



HAL
open science

Dark Energy Survey Year 3 Results: Redshift Calibration of the MagLim Lens Sample from the combination of SOMPZ and clustering and its impact on Cosmology

G. Giannini, A. Alarcon, M. Gatti, A. Porredon, M. Crocce, G.M. Bernstein, R. Cawthon, C. Sánchez, C. Doux, J. Elvin-Poole, et al.

► **To cite this version:**

G. Giannini, A. Alarcon, M. Gatti, A. Porredon, M. Crocce, et al.. Dark Energy Survey Year 3 Results: Redshift Calibration of the MagLim Lens Sample from the combination of SOMPZ and clustering and its impact on Cosmology. *Monthly Notices of the Royal Astronomical Society*, 2024, 527 (2), pp.2010-2036. <10.1093/mnras/stad2945>. <hal-03791422>

HAL Id: hal-03791422

<https://hal.science/hal-03791422v1>

Submitted on 20 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Dark Energy Survey Year 3 results: redshift calibration of the MAGLIM lens sample from the combination of SOMPZ and clustering and its impact on cosmology

G. Giannini^{1,2,3*}, A. Alarcon⁴, M. Gatti⁵, A. Porredon^{6,7}, M. Crocce^{8,9}, G. M. Bernstein⁵, R. Cawthon¹⁰, C. Sánchez³, C. Doux⁵, J. Elvin-Poole^{6,7}, M. Raveri⁵, J. Myles^{11,12,13}, H. Lin¹⁴, A. Amon^{15,16}, S. Allam¹⁴, O. Alves¹⁷, F. Andrade-Oliveira¹⁸, E. Baxter¹⁹, K. Bechtol²⁰, M. R. Becker⁴, J. Blazek²¹, H. Camacho^{18,22}, A. Campos²³, A. Carnero Rosell^{24,25}, M. Carrasco Kind^{26,27}, A. Choi^{6,7}, J. Cordero²⁸, J. De Vicente²⁹, J. DeRose³⁰, H. T. Diehl¹⁴, S. Dodelson^{23,31}, A. Drlica-Wagner^{32,14,33}, K. Eckert⁵, X. Fang^{34,35}, A. Farahi³⁶, P. Fosalba^{8,9}, O. Friedrich¹⁷, D. Gruen³⁷, R. A. Gruendl^{26,27}, J. Gschwend^{22,38}, I. Harrison³⁹, W. G. Hartley³⁸, E. M. Huff³², M. Jarvis⁵, E. Krause⁴⁰, N. Kuropatkin¹⁴, P. Lemos^{39,41}, N. MacCrann⁴², J. McCullough¹⁴, J. Muir¹¹, S. Pandey⁵, J. Prat^{1,2}, M. Rodriguez-Monroy²⁹, A. J. Ross⁶, E. S. Rykoff^{12,13}, S. Samuroff²³, L. F. Secco^{1,2}, I. Sevilla-Noarbe²⁹, E. Sheldon⁴³, M. A. Troxel⁴⁴, D. L. Tucker¹⁴, N. Weaverdyck¹⁷, B. Yanny¹⁴, B. Yin²³, Y. Zhang⁴, T. M. C. Abbott⁴⁵, M. Aguena²², D. Bacon⁴⁶, E. Bertin^{47,48}, S. Bocquet³⁵, D. Brooks³⁹, D. L. Burke¹¹, J. Carretero³, F. J. Castander^{8,9}, M. Costanzi^{49,50,51}, L. N. da Costa²², M. E. S. Pereira⁵², S. Desai⁵³, P. Doel³⁹, I. Ferrero⁵⁴, B. Flaugher¹⁴, D. Friedel²⁴, J. Frieman^{14,1,2}, J. García-Bellido⁵⁵, D. W. Gerdes^{56,17}, G. Gutierrez¹⁴, S. R. Hinton⁵⁷, D. L. Hollowood⁵⁸, K. Honscheid^{6,7}, D. J. James⁵⁹, S. Kent^{14,31}, K. Kuehn^{60,61}, O. Lahav³⁹, C. Lidman^{62,63}, M. Lima^{64,21}, P. Melchior⁶⁵, J. Mena-Fernández²⁹, F. Menanteau^{26,27}, R. Miquel^{3,66}, R. L. C. Ogando³⁶, M. Paterno¹⁴, F. Paz-Chinchón^{26,67}, A. Pieres^{22,36}, A. A. Plazas Malagón⁶⁵, A. Roodman¹¹, E. Sanchez²⁹, V. Scarpine¹⁴, M. Smith⁶⁸, E. Suchyta⁶⁹, M. E. C. Swanson³⁹, G. Tarle¹⁷, D. Thomas⁴⁶, C. To⁶ and M. Vincenzi^{46,68} (DES Collaboration)

Affiliations are listed at the end of the paper

Accepted 2023 September 20. Received 2023 September 6; in original form 2023 February 1

ABSTRACT

We present an alternative calibration of the MAGLIM lens sample redshift distributions from the Dark Energy Survey (DES) first 3 yr of data (Y3). The new calibration is based on a combination of a self-organizing-map-based scheme and clustering redshifts to estimate redshift distributions and inherent uncertainties, which is expected to be more accurate than the original DES Y3 redshift calibration of the lens sample. We describe in detail the methodology, and validate it on simulations and discuss the main effects dominating our error budget. The new calibration is in fair agreement with the fiducial DES Y3 $n(z)$ calibration, with only mild differences ($<3\sigma$) in the means and widths of the distributions. We study the impact of this new calibration on cosmological constraints, analysing DES Y3 galaxy clustering and galaxy–galaxy lensing measurements, assuming a Lambda cold dark matter cosmology. We obtain $\Omega_m = 0.30 \pm 0.04$, $\sigma_8 = 0.81 \pm 0.07$, and $S_8 = 0.81 \pm 0.04$, which implies a $\sim 0.4\sigma$ shift in the $\Omega - S_8$ plane compared to the fiducial DES Y3 results, highlighting the importance of the redshift calibration of the lens sample in multiprobe cosmological analyses.

Key words: gravitational lensing: weak – galaxies: distances and redshifts – dark energy.

1 INTRODUCTION

The Dark Energy Survey (DES, Flaugher et al. 2015) is currently the largest photometric galaxy survey to date, spanning 5000 deg² of the

Southern hemisphere and having detected hundreds of millions of galaxies. Together with other ongoing and future galaxy surveys (e.g. Kilo-Degree Survey KIDS, Kuijken et al. 2015; Hyper Suprime-Cam HSC, Aihara et al. 2018; Vera Rubin Observatory Legacy Survey of Space and Time (LSST), LSST Science Collaboration 2009; Euclid, Laureijs et al. 2011), DES can achieve competitive constraints on

* E-mail: giulia.giannini15@gmail.com

cosmological parameters by studying both the spatial distribution of the detected galaxies and by measuring the tiny distortions in their observed shapes due to gravitational lensing effects induced by the large-scale structure of the Universe. For instance, the analysis of the first 3 yr (Y3) of DES data (DES Collaboration 2022) placed tight constraints on cosmological parameters combining three different measurements of the two-point (3x2pt) correlation functions that involved galaxy positions and measured galaxy shapes. These measurements are as follows:

- (i) Cosmic shear, i.e. the two-point correlation function of galaxy shapes; the DES Y3 measurements (Amon et al. 2022; Secco et al. 2022) involve the angular correlation of 10^8 galaxy shapes from the weak lensing sample (Gatti et al. 2021), divided into four tomographic bins. We refer to this as the ‘source’ sample;
- (ii) galaxy clustering: the two-point correlation function of the positions of bright galaxies (which we refer to as the ‘lens’ sample) (Rodríguez-Monroy et al. 2022).
- (iii) galaxy–galaxy lensing: the cross-correlation function of galaxy shapes and the position of the galaxies of the lens sample (Prat et al. 2022).

The modelling of each of these correlation functions requires knowledge of the redshift distributions (from hereafter $n(z)$) of the two samples (lens and source galaxies), which have to be estimated with great accuracy in order to avoid biased cosmological results (Huterer et al. 2006; Cunha et al. 2012; Benjamin et al. 2013; Huterer, Cunha & Fang 2013; Bonnett et al. 2016; Hildebrandt et al. 2017, 2021; Hoyle et al. 2018; Joudaki et al. 2020; Tessore & Harrison 2020). The optimal solution would be to avail ourselves of spectroscopic observations, providing an accurate redshift measurement of each targeted galaxy. Unfortunately, it is not feasible to obtain said spectra other than for a small fraction of the science sample, due to the required time and cost of the observing campaign. Cosmological surveys like DES therefore have to use for their redshift estimation measurements only a few, noisy, broad-band fluxes, requiring inventive methods to create robust and unbiased redshift calibration pipelines.

For the DES Y3 3x2pt analysis, two different lens samples were used. The first sample is defined by selecting luminous red galaxies through the RedMaGiC algorithm (Rozo et al. 2016), which retains galaxies with high-quality photometric redshift, by fitting each galaxy to a red-sequence template. The galaxies passing the RedMaGiC selection have, however, a low number density, and the final sample comprises roughly 3000 000 galaxies. The second sample slightly compromises on the redshift accuracy to the benefit of a larger number density. The MAGLIM sample (Porredon et al. 2021) is a magnitude-limited sample with a number density more than three times greater than RedMaGiC, comprising roughly 10 000 000 galaxies. In the fiducial DES 3x2pt (DES Collaboration 2022) and 2x2pt analyses (Porredon et al. 2022) that rely on the MAGLIM sample; the redshift distributions of the sample have been characterized using the machine learning photometric redshift code Directional Neighbourhood Fitting (DNF, De Vicente, Sánchez & Sevilla-Noarbe 2016). In its current implementation, the DNF code provides per-galaxy redshift estimates using nearest neighbour techniques. The redshift distributions were then further calibrated using clustering redshift (hereafter WZ), which relies on cross-correlation measurements with spectroscopic samples (Cawthon et al. 2022). This calibration step also placed uncertainties on the redshift distribution estimates, which were modelled by ‘shifting’ and ‘stretching’ the redshift distributions.

This work presents an additional and more sophisticated calibration of the redshift distributions of the lens sample, and studies the impact of these new redshift distribution estimates on the cosmological constraints using DES Y3 galaxy clustering and galaxy–galaxy lensing measurements (2x2pt). In particular, we adopt an approach similar to the one adopted to characterize the redshift distributions of the DES Y3 weak lensing (WL) sample, presented in Myles et al. (2021) and Gatti et al. (2022). This methodology also combines photometric and clustering constraints to produce redshift estimates, and it is more powerful than the fiducial redshift calibration adopted for the lenses for a number of reasons. The photometric information is used to produce redshift estimates using a self-organizing-map-based scheme (hereafter SOMPZ), which allows a meticulous control over all the (known) potential sources of uncertainties affecting the estimates. The SOMPZ method works by leveraging the DES deep fields, which have deeper observations with additional photometric bands and overlap with many-band redshift surveys available. It is possible to reproduce realistic selection functions in the deep fields from the injection of galaxies into actual DES images using the sophisticated image simulation tool BALROG (Everett et al. 2022). The SOMPZ method provides an ensemble of redshift $n(z)$ for a given galaxy sample, which captures the uncertainties in the redshift distributions at all orders (i.e. not only in the mean or width of the distributions). The clustering constraints are then incorporated through a rigorous joint likelihood framework where the clustering data are forward modelled as a function of the input $n(z)$, and the specific WZ systematics are marginalized over. This scheme allows to draw $n(z)$ samples conditioned on both clustering and photometric measurements, improving the $n(z)$ estimates by correctly taking into account the significance of the information provided by each source of information. This combined approach has proven to be more robust than SOMPZ or WZ applied individually (Gatti et al. 2022), as the combination exploits the complementarity of both methods and reduces the overall $n(z)$ uncertainty.

The paper is organized as follows. In Section 2 we introduce all the samples used in this work, both on data and simulations. Simulated samples are used to validate the methodology. Section 3 summarizes the SOMPZ+WZ methodology adopted in this paper, also outlining the differences with the ‘standard’ SOMPZ+WZ methodology used to model the DES Y3 source redshift distributions (Myles et al. 2021; Gatti et al. 2022). Section 4 is devoted to the characterization of the method’s uncertainties. Section 5 presents the redshift distributions MAGLIM sample produced using the techniques described in this work. Section 6 describes the impact of this new redshift calibration on cosmological parameters estimation and compares it to the ‘fiducial’ constraints obtained using the DNF+WZ redshift calibration (Porredon et al. 2022). In Appendix A we provide details on the construction of the MAGLIM sample in simulations. Appendix B complements the paper with a validation of the methodology in simulations. In Appendix C are listed the values of parameters and the prior functions used in the cosmological inference; Appendix D discusses the impact of different redshift uncertainties marginalization techniques on the cosmological parameters estimation.

2 DATA

We describe in this section the data and simulated products used in this work. The samples used in this work are the following:

- (i) the DES MAGLIM sample, used as lenses in the DES cosmological analysis. Characterizing its redshift distribution is the main goal of this work;

(ii) the DES deep field samples, which are observed in small fields by DES with deeper observations than wide field ones and where information from additional photometric bands are available. Deep fields are a key element of the SOMPZ methodology;

(iii) the DES BALROG sample; this sample consists of software-injected deep field galaxies into DES wide field images and is a key element of the SOMPZ methodology;

(iv) the ‘redshift’ samples, which are a collection of either spectroscopic or multiband photometric samples collected by other surveys in the DES deep field region. The redshift samples are a key element in the SOMPZ methodology;

(v) BOSS/eBOSS spectroscopic galaxy catalogues; these are galaxies with spectroscopic redshift overlapping with the DES wide field footprint used for the WZ measurement;

(vi) the DES WL sample, used as sources in the DES cosmological analysis; we use the WL sample here when presenting the impact of MAGLIM SOMPZ redshift distributions on the cosmological analysis results.

All of these samples in data have also been reproduced in simulation for testing purposes.

2.1 DES Year 3 data

DES (Flaugher et al. 2015) is a five broad-band (*grizY*) photometric survey that mapped roughly 5000 deg² of the southern sky, using a 570 megapixel camera (DECam; Flaugher et al. 2015) mounted on the 4 m Blanco telescope at the Cerro Tololo Inter-American Observatory (CTIO) in Chile. In this work we use data from the first 3 yr (out of six) of observations (Y3), which were taken from 2013 August to 2016 February. The DES Data Management (DESDM) team was in charge of processing the raw images (Sevilla et al. 2011; Abbott et al. 2018; Morganson et al. 2018); full details are provided in Sevilla-Noarbe et al. (2021) and Gatti et al. (2021). The main catalogue upon which all the DES samples are built is the DES gold catalogue, obtained using observations in the *griz* bands. Objects belonging to the gold catalogue have passed a number of selections aimed at removing objects in problematic regions of the sky or anomalous detections (e.g. objects with pixels affected by saturation or truncation issues). The gold catalogue consists of 388 millions objects (Sevilla-Noarbe et al. 2021). Each object comes with morphological and photometric measurements based on two different pipelines, the Multi-Object Fitting pipeline (MOF) and the Single-Object Fitting pipeline (SOF). The former performs a simultaneous multi-object, multi-epoch, multiband fit to estimate morphology and photometric information; the latter does not perform the multi-object fit when it comes to crowded objects. The DES Y3 SOF implementation is faster and less prone to fit failures compared to the MOF pipeline, and it does not suffer from any significant loss in terms of accuracy (Sevilla-Noarbe et al. 2021).

2.2 MAGLIM sample

The main galaxy sample considered in this work is the MAGLIM sample. The MAGLIM sample is a subset of the DES gold catalogue and consists of bright galaxies selected with an ad hoc selection that optimizes the number density and the redshift accuracy of the sample (Porredon et al. 2021). The MAGLIM sample spans the full DES Y3 wide field footprint, for a total of ~ 4143 deg². SOF magnitudes in the *riz* bands¹ are used for the selection and photometry. The selection

¹We exclude the *g* band as its photometry is known to be affected by PSF estimation issues (Jarvis et al. 2021).

Table 1. Summary of the MAGLIM sample. We have outlined for each tomographic bin the redshift range (selected using DNF z_{mean}), the number of galaxies, the number density, and the magnification coefficient as measured in Elvin-Poole et al. (2023).

Bin	z range	N galaxies	n density	C_{flux}
1	[0.20, 0.40]	2236 473	0.150	0.43
2	[0.40, 0.55]	1599 500	0.107	0.30
3	[0.55, 0.70]	1627 413	0.109	1.75
4	[0.70, 0.85]	2175 184	0.146	1.94
5	[0.85, 0.95]	1583 686	0.106	1.56
6	[0.95, 1.05]	1494 250	0.100	2.96

is meant to be linear in redshift and magnitude, and reads

$$i < 4 * z_{\text{mean}} + 18 \quad i > 17.5, \quad (1)$$

where m_i is the i -band SOF magnitude and z_{mean} is a per-object redshift estimate from the photo- z code DNF (De Vicente et al. 2016; see also the next subsection). The sample is then further limited to the redshift range $0.2 < z_{\text{mean}} < 1.05$. This leads to a sample in the range $18.8 < i_{\text{mag}} < 22.2$. The MAGLIM sample is divided into six tomographic bins using DNF z_{mean} and considering the following bin edges: [0.2, 0.4, 0.55, 0.7, 0.85, 0.95, 1.05], with a total of a 10 716 506 galaxies, distributed across bins as summarized in Table 1. The MAGLIM sample is used as lens sample in the galaxy–galaxy lensing and galaxy clustering measurements of the DES Y3 2x2 cosmological analysis (Porredon et al. 2022).

2.2.1 DNF

The photo- z code DNF is used to define the MAGLIM selection and to define the MAGLIM tomographic bins. The DNF algorithm computes a point estimate z_{mean} of redshift of the galaxies by performing a fit to a hyper-plane in colour and magnitude space using up to 80 nearest neighbours taken from a reference sample made of spectroscopic galaxies with secure redshift information. For this purpose, a large number of spectroscopic catalogues collected by Gschwend et al. (2018) has been used, including spectra from SDSS DR4 (Abolfathi et al. 2018), OzDES (Lidman et al. 2020), VIPERS (Garilli et al. 2014), and from the PAU spectro-photometric catalogue (Eriksen et al. 2019). The total number of spectra used for training is $\sim 10^5$. DNF also provides a redshift estimate z_{DNF} drawn from the redshift PDF for each individual galaxy, although only the quantity z_{mean} (used for the selection and for the binning) is of interest in this work.

2.3 Deep fields sample

The Deep fields catalogue is a key element of the SOMPZ methodology. We provide here a few key details, but we refer the reader to Hartley et al. (2022) for extensive details and the characterization of the sample.

This work uses four different deep fields, i.e. E2, X3, C3, and COSMOS (COS) covering 3.32, 3.29, 1.94, and 1.38 square degrees, respectively. Each deep field has undergone a scrupulous masking procedure aimed at removing artefacts (e.g. cosmic rays, meteors, saturated pixels, etc.). Considering the final unmasked area overlapping with the UltraVISTA and VIDEO near-infrared (NIR) surveys (McCracken et al. 2012; Jarvis et al. 2013), which is needed to provide photometric information in additional bands, we are left with 5.2 square degrees of area for a total of 267 229 galaxies with measured u, g, r, i, z, J, H, K_s photometry with limiting magnitudes 24.64, 25.57, 25.28, 24.66, 24.06, 24.02, 23.69, and 23.58. Note that

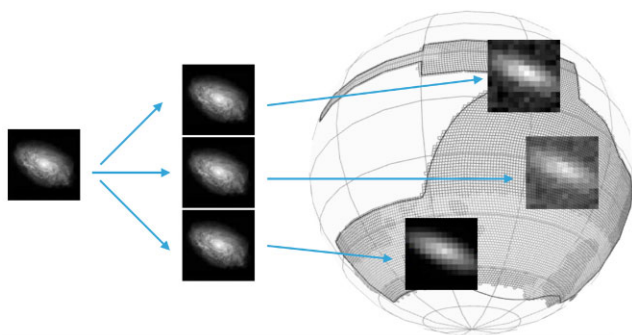


Figure 1. Scheme illustrating the operation of BALROG: the practically noiseless deep field galaxies are injected many times in the DES real wide field images; those dichotomous images are then processed through the fiducial DES detection pipeline, to construct a sample containing several noisy representations of the same deep galaxies.

deep field galaxies have deeper photometry and photometry available in more bands compared to the wide field galaxies; this is key for a good performance of the SOMPZ method as it reduces the colour-redshift degeneracy.

2.4 BALROG sample

The BALROG sample is another key element of the SOMPZ methodology. It is used to relate galaxies with given deep photometry to observed galaxies with wide field photometry, which are noisier. To this aim we rely on BALROG (Suchyta et al. 2016), a software which injects ‘fake’ galaxies into real images. For this analysis, BALROG was used to inject deep field galaxies into the broader wide field footprint (Everett et al. 2022). After injecting galaxies into images, the output BALROG images are passed into the DES Y3 photometric pipeline and injected galaxies are detected equivalently to real galaxies, yielding multiple realizations of each injected galaxy. The BALROG sample spans ~ 20 per cent of the DES Y3 footprint. We further select injected galaxies using the MAGLIM selection. We then construct a matched catalogue matching BALROG injected wide field MAGLIM galaxies with their deep field counterparts, for a total of 351 165 galaxies with both deep and wide photometric information. The resulting catalogue is called the BALROG sample. Fig. 1 shows a schematic illustration of Balrog.

2.5 Redshift samples

The redshift samples used for the SOMPZ section of the analysis consist of galaxies with secure redshift information (either spectroscopic or high-quality multiband photometric) observed in the deep fields. These samples are key to characterize the redshifts of the deep field sample and, in turn, to transfer the redshift information to the wide field MAGLIM sample.

We consider three separate redshift selections, similarly to what has been used in source sample redshift characterization (Myles et al. 2021):

(i) a collection of spectra from a number of different public and private spectroscopic samples, from the spectroscopic compilation by Gschwend et al. (2018). We have not restricted ourselves to a few, selected surveys as in the case of the DES Y3 weak lensing sample (Myles et al. 2021), where only zCOSMOS (Lilly et al. 2009), C3R2 (Masters et al. 2017, 2019), VVDS (Le Fèvre et al. 2013), and VIPERS (Scodreggio et al. 2018) were considered, because due

Table 2. Number of unique galaxies belonging to each of the three redshift catalogues (spectroscopic collection, COSMOS, and PAU) for each of the samples SPC (composed by galaxies from spectra, PAU, COSMOS in this order), SC (spectra, COSMOS), PC (PAU, COSMOS). The sample selection for the MAGLIM sample applied to the corresponding BALROG injections reduces greatly the size of all samples. For more information, see Section 2.5.

	Raw			After MagLim selection		
	SPC	SC	PC	SPC	SC	PC
Spec-z	35 826	35 826	–	10 429	10 429	–
PAU	18 780	–	28 780	3950	–	7015
COSMOS	64 139	82 856	69 686	3299	7231	3721
Total	118 745	118 682	98 466	17 678	17 660	10 736

the bright nature of the MAGLIM sample we would mostly select high signal-to-noise galaxies. Furthermore, using more spectra from different surveys allows us to simultaneously reduce the shot noise and improve the completeness of the sample, while minimizing the impact of possible outliers;

(ii) multiband photo-z galaxies from the COSMOS field; in particular, we used the COSMOS2015 30-band photometric redshift catalogue (Laigle et al. 2016), which is equipped with narrow, intermediate, and broad-bands covering the IR, optical, and UV regions of the electromagnetic spectrum;

(iii) redshifts from the PAUS + COSMOS 66-band photometric redshift catalogue (Alarcon et al. 2021), which adds 40 narrow-band filters from the PAU Survey.

We match these redshift catalogues to our deep field galaxies and keep only those that are selected at least once into our MAGLIM selection according to their BALROG injections. Due to the bright nature of the MAGLIM sample, the number of galaxies in our final redshift samples is greatly reduced: for the SPC sample, for example, the unique total number of galaxies passes from 118 745 to 17 718, a reduction of around 85 per cent.

In some cases, the same galaxy might have redshift information from multiple surveys. Following Myles et al. (2021), we created three slightly different redshift samples, where in case of multiple information from different surveys we use as fiducial the redshift from a specific survey. The samples are:

(1) *SPC*, where in case of multiple information available we first use the spectroscopic catalogue (S), then PAUS + COSMOS (P), and finally COSMOS2015 (C);

(2) *PC*, where we rank first the PAUS + COSMOS catalogue before COSMOS2015, and we do not include spectroscopic redshifts;

(3) *SC*: where we first use the spectroscopic catalogue before COSMOS2015, but we do not include the PAUS + COSMOS catalogue.

Table 2 summarizes the number of unique galaxies appearing in each of the three redshift samples, before and after performing the MAGLIM sample selection. The fiducial ensemble of redshift distributions is generated by marginalizing over all three of these redshift samples (SPC, PC, SC) with equal prior, which in practice is achieved by simply merging the $n(z)$ samples produced from the three redshift samples, creating a three times larger pool of $n(z)$. In such a way we marginalize over potential uncertainties and biases in the different redshift catalogues (S, P, and C).

Table 3. List of the spectroscopic samples from BOSS/eBOSS overlapping with the DES Y3 footprint used as reference galaxies for clustering redshifts in this work.

Spectroscopic samples			
Name	Redshifts	N_{gal}	Area
LOWZ (BOSS)	$z \in [0.0, 0.5]$	45 671	$\sim 860 \text{ deg}^2$
CMASS (BOSS)	$z \in [0.35, 0.8]$	74 186	$\sim 860 \text{ deg}^2$
LRG (eBOSS)	$z \in [0.6, 1.0]$	24 404	$\sim 700 \text{ deg}^2$
ELG (eBOSS)	$z \in [0.6, 1.1]$	89 967	$\sim 620 \text{ deg}^2$
QSO (eBOSS)	$z \in [0.8, 1.1]$	7 759	$\sim 700 \text{ deg}^2$

2.6 BOSS/eBOSS galaxy catalogues

The BOSS/eBOSS galaxy catalogue is our reference sample for the WZ measurement. It consists of a number of spectroscopic samples from the Sloan Digital Sky Survey (SDSS, Gunn et al. 2006; Eisenstein et al. 2011; Blanton et al. 2017), and combines SDSS galaxies from Baryonic Oscillation Spectroscopic Survey (BOSS, Dawson et al. 2013; Smee et al. 2013) and from extended-Baryon Oscillation Spectroscopic Survey (eBOSS, Dawson et al. 2016; Ahumada et al. 2020; Alam et al. 2021). In particular, the BOSS sample includes the LOWZ and CMASS catalogues from the SDSS DR 12 (Reid et al. 2016), while we included the large-scale structure catalogues from emission-line galaxies (ELGs, Raichoor et al. 2017), luminous red galaxies (LRGs, Prakash et al. 2016), and quasi-stellar objects (QSOs) (in preparation) from eBOSS. Following Cawthon et al. (2022) and Gatti et al. (2022), we stack together the different samples and use them as a single reference sample. We also create a single random catalogue by stacking all the random catalogues of each individual samples. The BOSS/eBOSS sample is divided into 50 bins spanning the $0.1 < z < 1.1$ range of the catalogue (width $\Delta z = 0.02$). The number of galaxies for each sample is listed in Table 3, with the final sample consisting of 241 987 objects and covering an area ranging from 14 to 17 per cent of the total DES footprint.

We note that estimates of the magnification coefficients are not available for BOSS/eBOSS galaxies. For our fiducial analysis we assumed magnification values for the BOSS/eBOSS sample to be set to zero. We are confident about this choice for the narrow shape of the MAGLIM tomographic bins, since the magnification is usually significant in the tails of the distribution when the clustering kernel due to selection effects is larger. We none the less verify in this work that our analysis is not very sensitive to the particular choice of the values of the magnification parameters (see Section 6.1.2).

2.7 Weak lensing catalogue

The DES Y3 WL sample is used in this work as a source in the galaxy–galaxy lensing measurement with the MAGLIM sample. The WL sample is created using the METACALIBRATION pipeline (described and tested in Huff & Mandelbaum 2017 and Sheldon & Huff 2017 and applied to the Y3 data in Gatti et al. 2021) and it is a subset of the gold catalogue. The METACALIBRATION pipeline provides a per-galaxy self-calibrated shape measurement, which is free from shear and selection biases. An additional, small calibration based on image simulations (MacCrann et al. 2022) accounts for blending and detection biases. The final catalogue consists of ~ 100 million galaxies, spanning the full DES Y3 wide field footprint and with an effective number density of $n_{\text{eff}} = 5.59 \text{ gal arcmin}^{-2}$. The WL sample is divided into four tomographic bin using the SOMPZ method (Myles et al. 2021).

2.8 Simulated galaxy catalogues

Our methodology is thoroughly validated using simulated catalogues. In particular, we use one realization of the sets of the Buzzard N -body simulations (DeRose et al. 2022). All the catalogues we used in data have their simulated counterparts, although we adopted some reasonable simplifications, when needed. We give here a brief summary of the Buzzard simulation and the simulated catalogue we had to create for this project, i.e. the simulated MAGLIM sample. The simulated BOSS/eBOSS catalogue description is provided in Gatti et al. (2022), whereas the simulated WL sample is described in DeRose et al. (2022).

Buzzard is a synthetic galaxy catalogue built starting from N -body light-cones produced by L-GADGET2 (Springel 2005). Galaxies are incorporated in the dark matter light-cones using the ADDGALS algorithm (DeRose et al. 2019). Buzzard spans 10 313 square degrees. The cosmological parameters chosen are $\Omega_m = 0.286$, $\sigma_8 = 0.82$, $\Omega_b = 0.047$, $n_s = 0.96$, $h = 0.7$. The simulations are created starting from three light-cones with different resolutions and size (1050^3 , 2600^3 , and $4000^3 \text{ Mpc}^3 h^{-3}$ boxes and 1400^3 , 2048^3 , and 2048^3 particles), to accommodate the need of a larger box at high redshift. Haloes are identified using the public code ROCKSTAR (Behroozi, Wechsler & Wu 2013) and they are populated with galaxies using ADDGALS (DeRose et al. 2019), which provides positions, velocities, magnitudes, spectral energy distributions (SEDs), and ellipticities. Galaxies are assigned their properties based on the relation between redshift, r -band absolute magnitude, and large-scale density from a subhalo abundance matching model (Conroy, Wechsler & Kravtsov 2006; Lehmann et al. 2017) in higher resolution N -body simulations. SEDs are assigned to galaxies by imposing the matching with the SED–luminosity–density relationship measured in the SDSS data. SEDs are K -corrected and integrated over the DES filter bands to generate DES *grizY* magnitudes. Ray-tracing is performed through the CALCLENS algorithm (Becker 2013), to introduce lensing effects, in order to provide weak-lensing shear, magnification, and lensed galaxy positions for the light-cone outputs. CALCLENS is run on to the sphere, masked with the DES Y3 footprint, using the HEALPIX algorithm (Górski & Hivon 2011) and is accurate to ~ 6.4 arcsec.

2.8.1 Simulated MAGLIM sample

In order to define a simulated MAGLIM sample, the photo- z code DNF has been run on a subset of the Buzzard simulations, restricted to i -band magnitudes $i < 23$, so as to reduce the running time without affecting the final result (note that the MAGLIM selection presents a cut at $i < 22.2$). The goal is to attain similar number density and colour distributions as in data. We provide more detailed information on the adaptation to the sample selection for Buzzard in Appendix A.

2.8.2 Simulated deep catalogue

The simulated true fluxes from Buzzard are used as the deep measurements, but we further assign a realistic error using the limiting flux for each mock deep band. We use the same uncertainties as in data, but as the Buzzard simulation has a different zero-point, those values have to be converted in magnitude using zero-point of 30, and then is converted to a flux uncertainty for a zero-point of 22.5, which is the zero-point of the Buzzard fluxes. We do not differentiate between fields, as it has been proven in Myles et al. (2021) that this had no impact on the simulated redshift distribution. The size of the sample is 968 759 galaxies. We use the true redshift for the redshift

sample and to compare our inferred redshift distributions to the true ones.

2.8.3 Simulated BALROG catalogue

We mimic the BALROG algorithm by randomly selecting positions over the full Y3 footprint and run the corresponding error model on the galaxies of the simulated deep catalogue to obtain noisy versions, according to the exposure times of each location. The deep galaxies can be injected an arbitrary number of times and we set this at 10. Only the wide counterparts of the deep galaxies that respect the MAGLIM selection defined in the Buzzard simulation are then included in the sample, yielding the final number of 250 193 galaxies.

3 REDSHIFT INFERENCE METHODOLOGY

We describe in this section the methodology adopted in this work to infer the redshift distributions of the lens sample. The methodology is similar to the one adopted for the weak lensing sample (Myles et al. 2021) and relies on two key techniques:

(i) photometric classification with self-organizing maps (SOM), known as the SOMPZ method (Buchs et al. 2019; Myles et al. 2021). The SOMPZ method takes advantage of the deeper photometry of eight bands (ugrizJHKs) available in the DES deep fields, where galaxies with high-quality redshifts can be accurately classified in the deep colour space, to ensure small selection biases, and well-characterized redshift estimates and uncertainties of DES wide field galaxies;

(ii) clustering-based or clustering redshift techniques (WZ), more established in cosmology (Newman 2008; Ménard et al. 2013). The redshift distributions calibration is based on angular correlation with a reference sample with high-quality redshift estimates. This method is affected by systematic biases different than photometric methods, which makes this combination interesting and improves the robustness of our redshift estimates. For example, it does not require the spectroscopic sample used for calibration to be representative of the target sample. On the other hand, the galaxy bias evolution of the galaxy samples is involved, and magnification effects have to be taken into account.

These two techniques are combined together to provide an estimate of the redshift distributions of the lens sample. Such a combination is powerful because it exploits the complementarity of the two methods, which are affected by two very different sets of biases and uncertainties. We provide the key ingredients of these two techniques in the following sections, followed by a description of how the two are combined together.

We note that this method is an alternate method compared to the one presented in Porredon et al. (2022) and Cawthon et al. (2022), which provides redshift estimates combining photometric estimates from the photo- z code DNF (De Vicente et al. 2016) and clustering constraints from Cawthon et al. (2022). We delay the comparison between the two methods to Section 5.1. Fig. 2 presents a flowchart illustration of MagLim redshift distributions calibration scheme.

3.1 SOMPZ methodology

The SOMPZ methodology estimates wide field redshift distributions by exploiting a mapping between wide field galaxies and deep field galaxies with deeper and more precise photometry. Extracting the redshift information from deep, several band photometry in order

to estimate the redshift of an observed wide field galaxy amounts to marginalizing over deep photometric information (Buchs et al. 2019). Let us consider the probability distribution function for the redshift of a galaxy $p(z)$; let us assume such a probability to be conditioned on observed wide field colour–magnitude $\hat{\mathbf{x}}$ and covariance matrix $\hat{\Sigma}$. The probability can be written by marginalizing over deep photometric colour \mathbf{x} as follows:

$$p(z|\hat{\mathbf{x}}, \hat{\Sigma}) = \int d\mathbf{x} p(z|\mathbf{x}, \hat{\mathbf{x}}, \hat{\Sigma})p(\mathbf{x}|\hat{\mathbf{x}}, \hat{\Sigma}). \quad (2)$$

The large dimensionality of this form prevents us from applying it to real situations. This problem can be circumvented by discretizing the colour space \mathbf{x} and $(\hat{\mathbf{x}}, \hat{\Sigma})$ in cells c and \hat{c} , each spanning a portion of the whole and representing a specific galaxy phenotype, respectively, of the deep and wide field. The galaxy samples are arranged in cells/phenotypes using SOM (Kohonen 1982), which is an unsupervised machine learning technique used to produce a lower dimensional representation of a complex data set, while preserving its core properties. The choice of the topology of the cells follows Buchs et al. (2019), where a two-dimensional representation of the colour space was chosen as it ensures an immediate visualization of the data not possible otherwise. Once we compressed our data in a more manageable set of information, we can write the $p(z)$ for the group of galaxies living in a particular wide cell \hat{c} . Since the MAGLIM tomographic bins \hat{b} are already defined, we are going to construct one set of SOMs (one deep and one wide) for each bin. Assigning all galaxies belonging to a tomographic bin to a wide SOM is straightforward. In order to construct the deep SOM we have to use our BALROG sample, consisting of all detected and selected BALROG realizations of the galaxies in the wide field, each associated with its own ‘noiseless’ replica in the deep sample. We therefore can assign to the deep SOM associated with a tomographic bin, galaxies whose BALROG wide replica is selected in that specific wide bin. Therefore, we can marginalize over deep field phenotypes c as

$$p(z|\hat{c}, \hat{b}) = \sum_c p(z|c, \hat{c}, \hat{b})p(c|\hat{c}, \hat{b}). \quad (3)$$

At this point we want to marginalize over all wide cells \hat{c} belonging to each tomographic bin. Again, we are computing $p(z|\hat{b})$ for each bin separately from different sets of SOMs:

$$p(z|\hat{b}) \approx \sum_{\hat{c}} \sum_c p(z|c, \hat{c}, \hat{b})p(c|\hat{c}, \hat{b})p(\hat{c}, \hat{b}). \quad (4)$$

Unfortunately there are very few galaxies for each (c, \hat{c}) pair, and in many cases there are none. This makes the term $p(z|c, \hat{c})$ quite difficult to estimate. However, we can reasonably assume that the $p(z)$ for galaxies assigned to a given deep cell c should not depend on the noisy wide photometry of that galaxy. Therefore, we can relax the selection:

$$p(z|\hat{b}) \approx \sum_{\hat{c}} \sum_c p(z|c, \hat{b})p(c|\hat{c}, \hat{b})p(\hat{c}, \hat{b}). \quad (5)$$

We use this approximation for our fiducial result. We obtain each of the terms appearing in equation (3.1) by placing galaxy samples to the SOM cells, as follows:

(i) $p(\hat{c})$ is computed collecting *wide* field galaxies from the MAGLIM sample into a wide field SOM (one per tomographic bin);

(ii) $p(c|\hat{c})$ is computed from the deep/BALROG sample. It consists of all detected and selected BALROG replicas of the deep galaxies injected in the wide field. We therefore can arrange the deep/BALROG sample simultaneously into a wide and deep SOMs. We call this term the *transfer function*. We weight the deep field galaxies according

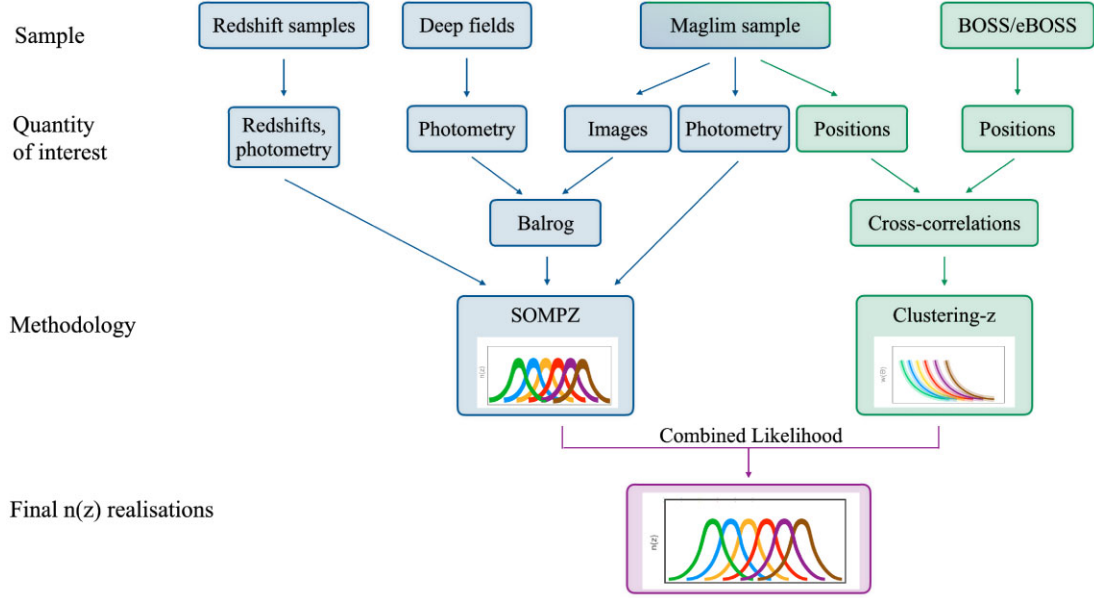


Figure 2. Flowchart illustrating the MAGLIM redshift distributions calibration scheme. The two methodologies included in the analysis are SOMPZ and clustering redshifts, which are inspired by the flowchart in Myles et al. (2021).

to their detection rate measured from BALROG. An alternative to BALROG would be using a subsection of the wide field and deep fields overlap, giving us both deep and wide photometry for a limited number of galaxies. However, the area of overlap is small and the particular observing conditions found in this area will not be representative of the overall observing conditions found in the Y3 footprint as highlighted in Myles et al. (2021).

(iii) $p(z|c)$ is computed from the *redshift* sample, which is a subset of the *deep* sample, for which we have both credible redshifts, eight-band deep photometry, and due to BALROG also wide-field realizations.

3.1.1 SOM properties

As in Buchs et al. (2019) and Myles et al. (2021), we use squared-shaped SOMs with n cells for each side (for a total of $n \times n$ cells) and periodic boundaries, which makes the visualization easier without compromising the efficiency. We parametrize the SOMs using luptitudes and lupticolours, following Buchs et al. (2019). Luptitudes are defined in Lupton, Gunn & Szalay (1999) as inverse hyperbolic sine transformation of fluxes:

$$\mu = \mu_0 - a \sinh^{-1} \frac{f}{2b} \quad \mu_0 = m_0 - 2.5 \log b, \quad (6)$$

where m are magnitudes, f are fluxes, $a = 2.5 \log b$, and b is a softening parameter that defines at which scale luptitudes transition between logarithmic and linear behaviour. For the deep SOM we compute seven lupticolours with respect to the i band

$$\mu = (\mu_1 - \mu_i, \dots, \mu_7 - \mu_i), \quad (7)$$

where the index from one to seven runs over the deep bands *urgzJHK*. We avoid using the g band for the wide field galaxies, as any observational systematics and chromatic effects are more evident in the g band. With only two lupticolours available in the wide SOM, we decided to add the i -band luptitude, as Buchs et al. (2019) find empirically that addition of the luptitude improves the training

performance:

$$\mu = (\mu_i, \mu_r - \mu_i, \mu_z - \mu_i). \quad (8)$$

The resolutions of the SOMs are 32×32 cells for the wide, and 12×12 cells for the deep. The reason behind the fewer cells in the deep SOM lies in the MAGLIM selection: the bright magnitude–redshift cuts must be applied also to the wide-component of the deep and redshift samples, and only the deep galaxies whose wide component is selected are included in the sample. This results in smaller deep and redshift samples covering a very small portion of the colour space, compared to the weak lensing source sample Myles et al. (2021). Also, reducing the number of cells means yielding more galaxies in each one. This is necessary in order to minimize the number of wide field galaxies assigned by the transfer function to a deep SOM cell with no redshift information. Reducing this number under 1 per cent is crucial to ensure that we get a correctly estimated redshift distribution for our sample. We note that shot noise caused by a small number of redshifts in a deep cell can play a significant role in biasing the estimate. We therefore performed a test to identify the optimal SOM size which would minimize these issues. We first computed several estimates in the Buzzard simulations using different resolutions for the deep SOM. We then evaluated which setting produced the smallest shift on the mean redshift with respect to the true value. As mentioned at the beginning of this section, SOMs require to be trained before being able to classify galaxies. After ensuring that the redshift samples and the MAGLIM sample span the same luptitude–lupticolour space (achieved using BALROG to obtain the redshift samples wide photometry), we decided to use the redshift sample for the deep SOM training. We instead use the MAGLIM sample itself to train the wide SOM.

3.2 WZ

Clustering redshift is a widely used method (Newman 2008; Ménard et al. 2013; Davis et al. 2017; Johnson et al. 2017; Morrison et al. 2017; Gatti et al. 2018, 2022; Scottez et al. 2018; van den Busch et al. 2020; Hildebrandt et al. 2021; Cawthon et al. 2022) to infer

or calibrate redshift distributions of galaxy samples. It relies on the assumption that the cross-correlation between two samples of objects is non-zero only in the case of overlap of the distribution of objects in physical space, due to their mutual gravitational influence.

Various implementations of the clustering redshift methodology differ in their details, but they all agree on one key aspect: the ‘target’ sample (hereafter dubbed ‘unknown’ sample), which has to be calibrated, has to be cross-correlated with a ‘reference’ sample divided into thin redshift bins. The reference sample consists of either high-quality photometric or spectroscopic redshift galaxies, and has to spatially overlap with the unknown sample.

Assuming linear galaxy-matter bias, we can express the clustering w_{ur} between the unknown sample and each of the reference sample thin bins as a function of the separation angle θ between the unknown and reference sample:

$$w_{\text{ur}}(\theta) = \int dz' n_r(z') n_u(z') b_r(z') b_u(z') w_{\text{DM}}(\theta, z') + M(\theta), \quad (9)$$

where n_r and n_u are the redshift distributions of the reference and unknown sample, b_r and b_u are the galaxy-matter biases of both samples, w_{DM} is the clustering of dark matter, and $M(\theta)$ denotes contributions due to magnification. Note that we are assuming Limber approximation (Limber 1953), but this has been shown to have no impact on the results (McQuinn & White 2013).

In our methodology, we use a single estimated value from the cross-correlation signal for each thin redshift bin. In practice, we do this by measuring the correlation function as a function of angular separation and then averaging it with a weight function to produce the single estimate:

$$\bar{w}_{\text{ur}} = \int_{\theta_{\text{min}}}^{\theta_{\text{max}}} d\theta W(\theta) w_{\text{ur}}(\theta), \quad (10)$$

where $W(\theta) \propto \theta^{-1}$ is a weighting function (Gatti et al. 2022). The integration limits in the integral in equation (10) are set to fixed physical scales (1.5 to 5 Mpc).

Since the n_r are binned in narrow bins we can approximate the number density of the sample of reference as a Dirac delta, and the revised expression becomes:

$$\bar{w}_{\text{ur}} \approx n_u b_r b_u \bar{w}_{\text{DM}} + \bar{M}. \quad (11)$$

The above equation relates the redshift distribution of the unknown sample to the measured clustering signal \bar{w}_{ur} . The galaxy-matter biases of the reference can be estimated from the autocorrelation of the reference sample. Usually the galaxy-matter bias of the unknown sample cannot be inferred and is treated as nuisance parameter. In this work, however, due to the relatively good redshift provided by DNF for the MAGLIM sample, we also use the autocorrelation of the latter as a prior for b_u (see Section 4.2). The other terms in the above equation are the clustering of dark matter \bar{w}_{DM} , which can be estimated from theory and it is not very sensitive to the cosmological parameters (Gatti et al. 2022), and the magnification term, which is expected to have a little impact (Gatti et al. 2022) and can be estimated if magnification coefficients for the samples are provided.

The angular scales considered have been chosen to span the physical interval between 1.5 and 5.0 Mpc. These bounds, applied to data as well as simulations, are selected so that the upper bound is below the range used for the galaxy clustering cosmological analyses, therefore granting the WZ likelihoods to be essentially independent of the assumed cosmology, and allowing us to produce $n(z)$ samples in an MCMC chain that runs independently of the cosmological ones. We perform the cross-correlations of MAGLIM with each of the 50 bins of width $\Delta z = 0.02$ of the BOSS/eBOSS catalogue, which

spans in the range $0.1 < z < 1.1$ as previously mentioned. We also weigh each galaxy of the MAGLIM sample by the clustering weights computed in Rodríguez-Monroy et al. (2022).

We use the Davis & Peebles (1983) estimator for the cross-correlation signal,

$$w_{\text{ur}}(\theta) = \frac{N_{\text{Rr}} D_u D_r(\theta)}{N_{\text{Dr}} D_u R_r(\theta)} - 1, \quad (12)$$

where $D_u D_r(\theta)$ and $D_u R_r(\theta)$ represent data–data and data–random pairs. The pairs are normalized through N_{Dr} and N_{Rr} , which is the total number of galaxies in the reference sample and in the reference random catalogue. The correlation estimates were computed using `treecorr`.²

4 CHARACTERIZATION OF SOURCES OF UNCERTAINTY

In this section, we present the characterization of the systematic uncertainties of our methodology. The dominant sources of uncertainties for the SOMPZ method are sample variance and shot noise. In the clustering redshift method, the main uncertainty is caused by the lack of prior knowledge on the redshift evolution of the galaxy-matter bias of the MAGLIM sample. This is modelled by a flexible systematic function, informed by a measurement of the MAGLIM autocorrelation function in data. Other minor sources of uncertainties are related to magnification effects and the approximation of linear bias (Gatti et al. 2022). We provide further details on each source of uncertainty in the following subsections. A full catalogue-to-cosmology validation of the method (in simulations) is then presented in Appendix B.

4.1 SOMPZ uncertainties

For the SOMPZ method we consider the following sources of uncertainty:

(i) *sample variance of the deep fields*: main uncertainty, caused by the limited area of the deep fields. We model the effect of sample variance by means of the *three step Dirichlet* (3sDir) analytical model described in Section 4.1.1;

(ii) *shot noise in the deep and redshift samples*: this is induced by the limited number of galaxies available in the deep and redshift samples. We model the effect of shot noise by means of the 3sDir analytical model described in Section 4.1.1;

(iii) *SOMPZ method uncertainty*: this uncertainty stems from discretizing the colour space in the SOMPZ mapping. We do estimate its impact on the SOMPZ estimates by replicating the SOMPZ methods multiple times in simulations, and incorporate its effects using probability integral transforms (PITs) (Section 4.1.2);

(iv) *photometric calibration*: related to uncertainties in the calibration of the deep fields zero-point, it is accounted for in the SOMPZ estimates by means of PITs (Section 4.1.3).

(v) *redshift sample biases*: these biases stem from uncertainties and biases in the redshift estimates of the redshift samples. Their impact is accounted for in our methodology by marginalizing over three different combinations of redshift samples (Section 4.1.4);

(vi) *transfer function*: any bias induced by an erroneous estimation of the transfer function due to a size-limited BALROG sample; we anticipate this to be negligible following the results from Myles et al. (2021) (Section 4.1.5).

²<https://github.com/rmjarvis/TreeCorr>

In the following sections we will proceed to describe in detail how we account for each of the items listed above.

4.1.1 Sample variance and shot noise (3sDir)

Sample variance is the dominant uncertainty affecting our SOMPZ estimates, and stems from the limited size and area coverage of the redshift and deep samples, with respect to the whole wide field. The deep fields only cover $\sim 9 \text{ deg}^2$, which means we could be learning the colour/redshift relation from a non-representative sample of the sky due to fluctuations in the matter density field; moreover, the finite size of the redshift sample can introduce shot noise effects, preventing a correct sampling of the quantities required for the redshift inference.

Generally the impact of sample variance can be evaluated estimating the redshift distributions in simulations multiple times using different line of sights for the deep fields (e.g. Hildebrandt et al. 2017, 2021; Hoyle et al. 2018; Buchs et al. 2019; Wright et al. 2020). Although we also performed a test where we evaluated the impact of sample variance using the Buzzard simulation, in our standard procedure we use the *three step Dirichlet* (3sDir) approach 3sDir presented in Sánchez et al. (2020) and applied to the redshift calibration of the DES Year 3 source sample (Myles et al. 2021).

The 3sDir method consists of an analytical sample variance model predicting what the redshift-colour distribution would be from the observed individual redshift and galaxy phenotypes (colours) of galaxies coming from smaller deep fields. Using this model we can build an ensemble of redshift distributions realizations whose fluctuations realistically represent the effect of sample variance. During the cosmological inference, by sampling over these realizations, one can effectively marginalize over the effect of sample variance. Here, we provide a short description of the 3sDir method, but we direct the reader to Myles et al. (2021) and Sánchez et al. (2020) for more details. The 3sDir method assumes the probability $p(z, c)$ that galaxies belong to a redshift bin z and colour phenotype c to be described by a probability histogram with coefficients f_{zc} (with $\sum f_{zc} = 1$ and $0 \leq f_{zc} \leq 1$). Under this assumption, the expected number counts of galaxies in a deep SOM cell given the coefficients f_{zc} are described by a multinomial distribution; if we assume a Dirichlet function for the prior on f_{zc} , the posterior of f_{zc} given the observed number count will also be described by a Dirichlet function. Such a Dirichlet posterior can be used to draw samples and naturally accounts for the effect of shot noise in the data. The effect of sample variance can be introduced by tuning the width of the prior on f_{zc} , which does not change the expected value for f_{zc} in the Dirichlet distribution, but does change its variance to simultaneously account for shot noise and sample variance.

If all the galaxies belonging to the redshift sample were independently drawn, then a Dirichlet distribution parametrized by the redshift sample counts in each couple of redshift bin z and phenotype c , N_{zc} , would fully characterize f_{zc} . However, one subtlety is that sample variance correlates with redshifts; to increase the variance with the correct redshift dependence one can use the fact that two different phenotypes (deep SOM cells) overlapping in redshift are correlated due to the same underlying large-scale structure fluctuations. The 3sDir model assumes that phenotypes at the same redshift share the same sample variance, and therefore groups cells with similar redshifts in *superphenotypes* T . One can then express the f_{zc} as

$$f_{zc} = \sum f_c^{zT} f_z^T f_T. \quad (13)$$

The 3sDir method consists of drawing values of these three sets of

coefficients with three Dirichlet functions. In this way, it is possible to include a redshift-dependent variance while conserving the expected value of f_{zc} .

The validation of the 3sDir method has been carried out in Myles et al. (2021), applied to the weak lensing source sample. The only difference with this work stands in the fact we are performing the 3sDir estimation independently for each tomographic bin, due to their definition.

As reported in Table 4, this uncertainty is dominant, both on the mean and width values of the $n(z)$ distributions, computed from the ensemble of realizations provided by the 3sDir method.

4.1.2 SOMPZ method uncertainty

The SOMPZ method relies on the discretization on the colour space spanned by our deep field sample, and this is an approximation that can lead to small biases or additional uncertainties. In order to estimate these, we compute our SOMPZ $n(z)$ a large number of times in the Buzzard simulations. In order to factor out sample variance, each time we randomly select patches of the Buzzard footprint to construct the mock deep fields. In this way, by averaging over all the final $n(z)$ realizations, we can produce an estimate of the $n(z)$ only minimally biased by sample variance, and test the agreement with the true $n(z)$. Due to the computational cost of the SOMPZ pipeline, we decided to produce 300 $n(z)$ replicas. To perform this test, we assumed that the redshift sample would only be limited to one of our four fields, of the size of COSMOS.

We computed the mean redshift offset of the ensemble with respect to the true value, for each tomographic bin. As reported in Table 4, these values are smaller than the effect of sample variance. These values are incorporated into our final $n(z)$ ensemble using the PIT method described in the following section, by additionally shifting each PIT (used to correct for the zero-point uncertainties) by a value drawn from a Gaussian centred at zero with standard deviation equal to the root-mean-square of the aforementioned mean offset values.

4.1.3 Deep fields photometric calibration uncertainty

Although the uncertainty in the photometry of each individual galaxy is implicitly accounted for in the SOM training, the uncertainty on the photometric calibrations as a whole must be evaluated by testing how the measured $n(z)$ are affected by changes in the photometric zero-point in each band. This is relevant for the deep fields, where the relatively precise fluxes are key to constraining reliable $p(z)$ in parts of parameter space that are not subject to selection biases. Ideally, this would be tested by rerunning the full analysis for an ensemble of perturbations of the photometric zero-point according to the zero-point uncertainty, but the computational requirements of the BALROG injection procedure make this infeasible. Instead, we produce an analogous ensemble of realizations in simulations, where the BALROG mock photometric survey is reduced to a computationally simpler procedure of adding Gaussian noise to true magnitudes. For each realization of this ensemble, we perturb all deep field magnitudes by a draw from a Gaussian whose width is determined by the photometric zero-point uncertainty in the Y3 deep fields catalogue in a specified band, as computed in Hartley et al. (2022). We then ‘inject’ these perturbed deep field fluxes with a mock BALROG procedure to generate wide field realizations of the galaxies and measure the corresponding $n(z)$. In this way we generate a full ensemble of $n(z)$ realizations reflecting the uncertainty in our redshift calibration due to the photometric calibration. We apply

Table 4. Summary of values for systematic uncertainties and centre values for mean (top panel) and width (bottom panel) for the $n(z)$ distributions. The various components are computed as described in Section 4 and as they are not completely independent it is not expected that they sum up to the total value. The values related to SOMPZ and SOMPZ+WZ refer to Fig. 5, and include only the 3sDir uncertainty due to sample variance and shot noise (and the redshift samples uncertainty), because it was logistically not possible to add the SOMPZ method and the zero-point sources of uncertainty before the combination with WZ. As a comparison, the ‘SOMPZ (with all unc)’ includes all uncertainties. The final $n(z)$ which has been used in the cosmological analysis is the bottom line.

Uncertainty	Bin 1	Bin 2	Mean	Bin 4	Bin 5	Bin 6
	$z \in [0.2, 0.4]$	$z \in [0.4, 0.55]$	Bin 3 $z \in [0.55, 0.7]$	$z \in [0.7, 0.85]$	$z \in [0.85, 0.95]$	$z \in [0.95, 1.05]$
Sample Variance and shot noise	0.015	0.010	0.010	0.008	0.009	0.009
SOMPZ method	0.004	0.003	0.005	0.001	0.007	0.005
Redshift samples	0.009	0.001	0.006	0.003	0.004	0.007
Zero-point	0.008	0.007	0.004	0.005	0.005	0.005
SOMPZ	0.315 ± 0.015	0.445 ± 0.010	0.630 ± 0.010	0.776 ± 0.008	0.895 ± 0.009	0.983 ± 0.012
SOMPZ+WZ	0.316 ± 0.014	0.456 ± 0.008	0.632 ± 0.008	0.780 ± 0.007	0.893 ± 0.008	0.985 ± 0.010
SOMPZ (with all unc)	0.317 ± 0.020	0.447 ± 0.012	0.634 ± 0.013	0.778 ± 0.010	0.897 ± 0.011	0.988 ± 0.013
SOMPZ+WZ (with all unc)	0.315 ± 0.016	0.463 ± 0.010	0.633 ± 0.009	0.781 ± 0.008	0.893 ± 0.009	0.990 ± 0.012
	Width					
Sample variance and shot noise	0.007	0.005	0.003	0.003	0.004	0.009
SOMPZ method	0.003	0.003	0.0007	0.0003	0.002	0.0001
Redshift samples	0.001	0.005	0.0007	0.0006	0.0003	0.001
Zero-point	0.003	0.004	0.001	0.0004	0.001	0.001
SOMPZ	0.077 ± 0.007	0.093 ± 0.007	0.065 ± 0.004	0.081 ± 0.004	0.071 ± 0.004	0.096 ± 0.009
SOMPZ+WZ	0.080 ± 0.004	0.089 ± 0.004	0.060 ± 0.002	0.077 ± 0.003	0.074 ± 0.004	0.105 ± 0.006
SOMPZ (with all unc)	0.081 ± 0.008	0.096 ± 0.007	0.067 ± 0.005	0.081 ± 0.004	0.073 ± 0.005	0.098 ± 0.009
SOMPZ+WZ (with all unc)	0.080 ± 0.005	0.081 ± 0.005	0.060 ± 0.002	0.073 ± 0.003	0.074 ± 0.004	0.101 ± 0.007

PITs as in Myles et al. (2021) to transfer the variation encoded in the ensemble from simulated $n(z)$ to our fiducial data result. Essentially, this process involves calculating the inverse cumulative distribution function (CDF) for each simulated realization $n_i(z)$ in the ensemble. The PIT is then obtained by computing the difference between the CDF of each realization and the average CDF of the entire ensemble. To apply these transformations to the data, the PIT value is added to the inverse CDF of the fiducial data $n(z)$. The PIT resulting from a single draw of zero-point offsets is determined and collectively applied to all tomographic bins. More details on this new implementation of the PIT can be found in Myles et al. (2023).

4.1.4 Redshift sample uncertainty

As mentioned in Section 2.5, we decided to choose three different catalogues to infer our redshift distributions from: a collection of spectroscopic surveys galaxies (Gschwend et al. 2018), PAU + COSMOS redshift as in Alarcon et al. (2021), and COSMOS30 photometric redshifts (Laigle et al. 2016). The reason for availing ourselves of more than one catalogue lies in the fact neither of these are exempt from systematic uncertainties: each survey uses different photometry, different model assumptions, and can be affected systematically by selection effects, incorrect templates, photometric outliers, etc. Since there is a considerable overlap in the number of galaxies belonging to more than one of the redshift catalogues selected for this work, to account for the intrinsic biases we decided to build three samples which are combinations of the aforementioned catalogues. We ranked the redshift catalogues differently for each sample: if a galaxy has information from multiple origins, we assign the redshift from the highest ranked catalogue. The three redshift samples, *SPC*, *PC*, *SC*, are described in Section 2.5.

For each of these, we will perform the complete pipeline, and the final set of realization will be constructed by an equal fraction $p(R) = 1/3$ from each survey. By placing equal prior probability to each sample, this is equivalent as saying that we do not believe any of the samples is more likely to be correct. But note that for galaxies from which we have information from only one catalogue, we are assuming that information to be true, and this is a caveat of this approach. Fig. 3 shows the uncertainty in the mean redshift of each tomographic bin, for each of the redshift samples and for their combination.

4.1.5 Transfer function uncertainty

One of the key points in this redshift calibration is the transfer function $p(c|\hat{c})$, the intermediate step necessary to assign redshifts from deep field galaxies to the whole wide field. If the transfer function is inaccurate, regardless of how precise the colour/redshift characterization is in the deep SOM, it can bias the final $n(z)$ distributions. $p(c|\hat{c})$ depends on the observation conditions in that location, determining if the galaxy is detected or not. Observing conditions vary across the wide field, but for our analysis we are interested in redshift distributions estimated across all the footprint. BALROG injects the same deep galaxies in random wide tiles, and despite these covering only around ~ 20 per cent of the DES footprint, in Myles et al. (2021) it was verified that BALROG is adequately sampling the observing conditions in the wide field. They bootstrapped the sample by the injected position and recomputed 1000 different transfer functions. They concluded that the dispersion in the final $n(z)$ mean redshift from repeating the analysis using each time a different transfer function was completely negligible. Here, we repeated that test, since our deep field sample has less galaxies and might impact differently the transfer function. We found that this is

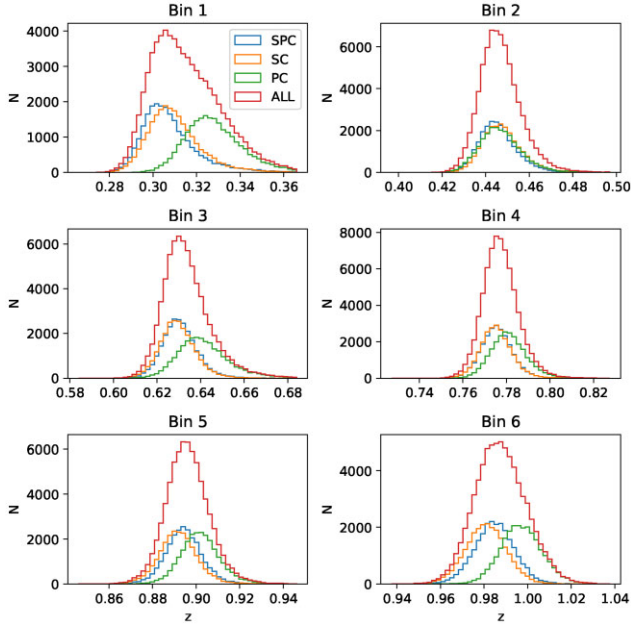


Figure 3. Uncertainty on the mean redshift represented by the number counts of the three redshift samples: SPC (prioritizes spectra, then PAU photo-z, then COSMOS30), PC (prioritizes PAU photo-z, then COSMOS30), and SC (prioritizes spectra, then COSMOS30). In red the total uncertainty given by their combination is shown.

also negligible for our case, with variations on the $n(z)$ mean $< 10^{-3}$ in each tomographic bin, and therefore decided not to propagate this in the final $n(z)$ estimate.

4.2 WZ uncertainties

The WZ systematic uncertainties have been identified and characterized in detail for the WL sample in Gatti et al. (2022). Namely, the systematic budget was found to be dominated by our lack of prior knowledge of the redshift evolution of the galaxy-matter bias of the unknown sample. This is also expected to be the case for the MAGLIM sample, although the amplitude of the effect might differ from the WL sample (ideally, since the MAGLIM redshift distributions are narrower, we might expect a smaller impact due to systematics slowly varying with redshift like the galaxy-matter bias of the unknown sample).

Similarly to Gatti et al. (2022), we model our systematics by means of a flexible function, $\text{Sys}(\mathbf{s})$, which mostly captures the redshift evolution of the galaxy-matter of the unknown sample. The $\text{Sys}(\mathbf{s})$ function is parametrized by $\mathbf{s} = \{s_1, s_2, \dots\}$ that we will marginalize over and is given by

$$\log[\text{Sys}(z, \mathbf{s})] = \sum_{k=0}^M \frac{\sqrt{2k+1}}{0.85} s_k P_k(u), \quad (14)$$

$$u \equiv 0.85 \frac{z - 0.5(z_{\max} + z_{\min})}{(z_{\max} - z_{\min})/2}, \quad (15)$$

with $P_k(z_i)$ being the k -th Legendre polynomial and $M = 6$ is the maximum order. In this work, we set the prior $p(\mathbf{s})$ to be a simple diagonal normal distribution, with the standard deviations $\{\sigma_{s_0}, \dots, \sigma_{s_M}\}$ and means informed by the measured autocorrelation of the MAGLIM sample.

In Gatti et al. (2022), such a systematic function was let to vary by the typical amplitude of the redshift evolution of the galaxy-matter

bias of the WL sample we measured in simulations. In practice, this was achieved by imposing a Gaussian prior with zero mean $p(\mathbf{s})$ on the coefficients \mathbf{s} of the systematic function.

In the case of the MAGLIM sample, we can use a more informative prior $p(\mathbf{s})$ that uses the information we have from the data about the galaxy-matter bias evolution of the sample. In particular, we rely on the fact that the MAGLIM sample has good per-galaxy redshift estimates, which allows us to divide the sample in relatively small bins and measure the autocorrelation of such bins. This was not possible for WL sample, due to the poor per-galaxy redshift accuracy.

To this aim, we use DNF one-point estimates z_{mean} to further divide the MAGLIM sample in bins of width of $\Delta_z = 0.02$, and we measure the autocorrelation of each bin. We note that the true width of each bin will be much larger than $\Delta_z = 0.02$, as the DNF photo- z are uncertain. Under the approximation of negligible redshift evolution of the galaxy-matter bias of the MAGLIM sample over each thin bin, the measured autocorrelation can be related to the galaxy-matter bias by knowing how broad the true $n(z)$ distribution of each bin is (Gatti et al. 2018; Cawthon et al. 2022):

$$w_{uu}(z_i) = b_u^2(z_i) w_{\text{DM}}(z_i) \int dz' n_{u,i}^2(z'), \quad (16)$$

where $n_{u,i}(z')$ is indeed the true distribution of the thin bin MAGLIM sample. Such a quantity is estimated using the PDF estimate from DNF z_{PDF} .

From this measurement performed in data we can then retrieve the galaxy bias $b_u(z)$ by inverting equation (16). We fit the $\text{Sys}(\mathbf{s})$ function presented in equation (14) to the measured $b_u(z)$ and obtain the best-fitting \mathbf{s} values, which we show in Fig. 4. These best-fitting coefficients are then used as the mean value of the Gaussian prior $p(\mathbf{s})$. The best-fitting $\text{Sys}(\mathbf{s})$ function to the data is shown in the right panel of Fig. 4.

To estimate the width of the prior $p(\mathbf{s})$ we took a different approach. First, we estimate the bias evolution in simulations by dividing galaxies into thin redshift bins using (i) the true redshifts from the simulation; and (ii) the photo- z estimated from the DNF code. When dividing the galaxies with the photo- z from DNF, we further correct the measured autocorrelation using equation (16). These measurements are shown in the left panel of Fig. 4. The discrepancy between the measured bias evolution from photo- z (equivalent to the application with real data) relative to the measured bias evolution with true redshifts (equivalent to the truth) is a systematic bias. We use the sum in quadrature of this difference with the statistical uncertainty of the bias measurement as the prior width of s_0 . For the higher order parameters we estimate the standard deviation of the prior by summing in quadrature the ratio between the two biases and the statistical uncertainty from the bias measurement in data. This allows to best capture the RMS variations of the bias function itself. As can be seen in Fig. 4, the 68 percent confidence interval spanned by the $\text{Sys}(\mathbf{s})$ function both brackets the ideal and real world measurements. The values for the mean and width of the prior are displayed in Table 5. Both the width of the prior on the 0-th and higher order coefficients are much tighter than in Myles et al. (2021), where $s_0 = 0.6$ and $s_{1,4} = 0.15$. As already explained, the difference lies in the initial accuracy of the photo- z estimates, which enables the measurement of the autocorrelation of the galaxy sample in thin redshift bins. For the weak lensing source sample such information was not available, and therefore a more conservative prior was deemed appropriate. In the MAGLIM sample case instead, the greater accuracy on its photo- z allows to extract more information from the autocorrelation.

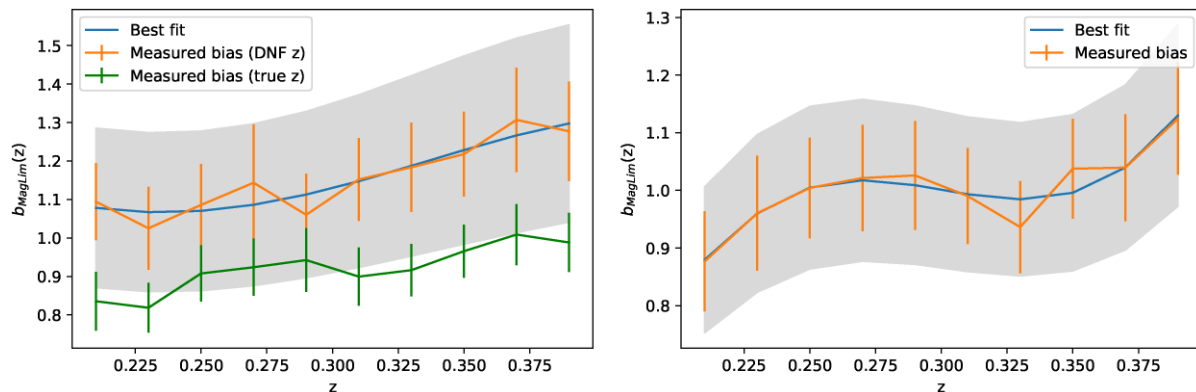


Figure 4. *Left panel:* galaxy-matter bias of Bin 1 or the MAGLIM sample ($0.2 < z < 0.4$) as estimated in simulation following the methodology outlined in Section 4.2. The green points are obtained by dividing the sample into thin bins using the true redshifts, while the orange ones are obtained by binning the sample using the DNF redshift estimates. The grey band encompasses the 68 per cent confidence interval of the Sys(s) function. *Right panel:* galaxy-matter bias of Bin 1 of the MAGLIM sample ($0.2 < z < 0.4$) as measured from the data (orange points); the blue line shows the best-fitting Sys(s) function, and the grey band encompasses its 68 per cent confidence interval.

Table 5. Means and widths of the Gaussian prior function $p(\mathbf{s})$ appearing in equation (18).

	Mean					
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
$\langle s_0 \rangle$	-0.028	-0.085	-0.319	-2.630	-0.119	-2.249
$\langle s_1 \rangle$	0.186	0.559	0.007	1.161	-1.660	0.819
$\langle s_2 \rangle$	0.046	0.139	-0.120	0.202	0.134	0.033
$\langle s_3 \rangle$	0.035	0.105	-0.130	0.314	0.293	0.174
$\langle s_4 \rangle$	0.037	0.111	-0.112	-0.197	0.211	0.279
$\langle s_5 \rangle$	-0.062	-0.189	-0.203	-0.210	1.408	0.569
	Width					
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
σ_{s_0}	0.107	0.216	0.123	0.072	0.067	0.198
$\sigma_{s_{1..5}}$	0.029	0.053	0.041	0.052	0.081	0.044

Last, we mention that an additional source of uncertainties for the WZ measurement is related to the impact of magnification. We do model magnification effects, but the accuracy of that model is limited by our knowledge of the magnification coefficients for the two samples. In particular, we do not have any prior knowledge of such coefficients for the BOSS/eBOSS sample. Those coefficients are set to 0 for our fiducial analysis (on the contrary, estimates for the magnification coefficient of the MAGLIM sample are available). We expect magnification to have a small impact, based on tests performed in Gatti et al. (2022), but we none the less test in the following section the impact of having a non-null magnification coefficient for the BOSS/eBOSS sample.

4.3 Combination of SOMPZ and WZ

In order to combine SOMPZ and WZ constraints, we follow Gatti et al. (2022) and write the clustering likelihood by forward modelling the full clustering signal as a function of the SOMPZ redshift distributions estimates $n(z)_{pz}$. Moreover, we include the systematic function Sys(s) introduced in the previous section, which describes the uncertainties on the WZ measurement, mostly driven by the lack

of knowledge of b_u and its redshift dependence:

$$\hat{w}_{ur}(z_i) = n(z)_{pz}(z_i)b_r(z_i)w_{DM}(z_i) \times Sys(z_i, \mathbf{s}) + M(\alpha_u, \alpha_r, b_u, n(z)_{pz}). \quad (17)$$

In the above equation, the quantities $\alpha_u(z_i)$ and $\alpha_r(z_i)$ are the magnification coefficients for the unknown and reference samples. See Gatti et al. (2022) for full description of the magnification term M . The clustering of dark matter $w_{DM}(z_i)$ is estimated from theory assuming fixed cosmology. We tested that varying cosmology has a negligible impact on our methodology.

The likelihood of the WZ data conditioned on the target $n(z)$ and all the systematic parameters reads as

$$\mathcal{L}[\text{WZ}|n_u(z), b_r(z), \alpha_r(z), w_{DM}(z)] \propto \int d\mathbf{s} d\mathbf{p} \exp\left[-\frac{1}{2}(w_{ur} - \hat{w}_{ur})^T \Sigma_w^{-1}(w_{ur} - \hat{w}_{ur})\right] p(\mathbf{s})p(\mathbf{p}), \quad (18)$$

where Σ_w is the clustering covariance, estimated through jackknife, and $\mathbf{p} = b_u, \alpha_u$. We implemented a Hamiltonian Monte Carlo sampler (HMC) that simultaneously samples the SOMPZ and WZ likelihood. The HMC does directly take as input the SOMs output of the sample variance estimation (described in 4.1.1), and it perturbs selectively the number counts in the SOMs in such a way to produce realizations that are already more likely to match the clustering redshift data.

5 RESULTS IN DATA

In this section, we present the final redshift distributions for the MAGLIM sample as obtained in data. We also compare the SOMPZ+WZ redshift distributions with the fiducial DNF+WZ estimates used for the same sample and adopted in the cosmological analysis presented in Porredon et al. (2022). A complete validation of the method in simulations is presented in Appendix B.

We first compare in Fig. 5 the redshift estimates obtained using the 3sDir method and the estimates obtained including the WZ information as described in Section 4. Due to logistics, the combination of the two methods was performed before incorporating the SOMPZ and zero-point errors. As here we are just displaying the effect of the combination, we are showing only how the 3sDir uncertainty from sample variance and shot noise (from the three redshift samples) varies once we add the information from WZ. The combination

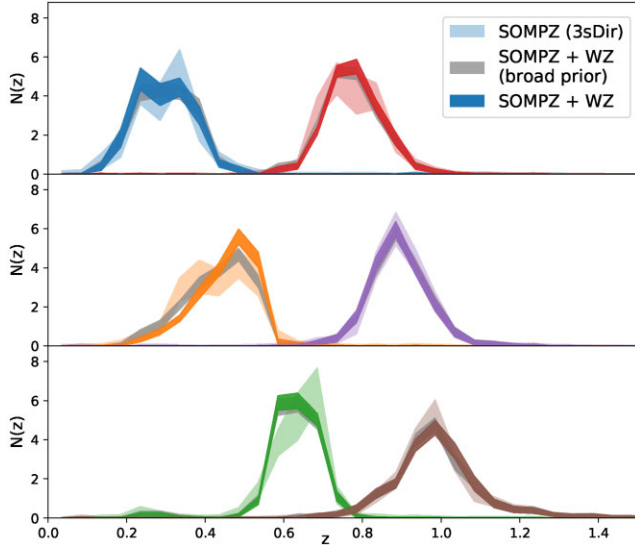


Figure 5. 3sDir distributions before (lighter shades) and after the combination with clustering-z (solid shades), and after the combination with clustering-z but using a broader prior on the parameters of the galaxy-matter bias function $\text{Sys}(s)$ (the same values of the width of the prior $p(s)$ that were used in Gatti et al. 2022). In the top row we have bins 1 and 4, in the middle row bins 2 and 5, and in the bottom row bins 3 and 6. The bands represent the 1σ error from the central value. Note how the combination with WZ tightens the constraint on the shape of the $n(z)$.

of the two methods results in stronger constraints on the shape of the $n(z)$, due to the complementarity in the information provided by each SOMPOZ and WZ. Particularly, the WZ signal strongly correlates across adjacent bins, excluding large portions of possible $n(z)$ shapes allowed by the SOMPOZ likelihood alone, which are affected by sample variance fluctuations from the small calibration fields, and resulting in a smoother distribution. The improvement on the uncertainty on the mean is more modest, but not null, as reported in Table 4. Usually, WZ data provide limited information on the mean redshift, especially compared to SOMPOZ, as the systematic uncertainty on the galaxy bias evolution of the target sample is large and directly degenerate with the mean redshift, as is the case in Gatti et al. (2022). However, in this work we have included a tighter prior on the $\text{Sys}(s)$ function describing the galaxy bias evolution uncertainty by measuring it directly from the MAGLIM autocorrelation function. The addition of the WZ information has a modest impact on the values of the mean and width of the redshift distributions, at most at the 1σ level (see Table 4); this is somewhat expected, as the WZ and SOMPOZ information is independent, but consistent with each other.

5.1 Comparison with DNF

We find it interesting to compare the final SOMPOZ+WZ redshift distributions with the fiducial ones used for DES Y3, obtained using DNF photometric estimates and clustering constraints (hereafter DNF+WZ). Since the two sets of distributions have been obtained with two different methods, we also briefly discuss the major differences between the two pipelines. The DNF code presented in Section 2.2.1 produces per-galaxy redshift estimates; these are stacked to produce the redshift distributions for the lens samples. Then, following Cawthon et al. (2022), a clustering redshift measurement is performed, using BOSS/eBOSS galaxies as reference sample, similarly to this work. The DNF $n(z)$ are matched to the WZ-estimated $n(z)$ through a chi-square fitting; in particular, the DNF $n(z)$

are allowed to shift and stretch to improve the χ^2 . The maximum a posteriori values of the shift and stretch and related uncertainties obtained through this matching procedure are used as a prior for the DNF $n(z)$ shift and stretch used in the cosmological inference.

Despite the DNF+WZ and SOMPOZ+WZ methods using the same photometric and clustering measurements, the methodologies differ in a number of aspects:

(i) **SOMPOZ versus DNF uncertainties:** SOMPOZ and DNF are both machine learning methods, but they are substantially different in spirit and implementation. DNF is a traditional supervised machine learning code where the likelihood (directional neighbourhood) between wide field magnitudes/colours and redshift is learned from training with a subsample of galaxies with both reliable redshift information and measured wide field photometry. On the other hand, in SOMPOZ machine learning is only used in an unsupervised fashion (without knowledge of redshift), to group self-similar parts of wide field magnitude/colour space together. Then, these groups (wide SOM cells) are probabilistically related using Bayes theorem to the colour-redshift relation measured empirically in the calibration deep fields, where much better information is available. The likelihood between each set of wide and deep field photometry is also measured empirically by injecting galaxies of the latter into images of the former. Furthermore, SOMPOZ provides a comprehensive list of statistical as well as systematic uncertainties affecting the calibration samples which are rigorously propagated through the $n(z)$. On the other hand, DNF only describes statistical uncertainties related to the residual differences to the closest training neighbours to the fitted hyperplane of the target galaxies.

(ii) **Combination:** The clustering information is included and combined with the photometric estimates in a substantially different way. In this work, SOMPOZ and WZ are combined by sampling from the joint posterior using the HMC method. No approximation is performed when combining the two likelihoods. On the other hand, matching DNF $n(z)$ to the WZ measurements it has been implicitly assumed that the DNF $n(z)$ estimates can only be biased at the level of their mean and width, and that inaccuracies in the higher order moments of the $n(z)$ can be neglected (or do not affect the matching procedure with the WZ measurements). However, if the DNF and WZ $n(z)$ estimates are substantially different beyond their first two moments, the matching might cause biases (Gatti et al. 2018) also in the first and second moments. Furthermore, in the combination of the fiducial method, the DNF shape is only allowed to be modified by shifting and stretching it. Therefore, the shift and stretch parameters are centred at the WZ values. This means that the photo- z priors for the cosmological inference only carry uncertainty from the WZ measurement, as this method does not propagate any systematic uncertainties related to uncertainty from the accuracy of DNF or the quality of its training sample photometry. In comparison, SOMPOZ+WZ properly combines the statistical significance from SOMPOZ and WZ yielding a final uncertainty that truly combines the information from each of them separately. Finally, the SOMPOZ+WZ $n(z)$ samples also capture the uncertainties in the higher moments of the redshift distributions, whereas the DNF+WZ uncertainties are only relative to the mean and width.

(iii) **WZ distribution tails:** The WZ measurements used to calibrate the DNF $n(z)$ have clipped tails, since the measurements were performed in a restricted redshift window to avoid biases related to un-modelled magnification effects in the tails of the redshift distribution. On the other hand, in this work, when combining the clustering information with SOMPOZ estimates, we use the WZ measurements over all the redshift range, since we also marginalize over magnification effects.

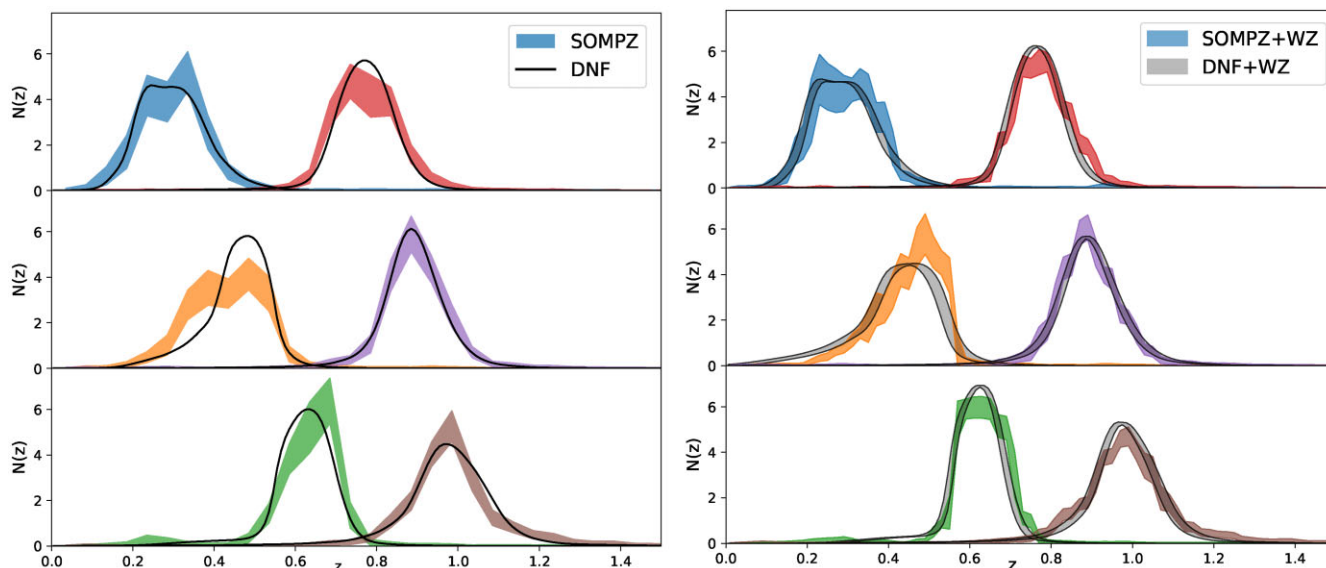


Figure 6. Left panel: final $n(z)$ realizations obtained from the SOMPZ methodology alone compared to the fiducial DNF distribution for MAGLIM (in black). Right panel: final $n(z)$ realizations obtained from both SOMPZ and WZ methodology compared to the fiducial DNF distribution for MAGLIM (grey bands) after shifting and stretching them to fit WZ measurement. Since in the inference the shift and stretch values are marginalized over, the uncertainties of the gray bands are obtained by sampling over the allowed ranges of shift and stretch defined by the prior, and applied, respectively, to the DNF estimate. Note that for a fairer comparison of the methods, the two remaining uncertainties were applied to the SOMPZ ensemble (zero-point and SOMPZ intrinsic), to include all the SOMPZ-related uncertainties. For both plots, in the top row we have bins 1 and 4, in the middle row bins 2 and 5, and in the bottom row bins 3 and 6.

Table 6. Values of mean and width of the SOMPZ+WZ final ensemble of distributions and the DNF estimate. The statistical difference $\Delta_{\langle z \rangle}$ is computed by considering the uncertainties of both methods summed in quadrature, as in $\Delta_{\langle z \rangle} = (\langle z \rangle_{\text{SOMPZ}} - \langle z \rangle_{\text{DNF}}) / \sqrt{\sigma(\langle z \rangle_{\text{SOMPZ}})^2 + \sigma(\langle z \rangle_{\text{DNF}})^2}$. We refer to these as lower limits. Because the WZ measurement is very similar in the two cases, and the uncertainties summed in quadrature are correlated and therefore we are likely underestimating $\Delta_{\langle z \rangle}$.

		Bin 1 $z \in [0.2, 0.4]$	Bin 2 $z \in [0.4, 0.55]$	Bin 3 $z \in [0.55, 0.7]$	Bin 4 $z \in [0.7, 0.85]$	Bin 5 $z \in [0.85, 0.95]$	Bin 6 $z \in [0.95, 1.05]$
$\langle z \rangle$	SOMPZ+WZ	0.315 ± 0.016	0.463 ± 0.010	0.633 ± 0.009	0.781 ± 0.008	0.893 ± 0.009	0.990 ± 0.012
	DNF+WZ	0.292 ± 0.007	0.422 ± 0.011	0.616 ± 0.006	0.762 ± 0.006	0.887 ± 0.007	0.969 ± 0.008
	$\Delta_{\langle z \rangle}$	1.3	2.7	1.7	1.9	0.5	1.5
σ_z	SOMPZ+WZ	0.080 ± 0.005	0.081 ± 0.005	0.060 ± 0.002	0.073 ± 0.003	0.074 ± 0.004	0.102 ± 0.007
	DNF+WZ	0.078 ± 0.005	0.094 ± 0.007	0.055 ± 0.003	0.062 ± 0.003	0.075 ± 0.004	0.080 ± 0.007
	Δ_{σ_z}	0.2	1.6	1.3	2.2	0.3	2.3

(iv) **WZ galaxy-matter bias:** The WZ measurements used in the DNF+WZ estimates are corrected for the redshift evolution of the galaxy-matter bias of the MAGLIM sample computed from autocorrelations measurements following equation (16) (Cawthon et al. 2022). As for this work we use the forward modelling approach described in Section 4.2, we instead do not correct directly for the bias, but from the MAGLIM autocorrelations we determine prior values of the parameters of our Sys(s), and then marginalize over possible bias functions in the sampling from the joint likelihood. We are therefore assuming an uncertainty on the galaxy-matter bias and validating the central value using SOMPZ data.

We must highlight that in Porredon et al. (2022) and Cawthon et al. (2022) several tests were performed to test the robustness of the DNF+WZ method. In particular, Cawthon et al. (2022) tested the performance of the clustering measurements in simulations, whereas Porredon et al. (2022) tested that matching DNF $n(z)$ to the WZ measurements was not introducing biases in the cosmological constraints, and that modelling only the uncertainties in the mean and width of the distributions was sufficient for the DES Y3 cosmological

analysis. These tests should cover potential worries raised in points (ii), (iii), and (iv) above for the DNF+WZ method. Having said this, any discrepancy between the SOMPZ+WZ $n(z)$ and the DNF+WZ $n(z)$ should boil down to the points listed above.

In Fig. 6, the shapes and uncertainties of the two methodologies are compared, before and after the inclusion of WZ information, respectively, in the left and right panel. Visually the DNF+WZ $n(z)$ look very similar to the SOMPZ+WZ ones, although some discrepancies can be noticed (e.g. in the second bin). We report in Table 6 the redshift means and widths of the two sets of distributions, and their agreement. The means and widths are also visually compared in Fig. 7. The agreement is computed assuming the uncertainties of the two methods to be uncorrelated, which is likely not true; therefore, the reported agreements are optimistic. Computing the level of correlation between the two redshift estimates is not trivial. The DNF+WZ estimates and uncertainties are driven only by the WZ measurements in the range where WZ measurements are available and magnification effects are negligible; the tails of the distribution, on the other hand, are described by the DNF

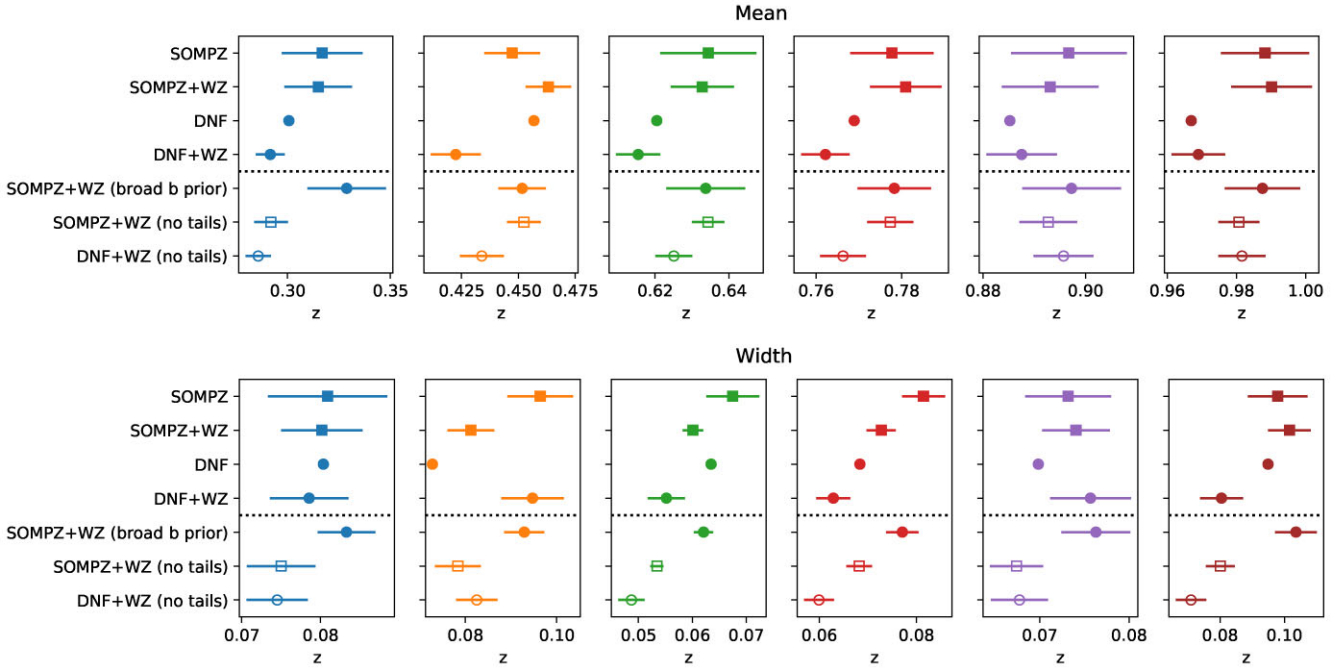


Figure 7. Visual representation of the uncertainties on mean (above) and width (below) of the redshift distributions estimated using the SOMPZ (square markers) and DNF (round markers) methods, before and after including the WZ information, for each tomographic bin. Below the dashed line is the comparison of the values computed in the redshift range used for the χ^2 fit of the DNF estimate with the smoothed WZ $n(z)$.

estimates. The SOMPZ+WZ estimates receive contributions from both SOMPZ and WZ; if the SOMPZ method was to completely drive our estimates, then the SOMPZ+WZ and the DNF+WZ estimates could be assumed to be independent. This is likely the case for the mean redshift estimates, as we have seen that WZ is not particularly constraining on the mean redshift (see Fig. 7). The width estimates are inferred more by the WZ measurements, and this might indicate that our tensions are underestimated, because we know that the two calibration methods share part of the WZ information. With this in mind, large tensions between means/widths of the two methods might indicate that either the DNF+WZ uncertainties are underestimated, or there are some real differences between the two methods (one or both are biased). The reported values in Table 6 does not point to dramatic differences between the two methods: the most extreme statistical distance is 2.7σ between means of Bin 2, and 2.3σ between widths of Bin 6.

From Table 6 we note that SOMPZ+WZ uncertainties on the mean are larger than the DNF+WZ ones, while uncertainties on the widths are comparable. This is due to the fact that the uncertainties in the mean redshifts for the SOMPZ estimates are very sensitive to contributions from outliers at high redshift. The DNF+WZ mean redshift estimates (and uncertainties), on the other hand, are driven by the match with the WZ measurements with clipped tails, i.e. they do not take into account uncertainties in the tails, and are therefore smaller. The fact that the modelling of the tails is different between the two methodologies is also responsible for the slightly higher mean redshifts of the SOMPZ+WZ estimates compared to the DNF+WZ estimates. If we restrict the comparison of the aforementioned quantities in redshift intervals that exclude the tails of the distributions, the match between SOMPZ+WZ and DNF+WZ improves (Fig. 7). We further investigate the importance of the tails on the cosmological constraints in Appendix D1, finding that, despite them being important, they do not drive the main difference between the SOMPZ+WZ and DNF+WZ constraints.

5.1.1 Galaxy-matter bias prior from WZ autocorrelation

We tested the impact on the Lambda cold dark matter (Λ CDM) cosmological parameters of using the same broad prior on the Sys(s) function describing the galaxy-matter bias as was done for the WL sample (Gatti et al. 2022). In this work we used more informative values computed from the clustering autocorrelation of the MAGLIM sample, the application of which is explained in more detail in Section 4.2. It is particularly interesting to look at the shape of distributions, especially for bin 2. Fig. 5 shows in grey the 1σ bands for the case without using the autocorrelation, and leaving a much broader prior. While in most bins the difference is not appreciable, and the grey bands are very similar to the solid bands, in bin 2 there is an evident difference. It is therefore suggested that this implementation of the autocorrelation information used as priors in the SOMPZ+WZ combination is able to help us constraining the galaxy-matter bias value, in a way that otherwise would not have been possible with traditional methods. In Fig. 7 is shown the comparison over mean redshift and width of the distributions between SOMPZ+WZ with the more informative prior from the autocorrelation, against the broad prior (labelled as ‘SOMPZ+WZ (broad prior)’). The means and widths are well compatible with the standard SOMPZ+WZ results, and for bins 2 and 3 they are slightly closer to the DNF+WZ results. Even in bin 2, where the shape of the $n(z)$ is substantially different, the values of mean and width do not differ greatly from the standard case, reinforcing the notion that mean and width alone are not sufficient to fully characterize redshift distributions of a lens sample.

6 COSMOLOGICAL RESULTS

In this section, we show the constraints on cosmological and nuisance parameters obtained using the DES Y3 measurements for galaxy-galaxy lensing and galaxy clustering (Prat et al. 2022; Rodríguez-

Monroy et al. 2022) (a.k.a. 2x2pt), and the $n(z)$ from this paper. As in Porredon et al. (2022), we also include in our analysis an additional likelihood constructed with the Shear Ratio (SR) measurements (Sánchez et al. 2022). This exploits galaxy–galaxy lensing signal at small scales ($<6 \text{ Mpc h}^{-1}$) to provide further constraint to the redshift distributions and intrinsic alignment parameters. The ratio of a galaxy–galaxy lensing signal of each lens sample redshift bin computed with respect to two source sample bins results in a primarily geometric measurement, which has been proven to be a powerful method for constraining systematics and nuisance parameters. This adds independent information from SOMPSZ and WZ to the source redshift calibration. The posterior distribution obtained follows the Bayes theorem:

$$P(p|D, M) \propto \mathcal{L}(D, p, M)\Pi(p|M), \quad (19)$$

where $\Pi(p|M)$ is the prior distribution for all the parameters of the model M . For the cosmological inference we use the COSMOSIS pipeline (Zuntz et al. 2015), and we sample the parameter posteriors using the POLYCHORD sampler (Handley, Hobson & Lasenby 2015a,b).

Our data vector $D = \{w(\theta), \gamma_r(\theta)\}$ is compared to theoretical predictions $T(p) = \{w(\theta, p), \gamma_r(\theta, p)\}$ in a Bayesian fashion, and the posterior of the parameters conditional on the data is evaluated assuming a Gaussian likelihood for the data:

$$\log \mathcal{L} \propto -\frac{1}{2}(D - T(p))^T C^{-1}(D - T(p)), \quad (20)$$

where C is the measurement covariance. In our analysis, we vary five (or six) cosmological parameters assuming a Λ CDM (or w CDM) cosmology: $\Omega_m, \sigma_8, n_s, \Omega_b, h_{100}$, and w for the w CDM case. Moreover, we also marginalize over ‘astrophysical’ nuisance parameters (describing intrinsic alignment effects and the galaxy–matter bias of the lens sample), and calibration parameters (redshift uncertainties, shear measurement uncertainties). In short, our set-up (covariance, parameters varied, prior ranges, etc.) is the same as the one adopted in Porredon et al. (2022), except for the redshift $n(z)$ and uncertainties priors of the lens sample, where the ones obtained in this work have been assumed, and other minor changes that we describe below. All modelling and analysis choices, together with the calculations of the theoretical two-point functions, are described in detail in Krause et al. (2021).

Our analyses were not ‘blinded’, since this work occurred after the ‘unblinding’ of the DES Y3 3x2pt results. We did not perform any cosmological analysis until the redshift distributions were frozen; no changes to the redshift distributions (and uncertainties prior) have been performed after looking at the cosmological constraints. To ensure the robustness of our final estimates, we adopted a p -value criteria on the best-fitting models to our data vector. Following Porredon et al. (2022), we required the goodness-of-fit p -value on unblinded data vectors was larger than 1 per cent. The goodness-of-fit has been computed using the predictive posterior distribution (PPD, Doux et al. 2021) and adopted in the main DES Y3 3x2pt analysis. The PPD methodology derives a calibrated probability-to-exceed p ; in the case of goodness-of-fit tests, this is achieved by drawing realizations of the data vector for parameters drawn from the posterior under study which are then compared to actual observations. The distance metric (χ^2) is computed in data space, which is then used to compute the p -value.

Concerning the redshift uncertainties, as it is the primary goal of this work, we proceeded using the fiducial DES Y3 methodology: we parametrize the redshift uncertainties with two parameters for each tomographic bin, which modify a fiducial $n(z)$ distribution with

a shift on the mean and a stretch on the width. The fiducial $n(z)$ is estimated by averaging the SOMPSZ+WZ $n(z)$ realizations. The Gaussian priors on the mean and stretch parameters are centred at the mean and width of the fiducial $n(z)$, while the Gaussian priors width are measured from the variance in the mean and width of the $n(z)$ ensemble. This parametrization can be compared directly to the fiducial DES Y3 2x2pt analysis (Porredon et al. 2022). In Appendix D we describe an alternative marginalization of the redshift uncertainties, by marginalizing over the full sets of $n(z)$ realizations provided by the SOMPSZ+WZ method. In principle, this latter method describes better the redshift uncertainties of our method. However, we find that the currently available techniques that marginalize over the full ensemble of realizations during cosmology inference are prohibitively computationally expensive. Therefore, we defer its application to future work.

Besides the different $n(z)$, we also ran a few analyses where we marginalized over magnification parameters of the lens samples over wide priors. This is different from Porredon et al. (2022), where magnification parameters have been fixed.

For the fiducial DES Y3 2x2pt analysis, the p -value from the data-model χ^2 using all six bins of MAGLIM was not sufficient to pass the 1 per cent criteria. After a series of tests the consensus was that the two highest redshift tomographic bins were responsible for worsening the fit. Therefore, the analysis in Porredon et al. (2022) included only the first four MAGLIM bins. Here, we perform the analyses using all the six bins of the MAGLIM sample, but also using only the first four bins, to verify if the same applies also to this work using different redshift distributions.

In particular, we consider the following scenarios:

- (i) Λ CDM (w CDM); four and six lens bins, fixed magnification. This is the fiducial analysis that mirrors the one presented in Porredon et al. (2022). Five (six) cosmological parameters are varied, including $\Omega_m, \sigma_8, n_s, \Omega_b, h_{100}$ (and w for the w CDM case). Intrinsic alignment, shear measurement, and redshift uncertainties parameters (of both lenses and sources) and galaxy–matter linear biases of the lenses also are marginalized over. The magnification coefficients of the lens sample, however, are fixed to the values estimated from BALROG (Everett et al. 2022). Uncertainties in the redshift distributions of the lens sample are modelled as a shift and stretch in the distributions.
- (ii) Λ CDM (w CDM); four and six lens bins, free magnification. Same parameters as the ones above, but magnification parameters are marginalized over using Gaussian priors. This is an additional set-up considered only after analysing the results from the aforementioned fixed magnification set-up.

In what follows, we will also quote results in terms of the S_8 parameter, defined as $S_8 \equiv \sigma_8(\Omega_m/0.3)^{0.5}$. In Table 7 we summarize the best-fitting values of $S_8, \Omega_m, \sigma_8, w$, and the computed PPD goodness-of-fit p -value for all the different analyses.

6.1 Λ CDM results

6.1.1 Fiducial results: four bins, fixed magnification, and comparison with DNF results

The first cosmological constraints we analyse are the ones obtained assuming a Λ CDM cosmology, using four lens bins and fixed magnification parameters. The decision on which set of results will be quoted as ‘fiducial’ for this work had to be made before conducting any cosmological analysis on data. We initially planned to only run the fiducial analyses with fixed magnification, as in Porredon et al. (2021). The choice between four or six lens bins would depend on the

Table 7. Constraints on the cosmological parameters Ω_m , S_8 , and σ_8 . For each parameter we report the mean of the posterior and the 68 per cent confidence interval. We also report the PPD goodness-of-fit p -value and the probability of the parameter difference (computed over the full parameter space) between the analyses considered in this work and *Planck* TTTEEE lowl lowE (Aghanim et al. 2020). The fiducial results from this work is reported in bold in the first row, while the official, fiducial results of DES Y3 are reported in bold in the second to last row.

$n(z)$	Model	bins	Magnif.	Ω_m	S_8	σ_8	w	p -value	<i>Planck</i>
SOMPZ+WZ SOMPZ (broad prior)	Λ CDM	4	Fixed	0.30 ± 0.04	0.81 ± 0.04	0.81 ± 0.07	–	0.029	1.15σ
SOMPZ+WZ	Λ CDM	4	Gauss.	0.31 ± 0.04	0.76 ± 0.06	0.76 ± 0.09	–	0.037	–
SOMPZ+WZ	Λ CDM	4	Gauss.	0.29 ± 0.04	0.81 ± 0.04	0.83 ± 0.08	–	0.035	1.11 σ
SOMPZ+WZ	Λ CDM	6	Fixed	–	–	–	–	0.008	–
SOMPZ+WZ	Λ CDM	6	Gauss.	0.28 ± 0.03	0.79 ± 0.03	0.82 ± 0.06	–	0.065	2.41 σ
SOMPZ+WZ	w CDM	4	Fixed	0.29 ± 0.04	0.79 ± 0.06	0.81 ± 0.08	-1.2 ± 0.3	0.032	0.46 σ
SOMPZ+WZ	w CDM	4	Gauss.	0.29 ± 0.04	0.79 ± 0.06	0.81 ± 0.07	-1.0 ± 0.3	0.035	0.46 σ
SOMPZ+WZ	w CDM	6	Fixed	0.30 ± 0.04	0.78 ± 0.04	0.78 ± 0.06	-0.9 ± 0.3	0.012	2.29 σ
SOMPZ+WZ	w CDM	6	Gauss.	0.31 ± 0.03	0.83 ± 0.04	0.82 ± 0.05	-0.7 ± 0.2	0.059	2.21 σ
DNF+WZ	Λ CDM	4	Fixed	0.32 ± 0.04	0.78 ± 0.04	0.76 ± 0.07	–	0.019	1.0 σ
DNF+WZ	w CDM	4	Fixed	0.32 ± 0.05	0.78 ± 0.05	0.76 ± 0.07	-1.0 ± 0.3	0.024	–

p -value criteria: if the Λ CDM, six bins, fixed magnification scenario were to yield a p -value above the specified threshold, then we would favour that configuration. This analysis though did not fulfil our p -value criteria (p -value = 0.008, see Table 7), similarly as for the analysis ran with the same settings but using the fiducial redshift distributions from DNF; hence, we do not show those results here. We then chose as fiducial the Λ CDM, four bins, fixed magnification analysis, which is equivalent to the ‘fiducial’ set-up assumed in Porredon et al. (2021), which also allows us to compare our results directly to the ones obtained using the DNF+WZ $n(z)$. The posterior on the cosmological parameters Ω_m and S_8 is shown in the left panel of Fig. 8; the marginalized mean values of S_8 , Ω_m , and σ_8 , along with the 68 per cent confidence intervals, are

$$\Omega_m = 0.30 \pm 0.04, \quad (21)$$

$$\sigma_8 = 0.81 \pm 0.07, \quad (22)$$

$$S_8 = 0.81 \pm 0.04. \quad (23)$$

The PPD goodness-of-fit test for this analysis results into p -value = 0.029, well above our threshold (see also Table 7). In the left panel of Fig. 8 we also compare our results with the constraints obtained using the fiducial DNF+WZ $n(z)$. The size of the posteriors is similar for the two cases, but the two posteriors are slightly shifted; the distance between the posteriors’ peaks in the 2D $\Omega_m - S_8$ plane is $d \sim 0.4\sigma$. In DES Y3 we impose a 0.3σ threshold for differences in the $\Omega_m - S_8$ plane induced by different analysis choices, as larger statistical distances would indicate the presence of systematic uncertainties unaccounted for; these results would apparently violate this criteria. We note, however, that the (arbitrary) 0.3σ threshold adopted by DES refers to differences in the $\Omega_m - S_8$ plane when *noiseless* theory data vectors are assumed. In the presence of noisy data vectors these differences can become larger, without invalidating our criteria. Having said this, a $d \sim 0.4\sigma$ difference none the less show the large impact a different redshift calibration of the lens sample can have on the cosmological constraints. This is somewhat different from the results obtained for the source sample $n(z)$ (Amon et al. 2022), where uncertainties in the redshift calibration had a negligible impact on the cosmological constraints.

In Section 4.2 we explained how for the combination of the two methods we marginalize over possible functional forms of the unknown galaxy-matter bias of the MAGLIM sample, by means of the systematic function $\text{Sys}(\mathbf{s})$ in our clustering model. The prior on

the parameters \mathbf{s} is inferred from the clustering autocorrelation. We tested the impact on the redshift distributions of using a broader prior (the same used in Myles et al. 2021) in Section 5. We have tested the impact of using these $n(z)$ for the cosmological inference, and found that there is no change in constraining power and no shift for Ω_m , but there is a shift on S_8 such to overlap with the fiducial results from DNF+WZ. Therefore, it is clear that the information carried by the autocorrelation is crucial in our cosmological analysis.

6.1.2 Four and six bins, free magnification

As supplementary analyses, we then proceed to relax the fixed priors on the magnification parameters for the lens sample. Instead of fixing them to the values estimated from Elvin-Poole et al. (2023) (as done in the previous section), we leave them as free parameters, using Gaussian priors. In short, Elvin-Poole et al. (2023) estimate the magnification parameters using `BALROG`, by injecting fake galaxies into the wide field with and without applying a small magnification; the difference between the number of galaxies passing the selection in the two cases is then used to estimate the magnification parameters of the sample. These parameters come with a small uncertainty, which is however ignored in the fiducial analysis, as the magnification parameters are assumed to be fixed to the mean `BALROG` value. The central values and the uncertainties are reported in Table C1 in Appendix C. One of the main reasons the DES Y3 fiducial analysis did not vary the magnification parameters was merely computational, as four (or six) additional parameters lengthen the parameter inference process. In principle there is no reason to doubt these estimates. Differences might be caused by the fact that the `BALROG` injections do not completely sample the full DES Y3 footprint, or in case our injections were not fully representative of the DES sample we are analysing.

When varying these parameters in our analyses, we find that the p -value computed using PPD indicates a good fit of the model to the data not only for the four bins case, but also for six bins case (see Table 7). Adding the last two lens bins significantly improves the constraining power on Ω_m by 30 per cent compared to the four bins case, whereas the constraints on S_8 are 20 per cent tighter.

6.2 w CDM results

We then proceed to analyse the results obtained with w CDM, for all four cases: four and six bins, fixed and free magnification, as described in the previous section. Parameter posteriors are shown in

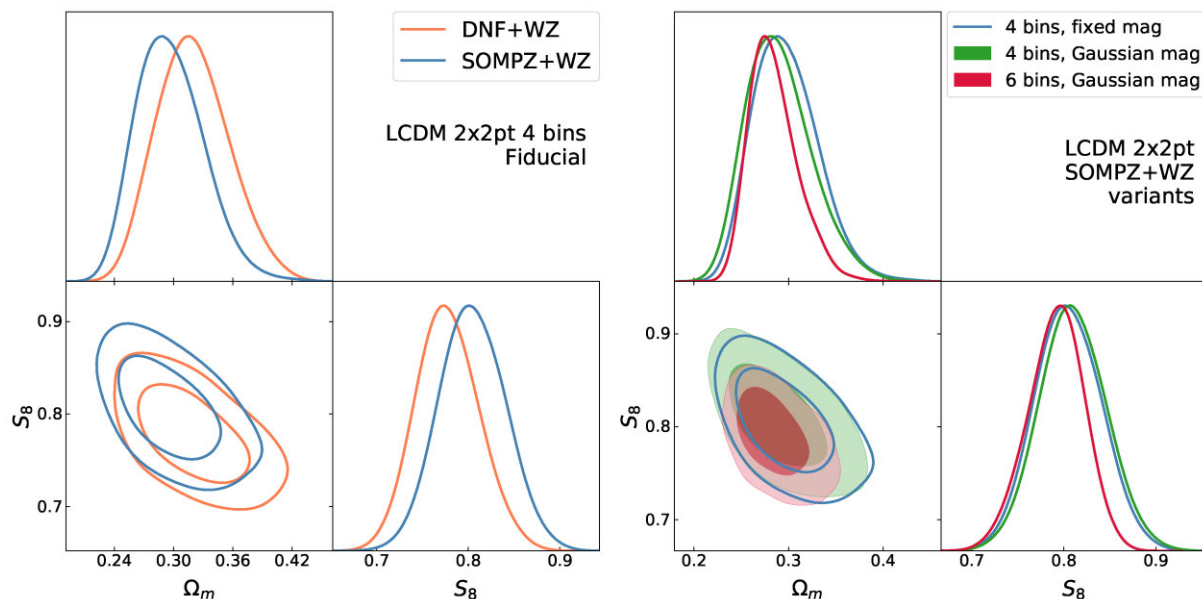


Figure 8. *Left panel:* posterior distributions of the cosmological parameters Ω_m and S_8 for the Λ CDM analysis involving four bins and fixed magnification parameters. The ‘fiducial’ posteriors have been obtained using the DNF+WZ redshift distributions, and they are compared to the ones obtained using the SOMPZ+WZ redshift distributions. *Right panel:* posterior distributions of the cosmological parameters Ω_m and S_8 for the Λ CDM analysis for three different cases: (1) four bins and fixed magnification parameters (the blue contours in the two plots share the same analysis choices); (2) four bins and marginalized over magnification parameters (in solid green); (3) six bins and marginalized over magnification parameters (in solid red). The 2D marginalized contours in both of these figures show the 68 per cent and 95 per cent confidence levels.

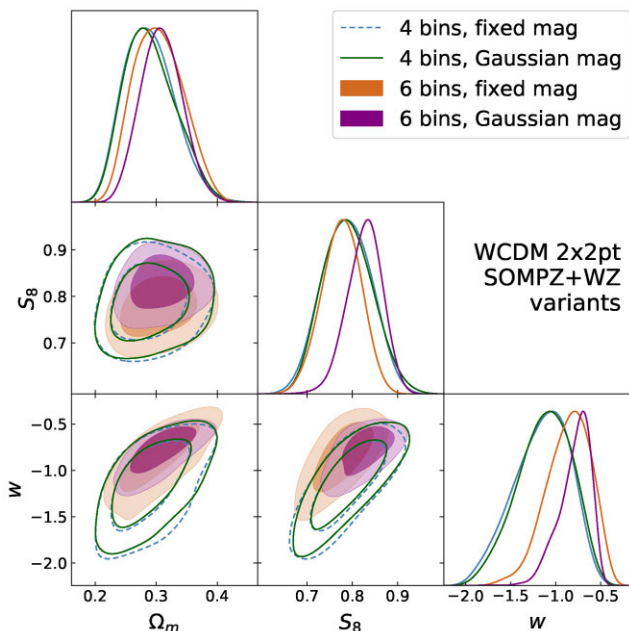


Figure 9. Posterior distributions of the cosmological parameters Ω_m and S_8 and w for four different cases: (1) wCDM, four bins and fixed magnification parameters; (2) wCDM, six bins and fixed magnification parameters, (3) wCDM, four bins and free magnification parameters; (4) wCDM, six bins and free magnification parameters. The 2D marginalized contours in these figures show the 68 per cent and 95 per cent confidence levels. We note that the posteriors of w for the six bins cases are partially affected by the prior edge ($w \in [-2, -0.33]$, Table C1); see the text for more details.

Fig. 9, whereas p -values and parameters constraints are reported in Table 7. All the reported p -values are above our $p = 0.01$ threshold.

In general, the 2x2pt constraints on w are loose and affected by the prior ($-2 < w < -0.3$), but compatible with a Λ CDM scenario. With respect to Λ CDM four bins case, freeing w loosens the constraint on S_8 (both with fixed and with free magnification) by ~ 30 per cent, while leaves it unvaried for Ω_m . For the six bins, we are unable to directly compare to the fixed magnification case, but for the free magnification the constraint on S_8 is ~ 25 per cent looser, while, similarly to the four bins case, it is unvaried for Ω_m .

Passing from the four bins to the six bins configuration, besides increasing the constraints on S_8 , also the constraints on w improves (by ~ 20 per cent), although part of the improvement is due to the posterior partially hitting the prior edge.

Freeing the magnification parameters slightly shifts w towards the upper edge of the prior ($w = -0.3$), and S_8 slightly towards higher values, due to a degeneracy between w , S_8 , and the magnification parameters of the two highest lens bins, which are now fairly broad (see Table C1). Such a shift is not present in the case of four bins, as the Gaussian priors used for the first four magnification parameters are much tighter.

6.3 Statistical distance to Planck

We compute here the statistical distances between our cosmological constraints and the early Universe ones from the *Planck* satellite (Aghanim et al. 2020). To this aim, we used the algorithm presented in Raveri & Doux (2021), which estimates the probability of tension between parameters via Monte Carlo approximation. In particular, the probability of tension between parameters can be expressed as follows:

$$\mathcal{P}(\Delta\theta) = \int_{V_p} \mathcal{P}_A(\theta) \mathcal{P}_B(\theta - \Delta\theta) d\theta, \quad (24)$$

where V_p represents the prior volume, while \mathcal{P}_A and \mathcal{P}_B represent two posterior parameter distributions under study. The probability of having a shift in the parameter space is described by the parameter shifts density:

$$\Delta = \int_{\mathcal{P}(\Delta\theta) > \mathcal{P}(0)} \mathcal{P}(\Delta\theta) d\Delta\theta. \quad (25)$$

This refers to the posterior portion beyond the constant probability contour for no shift, $\Delta\theta = 0$. The integration in equation (25) is performed via Monte Carlo techniques.

The comparison between the results has been performed considering all the parameters shared by our analyses and *Planck*. The values are reported in the last column of Table 7; we find no sign of significant tension ($<3\sigma$) in any of the analysis set-ups considered. In particular, we find that for the four bins case for Λ CDM (both fixed and free magnification) there is good agreement (1.15σ , 1.11σ), similarly for w CDM with four bins we have 0.46σ for both fixed and free magnification. For the six bins cases the values are larger (2.2 – 2.4σ), but still below the 3σ threshold.

7 CONCLUSIONS

In this paper, we presented an alternative calibration of the MAGLIM lens sample redshift distributions from the DES first 3 yr of data (Y3). This new method, which has already been applied to the DES Y3 weak lensing sample (Myles et al. 2021), is based on a combination of an SOMPZ and clustering redshifts (WZ) to estimate redshift distributions and inherent uncertainties. The original redshift calibration of the MAGLIM sample (and cosmological results obtained adopting that calibration) have been originally presented in Porredon et al. (2022), and has been based on the photo- z code DNF (De Vicente et al. 2016) and WZ constraints (Cawthon et al. 2022). The methodology presented in this paper is meant to be more accurate than the original one. First, the SOMPZ method allows a better control over all the potential sources of uncertainties affecting the estimates compared to DNF; second, the clustering constraints (WZ) are incorporated through a rigorous joint likelihood framework which allows to draw $n(z)$ samples conditioned on both clustering and photometric measurements, improving the $n(z)$ estimates (e.g. the final ‘SOMPZ+WZ’ $n(z)$ have a smaller scatter, or uncertainty, compared to the SOMPZ ones, see Fig. 5).

We described in detail the methodology followed to produce the alternative MAGLIM $n(z)$ based on the SOMPZ+WZ approach, together with a detailed report on the main systematics dominating our calibration error budget. Our redshift uncertainties, in particular, are dominated by the impact of sample variance on the SOMPZ estimate (due to the limited area spanned by the deep field sample used in the calibration) and by the effect of the redshift evolution of the galaxy-matter bias of the MAGLIM sample on the WZ constraints. We then compared our SOMPZ+WZ $n(z)$ with the fiducial DNF+WZ $n(z)$ estimates; the means and widths of the 6 MAGLIM tomographic bins show moderate statistical distances, with the largest deviation of 2.7σ in bin 2 (see Table 6). We also found the uncertainties on mean of the redshift distributions of the SOMPZ+WZ method to be slightly larger than the ones of the DNF+WZ method, due to a more conservative calibration of the tails of the redshift distributions. On the other hand, we found the two methods to have a similar constraining power on the widths of the distributions.

We then proceeded investigating the impact on the cosmological constraints of our new redshift calibration. In particular, we used the DES Y3 galaxy–galaxy lensing and galaxy clustering measurements (Prat et al. 2022; Rodríguez-Monroy et al. 2022) (a.k.a. 2x2pt), and

the $n(z)$ from this work, and compared to the results from Porredon et al. (2022). In the ‘fiducial’ configuration, which involves using the first four lens bins and assuming a Λ CDM cosmology, we obtained as marginalized mean values $\Omega_m = 0.30 \pm 0.04$, $\sigma_8 = 0.81 \pm 0.07$, and $S_8 = 0.81 \pm 0.04$. We noted a $\sim 0.4\sigma$ shift in the $\Omega - S_8$ plane compared to the Porredon et al. (2022) results, but no change in terms of constraining power. The shift indicates that the redshift calibration of the lens sample plays a key role on cosmological constraints from the 2x2pt analysis, contrary to the redshift calibration of the source sample (Amon et al. 2022). Subsequently, we explored different analysis set-ups; we tested the case where all the six MAGLIM redshift bins were included, a scenario where the magnification coefficients of the lens sample were marginalized during the inference, and last, we assumed a w CDM cosmology. We found that the inclusion of the last two redshift bins of the MAGLIM sample help improving the constraints on Ω_m by ~ 25 per cent, and on S_8 by ~ 20 per cent.

We also compared our results to the cosmological constraints from *Planck* (Aghanim et al. 2020), finding a no-tension of 1.15σ between the results when four lens bins were considered. We did find a statistical distance of 2.41σ in Λ CDM with free magnification coefficients when including in the analysis the two high-redshift bins ($z > 0.85$), which have not been included in the fiducial DES Y3 analysis (Porredon et al. 2022).

As a final comment, despite the SOMPZ+WZ method’s ability to produce $n(z)$ samples capturing the redshift uncertainties of our estimates, we could not efficiently marginalize over these realization during the cosmological inference, due to computational constraints. Our marginalization strategy followed the one adopted in Porredon et al. (2022): we adopted the mean of the SOMPZ+WZ samples as our fiducial $n(z)$, and marginalized over a shift in the mean and a stretch of the width of the distribution, using as priors the variances in the mean and widths of the SOMPZ+WZ $n(z)$ samples. While this strategy was deemed sufficient for this work, we plan to implement the full marginalization scheme for subsequent analyses of the lens samples with DES Y6 data.

ACKNOWLEDGEMENTS

The project leading to these results has received funding from ‘la Caixa’ Foundation (ID 100010434), under the fellowship LCF/BQ/DI17/11620053 and has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 713673.

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

The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College

London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, the National Optical Astronomy Observatory, the University of Nottingham, The Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, Texas A&M University, and the OZDES Membership Consortium.

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

The DES data management system is supported by the National Science Foundation under Grant Numbers AST-1138766 and AST-1536171. The DES participants from Spanish institutions are partially supported by MINECO under grants AYA2015-71825, ESP2015-66861, FPA2015-68048, SEV-2016-0588, SEV-2016-0597, and MDM-2015-0509, some of which include ERDF funds from the European Union. IFAE is partially funded by the CERCA program of the Generalitat de Catalunya. Research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013) including ERC grant agreements 240672, 291329, and 306478. We acknowledge support from the Brazilian Instituto Nacional de 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.

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS web site is www.sdss.org.

SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, Center for Astrophysics| Harvard & Smithsonian (CfA), the Chilean Participation Group, the French Participation Group, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU)/University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatorio Nacional/MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

8 DATA AVAILABILITY

The DES Y3 data products used in this work, as well as the full ensemble of DES Y3 MAGLIM sample redshift distributions described by this work, are publicly available at <https://des.ncsa.illinois.edu/releases>. As cosmology likelihood sampling software we use `cosmosis`, available at <https://github.com/joezuntz/cosmosis>.

REFERENCES

- Abbott T. M. C. et al., 2018, *ApJS*, 239, 18
 Abolfathi B. et al., 2018, *ApJS*, 235, 42
 Aghanim N. et al., 2020, *A&A*, 641, A6
 Ahumada R. et al., 2020, *ApJS*, 249, 3
 Aihara H. et al., 2018, *PASJ*, 70, S4
 Alam S. et al., 2021, *MNRAS*, 504, 4667
 Alarcon A. et al., 2021, *MNRAS*, 501, 6103
 Amon A. et al., 2022, *Phys. Rev. D*, 105, 023514
 Becker M. R., 2013, *MNRAS*, 435, 115
 Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013, *ApJ*, 762, 109
 Benjamin J. et al., 2013, *MNRAS*, 431, 1547
 Blanton M. R. et al., 2017, *AJ*, 154, 28
 Bonnett C. et al., 2016, *Phys. Rev. D*, 94, 042005
 Buchs R. et al., 2019, *MNRAS*, 489, 820
 Cawthon R. et al., 2022, *MNRAS*, 513, 5517
 Conroy C., Wechsler R. H., Kravtsov A. V., 2006, *ApJ*, 647, 201
 Cordero J. P. et al., 2022, *MNRAS*, 511, 2170
 Cunha C. E., Huterer D., Busha M. T., Wechsler R. H., 2012, *MNRAS*, 423, 909
 Davis M., Peebles P. J. E., 1983, *ApJ*, 267, 465
 Davis C. et al., 2017, preprint ([arXiv:1710.02517](https://arxiv.org/abs/1710.02517))
 Dawson K. S. et al., 2013, *AJ*, 145, 10
 Dawson K. S. et al., 2016, *AJ*, 151, 44
 De Vicente J., Sánchez E., Sevilla-Noarbe I., 2016, *MNRAS*, 459, 3078
 DeRose J. et al., 2019, preprint ([arXiv:1901.02401](https://arxiv.org/abs/1901.02401))
 DeRose J. et al., 2022, *Phys. Rev. D*, 105, 123520
 DES Collaboration, 2022, *Phys. Rev. D*, 105, 023520
 Doux C. et al., 2021, *MNRAS*, 503, 2688
 Eisenstein D. J. et al., 2011, *AJ*, 142, 72
 Elvin-Poole J. et al., 2023, *MNRAS*, 523, 3649
 Eriksen M. et al., 2019, *MNRAS*, 484, 4200
 Everett S. et al., 2022, *ApJS*, 258, 15
 Flaughner B. et al., 2015, *AJ*, 150, 150
 Garilli B. et al., 2014, *A&A*, 562, A23
 Gatti M. et al., 2018, *MNRAS*, 477, 1664
 Gatti M. et al., 2021, *MNRAS*, 504, 4312
 Gatti M. et al., 2022, *MNRAS*, 510, 1223
 Górski K. M., Hivon E., 2011, *Astrophysics Source Code Library*, record ascl: 1107.018
 Gschwend J. et al., 2018, *Astron. Comput.*, 25, 58
 Gunn J. E. et al., 2006, *AJ*, 131, 2332
 Handley W. J., Hobson M. P., Lasenby A. N., 2015a, *MNRAS*, 450, L61
 Handley W. J., Hobson M. P., Lasenby A. N., 2015b, *MNRAS*, 453, 4384
 Hartley W. G. et al., 2022, *MNRAS*, 509, 3547
 Hildebrandt H. et al., 2017, *MNRAS*, 465, 1454
 Hildebrandt H. et al., 2021, *A&A*, 647, A124
 Hoyle B. et al., 2018, *MNRAS*, 478, 592
 Huff E., Mandelbaum R., 2017, preprint ([arXiv:1702.02600](https://arxiv.org/abs/1702.02600))
 Huterer D., Takada M., Bernstein G., Jain B., 2006, *MNRAS*, 366, 101
 Huterer D., Cunha C. E., Fang W., 2013, *MNRAS*, 432, 2945
 Jarvis M. J. et al., 2013, *MNRAS*, 428, 1281
 Jarvis M. et al., 2021, *MNRAS*, 501, 1282
 Johnson A. et al., 2017, *MNRAS*, 465, 4118
 Joudaki S. et al., 2020, *A&A*, 638, L1
 Kohonen T., 1982, *Biol. Cybern.*, 43, 59
 Krause E. et al., 2021, preprint ([arXiv:2105.13548](https://arxiv.org/abs/2105.13548))
 Kuijken K. et al., 2015, *MNRAS*, 454, 3500

- LSST Science Collaboration, 2009, preprint (arXiv:0912.0201)
- Laigle C. et al., 2016, *ApJS*, 224, 24
- Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
- Le Fèvre O. et al., 2013, *A&A*, 559, A14
- Lehmann B. V., Mao Y.-Y., Becker M. R., Skillman S. W., Wechsler R. H., 2017, *ApJ*, 834, 37
- Lidman C. et al., 2020, *MNRAS*, 496, 19
- Lilly S. J. et al., 2009, *ApJS*, 184, 218
- Limber D. N., 1953, *ApJ*, 117, 134
- Lupton R. H., Gunn J. E., Szalay A. S., 1999, *AJ*, 118, 1406
- MacCrann N. et al., 2022, *MNRAS*, 509, 3371
- Masters D. C., Stern D. K., Cohen J. G., Capak P. L., Rhodes J. D., Castander F. J., Paltani S., 2017, *ApJ*, 841, 111
- Masters D. C. et al., 2019, *ApJ*, 877, 81
- McCracken H. J. et al., 2012, *A&A*, 544, A156
- McQuinn M., White M., 2013, *MNRAS*, 433, 2857
- Ménard B., Scranton R., Schmidt S., Morrison C., Jeong D., Budavari T., Rahman M., 2013, preprint (arXiv:1303.4722)
- Morganson E. et al., 2018, *PASP*, 130, 074501
- Morrison C. B., Hildebrandt H., Schmidt S. J., Baldry I. K., Bilicki M., Choi A., Erben T., Schneider P., 2017, *MNRAS*, 467, 3576
- Myles J. et al., 2021, *MNRAS*, 505, 4249
- Myles J. et al., 2023, *MNRAS*, 519, 1792
- Newman J. A., 2008, *ApJ*, 684, 88
- Porredon A. et al., 2021, *Phys. Rev. D*, 103, 043503
- Porredon A. et al., 2022, *Phys. Rev. D*, 106, 103530
- Prakash A. et al., 2016, *ApJS*, 224, 34
- Prat J. et al., 2022, *Phys. Rev. D*, 105, 083528
- Raichoor A. et al., 2017, *MNRAS*, 471, 3955
- Raveri M., Doux C., 2021, *Phys. Rev. D*, 104, 043504
- Reid B. et al., 2016, *MNRAS*, 455, 1553
- Rodríguez-Monroy M. et al., 2022, *MNRAS*, 511, 2665
- Rozo E. et al., 2016, *MNRAS*, 461, 1431
- Sánchez C., Raveri M., Alarcon A., Bernstein G. M., 2020, *MNRAS*, 498, 2984
- Sánchez C. et al., 2022, *Phys. Rev. D*, 105, 083529
- Scodreggio M. et al., 2018, *A&A*, 609, A84
- Scottez V., Benoit-Lévy A., Coupon J., Ilbert O., Mellier Y., 2018, *MNRAS*, 474, 3921
- Secco L. F. et al., 2022, *Phys. Rev. D*, 105, 023515
- Sevilla-Noarbe I. et al., 2021, *ApJS*, 254, 24
- Sevilla I. et al., 2011, preprint (arXiv:1109.6741)
- Sheldon E. S., Huff E. M., 2017, *ApJ*, 841, 24
- Smee S. A. et al., 2013, *AJ*, 146, 32
- Springel V., 2005, *MNRAS*, 364, 1105
- Suchyta E. et al., 2016, *MNRAS*, 457, 786
- Tessore N., Harrison I., 2020, *Open J. Astrophys.*, 3, 6
- van den Busch J. L. et al., 2020, *A&A*, 642, A200
- Wright A. H., Hildebrandt H., van den Busch J. L., Heymans C., Joachimi B., Kannawadi A., Kuijken K., 2020, *A&A*, 640, L14
- Zuntz J. et al., 2015, *Astron. Comput.*, 12, 45

APPENDIX A: MAGLIM SAMPLE IN SIMULATIONS

Due to the small but existing differences in magnitude/colour space between the Buzzard simulation and the DES data (DeRose et al. 2019), we expect the simulated sample to not be a perfect copy of the data sample, although we do not expect this to have a sensible impact on any of the conclusions drawn in this work. The direct application of the fiducial MAGLIM selection (equation 1) to the Buzzard catalogue leads to slightly different number densities and colour distributions with respect to data. We therefore redefine a more adequate MAGLIM selection for Buzzard, with the goal of achieving the same number density as the data sample. The new Buzzard MAGLIM selection is a piece-wise linear selection in redshift and magnitude, similar to equation (1) but with coefficients redefined by minimizing the quadratic sum of the difference in number density with the values in data, for each tomographic bin, in order to avoid discontinuities in the selection. Such a redefined selection guarantees similar number densities as the data sample. We then ensure similar colour distributions by an additional reweighting procedure of the mock catalogue, so as to resemble the colour distributions of the data sample. In particular, we iteratively reweight based on i, r magnitudes and $i-r$ colours, with the final distributions matching closely the data ones, as shown in Fig. A1.

The new MAGLIM selection in Buzzard for each tomographic bin is then the following:

- (i) Bin 1: $i < 2.017 * z_{\text{mean}} + 18.882$
- (ii) Bin 2: $i < 2.687 * z_{\text{mean}} + 18.614$
- (iii) Bin 3: $i < 5.705 * z_{\text{mean}} + 16.954$
- (iv) Bin 4: $i < 2.399 * z_{\text{mean}} + 19.268$
- (v) Bin 5: $i < 9.455 * z_{\text{mean}} + 13.271$
- (vi) Bin 6: $i < -0.960 * z_{\text{mean}} + 23.165$

We list in Table A1 the number densities of MAGLIM in Buzzard, obtained with the fiducial selection and with the adapted in simulations.

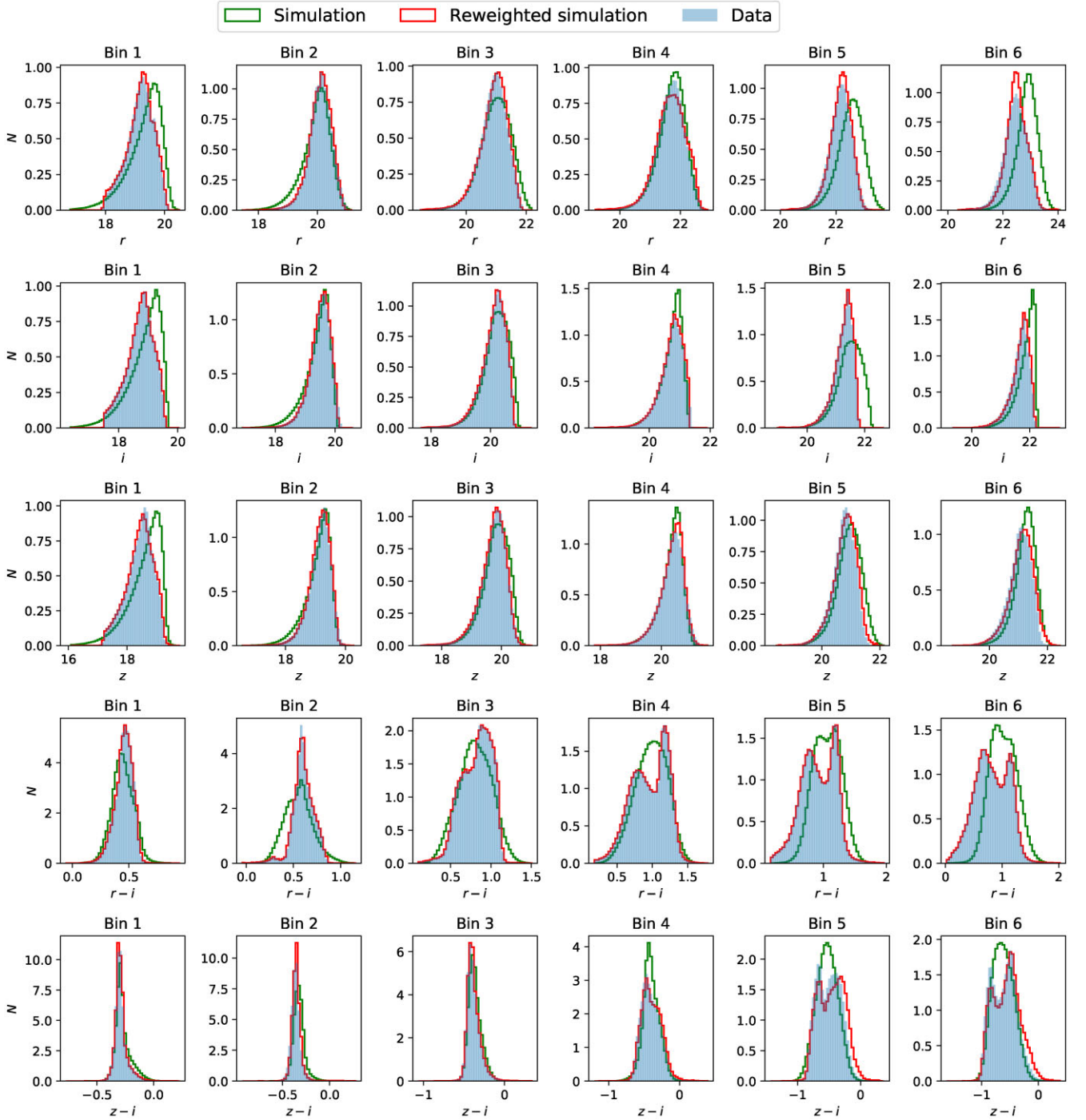


Figure A1. Comparison of riz -band magnitudes and $r-i$, $z-i$ colours of the six bins of the MAGLIM sample, between data (blue) and simulations, before (green) and after reweighting (red). The re-weighting process has proven successful in yielding magnitude distributions that closely resemble those observed in the actual data.

Table A1. Number densities of the MAGLIM sample in Buzzard as obtained with the fiducial MAGLIM selection, and with the one adapted for Buzzard.

Number density	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Before (Fiducial MAGLIM selection)	1.10	0.90	1.12	0.97	0.69	0.76
After (Buzzard MAGLIM selection)	0.98	0.99	0.99	0.99	0.99	0.98

APPENDIX B: VALIDATION IN SIMULATIONS

The validity of our methodology and pipeline has been tested in the Buzzard N -body simulation, introduced in Section 2.8. The measurements of redshift distributions using both phenotypes and clustering were validated in simulations to ensure unbiased estimates with respect to the true redshift distributions. The MAGLIM sample has been recreated in the Buzzard simulations, as described in Section 2. The sample selection has been altered to reproduce as faithfully as possible the number density and colour distributions of the data.

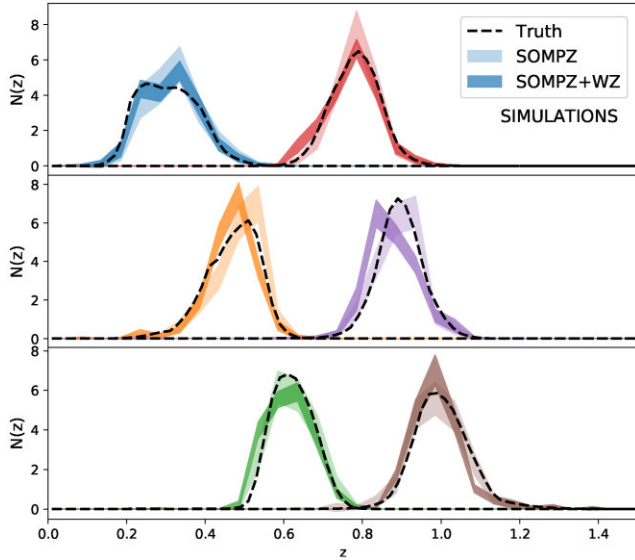


Figure B1. Estimated $n(z)$ in four tomographic bins using a 12×12 cell deep SOM and 32×32 cell wide SOM trained on Buzzard simulations. In the top row we have bins 1 and 4, in the middle row bins 2 and 5, and in the bottom row bins 3 and 6. The redshift sample used here has 100 000 galaxies drawn from 1.38 deg^2 , such that after the MAGLIM selection it yields $\sim 15\,000$ unique galaxies, which is the same order of magnitude as the redshift samples in data, see Table 2. The deep sample is drawn from three fields of size 3.32 , 3.29 , and 1.94 deg^2 , respectively, from the Buzzard simulated sky catalogue. The black dashed line marks the true value, the transparent bands are the $3sDir$ set of $n(z)$, and the solid bands are the realizations once combined with clustering redshifts. We can appreciate the effect of the combined likelihood, resulting in distributions more constrained in terms of shape, and still consistent with the truth.

Table B1. SIMULATIONS: Summary of values for centre values for mean (top panel) and width (bottom panel) for the $n(z)$ distributions as measured in the Buzzard simulations. The values related to SOMPZ and SOMPZ+WZ refer to Fig. B1. Note that the uncertainties quoted here only include sample variance and shot noise.

		Bin 1 $z \in [0.2, 0.4]$	Bin 2 $z \in [0.4, 0.55]$	Bin 3 $z \in [0.55, 0.7]$	Bin 4 $z \in [0.7, 0.85]$	Bin 5 $z \in [0.85, 0.95]$	Bin 6 $z \in [0.95, 1.05]$
$\langle z \rangle$	SOMPZ	0.319 ± 0.009	0.484 ± 0.007	0.623 ± 0.006	0.784 ± 0.006	0.891 ± 0.007	0.993 ± 0.010
	SOMPZ+WZ	0.313 ± 0.008	0.466 ± 0.006	0.613 ± 0.005	0.774 ± 0.007	0.876 ± 0.007	0.988 ± 0.007
$\Delta_{\langle z \rangle}$	SOMPZ	1.46	2.55	0.20	0.28	0.55	1.08
	SOMPZ+WZ	0.83	0.42	1.81	1.28	2.49	2.10
σ_z	SOMPZ	0.075 ± 0.010	0.064 ± 0.007	0.062 ± 0.006	0.056 ± 0.005	0.060 ± 0.005	0.068 ± 0.007
	SOMPZ+WZ	0.077 ± 0.005	0.057 ± 0.005	0.064 ± 0.004	0.068 ± 0.005	0.064 ± 0.005	0.060 ± 0.003
Δ_{σ_z}	SOMPZ	0.53	0.59	1.17	1.42	0.79	0.10
	SOMPZ+WZ	0.46	2.08	2.13	0.93	1.50	2.09

As described in Section 4.1.1, we generated 300 simulated deep field realizations that we used to estimate the SOMPZ method uncertainty, which we report in Table 4, and add into our overall error budget. Here, we illustrate that the uncertainty predicted by the $3sDir$ and the $3sDir + WZ$ models is consistent with the true $n(z)$ in one of these simulated realizations. We start by selecting one of these simulated realizations, which includes the four deep fields and their corresponding Balrog and redshift samples. We then proceeded to perform the $3sDir$ analytical sample variance estimation for that one specific realization. The geometry and resolution of the SOM used in simulations are the same as the ones used in data. There are two differences between our simulated scenario and real data: (1) we use the true redshifts from the Buzzard simulations; (2) we assume all redshift information comes from one of our four deep fields. This latter point matches the modelling assumption of $3sDir$, which also assumes that the redshift information only comes from one out of four fields. This is a conservative choice that inflates the modelled error due to sample variance in real data for the term $p(z|c)$, and it avoids modelling the highly non-trivial selection function of spectroscopic samples coming from fields other than the COSMOS field. We note that the sample variance contribution to the colour distribution $p(c)$ is modelled correctly as coming from all four fields. The SOMPZ redshift distributions, and their uncertainties estimated through the $3sDir$ method, are in agreement with the true distribution, as shown in Fig. B1. In Table B1 we summarize the mean and width of the simulated $n(z)$ of the SOMPZ and SOMPZ+WZ methods in each tomographic bin, and of the true $n(z)$, together with the respective statistical distances from the truth.

We also repeated in simulations the same procedure as for data also for the WZ estimates. We created a mock BOSS/eBOSS catalogue to use as a reference sample. As in data, also in simulations the BOSS/eBOSS sample is divided into 50 bins spanning the $0.1 < z < 1.1$ range of the catalogue (width $\Delta z = 0.02$). Before proceeding with combining the SOMPZ and WZ information through the combined likelihood, the compatibility between SOMPZ and WZ was checked. This was tested by inferring the windowed means and widths of the WZ and SOMPZ redshift estimates, following Gatti et al. (2022). The window has been determined such that magnification effects related to the WZ measurements can be neglected. As for WZ, we used a ‘simple’ estimator for the redshift distribution, inverting equation (17) and ignoring magnification effects (this is possible as we are considering only windowed quantities). The means and widths computed in this way for the two methods were compatible

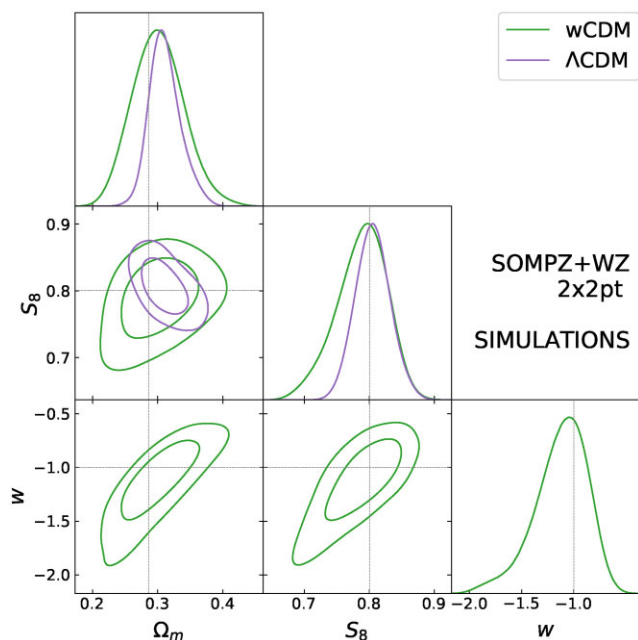


Figure B2. Posterior distributions of the cosmological parameters Ω_m , S_8 , and w for the Λ CDM and w CDM analyses. These have been run with six bins and fixed magnification parameters.

within statistical (and systematic) errors, hence the SOMPZ and WZ could be combined together.

The posterior obtained in simulations from multiplying the two likelihoods is shown in Fig. B1, in which the effect of the combination immediately stands out: the additional information from clustering redshifts places a tight constraint on the shape of the $n(z)$, while still being in agreement with the true distribution. This larger constraining power derives from the fact that in clustering the number density for each redshift bin correlates across neighbouring bins, which restrains the joint likelihood to prefer smoother realizations and reject the ones with more uncorrelated values of clustering.

As the second phase of the validation process, a full 2x2pt cosmological analysis was performed. We utilized the data vector consisting of the two-point measurements from the Buzzard simulations and the redshift distributions obtained from the SOMPZ+WZ method, obtained as described in the previous paragraph. We considered both Λ CDM and w CDM models, fixing magnification parameters and including all six MAGLIM tomographic bins. Additionally, we fixed the source galaxies redshift distributions, to ensure any deviation from the true parameter values of the simulation would be caused by the lens $n(z)$ alone. The mean values of S_8 , Ω_m (and w), with their respective 68 percent confidence intervals, are

- (i) Λ CDM: $S_8 = 0.73 \pm 0.18$, $\Omega_m = 0.31 \pm 0.07$;
- (ii) w CDM: $S_8 = 0.71 \pm 0.18$, $\Omega_m = 0.30 \pm 0.08$, $w = -1.3 \pm 0.4$.

For both analyses, the posterior distributions successfully recovered the input parameters (see Section 2), as displayed in Fig. B2.

APPENDIX C: COSMOLOGICAL PARAMETERS

In Table C1 are listed all the cosmological parameters included in our fiducial analysis.

Table C1. The parameters and their priors used in the fiducial MAGLIM Λ CDM, and w CDM analyses. The parameter w is fixed to -1 in Λ CDM. The square brackets denote a flat prior, while the parentheses denote a Gaussian prior of the form $\mathcal{N}(\mu, \sigma)$.

Parameter	Fiducial	Prior
Cosmology		
Ω_m	0.3	[0.1, 0.9]
$A_s 10^9$	2.19	[0.5, 5.0]
n_s	0.97	[0.87, 1.07]
w	-1.0	$[-2, -0.33]$
Ω_b	0.048	[0.03, 0.07]
h_0	0.69	[0.55, 0.91]
$\Omega_b h^2 10^3$	0.83	[0.6, 6.44]
Linear galaxy bias		
b_i	1.5, 1.8, 1.8, 1.9, 2.3, 2.3	[0.8, 3.0]
Lensmagnification		
C_1	0.43	(0.43, 0.51)
C_2	0.30	(0.30, 0.48)
C_3	1.75	(1.75, 0.39)
C_4	1.94	(1.94, 0.35)
C_5	1.56	(1.56, 0.71)
C_6	2.96	(2.96, 0.95)
Lens photo-z		
Δz_1^1	0.0	(0.0, 0.0164)
Δz_1^2	0.0	(0.0, 0.0100)
Δz_1^3	0.0	(0.0, 0.0085)
Δz_1^4	0.0	(0.0, 0.0084)
Δz_1^5	0.0	(0.0, 0.0094)
Δz_1^6	0.0	(0.0, 0.0116)
σz_1^1	1.0	(1.0, 0.0639)
σz_1^2	1.0	(1.0, 0.0624)
σz_1^3	1.0	(1.0, 0.0315)
σz_1^4	1.0	(1.0, 0.0409)
σz_1^5	1.0	(1.0, 0.0515)
σz_1^6	1.0	(1.0, 0.0650)
Intrinsic alignment		
a_i ($i \in [1, 2]$)	0.7, -1.36	$[-5, 5]$
η_i ($i \in [1, 2]$)	$-1.7, -2.5$	$[-5, 5]$
b_{TA}	1.0	[0, 2]
z_0	0.62	Fixed
Source photo-z		
Δz_s^1	0.0	(0.0, 0.018)
Δz_s^2	0.0	(0.0, 0.013)
Δz_s^3	0.0	(0.0, 0.006)
Δz_s^4	0.0	(0.0, 0.013)
Shear calibration		
m^1	-0.006	$(-0.006, 0.008)$
m^2	-0.010	$(-0.010, 0.013)$
m^3	-0.026	$(-0.026, 0.009)$
m^4	-0.032	$(-0.032, 0.012)$

APPENDIX D: REDSHIFT UNCERTAINTIES SAMPLING STRATEGY

How redshift uncertainties are propagated in the cosmological analysis can have an impact on the final result. In this section we discuss different strategies to marginalize over the redshift uncertainties of our sample during the cosmological inference. Because we have can rely on a full ensemble of $n(z)$ shapes capturing our redshift uncertainties, we can compare three different sampling methods:

- (i) **Shift**: we compress the realizations by computing their average, and marginalize over a shift on the mean;
- (ii) **Shift and stretch**: we compress the realizations by computing their average, and marginalize over both a shift on the mean and on a stretch on the width;
- (iii) **Full shape**: we provide as input all the produced realizations and we rank them by one of their properties using the *Hyperrank* method (Cordero et al. 2022), marginalizing over the full shape of the distributions.

Using only *shifts* is the methodology usually adopted to model redshift uncertainties in weak lensing sample, as the weak lensing kernel is mostly sensitive to the mean of the redshift distributions. On the other hand, clustering and galaxy–galaxy lensing measurements are also very sensitive to the width of the lens redshift distributions; therefore, the *shift and stretch* approach is preferred. The full shape marginalization, in theory, is more accurate, because it accounts for the uncertainties in the higher order moments of the distribution; however, depending on the science case, it might not make a huge impact on the final constraints. The full shape marginalization is implemented via hyperrank (Cordero et al. 2022), which is an algorithm that orders realizations of the ensemble according to a parameter, which facilitates the sampling and marginalization over the $n(z)$ ensemble within the cosmological likelihood Markov chains. Hyperrank was also implemented for the WL sources, although it had a negligible impact on the results. The quantity chosen for the ranking in that case was the mean. We decided for this case it would be more appropriate to perform the optimized ranking of the realization by the 68 per cent sigma rather than the mean, and we tested it indeed improved the performance of the sampling. To test the different sampling strategies, we built a synthetic noiseless data vector based on theory predictions at fixed cosmology and we used as $n(z)$ the realizations average of the SOMPZ+WZ estimates in data. We then marginalized over redshift uncertainties using the three approaches aforementioned. We performed this test both using four or six lens bins, although here we are just going to show the posteriors obtained with four bins as they are not qualitatively different from the ones with six bins. The results of this test are shown in Fig. D1, where we show the posterior of σ_8 , Ω_m and for sake of simplicity, two out of the four galaxy–matter linear biases.

Focusing on the shift and shift+stretch contours, one can notice that the width of the contour in the direction perpendicular to the degeneration axis is larger for the shift + stretch. This is related to impact of the additional marginalization over the width of the distributions. One caveat is that in our marginalization scheme (as adopted in the main DES Y3 2x2pt analysis), we are implicitly neglecting correlations between the uncertainties in the mean and widths of the distributions, which usually show a certain degree of correlation (from ~ 10 per cent to ~ 30 per cent, depending from the tomographic bin). These are neglected, which might translate in a slight overestimation of our constraints. When marginalizing over the uncertainties using the hyperrank framework, on the other hand, such correlations are implicitly accounted for. Indeed, one can notice that the hyperrank posteriors are slightly tighter than the shift or shift–stretch posteriors.

Unfortunately, we did not manage to successfully apply hyperrank to the data. When performing the cosmological analysis on data using hyperrank, we found significantly less smooth posteriors compared to our tests on simulations. A similar behaviour has also been found when applying hyperrank to the DES Y3 source sample (Amon et al. 2022), and it has been interpreted as a consequence of a possible larger degree of complexity of the redshift distributions of

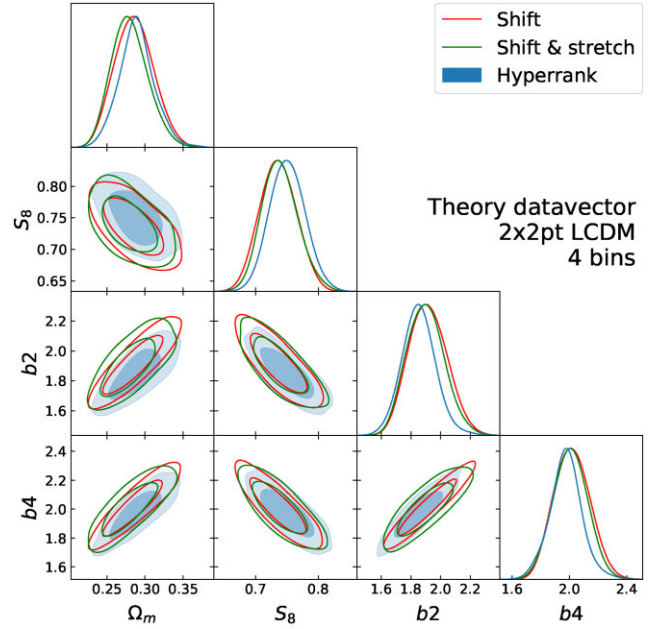


Figure D1. Posterior distributions of the cosmological parameters Ω_m , S_8 , and two out of four of the galaxy–matter biases (b_2 , b_4) for the Λ CDM analysis involving four bins and fixed magnification parameters. These analyses have been obtained assuming a theoretical data vector and adopting different marginalization schemes on the redshift distribution of the lens sample.

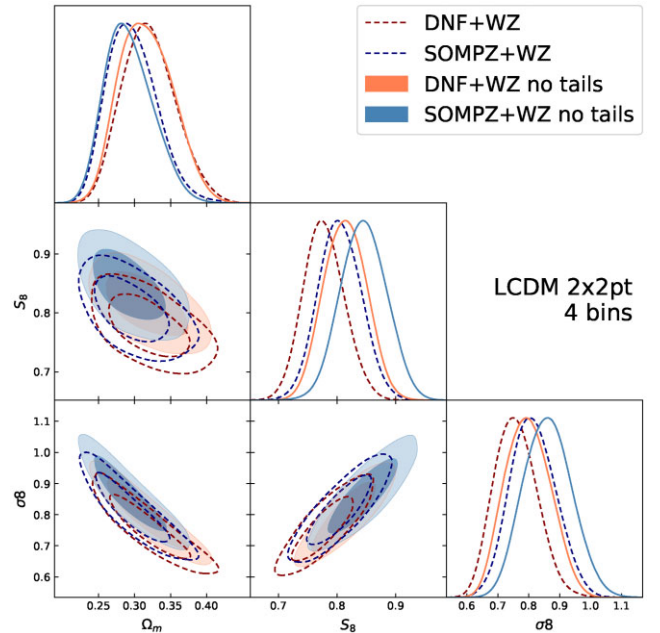


Figure D2. Same as the left panel of Fig. 8, but with two additional posteriors overplotted representing the constraints obtained using the redshift distributions with ‘clipped’ tails.

our data compared to simulations. We attempted both to artificially smooth our $n(z)$ and to increase the number of samples from the SOMPZ+WZ method, without reaching a satisfactory level. Due to the very high computational cost of running a cosmological chain using hyperrank, we could only test a few different levels of smoothing before deciding to abandon hyperrank for this work, and choose the shift + stretch as photo- z uncertainty marginalization

methodology. For DES Y6, we plan to apply several tools that will speed up our cosmological inference, enabling more tests on hyperrank, which has great potential and whose implementation is a goal for the DES Y6 analysis.

D1 Cosmological constraints with clipped $n(z)$ tails

Here, we test whether the difference between DNF+WZ and SOMPZ+WZ constraints (Fig. 8) were only due to the different treatment of redshift outliers and of the tails of the redshift distributions. We artificially removed the tails from the DNF+WZ and SOMPZ+WZ $n(z)$ (i.e. we set the distributions to zero), and repeated our cosmological analysis. We used as definition of the tails the same interval used to calibrate the DNF distribution with the WZ constraints adopted in Porredon et al. (2022). Results for the Λ CDM case, four bins, and fixed magnification are shown in Fig. D2. By removing the tails, both posteriors are shifted, which means that the calibration of the tails of the redshift distribution is important for our cosmological analysis. Since the two posteriors are shifted but they still do not overlap, we can assume that the differences in the bulk of the redshift distributions inferred by two methods is also crucially driving the differences at the constraints level seen in Fig. 8.

¹Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA

²Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA

³Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, E-08193 Bellaterra (Barcelona) Spain

⁴Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439, USA

⁵Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA

⁶Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA

⁷Department of Physics, The Ohio State University, Columbus, OH 43210, USA

⁸Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain

⁹Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain

¹⁰Physics Department, William Jewell College, Liberty, MO 64068, USA

¹¹Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA

¹²Kavli Institute for Particle Astrophysics & Cosmology, PO Box 2450, Stanford University, Stanford, CA 94305, USA

¹³SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA

¹⁴Fermi National Accelerator Laboratory, PO Box 500, Batavia, IL 60510, USA

¹⁵Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

¹⁶Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

¹⁷Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA

¹⁸Instituto de Física Teórica, Universidade Estadual Paulista, São Paulo, Brazil

¹⁹Institute for Astronomy, University of Hawai'i, 2680 Woodlawn Drive, Honolulu, HI 96822, USA

²⁰Physics Department, University of Wisconsin-Madison, 2320 Chamberlin Hall, 1150 University Avenue Madison, WI 53706-1390, USA

²¹Department of Physics, Northeastern University, Boston, MA 02115, USA

²²Laboratório Interinstitucional de e-Astronomia - LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ 20921-400, Brazil

²³Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15312, USA

²⁴Instituto de Astrofísica de Canarias, E-38205 La Laguna, Tenerife, Spain

²⁵Departamento Astrofísica, Universidad de La Laguna, E-38206 La Laguna, Tenerife, Spain

²⁶Center for Astrophysical Surveys, National Center for Supercomputing Applications, 1205 West Clark St, Urbana, IL 61801, USA

²⁷Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801, USA

²⁸Jodrell Bank Center for Astrophysics, School of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK

²⁹Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain

³⁰Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

³¹NSF AI Planning Institute for Physics of the Future, Carnegie Mellon University, Pittsburgh, PA 15213, USA

³²Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr, Pasadena, CA 91109, USA

³³Department of Astronomy, University of California, Berkeley, 501 Campbell Hall, Berkeley, CA 94720, USA

³⁴Departments of Statistics and Data Science, University of Texas at Austin, Austin, TX 78757, USA

³⁵Faculty of Physics, University Observatory, Ludwig-Maximilians-Universität, Scheinerstr 1, D-81679 Munich, Germany

³⁶Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ 20921-400, Brazil

³⁷School of Physics and Astronomy, Cardiff University, Cardiff CF24 3AA, UK

³⁸Department of Astronomy, University of Geneva, ch. d'Écogia 16, CH-1290 Versoix, Switzerland

³⁹Department of Physics & Astronomy, University College London, Gower Street, London WC1E 6BT, UK

⁴⁰Department of Astronomy/Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721-0065, USA

⁴¹Department of Physics and Astronomy, Pevensey Building, University of Sussex, Brighton BN1 9QH, UK

⁴²Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, UK

⁴³Brookhaven National Laboratory, Bldg 510, Upton, NY 11973, USA

⁴⁴Department of Physics, Duke University, Durham NC 27708, USA

⁴⁵Cerro Tololo Inter-American Observatory, NSF's National Optical-Infrared Astronomy Research Laboratory, Casilla 603, La Serena, Chile

⁴⁶Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK

⁴⁷Institut d'Astrophysique de Paris, CNRS, Sorbonne Université, UMR 7095, F-75014, Paris, France

⁴⁸Institut d'Astrophysique de Paris, Sorbonne Universités, UPMC Univ Paris 06, UMR 7095, F-75014 Paris, France

⁴⁹Astronomy Unit, Department of Physics, University of Trieste, via Tiepolo 11, I-34131 Trieste, Italy

⁵⁰INAF-Osservatorio Astronomico di Trieste, via G. B. Tiepolo 11, I-34143 Trieste, Italy

⁵¹Institute for Fundamental Physics of the Universe, Via Beirut 2, I-34014 Trieste, Italy

⁵²Hamburger Sternwarte, Universität Hamburg, Gojenbergsweg 112, D-21029 Hamburg, Germany

⁵³Department of Physics, IIT Hyderabad, Kandi, Telangana 502285, India

⁵⁴Institute of Theoretical Astrophysics, University of Oslo. PO Box 1029 Blindern, NO-0315 Oslo, Norway

⁵⁵Instituto de Física Teórica UAM/CSIC, Universidad Autónoma de Madrid, E-28049 Madrid, Spain

⁵⁶Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA

⁵⁷School of Mathematics and Physics, University of Queensland, Brisbane, QLD 4072, Australia

⁵⁸Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA

⁵⁹Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA

⁶⁰Australian Astronomical Optics, Macquarie University, North Ryde, NSW 2113, Australia

⁶¹*Lowell Observatory, 1400 Mars Hill Rd, Flagstaff, AZ 86001, USA*

⁶²*Centre for Gravitational Astrophysics, College of Science, The Australian National University, ACT 2601, Australia*

⁶³*The Research School of Astronomy and Astrophysics, Australian National University, ACT 2601, Australia*

⁶⁴*Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, CP 66318, São Paulo, SP 05314-970, Brazil*

⁶⁵*Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA*

⁶⁶*Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain*

⁶⁷*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*

⁶⁸*School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK*

⁶⁹*Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA*

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