OPEN ACCESS



Emulating Power Spectra for Prereconstructed and Postreconstructed Galaxy Samples

Yuting Wang^{1,2}, Ruiyang Zhao^{1,3,4}, Zhongxu Zhai^{5,6,7,8}, Kazuya Koyama⁴, Will J. Percival^{7,8,9}, Hong Guo¹⁰, Yin Li¹¹, Gong-Bo Zhao^{1,2,3}, Takahiro Nishimichi^{12,13,14}, Héctor Gil-Marín¹⁵, Yonghao Feng^{1,3}, Hanyu Zhang^{7,8}, and

Yi Wu^{1,3}

¹National Astronomical Observatories, Chinese Academy of Sciences, Beijing, 100101, People's Republic of China; ytwang@nao.cas.cn

Institute for Frontiers in Astronomy and Astrophysics, Beijing Normal University, Beijing 102206, People's Republic of China

³ School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China

Institute of Cosmology and Gravitation, University of Portsmouth, Dennis Sciama Building, Portsmouth PO1 3FX, UK

⁵ Department of Astronomy, School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China ⁶ Shanghai Koy Labortory for Particle Physics and Cosmology, Shanghai 200240, People's Republic of China

Shanghai Key Laboratory for Particle Physics and Cosmology, Shanghai 200240, People's Republic of China ⁷ Waterloo Centre for Astrophysics, University of Waterloo, 200 University Ave. W, Waterloo, ON N2L 3G1, Canada

⁸ Department of Physics and Astronomy, University of Waterloo, 200 University Ave. W, Waterloo, ON N2L 3G1, Canada

Perimeter Institute for Theoretical Physics, 31 Caroline St. N, Waterloo, ON N2L 2Y5, Canada

¹⁰ Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, People's Republic of China

¹¹ Department of Mathematics and Theory, Peng Cheng Laboratory, Shenzhen, Guangdong 518066, People's Republic of China

¹² Department of Astrophysics and Atmospheric Sciences, Faculty of Science, Kyoto Sangyo University, Motoyama, Kamigamo, Kitaku, Kyoto 603-8555, Japan

¹³ Center for Gravitational Physics and Quantum Information, Yukawa Institute for Theoretical Physics, Kyoto University, Kyoto 606-8502, Japan

¹⁴ Kavli Institute for the Physics and Mathematics of the Universe (WPI), The University of Tokyo Institutes for Advanced Study (UTIAS), The University of Tokyo, Chiba 277-8583, Japan

¹⁵ ICC, University of Barcelona, IEEC-UB, Martí i Franquès, 1, E-08028 Barcelona, Spain

Received 2023 November 16; revised 2024 February 13; accepted 2024 February 23; published 2024 April 24

Abstract

The small-scale linear information in galaxy samples typically lost during nonlinear growth can be restored to a certain level by the density field reconstruction, which has been demonstrated for improving the precision of the baryon acoustic oscillation (BAO) measurements. As proposed in the literature, a joint analysis of the power spectrum before and after the reconstruction enables an efficient extraction of information carried by high-order statistics. However, the statistics of the postreconstruction density field are difficult to model. In this work, we circumvent this issue by developing an accurate emulator for the prereconstructed, postreconstructed, and cross-power spectra (P_{pre} , P_{post} , P_{cross}) up to $k = 0.5 h \text{ Mpc}^{-1}$ based on the DARK QUEST *N*-body simulations. The accuracy of the emulator is at the percent level; namely, the error of the emulated monopole and quadrupole of the power spectra is less than 1% and 10% of the ground truth, respectively. A fit to an example power spectrum using the emulator shows that the constraints on cosmological parameters get largely improved using $P_{\text{pre}} + P_{\text{post}} + P_{\text{cross}}$ with $k_{\text{max}} = 0.25 h \text{ Mpc}^{-1}$, compared to that derived from P_{pre} alone; namely, the constraints on $(\Omega_m, H_0, \sigma_8)$ are tightened by ~41%–55%, and the uncertainties of the derived BAO and RSD parameters ($\alpha_{\perp}, \alpha_{\parallel}, f\sigma_8$) shrink by ~28%–54%, respectively. This highlights the complementarity among P_{pre} , P_{post} , and P_{cross} , which demonstrates the efficiency and practicability of a joint P_{pre} , P_{post} , and P_{cross} analysis for cosmological implications.

Unified Astronomy Thesaurus concepts: Large-scale structure of the universe (902); Cosmological parameters (339); Cosmology (343)

1. Introduction

Wide-area spectroscopic surveys are fundamental tools for cosmological studies since they enable us to probe the Universe both geometrically and dynamically. In particular, the observed baryon acoustic oscillations (BAOs) and redshift-space distortions (RSDs), which are specific 3D clustering patterns of galaxies, can be used to reconstruct the cosmic expansion history and the growth rate of the cosmic structure. Over the last few decades, massive spectroscopic surveys, including the Sloan Digital Sky Survey (SDSS; York et al. 2000), the Twodegree Field Galaxy Redshift Survey (2dFGRS; Colless et al. 2001), WiggleZ (Drinkwater et al. 2010), the SDSS-III Baryon Oscillation Spectroscopic Survey (BOSS; Dawson et al. 2013), and the SDSS-IV extended Baryon Oscillation Spectroscopic Survey (eBOSS; Dawson et al. 2016) have proven to be

Original content from this work may be used under the terms (cc) of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

powerful probes for cosmology (Peacock et al. 2001; Cole et al. 2005; Eisenstein et al. 2005; Percival et al. 2007; Blake et al. 2011; Alam et al. 2017, 2021).

In Fourier space, the BAO feature manifests itself as a set of wiggles in the power spectrum, which can be used as a standard ruler to measure the cosmic expansion history. Unfortunately, the BAO feature is generally blurred by the nonlinear evolution of the cosmic structure, reducing its strength as a cosmic probe. To sharpen the BAO feature, the reconstruction scheme was proposed (Eisenstein et al. 2007), which effectively restores the linearity of the density field to a certain extent by partially undoing the nonlinear structure evolution. This process brings the high-order information dominated by the three- and fourpoint statistics back to two-point statistics (Schmittfull et al. 2015), such that it is not only useful for boosting the BAO signal but also helpful for a general full-shape analysis of the power spectrum (Hikage et al. 2020).

Recently, a novel method was proposed (Wang et al. 2022) to extract information carried by high-order statistics from a joint analysis of the power spectrum of the prereconstructed density field (P_{pre}), the postreconstructed field (P_{post}), and the cross-power spectrum between the pre- and postreconstructed fields (P_{cross}). Their analysis, based on the Fisher matrix method, showed that a joint analysis using P_{pre} , P_{post} , and P_{cross} can tighten the constraints on the cosmological parameters compared to that using P_{post} alone, as part of the information from three-point and four-point statistics of the density field can be efficiently extracted (Wang et al. 2022).

In order to exploit the information content from the galaxy clustering, an accurate model for the statistics of the density field before and after the reconstruction is required. Traditional methods for the model building rely on the perturbation theory (PT). For Ppre, PT-based models can work up to scales of k = 0.2 or $0.25 h \text{ Mpc}^{-1}$, depending on the effective redshift of the galaxy sample (Taruya et al. 2010; Carrasco et al. 2012; Beutler et al. 2014; d'Amico et al. 2020; Ivanov et al. 2020; Chen et al. 2022). However, it is much more challenging to build PT-based models that can work on the same scales for P_{post} or P_{cross} due to complexities brought in by the reconstruction process (Hikage et al. 2020). One alternative to building PT-based models is to develop simulation-based models, i.e., the emulators, which have been extensively studied and developed for statistics for the prereconstructed density fields (Wibking et al. 2019; Winther et al. 2019; Zhai et al. 2019; Kobayashi et al. 2020; Donald-McCann et al. 2022; Yuan et al. 2022; Cuesta-Lazaro et al. 2023a, 2023b; Kwan et al. 2023).

In this work, we develop an emulator for P_{pre} , P_{post} , and P_{cross} up to $k = 0.5 h \text{ Mpc}^{-1}$, which is trained using the DARK QUEST simulations (Nishimichi et al. 2019) and a halo occupation distribution (HOD) model (Zheng et al. 2007). Our emulator is then validated using simulations that are not used for the training. Using our emulator, we perform a likelihood analysis using the monopole and quadrupole of galaxy power spectra up to the scale of $k = 0.25 h \text{ Mpc}^{-1}$ and find a significant information gain by a joint { P_{pre} , P_{post} , P_{cross} } analysis, compared to using P_{pre} alone.

This paper is structured as follows. The next section is a description of the simulations and galaxy mocks used for the training and validation, and Section 3 presents the details for creating the emulator. In Section 4, we perform a likelihood analysis using various types of power spectra and show the main result of this work before concluding in Section 5.

2. The DARK QUEST Simulations and Galaxy Mocks

The DARK QUEST simulations that we use to develop our emulator are a suite of *N*-body simulations with 2048³ dark matter particles in a $2h^{-1}$ Gpc side-length box (Nishimichi et al. 2019). The emulator is built using a single DARK QUEST snapshot at z = 0.549. The cosmologies used in the DARK QUEST simulations cover the 100 spatially flat wCDM models¹⁶ with six variable parameters and one spatially flat Λ CDM model with the best-fit value of Planck 2015 (Planck Collaboration et al. 2016) presented in Table 1, where $\omega_b \equiv \Omega_b h^2$ and $\omega_c \equiv \Omega_c h^2$ are the physical density parameters of the baryon and cold dark matter, respectively. Ω_{de} is the dimensionless dark energy density parameter. A_s and n_s are the amplitude and slope of the primordial power spectrum, respectively. w is the equation-of-state parameter of dark

 Table 1

 Cosmological and HOD Parameters Used in Our Emulator

Parameter	Fiducial Value	Sampling Range		
ω_b	0.02225	[0.0211375, 0.0233625]		
ω_c	0.1198	[0.10782, 0.13178]		
Ω_{de}	0.6844	[0.54752, 0.82128]		
$\ln(10^{10}A_s)$	3.094	[2.4752, 3.7128]		
n _s	0.9645	[0.916275, 1.012725]		
w	-1	[-1.2, -0.8]		
$\sigma_{\log M}$	0.596	[0.05, 1.2]		
M_0/M_1	0.1194	[0.0, 0.4]		
α	1.0127	[0.2, 1.5]		
$M_1/M_{\rm min}$	8.1283	[5, 15]		

Note. The fiducial cosmological values are from Planck 2015 (Planck Collaboration et al. 2016). We take the fiducial HOD parameters based on the fitting to the CMASS galaxy sample (Manera et al. 2013). The sampling ranges represent the bounds on the emulator training set.

energy. In addition, the total neutrino mass is fixed to $\sum m_{\nu} = 0.06 \text{ eV}$. The effect of massive neutrinos was included in simulations at the level of linear transfer function. Cosmological parameters are sampled over the parameter range presented in Table 1 using optimal maximum distance sliced Latin hypercube designs (Ba et al. 2015) so that parameter samplings can cover the parameter space as uniformly as possible (Nishimichi et al. 2019). We have 15 realizations for the fiducial Λ CDM cosmology.

The halos were identified using the phase-space temporal friends-of-friends halo finder, ROCKSTAR (Behroozi et al. 2013). The center of each halo is given as the center-of-mass location of a subset of member particles in the inner part of that halo, i.e., "core particles," and the velocity of each halo is defined as the center-of-mass velocity of the core particles. $M_{200m} = (4\pi/3)200\bar{\rho}_{\rm m0}R_{200m}^2$ is used as the halo mass definition in DARK QUEST, where R_{200m} is the spherical halo boundary radius within which the mean mass density is 200 times the mean mass density today, $\bar{\rho}_{\rm m0}$. The direct outputs of ROCKSTAR contain both distinct "host" halos and substructures. For the subsequent analyses, we remove substructures, which are found within the R_{200m} of a more massive nearby halo.

Galaxy mock catalogs are constructed from the DARK QUEST halo catalogs using the HOD framework, which is implemented in Halotools (Hearin et al. 2017). We use the functional form of HOD as proposed in Zheng et al. (2007) to model the mean number $\langle N(M) \rangle$ of galaxies in halos of mass *M*. The mean occupation functions of central and satellite galaxies are parameterized as

$$\langle N_{\rm c}(M) \rangle = \frac{1}{2} \Biggl[1 + \operatorname{erf} \Biggl(\frac{\log M - \log M_{\min}}{\sigma_{\log M}} \Biggr) \Biggr],$$
 (1)

$$\langle N_{\rm s}(M) \rangle = \langle N_{\rm c}(M) \rangle \left(\frac{M - M_0}{M_1} \right)^{\alpha},$$
 (2)

and $\langle N_{\rm s}(M) \rangle = 0$ when $M < M_0$. $M_{\rm min}$ is the cutoff halo mass scale for hosting central galaxies, with $\langle N_{\rm c}(M_{\rm min}) \rangle = 0.5$. $\sigma_{\log M}$ describes the profile for the halo mass cutoff, making $\langle N_{\rm c}(M) \rangle$ smoothly transit from 0 to 1. M_0 is the minimum halo mass to host satellite galaxies. M_1 is the normalization mass scale. α is

 $[\]frac{16}{16}$ These 100 cosmological simulations are generated using different random number seeds (Nishimichi et al. 2019).

the power-law slope of the satellite HOD at the massive end. The occupations of central and satellite galaxies are drawn from Bernoulli and Poisson distributions, respectively. Central galaxies are placed at the halo centers with the same velocities as their host halos, where we have ignored the effect of galaxy velocity bias (Guo et al. 2015, 2016). We also assume that the satellite galaxy distribution within the halos follows the Navarro–Frenk–White profile (Navarro et al. 1996).

We adopt the fiducial values of HOD parameters based on the best-fit values (log $M_{\min} = 13.09$, $\sigma_{\log M} = 0.596$, log $M_0 = 13.077$, log $M_1 = 14.0$, and $\alpha = 1.0127$) obtained by fitting to the CMASS ("constant mass") galaxy sample (Manera et al. 2013). The number density *n* can be derived by performing an integral over the mass function,

$$n = \int dM \frac{dn}{dM} (M) \langle N(M) \rangle, \qquad (3)$$

where dn/dM(M) is the halo mass function. We take its fitting formula from Tinker et al. (2008). The resulting HOD catalog has a number density of $n = 5.6 \times 10^{-4} h^3 \,\mathrm{Mpc}^{-3}$. In our work, we choose to fix the number density,¹⁷ then we sample four out of the five HOD parameters of our model. Here we reparameterize the HOD parameters as $\{\sigma_{\log M}, M_0/M_1, \alpha, M_1/M_{\min}\}$ as used in Wibking et al. (2020). Their fiducial values and flat prior ranges are presented in Table 1. We utilize the (randomized) quasi-Monte Carlo method to sample reparameterized HOD parameters in the prior range. Specifically, we generate 2450 points in 4D using the Sobol sequence (Sobol' 1967) utility in the scipy.stats.gmc package (Virtanen et al. 2020). We scramble the Sobol sequence with a random seed searched among integers from 0 to 65535 to minimize the mixture discrepancy (Zhou et al. 2013) as the uniformity measure. The first 2400 HOD samples are assigned to 80 cosmologies for training; i.e., each training cosmology is assigned 30 HODs. The remaining 50 HODs are assigned to each testing cosmology, yielding a testing set of 1000 models. For each sampling, we use Equation (3) to find the value of $\log M_{\min}$ that yields the fixed n. We include shifts due to RSD along the z-axis to generate the simulated galaxy samples in the redshift space; i.e., the simulated galaxies in the real space are shifted by $v_z/(aH)$, where v_{τ} is the peculiar velocity of the simulated galaxies along the line of sight and *a* is the scale factor.

3. Emulating Pre- and Postreconstructed Galaxy Power Spectra

In this section, we use the galaxy samples described in the previous section to emulate the P_{pre} , P_{post} , and P_{cross} of galaxies. We first present the measurement of power spectra with and without the density field reconstruction, then detail the training process of our emulator, and finally discuss the performance of the emulator.

3.1. The Density Field Reconstruction and the Power Spectrum Measurement

Before performing the density field reconstruction, we implement the Alcock–Paczynski (AP) effect (Alcock & Paczynski 1979), which arises from the discrepancy between the fiducial cosmology used for redshift–distance conversion and the underlying true cosmology. Although the equation relating the power spectrum before and after applying the AP effect is analytically known (Ballinger et al. 1996), including this effect in the reconstruction is complicated and requires nontrivial modeling (Sherwin & White 2019). An easier way to account for the AP effect is to manipulate the catalog by changing the coordinates of the samples. Specifically, we convert the galaxy positions in the true coordinate x' to the "observed" coordinate x and stretch the size lengths of simulation box L using the relations of $x = A^{-1}x'$ and $L \rightarrow A^{-1}L$ with

$$\boldsymbol{A} = \begin{pmatrix} \alpha_{\perp} & 0 & 0\\ 0 & \alpha_{\perp} & 0\\ 0 & 0 & \alpha_{\parallel} \end{pmatrix}, \ \alpha_{\perp} \equiv \frac{D_A(z)}{D_{A,\text{fid}}(z)}, \ \alpha_{\parallel} \equiv \frac{H_{\text{fid}}(z)}{H(z)}, \qquad (4)$$

where D_A and H are the comoving angular diameter distance and Hubble parameter, and quantities with subscript "fid" denote those in the fiducial cosmology. The galaxy density field is smoothed by convolving with the kernel $K(k) = \exp\left[-(k\Sigma_s)^2/2\right]$ in Fourier space, where k is the modulus of the conjugate wavenumber k of the observed coordinate x, and we set the smoothing scale to be $\Sigma_s = 10 h^{-1}$ Mpc, which is close to the optimal smoothing scale for the reconstruction efficiency (Seo et al. 2016; Vargas-Magaña et al. 2017). The displacement field is then estimated using the Zeldovich approximation, i.e.,, $\tilde{s}(\mathbf{k}) = -i\frac{\mathbf{k}}{k^2} \frac{\delta(\mathbf{k})}{b + f\mu^2} K(k)$, where $\delta(\mathbf{k})$ denotes the nonlinear redshift-space galaxy overdensity in the observed coordinate, b_{in} is the input linear bias of the galaxy sample, and f_{in} is the input logarithmic growth rate. An inverse Fourier transformation on \tilde{s} returns the configuration-space shift field s(x), which is used to move both galaxies and randoms. Although it is natural to use the true (fiducial) values of b and f as $b_{\rm in}$ and $f_{\rm in}$ for the reconstruction, this does not have to be the choice. Actually, the true values of b and f are not known before performing the analysis. As we shall demonstrate later, the final parameter estimation is largely insensitive to the choice of b_{in} and $f_{\rm in}$. In what follows, we use the fiducial b and f to start with and repeat the analysis with a significantly different set of $b_{\rm in}$ and $f_{\rm in}$ to demonstrate the robustness of the final result against the choice of these input parameters.

We measure the multipoles of $P_{\rm pre}$, $P_{\rm post}$, and $P_{\rm cross}$ using a fast-Fourier-transform-based estimator (Hand et al. 2017) implemented in nbodykit (Hand et al. 2018). The number density field of galaxies is constructed using the cloud-in-cell scheme to assign galaxies to the grid, and we correct for the aliasing effect using the interlacing scheme in Sefusatti et al. (2016). For the monopole of the auto power spectrum before and after density field reconstruction, the shot noise is removed as a constant. Note that the shot noise of $P_{\rm cross}$ is scale-dependent (Wang et al. 2022), which is estimated using the "half-sum half-difference" approach and then subtracted off, as in Ando et al. (2018) and Wang et al. (2022). The *k*-bin width is set to be $\Delta k = 0.01 h \,{\rm Mpc}^{-1}$ for all P(k) measurements.

¹⁷ Another way is allowing the number density *n* to vary, then including the information in *n* by adding a Gaussian prior for *n* into the likelihood (Donald-McCann et al. 2022; Lange et al. 2022). Varying *n* would weaken the constraints on the HOD parameters to some extent, depending on the uncertainty on *n*, but has a negligible impact on cosmological parameters (Donald-McCann et al. 2022).

THE ASTROPHYSICAL JOURNAL, 966:35 (15pp), 2024 May 1

3.2. Emulating the Power Spectra

In order to avoid the emulated quantities spanning several orders of magnitude, we choose to normalize the power spectrum multipoles using the linear Kaiser power spectrum (Kaiser 1987) with the BAO feature removed in the fiducial cosmology, i.e.,

$$R_0^{\rm X} = \frac{P_0^{\rm X}}{(b^2 + 2/3bf + 1/5f^2)P_{\rm nw,lin}},\tag{5}$$

$$R_2^{\rm X} = \frac{P_2^{\rm X}}{(4/3bf + 4/7f^2)P_{\rm nw,lin}},\tag{6}$$

$$R_4^{\rm X} = \frac{P_4^{\rm X}}{(8/15f^2)P_{\rm nw,lin}},\tag{7}$$

where $P_{nw,lin}$ is the linear power spectrum without the BAO feature (Eisenstein & Hu 1998). The superscript "X" runs for "{pre, post, cross}." To well capture the BAO wiggles in the monopole, we decompose R_0^X into two parts, i.e., the smoothed broadband shape (S) part and the BAO wiggles (W) part. The S part is obtained by applying a Savitzky–Golay filter (Savitzky & Golay 1964) to R_0^X , i.e., fitting to a certain number of data points (N) with a polynomial of pth order, and we find that N = 41 and p = 4 is a reasonable choice for the filtering. Then the BAO wiggles are extracted, i.e., $W_0^X = R_0^X - S_0^X$. Figure 6 in the Appendix shows the observables (2400 in total) used for training the emulator.

We follow Zhai et al. (2019, 2023) to construct the emulator, based on the George (Ambikasaran et al. 2016) code. In the Gaussian process (GP) modeling, the correlation between different training data points is modeled by a covariance matrix generated by a kernel function. This is of critical importance in the GP modeling since it defines the function we wish to learn. Due to the lack of prior knowledge of the correlation between training data points, the definition of the kernel function can be arbitrary. For the modeling of the galaxy power spectrum in this work, we adopt a Matern class kernel (\mathcal{K}), as we find that it produces sufficiently accurate predictions. In this model, the hyperparameters in the kernel define the strength of the correlation between neighboring points. The following process of training is to optimize the hyperparameters in the kernel function:

$$\ln \mathcal{L} = -\frac{1}{2} \mathbf{P}^T M^{-1} \mathbf{P} - \frac{1}{2} \ln|M| - \frac{1}{2} N \ln 2\pi, \qquad (8)$$

where $M = \mathcal{K} + \sigma^2 I$, \mathcal{K} is the covariance matrix populated by the kernel function, and σ represents the error of the training data P. Since each cosmology in the training data has only one realization, we estimate the uncertainty of the training data using the fiducial cosmology with 15 realizations. We find that 15 simulations gives a good approximation of σ , but more simulations might improve the emulator, which would be interesting to check in the future. With the optimized hyperparameters fed into the kernel function, we can obtain the power spectra for an arbitrary point in the parameter space.

3.3. Covariance Matrix

The DARK QUEST simulations only have 15 realizations in the fiducial cosmology, which is insufficient to construct a robust covariance matrix for galaxy clustering analysis. We therefore compute the correlation matrix using GLAM simulations, for which there are 986 independent realizations in the Planck cosmology¹⁸ (Klypin & Prada 2018). We adopt the best-fit HOD parameters for the $M_i < -21.6$ CMASS samples in Guo et al. (2014), leading to a contribution of shot noise ($\sim 2 \times 10^{-4} h^3$ Mpc⁻³) to the covariance. The side length of the GLAM simulation box is $1 h^{-1}$ Gpc. Simulations with a larger box size can help investigate the effect of supersample covariance (SSC) (Bayer et al. 2023). Limited by the simulation suite, the covariance determination in this work neglects supersample variance components.¹⁹ To be close to the volume of the BOSS survey (Alam et al. 2017), the data covariance matrix C_{data} is rescaled by a factor of 3. Specifically, we derive the C_{data} from GLAM mocks, i.e.,

$$(C_{i,j}^{\ell,\ell'})_{\text{data}} = \frac{1}{N_s - 1} \sum_{n=1}^{N_s} [P_\ell^n(k_i) - \overline{P_\ell}(k_i)] \times [P_{\ell'}^n(k_j) - \overline{P_{\ell'}}(k_j)], \qquad (9)$$

where the mean of the power spectra is defined as

$$\overline{P_{\ell}}(k_i) = \frac{1}{N_s} \sum_{n=1}^{N_s} P_{\ell}^n(k_i), \qquad (10)$$

and $N_s = 986$ is the number of mocks. Note that the effect of the error induced by the estimation of the covariance matrix from the finite number of mocks will be corrected as described later in this work.

3.4. Emulator Validation

In Figure 1, we show the prediction from our emulator for multipole moments of P_{pre} , P_{post} , and P_{cross} for the fiducial cosmology that is not used for the training. The symbols in the upper panels are the average of the power spectra measured from 15 realizations in the fiducial cosmology. The error bars are the statistical errors computed using Equation (9).

The lower panels of Figure 1 show the fractional difference between the emulated and the measured power spectra from the galaxy mocks. It indicates that the monopole and quadrupole measured from the galaxy mocks can be well described by our emulator by better than 1%-2% for the monopole and 2%-10%for the quadrupole at most scales. The quadrupole error can sometimes be >10\%, particularly around the scales where the quadrupole happens to cross 0. For the hexadecapole, the fractional difference is noisy because the amplitudes of the hexadecapole are close to 0. Within the statistical errors, our emulator gives an excellent prediction for the hexadecapole as well.

We quantify the accuracy of our emulator using 1000 test galaxy mocks that are not used in the training set. The three columns of Figure 2 from left to right show the performance of our emulator for $P_{\rm pre}$ (left), $P_{\rm post}$ (middle), and $P_{\rm cross}$ (right).

 $[\]frac{18}{18}$ The cosmological parameters of the GLAM simulations are from Planck 2013 (Ade et al. 2014), which is slightly different from the fiducial cosmology used in the DARK QUEST simulations. The minor difference between them is ignored in this work.

¹⁹ Hikage et al. (2020) found that the improvement of the error on the growth rate by the reconstruction was comparable between the cases with and without the SSC effect. It would be interesting to explore the impact of the SSC on cosmological parameters in future work.



Figure 1. Upper panels: the prediction from our emulator for multipole moments of P_{pre} , P_{post} , and P_{cross} for the fiducial cosmology that is not used for the training. The symbols are the average of 15 realizations in the fiducial cosmology. The errors are the statistical errors for a volume of 3 h^{-3} Gpc³. Lower panels: the fractional difference between the emulator prediction and the measured power spectra from mocks in the fiducial cosmology.



Figure 2. Upper panels: the average of the fractional difference between the emulated and the measured monopole power spectrum from 1000 testing mocks. The black solid lines show the inverse signal-to-noise ratio of the mean fiducial monopole measurement. The statistical error for the monopole power spectrum σ_{P_0} is computed using Equation (9). Middle panels: the average of the difference between the emulated and the measured quadrupole power spectrum relative to the statistical error. Lower panels: the average of the difference between the emulated and the measured power spectrum relative to the statistical error.

The symbols in the upper panels of Figure 2 show the average fractional error of the monopole power spectrum obtained by comparing the emulator predictions with the P_0 measurements from 1000 test mocks. The error bars are the standard deviation estimated from 1000 test mocks. The fractional error is within ~1%-2% over most scales. The solid lines show the inverse signal-to-noise ratio computed using the

average of the monopole measurements among 15 realizations from the fiducial cosmology.

Because the quadrupole and hexadecapole moments can cross 0, leading to large fractional errors, we instead show the absolute error between the emulator prediction and the measurement from the testing mocks relative to the statistical error in the middle and lower panels of Figure 2. We find that the emulator error for P_2 and P_4 is subdominant, roughly 50%–70% of the statistical error for a volume of 3 h^{-3} Gpc³.

4. Cosmological Application to Mock Catalogs

In this section, we test our emulator by applying it to the power spectrum measurements from mock galaxy catalogs in the fiducial cosmology, which are not in the training set. We use Cobaya (Torrado & Lewis 2021) to perform a Markov Chain Monte Carlo sampling of the 9D parameter space within the flat Λ CDM framework; i.e., the *w* parameter is fixed to -1. The following χ^2 gets minimized in the fitting,

$$\chi^2 \equiv [P_{\rm emu}(k) - P_{\rm mea}(k)]^T C^{-1} [P_{\rm emu}(k) - P_{\rm mea}(k)], \quad (11)$$

and we add a Gaussian prior for ω_b centered on 0.0223 with the width 0.00036 from Big Bang Nucleosynthesis (BBN) constraints (Mossa et al. 2020) and a Gaussian prior for n_s parameters centered on 0.965 with the width 0.0042 from Planck constraints (Aghanim et al. 2020). For other parameters, a flat prior over the range shown in Table 1 is used. Here P_{emu} is the prediction of the emulator, and P_{mea} denotes the average of the power spectra measured from 15 realizations in the fiducial cosmology. *C* is the covariance matrix consisting of two terms,

$$C = C_{\text{data}} + \sigma_{\text{emu}}^2 I, \qquad (12)$$

where C_{data} (computed in Section 3.3) is the contribution of the sample statistics, and σ_{emu} corresponds to the uncertainty due to the emulating error in the model prediction. Since the emulator is constructed for individual scale bins, here we assume that the emulating error is independent among different scale bins, which is computed using the testing set as discussed in Section 3.4. Since the covariance matrix C_{data} is estimated from a finite number of mocks, C_{data} is generally biased. To correct, we multiply C_{data} by a factor of \mathcal{P} (Percival et al. 2022),

$$\mathcal{P} = \frac{(N_s - 1)[1 + B(N_d - N_\theta)]}{N_s - N_d + N_\theta - 1},$$
(13)

with

$$B = \frac{N_s - N_d - 2}{(N_s - N_d - 1)(N_s - N_d - 4)}.$$
 (14)

Here, N_s is the number of simulations used to estimate the covariance, N_d is the number of the data vector, and N_{θ} is the number of parameters that are being fitted. Note that the \mathcal{P} -factor generally dilutes the constraints on the parameters being fitting by rescaling the covariance; namely, when using P_{pre} , P_{post} , or P_{cross} alone, the \mathcal{P} -factor increases the covariance by 9%. For joint analyses of $P_{\rm pre} + P_{\rm post}$ and $P_{\rm pre} + P_{\rm post} +$ P_{cross} (i.e., P_{all}), the \mathcal{P} -factor enlarges the covariance by 22% and 36%, respectively. These $\mathcal{O}(10\%)$ enlargements of the posteriors are a first-order way to correct for the lack of convergence of the covariance due to using a small number of simulations. Running more simulations would reduce the size of this correction and provide more robust contours. We plan to do this in future work. When a covariance matrix is constructed using Equation (9), we have, in effect, drawn the matrix as a random variable from a Wishart distribution. Using this



Figure 3. The correlation matrix for the monopoles and quadrupoles of power spectra (P_{pre} , P_{post} , P_{cross}).

knowledge, we can consider the set of covariance matrices that could have been drawn and determine the average effects on results from them. It is often found that the results are biased because the values of interest are skewed by the errors in the covariance. Examples of such effects include a skewed inverse covariance matrix (Hartlap et al. 2007) or skewed parameter errors (Dodelson & Schneider 2013). The effects can be modeled using PT, leading to correction terms such as those in Equation (13). The PT-based derivation assumes a linear model, but this has been shown to work well for typical cosmological problems (Percival et al. 2022). Nevertheless, this is only a first-order correction in terms of the link from the likelihood to model parameters, and having more mocks will always be better. The correlation matrix being estimated for the combined power spectra ($P_{\rm nre}$, $P_{\rm cross}$) is shown in Figure 3.

combined power spectra (P_{pre} , P_{post} , P_{cross}) is shown in Figure 3. Using k modes at $k \le 0.25 h \text{ Mpc}^{-1}$ for both the monopole and quadrupole of the power spectra,²⁰ we obtain the 1D posterior distributions and 2D contour plots for the derived cosmological parameters Ω_m , H_0 , and σ_8 , as shown in Figure 4. The mean values with 68% credible intervals of Ω_m , H_0 , and σ_8 are presented in Table 2. The left contour plot in Figure 4 shows a comparison of the fitting results using the prereconstructed power spectrum alone (gray), the postreconstructed power spectrum alone (red), the cross-power spectrum alone (green), and the joint fitting of prereconstructed, postreconstructed, and cross-power spectra (blue) for $k_{\text{max}} =$ $0.25 h \text{ Mpc}^{-1}$. Figure 4 shows that our emulator-based analysis can recover the expected values of cosmological parameters within statistical errors. The postreconstructed power spectrum alone is more informative, tightening the constraints on Ω_m , H_0 , and σ_8 by 10.9%, 35.7%, and 23.7%, respectively, compared to that from P_{pre} alone. It is found that the joint fit of the prereconstructed, postreconstructed, and cross-power spectra, denoted as P_{all} , gives the tightest constraint; namely, the constraints on Ω_m , H_0 , and σ_8 from P_{all} are improved by 44.5%, 41.7%, and 55.3%, respectively, compared to that from $P_{\rm pre}$ alone.

²⁰ Unless otherwise mentioned, k_{max} is 0.25 h Mpc⁻¹ as a default setting for the analysis in this work.

THE ASTROPHYSICAL JOURNAL, 966:35 (15pp), 2024 May 1



Figure 4. Left: the 1D posterior distribution and 2D contour plots showing 68% and 95% credible regions for Ω_m , H_0 , and σ_8 using the prereconstructed power spectrum alone (gray), the postreconstructed power spectrum alone (red), the cross-power spectrum alone (green), and the joint result of the prereconstructed, postreconstructed, and cross-power spectra (blue). Right: the same plot derived from the postreconstructed power spectrum alone with two choices of $k_{\text{max}} = 0.25 \text{ h Mpc}^{-1}$ (red) and $k_{\text{max}} = 0.5 \text{ h Mpc}^{-1}$ (blue).

 Table 2

 The Constraints on Derived Cosmological Parameters (Ω_m , H_0 , σ_8) and BAO and RSD Parameters (α_{\perp} , α_{\parallel} , $f\sigma_8$) Using Different Data Sets

	$k_{\rm max} = 0.25$						
$b_{\rm in}, f_{\rm in}$			$(b_{\rm fid})$	$(0.9 b_{\rm fid}, 0.7 f_{\rm fid})$	(b_{6d}, f_{6d})		
	$P_{\rm pre}$	P _{post}	P _{cross}	$P_{\rm pre+post}$	P _{all}	$P_{\rm all}$	P_{post}
P-fa-							
ctor	1.09	1.09	1.09	1.22	1.36	1.36	1.21
Ω_m	0.318 ± 0.0110	0.315 ± 0.0098	0.317 ± 0.0100	0.318 ± 0.0082	0.320 ± 0.0061	0.320 ± 0.0080	0.317 ± 0.0078
H_0	67.13 ± 0.84	67.00 ± 0.54	67.05 ± 0.69	67.15 ± 0.52	67.06 ± 0.49	67.12 ± 0.49	67.02 ± 0.57
σ_8	0.849 ± 0.038	0.833 ± 0.029	0.842 ± 0.040	0.834 ± 0.018	0.834 ± 0.017	0.834 ± 0.018	0.831 ± 0.016
α_{\perp}	1.0008 ± 0.0115	1.0039 ± 0.0088	1.0022 ± 0.0097	1.0004 ± 0.0083	1.0013 ± 0.0076	1.0003 ± 0.0081	1.0028 ± 0.0090
$\alpha_{ }$	1.0000 ± 0.0118	1.0041 ± 0.0106	1.0016 ± 0.0104	0.9996 ± 0.0097	0.9999 ± 0.0084	0.9987 ± 0.0098	1.0024 ± 0.0103
$f\sigma_8$	0.497 ± 0.023	0.487 ± 0.018	0.492 ± 0.024	0.488 ± 0.0116	0.488 ± 0.0104	0.488 ± 0.0114	0.486 ± 0.010

Note. The fiducial values of the parameters derived are $\Omega_m = 0.3156$, $H_0 = 67.24$, $\sigma_8 = 0.831$, $\alpha_{\perp} = 1$, $\alpha_{\parallel} = 1$, and $f\sigma_8(z = 0.549) = 0.485$, which are well recovered in all cases. The factor \mathcal{P} here is calculated using Equation (13). Our default choices of (b, f) parameters for reconstruction are $b_{\text{fid}} = 1.824$ and $f_{\text{fid}} = 0.778$ determined in the fiducial cosmology. To explore the effect of these inputs, we vary the bias by -10% (i.e., $0.9 b_{\text{fid}}$) and the f by -30% (i.e., $0.7 f_{\text{fid}}$).

The relative information gain from $P_{\rm all}$ compared to that from $P_{\rm pre}$ is expected to be greater in the nonlinear regime, i.e., including modes with $k > 0.25 h \,\mathrm{Mpc}^{-1}$. However, given the number of mocks and data points, we do not go further than $k_{\rm max} = 0.25 h \,\mathrm{Mpc}^{-1}$ for a $P_{\rm all}$ analysis. Instead, we perform a $P_{\rm post}$ -alone analysis with $k_{\rm max} = 0.5 h \,\mathrm{Mpc}^{-1}$ for a demonstration. We compare the constraints on Ω_m , H_0 , and σ_8 using $P_{\rm post}$ alone for $k_{\rm max} = 0.25$ and $0.5 h \,\mathrm{Mpc}^{-1}$ in the right panel of Figure 4. As shown, adding modes on smaller scales helps to constrain σ_8 ; namely, its uncertainty gets reduced by 44.8% as $k_{\rm max}$ increases from 0.25 to 0.5 $h \,\mathrm{Mpc}^{-1}$. The galaxy clustering on smaller scales is more sensitive to the amplitude-related parameter σ_8 compared to Ω_m and H_0 . In addition, we perform the analysis for different $k_{\rm max}$ using $P_{\rm post}$ alone. The fitting results as a function of $k_{\rm max}$ are presented in Figure 7 in the Appendix. Also, adding more modes does not generate bias in the posteriors, demonstrating the robustness of our emulator.

We then derive the BAO and RSD parameters $(\alpha_{\perp}, \alpha_{\parallel}, f\sigma_8)$ and show the 1D posterior distributions and 2D contour plots in Figure 5, with the mean values and 68% credible intervals of the BAO and RSD parameters listed in Table 2. Compared to P_{pre} alone, the constraints on the $(\alpha_{\perp}, \alpha_{\parallel}, f\sigma_8)$ parameters from P_{all} are improved by 33.9%, 28.8%, and 54.8%, respectively. P_{post} alone gives a tighter constraint than that using P_{pre} only but is outnumbered by P_{all} by 13.6% for α_{\perp} , 20.8% for α_{\parallel} , and 42.2% for $f\sigma_8$. The right panel of Figure 5 shows the contours of the derived BAO and RSD parameters with two different choices of k_{max} , as in Figure 4. As expected, adding small-scale modes ($k \in [0.25, 0.5] h \text{ Mpc}^{-1}$) helps to tighten the constraint on $f\sigma_8$ significantly; namely, the uncertainty gets reduced by



Figure 5. Left: the 1D posterior distribution and 2D contour plots showing 68% and 95% credible regions for the derived α_{\perp} , $\alpha_{||}$, and $f\sigma_8$ parameters using the pereconstructed power spectrum alone (gray), the postreconstructed power spectrum alone (red), the cross-power spectrum alone (green), and the joint result of the prereconstructed, postreconstructed, and cross-power spectra (blue). Right: the result using the postreconstructed power spectrum alone with two choices of $k_{\text{max}} = 0.25 h \text{ Mpc}^{-1}$ (red) and $k_{\text{max}} = 0.5 h \text{ Mpc}^{-1}$ (blue).

44.4%. Note that this level of constraint can be achieved by using P_{all} with $k_{\text{max}} = 0.25 h \text{ Mpc}^{-1}$.

We confirm that the information content in $P_{\rm cross}$ is complementary to that in $P_{\rm pre}$ and $P_{\rm post}$, as claimed in Wang et al. (2022). Specifically, adding $P_{\rm cross}$ to our joint analysis using $P_{\rm pre}$ and $P_{\rm post}$ improves the constraints on $(\Omega_m, H_0, \sigma_8)$ and $(\alpha_{\perp}, \alpha_{\parallel}, f\sigma_8)$ by 5.5%–25.6%, as presented in Table 2.

Since the BAO reconstruction process requires a pair of input b and f, denoted as b_{in} and f_{in} , the reconstructed power spectrum depends on b_{in} and f_{in} . One natural question is whether and how much the final posterior depends on b_{in} and $f_{\rm in}$. To investigate, we use a set of $b_{\rm in}$ and $f_{\rm in}$ that are significantly different from the fiducial b and f, namely, $b_{\rm in} = 0.9 b_{\rm fid}$ and $f_{\rm in} = 0.7 f_{\rm fid}$. Note that this level of deviation from the true values is much greater than that constrained by a typical galaxy survey such as BOSS (Beutler et al. 2017) and thus is sufficient to study the impact of using the "wrong" cosmological parameters for the reconstruction on the final result (Sherwin & White 2019). We repeat our analysis using this set of b_{in} and f_{in} and show the parameter constraint from P_{all} in this case in Table 2 and in Figure 8 in the Appendix. As shown, the constraint is largely unchanged, demonstrating the robustness of our method against the choice of b_{in} and f_{in} .

For completeness, we show the full contour plot for all parameters, including the cosmological and HOD parameters, in Figure 9 in the Appendix using different combinations of the power spectra. As expected, $P_{\rm all}$ provides the tightest constraint for all parameters, as predicted by the Fisher matrix analysis (Wang et al. 2022).

The results presented so far do not include information from P_4 , the hexadecapole, so it is useful to explore how P_4 can help to reduce the uncertainties. We perform an additional analysis using P_{all} including P_4 for all types of power spectra with $k_{\text{max}} = 0.25 h \text{ Mpc}^{-1}$ and find that P_4 can barely further improve the constraints on the cosmological parameters, as

shown in Figure 10 in the Appendix, because the hexadecapole has a relatively lower signal-to-noise ratio compared to P_0 and P_2 . A similar conclusion is found when only using P_{pre} , and the contour plot is presented in Figure 11.

5. Conclusion and Discussions

In this work, we develop an emulator for galaxy power spectra for catalogs with and without the BAO reconstruction based on the DARK QUEST simulations with HOD models to populate galaxies. The theoretical predictions of power spectra derived from our emulator are in excellent agreement with the ground truth (with a deviation of less than 10%). Our emulator-based likelihood analysis of mock galaxy catalogs demonstrates that input cosmological parameters can be accurately recovered from power spectra up to scales of k = 0.5 h Mpc⁻¹.

Our analysis shows that P_{pre} , P_{post} , and P_{cross} are highly complementary; thus, jointly using these power spectra can significantly improve constraints on cosmological parameters, which is consistent with the claim based on a Fisher matrix analysis (Wang et al. 2022). Specifically, the uncertainty of $(\Omega_m, H_0, \sigma_8)$ derived from $P_{\text{pre}} + P_{\text{post}} + P_{\text{cross}}$ gets tightened by 44.5%, 41.7%, and 55.3%, respectively, compared to that derived from P_{pre} ($k_{\text{max}} = 0.25 \text{ h} \text{ Mpc}^{-1}$ in all cases). The derived BAO and RSD parameters, α_{\perp} , α_{\parallel} , and $f\sigma_8$, are better determined by 33.9%, 28.8%, and 54.8%, respectively. Adding small-scale modes to the analysis helps to constrain parameters related to the amplitude of the power spectra. For example, extending $k_{\text{max}} = 0.25$ to $0.5 \, h \, \text{Mpc}^{-1}$ for P_{post} reduces the uncertainty on σ_8 and $f\sigma_8$ by 44.8% and 44.4%, respectively. We also find that the posteriors of the parameters are largely insensitive to input values of b and f, which are required for the BAO reconstruction process.

The methodology and pipeline developed in this work make it possible to extract high-order information from two-point statistics, which is of significance for cosmological studies. Our method and emulator can be directly applied to existing and forthcoming galaxy surveys including BOSS (Dawson et al. 2013), eBOSS (Dawson et al. 2016), the Dark Energy Spectroscopic Instrument (Aghamousa et al. 2016a, 2016b), the Prime Focus Spectrograph (Takada et al. 2014), and so forth after the required tuning in the emulation process for the number density, effective redshifts of the galaxy samples, etc., which is technically straightforward.

Acknowledgments

We thank Hee-Jong Seo for helpful discussions. Y.W. is supported by the National Key R&D Program of China (2022YFF0503404, 2023YFA1607800, 2023YFA1607803); NSFC grants 12273048, 11890691, and 11720101004; the CAS Project for Young Scientists in Basic Research (No. YSBR-092); and the Youth Innovation Promotion Association CAS. R.Z., Y.F., Y.W., and G.B.Z. are supported by NSFC grants 11925303 and 11890691. R.Z. is also supported by the Chinese Scholarship Council (CSC) and the University of Portsmouth. Z.Z. is supported by NSFC(12373003) and acknowledges the generous sponsorship from the Yangyang Development Fund. K.K. is supported by the STFC grant ST/ W001225/1. H.G. is supported by the National SKA Program of China (No. 2020SKA0110100), NSFC grants 11922305 and 11833005, and the science research grants from the China Manned Space Project with No. CMS-CSST-2021-A02. G.B. Z. is also supported by the CAS Project for Young Scientists in Basic Research (No. YSBR-092), the science research grants from the China Manned Space Project with No. CMS-CSST-2021-B01, and the New Cornerstone Science Foundation through the XPLORER prize. T.N. is supported by MEXT/ JSPS KAKENHI grant Nos. JP19H00677, JP20H05861, JP21H01081, and JP22K03634. Numeric work was performed on the UK Sciama High Performance Computing cluster supported by the ICG, University of Portsmouth. Y.W. thanks the ICG, University of Portsmouth, for their hospitality.

Appendix

This appendix includes six figures (Figures 6-11), with information detailed in the figure captions.



Figure 6. The complete training set for our emulator, which consists of 2400 power spectrum multipoles for P_{pre} (left column), P_{post} (middle), and P_{cross} (right column). All spectra have been properly normalized by power spectra derived from the linear Kaiser formula, so that their amplitudes are within a narrow range. The normalized monopole, R_0^X , is divided into a smoothed shape part (S_0^X) and a BAO "wiggles" part (W_0^X). More details are presented in the main text and Equation (5).



Figure 7. The mean and standard deviation of $(\Omega_m, H_0, \sigma_8)$ parameters normalized by the fiducial values as a function of k_{max} using P_{post} alone.



Figure 8. The 1D posterior distribution and 2D contour plots showing 68% and 95% credible regions for the derived parameters (Ω_m , H_0 , σ_8 , α_{\perp} , α_{\parallel} , $f\sigma_8$) from P_{all} reconstructed using two different sets of b_{in} and f_{in} shown in the legend. The dashed lines show the expected values of the parameters.



Figure 9. The 1D posterior distribution and 2D contour plots showing 68% and 95% credible regions for cosmological and HOD parameters derived from combinations of different types of power spectra shown in the legend. The dashed lines show the expected values of the parameters.



Figure 10. The 1D posterior distribution and 2D contour plots showing 68% and 95% credible regions for $(\Omega_m, H_0, \sigma_8, \alpha_{\perp}, \alpha_{\parallel}, f\sigma_8)$ using P_{all} with and without the hexadecapole. The dashed lines show the expected values of the parameters.



Figure 11. Same as Figure 10 but using P_{pre} alone.

ORCID iDs

Yuting Wang (b) https://orcid.org/0000-0001-7756-8479 Ruiyang Zhao (b) https://orcid.org/0000-0002-7284-7265 Zhongxu Zhai (b) https://orcid.org/0000-0001-7984-5476 Will J. Percival (b) https://orcid.org/0000-0002-0644-5727

Hong Guo [®] https://orcid.org/0000-0003-4936-8247 Yin Li [®] https://orcid.org/0000-0002-0701-1410 Gong-Bo Zhao [®] https://orcid.org/0000-0003-4726-6714 Takahiro Nishimichi [®] https://orcid.org/0000-0002-9664-0760

References

Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., et al. 2014, A&A, 571, A16 Aghamousa, A., Aguilar, J., Ahlen, S., et al. 2016a, arXiv:1611.00036 Aghamousa, A., Aguilar, J., Ahlen, S., et al. 2016b, arXiv:1611.00037 Aghanim, N., Akrami, Y., Ashdown, M., et al. 2020, A&A, 641, A6 Alam, S., Ata, M., Bailey, S., et al. 2017, MNRAS, 470, 2617 Alam, S., Aubert, M., Avila, S., et al. 2021, PhRvD, 103, 083533

Alcock, C., & Paczynski, B. 1979, Natur, 281, 358

Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W., & O'Neil, M. 2016, ITPAM, 38, 252

Ando, S., Benoit-Lévy, A., & Komatsu, E. 2018, MNRAS, 473, 4318
 Ba, S., Myers, W. R., & Brenneman, W. A. 2015, Technometrics, 57, 479
 Ballinger, W. E., Peacock, J. A., & Heavens, A. F. 1996, MNRAS, 282, 877

- Bayer, A. E., Liu, J., Terasawa, R., et al. 2023, PhRvD, 108, 043521
- Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013, ApJ, 762, 109
- Beutler, F., Saito, S., Seo, H.-J., et al. 2014, MNRAS, 443, 1065
- Beutler, F., Seo, H.-J., Saito, S., et al. 2017, MNRAS, 466, 2242
- Blake, C., Kazin, E. A., Beutler, F., et al. 2011, MNRAS, 418, 1707
- Carrasco, J. J. M., Hertzberg, M. P., & Senatore, L. 2012, JHEP, 2012, 82 Chen, S.-F., Vlah, Z., & White, M. 2022, JCAP, 02, 008
- Cole, S., Percival, W. J., Peacock, J. A., et al. 2005, MNRAS, 362, 505
- Colless, M., Dalton, G., Maddox, S., et al. 2001, MNRAS, 328, 1039
- Cuesta-Lazaro, C., Nishimichi, T., Kobayashi, Y., et al. 2023a, MNRAS, 523, 3219
- Cuesta-Lazaro, C., Paillas, E., Yuan, S., et al. 2023b, arXiv:2309.16539
- d'Amico, G., Gleyzes, J., Kokron, N., et al. 2020, JCAP, 2020, 005
- Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, AJ, 151, 44
- Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, AJ, 145, 10
- Dodelson, S., & Schneider, M. D. 2013, PhRvD, 88, 063537
- Donald-McCann, J., Beutler, F., Koyama, K., & Karamanis, M. 2022, MNRAS, 511, 3768
- Drinkwater, M. J., Jurek, R. J., Blake, C., et al. 2010, MNRAS, 401, 1429 Eisenstein, D. J., & Hu, W. 1998, ApJ, 496, 605
- Eisenstein, D. J., Seo, H.-J., Sirko, E., & Spergel, D. N. 2007, ApJ, 664, 675
- Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al. 2005, ApJ, 633, 560
- Guo, H., Zheng, Z., Behroozi, P. S., et al. 2016, MNRAS, 459, 3040
- Guo, H., Zheng, Z., Zehavi, I., et al. 2014, MNRAS, 441, 2398
- Guo, H., Zheng, Z., Zehavi, I., et al. 2015, MNRAS, 453, 4368
- Hand, N., Feng, Y., Beutler, F., et al. 2018, AJ, 156, 160
- Hand, N., Li, Y., Slepian, Z., & Seljak, U. 2017, JCAP, 2017, 002
- Hartlap, J., Simon, P., & Schneider, P. 2007, A&A, 464, 399
- Hearin, A. P., Campbell, D., Tollerud, E., et al. 2017, AJ, 154, 190
- Hikage, C., Koyama, K., & Takahashi, R. 2020, PhRvD, 101, 043510
- Hikage, C., Takahashi, R., & Koyama, K. 2020, PhRvD, 102, 083514
- Ivanov, M. M., Simonović, M., & Zaldarriaga, M. 2020, PhRvD, 101, 083504
- Kaiser, N. 1987, MNRAS, 227, 1
- Klypin, A., & Prada, F. 2018, MNRAS, 478, 4602
- Kobayashi, Y., Nishimichi, T., Takada, M., Takahashi, R., & Osato, K. 2020, PhRvD, 102, 063504
- Kwan, J., Saito, S., Leauthaud, A., et al. 2023, ApJ, 952, 80
- Lange, J. U., Hearin, A. P., Leauthaud, A., et al. 2022, MNRAS, 509, 1779

- Manera, M., Scoccimarro, R., Percival, W. J., et al. 2013, MNRAS, 428, 1036
- Mossa, V., Stöckel, K., Cavanna, F., et al. 2020, Natur, 587, 210
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563
- Nishimichi, T., Takada, M., Takahashi, R., et al. 2019, ApJ, 884, 29
- Peacock, J. A., Cole, S., Norberg, P., et al. 2001, Natur, 410, 169
- Percival, W. J., Friedrich, O., Sellentin, E., & Heavens, A. 2022, MNRAS, 510, 3207
- Percival, W. J., Nichol, R. C., Eisenstein, D. J., et al. 2007, ApJ, 657, 645
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13
- Savitzky, A., & Golay, M. J. E. 1964, AnaCh, 36, 1627
- Schmittfull, M., Feng, Y., Beutler, F., Sherwin, B., & Chu, M. Y. 2015, PhRvD, 92, 123522
- Sefusatti, E., Crocce, M., Scoccimarro, R., & Couchman, H. M. P. 2016, MNRAS, 460, 3624
- Seo, H.-J., Beutler, F., Ross, A. J., & Saito, S. 2016, MNRAS, 460, 2453
- Sherwin, B. D., & White, M. 2019, JCAP, 02, 027
- Sobol', I. 1967, USSR CMMPh, 7, 86
- Takada, M., Ellis, R. S., Chiba, M., et al. 2014, PASJ, 66, R1
- Taruya, A., Nishimichi, T., & Saito, S. 2010, PhRvD, 82, 063522
- Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, ApJ, 688, 709
- Torrado, J., & Lewis, A. 2021, JCAP, 2021, 057
- Vargas-Magaña, M., Ho, S., Fromenteau, S., & Cuesta, A. J. 2017, MNRAS, 467, 2331
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261
- Wang, Y., Zhao, G.-B., Koyama, K., et al. 2022, arXiv:2202.05248
- Wibking, B. D., Salcedo, A. N., Weinberg, D. H., et al. 2019, MNRAS, 484, 989
- Wibking, B. D., Weinberg, D. H., Salcedo, A. N., et al. 2020, MNRAS, 492, 2872
- Winther, H. A., Casas, S., Baldi, M., et al. 2019, PhRvD, 100, 123540
- York, D. G., Adelman, J., Anderson, J. E. J., et al. 2000, AJ, 120, 1579
- Yuan, S., Garrison, L. H., Eisenstein, D. J., & Wechsler, R. H. 2022, MNRAS, 515, 871
- 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
- Zheng, Z., Coil, A. L., & Zehavi, I. 2007, ApJ, 667, 760
- Zhou, Y.-D., Fang, K.-T., & Ning, J.-H. 2013, J. Complex., 29, 283