BAO from angular clustering: optimization and mitigation of theoretical systematics

K. C. Chan,[1,2,3]★ M. Crocce,[2,3] A. J. Ross,[4] S. Avila,[5,6] J. Elvin-Poole,[7] M. Manera,[8]
W. J. Percival,[5] R. Rosenfeld,[9,10] T. M. C. Abbott,[11] F. B. Abdalla,[8,12] S. Allam,[13]
E. Bertin,[14,15] D. Brooks,[8] D. L. Burke,[16,17] A. Carnero Rosell,[10,18]
M. Carrasco Kind,[19,20] J. Carretero,[21] F. J. Castander,[2,3] C. E. Cunha,[16]
C. B. D'Andrea,[22] L. N. da Costa,[10,18] C. Davis,[16] J. De Vicente,[23] T. F. Eifler,[24,25]
J. Estrada,[13] B. Flaugher,[13] P. Fosalba,[2,3] J. Frieman,[13,26] J. García-Bellido,[6]
E. Gaztanaga,[2,3] D. W. Gerdes,[27,28] D. Gruen,[16,17] R. A. Gruendl,[19,20] J. Gschwend,[10,18]
G. Gutierrez,[13] W. G. Hartley,[8,29] K. Honscheid,[4,30] B. Hoyle,[31,32] D. J. James,[33]
E. Krause,[24,25] K. Kuehn,[34] O. Lahav,[8] M. Lima,[10,35] M. March,[22] F. Menanteau,[19,20]
C. J. Miller,[27,28] R. Miquel,[21,36] A. A. Plazas,[25] K. Reil,[17] A. Roodman,[16,17]
E. Sanchez,[23] V. Scarpine,[13] I. Sevilla-Noarbe,[23] M. Smith,[37] M. Soares-Santos,[13]
F. Sobreira,[10,38] E. Suchyta,[39] M. E. C. Swanson,[20] G. Tarle,[28] D. Thomas,[5]
A. R. Walker[11] and DES Collaboration

Affiliations are listed at the end of the paper

ABSTRACT

We study the methodology and potential theoretical systematics of measuring baryon acoustic oscillations (BAO) using the angular correlation functions in tomographic bins. We calibrate and optimize the pipeline for the Dark Energy Survey Year 1 data set using 1800 mocks. We compare the BAO fitting results obtained with three estimators: the Maximum Likelihood Estimator (MLE), Profile Likelihood, and Markov Chain Monte Carlo. The fit results from the MLE are the least biased and their derived 1σ error bar are closest to the Gaussian distribution value after removing the extreme mocks with non-detected BAO signal. We show that incorrect assumptions in constructing the template, such as mismatches from the cosmology of the mocks or the underlying photo-z errors, can lead to BAO angular shifts. We find that MLE is the method that best traces this systematic biases, allowing to recover the true angular distance values. In a real survey analysis, it may happen that the final data sample properties are slightly different from those of the mock catalogue. We show that the effect on the mock covariance due to the sample differences can be corrected with the help of the Gaussian covariance matrix or more effectively using the eigenmode expansion of the mock covariance. In the eigenmode expansion, the eigenmodes are provided by some proxy covariance matrix. The eigenmode expansion is significantly less susceptible to statistical fluctuations relative to the direct measurements of the covariance matrix because of the number of free parameters is substantially reduced.

Key words: cosmology: observations – large-scale structure of Universe.

1 INTRODUCTION

Baryon acoustic oscillations (BAO), generated in the early universe, leave their imprint in the distribution of galaxies (Peebles & Yu 1970; Sunyaev & Zeldovich 1970). At early times ($z \gtrsim 1100$),

★ E-mail: chan@ice.cat

photons and baryons form a tightly coupled plasma, and sound waves propagate in this plasma. After the recombination of hydrogen, photons can free stream in the universe. The acoustic wave pattern remains frozen in the baryon distribution. The sound horizon at the drag epoch is close to 150 Mpc. The formation of the BAO in the early universe is governed by the well-understood linear physics, see Bond & Efstathiou (1984), Bond & Efstathiou (1987), Hu & Sugiyama (1996), Hu, Sugiyama & Silk (1997), and Dodelson (2003) for the details of the cosmic microwave background physics. Given that this scale is relatively large, it is less susceptible to astrophysical contamination and other non-linear effects. Another reason for its robustness is that it exhibits sharp features in two-point correlations, while non-linearity tends to produce changes in the broad-band power. The BAO can serve as a standard ruler (Eisenstein & Hu 1998; Meiksin, White & Peacock 1999; Eisenstein, Seo & White 2007). Observations of the BAO feature in the distribution of galaxies has been recognized as one of the most important cosmological probes that enables us to measure the Hubble parameter and the equation of state of the dark energy [see e.g. Weinberg et al. (2013); Aubourg et al. (2015)]. The potential distortions of BAO due to non-linear evolution and galaxy bias are less than 0.5 per cent (Crocce & Scoccimarro 2008; Padmanabhan & White 2009).

Using spectroscopic data, the BAO was first clearly detected in Sloan Digital Sky Survey (SDSS) (Eisenstein et al. 2005) and 2dF Galaxy Survey (2dFGS) (Cole et al. 2005), and subsequently in numerous studies, e.g. Gaztanaga, Cabre & Hui (2009), Percival et al. (2010), Beutler et al. (2011), Kazin et al. (2014), Ross et al. (2015), Alam et al. (2017), Bautista et al. (2017), and Ata et al. (2017). Spectroscopic data give precise redshift information, but such surveys are relatively expensive, as they require spectroscopic observations of galaxies targeted from existing imaging surveys. On the other hand, multiband imaging surveys relying on the use of photometric redshift (photo-z) for radial information (Koo 1985) can cheaply survey large volumes. There are several (generally weaker) detections of the BAO feature in the galaxy distribution using photometric data (Padmanabhan et al. 2007; Estrada, Sefusatti & Frieman 2009; Hütsi 2010; Seo et al. 2012; Carnero et al. 2012; de Simoni et al. 2013). The Dark Energy Survey (DES) and future surveys such as The Large Synoptic Survey Telescope (LSST) (LSST Science Collaboration et al. 2009) will deliver an enormous amount of photometric data with well calibrated photo-z's, thus we expect that accurate BAO measurements will be achieved from these surveys.

In this work, we investigate the BAO detection using the angular correlation function of a galaxy sample optimally selected from the first year of DES data (DES Y1, Crocce et al. 2017). DES is one of the largest ongoing galaxy surveys, and its goal is to reveal the nature of the dark energy. One of the routes to achieve this goal is to accurately measure the BAO scale in the distribution of galaxies as a function of redshift. As it is a photometric survey, its redshift information is not so precise, but it can cover a large volume. This is advantageous to the BAO measurement: the sound horizon scale is large, it requires large survey volume to get good statistics. The BAO sample (Crocce et al. 2017) derived from the first year DES data (Drlica-Wagner et al. 2017) already consists of about 1.3 million galaxies covering more than 1318 deg^2. This is only 1 per cent of the total number of galaxies identified in DES Y1, which can be used for science analyses such as Abbott et al. (2017a).

A measurement of the BAO at the effective redshift (or mean redshift) of the survey, $z_{eff} = 0.8$ was presented in Abbott et al. (2017b) using this sample. This effective redshift is less explored by other existing surveys. In Abbott et al. (2017b), three statistics:

the angular correlation function w, the angular power spectrum C_ℓ, and the 3D correlation function ξ were used. See Camacho et al. (2018) and Ross et al. (2017b) for the details on the C_ℓ and ξ analyses. These statistics are sensitive to different systematics and they provide important cross checks for the analyses. By measuring angular correlation w from the data divided into a number of redshift bins (or tomography), no precise redshift information is required, thus this statistics is well suited for extracting BAO information from the photometric sample. In the current paper, we present the details of the calibrations and optimization applied when using w to measure the BAO. Although the fiducial setup is tailored to DES Y1, the analysis and apparatus developed here will be useful for upcoming DES data and other large-scale imaging surveys.

This paper is organized as follows. In Section 2, we introduce the theory for the BAO modelling and for the Gaussian covariance matrix. We also describe the mock catalogues used to test the pipeline. In Section 3, we discuss the extraction of the angular diameter distance using BAO template-fitting methods, and we present three different estimators for such procedure: Maximum Likelihood Estimator (MLE), Profile Likelihood (PL), and Markov Chain Monte Carlo (MCMC). Potential systematics errors in the angular diameter distance scale due to the assumed template used for BAO extraction are studied in Section 4, such as the BAO damping scale, the assumed cosmological model or the propagation of photo-z errors. In Section 5, we discuss various optimizations for the analysis. Section 6 is devoted to issues related to the covariance; in particular we present an eigenmode expansion that allows to adapt the covariance to changes in the underlying template or sample assumptions. We present our conclusions in Section 7. In Appendix A, we further compare the results obtained with the three estimators.

2 THEORY AND MOCK CATALOGUES

To detect the BAO signal in the data, we employ a template-fitting method. A template encodes the expected shape and amplitude of the BAO feature. It is computed using the expected properties, e.g. the survey and galaxy sample characteristics. A large set of mock catalogues are constructed to mimic those detailed characteristics. The template is then fitted to the correlation functions measured on the mock catalogues to extract the BAO distance scale, and to study different systematic and statistical effects.

In this section, we discuss how the template is constructed. We also introduce a Gaussian theory covariance that is employed at different moments in the paper. The mock catalogues used in this study are briefly described in Section 2.3, with full details given in Avila et al. (2017).

2.1 The BAO template

The angular correlation function $w(\theta)$ measures the correlation between two points separated by an angle θ. We use linear theory for the angular correlation function, except that we include the effect where non-linear structure growth smooths the BAO feature by including one additional 'BAO damping' factor. This is sufficient for the angular scales we consider, $\theta > 0.5$. For example, we have checked that it can fit the mean result from the mocks well. See Section 4.1 for more details. In this work, the template is computed in configuration space directly as (Crocce, Cabré & Gaztañaga 2011)

$$w(\theta) = \int dz_1 \int dz_2\, g(z_1) g(z_2) \xi_s\big(s(z_1, z_2, \theta), \hat{s} \cdot \hat{l}\big), \quad (1)$$

with $g(z) \equiv D(z)\phi(z)$, $\phi(z)$ is the redshift distribution of the sample, and D is the linear growth factor. The redshift space correlation function ξ_s depends on the separation vector s in redshift space and the dot product between the line of sight direction $\hat{l}$ and $\hat{s}$, which is given by

$$\hat{s} \cdot \hat{l} = \frac{r(z_2) - r(z_1)}{s} \cos\frac{\theta}{2}, \tag{2}$$

with $r(z)$ being the comoving distance to redshift z. In the linear regime, the redshift–space correlation function is related to the real space power spectrum as (Kaiser 1987; Hamilton 1992; Cole, Fisher & Weinberg 1994)

$$\xi_s(s) = \sum_{\ell=0,2,4} i^\ell A_\ell P_\ell(\hat{s}\cdot\hat{l}) \int \frac{\mathrm{d}k k^2}{2\pi^2} j_\ell(ks) P_\mathrm{m}(k), \tag{3}$$

where P_ℓ is the Legendre polynomial of order ℓ and j_ℓ the spherical Bessel function. In our model, the power spectrum $P_\mathrm{m}(k)$ including the BAO damping is parametrized as

$$P_\mathrm{m}(k) = (P_\mathrm{lin} - P_\mathrm{nw})e^{-k^2\Sigma^2} + P_\mathrm{nw}, \tag{4}$$

where P_lin is the linear matter power spectrum computed by CAMB (Lewis, Challinor & Lasenby 2000) and P_nw is the linear one without BAO wiggles given in Eisenstein & Hu (1998). The damping scale should be anisotropic; but given the accuracy of our data we take the isotropic average $\Sigma = 5.2\,\mathrm{Mpc}\,h^{-1}$. In Section 4.1, we show how this value is determined. The multipole coefficients A_ℓ are given by

$$A_\ell = \begin{cases} b^2 + \frac{2}{3}bf + \frac{1}{5}f^2 & \text{for } \ell = 0, \\ \frac{4}{3}bf + \frac{4}{7}f^2 & \text{for } \ell = 2, \\ \frac{8}{35}f^2 & \text{for } \ell = 4, \end{cases} \tag{5}$$

where b is the linear galaxy bias and $f \equiv \mathrm{d}\ln D/\mathrm{d}\ln a$ with a being the scale factor. We use $f = \Omega_\mathrm{m}^{0.55}$, where Ω_m is the density parameter of matter (Peebles & Groth 1975; Linder 2005) evaluated at the mean redshift of the photo-z distribution. The bias parameters b are assumed to be constant in each redshift bin, which is sufficient given the narrow bins used here, but they can vary from bin to bin. They are determined by fitting, for each mock and redshift bin, the model w to the mock correlation function measurement in the θ range [0°.5, 2°.5]. We have checked that using a smaller range of [0°.5, 1°.5] results in less than 0.001 fractional variation of the best-fitting α (except a few extreme cases).

A template-fitting method is employed to detect the BAO feature in the angular correlation function. Analogous to that in Seo et al. (2012), we use the following template

$$T_\alpha(\theta) = B\, w(\alpha\theta) + A_0 + \frac{A_1}{\theta} + \frac{A_2}{\theta^2}, \tag{6}$$

where $w(\theta)$ is the angular correlation function computed in some cosmology (the fiducial setting is the Marenostrum Institut de Ciencias de l'Espai Simulations (MICE) cosmology, see below). The parameter α gives the shift in the model BAO position relative to the fiducial one. The parameter B allows for a shift in the overall amplitude. Its value is expected to be close to 1 as we have determined the physical bias parameter b by fitting to the data first. The polynomial in $1/\theta$ gives a smooth contribution and is not expected to lead to strong features in the BAO range. We will also test the model with varying number of A_i: with A_0 only (denoted as $N_\mathrm{p} = 1$), A_0 and A_1 ($N_\mathrm{p} = 2$), and A_0, A_1, and A_2 ($N_\mathrm{p} = 3$). We fit the template over a range of angles, from 0°.5 to 5°.

The MICE cosmology is the reference cosmology adopted in the DES Y1 BAO analysis (Abbott et al. 2017b), and thus is in

this paper as well. It was chosen primarily because of the MICE simulation set (Fosalba et al. 2015), a large high-resolution galaxy light-cone simulation tailored in part to reproduce DES observables, and accessible to us. In particular, MICE halo catalogues were used to calibrate the Halogen mock catalogues used in DES Y1 BAO analysis (Abbott et al. 2017b) and all the supporting papers including this one. In summary, the cosmological parameters in MICE cosmology are $\Omega_\mathrm{m} = 0.25$, $\Omega_\Lambda = 0.75$, $\Omega_\mathrm{b} = 0.044$, $h = 0.7$, $\sigma_8 = 0.8$, and $n_\mathrm{s} = 0.95$. Such a low matter density is no longer compatible with the current accepted value by *Planck* [$\Omega_\mathrm{m} = 0.31$ (Ade et al. 2016)]. We investigate in Section 4.2 how the BAO fit is affected when there is mismatch between the template cosmology and the cosmology of the mocks.

2.2 Theory covariance matrix

We consider both the covariance derived from the mock catalogues and an analytic Gaussian covariance model in this work. Here, we derive an expression for the Gaussian covariance matrix for two point function observables between different redshift bins, $w_{ij}(\theta)$, accounting for shot noise and angular binning. The cross-correlation between bins i and j can be expressed in terms of angular power spectra C_ℓ through a Legendre transform as follows (Peebles 1980)

$$w_{ij}(\theta) = \sum_\ell \frac{(2\ell+1)}{4\pi} P_\ell(\cos\theta) C_\ell^{ij}. \tag{7}$$

Note we do not include a term $1/n$ because it only contributes to the zero separation limit (Ross et al. 2011). In this work, we used the public code CAMB sources[1] to compute the harmonic spectra C_ℓ^{ij}. The covariance matrix between two-point correlations at different pairs of bins is then given by

$$\mathrm{Cov}[w_{ij}(\theta), w_{mn}(\theta')] = \sum_{\ell_1,\ell_2} \frac{(2\ell_1+1)(2\ell_2+1)}{(4\pi)^2} P_{\ell_1}(\cos\theta) P_{\ell_2}$$
$$\times (\cos\theta')\mathrm{Cov}[C_{\ell_1}^{ij}, C_{\ell_2}^{mn}], \tag{8}$$

which under the assumptions that the covariance scales inversely with the sky fraction in consideration (f_sky) and that in the all-sky limit the spectra band powers are diagonal, can be further written as (Crocce et al. 2011)

$$\mathrm{Cov}[w_{ij}(\theta), w_{mn}(\theta')] = \sum_\ell \frac{(2\ell+1)}{f_\mathrm{sky}(4\pi)^2} P_\ell(\cos\theta) P_\ell(\cos\theta')$$
$$\times \left[(C_\ell^{im} + \frac{\delta_\mathrm{K}^{im}}{\bar{n}_i})(C_\ell^{jn} + \frac{\delta_\mathrm{K}^{jn}}{\bar{n}_j}) + (C_\ell^{in} + \frac{\delta_\mathrm{K}^{in}}{\bar{n}_i})(C_\ell^{jm} + \frac{\delta_\mathrm{K}^{jm}}{\bar{n}_j}) \right], \tag{9}$$

where δ_K^{ab} is the Kronecker delta and $\bar{n}_i$ is the projected galaxy number density in bin i, and we have assumed Poisson shot noise.

The fact that w is measured over a finite angular binning can be taken into account if we express the above summations in terms of the bin-averaged Legendre polynomial defined as (Salazar-Albornoz et al. 2017)

$$\bar{P}_\ell = \frac{\int_{\theta_-}^{\theta_+} P_\ell(\cos\theta)\sin\theta\,\mathrm{d}\theta}{\int_{\theta_-}^{\theta_+} \sin\theta\,\mathrm{d}\theta}$$
$$= \frac{P_{\ell+1}(x_+) - P_{\ell+1}(x_-) - P_{\ell-1}(x_+) + P_{\ell-1}(x_-)}{(2\ell+1)(x_+ - x_-)}, \tag{10}$$

[1]http://camb.info/sources/

where θ_+ and θ_- (x_+ and x_-) denote the upper and lower limits of the bin (the cosine of the upper and lower limits), respectively. It is important to use the bin-averaged $\bar{P}_\ell$ in equation (9), otherwise the error is overestimated [see similar arguments in Cohn (2006), Smith, Scoccimarro & Sheth (2008), Sanchez, Baugh & Angulo (2008), and Salazar-Albornoz et al. (2017)]. The bin-averaged P_ℓ is substantially smaller than the unaveraged P_ℓ for high $\ell \gtrsim 1000$ for the typical bin width we consider $\Delta\theta \sim 0°.1\$$. This is because for high ℓ multipoles, the variation of P_ℓ across the bin width is not negligible. The effect of the bin averaging is small for the mean [i.e. equation (7)] because although the high ℓ part is inaccurate, it is rapidly oscillatory, the net effect is small. For the diagonal of the covariance, the terms are all positive, and so the effect is large.

In addition, care must be taken with the shot-noise terms, which are scale independent and hence factor out in the infinite sum over Legendre polynomials. One is left with a sum of the type $\sum_\ell (2\ell + 1) P_\ell(x) P_\ell(x^{'}) = 2\delta_D(x - x^{'})$ which is formally infinite. The fact that these are over averaged multipoles regularizes them to $\sum_\ell (2\ell + 1) \bar{P}_\ell(x) \bar{P}_\ell(x') = 2/(x_- - x_+)$ for autocorrelations (same angular bin) and zero otherwise. On the other hand, the cross terms of the form $C_\ell / \bar{n}$ do converge when summed over because C_ℓ decays at high ℓ. Hence, we treat the pure noise terms separately and perform the sum analytically. There are only two relevant cases

$$\mathrm{Cov}[w_{ij}(\theta), w_{ij}(\theta')] = \frac{\delta_K^{\theta\theta'}}{8\pi^2 f_{sky} \bar{n}_i \bar{n}_j (x_- - x_+)} + \sum_\ell \frac{2\ell + 1}{(4\pi)^2 f_{sky}}$$

$$\times \bar{P}_\ell(\cos\theta) \bar{P}_\ell(\cos\theta') \left(C_\ell^{ii} C_\ell^{jj} + + C_\ell^{ij} C_\ell^{ij} + \frac{C_\ell^{ii}}{\bar{n}_j} + \frac{C_\ell^{jj}}{\bar{n}_i} \right) \quad (11)$$

for $i \neq j$, and,

$$\mathrm{Cov}[w_{ii}(\theta), w_{ii}(\theta')] = \frac{\delta_K^{\theta\theta'}}{4\pi^2 f_{sky} \bar{n}_i^2 (x_- - x_+)} + \sum_\ell \frac{2\ell + 1}{8\pi^2 f_{sky}}$$

$$\times \bar{P}_\ell(\cos\theta) \bar{P}_\ell(\cos\theta') \left(C_\ell^{ii} C_\ell^{ii} + 2 \frac{C_\ell^{ii}}{\bar{n}_i} \right). \quad (12)$$

The remaining combinations of the indices (i, j, m, n) in equation (9) do not give rise to scale independent shot noise term, and can therefore be obtained from equation (9) without the need to sum them separately.

Finally, we note that the survey angular geometry mask does not appear explicitly in the Gaussian covariance, only the survey area through f_{sky}. In configuration space, the effect of the mask is less severe than in Fourier/harmonic space. None the less, the geometry of the mask, including the fact that it has holes, makes the number of random pairs as a function of separation not to simply scale with the effective area of the survey. Hence, the shot noise term will not exactly follow equation (9) but acquire a scale dependence, see Krause et al. (2017). We have checked that this effect is negligible for our BAO sample. Furthermore, Avila et al. (2017) compared the covariance matrix obtained from mocks to the Gaussian covariance matrix and found agreement to within 10 per cent. In Section 6, we study this issue in greater detail. Another effect, the supersample covariance due to the coupling of the small-scale modes inside the survey with the long mode when the window function is present and is only important for small scales (Hamilton, Rimes & Scoccimarro 2006; Takada & Hu 2013; Li, Hu & Takada 2014; Chan, Moradinezhad Dizgah & Noreña 2018).

2.3 Mock catalogues

We calibrate our methodology using a sample of 1800 Halogen mocks (Avila et al. 2017) that match the BAO sample of DES Y1. We outline the basic information here, and refer the readers to Avila et al. (2017), and Avila et al. (2015), for more details. For each mock realization, the dark matter particle distribution is created using second-order Lagrangian perturbation theory. Each mock run uses 1280^3 particles in a box size of $3072\,\mathrm{Mpc}\,h^{-1}$. Haloes are then placed in the dark matter density field based on the prescriptions described in Avila et al. (2015). The halo abundance, bias as a function of halo mass, and velocity distribution are matched to those in the MICE simulation (Fosalba et al. 2015). Haloes are arranged in the light-cone with the observer placed at one corner of the simulation box. The light-cone is spanned by 12 snapshots from $z = 0.3$ to 1.3. From this octant, the full sky mock is formed by replicating it eight times with periodic boundary conditions. Galaxies are placed in the haloes using a hybrid halo occupation distribution–halo abundance matching prescription that allows for galaxy bias and number density evolution.

The mocks match to the properties of the DES Y1 BAO sample. The angular mask and sample properties including the photo-z distribution, the number density, and the galaxy bias are matched. Redshift uncertainties are accurately modelled by fitting a double skewed Gaussian curve to photo-z distribution measured from the data, and this relation is then applied to the mocks. The final mock catalogue covers an area of 1318 deg² on the sky as in DES Y1. We consider the mock data in the photo-z redshift range [0.6,1], and there are close to 1.3 million galaxies per mock in this range. In the fiducial setting, the sample is further divided into four redshift bins of width $\Delta z = 0.1$. In total, we produce 1800 realizations, and we use them to calibrate the pipeline and estimate the covariance matrix. Unless otherwise stated, the mock covariance is used.

We measure the angular correlation function from the mocks with the Landy–Szalay estimator (Landy & Szalay 1993)

$$w(\theta) = \frac{DD(\theta) - 2DR(\theta) + RR(\theta)}{RR(\theta)}, \quad (13)$$

where DD, DR, and RR are the pair counts between the data–data, data–random, and random–random catalogues, respectively, normalized based on the size of catalogue. The number of objects in the random catalogue is 20 times those in the data. w is computed using the public code CUTE (Alonso 2012).

3 BAO FITTING METHODS

3.1 Methods overview

For high signal-to-noise BAO data, one expects to recover Gaussian likelihoods e.g. in the case of Alam et al. (2017). For such data, many methods for the BAO fitting would be expected to yield consistent results. But this might not be our situation, as the expected signal to noise is close to 2.

Here, we compare three methods to derive the BAO angular scale from the data, testing them thoroughly with the mocks: MLE, PL, and MCMC. We define each below. These methods differ in how they define the probability distribution of the interested variables, and how the best-fitting values and errors are computed. For a review of these statistical methods, see e.g. Press et al. (2007), Hogg, Bovy & Lang (2010), and Trotta (2017).

In the BAO fitting even though the full likelihood is multidimensional (with all the parameters in the template) we are ultimately

interested in only one, the BAO dilation parameter α. Here, we are mostly interested in which estimator gives the most reliable result for α.

Throughout, we use $\bar{\alpha}$ and σ_α to denote the best fit and the error obtained from the fitting method for an individual realization. We will use angular brackets to represent the ensemble average over the mocks: e.g. $\langle \bar{\alpha} \rangle$ is the mean of the best-fitting distribution, while std($\bar{\alpha}$) is the standard deviation of the $\bar{\alpha}$ distribution.

3.1.1 Maximum Likelihood Estimator

MLE is a point estimator for some parameters $\boldsymbol{\lambda}$, and it seeks the best fit by maximizing the likelihood function $L(D|\boldsymbol{\lambda})$, where D denotes the data set and $\boldsymbol{\lambda}$ the parameters. If the data are Gaussian distributed, and we can relate the likelihood to the χ^2 as (Press et al. 2007)

$$L \propto \exp(-\chi^2/2), \tag{14}$$

with the χ^2 defined as

$$\chi^2(\boldsymbol{\lambda}) = \sum_{i,j} [D_i - x_i(\boldsymbol{\lambda})][C^{-1}]_{ij}[D_j - x_j(\boldsymbol{\lambda})], \tag{15}$$

where C is the covariance matrix and $\boldsymbol{x}$ is the model. The best fit can be obtained by minimizing the χ^2. MLE itself does not require the likelihood to be Gaussian, and in that case χ^2 minimization can still be used to find the best fit although its connection to the probability distribution is not direct.

For the model equation (6), although there are large number of nuisance parameters B and A_i, they appear *linearly* and can be fit analytically using the least-squares fit method. Suppose that the model is given by

$$\boldsymbol{x}_i = \sum_\rho A_{i\rho} \boldsymbol{\lambda}_\rho, \tag{16}$$

where A is called the design matrix, then the best-fitting model parameters $\boldsymbol{\lambda}$ obtained by minimizing the χ^2 reads (Cowan 1998)

$$\boldsymbol{\lambda} = (A^T C^{-1} A)^{-1} A^T C^{-1} \boldsymbol{D}. \tag{17}$$

Thus in principle, we can end up with only one parameter α and its best fit can be found by a grid search. In practice, we find that sometimes we get the unphysical result $B < 0$. To avoid this unsavory situation, we impose the prior that $B > 0$. To do so, the $\chi^2(\alpha, B_i, A_i)$ is first analytically minimized with respect to the parameters A_i using equation (17) to get the best-fitting A_i^{bf}. We then numerically search for the best-fitting B_i^{bf} with the prior that $B_i > 0$ such that $\chi^2(\alpha, B_i, A_i^{\text{bf}})$ is minimized with respect to B_i. Finally, we are left with a one-parameter function $\chi^2(\alpha, B_i^{\text{bf}}, A_i^{\text{bf}})$. MLE has been adopted as a convenient choice for BAO fits in numerous recent studies [e.g. Anderson et al. (2014)].

For MLE, we use the 1σ error bar derived from the deviation from the minimum χ^2_{min} by $\Delta\chi^2 = 1$ (Lampton, Margon & Bowyer 1976; Press et al. 2007). This does assume that the likelihood of $\boldsymbol{\lambda}$ is Gaussian distributed [or $\chi^2(\boldsymbol{\lambda})$ is a quadratic function of $\boldsymbol{\lambda}$] and there is only one parameter $\boldsymbol{\lambda}$. It also applies to our case when the other parameters are maximized, see Press et al. (2007). This rule can be obtained as follows. We can expand the log of the likelihood $\ln L$ about $\alpha = \alpha_0$ where the maximum of the likelihood is attained as

$$\ln \mathcal{L}(\alpha) \approx \ln \mathcal{L}_0 + \frac{1}{2} \frac{\partial^2 \ln \mathcal{L}}{\partial \alpha^2} \bigg|_{\alpha_0} (\alpha - \alpha_0)^2. \tag{18}$$

When the Cramér–Rao bound [see Heavens (2009) for a review] is saturated, the variance of α, σ_α^2 is given by

$$\frac{1}{\sigma_\alpha^2} = -\frac{\partial^2 \ln \mathcal{L}}{\partial \alpha^2} \bigg|_{\alpha_0}. \tag{19}$$

I.e. the curvature of $\ln L$ encodes the error bar on the parameter. Hence, at 1σ from α_0, $\alpha_0 \pm \sigma_\alpha$

$$\ln \mathcal{L}(\alpha \pm \sigma_\alpha) = \ln \mathcal{L}_0 - \frac{1}{2}. \tag{20}$$

From equation (14), this is equivalent to $\Delta\chi^2 = 1$ rule for the 1σ error bar in MLE.

We will further take the symmetric error bar by averaging over the lower and upper bars. In the frequentist's interpretation, because α is a parameter, we can interpret the error bars only when the experiments are repeated. Suppose N independent measurements are repeated, we expect to have 68 per cent of the time the 1σ bars enclosing the true value ($\alpha = 1$ for the unbiased case) for a Gaussian distribution.

We will consider the likelihood in the range of $\alpha \in [0.8, 1.2]$, and BAO is regarded as being detected only if the 1σ interval can be constructed within the interval [0.8,1.2].

3.1.2 Profile Likelihood

Instead of only using the information at the maximum of the likelihood, we can compute the weighted mean and the standard deviation using the weight W

$$W(\alpha) = \frac{L(D|\alpha)}{\int d\alpha' \, L(D|\alpha')}. \tag{21}$$

Here, the partial likelihood $L(D|\alpha)$ is obtained by partially maximizing the likelihood with respect to all the other nuisance parameters except α following the procedures outlined in Section 3.1.1. The mean and variance are given by

$$\bar{\alpha} = \int d\alpha \, W(\alpha)\alpha, \tag{22}$$

$$\sigma_\alpha^2 = \int d\alpha \, W(\alpha)(\alpha - \bar{\alpha})^2. \tag{23}$$

This method sits between the MLE and MCMC in concept: it adopts MLE for parameters other than α, but performs a full marginalization for α as in the MCMC approach. The integration range is taken to be [0.8,1.2]. For PL, $\bar{\alpha}$ and σ_α^2 can always be defined, and it does not require a minimum in χ^2 to be found in the α range.

3.1.3 Markov Chain Monte Carlo

The MCMC has been widely used in cosmological parameter fitting in last couple of decades, e.g. Hinshaw et al. (2013), Ade et al. (2016), and Abbott et al. (2017a), mainly because it can efficiently sample large number of parameters. MCMC is based on the Bayes theorem

$$P(\boldsymbol{\lambda}|D) = \frac{P(\boldsymbol{\lambda})L(D|\boldsymbol{\lambda})}{P(D)}, \tag{24}$$

where $P(\boldsymbol{\lambda})$ is the prior distribution, and $P(D)$ is the probability of the data (often called the evidence).

In the Bayesian approach, α is a random variable, we can talk about the chance that the α value lies in the 1σ interval. Strict

Bayesians will stop at the posterior distributions as their final product, but to compare with other methods, we will deviate from the strict Bayesianism and use the posterior distributions to compute the summary statistics (Hogg et al. 2010). If the distribution is Gaussian, we expect 68 per cent of chance. We use the median of the MCMC chain for $\bar{\alpha}$, and σ_α is derived from 16 and 84 percentiles of the chain. Again to facilitate the comparison with other methods, we average over the left and right error bars to get a symmetric one. Alternatively, we can use the mean and the standard deviation for the best fit and its error bar. We opt for the median and the percentiles because we find that the results encloses $\langle \bar{\alpha} \rangle$ closer to the Gaussian expectation. The prior on α is taken to be [0.6,1.4]. Similar to PL, $\bar{\alpha}$ and σ_α can always be defined.

We use the MCMC implementation EMCEE (Foreman-Mackey et al. 2013), in which multiple walkers are employed and the correlation among the walkers are reduced by using the information among them. We use 100 walkers, 3000 burn-in steps, and 2000 steps for the run. In this setting, the MCMC fitting code takes about 50 times longer than the MLE code does. We have conducted some convergence test on the number of steps required. We took a sample of steps in the range from 1500 to 50 000 and find that fluctuations of $\bar{\alpha}$ are within 0.1 per cent from the convergent value (assuming convergence attained with 50 000 steps) and those of σ_α within 2 per cent. Thus, 2000 steps are sufficient to make sure that it does not impact the results later on. We note that the usage of the percentiles of the distribution is much less susceptible to statistical fluctuations than using the mean and the variance.
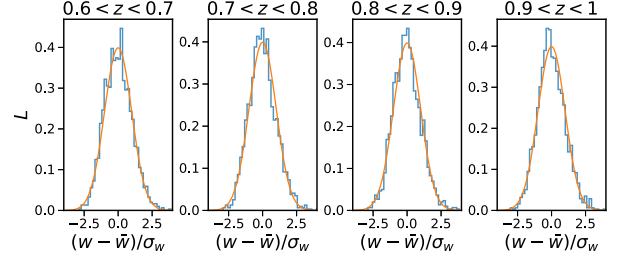
3.1.4 Comparison criteria

We will test these methods against the mock catalogues. We will check how stable the fit results are, especially how small a bias (in comparison to the known true value) that each method yields.

The full information is in the likelihood/posterior. The best fit and its error matter because we want to effectively represent the distribution by these two numbers. Here are some features that the desirable summary statistics should have. First, we want the estimator to yield an unbiased mean result. Second, in order to have the proper probabilistic interpretation of the error bar, it is desirable that the error bar encloses the true answer close to the Gaussian expectation, which is 68 per cent for 1σ. Hence, we will check the fraction of times that the error bar derived encloses the true answer. Another useful metric is that the standard deviation of the distribution, std($\bar{\alpha}$) agrees with the mean of the error derived $\langle \sigma_\alpha \rangle$. One way to compare the deviation of $\bar{\alpha}$ from the expected true answer (the spread of the $\bar{\alpha}$ distribution) and the error derived σ_α is to consider the normalized variable

$$d_{norm} = \frac{\bar{\alpha} - \langle \bar{\alpha} \rangle}{\sigma_\alpha}. \tag{25}$$

The distribution of this variable can be compared to a unit normal distribution as a test of the Gaussianity of the recovered results.

Note that we do not enforce the error distribution to be Gaussian, indeed they are not (see Fig. 3). We only want the 1σ error bar to enclose the true answer close to the Gaussian expectation. In principle, for MLE, we can adjust the value of $\Delta\chi^2$ so that it encloses the true value (e.g. $\alpha = 1$) 68 per cent of the time. Similar adjustments can be done for other estimators. By doing so, the derived error bar yields the desired Gaussian probability expectation. However, the $\Delta\chi^2 = 1$ rule works well for us, and no adjustment is needed.

Figure 1. The likelihood distribution of w at the bin $\theta = 2°$ (histogram). The standard normal variable is used. The Gaussian distribution with zero mean and unity variance (solid line) is plotted for comparison.

3.2 Comparison of the BAO fit results by MLE, PL, and MCMC

Following Abbott et al. (2017b), we consider those fits with their 1σ intervals of α falling outside the range [0.8,1.2] as non-detections. These non-detections are poorly fit by our template, and they cause the distribution of $\bar{\alpha}$ to be highly non-Gaussian. Thus, we will remove the non-detection mocks first. We will comment more on this at the end of the section.

Before proceeding to the comparison, we first verify the Gaussian likelihood assumption equation (14) using the mock catalogues. The likelihood tends to be Gaussian distributed thanks to the central limit theorem. Because the Gaussian likelihood assumption is central to the analysis, we need to check it [e.g. Scoccimarro (2000); Hahn et al. (2018)]. In Fig. 1, we show the likelihood distribution of the values of w measured from the mocks. We have shown the results for the bin $\theta = 2°$. We have used the standard normal variable ($w - \bar{w})/\sigma_w$, where $\bar{w}$ and σ_w are the mean and the standard deviation of the distribution of w. We find that the likelihood indeed follows the Gaussian distribution well.

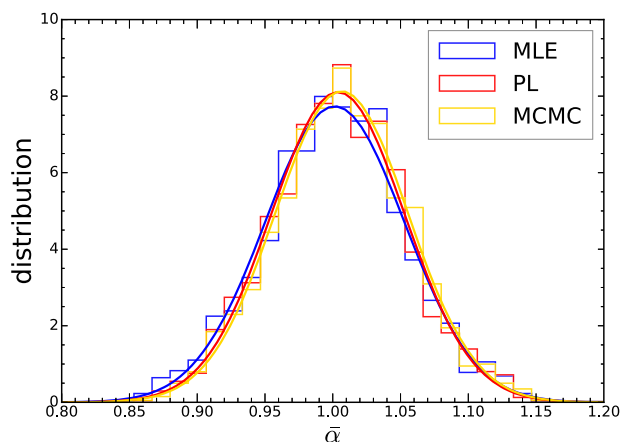In Table 1, we show the fit with MLE, PL, and MCMC for two detection criteria. First, for MLE with error derived from $\Delta\chi^2 = 1$, there are 91 per cent of the mocks with their 1σ error bars fall within the interval of [0.8,1.2], while for MCMC the fraction is 84 per cent. For PL, it is also almost 100 per cent.

In Fig. 2, we show the distribution of the best-fitting α obtained with these three methods after pruning the non-detection mocks. As a comparison we have also plotted the Gaussian distribution with the same mean and variance, we find that $\bar{\alpha}$ follows the Gaussian distribution well. The mean of the best fit from MLE is the least biased among the three methods. The distributions of $\bar{\alpha}$ from PL and MCMC are quite similar. They tend to be more skewed towards $\bar{\alpha} > 1$, and this can be seen from their corresponding Gaussian distribution and the $\langle \bar{\alpha} \rangle$ shown in Table 1. In particular for MCMC, $\langle \bar{\alpha} \rangle$, it is larger than 1 by 0.007. In Table 1, the fraction of mocks with the 1σ error bar enclosing the $\langle \bar{\alpha} \rangle$ is also shown. We find that MLE with $\Delta\chi^2 = 1$ prescription encloses $\langle \bar{\alpha} \rangle$ 0.69 of the time, which is very close to the Gaussian expectation, while PL and MCMC are higher than the Gaussian expectation by 9 per cent and 6 per cent, respectively.
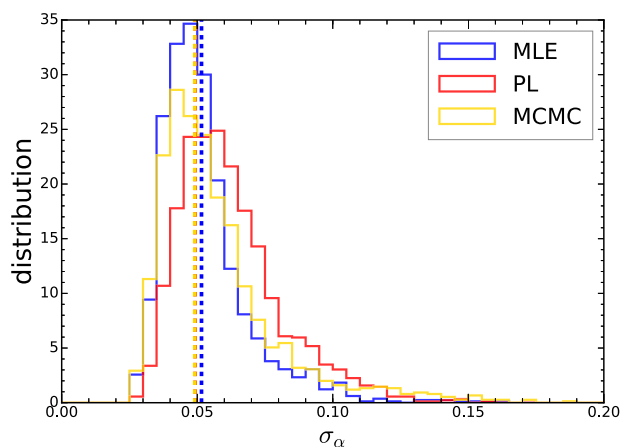
In order to see the error derived more clearly, in Fig. 3, the distribution of the errors derived from these three methods are shown. As a comparison, we also show the std($\bar{\alpha}$). In the caption of Fig. 3 we have given the numbers for the $\langle \sigma_\alpha \rangle$ and std($\bar{\alpha}$) for MLE, PL, and MCMC. The $\langle \sigma_\alpha \rangle$ obtained from the MLE method coincides with std($\bar{\alpha}$) well, while MCMC gives slightly larger error σ_α. PL tends to give the largest error with $\langle \sigma_\alpha \rangle$/std($\bar{\alpha}$) = 1.27.

Table 1. The BAO fit with MLE, PL, and MCMC obtained with two sets of detection criteria (left and right). For Detection Criterion 1, we only consider those mocks whose 1σ interval $\bar{\alpha} \pm \sigma_\alpha$ fall within the interval [0.8,1.2]. For Detection Criterion 2, we use the same set of mocks for all three methods, those having 1σ interval falling within [0.8,1.2] using the MLE estimate of $\bar{\alpha}$ and σ_α. For each criterion, the first fraction is normalized with respect to the total number of mocks (1800), while the second fraction is normalized with respect to the number of mocks satisfying the selection criterion.

	Detection Criterion 1			Detection Criterion 2		
	$\bar{\alpha} \pm \sigma_\alpha$ in [0.8,1.2] (Fraction selected)	$\langle\bar{\alpha}\rangle \pm \mathrm{std}(\bar{\alpha})$	Fraction with $\bar{\alpha} \pm \sigma_\alpha$ enclosing $\langle\bar{\alpha}\rangle$	$\bar{\alpha} \pm \sigma_\alpha$ in [0.8,1.2] MLE (Fraction selected)	$\langle\bar{\alpha}\rangle \pm \mathrm{std}(\bar{\alpha})$	Fraction with $\bar{\alpha} \pm \sigma_\alpha$ enclosing $\langle\bar{\alpha}\rangle$
MLE	0.91	1.001 ± 0.052	0.69	0.91	1.001 ± 0.052	0.69
PL	0.99	1.004 ± 0.049	0.77	0.91	1.003 ± 0.046	0.78
MCMC	0.84	1.007 ± 0.049	0.74	0.91	1.007 ± 0.059	0.74

Figure 2. The histograms show the distribution of the best-fitting $\bar{\alpha}$ obtained using MLE (blue), PL (red), and MCMC (yellow). The solid lines (blue for MLE, red for PL, and yellow for MCMC) are the Gaussian distributions with the same mean and variance as the corresponding histograms.

Figure 4. The distribution of the normalized variable d_{norm}. The results obtained with MLE (blue), PL (red), and MCMC (yellow) are shown. The Gaussian distribution with zero mean and unity variance (solid, black) are shown for reference.

Figure 3. The distribution of the error derived from each individual mock (blue for MLE, red for PL, and yellow for MCMC). The vertical dashed lines are the standard deviations of the best-fitting $\bar{\alpha}$ shown in Fig. 2. The $\langle\sigma_\alpha\rangle$ and $\mathrm{std}(\bar{\alpha})$ for MLE, PL, and MCMC are (0.053, 0.052), (0.062, 0.0492), and (0.057, 0.049) respectively. While for MLE $\langle\sigma_\alpha\rangle$ and $\mathrm{std}(\bar{\alpha})$ coincide, PL and MCMC yield larger $\langle\sigma_\alpha\rangle$.

We plot the distributions of the normalized variable for these methods in Fig. 4. By comparing with the Gaussian distribution with zero mean and unity variance, we find that MLE with $\Delta\chi^2 = 1$, d_{norm} is more Gaussian, while MCMC is slightly worse, and PL is the least Gaussian.

In the detection criterion we have applied, the non-detections are different for each of the methods. We therefore wish to check if the differences we have found in the results are simply due to that. We now apply the same selection criterion that the 1σ interval obtained from MLE falls within [0.8,1.2] to all the three methods. The results are shown in Table 1 as Detection Criterion 2. We find that the results are essentially the same as the previous one, ruling out that the differences are due to the selection criterion. Note that for MCMC, the $\mathrm{std}(\bar{\alpha})$ has increased quite appreciably because using the MLE detection criterion, a large fraction of extreme mocks are retained and they cause a big increase in the $\mathrm{std}(\bar{\alpha})$. In the rest of the analysis, we will stick to the original pruning criterion.

The choice of the interval [0.8,1.2] is somewhat arbitrary, but based on past results [e.g. Anderson et al. (2014); Ross et al. (2017a)] that suggest it is a reasonable choice (though, admittedly, more a rule of thumb than anything else). To further justify this choice, in the Appendix A, we show the comparison of results obtained with a larger interval [0.6,1.4]. We find that by including the extreme mocks, the distribution of the best-fitting $\bar{\alpha}$ exhibits strong tails, and it does not agree with the Gaussian distribution with the same mean and variance. These extreme mocks not only enlarge the mean of the error bars $\langle\sigma_\alpha\rangle$, they also cause bias in $\langle\bar{\alpha}\rangle$. In particular for MLE, $|\langle\bar{\alpha}\rangle - 1|$ changes from 0.001 (Table 1) to 0.029 (Table A1). For MCMC, it only changes mildly from 0.007 to 0.010. Overall, we find that when the wider interval [0.6,1.4] is adopted, the MCMC is superior to the other two methods because it gets less biased results, and the derived 1σ intervals which enclose $\langle\bar{\alpha}\rangle$ is

71 per cent of the mocks, the closest to the Gaussian expectation. Moreover, $\langle \sigma_\alpha \rangle / \mathrm{std}(\bar{\alpha}) = 0.92$ is close to 1.

Because the small fraction of the mocks without a BAO detection highly bias the distributions of the best-fitting parameters, and since it would be difficult to extract useful BAO information from them, we adopt a smaller interval to get rid of them. In practice, if the best fit with its 1σ interval is outside the range [0.8,1.2] in our actual data, the data would be poorly fitted (or the error bar poorly estimated) by our existing methodology and we can hardly claim to detect the BAO signals in the data. In this case, we may need to change the fiducial cosmological model or wait for additional data.

Thus, overall we find that after pruning the non-detections, MLE yields the estimate with the least bias and the error bar using $\Delta \chi^2 = 1$ results in the 1σ error closest to the Gaussian expectation. Therefore, if we prune the data, MLE furnishes an effective BAO fitting method. For the rest of the analysis, we always use the pruned data.

4 TEMPLATE SYSTEMATICS

In this section, we study the potential systematics associated with the template and investigate how they can affect the BAO fit results. We will first determine the physical damping scale Σ in the template by fitting to the mock catalogue. Although the damping scale does not bias the mean for the BAO fit, it can strongly affect the error bar that is derived. We then examine how the BAO fit is affected when the template does not coincide precisely with the data, i.e. the BAO scale in the template is different from that in the data. The parameter α is introduced to allow the shift in the BAO scale, it is crucial to check how successful it deals with the mismatch.

The BAO angular scale θ_{BAO} can be estimated as
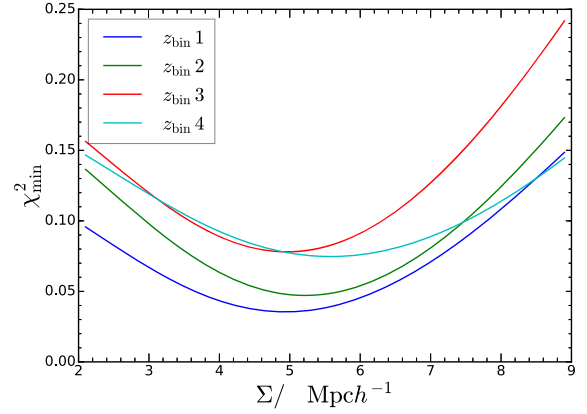
$$\theta_{\mathrm{BAO}} = \frac{r_s}{D_A}, \tag{26}$$

where r_s is the comoving sound horizon at the drag epoch and D_A denotes the *comoving* angular diameter distance [as in Weinberg et al. (2013)] and in a flat cosmology it is given by

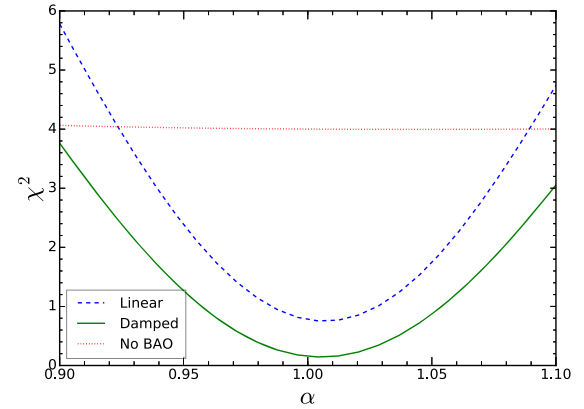$$D_A = \int_0^z \frac{c\,\mathrm{d}z'}{H(z')}, \tag{27}$$

where H is the Hubble parameter and c is the speed of light. Here, we consider two ways that the template mismatches the data. First, the cosmology for the data could be different from that of the fiducial cosmology. Another possibility is that there could be photo-z error causing $\phi(z)$ to be systematically biased. Both possibilities can cause shifts in θ_{BAO}. Using a wrong cosmology changes both r_s and D_A.

4.1 The BAO damping scale

To determine the correct physical damping scale, we fit the templates with only one fitting parameter Σ to the mean w of the mocks (i.e. $\alpha = 1$, $B = 1$, and $A_i = 0$). We have considered four redshift bins and fitted to each redshift bin separately. The minimum of the χ^2 obtained with MLE is plotted against the damping scale in Fig. 5. The best-fitting damping scale is in between 5 and 6 Mpc h^{-1} across the redshift bins. However, we note that the highest redshift bin, bin 4, requires the largest damping. This is contrary to the expectation that the damping scale should decrease as redshift increases because the non-linearity becomes weaker. This is not due to the photo-z distribution because of the following reason. Although bin 4 has the largest photo-z uncertainty, which can blur the BAO, it is taken into

Figure 5. The minimum of the χ^2 against the BAO damping scale Σ for the four redshift bins used.

Figure 6. The χ^2 as a function of α when the template without damping (dashed, blue) and with the damping scale $\Sigma = 5.2\,\mathrm{Mpc}\,h^{-1}$ (solid, green) are used. The fit with a no-BAO template (dotted, red) is also shown.

account in $\phi(z)$ already and does not affect the more fundamental 3D damping scale. Throughout this work, we simply take the mean of the best fit of the four redshift bins, which is $\Sigma = 5.2\,\mathrm{Mpc}\,h^{-1}$. We have checked that the differences between using the mean damping scales and the individual best fit results in no detectable change in α.

It is instructive to see how the fit is affected when the template without damping is used. To check this, we compare the fits with $\Sigma = 5.2\,\mathrm{Mpc}\,h^{-1}$ against the one without damping i.e. $\Sigma = 0\,\mathrm{Mpc}\,h^{-1}$. We jointly fit to the mean of the four bins and the results are shown in Fig. 6. The linear case yields a larger overall value of χ^2 than the damped one. As a comparison, we also show the fit using a template without BAO, and it is clearly disfavoured compared to the BAO model. We also see that the χ^2 bound is narrower for the linear template, and hence using it we would get an artificially tighter bound on α. On the other hand, the extreme case of no BAO, there is no bound on α at all.

In Eisenstein et al. (2007) [see also Seo & Eisenstein (2007)], the BAO damping is modelled using the differential Lagrangian displacement field between two points at a separation of the sound horizon. Under the Zel'dovich approximation (Zel'dovich 1970), at $z_{\mathrm{eff}} = 0.8$ the spherically averaged damping scale is $5.35\,\mathrm{Mpc}\,h^{-1}$, which is close to our recovered value. Note that the damping scale is obtained by integrating over the linear power spectrum [eq. (9) in Eisenstein et al. (2007)], and in *Planck* cosmology, we get $4.96\,\mathrm{Mpc}\,h^{-1}$ instead. Thus in principle, we can allow the damp-

ing scale to be a free parameter. In Abbott et al. (2017b), we have tested the results obtained using different damping scales (2.6 and 7.8 Mpc h^{-1}), consistent with the trend shown in Fig. 6, the damping scale has only small effect on the best fit, but a smaller damping scale results in a smaller error bar. Given the quality of the current data, we fix the BAO damping scale to be 5.2 Mpc h^{-1}.

4.2 Incorrect fiducial cosmology

The mock catalogues were created using the MICE cosmology. To test the cosmology dependence, we fit a template computed using the *Planck* cosmology (Ade et al. 2016). The *Planck* cosmology should be closer to the current cosmology and it is used in the other analyses, such as the Baryon Oscillation Spectroscopic Survey (BOSS) DR12 (Ross et al. 2017a). For the *Planck* cosmology, we use $\Omega_m = 0.309$, $\sigma_8 = 0.83$, $n_s = 0.97$, and $h = 0.676$. From CAMB (Lewis et al. 2000), in the MICE cosmology, the sound horizon at the drag epoch is 153.4 Mpc, while in the *Planck* cosmology it is 147.8 Mpc. For the effective redshift $z_{eff} = 0.8$, D_A in MICE cosmology is larger than that in the *Planck* one by 3.1 per cent. Thus, we expect the BAO in the *Planck* cosmology to be smaller than that in the MICE by 4.1 per cent in angular scale.

In Fig. 7, we plot the matter angular correlation function obtained using the MICE and *Planck* cosmology. To check how well the rescaling parameter α works, we rescale the angular correlation by $w(\alpha\theta)$ and then match the amplitude at the BAO scale with that of the MICE one. We find that the rescaling results in a good match with the MICE one around the BAO scale. However, away from the BAO scale, the disagreement with the MICE template increases. Since we use the angular scale in the range of $[0°.5, 5°]$ in the fitting, it is not clear if we can recover the true cosmology.

In Table 2, we compare the fit obtained using the MICE template with that from the *Planck* cosmology one. We have displayed the fits using the mean of the mocks and the individual mocks. As the mean result is obtained by averaging over 1800 realizations, the covariance is reduced by a factor of 1800. We have shown four decimal places as the error bars are small for the mean fit. In this very high signal-to-noise setting, there is still $\bar{\alpha} - 1 \sim 0.004$ bias in the best fit for all these estimators. It is much larger than the estimated error bar $\sigma_\alpha \sim 0.0013$, thus the bias is systematic. This can arise from the non-linear effects such as the dark matter non-linearity and galaxy bias (Crocce & Scoccimarro 2008; Padmanabhan & White 2009). When the *Planck* template is used, the best-fitting α is about 0.9674. Thus, the ratio of the dilation parameter is $\bar{\alpha}_{Planck}/\bar{\alpha}_{MICE} = 0.9634$ and it is still smaller than the expectation 0.959 by $\sim 3\sigma_\alpha$. Of course this high signal-to-noise case occurs if we have 1800 times the DES Y1 volume. For the present DES volume, it is still statistical error dominated as can be seen below.

Now we turn to the distribution of $\bar{\alpha}$. Since the expected $\bar{\alpha} \approx 0.96$ is not symmetric about the interval $[0.8, 1.2]$ for the case of *Planck* cosmology. This asymmetry can bias the distribution of $\bar{\alpha}$. To prevent this bias, we use the detection criterion that the 1σ error bounds fall within the boundary $[0.76, 1.16]$ for the case of *Planck* cosmology. For MCMC, the prior on α is changed to $[0.56, 1.36]$, although this results in no detectable significance compared to the fiducial choice $[0.6, 1.4]$. $\langle\bar{\alpha}\rangle$ shows larger variations than the mean fit case because the covariance is larger by a factor of 1800. The ratio of the dilation parameter $\langle\bar{\alpha}_{Planck}\rangle/\langle\bar{\alpha}_{MICE}\rangle$ are 0.964, 0.965, and 0.967 for MLE, PL, and MCMC, respectively. The ratio is close to what we get for the mean fit. Note that if we keep the boundary of α to be $[0.8, 1.2]$ for the *Planck* case, we get $\langle\bar{\alpha}\rangle = 0.970$, 0.976, and 0.979, resulting in larger difference from the mean fit results. It

is worth mentioning that this adjustment of the interval only affects the mocks with more extreme values of the best fit.

The small difference between the fit results and the estimation from equation (26) could be because we have used the information not just about the BAO scale, but also the shape of the correlation around it. This is related to the polynomial A_i used to absorb the broad-band power dependence. We have checked that the bias in the fit reduces with the order of polynomial in A_i. For example, for the case of mean fit, for $N_p = 1, 2$, and 3, $\bar{\alpha} - 1$ are -0.095, -0.054, and -0.034, respectively. For large range in the angles, the polynomial cannot perfectly remove the dependence. However for $N_p = 3$, the bias in the best-fitting $\bar{\alpha}$ is reduced to a basically negligible level given the signal-to-noise ratio of our data, suggesting our results should be robust at least for true cosmologies between MICE and *Planck*. Of course one can use a template that is closer to the currently accepted cosmology, although at the risk of the confirmation bias.

4.3 Photo-z error

Suppose that the true photo-z distribution is ϕ_0, but because of some photo-z error $\delta\phi$, which can arise from systematic errors in the photo-z calibration, the photo-z distribution $\phi = \phi_0 + \delta\phi$ is used. In this section, we test how the photo-z error affects the BAO fit.

We will investigate two types of photo-z errors in the templates: in one case, the means of the photo-z distributions are systematically shifted by 3 per cent in each of the four tomographic redshift bins, for the second case, the standard deviations of the photo-z distributions are increased by 20 per cent. While the typical relative error in the mean of the redshift distributions for the DES Y1 BAO sample is about 1 per cent (Gaztañaga et al. 2018), 3 per cent is used to increase the signal-to-noise ratio of our systematic test. The 20 per cent error in the spread is the upper bound of the relative error in the width of those redshift distributions (Gaztañaga et al. 2018).
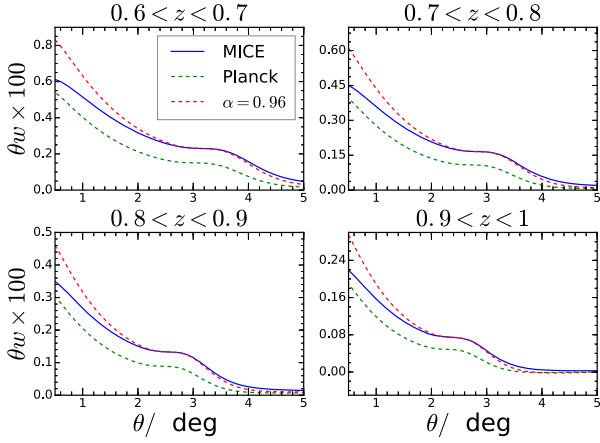
As we can see from Fig. 8, the photo-z distribution is approximately Gaussian. If ϕ_0 is Gaussian, we parametrize $\delta\phi$ as variations in the mean μ and variance σ^2 by an amount of $\delta\mu$ and $\delta\sigma^2$ as

$$\delta\phi(z) = \phi_0(z)\left[\frac{z-\mu}{\sigma^2}\delta\mu + \left(\frac{(z-\mu)^2}{\sigma^2} - 1\right)\frac{\delta\sigma}{\sigma}\right]. \quad (28)$$

Note that $\int dz\, \delta\phi(z) = 0$, thus $\phi_0 + \delta\phi$ is properly normalized. However, $\phi_0 + \delta\phi$ is not always positive in the whole range of z, so it is not really a legitimate probability distribution. None the less, since the variations we consider are small, the negative part is small. Furthermore, the $\delta\mu$ term, which is odd about μ, causes skewness about the mean in $\phi + \delta\phi$. All these suggest that $\phi_0 + \delta\phi$ may not be close to Gaussian any more. If so the parameter $\delta\mu$ ($\delta\sigma$) may not be the variation of the mean (standard deviation) of $\phi_0 + \delta\phi$ relative to that of ϕ_0. Indeed, we find that the output variation (by direct computation) from equation (28) does not match that from the input ($\delta\mu$ and $\delta\sigma$). Thus, for the case of mean shift, we will simply translate the distribution by certain amount instead of using equation (28). For the variation of the standard deviation, we adjust the value of $\delta\sigma$ in equation (28) to get the desired variation.

Because the photo-z error causes variation in the mean redshift of the slice and hence D_A, from equation (26), we can get a simple estimate for the shift of θ_{BAO} due to a shift in z by an amount of $\delta\mu$

$$\frac{\delta\theta_{BAO}}{\theta_{BAO}} = -\frac{c\,\delta\mu}{H(z)D_A(z)}. \quad (29)$$

Figure 7. The angular correlation obtained using the MICE (solid, blue) and *Planck* (dashed, green) cosmologies. The *Planck* correlation rescaled by $\alpha = 0.96$ (red) and shifted in amplitude to match the MICE one.

At the effective redshift $z_{\mathrm{eff}} = 0.8$, when $\delta\mu$ increases by 3 per cent, $\delta\theta_{\mathrm{BAO}}/\theta_{\mathrm{BAO}}$ changes by -2.4 per cent. This simple argument suggests that the leading effect comes from the systematic shift in the mean, while the variation in the standard deviation is symmetric about the mean to the lowest order and does not shift the BAO.

In Table 3, we compare the fit results obtained using the fiducial template, the template with 3 per cent increase in the mean of $\phi(z)$, and the one with 20 per cent increase in σ. For the fit to the mean, there is a shift of -0.0254, -0.0251, and -0.0239 for MLE, PL, and MCMC, respectively, when the mean of ϕ is systematically shifted by 3 per cent. When ϕ is systematically shifted by 3 per cent, we use the interval [0.776, 1.176] for α instead. The shift for $\langle\bar{\alpha}\rangle$ are -0.024 relative to the fiducial case for all the estimators. On the other hand, when the standard deviation of the distributions are increased by 20 per cent, there is no significant systematic trend for the best fits. Although not shown explicitly here, there is about 2 per cent increase in $\langle\sigma_\alpha\rangle$ for all the estimators, signalling that the error indeed increases.

Good agreement between the full fit results and the estimation by equation (29) validates the simple argument. This is useful because it provides a convenient estimate for the accuracy of the photo-z required for the BAO fit. For example, from the redshift validation (Gaztañaga et al. 2018), the mean of the photo-z error after sample variance correction is about 1 per cent, thus the shift in the BAO position is about 0.8 per cent. This is still marginal compared to other potential systematic shifts and is less than 20 per cent of the statistical uncertainty. For DES Y3, the amount of data is expected to increase roughly by a factor of 3, and so the error on α is expected to reduce by almost a factor of $\sqrt{3}$. Hence, the photo-z uncertainty is still expected to be subdominant for Year 3. The shift in the mean of the photo-z distribution has similar effects as the parameter α,

and hence they would be degenerate with each other. It is important that the photo-z distribution is calibrated in an independent means e.g. using the external spectroscopic data.

5 OPTIMIZING THE ANALYSIS

To derive the fit, an accurate precision matrix, i.e. the inverse of the covariance matrix, is necessary. In Abbott et al. (2017b), the workhorse covariance is derived from the mock catalogues. To get an accurate precision matrix from the mocks, we want the dimension of the covariance matrix to be small relative to the number of mocks (Anderson 2003; Hartlap, Simon & Schneider 2007). Hence, it is desirable to reduce the number of data bins while preserving the information content. To this end, we consider optimizing of the number of angular bins, the number of redshift bins, and the inclusion of cross redshift bins.
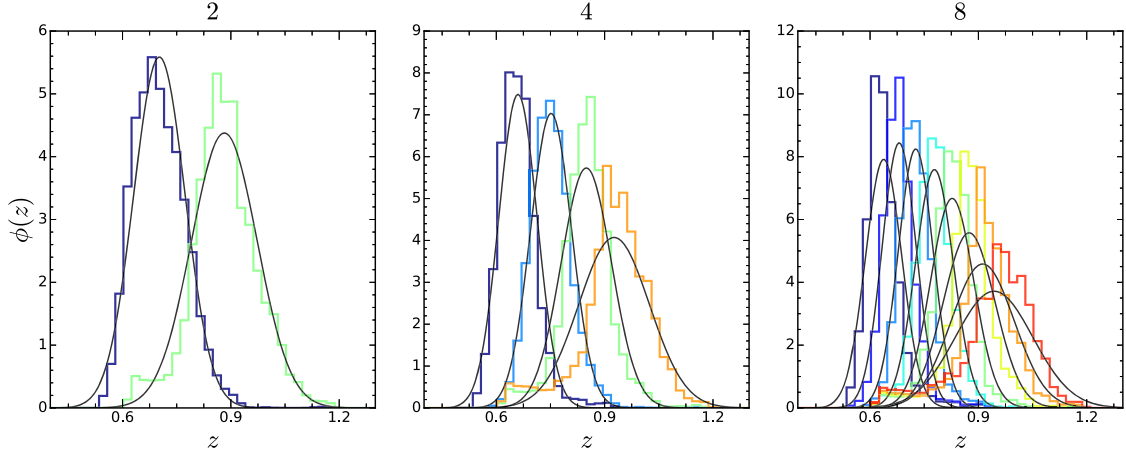
5.1 The angular bin width

We use the following method to check the dependence of the BAO fit results on the angular bin size. We generate a theory template with fine angular bin width ($0\overset{\circ}{.}01$ here). A mock data vector with coarser bin width is created by bin averaging the theory template over the bin width. The mock data vector is then fitted using the fine template. In the analysis, we use the Gaussian covariance equation (12). As we see in equation (20), for the error analysis, the magnitude of the χ^2 does not matter, and only the deviation from the best fit does. This method enables us to explore the likelihood about its maximum, and hence derive the strength of the constraint. It is similar to the often used Fisher forecast (Tegmark, Taylor & Heavens 1997; Dodelson 2003).

In Fig. 9, we show the distribution of χ^2 as a function of the dilation parameter α for a number of angular bin widths. In this ideal noiseless setting, the best fit can fit the mock data extremely well, as manifested with $\chi^2 \approx 0$ at $\alpha = 1$. The $\Delta\chi^2 = 1$ rule can give the 1σ constraint on α. For the angular bin width $\Delta\theta = 0\overset{\circ}{.}05$, $0\overset{\circ}{.}1$, $0\overset{\circ}{.}15$, $0\overset{\circ}{.}25$, $0\overset{\circ}{.}3$, and $0\overset{\circ}{.}4$, the 1σ error bars are 0.0522, 0.0521, 0.0520, 0.0514, 0.0509, and 0.0503, respectively. The mock data with coarser bin width yields slightly smaller 1σ error bar because they give a slightly less precise representation of the underlying model. This is often the case when a poor model is used, it gives larger χ^2 (the definition of a poor model) and an artificially stringent constraint.

For BAO fitting, it is preferable to have the bin width to be fine enough so that there are a few data points in the BAO dip-peak range to delineate the BAO feature. Thus for the rest of the study, we shall stick to $\Delta\theta = 0\overset{\circ}{.}15$. In Abbott et al. (2017b), the covariance is derived from the mock catalogues. As we find that the differences between different bin widths results are negligible, to reduce the size of the data vector, $\Delta\theta = 0\overset{\circ}{.}3$ was adopted in Abbott et al. (2017b).

Table 2. The BAO fit obtained with MLE, PL, and MCMC for the MICE and *Planck* cosmologies. The fits to the mean of the mocks and the individual mocks are shown.
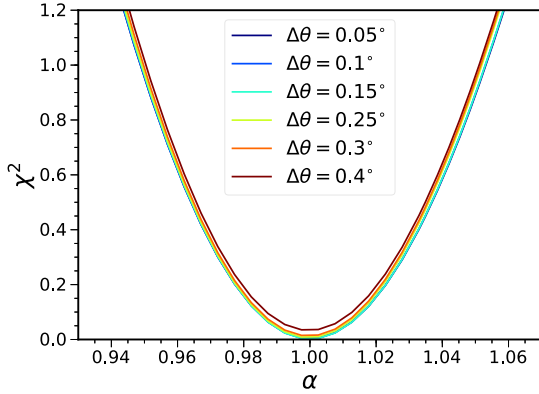
	MICE			*Planck*		
	Best fit to mean of mocks	$\langle\bar{\alpha}\rangle \pm \mathrm{std}(\bar{\alpha})$ (all mocks)	Fraction with $\bar{\alpha} \pm \sigma_\alpha$ enclosing $\langle\bar{\alpha}\rangle$	Best fit to mean of mocks	$\langle\bar{\alpha}\rangle \pm \mathrm{std}(\bar{\alpha})$ (all mocks)	Fraction with $\bar{\alpha} \pm \sigma_\alpha$ enclosing $\langle\bar{\alpha}\rangle$
MLE	1.0043 ± 0.0013	1.001 ± 0.052	0.69	0.9675 ± 0.0014	0.965 ± 0.050	0.70
PL	1.0043 ± 0.0013	1.004 ± 0.049	0.77	0.9673 ± 0.0014	0.969 ± 0.048	0.78
MCMC	1.0030 ± 0.0012	1.007 ± 0.049	0.74	0.9674 ± 0.0014	0.974 ± 0.047	0.75

Figure 8. The photo-z distribution from the mock catalogue as a function of the true redshift z for a number of redshift bins $N_z = 2$, 4, and 8 bins, respectively. The solid black lines show the Gaussian distribution with the same mean and variance as the photo-z distribution.

Table 3. The BAO fit using the fiducial photo-z distribution, the distribution with mean shift by 3 per cent, and the distribution with standard deviation increased by 20 per cent. The fits to the mean of the mocks and the individual mocks are shown.

	Fiducial $\phi(z)$		3 per cent increase in the mean of $\phi(z)$		20 per cent increase in the std of $\phi(z)$	
	Best fit to	$\langle \bar{\alpha} \rangle \pm \text{std}(\bar{\alpha})$	Best fit to	$\langle \bar{\alpha} \rangle \pm \text{std}(\bar{\alpha})$	Best fit to	$\langle \bar{\alpha} \rangle \pm \text{std}(\bar{\alpha})$
	mean of mocks	(all mocks)	mean of mocks	(all mocks)	mean of mocks	(all mocks)
MLE	1.0043 ± 0.0013	1.001 ± 0.052	0.9789 ± 0.0014	0.977 ± 0.051	1.0043 ± 0.0012	1.001 ± 0.051
LP	1.0043 ± 0.0013	1.004 ± 0.049	0.9792 ± 0.0013	0.980 ± 0.049	1.0043 ± 0.0013	1.004 ± 0.049
MCMC	1.0030 ± 0.0012	1.007 ± 0.049	0.9791 ± 0.0013	0.983 ± 0.049	1.0042 ± 0.0013	1.007 ± 0.049

Figure 9. χ^2 as a function of α for different angular bin widths. The models with different bin width $\Delta\theta$ yield similar results.

Figure 10. The χ^2 as a function of α for different number of redshift bins $N_z = 2$, 4, 5, 6, and 8, respectively. The black dash line indicates the $\Delta\chi^2 = 1$ threshold.

5.2 Number of photo-z bins

For a sample that relies on photo-zs, there can be large overlaps among the photo-z distributions from different redshift bins, and thus substantial covariance between different redshift bins. Here, we test how the constraints depend on the number of redshift bins.

Based on the assigned photo-z in the mock, we divide the samples in the redshift range [0.6,1] into N_z redshift bins, with equal width in z. For example, we have shown the photo-z distribution $\phi(z) \equiv P(z|z_{\text{photo}})$ for $N_z = 2$, 4, and 8. We see that there is indeed large overlap in the photo-z distribution. We also plotted the Gaussian distribution with the same mean and variance as the photo-z distribution. The photo-z distribution is moderately Gaussian and the deviation from Gaussianity increases with z. In fact, Avila

et al. (2017) find that a double Gaussian distribution offers a better fit to ϕ.

Similar to that in Section 5.1, using the photo-z distribution shown in Fig. 8, we generate templates with fine bin width ($\Delta\theta = 0°.01$), data vectors with coarser bin width ($\Delta\theta = 0°.15$), and the Gaussian covariances. The template needs to smoothly represent the theory, and $\Delta\theta = 0°.01$ is sufficient. The χ^2 as a function of α for different N_z are displayed in Fig. 10. We see that for $N_z = 2$, the data are undersampled and the 1σ error bar on α is weak (at $\Delta\chi^2 = 1$, it is 0.07). When N_z is increased to 4, the error bar is tightened to 0.05 at $\Delta\chi^2 = 1$. Further increasing the number of redshift bins, there is

little gain, and the change in 1σ error bar is less than 0.005 for N_z = 8. To strike a balance between retaining as much information as possible and keeping the size of the data vector small, we use N_z = 4 for the rest of the work and in Abbott et al. (2017b).

In Abbott et al. (2017b), different combinations of redshift bins of width 0.1 in the range [0.6–1] are used to do the BAO fit. In Table 4, we examine how the results vary with the number of redshift bins used. We consider the fit using four redshift bins (bins 1, 2, 3, and 4), three redshift bins (bins 2, 3, and 4), and two redshift bins (bins 3 and 4). Recall that the redshift bins are in ascending order, e.g. bin 1 refers to [0.6,0.7], etc. When the number of redshift bins is reduced, the constraining power of the data is expected to decrease. The fraction of mocks with 1σ error bar in the range [0.8,1.2] is expected to decrease with the number of redshift bins. The trends for both the MLE and MCMC are consistent with this expectation. PL does not show any clear changes mainly because the mean and error of PL can be defined within any range, which is taken to be [0.8,1.2] here. The mean of the best-fitting $\langle \bar{\alpha} \rangle$ is essentially unchanged, but the spread $\mathrm{std}(\bar{\alpha})$ increases mildly by 12 per cent, 6 per cent, and 10 per cent for MLE, PL, and MCMC, respectively. On the other hand, the error derived from the each realization show much larger increase. We find that $\langle \sigma_\alpha \rangle$ increases by 21 per cent, 35 per cent, and 33 per cent for MLE, PL, and MCMC, respectively. We also note that $\langle \sigma_\alpha \rangle / \mathrm{std}(\bar{\alpha})$ is closest to 1 among these fitting methods.

5.3 Adding cross-correlations

So far we considered only the autocorrelation in redshift bin w_{ii}. Here, we test the gain on the constraint in α when the cross-correlations among different redshift bins are included.

In Fig. 11, the auto and cross-correlation functions are plotted. We see that only the cross-correlation between the adjacent redshift bins show any BAO signal, while the redshift bins that are further apart have little signal in their cross-correlation. Thus, the BAO constraint improvement if any, can only come from the cross-correlation of adjacent redshift bins.

In Fig. 12, the χ^2 fit using only the auto correlation and when the cross-correlations are included are compared. In the fit, for each of the w_{ij}, a new set of parameters B and A_a are introduced. Hence for our case, there are altogether $10 \times 4 + 1 = 41$ free parameters. We find that the gain in the BAO constraint when the cross redshift bin correlations are included is very limited. In this estimate, Gaussian covariance is used. As it is noiseless, there is no problem of covariance matrix inversion. In practice, when the covariance is estimated from the mocks, we need to reduce the dimension of the covariance matrix. Of course, almost all the improvement on the α constraint is expect to come from the adjacent redshift bins (although this is not shown explicitly here), so in any real analysis, we should include only the nearest redshift bins. Still the dimension of the covariance matrix is increased from $4N_\theta$ to $7N_\theta$, where N_θ is the number of angular bin for each redshift, when the cross-correlations between adjacent redshift bins are included. Thus, we recommend that only the autocorrelations be used.

5.4 Testing the order of the polynomial

As we mentioned after equation (6), we test the number of free parameters A_i in the template. In Table 5, we show the fit results for N_p = 1, 2, and 3 in equation (6). From the limited data set, it is hard to draw a solid conclusion. Overall, there is no clear systematic trend with N_p for these three methods, thus it shows that the fiducial order of polynomial adopted although somewhat arbitrary, it does

not lead to systematics bias in the analysis. Assuringly, MLE with N_p = 3, which is the workhorse model adopted in Abbott et al. (2017b), is the method that performs best in terms of being largely unbiased and recovering close to 68 per cent of the true answer (i.e. the Gaussian expectation).

6 COVARIANCE

In this section, we consider the issues of covariance in more detail. We first study how the Gaussian covariance impacts the BAO fit, and then investigate how to improve the covariance derived from a set of mock catalogues. One of the potential issues that arises in the survey analysis pipeline is that sample properties such as the $n(z)$ and galaxy bias could be different from those used to create the mocks. We show that the correction due to these property changes can be mitigated with the help of the Gaussian covariance. We also investigate expanding the covariance matrix and precision matrix using the eigenmodes from some proxy covariance matrix. We demonstrate that this approach can substantially reduce the influence of the noise because the number of free parameters are significantly reduced, and that it can effectively mimic the effect of small changes in the sample properties.

6.1 BAO fit with the Gaussian covariance

In Fig. 13, we show the covariance matrix obtained from the mock catalogue and the Gaussian theory. We arrange the data vector in ascending order in redshift. To see the difference more clearly, we plot four different rows of the covariance matrix in Fig. 14. We have shown the results for two samples (proxy and target, see below). The symbols and lines show the mock and the Gaussian covariances, respectively. From Figs 13 and 14, it is clear that the Gaussian covariance captures most of the features well. However, the Gaussian covariance exhibits only correlation between the neighbouring redshift bins, while the mock covariances show correlation beyond the neighbouring redshift bins.

We summarize the BAO fit results using the mock and the Gaussian covariance in Table 6. The distribution of $\bar{\alpha}$ is similar to that from the mock and $\mathrm{std}(\bar{\alpha})$ is only systematically larger than that from the mock by a couple of percent. However, $\langle \sigma_\alpha \rangle$ from the Gaussian covariance is only about 80 per cent of $\langle \sigma_\alpha \rangle$ derived from the mock.
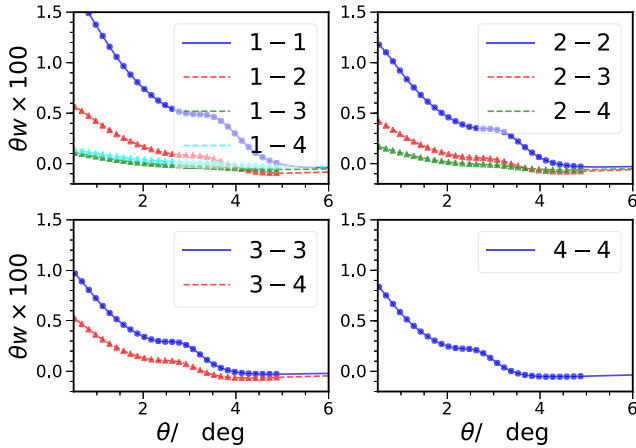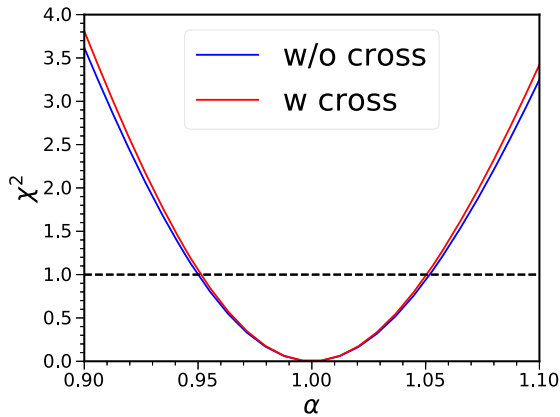
In Fig. 15, we plot the $\chi^2_{\min}$ per degree of freedom for the BAO fit using different prescriptions for the covariance. The histograms are obtained by fitting to 1800 mock data vectors, and they only differ in the covariance used in the fit. The mock covariance by construction gives the $\chi^2_{\min}$ per degree of freedom ~ 1, while we find that the Gaussian covariance (with the correct sample properties) gives a higher value, ~ 1.4. We will use the $\chi^2_{\min}$ per degree of freedom as the metric to decide which prescription of the covariance gives a better approximation to the correct mock covariance.

6.2 Correcting sample variation using Gaussian covariance

As mocks take time to produce, they are often created using some expected data properties. If these differ from the actual data properties, the mocks do not perfectly match the data. Here, we investigate correcting these changes in the mock covariance matrix using the Gaussian covariance. These sample changes can result from variation in the bias parameter and number density of the samples. To be concrete, let us call the mock that we created using the expected

Table 4. The BAO fit obtained using varying number of redshift bins: four, three, and only two bins. The results from MLE, PL and MCMC are shown.

	Bins 1–4			Bins 2–4			Bins 3 and 4		
	$\langle\bar{\alpha}\rangle \pm \mathrm{std}(\bar{\alpha})$	Fraction $\bar{\alpha} \pm \sigma_\alpha$ in [0.8,1.2]	$\langle\sigma_\alpha\rangle$ /std$(\bar{\alpha})$	$\langle\bar{\alpha}\rangle \pm \mathrm{std}(\bar{\alpha})$	Fraction $\bar{\alpha} \pm \sigma_\alpha$ in [0.8,1.2]	$\langle\sigma_\alpha\rangle$ /std$(\bar{\alpha})$	$\langle\bar{\alpha}\rangle \pm \mathrm{std}(\bar{\alpha})$	Fraction $\bar{\alpha} \pm \sigma_\alpha$ in [0.8,1.2]	$\langle\sigma_\alpha\rangle$ /std$(\bar{\alpha})$
MLE	1.001 ± 0.052	0.91	1.02	1.001 ± 0.053	0.87	1.06	0.998 ± 0.058	0.79	1.11
PL	1.004 ± 0.049	0.99	1.27	1.005 ± 0.051	0.99	1.32	1.003 ± 0.052	0.99	1.51
MCMC	1.007 ± 0.049	0.84	1.17	1.006 ± 0.051	0.75	1.22	1.004 ± 0.054	0.63	1.41

Figure 11. The auto (solid, circles, blue) and cross-correlation (dashed, triangles, other colours) between different redshift bins.

Figure 12. The χ^2 for the constraint on α for only using the autocorrelation (blue) and including also the cross-correlations (red).

properties the 'proxy mocks', and the ones with the final correct properties the 'target mocks'.

As the Gaussian covariance works reasonably well, here we consider correcting the changes in the mock covariance using the Gaussian one as

$$C_{\mathrm{Corrected}} = C_{\mathrm{Proxy}} + C_{\mathrm{TargetGauss}} - C_{\mathrm{ProxyGauss}}. \quad (30)$$

i.e. we only correct the Gaussian part in the mock covariance, assuming that the non-Gaussian part is a small correction.

In Fig. 14, we have plotted the $C_{\mathrm{Corrected}}$ and we find that the long-range correlation beyond the neighbouring bins are preserved. In Fig. 15, we have also plotted the χ^2_{min} per degree of freedom for the BAO fit using the proxy mock covariance with the Gaussian correction. It results in values smaller than those obtained using the proxy mock covariance or the target Gaussian covariance, and

hence it signals that the corrected covariance is closer to the true one.

Overall, by combining the proxy mock covariance with the Gaussian correction, we get better agreement with the target mock covariance than using solely the proxy or the Gaussian covariance. This composite approach is expected to work when the variation of the sample is small, e.g. the variation in $n(z)$ is small, and the non-Gaussian contribution is weak because we have not corrected the part due to the non-Gaussian covariance. It offers a means to correct the property changes in covariance matrix without having to re-run the mocks.

6.3 Eigenmode expansion of the covariance matrix

The covariance determined from mocks is susceptible to statistical noise. In reality, we are often not interested in the covariance per se, but its inverse, the precision matrix. For example, in the likelihood, what we really need is the precision matrix. In this case, the problem is even more acute. The noise causes bias in the precision matrix (Anderson 2003; Hartlap et al. 2007; Dodelson & Schneider 2013; Percival et al. 2014). Given a covariance matrix of size $p \times p$ determined from N_{mock} mock catalogues, for it to be invertible, we must have $N_{\mathrm{mock}} > p$. Even if it is invertible, the precision matrix so determined Ψ_{mock} is biased relative to the true precision matrix Ψ_{unbiased} as (Anderson 2003; Hartlap et al. 2007)
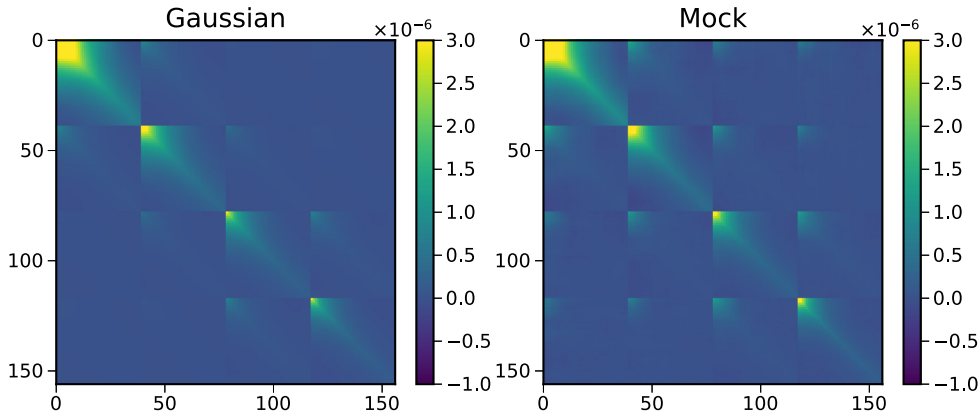
$$\Psi_{\mathrm{unbiased}} = \frac{N_{\mathrm{mock}} - p - 2}{N_{\mathrm{mock}} - 1} \Psi_{\mathrm{mock}}. \quad (31)$$

The correction factor has been checked to work very well for the case of power spectrum (Blot et al. 2016) and bispectrum (Chan & Blot 2017). None the less, although the bias can be modelled and removed, the fluctuations due to noise is unavoidable (Sellentin & Heavens 2016). Thus, it is highly desirable to reduce the number of free parameters to be determined in the covariance matrix.
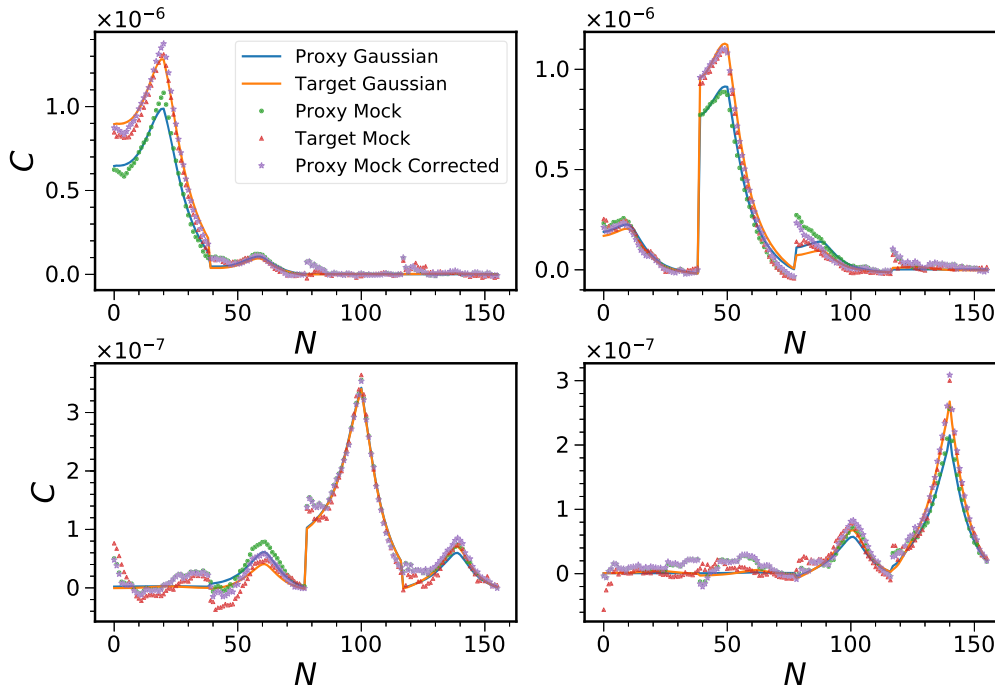
There are methods that have been proposed to combine mocks with theory covariance to reduce the impact of the noise fluctuations (Pope & Szapudi 2008; Taylor & Joachimi 2014; Paz & Sánchez 2015; Pearson & Samushia 2016; Friedrich & Eifler 2017). Here, we consider the expansion of the covariance and precision matrices using the eigenmodes or the principal components of a given covariance matrix. The effects due to noise can be mitigated because the number of free parameters is substantially reduced when the eigenmodes are given. The study of the eigenmode of the covariance matrix is known as principal component analysis (PCA) and it is widely used in many different fields. PCA identifies the most rapidly varying direction in the data space with many variables, and hence find a more effective way to describe the data. In the cosomological covariance context, it has been used in Scoccimarro (2000), Harnois-Déraps & Pen (2012), and Mohammed, Seljak & Vlah (2017).

Table 5. The BAO fit with different number N_p (=1, 2, 3) in the template equation (6). The results for MLE, PL, and MCMC are compared.

	$N_p = 1$		$N_p = 2$		$N_p = 3$	
	$\langle\bar{\alpha}\rangle \pm \text{std}(\bar{\alpha})$	Fraction with $\bar{\alpha} \pm \sigma_\alpha$ enclosing $\langle\bar{\alpha}\rangle$	$\langle\bar{\alpha}\rangle \pm \text{std}(\bar{\alpha})$	Fraction with $\bar{\alpha} \pm \sigma_\alpha$ enclosing $\langle\bar{\alpha}\rangle$	$\langle\bar{\alpha}\rangle \pm \text{std}(\bar{\alpha})$	Fraction with $\bar{\alpha} \pm \sigma_\alpha$ enclosing $\langle\bar{\alpha}\rangle$
MLE	0.996 ± 0.047	0.72	0.995 ± 0.051	0.68	1.001 ± 0.052	0.69
PL	0.993 ± 0.046	0.77	1.000 ± 0.050	0.79	1.004 ± 0.049	0.77
MCMC	1.005 ± 0.046	0.75	0.994 ± 0.050	0.71	1.007 ± 0.049	0.74

Figure 13. The Gaussian covariance (left-hand panel) and the mock covariance (right-hand panel).

Figure 14. Four rows of the covariance matrix (the row number corresponds to the peak position) are shown. The results for Gaussian covariance (blue curve for the proxy and orange one for the target) and the mock covariance (green circles for the proxy and red triangles for the target) are compared. The composite covariance (violet stars) obtained by combining the proxy mock results with the Gaussian correction is also displayed.

Table 6. The BAO fit results obtained using the mock and the Gaussian covariance. The results for MLE, PL, and MCMC are shown.

	Mock covariance			Gaussian covariance		
	$\langle \bar{\alpha} \rangle \pm \text{std}(\bar{\alpha})$	$\langle \sigma_\alpha \rangle / \text{std}(\bar{\alpha})$	Fraction with 1σ enclosing $\langle \bar{\alpha} \rangle$	$\langle \bar{\alpha} \rangle \pm \text{std}(\bar{\alpha})$	$\langle \sigma_\alpha \rangle / \text{std}(\bar{\alpha})$	Fraction with 1σ enclosing $\langle \bar{\alpha} \rangle$
MLE	1.001 ± 0.052	1.02	0.69	1.002 ± 0.054	0.84	0.60
PL	1.004 ± 0.049	1.27	0.77	1.004 ± 0.052	1.04	0.68
MCMC	1.007 ± 0.049	1.17	0.74	1.007 ± 0.051	0.96	0.64

Figure 15. The distributions of the $\chi^2_{\min}$ per degree of freedom for the BAO fit using various covariances. The results obtained using the target mock covariance (blue, dashed), Gaussian covariance (red, dotted), proxy mock covariance (green, dotted–dashed), the proxy mock corrected with the Gaussian covariance (cyan, solid), the direct eigenmode expansion on the proxy mocks (magenta, solid), and the eigenmode expansion combining the Gaussian covariance eigenmodes with the proxy mock ones (yellow, solid). The covariance yielding the best approximation to the mock covariance is expected to give $\chi^2_{\min}$ per degree of freedom closest to that obtained using the mock results.

Suppose that the eigenmodes $\boldsymbol{v}^{(i)}$ and the eigenvalues $\lambda^{(i)}$ of a covariance matrix C^{proxy} are given by

$$C^{\text{proxy}} \boldsymbol{v}^{(i)} = \lambda^{(i)} \boldsymbol{v}^{(i)}. \qquad (32)$$

We can express the matrix C^{proxy} in terms of its eigenmodes and eigenvalues as

$$C^{\text{proxy}}_{ab} = \sum_{i=1}^{p} \lambda^{(i)} \boldsymbol{v}^{(i)}_a \boldsymbol{v}^{(i)}_b. \qquad (33)$$

The precision matrix Ψ^{proxy} can be written as

$$\Psi^{\text{proxy}}_{ab} = \sum_{i=1}^{p} \frac{1}{\lambda^{(i)}} \boldsymbol{v}^{(i)}_a \boldsymbol{v}^{(i)}_b. \qquad (34)$$

We can expand a covariance matrix C^{target} in terms of these eigenmodes $\boldsymbol{v}^{(i)}$, treating $\lambda^{(i)}$ as the fitting parameters. The mode parameters can be extracted as

$$C^{\text{target}} \boldsymbol{v}^{(i)} \approx \kappa^{(i)} \boldsymbol{v}^{(i)}. \qquad (35)$$

Here, we assume that the eigenmode $\boldsymbol{v}^{(i)}$ approximates that of C^{target} well. Then we can express C^{target} as

$$C^{\text{target}}_{ab} \approx \sum_{i=1}^{m} \kappa^{(i)} \boldsymbol{v}^{(i)}_a \boldsymbol{v}^{(i)}_b. \qquad (36)$$

Note that we use the first m eigenmodes with $\lambda^{(i)}$ ranked in descending order to approximate the covariance matrix. For the precision

Figure 16. The eigenvalues of the covariance matrix measured from the mocks (red triangles) and the Gaussian covariance (blue circles).

matrix, we have

$$\Psi^{\text{target}}_{ab} \approx \sum_{i=1}^{m} \frac{1}{\kappa^{(i)}} \boldsymbol{v}^{(i)}_a \boldsymbol{v}^{(i)}_b. \qquad (37)$$
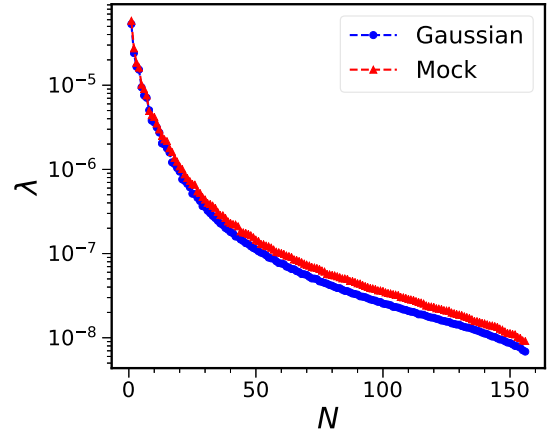
For the covariance matrix, it is clear that we can safely ignore the modes with small eignevalues. However, for the inverse, the modes with small eigenvalues in fact contribute more.

The full covariance matrix of dimension p has $p(p + 1)/2$ independent elements, while fitting using a given basis of eigenmodes has only p free parameters. Thus, this basis cannot allow for all variations in the covariance and they cannot fit a general symmetric matrix. The success of this method hinges on how well the given eigenmodes approximate those of the target covariance matrix. We can obtain these eigenmodes using theory or approximate methods, e.g. Second-order Lagrangian Perturbation Theory (2LPT) mock in Scoccimarro (2000). In the following, we use both the Gaussian theory covariance and the covariance derived from mocks with slightly different sample properties. We consider two samples (proxy and target) which differ slightly in the $n(z)$ and bias. The bias of these two samples differ by between a few per cent to 10 per cent, while the mean of the two photo-z distributions differs by up to 7 per cent. There are 1800 realizations for both samples.

In Fig. 16, we first compare the eigenvalues measured from the mock and Gaussian covariances (for the target sample). We find that for the largest eigenvalues the results from the Gaussian covariance agree with those from the mocks well, while the mock ones are larger than the Gaussian ones for the relatively small ones ($\lambda \lesssim 2 \times 10^{-6}$). We can also study the overlap between the eigenmodes from the mock covariance with those from the Gaussian one. In Fig. 17, we show the dot product between the jth eigenmode from the mock covariance $\boldsymbol{v}^{\text{mock}}_j$ and the ith eigenmode from the Gaussian covariance, $\boldsymbol{v}^{\text{Gauss}}_i$. In this plot, for each curve j is fixed, while i runs over all the modes. When the overlap is perfect, the dot prod-
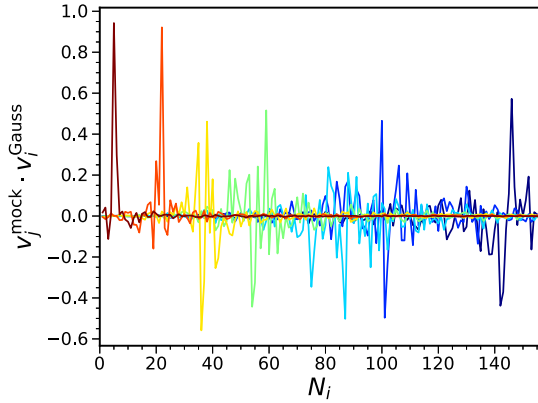
Figure 17. The dot product between the eigenmodes from the mock covariance and those from the Gaussian covariance. The eigenmodes are arranged in the same order as the eigenvalues shown in Fig. 16. For each curve, the eigenmode from the mock covariance $\boldsymbol{v}_j^{\mathrm{mock}}$ is fixed, while all the modes from the Gaussian covariance are run over.

uct is 1 (or −1). When eigenvalue is large, the overlap between the eigenmodes are large and well peaked. For relatively smaller eigenvalue modes, they do not match each other well. This shows that the eigenmode from the Gaussian covariance is a good approximation to that of the mock covariance only for the ones with large eigenvalues. Since the modes with the largest eigenvalues are close to Gaussian covariance prediction, we can call them the Gaussian modes. The fact that eigenvalues of the non-Gaussian modes from the Gaussian covariance are smaller than those from the mock covariance suggests that the Gaussian covariance underestimates the importance of the non-Gaussian modes.

We now check how important the non-Gaussian modes are for the covariance matrix and the precision matrix. In Fig. 18, we compare the covariance matrix and the precision matrix obtained using the eigenmode expansion by keeping terms with eigenvalues larger than the cut-off values 10^{-6}, 10^{-7}, and 2×10^{-8}, respectively. They corresponding to the largest 19, 53, and 112 eigenvalues. Here and for the rest of the analysis, it is useful to note that the size of the covariance matrix is 156×156. In this exercise, we have used the Gaussian covariance. It is clear that by keeping only the terms with large eigenvalues, e.g. larger than 10^{-6}, we get a reasonably good approximation to the original covariance matrix already. This is evident from equation (36). On the other hand, we find that keeping only terms with large eigenvalues results in a poor approximation to the precision matrix. This is not surprising because as we mentioned, the small eigenvalue modes in fact contribute more to the inverse, thus ignoring them yields an unsatisfactory approximation. Henceforth, we will keep all the modes.

Because the eigenmode expansion with given basis modes has far fewer free parameters than the direct measurement, we now test how effective the eigenmode expansion is in reducing the impact of noise on the measurement of the covariance and the precision matrix. In Fig. 19, we show two rows of the covariance matrices obtained using different prescriptions. The results from the mocks and the Gaussian theory are displayed. The eigenmodes are obtained from the 1800 proxy mocks. Using these proxy eigenmodes, we fit the 'eigenvalue' parameters from the target mock covariance estimated with N_{mock} target mocks. We have shown the results using $N_{\mathrm{mock}} = 200$, 400, and 1800 (all the mocks) realizations. As expected, the direct measurement results slowly become less noisy as the number

of realizations used increases. On the other hand, the eigenmode expansion results are far less noisy than the direct measurement ones. In fact, there are no marked variations among the different number of realizations shown. However, even for $N_{\mathrm{mock}} = 1800$, there are small deviations between the eigenmode expansion result and the direct measurement one with $N_{\mathrm{mock}} = 1800$. The systematic bias in the eigenmode expansion comes from the fact that the basis modes used are not the exact ones.

We present the corresponding results for the precision matrix in Fig. 20. For Ψ from direct measurement, we have applied the bias correction factor equation (31). Similar to the case shown in Fig. 19, the eigenmodes are estimated from 1800 proxy mocks, and the fitting parameters are extracted using the covariance estimated from N_{mock} target mocks. While there are some improvements in the noise suppression using the eigenmode expansion for the covariance, the gain for the precision matrix is impressive. This is especially true when N_{mock} is not substantially larger than p. For example, for the case of $N_{\mathrm{mock}} = 200$, the precision matrix from direct measurement is very noisy, while the results obtained from the eigenmode expansion is already close to the asymptotic results (the $N_{\mathrm{mock}} = 1800$ direct measurement is taken to be the asymptotic one). However, similar to the case of the covariance matrix, although the results converge quickly, they are slightly biased compared to the direct measurement with $N_{\mathrm{mock}} = 1800$. This is again attributed to the proxy basis modes used being not exact. In Fig. 15, we have plotted the results obtained using the eigenmode expansion. We have used the results determined using $N_{\mathrm{mock}} = 200$. The resultant χ^2_{min} per degree of freedom is one of the closest to the true mock covariance results among all the prescriptions shown. The other one is also eigenmode expansion based, to be commented on shortly.

We have tried even lower of number of realizations, e.g. 30. In this case, the covariance from the direct measurement is not invertible, but the eigenmode expansion still gives a valid result with a small amount of noise only. Thus, the eigenmode expansion is much more robust to noise than the direct measurement.

If the basis modes are given by the Gaussian theory, then the convergent result (i.e. when sufficient number of N_{mock} are used) gets close to the Gaussian one, and so it does not yield a good approximation to the true covariance. Alternatively, we can combine the eigenmodes from the Gaussian covariance with those from the proxy mocks to form a composite covariance. The idea is that because the large eigenvalue modes are close to the Gaussian covariance results, we can use the Gaussian theory eigenmodes for them, and this has the advantage that the sample properties can be taken into account in theory easily. For the non-Gaussian modes (with small eigenvalues), we use the basis modes determined from the proxy mock covariance. Using the composite basis, the eigenvalue parameters are fitted to a covariance estimated from a relatively small set of target mocks as in the previous method. This composite approach performs similar to the one using only the basis modes from the proxy mocks, as can be seen from Fig. 15. If the number of proxy mocks available is not large, this hybrid approach is preferred because the influence of noise on the estimate of the proxy eigenmodes is reduced.

In summary, we find that using the eigenmode expansion, the impact of the noise on the measurement of the covariance and especially the precision matrix is significantly reduced. Thus, this method enables us to substantially reduce the number of mock catalogues required. The success of the methods relies on how well the given proxy basis modes approximate the true ones. We find that although the expansion converges quickly, the results are slightly biased. This is due to the basis modes being not exact. One of
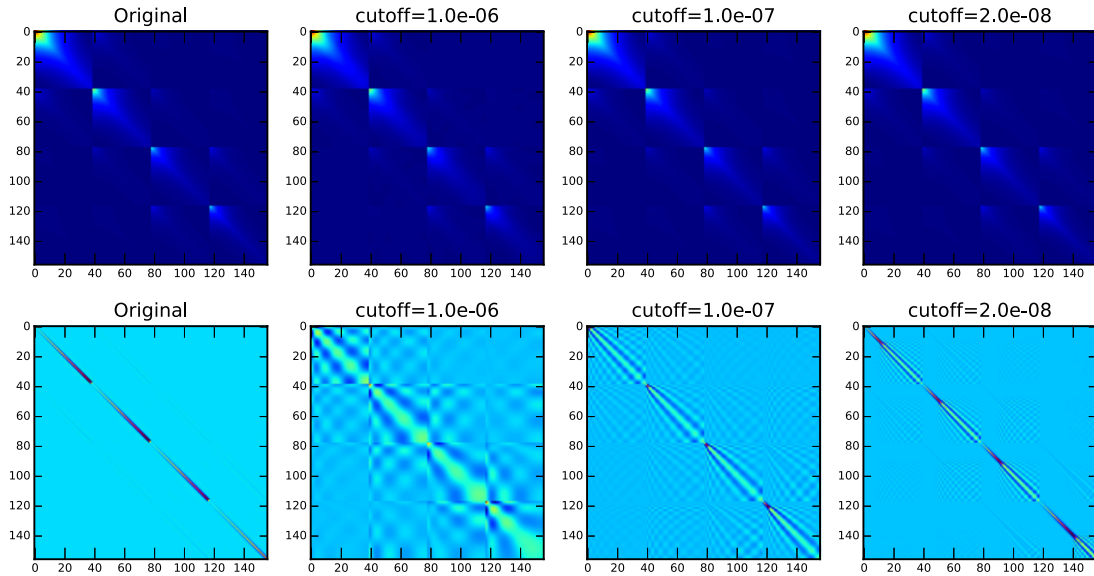
Figure 18. The covariance matrix (upper panels) and the precision matrix (lower panels) are shown. The leftmost ones show the original ones, and the rest are obtained using the eigenmode expansion by keeping the modes with eigenvalue larger than cutoff values 10^{-6}, 10^{-7}, and 2×10^{-8}, respectively.
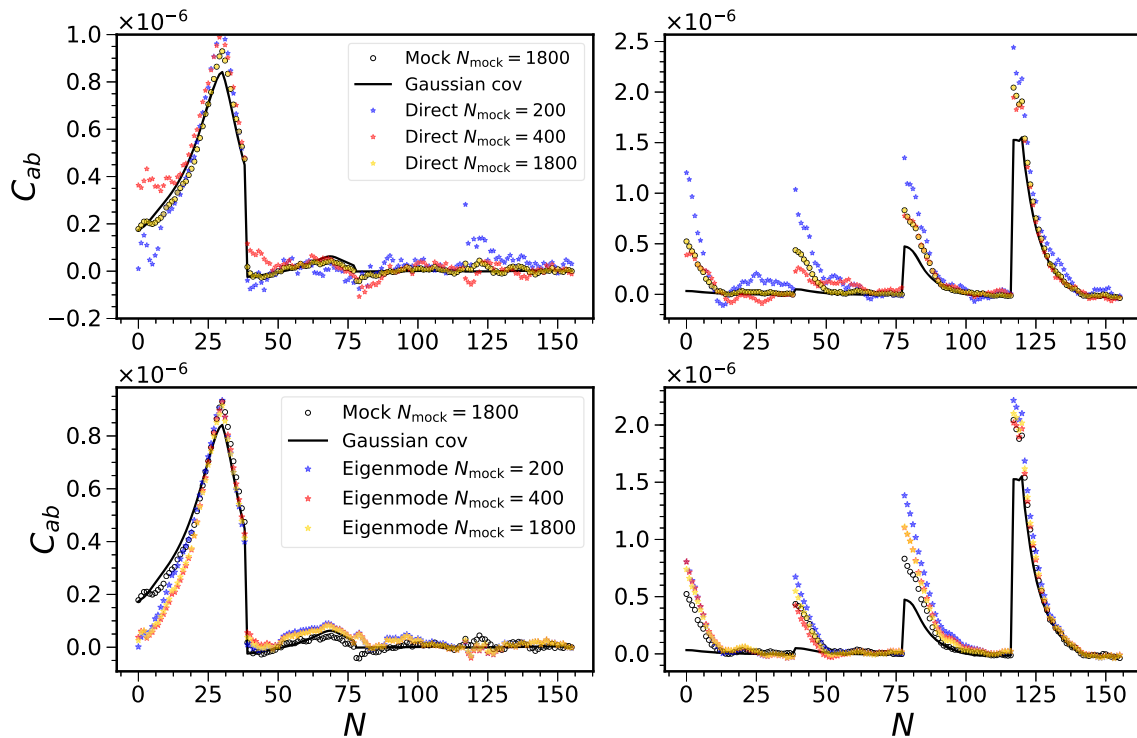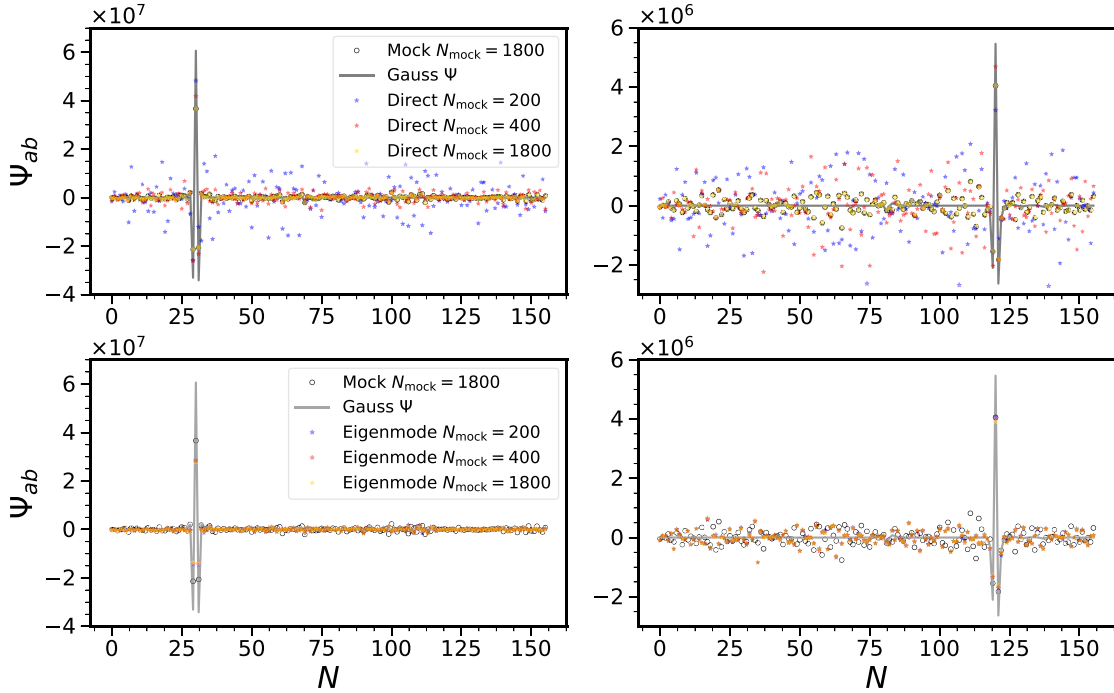

Figure 19. Two rows of the covariance matrix (left and right) obtained from the mock (direct measurement with $N_{mock} = 1800$, circles, in the upper panels coincide with the yellow stars), Gaussian covariance (solid black lines) are shown. In the upper panels, we show the results from direct measurement (stars) using $N_{mock} = 200$ (blue), 400 (red), and 1800 (yellow). We display the corresponding results from the eigenmode expansion in the lower panels.

the applications is to correct for sample changes. In this case, as we have demonstrated, the original mock samples can provide the required proxy eigenmodes. We have demonstrated that this method performs even better than the correction using Gaussian covariance proposed in Section 6.2.

7 CONCLUSIONS

Measurement of the BAO scale in the distribution of galaxies has been recognized as one of the most important current cosmological probes. DES already delivered encouraging BAO measurement using photometric data and future DES data, or surveys such as

Figure 20. Two rows of the precision matrix obtained by inverting the covariance matrix shown in Fig. 19 are displayed. The symbols are the same as those in Fig. 19. The upper panels are obtained from direct measurements and the lower ones are from eigenmode expansion.

LSST, are expected to yield more exciting results. To meet the big improvements in the data, the analysis pipeline must be optimized and theoretical systematics must be under control. In this paper, we study the theoretical systematics and optimization of the BAO detection using the angular correlation function. Although the fiducial setup is for DES Y1 in this paper, the techniques developed here are useful for other large-scale structure survey analysis as well.

To extract the BAO from the data, some estimators are required to produce the best fit and the error bar. We have compared three common fitting methods: MLE, PL, and MCMC. Among these methods, MLE yields the least bias and the error bar derived using $\Delta \chi^2 = 1$ is closest to the Gaussian distribution (Table 1). For MLE, we also find that the std($\bar{\alpha}$) agrees with $\langle \sigma_\alpha \rangle$ well ($\langle \sigma_\alpha \rangle$/std($\bar{\alpha}$) columns in Table 4). An advantage of MLE (and PL) is that it takes far less time than MCMC.

Because MLE is completely determined by the likelihood, it is more efficient in picking up small features such as those arising from mismatch between the template and the data. We regarded the mock catalogues whose 1σ interval of the best-fitting α fall outside the interval [0.8,1.2] as non-detections. Those extreme mocks are poorly fitted by our existing methodology and we have demonstrated that they cause bias in distribution of the best-fitting parameter in Appendix A.

We studied when there is mismatch between the template and the data either due to incorrect fiducial cosmology (Table 2) or the photo-z error (Table 3) whether the fitting pipeline can recover the mismatch correctly. This is especially relevant for DES Y1 as the mocks are matched to the MICE cosmology, which is quite different from the currently accepted *Planck* cosmology. Differences in cosmology lead to variation in both the sound horizon and the angular diameter distance, while photo-z error shifts the angular diameter distance only. We find that only the photo-z error leading to shift in the mean of the photo-z distribution causes bias in the

BAO fit, while reasonable variation in the width of the distribution does not produce any significant effect. The shift in BAO obtained from direct fitting is consistent with the estimate based on its effects on the sound horizon and/or the angular diameter distance.

A number of optimizations in the context of the DES Y1 BAO measurement were investigated: the angular bin width, the number of redshift bins, and the effect of the cross-correlation. We find that the optimal binning is $\Delta \theta = 0°.15$ (to $0°.3$), and four redshift bins are sufficient. Adding the cross-correlation does not enhance the BAO constraint substantially.

Sometimes the mocks are created with certain sample properties, but the finalized samples end up slightly different from those envisioned at the beginning. We proposed to correct these sample changes in the covariance matrix using the Gaussian covariance and the eigenmode expansion method. We show that the covariance with Gaussian covariance correction outperforms the pure Gaussian covariance or the proxy mock covariance; however, the eigenmode expansion works even better in incorporating the changes (Fig. 15). The success of the eigenmode expansion relies on how well the proxy eigenvectors approximate those of the target covariance. With the basis vectors given, as the number of free parameters to be determined is significantly reduced [p parameters versus $p(p + 1)/2$ for direct measurement] and we demonstrate that the effect of noise is substantially mitigated relative to the direct measurements. In this case, it does not rely on the Gaussian covariance, but a small number of target mocks are required. Thus we expect this method to work even when the covariance is highly non-Gaussian.

As this work was finished, the DES Y3 data became available and science ready. Although the depth of the Y3 data is similar to that of Y1, its area is roughly three times larger. Hence, we expect that the constraint on the BAO angular scale will be tightened by a factor of $\sim 1/\sqrt{3}$. The analysis pipeline developed here, and the lessons learned, can be easily extended to the Y3 data. Hence, we look forward to that analysis.

ACKNOWLEDGEMENTS

REFERENCES

Abbott T. M. C. et al., 2017a, preprint (arXiv:1708.01530)
Abbott T. M. C. et al., 2017b, preprint (arXiv:1712.06209)
Ade P. A. R. et al., 2016, A&A, 594, A13
Alam S. et al., 2017, MNRAS, 470, 2617
Alonso D., 2012, preprint (arXiv:1210.1833)
Anderson T. W., 2003, An Introduction to Multivariate Statistical Analysis, 3rd edn. Wiley-Interscience, New York
Anderson L. et al., 2014, MNRAS, 441, 24
Ata M. et al., 2017, MNRAS, 473, 4773
Aubourg É. et al., 2015, Phys. Rev., D92, 123516

Avila S., Murray S. G., Knebe A., Power C., Robotham A. S. G., Garcia-Bellido J., 2015, MNRAS, 450, 1856
Avila S. et al., 2017, MNRAS, 479, 94
Bautista J. E. et al., 2017, A&A, 603, A12
Beutler F. et al., 2011, MNRAS, 416, 3017
Blot L., Corasaniti P. S., Amendola L., Kitching T. D., 2016, MNRAS, 458, 4462
Bond J. R., Efstathiou G., 1984, ApJ, 285, L45
Bond J. R., Efstathiou G., 1987, MNRAS, 226, 655
Camacho A. et al., 2018, preprint (arXiv:1807.10163)
Carnero A., Sánchez E., Crocce M., Cabré A., Gaztañaga E., 2012, MNRAS, 419, 1689
Chan K. C., Blot L., 2017, Phys. Rev., D96, 023528
Chan K. C., Moradinezhad Dizgah A., Noreña J., 2018, Phys. Rev., D97, 043532
Cohn J. D., 2006, New Astron., 11, 226
Cole S., Fisher K. B., Weinberg D. H., 1994, MNRAS, 267, 785
Cole S. et al., 2005, MNRAS, 362, 505
Cowan G., 1998, Statistical Data Analysis. Oxford Univ. Press, Oxford
Crocce M., Scoccimarro R., 2008, Phys. Rev. D, 77, 023533
Crocce M., Cabré A., Gaztañaga E., 2011, MNRAS, 414, 329
Crocce M. et al., 2017, preprint (arXiv:1712.06211)
de Simoni F. et al., 2013, MNRAS, 435, 3017
Dodelson S., 2003, Modern Cosmology, Academic Press, San Diego
Dodelson S., Schneider M. D., 2013, Phys. Rev. D, 88, 063537
Drlica-Wagner A. et al., 2017, ApJS, 235, 33
Eisenstein D. J., Hu W., 1998, ApJ, 496, 605
Eisenstein D. J. et al., 2005, ApJ, 633, 560
Eisenstein D. J., Seo H.-J., White M., 2007, ApJ, 664, 660
Estrada J., Sefusatti E., Frieman J. A., 2009, ApJ, 692, 265
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Fosalba P., Crocce M., Gaztañaga E., Castander F. J., 2015, MNRAS, 448, 2987
Friedrich O., Eifler T., 2017, MNRAS, 473, 4150
Gaztanaga E., Cabre A., Hui L., 2009, MNRAS, 399, 1663
Gaztañaga E. et al., 2018, preprint (arXiv:e-prints)
Hahn C., Beutler F., Sinha M., Berlind A., Ho S., Hogg D. W., 2018, preprint (arXiv:1803.06348)
Hamilton A. J. S., 1992, ApJ, 385, L5
Hamilton A. J. S., Rimes C. D., Scoccimarro R., 2006, MNRAS, 371, 1188
Harnois-Déraps J., Pen U.-L., 2012, MNRAS, 423, 2288
Hartlap J., Simon P., Schneider P., 2007, A&A, 464, 399
Heavens A., 2009, preprint (arXiv:0906.0664)
Hinshaw G. et al., 2013, ApJS, 208, 19
Hogg D. W., Bovy J., Lang D., 2010, preprint (arXiv:1008.4686)
Hu W., Sugiyama N., 1996, ApJ, 471, 542
Hu W., Sugiyama N., Silk J., 1997, Nature, 386, 37
Hütsi G., 2010, MNRAS, 401, 2477
Kaiser N., 1987, MNRAS, 227, 1
Kazin E. A. et al., 2014, MNRAS, 441, 3524
Koo D. C., 1985, AJ, 90, 418
Krause E. et al., 2017, Phys. Rev. D, preprint (arXiv:1706.09359)
Lampton M., Margon B., Bowyer S., 1976, ApJ, 208, 177
Landy S. D., Szalay A. S., 1993, ApJ, 412, 64
Lewis A., Challinor A., Lasenby A., 2000, ApJ., 538, 473
Li Y., Hu W., Takada M., 2014, Phys. Rev., D89, 083519
Linder E. V., 2005, Phys. Rev. D, 72, 043529
LSST Science Collaboration et al., 2009, preprint (arXiv:0912.0201)
Meiksin A., White M., Peacock J. A., 1999, MNRAS, 304, 851
Mohammed I., Seljak U., Vlah Z., 2017, MNRAS, 466, 780
Padmanabhan N., White M., 2009, Phys. Rev. D, 80, 063508
Padmanabhan N. et al., 2007, MNRAS, 378, 852
Paz D. J., Sánchez A. G., 2015, MNRAS, 454, 4326
Pearson D. W., Samushia L., 2016, MNRAS, 457, 993
Peebles P. J. E., 1980, The Large-Scale Structure of the Universe. Princeton Univ. Press, Princeton, NJ
Peebles P. J. E., Groth E. J., 1975, ApJ, 196, 1

Peebles P. J. E., Yu J. T., 1970, ApJ, 162, 815
Percival W. J. et al., 2010, MNRAS, 401, 2148
Percival W. J. et al., 2014, MNRAS, 439, 2531
Pope A. C., Szapudi I., 2008, MNRAS, 389, 766
Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 2007, Numerical Recipes: The Art of Scientific Computing, 3rd edn. Cambridge Unive. Press, New York
Ross A. J., Percival W. J., Crocce M., Cabré A., Gaztañaga E., 2011, MNRAS, 415, 2193
Ross A. J., Samushia L., Howlett C., Percival W. J., Burden A., Manera M., 2015, MNRAS, 449, 835
Ross A. J. et al., 2017a, MNRAS, 464, 1168
Ross A. J. et al., 2017b, MNRAS, 472, 4456
Salazar-Albornoz S. et al., 2017, MNRAS, 468, 2938
Sanchez A. G., Baugh C. M., Angulo R., 2008, MNRAS, 390, 1470
Scoccimarro R., 2000, ApJ, 544, 597
Sellentin E., Heavens A. F., 2016, MNRAS, 456, L132
Seo H.-J., Eisenstein D. J., 2007, ApJ, 665, 14
Seo H.-J. et al., 2012, ApJ, 761, 13
Smith R. E., Scoccimarro R., Sheth R. K., 2008, Phys. Rev., D77, 043525
Sunyaev R. A., Zeldovich Y. B., 1970, Astrophys. Space Sci., 7, 3
Takada M., Hu W., 2013, Phys. Rev., D87, 123504
Taylor A., Joachimi B., 2014, MNRAS, 442, 2728
Tegmark M., Taylor A., Heavens A., 1997, ApJ, 480, 22
Trotta R., 2017, preprint (arXiv:1701.01467)
Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, Phys. Rep., 530, 87
Zel'dovich Y. B., 1970, A&A, 5, 84

## APPENDIX A: COMPARISON OF THE ESTIMATORS (ALMOST) WITHOUT PRUNING

In the main text, we presented the results with the pruning criterion that the best-fitting $\bar{\alpha}$ with its 1σ error bar must fall within the interval [0.8,1.2]. Although the interval [0.8,1.2] seems reasonable, this choice is to some degree arbitrary. In this appendix, we show the results when a much more generous interval [0.6,1.4] is used. I.e. for MLE and MCMC, we require that the best fit with its 1σ interval falls within [0.6,1.4]. For PL, it is computed in this interval.

In Table A1, we show the results obtained with these three estimators. The results should be compared to Table 1. For MLE and MCMC, the fraction of mocks satisfying this criterion has increased relative to the cases shown in Table 1, and it is almost 100 per cent for MCMC. For PL, it does not change significantly. The extreme mocks not only increase std($\bar{\alpha}$), they also result in appreciable bias in the mean of the MLE and PL estimators. However, the bias of the MCMC estimator is more mild, and $|\langle\bar{\alpha}\rangle - 1| = 0.001$ only. The fractions of time that the 1σ interval enclosing $\langle\bar{\alpha}\rangle$ are only 3 per cent from the Gaussian expectation for both MLE and MCMC. Furthermore, $\langle\sigma_\alpha\rangle/\mathrm{std}(\bar{\alpha}) = 0.92$ for MCMC. From these statistics, we find that when the pruning is weak, the MCMC is the most desirable estimator based on the criteria outlined in Section 3.1.4.

In Fig. A1, we plot the distributions of the best-fitting $\bar{\alpha}$. The

Table A1. The BAO fit using MLE, PL, and MCMC. Only those mocks whose 1σ interval $\bar{\alpha} \pm \sigma_\alpha$ fall within the interval [0.6,1.8] are considered.

|	$\bar{\alpha} \pm \sigma_\alpha$ in [0.6,1.4] (Fraction selected)	$\langle\bar{\alpha}\rangle \pm \mathrm{std}(\bar{\alpha})$	Fraction with $\bar{\alpha} \pm \sigma_\alpha$ enclosing $\langle\bar{\alpha}\rangle$	$\langle\sigma_\alpha\rangle/\mathrm{std}(\bar{\alpha})$
MLE	0.96	0.971 ± 0.074	0.65	0.75
PL	0.98	0.975 ± 0.069	0.83	1.34
MCMC	1.00	0.990 ± 0.078	0.71	0.92
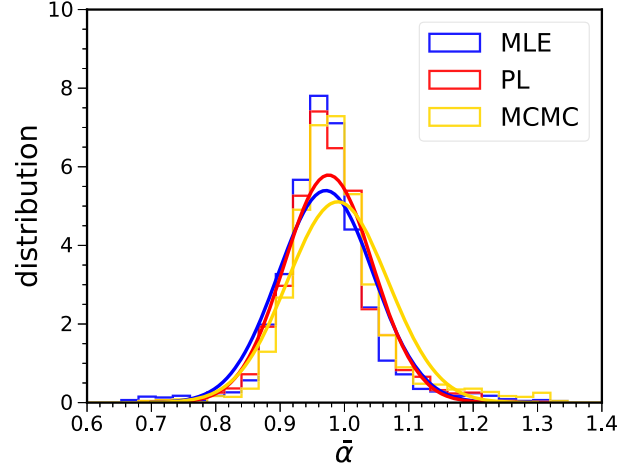
Figure A1. Similar to Fig. 2, except the interval of α is limited to [0.6,1.4] instead of [0.8,1.2].
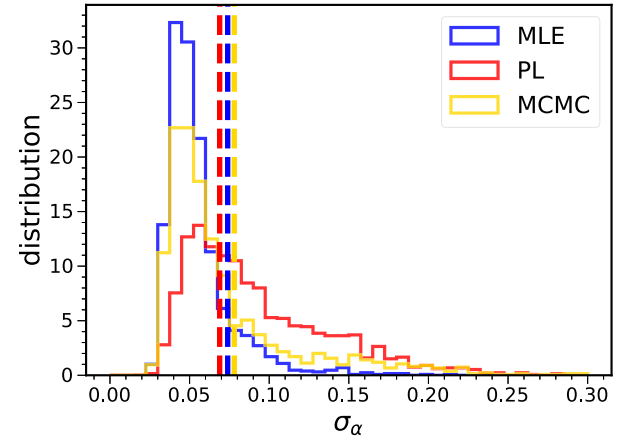
Figure A2. Similar to Fig. 3, except the interval of α is limited to [0.6,1.4] instead of [0.8,1.2].

Gaussian distributions with the same mean and variance as the histograms are also shown. Because of the strong tail in the histogram, the variance is not a good proxy for its width, a Gaussian distribution with the same variance as the histogram does not agree with the histogram. From the Gaussian distribution, we also see that only the MCMC one is close to be unbiased. In contrast the distributions in Fig. 2 are much closer to the Gaussian distributions. The distributions of the error are shown in Fig. A2. In comparison with Fig. A2, the derived errors have much stronger tails when the interval [0.6,1.4] is adopted. In particular, PL and MCMC show appreciable increase in the tail of the distribution. We plot the distributions of d_{norm} in Fig. A3. Besides thicker tails than those in Fig. 4, we also note that the distribution for MCMC is skewed aside and the PL case shows large deviation from the unit Gaussian distribution.

When the allowed interval is enlarged, the mocks with $\bar{\alpha}$ far from $\alpha = 1$ are included. They are often badly fitted by our model and cause strong tails in the distribution of the best-fitting parameters. We show that these strong tails cause significant deviation from the Gaussian distribution. They not only enlarge the derived error bars but also cause bias in the mean although it is relatively mild
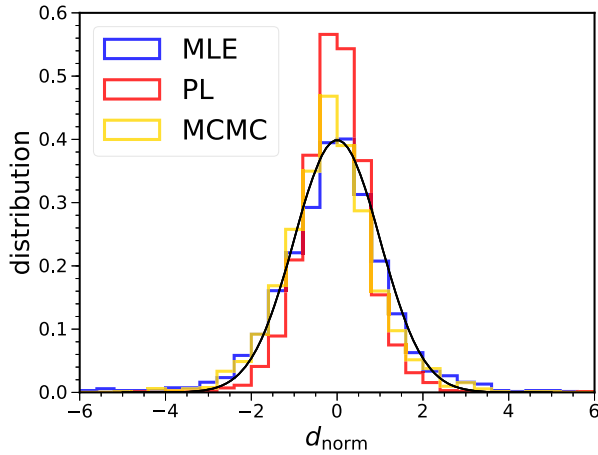
Figure A3. Similar to Fig. 4, except the interval of α is limited to [0.6,1.4] instead of [0.8,1.2].

for the MCMC. Those extreme mocks are poorly handled by our existing methodology, and little BAO information can be extracted from them. Therefore, we choose a smaller interval in the main text to get rid of those cases.

[1]*School of Physics and Astronomy, Sun Yat-Sen University, Guangzhou 510275, China*

[2]*Institut d'Estudis Espacials de Catalunya (IEEC), E-08193 Barcelona, Spain*

[3]*Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain*

[4]*Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA*

[5]*Institute of Cosmology, Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK*

[6]*Instituto de Fisica Teorica UAM/CSIC, Universidad Autonoma de Madrid, E-28049 Madrid, Spain*

[7]*Jodrell Bank Center for Astrophysics, School of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK*

[8]*Department of Physics, Astronomy, University College London, Gower Street, London WC1E 6BT, UK*

[9]*ICTP South American Institute for Fundamental Research Instituto de Física Teórica, Universidade Estadual Paulista, São Paulo, Brazil*

[10]*Laboratório Interinstitucional de e-Astronomia - LIneA, Rua Gal. José Cristino, Rio de Janeiro, RJ – 20921-400, Brazil*

[11]*Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, Casilla 603, La Serena, Chile*

[12]*Department of Physics and Electronics, Rhodes University, PO Box 94, Grahamstown 6140, South Africa*

[13]*Fermi National Accelerator Laboratory, PO Box 500, Batavia, IL 60510, USA*

[14]*CNRS, UMR 7095, Institut d'Astrophysique de Paris, F-75014, Paris, France*

[15]*Institut d'Astrophysique de Paris, Sorbonne Universités, UPMC Univ Paris 06, UMR 7095, F-75014, Paris, France*

[16]*Kavli Institute for Particle Astrophysics, Cosmology, PO Box 2450, Stanford University, Stanford, CA 94305, USA*

[17]*SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA*

[18]*Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ – 20921-400, Brazil*

[19]*Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801, USA*

[20]*National Center for Supercomputing Applications, 1205 West Clark St., Urbana, IL 61801, USA*

[21]*Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, E-08193 Bellaterra (Barcelona), Spain*

[22]*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*

[23]*Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain*

[24]*Department of Astronomy/Steward Observatory, 933 North Cherry Avenue, Tucson, AZ 85721-0065, USA*

[25]*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109, USA*

[26]*Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA*

[27]*Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA*

[28]*Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA*

[29]*Department of Physics, ETH Zurich, Wolfgang-Pauli-Strasse 16, CH-8093 Zurich, Switzerland*

[30]*Department of Physics, The Ohio State University, Columbus, OH 43210, USA*

[31]*Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, D-85748 Garching, Germany*

[32]*Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstr. 1, D-81679 München, Germany*

[33]*Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA*

[34]*Australian Astronomical Observatory, North Ryde, NSW 2113, Australia*

[35]*Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, CP 66318, São Paulo, SP 05314-970, Brazil*

[36]*Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain*

[37]*School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK*

[38]*Instituto de Física Gleb Wataghin, Universidade Estadual de Campinas, 13083-859, Campinas, SP, Brazil*

[39]*Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA*

This paper has been typeset from a TEX/LATEX file prepared by the author.