Clustering with general photo-z uncertainties: application to Baryon Acoustic Oscillations

Kwan Chuen Chan 1,2, Ismael Ferrero 2,3,4,5, Santiago Avila 2,6,7, Ashley J. Ross 8, Martin Crocce 2,4,5 and Enrique Gaztañaga 4,5

1School of Physics and Astronomy, Sun Yat-sen University, 2 Daxue Road, Tanguia, Zhuhai 519082, China
2CSST Science Center for the Guangdong-Hongkong-Macau Greater Bay Area SYSU, Zhuhai 519082, China
3Institute of Theoretical Astrophysics, University of Oslo, PO Box 1029 Blindern, NO-0315 Oslo, Norway
4Institut d’Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain
5Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain
6Instituto de Física Teórica UAM/CSIC, Universidad Autonoma de Madrid, E-28049 Madrid, Spain
7Departamento de Física Teórica, Universidad Autónoma de Madrid, E-28049 Madrid, Spain
8Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA

Accepted 2022 February 4. Received 2022 January 31; in original form 2021 October 25

ABSTRACT
Photometric data can be analysed using the 3D correlation function $\xi_p$ to extract cosmological information via e.g. measurement of the Baryon Acoustic Oscillations (BAO). Previous studies modeled $\xi_p$ assuming a Gaussian photo-z approximation. In this work we improve the modeling by incorporating realistic photo-z distribution. We show that the position of the BAO scale in $\xi_p$ is determined by the photo-z distribution and the Jacobian of the transformation. The latter diverges at the transverse scale of the separation $s_\perp$, and it explains why $\xi_p$ traces the underlying correlation function at $s_\perp$, rather than $s$, when the photo-z uncertainty $\sigma_z(1+z) \gtrsim 0.02$. We also obtain the Gaussian covariance for $\xi_p$. Due to photo-z mixing, the covariance of $\xi_p$ shows strong off-diagonal elements. The high correlation of the data causes some issues to the data fitting. None the less, we find that either can be solved by suppressing the largest eigenvalues of the covariance or it is not directly related to the BAO. We test our BAO fitting pipeline using a set of mock catalogs. The data set is dedicated for Dark Energy Survey Year 3 (DES Y3) BAO analyses and includes realistic photo-z distributions. The theory template is in good agreement with mock measurement. Based on the DES Y3 mocks, $\xi_p$ statistic is forecast to constrain the BAO shift parameter $\sigma$ to be $1.001 \pm 0.023$, which is well consistent with the corresponding constraint derived from the angular correlation function measurements. Thus, $\xi_p$ offers a competitive alternative for the photometric data analyses.

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

1 INTRODUCTION
Imaging surveys infer the redshift of galaxy samples, photo-z, using broadband filters. Ongoing and future large-scale structure surveys that collect enormous amount of photo-z data include Kilo-Degree Survey (KiDS),1 Dark Energy Survey (DES),2 Hyper Suprime-Cam (HSC),3 Rubin Observatory’s Legacy Survey of Space and Time (LSST),4 Euclid,5 and the Chinese Survey Space Telescope (CSST).6 These surveys typically use several bandpasses, e.g. DES has grizY and the accuracy of the resultant photo-z is about $\sigma \sim 0.03(1+z)$ [see e.g. Crocce et al. (2019), Carnero Rosell et al. (2021), but it can be improved if more stringent criteria are applied (Rozo et al. 2016)]. Although photometric surveys yield less precise redshift than spectroscopic ones ($\sigma \sim 10^{-3}$), they are more efficient to collect a large volume of survey data with deep magnitude.

Baryonic Acoustic Oscillations (BAO; Peebles & Yu 1970; Sunyaev & Zeldovich 1978) are the primordial acoustic features imprinted in the distribution of the large-scale structure. In the early universe, photons and baryons form a tightly coupled plasma, and acoustic oscillations are excited in it. After the recombination, hydrogen atoms form, the acoustic waves get stalled because there is no medium for propagation. The acoustic patterns are preserved in the distribution of galaxies and they correspond to the sound horizon at the drag epoch, which is about 150 Mpc. Formation of the BAO is governed by 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 its robustness, BAO has been regarded as a standard ruler in cosmology, see e.g. Weinberg et al. (2013) and Aubourg et al. (2015). BAO has been measured in numerous spectroscopic data analyses. It was first clearly detected by Eisenstein et al. (2005) and Cole et al. (2005), and followed up

© 2022 The Author(s)
Published by Oxford University Press on behalf of Royal Astronomical Society
in many subsequent studies, e.g. Gaztanaga, Cabre & Hui (2009), Percival et al. (2010), Beutler et al. (2011), Blake et al. (2012), Anderson et al. (2012), Kazin et al. (2014), Ross et al. (2015), and Alam et al. (2017, 2021).

The sample size in terms of volume and depth from photometric surveys may compensate its imperfection in redshift information and give competitive measurements (Seo & Eisenstein 2003; Blake & Bridle 2005; Amendola, Quercellini & Giallongo 2005; Benitez et al. 2009; Zhan, Knox & Tyson 2009; Chaves-Montero, Angulo & Hernández-Monteagudo 2018). Indeed, there are several detections of the BAO feature in the galaxy distribution using photometric data from SDSS (Padmanabhan et al. 2007; Estrada, Sefusatti & Frieman 2009; Hütsi 2010; Carnero et al. 2012; Seo et al. 2012; de Simoni et al. 2013), DES (Y1 (Abbott et al. 2019) and Y3 (Abbott et al. 2022), hereafter DES Y3), and DECaLS (Sridhar et al. 2020).

Because of the photo-z uncertainty, it is natural to perform the analyses of the photometric data using angular statistics such as the angular correlation function (Coupon et al. 2012; Crocce et al. 2016; Abbott et al. 2019; Elvin-Poole et al. 2018; Valki et al. 2020) or the angular power spectrum (Padmanabhan et al. 2007; Seo et al. 2012; Camacho et al. 2019; Nicola et al. 2020). In particular, these two statistics were adopted as the fiducial statistical tools to measure BAO in DES Y3 analyses. However, given enormous amount photo-z data available, it is worthwhile exploring new methods to enhance information extraction, especially from the BAO. Ross et al. (2017) proposed to measure the BAO using the 3D correlation function characterized by the transverse scale of the separation. This statistic has been applied to photo-z data to get interesting BAO measurements (Abbott et al. 2019; Sridhar et al. 2020). The photo-z uncertainty of the sample must be folded in the theory template. To simplify the modelling, the photo-z distribution was assumed to be Gaussian in Ross et al. (2017). The actual photo-z distribution has non-negligible deviation from Gaussian distribution, e.g. Crocce et al. (2019), Zhou et al. (2021), and Carnero Rosell et al. (2021). This assumption may risk introducing bias, and so this statistic was not adopted in the fiducial DES Y1 and Y3 analyses. In this work, we improve the modelling of the 3D clustering by incorporating general photo-z distribution and present a Gaussian covariance for it.

This paper is organized as follows. In Section 2, we first review the computation of the angular correlation and then show that the general angular correlation can be turned into the 3D correlation function $\xi_p$. We consider the Gaussian photo-z limit and use it to understand the effect of photo-z uncertainties on $\xi_p$ in Section 3. In Section 4, we derive the Gaussian covariance for $\xi_p$ using the Gaussian covariance for the angular correlation function. In Section 5, we present the numerical results on the template and covariance, and compare them with mock results whenever possible. We discuss the issue in data fitting caused by highly correlated data in Section 5.2.1, and then present the BAO fit results in Sections 5.2.2. We conclude in Section 6. In Appendix A, we test our pipeline for the case of spectroscopic data. The default cosmology adopted in this work is the MICE cosmology (Crocce et al. 2015; Fosalba et al. 2015), which is a flat $\Lambda$CDM with $\Omega_m = 0.25$, $\Omega_{\Lambda} = 0.75$, $h = 0.7$, and $\sigma_8 = 0.8$.

2 ANGULAR AND THREE-DIMENSIONAL TWO-POINT CORRELATION

In this section, we first review the computation of the angular correlation function, $w$, in the presence of general photo-z uncertainty. We then use the general angular correlation function to derive the 3D two-point correlation $\xi_p$.

2.1 Angular correlation function $w_p$

Because in imaging surveys, angles are well determined but there are typically significant redshift uncertainties, it is natural to use angular quantities. It is useful to first clarify the relation between volume number density and tomographic bin angular number density.

Suppose that there are $dN$ objects in a volume element parameterized by the redshift interval $dz$ and the solid angle $d\Omega$, the volume number density $n(z)$ is given by

$$n(z) = \frac{dN}{dz d\Omega}.$$  \hfill (1)

The tomographic angular number density $n_{2}(z)$ is defined as

$$n_{2}(z, \hat{r}) = \frac{dN}{d\Omega} = n(z) dz.$$  \hfill (2)

We can go on to define the volume density contrast $\delta(z)$ and angular density contrast $\delta_{2}(z, \hat{r})$

$$\delta(z) = \frac{n(z) - \bar{n}(z)}{\bar{n}(z)},$$  \hfill (3)

$$\delta_{2}(z, \hat{r}) = \frac{n_{2}(z, \hat{r}) - \bar{n}_{2}(z)}{\bar{n}_{2}(z)},$$  \hfill (4)

where $\bar{n}(z)$ and $\bar{n}_{2}(z)$ are their respective mean values. From equation (2), it is clear that $\delta(z) = \delta_{2}(z, \hat{r})$.

We start with the effect of the photo-z on the galaxy number density:

$$n_p(z_p, \hat{r}) = \int d\Omega f(z|z_p)n_{2}(z, \hat{r}),$$  \hfill (5)

where $f(z|z_p)dz$ is the probability that the true redshift of the galaxy falls within $[z, z + dz]$ given that its photo-z is $z_p$, and $n_p$ denotes the number density in photo-z space. For photometric data, $f(z|z_p)$ is of central importance and it can be estimated by the photo-z codes using template fitting or machine learning methods [see e.g. Benitez (2000); Ilbert et al. (2006); De Vicente, Sánchez & Sevilla-Noarbe (2016); Sadeh, Abdalla & Lahav (2016)].

At the background level, we have

$$\bar{n}_p(z_p) = \int d\Omega f(z|z_p)\bar{n}(z),$$  \hfill (6)

where $\bar{n}_p$ and $\bar{n}(z)$ are the angular mean number density in the photo-z and true redshift space, respectively. The perturbation is given by

$$n_p(z_p, \hat{r}) = \bar{n}_p(z_p)[1 + \delta_p(z_p, \hat{r})] = \int d\Omega f(z|z_p)n_{2}(z)[1 + \delta_{2}(z, \hat{r})],$$  \hfill (7)

and we get

$$\delta_p(z_p, \hat{r}) = \int d\Omega f(z|z_p)\frac{\bar{n}_{2}(z)}{\bar{n}_p(z_p)}\delta_{2}(z, \hat{r}).$$  \hfill (8)

Therefore, the appropriate true redshift distribution for the density contrast is

$$\phi(z|z_p) = f(z|z_p)\frac{\bar{n}_p(z_p)}{\bar{n}_p(z_p)}.$$  \hfill (9)

It is easy to see that $\phi(z|z_p)$ indeed furnishes a probability density. Here, we assume that there are sufficiently many particles in the coarse-grained scale of interest so that the photo-z effect can be modelled by the distribution $\phi$. In this regime, the effect of the photo-z is to average $\delta_{2}$ over $z$. 

MNRAS 511, 3965–3982 (2022)
Using equation (9), the photo-z angular correlation function $w_p$ can be written as
\[
w(\theta, z_p, z'_p) \equiv \langle h(\theta, \psi) h(\theta', \psi') \rangle = \int dz \phi(z|z_p) \frac{d}{dz} \phi(z'|z'_p) \langle \delta_{\Omega}(z, \theta) \delta_{\Omega}(z', \theta') \rangle,
\]
with $\theta \equiv \cos^{-1}(\hat{r} \cdot \hat{r}')$. By expressing $\delta_{\Omega}$ in terms of its Fourier modes, we have
\[
w(\theta, z_p, z'_p) = \int dz \phi(z|z_p) \frac{d}{dz} \phi(z'|z'_p) \int \frac{d^3k}{(2\pi)^3} \times e^{i(k(\hat{r} - \hat{r}'))} P(k, z, z').
\]
Utilizing the variable $s = r - r'$ and applying the Rayleigh expansion of the plane wave
\[
e^{iks} = \sum_{l,m=-\infty}^{\infty} \sum_{\ell} j_l(kr) Y_{lm}(\hat{k}) Y^*_{lm}(\hat{k}),
\]
we have
\[
w(\theta, z_p, z'_p) = \int dz \phi(z|z_p) \frac{d}{dz} \phi(z'|z'_p) \int \frac{d^3k}{(2\pi)^3} \times \sum_{l,m} 4\pi \ell |j_l(kr) Y_{lm}(\hat{k})|^2 P(k, z, z').
\]
Note that we have only one integral involving the spherical Bessel function because we expand the plane wave for $s$.

Equation (13) accounts for the photo-z uncertainty exactly and it is valid for the curved sky. We only need to plug in the appropriate power spectrum $P(k, z, z')$. However, the redshift-space power spectrum in the distant observer limit is often used, and so this effectively limits the results to the flat sky case. As long as the local line of sight (LOS) is used, the effect is mild, e.g. Yoo & Seljak (2015).

In general, because of the azimuthal symmetry about the LOS direction $\hat{e}$, the power spectrum can be expanded about the LOS direction as
\[
P(k, z, z') = \sum_{\text{even}} P_l(k, z, z') \mathcal{L}_l(\hat{k} \cdot \hat{e}),
\]
where $\mathcal{L}_l$ is the Legendre polynomial and $P_l$ is the multipole coefficient. Parity invariance further restricts the sum to even $\ell$ only. Then the angular correlation can be written as
\[
w(\theta, z_p, z'_p) = \sum_{\ell} \ell^2 \int dz \phi(z|z_p) \frac{d}{dz} \phi(z'|z'_p) \mathcal{L}_l(\hat{k} \cdot \hat{e})
\]
\[\times \int \frac{d^3k}{(2\pi)^3} j_l(kr) P_l(k, z, z'),
\]
where $j_l$ is the spherical Bessel function. To speed up the computation, the $j_l$ integrals are evaluated using FFTLog (Hamilton 2000).

For the linear redshift-space power spectrum with linear bias (Kaiser 1987)
\[
P(k, z, z') = [b + f(\hat{k} \cdot \hat{e})^2][b' + f'(\hat{k} \cdot \hat{e})^2]
\]
\[\times \langle D(z) D(z') \rangle P_m(k, 0, 0),
\]
where $b$ is the linear bias, $f$ is $dnD/dlna$ with $D$ being the linear growth factor, and $P_m(k, 0, 0)$ is the matter power spectrum evaluated at $z = 0$; we have (Hamilton 1992; Cole, Fisher & Weinberg 1994; Crocce, Cabré & Gaztañaga 2011)
\[
w(\theta, z_p, z'_p) = \int dz \phi(z|z_p) D(z) \int dz' \phi(z'|z'_p) D(z') \times \sum_{\ell=0,2,4} i^\ell A_\ell \mathcal{L}_\ell(\hat{s} \cdot \hat{e}) \int \frac{d^2k}{2\pi^2} J_\ell(ks) P_m(k, 0, 0),
\]
with
\[
A_\ell(z_p, z'_p) = \begin{cases} \frac{1}{3} (b'' + b')^2 + \frac{1}{3} f f' & \text{for } \ell = 0, \\
\frac{1}{3} (b'' + b')^2 + \frac{1}{3} f f' & \text{for } \ell = 2, \\
\frac{8}{3} f f' & \text{for } \ell = 4. \end{cases}
\]
In DES Y3, an anisotropic BAO damping factor is applied to the BAO feature, i.e.
\[
P(k, z, z') = [b + f(\hat{k} \cdot \hat{e})^2][b' + f'(\hat{k} \cdot \hat{e})^2] \times \langle D(z) D(z') \rangle \times \{P_{\text{lin}} - P_{\text{mw}} e^{-\sqrt{2} \xi(z)} \} + P_{\text{mw}}.
\]
where $P_{\text{lin}}$ and $P_{\text{mw}}$ denotes the linear power spectrum with and without BAO wiggles, respectively. Because $\xi$ is a function of $\mu$, we can no longer express the first three multipoles analytically, instead we compute them numerically. As our primary goal here is to model the BAO feature, we shall use equation (19). But we will use equations (17) and (18) to measure the bias parameters.

### 2.2 From $w$ to $\xi_p$

Because of the statistical rotational invariance, the angular correlation function $w(\theta, z, z')$ for the sample at the effective redshift $z_{\text{eff}}$ can be parametrized by the angle $\theta$ subtended by two points on the sky at the observer. If in addition, the redshifts (or photo-z) of these two points are also specified, e.g. in the form of tomographic bins, $w(\theta, z, z')$ furnishes a representation of the 3D correlation function. On the other hand, the 3D correlation function $\xi$ in real space depends only on the distance between the two points, $s = |r_1 - r_2|$, thanks to the statistical translational and rotational invariance. The redshift space distortion, along the LOS direction, breaks the rotational invariance in redshift space and $\xi$ is usually parametrized as a function of $s$ and $\mu$, the latter is the dot product between direction of the separation, $\hat{s}$ and the LOS $\hat{e}$, or in terms of $s_1$ and $s_2$, the component of $s$ perpendicular and parallel to the LOS. In practice, both $w$ and $\xi$ can be computed by assigning density to a grid (2D pixelized grid cells for $w$ and 3D grid cells for $\xi$) or by comparing pair counts between the data catalog and the random one.

Because $\delta_2 = \delta_3$, it is clear that the general cross angular correlation function and the 3D correlation are equal provided a cosmology is assumed to convert angle and redshift to distance and the redshift bin is fine enough so that the projection effect along the LOS direction is negligible. Depending on how it is parametrized, the correlation function goes by different names. We shall consider the 3D correlation function $\xi_p(s, \mu) = \langle \delta_b(r_1) \delta_b(r_2) \rangle$ and its projective version characterized by the transverse scale, $\xi_p(s_1, \mu) \equiv \langle \delta_b(r_1) \delta_b(r_2) \rangle_{||}$, where the notation (||) means that we average over $s_1 = s \sqrt{1 - \mu^2}$. We obtain $\xi_p(s, \mu)$ by mapping from $w$. We first evaluate the general cross correlation between redshift bin $i$ and $j$ separated by angle $\theta_{ij}$ bin, $w_{ij}(\theta_{ij})$. From its redshift bins and angular separation, $s$ and $\mu$ for this pair can be obtained. This allows us to allocate this angular correlation into the appropriate $s$ and $\mu$ bins. The relation between $w$ and $\xi_p$ is schematically shown on the left-hand panel of
Fig. 1. Left-hand panel: the relation between the angular correlation function specified by two true redshift distributions conditional on photo-z \( z_p \) and \( z_p' \), subtending an angle \( \theta \) at the observer. The corresponding 3D correlation function is described by the separation vector \( s \) relative to the LOS direction \( \hat{e} \). Right-hand panel: The integral over the two distributions can be approximated by a 1D integral over \( \Delta z_\ell \) or the separation \( s' \).

We have checked that these two definitions are essentially identical for our purpose. For example, at \( z_{\text{eff}} = 0.8 \) and \( s = 100 \text{ Mpc} \, h^{-1} \), the agreement between these two definitions are better than \( 10^{-3} \). Thus, we can use them interchangeably.

3 UNDERSTANDING THE IMPACT OF PHOTO-z BY A GAUSSIAN CASE STUDY

Ross et al. (2017) used a simple Gaussian setup to study the impact of photo-z uncertainty on the BAO information. The apparatus can be constructed by the following arguments. Suppose that the position vectors of two points are \( r_1 \) and \( r_2 \) and due to photo-z error, they have additional displacement \( \Delta r_1 \) and \( \Delta r_2 \). The total separation vector is

\[
s_p = (r_2 - r_1) + (\Delta r_2 - \Delta r_1) = s_1 + \Delta r_2 - \Delta r_1 + s_\perp,
\]

where \( s_\perp \) and \( s_1 \) denote the perpendicular and parallel components of \( r_2 - r_1 \) w.r.t. the LOS direction without photo-z error. The photo-z error only affects the direction parallel to the LOS in the plane-parallel limit. If we assume that \( \Delta r_1 \) and \( \Delta r_2 \) are independent Gaussian random variables with zero mean and constant variance \( \sigma^2 \), then \( \Delta r_2 - \Delta r_1 \) is Gaussian distributed with zero mean and variance \( 2\sigma^2 \). Thus, we can model the effects of the Gaussian photo-z error on the 3D correlation function by a 1D Gaussian distribution.

It is instructive to check under what conditions our general expression would reduce to this result. We will see that this exercise gives valuable insight into our results.

We now assume that the conditional true redshift distributions in equation (10) are Gaussian, namely

\[
\phi(z | z_p) = P_0(z - z_p; 0, \sigma^2) \equiv \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(z - z_p)^2}{2\sigma^2}}.
\]

By adopting the variables

\[
z = z_p + \Delta z, \quad z' = z_p' + \Delta z',
\]

\( w_p \) can be written as

\[
w_p(\theta, z_p, z_p') = \frac{1}{2\pi\sigma^2} \int \frac{d\Delta z}{\Delta z^2} e^{-\frac{\Delta z^2}{2\sigma^2}} \int \frac{d\Delta z'}{\Delta z'} e^{-\frac{\Delta z'^2}{2\sigma^2}} \times w(\theta, z, z_p, z_p' + \Delta z').
\]

After further changing to the ‘Centre-of-Mass’ variables

\[
\Delta z_c = \frac{\Delta z + \Delta z'}{2}, \quad \Delta z_t = \Delta z - \Delta z',
\]

we have

\[
w_p(\theta, z_p, z_p') \approx \frac{1}{2\pi\sigma^2} \int d\Delta z_t \int d\Delta z_i \exp \left(-\frac{\Delta z_i^2}{\sigma^2} - \frac{\Delta z_t^2}{4\sigma^2}\right) \times w(\theta, \Delta z_c + \Delta z_t, \Delta z_p' + \frac{\Delta z_t}{2}).
\]

To proceed, we neglect the \( \Delta z_t \) dependence in \( w \) as it gives an overall shift of the pair in redshift\(^7\) and we arrive at

\[
w_p(\theta, z_p, z_p') \approx \int d\Delta z_t \frac{1}{\sqrt{4\pi\sigma^2}} \exp \left(-\frac{\Delta z_t^2}{4\sigma^2}\right) \times w(\theta, z_p + \frac{\Delta z_i}{2}, z_p' + \frac{\Delta z_t}{2}).
\]

\(^7\)The standard 3D correlation analysis neglects the redshift evolution within the redshift bin.
By mapping to $s$ and $\mu$, we can get the corresponding equation for $\xi_p$, i.e.,

$$\xi_p(s(\theta, z_p, \zeta_p'), \mu(\theta, z_p, \zeta_p')) \approx \int d\Delta z_\xi P_G(\Delta z_\xi; 0, 2\sigma^2)$$

$$\times \xi \left( \frac{\Delta z_\xi}{\zeta_p}, \frac{\Delta z_\xi}{\zeta_p'}, \mu(\theta, z_p + \frac{\Delta z_\xi}{2}, \zeta_p') - \frac{\Delta z_\xi}{2} \right).$$  

(30)

This equation invites the interpretation on the right-hand panel of Fig. 1.

Alternatively, we can use the separation distance $s'$ (shown on the right-hand panel of Fig. 1) as the integration variable

$$\xi_p(s(\theta, z_p, \zeta_p'), \mu(\theta, z_p, \zeta_p')) \approx \int ds' \left( \frac{ds'}{d\Delta z_\xi} \right)^{-1} P_G(\Delta z_\xi(s'); 0, 2\sigma^2)\xi \left( s', \mu' \right).$$  

(31)

We can think of $s'$ as a function of $\Delta z_\xi$. When $\Delta z_\xi$ increases from the negative end, $s'$ first decreases and it reaches the minimum value $s' = s_{\perp}$ before starting to increase. The scale $s' = s_{\perp}$ divides the integral into two parts, and they should be evaluated separately.

The inverse of the Jacobian is given by

$$\frac{ds'}{d\Delta z_\xi} = \frac{e}{2s'} \left[ \frac{r(z)}{H(z)} - \frac{r(z)'}{H(z)} - \cos \theta \left( \frac{r(z)'}{H(z)} - \frac{r(z)}{H(z)} \right) \right],$$  

(32)

where $r(z)$ and $H(z)$ respectively represent the comoving distance and Hubble rate evaluated at redshift $z$, $e$ is the light speed, and $z_{\perp}$ denotes $z_p = \frac{\Delta z_\xi}{\zeta_p} (\zeta_p' + \frac{\Delta z_\xi}{2})$. It is important to note that this derivative vanishes when $z_{\perp} = z_{\perp}$, which corresponds to $s' = s_{\perp}$.

Equation (30) reveals that the photo-$z$ correlation function is a projection of the underlying true correlation with a Gaussian weight of zero mean and variance $2\sigma^2$, in agreement with equation (6) of Ross et al. (2017). In Fig. 2, we show the wedge correlation function obtained with the Gaussian photo-$z$ distribution. The effective redshift of the two conditional true redshift distributions centred at $z_p$ and $z_p'$, respectively, is taken to be $z_{\text{eff}} = 0.8$. The standard deviation of the distribution is assumed to be $\sigma = (1 + z_{\text{eff}})h^{-1}$ and a suite of $\sigma$ values are considered: 0.001, 0.01, 0.02, and 0.04. In this plot, for simplicity the linear real-space correlation is used. We have compared the results computed with the approximation equation (31) and by directly integrating over the 2D integral. We find that the 1D approximation is in good agreement with the exact result, but deviations become apparent for high $\sigma_z$.

Interestingly, while for small $\sigma_z$ such as 0.001, the BAO bump peaks at the true sound horizon size, $s_{\perp}$, it shifts to larger values as $\sigma_z$ increases. Also for large $\sigma_z$, it is clear that the BAO bump moves to larger $s$ as $\mu$ increases. In the lower panels of Fig. 2, we show $\hat{\xi}_p$ as a function of the transverse scale $s_{\perp} = s \sqrt{1 - \mu^2}$. As observed in Ross et al. (2017), when $\hat{\xi}_p$ is plotted against $s_{\perp}$, the BAO position appears to line up at $s_{\perp} = s_{\perp}$ for $\sigma_z \geq 0.02$. This plot suggests that the BAO peak in the upper panel (the usual way to plot $\hat{\xi}_p$) is directly related to the scale $s_{\perp}$ for $\sigma_z \geq 0.02$.

To shed light on these results, we show in Fig. 3 the three factors of the integrand of equation (31): the absolute value of the Jacobian $|d\Delta z/ds'|$, the Gaussian weight $P_G$ and the underlying correlation function $\xi$. We shall shortly explain these terms in details. In this illustration, we take $s(z_p, \zeta_p') = 100 \text{Mpc} \, h^{-1}$.

For small $\sigma_z$ such as 0.001, which is close to the spectroscopic redshift case, the Gaussian weight is narrowly distributed about the scale $s(z_p, \zeta_p')$. The distribution $P_G$ as a function of $s'$ is in fact mildly tilted for small $\mu$ and it becomes more symmetric when $\mu$ increases. As we mentioned, equation (32) vanishes at $s' = s_{\perp}$, and hence the Jacobian diverges. Except for small $\mu$, such as $\mu \leq 0.1$, the divergence occurs at the scale, where $P_G$ is essentially zero and hence it is irrelevant. Therefore, in this case the BAO scale in $\xi_p$ is controlled by $P_G$, and it corresponds to the true sound horizon well.

As $\sigma_z$ increases, the $P_G$ distribution becomes more tilted. Among the cases shown, for $\sigma_z \geq 0.01$, $s'$ is clearly not a monotonic function of $\Delta z_\xi$, as evident from the multivalued $P_G$ curves. The scale where all the $\mu$-curves pass through corresponds to $s' = s$. The smallest $s'$ that each $\mu$-curve attains corresponds to the transverse scale $s_{\perp} = s \sqrt{1 - \mu^2}$. The Jacobian formally diverges at $s' = s_{\perp}$. For small $\mu$, as $s_{\perp}$ coincides with $s$ well, the Jacobian and the Gaussian weight both peak at $s' = s$, the value of the integral is essentially equal to $\xi(s)$.

For larger $\mu$, it depends on the value of $\sigma_z$. While the Jacobian is sharply peaked at $s' = s_{\perp}$ and $\xi$ is also rapidly decreasing with $s'$. For $\sigma_z \approx 0.01$, the Gaussian weight at $s' = s_{\perp}$ is suppressed relative to its maximum value at $s$. Thus, the integral is not well peaked at $s' = s_{\perp}$. The distribution is more tilted as $\sigma_z$ further increases and becomes almost horizontally laid down for $\sigma_z \geq 0.04$. Because the Gaussian weight suppression is much more mild for $\sigma_z \geq 0.02$, the integral is essentially given by $\xi(s_{\perp})$. This explains why the scale $s_{\perp}$ but not $s'$ is a good indicator of the underlying correlation function when $\sigma_z \geq 0.02$.

Although we have carried out the calculations explicitly using Gaussian conditional true redshift distribution, the lessons learned are more general. The approximation that the $z_{\perp}$ dependence can be integrated out and the final results can be approximated by a 1D projection integral should be valid for well-behaved unimodal distributions (modulo small random fluctuations). In particular, the fact that the underlying correlation is reflected by $s_{\perp}$ primarily stems from the divergence of the Jacobian, which is not dependent on the precise form of the distribution.

We mention that the divergent Jacobian plays a similar role as the Limber approximation (Limber 1953; LoVerde & Afshordi 2008), which is phrased as a linear relation between the spherical wavenumber $\ell$ and the Fourier mode $k$ and is effective in reducing the dimension of the projection integrals. In this approximation, although only the Fourier modes transverse to the LOS are included, it works very well for reasonably large $\ell$ because those are the only significant contributions to the integral.

As $\hat{\xi}_p$ corresponds to the underlying correlation function at the transverse scale $s_{\perp}$ for $\sigma_z \geq 0.02$, we can stack all the modes of $\hat{\xi}_p$ with the same $s_{\perp}$ together to increase the signal to noise. Thus in this sense, the 3D $\hat{\xi}_p$ characterized by $s_{\perp}$ is effectively a projected correlation function. Consistent with Seo & Eisenstein (2003), Blake & Bridle (2005), $\xi_p(s_{\perp})$ for $\sigma_z \geq 0.02$ only measures the transverse BAO and cannot be used to probe the radial BAO and hence the Hubble parameter directly.

4 GAUSSIAN COVARIANCE

In this section, after briefly reviewing the Gaussian covariance for the angular correlation function, we then employ it to derive the Gaussian covariance for $\xi_p$. 

MNRAS 511, 3965–3982 (2022)
In the case of complete sky coverage, $\delta_p$ in equation (9) can be expressed in terms of the spherical harmonics as

$$\delta_p(z_p, r) = \sum_{\ell m} a_{\ell m}(z_p) Y_{\ell m}(\hat{r}).$$

(33)

We can define an estimator for angular correlation function:

$$\hat{w}_{ij} \left( \cos^{-1}(\hat{r} \cdot \hat{r}') \right) = \sum_{\ell m} \sum_{\ell' m'} Y_{\ell m}(\hat{r}) Y_{\ell' m'}(\hat{r}') \times a_{\ell m}(z_p) a_{\ell' m'}(z_p).$$

(34)

Figure 2. The wedge correlation function computed with Gaussian photo-z distributions. The upper and low panels show the results for using the separation $s$ and the transverse scale $s_\perp$ as the independent variables, respectively. The 1D approximation equation (31; solid) and the exact 2D integral results (dashed) agree with each other well, but small deviation occurs for large value of $\sigma_z$. Four values of $\sigma_z$ are displayed (left-hand to right-hand panels). Different colours indicate wedge correlation of different values of $\mu$.

Figure 3. The factors in the photo-z projection integral (equation 31) as a function of the separation $s'$. As shown on the right-hand panel of Fig. 1, as $\Delta z_r$ increases, $s'$ first decreases until it reaches the minimum value $s' = s_\perp$ and then it increases. Both the Gaussian weight $P_G$ (solid, colour) and absolute value of the Jacobian $|d\Delta z_r/ds'|$ (multiplied by a factor of 10, dotted, colour) correspond to the paths traversed as $\Delta z_r$ increases. The point where all the $P_G$ curves with different $\mu$ meet corresponds to $s' = s_{\perp}(z_p, \zeta_p')$ (chosen to be 100 Mpc $h^{-1}$ here). The Jacobian diverges at $s' = s_\perp$, and this indicates as the vertical line in the plot. The vertical part for the Jacobian touches the $P_G$ curve as the latter turns at $s' = s_\perp$. The underlying correlation function (multiplied by a factor 10$^3$, solid black) is also overplotted. See the text for further explanations.
Its expectation value gives (Peebles 1980)

\[ w_{ij}(\theta) = \sum_{\ell} \frac{2\ell + 1}{4\pi} L_{\ell}(\cos \theta) C^{ij}_{\ell}, \]  

(35)

where \( C^{ij}_{\ell} \) is the cross angular power spectrum defined as

\[ \langle a_{lm}(z_p) a_{lm'}^*(z_{p'}) \rangle = \left( C^{ij}_{\ell} + \delta^{ij}_{\ell} \right) \delta_{k}^{m} \delta_{k'}^{m'}, \]  

(36)

with \( \delta^{ij}_{\ell} \) being the Kronecker delta. Because galaxies are discrete tracers, we have included the shot noise contribution with \( \bar{n}_i \) being the tomographic angular number density given by

\[ \bar{n}_i = \int_{\text{Bin}_i} d^3x \bar{n}_p(z_p). \]  

(37)

Similarly, we can derive its Gaussian covariance (Crocco et al. 2011)

\[ \text{Cov}[\hat{w}_{ij}(\theta), \hat{w}_{lm}(\theta')] = \sum_{\ell} \frac{(2\ell + 1)}{(4\pi)^2} P_l(\cos \theta) P_{l'}(\cos \theta') \]  

\[ \times \left[ \left( C^{ij}_{\ell} + \delta^{ij}_{\ell} \right) \left( C^{lm}_{\ell} + \delta^{lm}_{\ell} \right) + \left( C^{ij}_{\ell} + \delta^{ij}_{\ell} \right) \left( C^{lm}_{\ell} + \delta^{lm}_{\ell} \right) \right]. \]  

(38)

Poisson shot noise is assumed in this equation. To obtain the covariance for partial sky coverage, we can apply the so-called \( f_{\text{sky}} \) approximation, i.e. the covariance scaling inversely with the fraction of sky coverage \( f_{\text{sky}} \) (Scott, Srednicki & White 1994).

A few comments are in order. We evaluate the cross harmonic power spectra \( C^{ij}_{\ell} \) using \textsc{camb sources} (Lewis & Challinor 2007). To account for the fact that the data vector is binned into angular bins, we need to use the bin-averaged Legendre polynomial (Salazar-Albornoz et al. 2017). In addition, as the terms are computed numerically, we need to evaluate some of the shot noise terms analytically to ensure numerical convergence (Chan et al. 2018).

Using equation (20), we can then write the covariance of \( \xi_p \) in terms of that of \( w \):

\[ \text{Cov}[\xi_p(x, \mu), \xi_p(x', \mu')] = \langle \xi_p(x, \mu), \xi_p(x', \mu') \rangle - \langle \xi_p(x, \mu) \rangle \langle \xi_p(x', \mu') \rangle \]  

\[ = \sum_{ijk} \sum_{lmn} f_{ijk} f_{lmn} \text{Cov}(w_{ij}(\bar{z}, \Delta z, \theta_k), w_{lm}(\bar{z}, \Delta z', \theta'_m)). \]  

(39)

This enables us to map the Gaussian covariance for \( w \) to that for \( \xi_p \). Similarly, we can get the covariance for \( \xi_p(z, z') \) using the corresponding weight.

It is worth discussing more about the shot noise term. The shot noise arises from the self-correlation of the discrete particles. In the Poisson model, the shot noise contribution to the correlation is given by

\[ \frac{1}{n_p(z_p) n_p(z_{p})} \int d\mathbf{z} f(z_p) (n(z, \hat{F}) - n_p(z_p)) \delta_D(\mathbf{z} - \mathbf{z}_p) \delta_D(\hat{F} - \hat{F}') \]  

\[ = \frac{1}{n_p(z_p)} \delta_D(z_p - z'_p) \delta_D(\hat{F} - \hat{F}'), \]  

(40)

in which the two-point distribution is converted to a product of a one-point distribution and the Dirac delta distribution. Thus, the shot noise contribution to the 3D correlation function depends on the 3D number density \( n_p \). Although the true redshift distribution of the neighbouring bins may overlap with each other for photometric galaxies, there is no cross photo-\( z \) (or cross-bin in tomographic analysis) shot noise contribution because shot noise arises from self-correlation and each galaxy is assigned one and only one photo-\( z \).

Our method approximates the 3D correlation and the shot noise by the tomographic version. For the signal, when the tomographic bin size is small compared to photo-\( z \) uncertainty, the results are convergent. For the shot noise, we approximate the Dirac delta in redshift in equation (40) by a Kronecker delta and the volume density by the tomographic one.

## 5 MOCK TESTS

In this section, we first test the theory template and the Gaussian covariance for \( \xi_p \) against mock catalogs in Section 5.1 and then test the BAO fit pipeline, as well as present the fit results in details in Section 5.2. To do so, we use the ICE-COLA mocks (Ferrero et al. 2021), which is a dedicated mock catalog for the DES Y3 BAO analyses. We briefly describe it here and refer readers to Ferrero et al. (2021) for more details.

The ICE-COLA mocks are derived from the COLA simulations which is based on the COLA method (Tassev, Zaldarriaga & Eisenstein 2013) and implemented by the ICE-COLA code (Izard, Crocco & Fosalba 2016). The COLA method combines the second-order Lagrangian perturbation theory with the particle-mesh simulation technique to ensure that the large-scale modes remain accurate when coarse simulation time steps are used. The simulation consists of 2048³ particles in a cube of side length 1536 Mpc \( h^{-1} \) matching to the mass resolution of MICE Grand Challenge N-body simulations (Crocco et al. 2015; Fosalba et al. 2015). To generate the whole sky light-cone simulation at \( z \sim 1.4 \), the simulation box is replicated three times in each Cartesian direction (altogether 64 copies). The replication has important consequences for the covariance measurements, and we will further comment on it later on. Mock galaxies are assigned to the ICE-COLA haloes using a hybrid Halo Occupation Distribution and Halo Abundance Matching recipe, following a similar approach in Avila et al. (2018). The mocks are calibrated with the redshift distribution and the angular clustering amplitude of the true galaxy samples. From the full-sky light-cone mocks, four maps with the DES Y3 footprint are extracted. Thus, there are 1952 mock catalogs in total, although we will only use a subset of them.

A key challenge is to assign realistic photo-\( z \)-s to the mock galaxies. To do so, the photo-\( z \) distribution is estimated using 8362 galaxies with known spectroscopic redshift from VIPERS survey (Guzzo et al. 2014). Using these galaxies, which have both the spectroscopic redshift \( z_s \) and the photo-\( z \) from DES, the 2D distribution \( p(z_s, z_p) \) can be estimated. By sampling \( p(z_s, z_p) \) conditioned on the true redshift \( z_s \), the candidate mock galaxies are put to the appropriate photometric bins. A further requirement is that the mock galaxies must match the abundance of the observed photometric galaxies. Suppose there are \( N(z_p) \) galaxies in the photometric bin \( z_p \), the most luminous \( N(z_p) p(z_s, z_p) \) galaxies from the bin \( (z_p, z_q) \) is taken as the mock sample. By construction, the mock galaxies will match the observed galaxy abundance and the conditional true redshift distribution.

As the cosmology adopted by the mock catalog is the MICE cosmology, the theory templates, i.e. the theoretical model described in Section 2.2, and the Gaussian covariance are computed in this cosmology, although we will also consider the template computed in Planck cosmology (Planck Collaboration 2020).

### 5.1 Template and covariance

In this section, we use the ICE-COLA mocks to test the theories developed in Sections 2 and 4. The theory calculations require
the conditional true redshift distribution and the bias parameters measured from mocks. Thus we shall first present the measurement of these quantities before confronting the theories with measurements. In Appendix A, we check the template calculation using the spectroscopic correlation function.

The spread of the conditional true redshift distribution generally increases with redshift. As \( \phi \) depends on the number density in true redshift and the spectroscopic data is not available outside \((0.6, 1.1)\), we cannot reliably calibrate \( \phi \) for all bins. Here, we approximate \( \phi \) by \( f \). The preliminary estimate suggests that the effect is small, but when the spectroscopic density is rapidly changing such as for the high redshift bins, it should be taken into account. We shall present more detailed analysis in future work. In the standard tomographic analysis, the samples are divided into a number of redshift bins and their auto (and cross) correlation functions are considered. For example, in the DES Y3 BAO analyses, the samples in the photo-z range \((0.6, 1.1)\) are divided into five tomographic bins with bin width \( \Delta z = 0.1 \). For 3D correlation function, because \( \phi \) is continuously sampled by the galaxy pair counts, we have to use fine photo-z bins in order to sample the \( \phi \) accurately and smoothly.

To investigate the number of photo-z bin width required, we consider a number of \( \phi \)'s measured from the ICE-COLA mocks in the photo-z range \((0.6, 1.1)\). We divide this photo-z range into redshift bins of equal spacing. In Fig. 4, we show a sample of \( \phi \) obtained with 5 and 50 photo-z bins. We also overplot the corresponding Gaussian approximation with the same mean and variance as the true \( \phi \). The Gaussian approximation is more accurate for the five-bin case as it is averaged over a larger redshift range. At high redshift, \( \phi \) is more irregular and the Gaussian approximation is less accurate.

The photo-z spread is often characterized by

\[
\sigma_z = \frac{\sigma_\phi}{1 + z}
\]

(41)

where \( \sigma_\phi \) is the standard deviation of \( \phi \). We measure \( \sigma_z \) from \( \phi \) for different number of photo-z bins and the results are displayed in Fig. 5. While \( \sigma_z \) is overestimated in the 5-bin case, the 20-bin results are consistent with the ones from the 50-bin case except that the 50-bin results are more noisy. This suggests that when the bin width \( \Delta z = 0.025 \lesssim \sigma_z \), the photo-z bin is fine enough to probe the intrinsic photo-z uncertainty. We emphasize that \( \sigma_z \) is used for sanity checks only, we shall use the full distribution in the test of the analysis.

Another ingredient required is the bias parameters. We measure the linear bias parameters by fitting the auto-angular correlation (equations 17 and 18) to the mock measurements as in Chan et al. (2018) and Abbott et al. (2019). The fit is performed by fitting to the mean of mocks in the angular range of \((0.5, 1.5)\) degree. We show the bias parameters obtained with different number of tomographic bins in Fig. 6. Similar to the case of \( \sigma_z \), the 20-bin and 50-bin cases show consistent result with the 50-bin ones showing more random fluctuations, while the five-bin results are slightly higher.

We compare the theory template against the mock measurements in Fig. 7. The 3D correlation \( \xi_3 \) is measured using the Landy-Szalay estimator (Landy & Szalay 1993) via the public code CUTE (Alonso 2012). Limited by the computing power, we only measure...
the correlation function from 384 ICE-COLA mocks. Judging from the size of the error bars, which are estimated by the standard error of the mean, the fidelity of the results are well sufficient for our purpose here. We estimate the correlation by considering the pairs with maximum radial separation 120 Mpc h\(^{-1}\) and transverse separation 175 Mpc h\(^{-1}\). To ensure good sampling of the projected correlation within the maximum radial separation, we only consider \(\xi(s)\) up to \(\mu = 0.8\). We further divide the \(\mu\) range (0, 0.8) into three equal spacing 'wedges'. Both \(\xi_P\) versus \(s\) and \(s_L\) are shown.

For the theory template, we first compute \(w_P(\theta)\) in 50 photo-z bins with angular bin width of 0.02 degree. Because the bin width in the radial direction is wide (\(\Delta r \sim 20\) Mpc h\(^{-1}\)), we linearly interpolate the resultant angular correlation to a larger redshift grid of 500 bins (\(\Delta z = 0.005\)) before binning all the pairs \(w_P(\theta_i)\) into the appropriate \(\xi(s, \mu)\) bin. It is important, especially for the projected correlation function, to ensure that the maximum ranges in the radial and transverse directions used in template construction match those in the measurements. We find that the template obtained is of slightly lower amplitude than the mock measurement (by about 3 per cent), the amplitude deficit is probably due to sparse sampling in the radial direction. The template shown in Fig. 7 has been multiplied by an overall amplitude factor. Besides this amplitude adjustment, the theory templates are in good agreement with mock results.

In the conventional angular tomographic analysis, the BAO features manifest in different angular scales, which is less direct to see the photo-z effect on the BAO. To take advantage of the fact that the BAO information is condensed to a single bump in \(\xi_P\), we plot in Fig. 8 the photometric correlation \(\xi_P\) against \(s_L\) and the spectroscopic \(\xi\) against \(s\). We have shown the results for both MICE and Planck cosmology. The spectroscopic correlation \(\xi\) is the dark matter real space correlation with isotropic BAO damping, obtained by an inverse Fourier transform of an analog of equation (19) at

\[ z_{e0} = 0.83. \]

We also show the sound horizon \(r_s\) in these two cases for comparison. Relative to the spectroscopic correlation, the BAO feature in \(\xi_P\) is less strong. Note that the BAO peak does not precisely correspond to \(r_{s0}\), and it is on different sides relative to the peak for the spectroscopic and the photometric cases. A small bias could arise because of the subtle shape differences between these two cosmologies. For example, the broadband terms introduced to take care of the shape difference, if their contribution is rapidly changing, it may slightly shift the BAO position. This is more likely to occur in the photometric case given the features are less sharp than in the spectroscopic one.

We now turn to the covariance. In Fig. 9, we show the correlation coefficient matrix for \(\xi_P\)

\[ \rho_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}, \]

where \(C_{ij}\) is an element of the covariance matrix. Both the results for \(s\) and \(s_L\) are shown. In this plot, we have used the Gaussian covariance. We have considered a single \(\mu\) bin in the range of (0, 0.8). These plots demonstrate that the covariance of \(\xi_P\) has strong off-diagonal elements, especially for \(s\). Real space results are known to be more correlated than the Fourier space ones. We argue that the situation for \(\xi_P\) is exacerbated by the photo-z mixing.

It is instructive to contrast the results with the Gaussian covariance for the angular correlation function, which is plotted in Fig. 10. The setup of the calculations are the same as those in the fiducial DES Y3 analyses. There are five photo-z bins, and in each bin, \(\theta\) ranges from 0.5 to 5 degree with angular bin width of 0.2 degree. The photo-z mixes modes in the radial direction. Because the width of the redshift bin (\(\Delta z = 0.1\)) is wide compared to the photo-z uncertainty \(\sigma [\sim 0.03(1 + z)]\), the photo-z mixing does not affect the covariance in the same bin much. The cross-correlation between different redshift bins is mainly caused by photo-z mixing, and the only significant correlation is between neighbouring redshift bins. Again, because the redshift bin is wide compared to \(\sigma\), the mixing between the neighbouring bins is mild, and so even the neighbouring bin covariance is weak. On the other hand, as photo-z mixing scale

Figure 7. The wedge correlation function \(\xi_P\) as a function of the separation \(s\) (upper panel) and the transverse separation \(s_L\) (lower panel). The \(\mu\) range in (0, 0.8) is divided into three equal spacing bin, (0,0.27; blue), (0.27,0.53; red), and (0.53,0.8; green). The measurements (filled circles) are the mock mean and the error bars estimated by the standard error are too small to be seen. The theory templates are the solid curves.

Figure 8. The photometric correlation function \(\xi_P\) versus \(s_L\) (dotted line with filled circles) and the dark matter real space spectroscopic correlation \(\xi\) against \(s\) (solid) in MICE (blue) and Planck (red) cosmology, respectively. The sound horizons at the drag epoch in these cosmologies are also overplotted as dashed vertical lines (in the same colour code). To enhance the BAO feature, they have been multiplied by the appropriate distance squared.
As we mentioned, construction of the ICE-COLA mock catalogs involves replicating the simulation box to form a full-sky light-cone mock. The repeated structures cause the covariance to be anomalously higher than the correct one, and this prevents us from using the mock covariance as the benchmark to test our Gaussian theory covariance. See Ferrero et al. (2021) for more discussions. Thus in DES Y3, the analytic Gaussian covariance is the default choice. Although we cannot directly test the $\xi_p$ covariance against mock results, the code used to compute the Gaussian covariance for $w$, which is the backbone of the $\xi_p$ results, has been tested in Chan et al. (2018), Avila et al. (2018), and Abbott et al. (2022). Moreover, we have verified that the amplitude ratio difference between the Gaussian theory covariance and the mock covariance are at a similar level for $w$ and $\xi_p$, albeit quantitative comparison is not possible.

5.2 BAO fit

In this section, we apply our theory template and covariance to extract the BAO scale from the mock catalogs. With the variable $s_\perp$, the BAO scale is almost equal to the true one, we adopt $\xi_\perp$ as a function of $s_\perp$ in the BAO fit. To measure the BAO scale, as in standard BAO analyses [e.g. Xu et al. (2012), Anderson et al. (2014), Ross et al. (2017)], we apply the following general model

$$M(s_\perp) = B T(\alpha s_\perp) + \sum_i \frac{A_i}{s_\perp^i}$$

where $T$ denotes the theory template and $B$ gives the freedom to adjust the amplitude. The parameter $\alpha$ in the argument of $T$ is the key parameter we are after. It allows us to shift the BAO scale computed in the fiducial cosmology to match that in the data. The polynomial in $1/s_\perp$ is designed to absorb residual broadband effects due to imperfectness in modelling, changes in correlation function shape as a result of the differences between the fiducial cosmology and the actual one, and residual broadband systematic effects.

The default configuration for the BAO fit is as follows: the broadband terms consisting of $\sum_i A_i/s_\perp^i$ with $i = 0, 1, 2$; the fit range $[40, 140]$ Mpc $h^{-1}$; the bin width $\Delta s = 3$ Mpc $h^{-1}$, and the Gaussian theory covariance. We use 384 mocks for the BAO fit test. Different configurations for BAO fit are checked.

We look for the best-fitting $\alpha$ by the maximum likelihood estimator following the procedures in Chan et al. (2018). We first analytically fit the linear parameter $A_i$, and then minimize the resultant $\chi^2$ w.r.t. $B$ subject to the constraint $B > 0$. At the end of the operations, we are left with the residual $\chi^2$ as a function of $\alpha$. The best-fitting $\alpha$ corresponds to the point where the residual $\chi^2$ attains its minimum and the symmetric 1σ error bar is derived from the $\Delta \chi^2 = 1$ criterion. Besides, we also cross check our results with the Markov Chain Monte Carlo (MCMC) method, in which all the parameters are fitted simultaneously.

5.2.1 Highly correlated data fit

The photo-z mixing causes $\xi_p$ to be strongly correlated, as evident in Fig. 9. Here, we first describe some fitting problems caused by the high correlation, namely the apparently bad fit results and the alarmingly large $\chi^2$ per degree of freedom ($\chi^2$/dof). The former issue is caused by the largest eigenvalues of the covariance and the latter is due to the small ones. We discuss a few approaches to modify the covariance, and find that the first issue can be resolved by suppressing the largest eigenvalues. For the second issue, while it is not practical to solve it by adjusting the eigenvalues, we demonstrate
The BAO fit results to the mean of the mock measurements (grey data points with error bars). The fit obtained with the MICE template without broadband terms (solid, red) is compared with the Planck template without broadband terms result (solid, green). The latter case is below all the data points. With the inclusion of the broadband terms, the Planck template (dashed, yellow) also offers good fit. The results for the modified covariance are also shown [No broadband terms (solid, green) and with broadband terms (dashed, violet)].

that the BAO feature is well-fitted by our model and the problem is not directly related to the BAO.

To begin with, we present some peculiar fit results. One of the odd examples appears in the BAO fit to the mean of the mocks, which are shown in Fig. 11. When the MICE cosmology template is used, the resultant fit is very good and it is not surprising that the \( \chi^2/\text{dof} \) is very small (0.009). However, when the template computed in Planck cosmology (Planck Collaboration 2020) is used, we find that the resultant \( \chi^2/\text{dof} \) is still small (0.11), but the fit is manifestly poor by eye because the best-fitting model is below all the data points and most of the data points are more than 1\( \sigma \) away from the best fit.

Although this issue only appears when the model equation (43) does not include the broadband terms at all and disappears otherwise, it does indicate that something abnormal happens to the fit.

In the other extreme, when we fit to the individual mocks, we find that the fit results are often fine by eye as the best-fitting model fall within all the 1\( \sigma \) errors, but \( \chi^2/\text{dof} \) is alarmingly large, some of them reach as high as 2. Note that the second issue we describe here appears for generic fit configurations, in particular for both Gaussian and mock covariance fit. We point out that the second problem is already implicit in Abbott et al. (2019) and Sridhar et al. (2020), as their \( \xi_p \) fit results look decent by eye but appreciably large \( \chi^2/\text{dof} \) were reported. We attribute these problems to the high correlation of the data induced by the photo-z mixing.

Although the covariance of \( \xi_p \) has strong off-diagonal elements, its inverse, the precision matrix \( \Psi \) could be approximated by a band matrix with a few non-vanishing elements near the diagonal. The upper panel of Fig. 12 shows the elements of \( \Psi \), and it is approximately a pentadiagonal band matrix, i.e. a band matrix has non-zero elements only for the main diagonal, and the first two upper and two lower diagonals. We can use this to intuitively understand why the best fit may not be driven to the data points as close as

possible. By assuming \( \Psi \) to be pentadiagonal, the \( \chi^2 \) can be written as

\[
\chi^2 = \sum_{j=1}^{N} \left( \Delta_{j-2}\Psi_{j-2,j} + \Delta_{j-1}\Psi_{j-1,j} + \Delta_{j}\Psi_{j,j} \\
+ \Delta_{j+1}\Psi_{j+1,j} + \Delta_{j+2}\Psi_{j+2,j} \right) \Delta_j,
\]

where \( \Delta \) denotes the difference between the model and the data. This expression can accommodate the results near the ‘boundaries’ as well if \( \Psi_{ij,j} \equiv 0 \) whenever \( i, j < 1 \) or \( > N \). Now suppose \( \Delta > 0 \), because \( \Psi_{ij,j} > 0 \), to minimize \( \chi^2 \), \( \Delta_j \) is attracted to zero. For diagonal covariance, this is the sole contribution. However, because \( \Psi_{j-2,j}, \Psi_{j-1,j}, \Psi_{j+1,j} \), and \( \Psi_{j+2,j} \) are negative, it is favourable for \( \Delta_{j-2}, \Delta_{j-1}, \Delta_{j+1}, \) and \( \Delta_{j+2} \) to be repelled to as negative as possible. Of course these expressions are coupled, more careful analysis is required, but it is likely that the competition between the diagonal terms and the off-diagonal ones lead to a best fit that does not precisely fall on the data.

To look into the covariance, we plot its eigenvalues in Fig. 13. We find that the spectrum spans three orders of magnitude, and most of them are of relatively small values. As a comparison we also show the diagonal elements of \( \Psi \), its range is much more limited. For the fit to the mean of mocks using the Planck template, we find that the fit is dominated by the largest eigenvalue mode. Because of its (excessively) large eigenvalue, its contribution to the \( \chi^2 \) is very small, so that even the fit is visually bad, \( \chi^2/\text{dof} \) is still small. On the other hand, for the fit to individual mocks, the large \( \chi^2/\text{dof} \) values are caused by significant contributions from modes with small eigenvalues.

The problem of fitting highly correlated data is also encountered in the lattice field theory because lattice simulation data are highly
correlated. Yoon et al. (2013) surveyed a few methods proposed by the lattice field theory community to tackle this problem. Michael (1994) advocated ignoring the off-diagonal elements in the covariance, but this is clearly a too violent modification. A more mild method is to remove the eigenmodes deemed problematic. Both removing the largest eigenvalue modes (Thacker & Lepage 1991) and the smallest ones (Drummond & Horgan 1993; Kilcup 1994) have been proposed. For the largest eigenvalue modes removal, it was argued that these modes have small contribution to $\chi^2$. The smallest eigenvalue modes removal is often favoured in the spirit of the singular value decomposition (Press et al. 2007). A physical argument in support of it is that those modes are generally highly oscillatory and so they are not conducive to getting a smooth fit. In a similar vein, Bernard et al. (2002) proposed to modify instead the correlation coefficient equation (42). They replace the eigenvalues of $\rho$ smaller than a certain threshold $\lambda_{\rho_{\text{th}}}$ by $\lambda_{\rho_{\text{th}}}$: A new covariance matrix can be constructed using the original eigenvectors and the modified eigenvalues. Modifying the eigenmodes is more attractive as it keeps all the directions in the covariance space constrained.

We test the results obtained with different covariance recipes. While for most of the fit, different approaches give similar results, in a small fraction (order of 10 percent) the differences are quite pronounced. In Fig. 14, we show the fit results from one of the mocks, in which the differences are clearly visible. The one obtained with the original covariance (blue) is manifestly not optimal. The best fit is $1.006 \pm 0.017$ in this case. In the other extreme, when the diagonal approximation (red) is used, it is not surprising that the best fit is driven close to data points, but the best fit we get is $1.002 \pm 0.032$, with the error bar substantially larger than the original one. By suppressing the largest eigenvalues of the covariance (green) or $\rho$ (yellow), we can also get an apparently better fit. We replace the eigenvalues larger than certain threshold $\lambda_{\text{th}}$ by $\lambda_{\text{th}}$. We take $\lambda_{\text{th}}$ to be $3 \times 10^{-8}$ and 1 for the covariance and correlation coefficient modification respectively. The thresholds are chosen so that as few eigenvalues are affected as possible and the affected ones account for about 10 percent of them. The best fits we get are $1.003 \pm 0.016$ and $1.005 \pm 0.017$, respectively.

In Fig. 12, we also show the elements of $\Psi$ obtained by suppressing the largest eigenvalues of $\rho$. Overall, the size of the modification is small, of the order $10^{-3}$. The main effect of the covariance modification is to shift $\Psi$ to the positive side so that the attraction to zero is enhanced and the repulsion to negative infinity is weakened. Note that the large eigenvalue modes correspond to eigenvectors that are oscillatory in nature with long ‘wavelength’, while the small eigenvalue ones tend to be well localized. Thus the modification also weakly affects the the zero far from the diagonal. Furthermore, we find that the $\rho$ modification recipe often results in better agreement with the original one compared to the covariance modification, hence we will adopt $\rho$ modification in the following analysis. In Fig. 11, we also show the results obtained with $\rho$ modification. Without the broadband terms, the best fit now passes through the data points, although the BAO position appears to be biased. With the default broadband terms ($A$, with $i = 0, 1, 2,$ and 3), the bias is removed.

On the other hand, the large $\chi^2$/dof problem is caused by the small eigenvalues. The nominal $\chi^2$/dof for the example shown in Fig. 14 are similar for different recipes, around 2.7, except for the diagonal approximation, which gives $\chi^2$/dof $\sim 0.06$. Since only the largest eigenvalues are modified, the $\chi^2$ is only increased slightly compared to the original one. As it is clear from the spectrum in Fig. 13, most of the eigenvalues are small and of similar values, we need to remove or modify a large fraction of eigenmodes before the effect becomes noticeable. Removing 40 percent of the eigenmodes with the smallest eigenvalues, the naive $\chi^2$/dof can be reduced to roughly 1. This represents a very substantial modification of the covariance matrix. Moreover, after eigenmodes removal or modification, there is no proper theoretical basis for $\chi^2$/dof anymore (Bernard et al. 2002). Thus, we shall not pursue to reduce $\chi^2$/dof by adjusting the covariance matrix.

Using the example in Fig. 14 again, we check directly if the BAO scale is well fitted by the model in Fig. 15. To decorrelate the data, we rotate the data vector by the matrix $U$, which is the orthogonal matrix that diagonalizes the covariance, i.e. $C = U^T A U$. To highlight the BAO feature in the rotated coordinates, the no-BAO template has been subtracted from both the data vector and the best-fitting model. Note that as the data vector is mixed by $U$, they no longer strictly correspond to the original distance $r$. Here, we only use it as a convenient label for the x-axis. None the less, from the difference
with the no-BAO template, the location of the BAO feature can be inferred. In this basis, the BAO feature shows up as a series of wiggles in $r \gtrsim 110 \text{ Mpc} \, h^{-1}$ instead of a single peak in the usual configuration representation. The error bars on the data are given by the eigenvalues in $\Lambda$ and they are not correlated in this basis. We have also shown error bars from the modified covariance according to the $\rho$ modification recipe. We see that the modification reduces the error bars around the BAO scale. To highlight the source of large $\chi^2$/dof, we show the ratio with respect to the error bar in the lower panel of Fig. 15. Both the results obtained with the original error bars and the modified ones are shown, which coincide with each other except for $r > 130 \text{ Mpc} \, h^{-1}$. Significant deviations of the best-fitting model from the data points occur for $r \lesssim 100 \text{ Mpc} \, h^{-1}$, while in the BAO zone, the model and the data agree with each other well for both types of error bars. Thus, we conclude that the large $\chi^2$/dof is caused by the part of the fit that is not directly related to the BAO.

5.2.2 Fit results

In this section, we present the BAO fit results in details. We will consider various metrics to quantify the accuracy of the fit and different variation of the fit configurations. Although these are important, they can be tedious to read. The key default result is given in the first line in Table 1. In particular, the $\xi_p$ constraint on $\alpha$ is $\langle \alpha \rangle \pm \sigma_{\alpha} = 1.001 \pm 0.023$. We have checked that the results are robust against variation of the fit configurations. The default setup uses the Gaussian covariance modified with the $\rho$ modification prescription. We shall also contrast them with the corresponding results for the angular correlation. They are also consistent with the $w$ results even though $\xi_p$ statistic has considerably larger $(\chi^2)$/dof.

The default case results for $\xi_p$ fit are shown in the first row in Table 1. As a reference, we also show the corresponding results for the angular correlation function, which are obtained from a joint five-tomographic fit. See Table II in Abbott et al. (2022) for more details on the $w$ fit. The mean of the best-fitting ($\langle \alpha \rangle$) is 1.001, with small bias which can be attributed to nonlinear evolution (Crocce & Scoccimarro 2008; Padmanabhan & White 2009). Two measures of the spread of the distribution of the best-fitting $\alpha$ are shown: the standard deviation $\sigma_{\alpha}$ and the half of the width between 16 and 84 percentile of the distribution, $\sigma_{\alpha,1}$, which is less sensitive to the tails of the distribution. We find that $\sigma_{\alpha}$ is larger than $\sigma_{\alpha,1}$ by a small amount (0.001), suggesting the influence by the tails. As mentioned, the 1σ error bar is derived from the $\Delta \chi^2 = 1$ condition, and for the error bars to be meaningful, $\langle \sigma_{\alpha} \rangle$ should agree with the measures of the spread of the distribution. In the case of Gaussian likelihood, they are equal to each other. We find that the mean of the error bar is slightly smaller than the spread of the distribution. These numbers are marginally better than those for $w$. The bias in the best-fitting $\langle \alpha \rangle$ is smaller (1.004 for $w$) and the measures of spread and error are smaller by $\sim 0.001$. These (mild) improvements could arise from the fact that the $\xi_p$ statistic makes use of the cross-correlation in the data as well, not only the tomographic bin auto-correlation for $w$ as in the study of DES Y3.

Another indicator for the accuracy of 1σ error bar is the fraction of times that it encloses $\langle \alpha \rangle$, which is 68% per cent in the case of Gaussian likelihood. In line with the error bar being slightly smaller than the spread, the fraction enclosing $\langle \alpha \rangle$ is slightly lower than the Gaussian expectation. The normalized deviation, $d_{\text{norm}} = (\alpha - \langle \alpha \rangle)/\sigma_{\alpha}$ is designed to study the correlation between the deviation of the best-fitting $\alpha$ from the ensemble mean and the error bar derived. The mean $\langle d_{\text{norm}} \rangle$ is close to zero, but the standard deviation $\sigma_{d_{\text{norm}}}$ exceeds unity by 18 per cent. This trend is consistent with the previous observation that the spread of the distribution is slightly larger than the error estimate. In contrast, for $w$, the fraction of time enclosing $\langle \alpha \rangle$ is 62 per cent, and $\langle d_{\text{norm}} \rangle = 0.02$ and $\sigma_{d_{\text{norm}}} = 1.12$. Both suggest that $\xi_p$ results are less Gaussian than the $w$ ones.

As a reference, we also show the fit to the mean measurement of the mocks. To show the results for a ‘typical’ fit, we have used the covariance corresponding to a single survey volume. The results are similar to $\langle \alpha \rangle \pm \sigma_{\alpha}$. However, we note that for $w$ these different measurements are almost always equal to each other, suggesting that the averaging and fitting operations are more commutative for $w$.

Finally, the $(\chi^2)$/dof $(\sim 2)$ is also shown for reference, but we should bear in mind the complications mentioned in the previous section. On the other hand, for $w$, $(\chi^2)$/dof is 1.05. We attribute this to the fact that the covariance for $w$ is much less correlated.

A number of tests on the variation of the fitting configurations have been performed. We have tested the number of the broadband terms $\sum_i A_i/s_i^2$ with $i = 0, 1, 0, 1,$ and $i = 0, 1, 2, -1$. Besides minor fluctuations, the fit results are not sensitive to the number of broadband terms used. This is because the broadband terms are not degenerate with $\alpha$, but they are degenerate among themselves. When a narrower fitting range $[70, 130] \text{ Mpc} \, h^{-1}$ is used, the constraining power is reduced and so the spread of the distribution is enlarged, especially for $\sigma_{\alpha,1}$. As the error bar $\sigma_{\alpha}$ increases only mildly, the difference
between the error bar and the spread of the distribution increases.

The bin width size $\Delta s = 10, 5, 2,$ and $1$ Mpc $h^{-1}$ are compared with the fiducial setting $\Delta s = 3$ Mpc $h^{-1}$. Among the cases shown, $\sigma_{\text{fid}}$ for $\Delta s = 3$ Mpc $h^{-1}$ is the lowest and it is also more consistent with $\sigma_{\text{fid}}$ and $\langle \sigma_\alpha \rangle$. Note that $\chi^2/\text{dof}$ are smaller for $\Delta s = 2$ and $1$ Mpc $h^{-1}$, but this is not vital here as we already mentioned that the BAO region is well-fitted by our model, the reduction in $\chi^2/\text{dof}$ is primarily due to the region lying outside the BAO.

To test if the expected shift in $\alpha$ can be recovered when the cosmology of the template is different from that of the mock, we generate a template computed with the Planck cosmology (Planck Collaboration 2020). We can estimate the expected $\alpha$ by

$$\alpha = \frac{\Delta M_{\text{Planck}}}{\Delta M_{\text{MICE}}} = 0.959,$$

where $\Delta r$ is the sound horizon at the drag epoch and $\Delta M$ is the comoving angular diameter distance evaluated at the effective redshift $= 0.835$ following DES Y3. For MICE cosmology, $\Delta r = 153.4$ Mpc and $\Delta M (\varpi_{\text{eff}}) = 2959.7$ Mpc, while for Planck cosmology $\Delta r = 147.6$ Mpc and $\Delta M (\varpi_{\text{eff}}) = 2967.0$ Mpc.

Unlike other tests, when an alternative fiducial cosmology is assumed, the measurement of $\xi_\mu$ should be performed again as the distances has to be computed in the new cosmology. Furthermore, we mentioned that the Hubble parameter $h$ is contained in the unit Mpc $h^{-1}$ and the distance does not explicitly depend on it.

In Table 2, we compare the values of the $\alpha$ parameter obtained by the simple estimate using equation (45) and the mock measurement results. In contrast, we also show similar results for $\mu$ from Abbott et al. (2022). Interestingly, we find that the $\xi_\mu$ results are closer to the simple estimate than the ones from $\mu$. The deviation from the simple estimate could stem from the nonlinear correction and other systematic effects, and they affect the results from both templates in a similar manner.

We have shown the results obtained with the original Gaussian covariance. The $\rho$ modification prescription only affects a small fraction of the mocks and the results are statistically similar to the original Gaussian case. Moreover, although the COLA covariance suffers from the overlap issue, we also show its results for reference.
could be underestimated as the likelihood is less smooth and the curvature around the minimum is increased.

In Table 1, we already compared the $\xi_p$ results in the photo-$z$ range (0.6,1.1) against the $w$ results obtained using the joint fit with five tomographic bins, each of bin width $\Delta z_p = 0.1$. It is instructive to compare the performance of $\xi_p$ and $w$ for individual tomographic bins, and the results are shown in Table 3. In this test, the intrinsic photo-$z$ distribution is comparable to the bin width. Unlike the joint-bin fit, there is no cross-bin information. The difference between these statistics is that the data are further projected onto a sphere for $w$. Because of the limitation in computing power, only 100 mocks are used in this comparison. Overall, the results for $w$ are more stable across redshift bins. Except for the first bin, the measures of the spread of the distribution are larger for $w$, and this can be interpreted as information loss due to projection. However, we note that for $\xi_p$, $\langle \sigma_w \rangle$ tends to be smaller than the measures of the spread by a larger amount. Besides, $\sigma_{\text{dum}}$ is also more significantly larger than 1. These are caused by the increase in noise as the data size reduces and are consistent with the trends for $N_{\text{min}} = 3$ and $N_e = 2$ in Table 1. The $\chi^2$/dof for $\xi_p$ is still substantially larger than 1 although it is smaller than the joint fit case, while for $w$, it remains close to 1. In summary, $\xi_p$ makes use of more information and hence potentially more constraining, but it is more susceptible to noise in the data.

Before closing this section, we would like to comment further on the dual roles of $\chi^2$: the characterization of the goodness of fit and the model differentiation. The $\chi^2$/dof is often used to indicate the goodness of the fit.\textsuperscript{9} We discussed previously that the $\chi^2$/dof is substantially larger than 1 although the model offers a good fit to the BAO. A separate question is if we can use the $\chi^2$ value for model differentiation, especially for claiming the significance of the BAO detection. Although we may not be able to use the $\chi^2$ distribution results, in principle we can use the mocks to answer the question that if there are no BAO signals in the data, how much the chance is that the BAO template yields a $\chi^2$ value smaller than certain threshold.

6 CONCLUSIONS

There are numerous large-scale structure photometric surveys ongoing or upcoming. One of the important probes to extract cosmological information from these surveys is via the measurement of the BAO. A common way to measure BAO in photometric data is to divide the sample into tomographic bins and measure their auto (and cross) correlation.

It has been suggested to analyse the photometric data as the spectroscopic ones through the 3D correlation function $\xi_p$. Previous modelling was limited to a Gaussian photo-$z$ approximation. In this work, we generalize the modelling to accommodate general photo-$z$ distribution. This eliminates the possibility of biasing the results by assuming Gaussian photo-$z$ approximation. To do so, we bin the general cross angular correlation between different redshift bins $w_{ij}(\theta_{ij})$ using the variables $s$ and $\mu$ (or $s_\perp$ and $\mu$) like in usual 3D correlation function analysis. Our approach highlights the connection between the angular correlation function and 3D correlation function.

To help understanding the impact of the photo-$z$ uncertainties on the correlation function, especially the BAO, we specialize the calculations to the Gaussian photo-$z$ case. When $\alpha \geq 0.2$, the BAO peak in the correlation shows up at the true BAO scale only if the transverse scale of the separation, $s_\perp$, is used as the independent variable. We show that this results from the interplay between the photo-$z$ distribution and the Jacobian of the transformation. The latter formally diverges at $s_\perp$ and for $\alpha \geq 0.2$, it dominates the integral, causing $\xi_p$ to trace the underlying correlation function at the scale $s_\perp$. Large-scale photometric surveys equipped with broadband filters, such as DES, typically have $\alpha \geq 0.2$. In this case, it is favourable to adopt the parametrization $\xi_p(s_\perp, \mu)$, which effectively projects the 3D correlation function to the transverse direction. Unfortunately, this also means that $\xi_p$ only probes the transverse BAO and cannot be used to directly constrain the Hubble parameter.

Our approach also enables us to derive the Gaussian covariance for $\xi_p$ from the Gaussian covariance for the angular correlation function $w$. Due to photo-$z$ mixing, the covariance of the $\xi_p$ statistics has strong off-diagonal elements. This high correlation causes problems to the data fitting. One of the problems is that some of the best-fitting results are manifestly poor, but we find that they can be resolved by suppressing the large eigenvalue modes in the covariance. Another issue is that the mean $\chi^2$/dof for the fit is large, even though the fit is apparently good. While this second issue cannot be easily resolved by adjusting the covariance, we find that the issue is primarily caused by the small eigenvalue modes and they are not directly related to the BAO scale. We conclude that the BAO scale is well-fitted by our model (Fig. 15).

We have used a set of dedicated DES Y3 mocks to test our pipeline. This set of mocks include realistic photo-$z$ distribution among other things. We find that the theory template is in good agreement with the mock measurement. After verifying our methodology, we apply our pipeline to the BAO fit in the mock catalog. Overall, we find that the results are consistent with the ones derived from $w$ even with small improvements because $\xi_p$ makes use of the cross redshift

\textsuperscript{9}Strictly speaking, the degrees of freedom is only defined for linear models as stressed in Andrae, Schulze-Hartung & Melchior (2010), while the parameter $\alpha$ (and $B$) appears nonlinearly. Here, we followed the usual practice to assume that the degrees of freedom can be approximated defined for our model.
bin correlation as well. However, the angular correlation function measurements are less correlated with $\chi^2/{\text{dof}} \sim 1$. As $\xi_p$ offers an alternative way to extract the BAO information, we shall apply it to the DES Y3 samples. This can serve as an important cross check as different statistics have different sensitivities to the potential systematics. Some preliminary results on $\xi_p$ have been presented in Abbott et al. (2022) on the BAO samples, but a detailed quantitative analysis is still lacking. This is particularly interesting, as the $D_M$ measurement in Abbott et al. (2022) shows stimulating deviation from the Planck results. Measuring $D_M$ using alternative statistics on various samples helps establish the results.

ACKNOWLEDGEMENTS

We thank Fengji Hou for useful discussions. KCC acknowledges the support from the National Science Foundation of China under the grant 11873102, the science research grants from the China Manned Space Project with NO. CMS-CSST-2021-B01, and the Science and Technology Program of Guangzhou, China (no. 202002030360). SA is supported by the MICUES project, funded by the EU H2020 Marie Sklodowska-Curie Actions grant no. 713366 (InterTalent UAM) and ‘EU-HORIZON-2020-776247 Enabling Weak Lensing Cosmology (EWC)’.

Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey.

The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l’Espai (IEEC/CSIC), the Institut de Física d’Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, NSF’s NOIRLab, the University of Nottingham, The Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, Texas A&M University, and the OzDES Membership Consortium.

Based in part on observations at Cerro Tololo Inter-American Observatory at NSF’s NOIRLab (NOIRLab Prop. ID 2012B-0001; PI: J. Frieman), which is managed by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

The DES data management system is supported by the National Science Foundation under grant numbers AST-1138766 and AST-1536171. The DES participants from Spanish institutions are partially supported by MICINN under grants ESP2017-89383, PGC2018-094773, PGC2018-102021, SEV-2016-0588, SEV-2016-0597, and MDM-2015-0509, some of which include ERDF funds from the European Union. IFAE is partially funded by the CERCA program of the Generalitat de Catalunya. Research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Program (FP7/2007-2013), including ERC grant agreements 240672, 291329, and 306478. We acknowledge support from the Brazilian Instituto Nacional de Ciência e Tecnologia (INCT) do e-Universo (CNpq grant 465376/2014-2).

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Abbott T. et al., 2019, MNRAS, 483, 4866
Abbott T. et al., 2022, Phys. Rev. D, 105, 043512
Alam S. et al., 2017, MNRAS, 470, 2617
Alam S. et al., 2021, Phys. Rev.D, 103, 083533
Alonso D., 2012, preprint (arXiv:1210.1833)
Amendola L., Quercellini C., Giallongo E., 2005, MNRAS, 357, 429
Anderson L. et al., 2012, MNRAS, 427, 3435
Anderson L. et al., 2014, MNRAS, 441, 244
Andrae R., Schulze-Hartung T., Melchior P., 2010, preprint (arXiv:1002.3754)
Aubourg É. et al., 2015, Phys. Rev. D, 92, 123516
Avila S. et al., 2018, MNRAS, 479, 94
Benitez N., 2000, ApJ, 536, 571
Benitez N. et al., 2009, ApJ, 691, 241
Bernard C. et al., 2002, Phys. Rev. D, 66, 094051
Beutler F. et al., 2011, MNRAS, 416, 3017
Blake C., Bridle S., 2005, MNRAS, 363, 1329
Blake C. et al., 2012, MNRAS, 425, 405
Bond J. R., Efstathiou G., 1984, ApJ, 285, L45
Bond J. R., Efstathiou G., 1987, MNRAS, 226, 655
Camacho H. et al., 2019, MNRAS, 487, 3870
Camero Rosell A. et al., 2021, MNRAS, 509, 778
Camero A., Sánchez E., Crocce M., Cabré A., Gaztañaga E., 2012, MNRAS, 419, 1689
Chan K. C. et al., 2018, MNRAS, 480, 3031
Chaves-Montero J., Angulo R. E., Hernández-Monteagudo C., 2018, MNRAS, 477, 3892
Cole S., Fisher K. B., Weinberg D. H., 1994, MNRAS, 267, 785
Cole S. et al., 2005, MNRAS, 362, 505
Coupon J. et al., 2012, A&A, 542, A5
Crocce M., Scoccimarro R., 2008, Phys. Rev. D, 77, 023533
Crocce M., Cabré A., Gaztañaga E., 2011, MNRAS, 414, 329
Crocce M., Castander F. J., Gaztañaga E., Fosalba P., Carretero J., 2015, MNRAS, 453, 1513
Crocce M. et al., 2016, MNRAS, 455, 4301
Crocce M. et al., 2019, MNRAS, 482, 2807
de Simoni F. et al., 2013, MNRAS, 435, 3017
De Vicente J., Sánchez E., Sevilla-Noarbe I., 2016, MNRAS, 459, 3078
Dodelson S., 2003, Modern Cosmology. Academic Press, Amsterdam
Drummond I. T., Horgan R. R., 1993, Phys. Lett. B, 302, 271
Eisenstein D. J. et al., 2005, ApJ, 633, 560
Elvin-Poole J. et al., 2018, Phys. Rev. D, 98, 042006
to verify our template method, we also compute the standard linear redshift-space 3D correlation function template directly, which can be obtained by replacing $\phi$ in equation (17) by a Dirac delta distribution. The real-space template can be retrieved by further setting the $f$-terms in equation (18) to zero. In Fig. A1, we show the real-space correlation function measurement and the corresponding best-fitting template with the bias as the fitting parameter. The real-space template is in good agreement with the measurements for $r \gtrsim 20 \text{ Mpc h}^{-1}$. Note that the BAO feature is stronger in the template as the linear power spectrum is used. The best-fitting effective bias is 1.63, which roughly corresponds to the bias value at the effective redshift 0.83 shown in Fig. 6 (or Fig. A2 below). Moreover, we compare the redshift-space correlation measurement with the corresponding template from the standard direct computation. The wedge correlation function is averaged over $\mu$ in the same way as $\xi_B$ in the main text. We find that the agreement is mildly worse than the real-space result, and the BAO feature damping appears to be weaker than in the real-space case. In the computation, we have used the best-fitting effective bias parameter from the real-space result so as to compare with the template obtained by mapping $w$ to $\xi$ below. We could have used the bias parameter obtained by fitting to the redshift-space correlation and the agreement on small scales would be slightly improved.

We also compute the spectroscopic template following the method of mapping $w$ to $\xi$ discussed in the main text. Similar to the analysis for the photo-z sample, we first measure the bias parameters of the redshift-space sample using the angular correlation function. In Fig. A2, we compare the bias parameters measured from the photo-z angular correlation function and the spectroscopic angular correlation. In both cases, we divide the data in the redshift range $(0.6,1.1)$ into 50 equal redshift bins. For the spectroscopic sample, the window function is a top-hat of width $\Delta z = 0.01$. We find that the recovered tomographic bin bias parameters are consistent with each other, and this demonstrates that fine photo-z bin measurements can indeed recover the underlying intrinsic bias.

We go on to use the bias parameters to compute the 3D correlation. For the case of spectroscopic redshift, we compute the cross angular correlation function $w_{ij}(\theta)$ with the the sampling width $\Delta z = 0.00333$ and $\Delta \theta = 0.02$. We find that finer sampling in $z$ does not improve the agreement.

**APPENDIX A: CHECKS ON THE SPECTROSCOPIC SAMPLE**

Although our goal is to consistently include photo-z in our modelling, it is helpful to check our methodology on the redshift-space correlation function.

To test our pipeline, we use the real-space and redshift-space correlation function measured from the ICE-COLA mocks. In order
the results. Note that in this case, we have used the linear power spectrum with BAO damping. The template is in decent agreement with the measurement, but in the intermediate range 30–80 Mpc $h^{-1}$ it is slightly worse than the direct computation result. The difference could be caused by the extra binning and averaging in the $w$ to $\xi$ mapping and these may result in smoothing of the correlation function. However, this is less of an issue in the photo-z case as the intrinsic photo-z mixing window is much stronger.

In Table A1, we show the BAO fit results for the spectroscopic samples. We have shown four scenarios with different template and covariance combinations. The template is either the theory one or the mock mean. The theory template is obtained by mapping $w$ to $\xi$. As there is small difference between the mock measurement and the theory, we also consider directly using the mean of the mock measurement as the template. In passing, in the main text, this possibility is not presented because we found that using the photo-z $\xi_p$ mock mean as the template leads to a non-smooth likelihood. Two covariances are considered, the mock covariance and the Gaussian one. The Gaussian covariance is computed in a similar way as the photo-z case with 50 bins in redshift.

When both the template and the covariance are from mocks, it could be the best case scenario. But we still find that the error estimate is smaller than the spread of the distribution and $d_{\text{norm}}$ deviates from Gaussianity significantly. This suggests that the likelihood is non-Gaussian. Moreover, the $\langle \chi^2 \rangle/\text{dof}$ is small relative to 1. This could be caused by issues of the mocks such as the overlapping issue discussed in the main text. When the mock covariance is replaced by the Gaussian one, $\langle \sigma_\alpha \rangle$ is further reduced and the fraction enclosing $\langle \alpha \rangle$ is lowered. Interestingly, $\langle \chi^2 \rangle/\text{dof}$ is closest to unity in this case. When the theory template is used with the mock covariance, the spread of the best-fitting $\alpha$ and $\langle \chi^2 \rangle/\text{dof}$ increase. These can be attributed to the difference between the template and the mock mean. Finally, when both the theory template and the Gaussian covariance are used, the $\langle \chi^2 \rangle/\text{dof}$ is the highest. We note that it is still smaller than the photo-z case by 0.5.

**Table A1.** The BAO fit results for the spectroscopic sample. For each of the case shown, the first entry denotes the type of template, mock mean, or theory and the second one represents the type of covariance, mock, or Gaussian covariance.

| Case         | $\langle \alpha \rangle$ | $\sigma_{\text{std}}$ | $\sigma_{68}$ | $\langle \sigma_\alpha \rangle$ | $\text{frac. encl.}(\alpha)$ | $\langle d_{\text{norm}} \rangle$ | $\sigma_{d_{\text{norm}}}$ | $\text{mean of mocks}$ | $\langle \chi^2 \rangle/\text{dof}$ |
|--------------|--------------------------|------------------------|---------------|-------------------------------|-------------------------------|---------------------------------|---------------------------|-------------------------|-------------------------------|
| Mock, Mock   | 1.000                    | 0.010                  | 0.010         | 0.007                         | 70%                           | -0.029                         | 1.60                      | 1.000 ± 0.009           | 24.0/29 (0.83)               |
| Mock, Gaussian| 1.000                   | 0.010                  | 0.011         | 0.005                         | 52%                           | -0.022                         | 1.43                      | 1.001 ± 0.011           | 30.9/29 (1.07)               |
| Theory, Mock | 1.001                   | 0.013                  | 0.014         | 0.011                         | 74%                           | -0.005                         | 1.25                      | 1.001 ± 0.012           | 36.6/29 (1.26)               |
| Theory, Gaussian | 1.002               | 0.014                  | 0.014         | 0.007                         | 44%                           | -0.023                         | 1.38                      | 1.002 ± 0.007           | 43.1/29 (1.49)               |

This paper has been typeset from a TeX/LaTeX file prepared by the author.