Dark Energy Survey Year 3 Results: clustering redshifts – calibration of the weak lensing source redshift distributions with redMaGiC and BOSS/eBOSS

M. Gatti, G. Giannini, G.M. Bernstein, A. Alarcon, J. Myles, A. Amon, R. Cawthon, M. Troxel, J. Derose, S. Everett, et al.

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **510,** 1223–1247 (2022) https://doi.org/10.1093/mnras/stab3311
Advance Access publication 2021 November 18

Dark Energy Survey Year 3 Results: clustering redshifts – calibration of the weak lensing source redshift distributions with *redMaGiC* and BOSS/eBOSS

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

Affiliations are listed at the end of the paper

Accepted 2021 November 11. Received 2021 November 11; in original form 2021 January 11

ABSTRACT

We present the calibration of the Dark Energy Survey Year 3 (DES Y3) weak lensing (WL) source galaxy redshift distributions $n(z)$ from clustering measurements. In particular, we cross-correlate the WL source galaxies sample with *redMaGiC* galaxies (luminous red galaxies with secure photometric redshifts) and a spectroscopic sample from BOSS/eBOSS to estimate the redshift distribution of the DES sources sample. Two distinct methods for using the clustering statistics are described. The first uses the clustering information independently to estimate the mean redshift of the source galaxies within a redshift window, as done in the DES Y1 analysis. The second method establishes a likelihood of the clustering data as a function of $n(z)$, which can be incorporated into schemes for generating samples of $n(z)$ subject to combined clustering and photometric constraints. Both methods incorporate marginalization over various astrophysical systematics, including magnification and redshift-dependent galaxy-matter bias. We characterize the uncertainties of the methods in simulations; the first method recovers the mean z of tomographic bins to RMS (precision) of $\sim$0.014. Use of the second method is shown to vastly improve the accuracy of the shape of $n(z)$ derived from photometric data. The two methods are then applied to the DES Y3 data.

Key words: galaxies: distances and redshifts – cosmology: observations.

1 INTRODUCTION

The Dark Energy Survey (DES) is a photometric survey that has imaged 5000 deg^2 of the sky. The DES Y3 '3x2' analysis (DES Collaboration 2021) using data taken during the first three seasons of observations constrains cosmological parameters by combining three different measurements of two-point correlation functions: cosmic shear (Amon et al. 2021; Secco et al. 2021), galaxy–galaxy lensing (Prat et al. 2020), and galaxy clustering (Rodríguez-Monroy et al. 2020). The cosmic shear measurement probes the angular correlation of more than 100 000 000 galaxy shapes from the weak lensing (WL) sample (Gatti et al. 2021), divided into four tomographic bins. The cross-correlation of galaxy shapes and the positions of red luminous

★ E-mail: marcogatti29@gmail.com (MG); ggiannini@ifae.es (GG)

galaxies identified by the *redMaGiC* algorithm (Rozo et al. 2016) is measured by galaxy–galaxy lensing. Lastly, galaxy clustering measures the autocorrelation of the positions of *redMaGiC* galaxies. A magnitude-limited sample (Porredon et al. 2021) will be also used as lens sample alternatively to *redMaGiC* in a second analysis (Porredon et al. 2021), with the goal of improving the cosmological constraints.

The correct cosmological interpretation of these measurements relies on an accurate estimate of the redshift distributions of the samples, which can otherwise lead to biases in the inferred cosmological parameters (e.g. Huterer et al. 2006; Hildebrandt et al. 2012; Choi et al. 2016; Hoyle et al. 2018). Photometric surveys have been relying on different methodologies to derive redshift distributions (Hildebrandt et al. 2010; Sánchez et al. 2014), mostly based on galaxies' multiband photometry (photo-z methods, or PZ). However, these methods are ultimately limited by the redshift ambiguities in few-band colours, and the limited and incomplete spectroscopic samples available to calibrate the colour–redshift relations.

Clustering-z methods (Newman 2008; Ménard et al. 2013; Choi et al. 2016; Davis et al. 2017; Johnson et al. 2017; Morrison et al. 2017; Gatti et al. 2018; van den Busch et al. 2020) offer an alternative to standard photo-z methods to infer redshift distributions. In short, clustering-z methods exploit the two-point correlation signal between a photometric 'unknown' sample and a 'reference' sample of high-fidelity redshift galaxies divided into thin bins, to infer the redshift distributions of the photometric sample. One of the biggest advantages of clustering-z methods is that the reference sample does not have to be representative of the photometric sample. Clustering-z methods (or WZ) have been in the past years successfully applied to both data (Rahman et al. 2015, 2016a, b; Scottez et al. 2016; Davis et al. 2017, 2018; Hildebrandt et al. 2017, 2021; Johnson et al. 2017; Cawthon et al. 2018; Bates et al. 2019; van den Busch et al. 2020) and simulations (McQuinn & White 2013; Schmidt et al. 2013; Scottez et al. 2017; Gatti et al. 2018), and they represent one credible supplement to standard photo-z methods for the new, upcoming generation of data sets (Scottez et al. 2017).

Clustering-z methods have been used both to provide an independent redshift distribution estimate and to calibrate distributions inferred from photo-z methods. In the DES Y1 cosmological analysis, we opted for the latter approach (Davis et al. 2017; Hoyle et al. 2018). In particular, we used high-quality photometric redshifts provided by *redMaGiC* galaxies (Rozo et al. 2016) to measure the clustering-z signal with the WL source-galaxy sample. The use of high-quality photometric redshifts rather than spectroscopic redshifts was motivated by the higher statistical power of the *redMaGiC* sample, owing to the large number of *redMaGiC* galaxies (650 000 for DES Y1) in the DES footprint. Due to the limited redshift range of the *redMaGiC* sample, clustering-z estimates could not have been used to determine $n(z)$ in its entirety on their own, but they have been used to calibrate the mean redshift of the distributions measured by other DES photo-z methods (with the mean taken over the *redMaGiC* z bounds). A similar approach has been implemented by the KiDS team in their recent cosmological analysis (van den Busch et al. 2020; Hildebrandt et al. 2021), where they used cross-correlation estimates to calibrate the mean redshifts inferred from other photo-z methods. They used a number of different spectroscopic samples as a reference sample, which guaranteed a greater redshift coverage but less statistical power compared to the use of *redMaGiC* galaxies.

The strategy for calibration of the WL redshift distributions for DES Y3 improves in multiple respects on the Y1 strategy outlined in Gatti et al. (2018). From the clustering-redshift side, we execute two different methods to combine clustering information with redshift distributions from photometry. The first approach is to use clustering-z to estimate the mean redshift $\langle z\rangle_{\rm wz}$, and assign a clustering-z likelihood to any candidate $n(z)$ from photo-z techniques based on the value of its mean $\langle z\rangle_{\rm pz}$ (similar to the DES Y1 analysis). We will refer to this as the 'mean-matching' approach. The second, new method is to pose both the clustering-z and the photo-z measurements as probabilities $p[D|n(z)]$ of the observational data D given redshift distributions $n(z)$; then to sample the full $n(z)$ from the posterior $p[n(z)]$ implied by multiplying these probabilities. We will refer to this as the 'full-shape' method.

We furthermore improve over Y1 in the modelling of the clustering signal, accounting for the redshift evolution of the galaxy-matter bias and the clustering of the underlying dark matter density field, which were neglected in the DES Y1 analysis. In the second method that calibrates the shape of the redshift distributions, we also marginalize over magnification effects. Finally, we use a combination of two different reference samples: *redMaGiC* galaxies with high-quality photometric redshifts; and a spectroscopic sample from the combined BOSS (Baryonic Oscillation Spectroscopic Survey, Dawson et al. 2013) and eBOSS (extended-Baryon Oscillation Spectroscopic Survey, Dawson et al. 2016; Ahumada et al. 2020; Alam et al. 2021) catalogues. Only *redMaGiC* galaxies were used in DES Y1. On one hand, *redMaGiC* galaxies span the full DES Y3 footprint (Rodríguez-Monroy et al. 2020) and are characterized by a higher number density than BOSS/eBOSS galaxies, which cover only ≈ 17 per cent of the DES Y3 footprint. On the other hand, the latter sample spans a wider redshift range and has better redshift estimates, which makes the combination of the two samples desirable.

The fiducial photo-z estimates for the DES Y3 WL sample are provided by a self-organizing map-based scheme (hereafter SOMPZ, Buchs et al. 2019; Myles et al. 2021). The SOMPZ method provides a means to generate samples of the $n(z)$ for all tomographic bins that encompass the uncertainties in the photometric inference of the distributions. The mean-matching clustering-z method may be used to confirm or adjust the $n(z)$ samples generated by SOMPZ. We use the full-shape method as the fiducial method for DES Y3, generating samples of $n(z)$ from the combined SOMPZ and clustering likelihoods. In either route, the DES Y3 cosmological analysis is done by sampling over the finite set of realizations generated by SOMPZ+clustering-z.

We note that there exist other strategies to combine clustering-z and photo-z estimates. For example, Sánchez & Bernstein (2019) and Alarcon et al. (2020) show how to combine photo-z and clustering-z estimates using a hierarchical Bayesian model (Leistedt, Mortlock & Peiris 2016). The application of these methods to DES data is left for future work.

This paper is organized as follows. In Section 2, we describe the two different methodologies used in DES Y3 to calibrate photo-z posteriors using clustering-z estimation, and explain how to assign a likelihood to the cross-correlation information. The simulations and the data sets used in this paper are described and compared in Section 3. In Section 4, we perform extended tests in simulations assessing the systematic uncertainty of the methods. The calibration on DES Y3 data is presented in Section 5, and in Section 6 we discuss future prospects for this method and present our conclusions.

2 METHODOLOGY

We describe the clustering-z (WZ) methodology as generally as possible in this section, deferring to Section 3 the description (and the choice of the binning) of the particular samples adopted for DES Y3.

2.1 Modelling and measuring the correlation signal

Clustering-z methods rely on the assumption that the cross-correlation between two samples of objects is non-zero only in the case of overlap of the distribution of objects in physical space, due their mutual gravitational influence. Let us consider two samples:

(i) An *unknown* sample, whose redshift distribution $n_{\rm u}(z)$ has to be measured, namely our WL source sample, and
(ii) A *reference* sample, whose redshift distribution $n_{\rm r}(z)$ is known (either from spectroscopic redshifts or from high-precision photometric redshifts).

We compute the angular clustering signal $w_{\rm ur}$ as a function of the separation angle θ between the unknown sample and the reference population. Under the assumption of linear biasing and the Limber approximation (Limber 1953), the clustering signal can be written as (e.g. Krause et al. 2017):

$$w_{\rm ur}(\theta) = \int {\rm d}z' n_{\rm u}(z') n_{\rm r}(z') b_{\rm u}(z') b_{\rm r}(z') w_{\rm DM}(\theta, z') + M(\theta), \quad (1)$$

where $n_{\rm u}(z^{'})$ and $n_{\rm r}(z^{'})$ are the unknown- and reference-sample redshift distributions (normalized to unity over the full redshift interval), $b_{\rm u}(z^{'})$ and $b_{\rm r}(z^{'})$ are the linear galaxy-matter biases of the two samples, and $w_{\rm DM}(\theta, z^{'})$ is the dark-matter two-point angular correlation function. The term $M(\theta)$ refers to the contribution of lensing magnification effects; description and full expressions for the terms $w_{\rm DM}(\theta, z^{'})$ and $M(\theta)$ are detailed below (equation 7 and equation A1). Note that while we acknowledge that the assumption of linear biasing is not expected to hold at small scales, we are nevertheless confident to be able to estimate the systematic bias introduced by this premise, as explained in Section 2.2. We also note that the Limber approximation is a standard assumption in clustering-z works, and it is expected to have a minimal impact on our results (e.g. McQuinn & White 2013).

Following Ménard et al. (2013), the correlation function is measured as a function of angle, and averaged over angular scales to produce a 'scalar' value via

$$\bar{w}_{\rm ur} = \int_{\theta_{\rm min}}^{\theta_{\rm max}} {\rm d}\theta \; W(\theta) w_{\rm ur}(\theta), \quad (2)$$

where $W(\theta) \propto \theta^{-\gamma}$ is a weighting function. We adopt $\gamma = 1$ to yield optimal S/N on the scalar in the presence of shot noise. The integration limits in the integral in equation (2) correspond to fixed physical scales. In this work, we choose to span the physical interval between 1.5 and 5.0 Mpc (Section 4). We use the Davis & Peebles (1983) estimator for the cross-correlation signal,

$$w_{\rm ur}(\theta) = \frac{N_{\rm Rr}}{N_{\rm Dr}} \frac{D_{\rm u} D_{\rm r}(\theta)}{D_{\rm u} R_{\rm r}(\theta)} - 1, \quad (3)$$

where $D_{\rm u}D_{\rm r}(\theta)$ and $D_{\rm u}R_{\rm r}(\theta)$ are, respectively, data–data and data–random pairs. The pairs are properly normalized through $N_{\rm Dr}$ and $N_{\rm Rr}$, corresponding to the total number of galaxies in the reference sample and in the reference random catalogue. If weights for the reference catalogue of galaxies (or for the catalogue of randoms) are provided, $N_{\rm Dr}$ (or $N_{\rm Rr}$) is the sum of the weights of the catalogue, and $D_{\rm u}D_{\rm r}(\theta)$ (or $D_{\rm u}R_{\rm r}(\theta)$) is the weighted number of pairs. Note that weights can also be assigned to the unknown sample; in that case, the weighted number of pairs $D_{\rm u}D_{\rm r}(\theta)$ (or $D_{\rm u}R_{\rm r}(\theta)$) also accounts for the weights of the unknown sample. As in Gatti et al. (2018), we use the Davis & Peebles estimator rather than the Landy & Szalay (1993) estimator since the former involves using a catalogue of random points for just one of the two samples. This allows us to avoid creating high-fidelity random catalogues for the DES Y3 source galaxy sample, whose selection function is very complex and non-trivial to replicate, besides being computationally very costly. For our analysis, we only rely on random points for the reference sample, whose selection function and mask are well understood. We note that in the rest of the paper we adopted the Davis & Peebles estimator even when measuring the autocorrelation of the reference samples, but we checked that using the Landy & Szalay estimator lead to negligible variations.

Now we assume that the reference sample is divided into redshift bins centred at z_i, each narrow enough that we can approximate $n_{{\rm r},\,i}(z) \approx \delta_D(z - z_i)$, with δ_D being Dirac's delta distribution and the integrands in equation (1) other than $n_{\rm r}$ can be treated as constant. Equations (1) and (2) become:

$$\bar{w}_{\rm ur}(z_i) \approx n_{\rm u}(z_i) b_{\rm u}(z_i) b_{\rm r}(z_i) \bar{w}_{\rm DM}(z_i) + \bar{M}(z_i), \quad (4)$$

where barred quantities indicate they have been averaged over angular scales as per equation (2). In what follows we will, for simplicity, drop the bar. The above quantity is always estimated at the redshift z_i of the i-th thin reference sample bin.

The goal is to use equation (4) to infer $n_{\rm u}(z)$, the unknown redshift distribution, from the multiple measures $w_{\rm ur}(z_i)$. But it is important to note that this equation follows from a simplifying assumption. We assumed the galaxy-matter bias to be described by a single number at all scales; this is true at large scales in the linear regime, but we do not expect this to hold at the small scales used in this work (1.5 to 5.0 Mpc). In the non-linear regime, even the fact that the terms inside the integral factorizes into $b_{\rm r}(z_i) b_{\rm u}(z_i) w_{\rm DM}(z_i)$ is not guaranteed (Bernardeau et al. 2002; Desjacques, Jeong & Schmidt 2018). The linear-bias assumption introduces a systematic uncertainty that depends on the scales adopted and the samples under study and that will be quantified in the following sections.

The evolution of the quantities $b_{\rm r}(z_i)$, $b_{\rm u}(z_i)$, $w_{\rm DM}(z_i)$ and $M(z_i)$ needs to be characterized to correctly recover the redshift distribution of the unknown sample. We turn now to how to model or estimate these terms.

(i) **The galaxy-matter bias evolution of the reference sample $b_{\rm r}(z)$.** As long as the redshifts of the reference sample are accurate enough, and we assume linear biasing, we can estimate $b_{\rm r}(z)$ by measuring the angle-averaged estimate of the autocorrelation function of the reference sample divided into thin redshift bins ($\delta z = 0.02$) centred at z_i:

$$w_{\rm rr}(z_i) = \int {\rm d}z' \left[b_{\rm r}(z') n_{\rm r,i}(z') \right]^2 w_{\rm DM}(z'). \quad (5)$$

If the bins are sufficiently narrow so as to consider the biases and $w_{\rm DM}$ constant over the distributions, they can be pulled out of the above integrals:

$$w_{\rm rr}(z_i) = b_{\rm r}^2(z_i) w_{\rm DM}(z_i) \int {\rm d}z' n_{\rm r,i}^2(z'). \quad (6)$$

Knowledge of the redshift distributions of the narrow bins is then required to use equation (6) to estimate $b_{\rm r}(z_i)$. Lastly, we need to model $w_{\rm DM}(z)$ to correctly recover $b_{\rm r}(z)$.

(ii) **The galaxy-matter bias evolution of the unknown sample $b_{\rm u}(z)$.** In principle, the autocorrelation of the unknown sample constrains this. However in our case, $n_{\rm u}(z)$ is broad and unknown, and $b_{\rm u}$ likely varies substantially across the sample, so the information on $b_{\rm u}$ from the autocorrelation is weak and entangled with $n_{\rm u}$ itself. The degeneracy between $b_{\rm u}$ and $n_{\rm u}$ is the fundamental limiting factor of clustering-z methods. Mitigation schemes exist, based on the use of additional information to constrain the evolution of $b_{\rm u}$: e.g.

Matthews & Newman (2010) use the additional constraints coming from the autocorrelation function of the tomographic bins (without dividing the samples into thin bins); or the method implemented in van den Busch et al. (2020), who use the additional constraint coming from the normalization of the redshift distribution of the full unknown catalogue not divided into tomographic bins. However, these methods are not free from shortcomings, so we decided not to attempt correcting for b_u. Since it is difficult to place a priori constraints on b_u, when forward modelling the clustering signal we chose to parametrize it in a flexible way (see Section 2.2.2), effectively treating it as a free function.

(iii) **The dark matter two-point correlation function $w_{DM}(z)$.** This can be modelled assuming a given cosmology and a non-linear power spectrum. At fixed z_i, this can be written as:

$$w_{DM}(z_i) = \int d\theta W(\theta) \sum \frac{2\ell+1}{4\pi} P_\ell(\cos\theta) \times \frac{1}{\chi(z_i)^2 H(z_i)} P_{NL}\left(\frac{l+1/2}{\chi(z_i)}, z_i\right), \quad (7)$$

where χ is the comoving distance and $H(z_i)$ is the Hubble expansion rate at redshift z_i. $P_\ell(x)$ is the Legendre polynomial of order ℓ. $P_{NL}(k, \chi)$ is the 3D non-linear matter power spectrum at wavenumber k (which, in the Limber approximation, is set equal to $(l + 1/2)/\chi(z_i)$) and at the cosmic time associated with redshift z_i. We find that the redshift evolution of $w_{DM}(z_i)$ depends little on the particular value of cosmological parameters, whereas the dependence of the overall amplitude of $w_{DM}(z_i)$ with respect to cosmology is absorbed by our systematic functions. Based on this, we hold cosmology fixed when computing $w_{DM}(z_i)$, assuming the values in Planck Collaboration VI (2020). We then verify *a posteriori* that this approximation is valid by repeating our analysis using very different values for the cosmological parameters ($\Omega_m = 0.4$, $\sigma_8 = 0.7$), finding that the impact on our conclusions is negligible. Note that some of the mitigation schemes adopted in literature to correct the galaxy-matter bias evolution of the unknown sample also automatically estimate $w_{DM}(z_i)$ from the data (Matthews & Newman 2010; van den Busch et al. 2020), but they are not adopted in this work.

(iv) **Magnification signal $M(z_i)$.** WL magnification (Narayan 1989; Villumsen, Freudling & da Costa 1997; Moessner & Jain 1998) changes the observed spatial density of galaxies: the enhancement in the flux of magnified galaxies can locally increase the number density, as more galaxies pass the selection cuts/detection threshold of the sample; at the same time, the same volume of space appears to cover a different solid angle on the sky, generally causing the observed number density to decrease. For a flux-limited sample, the net effect is driven by the slope of the luminosity function of the sample, here conveniently parametrized through the parameter α, and it has an impact on the measured clustering signal. Formally, the magnification term depends on the galaxy-matter bias and parameter α of the two samples, as well as on the redshift distribution of the unknown sample: $M(z_i; \alpha_r, \alpha_u, b_r, b_u, n_u)$. More details about our modelling of the magnification effects are given in Appendix A, although we anticipate magnification effects have a negligible impact on our analysis, due to our analysis choices. To keep our notation light, when possible, we will simply indicate magnification effects as $M(z_i)$, dropping the dependence on other factors.

Under the assumption of thin reference bins, linear galaxy-matter bias, and using the linearized version of the equation describing magnification effects (Appendix A), equation (4) becomes a linear system of equations, and can be solved to obtain an estimate of $n_u(z_i)$. This would be similar to standard clustering-z methods which use the cross-correlation signal as a starting point to infer the redshift distributions of the unknown sample (Newman 2008; McQuinn & White 2013; Ménard et al. 2013; Schmidt et al. 2013).

Alternatively, if an estimate of the $n_u(z_i)$ is provided by e.g. a photo-z method, equation (4) can be used to evaluate the expected correlation signal $w_{ur}(z_i)$ and compare it to the one measured in data, i.e. a forward modelling approach (see e.g. Choi et al. 2016).

This work represents a significant advancement over DES Y1, because in the Y1 analysis none of the terms described above were modelled. We assumed $b_r(z_i)$, $b_u(z_i)$, and $w_{DM}(z_i)$ to be constant within each photo-z bin, and used the simulations to estimate the systematic error induced by this assumption. In DES Y1 we also did not model $M(z_i)$, but we decided to exclude the redshift range (i.e. the tails of the redshift distributions) where magnification effects are expected to have a non-negligible impact. On the contrary, in this work we model $b_r(z_i)$, $w_{DM}(z_i)$, and, depending on the method, $M(z_i)$.

2.2 Assigning likelihood to the cross-correlation information

We use the clustering data $\{w_{ur}(z_i), w_{rr}(z_i)\}$, to place a likelihood $\mathcal{L}[WZ|n_u(z)]$ of obtaining the clustering-z data given some estimate of the true $n_u(z)$. The clustering-z data will be used to evaluate the likelihood of many candidate $n_u(z)$ functions, typically drawn from some combination of PZ and spectroscopic data. In the DES Y1 analysis, such realizations were taken as $n_u(z) = n_{pz}(z + \Delta z)$, where $n_{pz}(z)$ was a single 'best' photo-z estimate and Δz a free parameter. The Y3 approach is more general, with many realizations of the full function $n_u(z)$ being drawn. In any case we need only to define $\mathcal{L}[WZ|n_u(z)]$. To do so, we make use of two approaches, described below.

2.2.1 *Mean-matching method*

This method works by compressing the $n(z)$ functions to a single statistic, their mean $\langle z\rangle$. In this 'simpler' method, we do not model magnification effects, so the mean is taken over a restricted range of z, where a reference sample is available and $w_{ur}(z) \gg M(z)$, such that we can neglect magnification effects. For this method, cutting the tails can be preferable even when estimates of magnification effects in the tails are available. This is due to the fact that small errors in the magnification estimates in the tails can have a large impact on the mean of the redshift distribution, lowering the capability of the method to constrain the mean redshift.

Following the DES Y1 analysis, we choose a fixed interval $[z_{min}, z_{max}] = [\langle z\rangle_{pz} - 2\sigma_{pz}, \langle z\rangle_{pz} + 2\sigma_{pz}]$, where $\langle z\rangle_{pz}$ and σ_{pz} are the mean and root mean square of a canonical $n_{pz}(z)$. In case the fixed interval includes a range where there is no reference sample coverage, it is further reduced to ensure there are enough galaxies in the reference sample to provide a meaningful clustering-z estimate (see Section 4.1 for more details). We first create a nominal 'naive' estimator $\tilde{n}_u(z)$ using equation (4) which would be proportional to an unbiased estimator if linear bias holds and $b_u(z)$ is constant:

$$\tilde{n}_u(z_i) \propto \frac{w_{ur}(z_i)}{b_r(z_i) w_{DM}(z_i)}, \quad (8)$$

Then we define mean redshifts for the clustering-z data and the proposed $n_{pz}(z)$ as

$$\langle z\rangle_{wz} = \frac{\int_{z_{min}}^{z_{max}} dz\, z\, \tilde{n}_u(z)}{\int_{z_{min}}^{z_{max}} dz\, \tilde{n}_u(z)} \quad (9)$$

$$\langle z\rangle_{\rm pz} = \frac{\int_{z_{\rm min}}^{z_{\rm max}} {\rm d}z\, z\, n_{\rm pz}(z)}{\int_{z_{\rm min}}^{z_{\rm max}} {\rm d}z\, n_{\rm pz}(z)} \tag{10}$$

The likelihood of the WZ data given a proposed $n_{\rm u}(z)$ is then taken to be a Gaussian distribution in the $\langle z\rangle$ values:

$$\mathcal{L}\left[{\rm WZ}|n_{\rm u}(z)\right] \equiv \mathcal{N}\left(\langle z\rangle_{\rm pz} - \langle z\rangle_{\rm wz}, \sigma_{\langle z\rangle}\right) \tag{11}$$

The uncertainty $\sigma_{\langle z\rangle}$ must incorporate the estimated measurement noise and also systematic errors from shortcomings of the underlying model. Section 4.1 gives the results of using simulations to set these uncertainties. The assumption of Gaussianity is a reasonable choice even in absence of systematics, as per the central limit theorem (the mean redshift compresses the information from many different redshifts). Moreover, we parametrize the impact of systematics effects in such a way they can be described by a Gaussian likelihood, and systematic effects dominate our total error budget.

2.2.2 *Full-shape method*

This method dispenses with the mean statistic and simply compares the observed $w_{\rm ur}(z_i)$ data to a model $\hat{w}_{\rm ur}[z_i; n_{\rm u}(z), b_{\rm r}(z), b_{\rm u}(z), \alpha_{\rm r}(z), \alpha_{\rm u}(z), \mathbf{s}]$ that incorporates potential systematic effects. The model is an alteration of equation (4):

$$\hat{w}_{\rm ur}(z_i) = n_{\rm u}(z_i) b_{\rm r}(z_i) w_{\rm DM}(z_i) \times {\rm Sys}(z_i, \mathbf{s}) + M(z_i). \tag{12}$$

The functions $n_{\rm u}(z)$, $b_{\rm r}(z)$, and $M(z)$ are assumed to be given beforehand, and $w_{\rm DM}$ is calculated from theory as described in equation (7). The Sys function multiplies the clustering signal by some redshift-dependent value that is parameterized by $\mathbf{s} = \{s_1, s_2, \ldots\}$ that we will marginalize over. The role of the Sys function is to absorb all uncertainties in $b_{\rm u}$ and its redshift dependence, as well as uncertainties due to failures in the linear bias model itself, and in the determination of $b_{\rm r}(z)$. The choice of Sys function and the priors on its parameters are guided by simulations as described in Section 4.2. As a rule of thumb, we expect the Sys function amplitude to slowly vary across redshift, and to be of the same magnitude of a typical galaxy-matter bias (i.e. around unity). We note that in principle we could also have absorbed the redshift dependence of $w_{\rm DM}$, or the magnification contribution $M(z)$, into the Sys function. We did not proceed this way since we know how to model these contributions, although this comes at the expense of a more complex model. Lastly, we note that formally the magnification contribution also depends on the bias $b_{\rm u}$; this is marginalized separately, together with the magnification parameter of the unknown sample α_u (more details are given in Appendix A).

With a model for $w_{\rm ur}$ in hand, we assume that the measurement errors in the data are Gaussian and define a likelihood

$$\mathcal{L}\left[{\rm WZ}|n_{\rm u}(z), b_{\rm r}(z), \alpha_{\rm r}(z), w_{\rm DM}(z)\right] \propto \int {\rm d}\mathbf{s}\, {\rm d}\mathbf{p}\, \exp\left[-\frac{1}{2}(w_{\rm ur} - \hat{w}_{\rm ur})^T \Sigma_w^{-1} (w_{\rm ur} - \hat{w}_{\rm ur})\right] p(\mathbf{s}) p(\mathbf{p}), \tag{13}$$

where $\mathbf{p} = \{b_{\rm u}, \alpha_{\rm u}\}$ enters in the modelling of the magnification term. The data and model for $w_{\rm ur}$ are taken here to be vectors over z_i, and Σ_w is the covariance matrix of the data (from shot noise and sample variance). The nuisance parameter sets $\mathbf{s}$ and $\mathbf{p}$ each have their own priors. It is the extent of these priors that regulates the level of systematic error allowed for in the inference of $n_{\rm u}(z)$ from the clustering-z data. The systematic function and these priors are quantified in Section 4.2.

The covariance matrix Σ_w is estimated from simulated data through a jackknife (JK) approach, using the following expression (Quenouille 1949; Norberg et al. 2009):

$$\hat{\Sigma}(x_i, x_j) = \frac{(N_{\rm JK} - 1)}{N_{\rm JK}} \sum_{k=1}^{N_{\rm JK}} \left(x_i^k - \bar{x_i}\right)\left(x_j^k - \bar{x_j}\right), \tag{14}$$

where the sample is divided into $N_{\rm JK} = 1000$ subregions of roughly equal area, x_i is a measure of the statistic of interest ($=w_{\rm ur}$) in the i-th bin of the k-th sample, and $\bar{x_i}$ is the mean of the resamplings. The jackknife regions are safely larger than the maximum scale considered in our clustering analysis. The correction from Percival et al. (2021) is implemented when computing the inverse covariance, although it has a modest impact ($\sim$10 per cent on the amplitude of the covariance) given the number of jackknife regions and the data vector length.

Note that the clustering-z likelihood in equation (13) depends explicitly on the estimated bias and magnification coefficient $b_{\rm r}$ and $\alpha_{\rm r}$ of the reference sample, and depends implicitly on the cosmological model through the dark-matter clustering $w_{\rm DM}$. Thus in principle, this likelihood and the inferences on $n_{\rm u}(z)$ must be recalculated for each change in cosmological model. We have, however, tested numerically that the full expression for $\mathcal{L}[{\rm WZ}|n(z)]$ has negligible dependence on the cosmological parameters or the reference-sample properties once the marginalization over systematic nuisances $\mathbf{s}$ and $\mathbf{p}$ are done. This is because the systematic variables have enough freedom to absorb the small changes in the model wrought by changes in cosmology. It is therefore allowable for us to compute equation (13) using a fiducial cosmology and fiducial values of $b_{\rm r}$ and $\alpha_{\rm r}$, and use the inferred redshift distributions in a cosmological inference that might vary these parameters.

3 DATA AND SIMULATED DATA

This section describes the various photometric and spectroscopic catalogues that feed into the clustering-z measurements. The full analysis is also conducted on simulated catalogues; for each element of the real analysis, we also describe how its simulated counterpart was generated.

3.1 DES Y3 data

The DES observed $\sim$5000 square degrees of the Southern hemisphere in five different broad photometric bands (*grizY*) over 6 yr using the Dark Energy Camera (DECam, Flaugher et al. 2015), a 570-megapixel camera built by the DES Collaboration and stationed at the Cerro Tololo Inter-American Observatory (CTIO) 4-m Blanco telescope. DES will measure the shapes of about 300 million galaxies up to redshift $z \sim 1.4$. In this paper, we focus on the analysis of the first 3 yr (Y3) of observations. DES Y3 data span the full area of the survey, 4143 deg^2 after masking for foregrounds and problematic regions, a major advance over the 1321 deg^2 of DES Y1 (Drlica-Wagner et al. 2018; Troxel et al. 2018). The complete DES (Y6) reaches greater depth than Y3 data; furthermore, the data are more uniform in depth.. The total number of objects detected in DES Y3 is $\approx$ 390 000 000. Object detection and measurements are described in Sevilla-Noarbe et al. (2021).

3.2 Buzzard *N*-body simulation

We use one realization of the DES Y3 Buzzard catalogue v2.0 (DeRose et al. 2019). Initial conditions were generated using 2LPTIC

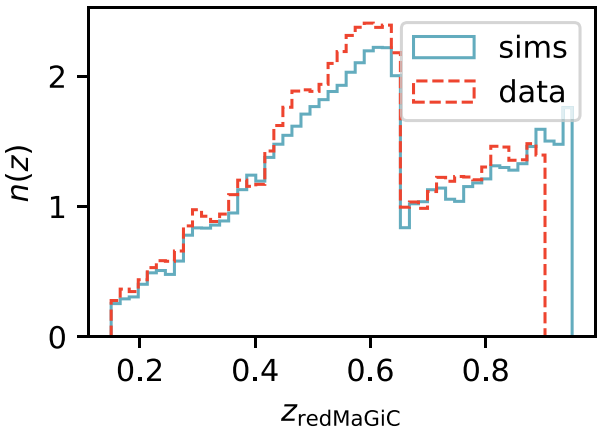

Figure 1. Redshift distributions of the *redMaGiC* samples, binned using the *redMaGiC* photo-z estimates, in data and in simulations.

(Crocce, Pueblas & Scoccimarro 2006) and the N-body run using L-GADGET2 (Springel 2005). Cosmological parameters have been chosen to be $\Omega_{\rm m} = 0.286$, $\sigma_8 = 0.82$, $\Omega_b = 0.047$, $n_s = 0.96$, $h = 0.7$. Light-cones are generated on the fly starting from three boxes with different resolutions and size (1050^3, 2600^3, and 4000^3 Mpc$^3h^{-3}$ boxes and 1400^3, 2048^3, and 2048^3 particles), to accommodate the need of a larger box at high redshift. Haloes are identified using the public code ROCKSTAR (Behroozi, Wechsler & Wu 2013) and they are populated with galaxies using ADDGALS (DeRose et al. 2019). Galaxies are assigned magnitudes and positions based on the relation between redshift, r-band absolute magnitude, and large-scale density found in a subhalo abundance matching model (Conroy, Wechsler & Kravtsov 2006; Lehmann et al. 2017) in higher resolution N-body simulations. SEDs are assigned to galaxies from the SDSS DR7 Value Added galaxy catalog (Blanton et al. 2005) by imposing the matching with the SED–luminosity–density relationship measured in the SDSS data. SEDs are K-corrected and integrated over the DES filter bands to generate DES $grizY$ magnitudes. Lensing effects are calculated using the multiple plane ray-tracing algorithm CALCLENS (Becker 2013), which provides weak-lensing shear, magnification, and lensed galaxy positions for the light-cone outputs.

3.3 Reference sample 1: *redMaGiC* galaxies

The first reference sample used in this clustering-z analysis consists of DES *redMaGiC* galaxies. The *redMaGiC* algorithm selects red luminous galaxies with high-quality photometric redshift estimates (Rozo et al. 2016). This is achieved by fitting each galaxy to a red sequence template; galaxies are then selected only if they pass a goodness of fit and luminosity threshold. In DES, *redMaGiC* galaxies are used as lens sample in the galaxy–galaxy lensing analysis and in the clustering analysis (Prat et al. 2020; Rodríguez-Monroy et al. 2020). Two samples are selected with different number density by means of two distinct luminosity thresholds: a first sample called 'high density' selected with a cut L/L$*$ > 0.5 and a sample called 'high luminosity' selected with a cut L/L$*$ > 1. A combined sample is then obtained by joining these two samples, using the high-density sample for redshifts $z < 0.65$, the high-luminosity sample for higher redshifts.

In simulations, the *redMaGiC* sample is selected with the same algorithm used in the data. A comparison between the redshift distributions for the *redMaGiC* samples in data and in simulations is shown in Fig. 1, illustrating the good agreement between the two. Small differences are due to small discrepancies in the evolution of the red-sequence between the simulation and the data. Both in simulations and in data, the *redMaGiC* sample is divided into 40 bins of width $\Delta z = 0.02$ spanning the $0.14 < z < 0.94$ range of the *redMaGiC* catalogue.[1] The particular choice of the bin width is not expected to impact our conclusions, as long as bins are small enough compared to the typical variation scales of the WL $n(z)$ and the galaxy-matter biases of the two samples. The total number of *redMaGiC* galaxies is 3041 935 in the data, and 2594 036 in the simulation. The difference in the number density is due to the aforementioned discrepancy in the evolution of the red-sequence between data and simulations. This implies that the statistical uncertainties of the clustering-z estimates obtained using the *redMaGiC* sample are larger in simulations compared to data. We do not expect this to be important, as we show in Section 4.1 that the clustering-z methodology is dominated by systematic uncertainties, and the statistical uncertainties are negligible.

We compare the typical *redMaGiC* photo-z scatter and bias found in data versus in simulations in Fig. 2. Since only a portion of the data have spec-z information, we reweight the magnitude distribution of the spectroscopic sample such that it matches the magnitude distribution of the *redMaGiC* galaxies before computing the statistics shown in Fig. 2. This reweighting is performed separately for each redshift bin. Note that the typical scatter of *redMaGiC* photo-z is similar to our bin width, which might call into question the choice of bin width for *redMaGiC* galaxies. However, we verify in Section 4.1 that even with this set-up, *redMaGiC* photo-z uncertainties are not a dominant source of systematic error for our methodology. Therefore, we decided that using a larger bin width for *redMaGiC* galaxies was not necessary.

Using cross-correlation techniques, Cawthon et al. (2020) noted that photo-z uncertainties in *redMaGiC* galaxies at $z > 0.8$ might be underestimated. We do not think this constitutes a problem for the current analysis, as *redMaGiC* photo-z uncertainties are a subdominant systematic in our methodology (Section 4.1), and clustering-z constraints at $z > 0.8$ are driven by the BOSS/eBOSS sample (Section 4.2.2).

A catalogue of random points for *redMaGiC* galaxies is generated uniformly over the footprint. Both in data and in simulations, weights are assigned to *redMaGiC* galaxies such that spurious correlations with observational systematics are cancelled. Note that due to low-statistics issues, the weights do not resolve fluctuations on scales relevant for this work, but only capture large-scale spurious correlations. The methodology used to assign weights is described in Rodríguez-Monroy et al. (2020), and it is the same for data and simulations. The main difference between data and Buzzard simulations is that the latter only models depth variations across the footprint, while data are subject to a larger number of systematics which are not modelled in simulations. This should not affect any conclusion drawn here: the weights remove the spurious dependence of the number density with respect to any systematic, regardless of their number, at least at the level needed for two-point correlation functions to be unbiased (Rodríguez-Monroy et al. 2020). This of course holds as long as all the systematics affecting the data are taken into account when producing the weights.

[1] We note that the simulated *redMaGiC* sample spans a slightly wider range in redshift; we none the less cut the redshift interval at $z = 0.90$ to be consistent with the data.

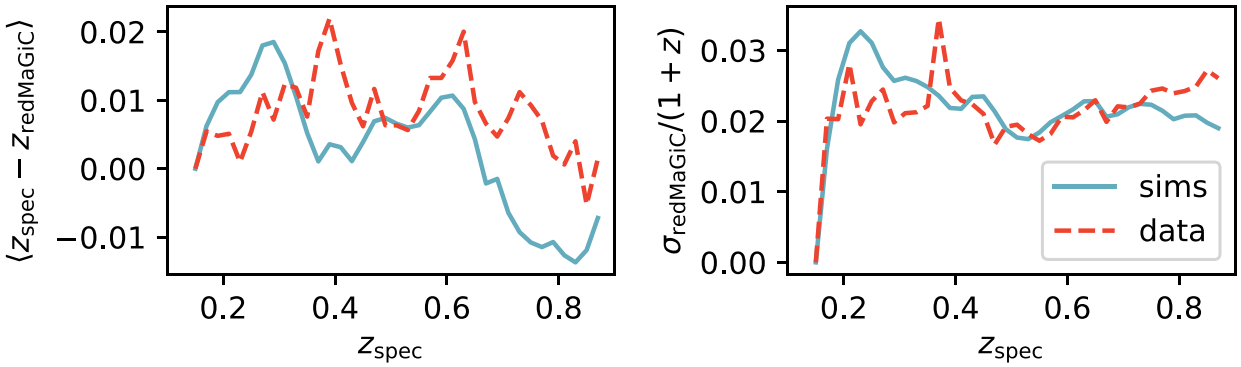

Figure 2. The bias (left) and scatter (right) of $z_{redMaGiC}$ for the simulated *redMaGiC* sample (solid lines) compared to the data (dashed lines).

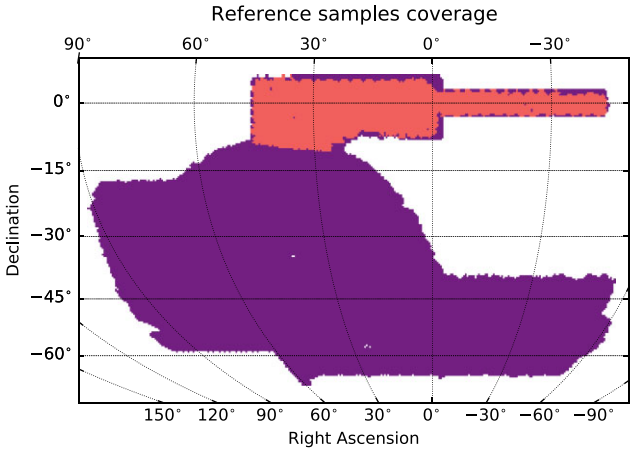

Figure 3. Spatial coverage of the two reference samples used in this work. Purple indicates the coverage by *redMaGiC* galaxies, pink indicates the coverage by BOSS and eBOSS galaxies.

3.4 Reference sample 2: spectroscopic galaxies

The second reference sample used in this work is a combination of spectroscopic samples from the Sloan Digital Sky Survey (SDSS, Gunn et al. 2006; Eisenstein et al. 2011; Blanton et al. 2017). In particular, we combine SDSS galaxies from BOSS (Dawson et al. 2013; Smee et al. 2013) and from eBOSS (Dawson et al. 2016; Ahumada et al. 2020; Alam et al. 2021). The BOSS sample includes the LOWZ and CMASS catalogues from the SDSS DR 12, fully described in Reid et al. (2016), while we included the large-scale structure catalogues from emission-line galaxies (ELGs, see Raichoor et al. 2017 for the target selection description), luminous red galaxies (LRGs, target selection described in Prakash et al. 2016), and quasi-stellar objects (QSOs) (Hou et al. 2021) from eBOSS, which were provided to DES for clustering-zs usage by agreement between DES and eBOSS. The different samples are stacked together, and used as one single reference sample in this work. Each sample comes with its own catalogue of random points, which account for selection effects. Different catalogues of random points are stacked together. We made sure the ratio of the number of randoms with respect to the number of galaxies was the same for each random catalogue before combining them. Both in simulations and in data, the BOSS/eBOSS sample is divided into 50 bins spanning the $0.1 < z < 1.1$ range of the catalogue (width $\Delta z \sim 0.02$). The area coverage is smaller compared to *redMaGiC* galaxies, as shown in Fig. 3. The redshift distribution of the samples is shown in Fig. 4, and the area coverage and number of objects of each sample are summarized in Table 1 . Note that some of the galaxies in the BOSS/eBOSS sample are also in the *redMaGiC* catalogue: ∼1 per cent of the *redMaGiC* galaxies are matched to ∼10 per cent of the BOSS/eBOSS galaxies, within 1 arcsec. We did not remove these galaxies from the *redMaGiC* sample, as they have a negligible impact both on our constraints and on the covariance between the two samples (as it will be clear in the following sections, the constraints from both samples are systematic-dominated).

Table 1. List of the spectroscopic samples from BOSS/eBOSS overlapping with the DES Y3 footprint used as reference galaxies for clustering-zs in this work.

Spectroscopic samples			
Name	Redshifts	N_{gal}	Area
LOWZ (BOSS)	$z \sim [0.0, 0.5]$	45 671	∼860 deg^2
CMASS (BOSS)	$z \sim [0.35, 0.8]$	74 186	∼860 deg^2
LRG (eBOSS)	$z \in [0.6, 1.0]$	24 404	∼700 deg^2
ELG (eBOSS)	$z \in [0.6, 1.1]$	89 967	∼620 deg^2
QSO (eBOSS)	$z \in [0.8, 1.1]$	7759	∼700 deg^2

To replicate the spectroscopic BOSS/eBOSS sample in simulations, we selected bright galaxies with similar sky coverage and redshift distribution as the ones in data. We did not try to further match other properties of the sample, e.g. the galaxy-matter bias likely differs from that of the real data. We note that the clustering-z methodology corrects for the reference bias, so at no point in the analysis of the real data are we assuming that the simulations have the same bias.

3.5 WL sample

The WL sample in data is created using the METACALIBRATION pipeline, which is fully described in Gatti et al. (2021). After creation of the DES Y3 'Gold' catalogue (Sevilla-Noarbe et al. 2021), the METACALIBRATION pipeline measures the shapes of each detected object. Selection cuts for the sample are described in Gatti et al. (2021) and are chosen from results of tests on both sky data and image simulations (MacCrann et al. 2022), and are designed to minimize systematic biases in the shear measurement. Galaxies are weighted by the inverse variance of shear measurement, which increases the statistical power of the catalogue. The final sample comprises 100 204 026 objects, for an effective number density of $n_{eff} = 5.59$ gal arcmin^{-2}. Galaxies are further divided into four tomographic

bins, and redshift distribution estimates for each of the tomographic bins are provided by the SOMPZ method (Buchs et al. 2019; Myles et al. 2021). The tomographic bins are selected such that they have roughly equal raw number density.

The WL sample is reproduced with high fidelity in the Buzzard simulation by applying flux and size cuts to the simulated galaxies that mimic the DES Y3 source selection thresholds. The WL galaxy sample in Buzzard is selected with the aim of reproducing the same selection applied in DES Y3 data in terms of size, signal-to-noise ratio, and colours. Shape noise has been added to the galaxies to match the measured shape noise of the DES Y3 WL sample.

3.5.1 *Photo-z estimates: SOMPZ*

The SOMPZ method uses spectroscopic and multiband photometric information, and data from a number of deep fields (Hartley et al. 2022) where additional photometry in the infrared bands *YJKs* and *u*-band is available, besides the standard 5-band (*grizY*) photometry available in the DES wide field. This additional information is used to break the degeneracies in the photo-z estimates of the DES wide-field galaxies (which have fewer bands available). This is achieved by creating two Self-Organizing Maps (SOM, Kohonen 1982), one mapping the deep/spectroscopic galaxies into a 2D grid of cells using their 8-band fluxes, and another mapping the WL sample galaxies into a 2D grid using the *riz* photometry. A probabilistic mapping from the wide-field SOM to the deep-field SOM is generated using the 'Balrog' source-injection simulations (Everett et al. 2020) and a map from the deep-field SOM to redshift is estimated using the spectroscopic data.

The tomographic bins are constructed as follows: a first set of edge values are arbitrarily selected. Each galaxy of the redshift sample is then assigned to the tomographic bin in which its redshift estimate falls. A number of galaxies at this point share the same photometry cell of the wide-field SOM and same tomographic bin, so the cell in its entirety is assigned to the bin to which the majority of its galaxies live. The initial bin edges are adjusted to yield approximately the same number of galaxies, and finally the whole procedure is repeated with the new bin edges. After completing this procedure, the final bin edges are [0.0, 0.358, 0.631, 0.872, 2.0] for the Y3 WL source catalogue.

The full Y3 SOMPZ procedure is described in Myles et al. (2021). A number of factors contribute to the error budget of the method: (1) shot noise (i.e. the limited number of galaxy redshifts available); (2) sample variance (i.e. the fact that the spectroscopic and deep fields span a limited area); (3) systematic uncertainties in the spectroscopic/multiband photometry samples; (4) uncertainty in the methodology in general; (5) photometric calibration uncertainties in the Y3 deep fields, i.e. the uncertainty on the zero-point calibration in each band.

The total error budget is dominated by the photometric calibration uncertainty in the low-redshift bin, while it is dominated by sample variance and biases in the spectroscopic/multiband photometric samples in the high-redshift bins (Myles et al. 2021).

The SOMPZ method incorporates methods for assessing the likelihood $\mathcal{L}[\mathrm{PZ}|n_u(z)]$ of obtaining the various SOMPZ data elements (SOM cell counts, etc.) given a candidate set of $n_u(z)$ redshift distributions for the tomographic bins, which account for shot noise and sample variance in the various catalogues used by SOMPZ. The construction of this likelihood and the methods for sampling candidate $n(z)$ distributions from it are given by Sánchez & Bernstein (2019). Potential selection biases in the spectroscopic redshift

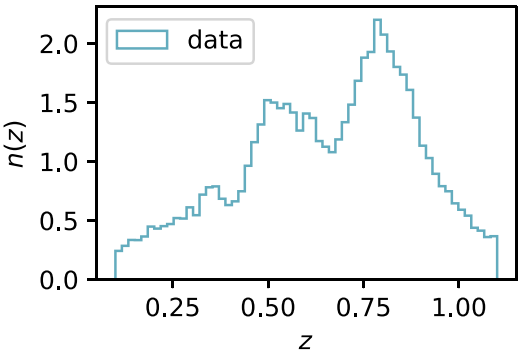

Figure 4. Redshift distribution of the BOSS/eBOSS sample in data.

assignments are estimated by compiling $n(z)$ realizations obtained by calibrating with three different sets of spectroscopic/multiband photometric samples. Redshift uncertainties related to the zero-point calibration are added *after* the SOMPZ realizations are informed by the clustering measurements (Myles et al. 2021). This is done for efficiency reasons and it does not affect the main results of this work.

The SOMPZ process is completely reproduced in simulations, including the creation of spectroscopic catalogues from small-area surveys, but these simulations do not take into account the uncertainties related to unknown redshift selection biases in the spectroscopic/multiband samples. As a result of the slight differences of the simulated Y3 source sample data equivalent, the bin edges in the equivalent Buzzard catalogue are [0.0, 0.346, 0.628, 0.832, 2.0]. Estimates of the $n(z)$ obtained in simulations are shown in Fig. 5.

4 RESULTS ON SIMULATIONS AND SYSTEMATIC ERRORS

In this section, we present the results of our two calibration strategies performed in simulations. In particular, we aim to evaluate the systematic uncertainties of each method, and verify that the calibration procedure in simulations works as expected. Note that at no point are the simulations used to make corrections to the data; rather the simulations are used to (1) estimate the level of uncertainty to assign to various systematic errors, and (2) validate that the method yields results for $n(z)$ consistent with truth.

Before focusing on the details of the two calibration procedures, we show in Fig. 6 the redshift distributions estimates obtained using the clustering-z $n_u(z)$ estimator (following equation 8) on simulations, compared to the true distributions. The angular scales considered in the clustering measurements have been chosen to span the physical interval between 1.5 and 5.0 Mpc. These bounds (which are applied to the data as well) are selected so that the upper bound is below the range used for the $w(\theta)$ statistics used in cosmological analyses, thus allowing the clustering-z likelihoods to be essentially statistically independent of cosmology, and permitting us to produce $n(z)$ samples in an MCMC chain that runs before, and independent of, the cosmology. The values of b_r in the clustering-z analysis are not required to match those used in the cosmological analyses. The lower bound is chosen to produce high signal-to-noise ratio *S/N* while mitigating failures of the linear bias model.

We start with an idealized case: the distributions shown in Fig. 6 are obtained using *redMaGiC* galaxies as a reference binned using true redshift. In simulations we also have an accurate estimate of $b_u(z)$, obtained from the autocorrelations of each of the tomographic bins of

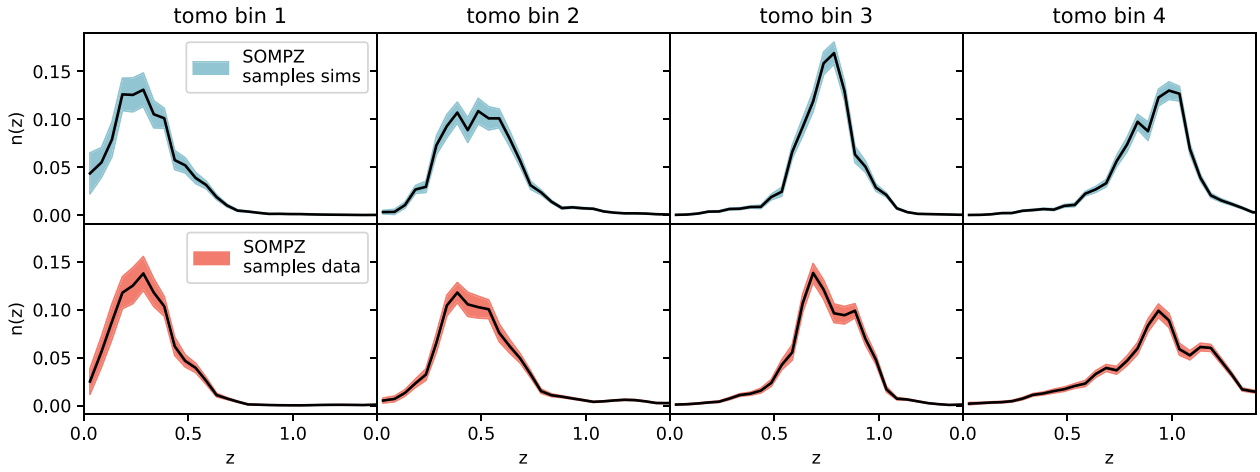

Figure 5. SOMPZ redshift distributions, as estimated in simulations (upper panels) and in data (lower panels), for the four tomographic bins considered in this analysis. The bands represent the 68 per cent confidence interval spanned by the SOMPZ n(z) realizations.

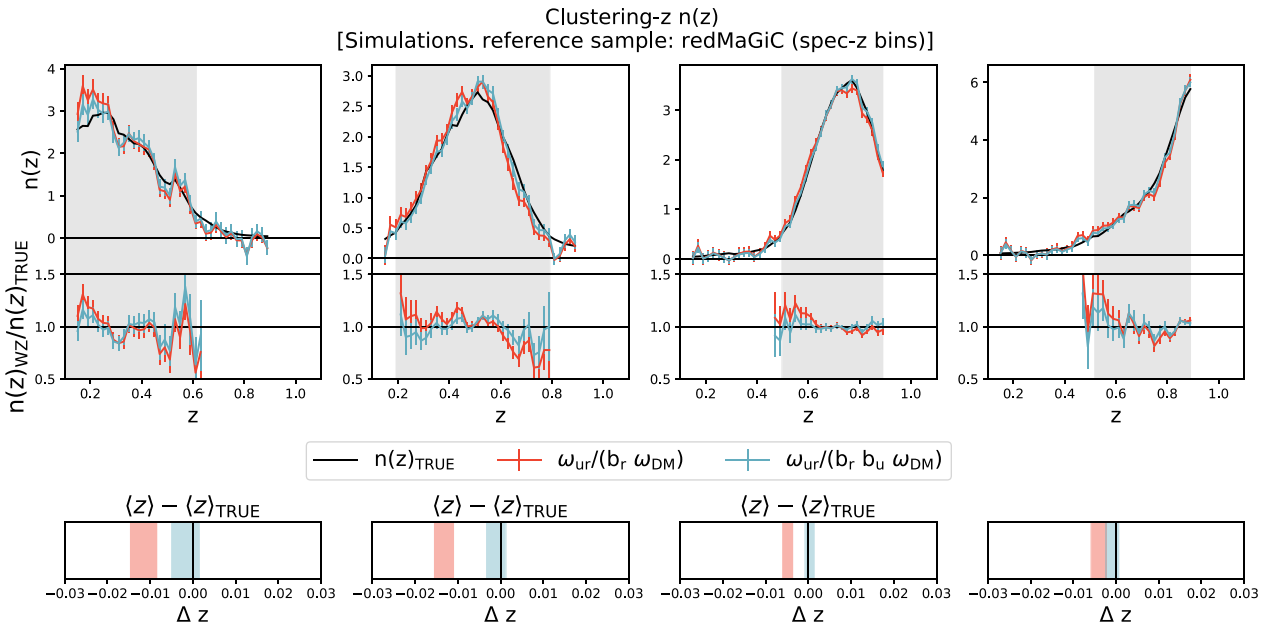

Figure 6. Sources redshift distributions estimated using clustering-z in simulations for an **idealized set-up** (see the text in Section 4), compared to the truth (black lines). The top panels show the redshift distributions; the middle panels show the ratio between the true $n(z)$ and the $n(z)$ estimated using clustering-z; and the bottom panels show the mean of the redshift distributions. The red lines represent the clustering-z estimates obtained using the estimator introduced by equation (8). The blue lines represent the clustering-z estimated obtained further correcting for the term b_u, which is only possible in simulations. The four different tomographic bins used in the DES Y3 cosmological analysis are shown. We used *redMaGiC* galaxies as the reference sample, binned using true redshifts. For this plot, we also subtracted from the clustering-z $n(z)$ estimates the expected magnification contribution in simulations (Appendix A); this has only a mild effect at high redshift ($z > 0.6$) for the first two bins. The redshift distributions are normalized over the same interval. The grey shaded regions indicate the interval considered for the mean matching method. The mean of the distributions shown in the bottom panels is computed only considering the grey intervals. Error bars only include statistical uncertainties.

the unknown sample, divided into thin bins of width $\Delta z = 0.02$.[2] This is not possible in data since the precision of the photometric redshift is not sufficient to divide the sample in bins of adequate width. Fig. 6 shows the impact on the estimated $n(z)$'s of assuming we know $b_u(z)$ with good accuracy (in cyan), dividing equation (8) by $b_u(z)$. We note that correcting for b_u drives both the shape of the distributions and the mean value closer to the truth, which are otherwise biased. As we cannot estimate b_u in data, this highlights that variation in b_u introduces a systematic uncertainty that has to be quantified. Note that the errors bars in Fig. 6 *only include statistical uncertainties*.

[2]In order to measure the autocorrelations, we generated randoms properly accounting for the WL mask. We also created systematic weights for the WL sample using the same procedure used for *redMaGiC* galaxies (although we found they have a negligible impact).

Table 2. Mean-matching method: We display the total systematic error budget on the mean redshift, for the two reference samples used in this work. We also report the contribution due to each single source of systematic uncertainty, as a function of tomographic bin. As for the *redMaGiC* systematic, we also report in parentheses the values of the uncertainties we would have obtained if we had not included the correction factor in the bias estimation (see Section 4.1).

Systematic	tomo bin 1	tomo bin 2	tomo bin 3	tomo bin 4
methodology:	0.002 ± 0.003	0.001 ± 0.002	0.000 ± 0.001	0.001 ± 0.002
magnification:	0.004	0.005	0.003	0.004
WL galaxy bias unc:	0.013	0.013	0.013	0.013
redMaGiC syst:	0.000 (0.014)	0.001 (0.007)	0.002 (0.000)	0.005 (0.003)
total systematic *redMaGiC*:	0.014	0.014	0.014	0.015
statistical *redMaGiC*:	0.003	0.002	0.001	0.002
total systematic BOSS/eBOSS:	0.014	0.014	0.014	0.014
statistical BOSS/eBOSS:	0.007	0.006	0.004	0.006

In the following subsections, we tested the accuracy of our calibration procedure using the two different approaches outlined in Section 2, i.e. the mean-matching and the full-shape methods.

4.1 Method 1 (mean-matching): systematic uncertainties estimation in simulations

We test in this section the mean-matching clustering-z photo-z calibration method. The metric used here to assess the accuracy of our methodology is the difference between the mean of the recovered redshift distribution and the true mean, as follows:

$$\Delta\langle z\rangle \equiv |\langle z\rangle_{\rm true} - \langle z\rangle_{\rm WZ}|. \quad (15)$$

As described in Section 2.2, $\langle z\rangle$ is calculated over a restricted redshift interval $\langle z\rangle_{\rm SOMPZ} - 2\sigma_{\rm SOMPZ} < z < \langle z\rangle_{\rm SOMPZ} + 2\sigma_{\rm SOMPZ}$ to reduce the impact of magnification[3]. The redshift intervals are of course also truncated at the bounds of the reference sample. The same redshift intervals are used for simulations as for data (see Fig. 6). The intervals used are [0.14,0.62], [0.18,0.80], [0.46, 0.90], [0.48, 0.90] for *redMaGiC* and [0.10,0.62], [0.18,0.80], [0.46, 0.98], [0.48, 1.06] for BOSS/eBOSS.

4.1.1 Systematic uncertainties

We quantify here the systematic uncertainties of the mean matching method. Since the mean-matching method reduces each $n(z)$ to its windowed mean $\langle z\rangle$, the systematic errors will be quantified by the uncertainties that they imply should be added (in quadrature) to the σ_z values of equation (11). We note that the absolute value of the terms in equation (8) are irrelevant for this method, as we are only interested in how they evolve with redshift. In principle, in the absence of magnification, assuming perfect reference sample redshift accuracy (e.g. *redMaGiC* redshifts to be exact), assuming that we are able to successfully estimate all the terms in equation (8), and assuming that we know the galaxy-matter bias evolution of the unknown sample, we should correctly recover the mean of the unknown redshift distributions. The above assumptions might not hold when applying this methodology in data, causing a systematic bias in the calibration, In particular, $\Delta\langle z\rangle$ can differ from zero because of the following reasons:

(i) The approximations that allowed us to factorize the integral in equation (1) into $b_{\rm r}(z)b_{\rm u}(z)w_{\rm DM}(z)$ might not hold (e.g. linear bias model, infinitesimally thin bins), leading to inaccuracies in the modelling at small scales. We will quote these as **methodology systematics**. This systematic does not depend on the reference sample used.

(ii) Magnification contribution. In the mean matching approach, we do not correct for magnification effects, as we cut the tails of the redshift distributions. This systematic quantifies how effective our cut is. We will refer to this as **magnification systematic**. This is a subdominant effect in our total error budget. We defer a detailed description of magnification effects and how they are evaluated to Appendix A.

(iii) The clustering-z estimator ignores the redshift evolution of the galaxy-matter bias of the unknown sample (**WL galaxy bias uncertainty**). This systematic does not depend on the reference sample used.

(iv) The reference sample is binned using photometric redshifts and not spectroscopic redshifts. This only applies to the *redMaGiC* case. We will refer to this as *redMaGiC* **systematic**.

We studied the performance of the estimator described in equation (8) for four cases, starting from an ideal environment free from the effects of systematics and introducing one uncertainty at the time, leading to a more complex, realistic case. This allows us to estimate separately the magnitude of each systematic independently. In the following tests, we will only use the *redMaGiC* galaxies as a reference sample to estimate the systematic uncertainties. Indeed, the BOSS/eBOSS sample should be affected by the same systematic uncertainties as the *redMaGiC* sample, except for the *redMaGiC* systematic.

We begin with the most ideal case possible, shown in Fig. 6, which we already described at the beginning of this section. Recall that for this case we used *redMaGiC* galaxies as a reference binned using true redshifts, we corrected for the bias evolution of the unknown and reference sample, and we corrected for the redshift evolution of the clustering of dark matter. The $\Delta\langle z\rangle$ mean for this case provides an estimate of the methodology systematic, and it is reported in the first line of Table 2. This value is compatible with zero within statistical uncertainty (estimated through jackknife resampling), indicating that for the scales considered in this work (1.5–5.0 Mpc), the approximation of linear bias model, and infinitesimally thin redshift bins are good enough for the purpose of calibrating the mean with clustering information.

[3]In principle, performing a symmetric cut in comoving distance rather than in redshift should reduce the impact of magnification effects more efficiently. We followed the DES Y1 prescription, which implements a symmetric cut in redshift. We note, however, that a symmetric cut in distance (rather than in redshift) would have changed the location of the interval edges by at most $\Delta z \sim 0.03$, hence it would have had a negligible impact on our methodology.

We next quantify the impact of ignoring the redshift evolution of the galaxy-matter bias of the unknown sample $b_u(z)$, as this cannot be measured in data. We estimate the size of this effect in the simulation by assuming a constant b_u for each tomographic bin, and we obtain the resultant shifts of $\Delta\langle z\rangle$ of (0.010,0.013,0.006,0.001). The effects of redshift-dependent $b_u(z)$ on the mean and on the shape of the clustering-z $n(z)$ are shown in Fig. 6: the red and blue values differ only in the presence of the $b_u(z)$ term in the latter. Given that the WL galaxy bias uncertainty is the dominant uncertainty of the clustering-z method, we take the conservative approach of assigning an RMS systematic value to every bin that is equal to largest $\Delta\langle z\rangle$ found in Buzzard, i.e. $\Delta\langle z\rangle = 0.013$ estimated for the second bin. This σ_z contribution is listed in the third row of Table 2.

Finally, we estimate the systematic uncertainty in $\langle z\rangle$ due to inaccuracies in the bin-shape integral in equation (6) for *redMaGiC* galaxies when they are placed into thin bins using their photo-z estimates. This is done in the simulation by comparing the $\langle z\rangle$ estimates obtained when binning the *redMaGiC* galaxies using true redshifts to estimates obtained when binning using *redMaGiC* photo-z. The photo-z accuracies of *redMaGiC* galaxies are better than those of the WL sample, but not as good as those of a spectroscopic sample. This can introduce two kinds of errors in $\langle z\rangle$: first, if all *redMaGiC* photo-z estimates were biased towards lower redshift, we would infer a similarly biased $n(z)$. Second, the change in shape or width of the $n_r(z)$ because of photo-z errors can cause $\int \mathrm{d}z n_r(z)^2$ to be wrong which propagates to a shift in $\langle z\rangle$.

The shifts $\Delta\langle z\rangle$ that result from binning the *redMaGiC* galaxies using *redMaGiC* photo-z rather than true redshifts are given in the fourth row of Table 2. We do not report statistical uncertainties, as they are negligible, since the shifts are computed taking the difference of two highly correlated measurements. The shifts are relatively small and unimportant in comparison to the b_u uncertainties. We also report in parentheses the errors in $\langle z\rangle$ we would have obtained had we not included the correction factor of equation (6) when estimating the galaxy-matter bias of *redMaGiC* galaxies. Given the difference between the two estimates, the correction due to the $n_r^2(z)$ integral clearly cannot be neglected when applying the methodology to data. Lastly, we also estimated the *redMaGiC* $\Delta\langle z\rangle$ using theory data vectors of the cross-correlation signal w_{ur}, and modelling the *redMaGiC* redshift distributions in each reference bin assuming the *redMaGiC* photo-z uncertainties estimated from data (Fig. 2), rather than the ones from the Buzzard simulation. This test delivered $\Delta\langle z\rangle$ of the same order of magnitude as the ones estimated directly in Buzzard and reported in Table 2.

Before reporting the total error budget for the mean matching method, we validate the assumption that we can assume a fixed cosmology when calculating the clustering of dark matter, $w_{DM}(z)$. Assuming different values for the cosmological parameters ($\Omega_m = 0.4$, $\sigma_8 = 0.7$) results in a negligible shift, $\Delta\langle z\rangle < 10^{-3}$.

The total error budget is reported at the end of Table 2, and is obtained by adding in quadrature all the single sources of errors, assuming they are independent. The dominant source of uncertainty is the potential redshift evolution of the WL sample, which we do not model in the mean-matching analysis of the real data or in the validation analyses of the simulations, which are described next.

4.1.2 *Application of the method in simulations*

In order to apply the mean matching method in simulations, we run our clustering measurements using a realistic set-up, for the two reference samples considered in this work. Fig. 7 compares the $n(z)$ distributions obtained from simulations with *redMaGiC* and BOSS/eBOSS as reference samples. In particular, *redMaGiC* galaxies have been binned using the *redMaGiC* photo-z estimates rather than the true redshifts and we did not correct for the bias evolution of the unknown sample. This plot highlights the differences between the two samples: *redMaGiC* has a smaller statistical uncertainty, but the BOSS/eBOSS sample has a wider coverage in redshift, helping especially at higher z. The distributions are compatible within errors. We note that in order to correct for the bias evolution of the reference sample when using *redMaGiC* galaxies as a reference, we have to apply a correction to the width of *redMaGiC* bins, as described in equation (6), to account for the broader distributions that *redMaGiC* bins have compared to a top-hat bin. This correction is shown in Fig. 8.

Once we have $n(z)$ clustering-z estimates, we first verify that the clustering-z windowed mean redshift estimates obtained using the two reference samples are both compatible within uncertainties (including systematic and statistical) with the truth, and with SOMPZ estimates. This is shown in the lower panel of Fig. 7. Note that the clustering-z windowed means are compatible by construction with the truth, given our modelling of the systematic uncertainties of the method.

We can then proceed combining the clustering-z information with the SOMPZ method. Recall that the SOMPZ method can provide samples of the $n_u(z)$'s from its posterior distribution. We can importance-sample these SOMPZ samples by assigning each a weight through the likelihood given by equation (11). As we have two reference samples, we multiply the likelihoods obtained using the *redMaGiC* and BOSS/eBOSS samples; we assume the two likelihoods share the WL galaxy bias uncertainty but are otherwise considered independent, which is a reasonable assumption given the fact the total error budget of the methodology is systematic dominated and the overlap between the two sample is minimal.

Fig. 9 shows, in red, the distributions of $\langle z\rangle$ over SOMPZ realizations, one panel for each tomographic bin. Note that in this case, $\langle z\rangle$ is taken over $0 < z < 4$, not restricted to narrower ranges where the clustering-z signal is measured and large. The blue curves show the distributions of $\langle z\rangle$ after having being weighted by the clustering-z likelihood. The means and standard deviations of $\langle z\rangle$ of the SOMPZ realizations are also reported in Table 3, with and without the importance weighting by mean-matching. The importance-weighted $\langle z\rangle$ values are fully consistent with unweighted SOMPZ realizations, and with the truth for the simulations.

The clustering-z information in fact offers little improvement in the constraints from the SOMPZ realization. The systematic errors we derive on $\langle z\rangle$ are larger than the statistical errors with DES Y3 data (Table 2), and also larger than the total errors estimated for the SOMPZ method (Fig. 7). This means that for the DES Y3 analysis, the mean-matching method can be useful as an independent cross-check of the SOMPZ methodology, but it does not significantly improve the constraints on the mean of the redshift distributions.

This is not entirely surprising, because we have seen that the dominant systematic error in the mean-matching method (indeed for clustering-z in general) is the uncertainty in the redshift evolution of the bias of the unknown sample, $b_u(z)$. Even a simple linear slope to $b_u(z)$ will be imprinted on the inferred $n_u(z)$ and shift $\langle z\rangle$, meaning that the dominant systematic error has its largest effect on this lowest order moment of $n_u(z)$. Thus in some sense, $\langle z\rangle$ is the statistic for which we should expect clustering-z techniques to be least informative. On the other hand, we expect $b_u(z)$, and other sources of systematic error in the clustering-z method, to be smooth, low-order functions of z. We will therefore look next into the ability of clustering-z data to constrain the full shape of $n_u(z)$.

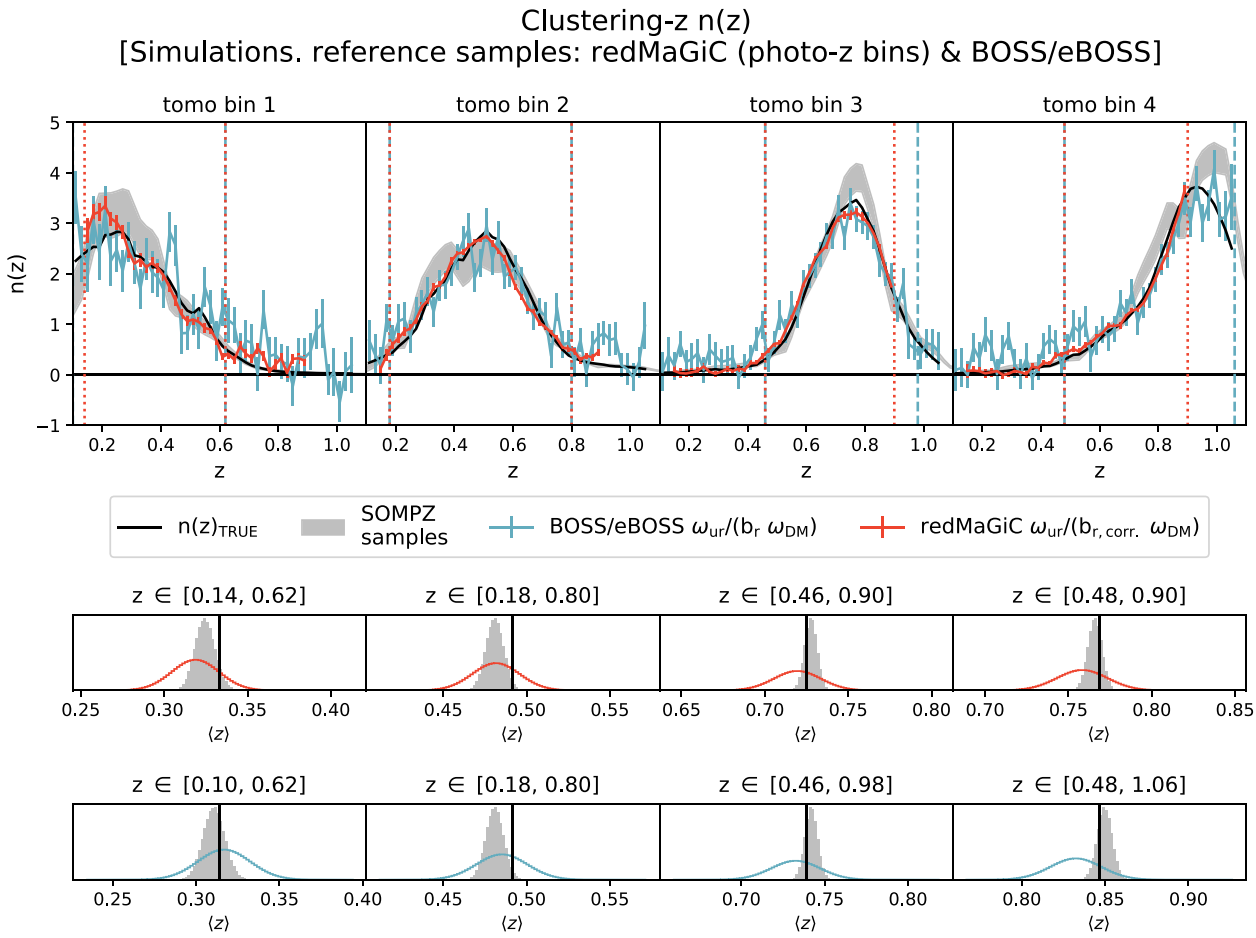

Figure 7. *Upper panels*: the redshift distributions estimated per equation (8) for a realistic set-up (see the text in Section 4), compared to the truth (black lines). We show both the redshift distributions obtained using *redMaGiC* galaxies as a reference sample, binned using their *redMaGiC* photo-z estimates, and the ones obtained using BOSS/eBOSS galaxies as a reference sample. The grey bands show, as a comparison, the 1σ region encompassed by the SOMPZ realizations. The vertical dotted (dashed) lines indicate the intervals where the windowed means of the *redMaGiC* (BOSS/eBOSS) have been computed. *Central and bottom panels*: windowed mean of the redshift distributions. The clustering-z estimates are represented by Gaussian histograms with mean equal to $\langle z\rangle_{\rm WZ}$ and σ equal to the uncertainty of the method. The error budget of the clustering-z mean redshift estimates includes both statistical and systematic uncertainties (estimated in Section 4.1 and reported in Table 2), contrary to what was shown in Fig. 6 that only reported statistical uncertainties.

4.2 Method 2 (full-shape): systematic uncertainty estimation in simulations

In the full-shape likelihood of Section 2.2, we produce a model for the $w(z)$ signal across the full redshift range covered by the reference samples (i.e. including the tails of the distributions) and produce a likelihood for the observed $w(z)$ data. In practice, this allows us to constrain the full shape of the redshift distributions, not only the mean. Here we use the Buzzard simulations to set the priors for the systematic-error parameters within this model.

4.2.1 Systematic uncertainty determination

Recall that in Section 2, specifically equation (12), the cross-correlation signal is modelled starting from a proposed value for $n_{\rm u}(z)$ (e.g. provided by SOMPZ), the (measurable) reference-population properties $b_{\rm r}(z)$ and $\alpha_{\rm r}(z)$, and nuisance parameters for the (poorly known) bias and magnification properties of the source population $b_{\rm u}(z)$ and $\alpha_{\rm u}(z)$. We will set these last two as constant over redshift and marginalize over broad priors on these constants, to flexibly model the magnification signal. The underlying function $w_{\rm DM}(z)$ is estimated assuming a cosmological model.

The final component of the $\hat{w}_{\rm ur}$ model is a function Sys(z, **s**) that multiplies the true clustering signal and will absorb the systematic errors described for the mean-matching method: failures of the linear-bias model itself; the unknown and redshift-dependent $b_{\rm u}(z)$; and possible errors in the $n_{\rm r}(z)$ functions for *redMaGiC* bins. The parameters **s** of this systematic function will be marginalized as well, as per equation (13).

Our strategy will be to determine what the Sys(z) function is in the Buzzard simulation, and then produce a prior on the **s** parameters which allows marginalization over a broad family of functions with similar form of deviation from unity. The Sys(z) function is given substantial freedom for low-order, smooth variation with z, as we expect from all of the systematic errors, leaving the finer-scale information in $w_{\rm ur}(z)$ to constrain fine-scale behaviour in $n_{\rm u}(z)$, i.e. the shape of $n_{\rm u}(z)$.

The blue data points in Fig. 10 plot the Sys(z) functions observed in the Buzzard simulations, for both reference samples. Namely they

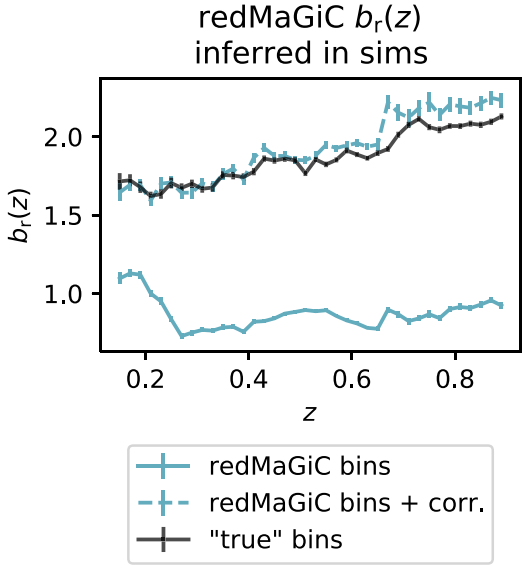

Figure 8. Redshift evolution of the galaxy-matter bias b_r of simulated *redMaGiC* galaxies, estimated with different binning. In particular, the black line has been obtained binning *redMaGiC* galaxies using the true redshift, and the solid light blue line has been obtained binning *redMaGiC* galaxies using *redMaGiC* photo-z. The lower amplitude is due to the larger effective bin width due to the photometric uncertainties. The light blue dashed lines is computed from the light blue solid line after correcting for the larger width of the bins, following equation (6).

plot

$$\mathrm{Sys}_{\mathrm{sim}}(z_i) = \frac{w_{\mathrm{ur}}(z_i)}{\hat{w}_{\mathrm{ur}}(z_i))}, \tag{16}$$

where the model uses the true $n_{\mathrm{u}}(z)$, $b_{\mathrm{r}}(z)$, and $n_{\mathrm{r}}(z)$ values. We evaluate and plot this ratio only in the z interval where the w_{ur} signal is large enough to have good signal-to-noise ratio and subdominant magnification contribution. The *redMaGiC* $w_{\mathrm{ur}}(z)$ uses *redMaGiC* photo-z's for binning, just as the real data do.

The $\mathrm{Sys}_{\mathrm{sim}}$ ratio deviates from unity due to systematic effects, as expected. We quantify this by the RMS of $\log\left[\mathrm{Sys}_{\mathrm{sim}}(z)\right]$, which are measured to be (0.11, 0.07, 0.07, 0.11) for the *redMaGiC* tomographic bins and (0.18, 0.15, 0.10, 0.15) for BOSS/eBOSS. From this we conservatively decide that the Sys function needs to have the freedom to have RMS (log) fluctuations of $\approx$0.15 as 1σ deviations under its $p(\mathbf{s})$ function.

We seek a parametric function $\mathrm{Sys}(z;\mathbf{s})$ and a prior $p(\mathbf{s})$ which have these desired properties:

(i) The function and prior yield a good fit to the $\mathrm{Sys}_{\mathrm{sim}}$ measured in Buzzard.

(ii) The prior can be tuned to yield typical RMS variations in $\log\left[\mathrm{Sys}(z)\right]$ at similar level to that seen in Buzzard.

(iii) The parametric form allows a similar smoothness of variation as seen in Buzzard, i.e. similar number of 'wiggles' across the $0 < z < 1.2$ range where the WL source galaxies lie.

(iv) The RMS of $\log\left[\mathrm{Sys}(z;\mathbf{s})\right]$ as we vary $\mathbf{s}$ under the prior $p(\mathbf{s})$ is a flat function of z.

(v) The prior on $\mathbf{s}$ is simple to construct and to use in a Hamiltonian Monte Carlo chain.

We chose the Sys (z, s) function to be given by:

$$\log[\mathrm{Sys}\,(z,\mathbf{s})] = \sum_{k=0}^{M} \frac{\sqrt{2k+1}}{0.85} s_k P_k(u), \tag{17}$$

$$u \equiv 0.85 \frac{z - 0.5(z_{\max} + z_{\min})}{(z_{\max} - z_{\min})/2}. \tag{18}$$

with $P_k(z_i)$ being the k-th Legendre polynomial, M is the maximum order, and the second line linearly remaps the z interval $[z_{\min}, z_{\max}]$ to $[-0.85, 0.85]$. The fraction under the summation makes the basis functions close to orthonormal so that the RMS of log(Sys) is $|\mathbf{s}|^2$. The prior $p(\mathbf{s})$ is chosen to be a simple diagonal normal distribution with standard deviations $\{\sigma_{s0}, \ldots, \sigma_{sM}\}$ and means of zero. Mathematical details of this choice for the systematic function and its prior are given in Appendix B.

A distinct set of nuisance parameters $\mathbf{q} = \{\mathbf{p}, \mathbf{s}\}$ (with $\mathbf{p} = \{b_{\mathrm{u}}', \alpha_{\mathrm{u}}'\}$) are assigned to each combination of tomographic bin and reference sample, and each of these eight sets of w_{ur} measurements are fit independently. We set $[z_{\min}, z_{\max}]$ to span the full range of the reference catalogue, [0.14,0.90] for *redMaGiC* and [0.10,1.06] for BOSS/eBOSS. We set $M = 5$ and we set the σ_{s_i} to yield an expectation value of 0.15 for the RMS of $\log\left[\mathrm{Sys}(z)\right]$. The order M was chosen by finding the value beyond which the RMS residual stopped decreasing for a fit of equation (17) to the $\mathrm{Sys}(z)$ function found in the simulated *redMaGiC* $w_{ur}(z)$ data. The σ_{s_i} prior is set to make the simulated $\mathrm{Sys}(z)$ functions be $\approx 1\sigma$ fluctuations from a constant. Since e^{s_0} is approximately the mean bias of the unknown sample, and we expect the mean bias b_r to be more uncertain than the variation with redshift, we treat the prior on s_0 somewhat differently, giving it a wide prior $\sigma_{s0} = 0.6$. The RMS of 0.15 is then allocated among the remaining elements $k \geq 1$ of $\mathbf{s}$ which model redshift-dependent systematic errors.

The nuisance parameter b_{u}' used in magnification estimation is given a Gaussian prior with $(\mu, \sigma) = (1., 1.5)$ (which encompasses the bias of the WL sample as measured in simulation). The other magnification nuisance α_{u}' is given a mean estimated from image-injection simulations (Appendix A) and a conservatively large uncertainty of $\sigma = 1$.

The dashed curves in Fig. 10 plot the Sys functions obtained from the maximum-posterior fits to the simulations' $w_{\mathrm{ur}}(z)$ data, combining the priors on the nuisance parameters with the likelihood of equation (13). In all cases, the best fit models succeed in capturing the slowly varying component of the systematic. In some bins, some of the rapid variations in redshift are not well captured – this is expected, as we truncate the polynomial of the Sys function to order $M = 5$. While this could be improved by increasing the maximum order M, we find in practice that these small discrepancies cause no significant bias in the recovered redshift distributions when the method is applied in simulations (see below). The fitted functions remain well behaved over the full w_{ur} redshift range even though the fit is done only for redshifts with strong signals. We conclude that this formulation of the systematic errors is sufficient to model the systematic errors in our clustering-z measurement in the Buzzard simulation, and we assume that marginalization over $\mathbf{q}$ will allow us to capture the uncertainties present in the real data as well.

The grey curves in Fig. 10 show a few examples of $\mathrm{Sys}(z;\mathbf{s})$ functions obtained by random sampling of the prior $p(\mathbf{s})$. This illustrates the flexibility of our model for the systematic uncertainty, which is able to model a large variety of curves.

It is useful to ask whether this implementation of systematic errors in the full-shape method is consistent with the systematic uncertainties derived for the mean-matching method. This can be

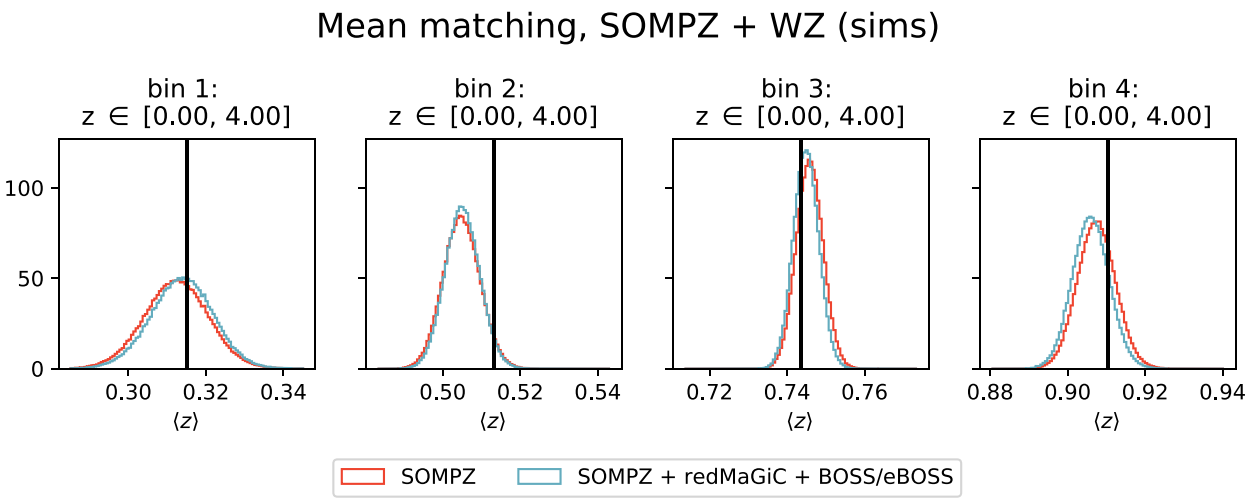

Figure 9. Mean redshift posteriors for the four tomographic bins obtained using the mean matching method in simulations. The red histograms represent the distribution of the mean redshift of the SOMPZ realizations, whereas the light-blue histograms show the mean redshift posteriors of the SOMPZ realizations using the clustering-z likelihood. The mean redshift of the SOMPZ realizations has been computed over a wide redshift interval ($0 < z < 4$), also including the redshift range where there is no clustering-z information.

Table 3. Simulations. The mean redshift estimates of the SOMPZ distributions with and without clustering-z information, in simulations.

Case	tomo bin 1	tomo bin 2	tomo bin 3	tomo bin 4
True $\langle z \rangle$:	0.315	0.513	0.743	0.910
SOMPZ $\langle z \rangle$:	0.312 ± 0.008	0.505 ± 0.005	0.746 ± 0.003	0.907 ± 0.005
SOMPZ + WZ (mean-matching) :	0.314 ± 0.008	0.505 ± 0.004	0.745 ± 0.003	0.906 ± 0.005
SOMPZ + WZ (full-shape) :	0.312 ± 0.009	0.507 ± 0.005	0.747 ± 0.004	0.907 ± 0.005

done by drawing many realizations of $\mathbf{s}$ from its prior, constructing a model $\hat{w}_{\rm ur}$ data vector using each realization of Sys$(z, \mathbf{s})$, and then treating this model as data input to the mean-matching method. Each realization of $\mathbf{s}$ then yields an estimate of $\Delta\langle z \rangle$ with respect to the true distribution. We obtained a typical $|\Delta\langle z \rangle|$ in the range 0.010–0.015 depending on the tomographic bin, in very good agreement with the total systematic uncertainties estimated in Table 2 for the mean-matching method.

4.2.2 *Application of the method in simulations*

Once our family of systematic functions is determined for the full-shape method, we may proceed to validating the performance of the combination of SOMPZ and the full-shape clustering-z method on the Buzzard simulations. This combination is implemented (both in simulations and in data) by sampling the $n_{\rm u}(z)$ functions for all four tomographic bins from a posterior defined by the product of:

(i) the SOMPZ probability defined by Myles et al. (2021);
(ii) the clustering-z probability defined by equation (13) for the $w_{\rm ur}(z_i)$ measured against the *redMaGiC* sample, marginalized over $\mathbf{q}$ as described in Appendix B;
(iii) and likewise, the marginalized clustering-z probability derived for the BOSS/eBOSS sample, marginalized over $\mathbf{q}$ as described in Appendix B;

The clustering-z probabilities use $w_{\rm ur}(z)$ over the full redshift range of their respective reference samples. The reference-sample magnification coefficients $\alpha_{\rm r}$ and the cosmology used to derive $w_{\rm DM}(z)$ are held fixed to nominal values. We verify below that the choices of $\alpha_{\rm r}$ and cosmology have insignificant impact on the outcome of the full-shape method. In this case, contrary to the mean-matching method, we consider the *redMaGiC* and BOSS/eBOSS likelihoods independent, i.e. they do not share the WL galaxy bias uncertainty. We did this because in the full-shape case we did not split our systematic function into different source of errors, owing to an increasing complexity in the modelling. Given the flexibility of Sys $(z, \mathbf{s})$ and the conservative choice on the RMS of log $\left[\text{Sys}(z; \mathbf{s})\right]$, considering the *redMaGiC* and BOSS/eBOSS likelihoods independent should not be an issue for the methodology. The sampling of the joint SOMPZ+WZ posterior is done using a Hamiltonian Monte Carlo method described in Bernstein (in preparation).

Fig. 11 compares the Buzzard true redshift distribution to the distributions drawn from only the SOMPZ likelihood and the distributions drawn from the joint SOMPZ+WZ posterior. The distributions of the mean redshifts per bin in the lower panels are not shown, but it is reported in Table 3. It shows that the full-shape clustering-z likelihood adds little information on these mean z's. This is as we expect from the results and discussion of the mean-matching method in Section 4.1.2. The plots in Fig. 11, however, show that the addition of full-shape clustering-z likelihood produces a remarkable improvement in the fidelity of the shape of $n_{\rm u}(z)$ to the truth. To better quantify the improvement, we also show the signal-to-noise ratio (S/N) of the $n_{\rm u}(z)$ estimates, defined as the ratio between the SOMPZ $n_{\rm u}(z)$ and the 68 per cent confidence interval of the SOMPZ realizations. The S/N is generally increased by the inclusion of the clustering-z information; in particular, the S/N is increased up to a factor of 3 in the relevant redshift range where $n(z)$

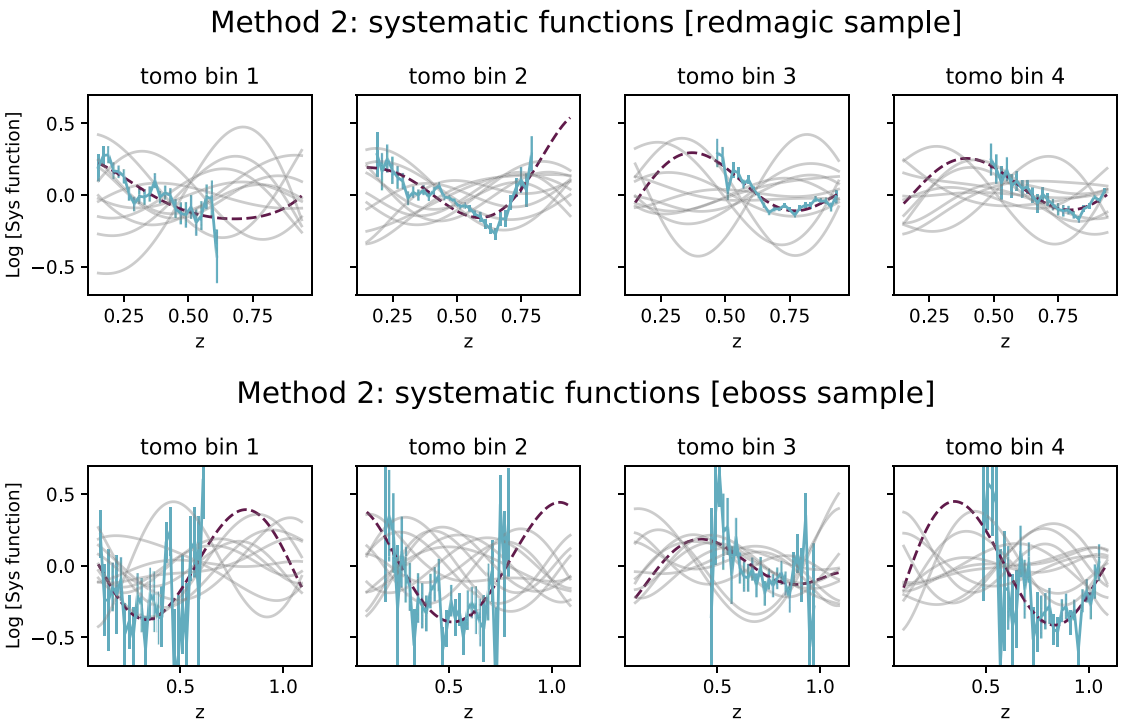

Figure 10. Systematic uncertainties of the full-shape method as measured in simulations following equation (16), for the four tomographic bins and for the two reference samples considered (*redMaGiC* upper panels, BOSS/eBOSS, lower panels). The measured systematic uncertainties are represented by the light blue lines; the purple dashed lines represent the best fitting model. The grey lines represent 10 random realizations of the systematic uncertainty model assumed for the full-shape method and described by equation (17). Note that the rapid upturn of the systematic function in bin 2 in the case of the *redMaGiC* sample is due to a rapid evolution of the galaxy-matter bias of the unknown sample, related to a strong evolution of the properties of the galaxy population. Such an evolution is also present in other bins, but it is milder. When the BOSS/eBOSS sample is used, the lower sensitivity does not allow to appreciate this rapid change in the slope of the systematic function.

is substantially different from 0. In the same S/N panels of Fig. 11, we also show the contribution to the S/N increment due to *redMaGiC* galaxies or BOSS/eBOSS galaxies alone. The latter sample mostly contributes in the redshift range $0.8 < z < 1.0$, whereas most of the clustering-z information at lower redshift comes from *redMaGiC* galaxies.

The SOMPZ method has strong fine-scale fluctuations in $n_{\rm u}(z)$ due to sample variance on the small regions of sky used for its deep imaging and spectroscopy. The clustering-z correlation functions, on the other hand, are measured over the full DES Y3 footprint and have high *S/N* level. Although the clustering signal has a strong systematic uncertainty from the unknown WL bias, this systematic is slowly varying as a function of redshift and has less fine-scale fluctuations. The clustering-z likelihood is thus able to drive the $n_{\rm u}(z)$ outputs to a smooth distribution, at least over redshifts where clustering-z reference samples are available.

We remind the reader that the clustering information alone cannot be used to infer the $n_{\rm u}(z)$, as the reference samples used in this work do not span the whole redshift range relevant for the DES Y3 $n_{\rm u}(z)$. None the less, we can try to understand in simulations if the full-shape method would be unbiased independently of the SOMPZ information. We did this by importance-sampling realizations of the true $n_{\rm u}(z)$s shifted around their mean redshift, and by assigning to each sample a weight through the likelihood given by equation (13). This test allowed us to recover the true $n_{\rm u}(z)$ within uncertainties, hence proving the method to be unbiased; for more details, see Appendix B.

Finally, we verify that the choices of the parameters $\alpha_{\rm r}$ or the cosmology assumed to compute $w_{\rm DM}$ do not impact the methodology. We find that assuming different values for the cosmological parameters ($\Omega_{\rm m} = 0.4$, $\sigma_8 = 0.7$) results in a shift in $\Delta\langle z\rangle < 10^{-3}$ on the calibrated SOMPZ redshift distributions. Concerning magnification, in order to roughly asses the impact of the exact values of the magnification coefficients $\alpha_{\rm r}$, we verified that assuming values for $\alpha_{\rm r}$ that are $-1\times$ the fiducial ones resulted in shifts $\Delta\langle z\rangle < 10^{-3}$ (see Appendix A for more details). We conclude that the full-shape likelihoods, like the mean-matching, can be calculated in advance of and independent from the cosmology chains.

5 APPLICATION TO DATA

We apply the clustering-z methods to DES Y3 data by first measuring the angle-averaged $w_{\rm ur}(z_i)$ (equation 2) of each WL source tomographic bin sample against the *redMaGiC* and BOSS/eBOSS samples described in Section 3. These cross-correlation data are plotted in Fig. 12. Note the exceptionally high *S/N* level of the *redMaGiC* data in particular, even at the rather fine binning of $\Delta z = 0.02$ that we use throughout. Bin-by-bin estimates of the reference bias $b_{\rm r}(z_i)$ are obtained using equation (6), with a dark-matter $w_{\rm DM}(z_i)$ predicted from theory for nominal cosmological parameters (Planck Collaboration VI 2020).

Note that for the *redMaGiC* galaxies we calculated $b_{\rm r}(z_i)$ applying the correction to the galaxy-matter bias of the reference sample described by equation (5), using the fraction of the *redMaGiC* galaxies

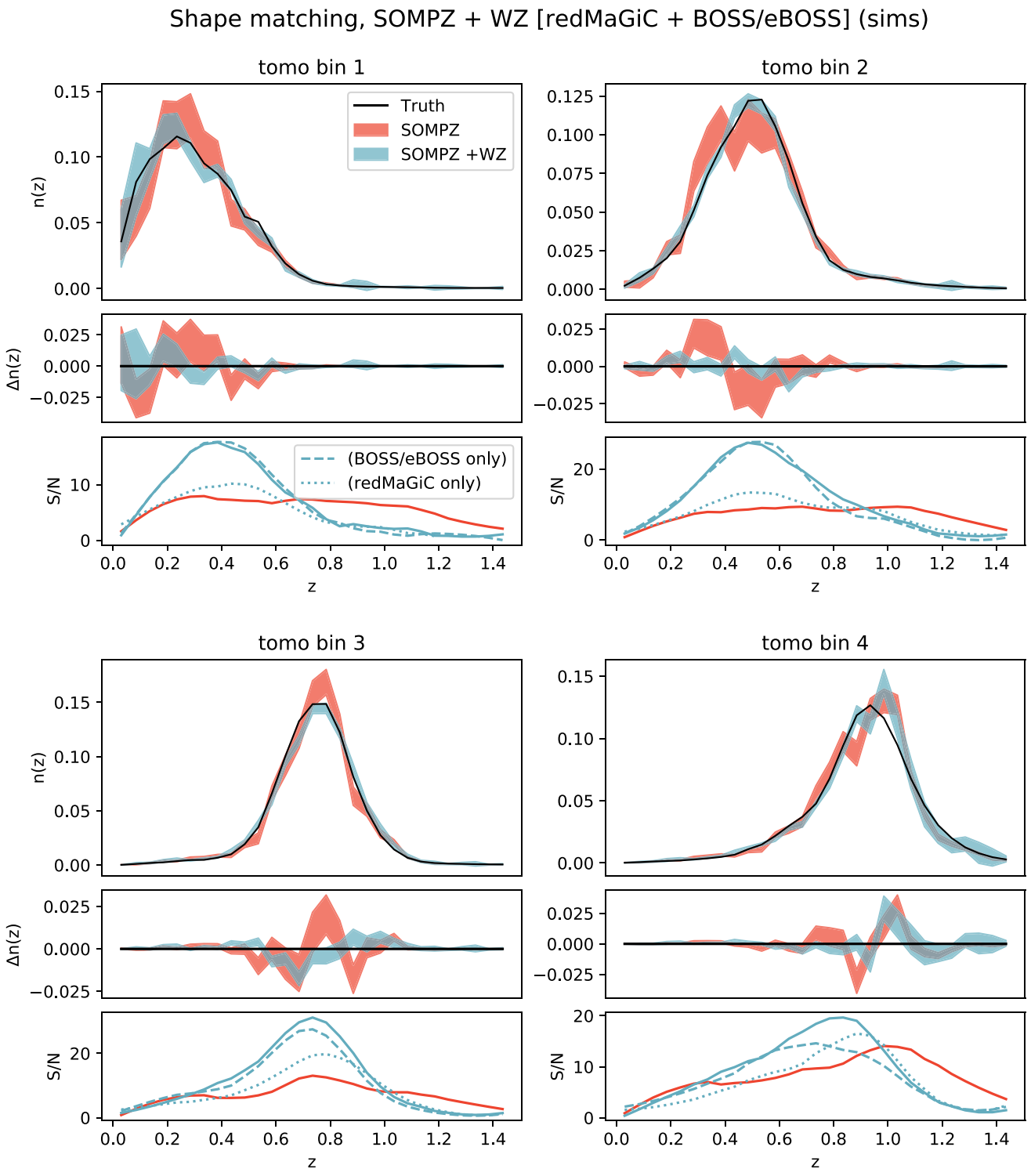

Figure 11. For each tomographic bin, three panels are shown. *Upper panels:* SOMPZ redshift distributions, as estimated in simulations, with and without clustering information (full-shape method). The bands encompass 68 per cent confidence interval of the SOMPZ $n(z)$ realizations. *Central panels:* difference between the recovered $n(z)$ and the true $n(z)$ in simulations. *Lower panels:* S/N, defined as the ratio between the SOMPZ $n_u(z)$ and its 68 per cent confidence interval of the SOMPZ realizations, with and without clustering information. The dashed (dotted) line has been obtained only using clustering-z constraints from *redMaGiC* (BOSS/eBOSS) galaxies.

which have a spectroscopic redshift. As *redMaGiC* galaxies with spec-z counterparts tend to have brighter magnitudes compared to the full *redMaGiC* sample, we have applied a magnitude reweighting to those galaxies before computing the correction, so as to up-weigh (down-weigh) *redMaGiC* galaxies under (over) represented in the spec-z subsample. After the reweighting, the spec-z sample had the same magnitude distribution of the full *redMaGiC* sample. Imperfections in this process should be small based on the tests in

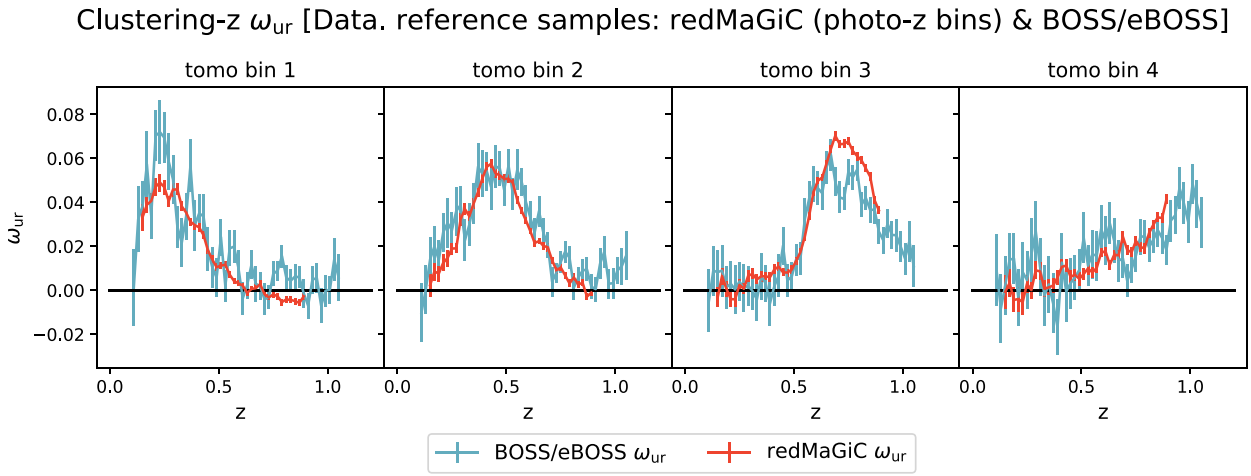

Figure 12. The measured $w_{ur}(z)$ for the DES Y3 data are plotted for each of the four tomographic bins, using reference samples from BOSS/eBOSS (blue) and *redMaGiC* (red).

previous sections and are included in the systematic uncertainties of the two methods.

Lastly, we note that the *redMaGiC* estimates show a small, negative tail at high redshift, for the first WL tomographic bin. We believe this is due to a systematic effect not corrected by our lens weights, rather than magnification, which should be positive at those redshifts, according to our estimates. The Balrog estimates of the magnification coefficients should also include realistic systematic and observational biases, which might lead to negative magnification; the fact that our estimates are none the less positive indicates that this effect is due to some systematic that affects the *redMaGiC* number density and that anticorrelates with the WL density distribution. We know, indeed, that the *redMaGiC* sample is affected by some residual systematics, which does not affect cosmology (DES Collaboration 2021; Pandey et al. 2021), but manifests as a scale-, redshift-, and sky-area-independent phenomenological decorrelation parameter. Given the small amplitude of this effect, the fact that we also have constraints from another independent sample (BOSS/eBOSS), and that our clustering-z constraints are compatible with SOMPZ and shear-ratio (Sánchez et al. 2021) prior to combination, we believe this should have a negligible impact on our results.

5.1 Mean-matching results

We use the mean-matching method as an independent check on the SOMPZ estimates of $n_u(z)$ in each tomographic bins. This begins by calculating the naive (linear-theory, no-magnification, constant-b_u) redshift distribution $\tilde{n}_u(z_i)$ from equation (8), plotted in Fig. 13. We show the distributions obtained with the two reference samples, and, for comparison, the 1σ region encompassed by the SOMPZ realizations.

Following the prescription for mean-matching in equation (9), we first compute the mean of the *redMaGiC* and BOSS/eBOSS clustering-z distributions in the redshift interval where they overlap, also excluding the tails (as detailed at the beginning of Section 4.1). We measure differences in $\langle z \rangle$ of $(-0.009 \pm 0.010,\ 0.006 \pm 0.009,\ 0.005 \pm 0.006,\ 0.022 \pm 0.014)$, for the four tomographic bins. The quoted uncertainties take into account the statistical and systematic uncertainties as reported in Table 2, except for the WL galaxy bias uncertainty that is assumed to be shared by the two samples. The statistical uncertainties are estimated through jackknife resampling. Statistical and systematic uncertainties are added in quadrature. We then compare the $\langle z \rangle$ values derived for the clustering-z with two reference samples and the SOMPZ estimates of $n_u(z)$: this is shown in the lower panels of Fig. 13. In this case the full systematic mean-matching uncertainty from Table 2 has been included in the clustering-z values. The clustering-z values are fully consistent with the SOMPZ values in the mean-matching statistic, although they are weaker. The behaviour is very similar to what was seen in simulations.

5.2 Full-shape results

Following the procedure used on the simulations, we define a full-shape clustering-z likelihood using equations (12) and (13). We assume fiducial values for the magnification parameters for the *redMaGiC* sample, as estimated using Balrog (Suchyta et al. 2016; Everett et al. 2020). We do not have an estimate of the magnification parameters for BOSS/eBOSS galaxies available, so we assumed the same values used for *redMaGiC* galaxies. We confirm, however, that assuming values for the magnification parameters that are -1 $\times$ the fiducial ones resulted in no relevant effect on the mean of the resultant redshift distributions. The nuisance-parameter priors derived from simulations in Section 4.2 are used, including those specifying the allowed variation with z in $b_u(z)$ and other elements of the Sys(z) function.

Before applying the full-shape method, we checked that the fiducial $\hat{w}_{ur}$ model on data (obtained using SOMPZ $n_u(z)$ as baseline) was compatible with the measured w_{ur} marginalized over the systematic function Sys(z). This check has been performed separately for *redMaGiC* and BOSS/eBOSS. We then use the Hamiltonian Monte Carlo method to draw samples from the joint posterior distribution of the SOMPZ likelihood and the clustering-z likelihoods for both *redMaGiC* and BOSS/eBOSS data. Fig. 14 shows the 68 per cent confidence interval of the $n_u(z)$ samples from the SOMPZ+WZ posterior, as well as those from the pure SOMPZ posterior. At redshifts where clustering-z information is available, it greatly reduces the point-by-point uncertainties in $n_u(z)$, just as in the simulations. The clustering-z full-shape method is thus very successful at reducing the impact sample variance on SOMPZ estimators. This combined estimator also shows no sign of negative tail at high redshift in the first tomographic bin (as seen, instead, in

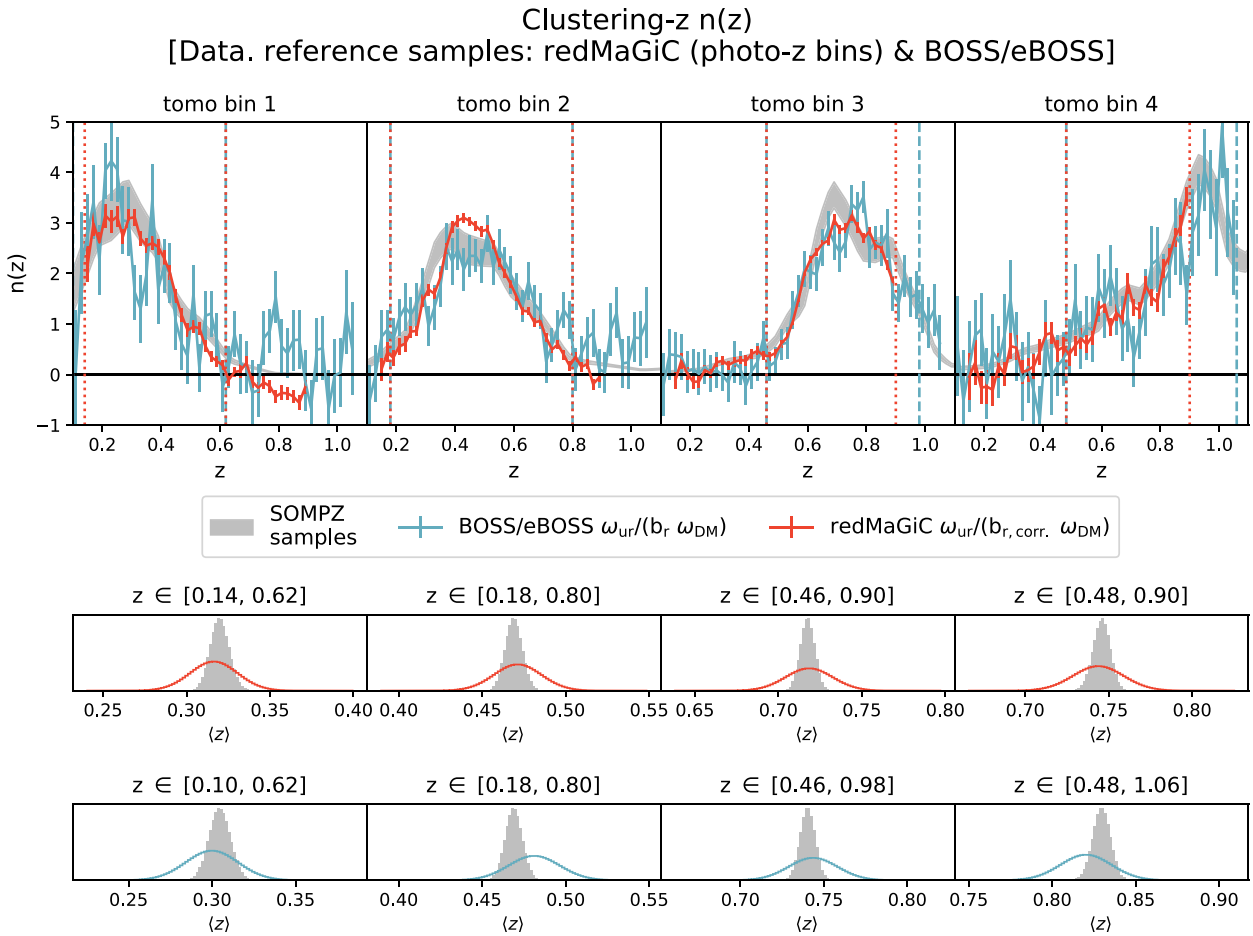

Figure 13. Illustration of the agreement among the SOMPZ $n(z)$ and the clustering-z $n(z)$ obtained using the naive redshift estimator and *redMaGiC* and BOSS/eBOSS galaxies as a reference sample. *Upper panels*: the naive redshift distributions estimated per equation (8) using clustering-z in data (i.e. no corrections for systematic errors or magnification). That is, these are the distributions employed in the mean-matching method. The redshift distributions obtained using *redMaGiC* galaxies as a reference sample, binned using their *redMaGiC* photo-z estimates, are in red. Those using BOSS/eBOSS galaxies as a reference sample are in blue. The grey bands show the 1σ region encompassed by the SOMPZ realizations. The vertical dotted (dashed) lines indicate the intervals where the windowed means of the *redMaGiC* (BOSS/eBOSS) have been computed. The lower panels plot the windowed mean redshifts $\langle z \rangle$ for each bin, as per equation (9), for the two clustering-z reference samples and for the SOMPZ samples. The clustering-z estimates are represented by Gaussian histograms with mean equal to $\langle z \rangle_{\rm WZ}$ and σ equal to the uncertainty of the method. The SOMPZ histograms are obtained from the mean redshift of the SOMPZ $n(z)$ realizations. Good agreement is seen among all three estimators.

the clustering measurement, Fig. 12). This stresses the importance of a combined analysis, which is more robust and is able to remove some of the potential problems or systematics affecting each of the two estimators when used individually.

The averages and standard deviations of the mean-z distributions of the SOMPZ and SOMPZ+WZ posteriors are listed in Table 4, along with the results of importance-weighting the SOMPZ samples with the mean-matching likelihood in equation (11). As expected from the simulations, the clustering-z information does not substantially alter the bin means derived from photo-z methods, in both the mean-matching and full-shape methods. The significant improvement in shape accuracy, as seen in Fig. 14, is the principal product of the clustering-z method for DES Y3 analyses.

6 CONCLUSIONS

This work describes the use of clustering measurements to constrain the WL source galaxy redshift distributions for the Dark Energy Survey Year 3 (DES Y3) cosmological analyses. We cross-correlate the WL source galaxies (the 'unknown' sample u) with 'reference' samples (r) from both the DES Y3 *redMaGiC* catalogue (LRGs with secure photometric redshifts) and BOSS/eBOSS galaxies (with spec-z estimates). The reference samples are divided into thin redshift bins centred at $\{z_i\}$ to yield two-point angular cross-correlation measurements $w_{\rm ur}(z_i)$, for each combination of reference sample and WL tomographic bin, following now-standard practices for clustering-z (WZ) methods. The $w_{\rm ur}(z_i)$ measurements are weighted over angular separation to maximize the overall *S/N* ratio while avoiding the large angular scales used for cosmological measurements, in order to keep the clustering-z inferences statistically independent of the cosmological data vectors.

We describe two distinct methods to constrain the redshift distributions $n_{\rm u}(z)$ of the unknown samples using the $w_{\rm ur}(z_i)$ data. The 'mean-matching' method focuses on the mean $\langle z \rangle$ of the redshift distribution over a redshift window bounded by the redshift range of the reference sample and the 2σ extent of $n_{\rm u}(z)$. This method, similar to what was used in DES Y1 analyses (Davis et al. 2018;

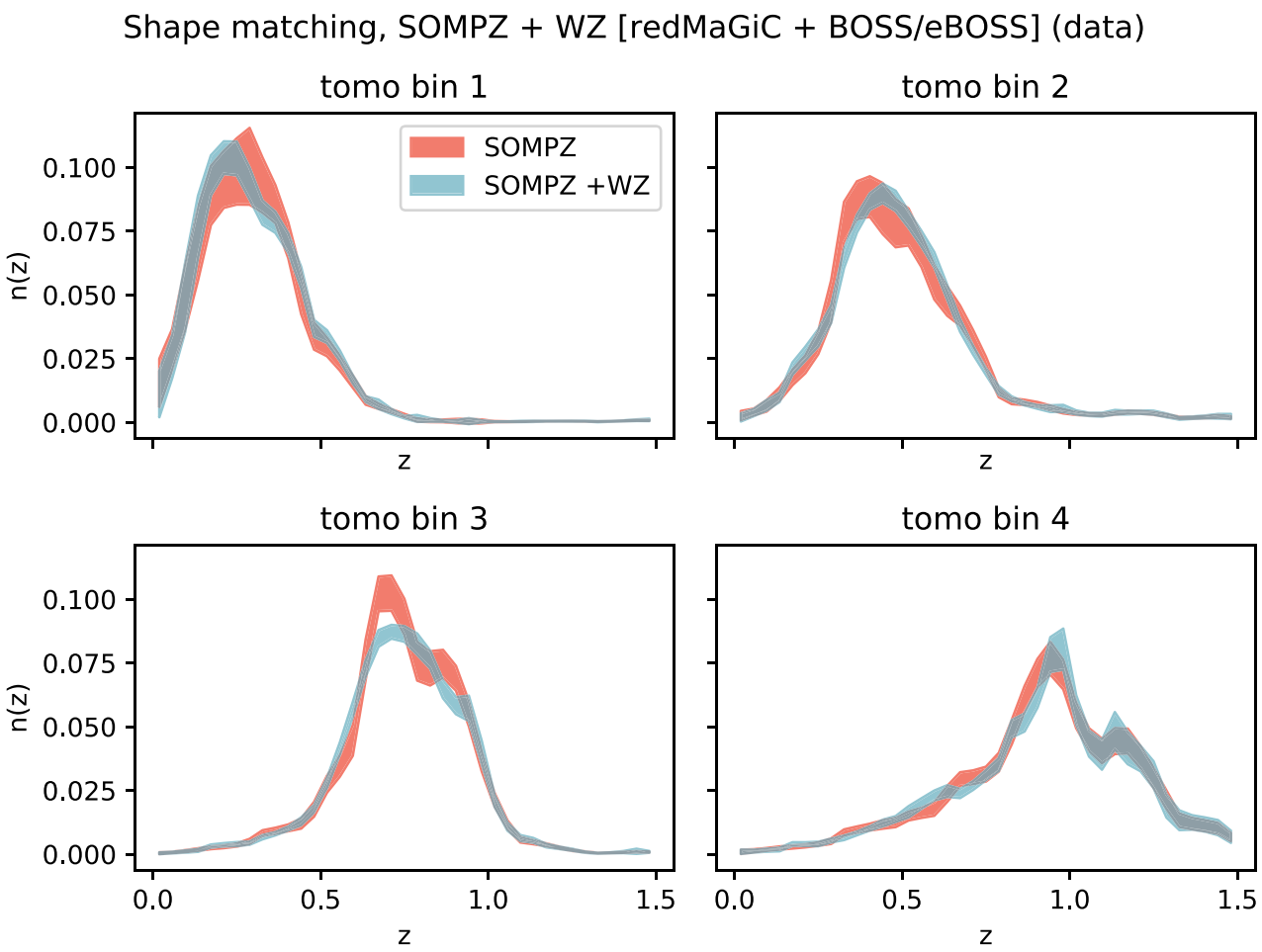

Figure 14. SOMPZ redshift distributions, as estimated in data, with and without clustering information (full-shape method). The bands encompass the statistical and systematic uncertainties of the distributions.

Table 4. Data. The mean redshift estimates of the SOMPZ distributions with and without clustering-z information.

Case	tomo bin 1	tomo bin 2	tomo bin 3	tomo bin 4
SOMPZ $\langle z \rangle$:	0.318 ± 0.009	0.513 ± 0.006	0.750 ± 0.005	0.942 ± 0.011
SOMPZ + WZ (mean-matching) :	0.317 ± 0.008	0.514 ± 0.006	0.750 ± 0.005	0.941 ± 0.011
SOMPZ + WZ (full-shape) :	0.321 ± 0.008	0.517 ± 0.006	0.749 ± 0.005	0.940 ± 0.010

Gatti et al. 2018), starts by computing the $\langle z \rangle$ of a naive clustering-z estimate $\tilde{n}_{\rm u}(z_i)$ (per equation 8) that assumes linear biasing with constant $b_{\rm u}$ and no magnification. From simulations, we estimate the additional uncertainty on $\langle z \rangle$ that arises from systematic errors in the naive estimator, which we conservatively take as 0.014 and are dominated by the unknown redshift dependence of $b_{\rm u}(z)$. Finally, we can compare this clustering-z estimate of $\langle z \rangle$ to that of the $n_{\rm u}(z)$ inferred from photo-z or some other independent method. For the DES Y3 data, we find the mean-matching method indicates full consistency between the SOMPZ photometric estimator and the clustering-z estimators, for all combinations of tomographic bin and reference sample.

The systematic errors we derive on $\langle z \rangle$ are larger than the statistical errors (estimated through jackknife) with DES Y3 data (Table 2), and also larger than the total errors estimated for the SOMPZ method (Fig. 13). Thus, this mean-matching approach has reached the limits of its usefulness, unless future experiments obtain narrower WL tomographic bins, and/or obtain external information on the relative bias of the unknown sample against the reference samples. Indeed the degeneracy between $n_{\rm u}(z)$ and $b_{\rm u}(z)$ in the observable $w_{\rm ur}(z)$ is the fundamental limitation of the clustering-z approach. This does not, however, mean that we have exhausted the information available from the clustering-z data in general. As discussed at the end of Section 4.1, the mean z is probably the summary statistic of $n_{\rm u}(z)$ that is *most* degraded by the dominant systematic error, redshift-dependent bias $b_{\rm u}(z)$, because this unmodelled multiplicative contribution to $w_{\rm ur}(z)$ is a smooth function of redshift. Higher order moments, or more generally the detailed shape of $n_{\rm u}(z)$, are less susceptible to clustering-z systematic errors, which are all expected to be smooth functions of z.

To extract this information, we apply the 'full-shape' method, developed in Section 4.2, using $w_{\rm ur}(z)$ data to inform $n_{\rm u}(z)$. We allow our model $\hat{w}_{\rm ur}(z)$ to incorporate an arbitrary multiplicative function

Sys(z, **s**) of redshift and nuisance parameters **s** that will mimic the effects of smooth systematic errors, such as $b_{\rm u}(z)$ and failure of the linear-bias model at small scales. Using simulations to choose this function and an appropriately flexible prior on **s**, we can now define a clustering-z likelihood for an arbitrary choice of $n_{\rm u}(z)$ that marginalizes over these nuisance parameters, as well as nuisances associated with lensing magnification signals that contaminate $w_{\rm ur}(z)$ (equation 13).

We note that this method improves on previous applications of clustering-z methods to WL cosmology (e.g. Choi et al. 2016; Davis et al. 2017; van den Busch et al. 2020), which mostly have used the former to constrain simple shifts of photo-z-derived redshift distributions, i.e. $n_{\rm u}(z) = n_{\rm PZ}(z + \Delta z)$. These approaches can lead to biased results if the shape of the photo-z posterior differs from the truth, or if clustering systematics are not taken into account by a sufficiently flexible model, as noted by Gatti et al. (2018) and van den Busch et al. (2020). We improve on these approaches for DES Y3 by defining likelihoods for $n_{\rm u}(z)$ arising from both SOMPZ and clustering-z methods, and using Hamiltonian Monte Carlo (HMC) to sample $n_{\rm u}(z)$ realizations from the product of these independent likelihoods. This also allows us to combine the information of the *redMaGiC* and BOSS/eBOSS references into a single inference. Note that each sample of the chain specifies redshift distributions for all four tomographic bins, capturing any inter-bin correlations that arise from the SOMPZ inference. This SOMPZ+WZ technique is extremely successful at reducing the point-by-point uncertainty in $n_{\rm PZ}(z)$ that arises from sample variance in the small surveys typically used to calibrate photo-z methods. The results for DES Y3 data can be seen in Fig. 14. The addition of the full-shape clustering-z information to SOMPZ yields $n_{\rm u}(z)$ samples that are much smoother and more realistic, taking advantage of the very high *S/N* that we have in $w_{\rm ur}(z)$ from the full footprint of DES Y3. This benefit is present despite the fact that the full-shape method does little to improve the SOMPZ's estimate of the mean redshift of each bin.

The final DES Y3 redshift calibration strategy includes a few additional minor tweaks to the SOMPZ+WZ samples, not addressed here. The $n(z)$ realizations are modified to account for uncertainties in the photometric calibration of the SOMPZ inputs, and the $z < 0.1$ behaviour (which is not constrained by clustering-z data) is smoothed to a physically reasonable form. These steps mostly affect the first tomographic bin (Myles et al. 2021). An additional correction to all the $n(z)$ realizations is performed to account for the effects of blending, based on the work on image simulations described in MacCrann et al. (2022). Then, ideally, the realizations are sampled over during the cosmological analysis, using the *hyperrank* technique (Cordero et al. 2021). In practice, however, in our fiducial cosmological run, we decided to parametrize the $n(z)$ uncertainties by shifts around their mean with a shift parameter Δz. This choice was dictated by efficiency reasons, and by the fact that we verified in Cordero et al. (2021) that marginalizing over the mean of the redshift distributions rather than sampling over the multiple $n(z)$ realizations was sufficient for the DES Y3 analysis. The prior on Δz is naturally provided by the scatter on the mean of the $n(z)$ realizations. Finally, when sampling the cosmological parameters, further constraints on the $n(z)$ are provided by the 'shear-ratio' test (Sánchez et al. 2021). The shear-ratio test uses small-scale galaxy–galaxy lensing measurements to further inform the shifts Δz. In practice, when running the cosmological analysis, the shear-ratio likelihood is simply multiplied by the cosmological likelihood, since the two are independent. Having combined these sources of information on $n(z)$, we find in DES Collaboration (2021) that its uncertainties are insignificant contributors to the Y3 cosmological uncertainty, despite these data having the smallest statistical uncertainties of any photometric cosmology survey to date.

The techniques used in this paper are applicable to other large imaging surveys. Further improvements in accuracy could be possible from having a reference sample that has spectroscopic redshifts like BOSS/eBOSS (eliminating one systematic error source) but large area and very high *S/N* like the DES Y3 *redMaGiC* sample. Improved prior knowledge of the magnification coefficients $\alpha_{\rm u}$, $\alpha_{\rm r}$ would also be of use. Mitigation schemes to reduce the impact of the bias evolution of the target sample could also be implemented (Matthews & Newman 2010; van den Busch et al. 2020). Importantly, the impact of bias evolution on clustering-z measures scales as $(\Delta z)^2$, where Δz is the rough width of each tomographic source bin, so improved binning accuracy from photo-z's will increase the value of clustering-zs. Ultimately the scheme of Sánchez & Bernstein (2019) and Alarcon et al. (2020), where one samples the posterior of the actual mass density field, individual source $z's$, and bias functions as constrained by the full catalogues, may offer stronger information than clustering-z methods that reduce the catalogues to the summary two-point statistics $w_{\rm ur}(z)$. But the methods applied to DES Y3 do make more complete use of the clustering-z data at summary-statistic level than has been done in the past.

ACKNOWLEDGEMENTS

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

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

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

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

The DES data management system is supported by the National Science Foundation under Grant Numbers AST-1138766 and AST-1536171. The DES participants from Spanish institutions are partially supported by MINECO under grants AYA2015-71825, ESP2015-66861, FPA2015-68048, SEV-2016-0588, SEV-2016-0597, and MDM-2015-0509, some of which include ERDF funds from the European Union. IFAE is partially funded by the CERCA program of the Generalitat de Catalunya. Research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013) including ERC grant agreements 240672, 291329, and 306478. We acknowledge support from the Brazilian Instituto Nacional de Ciência e Tecnologia (INCT) e-Universe (CNPq grant 465376/2014-2).

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

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

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

DATA AVAILABILITY

The full METACALIBRATION and *redMaGiC* catalogues will be made publicly available following publication, at the URL https://des.ncsa.illinois.edu/releases. The code used to perform the tests in this manuscript will be made available upon reasonable request to the authors.

REFERENCES

Ahumada R. et al., 2020, ApJS, 249, 3
Alam S. et al., 2021, MNRAS, 504, 4667
Alarcon A., Sánchez C., Bernstein G. M., Gaztañaga E., 2020, MNRAS, 498, 2614
Amon A. et al., 2021, preprint (arXiv:2105.13543)
Bates D. J., Tojeiro R., Newman J. A., Gonzalez-Perez V., Comparat J., Schneider D. P., Lima M., Streblyanska A., 2019, MNRAS, 486, 3059
Becker M. R., 2013, MNRAS, 435, 115
Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013, ApJ, 762, 109
Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, Phys. Rep., 367, 1
Blanton M. R. et al., 2005, AJ, 129, 2562
Blanton M. R. et al., 2017, AJ, 154, 28
Buchs R. et al., 2019, MNRAS, 489, 820
Cawthon R. et al., 2018, MNRAS, 481, 2427
Cawthon R. et al., 2020, preprint (arXiv:2012.12826)
Choi A. et al., 2016, MNRAS, 463, 3737
Conroy C., Wechsler R. H., Kravtsov A. V., 2006, ApJ, 647, 201
Cordero J. P. et al., 2021, preprint (arXiv:2109.09636)
Crocce M., Pueblas S., Scoccimarro R., 2006, MNRAS, 373, 369
Davis C. et al., 2017, preprint (arXiv:1710.02517)
Davis C. et al., 2018, MNRAS, 477, 2196
Davis M., Peebles P. J. E., 1983, ApJ, 267, 465
Dawson K. S. et al., 2013, AJ, 145, 10
Dawson K. S. et al., 2016, AJ, 151, 44
DeRose J. et al., 2019, preprint (arXiv:2105.13547)
DES Collaboration, 2021, preprint (arXiv:2105.13549)
Desjacques V., Jeong D., Schmidt F., 2018, Phys. Rep., 733, 1
Drlica-Wagner A. et al., 2018, ApJS, 235, 33
Eisenstein D. J. et al., 2011, AJ, 142, 72
Elvin-Poole J. et al., 2021, MNRAS, to be submitted
Everett S. et al., 2020, preprint (arXiv:2012.12825)
Flaugher B. et al., 2015, AJ, 150, 150
Gatti M. et al., 2018, MNRAS, 477, 1664
Gatti M. et al., 2021, MNRAS, 504, 4312
Gunn J. E. et al., 2006, AJ, 131, 2332
Hartley W. G. et al. 2022, MNRAS, 509, 3547
Hildebrandt H. et al., 2010, A&A, 523, A31
Hildebrandt H. et al., 2012, MNRAS, 421, 2355
Hildebrandt H. et al., 2017, MNRAS, 465, 1454
Hildebrandt H. et al., 2021, A&A, 647, A124
Hou J. et al. 2021 MNRAS, 500, 1201
Hoyle B. et al., 2018, MNRAS, 478, 592
Huterer D., Takada M., Bernstein G., Jain B., 2006, MNRAS, 366, 101
Johnson A. et al., 2017, MNRAS, 465, 4118
Kohonen T., 1982, Biol. Cybern., 43, 59
Krause E. et al., 2017, preprint (arXiv:1706.09359v1)
Landy S. D., Szalay A. S., 1993, ApJ, 412, 64
Lehmann B. V., Mao Y.-Y., Becker M. R., Skillman S. W., Wechsler R. H., 2017, ApJ, 834, 37
Leistedt B., Mortlock D. J., Peiris H. V., 2016, MNRAS, 460, 4258
Limber D. N., 1953, ApJ, 117, 134
MacCrann N. 2022, MNRAS, 509, 3371
Matthews D. J., Newman J. A., 2010, ApJ, 721, 456
McQuinn M., White M., 2013, MNRAS, 433, 2857
Ménard B., Scranton R., Schmidt S., Morrison C., Jeong D., Budavari T., Rahman M., 2013, preprint (arXiv:1303.4722v2)
Moessner R., Jain B., 1998, MNRAS, 294, L18
Morrison C. B., Hildebrandt H., Schmidt S. J., Baldry I. K., Bilicki M., Choi A., Erben T., Schneider P., 2017, MNRAS, 467, 3576
Myles J. et al., 2021, MNRAS, 505, 4249
Narayan R., 1989, ApJ, 339, L53
Newman J. A., 2008, ApJ, 684, 88
Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, MNRAS, 396, 19
Pandey S. et al., 2021, preprint (arXiv:2105.13545)
Percival W. J., Friedrich O., Sellentin E., Heavens A., 2021, preprint (arXiv:2108.10402)
Planck Collaboration VI, 2020, A&A, 641, A6
Porredon A. et al., 2021, Phys. Rev. D, 103, 043503

Porredon A. et al. 2021, preprint (arXiv:2105.13546)
Prakash A. et al., 2016, ApJS, 224, 34
Prat J. et al., 2020, MNRAS, preprint(arXiv:2105.13541)
Quenouille M. H., 1949, Proc. Camb. Phil. Soc., 45, 483
Rahman M., Ménard B., Scranton R., Schmidt S. J., Morrison C. B., 2015, MNRAS, 447, 3500
Rahman M., Mendez A. J., Ménard B., Scranton R., Schmidt S. J., Morrison C. B., Budavári T., 2016a, MNRAS, 457, 3912
Rahman M., Mendez A. J., Ménard B., Scranton R., Schmidt S. J., Morrison C. B., Budavári T., 2016b, MNRAS, 460, 163
Raichoor A. et al., 2017, MNRAS, 471, 3955
Reid B. et al., 2016, MNRAS, 455, 1553
Rodríguez-Monroy M. et al., 2020, MNRAS,preprint (arXiv:2105.13540)
Rozo E. et al., 2016, MNRAS, 461, 1431
Sánchez C. et al., 2014, MNRAS, 445, 1482
Sánchez C. et al., 2021, preprint (arXiv:2105.13542)
Sánchez C., Bernstein G. M., 2019, MNRAS, 483, 2801
Schmidt S. J., Ménard B., Scranton R., Morrison C., McBride C. K., 2013, MNRAS, 431, 3307
Scottez V. et al., 2016, MNRAS, 462, 1683
Scottez V., Benoit-Lévy A., Coupon J., Ilbert O., Mellier Y., 2017, MNRAS, 474, 3921
Secco L. F. et al., 2021, preprint (arXiv:2105.13544)
Sevilla-Noarbe I. et al., 2021, ApJS, 254, 24
Smee S. A. et al., 2013, AJ, 146, 32
Springel V., 2005, MNRAS, 364, 1105
Suchyta E. et al., 2016, MNRAS, 457, 786
Troxel M. A. et al., 2018, Phys. Rev. D, 98, 043528
van den Busch J. L. et al., 2020, A&A, 642, A200
Villumsen J. V., Freudling W., da Costa L. N., 1997, ApJ, 481, 578

APPENDIX A: MAGNIFICATION EFFECTS

We provide in this Appendix more details about the modelling of magnification effects $M(z_i)$ in the cross-correlation signal between the unknown and reference samples. Considering only the dominant terms (which account for the magnification of the unknown sample by the reference sample and the magnification of the reference sample by the unknown sample) and assuming linear bias, this can be written as:

$$M(z_i) = \int \mathrm{d}\theta\, W(\theta) \int \frac{\mathrm{d}l\, l}{2\pi} J_0(l\theta) \int \frac{\mathrm{d}\chi}{\chi^2} \times \left[b_\mathrm{r}\alpha_\mathrm{u} q_\delta^\mathrm{r} q_\kappa^\mathrm{u} + b_\mathrm{u}^{'}\alpha_\mathrm{r} q_\delta^\mathrm{u} q_\kappa^\mathrm{r} \right] P_\mathrm{NL}\left(\frac{l+1/2}{\chi}, z(\chi) \right), \quad \text{(A1)}$$

where the terms q_δ and q_κ read:

$$q_\delta(\chi) = n[z(\chi')] \frac{\mathrm{d}z}{\mathrm{d}\chi'}, \quad \text{(A2)}$$

$$q_\kappa(\chi) = \frac{3H_0^2\Omega_\mathrm{m}\chi}{c^2 a(\chi)} \int_\chi^{\chi(z=\infty)} \mathrm{d}\chi' n(z(\chi')) \frac{\mathrm{d}z}{\mathrm{d}\chi'} \frac{\chi'-\chi}{\chi'}. \quad \text{(A3)}$$

In the above equations, $n[z(\chi)]$ is either $n_\mathrm{u}(z)$ or $n_{\mathrm{r,\,i}}(z)$. Under the approximation of thin redshift bins, we can write equation (A1) as a discrete summation over redshift bins of width $\Delta\chi$:

$$M(z_i) = b_\mathrm{r}(z_i)\alpha_\mathrm{u}(z_i) \sum_{j>i} \left[D_{ij} n_\mathrm{u}(z_j) \right] + b_\mathrm{u}^{'}(z_i)\alpha_\mathrm{r}(z_i) \times \sum_{j>i} \left[D_{ij} n_\mathrm{u}(z_i) \right], \quad \text{(A4)}$$

with

$$D_{ij} = \frac{3H_0^2\Omega_\mathrm{m}}{c^2} w_\mathrm{DM}(z_i) \frac{\chi(z_i)}{a(z_i)} \frac{\chi(z_j)-\chi(z_i)}{\chi(z_j)} \Delta\chi_j. \quad \text{(A5)}$$

The magnification coefficient α, for an ideal flux-limited sample, can be related to the slope s of the cumulative number counts evaluated at flux limit: $\alpha \equiv 2.5s - 1$, with the slope formally defined as

$$s = \frac{\mathrm{d}}{\mathrm{d}m} \log_{10} n(<m), \quad \text{(A6)}$$

where $n(<m)$ is the cumulative number count as a function of magnitude m, and s is to be evaluated at the flux limit of the sample. For a sample which is not flux limited, evaluating the coefficient s is more complicated, and equation (A6) cannot be used. We use two different methods to estimate such coefficients for our samples, depending on whether we estimate them on data or on simulations (see below for further details). Estimates of α for both the reference and unknown samples are needed to properly model magnification effects.

When adopting the mean-matching method, magnifications effects are not modelled, but the tails of the distributions where magnification effects should be relevant are removed. We verify below that this method is efficient. On the other hand, in the full-shape method we do model magnification effects, according to equation (A4). In this latter case, while we absorb the contribution due to b_u to the clustering signal into the Sys function, we leave $b_\mathrm{u}^{'}$ as a free parameter in the magnification term. We also leave α_u as a free parameter, and marginalize over both parameters analytically when computing the likelihood. By doing so, we absorb uncertainties not only in these values but also in b_r, α_r, and in the linear-bias model adopted for magnification. Hence, formally, the $b_\mathrm{u}^{'}$ value appearing in the magnification is not assumed to equal the b_u that might multiply w_DM. We do not implement redshift dependence of $\mathbf{p} = \{b_\mathrm{u}, \alpha_\mathrm{u}\}$ (although the formalism would allow it) because magnification signals are important only over limited ranges of z (i.e. in the tails, see e.g. Gatti et al. 2018) for a given tomographic bin of the WL sources.

A1 Magnification coefficients estimates

In order to estimate the magnification coefficients of our samples, we adopt two different strategies. For the coefficients in data we use Balrog image simulations (Suchyta et al. 2016; Everett et al. 2020) in a process briefly described here. Galaxy profiles are drawn from the DES deep fields (Hartley et al. 2022) and injected into real DES images. The full photometry pipeline (Sevilla-Noarbe et al. 2021), the *redMaGiC*, and WL sample selection are applied to the new images to produce simulated *redMaGiC* and WL samples with the same selection effects as the real data. To compute the impact of magnification, the process is repeated, this time applying a constant magnification to each injected galaxy. The magnification coefficients are then derived from the fractional increase in number density when magnification is applied. This method captures the impact of magnification on both the galaxy magnitudes and the galaxy sizes, including all sample selection effects and potential observational and systematic effects. See Everett et al. (2020) and Elvin-Poole et al. (2021) for further details. The coefficients have been estimated for *redMaGiC* in five wide redshift bins, centred at z = (0.25, 0.425, 0.575, 0.75, 0.9), yielding the magnification coefficients $\boldsymbol{\alpha_\mathrm{r}}$ **= (0.3 ± 0.7, −1.5 ± 0.5, −0.7 ± 0.4, 1.2 ± 0.5, 1.0 ± 0.5).** The accuracy of these estimates is limited by the number of Balrog injections, which are scarce for a sample as bright as *redMaGiC*. Since the full-shape matching method formally requires values of the magnification coefficients for each of the 40 bins of the *redMaGiC* sample, we interpolate these values in z using the `scipy` routine `interp1d`. Although this procedure might not be too accurate

given the large uncertainties of the values of α_r, magnification effects are largely negligible, such that the interpolation details should not impact our main results. For the WL sample, using the same methodology, we infer $\alpha_u = (-0.4 \pm 0.2, -0.21 \pm 0.10, 0.00 \pm 0.10, 0.31 \pm 0.07)$, for the four tomographic bins, respectively. Note that these values do not have to be interpolated.

For the values estimated for the samples in simulations we adopt a different strategy. In particular, we use the estimated convergence, κ, computed at the location of each galaxy, to apply a small magnification to the galaxy magnitudes (Δm), and then select our samples with and without this Δm applied and compute the fractional change of objects passing the selection $\Delta N/N$ in 10 equally spaced κ bins. The gradient of this relation is then related to the magnification coefficient (Elvin-Poole et al. 2021). This method only captures the effect of magnification on the galaxy fluxes, as it is the only effect expected in simulations. We estimate $\alpha_r = (0.2 \pm 0.4, 0.05 \pm 0.15, 0.00 \pm 0.08, 1.11 \pm 0.12, 1.18 \pm 0.06)$ for the *redMaGiC* sample, and $\alpha_u = (-0.365 \pm 0.002, -0.655 \pm 0.002, -0.447 \pm 0.002, 0.836 \pm 0.002)$ for the WL sample.

Lastly, we note that estimates of the magnification coefficients are not available for BOSS/eBOSS galaxies, as we did not try to reproduce the complex BOSS/eBOSS selection function within Balrog image simulations. We also did not estimate these coefficients for the simulated BOSS/eBOSS sample. This is not a problem, as we verify below that BOSS/eBOSS does not have the sensitivity to measure magnification effects. When formally needed (for the full-shape method), though, we adopted the same coefficients as the *redMaGiC* sample.

A2 Magnification impact on the clustering measurements

We show the impact of magnification effects on the clustering signal in Fig. A1. The figure has been produced assuming the expected clustering signal for the redshift distributions adopted in the Buzzard simulations, and for the magnification coefficients as computed in simulations.The magnification contributions are barely detectable: only for the first two tomographic bins, at high redshift, the deviations from the 'clustering only' signal due to magnification are slightly larger than the 68 per cent confidence interval of the *redMaGiC* measurement.

Concerning the mean-matching method, it can be noted how the contribution is always smaller than the 68 per cent confidence interval of the measurement when confined within the 2σ interval of the mean matching method. The impact of magnification effects on the windowed mean when using the coefficients estimated for the simulations is at most $\Delta\langle z\rangle \approx 0.002$. Thus, it is of the same order of magnitude of the statistical uncertainty of the measurement. To be more conservative, for the mean matching approach we estimated the impact on $\Delta\langle z\rangle$ if the data had different (and potentially larger) values of α_r and α_u than the ones estimated in simulations. We computed the magnification term $M(\theta)$ assuming Gaussian priors $\alpha_r \sim \mathcal{N}(0, 2)$ and $\alpha_u \sim \mathcal{N}(0, 2)$, and measured the resultant scatter in $\Delta\langle z\rangle$. These priors are rather wide, but even with these broad priors, magnification is a negligible component of our final error model. Indeed, we obtained an RMS scatter on this metric of $\Delta\langle z\rangle_{\rm RMS} = (0.004, 0.005, 0.003, 0.004)$ for the four tomographic bins. We note that these values are up to a factor 10 smaller than what we would have obtained by including the tails of the redshift distributions, justifying the 2σ cut. These values, in the second row of Table 2, are taken as the magnification contribution to σ_z.

As for the full-shape method, magnification effects are modelled over the full range of redshift, using as input the estimated magnification coefficients. Nevertheless, their impact is strongly reduced by the combination with the SOMPZ likelihood, which enforces the tails of the redshift distributions to have a small amplitude. To roughly asses the impact of the exact values of the magnification coefficients α_r and α_u, we performed the following test, both in simulations and on data: we verified that assuming values for α_r or α_u that are $-1\times$ the fiducial ones resulted in shifts $\Delta\langle z\rangle < 10^{-3}$. This highlights the importance of combining SOMPZ and clustering information to achieve a more robust estimator of the redshift distributions.

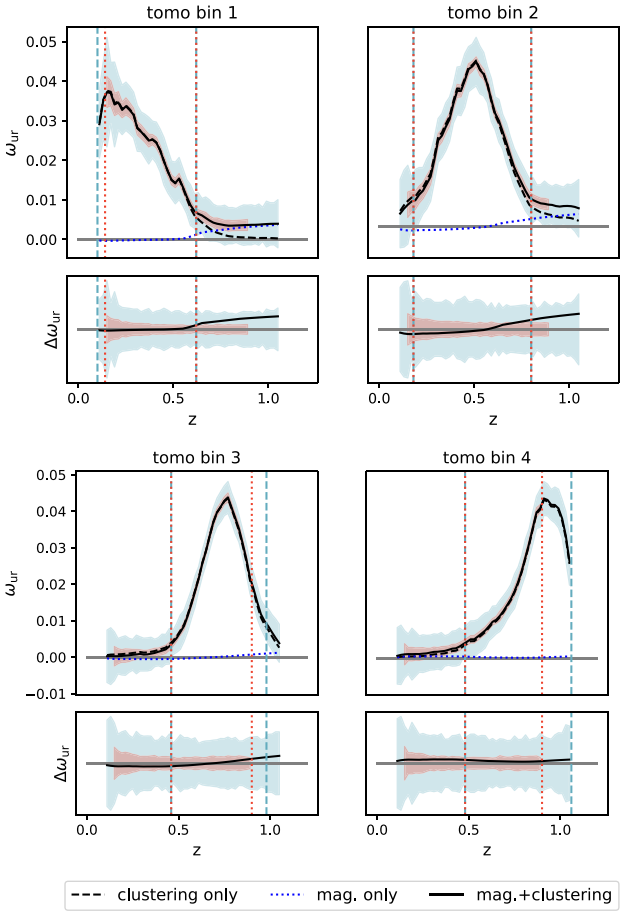

Figure A1. Simulated clustering amplitude for the four tomographic bins. The blue and red coloured bands encompass the 68 confidence interval of the clustering measurement, obtained using BOSS/eBOSS and *redMaGiC* galaxies as a reference, respectively. The dashed, dotted, and solid lines represent the simulated clustering only signal, magnification only signal, and clustering with magnification signal, respectively. The vertical dashed lines indicate the 2σ interval used in the mean matching approach. The four small panels show the deviations Δw_{ur} from the clustering only signal when magnification effects are included. Magnification effects are estimated using the values for the magnification coefficients as estimated in simulation.

APPENDIX B: FULL $\hat{w}_{ur}$ MODEL AND ANALYTICAL MARGINALIZATION

We provide here more details about the implementation of the full-shape method. The method assigns a likelihood (equation 13) of the observed $w_{ur}(z_i)$ given a proposal for the redshift distributions $\{n_u(z_i)\}$ along with a set of other relevant parameters. The likelihood uses the model in equation (12). We will assume that the values of the dark-matter correlation $w_{DM}(z_i)$, the reference-sample properties

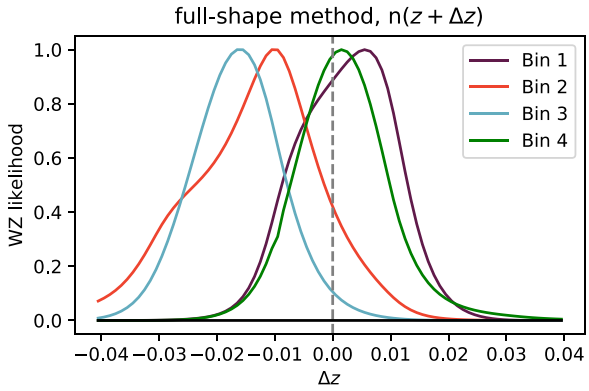

Figure B1. Likelihood of the shift of the mean of the redshift distributions obtained using the full shape method in simulations, and using true redshift distributions (but shifted around their mean) as proposals distributions.

$b_{\rm r}(z)$ and $\alpha_{\rm r}(z)$, and the magnification coefficients D_{ij} are provided along with $n_{\rm u}(z_i)$. We will consider as nuisance parameters the properties of the unknown population, namely the $\alpha_{\rm u}(z)$ and $b_{\rm u}(z)$ used in magnification terms; plus any parameters $\mathbf{s}$ of the Sys(z) function that allows for systematic errors.

We will assume here that $\alpha_{\rm u}$ and $b_{\rm u}$ are independent of redshift, though in principle a more general function, linear in some parameters, can be used without altering any of the methods herein. We note that we did not multiply the magnification terms by the systematic function: despite the fact that the magnification terms are not immune to systematic errors, we assumed that it was not necessary to further modelling those, as the $\alpha_{\rm r}$, $\alpha_{\rm u}$, and $\bar{b}_{\rm u}$ parameters provide enough flexibility to the model and the magnification signal is much smaller than clustering to start with. We also note that the $b_{\rm u}$ parameter is used only in the magnification term, and hence can be independent of the bias for clustering that is absorbed into the Sys(z) function. This allows for the systematic errors in the magnification term to differ from those in the clustering term.

The systematic-error function for clustering is given in the exponentiated polynomial form in equations (17) and (18). Tuning the order M allows us to adjust the smoothness of the function, and exponentiation allows us to draw the coefficients $\mathbf{s}$ from 1D Gaussian priors while maintaining positive Sys(z). Adjusting the σ_s values of these priors tunes the RMS of the systematic variations, in a way made predictable by the orthogonality of the Legendre polynomials. We wish for independent, uniform Gaussian priors on the s_i to propagate into RMS variation of $\log {\rm Sys}(z)$ that is approximately independent of z over $[z_{\rm min}, z_{\rm max}]$. The Legendre polynomials have this property over most of their nominal domain $u \in [-1, 1]$, but not near the edges of this range. For this reason we map $[z_{\rm min}, z_{\rm max}] \rightarrow [-0.85, 0.85]$, as indicated by equation (18).

Equation (13) requires us to marginalize over the nuisance-parameter vector $\mathbf{q} = \{\mathbf{p}\,\mathbf{s}\}$ (with $\mathbf{p} = b_{\rm u}, \alpha_{\rm u}$). Doing so as part of a Markov chain would be unwieldy, as we would have to introduce eight free parameters for each of the four tomographic bins times 2 reference samples. It is far better to execute the marginalization on the fly during sampling if possible. The log-likelihood is not quite quadratic in $\mathbf{q}$ – the exponentiation of the polynomial in Sys(z; $\mathbf{s}$) makes the model $\hat{w}_{\rm ur}$ non-linear in $\mathbf{s}$. We opt to linearize the model about its maximum $\mathbf{s}_0 = \{s_{k,0}\}$:

$$
\begin{aligned}
{\rm Sys}\,(z_i\,,\mathbf{s}) &\approx {\rm Sys}\,(z_i\,,\mathbf{s}_0) \\
&\times \left[1 + \sum_{k=0}^{M} \frac{\sqrt{2k+1}}{0.85} P_k(u) s_{k,0}\left(s_k - s_{k,0}\right)\right]. \qquad \text{(B1)}
\end{aligned}
$$

The deviation of the data from the model can then be rewritten in linear form, with $\mathbf{w}_{\rm ur}$ being a vector over redshifts, as

$$
\mathbf{w}_{\rm ur} - \hat{\mathbf{w}}_{\rm ur} = \mathbf{c}(\mathbf{q}_0) - A\mathbf{q}, \qquad \text{(B2)}
$$

where $\mathbf{c}$ is a vector independent of $\mathbf{q}$ and A is a matrix composed of the linear terms in equation (B1) and elements of the magnification terms.

If we assume the nuisance parameters we want to marginalize over to have a Gaussian prior $\mathbf{q} \sim \mathcal{N}(\mu_q, \Sigma_q)$, we can write the full likelihood as follows:

$$
\begin{aligned}
\mathcal{L}_{\rm WZ} &\approx |2\pi\,\Sigma_{\rm wz}|^{-1/2} |2\pi\,\Sigma_p|^{-1/2} \\
&\times \int {\rm d}\mathbf{q}\, \exp\left[-\frac{1}{2}(\mathbf{c} - A\mathbf{q})^T \hat{\Sigma}_{\rm wz}^{-1} (\mathbf{c} - A\mathbf{q})\right] \\
&\times \exp\left[-\frac{1}{2}(\mathbf{q} - \mu_q)^T \hat{\Sigma}_q^{-1} (\mathbf{q} - \mu_q)\right]. \qquad \text{(B3)}
\end{aligned}
$$

This is a Gaussian integral that can be reduced to linear algebra.

In summary, the algorithm for the marginalization in equation (13) is as follows:

(i) Find the values $\mathbf{q}_0$ which maximize the integrand. This is done using Newton iterations.
(ii) Evaluate the vector $\mathbf{c}$ and matrix A at this value of $\mathbf{q}_0$.
(iii) Substitute these and the known $\Sigma_{\rm wz}$, μ_q, and Σ_q into the analytic result for the Gaussian integral above.

Although this marginalization is approximate, it does not actually need to be exact, because the chosen functional form for Sys(z, $\mathbf{s}$) is somewhat arbitrary. All that is necessary is that the algorithm yields a likelihood $\mathcal{L}$ of the clustering-z data given a proposed $n_{\rm u}(z)$ that decreases in a meaningful and robust way as the data move away from the naive linear model. We prove that the full-shape method recovers the true $n(z)$ within uncertainties in Section 4.2.2, assuming the SOMPZ realizations as $n_{\rm u}(z)$ proposals. Here, we also show the result of a simpler test, performed in simulations, where the $n_{\rm u}(z)$ proposals are simply taken to be true redshift distributions shifted around their mean. This is a useful test because it is shows that the methodology is unbiased independently of the SOMPZ information. We use equation (B3) to assign each true $n_{\rm u}(z)$s (shifted around their mean) a weight, using the clustering measurement and the magnification coefficients from the simulations. The key result is then the likelihood of the shifts Δz, which has to be statistically compatible with 0. This is shown in Fig. B1; in particular, we obtain $\Delta z = 0.002 \pm 0.008$, -0.013 ± 0.011, -0.016 ± 0.008, 0.002 ± 0.008 for the four tomographic bins, which indicates statistical compatibility with the truth. The models are a good fit to the data, with $\chi^2 = 1.29, 0.67, 0.72, 0.63$ for the *redMaGiC* sample, and $\chi^2 = 1.19, 1.20, 0.58, 0.88$ for the BOSS/eBOSS sample.

[1] *Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, E-08193 Bellaterra (Barcelona), Spain*
[2] *Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*

[3]*Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439, USA*
[4]*Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA*
[5]*Kavli Institute for Particle Astrophysics & Cosmology, PO Box 2450, Stanford University, Stanford, CA 94305, USA*
[6]*Physics Department, 2320 Chamberlin Hall, University of Wisconsin-Madison, 1150 University Avenue Madison, WI 53706-1390, USA*
[7]*Department of Physics, Duke University Durham, NC 27708, USA*
[8]*Department of Astronomy, University of California, Berkeley, 501 Campbell Hall, Berkeley, CA 94720, USA*
[9]*Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA*
[10]*Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA*
[11]*SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA*
[12]*Department of Physics, The Ohio State University, Columbus, OH 43210, USA*
[13]*Jodrell Bank Center for Astrophysics, School of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK*
[14]*Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA*
[15]*Fermi National Accelerator Laboratory, PO Box 500, Batavia, IL 60510, USA*
[16]*Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain*
[17]*Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain*
[18]*Department of Physics, University of Arizona, Tucson, AZ 85721, USA*
[19]*Cerro Tololo Inter-American Observatory, NSF's National Optical-Infrared Astronomy Research Laboratory, Casilla 603, La Serena 1700000, Chile*
[20]*Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, CP 66318, São Paulo, SP 05314-970, Brazil*
[21]*Laboratório Interinstitucional de e-Astronomia - LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ 20921-400, Brazil*
[22]*Instituto de Fisica Teorica UAM/CSIC, Universidad Autonoma de Madrid, E-28049 Madrid, Spain*
[23]*Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK*
[24]*CNRS, UMR 7095, Institut d'Astrophysique de Paris, F-75014 Paris, France*
[25]*Institut d'Astrophysique de Paris, Sorbonne Universités, UPMC Univ Paris 06, UMR 7095, F-75014 Paris, France*
[26]*Department of Physics & Astronomy, University College London, Gower Street, London WC1E 6BT, UK*
[27]*Instituto de Astrofisica de Canarias, E-38205 La Laguna, Tenerife, Spain*
[28]*Dpto. Astrofisica, Universidad de La Laguna, E-38206 La Laguna, Tenerife, Spain*
[29]*Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801, USA*
[30]*National Center for Supercomputing Applications, 1205 West Clark St, Urbana, IL 61801, USA*
[31]*School of Physics and Astronomy, University of Nottingham, Nottingham NG7 2RD, UK*
[32]*INAF-Osservatorio Astronomico di Trieste, via G. B. Tiepolo 11, I-34143 Trieste, Italy*
[33]*Institute for Fundamental Physics of the Universe, Via Beirut 2, I-34014 Trieste, Italy*
[34]*Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil*
[35]*Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA*
[36]*Department of Physics and Astronomy, University of Utah, 115 S. 1400 E., Salt Lake City, UT 84112, USA*
[37]*Department of Physics, IIT Hyderabad, Kandi, Telangana 502285, India*
[38]*Department of Astronomy/Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721-0065, USA*
[39]*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109, USA*
[40]*Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA*
[41]*Institute of Theoretical Astrophysics, University of Oslo. PO Box 1029 Blindern, NO-0315 Oslo, Norway*
[42]*IKavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA*
[43]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[44]*Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[45]*School of Mathematics and Physics, University of Queensland, Brisbane, QLD 4072, Australia*
[46]*Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr 1, D-81679 Munich, Germany*
[47]*Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, D-85748 Garching, Germany*
[48]*Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstr 1, D-81679 München, Germany*
[49]*Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA*
[50]*Australian Astronomical Optics, Macquarie University, North Ryde, NSW 2113, Australia*
[51]*Lowell Observatory, 1400 Mars Hill Rd, Flagstaff, AZ 86001, USA*
[52]*Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, UK*
[53]*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*
[54]*Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA*
[55]*Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain*
[56]*Waterloo Centre for Astrophysics, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada*
[57]*Department of Physics and Astronomy, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada*
[58]*Perimeter Institute for Theoretical Physics, 31 Caroline St North, Waterloo, ON N2L 2Y5, Canada*
[59]*Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), E-28040 Madrid, Spain*
[60]*Department of Physics and Astronomy, Sejong University, Seoul 143-747, Korea*
[61]*Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15312, USA*
[62]*School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK*
[63]*Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA*
[64]*Department of Physics and Astronomy, Pevensey Building, University of Sussex, Brighton BN1 9QH, UK*

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