THE EUROPEAN
PHYSICAL JOURNAL C

Regular Article - Theoretical Physics

Angular correlation function from sample covariance with BOSS and eBOSS LRG

Paula S. Ferreira[1,a] ⬤, **Ribamar R. R. Reis**[1,2,b]

[1] Instituto de Física, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ CEP 21941-972, Brazil
[2] Observatòrio do Valongo, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ CEP 20080-090, Brazil

Abstract The Baryon Acoustic Oscillations (BAO) are one of the most used probes to understand the accelerated expansion of the Universe. Traditional methods rely on fiducial model information within their statistical analysis, which may be a problem when constraining different families of models. This work aims to provide a method that constrains θ_{BAO} through a model-independent approach using the covariance matrix from the galaxy sample from thin redshift bins, later validated with a mock sample covariance matrix. We used widths of $\delta z = 0.002$ separation for all samples as the basis for a sample covariance matrix weighted by the statistical importance of the redshift bin. Each sample belongs to the Sloan Digital Sky Survey: BOSS1, BOSS2, and eBOSS, with effective redshift z_{eff}: 0.35, 0.51, 0.71, and different numbers of bins with 50, 100, and 200. To get θ_{BAO}, we correct the angular separation from the polynomial fit (θ_{fit}) by comparing each bin correlation function with the correlation function of the whole set, a parameter named $\tilde{\alpha}$. We also tested such correction by choosing the bin at z_{eff} and found that for eBOSS θ_{BAO} is in 1σ agreement with the Planck 18 model. Finally, we found that the sample covariances are noisy compared to the mocks for lower z samples, something expected due to nonlinear effects. Such noise impact can be seen in the parameter constraints but does not affect the eBOSS covariance sample. It is shown that mocks' results do tend to its chosen fiducial cosmology θ_{BAO}. BOSS1 and BOSS2 showed agreement with Planck 18 and an agreement with Pantheon + SH_0ES when $\tilde{\alpha}$ is based on the bin $z = z_{eff}$.

1 Introduction

The Baryon Acoustic Oscillations (BAO) are one of the most used probes to understand the accelerated expansion of the Universe. Cosmological information can be extracted through the two-point correlation function and power spectra estimated with the sky distribution and redshift of standard tracers [39]. Among the tracers, the most used are the luminous red galaxy (LRG) first used by [18,40]. Now, a multi-tracer analysis is possible with emission line galaxies [16,53], quasars [25] and Lyman-α forests [17].

Future and current surveys will reach a larger number of observed objects, such as the Dark Energy Spectroscopic Instrument (DESI) [20], the Dark Energy Survey (DES) [45], the Large Synoptic Survey (LSST) survey [4], the Javalambre-Physics of the Accelerated Universe Astrophysical Survey (J-PAS) [8], Euclid [49]. Larger samples have the advantage of being statistically robust and less susceptible to cosmic variance [35]. Their goal is to reach higher precision in order to test different cosmological models. This is only achievable through a template analysis that can be applied to any model.

The traditional methods rely on fiducial cosmology and *ad hoc* parameters to fix nonlinear effects [51]. Moreover, the statistical analysis carries fiducial model information within its template. A possible issue with these methods is the applicability to test other families of models. References [47,48] analysed the angular and radial correlation functions using a polynomial fit, also independent of the model. References [32,33] followed this construction using the Sloan Digital Sky Server (SDSS) precise surveys. Other alternative approaches have been proposed in the literature such as the new standard ruler called Linear Point [5,6], which can be used to derive a purely geometrical BAO distance [7].

Parameter estimation from Large Scale Structure (LSS) galaxy counting requires a covariance matrix. Usually, the

community uses hundreds of mock catalogs that mimic surveys. This can be done by N-body simulations such as the MultiDark-Patchy Mocks by [28,44], and the N-body Parallel Particle-Mesh GLAM code (PPMGLAM) by [29], which is the core of GLAM (GaLAxy Mocks). Simulation mocks are highly computationally expensive because they solve the matter density field evolution. The solution for many collaborations was to use a modest mock construction like the Log-Normal mocks, that can be found in `CoLoRe` [41], `FLASK` [55], `nbodykit` [23] `LogNormalCatalog`. Both approaches require fiducial cosmology, something that would be desirable to avoid.

Reference [50] estimated the covariance matrix assuming that the error distribution is Gaussian. References [57,58] used the so-called jackknife error estimate, by dividing a larger sample into sub-samples in order to find the covariance matrix. Their analysis concluded that the results were representative error estimators. Reference [46] also used subsamples to find the covariance matrix and error bars from the data set. However, there was a lack of observed objects. The latest data sets from spectroscopic surveys must provide even more representative errors and covariance matrices once we are provided with more objects, thus thinner redshift subsamples.

This work aims to provide a pipeline to constrain θ_{BAO} through a model-independent estimation of the angular correlation function using the covariance matrix from the data sample (sample covariance). We use thin bins of redshift with width $\delta z = 0.002$ as the basis for a such matrix weighted by the statistical importance of the redshift bin. Our analysis compares the sample covariance to the covariance matrix originating from mock catalogs and also the traditional covariance from mocks, without the binning. We compare samples with different effective redshift values, from lowest z to highest z and check the consistency according to each scale. Furthermore, we correct the angular separation that encodes the BAO feature of the [47] polynomial fit by a bias function that compares the correlation function of each bin with the whole set.

Our analysis is first described with our methodology in Sect. 2 where we describe the data used in Sect. 2.1, how it was divided into redshift bins in Sect. 2.2, and finally its correlation function estimator in Sect. 2.3. Next, we discuss the construction of the covariance matrix with the data and with mocks in Sect. 3. The polynomial function is described in Sect. 4.2 and our θ_{BAO} method in Sect. 3.2. The results and conclusion are found in Sects. 4 and 5, respectively.

2 Methods

2.1 Data and mocks

In this work, we used two data sets from the SDSS. The LRG and the LRG CMASS data from the Data Release 16 (DR16) [53] of the extended Baryon Oscillation Spectroscopic Survey (eBOSS) spectroscopic observations from the SDSS IV. The sample is distributed in Northern Galactic Cap (NGC) and Southern Galactic Cap (SGC) with a redshift range of $0.6 \leq z \leq 1.0$ and a total of 552,274 galaxies. Its random catalogue contains fifty times more galaxies than the real one.

The other set consists of galaxies from the BOSS sample of the SDSS-III DR12. We separated a redshift range $0.3 \leq z < 0.4$ called the BOSS1 set and BOSS2 range is $0.4 \leq z < 0.6$. The summary of the data specification can be found in Table 1. We chose this minimum redshift because, according to our tests, smaller redshift ranges did not show a significant signal for the BAO feature, as expected.

For the eBOSS set, we used $N = 1000$ realisations of realistic mocks catalogs from [60], for each galactic cap, based on the effective Zel'dovich approximacovariancetion mock generator (EZmock) by [15]. These mocks were constructed using the Zel'dovich approximation [59] which is faster than producing mocks using N-body simulations and accurate in clustering statistics.

The BOSS set was validated with $N = 1000$ realisations of the MultiDark-Patchy mocks (MD-Patchy mocks). These mocks were constructed for both the LOWZ redshift range [28] and the CMASS part [44]. We must keep in mind there is a difference between the two mocks' construction, MD-Patchy mocks were written from the N-body simulation Big-MultiDark simulations [30] using the Patchy code [26,27] which considers a biasing model. EZmock also used Big-MultiDark, but their approach is not fully N-body based, and different from PATCHY, they avoid a strict biasing model.

We cut the mocks to match the survey's redshift range. The same routine was applied to calculate the angular correlation function for the mocks using the same random catalogue for all mocks.

2.2 Redshift bins

In order to get a direct BAO measurement, we used an approach similar to the one developed by [47].

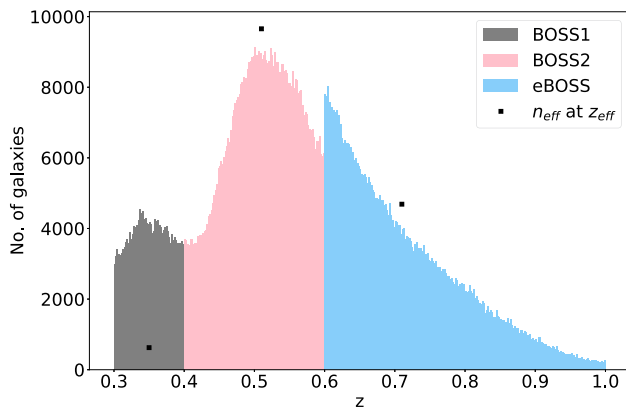eBOSS sample was divided into 200 redshift bins with

Fig. 1 Each sample bin separation. The black squares are the n_{eff} of each sample. BOSS1, BOSS2, and eBOSS are in gray, pink, and blue, respectively

width $\delta z = 0.002$. The choice was made considering the distribution of galaxies in the sky. One wants many bins to construct a fair covariance matrix and enough galaxies in each bin to get the BAO feature through angular counting. The angular pairs of each bin were computed using the code `corrfunc` developed by [52]. The data-data (DD), the data-random (DR), and random-random (RR) pairs counts calculated for angular pixels from $0.8°$ to $10°$ with width $\delta\theta \sim 0.4°$. This was also a convenient choice to find the BAO feature. Larger angles would provide a smeared function, while smaller ones would have too few galaxies.

The BOSS1 set was divided into 50 bins with width $\delta z = 0.002$, angular range $1° < \theta < 12°$ with $\delta\theta \sim 0.31°$. BOSS2 was divided into 100 bins with the same redshift separation and the same angular configuration as eBOSS. The redshift distribution for the chosen binning can be visualized in Fig. 1.

With the pairs counts, we computed the angular two-point correlation function for each redshift bin using the Landy-Szalay estimator [31]. Each angular function was normalised according to its corresponding i-th bin length. The angular correlation function (w_i) of the i-th bin is written in the following equation:

$$w_i(\theta) = \left(\frac{N_{i,rand}}{N_{i,data}}\right)^2 \frac{DD_i(\theta)}{RR_i(\theta)} - 2\frac{N_{i,rand}}{N_{i,data}}\frac{DR_i(\theta)}{RR_i(\theta)} + 1, \tag{1}$$

where $N_{i,rand}$ is the number of galaxies in the random bin, and $N_{i,data}$ in a data bin.

2.3 Weighting-scheme and correlation function

In addition to normalisation, it is important to account for the statistical significance of each bin. Our approach was to use the total weights ($w_{tot,a}$) that account for the systematic effects of the spectrographs. For the LRG sample, the weights

given by

$$w_{tot}^{LRG} = w_{sys,a} \cdot w_{cp,a} \cdot w_{noz,a}, \tag{2}$$

which is described in [22]. We used the prescription described in [42] with the pCMASS, eBOSS and CMASS BOSS samples:

$$w_{tot,a} = w_{sys,a}(w_{cp,a} + w_{noz,a} - 1), \tag{3}$$

where the total galaxy weights $w_{tot,a}$ depends on the total angular systematic weights $w_{systot,a}$, the weight for close pair correction $w_{cp,a}$, and the total weight to the nearest neighbour of the redshift failure $w_{noz,a}$. `corrfunc` [52] weights the pairs using the `pair_product` method, which is simply multiplying the weights of the galaxies in a pair.

We computed the effective redshift of the three samples according to the following equation based on [10]:

$$z_{eff} = \frac{\sum_{a<b}\frac{w_a w_b(z_a+z_b)}{2}}{\sum_{a<b} w_a w_b}, \tag{4}$$

where a and b represent galaxies in a sample, and $w_{a|b}$ is their w_{tot}. The values are shown in Table 1.

Random and mock bins do not carry the systematic effects of fibres on the spectrographs. Since the available FKP weight by [19] accounts for the number of galaxies in a volume, we use unit weights $w = 1.0$ for the angular correlation function. Although EZmock contains synthetic systematics within their weights we kept them unitary as a matter of consistency with MD-Patchy mocks.

As the bins have different sizes and redshifts, the bins also need to be weighted. For that, we consider a bin weighting revisiting [32]. Now, redshift bins are weighted instead of pixels.

The weighting of each bin is similar to the variance of the weights of each random galaxy. For each bin, the weight rr_i is written as:

$$rr_i = \frac{\left(\sum_j RR_{ij}\right)^2 - \sum_j RR_{ij}^2}{2} \tag{5}$$

where i stands for the redshift bin and j for j-angular separation. Reference [32] wrote rr_α as a weight for each pixel; here we use the pair counting RR because all of them are uncorrelated with the observed points but still an accurate representation of them in terms of redshift. We want to weigh the bins according to statistical significance and that is related to how many galaxies are in the ij box.

Finally, the correlation function $W(\theta)$ is computed as a weighted mean of the functions for each bin

$$W(\theta) = \frac{\sum_i w_i(\theta) \cdot rr_i}{\sum_i rr_i}. \tag{6}$$

Table 1 Data specifications

Sample	z range	# of Galaxies	z_{eff}	No. of bins
BOSS1	$0.3 \leq z < 0.4$	$192,285$	0.35	50
BOSS2	$0.4 \leq z < 0.6$	$686,370$	0.51	100
eBOSS	$0.6 \leq z \leq 1.0$	$552,274$	0.71	200

Table 2 Priors used in the MCMC, both with ζ_{ml} and C_{ml}

	C	θ_{fit}	σ
BOSS1	$\mathcal{N}(5 \times 10^{-3}, 1 \times 10^{-5})$	$\mathcal{N}(6.00, 0.01)$	$\mathcal{N}(1.00, 0.01)$
BOSS2	$C > 0$	$\mathcal{N}(4.2, 0.1)$	$\mathcal{N}(0.04, 0.01)$
eBOSS	$\mathcal{N}(7.60 \times 10^{-3}, 1 \times 10^{-5})$	$\mathcal{N}(2.70, 0.01)$	$\sigma > 0$

3 Covariance matrix and θ_{BAO}

The whole sample has an effective number of galaxies per bin (n_{eff}):

$$n_{eff} = \frac{\left(\sum_i rr_i \right)^2}{\sum_i rr_i^2}. \tag{7}$$

We consider a good sample the one that has an n_{eff} close to the number of galaxies at the $z_i \sim z_{eff}$. This can be seen in Fig. 1, where we display the n_{eff} of each sample as black squares. The BOSS2 and eBOSS present this characteristic, while the smaller sample with fewer bins shows disagreement with the size of the bin.

Equation (1) works as a matrix of many correlation function realisations, similar to the usual usage of mocks. We use each $w_i(\theta)$ compared to $W(\theta)$ from Eq. (6) to construct the sample covariance matrix ζ_{ml}:

$$\zeta_{ml} = \frac{n_{eff}^{-1}}{(N_i - 1)} \sum_{i=1}^{N_i} [w_i(\theta_m) - W(\theta_m)]$$
$$\times [w_i(\theta_l) - W(\theta_l)]. \tag{8}$$

Basically, we compute the bias between each $w_i(\theta)$ and $W(\theta)$, but because $w_i(\theta)$ is not weighted according to Eq. (3), so we must correct this by n_{eff}.

In order to validate our method we require realisations of the survey catalogues from mock catalogues. Mocks are an approximation of a survey concerning its footprint and redshift. However, when analysing the clustering statistics we need to remember that mocks have the challenge of accounting for nonlinear effects which are inevitable for lower redshift values, which is mentioned by other studies. We use mocks to compute $W_k(\theta)$ of the k-mock and test the method. Now, the covariance matrix C_{ml}:

$$C_{ml} = \frac{n_{eff}^{-1}}{N - 1} \sum_{k=1}^{N=1000} [W_k(\theta_m) - \bar{W}(\theta_m)]$$
$$\times [W_k(\theta_l) - \bar{W}(\theta_l)], \tag{9}$$

where $\bar{W}$ is the average over the correlation function $W_k(\theta)$ of each mock. Nevertheless, Eq. (9) is not model-independent,

because the mocks were constructed assuming a fiducial cosmology.

3.1 Polynomial function

We used [47] expression below for the angular two-point correlation function to fit our estimated $W(\theta)$.

$$\mathcal{W}(\theta) = A + B\theta^\gamma + Ce^{-\frac{(\theta - \theta_{fit})^2}{2\sigma^2}} \tag{10}$$

The parameter A is related to the behaviour of the function after the BAO peak. B and C weigh the importance of their respective terms, if there is no BAO peak, for instance, $C = 0$. γ is the power law of the function's overall shape. The physical parameters are θ_{fit} and σ, θ_{fit} gives the position of the BAO, while σ is the width of the BAO.

We applied Gaussian priors to the parameters that are related to the BAO: C, θ_{fit}, and σ, the priors are written in Table 2. The parameter estimation was done using the maximum likelihood estimator method through the emcee's Affine Invariant Markov Chain Monte Carlo (MCMC) Ensemble sampler software by [21]. The fitting is done with the $W(\theta)$ from the data, we will compare the best-fit results alternating between ζ_{ml} and C_{ml}.

The Chi-squared function is:

$$\chi^2 = \sum_{ml} [W(\theta_m) - \mathcal{W}(\theta_m)] \zeta_{ml}^{-1} [W(\theta_l) - \mathcal{W}(\theta_l)], \tag{11}$$

the same equation is used with C_{ml}.

3.2 The BAO scale θ_{BAO}

Should the estimator and its covariance matrix be representative of the angular pairs counting of the LSS, we can constrain cosmological parameters from fitting a model to the data. For that we need to find the BAO signal, θ_{BAO} as a function of θ_{fit}. Here, we will focus on the use of a totally model-independent procedure, leaving a model that gets the advantage of our estimator for future work.

We propose a similar method to [13] to find θ_{BAO}. We take the correlation function estimation for the whole set, as in Eq. (1) without weighting according to the random bins, to find $w_{\delta z=0}(\theta)$ and compute a bias function, the difference between $w_{\delta z=0}(\theta)$ and the $w_i(\theta)$ not normalised by rr_i:

$$\Delta w_i(\theta) = w_i(\theta) - w_{\delta z=0}(\theta), \tag{12}$$

Next, we take this bias function and compute the 20th percentile and the median over the bins to get our shift parameter $\tilde{\alpha}$ as a function of θ:

$$\tilde{\alpha}_p(\theta) = \frac{\sum_i^{N_p} \Delta w_i(\theta)}{N_p} \geq \frac{p}{100}, \qquad (13)$$

$\tilde{\alpha}_p$ is normalised by $\sqrt{\int |w_i(\theta) - w_{\delta z=0}(\theta)|^2 d\theta}$, which is the contribution of all dispersion from $w_{\delta z=0}$. Here, p stands for percentile, $\tilde{\alpha}_{20}(\theta)$ represents the bias function of a bin that is greater or equal to 20% of the bias function of the other bins, $\tilde{\alpha}_{50}(\theta)$ is 50% greater or equal to the other bias functions. In the case of the bias function for the bins was a normal distribution, the average value would be the percentile 50. Knowing a perfect normal distribution is not achieved we also test the percentile 20, which indicates whether the most statistically significant bias is dislocated from the average.

Finally, θ_{BAO} is the correction of the model-independent θ_{fit} according to the bias of each with respect to the correlation function if the binning was not applied:

$$\theta_{BAO}(z_{eff}) = \theta_{fit} + \tilde{\alpha}(\theta_{fit})\theta_{fit}. \qquad (14)$$

$\tilde{\alpha}(\theta_{fit})$ is the value of the function $\tilde{\alpha}$ at the angle found from the MCMC fitting results, both from sample covariance and mock covariance. We used $\tilde{\alpha}_{20}$ and $\tilde{\alpha}_{50}$ in Eq. (14) and will present the results for both.

We also tested $\tilde{\alpha}$ in Eq. (14) as the normalized dispersion between the whole set correlation function and the one for the bin which contains z_{eff}. This bin is expected to be representative of the whole sample as described in Sect. 2.3. We can write it as

$$\tilde{\alpha}_{eff}(\theta) = \frac{w_{eff}(\theta) - w_{\delta z=0}(\theta)}{\sqrt{\int |w_{eff}(\theta) - w_{\delta z=0}(\theta)|^2 d\theta}}, \qquad (15)$$

4 Results

4.1 Covariance matrix

The correlation matrices of both mocks and data sets are shown in Fig. 5. The top left panel shows the correlation matrix of the BOSS1 sample, while the MultiDark Patchy Mocks results are shown on the top right one. Both plots show a higher correlation at smaller scales, from $\sim 1°$ to $\sim 5°$, and the data matrix is more noisy than the mocks, as one could expect.

In the middle left panel, we can see that BOSS2 has more low correlation patches when compared to BOSS1, possibly due to the bigger number of objects and depth of this sample. Its mock counterpart shows much less noise, as expected, and similar over-correlated regions on both extremes of the angular separation. Additionally, the BOSS2 sample shows
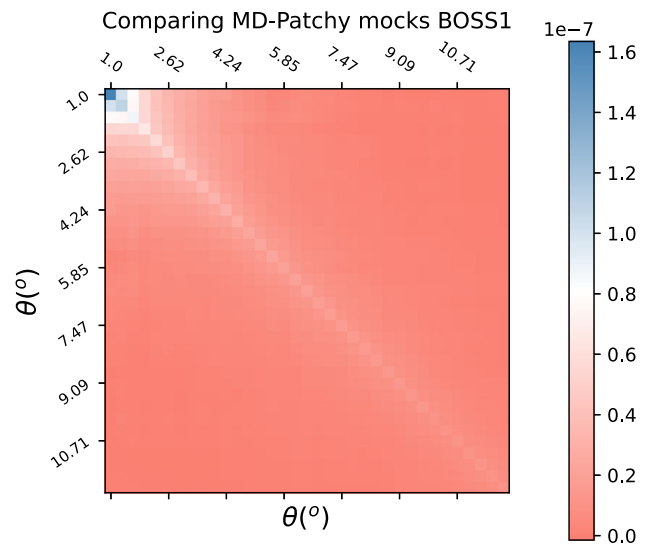

Fig. 2 Comparison between our method and a covariance matrix without tomography

a high correlation between small and big scales that is not visible in BOSS1 and the mocks.

Lastly, the results for eBOSS, the larger and deeper sample, are shown in the bottom panels of Fig. 5. The data correlation matrix is clearly more noisy, and it does not show regions of stronger or weaker correlation, except for the diagonal. The EZ mocks appear to have a similar pattern to that of the other mocks, but there is a stronger correlation region on larger scales, unlike the other mocks.

As a means of validation, we computed the angular correlation function for each mock without our binning method. This full covariance equation is described below:

$$Cov_{ml}^{full} = \frac{n_{eff}^{-1}}{N-1} \sum_{k=1}^{N=1000} [W^{full}_k(\theta_m) - \bar{W}^{full}(\theta_m)]$$
$$\times [W^{full}_k(\theta_l) - \bar{W}^{full}(\theta_l)]. \qquad (16)$$

In Figs. 2, 3, and 4, we show $C_{ml} - Cov_{ml}^{full}$. The difference is of the order 10^{-7} for BOSS1 and eBOSS and 10^{-11} for BOSS2. For all samples, the difference is negligible near the BAO angular feature for the lower-z sample. However, this difference is larger for eBOSS, but still a small difference. This validates our method as a robust one (Fig. 5).

We performed the minimum likelihood estimation with Cov_{ml}^{full}, the results are in Table 7, the results show that they agree in 1σ with the constraints from C_{ml}.

4.2 Polynomial fit

After taking the steps described in the previous section, we get the estimated correlation function, where the results of the physical parameters of the fitting formula are compiled
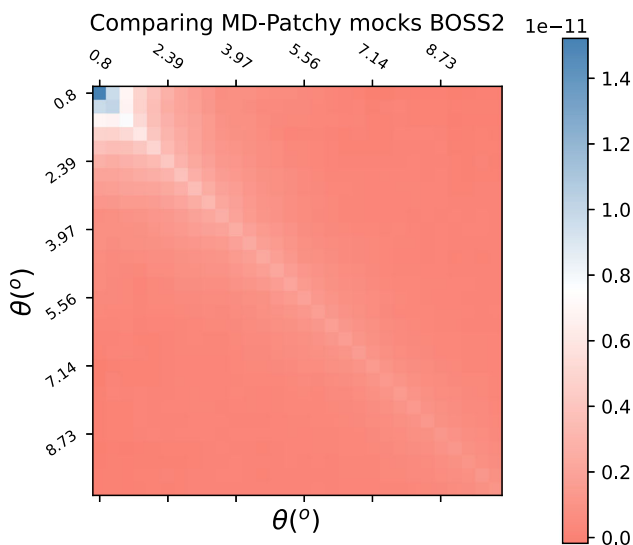
Fig. 3 Comparison between our method and a covariance matrix without tomography
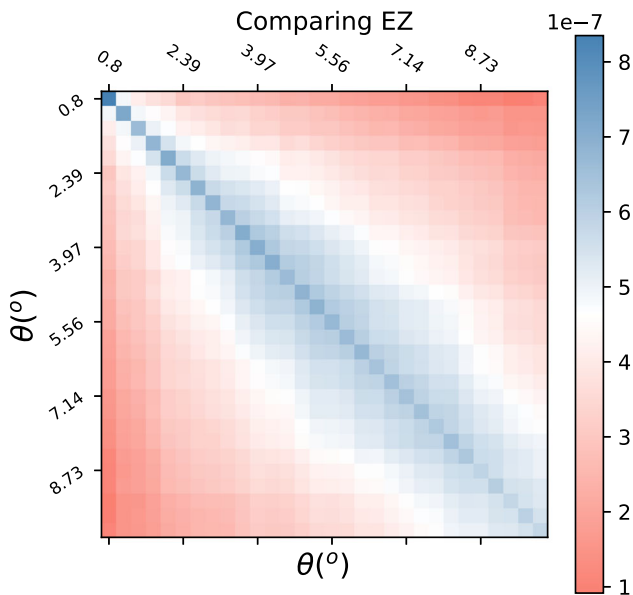

Fig. 4 Comparison between our method and a covariance matrix without tomography
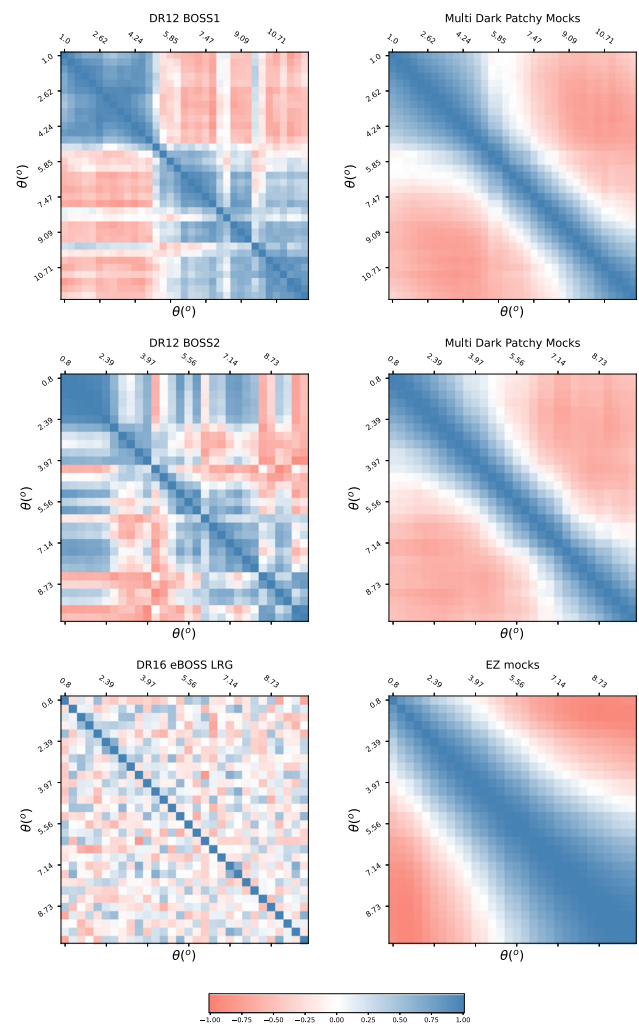

Fig. 5 Correlation matrices relative to ζ_{ml}, left column, and C_{ml}, right column. Negative correlation in salmon, a positive correlation in blue, and a zero correlation in white colour

in Table 3. The comparison between mock and data fitting is shown in Figs. 6, 7, and 8. The contour indicates the 68%, 95%, and 99.7% confidence level regions, from darker to lighter colours.

We chose Gaussian priors for the physical parameters for all samples. But instead of using traditional random walkers, we used a mixture of moves in our code, in order to improve the performance of the MCMC walkers. 80% of the moves are a "differential evolution" algorithm Differential Evolution MCMC (DEMCMC) by [36], available in `emcee` as `DEMove`. The other 20% of the moves are done with `KDEMove`.

Table 2 shows the priors chosen for all cases. For all results shown in Figs. 6, 7 and 8, and Table 3, we used the real data and changed the covariance as indicated. BOSS1 and BOSS2 results for the parameter C did not agree at $\sim 5\sigma$ when we change the covariance from data to their respective mocks, while eBOSS does not show such issue.

eBOSS agrees within 1σ compared to their respective mocks for all parameters, as seen in Fig. 8, with the mocks covariance presenting tighter constraints as expected due to the lower noise. The posteriors are not close to a Gaussian distribution, showing a long tail towards higher values of C.

BOSS1 did not show the same behaviour; the main reason is the size of the sample. It is reasonable to expect that the mocks do not include all possible nonlinear effects of the real LSS. This should be particularly important for lower redshift samples, where such effects will be greater. The method used here does not account for this depth dependence. These

Table 3 Best-fit parameters

Sample	C	$\theta_{fit}(°)$	$\sigma(°)$
BOSS1(ζ_{ml})	0.010 ± 0.001	5.66 ± 0.02	0.15 ± 0.03
BOSS1(C_{ml})	0.0047 ± 0.0001	5.79 ± 0.006	0.198 ± 0.005
BOSS2(ζ_{ml})	0.002 ± 0.0001	4.24 ± 0.02	0.30 ± 0.01
BOSS2(C_{ml})	0.02 ± 0.01	4.21 ± 0.02	0.030 ± 0.009
eBOSS(ζ_{ml})	0.008 ± 0.002	2.72 ± 0.09	0.13 ± 0.05
eBOSS(C_{ml})	0.009 ± 0.001	2.77 ± 0.04	0.12 ± 0.03
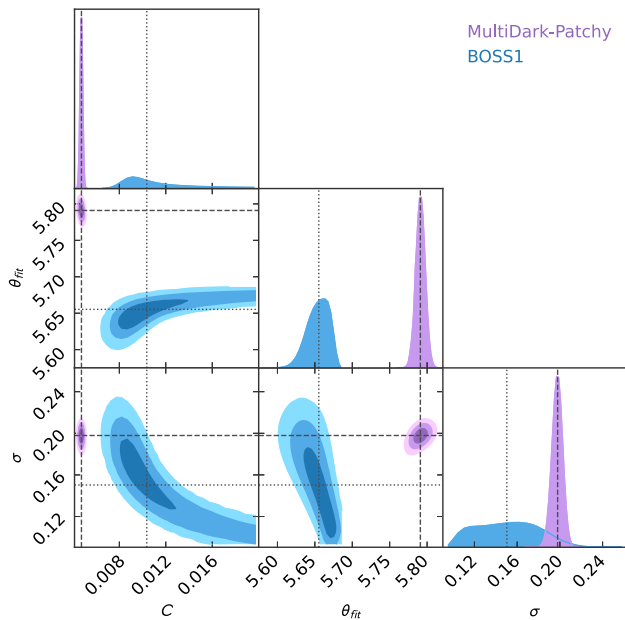
Fig. 6 Comparison between results for the BOSS1 using using ζ_{ml} in blue, and C_{ml} in purple, using pygtc [11] for the parameters C, θ_{fit}, and σ

Fig. 7 Comparison between results for the BOSS2 using using ζ_{ml} in blue, and C_{ml} in purple, using pygtc [11] for the parameters C, θ_{fit}, and σ
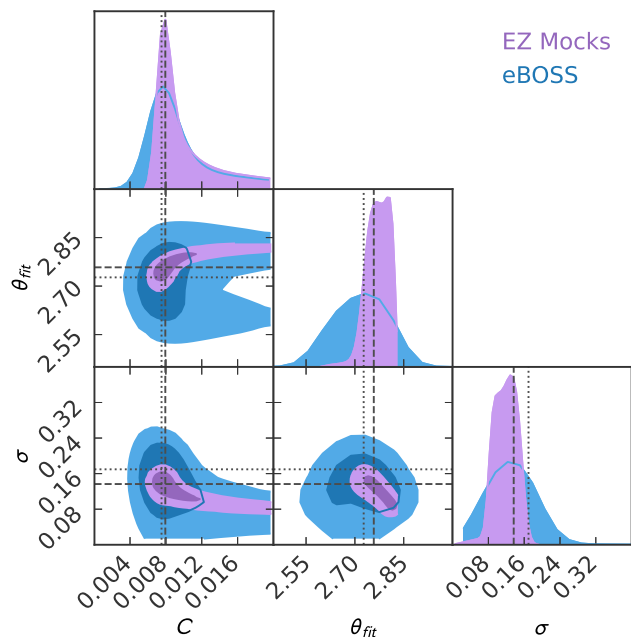
Fig. 8 Comparison between results for the eBOSS using using ζ_{ml} in blue, and C_{ml} in purple, using pygtc [11] for the parameters C, θ_{fit}, and σ

struggles with lower redshift and modelled solutions have been discussed extensively in the literature, especially in the 3D power spectrum case as seen in [51,56].

We see in Fig. 6 that, despite the tension in the results, both C_{ml} and ζ_{ml} show the BAO feature, with tighter 1σ CL for C_{ml} results. The first sample is affected by two peculiarities. As a consequence, we expect that the precision of the angular diameter distance will be affected.

BOSS2 presented better results based on ζ_{ml} (Fig. 7), while for C_{ml} (Fig. 7) we obtained broader posteriors. This again reflects that this method does not account for nonlinear effects and that the mocks are not perfect idealizations of the LSS.

4.3 θ_{BAO} results

The bias function from Eq. (12) can be visualized as a scatter matrix between the redshift bins. Some bin distributions can be found in Appendix A, in which we chose 10 bins close to

the effective redshift to understand this bias relation. If the bias function is close to zero, the difference between the CFs of each bin and the whole set is supposed to be close to a symmetric distribution with mean zero.

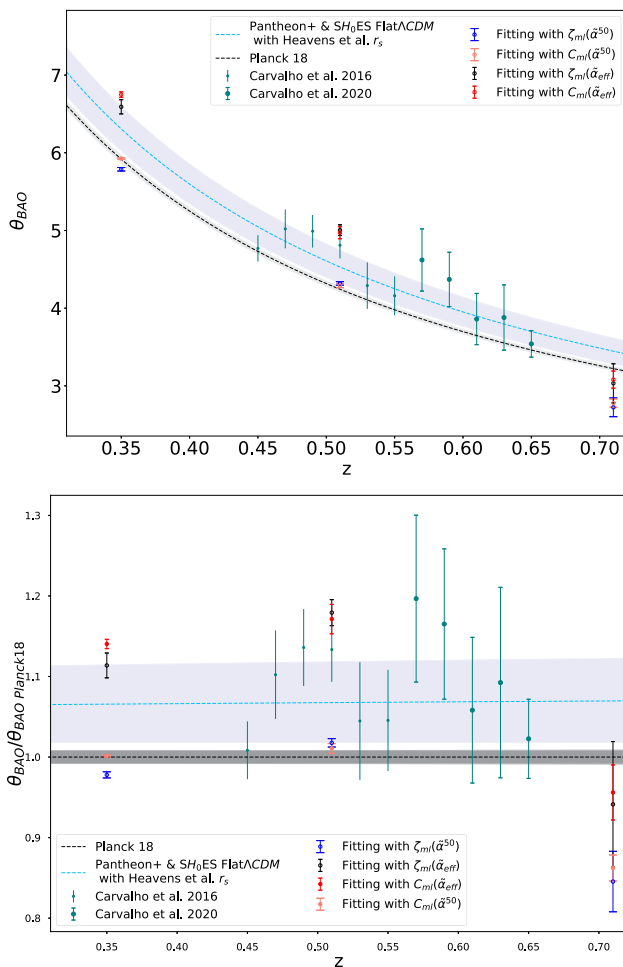In Fig. 9, top panel, we show the θ_{BAO} results from Eq. (14). The sample covariance results are shown in blue,

Fig. 9 *Top*: θ_{BAO} as a function of redshift with α from Eq. (13) and (15). Planck 18 [3] θ_{BAO} is shown as the black line and SH_0ES [12] as the dashed blue region. We kept the blue colour for the results with ζ_{ml} and salmon for C_{ml}. The shaded regions are the CL of Planck 18 in grey and SH_0ES in purple 68% CL. *Bottom*: Same result of the top panel divided by the Planck 18 one ($\theta_{BAO}/\theta_{BAO}^{Planck18}$)

Table 4 Parameters for FlatΛCDM cosmology for Planck 18 [3] and Pantheon+ & SH_0ES [12]

Parameters	Planck 18	Pantheon+ & SH_0ES
Ω_m	0.3153 ± 0.0073	0.334 ± 0.018
Ω_Λ	0.6847 ± 0.0073	0.666 ± 0.018
H_0(km s^{-1} Mpc^{-1})	67.36 ± 0.54	73.6 ± 1.1

Table 5 θ_{BAO}

	$\theta_{BAO}(°)$
BOSS1 with $\tilde{\alpha}^{50}$	5.79 ± 0.02
BOSS2 with $\tilde{\alpha}^{50}$	4.32 ± 0.02
eBOSS with $\tilde{\alpha}^{50}$	2.72 ± 0.12
BOSS1 mocks with $\tilde{\alpha}^{50}$	5.93 ± 0.01
BOSS2 mocks with $\tilde{\alpha}^{50}$	4.29 ± 0.03
eBOSS mocks with $\tilde{\alpha}^{50}$	2.79 ± 0.05
BOSS1 with $\tilde{\alpha}_{eff}$	6.59 ± 0.09
BOSS2 with $\tilde{\alpha}_{eff}$	5.00 ± 0.07
eBOSS with $\tilde{\alpha}_{eff}$	3.03 ± 0.25
BOSS1 mocks with $\tilde{\alpha}_{eff}$	6.75 ± 0.03
BOSS2 mocks with $\tilde{\alpha}_{eff}$	4.97 ± 0.08
eBOSS mocks with $\tilde{\alpha}_{eff}$	3.08 ± 0.11

while the mock covariance results are in salmon. The black and red points represent the $\tilde{\alpha}_{eff}$ results of the sample covariance and mock covariance respectively. The sea-green points are results from [13,14].

As a means of comparison, we use two different cosmological parameter references to check our results. The first is the Planck 2018 Collaboration [3], shown in black with a gray-shaded region. We considered the TT, TE, EE+lowE+lensing constraints. The second model is based on the Pantheon+ & SH_0ES [12] FlatΛCDM in light blue with a purple shaded region. The cosmological parameters can be found in Table 4. The shaded regions are the 68% (darker) and 95% (lighter) confidence levels (CL) for the results (Tables 5, 6). The curves are plotted using the following relation as a function of redshift:

$$\theta_{BAO} = \frac{r_s}{(1+z)D_A(z)}, \qquad (17)$$

where $D_A(z)$ is the angular diameter distance as a function of redshift and $r_s = 147.090 \pm 0.026$ Mpc is the sound horizon at drag epoch which is the Planck 18 [3] TT, TE, EE, lowE, lensing result for the black region. For SH_0ES fiducial model, we chose [24] as the sound horizon, this gives $r_s = 142.8 \pm 3.7$ Mpc. Figure 9, bottom panel, shows θ_{BAO} normalised by Planck 18 cosmology $\theta_{BAO}^{Planck18}$, in order to ease the comparison.

We see in Fig. 9, top panel, that the highest z_{eff} sample, eBOSS, has the larger error bars, this is due to the incompleteness of the sample. Since Eq. (6) weights the correlation function of each bin according to its random catalog size, this affects θ_{fit} and consequently θ_{BAO}, as the number of galaxies per bin decreases with redshift for eBOSS, so the n_{eff} is smaller compared to the other samples which increase the error bars. For all samples, the mock results have smaller error bars due to a larger number of galaxies per bin. Since we chose the same bin size ($\delta z = 0.002$) for all samples this results in fewer galaxies per bin for BOSS1 when compared to BOSS2 which leads to greater error bars for BOSS1. A summary of θ_{BAO} results is in Table 5

It is shown in Fig. 9, top panel, that when using $\tilde{\alpha}_{50}$, BOSS1 and BOSS2 agree with Planck 18 in 2 σ and 1 σ respectively, while eBOSS is in tension with Planck 18. Using mocks covariance, the results for BOSS1 and BOSS2 both agree with Planck 18 results, eBOSS constraints with C_{ml}

Table 6 The difference between the 20th percentile and the 50th percentile

	$\theta_{BAO}^{20} - \theta_{BAO}^{50}$	$\theta_{BAO}^{20}/\theta_{BAO}^{50}$
BOSS1	-6.97×10^{-4}	99.98%
BOSS2	-2.81×10^{-4}	99.99%
eBOSS	-7.22×10^{-3}	99.73%
BOSS1 mocks	-1.77×10^{-3}	99.97%
BOSS2 mocks	-6.98×10^{-3}	99.98%
eBOSS mocks	-7.22×10^{-3}	99.73%

also show a tension with Planck 18. This shows a tendency of the mocks' results to "reach" Planck 18 results.

When using $\tilde{\alpha}_{eff}$ the results change significantly. First, the error increases as a result of statistical loss, now we compare a single bin with $w_{\delta=0}$. The θ_{BAO} now have higher values, BOSS1 and BOSS2 (in black) agree with SH_0ES results in 1σ and 2σ, respectively. eBOSS match Planck 18 in 1σ. BOSS2 has the same z_{eff} as one of the bins from [13] that, now, they agree in 1σ. We must stress that [13,14] show larger error bars as a choice to avoid overlapping between bins.

The higher difference for the lower redshift samples can be explained by non-linearity influence, especially when we see the BOSS1 mocks' results (in red). It is hard to model all the necessary details of local galaxies clustering, as a result, θ_{BAO} is not in agreement with the sample covariance constraints.

Reference [33] used the same data set, but chose a different approach from [32] and a different selection cut than ours. Their method is based on [32] as well as the present study but maintained the mocks as the only covariance matrix estimator. We had a similar pattern of θ_{BAO} as a function of the redshift. Like our results, many of their bins did not agree with Planck 18 cosmology. Moreover, their findings were capable of extending to cosmological parameter inference because we are just interested in the fiducial model-independent analysis itself.[1]

It has been discussed by [38] that 15 transversal BAO measurements combined with Planck 18 alleviates the tension between Planck 18 and [43] Cepheid calibration. Reference [37] also showed that combining the same measurements with Big Bang Nucleosynthesis (BBN) information [2] and strongly lensed quasars from H0LiCOW [54], and BBN and cosmic chronometers (CC) [34] agree in 1σ with SH_0ES [43] like our results using $\tilde{\alpha}_{eff}$. Still concerning tensions with Planck 18 cosmology, in [9], the authors showed that the combination of Planck 18 and the 15 transversal measurements

solves the H_0 tension by pushing to higher values of the Hubble parameter when considering an Interacting Dark Energy (IDE) model. Moreover, a considerable mismatch between Planck 18 cosmology and this study, [13,14,33] could be explained by the influence of non-linearity is much higher in the lower redshift samples as compared to CMB experiments.

5 Conclusions

We used BOSS and eBOSS LRG samples to obtain the angular feature of the BAO using thin bins to construct its covariance matrix. Each sample BOSS1, BOSS2, and eBOSS had 50, 100, and 200 thin redshift bins with $\delta z = 0.002$ width. We adapted [32,47] methodology and wrote an angular correlation function estimator through those thin bins using a weighting scheme based only on the random catalog. We also performed the full analysis using both covariance matrices from mocks and real data. We considered the random binning as the reference to the weight of each bin according to the number of galaxies Poisson distributed in the sky.

The comparison between mocks (C_{ml}) and sample covariance (ζ_{ml}) best-fit showed that they agree at least for higher redshift. BOSS1 and BOSS2 showed disagreement with the physical parameters C and σ. We must remind ourselves that the mocks are idealisations of the LSS and its non-linearity, but they may not represent the exact structure of the data, which could explain the disagreements in BOSS1 and BOSS2 results. We compared the method we proposed to traditional mocks without binning the data, the maximum difference is of the order 10^{-7} and is even smaller for the scale close to the feature of interest.

Furthermore, our approach is purely statistical, the way we divide the bins and the amount of galaxies for each bin also changes the covariance matrix. Therefore, one must divide the bins to hold statistical significance (many galaxies in one bin) and good correlation significance (as many bins as possible to a robust mean to represent the whole sample).

Considering statistical analysis, eBOSS performed better in our analysis, first, because it is a deeper sample. Second, it has more thin bins than the other samples, a closer approximation to a redshift-space correlation function, but losing statistical significance. The results showed agreement between the data and mock covariance estimation for all physical parameters.

We provided a model-free approach to estimate θ_{BAO} from θ_{fit} revisiting [13], we compute the correlation function of the whole sample and then find a bias compared to each bin correlation function. This is used to write a bias function $\tilde{\alpha}$ that shifts θ_{fit} closer to θ_{BAO}. Instead of averaging over the bins, we get $\tilde{\alpha}(\theta)$ through the 20th and 50th percentile.

The results were compared to the fiducial models of Planck 18 [3] and Pantheon+ & SH_0ES [12] FlatΛCDM with [24]

---

[1] In a recent draft of the Dark Energy Survey's final release, [1], their results show a larger discrepancy from Planck 18 cosmology compared to their Y3 results (Fig. 6). Moreover, in Fig. 10, there is a considerable shift in the BAO feature compared to Planck 18 results.

r_s. Regarding the percentile approach. our findings indicated the samples with lower redshift agree at least in 2σ with Planck 18 results. eBOSS, on the other hand, did not agree with any of the fiducial cosmology described here.

When assuming $\tilde{\alpha}_{eff}$, the results change, we see that eBOSS matched Planck 18 results. BOSS1 and BOSS2 agree at least in 2σ with Pantheon+ & SH_0ES fiducial cosmology. We also found agreement with the same sample as BOSS2 from [13] z=0.51 bin. This indicates that a loss of statistical significance interferes with how the measurement behaves provided that [13] chose to lose many bins to avoid overlapping between them.

The mocks showed a tendency of getting θ_{BAO} closer to Planck 18 results than the real data. The only exception was BOSS1, the lower redshift sample which is more susceptible to non-linearity effects, something known in the literature as a challenging characteristic for obtaining survey mocks.

Next-generation surveys with large samples are suitable for the method. DESI, for instance, promises 8 million LRGs [61] with $0.4 < z < 1.0$, ideal to reduce the nonlinearity noise in our covariance matrix. These future samples should allow us to use thinner bins without reducing the number of galaxies per bin, which seems to be a key feature of this method. A further study is required to construct a test for cosmological models from that methodology, this would require a larger sample with higher redshift distribution, something interesting to future photometric surveys. This, however, also comes with the price of adding the photometric redshift uncertainties and even probability distribution functions of each object.

Data Availability Statement Data will be made available on reasonable request. [Author's comment: This is a theoretical study and no experimental data.]

Code Availability Statement This manuscript has no associated code/ software. [Author's comment: The data comes from the SDSS survey, that can be downloaded through https://data.sdss.org/sas/.]

Appendix A: Bias function scatter matrices

In Figs. 10, 11, and 12, we present the scatter matrix of the bias function from Eq. (12) for 10 bins close to the z_{eff} for the three samples we chose to analyse. It is clear that BOSS1 and BOSS2 (Figs. 10, 11) show a high bias between the bins and whole set correlation function, in other words, they are not a symmetric distribution with mean zero. eBOSS (Fig. 12), on the other hand, maintains symmetric distributions even between bins distant from z_{eff}. This reflects the strong correlation between bins for low-z samples due to nonlinearities.
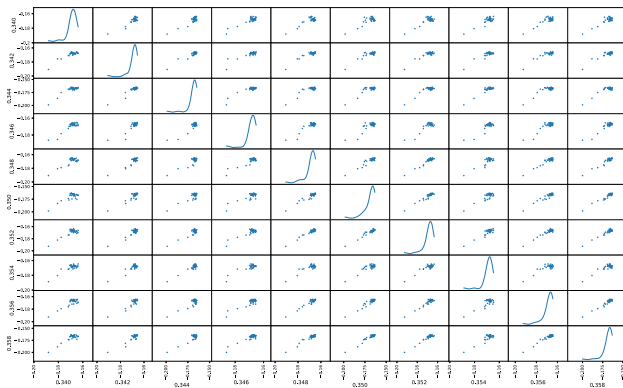
Fig. 10 BOSS1 scatter matrix of the bias relation between 10 neighbour bins. Both axes show the average z for each bin
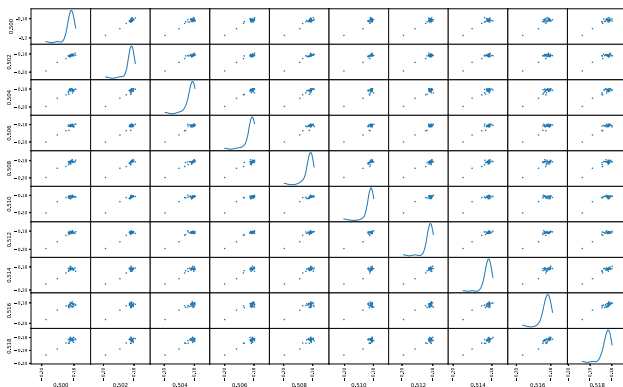

Fig. 11 BOSS2 scatter matrix of the bias relation between 10 neighbour bins. Both axes show the average z for each bin
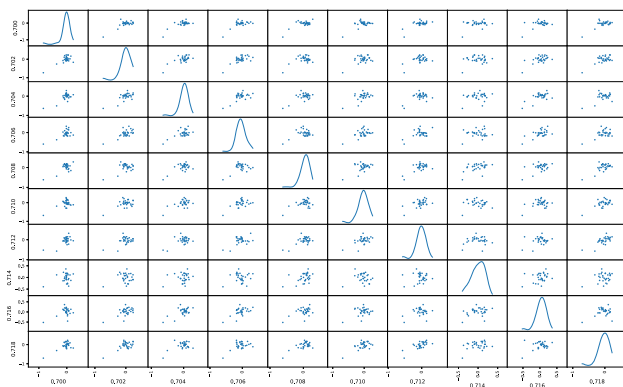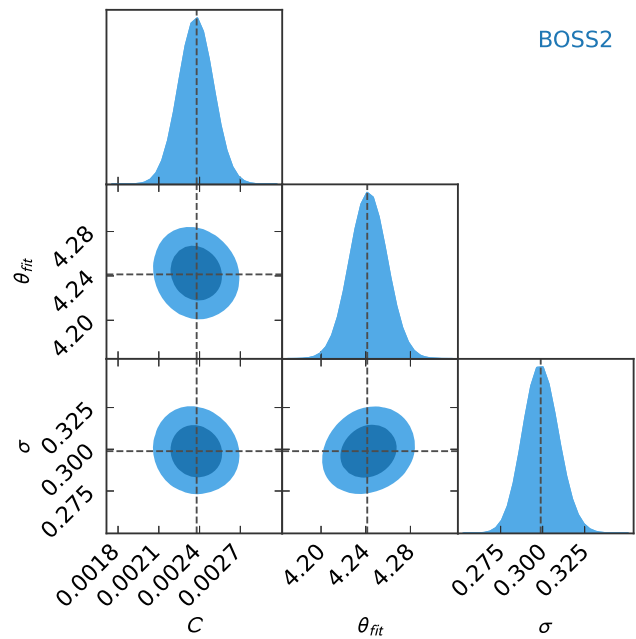

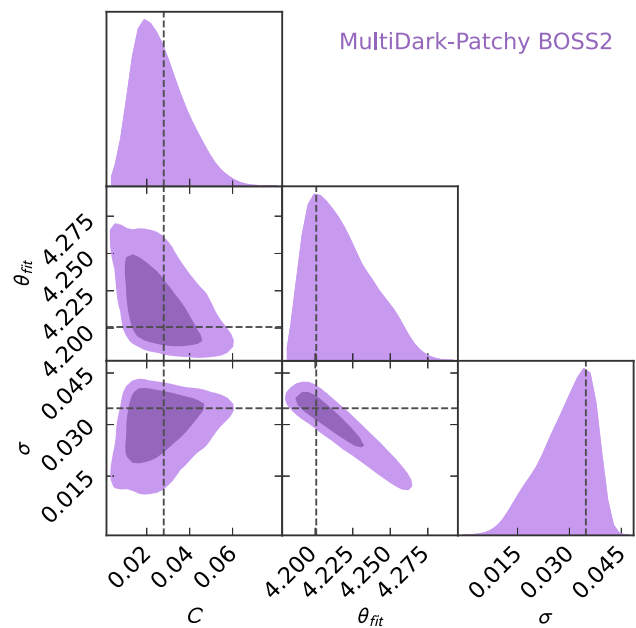**Fig. 12** eBOSS scatter matrix of the bias relation between 10 neighbour bins. Both axes show the average z for each bin

(a) BOSS2 results with ζ_{ml}.

(b) BOSS2 results with ζ_{ml}.

Fig. 13 BOSS2 results using `pygtc` [11]

Appendix B: BOSS2 triangle plots

In Fig. 13a and b, there are constraints for the separated physical parameters. The C_{ml} results did not perform as well as the real data inference which agrees with the discussion of Sect. 4.2. The mocks should be a good representation, but this is more challenging for a shallower sample. The very different results are a combination of nonlinear effects representation for the mocks and our method not being successful in fixing such issues.

Appendix C: Comparing with full mocks

In Table 7, we show the constraints with the full mocks samples.

Table 7 Best-fit parameters using Cov_{ml}^{full}

Sample	C	$\theta_{fit}(°)$	$\sigma(°)$
BOSS1	0.0065 ± 6 $\times 10^{-10}$	5.798 ± 2.0 $\times 10^{-8}$	2.7857 ± 2 $\times 10^{-8}$
BOSS2	(9.11 ± 0.00005) $\times 10^{-4}$	4.2 ± 15 $\times 10^{-7}$	$1.83 \times 10^{-1} \pm 1$ $\times 10^{-7}$
eBOSS	(8.090 ± 0.0001) $\times 10^{-3}$	2.74 ± 6 $\times 10^{-5}$	(162.301 ± 0.004) $\times 10^{-3}$

References

1. T. Abbott et al., (2024). arXiv preprint arXiv:2402.10696
2. E.G. Adelberger et al., Rev. Mod. Phys. **83**, 195 (2011)
3. N. Aghanim et al., Astron. Astrophys. **641**, A6 (2020)
4. R. Ansari et al., Astron. Astrophys. **623**, A76 (2019)
5. S. Anselmi, G.D. Starkman, R.K. Sheth, Mon. Not. R. Astron. Soc. **455**, 2474 (2016)
6. S. Anselmi, P.-S. Corasaniti, G.D. Starkman, R.K. Sheth, I. Zehavi, Phys. Rev. D **98**, 023527 (2018)
7. S. Anselmi, P.-S. Corasaniti, A.G. Sanchez, G.D. Starkman, R.K. Sheth, I. Zehavi, Phys. Rev. D **99**, 123515 (2019)
8. N. Benitez et al., (2014). arXiv preprint arXiv:1403.5237
9. A. Bernui, E. Di Valentino, W. Giarè, S. Kumar, R.C. Nunes, Phys. Rev. D **107**, 103531 (2023)
10. F. Beutler et al., Mon. Not. R. Astron. Soc. **416**, 3017 (2011)
11. S. Bocquet, F.W. Carter, J. Open Sour. Softw. **1**, 46 (2016)
12. D. Brout et al., Astrophys. J. **938**, 110 (2022)
13. G. Carvalho, A. Bernui, M. Benetti, J. Carvalho, J. Alcaniz, Phys. Rev. D **93**, 023530 (2016)
14. G. Carvalho, A. Bernui, M. Benetti, J. Carvalho, E. de Carvalho, J. Alcaniz, Astropart. Phys. **119**, 102432 (2020)
15. C.-H. Chuang, F.-S. Kitaura, F. Prada, C. Zhao, G. Yepes, Mon. Not. R. Astron. Soc. **446**, 2621 (2015)
16. A. De Mattia et al., Mon. Not. R. Astron. Soc. **501**, 5616 (2021)
17. H. Bourboux, D.M. Des et al., Astrophys. J. **901**, 153 (2020)
18. D.J. Eisenstein et al., Astrophys. J. **633**, 560 (2005)
19. H.A. Feldman, N. Kaiser, J.A. Peacock, (1993). arXiv preprint arXiv:astro-ph/9304022
20. B. Flaugher, C. Bebek, in *Ground-based and Airborne Instrumentation for Astronomy V* (2014), p. 282–289
21. D. Foreman-Mackey, D.W. Hogg, D. Lang, J. Goodman, [PASP] **125**, 306 (2013)
22. H. Gil-Marín et al., Mon. Not. R. Astron. Soc. **498**, 2492 (2020)
23. N. Hand, Y. Feng, F. Beutler, Y. Li, C. Modi, U. Seljak, Z. Slepian, Astron. J. **156**, 160 (2018)
24. A. Heavens, R. Jimenez, L. Verde, Phys. Rev. Lett. **113**, 241302 (2014)
25. J. Hou et al., Mon. Not. R. Astron. Soc. **500**, 1201 (2021)
26. F.-S. Kitaura, G. Yepes, F. Prada, Mon. Not. R. Astron. Soc. Lett. **439**, L21 (2014)
27. F.-S. Kitaura, H. Gil-Marín, C.G. Scóccola, C.-H. Chuang, V. Müller, G. Yepes, F. Prada, Mon. Not. R. Astron. Soc. **450**, 1836 (2015)
28. F.-S. Kitaura et al., Mon. Not. R. Astron. Soc. **456**, 4156 (2016)
29. A. Klypin, F. Prada, Mon. Not. R. Astron. Soc. **478**, 4602 (2018)
30. A. Klypin, G. Yepes, S. Gottlöber, F. Prada, S. Hess, Mon. Not. R. Astron. Soc. **457**, 4340 (2016)
31. S.D. Landy, A.S. Szalay, Astrophys. J. **412**, 64 (1993)
32. V. Marra, E. Isidro, G. Chirinos, Mon. Not. R. Astron. Soc. **487**, 3419 (2019)
33. R. Menote, V. Marra, Mon. Not. R. Astron. Soc. **513**, 1600 (2022)
34. M. Moresco et al., J. Cosmol. Astropart. Phys. **2016**, 014 (2016)
35. B.P. Moster, R.S. Somerville, J.A. Newman, H.-W. Rix, Astrophys. J. **731**, 113 (2011)
36. B. Nelson, E.B. Ford, M.J. Payne, Astrophys. J. Suppl. Ser. **210**, 11 (2013)
37. R.C. Nunes, A. Bernui, Eur. Phys. J. C **80**, 1 (2020)
38. R.C. Nunes, S.K. Yadav, J. Jesus, A. Bernui, Mon. Not. R. Astron. Soc. **497**, 2133 (2020)
39. P. Peebles, (2001). arXiv preprint arXiv:astro-ph/0103040
40. W.J. Percival et al., Mon. Not. R. Astron. Soc. **327**, 1297 (2001)
41. C. Ramírez-Pérez, J. Sanchez, D. Alonso, A. Font-Ribera, J. Cosmol. Astropart. Phys. **2022**, 002 (2022)
42. B. Reid et al., Mon. Not. R. Astron. Soc. **455**, 1553 (2016)
43. A.G. Riess, S. Casertano, W. Yuan, L.M. Macri, D. Scolnic, Astrophys. J. **876**, 85 (2019)
44. S.A. Rodríguez-Torres et al., Mon. Not. R. Astron. Soc. **460**, 1173 (2016)
45. A.C. Rosell et al., Mon. Not. R. Astron. Soc. **509**, 778 (2022)
46. N.P. Ross et al., Mon. Not. R. Astron. Soc. **381**, 573 (2007)
47. E. Sánchez et al., Mon. Not. R. Astron. Soc. **411**, 277 (2011)
48. E. Sánchez, D. Alonso, F. Sánchez, J. García-Bellido, I. Sevilla, Mon. Not. R. Astron. Soc. **434**, 2008 (2013)
49. R. Scaramella et al., Astron. Astrophys. **662**, A112 (2022)
50. R. Scranton et al., Astrophys. J. **579**, 48 (2002)
51. H.-J. Seo, E.R. Siegel, D.J. Eisenstein, M. White, Astrophys. J. **686**, 13 (2008)
52. M. Sinha, L.H. Garrison, [MNRAS], **491**, 3022 (2020). https://doi.org/10.1093/mnras/stz3157
53. Y. Wang et al., Mon. Not. R. Astron. Soc. **498**, 3470 (2020)
54. K.C. Wong et al., Mon. Not. R. Astron. Soc. **498**, 1420 (2020)
55. H.S. Xavier, F.B. Abdalla, B. Joachimi, Mon. Not. R. Astron. Soc. **459**, 3693 (2016)
56. X. Xu, N. Padmanabhan, D.J. Eisenstein, K.T. Mehta, A.J. Cuesta, Mon. Not. R. Astron. Soc. **427**, 2146 (2012)
57. I. Zehavi et al., Astrophys. J. **571**, 172 (2002)
58. I. Zehavi et al., Astrophys. J. **608**, 16 (2004)
59. Y.B. Zel'Dovich, Astron. Astrophys. **5**, 84 (1970)
60. C. Zhao et al., Mon. Not. R. Astron. Soc. **503**, 1149 (2021)
61. R. Zhou et al., Astron. J. **165**, 58 (2023)