Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **511,** 2665–2687 (2022) https://doi.org/10.1093/mnras/stac104
Advance Access publication 2022 January 28

Dark Energy Survey Year 3 results: galaxy clustering and systematics treatment for lens galaxy samples

M. Rodríguez-Monroy[1★] N. Weaverdyck[2★] J. Elvin-Poole[3,4★] M. Crocce[5,6] A. Carnero Rosell[7,8,9] F. Andrade-Oliveira[8,10] S. Avila[11] K. Bechtol[12] G. M. Bernstein[13] J. Blazek[14,15] H. Camacho[8,10] R. Cawthon[12] J. De Vicente[1] J. DeRose[16] S. Dodelson[17,18] S. Everett[19] X. Fang[20] I. Ferrero[21] A. Ferté[22] O. Friedrich[23] E. Gaztanaga[5,6] G. Giannini[24] R. A. Gruendl[25,26] W. G. Hartley[27] K. Herner[28] E. M. Huff[22] M. Jarvis[13] E. Krause[20] N. MacCrann[29] J. Mena-Fernández[1] J. Muir[30] S. Pandey[13] Y. Park[31] A. Porredon[3,4] J. Prat[32,33] R. Rosenfeld[8,34] A. J. Ross[3] E. Rozo[35] E. S. Rykoff[30,36] E. Sanchez[1] D. Sanchez Cid[1] I. Sevilla-Noarbe[1] M. Tabbutt[12] C. To[30,36,37] E. L. Wagoner[35] R. H. Wechsler[30,36,37] M. Aguena[8] S. Allam[28] A. Amon[30] J. Annis[28] D. Bacon[38] E. Baxter[39] E. Bertin[40,41] S. Bhargava[42] D. Brooks[43] D. L. Burke[30,36] M. Carrasco Kind[25,26] J. Carretero[24] F. J. Castander[5,6] A. Choi[3] C. Conselice[44,45] M. Costanzi[46,47,48] L. N. da Costa[8,49] M. E. S. Pereira[2] S. Desai[50] H. T. Diehl[28] B. Flaugher[28] P. Fosalba[5,6] J. Frieman[28,33] J. García-Bellido[11] T. Giannantonio[51,23] D. Gruen[30,36,37] J. Gschwend[8,49] G. Gutierrez[28] S. R. Hinton[52] D. L. Hollowood[19] K. Honscheid[3,4] D. Huterer[2] B. Jain[13] D. J. James[53] K. Kuehn[54,55] N. Kuropatkin[28] M. Lima[8,56] M. A. G. Maia[8,49] M. March[13] J. L. Marshall[57] P. Melchior[58] F. Menanteau[25,26] C. J. Miller[59,2] R. Miquel[24,60] J. J. Mohr[61,62] R. Morgan[12] A. Palmese[28,33] F. Paz-Chinchón[25,51] A. Pieres[8,49] A. A. Plazas Malagón[58] A. Roodman[30,36] V. Scarpine[28] S. Serrano[5,6] M. Smith[63] M. Soares-Santos[2] E. Suchyta[64] G. Tarle[2] D. Thomas[38] and T. N. Varga[62,65] (DES Collaboration)

Affiliations are listed at the end of the paper

Accepted 2021 December 14. Received 2021 October 25; in original form 2021 June 9

ABSTRACT

In this work, we present the galaxy clustering measurements of the two DES lens galaxy samples: a magnitude-limited sample optimized for the measurement of cosmological parameters, MAGLIM, and a sample of luminous red galaxies selected with the REDMAGIC algorithm. MAGLIM/REDMAGIC sample contains over 10 million/2.5 million galaxies and is divided into six/five photometric redshift bins spanning the range $z \in [0.20, 1.05]/z \in [0.15, 0.90]$. Both samples cover 4143 deg^2 over which we perform our analysis blind, measuring the angular correlation function with an S/N $\sim$ 63 for both samples. In a companion paper, these measurements of galaxy clustering are combined with the correlation functions of cosmic shear and galaxy–galaxy lensing of each sample to place cosmological constraints with a 3 $\times$ 2pt analysis. We conduct a thorough study of the mitigation of systematic effects caused by the spatially varying survey properties and we correct the measurements to remove artificial clustering signals. We employ several decontamination methods with different configurations to ensure the robustness of our corrections and to determine the systematic uncertainty that needs to be considered for the final cosmology analyses. We validate our fiducial methodology using lognormal mocks, showing that our decontamination procedure induces biases no greater than 0.5σ in the ($\Omega_{\rm m}$, b) plane, where b is the galaxy bias.

Key words: cosmological parameters – cosmology: observations – dark energy – large-scale structure of the Universe.

1 INTRODUCTION

The current Standard Model of Cosmology, ΛCDM, provides an excellent fit to current observations, including distance measurements to Type Ia supernovae (SN Ia; Riess et al. 1998; Perlmutter et al. 1999), the cosmic microwave background (CMB) fluctuations (Spergel et al. 2003; Planck Collaboration XI 2020), and the large-scale structure of the Universe (Alam et al. 2017; Abbott et al. 2019; Alam et al. 2021), with only six free parameters. In addition, photometric galaxy surveys, such as the Kilo-Degree Survey (KiDS, de Jong et al. 2013), Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP; Aihara et al. 2018) and the Dark Energy Survey (DES,

★ E-mail: rodriguez-monroy@ijclab.in2p3.fr (MR); nweaverd@umich.edu (NW); elvin-poole.1@osu.edu (JE)

The Dark Energy Survey Collaboration 2005) are now reaching a level of sensitivity that competes and complements the precise determinations from CMB observatories. The comparison of the measurements of the late Universe, provided by galaxy surveys, and the early Universe, provided by CMB measurements, allows for powerful tests of the nature of cosmic acceleration and general relativity. The precision that photometric surveys are able to reach in the determination of cosmological parameters comes from the combination of different observables, mainly from weak lensing and clustering of galaxies, in the so-called 3 × 2pt analysis, whose methodology is described in DES Collaboration (2018a), van Uitert et al. (2018), Joudaki et al. (2018), Heymans et al. (2021), and Krause et al. (2021) (and references therein).

In this work, we present the clustering measurements of the lens galaxy samples that enter in the DES Year 3 (Y3) 3 × 2pt (DES Collaboration et al. 2022) and the 2 × 2pt (Elvin-Poole et al. 2021; Pandey et al. 2021; Porredon et al. 2021a; Prat et al. 2021, in combination with the shear field or galaxy–galaxy lensing) analyses. The cosmological information is extracted from the large-scale structure (LSS) measurements using the angular two-point correlation function that characterizes the spatial distribution of galaxies in tomographic photometric redshift bins. However, the measurement of the angular correlation function is affected by spatially varying survey properties that must be taken into account and corrected to extract the full cosmological power of DES. These systematic effects come from the observing conditions and translate into changes in the selection function across the observed footprint or with redshift.

As photometric surveys have become more extended in area, both the impact of these survey properties or observational effects, and the diminishing statistical errors, have spurred the development of a variety of techniques to correct for them in clustering measurements. Already in SDSS (Scranton et al. 2002; Myers et al. 2006) and 2MASS (Maller et al. 2005), cross-correlations with different survey properties and masking were used to check for possible sources of systematic error, which were deemed to be insignificant given the statistical errors. Ross et al. (2011) compared several methodologies (masking, cross-correlation correction and computing weights for the data) in SDSS-III. The cross-correlation correction method was applied to early DES data (DES-SV) in Crocce et al. (2016), and was studied by Elsner, Leistedt & Peiris (2016) (there called 'template subtraction') who derived its characteristic bias. The application of weights have increasingly become a popular method, applied for instance in BOSS (Ross et al. 2017, 2020), eBOSS (Laurent et al. 2017; Raichoor et al. 2021), DES-SV (Kwan et al. 2017, comparing with the cross-correlation method), DES Y1 data (Elvin-Poole et al. 2018), and DESI targets (Kitanidis et al. 2020). Rather than applying weights to the observed data, the inverse-weights can be applied to the random sample used for correlation function analyses, as shown in Morrison & Hildebrandt (2015) and applied to eBOSS data via a multilinear regression analysis in (Bautista et al. 2018; Icaza-Lizaola et al. 2020). These approaches have been refined in recent years as the importance of addressing these spatial systematics has grown (Vakili et al. 2020; Wagoner et al. 2021; Weaverdyck & Huterer 2021), including the development of machine learning approaches using neural networks (NNs; Rezaie et al. 2020) or self-organizing maps (Johnston et al. 2021). Some approaches have operated only at the level of the power spectrum, including mode projection methods (Rybicki & Press (1992) with examples of applications and further developments shown in Leistedt et al. (2013), Leistedt & Peiris (2014), Elsner et al. (2016), and Elsner, Leistedt & Peiris (2017). Weaverdyck & Huterer (2021) reviewed several of the above techniques and showed how mode projection methods operating on the pseudo-power spectrum are related to multilinear regression methods, identifying residual biases in both approaches.

We present the methods we apply to DES-Y3 data in order to mitigate these effects, the full set of validation tests we perform, both on data and on simulations, and its final implementation on the data. These corrections enable robust measurements of the clustering amplitude of lens galaxies. The results of this analysis are used as the clustering input for the full 3 × 2pt cosmological analysis in DES-Y3 (DES Collaboration et al. 2022).

This paper is organized as follows: In Section 2, we describe the modeling of the galaxy clustering angular correlation function used throughout the Y3 analysis. In Section 3, we introduce the Y3 data and the galaxy samples derived from it. In Section 4, we present the description of different observing conditions and their representation. In Section 5, we present the methodology, with special attention to the decontamination pipeline (Sections 5.3.1 and 5.3.2). In Section 6, we show the galaxy clustering results after applying the correction methods. This correction is validated in Section 7. In Section 8, we discuss the post-unblinding findings about the amplitude of the angular correlation functions in terms of the considered survey properties. Finally, we present the conclusions in Section 9.

2 MODELLING

The observed projected galaxy density contrast $\delta^i_{\rm obs}(\hat{\boldsymbol{n}})$ of galaxies in tomography bin i at position $\hat{\boldsymbol{n}}$ can be written as

$$\delta^i_{\rm g,obs}(\hat{\boldsymbol{n}}) = \underbrace{\int {\rm d}\chi\, W^i_\delta(\chi)\, \delta^{\rm (3D)}_{\rm g}(\hat{\boldsymbol{n}}\chi, \chi)}_{\delta^i_{\rm g,D}(\hat{\boldsymbol{n}})} + \delta^i_{g,\rm RSD}(\hat{\boldsymbol{n}}) + \delta^i_{{\rm g},\mu}(\hat{\boldsymbol{n}}), \quad (1)$$

with χ the comoving distance, $W^i_\delta = n^i_{\rm g}(z)\,{\rm d}z/{\rm d}\chi$ the normalized selection function of galaxies in tomographic bin i. Here the first term is the line-of-sight projection of the three-dimensional galaxy density contrast, $\delta^{\rm (3D)}_{\rm g}$; the remaining terms are the contributions from linear redshift-space distortions (RSDs) and magnification (μ), which are described in Krause et al. (2021).

We model the galaxy density assuming a local, linear galaxy bias model (Fry & Gaztanaga 1993), where the galaxy and matter density fluctuations are related by $\delta_{\rm g}(\boldsymbol{x}) = b\delta_{\rm m}(\boldsymbol{x})$, with density fluctuations defined by $\delta \equiv (n(\boldsymbol{x}) - \bar{n})/\bar{n}$. We model the linear galaxy bias to be constant across each tomographic bin, denoted as b^i. The validity of these assumptions to the accuracy of the Y3 3 × 2pt analysis is demonstrated in Krause et al. (2021) (see section V.B.2 and also DeRose et al. 2021, where it is determined that the redshift evolution of linear galaxy bias within redshift bins is negligible for the clustering and galaxy–galaxy lensing combined analyses).

The angular power spectrum consists of six different terms, corresponding to auto- and cross-power spectra of galaxy density, RSD and magnification. For Y3, we use the exact (non-Limber) computation for angular clustering. For a quantitative analysis of the impact of the Limber approximation on near-future data sets, see Fang et al. (2020). For example, the exact expression for the density–density contribution to the angular clustering power spectrum is

$$C^{ij}_{\delta_{\rm g,D}\delta_{\rm g,D}}(\ell) = \frac{2}{\pi}\int {\rm d}\chi_1\, W^i_\delta(\chi_1) \int {\rm d}\chi_2\, W^j_\delta(\chi_2) \times \int \frac{{\rm d}k}{k} k^3 P_{gg}(k, \chi_1, \chi_2) j_\ell(k\chi_1) j_\ell(k\chi_2), \quad (2)$$

with $j_\ell(k\chi)$ Bessel functions of order ℓ and $P_{\rm gg}(k, z_1, z_2)$ the 3D galaxy power spectrum in real space. The full expressions including magnification and RSD are given in Fang et al. (2020). Schematically, the integrand in equation (2) is split into the contribution from non-linear evolution, for which unequal time contributions are negligible so that the Limber approximation is sufficient, and the linear-evolution power spectrum, for which time evolution factorizes.[1]

The angular correlation function is then given by

$$w^i(\theta) = \sum_\ell \frac{2\ell+1}{4\pi} P_\ell(\cos\theta) C^{ii}_{\delta_{\rm g,obs}\delta_{\rm g,obs}}(\ell)\,, \tag{3}$$

where P_ℓ are the Legendre polynomials.

Throughout this paper, we use the COSMOSIS framework[2] (Zuntz et al. 2015) to compute correlation functions, and to infer cosmological parameters. The evolution of linear density fluctuations is obtained using the CAMB (Lewis & Bridle 2002) module[3] and then converted to a non-linear matter power spectrum $P_{NL}(k)$ using the updated HALOFIT recipe (Takahashi et al. 2012). Nevertheless, the baseline model used for this analysis assumes linear galaxy bias, so the relation between galaxy and matter power spectra is given by $P_{\rm gg} = b^2 P_{\rm mm}(k)$ (see Krause et al. 2021, for a more general expression and for the validation of this model).

We model (and marginalize over) photometric redshift bias uncertainties as an additive shift Δz^i in the galaxy redshift distribution $n^i_{\rm g}(z)$ for each redshift bin i:

$$n^i_{\rm g}(z) \to n^i_{\rm g}(z - \Delta z^i), \tag{4}$$

and a stretch parameter to characterize the uncertainty on the width for some of the tomographic bins and samples:

$$n^i_{\rm g}(z) \to n^i_{\rm g}\left(\sigma^i_z[z - \langle z\rangle] + \langle z\rangle\right). \tag{5}$$

The priors on the Δz^i and σz^i nuisance parameters are measured and calibrated directly using the angular cross-correlation between the DES sample and a spectroscopic sample, as described in Cawthon et al. (2020). We use the same Δz^i and σz^i as in the Y3 3 × 2pt analysis for all tests of robustness of the parameter constraints, as listed in Table 3.

3 DATA

The Dark Energy Survey collected imaging data with the Dark Energy Camera (DECam; Flaugher et al. 2015) mounted on the Blanco 4m telescope at the Cerro Tololo Inter-American Observatory (CTIO) in Chile during six years, from 2013 to 2019. The observed sky area covers ∼5000 deg^2 in five broad-band filters, *grizY*, covering near-infrared and visible wavelengths. This work uses data from the the first 3 yr (from 2013 August to 2016 February), with approximately four overlapping exposures over the full wide-field area, reaching a limiting magnitude of $i \sim 23.3$ for signal-to-noise ratio (S/N) = 10 point sources. The data were processed by the DES Data Management system (Morganson et al. 2018) and, after a complex reduction and vetting procedure, compiled into object catalogues. The catalogue used here amounts to nearly 400 million sources (available publicly as Data Release 1[4]; DES Collaboration 2018b). We calculate additional metadata in the form of quality flags, survey flags, survey property (SP) maps, object classifiers, and photometric redshifts to build the Y3 GOLD data set (Sevilla-Noarbe et al. 2021).

From this catalogue, we build the different galaxy samples for LSS studies. For robustness, we decided to use two different types of lens galaxies, MAGLIM and REDMAGIC, which are used as lens samples for galaxy clustering and for combination with weak lensing for the 3 × 2pt analysis. These two samples are described in the following subsections.[5]

3.1 Y3 MAGLIM sample

The main lens sample considered in this work, MAGLIM, is the result of the optimization carried out in Porredon et al. (2021b). The sample is designed to maximize the cosmological constraining power of the combined clustering and galaxy–galaxy lensing analysis (also known as 2 × 2pt) keeping the selection criterion as simple as possible. The selection cuts, based on the table columns from Sevilla-Noarbe et al. (2021), are as follows:

(i) flags_foreground=0 & flags_footprint=1 & bitand(flags_badregions,2)=0 & bitand(flags_gold,126)=0;
(ii) star-galaxy separation with EXTENDED_CLASS_MASH_SOF = 3;
(iii) $i < 4 \cdot z_{\rm phot} + 18$;
(iv) $i > 17.5$;

The first cut is a quality flag to remove badly measured objects or objects with issues in the processing steps. It also removes problematic regions due to astrophysical foregrounds. The second cut removes stars from the galaxy sample. The faint magnitude cut in the *i* band depends linearly on the photometric redshift, $z_{\rm phot}$, and selects bright galaxies. The photometric redshift estimator used for this sample is the Directional Neighbourhood Fitting (DNF; De Vicente, Sánchez & Sevilla-Noarbe 2016) algorithm (see also Porredon et al. 2021a), in particular its mean estimate using 80 nearest neighbours in colour and magnitude space, by performing a hyperplane fit. The brighter magnitude cut removes residual stellar contamination from binary stars and other bright objects.

The number and width of the redshift bins is studied in Porredon et al. (2021b), where they evaluate the impact of this kind of choices on the 2 × 2pt constraining power in wCDM (Fisher forecasts and MCMC sampling of the posterior distributions of $\Omega_{\rm m}$, σ_8, and w). We split the sample into six tomographic lens bins, with bin edges $z_{\rm phot} = [0.20, 0.40, 0.55, 0.70, 0.85, 0.95, 1.05]$. These edges have been slightly modified with respect to Porredon et al. (2021b) in order to improve the photometric redshift calibration (De Vicente et al. 2016). We refer the reader to Porredon et al. (2021b) for more details about the optimization of this sample and its comparison with REDMAGIC and other flux-limited samples. The main properties of the sample are summarized at the top panel of Table 1.

3.2 Y3 REDMAGIC sample

The REDMAGIC algorithm selects luminous red galaxies (LRGs) according to the magnitude–colour–redshift relation of red sequence galaxy clusters, calibrated using an overlapping spectroscopic sample. This sample is defined by an input threshold luminosity $L_{\rm min}$

[1] https://github.com/xfangcosmo/FFTLog-and-beyond.
[2] https://bitbucket.org/joezuntz/cosmosis.
[3] http://camb.info.
[4] https://des.ncsa.illinois.edu/releases/dr1.

[5] Moreover, from Y3 GOLD. we also define the BAO SAMPLE, a galaxy sample especially defined for studies on the baryonic acoustic oscillation scales (Carnero Rosell et al. 2021), which is not used here, but undergoes an analogous treatment of its spatial systematics.

Table 1. MAGLIM (top table) and REDMAGIC (bottom table) characterization parameters: number of galaxies, N_g, and number density, $\langle n_g \rangle$, blind galaxy bias, b^i, and scales excluded per redshift bin.

Redshift bin	N_g	$\langle n_g \rangle$	b^i	$\theta >$ [arcmin]
		MAGLIM		
$0.20 < z < 0.40$	2236 462	0.150	1.5	33.88
$0.40 < z < 0.55$	1599 487	0.107	1.8	24.35
$0.55 < z < 0.70$	1627 408	0.109	1.8	17.41
$0.70 < z < 0.85$	2175 171	0.146	1.9	14.49
$0.85 < z < 0.95$	1583 679	0.106	2.3	12.88
$0.95 < z < 1.05$	1494 243	0.100	2.3	12.06
		REDMAGIC		
$0.15 < z < 0.35$	330 243	0.022	1.7	39.23
$0.35 < z < 0.50$	571 551	0.038	1.7	24.75
$0.50 < z < 0.65$	872 611	0.059	1.7	19.66
$0.65 < z < 0.80$	442 302	0.030	2.0	15.62
$0.80 < z < 0.90$	377 329	0.025	2.0	12.40

Notes. The number densities are in units of arcmin^{-2} and the scales excluded correspond to 8 Mpc h^{-1} for both samples, as described in Krause et al. (2021). The blind galaxy bias values correspond to the fiducial values that were assumed to create the lognormal mocks used in this analysis, not the best-fitting values from 3 × 2pt.

and constant co-moving density. The full REDMAGIC algorithm is described in Rozo et al. (2016). REDMAGIC is the algorithm used for the fiducial clustering sample of the DES Y1 3 × 2pt cosmology analyses (DES Collaboration 2018a; Elvin-Poole et al. 2018), with some updates improving the redshift estimates and selection uniformity, besides the usage of new photometry from Y3 GOLD.

We define the Y3 REDMAGIC sample in five tomographic lens bins, selected on the REDMAGIC redshift point estimate quantity zredmagic. The bin edges used are $z_{\rm REDMAGIC} = [0.15, 0.35, 0.50, 0.65, 0.80, 0.90]$. The first three bins use a luminosity threshold of $L_{\rm min} > 0.5L_*$ and are known as the high density or *highdens* sample. The last two redshift bins use a luminosity threshold of $L_{\rm min} > 1.0L_*$ and are known as the high luminosity or *highlum* sample.

The REDMAGIC selection also includes the following cuts on quantities from the Y3 GOLD catalogue and REDMAGIC calibration:

(i) Removed objects with FLAGS_GOLD in 8|16|32|64.
(ii) Star galaxy separation with EXTENDED_CLASS_MASH_SOF ≥2.
(iii) Cut on the red-sequence goodness of fit $\chi^2 < \chi^2_{\rm max}(z)$.

The main properties of the sample are summarized in the bottom part of Table 1. See Sevilla-Noarbe et al. (2021) for further details on these quantities.

3.3 Angular mask

The total sky area covered by the Y3 GOLD catalogue footprint is 4946 deg^2. We then mask regions where astrophysical foregrounds (bright stars or large nearby galaxies) are present, or where there are known processing problems ('bad regions'), reducing the total area by 659.68 deg^2 (Sevilla-Noarbe et al. 2021). The angular mask is defined as a HEALPIX[6] (Górski et al. 2005) map of resolution $N_{\rm side} = 4096$. Pixels with fractional coverage smaller than 80 per cent are removed. In addition, we require homogeneous depth across the footprint for both galaxy samples, removing too shallow or incomplete regions. As a summary, we use the following Y3 GOLD and REDMAGIC specific map quantities to define the final common area:

(i) footprint = 1;
(ii) foregrounds = 0;
(iii) badregions ≤1;
(iv) fracdet > 0.8;
(v) depth *i* band ≥22.2;
(vi) $z_{\rm MAX,\ highdens} \geq 0.65$;
(vii) $z_{\rm MAX,\ highlum} \geq 0.95$.

where the depth for the *i*-band magnitude is obtained using the SOF photometry (detailed in Sevilla-Noarbe et al. 2021) (as used in MAGLIM) and the conditions on ZMAX are inherited from the REDMAGIC redshift span. The $z_{\rm MAX}$ quantity is the maximum redshift at which a REDMAGIC galaxy can be detected with the luminosity threshold employed ($0.5L^*$ for REDMAGIC *highdens* and $1.0L^*$ for *highlum*), given the depth of the survey at that location. We only remove pixels where $z_{\rm MAX}$ is lower than the upper edge of the redshift bin, so we have $Z_{\rm MAX,\ highdens} < 0.65$ and $z_{\rm MAX,\ highlum} < 0.95$ for this sample. Initially (well before unblinding), we considered an upper redshift bin edge of 0.95, but this was reduced to 0.90 due to poor coverage in the spectroscopic sample employed for validation (Cawthon et al. 2020). However, we decided to keep the definition of the angular mask at $Z_{\rm MAX,\ highlum} \geq 0.95$ for compatibility with other parts of the DES Y3 analysis for which the angular mask was already fixed. This is also a more conservative cut, since it removes pixels at the edge of the depth cut. The final analysed sky area is 4143 deg^2.

4 SURVEY PROPERTIES

4.1 SP maps

Through their impact on the galaxy selection function, survey properties can modify the observed galaxy density field. In order to correct these effects, we develop spatial templates for potential contaminants by creating HEALPIX sky maps of survey properties ('SP maps'), which we then use to characterize and remove contamination from the observed density fields (see Leistedt et al. 2016, for the details of the original implementation of this mapping in DES). Each pixel of a given SP map corresponds to a summary statistic that characterizes the distribution of values of the measured quantity over multiple observations. Table 2 summarizes the survey properties considered in this analysis along with the summary statistics used to produce the SP maps. As foreground sources of contamination we use a star map created with bright DES point sources, labeled *stellar_dens* (the star map from which the stellar density map is made, *stars_1620*, has the cut $16 < i < 20$), and the interstellar extinction map from Schlegel, Finkbeiner & Davis (1998), *sfd98*.[7] More detailed information on the construction of these maps can be found in Sevilla-Noarbe et al. (2021). Hereafter, we will use SP map to refer to SP and foreground maps generically.

[6] https://healpix.sourceforge.io.

[7] We have verified that substituting the DES point sources map with the Gaia EDR3 star map (Gaia Collaboration 2020) and the *sfd98* map with the Planck 2013 thermal dust emission map (Planck CollaborationVI 2014) has no significant impact on the results.

Table 2. Survey properties used for the systematics mitigation effort of the DES Y3 Key Project, along with their physical units and the statistics used to generate SP maps from the stacking of images.

Quantity	Units	Statistics
airmass	∅	WMEAN, MIN, MAX
fwhm	arcsec	WMEAN, MIN, MAX
fwhm_fluxrad	arcsec	WMEAN, MIN, MAX
exptime	s	SUM
t_eff	∅	WMEAN, MIN, MAX
t_eff_exptime	s	SUM
skybrite	electrons/CCD pixel	WMEAN
skyvar	(electron s/CCD pixel)^2	WMEAN, MIN, MAX
skyvar_sqrt	electrons/CCD pixel	WMEAN
skyvar_uncertainty	electrons/ s × coadd pixel	
sigma_mag_zero	mag	QSUM
fgcm_gry	mag	WMEAN, MIN
maglim	mag	
sof_depth	mag	
magauto_depth	mag	
stars_1620	# stars	
stellar_dens	stars/deg^2	
sfd98	mag	

Notes. As foreground sources of contamination, we use a DES bright stars map and the dust extinction map from Schlegel et al. (1998). We use both the raw number count of DES point sources, *stars_1620*, and the density, *stellar_dens*. We use an SP map for each statistic in each photometric band in {*g*, *r*, *i*, *z*} (with the exception of *stars_1620*, *stellar_dens*, and *sfd98*), resulting in 107 total SP maps.

4.2 Reduced PCA map basis

The Y1 analysis used 21 SP maps selected a priori. However, a reduced set of SP maps is equivalent to setting a hard prior of no contamination from those SP maps that are unused, so we should be careful to not discard spatial templates that carry unique information about potential systematics (Weaverdyck & Huterer 2021). For Y3 we have initially increased the number of SP maps considered to 107. By expanding the library of SP maps used for cleaning, we relax the implicit priors and adopt a more data-driven approach to cleaning observational systematics from the clustering data.

Many of the Y3 additional SP maps we use are alternative summary statistics for characterizing the observed quantity, such as MIN and MAX instead of the weighted mean (WMEAN), which results in a high correlation between SP maps. We therefore create an orthogonal set of SP maps by using the principal components of the pixel covariance matrix across all 107 SP maps (standardized to zero mean and unit variance) at $N_{\rm side} = 4096$.[8] This provides an orthornormal basis set of SP maps that can be ordered according to the total variance they capture in the space spanned by the 107 SP maps. We will refer to these principal component maps as PC maps to differentiate from SP maps in the standard (STD) basis, where each map represents a single SP (e.g. *exptime*). From this point forward, we will use 'SP' map to more generically refer to maps that may be in either the PC or STD basis. We retain the first 50 PC maps, which account for ∼98 per cent of the variance of the full 107 map basis. This allows us to capture the dominant features of the additional maps while reducing the risk of removing real LSS signal from overfitting (we note that we use PCA to decorrelate the STD maps and after that we find it convenient to reduce their number, so we employ this same formalism for this task). We test the impact of adjusting the number of PC maps used in Section 8 and Appendix D, finding that the full set of 107 maps results in galaxy weights that overcorrect and correlate significantly with LSS. The fiducial set of maps employed to decontaminate the data are these first 50 PC maps, although we have also run validation tests with the STD maps, as we explain in the next sections.

[8]We use this resolution because we wish to apply the correction to the data at the maximum resolution available. This is provided by DESDM (Morganson et al. 2018), which generates the SP maps at $N_{\rm side} = 4096$, which is a good compromise between computational speed and needed resolution for this cosmological analysis. We verify that the difference in the variance explained by A) the principal component maps at $N_{\rm side} = 4096$ and then degraded to 512 and B) the maps obtained after performing a PCA at $N_{\rm side} = 512$ is negligible.

5 ANALYSIS TOOLS AND METHODOLOGY

5.1 Clustering estimator

The analysis of the galaxy clustering is performed by measuring the angular two-point correlation function, $w(\theta)$, in photometric redshift bins. In this analysis, we work with HEALPIX (Górski et al. 2005) maps of the SPs and galaxy density from lognormal mock catalogues. The decontamination methods generate HEALPIX weight maps as well. Weights are actually obtained for each SP pixel, so we also work with pixelized versions of our galaxy samples, and use a pixel-based version of the Landy–Szalay estimator (Landy & Szalay 1993), following the notation of Crocce et al. (2016):

$$\hat{w}(\theta) = \sum_{i=1}^{N_{\rm pix}} \sum_{j=1}^{N_{\rm pix}} \frac{(N_i - \bar{N}) \times (N_j - \bar{N})}{\bar{N}^2} \Theta_{i,\,j}, \tag{6}$$

where N_i is the galaxy number density in pixel i, $\bar{N}$ is the mean galaxy number density over all pixels within the footprint, and $\Theta_{i,\,j}$ is a top-hat function that is equal to 1 when pixels i and j are separated by an angle θ within the bin size $\Delta\theta$. The fractional coverage of each pixel is taken into account in the calculation of N_i and $\bar{N}$. These correlation functions are calculated using TREECORR[9] (Jarvis, Bernstein & Jain 2004). We verify on the data that the difference between this pixel version of the estimator at both $N_{\rm side} = 4096$ and 512 and that using random points is negligible for the angular scales we consider.

5.2 Lognormal mocks

We rely on a set of lognormal mock realizations of the observed data to evaluate the significance of the correlation between data and SP maps following the methodology of Elvin-Poole et al. (2018) and Xavier, Abdalla & Joachimi (2016). For each of our galaxy samples, we create a set of 1000 mocks that matches their mean galaxy number density and power spectrum. We generate full-sky mock catalogues at a HEALPIX resolution of $N_{\rm side} = 512$, corresponding to ∼0.°11 pixels. We then apply the DES-Y3 angular mask. This angular resolution is small enough to be used for the scales employed in the cosmology analysis. The usage of these mocks is covered in Section 5.3.1. We also create sets of contaminated lognormal mocks that we later use to validate our decontamination methods. These mocks incorporate the effect of SP maps observed on the data. Appendix A contains more details about their creation and contamination.

[9]https://rmjarvis.github.io/TreeCorr.

5.3 Correction methods

The observed galaxy sample has contamination from observing conditions and foregrounds, which modify the selection function across the survey footprint. Our goal is to correct these effects in the lens galaxy samples. To do so, we create a set of weights to apply to the galaxy samples, constructed from a list of SP maps. The weighted sample is then used for measurements of $w(\theta)$ and for combination with weak lensing measurements (DES Collaboration et al. 2022; Elvin-Poole et al. 2021; Pandey et al. 2021; Porredon et al. 2021a). This approach has been successfully applied to the angular correlation function of the DES Year 1 clustering measurements (Elvin-Poole et al. 2018), as well as in SDSS-III (e.g. in Ross et al. 2011, 2017), eBOSS (Laurent et al. 2017; Bautista et al. 2018; Icaza-Lizaola et al. 2020; Ross et al. 2020; Raichoor et al. 2021), and KiDS (Vakili et al. 2020).

Most correction procedures can be interpreted as regression methods, with the true overdensity field corresponding to the residuals after regressing the observed density field against a set of SP maps. Adding SP maps is equivalent to adding additional explanatory variables to the regression, which increases the chance of overfitting. Such overfitting will reduce the magnitude of the inferred overdensity field (i.e. shrink the size of regression residuals), and thus overfitting will generically lead to a reduced clustering signal.

There are several approaches to address this. One can a priori restrict the number of SP maps to reduce the level of false correction. This is equivalent to asserting that there is no contamination from the discarded SP maps, which risks biasing the data from unaccounted-for systematic effects. A second option is to clean with all of the SP maps and then debias the measured clustering based on an estimate of the expected level of false correction (e.g. pseudo-C_ℓ mode projection; Elsner et al. 2016, 2017; Alonso et al. 2019). This approach can be interpreted as a simultaneous ordinary least-squares regression with a step to debias the power spectrum. Map-level weights that may enter in the analysis of other observables, such as galaxy–galaxy lensing, can be produced from this approach, but they will be overly aggressive if the number of SP maps is large. Wagoner et al. (2021) extend this approach by incorporating the pixel covariance and using Markov chain Monte Carlo to include map-level error estimates, but this again becomes less feasible if the number of SP maps is too large. Finally, one can take an approach between these extremes, reducing the number of SP maps used for fitting, but doing so in a data-driven manner. We apply two different methods that take this third approach. They make different assumptions, but were both found to perform well in simulated tests in Weaverdyck & Huterer (2021). The SP maps we run these two methods on is our fiducial set of 50 PC maps that we introduced in Section 4. In addition, we present a third method that we use to test linearity assumptions made by the other two.

5.3.1 Iterative Systematics Decontamination (ISD)

In this subsection, we describe the fiducial correction method that we use for DES Y3, called ISD. It is an extension of the methodology applied in Y1 (Elvin-Poole et al. 2018).

ISD is organized as a pipeline that corrects the PC map (or any generic SP map) effects by means of an iterative process whose steps can be summarized as (i) identify the most significant PC map, (ii) obtain a weight map from it, (iii) apply it to the data, and (iv) go back to (i). The algorithm stops when there are no more maps with an effect larger than an a priori fixed threshold. Each step is described in more detail in the following lines.

To begin with, we degrade each PC map to $N_{\rm side}= 512$ and then we compute the relation between their values and $n_{\rm o}/\langle n_{\rm o}\rangle$, where $n_{\rm o}$ is the observed density of galaxies at a given part of the sky and $\langle n_{\rm o}\rangle$ is the average density over the full footprint. In the following, we refer to this as the 1D relation. To obtain the statistical significance of the observed correlations, we bin the 1D relation into 10 equal-sky areas for each PC map and estimate a covariance matrix for the 1D relation bin means of that PC map using the set of 1000 uncontaminated mocks described in Section 5.2. Since the bins are defined as equal area, the statistical error associated with each bin is similar and no one region dominates the fit. We use this covariance matrix for determining the best-fitting parameters of a function to approximate the 1D relation, as well as to assess its goodness of fit.

We fit the 1D relation to a linear function of the PC map values

$$\frac{n_{\rm o,i}}{\langle n_{\rm o}\rangle} = m \times s_i + c, \tag{7}$$

by minimizing χ^2, which we then denote $\chi^2_{\rm model}$. The index i runs over the PC map bins. Similarly, we compute the goodness of fit for the case where $n_{\rm o}/\langle n_{\rm o}\rangle$ is a constant function $f(s) = 1$ labelled $\chi^2_{\rm null}$. Finding that $n_{\rm o}/\langle n_{\rm o}\rangle$ fits well to this constant function is equivalent to finding that this particular PC has no impact on the galaxy density field. To calculate both χ^2 definitions, we make use of the (10×10) covariance matrix obtained from the lognormal mocks.

The degree of impact of a given PC map on the data is evaluated using

$$\Delta\chi^2 = \chi^2_{\rm null} - \chi^2_{\rm model}. \tag{8}$$

To decide whether this impact is statistically significant or not, we run the exact same procedure described above on 1000 lognormal mock realizations. In this way, we obtain the probability distribution of $\Delta\chi^2$. We define $\Delta\chi^2(68)$ as the value below which are 68 per cent of the $\Delta\chi^2$ values from the mocks. Then, we consider an SP map significant if

$$S_{\rm 1D} = \frac{\Delta\chi^2}{\Delta\chi^2(68)} > T_{\rm 1D}, \tag{9}$$

where $T_{\rm 1D}$ is a significance threshold that is fixed beforehand. The square root of this quotient is proportional to the significance in terms of σ.

After identifying the most contaminating map, s_i, the next step is to obtain a weights map, $w_{s,\,i}$, to correct its impact. We compute this weights map as

$$w_{s,i} = \frac{1}{F(s_i)}, \tag{10}$$

where $F(s_i)$ is a linear function of s_i with which its 1D relation is fitted. In general, this function depends on the nature of the SP map, although the aim is to use functions as simple as possible to prevent overfitting. In the case of PC maps, we find no significant deviations from linearity in the 1D relations (see Appendix E).

After obtaining the weight map, the pipeline normalizes it to $\bar{w}_s = 1$. Then, it is applied to the data in such a way that $N^p_{\rm gal} \rightarrow N^p_{\rm gal} \times w^p_s$, where p is an index that runs over the footprint pixels at $N_{\rm side} = 4096$. The process is repeated iteratively, identifying at each iteration the most significant PC map and correcting for it until all the PC maps have a significance lower than $T_{\rm 1D}$. At iteration i, the weights from iterations 1 to i have been applied. Fig. 1 shows the 1D relation of a given PC map that has been identified as a significant contaminant (dots) and after correcting for it (triangles).

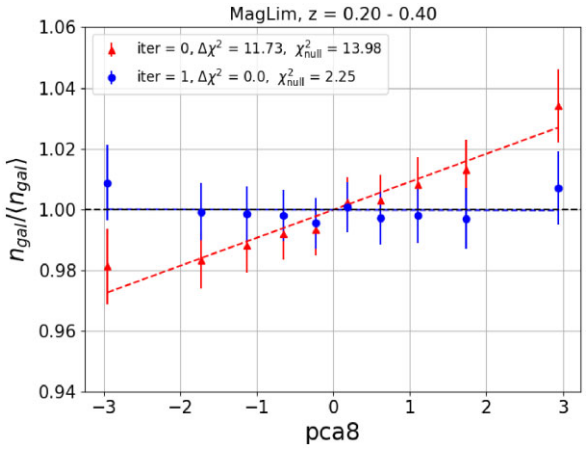

Figure 1. Example of how ISD works. We illustrate this by showing the observed pixel number density (relative to the mean over the full footprint) as a function of a PC map pixel value, evaluated in 10 equal area bins. We refer to this as 1D relation. The method identifies the PC map *pca*8 as the most significant one at iteration 0 (i.e. no weights have been applied yet) at the first redshift bin of MAGLIM. The corresponding 1D relation is depicted by the red triangles and the red line corresponds to their best-fitting linear function. After correcting for the contaminating template with weights (given by equation 10) at iteration 1, the impact of this PC map on the data is highly reduced. The blue points and their best-fitting linear function (blue line) show that the 1D relation is now compatible with no effect.

The weights associated to each significant PC map are incorporated multiplicatively to the total weight map, $w_{\rm T}$, that is

$$w_{\rm T} = \prod_{i=1}^{f} w_{s,i}, \tag{11}$$

where i runs over the number of PC maps it is necessary to weight for. $w_{\rm T}$ is then the total weight map that contains the information about the individual contaminants. These are the weights we apply to the data to mitigate the contamination. This total weight map is also normalized so its mean value over the full footprint is 1. The pipeline runs this procedure for each redshift bin independently.

5.3.2 *Elastic Net (ENET)*

We also generated sets of weights using the ENET method described in Weaverdyck & Huterer (2021) on the list of 50 PC maps. In this work, ENet has been used to perform robustness tests. Recall that the ISD method estimates contamination via a series of 1D regressions that are used to construct a total weight map via equation (11). In contrast, ENET estimates the amplitude of contamination for all PC maps simultaneously, by maximizing the following log-posterior over α:

$$\mathcal{P}(\alpha) \propto -\frac{1}{2N_{\rm pix}}||\delta_{\rm obs} - \boldsymbol{S}\alpha||_2^2 - \lambda_1||\alpha||_1 - \frac{\lambda_2}{2}||\alpha||_2^2, \tag{12}$$

where α_i is the contamination amplitude for PC map s_i, $\boldsymbol{S}$ is a matrix with the pixelated PC maps as columns,[10] and

$$\delta_{{\rm obs},j} = \frac{f_{{\rm det,j}} N_j}{\sum_j^{N_{\rm pix}} (f_{{\rm det,j}} N_j)/N_{\rm pix}} - 1, \tag{13}$$

[10]In practice, we standardize PC maps to have mean 0 and unit standard deviation before computing equation (12).

where $f_{{\rm det,\,j}}$ is the fraction of pixel j that is not masked. The first term in equation (12) corresponds to the standard Gaussian likelihood that is maximized for an ordinary least-squares regression. The regularizing terms act as components of a mixed, zero-centred prior on the elements of α. The mixture consists of a Laplace and Gaussian distribution, with their precisions controlled by λ_1 and λ_2. The Laplace component is sharply peaked at zero, encouraging sparsity in the coefficients. We determine the values of λ_1 and λ_2 by minimizing the mean squared error of the predictions on held-out portions of the footprint via five-fold cross-validation. This allows the data to pick the precision and form of the prior based on predictive power.

We use the `scikit-learn` (Pedregosa et al. 2011) implementation of `ElasticNetCV`, with a hyperparameter space of $\lambda_1/(\lambda_1 + \lambda_2) \in \{0.1, 0.5, 0.9\}$ and 20 values of $(\lambda_1 + \lambda_2)$ spanning four orders of magnitude (automatically determined from the input data). We degrade all maps to $N_{\rm side} = 512$, and compute equation (12) using a *training mask* that only includes pixels with $f_{\rm det} \geq 0.1$ (detection fraction from the `Y3 GOLD` STD maps that is inherited by the PC maps). We performed many subsequent tests changing the definition of this training mask, with little observed impact on the final $w(\theta)$. Using ENET on the STD maps we also extended $\boldsymbol{S}$ to include quadratic terms of form s_i^2, and/or terms of form $s_i s_{\rm stellardens}$, but these showed decreased predictive power on held-out samples, suggesting that the risk of overfitting from these additional maps dominates over additional contamination they identify.

The total weight map is computed (still at $N_{\rm side} = 512$) as

$$w_{\rm T}^{\rm ENET} = [F_{\rm ENET}(\boldsymbol{S})]^{-1} = (1 + \boldsymbol{S}\hat{\alpha})^{-1}. \tag{14}$$

The ISD and ENET methods make different assumptions and take significantly different approaches to select important SP maps while minimizing the impact of overcorrection. ENET neglects the covariance of pixels, as well as the differing clustering properties of the SP maps, but it is less dependent on the basis of SP maps than is ISD. It avoids some of the difficulties the ISD method has when SP maps are highly correlated or contamination is distributed weakly across a combination of many maps, and hence missed by 1D marginal projections. We therefore expect the ENET method to be a useful robustness test of the fiducial ISD method, and it is also used to estimate the systematic contribution to the $w(\theta)$ covariance (see Section 6).

5.3.3 *Neural net weights (NN-weights)*

To evaluate the robustness of the assumptions made and codes used in producing galaxy-density weights, we created a third alternative process with different choices and independent code – in particular, abandoning the assumption that the mean galaxy density is a linear or polynomial function of all SP maps. The basic principle remains the same, namely that a function $w(\boldsymbol{s})$ of the vector $\boldsymbol{s}$ of SP values is found that maximizes the uniformity of the observed catalogue. In this case, however, the function is realized by an NN, in a manner very similar to that of Rezaie et al. (2020).

In contrast to ISD and ENET, we apply this method on the STD basis of maps. In addition, two important changes to the weighting procedure were made to avoid having the NN overtrain, in the sense of absorbing true cosmological density fluctuations into the observational density factor w. First, the input STD maps were limited to those that should in principle fully describe the characteristics of the coadd images: the *fwhm*, *skyvar_uncertainty*, *exptime* and *fgcm_gry* exposure-averaged values for each of the *griz* bands, the *sfd98* extinction estimate, and a *gaia_density* estimate of local stellar

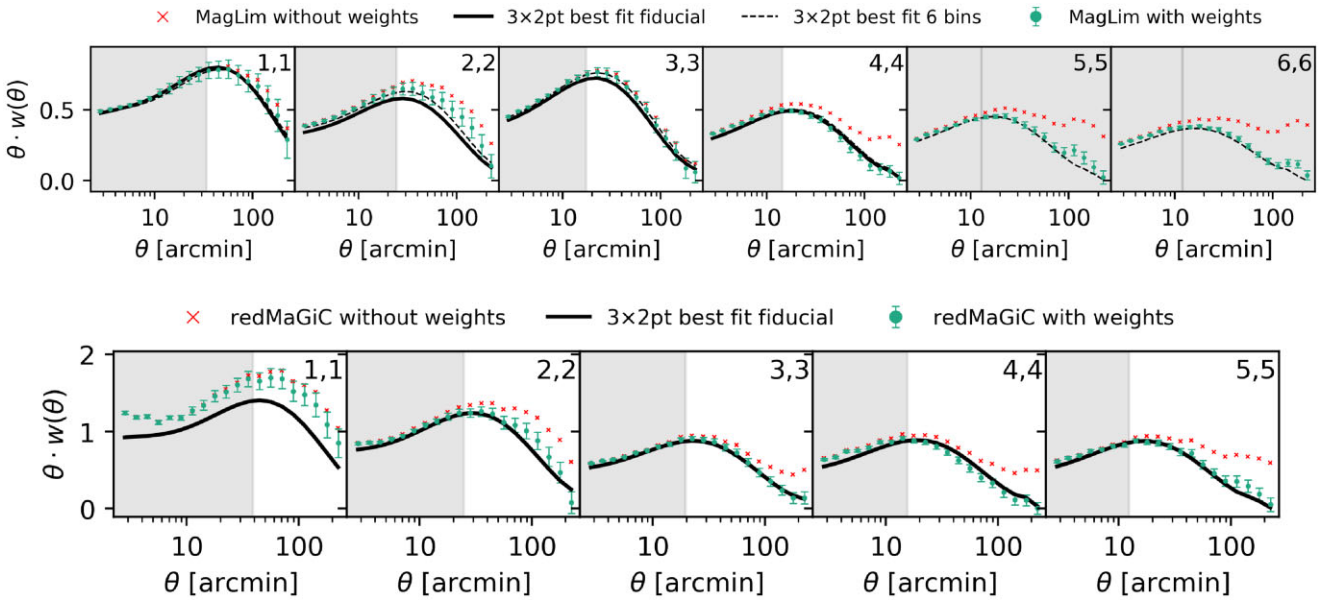

Figure 2. DES Y3 galaxy clustering results for MAGLIM (top panel) and REDMAGIC (bottom panel). The green points correspond to the angular correlation function of the ISD-PC<50 weighted data, while the red points correspond to the uncorrected data. The solid black line shows the best-fitting theory prediction from the DES Y3 3 × 2pt ΛCDM results of each sample (DES Collaboration et al. 2022). Note that for MAGLIM, we also show the best fit from the analysis including all six redshift bins (dashed black line), although the fiducial 3 × 2pt cosmology results from this sample only include its first four bins. The shaded regions correspond to the scales that are excluded for cosmological constraints.

density constructed from Gaia EDR3 (Gaia Collaboration 2020). We confirm that weights constructed with these STD maps eliminate any correlation of galaxy density on *airmass* or *depth*, and additionally find that *fgcm_gry* has no significant effect, so it is dropped, leaving 14 STD maps. The second major change to avoid overtraining is to institute N-fold cross-validation: the footprint is divided into healpixels at $N_{\rm side} = 16$, which are randomly divided into N distinct 'folds'. The weights for each fold are determined by training the NN on the other $N - 1$ folds, halting the training when the loss function for the target fold stops improving. We use $N = 3$.

The weights are created on a healpixelization at $N_{\rm side} = 4096$. With n_i, f_i, and w_i being the galaxy counts, useful-area fraction, and weight estimate for each healpixel, respectively, the NN is trained to minimize the binary cross-entropy :

$$S \equiv \sum_{n_i>0} \log \bar{n}\, f_i w_i + \sum_{n_i=0} \log\left(1 - \bar{n}\, f_i w_i\right). \tag{15}$$

In a further departure from the standard weighting scheme, we take the input vector $\boldsymbol{s}$ to be the logarithm of each input STD map (except for *sfd98*, which is already a logarithmic quantity), then linearly rescale each dimension to have its 1–99 percentile range span (0,1). We mask the < 1 per cent of survey area for which any such rescaled SP has s_i outside the range (− 0.5, 1.5), knowing that the NN will fail to train properly on rare values of STD maps.

Using the KERAS software,[11] we define the weight function for a given galaxy bin as

$$\log w(\boldsymbol{s}) = \alpha \cdot \boldsymbol{s} + NN(\boldsymbol{s}), \tag{16}$$

where α defines a nominal power-law relationship between the STD maps and the expected galaxy density, and NN is a three-layer perceptron describing deviations from pure power-law behaviour. The training of all folds for all redshift bins can be done overnight on a single compute node.

6 RESULTS

ISD returns a list of maps with significant impact on galaxy clustering and that we need to weight for in each redshift bin of the samples. We studied the impact of observing conditions at three different significance threshold values, $T_{\rm 1D} = 2,\ 4,\ 9$. Increasing this threshold is equivalent to relaxing the strictness of the decontamination, decreasing the number of significant SP maps. After testing for over and undercorrection on mocks, the fiducial choice of significance threshold is $T_{\rm 1D} = 2$ (see Sections 7 and 8 for more details).

We find that, in general, both samples show a similar trend and they are more impacted by observing conditions at higher redshift. Generally, more SP maps are significant for the MAGLIM sample than for REDMAGIC. The measured angular 2pt correlation functions on the weighted samples can be seen in Fig. 2. The S/N[12] of this detection is ∼63 for both samples (using only the first four bins of MAGLIM). The data have been corrected for systematic contamination by applying the ISD-PC<50 weights. After the correction, they are in good agreement (green points) with the best-fitting cosmology from 3 × 2pt. The deviation in the first redshift bin for REDMAGIC is known to come from an inconsistency between clustering results and galaxy–galaxy lensing in this sample. We defer the discussion of this important result from the point of view of observational systematics to Section 8. We note also that for MAGLIM we depict two best-fitting correlation functions: the best-fitting model from 3 × 2pt analysis using its six redshift bins (dashed black lines) and excluding

[11] https://keras.io.

[12] The signal-to-noise ratio is defined as $S/N \equiv \frac{w_{\rm data}(\theta) C^{-1} w_{\rm model}(\theta)}{\sqrt{w_{\rm model}(\theta) C^{-1} w_{\rm model}(\theta)}}$, where C is the $w(\theta)$ part of the covariance matrix and $w_{\rm model}(\theta)$ is the best-fitting model from 3 × 2pt.

its last two bins (solid black lines). The DES fiducial constraints are obtained without the last two bins (shaded in grey in Fig. 2), as explained in Porredon et al. (2021a). The shaded regions in this figure depict the scales excluded (see Table 1) from our data vectors. These regions are not used to obtain constraints on cosmological parameters. The uncorrected $w(\theta)$ are shown as red crosses. We note that the impact of systematic corrections is easily larger than the statistical uncertainty in the measurements, and are therefore necessary for unbiased cosmological inference, as we will illustrate below. These corrections are more important at higher redshift bins in both galaxy samples. For a comparison of this correction with respect to DES Y1 galaxy clustering, see Elvin-Poole et al. (2018).

In Fig. 3, we explicitly demonstrate the importance of our systematics correction by placing constraints on $\Omega_{\rm m}$ and the clustering biases b^i from the galaxy clustering correlation function alone. We do this by fitting the theory model presented in Section 2 to the data using COSMOSIS and the POLYCHORD sampling software (Handley, Hobson & Lasenby 2015a,b). The covariance that we employ is given by COSMOLIKE (Krause & Eifler 2017) and it includes the systematic contributions that we introduce in Section 8.4. We again marginalize over shifts in the photometric redshift distributions and over their widths. These nuisance parameters are sensitive to the clustering amplitude. For both samples, the rest of the cosmological parameters are fixed to their respective DES Y3 fiducial best-fitting cosmology (note that for MAGLIM this only considers the first four redshift bins). For this reason, this constraint on $\Omega_{\rm m}$ should not be taken as a true constraint, but this illustrates how the changes in the measured $w(\theta)$ can impact cosmology constraints. The priors for these cosmological and nuisance parameters are given in Table 3. We obtain these contours for the unweighted and ISD-weighted data. As evidence of robustness of our choice of SP maps, we also show contours for another configuration of ISD (ISD-STD34), where only 34 STD maps are considered (see Section 8.1 and appendix B of Carnero Rosell et al. (2021) for more details on this selection of SP maps). The corrections for the two ISD configurations are equivalent within the statistical uncertainty. In Fig. 3, we focus on the redshift bins that show the most prominent systematic shift in the $w(\theta)$, namely the fourth and the fifth bins of the MAGLIM and REDMAGIC samples, respectively. For these bins, we find a difference in the mean of the posteriors of $\Omega_{\rm m}$ from uncorrected (red contours) and corrected data (blue contours) of 4.03σ for MAGLIM and 6.79σ for REDMAGIC, where σ is the standard deviation of the posterior distribution of this parameter for the corrected data. Failing to correct for the systematic impact of the SP maps would result in shifting the inferred galaxy bias parameters to higher values while significantly lowering $\Omega_{\rm m}$. The significance of these shifts is somewhat larger than that obtained from the 3×2pt analysis, as we fix the rest of the cosmological parameters (while still varying nuisance parameters) such that the uncertainty is reduced. Note that because of correlations between the galaxy bias parameters and $\Omega_{\rm m}$, a given redshift bin with relatively little change in $w(\theta)$ due to weighting (e.g. bin 3 of MAGLIM) can still have a significant shift in its inferred galaxy bias.

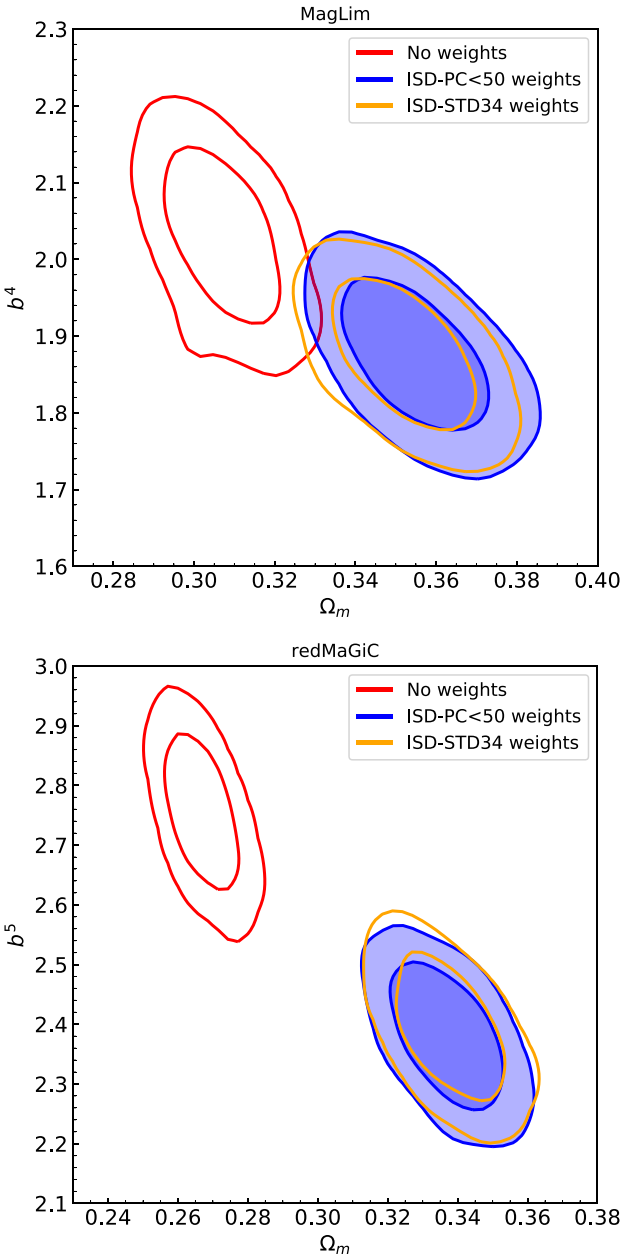

Figure 3. Constraints on $\Omega_{\rm m}$ and galaxy bias before and after applying our weighting methodology to the data for the fourth redshift bin of MAGLIM (top panel) and the fifth bin of REDMAGIC (bottom panel). We focus on the redshift bins where the impact of the systematic effects is more relevant in $w(\theta)$ (see Fig. 2). Red contours correspond to the uncorrected data, while blue contours correspond to the corrected data. The absence of correction strongly biases our estimations. We also show constraints for ISD-STD34 weighted data (orange contours). We obtain similar behaviours for the rest of the redshift bins of both samples. The goodness of fit for the no weights, ISD-PC<50 and ISD-STD34, cases are $65.23/30\,(p = 2\times10^{-4})$, $42.25/30\,(p = 0.07)$, and $38.73/30\,(p = 0.13)$ for MAGLIM and $156.05/42\,(p = 5\times10^{-15})$, $66.10/42\,(p = 0.01)$, and $68.91/42\,(p = 0.01)$ for REDMAGIC, respectively. There is an improvement in the p-value from $\sim$0.02 (no weights applied case) to $\sim$10 per cent (weights applied case). These χ^2 values correspond to galaxy clustering-only fits.

7 WEIGHTS VALIDATION

We validate our methodology on simulated catalogues to ensure that no biases are induced. We use unaltered lognormal mocks and also mocks that are artificially contaminated by our SP maps (see Appendix A for details on how we apply this contamination). We contaminate these mocks by applying the inverse of the weights determined from the data using ENET on the full list of 107 STD maps. Decontamination, however, is performed using weights

Table 3. List of prior values used to constrain Ω_m and the sample galaxy biases b^i per redshift bin.

Redshift biy	Δz	σ_z
	MAGLIM	
$0.20 < z < 0.40$	$(-0.009, 0.007)$	(0.975,0.062)
$0.40 < z < 0.55$	$(-0.035, 0.011)$	(1.306,0.093)
$0.55 < z < 0.70$	$(-0.005, 0.006)$	(0.87,0.054)
$0.70 < z < 0.85$	$(-0.007, 0.006)$	(0.918,0.051)
$0.85 < z < 0.95$	(0.002, 0.007)	(1.08,0.067)
$0.95 < z < 1.05$	(0.002, 0.008)	(0.845,0.073)
	REDMAGIC	
$0.15 < z < 0.35$	(0.006,0.004)	Fixed to 1
$0.35 < z < 0.50$	(0.001,0.003)	Fixed to 1
$0.50 < z < 0.65$	(0.006,0.004)	Fixed to 1
$0.65 < z < 0.80$	$(-0.002, 0.005)$	Fixed to 1
$0.80 < z < 0.90$	$(-0.007, 0.010)$	(1.23,0.054)
	Both samples	
	Ω_m	b^i
All redshifts	[0.1,0.9]	[0.8,3.0]

Notes. The other cosmological parameters have been fixed to the fit values in the 3 × 2pt analysis as described in the text. Square brackets denote a flat prior, while parentheses denote a Gaussian prior of the form $\mathcal{N}(\mu, \sigma)$. The shift Δz and stretch σ_z parameters are defined in equations (4) and (5). In some cases, the latter is not marginalized over (fixed). The redshift priors were determined in Cawthon et al. (2020).

determined by ISD-PC<50. This procedure adds an additional layer of protection: if we contaminate mocks with the weights from one method and decontaminate by the same method, the test is only checking sensitivity to forms of contamination to which we a priori know the method is sensitive to. Generating an equally plausible realization of contamination from an alternative method adds the benefit of potentially revealing blind spots in the method that is being validated. In Appendix B, we also perform a sanity check to confirm that we recover unbiased $w(\theta)$ measurements at all angular scales under idealized circumstances, that is, contaminating and correcting for the exact same set of SP maps.

We calculate $\bar{w}_{\rm dec}(\theta)$ and $\bar{w}_{\rm unc}(\theta)$ as the mean correlation function of 400 decontaminated and 400 uncontaminated mocks, respectively. Since the lognormal mocks are generated at $N_{\rm side} = 512$, which corresponds to separation angles of ∼6.9 arcmin between pixels, we compute the correlation functions at the 14 fiducial angular scales that are larger than this limit. Then we estimate the impact of the different biases (see next two Sections) on $w(\theta)$ by means of the true mean in uncontaminated mocks, $\bar{w}_{\rm unc}(\theta)$:

$$\chi^2 = (\bar{w}_{\rm dec}(\theta) - \bar{w}_{\rm unc}(\theta))^\top \times C^{-1} \times (\bar{w}_{\rm dec}(\theta) - \bar{w}_{\rm unc}(\theta)). \quad (17)$$

The covariance matrix, C, is the galaxy clustering part of the analytical covariance given by COSMOLIKE, and it is also used for the clustering part of the 3 × 2pt cosmological analysis. If we find that any bias causes a change in the joint fit to all redshift bins according to the definition above, equivalent to $\chi^2 > 3$, then we marginalize over this bias in our final analysis. This threshold was chosen such that the impact on χ^2 would be a small compared to the expected width of the χ^2 distribution of the $w(\theta)$ data vector. As we detail in Section 8.4, we marginalize over biases by modifying the covariance matrix to account for these sources of systematic uncertainty. The fiducial covariance matrix for DES Y3 3 × 2pt analysis includes these systematic terms.

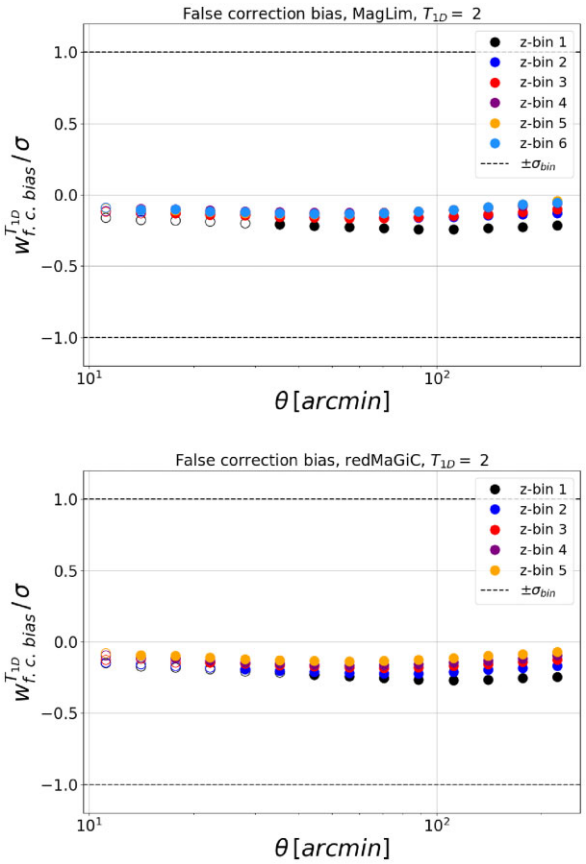

Figure 4. False correction bias, $w^{T_{\rm 1D}}_{\rm f.\,c.\,bias}(\theta)$, for MAGLIM (top panel) and REDMAGIC (bottom panel) relative to the $w(\theta)$ error from the unaltered COSMOLIKE covariance diagonal elements. Negative values are indicative of overcorrection. Both samples show negligible levels of overcorrection, weak dependence with the angular scale and at most ∼20 per cent of the statistical error. The values depicted here have been calculated with significance threshold $T_{\rm 1D} = 2$. Empty dots correspond to the angular scales not considered for each redshift bin of the samples.

7.1 False correction test

Since we consider a large number of SP maps in this analysis, chance correlations between the data and some of these maps could arise, even after reducing our number of SP maps. This is more important when using a strict significance threshold. These purely random correlations could cause overcorrections, therefore biasing the measured value of $w(\theta)$ and the inferred cosmological parameters. To characterize this effect, we run ISD with $T_{\rm 1D} = 2$ on a set of 400 uncontaminated mocks and then we obtain their correlation functions, $w^{T_{\rm 1D}}_{{\rm dec},\,i}$. The false correction bias is defined as

$$w^{T_{\rm 1D}}_{\rm f.\,c.\,bias}(\theta) = \frac{1}{400}\left(\sum_{i=1}^{400} w^{T_{\rm 1D}}_{{\rm dec},\,i}(\theta) - \sum_{j=1}^{400} w_{{\rm unc},\,j}(\theta)\right), \quad (18)$$

where $w_{{\rm unc},\,j}$ are the correlation functions measured on the unaltered uncontaminated mocks.

In general, the effect of removing the systematic effects is to diminish the amplitude of $w(\theta)$. Thus, a negative value of this estimator indicates overcorrection. In Fig. 4, we show the results of $w^{T_{\rm 1D}}_{\rm f.\,c.\,bias}(\theta)/\sigma$ for $T_{\rm 1D} = 2$, where σ is the diagonal of the unmodified covariance matrix. We find a very marginal indication of overcorrection, always well below the statistical error. We also note

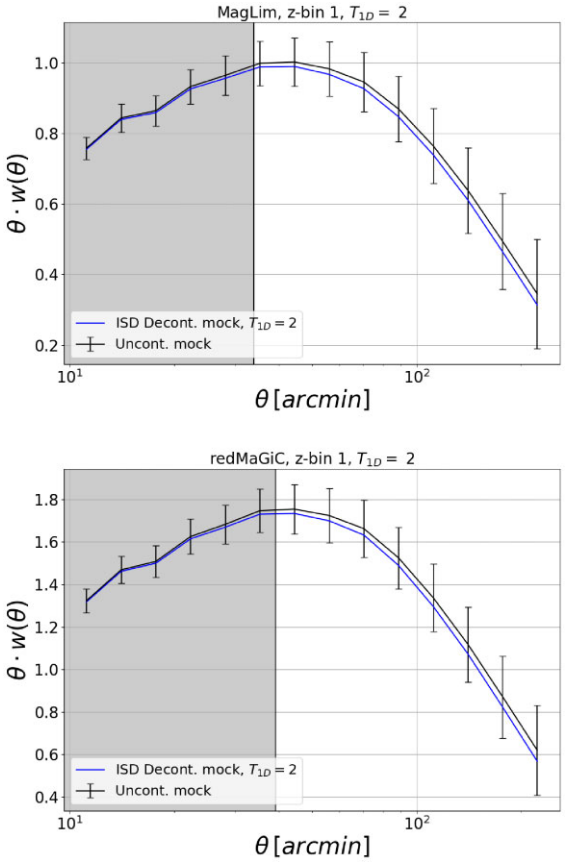

Figure 5. Mean angular correlation function, $w(\theta)$, from raw uncontaminated lognormal mocks (black lines) and decontaminated uncontaminated mocks (blue lines) for MAGLIM (top panel) and REDMAGIC (bottom panel) at their lowest redshift bins. The shaded region corresponds to the scales excluded at this redshift. In this redshift bin there is $\sim$20 per cent of false correction with respect to the statistical error due to chance correlations between PC maps and mocks. The error bars correspond to the diagonal of the covariance matrix with systematic terms added.

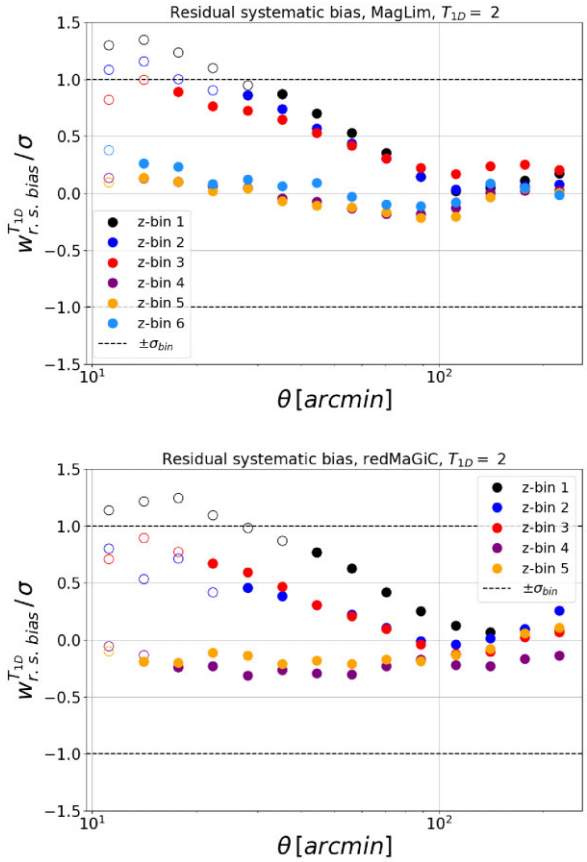

Figure 6. Residual systematic bias, $w^{T_{1D}}_{r.s.\,bias}(\theta)$, for MAGLIM (top panel) and REDMAGIC (bottom panel) relative to the $w(\theta)$ error from the unaltered COSMOLIKE covariance diagonal. The empty dots represent the scales excluded at each bin. Both samples show similar trends: The highest redshift bins present lower biases, while the lowest ones show important levels of undercorrection at the smallest scales. On the other hand, the largest scales are recovered nearly unbiased. Since the χ^2 of the total residual bias in all bins is higher than 3, we add a systematic term to the covariance matrix to marginalize over this effect.

that this ratio has small angular dependence, as can be seen in Fig. 5, which compares the mean true $w(\theta)$ (black line) with the mean of the decontaminated correlation functions (blue line). Therefore, we do not consider any contribution from the false correction bias to the final covariance matrix. The small impact of this effect on the cosmological parameters is highlighted in Section 7.3. Nevertheless, we note that the error bars shown in Fig. 5 correspond to the diagonal of the covariance matrix that has been modified to account for systematic uncertainties, as it is explained in Section 8.4.

7.2 Residual systematic test

Here we demonstrate that ISD effectively recovers the true correlation function from a contaminated sample. We can then verify if our approach (with $T_{1D} = 2$) meets the requirements for the Y3 cosmology analysis or whether it is necessary to account for any bias due to uncorrected contamination.

We define the residual systematic bias as

$$w^{T_{1D}}_{r.\,s.\,bias}(\theta) = \frac{1}{400}\left(\sum_{i=1}^{400} w^{T_{1D}}_{dec,\,i}(\theta) - \sum_{j=1}^{400} w_{unc,\,j}(\theta)\right), \qquad (19)$$

where the $w^{T_{1D}}_{dec,\,i}$ are the correlation functions measured on mocks that have had systematic contamination added and then have been decontaminated using ISD.

Because we are interested in the level of *residual* systematics that are insufficiently captured by the weighting method, we use the alternative method ENET with all 107 maps in the standard basis to generate an aggressive level of contamination. We observe that both ISD-PC107 and ENET-STD107 significantly overcorrect at the lowest redshift bins of both galaxy samples (see Section 8), so when using the corresponding weights to contaminate the mocks we are introducing excessive contamination. Therefore, we expect some degree of undercorrection when later running ISD with a subset of PC maps such as with ISD-PC<50. Furthermore, by using ENET to estimate the contamination instead of ISD, the contaminated mocks will include possible contamination modes to which ENET is sensitive but to which ISD may not be.

In Fig. 6, we show the results for this bias with respect to the diagonal of the unaltered analytical errors. While the highest redshift bins of both MAGLIM and REDMAGIC present moderate levels of overcorrection, the lowest redshift bins of the two samples show a trend to under-correct at the small angular scales, but still above the scales we exclude. As already mentioned, we expect some level

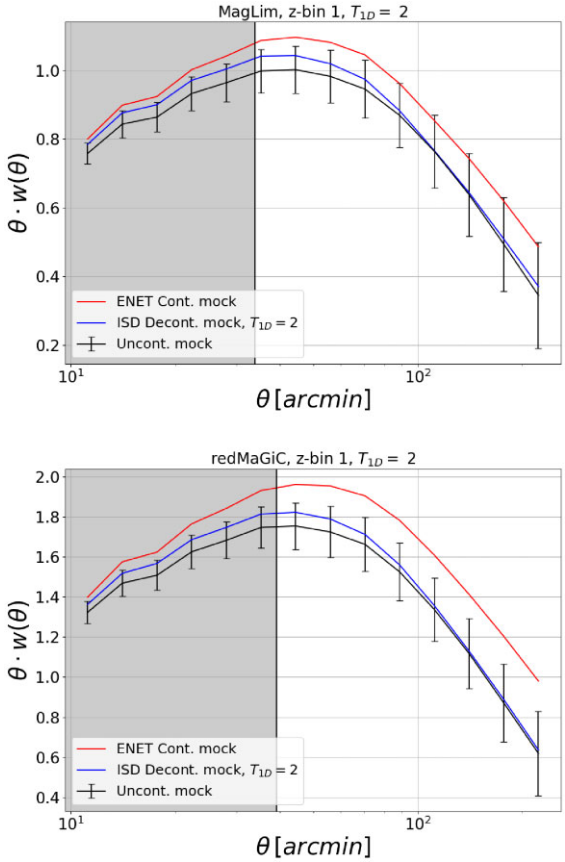

Figure 7. Mean angular correlation function, $w(\theta)$, from uncontaminated mocks (black line) and from decontaminated mocks (blue line) for MAGLIM (top panel) and REDMAGIC (bottom panel). The red line corresponds to the mean of the mocks with contamination added from ENET, and the shaded regions represent the scales not used for cosmological constraints. While ISD recovers a nearly unbiased clustering at the largest angular scales, there is an important bias at the smallest ones. For this reason, this effect is marginalized over by adding it a systematic contribution to the error budget. The error bars shown take into account this contribution.

of undercorrection due to the aggressive contamination imprinted on the mocks. Even under this consideration, these bins cause the χ^2 of the joint fit to exceed our limit, so we incorporate this bias as a systematic contribution to our covariance matrix. This is covered in Section 8.4. In Fig. 7, we depict the mean recovered clustering (blue lines) compared to the true clustering (black lines). We also show the mean contaminated correlation function (red lines). It can be seen that ISD performs a nearly unbiased decontamination at the largest angular scales. The error bars in this figure include the systematic terms added to the covariance (see Section 8.4 for a comparison of the error bars with and without the systematic contributions).

7.3 Impact on parameter estimation

Finally, as an additional evidence of robustness we check the impact of the decontamination procedure on the estimation of cosmological parameters. We use as data vectors (i) the mean correlation function over 400 uncontaminated mocks, (ii) the mean correlation function biased by our overcorrection estimate (Section 7.1), and (iii) the mean correlation function biased as by the residual systematic uncertainty estimate (Section 7.2). To test the influence of these analysis modifications on cosmology, we recalculate the constraints on the parameters Ω_m and b^i, marginalizing as before over redshift-bin centroid positions and widths of the redshift distributions. We use the same priors from Table 3 and the rest of the parameters are fixed to the values used to generate the mocks. The results that we obtain are shown in Fig. 8. It can be seen that the recovered contours from the false correction bias case (run on uncontaminated mocks) are in good agreement with those from the reference case, demonstrating that biases from overcorrection in inferred cosmological parameters are negligible. The contours corresponding to the residual systematic bias (run on ENET contaminated mocks) show a small level of undercorrection that is translated to slightly higher galaxy bias values, though this mismatch is also within the statistical uncertainties given by our analytical covariance. This covariance includes a systematic uncertainty correction that is explained in Section 8.4. In Table 4, we present the difference in the Ω_m and b^i mean posteriors in units of σ from uncontaminated mock contours. We note that all differences are smaller than 0.5σ. It must be taken into account that, since the rest of the cosmological parameters are fixed, the 1σ contours are smaller than for any of the final DES cosmology analyses, making this test more stringent. We found that the mean $w(\theta)$ of the lognormal mocks is slightly shifted to lower amplitudes from the theory prediction with the same input values. This causes some shifting of the contours as well, but we have verified that this does not affect our conclusions from the decontamination methodology.

8 POST-UNBLINDING INVESTIGATIONS OF THE IMPACT OF OBSERVATIONAL SYSTEMATICS ON $w(\theta)$

The DES 3 × 2pt analysis combines the correlation functions from galaxy clustering, $w(\theta)$, galaxy–galaxy lensing (for short, gg-lensing), $\gamma_t(\theta)$, and cosmic-shear, $\xi_\pm(\theta)$, in order to improve the individual constraining powers of each probe and to break degeneracies in some cosmological parameters. In addition, since each of these 2pt functions is potentially affected by different systematic effects, it allows for consistency checks comparing different results. The consideration of two different lens galaxy samples for $w(\theta)$ and $\gamma_t(\theta)$ allows us to further assess the robustness of the whole cosmology analysis. The cosmology analysis is performed blindly, that is, we only look at the cosmology results once a set of predefined criteria are fulfilled, as is described in DES Collaboration et al. (2022). During the unblinding process of REDMAGIC we found that this sample passed all the consistency tests we had a priori decided were required for unblinding. However, after unblinding, we identified a potential inconsistency between the amplitudes of galaxy clustering and gg-lensing: Either the former has an anomalously high amplitude or the latter has an anomalously low one. This inconsistency is explored in detail in Pandey et al. (2021).

Observational systematics from survey properties tend to increase the amplitude of $w(\theta)$ and so one possible explanation is that the clustering amplitude is anomalously high due to the decontamination procedure failing to fully capture all contamination in the data. Thus, the true underlying galaxy correlation function in the data would not be correctly recovered. This led us to perform a variety of additional tests as we describe below. It was during these tests when some of the methods described in Sections 4 and 5 were incorporated, such as the change in SP map basis (both expanding the number of SP maps and decorrelating them) and the robustness checks using ENET and the neural net. Ultimately, we found that the difference between galaxy clustering and lensing observables in REDMAGIC remained robust to different choices in the decontamination procedure. We also applied these additional tests to the MAGLIM sample before it

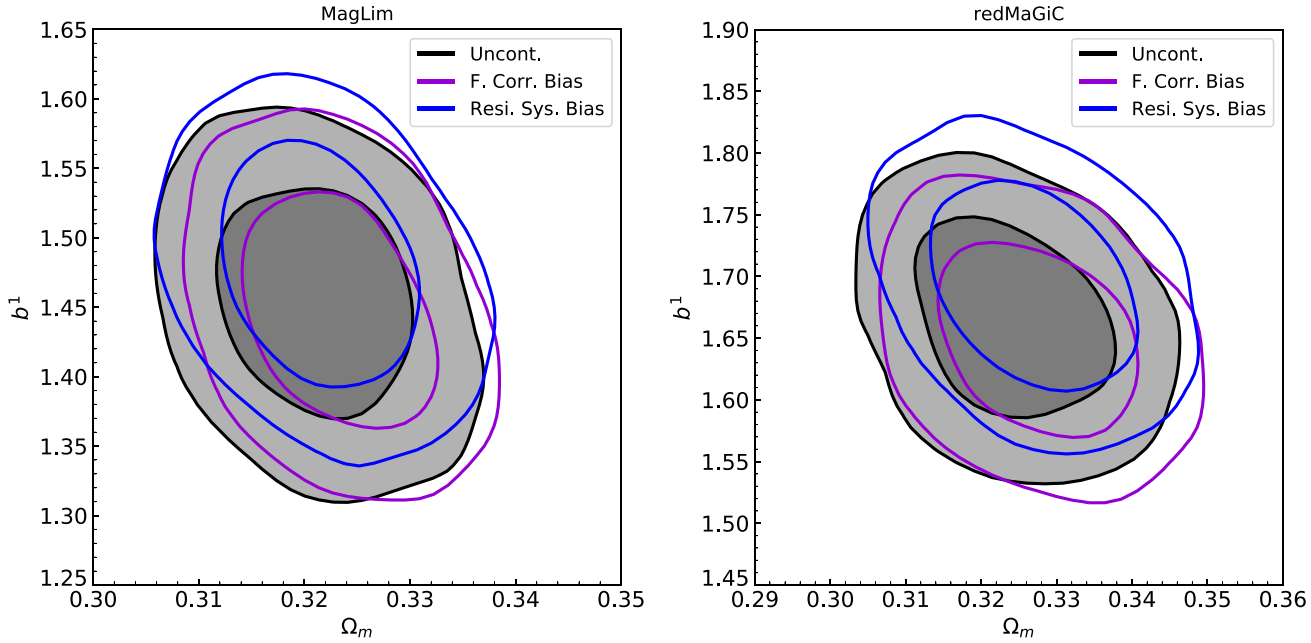

Figure 8. Constraints in the $\Omega_m - b^i$ parameter space at fixed σ_8 from the mean $w(\theta)$ of uncontaminated mocks (black contours) and from decontaminated mocks according to the false correction bias (violet contours) and to the residual systematic bias (blue contours). MAGLIM is shown in the left-hand panel and REDMAGIC in the right-hand one. It can be seen how both the false correction bias and the residual systematic bias lead to small shifts from the reference mocks relative to the error given by the COSMOLIKE analytical covariance, which includes the systematic uncertainty contributions. We only show contours for the first redshift bins of the two galaxy samples in this figure, but we verify that the shifts at the other bins are smaller or smaller. Because σ_8 and other cosmological parameters are fixed in this test, the posterior is smaller than from any of the DES final cosmological analyses that use the $w(\theta)$ data.

Table 4. Relative difference in the Ω_m and b^i mean of the posteriors for the two tests on decontaminated mocks in units of σ.

Parameter	False correction bias	Residual systematic bias
	MAGLIM	
Ω_m	0.36	0.08
b^1	− 0.09	0.43
b^2	− 0.06	0.40
b^3	− 0.25	0.12
b^4	0.05	0.16
b^5	− 0.15	− 0.02
b^6	− 0.06	− 0.04
	REDMAGIC	
Ω_m	0.39	0.31
b^1	− 0.29	0.50
b^2	− 0.33	0.11
b^3	− 0.30	0.27
b^4	− 0.32	− 0.35
b^5	− 0.19	− 0.21

Notes. All values are below half a σ. Note that the posteriors in this test are much smaller than in any of the final DES cosmology analyses because all the other parameters are fixed.

was unblinded. In contrast to our results with the REDMAGIC sample, once we unblinded the MAGLIM sample we found that its lensing and clustering signals were consistent with one another. For this reason, MAGLIM is the fiducial choice for our cosmological constraints (DES Collaboration et al. 2022). The fiducial MAGLIM cosmology results use only the first four redshift bins, as the two highest redshift bins gave inconsistent results, while adding little constraining power. Porredon et al. (2021a) investigates these results in detail.

8.1 ISD and ENET at the STD map basis

Before unblinding, ISD weights were obtained from a selection of STD maps performed by setting a limit for the Pearson's correlation coefficient between them. This selection gave 34 representative STD maps that were used to obtain weights with ISD (ISD-STD34). More details on this selection can be found in appendix B of Carnero Rosell et al. (2021). To check whether the clustering-lensing inconsistency found in REDMAGIC was caused by an STD map not selected in the STD34 set, we ran ISD on the full list of STD maps, and verified that derived weights did not significantly impact the resulting clustering signal. In Fig. 9, we show the correlation functions at the first bin of REDMAGIC obtained for these two configurations of ISD with STD maps.

We also checked the subtle possibility of a combination of STD maps leading to a large systematic contribution despite no single map being individually significant. For this reason, we ran ENET-STD107 on REDMAGIC, which simultaneously fits to all template maps, finding a significant decrease of $\sim 1\sigma$ in the amplitude of the correlation function in the first three redshift bins. This motivated further investigation to determine whether there could be significant residual contamination in the form of low-significance linear combinations of STD maps that eluded the initial decontamination procedure. We found that decorrelating the STD maps via PCA before running the ISD method and using the 107 components resulted in much better agreement between ISD and ENET, which motivated the change to the PC basis that has been used for the results presented in this paper (see ISD-PC107 in Fig. 9). We also found that there are no significant changes when running ENET on the PC basis of maps (this method is less basis-dependent, since it performs a simultaneous fit to all maps).

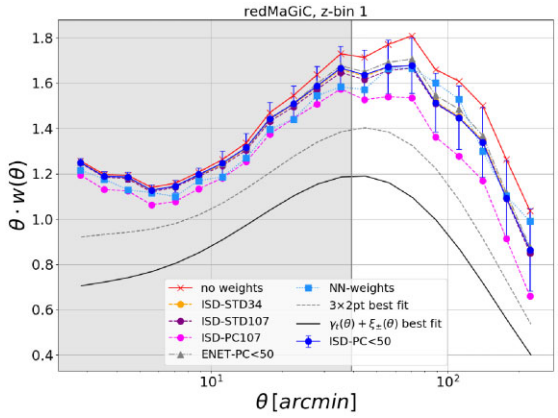

Figure 9. Comparison of the clustering amplitude recovered from several methods and configurations for the first redshift bin of REDMAGIC. All methods agree within the statistical uncertainty given by the analytical covariance. The solid red line corresponds to the unweighted data and the dashed purple line corresponds to the ISD-PC107 configuration. The difference between this configuration and the rest of methods is consistent with the overcorrection observed on contaminated mocks (see Fig. 10). The solid and dashed black lines are the best-fitting cosmology from cosmic-shear and gg-lensing only and from the 3 × 2pt analysis, respectively. The grey shaded region represents the scales that are not used for cosmological analysis. None of the various configurations produce values of $w(\theta)$ approaching the best-fitting prediction from cosmic-shear and gg-lensing.

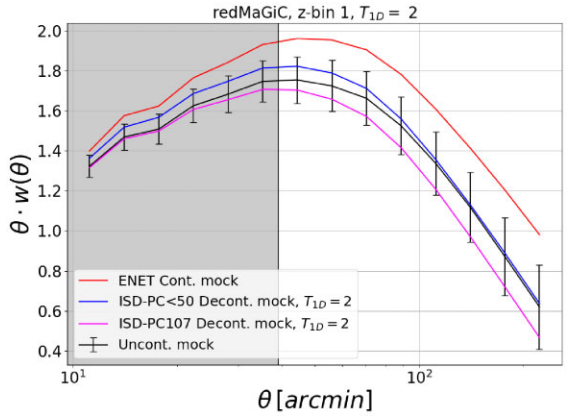

Figure 10. Effect of considering different numbers of PC maps on the two-point angular correlation function: weights obtained from 107 PC maps cause overcorrection on $w(\theta)$ (magenta line). This overcorrection ranges from $\sim 0.5\sigma$ to 1σ and is most prominent at large angular scales. This overcorrection can explain most of the difference in clustering between ISD-PC<50 and ISD-PC107 observed in Fig. 9. On the other hand, weights obtained from the first 50 PC maps yield a clustering amplitude (blue line) that is in good agreement with the mean $w(\theta)$ from uncontaminated mocks (black line), especially at the largest scales. The difference between the amplitudes from uncontaminated and ISD-PC<50 decontaminated mocks is included as a systematic contribution to the covariance (error bars in this figure already include that term). The red line corresponds to the ENET-STD107 contaminated mocks.

8.2 ISD and ENET in the PC map basis

We evaluated the impact of the ISD-PC107 weights on both uncontaminated and ENET contaminated mocks, similar to the tests from Sections 7.1 and 7.2. These tests revealed a significant level of overcorrection when using the full list of PC maps with ISD, especially when evaluated on contaminated mocks, indicating that true LSS fluctuations were being removed in the decontamination process. This effect can be seen in Fig. 10. We observed a similar overcorrection effect on MAGLIM with these ISD settings. The overcorrection is most prominent in lower redshift bins where the intrinsic clustering signal is larger, losing significance at higher redshift for both samples.

These results suggest that there is a higher likelihood of chance correlation in the PC107 basis than in the STD107 basis. We also found that PC107 weights obtained from the data showed significant correlations with DES κ maps (see Appendix D for details). We therefore conclude that using all 107 principal components results in removing not only actual systematic contamination from the data, but also cosmological signal, causing a lower $w(\theta)$ amplitude.

We therefore applied a cut-off to the number of PC maps to be used. To select this cut-off, we required that the weight map resulting from running ISD with the set of the first n PC maps should not induce a significant overcorrection on contaminated mocks (as we observed with ISD-PC107 weights), while still removing the contamination that was applied using ENET-STD107. We found that $n = 50$ principal component maps meets this requirement. The impact of the ISD-PC<50 weights on contaminated mocks and finally on the data can be seen in Figs 10 (blue line) and 9, respectively. Then, we calculated ENET-PC<50 weights as well, finding good agreement between the two methods with this configuration (see Fig. 9). Our adoption of this configuration was further supported by the desire to have a comparatively small number of maps to avoid overcorrection, as with the 107 PC maps, while still preserving most of the variance present in the full set of 107 STD maps. We point the reader to Appendix D for more details on the selection of this cut-off. We found that the difference between $w(\theta)$ functions given by ISD-PC<50 and ENET-PC<50 yields a χ^2 for the joint fit to all redshift bins smaller than 3. Nevertheless, we found some map configurations for the two methods that yield $\chi^2 > 3$. Thus, in order to be conservative, we consider this difference as an additional systematic uncertainty to be marginalized over, similar to the difference between uncontaminated and decontaminated mocks from Section 7.2.

For these reasons, we used ISD-PC<50 as the fiducial correction method, as described in the previous sections of this paper. In Fig. 9, we summarize the clustering amplitudes obtained from each of the methods and configurations described in the first redshift bin of REDMAGIC. None of the methods produce a $w(\theta)$ consistent with the best-fitting prediction from cosmic-shear and gg-lensing (solid black line). For reference, the dashed grey line shows the best-fitting prediction from the combined 3 × 2pt analysis.

The tests conducted to determine this cut-off were focused on the first redshift bin of REDMAGIC, but we verified that the impact of this choice on the rest of the bins is similar, although milder, since the overcorrection observed at higher bins is less significant. We also ran these tests on MAGLIM, obtaining similar conclusions for the same cut-off.

8.3 Tests with neural net weights

As noted in Section 8.3, we developed an independent, non-linear correction method using NNs. This was applied post-unblinding to test the robustness of the weights, in particular to the assumption of linearity between galaxy number density and the systematic maps. If there is excess clustering due to non-linear functions of the STD maps, then we expect it to be captured by the NN-weights. Because of the significant time required to run the method, we did not subject it to the full extent of validation tests on contaminated and uncontaminated mocks as we did for the ISD and ENET methods.

However, as Fig. 9 shows, the changes to $w(\theta)$ are small when using the NN-weights, suggesting that residual non-linear contamination from the existing set of STD maps is not driving a spuriously high estimate of $w(\theta)$.

8.4 Modifications to the covariance matrix

In this analysis, we consider the systematic uncertainty in the correction method from two sources: from the choice of correction method, and the bias measured in contaminated mocks (as mentioned in Section 7.2). As noted in the previous section, the NN-weights method did not undergo the extensive validation process that the ISD and ENET weights did. For this reason, we focused on the systematic uncertainty associated to the differences between ISD-PC<50 and ENET-PC<50.

The two systematics considered are each analytically marginalized over through an additional term in the $w(\theta)$ covariance matrix following the methodology of Bridle et al. (2002) summarized here. If one takes an arbitrary data vector $\mathbf{y}$ that is biased by an additive systematic effect $\boldsymbol{s}$,

$$\boldsymbol{y}' = \boldsymbol{y} + A\boldsymbol{s}, \tag{20}$$

where A is the amplitude of the systematic error. If the amplitude A has a Gaussian prior of zero-mean and width σ_A (which can be determined by external constraints), the parameter A can be analytically marginalized over in the covariance matrix of $\boldsymbol{y}$ with

$$\mathrm{Cov}(\boldsymbol{y}', \boldsymbol{y}') = \mathrm{Cov}(\boldsymbol{y}, \boldsymbol{y}) + \sigma_A^2 \boldsymbol{s}\boldsymbol{s}^T. \tag{21}$$

In this analysis, we model the impact of the systematic uncertainty in the correction as

$$w'(\theta) = w(\theta) + A_1 \Delta w_{\mathrm{method}}(\theta) + A_2 w^{T_{\mathrm{ID}}}_{\mathrm{r.\,s.\,bias}}(\theta), \tag{22}$$

where $\Delta w_{\mathrm{method}}(\theta)$ is the difference between the ISD and ENET methods, both using the PC<50 basis of maps as shown in Fig. 11; $w^{T_{\mathrm{ID}}}_{\mathrm{r.\,s.\,bias}}(\theta)$ is the residual systematic bias measured on lognormal mocks in Section 7.2, and A_1 and A_2 are two arbitrary amplitudes that describe the size of the systematic error in the correction.

We analytically marginalize over these terms assuming a unit Gaussian as the prior on the amplitudes A_1 and A_2 such that the measured systematic size is a 1σ deviation from the prior centre, and the systematic can move $w(\theta)$ in either direction. The final additional covariance term is

$$\Delta\mathrm{Cov}(\boldsymbol{w}', \boldsymbol{w}') = \Delta\boldsymbol{w}_{\mathrm{method}} \Delta\boldsymbol{w}_{\mathrm{method}}{}^T + \boldsymbol{w}^{T_{\mathrm{ID}}}_{\mathrm{r.\,s.\,bias}} \boldsymbol{w}^{T_{\mathrm{ID}}}_{\mathrm{r.\,s.\,bias}}{}^T. \tag{23}$$

The method difference term $\Delta w_{\mathrm{method}}(\theta)$ is measured on real data and therefore contains the same noise as the $w(\theta)$ data vector being used for cosmological inference. To avoid adding this noise to the covariance term, we fit a flexible polynomial to the two $w(\theta)$ measurements described in Appendix C. $\Delta w_{\mathrm{method}}(\theta)$ is the difference between these two polynomial fits.

The mock bias term $w^{T_{\mathrm{ID}}}_{\mathrm{r.\,s.\,bias}}(\theta)$ is averaged over 400 mocks so is a smooth function of θ and does not require any additional fitting. The impact of the additional covariance terms is shown in the error bars of Fig. 11. The systematic contribution to each tomographic bin is treated as independent so the covariance between bins is not modified.

8.5 Tests with BALROG

BALROG (Suchyta et al. 2016; Everett et al. 2022) is a software package that beds fake objects in real images in order to accurately characterize measurement effects.

BALROG simulated galaxies are created using real objects from the DES deep fields (Hartley et al. 2021), which can be considered as approximations to noiseless astrophysical sources due to the depth of the images they come from with respect to the wide field imaging. These objects have been measured using the same instrument and filters as the Y3 data set. This collection is sampled and injected into the individual single epoch images, which are then processed and coadded again with the same Y3 DES Data Management pipeline. Therefore their detectability is subject to the same conditions as the real galaxies from the Y3 wide field survey, as they inherit the background and noise properties of the real images.

BALROG is a useful tool to make independent consistency tests of the decontamination methods: while the galaxy samples trace the actual large-scale structure, the BALROG samples are formed by galaxies that are artificially injected on a uniform grid, that is, they are non-LSS distributed. What both real and BALROG samples have in common is the impact of systematics. Therefore, any correlation between the two after applying the weights would mean the presence of a common systematic. For this reason, we used the cross-correlation of REDMAGIC and MAGLIM with their associated BALROG samples to test for the presence of an extraneous signal that would indicate a pending, unknown systematic that is not being corrected by the applied weights. These results are presented in Fig. 12. The cross-correlations are calculated in $\sim 1000\,\mathrm{deg}^2$ (available area of the BALROG samples). We find that the cross-correlation with the weights applied is consistent with zero signal within the statistical errors. These errors are computed with jackknife re-sampling using 100 patches for MAGLIM and 50 for REDMAGIC. However, the signal itself is small but non-zero, growing in magnitude towards larger scales. We note that, due to its lower number density, the points for REDMAGIC are noisier than those for MAGLIM. The reduced χ^2 for a constant cross-correlation of 0 are 0.46, 0.96, 1.25, 3.60, 1.18 for REDMAGIC and 1.13, 0.71, 0.78, 0.94, 0.65, 0.69 for MAGLIM. The relative strength of the cross-correlation signal with respect to the auto-correlation signal can be seen in the bottom rows of each panel. In general, it is at or below 5 per cent for the five lowest angular bins at all redshift bins, and it is lower than 10 per cent for scales smaller than $\sim$30 arcmin. This relative strength gives us an indication of the size of a systematic effect that could be still unaccounted for. Even if the REDMAGIC results are noisy, those for MAGLIM do not show a clear indication of uncorrected effects from imaging systematics.

8.6 Summary of findings

We performed a series of tests post-unblinding to determine if the observed inconsistency between the galaxy clustering and gg-lensing signals in REDMAGIC is due to residual systematic contamination of the galaxy clustering signal. In particular, we investigated whether expanding the set of SP maps, adjusting the contamination model, or changing a variety of methodological choices for the decontamination procedure resulted in a significantly different inferred galaxy clustering signal. We largely performed these tests at the level of $w(\theta)$, without further looking at the impact of these decisions on cosmological parameters. The following list is a summary of the obtained results:

(i) Expanding the list of 34 to all 107 STD maps has negligible impact on the resulting amplitude of $w(\theta)$ using the fiducial ISD decontamination procedure. We thus conclude that the discrepancy is not due to residual contamination from one of the previously-discarded STD maps.

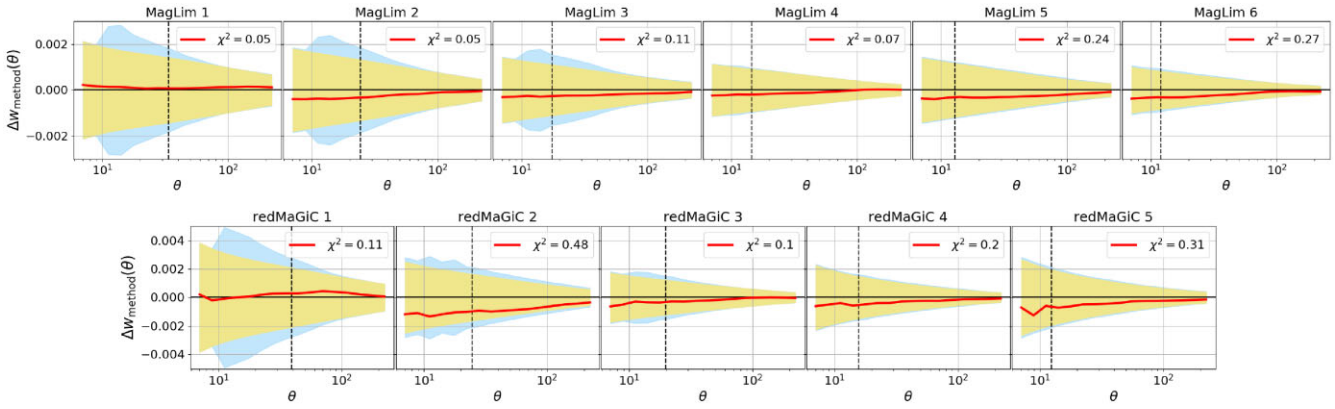

Figure 11. Method difference term $\Delta w_{\rm method}(\theta)$ in real data for MAGLIM (top row) and REDMAGIC (bottom row). The methods compared are ISD-PC<50 and ENET-PC<50 (red line). The light blue error bands correspond to the diagonal of the covariance with the additional systematic terms included, while the yellow ones correspond to the original analytical covariance.

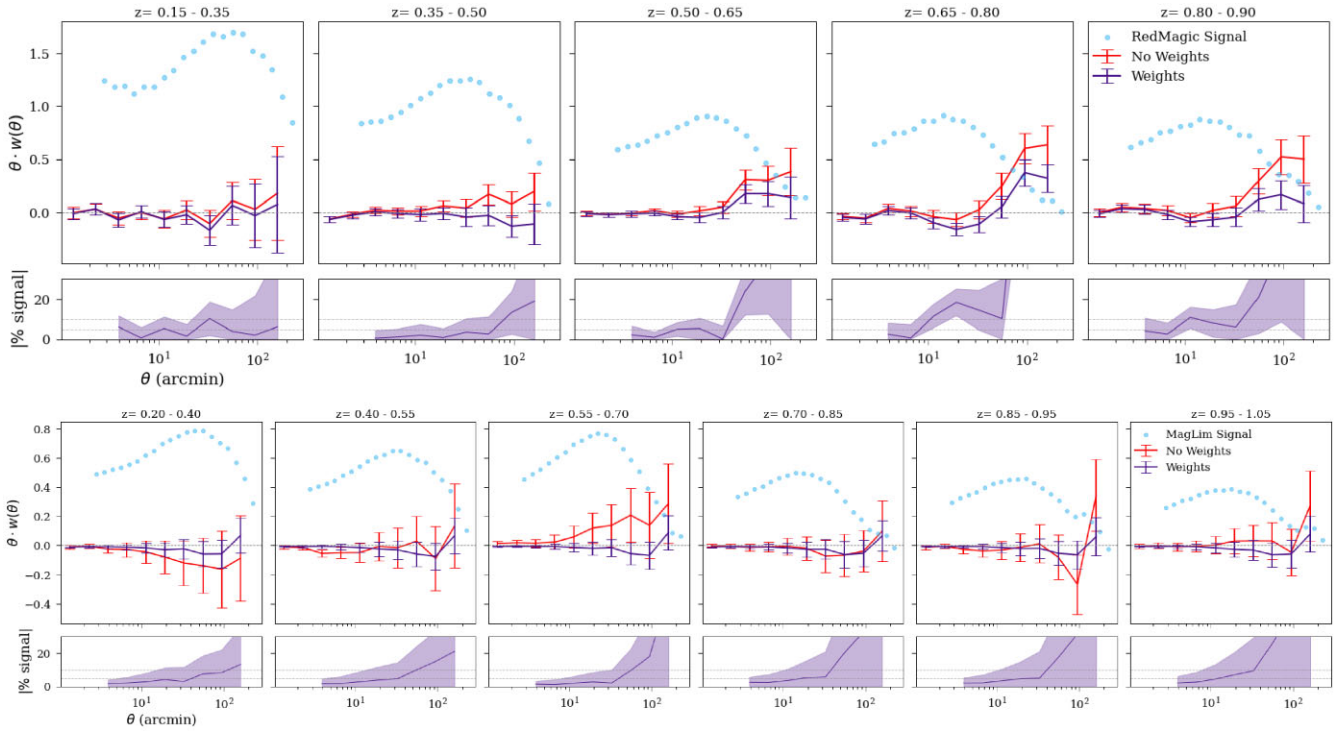

Figure 12. Cross-correlation between REDMAGIC (top panel) and MAGLIM (bottom panel) samples selected in data and produced with BALROG. The cross-correlations are shown in the top row of each panel, before weighting (red line) and after weighting (purple line) by SP maps effects, compared to the data $w(\theta)$ (blue points). The error bars have been obtained by jackknife re-sampling. The bottom row of each panel shows the relative difference (in per cent) between the cross-correlation signal and the auto-correlation one. A non-zero cross-correlation between the data samples and BALROG samples (which are injected and non-LSS distributed), would imply a pending, unknown systematic in the images, which would not have been corrected for. We see that the cross-correlation is zero within statistical errors. In general, all differences are compatible with zero and well below the statistical errors showing no clear evidence of uncorrected effects from imaging systematics, though we note that the points for REDMAGIC are noisier due to its lower number density.

(ii) We performed a principle component analysis of the 107 STD maps and used the principle components as an orthonormal basis for the decontamination procedure, i.e. ran ISD-PC107. We found good agreement with ENET-STD107 (and ENET-PC107), resulting in a reduction of the $w(\theta)$ amplitude. This was most pronounced in the first redshift bin of REDMAGIC, with a decrease in $w(\theta)$ of $\sim 1\sigma$.

(iii) We observed a significant overcorrection of $w(\theta)$ when computing ISD-PC107 weights from contaminated mocks. For this reason, we applied a cut-off to the number of PC maps, limiting it to the 50 PC maps with the highest S/N. We found that the resultant ISD-PC<50 weights produce little overcorrection and we add a systematic contribution to our error budget corresponding to the difference between ISD-PC<50 and ENET-PC<50. We also add a systematic contribution for the undercorrection observed on contaminated mocks using only the first 50 PC maps assuming the true contamination corresponds to the estimate of ENET-STD107.

(iv) We implemented a non-linear decontamination procedure using an NN, which also used different choices for the mask and

base set of STD maps. This resulted in differences in $w(\theta)$ that were much smaller than the observed discrepancy between galaxy clustering and gg-lensing.

(v) We cross-correlated both REDMAGIC and MAGLIM with their corresponding BALROG samples and we found no clear evidence of uncorrected contamination of known systematic templates common to both types of samples.

We note that the ISD-STD34 weights passed an extensive battery of validation tests, described in Section 7. However, after our findings and comparisons between ENET and ISD, we decided to use the ISD-PC<50 weights in the fiducial analysis.

Given these findings, we conclude that the anomalous high clustering amplitude of REDMAGIC sample is unlikely to be due to uncorrected contamination coming from any of our known templates nor from a linear combination of them. Because the clustering remains high when using higher order STD maps with ENET (after accounting for false correction bias) as well as using the neural net, we are unable to identify non-linear contamination from our SP maps as the cause (see Appendix E for additional tests). We performed a number of further exploratory tests such as more aggressive masking, including based on the leverage statistic (cf. Weaverdyck & Huterer 2021) and found $w(\theta)$ to be robust to these choices. Applying our fiducial decontamination procedure to MAGLIM does not show the same discrepancy between probes as does REDMAGIC.

9 CONCLUSIONS

We measure the angular two-point correlation of DES Y3 lens galaxies, and study the impact of systematic errors on these measurements. We use two lens samples: MAGLIM, a magnitude-limited sample with enhanced number density and reliable photometric redshifts (Porredon et al. 2021b), and REDMAGIC, a sample of luminous red galaxies (LRGs) selected by the algorithm described in Rozo et al. (2016), which also provides high-quality photometric redshifts. We extend the methodology employed in DES Y1 (Elvin-Poole et al. 2018), both for correcting the data and to ensure its robustness. A more thorough set of SP maps is used and we employ them directly and through the application of principal components analysis to the map set. Additionally, a new weight estimation method is used in parallel (ENET; Weaverdyck & Huterer 2021) and a cross-check of linearity assumptions is made with an NN framework based on recent literature (Rezaie et al. 2020). These steps help us to avoid possible blind spots in our validation methodology.

Our findings are as follows:

(i) The updated DES Y1 methodology, dubbed ISD, is able to successfully remove systematic contamination, as shown by validation tests on lognormal mocks (Figs 5 and 7) and data.

(ii) The ENET method is a viable alternative correction method to ISD. We evaluate several configurations and demonstrate that both methods are in agreement within statistical precision. To be sure that any residual difference is taken into account, we include a systematic uncertainty in the covariance matrix as the difference between the two results. This uncertainty is included in the final covariance that is used for cosmological constraints, after checking that it does not bias our results.

(iii) The decontamination procedure does not produce a significant bias in $w(\theta)$ or in the $\Omega_{\rm m} - b^i$ parameter space.

(iv) We find that survey properties have a significant impact on the recovered galaxy clustering signal, particularly at high redshifts, as compared to REDMAGIC Y1 results (Elvin-Poole et al. 2018). This contamination is corrected by applying the ISD method together with a principal component analysis of our survey sroperty maps. The same methodology is applied to both samples.

(v) We find an inconsistent clustering amplitude for the REDMAGIC sample when combined with other 2pt lensing probes. We study it from the point of view of the impact of SP maps, considering different methods, such as ISD and ENET, and different numbers, types and bases of SP maps. We find agreement between the weighted correlation functions yielded by each method within our errors. We also investigate weights from an NN weighting scheme. All our tests confirm that our systematics corrections are robust and the template maps used in this analysis do not explain the REDMAGIC internal inconsistency.

The results presented in this work have been optimized to be used for their combination with galaxy–galaxy lensing (Elvin-Poole et al. 2021; Pandey et al. 2021; Prat et al. 2021; Porredon et al. 2021a) and cosmic-shear (Amon et al. 2022; Secco et al. 2022) measurements to obtain the 3×2pt cosmological results from the DES Year 3 data (DES Collaboration et al. 2022), and constitutes one of the basic pillars for this measurement.

This work highlights the importance of adequate validation and cross-checking of this highly relevant step in the estimation of galaxy clustering, and builds upon several developments within the DES project and in the literature. For Y6, given the rapid developments in the field, we plan to approach the problem from the beginning with a variety of methodologies in mind, possibly considering multiregression approaches or assessing the feasibility of using a wider BALROG sample, making it part of the pipeline from the start now that the algorithm is fully developed. This will be coupled with possibly a multitiered unblinding approach with additional steps to be able to make decisions on investigating unusual results in internal consistency tests at different stages of the process. Additional work in parallel on the Y3 samples and SP maps will shed some light on possible details that the Y6 methodology will have to address, such as understanding the overcorrection produced by some maps or issues with the galaxy samples.

ACKNOWLEDGEMENTS

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

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

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

This 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.

DATA AVAILABILITY

The main product of this work is the maps of correcting weights applied to the galaxy clustering measurements of the DES Y3 lens galaxy samples. These weights are part of the data vectors that will be made available as part of the DES Y3 coordinated release at https://des.ncsa.illinois.edu/releases following publication of the DES Y3 Cosmology Results papers (https://www.darkenergysurvey.org/des-year-3-cosmology-results-papers/). The COSMOSIS software (Zuntz et al. 2015) is available at https://bitbucket.org/joezuntz/cosmosis/wiki/Home and the TREECORR package (Jarvis et al. 2004) can be found at https://rmjarvis.github.io/TreeCorr.

REFERENCES

Abbott T. M. C. et al., 2019, Phys. Rev. Lett., 122, 171301
Aihara H. et al., 2018, PASJ, 70, S4
Alam S. et al., 2017, MNRAS, 470, 2617
Alam S. et al., 2021, Phys. Rev. D, 103, 083533
Alonso D. et al., 2019, MNRAS, 484, 4127
Amon A. et al., 2022, Phys. Rev. D, 105, 023514
Bautista J. E. et al., 2018, ApJ, 863, 110
Bridle S. L. et al., 2002, MNRAS, 335, 1193
Carnero Rosell A. et al., 2021, MNRAS, 509, 778
Cawthon R. et al., 2020, preprint (arXiv:2012.12826)
Crocce M. et al., 2016, MNRAS, 455, 4301
de Jong J. T. A. et al., 2013, Exp. Astron, 35, 25
De Vicente J., Sánchez E., Sevilla-Noarbe I., 2016, MNRAS, 459, 3078
DeRose J. et al., 2021, MNRAS, preprint (arXiv: 2105.13547)
DES Collaboration, 2018a, Phys. Rev. D, 98, 043526
DES Collaboration, 2018b, ApJS, 239, 18
DES Collaboration et al., 2022, Phys. Rev. D, 105, 023520
Elsner F., Leistedt B., Peiris H. V., 2016, MNRAS, 456, 2095
Elsner F., Leistedt B., Peiris H. V., 2017, MNRAS, 465, 1847
Elvin-Poole J. et al., 2018, Phys. Rev. D, 98, 042006
Elvin-Poole J., MacCrann N. et al., 2021, MNRAS
Everett S. et al., 2022, Astrophys. J. Supp., 258, 15
Fang X. et al., 2020, J. Cosmol. Astropart. Phys., 2020, 010
Flaugher B. et al., 2015, AJ, 150, 150
Friedrich O. et al., 2021, MNRAS, 508, 3125
Fry J. N., Gaztanaga E., 1993, ApJ, 413, 447
Gaia Collaboration, 2020, A&A, 649, id.A1
Górski K. M. et al., 2005, ApJ, 622, 759
Handley W. J., Hobson M. P., Lasenby A. N., 2015a, MNRAS, 450, L61
Handley W. J., Hobson M. P., Lasenby A. N., 2015b, MNRAS, 453, 4384
Hartley W. G. et al., 2021, MNRAS, 509, 3547
Heymans C. et al., 2021, A&A, 646, A140
Howlett C. et al., 2012, J. Cosmol. Astropart. Phys., 1204, 027
Icaza-Lizaola M. et al., 2020, MNRAS, 492, 4189
Jarvis M., Bernstein G., Jain B., 2004, MNRAS, 352, 338
Jeffrey A. et al., 2021, MNRAS, 505, 4626
Johnston H. et al., 2021, A&A, 648, A98
Joudaki S. et al., 2018, MNRAS, 474, 4894
Kitanidis E. et al., 2020, MNRAS, 496, 2262
Krause E., Eifler T., 2017, MNRAS, 470, 2100
Krause E. et al., 2021, Phys. Rev. D, preprint (arXiv:2105.13548)
Kwan J. et al., 2017, MNRAS, 464, 4045
Landy S. D., Szalay A. S., 1993, ApJ, 412, 64
Laurent P. et al., 2017, J. Cosmol. Astropart. Phys., 2017, 017
Leistedt B., Peiris H. V., 2014, MNRAS, 444, 2
Leistedt B. et al., 2013, MNRAS, 435, 1857
Leistedt B. et al., 2016, ApJS, 226, 24
Lewis A., Bridle S., 2002, Phys. Rev. D, 66, 103511
Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473
Maller A. H. et al., 2005, ApJ, 619, 147
Morganson E. et al., 2018, PASP, 130, 074501
Morrison C. B., Hildebrandt H., 2015, MNRAS, 454, 3121
Myers A. D. et al., 2006, ApJ, 638, 622
Pandey S. et al., 2021, MNRAS, preprint (arXiv: 2105.13545)
Pedregosa F. et al., 2011, J. Machine Learning Res., 12, 2825
Perlmutter S. et al., 1999, ApJ, 517, 565
Planck CollaborationVI, 2014, A&A, 571, A11
Planck CollaborationXI, 2020, A&A, 641, A6
Porredon A. et al., 2021a, Phys. Rev. D, preprint (arXiv: 2105.13546)
Porredon A. et al., 2021b, Phys. Rev. D, 103, 043503
Prat J. et al., 2021, Phys. Rev. D, preprint (arXiv: 2105.13541)
Raichoor A. et al., 2021, MNRAS, 500, 3254
Rezaie M. et al., 2020, MNRAS, 495, 1613
Riess A. G. et al., 1998, AJ, 116, 1009
Ross A. J. et al., 2011, MNRAS, 417, 1350
Ross A. J. et al., 2017, MNRAS, 464, 1168
Ross A. J. et al., 2020, MNRAS, 498, 2354
Rozo E., et al., 2016, MNRAS, 461, 1431
Rybicki G. B., Press W. H., 1992, ApJ, 398, 169
Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525
Scranton R. et al., 2002, ApJ, 579, 48
Secco L. F. et al., 2022, Phys. Rev. D, 105, 023515
Sevilla-Noarbe I. et al., 2021, Astrophys. J. Suppl., 254, 24
Spergel D. N. et al., 2003, ApJS, 148, 175
Suchyta E. et al., 2016, MNRAS, 457, 786
Takahashi R. et al., 2012, ApJ, 761, 152
The Dark Energy Survey Collaboration, 2005, preprint(astro-ph/0510346)
Vakili M. et al., 2020, preprint (arXiv:2008.13154)
van Uitert E. et al., 2018, MNRAS, 476, 4662
Wagoner E. L. et al., 2021, MNRAS, 503, 4349
Weaverdyck N., Huterer D., 2021, MNRAS, 503, 5061
Xavier H. S., Abdalla F. B., Joachimi B., 2016, MNRAS, 459, 3693
Zonca A. et al., 2019, J. Open Source Softw., 4, 1298
Zuntz J. et al., 2015, Astron. Comput., 12, 45

APPENDIX A: LOGNORMAL MOCKS

The mocks used for the systematic analysis are 2D lognormal fields generated at a given power spectrum. We start by using CAMB (Lewis, Challinor & Lasenby 2000; Howlett et al. 2012) to obtain a matter power spectrum and project into a galaxy clustering angular power spectrum, $C_i^{gg}(l)$, following the theory modelling described in Krause et al. (2021). To produce this power spectrum, we assume our fiducial cosmology and fix the galaxy bias for each redshift bin to the values from the blind bias analysis (Table 1). Then, we use this power spectrum to generate a Gaussian random field of δ_g for each mock realization on a HEALPIX map (Górski et al. 2005) using the HEALPY package (Zonca et al. 2019). We then apply a lognormal transformation to the field following the methodology of Xavier et al. (2016). This uses a skewness parameter that was derived in Friedrich et al. (2021). We then transform the lognormal δ_g field to a galaxy number counts field, $N_{\rm gal}$, using the observed number count, $\bar{N}_o$, from the galaxy sample we want to reproduce and the relation:

$$N_{\rm gal} = \bar{N}_o \times (1 + \delta_g). \tag{A1}$$

We apply the angular mask to the full-sky realizations. In this way, the covariance matrices built from these mocks incorporate the same mask effects as the real data. In order to add shot noise, we finally Poisson sample the $N_{\rm gal}$ field.

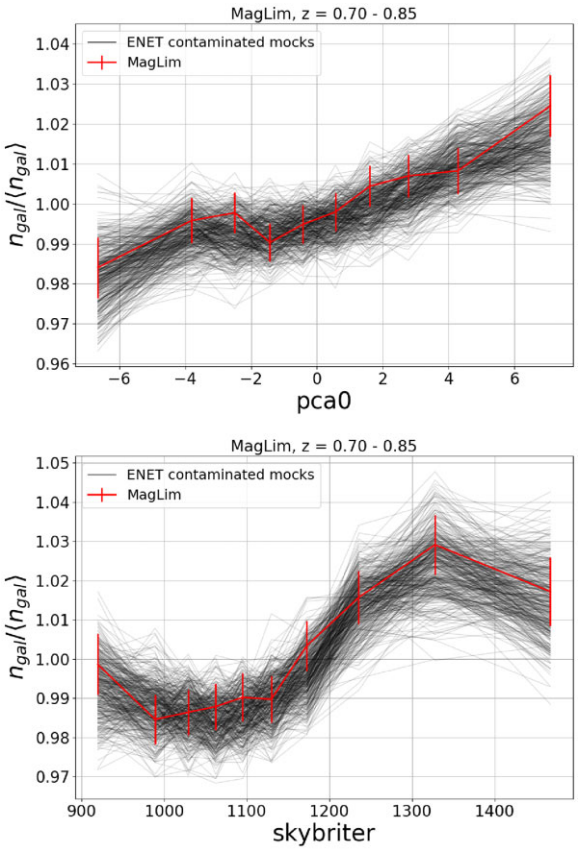

Figure A1. 1D relations for 400 MAGLIM ENET contaminated lognormal mocks (shaded black lines) compared with the data (red line). The top panel shows the 1D relations with the *pca*0 map at the fourth redshift bin of this sample, whereas the bottom panel shows the 1D relations with *skybrite* in the *r* band. The contamination observed on the data is well reproduced by these mocks. The error bars are obtained from the uncontaminated mocks used to calculate the 1D significance.

As we mention in Section 7, we also create a set of lognormal mocks contaminated by survey properties systematics, so we can look for biases introduced by ISD and check their impact on the measurements. We imprint contamination on the lognormal mocks by multiplying the galaxy number counts field by the inverse of the weight map derived from the data, which is

$$N^p_{\rm gal,\,mock} \rightarrow N^p_{\rm gal,\,mock} \times \frac{1}{w^p}. \tag{A2}$$

This step is applied before Poisson sampling the galaxy field. We produce a set of 400 contaminated lognormal mocks following this procedure using weights derived from ENET-STD107, as is mentioned in Section 7. We check that the 1D relations of these mocks reproduce in shape and amplitude those observed on the data. An example of this can be seen in Fig. A1.

APPENDIX B: INTERNAL CONSISTENCY TESTS: ESTIMATOR BIAS TEST

In addition to the tests described in Section 7, we perform an internal consistency test that seeks to confirm no bias in $w(\theta)$ is introduced by ISD under idealized circumstances. For this test, we contaminate and correct for the same list of SP maps, demonstrating the Landy–Szalay estimator can recover a negligibly biased signal. Since the focus of this test is the $w(\theta)$ estimator itself when applied to weighted

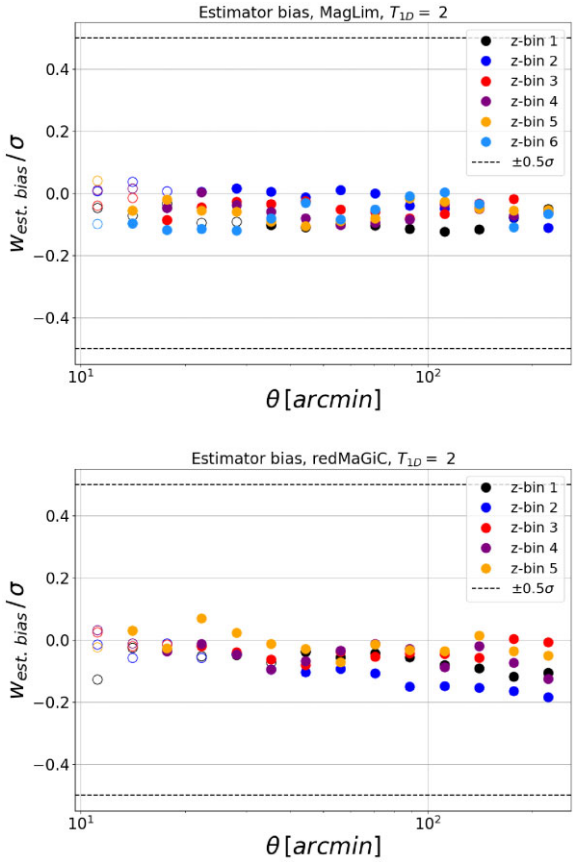

Figure B1. Estimator bias for MAGLIM (top panel) and REDMAGIC (bottom panel). The negative values are due to small level of overcorrection. Empty dots correspond to the scales excluded for each redshift bin. As can be seen, we find no evidence of bias in $w(\theta)$ introduced by the ISD methodology and the Landy–Szalay estimator under idealized circumstances at any angular scale nor any redshift bin of both galaxy samples.

data, independent of the origin of these weights, we conduct it using weights from a preliminary run of ISD on the standard SP maps, with the same threshold that we use to obtain the weights from the data, $T_{1D} = 2$. To get the magnitude of this potential bias, we defined

$$w_{\text{est. bias}}(\theta) = \frac{1}{N}\left(\sum_{i=n}^{N} w_{\text{dec},\,i} - \sum_{j=1}^{N} w_{\text{unc},\,j}\right)(\theta), \tag{B1}$$

where $w_{\text{unc},\,i}$ are the correlation functions from uncontaminated mocks, and $w_{\text{dec},\,i}$ are those from decontaminated mocks and $N = 1000$ mock realizations. Fig. B1 showcases the values of $w_{\text{est. bias}}(\theta)$. As it can be seen, we see no indication of estimator bias for both lens samples at every redshift bin. This demonstrates that the combination of our weighting methodology with the Landy–Szalay estimator for $w(\theta)$ does not induce any bias on our measurements when the list of contaminating SP maps is known.

APPENDIX C: POLYNOMIAL FITS FOR $\Delta w_{\text{method}}(\theta)$

The additional covariance term described in Section 8.4 depends on the difference between $w(\theta)$ measured with two different systematics correction methods, $\Delta w_{\text{method}}(\theta)$. As $\Delta w_{\text{method}}(\theta)$ is measured on real data, it contains the same noise as the $w(\theta)$ data vector being used for cosmological inference. To avoid adding this noise to the covariance term, we fit a flexible polynomial to the two $w(\theta)$ measurements in the form

$$w_{\text{polyfit}}(\theta) = \sum_{i=-3}^{+3} B_i \theta^i, \tag{C1}$$

where B_i are the coefficients to be fitted. The best-fitting polynomials are shown in Fig. C1. We find this polynomial to be a good fit to the data, and the difference between measured correlation functions matches the difference in fitted polynomials well.

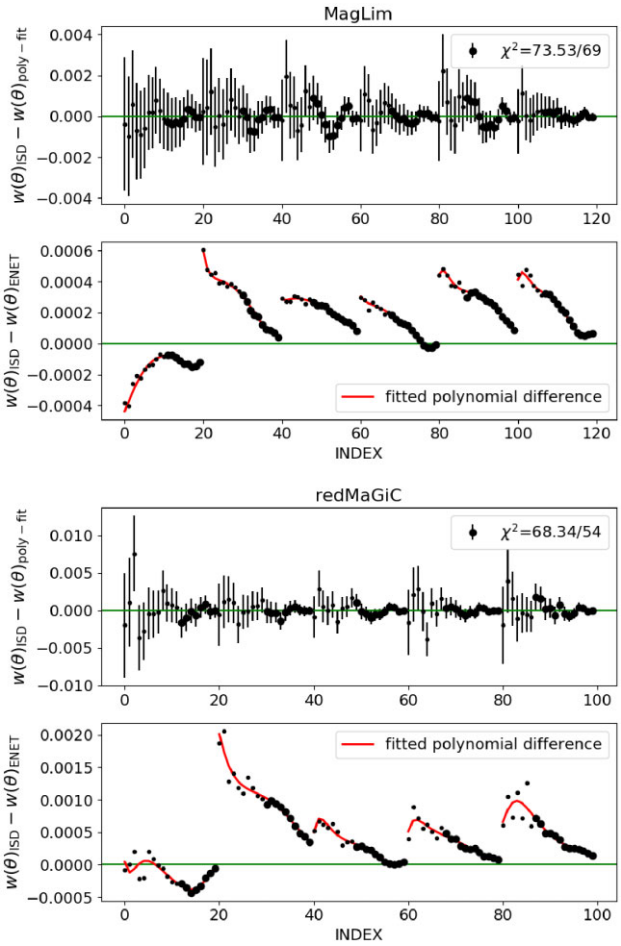

Figure C1. Polynomial fits to $w(\theta)$ data used in estimating the systematic terms in the $w(\theta)$ covariance in Section 8.4. The first and third panels show the fit residuals to the fiducial $w(\theta)$ measurements for each sample. The second and fourth panels show the difference between the polynomial fits of the two correction methods considered in these terms, ISD and ENET, both with the first 50 principle component template maps. The bold points are the data included by the scale cuts and included in the fit and χ^2 calculations.

APPENDIX D: PRINCIPAL COMPONENT MAPS CUT-OFF

In Section 8, we describe a set of systematics weights using only the first 50 principle component maps labelled ISD-PC<50, which are used as the fiducial weights in the cosmology analysis. In this Appendix, we provide some further justification for this choice.

In order to test for the correlation of real large-scale structure with the weight maps, we cross-correlate the convergence, κ, maps from Jeffrey et al. (2021) with the weight maps obtained using different methods, ISD-STD34, ISD-PC107, and ISD-PC<50. We correlate with the convergence map for the third tomographic source bin due to the large overlap between its lensing kernel and the lens sample. In the absence of systematics in the κ maps, we do not expect there to be correlations between the SP or weight maps and the convergence maps. We show these correlations in Fig. D2 for the five REDMAGIC tomographic bins (the error bars are estimated using jackknife methodology using 150 patches). We find that while ISD run on only the 34 representative STD maps does not correlate with the convergence maps, we obtain a large correlation with the weight maps

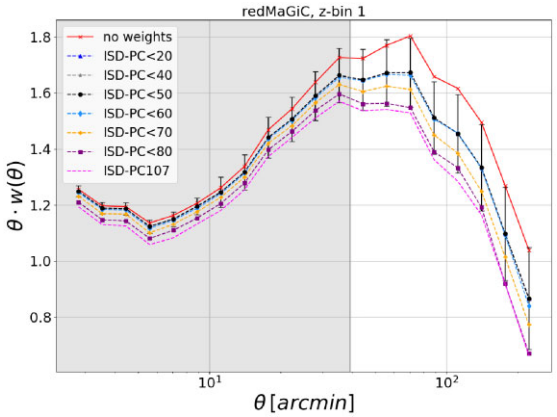

Figure D1. Clustering amplitude at the first redshift bin of REDMAGIC for several PC cut-offs, ISD-PC<n. The solid red line corresponds to the unweighted data and the dashed magenta one to the weights obtained from ISD-PC107, which lead to overcorrection. It can be seen how around $n = 50$ the $w(\theta)$ amplitudes converge.

using all the PC maps, pointing to potential leakage of cosmological structure in these weights, either from chance correlation or real large-scale structure leaking into the high PC maps. These cross-correlations are calculated with the weight maps at $N_{\rm side} = 4096$ and κ at $N_{\rm side} = 1024$, so small-scale noise correlations are ruled out (moreover, the cross-correlations have been evaluated at coarser resolutions, finding similar results).

To mitigate any correlation with real large-scale structure, we restrict the weight estimation to use only the first n PC maps. First of all, to ensure that all dominant features of the SP maps are taken into account, we look at the amount of variance captured up to each component. This is shown on Fig. D3. Based on this, we use $n = 50$ as a starting point. PC maps up to this component explain $\sim$98 per cent of the total variance and we consider that it represents a balance between including too many maps, resulting in overcorrection, and discarding too many of them, so we risk not accounting for enough contaminants. Then, we obtain the ISD-PC<50 weights and we observe that these weights cause no significant overcorrection on contaminated mocks, as explained in Section 8. After this, we verify that the ISD-PC<50 weights show negligible levels of cross-correlation with κ, similar to those from ISD-STD34. Moreover, the recovered correlation function from these weights is in excellent agreement with that from ISD-STD34 weights, as it is shown on Fig. 9.

In order to make the rejection of PC maps that could be causing the overcorrection as specific as possible, we cross-correlate κ directly with the maps that contribute to the overcorrecting ISD-PC107 weights (according to the multiplicative way of ISD to make weights). However, we do not identify any individual map or family of maps clearly causing the excess correlation. In general, the PC maps that have the highest κ correlation are the highest principal components (which have the smallest contribution to the total variance of the STD maps). Given this, we decide to test removing all PC maps above a given component. We test multiple cut-offs with PC<n, evaluating their clustering amplitudes, as it is shown in Fig. D1. We find that the clustering amplitudes yielded by the ISD-PC<n weights with n between 20 and 60 converge to similar values, while for higher n, it jumps abruptly to lower amplitudes. This result, together with the large amount of variance contained up to PC<50 and the impossibility of flagging a specific set of PC maps among the highest components as the culprit ones of the overcorrection, motivates the choice of $n = 50$ as our final cut-off.

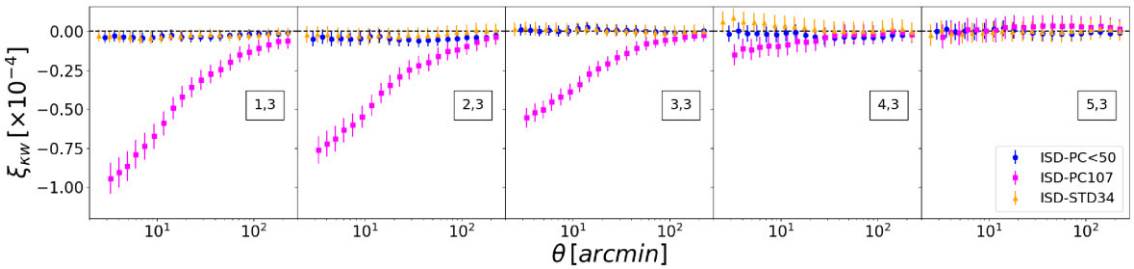

Figure D2. Cross-correlation of weight maps from different configurations of ISD with the convergence field, κ. The error bars are calculated using jackknife with 150 patches. It can be seen how the ISD-PC107 weights cross-correlate significantly with κ, while the weights from the other two configurations do not. This suggests that the high PC template maps may correlate with LSS. An off-set has been added to the x-axis points for better visualization.

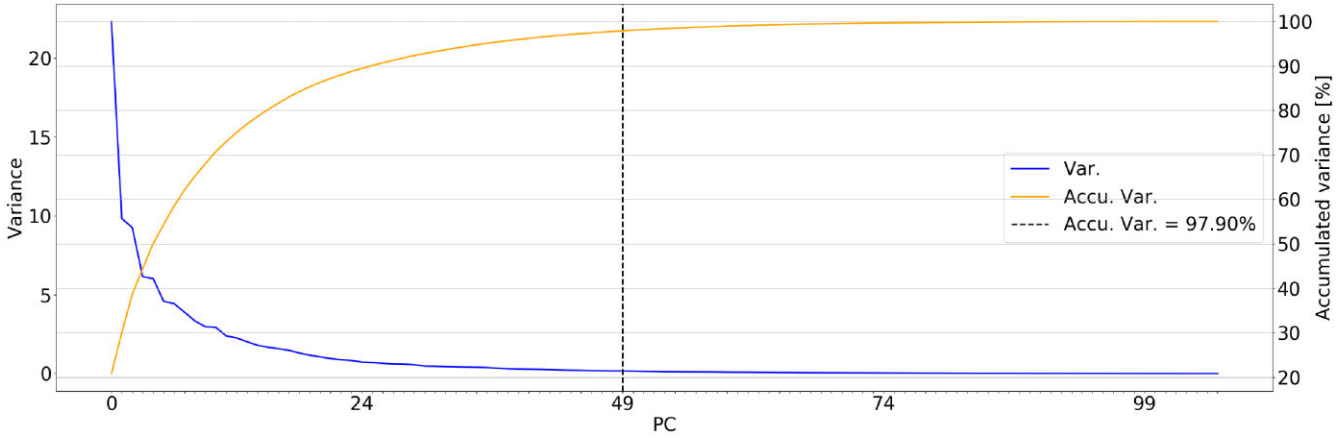

Figure D3. Variance of each PC map (blue line) and per cent of accumulated variance (orange line). For the principal component map, 49 the accumulated variance is ∼98 per cent, so the remaining maps are compatible with noise.

APPENDIX E: NON-LINEAR CONTAMINATION WITH ISD

In order to look for non-linear contamination still present on the data after applying weights, we evaluate the distribution of $\chi^2_{\rm null}$ values from the 1D relations of the ISD-PC<50 weighted data. This kind of contamination could be undetected when using a linear model, as ISD does, and would result on high $\chi^2_{\rm null}$ values. In Fig. E1, we show the values obtained for REDMAGIC. The distributions obtained for each redshift bin are not significantly different from a χ^2 with 10 degrees of freedom (number of 1D bins used). We obtain similar results for the MAGLIM sample. Therefore, we find no clear evidence of the presence of non-linear contamination in our weighted data that could have been unaccounted for.

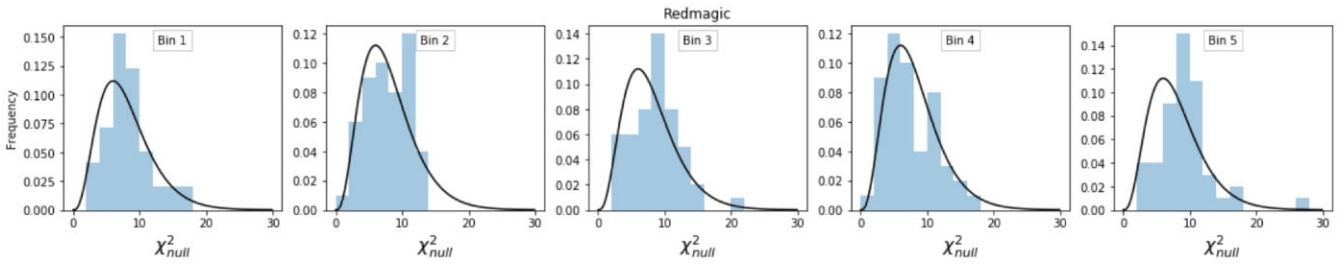

Figure E1. $\chi^2_{\rm null}$ distributions (blue histograms) for the ISD-PC<50 weighted REDMAGIC sample compared with a χ^2 with 10 degrees of freedom (black lines). Given the good agreement between both distributions, we find no clear evidence of deviations from linearity in the 1D relations of the weighted data. We find similar results for MAGLIM sample.

[1]*Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Av. Complutense, 22, E- 28040 Madrid, Spain*
[2]*Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA*
[3]*Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA*
[4]*Department of Physics, The Ohio State University, Columbus, OH 43210, USA*
[5]*Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain*
[6]*Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain*
[7]*Instituto de Astrofisica de Canarias, E-38205 La Laguna, Tenerife, Spain*
[8]*Laboratório Interinstitucional de e-Astronomia - LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ – 20921-400, Brazil*
[9]*Universidad de La Laguna, Dpto. Astrofísica, E-38206 La Laguna, Tenerife, Spain*
[10]*Instituto de Física Teórica, Universidade Estadual Paulista, São Paulo, Brazil*
[11]*Instituto de Fisica Teorica UAM/CSIC, Universidad Autonoma de Madrid, E-28049 Madrid, Spain*
[12]*Physics Department, 2320 Chamberlin Hall, University of Wisconsin-Madison, 1150 University Avenue Madison, WI 53706-1390, USA*
[13]*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*
[14]*Department of Physics, Northeastern University, Boston, MA 02115, USA*
[15]*Laboratory of Astrophysics, École Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, CH-1290 Versoix, Switzerland*
[16]*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*
[17]*Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15312, USA*
[18]*NSF AI Planning Institute for Physics of the Future, Carnegie Mellon University, Pittsburgh, PA 15213, USA*
[19]*Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA*
[20]*Department of Astronomy/Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721-0065, USA*
[21]*Institute of Theoretical Astrophysics, University of Oslo. PO Box 1029 Blindern, NO-0315 Oslo, Norway*
[22]*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109, USA*
[23]*Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[24]*Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, E-08193 Bellaterra (Barcelona), Spain*
[25]*Center for Astrophysical Surveys, National Center for Supercomputing Applications, 1205 West Clark St, Urbana, IL 61801, USA*
[26]*Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801, USA*
[27]*Department of Astronomy, University of Geneva, ch. d'Écogia 16, CH-1290 Versoix, Switzerland*
[28]*Fermi National Accelerator Laboratory, PO Box 500, Batavia, IL 60510, USA*
[29]*Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, UK*
[30]*Kavli Institute for Particle Astrophysics & Cosmology, PO Box 2450, Stanford University, Stanford, CA 94305, USA*
[31]*Kavli Institute for the Physics and Mathematics of the Universe (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan*
[32]*Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA*
[33]*Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA*
[34]*ICTP South American Institute for Fundamental Research Instituto de Física Teórica, Universidade Estadual Paulista, Al. Santos 647, São Paulo, Brazil*
[35]*Department of Physics, University of Arizona, Tucson, AZ 85721, USA*
[36]*SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA*
[37]*Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA*
[38]*Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth, PO1 3FX, UK*
[39]*Institute for Astronomy, University of Hawai'i, 2680 Woodlawn Drive, Honolulu, HI 96822, USA*
[40]*CNRS, UMR 7095, Institut d'Astrophysique de Paris, F-75014 Paris, France*
[41]*Sorbonne Universités, UPMC Univ Paris 06, UMR 7095, Institut d'Astrophysique de Paris, F-75014 Paris, France*
[42]*Department of Physics and Astronomy, Pevensey Building, University of Sussex, Brighton BN1 9QH, UK*
[43]*Department of Physics and Astronomy, University College London Gower Street, London WC1E 6BT, UK*
[44]*Jodrell Bank Center for Astrophysics, School of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK*
[45]*University of Nottingham, School of Physics and Astronomy, Nottingham NG7 2RD, UK*
[46]*Astronomy Unit, Department of Physics, University of Trieste, via Tiepolo 11, I-34131 Trieste, Italy*
[47]*INAF-Osservatorio Astronomico di Trieste, via G. B. Tiepolo 11, I-34143 Trieste, Italy*
[48]*Institute for Fundamental Physics of the Universe, Via Beirut 2, I-34014 Trieste, Italy*
[49]*Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro RJ 20921-400, Brazil*
[50]*Department of Physics, IIT Hyderabad, Kandi, Telangana 502285, India*
[51]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[52]*School of Mathematics and Physics, University of Queensland, Brisbane, QLD 4072, Australia*
[53]*Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA*
[54]*Australian Astronomical Optics, Macquarie University, North Ryde, NSW 2113, Australia*
[55]*Lowell Observatory, 1400 Mars Hill Rd, Flagstaff, AZ 86001, USA*
[56]*Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, CP 66318, São Paulo, SP 05314-970, Brazil*
[57]*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*
[58]*Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA*
[59]*Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA*
[60]*Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain*
[61]*Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, D-81679 Munich, Germany*
[62]*Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, D-85748 Garching, Germany*
[63]*School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK*
[64]*Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA*
[65]*Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstr. 1, D-81679 München, Germany*

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