

Dark Energy Survey year 3 results: covariance modelling and its impact on parameter estimation and quality of fit

O. Friedrich,^{1,2★} F. Andrade-Oliveira,^{3,4} H. Camacho^{id},^{3,4} O. Alves,^{3,5} R. Rosenfeld,^{4,6} J. Sanchez,⁷ X. Fang^{id},⁸ T. F. Eifler^{id},^{8,9} E. Krause,⁸ C. Chang,^{10,11} Y. Omori,¹² A. Amon,¹² E. Baxter,¹³ J. Elvin-Poole,^{14,15} D. Huterer,⁵ A. Porredon,^{14,16,17} J. Prat,¹⁰ V. Terra,³ A. Troja,^{4,6} A. Alarcon,¹⁸ K. Bechtol,¹⁹ G. M. Bernstein,²⁰ R. Buchs,²¹ A. Campos,²² A. Carnero Rosell,^{23,24} M. Carrasco Kind,^{25,26} R. Cawthon,¹⁹ A. Choi,¹⁴ J. Cordero,²⁷ M. Crocce,^{16,17} C. Davis,¹² J. DeRose,^{28,29} H. T. Diehl,⁷ S. Dodelson,²² C. Doux,²⁰ A. Drlica-Wagner,^{7,10,11} F. Elsner,³⁰ S. Everett,²⁹ P. Fosalba,^{16,17} M. Gatti,³¹ G. Giannini,³¹ D. Gruen,^{12,21,32} R. A. Gruendl,^{25,26} I. Harrison,²⁷ W. G. Hartley,³³ B. Jain,²⁰ M. Jarvis,²⁰ N. MacCrann,³⁴ J. McCullough,¹² J. Muir,¹² J. Myles,³² S. Pandey,²⁰ M. Raveri,¹¹ A. Roodman,^{12,21} M. Rodriguez-Monroy,³⁵ E. S. Rykoff,^{12,21} S. Samuroff,²² C. Sánchez,²⁰ L. F. Secco,²⁰ I. Sevilla-Noarbe,³⁵ E. Sheldon,³⁶ M. A. Troxel,³⁷ N. Weaverdyck,⁵ B. Yanny,⁷ M. Aguena,^{4,38} S. Avila,³⁹ D. Bacon,⁴⁰ E. Bertin,^{41,42} S. Bhargava,⁴³ D. Brooks,³⁰ D. L. Burke,^{12,21} J. Carretero,³¹ M. Costanzi,^{44,45} L. N. da Costa,^{4,46} M. E. S. Pereira,⁵ J. De Vicente,³⁵ S. Desai,⁴⁷ A. E. Evrard,^{4,48} I. Ferrero,⁴⁹ J. Frieman,^{7,11} J. García-Bellido,³⁹ E. Gaztanaga,^{16,17} D. W. Gerdes,^{5,48} T. Giannantonio,^{1,50} J. Gschwend,^{4,46} G. Gutierrez,⁷ S. R. Hinton,⁵¹ D. L. Hollowood,²⁹ K. Honscheid,^{14,15} D. J. James,⁵² K. Kuehn,^{53,54} O. Lahav,³⁰ M. Lima,^{3,38} M. A. G. Maia,^{4,46} F. Menanteau,^{25,26} R. Miquel,^{31,55} R. Morgan,¹⁹ A. Palmese,^{7,11} F. Paz-Chinchón,^{26,50} A. A. Plazas,⁵⁶ E. Sanchez,³⁵ V. Scarpine,⁷ S. Serrano,^{16,17} M. Soares-Santos,⁵ M. Smith,⁵⁷ E. Suchyta,⁵⁸ G. Tarle,⁵ D. Thomas,⁴⁰ C. To,^{12,21,32} T. N. Varga,^{59,60} J. Weller,^{59,60} and R. D. Wilkinson⁴³ (DES Collaboration)

Affiliations are listed at the end of the paper

Accepted 2021 August 10. Received 2021 July 30; in original form 2020 December 23

ABSTRACT

We describe and test the fiducial covariance matrix model for the combined two-point function analysis of the Dark Energy Survey Year 3 (DES-Y3) data set. Using a variety of new ansatzes for covariance modelling and testing, we validate the assumptions and approximations of this model. These include the assumption of Gaussian likelihood, the trispectrum contribution to the covariance, the impact of evaluating the model at a wrong set of parameters, the impact of masking and survey geometry, deviations from Poissonian shot noise, galaxy weighting schemes, and other sub-dominant effects. We find that our covariance model is robust and that its approximations have little impact on goodness of fit and parameter estimation. The largest impact on best-fitting figure-of-merit arises from the so-called f_{sky} approximation for dealing with finite survey area, which on average increases the χ^2 between maximum posterior model and measurement by 3.7 per cent ($\Delta\chi^2 \approx 18.9$). Standard methods to go beyond this approximation fail for DES-Y3, but we derive an approximate scheme to deal with these features. For parameter estimation, our ignorance of the exact parameters at which to evaluate our covariance model causes the dominant effect. We find that it increases the scatter of maximum posterior values for Ω_m and σ_8 by about 3 per cent and for the dark energy equation-of-state parameter by about 5 per cent.

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

1 INTRODUCTION

Our understanding of the Universe has become much more accurate in the past decades due to a massive amount of observational data collected through different probes, such as the cosmic microwave background (CMB; see e.g. Planck Collaboration VI 2020), big bang nucleosynthesis

* E-mail: of259@ast.cam.ac.uk

(see e.g. Fields et al. 2020), Type Ia supernovae (see e.g. Riess 2017; Smith et al. 2020), number counts of clusters of galaxies (see e.g. Mantz et al. 2014; Costanzi et al. 2019; Abbott et al. 2020), the correlation of galaxy positions, and that of their measured shape (see e.g. Abbott et al. 2018; Heymans et al. 2020). From the study of that data, a standard cosmological model has emerged characterized by a small number of parameters (see e.g. Frieman, Turner & Huterer 2008; Peebles 2012; Blandford et al. 2020). Current spectroscopic and photometric surveys of galaxies such as the Extended Baryon Oscillation Spectroscopic Survey¹ and earlier phases of the Sloan Digital Sky Survey, the Hyper Suprime-Cam Subaru Strategic Program,² the Kilo-Degree Survey (KiDS³), and the Dark Energy Survey (DES⁴) have become instrumental in testing this standard model at a new front: the growth of density perturbations in the late-time Universe. Also, future surveys, such as the Dark Energy Spectroscopic Instrument,⁵ the Vera Rubin Observatory Legacy Survey of Space and Time,⁶ *Euclid*,⁷ and the *Nancy Grace Roman Space Telescope*,⁸ will push this test to a precision exceeding that provided by other cosmological probes.

An important part of this program is the DES, a state-of-the-art galaxy survey that completed its 6-yr observational campaign in 2019 January (Diehl et al. 2019) collecting data on position, colour, and shape for more than 300 million galaxies. This makes DES the most sensitive and comprehensive photometric galaxy survey ever performed. The main cosmological analyses of the first year (Y1) of DES data have been concluded (Abbott et al. 2018, 2019b) and analyses of the first 3 yr of data (Y3) are under way. The study of the large-scale structure (LSS) of the Universe based on the DES-Y3 data set has the potential to become the most stringent test of our understanding of cosmological physics to date.

To achieve this goal, the DES team is comparing different theoretical models characterized by a range of cosmological parameters to the measured statistics of the LSS in order to determine the model and range of parameters that are in best agreement with the data. The statistics of the LSS considered in the main DES-Y3 analysis are two-point correlation functions of the galaxy density field (galaxy clustering), the weak gravitational lensing field (cosmic shear), and the cross-correlation functions between these fields (galaxy–galaxy lensing) in real space and measured in different redshift bins. These three types of two-point correlation functions are combined into one data vector – the so-called 3×2 pt data vector.

A key ingredient in analysing these statistics is a model for the likelihood of a cosmological model given the measured correlation functions. Under the assumption of Gaussian statistical uncertainties (which is to be validated), this likelihood is completely characterized by the covariance matrix that describes how correlated the uncertainties of different data points in the 3×2 pt data vector are. Validating the quality of the covariance model for the DES-Y3 two-point analyses is the main focus of this paper.

There are several methods to estimate covariance matrices that can roughly be divided into four main categories: covariance estimation from the data itself (e.g. through jackknife or sub-sampling methods; cf. Norberg et al. 2009; Friedrich et al. 2016), covariance estimation from a suite of simulations (e.g. Hartlap, Simon & Schneider 2007; Dodelson & Schneider 2013; Taylor, Joachimi & Kitching 2013; Percival et al. 2014; Taylor & Joachimi 2014; Joachimi 2017; Sellentin & Heavens 2017; Avila et al. 2018; Shirasaki et al. 2019), theoretical covariance modelling (e.g. Schneider et al. 2002; Eifler, Schneider & Hartlap 2009; Krause et al. 2017), or hybrid methods combining both simulations and theoretical covariance models (e.g. Pope & Szapudi 2008; Friedrich & Eifler 2018; Hall & Taylor 2019).

For the DES-Y3 3×2 pt analysis, we adopt a theoretical covariance model as our fiducial covariance matrix. This fiducial covariance model is based on a halo model and includes a dominant Gaussian component, a non-Gaussian component (trispectrum and supersample covariance), redshift space distortions (RSDs), curved sky formalism, finite angular bin width, non-Limber computation for the clustering part, Gaussian shape noise, Poissonian shot noise, and f_{sky} approximation to treat the finite DES-Y3 survey footprint (although taking into account the exact survey geometry when computing sampling noise contributions to the covariance). In order to assess the accuracy of that model, we study the impact of several approximations and assumptions that go into it (and into two-point function covariance models in general):

- (i) the Gaussian likelihood assumption, i.e. whether knowledge of the covariance is sufficient to calculate the likelihood;
- (ii) robustness with respect to the modelling of the non-Gaussian covariance contributions, i.e. contributions from the trispectrum and supersample covariance;
- (iii) treatment of the fact that two-point functions are measured in finite angular bins;
- (iv) cosmology dependence of the covariance model;
- (v) random point shot noise;
- (vi) the assumption of Poissonian shot noise;
- (vii) survey geometry and the f_{sky} approximation;
- (viii) other covariance modelling details such as flat sky versus curved sky calculations, Limber approximation, and RSDs.

We generate different types of mock data and/or analytical estimates to determine how each of these effects impacts the quality of the fit between measurements of the 3×2 pt data vector and maximum posterior models (quantified by the distribution of χ^2 between the two).

¹www.sdss.org/surveys/eboss

²hsc.mtk.nao.ac.jp/ssp

³kids.strw.leidenuniv.nl

⁴www.darkenergysurvey.org

⁵www.desi.lbl.gov

⁶www.lsst.org

⁷www.euclid-ec.org

⁸nasa.gov/content/goddard/nancy-grace-roman-space-telescope

We also show how they impact cosmological parameter constraints derived from measurements of the 3×2 pt data vector. For most of these tests, we employ a linearized Gaussian likelihood framework that allows us to analytically quantify the impact of covariance errors on the χ^2 distribution and parameter constraints. This is complemented by a set of lognormal simulations and importance sampling techniques to quickly assess large numbers of mock (non-linear) likelihood analyses.

This paper is part of a larger release of scientific results from year-3 data of the DES and our analysis is informed by the (in some cases preliminary) analysis choices of the other DES-Y3 studies. In addition to carving out the most stringent constraints on cosmological parameters from late-time two-point statistics of galaxy density and cosmic shear yet, the year-3 analysis of the DES collaboration is introducing and testing numerous methodological innovations that pave the way for future experiments. Details of the DESY3 galaxy catalogues and the photometric estimation of their redshift distribution are presented by Sevilla-Noarbe et al. (2020), Hartley et al. (2020), Everett et al. (2020), Myles et al. (2021), Gatti et al. (2020a), Cawthon et al. (2020), Buchs et al. (2019), and Cordero et al. (2020). The measurements of galaxy shapes and the calibration of these measurements for the purpose of cosmic weak gravitational lensing analyses are detailed by Gatti et al. (2020b), Jarvis et al. (2020), and MacCrann et al. (2020). Krause et al. (2021) develop and test the theoretical modelling pipeline of the DES-Y3 3×2 pt analysis, Pandey et al. (2021) outline how galaxy bias is incorporated in this pipeline, DeRose et al. (2020) validate this pipeline with the help of simulated data, and Muir et al. (2020) describe how we have blinded our analysis to focus our efforts on model-independent validation criteria and reduce the chance for confirmation bias. The DESY3 methodology to sample high-dimensional likelihoods and to characterize external and internal tensions is outlined by Lemos et al. (2021) and Doux et al. (2020). Measurements of cosmic shear two-point correlation functions and analyses thereof are presented by Amon et al. (2021) and Secco et al. (2021), the measurement and analysis of galaxy clustering two-point statistics are carried out by Rodríguez-Monroy et al. (2021), and two-point cross-correlations between galaxy density and cosmic shear (galaxy–galaxy lensing) are measured and analysed by Prat et al. (2021), with additional analyses of lensing magnification and shear ratios carried out by Elvin-Poole et al., (in preparation) and Sánchez et al. (2021) and results for an alternative lens galaxy sample presented by Porredon et al. (2020) and Porredon et al., (2021). Finally, in DES Collaboration (2020) we present our cosmological analysis of the full 3×2 pt data vector.

Our paper is structured as follows. We start by presenting a discussion of our validation strategy in Section 2, where we also summarize our main findings before plunging into the details in the remaining of the paper. In Section 3, we review the modelling and structure of the 3×2 pt data vector. Section 4 describes our fiducial covariance model as well as two alternatives to it that are used to validate several modelling assumptions. In Section 5, we describe our linearized likelihood formalism and derive analytically how different covariance matrices impact parameter constraints and maximum posterior χ^2 within that formalism (including the presence of nuisance parameters and allowing for Gaussian priors on these parameters). In Section 6, we present the details of each step in our validation strategy followed by a short Section 7 presenting a simple test to corroborate some of the results from the linearized framework. We conclude with a discussion of our results in Section 8. Seven appendices describe in more detail some results used in this work.

2 COVARIANCE VALIDATION STRATEGY AND SUMMARY OF THE RESULTS

How should one validate the quality of a covariance model (and the associated likelihood model) for the purpose of constraining cosmological model parameters from a measured statistic? A straightforward answer seems to be that one should run a large number of accurate cosmological simulations, then measure and analyse the statistic at hand in each of the simulated data sets, and test whether the true parameters of the simulations are located within the, say, 68.3 per cent quantile of the inferred parameter constraints in 68.3 per cent of the simulations. There are, however, at least two problems with such an approach.

The first one is a conceptual problem. Consider a Bayesian analysis of a measured statistic $\hat{\xi}$ with a model $\xi[\boldsymbol{\pi}]$ that is parametrized by model parameters $\boldsymbol{\pi}$. For each value of $\boldsymbol{\pi}$, the statistical uncertainties in the measurement $\hat{\xi}$ will have some distribution

$$\mathcal{L}(\boldsymbol{\pi}|\hat{\xi}) \equiv p(\hat{\xi}|\boldsymbol{\pi}), \quad (1)$$

which is also called the likelihood of the parameters given the data. If this function is known, then a Bayesian analysis will assign a posterior probability distribution to the parameters as

$$p(\boldsymbol{\pi}|\hat{\xi}) = \frac{1}{\mathcal{N}} \mathcal{L}(\boldsymbol{\pi}|\hat{\xi}) \text{pr}(\boldsymbol{\pi}). \quad (2)$$

Here, $\text{pr}(\boldsymbol{\pi})$ is a prior probability distribution that parametrizes prior knowledge from other experiments (or theoretical constraints) and the normalization constant \mathcal{N} is fixed by demanding that $p(\boldsymbol{\pi}|\hat{\xi})$ be a probability distribution. The 68.3 per cent confidence region for the parameters $\boldsymbol{\pi}$ would then, e.g. be stated as a volume $V_{68.3 \text{ per cent}}$ in parameter space that contains 68.3 per cent of the probability. To unambiguously define that volume, one can e.g. impose the additional condition that

$$\min_{\boldsymbol{\pi} \in V_{68.3 \text{ per cent}}} p(\boldsymbol{\pi}|\hat{\xi}) \geq \max_{\boldsymbol{\pi} \notin V_{68.3 \text{ per cent}}} p(\boldsymbol{\pi}|\hat{\xi}) \quad (3)$$

or, more frequently, one would directly define one-dimensional intervals that satisfy the above conditions for the marginalized posterior distributions on the individual parameter axes. Unfortunately, if one performs such an analysis many times one is not guaranteed that the true parameters (e.g. of a simulation) are located within $V_{68.3 \text{ per cent}}$ in 68.3 per cent of the times. This has recently been referred to as *prior volume effect* [this issue is discussed in, e.g. Raveri & Hu (2019) and Abbott et al. (2019a)]. One may argue that a Bayesian posterior should not be

interpreted in terms of frequencies but that does not help for the task of validating this posterior on the basis of a large number of simulated data sets.⁹

Another more practical problem is the fact that it is not (yet) feasible to generate enough sufficiently accurate mock data sets to validate covariance matrices of large data vectors with high precision. We recall that for the DES-Y1 analysis, a total of 18 realistic simulated data sets were available to validate the inference pipeline (MacCrann et al. 2018). At the same time, the main reason why N -body simulations would be required to test the accuracy of covariance (and likelihood) models is to capture contributions to the covariance coming from the trispectrum (connected four-point function) of the cosmic density field. However, for DES-like analyses it has been shown that this contribution is negligible (see e.g. Krause et al. 2017; Barreira, Krause & Schmidt 2018). The reason for this is twofold: First, very small scales (where the trispectrum contribution to the covariance would matter most) are often cut off from analyses because on these scales already the modelling of the data vector, $\xi[\boldsymbol{\pi}]$, is inaccurate. Secondly, on small scales the covariance matrix is often dominated by effects coming from sparse sampling such as shot noise and shape noise. These covariance contributions are typically easy to model (although one has to be careful when estimating effective number densities and shape-noise dispersions or when estimating the number of galaxy pairs in the presence of complex survey footprints; see Troxel et al. 2018a, b).

As a result of the above considerations, we base our covariance validation strategy mostly on the use of a linearized likelihood (where the model $\xi[\boldsymbol{\pi}]$ is linear in the parameters $\boldsymbol{\pi}$). In this framework, the Bayesian likelihood allows for an interpretation in terms of frequencies – both for total and marginalized constraints. Also, this allows us to perform large numbers of simulated likelihood analyses very efficiently, without the need to run computationally expensive Markov chain Monte Carlo (MCMC) codes. In addition, any leading-order deviation from a linearized likelihood will be next-to-leading order for the purpose of studying the impact of covariance errors (i.e. errors on errors) on our analysis.

Within the linearized likelihood formalism, we confirm the findings of Krause et al. (2017) and Barreira et al. (2018) for the DES-Y3 set-up: Both supersample covariance and trispectrum have a negligible impact on our analysis. This allows us to estimate the impact of other assumptions in our covariance and likelihood model either analytically or by the means of simplified mock data such as lognormal simulations (as opposed to full N -body simulations; cf. Section 4.3).

We summarize our main findings in Fig. 1 and Table 1 for the busy reader. For the combined data vector of the DES-Y3 two-point function analysis (the 3×2 pt data vector; see details in Section 3), the left-hand panel of Fig. 1 shows the impact of different assumptions in our likelihood model on the mean and scatter of χ^2 between maximum posterior model and measurements. To obtain the maximum posterior model, we are fitting for all the 28 parameters listed in Table 3 within the linearized likelihood framework described in Section 5.1. Since we assume Gaussian priors on 13 nuisance parameters, the effective number of parameters in that fit will be between 28 and 15. Within the linearized likelihood approach, we find that with a perfect covariance model the average χ^2 is expected to be about 507.6; i.e. the effective number of degrees of freedom in the fit is $N_{\text{param,eff}} \approx 23.4$. The right-hand panel of Fig. 1 shows the equivalent results when cosmic shear correlation functions are excluded from the data vector (the 2×2 pt data vector). The green points in both panels denote effects that have been already accounted for in the previous year-1 analysis of DES.

What stands out in our analysis is the large effect of finite angular bin sizes on the cosmic variance and mixed terms of our covariance model (cf. Section 4 for this terminology, where we also show that it is unavoidable to take into account finite bin width in the pure shot noise and shape noise terms of the covariance). In DES-Y1, this has been dealt with in an approximate manner, by computing the covariance model for a very fine angular binning and then re-summing the matrix to obtain a coarser binning (Krause et al. 2017). This time we incorporate the exact treatment of finite angular bin size for all the three two-point functions into our fiducial covariance model (cf. Section 4). The blue points in Fig. 1 denote improvements that have been made in the year-3 analysis compared to the year-1 covariance model, and the red points are estimates of effects that are not taken into account in the fiducial DES-Y3 likelihood – either because they are negligible or because an exact treatment is unfeasible (cf. Section 6 for details). Adding these effects in quadrature, our results suggest that the maximum posterior χ^2 of the DES-Y3 3×2 pt analysis should be on average ≈ 4 per cent ($\Delta\chi^2 \approx 20.3$) higher than expected if the exact covariance matrix of our data vector was known.

Table 1 summarizes the offsets in χ^2 displayed in the left-hand panel of Fig. 1 and also shows how parameter constraints based on the 3×2 pt data vector are impacted by assumptions of our covariance and likelihood model. We distinguish two effects here: the scatter of a maximum posterior parameter π (which we denote by $\sigma[\hat{\pi}]$) and the width of posterior constraints inferred from our likelihood model (which we denote by σ_π). For our tests of likelihood non-Gaussianity, we state the changes in the difference between the fiducial parameter values and the upper (high) and lower (low) boundaries of the 68.3 per cent quantile with respect to the standard deviation of the Gaussian likelihood. For our tests of the impact of covariance cosmology, we show the mean of all σ_π obtained from our 100 different covariances and also indicate the scatter of these σ_π values.

The effect that has the dominant impact on parameter constraints is that of evaluating the covariance model at a set of parameters that do not represent the exact cosmology of the Universe. When computing the covariance at 100 different cosmologies that were randomly drawn from a Monte Carlo Markov chain (run around a fiducial model data vector; see Section 6.8 for details), we find that the differences between these covariances introduce an additional scatter in maximum posterior parameter values. This scatter increases by about 3 per cent for Ω_m

⁹In order to deal with the prior volume effect, Joachimi et al. (2020) proposed to report parameter constraints through what they call projected joint highest posterior density. This topic will be addressed in a separate DES paper (Raveri et al., in preparation).

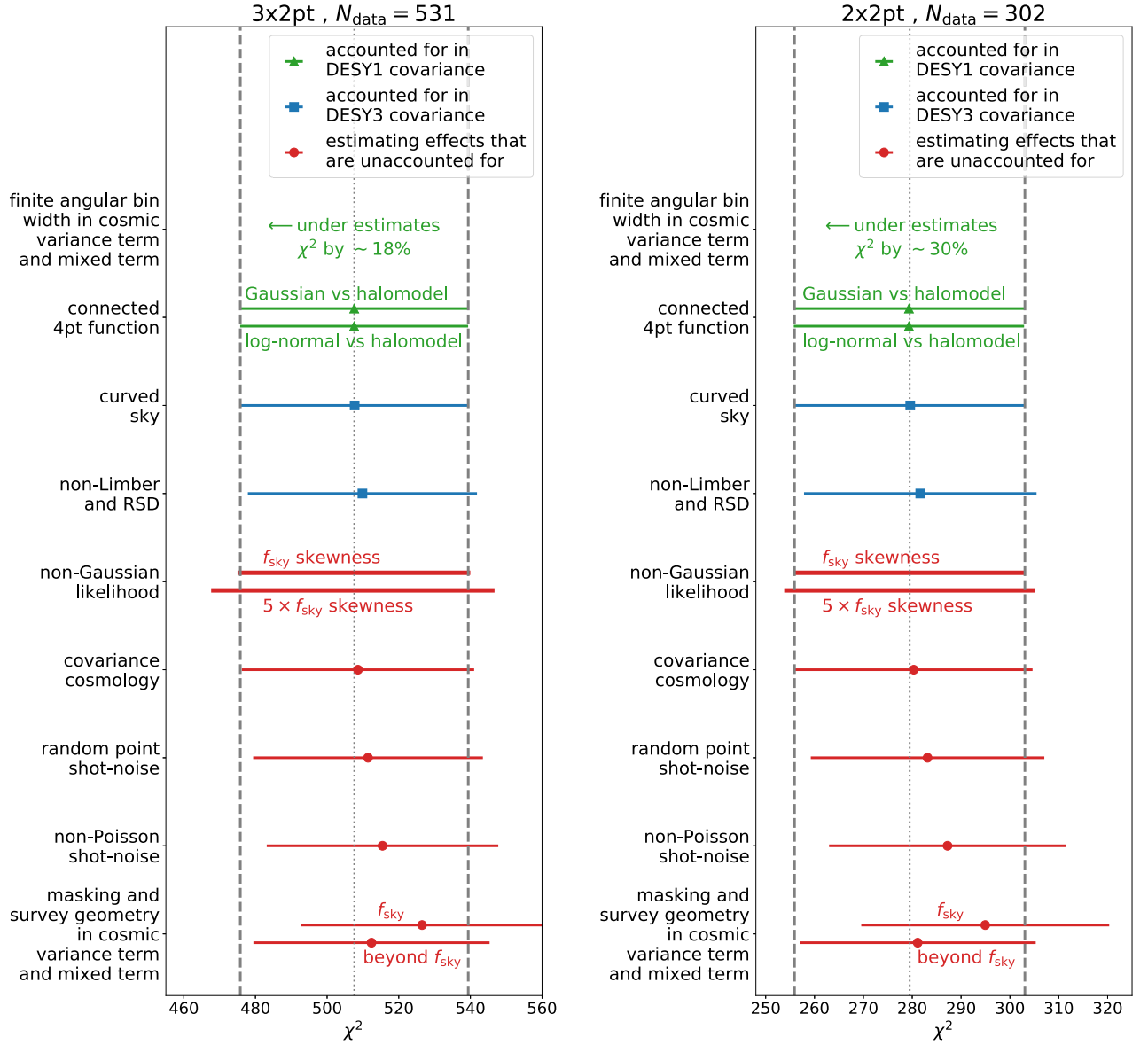


Figure 1. Impact of different covariance modelling choices on χ^2 between measured 3×2 pt (left-hand panel) and 2×2 pt (right-hand panel) data vectors and maximum posterior models. The dashed vertical lines and error bars indicate the 1σ fluctuations expected in χ^2 . See the main text for details.

Table 1. Summary of the impact of the different effects tested here on the distribution of χ^2 between measurement and maximum posterior model, on the scatter $\sigma[\hat{\pi}]$ of maximum posterior parameters $\hat{\pi}$, and on the standard deviations σ_{π} on these parameters inferred from the likelihood. See the text for details.

Effect	$\langle \chi^2 \rangle$	$\sigma(\chi^2)$	$\sigma(\hat{\Omega}_m)$	σ_{Ω_m}	$\sigma(\hat{\sigma}_8)$	σ_{σ_8}	$\sigma(\hat{w})$	σ_w
Fiducial	507.6	31.8	0.0509	0.0509	0.0975	0.0975	0.244	0.244
Angular bin width	402.1	26.0	+0.8 per cent	+7.4 per cent	+0.8 per cent	+8.3 per cent	+1.0 per cent	+7.4 per cent
Connected four-point function	507.6	31.8	+0.1 per cent	-0.8 per cent	+0.1 per cent	-0.9 per cent	+0.1 per cent	-0.8 per cent
Curved sky	507.7	31.8	+0.0 per cent	-0.0 per cent	+0.0 per cent	-0.0 per cent	+0.0 per cent	-0.0 per cent
Non-Limber and RSD	511.4	32.1	+0.1 per cent	-0.6 per cent	+0.1 per cent	-0.6 per cent	+0.3 per cent	-1.4 per cent
Non-Gauss. likelihood	-	32.6	+0.8 per cent (low)	-	+0.4 per cent (low)	-	+0.5 per cent (low)	-
			-0.9 per cent (high)		-0.4 per cent (high)		+0.05 per cent (high)	
Covariance cosmology	508.6	32.4	+2.9 per cent	+(0.1 ± 0.06) per cent	+2.8 per cent	+(0.1 ± 0.05) per cent	+4.7 per cent	+(0.1 ± 0.06) per cent
Random point shot noise	511.3	32.0	+0.0 per cent	-0.5 per cent	+0.0 per cent	-0.6 per cent	+0.0 per cent	-0.2 per cent
Non-Poisson shot noise	515.0	32.3	+0.0 per cent	-0.7 per cent	+0.0 per cent	-0.8 per cent	+0.0 per cent	-0.6 per cent
Masking and survey geometry	526.5	33.8	+0.6 per cent	-0.8 per cent	+0.7 per cent	-0.3 per cent	+0.3 per cent	-1.3 per cent

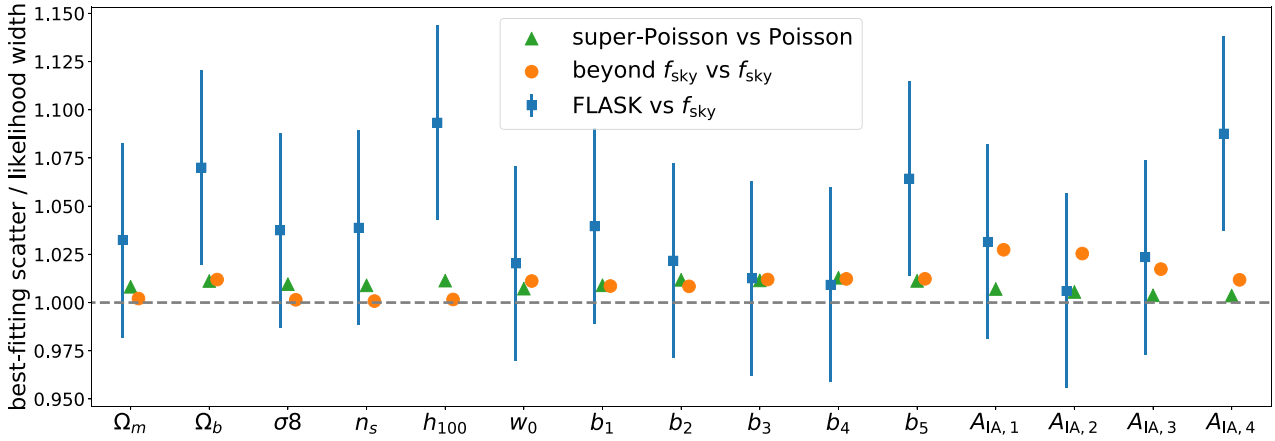


Figure 2. Impact of covariance errors on the ratio of the standard deviation of maximum posterior parameters to the width of the posterior derived from the erroneous covariance. Green triangles show the effect caused by non-Poissonian shot-noise and orange circles show the effect caused by the f_{sky} approximation (cf. Appendix C for our beyond- f_{sky} treatment). These ratios have been calculated purely on the base of different analytic covariance matrices and within the linearized likelihood framework discussed in Section 5.1. We also show the ratio of maximum posterior parameter scatter observed from the 197 FLASK simulations to the statistical uncertainties expected from a lognormal covariance matrix matching the FLASK configuration. Within the statistical uncertainties, these ratios are consistent with 1.

and σ_8 and by about 5 per cent for the dark energy equation-of-state parameter w . This increased scatter is in fact the dominant effect, since the width of the derived parameter constraints hardly changes between the different covariance matrices. Note especially that rerunning the analysis with a covariance updated to the best-fitting parameters does not mitigate this effect.

In Fig. 2, we take the two effects that had the largest impacts on χ^2 and show the resulting mismatch between scatter of maximum posterior values and width of the inferred contours for a wider range of parameters. All of our results take into account marginalization over nuisance parameters (and all other parameters).

Our reason for exclusively investigating the impact of covariance errors on χ^2 and parameter constraints is that those are the two measures by which our final (on-shot) data analysis will be interpreted and judged.¹⁰ In the remainder of this paper, we detail how the above results were obtained.

3 THE 3×2 -POINT DATA VECTOR

The combined 3×2 pt data vector of the DES-Y3 analysis consists of measurements of the following two-point correlations:

- (i) the angular two-point correlation function $w(\theta)$ of galaxy density contrast measured for luminous red galaxies in five different redshift bins (see e.g. Rodríguez-Monroy et al. 2021; Cawthon et al. 2020, as well as other relevant references given in Section 1),
- (ii) the auto- and cross-correlation functions $\xi_+(\theta)$ and $\xi_-(\theta)$ between the galaxy shapes of four redshift bins of source galaxies (see e.g. Amon et al. 2021; Gatti et al. 2020a; Myles et al. 2021; Secco et al. 2021),
- (iii) the tangential shear $\gamma(\theta)$ imprinted on source galaxy shapes around positions of foreground redMaGiC galaxies (see e.g. Prat et al. 2021).

At the time of writing this paper, the exact choices for redshift intervals and angular bins considered for each two-point function are still being determined by a careful study of their impact on the robustness of DES-Y3 parameter constraints (DES Collaboration 2020; Krause et al. 2021). For the purposes of testing the modelling of the covariance matrix, we will use the most recent but possibly not final DES-Y3 analysis choices. We do not expect that our tests and conclusions will change in a significant manner with further updated analysis choices. We assume that each of the correlation functions is measured in 20 logarithmically spaced angular bins between $\theta_{\min} = 2.5$ arcmin and $\theta_{\max} = 250$ arcmin. Some of these bins in some of the measured two-point functions are being cut from the analysis to ensure unbiased cosmological results, resulting in a total of 531 data points when using the preliminary DES-Y3 scale cuts.

Our starting point of modelling the different two-point functions in the 3×2 pt data vector is the three-dimensional (3D) non-linear matter power spectrum $P(k, z)$ at a given wavenumber k and redshift z . We obtain it by using either of the Boltzmann solvers CLASS¹¹ or CAMB¹² to calculate the linear power spectrum and the HALOFIT fitting formula (Smith et al. 2003) in its updated version (Takahashi et al. 2012) to turn this into the late-time non-linear power spectrum. From this 3D power spectrum, the angular power spectra required for our three two-point functions [cosmic shear ($\kappa\kappa$), galaxy–galaxy lensing ($\delta_g\kappa$), and galaxy–galaxy clustering ($\delta_g\delta_g$)] in the Limber approximation are given by (e.g. Limber 1953; Krause et al. 2017):

¹⁰Alternatively, one could investigate the distribution of p -values (or *probability to exceed*; cf. Hall & Taylor 2019) as opposed to the distribution of χ^2 .

¹¹www.class-code.net

¹²camb.info

$$C_{\kappa\kappa}^{ij}(\ell) = \int d\chi \frac{q_{\kappa}^i(\chi)q_{\kappa}^j(\chi)}{\chi^2} P\left(\frac{\ell + \frac{1}{2}}{\chi}, z(\chi)\right), \quad (4)$$

$$C_{\delta_{\text{g}\kappa}}^{ij}(\ell) = \int d\chi \frac{q_{\delta}^i\left(\frac{\ell + \frac{1}{2}}{\chi}, \chi\right)q_{\kappa}^j(\chi)}{\chi^2} P\left(\frac{\ell + \frac{1}{2}}{\chi}, z(\chi)\right), \quad (5)$$

$$C_{\delta_{\text{g}\delta_{\text{g}}}}^{ij}(\ell) = \int d\chi \frac{q_{\delta}^i\left(\frac{\ell + \frac{1}{2}}{\chi}, \chi\right)q_{\delta}^j\left(\frac{\ell + \frac{1}{2}}{\chi}, \chi\right)}{\chi^2} P\left(\frac{\ell + \frac{1}{2}}{\chi}, z(\chi)\right), \quad (6)$$

where χ is the comoving radial distance, i and j denote different combinations of pairs of redshift bins, and the lensing efficiency q_{κ}^i and the radial weight function for clustering q_{δ}^i are, respectively, given by

$$q_{\kappa}^i(\chi) = \frac{3H_0^2\Omega_m\chi}{2a(\chi)} \int_{\chi}^{\chi_h} d\chi' \left(\frac{\chi' - \chi}{\chi}\right) n_{\kappa}^i(z(\chi')) \frac{dz}{d\chi'},$$

$$q_{\delta}^i(k, \chi) = b^i(k, z(\chi)) n_{\delta}^i(z(\chi)) \frac{dz}{d\chi}. \quad (7)$$

Here, H_0 is the Hubble parameter today, Ω_m is the ratio of today's matter density to today's critical density of the Universe, $a(\chi)$ is the Universe's scale factor at comoving distance χ , and $b^i(k, z)$ is a scale- and redshift-dependent galaxy bias. Furthermore, $n_{\kappa, \text{g}}^i(z)$ denote the redshift distributions of the different DES-Y3 redshift bins of source and lens galaxies, respectively, normalized such that

$$\int dz n_{\kappa, \text{g}}^i(z) = 1. \quad (8)$$

Note that on large angular scales the DES-Y3 analysis does not make use of the Limber approximation for galaxy clustering but instead employs the method derived in Fang, Eifler & Krause (2020a).

The above angular power spectra are now related to the real space correlation functions $w(\theta)$, $\gamma_r(\theta)$, and $\xi_{\pm}(\theta)$ as

$$w^i(\theta) = \sum_{\ell} \frac{2\ell + 1}{4\pi} P_{\ell}(\cos\theta) C_{\delta_{\text{g}\delta_{\text{g}}}}^{ii}(\ell),$$

$$\gamma_r^{ij}(\theta) = \sum_{\ell} \frac{2\ell + 1}{4\pi} \frac{P_{\ell}^2(\cos\theta)}{\ell(\ell + 1)} C_{\delta_{\text{g}\kappa}}^{ij}(\ell),$$

$$\xi_{\pm}^{ij}(\theta) = \sum_{\ell \geq 2} \frac{2\ell + 1}{4\pi} \frac{2(G_{\ell, 2}^+(x) \pm G_{\ell, 2}^-(x))}{\ell^2(\ell + 1)^2} C_{\kappa\kappa}^{ij}(\ell). \quad (9)$$

Here, P_{ℓ} are the Legendre polynomials of order ℓ , P_{ℓ}^m are the associated Legendre polynomials, $x = \cos\theta$, and the functions $G_{\ell, 2}^{\pm}(x)$ are given in Appendix A (see also Stebbins 1996). Note that we only consider the autocorrelations $w^i(\theta)$ for each tomographic bin since in the Y1 analysis it was shown that the cross-correlations do not carry significant information (Elvin-Poole et al. 2018).

The above relations between angular power spectra and real space correlation functions can all be written in the form

$$\xi^{AB}(\theta) = \sum_{\ell=0}^{\infty} \frac{2\ell + 1}{4\pi} F_{\ell}^{AB}(\theta) C_{\ell}^{AB}. \quad (10)$$

This is particularly useful when deriving covariance expressions and when performing averages over finite bins in the angular scale θ . To achieve the latter, one can simply derive analytical averages of the functions $F_{\ell}^{AB}(\theta)$. Both of these points will be considered in the next sections.

For most of our tests, we consider the 3×2 pt data vector and its covariance matrix at the fiducial cosmology described in Section 5, where we also show the Gaussian priors assumed on some of these parameters when assessing the impact of covariance modelling on parameter constraints and maximum posterior χ^2 .

4 COVARIANCE MATRICES FOR THE 3×2 PT DATA VECTOR

The covariance matrix of measurements of cosmological two-point statistics typically contains three contributions (cf. Krause & Eifler 2017; Krause et al. 2017),

$$\mathbf{C} = \mathbf{C}_{\text{G}} + \mathbf{C}_{\text{nG}} + \mathbf{C}_{\text{SSC}}. \quad (11)$$

Here, \mathbf{C}_{G} is the contribution to the covariance that would be present if the cosmic matter density and cosmic shear fields were pure Gaussian random fields (see also Schneider et al. 2002; Crocce, Cabré & Gaztañaga 2011), \mathbf{C}_{nG} are contributions involving the connected four-point function of these fields (the trispectrum), and \mathbf{C}_{SSC} is the so-called supersample covariance contribution resulting from the fact that any survey

only observes a finite volume of the Universe and that the mean density in that volume is subject to fluctuations due to long wavelength modes (Takada & Hu 2013; Schaan, Takada & Spergel 2014).

In the fiducial DES-Y3 analysis, we model all of these covariance contributions analytically. This fiducial model is described in Section 4.1. In Section 4.2, we describe an alternative model for the non-Gaussian covariance contributions that is used to test the robustness of our analysis with respect to the modelling of the trispectrum contribution. Finally, Section 4.3 describes a set of lognormal simulations (Xavier, Abdalla & Joachimi 2016) and the covariance matrix of the 3×2 pt data vector estimated from them. These simulations also allow us to test the accuracy of our Gaussian likelihood assumption and the treatment of masking and finite survey area in our fiducial covariance model.

4.1 Fiducial DES-Y3 covariance

In our fiducial covariance matrix, we model the non-Gaussian covariance contributions C_{nG} and C_{SSC} using a halo model combined with leading-order perturbation theory to approximate the trispectrum of the cosmic density field and to compute the mode coupling between scales larger than the considered survey volume with scales inside that volume. These calculations are carried out using the `CosmoCov` code package (Fang et al. 2020a) based on the `CosmoLike` framework (Krause & Eifler 2017). Our modelling of these contributions has not changed with respect to the year-1 analysis of DES and we refer the reader to Krause et al. (2017) as well as to the `CosmoLike` papers for details. However, the modelling of the Gaussian contribution has changed as described in the following.

4.1.1 Gaussian covariance

Our modelling of the Gaussian covariance part has changed with respect to the year-1 analysis in the following ways:

- (i) we use (and present for the first time¹³) analytical expression for the angular bin averaging of the functions $F_\ell^{AB}(\theta)$ (cf. equation 10) for all four types of two-point functions present in our data vector (see Section 6.3; this is especially relevant for the sampling-noise contribution to the covariance; cf. Troxel et al. 2018b);
- (ii) we account for RSDs and also use a non-Limber calculation to obtain the galaxy–galaxy clustering power spectrum $C_{\delta_g \delta_g}(\ell)$ (see Section 6.5);
- (iii) we do not make use of the flat-sky approximation anymore (see Section 6.4).

To derive expressions for the Gaussian covariance part, let us first consider an all-sky survey. If a two-point function measurement $\hat{\xi}^{AB}(\theta)$ could be obtained from data on the entire sky, then for most types of two-point correlations it would be related to power spectrum measurements C_ℓ^{AB} from a spherical harmonics decomposition of the same all-sky data through equation (10), i.e.

$$\hat{\xi}^{AB}(\theta) = \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} F_\ell^{AB}(\theta) \hat{C}_\ell^{AB}. \quad (12)$$

A notable exception to this is the cosmic shear two-point functions $\hat{\xi}_\pm$ that obtain contributions from both the so-called E-mode and B-mode power spectra (Schneider et al. 2002). For these functions, equation (12) in the curved sky formalism becomes

$$\hat{\xi}_\pm^{ij}(\theta) = \sum_{\ell \geq 2} \frac{2\ell+1}{4\pi} \frac{2(G_{\ell,2}^+(x) \pm G_{\ell,2}^-(x))}{\ell^2(\ell+1)^2} (\hat{C}_{\gamma\gamma}^{E,ij}(\ell) \pm \hat{C}_{\gamma\gamma}^{B,ij}(\ell)), \quad (13)$$

where in the absence of shape-measurement systematics (and ignoring post-Born corrections) $\langle \hat{C}_{\gamma\gamma}^{E,ij}(\ell) \rangle = C_{\kappa\kappa}^{ij}(\ell)$ and $\langle \hat{C}_{\gamma\gamma}^{B,ij}(\ell) \rangle = 0$.

Since this is a linear equation in $C(\ell)$'s, the covariance of two different two-point function measurements $\hat{\xi}^{AB}$ and $\hat{\xi}^{CD}$ at two different angular scales θ_1 and θ_2 would be given in terms of the covariance of the corresponding power spectrum measurements by

$$\text{Cov} [\hat{\xi}^{AB}(\theta_1), \hat{\xi}^{CD}(\theta_2)] = \sum_{\ell_1, \ell_2} \frac{(2\ell_1+1)(2\ell_2+1)}{(4\pi)^2} F_{\ell_1}^{AB}(\cos \theta_1) F_{\ell_2}^{CD}(\cos \theta_2) \text{Cov} [\hat{C}_{\ell_1}^{AB}, \hat{C}_{\ell_2}^{CD}]. \quad (14)$$

Again, for $\hat{\xi}^{AB}(\theta) = \hat{\xi}_\pm(\theta)$ one would have to use $C_\ell^{AB} = \hat{C}_{\gamma\gamma}^E(\ell) \pm \hat{C}_{\gamma\gamma}^B(\ell)$ in this sum.

For the autopower spectrum of galaxy density contrast in one of our redshift bins, the harmonic space covariance would be (Crocce et al. 2011)

$$\text{Cov} [\hat{C}_{\delta_g \delta_g}^{ii}(\ell_1), \hat{C}_{\delta_g \delta_g}^{ii}(\ell_2)] = \frac{2\delta_{\ell_1 \ell_2}}{(2\ell_1+1)} \left(C_{\delta_g \delta_g}^{ii}(\ell_1) + \frac{1}{n_g} \right)^2. \quad (15)$$

Here, n_g is the number density of the galaxies and $\delta_{\ell_1 \ell_2}$ is the Kronecker symbol. To account for partial-sky surveys (such as DES), we simply divide this expression (and similar ones for the other two-point functions) by the observed sky fraction f_{sky} . This so-called f_{sky} approximation leads to the following harmonic space Gaussian covariances of (see also Krause et al. 2017):

¹³We have shared our results with Fang et al. (2020a), who have used them for their covariance calculations.

$$\text{Cov} \left[\hat{C}_{\text{gg}}^{ij}(\ell_1), \hat{C}_{\text{gg}}^{kl}(\ell_2) \right] = \frac{\delta_{\ell_1 \ell_2} \left[\left(C_{\text{gg}}^{ik}(\ell_1) + \frac{\delta_{ik}}{n_g^j} \right) \left(C_{\text{gg}}^{jl}(\ell_1) + \frac{\delta_{jl}}{n_g^j} \right) + \left(C_{\text{gg}}^{il}(\ell_1) + \frac{\delta_{il}}{n_g^j} \right) \left(C_{\text{gg}}^{jk}(\ell_1) + \frac{\delta_{jk}}{n_g^j} \right) \right]}{(2\ell_1 + 1)f_{\text{sky}}} \quad (16)$$

$$\text{Cov} \left[\hat{C}_{\gamma\gamma}^{E,ij}(\ell_1), \hat{C}_{\gamma\gamma}^{E,kl}(\ell_2) \right] = \frac{\delta_{\ell_1 \ell_2} \left[\left(C_{\kappa\kappa}^{ik}(\ell_1) + \frac{\delta_{ik}\sigma_{\epsilon,i}^2}{n_s^j} \right) \left(C_{\kappa\kappa}^{jl}(\ell_1) + \frac{\delta_{jl}\sigma_{\epsilon,j}^2}{n_s^j} \right) + \left(C_{\kappa\kappa}^{il}(\ell_1) + \frac{\delta_{il}\sigma_{\epsilon,i}^2}{n_s^j} \right) \left(C_{\kappa\kappa}^{jk}(\ell_1) + \frac{\delta_{jk}\sigma_{\epsilon,j}^2}{n_s^j} \right) \right]}{(2\ell_1 + 1)f_{\text{sky}}} \quad (17)$$

$$\text{Cov} \left[\hat{C}_{\gamma\gamma}^{B,ij}(\ell_1), \hat{C}_{\gamma\gamma}^{B,kl}(\ell_2) \right] = \frac{\delta_{\ell_1 \ell_2} \left[\frac{\delta_{ik}\sigma_{\epsilon,i}^2}{n_s^j} \frac{\delta_{jl}\sigma_{\epsilon,j}^2}{n_s^j} + \frac{\delta_{il}\sigma_{\epsilon,i}^2}{n_s^j} \frac{\delta_{jk}\sigma_{\epsilon,j}^2}{n_s^j} \right]}{(2\ell_1 + 1)f_{\text{sky}}} \quad (18)$$

$$\text{Cov} \left[\hat{C}_{\text{gk}}^{ij}(\ell_1), \hat{C}_{\text{gk}}^{kl}(\ell_2) \right] = \frac{\delta_{\ell_1 \ell_2} \left[\left(C_{\text{gg}}^{ik}(\ell_1) + \frac{\delta_{ik}}{n_g^j} \right) \left(C_{\kappa\kappa}^{jl}(\ell_1) + \frac{\delta_{jl}\sigma_{\epsilon,j}^2}{n_s^j} \right) + C_{\text{gk}}^{il}(\ell_1) C_{\text{gk}}^{kj}(\ell_1) \right]}{(2\ell_1 + 1)f_{\text{sky}}} \quad (19)$$

$$\text{Cov} \left[\hat{C}_{\text{gg}}^{ij}(\ell_1), \hat{C}_{\gamma\gamma}^{E,kl}(\ell_2) \right] = \frac{\delta_{\ell_1 \ell_2} \left[C_{\text{gk}}^{ik}(\ell_1) C_{\text{gk}}^{jl}(\ell_1) + C_{\text{gk}}^{il}(\ell_1) C_{\text{gk}}^{jk}(\ell_1) \right]}{(2\ell_1 + 1)f_{\text{sky}}} \quad (20)$$

$$\text{Cov} \left[\hat{C}_{\text{gg}}^{ij}(\ell_1), \hat{C}_{\text{gk}}^{kl}(\ell_2) \right] = \frac{\delta_{\ell_1 \ell_2} \left[\left(C_{\text{gg}}^{ik}(\ell_1) + \frac{\delta_{ik}\sigma_{\epsilon,i}^2}{n_s^j} \right) C_{\text{gk}}^{jl}(\ell_1) + C_{\text{gk}}^{il}(\ell_1) \left(C_{\text{gg}}^{jk}(\ell_1) + \frac{\delta_{jk}}{n_g^j} \right) \right]}{(2\ell_1 + 1)f_{\text{sky}}} \quad (21)$$

$$\text{Cov} \left[\hat{C}_{\text{gk}}^{ij}(\ell_1), \hat{C}_{\gamma\gamma}^{E,kl}(\ell_2) \right] = \frac{\delta_{\ell_1 \ell_2} \left[C_{\text{gk}}^{ik}(\ell_1) \left(C_{\kappa\kappa}^{jl}(\ell_1) + \frac{\delta_{jl}\sigma_{\epsilon,j}^2}{n_s^j} \right) + C_{\text{gk}}^{il}(\ell_1) \left(C_{\kappa\kappa}^{jk}(\ell_1) + \frac{\delta_{jk}\sigma_{\epsilon,j}^2}{n_s^j} \right) \right]}{(2\ell_1 + 1)f_{\text{sky}}} \quad (22)$$

$$\text{Cov} \left[\hat{C}_{\text{gg}}^{ij}(\ell_1), \hat{C}_{\gamma\gamma}^{B,kl}(\ell_2) \right] = 0 \quad (\text{as are all other covariances with only one } \hat{C}_{\gamma\gamma}^B). \quad (23)$$

At this point, let us introduce the following nomenclature: We will denote the terms that contain two power spectra as *cosmic variance* contribution to the covariance, the terms that contain no power spectrum at all as the *sampling noise* contributions (or *shape-noise* and *shot-noise* contributions), and the terms that contain contribution from one power spectrum and a sampling noise as the *mixed terms*. We test the accuracy of the f_{sky} approximation that results in equations (16–23) in Section 6.6 by comparing it to more accurate expressions.

4.2 Analytical lognormal covariance model

To test the robustness of the CosmoLike covariance, we also employ an alternative model for the connected four-point function part of the covariance – the lognormal model. Hilbert, Hartlap & Schneider (2011) originally derived this as a model for the covariance of cosmic shear correlation function, assuming that the lensing convergence κ can be written in terms of a Gaussian random field n as (see also Xavier et al. 2016)

$$\kappa = \lambda \left(e^{n+\mu} - 1 \right), \quad (24)$$

where it is assumed that $\langle n \rangle = 0$. For given values $\lambda > 0$ and μ the power spectrum of n can be chosen such as to reproduce a desired two-point correlation function ξ_κ (see Xavier et al. 2016, for caveats). Furthermore, for any given value $\lambda > 0$ one can choose μ such that $\langle \kappa \rangle = 0$. This makes λ the only free parameter of the lognormal covariance model. Hilbert et al. (2011) show that this model leads to a number of correction terms to the Gaussian covariance model, and identify the most dominant of these terms to be

$$C_{\text{LN}}[\hat{\xi}_\kappa(\theta_1), \hat{\xi}_\kappa(\theta_2)] \approx C_{\text{G}}[\hat{\xi}_\kappa(\theta_1), \hat{\xi}_\kappa(\theta_2)] + \frac{4 \xi_\kappa(\theta_1) \xi_\kappa(\theta_2)}{A_{\text{S}} \lambda^2} \text{Var}_{\text{S}}(\kappa). \quad (25)$$

Here, A_{S} is the area of the considered survey footprint and $\text{Var}_{\text{S}}(\kappa)$ is the variance of κ when averaged over the footprint. We generalize this to the covariance of two-point correlations $\hat{\xi}_{AB}$ and $\hat{\xi}_{CD}$ between arbitrary scalar fields δ_A , δ_B , δ_C , and δ_D as

$$C_{\text{LN}}[\hat{\xi}_{AB}(\theta_1), \hat{\xi}_{CD}(\theta_2)] - C_{\text{G}}[\hat{\xi}_{AB}(\theta_1), \hat{\xi}_{CD}(\theta_2)] \approx \frac{\xi_{AB}(\theta_1) \xi_{CD}(\theta_2)}{A_{\text{S}}} \left\{ \frac{\text{Cov}_{\text{S}}(\delta_A, \delta_C)}{\lambda_A \lambda_C} + \frac{\text{Cov}_{\text{S}}(\delta_A, \delta_D)}{\lambda_A \lambda_D} + \frac{\text{Cov}_{\text{S}}(\delta_B, \delta_C)}{\lambda_B \lambda_C} + \frac{\text{Cov}_{\text{S}}(\delta_B, \delta_D)}{\lambda_B \lambda_D} \right\}. \quad (26)$$

Here, $\text{Cov}_{\text{S}}(\delta_A, \delta_C)$ is the covariance of δ_A and δ_C after the two fields have been averaged over the entire survey footprint (and likewise for the other terms appearing above). Following Hilbert et al. (2011), we use this expression even when considering non-scalar fields (i.e. the shear field) by replacing $\xi_{XY}(\theta)$ by the appropriate two-point functions $\xi_+(\theta)$, $\xi_-(\theta)$, $\gamma_t(\theta)$ [or $w(\theta)$, for the scalar galaxy density contrast].

To choose the parameters λ_X (also called the lognormal shift parameters; cf. Xavier et al. 2016), we follow a procedure similar to the one outlined in Friedrich et al. (2018). There, it is shown how the value of λ_X can be adjusted in order to match the re-scaled cumulant

$$S_3(\vartheta) \equiv \frac{\langle \delta_X(\vartheta)^3 \rangle}{\langle \delta_X(\vartheta)^2 \rangle^2} \quad (27)$$

of the random field δ_X smoothed with a top-hat filter of angular radius ϑ to the value of S_3 predicted by leading-order perturbation theory for that same smoothing scale. Since the focus in our paper is the covariance matrix of two-point statistics (hence a four-point function), we

Table 2. Number density, galaxy bias (relevant for lenses only), shape-noise dispersion (per shear component; relevant for sources only), and the lognormal shift parameters obtained from the procedure described in Section 4.2.

z -bin	n_g (arcmin $^{-2}$)	Bias	σ_ϵ	Lognormal shift
Lenses 1	0.0221	1.7	—	1.089
Lenses 2	0.0381	1.7	—	1.106
Lenses 3	0.0583	1.7	—	1.047
Lenses 4	0.0295	2.0	—	1.252
Lenses 5	0.0251	2.0	—	1.177
Sources 1	1.7971	—	0.2724	0.004 53
Sources 2	1.5521	—	0.2724	0.008 85
Sources 3	1.5967	—	0.2724	0.019 18
Sources 4	1.0979	—	0.2724	0.032 87

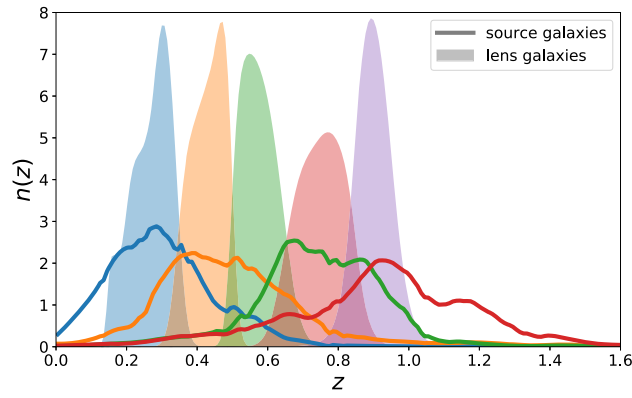


Figure 3. Redshift distributions of lens galaxies (shaded regions) and source galaxies (solid lines) in our fiducial test configuration.

modify their method to match instead the value of reduced fourth-order cumulant

$$S_4(\vartheta) \equiv \frac{\langle \delta_X(\vartheta)^4 \rangle - 3 \langle \delta_X(\vartheta)^2 \rangle^2}{\langle \delta_X(\vartheta)^2 \rangle^3}. \quad (28)$$

The field δ_X here will be either projections of the 3D matter density contrast along the line-of-sight distribution of our lens galaxies or the lensing convergence fields corresponding to our four source redshift bins. The smoothing scale ϑ at which we use the λ_X to match S_4 to its perturbation theory value is chosen such that it corresponds to about 10 Mpc h^{-1} at the mean redshift of the line-of-sight projection kernels corresponding to the different δ_X . This is approximately the scale at which Friedrich et al. (2018) found the shifted lognormal model to be a good approximation of the overall PDF of density fluctuations in N -body simulations (cf. their fig. 5).

Our results are shown in Table 2, where we present the number density, galaxy bias (relevant for lenses only), shape-noise dispersion (per shear component; relevant for sources only), and the lognormal shift parameters obtained from the procedure described above. Note that for the source galaxy samples, the relevant line-of-sight projection kernel used to derive the shift parameter is the lensing kernel (and not the redshift distribution of the source galaxies). For the lens galaxies, all shift parameters come out to be >1 . As a consequence, there will be pixels with negative density in our lognormal simulations. However, the fraction of such pixels is <0.01 for all runs and all bins and setting $\delta_g = -1$ in these bins has an unnoticeable effect on the statistics measured in these maps (e.g. for bin 4, which is affected most, the standard deviation of δ_g changes by 0.053 per cent). Note further that at the time of completing the simulation runs presented in Section 4.3, the DES-Y3 shear catalogue and redshift distribution were not finalized. As a consequence, the shape-noise dispersion values used for simulations differ from the values in this table. We display the projection kernels assumed for our analysis in Fig. 3.

4.3 Lognormal covariance from simulations

We also produce a test DES-Y3 covariance matrix from a set of simulations. We use the publicly available code FLASK (Full sky Lognormal Astro fields Simulation Kit; Xavier et al. 2016¹⁴) to generate 800 DES-Y3 footprint sky maps of density, convergence, and

¹⁴<http://www.astro.iag.usp.br/flask/>

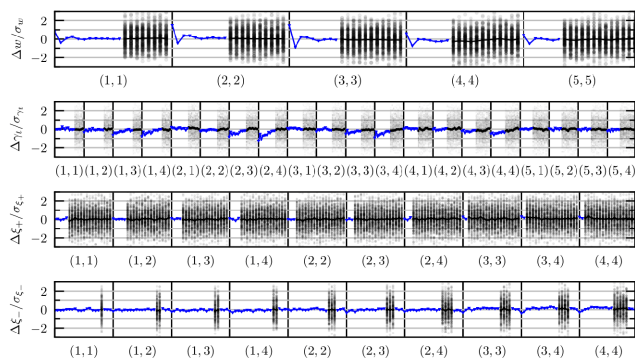


Figure 4. Validation of FLASK simulations. Each panel shows the absolute difference of three two-point correlations measured on FLASK realizations and the predicted correlation functions from input $C(\ell)$ s normalized to the statistical error given by the standard deviation along FLASK realizations ($\Delta X/\sigma_X$, where $X = w, \gamma_t, \xi_+, \xi_-$). Grey dots are single realizations and blue dots its mean.

shear healpix maps (Górski et al. 2005) with $\text{NSIDE} = 8192$, as well as galaxy positions catalogues, used to reproduce the DES-Y3 properties. FLASK is able to quickly produce tomographic correlated simulations of clustering and weak lensing lognormal fields based on the DES-Y3 lens and sources samples. The lognormal distribution of cosmological fields has been shown to be a good approximation (Coles & Jones 1991; Wild et al. 2005; Clerkin et al. 2017) but much less computationally expensive to generate than full N -body simulations.

As input for the simulations, we used a set of auto- and cross-correlated power spectrum and the lognormal field shift parameters. The theoretical input power spectrum was generated using CosmoLike, and the lognormal shifts are the ones listed in Table 2. In order to reproduce the properties of shear fields, we added the shape-noise term by sampling each pixel of the simulated maps to match the correspondent shape-noise dispersion σ_ϵ and number density n_g of the tomographic bin. At the time of completing the simulation runs, the DES-Y3 shear catalogue and redshift distribution were not finalized. For this reason, the values used in the simulations are slightly different from the values in Table 2. For the simulations, we set the number density for the five tomographic lens bins as 0.0227, 0.0392, 0.0583, 0.0451, and 0.0278 (arcmin^{-2}). The shape-noise dispersion values for the four tomographic bins of sources were set to 0.270 49, 0.332 12, 0.325 37, and 0.350 37. The cosmology adopted for the theoretical power spectra is set as $\Omega_m = 0.3$, $\sigma_8 = 0.823\ 55$, $n_s = 0.97$, $\Omega_b = 0.048$, $h_0 = 0.69$, and $\Omega_b h_0^2 = 0.000\ 83$.

We use the publicly available code TREECORR¹⁵ (Jarvis, Bernstein & Jain 2004) to measure the 3×2 point correlation measurements for 200 DES-Y3 realizations. For all measurements, we used 20 log-spaced angular separation bins on scales between 2.5 and 250 arcmin. We set the `bin_slop TreeCorr` parameter to zero, essentially setting all estimators to brute-force computation. In Fig. 4, we show the validation of the measurements comparing with the theoretical input.

We will use the FLASK covariance mainly to estimate the impact of the survey geometry.

4.4 Comparisons among covariances

Here, we present some comparisons between the different covariance matrices. In Fig. 5, we show the ratio of the diagonal elements of the different covariance matrices introduced in this section displaying both the variances of the measurements of $\xi_+(\theta)$ of $w(\theta)$.

In Fig. 6, we compare the covariance matrices obtained from the FLASK simulations and the analytical halo model covariance.

5 IMPACT OF COVARIANCE ERRORS ON A LINEARIZED GAUSSIAN LIKELIHOOD

As discussed earlier, a full assessment of the impact of using different covariance matrices to parameter estimation becomes unfeasible due to the computational demand of running a large number of MCMC chains. Since the covariance matrices studied in this work differ by subdominant effects, we do not expect large modifications in the results of the estimation of the parameters. Therefore, we will bypass this difficulty by using a linearized approximation of the model data vector as a function of the parameters. The measured data are assumed to be a Gaussian multivariate variable characterized by a covariance matrix and a given prior matrix. This approach is called the Gaussian linear model (Seehars et al. 2014, 2016; Raveri & Hu 2019).

Within this approach, we study the following impacts of different covariances:

- (i) error in the parameter estimation, characterized by the width of the contours;

¹⁵<https://github.com/rmjarvis/TreeCorr>

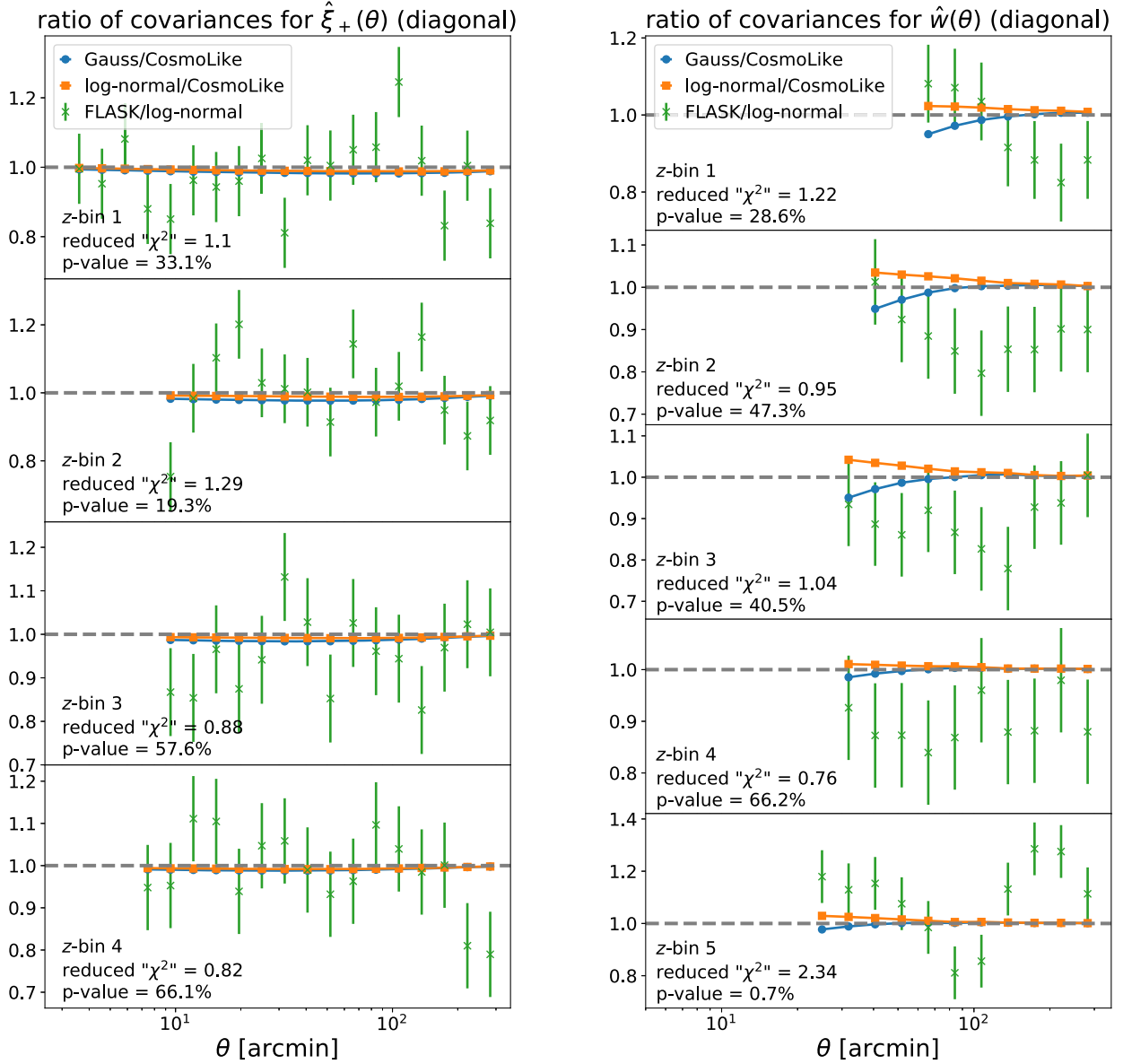


Figure 5. Ratio of the diagonal elements of the different covariance matrices introduced in this section with respect to each other. The left-hand panel compares the variances of measurements of $\xi_+(\theta)$ while the right-hand panel compares the variances of measurements of $w(\theta)$. To give a sense of the goodness of fit between the covariance estimated from FLASK and our fiducial analytic matrix, we treat the diagonal elements of the FLASK covariance as a multivariate Gaussian whose covariance can be inferred from the properties of the Wishart distribution (Taylor et al. 2013). The low p -value for the highest redshift bin of $w(\theta)$ most likely results from our incomplete treatment of the survey mask (cf. discussion in Section 6 and Appendix C).

- (ii) the scatter of the best-fitting (maximum posteriors) parameters;
- (iii) change in the maximum posterior χ^2 value;
- (iv) error in the maximum posterior χ^2 value.

In the remainder of this section, we detail this method.

5.1 Linearized likelihoods

To speed up our simulated likelihood analyses, we employ a linearized model of the data vector ξ (e.g. the DES-Y3 3×2 -point function data vector). This can be considered a linear Taylor expansion of our full model around a fiducial set of parameters π^0 that is summarized in Table 3. In this approximation, our model data vector becomes

$$\hat{\xi}(\pi) = \xi(\pi^0) + \sum_{\alpha} (\pi_{\alpha} - \pi_{\alpha}^0) \left. \frac{\partial \xi(\pi)}{\partial \pi_{\alpha}} \right|_{\pi = \pi^0}, \quad (29)$$

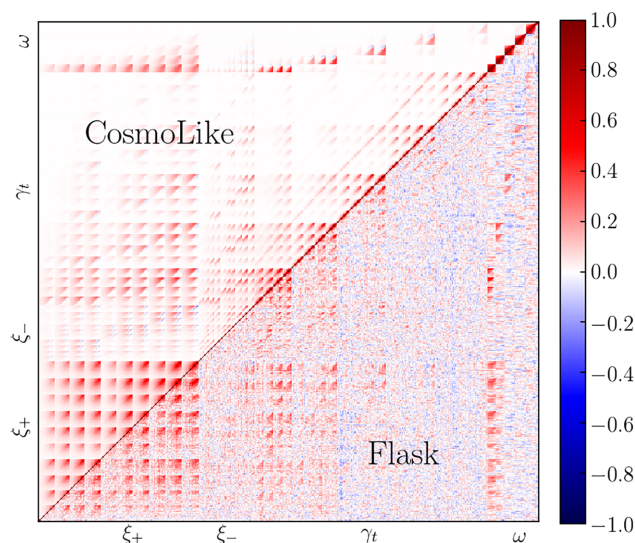


Figure 6. FLASK (lower diagonal) versus CosmoLike halo model (upper diagonal) correlation matrix.

Table 3. Fiducial cosmology and standard deviation of Gaussian parameter priors used in our mock likelihood analyses. $A_{IA, i}$ is the intrinsic alignment amplitude in the i -th source redshift bin, m_i is the multiplicative shear bias, and $\Delta z_{s, i}$ parametrizes systematic shifts in the photometric redshift distribution of that bin. $\Delta z_{l, i}$ parametrizes systematic shifts in the photometric redshift distribution of the i -th lens redshift bin. The Gaussian priors we choose for the parameters follow the analysis choices of Abbott et al. (2018) and we assume infinite flat priors for all other parameters.

Parameter	Fiducial value	σ_{prior}
Cosmology		
Ω_m	0.3	–
σ_8	0.823 55	–
h_{100}	0.69	–
n_s	0.97	–
w_0	–1	–
Ω_b	0.048	–
Ω_ν	0.001 743	–
Ω_Λ	$1 - \Omega_m - \Omega_\nu$	–
b_1	1.7	–
b_2	1.7	–
b_3	1.7	–
b_4	2.0	–
b_5	2.0	–
$\Delta z_{l, 1}$	0.0	0.04
$\Delta z_{l, 2}$	0.0	0.04
$\Delta z_{l, 3}$	0.0	0.04
$\Delta z_{l, 4}$	0.0	0.04
$\Delta z_{l, 5}$	0.0	0.04
$\Delta z_{s, 1}$	0.0	0.08
$\Delta z_{s, 2}$	0.0	0.08
$\Delta z_{s, 3}$	0.0	0.08
$\Delta z_{s, 4}$	0.0	0.08
$A_{IA, 1}$	0.0	–
$A_{IA, 2}$	0.0	–
$A_{IA, 3}$	0.0	–
$A_{IA, 4}$	0.0	–
m_1	0.0	0.03
m_2	0.0	0.03
m_3	0.0	0.03
m_4	0.0	0.03

where the sum is over all components π_α of the parameter vector $\boldsymbol{\pi}$ (we will use Latin indices for the components of the data vector and Greek indices for the components of the parameter vector). Given a two-point function measurement $\hat{\boldsymbol{\xi}}$ and abbreviating

$$\begin{aligned}\xi^0 &= \boldsymbol{\xi}(\boldsymbol{\pi}^0) \\ \delta\boldsymbol{\xi} &= \hat{\boldsymbol{\xi}} - \xi^0 \\ \delta\boldsymbol{\pi} &= \boldsymbol{\pi} - \boldsymbol{\pi}^0 \\ \partial_\alpha \boldsymbol{\xi} &= \left. \frac{\partial \boldsymbol{\xi}(\boldsymbol{\pi})}{\partial \pi_\alpha} \right|_{\boldsymbol{\pi}=\boldsymbol{\pi}^0}\end{aligned}$$

our figure of merit χ^2 as a function of the parameters becomes in the linearized approximation

$$\begin{aligned}\chi^2[\delta\boldsymbol{\pi}] &= \left(\delta\boldsymbol{\xi} - \sum_\alpha \delta\pi_\alpha \partial_\alpha \boldsymbol{\xi} \right)^T \mathbf{C}^{-1} \left(\delta\boldsymbol{\xi} - \sum_\alpha \delta\pi_\alpha \partial_\alpha \boldsymbol{\xi} \right) \\ &+ (\boldsymbol{\pi} - \boldsymbol{\pi}^{\text{prior}})^T \mathbf{P} (\boldsymbol{\pi} - \boldsymbol{\pi}^{\text{prior}}).\end{aligned}\quad (30)$$

Here, we have allowed for a Gaussian prior with covariance matrix \mathbf{P}^{-1} and central value $\boldsymbol{\pi}^{\text{prior}}$. To find the deviation $\delta\boldsymbol{\pi}^{\text{MP}} = \boldsymbol{\pi}^{\text{MP}} - \boldsymbol{\pi}^0$ from our fiducial parameters that minimizes this function (the maximum posterior value of the parameters is denoted by $\boldsymbol{\pi}^{\text{MP}}$), we have to solve

$$\left. \frac{\partial \chi^2}{\partial (\delta\pi_\beta)} \right|_{\delta\boldsymbol{\pi}=\delta\boldsymbol{\pi}^{\text{MP}}} = 0.\quad (31)$$

Defining a vector \mathbf{x} such that $x_\beta = \delta\boldsymbol{\xi}^T \mathbf{C}^{-1} \partial_\beta \boldsymbol{\xi}$ as well as the Fisher matrix $F_{\alpha\beta} = \partial_\beta \boldsymbol{\xi}^T \mathbf{C}^{-1} \partial_\alpha \boldsymbol{\xi}$, this becomes

$$\begin{aligned}(\mathbf{F} + \mathbf{P}) \delta\boldsymbol{\pi}^{\text{MP}} &= \mathbf{x} + \mathbf{P} (\boldsymbol{\pi}^{\text{prior}} - \boldsymbol{\pi}^0) \\ \Rightarrow \boldsymbol{\pi}^{\text{MP}} &= \boldsymbol{\pi}^0 + (\mathbf{F} + \mathbf{P})^{-1} \mathbf{x} + (\mathbf{F} + \mathbf{P})^{-1} \mathbf{P} (\boldsymbol{\pi}^{\text{prior}} - \boldsymbol{\pi}^0).\end{aligned}\quad (32)$$

We now want to consider the situation when a model covariance matrix \mathbf{C}_{mod} is used to calculate the likelihood in equation (30) that is different from the true covariance matrix \mathbf{C}_{true} of the statistical uncertainties in the data vector $\hat{\boldsymbol{\xi}}$. In that case, our linearized likelihood will be a Gaussian centred around $\boldsymbol{\pi}^{\text{MP}}$ and with parameter covariance matrix

$$\mathbf{C}_{\boldsymbol{\pi}, \text{like}} = (\mathbf{F}_{\text{mod}} + \mathbf{P})^{-1},\quad (33)$$

where $F_{\text{mod}, \alpha\beta} = \partial_\beta \boldsymbol{\xi}^T \mathbf{C}_{\text{mod}}^{-1} \partial_\alpha \boldsymbol{\xi}$ is the Fisher matrix calculated from the model covariance.

The actual covariance matrix of $\boldsymbol{\pi}^{\text{MP}}$ includes two sources of noise. First, statistical uncertainties in the measurement $\hat{\boldsymbol{\xi}}$ that are described by the covariance matrix \mathbf{C}_{true} and are represented by the first term in equation (32) that is proportional to \mathbf{x} , and secondly, statistical uncertainties in our choice of the prior centre that are described by the prior covariance matrix \mathbf{P}^{-1} and are represented by the second term in equation (32) that is proportional to $\boldsymbol{\pi}^{\text{prior}}$. The latter term has the covariance matrix $(\mathbf{F}_{\text{mod}} + \mathbf{P})^{-1} \mathbf{P} (\mathbf{F}_{\text{mod}} + \mathbf{P})^{-1}$ (because the covariance matrix of $\boldsymbol{\pi}^{\text{prior}}$ is \mathbf{P}^{-1}). Hence, the total covariance matrix of $\boldsymbol{\pi}^{\text{MP}}$ can be written as

$$\begin{aligned}(\mathbf{C}_{\boldsymbol{\pi}, \text{MP}})_{\alpha\beta} &\equiv \text{Cov} [\pi_\alpha^{\text{MP}}, \pi_\beta^{\text{MP}}] \\ &= (\mathbf{F}_{\text{mod}} + \mathbf{P})^{-1} \mathbf{P} (\mathbf{F}_{\text{mod}} + \mathbf{P})^{-1} + \sum_{\kappa, \lambda} (\mathbf{F}_{\text{mod}} + \mathbf{P})_{\alpha\kappa}^{-1} (\mathbf{F}_{\text{mod}} + \mathbf{P})_{\lambda\beta}^{-1} \sum_{i, k} \partial_\kappa \xi_i (\mathbf{C}_{\text{mod}}^{-1} \mathbf{C}_{\text{true}} \mathbf{C}_{\text{mod}}^{-1})_{ik} \partial_\lambda \xi_k.\end{aligned}\quad (34)$$

For $\mathbf{C}_{\text{mod}} = \mathbf{C}_{\text{true}}$, it is easy to see that this parameter covariance $\mathbf{C}_{\boldsymbol{\pi}, \text{MP}}$ equals the covariance $\mathbf{C}_{\boldsymbol{\pi}, \text{like}}$ that describes the shape of our likelihood (as it should).

5.2 Impact on the width of the likelihood and scatter of best-fitting parameters

We can use the above findings to study the impact of different effects in covariance modelling on parameter constraints. If a covariance matrix \mathbf{C}_1 contains a noise contribution that is missing in another covariance matrix \mathbf{C}_2 , then we quantify the difference between these matrices by considering the following two effects:

(i) *Width of likelihood contours*: Denoting the Fisher matrices obtained from \mathbf{C}_1 or \mathbf{C}_2 as \mathbf{F}_1 and \mathbf{F}_2 , respectively, the widths of likelihood contours drawn from the different covariances are given by

$$\begin{aligned}\mathbf{C}_{\boldsymbol{\pi}, \text{like}, 1} &= (\mathbf{F}_1 + \mathbf{P})^{-1} \\ \mathbf{C}_{\boldsymbol{\pi}, \text{like}, 2} &= (\mathbf{F}_2 + \mathbf{P})^{-1}.\end{aligned}\quad (35)$$

Hence, if the difference $\mathbf{C}_1 - \mathbf{C}_2 = \mathbf{E}$ represents noise contributions missing from (or miss-estimated in \mathbf{C}_2), then a comparison of $\mathbf{C}_{\boldsymbol{\pi}, \text{like}, 1}$ and $\mathbf{C}_{\boldsymbol{\pi}, \text{like}, 2}$ quantifies the impact of this on the width of parameter contours.

(ii) *Scatter in the centre of likelihood contours*: If the data vector $\hat{\boldsymbol{\xi}}$ had \mathbf{C}_1 as its true covariance matrix but \mathbf{C}_2 would be used to derive the maximum posterior parameters $\boldsymbol{\pi}^{\text{MP}}$ from it, then the maximum posterior parameter covariance would be given by

$$(\mathbf{C}_{\boldsymbol{\pi}, \text{MP}, 2})_{\alpha\beta} = (\mathbf{F}_2 + \mathbf{P})^{-1} \mathbf{P} (\mathbf{F}_2 + \mathbf{P})^{-1} + \sum_{\kappa, \lambda} (\mathbf{F}_2 + \mathbf{P})_{\alpha\kappa}^{-1} (\mathbf{F}_2 + \mathbf{P})_{\lambda\beta}^{-1} \sum_{i, k} \partial_\kappa \xi_i (\mathbf{C}_2^{-1} \mathbf{C}_1 \mathbf{C}_2^{-1})_{ik} \partial_\lambda \xi_k.\quad (36)$$

If the difference $\mathbf{C}_1 - \mathbf{C}_2 = \mathbf{E}$ represents noise contributions missing from (or miss-estimated in \mathbf{C}_2), then a comparison of $\mathbf{C}_{\pi, \text{MP}, 2}$ and $\mathbf{C}_{\pi, \text{MP}, 1} \equiv \mathbf{C}_{\pi, \text{like}, 1}$ quantifies the impact of this on the scatter in the location of parameter contours.

An inaccurate covariance model will in general have a different impact on the width and the location of parameter contours. Hence, in order to quantify the importance of different effects in covariance modelling for parameter estimation, we compare both the pair $\mathbf{C}_{\pi, \text{like}, 1} / \mathbf{C}_{\pi, \text{like}, 2}$ and the pair $\mathbf{C}_{\pi, \text{MP}, 1} / \mathbf{C}_{\pi, \text{MP}, 2}$.

5.3 Distribution of χ^2 when fitting for parameters

Within the linearized likelihood model developed in the previous section, we now investigate how errors in the covariance model impact the distribution of χ_{MP}^2 between measured data vector $\hat{\xi}$ and a maximum posterior model $\xi_{\text{MP}} = \xi(\pi^{\text{MP}})$,

$$\hat{\chi}_{\text{MP}}^2 = (\hat{\xi} - \xi_{\text{MP}})^T \mathbf{C}^{-1} (\hat{\xi} - \xi_{\text{MP}}). \quad (37)$$

We start with the case that

- (i) the true covariance \mathbf{C} of $\hat{\xi}$ is known;
- (ii) no parameter priors are used when determining the best-fitting model ξ_{MP} ;
- (iii) the true expectation value $\bar{\xi} \equiv \langle \hat{\xi} \rangle$ lies within our parameter space; i.e. there are parameters π^{true} such that $\xi(\pi^{\text{true}}) = \bar{\xi}$.

We will show that, as expected, in this case $\hat{\chi}_{\text{MP}}^2$ should follow a χ^2 distribution with $N_{\text{data}} - N_{\text{param}}$ degrees of freedom.

Using equations (29) and (32) (and setting again $\delta\xi \equiv \hat{\xi} - \xi^0$), one can see that the maximum posterior data vector is given by

$$\begin{aligned} \xi_{\text{MP}} &= \xi^0 + \sum_{\alpha\beta} \partial_\alpha \xi (\mathbf{F}^{-1})_{\alpha\beta} (\delta\xi^T \mathbf{C}^{-1} \partial_\beta \xi) \\ &= \bar{\xi} + \sum_{\alpha\beta} \partial_\alpha \xi (\mathbf{F}^{-1})_{\alpha\beta} \left((\hat{\xi} - \bar{\xi})^T \mathbf{C}^{-1} \partial_\beta \xi \right) \\ &= \bar{\xi} + \sum_{\alpha\beta} \sum_{kl} \partial_\alpha \xi (\mathbf{F}^{-1})_{\alpha\beta} (\hat{\xi}_k - \bar{\xi}_k) (\mathbf{C}^{-1})_{kl} \partial_\beta \xi_l \\ &\equiv \bar{\xi} + \mathcal{P} \cdot (\hat{\xi} - \bar{\xi}). \end{aligned} \quad (38)$$

Here, the second line follows from the fact that $\bar{\xi} = \langle \hat{\xi} \rangle = \langle \xi_{\text{MP}} \rangle$ and we have defined the matrix

$$\mathcal{P}_{ij} = \sum_{\alpha\beta} \partial_\alpha \xi_i \sum_l (\mathbf{F}^{-1})_{\alpha\beta} (\mathbf{C}^{-1})_{lj} \partial_\beta \xi_l. \quad (39)$$

It can be shown that \mathcal{P} is an idempotent matrix ($\mathcal{P}^2 = \mathcal{P}$) and furthermore that

$$\begin{aligned} \text{Trace}(\mathcal{P}) &= N_{\text{param}} \\ \mathbf{C}^{-1} \mathcal{P} \mathbf{C} &= \mathcal{P}^T. \end{aligned} \quad (40)$$

The residual between the measurement $\hat{\xi}$ and the best-fitting model ξ_{MP} can be written in terms of \mathcal{P} as

$$\begin{aligned} \hat{\xi} - \xi_{\text{MP}} &= (\hat{\xi} - \bar{\xi}) - (\xi_{\text{MP}} - \bar{\xi}) \\ &= (\mathbb{1} - \mathcal{P}) \cdot (\hat{\xi} - \bar{\xi}). \end{aligned} \quad (41)$$

Hence, the covariance matrix of $\hat{\xi} - \xi_{\text{MP}}$ is given by

$$\mathbf{C}_{\mathcal{P}} \equiv \langle (\hat{\xi} - \xi_{\text{MP}})^T (\hat{\xi} - \xi_{\text{MP}}) \rangle = (\mathbb{1} - \mathcal{P}) \mathbf{C} (\mathbb{1} - \mathcal{P})^T. \quad (42)$$

This makes it straightforward to find the expectation value

$$\begin{aligned} \langle \chi_{\text{MP}}^2 \rangle &= \langle (\hat{\xi} - \xi_{\text{MP}})^T \mathbf{C}^{-1} (\hat{\xi} - \xi_{\text{MP}}) \rangle \\ &= \text{Trace}(\mathbf{C}_{\mathcal{P}} \mathbf{C}^{-1}) \\ &= \sum_{jk} C_{kj} (\mathbf{C}^{-1})_{jk} - \sum_k \mathcal{P}_{kk} \\ &= N_{\text{data}} - N_{\text{param}}. \end{aligned} \quad (43)$$

Similarly, the variance of χ_{MP}^2 can be shown to be

$$\begin{aligned} \text{Var}(\chi_{\text{MP}}^2) &= \langle (\chi_{\text{MP}}^2)^2 \rangle - \langle \chi_{\text{MP}}^2 \rangle^2 \\ &= 2 \text{Trace}([\mathbf{C}_{\mathcal{P}} \mathbf{C}^{-1}]^2) \\ &= 2(N_{\text{data}} - N_{\text{param}}). \end{aligned} \quad (44)$$

So far, we have only re-derived textbook results (Anderson 2003). Now how do $\langle \chi_{\text{MP}}^2 \rangle$ and $\text{Var}(\chi_{\text{MP}}^2)$ change if the covariance model \mathbf{C}_{mod} we use to find the best-fitting model ξ_{MP} and to compute χ_{MP}^2 is different from the true covariance matrix \mathbf{C} of $\hat{\xi}$?

Following similar steps as equations (38) and (39), one can show that

$$\xi_{\text{MP}} = \bar{\xi} + \mathcal{P}_{\text{mod}} \cdot (\hat{\xi} - \bar{\xi}), \quad (45)$$

where

$$(\mathcal{P}_{\text{mod}})_{ij} = \sum_{\alpha\beta} \partial_{\alpha} \xi_i \sum_l (\mathbf{F}_{\text{mod}}^{-1})_{\alpha\beta} (\mathbf{C}_{\text{mod}}^{-1})_{lj} \partial_{\beta} \xi_l \quad (46)$$

and where the Fisher matrix \mathbf{F}_{mod} is computed from the model covariance \mathbf{C}_{mod} . Equation (45) especially shows that ξ_{MP} is still an unbiased estimator of $\bar{\xi}$ even when $\mathbf{C}_{\text{mod}} \neq \mathbf{C}$. When deriving the moments of χ_{MP}^2 , we will still come across expectation values like (cf. equation 43)

$$\langle (\hat{\xi}_i - \bar{\xi}_i)(\hat{\xi}_j - \bar{\xi}_j) \rangle \equiv (\mathbf{C})_{ij} \neq (\mathbf{C}_{\text{mod}})_{ij}. \quad (47)$$

Hence, the expectation value and variance of χ_{MP}^2 are given by

$$\langle \chi_{\text{MP}}^2 \rangle = \text{Trace} (\mathbf{C}_{\mathcal{P}_{\text{mod}}} \mathbf{C}_{\text{mod}}^{-1}) \quad (48)$$

$$\text{Var} (\chi_{\text{MP}}^2) = 2 \text{Trace} \left([\mathbf{C}_{\mathcal{P}_{\text{mod}}} \mathbf{C}_{\text{mod}}^{-1}]^2 \right), \quad (49)$$

where

$$\mathbf{C}_{\mathcal{P}_{\text{mod}}} = (\mathbb{1} - \mathcal{P}_{\text{mod}}) \mathbf{C} (\mathbb{1} - \mathcal{P}_{\text{mod}})^T. \quad (50)$$

Now we are left to investigate how equations (48) and (49) change when a Gaussian parameter prior \mathbf{P} is included in the likelihood function (cf. equation 30). A complication in this case is that now ξ_{MP} is not necessarily an unbiased estimate of $\bar{\xi}$ anymore. This is because in equation (30) we have centred our prior around the model parameters π^{prior} that may be different from the true parameters π^{true} . Inserting the full expression for the maximum posterior parameters (equation 32) into our linearized model, we now get

$$\xi_{\text{MP}} = \xi^0 + \mathcal{P}_{\text{mod}} \cdot (\hat{\xi} - \xi^0) + \zeta \quad (51)$$

with

$$\begin{aligned} (\mathcal{P}_{\text{mod}})_{ij} &= \sum_{\alpha\beta} \partial_{\alpha} \xi_i \sum_l (\mathbf{F}_{\text{mod}} + \mathbf{P})_{\alpha\beta}^{-1} (\mathbf{C}_{\text{mod}}^{-1})_{lj} \partial_{\beta} \xi_l \\ \zeta &= \sum_{\alpha} [(\mathbf{F}_{\text{mod}} + \mathbf{P})^{-1} \mathbf{P} (\pi^{\text{prior}} - \pi^0)]_{\alpha} \partial_{\alpha} \xi. \end{aligned} \quad (52)$$

The residual between $\hat{\xi}$ and ξ_{MP} hence becomes

$$\hat{\xi} - \xi_{\text{MP}} = (\mathbb{1} - \mathcal{P}_{\text{mod}}) \cdot (\hat{\xi} - \xi^0) - \zeta. \quad (53)$$

Treating the prior centre π^{prior} again as a random vector centred around π^{true} , ζ also becomes a random vector with covariance

$$\begin{aligned} (\mathbf{C}_{\zeta})_{ij} &\equiv \text{Cov}[\zeta_i, \zeta_j] \\ &= \sum_{\alpha\beta\gamma\delta} \partial_{\alpha} \xi_i (\mathbf{F}_{\text{mod}} + \mathbf{P})_{\alpha\beta}^{-1} \mathbf{P}_{\beta\gamma} (\mathbf{F}_{\text{mod}} + \mathbf{P})_{\gamma\delta}^{-1} \partial_{\delta} \xi_j. \end{aligned} \quad (54)$$

Hence, along lines similar to the case without a prior, we can write the moments of χ_{MP}^2 for a given model covariance as

$$\langle \chi_{\text{MP}}^2 \rangle = \text{Trace} \left(\{ \mathbf{C}_{\mathcal{P}_{\text{mod}}} + \mathbf{C}_{\zeta} \} \mathbf{C}_{\text{mod}}^{-1} \right) \quad (55)$$

$$\text{Var} (\chi_{\text{MP}}^2) = 2 \text{Trace} \left(\left[\{ \mathbf{C}_{\mathcal{P}_{\text{mod}}} + \mathbf{C}_{\zeta} \} \mathbf{C}_{\text{mod}}^{-1} \right]^2 \right). \quad (56)$$

Notice that in the absence of priors $\mathbf{C}_{\zeta} = 0$ and for the true covariance \mathbf{C} , we recover equations (43) and (44) as expected. Equations (55) and (56) are used to produce our main result shown in Fig. 1 for different covariance matrices.

6 EXPLORING DIFFERENT EFFECTS IN THE COVARIANCE MODELLING

Our main goal is to study the impact of including different effects in the covariance modelling on the estimation of parameters. Several covariance matrices were generated and tested under different assumptions and approximations. The main results were already shown in Section 2. We now present the details of each step in the validation strategy that was outlined in Section 5.

6.1 Gaussian likelihood assumption

A basic assumption of our framework of testing different covariance matrices is that the likelihood function of the data is Gaussian. One simple reason of why the sampling distribution of the correlation functions cannot be an exact multivariate Gaussian is that this violates the positivity constraint of the power spectrum (Schneider & Hartlap 2009). There are also other reasons described below. The purpose of this subsection is to assess the impact of non-Gaussianity of the likelihood of two-point functions in the parameter estimation. In this sense, checking this basic assumption is a test of the whole framework and is different from the robustness tests for the covariance matrix modelling described in the remaining subsections of this section.

The impact of a non-Gaussian likelihood in parameter estimation of weak lensing correlation functions has been recently studied in Lin et al. (2020) where no significant biases were found in one-dimensional posteriors of Ω_m and σ_8 between the multivariate Gaussian likelihood model and more complex non-Gaussian likelihood models. Also, in Sellentin, Heymans & Harnois-Déraps (2018) the skewed distributions of weak lensing shear correlation functions are used to derive an analytical expression for a non-Gaussian likelihood.

We first consider a full-sky survey such that each of our two-point function estimators $\hat{\xi}^{AB}(\theta)$ is a harmonic transform of a harmonic space estimator \hat{C}_ℓ^{AB} (cf. equation 12), i.e.

$$\hat{\xi}^{AB}(\theta) = \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} F_\ell^{AB}(\theta) \hat{C}_\ell^{AB}. \quad (57)$$

Each \hat{C}_ℓ^{AB} is given in terms of the spherical harmonics coefficients $a_{\ell m}$, $b_{\ell m}$ of two Gaussian random fields as

$$\hat{C}_\ell^{AB} = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} a_{\ell m} b_{\ell m}^*. \quad (58)$$

The product of two Gaussian random variables does not follow a Gaussian distribution. Therefore, in principle one would not expect \hat{C}_ℓ^{AB} [and consequently $\hat{\xi}^{AB}(\theta)$] to have a Gaussian likelihood. However, at small scales, i.e. at high multipoles ℓ , the sum of the random variables $a_{\ell m} b_{\ell m}^*$ in equation (58) will approach a Gaussian distribution by means of the central limit theorem, since there are a large number ($2\ell+1$) of independent modes. It should be pointed out that at these small scales the galaxy density and shear fields characterized by $a_{\ell m}$ and $b_{\ell m}$ are themselves non-Gaussian due to the non-linear evolution of gravity.

It is hence our working hypothesis that non-Gaussianity of \hat{C}_ℓ^{AB} only matters at the largest scales (small ℓ 's) where both $a_{\ell m}$ and $b_{\ell m}$ can be considered Gaussian random variables but not their product. In the full-sky case, it can then be shown that the second and third central moments of \hat{C}_ℓ^{AB} are given by

$$\langle (\hat{C}_\ell^{AB} - C_\ell^{AB})^2 \rangle = \frac{[(C_\ell^{AB})^2 + C_\ell^{AA} C_\ell^{BB}]}{2\ell+1} \quad (59)$$

$$\langle (\hat{C}_\ell^{AB} - C_\ell^{AB})^3 \rangle = \frac{2[(C_\ell^{AB})^3 + 3C_\ell^{AA} C_\ell^{BB} C_\ell^{AB}]}{(2\ell+1)^2}. \quad (60)$$

If only a fraction f_{sky} of the sky is being observed, these moments get divided by f_{sky} and f_{sky}^2 , respectively.

Assuming different multipoles to be uncorrelated, the corresponding moments of $\hat{\xi}^{AB}(\theta)$ can be computed as

$$\langle (\hat{\xi}^{AB}(\theta) - \xi^{AB}(\theta))^2 \rangle = \sum_{\ell=0}^{\infty} \left(\frac{2\ell+1}{4\pi} F_\ell^{AB}(\theta) \right)^2 \langle (\hat{C}_\ell^{AB} - C_\ell^{AB})^2 \rangle \quad (61)$$

$$\langle (\hat{\xi}^{AB}(\theta) - \xi^{AB}(\theta))^3 \rangle = \sum_{\ell=0}^{\infty} \left(\frac{2\ell+1}{4\pi} F_\ell^{AB}(\theta) \right)^3 \langle (\hat{C}_\ell^{AB} - C_\ell^{AB})^3 \rangle. \quad (62)$$

Equation (61) is of course nothing but the diagonal of the covariance matrix (cf. equation 14).

The dominant effect of the non-Gaussianity of the C_ℓ 's is a positive skewness in the distribution of our data vectors (Sellentin et al. 2018). To estimate its impact on our parameter constraints, we approximate the entire distribution of our 3×2 pt data vector by a multivariate lognormal distribution. The covariance of our data vector and the skewness of each data point as given by equation (62) are sufficient to fix the parameters of a shifted lognormal distribution. We have already discussed this in Sections 4.2 and 4.3, though with a conceptual difference: In that section, we describe how to configure lognormal simulations of the cosmic density field, while here we assume measurements of the 3×2 -point functions to have a multivariate lognormal distribution. To be explicit, we fix the shift parameters $\lambda(\theta)$ that enter the lognormal PDF of the measurements $\hat{\xi}^{AB}(\theta)$ in the different angular bins (cf. equation 24 for the definition of λ) via the equation

$$\langle (\hat{\xi}^{AB}(\theta) - \xi^{AB}(\theta))^3 \rangle = \frac{3 \langle (\hat{\xi}^{AB}(\theta) - \xi^{AB}(\theta))^2 \rangle^2}{\lambda(\theta)} + \frac{\langle (\hat{\xi}^{AB}(\theta) - \xi^{AB}(\theta))^2 \rangle^3}{\lambda(\theta)^3}, \quad (63)$$

which relates the second and third central moments of lognormal random variables (Hilbert et al. 2011).

In the top panel of Fig. 7, we show the impact of this non-Gaussianity on the distribution of maximum posterior χ^2 . For that figure, we generated 300 000 random realizations of our fiducial data vector from a multivariate Gaussian distribution, 300 000 random realizations of that data vector from a multivariate lognormal distribution, and 300 000 random realizations from another lognormal distribution, whose skewness

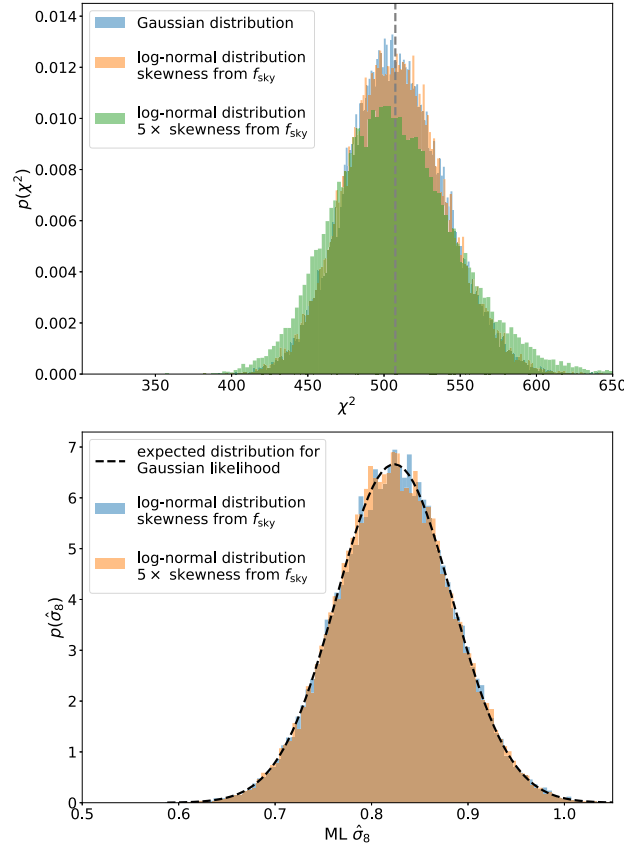


Figure 7. Top panel: Distribution of χ^2 when drawing 3×2 pt data vectors from a Gaussian distribution (blue histogram), from a shifted lognormal distribution where the skewness of each data point was computed in the f_{sky} approximation (red histogram) and when assuming that the skewness of the data points is 5 times that of the f_{sky} approximation (green histogram). Bottom panel: Distribution of maximum posterior σ_8 when fitting the linearized model to Section 5.1 Gaussian realizations of our fiducial data vector, to lognormal realizations of our fiducial data vector (blue histogram), and to lognormal realizations with 5 times the skewness of the f_{sky} approximation employed in Section 6.1 (orange histogram).

in each data point was increased by a factor of 5. For each of these random realizations, we analytically determined the maximum posterior model within the linearized likelihood formalism of Section 5.1 and then computed the χ^2 between that model and the random realization. The blue histogram in the top panel of Fig. 7 shows the distribution of these χ^2 values for the Gaussian random realizations and the red histogram corresponds to the lognormal random realizations. The two histograms are almost identical. Hence, within the f_{sky} approximation employed above non-Gaussianity in the likelihood does not seem to affect our analysis. Also, even in the extreme scenario of enhancing the skewness of the data vector by a factor of 5 (green histogram) the increase in the scatter of χ^2 remains smaller than about 3 per cent of the average χ^2 – which still would not dominate over the other effects discussed in subsequent sections (cf. Fig. 1).

The impact of non-Gaussianity on the likelihood becomes even more negligible when directly considering the distribution of maximum posterior parameters. We demonstrate this in the bottom panel of Fig. 7 for the best-fitting values of σ_8 but find similar results for our other key cosmological parameters Ω_m and w_0 . Therefore, we conclude that it is safe to assume a Gaussian distribution for the statistical uncertainties of the DES-Y3 two-point function measurements.

6.2 Modelling of connected four-point function in covariance

The connected four-point contribution to the covariance is the part that is most challenging to model analytically (Schneider et al. 2002; Hilbert et al. 2011; Sato et al. 2011; Takada & Hu 2013). This contribution is most relevant at small scales and turns out to be a small one for current LSS analyses (Krause et al. 2017; Barreira et al. 2018). This is for two reasons: (1) such analyses typically cut away their smallest scales because of uncertainties in the modelling of their data vectors and (2) at small scales the covariance matrix is often dominated by shape noise and shot noise that are believed to be well understood.

We test whether the non-Gaussian covariance parts (by which we mean both the connected four-point function and supersample covariance) are a relevant contribution to our error budget by either

- (i) replacing the non-Gaussian contributions from the fiducial halo model with the lognormal covariance described in Section 4.2.
- (ii) or setting it to zero, i.e. using only a Gaussian covariance matrix.

Fig. 1 and Table 1 show that neither of these changes has a significant impact on the distribution of χ^2 and our parameter constraints. Assuming that our halo model and lognormal recipes do not underestimate the non-Gaussian covariance parts by orders of magnitude [see e.g. Sato et al. (2009) and Hilbert et al. (2011) for justifications of this assumption], this demonstrates that we are insensitive to the exact modelling of these contributions. At the same time, we want to stress that this finding holds for the specific scale cuts, redshift distributions, and tracer densities of the DESY3 3×2 pt analysis and cannot necessarily be generalized to other analysis set-ups.

6.3 Exact angular bin averaging

Equation (14) holds when measuring the two-point correlation functions in infinitesimally small bins around the angular scales θ_1 and θ_2 . This is unfeasible in practice and in fact also leads to divergent covariance matrices. This can, for example, be seen for the galaxy clustering correlation functions, where the constant term proportional to $1/n_g^2$ in the harmonic space covariance gives a contribution to the real space covariance of

$$\frac{1}{4\pi^2 n_g^2 f_{\text{sky}}} \lim_{N \rightarrow \infty} \sum_{\ell=1}^N \frac{(2\ell+1)}{2} P_\ell(\cos\theta)^2 \rightarrow \frac{1}{4\pi^2 n_g^2 f_{\text{sky}}} \delta_D(\cos\theta - \cos\theta) (= \infty).$$

The reason for this divergence is simply the fact that the number of galaxy pairs found in an infinitesimal bin vanishes, leading to infinite shot noise. This problem disappears when considering finite angular bins.

To analytically average over a finite angular bin $[\theta_{\min}, \theta_{\max}]$, we assume that the number of galaxy pairs with angular separation θ is proportional to $\sin\theta$ (corresponding to a uniform distribution of galaxies on the sky). We then replace the functions $F_\ell^{AB}(\theta)$ in equations (9) and (10) by

$$F_\ell^{AB}(\theta) \rightarrow \frac{1}{\cos\theta_{\min} - \cos\theta_{\max}} \int_{\theta_{\min}}^{\theta_{\max}} d\theta \sin\theta F_\ell^{AB}(\theta). \quad (64)$$

For the galaxy clustering correlation function $w(\theta)$, this leads to

$$P_\ell(\cos\theta) \rightarrow \frac{[P_{\ell+1}(x) - P_{\ell-1}(x)]_{\cos\theta_{\max}}^{\cos\theta_{\min}}}{(2\ell+1)(\cos\theta_{\min} - \cos\theta_{\max})}. \quad (65)$$

The corresponding expressions for the galaxy–galaxy lensing correlation function $\gamma_\ell(\theta)$ and for the cosmic shear correlation functions ξ_\pm are presented (together with derivations of all the bin averaged expressions) in Appendix B.

We show below how the bin averaging solves the problem of diverging diagonal values of the covariance for $w(\theta)$. This can be seen from

$$\begin{aligned} \sum_\ell \frac{([P_{\ell+1}(x) - P_{\ell-1}(x)]_{\cos\theta_{\max}}^{\cos\theta_{\min}})^2}{2(2\ell+1)f_{\text{sky}}(n_g A_{\text{bin}})^2} &= \int_{\cos\theta_{\max}}^{\cos\theta_{\min}} dx_1 \int_{\cos\theta_{\max}}^{\cos\theta_{\min}} dx_2 \sum_\ell \frac{2\ell+1}{2} \frac{P_\ell(x_1) P_\ell(x_2)}{f_{\text{sky}}(n_g A_{\text{bin}})^2} \\ &= \int_{\cos\theta_{\max}}^{\cos\theta_{\min}} dx_1 \int_{\cos\theta_{\max}}^{\cos\theta_{\min}} dx_2 \frac{\delta_D(x_1 - x_2)}{f_{\text{sky}}(n_g A_{\text{bin}})^2} \\ &= \frac{\cos\theta_{\min} - \cos\theta_{\max}}{f_{\text{sky}}(n_g A_{\text{bin}})^2} \\ &= \frac{2}{A_{\text{survey}} A_{\text{bin}} n_g^2} \\ &= \frac{1}{N_{\text{pair}}}, \end{aligned} \quad (66)$$

where $A_{\text{survey}} = 4\pi f_{\text{sky}}$ is the total survey area, $A_{\text{bin}} = 2\pi(\cos\theta_{\min} - \cos\theta_{\max})$ is the bin area, and N_{pair} the total number of galaxy pairs used to estimate \hat{w} . The above expression is the usual formula for the shot-noise part of the real space covariance.

The impact of the exact angular bin averaging for the noise and mixed terms in the Gaussian part of the covariance matrix is included for all four types of two-point functions present in the DES-Y3 data vector and the DES-Y3 fiducial covariance.

6.4 Flat versus curved sky

For the Y1 analysis, it was shown that the flat-sky approximation was valid for the galaxy–galaxy shear and shear–shear two-point correlation function (Krause et al. 2017). In Y3, the fiducial covariance computes the full sky correlations; see equations (12) and (13). We show in Fig. 1 that the effect of including curved sky results has negligible impact on the χ^2 distribution. Table 1 shows that this is also true for parameter constraints.

6.5 RSD and Limber approximation and RSD effects

The modelling of the angular power spectrum of two tracers involves a projection from the 3D power spectrum that requires integrals with integrands containing the product of two spherical Bessel functions, which are highly oscillatory. The inclusion of RSD effects in a simple linear

modelling (Kaiser 1987) involves the computation of those integrals with derivatives of the Bessel functions. These integrals are notoriously difficult to perform numerically and it is usual to apply the so-called Limber approximation (Limber 1953; LoVerde & Afshordi 2008). An efficient computation of these integrals without resorting to the Limber approximation was recently implemented in the case of the angular power spectrum for galaxy clustering in Fang et al. (2020b). We use their approach to study the impact of taking into account both non-Limber computations and RSD effects in the covariance matrix. Fig. 1 and Table 1 show that not taking these effects into account leads to an increase in average χ^2 of about 0.5 per cent and an underestimation of uncertainties in key cosmological parameters by 0.6 per cent to 1.4 per cent.

6.6 Effect of the mask geometry

The analytical covariance models described in Section 4 make use of the so-called f_{sky} approximation; i.e. they take the covariance of an all-sky survey and divide this by the sky fraction of DES-Y3 to approximate the covariance of our partial sky data. In Appendix C, we show how to go beyond this approximation. First, we note there that the covariance of the two-point function measurements between pairs of scalar random fields (δ_a, δ_b) and (δ_c, δ_d) within angular bins $[\theta_-^{ab}, \theta_+^{ab}]$ and $[\theta_-^{cd}, \theta_+^{cd}]$ is given by

$$\begin{aligned} \text{Cov} \{ \hat{\xi}^{ab} [\theta_-^{ab}, \theta_+^{ab}], \hat{\xi}^{cd} [\theta_-^{cd}, \theta_+^{cd}] \} &= \frac{N_{\text{pair}}^{ab} [\theta_-^{ab}, \theta_+^{ab}] N_{\text{pair}}^{cd} [\theta_-^{cd}, \theta_+^{cd}]}{n_a n_b n_c n_d} \\ &= (2\pi)^2 \sum_{\ell_1 \ell_2} [P_{\ell_1+1}(x) - P_{\ell_1-1}(x)]_{\theta_+^{ab}}^{\theta_-^{ab}} [P_{\ell_2+1}(x) - P_{\ell_2-1}(x)]_{\theta_+^{cd}}^{\theta_-^{cd}} \cdot \text{Cov} \{ \tilde{C}_{\ell_1}^{ab}, \tilde{C}_{\ell_2}^{cd} \}. \end{aligned} \quad (67)$$

Here, P_ℓ are again the Legendre polynomials and the angular bin averaging was already evaluated. The factor $N_{\text{pair}}^{ab} [\theta_-^{ab}, \theta_+^{ab}]$ (resp. $N_{\text{pair}}^{cd} [\theta_-^{cd}, \theta_+^{cd}]$) is the average number of pairs of random points that uniformly sample the footprint with densities n_a, n_b (resp. n_c, n_d) and whose separation falls into the angular bin $[\theta_-^{ab}, \theta_+^{ab}]$ (resp. $[\theta_-^{cd}, \theta_+^{cd}]$). Hence, these factors describe how the exact survey geometry suppresses the number of pairs of positions in the bins $[\theta_-^{ab}, \theta_+^{ab}]$ and $[\theta_-^{cd}, \theta_+^{cd}]$ with respect to the f_{sky} approximation. Also, finally, $\text{Cov}\{\tilde{C}_{\ell_1}^{ab}, \tilde{C}_{\ell_2}^{cd}\}$ is the covariance of pseudo- C_ℓ estimates of the power spectra between the fields (δ_a, δ_b) and (δ_c, δ_d) (see Appendix C for more details). Note that the full survey footprint modifies the covariance with respect to the f_{sky} approximation used in Section 3 both through the factors $N_{\text{pair}}^{ab} [\theta_-^{ab}, \theta_+^{ab}]/n_a n_b$, $N_{\text{pair}}^{cd} [\theta_-^{cd}, \theta_+^{cd}]/n_c n_d$ and by changing $\text{Cov}\{\hat{C}_{\ell_1}^{ab}, \hat{C}_{\ell_2}^{cd}\}$ compared to equations (16)–(23).

One can determine the factors $N_{\text{pair}}^{ab} [\theta_-^{ab}, \theta_+^{ab}]/n_a n_b$ and $N_{\text{pair}}^{cd} [\theta_-^{cd}, \theta_+^{cd}]/n_c n_d$ either by counting pairs in a set of random points that trace the survey footprint homogeneously or they can be calculated analytically from the power spectrum of the survey mask (see our Appendix C as well as Troxel et al. 2018b). This will generally lead to an enhancement of statistical uncertainties (i.e. of the covariance matrix) with respect to the f_{sky} approximation. To calculate $\text{Cov}\{\tilde{C}_{\ell_1}^{ab}, \tilde{C}_{\ell_2}^{cd}\}$, one could e.g. follow approximations made by Efstathiou (2004). We slightly modify their arguments in Appendix C to arrive at

$$\text{Cov} \{ \tilde{C}_{\ell_1}^{ab}, \tilde{C}_{\ell_2}^{cd} \} \approx \frac{1}{4} \left(\frac{C_{\ell_1}^{ac} C_{\ell_2}^{bd} + C_{\ell_2}^{ac} C_{\ell_1}^{bd} + C_{\ell_1}^{ac} C_{\ell_1}^{bd} + C_{\ell_2}^{ac} C_{\ell_2}^{bd}}{(2\ell_1 + 1)(2\ell_2 + 1)} + \frac{C_{\ell_1}^{ad} C_{\ell_2}^{bc} + C_{\ell_2}^{ad} C_{\ell_1}^{bc} + C_{\ell_1}^{ad} C_{\ell_1}^{bc} + C_{\ell_2}^{ad} C_{\ell_2}^{bc}}{(2\ell_1 + 1)(2\ell_2 + 1)} \right) \mathcal{M}_{\ell_1 \ell_2}. \quad (68)$$

Here, the mode coupling matrix $\mathcal{M}_{\ell_1 \ell_2}$ again depends on the power spectrum of the survey mask and is also detailed in Appendix C. Note that in order to keep our notation brief, we have assumed that the power spectra $C_{\ell_1}^{ac}$ etc. in the above equation include both the underlying cosmological power spectra and contributions to the power spectra from sampling noise, such as shape noise and shot noise.

In practice, equation (68) and the approximations proposed by Efstathiou (2004) yield very similar results and they are both valid on scales ℓ_1, ℓ_2 that are much smaller than the typical scales of the mask W . Unfortunately, the DES-Y3 analysis mask has features and holes over a large range of scales. Hence, the angular scales of interest in the $3 \times 2pt$ analysis are never strictly smaller than the scales of our mask. Hence, equation (68) is not sufficiently accurate in our case and in fact significantly overestimates our statistical uncertainties. In Fig. 8, we explain a simple scheme that can be used to correct for this. To motivate this procedure, consider how one would compute the Gaussian covariance model directly from the real space two-point correlation functions, i.e. without taking the detour to Fourier space that was used in Section 3. For the covariance of $\hat{\xi}^{ab} [\theta_-^{ab}, \theta_+^{ab}]$ and $\hat{\xi}^{cd} [\theta_-^{cd}, \theta_+^{cd}]$, this would amount to integration over all pairs of locations within our survey mask that fall into the angular bins $[\theta_-^{ab}, \theta_+^{ab}]$ and $[\theta_-^{cd}, \theta_+^{cd}]$. Schematically, this leads to terms of the form

$$\text{Cov} \propto \int_{(ab) \in \text{mask, bin}} d\Omega^a d\Omega^b \int_{(cd) \in \text{mask, bin}} d\Omega^c d\Omega^d \xi^{ac}(\theta^{ac}) \xi^{bd}(\theta^{bd}) + \dots \quad (69)$$

Here, $\Omega^a \dots \Omega^d$ are four locations inside the survey mask such that the distance between Ω^a and Ω^b lies inside the angular bin $[\theta_-^{ab}, \theta_+^{ab}]$ and the distance between Ω^c and Ω^d lies inside the angular bin $[\theta_-^{cd}, \theta_+^{cd}]$. Now the approximation of Efstathiou (2004) assumes that the two-point functions $\xi^{ac}(\theta)$, $\xi^{bd}(\theta)$ are negligible on scales θ comparable to the small-scale features in the mask. Schematically, this amounts to making the approximation

$$\int d\Omega^a \dots W(\Omega^a) \xi^{ac}(\theta^{ac}) \approx \bar{\xi}^{ac} \int d\Omega^a \dots W(\Omega^a) \delta_{\text{Dirac}}^2(\Omega^a - \Omega^c), \quad (70)$$

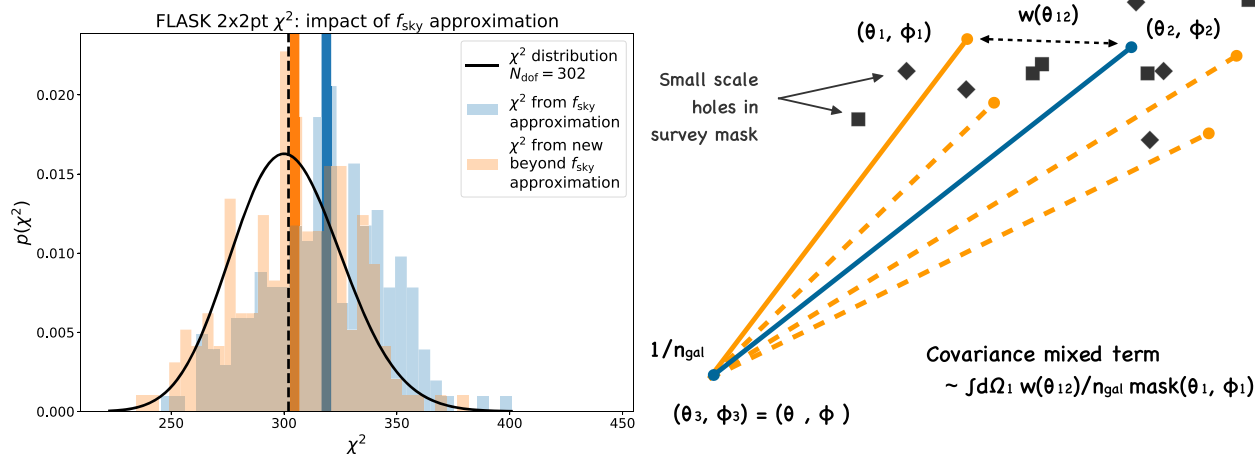


Figure 8. The impact of masking on the DES-Y3 covariance. The blue histogram in the right-hand panel shows the distribution of χ^2 obtained from our FLASK data vectors when using the f_{sky} approximation. We restrict this figure to the 2×2 pt function part of the data vector since it is this part that suffers the most from masking effects (cf. Fig. 1). Ansatzes in the CMB literature (e.g. Efstathiou 2004) are not sufficient to correct for this, because the DES footprint has features down to very small scales. In the main text, we have motivated a possible way to correct for these small-scale masking features and the orange histogram in the left-hand panel shows that this ansatz indeed significantly improves the χ^2 obtained from our FLASK measurements. The sketch in the right-hand panel visualizes how small-scale features in the mask lead to an overestimation in the covariance when using common ways to treat the impact of survey geometry on the two-point function covariance (see the main text for explanation).

where $\bar{\xi}^{ac}$ is a suitable average of the two-point function over different scales. Our understanding of the approximation of Efstathiou (2004) via equation (70) is based on findings that we present in Appendix D. This approximation fails when the mask contains features (e.g. holes) on scales where the two-point function has not yet decayed. Assuming that such small-scale holes cover a fraction of f_{mask} of a more coarse version of the footprint, this can roughly be corrected for with a multiplicative factor, i.e. by instead using the approximation

$$\int d\Omega^a \dots W(\Omega^a) \bar{\xi}^{ac}(\theta^{ac}) \approx f_{\text{mask}} \bar{\xi}^{ac} \int d\Omega^a \dots W(\Omega^a) \delta_{\text{Dirac}}^2(\Omega^a - \Omega^c). \quad (71)$$

The right-hand panel of Fig. 8 visualizes this for the mixed terms in the covariance, where one of the correlation functions ξ^{ac} or ξ^{bd} is due to sampling noise such as shape noise or shot noise and is hence exactly proportional to a Dirac delta function. In that case, the integration is over pairs that share one end point. Now, the approximation made e.g. in Efstathiou (2004) or by our equation (68) assumes that also the correlation function between the other two end points effectively acts as a delta function with respect to the smallest scale features in the survey mask (cf. equation 70). We find that this is not the case for the DES-Y3 mask and that it contains features on all scales relevant to our analysis. However, as indicated in equation (71), one can approximately correct for this by multiplying the mixed terms in the covariance by the fraction f_{mask} of the coarser survey geometry that is covered by small-scale holes in the mask. This can be considered a next-to-leading-order correction to our equation (68).

By applying equation (71) twice, one can see that the cosmic variance terms (terms where neither of the two-point functions ξ^{ac} or ξ^{bd} are exactly proportional to delta functions) can be corrected by multiplication with f_{mask}^2 . To implement this correction in practice, we draw circles within the DES-Y3 survey footprint with radii ranging from 5 to 20 arcmin and measure the masking fraction in these circles. We find that this fraction is ≈ 90 per cent across the considered scales. Multiplying the mixed terms in the covariance by that fraction and the cosmic variance terms by the square of that fraction (together with using equation 68), we indeed find significant improvement of the maximum posterior χ^2 obtained for the FLASK simulations – as is shown in the left-hand panel of Fig. 8 (as well as in Fig. 1).

In Fig. 9, we use our FLASK measurements together with the technique of *precision matrix expansion* (PME; from inverse covariance = precision matrix; Friedrich & Eifler 2018) and perform a consistency of the modelling ansatz described above by investigating the impact of masking on individual covariance terms. We find both with the PME methods and with our analytic ansatz that masking effects are most impactful in the covariance terms that depend on shape noise of the weak lensing source galaxies (i.e. in what we called mixed terms in Section 3). This also agrees with the findings of Joachimi et al. (2020) and it further motivates the modelling of masking effects that we have described here. Nevertheless, we do not elevate this modelling ansatz to our fiducial covariance model because its motivation remains rather heuristic. However, we consider it a realistic estimate for the error made by the f_{sky} approximation and can hence use it to estimate the impact of that approximation on parameter constraints. In Fig. 2, we have already shown that this impact is below the 1 per cent level; i.e. we underestimate the scatter of maximum posterior parameters by less than 1 per cent when making the f_{sky} approximation in our fiducial covariance model.

Note that Kilbinger & Schneider (2004), Sato et al. (2011), Shirasaki et al. (2019), and Philcox & Eisenstein (2019) have devised and promoted an alternative method to correct for masking, which amounts to direct Monte Carlo integration of expressions like equation (69). Given the large area of DES-Y3 and its numerous combinations of redshift bins, we did not find this to be feasible.

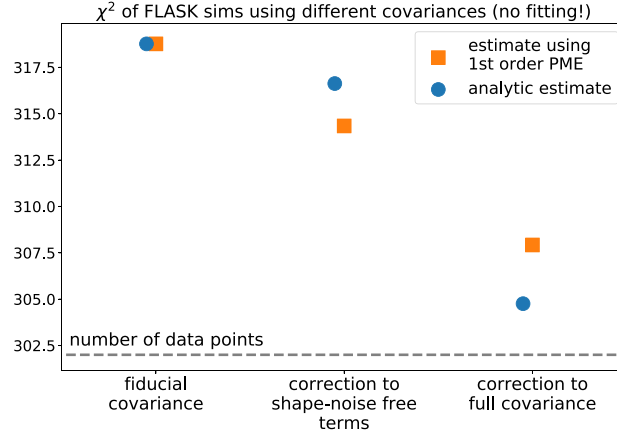


Figure 9. The method of PME (Friedrich & Eifler 2018) allows us to estimate the impact of individual covariance terms on χ^2 even when only few simulated measurements are available. The orange squares show the average χ^2 between our FLASK measurements and their mean [re-scaled by a factor of $N_{\text{FLASK}}/(N_{\text{FLASK}} - 1)$ to account for the correlation of individual measurements and mean] when using either no PME at all or when using PME estimates from shape-noise free sims or from the full sims. The blue dots show the corresponding χ^2 values when using the heuristically motivated analytical treatment of masking and survey geometry presented in the main text. The grey dashed line represents the number of data points and should be the average χ^2 if we had a perfect covariance model (note that for this comparison we have not performed any parameter fitting).

6.7 Non-Poissonian shot noise

In the Poissonian limit and in a complete region of the sky, the power spectrum of shot noise is scale independent and given by

$$N_\ell^{\text{complete}} = \frac{1}{\bar{n}}, \quad (72)$$

where \bar{n} is the galaxy density per steradian. As noted in the previous subsection, in galaxy surveys not every region of the sky is fully accessible; i.e. the presence of bright stars, satellite trails, etc. leads to artificial changes in the measured galaxy density. These density changes can potentially modify the observed galaxy power spectrum and bias any cosmological analyses derived from them, and thus, they are avoided by removing certain regions of the sky where artefacts may be found. These regions are usually smaller than the resolution of the (pixelated) survey mask used to determine whether a region of the sky is within the footprint or not, since it is computationally expensive to increase the resolution. This can be described through a fractional mask $W_i = 1/f_i$, where i is a given pixel of the mask and f_i is the fractional area of the pixel unaffected by the presence of artefacts. If we compute the galaxy overdensity as $\delta_{g,i} = N_i/(\bar{N}W_i) - 1$, with $\bar{N} = \sum_i N_i / \sum_i W_i$, the mean number of sources per pixel, we can estimate the Poissonian noise power spectrum as (Nicola et al. 2020)

$$N_\ell = \Omega_{\text{pix}} \frac{\bar{W}}{\bar{N}}, \quad (73)$$

where \bar{w} is the mean of the weights w_i across the footprint, and Ω_{pix} is the area of the pixels from the mask in steradians. In the case where all the pixels in the footprint are fully complete, we recover equation (72) since $\bar{n} = \bar{N}/\Omega_{\text{pix}}$, and $\bar{N} = \sum_i N_i / \sum_i 1$. However, in the case that any of the pixels of the mask are not fully complete we obtain an increased shot-noise contribution by a factor $\bar{W} \geq 1$ (since $0 \leq f_i \leq 1$).

In previous studies, the DES-Y1 lens galaxies were shown to prefer a super-Poissonian variance (Friedrich et al. 2018; Gruen et al. 2018) that might be a consequence of their complex selection criteria or due to the nature of their formation and evolution (see e.g. Baldauf et al. 2013; Dvornik et al. 2018). This super-Poissonian variance leads to an enhancement in shot noise. In order to test for this effect, we proceeded to estimate the angular power spectrum, $C_\ell \approx C_{\ell, \text{galaxies}} + N_\ell + \delta C_\ell$, of DES-Y1 REDMAGIC galaxies selected in Elvin-Poole et al. (2018) using NaMaster (Alonso et al. 2019), where N_ℓ is the shot-noise contribution from equation (73) and δC_ℓ is the excess power that can be due to a number of factors (variations in completeness not captured by the mask, super-Poissonian shot noise, observational systematics, etc.). We also compute the power spectrum, $C_{\ell, \text{rnd}}$ of a random field with the same number of objects as the galaxy sample considered, and the probability of populating a pixel i is proportional to its weight, W_i . We find that $C_{\ell, \text{rnd}}$ is statistically consistent with N_ℓ . We then compute the ratio as follows:

$$r_\ell = \frac{C_\ell - C_{\ell, \text{rnd}}}{N_\ell} \approx \frac{C_{\ell, \text{galaxies}}}{N_\ell} + \frac{\delta C_\ell}{N_\ell}. \quad (74)$$

In Fig. 10, we show r_ℓ compared to the theoretical expectation for $C_{\ell, \text{galaxies}}/N_\ell = C_{\ell, \text{th}}/N_\ell$, where $C_{\ell, \text{th}}$ is the theoretical power spectrum computed using the best-fitting parameters found in Elvin-Poole et al. (2018). We also allow for a 20 per cent variation in the linear galaxy bias, which is much larger than the uncertainty found in Elvin-Poole et al. (2018). We find that there is an excess power at $\ell \geq 3000$ that cannot be explained by an excess galaxy clustering (i.e. a larger than measured linear bias or a non-linear bias component). We identify this excess (between 2 per cent and 6 per cent) with $\frac{\delta C_\ell}{N_\ell}$ in equation (74).

This excess will translate into an extra shot-noise-like contribution to the covariance matrix (Philcox et al. 2020). The way we include this is by fitting a correction to the number density α_n such that the excess power is compatible with zero. In order to do so, we minimize the

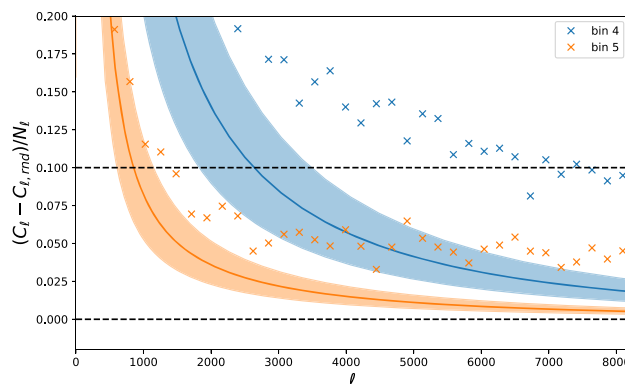


Figure 10. Measured ratio $r_\ell = \frac{C_\ell - C_{\ell,rand}}{N_\ell}$ (crosses) compared to predicted contribution of the galaxy power spectra over the shot noise (solid line) for the fiducial parameters at Elvin-Poole et al. (2018) allowing for a 20 per cent uncertainty in the galaxy bias (shaded regions) for two redshift bins (bin 4 in blue and bin 5 in orange). Horizontal dashed lines are just to guide the eye. If the shot noise were to be completely Poissonian, the measured and predicted ratios would agree; however, we find an excess between 2 per cent and 6 per cent.

Table 4. Best-fitting values of α_n to correct for the excess shot noise with the DES-Y1 REDMAGIC galaxies.

Bin number	α
1	1.042 ± 0.002
2	1.069 ± 0.003
3	1.072 ± 0.003
4	1.057 ± 0.003
5	1.021 ± 0.001

following χ^2 :

$$\chi^2(\alpha_n) = \sum_\ell \left(\frac{C_\ell - C_{\ell,th}}{\alpha_n N_\ell} - \frac{C_{\ell,rand}}{N_\ell} \right)^2 \left(\frac{\Delta C_{\ell,th}}{\alpha_n N_\ell} \right)^{-2}, \quad (75)$$

where $\Delta C_{\ell,th}$ is varied in the range of $1.2^2 C_{\ell,th} - 0.8^2 C_{\ell,th}$ (so we are allowing for 20 per cent uncertainty in the bias for the fit). The resulting values for α_n can be found in Table 4. In Figs 1 and 2 and Table 1, one can see that depleting the lens galaxy densities in our fiducial covariance model by these factors has a negligible effect on both maximum posterior χ^2 and parameters constraints.

6.8 Cosmology dependence of the covariance model

In order to evaluate our covariance model, we choose a particular set of cosmological parameters. We do not vary these parameters when sampling our parameter posterior and this may impact the width of our parameter constraints (Hamimeche & Lewis 2008; Eifler et al. 2009; White & Padmanabhan 2015; Kalus, Percival & Samushia 2016). Our main reason for not sampling the covariance model along with the data model is that computing a covariance matrix is computationally too costly for this to be feasible. Recently, Carron (2013) has also indicated that it may indeed be incorrect to vary the covariance cosmology when running MCMC chains.

It is only after running the MCMC chains that we can recompute the covariance at our best-fitting parameters and re-derive our parameter constraints – repeating this process until our constraints have converged (cf. Abbott et al. 2018, for the application of this procedure in the DES Y1 data). Therefore, the cosmology at which we compute our covariance is expected to be off from the best-fitting cosmology. In this subsection, we investigate how χ^2 , as well as cosmological parameter constraints, shifts when computing the covariance at cosmologies that are randomly drawn from the DESY3-like posterior.

We test the robustness of our constraints against the choice of cosmological parameters at which we evaluate the covariance model by taking a set of 100 different cosmologies drawn randomly from the simulated DES-Y3 3×2 pt posterior and generating 100 lognormal covariance matrices. Using each of these covariances, we estimate posteriors for a given realization of simulated DES-Y3 3×2 pt data with noise drawn from a fiducial lognormal covariance.

Since it is prohibitively expensive to perform simulated analyses running MCMC chains for each covariance matrix, we use the technique of importance sampling. That allows us to quickly evaluate how these different likelihood modelling choices impact the derived parameter constraints without repeatedly running expensive sampling algorithms. In our importance sampling pipeline, we take a fiducial analysis as a

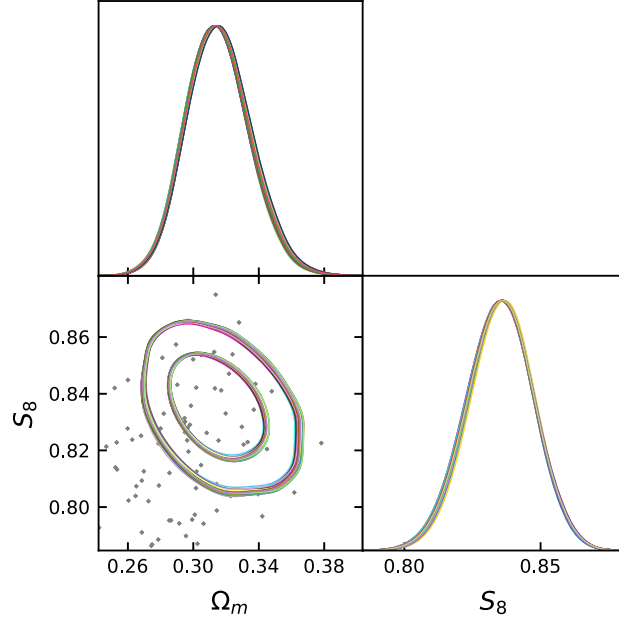


Figure 11. (S_8 , Ω_m) constraints for a given noisy realization of the DES-Y3 3×2 pt data vector analysed using 100 lognormal covariance matrices, each computed from a different cosmology drawn from a simulated DES-Y3 3×2 pt posterior. The 100 contours are superimposed in the plot, showing very small change in constraints. The points indicate the cosmologies at which the covariances were evaluated.

proposal distribution, re-evaluate the likelihoods using the alternative covariance matrix, and compute importance weights as:

$$w_i = \frac{\mathcal{L}(\pi_i | \hat{\xi}, \mathbf{C}_{\text{alt}})}{\mathcal{L}(\pi_i | \hat{\xi}, \mathbf{C}_{\text{fid}})}, \quad (76)$$

where \mathbf{C}_{fid} is the fiducial covariance in the analysis and \mathbf{C}_{alt} is the alternative one. If the changes induced by the new covariance matrix in the posterior are not too large, the re-weighted samples represent the target distribution (i.e. the posterior for the alternative covariance matrix). So we have

$$E_{\mathbf{p}}[f(X_i)] = \sum_i p_i f(X_i) = \sum_i q_i \frac{p_i}{q_i} f(X_i) = E_{\mathbf{q}}[w_i f(X_i)], \quad (77)$$

for a function $f(X_i)$ of the posterior samples. Here, p_i is the probability of X_i under the target distribution \mathbf{p} and q_i is the probability of X_i under the proposal distribution \mathbf{q} (see e.g. MacKay 2002; Owen 2013).

To diagnose the performance of our importance sampling estimates, we use the effective sample size (ESS):

$$\text{ESS} = \frac{\langle w_i \rangle^2}{\langle w_i^2 \rangle} N_{\text{samples}}, \quad (78)$$

where N_{samples} is the total number of posterior samples used in the estimation. The ESS as defined above quantifies the statistical power of the sample set after the re-weighting process (assuming uncorrelated samples). It is equal to the original sample size re-scaled by the ratio of the variances under each of the distributions (Martino, Elvira & Louzada 2017), such that the error of the mean of a quantity x with standard deviation σ_x under the target distribution can be estimated as $\sigma_x / \sqrt{\text{ESS}}$. Additionally, since our proposal distribution is itself a weighted sample set, we incorporate both the original and the importance weights in our ESS estimate.

Using the fiducial lognormal covariance matrix, we run the nested sampling algorithm MultiNest (Feroz & Hobson 2008; Feroz, Hobson & Bridges 2009; Feroz et al. 2019), and perform the importance sampling procedure to estimate parameters using each of the 100 covariance matrices randomly sampled in parameter space. The (S_8 , Ω_m) contours can be seen in Fig. 11. The ESSs for the importance sampled estimates range from 16 446 to 18 329 (implying a standard error of the mean within 0.78 per cent of the standard deviation for all cases), and the contours show good statistics. As the impact of covariance cosmology is barely noticeable for this range of tested parameters, we repeat the analysis for a few more extreme (and unlikely) cosmologies in Appendix F.

These results all confirm that we can safely neglect the impact of the choice of covariance cosmology in DES-Y3 3×2 pt analysis. One caveat of this conclusion is that we have indeed only varied cosmological parameters (including galaxy bias parameters) but not nuisance parameters (multiplicative shear bias, photometric redshift uncertainties) or parameters that describe intrinsic alignment. However, the DES-Y3 shear and photo- z calibration yield tight Gaussian priors on the corresponding nuisance parameters. Also, intrinsic alignment is relevant only on small angular scales where the covariance matrix is dominated by sampling noise contributions. Hence, we do not expect the results of this section to change significantly had all parameters been varied.

6.9 Random point shot noise

We also consider the effect of additional shot noise in the measurements of galaxy clustering resulting from the use of finite numbers of random points. The Landy–Szalay estimator (Landy & Szalay 1993) is estimating the galaxy clustering correlation function inside an angular bin $[\theta_1, \theta_2]$ as

$$\hat{w}[\theta_1, \theta_2] = \frac{DD[\theta_1, \theta_2] - 2DR[\theta_1, \theta_2] + RR[\theta_1, \theta_2]}{RR[\theta_1, \theta_2]}, \quad (79)$$

where $DD[\theta_1, \theta_2]$ is the number of galaxy pairs found within the angular bin, $RR[\theta_1, \theta_2]$ is the (normalized) number of pairs of random points that uniformly samples the survey footprint, and $DR[\theta_1, \theta_2]$ is the (normalized) number of galaxy-random-point pairs within the angular bin. If the number density of random points n_r is much larger than the number density of the galaxies n_g (as is recommended for reduce sampling noise), then both RR and DR must be re-scaled by factors of $(n_g/n_r)^2$ and (n_g/n_r) respectively.

We stress that the Landy–Szalay estimator was devised at a time of very limited computational resources, where it was prohibitively costly to measure galaxy pair in a large number of random points. Hence, it was vital to minimize random point shot noise. Nowadays, footprint geometries of photometric surveys are typically characterized by high-resolution healpix maps. The most straightforward way to calculate galaxy clustering correlation function is to simply assign a value of galaxy density contrast to each of these pixels and then measure the scalar autocorrelation function of the unmasked pixels. This way, one is avoiding random point shot noise completely.

Nevertheless, it is still very common to measure $w(\theta)$ by means of equation (79). So we also tested what impact a finite number of random points would have on our analysis. To do so, we extended expressions of Cabré & Gaztañaga (2009; see their appendix A) to the case where the same random points are used to estimate $w(\theta)$ in each of our redshift bins and also to subtract shear around random points from our galaxy–galaxy lensing correlation functions. Note that this causes a noise contribution to the two-point function measurements that is correlated among different redshift bins. We assumed a random point density of 1.36 arcmin^{-2} , which is more than 20 times larger than the density of our most dense lens galaxy sample. From Table 1, it can be seen that not accounting for the random point shot noise in the covariance leads to an increase in average χ^2 of $\lesssim 1$ per cent and to an underestimation of parameter uncertainties by ≈ 0.5 per cent. Hence, this effect can be ignored for our analysis.

6.10 Effective densities and effective shape noise

We are closing this section by spelling out an aspect of covariance modelling that may seem straightforward but which has repeatedly come up in covariance discussions.

If the tracer galaxies used to estimate two-point correlation functions are weighted according to some weighting scheme, then this may change the effective number densities and the effective shape noise that should be used when evaluating the covariance expressions in Section 4. In the following, we will derive how this can be done for each of the two-point functions in the DES Y3 $3 \times 2\text{pt}$ data vector.

6.10.1 Galaxy clustering

We start with the galaxy clustering correlation function $w(\theta)$. We assume a weighting scheme that is aimed at correcting for non-cosmological density fluctuations resulting from spatially varying observing conditions (as e.g. in Elvin-Poole et al. 2018). This means that the weights assigned to each galaxy in fact sample a weight map that spans the entire footprint.

Instead of measuring $w(\theta)$ from the weighted galaxies by means of, say, the Landy–Szalay estimator (Landy & Szalay 1993), it will be more convenient to think of the galaxy density contrast as a pixelized field on the sky. Furthermore, we will assume that the weight map has been normalized such that $\langle w \rangle = 1$ (which can always be done without changing the outcome of the weighted measurement). Consider pixel i with galaxy count $N_{g,i}$ and weight w_i . If n_g is the average galaxy density of the unweighted sample, then by taking expectation values with respect to many Poissonian shot-noise realizations (and hence ignoring fluctuations of the underlying matter density field) we get

$$\langle N_{g,i} \rangle = \frac{A_{\text{pix}} n_g}{w_i} \quad (80)$$

$$\text{Var}(N_{g,i}) = \frac{A_{\text{pix}} n_g}{w_i} \quad (81)$$

$$\text{Var}(w_i N_{g,i}) = w_i A_{\text{pix}} n_g \quad (82)$$

$$\begin{aligned} \text{Var}\left(\frac{w_i N_{g,i}}{A_{\text{pix}} n_g} - 1\right) &\equiv \text{Var}(\delta_{g,i}) \\ &= \frac{w_i}{A_{\text{pix}} n_g}, \end{aligned} \quad (83)$$

where A_{pix} is the area of each pixel and the second to last line serves as definition of $\delta_{g,i}$ and needs the fact that we demanded $\langle w \rangle = 1$. Note that these equations are only valid for an ensemble of observations that shares the same weight maps and differs only in their shot-noise realizations.

From the set of all pixels, we can now estimate $w(\theta)$ within a finite angular bin $[\theta_1, \theta_2]$ as

$$\hat{w}[\theta_1, \theta_2] = \frac{\sum_{\text{pxls } i>j} \Delta_{[\theta_1, \theta_2]}^{ij} \delta_{g,i} \delta_{g,j}}{\sum_{\text{pxls } i>j} \Delta_{[\theta_1, \theta_2]}^{ij}}, \quad (84)$$

where the symbol $\Delta_{[\theta_1, \theta_2]}^{ij}$ in the double sum over all pixels is 1 when the distance of the pixel pair i, j is within $[\theta_1, \theta_2]$ and 0 otherwise. Note that we assume an enumeration of the pixels and that we demand $i > j$ in the sum in order to not count any pair of pixels twice.

If shot noise is the only source of noise, then it is straightforward to calculate the variance of this measurement as

$$\begin{aligned} \text{Var}(\hat{w}[\theta_1, \theta_2]) &= \frac{\sum_{\text{pxls } i>j} \Delta_{[\theta_1, \theta_2]}^{ij} \langle \delta_{g,i}^2 \rangle \langle \delta_{g,j}^2 \rangle}{\left[\sum_{\text{pxls } i>j} \Delta_{[\theta_1, \theta_2]}^{ij} \right]^2} \\ &= \frac{1}{(A_{\text{pix}} n_g)^2} \frac{\sum_{\text{pxls } i>j} \Delta_{[\theta_1, \theta_2]}^{ij} w_i w_j}{\left[\sum_{\text{pxls } i>j} \Delta_{[\theta_1, \theta_2]}^{ij} \right]^2} \\ &= \frac{1}{N_{\text{pair},g}[\theta_1, \theta_2]} \frac{\sum_{\text{pxls } i>j} \Delta_{[\theta_1, \theta_2]}^{ij} w_i w_j}{\sum_{\text{pxls } i>j} \Delta_{[\theta_1, \theta_2]}^{ij}}. \end{aligned} \quad (85)$$

In the last line, $N_{\text{pair},g}[\theta_1, \theta_2]$ is the number of unweighted galaxy pairs within the angular bin $[\theta_1, \theta_2]$. Note that in the presence of clustering, this should be calculated from a set of random points instead of from the actual galaxy catalogue.

The first factor on the right-hand side of equation (85) is what the shot-noise variance of \hat{w} should be in the absence of a weighting scheme. The second term is a two-point function of the weight map itself. If the weight map has a white-noise power spectrum, then this factor will be close to 1 in any angular bin that does not include angular distances of 0. This means that at large enough scales the last line of equation (85) looks like the covariance for plain Poissonian shot noise without any notion of an effective number density. This may be surprising, but it stems from the fact that the weighting scheme we assumed does not simply multiply the galaxy density contrast field. Instead, it reverses an already existing depletion of galaxy density from non-cosmological density fluctuations.

In conclusion, the effective number density that should be used to compute the covariance of $\hat{w}[\theta_1, \theta_2]$ is

$$n_{g,\text{eff}}[\theta_1, \theta_2] = n_g \sqrt{\frac{\sum_{\text{pxls } i>j} \Delta_{[\theta_1, \theta_2]}^{ij}}{\sum_{\text{pxls } i>j} \Delta_{[\theta_1, \theta_2]}^{ij} w_i w_j}}. \quad (86)$$

6.10.2 Galaxy–galaxy lensing

We move on to consider the galaxy–galaxy lensing correlation function $\gamma_t[\theta_1, \theta_2]$. We assume that the lens galaxy sample comes with weights derived from a weight map w^l as in the previous subsection while each source galaxy j has a weight w_j^s that does not come from an entire weight map but is instead the result of the individual quality of shape measurement for this galaxy. A measurement of γ_t can be constructed as

$$\hat{\gamma}_t[\theta_1, \theta_2] = \frac{\sum_{\text{pxl } i, \text{ source } j} \Delta_{[\theta_1, \theta_2]}^{ij} \delta_{l,i} \epsilon_{t,j \rightarrow i} w_j^s}{\sum_{\text{pxl } i, \text{ source } j} \Delta_{[\theta_1, \theta_2]}^{ij} w_j^s}. \quad (87)$$

Here, $\delta_{l,i}$ is the galaxy density contrast of the lenses defined in analogy to the previous subsection, $\epsilon_{t,j \rightarrow i}$ is the tangential component of the shear of source j with respect to lens galaxy i , and w_j is the weight of source galaxy j . Note that due to our definition of the lens galaxy density contrast this estimator already includes subtraction of shear around random points.

If shot noise and shape noise are the only sources of noise, then it can be readily shown that

$$\begin{aligned} \text{Var}(\hat{\gamma}_t[\theta_1, \theta_2]) &= \frac{\sum_{\text{pxl } i, \text{ source } j} \Delta_{[\theta_1, \theta_2]}^{ij} \langle \delta_{l,i}^2 \rangle \langle (\epsilon_{t,j \rightarrow i} w_j^s)^2 \rangle}{\left[\sum_{\text{pxl } i, \text{ source } j} \Delta_{[\theta_1, \theta_2]}^{ij} w_j^s \right]^2} \\ &\approx \frac{1}{N_{\text{pair},ls}[\theta_1, \theta_2]} \frac{\sum_{\text{pxl } i, \text{ source } j} \Delta_{[\theta_1, \theta_2]}^{ij} \langle (\epsilon_{t,j \rightarrow i} w_j^s)^2 \rangle}{\sum_{\text{pxl } i, \text{ sources } j} \Delta_{[\theta_1, \theta_2]}^{ij} w_j^s} \\ &\approx \frac{1}{2N_{\text{pair},ls}[\theta_1, \theta_2]} \frac{\sum_{\text{source } j} \langle |\epsilon_j|^2 (w_j^s)^2 \rangle}{N_s} \quad (\text{only with } \langle w_j^s \rangle = 1 = \langle w_j^l \rangle!). \end{aligned} \quad (88)$$

Here, $N_{\text{pair},ls}[\theta_1, \theta_2]$ is the number of unweighted lens–source pairs in the angular bin $[\theta_1, \theta_2]$, N_s is the total number of source galaxies and $\epsilon_j = \epsilon_{1,j} + i\epsilon_{2,j}$ is the complex intrinsic ellipticity of source galaxy j .

Note that the final expression in equation (88) explicitly allows for the possibility that the source weights w_j^s are correlated with the intrinsic ellipticities ϵ_j of the source galaxies. One can interpret equation (88) as

$$\text{Var}(\hat{\gamma}_t[\theta_1, \theta_2]) = \frac{\sigma_{\epsilon, \text{eff}}^2}{N_{\text{pair}, ts}[\theta_1, \theta_2]} \quad (89)$$

with the effective dispersion of intrinsic ellipticity per shear component given by

$$\sigma_{\epsilon, \text{eff}}^2 = \frac{1}{2} \frac{\sum_{\text{source } j} |\epsilon_j|^2 (w_j^s)^2}{N_s}. \quad (90)$$

One subtlety here is that the above derivation requires $\langle w_j^s \rangle = 1$. The above expressions must be modified as this is not the case or when taking into account responses R_j of a shape catalogue generated with metacalibration (Sheldon & Huff 2017). We detail what to do in the latter case in Appendix G.

6.10.3 cosmic shear

For cosmic shear, we follow Schneider et al. (2002) and construct a measurement of ξ_+ from a set of sources as

$$\hat{\xi}_+[\theta_1, \theta_2] = \frac{\sum_{i>j} \Delta_{[\theta_1, \theta_2]}^{ij} w_i w_j (\epsilon_{1,i} \epsilon_{1,j} + \epsilon_{2,i} \epsilon_{2,j})}{\sum_{i>j} \Delta_{[\theta_1, \theta_2]}^{ij} w_i w_j}. \quad (91)$$

If shape noise is the only source of noise and if the intrinsic ellipticities of galaxies are not correlated with their weights, then the variance of $\hat{\xi}_+$ is given by

$$\begin{aligned} \text{Var}(\hat{\xi}_+[\theta_1, \theta_2]) &= \frac{\sum_{i>j} \Delta_{[\theta_1, \theta_2]}^{ij} \langle \epsilon_{1,i}^2 w_i^2 \rangle \langle \epsilon_{1,j}^2 w_j^2 \rangle + \langle \epsilon_{2,i}^2 w_i^2 \rangle \langle \epsilon_{2,j}^2 w_j^2 \rangle}{\left[\sum_{i>j} \Delta_{[\theta_1, \theta_2]}^{ij} w_i w_j \right]^2} \\ &\approx \frac{2\sigma_{\epsilon, \text{eff}}^4}{N_{\text{pair}}[\theta_1, \theta_2]}. \end{aligned} \quad (92)$$

Here, $N_{\text{pair}}[\theta_1, \theta_2]$ is the number of source galaxy pairs in the bin $[\theta_1, \theta_2]$ and we have replaced each expectation value $\langle \epsilon_{1/2,i}^2 w_i^2 \rangle$ by $\sigma_{\epsilon, \text{eff}}^2$ from equation (90). Note that we again assumed $\langle w_j \rangle = 1$ and that this may require re-scaling of both weights and σ_{ϵ} when using shape measurements from metacalibration.

6.10.4 Testing validity of effective shape noise

To test the validity of our expression for effective shape noise in equation (90), we run a sub-sample covariance estimator on our data (see e.g. Friedrich et al. 2016). In particular, we divide all of our source and lens galaxy samples into 200 randomly chosen sub-samples and measure the galaxy–galaxy lensing correlation function of each source–lens bin combination. As a result, we obtain 200 measurements of $\hat{\gamma}_t$ in each source–lens bin combination. Since we employ completely random sub-sampling, i.e. without any regard for e.g. a division of our footprint into sub-regions, the sample covariance of these 200 measurements will almost exclusively be dominated by shape noise and shot noise. This is even more so, because the lens and source densities of the sub-samples are very low.

In Fig. 12, we show the ratio of the variances of the 200 galaxy–galaxy lensing measurements $\hat{\gamma}_t$ in the different lens–source bin combinations to equation (89). Assuming that the sub-sample covariances follow a Wishart distribution, we find that these ratios are perfectly consistent with 1. This indicates that equation (90) indeed yields an accurate effective shape-noise dispersion, and that one should indeed use the plain density of lens galaxies (as opposed to any notion of effective density) when evaluating covariance expressions.

7 A SIMPLE χ^2 TEST

In this short section, we present a simple χ^2 test that does not rely on the linearized framework. However, it has the disadvantage of not addressing the impact on the estimation of parameters. Here, we generate a large number of ‘contaminated’ data vectors (we use 1000) by a Gaussian sampling of a given covariance matrix that includes different effects and to compute a χ^2 distribution from these data vectors using a fiducial covariance matrix. The resulting shifts in the mean value of χ^2 and their standard deviations give another benchmark for the importance of the different effects considered here. We show the results of this test in Fig. 13. Note that the relative increases in χ^2 follow closely what we obtained within the linearized likelihood framework in Fig. 1. This indicates that the dominant way in which covariance errors cause χ^2 offsets is not through the altered scatter of maximum posterior parameter locations but simply through using an erroneous inverse covariance when computing χ^2 . That also justifies our usage of the linearized likelihood framework since any impact of non-linear parameter dependences on parameter fitting can be expected to be even less relevant than linear fitting in the first place.

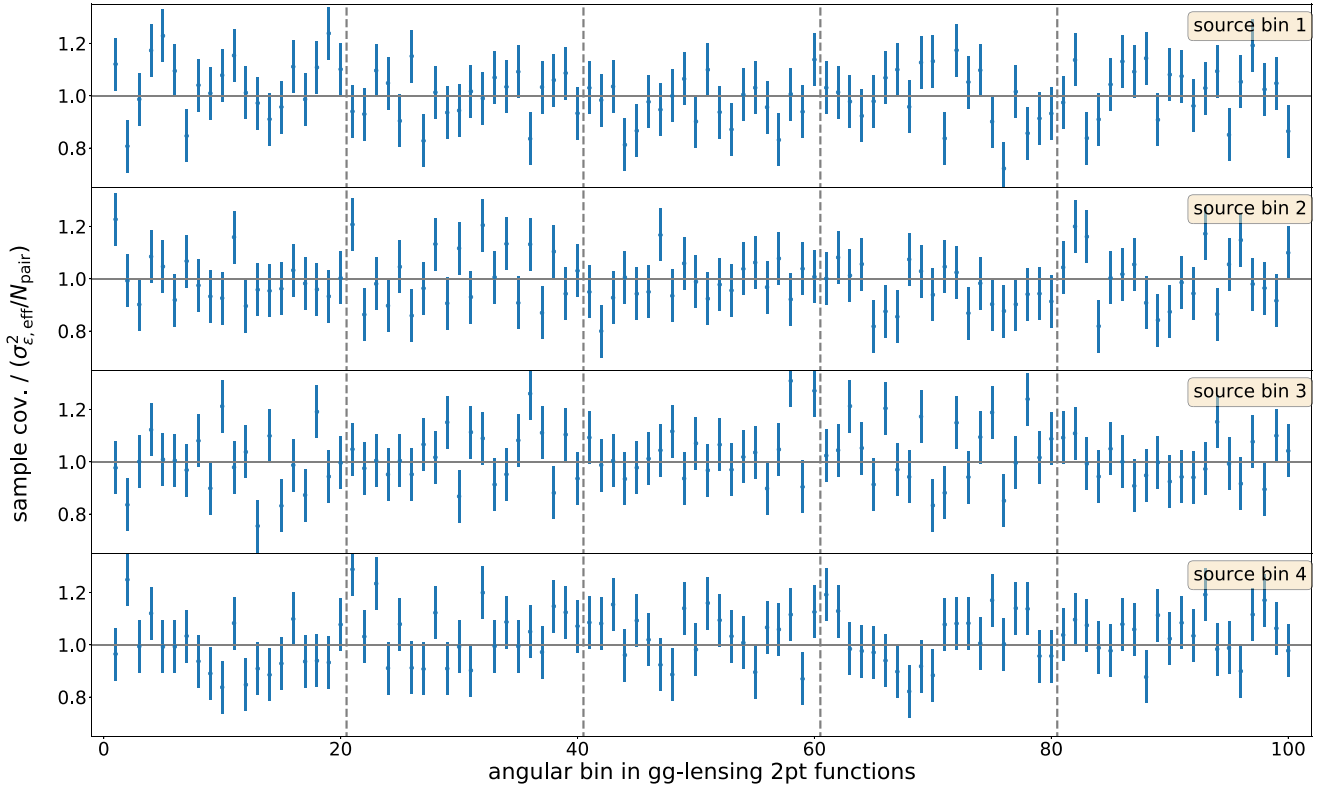


Figure 12. Ratio between the sample variance of $\hat{\gamma}_l$ measured in 200 randomly selected sub-samples of the DESY3 lens and sources catalogues and equation (89) for the shape-noise contribution to the covariance (again using equation 90 to calculate the effective shape-noise dispersion $\sigma_{\epsilon,\text{eff}}$). Each row displays the variances measured for a different source redshift bin and vertical dashed lines separate points belonging to different lens redshift bins (1–5 from left to right). Assuming that the covariance estimates have a Wishart distribution, we calculate the covariance matrix of these ratios (cf. Taylor et al. 2013) and find that they are consistent with 1 (both for the cosmic shear and galaxy–galaxy lensing variances).

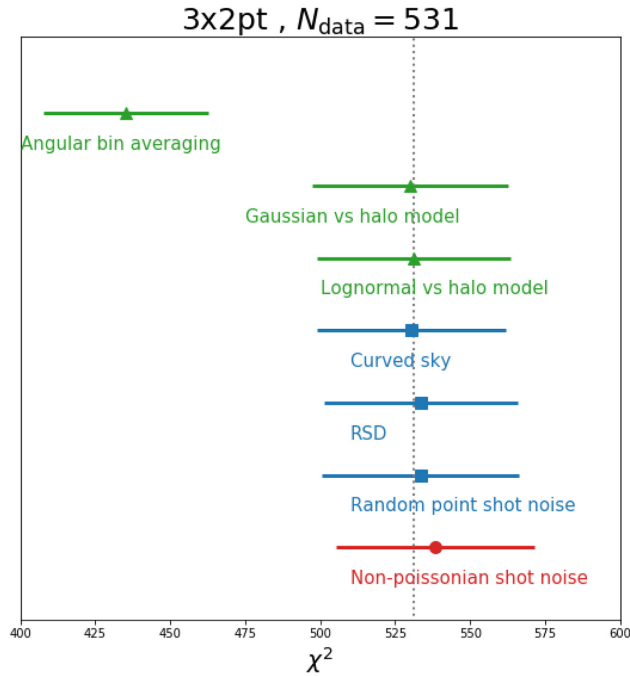


Figure 13. χ^2 tests taking into account different effects. Colours follow the scheme of Fig. 1.

8 DISCUSSIONS AND CONCLUSIONS

In this paper, we have presented the fiducial covariance model of the DES-Y3 joint analysis of cosmic shear, galaxy–galaxy lensing, and galaxy clustering correlation functions (the 3×2 pt analysis). We then investigated how the assumptions and approximations of that model (including the assumption of Gaussian statistical uncertainties) impact the distribution of maximum posterior χ^2 and maximum posterior estimates of cosmological parameters.

The fiducial covariance matrix of the DES-Y3 3×2 pt analysis uses the formalism of Krause & Eifler (2017) to model supersample covariance as well as the trispectrum contribution to the covariance. The model for the Gaussian covariance part (i.e. the contributions from the disconnected four-point function) correctly takes into account sky curvature and includes analytical averaging over the finite angular bins in which the two-point functions are measured. Furthermore, the galaxy clustering power spectra that enter our calculation of the Gaussian covariance part are computed using the non-Limber formalism of Fang et al. (2020b) and also include modelling of RSDs. The finite survey area of DES-Y3 is incorporated in the covariance model via the f_{sky} approximation (except in the pure shape-noise and shot-noise terms where we follow Troxel et al. 2018b).

In order to perform our validation tests for the DES-Y3 covariance matrix, we developed a plethora of new modelling ansatzes and testing strategies that are applicable in general. These new techniques are as follows:

(i) We have motivated and devised a way of drawing realizations of the 3×2 pt data vector from a non-Gaussian distribution in order to test the accuracy of our Gaussian likelihood assumption.

(ii) We have derived analytical expressions for angular bin averaging of all four types of two-point correlation functions included in the 3×2 pt vector (ξ_+ , ξ_- , γ_t , and w). These expressions correctly account for sky curvature. To the best of our knowledge, an analytical treatment of bin averaging for cosmic shear two-point function has not been presented before (though we have shared our results with Fang et al. 2020a, who have used them for the fiducial covariance computations).

(iii) We have extended the lognormal analytical model for the covariance of cosmic shear two-point function of Hilbert et al. (2011) to the other two-point functions present in the 3×2 pt data vector.

(iv) Within a linearized likelihood formalism, we have analytically derived how covariance model inaccuracies influence the distribution of maximum-posterior χ^2 and of maximum-posterior parameter estimates. The results we presented also allow for the possibility of including the Gaussian priors on certain model parameters and can be used to analytically estimate the impact of covariance errors on cosmological likelihood analyses.

(v) By fitting an effective number density to the high- ℓ plateau of galaxy clustering C_ℓ measurements, we have estimated how much the assumption of Poissonian shot noise influences our likelihood analysis. This is similar in spirit to the RASCALC technique presented by Philcox et al. (2020), and we agree with those authors that non-Poissonian shot noise can be viewed as an effective description of how short-scale non-linearities in galaxy clustering influence the covariance.

(vi) We calculated covariance matrices for 100 different sets of cosmological and nuisance parameters randomly drawn from a simulated likelihood chain. This allowed us to investigate whether calculating our covariance model at a reasonable, but wrong point in parameter space significantly impacts our analysis. This was done both within our linearized likelihood framework and by using importance sampling to quickly evaluate the 100 non-linear likelihoods.

(vii) We have derived how the two-point correlation function of weight maps influences the covariance of galaxy clustering two-point function measurements. In that context, we have also found that traditional ways of deriving an effective number density for a given set of galaxy weights are erroneous when those weights are aimed at undoing a suspected depletion of galaxy density (e.g. due to variations in observing conditions).

(viii) We have derived an expression for the effective dispersion of intrinsic source shapes for the situation when source galaxy weights are correlated with galaxy ellipticity. We have also shown how metacalibration responses (Sheldon & Huff 2017) enter that expression for effective shape noise.

(ix) We have described a clean sub-sample covariance estimation scheme that directly measures the sampling noise contributions to the covariance from a given data set. We then used the resulting covariance estimates to test the validity of our assumed effective shape noise values.

(x) We have employed the hybrid covariance estimation technique PME (Friedrich & Eifler 2018) to efficiently evaluate the importance of individual contributions to the covariance from only a limited set of simulated data (in our case: 200 realizations of the 3×2 pt data vector including shape noise and 100 realizations without shape noise).

(xi) We have devised a treatment of survey geometry in covariance modelling that improves upon existing approximations (of e.g. Efstathiou 2004) and we have demonstrated how to carry those approximations from harmonic space to real space.

Using these results, we perform several tests for the fiducial DES-Y3 3×2 pt covariance matrix and likelihood model, with the following conclusions:

(i) The assumption of Gaussian statistical uncertainties is sufficiently accurate (cf. Section 6.1). Hence, knowledge of the covariance of the 3×2 pt data vector is sufficient to model our statistical uncertainties. The main assumption made to arrive at this conclusion is that non-Gaussian error bars are primarily a large-scale problem and that at small scales the number of modes present within the DES-Y3 survey volume converges to a Gaussian distribution via the central limit theorem.

(ii) The non-Gaussian part of the covariance has a negligible impact on both maximum posterior χ^2 and parameter constraints (cf. Section 6.2). This statement is not a general one but only holds for the specific DES-Y3 3×2 pt analysis set-up. The main assumption made to arrive at that conclusion is that the CosmoLike model (Krause, Eifler & Blazek 2016) or the lognormal model (Hilbert et al. 2011) for the non-Gaussian covariance does not vastly underestimate the true covariance. Given the results of Sato et al. (2009) and Hilbert et al. (2011), we find that this a safe assumption.

(iii) Of all covariance modelling assumptions investigated in this paper, the f_{sky} approximation (made in the mixed term and cosmic variance term of our covariance model) has the largest effect on maximum posterior χ^2 . On average, it increases χ^2 between measurement and maximum posterior model by about 3.7 per cent ($\Delta\chi^2 \approx 18.9$) for the 3×2 pt data vectors and by about 5.7 per cent ($\Delta\chi^2 \approx 16.0$) for the 2×2 pt data vector (cf. Table 1).

(iv) However, neither f_{sky} approximation nor any other covariance modelling detail tested in this paper (cf. Table 1; with the exception of finite bin width; see the next point) has any significant impact on the location and width of constraints on the parameters Ω_m , σ_8 , and w .

(v) The only exception to this statement is finite angular bin width that – if not taken into account in the mixed term and cosmic variance term of the covariance model – significantly increases the scatter of maximum posterior parameters (without increasing the inferred constraints accordingly). However, finite bin width has been taken into account in the past in an approximate manner – see e.g. Krause et al. (2017).

(vi) The fact that we do not know the true cosmological parameters of the Universe forces us to evaluate our covariance model at a wrong set of parameters. Even when iteratively adjusting those parameters to the maximum posterior parameters of the analysis, the parameters of the covariance model will scatter around the ‘true’ cosmological/nuisance parameters. We consider this an irreducible covariance error and find that it increases the maximum posterior scatter of Ω_m and σ_8 by about 3 per cent and that of the dark energy equation-of-state parameter w by about 5 per cent (cf. Section 6.8 and Table 1). At the same time, we find this effect to have a negligible impact on maximum posterior χ^2 .

In summary, we have shown that our fiducial covariance and likelihood model underestimates the scatter of maximum posterior parameters by about 3–5 per cent, which is mostly caused by uncertainty in the set of cosmological and nuisance parameters at which we evaluate that model. On average, the χ^2 between maximum posterior model and measurement of the 3×2 pt data vector will be ~ 4 per cent higher than expected with perfect knowledge of the covariance matrix. This is mainly caused by our use of the f_{sky} approximation. We have devised an improved treatment of the full survey geometry (cf. Section 6.6) but for the reason mentioned above this was only used to test the impact of the f_{sky} approximation on parameter constraints.

Given the small impact that we estimated from the unaccounted effects in the covariance modelling, we conclude that the fiducial covariance model is adequate to be used in the 3×2 pt DES-Y3 analysis. While our validation of this covariance model has been carried out with a preliminary set of scale cuts and redshift distributions, we do not expect qualitative changes for the final DES-Y3 analysis set-up.

While the DESY3 specific outcomes of our study cannot straightforwardly be transferred to other surveys and analyses, our methodological innovations will be useful tools in the covariance and likelihood validation of future experiments.

ACKNOWLEDGEMENTS

OF gratefully acknowledges support by the Kavli Foundation and the International Newton Trust through a Newton-Kavli-Junior Fellowship and by Churchill College Cambridge through a postdoctoral By-Fellowship. The authors thank Henrique Xavier for very helpful discussions about mock simulations and the FLASK code. This research was partially supported by the Laboratório Interinstitucional de e-Astronomia (LIneA), the Brazilian Funding agency CNPq, the INCT of the e-Universe, and the Sao Paulo State Research Agency (FAPESP). The authors acknowledge the use of computational resources from LIneA, the Center for Scientific Computing (NCC/GridUNESP) of the Sao Paulo State University (UNESP), and from the National Laboratory for Scientific Computing (LNCC/MCTI, Brazil), where the SDumont supercomputer (`sdumont.lncc.br`) was used.

This paper has gone through internal review by the DES collaboration. 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 DES.

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, the National Optical Astronomy Observatory, 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 MINECO under grants AYA2015-71825, ESP2015-66861, FPA2015-68048, 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 Ciéncia e Tecnologia (INCT) e-Universe (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.

This work made use of the software packages GETDIST (Lewis 2019), CHAINCONSUMER (Hinton 2016), MATPLOTLIB (Hunter 2007), and NUMPY (Harris et al. 2020).

We would like to thank the anonymous journal referee for their helpful comments.

DATA AVAILABILITY

The DES-Y3 3×2 pt covariance matrix and likelihoods will be made public upon publication of our final data analysis. C++ and PYTHON tools to configure FLASK as described in Section 4.3 are available at <https://github.com/OliverFHD/CosMomentum>. Tools to compute Gaussian and halomodel covariance are available at <https://github.com/CosmoLike/CosmoCov>.

REFERENCES

- Abbott T. M. C. et al., 2018, *Phys. Rev. D*, 98, 043526
 Abbott T. M. C. et al., 2019a, *Phys. Rev. D*, 99, 123505
 Abbott T. M. C. et al., 2019b, *Phys. Rev. Lett.*, 122, 171301
 Abbott T. M. C. et al., 2020, *Phys. Rev. D*, 102, 023509
 Alonso D., Sanchez J., Slosar A., LSST Dark Energy Science Collaboration, 2019, *MNRAS*, 484, 4127
 Amon A. et al., 2021, ([arXiv:2105.13543](https://arxiv.org/abs/2105.13543))
 Anderson T., 2003, *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Statistics. Wiley, Hoboken, NJ, available at <https://books.google.de/books?id=Cmm9QgAACAAJ>
 Avila S. et al., 2018, *MNRAS*, 479, 94
 Baldauf T., Seljak U., Smith R. E., Hamaus N., Desjacques V., 2013, *Phys. Rev. D*, 88, 083507
 Barreira A., Krause E., Schmidt F., 2018, *J. Cosmol. Astropart. Phys.*, 10, 053
 Blandford R., Dunkley J., Frenk C., Lahav O., Shapley A., 2020, *Nat. Astron.*, 4, 122
 Buchs R. et al., 2019, *MNRAS*, 489, 820
 Cabré A., Gaztañaga E., 2009, *MNRAS*, 393, 1183
 Carron J., 2013, *A&A*, 551, A88
 Cawthon R. et al., 2020, *MNRAS*, ([arXiv:2012.12826](https://arxiv.org/abs/2012.12826))
 Challinor A., Chon G., 2005, *MNRAS*, 360, 509
 Clerkin L. et al., 2017, *MNRAS*, 466, 1444
 Coles P., Jones B., 1991, *MNRAS*, 248, 1
 Cordero J. P. et al., 2020, *MNRAS*, ([arXiv:2109.09636](https://arxiv.org/abs/2109.09636))
 Costanzi M. et al., 2019, *MNRAS*, 488, 4779
 Crocce M., Cabré A., Gaztañaga E., 2011, *MNRAS*, 414, 329
 de Putter R., Takada M., 2010, *Phys. Rev. D*, 82, 103522
 DeRose J. et al., 2020, ([arXiv:2105.13547](https://arxiv.org/abs/2105.13547))
 DES Collaboration, 2020, ([arXiv:2105.13549](https://arxiv.org/abs/2105.13549))
 Diehl H. T. et al., 2019, FERMILAB-TM-2720-AE
 Dodelson S., Schneider M. D., 2013, *Phys. Rev. D*, 88, 063537
 Doux C. et al., 2020, *MNRAS*, 503, 2688
 Dvornik A. et al., 2018, *MNRAS*, 479, 1240
 Efstathiou G., 2004, *MNRAS*, 349, 603
 Eifler T., Schneider P., Hartlap J., 2009, *A&A*, 502, 721
 Elvin-Poole J. et al., 2018, *Phys. Rev. D*, 98, 042006
 Everett S. et al., 2020, ([arXiv:2012.12825](https://arxiv.org/abs/2012.12825))
 Fang X., Eifler T., Krause E., 2020a, *MNRAS*, 497, 2699
 Fang X., Krause E., Eifler T., MacCrann N., 2020b, *J. Cosmol. Astropart. Phys.*, 2020, 010
 Feroz F., Hobson M. P., 2008, *MNRAS*, 384, 449
 Feroz F., Hobson M. P., Bridges M., 2009, *MNRAS*, 398, 1601
 Feroz F., Hobson M. P., Cameron E., Pettitt A. N., 2019, *Open J. Astrophys.*, 2
 Fields B. D., Olive K. A., Yeh T.-H., Young C., 2020, *J. Cosmol. Astropart. Phys.*, 2020, 010
 Friedrich O., Eifler T., 2018, *MNRAS*, 473, 4150

- Friedrich O., Seitz S., Eifler T. F., Gruen D., 2016, *MNRAS*, 456, 2662
- Friedrich O. et al., 2018, *Phys. Rev. D*, 98, 023508
- Frieman J., Turner M., Huterer D., 2008, *Ann. Rev. Astron. Astrophys.*, 46, 385
- Galassi M., Davies J., Theiler J., Gough B., Jungman G., Alken P., Booth M., Rossi F., 2009, GNU Scientific Library: Reference Manual for GSL Version 1.12. Network Theory, available at <http://www.gnu.org/software/gsl/>
- Gatti M. et al., 2020a, *MNRAS*, [arXiv:2012.08569](https://arxiv.org/abs/2012.08569)
- Gatti M. et al., 2020b, *MNRAS*, 504, 4312
- Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, *ApJ*, 622, 759
- Gruen D. et al., 2018, *Phys. Rev. D*, 98, 023507
- Hall A., Taylor A., 2019, *MNRAS*, 483, 189
- Hamimeche S., Lewis A., 2008, *Phys. Rev. D*, 77, 103013
- Harris C. R. et al., 2020, *Nature*, 585, 357
- Hartlap J., Simon P., Schneider P., 2007, *A&A*, 464, 399
- Hartley W. G. et al., 2020, *MNRAS*, [arXiv:2012.12824](https://arxiv.org/abs/2012.12824)
- Heymans C. et al., 2020, *A&A*, 646, A140
- Hilbert S., Hartlap J., Schneider P., 2011, *A&A*, 536, A85
- Hinton S. R., 2016, *J. Open Source Softw.*, 1, 00045
- Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
- Jarvis M., Bernstein G., Jain B., 2004, *MNRAS*, 352, 338
- Jarvis M. et al., 2020, preprint([arXiv:2011.03409](https://arxiv.org/abs/2011.03409))
- Joachimi B., 2017, *MNRAS*, 466, L83
- Joachimi B. et al., 2020, *Astronomy & Astrophysics* 646 preprint([arXiv:2007.01844](https://arxiv.org/abs/2007.01844))
- Kaiser N., 1987, *MNRAS*, 227, 1
- Kalus B., Percival W. J., Samushia L., 2016, *MNRAS*, 455, 2573
- Kilbinger M., Schneider P., 2004, *A&A*, 413, 465
- Krause E., Eifler T., 2017, *MNRAS*, 470, 2100
- Krause E., Eifler T., Blazek J., 2016, *MNRAS*, 456, 207
- Krause E. et al., 2017, preprint([arXiv:1706.09359](https://arxiv.org/abs/1706.09359))
- Krause E. et al., 2021, [arXiv:2105.13548](https://arxiv.org/abs/2105.13548)
- Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64
- Lemos P. et al., 2021, *MNRAS*, 505, 4
- Lewis A., 2019, preprint([arXiv:1910.13970](https://arxiv.org/abs/1910.13970))
- Limber D. N., 1953, *ApJ*, 117, 134
- Lin C.-H., Harnois-Déraps J., Eifler T., Pospisil T., Mandelbaum R., Lee A. B., Singh S., LSST Dark Energy Science Collaboration, 2020, *MNRAS*, 499, 2977
- LoVerde M., Afshordi N., 2008, *Phys. Rev. D*, 78, 123506
- MacCrann N. et al., 2018, *MNRAS*, 480, 4614
- MacCrann N. et al., 2020, [arXiv:2012.08567](https://arxiv.org/abs/2012.08567)
- MacKay D. J. C., 2002, *Information Theory, Inference & Learning Algorithms*. Cambridge Univ. Press, USA
- Mantz A. B., Allen S. W., Morris R. G., Rapetti D. A., Applegate D. E., Kelly P. L., von der Linden A., Schmidt R. W., 2014, *MNRAS*, 440, 2077
- Martino L., Elvira V., Louzada F., 2017, *Signal Process.*, 131, 386
- Muir J. et al., 2020, *MNRAS*, 494, 4454
- Myles J. T. et al., 2021, *MNRAS* 505 3
- N. Bronstein I., A. Semendjajew K., 1979, in Teubner B. G., ed., *Taschenbuch der Mathematik*, 19th edn. BSB. 109, 116, 526, 911, Verlagsgesellschaft, Nauka-Verlag, Leipzig, Moskau
- Nicola A. et al., 2020, *J. Cosmol. Astropart. Phys.*, 03, 044
- Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, *MNRAS*, 396, 19
- Owen A. B., 2013, Monte Carlo Theory, Methods and Examples, available at <https://statweb.stanford.edu/~owen/mc/>
- Pandey S. et al., 2021([arXiv:2105.13545](https://arxiv.org/abs/2105.13545))
- Peebles P. J. E., 2012, *ARA&A*, 50, 1
- Percival W. J. et al., 2014, *MNRAS*, 439, 2531
- Philcox O. H. E., Eisenstein D. J., 2019, *MNRAS*, 490, 5931
- Philcox O. H. E., Eisenstein D. J., O'Connell R., Wiegand A., 2020, *MNRAS*, 491, 3290
- Planck Collaboration VI, 2020, *A&A*, 641, A6
- Pope A. C., Szapudi I., 2008, *MNRAS*, 389, 766
- Porredon A. et al., 2020, *Phys. Rev. D*, 103, 043503
- Porredon A. et al., 2021, ([arXiv:2105.13546](https://arxiv.org/abs/2105.13546))
- Prat J. et al., 2021, ([arXiv:2105.13541](https://arxiv.org/abs/2105.13541))
- Raveri M., Hu W., 2019, *Phys. Rev. D*, 99, 043506
- Riess A. G., 2017, *Handbook of Supernovae Confirming Cosmic Acceleration in the Decade that Followed from SNe Ia at $z > 1$* . p. 2615 Springer International Publishing AG, Basel.
- Rodríguez-Monroy M. et al., 2021, ([arXiv:2105.13540](https://arxiv.org/abs/2105.13540))
- Ross A. J., Percival W. J., Crocce M., Cabré A., Gaztañaga E., 2011, *MNRAS*, 415, 2193
- Sánchez C. et al., 2021, ([arXiv:2105.13542](https://arxiv.org/abs/2105.13542))
- Sato M., Hamana T., Takahashi R., Takada M., Yoshida N., Matsubara T., Sugiyama N., 2009, *ApJ*, 701, 945
- Sato M., Takada M., Hamana T., Matsubara T., 2011, *ApJ*, 734, 76
- Schaan E., Takada M., Spergel D. N., 2014, *Phys. Rev. D*, 90, 123523
- Schneider P., Hartlap J., 2009, *A&A*, 504, 705
- Schneider P., van Waerbeke L., Kilbinger M., Mellier Y., 2002, *A&A*, 396, 1

- Secco L. F. et al., 2021, (arXiv:2105.13544)
- Seehars S., Amara A., Refregier A., Paranjape A., Akeret J., 2014, *Phys. Rev. D*, 90, 023533
- Seehars S., Grandis S., Amara A., Refregier A., 2016, *Phys. Rev. D*, 93, 103507
- Sellentin E., Heavens A. F., 2017, *MNRAS*, 464, 4658
- Sellentin E., Heymans C., Harnois-Déraps J., 2018, *MNRAS*, 477, 4879
- Sevilla-Noarbe I. et al., 2020, *ApJS*, 254, 24
- Sheldon E. S., Huff E. M., 2017, *ApJ*, 841, 24
- Shirasaki M., Hamana T., Takada M., Takahashi R., Miyatake H., 2019, *MNRAS*, 486, 52
- Smith R. E. et al., 2003, *MNRAS*, 341, 1311
- Smith M. et al., 2020, *MNRAS*, 494, 4426
- Stebbins A., 1996, preprint (astro-ph/9609149)
- Takada M., Hu W., 2013, *Phys. Rev. D*, 87, 123504
- Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, *ApJ*, 761, 152
- Taylor A., Joachimi B., 2014, *MNRAS*, 442, 2728
- Taylor A., Joachimi B., Kitching T., 2013, *MNRAS*, 432, 1928
- Troxel M. et al., 2018a, *Phys. Rev. D*, 98, 043528
- Troxel M. A. et al., 2018b, *MNRAS*, 479, 4998
- Varshalovich D. A., Moskalev A. N., Khersonskii V. K., 1988, *Quantum Theory of Angular Momentum*, World Scientific Publishing Co. Pte. Ltd., Singapur
- White M., Padmanabhan N., 2015, *J. Cosmol. Astropart. Phys.*, 2015, 058
- Wild V. et al., 2005, *MNRAS*, 356, 247
- Xavier H. S., Abdalla F. B., Joachimi B., 2016, *MNRAS*, 459, 3693

APPENDIX A: CURVED SKY FORMALISM

The angular clustering correlation function of galaxies $w(\theta)$ is given in terms of the galaxy clustering power spectrum C_ℓ^{gg} as

$$w(\theta) = \sum_{\ell} \frac{2\ell + 1}{4\pi} P_{\ell}(\cos \theta) \left(C_{\ell}^{gg} + \frac{1}{n} \right), \quad (\text{A1})$$

where n is the galaxy density per steradian. The term proportional to $\frac{1}{n}$ is usually omitted (cf. Ross et al. 2011) since it sums up to¹⁶

$$\begin{aligned} \frac{1}{2\pi n} \sum_{\ell} \frac{2\ell + 1}{2} P_{\ell}(\cos \theta) &= \frac{1}{2\pi n} \delta_D(\cos \theta - 1) \\ &= \frac{\delta_D(\theta)}{2\pi n \sin \theta}, \end{aligned} \quad (\text{A2})$$

which has to be interpreted as a two-dimensional Dirac delta function on the sphere.

According to de Putter & Takada (2010) (see also Stebbins 1996), the galaxy–galaxy lensing correlation function $\gamma_{\ell}(\theta)$ is given in terms of the galaxy-convergence cross-power spectrum $C_{\ell}^{g\kappa}$ as

$$\gamma_{\ell}(\theta) = \sum_{\ell} \frac{2\ell + 1}{4\pi} \frac{P_{\ell}^2(\cos \theta)}{\ell(\ell + 1)} C_{\ell}^{g\kappa}, \quad (\text{A3})$$

where P_{ℓ}^m are the associated Legendre Polynomials.

Finally, the cosmic shear correlation functions $\xi_{\pm}(\theta)$ are given by

$$\xi_{\pm}(\theta) = \sum_{\ell \geq 2} \frac{2\ell + 1}{4\pi} \frac{2(G_{\ell,2}^{+}(x) \pm G_{\ell,2}^{-}(x))}{\ell^2(\ell + 1)^2} C_{\ell}^E, \quad (\text{A4})$$

where $x = \cos \theta$ and C_{ℓ}^E is the E-mode power spectrum of shear, and we assume B modes to vanish. The functions $G_{\ell,2}^{\pm}(x)$ are defined in equation (4.18)¹⁷ of Stebbins (1996). Equation (A4) can be expressed in terms of associated Legendre polynomials by using equation (4.19) of Stebbins (1996), which gives

$$G_{\ell,2}^{+}(x) \pm G_{\ell,2}^{-}(x) = P_{\ell}^2(x) \left\{ \frac{4 - \ell \pm 2x(\ell - 1)}{1 - x^2} - \frac{\ell(\ell - 1)}{2} \right\} + P_{\ell-1}^2(x) \frac{(\ell + 2)(x \mp 2)}{1 - x^2}. \quad (\text{A5})$$

In Appendix B, we show how to obtain A4 from the notation in Stebbins (1996).

It can be seen from above that each of the considered two-point correlation functions can be written in terms of the corresponding power spectra as

$$\xi^A(\theta) = \sum_{\ell} \frac{2\ell + 1}{4\pi} F_{\ell}^A(\cos \theta) C_{\ell}^A. \quad (\text{A6})$$

¹⁶from $\sum_{\ell} \frac{2\ell+1}{2} P_{\ell}(x)P_{\ell}(y) = \delta_D(x - y)$ - see N. Bronstein & A. Semendjajew (1979) for this and other properties of Legendre polynomials.

¹⁷Note that a factor of $1/\sin(\theta)$ is missing in the second line of this equation.

APPENDIX B: AVERAGING CORRELATION FUNCTIONS OVER FINITE BINS

The area average can be performed for γ_t by replacing $P_\ell^2(\cos\theta)$ with

$$\begin{aligned} \frac{\int_{\theta_{\min}}^{\theta_{\max}} d\theta \sin\theta P_\ell^2(\cos\theta)}{\cos\theta_{\min} - \cos\theta_{\max}} &= \frac{\int_{\cos\theta_{\max}}^{\cos\theta_{\min}} dx P_\ell^2(x)}{\cos\theta_{\min} - \cos\theta_{\max}} \\ &= \frac{\int_{\cos\theta_{\max}}^{\cos\theta_{\min}} dx (1-x^2) \frac{d^2}{dx^2} P_\ell(x)}{\cos\theta_{\min} - \cos\theta_{\max}}. \end{aligned} \quad (\text{B1})$$

Using integration by parts and various recursion relations of Legendre polynomials (cf. N. Bronstein and A. Semendjajew 1979), this becomes

$$\frac{\int_{\theta_{\min}}^{\theta_{\max}} d\theta \sin\theta P_\ell^2(\cos\theta)}{\cos\theta_{\min} - \cos\theta_{\max}} = \frac{1}{\cos\theta_{\min} - \cos\theta_{\max}} \left\{ \left(\ell + \frac{2}{2\ell+1} \right) [P_{\ell-1}(x)]_{\cos\theta_{\max}}^{\cos\theta_{\min}} + (2-\ell) [xP_\ell(x)]_{\cos\theta_{\max}}^{\cos\theta_{\min}} - \frac{2}{2\ell+1} [P_{\ell+1}(x)]_{\cos\theta_{\max}}^{\cos\theta_{\min}} \right\}. \quad (\text{B2})$$

In his equation (4.26), Stebbins (1996) defines the shear correlation function $C_\gamma(\theta, \phi_1, \phi_2)$ that, by inspection of equation 4.27 and Fig. 2 of this work, translates into the shear correlation functions $\xi_\pm(\theta)$ as

$$\xi_\pm(\theta) = C_\gamma(\theta, 0, 0) \pm C_\gamma(\theta, \pi/4, \pi/4). \quad (\text{B3})$$

This way, one directly arrives at our expression for the shear correlation functions, equation (A4).

To account for finite bin width in equation (A4) and in the covariance of $\hat{\xi}_\pm$, one has to perform the area-weighted bin average of the functions $[G_{\ell,2}^+(\cos\theta) \pm G_{\ell,2}^-(\cos\theta)]$. To do so, one can insert the relations

$$\begin{aligned} \int_{x_1}^{x_2} dx \frac{x P_\ell^2(x)}{1-x^2} &= \left[x \frac{dP_\ell(x)}{dx} \right]_{x_1}^{x_2} - [P_\ell(x)]_{x_1}^{x_2} \\ \int_{x_1}^{x_2} dx \frac{P_\ell^2(x)}{1-x^2} &= \left[\frac{dP_\ell(x)}{dx} \right]_{x_1}^{x_2} \end{aligned} \quad (\text{B4})$$

into equation (A5). In summary, this means that one has to exchange the functions $[G_{\ell,2}^+(\cos\theta) \pm G_{\ell,2}^-(\cos\theta)]$ as follows:

$$\begin{aligned} \frac{\int_{\cos\theta_{\max}}^{\cos\theta_{\min}} dx (G_{\ell,2}^+(x) \pm G_{\ell,2}^-(x))}{\cos\theta_{\min} - \cos\theta_{\max}} &= \left\{ - \left(\frac{\ell(\ell-1)}{2} \right) \left(\ell + \frac{2}{2\ell+1} \right) [P_{\ell-1}(x)]_{\cos\theta_{\max}}^{\cos\theta_{\min}} \right. \\ &\quad - \frac{\ell(\ell-1)(2-\ell)}{2} [xP_\ell(x)]_{\cos\theta_{\max}}^{\cos\theta_{\min}} + \frac{\ell(\ell-1)}{2\ell+1} [P_{\ell+1}(x)]_{\cos\theta_{\max}}^{\cos\theta_{\min}} + (4-\ell) \left[\frac{dP_\ell(x)}{dx} \right]_{\cos\theta_{\max}}^{\cos\theta_{\min}} \\ &\quad \left. + (\ell+2) \left\{ \left[x \frac{dP_{\ell-1}(x)}{dx} \right]_{\cos\theta_{\max}}^{\cos\theta_{\min}} - [P_{\ell-1}(x)]_{\cos\theta_{\max}}^{\cos\theta_{\min}} \right\} \right. \\ &\quad \left. \pm 2(\ell-1) \left\{ \left[x \frac{dP_\ell(x)}{dx} \right]_{\cos\theta_{\max}}^{\cos\theta_{\min}} - [P_\ell(x)]_{\cos\theta_{\max}}^{\cos\theta_{\min}} \right\} \right. \\ &\quad \left. \mp 2(\ell+2) \left[\frac{dP_{\ell-1}(x)}{dx} \right]_{\cos\theta_{\max}}^{\cos\theta_{\min}} \right\} \frac{1}{\cos\theta_{\min} - \cos\theta_{\max}}. \end{aligned} \quad (\text{B5})$$

These expressions can be very efficiently calculated and pre-tabulated e.g. with the help of the `gnu scientific library` (Galassi et al. 2009).

APPENDIX C: MASKING IN REAL SPACE COVARIANCES

DES observations do not cover the entire sky, but are located within a survey mask, described by a function $W(\hat{\mathbf{n}})$ that is =1 if we have observed the sky at location $\hat{\mathbf{n}}$ and zero otherwise.¹⁸ In this appendix, we derive how this masking changes the covariance of any measured two-point statistics.

¹⁸A map of the mask will come with a finally resolution, in which case the fractional values $0 < W < 1$ of the mask will describe the completeness of observations within the map resolution.

We start by computing, how many galaxy pairs we expect to find within our mask (assuming that galaxies do not cluster). If n_1 and n_2 are the number densities of two different tracer samples, then the expected number of pairs $dN_{\text{pair}}(\theta)$ with angular separation within $[\theta, \theta + d\theta]$ is given by (cf. Troxel et al. 2018b)

$$\begin{aligned} \frac{dN_{\text{pair}}(\theta)}{n_1 n_2} &= d\theta \int d\Omega_1 d\Omega_2 W(\hat{\mathbf{n}}_1) W(\hat{\mathbf{n}}_2) \delta_D(\arccos[\hat{\mathbf{n}}_1 \cdot \hat{\mathbf{n}}_2] - \theta) \\ &= d\theta \int d\Omega_1 d\Omega_2 W(\hat{\mathbf{n}}_1) W(\hat{\mathbf{n}}_2) \sqrt{1 - x_{12}^2} \delta_D(x_{12} - \cos \theta) \\ &= d\theta \sin \theta \int d\Omega_1 d\Omega_2 W(\hat{\mathbf{n}}_1) W(\hat{\mathbf{n}}_2) \delta_D(x_{12} - \cos \theta) \\ &= d\theta \sin \theta \sum_{\ell} \frac{2\ell + 1}{2} P_{\ell}(\cos \theta) \int d\Omega_1 d\Omega_2 W(\hat{\mathbf{n}}_1) W(\hat{\mathbf{n}}_2) P_{\ell}(x_{12}). \end{aligned} \quad (\text{C1})$$

Using the fact that

$$P_{\ell}(\hat{\mathbf{n}} \cdot \hat{\mathbf{m}}) = \frac{4\pi}{2\ell + 1} \sum_{m=-\ell}^{\ell} Y_{\ell m}(\hat{\mathbf{n}}) Y_{\ell m}^*(\hat{\mathbf{m}}) \quad (\text{C2})$$

$$\left(\Rightarrow \xi(\hat{\mathbf{n}} \cdot \hat{\mathbf{m}}) = \sum_{\ell} \frac{2\ell + 1}{4\pi} C_{\ell} P_{\ell}(\hat{\mathbf{n}} \cdot \hat{\mathbf{m}}) = \sum_{\ell m} C_{\ell} Y_{\ell m}(\hat{\mathbf{n}}) Y_{\ell m}^*(\hat{\mathbf{m}}) \right) \quad (\text{C3})$$

this becomes

$$\begin{aligned} \frac{dN_{\text{pair}}(\theta)}{n_1 n_2} &= 2\pi \sin \theta d\theta \sum_{\ell m} P_{\ell}(\cos \theta) \\ &\quad * \left(\int d\Omega_1 W(\hat{\mathbf{n}}_1) Y_{\ell m}^*(\hat{\mathbf{n}}_1) \right) \left(\int d\Omega_2 W(\hat{\mathbf{n}}_2) Y_{\ell m}(\hat{\mathbf{n}}_2) \right) \\ &= 2\pi \sin \theta d\theta \sum_{\ell m} P_{\ell}(\cos \theta) |W_{\ell m}|^2 \\ &= 2\pi \sin \theta d\theta \sum_{\ell} (2\ell + 1) P_{\ell}(\cos \theta) C_{\ell}^W \\ &= 8\pi^2 \sin \theta d\theta \xi^W(\theta), \end{aligned} \quad (\text{C4})$$

where in the last steps we have defined the angular power spectrum of the mask through $(2\ell + 1)C_{\ell}^W = \sum_m |W_{\ell m}|^2$ as well as the angular two-point function of the mask, $\xi^W(\theta)$. The total number of galaxy pairs in a finite angular bin $[\theta_{\min}, \theta_{\max}]$ is then

$$\begin{aligned} N_{\text{pair}}[\theta_{\min}, \theta_{\max}] &= (8\pi^2 n_1 n_2) \int_{\theta_{\min}}^{\theta_{\max}} d\theta \sin \theta \xi^W(\theta) \\ &= (8\pi^2 n_1 n_2) \sum_{\ell} \frac{[P_{\ell+1}(x) - P_{\ell-1}(x)]_{\theta_{\max}}^{\theta_{\min}}}{4\pi} C_{\ell}^W. \end{aligned} \quad (\text{C5})$$

The two-point correlation function of two scalar random fields, $\xi^{ab}(\theta) = \langle \delta_a(\hat{\mathbf{n}}_a) \delta_b(\hat{\mathbf{n}}_b) | \hat{\mathbf{n}}_a \cdot \hat{\mathbf{n}}_b = \cos \theta \rangle$, is in practice estimated within a finite angular bin as

$$\begin{aligned} \hat{\xi}^{ab}[\theta_{\min}, \theta_{\max}] \frac{N_{\text{pair}}[\theta_{\min}, \theta_{\max}]}{n_a n_b} &= \int_{\theta_{\min}}^{\theta_{\max}} d\theta \sin \theta \int d\Omega_a d\Omega_b W(\hat{\mathbf{n}}_a) W(\hat{\mathbf{n}}_b) \\ &\quad * \delta_D(x_{ab} - \cos \theta) \delta_a(\hat{\mathbf{n}}_a) \delta_b(\hat{\mathbf{n}}_b) \\ &= 2\pi \sum_{\ell m} \int_{\theta_{\min}}^{\theta_{\max}} d\theta \sin \theta P_{\ell}(\theta) \\ &\quad * \int d\Omega_a d\Omega_b W(\hat{\mathbf{n}}_a) W(\hat{\mathbf{n}}_b) Y_{\ell m}(\hat{\mathbf{n}}_a) Y_{\ell m}^*(\hat{\mathbf{n}}_b) \delta_a(\hat{\mathbf{n}}_a) \delta_b(\hat{\mathbf{n}}_b) \\ &= 2\pi \sum_{\ell} [P_{\ell+1}(x) - P_{\ell-1}(x)]_{\theta_{\max}}^{\theta_{\min}} \\ &\quad * \frac{1}{2\ell + 1} \sum_m \int d\Omega_a W(\hat{\mathbf{n}}_a) \delta_a(\hat{\mathbf{n}}_a) Y_{\ell m}(\hat{\mathbf{n}}_a) \\ &\quad * \int d\Omega_b W(\hat{\mathbf{n}}_b) Y_{\ell m}^*(\hat{\mathbf{n}}_b) \delta_b(\hat{\mathbf{n}}_b). \end{aligned} \quad (\text{C6})$$

From the last line, it can be seen that $\hat{\xi}^{ab}$ – apart from the normalization by $N_{\text{pair}}[\theta_{\min}, \theta_{\max}]/n_a n_b$ – is exactly the angular space counterpart of the pseudo-cell estimator in the corresponding harmonic space (cf. Efstathiou 2004). Why this is the case can be understood most easily in

the limit of an infinitesimal angular bin. In this limit

$$N_{\text{pair}}(\theta) \propto \sin \theta \xi^W(\theta)$$

$$N_{\text{pair}}(\theta) \langle \hat{\xi}^{ab}(\theta) \rangle \propto \sin \theta \xi^W(\theta) \xi^{ab}(\theta), \quad (\text{C7})$$

where the second line shows that the convolution of mask and signal power spectrum in harmonic space (Efstathiou 2004) becomes a simple multiplication in angular space. Normalization by $N_{\text{pair}}(\theta)$ especially is the angular analogue of multiplication with the inverse mode-coupling matrix that appears in harmonic space.

Let \hat{C}_ℓ^{ab} be the pseudo- C_ℓ estimator we identified in equation (C6); i.e. we write

$$\hat{\xi}^{ab}[\theta_{\min}, \theta_{\max}] \frac{N_{\text{pair}}[\theta_{\min}, \theta_{\max}]}{n_a n_b} = 2\pi \sum_{\ell} [P_{\ell+1}(x) - P_{\ell-1}(x)]_{\theta_{\max}}^{\theta_{\min}} \hat{C}_\ell^{ab} \quad (\text{C8})$$

with

$$\hat{C}_\ell^{ab} \equiv \frac{1}{2\ell+1} \sum_m (W \delta_a)_{\ell m} (W \delta_b)_{\ell m}^*. \quad (\text{C9})$$

Then, the covariance of the two-point function measurements between the fields δ_a & δ_b and δ_c & δ_d within angular bins $[\theta_-^{ab}, \theta_+^{ab}]$ and $[\theta_-^{cd}, \theta_+^{cd}]$ is given by

$$\text{Cov} \left\{ \hat{\xi}^{ab}[\theta_-^{ab}, \theta_+^{ab}], \hat{\xi}^{cd}[\theta_-^{cd}, \theta_+^{cd}] \right\} \frac{N_{\text{pair}}^{ab}[\theta_-^{ab}, \theta_+^{ab}] N_{\text{pair}}^{cd}[\theta_-^{cd}, \theta_+^{cd}]}{n_a n_b n_c n_d} = (2\pi)^2 \sum_{\ell_1 \ell_2} [P_{\ell_1+1}(x) - P_{\ell_1-1}(x)]_{\theta_+^{ab}}^{\theta_-^{ab}} [P_{\ell_2+1}(x) - P_{\ell_2-1}(x)]_{\theta_+^{cd}}^{\theta_-^{cd}} \cdot \text{Cov} \left\{ \hat{C}_{\ell_1}^{ab}, \hat{C}_{\ell_2}^{cd} \right\}. \quad (\text{C10})$$

Assuming that δ_a , δ_b , δ_c , and δ_d are Gaussian random fields and defining the symbols

$$W_{\ell_1 \ell_2 m_1 m_2} \equiv \int d\Omega W(\hat{\mathbf{n}}) Y_{\ell_1 m_1}(\hat{\mathbf{n}}) Y_{\ell_2 m_2}^*(\hat{\mathbf{n}}), \quad (\text{C11})$$

it is straightforward to show that $\text{Cov} \left\{ \hat{C}_{\ell_1}^{ab}, \hat{C}_{\ell_2}^{cd} \right\}$ is given by (cf. Efstathiou 2004)

$$\begin{aligned} \text{Cov} \left\{ \hat{C}_{\ell_1}^{ab}, \hat{C}_{\ell_2}^{cd} \right\} &= \frac{1}{(2\ell_1+1)(2\ell_2+1)} \sum_{m_1 m_2} \int d\Omega_a d\Omega_b d\Omega_c d\Omega_d \\ &\quad * W(\hat{\mathbf{n}}_a) W(\hat{\mathbf{n}}_b) W(\hat{\mathbf{n}}_c) W(\hat{\mathbf{n}}_d) \\ &\quad * Y_{\ell_1 m_1}(\hat{\mathbf{n}}_a) Y_{\ell_1 m_1}^*(\hat{\mathbf{n}}_b) Y_{\ell_2 m_2}(\hat{\mathbf{n}}_c) Y_{\ell_2 m_2}^*(\hat{\mathbf{n}}_d) \\ &\quad * \left\{ \xi^{ac}(\theta^{ac}) \xi^{bd}(\theta^{bd}) + \xi^{ad}(\theta^{ad}) \xi^{bc}(\theta^{bc}) \right\} \\ &= \frac{1}{(2\ell_1+1)(2\ell_2+1)} \sum_{m_1 m_2} \sum_{\ell_3 m_3} \sum_{\ell_4 m_4} (C_{\ell_3}^{ac} C_{\ell_4}^{bd} + C_{\ell_3}^{ad} C_{\ell_4}^{bc}) \\ &\quad * W_{\ell_1 \ell_3 m_1 m_3} W_{\ell_3 \ell_2 m_3 m_2} W_{\ell_2 \ell_4 m_2 m_4} W_{\ell_4 \ell_1 m_4 m_1}. \end{aligned} \quad (\text{C12})$$

At this point, Efstathiou (2004) and Varshalovich, Moskalev & Khersonskii (1988) follow with the approximation

$$\begin{aligned} &\approx \frac{1}{2} \frac{C_{\ell_1}^{ac} C_{\ell_2}^{bd} + C_{\ell_2}^{ac} C_{\ell_1}^{bd} + C_{\ell_1}^{ad} C_{\ell_2}^{bc} + C_{\ell_2}^{ad} C_{\ell_1}^{bc}}{(2\ell_1+1)(2\ell_2+1)} \\ &\quad * \sum_{m_1 m_2} \sum_{\ell_3 m_3} \sum_{\ell_4 m_4} W_{\ell_1 \ell_3 m_1 m_3} W_{\ell_3 \ell_2 m_3 m_2} W_{\ell_2 \ell_4 m_2 m_4} W_{\ell_4 \ell_1 m_4 m_1}. \end{aligned} \quad (\text{C13})$$

We, however, find that for fixed values of ℓ_1 and ℓ_2 the expression

$$\sum_{m_1 m_2 m_3 m_4} W_{\ell_1 \ell_3 m_1 m_3} W_{\ell_3 \ell_2 m_3 m_2} W_{\ell_2 \ell_4 m_2 m_4} W_{\ell_4 \ell_1 m_4 m_1} \quad (\text{C14})$$

has four maxima at $[\ell_3 = \ell_1, \ell_4 = \ell_2]$, $[\ell_3 = \ell_2, \ell_4 = \ell_1]$, $[\ell_3 = \ell_1, \ell_4 = \ell_1]$, and at $[\ell_3 = \ell_2, \ell_4 = \ell_2]$ and that the area around each of these maxima contributes a similar amount to the sums over ℓ_3 and ℓ_4 . Hence, we opt instead for the approximation

$$\begin{aligned} \text{Cov} \left\{ \hat{C}_{\ell_1}^{ab}, \hat{C}_{\ell_2}^{cd} \right\} &\approx \frac{1}{4} \left(\frac{C_{\ell_1}^{ac} C_{\ell_2}^{bd} + C_{\ell_2}^{ac} C_{\ell_1}^{bd} + C_{\ell_1}^{ad} C_{\ell_2}^{bc} + C_{\ell_2}^{ad} C_{\ell_1}^{bc}}{(2\ell_1+1)(2\ell_2+1)} + \frac{C_{\ell_1}^{ad} C_{\ell_2}^{bc} + C_{\ell_2}^{ad} C_{\ell_1}^{bc} + C_{\ell_1}^{ac} C_{\ell_2}^{bd} + C_{\ell_2}^{ac} C_{\ell_1}^{bd}}{(2\ell_1+1)(2\ell_2+1)} \right) \\ &\quad * \sum_{m_1, m_2, m_3, m_4} W_{\ell_1 \ell_3 m_1 m_3} W_{\ell_3 \ell_2 m_3 m_2} W_{\ell_2 \ell_4 m_2 m_4} W_{\ell_4 \ell_1 m_4 m_1}. \end{aligned} \quad (\text{C15})$$

In practice, both approximations yield very similar results and they are valid on scales ℓ_1, ℓ_2 that are much smaller than the typical scales of the mask W (Efstathiou 2004; Varshalovich et al. 1988). Unfortunately, the DES-Y3 analysis mask has features and holes over a large range of scales. Hence, the angular scales of interest in the 3×2 pt analysis are never strictly smaller than the scales of our mask. Hence, equation (C15)

is not sufficiently accurate in our case and infact significantly overestimates our covariance matrix. In Fig. 8, we explain a simple scheme that can be used to correct for this: Calculating the covariance of $\hat{\xi}^{ab}[\theta_{-}^{ab}]$ and θ_{+}^{ab} , $\hat{\xi}^{cd}[\theta_{-}^{cd}]$ and θ_{+}^{cd} within the Gaussian covariance model (see also Section 3) requires integration over all pairs of locations within our survey mask that fall into the angular bins $[\theta_{-}^{ab}, \theta_{+}^{ab}]$ and $[\theta_{-}^{cd}, \theta_{+}^{cd}]$. Schematically, the covariance then depends on expressions of the form

$$\text{Cov} \propto \int_{(ab) \in \text{mask, bin}} d\Omega^a d\Omega^b \int_{(cd) \in \text{mask, bin}} d\Omega^a d\Omega^b \xi^{ac}(\theta^{ac}) \xi^{bd}(\theta^{bd}) + \dots \quad (\text{C16})$$

Fig. 8 visualizes this for the mixed terms in the covariance, where one of the correlation functions ξ^{ac} or ξ^{bd} is due to sampling noise such as shape noise of shot noise and hence is proportional to a Dirac delta function. In that case, the integration is over pairs that share one end point. Now, the approximation made e.g. in Efstathiou (2004) or by our equation (C15) assumes that also the correlation function between the other two end points effectively acts as a delta function – at least with respect to the smallest scale features in the survey mask. We find that this is not the case for the DES-Y3 mask and that it contains features on all scales relevant to our analysis. However, Fig. 8 also indicates a simple way to fix this: Approximating the integrand over the distance of the two remaining end points by a delta function roughly overestimates the integral by a factor equal to one over the fraction of small-scale hole in the survey footprint compared to the average scale at which the two-point function between the two end points decays. Also, multiplying the mixed terms in the covariance by this fraction can serve as a next-to-leading-order correction to our equation (C15). By similar arguments, one can deduce that the cosmic variance terms (terms where none of the end points must be joined) can be corrected by performing this multiplication twice.

To implement this correction, we draw circles within the DES-Y3 survey footprint with radii ranging from 5 to 20 arcmin and measure the masking fraction in these circles. We find that this fraction is ≈ 90 per cent across the considered scales. Multiplying the mixed terms in the covariance by that fraction and the cosmic variance terms by the square of that fraction (and using equation C15), we indeed find significant improvement of the maximum posterior χ^2 obtained for the FLASK simulations (cf. lower panel of Fig. 8 as well as Fig. 1).

We end this appendix by further simplifying equation (C15). Using the completeness of the $Y_{\ell m}$ as well as the fact that $W(\hat{\mathbf{n}})^2 = W(\hat{\mathbf{n}})$, one can show that (Efstathiou 2004)

$$\sum_{m_1, m_2, m_3, m_4} W_{\ell_1 \ell_3 m_1 m_3} W_{\ell_3 \ell_2 m_3 m_2} W_{\ell_2 \ell_4 m_2 m_4} W_{\ell_4 \ell_1 m_4 m_1} = |W_{\ell_1 \ell_2 m_1 m_2}|^2. \quad (\text{C17})$$

Then, rewriting $W_{\ell_1 \ell_2 m_1 m_2}$ as

$$\begin{aligned} W_{\ell_1 \ell_2 m_1 m_2} &= \int d\Omega W(\hat{\mathbf{n}}) Y_{\ell_1 m_1}(\hat{\mathbf{n}}) Y_{\ell_2 m_2}^*(\hat{\mathbf{n}}) \\ &= \sum_{\ell m} W_{\ell m} \int d\Omega Y_{\ell m}(\hat{\mathbf{n}}) Y_{\ell_1 m_1}(\hat{\mathbf{n}}) Y_{\ell_2 m_2}^*(\hat{\mathbf{n}}) \\ &= (-1)^{m_2} \sum_{\ell m} W_{\ell m} \sqrt{\frac{(2\ell+1)(2\ell_1+1)(2\ell_2+1)}{4\pi}} \\ &\quad * \begin{pmatrix} \ell & \ell_1 & \ell_2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \ell & \ell_1 & \ell_2 \\ m & m_1 & -m_2 \end{pmatrix} \end{aligned} \quad (\text{C18})$$

and using orthogonality properties of Wigner 3j symbols, one can see that

$$\begin{aligned} \frac{\sum_{m_1 m_2} |W_{\ell_1 \ell_2 m_1 m_2}|^2}{(2\ell_1+1)(2\ell_2+1)} &= \sum_{\ell m} \frac{|W_{\ell m}|^2}{4\pi} \begin{pmatrix} \ell & \ell_1 & \ell_2 \\ 0 & 0 & 0 \end{pmatrix}^2 \\ &= \sum_{\ell} \frac{2\ell+1}{4\pi} C_{\ell}^W \begin{pmatrix} \ell & \ell_1 & \ell_2 \\ 0 & 0 & 0 \end{pmatrix}^2. \end{aligned} \quad (\text{C19})$$

$$\begin{aligned} &\Rightarrow \text{Cov} \{ \hat{C}_{\ell_1}^{ab}, \hat{C}_{\ell_2}^{cd} \} \approx \frac{1}{4} (C_{\ell_1}^{ac} C_{\ell_2}^{bd} + C_{\ell_2}^{ac} C_{\ell_1}^{bd} + C_{\ell_1}^{ac} C_{\ell_1}^{bd} + C_{\ell_2}^{ac} C_{\ell_2}^{bd} \\ &\quad + C_{\ell_1}^{ad} C_{\ell_2}^{bc} + C_{\ell_2}^{ad} C_{\ell_1}^{bc} + C_{\ell_1}^{ad} C_{\ell_1}^{bc} + C_{\ell_2}^{ad} C_{\ell_2}^{bc}) \\ &\quad * \sum_{\ell} \frac{2\ell+1}{4\pi} C_{\ell}^W \begin{pmatrix} \ell & \ell_1 & \ell_2 \\ 0 & 0 & 0 \end{pmatrix}^2 \\ &\equiv \frac{1}{4} (C_{\ell_1}^{ac} C_{\ell_2}^{bd} + C_{\ell_2}^{ac} C_{\ell_1}^{bd} + C_{\ell_1}^{ac} C_{\ell_1}^{bd} + C_{\ell_2}^{ac} C_{\ell_2}^{bd} + C_{\ell_1}^{ad} C_{\ell_2}^{bc} + C_{\ell_2}^{ad} C_{\ell_1}^{bc} + C_{\ell_1}^{ad} C_{\ell_1}^{bc} + C_{\ell_2}^{ad} C_{\ell_2}^{bc}) \mathcal{M}_{\ell_1 \ell_2}. \end{aligned} \quad (\text{C20})$$

For an efficient numerical evaluation of the above sum, we point out the following useful relation of Wigner-3j symbols (following from functions.wolfram.com/HypergeometricFunctions/ThreeJSymbol/):

$$\begin{pmatrix} \ell & \ell_1 & \ell_2 \\ 0 & 0 & 0 \end{pmatrix}^2 / \begin{pmatrix} \ell-2 & \ell_1 & \ell_2 \\ 0 & 0 & 0 \end{pmatrix}^2 = \frac{(\ell_2 - \ell_1 + \ell - 1)(\ell_1 - \ell_2 + \ell - 1)(\ell_1 + \ell_2 - \ell + 2)(\ell_1 + \ell_2 + \ell)}{(\ell_2 - \ell_1 + \ell)(\ell_1 - \ell_2 + \ell)(\ell_1 + \ell_2 - \ell + 1)(\ell_1 + \ell_2 + \ell + 1)}. \quad (\text{C21})$$

APPENDIX D: MOTIVATION FOR OUR RE-SCALING ANSATZ FOR MASKING EFFECTS

In this appendix, we make some of the arguments presented in Section 6.6 more precise. We have stated there that calculating the covariance of $\hat{\xi}^{ab}[\theta_-^{ab}, \theta_+^{ab}]$ and $\hat{\xi}^{cd}[\theta_-^{cd}, \theta_+^{cd}]$ amounts to integration over all pairs of locations within our survey mask that fall into the angular bins $[\theta_-^{ab}, \theta_+^{ab}]$ and $[\theta_-^{cd}, \theta_+^{cd}]$. Schematically, this leads to terms of the form

$$\text{Cov} = \frac{1}{\mathcal{N}} \int_{(ab) \in \text{mask, bin}} d\Omega^a d\Omega^b \int_{(cd) \in \text{mask, bin}} d\Omega^c d\Omega^d \xi^{ac}(\theta^{ac}) \xi^{bd}(\theta^{bd}) + \dots, \quad (\text{D1})$$

where compared to equation (69) we have now explicitly included a normalization factor that is proportional to the product of the number of pairs of locations within our mask that fall into the angular bins $[\theta_-^{ab}, \theta_+^{ab}]$ and $[\theta_-^{cd}, \theta_+^{cd}]$,

$$\mathcal{N} \propto N_{\text{pair}}^{ab}[\theta_-^{ab}, \theta_+^{ab}] N_{\text{pair}}^{cd}[\theta_-^{cd}, \theta_+^{cd}]. \quad (\text{D2})$$

We have argued in Section 6.6 that the approximation of Efstathiou (2004) for evaluating the impact of masking on the two-point function covariance can roughly be understood as making the replacements

$$\int d\Omega^a \dots W(\Omega^a) \xi^{ac}(\theta^{ac}) \approx \bar{\xi}^{ac} \int d\Omega^a \dots W(\Omega^a) \delta_{\text{Dirac}}^2(\Omega^a - \Omega^c) \quad (\text{D3})$$

and

$$\int d\Omega^b \dots W(\Omega^b) \xi^{bd}(\theta^{bd}) \approx \bar{\xi}^{bd} \int d\Omega^b \dots W(\Omega^b) \delta_{\text{Dirac}}^2(\Omega^b - \Omega^d), \quad (\text{D4})$$

where $\bar{\xi}^{ac}$ is the integral of $\xi^{ac}(\theta^{ac})$ over the two-dimensional angular distance vector θ^{ac} and $\bar{\xi}^{bd}$ is the integral of $\xi^{bd}(\theta^{bd})$ over θ^{bd} .

Within these approximations, the right-hand side of equation (D1) can only be non-zero if the angular bins $[\theta_-^{ab}, \theta_+^{ab}]$ and $[\theta_-^{cd}, \theta_+^{cd}]$ are identical. If that is the case, then the covariance becomes

$$\text{Cov} \approx \frac{\bar{\xi}^{ac} \bar{\xi}^{bd}}{\mathcal{N}} \int_{(ab) \in \text{mask, bin}} d\Omega^a d\Omega^b + \dots \quad (\text{D5})$$

However, the integral on the right-hand side of this equation is nothing but $N_{\text{pair}}^{ab}[\theta_-^{ab}, \theta_+^{ab}]$, so we have

$$\text{Cov} \{ \hat{\xi}^{ac}[\theta_-, \theta_+], \hat{\xi}^{bd}[\theta_-, \theta_+] \} \propto \frac{1}{N_{\text{pair}}^{ab}[\theta_-, \theta_+]}, \quad (\text{D6})$$

with proportionality coefficients that do not depend on the survey mask. So if our above understanding of the approximation proposed by Efstathiou (2004) is (at least approximately) correct, then its ratio with respect to the f_{sky} approximation (i.e. the approximation where one computes the covariance of a full-sky survey and then re-scales it with the sky fraction f_{sky} of the survey footprint) should be given by the inverse ratio of the exact value of $N_{\text{pair}}^{ab}[\theta_-, \theta_+]$ to the f_{sky} calculation

$$N_{\text{pair}, f_{\text{sky}}}^{ab}[\theta_-, \theta_+] = 4\pi^2(\theta_+^2 - \theta_-^2) f_{\text{sky}} n_a n_b. \quad (\text{D7})$$

In Fig. D1, we show the ratio of N_{pair} to $N_{\text{pair}, f_{\text{sky}}}$ (green dashed lines) and compare it to the ratios of different covariance terms when computed with either the f_{sky} approximation or the approximation of Efstathiou (2004) (cosmic variance term: blue dotted lines; mixed term: solid orange lines). The upper panel of the figure computes these ratios for the galaxy clustering two-point function $w(\theta)$ in the first lens bin of our fiducial configuration while the lower panel considers the cosmic shear correlation function $\xi_+(\theta)$ for our first source bins (other redshift bins and two-point functions behave similarly). For $w(\theta)$, these different ratios indeed closely agree with each other. The agreement between the ratio of pair counts and the ratio of the different approximations for the mixed terms is especially striking. It is most likely caused by the fact that for the mixed covariance terms one of the two replacements in equations (D3) and (D4) is actually exact. Even for $\xi_+(\theta)$, the ratio of the different approximations for the mixed term is well described by the ratio of pair counts on most scales. For the cosmic variance of $\xi_+(\theta)$, one can on the other hand observe a strong deviation. This does not necessarily signify a breakdown of our arguments but may be caused by the fact that cosmic shear is a spin-2 field. The calculations of Efstathiou (2004) do in fact only hold for scalar fields and our extension of their formulae to shear correlation functions is only approximate (see e.g. Challinor & Chon 2005, for more general calculations). Since we have identified the mixed terms to carry the strongest impact of masking on the total covariance, we do not address this any further. Instead, we consider the agreement between pair count ratios and mixed term ratios observed in Fig. D1 as sufficient justification for the re-scaling ansatz of the different covariance terms presented in Section 6.6.

APPENDIX E: PME TO INVESTIGATE THE IMPACT OF MASKING ON INDIVIDUAL COVARIANCE TERMS

In this appendix, we briefly summarize the PME method that went into Fig. 9. The covariance of the 2×2 pt (i.e. non-cosmic shear) part of our data vector has contributions from shape noise because of the presence of the mixed term described in Section 4. To pinpoint further which parts of our analytical covariance contribute to the elevation in χ^2 (and to further motivate our heuristic modelling ansatz for masking effect in the covariance presented in Section 6.6), we rerun 100 of the FLASK simulations with shape noise turned off. We then use the covariances

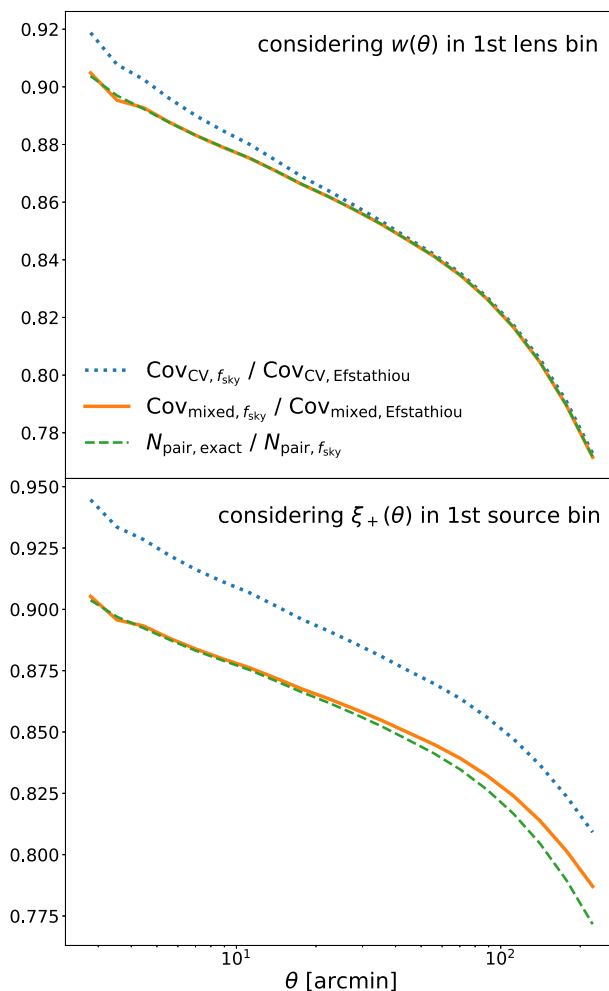


Figure D1. Ratio of exact galaxy pair counts N_{pair} to pair counts computed using equation (D7) (green dashed lines) compared to the ratios of different covariance terms when computed with either the f_{sky} approximation or the approximation of Efstathiou (2004) (cosmic variance term: blue dotted lines; mixed term: solid orange lines).

estimated from the different FLASK runs to derive corrections to our covariance model. This can be done – even with only a limited number of simulations – with the method of PME that was described by Friedrich & Eifler (2018). At the first order, their expansion estimates the precision matrix Ψ (i.e. the inverse covariance matrix) as

$$\hat{\Psi} = \mathbf{C}_{\text{model}}^{-1} - \mathbf{C}_{\text{model}}^{-1}(\hat{\mathbf{B}} - \mathbf{B}_{\text{model}})\mathbf{C}_{\text{model}}^{-1}. \quad (\text{E1})$$

Here, the matrix $\mathbf{B}_{\text{model}}$ can be the full covariance model, in which case $\hat{\mathbf{B}}$ is the full covariance estimated from FLASK, or it could be the shape-noise free part of the covariance, in which case $\hat{\mathbf{B}}$ will be the covariance estimated from the shape-noise free FLASK simulations. Friedrich & Eifler (2018) have also derived a second-order correction to equation (E1), but given the small magnitude of our observed χ^2 elevation we restrict ourselves to the first-order expansions, which should also reduce the noise of the PME. Note that equation (E1) does not contain the inverse of any noisy matrix. This is why PME works well even in the presence of only few numerical simulations (a benefit that is even further boosted because the matrix \mathbf{B} can be chosen to represent only sub-parts of the covariance).

For each FLASK measurement of the 2×2 pt data vector, we estimate the first-order PME from the remaining 196 FLASK data vectors (respectively from the ~ 100 shape-noise free data vectors). The average resulting χ^2 values between each data vector and the mean of all data vectors are displayed in Fig. 9 and compared to the χ^2 values obtained when applying the analytical masking corrections presented in Section 6.6 to either the shape-noise free covariance terms or the full covariance. The average χ^2 when using the analytical, best-guess covariance matrix is ≈ 318.8 for a total of 302 data points in the 2×2 pt data vector. This corresponds to a bias in χ^2 of about 5.5 per cent. The PME estimate of the inverse covariance manages to push this down to ≈ 307.9 (≈ 304.8 with our analytic ansatz), hence decreasing the bias in χ^2 to about 1.9 per cent (< 1 per cent for the analytic ansatz). If the PME correction term is computed with the shape-noise free FLASK covariance, then the bias only slightly reduces to $\langle \chi^2 \rangle \approx 314.3$ (≈ 316.6 with our analytic ansatz). Hence, the shape-noise-dependent mixed terms in the covariance indeed seem to be the main cause of our remaining χ^2 offset. This was also found by Joachimi et al. (2020) for the latest analysis of the KiDS. These mixed terms do not depend on the connected four-point function of the density field (cf. Section 3) and the

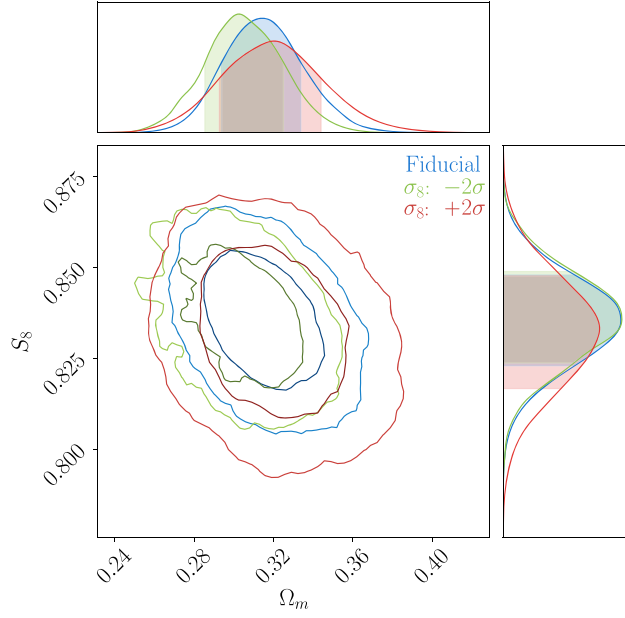


Figure F1. (S_8, Ω_m) constraints for a given noisy realization of the DES-Y3 3×2 pt data vector analysed using: a fiducial covariance matrix evaluated at $\sigma_8 = \sigma_8^{\text{fiducial}} - 2\sigma_{\sigma_8}$ (green); and a covariance matrix evaluated at $\sigma_8 = \sigma_8^{\text{fiducial}} + 2\sigma_{\sigma_8}$ (red). The green and red posteriors were obtained by importance sampling the fiducial samples.

only approximation we make in their calculation is the treatment of our survey footprint through the f_{sky} approximation. Hence, we follow Joachimi et al. (2020) in our conclusion that this approximation is the main driver of the residual errors in our covariance model.

APPENDIX F: IMPACT OF EXTREME COSMOLOGIES ON PARAMETER CONSTRAINTS

To demonstrate that the importance sampling technique employed in Section 6.8 indeed manages to capture even strong changes in the likelihood, we repeat the tests presented there with covariance matrices that drastically differ from our fiducial covariance model. In particular, we shift the value of σ_8 for which the covariance model is evaluated by $\pm 2\sigma$ of the marginalized σ_8 constraints expected from DES-Y3. Note that it is a radical change because it ignores parameter degeneracies; i.e. such a shift of σ_8 without changes in other parameters would be detected at high significance. Fig. F1 shows the likelihood contours in the S_8 – Ω_m plane obtained both from our fiducial covariance and from importance sampling with the altered covariance matrices. One can now clearly see a change in contour width. However, as we have shown in Section 6.8, this effect is far less significant for realistic parameter uncertainties in the covariance model.

APPENDIX G: EFFECTIVE SHAPE NOISE WHEN USING METACALIBRATION

In Section 6.10, we have considered how the sampling noise contribution to covariance of the galaxy–galaxy lensing two-point function can be expressed in terms of an effective shape noise when each source galaxy is weighted by a certain weight (e.g. weight w_j for the j th galaxy, with the average weight $\langle w_j \rangle_j = 1$). The expressions derived there have to change when taking into account responses R_j of a shape catalogue generated with metacalibration (Sheldon & Huff 2017). In that case, a measurement of $\hat{\gamma}_t[\theta_1, \theta_2]$ becomes

$$\hat{\gamma}_t[\theta_1, \theta_2] = \frac{\sum_{\text{pxl } i, \text{ source } j} \Delta_{[\theta_1, \theta_2]}^{ij} \delta_{l,i} \epsilon_{t,j \rightarrow i} w_j^s}{\sum_{\text{pxl } i, \text{ source } j} \Delta_{[\theta_1, \theta_2]}^{ij} w_j^s R_j}. \quad (\text{G1})$$

One can rewrite this to conform with the derivations of Section 6.10 by defining

$$\begin{aligned} \hat{\gamma}_t[\theta_1, \theta_2] &= \frac{\sum_{\text{pxl } i, \text{ source } j} \Delta_{[\theta_1, \theta_2]}^{ij} \delta_{l,i} \frac{\epsilon_{t,j \rightarrow i}}{R_j} w_j^s R_j}{\sum_{\text{pxl } i, \text{ source } j} \Delta_{[\theta_1, \theta_2]}^{ij} w_j^s R_j} \\ &\equiv \frac{\sum_{\text{pxl } i, \text{ source } j} \Delta_{[\theta_1, \theta_2]}^{ij} \delta_{l,i} \tilde{\epsilon}_{t,j \rightarrow i} \tilde{w}_j^s}{\sum_{\text{pxl } i, \text{ source } j} \Delta_{[\theta_1, \theta_2]}^{ij} \tilde{w}_j^s}. \end{aligned} \quad (\text{G2})$$

Now the transformed weights \tilde{w}_j^s should be normalized to $\langle \tilde{w}_j^s \rangle = 1$ and then be used together with the transformed ellipticities $\tilde{\epsilon}_j$ to calculate $\sigma_{\epsilon, \text{eff}}$ from equation (90). This is what we have done for Fig. 12.

¹Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

- ²Churchill College, University of Cambridge, Cambridge CB3 0DS, UK
- ³Instituto de Física Teórica, Universidade Estadual Paulista, São Paulo, 01140-070, Brazil
- ⁴Laboratório Interinstitucional de e-Astronomia – LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ 20921-400, Brazil
- ⁵Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA
- ⁶ICTP South American Institute for Fundamental Research Instituto de Física Teórica, Universidade Estadual Paulista, São Paulo, 01140-070, Brazil
- ⁷Fermi National Accelerator Laboratory, PO Box 500, Batavia, IL 60510, USA
- ⁸Department of Astronomy/Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721-0065, USA
- ⁹Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
- ¹⁰Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA
- ¹¹Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA
- ¹²Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, PO Box 2450, Stanford, CA 94305, USA
- ¹³Department of Physics and Astronomy, University of Hawaii, Watanabe 416, 2505 Correa Road, Honolulu, HI 96822, USA
- ¹⁴Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA
- ¹⁵Department of Physics, The Ohio State University, Columbus, OH 43210, USA
- ¹⁶Institut d’Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain
- ¹⁷Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain
- ¹⁸Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439, USA
- ¹⁹Physics Department, 2320 Chamberlin Hall, University of Wisconsin-Madison, 1150 University Avenue, Madison, WI 53706-1390, USA
- ²⁰Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA
- ²¹SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA
- ²²Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15312, USA
- ²³Instituto de Astrofísica de Canarias, E-38205 La Laguna, Tenerife, Spain
- ²⁴Departamento Astrofísica, Universidad de La Laguna, E-38206 La Laguna, Tenerife, Spain
- ²⁵Department of Astronomy, University of Illinois at Urbana–Champaign, 1002 West Green Street, Urbana, IL 61801, USA
- ²⁶National Center for Supercomputing Applications, 1205 West Clark Street, Urbana, IL 61801, USA
- ²⁷Jodrell Bank Center for Astrophysics, School of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK
- ²⁸Department of Astronomy, University of California, Berkeley, 501 Campbell Hall, Berkeley, CA 94720, USA
- ²⁹Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA
- ³⁰Department of Physics & Astronomy, University College London, Gower Street, London WC1E 6BT, UK
- ³¹Institut de Física d’Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, E-08193 Bellaterra (Barcelona), Spain
- ³²Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA
- ³³Département de Physique Théorique and Center for Astroparticle Physics, Université de Genève, 24 Quai Ernest Ansermet, CH-1211 Geneva, Switzerland
- ³⁴Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, UK
- ³⁵Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, 28040, Spain
- ³⁶Brookhaven National Laboratory, Bldg 510, Upton, NY 11973, USA
- ³⁷Department of Physics, Duke University Durham, Durham, NC 27708, USA
- ³⁸Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, CP 66318, São Paulo, SP 05314-970, Brazil
- ³⁹Instituto de Física Teórica UAM/CSIC, Universidad Autónoma de Madrid, E-28049 Madrid, Spain
- ⁴⁰Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK
- ⁴¹CNRS, UMR 7095, Institut d’Astrophysique de Paris, F-75014 Paris, France
- ⁴²Sorbonne Universités, UPMC Univ Paris 06, UMR 7095, Institut d’Astrophysique de Paris, F-75014 Paris, France
- ⁴³Department of Physics and Astronomy, Pevensey Building, University of Sussex, Brighton BN1 9QH, UK
- ⁴⁴INAF – Osservatorio Astronomico di Trieste, Via G. B. Tiepolo 11, I-34143 Trieste, Italy
- ⁴⁵Institute for Fundamental Physics of the Universe, Via Beirut 2, I-34014 Trieste, Italy
- ⁴⁶Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ 20921-400, Brazil
- ⁴⁷Department of Physics, IIT Hyderabad, Kandi, Telangana 502285, India
- ⁴⁸Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA
- ⁴⁹Institute of Theoretical Astrophysics, University of Oslo, PO Box 1029 Blindern, NO-0315 Oslo, Norway
- ⁵⁰Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
- ⁵¹School of Mathematics and Physics, University of Queensland, Brisbane, QLD 4072, Australia
- ⁵²Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA
- ⁵³Australian Astronomical Optics, Macquarie University, North Ryde, NSW 2113, Australia
- ⁵⁴Lowell Observatory, 1400 Mars Hill Road, Flagstaff, AZ 86001, USA
- ⁵⁵Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain
- ⁵⁶Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA
- ⁵⁷School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK
- ⁵⁸Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
- ⁵⁹Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, D-85748 Garching, Germany
- ⁶⁰Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstr 1, D-81679 München, Germany

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.