DES-2019-0485
FERMILAB-PUB-20-659-AE

MNRAS **000**, 1–29 (2020) Preprint 17 December 2020 Compiled using MNRAS LaTeX style file v3.0

Dark Energy Survey Year 3 Results: Redshift Calibration of the Weak Lensing Source Galaxies

J. Myles,[1,2,3★] A. Alarcon,[4,65,66†] A. Amon,[2] C. Sánchez,[5] S. Everett,[6] J. DeRose,[7,6] J. McCullough,[2] D. Gruen,[1,2,3] G. M. Bernstein,[5] M. A. Troxel,[8] S. Dodelson,[9] A. Campos,[9] N. MacCrann,[10] B. Yin,[9] M. Raveri,[11] A. Amara,[12] M. R. Becker,[4] A. Choi,[13] J. Cordero,[14] K. Eckert,[5] M. Gatti,[15] G. Giannini,[15] J. Gschwend,[16,17] R. A. Gruendl,[18,19] I. Harrison,[20,14] W. G. Hartley,[21] E. M. Huff,[22] N. Kuropatkin,[23] H. Lin,[23] D. Masters,[24] R. Miquel,[25,15] J. Prat,[26] A. Roodman,[2,3] E. S. Rykoff,[2,3] I. Sevilla-Noarbe,[27] E. Sheldon,[28] R. H. Wechsler,[1,2,3] B. Yanny,[23] T. M. C. Abbott,[29] M. Aguena,[30,16] S. Allam,[23] J. Annis,[23] D. Bacon,[12] E. Bertin,[31,32] S. Bhargava,[33] S. L. Bridle,[14] D. Brooks,[34] D. L. Burke,[2,3] A. Carnero Rosell,[35,36] M. Carrasco Kind,[18,19] J. Carretero,[15] F. J. Castander,[37,38] C. Conselice,[14,39] M. Costanzi,[40,41] M. Crocce,[37,38] L. N. da Costa,[16,17] M. E. S. Pereira,[42] S. Desai,[43] H. T. Diehl,[23] T. F. Eifler,[44,22] J. Elvin-Poole,[13,45] A. E. Evrard,[46,42] I. Ferrero,[47] A. Ferté,[22] B. Flaugher,[23] P. Fosalba,[37,38] J. Frieman,[23,11] J. García-Bellido,[48] E. Gaztanaga,[37,38] T. Giannantonio,[49,50] S. R. Hinton,[51] D. L. Hollowood,[6] K. Honscheid,[13,45] B. Hoyle,[52,53,54] D. Huterer,[42] D. J. James,[55] E. Krause,[44] K. Kuehn,[56,57] O. Lahav,[34] M. Lima,[30,16] M. A. G. Maia,[16,17] J. L. Marshall,[58] P. Martini,[13,59,60] P. Melchior,[61] F. Menanteau,[18,19] J. J. Mohr,[52,53] R. Morgan,[62] J. Muir,[2] R. L. C. Ogando,[16,17] A. Palmese,[23,11] F. Paz-Chinchón,[49,19] A. A. Plazas,[61] M. Rodriguez-Monroy,[27] S. Samuroff,[9] E. Sanchez,[27] V. Scarpine,[23] L. F. Secco,[5] S. Serrano,[37,38] M. Smith,[63] M. Soares-Santos,[42] E. Suchyta,[64] M. E. C. Swanson,[19] G. Tarle,[42] D. Thomas,[12] C. To,[1,2,3] T. N. Varga,[53,54] J. Weller,[53,54] and W. Wester[23]

(DES Collaboration)

The authors' affiliations are shown at the end of this paper.

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

Determining the distribution of redshifts of galaxies observed by wide-field photometric experiments like the Dark Energy Survey is an essential component to mapping the matter density field with gravitational lensing. In this work we describe the methods used to assign individual weak lensing source galaxies from the Dark Energy Survey Year 3 Weak Lensing Source Catalogue to four tomographic bins and to estimate the redshift distributions in these bins. As the first application of these methods to data, we validate that the assumptions made apply to the DES Y3 weak lensing source galaxies and develop a full treatment of systematic uncertainties. Our method consists of combining information from three independent likelihood functions: Self-Organizing Map $p(z)$ (SOMPZ), a method for constraining redshifts from galaxy photometry; clustering redshifts (WZ), constraints on redshifts from cross-correlations of galaxy density functions; and shear ratios (SR), which provide constraints on redshifts from the ratios of the galaxy-shear correlation functions at small scales. Finally, we describe how these independent probes are combined to yield an ensemble of redshift distributions encapsulating our full uncertainty. We calibrate redshifts with combined effective uncertainties of $\sigma_{\langle z \rangle} \sim 0.01$ on the mean redshift in each tomographic bin.

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

★ E-mail: jmyles@stanford.edu
† E-mail: alexalarcongonzalez@gmail.com

1 INTRODUCTION

The matter density fluctuations present in the Universe, and their evolution over time under the impact of gravity and cosmic expansion, are sensitive to cosmological physics, including the nature of dark energy, neutrino masses, and the nature of dark matter. Galaxy surveys like the Dark Energy Survey (DES, Troxel et al. 2018; Abbott et al. 2018), the Kilo-Degree Survey (KiDS, Heymans et al. 2020), Hyper Suprime-Cam survey (HSC, Hikage et al. 2019), the Legacy Survey of Space and Time (LSST, LSST Dark Energy Science Collaboration 2012), or the *Euclid* mission (Laureijs et al. 2011) use this to achieve competitive constraints on cosmological parameters from observable proxies of the matter density field. In particular, the DES first three years of observation data are used, among other purposes, to measure three two-point (3×2pt.) correlation functions (DES Collaboration et al. 2020):

(i) Cosmic shear: the correlation function of the shapes of "source" galaxies divided into four tomographic bins (Gatti & Sheldon et al., 2020b; Amon et al. 2020; Secco, Samuroff et al. 2020).
(ii) Galaxy clustering: the auto-correlation function of the positions of luminous red "lens" galaxies selected by the RedMaGiC algorithm (Rozo et al. 2016; Rodríguez-Monroy et al. 2020), or alternatively the positions of an optimized magnitude-limited sample (Porredon et al. 2020, prep).
(iii) Galaxy-galaxy lensing: the cross-correlation function of source galaxy shapes around lens galaxy positions (Prat et al. 2020).

The use of gravitational lensing signals is indispensable in this approach: In a photometric survey, while the positions of galaxies can be used as tracing matter density, the only direct connection to the underlying density field is through its effect on the images of distant galaxies by means of gravitational lensing. In order to draw conclusions on the physical density fluctuations from observations of gravitational lensing, however, the distances to the lensed background sources must be known.

Any gravitational lensing measurement, including the interpretation of the cosmic shear and galaxy-galaxy lensing correlation functions, therefore relies on a robust characterization of the distribution $n(z)$ of redshifts z of the respective source galaxy samples (Huterer et al. 2006; Hildebrandt et al. 2012; Benjamin et al. 2013; Huterer et al. 2013; Samuroff et al. 2017; Joudaki et al. 2019; Tessore & Harrison 2020). While ideally this could be accomplished by measuring the spectrum of each galaxy in a given catalogue, it is so far only feasible to gather spectra for small, possibly non-representative subsets of galaxies. As a consequence, large optical imaging surveys with measurements of tens or hundreds of millions of galaxies must rely on relatively few, noisy photometric bands to constrain redshifts. The key challenge in doing this is the presence of degeneracies in the statistical colour-redshift relation, making it commonly impossible to uniquely determine the redshift of any given galaxy from wide-band photometry. One can address this challenge by determining a prescription for reweighting the $n(z)$ of a sample with credible, known redshifts according to those galaxies' relative abundance in the overall sample detected and selected by a photometric survey (e.g. Lima et al. 2008; Cunha et al. 2012; Bonnett et al. 2016; Speagle & Eisenstein 2017a,b; Hoyle & Gruen et al., 2018; Tanaka et al. 2018; Hildebrandt et al. 2020a; Wright et al. 2020a; Euclid Collaboration 2020; Schmidt et al. 2020). The problem of degeneracies in the statistical colour-redshift relation in this case manifests as uncertainty on the measured redshift distribution, often quantified in terms of uncertainties on the moments of the measured $n(z)$. Much of the work in estimating redshift distributions is dedicated to understanding how measured $n(z)$ are biased due to sample variance and selection biases in the sample of galaxies with credible redshifts (Gruen & Brimioulle 2017; Hartley & Chang et al., 2020b). In this work, we describe the analysis used to characterize the redshift distributions of the DES Year 3 (the first 3 seasons of observations) source galaxy sample from their photometry, validate this methodology on realistic simulations of the survey data, and present the results of the analysis on the DES data.

A challenge to determination of $n(z)$ is the combination of incompleteness in the spectroscopic samples and inaccuracies in many-band photometric redshifts used to calibrate the colour-redshift maps. Our work ameliorates this challenge by weighting the redshift-calibration sample to match the abundance of the target sample in a high-dimensional colour space (Buchs & Davis et al., 2019). Differences in reweighting procedures are known to result in scientifically meaningfully different constraints on the matter clustering parameter σ_8 (Troxel & MacCrann et al., 2018; Joudaki & Hildebrandt et al., 2019), highlighting the critical importance of properly accounting for the impact of selection biases on redshift distribution measurement.

A robust redshift analysis should be validated on simulations, rely on multiple independent data sets and methodologies, and have well-characterized uncertainties. Besides the work presented in this paper on photometric redshifts, we accomplish this by combining photometric information with galaxy clustering and shear ratios to constrain redshift distributions. Clustering redshifts (WZ) and shear ratios (SR) play the essential role of providing additional, independent constraining power to validate and further constrain photometric redshift distributions (Gatti, Giannini et al. 2020a; Sánchez, Prat et al. 2020a).

We describe this overall DES Year 3 redshift inference scheme in §2. In §3 we describe the data used in this analysis. We develop the methodology for determining $n(z)$ from galaxy magnitude and colours and the uncertainty on those $n(z)$ in §4 and §5, respectively. We present our results in §6 and discuss their implications in §7.

2 DES Y3 REDSHIFT SCHEME

The overarching DES Year 3 redshift inference scheme uses multiple, independent analyses to robustly characterize the weak lensing source galaxy redshift distributions. As illustrated in Fig. 1, the three likelihood functions computed rely on three independent methods and data: SOMPZ, clustering redshifts, and shear ratios.

(i) Self-Organizing Map $p(z)$ (SOMPZ) leverages the Y3 DES Deep Fields (Hartley, Choi et al. 2020a) to accurately determine the number density of galaxies in deep $ugrizJHK_s$ colour space. Since redshifts are well-constrained at a given $ugrizJHK_s$ colour, this number density can be used to properly weigh galaxies within a sample of credible redshifts in a way that is not subject to selection biases. In brief, this method relies on determining the $p(z)$ at a given cell in 8-band colour space from galaxies with deep 8-band coverage, the probability of each cell in 8-band colour space contributing to the galaxies in a given cell in noisy 3-band colour-magnitude space, and the abundance of galaxies in 3-band colour-magnitude space, to compute the overall redshift distribution of the Year 3 lensing source galaxy sample. The validation of this method and the characterization of its sources of uncertainty are outlined in detail in this work.
(ii) Clustering redshifts constrain the distances to source galaxies from their angular galaxy clustering with samples of reference galaxies within narrow redshift ranges (Newman 2008; Ménard et al.

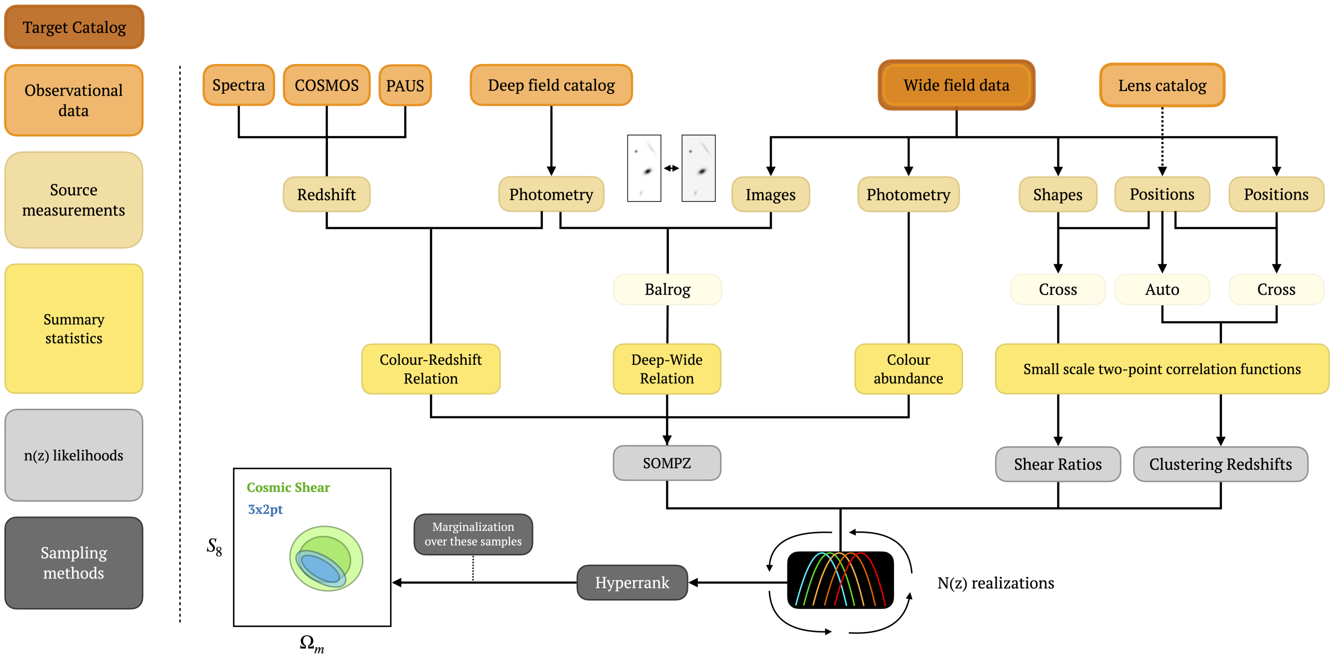

Figure 1. Flowchart illustrating the weak lensing redshift distributions calibration scheme. The three main $n(z|\text{model})$ likelihood functions of the analysis, shown in gray, are SOMPZ, clustering redshifts, and shear ratios. Note that the parameter constraint plot is only an illustration and is not a result from real measurements.

2013; McQuinn & White 2013; Johnson et al. 2017; Morrison et al. 2017; Davis & Gatti et al., 2017; Gatti & Vielzeuf et al., 2018; Cawthon et al. 2017; van den Busch et al. 2020; Hildebrandt et al. 2020a). This method is based on the fact that the amplitude of this correlation function is proportional to the fraction of source galaxies in physical proximity to those reference galaxies. Clustering redshifts validate and refine photometric $n(z)$ with the key benefit of avoiding any reliance on the statistical colour-redshift relation and bypassing the completeness issues associated with spectroscopic survey coverage. The details of this analysis are described fully in Gatti, Giannini et al. (2020a).

(iii) Shear ratios (Jain & Taylor 2003; Mandelbaum et al. 2005; Heymans et al. 2012; Prat & Sánchez et al., 2018; Prat & Baxter et al., 2019; Hildebrandt et al. 2020b) provide additional constraining power and validation by measuring the galaxy-galaxy lensing signal of a lens galaxy redshift bin at small scales. The ratio of this signal from two source bins reflects the ratio of mean lensing efficiencies of objects in those source bins with respect to the lens bin redshift. This, in turn, depends on the redshift distribution of the sources. Because this methodology utilizes lensing signals, it is virtually independent from SOMPZ and clustering redshifts. The methodology of this analysis is described fully in Sánchez, Prat et al. (2020a). Both the clustering and shear ratio redshift constraints are derived from data on small angular scales, which allows the redshift constraints to remain largely statistically independent of cosmological constraints based on larger-scale signals.

In summary, we use galaxy photometry to constrain $n(z)$ with SOMPZ, galaxy positions to constrain $n(z)$ with clustering redshifts, and galaxy shapes to constrain $n(z)$ with shear ratios. As in past work, we assess consistency of these measurements. We further subsequently *combine* these measurements. The final result of this analysis is an ensemble of redshift distributions whose variation encodes the combined uncertainties on the $n(z)$ due to all sources of information. Any DES Y3 lensing likelihood that uses the same redshift bins can be estimated by sampling from this ensemble. Specifically, the $n(z)$s in this ensemble are ordered with an algorithm called HYPERRANK, which facilitates sampling and marginalization over the $n(z)$ ensemble within the cosmological likelihood Markov chains (Cordero et al. 2020).

3 DATA

3.1 DES Wide Field Survey

This work presents tomographic redshift distributions for the DES Year 3 weak lensing source catalogue, described in Gatti & Sheldon et al., (2020b). The source catalogue is a subset of the DES Year 3 Gold catalogue of photometric objects (Sevilla-Noarbe et al. 2020). After the applied selections, it consists of 100,208,944 galaxies with measured r, i, and z METACALIBRATION photometry and shapes (Sheldon & Huff 2017). We note that a subset of the selections defined in Gatti & Sheldon et al., (2020b) were motivated by achieving a more homogeneous photometric catalogue, and therefore more accurate redshift calibration. These cuts on METACALIBRATION photometry are as follows:

(i) $18 < m_i < 23.5$
(ii) $15 < m_r < 26$
(iii) $15 < m_z < 26$
(iv) $-1.5 < m_r - m_i < 4$
(v) $-4 < m_z - m_i < 1.5$

The bright limits of selections (i), (ii), and (iii) remove nearby galaxies for which no lensing signal is expected. They also remove some remaining stars that were incorrectly included in the source galaxy sample. The faint limit of these selections excludes the region of magnitude space where COSMOS-30 thirty-band photometric redshifts are found to be more biased (Laigle et al. 2016; Joudaki et al. 2019). Selections (iv) and (v) remove unphysical colors which are assumed to be caused by catastrophic flux measurement failures.

In this work, we frequently refer to this sample and its photometry as *wide (field)* data. For further details on this catalogue, we refer the reader to Gatti & Sheldon et al., (2020b).

For the DES Y3 weak lensing analysis, we exclude DES wide-field g-band data due to biases caused by difficulties in modeling the g-band point-spread function (PSF). In particular, METACALIBRATION requires an accurate PSF model to deconvolve (and subsequently reconvolve) a galaxy image from the PSF in order to determine how a galaxy image responds to artificial shear. Inadequate modeling of the PSF would lead to an imprecise constraint on the shear response R_s of each galaxy. In the g-band, such model inaccuracies are expected to result e.g. from chromatic effects on the PSF (Plazas & Bernstein 2012). Our diagnostics indeed show that PSF models are significantly less accurate in the g-band than in the redder DES filters. As a result, we do not use g-band data for any purpose that requires accurate PSF deconvolution, including the METACALIBRATION correction for selection biases. This problem precludes the use of g-band for defining redshift bins, since selection biases can only be corrected within METACALIBRATION when all selections (including the selection into a redshift bin) are made based on properties also measured on artificially sheared images, which are not available in the g-band. For further details on this challenge, see Gatti & Sheldon et al., (2020b).

3.2 DES Deep Field Survey and Artificial Wide Field Photometry

The DES Y3 Deep Fields and mock wide-field photometry for the deep-field detections are the cornerstone of SOMPZ. Full characterization of these data products are provided in Hartley, Choi et al. (2020a) and Everett et al. (2020), respectively, and we summarize requisite details here. Our inference method relies on extracting source density information from four *deep* fields named E2, X3, C3, and COSMOS (COS) covering areas of 3.32, 3.29, 1.94, and 1.38 square degrees, respectively, as shown in Fig. 2. After masking regions with artefacts such as cosmic rays, artificial satellites, meteors, asteroids, and regions of saturated pixels, 5.2 square degrees of overlap with the UltraVISTA and VIDEO near-infrared (NIR) surveys (McCracken et al. 2012; Jarvis et al. 2013) remain. This yields 2.8M detections with measured $ugrizJHK_s$ photometry with limiting magnitudes 24.64, 25.57, 25.28, 24.66, 24.06, 24.02, 23.69, and 23.58, substantially fainter than the faintest galaxies in the sample of source galaxies. In this work we frequently refer to this sample and its photometry as *deep (field)* data.

In order to relate galaxies with given deep photometry to observed lensing sources with wide photometry, we rely on the BALROG (Suchyta et al. 2016) software which injects simulated galaxies, based on the deeper photometry from the DES deep fields, into real images. For this analysis, BALROG was used to inject model profiles fit to deep-field galaxies into the broader wide-field footprint (Everett et al. 2020). After injecting galaxies into images, the output is passed into the DES Y3 photometric pipeline. Each deep-field galaxy is injected multiple times at different positions, and injected galaxies are detected equivalently to real galaxies, yielding multiple realisations of each deep-field galaxy. The output matched catalogue of 2,417,437 injection-realisation pairs containing both deep and wide photometric information is a key part of our redshift calibration inference method. This catalogue is called the *Deep/BALROG*

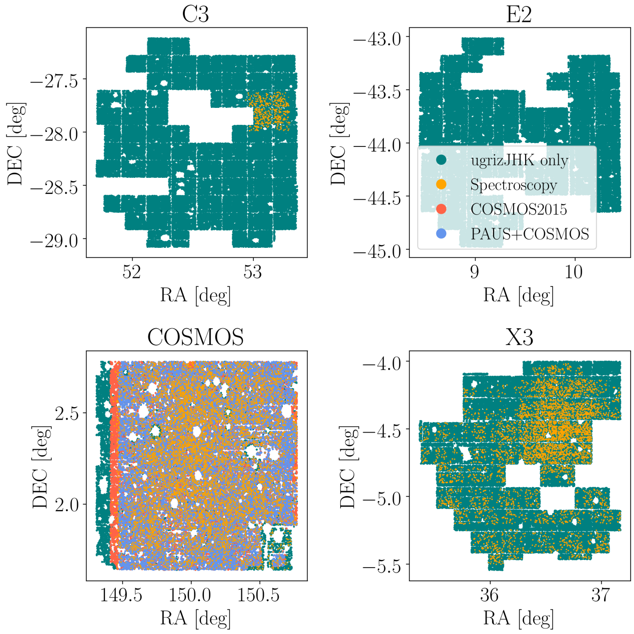

Figure 2. The four DES deep fields used for our redshift analysis. Each field has overlapping deep DES *ugriz* bands and archival *JHK* bands from the VIDEO or UltraVISTA surveys. Green points indicate DES deep-field galaxies with no spectroscopic or many-band photometric redshifts. Yellow (S), blue (C), and red (P) indicate deep-field galaxies with redshifts from spectroscopy, COSMOS2015, or PAUS+COSMOS, respectively. Missing rectangular regions are DECam CCDs on which scattered light hampered precision deep photometry.

Sample. Note that this sample contains a total of 267,229 unique deep-field galaxies having $\geq$ 1 BALROG realisation that passes the wide-field selection criteria.

3.3 Redshift samples

Our analysis relies on the use of galaxy samples with known redshift and deep-field photometry. To this end, we use catalogues of both high-resolution spectroscopic and multi-band photometric redshifts and develop an experimental design that allows us to test uncertainty in our redshift calibration due to biases in these samples. The spectroscopic catalogue we use contains both public and private spectra from the following surveys: zCOSMOS (Lilly et al. 2009), C3R2 (Masters et al. 2017, 2019), VVDS (Le Fèvre et al. 2013), and VIPERS (Scodeggio et al. 2018). We use two multi-band photo-z catalogues from the COSMOS field (Scoville et al. 2007): the COSMOS2015 30-band photometric redshift catalogue (Laigle et al. 2016), which includes 30 broad, intermediate, and narrow bands covering the UV, optical, and IR regions of the electromagnetic spectrum, and the PAUS+COSMOS 66-band photometric redshift catalogue (Alarcon et al. 2020a) from the combination of PAU Survey data (Padilla et al. 2019; Eriksen et al. 2019) in 40 narrow-band filters and 26 COSMOS2015 bands excluding the mid-infrared.

Fig. 2 shows the DES deep-field footprints (Hartley, Choi et al. 2020a) and highlights the footprint of each of the different redshift catalogues. While the two photo-z catalogues are limited to the COSMOS field, our spectroscopic compilation partially covers the COSMOS, X3 and C3 fields. Fig. 3 shows the DES i-band magnitude distribution for all galaxies with $ugrizJHK_s$ photometry and

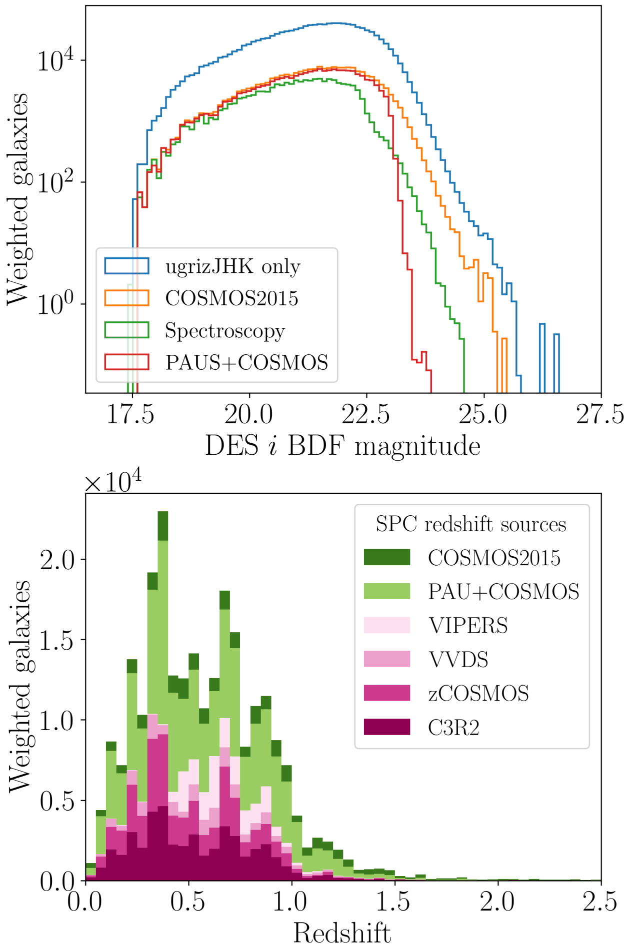

Figure 3. *Top*: Distribution of redshift samples as a function of DES i-band magnitude. Each galaxy in this histogram is weighted by all weights used in the cosmological analyses: probability of detection from Balrog, metacalibration response, and lensing weight (see § 4.1). For details on the definition of the 'Bulge Plus Disk, Fixed Ratio' (BDF) galaxy profile see Hartley, Choi et al. (2020a). *Bottom*: Distribution of redshifts used in our analysis, for one of our redshift samples SPC. This sample is defined to preferentially use redshift from spectroscopy, then `PAUS+COSMOS`, then `COSMOS2015`. Each galaxy in this stacked histogram is weighted by all weights used in the analysis: probability of detection from Balrog, metacalibration response, and lensing weight (see § 4.1).

for each of the redshift samples (for a definition of BDF magnitude see Hartley, Choi et al. 2020a). Each galaxy has been weighted by the same weight used in the cosmological analysis, which includes the galaxy detection probability from Balrog, the metacalibration response and a lensing weight (see section 4.1 for more details on these weights). While the spectroscopic compilation spans the largest area among the redshift catalogues, it is also the shallowest. The `COSMOS2015` catalogue is the deepest, but also has the lowest redshift precision. Finally, the `PAUS+COSMOS` catalogue is more precise than `COSMOS2015` and, unlike spectroscopic samples, is nearly complete in the highly relevant magnitude range of up to $i \approx 23$ but has the lowest areal coverage at faint magnitudes.

To estimate the redshift distribution of each tomographic bin, we compose three main redshift samples for which we rank the redshift information differently, meaning that for an object with redshift information from multiple origins, we choose the estimation from the highest ranked one. These redshift samples are:

- *SPC*: This sample ranks first the spectroscopic catalogue (S), then `PAUS+COSMOS` (P), and finally `COSMOS2015` (C). This sample is designed to inform an understanding of cosmological results that is minimally reliant on the `COSMOS2015` data without introducing potential selection biases such as those discussed by Gruen & Brimioulle (2017).
- *PC*: This sample ranks first the `PAUS+COSMOS` catalogue before `COSMOS2015`, and does not include spectroscopic redshifts. This sample is designed to inform an understanding of cosmological results that are maximally reliant on many-band photometric redshifts, and thus not affected by selection effects resulting from spectroscopic survey selection functions.
- *SC*: This sample ranks first the spectroscopic catalogue before `COSMOS2015`, and does not include the `PAUS+COSMOS` catalogue. This sample is designed to inform an understanding of cosmological results that are not reliant on PAU multi-band photometric redshifts.

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 concatenating the $n(z)$ samples produced from these three redshift samples. In addition to the three samples used for our fiducial analysis, we define the following alternative redshift samples that we deem less reliable. These samples are used to test the robustness of our redshift information:

- C: This sample includes only information from the `COSMOS2015` catalogue and would therefore suffer most strongly from systematic biases in these photometric redshifts.
- SPC-MB: The redshift information used in the fiducial $n(z)$ still preserves 10 per cent of the effective information from `COSMOS2015`, primarily at the faintest magnitudes, due to the relative paucity of spectroscopic redshifts of galaxies at these fainter magnitudes. To test the impact of this, our SPC-MB (SPC, Magnitude-Biased) sample is artificially constructed to bias the `COSMOS` galaxies that are not matched to spectroscopic information with a magnitude-and-redshift-dependent prescription. The prescription is as follows: we bin galaxies for which we have a spectroscopic/PAU redshift and `COSMOS2015` photometric redshift into magnitude-redshift bins with lower magnitude bin limits $[18, 21, 22.4]$ and a redshift bin width of 0.01. For each of these galaxies, we compute the redshift bias $\Delta = z_{\rm SPC} - z_{\tt COSMOS2015}$. We remove all outlying galaxies with $\Delta > 0.15$. For each magnitude-redshift bin, the mean bias $\langle\Delta\rangle$ is applied to all `COSMOS2015` galaxies for which we *do not* have a spectroscopic/PAU redshift, to yield a realistic mock spectroscopic/PAU redshift. In this way, we generate a sample of redshifts which is maximally different than a purely `COSMOS2015` sample.

These variant samples are detailed in Table 1. The impact of using these respective samples to produce redshift distributions is discussed in § 5.2. Note that we do not attempt a calibration of the DES Y3 lensing source redshift distribution that is solely informed by spectroscopic redshifts. The sample of available spectroscopic redshifts in the deep fields does not span the full $ugrizJHK_s$ color-space of the DES data. If any cell in deep color space were to only include the subset of galaxies with successful spectroscopic redshift, we expect the resulting estimates of the redshift distributions would suffer from unquantified selection biases. However, comparisons of redshift calibration between the samples used, some of which are almost a 1:1 mixture of spectroscopic and high-quality photometric

Name	% Spectra	% PAUS+COSMOS	% COSMOS2015
SC	47	0	53
PC	0	87	13
SPC	47	43	10
C	0	0	100
SPC-MB	47	43	10*

Table 1. Redshift samples used in our analysis, both in the fiducial case (SC, PC, and SPC) and in the less reliable cases (C and SPC-MB) and their relative contribution from spectroscopic data, PAUS+COSMOS and COSMOS2015. The relative contribution includes all galaxy weights used in the analysis: probability of detection from Balrog, metacalibration response, and lensing weight (see § 4.1)

redshifts, should provide robust indications of any relevant biases in the PAU or COSMOS2015 photometric redshift samples.

3.4 Simulated Galaxy Catalogues

We use the Buzzard cosmological simulations to validate aspects of our analysis. These simulations are briefly described here, and discussed comprehensively in DeRose et al. (2020a), as well as additional validation tests of the photometry in these simulations in DeRose et al. (2019).

The Buzzard simulations are galaxy catalogues that have been populated in N-body lightcones by applying the Addgals algorithm. They make use of a set of 3 independent N-body lightcones with box sizes of $[1.05, 2.6, 4.0]$ ($h^{-3}\,\mathrm{Gpc}^3$), with mass resolutions of $[0.33, 1.6, 5.9] \times 10^{11}\, h^{-1} M_{\odot}$, and spanning redshift ranges in the intervals $[0.0, 0.32, 0.84, 2.35]$ respectively. This produces a simulation that spans $10,313$ square degrees. We use the L-Gadget2 N-body code, a memory-optimized version of Gadget2 (Springel 2005), with initial conditions generated using 2LPTIC at $z = 50$.

Addgals provides simulated galaxy positions, velocities, absolute magnitudes, spectral energy distributions (SEDs), ellipticities and half-light radii for each galaxy. Positions and absolute magnitudes are assigned such that the simulated galaxies reproduce projected clustering measurements in the Sloan Digital Sky Survey Main Galaxy Sample (SDSS MGS). Likewise, SEDs are assigned from SDSS MGS using a conditional abundance-matching model DeRose et al. (2020b), that reproduces the color-and-luminosity-dependent clustering in SDSS MGS. Broad band photometry is produced from these SEDs by k-correcting them to each galaxy's rest frame, and integrating over the DES and VISTA bandpasses to produce $ugrizJHK_s$ photometry. While we find reasonably good agreement between the Buzzard photometry and that observed in our deep and wide fields, the match is by no means perfect, particularly in bluer bands and for redshifts $z > 1.2$, as illustrated in fig. 1 of DeRose et al. (2020a).

The simulations are ray-traced using Calclens using an $N_{\mathrm{side}} = 8192$ HealPix grid (Becker 2013), and angular deflections, shear, and magnification quantities are computed for each galaxy. The DES Y3 footprint mask is applied to the ray-traced simulations, resulting in a footprint with an area of 4305 square degrees. We apply a photometric error model to the mock wide-field photometry in our simulations based on a relation measured from Balrog. A weak lensing source selection is applied to the simulations using the PSF-convolved sizes and i-band SNR in order to match the non-tomographic source number density, 5.84 arcmin^{-2}, in the metacalibration source catalogue. In order to simulate a lens galaxy catalogue, we also apply the redMaGiC selection algorithm on the simulations using the same configuration as used in the Y3 data.

4 SOMPZ METHODOLOGY

We aim to determine the redshift distribution $n(z)$ of the weak lensing galaxy sample, proportional to the probability $p(z)$ of a galaxy in that sample to be at a given redshift z, by reweighting the distribution of redshifts of a sample with reliable redshift information in a suitable way that prevents selection bias and reduces sample variance. A sample of galaxies with both well-constrained redshift and deep photometry in several bands, and an additional, larger sample of galaxies with deep photometry in the same set of bands provide crucial information on how to accurately perform that weighting. In this section, we provide details of the methodology and, in addition, brief descriptions of the additional steps of DES Y3 redshift distribution calibration related to clustering redshifts (Gatti, Giannini et al. 2020a), image simulations (MacCrann et al. 2020), and shear ratios (Sánchez, Prat et al. 2020a).

4.1 Redshift Distribution Inference Formalism

Extracting the redshift information from deep, several-band photometry to estimate the redshift of an observed wide-field galaxy amounts to marginalizing over deep photometric information (Buchs et al. 2019). The probability distribution function for the redshift of a galaxy, conditioned on observed wide-field colour-magnitude $\hat{\mathbf{x}}$ and covariance matrix $\hat{\Sigma}$, and on passing a selection function $\hat{s}$, can be written by marginalizing over deep photometric colour $\mathbf{x}$ as follows:

$$p(z|\hat{\mathbf{x}}, \hat{\mathbf{\Sigma}}, \hat{s}) = \int d\mathbf{x}\; p(z|\mathbf{x}, \hat{\mathbf{x}}, \hat{\mathbf{\Sigma}}, \hat{s})\, p(\mathbf{x}|\hat{\mathbf{x}}, \hat{\mathbf{\Sigma}}, \hat{s}). \quad (1)$$

The large number of dimensions of the variables on the right-hand side of Equation 1 make these probabilities unfeasible to evaluate directly. We instead must discretize the smooth colour and colour-magnitude spaces spanned by $\mathbf{x}$ and $(\hat{\mathbf{x}}, \hat{\mathbf{\Sigma}})$ into categories c and $\hat{c}$. These c and $\hat{c}$, which we call cells, define a set of galaxy photometric *phenotypes* (Sánchez & Bernstein 2018; Buchs et al. 2019). While any of the many existing unsupervised classification or clustering algorithms can be used to categorize galaxies in this way, we use the *Self-Organizing Map* because it allows for a two-dimensional representation of the data set whose continuity facilitates interpolation and easily interpretable visualizations (Kohonen 1982, 2001; Carrasco Kind & Brunner 2014; Greisel et al. 2015; Masters et al. 2015). With this compressed information, we can marginalize over deep-field information c to write the $p(z)$ for the ensemble of galaxies associated with a particular cell $\hat{c}$ as:

$$p(z|\hat{c}, \hat{s}) = \sum_{c} p(z|c, \hat{c}, \hat{s})\, p(c|\hat{c}, \hat{s}). \quad (2)$$

After associating $\hat{c}$ with tomographic bins according to a given binning algorithm (discussed in detail in §4.3), the $n(z)$ in each tomographic bin $\hat{b}$ can be constructed by marginalizing over (i.e. summing) the constituent cells $\hat{c} \in \hat{b}$ of the tomographic bin:

$$p(z|\hat{b}, \hat{s}) = \sum_{\hat{c}\in\hat{b}} p(z|\hat{c}, \hat{s})\, p(\hat{c}|\hat{s}, \hat{b}) \quad (3)$$

$$= \sum_{\hat{c}\in\hat{b}} \sum_{c} p(z|c, \hat{c}, \hat{s})\, p(c|\hat{c}, \hat{s})\, p(\hat{c}|\hat{s}, \hat{b}). \quad (4)$$

Each galaxy is assigned to exactly one wide SOM cell and each wide SOM cell $\hat{c}$ is assigned to exactly one tomographic bin.

The redshift probability conditioned on both c and $\hat{c}$ is statistically difficult to estimate because very few galaxies will meet both conditions simultaneously. In other words, because the number of pairs $(c, \hat{c})$ is so large, each pair will have very few, if any, galaxies. However, under the assumption that the $p(z)$ for galaxies assigned to a given deep photometric cell c should not depend sensitively on the noisy wide photometry of that galaxy, we can relax the selection condition $\hat{c}$ to $\hat{b}$ (as in Equation 5) or remove this selection entirely (as in Equation 6):

$$p(z|\hat{b},\hat{s}) \approx \sum_{\hat{c}\in\hat{b}} \sum_{c} p(z|c,\hat{b},\hat{s}) p(c|\hat{c},\hat{s}) p(\hat{c}|\hat{s}) \tag{5}$$

$$\approx \sum_{\hat{c}\in\hat{b}} \sum_{c} p(z|c,\hat{s}) p(c|\hat{c},\hat{s}) p(\hat{c}|\hat{s}). \tag{6}$$

We use the approximations in Equations 5 and 6 for our fiducial measurement on the Y3 weak lensing source catalogue. In particular, for each tomographic bin, we use Equation 5 when possible (i.e. in cases for which at least one galaxy satisfies both c and $\hat{b}$), and Equation 6 otherwise. For our tests on the equivalent simulated catalogue, we use Equation 5 exclusively, discarding cases for which there is no galaxy satisfying both c and $\hat{b}$. We illustrate each factor in this equation in Fig. 4 and show the fiducial Self-Organizing Maps in Fig. 5. The validity and impact of these assumptions are discussed in §5.1.1.

The terms in this equation are estimated from the following different samples of galaxies:

(i) $p(\hat{c}|\hat{s})$ is computed from our Wide Sample, which consists of all galaxies in the DES Year 3 weak lensing source catalogue.

(ii) $p(c|\hat{c},\hat{s})$ is computed from our Deep and Balrog Samples, which consist of all detected and selected Balrog realisations of the galaxies in the Deep Sample. We call this term the *transfer function*.

(iii) $p(z|c,\hat{b},\hat{s})$ is computed from the Redshift Sample subset of the Deep Sample, for which we have reliable redshifts, 8-band deep photometry, and wide-field Balrog realisations. [1]

4.2 Weighting Redshift Distributions for Lensing Analyses

Under weak lensing shear γ, the measured galaxy ellipticity transforms as $e \to e + R\gamma$ with a shear response R. Average quantities like mean tangential shear or two-point correlation functions are thus implicitly weighted by R.

Additionally, each galaxy has an explicit lensing weight w defined to reduce the variance of the measured shear (for more detail, see Gatti & Sheldon et al., 2020b). When predicting any shear signal, the $n(z)$ must be weighted by the product of response and explicit weight, $R \times w$ (see §3.3 in MacCrann et al. 2020 for details and blending-related limitations of this approach).

4.2.1 Lensing Weighted Wide SOM Cell Occupation

The contribution of a wide cell $\hat{c}$ to the lensing signal measured by some selection $\hat{s}$ of galaxies needs to take into account the response and lensing weights of individual galaxies in $\hat{c}$. Thus, the weight of wide SOM cell $\hat{c}$ is computed with the following sum over all galaxies i assigned to that cell:

$$p(\hat{c}|\hat{s}) = \sum_{i\in\hat{c}} \frac{w_i R_i}{\sum_{j\in\hat{s}} w_j R_j}. \tag{7}$$

4.2.2 Lensing Weighted $p(z|c,\hat{b},\hat{s})$

In addition to the response and lensing weightings, each selected galaxy in the Balrog Sample must be weighted by the number of times it was detected, passed the selection $\hat{s}$, and was assigned to the same bin $\hat{b}$; this weight must also be normalized by the number of times $N_{\rm inj}$ it was injected with Balrog.

The lensing weighted $p(z)$ for a galaxy i in the Deep Sample given its assignment to a deep cell c and a wide bin $\hat{b}$ is:

$$p(z|c,\hat{b},\hat{s}) \propto \sum_{i\in(c,\hat{b})} \frac{w_i R_i p_i(z)}{N_{i,\rm inj}}, \tag{8}$$

where the sum runs over Balrog realisations i of Redshift Sample galaxies that are assigned to deep-field cell c and tomographic bin $\hat{b}$, and $p_i(z)$ is either the spectroscopic or many-band photometric redshift posterior for that galaxy.

4.2.3 Lensing Weighted Transfer Matrix

Finally, the lensing weighted transfer matrix $p(c|\hat{c},\hat{s})$ is found by similarly weighting the counts of $(c,\hat{c})$ pairs among Balrog realisations:

$$p(c|\hat{c},\hat{s}) = \frac{p(c,\hat{c}|\hat{s})}{p(\hat{c}|\hat{s})}. \tag{9}$$

The respective sums over Balrog realisations i to compute the numerator and denominator of this term are:

$$p(c,\hat{c}|s) \propto \sum_{i\in\hat{s}} \delta_{c,c_i} \delta_{\hat{c},\hat{c}_i} w_i R_i / N_{i,\rm inj} \tag{10}$$

$$p(\hat{c}|\hat{s}) \propto \sum_{i\in\hat{s}} \delta_{\hat{c},\hat{c}_i} w_i R_i / N_{i,\rm inj}. \tag{11}$$

Note that the transfer function is computed from Balrog realisations, not the full wide galaxy sample, since only for the former are both wide-field and deep-field photometry available.

4.2.4 Smooth Response Weights

As a consequence of using response to weight on a per-galaxy basis, the derived redshift distribution can carry the noise inherent in the responses themselves. This may even generate a non-physical negative distribution at some redshifts. To remedy this, the response weights are smoothed over a grid of galaxy size and signal-to-noise according to the treatment in MacCrann et al. (2020, see their appendix D). As demonstrated there on the simulated sample, this introduces an error in mean redshift (per tomographic bin) of the order of $|\Delta\bar{z}| \approx 10^{-3}$. By contrast, the effect of response weighting overall is an order of magnitude larger at $|\Delta\bar{z}| \approx 0.01$. Therefore, we can conclude that the uncertainty introduced due to smoothing the response weights is negligible with respect to the other effects at play, and that the resulting redshift distributions benefit from the reduced noise in response.

[1] This term could, in principle, be computed from the overlapping photometry of the deep and wide fields, but is much more well sampled by making use of Balrog.

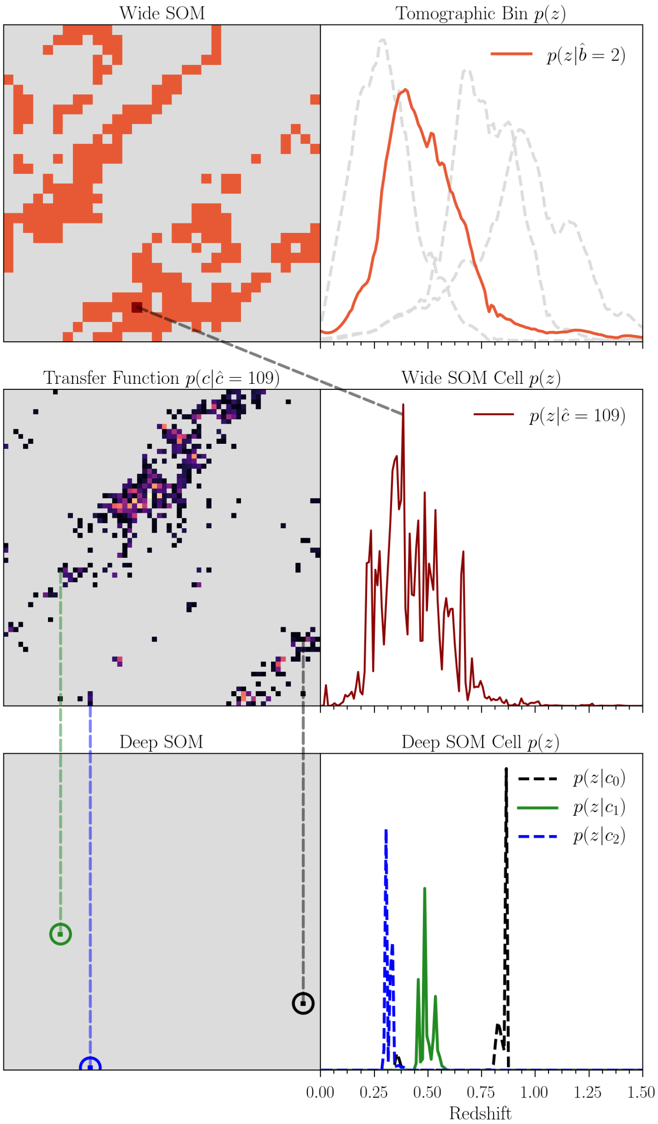

Figure 4. Visual representation of each term in the SOMPZ inference methodology. *Top left*: Wide SOM cells assigned to the second tomographic bin. *Middle left*: Transfer Function $p(c|\hat{c})$ for the selected wide SOM cell $\hat{c}$. Lighter color indicates higher values of $p(c|\hat{c})$, which corresponds to deep SOM cells with a larger number of Balrog draws in the selected $\hat{c}$. *Bottom left*: Three selected deep SOM cells c with non-zero $p(c|\hat{c})$. Different colors indicate different deep SOM cells. *Top right*: The redshift distribution of a tomographic bin. *Middle right*: One wide SOM cell in that bin. *Bottom right*: Three deep SOM cells associated with the highlighted wide SOM cell.

4.3 Construction of Tomographic Bins

Once galaxies have been categorized into phenotypes based on their photometric observations, we construct tomographic bins and assign each phenotype $\hat{c}$ to a bin. For our fiducial result, we construct these bins according to the following procedure:

(i) To construct a set of n tomographic bins $\hat{b}$, begin with an arbitrary set of $n+1$ bin edge values e_j.

(ii) Assign each galaxy in the Redshift Sample to the tomographic bin $\hat{b}$ in which the best-estimate median redshift value of its $p(z)$ (or its spectroscopic redshift z) falls. This yields an integral number of galaxies $N_{\mathrm{spec},(\hat{c},\hat{b})}$ satisfying the dual condition of membership in a wide SOM cell $\hat{c}$ and a tomographic bin $\hat{b}$. This can be written as a sum over Balrog realizations i of redshift galaxies:

$$N_{\mathrm{spec},(\hat{c},\hat{b})} = \sum_i \delta_{\hat{c},\hat{c}_i}\delta_{\hat{b},\hat{b}_i}. \tag{12}$$

(iii) Assign each wide cell $\hat{c}$ to the bin $\hat{b}$ to which a plurality of its constituent Redshift Sample galaxies are assigned:

$$\hat{b} = \{\hat{c}|\operatorname*{argmax}_{\hat{b}} N_{\mathrm{spec},(\hat{c},\hat{b})}\}. \tag{13}$$

(iv) Adjust the edge values e_j post hoc such that the numbers of galaxies in each tomographic bin $\hat{b}$ are approximately equal and repeat the procedure from step (ii) with the final edges e_j.

This procedure yields bin edges of [0.0, 0.358, 0.631, 0.872, 2.0]

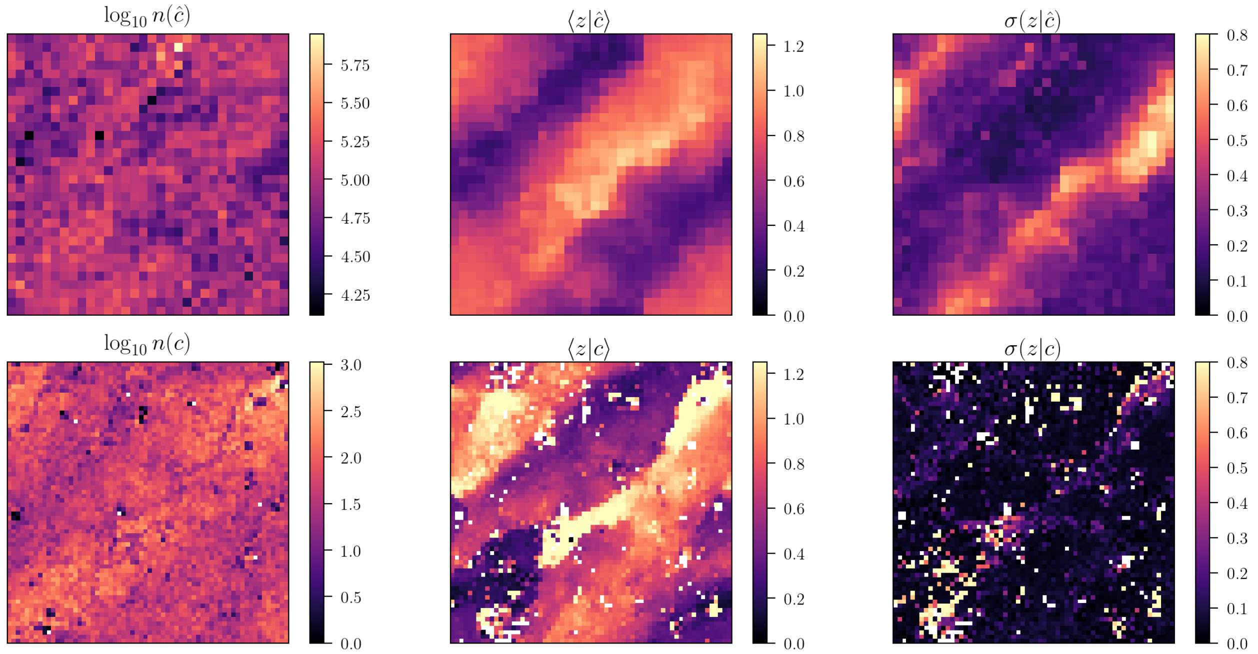

Figure 5. Visualization of the wide (top panel) and deep (bottom panel) field Self-Organizing Maps. Shown here are the total number of unique galaxies assigned to each SOM (left), the mean redshift of each cell (middle), and the standard deviation of the redshift distribution of each cell (right). White cells in the deep SOM are parts of color space for which there are no galaxies in the `COSMOS2015` sample.

for the Y3 weak lensing source catalogue. As an inconsequential result of the slight differences in the Y3 source galaxy catalogue and the simulated equivalent, the bin edges in the equivalent BUZZARD catalogue are [0.0, 0.346, 0.628, 0.832, 2.0].

4.4 Clustering redshift information

Fully independent information on the redshift distribution of the tomographic bins of our source sample is provided by its angular cross-correlation with galaxy samples of known redshift (Newman 2008; Ménard et al. 2013). Previous experiments have used this type of information to validate and/or further constrain the mean redshift of their sources (e.g. Hildebrandt et al. 2017; Davis et al. 2017; Hildebrandt et al. 2020a). A dominant confounding factor in this approach is the redshift evolution, within the tomographic bin, of the clustering bias of the source galaxies, which is highly degenerate with the mean redshift of a tomographic bin (e.g. Gatti et al. 2018; van den Busch et al. 2020).

The full description of the DES Y3 source galaxy clustering redshift analysis is given by Gatti, Giannini et al. (2020a). In brief, as reference galaxies we use the combination of redMaGiC luminous red galaxies with high quality photometric redshifts (Rozo et al. 2016; Rodríguez-Monroy et al. 2020) and spectroscopic galaxies from BOSS and eBOSS (Smee et al. 2013; Dawson et al. 2013, 2016; Ahumada et al. 2019) where they overlap the DES survey area.

There are two ways in which the clustering redshift data is used to validate and inform the redshift calibration. From comparing the clustering signal to the signal expected for a fiducial redshift distribution within a redshift range where the former exists, and assuming that clustering bias is constant as a function of source redshift, one can determine the best shift Δz of the fiducial redshift distribution and compare it to zero within its statistical and systematic uncertainty. This first method is only used as cross-check to validate the photometric estimate of $n(z)$. Alternatively, one can include the clustering redshift information in a likelihood analysis, jointly with sample variance and shot noise, that returns samples of probable redshift distributions, while marginalizing over a flexible model of source clustering bias redshift evolution. This second method is used to generate the ensemble of redshift distributions in this paper (see § 5.1.1 and § D5), and it is shown to vastly improve the accuracy of the shape of $n(z)$ derived from photometric data alone. For details of both approaches, we refer the reader to Gatti, Giannini et al. (2020a).

4.5 Image simulations and the effect of blending

The calibration as described thus far is aimed at recovering the distribution of redshifts of the dominant galaxies associated with an ensemble of detections in the DES Y3 METACALIBRATION catalogue, weighted by the individual detections' shear response. However, the measurement of a detection's shape commonly depends not just on the shear of the dominant associated galaxy, but also on the shear applied to galaxies blended with it. As MacCrann et al. (2020) show, this leads to significant response to the shear of light at other redshifts. This is best accounted for by a modification of the redshift distribution to be used for predicting lensing signals. In MacCrann et al. (2020), such a modification is derived for the DES Y3 source galaxy bins defined here. This modification reduces the mean redshift of the bins (see §2) and is calibrated with an uncertainty shown in Table 2.

We note that this correction to the $n(z)$ calibrated by photometry and clustering is expected to have non-zero shifts on the mean redshift in each tomographic bin. Additionally, several aspects of our photometric calibration strategy are validated in image simulations (see MacCrann et al. 2020, e.g. recovered true redshift distributions and their appendix D).

4.6 Shear ratio information

For physically separated pairs of a gravitational lens with sources from two bins, the ratio of the shear signals is indicative of the redshift distributions of the sources for fixed parameters of the cosmological model. In the DES Y3 lensing analyses, we use this information as an additional term in the likelihood of the lensing signals. We provide a brief summary here and refer readers to Sánchez, Prat et al. (2020a) for details of the methodology.

Gravitational shear signals on small to moderate scales are calculated for the source bins defined here around samples of redMaGiC lens galaxies. The ratio of these signals between pairs of source bins is used as the data over which likelihoods are calculated. The use of a ratio removes sensitivity of the measured shear signal to the mean matter overdensity profile around the lens galaxies but magnification, intrinsic alignments of sources relative to physically nearby lenses, and a mild dependence of the geometric shear ratio to cosmology require the likelihood to be evaluated alongside the cosmological and nuisance parameters of the Y3 lensing analyses. The shear ratio information provides constraints on this multi-dimensional parameter space in addition to, and somewhat degenerate with, the source redshift information.

For consistency tests in this paper, we use constraints from a shear-ratio-only chain to judge the consistency of the $n(z)$s with the lensing signals, from a free parameter with flat prior for the shift of the fiducial redshift distribution, at fixed cosmological parameters (see Sánchez, Prat et al. 2020a, for details). Note that for the reasons described in § 4.5, perfect agreement of the shear ratio constraint and the redshift distribution derived by means of photometry and clustering is not expected.

5 CHARACTERIZATION OF SOURCES OF UNCERTAINTY IN PHOTOMETRIC N(Z)

In this section we will characterize the uncertainties in our measurement of redshift distributions from galaxy photometry. In brief, our method consists in using secure redshifts to determine $p(z)$ in 8-band colour-space, and using the DES deep fields to determine the abundance of galaxies in 8-band color-space in the 3-band magnitude and colour-space of the lensing source galaxy sample. As a result, we must incorporate uncertainties in the redshifts used and in the estimated abundances of galaxies in each region of colour-space. The fully enumerated list of contributing sources of uncertainty is thus:

(i) Sample Variance: fluctuations in the underlying matter density field determine the abundance of observed deep field galaxies of a given 8-band colour and at a given redshift (§5.1)

(ii) Shot Noise: shot noise in the counts of deep field galaxies of a given 8-band colour and at a given redshift (§5.1)

(iii) Redshift Sample Uncertainty: biases in the redshifts of the secure redshift galaxy samples used (§5.2)

(iv) Photometric Calibration Uncertainty: uncertainty in the 8-band colour of deep field galaxies (§5.3)

(v) Balrog uncertainty: imperfections in the procedure of simulating the wide field photometry of deep field galaxies (§5.4)

(vi) SOMPZ Method Uncertainty: bias in the estimated redshift distributions relative to truth inherent to the methodology (§5.5)

We now turn to developing the formalism necessary to describe each of these uncertainties and how they affect our measured $n(z)$.

Our ultimate goal is to characterize the uncertainty in our estimation of the redshift distribution of each tomographic bin $p(z|\hat{b},\hat{s})$. It is useful to rewrite this probability (following Equation 5 and Equation 9) explicitly as a function of the four galaxy samples involved in its estimation:

$$p(z|\hat{b},\hat{s}) \approx \sum_{\hat{c}\in\hat{b}} \sum_{c} \underbrace{p(z|c)}_{\text{Redshift}} \underbrace{p(c)}_{\text{Deep}} \underbrace{\frac{p(c,\hat{c})}{p(c)p(\hat{c})}}_{\text{Balrog}} \underbrace{p(\hat{c})}_{\text{Wide}}, \tag{14}$$

where the right-hand-side terms are implicitly conditioned on the selections $\hat{b}, \hat{s}$ (not shown in Equation 14 for clarity). Note that the Balrog Sample does not inform the marginal distributions of either the deep nor the wide SOM cells, *i.e.* the Balrog Sample is not used to compute $p(c)$ or $p(\hat{c})$.

First, there is uncertainty because the galaxy samples involved are finite in both number and area. The finite area and size of the Redshift and Deep samples introduce shot noise and sample variance, which we model analytically, as explained in § 5.1. Moreover, as mentioned in § 4.1, the current finite size of the combined Redshift and Balrog Samples makes it difficult to empirically estimate $p(z|c,\hat{c})$ for all $(c,\hat{c})$ pairs, so we implement an approximate estimate for this term (Equations 5 & 6). We describe and explore the effects of this approximation on the $n(z)$ in § 5.1.1, where we validate the methodology using simulated mock catalogues.

Second, the z values of the Redshift Sample carry uncertainty. In § 5.2 we compare the redshift information that we have available from different sources in the deep fields (from spectroscopy and many-band photometry) and discuss the limitations of each. Third, the cell assignments are stochastic and thus their rate estimates are subject to shot noise as well as systematic biases. In § 5.4 we test the robustness of the Balrog transfer function against variable observing conditions across the footprint and by comparing to an alternative transfer function estimated directly with actual wide and deep photometry. Finally, in § 5.3 we examine the photometric zero-point uncertainty across the deep fields which introduces noise in the deep field colours, and we describe the method used to propagate that noise to each estimated $p(z|\hat{b},\hat{s})$.

5.1 Sample Variance and Shot Noise

The SOMPZ Bayesian formalism described in § 4.1 makes it very explicit how we estimate the redshift distribution of our four source weak lensing tomographic bins. As highlighted by Equation 14, we use the sample with the best statistics to infer each particular probability that is needed to determine the $n(z)$. Therefore, quantifying the $n(z)$ uncertainty means describing the limitations of each sample at determining each of these probabilities.

In this subsection we discuss some of the limitations of the Redshift and Deep Sample in estimating the redshift and color probability $p(z,c)$. Common limitations in redshift calibration samples are: shot noise due to finite sample size; sample variance due to large scale structure fluctuations; photometric selection effects; photometric calibration errors; spectroscopic selection effects and incompleteness; and photometric redshift errors. We explore systematic errors due to spectroscopic redshift biases or photometric redshift biases in § 5.2, and discuss errors in the deep field photometric zero-point calibration in § 5.3. We match the photometric selection effects from the wide field by injecting deep field galaxies into wide-field images using Balrog (Everett et al. 2020) and calculating the rate at which deep field galaxies would be detected and selected for the weak lensing sample. Since the deep fields are $\sim$ 1.5 magnitudes deeper than the wide field (Hartley, Choi et al. 2020a), deep-field depth variations are negligible.

Here we focus on how to estimate the shot noise and sample variance uncertainty in our deep-field samples. Typically this has been achieved by performing the same redshift estimation analysis on mocked realisations of the redshift calibration samples at different line-of-sight positions. Then, the variance and correlations in the mean redshift of the tomographic bin $n(z)$ are obtained from the variations across simulated versions of the data (*e.g.* Hildebrandt et al. 2017, 2020a; Hoyle et al. 2018; Buchs et al. 2019; Wright et al. 2020a). While we also run all methods in multiple simulated deep field realisations (§ 5.1.1), we do so as a validation and to verify if there are any remaining systematic uncertainties intrinsic to the methods themselves, but not to get an estimate of sample variance for real data. Instead, we build an analytical model of sample variance that predicts the distribution of the redshift-colour distribution in the deep fields given the data that we have observed, *i.e.* we write the distribution of a distribution: $P(p(z, c)|\text{data})$. Given this, one can propagate this distribution of uncertainties with Equation 14 and calculate the distribution of plausible $n(z)$ shapes allowed by sample variance and shot noise.

To analytically model sample variance we use a model involving *three-step Dirichlet* sampling, labelled 3sDir in this work. This approximate model of sample variance was introduced in Sánchez et al. (2020b), and is the product of three independent Dirichlet distributions. Sánchez et al. (2020b) showed in simulations that 3sDir predicted well the levels of uncertainty due to sample variance and shot noise in the first two moments of the $n(z)$ for a non-tomographic galaxy sample. We explore its performance at describing the sample variance of our four tomographic bins using the Buzzard simulations and discuss the results in § 5.1.1. We give extensive technical details of the model's mathematical formalism and application to DES Y3 in Appendix D.

In short, the 3sDir model describes the probability that galaxies belong to a redshift bin z and colour phenotype c, given that a number of galaxies have been observed to be at redshift bin z and colour phenotype c. We describe the probability in redshift and deep color $p(z, c)$ with a finite set of coefficients $\{f_{zc}\}$ indicating the probability in redshift bin z and color phenotype c, where $\sum_{zc} f_{zc} = 1$ and $0 \leq f_{zc} \leq 1$. If each Redshift Sample galaxy were representative and independently drawn, then a Dirichlet distribution parameterized by the Redshift Sample counts N_{zc} would fully characterize $p(\{f_{zc}\})$. Sample variance correlates the redshifts, however, and the more complex 3sDir model incorporates this, *i.e.* $p(\{f_{zc}\}|\{N_{zc}\}) \approx$ 3sDir.

5.1.1 Methodology Validation

In order to validate the methodology we use the suite of Buzzard simulations (§ 3.4), where we simulate the DES Y3 Wide, Deep, Balrog and Redshift Samples. First, we want to test the accuracy of the SOMPZ methodology in estimating the wide field $n(z)$ using the 8-band colour and complete redshift information available in the DES deep fields, in the same spirit as the Buchs et al. (2019) and Wright et al. (2020a) analyses. Secondly, we want to test the accuracy of the 3sDir method in describing the sample variance uncertainty in the estimated $n(z)$ within the SOMPZ framework. Buchs et al. (2019) validated the SOMPZ methodology in the context of DES, but here we use a more realistic set of simulated samples, and we also introduce 'bin conditionalization' (see Equation 5), which reduces the intrinsic bias in mean redshift of the estimated $n(z)$. Sánchez et al. (2020b) validated the 3sDir method but in a different context: for a non-tomographic sample with a different selection than the DES Y3 source sample, where all galaxies in the deep fields had redshift information, and without a transfer function to reweight the colours in the deep field.

The effect of sample variance in our estimated $n(z)$ is of particular interest. To this end we generate 300 versions of the four DES Deep Samples (where one of the four has perfect redshift information) at different random line-of-sight positions in the Buzzard simulations. For each of the 300 realisations of the deep fields, we run the SOMPZ algorithm and we obtain a $n(z)$ estimate for each tomographic bin by fixing the probabilities to the observed redshift and colour phenotype number counts. Fig. 6 shows the 300 $n(z)$s estimated by SOMPZ for each tomographic bin (light solid lines), together with their average (dark dashed lines) and the true wide field $n(z)$ in the simulation (dark solid lines with colour). We find the average simulated $n(z)$ to be extremely close to the truth. For comparison, we show the estimated $n(z)$ from data (grey dashed lines), which shows a reasonable agreement to the simulated ones. Note that the averaged $n(z)$ in simulations looks much more smooth than that from data as we are averaging out sample variance, while the $n(z)$ from data corresponds to a single realisation observed in the DES deep fields, which is affected by sample variance. In addition, to test the performance of the 3sDir method, we calculate multiple samples of $n(z)$ in each Buzzard realisation by drawing from the 3sDir likelihood; with the range of $n(z)$ samples spanning the sample variance uncertainty allowed by the 3sDir model in the redshift-colour probability.

We show technical details and specific figures of the methods validation in Appendix E, and highlight the main findings here. We find the average mean redshift (average $\bar{z}$ or $\langle\bar{z}\rangle$) across the 300 Buzzard realisations to be consistent between the SOMPZ and 3sDir methods. However, when compared to the truth we find a residual offset of $\Delta_z = [0.0051, 0.0024, -0.0013, -0.0024]$ in each bin, where $\Delta_z \equiv \langle\bar{z}^{\text{SOMPZ}}\rangle - \bar{z}^{\text{true}}$.

We take this nonzero offset as a systematic error intrinsic to the method and due to the assumption of bin conditionalization (Equation 5); we describe how we propagate this uncertainty in § 5.5.

Using the 3sDir model one can compute, in each Buzzard realisation, a distribution of mean redshift values, or $\bar{z}$, with the values allowed by sample variance and shot noise uncertainty. We find the expected value of that distribution to be unbiased with the mean redshift value from SOMPZ in individual Buzzard realisations, and in each tomographic bin. We also compare the width of the $\bar{z}$ distribution from 3sDir in each Buzzard realisation, and the width across the 300 $\bar{z}$ from SOMPZ in all realisations. We find the width predicted by 3sDir to be within 10 per cent from the width estimated with SOMPZ in the 3 lower redshift tomographic bins, but 50 per cent wider in the last tomographic bin. This is a feature of the 3sDir model, which gives an unbiased likelihood at the expense of slightly underestimating the uncertainty due to sample variance at lower redshifts, and overestimating it at higher redshifts.

We have taken great care to validate that 3sDir provides a likelihood of $n(z)$ whose mean redshift is fully compatible with the mean redshift from SOMPZ. The mean redshift serves as the leading order statistic of the $n(z)$ affecting the cosmological constraints of cosmic shear analysis, and historically the $n(z)$ has been parameterized with a fiducial $n(z)$ fixed from galaxy counts and a shift

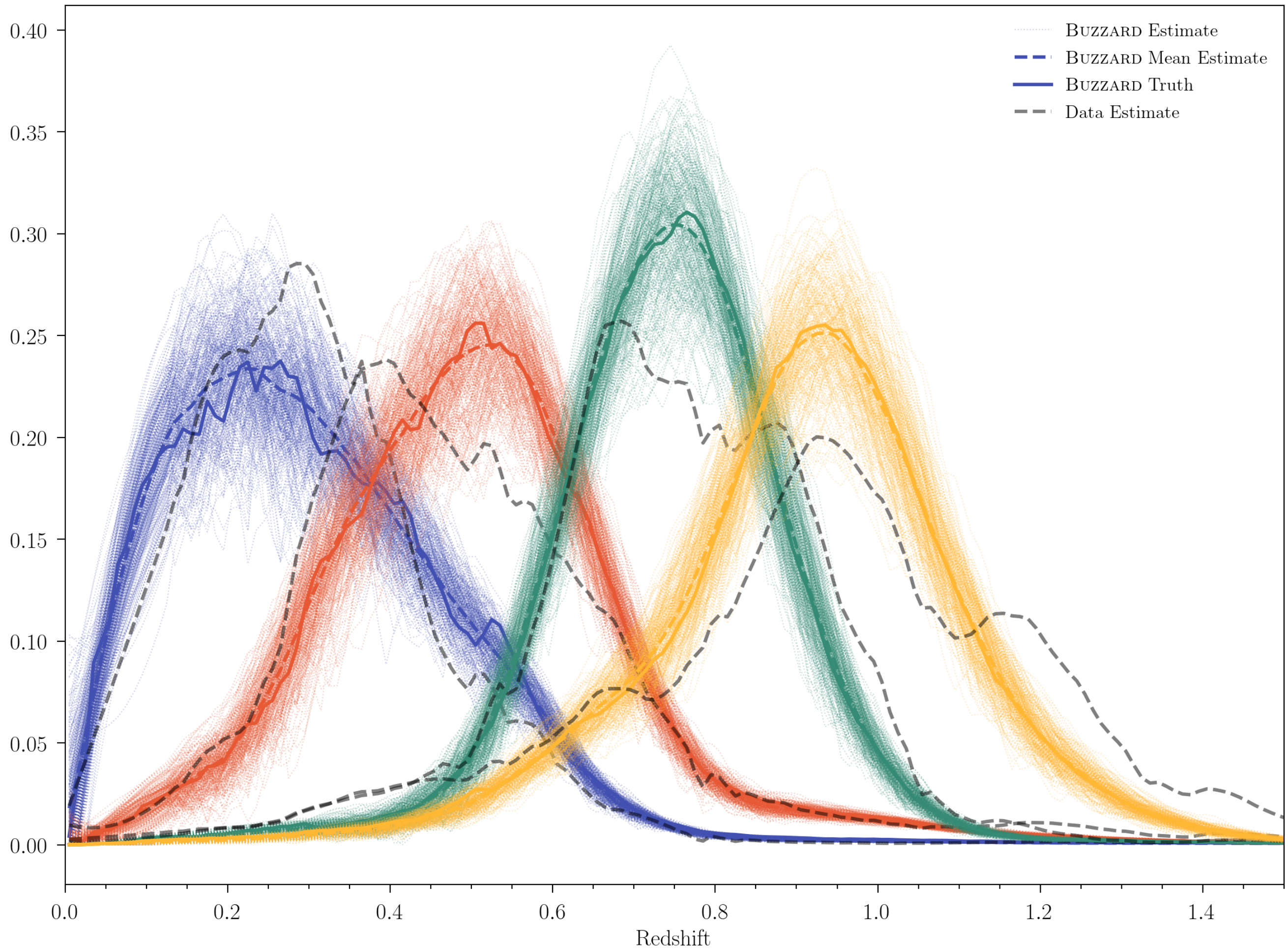

Figure 6. Estimated $n(z)$ in four tomographic bins from the Buzzard simulations using an ensemble of 300 different sets of deep fields on the Buzzard sky (colourful fine dashed lines). The similarity of the mean of the estimated $n(z)$ (colourful broad dashed lines) relative to the truth (colour broad solid lines) is a basic illustrative validation of the method. The Redshift Sample used here has 100000 galaxies drawn from 1.38 deg^2, the Deep Sample in each realisation is drawn from three fields of size 3.32, 3.29, and 1.94 deg^2, respectively from the Buzzard simulated sky catalogue. The variation in estimated $n(z)$ reflects the uncertainty of the SOMPZ method primarily due to sample variance in the deep fields. The similarity of the $n(z)$ from simulations to the fiducial result in data (gray broad dashed line) reflects the similarity of the simulated catalogue to the data.

parameter incorporating the uncertainty information (*e.g.* Bonnett et al. 2016; Hoyle et al. 2018; Troxel et al. 2018; Tanaka et al. 2018; Wright et al. 2020a,b; Hildebrandt et al. 2020a). However, here we present a change of paradigm and write a full likelihood function for the redshift distribution. Therefore, we want to make sure that no intrinsic biases are introduced in $\bar{z}$ with respect to the mean redshift of the SOMPZ methodology. There are a number of advantages to preferring a full likelihood function to a fixed $n(z)$ with a shift to its mean: it more accurately represents our uncertainty from photometry and the redshift-colour relation, it propagates higher order moment uncertainties of the redshift distribution, and it is more suitable to be combined with other sources of redshift information like clustering redshifts. As shown in van den Busch et al. (2020) and Hildebrandt et al. (2020a), combining a clustering redshift likelihood function with a fixed $n(z)$ from photometry parametrized with a shift can introduce a bias in the values of the shift parameter when the $n(z)$ is inaccurate. However, having a full likelihood over $n(z)$ presents the full set of possible $n(z)$ distributions spanning our uncertainty from photometry, with variable shapes, which can be combined, for example, with a clustering redshifts likelihood function.

In order to combine the 3sDir likelihood with a clustering redshifts (or WZ) likelihood, one can draw 3sDir $n(z)$ samples and importance sample them by the value of their WZ likelihood with each $n(z)$ draw. Even though drawing from 3sDir is very fast, this is an extremely inefficient process as the drawn $n(z)$ samples very often contain sample variance fluctuations that deliver a low WZ likelihood. By contrast, a Hamiltonian Monte-Carlo (HMC) sampler has the ability to draw from the joint combination of both likelihoods and, although drawing individual samples is slower, sampling the joint space becomes much more efficient and fast. We have defined a modified version of the 3sDir likelihood that we use in a HMC chain to sample together with the WZ likelihood. For further details on this HMC chain, see Bernstein (2020). This modified likelihood, or 3sDir-MFWZ, is defined in Appendix D5, and is by construction more sensitive to sample variance. In short, it is using less information of the colour distribution observed in the deep fields. As a result, the width of $\bar{z}$ values from 3sDir-MFWZ is larger in all redshift bins – $[78, 31, 23, 39]$ per cent larger than 3sDir in each bin, respectively (see Appendix E for further details).

5.2 Redshift Sample Uncertainty

If all galaxies with redshift information were selected independently and representatively from the source population, with no systematic uncertainties on z, then we could simply merge them all into a single sample regardless of their origin. In reality, we have overlapping redshift information from several surveys, each with unique selection criteria and biases, as described in § 3.3. We label the different redshift surveys (or combinations thereof) with R. There are different ways that we could combine information from multiple surveys. One limit is to state that one combination R is correct, but we only have some prior guess $p(R)$ about which one it is. Sampling the Bayesian posterior for $n(z)$ under this assumption is simple: we simply produce samples of f_{zc} from each survey independently by the methods of the previous subsections; and then make a final set of samples for which a fraction $p(R)$ comes from each survey. In our case we do not know that any of R is correct, but we none the less execute this marginalization over R under the principle that it is still likely to now contain the truth and also span the range of uncertainty that we have from our ignorance of the quantitative errors in different surveys.

As each of P≡`PAUS+COSMOS`, C≡`COSMOS2015`, S≡`SPEC` do not span the same region of colour space (or deep SOM cells c), as detailed in § 3.3, we define three redshift samples (SPC, PC, SC) to maximize the completeness of the redshift coverage in any sample by combining information from different sources, but as a consequence the different samples also become correlated. We still sample them separately, assigning them an equal prior probability, $p(R) = \frac{1}{3}$. We note that for those cells c that only have redshift information from one catalogue, we assume that information to be correct. Although the spectroscopic samples `SPEC` technically span a larger area than the COSMOS field, and are therefore not completed by photometric data outside this area, they are comprised of several catalogues with different selection functions in redshift and different footprints. For simplicity, we use a sample variance theory prediction which assumes an area equal to the COSMOS area in all redshift samples, which is a conservative approach.

Both `COSMOS2015` and `PAUS+COSMOS` multi-band photometric redshift catalogues report an individual redshift function $q(z)$ for each galaxy, which is not a proper posterior, but a marginalized likelihood function for different templates of galaxies. If the full likelihood of redshift, templates, and c were known, one could simultaneously and hierarchically infer the underlying f_{zc} and the redshift posterior for each galaxy (note that f_{zc} is at the same time the prior for each galaxy). However, we have found that the width of these $q(z)$ is so small compared to the redshift resolution that we have with the DES riz bands that the SOMPZ mean redshift changes by less than 10^{-3} in all tomographic bins if we treat $q(z)$ as a delta function centred at the mode of the distribution. This is also a much smaller effect than both the uncertainty from different redshift samples R and that from sample variance, so we decide to completely neglect it and treat $q(z)$ as a delta function when generating the 3sDir f_{zc} samples.

Fig. 7 shows the distribution of mean redshift values predicted by 3sDir-MFWZ for each of the samples SPC, PC, SC; which we find to be generally in agreement. We find a lower mean redshift for samples coming from SC, while samples from SPC and PC agree

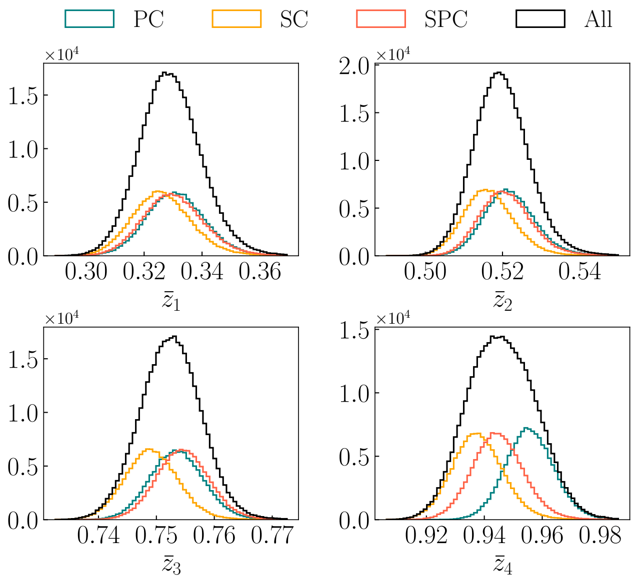

Figure 7. The distribution of mean redshift values $\bar{z}$ from 3sDir-MFWZ for each of the three redshift samples – SPC, PC and SC– on real data. We assume that their combination (shown as black histograms) contains the truth and also spans the range of uncertainty that we have from biases of the redshift samples.

very well with each other. This is in agreement with Alarcon et al. (2020a), which finds `PAUS+COSMOS` to be unbiased compared to spectra, but finds `COSMOS2015` to be systematically biased towards lower redshifts.

The small differences between SPC and PC (Fig. 7 and 8) show that our best photometric redshift and spectroscopic redshift information produce $n(z)$ samples that are largely in agreement. We check for additional robustness using the SPC-MB sample, to test the impact of the faintest galaxies whose redshift information is dominated by C. We find the $\bar{z}$ shift between SPC-MB and SPC to be smaller than ~ 0.006, adding confidence that our faintest galaxies, for which we do not have redundant redshift information, are not significantly biasing our mean redshift. This test is limited in that the applied bias as a function of magnitude (described in §3) is inferred from the available overlap between S, P and C, which is limited for faint galaxies. Fig. 8 shows the difference in $\bar{z}$ values between several samples: SPC, PC, SC, C, SPC-MB; and the average $\bar{z}$ value of SPC, PC and SC. The size of the offset in mean redshift due to using these different underlying redshift samples, as shown in Fig. 8, illustrates the value of additional follow-up spectroscopic and narrow-band photometric observational campaigns. As shown in Fig. 10, this uncertainty is a significant contributor to our overall error budget, however we find a smaller uncertainty due to this effect than Joudaki et al. (2019), who report mean offsets due to varying the redshift sample of $[0.014, -0.053, -0.020, -0.035]$ (see their table 1) in a reanalysis of the Dark Energy Survey's Year 1 analysis Hoyle et al. (2018). We attribute these differences to uncertainty on their method when applied to few-band data, as indicated by systematic offsets of 0.01–0.03 found in their MICE2 simulated analysis (see their appendix), with spectroscopic selection effects primarily responsible. In this work, we mitigate these with the inclusion of multi-band data from the Deep Fields and by creating redshift samples that are complete. The use of a larger number of bands on its own is likely to significantly reduce the systematic error

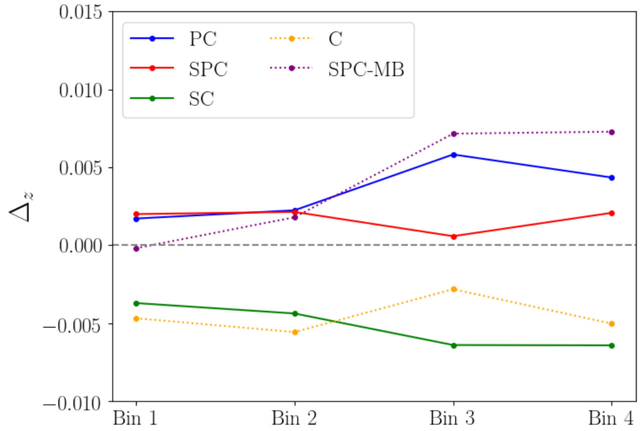

Figure 8. Mean redshift difference Δ_z in each tomographic bin for each Redshift Sample being tested: SPC, SC, PC, SPC-MB, C, relative to the average mean redshift of SPC, SC and PC. Spectroscopic catalogues are labelled as S, the `PAUS+COSMOS` catalogue as P and the `COSMOS2015` as C. SPC, SC and PC (solid lines) are the three redshift samples used in this work. SPC-MB shows the effect of extrapolating the bias between S and C to all galaxies that are still using redshift information from C in SPC (mainly faint galaxies) (see section 3.3 for a definition of the different samples used in this figure). The mean redshift is obtained by computing the $n(z)$ with SOMPZ using each sample.

due to spectroscopic selection effects (Masters et al. 2015; Gruen & Brimioulle 2017; Wright et al. 2020a).

5.3 Photometric Calibration Uncertainty

We now turn to testing the sensitivity of our measured $n(z)$ to the uncertainty in the photometric zero-points of the deep fields.

These are uncertain in large part due to field-to-field variations in photometry (for more detail see §6.4 of Hartley, Choi et al. 2020a). Consistent colours across the four deep fields to assign galaxies to deep SOM cells c, and to set the fluxes of their artificial wide field renderings in the Balrog procedure, are assumed by our SOMPZ formalism. We must propagate this uncertainty into our resulting $n(z)$.

We draw samples of deep-field magnitude zeropoint offsets from a Gaussian with standard deviation equal to the photometric zero point uncertainty in the Y3 deep fields catalogue in the relevant band as measured by Hartley, Choi et al. (2020a). For each zeropoint-error realization, we perturb all magnitudes in the mock Balrog catalogue with these zeropints and rerun the SOMPZ procedures to generate a perturbed $n(z)$. In this way we generate a full ensemble of $n(z)$s reflecting the uncertainty of our redshift calibration due to the photometric calibration.

It then remains to transfer the variation among the $n(z)$s in this simulation-based ensemble to a corresponding data-based ensemble of $n(z)$ distributions. We implement a novel application of Probability Integral Transforms (PITs) to achieve this. This PIT method transfers the variation encoded in the ensemble from simulated $n(z)$ (ensemble A) to our fiducial data result to ultimately yield a second ensemble (ensemble B). In brief, we achieve this by transferring the difference between the values of the quantile function of each realisation. For the details of this implementation, see Appendix C. The impact of this source of uncertainty is shown in Fig. 9 and 10 and documented in Table 2.

5.4 Balrog Uncertainty

Recall that we use the Balrog software (see § 3.2) to empirically estimate the relation between wide and deep field colours, $p^{\mathcal{B}}(c, \hat{c})$. The marginal distributions $p^{\mathcal{B}}(\hat{c})$ and $p^{\mathcal{B}}(c)$ from Balrog are not important, (they are measured from the Deep and Wide Samples), but the transfer function, $p^{\mathcal{B}}(c, \hat{c})/(p^{\mathcal{B}}(c)p^{\mathcal{B}}(\hat{c}))$, is a potentially important source of uncertainty. The probability of observing certain wide colours $\hat{c}$ given deep colours c depends in general on the observing conditions present in the wide field. Observing conditions vary across the DES Y3 wide-field footprint, but for our cosmic shear analysis we are interested in the *average* $n(z)$ across the footprint. Since Balrog injects galaxies with tiles placed at random across the DES Y3 wide-field footprint (covering about ~ 20 per cent of it), we are fairly sampling the distribution of observing conditions present in the wide field.

To verify that the average transfer function from Balrog is well-estimated, we bootstrap the Balrog galaxies by their injected position in the wide field. First, we create 100 subsamples by grouping the injected position using the kmeans_radec[2] software. Then, we draw the same number of subsamples with replacement, use them to recompute the average transfer function and calculate the SOMPZ $n(z)$. We repeat this process 1000 times and find the dispersion in mean redshift to be smaller than 10^{-3} in all tomographic bins. Therefore, we conclude that the internal noise in the average Balrog transfer function is negligible, and consider $f^{\mathcal{B}}_{c\hat{c}} = N^{\mathcal{B}}_{c\hat{c}}/N^{\mathcal{B}}$ to be true (with $f^{\mathcal{B}}_{c\hat{c}}$ from Equation D2).

Three of the DES deep fields (C3, E2, X3) overlap with the DES Y3 wide field, which we can use to construct a galaxy sample of position-matched wide-deep photometry pairs. We refer to this galaxy sample as wide-deep. We can empirically estimate the transfer function using the deep and wide colours observed in this catalogue. We do not use this transfer function for our fiducial result because it is computed from one realisation of the deep and wide mapping that happens with the particular wide-field observing conditions found in the deep fields, which are a much smaller area than the overall wide-field footprint. However, we can compare the Balrog and wide-deep transfer functions and their impact on the mean redshift to see if they are reasonably in agreement, subject to the limitations just mentioned.

For this test, we estimate the Balrog transfer function using only injected deep field galaxies that are also present in the wide-deep Sample. We can simulate the uncertainty due to varying observing conditions of the wide-deep transfer function using Balrog subsamples similar to the wide-deep Sample. However, Balrog galaxies are injected at one-fifth of the density of real galaxies, so we can either reproduce a wide-deep-like Sample with the same number of objects and five times the area, or the same area but one fifth of the number of objects. The uncertainty of the first will be smaller than the real uncertainty of the wide-deep, while the uncertainty of the second case would be larger. We choose the former, which yields a lower limit on the uncertainty due to variable observing conditions.

We find the difference in mean redshift $\Delta_{\bar{z}}$ between using the Balrog or the wide-deep transfer functions to be within $\sim 2\sigma_{\bar{z}}$ of the distribution of simulated wide-deep Samples: $(\Delta_{\bar{z}} \pm \sigma_{\bar{z}}) \times 10^3 = [-2.8 \pm 1.8; 3.6 \pm 1.4; 3.1 \pm 1.4; 8.2 \pm 4.8]$. Since the estimated value of $\sigma_{\bar{z}}$ is a lower limit, we conclude the difference is consistent with the expected variance from observing conditions.

[2] `https://github.com/esheldon/kmeans_radec`

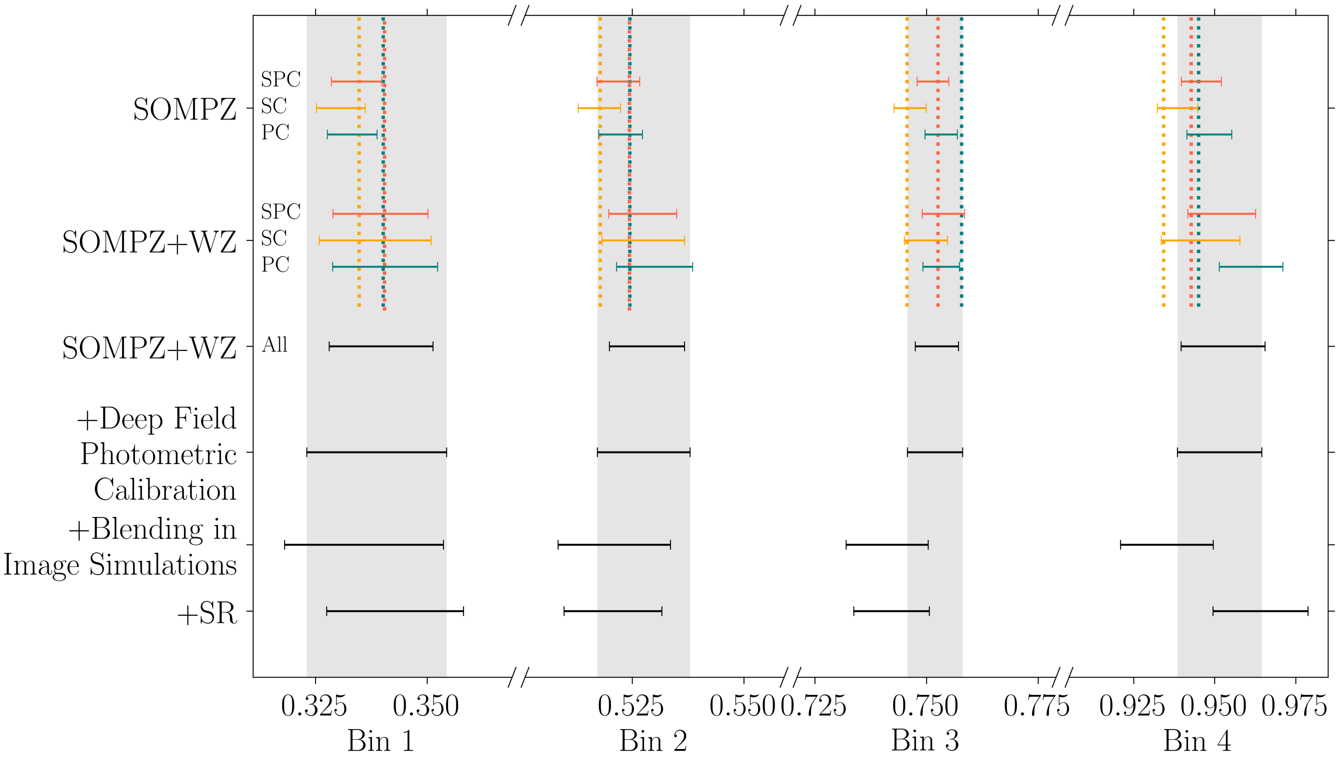

Figure 9. Mean redshifts of each tomographic bin for each of the fiducial redshift samples at each stage of the analysis. Vertical dotted lines indicate the mean redshift in each bin from the $n(z)$ output of SOMPZ, given a particular Redshift Sample. The horizontal intervals indicate the 68 per cent confidence intervals on the mean as estimated according to the methods described in §5, some of which shift the mean redshift. The larger uncertainty on the mean from the SOMPZ +WZ ensemble relative to the SOMPZ ensemble can be attributed to the different sample variance model used to combine SOMPZ with WZ (3sDIR-MFWZ, rather than 3sDIR, see §D5). For details on the modification to incorporate the effect of blending as measured by image simulations see MacCrann et al. (2020).

5.5 SOMPZ Method Uncertainty

As shown in § 5.1.1, we find an intrinsic error on the mean redshift predicted by SOMPZ when we compare it to the true mean redshift across 300 BUZZARD realisations. This inherent method uncertainty, like our zero-point calibration uncertainty, is incorporated into our $n(z)$ ensemble using the PIT method, albeit in a much simpler way: we can incorporate this uncertainty by shifting each probability integral transform by a value drawn from a Gaussian with zero mean and a standard deviation equal to the root-mean-square of these mean offset values, 0.003.

We note that this ensemble is made with an assignment of wide SOM cells to tomographic bins that is fixed for all realizations. Additionally, this method uncertainty is necessarily produced from runs with finite sample sizes, meaning there is some statistical contribution to the resulting estimate of systematic uncertainty.

5.6 Summary of sources of uncertainty

In summary, we incorporate uncertainties due to the following sources into a final ensemble of redshift distributions. These results are summarized in Table 2 and illustrated visually in Fig. 9 and 10. We note that the individual contributing sources of uncertainty do not combine linearly. We report our best estimate of the uncertainty due to each factor considered in this section in Table 2, but note that the combined uncertainty is less than the quadrature sum of these individual approximations. Fig. 10 illustrates the relative magnitude of each source of uncertainty for each bin, and the relative importance of each source of uncertainty as a function of redshift.

(i) Sample Variance: This uncertainty is estimated and incorporated into our result as part of the 3sDIR formalism. This uncertainty is a main contributor to the uncertainty budget in all of our tomographic bins.

(ii) Shot Noise: This uncertainty is estimated and incorporated into our result as part of the 3sDIR formalism.

(iii) Redshift Sample Uncertainty: This uncertainty is estimated by performing our inference with multiple different underlying redshift samples, and marginalizing over these choices by compiling their resultant $n(z)$ samples into a single ensemble. The uncertainty added by this marginalization is non-negligible in the third and fourth tomographic bins and dominant in the third tomographic bin.

(iv) Photometric Calibration Uncertainty: This uncertainty is estimated by running many times in simulations with offsets introduced to the galaxy photometry, and is incorporated into our result using PIT. This uncertainty is non-negligible in the first tomographic bin.

(v) BALROG Uncertainty: This uncertainty is estimated by replacing the transfer function $p(c|\hat{c})$ with an equivalent term estimated directly from galaxies for which we have independent deep and wide photometry, rather than using BALROG. This uncertainty is found to be negligible in all bins and is thus not propagated into our final resulting ensemble.

(vi) SOMPZ method uncertainty: This uncertainty is estimated by running many times in simulations, and is incorporated into

z^{PZ} range		Bin 1 0.0—0.358	Bin 2 0.358—0.631	Bin 3 0.631—0.872	Bin 4 0.872—2.0
$\langle z \rangle$ SOMPZ		0.332	0.520	0.750	0.944
$\langle z \rangle$ SOMPZ + WZ		0.339	0.528	0.752	0.952
Effective $\langle z \rangle$ SOMPZ + WZ + Blending [a]		0.336	0.521	0.741	0.935
Effective $\langle z \rangle$ SOMPZ + WZ + SR + Blending [b]		0.343	0.521	0.742	0.964
Uncertainty	**Method**				
Shot Noise & Sample Variance	3sDir	0.006	0.005	0.004	0.006
Redshift Sample Uncertainty	Sampling	0.003	0.004	0.006	0.006
Balrog Uncertainty	None	<0.001	<0.001	<0.001	<0.001
Photometric Calibration Uncertainty	PIT	0.010	0.005	0.002	0.002
Inherent SOMPZ Method Uncertainty	PIT	0.003	0.003	0.003	0.003
Combined Uncertainty: SOMPZ (from 3sDir)	-	0.012	0.008	0.006	0.009
Shot Noise & Sample Variance	3sDir MFWZ	0.011	0.007	0.005	0.010
Combined Uncertainty: SOMPZ (from 3sDir-MFWZ)	-	0.015	0.010	0.007	0.012
Combined Uncertainty: SOMPZ + WZ	-	0.016	0.012	0.006	0.015
Effective Combined Uncertainty: SOMPZ + WZ + Blending [a]	-	0.018	0.015	0.011	0.017
Effective Combined Uncertainty: SOMPZ + WZ + SR + Blending [b]	-	0.015	0.011	0.008	0.015

[a] These values correspond to the $n(z)$ prior used in subsequent cosmological analyses.
[b] These values correspond to the $n(z)$ posterior from a SR-only chain with fixed cosmology parameters. SR information is included in the cosmology analysis as an additional modelled data vector (see §4.6 for more details).

Table 2. Values of and approximate error contributions to the mean redshift of each tomographic bin at each stage of the analysis. We find that Sample Variance in the deep fields is the greatest contributor to our overall uncertainty for our fiducial result. The Shot Noise & Sample Variance term here is computed with the SPC sample. At low redshifts, the photometric calibration uncertainty is also significant, motivating improved work on the deep field photometric calibration. As expected, the uncertainty due to choice in Redshift Sample is a leading source of uncertainty for the third and fourth bins, motivating follow-up spectroscopic and narrow-band photometric observations. Note, the uncertainties combine non-linearly, so the combined uncertainties are not necessarily the quadrature sum of the contributing factors. Note, we label all results that incorporate blending as 'Effective' because we expect non-zero shifts on the mean redshift due to blending (as discussed in §4.5), but we do not expect non-zero shifts on the mean redshift between SOMPZ and WZ.

our result using PIT. This uncertainty is found to be negligible in all tomographic bins but it nevertheless propagated into our final resulting ensemble.

6 RESULTS

6.1 Redshift Distribution Ensembles

The results of the combined redshift calibration techniques are shown in Fig. 11. We show the ensemble produced by SOMPZ as well as the ensemble constrained by the addition of WZ and SR. Notably, our knowledge of the uncertainty on our measurement is not limited to the mean redshift, or any other finite set of moments of the distributions. Rather, the ensemble of redshift distributions effectively defines a full probability distribution function for the $p(z)$ of each histogram bin, as illustrated by the violin plots of Fig 11. Visual inspection of the SOMPZ-only distributions show that they are often not smooth functions of z. This is expected because the 3sDir likelihood (and similar 3sDir-MFWZ likelihood) aims to raise the uncertainties to the levels expected from sample variance, but does not force the resultant distributions to be smooth. The filled violins include WZ information, which heavily favors smooth $n(z)$ in the $0.1 < z < 1.0$ region where WZ data are available. The smoother nature of the ensemble after incorporating WZ demonstrates the valuable independence of that probe and its lesser reliance on biased redshift samples. SR information is included in the ensemble at the level of shifts on the mean redshift.

6.2 Consistency Of Independent Redshift Distribution Measures

Fig. 12 demonstrates consistency among the distinct sources of information used to determine $n(z)$, namely SOMPZ (colour-magnitude), WZ (clustering), and SR (shear ratios). A formal consistency check is complicated by the fact that the methods do not constrain common directions in the space of all possible $n(z)$'s. We choose to compare $\bar{z}$, the mean of $n(z)$, and define Δz here as the shift of $\bar{z}$ relative to the mean $\bar{z}$ of the SOMPZ +WZ ensemble. Even with this simplification there are complications, e.g. WZ can only constrain $n(z)$ (and hence its mean) as restricted to the range in z where adequate reference samples exist. Similarly, SR data measure redshift with an implicit weight related to lensing efficiency functions. The Δz values are plotted by always applying matching redshift windows to both SOMPZ and the sample under study.

On this basis, we find consistency between the three methods, as well as the combinations thereof. While the constraints on the mean redshift in each tomographic bin from shear ratios are broader than from SOMPZ, the relative independence of this information yields significantly more precise combined constraints on these means. The WZ constraints on $\bar{z}$ are weaker than those from SOMPZ, but as detailed in Gatti et al. (2020a), the WZ data are much more powerful in constraining the shape and smoothness of $n(z)$ than in constraining the mean. This is illustrated directly by comparing the SOMPZ ensemble to the SOMPZ +WZ ensemble in Fig. 11.

Further, because the shear signals measured in the SR analysis are subject to systematic observational effects described in MacCrann et al. (2020), we expect a certain degree of inconsistency between SR and SOMPZ. Overall, however, within the reported uncertainties we find that these three likelihood functions can be combined. As described in §4.6, the SR information is included in

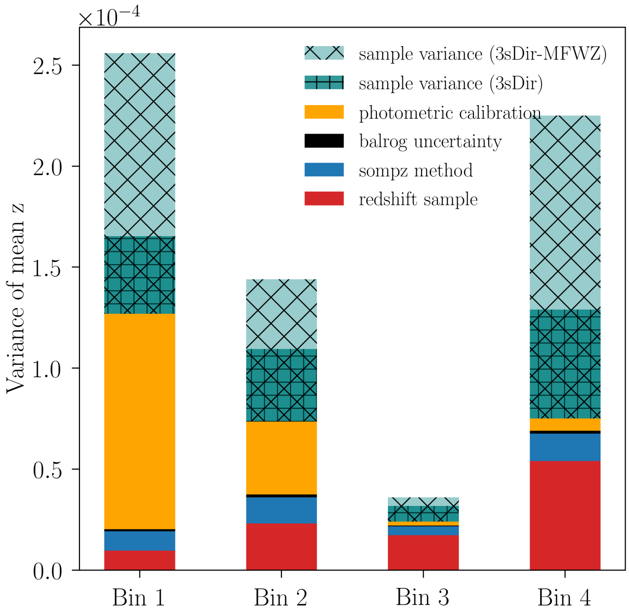

Figure 10. Variance of each source of uncertainty in each tomographic bin. Note, the bar symbols indicating contributions from 3sDir and 3sDir-MFWZ start at the same value for each bin, but 3sDir-MFWZ extends to a larger total uncertainty. The larger uncertainty estimated by 3sDir-MFWZ is an artefact of the likelihood we must use to combine $n(z)$ constraints from SOMPZ and WZ (see §D5). As shown here the Redshift Sample uncertainty becomes a larger contributor to the uncertainty for higher redshift tomographic bins. Note, the contributing sources of uncertainty combine non-linearly. As a result, to illustrate the relative magnitude of each source of uncertainty in each bin, and the relative importance of each contributing source of uncertainty as a function of redshift, we rescale the total variance in this figure to match the combined uncertainty (see Table 2).

the cosmology chains as an additional data vector, where the SR model is evaluated alongside the cosmological and nuisance parameters of the Y3 lensing analyses. As a result, the uncertainty we report in Table 2 is not a prior on the uncertainty in the mean directly used in the cosmological Markov chains, but the posterior from a SR only chain where SOMPZ+WZ is used as the $n(z)$ prior, the cosmological parameters are fixed, and the nuisance parameters are varied within their priors.

7 DISCUSSION

We derive constraints on the redshift distributions of the DES Y3 lensing source sample from the combination of wide field photometry (Sevilla-Noarbe & Bechtol et al., 2020; Gatti & Sheldon et al., 2020b), deep field photometry (Hartley, Choi et al. 2020a), artificial DES wide field photometry (Everett et al. 2020), and high quality photometric and spectrocopic redshifts, using and updating the methodology of Buchs et al. (2019). When quantifying the full uncertainty, including sample variance, the choice of Redshift Sample, calibration uncertainty of the photometric deep fields, and necessary assumptions made in the method, we find small errors ($\sigma_{\langle z \rangle} \sim 0.01$) on the mean redshift of each of the four tomographic bins. Within their joint errors, these redshift distributions are consistent with estimates from cross-correlation of galaxies with high-quality redshift reference samples (Gatti, Giannini et al. 2020a) and with the ratios of small-scale galaxy-galaxy lensing signals (Sánchez, Prat et al. 2020a), which we incorporate for a joint estimate of $n(z)$s. We also quantify the full uncertainty in $n(z)$ shape which, while for many applications being subdominant to the uncertainty in mean redshift, can now be fully propagated to parameter constraints from DES Y3 lensing analyses (Cordero et al. 2020).

While these results are encouraging, it is useful to consider the limiting factors of our analysis to inform future work. There is not one single effect. Rather, we find that our uncertainty is dominated by photometric calibration uncertainty of the deep fields at the low redshift end of our sample, and that sample variance in the deep fields and biases in the redshift samples dominate at higher redshifts. Future work should address these sources of uncertainties with targeted observing campaigns and development of new methods. In particular, the LSST science requirement specification for error on the mean redshift below 0.003 will require improvements on all counts. As discussed by Speagle & Eisenstein (2017a,b), the overlap of LSST photometry with NIR photometry from the *Euclid* survey (Laureijs et al. 2011), especially over joint deep fields, will enable methods like those used in our work for future lensing surveys (see also Capak & Cuillandre et al., (2019); Rhodes & Nichol et al., (2017)). We enumerate several opportunities for improving weak lensing redshift calibration below:

(i) **Spectroscopic follow-up targeting SOM cells**: The deep SOM constructed for this work defines a map of 8-band colour space which can be used to design future spectroscopic surveys. Many, but not all, cells defined by this SOM are populated with spectroscopic redshifts. Particularly for the 8-band or 9-band (including the NIR *Y* band, for example) colour space spanned by deeper lensing samples, the fraction of cells covered by spectroscopy is expected to decrease, the number of spectroscopic redshifts per cell is expected to decrease, and the magnitude range, at fixed colour, spanned by spectroscopic observations is expected to not match the magnitude range of the lensing sources. Follow-up observations should prioritize deep SOM cells with few redshifts, or with highly discrepant redshifts, as done by Masters et al. (2019). Larger samples per cell will be required to address any degeneracies remaining at fixed colour, and to calibrate the effects of magnitude-dependent incompleteness at fixed colour.

(ii) **Narrow-band imaging**: Narrow-band imaging can serve as a valuable complement to broadband imaging and spectroscopic surveys. Narrow-band imaging offers the benefit of measuring relatively high wavelength resolution data for surveys of large fields without selection biases. Given the intractable nature of selection biases in spectroscopic redshift samples, narrow-band imaging can serve a key role in breaking degeneracies of the colour-redshift relation for the large regions of colour-magnitude space sampled by weak lensing surveys. Given the dominance of redshift sample uncertainty in the most cosmologically constraining bins, redshifts informed by narrow-band imaging may prove key to meeting the LSST redshift calibration requirements. (see e.g. Alarcon et al. (2020a); Benitez et al. (2014))

(iii) **Improved transfer function**: A key innovation of this work is the construction of a transfer function encoding the probabilistic relation between deep and wide field photometry. While this transfer function was validated to be a negligible contribution to our uncertainty, it could be improved by injecting across a larger fraction of the wide-field survey footprint to probe more variation in survey properties. Further, as described in Everett et al. (2020), the Balrog injection procedure itself could be improved by using galaxy image cutouts, rather than CModel fits, to account for the full

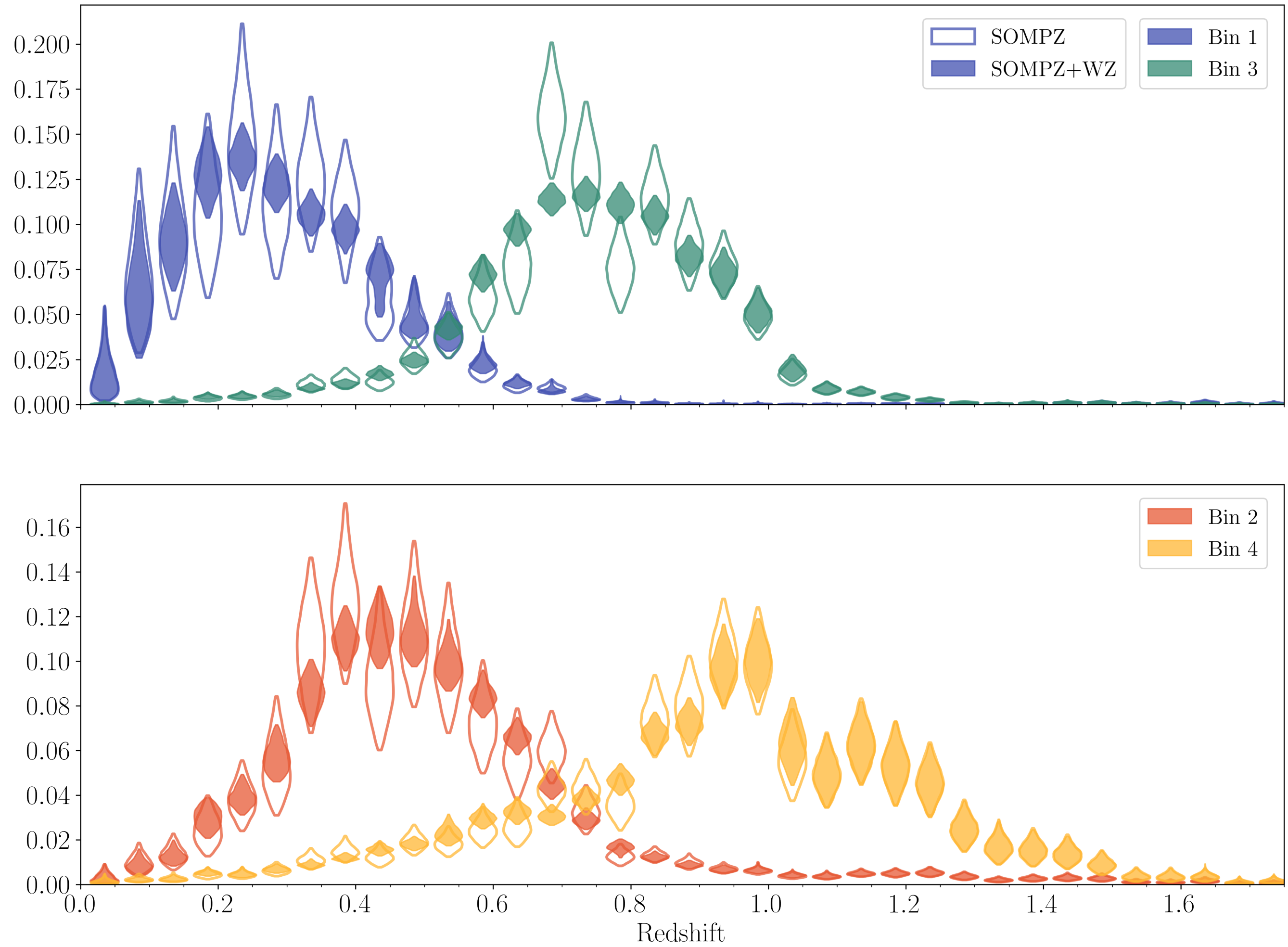

Figure 11. Visualization of the ensemble of redshift distributions in four tomographic bins, as inferred from SOMPZ only (open), and from SOMPZ combined with WZ (filled). Each violin symbol shows the 95 per cent credible interval of the probability of a galaxy in the weak lensing source sample and assigned to a given tomographic bin to have redshift z. The width at any part of a violin indicates the relative likelihood of $p(z)$ in that histogram bin. The uncertainty on p(z) is due to biases in the secure redshifts used in the analysis, sample variance and shot noise in the galaxies in the DES deep fields, photometric calibration uncertainty for the DES deep fields, and the inherent uncertainty of the methods applied. SR information is included in the ensemble at the level of shifts on the mean redshift. The low probability region of SOMPZ-only near $z \sim 0.75$ is due to an imprint of large-scale structure in the COSMOS field, as illustrated by the abundance of spectroscopic and photometric redshifts available in that region in Fig. 3.

diversity in galaxy image properties that exceeds what the CModel galaxy profile is able to describe.

(iv) **Photometric calibration** uncertainty leads to redshift uncertainty that, at low redshift, is dominated by the DES deep field u-band calibration. Reducing the uncertainty on the u-band zero point can significantly aid redshift calibration. Additional u-band data collected after the DES Y3 Deep Fields effort will enable an improved photometric calibration in future work.

(v) **Improved optimization schemes for incorporating magnitude** to the photometric information used in the Deep SOM: We construct the deep SOM with colour only, rather than colour-magnitude, following the finding by (Buchs et al. 2019) that the addition of total flux (or magnitude) to the deep SOM does not improve the performance of SOMPZ (see their section 5.1). Depending on the survey photometric noise, it is in principle possible for there to be residual correlation between redshift and magnitude at fixed 8-band colour, as shown in fig. 4 of Speagle et al. (2019), but also possible for the addition of total flux (or magnitude) to worsen results because magnitude correlates more weakly with redshift than colour. We leave it to future work to perform additional tests including magnitude in the information used in the Deep SOM.

(vi) In our analysis, the sample variance on the abundance of an 8-band colour in our Deep Field Sample is propagated throughout, but the abundance is not updated from wide-field information. A **hierarchical Bayesian model** can significantly reduce the Sample Variance on $p(c)$ by using $p(\hat{c})$ and the transfer function $p(c|\hat{c})$ to update and constrain $p(c)$ (Leistedt et al. 2016; Sánchez & Bernstein 2018). Likewise, $p(z|c)$ can be further constrained with a hierarchical Bayesian model that includes clustering information from wide field galaxies (Alarcon & Sánchez et al., 2020b).

(vii) Modeling $n(z)$ with **dependence on observing conditions**: variations in observing conditions over surveys remains a barrier to using the full non-homogeneous photometric data set collected by a given galaxy survey. To enable analysis of cosmic shear two-point functions in a survey with non-uniform depth, future work may require modeling lensing survey $n(z)$ from non-uniform catalogues (see e.g. Hoyle & Gruen et al., 2018, appendix B; Heydenreich et al. 2020). Our formalism, by explicitly evaluating the observing-

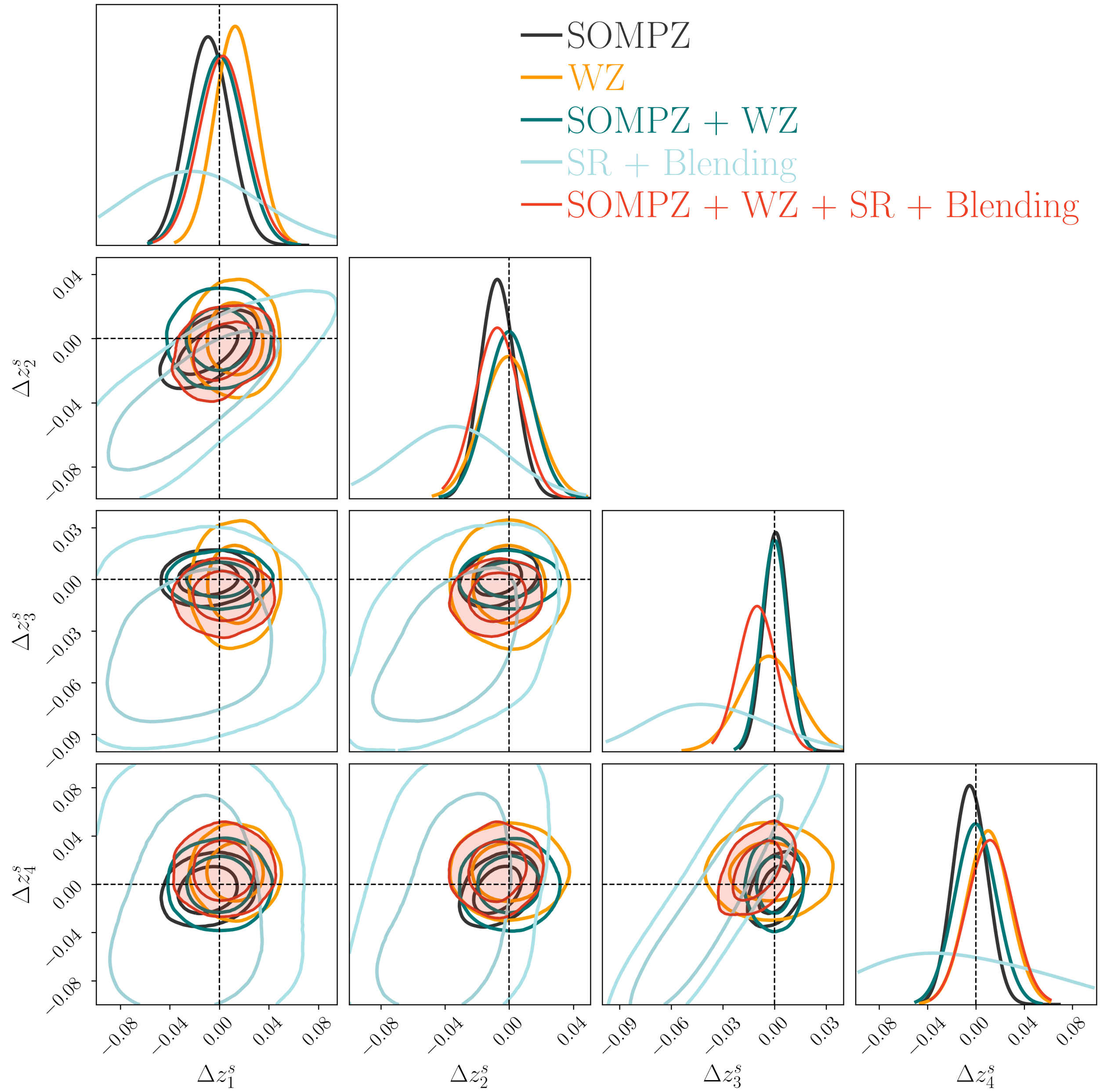

Figure 12. Consistency of the measured mean redshift in each tomographic bin from the three inference likelihoods on data. Each axis represents the difference $\Delta\bar{z}$ in mean redshift $\bar{z}$ for a particular bin relative to the mean value of $\bar{z}$ in the SOMPZ +WZ ensemble. As noted in the text, $\bar{z}$ can only be calculated from WZ and SR information using a windowed (or weighted) average over z, so this plot makes use of such windows where necessary. As shown by the light-blue contour, the inclusion of information from the ratios of the shear-position correlation functions at small scales significantly reduces the uncertainty on the mean redshift in each tomographic bin. Note the contours including SR information have additional uncertainty due to incorporating the effect of blending, thus leading to the false appearance that our combined SOMPZ +WZ+SR constraint is less constraining than SOMPZ, WZ, and SR individually.

condition-dependent transfer function $p(c|\hat{c})$ naturally extends toward this goal. Future work can use Balrog to mock galaxies at varying levels of survey depth to match non-uniform surveys.

DES Y3 has developed several new redshift calibration methods to facilitate advanced quantification of our uncertainties. Given our results, we conclude that future work for deeper lensing surveys such as DES Y6 and Stage-IV experiments such as the Legacy Survey of Space and Time (LSST) (LSST Dark Energy Science Collaboration 2012) must address these challenges to achieve the stated LSST science goal of uncertainty on the mean redshift below 0.003. In particular, we highlight the need for targeted spectroscopic and narrow-band photometric observations overlapping the LSST footprint. We emphasize the utility of our constructed SOMs to facilitate effective experimental design for such observations. Work to achieve these goals is underway, see e.g. Euclid Collaboration (2020); Masters et al. (2017).

ACKNOWLEDGEMENTS

This work was supported by the Department of Energy, Laboratory Directed Research and Development program at SLAC National Accelerator Laboratory, under contract DE-AC02-76SF00515 and as part of the Panofsky Fellowship awarded to DG. JM thanks the LSSTC Data Science Fellowship Program, which is funded by LSSTC, NSF Cybertraining Grant #1829740, the Brinson Foundation, and the Moore Foundation; his participation in the program has benefited this work. Argonne National Laboratory's work was supported by the U.S. Department of Energy, Office of High Energy Physics. Argonne, a U.S. Department of Energy Office of Science Laboratory, is operated by UChicago Argonne LLC under contract no. DE-AC02-06CH11357.

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, NFS's NOIRLab, the University of Nottingham, The Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, Texas A&M University, and the OzDES Membership Consortium.

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

The DES data management system is supported by the National Science Foundation under Grant Numbers AST-1138766 and AST-1536171. The DES participants from Spanish institutions are partially supported by MICINN under grants ESP2017-89838, PGC2018-094773, PGC2018-102021, SEV-2016-0588, SEV-2016-0597, and MDM-2015-0509, some of which include ERDF funds from the European Union. IFAE is partially funded by the CERCA program of the Generalitat de Catalunya. Research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013) including ERC grant agreements 240672, 291329, and 306478. We acknowledge support from the Brazilian Instituto Nacional de Ciência e Tecnologia (INCT) do e-Universo (CNPq grant 465376/2014-2).

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

8 DATA AVAILABILITY

The DES Y3 data products used in this work, as well as the full ensemble of DES Y3 source galaxy redshift distributions described by this work, will be made publicly available following publication, at the URL `https://des.ncsa.illinois.edu/releases`.

REFERENCES

Abbott T. M. C., et al. 2018, Phys. Rev. D, 98, 043526
Ahumada R., et al. 2019, arXiv e-prints, p. arXiv:1912.02905
Alarcon A., et al. 2020a, arXiv e-prints, p. arXiv:2007.11132
Alarcon A., et al. 2020b, MNRAS, 498, 2614
Amon A., et al., 2020, To be submitted to
Becker M. R., 2013, MNRAS, 435, 115
Benitez N., et al. 2014, preprint, (arXiv:1403.5237)
Benjamin J., et al. 2013, Monthly Notices of the Royal Astronomical Society, 431, 1547
Bernstein G., 2020, in preparation
Bonnett C., et al. 2016, Phys. Rev. D, 94, 042005
Buchs R., et al. 2019, MNRAS, 489, 820
Capak P., et al. 2019, arXiv e-prints, p. arXiv:1904.10439
Carrasco Kind M., Brunner R. J., 2014, MNRAS, 438, 3409
Cawthon R., et al. 2017, preprint, (arXiv:1712.07298)
Cordero J. P., Harrison I., et al., 2020, To be submitted to MNRAS
Cunha C. E., et al. 2012, MNRAS, 423, 909
DES Collaboration et al., 2020, To be submitted to
Davis C., et al. 2017, preprint, (arXiv:1710.02517)
Dawson K. S., et al. 2013, AJ, 145, 10
Dawson K. S., et al. 2016, AJ, 151, 44
DeRose J., et al. 2019, arXiv e-prints, p. arXiv:1901.02401
DeRose J., et al., 2020a, To be submitted to
DeRose J., et al., 2020b, To be submitted to
Eriksen M., et al. 2019, MNRAS, 484, 4200
Euclid Collaboration 2020, A&A, 642, A192
Everett S., et al., 2020, To be submitted to MNRAS
Gatti M., et al. 2018, MNRAS, 477, 1664
Gatti M., Giannini G., et al., 2020a, To be submitted to MNRAS
Gatti M., et al. 2020b, arXiv e-prints, p. arXiv:2011.03408
Greisel N., et al. 2015, MNRAS, 451, 1848
Gruen D., Brimioulle F., 2017, MNRAS, 468, 769
Hartley W. G., Choi A., et al., 2020a, To be submitted to MNRAS
Hartley W. G., et al. 2020b, MNRAS, 496, 4769
Heydenreich S., et al. 2020, A&A, 634, A104
Heymans C., et al. 2012, Monthly Notices of the Royal Astronomical Society, 427, 146–166
Heymans C., et al. 2020, arXiv e-prints, p. arXiv:2007.15632
Hikage C., et al. 2019, PASJ, 71, 43
Hildebrandt H., et al. 2012, MNRAS, 421, 2355
Hildebrandt H., et al. 2017, MNRAS, 465, 1454
Hildebrandt H., et al. 2020a, arXiv e-prints, p. arXiv:2007.15635
Hildebrandt H., et al. 2020b, A&A, 633, A69

Hoyle B., et al. 2018, MNRAS, 478, 592
Huterer D., et al. 2006, MNRAS, 366, 101
Huterer D., Cunha C. E., Fang W., 2013, Monthly Notices of the Royal Astronomical Society, 432, 2945
Jain B., Taylor A., 2003, Phys. Rev. Lett., 91, 141302
Jarvis M. J., et al. 2013, MNRAS, 428, 1281
Johnson A., et al. 2017, MNRAS, 465, 4118
Joudaki S., et al. 2019, arXiv e-prints, p. arXiv:1906.09262
Kohonen T., 1982, Biological Cybernetics, 43, 59
Kohonen T., 2001, Self-organizing maps, 3rd edn. Springer series in information sciences, 30, Springer-Verlag, Berlin, Heidelberg, doi:10.1007/978-3-642-56927-2
LSST Dark Energy Science Collaboration 2012, arXiv e-prints, p. arXiv:1211.0310
Laigle C., et al. 2016, ApJS, 224, 24
Laureijs R., et al. 2011, arXiv e-prints, p. arXiv:1110.3193
Le Fèvre O., et al. 2013, A&A, 559, A14
Leistedt B., Mortlock D. J., Peiris H. V., 2016, MNRAS, 460, 4258
Lilly S. J., et al. 2009, ApJS, 184, 218
Lima M., et al. 2008, MNRAS, 390, 118
Lupton R. H., Gunn J. E., Szalay A. S., 1999, AJ, 118, 1406
MacCrann N., et al. 2020, To be submitted to MNRAS
Mandelbaum R., et al. 2005, Monthly Notices of the Royal Astronomical Society, 361, 1287–1322
Masters D., et al. 2015, ApJ, 813, 53
Masters D. C., et al. 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., et al. 2013, preprint, (arXiv:1303.4722)
Morrison C. B., et al. 2017, MNRAS, 467, 3576
Newman J. A., 2008, ApJ, 684, 88
Padilla C., et al. 2019, AJ, 157, 246
Plazas A. A., Bernstein G., 2012, PASP, 124, 1113
Porredon A., et al. 2020, arXiv e-prints, p. arXiv:2011.03411
Porredon A., et al., in prep., To be submitted to PRD
Prat J., et al. 2018, Phys. Rev. D, 98, 042005
Prat J., et al. 2019, MNRAS, 487, 1363
Prat J., et al., 2020, To be submitted to MNRAS
Rhodes J., et al. 2017, ApJS, 233, 21
Rodríguez-Monroy M., et al., 2020, To be submitted to MNRAS
Rozo E., et al. 2016, MNRAS, 461, 1431
Samuroff S., et al. 2017, MNRAS, 465, L20
Sánchez C., Bernstein G. M., 2018, preprint, (arXiv:1807.11873)
Sánchez C., Prat J., et al., 2020a, To be submitted to MNRAS
Sánchez C., et al. 2020b, arXiv e-prints, p. arXiv:2004.09542
Schmidt S. J., et al. 2020, MNRAS, 499, 1587
Scodeggio M., et al. 2018, A&A, 609, A84
Scoville N., et al. 2007, ApJS, 172, 1
Secco L. F., Samuroff S., et al., 2020, To be submitted to
Sevilla-Noarbe I., et al. 2020, arXiv e-prints, p. arXiv:2011.03407
Sheldon E. S., Huff E. M., 2017, ApJ, 841, 24
Smee S. A., et al. 2013, AJ, 146, 32
Speagle J. S., Eisenstein D. J., 2017a, MNRAS, 469, 1186
Speagle J. S., Eisenstein D. J., 2017b, MNRAS, 469, 1205
Speagle J. S., et al. 2019, MNRAS, 490, 5658
Springel V., 2005, MNRAS, 364, 1105
Suchyta E., et al. 2016, MNRAS, 457, 786
Tanaka M., et al. 2018, PASJ, 70, S9
Tessore N., Harrison I., 2020, The Open Journal of Astrophysics, 3, 6
Troxel M. A., et al. 2018, Phys. Rev. D, 98, 043528
Wright A. H., et al. 2020a, A&A, 637, A100
Wright A. H., et al. 2020b, A&A, 640, L14
van den Busch J. L., et al. 2020, arXiv e-prints, p. arXiv:2007.01846

APPENDIX A: APPENDIX ON SOM

Fig. A1 and Fig. A2 show the i-band magnitude and colours of each wide and deep SOM cell, respectively. Given that the SOM training algorithm attempts to construct a smooth map in the full parameter space of the training inputs, we can interpret stark differences in adjacent cells as indirect indicators of degeneracies in the colour-redshift relation. Comparison with the upper right panel of Fig. 5 indicates that the wide SOM cells with the broadest $p(z|\hat{c})$ tend to be cells with overall fainter galaxies, supporting the intuitive conclusion that our redshift constraints are weaker for fainter galaxies.

APPENDIX B: SOMPZ IMPLEMENTATION DETAILS

We enumerate here several technical details about the implementation of SOMPZ:

B1 SOM Training

We note a few details about the SOM training algorithm here and refer the reader to Buchs et al. (2019) for a full treatment. We use the magnitude scale defined by Lupton et al. (1999) for our SOM training, which we call 'luptitude'. The input vector of the Deep SOM is chosen to be a list of lupticolors with respect to the luptitude in i-band:

$$\mathbf{x} = (\mu_{x_1} - \mu_i, ..., \mu_{x_7} - \mu_i),$$

where the bands x_1 to x_7 are $ugrzJHK$. For the input vector of the Wide SOM, we also use lupticolors with respect to the luptitude in i band, and we add the luptitude in i band:

$$\hat{\mathbf{x}} = (\mu_i, \mu_r - \mu_i, \mu_z - \mu_i).$$

In the case of the wide field, where only few colors are measured, Buchs et al. (2019) find empirically that addition of the luptitude improves the performance of the scheme.

B2 Deep SOM Training Sample

We find that training the deep SOM only on deep galaxies whose Balrog realisations are detected and selected by the weak lensing source selection function leads to a SOM with more precise $p(z)$.

B3 Smoothing

The $n(z)$ output from SOMPZ includes extreme unphysical differences in $n(z)$ between adjacent redshift bins due to shot noise in the sample of redshifts used. While this accurately reflects our estimate of the $n(z)$ given the information available, it inconveniently increases the computation time needed to integrate over the $n(z)$ in cosmological likelihood Markov chains. As a result, we smooth the SOMPZ output $n(z)$ with a Savitzky–Golay filter.

B4 High Redshift Pile-up

The redshift samples used contain galaxies with $p(3 < z < 6) > 0$. Although the resulting SOMPZ $n(z)$ with probability density at redshifts greater than three accurately reflect our estimate of the $n(z)$ given the information available, the relatively small probability in

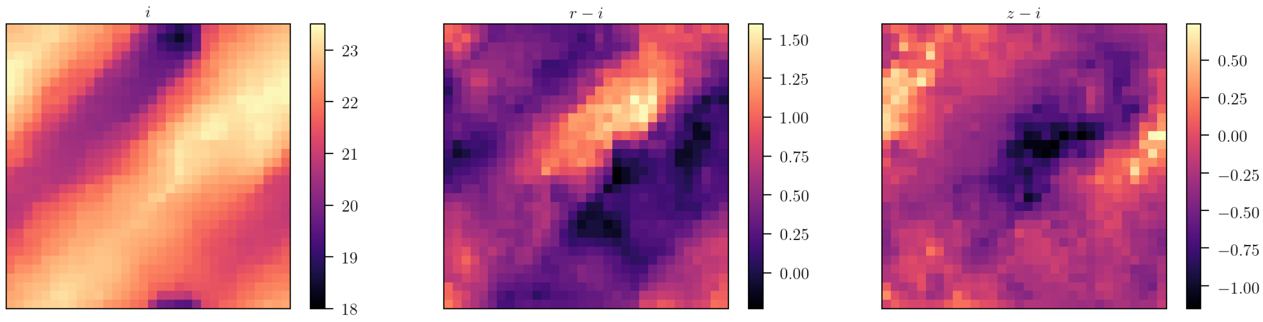

Figure A1. Visualization of the wide Self-Organizing Map. Shown here are the mean i-band magnitude (left), the mean $r - i$ colour (middle), and the mean $z - i$ colour (right) of each cell in the wide SOM. The implementation of SOMs used in our analysis generates a toroidal map; in other words, the left and right edges of each map correspond to the same region of colour-magnitude space, as do the upper and lower edges.

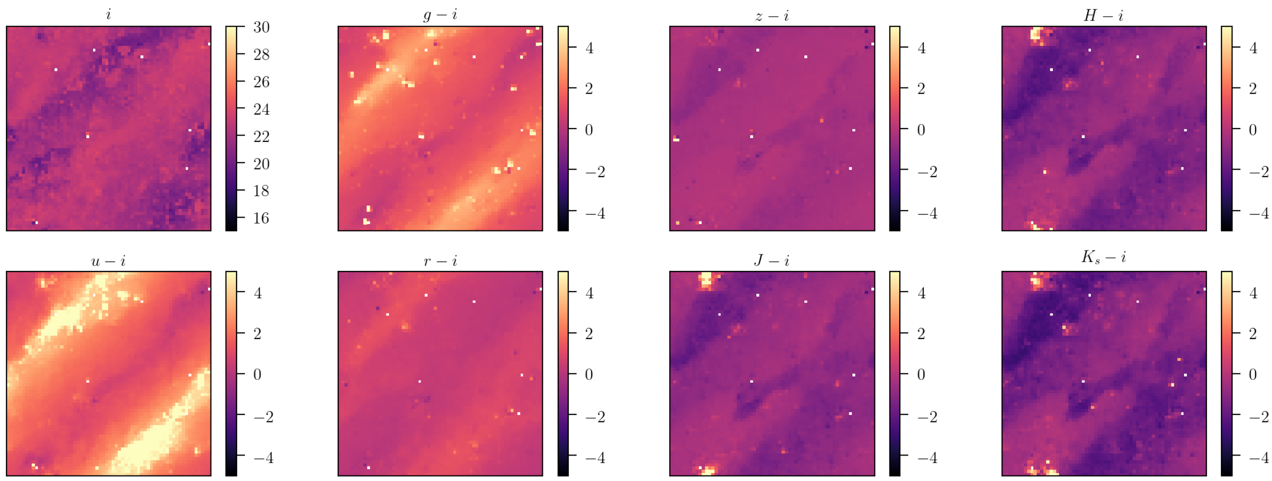

Figure A2. Visualization of the deep field Self-Organizing Map. Shown here are the mean i-band magnitude (upper left) of each cell of the deep SOM, as well as each colour used in the deep SOM training. The implementation of SOMs used in our analysis generates a toroidal map; in other words, the left and right edges of each map correspond to the same region of colour-magnitude space, as do the upper and lower edges.

this high redshift region inconveniently increases the computation time needed to integrate over the $n(z)$ in cosmological likelihood Markov chains. To mitigate this effect, we shift all probability greater than a cut-off value of three to the final redshift bin at 3. The amount of probability beyond redshift 2 is less than one per cent in all cases: [0.0096, 0.0062, 0.0021, 0.0077].

B5 Ramping

The DES Y3 3×2pt. cosmological analyses sample over this ensemble in cosmological likelihood inference Markov chains. Importantly, we find that non-zero probability density near zero redshift ($p(z \approx 0) > 0$) significantly increases the computation time necessary to efficiently sample parameter space due to high sensitivity of the intrinsic galaxy alignment (IA) model. This effect is most pronounced for the lowest redshift bin because this bin has the greatest $p(z \approx 0)$. Given the relatively little cosmological information in this bin, we alter the ensemble of redshift distributions post hoc to manually reduce $p(z \approx 0)$ by multiplying the $p(z)$ up to $z = 0.055$ with a linear function that serves as an effective prior.

B6 Deep Field Noise Differentiation

We note a test run on simulations in which the deep field photometric noise is set to different levels in COSMOS and the other DES deep fields. As for all runs in simulations, we set the noise levels by measuring the median noise levels in the corresponding data catalogues. We find no measurable difference on the mean redshift with this added realism relative to previous work.

APPENDIX C: PIT IMPLEMENTATION

This subsection (5.3) is dedicated to describing this novel method for transferring the variation in $n(z)$ as a result of our photometric calibration uncertainty and how we implement this method in practice.

The conceptual procedure for achieving this is described below and described in greater detail in Myles et al. in prep. We begin by computing the inverse cumulative distribution function (i.e. quantile function) F_i^{-1} for each simulated realisation $n_i(z)$ in the ensemble labelled A. For a tomographic bin $\hat{b}$, this can be written as

$$F^{-1}_{i,\hat{b}}(p) = \{z : F_{i,\hat{b}}(z) = p\} \quad \text{with} \quad F_{i,\hat{b}}(z) = \int_{-\infty}^{z} n_{i,\hat{b}}(z')dz' \, . \tag{C1}$$

We construct each PIT by computing the difference of the inverse CDF of a given realisation $F^{-1}_{i,\hat{b}}$ with the average inverse CDF of the ensemble:

$$\mathrm{PIT}_{i,\hat{b}} = F^{-1}_{i,\hat{b}} - \langle F^{-1}_{\hat{b}} \rangle. \tag{C2}$$

Subtracting the average inverse CDF ensures that the mean redshift is not changed by the PIT. This is necessary because each realisation, in addition to having some zero-point offset introduced, is drawn from a noisy distribution due to (i) deep field photometric noise and (ii) mock-Balrog realisation noise in simulations. As a result of (i) and (ii), there would be a non-zero mean shift of the mean z shifts of the ensemble of PITs if not for subtracting $\langle F^{-1}_{\hat{b}} \rangle$.

We apply these transformations to the data by simply adding each PIT to the inverse CDF of the fiducial data $n(z)$, $F^{-1}_{\text{data, fiducial}}$. The PIT due to one draw of zero-point offsets is determined and applied jointly to all tomographic bins:

$$F^{-1}_{i,\hat{b},\text{data}} = F^{-1}_{\hat{b},\text{data, fiducial}} + \mathrm{PIT}_{i,\hat{b}}. \tag{C3}$$

Given this ensemble of inverse CDFs of the data $n(z)$, we construct the corresponding ensemble of data $n(z)$ by taking the inverse to yield CDFs, then differentiating:

$$n_{i,\hat{b}}(z) = \frac{\mathrm{d}}{\mathrm{d}z}\left(F_{i,\hat{b},\text{data}}\right). \tag{C4}$$

Implementing the PIT offers two insights into our calibration uncertainty: first, our uncertainty is driven by the u-band calibration, and second, we find n(z) uncertainty increases at wavelengths corresponding to photometric filter transitions of the 4000Å break, as shown in Fig. C1.

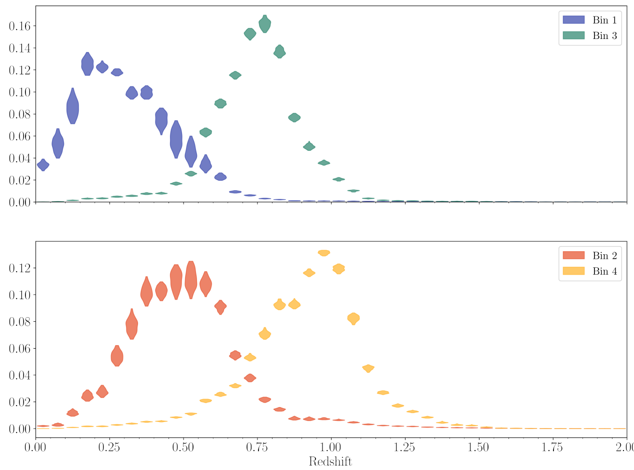

Figure C1. Impact of the deep field photometric zero-point offset on estimated $n(z)$. The spread in values of $n(z)$ for any given histogram bin here reflect the propagated impact of the photometric zero-point uncertainty on $n(z)$. We determine the offset used in each band for each realisation by sampling a multivariate Normal distribution with standard deviations set to the zero-point uncertainty in each band. As shown here, some redshifts have much larger spread in $n(z)$ than others. The uncertain region in the interval $0 < z < 0.2$ corresponds to a redshifted 4000Å-break between 400 and 480nm, which is in the DES g-band filter. Likewise, the interval $0.4 < z < 0.6$ corresponds to a redshifted 4000Å-break between 560 and 640nm, which is in the DES r-band filter, and the transition to the i-band filter occurs at $z \sim 0.75$.

APPENDIX D: 3sDir MODEL

Here we describe the formalism we use to model shot noise and sample variance in the redshift-colour probability from the deep fields to propagate it through to the redshift distribution of a tomographic bin.

Regarding the notation for the probability of redshift and colour space, note that c and $\hat{c}$ are discrete variables, denoting regions in a partitioning of colour space. Following Leistedt et al. (2016, see Eq.1-2 and §2), we will also adopt a piece-wise constant representation of probabilities in redshift space (essentially a probability histogram). In other words, we define any probability of a galaxy in our sample having redshift z, $p(z)$, with a finite set of coefficients f_i of step functions Θ,

$$p(z) \equiv \sum_i \frac{f_i}{z_i - z_{i-1}} \times \Theta(z - z_{i-1})\Theta(z_i - z). \tag{D1}$$

In the remainder of the work, we will use the symbol z to represent the discrete index of redshift bins divided at the z_i bin edge values. Given this notation, we can represent the joint probability of colour and redshift with the set of coefficients $\{f_{zc}\}$.

We denote each dataset D as: $\mathcal{W}$ for Wide, $\mathcal{D}$ for Deep, $\mathcal{B}$ for Balrog and $\mathcal{R}$ for Redshift. Note that $\mathcal{R} \subset \mathcal{D}$, and that B contains several mock realisations of galaxies from $\mathcal{D}$ which have been injected and measured as in $\mathcal{W}$ using Balrog (see § 3). We denote a set of coefficients f that has been inferred from a dataset D as f^D. For example, the coefficients of the joint redshift and colour distribution inferred from the Redshift Sample is denoted by $f^{\mathcal{R}}_{zc}$.

D1 Shot Noise

We begin by rewriting Equation 14 using the f coefficients notation, indicating which sample is used to infer each of the coefficients:

$$p(z|\hat{b},\hat{s}) \approx \sum_{\hat{c}\in\hat{b}} \sum_{c} \frac{f^{\mathcal{R}}_{zc}}{f^{\mathcal{R}}_{c}} f^{\mathcal{D}}_{c} \frac{f^{\mathcal{B}}_{c\hat{c}}}{f^{\mathcal{B}}_{c} f^{\mathcal{B}}_{\hat{c}}} f^{\mathcal{W}}_{\hat{c}}, \tag{D2}$$

where $f^{\mathcal{R}}_{c} = \sum_z f^{\mathcal{R}}_{zc}$, $f^{\mathcal{B}}_{c} = \sum_{\hat{c}} f^{\mathcal{B}}_{c\hat{c}}$ and $f^{\mathcal{B}}_{\hat{c}} = \sum_c f^{\mathcal{B}}_{c\hat{c}}$. Following Leistedt et al. (2016, see Section 3.1), we want to infer the parameters $(\{f^{\mathcal{R}}_{zc}\}, \{f^{\mathcal{B}}_{\hat{c}c}\}, \{f^{\mathcal{D}}_{c}\})$ from the following sets of galaxy data

- $D^{\mathcal{R}} = \{z_g, c_g\}^{\mathcal{R}}$ for $g = 1\ldots N^{\mathcal{R}}$,
- $D^{\mathcal{B}} = \{\hat{c}_g, c_g\}^{\mathcal{B}}$ for $g = 1\ldots N^{\mathcal{B}}$,
- $D^{\mathcal{D}} = \{c_g\}^{\mathcal{D}}$ for $g = 1\ldots N^{\mathcal{D}}$, and
- $D^{\mathcal{W}} = \{\hat{c}_g\}^{\mathcal{W}}$ for $g = 1\ldots N^{\mathcal{W}}$,

Let's start by assuming that the properties of these galaxies are known (*i.e.* they are noiseless). Consider a scenario in which we ignore line-of-sight density variance, redshift errors, zero-point errors and other systematic uncertainties. In this scenario, a sufficient statistic for inferring the coefficients $(\{f^{\mathcal{R}}_{zc}\}, \{f^{\mathcal{B}}_{\hat{c}c}\}, \{f^{\mathcal{D}}_{c}\})$, is the count of galaxies in each of the joint bins of redshift and SOM cells. Let's take for example the Redshift Sample, which we can reduce to the counts $D^{\mathcal{R}} \to \{N^{\mathcal{R}}_{zc}\}$ of galaxies detected in redshift bin z and deep SOM cell c.

From Bayes' Theorem, the probability of these coefficients

$\boldsymbol{f}^{\mathcal{R}} \equiv \{f^{\mathcal{R}}_{zc}\}$ given the observed galaxy counts can be written as

$$\begin{aligned} p(\boldsymbol{f}^{\mathcal{R}}|D^{\mathcal{R}}) &\propto p(D^{\mathcal{R}}|\boldsymbol{f}^{\mathcal{R}})\, p(\boldsymbol{f}^{\mathcal{R}}) \\ &= p(\{N^{\mathcal{R}}_{zc}\}|\boldsymbol{f}^{\mathcal{R}}) p(\boldsymbol{f}^{\mathcal{R}}) \\ &= p(\boldsymbol{f}^{R}) \prod_{zc} \left(f^{\mathcal{R}}_{zc}\right)^{N^{\mathcal{R}}_{zc}}, \end{aligned} \tag{D3}$$

and similarly for the other two sets of coefficients. The likelihood function of the binned data $p(\{N^{\mathcal{R}}_{zc}\}|\boldsymbol{f}^{\mathcal{R}})$ is a multinomial distribution by definition, under the assumption of independent selection of each galaxy. The conjugate prior for a multinomial likelihood function is a Dirichlet distribution, so if we choose our prior $p(\boldsymbol{f}^{\mathcal{R}})$ to be a Dirichlet distribution with rate parameters $\alpha^{\mathcal{R}}_{zc} = \epsilon$, then the posterior is also a Dirichlet distribution which depends only on the galaxy number counts (which makes the posterior analytical and easy to sample from). The Dirichlet distribution is also a natural prior because it enforces the constraints that $f^{\mathcal{R}}_{zc} > 0$ for all zc, $\sum f^{\mathcal{R}}_{zc} = 1$ (required for any probability), and is invariant under any rearrangement of the f's (it is agnostic to the meaning of the bins). The posterior Dirichlet distribution is

$$\begin{aligned} p(\boldsymbol{f}^{\mathcal{R}}|\{N^{\mathcal{R}}_{zc}\}) &= \mathrm{Dir}\left(\boldsymbol{f}^{\mathcal{R}}; \{\alpha_{zc}\}\right) \\ &= \mathrm{Dir}\left(\boldsymbol{f}^{\mathcal{R}}; \{\alpha_{zc} = N^{\mathcal{R}}_{zc} + \epsilon\}\right) \\ &\propto \delta\left(\sum_{zc} f^{\mathcal{R}}_{zc} - 1\right) \prod_{zc} (f^{\mathcal{R}}_{zc})^{N^{\mathcal{R}}_{zc}-1+\epsilon}, \end{aligned} \tag{D4}$$

where ϵ is a positive, small number to ensure that the Dirichlet distribution cannot get zero or negative counts as input (some of the z, c counts will be zero).

For a large number of galaxies, the marginalized mean and variance of $f^{\mathcal{R}}_{zc}$ reduce to $N^{\mathcal{R}}_{zc}/N^{\mathcal{R}}$, which is the classical approximate histogram estimator (note that Dirichlet is the correct posterior distribution for a histogram, but a Gaussian distribution with $N^{\mathcal{R}}_{zc}/N^{\mathcal{R}}$ mean and variance become a good approximation). Equivalent expressions arise for the $f^{\mathcal{B}}$ and $f^{\mathcal{D}}$ coefficients. We note that, for the wide sample, we can consider $f^{\mathcal{W}}_{\hat{c}} = N^{\mathcal{W}}_{\hat{c}}/N^{\mathcal{W}}$ to be an exact result (not stochastic), because we are interested in the $p(z)$ for the realisation of the wide-field survey that we have, not the redshift distribution for a hypothetical infinite survey.

D2 Sample Variance

Both Deep and Redshift Samples span a much smaller area than that of the DES Y3 wide source sample. Therefore, the underlying redshift distribution measured in the deep fields – and since they are correlated, the measured colour distribution – contain random large-scale structure fluctuations particular to that volume, commonly referred to as *sample variance*. We can describe the observed redshift distribution in the deep field as

$$N^{\mathcal{D}}_{z} = \mathrm{Poisson}\left[N^{\mathcal{D}} f^{\mathcal{W}}_{z} (1 + \Delta^{\mathcal{D}}_{z})\right], \tag{D5}$$

with $f^{\mathcal{W}}_{z}$ as the underlying redshift distribution in the wide field, and $\Delta^{\mathcal{D}}_{z}$ the particular redshift fluctuation found in the deep field with respect to the wide field, with $\mathrm{Var}(\Delta^{\mathcal{D}}_{z})$ the sample variance.

The Dirichlet sampling (Equation D4, $\mathrm{Dir}(\boldsymbol{f}^{\mathcal{R}}; \alpha_{zc} = N^{\mathcal{R}}_{zc} + \epsilon)$) described in § D1 only reproduces the variance expected from Poisson noise, but does not account of the additional uncertainty due to sample variance. In order to increase the variance of the Dirichlet sample, one can perform the transformation $\alpha_i \to \alpha_i/\lambda$, which does not change the expected value of f_i in the Dirichlet distribution, but does change its variance roughly as $\mathrm{Var}(f_i) \to \lambda \mathrm{Var}(f_i)$, for those coefficient indices i for which $\alpha_i \ll \sum_i \alpha_i$. When the value of λ is the ratio between total noise (sample variance and shot noise) over shot noise we obtain samples of f_i with the larger, correct variance. However, for equally spaced redshift bins $\mathrm{Var}(\Delta^{\mathcal{D}}_{z})$ is a function of redshift (*i.e.* sample variance becomes larger at lower redshift where the volume is smaller). A constant value λ cannot increase the variance as a function of redshift as needed, but a value of λ that changes as a function of redshift would bias the expected value of f_i.

However, we do not sample in the f_z space, but in f_{zc} space. Two phenotypes (different deep SOM cells c) that overlap in redshift have correlations due to the same underlying large-scale structure fluctuations. We will work under the assumption that different phenotypes at the same redshift have the same sample variance. As phenotypes are defined in observed colour space we do not expect large differences in their galaxy bias when they overlap in redshift, and we defer a more detailed study to future work. Sánchez et al. (2020b, S20 hereafter) introduced a three-step Dirichlet sampling method (3sDir) which produces samples of f_{zc} that can incorporate the correct level of sample variance as a function of redshift and correlate phenotypes that overlap in redshift. S20 validated in simulations that 3sDir correctly reproduces the amount of sample variance expected in both the mean redshift and width of the redshift distribution of a wide field estimated from a smaller patch of the sky.

D2.1 3sDir method

The 3sDir method from S20 assumes the coefficients f_{zc} are inferred from a single Redshift Sample, $f^{\mathcal{R}}_{zc}$. The 3sDir method introduces the concept of a *superphenotype* T as a group of deep SOM cells that are close in redshift, such that the superphenotypes become nearly disjoint in redshift space. This allows us to introduce a redshift dependent parameter λ (one λ for each T) and correlate phenotypes that are close in redshift (different c in the same T are correlated). Following S20, we can write the probability of redshift and colour, with the superphenotypes T, as

$$p(z, c) = \sum_T p(c|z, T) p(z|T) p(T) \tag{D6}$$

$$f_{zc} = \sum_T f^{zT}_{c}\, f^{T}_{z}\, f_T\,. \tag{D7}$$

To produce a sample of the coefficients $\{f_{zc}\}$ we need to produce a sample of the coefficients $(\{f^{zT}_c\}, \{f^{T}_z\}, \{f_T\})$, which we infer from the observed Redshift Sample number counts in each zcT bin, $N^{\mathcal{R}}_{zcT}$. Note that $\{f^{T}_z\}$ are independent from $\{f_T\}$, since the former is conditioned on T (indicated by the superscript). Similarly, $\{f^{zT}_c\}$ are independent from both $\{f^{T}_z\}$ and $\{f_T\}$). Therefore, we can sample them separately from the observed counts. 3sDir consists of drawing, in sequence, values of $\{f_T\}$, then of $\{f^{T}_z\}$, then of $\{f^{zT}_c\}$ with individual Dirichlet distributions from the appropriate galaxy counts, $\{N^{\mathcal{R}}_T\}$, $\{N^{\mathcal{R}}_{zT}\}$ and $\{N^{\mathcal{R}}_{zcT}\}$, respectively. However, we will rescale the counts used to infer the samples of both $\{f_T\}$ and $\{f^{T}_z\}$. This process increases the variance of the final $\{f_{zc}\}$ sample to the level expected for the sum of shot noise and sample variance, while keeping its expected value. In other words, we draw from the following distributions

$$p(\{f_T\}|\{N_T^{\mathcal{R}}\}) \sim \mathrm{Dir}\left(\{f_T\}; \{\alpha_T = \frac{N_T^{\mathcal{R}}}{\bar{\lambda}} + \epsilon\}\right); \tag{D8a}$$

$$p(\{f_z^T\}|\{N_{zT}^{\mathcal{R}}\}) \sim \mathrm{Dir}\left(\{f_z^T\}; \{\alpha_z = \frac{N_{zT}^{\mathcal{R}}}{\lambda_T} + \epsilon\}\right) \quad \text{for each } T; \tag{D8b}$$

$$p(\{f_c^{zT}\}|\{N_{zcT}^{\mathcal{R}}\}) \sim \mathrm{Dir}\left(\{f_c^{zT}\}; \{\alpha_c = N_{zcT}^{\mathcal{R}} + \epsilon\}\right) \quad \text{for each } z, T. \tag{D8c}$$

where

$$\bar{\lambda} \equiv \sum_z \lambda_z \frac{N_z^{\mathcal{R}}}{N^{\mathcal{R}}}, \tag{D9}$$

$$\lambda_T \equiv \sum_z \lambda_z \frac{N_{zT}^{\mathcal{R}}}{N_T^{\mathcal{R}}}, \tag{D10}$$

$$\lambda_z \equiv \frac{\mathrm{Var}(N_z^{\mathcal{R}})}{N_z^{\mathcal{R}}} = 1 + N_z^{\mathcal{R}}\mathrm{Var}(\Delta_z^{\mathcal{R}}). \tag{D11}$$

Equation D11, λ_z, is the ratio of the total variance (shot noise and sample variance) to only the shot noise variance. When we infer $\{f_z^T\}$, the redshift counts for each superphenotype, $\{N_{zT}^{\mathcal{R}}\}$, are rescaled by a constant value equal to the average λ_z ratio weighted by the superphenotype's redshift distribution: λ_T (Equation D10). When we infer $\{f_T\}$, the counts $\{N_T^{\mathcal{R}}\}$ get rescaled by the average λ_z weighted by the sample redshift distribution, $\bar{\lambda}$ (Equation D9). Overall, this noise-inflated Dirichlet sampling scheme (Equation D8) is an approximate model of how sample variance affects the joint redshift and colour redshift distribution, which allows one to increase the variance as a function of redshift without introducing any bias (as noted in S20).

Finally, we estimate the sample variance term, $\mathrm{Var}(\Delta_z^{\mathcal{R}})$ (Equation D11), from theory following the same assumptions as in S20, which assumed a circular footprint of the same area as the Redshift Sample, which gives a prediction which is good at the 10 – 20 per cent level, mostly due to the galaxy bias modeling (see S20 for more details, including small dependence of the prediction on cosmology). S20 validated the method in simulations, and then applied 3sDir to the COSMOS field, which is the field of our Redshift Sample, so we directly use the sample variance prediction from S20.

D2.2 Sample variance in the Deep Sample

In this analysis the Redshift Sample spans a smaller area than the whole deep field area, which carries additional information of the marginal distribution of colours, $p(c)$. We have four deep fields, F = {COSMOS=COS, C3, E2, X3}, so we can write the probability of f_z conditioned on the counts from the four fields as

$$\begin{aligned}
p(f_z|N_z^{\mathrm{COS}}, N_z^{\mathrm{C3}}, N_z^{\mathrm{E2}}, N_z^{\mathrm{X3}}) &\propto p(N_z^{\mathrm{COS}}, N_z^{\mathrm{C3}}, N_z^{\mathrm{E2}}, N_z^{\mathrm{X3}}|f_z)p(f_z) \\
&\approx p(N_z^{\mathrm{COS}}|f_z)p(N_z^{\mathrm{C3}}|f_z) \\
&\quad \times p(N_z^{\mathrm{E2}}|f_z)p(N_z^{\mathrm{X3}}|f_z)p(f_z) \\
&\propto \mathrm{Dir}\left(\alpha_z = \sum_F \frac{N_z^F + \epsilon}{1 + N_z^F\,\mathrm{Var}(\Delta_z^F)}\right),
\end{aligned} \tag{D12}$$

where in the second line of Equation D12 we approximate that the observed redshift number counts of each field N_z^F are independent of each other. However we do not have complete redshift information in all fields: we have complete high-quality photometric redshift information in the COSMOS field, while we have incomplete and inhomogeneous spectroscopic coverage in all fields. For the purpose of modeling sample variance, one limit is to ignore the redshift information in the C3, E2 and X3 fields; and assume that the Redshift Sample is self-contained in the COSMOS field. Then, one can define the redshift number counts in any field by re-weighting the redshift information in the COSMOS field. In other words, we use

$$N_z^F \equiv \sum_c \left(\frac{N_{zc}^{\mathrm{COS}}}{\sum_{z'} N_{z'c}^{\mathrm{COS}}} N_c^F\right) \qquad \text{for } F \in \{\mathrm{COS, C3, E2, X3}\}. \tag{D13}$$

The N_z^F are independent from each other in the limit where there is a tight relation between redshift and deep color (*i.e.* $p(z|c)$ is narrow) that is well determined in the Redshift Sample; and when the noise is dominated by the sample variance in the color distribution in each field, N_c^F.

We define the effective ratio of the total variance to only the shot noise in all the deep fields, λ_z^{eff}, from Equation D12 as

$$\frac{\sum_F N_z^F}{\lambda_z^{\mathrm{eff}}} \equiv \sum_{F\in\{\mathrm{COS,C3,E2,X3}\}} \frac{N_z^F}{1 + N_z^F\,\mathrm{Var}(\Delta_z^F)}, \tag{D14}$$

where $\mathrm{Var}(\Delta_z^F)$ is defined by using the correct area of each field. We define $\bar{\lambda}^{\mathrm{eff}}$ as

$$\bar{\lambda}^{\mathrm{eff}} \equiv \sum_z \lambda_z^{\mathrm{eff}} \frac{\sum_F N_z^F}{\sum_F N^F}. \tag{D15}$$

In practice, the value of $\bar{\lambda}$ and $\bar{\lambda}^{\mathrm{eff}}$ is similar, since the decrease in sample variance (roughly inversely proportional to the area) is in part compensated by the increase in number counts (proportional to the area).

D2.3 Application of 3sDir to DES Y3

From Equation D2, we want to sample the following coefficients

$$f_{zc} \equiv \frac{f_{zc}^{\mathcal{R}}}{\sum_z f_{zc}^{\mathcal{R}}} f_c^{\mathcal{D}}. \tag{D16}$$

First, we sample the coefficients $\{f_{zc}^{\mathcal{R}}\}$ using only the Redshift Sample with the same 3sDir formalism from § D2.1. Then, we separately sample the coefficients $\{f_c^{\mathcal{D}}\}$ using only the Deep Sample with the formalism that we now describe. Finally, we can compute the sample of coefficients $\{f_{zc}\}$ using Equation D16, which replaces the sample from Equation D7.

To sample the coefficients $\{f_c^{\mathcal{D}}\}$, we write the probability of colour with the superphenotypes T as

$$p(c) = \sum_T p(c|T)p(T); \tag{D17}$$

$$f_c = \sum_T f_c^T f_T; \tag{D18}$$

similar to Equation D7. Then, we sample the coefficients $\{f_c^T\}$ and $\{f_T\}$ with

$$p(\{f_T\}|\{N_T^{\mathcal{D}}\}) \sim \mathrm{Dir}\left(\{f_T\}; \alpha_T = \frac{N_T^{\mathcal{D}}}{\bar{\lambda}^{\mathrm{eff}}} + \epsilon\right); \tag{D19a}$$

$$p(\{f_c^T\}|\{N_{cT}^{\mathcal{D}}\}) \sim \mathrm{Dir}\left(\{f_c^T\}; \alpha_c = N_{cT}^{\mathcal{D}} + \epsilon\right) \quad \text{for each } T. \tag{D19b}$$

with $\bar{\lambda}^{\rm eff}$ from Equation D15.

D3 Bin conditionalization

The sampling process described so far consists of drawing values for f_{zc} (Equation D16) which represents the term $f_{zc} = \frac{f^{\mathcal{R}}_{zc}}{f^{\mathcal{R}}_{c}} f^{\mathcal{D}}_{c}$ from Equation D2 because it includes information from both the Deep and Redshift Samples. We already include the probablity that a galaxy is selected into the weak lensing sample in the counts $N^{\mathcal{R}}_{zc}$ and $N^{\mathcal{D}}_{c}$ that we input to the 3sDir method (*i.e.* each galaxy counts as a fraction equal to its Balrog detection probability). We draw one sample of f_{zc} for all four tomographic bins, and we add the bin conditionalization (Equation 5) by multiplying the fractional probability g_{zc} that each (z, c) bin is assigned to a tomographic bin $\hat{b}$ as measured from the counts:

$$f^{\mathcal{R},\hat{b}}_{zc} \equiv g^{\mathcal{R},\hat{b}}_{zc} \times f^{\mathcal{R}}_{zc}, \tag{D20}$$

where $g^{\mathcal{R},\hat{b}}_{zc}$ is the fractional probability that galaxies from the Redshift sample end up in each tomographic bin according to Balrog,

$$g^{\mathcal{R},\hat{b}}_{zc} \equiv \frac{\sum_{\hat{c}\in\hat{b}} N^{\mathcal{R}}_{z,c,\hat{c}}}{\sum_{\hat{c}} N^{\mathcal{R}}_{z,c,\hat{c}}}, \quad \text{and} \quad \sum_{\hat{b}} f^{\mathcal{R},\hat{b}}_{zc} = f^{\mathcal{R}}_{zc}. \tag{D21}$$

Similarly, we can also define for the deep sample

$$f^{\mathcal{D},\hat{b}}_{c} \equiv g^{\mathcal{D},\hat{b}}_{c} \times f^{\mathcal{D}}_{c}; \qquad g^{\mathcal{D},\hat{b}}_{c} \equiv \frac{\sum_{\hat{c}\in\hat{b}} N^{\mathcal{D}}_{c,\hat{c}}}{\sum_{\hat{c}} N^{\mathcal{D}}_{c,\hat{c}}}; \qquad \sum_{\hat{b}} f^{\mathcal{D},\hat{b}}_{c} = f^{\mathcal{D}}_{c}. \tag{D22}$$

We define an effective tomographic bin weight that we can apply to our sample f_{zc}, $g^{\hat{b}}_{zc}$, as

$$g^{\hat{b}}_{zc} \equiv \frac{g^{\mathcal{R},\hat{b}}_{zc}}{\sum_z g^{\mathcal{R},\hat{b}}_{zc}} g^{\mathcal{D},\hat{b}}_{c} \quad \text{and then} \quad f^{\hat{b}}_{zc} = g^{\hat{b}}_{zc} \times f_{zc}. \tag{D23}$$

Whenever there are no Redshift galaxies measured in a bin and cell, we set the redshift distribution to the non-tomographic one (following Equation 6 and the discussion in section 4.1). To summarize, we draw one sample of f_{zc} and use the weight $g^{\hat{b}}_{zc}$ to compute the four tomographic bin samples $f^{\hat{b}}_{zc}$ (Equation D23).

D4 Lensing weights

Similarly, to include the lensing and response weights from § 4.2 we define an averaged weight for each (z, c) pair in the Redshift sample:

$$\langle w^{\mathcal{R}}_{zc} \rangle \propto g^{\mathcal{R},\hat{b}}_{zc} \sum_{i\in(z,c)} \left(\frac{1}{M_i} \sum_j w_{ij} \right), \tag{D24}$$

where w_{ij} is the lensing weight for the j-th detection that passes metacalibration selection of the i-th deep field galaxy with redshift information; M_i is the number of times galaxy i has been injected into Balrog; $g^{\mathcal{R},\hat{b}}_{zc}$ is the conditioned probability of each tomographic bin (Equation D21). We are also interested in the lensing averaged weight for each deep cell in the Redshift Sample:

$$\langle w^{\mathcal{R}}_{c} \rangle \propto \left(\sum_z g^{\mathcal{R},\hat{b}}_{zc} \right) \sum_{i\in(c)} \left(\frac{1}{M_i} \sum_j w_{ij} \right). \tag{D25}$$

Analogously, we define an averaged lensing weight for the Deep Sample:

$$\langle w^{\mathcal{D}}_{c} \rangle \propto g^{\mathcal{D},\hat{b}}_{c} \sum_{i\in(c)} \left(\frac{1}{M_i} \sum_j w_{ij} \right), \tag{D26}$$

with $g^{\mathcal{D},\hat{b}}_{c}$ from Equation D22. Finally, we define the effective weight as

$$\langle w_{zc} \rangle \equiv \frac{\langle w^{\mathcal{R}}_{zc} \rangle}{\langle w^{\mathcal{R}}_{c} \rangle} \langle w^{\mathcal{D}}_{c} \rangle, \quad \text{so that} \quad f^{\hat{b}}_{zc} \to \langle w_{zc} \rangle f^{\hat{b}}_{zc}. \tag{D27}$$

with $f^{\hat{b}}_{zc}$ from Equation D23.

In summary, we obtain a sample of f_{zc} from Equation D16 from Balrog-weighted counts of the Redshift and Deep fields, to which we apply a tomographic bin selection probability weight to obtain the coefficients for each tomographic bin, $f^{\hat{b}}_{zc}$, (Equation D23) and finally apply the lensing and response weight (Equation D27).

D5 3sDir Modified for WZ (MFWZ)

To jointly sample from the 3sDir likelihood from photometry from this paper and the clustering redshifts (WZ) likelihood from Gatti, Giannini et al. (2020a) we have implemented a Hamiltonian Monte Carlo (HMC) algorithm, which is far more efficient than importance sampling 3sDir samples with the WZ likelihood (see Gatti, Giannini et al. (2020a) for details). However, we have implemented a modified version of the 3sDir likelihood (MFWZ) for the HMC algorithm that we describe here.

The 3sDir MFWZ likelihood samples using the equations for the Redshift Sample, Equation D7 and Equation D8, and only incorporates the information from the deep-field colour counts during step 1 (Equation D8a). Accordingly, we also update the value of $\bar{\lambda}$ in Equation D9 with $\bar{\lambda}^{\rm eff}$ from Equation D15. We sample $\{f_T\}$ from the colour counts from the deep field $\{N^{\mathcal{D}}_T\}$ with

$$p(\{f_T\}|\{N^{\mathcal{D}}_T\}) \sim \mathrm{Dir}\left(\{f_T\}; \alpha_T = \frac{N^{\mathcal{D}}_T}{\bar{\lambda}^{\rm eff}} + \epsilon \right). \tag{D28}$$

The samples of $\{f^{zT}_c\}$ and $\{f^T_z\}$ are obtained from Equation D8b and Equation D8c. Finally one obtains the $\{f_{zc}\}$ sample from $(\{f^{zT}_c\}, \{f^T_z\}, \{f_T\})$ using Equation D7.

Although 3sDir MFWZ is using less information from the deep fields, we find it easier to implement in a HMC together with the WZ likelihood.

D6 Known errors

During the processing of the 3sDir and 3sDir-MFWZ samples the following error was made. In bin conditionalization, when there is no Redshift galaxy that satisfies both c and $\hat{b}$, we instead use the redshift information from any tomographic bin in that cell. In other words, we use Equation 6 instead of Equation 5 as discussed in section 4.1. When implementing the lensing responses in 3sDir we did not properly implement this last change, and in practice we always used Equation 5. This produces a shift in the $n(z)$ average mean redshift equal to $\Delta_z = [0.003, 0.003, \sim 0, -0.004]$ (difference between the correct implementation minus the actual implementation). We note that the effect of this error is small compared to all other uncertainties included in the analysis.

APPENDIX E: VALIDATION OF SOMPZ AND 3sDir

In order to validate the methodology of SOMPZ and 3sDir we use the suite of Buzzard simulations. We note that the Buzzard simulations do not include simulated images, so we cannot test the lensing and response weight methods from § 4.2 in SOMPZ, nor § D4 in 3sDir. The validation of such weights is explored in (MacCrann et al. 2020), and we have verified that both the SOMPZ and 3sDir weight implementations are consistent: the SOMPZ weights are applied individually to galaxies, while in 3sDir they are applied as averaged quantities to f_{zc}. We have verified this change does not introduce biases larger than 10^{-3} in the mean redshift in any of the tomographic bins.

We generate 300 versions of the four DES Deep Samples (where one of the four has perfect redshift information) at different random line-of-sight positions in the Buzzard simulations. For each of the 300 realisations of the deep fields, we run the SOMPZ algorithm and estimate the different simulated number counts, $N_c^{\mathcal{D}}$, $N_{zc}^{\mathcal{R}}$ and $N_{c\hat{c}}^{\mathcal{B}}$, while the wide field remains constant. Then we obtain an $n(z)$ estimate for each tomographic bin by fixing the probabilities to the observed number counts.

To test the performance of the 3sDir method we perform the following procedure in each of the 300 Buzzard realisations of the deep fields. We draw 10^4 samples from Equation D16, $\{f_{zc}^i;\, i = 1, \ldots, 10^4\}$ to which we apply the bin conditionalization using Equation D23, and use Equation D2 to obtain the 10^4 $\{f_z^i;\, i = 1, \ldots, 10^4\}$ samples for each tomographic bin. From it, we estimate the mean redshift of each f_z^i sample, $\bar{z}^i = \sum_z z f_z^i$, and its average value $\bar{z}^{\text{3sDir}} \equiv \langle \bar{z}^i \rangle$ in each Buzzard realisation. We also compute the $\bar{z}^{\text{SOMPZ}}$ value of the single $n(z)$ from SOMPZ in each realisation, which we obtain by fixing the probabilities to the number counts. In summary, we have 300 values of $\bar{z}^{\text{SOMPZ}}$ and $\bar{z}^{\text{3sDir}}$, and a total of 300×10^4 values of $\bar{z}^i$ whose variance reflects the uncertainty on the mean redshift per tomographic bin as estimated from 3sDir.

Figure E1 shows the distribution of the 300 values of $\bar{z}^{\text{SOMPZ}}$, $\bar{z}^{\text{3sDir}}$ and $\bar{z}^{\text{3sDir-MFWZ}}$ compared to the true $\bar{z}^{\text{true}}$ (shown as dotted lines). First, we find the $\bar{z}^{\text{SOMPZ}}$ distribution to be centred offset from the truth by $\Delta_z = [0.0051, 0.0024, -0.0013, -0.0024]$ in each bin, where $\Delta_z \equiv \langle \bar{z}^{\text{SOMPZ}} \rangle - \bar{z}^{\text{true}}$. As discussed in § 5.1.1, we expect a nonzero offset due to the bin conditionalization approximation, and we include this nonzero offset as an intrinsic systematic error to the mean redshift (see § 5.5). On the other hand, we find the averages over 300 realisations, $\langle \bar{z}^{\text{3sDir}} \rangle$ and $\langle \bar{z}^{\text{SOMPZ}} \rangle$, to be within 0.001 of each other in redshift in Figure E1, meaning that 3sDir is on average unbiased with respect to the SOMPZ mean redshift. We also find the width of both distributions to agree. However, we find the distribution of $\bar{z}^{\text{3sDir-MFWZ}}$ to have more scatter than the $\bar{z}^{\text{SOMPZ}}$ and $\bar{z}^{\text{3sDir}}$ distributions. This is a consequence of 3sDir-MFWZ not fully exploiting the information available in the Deep Sample on the colour abundance $p(c)$, since we only use it to inform the superphenotype distribution $p(T)$ (§ D5). The 3sDir-MFWZ likelihood is more suitable to be sampled efficiently together with the clustering redshifts likelihood using an HMC (Gatti, Giannini et al. 2020a).

To test if the predicted distribution on $\bar{z}$ values from 3sDir is consistent with $\bar{z}^{\text{SOMPZ}}$ we compute in each Buzzard realisation the pull distribution as $\Delta_z^i = (\bar{z}^i - \bar{z}^{\text{SOMPZ}})/\sigma(\bar{z}^i)$, with $\sigma(\bar{z}^i)$ the standard deviation of the $\bar{z}^i$ values from 3sDir. Fig. E2 presents the stacked pull distributions from all 300 Buzzard realisations. We find this distribution to be centred at zero and very similar to a Gaussian distribution with zero mean and unit variance, illustrating more rigorously that 3sDir predicts samples of $\bar{z}$ which are fully compatible with the SOMPZ mean redshift. We also do the same test with 3sDir-MFWZ, finding the same conclusion.

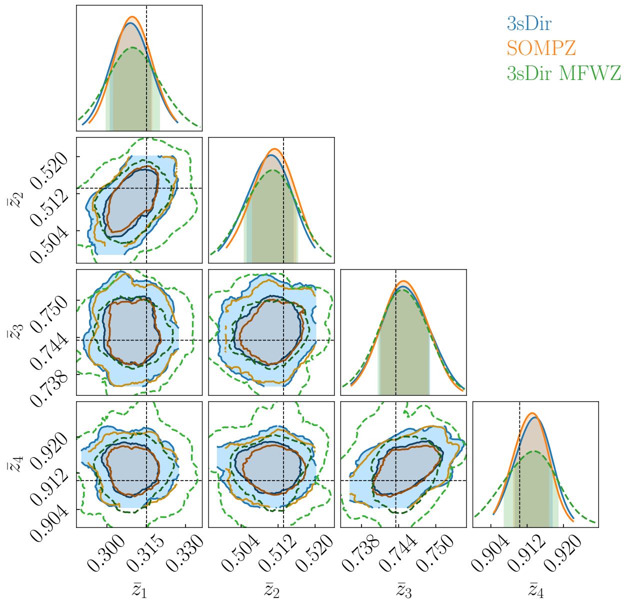

Figure E1. Distribution of mean redshift, $\bar{z}$, values in each of the 300 realisations of the deep fields in Buzzard compared to the truth (cross-hatches). In each deep field realisation we run the SOMPZ code, obtain an $n(z)$ by fixing the probabilities to the number count measurements, and calculate the mean redshift of each tomographic bin. Similarly, we draw 10^4 samples of f_{zc} with the 3sDir and 3sDir-Alt method, compute the redshift distribution of each tomographic bin for each sample, their mean redshift, and we finally compute the average $\bar{z}$ value. The $\bar{z}$ distribution from 3sDir is wider because it is not using all the colour information from the deep fields.

Fig. E3 addresses the width predicted by 3sDir or 3sDir-MFWZ in each Buzzard realisation, compared to the scatter in $\bar{z}$ from SOMPZ across the 300 Buzzard realisations. The vertical line in each panel shows the spread of $\bar{z}^{\text{SOMPZ}}$ across the 300 Buzzard realisations (*i.e.* the spread of SOMPZ in Figure E1). While SOMPZ only produces one estimate of $\bar{z}$ in each realisation, the 3sDir and 3sDir-MFWZ models produce a distribution of $\bar{z}$ values in each Buzzard realisation. In each realisation, we compute the standard deviation of $\bar{z}$ for both 3sDir and 3sDir-MFWZ, and we show the histogram of these 300 values. As expected, the predicted $\sigma(\bar{z})$ values from 3sDir-MFWZ are [78, 31, 23, 39] per cent larger than 3sDir in each bin, since the former is using less information from the deep fields. We find the $\sigma(\bar{z})$ from 3sDir to be in reasonable agreement with SOMPZ, although we find them to be slightly underestimated at lower redshift and overestimated at higher redshift, finding $[-11, 4, 8, 53]$ per cent difference in each bin. This is in agreement with Sánchez et al. (2020b) (see their fig. 12) which shows that 3sDir tends to underpredict the variance at low redshift, and the opposite at high redshift.

AFFILIATIONS

[1] Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA
[2] Kavli Institute for Particle Astrophysics & Cosmology, P. O. Box

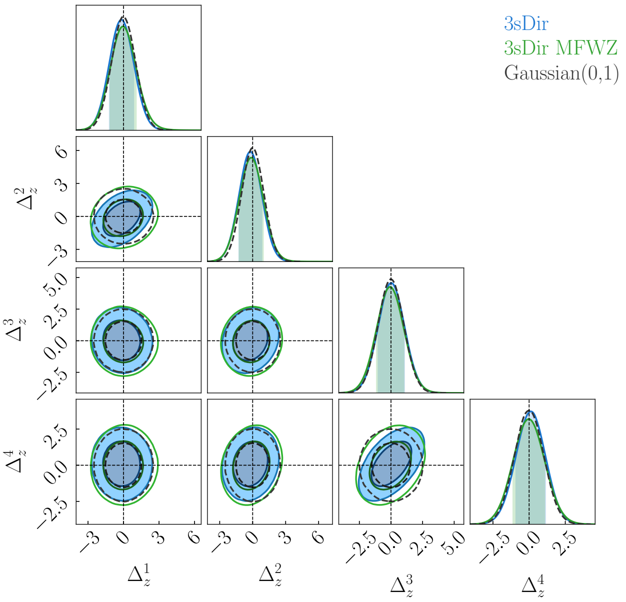

Figure E2. Distribution of residual 3sDir or 3sDir-Alt samples across 300 realisations of the deep fields on Buzzard. A residual sample is defined as $\Delta_z^i = (\bar{z}^i - \bar{z}^{\mathrm{SOMPZ}})/\sigma(\bar{z}^i)$, with $\sigma(\bar{z}^i)$ the standard deviation of the $\bar{z}^i$ values from 3sDir in each realisation. The distributions agree with a Gaussian distribution with zero mean and unit variance (shown as dashed lines), which shows that the mean redshifts from 3sDir and 3sDir-Alt are statistically in agreement with $\bar{z}^{\mathrm{SOMPZ}}$ across the 300 Buzzard realisations.

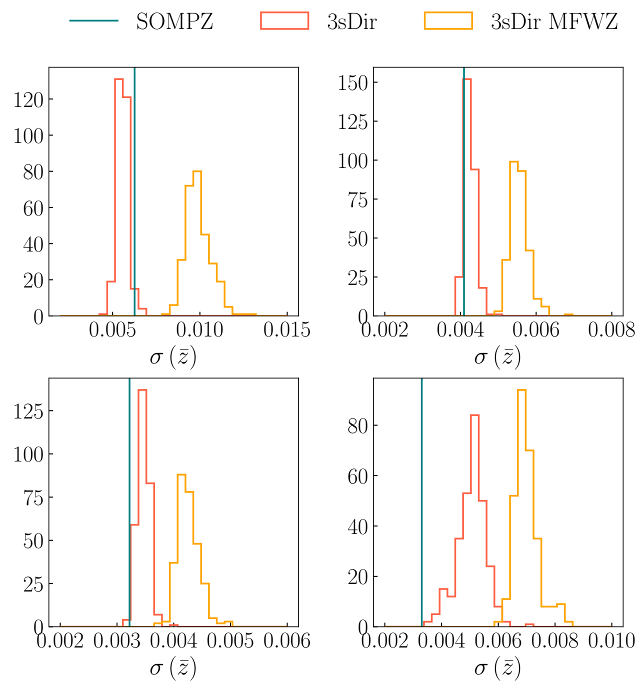

Figure E3. Vertical line: Standard deviation of $\bar{z}^{\mathrm{SOMPZ}}$ across the 300 realisations. Histograms: standard deviation of individual $\bar{z}$ values drawn with 3sDir or 3sDir-Alt in each of the 300 realisations.

2450, Stanford University, Stanford, CA 94305, USA
[3] SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA
[4] Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439, USA
[5] Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA
[6] Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA
[7] Department of Astronomy, University of California, Berkeley, 501 Campbell Hall, Berkeley, CA 94720, USA
[8] Department of Physics, Duke University Durham, NC 27708, USA
[9] Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15312, USA
[10] Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, UK
[11] Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA
[12] Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth, PO1 3FX, UK
[13] Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA
[14] Jodrell Bank Center for Astrophysics, School of Physics and Astronomy, University of Manchester, Oxford Road, Manchester, M13 9PL, UK
[15] Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona) Spain
[16] Laboratório Interinstitucional de e-Astronomia - LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil
[17] Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil
[18] Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801, USA
[19] National Center for Supercomputing Applications, 1205 West Clark St., Urbana, IL 61801, USA
[20] Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK
[21] Département de Physique Théorique and Center for Astroparticle Physics, Université de Genève, 24 quai Ernest Ansermet, CH-1211 Geneva, Switzerland
[22] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109, USA
[23] Fermi National Accelerator Laboratory, P. O. Box 500, Batavia, IL 60510, USA
[24] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
[25] Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain
[26] Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA
[27] Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain
[28] Brookhaven National Laboratory, Bldg 510, Upton, NY 11973, USA
[29] Cerro Tololo Inter-American Observatory, NSF's National Optical-Infrared Astronomy Research Laboratory, Casilla 603, La Serena, Chile
[30] Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, CP 66318, São Paulo, SP, 05314-970, Brazil

[31] CNRS, UMR 7095, Institut d'Astrophysique de Paris, F-75014, Paris, France
[32] Sorbonne Universités, UPMC Univ Paris 06, UMR 7095, Institut d'Astrophysique de Paris, F-75014, Paris, France
[33] Department of Physics and Astronomy, Pevensey Building, University of Sussex, Brighton, BN1 9QH, UK
[34] Department of Physics & Astronomy, University College London, Gower Street, London, WC1E 6BT, UK
[35] Instituto de Astrofisica de Canarias, E-38205 La Laguna, Tenerife, Spain
[36] Universidad de La Laguna, Dpto. Astrofísica, E-38206 La Laguna, Tenerife, Spain
[37] Institut d'Estudis Espacials de Catalunya (IEEC), 08034 Barcelona, Spain
[38] Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, 08193 Barcelona, Spain
[39] University of Nottingham, School of Physics and Astronomy, Nottingham NG7 2RD, UK
[40] INAF-Osservatorio Astronomico di Trieste, via G. B. Tiepolo 11, I-34143 Trieste, Italy
[41] Institute for Fundamental Physics of the Universe, Via Beirut 2, 34014 Trieste, Italy
[42] Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA
[43] Department of Physics, IIT Hyderabad, Kandi, Telangana 502285, India
[44] Department of Astronomy/Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721-0065, USA
[45] Department of Physics, The Ohio State University, Columbus, OH 43210, USA
[46] Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA
[47] Institute of Theoretical Astrophysics, University of Oslo. P.O. Box 1029 Blindern, NO-0315 Oslo, Norway
[48] Instituto de Fisica Teorica UAM/CSIC, Universidad Autonoma de Madrid, 28049 Madrid, Spain
[49] Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
[50] Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
[51] School of Mathematics and Physics, University of Queensland, Brisbane, QLD 4072, Australia
[52] Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, 81679 Munich, Germany
[53] Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, 85748 Garching, Germany
[54] Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstr. 1, 81679 München, Germany
[55] Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA
[56] Australian Astronomical Optics, Macquarie University, North Ryde, NSW 2113, Australia
[57] Lowell Observatory, 1400 Mars Hill Rd, Flagstaff, AZ 86001, USA
[58] George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, and Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA
[59] Department of Astronomy, The Ohio State University, Columbus, OH 43210, USA
[60] Radcliffe Institute for Advanced Study, Harvard University, Cambridge, MA 02138
[61] Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA
[62] Physics Department, 2320 Chamberlin Hall, University of Wisconsin-Madison, 1150 University Avenue Madison, WI 53706-1390
[63] School of Physics and Astronomy, University of Southampton, Southampton, SO17 1BJ, UK
[64] Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831
[65]Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, 08193 Barcelona, Spain
[66]Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain

This paper has been typeset from a TEX/LATEX file prepared by the author.