Emulating power spectra for pre- and post-reconstructed galaxy samples

Authors: Yuting Wang, Ruiyang Zhao, Zhongxu Zhai, Kazuya Koyama, Will J. Percival, Guo Hong, Yin Li, Gong-Bo Zhao, Takahiro Nishimichi, Héctor Gil-Marín, Yonghao Feng, Hanyu Zhang, Yi Wu

Date: 2023-12-25T00:00:00+00:00

Abstract

The small-scale linear information in galaxy samples typically lost during non-linear 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 oscillations (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 post#2;reconstruction density field are difficult to model. In this work, we circumvent this issue by developing an accurate emulator for the pre-reconstructed, post-reconstructed, and cross power spectra (Ppre, Ppost, Pcross) up to k = 0.5 h Mpc−1 based on the Dark Quest N-body simulations. The accuracy of the emulator is at percent level, namely, the error of the emulated monopole and quadrupole of the power spectra is less than 1% and 5% of the ground truth, respectively. A fit to an example power spectra using the emulator shows that the constraints on cosmological parameters get largely improved using Ppre+Ppost+Pcross with kmax = 0.25 h Mpc−1 , compared to that derived from Ppre alone, namely, the constraints on (Ωm, H0, \$ 8)$aretightenedby \sim 41$ \$, α , f\$ \$8) shrink by 28% − 54%, respectively. This highlights the complementarity among Ppre, Ppost and Pcross, which demonstrates the efficiency and practicability of a joint Ppre, Ppost and Pcross analysis for cosmological implications.

Full Text

Preamble

Draft version November 13, 2023
Typeset using LATEX twocolumn style in AASTeX62

Emulating power spectra for pre- and post-reconstructed galaxy samples

Yuting Wang,[1,2] Ruiyang Zhao,[1,3,4] Zhongxu Zhai,[5,6,7,8] Kazuya Koyama,[4] Will J. Percival,[7,8,9] Hong Guo,[10] Yin Li,[11] Gong-Bo Zhao,[1,3,2] Takahiro Nishimichi,[12,13,14] Héctor Gil-Marín,[15] Yonghao Feng,[1,3] Hanyu Zhang,[7,8] and Yi Wu[1,3]

[1]National Astronomical Observatories, Chinese Academy of Sciences, Beijing, 100101, P.R.China
[2]Institute for Frontiers in Astronomy and Astrophysics, Beijing Normal University, Beijing 102206, China
[3]School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, P.R.China
[4]Institute of Cosmology and Gravitation, University of Portsmouth, Dennis Sciama Building, Portsmouth PO1 3FX, United Kingdom
[5]Department of Astronomy, School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai 200240, China
[6]Shanghai Key Laboratory for Particle Physics and Cosmology, Shanghai 200240, 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
[9]Perimeter Institute for Theoretical Physics, 31 Caroline St. North, Waterloo, ON, N2L 2Y5, Canada
[10]Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China
[11]Department of Mathematics and Theory, Peng Cheng Laboratory, Shenzhen, Guangdong 518066, 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

Abstract

The small-scale linear information in galaxy samples typically lost during non-linear growth can be restored to a certain level by density field reconstruction, which has been demonstrated to improve the precision of baryon acoustic oscillation (BAO) measurements. As proposed in the literature, a joint analysis of the power spectrum before and after reconstruction enables efficient extraction of information carried by high-order statistics. However, the statistics of the post-reconstruction density field are difficult to model. In this work, we circumvent this issue by developing an accurate emulator for the pre-reconstructed, post-reconstructed, and cross power spectra ($P_{\rm pre}$, $P_{\rm post}$, $P_{\rm cross}$) up to $k = 0.5\,h\,{\rm Mpc}^{-1}$ based on the Dark Quest N-body simulations. The accuracy of the emulator is at the percent level: the error of the emulated monopole and quadrupole of the power spectra is less than 1% and 5% of the ground truth, respectively. A fit to example power spectra using the emulator shows that constraints on cosmological parameters are substantially improved using $P_{\rm pre} + P_{\rm post} + P_{\rm cross}$ with $k_{\rm max} = 0.25\,h\,{\rm Mpc}^{-1}$ compared to those derived from $P_{\rm pre}$ alone. Specifically, the constraints on $(\Omega_m, H_0, \sigma_8)$ are tightened by 55%, and the uncertainties of the derived BAO and RSD parameters $(\alpha_\perp, \alpha_\parallel, f\sigma_8)$ shrink by 54%, respectively. This highlights the complementarity among $P_{\rm pre}$, $P_{\rm post}$, and $P_{\rm cross}$, demonstrating the efficiency and practicability of a joint $P_{\rm pre}$, $P_{\rm post}$, and $P_{\rm cross}$ analysis for cosmological inference.

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 (BAO) and redshift-space distortions (RSD), which are specific three-dimensional clustering patterns of galaxies, can be used to reconstruct the cosmic expansion history and the growth rate of cosmic structure.

Over the last few decades, massive spectroscopic surveys, including the Sloan Digital Sky Survey (SDSS) (York et al. 2000), the Two-Degree-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 powerful probes for cosmology (Peacock et al. 2001; Eisenstein et al. 2005; Cole 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 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 nonlinear structure evolution. This process brings information from high-order statistics back to two-point statistics, making it not only useful for boosting the BAO signal but also helpful for 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 pre-reconstructed density field ($P_{\rm pre}$), the post-reconstructed field ($P_{\rm post}$), and the cross-power spectrum between pre- and post-reconstructed fields ($P_{\rm cross}$). Their analysis, based on the Fisher matrix method, showed that a joint analysis using $P_{\rm pre}$, $P_{\rm post}$, and $P_{\rm cross}$ can tighten constraints on cosmological parameters compared to using $P_{\rm post}$ alone, as part of the information from 3-point and 4-point statistics of the density field can be efficiently extracted (Wang et al. 2022).

To exploit the information content from galaxy clustering, an accurate model for the statistics of the density field before and after reconstruction is required. Traditional methods for model building rely on perturbation theory (PT). For $P_{\rm pre}$, PT-based models can work up to scales of $k = 0.2$ or $0.25\,h\,{\rm Mpc}^{-1}$, 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 work on the same scales for $P_{\rm post}$ or $P_{\rm cross}$, due to complexities introduced by the reconstruction process (Hikage et al. 2020). One alternative to building PT-based models is to develop simulation-based models, i.e., emulators, which have been extensively studied and developed for statistics of pre-reconstructed density fields (Zhai et al. 2019; Wibking et al. 2019; Kobayashi et al. 2020; Yuan et al. 2022; Winther et al. 2019; Donald-McCann et al. 2022; Cuesta-Lazaro et al. 2023; Kwan et al. 2023).

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

This paper is structured as follows: the next section describes the simulations and galaxy mocks used for training and validation, Section 3 presents the details of creating the emulator, Section 4 performs a likelihood analysis using various types of power spectra and shows the main result of this work, and we conclude in Section 5.

2. The Dark Quest Simulations and Galaxy Mocks

The Dark Quest simulations used to develop our emulator are a suite of N-body simulations with 2048^3 dark matter particles in a $2\,h^{-1}\,{\rm Gpc}$ 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 100 spatially flat wCDM models with six variable parameters and one spatially flat ΛCDM model with the best-fit values from Planck 2015 (Planck Collaboration et al. 2016) presented in Table 1 , where ω_b and $\Omega_c h^2$ are the physical density parameters of baryons 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 energy. In addition, the total neutrino mass is fixed to $\sum m_\nu = 0.06\,\mathrm{eV}$. The effect of massive neutrinos was included in simulations at the level of the 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 cover the parameter space as uniformly as possible (Nishimichi et al. 2019). We have 15 realizations for the fiducial ΛCDM cosmology.

Halos were identified using the phase-space 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. $M_{200m} = (4\pi/3)200\bar{\rho}_{m0}R_{200m}^3$ 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 $\bar{\rho}_{m0}$. The direct outputs of Rockstar contain both distinct "host" halos and substructures. For subsequent analyses, we remove substructures found within R_{200m} of a more massive nearby halo.

Galaxy mock catalogs are constructed from the Dark Quest halo catalogs using the halo occupation distribution (HOD) framework, 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 of galaxies in halos of mass M. The mean occupation functions of central and satellite galaxies are parameterized as:

$$N_c(M) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\log M - \log M_{\mathrm{min}}}{\sigma_{\log M}}\right)\right]$$

$$N_s(M) = N_c(M)\left(\frac{M - M_0}{M_1}\right)^{\alpha}$$

where $N_s(M) = 0$ when $M < M_0$. M_{min} is the cutoff halo mass scale for hosting central galaxies, with $\sigma_{\log M}$ describing the profile for $N_c(M_{\mathrm{min}})$, making the halo mass cutoff transition smoothly from 0 to 1. M_0 is the minimum halo mass to host satellite galaxies. M_1 is the normalization mass scale. α is 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 halo centers with the same velocities as their host halos, ignoring the effect of galaxy velocity bias (Guo et al. 2015; Guo et al. 2016). We assume that the satellite galaxy distribution within halos follows the Navarro-Frenk-White profile (Navarro et al. 1996).

We adopt fiducial HOD parameter values based on the best-fit values ($\log M_{\mathrm{min}} = 13.09$, $\sigma_{\log M} = 0.596$, $\log M_0 = 13.077$, $\log M_1 = 14.0$, 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:

$$n = \int dM \frac{dn}{dM}(M)N(M)$$

where $dn/dM(M)$ is the halo mass function. We use its fitting formula from Tinker et al. (2008). The resulting HOD catalog has a number density of $n = 10^{-4}\,h^3\,\mathrm{Mpc}^{-3}$. In our work, we choose to fix the number density, then sample four of the five HOD parameters of our model. Here we re-parameterize HOD parameters as used in Wibking et al. (2020). Their fiducial values and flat prior ranges are presented in Table 1.

We utilize a (randomized) quasi-Monte Carlo method to sample re-parameterized HOD parameters in the prior range. Specifically, we generate 2450 points in 4D using the Sobol sequence (Sobol' 1967) utility in the scipy.stats.qmc 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 the integral equation to find the value of $\log M_{\mathrm{min}}$ that yields the fixed number density.

3. Emulating Pre- and Post-Reconstructed Galaxy Power Spectra

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

3.1. The Density Field Reconstruction and Power Spectrum Measurement

Before performing 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 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 galaxy positions from true coordinates $\mathbf{x}'$ to "observed" coordinates $\mathbf{x}$ and stretch the simulation box side length L using the relations $\mathbf{x} = A^{-1}\mathbf{x}'$ and $L \to A^{-1}L$ with

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

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 $K(k) = \exp[-(k\Sigma_s)^2/2]$ in Fourier space, where k is the modulus of the conjugate wavenumber $\mathbf{k}$ of the observed coordinate $\mathbf{x}$, and we set the smoothing scale to $\Sigma_s = 10\,h^{-1}\,\mathrm{Mpc}$. The displacement field is then estimated using the Zeldovich approximation:

$$\tilde{\mathbf{s}}(\mathbf{k}) = \frac{i\mathbf{k}}{k^2}\frac{\delta(\mathbf{k})}{b_{\mathrm{in}} + f_{\mathrm{in}}\mu^2}K(k)$$

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 transform on $\tilde{\mathbf{s}}$ returns the configuration-space shift field $\mathbf{s}(\mathbf{x})$, 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_{in} and f_{in} for reconstruction, this does not have to be the choice. In practice, the true values of b and f are not known before performing the analysis. As we demonstrate later, the final parameter estimation is largely insensitive to the choice of b_{in} and f_{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_{in} and f_{in} to demonstrate the robustness of the final result against the choice of these input parameters.

We measure the multipoles of P_{pre}, P_{post}, and P_{cross} using a fast Fourier transform (FFT)-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 (CIC) scheme to assign galaxies to the grid, and we correct for aliasing effects using the interlacing scheme from 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_{cross} is scale-dependent (Wang et al. 2022), which is estimated using the "half-sum (HS) half-difference (HD)" approach and then subtracted, as in Ando et al. (2018) and Wang et al. (2022). The k-bin width is set to $\Delta k = 0.01\,h\,\mathrm{Mpc}^{-1}$ for all $P(k)$ measurements.

3.2. Emulating the Power Spectra

To avoid emulated quantities spanning several orders of magnitude, we normalize the power spectrum multipoles using the linear Kaiser power spectrum (Kaiser 1987) with the BAO feature removed:

$$R_\ell^X(k) \equiv \frac{P_\ell^X(k)}{P_{\mathrm{nw},\ell}^{\mathrm{Kaiser}}(k)}$$

where X runs for "pre", "post", "cross", and

$$P_{\mathrm{nw},\ell}^{\mathrm{Kaiser}}(k) = \begin{cases} (b^2 + \frac{2}{3}bf + \frac{1}{5}f^2)P_{\mathrm{nw,lin}}(k) & (\ell = 0) \\ (\frac{4}{3}bf + \frac{4}{7}f^2)P_{\mathrm{nw,lin}}(k) & (\ell = 2) \\ (\frac{8}{35}f^2)P_{\mathrm{nw,lin}}(k) & (\ell = 4) \end{cases}$$

with $P_{\mathrm{nw,lin}}$ being the linear power spectrum without the BAO feature (Eisenstein & Hu 1998). To well capture the BAO wiggles in the monopole, we decompose R_0^X into two parts: 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, fitting to a certain number of data points (N) with a polynomial of p-th order. We find that $N = 41$ and $p = 4$ is a reasonable choice for the filtering. Then the BAO wiggles are extracted as $W_0^X = R_0^X - S_0^X$. Figure 5 [Figure 5: see original paper] 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 code (Ambikasaran et al. 2016). In Gaussian Process (GP) modeling, the correlation between different training data points is modeled by a covariance matrix generated by a kernel function. This is critically important in GP modeling since it defines the function we wish to learn. Due to the lack of prior knowledge about the correlation between training data points, the definition of the kernel function can be arbitrary. It turns out that for modeling the galaxy power spectrum in this work, a Matern class kernel is a proper choice for accurate predictions.

In this model, the hyperparameters in the kernel define the strength of correlation between neighboring points. The training process optimizes these hyperparameters by maximizing the log-likelihood, which involves terms like $\mathbf{P}^T M^{-1}\mathbf{P}$ and $N \ln 2\pi$, where M is the covariance matrix populated by the kernel function plus $\sigma^2 I$, and σ represents the error of the training data $\mathbf{P}$. 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. With the optimized hyperparameters fed into the kernel function, we can obtain power spectra for arbitrary points in the parameter space.

3.3. Emulator Validation

In Figure 1 [Figure 1: see original paper], we show predictions from our emulator for multipole moments of P_{pre}, P_{post}, and P_{cross} for the fiducial cosmology that was not used for training. The symbols in the upper panels are the average of power spectra measured from 15 realizations in the fiducial cosmology. The error bars are the statistical errors computed using Equation 11.

The lower panels of Figure 1 [Figure 1: see original paper] show the fractional difference between the emulated and measured power spectra from galaxy mocks. The monopole and quadrupole measured from the galaxy mocks can be well described by our emulator to better than 1-2% for the monopole and 2-5% for the quadrupole at most scales. For the hexadecapole, the fractional difference is noisy because the amplitudes are close to zero. Within the statistical errors, our emulator gives an excellent prediction for the hexadecapole as well.

We quantify the accuracy of our emulator by comparing emulator predictions with power spectrum multipoles measured from mocks in the testing set, which includes 20 cosmologies not used in the training set, with each cosmology assigned 50 HODs. Thus the testing set has 1000 measurements for each type of power spectrum.

Figure 2 [Figure 2: see original paper] shows the performance of our emulator for P_{pre} (left), P_{post} (middle), and P_{cross} (right). The symbols in the upper panels show the average fractional errors of the monopole power spectrum obtained by comparing emulator predictions with measurements from 1000 testing 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 15 fiducial monopole measurements. Since quadrupole and hexadecapole moments can cross zero, we instead show the difference between emulator predictions and measurements from testing mocks relative to the statistical error in the middle and lower panels. We find that the emulator error is subdominant, roughly 50-70% of the statistical error for a volume of $3\,h^{-3}\,\mathrm{Gpc}^3$.

4. Cosmological Application to Mock Catalogs

In this section, we test our emulator by applying it to 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 Markov chain Monte Carlo (MCMC) sampling of the 9-dimensional parameter space within the flat ΛCDM framework (i.e., the w parameter is fixed to -1). The following χ^2 is minimized in the fitting:

$$\chi^2 = [\mathbf{P}_{\mathrm{emu}}(k) - \mathbf{P}_{\mathrm{mea}}(k)]^T C^{-1} [\mathbf{P}_{\mathrm{emu}}(k) - \mathbf{P}_{\mathrm{mea}}(k)]$$

We add a Gaussian prior for ω_b centered on 0.0223 with width 0.00036 from BBN constraints (Mossa et al. 2020) and a Gaussian prior for n_s centered on

0.965 with width 0.0042 from Planck constraints (Aghanim et al. 2020). Here $\mathbf{P}_{\rm emu}$ is the emulator prediction, and $\mathbf{P}_{\rm mea}$ denotes the average of power spectra measured from 15 realizations in the fiducial cosmology. C is the covariance matrix consisting of two terms: $C = C_{\rm data} + \sigma^2_{\rm emu} I$, where $C_{\rm data}$ is the contribution from sample statistics and $\sigma_{\rm emu}$ corresponds to uncertainty due to emulation error in the model prediction. Since the emulator is constructed for individual scale bins, we assume that emulation error is independent among different scale bins, computed using the testing set as discussed in Section 3.3.

As the Dark Quest simulations only have 15 realizations in the fiducial cosmology, which are insufficient to construct a robust covariance matrix for galaxy clustering analysis, we compute the correlation matrix using GLAM simulations (Klypin & Prada 2018), which have 986 independent realizations. We adopt the best-fit HOD parameters for the CMASS samples from Guo et al. (2014), leading to a shot noise contribution to the covariance. The side length of the GLAM simulation box is $1\,h^{-1}\,\rm Gpc$. To approximate the volume of the BOSS survey (Alam et al. 2017), the data covariance matrix $C_{\rm data}$ is rescaled by a factor of 3. Specifically, we derive $C_{\rm data}$ from GLAM mocks:

$$(C^{ij}_{\ell,\ell'})_{\rm data} = \frac{1}{N_s - 1}\sum_{n=1}^{N_s}[P^n_\ell(k_i) - \bar{P}_\ell(k_i)][P^n_{\ell'}(k_j) - \bar{P}_{\ell'}(k_j)]$$

where the mean power spectra are defined as $\bar{P}_\ell(k_i) = \frac{1}{N_s}\sum_{n=1}^{N_s} P^n_\ell(k_i)$ and $N_s = 986$ is the number of mocks. Since $C_{\rm data}$ is estimated from a finite number of mocks, it is generally biased. To correct this, we multiply $C_{\rm data}$ by a factor (Percival et al. 2022):

$$\frac{(N_s - 1)}{(N_s - N_d - N_\theta - 1)}\frac{(N_s - N_d + N_\theta - 1)}{(N_s - N_d + N_\theta - 2)}$$

where N_s is the number of simulations used to estimate the covariance, N_d is the number of data vector elements, and N_θ is the number of parameters being fitted. The correction factor generally dilutes the constraints on parameters.

Using k modes at $k \leq 0.25\,h\,\rm Mpc^{-1}$ for both monopole and quadrupole of the power spectra, we obtain 1D posterior distributions and 2D contour plots for derived cosmological parameters Ω_m, H_0, and σ_8, as shown in Figure 3 [Figure 3: see original paper]. The mean values with 68% credible intervals are presented in Table 2 . The left panel of Figure 3 shows a comparison of fitting results using the pre-reconstructed power spectrum alone (grey), post-reconstructed power spectrum alone (red), and joint fitting of pre-, post-, and cross-power spectra (blue) for $k_{\rm max} = 0.25\,h\,\rm Mpc^{-1}$. Our emulator-based analysis recovers the expected values of cosmological parameters within statistical errors. The post-reconstructed power spectrum alone is more informative, tightening constraints on Ω_m, H_0, and σ_8 by 10.9%, 35.7%, and 23.7%, respectively, compared to P_{pre} alone. The joint fit of pre-, post-, and cross-power spectra, denoted P_{all}, provides the tightest constraints: uncertainties on Ω_m, H_0, and σ_8 from P_{all} are reduced by 44.5%, 41.7%, and 55.3%, respectively, compared to P_{pre} alone.

The relative information gain from P_{all} compared to P_{pre} is expected to be greater as we include more modes on smaller scales. However, given the number of mocks and data points, we do not extend beyond $k_{\text{max}} = 0.25\,h\,\text{Mpc}^{-1}$ for a P_{all} analysis. Instead, we perform a P_{post}-alone analysis with $k_{\text{max}} = 0.5\,h\,\text{Mpc}^{-1}$ for demonstration. The right panel of Figure 3 compares constraints on Ω_m, H_0, and σ_8 using P_{post} alone for $k_{\text{max}} = 0.25$ and $0.5\,h\,\text{Mpc}^{-1}$. Adding modes on smaller scales helps constrain σ_8, reducing its uncertainty by 44.8% as k_{max} increases from 0.25 to $0.5\,h\,\text{Mpc}^{-1}$. Adding more modes does not generate bias in the posteriors, demonstrating the robustness of our emulator.

We then derive BAO and RSD parameters ($\alpha_\perp$, $\alpha_\parallel$, $f\sigma_8$) and show 1D posterior distributions and 2D contour plots in Figure 4 [Figure 4: see original paper], with mean values and 68% credible intervals listed in Table 2 . Compared to P_{pre} alone, constraints on ($\alpha_\perp$, $\alpha_\parallel$, $f\sigma_8$) from P_{all} are improved by 33.9%, 28.8%, and 54.8%, respectively.

P_{post} alone gives tighter constraints than P_{pre} only but is outperformed by P_{all} by 13.6% for $\alpha_\perp$, 20.8% for $\alpha_\parallel$, and 42.2% for $f\sigma_8$. The right panel of Figure 4 shows contours of derived BAO and RSD parameters with two different choices of k_{max}, as in Figure 3. As expected, adding small-scale modes ($[0.25, 0.5]\,h\,\text{Mpc}^{-1}$) helps tighten the constraint on $f\sigma_8$ significantly, reducing its uncertainty by 44.4%. This level of constraint can be achieved using P_{all} with $k_{\text{max}} = 0.25\,h\,\text{Mpc}^{-1}$.

We confirm that the information content in P_{cross} is complementary to that in P_{pre} and P_{post}, as claimed in Wang et al. (2022). Specifically, adding P_{cross} to our joint analysis using P_{pre} and P_{post} improves constraints on (Ω_m, H_0, σ_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 input values of b and f, denoted b_{in} and f_{in}, the reconstructed power spectrum depends on these choices. One natural question is whether and how much the final posterior depends on b_{in} and f_{in}. To investigate, we use a set of b_{in} and f_{in} significantly different from the fiducial values: $b_{\text{in}} = 0.9\,b_{\text{fid}}$ and $f_{\text{in}} = 0.7\,f_{\text{fid}}$. This level of deviation from the true values is much greater than that constrained by typical galaxy surveys such as BOSS (Beutler et al. 2017), thus sufficient to study the impact of using "wrong" cosmological parameters for 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 Figure 6 [Figure 6: see original paper] in the Appendix. 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 cosmological and HOD parameters, in Figure 7 [Figure 7: see original paper] in the Appendix using different combinations of 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).

Results presented so far do not include information from P_4 (the hexadecapole), so it is useful to explore how P_4 can help reduce uncertainties. We perform an additional analysis using $P_{\rm all}$ including P_4 for all types of power spectra with $k_{\rm max} = 0.25\,h\,{\rm Mpc}^{-1}$ and find that P_4 barely improves constraints on cosmological parameters, as shown in Figure 8 [Figure 8: see original paper] in the Appendix.

5. Conclusion and Discussions

In this work, we develop an emulator for galaxy power spectra for catalogs with and without 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 deviations less than 5%). Our emulator-based likelihood analysis on mock galaxy catalogs demonstrates that input cosmological parameters can be accurately recovered from power spectra up to scales of $k = 0.5\,h\,{\rm Mpc}^{-1}$.

Our analysis shows that $P_{\rm pre}$, $P_{\rm post}$, and $P_{\rm cross}$ are highly complementary, thus jointly using these power spectra can significantly improve constraints on cosmological parameters, consistent with claims based on Fisher matrix analysis (Wang et al. 2022). Specifically, uncertainties in (Ω_m, H_0, σ_8) derived from $P_{\rm pre} + P_{\rm post} + P_{\rm cross}$ are reduced by 44.5%, 41.7%, and 55.3%, respectively, compared to those from $P_{\rm pre}$ alone ($k_{\rm max} = 0.25\,h\,{\rm Mpc}^{-1}$ in all cases). The derived BAO and RSD parameters $\alpha_\perp$, $\alpha_\parallel$, and $f\sigma_8$ are better determined by 33.9%, 28.8%, and 54.8%, respectively.

Adding small-scale modes to the analysis helps constrain parameters related to the amplitude of power spectra. For example, extending $k_{\rm max}$ from 0.25 to $0.5\,h\,{\rm Mpc}^{-1}$ for $P_{\rm post}$ reduces uncertainties on σ_8 and $f\sigma_8$ by 44.8% and 44.4%, respectively. We also find that parameter posteriors 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 significant 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), DESI (Dark Energy Spectroscopic Instrument; Aghamousa et al. 2016a,b), PFS (Prime Focus Spectrograph; Takada et al. 2014), and others, after the required tuning in the emulation process for number density, effective redshifts of galaxy samples, etc., which is technically straightforward.

Acknowledgments

We thank Hee-Jong Seo for helpful discussions. YW is supported by National Key R&D Program of China No. 2022YFF0503404, NSFC Grants 12273048, 11890691, 11720101004, the CAS Project for Young Scientists in Basic Research (No. YSBR-092), and the Youth Innovation Promotion Association CAS. RZ, YF, YW, and GBZ are supported by NSFC grants 11925303 and 11890691. RZ is also supported by the Chinese Scholarship Council (CSC) and the University of Portsmouth. KK is supported by the STFC grant ST/W001225/1. HG is supported by National SKA Program of China (No. 2020SKA0110100), NSFC Grants 11922305, 11833005, and the science research grants from the China Manned Space Project with No. CMS-CSST-2021-A02. GBZ 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. TN is supported by MEXT/JSPS KAKENHI Grant Numbers JP19H00677, JP20H05861, JP21H01081, and JP22K03634. Numerical work was performed on the UK Sciama High Performance Computing cluster supported by the ICG, University of Portsmouth. YW thanks the ICG, University of Portsmouth for their hospitality.

References

Aghamousa, A., et al. 2016a. https://arxiv.org/abs/1611.00036

Aghanim, N., et al. 2020, Astron. Astrophys., 641, A6, doi: 10.1051/0004-6361/201833910

Alam, S., Ata, M., Bailey, S., et al. 2017, MNRAS, 470, 2617, doi: 10.1093/mnras/stx721

Alam, S., Aubert, M., Avila, S., et al. 2021, PhRvD, 103, 083533, doi: 10.1103/PhysRevD.103.083533

Alcock, C., & Paczynski, B. 1979, Nature, 281, 358, doi: 10.1038/281358a0

Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W., & O'Neil, M. 2016, IEEE Transactions on Pattern Analysis and Machine Intelligence, 38, 252, doi: 10.1109/TPAMI.2015.2448083

Ando, S., Benoit-Lévy, A., & Komatsu, E. 2018, MNRAS, 473, 4318, doi: 10.1093/mnras/stx2634

Ba, S., Myers, W. R., & Brenneman, W. A. 2015, Technometrics, 57, 479, doi: 10.1080/00401706.2014.957867

Ballinger, W. E., Peacock, J. A., & Heavens, A. F. 1996, MNRAS, 282, 877, doi: 10.1093/mnras/282.3.877

Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013, ApJ, 762, 109, doi: 10.1088/0004-637X/762/2/109

Beutler, F., Saito, S., Seo, H.-J., et al. 2014, MNRAS, 443, 1065, doi: 10.1093/mnras/stu1051

Beutler, F., et al. 2017, MNRAS, 466, 2242, doi: 10.1093/mnras/stw3298

Blake, C., Kazin, E. A., Beutler, F., et al. 2011, MNRAS, 418, 1707, doi:

10.1111/j.1365-2966.2011.19592.x
Carrasco, J. J. M., Hertzberg, M. P., & Senatore, L. 2012, JHEP, 2012, 82, doi: 10.1007/JHEP09(2012)082
Chen, S.-F., Vlah, Z., & White, M. 2022, JCAP, 02, 008, doi: 10.1088/1475-7516/2022/02/008
Cole, S., Percival, W. J., Peacock, J. A., et al. 2005, MNRAS, 362, 505, doi: 10.1111/j.1365-2966.2005.09318.x
Colless, M., Dalton, G., Maddox, S., et al. 2001, MNRAS, 328, 1039, doi: 10.1046/j.1365-8711.2001.04902.x
Cuesta-Lazaro, C., Nishimichi, T., Kobayashi, Y., et al. 2023, MNRAS, 523, 3219, doi: 10.1093/mnras/stad1207
Cuesta-Lazaro, C., et al. 2023. https://arxiv.org/abs/2309.16539
d'Amico, G., Gleyzes, J., Kokron, N., et al. 2020, JCAP, 2020, 005, doi: 10.1088/1475-7516/2020/05/005
Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, AJ, 145, 10, doi: 10.1088/0004-6256/145/1/10
Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, AJ, 151, 44, doi: 10.3847/0004-6256/151/2/44
Donald-McCann, J., Beutler, F., Koyama, K., & Karamanis, M. 2022, MNRAS, 511, 3768, doi: 10.1093/mnras/stac239
Drinkwater, M. J., Jurek, R. J., Blake, C., et al. 2010, MNRAS, 401, 1429, doi: 10.1111/j.1365-2966.2009.15754.x
Eisenstein, D. J., & Hu, W. 1998, ApJ, 496, 605, doi: 10.1086/305424
Eisenstein, D. J., Seo, H.-J., Sirko, E., & Spergel, D. N. 2007, ApJ, 664, 675, doi: 10.1086/518712
Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al. 2005, ApJ, 633, 560, doi: 10.1086/466512
Guo, H., Zheng, Z., Zehavi, I., et al. 2014, MNRAS, 441, 2398, doi: 10.1093/mnras/stu763
Guo, H., et al. 2015, MNRAS, 453, 4368, doi: 10.1093/mnras/stv1966
Guo, H., Zheng, Z., Behroozi, P. S., et al. 2016, MNRAS, 459, 3040, doi: 10.1093/mnras/stw845
Hand, N., Feng, Y., Beutler, F., et al. 2018, AJ, 156, 160, doi: 10.3847/1538-3881/aadae0
Hand, N., Li, Y., Slepian, Z., & Seljak, U. 2017, JCAP, 2017, 002, doi: 10.1088/1475-7516/2017/07/002
Hearin, A. P., Campbell, D., Tollerud, E., et al. 2017, AJ, 154, 190, doi: 10.3847/1538-3881/aa859f
Hikage, C., Koyama, K., & Takahashi, R. 2020, Phys. Rev. D, 101, 043510, doi: 10.1103/PhysRevD.101.043510
Hikage, C., Takahashi, R., & Koyama, K. 2020, PhRvD, 102, 083514, doi: 10.1103/PhysRevD.102.083514
Ivanov, M. M., Simonović, M., & Zaldarriaga, M. 2020, PhRvD, 101, 083504, doi: 10.1103/PhysRevD.101.083504
Kaiser, N. 1987, MNRAS, 227, 1, doi: 10.1093/mnras/227.1.1
Klypin, A., & Prada, F. 2018, MNRAS, 478, 4602, doi: 10.1093/mnras/sty1340

Kobayashi, Y., Nishimichi, T., Takada, M., Takahashi, R., & Osato, K. 2020, Phys. Rev. D, 102, 063504, doi: 10.1103/PhysRevD.102.063504
Kwan, J., Saito, S., Leauthaud, A., et al. 2023, ApJ, 952, 80, doi: 10.3847/1538-4357/acd92f
Lange, J. U., Hearin, A. P., Leauthaud, A., et al. 2022, MNRAS, 509, 1779, doi: 10.1093/mnras/stab3111
Manera, M., Scoccimarro, R., Percival, W. J., et al. 2013, MNRAS, 428, 1036, doi: 10.1093/mnras/sts084
Mossa, V., et al. 2020, Nature, 587, 210, doi: 10.1038/s41586-020-2878-4
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563, doi: 10.1086/177173
Nishimichi, T., Takada, M., Takahashi, R., et al. 2019, ApJ, 884, 29, doi: 10.3847/1538-4357/ab3719
Peacock, J. A., et al. 2001, Nature, 410, 169, doi: 10.1038/35065528
Percival, W. J., Friedrich, O., Sellentin, E., & Heavens, A. 2022, MNRAS, 510, 3207, doi: 10.1093/mnras/stab3540
Percival, W. J., Nichol, R. C., Eisenstein, D. J., et al. 2007, ApJ, 657, 645, doi: 10.1086/510615
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13, doi: 10.1051/0004-6361/201525830
Savitzky, A., & Golay, M. J. E. 1964, Analytical Chemistry, 36, 1627, doi: 10.1021/ac60214a047
Sefusatti, E., Crocce, M., Scoccimarro, R., & Couchman, H. M. P. 2016, MNRAS, 460, 3624, doi: 10.1093/mnras/stw1229
Sherwin, B. D., & White, M. 2019, JCAP, 02, 027, doi: 10.1088/1475-7516/2019/02/027
Sobol', I. 1967, USSR Computational Mathematics and Mathematical Physics, 7, 86, doi: https://doi.org/10.1016/0041-5553(67)90144-9
Takada, M., Ellis, R. S., Chiba, M., et al. 2014, PASJ, 66, R1, doi: 10.1093/pasj/pst019
Taruya, A., Nishimichi, T., & Saito, S. 2010, PhRvD, 82, 063522, doi: 10.1103/PhysRevD.82.063522
Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, ApJ, 688, 709, doi: 10.1086/591439
Torrado, J., & Lewis, A. 2021, JCAP, 2021, 057, doi: 10.1088/1475-7516/2021/05/057
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261, doi: 10.1038/s41592-019-0686-2
Wang, Y., Zhao, G.-B., Koyama, K., et al. 2022, arXiv e-prints, arXiv:2202.05248, doi: 10.48550/arXiv.2202.05248
Wibking, B. D., Weinberg, D. H., Salcedo, A. N., et al. 2020, MNRAS, 492, 2872, doi: 10.1093/mnras/stz3423
Wibking, B. D., Salcedo, A. N., Weinberg, D. H., et al. 2019, MNRAS, 484, 989, doi: 10.1093/mnras/sty2258
Winther, H. A., Casas, S., Baldi, M., et al. 2019, PhRvD, 100, 123540, doi: 10.1103/PhysRevD.100.123540

York, D. G., Adelman, J., Anderson, John E., J., et al. 2000, AJ, 120, 1579, doi: 10.1086/301513
Yuan, S., Garrison, L. H., Eisenstein, D. J., & Wechsler, R. H. 2022, MNRAS, 515, 871, doi: 10.1093/mnras/stac1830
Zhai, Z., Tinker, J. L., Becker, M. R., et al. 2019, ApJ, 874, 95, doi: 10.3847/1538-4357/ab0d7b
Zhai, Z., Tinker, J. L., Banerjee, A., et al. 2023, ApJ, 948, 99, doi: 10.3847/1538-4357/acc65b
Zheng, Z., Coil, A. L., & Zehavi, I. 2007, ApJ, 667, 760, doi: 10.1086/521074
Zhou, Y.-D., Fang, K.-T., & Ning, J.-H. 2013, Journal of Complexity, 29, 283, doi: https://doi.org/10.1016/j.jco.2012.11.006

Appendix

This appendix includes four figures with information detailed in the figure captions.

Figure 5 [Figure 5: see original paper]. The complete training set for our emulator, which consists of 2400 power spectrum multipoles for $P_{\rm pre}$ (left column), $P_{\rm post}$ (middle), and $P_{\rm cross}$ (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 6 [Figure 6: see original paper]. The 1D posterior distribution and 2D contour plots showing 68% and 95% credible regions for derived parameters (Ω_m, H_0, σ_8, $\alpha_\perp$, $\alpha_\parallel$, $f\sigma_8$) from $P_{\rm all}$ reconstructed using two different sets of $b_{\rm in}$ and $f_{\rm in}$ shown in the legend. The dashed lines show the expected values of the parameters.

Figure 7 [Figure 7: see original paper]. 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 8 [Figure 8: see original paper]. The 1D posterior distribution and 2D contour plots showing 68% and 95% credible regions for (Ω_m, H_0, σ_8, $\alpha_\perp$, $\alpha_\parallel$, $f\sigma_8$) using $P_{\rm all}$ with and without the hexadecapole. The dashed lines show the expected values of the parameters.

Note: Figure translations are in progress. See original paper for figures.