cosmological information beyond that provided by typical statistics.

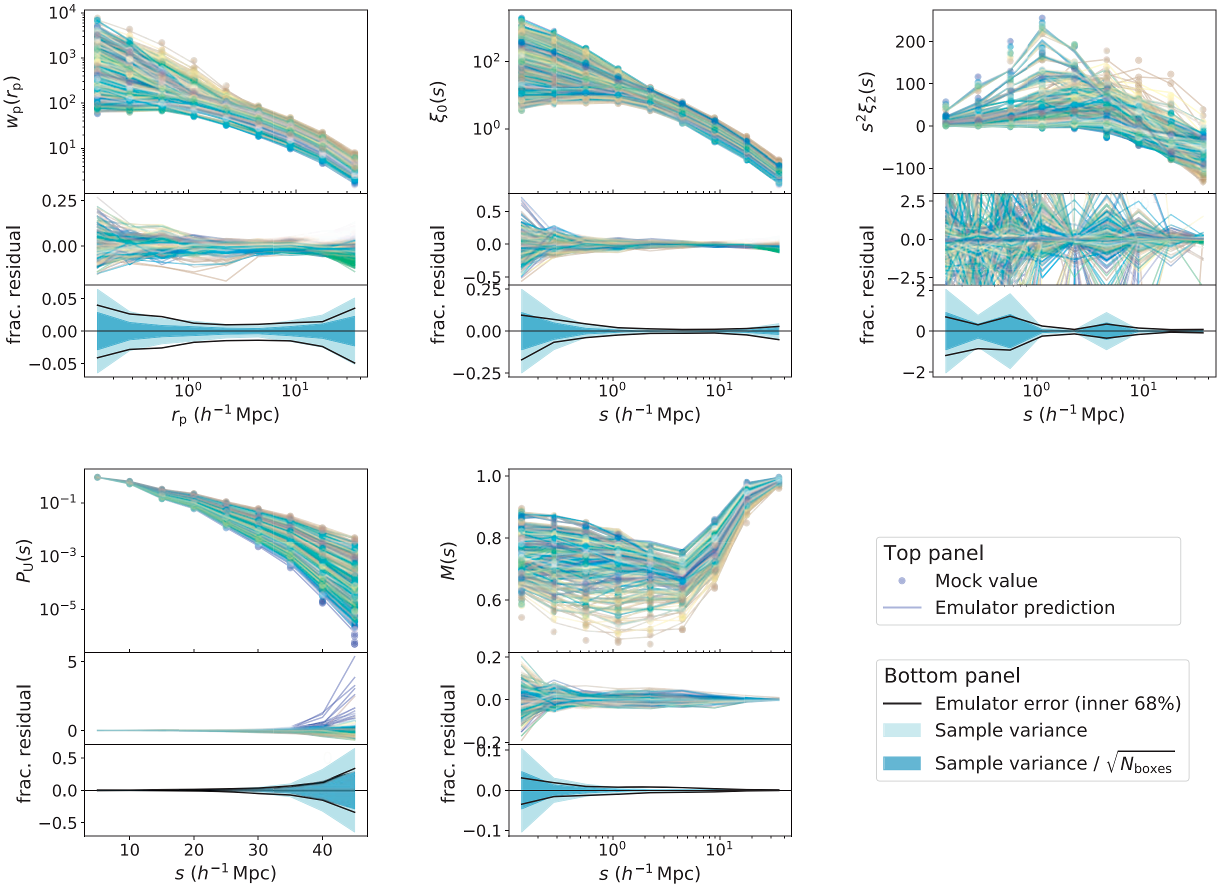
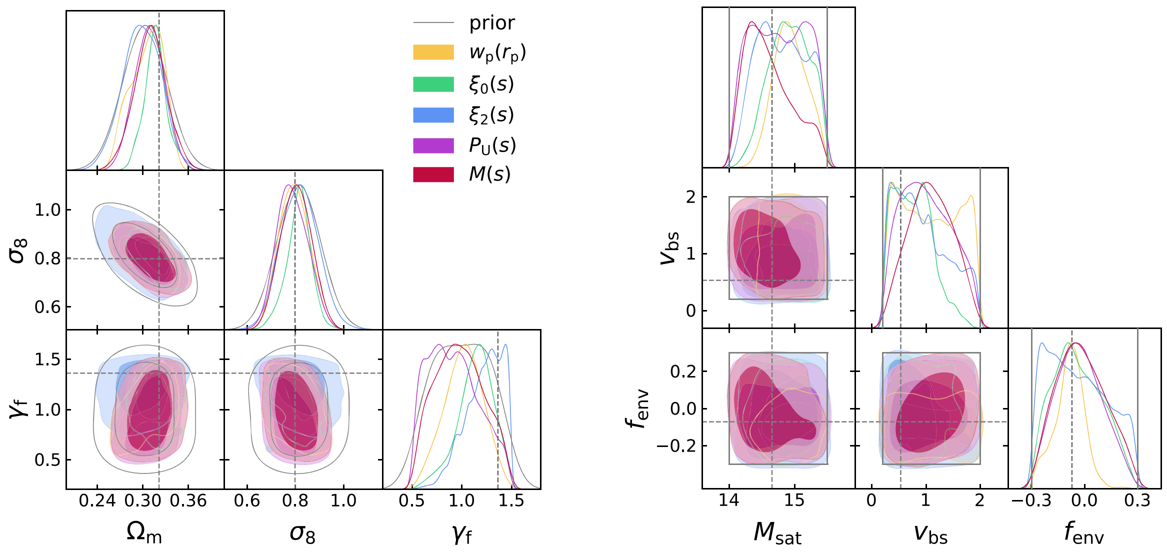


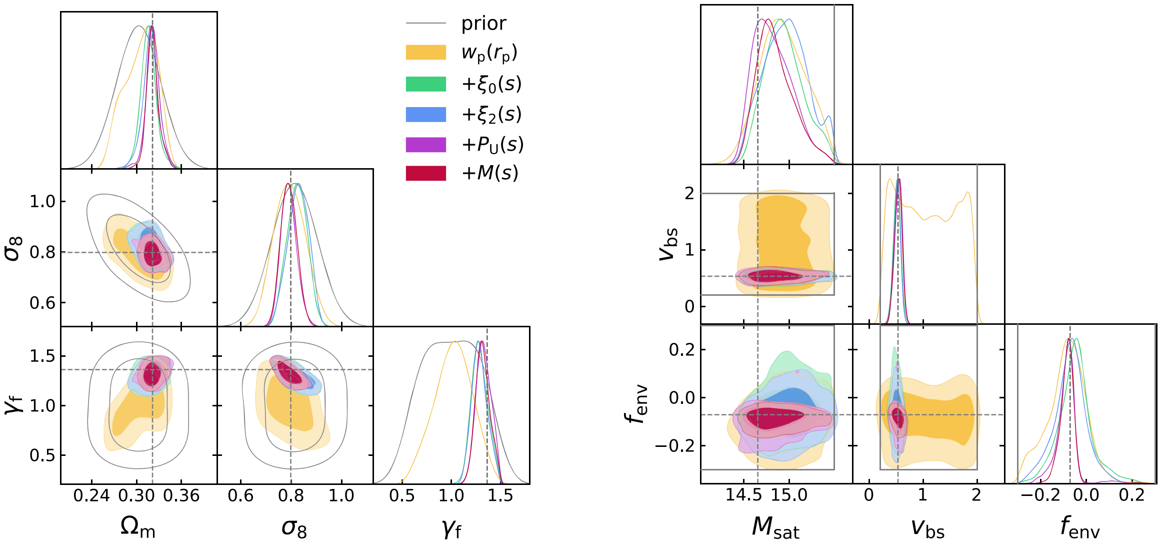
Figure 2. The accuracy of our Gaussian process emulator predictions for the projected correlation function $w_p(r_p)$, monopole and quadrupole of the two-point correlation function $\xi_o(s)$ and $\xi_2(s)$, underdensity probability function $P_U(s)$, and marked correlation function $M(s)$. Top panels show the measured statistics (circles), averaged over N_{box} test boxes for each model, and the corresponding emulator predictions (lines) for each cosmology+HOD model. The colors denote different cosmologies. The middle panels show the fractional error of each of the predictions. The bottom panels show the inner 68% region of the fractional errors (black line), compared to the sample variance of the simulations (light blue). The sample variance scaled by $\sqrt{N_{box}}$ adjusts for the effective increase in volume of comparing emulator predictions to the mean of N_{box} measurements.



(a)

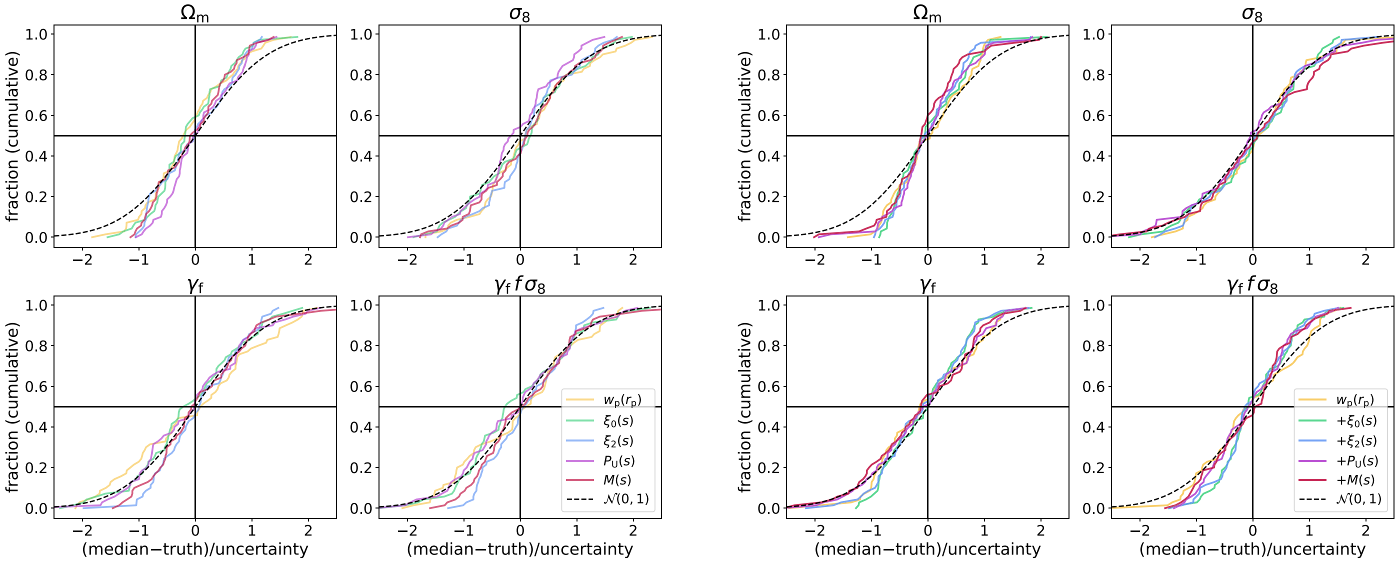
(b)

Figure 3. Recovery tests for a single cosmology+HOD model, using a single observable for each MCMC chain. Contours are shown for (a) key cosmological parameters and (b) key HOD and assembly bias parameters.



(a) (b)

Figure 4. Recovery tests for a single cosmology+HOD model, successively adding in the observables. Contours are shown for (a) key cosmological parameters and (b) key HOD and assembly bias parameters.



(a) (b)

Figure 5. Cumulative distribution functions (CDFs) of the differences between the true parameter value and the median of MCMC chain samples, divided by the uncertainty σ. Panel (a) shows CDFs for each of the observables on their own, and panel (b) adding in the observables successively; panel (b) excludes the two largest-scale $w_p(r_p)$ bins from all combinations, due to a bias discussed in the text. The dashed line shows the CDF of a unit normal distribution for comparison.

5.3. Statistical Results of Recovery Tests

We perform this MCMC inference for all 70 of our recovery test models (7 cosmologies populated with 10 unique HODs each, averaged over the 5 realizations). We first assess the accuracy of the recovered parameters by computing the cumulative distribution function (CDF) of the error on the inferred parameter (difference between the median and truth), normalized by the uncertainty, for the 70 recovery test models. Figure 5(a) shows this CDF for each of the observables used for inference on their own. We find that, for most of the parameters of interest, the CDF follows a unit normal distribution, which is an indication that the recovery is unbiased. (We note that the CDF is not an ideal statistic to measure bias, as the function values are dependent on all previous values, but a histogram with only 70 samples is too noisy to make statements about accuracy.) The exception is Ω_m when using $w_p(r_p)$ or $\xi_0(s)$; we find that the distribution is biased by $\sim 0.5\sigma$ to lower values of Ω_m. There is also a similar slight bias in the $\gamma_f f\sigma_8$ distribution, which we find to be from its dependence on Ω_m. This bias is small but surprising, as these are both such standard statistics.

To see if the issue could be a result of small number statistics, we run a larger set of recovery tests with $w_p(r_p)$ as the sole observable (including all bins), using the full 700 model test suite (each of the seven cosmologies populated with the same 100 HOD models). We compute the CDF of these 700 results and see that the same bias toward low Ω_m values persists. With this larger sample, the histogram is less noisy,

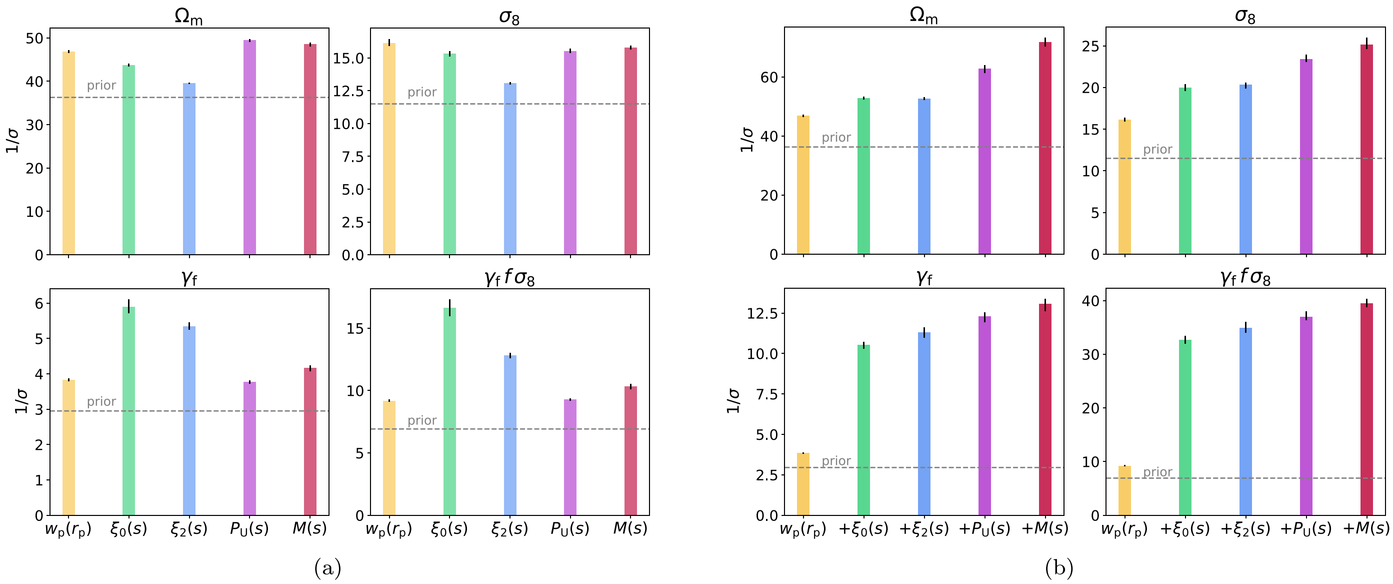


Figure 6. The precision of recovery tests for key parameters, averaged over the 70 test models. The quantity $1/\sigma$ is the inverse uncertainty on the posterior marginalized over the other parameters, with σ defined as the symmetrized inner 68% region. The precision using only the prior is shown by the gray dashed line. Black bars show the uncertainty on $1/\sigma$ using bootstrap estimation. Panel (a) shows the precision for tests with single observables, and panel (b) for successively adding in each observable.

and the bias is small but clearly visible in the histogram as well. One possibility is that there are degeneracies with other cosmological or HOD parameters that contribute to $w_{\rm p}(r_{\rm p})$ and $\xi_0(s)$ favoring lower $\Omega_{\rm m}$ values, but this is difficult to disentangle.

We further investigate this issue by excluding successively larger scales of $w_{\rm p}(r_{\rm p})$ and $\xi_0(s)$ from our analysis, as large-scale clustering should be the most affected by $\Omega_{\rm m}$. We find that removing the two largest-scale bins, above 12.5 h^{-1} Mpc (with logarithmic averages of 17.7 and 35.4 h^{-1} Mpc) results in an unbiased CDF of recovered $\Omega_{\rm m}$ values. We check the effect of this bias on the precision of the recovered parameters by rerunning our recovery tests excluding the two largest-scale bins for $w_{\rm p}(r_{\rm p})$ and $\xi_0(s)$ (but including these bins for the other observables that use them). We find that, when excluding these scales, the precision we obtain on $\Omega_{\rm m}$ using $w_{\rm p}(r_{\rm p})$ decreases by $\sim$ 13% and using $\xi_0(s)$ by 16% (averaged over 70 test models); this is similar when using the three standard statistics, and reduces to a precision decrease of only 9% when including all five statistics. For the quantity $\gamma_f f\sigma_8$, removing these two bins does significantly decrease the precision by 49% when using only $w_{\rm p}(r_{\rm p})$, but for $\xi_0(s)$, this only decreases it by 4%. For the three standard statistics, the decrease is 8%, and for all five, it is only 4%. This corresponds to very small changes to our main target result, the relative increased precision when including the beyond-standard statistics compared to the standard statistics, showing that our overall results are robust to this bin exclusion choice. The exception is that excluding these two large-scale bins actually increases the relative precision on $\Omega_{\rm m}$, perhaps because the beyond-standard statistics, likely $M(s)$, are capturing the large-scale information that we are excluding from $w_{\rm p}(r_{\rm p})$ and $\xi_0(s)$.

For the rest of the results in this paper, except where noted, we exclude the two largest-scale bins for both $w_{\rm p}(r_{\rm p})$ and $\xi_0(s)$. We show the CDF when using combinations of successively more observables in Figure 5(b). We find that these distributions are now generally unbiased for $w_{\rm p}(r_{\rm p})$, as well as other parameter combinations that include $w_{\rm p}(r_{\rm p})$ and $\xi_0(s)$, for all of the cosmological parameters; the CDFs generally follow the unit normal distribution. Both $\Omega_{\rm m}$ and $\gamma_f f\sigma_8$ show distributions slightly tighter than the normal distribution, indicating that we have overestimated our errors. This means that our errors may be conservative, but the difference is small and we do not expect this to have significant effects on our results.

We next show our results on the precision of the recovered parameters. For each parameter, we compute the uncertainty σ on the posterior, defined as the symmetrized inner 68% confidence region, marginalized over the other parameters. In Figure 6(a), we show the inverse uncertainty $1/\sigma$ for each of the key cosmological parameters, including the combined quantity $\gamma_f f\sigma_8$, averaged over all 70 test models, when using each of the statistics alone for the inference. It should be noted that larger bars indicate tighter constraints. We compare this to the uncertainty obtained when just using the prior. We see that all of the statistics on their own provide additional constraining power over the prior, for all parameters: $P_{\rm U}(s)$ provides the most information for $\Omega_{\rm m}$; all the statistics besides $\xi_2(s)$ constrain σ_8 similarly; and $\xi_0(s)$ constrains γ_f and $\gamma_f f\sigma_8$ the most strongly. The amount of information from $w_{\rm p}(r_{\rm p})$ is relatively high compared to that found by other analyses (e.g., Lange et al. 2022); we find that this is largely due to our choice to integrate out to only 40 h^{-1} Mpc along the line of sight, which preserves information in RSDs. We test integrating out to 80 h^{-1} Mpc and find much less information content in $w_{\rm p}(r_{\rm p})$ alone, though it still contains some. For the beyond-standard statistics, it is noteworthy that $P_{\rm U}(s)$ and $M(s)$ do provide information on their own, in particular on $\Omega_{\rm m}$ and σ_8.

We next perform recovery tests adding in each observable one at a time for the full test suite. We show the results in Figure 6(b), again for the mean of 70 test models. We see that the inverse uncertainty monotonically increases as we add in additional observables. Our main result is that the constraining power increases significantly between using only the combined standard observables, $w_{\rm p}(r_{\rm p})$+$\xi_0(s)$+$\xi_2(s)$ (blue), and when

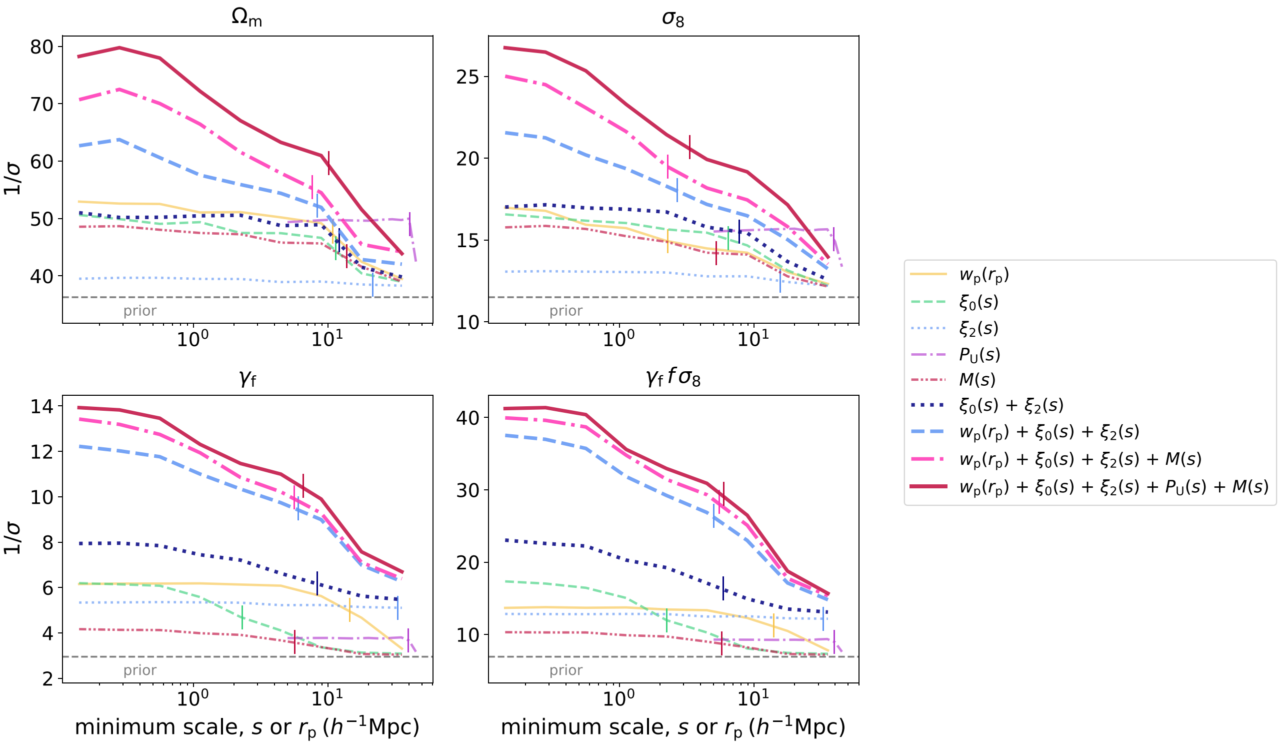


Figure 7. The precision of recovery tests as a function of the minimum scale used in the analysis, averaged over the 70 test models. The maximum scale remains fixed at the maximum bin value. The precision is shown for chains using a single observable, as well as for several multiobservable combinations. The vertical bars indicate the scale at which half of the constraining power for that observable is in larger scales and half in smaller scales. We note that $P_U(s)$ is measured on different scales than the other observables, from 5 to 45 h^{-1} Mpc, so at a minimum scale below 5 h^{-1} Mpc it results in an overall shift in precision.

adding in the beyond-standard statistics as well, $w_p(r_p)$+$\xi_0(s)$+$\xi_2(s)$+$P_U(s)$+$M(s)$ (red). The change in precision for these two cases tells us the amount of additional information contained in these new statistics: The precision increases (defined as the fractional decrease in the uncertainty σ) by 27% for Ω_m, 19% for σ_8, 13% for γ_f, and 12% for the combined growth of structure parameter $\gamma_f f\sigma_8$. These are significant increases, given the current precision of cosmological measurements.

5.4. Scale Dependence

We investigate the dependence of our parameter constraints on the scales used in the inference. To analyze the contribution of small scales, we vary the minimum scale bin used and rerun the MCMC chains, for each parameter individually as well as the five-observable combined constraint. The results are shown in Figure 7, averaged over the 70 test models. We note that the $P_U(s)$ uses a different binning scheme than the other observables, so it is only shown on the scales on which it is computed, 5–45 h^{-1} Mpc, and when it is included in combination with the other observables, it results in an overall shift in precision below 5 h^{-1} Mpc. For this reason, we add the $P_U(s)$ and $M(s)$ in the opposite order as the rest of this paper. We also include using just the combination $\xi_0(s) + \xi_2(s)$, as many analyses do. Similarly, we run recovery tests varying the maximum scale. The $1/\sigma$ lines for the minimum and maximum scale variation will cross each other at a particular scale; this scale is marked by a vertical bar and indicates the scale at which equal information is provided by scales smaller than and larger than this scale. Thus, a vertical bar far in the small-scale regime means that most of the information comes from small scales (as only the smallest scales are needed on their own to equal the information content in all the larger scales), and conversely, a vertical bar at large scales means that most information comes from large scales.

As we include smaller scales, the precision increases mostly monotonically. Using the vertical bars described above, we find that, for $\gamma_f f\sigma_8$, using the five-observable constraint, scales from 0.1 to 6 h^{-1} Mpc provide as much information as the scales 6–50 h^{-1} Mpc. We find that the information content continues to increase as we include smaller scales, until a scale of $\sim$0.5 h^{-1} Mpc. This is a remarkable finding, given that previous analyses either have not pushed to scales this small or did not find as significant of a contribution from small scales; we discuss this further in Section 6.

To understand this result, we look at the constraints from individual observables for $\gamma_f f\sigma_8$. For $\xi_0(s)$, half of the information comes from scales below $\sim$2.25 h^{-1} Mpc; for $M(s)$, below $\sim$5.75 h^{-1} Mpc; and for $w_p(r_p)$, below $\sim$15h^{-1} Mpc. Thus, $\xi_0(s)$ is driving the large amount of information on $\gamma_f f\sigma_8$ at small scales, with a contribution from $M(s)$. We also look at the constraints on the individual key cosmological parameters Ω_m, σ_8, and γ_f; for the five-observable constraint, half the information on σ_8 comes from scales below $\sim$ 3.3 h^{-1} Mpc; on f, below $\sim$ 6.5 h^{-1} Mpc; and on Ω_m, below $\sim$ 10 h^{-1} Mpc. Thus, the small-scale constraints on $\gamma_f f\sigma_8$ are driven mainly by the ability of small scales to constrain σ_8. Looking at the commonly used statistic combination $\xi_0(s) + \xi_2(s)$, we see that the precision nearly flattens out for scales below $\sim$10 h^{-1} Mpc for Ω_m and σ_8. Including $w_p(r_p)$ adds significant constraining power at small scales for all of the parameters; we discuss this further below. Finally, adding in $P_U(s)$ and $M(s)$ accesses a significant amount of additional information at smaller scales, in particular for Ω_m and σ_8.

Notably, the significant additional constraining power from adding in $w_p(r_p)$ to $\xi_0(s)+\xi_2(s)$ differs from the findings of Lange et al. (2022), who found that it only marginally improved constraints. Given that the effect of $w_p(r_p)$ is somewhat stronger for γ_f and $\gamma_f f\sigma_8$ in our analysis, and Lange et al. (2022) do not include this velocity field rescaling parameter, it seems that the increase in constraining power we find is due to the sensitivity of $w_p(r_p)$ to velocity information. Indeed, we only integrate out to $\pi_{\max} = 40\,h^{-1}$ Mpc, while Lange et al. (2022) uses a value of $\pi_{\max} = 80\,h^{-1}$ Mpc. We perform a test using this larger value and find that, as expected in this case, $w_p(r_p)$ does not add much more constraining power to either γ_f or $\gamma_f f\sigma_8$. We further investigate this by fixing the γ_f parameter in our inference to the true value for each test mock and rerunning the inference, with the goal of isolating the effect of this parameter. We find that, when we do this, the difference between the constraints using $w_p(r_p)+\xi_0(s)+\xi_2(s)$ and $\xi_0(s)+\xi_2(s)$ only is greatly reduced: with free γ_f, dropping $w_p(r_p)$ leads to a 63% reduction in precision on $\gamma_f f\sigma_8$, while with fixed γ_f, it is only a 27% reduction on $\gamma_f f\sigma_8$ (which in this case is determined only by the precision on f and σ_8). Thus, we conclude that our choice of $\pi_{\max}$ preserves significant velocity information that allows $w_p(r_p)$ to constrain the growth of structure parameter through its sensitivity to the halo velocity field rescaling parameter γ_f. Though this is a plausible explanation, the information content of $w_p(r_p)$ being independent of $\xi_0(s)$ and $\xi_2(s)$ is still a somewhat surprising result, and the combination of these statistics should be considered carefully in future work.

5.5. Recovery Tests on External Models

A key assumption on which our approach relies is that the HOD model is sufficiently flexible to span the space of observed data. The HOD is just one way of relating halo properties to the galaxy distribution, and it incorporates certain (physically and empirically motivated) assumptions about galaxy formation. This is notable in relation to perturbation theory approaches, which require a large number of nuisance parameters to model higher-order statistics such as the bispectrum (e.g., Philcox et al. 2022), while the HOD is a relatively compact parameterization that makes stronger assumptions while showing promise in still being able to model high-order statistics (e.g., Zhang et al. 2022a). To check that our HOD model is sufficiently flexible to model our chosen statistics, we test our approach on a catalog constructed with a different galaxy formation prescription that incorporates different assumptions than the HOD, namely Subhalo Abundance Matching (SHAM; e.g., Kravtsov et al. 2004; Vale & Ostriker 2004; Conroy et al. 2006). This is an important validation step before applying our emulators to real data. When we adapt our emulators for the full data analysis, we will perform additional tests in this vein to ensure that our framework encompasses the range of expected galaxy formation scenarios.

For this test, we use mock catalogs generated from the UNIT simulations[13] (Chuang et al. 2019) to check that our framework generalizes beyond the AEMULUS N-body simulations. The UNIT simulations have a mass resolution of $1.2 \times 10^9\,h^{-1}M_\odot$ and consist of two pairs of simulations constructed with the fixed-and-paired inverse-phase technique (Angulo & Pontzen 2016). Each simulation has a volume of $(1\,h^{-1}\,\mathrm{Gpc})^3$, leading to an effective volume significantly larger than the AEMULUS boxes, so the clustering statistic measurements should be more precise. To test that our approach is robust to our use of an HOD model with environment-dependent galaxy assembly bias, we instead populate the UNIT simulations using the SHAM approach. SHAM assigns galaxies to subhalos based on a rank-ordered relation between galaxy mass and subhalo mass, with some additional parameters to regulate the scatter, and it is able to reproduce galaxy assembly bias to some extent. We specifically use the SHAM method of Lehmann et al. (2016) to generate our UNIT mocks.

For this test, we require a data covariance matrix for the UNIT mock data, and we use the GLAM Particle-Mesh simulations (Klypin & Prada 2018) for this purpose. These simulations have many independent realizations; we use 986 boxes for our covariance estimate. They are all at the same cosmology; we consider the covariance of the fractional differences from the mean for each statistic. The GLAM boxes have a volume of $(1\,h^{-1}\,\mathrm{Gpc})^3$, so we rescale the covariance matrix for the effective UNIT volume. We use the emulator covariance matrix C_{emu} described in Section 4.2, add this to the data covariance, and then perform a Gaussian smoothing on this total covariance matrix to obtain the final covariance we use in the likelihood function.

We compute the statistics on the UNIT SHAM mocks in the same way as for the AEMULUS mocks, first including redshift-space distortions for the UNIT galaxies. All of the measured statistics are within 1σ of the mean of the training set mocks. Still, we find that, when we perform a full MCMC over all the parameters with the UNIT statistics, some of the parameter constraints are slightly biased (~ 1–2σ) when adding the two beyond-standard statistics to the data vector. We note that we also encountered a similar issue when attempting a recovery test using the Uchuu simulations (Ishiyama et al. 2021) populated with SHAM galaxies. Below, we will show results based on the UNIT simulation, but the main findings for the SHAM tests hold for Uchuu as well. This suggests that our HOD parameter space or parameterization may not be sufficiently flexible to encompass the details of the SHAM-distributed galaxies, and that the differences between the HOD and SHAM galaxies manifest in the beyond-standard statistics much more than in the standard statistics. This is an important issue to investigate further in future work.

For our SHAM recovery tests, we thus choose to focus our constraints on the growth-related parameters. We fix two of the cosmological parameters, w and h, to their fiducial values (similar to adopting a CMB prior on them), as they are not part of the goals of this analysis (nor other small-scale clustering analyses that use this approach). We leave all other cosmological parameters and all HOD parameters free in the MCMC exploration. We also exclude the two largest-scale bins for $w_p(r_p)$ and $\xi_0(s)$, as these showed a slight bias in the AEMULUS recovery test results (see Section 5.3). The results of the tests for the UNIT simulation are shown in Figure 8, both using just standard statistics and including the beyond-standard statistics. We find that, in both cases, we can accurately recover the UNIT cosmological parameters as well as γ_f. The inclusion of the beyond-standard statistics results in an increase in precision of 40% on Ω_m, 25% on σ_8, 30% on γ_f, and 17% on $\gamma_f f\sigma_8$; this is similar to (in fact, even better than) our findings with AEMULUS recovery tests. We do note that, for some of the parameters, the results become slightly more biased when adding in the beyond-standard statistics, but all parameters are still recovered to within 1σ. This may be related to the

[13] http://www.unitsims.org

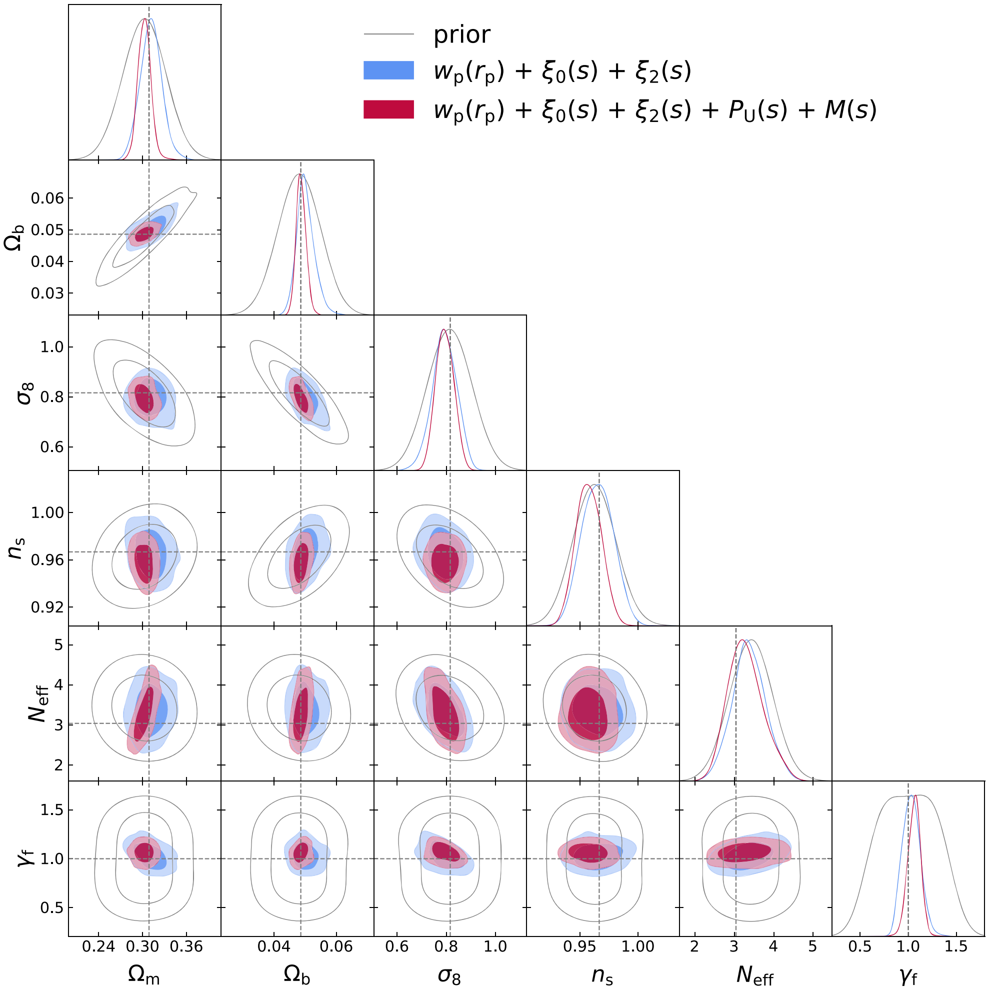


Figure 8. Recovery test on the UNIT mock catalog. Constraints are shown for the cosmological parameters and γ_f when using just the standard statistics (blue), and when including the $P_U(s)$ and $M(s)$ (red); the prior is shown for comparison (gray), and the true parameter values are shown in the dashed gray lines. We keep w and h fixed, as discussed in the text. The parameters are recovered accurately, with the beyond-standard statistics adding increased precision on the parameters of interest.

aforementioned issue with differences between the SHAM and HOD galaxies that are captured by those statistics. We also test fixing other combinations of cosmological parameters, and find that fixing w and N_{eff} (with h and the others free) or fixing w and Ω_b produces results similar to the case shown here. Fixing more than two of these parameters does not change the results, so we chose the two-fixed-parameter test as our fiducial case. While this SHAM test does reveal caveats to our approach, the results are still promising for the application of this framework to real data sets.

6. Discussion and Conclusions

We have constructed Gaussian process emulators for galaxy clustering statistics using the AEMULUS simulation suite, including the nonstandard statistics the underdensity probability function $P_U(s)$ and the marked correlation function $M(s)$, which we expect to contain additional information relevant to constraining cosmological parameters of interest. We achieve typical prediction errors of ∼2% with our emulator, depending on the scale and statistic. Using held-out test simulations, we perform recovery tests to determine how well we can constrain the input parameters. We find that including the beyond-standard statistics significantly increases the precision on the recovered parameters, by 19% on σ_8, 27% on Ω_m, and 12% on $\gamma_f f \sigma_8$. We confirm that our framework is robust to different simulations and galaxy bias models by testing it on mock catalogs constructed from the UNIT simulations and the SHAM method, on which we achieve unbiased constraints and a similar improvement in precision when including the beyond-standard statistics.

To follow this proof-of-concept work, we will apply these emulators to measure the growth of structure in a current galaxy sample (BOSS or DESI). We expect that our combination of beyond-standard statistics with small-scale emulation will improve constraints; for instance, Satpathy et al. (2019) used the marked correlation function to analyze the BOSS data and found that their results were limited by modeling RSD effects on small scales. This analysis will require a careful treatment of many issues and subtleties in real data. We will have to handle redshift evolution, by working in redshift slices with emulators trained at the proper redshift. We will require a sample constant in number density, both to match our emulators and because void- and density-based statistics are particularly sensitive to variations in number density. One of the main issues when applying to BOSS data will be fiber collisions, which lead to galaxies without measured redshifts, producing a nontrivial impact on clustering measurements especially at small scales (e.g., Zehavi et al. 2002).

Additionally, we will have to handle survey geometry effects including edges and bad fields. The underdensity probability function and the local density-based marks used for the marked correlation will both be especially sensitive to these issues; we will apply fiber collision weights to the statistics and volume corrections to the spheres used for the density computations, and perform robust tests to ensure that we can recover unbiased parameters.

The application of this work to the BOSS sample will extend the project of AEMULUS V (Zhai et al. 2023). The AEMULUS V analysis used $w_{\rm p}(r_{\rm p})$, $\xi_0(s)$, and $\xi_2(s)$, the standard statistics discussed in this paper, and obtained tight constraints on the growth of structure parameter $f\sigma_8$ in three redshift bins. The analysis obtained a low value of $f\sigma_8$ compared to Planck constraints based on a ΛCDM+GR model, adding to a recent wave of similarly low results based on small-scale clustering (Chapman et al. 2021; Lange et al. 2022; Yuan et al. 2022). These studies are also based on standard clustering statistics; bringing in additional statistics and thus additional constraining power will allow for clearer tests of internal consistency between these analyses, as well as testing the demonstrated tension with Planck results.

There are multiple effects that could be contributing to this $f\sigma_8$ tension. One is additional baryonic effects that influence galaxy formation and are unmodeled in the HOD, introducing errors; while these are unlikely to be relevant at current precision, in future surveys they may become important. Future work will incorporate additional flexibility in the galaxy bias and assembly bias models to test this hypothesis, and this will in turn require increased constraining power from the data. The complementary information provided by nonstandard statistics, as shown in this work, will be important in offsetting this flexibility to obtain high-precision constraints on $f\sigma_8$ and help confirm or rule out this explanation for the $f\sigma_8$ tension. Another potentially relevant effect is that of massive neutrinos, which suppress the growth of structure in a scale-dependent way. The next generation of the AEMULUS simulations (AEMULUS ν; DeRose et al. 2023) will incorporate massive neutrinos, and the emulation of nonstandard statistics will also be important in obtaining precise small-scale constraints from this updated model.

In this work, we have included a detailed handling of the covariances between the observables, incorporating both the data and emulator covariances in our inference. To estimate the relative contribution of these sources of uncertainty in our target analysis, we perform a volume-based scaling of the data covariance of the AEMULUS test boxes ($C_{\rm aemulus}$) to one of our target samples, the CMASS high-z bin. We find that this data covariance is of similar order to the emulator covariance, and the dominating source of uncertainty depends on the observable and scale; in either case, they are never more than a factor of $\sim$2 different. While this indicates that we are still theory-limited in some regimes, this is reasonable given the newness of the emulator approach. A comparison between the emulator and data covariances for the standard statistics is also shown in Figure 15 of AEMULUS V (Zhai et al. 2023), which similarly finds that the errors are comparable. Future iterations of this type of analysis will be able to reduce the theory uncertainty through a combination of more training simulations (as in AEMULUS ν), larger simulation volumes, and improved emulation techniques. These improvements will become increasingly important as the data uncertainty also gets reduced with future observations.

The effects of galaxy assembly bias are not yet a concern, given the current precision of our surveys, as shown in the Zhai et al. (2023) BOSS RSD analysis, but as both our data and constraining power of methods improve, this will become a key source of uncertainty. Previous works have found a small but significant dependence on halo environment (e.g., Zehavi et al. 2018; Yuan et al. 2021). The density-sensitive statistics we investigate here—namely the $M(s)$ with marks given by the galaxy number density on $10\,h^{-1}$ Mpc scales, and the $P_{\rm U}(s)$, which measures underdense regions across a large range of scales—target this environmental bias. We have shown that these statistics are well-positioned to improve constraints on cosmological parameters by breaking degeneracies between cosmological and environmental assembly bias effects. Other sources of assembly bias, such as halo formation time, concentration, and spin, could be analyzed with marked correlation functions based on these properties or other similarly targeted statistics; these can be readily incorporated into our emulation framework.

More broadly, this work confirms that additional, beyond-standard clustering statistics, namely the $P_{\rm U}(s)$ and $M(s)$, can increase the constraining power in existing data, with little added cost. This approach could be extended to include other statistics that depend on the goals of the analysis. These could include the three-point function (e.g., Takada & Jain 2003; McBride et al. 2011), the kNN-CDF (Banerjee & Abel 2021), and galaxy group statistics such as the group multiplicity function (e.g., Berlind & Weinberg 2002) and the group velocity dispersion. We will explore some of these in future work. It is important to note that these statistics may be more sensitive to the choice of galaxy bias model than standard statistics, as we found in our initial tests on SHAM galaxies (Section 5.5). This should be carefully checked when incorporating new statistics; in our case, we do find that including the $P_{\rm U}(s)$ and $M(s)$ result in slightly biased parameter constraints on the SHAM galaxies when all cosmological parameters are left free. This may point to the need for an even more flexible HOD parameterization, an investigation we leave for future work.

One of the primary goals of the AEMULUS project is to extract information from small-scale clustering, which is difficult to model theoretically and expensive to simulate fully. Here, we have shown that there is significant information at small scales for nearly all of the statistics we analyze. For the constraint on $\gamma_f f\sigma_8$, we find that scales from 0.1 to $6\,h^{-1}$ Mpc contribute half of the information content, and that there is additional information all the way down to $0.5\,h^{-1}$ Mpc. This confirms a similar result by Zhai et al. (2019), which uses $w_{\rm p}(r_{\rm p})$, $\xi_0(s)$, and $\xi_2(s)$, and includes the halo velocity field scaling parameter γ_f. Some recent analyses have not found as much additional information at these small scales. Lange et al. (2022) conclude that, for their low-redshift sample, which is closer in number density to the one analyzed here, scales between 1 and $2h^{-1}$ Mpc increase the constraining power on $f\sigma_8$ by a small amount, and scales below $\sim 1h^{-1}$ Mpc not at all. As discussed in Section 5.4, they do not incorporate a γ_f parameter to scale the velocity field, and they do not use $w_{\rm p}(r_{\rm p})$ as we do. This model flexibility, which the work of Zhai et al. (2019) also includes, combined with a statistic sensitive to velocity information, may allow us to extract additional information from small scales. The analysis by Lange et al. (2022) does include an assembly bias model using the decorated HOD framework (Hearin et al. 2016), but this is not as flexible as our three-parameter environmental assembly bias model. Our increased flexibility on this front may also contribute to the discrepancy, though future work should revisit these hypotheses.

Finally, in this work we built emulators at fixed redshift and scale. To apply to different data sets, we will require predictions at various redshifts, for which suites of emulators can be constructed and trained at the needed redshifts; an extension of this work could construct emulators that are able to make predictions as a continuous function of redshift. In a similar vein, here we emulated the clustering statistics at fixed scale, with a different model trained for each bin. In future work, we could train the model on all bins simultaneously to include the full covariance properties; even better, we could include scale as an input parameter and make predictions at any scale.

Acknowledgments

K.S.F. thanks Sean McLaughlin, Johannes Lange, Sihan Yuan, David W. Hogg, and the Astronomical Data group at the Center for Computational Astrophysics for helpful discussions. We are grateful to the anonymous referee, whose feedback has greatly improved this manuscript. This work received support from the U.S. Department of Energy under contract number DE-AC02-76SF00515 and from the Kavli Institute for Particle Astrophysics and Cosmology. K.S.F. is supported by the NASA FINESST program under award number 80NSSC20K1545. J.L.T. acknowledges support of NSF grant AST-2009291. J.D. is supported by the Lawrence Berkeley National Laboratory Chamberlain Fellowship. This research made use of computational resources at SLAC National Accelerator Laboratory and the NYU Department of Physics; the authors thank the SLAC computational team and Mulin Ding at NYU for computational 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-AC0205CH11231.

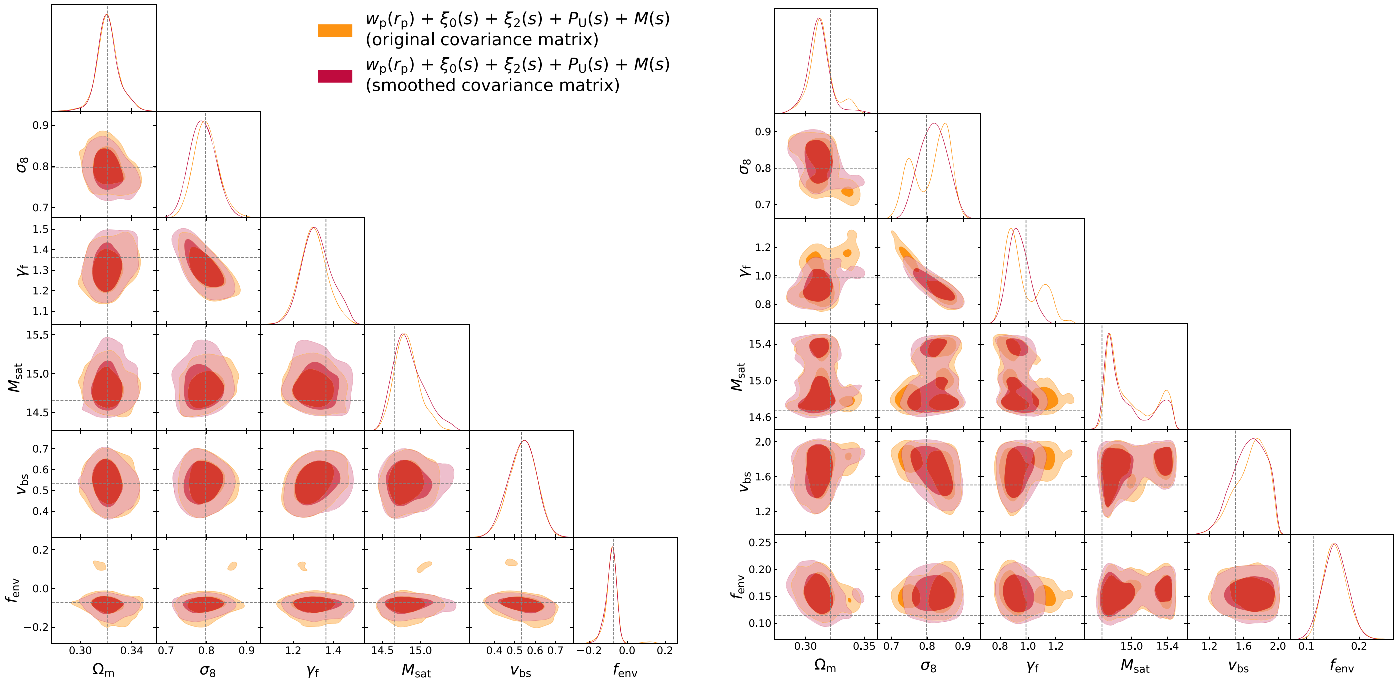
We thank Instituto de Astrofisica de Andalucia (IAA-CSIC), Centro de Supercomputacion de Galicia (CESGA), and the Spanish academic and research network (RedIRIS) in Spain for hosting Uchuu DR1, DR2, and DR3 in the Skies & Universes site for cosmological simulations. The Uchuu simulations were carried out on Aterui II supercomputer at Center for Computational Astrophysics, CfCA, of National Astronomical Observatory of Japan, and the K computer at the RIKEN Advanced Institute for Computational Science. The Uchuu Data Releases efforts have made use of the skun@IAA_RedIRIS and skun6@IAA computer facilities managed by the IAA-CSIC in Spain (MICINN EU-Feder grant EQC2018-004366-P).

The UNIT simulations were run at the MareNostrum Supercomputer hosted by the Barcelona Supercomputing Center (BSC), Spain, thanks to the PRACE project grant No. 2016163937.

Software: numpy (Van Der Walt et al. 2011), IPython (Perez & Granger 2007), scipy (Virtanen et al. 2020), matplotlib (Hunter 2007), corrfunc (Sinha & Garrison 2019, 2020), halotools (Hearin et al. 2017), dynesty (Speagle 2020), george (Ambikasaran et al. 2016), getdist (Lewis 2019). Our emulator and related data products are publicly accessible in living form at https://github.com/kstoreyf/aemulator and in archived form on Zenodo doi:10.5281/zenodo.8433111 (Storey-Fisher 2023a). The code used to compute the clustering statistics is available in living form at https://github.com/kstoreyf/clust and in archived form on Zenodo doi:10.5281/zenodo.8433095 (Storey-Fisher 2023b).

Appendix A
Covariance Matrix Comparison

We compare the posteriors of recovery tests when using the original noisy covariance matrix compared with the Gaussian-smoothed covariance matrix, as described in Section 4.2. The results are shown in Figure 9 for two different cosmology+HOD models for a mix of key cosmological and HOD parameters. We find that, for the generally well-behaved model (Figure 9(a)), the



(a)

(b)

Figure 9. A comparison of the effect of the covariance matrix on recovered parameters. Panels (a) and (b) show recovery tests of key parameters for two different cosmology+HOD models, using all five observables, with the original covariance matrix compared to the covariance matrix with a Gaussian smoothing.

posteriors are similar between the two covariance matrices, with the smoothed matrix resulting in slightly more accurate parameter estimates. For the less well-behaved model (Figure 9(b)), the posteriors are quite noisy with the original covariance matrix. Using the smoothed version cleans up some of the spurious modes in the posteriors, suggesting that the smoothing does help in avoiding issues related to noise in the covariance matrix. However, some of the modes persist even when using the smoothed matrix, indicating that perhaps we are still not properly sampling our parameter space, or that some of these regions of parameter space may be actual good fits to the observables and indicate true degeneracies in the parameters.

Appendix B
Recovery Test Results for HOD & Assembly Bias Parameters

We show the precision of our recovery tests for the HOD parameter $M_{\rm sat}$ and the three assembly bias parameters, $f_{\rm env}$, $\delta_{\rm env}$, and $\sigma_{\rm env}$, in Figure 10. Results are shown averaged over the 70 test models, when successively adding in each of our five observables. We see that, for all of the parameters, each of the observables provides additional information on the parameter, with the exception of $\xi_2(s)$. The two beyond-standard statistics $P_{\rm U}(s)$ and $M(s)$ provide significantly increased precision compared to the standard statistics alone. This indicates that the additional constraining power from these statistics for the cosmological parameters may be related to their heightened sensitivity to assembly bias, as well as the ability of the combination of many observables to constrain the flexible HOD model.

It is somewhat surprising that $w_{\rm p}(r_{\rm p})$ on its own provides significant constraining power over the prior on $f_{\rm env}$, the amplitude of environmental assembly bias. Investigating the relationship between these, we find that, with the rest of the parameters fixed, at large scales, $w_{\rm p}(r_{\rm p})$ decreases as $f_{\rm env}$ is increased. This makes sense because positive $f_{\rm env}$ values effectively transfer halos from high- to low-density regions, reducing overall clustering, which translates to a lower two-halo term. It is notable that this effect is significant enough to be able to constrain this parameter, highlighting the importance of including a flexible model of environmental assembly bias.

Appendix C
Full Posterior Plots

We show contour plots of all the recovered parameters for a single cosmology+HOD test model in Figure 11, when successively adding in our observables. In Figure 11(a), we show the cosmological parameters; in Figure 11(b), a combination of the key cosmological, HOD, and assembly bias parameters; and in Figure 11(c), all the HOD and assembly bias parameters. We can clearly see the degeneracies between many of the parameters here, and for many of these, including the beyond-standard statistics breaks the degeneracy. This is true for degeneracies between cosmological parameters and HOD parameters, as with σ_8 and $M_{\rm sat}$; between HOD parameters, as with $v_{\rm bs}$ and $\sigma_{\log M}$; and between assembly bias parameters, as with $f_{\rm env}$ and $\sigma_{\rm env}$. This helps explain how the combination of our flexible assembly bias model and the emulation of beyond-standard statistics improves our precision on cosmological parameter constraints.

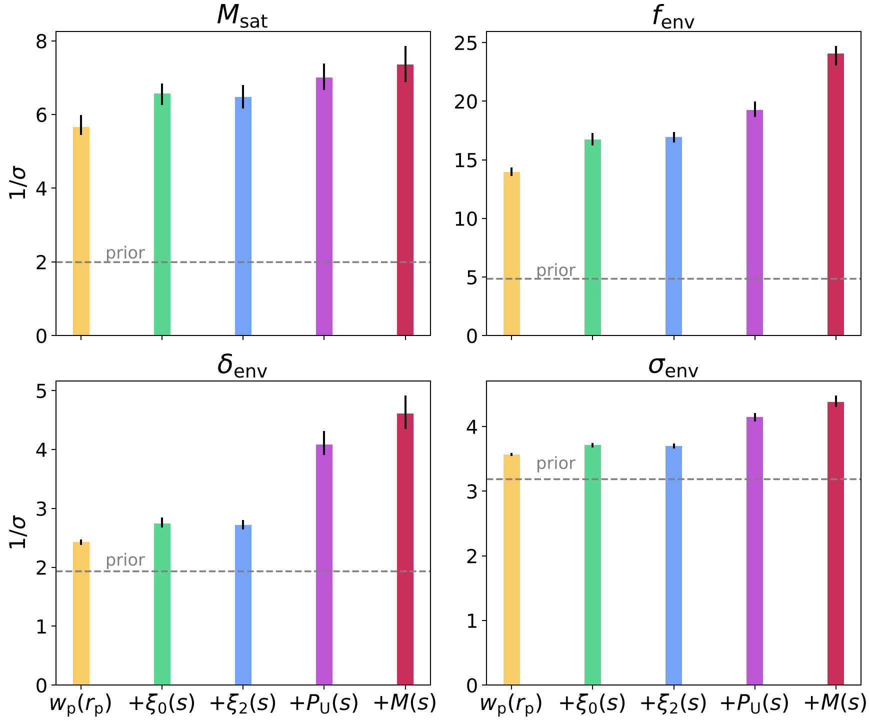
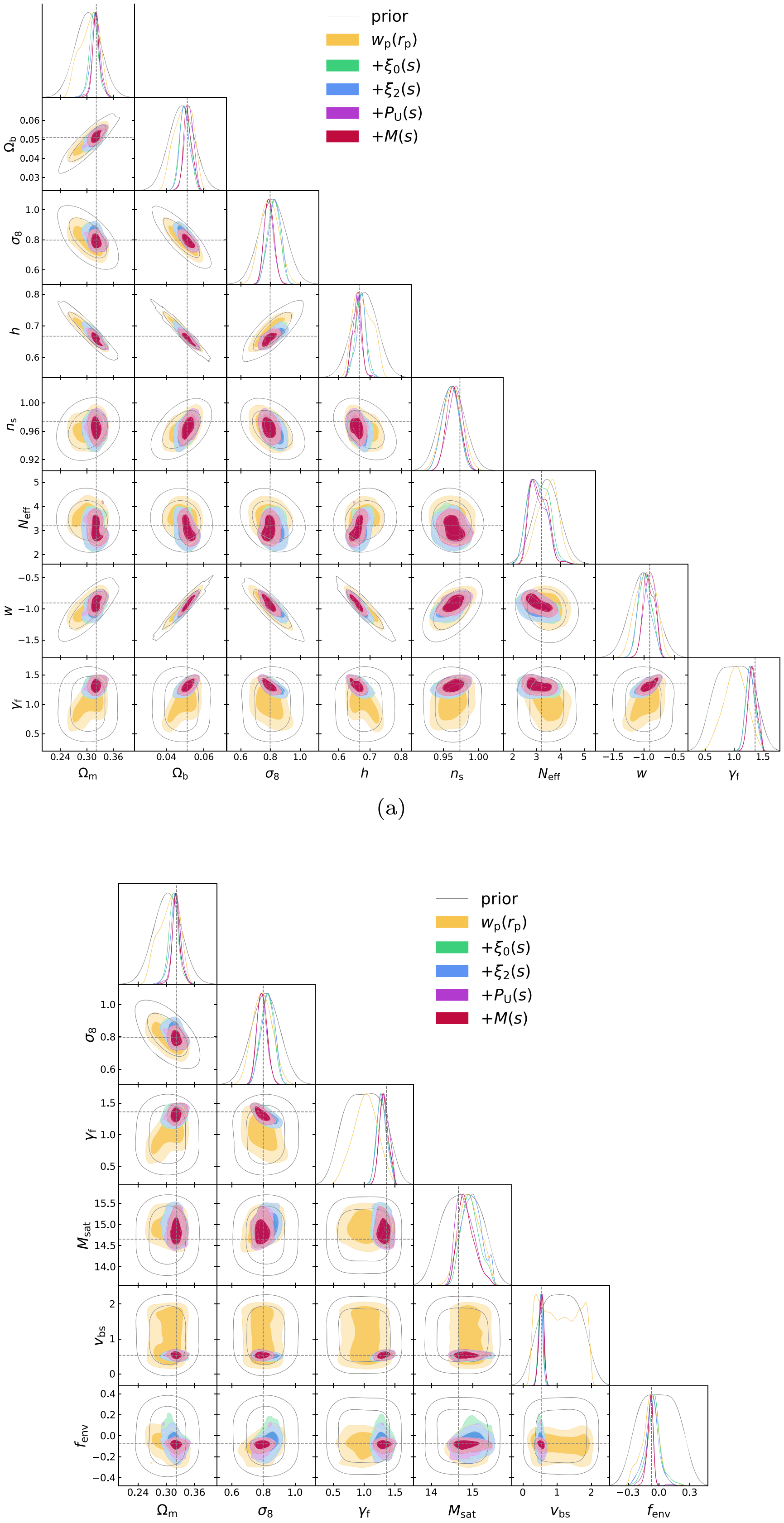


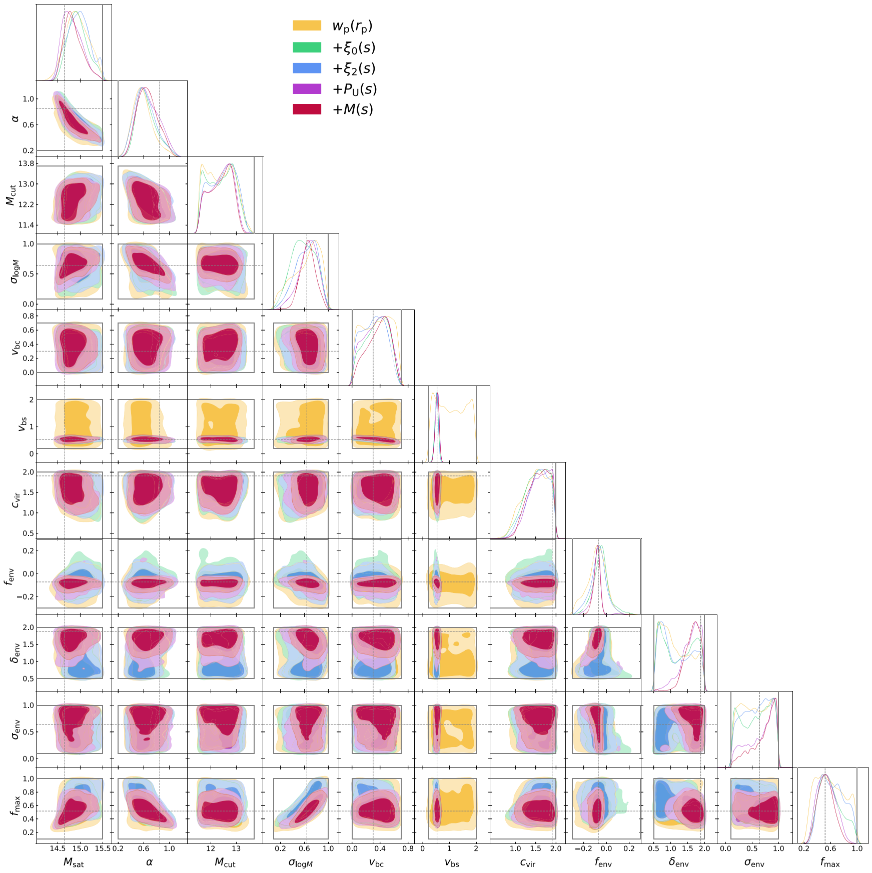
Figure 10. The precision of recovery tests when successively adding in observables, averaged over the 70 test models, for the HOD parameter $M_{\rm sat}$ and the three assembly bias parameters. Definitions are the same as in Figure 6(a).



(a)

(b)

Figure 11. Posteriors for all free parameters in our recovery test of a single cosmology+HOD model, when adding in observables successively. Contours are shown for (a) all cosmological parameters; (b) a mix of the key cosmological, HOD, and assembly bias parameters; and (c) all HOD and assembly bias parameters.



(c)

Figure 11. (Continued.)

ORCID iDs

Kate Storey-Fisher https://orcid.org/0000-0001-8764-7103
Jeremy L. Tinker https://orcid.org/0000-0003-3578-6149
Zhongxu Zhai https://orcid.org/0000-0001-7984-5476
Joseph DeRose https://orcid.org/0000-0002-0728-0960
Risa H. Wechsler https://orcid.org/0000-0003-2229-011X
Arka Banerjee https://orcid.org/0000-0002-5209-1173

References

Aghamousa, A., Aguilar, J., Ahlen, S., et al. 2016, arXiv:1611.00036
Alam, S., Ata, M., Bailey, S., et al. 2017, MNRAS, 470, 2617
Alcock, C., & Paczyński, B. 1979, Natur, 281, 358
Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W., & O'Neil, M. 2016, ITPAM, 38, 252
Angulo, R. E., & Pontzen, A. 2016, MNRAS Lett., 462, L1
Angulo, R. E., Zennaro, M., Contreras, S., et al. 2021, MNRAS, 507, 5869
Banerjee, A., & Abel, T. 2021, MNRAS, 500, 5479
Beltz-Mohrmann, G. D., Berlind, A. A., & Szewciw, A. O. 2020, MNRAS, 491, 5771
Berlind, A. A., & Weinberg, D. H. 2002, ApJ, 575, 587
Chapman, M. J., Mohammad, F. G., Zhai, Z., et al. 2022, MNRAS, 516, 617
Chuang, C. H., Yepes, G., Kitaura, F. S., et al. 2019, MNRAS, 487, 48
Colless, M., Peterson, B. A., Jackson, C., et al. 2003, arXiv:astro-ph/0306581
Conroy, C., Wechsler, R. H., & Kravtsov, A. V. 2006, ApJ, 647, 201
Cooray, A., & Sheth, R. K. 2002, PhR, 372, 1
Croton, D. J., Gao, L., & White, S. D. 2007, MNRAS, 374, 1303
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, 55
De Mattia, A., Ruhlmann-Kleider, V., Raichoor, A., et al. 2021, MNRAS, 501, 5616
DeRose, J., Chen, S. F., White, M., & Kokron, N. 2022, JCAP, 2022, 056
DeRose, J., Kokron, N., Banerjee, A., et al. 2023, JCAP, 2023, 054
DeRose, J., Wechsler, R. H., Tinker, J. L., et al. 2019, ApJ, 875, 69
eBOSS Collaboration, Alam, S., Aubert, M., et al. 2021, PhRvD, 103, 083533
Euclid Collaboration, Knabenhans, M., Stadel, J., et al. 2019, MNRAS, 484, 5509
Fletcher, R. R. 1987, Practical methods of optimization (Chichester, NY: Wiley)
Giblin, B., Cataneo, M., Moews, B., & Heymans, C. 2019, MNRAS, 490, 4826
Green, J., Schechter, P., Baltay, C., et al. 2012, arXiv:1208.4012
Guo, H., Yang, X., & Lu, Y. 2018, ApJ, 858, 30
Hartlap, J., Simon, P., & Schneider, P. 2007, A&A, 464, 399
Hearin, A. P., Campbell, D., Tollerud, E., et al. 2017, AJ, 154, 190
Hearin, A. P., Zentner, A. R., Van Den Bosch, F. C., Campbell, D., & Tollerud, E. 2016, MNRAS, 460, 2552
Heitmann, K., Higdon, D., White, M., et al. 2009, AJ, 705, 156
Heitmann, K., White, M., Wagner, C., Habib, S., & Higdon, D. 2010, AJ, 715, 104
Higson, E., Handley, W., Hobson, M., & Lasenby, A. 2019, S&C, 29, 891
Ho, M. F., Bird, S., & Shelton, C. R. 2022, MNRAS, 509, 2551
Hoshino, H., Leauthaud, A., Lackner, C., et al. 2015, MNRAS, 452, 998
Hoyle, F., & Vogeley, M. S. 2004, ApJ, 607, 751
Hunter, J. D. 2007, CSE, 9, 90
Ishiyama, T., Prada, F., Klypin, A. A., et al. 2021, MNRAS, 506, 4210
Joudaki, S., Hildebrandt, H., Traykova, D., et al. 2020, A&A, 638, L1
Klypin, A., & Prada, F. 2018, MNRAS, 478, 4602
Klypin, A. A., Trujillo-Gomez, S., & Primack, J. 2011, AJ, 740, 102
Knabenhans, M., Stadel, J., Potter, D., et al. 2021, MNRAS, 505, 2840
Kokron, N., DeRose, J., Chen, S. F., White, M., & Wechsler, R. H. 2021, MNRAS, 505, 1422
Kravtsov, A. V., Berlind, A. A., Wechsler, R. H., et al. 2004, ApJ, 609, 35
Kwan, J., Heitmann, K., Habib, S., et al. 2015, AJ, 810, 35
Landy, S. D., & Szalay, A. S. 1993, ApJ, 412, 64
Lange, J. U., Hearin, A. P., Leauthaud, A., et al. 2022, MNRAS, 509, 1779
Laureijs, R. J. 2011, arXiv:1110.3193
Lawrence, E., Heitmann, K., Kwan, J., et al. 2017, ApJ, 847, 50

Lawrence, E., Heitmann, K., White, M., et al. 2010, ApJ, 713, 1322

Leauthaud, A., Bundy, K., Saito, S., et al. 2016, MNRAS, 457, 4021

Leauthaud, A., Saito, S., Hilbert, S., et al. 2017, MNRAS, 467, 3024

Lehmann, B. V., Mao, Y. Y., Becker, M. R., Skillman, S. W., & Wechsler, R. H. 2016, ApJ, 834, 37

Lewis, A. 2019, arXiv:1910.13970

Macaulay, E., Wehus, I. K., & Eriksen, H. K. 2013, PhRvL, 111, 161301,

MacCrann, N., Zuntz, J., Bridle, S., Jain, B., & Becker, M. R. 2015, MNRAS, 451, 2877

Mandelbaum, R., Slosar, A., Baldauf, T., et al. 2013, MNRAS, 432, 1544

McBride, C. K., Connolly, A. J., Gardner, J. P., et al. 2011, ApJ, 739, 85

McClintock, T., Rozo, E., Becker, M. R., et al. 2019a, ApJ, 872, 53

McClintock, T., Rozo, E., Becker, M. R., et al. 2019b, arXiv:1907.13167

Miyatake, H., Sugiyama, S., Takada, M., et al. 2022, PhRvD, 106, 083520

Neveux, R., Burtin, E., Ruhlmann-Kleider, V., et al. 2022, MNRAS, 516, 1910

Nishimichi, T., Takada, M., Takahashi, R., et al. 2019, ApJ, 884, 29

Peebles, P. J. E., & Hauser, M. G. 1974, ApJS, 28, 19

Pellejero-Ibañez, M., Angulo, R. E., Aricó, G., et al. 2020, MNRAS, 499, 5257

Perez, F., & Granger, B. E. 2007, CSE, 9, 21

Philcox, O. H. E., Ivanov, M. M., Cabass, G., et al. 2022, PhRvD, 106, 043530

Rasmussen, C. E., & Williams, C. K. 2006, Gaussian Processes for Machine Learning (Cambridge, MA: MIT)

Reddick, R. M., Wechsler, R. H., Tinker, J. L., & Behroozi, P. S. 2013, ApJ, 771, 32

Reid, B. A., Seo, H. J., Leauthaud, A., Tinker, J. L., & White, M. 2014, MNRAS, 444, 476

Sánchez, A. G., Montesano, F., Kazin, E. A., et al. 2014, MNRAS, 440, 2692

Satpathy, S., Croft, R. A., Ho, S., & Li, B. 2019, MNRAS, 484, 2148

Seljak, U. 2000, MNRAS, 318, 203

Sheth, R. K., & Tormen, G. 2004, MNRAS, 350, 1385

Sinha, M., & Garrison, L. 2019, in Software Challenges to Exascale Computing, Second Workshop, SCEC 2018, CCIS 964, ed. A. Majumdar & R. Arora, 3

Sinha, M., & Garrison, L. H. 2020, MNRAS, 491, 3022

Skilling, J. 2006, BayAn, 1, 833

Speagle, J. S. 2020, MNRAS, 493, 3132

Spurio Mancini, A., Piras, D., Alsing, J., Joachimi, B., & Hobson, M. P. 2022, MNRAS, 511, 1771

Storey-Fisher, K. 2023a, kstoreyf/clust: Aemulus VI, v1.0.0, Zenodo, doi:10.5281/zenodo.8433094

Storey-Fisher, K. 2023b, kstoreyf/aemulator: Aemulus V, v1.0.0, Zenodo, doi:10.5281/zenodo.8433110

Szewciw, A. O., Beltz-Mohrmann, G. D., Berlind, A. A., & Sinha, M. 2022, ApJ, 926, 15

Takada, M., Ellis, R. S., Chiba, M., et al. 2014, PASJ, 66, R1

Takada, M., & Jain, B. 2003, MNRAS, 340, 580

Tinker, J. L., Conroy, C., Norberg, P., et al. 2008, ApJ, 686, 53

Tinker, J. L., Weinberg, D. H., & Warren, M. S. 2006, ApJ, 647, 737

Vale, A., & Ostriker, J. P. 2004, MNRAS, 353, 189

Van Der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, CSE, 13, 22

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261

Walsh, K., & Tinker, J. 2019, MNRAS, 488, 470

Wang, K., Mao, Y. Y., Zentner, A. R., et al. 2019, MNRAS, 488, 3541

Wechsler, R. H., & Tinker, J. L. 2018, ARA&A, 56, 435

Wechsler, R. H., Zentner, A. R., Bullock, J. S., Kravtsov, A. V., & Allgood, B. 2006, ApJ, 652, 71

White, M. 2016, JCAP, 2016, 057

White, M. J., & Padmanabhan, N. 2009, MNRAS, 395, 2381

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

Yuan, S., Garrison, L. H., Eisenstein, D. J., & Wechsler, R. H. 2022, MNRAS, 515, 871

Yuan, S., Hadzhiyska, B., Bose, S., Eisenstein, D. J., & Guo, H. 2021, MNRAS, 502, 3582

Zehavi, I., Blanton, M. R., Frieman, J. A., et al. 2002, ApJ, 571, 172

Zehavi, I., Contreras, S., Padilla, N., et al. 2018, ApJ, 853, 84

Zentner, A. R., Hearin, A. P., & van den Bosch, F. C. 2014, MNRAS, 443, 3044

Zhai, Z., Tinker, J. L., Banerjee, A., et al. 2023, ApJ, 948, 99

Zhai, Z., Tinker, J. L., Becker, M. R., et al. 2019, ApJ, 874, 95

Zhang, H., Samushia, L., Brooks, D., et al. 2022a, MNRAS, 515, 6133

Zhang, P., D'Amico, G., Senatore, L., Zhao, C., & Cai, Y. 2022b, JCAP, 2022, 036

Zheng, Z., Berlind, A. A., Weinberg, D. H., et al. 2005, ApJ, 633, 791