Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **477,** 1664–1682 (2018)
Advance Access publication 2018 February 22

doi:10.1093/mnras/sty466

Dark Energy Survey Year 1 results: cross-correlation redshifts – methods and systematics characterization

M. Gatti,[1⋆] P. Vielzeuf,[1⋆] C. Davis,[2] R. Cawthon,[3] M. M. Rau,[4] J. DeRose,[2,5] J. De Vicente,[6] A. Alarcon,[7] E. Rozo,[8] E. Gaztanaga,[7] B. Hoyle,[4] R. Miquel,[1,9] G. M. Bernstein,[10] C. Bonnett,[1] A. Carnero Rosell,[11,12] F. J. Castander,[7] C. Chang,[3] L. N. da Costa,[11,12] D. Gruen,[2,13] J. Gschwend,[11,12] W. G. Hartley,[14,15] H. Lin,[16] N. MacCrann,[17,18] M. A. G. Maia,[11,12] R. L. C. Ogando,[11,12] A. Roodman,[2,13] I. Sevilla-Noarbe,[6] M. A. Troxel,[17,18] R. H. Wechsler,[2,5,13] J. Asorey,[19,20] T. M. Davis,[19,20] K. Glazebrook,[21] S. R. Hinton,[20] G. Lewis,[19,22] C. Lidman,[19,23] E. Macaulay,[20] A. Möller,[19,24] C. R. O'Neill,[19] N. E. Sommer,[19,24] S. A. Uddin,[19,25] F. Yuan,[19,24] B. Zhang,[19,24] T. M. C. Abbott,[26] S. Allam,[16] J. Annis,[16] K. Bechtol,[27] D. Brooks,[14] D. L. Burke,[2,13] D. Carollo,[19,28] M. Carrasco Kind,[29,30] J. Carretero,[1] C. E. Cunha,[2] C. B. D'Andrea,[10] D. L. DePoy,[31] S. Desai,[32] T. F. Eifler,[33,34] A. E. Evrard,[35,36] B. Flaugher,[16] P. Fosalba,[7] J. Frieman,[3,16] J. García-Bellido,[37] D. W. Gerdes,[35,36] D. A. Goldstein,[38,39] R. A. Gruendl,[29,30] G. Gutierrez,[16] K. Honscheid,[17,18] J. K. Hoormann,[20] B. Jain,[10] D. J. James,[40] M. Jarvis,[10] T. Jeltema,[41] M. W. G. Johnson,[28] M. D. Johnson,[28] E. Krause,[2] K. Kuehn,[23] S. Kuhlmann,[42] N. Kuropatkin,[16] T. S. Li,[16] M. Lima,[11,43] J. L. Marshall,[31] P. Melchior,[44] F. Menanteau,[28,29] R. C. Nichol,[45] B. Nord,[16] A. A. Plazas,[34] K. Reil,[13] E. S. Rykoff,[2,13] M. Sako,[10] E. Sanchez,[6] V. Scarpine,[16] M. Schubnell,[36] E. Sheldon,[46] M. Smith,[47] R. C. Smith,[26] M. Soares-Santos,[16] F. Sobreira,[11,48] E. Suchyta,[49] M. E. C. Swanson,[30] G. Tarle,[36] D. Thomas,[45] B. E. Tucker,[19,24] D. L. Tucker,[16] V. Vikram,[42] A. R. Walker,[26] J. Weller,[4,50,51] W. Wester[16] and R. C. Wolf[10]

Affiliations are listed at the end of the paper

Accepted 2018 January 29. Received 2018 January 29; in original form 2017 September 5

ABSTRACT

We use numerical simulations to characterize the performance of a clustering-based method to calibrate photometric redshift biases. In particular, we cross-correlate the weak lensing source galaxies from the Dark Energy Survey Year 1 sample with *redMaGiC* galaxies (luminous red galaxies with secure photometric redshifts) to estimate the redshift distribution of the former sample. The recovered redshift distributions are used to calibrate the photometric redshift bias of standard photo-z methods applied to the same source galaxy sample. We apply the method to *two* photo-z codes run in our simulated data: Bayesian Photometric Redshift and Directional Neighbourhood Fitting. We characterize the systematic uncertainties of our

⋆E-mail: mgatti@ifae.es (MG); pvielzeuf@ifae.es (PV)

calibration procedure, and find that these systematic uncertainties dominate our error budget. The dominant systematics are due to our assumption of unevolving bias and clustering across each redshift bin, and to differences between the shapes of the redshift distributions derived by clustering versus photo-zs. The systematic uncertainty in the mean redshift bias of the source galaxy sample is $\Delta z \lesssim 0.02$, though the precise value depends on the redshift bin under consideration. We discuss possible ways to mitigate the impact of our dominant systematics in future analyses.

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

1 INTRODUCTION

Current and future large photometric galaxy surveys like the Dark Energy Survey (DES)[1] (The Dark Energy Survey Collaboration 2005), Kilo-Degree Survey (KiDS)[2] (de Jong et al. 2015), Hyper Suprime-Cam (HSC)[3] (Aihara et al. 2018), Large Synoptic Survey Telescope (LSST)[4] (LSST Science Collaboration 2009), *Euclid*[5] (Laureijs et al. 2011), and *Wide Field Infrared Survey Telescope* (*WFIRST*)[6] (Spergel et al. 2013) will map large volumes of the Universe, measuring the angular positions and shapes of hundreds of millions (or billions) of galaxies. This will allow cosmological measurements with an unprecedented level of precision, leading to a considerable step forward in our understanding of cosmology and particularly of the nature of dark energy. To capitalize on their statistical constraining power, these surveys require accurate characterization of the redshift distributions of selected galaxies, which presents a considerable challenge in the absence of complete spectroscopic coverage.

Given the large amount of forthcoming photometric data, obtaining a spectroscopic redshift for every individual source is unfeasible: spectroscopy of large samples is time-consuming and expensive, and it is usually restricted to the brightest objects of any given sample. Because of this limitation, photometric surveys provide redshift estimates for each galaxy based on that galaxy's multiband photometry, a technique called photometric redshift, or photo-z. There exists a large variety of photo-z methods (e.g. Hildebrandt et al. 2010; Sánchez et al. 2014). However, unrealistic spectral energy distribution (SED) templates, degeneracies between colours and redshift, and unrepresentative spectroscopic samples for both training and calibration ultimately limit the performance of photo-z methods (Lima et al. 2008; Cunha et al. 2009; Newman et al. 2015; Bezanson et al. 2016; Masters et al. 2017).

Clustering-based redshift estimation methods (Newman 2008; Matthews & Newman 2010; Ménard et al. 2013; Schmidt et al. 2013) constitute an interesting alternative to infer redshift distributions, since they are more general and do not suffer the above limitations. Briefly, one uses the fact that the correlation amplitude between a sample with unknown redshifts and a reference sample with known redshifts in some narrow redshift bin can be related to the fraction of galaxies in the unknown sample that lie within the redshift range of the reference sample.

[1] https://www.darkenergysurvey.org/
[2] http://kids.strw.leidenuniv.nl/
[3] https://subarutelescope.org/Projects/HSC/
[4] https://www.lsst.org/
[5] http://sci.esa.int/euclid/
[6] https://wfirst.gsfc.nasa.gov/

Clustering-based redshift estimators have been studied and applied both to simulations and to data (e.g. Schneider et al. 2006; McQuinn & White 2013; Ménard et al. 2013; Schmidt et al. 2013; Scottez et al. 2016, 2018; Hildebrandt et al. 2017; Davis et al. 2018; van Daalen & White 2018). Hildebrandt et al. (2017) cross-correlated the source galaxies used in the KiDS cosmological analysis with galaxies from zCOSMOS (Lilly et al. 2009) and DEEP2 (Newman et al. 2013). Unfortunately, the small ($\leq$1 deg^2) area covered by these reference galaxy samples severely limited the usefulness of the resulting cross-correlation analyses. Consequently, Hildebrandt et al. (2017) ultimately chose to rely on traditional photo-z methods in deriving the KiDS cosmological constraints. Small area effects were partially mitigated in a more recent work by Morrison et al. (2017), who recovered KIDS redshift distributions relying on a larger overlap ($\sim$170 deg^2) with the spectroscopic samples from Sloan Digital Sky Survey (SDSS; Alam et al. 2015) and Galaxy And Mass Assembly (GAMA; Liske et al. 2015).

The Dark Energy Survey Year 1 (DES Y1) cosmological analyses rely on a different strategy. Instead of using a spectroscopic sample as reference, we use red-sequence galaxies from the DES Y1 redSequence Matched-filter Galaxy Catalog (*redMaGiC*; Rozo et al. 2016). The *redMaGiC* algorithm is designed to select galaxies with high-quality photometric redshift estimates. While the reliance on *redMaGiC* photometric redshifts may be a source of concern for the cross-correlation program, the vastly superior statistical power of the sample renders the resulting cross-correlation constraints competitive with traditional photo-z methods.

The DES Y1 analysis attempts to combine traditional photo-z methods with cross-correlation techniques. In particular, motivated by the fact that the DES Y1 cosmological analyses are primarily sensitive to an overall redshift bias in the photometric redshift estimates (DES Collaboration 2017; Hoyle et al. 2017; Troxel et al. 2017), we have sought to use cross-correlation methods to verify and calibrate the redshift bias of traditional photo-z methods. By combining these two techniques we benefit from the strength of both methods, while ameliorating their respective weaknesses. This calibration strategy is fully implemented in the DES Y1 cosmic shear and combined two-point function analysis (DES Collaboration 2017; Krause et al. 2017; Troxel et al. 2017).

This paper characterizes the performance and systematic uncertainties of our method for calibrating photometric redshift biases in the DES Y1 source galaxy sample via cross-correlation with *redMaGiC* galaxies. Specifically, we implement our method on simulated data, introducing sources of systematic uncertainty one at a time to arrive at a quantitative characterization of the reliability and accuracy of our method. A companion paper (Davis et al. 2017) implements the photometric calibration method developed here to enable DES Y1 cosmology analyses, while a second companion paper (Cawthon et al. 2017) uses cross-correlations to validate the

photometric redshift performance of the *redMaGiC* galaxies themselves.

The paper is organized as follows. In Section 2 we present the methodology we use to calibrate photo-z posteriors using clustering-based redshift estimation. The simulations and the samples used are described in Section 3. Section 4 is devoted to the study and quantification of the systematic error of our method. In Section 5 we further discuss some aspects of clustering-based redshift estimation techniques and how our method could be improved upon in the future. Finally, we present our conclusions in Section 6.

2 METHODOLOGY

In DES Y1 we will use clustering-based redshift estimates to correct the photo-z posterior distributions of a given science sample. We defer the description (and the binning) of the particular samples (unknown and reference) adopted in this work to Section 3, while keeping the description of the methodology as general as possible. Here, 'unknown' always refers to the photometric galaxy sample for which we wish to calibrate photometric redshift biases, while 'reference' refers to the galaxy sample with known, highly accurate redshifts (be they spectroscopic or photometric).

Our methodology divides into two steps.

(i) We first estimate the redshift distribution of the unknown galaxy sample by cross-correlating with the reference sample. Note that the reference sample does not necessarily have to span the full redshift interval of the unknown sample.

(ii) We then use the recovered redshift distribution to calibrate the redshift bias of the unknown galaxy population by finding the shift Δz that brings the photo-z posterior in better agreement with the redshift distribution obtained through cross-correlations.

2.1 First step: clustering-based redshift estimates

In the literature a variety of methods to recover redshift distributions based on cross-correlation have been discussed (Newman 2008; McQuinn & White 2013; Ménard et al. 2013; Schmidt et al. 2013). The underlying idea shared by all methods is that the spatial cross-correlation between two samples of objects is non-zero only in the case of 3D overlap. Let us now consider two galaxy samples.

(i) An *unknown* sample, whose redshift distribution $n_{\rm u}(z)$ has to be recovered.

(ii) A *reference* sample, whose redshift distribution $n_{\rm r}(z)$ is known (either from spectroscopic redshifts or from high-precision photometric redshifts). The reference sample is divided into narrow redshift bins.

To calibrate the redshift distribution of the unknown sample we bin the reference sample in narrow redshift bins, and then compute the angular cross-correlation signal $w_{\rm ur}$ between the unknown sample and each of these reference redshift bins. Under the assumption of linear biasing, we find

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

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 biases of the two samples, and $w_{\rm DM}(\theta, z')$ is the dark matter two-point correlation function.

In this paper we implement three different clustering-based methods: Schmidt/Ménard's method (Ménard et al. 2013; Schmidt et al. 2013), a 'weighted' method, and Newman's method as explained in Matthews & Newman (2010). We briefly describe each of the three methods. A comparison of the three methods is presented in Section 4.6. At the end, we have opted for using Schmidt/Ménard's method for our fiducial analysis.

Schmidt/Ménard's method. The implementation of the method is discussed in details in Schmidt et al. (2013). The authors use a 'one-angular bin' estimate of the cross-correlation signal. This is achieved by computing the number of sources of the unknown sample in a physical annulus around each individual object of the reference sample, from a minimum comoving distance $r_{\rm min}$ to a maximum distance $r_{\rm max}$. Our fiducial choice for the scales is from 500 to 1500 kpc.[7] In addition, each object of the unknown sample is weighted by the inverse of the distance from the reference object, which has been shown to increase the signal-to-noise ratio (S/N) of the measurement (Schmidt et al. 2013). We use the Davis & Peebles (1983) estimator of the cross-correlation signal:

$$\bar{w}_{\rm ur} = \frac{N_{\rm Rr}}{N_{\rm Dr}} \frac{\int_{r_{\rm min}}^{r_{\rm max}} {\rm d}r' W(r') \left[D_{\rm u} D_{\rm r}(r')\right]}{\int_{r_{\rm min}}^{r_{\rm max}} {\rm d}r' W(r') \left[D_{\rm u} R_{\rm r}(r')\right]} - 1, \tag{2}$$

where $D_{\rm u}D_{\rm r}(r')$ and $D_{\rm u}R_{\rm r}(r')$ are, respectively, data–data and data–random pairs, and $W(r') \propto 1/r'$ the weight function. 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.

The Davis & Peebles (1983) estimator is less immune to window function contamination than the Landy & Szalay (1993) estimator, since it involves using a catalogue of random points for just one of the two samples. We choose to use the Davis & Peebles estimator so as to avoid creating high-fidelity random catalogues for the DES Y1 source galaxy sample (our unknown sample): the selection function depends on local seeing and imaging depth, resulting in a complex spatial selection function. We therefore decide to use a catalogue of random points only for the reference sample, whose selection function and mask are well understood (Elvin-Poole et al. 2017).

Assuming the reference sample is divided into sufficiently narrow bins centred at z, we can approximate $n_{\rm r}(z') \propto N_{\rm r}\delta_{\rm D}(z - z')$ (with $\delta_{\rm D}$ being Dirac's delta distribution, and $N_{\rm r}$ being the number of galaxies in the reference bin) and invert equation (1) to obtain the redshift distribution of the unknown sample:

$$n_{\rm u}(z) \propto \bar{w}_{\rm ur}(z) \frac{1}{b_{\rm u}(z)} \frac{1}{b_{\rm r}(z)} \frac{1}{\bar{w}_{\rm DM}(z)}, \tag{3}$$

where barred quantities indicate they have been 'averaged' over angular scales, reflecting the fact that we are using one-angular bin estimates of the correlation while weighting pairs by their inverse separation. The proportionality constant is obtained from the requirement that $n_{\rm u}(z)$ has to be properly normalized.

In principle, the redshift evolution of the galaxy–matter biases and of $\bar{w}_{\rm DM}(z)$ could be estimated by measuring the one-bin auto-correlation function of both samples as a function of redshift:

$$\bar{w}_{{\rm rr},z} = \int {\rm d}z' \left[b_{\rm r}(z') n_{{\rm r},z}(z')\right]^2 \bar{w}_{\rm DM}(z'), \tag{4}$$

$$\bar{w}_{{\rm uu},z} = \int {\rm d}z' \left[b_{\rm u}(z') n_{{\rm u},z}(z')\right]^2 \bar{w}_{\rm DM}(z'), \tag{5}$$

[7] Even though these scales are clearly non-linear, these non-linearities do not have a significant impact on the methodology, as demonstrated in Schmidt et al. (2013) and in this paper. See Section 4.6 for a discussion concerning the choice of scales.

where $n_{\mathrm{r},z}(z')$ and $n_{\mathrm{u},z}(z')$ are the redshift distributions of the reference and unknown samples binned into narrow bins centred in z. If the bins are sufficiently narrow so as to consider the biases and $\bar{w}_{\mathrm{DM}}$ to be constant over the distributions, they can be pull out of the above integrals. Knowledge of the redshift distributions of the narrow bins is then required to use equations (4) and (5) to estimate b_{r}, b_{u}, and $\bar{w}_{\mathrm{DM}}$.

In our fiducial analyses we do not attempt to correct for the redshift evolution of the galaxy–matter bias and of the dark matter density field. Rather, we assume b_{r}, b_{u}, and $\bar{w}_{\mathrm{DM}}$ to be constant within each photo-z bin, and use the simulations to estimate the systematic error induced by this assumption. This choice is motivated and discussed in more details in Section 5.1 and Appendix B. Under this assumption, equation (3) reduces to

$$n_{\mathrm{u}}(z) \propto \bar{w}_{\mathrm{ur}}(z), \tag{6}$$

where the proportionality constant is again obtained requiring a proper normalization for $n_{\mathrm{u}}(z)$.

'Weighted' method. This method is a modified version of the method described in Ménard et al. (2013), and differs from that of Schmidt/Ménard in how the one-angular bin estimate of the cross-correlation signal is measured. In particular, for the 'weighted' method the correlation function is measured as a function of angle, and the recovered correlation function is averaged over angular scales via

$$\bar{w}_{\mathrm{ur}}(z) = \int_{\theta_{\mathrm{min}}}^{\theta_{\mathrm{max}}} \mathrm{d}\theta\, W(\theta) w_{\mathrm{ur}}(\theta, z), \tag{7}$$

where $W(\theta) \propto \theta^{-\gamma}$ is a weighting function. We assume $\gamma = 1$ to increase the S/N. The integration limits in the integral in equation (7) correspond to fixed physical scales (500–1500 kpc). As can be seen, the primary difference between the 'weighted' method and Schmidt/Ménard's method is whether one computes the angular correlation function first followed by a weighted integral over angular scales, or whether one performs a weighted integral of pairs first, and then computes the angular correlation function.

Newman's method (Newman 2008; Matthews & Newman 2010). Following Matthews & Newman (2010) implementation, we assume that all the correlation functions can be described by power laws $\xi(r) = (r/r_0)^{-\gamma}$. The method requires measuring n cross-correlations between the unknown sample and the reference sample, n autocorrelations of the reference sample and one autocorrelation for the unknown sample, where n is the number of redshift bins into which the reference sample is divided. Adopting a linear bias model, this allows one to relate the measured cross-correlation signal to $n_{\mathrm{u}}(z)$ and to quantities computable from a given cosmological model. Specifically, one has

$$w_{\mathrm{ur}}(\theta, z) = \frac{n_{\mathrm{u}}(z) H(\gamma_{\mathrm{ur}}) r_{0,\mathrm{ur}}^{\gamma_{\mathrm{ur}}} \theta^{1-\gamma_{\mathrm{ur}}} D_{\mathrm{A}}^{1-\gamma_{\mathrm{ur}}}}{\mathrm{d}\chi/\mathrm{d}z}. \tag{8}$$

Here γ_{ur} corresponds to the power-law slope of the correlation function, while $H(\gamma_{\mathrm{ur}}) = \Gamma(1/2)\Gamma((\gamma_{\mathrm{ur}} - 1)/2)/\Gamma(\gamma_{\mathrm{ur}}/2)$. $D_{\mathrm{A}}(z)$ and $\chi(z)$ are, respectively, the angular size distance and the comoving distance at a given redshift.

We fit the observed cross-correlation signal using a function of the form $w_{\mathrm{ur}}(\theta, z) = A_{\mathrm{ur}}(z)\theta^{1-\gamma_{\mathrm{ur}}} - C_{\mathrm{ur}}$. With respect to Schmidt's and Ménard's methods, we note that this implementation introduces two extra degrees of freedom (γ_{ur} and C_{ur}). The index γ_{ur} is fixed and it is estimated from the arithmetic mean of the indexes of the unknown and reference autocorrelation functions (see below). The parameters $A_{\mathrm{ur}}(z)$ and C_{ur} are obtained through chi-square minimization; we estimate the covariance needed for the fit through jackknife resampling. Setting our two expressions for w_{ur} equal to each other, and solving for the redshift distribution, we arrive at

$$n_{\mathrm{u}}(z) = \frac{\mathrm{d}\chi/\mathrm{d}z}{D_{\mathrm{A}}^{1-\gamma_{\mathrm{ur}}} H(\gamma_{\mathrm{ur}}) r_{0,\mathrm{ur}}^{\gamma_{\mathrm{ur}}}} A_{\mathrm{ur}}(z). \tag{9}$$

Under the assumption of linear bias, both the index of the cross-correlation function and its correlation length can be calculated from the unknown and reference autocorrelations. One has $\gamma_{\mathrm{ur}} = (\gamma_{\mathrm{uu}} + \gamma_{\mathrm{rr}})/2$ and $r_{0,\mathrm{ur}}^{\gamma_{\mathrm{ur}}} = (r_{0,\mathrm{uu}}^{\gamma_{\mathrm{uu}}} r_{0,\mathrm{rr}}^{\gamma_{\mathrm{rr}}})^{1/2}$.[8] A first guess value for $r_{0,\mathrm{uu}}$ can be inserted in equation (9) to estimate the redshift distribution, which can be inserted back in equation (8) to refine the value of $r_{0,\mathrm{uu}}$. Following Matthews & Newman (2010), we assume $r_{0,\mathrm{uu}} \propto r_{0,\mathrm{rr}}$ and we take as a first guess $r_{0,\mathrm{uu}} = r_{0,\mathrm{rr}}$. The whole procedure is repeated until convergence.

2.2 Second step: correcting photo-z posterior

Given an unknown galaxy sample, one can readily use photo-z techniques to estimate the corresponding redshift distribution. Here, we seek to use the redshift distribution recovered via cross-correlation to calibrate the photometric redshift bias of the photo-z posterior. We investigated two approaches.

(i) Criteria I – shape matching. Let $n_{\mathrm{pz}}(z)$ be the photo-z posterior for the unknown galaxy sample and $n_{\mathrm{wz}}(z)$ the redshift distribution recovered via cross-correlations. The corrected photo-z posterior is defined as $n_{\Delta}(z) \equiv n_{\mathrm{pz}}(z - \Delta z)$, where Δz is the photometric redshift bias. The photo-z bias is calibrated matching the shapes of $n_{\Delta}(z)$ and $n_{\mathrm{wz}}(z)$ within the redshift interval covered by the reference sample.

(ii) Criteria II – mean matching. Let $\langle z \rangle_{\Delta}$ be the mean of $n_{\Delta}(z)$ and $\langle z \rangle_{\mathrm{wz}}$ the mean of $n_{\mathrm{wz}}(z)$. The photo-z bias is calibrated requiring $\langle z \rangle_{\Delta}$ and $\langle z \rangle_{\mathrm{wz}}$ to match. Note that the means have to be estimated over the same redshift range.

Quantitatively, the matching is done using a likelihood function to solve for the photometric redshift bias of the photo-z posterior. We recall that we do not attempt to debias higher order moments of the photo-z posterior as the cosmological probes in the accompanying DES Y1 analysis are primarily sensitive to the mean of this distribution (DES Collaboration 2017; Hoyle et al. 2017; Krause et al. 2017; Troxel et al. 2017). The log-likelihoods for the parameter Δz for each of the two matching criteria are defined via

$$\mathcal{L}_{\mathrm{I}} = -\frac{1}{2}\chi^2\left(e^k n_{\Delta};\, n_{\mathrm{wz}};\, \hat{\Sigma}_{\mathrm{wz}}^{-1}\right) + \mathrm{Prior}\,(k, \Delta z) \tag{10}$$

and

$$\mathcal{L}_{\mathrm{II}} = -\frac{1}{2}\chi^2\left(\langle z \rangle_{\Delta};\, \langle z \rangle_{\mathrm{wz}};\, \hat{\Sigma}_{\langle z \rangle_{\mathrm{wz}}}^{-1}\right) + \mathrm{Prior}\,(\Delta z)\,. \tag{11}$$

Note these likelihoods can account for the existence of a priori estimate of the photometric redshift bias Δz. In the above equations, e^k is a relative normalization factor that rescales $n_{\Delta}(z)$, which is properly normalized to unity over the full redshift interval, to a distribution that is normalized to unity over the range of $n_{\mathrm{wz}}(z)$.

The quantity $\hat{\Sigma}$ for each of the likelihoods is the appropriate covariance matrix from the cross-correlation analysis. They are

[8] We note that if we assume constant (scale-independent) bias, then $\gamma_{\mathrm{uu}} = \gamma_{\mathrm{rr}}$. Nonetheless, we compute γ_{ur} as the arithmetic mean of γ_{uu} and γ_{rr} to follow Matthew & Newman's original recipe.

estimated from simulated data through a jackknife (JK) approach, using the following expression (Norberg et al. 2009):

$$\hat{\Sigma}(x_i, x_j) = \frac{(N_{\rm JK}-1)}{N_{\rm JK}} \sum_{k=1}^{N_{\rm JK}} (x_i^k - \bar{x_i})(x_j^k - \bar{x_j}), \quad (12)$$

where the sample is divided into $N_{\rm JK} = 1000$ subregions of roughly equal area ($\sim$1 deg^2), x_i is a measure of the statistic of interest in the ith bin of the kth sample, and $\bar{x_i}$ is the mean of our resamplings. The JK regions are safely larger than the maximum scale considered in our clustering analysis. The Hartlap correction (Hartlap, Simon & Schneider 2007) is used to compute the inverse covariance.

Finally, despite the fact that our reference sample (*redMaGiC* galaxies) spans the redshift interval [0.15, 0.85] in our simulations, in practice, in criteria II (mean matching), we restrict ourselves to a narrower redshift range, defined by the intersection of [0.15, 0.85] and $[\langle z\rangle_{\rm wz} - 2\sigma_{\rm wz}, \langle z\rangle_{\rm wz} + 2\sigma_{\rm wz}]$, where $\sigma_{\rm wz}$ is the root-mean-square of the redshift distribution $n_{\rm wz}(z)$. We have found that this choice increases the accuracy and robustness of our method by minimizing systematics (e.g. lensing magnification) associated with regions in which there is little intrinsic clustering signal. Section 4.4 quantifies the impact of this choice on our results. We do not shrink the interval used for matching under criteria I (shape matching), as this procedure is inherently less sensitive to noise and biases in the tails.

One important feature of our analysis is that, when treating multiple weak lensing (WL) source redshift bins, we treat each bin independently. In practice, there are clear statistical correlations between bins, as revealed by significant off-diagonal elements in the JK covariance matrix. However, as we demonstrate below, our analysis is easily systematics dominated. This has an important consequence, as we have found that attempting a simultaneous fit to all WL source redshift bins clearly produces incorrect results: systematic biases in one bin get propagated into a different bin via these off-diagonal terms, throwing off the best-fitting models for the ensemble. By treating each bin independently, we find that we can consistently recover numerically stable, accurate (though systematics-dominated) estimates of the photometric redshift bias.[9]

We sampled the likelihood in equations (10) and (11) using the affine-invariant Markov chain Monte Carlo (MCMC) ensemble sampler EMCEE (Foreman-Mackey et al. 2013).[10] We assume non-informative flat priors for k and Δz.

3 SIMULATED DATA

3.1 Buzzard simulations

We test our calibration procedure on the BUZZARD-V1.1 simulation, a mock DES Y1 survey created from a set of dark-matter-only simulations. The simulation and creation of the mock survey data are detailed in DeRose et al. (in preparation), Wechsler et al. (in preparation), and MacCrann et al. (2018), so we provide only a brief summary of both. BUZZARD-V1.1 is constructed from a set of three N-body simulations run using L-GADGET2, a version of GADGET2 (Springel 2005) modified for memory efficiency. The simulation boxes ranged from 1 to 4 Gpc h^{-1}. Light cones from each box were constructed on the fly. Haloes were identified using ROCKSTAR (Behroozi, Wechsler & Wu 2013), and galaxies were added to the simulations using the Adding Density Dependent GAlaxies to Light-cone Simulations (ADDGALS) algorithm (Wechsler et al., in preparation). ADDGALS uses the large-scale dark matter density field to place galaxies in the simulation based on the probabilistic relation between density and galaxy magnitude. The latter is calibrated from subhalo abundance matching in high-resolution N-body simulations. SEDs are assigned to the galaxies from a training set of spectroscopic data from SDSS Data Release 7 (DR7; Cooper et al. 2011) based on local environmental density. The SEDs are integrated in the DES pass bands to generate *griz* magnitudes. Galaxy sizes and ellipticities are drawn from distributions fit to deep SuprimeCam i'-band data. Galaxies are added to the simulation to the projected apparent magnitude limit of DES Year 5 (Y5) data out to redshift $z = 2$. The galaxy positions, shapes and magnitudes are then lensed using the multiple-plane ray-tracing code Curved-sky grAvitational Lensing for Cosmological Light conE simulatioNS (CALCLENS; Becker 2013). Finally, the catalogue is cut to the DES Y1 footprint with RA > 0 using the footprint and bad region masking including bright stars, regions of high extinction, etc., used in the actual Y1 data, and photometric errors are added using the DES Y1 depth map (Rykoff, Rozo & Keisler 2015). This yields a total masked area of 1108.13 deg^2, 12 million WL source galaxies, and 102 120 galaxies in the higher luminosity *redMaGiC* sample used in this paper, as will be discussed in Sections 3.2 and 3.3.

3.2 Unknown sample in simulations – weak lensing source sample

We seek to mimic the selection and redshift distribution of the WL source galaxies included in the DES Y1 METACALIBRATION shear catalogue described in Zuntz et al. (2017). To do so, we apply flux and size cuts to the simulated galaxies that mimic the DES Y1 source selection thresholds. Each source has its redshift estimated, and is assigned a photometric redshift, the details depending on the photo-z code used (see below). These redshifts are used to divide the source galaxies into four redshift bins corresponding to [(0.2–0.43), (0.43–0.63), (0.63–0.9) and (0.9–1.3)], in agreement with the DES Y1 cosmological analysis choices concerning bins assignment (see Hoyle et al. 2017). We note, however, that our methodology does not depend on the chosen method to define source bins (e.g. colours could have been used instead of photometric redshifts).

Because of the limited redshift coverage of the *redMaGiC* reference sample, we only apply our method to the first three redshift bins. The number densities of the WL sample in the simulations are 1.25, 0.82, and 0.64 arcmin^{-2} for these source bins. The corresponding values of the DES Y1 shear catalogue are 1.45, 1.43, and 1.47 arcmin^{-2}. The lower number densities in simulation have a negligible impact on the recovered statistical uncertainty, as the latter is dominated by the shot noise of the reference sample. Importantly, the shape of $n_{\rm pz}(z)$ as estimated by the photo-z codes in the simulations match the data with good fidelity.

Two different photo-z codes have been run on the simulated WL source samples.

(i) *The Bayesian Photometric Redshifts* (BPZ; Benítez 2000; Coe et al. 2006). BPZ is a template-based method. It returns the full probability distribution $p(z)$ for each galaxy given its magnitudes and template libraries. Templates and priors used here are described in more details in Hoyle et al. (2017), as BPZ has been run on

[9] In principle, neglecting correlations between different bins should result in an underestimation of the statistical uncertainty. In practice, this effect is negligible.
[10] http://dan.iel.fm/emcee

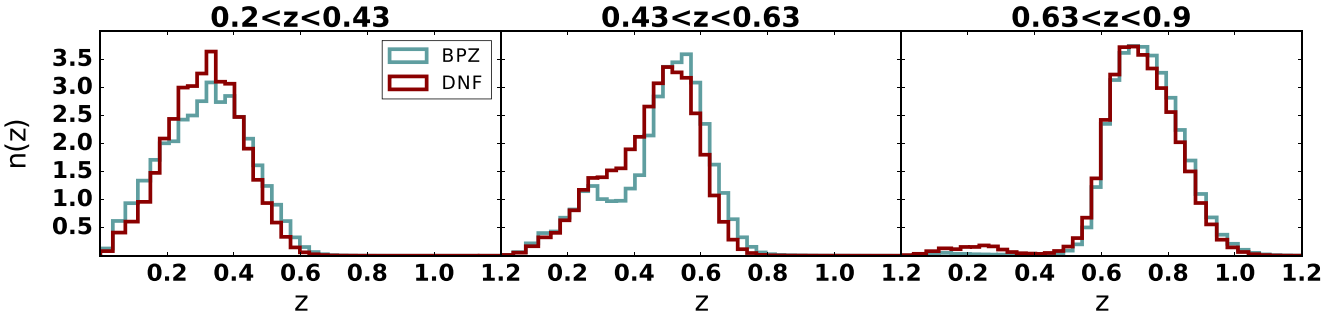

Figure 1. True redshift distributions for the simulated WL source samples obtained binning with different photo-z codes, as described in Section 3.2. The redshift distributions are normalized to unity over the full redshift interval.

simulations with the same set-up used on data. Briefly, templates are generated based on low-redshift models from Coleman, Wu & Weedman (1980) and Kinney et al. (1996), while the redshift evolution and calibration corrections to template fluxes have been computed by matching PRIsm MUlti-object Survey (PRIMUS) Data Release 1 (DR1; Coil et al. 2011) spectroscopic redshifts to DES photometry. The calibration sample comprises 72 176 galaxies. The luminosity prior used in BPZ takes the form of smooth exponential functions, which have been fitted using Cosmological Evolution Survey (COSMOS) galaxies with accurate photometric redshifts (Laigle et al. 2016).

(ii) *Directional Neighbourhood Fitting* (DNF; De Vicente, Sánchez & Sevilla-Noarbe 2016). DNF is a machine learning algorithm for galaxy photometric redshift estimation. Based on a training sample, DNF constructs the prediction hyperplane that best fits the neighbourhood of each target galaxy in multiband flux space. It then uses this hyperplane to predict the redshift of the target galaxy. This redshift is used to divide the WL sample into tomographic bins. The key feature of DNF is the definition of a new neighbourhood, the Directional Neighbourhood. Under this definition two galaxies are neighbours not only when they are close in the Euclidean multiband flux space, but also when they have similar relative flux in different bands, i.e. colours.

DNF does not directly provide the full photo-z posterior. We use the redshift of the nearest neighbour within the training sample as an approximation of a random sample from the photo-z posterior; this redshift is used to obtain the full WL redshift distribution.

DNF photo-zs require us to define a training/validation sample. The sample is first defined in data. A catalogue is built collecting high-quality spectra from more than 30 spectroscopic surveys overlapping the DES Y1 footprint and matching them to DES Y1 galaxies (Gschwend et al. 2017; Hoyle et al. 2017). This catalogue is then used to define the training/validation sample in simulations, by selecting the nearest neighbours in magnitude and redshift space. The selection algorithm is applied in HEALPIX pixels with resolution NSIDE = 128 (0.2 deg^2): if there is no galaxy in the catalogue in a given HEALPIX pixel, no simulated counterpart is selected. This roughly mimics the geometry and selection effects of the spectroscopic surveys. The final simulated training sample comprises roughly 225 000 spectra, covering a large fraction of the DES Y1 footprint.

The true redshift distributions of the sources binned according to each of the three photo-z codes are presented in Fig. 1. The two photo-z codes provide similar distributions, except in the second tomographic bin, where BPZ includes in the selection a population of low-redshift blue galaxies, responsible for the observed $z \sim 0.3$ bump. The characteristic scatter and bias for photometric redshifts obtained with BPZ are shown in Fig. 2.

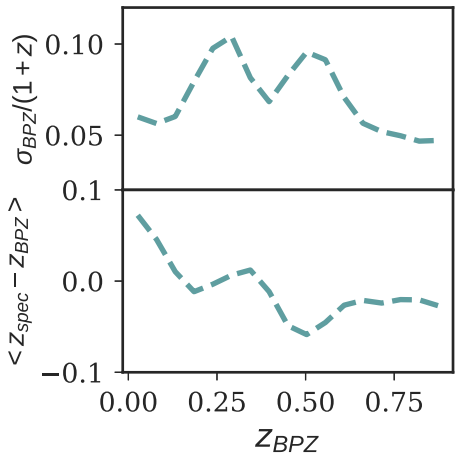

Figure 2. The scatter (top) and bias (bottom) of $z_{\rm BPZ}$ for the simulated WL sample. A similar estimate is not possible in the data due to the lack of a complete spectroscopic sample.

We note that a Random Forest (RF) code (e.g. Breiman 2001; Carrasco Kind & Brunner 2013; Rau et al. 2015) has also been run in simulations; however, results were not significantly different from the DNF results, so we decided not to include them in the paper.

3.3 Reference sample in simulations – *redMaGiC* galaxies

We use *redMaGiC* galaxies for our reference samples. These are luminous red galaxies selected as described in Rozo et al. (2016). The *redMaGiC* algorithm is designed to select galaxies with high-quality photometric redshift estimates. This is achieved by using the red-sequence model that is iteratively self-trained by the *redMaPPer* cluster finding algorithm (Rykoff et al. 2014). *redMaGiC* imposes strict colour cuts around this model to produce a luminosity-thresholded galaxy sample with a constant comoving density. The latter condition follows from the idea of trying to select the 'same' sample of galaxies at different redshift: under the approximation of no merging, red-sequence galaxies evolve passively, resulting in a constant comoving density sample.

The algorithm has only two free parameters: the desired comoving density of the sample, and the minimum luminosity of the selected galaxies. The result is a pure sample of red-sequence galaxies with nearly Gaussian photometric redshift estimates that are both accurate and precise.

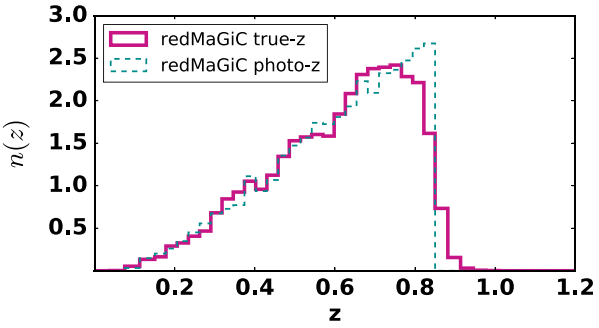

Figure 3. The redshift distribution of the simulated *redMaGiC* reference sample used in our analysis to measure cross-correlation redshifts. We show the distributions of both *redMaGiC* true-z and photo-z.

For this work we selected *redMaGiC* galaxies in the redshift interval $0.15 < z < 0.85$, applying the luminosity cut of $L > 1.5L_*$; the resulting redshift distribution is shown in Fig. 3. The reference sample is further split into 25 uniform redshift bins. In our simulation, the mask of the *redMaGiC* sample includes all the survey regions that reach sufficient depth to render the sample volume limited up to $z = 0.85$. The fact that the footprint in the simulation is ∼30 per cent smaller than in the data results in a *redMaGiC* sample that has ∼30 per cent less galaxies than the data. Moreover, due to small differences in the evolution of the red sequence between the simulation and the data, the sample reaches a maximum redshift of $z_{\rm max} = 0.85$ (instead of $z_{\rm max} = 0.9$ in data). We expect statistical errors in this work to be overestimated by ∼20 per cent with respect to data.[11] We note that to be consistent with the redshift interval considered here, the analysis in data has been performed cutting the *redMaGiC* sample at $z_{\rm max} = 0.85$ (Davis et al. 2017).

The characteristic scatter and bias of the *redMaGiC* photometric redshifts found in the data are very closely reproduced by the simulations as can be seen in Fig. 4. It should be noted that in the simulation we have the true redshifts of all *redMaGiC* galaxies, and thus can calculate the aforementioned statistics using the full sample, whereas in the data we only have an incomplete spectroscopic training set with which to make these measurements. Cawthon et al. (2017) discuss further validation of the robustness of these estimates in the data.

We also generate a catalogue of random points for *redMaGiC* galaxies. *redMaGiC* randoms are generated uniformly over the footprint, as observational systematics (e.g. airmass and seeing) are not included in the simulation and for the simulated *redMaGiC* sample used in this analysis, number density does not correlate with variation in the limiting magnitude of the galaxy catalogues.

4 SYSTEMATIC CHARACTERIZATION

In this section we test our clustering calibration of DES Y1 photo-z redshift distributions. To assess the accuracy of the methodology, we consider the mean of the redshift distributions, computed over the full redshift interval (i.e. without restricting to the matching interval where we have reference coverage). Any residual difference in the mean between the calibrated photo-z posterior and the

[11] As our methodology is systematic dominated, this has a negligible impact on the results drawn in this paper.

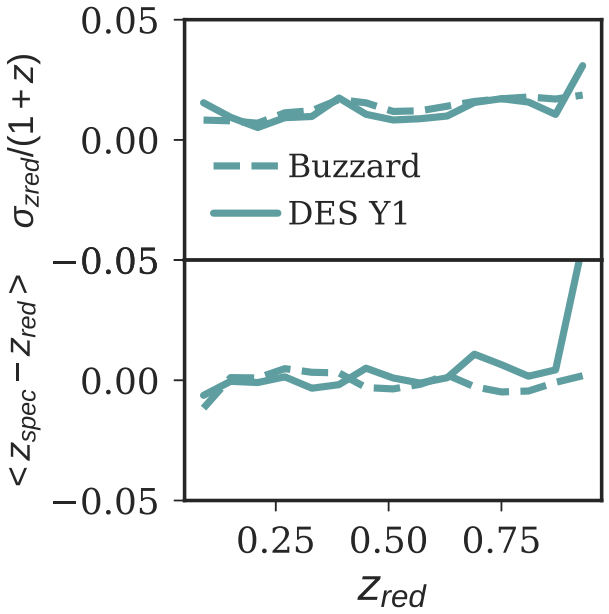

Figure 4. The scatter (top) and bias (bottom) of $z_{redMaGiC}$ for the simulated *redMaGiC* sample (dashed lines) compared to the data (solid lines).

true distribution is interpreted as a systematic uncertainty, which is quantified through the metric

$$\Delta\langle z\rangle \equiv \langle z\rangle_{\rm true} - \langle z\rangle_{\Delta}. \quad (13)$$

We will refer to this metric as the 'residual difference in the mean'. We recall that in the above equation $\langle z\rangle_{\Delta}$ is the mean of the photo-z posterior once the photo-z bias Δz has been calibrated.

Systematic errors can arise if the clustering-based redshift distribution differs from the truth, owing to the fact that:

(i) we are neglecting the redshift evolution of the galaxy–matter biases of both the WL and *redMaGiC* samples (and of the dark matter density field); hereafter, we will refer to this systematic as *bias evolution systematic*;

(ii) we are using photo-z as opposed to true redshift to bin the reference sample; hereafter, referred to as *redMaGiC photo-z systematic*.

Moreover, when we correct the photo-z posterior $n_{\rm pz}(z)$ using the clustering-based $n_{\rm wz}(z)$.

(i) If the 'shape matching' criterion is used, differences between the shapes of $n_{\rm pz}(z)$ and $n_{\rm wz}(z)$ could impact the recovered photometric redshift bias, as the criterion does not impose any requirement on the mean of $n_{\Delta}(z)$. An incorrect shape of the photo-z posterior could also affect the 'mean matching' criterion, as the matching is performed within $2\sigma_{\rm WZ}$ of $\langle z\rangle_{\rm wz}$, and the photo-$z$ posterior outside this interval cannot be calibrated. Hereafter we will refer to this systematic as *shape systematic*.

Below we introduce each of these systematics one at a time, computing each of their contributions to the total systematic error budget in our method. We will make the ansatz that our systematics can be treated as independent. We will come back to this assumption later in Section 4.5.

4.1 Bias evolution systematic

We first estimate errors due to bias evolution and evolution in the clustering amplitude of the density field. We apply our method

to a nearly ideal scenario in which the source galaxy distribution is binned in redshift bins according to the mean of the photo-z posterior as estimated by each of the photo-z codes we consider. We use the true redshifts of the *redMaGiC* reference sample when applying our cross-correlation method. We also assume that the $n_{\rm pz}(z)$ of each redshift bin is identical to the true redshift distribution.

Our results are shown in the upper panels of Fig. 5, labelled 'scenario A', and the residual shifts in the mean $\Delta\langle z\rangle_{\rm A}$ after the calibration are summarized in the first row of Tables 1 and 2. If the calibration procedure was not affected by systematic errors, we should recover residual shifts in the mean $\Delta\langle z\rangle_{\rm A}$ compatible with zero. However, the values in Tables 1 and 2 are as large as $|\Delta\langle z\rangle_{\rm A}| = 0.02$, owing to an incorrect $n_{\rm wz}(z)$ estimate. All the residual shifts are substantially larger than the typical statistical uncertainty of the measurement. The specific values of the shifts vary depending on the photo-z code (as they can select slightly different populations of galaxies) and redshift bins. The two matching criteria do not always lead to the same residual shifts: in our calibration procedure, matching the shapes of two different distributions is not expected to give the same photo-z bias obtained by matching their means.

We demonstrate in Section 5.1 that correcting for the redshift evolution of the biases and of the clustering amplitude of the density field accounts for the observed residual shifts $\Delta\langle z\rangle_{\rm A}$. This evolution can be readily estimated in our simulated data, but is difficult to account for in the real world. Therefore, the residual shifts reported in Tables 1 and 2 represent the systematic error on the photo-z bias calibration due to the bias evolution systematic.

Lastly, we note that in Fig. 5 the clustering-based estimate recovers a spurious signal (in the form of a positive tail at high redshift) for the first redshift bin, which may potentially be explained by lensing magnification effects. Disentangling the signal due to a population of photo-z outliers from lensing magnification can be non-trivial. This is not explicitly taken into account here but discussed more in detail in Section 5.2. Despite not being general, in our case the shape matching procedure is quite insensitive to biases in the tails, as the photo-z posterior correctly goes to zero. On the other hand, the mean matching within $2\sigma_{\rm WZ}$ of $\langle z\rangle_{\rm wz}$ is insensitive to the tails by construction.

4.2 *redMaGiC* photo-z systematic

Next, we relax the assumption that we have true redshifts for the reference *redMaGiC* sample. Naively, we expect that any photometric redshift biases in *redMaGiC* will imprint themselves into the clustering result. We repeat the same analysis as before, only now we use photometric rather than true redshifts for the *redMaGiC* galaxies. Since this run is affected by bias evolution, we are interested in the *change* of the residual shifts $\Delta\langle z\rangle$ relative to that in the previous section.

Results are shown in the central panels of Fig. 5, labelled 'scenario B', while the *changes* in the residual shifts $(\Delta\langle z\rangle_{\rm B} - \Delta\langle z\rangle_{\rm A})$ are summarized in the second row of Tables 1 and 2. These changes correspond to the values of the *redMaGiC* photo-z systematic. Note that we do not show the statistical uncertainty for this systematic: as the residual shifts for scenarios A and B are highly correlated (since they are estimated using similar data covariances), the statistical uncertainty of their difference is close to zero.

A comparison with the values obtained for the bias evolution systematic shows that *redMaGiC* photo-z systematic is subdominant.

4.3 Shape systematic

Relative to the previous run, in which the photo-z posterior was assumed to be the true redshift distribution of the source galaxies, we now replace the shape of the photo-z posterior by the photometrically estimated $n_{\rm pz}(z)$ from each of the photo-z codes we consider. This constitutes our most realistic scenario, as it suffers from all three systematics identified in this paper: bias evolution, *redMaGiC* photo-z, and shape systematic. Our results are shown in the lower panels of Fig. 5, labelled 'scenario C'. To disentangle the shape systematic from the other two, we compute the *change* of the residual shift in the mean $\Delta\langle z\rangle$ relative to that obtained in the previous section. The *changes* in the residual shifts $(\Delta\langle z\rangle_{\rm C} - \Delta\langle z\rangle_{\rm B})$ are summarized in the third row of Tables 1 and 2. As for the case of *redMaGiC* photo-z systematic, we do not show the statistical uncertainty, which should be close to zero.

We see from Tables 1 and 2 that the shape systematic has a much stronger impact on the shape matching criteria than on the mean matching criteria. This is particularly evident in the second redshift bin, where the differences in the shapes of the photo-z posterior and the true/cross-correlation redshift distributions are especially pronounced. Note in particular the absence of a secondary low redshift peak in the photo-z posteriors. Given the smaller systematic uncertainty associated with shape systematics in the mean matching criteria, we adopt it as our fiducial matching criteria for the DES Y1 analysis.

Given that this last run (scenario C) includes all systematic uncertainties, we also report in Tables 1 and 2 (fifth and sixth rows) the residual shift in the mean $\Delta\langle z\rangle_{\rm C}$ of the photo-z posterior before and after the calibration. Error bars only account for statistical uncertainty. In almost all the cases the calibration procedure greatly reduces the residual shifts in the mean. In particular, for many of the bins the corrected redshift distributions are consistent with zero photometric redshift bias. We note that in the second redshift bin, while it might seem by eye (Fig. 5) that the calibrated $n_{\Delta}(z)$ differs from the true distribution, their means are correctly matched.

While encouraging, this is partly due to cancelling systematic shifts (note that many of the shifts in Tables 1 and 2 have opposite signs): the final systematic error budget – which will be discussed in Section 4.5 – does not rely on such fortuitous cancellations.

4.4 Dependence of the mean matching criteria on the choice of redshift interval

We briefly discuss here our choice to apply the mean matching criteria only in the interval $\langle z\rangle_{\rm WZ} \pm 2\sigma_{\rm WZ}$. The interval has been chosen as it roughly covers most of the range sampled by the true distribution, minimizing the impact of possible systematics affecting the tails of the recovered distribution.

We estimate the values of each systematic for different interval choices, namely $\langle z\rangle_{\rm WZ} \pm 1.5\sigma_{\rm WZ}$, $\pm 2\sigma_{\rm WZ}$, $\pm 2.5\sigma_{\rm WZ}$ and a run using all reference redshift bins. The results are shown in Fig. 6. Variations in the values of the systematics are typically smaller than $\sim$0.005. However, there are two exceptions. In the first WL redshift bin, large intervals ($>2.5\sigma_{\rm WZ}$) include in the analysis the positive tail that appears in the clustering-based estimate at high redshift. This substantially affects the bias evolution systematic. In the second WL redshift bin, the redshift interval $\langle z_{\rm wz}\rangle \pm 1.5\sigma_{\rm wz}$ is narrow enough that the secondary peak in the redshift distribution is not included in the analysis. This omission introduces a larger shape systematic.

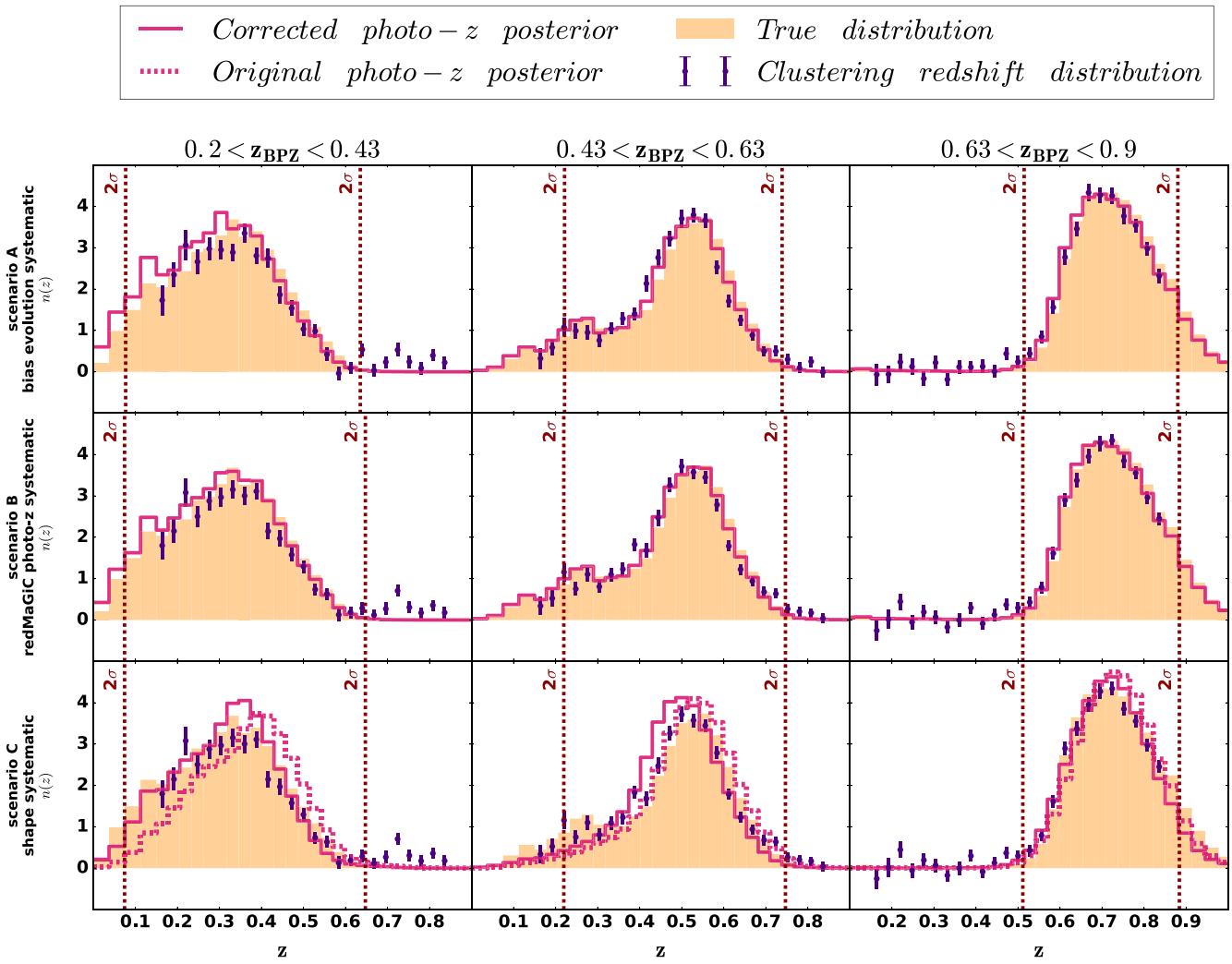

Figure 5. Calibration procedure of the photo-z posteriors of the three WL source redshift bins, for a number of different test scenarios. Points represent clustering-based estimate obtained using *redMaGiC* galaxies as a reference, while the true distribution is represented by the solid peach histogram. The pink histograms represent the corrected photo-z posterior; the uncorrected photo-z posterior is shown as dashed pink histogram, only when it does not correspond to the true distribution (therefore, it is only shown for scenario C). Only the results obtained by binning with BPZ and matching the mean are shown. The 2σ matching interval is also shown. Upper panels – scenario A: bias evolution systematic scenario. Scenario outlined in Section 4.1; reference sample is binned using *redMaGiC* true-z, and the true distribution is used as photo-z posterior. Central panels – scenario B: *redMaGiC* photo-z systematic scenario. Scenario outlined in Section 4.2; reference sample is binned using *redMaGiC* photo-z, and the true distribution is used as photo-z posterior. Lower panels – scenario C: shape systematic scenario. Scenario outlined in Section 4.3; reference sample is binned using *redMaGiC* photo-z, and the calibration is applied to the proper photo-z posterior distribution.

To accommodate the impact of the choice of interval in the cross-correlation measurement into our systematic error budget, for each WL source bin we have opted to estimate the systematic using both our $\pm 2\sigma_{\rm WZ}$ and $\pm 2.5\sigma_{\rm WZ}$ cuts, always adopting the largest of the two systematic error estimates.

4.5 Total systematic error budget

We choose not to correct for the biases found in Sections 4.1–4.3, thereby not taking advantage of the fortuitous cancellations measured in the simulations. Instead, we consider each of the shifts reported in Tables 1 and 2 as systematic errors, and proceed to add them all up in quadrature to produce our final systematic error estimate. This assumes the three sources of systematic error to be independent. We do not expect any correlation between *redMaGiC* photo-z errors, and the WL galaxy–matter bias or WL photo-z posterior. There might be slight (anti) correlations between the bias evolution systematic and the shape systematic, if the photo-z code misplaces a population of galaxies with a given bias outside the matching interval. However, assuming a correlation coefficient of $r = 0.5$ (or $r = -0.5$) between these two systematics has a negligible impact on the total error budget, so we ignored this effect.

Total systematic errors are provided in the fourth row of Tables 1 and 2, as a function of WL source redshift bin. The total error budget is dominated by the bias evolution and shape systematic, while the *redMaGiC* photo-z systematic is only responsible for a marginal contribution.

Values presented in Tables 1 and 2 are substantially larger than the typical statistical uncertainty, indicating that our calibration procedure is systematic dominated.

Table 1. BPZ systematic errors. Systematic errors for BPZ, as a function of WL redshift bin and matching procedure. $\Delta\langle z\rangle_{\rm A}$, $\Delta\langle z\rangle_{\rm B}$, and $\Delta\langle z\rangle_{\rm C}$ refers to the residual shifts in the mean relative to the scenarios A, B, and C, as outlined in Sections 4.1, 4.2, and 4.3. For the most realistic scenario (scenario C, Section 4.3), we also show the residual shifts in the mean before and after the calibration. When a value in the table is accompanied by an uncertainty, it refers to the statistical uncertainty, estimated from the posterior of the photo-z bias, as explained in Section 2.2.

	Bin 1		Bin 2		Bin 3	
	Mean match	Shape match	Mean match	Shape match	Mean match	Shape match
Bias evolution systematic: $\Delta\langle z\rangle_{\rm A}$	0.020 ± 0.006	0.019 ± 0.005	0.010 ± 0.004	0.020 ± 0.002	0.008 ± 0.003	0.007 ± 0.003
redMaGiC photo-z systematic: $\Delta\langle z\rangle_{\rm B} - \Delta\langle z\rangle_{\rm A}$	-0.009	-0.005	-0.001	-0.006	-0.001	-0.002
Shape systematic: $\Delta\langle z\rangle_{\rm C} - \Delta\langle z\rangle_{\rm B}$	-0.011	-0.017	-0.012	-0.032	0.004	0.008
Total systematic error	0.025	0.026	0.016	0.038	0.014	0.011
Uncalibrated photo-z posterior $\Delta\langle z\rangle_{\rm C}$	-0.048	-0.048	-0.040	-0.040	-0.002	-0.002
Calibrated photo-z posterior $\Delta\langle z\rangle_{\rm C}$	0.000 ± 0.006	-0.003 ± 0.005	-0.003 ± 0.003	-0.019 ± 0.003	0.011 ± 0.002	0.013 ± 0.002

Table 2. DNF systematic errors. Same as Table 1, but for DNF.

	Bin 1		Bin 2		Bin 3	
	Mean match	Shape match	Mean match	Shape match	Mean match	Shape match
Bias evolution systematic: $\Delta\langle z\rangle_{\rm A}$	0.007 ± 0.005	0.008 ± 0.004	0.012 ± 0.004	0.010 ± 0.003	0.010 ± 0.003	0.003 ± 0.002
redMaGiC photo-z systematic: $\Delta\langle z\rangle_{\rm B} - \Delta\langle z\rangle_{\rm A}$	-0.005	-0.002	-0.006	-0.001	-0.003	-0.001
Shape systematic: $\Delta\langle z\rangle_{\rm C} - \Delta\langle z\rangle_{\rm B}$	-0.007	-0.011	-0.002	-0.032	-0.015	-0.024
Total systematic error	0.012	0.013	0.013	0.034	0.019	0.025
Uncalibrated photo-z posterior $\Delta\langle z\rangle_{\rm C}$	-0.032	-0.032	-0.048	-0.048	-0.023	-0.023
Calibrated photo-z posterior $\Delta\langle z\rangle_{\rm C}$	-0.005 ± 0.005	-0.005 ± 0.003	0.004 ± 0.004	-0.024 ± 0.002	-0.009 ± 0.003	-0.023 ± 0.002

4.6 Choice of method and angular scales

Throughout this paper we have adopted as our fiducial clustering-based method the one introduced by Schmidt et al. (2013) and considered physical scales between 500 and 1500 kpc. We test the impact of the choice of angular scales by recomputing the residual shifts in the mean for one of the dominant systematic (the bias evolution systematic) with a different choice of physical scales and methods.

Fig. 7 shows the residual shifts using the Schmidt/Ménard, 'weighted', and Newman methods (outlined in Section 2.1), and for the following scales: 200–1250 kpc (i.e. small scales only), 1250–8000 kpc (i.e. large scales only), 200–8000 kpc and 500–1500 kpc (our fiducial choice for this work).

We find that the 'weighted' and Schmidt/Ménard methods perform similarly. Small differences arise because of how the two methods average over angular scales. The iterative bias correction procedure implemented in Newman's method does not lead to a reduction of the bias evolution systematic; this is due to the underlying assumption of $b_{\rm u} \propto b_{\rm r}$ in our implementation of the method, which does not hold in our case (see Section 5.1). We also find that the Newman method results in the noisiest estimates, as the implementation of Newman's method introduces two new degrees of freedom that the Schmidt/Ménard and 'weighted' approaches do not ($\gamma_{\rm ur}$ and $C_{\rm ur}$, see Section 2.1). For the largest angular scales (1250–8000 kpc), the reconstruction is so noisy that it fails to provide corrections to the photo-z posteriors, as the MCMC chains fail to converge. The noisier estimates are due to the power-law fits: some bins can have degeneracies among power-law parameters (especially in the tails of the $n(z)$), and in some cases the cross-correlation signal deviates from a pure power-law shape. We note that we could have fitted a parametrized form for the evolution of the amplitude of the reference autocorrelation function using coarser binning (similarly to what has been done in Hildebrandt et al. 2017); this could have led to a more stable recovery of the power-law parameters. As we did not use Newman's method as our fiducial method, we decided not to attempt this here.

Results are compatible within statistical errors for different choices of angular scales. Only the shifts obtained using Schmidt/Ménard and 'weighted' methods over our nominal scales

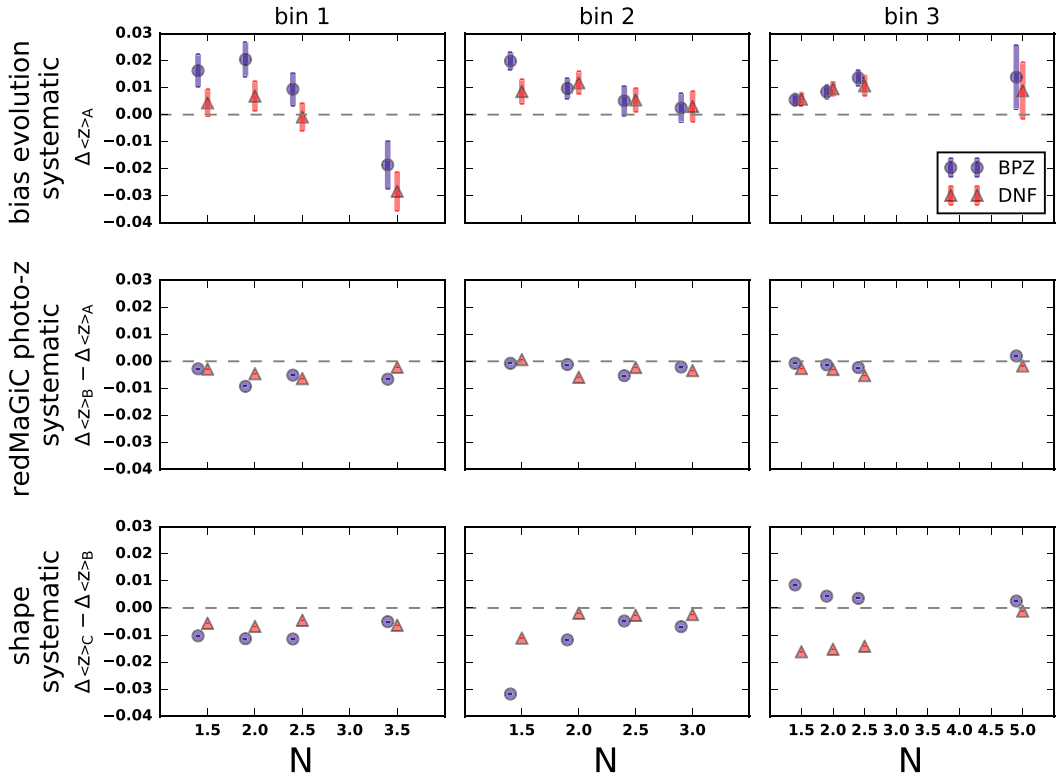

Figure 6. Mean matching procedure: inferred values of the three systematics as a function of matching interval, defined as $\langle z \rangle_{\rm WZ} \pm N\sigma_{\rm WZ}$. For each WL redshift bin, we show results for $N = 1.5$, 2, and 2.5; in addition, we show results obtained including all the redshift interval covered by *redMaGiC* galaxies. This corresponds to $N = 3.5$, 3, and 5 for the first, second, and third source redshift bin, respectively. For the third bin a larger N is needed to include the low-redshift tail down to $z = 0.15$.

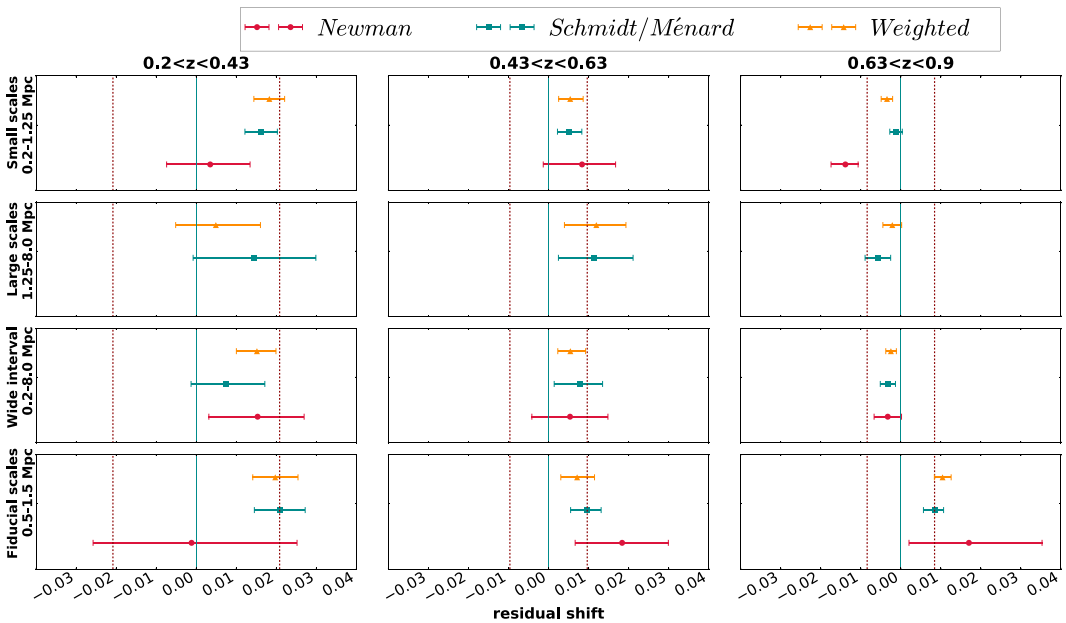

Figure 7. Residual shifts in the mean for the scenario A outlined in Section 4.1 (equivalent to the bias evolution systematic), for different choices of physical scales and different clustering-based methods. Error bars represent the statistical errors of the measurements. Vertical lines highlight our fiducial value for the bias evolution systematic, computed using the Schmidt/Ménard method over scales 500–1500 kpc. Only the values for the mean matching procedure and for BPZ are shown. In the large scales panel, red points are missing as the Newman method failed to provide a correction.

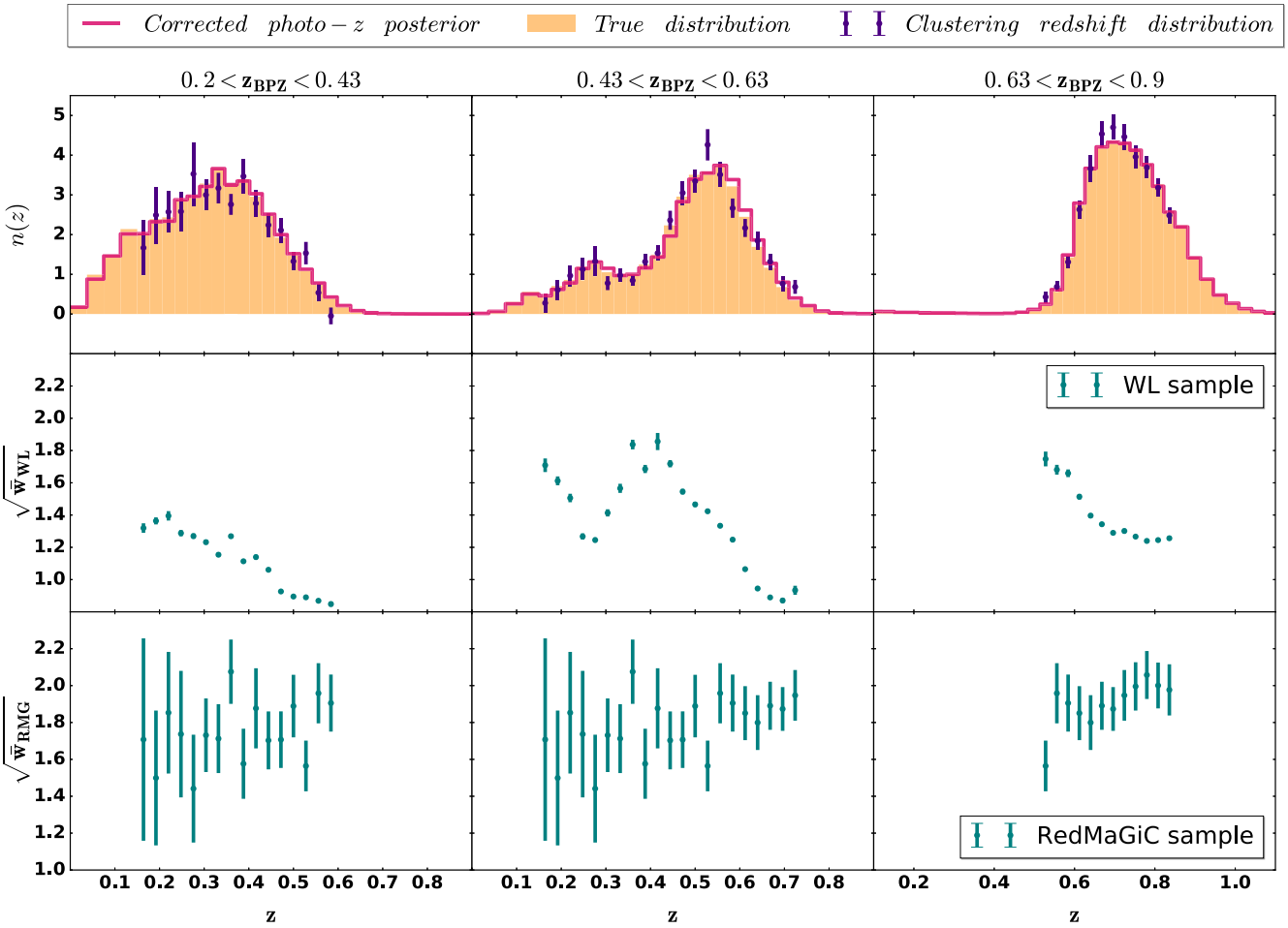

Figure 8. WL photo-z posterior calibration after correcting for the redshift evolution of the bias and dark matter density field of both samples (see Section 5.1). Top panels: clustering-based $n_{\rm wz}(z)$ and corrected photo-z posteriors $n_{\Delta}(z)$. The true redshift distributions have been used as photo-z posteriors. Middle and bottom panels: one-bin autocorrelation estimates for the WL and *redMaGiC* samples, as a function of redshift. We restricted the clustering-based estimate to the interval where it was possible to measure the autocorrelation function of the two samples. Only the results obtained by binning with BPZ and matching the mean are shown.

in the third photo-z bin are significantly different from the rest. However, the differences are safely within the bounds of our total systematic error. Fig. 7 suggests that non-linear galaxy–matter bias (e.g. Smith, Scoccimarro & Sheth 2007) seems to have a negligible impact on our methodology. At large scales, as the amplitude of the clustering signal decreases, the S/N deteriorates, so we have chosen not to use the largest scales (>1500 kpc) in our fiducial analysis. This is in line with the findings of Schmidt et al. (2013) and Ménard et al. (2013), which advocate the use of small scales to improve the S/N of the measurement. We note that systematic effects (like observational systematics) that might have an impact in real data cannot be responsible for the decrease in the S/N here: only depth variations are included in the simulation, but as the number density of the reference sample does not correlate with such variations (see Section 3.3), they should not impact the estimated $N(z)$ or its noise properties.

From Fig. 7, it is clear that using scales as small as 200 kpc appears to improve the statistical and systematic uncertainties of our method relative to the 500 kpc inner scale radius. However, the differences are small. We have opted to use the larger 500 kpc radius to avoid possible systematic uncertainties that may arise in the data but not in our simulation. In particular, photometric contamination from nearby galaxies will become more important as the inner scale radius is reduced. Several works have pointed out that blending becomes relevant on subarcminute scales (see Hartlap et al. 2011; Applegate et al. 2014; Morrison & Hildebrandt 2015; Simet & Mandelbaum 2015; Choi et al. 2016; MacCrann et al. 2017); in this respect, our choice to set $r_{\rm min} = 500$ kpc is meant to safeguard our results against any such kind of neighbouring biases.

5 DISCUSSION

5.1 Impact of galaxy–matter biases on clustering-based methods

We now demonstrate our previous claim that the bias evolution is responsible for the systematic shifts we observed in Section 4.1, when we used true redshifts for the reference sample and the true redshift distribution as $n_{\rm pz}(z)$.

In our standard methodology we have chosen not to correct for the redshift evolution of the galaxy–matter bias of both samples (and for the evolution of the dark matter density field, even if the effect is generally subdominant; Ménard et al. 2013). This approximation holds as long as the biases evolve on scales larger than the

Table 3. Residual shifts in the mean for the scenario A outlined in Section 4.1 (equivalent to the bias evolution systematic), using the estimator introduced in Section 5.1, which accounts for the redshift evolution of the bias and dark matter density field of both samples. We restricted the clustering-based estimate to the interval where it was possible to measure the autocorrelation function of the two samples.

	Bin 1		Bin 2		Bin 3	
	Mean match	Shape match	Mean match	Shape match	Mean match	Shape match
Bias evolution systematic (BPZ):	-0.004 ± 0.013	0.001 ± 0.007	-0.007 ± 0.007	0.004 ± 0.003	0.002 ± 0.003	0.001 ± 0.002
Bias evolution systematic (DNF):	0.000 ± 0.009	-0.001 ± 0.006	-0.002 ± 0.005	0.002 ± 0.004	0.004 ± 0.003	0.002 ± 0.003

typical width of the unknown distribution. In this sense, it is clear that binning the unknown sample into narrow bins through photo-z or colour selection helps to reduce the impact of bias evolution (Ménard et al. 2013; Schmidt et al. 2013; Rahman et al. 2015). If the bins are not sufficiently narrow, neglecting the bias evolution leads to systematic shifts.

We estimate the redshift evolution of the galaxy–matter bias and dark matter density field looking at the autocorrelation functions of the reference and the unknown samples, as suggested in Ménard et al. (2013) and explained in Section 2.1. We therefore bin both samples in 25 equally spaced thin bins from $z = 0.15$ to 0.85 using their true redshift and we then measure the one-bin autocorrelation functions of the samples. If the bins are sufficiently narrow, and each bin has a top-hat shape, the autocorrelation functions are simply proportional to $b^2 w_{\rm DM}$, as shown in Section 2.1. We caution that we use uniform randoms to compute the WL autocorrelation functions: even though the WL sample selection function used in simulation only roughly mimics the one applied to data, using uniform random is only approximately correct. With this caveat in mind, we estimate the redshift distribution using the following estimator:

$$n_{\rm u}(z) \propto \frac{\bar{w}_{\rm ur}(z)}{\sqrt{\bar{w}_{\rm uu}(z)\bar{w}_{\rm rr}(z)}}. \tag{14}$$

The results obtained using this estimator are shown in Fig. 8 and reported in Table 3.

By using the new estimator, we can obtain residual shifts that are compatible with zero (see the values reported in Table 3 and Fig. 8) within statistical uncertainty. The correction induced by including in the estimator the term $\sqrt{\bar{w}_{\rm uu}(z)}$ accounts for most of the bias evolution systematic, indicating that the major contribution to the systematic is due to the WL sample. Interestingly, in the second source bin, the low-redshift hump in the redshift distribution is associated with a dip in the bias evolution (see Fig. 8): this is due to a small population of low-redshift blue galaxies selected by BPZ. Concerning the reference sample, the correction induced by the term $\sqrt{\bar{w}_{\rm rr}(z)}$ is negligible. This is not unexpected: the *redMaGiC* sample is mainly composed of passively evolving, early-type, red-sequence galaxies, with similar luminosity/stellar mass, constituting a 'best-case' sample in terms of minimal bias evolution.

We emphasize that this estimator can be implemented only in simulations, since in data we do not have access to the true redshifts needed to bin the samples. In Appendix B we show an alternative correction obtained by binning the samples using their photo-z. The correction only works for the *redMaGiC* bias, but we decided not to implement it as its impact is negligible.

The bias evolution of the unknown sample constitutes one of the major issues for clustering-based methods, and it is one of the dominant sources of systematic in our work. It is worth noting that in our simulation the bias evolution can be complex (as it can be inferred from the middle panels of Fig. 8) and therefore not especially suited for correction using simple parametric approaches (e.g. Matthews & Newman 2010; Schmidt et al. 2013; Davis et al. 2018).

As the clustering amplitudes of galaxies have been found to depend on galaxy types, colours, and luminosities (Zehavi et al. 2002; Hogg et al. 2003; Blanton et al. 2006; Coil et al. 2006; Cresswell & Percival 2009; Marulli et al. 2013; Skibba et al. 2014), further splitting the unknown sample into smaller subsamples with similar colours/luminosity properties (together with thinner binnings in redshift space) should alleviate the impact of bias evolution (Scottez et al. 2018; van Daalen & White 2018). We also note that one could break the degeneracy between galaxy bias and redshift distribution using other probes, like galaxy–galaxy lensing (Prat et al. 2017).

5.2 Impact of lensing magnification on clustering-based methods

It is well known that lensing magnification (Narayan 1989; Villumsen, Freudling & da Costa 1997; Moessner & Jain 1998) can lead to a change in 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. The net effect is driven by the slope of the luminosity function, and it has an impact on the measured clustering signal (Bartelmann & Schneider 2001; Scranton et al. 2005; Ménard et al. 2010; Morrison et al. 2012). For the WL samples, size bias can also be important (Schmidt et al. 2009), but it will not be considered here.

In the context of clustering-based redshift estimates, lensing magnification effects have generally been neglected (Matthews & Newman 2010; Johnson et al. 2017; van Daalen & White 2018). Ménard et al. (2013) state that the amplitude of the magnification effect on arcminute scales is generally negligible compared to the clustering signal of overlapping samples, and it has a mild dependence with redshift. However, magnification may become dominant in the regimes where the unknown and reference samples do not overlap, as discussed in Matthews (2014) (see also McQuinn & White 2013; Choi et al. 2016).

We try here to estimate the impact of lensing magnification on our recovered clustering-based $n(z)$. The impact of lensing magnification on the galaxy overdensity is

$$\delta_{\rm obs} = \delta_{\rm g} + \delta_{\mu}, \tag{15}$$

where $\delta_{\rm g}$ is the galaxy overdensity, while δ_{μ} is the overdensity induced by lensing magnification effects. The cross-correlation signal

between two different samples is therefore

$$w_{\rm ur,tot}(\theta) = \langle \delta_{\rm g,u}, \delta_{\rm g,r} \rangle(\theta) + < \delta_{\rm g,u}, \delta_{\mu,\rm r} > (\theta) + \langle \delta_{\rm g,r}, \delta_{\mu,\rm u} \rangle(\theta) + \langle \delta_{\mu,\rm u}, \delta_{\mu,\rm r} \rangle(\theta). \quad (16)$$

The first term on the right-hand side of equation (16) is associated with the clustering due to gravitational interactions, and disappears in the case of no redshift overlap between the unknown and reference samples. All of our methodology described in Section 2.1 assumes this term to be the dominant one. The second and third terms correspond to the lensing magnification contribution. The fourth term is generally small and can be neglected (Duncan et al. 2014).

Using the Limber and flat-sky approximations (e.g. Hui, Gaztañaga & Loverde 2007; Loverde, Hui & Gaztañaga 2008; Choi et al. 2016), the first clustering term in the above expression can be modelled via

$$\langle \delta_{\rm g,u}, \delta_{\rm g,r} \rangle(\theta) = b_{\rm u} b_{\rm r} \int \frac{{\rm d}l\, l}{2\pi} J_0(l\theta) \int {\rm d}\chi \times \frac{[n_{\rm u}(z(\chi))]\,[n_{\rm r}(z(\chi))]}{\chi^2 H(z)} P_{\rm NL}\left(\frac{l+1/2}{\chi}, z(\chi)\right). \quad (17)$$

The terms $b_{\rm u}$ and $b_{\rm r}$ indicate the galaxy–matter bias of the two samples, χ is the comoving distance and $H(z)$ is the Hubble expansion rate at redshift z. J_0 is the zeroth-order Bessel function. $P_{\rm NL}(k, \chi)$ is the 3D matter power spectrum at wavenumber k (which, in the Limber approximation, is set equal to $(l + 1/2)/\chi$) and at the cosmic time associated with redshift z.

Under the approximation of weak gravitational lensing, the terms due to lensing magnification in equation (16) can be written as

$$\langle \delta_{\rm g,1}, \delta_{\mu,2} \rangle(\theta) = b_1(2.5s_2 - 1) \int \frac{{\rm d}l\, l}{2\pi} J_0(l\theta) \int {\rm d}\chi n_1(z(\chi)) \times \frac{q_2(\chi)}{\chi^2 H(z)} P_{\rm NL}\left(\frac{l+1/2}{\chi}, z(\chi)\right). \quad (18)$$

The subscripts 1 and 2 are such that equation (18) can refer either to the term $\langle \delta_{\rm g,\,u}, \delta_{\mu,\,\rm r} \rangle$ or to $\langle \delta_{\rm g,\,r}, \delta_{\mu,\,\rm u} \rangle$. The term s_2 is the slope of the cumulative number counts evaluated at flux limit of the sample '2'. The slope of the cumulative number counts is formally defined for a flux-limited sample as

$$s = \frac{{\rm d}\log_{10} n(>m)}{{\rm d}m}, \quad (19)$$

where $n(m)$ is the cumulative number counts as a function of magnitude m, and s is to be evaluated at the flux limit of the sample. The term $q_2(\chi)$ is the lensing redshift weight function defined as

$$q_2(\chi) = \frac{3H_0^2\Omega_{\rm m}\chi}{c^2 a(\chi)} \int_{\chi}^{\chi(z=\infty)} {\rm d}\chi' n_2(z(\chi')) \frac{{\rm d}z}{{\rm d}\chi'} \frac{\chi' - \chi}{\chi'}. \quad (20)$$

H_0 and $a(\chi)$ are, respectively, the Hubble constant today and the scale factor.

Knowing the redshift distribution, the bias evolution and the slope of the cumulative number counts for the two samples, theoretical predictions for the expected clustering-based $n(z)$ signal can be made through equations (17) and (18) and compared to the signal measured in simulations.

The true redshift distribution of the two samples is obtainable from the simulations. For the bias evolution, we make use of the one-point estimate measured in Section 5.1, appropriately corrected for the contribution due to the dark matter density field. For the sake of simplicity, we do not propagate to the theoretical predictions the statistical uncertainty of the one-point estimates of the two samples biases.

Concerning the slope of cumulative number counts, *redMaGiC* galaxies are in principle not a flux-limited sample (the sample is indeed volume limited up to $z = 0.85$, and on top of that, galaxies are required to belong to the red sequence and to have luminosity greater than a fixed threshold value; see Rozo et al. 2016). However, *redMaGiC* galaxies are binned in thin redshift bins; within each bin, the sample can be well approximated as flux limited ($m > M_{\rm lim} + 5\log_{10}(d(z_{\rm bin}))$. The thinner the bins, the better the approximation: this should be reflected as a sharp drop in the number counts as a function of magnitude. Therefore, for each bin, we evaluate the slope of the cumulative number count using equation (19) at the magnitude where the number counts start to drop.

For the WL sample the selection is way more complex and equation (19) cannot be directly applied. Fully characterizing the selection function for the WL sample goes beyond the scope of this paper. We consider the predicted lensing signal for two characteristic values of the amplitude parameter $2.5s - 1$, namely $2.5s - 1 = \pm 1.5$.

The results of this procedure are shown in Fig. 9. We see that the predicted magnification signal is qualitatively similar to the excess clustering observed in the simulations, suggesting that the excess shown at high redshift in the top left-hand panel of Fig. 5 is indeed due to magnification induced by *redMaGiC* galaxies at high redshift. Magnification due to the WL sample acting as a source is producing a noticeable effect only in the third bin, and the effect depends on the exact value of the amplitude parameter $2.5s - 1$. We note that to firmly exclude other possible sources for the excess signal seen in Fig. 5, we would need to re-run our analysis on a simulation with lensing effects turned off. This goes beyond the scope of this paper, but it might be needed in the near future, as survey requirements will become more stringent and lensing effects would need to be characterized at a few per cent level.

The result of this test shows that lensing magnification can have a non-negligible impact on the clustering-based $n(z)$, mostly on the tails of the recovered distribution, possibly hampering the detection of populations of photo-z outliers. It is worth stressing that the procedure presented in this paper is little affected by lensing magnification, as we cut out the tails from our analysis. We leave properly incorporating weak-lensing magnification effects into the analysis to future work.

6 CONCLUSIONS

Using numerical simulations, we characterize the performance of clustering-based calibration of the DES Y1 redshift distributions. Our standard calibration procedure is divided into two steps: a first step where the redshift distribution of a given science sample is estimated using a clustering-based method; a second step where this estimated redshift distribution is used to correct for an overall photometric redshift bias in the posterior of traditional photo-z algorithms.

We use *redMaGiC* galaxies as the reference sample for the clustering-based estimate. We show that our procedure could be applied in the case of partial overlap in redshift space between the reference sample and the science sample. As for the science sample, we consider a simulated version of DES Y1 WL source galaxies, divided in three redshift bins. We present the results for the photo-z posterior of two different photo-z codes (a template-based code, BPZ, and a machine learning code, DNF). The photo-z codes are also used to bin the WL source redshift bins, using their mean photo-z redshift.

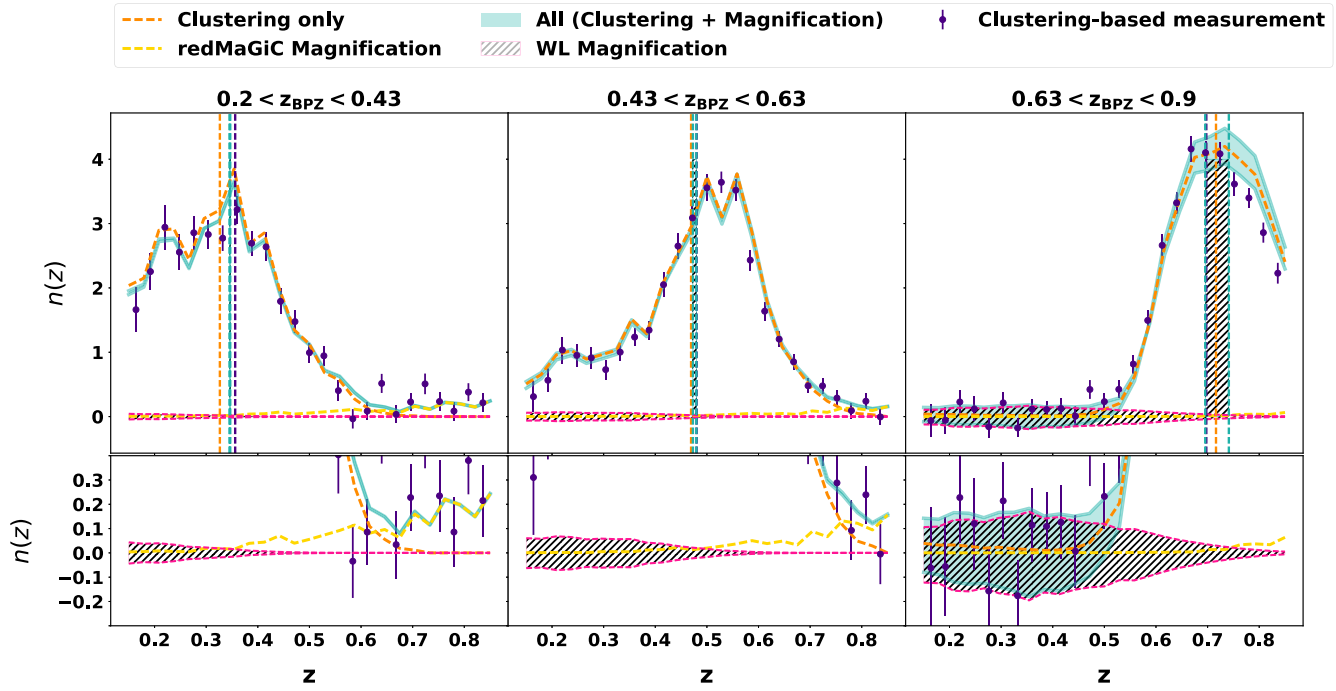

Figure 9. Upper panels: effect of lensing magnification on clustering-based estimate. Shaded regions represent theoretical predictions for the various components of the signal (lensing magnification due to *redMaGiC* and the WL sample, as well as the clustering due to gravitational interaction only), while the points represent the actual measurement in simulations. Vertical lines show the mean of the redshift distributions computed over the full redshift interval. Lower panels: as the upper panels, but zoomed in on the lensing magnification signal.

We identify and characterized in our procedure three main sources of systematic errors in our methodology.

(i) *Bias evolution systematic*. Systematic error induced by neglecting the redshift evolution of the galaxy–matter biases of the WL and *redMaGiC* samples and the evolution of the dark matter density field.

(ii) *redMaGiC photo-*z *systematic*. Systematic caused by not using a spectroscopic sample as a reference.

(iii) *Shape systematic*. Systematic due to an incorrect shape of the photo-z posterior. This systematic is exacerbated if there is only a partial overlap between the *redMaGiC* and WL samples.

We find the bias evolution systematic (particularly, the effect due to the bias evolution of the WL sample) and shape systematic to dominate the total error budget. We also find statistical uncertainties in our procedure to be subdominant with respect to systematic errors. Total systematic errors for our calibration procedure, as a function of WL source redshift bin and photo-z code, are provided in Section 4.5, and stand at the level of $\Delta\langle z\rangle \lesssim 0.02$.

We further address the impact of changing our fiducial choices concerning the angular scales and method used for the clustering-based estimate, and discuss how our methodology could be improved. In particular, future works have to efficiently deal with the problem of the redshift evolution of the galaxy–matter bias of the science sample. This could be achieved by further splitting the science sample in luminosity/colour cells. Other probes, like galaxy–galaxy lensing, could be also used to break the degeneracy between galaxy bias and redshift distribution. Lensing magnification, whose impact is marginal in this study, might no longer be negligible as survey requirements become more stringent. Lastly, we note that as clustering-based methods improve and systematic errors become subdominant with respect to statistical errors, full modelling of the cross-covariance between clustering-based $n(z)$ and other two-point correlation functions will be required so as not to bias the cosmological analysis.

The calibration strategy presented in this paper is fully implemented in the DES Y1 cosmic shear and combined two-point function analysis (DES Collaboration 2017; Troxel et al. 2017). Its direct application to DES Y1 data is discussed in two other companion papers (Cawthon et al. 2017; Davis et al. 2018). Even though we show systematic errors to dominate over statistical uncertainties for this calibration procedure, this does not have negative implications for the DES Y1 cosmological analysis, which remains statistically dominated.

ACKNOWLEDGEMENTS

This paper has gone through internal review by the DES collaboration. It has been assigned DES paper id DES-2017-0261 and FermiLab preprint number PUB-17-317-A-AE.

Support for DG was provided by NASA through Einstein Postdoctoral Fellowship grant number PF5-160138 awarded by the *Chandra* X-ray Center, which is operated by the Smithsonian Astrophysical Observatory for NASA under contract NAS8-03060. ER acknowledges support by the DOE Early Career Program, DOE grant DE-SC0015975, and the Sloan Foundation grant FG-2016-6443.

Funding for the DES Projects has been provided by the US Department of Energy, the US 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.

This study based in part on observations at Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

The DES data management system is supported by the National Science Foundation under Grant Numbers AST-1138766 and AST-1536171. The DES participants from Spanish institutions are partially supported by MINECO under grants AYA2015-71825, ESP2015-88861, FPA2015-68048, SEV-2012-0234, 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 Australian Research Council Centre of Excellence for All-sky Astrophysics (CAASTRO), through project number CE110001020.

This paper has been authored by *Fermi* Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the US Department of Energy, Office of Science, Office of High Energy Physics. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this paper, or allow others to do so, for United States Government purposes.

This research used computing resources at SLAC National Accelerator Laboratory, and at the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the US Department of Energy under Contract No. DE-AC02-05CH11231. This research was funded partially by the Australian Government through the Australian Research Council through project DP160100930.

REFERENCES

Aihara H. et al., 2018, PASJ, 70, S8
Alam S. et al., 2015, ApJS, 219, 12
Applegate D. E. et al., 2014, MNRAS, 439, 48
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Becker M. R., 2013, MNRAS, 435, 115
Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013, ApJ, 762, 109
Benítez N., 2000, ApJ, 536, 571
Bezanson R. et al., 2016, ApJ, 822, 30
Blanton M. R., Eisenstein D., Hogg D. W., Zehavi I., 2006, ApJ, 645, 977
Breiman L., 2001, Machine Learning, 45, 5
Carrasco Kind M., Brunner R. J., 2013, MNRAS, 432, 1483
Cawthon R. et al., 2017, MNRAS, preprint (arXiv:1712.07298)
Choi A. et al., 2016, MNRAS, 463, 3737
Coe D., Benítez N., Sánchez S. F., Jee M., Bouwens R., Ford H., 2006, AJ, 132, 926
Coil A. L., Newman J. A., Cooper M. C., Davis M., Faber S. M., Koo D. C., Willmer C. N. A., 2006, ApJ, 644, 671
Coil A. L. et al., 2011, ApJ, 741, 8
Coleman G. D., Wu C.-C., Weedman D. W., 1980, ApJS, 43, 393
Cooper M. C. et al., 2011, ApJS, 193, 14
Cresswell J. G., Percival W. J., 2009, MNRAS, 392, 682
Cunha C. E., Lima M., Oyaizu H., Frieman J., Lin H., 2009, MNRAS, 396, 2379
Davis M., Peebles P. J. E., 1983, ApJ, 267, 465
Davis C. et al., 2017, MNRAS, preprint (arXiv:1710.02517)
Davis C. et al., 2018, MNRAS, 477, 2196
de Jong J. T. A. et al., 2015, A&A, 582, A62
DES Collaboration 2017, preprint (arXiv:1708.01530)
De Vicente J., Sánchez E., Sevilla-Noarbe I., 2016, MNRAS, 459, 3078
Duncan C. A. J., Joachimi B., Heavens A. F., Heymans C., Hildebrandt H., 2014, MNRAS, 437, 2471
Elvin-Poole J. et al., 2017, preprint (arXiv:1708.01536)
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Gschwend J. et al., 2017, preprint (arXiv:1708.05643)
Hartlap J., Simon P., Schneider P., 2007, A&A, 464, 399
Hartlap J., Hilbert S., Schneider P., Hildebrandt H., 2011, A&A, 528, A51
Hildebrandt H. et al., 2010, A&A, 523, A31
Hildebrandt H. et al., 2017, MNRAS, 465, 1454
Hogg D. W. et al., 2003, ApJ, 585, L5
Hoyle B. et al., 2017, MNRAS, preprint (arXiv:1708.01532)
Hui L., Gaztañaga E., Loverde M., 2007, Phys. Rev. D, 76, 103502
Johnson A. et al., 2017, MNRAS, 465, 4118
Kinney A. L., Calzetti D., Bohlin R. C., McQuade K., Storchi-Bergmann T., Schmitt H. R., 1996, ApJ, 467, 38
Krause E. et al., 2017, preprint (arXiv:1706.09359)
Laigle C. et al., 2016, ApJS, 224, 24
Landy S. D., Szalay A. S., 1993, ApJ, 412, 64
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Lilly S. J. et al., 2009, ApJS, 184, 218
Lima M., Cunha C. E., Oyaizu H., Frieman J., Lin H., Sheldon E. S., 2008, MNRAS, 390, 118
Liske J. et al., 2015, MNRAS, 452, 2087
Loverde M., Hui L., Gaztañaga E., 2008, Phys. Rev. D, 77, 023512
LSST Science Collaboration, 2009, preprint (arXiv:0912.0201)
MacCrann N. et al., 2017, MNRAS, 465, 2567
MacCrann N. et al., 2018, preprint (arXiv:1803.09795)
McQuinn M., White M., 2013, MNRAS, 433, 2857
Marulli F. et al., 2013, A&A, 557, A17
Masters D. C., Stern D. K., Cohen J. G., Capak P. L., Rhodes J. D., Castander F. J., Paltani S., 2017, ApJ, 841, 111
Matthews D. J., 2014, PhD thesis, Univ. Pittsburgh, Pittsburgh, PA
Matthews D. J., Newman J. A., 2010, ApJ, 721, 456
Ménard B., Scranton R., Fukugita M., Richards G., 2010, MNRAS, 405, 1025

Ménard B., Scranton R., Schmidt S., Morrison C., Jeong D., Budavari T., Rahman M., 2013, preprint (arXiv:1303.4722)
Moessner R., Jain B., 1998, MNRAS, 294, L18
Morrison C. B., Hildebrandt H., 2015, MNRAS, 454, 3121
Morrison C. B., Scranton R., Ménard B., Schmidt S. J., Tyson J. A., Ryan R., Choi A., Wittman D. M., 2012, MNRAS, 426, 2489
Morrison C. B., Hildebrandt H., Schmidt S. J., Baldry I. K., Bilicki M., Choi A., Erben T., Schneider P., 2017, MNRAS, 467, 3576
Narayan R., 1989, ApJ, 339, L53
Newman J. A., 2008, ApJ, 684, 88
Newman J. A. et al., 2013, ApJS, 208, 5
Newman J. A. et al., 2015, Astropart. Phys., 63, 81
Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, MNRAS, 396, 19
Prat J. et al., 2017, preprint (arXiv:1708.01537)
Rahman M., Ménard B., Scranton R., Schmidt S. J., Morrison C. B., 2015, MNRAS, 447, 3500
Rau M. M., Seitz S., Brimioulle F., Frank E., Friedrich O., Gruen D., Hoyle B., 2015, MNRAS, 452, 3710
Rozo E. et al., 2016, MNRAS, 461, 1431
Rykoff E. S. et al., 2014, ApJ, 785, 104
Rykoff E. S., Rozo E., Keisler R., 2015, AJ, preprint (arXiv:1509.00870)
Sánchez C. et al., 2014, MNRAS, 445, 1482
Schmidt F., Rozo E., Dodelson S., Hui L., Sheldon E., 2009, Phys. Rev. Lett., 103, 051301
Schmidt S. J., Ménard B., Scranton R., Morrison C., McBride C. K., 2013, MNRAS, 431, 3307
Schneider M., Knox L., Zhan H., Connolly A., 2006, ApJ, 651, 14
Scottez V. et al., 2016, MNRAS, 462, 1683
Scottez V., Benoit-Lévy A., Coupon J., Ilbert O., Mellier Y., 2018, MNRAS, 474, 3921
Scranton R. et al., 2005, ApJ, 633, 589
Simet M., Mandelbaum R., 2015, MNRAS, 449, 1259
Skibba R. A. et al., 2014, ApJ, 784, 128
Smith R. E., Scoccimarro R., Sheth R. K., 2007, Phys. Rev. D, 75, 063512
Spergel D. et al., 2013, preprint (arXiv:1305.5422)
Springel V., 2005, MNRAS, 364, 1105
The Dark Energy Survey Collaboration, 2005, preprint (arXiv:e-print)
Troxel M. A. et al., 2017, preprint (arXiv:1708.01538)
van Daalen M. P., White M., 2018, MNRAS, 476, 4649
Villumsen J. V., Freudling W., da Costa L. N., 1997, ApJ, 481, 578
Zehavi I. et al., 2002, ApJ, 571, 172
Zuntz J. et al., 2017, MNRAS, preprint (arXiv:1708.01533)

APPENDIX A: RESULTS FOR A DIFFERENT *redMaGiC* GALAXY SAMPLE

In this paper we have adopted *redMaGiC* galaxies as a reference sample, as opposed to the more standard choice of using spectroscopic samples (e.g. Ménard et al. 2013; Schmidt et al. 2013; Choi et al. 2016; Hildebrandt et al. 2017). This choice has been mainly driven by the necessity of reducing the impact of shot noise and cosmic variance that the use of a small spectroscopic sample would have implied. We proved in Section 4.2 that the systematic error induced by *redMaGiC* photo-z is small compared to other source of systematics.

Despite statistical uncertainty being subdominant with respect to systematic errors, we note that the constant comoving density cut (together with luminosity threshold) used to select *redMaGiC* galaxies leads to large shot noise in the lowest redshift bins. We could select the *redMaGiC* sample imposing a lower luminosity threshold but a higher comoving density, so as to reduce shot noise. We therefore create a *combined redMaGiC* galaxy sample, made of three subsamples selected as follows: (1) high-density sample, $0.15 < z < 0.6$, $L > 0.5L*$; (2) high-luminosity sample, $0.6 < z < 0.75$, $L > L*$; (3) higher luminosity sample, $0.75 < z < 085$, $L > 1.5L*$. The latter corresponds to the sample used in the main analysis, but restricted to a smaller redshift interval.

Table A1. Total systematic error with *redMaGiC* combined sample as a reference. The table shows the total systematic error for the mean matching procedure, for the three WL source redshift bins and photo-z codes.

	Bin 1	Bin 2	Bin 3
BPZ	0.037	0.016	0.007
DNF	0.021	0.015	0.016

We repeated the full analysis for this new *redMaGiC* combined sample: results for the total systematic are summarized in Table A1. As compared to our fiducial analysis (Tables 1 and 2), we find larger systematics for the first WL source redshift bin and slightly smaller ones for the third bin. In general, lowering the luminosity threshold of the *redMaGiC* algorithm allows to select more galaxies, but at the same time, increases the photometric error (and the *redMaGiC* photo-z systematic). Moreover, being now the sample made of three subsamples each characterized by a different luminosity, we might expect a non-negligible bias evolution for the reference sample. The increase in the photometric error particularly affects the first bin (as it overlaps mainly with the high-density sample), and together with the stronger bias evolution, leads to a larger total systematic error. As for the third bin, the stronger bias evolution of *redMaGiC* cancels out with the bias evolution of the WL sample, reducing the bias evolution systematic and the total systematic error.

Given the larger impact of *redMaGiC* sample bias evolution and photometric errors on the total systematic budget, we preferred to use the higher luminosity sample for our analysis.

APPENDIX B: CORRECTING THE REDSHIFT EVOLUTION OF THE GALAXY–MATTER BIAS WITH AUTOCORRELATIONS WHEN SPECTROSCOPIC REDSHIFTS ARE NOT AVAILABLE

In Section 5.1 we showed that we could get rid of the bias evolution systematic within statistical errors if we could measure the autocorrelation functions of the two samples divided in thin top-hat redshift bins (i.e. using true-z). Unfortunately, it cannot be applied to data since we only have access to galaxies photo-z. Nonetheless, we could try to understand whether we could anyway correct for the redshift evolution of the galaxy–matter bias measuring the samples autocorrelation functions binned using photo-z.

In Fig. B1 we show what we would obtain if we binned the WL source samples using the one-point estimates of the photo-z codes and measure the autocorrelation functions. This is compared to the results shown in Section 5.1, where the autocorrelation functions are binned using galaxies true-z (quantities in the plot correspond to the one-bin version of the autocorrelation functions, averaged over angular scales as explained in Section 2.1).

Because of the poor quality of source galaxies photo-z, the measurements are completely different: not only they can span a different redshift range, but also the redshift dependence is completely dissimilar.

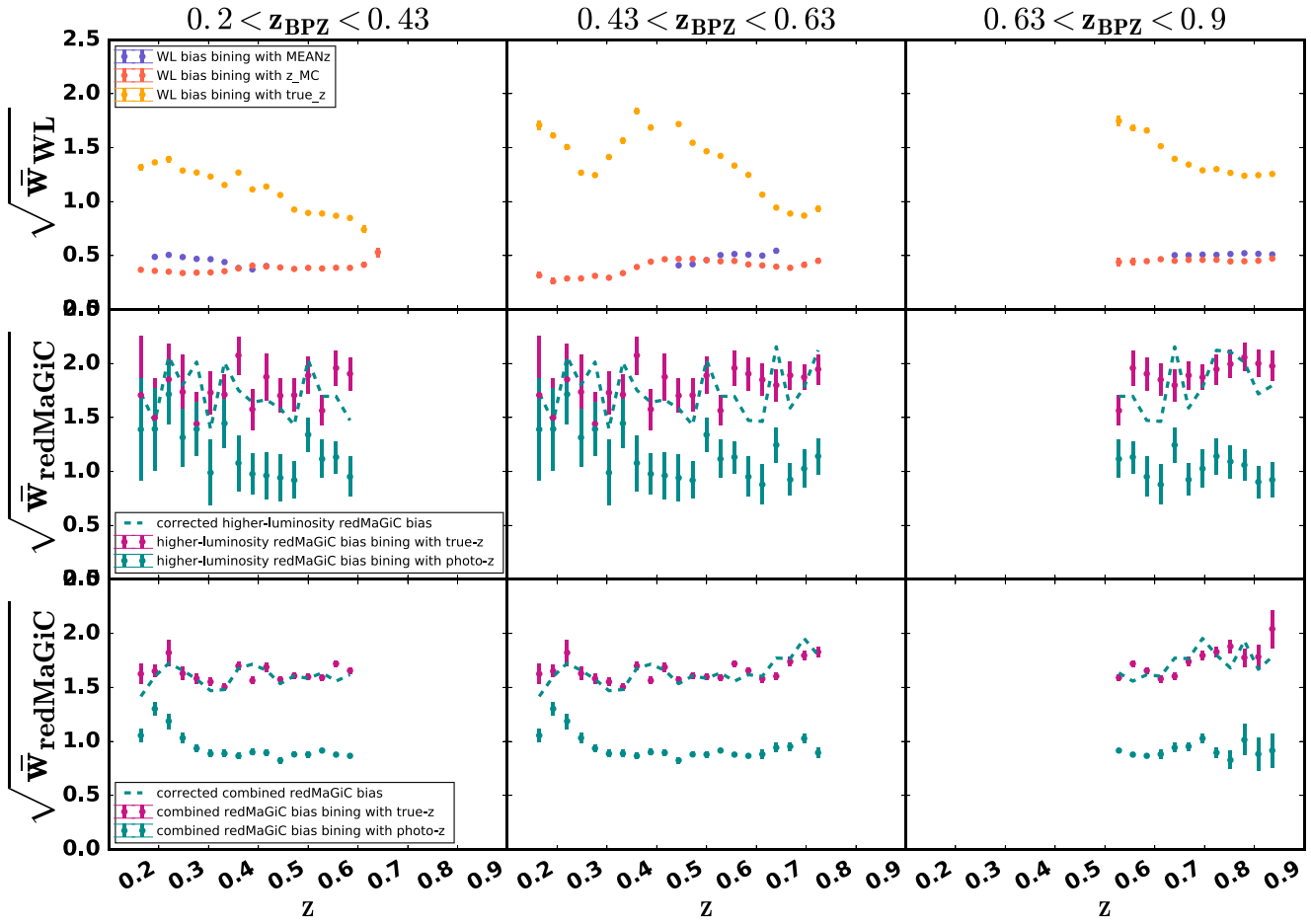

Figure B1. Top panels: redshift dependence of the one-bin estimates of the autocorrelation functions for the three WL source redshift bins. In yellow, the values obtained binning the samples using their true-z. In purple, the ones obtained binning the samples using their mean photo-z redshift (MEANz). In red, the values obtained binning the samples using a random draw from their photo-z posterior (z_{MC}). Middle and bottom panels: redshift dependence of the one-bin estimates of the autocorrelation functions for two *redMaGiC* samples – higher luminosity sample (middle panels) and combined sample (lower panels). For both samples, we display the autocorrelation functions obtained by binning the sample by true-z (pink) and photo-z (blue). We also display the corrected autocorrelation functions, computed starting from the estimates obtained with photo-z applying the corrections explained in Appendix B.

For the reference galaxies, the scenario is a bit different (see Fig. B1). In theory, *redMaGiC* galaxies have high quality, almost Gaussian photo-zs, and we could in principle try to relate the two measurements. This can be done as follows: starting from equation (1), the autocorrelation function included in the estimator proposed in Section 5.1 can be written as

$$w_{\rm RMG}(\theta) = \int {\rm d}z' b_{\rm r}^2(z') n_{\rm r}^2(z') w_{\rm DM}(\theta, z'), \tag{B1}$$

where $n_{\rm r}(z')$ is the redshift distribution of the *redMaGiC* galaxies in a given reference bin, and $b_{\rm r}(z')$ the reference sample galaxy–matter bias. If we assume the galaxy–matter bias (and the growth factor) to evolve as a function of redshift on scales larger than the reference bin width we can rewrite equation (B1) as

$$w_{\rm RMG}(\theta) = w_{\rm DM}(\theta, \langle z\rangle) b_{\rm r}^2(\langle z\rangle) \int {\rm d}z' n_{\rm r}^2(z'), \tag{B2}$$

where the quantities outside the integral are now computed at the median redshift $\langle z\rangle$ of the reference bin. This would allow us to relate the one-bin estimates of the *redMaGiC* autocorrelation functions computed binning by true-z and photo-z as follows:

$$\bar{w}_{{\rm RMG,spec}\text{-}z,\langle z\rangle} = \bar{w}_{{\rm RMG,photo}\text{-}z,\langle z\rangle} \frac{\int {\rm d}z' n^2_{{\rm r,spec}\text{-}z}(z')}{\int {\rm d}z' n^2_{{\rm r,photo}\text{-}z}(z')}. \tag{B3}$$

This correction requires knowledge of $n_{{\rm r,photo}\text{-}z}(z')$, which is the true distribution of the reference sample binned using *redMaGiC* photo-z. This is usually not available in data, but an estimate can be obtained looking at the subsample of *redMaGiC* galaxies with spectra.

In Fig. B1 one can see how $\bar{w}_{{\rm RMG,spec}\text{-}z,\langle z\rangle}$ is precisely recovered using this procedure. We note that this procedure cannot be applied to the WL sample: indeed, B3 is valid as long as we are able to select bins (using photo-z one-point estimates) thin enough so as to consider the bias constant over the whole bin width. This is not possible for the WL sample due to the poor quality of the photo-zs of the source galaxies. Increasing the bin size proved not to help in reconstructing the WL bias evolution.

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

[2]*Kavli Institute for Particle Astrophysics and Cosmology, PO Box 2450, Stanford University, Stanford, CA 94305, USA*
[3]*Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA*
[4]*Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstr. 1, D-81679 München, Germany*
[5]*Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA*
[6]*Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), E-28040 Madrid, Spain*
[7]*Institute of Space Sciences, IEEC-CSIC, Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain*
[8]*Department of Physics, University of Arizona, Tucson, AZ 85721, USA*
[9]*Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain*
[10]*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*
[11]*Laboratório Interinstitucional de e-Astronomia – LIneA, Rua Gal. José Cristino 77, 20921-400 Rio de Janeiro, RJ, Brazil*
[12]*Observatório Nacional, Rua Gal. José Cristino 77, 20921-400 Rio de Janeiro, RJ, Brazil*
[13]*SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA*
[14]*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK*
[15]*Department of Physics, ETH Zurich, Wolfgang-Pauli-Strasse 16, CH-8093 Zurich, Switzerland*
[16]*Fermi National Accelerator Laboratory, PO Box 500, Batavia, IL 60510, USA*
[17]*Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA*
[18]*Department of Physics, The Ohio State University, Columbus, OH 43210, USA*
[19]*ARC Centre of Excellence for All-sky Astrophysics (CAASTRO)*
[20]*School of Mathematics and Physics, University of Queensland, Brisbane, QLD 4072, Australia*
[21]*Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Victoria 3122, Australia*
[22]*Sydney Institute for Astronomy, School of Physics A28, The University of Sydney, NSW 2006, Australia*
[23]*Australian Astronomical Observatory, North Ryde, NSW 2113, Australia*
[24]*The Research School of Astronomy and Astrophysics, Australian National University, ACT 2601, Australia*
[25]*Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing, Jiangshu 210008, China*
[26]*Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, Casilla 603, La Serena, Chile*
[27]*LSST, 933 North Cherry Avenue, Tucson, AZ 85721, USA*
[28]*INAF – Osservatorio Astrofisico di Torino, via Osservatorio 20, I-10025 Pino Torinese, Italy*
[29]*Department of Astronomy, University of Illinois, 1002 W. Green Street, Urbana, IL 61801, USA*
[30]*National Center for Supercomputing Applications, 1205 West Clark St, Urbana, IL 61801, USA*
[31]*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*
[32]*Department of Physics, IIT Hyderabad, Kandi, Telangana 502285, India*
[33]*Department of Physics, California Institute of Technology, Pasadena, CA 91125, USA*
[34]*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr, Pasadena, CA 91109, USA*
[35]*Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA*
[36]*Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA*
[37]*Instituto de Fisica Teorica UAM/CSIC, Universidad Autonoma de Madrid, E-28049 Madrid, Spain*
[38]*Department of Astronomy, University of California, Berkeley, 501 Campbell Hall, Berkeley, CA 94720, USA*
[39]*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*
[40]*Astronomy Department, University of Washington, Box 351580, Seattle, WA 98195, USA*
[41]*Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA*
[42]*Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439, USA*
[43]*Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, CP 66318, 05314-970 São Paulo, SP, Brazil*
[44]*Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA*
[45]*Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK*
[46]*Brookhaven National Laboratory, Bldg 510, Upton, NY 11973, USA*
[47]*School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK*
[48]*Instituto de Física Gleb Wataghin, Universidade Estadual de Campinas, 13083-859 Campinas, SP, Brazil*
[49]*Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA*
[50]*Excellence Cluster Universe, Boltzmannstr. 2, D-85748 Garching, Germany*
[51]*Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, D-85748 Garching, Germany*

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