

# Dark Energy Survey Year 3 results: cosmology with moments of weak lensing mass maps – validation on simulations

M. Gatti,<sup>1★</sup> C. Chang,<sup>2,3</sup> O. Friedrich,<sup>4</sup> B. Jain,<sup>5</sup> D. Bacon,<sup>6</sup> M. Crocce,<sup>7,8</sup> J. DeRose,<sup>9,10</sup> I. Ferrero,<sup>11</sup> P. Fosalba,<sup>7,8</sup> E. Gaztanaga,<sup>7,8</sup> D. Gruen,<sup>9,10,12</sup> I. Harrison,<sup>13</sup> N. Jeffrey,<sup>14</sup> N. MacCrann,<sup>15,16</sup> T. McClintock,<sup>17</sup> L. Secco,<sup>5</sup> L. Whiteway,<sup>14</sup> T. M. C. Abbott,<sup>18</sup> S. Allam,<sup>19</sup> J. Annis,<sup>19</sup> S. Avila,<sup>20</sup> D. Brooks,<sup>14</sup> E. Buckley-Geer,<sup>19</sup> D. L. Burke,<sup>10,12</sup> A. Carnero Rosell,<sup>21,22</sup> M. Carrasco Kind,<sup>23,24</sup> J. Carretero,<sup>1</sup> R. Cawthon,<sup>25</sup> L. N. da Costa,<sup>22,26</sup> J. De Vicente,<sup>21</sup> S. Desai,<sup>27</sup> H. T. Diehl,<sup>19</sup> P. Doel,<sup>14</sup> T. F. Eifler,<sup>28,29</sup> J. Estrada,<sup>19</sup> S. Everett,<sup>30</sup> A. E. Evrard,<sup>31,32</sup> J. Frieman,<sup>3,19</sup> J. García-Bellido,<sup>20</sup> D. W. Gerdes,<sup>31,32</sup> R. A. Gruendl,<sup>23,24</sup> J. Gschwend,<sup>22,26</sup> G. Gutierrez,<sup>19</sup> D. J. James,<sup>30</sup> M. D. Johnson,<sup>24</sup> E. Krause,<sup>28</sup> K. Kuehn,<sup>33,34</sup> M. Lima,<sup>22,35</sup> M. A. G. Maia,<sup>22,26</sup> M. March,<sup>5</sup> J. L. Marshall,<sup>36</sup> P. Melchior,<sup>37</sup> F. Menanteau,<sup>23,24</sup> R. Miquel,<sup>1,38</sup> A. Palmese,<sup>3,19</sup> F. Paz-Chinchón,<sup>23,24</sup> A. A. Plazas,<sup>37</sup> C. Sánchez,<sup>5</sup> E. Sanchez,<sup>21</sup> V. Scarpine,<sup>19</sup> M. Schubnell,<sup>32</sup> S. Santiago,<sup>7,8</sup> I. Sevilla-Noarbe,<sup>21</sup> M. Smith,<sup>39</sup> M. Soares-Santos,<sup>40</sup> E. Suchyta,<sup>41</sup> M. E. C. Swanson,<sup>24</sup> G. Tarle,<sup>32</sup> D. Thomas,<sup>6</sup> M. A. Troxel,<sup>42</sup> and J. Zuntz,<sup>43</sup> (Des Collaboration)

*Affiliations are listed at the end of the paper*

Accepted 2020 August 17. Received 2020 July 20; in original form 2019 November 15

## ABSTRACT

We present a simulated cosmology analysis using the second and third moments of the weak lensing mass (convergence) maps. The second moment, or variances, of the convergence as a function of smoothing scale contains information similar to standard shear two-point statistics. The third moment, or the skewness, contains additional non-Gaussian information. The analysis is geared towards the third year (Y3) data from the Dark Energy Survey (DES), but the methodology can be applied to other weak lensing data sets. We present the formalism for obtaining the convergence maps from the measured shear and for obtaining the second and third moments of these maps given partial sky coverage. We estimate the covariance matrix from a large suite of numerical simulations. We test our pipeline through a simulated likelihood analyses varying 5 cosmological parameters and 10 nuisance parameters and identify the scales where systematic or modelling uncertainties are not expected to affect the cosmological analysis. Our simulated likelihood analysis shows that the combination of second and third moments provides a 1.5 per cent constraint on  $S_8 \equiv \sigma_8(\Omega_m/0.3)^{0.5}$  for DES Year 3 data. This is 20 per cent better than an analysis using a simulated DES Y3 shear two-point statistics, owing to the non-Gaussian information captured by the inclusion of higher order statistics. This paper validates our methodology for constraining cosmology with DES Year 3 data, which will be presented in a subsequent paper.

**Key words:** cosmology: observations – large-scale structure of Universe – gravitational lensing: weak.

## 1 INTRODUCTION

A map of the mass distribution of the Universe, or the large-scale structure (LSS), contains a vast amount of cosmological information. A given cosmological model predicts the spatial statistics of the mass distribution as well as its evolution over time. One of the cleanest ways to probe the mass distribution in the Universe is through weak (gravitational) lensing. Gravitational lensing refers to the phenomenon that light rays from distant galaxies bend as they travel through space–time, causing distortion of the observed galaxy images. This is because the space–time is perturbed by mass distribution between the galaxy and the observer according to Gen-

eral Relativity (Einstein 1936). Weak lensing is the regime where this perturbation is small; its effect is usually much smaller than the noise on a single galaxy basis, and the signal is extracted statistically using very large ensembles of galaxies. As lensing is a purely gravitational effect, it is directly sensitive to the total mass distribution compared to other cosmological probes that use galaxies as tracers of the mass density field, such as galaxy clustering (for a review of weak gravitational lensing see e.g. Bartelmann & Schneider 2001).

A key element of a weak lensing analysis is to have a large number of galaxies with well-measured shapes. This means that we need (1) cosmological surveys that collect photons from as many galaxies as possible, and (2) well-controlled systematic errors in the shape measurement of these galaxies. Motivated by the potential cosmological power of weak lensing, photometric galaxy surveys targeted at weak lensing science have been operating over the past

\* E-mail: mgatti@ifae.es

two decades. Today, unprecedented large galaxy surveys such as the Dark Energy Survey (DES; Flaugher 2005), the Hyper Suprime-Cam (HSC) Subaru Strategic Program (Aihara et al. 2018), the Kilo-Degree Survey (KiDS; de Jong et al. 2013) are all pushing the limits of weak lensing measurements.

Most of the current weak lensing analyses have focused on tomographic two-point correlation measurements (e.g. Hildebrandt et al. 2017; Troxel et al. 2018; Hikage et al. 2019). With the past two decades of work, the theoretical modelling of the shear two-point correlation function has matured significantly. Although there is still active research on, for example, the modelling of the small scales and of non-linear lensing corrections, the baseline theory of shear two-point correlation function is reasonably robust. State-of-the-art data sets from the first year (Y1) of DES currently give the tightest constraints from cosmic shear surveys on the Universe’s clustering amplitude under a Lambda cold dark matter ( $\Lambda$ CDM) cosmology,  $S_8 \equiv \sigma_8 \sqrt{\Omega_m/0.3} = 0.782^{+0.027}_{-0.027}$  (Troxel et al. 2018). The parameter  $S_8$  that is a combination of  $\sigma_8$  (the amplitude of structure in the present-day Universe, parametrized as the standard deviation of the linear overdensity fluctuations on a  $8 h^{-1}$  Mpc scale) and  $\Omega_m$  (the density of the total matter today) is designed to be approximately the parameter most constrained by weak lensing observations. We note that these constraints are at a level similar to those provided by the cosmic microwave background (CMB) from the *Planck* satellite  $S_8 = 0.841^{+0.027}_{-0.025}$ , when marginalizing over neutrino mass and considering the same parameter space as DES (see Troxel et al. 2018, table III).

However, there is much more information stored in the matter fields beyond what can be captured by two-point statistics. Two-point correlation functions only capture the Gaussian information stored in the field, while it is well known that the probability distribution function (PDF) of the galaxy density contrast in the late Universe has a one-point distribution that is approximated better as lognormal than Gaussian (Hubble 1934; Coles & Jones 1991; Wild et al. 2005). Over the years, efforts have been made to explore statistics beyond two-point for cosmology. These include three-point correlation functions and bi-spectrum (Takada & Jain 2003, 2004; Semboloni et al. 2011; Fu et al. 2014), weak lensing peak statistics (Dietrich & Hartlap 2010; Kratochvil, Haiman & May 2010; Liu et al. 2015; Kacprzak et al. 2016; Martinet et al. 2018; Peel et al. 2018; Shan et al. 2018), higher moments of the weak lensing convergence field (Van Waerbeke et al. 2013; Petri et al. 2015; Vicinanza et al. 2016; Chang et al. 2018; Peel et al. 2018; Vicinanza et al. 2018), the PDF of the weak lensing convergence field (Patton et al. 2016), density-split statistics (Friedrich et al. 2018; Gruen et al. 2018), Minkowski functionals (Kratochvil et al. 2012; Petri et al. 2015; Vicinanza et al. 2019; Parroni et al. 2020), and the Minimum Spanning Tree (MST; Naidoo et al. 2019). For some of these summary statistics (peak statistics, Minkowski functionals), one major challenge is that no analytical theoretical prediction of the target statistics exist and cosmological constraints must come from a large number of numerical simulations that span a range of cosmological parameters. In addition, these simulations also need to be closely matched to data and it is not clear what the requirements are for the matching between simulation and data (though there exists some work in systematically addressing this question, e.g. Bruderer et al. 2016; Kacprzak et al. 2019). With the increasingly large data sets, the demand for simulations for these statistics become increasingly hard to meet. For the other statistics where analytical forms exist (three-point function, higher moments, PDF, and density split statistics), most of the exploration work has been carried out with idealized simulations that in many respects do not represent the survey data. One of the reasons for this is

that once one moves beyond two-point statistics, the computation of the estimator and the theoretical modelling of the signal become more complicated. This means that the noise and systematic effects propagate non-trivially.

In this first paper of a series of two, we focus on developing the methodology of using the second and third moments of the weak lensing convergence field to constrain cosmology using the third year (Y3) of DES data. The modelling of second and third moments is based on theoretical predictions, rather than relying on large suites of  $N$ -body simulations. The goal of this first paper is to describe and validate the methodology using simulations, determining the lower bounds on scales where systematic or modelling uncertainties are not expected to affect the cosmological analysis. A companion paper applying this framework to the DES Y3 data will follow, further discussing observational systematic null tests not addressed here and testing the consistency of the constraints with the results from other DES Y3 probes or external data sets (e.g. *Planck*).

First, studied in Jain & Seljak (1997), the moments of the weak lensing convergence field is one of the simpler high-order statistics both in terms of the measurement and in terms of the theoretical modelling. Several papers (e.g. Gaztanaga & Bernardeau 1998; Fosalba et al. 2008; Van Waerbeke et al. 2013; Pujol et al. 2016; Chang et al. 2018) have performed various moments measurements on simulations and/or data and compared the results with theoretical predictions, although this information was not then used to place constraints on cosmological parameters. In Vafaei et al. (2010), the authors studied the tradeoff between different survey strategies in the Canada–France–Hawaii Telescope Lensing Survey (CFHTLenS) for combining two- and three-point statistics using simulations. They concluded that combining two- and three-point statistics of the convergence field could increase the cosmological constraints by 10–20 per cent, in the case of CFHTLenS data. In Petri et al. (2015), the authors used a set of simulations with different cosmological parameters to study how the moments of the convergence field can help constrain the cosmological parameters. They included up to the fourth moment and showed that the constraints improve by up to 20–30 per cent compared to the power spectrum-only constraints.

We build on the previous work and make several improvements. First, we use an analytical framework to incorporate the effect of masking, adapting a well-tested pseudo-angular power spectrum estimation formalism (pseudo- $C_\ell$  in the following). Secondly, we include several systematic effects that are commonly accounted for in shear two-point correlation function measurements and are key to obtaining unbiased cosmological constraints: namely, shear calibration bias, photometric redshift calibration uncertainty, and intrinsic alignment. Thirdly, we test how robust our statistics are to small scales, higher order lensing corrections such as reduced shear and source clustering, and to the effect of small-scales baryonic physics. Finally, we test our framework with two different sets of simulations (simple lognormal simulations and full  $N$ -body simulations that match the characteristics of the data set of interest), each suited for specific purposes. Although the simulations and analysis choices here are specific to the DES Y3 data, we note that the general approach in this paper can be easily transferred to a different data set.

The paper is organized as follows. In Section 2, we describe how to generate the weak lensing convergence maps from a shape catalogue using the Kaiser & Squires (1993) algorithm generalized to operate on a sphere. In the same section, we further show how the second and third moments of these convergence maps can be modelled, taking into account the effect of the mask as well as other systematics. In Section 3, we describe the characteristics and purpose of the two set of simulations used in this work. We test the validity of our

modelling with simulations in Section 4 and determine the regime where our model can correctly predict the second and third moments. In Section 5, we derive the final components needed for a cosmology analysis: the covariance matrix and the likelihood. We determine in Section 6, the final fiducial scale cuts by examining how the cosmological constraints are biased as a function of scale cuts, and we present a simulated likelihood analysis the cosmological constraints for DES Y3 and Y5 data. We summarize our findings in Section 7.

## 2 MAP MAKING AND THEORETICAL MODELLING

In order to extract cosmological information from weak lensing convergence maps, we need to first construct the convergence map  $\kappa$  from the observed weak lensing shear  $\gamma$ . The theoretical modelling of the moments measured from the convergence map depends on the particular procedure one took to construct the map. As such, we first describe in Section 2.1 our map construction procedure and next introduce in Section 2.2 the theoretical model of our moment measurements.

### 2.1 Map making: formalism and map making procedure

We implement a full-sky approach to obtain an estimate of the convergence field  $\kappa$  from the estimated shear  $\gamma$  (Castro, Heavens & Kitching 2005; Leistedt et al. 2017; Wallis et al. 2017). Such a full-sky formalism has been applied to both DES SV and Y1 data (Wallis et al. 2017; Chang et al. 2018), and it is a generalization of the flat-sky formalism developed in Kaiser & Squires (1993). In Wallis et al. (2017), the authors show that the convergence maps constructed using various flat-sky projection schemes could introduce up to 10 per cent error in the estimation of the curl-free modes (E modes) of the convergence and up to 20 per cent for divergence-free modes (B modes) of the convergence for an area approximately the DES Y3 footprint (5000 deg<sup>2</sup>). As a result, it is necessary that we use this full-sky formalism in this work.

At any position in comoving space  $(\chi, \theta, \phi)$ , one can relate the lensing potential  $\psi$  to the local Newtonian potential  $\Psi$  along the line of sight

$$\psi(\chi, \theta, \phi) = \frac{2}{c^2} \int_0^\chi d\chi' \frac{f_k(\chi - \chi')}{f_k(\chi)f_k(\chi')} \Psi(\chi', \theta, \phi), \quad (1)$$

where  $f_k$  assumes values of  $\sin\chi$ ,  $\chi$ ,  $\sinh\chi$  for a closed ( $k = 1$ ), flat ( $k = 0$ ), and open ( $k = -1$ ) Universe, respectively. Equation (1) implicitly assumes the Born approximation (i.e. the photons move along the unperturbed geodesics when computing their deflection angle). The lensing potential in equation (1) can be related to convergence  $\kappa$  and shear  $\gamma$  following Castro et al. (2005):

$$\kappa = \frac{1}{4}(\bar{\partial}\bar{\partial} + \bar{\partial}\bar{\partial})\psi, \quad (2)$$

$$\gamma = \gamma_1 + i\gamma_2 = \frac{1}{2}\bar{\partial}\bar{\partial}\psi, \quad (3)$$

where  $\bar{\partial}$  and  $\bar{\partial}$  are the raising and lowering operators acting on spin-weighted spherical harmonics defined in, e.g. Castro et al. (2005). Expanding  $\psi(\chi, \theta, \phi)$  in spherical harmonics leads to:

$$\psi(\chi, \theta, \phi) = \sum_{\ell m} \psi_{\ell m}(\chi)_0 Y_{\ell m}(\theta, \phi), \quad (4)$$

$$\psi_{\ell m}(\chi) = \int d\Omega \psi_{\ell m}(\chi, \theta, \phi)_0 Y_{\ell m}^*(\theta, \phi), \quad (5)$$

where  $_0Y_{\ell m}(\theta, \phi)$  are the spin-0 spherical harmonic basis set and  $\psi_{\ell m}(\chi)$  the harmonic coefficients at a given comoving distance. Analogously, we can expand  $\kappa$  and  $\gamma$

$$\kappa = \kappa_E + i\kappa_B = \sum_{\ell m} (\kappa_{E,\ell m} + i\kappa_{B,\ell m})_0 Y_{\ell m}, \quad (6)$$

$$\gamma = \gamma_1 + i\gamma_2 = 2 \sum_{\ell m} \gamma_{\ell m 2} Y_{\ell m}, \quad (7)$$

with  $_2Y_{\ell m}$  spin-2 spherical harmonics. We note that the convergence field has been divided into curl-free E modes and divergence-free B modes. One can relate the shear signal to the convergence field as follow:

$$\kappa_{E,\ell m} + i\kappa_{B,\ell m} = -\frac{1}{2}\ell(\ell+1)\Psi_{\ell m}, \quad (8)$$

$$\gamma_{\ell m} = \hat{\gamma}_{E,\ell m} + i\gamma_{B,\ell m} = \frac{1}{2}[\ell(\ell+1)(\ell-1)(\ell+2)]^{1/2}\Psi_{\ell m} \quad (9)$$

$$\gamma_{\ell m} = -\sqrt{\frac{(\ell+2)(\ell-1)}{\ell(\ell+1)}}(\kappa_{E,\ell m} + i\kappa_{B,\ell m}). \quad (10)$$

The shear field needs first to be decomposed into spherical harmonics; then E and B modes of the convergence field follows from applying equation (10). Curl-free E modes carry most of the cosmological signal. Divergence-free B modes can arise due to non-linear lensing corrections (such as deflection along the first-order Born approximation), clustering of the lenses, and reduced shear corrections (Schneider et al. 1998; Schneider, van Waerbeke & Mellier 2002; Krause & Hirata 2010). These effects are assumed to be small for current stage III weak lensing surveys (e.g. DES, KIDS, and HSC) and will not be modelled. Biases in the shear measurement pipeline or object selection biases can also produce B modes that can affect the parameters inference by few per cent (Hoekstra 2004; Asgari et al. 2018). Finally, partial sky coverage can induce mode mixing, producing spurious B modes in the reconstructed convergence maps due to E-mode leakage. This will be the only source of B modes that we will take into account in our modelling.

To apply equation (10) to data, we need an estimate of the shear field. In practice, the shear field cannot be directly measured. The observable is the reduced shear

$$g = \frac{\gamma}{1 - \kappa}. \quad (11)$$

Since galaxies have an intrinsic shape, what we actually measure is the ellipticity, or shape of the galaxy, which is a noisy estimate of the reduced shear

$$\epsilon = \frac{g + \epsilon_{\text{int}} + \epsilon_{\text{m}}}{1 + g(\epsilon_{\text{int}} + \epsilon_{\text{m}})}, \quad (12)$$

where  $\epsilon_{\text{int}}$  is the intrinsic shape of the galaxy, and  $\epsilon_{\text{m}}$  the shape measurement noise. The above equation holds for reduced shear less than 1, which is always satisfied in the weak flensing regime. The latter two quantities ( $\epsilon_{\text{int}}$  and  $\epsilon_{\text{m}}$ ) should average to zero for large number of galaxies (assuming no shear measurement biases). Moreover, in the weak lensing regime,  $\gamma, \kappa \ll 1$ , so the observed shape results in a noisy estimator for the shear field  $\epsilon \approx \gamma + \epsilon_{\text{int}} + \epsilon_{\text{m}}$  (this does not hold in case of intrinsic alignment).

We provide a short description of map making procedure here, as it has been detailed in previous DES papers (e.g. Chang et al. 2018). The maps are constructed using HEALPIX pixelization (Górski et al. 2005), with NSIDE = 1024, corresponding to a pixel size of 3.44 arcmin.

In the case where a mock galaxy catalogue with galaxy shapes were provided, the first step in the reconstruction of the mass map would involve making pixelized ellipticity maps  $\epsilon_1$  and  $\epsilon_2$  from the

catalogue. These can be obtained by averaging the two components of the shape estimate over all the galaxies belonging to a given HEALPIX pixel. We skip this step in this paper and directly work with noisy shear maps.

Next, we perform the spin transformation which converts the ellipticity maps into a curl-free E-mode convergence map  $\hat{\kappa}_E$  and a divergence-free B-mode convergence map  $\hat{\kappa}_B$ . We use the HEALPIX functions MAP2ALM to decompose the shear field in spherical harmonic space obtaining the coefficients  $\hat{\gamma}_{E,\ell m}$ ,  $\hat{\gamma}_{B,\ell m}$  and calculate  $\hat{\kappa}_{E,\ell m}$ ,  $\hat{\kappa}_{B,\ell m}$  following equation (10). Finally, we use the HEALPIX function ALM2MAP to convert these coefficients back to the real space  $\kappa_E$  and  $\kappa_B$  maps. Since the convergence reconstruction is only valid up to a constant due to the mass-sheet degeneracy, all the recovered maps are mean-subtracted before computing the moments of the maps.

The above theoretical derivation and map making procedure describes how to obtain the convergence maps from an estimate of the shear field. The method we implement in this paper does not assume any prior knowledge of the convergence field to be reconstructed. There exist methods, however, which implicitly or explicitly assume priors that improve the map reconstructions over a range of metrics (e.g. Jeffrey et al. 2018, 2019; Mawdsley et al. 2019). Some of these methods will be explored in a future DES Y3 Mass Maps paper (in preparation). We are not considering these methods here: in this paper, the convergence moments are modelled from theory, and including the effects of such priors on the maps moments will be difficult. On the other hand, these alternative methods are valuable when  $N$ -body simulations are used to model the observables (e.g. Petri et al. 2015; Fluri et al. 2018).

## 2.2 Theoretical modelling of convergence moments

We adopt the theoretical model for second and third moments (variance and skewness) of the convergence field using a non-linear extension of cosmological perturbation theory (Scoccimarro & Couchman 2001; Van Waerbeke et al. 2001; Bernardeau et al. 2002).

As we are interested in highlighting the features of our convergence field at different angular scales, we smooth our recovered convergence fields using a top-hat filter at different angular scales. We chose a top-hat filter to facilitate the analytical evaluation of third moments, but different filters with different properties can be chosen (e.g. Van Waerbeke et al. 2013 used a Gaussian filter). A top-hat filter  $W$  in harmonic space of smoothing length  $\theta_0$  is defined as

$$W_\ell(\theta_0) = \frac{P_{\ell-1}(\cos(\theta_0)) - P_{\ell+1}(\cos(\theta_0))}{(2\ell+1)(\ell - \cos(\theta_0))}, \quad (13)$$

where  $P_\ell$  are Legendre polynomials of order  $\ell$ . The variance of matter contrast  $\delta$  smoothed by such a filter at a given comoving distance  $\chi$  is

$$\langle \delta_{\theta_0, \text{NL}}^2 \rangle(\chi) = \sum_\ell \frac{2\ell+1}{4\pi} P_{\text{NL}}(\ell/\chi, \chi) F_\ell^2 W_\ell(\theta_0)^2, \quad (14)$$

where  $F_\ell$  is the pixel window function (modelled using the pixel window function provided by HEALPIX) and  $P_{\text{NL}}(\ell/\chi, \chi)$  the non-linear power spectrum. For the latter we used HALOFIT as detailed in Takahashi et al. (2014) and assumed in the fiducial DES Y3 analysis.

For the smoothed version of the third moment (or skewness) of the matter contrast, at leading order in perturbation theory it reads

$$\langle \delta_{\theta_0, \text{NL}}^3 \rangle(\chi) = S_3 [\langle \delta_{\theta_0, \text{NL}}^2 \rangle(\chi)]^2, \quad (15)$$

where  $S_3$  is the reduced skewness parameter. The analytical derivation of the reduced skewness parameter is performed to leading order, which is linear in the power spectrum, but as such predictions

perform well even in the mildly non-linear regime ( $k \approx 0.1 h^{-1}$  Mpc; Bernardeau et al. 2002), we assume their validity when a non-linear power spectrum (the HALOFIT from Takahashi et al. 2014) is used to compute the variance. We also implement a refinement (in the form of analytical fitting formulae) of the treatment of the skewness at small scales based on  $N$ -body, CDM-only simulations. In this paper, we focus on the analytical fitting formulae presented in Scoccimarro & Couchman 2001 (hereafter SC01) and Gil-Marín et al. 2012 (hereafter GM12), but we note that there are alternative formulae such as that recently presented in Takahashi et al. (2019). The SC01 and GM12 analytical refinements come with a modelling uncertainty (Van Waerbeke et al. 2001; Semboloni et al. 2011; Simon et al. 2015; Harnois-Déraps et al. 2016), which ultimately depends on the resolution of the  $N$ -body simulations that have been used to perform the fit and on the flexibility of the formulae to model all the configurations (e.g. equilateral, flattened, and squeezed) of the measured bispectrum. In this paper, we implement the analytical fitting formulae from SC01, as they provide a better fit to the  $N$ -body simulations used in this paper to validate the methodology. The analytical expression of the reduced skewness parameter is provided in Appendix A1.

The analytical expression of the second and third moments of the convergence field for a given redshift distribution are provided under the Limber approximation (Limber 1953). The Limber approximation allows us to relate the 3D spatial clustering properties of the density field to 2D projected quantities. The approximation usually breaks down at small scales and for narrow redshift distributions. Under such approximation, the second and third moments read

$$\langle \kappa_{\theta_0}^2 \rangle^{i,j} = \int d\chi \frac{q^i(\chi)q^j(\chi)}{\chi^2} \langle \delta_{\theta_0}^2 \rangle(\chi), \quad (16)$$

$$\langle \kappa_{\theta_0}^3 \rangle^{i,j,k} = \int d\chi \frac{q^i(\chi)q^j(\chi)q^k(\chi)}{\chi^4} \langle \delta_{\theta_0}^3 \rangle(\chi). \quad (17)$$

The superscripts  $i, j, k$  refer to different tomographic bins. We have dropped the subscript  $\text{NL}$  for brevity. The term  $q^i$  represents the lensing kernel and reads

$$q^i(\chi) = \frac{3H_0^2 \Omega_m}{2c^2} \frac{\chi}{a(\chi)} \int_\chi^{\chi_h} d\chi' n^i(z(\chi')) dz/d\chi' \frac{\chi' - \chi}{\chi}, \quad (18)$$

where  $H_0$  is the Hubble constant at present time,  $c$  the speed of light,  $n^i(z)$  the normalised redshift distribution of a given tomographic bin, and  $a(\chi)$  the scale factor.

We note that the variance and skewness of the convergence field have differing dependencies on the parameters  $\Omega_m$  and  $\sigma_8$  (Seljak & Zaldarriaga 1996; Bernardeau, van Waerbeke & Mellier 1997).

### 2.2.1 Effects of masking

One of the problems in estimating the convergence field from the observed shapes is that we observe only a portion of the sky. This means that the reconstruction will suffer edge effects, due to the convolution with a window function representing the survey footprint. Some methods deal with mask effects at the level of map making (Pires et al. 2009; Mawdsley et al. 2019), whereas in this work, we will account for the mask effects in our theoretical predictions using a pseudo- $C_\ell$  formalism (Brown, Castro & Taylor 2005; Hikage et al. 2011).

The pseudo- $C_\ell$  formalism correctly recovers the shear power spectrum estimated from the shear field in the case of partial sky coverage. It also predicts mode mixing (that is, part of the E modes leaks into B modes and vice versa). In particular, if we define

$$\hat{C}_\ell^{EE} = \frac{1}{2\ell+1} \sum_m |\hat{\gamma}_{E,\ell m}|^2, \quad (19)$$

$$\hat{C}_\ell^{EB} = \frac{1}{2\ell+1} \sum_m \hat{\gamma}_{E,\ell m} \hat{\gamma}_{B,\ell m}^*, \quad (20)$$

$$\hat{C}_\ell^{BB} = \frac{1}{2\ell+1} \sum_m |\hat{\gamma}_{B,\ell m}|^2, \quad (21)$$

one can write the masked (pseudo-) spectra as the convolution of the true spectra with a mode-coupling matrix

$$\hat{C}_\ell = \sum_{\ell'} \mathbf{M}_{\ell\ell'} \mathbf{C}_{\ell'}, \quad (22)$$

where we have introduced the vector  $\mathbf{C}_\ell \equiv (C_\ell^{EE}, C_\ell^{EB}, C_\ell^{BB})$ . The mode-mode coupling matrix  $\mathbf{M}$  is expressed in terms of  $M_{\ell\ell'}^{EE,EE}$ ,  $M_{\ell\ell'}^{BB,BB}$ ,  $M_{\ell\ell'}^{EB,EB}$ , and  $M_{\ell\ell'}^{EE,BB}$ . The mode-coupling matrices contain information about the survey geometry; analytical expressions for the mode-coupling matrices in terms of the window function can be found in Hikage et al. (2011) and in Appendix B. The pseudo- $C_\ell$  formalism can be incorporated in equation (16) as

$$\begin{aligned} \langle \kappa_{\theta_0}^2 \rangle^{i,j,EE/BB} &= \int d\chi \frac{q^i(\chi)q^j(\chi)}{\chi^2} \\ &\times \sum_{\ell} \frac{2\ell+1}{4\pi} f_\ell^{-1} W_\ell(\theta_0)^2 \sum_{\ell'} M_{\ell\ell'}^{EE/BB,EE} \\ &\times P_{\text{NL}}(\ell'/\chi, \chi) F_{\ell'}^2 f_{\ell'}. \end{aligned} \quad (23)$$

In the above equations, the factor  $f_\ell = [(\ell+2)(\ell-1)]/[\ell(\ell+1)]$  accounts for the fact that the mode-coupling matrix is applied to the shear field rather than to the convergence field directly. Depending on the mode-coupling matrix used ( $M_{\ell\ell'}^{EE,EE}$  or  $M_{\ell\ell'}^{BB,EE}$ ), with equation (23) we can predict the variance of both E and B modes of the recovered convergence field. As for the third moments, if we neglect the contribution of the masking to the reduced skewness parameter  $S_3$ , we can write

$$\begin{aligned} \langle \kappa_{\theta_0}^3 \rangle^{i,j,k,EE/BB} &= \int d\chi \frac{q^i(\chi)q^j(\chi)q^k(\chi)}{\chi^4} \\ &\times S_3 \left[ \langle \delta_{\theta_0, \text{NL}}^2 \rangle^{EE/BB}(\chi) \right]^2. \end{aligned} \quad (24)$$

We note that neglecting the effects of masking on  $S_3$  does not imply we are neglecting the effects on masking on the third moment  $\langle \kappa_{\theta_0}^3 \rangle$  but rather we assume that most of the effect of the mask is included in the term  $[\langle \delta_{\theta_0, \text{NL}}^2 \rangle(\chi)]^2$ . We note that Gil-Marín et al. (2015) made a similar assumption when modelling mask effects in the bispectrum predictions of SDSS DR11 BOSS galaxies, demonstrating its validity for modes smaller than the size of the footprint. We show in Section 4 that equation (24) captures the mask effects on the third moment well for the scales considered in this analysis.

### 2.2.2 Systematic effects

Astrophysical and measurement systematic effects are modelled through nuisance parameters. We marginalize over all the nuisance parameters when estimating the cosmological parameters. Values and priors are summarized in Table 1.

**2.2.2.1 Photometric redshift uncertainties** Photometric redshift uncertainties are parametrized through a shift  $\Delta z$  in the mean of the

**Table 1.** Cosmological, systematic, and astrophysical parameters. The cosmological parameters considered are  $\Omega_m$ ,  $\sigma_8$ ,  $\Omega_b$  (the baryonic density in units of the critical density),  $n_s$  (the spectral index of primordial density fluctuations), and  $h$  (the dimensionless Hubble parameter). The nuisance parameters are the multiplicative shear biases  $m_i$  and the mean photometric uncertainties of the weak lensing samples  $\Delta z_i$ . The astrophysical parameters  $A_{\text{IA},0}$  and  $\alpha_{\text{IA}}$  describe the intrinsic alignment model. We report the boundaries for both Flat and Gaussian priors. For Gaussian priors, we also report the mean and the  $1\sigma$  in the prior column. Priors are described in Section 2.2.2.

Parameter	Range	Prior
$\Omega_m$	0.1–0.9	Flat
$\sigma_8$	0.4–1.3	Flat
$h_{100}$	55–90	Flat
$n_s$	0.87–1.07	Flat
$\Omega_b$	0.03–0.07	Flat
$m_1$ – $m_4 \times 10^2$	–10.0 to 10.0	$0.0 \pm 2.3$
$\Delta z_1 \times 10^2$	–10.0 to 10.0	$0.0 \pm 1.6$
$\Delta z_2 \times 10^2$	–10.0 to 10.0	$0.0 \pm 1.3$
$\Delta z_3 \times 10^2$	–10.0 to 10.0	$0.0 \pm 1.1$
$\Delta z_4 \times 10^2$	–10.0 to 10.0	$0.0 \pm 2.2$
$A_{\text{IA},0}$	–5.0 to 5.0	Flat
$\alpha_{\text{IA}}$	–5.0 to 5.0	Flat

redshift distribution

$$n^i(z) = \hat{n}^i(z + \Delta z), \quad (25)$$

where  $\hat{n}^i$  is the original estimate of the redshift distribution coming from the photometric redshift code. We assume DES Y1 priors for the shift parameters.

**2.2.2.2 Multiplicative shear biases** Biases coming from the shear measurement pipeline are modelled through an average multiplicative parameter  $1 + m^i$  for each tomographic bin. Such parameter affects our moments in the following way:

$$\langle \kappa_{\theta_0}^2 \rangle^{i,j} \rightarrow (1 + m^i)(1 + m^j) \langle \kappa_{\theta_0}^2 \rangle^{i,j}, \quad (26)$$

$$\langle \kappa_{\theta_0}^3 \rangle^{i,j,k} \rightarrow (1 + m^i)(1 + m^j)(1 + m^k) \langle \kappa_{\theta_0}^3 \rangle^{i,j,k}. \quad (27)$$

Gaussian priors are assumed for each of the  $m^i$ .

**2.2.2.3 Intrinsic galaxy alignments (IA)** IA is modelled according to the non-linear alignment (NLA) model (Hirata & Seljak 2004; Bridle & King 2007). It can be incorporated in the modelling by modifying the lensing kernel

$$q^i(\chi) \rightarrow q^i(\chi) - A(z(\chi)) \frac{n^i(z(\chi))}{\langle n^i \rangle} \frac{dz}{d\chi}. \quad (28)$$

The NLA model is usually used in the context of two-point correlation statistics, but the above equation generalizes it to the third moments case as well. The amplitude of the IA contribution can be written as a power law

$$A(z) = A_{\text{IA},0} \left( \frac{1+z}{1+z_0} \right)^{\alpha_{\text{IA}}} \frac{c_1 \rho_{m,0}}{D(z)}, \quad (29)$$

with  $z_0 = 0.62$ ,  $c_1 \rho_{m,0} = 0.0134$  (Bridle & King 2007, Krause et al. 2017) and  $D(z)$  the linear growth factor. We marginalize over  $A_{\text{IA},0}$  and  $\alpha_{\text{IA}}$  assuming flat priors.

### 2.2.3 Moments estimator

To estimate the moments of a smoothed map, we use a simple estimator

$$\langle \hat{\kappa}_{\theta_0}^2 \rangle^{i,j} = \frac{1}{N_{\text{tot}}} \sum_{\text{pix}} \kappa_{\theta_0, \text{pix}}^i \kappa_{\theta_0, \text{pix}}^j, \quad (30)$$

$$\langle \hat{\kappa}_{\theta_0}^3 \rangle^{i,j,k} = \frac{1}{N_{\text{tot}}} \sum_{\text{pix}} \kappa_{\theta_0, \text{pix}}^i \kappa_{\theta_0, \text{pix}}^j \kappa_{\theta_0, \text{pix}}^k, \quad (31)$$

where  $i$ ,  $j$ , and  $k$  refers to different tomographic bins. The sum runs over all the pixel of the sky, also outside the footprint: the transformation from the shear field to the convergence field is non-local and some of the power is transferred outside the footprint, despite most of it remaining confined to the footprint. The lack of power outside the footprint (due to the fact that the shear field is not defined there) is taken into account by the mode-coupling matrices (equations 23, 24).

Due to the presence of shape noise, the measurement of galaxy shapes will be a noisy estimate of the shear field  $\gamma$ . This also means that our estimate of the convergence field will be noisy

$$\kappa_{E, \text{obs}} = \kappa_{E, \text{true}} + \kappa_{E, \text{noise}}, \quad (32)$$

$$\kappa_{B, \text{obs}} = \kappa_{B, \text{true}} + \kappa_{B, \text{noise}}. \quad (33)$$

In the above equations, we omitted the smoothing angle  $\theta_0$ . The contribution of the noise to the convergence field can be estimated by randomly rotating the shape of the galaxies and applying the full-sky spherical harmonics approach to obtain the convergence (Van Waerbeke et al. 2013; Chang et al. 2018). As the random rotation should completely erase the cosmological contribution, the resulting convergence signal will just contain noise and should average to 0 (but with a non-negligible variance).

It follows that when estimating second and third moments from noisy convergence maps it is necessary to properly de-noise the measured moments. Following Van Waerbeke et al. (2013):

$$\langle \hat{\kappa}^2 \rangle^{i,j} = \langle \kappa^2 \rangle^{i,j} - \langle \kappa \kappa_{\text{rand}} \rangle^{i,j} - \langle \kappa_{\text{rand}} \kappa \rangle^{i,j} - \langle \kappa_{\text{rand}}^2 \rangle^{i,j}, \quad (34)$$

$$\begin{aligned} \langle \hat{\kappa}^3 \rangle^{i,j,k} &= \langle \kappa^3 \rangle^{i,j,k} - \langle \kappa_{\text{rand}}^3 \rangle^{i,j,k} \\ &\quad - \left[ \langle \kappa_{\text{rand}}^2 \kappa \rangle^{i,j,k} - \langle \kappa_{\text{rand}} \kappa^2 \rangle^{i,j,k} + \text{cycl.} \right], \end{aligned} \quad (35)$$

where cycl. refers to the cyclic permutation of the indexes  $i$ ,  $j$ , and  $k$  for the terms in parenthesis. In the above equations, the term  $\langle \kappa_{\text{rand}}^2 \rangle^{i,j}$  is the noise-only contribution to the second moment of the tomographic bins  $i, j$ ; for  $i \neq j$  it vanishes. The map  $\kappa_{\text{rand}}$  represents the estimate of the shape noise contribution to the convergence map; it is estimated by randomly rotating the galaxy shapes. The intrinsic ellipticity distribution of observed galaxies is not expected to be perfectly Gaussian, but by the central limit theorem, it would be the correct distribution in the limit of large numbers of galaxies averaged in the pixels of the convergence map (Jeffrey et al. 2018). If this holds, also the term  $\langle \kappa_{\text{rand}}^3 \rangle^{i,j,k}$  (which is the noise-only contribution to the third moment of the tomographic bin  $i, j$ , and  $k$ ) would vanish. Additional checks will need to be performed on DES Y3 data, as we do not include potential sources of noise inhomogeneities (e.g. astrophysical or observational systematics) in this work. Finally, we note that if the convergence field and the shape noise term in a given map pixel are uncorrelated, mixed terms should be consistent with zero.

## 3 SIMULATIONS

Two different sets of simulations are used to validate our theoretical approach. These simulations differ in the complexity of the physics included, and are used to validate different parts of our methodology. In particular, we make use of:

(i) FLASK simulations (Xavier, Abdalla & Joachimi 2016). These are lognormal realizations, and are used to produce a large number of realizations (of the order of 1000) of the shear and convergence fields. They require input power spectra at the redshift of the observation for their predictions, so they cannot be used to test the modelling of the second and the third moments, as they are key ingredients to run the simulations. We use them to model the covariance matrices of our measurements and to test the modelling of mask effects.

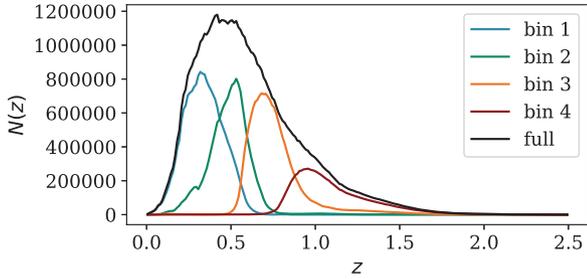
(ii) Takahashi et al. (2017, hereafter T17) mocks. We use 100 full-sky gravitational lensing convergence and shear maps obtained from full  $N$ -body simulations and a ray-tracing algorithm described in T17. We use these to validate the theoretical modelling of second and third moments over a large number of simulations. We also use them to check the effect of non-linear lensing corrections in our modelling.

Below we provide a more in-depth description of each of the simulations.

### 3.1 FLASK simulations

The FLASK software (Xavier et al. 2016) allows one to rapidly generate full-sky, lognormal realizations of a given field (in our case, the convergence field). In particular, FLASK assumes the convergence field to be described by a zero-mean shifted lognormal distribution, where the parameters of the lognormal probability distribution function (PDF) are chosen to match the variance and skewness of the input. The lognormal approximation is usually adopted for the density field (Hubble 1934; Coles & Jones 1991; Wild et al. 2005) and is not expected to exactly hold for the convergence field, as it is a weighted projection of the mass density field along the line of sight. Tests on numerical simulations showed a lognormal PDF to be a reasonable model (e.g. Taruya, Hamana & Kayo 2002; Hilbert, Hartlap & Schneider 2011), although generalized lognormal PDFs have been shown to improve the fit at the tails of the distribution (Das & Ostriker 2006; Joachimi, Taylor & Kiessling 2011; Takahashi et al. 2011). Observational evidences from DES SV (Clerkin et al. 2017) find that at intermediate scales between 10 and 20 arcmin, the convergence distributions are more lognormal than Gaussian (at larger scales noise dominates). We show in Section 6.1 that relying on the lognormal approximation to build our covariance matrix does not bias the recovery of the cosmological parameters.

The software requires as inputs a set of auto and cross power spectra and a lognormal shift parameter. This latter parameter is a combination of the variance and skewness (Xavier et al. 2016) and it is computed from theory and fixed to the value at no smoothing. Formally, this means that the third moment computed in FLASK should match theoretical predictions only at no smoothing. Slight variations can occur with a non-zero smoothing as the convergence field is not perfectly lognormal. The second moment should agree at every smoothing scale as the full power spectrum is provided. We generated theoretical predictions for the power spectra of the convergence field for four tomographic bins of our WL source sample. We used the true redshift distributions of the WL sample as measured in a fiducial DES simulated sample (DeRose et al. 2019). Redshift distributions are shown in Fig. 1. We fixed the cosmology



**Figure 1.** Redshift distributions of the 4 tomographic WL bins (and the full sample), from a fiducial DES simulated sample (DeRose et al. 2019). A bin width of  $\Delta_z = 0.01$  has been used for the histograms.

of our input power spectra to be  $\Omega_m = 0.286$ ,  $\sigma_8 = 0.82$ ,  $\Omega_b = 0.047$ ,  $n_s = 0.96$ , and  $h_{100} = 0.7$ .<sup>1</sup> We generated 1000 realizations of the convergence fields in the form of HEALPIX maps with NSIDE = 1024. This resolution is chosen based on the expected number density of the DES Y3 weak lensing sample. For each of the realizations, we cut out a DES Y3 footprint using a mask that is close to what will be used for the DES Y3 weak lensing analysis. We assign shape noise to each pixel  $\epsilon_{\text{int}}$  of the shear fields based on the expected per-component shape noise of the full DES Y3 WL sample ( $\sigma_\epsilon$ ) and the galaxy number density predicted by FLASK in each pixel ( $n_{g, \text{pix}}$ ), so as  $\epsilon_{\text{int, pix}} = \sigma_\epsilon / \sqrt{n_{g, \text{pix}} A_{\text{pix}}}$ , with  $A_{\text{pix}}$  the pixel area. The average number densities over the full footprint of each bin are, respectively, 1.38, 1.36, 1.35, 0.86 gal arcmin<sup>-2</sup>, while the  $\sigma_\epsilon$  (the standard deviation of the two components for the measured galaxy shapes) are 0.29, 0.29, 0.29, and 0.30. We use such FLASK mocks to validate our modelling of the mask effects and to generate covariance matrices for our measurements. In future sections, to quantify the offset between the third moments generated by FLASK and the theory predictions, we use the offset function defined as follows:

$$\text{offset}(\theta_0) \equiv \frac{\langle \kappa_{\theta_0}^3 \rangle_{\text{FLASK, full-sky}} - \langle \kappa_{\theta_0}^3 \rangle_{\text{theory, full-sky}}}{\langle \kappa_{\theta_0}^3 \rangle_{\text{theory, full-sky}}}, \quad (36)$$

such that  $\langle \kappa_{\theta_0}^3 \rangle_{\text{theory, full-sky}} (1 + \text{offset}(\theta_0)) = \langle \kappa_{\theta_0}^3 \rangle_{\text{FLASK, full-sky}}$ . The offset function is 0 at no smoothing (since the shift parameter provided to FLASK as input is only valid at no smoothing), and reaches values up to  $\sim 40$  per cent for  $\theta_0 \sim 200$  arcmin.

### 3.2 T17 N-body simulation

The simulations are a set of 108 full-sky lensing convergence and shear maps obtained for a range of redshifts between  $z = 0.05$  and  $5.3$  at intervals of  $150 h^{-1}$  Mpc comoving distance.

Initial conditions were generated using 2LPTIC (Croce, Puelbas & Scoccimarro 2006) and the  $N$ -body run using L-GADGET2 (Springel 2005), consistent with WMAP 9 yr results (Hinshaw et al. 2013):  $\Omega_m = 0.279$ ,  $\sigma_8 = 0.82$ ,  $\Omega_b = 0.046$ ,  $n_s = 0.97$ , and  $h = 0.7$ .

The simulations begin with 14 boxes with side lengths  $L = 450, 900, 1350, \dots, 6300 h^{-1}$  Mpc in steps of  $450 h^{-1}$  Mpc, with six independent copies at each box size and  $2048^3$  particles per box. Snapshots are taken at the redshift corresponding to the lens planes at intervals of  $150 h^{-1}$  Mpc comoving distance. The authors checked

<sup>1</sup>The values of the cosmological parameters used to compute the covariance are slightly different than the ones of the mocks used to validate the modelling of second and third moments. These values have been chosen to facilitate the comparison with other simulated cosmological analysis for DES Y3.

that the agreement of the average matter power spectra with the revised HALOFIT (Takahashi et al. 2012) was within 5 per cent for  $k < 1 h \text{ Mpc}^{-1}$  at  $z < 1$ , for  $k < 0.8 h \text{ Mpc}^{-1}$  at  $z < 3$ , and for  $k < 0.5 h \text{ Mpc}^{-1}$  at  $z < 7$ . Weak lensing quantities were estimated using the multiple plane ray-tracing algorithm GRAYTRIX (Hamana et al. 2015), and shear and convergence HEALPIX maps with resolution NSIDE = 4096 are provided. Haloes are identified in the simulation using the public code ROCKSTAR (Behroozi, Wechsler & Wu 2013). The simulations do not come with a galaxy catalogue. For each of the 108 realizations, we produced convergence maps for the 4 WL tomographic bins by stacking the convergence snapshots taking into account the redshift distributions of the bins. We used the same redshift distribution as that used in the FLASK simulations.

## 4 MODEL VALIDATION WITH SIMULATIONS

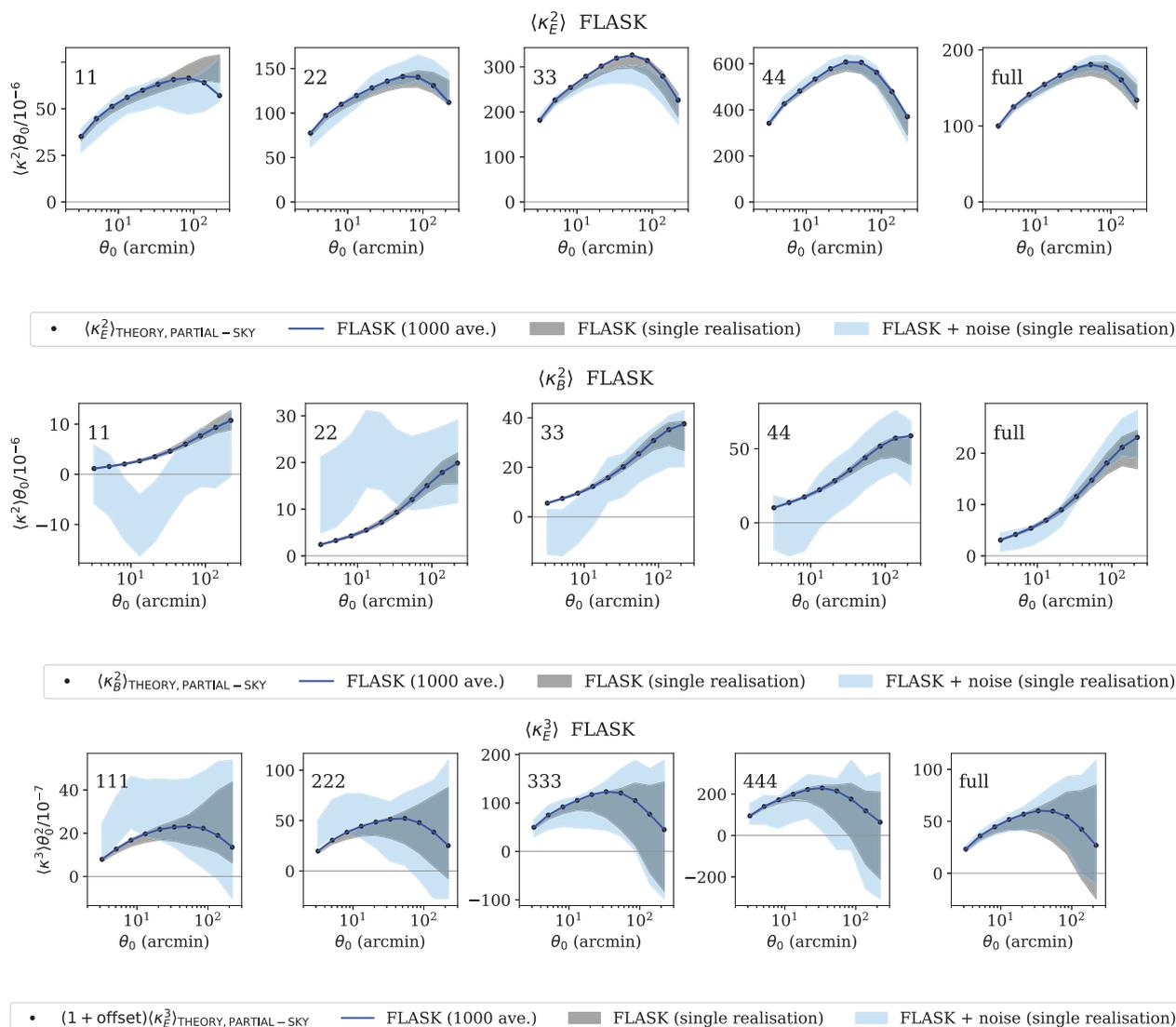
In this section, we present a series of validation tests with simulations to show that our model presented in Section 2.2 does indeed model the second and third moments of the convergence maps. We first validate our model for the effect of masking (i.e. the mode-coupling matrix approach) in Section 4.1, then validate the remaining components of the modelling of the second and third moments in Section 4.2. In Section 4.3, we estimate the potential impact of baryonic feedback at small scales; finally, in Section 4.4, we assess the impact of higher order lensing corrections (such as reduced shear or source crowding) not included in our modelling.

### 4.1 Testing mask effects

We first considered the case of no shape noise. We used 1000 FLASK realizations of the DES Y3 footprint, and measured the convergence field starting from the shear field using the method explained in Section 2.1. This has been done for the four tomographic bins and the non-tomographic sample. We then smoothed the map with a top hat filter at different smoothing scales. We choose as an interval  $\theta_0 \in [3.2, 220]$  arcmin, and we used 10 equally (logarithmic) spaced scales (even though we expect scales close to the pixel size, which is  $\approx 3.4$  arcmin, to not contain much information).

The (smoothed) second moments, both for the E and B modes, are shown in the top and middle panels of Fig. 2 and compared with theoretical predictions. In the figure, we just show automoments (i.e. moments obtained from maps of the same tomographic bin). We also show the average of the 1000 partial-sky FLASK realizations, which agrees to better than 0.5 per cent with the theoretical modelling. Without the mode-coupling matrices, we would have not been able to predict any B modes. Moreover, our theoretical predictions for the E modes would have been biased high, as no leakage of E modes into B modes would have occurred. We note that in Fig. 2 we do not show uncertainties for the average measurements, as they are negligibly small; on the other hand, the amplitude of the measurement uncertainty from a *single* realization is represented by the shaded region.

The third moments are shown in the lower panel of Fig. 2. We just show E modes as B modes are not measured at any statistical significance. We cannot directly compare third moments measured from partial-sky FLASK mocks to masked theory predictions: as explained in Section 3.1, flask simulations are expected to recover the input third moments only at no smoothing; for larger smoothing scales, we expect (and measure) an offset with respect to theoretical predictions such that  $\langle \kappa_{\theta_0}^3 \rangle_{\text{FLASK, full-sky}} \sim \langle \kappa_{\theta_0}^3 \rangle_{\text{theory, full-sky}} [1 + \text{offset}(\theta_0)]$ . To check that we correctly model third moments mask effects in the partial-sky predictions, we need then to verify that the



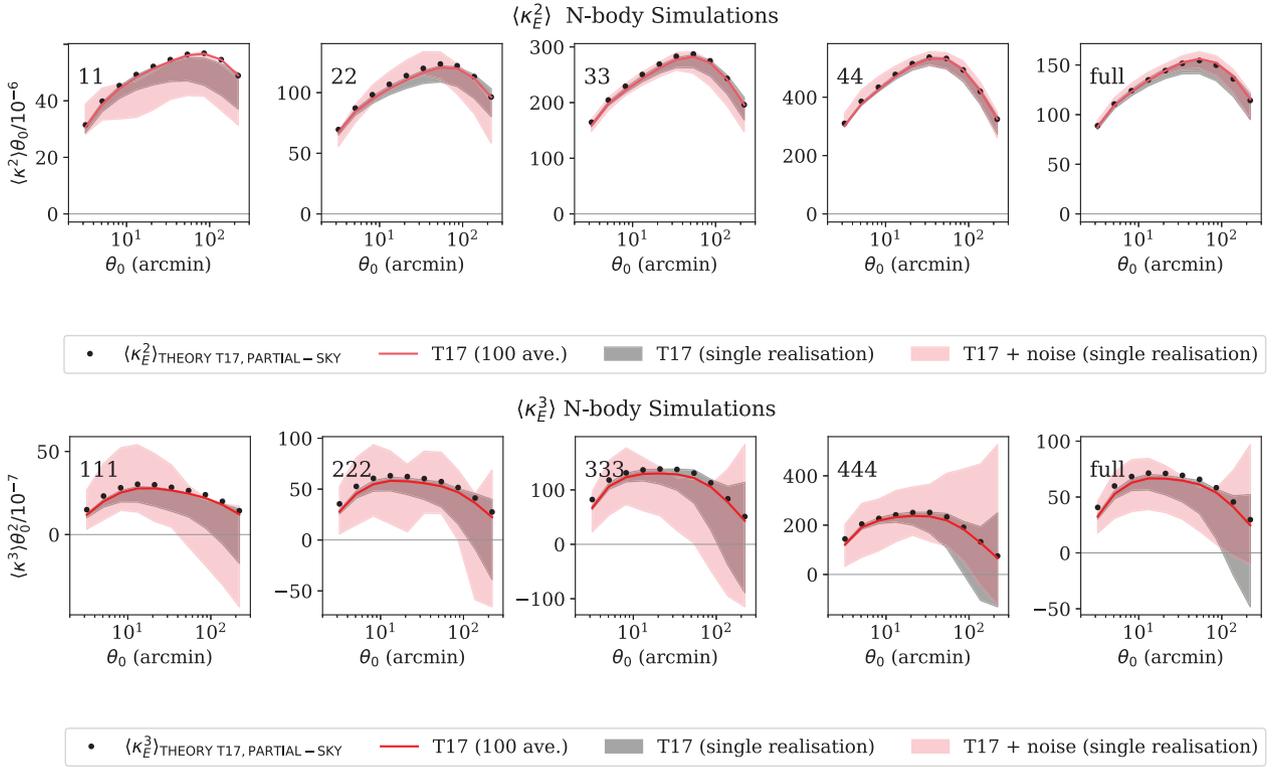
**Figure 2.** Second moments (E and B modes) and third moments (E only) measured in FLASK simulations from partial-sky coverage realizations of the DES Y3 footprint. The convergence maps are obtained from the realizations of the FLASK shear fields configured as explained in Section 3.1. Mask effects are included in the theory modelling (black dots, equations 23 and 24). The ‘offset’ term in the theory predictions for the third moments refers to the function that accounts for the (known) fact that FLASK fails to produce the correct third moments for a given input cosmology (see text for more details). Grey bands represent the measurement from one (taken at random) noiseless FLASK realization, together with its uncertainty (measurements uncertainties are estimated in Section 5.1). Light blue bands also include shape noise. The average of the measurement over 1000 FLASK realizations are shown by the blue lines (error bars are omitted). The numbers 11, 22, 33 etc. in each plot refers to the combination of tomographic bins considered to compute the moments, while ‘full’ refers to the non-tomographic case. Only autocorrelations are shown. *Upper panels:* second moments, E mode of the convergence maps. *Middle panels:* second moments, B mode of the convergence maps. B modes are much smaller than E modes and are due to mask effects. *Lower panels:* third moments, E mode of the convergence maps. Third moments measured in FLASK simulations are not expected to match the input theory perfectly (see text for more details); here, the theoretical predictions for the third moments are replaced by the average measurement of third moments in 1000 full-sky FLASK realizations.

third moments computed from partial-sky FLASK realizations follow  $\langle \kappa_{\theta_0}^3 \rangle_{\text{FLASK, partial-sky}} \sim \langle \kappa_{\theta_0}^3 \rangle_{\text{theory, partial-sky}} [1 + \text{offset}(\theta_0)]$ . This is shown in the lower panel of Fig. 2. The FLASK third moments theory lines include the offset function. These agree with the average of 1000 DES Y3 (partial-sky) FLASK realizations within 3 per cent, which is much smaller than the observational uncertainties. We conclude that our mode-coupling matrices deal efficiently with mask effects also for the third moments.

We next consider a more realistic scenario in which shape noise is included. In this case, we need to perform the de-noising procedure (equations 34 and 35), which subtracts the shape-noise contributions from the measured moments. For the second moments, we first

checked that the mixed terms ( $\langle \kappa_{\text{rand}} \kappa \rangle^{i,j}$  and  $\langle \kappa_{\text{rand}} \kappa \rangle^{j,i}$ ) averaged to zero, while the terms  $\langle \kappa_{\text{rand}}^2 \rangle^{i,i}$  (corresponding to the noise-only second moments) did not and needed to be subtracted. As for the third moments, we found out that mixed terms of the form  $\langle \kappa \kappa_{\text{rand}}^2 \rangle^{i,j,k}$  did not vanish for some choice of indices and needed to be subtracted. This is due to source galaxy density–convergence field correlations that do not vanish at third order. All the other terms, including  $\langle \kappa_{\theta_0, \text{rand}}^3 \rangle^{i,i,i}$ , averaged to zero and did not need to be subtracted.

The de-noised measurements are shown again in Fig. 2 (light blue shaded regions). The measurements are clearly noisier than the previous case, but we verified that when the averages over the 1000



**Figure 3.** Second moments and third moments (E modes) measured in the T17 simulations from partial-sky coverage realizations of the DES Y3 footprint. The convergence maps have been obtained starting from a realization of the DES Y3 shear field. Mask effects are included in the theory modelling (black dots, equations 23 and 24). The label of the theory modelling points specifies ‘theory T17’ to differentiate it from the `flask` theory lines, since the two sets of simulations have a slightly different cosmology. Grey bands represent the measurement from one (taken at random) noiseless T17 realization, together with its uncertainty (measurements uncertainties are estimated in Section 5.1). Red bands also include shape noise. The average of the measurement over 100 T17 realizations are shown by the red lines (error bars are omitted). The numbers 11, 22, 33 etc. in each plot refers to the combination of tomographic bins considered to compute the moments, while ‘full’ refers to the non-tomographic case. Only autocorrelations are shown. *Upper panels:* second moments, E modes of the convergence maps. *Lower panels:* third moments, E modes of the convergence maps.

FLASK realizations are considered, the match with the theory shows the same level of agreement as the noiseless case.

## 4.2 Testing second and third moments modelling

To validate our modelling of the second and third moments we need a full  $N$ -body simulation. In particular, we need to validate the E modes, as they will be used in the cosmological analysis (B modes have a low signal to noise, and they will be mainly used as a diagnostic). To do this, we use 100 realizations of the shear field obtained using the T17 simulations. The comparison with the theory (second and third moments, E modes) is shown in Fig. 3. In the same figure, we also show the average of the 100 realizations of the DES Y3 footprint. For the second moments, the match with the theory is better than 1 percent at large scales (comparable with the uncertainties in the modelling of mask effects) and it is at the level of 2–3 percent at small scales (comparable with the accuracy of the simulations at low redshift). The good match at large scales also justifies the use of the Limber approximation in our modelling.

For the third moments, the theory matches the measurement to better than 10 percent at all scales. The modelling at small scales is obtained including the SC01 analytical refinement based on  $N$ -body, CDM-only simulations. We note that without the SC01 formulae, the predictions of the third moments from perturbation theory only would start departing from the T17 measurement at  $\sim 30$ – $40$  arcmin,

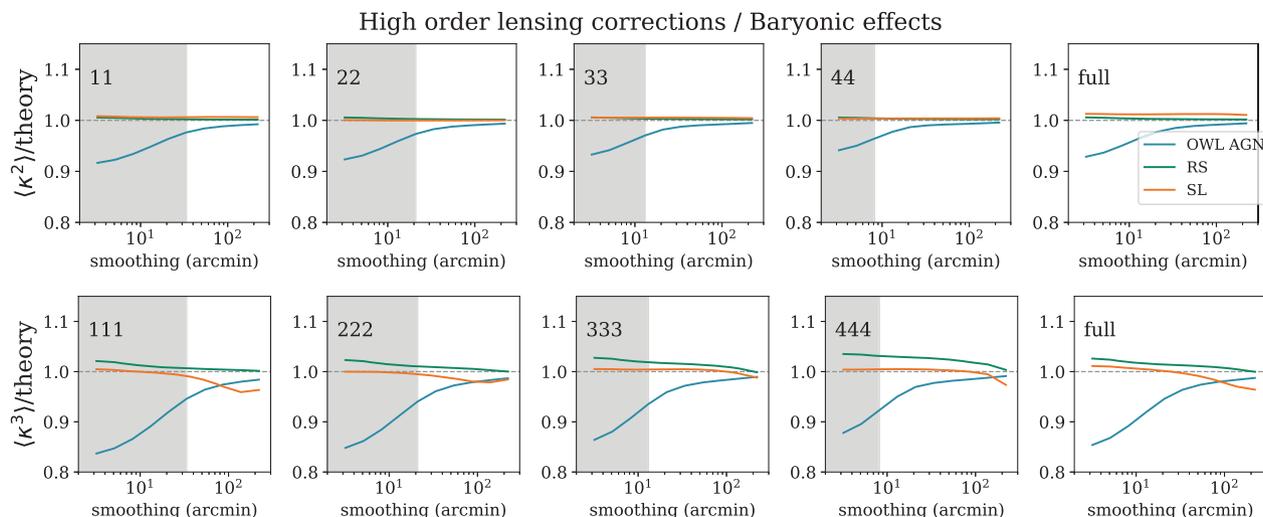
reaching a disagreement of 80 percent at 5 arcmin in the first tomographic bin.

## 4.3 Baryonic effects

We discuss in this and in the next subsection the impact of a number of effects not included in our fiducial modelling. Ultimately, the impact of these effects (together with the comparison with T17 sims from the previous section) will directly determine the scales to be used in the cosmological analysis.

We consider here the possible contamination of our data vector by baryonic feedback effects at small scales. Including baryonic feedback models in the theoretical modelling is an on-going issue in current cosmic shear analyses, due to the uncertainties in current baryonic feedback models. The strategy adopted by DES (in the Y1 and Y3 analyses) is to not model baryonic feedback effects, but to exclude the scales of the data vector possibly contaminated by baryonic feedback.

To this aim we contaminate a data vector by the effects of baryonic feedback as estimated from the OWLS ‘AGN’ simulations (Schaye et al. 2010; van Daalen et al. 2011). We note that the OWLS suite is not the only set of simulations including baryonic effects (see e.g. EAGLE simulation, Hellwing et al. 2016, IllustrisTNG simulations, Springel et al. 2018, Horizon simulations, Chisari et al. 2018). The impact of baryons on the dark matter power spectrum in the OWLS simulations is large compared to other simulations, though more



**Figure 4.** Impact of baryonic effects (from OWLS simulations) and two non-linear lensing corrections to (E modes) moments. The blue line (OWL) refers to the rescaled predicted moments including baryonic contributions from AGN feedback. The orange line (label RS) shows the contribution to reduced shear correction, as measured in simulations. The green line refers to source-lens clustering (label SL), as measured in simulations. The grey shaded regions represent the angular scales cut out from the analysis (see Section 6.1; as the scales cut is determined only for the tomographic version of the data vector, we do not show any shaded region for the non-tomographic case).

extreme models exist. We use the OWLS predictions to contaminate our data vector and use them as an upper limit on the magnitude of baryonic effects.

To contaminate the data vector, we proceed in similar fashion to what was done in the DES Y1 cosmic shear analysis (Troxel et al. 2018). We rescale the power spectrum so as to include contribution from the OWLS ‘AGN’ sub-grid prescriptions

$$P_{\text{NL}}(k, z) \rightarrow \frac{P_{\text{DM+baryons}}}{P_{\text{DM}}} P_{\text{NL}}(k, z), \quad (37)$$

where  $P_{\text{DM}}$  is the OWLS power spectrum due to dark matter, and  $P_{\text{DM+baryons}}$  is the OWLS power spectrum including the ‘AGN’ feedback prescription. Applying such contamination procedure to the power spectrum should account for most of the baryonic effects on the third moments as well (Foreman et al. 2019 show that baryonic contributions to the bispectrum go as  $P_{\text{DM+baryons}}^2/P_{\text{DM}}^2$ , at least for the scales under study here).

The effects of contaminating a theoretical data vector with baryonic feedback are shown in Fig. 4, where we show the ratio between a contaminated, theoretical data vector and one that is not contaminated. The OWLS power spectrum dampens the measured moments at small smoothing scales, whereas the effect is almost negligible at larger scales. This implies that a range of small scales needs to be excluded from the cosmological analysis: including such scales could introduce biases in the cosmological inference. The angular scales cut that safeguards against possible baryonic effects is shown in Fig. 4 (grey-shaded regions) and it is quantitatively determined with a simulated likelihood analysis in Section 6.1.

Last, we note that from Fig. 4 it is not straightforward to compare the smoothing scales at which the OWLS power spectrum starts affecting the moments with the angular scales used in the DES Y1 cosmic shear analysis (Troxel et al. 2018), as the two probes get contributions from the high multipoles in harmonic space differently.

#### 4.4 Higher order lensing corrections

We next verify the impact of a number of higher order lensing corrections to our theoretical modelling (Schneider et al. 1998,

2002; Schmidt et al. 2009; Krause & Hirata 2010). As we have not implemented theoretical modelling of the following effects, we resort to simulations to assess their impact on the data vector. We look at three different effects: reduced shear, source-lens clustering, and magnification bias. The first is due to the fact that we cannot directly observe the shear field, but rather we observe the reduced shear (equation 11). Source-lens clustering is due to the correlation between source galaxies and lensing potentials along the line of sight. The convergence field traces the integrated density contrast up to the position where the sources are detected. Since we estimate the convergence field from an ensemble of sources at different redshifts, and the source galaxies are not uniformly distributed along the line of sight, this affects the estimated convergence values. The effect is enhanced in case of broad redshift distributions. We note that fluctuations in the density field are also caused by magnification effects (magnification bias).

The simulation set-up of the tests shown in Section 4.2 did not include such higher order effects. In order to include the reduced shear contribution, we start from equation (11) and note that in the weak lensing limit  $1/(1 - \kappa) \sim 1 + \kappa$ . It follows that the observed shear has an additional contribution that can be modelled as

$$\gamma_{\text{obs}} \rightarrow \gamma(1 + \kappa). \quad (38)$$

Source-lens clustering and magnification effects can be modelled by accounting for the effect of the density fluctuations along the line of sight when estimating the shear field

$$\gamma_{\text{obs}} \rightarrow \gamma(1 + \delta_{\text{obs}}), \quad (39)$$

where the  $\delta_{\text{obs}} \equiv 1 - N_{\text{obs}}/\langle N \rangle$  is the estimated density contrast ( $N_{\text{obs}}$  is the number of galaxies along the line of sight and  $\langle N \rangle$  is the average number of galaxies). The fluctuations in the density field are due to source galaxies overdensities and lensing magnification effects. Lensing magnification enhances the flux of galaxies and this can locally increase the number density, as more galaxies pass the selection cuts/detection threshold of the sample; at the same time, the same volume of space appears to cover a different solid angle on the sky, causing the observed number density to decrease. At

first order, the impact of source galaxies overdensities and lensing magnification effects can be modelled as

$$\delta_{\text{obs}} = \delta_{\text{gal}} + q\kappa, \quad (40)$$

with  $q$  expected to be of order unity (see Schmidt et al. 2009 for an approximate description of the term  $q$ ). Summing up equations (38) and (39)

$$\gamma_{\text{obs}} = \gamma[1 + \delta_{\text{gal}} + (1 + q)\kappa]. \quad (41)$$

Reduced shear contributes as  $\approx 1 + \kappa$ , magnification effects as  $\approx 1 + q\kappa$ , lens-source clustering as  $\approx 1 + \delta_{\text{gal}}$ . To test the impact of these effects, we used the T17 simulations. Using the full-sky spherical harmonics approach laid out in Section 2.1, we generated for every redshift layer of the simulations: (1) shear field  $\gamma$  distributions starting from the convergence maps  $\kappa$ ; (2) shear field distributions with  $1 + \kappa$  and  $1 + \delta_{\text{gal}}$  contributions (equations 38 and 39); (3) density contrast field distributions  $1 + \delta_{\text{obs}}$ . We then stacked the redshift layers together according to the redshift distributions of the WL tomographic bins, and generated the following maps:

$$\langle \gamma \rangle_{\text{pix}(\theta)} \approx \frac{\int dz n(z) \gamma(z, \theta)}{\int dz n(z)}, \quad (42)$$

$$\langle \gamma \rangle_{\text{pix}(\theta)}^{\text{RS}} \approx \frac{\int dz n(z) (1 + \kappa(z, \theta)) \gamma(z, \theta)}{\int dz n(z)}, \quad (43)$$

$$\langle \gamma \rangle_{\text{pix}(\theta)}^{\text{SL}} \approx \frac{\int dz n(z) (1 + \delta(z, \theta)) \gamma(z, \theta)}{\int dz n(z) (1 + \delta(z, \theta))}. \quad (44)$$

Equations (42), (43) and (44) are, respectively, the shear fields with no non-linear lensing corrections, with reduced-shear contributions, and source-lens clustering. As for the latter, we divided by the integrated density field to mimic the map making process, where each pixel contains the average of the shear field along the line of sight.

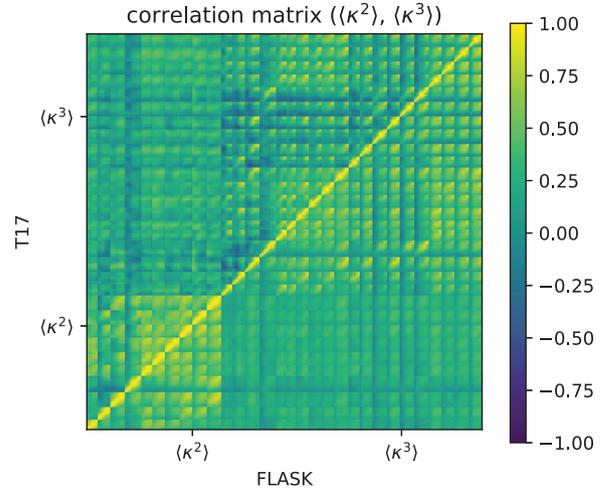
The impact of such corrections on E modes is shown in Fig. 4. We estimated the moments from a full-sky, noise-free realization of the simulation. For the reduced shear and source-lens clustering, we considered as a ‘theory’ the moments estimated from the same realization of the simulations using equation (42) to estimate the shear field. We do not show error bars for the moments measurement as we expect them to be much smaller than DES Y3 uncertainties.<sup>2</sup> We also do not show magnification effects as they are of the same order as the reduced-shear correction (assuming  $q$  of the order of unity). We find that these non-linear lensing corrections are much smaller than DES Y3 uncertainties and sub-dominant with respect to baryonic effects. We checked that the small bias due to source-lens clustering at very large scales of the third moments does not affect the cosmological analysis (at very large scales, the signal to noise for the third moments is much smaller than 1, thus, a 5 per cent bias on the signal does not bias the cosmological constraints).

## 5 COVARIANCE AND LIKELIHOOD

### 5.1 Covariance estimation

To correctly infer cosmological parameters from our data, we need an accurate estimate of the measurement uncertainty. We estimate

<sup>2</sup>First, since we are considering the full sky, we expect the covariance of the moments measurement to be roughly  $\approx 8$  times smaller. Secondly, as we are using the moments of the same realization with no non-linear lensing corrections as the ‘theory’, we can expect the measurements to be highly correlated, and the uncertainties in their ratio should be very small as cosmic variance would cancel.



**Figure 5.** Measured correlation matrix of second and third moments from 1000 FLASK simulations (lower right triangle) and from 100 T17 simulations (upper left triangle). A  $24 h^{-1}$  Mpc scale cut has been applied.

the covariance from 1000 independent realizations of the FLASK simulation. For each FLASK realization, we measure the second and third moments of the smoothed convergence field as explained in Section 4.1. We then build our covariance matrix as

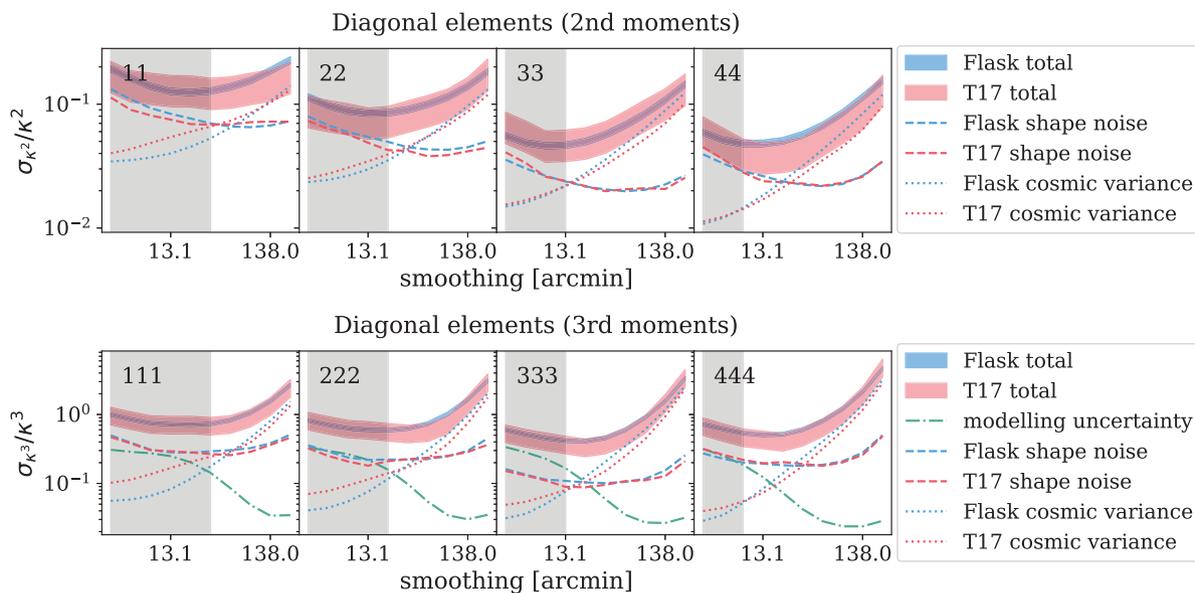
$$\hat{C} = \frac{1}{\nu} \sum_{i=1}^{N_s} (\hat{d}_i - \hat{d})(\hat{d}_i - \hat{d})^T, \quad (45)$$

where  $\nu = N_s - 1$  with  $N_s$  the number of realizations,  $\hat{d}_i$  the data vector measured in the  $i$ -th simulation and  $\hat{d}$  the sample mean. The data vector is made of a combination of second and third moments as measured at different smoothing scales.

Within single realizations, variations in the measured moments among different simulations are mostly due to two different contributions: (1) a combination of galaxy intrinsic shape and measurement noise, or ‘shape noise’, and (2) the cosmic density field inside the DES Y3 footprint is a random realization of the underlying cosmology, or ‘cosmic variance’. For third moments, we also include in our covariance a ‘modelling uncertainty’ related to the analytical fitting formulae describing the third moments at small scales. To this aim, we add to the diagonal part of the covariance the difference between SC01 and GM12 squared (this approach is similar to the one adopted by Simon et al. 2015, who included a 20 per cent r.m.s in the covariance to take into account small-scale modelling uncertainties in the bispectrum). We note that for the scale cuts used in this analysis, the contribution of such modelling uncertainty to the error budget is small (see Fig. 6 and Appendix A2).

The measured correlation matrix is shown in Fig. 5. The matrix was obtained using equation (45) and selecting the data vector elements passing a  $24 h^{-1}$  Mpc scale cut (we selected the scales  $\theta_0$  such that  $\theta_0 > R_0/\chi(z)$ , where  $\chi(z)$  is the comoving distance of the mean redshift of a given tomographic bin and  $R_0 = 24 h^{-1}$  Mpc). Values at different smoothing scales for the same moment are highly correlated. Fig. 5 also shows that second and third moments are not very correlated. This is mostly due to shape noise and third moment modelling uncertainties at small scales that wash out existing correlations.

The values of the diagonal elements of the covariance matrix, relative to values of their data vector entries, are shown in Fig. 6, for both FLASK and T17 simulations. We also show the errors due to the finite number of simulation realizations. One can see that for both second and third moments the intermediate scales are the ones



**Figure 6.** Diagonal elements of the covariance of the second and third moments estimated from FLASK simulations. The  $x$ -axis shows the corresponding data vector entries. We show separate contributions due to shape noise and cosmic variance. For the third moments, we also show the modelling uncertainties related to the small scales analytical fitting formulae. For comparison, we also show the covariance estimated from the T17 simulations. For the total covariance, we show uncertainty due to the finite number of simulation realizations. The shaded regions represent the scales excluded in the cosmological analysis (a  $24 h^{-1}$  Mpc scale cut has been applied).

with better signal to noise, and that in general second moments have a much better signal to noise than third moments.

The sample variance part of the covariance is cosmology dependent and dominates at large scales. We do not expect this cosmological dependence to significantly impact the recovery of cosmological parameters (see discussion in Section 6.1). We also note here that the lognormal approximation assumed by FLASK needs to be checked for the sample variance part of the covariance for third moments. However, the scales dominated by sample variance have a smaller signal to noise for the third moments; moreover, despite FLASK limitations, Fig. 6 shows that the FLASK and T17 covariances agree within uncertainties. In Appendix C, we provide further evidence that the uncertainties in the modelling of the third moments covariance have little effect on the cosmological inference.

We therefore decided to rely on FLASK simulations to build our fiducial covariance because the cosmological parameters can be easily changed and we can produce a large number of simulations. The T17 simulations have a fixed cosmology, and, above all, are limited in numbers, causing the inverse of the covariance matrix to be extremely noisy (and biased, see e.g. Hartlap, Simon & Schneider 2007). However, in the next section we show an implementation of a data-compression algorithm that greatly reduces the size of the data vector (and the noise in the covariance due to the paucity of simulations). The data compression algorithm is implemented in our fiducial analysis and in principle allows us to run our cosmological pipeline also using the T17 covariance (although it will still be noisier than the FLASK covariance). While we still use FLASK as our fiducial covariance, we show in Appendix C that the differences in the recovered cosmological parameters between using the T17 covariance (in combination with the data compression algorithm) or FLASK covariance are small.

## 5.2 Data compression

To reduce the noise in our covariance matrix estimated from FLASK mocks, we implement the MOPED data-compression algorithm

(Tegmark, Taylor & Heavens 1997; Heavens, Jimenez & Lahav 2000; Galdi et al. 2018). We follow Heavens et al. (2000) and include a data-compression scheme based on the Karhunen–Loève algorithm. The algorithm works by assigning weights to each element of the data vector that are proportional to the sensitivity of the element to the variation of a given model parameter. In such a way, it is possible to reduce the dimension of the data vector to the number of model parameters considered. The compressed data vector can be written as

$$d_i^{\text{compr}} = \langle d \rangle_i^T \hat{C}^{-1} d \equiv b_i d, \quad (46)$$

where  $d$  is the full-length data vector,  $\hat{C}$  is the measurement covariance and  $d_i^{\text{compr}}$  is the  $i$ -th element of the compressed data vector. The index  $i$  refers to the  $i$ -th model parameter  $p$  considered, and  $\langle d \rangle_i^T$  is the derivative of the model data vector with respect to that parameter.

The above equation assumes that the dependence of the covariance on cosmological parameters is mild ( $\partial \ln C / \partial \ln p_i \ll 1$ ). While being reasonable, we do not explicitly check the latter assumption as it would require producing many covariance matrices, which is computationally expensive. We also note that for the compression algorithm to be lossless, the likelihood of the non-compressed data vector must be Gaussian. We check this in Appendix C, and we show that the uncompressed data vector shows only mild deviations from Gaussianity. We note, however, that we expect the compressed data vector to have a more Gaussian distribution, due to the central limit theorem (Heavens et al. 2017). We show this in Section 6.1.

In general, if one or more assumptions underlying the data-compression algorithm are violated, we can expect the compression to be not optimal. In this case, the credible regions would be larger than they could be (Heavens et al. 2017; Alsing, Wandelt & Feeney 2018), but the parameter inference would still be valid.

To implement the algorithm described in equation (46), we use the FLASK covariance, and we estimate the derivative of the data vector using a five-point stencil derivative centred on the true value of the simulation parameters. As model parameters we use the five cosmological parameters and all the nuisance parameters described in

Section 2.2.2. The compressed covariance can be easily obtained as

$$\hat{C}_{ij}^{\text{compr}} = b_i^T \hat{C} b_j. \quad (47)$$

We defer the validation of the compression algorithm to Appendix C, where we compare the posterior distributions obtained with and without the data-compression algorithm. In general, we find smaller contours for the chains run with the compressed data vector, as expected by the lower noise in the covariance (we explain how we deal with the noise in the covariance in the next section).

### 5.3 Data vector and likelihood

The final data vector includes all the ‘auto’ moments of different tomographic bins (e.g. [1, 1], [1, 1, 1], [2, 2], and [2, 2, 2]) and the ‘cross’ moments (e.g. [1, 2], [1, 1, 2], and [1, 2, 2]), for a total of 10 combinations for second moments and 20 combinations for third moments. The full data vector is shown in Appendix D. The scale cuts are discussed in the next section.

We evaluate the posterior of the parameters conditional on the data by assuming a Gaussian likelihood for the data, i.e.

$$-2 \ln L = f_2 f_1 [\hat{d} - M(p)] \hat{C}^{-1} [\hat{d} - M(p)]^T \quad (48)$$

(see Section 6.1 for an investigation of this assumption). Here,  $M(p)$  is our theoretical model,  $\hat{d}$  is the data vector, and  $\hat{C}^{-1}$  is the inverse of our covariance estimate. The posterior is then the product of the likelihood and the priors. Note that the quantities  $M(p)$ ,  $\hat{d}$ , and  $\hat{C}^{-1}$  in equation (48) are to be considered compressed quantities, and we have dropped the superscript ‘compr’ for brevity. The terms  $f_1$  and  $f_2$  account for noise introduced when the covariance matrix is estimated from random realizations of the data. Even if a covariance estimate  $\hat{C}$  from  $N_{\text{sims}}$  random realizations is an unbiased estimate of the true covariance of the data, its inverse  $\hat{C}^{-1}$  is only a biased estimate of the true precision matrix  $C^{-1}$  (Hartlap et al. 2007). This bias can be corrected with the multiplicative factor

$$f_1 = \frac{N_{\text{sims}} - N_{\text{data}} - 2}{N_{\text{sims}} - 1}, \quad (49)$$

where in our case the number of independent realizations used to estimate the covariance is  $N_{\text{sims}} = 1000$ , and  $N_{\text{data}}$  is the length of the data vector. Note that this is just an approximate treatment of the noise in the covariance matrix, since the data likelihood depends on the precision matrix in a non-linear way. Sellentin & Heavens (2016) have devised a more accurate treatment, taking into account the impact of the covariance estimation noise on the entire likelihood. We investigate their alternative likelihood in Appendix C and find that after our data compression it has a negligible effect.

There is a second – and often more severe – problem in estimating the likelihood of data from a finite number of random realizations that is not solved by the likelihood of Sellentin & Heavens (2016). This problem is that the noise in a covariance estimate does not just change the width of parameter contours but also their location (Dodelson & Schneider 2013, see also fig. 1 in Friedrich & Eifler 2018 for a simple demonstration of the effect). An approximate way to take this into account is to multiply our loglikelihood by

$$f_2 = \left[ 1 + \frac{(N_{\text{data}} - N_{\text{par}})(N_{\text{sims}} - N_{\text{data}} - 2)}{(N_{\text{sims}} - N_{\text{data}} - 1)(N_{\text{sims}} - N_{\text{data}} - 4)} \right]^{-1}. \quad (50)$$

This correction (dubbed Dodelson–Schneider-factor by Friedrich & Eifler 2018) assumes the model to be linear in all the parameters and widens the contours to encompass the additional noise in the parameter estimates (Dodelson & Schneider 2013). We note that as

the data-compression greatly reduces the length of the data vectors,  $f_1$  and  $f_2$  become close to 1.

To sample the posteriors of our parameters, we generate Monte Carlo Markov chain (MCMC) samples that map out the posterior space leading to parameter constraints. To this aim, we use the public software package EMCEE (Foreman-Mackey et al. 2013), which is an affine-invariant ensemble sampler for MCMC. To test the convergence of our MCMC chains, we adopted the Gelman & Rubin (1992) test.

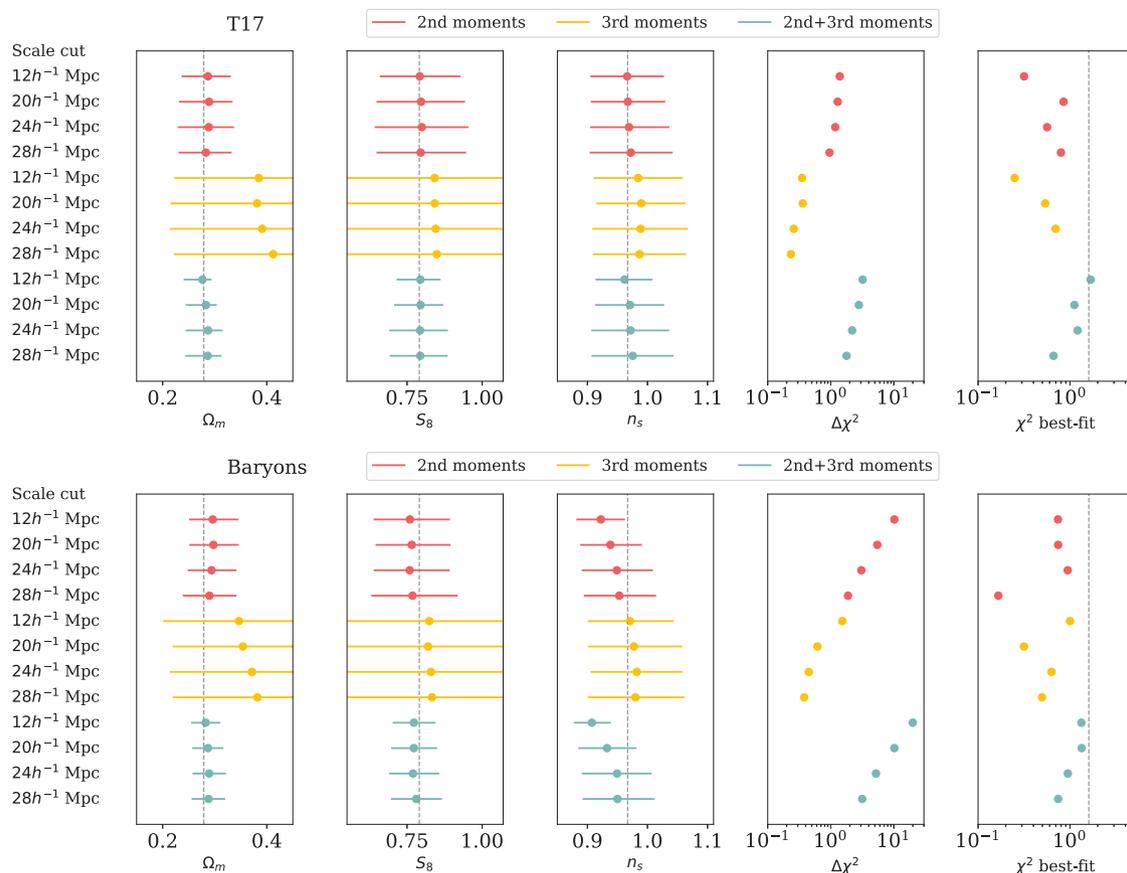
For the cosmological parameters, we assume a flat  $\Lambda$ CDM cosmology and vary five parameters:  $\Omega_m$ ,  $\sigma_8$ ,  $\Omega_b$  (the baryonic density in units of the critical density),  $n_s$  (the spectral index of primordial density fluctuations), and  $h$  (the dimensionless Hubble parameter). We assume wide flat priors on  $\Omega_m$  and  $\sigma_8$  and adopt the informative priors in  $h$ ,  $n_s$ , and  $\Omega_b$  that that were used in the DES Y1 two-point function analysis (see Table 1). When constraining cosmological parameters, we marginalize over nuisance parameters describing mean photo- $z$  uncertainties, multiplicative shear biases and IA effects in our measurements. The modelling of our nuisance parameters is described in Section 2.2.2. As at the time of finishing this work, the DES Y3 priors were not finalized yet, so we again assume DES Y1 priors for all the nuisance parameters (priors are summarized in Table 1). Photo- $z$  uncertainties are parametrized by a shift in the mean of the distribution (one for each tomographic bin). Priors for the shifts come from redshift distributions of a matched sample of galaxies in the COSMOS survey and angular cross-correlation with *redMaGiC* galaxies (Hoyle et al. 2018). Multiplicative shear bias priors are described in Zuntz et al. (2018). We also assume wide flat priors for intrinsic alignment.

Due to the fact that the theory predictions described in Section 2.2 can be quite time consuming to compute due to the large number of cross-correlations and integrations involved, we further implemented an emulator (Heitmann et al. 2006; Habib et al. 2007) to speed up the calculations. In our implementation, the emulator provides fast theoretical predictions by interpolating over a number of predictions computed at some training points spanning the parameter space of interest. In particular, the training of our emulator requires to interpolate over 500 training points, and it provides fast theoretical predictions as a function of five different (cosmological) parameters. The speedup stems from the fact that the time-consuming part of the calculation is substituted by an interpolation over few training points. The speedup achieved by using the emulator is of two orders of magnitudes, with a negligible impact on the accuracy of the theoretical predictions. We note that emulators are often implemented when  $N$ -body simulations are used to model the data vector (e.g. Knabenhans et al. 2019) due to the impracticability of producing a simulation for every point of the parameter space. In our case, even if we do not directly use any  $N$ -body simulation to model the data vector, we take advantage of the computational speed up provided by the emulation of our own theoretical model. More details are provided in Appendix E.

## 6 COSMOLOGICAL CONSTRAINTS FROM MOMENTS OF THE CONVERGENCE FIELD

### 6.1 Fiducial scale cuts

The last analysis choice to make before presenting the final cosmological constraints of the second and third moments of the convergence field is which scales are to be used for the analysis. The scale cuts we use are determined based on two tests. First, we check that our theoretical modelling is adequate to describe the data vectors



**Figure 7.** The  $1\sigma$  marginalized constraints on cosmological parameters for a number of different scale cuts. In the upper plot, the average of 100 **T17** simulations has been used as the data vector. In the lower plot, the constraints are obtained by using a theory data vector contaminated with the OWLS AGN power spectrum. Data points represent the mean of the 1D marginalized posterior, while for the confidence interval we show the two-tail symmetric intervals. The vertical dashed lines in the first three columns represent the input values of cosmological parameters. The column  $\Delta\chi^2$  represents the  $\chi^2$  of the data vector contaminated with baryonic effects or from the average of **T17** sims with respect to a theory data vector. The  $\chi^2$  best-fitting column represents the  $\chi^2$  of the best-fitting cosmology from the MCMC chain. The vertical line in the last column marks the  $\chi^2 = 1.6$  boundary.

as obtained from the average of many  $N$ -body simulations from **T17**. Secondly, we check that the impact of baryons on our data vector is not significant. We recall that we adopt a strategy to mitigate baryonic effects which aims at excluding the scales potentially affected by baryonic feedback, without trying to model such effects. For both tests, we run MCMC chains for different combinations of scale cuts. For the former test, we use a data vector from **T17** simulations; for the latter, we use a baryons-contaminated data vector (obtained using the outputs of the OWLS simulation, as explained in Section 4.3), which should represent a reasonable upper limit to the magnitude of baryonic feedback effects in real data. We vary the scales under study and we require the resulting constraints on cosmological parameters not to be biased against the truth, to check our modelling is adequate for the range of scales considered.

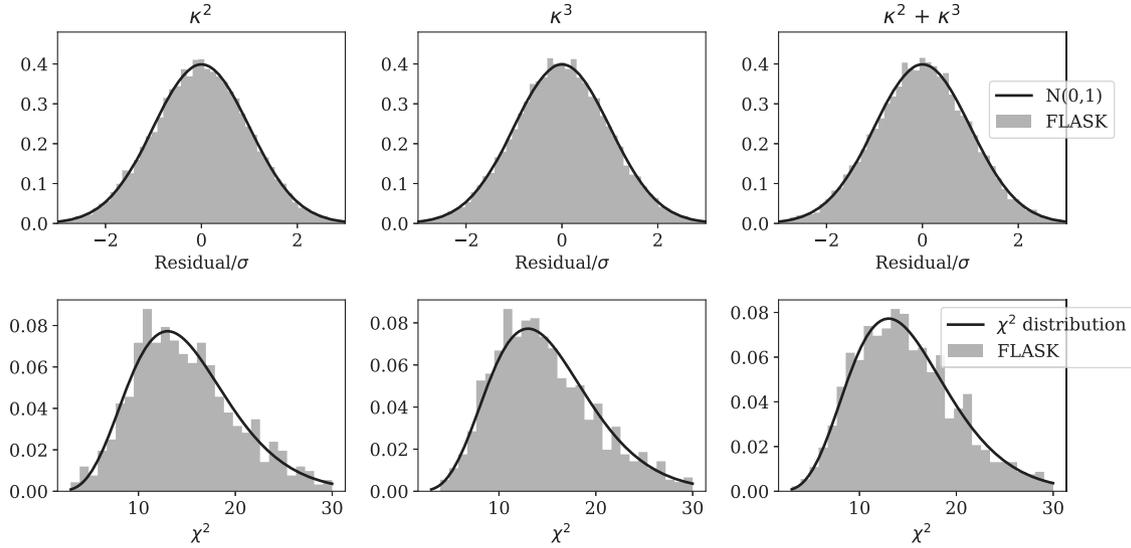
For a combination of scales to be acceptable, we require the mean of the marginalized 1D posterior of  $\Omega_m$ ,  $S_8 = \sigma_8(\Omega_m/0.3)^{0.5}$  to be within  $0.3\sigma$  of the values obtained with a ‘theory’ data vector. As we partially constrain  $n_s$ , we also require the posterior of  $n_s$  to be within  $0.5\sigma$  of the baseline value.

We also adopt a second criterion on the  $\chi^2$  of the best-fitting cosmology. When analysing the data, the best-fitting  $\chi^2$  is used for hypothesis testing and as a proxy of the adequacy of the data vector modelling. A bad best-fitting  $\chi^2$  implies that either our covariance or the parametrization of the data vector is not adequate to describe the measurement. Since we do not model baryonic effects or the small

discrepancies between our theoretical predictions and the data vector from **T17** simulations, we should expect the best-fitting  $\chi^2$  from the data to be biased. By adopting a criteria on the  $\chi^2$  of the best-fitting cosmology of the contaminated data vector we make sure the biases from these two effects are small. In particular, we require the  $\chi^2$  of the best-fitting cosmology obtained from a contaminated data vector to be within 0.3 of the expected spread of the  $\chi^2$  distribution. Therefore, since the length of the compressed data vector is 15, we require the best-fitting  $\chi^2 < 1.6$ . Ideally, for negligible contamination we expect a best-fitting  $\chi^2 = 0$ , as we are using a theory data vector as a baseline (whereas using a noisy data vector would give, on average,  $\chi^2 \sim \text{d.o.f.}$ ). We note that in what follows we never consider the *reduced*  $\chi^2$ , but only the  $\chi^2$  statistics; the reported  $\chi^2$  might seem small due to the small number of d.o.f (due to data compression) and due to the lack of measurement noise in the input data vectors.

In this section, scale cuts are expressed in terms of a specific comoving scale  $R_0$ ; the relation with the smoothing scale  $\theta_0$  is given by  $\theta_0 = R_0/\chi(z)$ , where  $\chi(z)$  is the comoving distance of the mean redshift of a given tomographic bin. In the case of moments from different tomographic bins, we took the average of the  $\langle z \rangle$  of the bins.

The tests run on the data vectors obtained from the average of the **T17** simulations are shown in the upper plot of Fig. 7. As we estimated the bias in the cosmological parameters induced by the emulator in Appendix E, we re-scaled the measured data vector by the ratio between an emulated theory data vector and a non-



**Figure 8.** *Top panels:* residuals (i.e. the difference between the measurement signal in a FLASK simulation and the simulations mean value) of individual data points in units of their expected standard deviation for a compressed data vector. We compare to a Gaussian with zero mean and unit standard deviation. *Bottom panels:* Distribution of the  $\chi^2$  of each realization of the FLASK simulations, compared to a theoretical  $\chi^2$  distribution.

approximated one predicted at the T17 cosmology. This assumes the emulator uncertainties propagate linearly to the data vector; this is justified as at the T17 cosmology the emulator accuracy is below the per cent level.

There are different known reasons why the data vector from the average of T17 might differ from our theoretical predictions: inaccuracies of the simulations or in the modelling of the third moments at small scales (Sections 3 and 4.2), inaccuracies in accounting for mask effects (Section 4.1), inaccuracies in the covariance modelling (Section 5.1), etc. In the past sections, we showed (or discussed) these differences to be small at the level of the data vector, but here we want to assess the impact on the cosmological parameters posteriors.

Fig. 7 shows the marginalized 1D posterior for three out of five cosmological parameters under study. We do not show constraints for  $\Omega_b$  and  $h_{100}$  because the posteriors are heavily prior dominated. For each parameter, we show the mean of the posterior and the symmetric  $1\sigma$  confidence interval. We note that  $n_s$  is mildly constrained and its posterior is partially dominated by the prior (which is assumed to be flat with  $n_s \in [0.87, 1.07]$ ; see Table 1). The constraints from second moments and from the combination of second and third moments are close to the input cosmology, and pass our  $0.3\sigma$  criteria at all scales. We note that the values of  $\Omega_m$  from the third moments are biased. This is due to the fact that the posterior is strongly asymmetric. We checked that the posterior of a theory data vector shows the same level of shifts in the mean value of  $\Omega_m$  for the third moments, and the difference with respect to the results from the T17 data vector are much smaller than  $0.3\sigma$ .

In Fig. 7, we show both the difference  $\Delta\chi^2$  of the T17 data vector and the theory data vector, and the  $\chi^2$  of the best-fitting cosmology. The former quantity gives a rough idea of the discrepancy of the data vector with respect to the truth: a variation of  $\Delta\chi^2 = 1$  could, in the worst case possible, cause a  $1\sigma$  shift in the marginalized 1D posterior of one of the parameters probed. Usually the difference is absorbed and shared across all the parameters probed (and this is the case). The values of best-fitting  $\chi^2$  for the T17 data vectors also pass our  $0.3\sigma$  criteria, being always  $\chi^2 < 1.6$ .

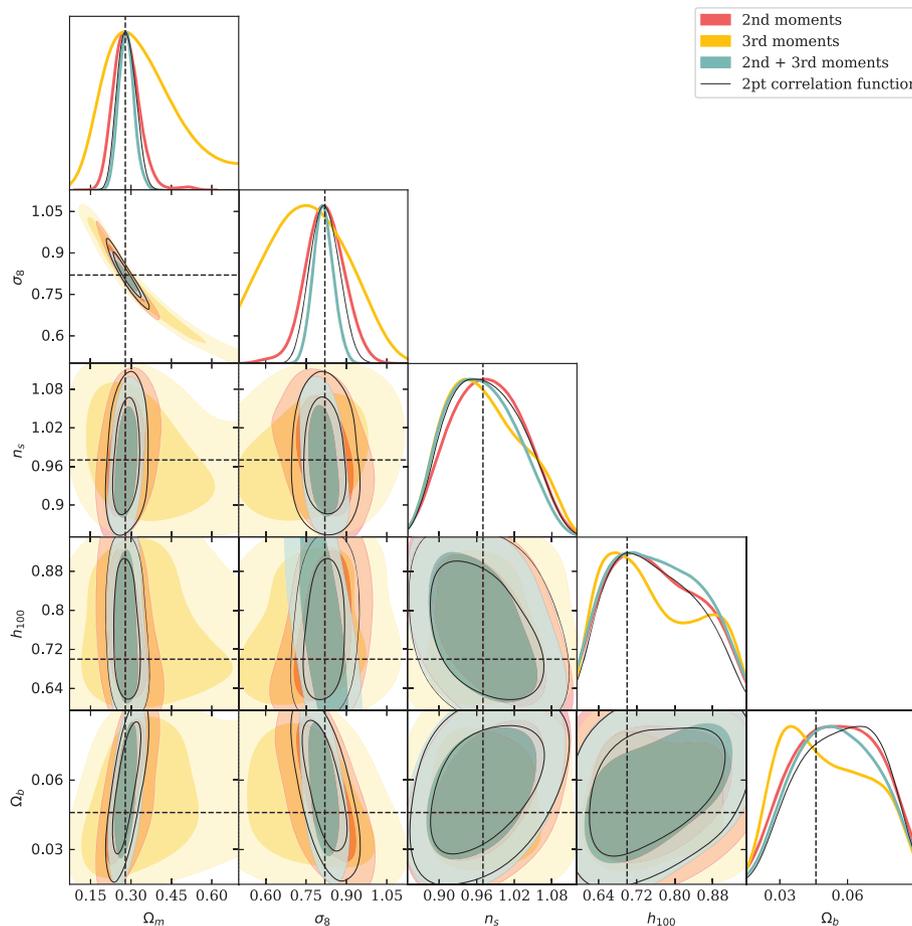
We next test the impact of baryonic effects, by contaminating a theory data vector with the effects from the OWLS AGN simulation,

as described in Section 4.3. The results are shown in the lower panel of Fig. 7. The impact on the data vector from baryons is more pronounced than from the T17 data vector, as shown by the larger  $\Delta\chi^2$  values, and it is more important at small scales. This translates in a bias in  $n_s$  at small scales. Second moments pass our scale cuts criteria starting from  $20 h^{-1}$  Mpc, while the combination of second and third moments from  $24 h^{-1}$  Mpc. As for the third moments, they pass our criteria at all the scales probed here (similarly to the T17 data vector test, the values of the mean of the  $\Omega_m$  posteriors show a negligible shift with respect to the values obtained using a theory data vector). At all scales and for the combinations of second and third moments, our criteria on the best-fitting  $\chi^2$  is passed.

We note that we performed these tests adopting a FLASK covariance, which has a slightly different cosmology with respect to the T17 data vector. This, however, did not significantly bias our posteriors, as shown in the upper panel of Fig. 7.

Based on these tests, we adopt the following fiducial scale cuts:  $20 h^{-1}$  Mpc as a minimum smoothing scale for second moments,  $12 h^{-1}$  Mpc for third moments, and  $24 h^{-1}$  Mpc when second and third moments are combined. We note that the scale  $24 h^{-1}$  Mpc translates into a cut at  $\approx 33$  (8) arcmin for the first (fourth) tomographic bin, while  $12 h^{-1}$  Mpc translates into a cut at  $\approx 16$  (4) arcmin for the first (fourth) tomographic bin. As there is no significant information below twice the pixel size (i.e.  $< 7$  arcmin) and most of the constraining power comes from the two high redshift tomographic bins, we have not considered scales smaller than  $12 h^{-1}$  Mpc in the above tests.

With the final scale cuts determined, we perform extra checks on the covariance and data vector. We checked that the mean  $\chi^2$  of the 1000 FLASK realizations agreed within errors with the number of degree of freedom of our data vector. The distributions of the measured  $\chi^2$  are shown in the bottom panels of Fig. 8. We also verified that the distribution of the residuals (i.e. the difference between the measurement signal in a FLASK simulation and the simulations mean value) for each entry of our data vector followed a Gaussian distribution. This is shown in the top panels of Fig. 8. We note that the data-compression algorithm surely helps in giving



**Figure 9.** Cosmological parameter posteriors obtained from the simulated likelihood analysis. We marginalize over nuisance parameters as explained in Section 5.3. We show constraints from second moments, third moments, and second and third moments combined, along with constraints from a shear two-point correlation function analysis.

the compressed data a more Gaussian distribution, due to the central limit theorem (Heavens et al. 2017).

## 6.2 Simulated likelihood analysis

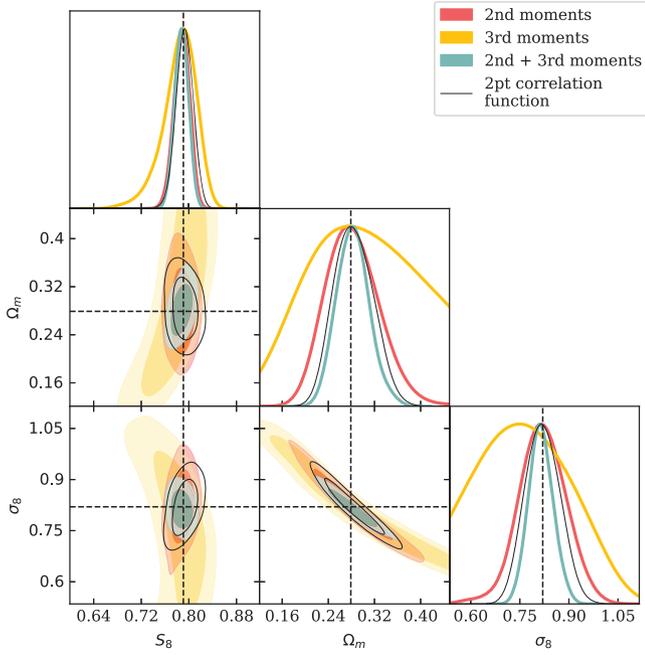
We simulate a DES Y3 likelihood analysis and show the expected constraints for DES Y3 in Fig. 9, for five cosmological parameters. All the tests shown in this section use a theory data vector that includes all the ‘auto’ moments of different tomographic bins (e.g. [1,1], [1,1,1], [2,2], [2,2,2]) and the ‘cross’ moments (e.g. [1,2], [1,1,2], [1,2,2]), for a total of 10 combinations for second moments and 20 combinations for third moments. The fiducial scale cuts determined in the previous section have been adopted ( $20 h^{-1}$  Mpc for second moments,  $12 h^{-1}$  Mpc for third moments, and  $24 h^{-1}$  Mpc when second and third moments are combined). For the data vector, a fiducial T17 cosmology is assumed, with nuisance and astrophysical parameters (photo- $z$  biases, multiplicative shear biases, intrinsic alignment IA) assumed to be null, and no baryonic contamination. We adopted the FLASK covariance described in Section 5.1, and compressed our data vector following Section 5.2. When estimating parameters posterior, we further marginalize over nuisance parameters as explained in Section 2.2.

As we commented in the previous section, second and third moments mostly constrain  $\Omega_m$  and  $\sigma_8$ , while  $n_s$  is partially affected by the prior and  $h_{100}$  and  $\Omega_b$  are prior dominated. In general,

third moments are less constraining than second moments; however, they contain additional non-Gaussian information and they have a slightly different degeneration axis in the  $\Omega_m$ – $\sigma_8$  plane compared to second moments. This helps breaking the degeneracy when the two are combined, delivering tighter constraints. This is also shown in Fig. 10, where we show results in the  $\Omega_m$ – $S_8$  plane.

We report in Table 2 the constraining power of moments for  $\Omega_m$ ,  $S_8$ , and  $n_s$ ; the level of improvement when the moments are combined (quantified as the ratio between predicted confidence intervals) is reported in Table 3. Secondly, third moments and their combination constrain  $\Omega_m$  to 17 per cent, 66 per cent, and 10 per cent, respectively, and  $S_8$  to 1.8 per cent, 3.6 per cent, and 1.5 per cent, respectively. These particular values are obtained specifically for DES Y3 and depend on the particular scales and the noise properties of the sample considered.

We also show in Table 3 how much we expect to improve our constraints when moving to the final DES release, which will include all the data from the 5 yr (Y5) of observations. This is quantified in terms of the ratio between predicted confidence intervals. The values have been obtained by assuming the expected DES Y5 number density (which should roughly double DES Y3 one) and the same DES Y3 scale cuts and tomographic binning. We did not take into account the possibility of having more than four tomographic bins, which would be possible having a deeper sample. In general, we



**Figure 10.** Same as Fig. 9, but now a zoomed in version of the  $S_8$ - $\sigma_8$ - $\Omega_m$  plane.

**Table 2.** DES Y3/Y5 simulated likelihood analysis comparison. Fractional accuracy ( $1\sigma$  marginalized posterior confidence intervals over input value) for  $\Omega_m$ ,  $S_8$ , and  $n_s$ . DES Y5 constraints are obtained with the expected DES Y5 number density and DES Y3 scale cuts and tomographic binning.

	$\Omega_m$	$S_8$	$n_s$
2 moments (Y3)	17 per cent	1.8 per cent	6.9 per cent
3 moments (Y3)	66 per cent	3.6 per cent	7.9 per cent
2 + 3 moments (Y3)	10 per cent	1.5 per cent	6.5 per cent
2pt function (Y3)	12 per cent	1.8 per cent	6.4 per cent
2 moments (Y5)	14 per cent	1.5 per cent	6.0 per cent
3 moments (Y5)	48 per cent	2.8 per cent	7.9 per cent
2 + 3 moments (Y5)	8 per cent	1.4 per cent	5.7 per cent

**Table 3.** DES Y3/Y5 simulated likelihood analysis comparison. We show the ratio between the  $1\sigma$  marginalized posterior confidence intervals for  $\Omega_m$ ,  $S_8$ , and  $n_s$  for a number of cases (a value of  $<1$  indicates improvement). DES Y5 constraints are obtained with the expected DES Y5 number density and the same DES Y3 scale cuts and tomographic binning.

	$\Omega_m$ (ratio)	$S_8$ (ratio)	$n_s$ (ratio)
2 (Y3) $\rightarrow$ 2 + 3 (Y3)	0.57	0.79	0.97
2 (Y5) $\rightarrow$ 2 + 3 (Y5)	0.58	0.89	1.01
2 (Y3) $\rightarrow$ 2 (Y5)	0.85	0.86	0.83
3 (Y3) $\rightarrow$ 3 (Y5)	0.78	0.80	0.99
2 + 3 (Y3) $\rightarrow$ 2 + 3 (Y5)	0.87	0.97	0.89

can expect to further improve our constraints by 10–20 per cent with respect to DES Y3.

We overlay in Figs 9 and 10, the expected posteriors from the DES Y3 shear two-point correlation function analysis. Such analysis includes the same combination of tomographic bins considered in the analysis of second moments ([1,1], [2,2], [1,2], ..., for a total of 10 combinations). Scale cuts for the two-point correlation function

analysis have been chosen by contaminating a shear two-point data vector with the effect of baryons and looking at the bias in the parameters' posteriors, in a fashion similar to what has been done in Section 6.1. The minimum angular scale depends on the particular pair of tomographic bins and whether we are considering  $\xi_+$  or  $\xi_-$ . In particular, the minimum scales considered for  $\xi_+$  are 10, 10, 10, 10, 10, 8, 8, 6, 4, and 4 arcmin, whereas the minimum scales for  $\xi_-$  are 170, 170, 120, 120, 120, 90, 90, 72, 72, and 72 arcmin for the tomographic bin pairs [1, 1], [1, 2], [1, 3], [1, 4], [2, 2], [2, 3], [2, 4], [3, 3], [3, 4], and [4, 4]. The minimum scales considered for  $\xi_-$  are substantially larger due to the fact that  $\xi_-$  probes higher multipoles in harmonic space compared to  $\xi_+$  and it is more affected by baryonic effects. We note that it is difficult to directly compare these scales to the minimum smoothing scales considered for the moments analysis, as these statistics probe scales differently.

The measurement covariance has been obtained using jackknife resampling and a fiducial DES Y3 simulation (DeRose et al. 2019). The shear two-point analysis delivers slightly tighter posteriors than second moments alone, but is less constraining than the combination of second and third moments. Indeed, we find it to constrain  $\Omega_m$  and  $S_8$  at the level of 12 per cent and 1.8 per cent; the combined second and third moments result is 20 per cent more constraining. Without measuring the cross-covariance between moments and shear two-point correlation function, it is hard to quantitatively explain why the latter is more constraining than second moments alone. One reason could be that they have access to the same information (the power spectrum), but they probe scales differently (the two-point correlation function is more localized in harmonic space, whereas moments get contributions from a broader range of multipoles, being prone to baryonic effects at all smoothing scales). A different sensitivity to the effects that drives the scale selection can limit the constraining power of a probe compared to others (see e.g. Asgari et al. 2019). More in general, a different sensitivity to angular scales might cause different observables to be only weakly correlated, even if they belong to the same category of two-point statistics (see e.g. Hamana et al. 2019). Future works will investigate further the correlation between two-point correlation function and second and third moments.

One relevant feature that can be observed from Figs 9 and 10, is that shear two-point correlation function has a similar degeneracy direction compared to second moments only. Combining shear two-point correlation function with any other probes sensitive to the bispectrum (such as the third moments) is likely to significantly improve the constraints due to the different degeneracy direction of their constraints.

## 7 SUMMARY

In this paper, we have presented a simulated cosmology analysis using the second and third moments of the weak lensing mass (convergence) maps. We targeted the analysis at the 3 yr (Y3) data from the DES, but the methods developed here are general and can be applied to other data sets. The goal of this paper was to describe and validate the methodology using simulations, determining the lower bounds on scales where systematic or modelling uncertainties are not expected to affect the cosmological analysis. A companion paper, applying the methodology to DES Y3 data, together with observational systematic checks (e.g. potential systematic effects such as modelling errors in the point spread function, inhomogeneities in the noise, and spurious dependencies of shear with observing conditions) and consistency checks with the results from other DES Y3 probes and external data sets, will follow.

The second moment of the convergence as a function of smoothing scale contains similar information as the standard two-point shear correlations. The third moment, or the skewness, contains additional non-Gaussian information of the field. We described how the convergence maps are constructed starting from the shear catalogue using a full-sky Kaiser–Squires (Kaiser & Squires 1993; Chang et al. 2018) formalism. We obtain analytical predictions for the second and third moments using perturbation theory. We included the effects of partial sky coverage in the theoretical modelling of the moments using the pseudo- $C_\ell$  formalism. We validated the modelling of the convergence moments using a large suite of simulations, including the effects of the survey mask and non-linear lensing corrections (such as reduced-shear and source crowding). We used the same simulations to estimate the covariance. We furthermore showed how the computation of theoretical predictions can be sped up without introducing biases in the cosmological analysis by implementing a 5-parameter emulator.

We tested our pipeline through simulated likelihood analyses varying five cosmological parameters ( $\Omega_m$ ,  $\sigma_8$ ,  $n_s$ ,  $\Omega_b$ , and  $h_{100}$ ) and 10 nuisance parameters (modelling redshift uncertainties, shear biases, and intrinsic alignments). We determined the scale cuts based on the impact of baryonic physics and modelling inaccuracies of the third moments at small scales.

We then simulated the constraints achievable with a DES Y3 analysis. We found that second moments, third moments, and their combination constrain  $\Omega_m$  to 17 per cent, 66 per cent, and 10 per cent, respectively, and  $S_8$  to 1.8 per cent, 3.6 per cent, and 1.5 per cent, respectively. The combination of second and third moments provides improved constraints with respect to second moments due to the extra non-Gaussian information probed by the third moments and the different inclination of the degeneracy axis in the  $\sigma_8$ – $\Omega_m$  plane of the two probes. For DES Y5, where we expect to have a data set with higher galaxy density, we forecast a further improvement in the constraining power at the level of 10–20 per cent.

We also compared with a simulated shear two-point analysis for DES Y3, which yields constraints at the level of 12 per cent and 1.8 per cent for  $\Omega_m$  and  $S_8$ . The combined second and third moments result is about 20 per cent more constraining. This analysis shows the importance of including in the analysis probes of higher order statistics to improve the cosmological constraints.

## ACKNOWLEDGEMENTS

The authors thank the anonymous referee for reading the paper carefully and providing many insightful comments and suggestions.

CC is supported by the Henry Luce Foundation. G. Longhi’s colour scheme suggestions are gratefully acknowledged.

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

The Collaborating Institutions are Argonne National Laboratory, the University of California, 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, National Optical Astronomy Observatory, which is operated 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 7th Framework Program (FP7/2007-2013) including ERC grant agreements 240672, 291329, and 306478. We acknowledge support from the Brazilian Instituto Nacional de Ciência e Tecnologia (INCT) 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.

## DATA AVAILABILITY

The simulated data used in this work have been either generated using the public code FLASK (<http://www.astro.iag.usp.br/~flask/>) or using the public T17 simulations ([http://cosmo.phys.hirosaki-u.ac.jp/takahasi/allsky\\_raytracing/](http://cosmo.phys.hirosaki-u.ac.jp/takahasi/allsky_raytracing/)). The derived simulated data and the code used in this article will be shared on reasonable request to the corresponding author.

## REFERENCES

- Aihara H. et al., 2018, *PASJ*, 70, S4
- Alsing J., Wandelt B., Feeney S., 2018, *MNRAS*, 477, 2874
- Amendola L., 1996, *Astrophys. Lett. Commun.*, 33, 63
- Asgari M., Taylor A., Joachimi B., Kitching T. D., 2018, *MNRAS*, 479, 454
- Asgari M. et al., 2019, *Astronomy and Astrophysics*, 624, 134
- Bartelmann M., Schneider P., 2001, *Phys. Rep.*, 340, 291
- Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013, *ApJ*, 762, 109
- Bernardeau F., van Waerbeke L., Mellier Y., 1997, *A&A*, 322, 1
- Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, *Phys. Rep.*, 367, 1
- Bridle S., King L., 2007, *New J. Phys.*, 9, 444
- Brown M. L., Castro P. G., Taylor A. N., 2005, *MNRAS*, 360, 1262

- Bruderer C., Chang C., Refregier A., Amara A., Bergé J., Gamper L., 2016, *ApJ*, 817, 25
- Castro P. G., Heavens A. F., Kitching T. D., 2005, *Phys. Rev. D*, 72, 023516
- Chang C. et al., 2018, *MNRAS*, 475, 3165
- Chisari N. E. et al., 2018, *MNRAS*, 480, 3962
- Clerkin L. et al., 2017, *MNRAS*, 466, 1444
- Coles P., Jones B., 1991, *MNRAS*, 248, 1
- Crocce M., Puelbas S., Scoccimarro R., 2006, *MNRAS*, 373, 369
- Das S., Ostriker J. P., 2006, *ApJ*, 645, 1
- de Jong J. T. A. et al., 2013, *The Messenger*, 154, 44
- DeRose J. et al., 2019, preprint ([arXiv:1901.02401](https://arxiv.org/abs/1901.02401))
- Dietrich J. P., Hartlap J., 2010, *MNRAS*, 402, 1049
- Dodelson S., Schneider M. D., 2013, *Phys. Rev. D*, 88, 063537
- Einstein A., 1936, *Science*, 84, 506
- Flaugher B., 2005, *Int. J. Mod. Phys. A*, 20, 3121
- Fluri J., Kacprzak T., Refregier A., Amara A., Lucchi A., Hofmann T., 2018, *Phys. Rev. D*, 98, 123518
- Foreman S., Coulton W., Villaescusa-Navarro F., Barreira A., 2020, *Astrophysical Journal*, 895, 113
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
- Fosalba P., Gaztañaga E., Castander F. J., Manera M., 2008, *MNRAS*, 391, 435
- Friedrich O., Eifler T., 2018, *MNRAS*, 473, 4150
- Friedrich O. et al., 2018, *Phys. Rev. D*, 98, 023508
- Fu L. et al., 2014, *MNRAS*, 441, 2725
- Gaztanaga E., Bernardeau F., 1998, *A&A*, 331, 829
- Gelman A., Rubin D. B., 1992, *Stat. Sci.*, 7, 457
- Gil-Marín H., Wagner C., Fragkoudi F., Jimenez R., Verde L., 2012, *J. Cosmol. Astropart. Phys.*, 2012, 047 (GM12)
- Gil-Marín H., Noreña J., Verde L., Percival W. J., Wagner C., Manera M., Schneider D. P., 2015, *MNRAS*, 451, 539
- 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
- Gualdi D., Manera M., Joachimi B., Lahav O., 2018, *MNRAS*, 476, 4045
- Habib S., Heitmann K., Higdon D., Nakhleh C., Williams B., 2007, *Phys. Rev. D*, 76, 083503
- Hamana T., Sakurai J., Koike M., Miller L., 2015, *PASJ*, 67, 34
- Hamana T. et al., 2020, *Publications of the ASJ*, 72, 16
- Harnois-Déraps J. et al., 2016, *MNRAS*, 460, 434
- Hartlap J., Simon P., Schneider P., 2007, *A&A*, 464, 399
- Heavens A. F., Jimenez R., Lahav O., 2000, *MNRAS*, 317, 965
- Heavens A. F., Sellentin E., de Mijolla D., Vianello A., 2017, *MNRAS*, 472, 4244
- Heitmann K., Higdon D., Nakhleh C., Habib S., 2006, *ApJ*, 646, L1
- Heitmann K., Higdon D., White M., Habib S., Williams B. J., Lawrence E., Wagner C., 2009, *ApJ*, 705, 156
- Hellwing W. A., Schaller M., Frenk C. S., Theuns T., Schaye J., Bower R. G., Crain R. A., 2016, *MNRAS*, 461, L11
- Hikage C., Oguri M., 2016, *MNRAS*, 462, 1359
- Hikage C., Takada M., Hamana T., Spergel D., 2011, *MNRAS*, 412, 65
- Hikage C. et al., 2019, *PASJ*, 71, 43
- Hilbert S., Hartlap J., Schneider P., 2011, *A&A*, 536, A85
- Hildebrandt H. et al., 2017, *MNRAS*, 465, 1454
- Hinshaw G. et al., 2013, *ApJS*, 208, 19
- Hirata C. M., Seljak U., 2004, *Phys. Rev. D*, 70, 063526
- Hoekstra H., 2004, *MNRAS*, 347, 1337
- Hoyle B., et al., 2018, *MNRAS*, 478, 592
- Hubble E., 1934, *ApJ*, 79, 8
- Jain B., Seljak U., 1997, *ApJ*, 484, 560
- Jeffrey N. et al., 2018, *MNRAS*, 479, 2871
- Jeffrey N., Lanusse F., Lahav O., Starck J.-L., 2020, *MNRAS*, 492, 5023
- Joachimi B., Taylor A. N., Kiessling A., 2011, *MNRAS*, 418, 145
- Kacprzak T. et al., 2016, *MNRAS*, 463, 3653
- Kacprzak T. et al., 2020, *PRD*, 101, 082003
- Kaiser N., Squires G., 1993, *ApJ*, 404, 441
- Knabenhans M. et al., 2019, *MNRAS*, 484, 5509
- Kratochvil J. M., Haiman Z., May M., 2010, *Phys. Rev. D*, 81, 043519
- Kratochvil J. M., Lim E. A., Wang S., Haiman Z., May M., Huffenberger K., 2012, *Phys. Rev. D*, 85, 103513
- Krause E., Hirata C. M., 2010, *A&A*, 523, A28
- Krause E., Eifler et al., 2017, preprint ([arXiv:1706.09359](https://arxiv.org/abs/1706.09359))
- Kwan J., Heitmann K., Habib S., Padmanabhan N., Lawrence E., Finkel H., Frontiere N., Pope A., 2015, *ApJ*, 810, 35
- Leistedt B., McEwen J. D., Büttner M., Peiris H. V., 2017, *MNRAS*, 466, 3728
- Limber D. N., 1953, *ApJ*, 117, 134
- Liu J., Petri A., Haiman Z., Hui L., Kratochvil J. M., May M., 2015, *Phys. Rev. D*, 91, 063507
- Martinet N. et al., 2018, *MNRAS*, 474, 712
- Mawdsley B. et al., 2020, *MNRAS*, 493, 5662
- McKay M. D., Beckman R. J., Conover W. J., 1979
- Naidoo K., Whiteway L., Massara E., Gualdi D., Lahav O., Viel M., Gil-Marín H., Font-Ribera A., 2019, *MNRAS*, 491, 1709
- Parroni C., Cardone V. F., Maoli R., Scaramella R., 2020, *A&A*, 633, A71
- Patton K., Blazek J., Honscheid K., Huff E., Melchior P., Ross A. J., Suchyta E., 2016, preprint ([arXiv:1611.01486](https://arxiv.org/abs/1611.01486))
- Peel A., Pettorino V., Giocoli C., Starck J.-L., Baldi M., 2018, *A&A*, 619, A38
- Petri A., Liu J., Haiman Z., May M., Hui L., Kratochvil J. M., 2015, *Phys. Rev. D*, 91, 103511
- Pires S., Starck J.-L., Amara A., Teyssier R., Réfrégier A., Fadili J., 2009, *MNRAS*, 395, 1265
- Pujol A. et al., 2016, *MNRAS*, 462, 35
- Rasmussen C. E., Williams C. K. I., 2006, *Gaussian Processes for Machine Learning*
- Schaye J. et al., 2010, *MNRAS*, 402, 1536
- Schmidt F., Rozo E., Dodelson S., Hui L., Sheldon E., 2009, *Phys. Rev. Lett.*, 103, 051301
- Schneider P., van Waerbeke L., Jain B., Kruse G., 1998, *MNRAS*, 296, 873
- Schneider P., van Waerbeke L., Mellier Y., 2002, *A&A*, 389, 729
- Scoccimarro R., Couchman H. M. P., 2001, *MNRAS*, 325, 1312 (SC01)
- Seljak U., Zaldarriaga M., 1996, *ApJ*, 469, 437
- Sellentin E., Heavens A. F., 2016, *MNRAS*, 456, L132
- Semboloni E., Schrabback T., van Waerbeke L., Vafaei S., Hartlap J., Hilbert S., 2011, *MNRAS*, 410, 143
- Shan H. et al., 2018, *MNRAS*, 474, 1116
- Simon P. et al., 2015, *MNRAS*, 449, 1505
- Springel V., 2005, *MNRAS*, 364, 1105
- Springel V. et al., 2018, *MNRAS*, 475, 676
- Takada M., Jain B., 2003, *MNRAS*, 344, 857
- Takada M., Jain B., 2004, *MNRAS*, 348, 897
- Takahashi R., Oguri M., Sato M., Hamana T., 2011, *ApJ*, 742, 15
- Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, *ApJ*, 761, 152
- Takahashi R., Soma S., Takada M., Kayo I., 2014, *MNRAS*, 444, 3473
- Takahashi R., Hamana T., Shirasaki M., Namikawa T., Nishimichi T., Osato K., Shirogama K., 2017, *ApJ*, 850, 24 (T17)
- Takahashi R., Nishimichi T., Namikawa T., Taruya A., Kayo I., Osato K., Kobayashi Y., Shirasaki M., 2019, preprint ([arXiv:1911.07886](https://arxiv.org/abs/1911.07886))
- Taruya A., Hamana T., Kayo I., 2002, in Ikeuchi S., Hearnshaw J., Hanawa T., eds, *Proc. IAU, Vol II, 8th Asian-Pacific Regional Meeting*. San Francisco, Calif : Astronomical Society of the Pacific, National Center of Sciences, Hitotsubashi Memorial Hall, Tokyo, p. 299
- Tegmark M., Taylor A. N., Heavens A. F., 1997, *ApJ*, 480, 22
- Troxel M. A. et al., 2018, *Phys. Rev. D*, 98, 043528
- Vafaei S., Lu T., van Waerbeke L., Semboloni E., Heymans C., Pen U.-L., 2010, *Astropart. Phys.*, 32, 340
- van Daalen M. P., Schaye J., Booth C. M., Dalla Vecchia C., 2011, *MNRAS*, 415, 3649
- Van Waerbeke L., Hamana T., Scoccimarro R., Colombi S., Bernardeau F., 2001, *MNRAS*, 322, 918
- Van Waerbeke L. et al., 2013, *MNRAS*, 433, 3373

- Vicinanza M., Cardone V. F., Maoli R., Scaramella R., Er X., 2016, preprint (arXiv:1606.03892)
- Vicinanza M., Cardone V. F., Maoli R., Scaramella R., Er X., 2018, Phys. Rev. D, 97, 023519
- Vicinanza M., Cardone V. F., Maoli R., Scaramella R., Er X., Tereno I., 2019, Phys. Rev. D, 99, 043534
- Wallis C. G. R., McEwen J. D., Kitching T. D., Leistedt B., Plouviez A., 2017, preprint (arXiv:1703.09233)
- Wandelt B. D., Hivon E., Górski K. M., 2001, Phys. Rev. D, 64, 083003
- Wild V. et al., 2005, MNRAS, 356, 247
- Xavier H. S., Abdalla F. B., Joachimi B., 2016, MNRAS, 459, 3693
- Zuntz J. et al., 2018, MNRAS, 481, 1149

## APPENDIX A: SKEWNESS PARAMETER

### A1 Skewness derivation using perturbation theory

In perturbation theory, the Fourier space equations of motion for the matter density contrast  $\delta$  and the divergence of the velocity field  $\theta = \nabla \cdot \mathbf{v}$  are (Bernardeau et al. 2002)

$$\begin{aligned} \frac{\partial \delta(\mathbf{k}, \tau)}{\partial \tau} + \theta(\mathbf{k}, \tau) \\ = - \int d^3 k_1 d^3 k_2 \delta_D(\mathbf{k} - \mathbf{k}_{12}), \alpha(\mathbf{k}_1, \mathbf{k}_2) \delta(\mathbf{k}_1, \tau) \theta(\mathbf{k}_2, \tau) \\ \equiv \alpha[\delta, \theta, \mathbf{k}], \end{aligned} \quad (\text{A1})$$

$$\begin{aligned} \frac{\partial \theta(\mathbf{k}, \tau)}{\partial \tau} + H\theta(\mathbf{k}, \tau) + \frac{3\Omega_m H_0^2}{2a} \delta(\mathbf{k}, \tau) \\ = - \int d^3 k_1 d^3 k_2 \delta_D(\mathbf{k} - \mathbf{k}_{12}), \beta(\mathbf{k}_1, \mathbf{k}_2) \theta(\mathbf{k}_1, \tau) \theta(\mathbf{k}_2, \tau) \\ \equiv \beta[\delta, \theta, \mathbf{k}], \end{aligned} \quad (\text{A2})$$

with  $\tau$  being the conformal time,  $a$  the scale factor,  $H = \frac{d}{d\tau} \ln a$ ,  $\mathbf{k}_{12} = \mathbf{k}_1 + \mathbf{k}_2$  and  $\alpha$  and  $\beta$  defined by

$$\alpha(\mathbf{k}_1, \mathbf{k}_2) = 1 + \frac{1}{2} \frac{\mathbf{k}_1 \mathbf{k}_2}{k_1 k_2} \left( \frac{k_1}{k_2} + \frac{k_2}{k_1} \right), \quad (\text{A3})$$

$$\beta(\mathbf{k}_1, \mathbf{k}_2) = \frac{1}{2} \frac{\mathbf{k}_1 \mathbf{k}_2}{k_1 k_2} \left( \frac{k_1}{k_2} + \frac{k_2}{k_1} \right) + \frac{(\mathbf{k}_1 \mathbf{k}_2)^2}{k_1^2 k_2^2}. \quad (\text{A4})$$

The matter density contrast and the divergence of the velocity field can be expanded as

$$\delta(\mathbf{k}, \tau) = \sum_{n=1} \delta_n(\mathbf{k}, \tau), \quad (\text{A5})$$

$$\theta(\mathbf{k}, \tau) = - \frac{\partial \ln D_+(\tau)}{\partial \tau} \sum_{n=1} \theta_n(\mathbf{k}, \tau), \quad (\text{A6})$$

**Table A1.** Values of the coefficients for the fitting formula described in equations (A14), (A15), and (A16) from SC01 and GM12.

Coefficient	SC01	GM12
$\alpha_1$	0.25	0.484
$\alpha_2$	3.5	3.740
$\alpha_3$	2	-0.849
$\alpha_4$	1	0.392
$\alpha_5$	2	1.013
$\alpha_6$	-0.2	-0.575
$\alpha_7$	1	0.128
$\alpha_8$	0	-0.722
$\alpha_9$	0	-0.926

where  $n$  indicates the order at which the fields are approximated and  $D_+$  is the linear growth factor. At linear order,  $\delta_1(\mathbf{k}, \tau) = \theta_1(\mathbf{k}, \tau) = D_+(\tau) \delta_1(\mathbf{k})$ .

At second order, the Fourier equations of motion are solved by

$$\delta_2(\mathbf{k}, \tau) = D_+^2(\tau) \alpha[\delta_1, \delta_1, \mathbf{k}] + D_2(\tau) (\beta[\delta_1, \delta_1, \mathbf{k}] - \alpha[\delta_1, \delta_1, \mathbf{k}]), \quad (\text{A7})$$

with  $D_2$  the solution of the following differential equation:

$$\frac{\partial^2 D_2(\tau)}{\partial \tau^2} + H \frac{\partial D_2(\tau)}{\partial \tau} - \frac{3\Omega_m H_0^2}{2a} D_2(\tau) = \left( \frac{\partial D_+(\tau)}{\partial \tau} \right)^2 \quad (\text{A8})$$

Lastly, we define the following quantity  $\mu$ , as it will enter in the modelling of the third moment:

$$\mu \equiv 1 - D_2/D_+^2. \quad (\text{A9})$$

At leading order in perturbation theory, one can compute the variance of the dark matter density field smoothed by a top hat filter as

$$\langle \delta_{\theta_0, \text{lin}}^2 \rangle(\tau) = \frac{1}{2\pi} \int dk k W(k, \theta_0)^2 P_{\text{lin}}(k, \tau); \quad (\text{A10})$$

while the skewness will be described by the following equation:

$$\begin{aligned} \langle \delta_{\theta_0, \text{lin}}^3 \rangle(\tau) = \frac{6}{(2\pi)^3} \int d^2 k_1 d^2 k_2 W(\mathbf{k}_1, \theta_0) W(\mathbf{k}_2, \theta_0) W(\mathbf{k}_1 + \mathbf{k}_2, \theta_0) \\ \times P_{\text{lin}}(\mathbf{k}_1, \tau), P_{\text{lin}}(\mathbf{k}_2, \tau) F_2(\mathbf{k}_1, \mathbf{k}_2, \tau), \end{aligned} \quad (\text{A11})$$

where  $P_{\text{lin}}(k, \tau)$  is the linear power spectrum and  $W(k, \theta_0)$  is the top hat filter described in equation (13). The term  $F_2(\mathbf{k}_1, \mathbf{k}_2, \tau)$  reads

$$\begin{aligned} F_2(\mathbf{k}_1, \mathbf{k}_2, \tau) = \frac{1}{2} \left[ \left( 1 + \frac{k_1}{k_2} \cos \phi \right) + \left( 1 + \frac{k_2}{k_1} \cos \phi \right) \right] \\ + [1 - \mu(\tau)] (\cos^2 \phi - 1), \end{aligned} \quad (\text{A12})$$

with  $\phi$  the angle between  $\mathbf{k}_1$  and  $\mathbf{k}_2$ . We implement here a refinement of the term  $F_2$  based on  $N$ -body simulations (while equation A12 has been obtained, so far, exclusively relying on perturbation theory). The refinement we are implementing here has been first obtained by Scoccimarro & Couchman (2001; hereafter SC01) and later on by Gil-Marín et al. (2012, hereafter GM12) fitting an analytical formula to the non-linear evolution of the bispectrum based on a suite of cold dark matter  $N$ -body simulations. Implementing such corrections, equation (A12) becomes

$$\begin{aligned} F_2(\mathbf{k}_1, \mathbf{k}_2, \tau) = \frac{1}{2} b_1 b_2 \left[ \left( 1 + \frac{k_1}{k_2} \cos \phi \right) + \left( 1 + \frac{k_2}{k_1} \cos \phi \right) \right] \\ + [1 - \mu(\tau)] c_1 c_2 (\cos^2 \phi - 1) + [a_1 a_2 \mu(\tau) - b_1 b_2] \\ + [1 - \mu(\tau)] c_1 c_2. \end{aligned} \quad (\text{A13})$$

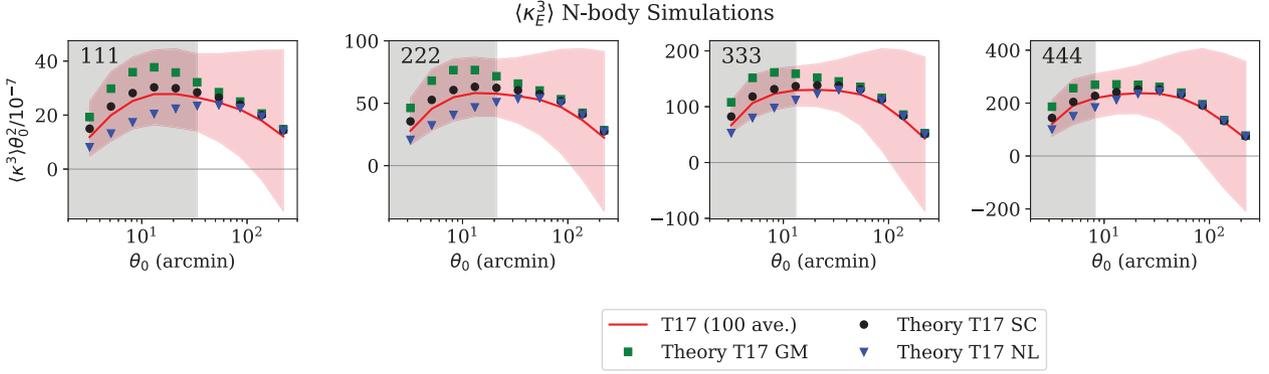
The terms  $a, b, c$  are taken from GM12; their subscripts in the above equations indicate if they refer to  $k_1$  or  $k_2$ . In particular

$$a(n, k, \tau) = \frac{1 + (\sigma_8 D_+)^{a_6} [0.7(4 - 2^n)/(1 + 2^{2n+1})]^{1/2} (q a_1)^{n+a_2}}{1 + (q a_1)^{n+a_2}}, \quad (\text{A14})$$

$$b(n, k, \tau) = \frac{1 + 0.2 a_3 (n + 3) (q a_7)^{n+3+a_8}}{1 + (q a_7)^{n+3.5+a_8}}, \quad (\text{A15})$$

$$c(n, k, \tau) = \frac{1 + 4.5 a_4 / [1.5 + (n + 3)^4 (q a_5)^{n+3+a_9}]}{1 + (q a_5)^{n+3.5+a_9}}. \quad (\text{A16})$$

In the above equations,  $n$  is the slope of the linear power spectrum at scale  $k$  and  $q \equiv k/k_{\text{NL}}$ , where  $k_{\text{NL}}$  is the scale where non-linearities



**Figure A1.** Comparison between measured third moments in **T17** simulations (red line) and theoretical predictions using different modelling choices for the small-scales regime: **SC01** fitting formulae (black points), **GM12** fitting formulae (green points), and the perturbation theory prediction obtained using the non-linear power spectrum (NL, blue points). The red shaded region corresponds to the  $1\sigma$  uncertainty of the measurement. The grey shaded regions indicate the angular scales excluded in the main cosmological analysis when combining with second moments (i.e. a  $24 h^{-1}$  Mpc scale cut).

start to be important and it is defined so that  $k_{\text{NL}}^3 P(k, \tau) / 2\pi^2 = 1$ . We report in Table A1, the values of the coefficients  $a_1, \dots, a_9$  as from **SC01** and **GM12**. Implementing these corrections in equation (A11) leads to

$$\begin{aligned} \langle \delta_{\theta_0, \text{lin}}^3 \rangle(\tau) &= \frac{6}{(4\pi^2)^2} \int dk_1 dk_2 W(k_1, \theta_0) W(k_2, \theta_0) \\ &\times P_{\text{lin}}(k_1, \tau), P_{\text{lin}}(k_2, \tau) \int d\phi W \\ &\times \left( \sqrt{k_1^2 + k_2^2 + 2k_1 k_2 \cos\phi}, \theta_0 \right) \\ &\times F_2(k_1, k_2, \phi, \tau). \end{aligned} \quad (\text{A17})$$

The integral on the angle  $\phi$  can be written as

$$\begin{aligned} &\int d\phi W \left( \sqrt{k_1^2 + k_2^2 + 2k_1 k_2 \cos\phi}, \theta_0 \right) F_2(k_1, k_2, \phi, \tau) \\ &= \frac{1}{2} b_1 b_2 \int d\phi W \left( \sqrt{k_1^2 + k_2^2 + 2k_1 k_2 \cos\phi}, \theta_0 \right) \\ &\times \left[ 2 + \left( \frac{k_1}{k_2} + \frac{k_2}{k_1} \right) \cos\phi \right] \\ &+ \int d\phi W \left( \sqrt{k_1^2 + k_2^2 + 2k_1 k_2 \cos\phi}, \theta_0 \right) \\ &\times [(1 - \mu) c_1 c_2 (\cos^2\phi - 1)] \\ &+ \int d\phi W \left( \sqrt{k_1^2 + k_2^2 + 2k_1 k_2 \cos\phi}, \theta_0 \right) \\ &\times [a_1 a_2 \mu - b_1 b_2 + (1 - \mu) c_1 c_2]. \end{aligned} \quad (\text{A18})$$

For brevity, we omitted the dependence on  $\tau$  in  $\mu$ . The three integrals in equation (A18) can be solved as

$$\begin{aligned} &b_1 b_2 [2\pi W(k_1, \theta_0) W(k_2, \theta_0) + \frac{\pi}{2} \frac{\partial}{\partial \theta_0} (W(k_1, \theta_0) W(k_2, \theta_0))] \\ &- c_1 c_2 [\pi(1 - \mu) W(k_1, \theta_0) W(k_2, \theta_0)] \\ &+ 2\pi [a_1 a_2 \mu - b_1 b_2 + (1 - \mu) c_1 c_2] W(k_1, \theta_0) W(k_2, \theta_0) \\ &= \frac{\pi}{2} b_1 b_2 \frac{\partial}{\partial \theta_0} [W(k_1, \theta_0) W(k_2, \theta_0)] \\ &+ \pi [2a_1 a_2 - (1 - \mu) c_1 c_2] W(k_1, \theta_0) W(k_2, \theta_0). \end{aligned} \quad (\text{A19})$$

After some algebra, one can express equation (A17) as

$$\begin{aligned} \langle \delta_{\theta_0, \text{lin}}^3 \rangle(\tau) &= 6 \left[ \int dk k a W(k, \theta_0)^2 P_{\text{lin}}(k, \tau) \right]^2 \\ &- 3 \left[ \int dk k (1 - \mu) c W(k, \theta_0)^2 P_{\text{lin}}(k, \tau) \right]^2 \\ &+ \frac{3}{4} \frac{\partial}{\partial \ln \theta_0} \left[ \int dk k b W(k, \theta_0)^2 P_{\text{lin}}(k, \tau) \right]^2, \end{aligned} \quad (\text{A20})$$

$$\begin{aligned} \langle \delta_{\theta_0, \text{lin}}^3 \rangle(\tau) &= 3[2(\langle \delta_{\theta_0, \text{lin}, a}^2 \rangle(\tau))^2 - (1 - \mu)(\langle \delta_{\theta_0, \text{lin}, c}^2 \rangle(\tau))^2] \\ &+ \frac{3}{2} \frac{\partial \langle \delta_{\theta_0, \text{lin}, b}^2 \rangle(\tau)}{\partial \ln \theta_0}. \end{aligned} \quad (\text{A21})$$

In the above equation, we have defined

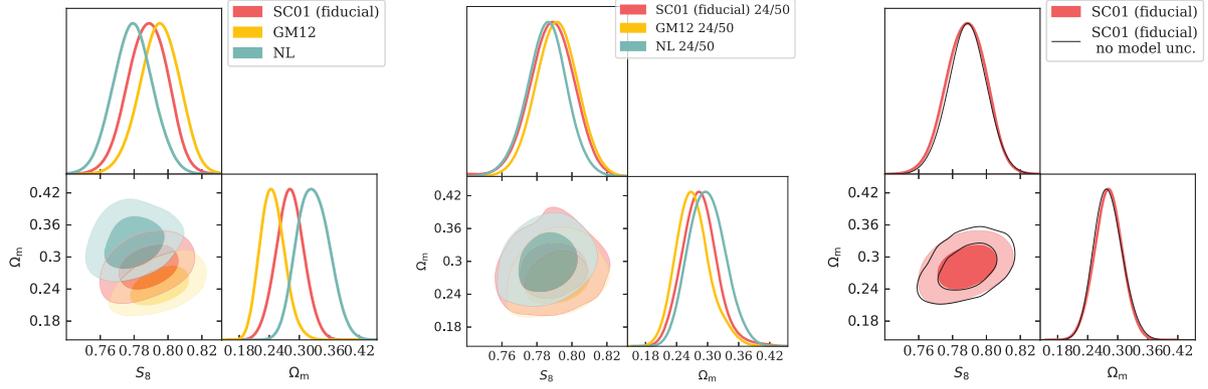
$$\langle \delta_{\theta_0, \text{lin}, x}^2 \rangle(\tau) = \frac{1}{2\pi} \int dk k X(k, \tau) W(k, \theta_0)^2 P_{\text{lin}}(k, \tau), \quad (\text{A22})$$

with  $X$  that can be either  $a$ ,  $b$ , or  $c$ . We finally define the reduced skewness parameter as

$$S_3 \equiv \frac{\langle \delta_{\theta_0, \text{lin}}^3 \rangle(\tau)}{[\langle \delta_{\theta_0, \text{lin}}^2 \rangle(\tau)]^2}. \quad (\text{A23})$$

The original perturbation theory result can be obtained noting that in the limit of  $a, b, c \rightarrow 1$  we have  $\delta_{\theta_0, \text{lin}, a}^2, \delta_{\theta_0, \text{lin}, b}^2, \delta_{\theta_0, \text{lin}, c}^2 \rightarrow \delta_{\theta_0, \text{lin}}^2$ ; in this case, the reduced skewness parameter assumes the following form:

$$S_3 \equiv \frac{\langle \delta_{\theta_0, \text{lin}}^3 \rangle(\tau)}{[\langle \delta_{\theta_0, \text{lin}}^2 \rangle(\tau)]^2} = 3(1 + \mu) + \frac{3}{2} \frac{\partial \ln \langle \delta_{\theta_0, \text{lin}}^2 \rangle(\tau)}{\partial \ln \theta_0}. \quad (\text{A24})$$



**Figure A2.** *Left-hand panel:* forecast posteriors obtained assuming three different theory data vectors using different modelling choices for the third moments small scales (the **SC01** and **GM12** models and the simple non-linear theory prediction). The three data vectors have been analysed assuming as true model the **SC01** model. The data vectors include second and third moments and assume the fiducial scale cut (i.e. a  $24 h^{-1}$  Mpc scale cut). *Central panel:* same as left-hand panel, but assuming a large scale cut ( $50 h^{-1}$  Mpc) for third moments. *Right-hand panel:* comparison between forecast posteriors obtained assuming a theory data vector, a fiducial scale cut ( $24 h^{-1}$  Mpc) and including (or not including) third moments modelling uncertainties in the covariance.

The equations above for the third moments hold in the linear regime, but they are usually extrapolated to the mild non-linear regime using predictions of the non-linear power spectrum.

## A2 Impact of different fitting formulae on the predicted third moments

The exact value of the predicted third moments at small scales depends on the particular fitting formulae chosen for the modelling of the skewness parameter. We note here that there is up to a 30 per cent difference between **SC01** and **GM12** fitting formulae at small scales ( $\sim 5$  arcmin for the first tomographic bin). This is shown in Fig. A1, along with the predicted third moments obtained without implementing the small-scales refinement (i.e. assuming the standard  $F_2$  kernel and the non-linear power spectrum). In our main analysis, we used the fitting formulae from **SC01** because they provide a better fit to our simulations, but in order to be conservative, we included the difference between the **SC01** and **GM12** models in our covariance in order to account for the small-scales modelling uncertainty of the skewness. In Fig. A2 (left-hand panel), we further show the level of bias we expect in the parameters posterior if the real Universe followed a model different than **SC01** for the third moments (which implies the simulations we used to validate our modelling are not accurate enough for validating third moments). For this test, we used a theory data vector that combines second and third moments: it includes all the ‘auto’ moments of different tomographic bins (e.g. [1,1], [1,1,1], [2,2], [2,2,2]) and the ‘cross’ moments (e.g. [1,2], [1,1,2], [1,2,2]) and it assumes a fiducial scale cut-off ( $24 h^{-1}$  Mpc). All the other nuisance parameters have been set to their nominal value (zero). We prepared three data vectors using the **SC01**, the **GM12** and the standard non-linear prediction for the third moments (NL). The level of bias in the parameters posterior is around  $\sim 1\sigma$  when analysing the **GM12** and NL data vector with the **SC01** model. We note that a more conservative scale cut for the third moments ( $50 h^{-1}$  Mpc) would strongly decrease the bias (central panel of Fig. A2). This would also lower the constraining power of the combined second and third moments, but the gain compared to using second moments only would still be considerable ( $\sim 30$  per cent for  $\Omega_m$ ,  $\sim 10$  per cent for  $S_8$ ).

Finally, the right-hand panel of Fig. A2 shows the change in the parameters posterior when including (or not) the third moments

modelling uncertainty in the covariance; the change in the posterior obtained combining second and third moments is minimal, for the scales used in the fiducial analysis.

## APPENDIX B: MODE–MODE COUPLING MATRICES

We provide here mathematical recipes for the mode–mode coupling matrices  $\mathbf{M}$  used in Section 2.2 to account for masking effects. Such matrices are developed in the context of pseudo-power spectrum estimators (e.g. Wandelt, Hivon & Górski 2001; Brown et al. 2005; Hikage et al. 2011; Hikage & Oguri 2016). In particular, we strictly follow here section 2.1 of Hikage et al. (2011).

In the presence of a window function (in our case, the DES Y3 footprint)  $K(\theta, \phi)$ , the shear field assumes the following expression:

$$\bar{\gamma}_1(\theta, \phi) + \bar{\gamma}_2(\theta, \phi) = K(\theta, \phi)(\gamma_1(\theta, \phi) + \gamma_2(\theta, \phi)). \quad (\text{B1})$$

When the shear field is transformed into its spherical harmonic counterpart (equation 7), it obtains an additional contribution due to the convolution with the footprint mask

$$\hat{\gamma}_{E,\ell m} \pm i \hat{\gamma}_{B,\ell m} = \int d\Omega [K(\theta, \phi)(\gamma_1(\theta, \phi) + \gamma_2(\theta, \phi))]_{\pm 2} Y_{\ell m}^*(\theta, \phi). \quad (\text{B2})$$

The quantities  $\hat{\gamma}_{E,\ell m}$  and  $\hat{\gamma}_{B,\ell m}$  are called pseudo E and B modes (as they are convolved with the footprint mask) and their relation with the true E and B modes can be written as

$$\hat{\gamma}_{E,\ell m} \pm i \hat{\gamma}_{B,\ell m} = \sum_{\ell' m'} (\hat{\gamma}_{E,\ell' m'} \pm i \hat{\gamma}_{B,\ell' m'})_{\pm 2} W_{\ell \ell' m m'}, \quad (\text{B3})$$

where  ${}_{\pm 2} W_{\ell \ell' m m'}$  is a convolution kernel

$$\begin{aligned} {}_{\pm 2} W_{\ell \ell' m m'} &= \int d\Omega {}_{\pm 2} Y_{\ell' m'}(\theta, \phi) K(\theta, \phi) {}_{\pm 2} Y_{\ell m}^*(\theta, \phi) \\ &= \sum_{\ell'' m''} K_{\ell'' m''} (-1)^m \sqrt{\frac{(2\ell + 1)(2\ell' + 1)(2\ell'' + 1)}{4\pi}} \\ &\quad \times \begin{pmatrix} \ell & \ell' & \ell'' \\ \pm 2 & \mp 2 & 0 \end{pmatrix} \begin{pmatrix} \ell & \ell' & \ell'' \\ m & m' & m'' \end{pmatrix}, \end{aligned} \quad (\text{B4})$$

with  $\begin{pmatrix} \ell & \ell' & \ell'' \\ m & m' & m'' \end{pmatrix}$  Wigner  $3j$  symbols and  $K_{\ell m} = \int d\Omega K(\theta, \phi) Y_{\ell m}^*(\theta, \phi)$ , the harmonic transform of the window function. Defining

$$\hat{C}_\ell^{EE} = \frac{1}{2\ell+1} \sum_m |\hat{\gamma}_{E,\ell m}|^2, \quad (\text{B5})$$

$$\hat{C}_\ell^{EB} = \frac{1}{2\ell+1} \sum_m \hat{\gamma}_{E,\ell m} \hat{\gamma}_{B,\ell m}^*, \quad (\text{B6})$$

$$\hat{C}_\ell^{BB} = \frac{1}{2\ell+1} \sum_m |\hat{\gamma}_{B,\ell m}|^2, \quad (\text{B7})$$

we can write the masked (pseudo-) spectra as the convolution of the true spectra with a mode–mode coupling matrix

$$\hat{C}_\ell = \sum_{\ell'} \mathbf{M}_{\ell\ell'} \mathbf{C}_{\ell'}, \quad (\text{B8})$$

where we introduced the vector  $\mathbf{C}_\ell \equiv (C_\ell^{EE}, C_\ell^{EB}, C_\ell^{BB})$ . The mode–mode coupling matrix  $\mathbf{M}$  is expressed in terms of  $M_{\ell\ell'}^{EE,EE}$ ,  $M_{\ell\ell'}^{BB,BB}$ ,  $M_{\ell\ell'}^{EB,EB}$ ,  $M_{\ell\ell'}^{EE,EB}$

$$\begin{aligned} M_{\ell\ell'}^{EE,EE} &= M_{\ell\ell'}^{BB,BB} \\ &= \frac{2\ell'+1}{8\pi} \sum_{\ell''} (2\ell''+1) K_{\ell''} [1 + (-1)^{\ell+\ell'+\ell''}] \\ &\quad \times \begin{pmatrix} \ell & \ell' & \ell'' \\ 2 & -2 & 0 \end{pmatrix}^2, \end{aligned} \quad (\text{B9})$$

$$\begin{aligned} M_{\ell\ell'}^{EE,BB} &= M_{\ell\ell'}^{BB,EE} \\ &= \frac{2\ell'+1}{8\pi} \sum_{\ell''} (2\ell''+1) K_{\ell''} [1 - (-1)^{\ell+\ell'+\ell''}] \\ &\quad \times \begin{pmatrix} \ell & \ell' & \ell'' \\ 2 & -2 & 0 \end{pmatrix}^2, \end{aligned} \quad (\text{B10})$$

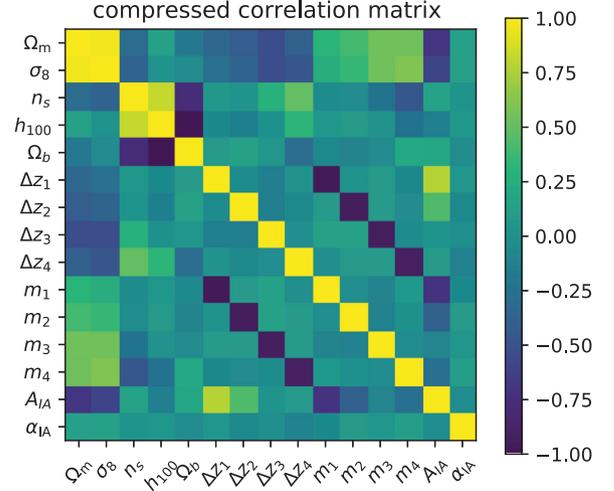
$$M_{\ell\ell'}^{EB,EB} = \frac{2\ell'+1}{4\pi} \sum_{\ell''} (2\ell''+1) K_{\ell''} \begin{pmatrix} \ell & \ell' & \ell'' \\ 2 & -2 & 0 \end{pmatrix}^2, \quad (\text{B11})$$

$$\text{with } K_\ell = \frac{1}{2\ell+1} \sum_m K_{\ell m} K_{\ell m}^*.$$

### APPENDIX C: CONSTRAINTS WITH DATA-COMPRESSION AND ALTERNATIVE COVARIANCE MATRIX

Our fiducial analysis has been carried out using a covariance matrix obtained from multiple FLASK realizations (see Section 3.1). FLASK is a lognormal simulation, where the only required inputs are the desired auto and cross power spectra of the convergence fields and the so-called lognormal shift parameters (which effectively set the skewness of the simulated fields at one scale, see e.g. Friedrich et al. 2018; Gruen et al. 2018). No additional physics is encoded in the FLASK maps. This means that our FLASK realizations reproduce the correct second moments set by our  $\Lambda$ CDM input spectra, but has only limited accuracy in its third moments. We have shown that this does not strongly bias the recovery of input cosmological parameters once applied to  $N$ -body simulations (see Section 6.1).

In this section, we show how to obtain cosmological constraints from our pipeline using the T17 covariance and compare them to the ones obtained from the flask covariance, using a data compression algorithm (described in Section 5.2). We also validate the efficiency



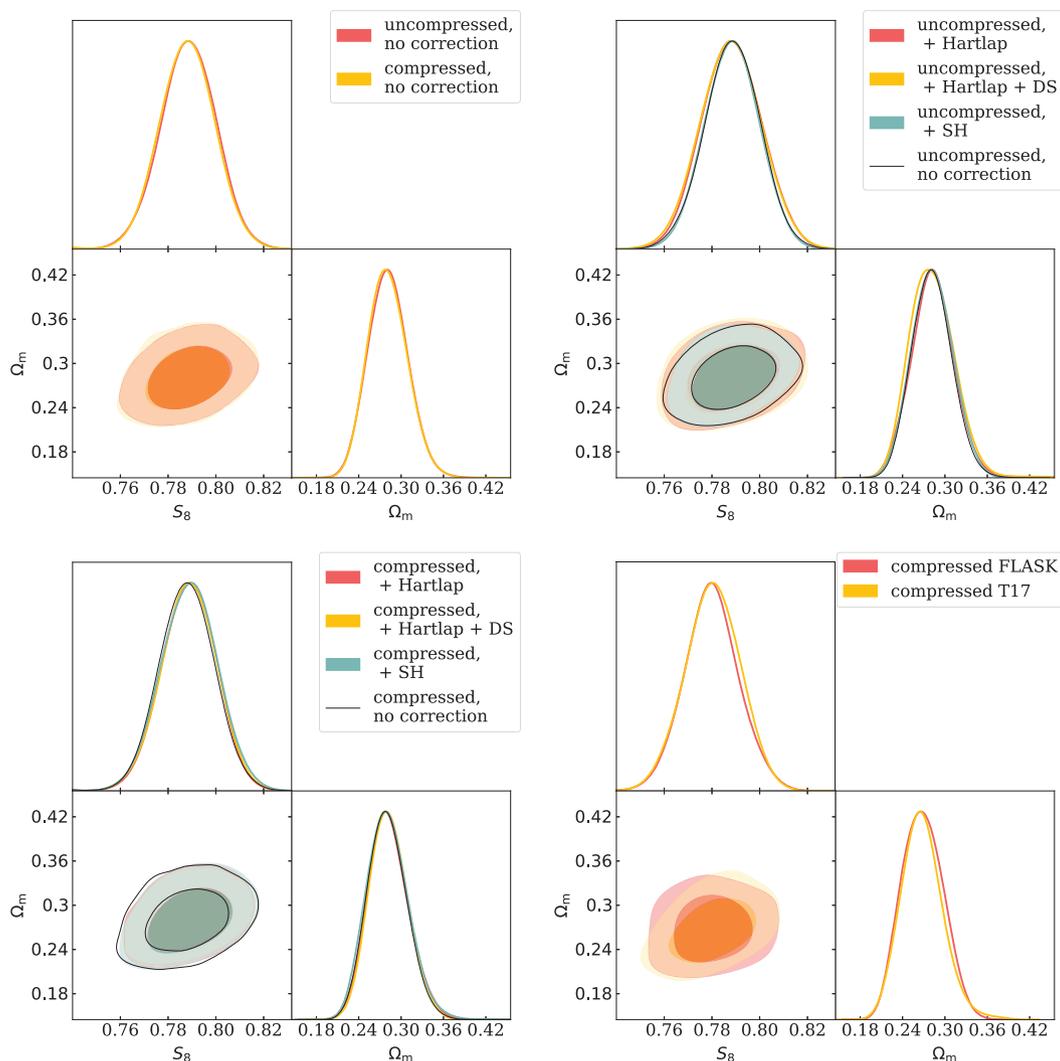
**Figure C1.** Measured compressed correlation matrix of second and third moments from 1000 FLASK simulations. A  $24 h^{-1}$  Mpc scale cut has been applied (see Section 6.1 for a definition of the scale cuts). The entries of the correlation matrix are shown with respect to the parameter used to compress the data vector.

of the data compression algorithm and show how it helps to reduce the noise in the inferred parameters caused by the paucity of simulations used to estimate the covariance matrix.

All the tests shown in this section use a theory data vector that includes all the ‘auto’ moments of different tomographic bins (e.g. [1,1], [1,1,1], [2,2], [2,2,2]) and the ‘cross’ moments (e.g. [1,2], [1,1,2], [1,2,2]), for a total of 10 combinations for second moments and 20 combinations for third moments. Depending on the test, we show results from second moments, third moments or the combination of the two. We use the fiducial scale cuts determined in Section 7, of  $20 h^{-1}$  Mpc for second moments,  $12 h^{-1}$  Mpc for third moments, and  $24 h^{-1}$  Mpc when second and third moments are combined. For the data vector, a fiducial T17 cosmology is assumed, with nuisance and astrophysical parameters (photo- $z$  biases, multiplicative shear biases, intrinsic alignment IA) assumed to be null, and no baryonic contamination. When estimating parameters posterior, we further marginalize over nuisance parameters as explained in Section 2.2.

We show the compressed correlation matrix in Fig. C1. The correlation matrix has now 15 entries, as many as the number of parameters we constrain in our analysis. Interestingly, the correlation between the different elements of the compressed data vector reflects the correlation between parameters. For instance,  $\Omega_m$  and  $\sigma_8$  show a significant correlation, as expected from Fig. 9. The anticorrelation between the mean shift  $\Delta z$  and the multiplicative shear biases  $m$  is due to the fact that the amplitude of the moments depends on the mean redshift of the source distribution (see e.g. Bernardeau et al. 1997); due to our definition and code implementation, a positive  $\delta z$  shifts the mean of the distribution to lower redshift and lowers the amplitude of the moments, while a positive  $m$  has the opposite effect.

We next perform here several tests to validate our compression algorithm. First, we run two forecast chains using the compressed and uncompressed FLASK covariance and compare the contours. This is shown in the top left panel of Fig. C2, for the  $\Omega_m$  and  $\sigma_8$  parameters. In this first test, we did not apply any correction for the noise in the inverse of the covariance (equations 49 and 50), as we are interested in validating the compression algorithm only. The marginalized 1D posteriors of  $\Omega_m$  and  $\sigma_8$  have similar width, showing that the data compression implemented is basically lossless.



**Figure C2.** Posterior of  $\Omega_m$  and  $S_8$  for four different cases. *Top left:* posteriors obtained using the uncompressed and compressed FLASK covariance, without applying any corrections due to noise (equations 49 and 50). *Top right:* posteriors obtained using uncompressed FLASK covariance, with a number of corrections to account for the noise in the estimated covariance matrix. ‘Hartlap’ refers to the Hartlap et al. (2007) correction (equation 49), ‘DS’ refers to the Dodelson & Schneider (2013) correction (equation 50), while ‘SH’ refers to the Sellentin & Heavens (2016) likelihood (see text in Appendix C for more details). *Bottom left:* same as the top right panel, but for compressed data vectors. *Bottom right:* posteriors obtained using the compressed FLASK and T17 covariances. 1000 FLASK simulations have been used for these tests.

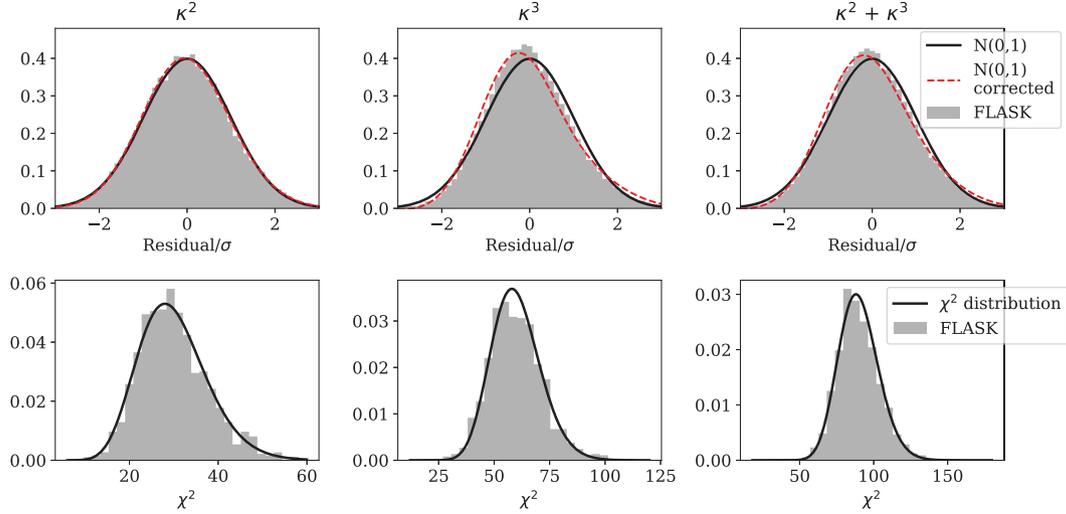
As a caveat, we remind the reader that we assume the likelihood to be Gaussian, which in the case of the uncompressed data vector is only an approximation (see below).

Second, we show in the top right panel of Fig. C2 how the constraints degrade once the uncertainties in the inverse of the covariance matrix are taken into account. The Hartlap et al. (2007) and Dodelson & Schneider (2013) corrections (equations 49 and 50) noticeably enlarge the contours, the net effect depending on the number of simulations used to estimate the covariance matrix. We also show, for comparison purposes, how the posteriors would look if the likelihood from Sellentin & Heavens (2016) was used. Sellentin & Heavens (2016) argue that when the covariance matrix is estimated from simulations, the likelihood is no longer Gaussian but rather is described by an adapted version of a multivariate  $t$ -distribution, a fact not taken into account by the Hartlap et al. (2007) correction. They suggest that marginalizing over the true covariance produces a tighter posterior close to the peak compared to the simple Hartlap et al. (2007) correction, which according to Sellentin & Heavens (2016) overestimates

its size. This is confirmed by the top right panel of Fig. C2. We note, however, that the additional scatter in the parameters posterior encoded by the Dodelson & Schneider (2013) correction is not accounted for in the Sellentin & Heavens (2016) framework.

The lower left panel of Fig. C2 is the same as the top right panel but for the compressed data vector. The compression greatly reduces the noise in the estimated covariance matrix and equations (49) and (50) approach  $\sim 1$ . Also the Sellentin & Heavens (2016) likelihood approaches a multivariate Gaussian, becoming almost indistinguishable from the no correction case.

Lastly, in the lower right panel of Fig. C2, we show the contours obtained using the compressed T17 covariance matrix. The Hartlap et al. (2007) and Dodelson & Schneider (2013) corrections are applied here, even though their effect is negligible, as shown above. We expect the shape of the posterior to be different when using the compressed T17 covariance and the FLASK covariance in two ways. First, the cosmology of the T17 simulations is slightly different from the FLASK one. Secondly, third moments should be more accurately



**Figure C3.** This figure is the same as Fig. 8 but for a uncompressed data vector. *Upper panels:* residuals (i.e. the difference between the measurement signal in a FLASK simulation and the simulations mean value) of individual data points in units of their expected standard deviation. We compare to a Gaussian with 0 mean and unit standard deviation; we also compare to a Gaussian corrected by the first term of the Edgeworth expansion of the likelihood (see text for more details). *Bottom panels:* Distribution of the  $\chi^2$  of each realization of the FLASK simulations, compared to a theoretical  $\chi^2$  distribution.

modelled in the T17 simulations a FLASK does not contain the physics to model the third moments beyond the lognormal shift. Differences in the widths between the two compressed covariances are smaller than 2 per cent, suggesting that the two factors considered above have a modest impact.

Finally, we comment on the more Gaussian nature of the compressed data vector compared to the uncompressed one. This is shown in Fig. C3. The residuals (i.e. the difference between the measurement signal in a FLASK simulation and the simulations mean value) of the uncompressed data vector appear much less Gaussian for the third moments and the combination of second and third moments compared to what we found for the compressed data vector in Fig. 8 (no significant difference in the distribution of the residuals is seen when only second moments are used). We compute how the distribution of residuals would look if the likelihood were not purely Gaussian, by means of a multivariate Edgeworth expansion of the

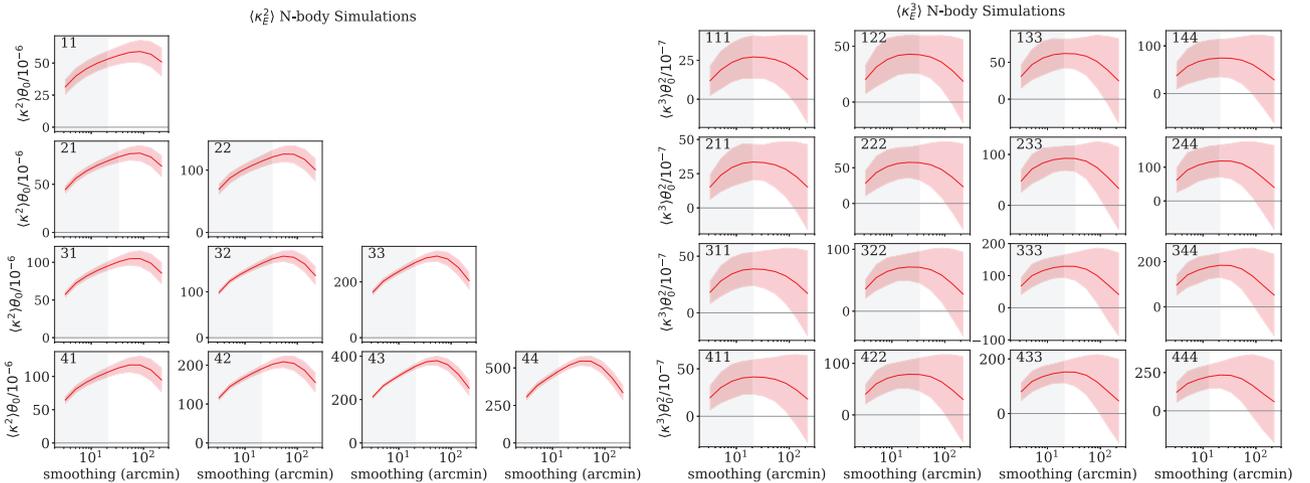
likelihood (e.g. Amendola 1996)

$$L = G(x, C) \left[ 1 + \frac{1}{6} k_x^{ijk} h_{ijk} + \dots \right], \quad (\text{C1})$$

with

$$h_{ijk} = (-1)^3 G^{-1}(x, C) \partial_{ijk} G(x, C), \quad (\text{C2})$$

where  $G(x, C)$  is the Gaussian part of the likelihood,  $x$  and  $C$  are the data vector and its covariance, respectively, and  $k_x^{ijk} = \langle x^i x^j x^k \rangle$  is the third-order cumulant of the data vector (which can be measured in simulations). The predicted distribution of residuals in Fig. C3 obtained with the first term of the Edgeworth expansion is in better agreement with the one measured in FLASK simulations.



**Figure D1.** Second and third moments data vector. The red line represents a theory data vector (obtained assuming the fiducial values for the nuisance parameters and a T17 cosmology) and the red shaded regions the  $1\sigma$  measurement uncertainty from the T17  $N$ -body simulations. Grey shaded regions show the angular scales not included in the fiducial analysis.

## APPENDIX D: FULL DATA VECTOR AND SIGNAL TO NOISE

We show in Fig. D1 the full data vector, including tomographic cross-correlation terms. The total signal-to-noise ratio (defined as  $\text{SNR} \equiv \sqrt{d^T C^{-1} d}$ ), for a  $24 h^{-1}$  Mpc scale cut, is 51.5, 10.4, and 66.3 for second, third, and combination of second and third moments, respectively. If all the scales are considered (down to 3 arcmin), the signal-to-noise ratio increases to 59.3, 15.4, and 106.0. Among the second moments bins, the 11, 22, 33, and 44 bins have signal-to-noise ratio of 8.8, 13.3, 25.2, and 24.5; as for the third moments, the 111, 222, 333, and 444 bins have signal-to-noise ratio of 2.1, 2.8, 4.6, and 3.9. The signal-to-noise ratio of the cross-moments is somewhat in between the values of the corresponding automoments.

## APPENDIX E: FAST EMULATOR THEORY PREDICTIONS

The theory prediction described in Section 2.2 can be quite time-consuming due to the large number of cross-correlations and integrations involved. In order to speed up this calculation, we implemented an emulator (Heitmann et al. 2006; Habib et al. 2007). Typically, emulators in the cosmology context are used when an expensive calculation is needed in a large parameter space, but the variation of the calculation over the parameter space is smooth. A primary example is predicting the dark matter power spectrum given cosmological parameters (Heitmann et al. 2009; Kwan et al. 2015). The power spectrum is computed to high accuracy at a given number of points in the cosmological parameter space and an interpolator is used to derive the power spectrum at some arbitrary point in the parameter space.

In our case, the quantities we wish to emulate are the second and third moments of the matter density field once the mask edges are accounted for, namely  $\langle \delta_{\theta_0, \text{NL}}^2 \rangle^{EE/BB}(\chi)$  and  $\langle \delta_{\theta_0, \text{NL}}^3 \rangle^{EE/BB}(\chi)$ , as a function of the 5 cosmological parameters under study and for a large number of (fixed) redshifts. We first compute their values at specific points in our parameter space, and then we build an interpolator that provides fast prediction at any point of the 5-parameter space (and as a function of redshift). We also build a separate emulator for each smoothing scale considered. The fact that such quantities are emulated at a fixed range of redshifts allows us to compute the second and third moments of the convergence field for a given tomographic bin after the emulation step, by integrating over the redshifts and taking into account the lensing kernel (and other nuisance parameters).

To decide which points to use for building the interpolator, we sampled our parameter space using a Latin hypercube (McKay, Beckman & Conover 1979), which is a scheme that provides good space-filling properties. We sampled the space delimited by the priors defined in Table 1, and chose 500 points. For each point of the Latin hypercube, we predicted the second and third moments of the dark matter density field (equations 14 and 15) with a resolution of  $\delta z = 0.01$  up to redshift 4, for 12 equally logarithmic spaced smoothing scales between  $\theta_0 = 0$  and  $\theta_0 = 220$  arcmin. For each smoothing scale, we organized the predictions of our second and third moments in a matrix of dimensionality  $n_z \times n_{\text{points}} = 400 \times 500$ . Since interpolating a  $400 \times 500$  matrix as a function of cosmological parameters would be impractical, we further reduce the dimensionality using the singular value decomposition. We define  $\eta = \mathbf{UBV}^T$ , where  $\mathbf{U}$  has dimensionality  $n_z \times n_z$  and  $\mathbf{V}$   $n_z \times n_{\text{points}}$ .  $\mathbf{B}$  is a diagonal matrix of singular values. We defined the basis vectors

$\Phi = \frac{1}{\sqrt{n_z}} \mathbf{UB}$  and weights  $\omega = \sqrt{n_z} \mathbf{V}^T$ . Then, we kept only the first  $p < n_z$  principal components of our basis vectors

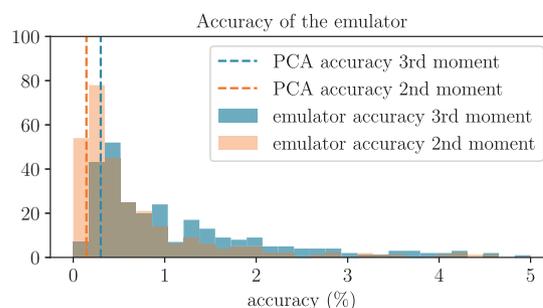
$$\begin{aligned} \langle \delta_{\theta_0, \text{NL}}^2 \rangle^{EE/BB}(\chi(z), \Omega_m, \Omega_b, \sigma_8, n_s, h_{100}) \\ = \sum_{i=0}^p \omega_i^{\delta^2, \theta_0}(\Omega_m, \Omega_b, \sigma_8, n_s, h_{100}) \Phi_i^{\delta^2, \theta_0}(\chi(z)), \end{aligned} \quad (\text{E1})$$

$$\begin{aligned} \langle \delta_{\theta_0, \text{NL}}^3 \rangle^{EE/BB}(\chi(z), \Omega_m, \Omega_b, \sigma_8, n_s, h_{100}) \\ = \sum_{i=0}^p \omega_i^{\delta^3, \theta_0}(\Omega_m, \Omega_b, \sigma_8, n_s, h_{100}) \Phi_i^{\delta^3, \theta_0}(\chi(z)), \end{aligned} \quad (\text{E2})$$

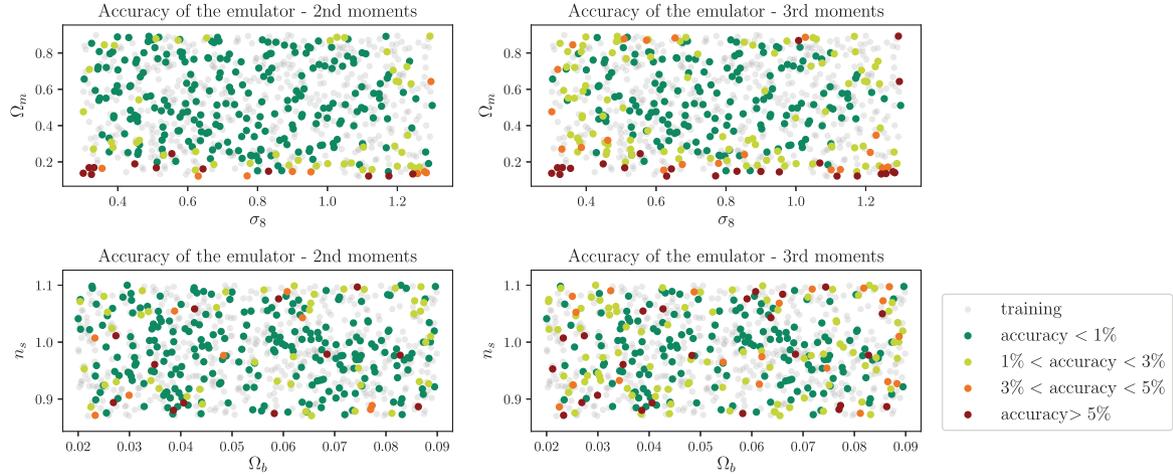
where the basis and weights are different for the second and third moments and depend on the smoothing scale. We found that setting  $p = 15$  and  $p = 45$  retains most of the information in the moments (99.9 percent and 99.7 percent for second and third moments, respectively), so we can neglect the other components. The third moments require more components due to the complex dependence on cosmological parameters at small scales.

After the singular value decomposition, we are left to interpolate, as a function of five cosmological parameters, 60 weight functions in total between  $\omega_i^{\delta^2, \theta_0}$  and  $\omega_i^{\delta^3, \theta_0}$  measured at 500 different points in our parameter space. We opted for a Gaussian process (Rasmussen & Williams 2006) interpolation scheme. A Gaussian process is a stochastic process where any finite subset forms a multivariate Gaussian distribution. At each reconstruction point  $x = (\Omega_m, \Omega_b, \sigma_8, n_s, \text{and } h_{100})$  of our parameter space, the weights  $\omega_i^{\delta^2, \theta_0}$ ,  $\omega_i^{\delta^3, \theta_0}$  are modelled as multivariate Gaussian distributions with a given mean value and Gaussian errors. The latter is determined by a covariance function  $k(x; x')$  that correlates the function at different points. The covariance function depends on only two hyperparameters (the amplitude and the typical scale of the correlation) which are fixed during the training phase.

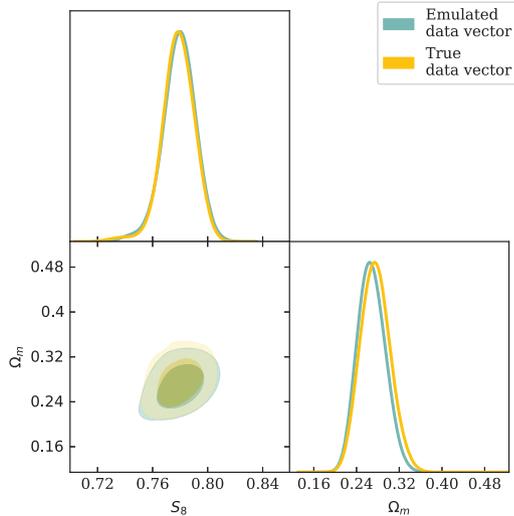
We tested the accuracy (defined as the mean relative discrepancy between the emulator predictions and the validation model over all the smoothing scales and redshifts considered) of our interpolation scheme by training and validating over two different sets of 500 points determined using two different Latin hypercubes. The resulting accuracy is shown in Figs E1 and E2. The performance of the emulator is generally better than 1 percent. The recovery gets worse close the edges of the priors. This is particularly evident for  $\sigma_8$  and  $\Omega_m$  (Fig. E2) as these two are the parameters to which



**Figure E1.** Accuracy of the emulator for the second and third moments. We tested the emulator using a validation sample of 500 points. Each entry of the histogram refers to the mean relative discrepancy between the emulator predictions and the validation model over all the smoothing scales and redshifts considered. The vertical dashed lines show the error introduced by selecting only 15 principal components.



**Figure E2.** Accuracy of the emulator for the second and third moments as a function of four cosmological parameters. We tested the emulator using a validation sample of 500 points. Each entry of the scatter plots refers to the mean relative discrepancy between the emulator predictions and the validation model over all the smoothing scales and redshifts considered. Training points are shown in grey.



**Figure E3.** Forecast posteriors for cosmological parameters, obtained with a theory data vector and an emulated data vector (see Section E). We marginalize over nuisance parameters as explained in Section 5.3. Constraints with the second and third moments combined are shown in the  $S_8$ – $\Omega_m$  plane.

our measurement is most sensitive. The emulator performs slightly worse for the third moment, due to a more complex dependence on the cosmological parameters. We note that Figs E1 and E2 report the mean accuracy of the emulator across smoothing scales and redshifts. While for the second moments, the accuracy does not strongly depend on the smoothing scales or redshift, we found that the emulator for the third moments performs slightly worse at low redshift and intermediate scales, where the accuracy is around  $\sim 3$  percent, still well below observational uncertainties. The speedup achieved by using the emulator is of two orders of magnitudes.

After predicting the masked second and third moments of the dark matter density field with the emulator, we took into account the lensing kernel of the samples and the nuisance parameters as described in Section 2.2. We checked that the emulated theory data vector causes small variations in the  $\chi^2$  with respect to a theory data vector obtained without approximations. For the fiducial cosmology,

such variations are of the order of  $\Delta\chi^2 \sim 0.2$ – $0.4$ , the exact value depending on the particular scale cut combination of second and third moments considered. We also verified that the difference between the maximum of the 1D marginalized posterior of the cosmological parameters obtained running an MCMC chain on an emulated theory data vector and on a non-approximated one are much smaller than the parameters’  $1\sigma$  confidence intervals. This is shown in Fig. E3, and the differences are at the level of  $< 1.5$  percent for  $\Omega_m$  and  $< 0.3$  percent for  $S_8 = \sigma_8(\Omega_m/0.3)^{0.5}$ .

<sup>1</sup>*Institut de Física d’Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, E-08193 Bellaterra, Spain*

<sup>2</sup>*Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA*

<sup>3</sup>*Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA*

<sup>4</sup>*Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*

<sup>5</sup>*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*

<sup>6</sup>*Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK*

<sup>7</sup>*Institut d’Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain*

<sup>8</sup>*Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain*

<sup>9</sup>*Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA*

<sup>10</sup>*Kavli Institute for Particle Astrophysics & Cosmology, Stanford University, PO Box 2450, Stanford, CA 94305, USA*

<sup>11</sup>*Institute of Theoretical Astrophysics, University of Oslo, PO Box 1029 Blindern, NO-0315 Oslo, Norway*

<sup>12</sup>*SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA*

<sup>13</sup>*Jodrell Bank Center for Astrophysics, School of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK*

<sup>14</sup>*Department of Physics & Astronomy, University College London, Gower Street, London WC1E 6BT, UK*

<sup>15</sup>*Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA*

<sup>16</sup>*Department of Physics, The Ohio State University, Columbus, OH 43210, USA*

<sup>17</sup>*Department of Physics, University of Arizona, Tucson, AZ 85721, USA*

<sup>18</sup>*Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, Casilla 603, La Serena, Chile*

<sup>19</sup>*Fermi National Accelerator Laboratory, PO Box 500, Batavia, IL 60510, USA*

<sup>20</sup>*Instituto de Física Teórica UAM/CSIC, Universidad Autónoma de Madrid, E-28049 Madrid, Spain*

<sup>21</sup>*Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain*

<sup>22</sup>*Laboratório Interinstitucional de e-Astronomia - LIneA, Rua Gal. José Cristino 77, Rio de Janeiro 20921-400, Brazil*

<sup>23</sup>*Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801, USA*

<sup>24</sup>*National Center for Supercomputing Applications, 1205 West Clark St, Urbana, IL 61801, USA*

<sup>25</sup>*Physics Department, University of Wisconsin-Madison, 2320 Chamberlin Hall, 1150 University Avenue Madison, WI 53706-1390, USA*

<sup>26</sup>*Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro 20921-400, Brazil*

<sup>27</sup>*Department of Physics, IIT Hyderabad, Kandi, Telangana 502285, India*

<sup>28</sup>*Department of Astronomy/Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721-0065, USA*

<sup>29</sup>*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109, USA*

<sup>30</sup>*Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA*

<sup>31</sup>*Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA*

<sup>32</sup>*Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA*

<sup>33</sup>*Australian Astronomical Optics, Macquarie University, North Ryde, NSW 2113, Australia*

<sup>34</sup>*Lowell Observatory, 1400 Mars Hill Rd, Flagstaff, AZ 86001, USA*

<sup>35</sup>*Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, CP 66318, São Paulo, SP 05314-970, Brazil*

<sup>36</sup>*George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, and Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA*

<sup>37</sup>*Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA*

<sup>38</sup>*Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain*

<sup>39</sup>*School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK*

<sup>40</sup>*Physics Department, Brandeis University, 415 South Street, Waltham MA, USA*

<sup>41</sup>*Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA*

<sup>42</sup>*Department of Physics, Duke University Durham, NC 27708, USA*

<sup>43</sup>*Institute for Astronomy, University of Edinburgh, Edinburgh EH9 3HJ, UK*

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