

Cosmological constraints from density-split clustering in the BOSS CMASS galaxy sample

Enrique Paillas^{1,2*}, Carolina Cuesta-Lazaro^{3,4,5}, Will J. Percival^{1,2,6}, Seshadri Nadathur⁷, Yan-Chuan Cai⁸, Sihan Yuan^{9,10,11}, Florian Beutler⁸, Arnaud de Mattia¹², Daniel J. Eisenstein³, Daniel Forero-Sanchez¹³, Nelson Padilla¹⁴, Mathilde Pinon¹², Vanina Ruhlmann-Kleider¹², Ariel G. Sánchez¹⁵, Georgios Valogiannis^{16,17}, and Pauline Zarrouk¹⁸

¹ Waterloo Centre for Astrophysics, University of Waterloo, Waterloo, ON N2L 3G1, Canada

² Department of Physics and Astronomy, University of Waterloo, Waterloo, ON N2L 3G1, Canada

³ Center for Astrophysics | Harvard & Smithsonian, 60 Garden St, Cambridge, MA 02138, USA

⁴ The NSF AI Institute for Artificial Intelligence and Fundamental Interactions

⁵ Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

⁶ Perimeter Institute for Theoretical Physics, 31 Caroline St North, Waterloo, ON N2L 2Y5, Canada

⁷ Institute of Cosmology and Gravitation, University of Portsmouth, Burnaby Road, Portsmouth, PO1 3FX, UK

⁸ Institute for Astronomy, University of Edinburgh, Blackford Hill, Edinburgh, EH9 3HJ, UK

⁹ Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 452 Lomita Mall, Stanford, CA 94305, USA

¹⁰ Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA

¹¹ SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA

¹² IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France

¹³ Institute of Physics, Laboratory of Astrophysics, École Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, CH-1290 Versoix, Switzerland

¹⁴ Instituto de Astronomía Teórica y Experimental (IATE), CONICET-Universidad Nacional de Córdoba, Laprida 854, X5000BGR, Córdoba, Argentina

¹⁵ Max-Planck-Institut für Extraterrestrische Physik, Postfach 1312, Giessenbachstr., D-85748 Garching, Germany

¹⁶ Department of Physics, Harvard University, Cambridge, MA, 02138, USA

¹⁷ Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL, 60637, USA

¹⁸ Sorbonne Université, Université Paris Diderot, Sorbonne Paris Cité, CNRS, Laboratoire de Physique Nucléaire et de Hautes Energies (LPNHE), 4 place Jussieu, F-75252, Paris Cedex 5, France

23 April 2024

ABSTRACT

We present a clustering analysis of the BOSS DR12 CMASS galaxy sample, combining measurements of the galaxy two-point correlation function and density-split clustering down to a scale of $1 h^{-1}$ Mpc. Our theoretical framework is based on emulators trained on high-fidelity mock galaxy catalogues that forward model the cosmological dependence of the clustering statistics within an extended- Λ CDM framework, including redshift-space and Alcock-Paczynski distortions. Our base- Λ CDM analysis finds $\omega_{\text{cdm}} = 0.1201 \pm 0.0022$, $\sigma_8 = 0.792 \pm 0.034$, and $n_s = 0.970 \pm 0.018$, corresponding to $f\sigma_8 = 0.462 \pm 0.020$ at $z \approx 0.525$, which is in agreement with Planck 2018 predictions and various clustering studies in the literature. We test single-parameter extensions to base- Λ CDM, varying the running of the spectral index, the dark energy equation of state, and the density of massless relic neutrinos, finding no compelling evidence for deviations from the base model. We model the galaxy-halo connection using a halo occupation distribution framework, finding signatures of environment-based assembly bias in the data. We validate our pipeline against mock catalogues that match the clustering and selection properties of CMASS, showing that we can recover unbiased cosmological constraints even with a volume 84 times larger than the one used in this study.

Key words: cosmological parameters, large-scale structure of Universe

1 INTRODUCTION

Within the vast cosmic structures we observe today, signatures of primordial features are intermixed with non-linear processes that shape the evolution of galaxies, constituting a ground that is challenging to model but rich in information. In our standard cosmological model, the Λ CDM paradigm, the present-day Universe is the result of a hierarchical structure formation scenario that started from primordial

density perturbations, which were amplified during inflation and continued to grow through gravitational collapse until today (Guth & Pi 1982; Hawking 1982). Galaxy clustering has been pivotal in testing this hypothesis from late-time Universe data by characterizing the way in which matter is distributed in space, using galaxies as biased tracers of the underlying dark matter distribution.

The most common way to approach galaxy clustering is through the two-point correlation function, or its Fourier pair, the power spectrum. These statistics provide a nearly-complete description of the galaxy density field on large scales and encode information about

* E-mail: enrique.paillas@uwaterloo.ca

physics of the early Universe (Peebles 1980). Baryon acoustic oscillations (BAO), which originate from sound waves in the photon-baryon plasma before recombination (Peebles & Yu 1970; Sunyaev & Zel'dovich 1970), leave an imprint on the matter distribution, which is detected as a bump in the correlation function or as wiggles in the power spectrum (Eisenstein & Hu 1998). This acoustic feature works as a standard ruler, allowing us to measure the expansion rate of the Universe at different epochs (Percival et al. 2001; Eisenstein et al. 2005; Cole et al. 2005; eBOSS Collaboration et al. 2020; Moon et al. 2023). In addition to BAO, the full shape of the galaxy correlation function and power spectrum contains an enormous wealth of information due to its sensitivity to the growth rate of cosmic structure (Blake et al. 2011; Reid et al. 2012; Alam et al. 2017; Brieden et al. 2021), dark energy, the physics of neutrinos (Zhang & Cai 2022), primordial non-Gaussianity (Moradinezhad Dizgah et al. 2021), and the galaxy-halo connection (Yuan et al. 2022a).

Models based on perturbation theory provide an accurate description of galaxy clustering data on linear and mildly non-linear scales, allowing the extraction of information from the full shape of the power spectrum (e.g., Sánchez et al. 2017; Grieb et al. 2017; d'Amico et al. 2020; Tröster et al. 2020; Bautista et al. 2021; Philcox & Ivanov 2022; Semenaite et al. 2022, 2023). The assumptions behind these models tend to break down in the highly non-linear regime ($\lesssim 20 h^{-1} \text{Mpc}$), which has motivated the development of models calibrated on N-body simulations that can accurately describe the galaxy field on scales where non-linear physical processes become relevant. This has recovered information from a portion of the survey data that is usually discarded from the standard clustering analyses, providing clues not only about cosmology but also in regards to galaxy evolution and its connection to the dark matter halo field (Kobayashi et al. 2020; Zhai et al. 2023; Chapman et al. 2022; Lange et al. 2022, 2023; Yuan et al. 2022a).

Two-point functions provide a complete description of Gaussian density fields. This is satisfied at large scales, where non-linear evolution is mild, and the Gaussianity of the primordial fluctuations are still largely preserved. The late-time Universe, however, is highly non-Gaussian at small scales due to non-linear evolution, and higher-order summary statistics are required to capture all the information. Thanks to substantial theoretical and algorithmic development over the last years, measurements of N-point correlation functions (Slepian et al. 2017; Philcox et al. 2021; Sugiyama et al. 2023) and polyspectra (Gil-Marín et al. 2017; Gualdi et al. 2021; Philcox & Ivanov 2022) are now being performed on data, which not only tightens the constraining power on ΛCDM , but also opens an avenue for testing potential signatures of physics beyond our fiducial model, such as parity violation (Philcox 2022; Hou et al. 2023a). However, the measurement of these N-point statistics remains challenging due to their high computational demands and the large volumes that are needed to detect them with enough statistical significance. This has motivated the development of robust, informative, and efficient clustering methods that can access the information that leaks into higher orders, which can then be complemented and cross-validated with the standard N-point clustering analysis.

Several alternative summary statistics that meet these criteria have been proposed in the literature, including k-th nearest neighbor statistics (Banerjee & Abel 2021), wavelet scattering transforms (Valogiannis & Dvorkin 2022a,b), void statistics (Lavaux & Wandelt 2012; Sutter et al. 2012; Hamaus et al. 2016; Mao et al. 2017; Hawken et al. 2017; Achitouv 2019), marked correlations (White 2016; Masara et al. 2022), skew spectra (Hou et al. 2023b), and Minkowski functionals (Lippich & Sánchez 2021). Neyrinck (2008) showed that conditioning a clustering analysis on local density brings sensitiv-

ity to non-Gaussian features in the large-scale structure, revealing the expansion or contraction of the BAO feature to local density. Along the same lines, Paillas et al. (2021) showed that the analysis of redshift-space distortions around regions split by local density is able to tighten the constraints on the growth rate of cosmic structure compared to a standard two-point clustering analysis. Paillas et al. (2023) expanded on this by quantifying the information content of the full shape of the correlation function split by density, forecasting that the method could deliver precise constraints on ΛCDM parameters, and could potentially be used to put upper limits on the sum of neutrino masses. Density-split statistics have also been successfully used in weak gravitational lensing analyses, where it has been shown that they are able to access higher-order information from the projected density field and provide cosmological constraints that are competitive with the standard analyses based on two-point statistics (Gruen et al. 2018; Friedrich et al. 2018; Burger et al. 2023).

Until now, a model that can capture the cosmological dependence of the full shape of the density-split correlation functions from three-dimensional large-scale structure maps was not available. In Cuesta-Lazaro et al. (2023), we have presented SUNBIRD, a simulation-based model for density-split and two-point clustering that can operate down to intra-halo scales, well into the non-linear regime, which has been validated against high-fidelity mock galaxy catalogues based on the ABACUS SUMMIT suite of simulations. In this work, we use SUNBIRD to carry out the first application of density-split clustering to observational data, applying it to the final data release of the Baryon Oscillation Spectroscopic Survey (Dawson et al. 2013). We fit a full-shape model of the galaxy two-point correlation function and density-split multipoles down to $1 h^{-1} \text{Mpc}$, putting constraints on the base- ΛCDM model, as well as on extensions to the base model that vary the dark energy equation of state, the density of relic neutrinos, and the running of the spectral index of the primordial power spectrum.

The paper is organized as follows. We define our observables in Sect. 2. The clustering modelling, as well as the simulations used to validate our pipeline are presented in Sect. 3. We present our main cosmological constraints and the validation tests in Sect. 4. Finally, we summarize and conclude in Sect. 5.

2 OBSERVATIONS

2.1 BOSS CMASS galaxy sample

We use data from the final data release (DR12) of the Baryon Oscillation Spectroscopic Survey (BOSS, Dawson et al. 2013). BOSS was a survey conducted as part of the third stage of the larger Sloan Digital Sky Survey (SDSS, York et al. 2000), which collected optical spectra from more than 1.5 million targets using the 2.5-m Sloan Telescope (Gunn et al. 2006) at Apache Point, New Mexico. BOSS covered roughly $10,000 \text{ deg}^2$ of the sky in two hemispheres, referred to as the North and the South Galactic caps (NGC and SGC, respectively).

Our analysis is focused on the CMASS galaxy sample, which is dominated by luminous red galaxies (LRG) that were selected on SDSS multicolour photometric observations (Gunn et al. 1998; Gunn et al. 2006). CMASS is nearly complete down to stellar mass of $M_* \approx 10^{11.3} M_\odot$ for $z > 0.45$ (Maraston et al. 2013), and covers a redshift range $0.4 \lesssim z \lesssim 0.7$. For this paper, we impose a more stringent redshift cut $0.45 \leq z \leq 0.6$ to avoid regions where the galaxy number density drops abruptly, which can potentially bias our model predictions. Additionally, we restrict the analysis to the NGC for simplicity. After imposing these restrictions, this results in a sample

with an average number density of $\bar{n} \approx 3.5 \times 10^{-4} (h\text{Mpc}^{-1})^3$, a total volume of $\approx 1.4 (h^{-1}\text{Gpc})^3$, and an *effective* volume of $\approx 1.1 (h^{-1}\text{Gpc})^3$, where we have defined the effective volume as

$$V_{\text{eff}} = \int \left[\frac{\bar{n}P_0}{1 + \bar{n}P_0} \right]^2 dV(z), \quad (1)$$

where $P_0 = 10000 h^{-3}\text{Mpc}^3$ (Reid et al. 2016).

We use the DR12 large-scale structure catalogues provided by the BOSS collaboration¹ (Reid et al. 2016). These catalogues include angles and redshifts for each galaxy, which we convert to comoving Cartesian coordinates by adopting a flat- Λ CDM fiducial cosmology characterized by a matter density parameter $\Omega_m = 0.315$, which closely matches the Planck 2018 best-fit cosmology, assuming base- Λ CDM (Planck Collaboration et al. 2020). The BOSS collaboration also provides a set of random catalogues that follow the footprint and radial selection of CMASS galaxies, but with no intrinsic clustering, which are used to estimate the overdensity field as described in the following sections.

2.2 Clustering measurements

2.2.1 Two-point clustering

Galaxy clustering is usually characterized in terms of the two-point correlation function (2PCF), $\xi^{\text{gg}}(r)$, which, in its simplest form, quantifies the excess probability dP of finding a galaxy in a volume dV , separated at a distance r from another galaxy, with respect to an unclustered Poisson distribution:

$$dP = n_g [1 + \xi^{\text{gg}}(r)] dV, \quad (2)$$

where n_g is the mean number density of galaxies in the sample. In the presence of redshift-space distortions (RSD, Jackson 1972; Kaiser 1987) or Alcock-Paczynski distortions (AP, Alcock & Paczynski 1979), the galaxy distribution appears anisotropic to the observer. To capture this anisotropy, a convenient choice is to bin the correlation function in terms of s and μ , where s is the redshift-space pair separation and μ is the cosine of the angle between the vector connecting the two galaxies and the observer's line of sight.

A number of estimators have been proposed to measure the 2PCF from observational data. A robust and commonly used estimator is the one introduced by Landy & Szalay (1993),

$$\xi^{\text{gg}}(s, \mu) = \frac{\text{GG} - 2\text{GR} + \text{RR}}{\text{RR}}, \quad (3)$$

where GG is the normalized number of galaxy pairs in the (s, μ) bin, while GR and RR are the normalized galaxy-random and random-random pairs, which make use of the unclustered random catalogues.

It is useful to separate out the different angular components of the (s, μ) correlation function by decomposing it into multipole moments, defined by

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

with P_ℓ the Legendre polynomials.

We measure $\xi^{\text{gg}}(s, \mu)$ in CMASS using PYCORR², which is a wrapper around a modified version of the CORRFUNC pair-counting code (Sinha & Garrison 2020). We focus the analysis on the monopole (ξ_0) and quadrupole (ξ_2) moments. The correlation functions are measured in 241 μ bins from -1 to 1 , and radial bins with scale-dependent

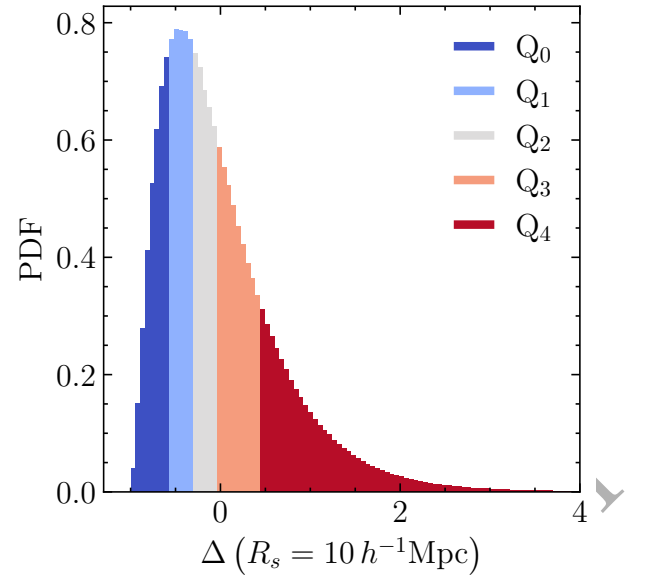


Figure 1. Probability distribution function (PDF) of the galaxy overdensity measured around random query points. The overdensity field has been smoothed with a Gaussian filter of width $R_s = 10 h^{-1}\text{Mpc}$. The colours represent the split of the PDF into quintiles.

widths: $1h^{-1}\text{Mpc}$ bins for $s \in [0, 4] h^{-1}\text{Mpc}$, $3h^{-1}\text{Mpc}$ bins for $s \in (4, 30] h^{-1}\text{Mpc}$, and $5h^{-1}\text{Mpc}$ bins for $s \in (30, 151] h^{-1}\text{Mpc}$. We have chosen to work with scale-dependent bin widths to have more resolution in those scales where we expect to see important features in the correlation functions, such as the BAO peak and the one- to two-halo term transition, and less resolution in scales where the correlation functions are featureless. This keeps the size of the data vector within a manageable size such that the dimensions of the covariance matrix do not increase too much, making our mock-based estimates of the covariance more reliable.

Galaxies are appropriately weighted during the pair counting to consider various observational systematics that can bias the clustering measurements. The total systematic weight for each galaxy is given by

$$w_{\text{sys,tot}} = w_{\text{sys}}(w_{\text{fc}} + w_{\text{zf}} - 1), \quad (5)$$

where w_{sys} , w_{fc} , and w_{zf} account for imaging systematics, fiber collisions, and redshift failures (Ross et al. 2016). This is additionally multiplied by a weight w_{FKP} , which optimally weights the contribution of galaxies based on their redshift-dependent number density (Feldman et al. 1994),

$$w_{\text{FKP}} = 1/(1 + n(z)P_0), \quad (6)$$

where $P_0 = 10^4 h^3\text{Mpc}^{-3}$. The total weights for each galaxy are then given by

$$w_{\text{tot}} = w_{\text{FKP}}w_{\text{sys,tot}}, \quad (7)$$

while points from the random catalogue are only weighted by w_{FKP} .

2.2.2 Density-split clustering

The density-split clustering (DSC) method (Paillas et al. 2021) characterizes galaxy clustering in bins of environmental density, with the aim of extracting and combining the cosmological information coming from distinct environments of the cosmic web. We apply the density-split algorithm to the CMASS galaxy sample using our

¹ <https://data.sdss.org/sas/dr12/booss/lss/>.

² <https://github.com/cosmodesi/pycorr>.

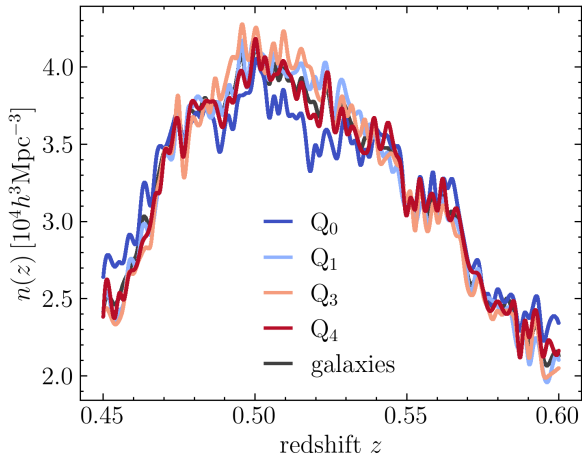


Figure 2. Comoving number density as a function of redshift for the density-split quintiles and galaxies from the BOSS DR12 CMASS sample. [🔗](#)

publicly available code³, with slight modifications to the algorithm presented in Paillas et al. (2023) to account for the non-uniform survey geometry.

We start by painting the CMASS galaxies and randoms to a rectangular grid that fully encompasses the survey volume, and we estimate the overdensity field as

$$\delta = \frac{G}{R} - 1 \quad (8)$$

where G and R are the normalized galaxy and random counts in each cell, weighted as given in Eq. (5). We smooth the overdensity field with a Gaussian filter of radius $R_s = 10 h^{-1} \text{Mpc}$, and then we sample it using cloud-in-cell interpolation at N_{query} query positions, which are taken from the CMASS random catalogue. Here we set N_{query} to 5 times the number of galaxies in the catalogue, and split those query points into 5 quintiles, according to the overdensity at each location.

Figure 1 shows the probability distribution function (PDF) of galaxy overdensity measured at the query positions. The overdensity field shows a non-Gaussian PDF with significant skewness and kurtosis. This shape is a consequence of the growth of structure being bounded at $\Delta = -1$ from below (regions completely devoid of galaxies), while no such constraint is present at the positive Δ end. The distribution peaks at negative overdensities, reflecting that the average region in the Universe is underdense due to the larger volume occupied by voids. The division into quintiles is demarcated by the different colours in the figure. We label the quintiles as Q_i , where i goes from 0 to 4 from lower to higher densities. In what follows, we discard Q_2 from the clustering analysis, since all five quintiles are not independent from each other.⁴

Figure 2 shows the comoving number density of quintile positions as a function of redshift, which is observed to closely follow that of the galaxies. This is given by construction, as the query positions are sampled from the clustering random catalogues, which in turn are constructed to match the footprint and radial selection of the galaxies.

Once the density quintiles are defined, we estimate the quintile-galaxy cross-correlation function using the estimator (Landy & Sza-

³ <https://github.com/epaillas/densitysplit>.

⁴ As the query points are random, the sum of the density-split cross-correlation functions over quintiles vanishes up to shot noise, and all the information in Q_2 is already contained in the remaining four quintiles, as shown in previous work (Paillas et al. 2023).

lay 1993)

$$\xi^{\text{qg}}(s, \mu) = \frac{\text{QG} - \text{QR} - \text{RG} + \text{RR}}{\text{RR}}, \quad (9)$$

where QG, QR, RG, RR are the normalized quintile-galaxy, quintile-random, galaxy-random, and random-random pair counts. We note that this assumes that the random catalogue is the same for galaxies and quintiles, which is justified by the good match between the $n(z)$ distributions from Fig. 2. The autocorrelation functions of DS quintiles are estimated as

$$\xi^{\text{qq}}(s, \mu) = \frac{\text{QQ} - 2\text{QR} + \text{RR}}{\text{RR}}, \quad (10)$$

with QQ being the normalized quintile-quintile pair counts.

We adopt the same binning scheme as for the galaxy 2PCF, as described in Sect. 2.2.1, and we weight the galaxy pairs according to Eq. (7).

Figure 3 shows the measured multipoles from DSC and the galaxy 2PCF. The DSC auto- and cross-correlation functions cover a broad range of amplitudes, as the quantiles trace the underlying matter density field in different ways. Q_0 & Q_1 are underdense regions with negative linear bias parameters that usually range from -3 to -1, akin to cosmic voids (Hamaus et al. 2014; Chan et al. 2014), whereas Q_3 & Q_4 tracer overdensities and have positive bias parameters that can range from 1 to 3 (Paillas et al. 2021). Two signature features are spotted in all profiles: a transition regime around $25 h^{-1} \text{Mpc}$, which is due to the scale that was used to smooth the overdensity field and define the quintiles, and the peak/valley around $100 h^{-1} \text{Mpc}$, which is an imprint of the BAO that originated from sound waves in the photon-baryon plasma prior to recombination.

The quadrupoles show a large degree of anisotropy in DSC. Two main factors contribute to this anisotropy. Firstly, there is an RSD effect caused by the dynamics of galaxies around different density environments. Kaiser-like motions on large scales and random motions on small scales (Fingers of God) cause distinct RSD patterns on the clustering of each quintile, similar to the well-known effects seen in the galaxy 2PCF. Secondly, there is an RSD effect imprinted on the quintile positions themselves, which is the product of identifying the density quintiles in redshift space. Paillas et al. (2023) showed that splitting densities in redshift space causes selection effects that induce distortions in the clustering of the quintiles, which manifests itself as a quadrupole moment in the DSC autocorrelation functions. This is similar in nature to the selection effect that is produced when cosmic voids are identified in redshift space (Chuang et al. 2017; Nadathur et al. 2019; Correa et al. 2020).

We reserve the discussion of the model fits (solid lines) for Sect. 4.1.

3 MODELLING

3.1 Mock galaxy catalogues

Throughout this work, we use different types of mock galaxy catalogues to study various aspects of the analysis. The MD-Patchy mocks are used to estimate the sample variance associated to our observed clustering measurements. The Nseries mocks are used to validate the theory model we apply to CMASS. The AbacusSummit mocks are used to train our simulation-based model for galaxy clustering, as presented in our companion paper (Cuesta-Lazaro et al. 2023). Each of these mocks are described in more detail below.

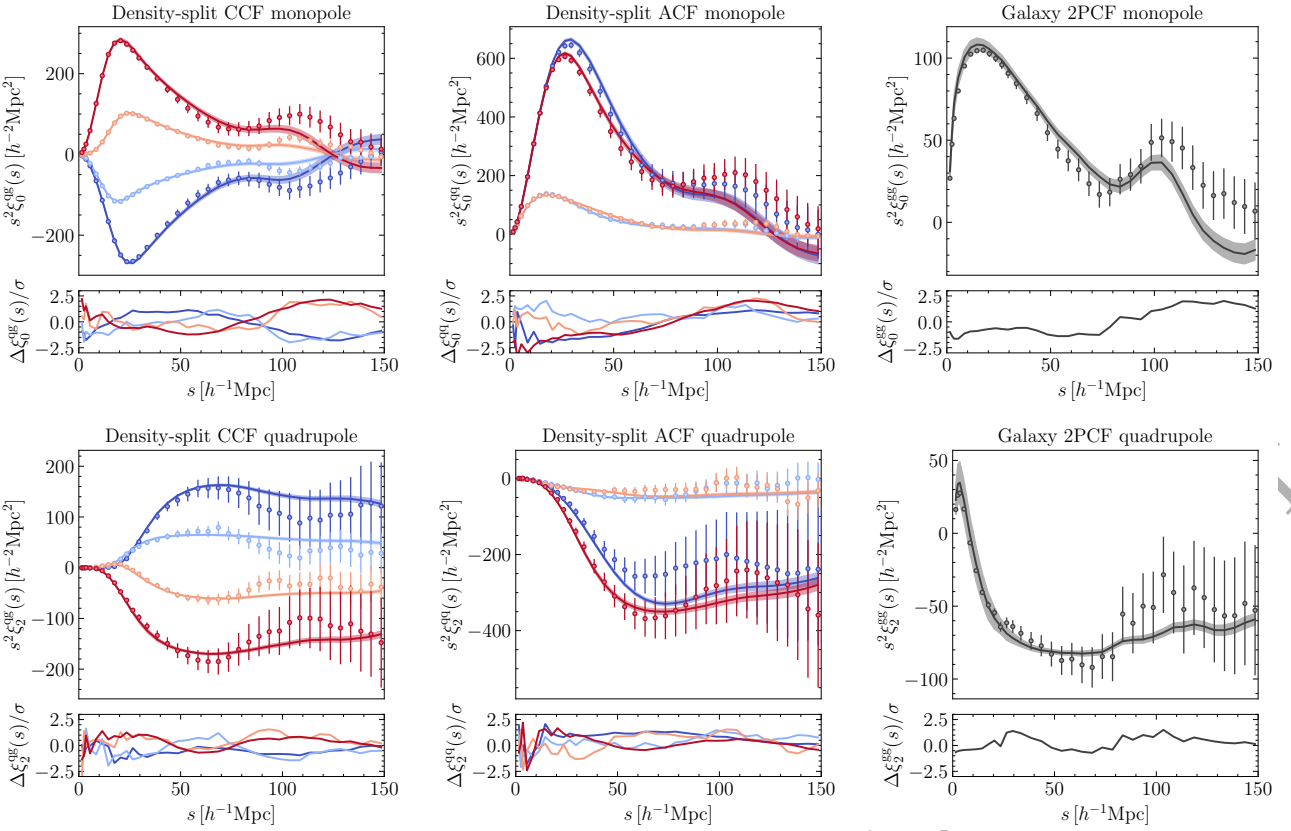


Figure 3. DSC and galaxy 2PCF multipoles measured from the BOSS DR12 CMASS catalogue (circles with error bars), along with the best-fit model from our emulator (solid lines with shaded bands). The columns show the multipoles of the quintile-galaxy cross-correlation function (left), the quintile autocorrelation function (middle), and the galaxy two-point correlation function (right). The lower sub-panels display the difference between the model and the data, in units of the standard deviation of the total error budget. Each colour corresponds to a different density quintile, as illustrated in Fig. 1.

The 68 per cent errors of the data are estimated from 2048 realizations of the MD-Pathchy mocks and represent the expected level of sample variance for the CMASS NGC volume. The emulator uncertainty is estimated by validating the predictions against a test set of simulations with a known cosmology. [☑](#)

3.1.1 MD-Pathchy mocks

To estimate the covariance of the data vector, we use the MultiDark-Pathchy mocks (MD-Pathchy, [Kitauro et al. 2016](#)), a suite of 2048 mock galaxy catalogues that were designed to match the footprint, redshift distribution and halo occupation distribution of the BOSS DR12 galaxy samples. The mocks are based on approximate lightcones generated with augmented Lagrangian perturbation theory (ALPT) using the PATCHY code ([Kitauro et al. 2014, 2015](#)). ALPT is based on a combination of second-order Lagrangian perturbation theory on large scales and the spherical collapse model on smaller scales. PATCHY populates dark matter haloes using a subhalo abundance matching prescription that is calibrated from N-body simulations from the Big MultiDark Suite ([Klypin et al. 2016](#)), and uses a Λ CDM model matched to the best fit to the Planck 2013 CMB measurements ([Planck Collaboration et al. 2014](#)), characterized by a matter density parameter $\Omega_m = 0.307115$, a baryon density parameter $\Omega_b = 0.048$, an amplitude of matter fluctuations in $8 h^{-1}$ Mpc spheres $\sigma_8 = 0.8288$, a tilt of the primordial power spectrum $n_s = 0.9611$, and a dimensionless Hubble parameter $h = 0.6777$.

3.1.2 Nseries mocks

To study systematic errors in our theory model, we use the Nseries cutsy mocks, a collection of 84 mock catalogues that were designed to match the clustering, footprint and radial selection of the CMASS

galaxy sample. The cutsy mocks are constructed from a base set of seven independent full N-body simulations, with a box side length of $2.6 h^{-1}$ Gpc and a mass resolution of $1.5 \times 10^{11} h^{-1} M_\odot$. Dark matter haloes are populated with a halo occupation distribution prescription, with parameters chosen to best model the clustering of the BOSS DR12 CMASS sample. Each box is trimmed and rotated in different ways to produce 12 cutsy mocks that match the geometry of the CMASS sample, which results in a total of 84 pseudo-independent cutsy catalogues. The mocks are then passed through the same fiber assignment code that was used for BOSS, so that they faithfully reproduce the angular variations of fiber collisions in the data ([Hahn et al. 2017](#)).

Nseries is characterized by a cosmology $\Omega_m = 0.286$, $\Omega_b = 0.0470$, $h = 0.70$, $\sigma_8 = 0.82$, and $n_s = 0.96$.

3.1.3 AbacusSummit mocks

To train and validate our clustering emulators, we use ABACUS-SUMMIT, a suite of cosmological N-body simulations ([Maksimova et al. 2021](#)) designed to meet the simulation requirements of the Dark Energy Spectroscopic Instrument ([Levi et al. 2019](#)). The simulations were run with the ABACUS N-body code ([Garrison et al. 2019, 2021](#)), comprising different volumes, resolutions, and cosmologies. The *base* simulations follow the evolution of 6912^3 dark matter particles in a $(2 h^{-1} \text{Gpc})^3$ volume, with a mass resolution of

Table 1. List of cosmological and HOD parameters used in our analysis. For each, we quote the parameter symbol, the prior distribution, the fixed value in the baseline model (where appropriate), and the physical interpretation. We note that the prior distribution for all parameters is uniform, with the exception of the baryon density, for which we adopt a normal distribution with a mean and dispersion as specified.

Parameter	Prior distribution	Baseline	Interpretation
ω_b	$\mathcal{N}(0.02268, 0.00038)$	—	Physical baryon density
ω_{cdm}	$\mathcal{U}[0.1032, 0.14]$	—	Physical cold dark matter density
σ_8	$\mathcal{U}[0.687, 0.938]$	—	Amplitude of matter fluctuations in $8 h^{-1} \text{Mpc}$ spheres
n_s	$\mathcal{U}[0.901, 1.025]$	—	Spectral index of the primordial power spectrum
$dn_s/d \ln k$	$\mathcal{U}[-0.038, 0.038]$	0.0	Running of the spectral index
N_{eff}	$\mathcal{U}[2.1902, 3.9022]$	3.0146	Number of ultra-relativistic species
w_0	$\mathcal{U}[-1.27, -0.70]$	-1.0	Present-day dark energy equation of state
w_a	$\mathcal{U}[-0.628, 0.621]$	0.0	Time evolution of the dark energy equation of state
M_{cut}	$\mathcal{U}[12.4, 13.3]$	—	Minimum halo mass to host a central
M_1	$\mathcal{U}[13.2, 14.4]$	—	Typical halo mass to host one satellite
$\log \sigma$	$\mathcal{U}[-3.0, 0.0]$	—	Slope of the transition from hosting zero to one central
α	$\mathcal{U}[0.7, 1.5]$	—	Power-law index for the mass dependence of the number of satellites
κ	$\mathcal{U}[0.0, 1.5]$	—	Parameter that modulates the minimum halo mass to host a satellite
α_c	$\mathcal{U}[0.0, 0.5]$	—	Velocity bias for centrals
α_s	$\mathcal{U}[0.7, 1.3]$	—	Velocity bias for satellites
B_{cen}	$\mathcal{U}[-0.5, 0.5]$	—	Environment-based assembly bias for centrals
B_{sat}	$\mathcal{U}[-1.0, 1.0]$	—	Environment-based assembly bias for satellites

$2 \times 10^9 h^{-1} \text{M}_\odot$. There are 97 cosmology variations in total, exploring an eight-dimensional parameter space around the Planck18 primary ΛCDM cosmology (PL18, [Planck Collaboration et al. 2020](#)):

$$\theta_{\text{ABACUSUMMIT}} = \{ \omega_{\text{cdm}}, \omega_b, \sigma_8, n_s, dn_s/d \ln k, N_{\text{eff}}, w_0, w_a \}. \quad (11)$$

Here, $\omega_{\text{cdm}} = \Omega_c h^2$ and $\omega_b = \Omega_b h^2$ are the physical cold dark matter and baryon densities, $dn_s/d \ln k$ is the running of the spectral tilt, N_{eff} is the effective number of ultra-relativistic species, w_0 is the present-day dark energy equation of state, and w_a captures the time evolution of the dark energy equation of state. The simulations assume a flat spatial curvature, and the dimensionless Hubble parameter h is calibrated to match the Cosmic Microwave Background (CMB) acoustic scale θ_* to the PL18 measurement.

For the training and validation of the emulators, we restrict to the following subset of simulations, which were all run with the same initial conditions:

c000 PL18 ΛCDM base cosmology, matching the mean of the base_plikHM_TTTEEE_low_lowE_lensing_likelihood.

c001-004 Secondary cosmologies, including a low ω_{cdm} choice (WMAP7, [Komatsu et al. 2011](#)), a $w\text{CDM}$ choice, a high N_{eff} choice, and a low σ_8 choice.

c013 Cosmology matching the Euclid Flagship2 ΛCDM simulation (Castander et al., in preparation).

c100-126 Linear derivative grid providing paired simulations with small negative and positive steps in the eight-dimensional cosmological parameter space from Eq. (11).

c130-181 An emulator grid around the c000 cosmology that provides a wider coverage of the cosmological parameter space.

In addition to the base simulations, there are multiple realizations of smaller boxes with a side length of $500 h^{-1} \text{Mpc}$ at the c000 cosmology, which can be used for covariance estimation. Throughout the rest of the paper, we will refer to these simulations as ABACUS_{SMALL}, and to the base simulations simply as ABACUS_{SUMMIT}.

Group finding is done on the fly, using a hybrid Friends-of-Friends/Spherical Overdensity algorithm, dubbed CompaSO

([Hadzhiyska et al. 2021](#)). As described in Sect. 3.3, we populate these halo catalogues with galaxies using a halo occupation distribution prescription, which are then used to obtain the clustering measurements for our training data.

3.2 Galaxy-halo connection model

In the current paradigm of cosmology, galaxies are thought to form and evolve within dark matter halos, which are large structures that form as a result of the gravitational collapse of overdensities in the Universe. The halo occupation distribution (HOD) is a statistical model that describes how galaxies are distributed within dark matter halos.

A well-suited model for LRG is the base halo model from [Zheng et al. \(2007\)](#), where the average number of central galaxies in a halo of mass M is given by

$$\langle N_c \rangle (M) = \frac{1}{2} \left(1 + \text{erf} \left(\frac{\log M - \log M_{\text{cut}}}{\sqrt{2}\sigma} \right) \right) \quad (12)$$

where $\text{erf}(x)$ denotes the error function, M_{cut} is the minimum mass required to host a central, and σ is the slope of the transition between having zero to one central galaxies. The average number of satellite galaxies is in turn given by

$$\langle N_s \rangle (M) = \langle N_c \rangle (M) \left(\frac{M - \kappa M_{\text{cut}}}{M_1} \right)^\alpha \quad (13)$$

where κM_{cut} gives the minimum mass required to host a satellite, M_1 is the typical mass that hosts one satellite, and α is the power law index for the number of galaxies.

We use the ABACUSHOD package, which is highly efficient and contains a wide range of HOD extensions ([Yuan et al. 2022a](#)). For this work, we extend the base model to modulate galaxy peculiar velocities via the parameters $\alpha_{\text{vel,c}}$, which parametrizes the velocity bias between the central galaxy and the halo centre, and $\alpha_{\text{vel,s}}$, which parametrizes the velocity bias between the satellite galaxies and the local dark matter particles. When no velocity bias is present, $\alpha_{\text{vel,c}} = 0$ and $\alpha_{\text{vel,s}} = 1$, in which case centrals perfectly follow the velocity

of halo centres, and satellites perfectly match the velocity of the underlying dark matter particles.

We introduce two additional parameters to account for galaxy assembly bias: B_{cen} and B_{sat} , which add environment-based secondary bias for centrals and satellites, respectively. Here, the environment is defined as the smoothed matter density around the halo centres, using a top-hat filter of radius $R_s = 5 h^{-1} \text{Mpc}$. When no secondary bias is present, $B_{\text{cen}} = B_{\text{sat}} = 0$. Positive/negative values of these parameters indicate a preference for galaxies to form in haloes around less/more dense environments, respectively.

3.3 Clustering emulators

In the context of galaxy clustering and cosmology, an emulator refers to a computational model or algorithm that approximates the predictions of expensive or time-consuming simulations or calculations. They are used to efficiently and accurately generate predictions from cosmological models without resorting to additional simulations.

Emulators are trained on a set of pre-computed simulations, often referred to as a training or calibration set, which cover a wide range of model parameter values and capture the desired properties of interest of an observable. As such, the emulator can learn how this observable responds to changes in cosmology or galaxy-halo connection parameters. Once the emulator is trained, it can rapidly predict the desired model predictions for any given set of input parameters.

In Cuesta-Lazaro et al. (2023), we present our galaxy 2PCF and DSC emulators, which are based on HOD catalogues constructed from the ABACUS-SUMMIT simulations. Here, we present a brief description of how the emulators are constructed.

We start from the dark matter halo catalogues from the ABACUS-SUMMIT snapshots at $z = 0.5$, spanning 85 different cosmologies within the eight-dimensional $w_0 w_d \Lambda \text{CDM}$ parameter space defined in Eq. (11). Using the ABACUSHOD code (Yuan et al. 2022a), we populate dark matter haloes with a nine-parameter extended HOD framework (Sect. 3.2),

$$\theta_{\text{HOD}} = \{M_{\text{cut}}, M_1, \sigma, \alpha, \kappa, \alpha_{\text{vel,c}}, \alpha_{\text{vel,s}}, B_{\text{cen}}, B_{\text{sat}}\}, \quad (14)$$

generating 100 unique HOD variations per cosmology, where the HOD parameters are sampled from a Latin Hypercube to ensure an optimal sampling of the parameter space. When the number density of an HOD catalogue is above the average number density of the CMASS sample ($n_{\text{gal}} \approx 3.5 \times 10^{-4} (h/\text{Mpc})^3$), we invoke an incompleteness parameter f_{ic} to downsample the catalogue down to the target number density.

Under the distant-observer approximation, we map the positions of galaxies to redshift space by perturbing their positions with their peculiar velocities along one cartesian axis of the simulation box chosen to represent the line of sight. We repeat this procedure for each of the three axes of the simulation boxes, effectively generating three pseudo-independent redshift-space catalogues for each HOD variation, from which we can average the clustering measurements to reduce cosmic variance later on.

For each of these mock catalogues, we run the DSC pipeline, using a mesh resolution of $R_{\text{cell}} = 5 h^{-1} \text{Mpc}$, a number of random query points N_{query} equal to five times the number of galaxies in each catalogue, a smoothing radius for the Gaussian filter of $R_s = 10 h^{-1} \text{Mpc}$, and five density quintiles. We compute the galaxy 2PCF and the density-split correlation functions in bins of s and μ , and we decompose them into their multipole moments.

We split the HOD catalogues into training, validation and test sets, and for each clustering statistic we construct separate fully-connected neural networks, which take the cosmological and HOD parameters

as an input, and return the monopole and quadrupole moments of the correlation functions. For training and validation, we use cosmologies [c100-126] and [c130-181], whereas the rest are reserved for the test set. The hyperparameters of the neural networks are calibrated to minimize the validation loss. Overall, we observe that the emulators produce model predictions with percent-level accuracy for the full range of scales, when tested against the test simulation boxes.

3.4 Likelihood

When fitting the emulator to the CMASS data, we define the log-likelihood as

$$\mathcal{L} = (\mathbf{d}^{\text{data}} - \mathbf{d}^{\text{model}}) \mathbf{C}^{-1} (\mathbf{d}^{\text{data}} - \mathbf{d}^{\text{model}})^{\text{T}} \quad (15)$$

where \mathbf{d}^{data} is the observed data vector, $\mathbf{d}^{\text{model}}$ is the emulator prediction, and \mathbf{C} is the covariance matrix, which includes three contributions to the error budget:

$$\mathbf{C} = \mathbf{C}^{\text{data}} + \mathbf{C}^{\text{emu}} + \mathbf{C}^{\text{sim}}. \quad (16)$$

Here, \mathbf{C}^{data} is the term associated with the sample variance of the data vector, which is estimated from multiple realizations of the MD-Patchy mocks:

$$\mathbf{C}^{\text{data}} = \frac{1}{N_{\text{patchy}} - 1} \sum_{k=1}^{N_{\text{patchy}}} (\mathbf{d}_k - \bar{\mathbf{d}}) (\mathbf{d}_k - \bar{\mathbf{d}})^{\text{T}}, \quad (17)$$

where $N_{\text{patchy}} = 2048$.

The simulations used for training are at a fixed phase, which could be different from the true underlying phase of the Universe. In other words, the cosmic variance is frozen in our emulator predictions. To account for this, we add an extra sample variance contribution to the error budget, associated with the finite size of the training simulations. We estimate this covariance using multiple mock realizations of the AbacusSummit fiducial cosmology with a fixed set of HOD parameters with high likelihood:

$$\mathbf{C}^{\text{sim}} = \frac{1}{N_{\text{sim}} - 1} \sum_{k=1}^{N_{\text{sim}}} (\mathbf{d}_k - \bar{\mathbf{d}}) (\mathbf{d}_k - \bar{\mathbf{d}})^{\text{T}}, \quad (18)$$

where $N_{\text{sim}} = 1800$.

\mathbf{C}^{emu} accounts for the intrinsic error in the model predictions due to an imperfect emulation. This term is calculated by computing a covariance matrix from the difference between the emulator predictions and measurements from a set of test simulations with known cosmologies and HOD parameters, $\Delta \mathbf{d}$:

$$\mathbf{C}^{\text{emu}} = \frac{1}{N_{\text{test}} - 1} \sum_{k=1}^{N_{\text{test}}} (\Delta \mathbf{d}_k - \overline{\Delta \mathbf{d}}) (\Delta \mathbf{d}_k - \overline{\Delta \mathbf{d}})^{\text{T}}, \quad (19)$$

where the overline denotes the mean across all test simulations, and $N_{\text{test}} = 600$.

When calculating the likelihood, both \mathbf{C}^{data} and \mathbf{C}^{sim} are multiplied by a factor \mathcal{P} before inversion (Percival et al. 2022):

$$\mathcal{P} = \frac{(N_s - 1)[1 + B(N_d - N_\theta)]}{N_s - N_d + N_\theta - 1}, \quad (20)$$

where

$$B = \frac{N_s - N_d - 2}{(N_s - N_d - 1)(N_s - N_d - 4)}. \quad (21)$$

Here, N_s is the number of simulations used to estimate the covariance, N_d is the length of the data vector and N_θ is the number of

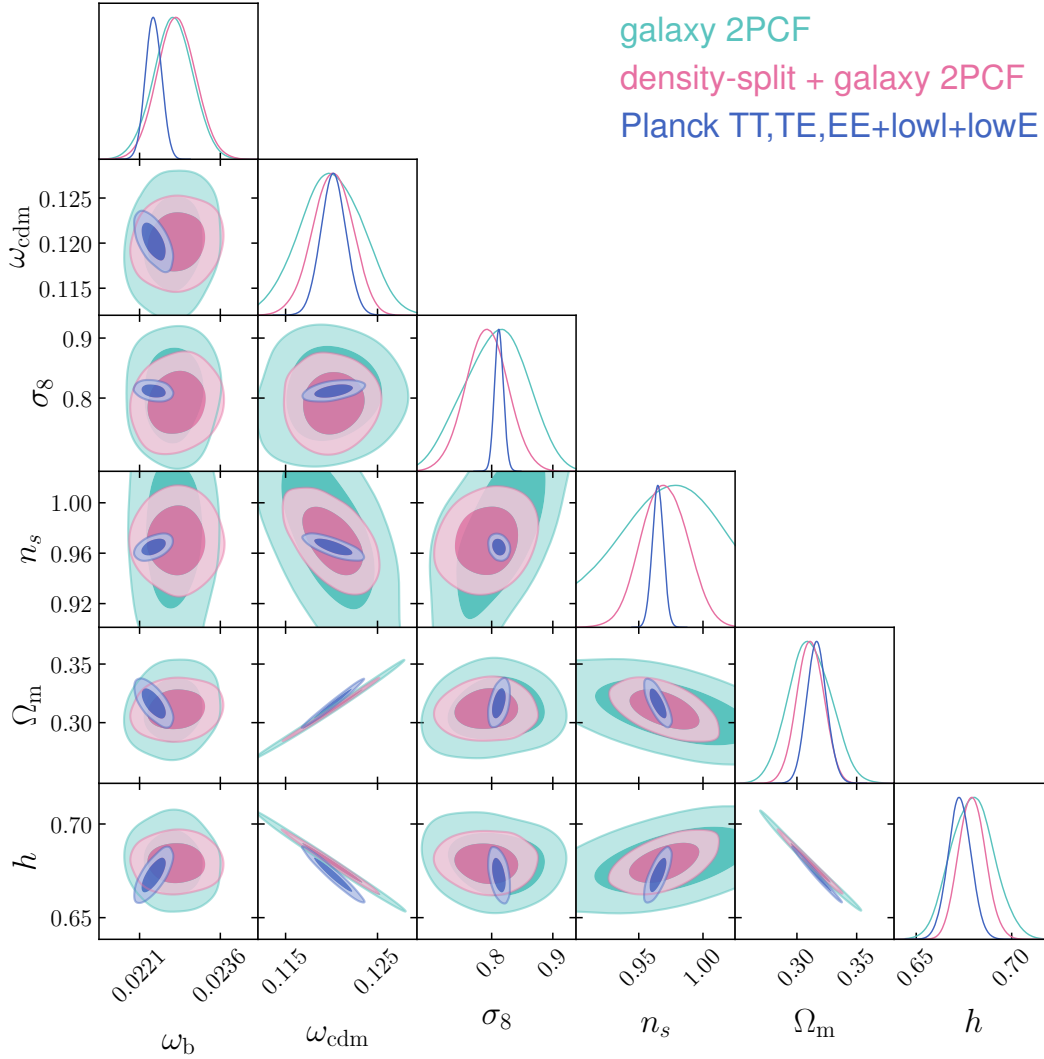


Figure 4. Posterior probability distributions of the base- Λ CDM parameters from the fits to the BOSS DR12 CMASS clustering data. The pink contours show results from the combination of density-split clustering and the galaxy 2PCF, while the aquamarine contours show results using only the latter. We also overplot constraints taken from the PLANCK_TTTEEE_LOWL_LOWE likelihood (Planck Collaboration et al. 2020) in blue. All 2D contours show 68 and 95 per cent confidence intervals. $\color{blue}{\bullet}$

parameters that are being fitted. This corrects for the fact that even though Eq. (17) & (18) are unbiased estimates of the covariance matrices, the inversion leads to biased parameter constraints.

We sample the posterior distribution of parameters using the DYNesty nested sampler (Speagle 2020), which also provides an estimation of the Bayesian evidence. As listed in Table 1, we assume flat priors for all parameters except for the baryon density, for which we adopt a BBN-informed Gaussian prior (Aver et al. 2015; Cooke et al. 2018; Schöneberg et al. 2019) for our baseline analysis:

$$\omega_b = 0.02268 \pm 0.00038. \quad (22)$$

In later sections, we also explore using a flat prior range for ω_b , finding that it has little impact on our conclusions and the reported constraints for other parameters.

3.5 θ_* prior

In the simulations used to train our emulator, the value of the dimensionless Hubble parameter h was chosen to fix the CMB acoustic

Table 2. The χ^2 , degrees of freedom, log-likelihood, and log-evidence for the best fits to the BOSS DR12 multipoles from the galaxy two-point correlation function (2PCF), density-split clustering (DSC) and the combination of the two. We simultaneously fit the monopole and quadrupole, using scales $1 h^{-1} \text{Mpc} < s < 150 h^{-1} \text{Mpc}$.

Statistic	χ^2	dof	χ^2/dof	$\log \mathcal{L}$	$\log \mathcal{Z}$
2PCF + DSC	1052	635	1.66	-257.58	-276.38
DSC	942	563	1.67	-253.15	-270.85
2PCF	63	59	1.06	-29.63	-38.03

scale θ_* ($100\theta_* = 1.041533$) matching the best-fit measurement from PL18 (Planck Collaboration et al. 2020; Maksimova et al. 2021). This means that effectively we apply a fixed θ_* constraint as a prior on our emulator results. This ensures that we only consider models within the part of parameter space where the emulator has been trained.

As a result, h is not a free parameter in our model, but is determined by the sampled parameters and the θ_* constraint. We use CLASS (Blas

et al. 2011) to derive h at each point in our chains, and we then use this to also obtain Ω_m , calculated as (Dodelson 2003)

$$\Omega_m = (\omega_b + \omega_{\text{cdm}} + \omega_\nu)/h^2, \quad (23)$$

where $\omega_\nu = 0.00064420$ accounts for 60 meV neutrinos (also the choice of the base PL18 cosmology), which is always fixed in our model.

A consequence of this prior on θ_* is that the parameter constraints we obtain below do not come exclusively from late-Universe clustering measurements but rather from a combination of galaxy clustering and information on the CMB acoustic scale, although they do not use other CMB information. The simple geometrical interpretation of θ_* means that it is one of the best-measured quantities in all cosmology (the PL18 measurement corresponds to a 0.03 per cent precision level), and this measurement is also extremely robust to changes in the cosmological model (Planck Collaboration et al. 2020). We therefore consider this to be a very well-justified prior.

4 RESULTS

In this section, we apply our emulator framework to the CMASS clustering data and present our main cosmological constraints, along with tests for model systematics. We focus on the constraints for cosmological parameters and reserve the discussion about HOD constraints for Appendix A.

4.1 Model fits

The solid lines in Fig. 3 show the best-fit base- Λ CDM model to the measured multipoles, with the χ^2 , likelihood, and Bayesian evidence values reported in Table 2. The χ^2 per degree of freedom for the galaxy 2PCF + DSC combination is 1.66, while for the 2PCF-only fit it is 1.06, which is comparable to the reported values for the small-scale 2PCF fit from Yuan et al. (2022b) and the configuration-space BAO fit from Ross et al. (2016).

We observe some hints of the models under-predicting (or over-predicting, in the case of negative density contrasts) the observed clustering at scales $s > 90 h^{-1}\text{Mpc}$. Such excess large-scale clustering has been observed before in the galaxy 2PCF analysis of BOSS data. In Ross et al. (2016), the authors find that the observed monopole shows an apparent excess with respect to the mean of the MD-Patchy mock catalogues, although it is argued that the mismatch is of low statistical significance. In Satpathy et al. (2017), the authors fit the correlation function using a model based on Convolutional Lagrangian Perturbation Theory, also finding that the model under-predicts the monopole data at large scales. Here, we observe a similar trend when comparing the measured multipoles to our emulator predictions, although the level of discrepancy is never greater than two standard deviations for all scales considered, irrespective of the summary statistic. As informed by the χ^2 values, the quality of the fit is better than might be guessed by the eye, since the separation bins are correlated. In Sect. 4.4, we show that the mean clustering signal of the Nseries mocks monopole is also lower than the CMASS data, although still consistent within 2σ . We also observe that the emulator is able to fit the mean of the mocks with much higher accuracy than the data, which we deem reasonable given that our model was trained to make predictions for the ensemble average of the clustering statistics.

Although some of the observed offsets between our best-fit model and CMASS could be partially attributed to statistical fluctuations,

we should bear in mind the possibility that there are residual systematic effects in the clustering data that are not fully corrected by the weighting procedure. Lavaux et al. (2019) performed a field-level inference of BOSS data, finding evidence of residual systematics on the spectroscopic data that so far remain unexplained, which can induce correlated modulations of the order of 30 per cent on the sky for CMASS. In addition, it is worth noting that the systematic weights for the large-scale structure catalogues provided by the BOSS collaboration were mainly validated for two-point statistics, and it is possible that they are sub-optimal for alternative clustering methods such as DSC. We plan to explore this topic further in future work.

4.2 Base- Λ CDM constraints

In this section, we explore the constraining power of DSC and the galaxy 2PCF on the base Λ CDM parameters, letting ω_b , ω_{cdm} , σ_8 , n_s , and the HOD parameters vary during the fit. To generate the model predictions, we query the emulator fixing the remaining cosmological parameters to their baseline values $w_0 = -1$, $w_a = 0$, $dn_s/d \ln k = 0$, and $N_{\text{eff}} = 3.0146$.

Figure 4 shows the 2D posterior distributions on Λ CDM, marginalized over the HOD parameters. We also over-plot the base- Λ CDM constraints from the PLANCK_TTTEEE_LOWL_LOWE likelihood from PL18 to facilitate comparison. The reported best-fit values, along with the means of the marginalized posteriors and their dispersion are listed in Table 3. It is worth noting that the best-fit values can be shifted with respect of the mean of the posterior due to its non-Gaussian shape.

The galaxy 2PCF by itself constrains ω_{cdm} , σ_8 , and n_s with a 2.7, 6.2, and 3.2 per cent precision, respectively. Adding the DSC multipoles tightens the constraining power to a precision of 1.7, 3.8 and 1.8, respectively. The constraints for baryon density are entirely dominated by the BBN prior. Overall, both posterior distributions are largely consistent with the PL18 results, with a 0.04σ and 0.6σ difference in the mean values of ω_{cdm} and σ_8 between our baseline fit and PL18.

As described in Sect. 3.5, we use the θ_* constraint to obtain h and Ω_m as derived parameters from our chains, and also show the results for them in Fig. 4. We obtain $h = 0.6793 \pm 0.0070$ and $\Omega_m = 0.311 \pm 0.011$ when fixing to the base Λ CDM model. As Fig. 4 shows, we do not observe any significant degeneracy between Ω_m and the baryon density (and our constraints on ω_b are also prior-dominated). This means that, to a good approximation, $\theta_* \propto \Omega_m h^{3.4}$, the expected behaviour for fixed $\Omega_b h^2$ within flat- Λ CDM models (Percival et al. 2002), and thus DSC and the galaxy 2PCF results both describe this narrow degeneracy direction in the Ω_m - h plane. This is close to but not the same as the similarly tight degeneracy obtained from the the CMB by PL18, which corresponds to constant $\Omega_m h^3$. The degeneracy direction for models with the same θ_* depends weakly on $\Omega_b h^2$ through the sound horizon. PL18 uses the CMB data to also constrain $\Omega_b h^2$, leading to an anti-correlation between constraints on Ω_b and Ω_m , which then results in a different degeneracy in the Ω_m - h plane.

Within Λ CDM, the linear growth rate of structure can be approximated as

$$f(z) \approx \Omega_m^{0.55}(z). \quad (24)$$

We use this expression to derive a value for $f\sigma_8$ at the effective redshift of CMASS from our chains.⁵ We obtain $f\sigma_8(z = 0.525) =$

⁵ As we are doing a direct fit of Λ CDM parameters and marginalizing over

Table 3. Parameter constraints for the base Λ CDM fits on the BOSS DR12 CMASS galaxy catalogue, using the combination of density-split clustering and the galaxy 2PCF, or using only the latter.

Parameter	density-split + galaxy 2PCF		density-split		galaxy 2PCF	
	best-fit	mean $\pm \sigma$	best-fit	mean $\pm \sigma$	best-fit	mean $\pm \sigma$
ω_b	0.0234	0.02279 ± 0.00035	0.0233	0.02283 ± 0.00035	0.0226	0.02272 ± 0.00036
ω_{cdm}	0.1209	0.1201 ± 0.0022	0.1187	$0.1191^{+0.0026}_{-0.0022}$	0.1190	0.1200 ± 0.0034
σ_8	0.8116	0.792 ± 0.034	0.7715	$0.768^{+0.037}_{-0.043}$	0.7591	$0.807^{+0.055}_{-0.049}$
n_s	0.9837	0.970 ± 0.018	0.9952	0.968 ± 0.023	0.9515	$0.969^{+0.043}_{-0.026}$
h	0.6774	0.6793 ± 0.0070	0.6846	$0.6828^{+0.0070}_{-0.0083}$	0.6832	0.679 ± 0.011
Ω_m	0.3157	0.311 ± 0.011	0.3044	0.306 ± 0.012	0.3047	0.311 ± 0.018
$f\sigma_8$	0.4747	0.462 ± 0.020	0.4482	$0.447^{+0.021}_{-0.025}$	0.4411	$0.470^{+0.033}_{-0.028}$

0.462 ± 0.020 , which is a 4.3 per cent constraint from the combination of DSC and the galaxy 2PCF. We contrast this constraint with the PL18 prediction and other clustering studies in Fig. 5. Our constraint is slightly lower than the mean of PL18, but consistent at the 1σ level. We also find good agreement with the results of the main BOSS clustering analysis (Alam et al. 2017), but we get a factor of 1.9 better precision using the DSC + galaxy 2PCF combination, compared to their middle redshift bin that covers $0.4 < z < 0.6$.

We find a higher mean $f\sigma_8$ than Yuan et al. (2022b) (from hereon Y22), although there is still consistency at the 0.7σ level. Interestingly, Y22 constrain $f\sigma_8$ to a precision similar to our baseline analysis, even though they only use the galaxy 2PCF at scales $< 30 h^{-1}\text{Mpc}$. Our 2PCF-only fit is a factor of 1.8 worse than theirs, which seems puzzling given that their model is based on the same set of simulations used in this study. The main reason we attribute this difference to is that while we estimate the intrinsic emulator error using a test set of simulations that covers the whole prior range in the cosmology & HOD parameter space, Y22 estimate the error using only HOD catalogues that have a high likelihood with respect to the measured data vector. Our emulator error will generally tend to be more conservative, as it also considers the error around regions near the edge of the priors, which are usually harder to emulate. It also dominates the error budget at small scales, precisely where Y22 extract most of their information. Effectively, this means we are discarding some small-scale information which they keep. We have verified that by artificially lowering or removing the emulator error, the precision of our galaxy 2PCF constraints closely matches that of Y22. Other reasons that could potentially add to this are that Y22 use Gaussian priors on the HOD parameters while we use flat priors, and that Y22 estimate the data covariance through jackknife resampling, while our covariance is estimated from mock catalogues.

Our results are 2.2σ higher than those reported by Zhai et al. (2023), who find $f\sigma_8(z = 0.55) = 0.396 \pm 0.022$. Zhai et al. (2023) train an emulator using the AEMULUS suite of simulations to fit the small-scale clustering of BOSS galaxies using the galaxy 2PCF between $0.1 h^{-1}\text{Mpc}$ and $60 h^{-1}\text{Mpc}$. Although we have not tested our emulator against AEMULUS directly, our companion paper (Cuesta-Lazaro et al. 2023) shows that our model can recover unbiased cosmological constraints when fitted to mock galaxy catalogues generated with a different N-body code and galaxy-halo connection model, specifically galaxies generated with the subhalo abundance matching technique by Zhai et al. (2023) from the Uchuu suite of simulations

h , our results are not affected by the problems of the standard template-based clustering analyses pointed out in Sánchez (2020).

Table 4. Constraints on single-parameter extensions to the base- Λ CDM model from the BOSS DR12 CMASS data, using the combination of density-split clustering and the galaxy 2PCF.

Parameter	best-fit	mean $\pm \sigma$
N_{eff}	3.0759	$3.02^{+0.24}_{-0.27}$
$dn_s/d \ln k$	-0.0022	-0.005 ± 0.013
w_0	-0.9444	$-0.956^{+0.046}_{-0.041}$

(Ishiyama et al. 2021; Dong-Páez et al. 2022; Oogi et al. 2023; Aung et al. 2023; Prada et al. 2023). Applying our emulator to AEMULUS could be a useful step to explore in future work to better understand the source of this discrepancy.

Yu et al. (2023) use Eulerian perturbation theory in combination with a halo model calibrated on N-body simulations to model the power spectrum multipoles from BOSS DR12 up to $k = 0.2 h\text{Mpc}^{-1}$, finding $f\sigma_8(z = 0.61) = 0.455 \pm 0.026$, which is in excellent agreement with our constraint, albeit being at an effective redshift slightly higher than our sample.

Comparing with a full-shape analysis based on the Effective Field Theory of Large-Scale Structures (EFTofLSS, Carrasco et al. 2012), we find an $f\sigma_8$ value than is 1.7σ higher than the one derived by d’Amico et al. (2020), who fit the power spectrum multipoles of BOSS DR12 galaxies up to $k = 0.2 h\text{Mpc}^{-1}$ and report $f\sigma_8(z = 0.55) = 0.399 \pm 0.031$, which is closer to the Zhai et al. (2023) estimate. Although our Ω_m constraints agree to within 0.1σ with d’Amico et al. (2020), their predicted amplitude of the primordial power spectrum, A_s , which is 2.3σ lower than PL18, drives their derived $f\sigma_8$ to lower mean values.

Finally, Semenaite et al. (2022) presented a clustering analysis of BOSS and eBOSS data, using information of the full shape of the BOSS clustering wedges presented by Sánchez et al. (2017), and the multipoles from eBOSS quasars from Hou et al. (2021). They derive $\sigma_8 = 0.815 \pm 0.044$ and $\Omega_m = 0.290^{+0.012}_{-0.014}$, which agree with our constraints at the 0.4σ and 1.2σ level, respectively.

4.3 Extended- Λ CDM constraints

To explore potential deviations from Λ CDM, we have analyzed a grid of well-motivated extensions to the base model.

Figure 6 and Table 4 summarize the results of single-parameter extensions to the base Λ CDM model. All these fits have been run using the baseline configuration of data vectors, scale cuts and model prescription, except for the addition of a single parameter to the cosmological model, which are incorporated one at a time. We do

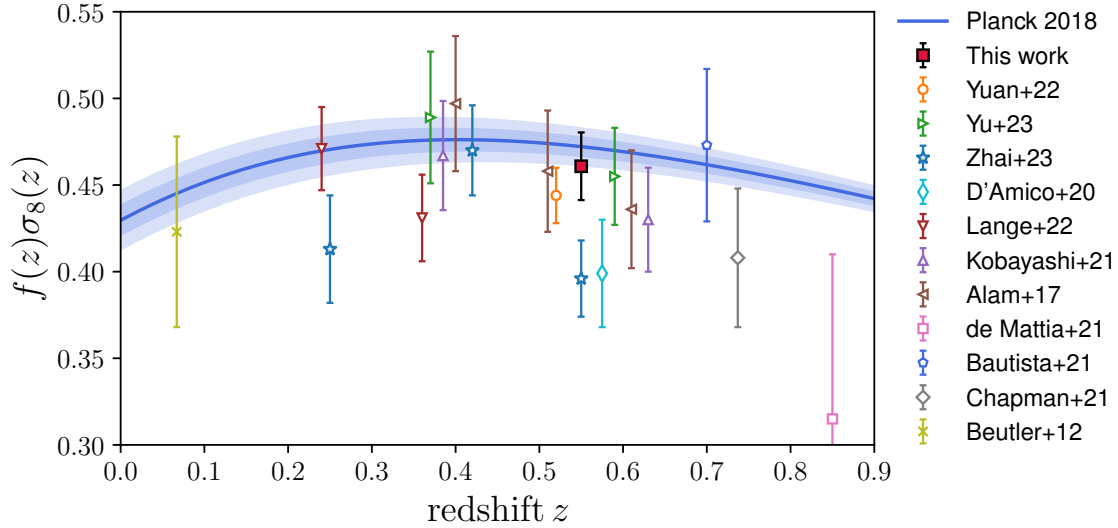


Figure 5. Growth rate of structure derived from our analysis of the BOSS DR12 CMASS sample, combining density-split clustering and galaxy 2PCF measurements. The base- Λ CDM prediction, based on the best-fit Planck 2018 cosmology, is shown in blue, where the darker and lighter shades represent 68 and 95 per cent confidence intervals. We also compare with other measurements from the literature at different redshifts. The data include 6dFGS (Beutler et al. 2012), eBOSS (de Mattia et al. 2021; Bautista et al. 2021; Chapman et al. 2022), as well as other studies performed on BOSS: the consensus DR12 analysis (Alam et al. 2017), a LOWZ simulation-based analysis (Lange et al. 2022), CMASS simulation-based analyses (Yuan et al. 2022b; Kobayashi et al. 2022; Zhai et al. 2023), an EFTofLSS analysis (d’Amico et al. 2020), and a halo perturbation theory analysis (Yu et al. 2023). In some cases, the redshifts of the measurements have been slightly shifted horizontally for visual clarity.

not find compelling proof supporting any of these extensions, as the marginalized posteriors for the additional parameters generally overlap with the base model within one standard deviation.

4.3.1 Running of the spectral index

The first extension we look at is in regard to the scale dependence of the primordial density fluctuations. Here we characterize the primordial power spectrum as a power law with a normalization amplitude A_s , a spectral index n_s and its first derivative with respect to $\ln k$ (also known as *the running*):

$$P(k) = A_s \left(\frac{k}{k_0} \right)^{n(k)} \quad (25)$$

$$n(k) = n_s - 1 + (1/2)(dn_s/d \ln k) \ln(k/k_0), \quad (26)$$

where k_0 is a pivot wavenumber used to specify the point at which the power spectrum is normalized. Our base- Λ CDM constraints from Sect. 3 have assumed a zero running for the spectral index, finding $n_s = 0.970 \pm 0.018$. CMB experiments also favour $n_s < 1$, which is predicted by common single-field slow-roll inflationary models (Mukhanov 2007). Such models also predict a very small running, since it is only second-order in inflationary slow-roll parameters (Kosowsky & Turner 1995), but it is possible to construct valid models that predict larger values. We find $dn_s/d \ln k = -0.005 \pm 0.013$, which is consistent with zero running. The 1D marginalized posterior is shown in the left-hand side panel of Fig. 6, where we also overplot the constrain from the PLANCK_TTEEE_LOWL_LOWE likelihood from PL18, who find -0.0055 ± 0.0067 . We warn the reader that the parameter range for the running shown in Fig. 6 corresponds to our full prior range, beyond which we do not have simulations to sample the parameter space with our model. The posterior distribution quickly approaches zero near the prior walls, giving us

confidence that our 1σ limits are not prior-dominated. However, we cannot confidently provide 3σ limits given this restriction.

4.3.2 Effective number of relativistic species

A relic neutrino background is a generic prediction of the standard hot Big Bang model (Lesgourgues & Pastor 2014). The constraints on the properties of relic neutrinos and other relativistic species beyond the Standard Model of particle physics is of special interest for large scale structure analyses. The combination of CMB experiments, galaxy and supernova surveys have put the tightest upper limits on the sum of neutrino masses (Emas & Wulandari 2019), whereas Planck has constrained the density of light relics with sub-percent level precision (Planck Collaboration et al. 2020).

In the instantaneous neutrino decoupling limit, the density of radiation in the Universe (besides photons) can be written as (Lesgourgues & Pastor 2014):

$$\frac{\rho_\nu}{\rho_\gamma} = \frac{7}{8} N_{\text{eff}} \left(\frac{4}{11} \right)^{4/3}, \quad (27)$$

where $N_{\text{eff}} = 3$, usually called the effective number of relativistic species, is a convenient parametrization of the relativistic energy density of the Universe beyond just photons, in units of a single neutrino. Detailed calculations that go beyond the instantaneous neutrino decoupling limit, including neutrino oscillations, predict $N_{\text{eff}} \approx 3.046$ (de Salas & Pastor 2016).

For the base- Λ CDM constraints from Sect. 3, we fixed N_{eff} to the baseline value of 3.046. In this section, we vary this parameter during the inference analysis, finding $N_{\text{eff}} = 3.02^{+0.24}_{-0.27}$, which is a 8.4 per cent precision constraint, in excellent agreement with the PL18 measurement.

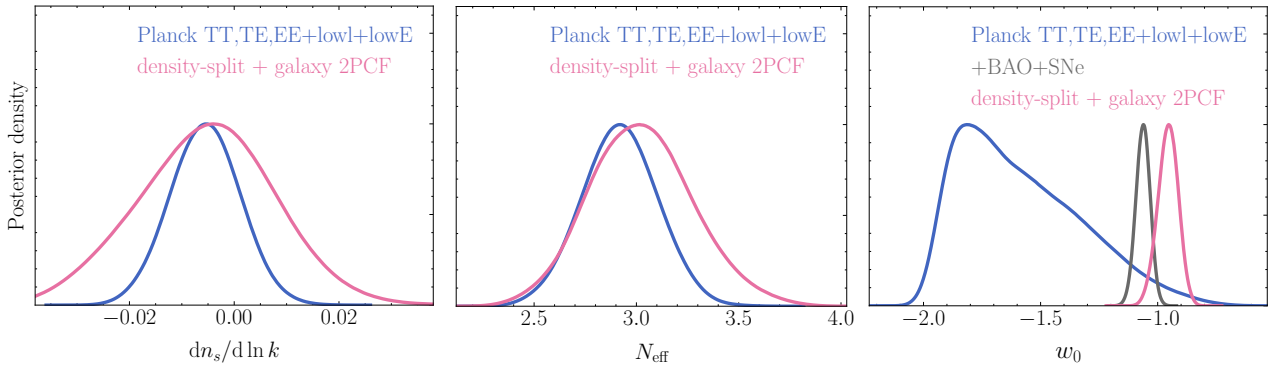



Figure 6. Constraints on single-parameter extensions to the base- Λ CDM model from the galaxy 2PCF + density-split clustering fits on BOSS DR12 CMASS (pink), Planck TT,TE,EE+lowl+lowE (blue), and Planck TT,TE,EE+lowl+lowE+BAO+SNe (grey). Extensions include variations in the running of the spectral index of the primordial power spectrum, $dn_s/d \ln k$, the effective number of relativistic species, N_{eff} , and the dark energy equation of state parameter, w_0 . 

4.3.3 Dark energy equation of state

One of the main goals of modern observational cosmology is elucidating the nature of the accelerated expansion of the Universe. In the base- Λ CDM model, dark energy takes the form of a cosmological constant that has an equation of state $w_0 \equiv p/\rho = -1$, where p and ρ represent the pressure and density of the fluid. In this section, we explore models with a constant w by letting it vary as a free parameter in the fit.

The right-hand side panel of Fig. 6 shows the marginalized posterior of w for the combination of the galaxy 2PCF and DSC. We find $w_0 = -0.956^{+0.046}_{-0.041}$, which is consistent with a fiducial cosmological constant at the 1σ level. CMB data alone (blue contour) does not put a very tight constraint on w_0 , as it is a $z \approx 1100$ measurement. The grey contours show the posterior resulting from the combination of the PLANCK_TTTEEE_LOWL_LOWE PL18 likelihood with late-time probes of the expansion rate, including BAO measurements from BOSS DR12 (Alam et al. 2017), SDSS-MGS (Ross et al. 2015), and 6dFGS (Beutler et al. 2011), as well as type Ia supernovae distance measurements from the PANTHEON sample (Scolnic et al. 2015), and local estimates of the Hubble parameter from Milky Way Cepheid variables from Riess et al. (2018). This combination tightens the marginalized posterior to -1.063 ± 0.031 , which agrees with our constraint at the 2σ level.

We emphasize that for our results from galaxy clustering, we have adopted the prior constraint for the acoustic scale θ_* from PL18, so our constraints on w_0 are not fully independent from the PLANCK_TTTEEE_LOWL_LOWE likelihood.

4.4 Test for systematics

In this section, we carry out tests on mock galaxy catalogues to look for systematics in the theoretical modelling, and we assess the robustness of our cosmological constraints to different choices in our inference pipeline.

4.4.1 Recovery tests on the Nseries mocks

We begin by testing our pipeline on the Nseries mock galaxy catalogues, which were calibrated onto the clustering of the CMASS NGC galaxy sample, matching its footprint and radial selection (Sect. 3.1.2). We measure the galaxy 2PCF and density-split multipoles from each of the 84 mock realizations, and analyze each mock using the baseline configuration of our pipeline as we did in

Sect. 3. We estimate the covariance matrix of the data vectors from the MD-Patchy mocks, following the same procedure as described in Eq. (17).

The left panel of Fig. 8 shows the distribution of recovered best-fit values from the 84 fits. The true cosmology of the simulations, which is shown by the vertical red lines, is well within the 68 per cent confidence region of the distribution, showing that our clustering pipeline is able to recover unbiased cosmological constraints even in the presence of complex survey masks, fiber collisions and non-uniform radial selection functions.

As a complementary test, we proceed to average the data vectors across the 84 mock realizations, and perform cosmological fits on the mean data vectors. The right panel of Fig. 8 shows the recovered cosmological parameters in this setup, using two different covariance matrices. The grey contours show results with the usual covariance matched to the volume covered by the CMASS sample used throughout the paper. The black contours show results where the covariance is rescaled to match the comoving volume covered by 84 realizations⁶ of the Nseries suite, which amounts to roughly $120 (h^{-3} \text{Gpc}^3)$, after applying the redshift cuts to match our data sample, $0.45 < z < 0.6$. Even using this large volume, the true cosmology of the simulations falls within one standard deviation of the marginalized parameter posterior distributions, highlighting the robustness of our pipeline even for datasets far larger than the one analysed in this paper.

Figure 10 shows the best fit to the galaxy 2PCF measured from the mean of the Nseries samples (using the covariance associated to a single CMASS volume), where we also overplot the measurement from the CMASS sample for comparison. Overall, the fit to the Nseries mocks is accurate at all scales, with the best-fit model always falling within one standard deviation of the error bars. This can be contrasted with the fit to the CMASS data seen earlier in Sect. 4.1, where the data measurement shows a lower and higher clustering amplitude at intermediate and large scales compared to the model, respectively. CMASS also shows a similar difference in clustering with respect to the mean of the Nseries mocks. The χ^2 between the mocks and data monopole is 30 for 23 degrees of freedom, which corresponds roughly to a 2σ shift. Although we do not show the other statistics for brevity, we have observed the same trends for DSC. However, it is worth keeping in mind that the cosmology of the

⁶ It should be noted however that the 84 realizations from Nseries are not fully independent from each other, since they were generated from only seven large cubic boxes that were rotated and trimmed in different ways to construct the cutsy mocks.

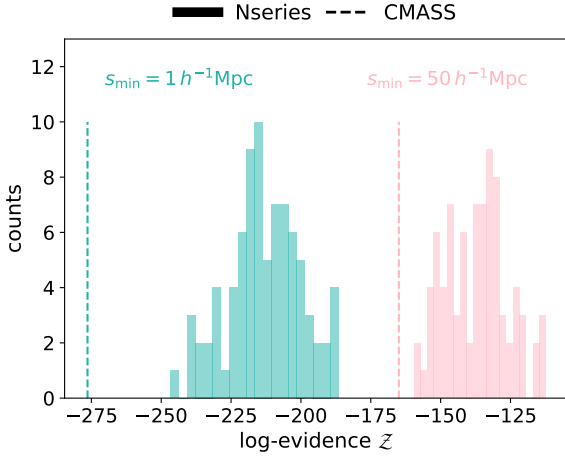


Figure 7. Logarithm of the Bayesian evidence obtained from individual fits to the Nseries mocks (histograms) and to the BOSS DR12 CMASS data (dashed lines). We show results for two different minimum scale cuts: $s_{\min} = 1 h^{-1}\text{Mpc}$ (green) and $s_{\min} = 50 h^{-1}\text{Mpc}$ (pink). The evidence of the CMASS fit is significantly lower than any of the Nseries mocks in both cases, but this discrepancy is reduced when excluding small scales. $\color{blue}\text{\textcircled{C}}$

Nseries simulations differs substantially from the fiducial cosmology we adopted to convert redshifts to distances, so the Nseries multipoles could be affected by more severe AP distortions than the CMASS data, if the true cosmology of the Universe is closer to our fiducial cosmology compared to the mocks.

Figure 7 shows the distribution of log-evidence from the fits to the 84 Nseries mocks and CMASS. We include the results obtained using two different scale cuts: $s_{\min} = 1 h^{-1}\text{Mpc}$ and $s_{\min} = 50 h^{-1}\text{Mpc}$. We see that regardless of the scale cut that is used, the evidence of the CMASS fit is significantly lower than for any of the Nseries mocks. This means that, overall, the model is a much better description of the mocks than it is of the data, even when marginalizing over all the parameter space of the model. This supports the hypothesis that there might be residual systematic effects in the CMASS clustering catalogues that are currently unaccounted for by the weighting procedure (Lavaux et al. 2019). While removing all the information below $s = 50 h^{-1}\text{Mpc}$ reduces the gap between Nseries and CMASS, it is still statistically unlikely that the large offsets between the baseline model and the data can be fully explained by a random fluctuation due to sample variance from the observations, if we assume that the mocks are a faithful reproduction of the CMASS data. This could be a combination of the above-mentioned systematic effects with insufficient models of the halo-galaxy connection on small scales. Given the small number of Nseries mocks, we cannot translate these findings into more quantitative statements.

4.4.2 Scale cuts

Another important aspect we study with the Nseries mocks is the choice of scales that are used in the main data analysis. Figure 9 shows the marginalized constraints on ω_{cdm} , σ_8 , n_s , and $f\sigma_8$, adopting eight different minimum scale cuts, ranging from $1 h^{-1}\text{Mpc}$ to $60 h^{-1}\text{Mpc}$. For these measurements, we use the mean of the 84 Nseries mocks as the data vector, and a covariance associated to a single CMASS volume. We observe that for Nseries, the mean values of the marginalized posteriors are very stable when changing the scale cuts for all parameters that are considered. It is interesting to note

that the size of the error bars does not shrink beyond $\approx 20 h^{-1}\text{Mpc}$. There are two main reasons for this. The first one is the inclusion of the model uncertainty in the covariance matrix that is used for the calculation of the likelihood [Eq. (16)]. At these scales, although the emulator error has a percent-level accuracy, it starts to dominate the total error budget for the monopole of the 2PCF and the density-split CCF. This ensures that the cosmological inference is always robust, even when including scales where the emulator is less accurate than the precision of the data, which comes at the expense of not being able to extract all the information there is available. The second reason is that even though we are imposing a minimum scale cut in the multipoles, the DSC quintiles are defined using small-scale information from the density field, which also propagates into the multipoles at larger scales, as shown in Paillas et al. (2023).

Figure 9 also shows how the constraints from CMASS data change depending on the minimum scale cut. Although the results we get using $s_{\min} = 1 h^{-1}\text{Mpc}$ are consistent with those obtained with more conservative cuts to within 1σ , we observe some interesting trends with scale. ω_{cdm} shows some tendency towards larger mean values the more small-scale information we include. Furthermore, σ_8 transitions to larger mean values going from $s_{\min} = 20 h^{-1}\text{Mpc}$ to $s_{\min} = 5 h^{-1}\text{Mpc}$. Interestingly, this transition takes place near the scale that is used to smooth the density field when defining the quintiles. The fact that these effects are not observed in the Nseries mocks could be partially attributed to noise in the CMASS data vector, residual observational systematic effects that the mocks do not capture, or misunderstandings in the galaxy-halo connection modelling, which is mostly constrained by small scales. One of the most common systematics that can affect small-scale clustering is fiber assignment, which artificially lowers the clustering amplitude on scales smaller than the fiber collision angular scale. The Nseries mocks are already infused with fiber collisions that should closely match the BOSS data, so any important effect coming from this should also be imprinted in the multipoles measured from the mocks. Regarding the potential impact of the smoothing scale, it is beyond the computational resources allocated for this work to explore the dependence of our cosmological constraints on the choice of smoothing scale, given that it would require the generation of new training sets with different scale choices, which is computationally demanding. However, we plan to explore this effect in more detail in future work.

Based on the fact that the constraints from Nseries are robust against variations in s_{\min} , and that the CMASS constraints with different scale cuts are consistent to within 1σ , we adopt $s_{\min} = 1 h^{-1}\text{Mpc}$ as the baseline for our analysis.

4.4.3 Systematic error budget

Based on the tests described in the previous section, we use the Nseries mocks to determine the contribution to the systematic error budget coming from our emulator. This was trained on periodic boxes at a fixed redshift, but we are applying it to fit a survey that has a non-uniform footprint and radial selection, along with fiber collision effects that artificially decrease the clustering on small scales. Thus, there is the possibility of errors in addition to the emulator error term determined from the test sample when constructing the emulator.

To look for such error terms, we calculate the offset between the expected value of each cosmological parameter and the mean of the marginalized posteriors from the fit to the mean of 84 mocks, using the covariance of a $120(h^{-3}\text{Gpc}^3)$ volume (solid contours from the right-hand side panel of Fig. 8). For the combination of DSC and the

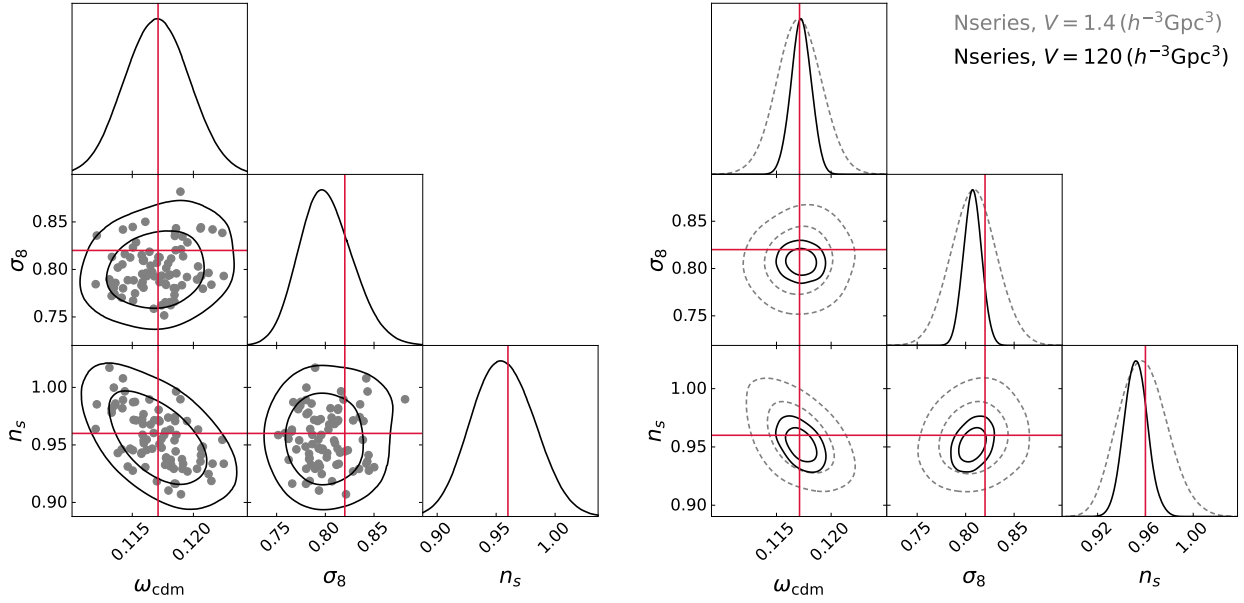


Figure 8. Marginalized constraints on ω_{cdm} , σ_8 , and n_s from fits to the Nseries mock catalogues, which were calibrated onto the clustering of the CMASS NGC sample, matching its number density and geometry. The fits were performed using the baseline configuration of our analysis, consisting on the combination of the density-split and the galaxy correlation functions at scales $1.0 h^{-1}\text{Mpc} < s < 150 h^{-1}\text{Mpc}$.

Left: the dots represent the best-fit values of individual fits to 84 realizations of the Nseries mocks. The contours show the 68 and 85 per cent confidence regions of the distribution of individual fits. Right: fits to data vectors that are averaged over the 84 Nseries mocks, using a covariance rescaled to match the volume of the CMASS sample used in our main analysis (dashed-grey) or the total volume of the Nseries suite (solid-black).

galaxy 2PCF, the offsets and their associated 2σ uncertainties are:

$$\begin{aligned}\Delta\omega_{\text{cdm}} &= 0.00015^{+0.0018}_{-0.0018} \\ \Delta\sigma_8 &= 0.01288^{+0.018}_{-0.018} \\ \Delta n_s &= 0.00801^{+0.020}_{-0.019} \\ \Delta f\sigma_8 &= 0.00657^{+0.011}_{-0.011} \quad (\text{DSC} + \text{galaxy 2PCF}),\end{aligned}$$

while the fits that only include the galaxy 2PCF give

$$\begin{aligned}\Delta\omega_{\text{cdm}} &= 0.00002^{+0.0040}_{-0.0039} \\ \Delta\sigma_8 &= 0.01151^{+0.046}_{-0.045} \\ \Delta n_s &= 0.00212^{+0.039}_{-0.038} \\ \Delta f\sigma_8 &= 0.00607^{+0.028}_{-0.027} \quad (\text{galaxy 2PCF}).\end{aligned}$$

We do not find statistically significant offsets from the expected values, with the shifts for all parameters appearing well within their 2σ limits. However, the level of precision to which we can carry out this test is set by the number of mocks that are available, which limits the total effective volume used in the test. We take the 2σ limit of the marginalized distributions of best-fit values (left-hand side of Fig. 8), divided by $\sqrt{84}$, as a conservative (maximum) estimate of the systematic error for each parameter, which are then added in quadrature to the associated statistical errors. We estimate this from the distribution of individual fits divided by $\sqrt{84}$ rather than from the fit to the mean, because the 2σ limits of the latter are dominated by the emulator error on the multipoles, which is already included in the covariance when we calculate the likelihood [Eq. (16)]. In this way, we avoid double counting the emulator uncertainty, only including the limiting precision that comes from the finite number of mocks used for the test. With this setup, the systematic error added to each parameter for the combination of DSC + galaxy 2PCF combination

is:

$$\begin{aligned}\sigma_{\omega_b}^{\text{sys}} &= 0.0000156 \\ \sigma_{\omega_{\text{cdm}}}^{\text{sys}} &= 0.00037 \\ \sigma_{\sigma_8}^{\text{sys}} &= 0.00370 \\ \sigma_{n_s}^{\text{sys}} &= 0.00354 \quad (\text{DSC} + \text{galaxy 2PCF}),\end{aligned}$$

while for the 2PCF-only fits we get

$$\begin{aligned}\sigma_{\omega_b}^{\text{sys}} &= 0.0000074 \\ \sigma_{\omega_{\text{cdm}}}^{\text{sys}} &= 0.00046 \\ \sigma_{\sigma_8}^{\text{sys}} &= 0.00519 \\ \sigma_{n_s}^{\text{sys}} &= 0.00348 \quad (\text{galaxy 2PCF}).\end{aligned}$$

These systematic errors are added in quadrature to the statistical error budget. The systematic errors are negligible compared to the statistical errors on the parameter constraints from CMASS, so the values reported in Table 3 are unaffected, up to the significant figures that are shown.

4.4.4 Robustness against pipeline settings

Here we explore the robustness of our clustering analysis against different choices of settings in the inference pipeline. The tests on this section are performed on the real data, and no longer using the Nseries mocks. As a reminder, our baseline configuration consists of:

- Data vector: Concatenation of the monopole and quadrupole moments of the DSC auto-correlation and cross-correlation functions, using four quintiles (Q_0 , Q_1 , Q_3 , and Q_4) and the galaxy 2PCF.

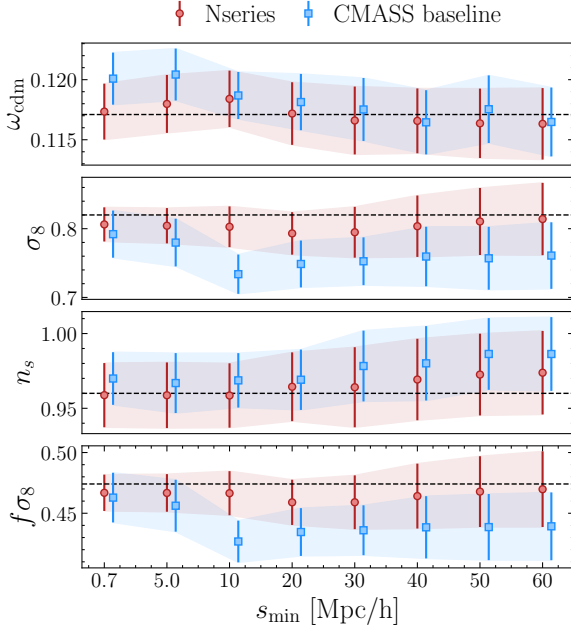


Figure 9. Impact of the minimum scale cut on the constraints on base- Λ CDM parameters and $f\sigma_8$ when fitting the mean of the Nseries mocks (red circles), or the BOSS DR12 CMASS data blue squares). The horizontal dashed line shows the true cosmology of the Nseries simulations. Note that the horizontal axes do not use a linear scale. \odot

- Parameter space: Base- Λ CDM model with an extended HOD framework, including velocity bias and environment-based assembly bias:

$$\theta_{\text{cosmo}} = \{\omega_{\text{cdm}}, \omega_b, \sigma_8, n_s\}$$

$$\theta_{\text{HOD}} = \{M_{\text{cut}}, M_1, \sigma, \alpha, \kappa, \alpha_{\text{vel,c}}, \alpha_{\text{vel,s}}, B_{\text{cen}}, B_{\text{sat}}\}$$

- Priors: Uniform priors for all parameters except the baryon density, for which we adopt a BBN-like Gaussian prior, as detailed in Table 1.

- Error budget: The covariance used in the likelihood calculation includes contributions from sample variance of the data vector, and model uncertainty associated with the emulator training.

Figure 11 shows the constraints that result from varying various aspects of these settings. Starting from top to bottom, we try a uniform prior for ω_b , finding that it shifts the means of the marginalized posteriors by less than 0.2σ , but degrading the constraining power on ω_{cdm} by 15 per cent. ω_b itself is basically unconstrained with our analysis under a uniform prior, which is the main motivation for adopting a BBN prior, following other clustering studies in the literature (e.g., Alam et al. 2017; Ivanov et al. 2020; Philcox & Ivanov 2022).

Using only the monopole of the correlation functions weakens the precision of the constraints on ω_{cdm} and σ_8 by 18 and 13 per cent, respectively, resulting in a 18 per cent degradation of the precision on $f\sigma_8$.

Using only the most underdense and overdense quintiles (Q_0 and Q_4 , respectively) degrades the precision on ω_{cdm} by 9 per cent. The precision on σ_8 increases by 15 per cent and its mean is shifted to slightly lower values. The increase in precision for σ_8 might seem counter intuitive, but can be explained by the fact that individual quintiles can sometimes lead to marginalized posteriors with slightly different mean values, which is exemplified by the rows where we fit Q_0 and Q_4 separately. This can produce wider contours when

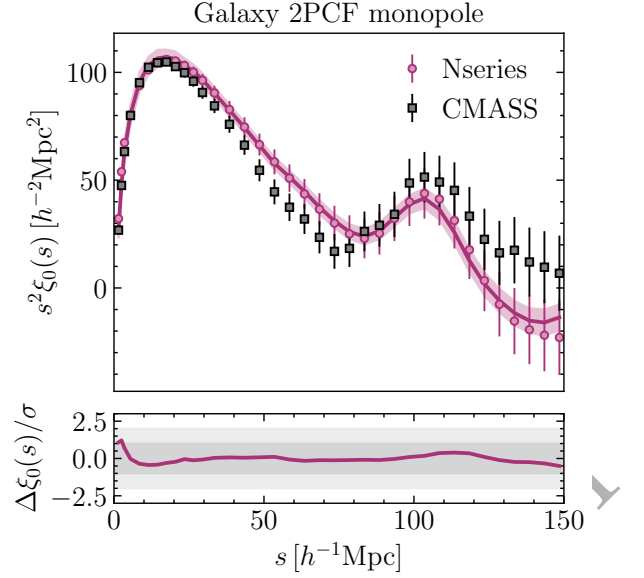


Figure 10. Monopole of the galaxy two-point correlation function, averaged over 84 realizations of the Nseries mocks (violet circles with error bars), along with the best-fit model and its associated uncertainty (violet solid line and bands). Also shown is the monopole from the BOSS CMASS galaxy sample (grey squares with error bars). The lower sub-panel shows the difference between the Nseries mocks and the best-fit model, in units of the error bars. The dark-grey and grey shaded regions demarcate 1σ and 2σ offsets, respectively. \odot

all quintiles are fitted simultaneously. Overall, the fact that the constraining power on most parameters is not severely degraded when we fit only the extreme quintiles (which effectively discards half of the DSC data set) agrees with the picture that most of the information from DSC comes from the very under-dense and over-dense regions, as suggested by Paillas et al. (2021).

DSC by itself predicts $\sigma_8 = 0.768^{+0.037}_{-0.043}$, which is slightly lower than the predicted value from the 2PCF-only fit. However, both the DSC-only and 2PCF-only fits are consistent within one standard deviation with the baseline fit.

We observe that the precision on ω_{cdm} degrades by a factor of 2.4 when we allow N_{eff} to vary, which is due to the strong correlation between these parameters. However, the precision for the other parameters remains relatively stable. If we also allow $dn_s/d\ln k$ and w_0 to vary, it does not significantly affect the constraining power on the base Λ CDM parameters.

We find an interesting shift of the estimated σ_8 to lower values when we fix the assembly bias or velocity bias parameters during the fit. As shown in Appendix A, our best-fit model constrains the environment-based assembly bias parameters to be negative, meaning that galaxies preferentially form in haloes in denser environments. We observe a mild correlation between σ_8 and α , $\alpha_{\text{vel,s}}$ and B_{sat} that is likely to explain the shifts in σ_8 towards lower values when removing assembly or velocity bias parameters.

Finally, we run a fit neglecting the contribution of the model uncertainty to the error budget, which can be achieved by removing \mathbf{C}^{emu} and \mathbf{C}^{sim} from the covariance used to calculate the likelihood. This has a drastic impact on the derived constraints, resulting in a precision that can be more than two times better than the baseline analysis. Although for this particular test we observe that the mean values of the recovered parameters do not shift significantly with respect to the baseline fit, we have explicitly verified that removing the

emulator error results in cosmological constraints that can be biased at a more than 3σ level when fitting the mean of the Nseries mocks down to $1 h^{-1}\text{Mpc}$.

5 DISCUSSION AND CONCLUSIONS

We have presented a clustering analysis of the DR12 BOSS CMASS galaxy sample at $0.45 < z < 0.6$, using simulation-based models of the galaxy two-point correlation function (2PCF) and density-split clustering (DSC). Our theory framework, which is presented in detail in our companion paper (Cuesta-Lazaro et al. 2023), is based on emulators trained on high-fidelity mock galaxy catalogues, which forward model the cosmological dependence of the full shape of the galaxy 2PCF and DSC multipoles, including redshift-space and Alcock-Paczynski distortions.

It should be noted that due to the limitations of the simulation data available for training the emulator, our model fits impose a fixed prior on the acoustic scale θ_* measured from the CMB (Planck Collaboration et al. 2020). This is, however, a very precise and model-independent measurement, so this prior does not significantly restrict our conclusions about the models analysed.

We have validated our theory model against the Nseries mock galaxy catalogues, which were calibrated onto the clustering and selection properties of the CMASS galaxy sample, finding that we can recover unbiased cosmological constraints even using a volume of $120(h^{-3}\text{Gpc}^3)$, which is 84 times larger than the volume examined in our data analysis.

For our base ΛCDM analysis, we find that the galaxy 2PCF constrains ω_{cdm} , σ_8 , and n_s with a precision of 2.8, 6.1, and 3.2 per cent, respectively, using a scale range $1 h^{-1}\text{Mpc} < s < 150 h^{-1}\text{Mpc}$. Adding the DSC multipoles using the same scale range tightens the constraining power to a precision of 1.8, 4.3 and 1.8 per cent, respectively, obtaining $\omega_{\text{cdm}} = 0.1201 \pm 0.0022$, $\sigma_8 = 0.792 \pm 0.034$, and $n_s = 0.970 \pm 0.018$. This is a factor of 1.6, 1.4, and 1.8 of improvement in precision with respect to the 2PCF-only constraints, respectively.

Combining the galaxy 2PCF and DSC multipoles, we derive $f\sigma_8 = 0.462 \pm 0.020$ at $z \approx 0.525$, which is a 4.3 per cent constraint. In comparison, the main BOSS clustering analysis presented in Alam et al. (2017) derived a 8.3 per cent constraint on $f\sigma_8$ using both Galatic caps for their $0.4 < z < 0.6$ redshift bin. We obtain 1.9 times better precision using only the Northern Galactic cap and a narrower redshift bin. This improvement mainly comes from the inclusion of higher-order clustering information that is captured by DSC, and the addition of non-linear scales in the fitting. Our $f\sigma_8$ constraint is largely consistent with Planck 2018 base- ΛCDM predictions (Planck Collaboration et al. 2020), and also agrees well with other clustering studies in the literature that use the same galaxy sample (Yuan et al. 2022b; Yu et al. 2023). Our base- ΛCDM cosmological constraints are summarized in Table 3.

We have also performed fits with single-parameter extensions to base- ΛCDM , where we vary the running of the spectral index of the primordial power spectrum ($dn_s/d\ln k$), the density of massless relic neutrinos (N_{eff}), and the dark energy equation of state parameter w_0 . Overall, we do not find compelling evidence for deviations from the base ΛCDM model, obtaining $N_{\text{eff}} = 3.02^{+0.24}_{-0.27}$, $dn_s/d\ln k = -0.005 \pm 0.013$, and $w_0 = -0.956^{+0.046}_{-0.041}$.

We have used an extended halo occupation distribution (HOD) framework to model the LRG galaxy-halo connection, using halo catalogues from the ABACUSSUMMIT suite of simulations. Our HOD constraints are largely consistent with the findings of Yuan et al.

(2022a) and Yuan et al. (2022b). We constrain the minimum halo mass for hosting centrals to be $\log M_{\text{cut}} = 12.65^{+0.08}_{-0.11}$, and the typical halo mass for hosting one satellite $\log M_1 = 13.69^{+0.10}_{-0.15}$. We find signs of environment-based assembly bias, suggesting a preference for galaxies to form in haloes around denser environments.

Overall, we find that our model is able to fit the Nseries mock galaxy catalogues much more accurately than the CMASS data itself. We deem that this is not a problem specific to our emulation framework, since the mean clustering signal of the Nseries mocks, which agrees with the clustering of the AbacusSummit simulations for similar cosmologies, shows a similar offset with respect to the CMASS data. This suggests the possibility that there might be residual systematic effects in the data that are currently not captured by the mocks and are not taken into account by the weighting procedure that we adopt when calculating the clustering.

It is also worth noting that our emulator has been trained on HOD catalogues at a fixed redshift, $z = 0.5$, and we do not include any possible effects of redshift dependence on the modelling. Furthermore, the effective redshift of our CMASS sample, $z_{\text{eff}} = 0.525$, differs slightly from the redshift of the HOD catalogues. Although the relatively narrow redshift range we are imposing in the CMASS catalogue ($0.45 < z < 0.6$) could alleviate some concerns regarding the redshift dependence of the theoretical model, a careful study of the impact of this simplification on the parameter constraints is going to become even more relevant when extending our framework to other datasets such as eBOSS (Dawson et al. 2016). We plan to study this in future work using the ABACUSSUMMIT lightcone simulations that have recently become available (Hadzhiyska et al. 2022).

On-going and upcoming galaxy redshift surveys, such as DESI (DESI Collaboration et al. 2016), Euclid (Laureijs et al. 2011) and the Nancy Grace Roman Space Telescope (Green et al. 2012) will open exciting avenues to use beyond-two-point statistics for cosmology, not only in terms of improved constraints on cosmological parameters, but also in regards to our understanding of the galaxy-halo connection, the treatment of observational systematics, and the possibility of finding surprises in the data that can challenge our preconceptions about the Cosmos.

ACKNOWLEDGEMENTS

The authors thank Etienne Burtin for helpful discussions throughout the development of this project. This research was enabled in part by the support provided by Compute Ontario (computeontario.ca) and the Digital Research Alliance of Canada (alliancecan.ca). WP acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [funding reference number RGPIN-2019-03908] and from the Canadian Space Agency. SN acknowledges support from an STFC Ernest Rutherford Fellowship, grant reference ST/T005009/2. FB is a University Research Fellow and has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grantagreement853291). YC acknowledges the support of the Royal Society through a University Research Fellowship.

Research at Perimeter Institute is supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Colleges and Universities. This work was supported by the U.S. Department of Energy through grant DE-SC0013718 and under DE-AC02-76SF00515 to SLAC National Accelerator Laboratory, and by the Kavli Institute for Particle Astrophysics and Cosmology. This project used resources of the Na-

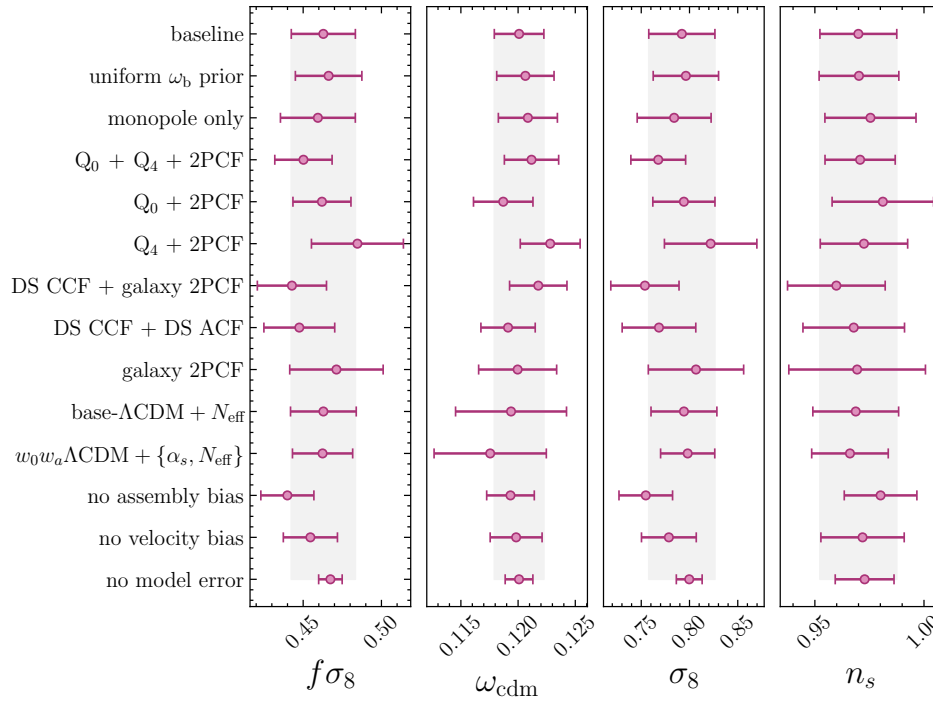


Figure 11. A comparison of the constraints on the base- Λ CDM parameters and the derived $f\sigma_8$ value from fits run with different analysis settings. Our baseline analysis, shown at the top, was run by simultaneously fitting the galaxy 2PCF and density-split multipoles with a base- Λ CDM + extended-HOD model, using scales $1 h^{-1}\text{Mpc} < s < 150 h^{-1}\text{Mpc}$. The other points represent variations to that baseline configuration, either by changing the data vector or the model prescription.

tional Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231. The ABACUSUMMIT simulations were run at the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725.

For the purpose of open access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript version arising.

DATA AVAILABILITY STATEMENT

The source code to generate the figures in this manuscript is available at <https://github.com/florpi/sunbird>. The data from the ABACUSUMMIT suite of simulations can be found at <https://abacusnbody.org>. Any other data necessary to reproduce the results of this work will be shared upon reasonable request to the corresponding author.

REFERENCES

Achitouv I., 2019, *Phys. Rev. D*, **100**, 123513
 Alam S., et al., 2017, *MNRAS*, **470**, 2617
 Alcock C., Paczynski B., 1979, *Nature*, **281**, 358
 Aung H., et al., 2023, *MNRAS*, **519**, 1648
 Aver E., Olive K. A., Skillman E. D., 2015, *Journal of Cosmology and Astroparticle Physics*, 2015, 011
 Banerjee A., Abel T., 2021, *MNRAS*, **500**, 5479
 Bautista J. E., et al., 2021, *MNRAS*, **500**, 376
 Beutler F., et al., 2011, *MNRAS*, **416**, 3017

Beutler F., et al., 2012, *MNRAS*, **423**, 3430
 Blake C., et al., 2011, *MNRAS*, **415**, 2876
 Blas D., Lesgourgues J., Tram T., 2011, *J. Cosmology Astropart. Phys.*, 2011, 034
 Brieden S., Gil-Marín H., Verde L., 2021, *J. Cosmology Astropart. Phys.*, 2021, 054
 Burger P. A., et al., 2023, *A&A*, **669**, A69
 Carrasco J. J. M., Hertzberg M. P., Senatore L., 2012, *Journal of High Energy Physics*, 2012, 82
 Chan K. C., Hamaus N., Desjacques V., 2014, *Phys. Rev. D*, **90**, 103521
 Chapman M. J., et al., 2022, *MNRAS*, **516**, 617
 Chuang C.-H., Kitaura F.-S., Liang Y., Font-Ribera A., Zhao C., McDonald P., Tao C., 2017, *Physical Review D*, 95
 Cole S., et al., 2005, *MNRAS*, **362**, 505
 Cooke R. J., Pettini M., Steidel C. C., 2018, *The Astrophysical Journal*, **855**, 102
 Correa C. M., Paz D. J., Sánchez A. G., Ruiz A. N., Padilla N. D., Angulo R. E., 2020, *MNRAS*,
 Cuesta-Lazaro C., et al., 2023, arXiv e-prints, p. arXiv:2309.16539
 DESI Collaboration et al., 2016, arXiv e-prints, p. arXiv:1611.00036
 Dawson K. S., et al., 2013, *AJ*, **145**, 10
 Dawson K. S., et al., 2016, *AJ*, **151**, 44
 Dodelson S., 2003, *Modern Cosmology*. Academic Press, Amsterdam
 Dong-Páez C. A., et al., 2022, arXiv e-prints, p. arXiv:2208.00540
 Eisenstein D. J., Hu W., 1998, *ApJ*, **496**, 605
 Eisenstein D. J., et al., 2005, *ApJ*, **633**, 560
 Emas N. P. A. P., Wulandari H., 2019, in *Journal of Physics Conference Series*. p. 012026, doi:10.1088/1742-6596/1231/1/012026
 Feldman H. A., Kaiser N., Peacock J. A., 1994, *The Astrophysical Journal*, **426**, 23
 Friedrich O., et al., 2018, *Phys. Rev. D*, **98**, 023508
 Garrison L. H., Eisenstein D. J., Pinto P. A., 2019, *Monthly Notices of the Royal Astronomical Society*, **485**, 3370
 Garrison L. H., Eisenstein D. J., Ferrer D., Maksimova N. A., Pinto P. A.,

- 2021, *Monthly Notices of the Royal Astronomical Society*, 508, 575
- Gil-Marín H., Percival W. J., Verde L., Brownstein J. R., Chuang C.-H., Kitaura F.-S., Rodríguez-Torres S. A., Olmstead M. D., 2017, *MNRAS*, 465, 1757
- Green J., et al., 2012, Wide-Field InfraRed Survey Telescope (WFIRST) Final Report, doi:10.48550/ARXIV.1208.4012, <https://arxiv.org/abs/1208.4012>
- Grieb J. N., et al., 2017, *MNRAS*, 467, 2085
- Gruen D., et al., 2018, *Phys. Rev. D*, 98, 023507
- Gualdi D., Gil-Marín H., Verde L., 2021, *Journal of Cosmology and Astroparticle Physics*, 2021, 008
- Gunn J. E., et al., 1998, *AJ*, 116, 3040
- Gunn J. E., et al., 2006, *The Astronomical Journal*, 131, 2332
- Guth A. H., Pi S. Y., 1982, *Phys. Rev. Lett.*, 49, 1110
- Hadzhiyska B., Eisenstein D., Bose S., Garrison L. H., Maksimova N., 2021, *Monthly Notices of the Royal Astronomical Society*, 509, 501
- Hadzhiyska B., Garrison L. H., Eisenstein D., Bose S., 2022, *MNRAS*, 509, 2194
- Hahn C., Scoccimarro R., Blanton M. R., Tinker J. L., Rodríguez-Torres S. A., 2017, *MNRAS*, 467, 1940
- Hamaus N., Wandelt B. D., Sutter P. M., Lavaux G., Warren M. S., 2014, *Phys. Rev. Lett.*, 112, 041304
- Hamaus N., Pisani A., Sutter P. M., Lavaux G., Escoffier S., Wandelt B. D., Weller J., 2016, *Phys. Rev. Lett.*, 117, 091302
- Hawken A. J., et al., 2017, *A&A*, 607, A54
- Hawking S. W., 1982, *Physics Letters B*, 115, 295
- Hou J., et al., 2021, *MNRAS*, 500, 1201
- Hou J., Slepian Z., Cahn R. N., 2023a, *MNRAS*, 522, 5701
- Hou J., Moradinezhad Dizgah A., Hahn C., Massara E., 2023b, *J. Cosmology Astropart. Phys.*, 2023, 045
- Ishiyama T., et al., 2021, *MNRAS*, 506, 4210
- Ivanov M. M., Simonović M., Zaldarriaga M., 2020, *J. Cosmology Astropart. Phys.*, 2020, 042
- Jackson J. C., 1972, *MNRAS*, 156, 1P
- Kaiser N., 1987, *MNRAS*, 227, 1
- Kitaura F. S., Yepes G., Prada F., 2014, *MNRAS*, 439, L21
- Kitaura F.-S., Gil-Marín H., Scóccola C. G., Chuang C.-H., Müller V., Yepes G., Prada F., 2015, *MNRAS*, 450, 1836
- Kitaura F.-S., et al., 2016, *MNRAS*, 456, 4156
- Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, *MNRAS*, 457, 4340
- Kobayashi Y., Nishimichi T., Takada M., Takahashi R., 2020, *Physical Review D*, 101
- Kobayashi Y., Nishimichi T., Takada M., Miyatake H., 2022, *Phys. Rev. D*, 105, 083517
- Komatsu E., et al., 2011, *The Astrophysical Journal Supplement Series*, 192, 18
- Kosowsky A., Turner M. S., 1995, *Phys. Rev. D*, 52, R1739
- Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64
- Lange J. U., Hearin A. P., Leauthaud A., van den Bosch F. C., Guo H., DeRose J., 2022, *MNRAS*, 509, 1779
- Lange J. U., Hearin A. P., Leauthaud A., van den Bosch F. C., Xhakaj E., Guo H., Wechsler R. H., DeRose J., 2023, *MNRAS*, 520, 5373
- Laureijs R., Amiaux J., Arduini S., Auguères J. L., Brinchmann J., Cole R., et al., 2011, preprint (arXiv:1110.3193)
- Lavaux G., Wandelt B. D., 2012, *ApJ*, 754, 109
- Lavaux G., Jasche J., Leclercq F., 2019, arXiv e-prints, p. arXiv:1909.06396
- Lesgourgues J., Pastor S., 2014, *New Journal of Physics*, 16, 065002
- Levi M., et al., 2019, in *Bulletin of the American Astronomical Society*. p. 57 (arXiv:1907.10688)
- Lippich M., Sánchez A. G., 2021, *MNRAS*, 508, 3771
- Maksimova N. A., Garrison L. H., Eisenstein D. J., Hadzhiyska B., Bose S., Satterthwaite T. P., 2021, *Monthly Notices of the Royal Astronomical Society*, 508, 4017
- Mao Q., Berlind A. A., Scherrer R. J., Neyrinck M. C., Scoccimarro R., Tinker J. L., McBride C. K., Schneider D. P., 2017, *ApJ*, 835, 160
- Maraston C., et al., 2013, *MNRAS*, 435, 2764
- Massara E., et al., 2022, arXiv e-prints, p. arXiv:2206.01709
- Moon J., et al., 2023, arXiv e-prints, p. arXiv:2304.08427
- Moradinezhad Dizgah A., Biagetti M., Sefusatti E., Desjacques V., Noreña J., 2021, *J. Cosmology Astropart. Phys.*, 2021, 015
- Mukhanov S., 2007, *Journal of Physics A Mathematical General*, 40, 6561
- Nadathur S., Carter P., Percival W. J., 2019, *MNRAS*, 482, 2459
- Neyrinck M. C., 2008, *MNRAS*, 386, 2101
- Oogi T., et al., 2023, *MNRAS*, 525, 3879
- Paillas E., Cai Y.-C., Padilla N., Sánchez A. G., 2021, *Monthly Notices of the Royal Astronomical Society*, 505, 5731–5752
- Paillas E., et al., 2023, *MNRAS*, 522, 606
- Peebles P. J. E., 1980, The large-scale structure of the universe
- Peebles P. J. E., Yu J. T., 1970, *ApJ*, 162, 815
- Percival W. J., et al., 2001, *MNRAS*, 327, 1297
- Percival W. J., et al., 2002, *MNRAS*, 337, 1068
- Percival W. J., Friedrich O., Sellentin E., Heavens A., 2022, *MNRAS*, 510, 3207
- Philcox O. H. E., 2022, *Phys. Rev. D*, 106, 063501
- Philcox O. H. E., Ivanov M. M., 2022, *Phys. Rev. D*, 105, 043517
- Philcox O. H. E., Hou J., Slepian Z., 2021, arXiv e-prints, p. arXiv:2108.01670
- Planck Collaboration et al., 2014, *A&A*, 571, A16
- Planck Collaboration et al., 2020, *A&A*, 641, A6
- Prada F., Behroozi P., Ishiyama T., Klypin A., Pérez E., 2023, arXiv e-prints, p. arXiv:2304.11911
- Reid B. A., et al., 2012, *MNRAS*, 426, 2719
- Reid B., et al., 2016, *MNRAS*, 455, 1553
- Riess A. G., et al., 2018, *ApJ*, 855, 136
- Ross A. J., Samushia L., Howlett C., Percival W. J., Burden A., Manera M., 2015, *MNRAS*, 449, 835
- Ross A. J., et al., 2016, *Monthly Notices of the Royal Astronomical Society*, 464, 1168
- Sánchez A. G., 2020, *Phys. Rev. D*, 102, 123511
- Sánchez A. G., et al., 2017, *MNRAS*, 464, 1640
- Satpathy S., et al., 2017, *MNRAS*, 469, 1369
- Schöneberg N., Lesgourgues J., Hooper D. C., 2019, *Journal of Cosmology and Astroparticle Physics*, 2019, 029
- Scolnic D., et al., 2015, *ApJ*, 815, 117
- Semenáite A., et al., 2022, *MNRAS*, 512, 5657
- Semenáite A., et al., 2023, *MNRAS*, 521, 5013
- Sinha M., Garrison L. H., 2020, *MNRAS*, 491, 3022
- Slepian Z., et al., 2017, *MNRAS*, 468, 1070
- Speagle J. S., 2020, *MNRAS*, 493, 3132
- Sugiyama N. S., et al., 2023, *MNRAS*, 523, 3133
- Sunyaev R. A., Zeldovich Y. B., 1970, *Ap&SS*, 7, 3
- Sutter P. M., Lavaux G., Wandelt B. D., Weinberg D. H., 2012, *ApJ*, 761, 187
- Tröster T., et al., 2020, *A&A*, 633, L10
- Valogiannis G., Dvorkin C., 2022a, *Phys. Rev. D*, 105, 103534
- Valogiannis G., Dvorkin C., 2022b, *Phys. Rev. D*, 106, 103509
- White M., 2016, *J. Cosmology Astropart. Phys.*, 2016, 057
- York D. G., et al., 2000, *AJ*, 120, 1579
- Yu B., Seljak U., Li Y., Singh S., 2023, *J. Cosmology Astropart. Phys.*, 2023, 057
- Yuan S., Garrison L. H., Hadzhiyska B., Bose S., Eisenstein D. J., 2022a, *MNRAS*, 510, 3301
- Yuan S., Garrison L. H., Eisenstein D. J., Wechsler R. H., 2022b, *MNRAS*, 515, 871
- Zhai Z., et al., 2023, *ApJ*, 948, 99
- Zhang P., Cai Y., 2022, *J. Cosmology Astropart. Phys.*, 2022, 031
- Zheng Z., Coil A. L., Zehavi I., 2007, *ApJ*, 667, 760
- d'Amico G., Gleyzes J., Kokron N., Markovic K., Senatore L., Zhang P., Beutler F., Gil-Marín H., 2020, *J. Cosmology Astropart. Phys.*, 2020, 005
- de Mattia A., et al., 2021, *MNRAS*, 501, 5616
- de Salas P. F., Pastor S., 2016, *J. Cosmology Astropart. Phys.*, 2016, 051
- eBOSS Collaboration et al., 2020, arXiv e-prints, p. arXiv:2007.08991

Table A1. HOD parameter constraints for the base- Λ CDM fits on the CMASS galaxy catalogue, using the combination of density-split clustering and the galaxy 2PCF, or using only the latter. In those cases where there is no constraining power other than that from the prior, we use an em dash.

Parameter	density-split + galaxy 2PCF		galaxy 2PCF	
	best-fit	mean $\pm \sigma$	best-fit	mean $\pm \sigma$
$\log M_1$	13.6662	$13.74^{+0.10}_{-0.20}$	13.9449	> 13.9
$\log M_{\text{cut}}$	12.6617	$12.698^{+0.074}_{-0.16}$	12.7406	$12.71^{+0.13}_{-0.18}$
α	1.3452	$1.18^{+0.24}_{-0.15}$	1.4275	$1.08^{+0.20}_{-0.26}$
$\alpha_{\text{vel},s}$	0.7555	< 0.857	0.8682	$0.963^{+0.093}_{-0.25}$
$\alpha_{\text{vel},c}$	0.0079	< 0.147	0.0856	—
$\log \sigma$	-0.5054	$-0.613^{+0.16}_{-0.013}$	-1.0385	< -1.24
κ	0.0312	< 0.524	0.6525	> 0.622
B_{cen}	-0.4359	$-0.306^{+0.040}_{-0.18}$	0.0119	$-0.20^{+0.12}_{-0.24}$
B_{sat}	-0.2179	-0.39 ± 0.30	-0.4332	$-0.18^{+0.28}_{-0.76}$

APPENDIX A: GALAXY-HALO CONNECTION CONSTRAINTS

Figure A1 shows the full posterior distribution, including HOD parameters, for the base- Λ CDM fit using the baseline configuration of our analysis, where we combine the DSC and galaxy 2PCF multipoles. The constraints are largely consistent with CMASS HOD constraints with the redshift-space 2PCF in Yuan et al. (2022a) and the cosmology-marginalized constraints in Yuan et al. (2022b). In terms of the vanilla parameters, both Yuan et al. (2022a) and Yuan et al. (2022b) found the best fit to favor $\log M_{\text{cut}} \approx 12.8$, $\log M_1 \approx 14.0$, and $\alpha \approx 1.0$, all in excellent agreement with our posterior constraints here. However, Yuan et al. (2022a) and Yuan et al. (2022b) found a lower $\log \sigma$, which is consistent with the 2PCF-only constraints in Table A1, but significantly less than the DSC + 2PCF results. In terms of velocity biases, we find a lower central and satellite velocity bias compared to the two previous studies.

In terms of galaxy assembly bias, we find negative B_{cen} and B_{sat} , which is qualitatively consistent with Yuan et al. (2022a). However, the amplitude of the inferred central assembly bias is significantly larger than that of Yuan et al. (2022a) and Yuan et al. (2022b). Physically, negative central assembly bias means that central galaxies prefer denser environments, consistent with our intuition of large red galaxies preferentially occupying highly biased cluster environments. Yuan et al. (2022b) also found a mild degeneracy between environment-based assembly bias and velocity bias. By allowing galaxies to preferentially occupy halos in denser environments, these galaxies would occupy deeper potential wells and thus have higher peculiar velocities, reducing the need to invoke additional velocity bias. This could potentially explain the lower inferred velocity bias we find.

We also notice that even though our priors on HOD parameters are fairly broad, the marginalized posteriors for some parameters hit the lower bound of the prior limits. This is the case for the velocity bias of satellites $\alpha_{\text{vel},s}$, and the velocity bias of centrals B_{cen} . Therefore, we advise the reader to interpret our results for these parameters as trends rather than reliable central values, error bars, or upper limits. However, our cosmology results are likely not significantly impacted, as there is weak to no correlation between these parameters and cosmological parameters. We plan to further explore different choices of priors on the HOD parameters in our upcoming work.

APPENDIX B: DEPENDENCE ON THE MODEL PARAMETERS

To explore the source of the cosmological constraining power of DSC, Fig. B1 illustrates how the model vector responds to changes in different cosmological parameters. For brevity, we only show the cross-correlation function between the first quintile, Q_0 and the galaxy field, but we have verified that the trends for other quintiles are similar in nature.

Increasing ω_{cdm} results in a decrease in the absolute value of the amplitude of the monopole and quadrupole. There are several effects that come into play here. First, changing ω_{cdm} has an effect on the matter clustering, which affects the shape of the galaxy density PDF, and consequently the identification of density quintiles. ω_{cdm} is also tightly correlated with Ω_{m} , which determines $f\sigma_8$, which in turn sets the amplitude of velocity fluctuations and affects RSD. Moreover, the changes in ω_{cdm} produce AP distortions, which affect the quintile-galaxy pair separations and become more severe the more it deviates from the fiducial cosmology used to convert redshift to distances. AP distortions not only change the amplitude of the multipoles, but also produce horizontal shifts in the monopole, changing the scale of the acoustic feature.

The amplitude of the quadrupole correlates positively with σ_8 , which is in agreement with our expectations of linear theory regarding how the amplitude of velocity fluctuations scales with this parameter. Interestingly, the monopole amplitude remains fixed regardless of the value of σ_8 . This is a consequence of the way in which we populate dark matter halos with galaxies: for those HOD parameter combinations where the resulting galaxy number density is greater than the CMASS number density, $\approx 3.5 \times 10^{-4} (h\text{Mpc}^{-1})^3$, the catalogues are downsampled to match the target. This effectively fixes $b\sigma_8$ in our training sample, where b is the linear galaxy bias. As we vary σ_8 , the galaxy bias also changes, keeping the amplitude of the monopole fixed.

Finally, we see that the spectral index of the primordial power spectrum, n_s , is anti-correlated with the amplitude of the monopole and quadrupole, but does not produce horizontal shifts in the profiles as in AP distortions.

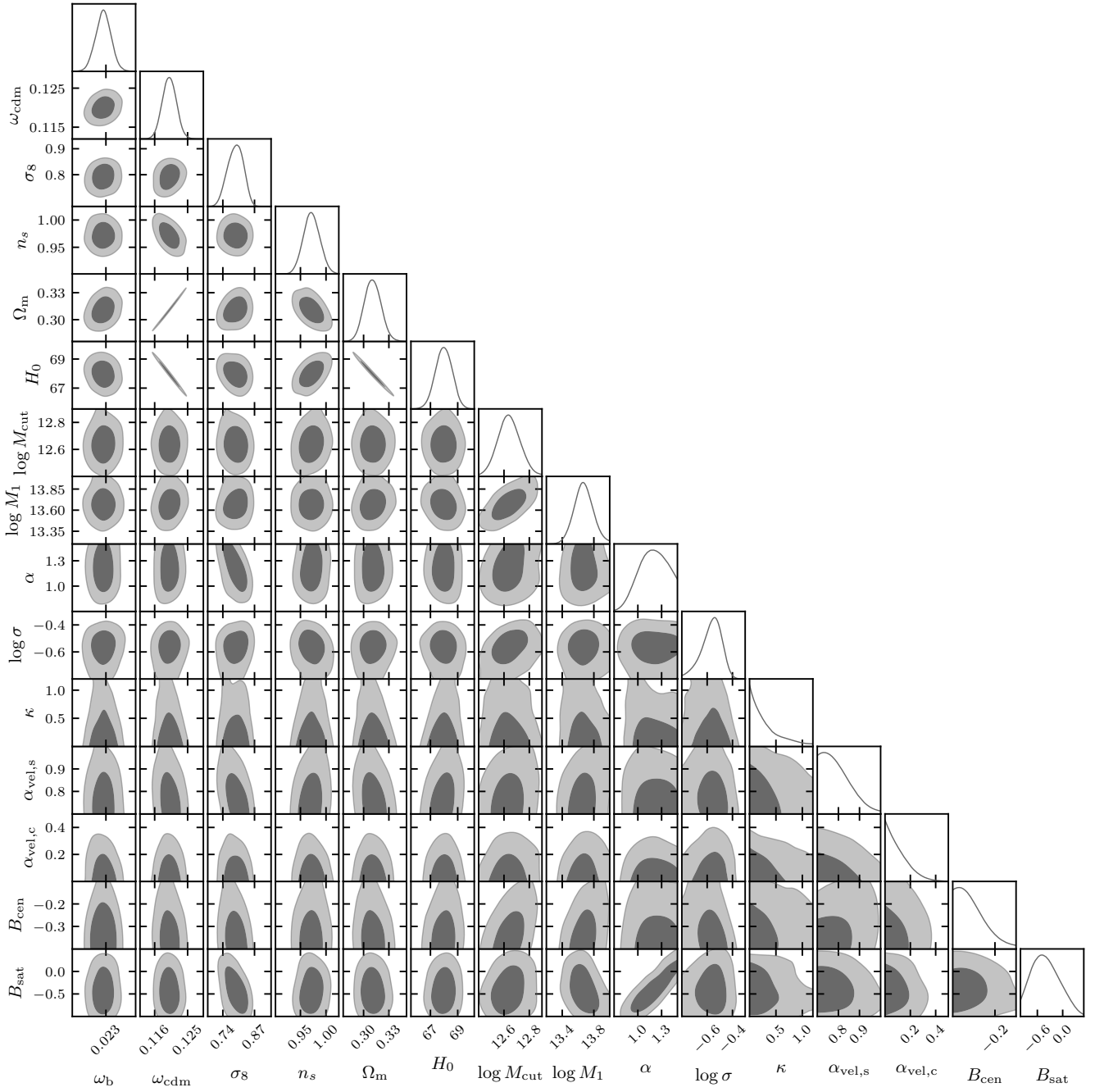



Figure A1. Similar to Fig. 4, but showing the full posterior distribution of cosmological and HOD parameters from the base- Λ CDM fits with our baseline configuration. 

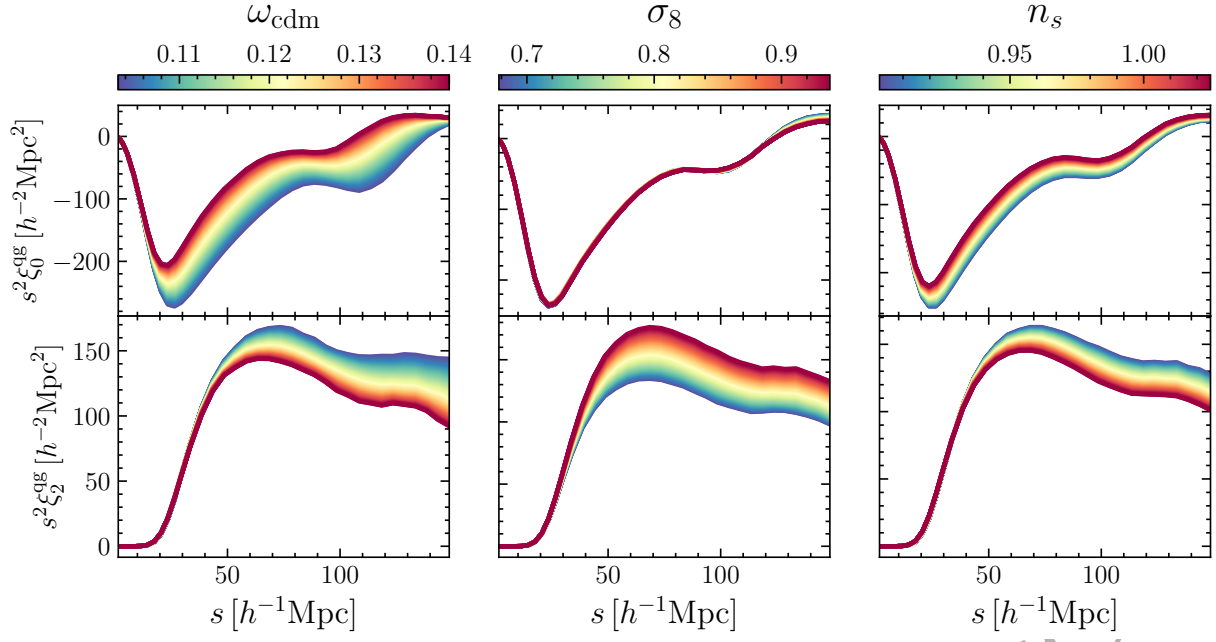


Figure B1. The dependence on model parameters as predicted by our emulator, for the cross-correlation function between the most underdense quintile (Q_0) and the galaxy field. Each subpanel shows the monopole and quadrupole moments, rescaled by s^2 to highlight the model features at large scales. The ranges that are plotted for each parameter correspond to the prior ranges used in the likelihood analysis. [🔗](#)

ORIGINAL UNEDITED MANUSCRIPT