Weak-lensing mass calibration of redMaPPer galaxy clusters in Dark Energy Survey Science Verification data

P. Melchior,[1][*] D. Gruen,[2,3][*][†] T. McClintock,[4] T. N. Varga,[5,6] E. Sheldon,[7] E. Rozo,[4]
A. Amara,[8] M. R. Becker,[9,2] B. A. Benson,[10,11] A. Bermeo,[12] S. L. Bridle,[13]
J. Clampitt,[14] J. P. Dietrich,[15,16] W. G. Hartley,[17,8] D. Hollowood,[18] B. Jain,[14]
M. Jarvis,[14] T. Jeltema,[18] T. Kacprzak,[8] N. MacCrann,[13] E. S. Rykoff,[2,3] A. Saro,[16]
E. Suchyta,[19] M. A. Troxel,[20,21] J. Zuntz,[13] C. Bonnett,[22] A. A. Plazas,[23]
T. M. C. Abbott,[24] F. B. Abdalla,[17,25] J. Annis,[10] A. Benoit-Lévy,[26,17,27]
G. M. Bernstein,[14] E. Bertin,[26,27] D. Brooks,[17] E. Buckley-Geer,[10]
A. Carnero Rosell,[28,29] M. Carrasco Kind,[30,31] J. Carretero,[32,22] C. E. Cunha,[2]
C. B. D'Andrea,[33,34] L. N. da Costa,[28,29] S. Desai,[35] T. F. Eifler,[23] B. Flaugher,[10]
P. Fosalba,[32] J. García-Bellido,[36] E. Gaztanaga,[32] D. W. Gerdes,[37] R. A. Gruendl,[30,31]
J. Gschwend,[28,29] G. Gutierrez,[10] K. Honscheid,[20,21] D. J. James,[38,24] D. Kirk,[17]
E. Krause,[2] K. Kuehn,[39] N. Kuropatkin,[10] O. Lahav,[17] M. Lima,[40,28]
M. A. G. Maia,[28,29] M. March,[14] P. Martini,[20,41] F. Menanteau,[30,31] C. J. Miller,[42,37]
R. Miquel,[43,22] J. J. Mohr,[5,15,16] R. C. Nichol,[33] R. Ogando,[28,29] A. K. Romer,[12]
E. Sanchez,[44] V. Scarpine,[10] I. Sevilla-Noarbe,[44] R. C. Smith,[24] M. Soares-Santos,[10]
F. Sobreira,[28,45] M. E. C. Swanson,[31] G. Tarle,[37] D. Thomas,[33] A. R. Walker,[24]
J. Weller,[5,6,15] Y. Zhang[10] and (The DES Collaboration)

Affiliations are listed at the end of the paper

ABSTRACT

We use weak-lensing shear measurements to determine the mean mass of optically selected galaxy clusters in Dark Energy Survey Science Verification data. In a blinded analysis, we split the sample of more than 8000 redMaPPer clusters into 15 subsets, spanning ranges in the richness parameter $5 \leq \lambda \leq 180$ and redshift $0.2 \leq z \leq 0.8$, and fit the averaged mass density contrast profiles with a model that accounts for seven distinct sources of systematic uncertainty: shear measurement and photometric redshift errors; cluster-member contamination; miscentring; deviations from the NFW halo profile; halo triaxiality and line-of-sight projections. We combine the inferred cluster masses to estimate the joint scaling relation between mass, richness and redshift, $\mathcal{M}(\lambda, z) \propto M_0 \lambda^F (1+z)^G$. We find $M_0 \equiv \langle M_{200\mathrm{m}} \,|\, \lambda = 30, z = 0.5 \rangle = [2.35 \pm 0.22 \,(\mathrm{stat}) \pm 0.12 \,(\mathrm{sys})] \times 10^{14} \, \mathrm{M_\odot}$, with $F = 1.12 \pm 0.20 \,(\mathrm{stat}) \pm 0.06 \,(\mathrm{sys})$ and $G = 0.18 \pm 0.75 \,(\mathrm{stat}) \pm 0.24 \,(\mathrm{sys})$. The amplitude of the mass–richness relation is in excellent agreement with the weak-lensing calibration of redMaPPer clusters in SDSS by Simet et al. and with the Saro et al. calibration based on abundance matching of SPT-detected clusters. Our results extend the redshift range over

which the mass–richness relation of redMaPPer clusters has been calibrated with weak lensing from $z \leq 0.3$ to $z \leq 0.8$. Calibration uncertainties of shear measurements and photometric redshift estimates dominate our systematic error budget and require substantial improvements for forthcoming studies.

Key words: gravitational lensing: weak – galaxies: clusters: general – cosmology: observations.

1 INTRODUCTION

The growth of massive structures and the cosmic expansion rate depend directly on the energy constituents of the Universe and the behaviour of gravity at a range of scales. In the currently favoured cosmological model, the energy density at the present epoch is dominated by dark matter and dark energy, with space–time evolving according to the standard theory of gravity, general relativity. Within this model, the number of haloes of a given mass and the history of cosmic expansion depend sensitively on the relative amount and detailed properties of both dark matter and dark energy. Alternative theories of gravity may make different predictions for the number density of haloes at a given cosmic epoch for the same expansion history (see e.g. Schmidt et al. 2009). Thus, much can be learned about cosmology, and potentially about gravity itself, by studying the abundance of massive structures as a function of their mass and redshift (Allen, Evrard & Mantz 2011; Weinberg et al. 2013).

Clusters of galaxies are thought to directly correspond to the largest dark matter haloes, the number density of which is particularly sensitive to the dark energy. However, clusters are typically identified not by the total mass of their halo, but by a related observable. Thus, cosmological inference from cluster abundance requires a cluster catalogue with measurements of some observable with a well-understood selection function, a theoretical prediction for the abundance of haloes as a function of mass and redshift for different cosmologies, and, crucially, a mass–observable relation (MOR) that connects the observable and true mass of a halo.

The abundance of dark matter haloes is, in principle, predictable for a given cosmological model purely from *N*-body simulations (e.g. Tinker et al. 2008; Bocquet et al. 2016). However, most cluster observables that are readily measured in large surveys are a manifestation of complex astrophysical processes. Some common observables are counts of galaxies above a threshold luminosity (Koester et al. 2007; Rykoff et al. 2014), X-ray emission from hot gas (Piffaretti et al. 2011; Mehrtens et al. 2012) and Compton scattering of cosmic microwave background photons off electrons in that same gas (Hasselfield et al. 2013; Bleem et al. 2015; Planck Collaboration XXVII 2016). Predicting the corresponding MOR for these observables from first principles, or from simulations, is not straightforward. At this time, it is more practical to determine the relationship empirically based on an observable proxy of the cluster mass. These empirical measurements are similarly challenging; uncertainties in the MOR have in fact become the main impediment to reliable inference of cosmological parameters from cluster abundances (e.g. Rozo et al. 2010; Bocquet et al. 2015; Mantz et al. 2015; Planck Collaboration XXIV 2016).

Currently, the conceptually cleanest method for calibrating cluster masses is gravitational lensing, the deflection of light from background objects due to all matter contained in a foreground cluster (e.g. Johnston et al. 2007b; Gruen et al. 2014; von der Linden et al.

2014b; Hoekstra et al. 2015; Okabe & Smith 2016; van Uitert et al. 2016; Simet et al. 2017b). In this work, we will use lensing to measure the mass of clusters found in the Dark Energy Survey (DES; Dark Energy Survey Collaboration 2005, 2016).

The Dark Energy Survey is a 5000 deg^2 survey of the southern sky using the 4-metre Blanco Telescope and the Dark Energy Camera (Flaugher et al. 2015) at the Cerro Tololo Inter-American Observatory. The primary goal of the survey is to constrain the distribution of dark matter in the Universe, and the amount and properties of dark energy, including its equation of state. Due to the large area, depth and image quality of DES, the data will support optical identification of a large number ($\approx$100 000) of galaxy clusters and groups up to a redshift $z \approx 1$. The potential for probing cosmology with these optically selected clusters can be realized only if the MOR is well understood.

In this paper, we present the first ensemble, or 'stacked', lensing measurements of optically selected clusters from DES. We use the redMaPPer cluster catalogue (Rykoff et al. 2016), generated from the Science Verification (SV) data taken before the first official DES observing season. For the lensing measurements, we use catalogues of lensed galaxies constructed from the same data (Jarvis et al. 2016).

The goal of this work is twofold. First, we measure the statistical relationship between the number of galaxies in a cluster and the underlying halo mass – i.e. the MOR – and compare our findings to the results in the literature. Secondly, we develop and test a new analysis pipeline, which we will apply to the much larger DES data sets currently being acquired and processed. We especially seek to fully account for systematic effects, considering biases in shear and photometric redshift measurements, cluster member contamination in the lensed background source sample, miscentring and triaxiality of clusters, projection of multiple clusters along the line of sight and deviations of halo profiles from the analytical form.

The structure of this paper is as follows. In Section 2, we introduce the data used in this work. In Section 3, we describe our methodology for obtaining ensemble cluster density profiles from stacked weak-lensing shear measurements. A comprehensive set of tests and corrections for systematic effects is presented in Section 4. The model of the lensing data and the inferred stacked cluster masses are given in Section 5. The main result, the mass–richness–redshift relation of redMaPPer clusters in DES, is presented in Section 6. We compare our results to other published works in the literature in Section 7 and conclude in Section 8.

For the purpose of this analysis, we assume a flat Λ cold dark matter (ΛCDM) cosmology with $\Omega_m = 0.3$ and $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$. Distances and masses, unless otherwise noted, are defined as physical quantities with this choice of cosmology, rather than in comoving coordinates. We denote the mass inside spheres around the cluster centre as M_{200m}, corresponding to an overdensity factor of 200 with respect to the mean matter density at the cluster redshift.
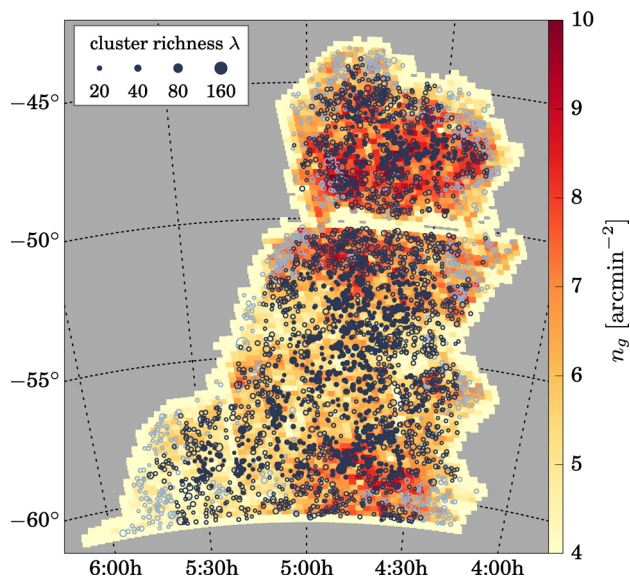
Figure 1. The distribution of redMaPPer clusters (Section 2.1) in the DES SV footprint, overlaid on the density of galaxies used for the weak-lensing analysis (Section 2.2). The size of the cluster markers is scaled by the richness λ, i.e. the number of member galaxies. Clusters that are surrounded by galaxies over the entire radial range of 30 Mpc are shown as filled circles (other clusters are denoted by open circles). For the purpose of clarity, clusters with $\lambda < 10$ are omitted.

2 THE DES SCIENCE VERIFICATION DATA

DES began taking official survey data in 2013 August (Diehl et al. 2014). Before this, a small SV survey was conducted from 2012 November to 2013 February. For this work, we restricted our measurements to the largest contiguous portion of the SV area, dubbed 'SPT-East' and shown in Fig. 1, an area of approximately 139 deg^2 in the eastern part of the region observed by the 2500-square-degree South Pole Telescope Sunyaev–Zel'dovich Survey (Carlstrom et al. 2011). We briefly introduce the three main data products utilized in this work and refer to the respective publications for details. We want to emphasize that all catalogues used in this work have already been made public[1] to facilitate external review and further scientific exploration.

2.1 Cluster catalogue

We rely on photometrically selected clusters identified in the DES SV data set using the redMaPPer cluster finding algorithm (Rykoff et al. 2014). The resulting cluster catalogue was presented in Rykoff et al. (2016), and the distribution in the footprint is shown in Fig. 1. We note that redMaPPer has had multiple public releases corresponding to different versions of the algorithm. The catalogue we employ corresponds to version 6.3.3 of the algorithm.

Briefly, redMaPPer identifies galaxy clusters as overdensities of red-sequence galaxies. It iteratively trains a model of the red sequence as a function of redshift, and utilizes this model to assign a cluster membership probability to every galaxy in the vicinity of a cluster. The cluster richness λ is the sum of the membership probabilities of the galaxies within a cluster radius R_λ. The radius scales with richness as $R_\lambda = 1.0(\lambda/100)^{0.2}h^{-1}$ Mpc, a choice that was found to minimize the scatter in the MOR (Rykoff et al. 2012).
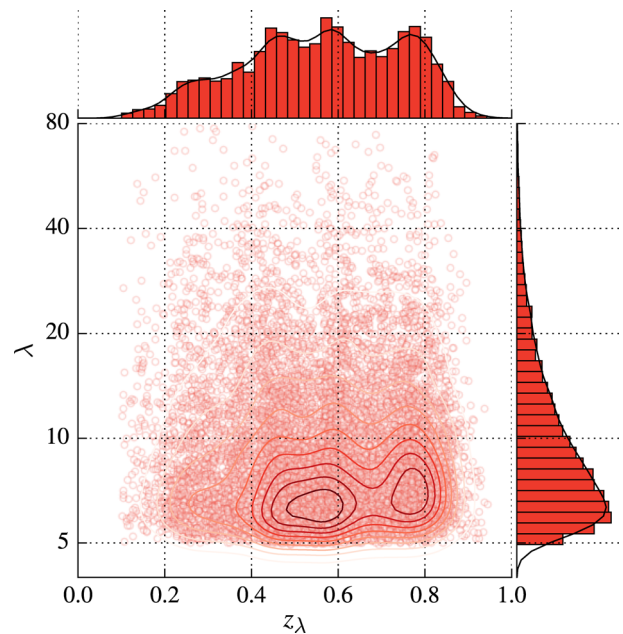
Figure 2. Redshift–richness distribution of redMaPPer clusters in the DES SV catalogue, overlaid with density contours to highlight the densest regions. At the top and on the right are histograms of the projected quantities, z_λ and λ, respectively, with smooth kernel density estimates (*black lines*).

Because of the interdependence of λ and R_λ, richness estimation involves finding the self-consistent pair of these parameters (cf. Rozo et al. 2009, their section 3.2).

Cluster central galaxies are chosen using a probabilistic approach that weights not just galaxy luminosity, but also local galaxy density, and demands consistency between the photometric redshift of the central galaxy and the cluster redshift. The mean of the redMaPPer centring probabilities is 0.81, i.e. we expect 81 per cent of the clusters to be properly centred, which is in good agreement with estimates from *XMM–Newton*, *Chandra* and SPT cluster detections (Rykoff et al. 2016, their section 6.2.4).

High-redshift clusters can be identified only in the deepest survey regions, so the redshift range probed by the catalogue varies from location to location. Specifically, at any given location, the maximum redshift z_{max} is set by requiring that the survey depth at that location be sufficient to ensure a 10σ detection of redMaPPer member galaxies brighter than $0.2\,L_\star$ for a cluster at redshift z_{max}.

The distribution of richness and redshift for the cluster sample is shown in Fig. 2. Rykoff et al. (2016) show that the photometric redshift performance of redMaPPer on the DES SV region is $\sigma_{z_\lambda}/(1+z) \lesssim 0.01$, while the abundance and redshift evolution of the sample suggest that a richness $\lambda = 20$ corresponds to a halo mass $M_{500c} \approx 10^{14}\,M_\odot$ or $M_{200m} \approx 1.8 \times 10^{14}\,M_\odot$, with a variation of mass with richness expected to be roughly linear.

2.2 Shear catalogue

We use the official DES shear catalogues presented in Jarvis et al. (2016). Two separate catalogues were created, NGMIX (Sheldon 2014) and IM3SHAPE (Zuntz et al. 2013). These catalogues were found to be consistent when selection effects were taken into account (Jarvis et al. 2016). Both catalogues are adequate for the purposes of estimating shear correlation functions, including tangential shear analyses such as the one pursued in this work. Here, we adopt the NGMIX catalogue, which has a higher effective number

density (5.7 arcmin^{-2}) of sources because it combines the image data from three bands (r, i, z) instead of relying on any single band.

The NGMIX catalogue was generated using an implementation of the 'lensfit' algorithm (Miller et al. 2007), a galaxy model fitting technique. An exponential model was fit to each galaxy image, and the full likelihood surface of the model parameter space was explored. The ellipticity statistic e was taken to be the mean $\langle e_\mu \rangle$ across this surface, where μ denotes the two ellipticity components.

During exploration of the likelihood surface, a centrally concentrated prior was applied to the ellipticity. This prior stabilizes and limits the exploration of the likelihood. However, application of an ellipticity prior results in a shear calibration error: a naive sum over the ellipticities does not recover the true applied shear for low-S/N measurements. Essentially, the sensitivity of the ellipticity as a shear estimator is reduced due to suppressing high ellipticity regions of the parameter space. A correction for this calibration bias was calculated, called the 'sensitivity' s (Jarvis et al. 2016, their equation 7.10), which yields a shear estimator of the following form:

$$\tilde{g}_\mu = \frac{\sum_i e_{i,\mu}}{\sum_i s_{i,\mu}}, \tag{1}$$

for all sources i that experienced the same constant shear. It is important to note that for any single galaxy, s is a very noisy quantity. The sensitivity correction is therefore applied only for an entire source ensemble. Also, while s has two components, it does not transform as a polarization. For tangential shear measurements, we rotate the galaxy shapes into the tangential frame around each cluster, but this cannot be done for the sensitivities. Instead, we simply average the two sensitivity components into one,

$$s_i = \frac{1}{2}(s_{i,1} + s_{i,2}). \tag{2}$$

We have verified in simulations that using the mean sensitivity gives equivalent results to using the individual components when recovering a constant shear. This reflects the fact that the two sensitivity components are equal in the mean, within our uncertainties.

Blinding

During the development of the analysis presented here, the NGMIX shear catalogue was blinded by an unknown factor between 0.9 and 1.0 (Jarvis et al. 2016, their section 7.5). Only after the data and modelling pipeline presented here had successfully passed all internal tests were the unblinded shears processed.

However, after unblinding we uncovered an inconsistency in the interpretation of radial bin limits between data and model. We note that this discovery was made by directly assessing whether the model properly approximates the data, and not by comparing our mass–richness relation to literature results or expectations. We repeated the blinded analysis with the corrected pipeline, checked the results for consistency again and only then continued with the unblinded analysis.

2.3 Photometric redshift catalogue

Inference of physical quantities such as mass from a lensing signal requires knowledge of the redshift distribution of the source galaxies being lensed. DES has explored a broad range of photometric redshift estimators (Sánchez et al. 2014). Four of the best-performing codes were selected for a detailed characterization of the impact of photometric redshift uncertainties on weak-lensing studies such

as this one (Bonnett et al. 2016). These were ANNZ2 (Collister & Lahav 2004; Sadeh, Abdalla & Lahav 2016), SKYNET (Graff et al. 2014), TPZ (Carrasco Kind & Brunner 2013) and BPZ (Benítez 2000; Coe et al. 2006). The first three of these are machine-learning codes, while BPZ is a template-based method. All four methods noted above were used to produce a full probability distribution $p(z)$ for every source galaxy in the DES shear catalogues, finding comparable performances.

We note that Bonnett et al. (2016) found it necessary to shift the probability distribution recovered from BPZ upwards by 0.05 in order to counteract intrinsic biases in BPZ that arise due to limitations in the template set employed. Since our analysis relies on the same photometric data and photometric redshift outputs as Bonnett et al. (2016), we follow these authors in applying a systematics shift of 0.05 to the photometric redshift distributions from BPZ for all following analyses, and in using the SKYNET photo-zs for the fiducial calculations in this work. We present a detailed characterization of photometric redshift uncertainties in Section 4.3.

3 STACKED LENSING MEASUREMENTS

We briefly summarize the methodology for stacked cluster lensing measurements, and refer the interested reader to Bartelmann & Schneider (2001) and Sheldon et al. (2004) for details.

Equation (1) defines an estimator for the 'reduced shear'

$$g \equiv \frac{\gamma}{1-\kappa}, \tag{3}$$

a non-linear combination of the gravitational shear γ and convergence κ. *Weak* lensing is characterized by $g \approx \gamma$, which is an accurate approximation for this study: we excise areas close to cluster centres to avoid high shear and difficulties of photometry and shape measurements in crowded regions (cf. Sections 4.2 and 5.7).

The weak shear by a foreground mass concentration induces correlations in the shapes of background galaxies, such that, on average, galaxies images are stretched tangentially with respect to the centre of mass. The mean tangential shear γ^T at a distance R from a cluster is related to the surface mass density Σ of the cluster via

$$\Delta\Sigma \equiv \overline{\Sigma}(<R) - \overline{\Sigma}(R) = \Sigma_{\text{crit}}\,\gamma^\mathsf{T}(R), \tag{4}$$

where $\Sigma(x, y) = \int_z \rho(x, y, z)$ is the line-of-sight projection of the physical mass density ρ, $\overline{\Sigma}(<R)$ is the mean surface density within R and $\overline{\Sigma}(R)$ is the azimuthally averaged surface density at radius R. We will take the source ellipticity, rotated to the tangential frame, as a noisy estimator

$$\gamma^\mathsf{T} \approx e^\mathsf{T} + \text{noise}, \tag{5}$$

where the noise is due to both intrinsic ellipticities of the source galaxies and measurement noise.

The strength of the lensing signal imprinted on background galaxies is modulated by the critical surface density

$$\Sigma_{\text{crit}}(z_s, z_l) = \frac{c^2}{4\pi G} \frac{D_s}{D_l D_{ls}}, \tag{6}$$

where D_s, D_l and D_{ls} refer to the angular diameter distances to the source, to the lens, and between lens and source, respectively. Equation (6) provides a mechanism to weight each lens–source pair to maximize the signal-to-noise ratio of $\Delta\Sigma$. For most lensing analyses (including this one), the distances of the sources must be estimated from photometric rather than spectroscopic redshifts, which demands the introduction of an effective critical density for
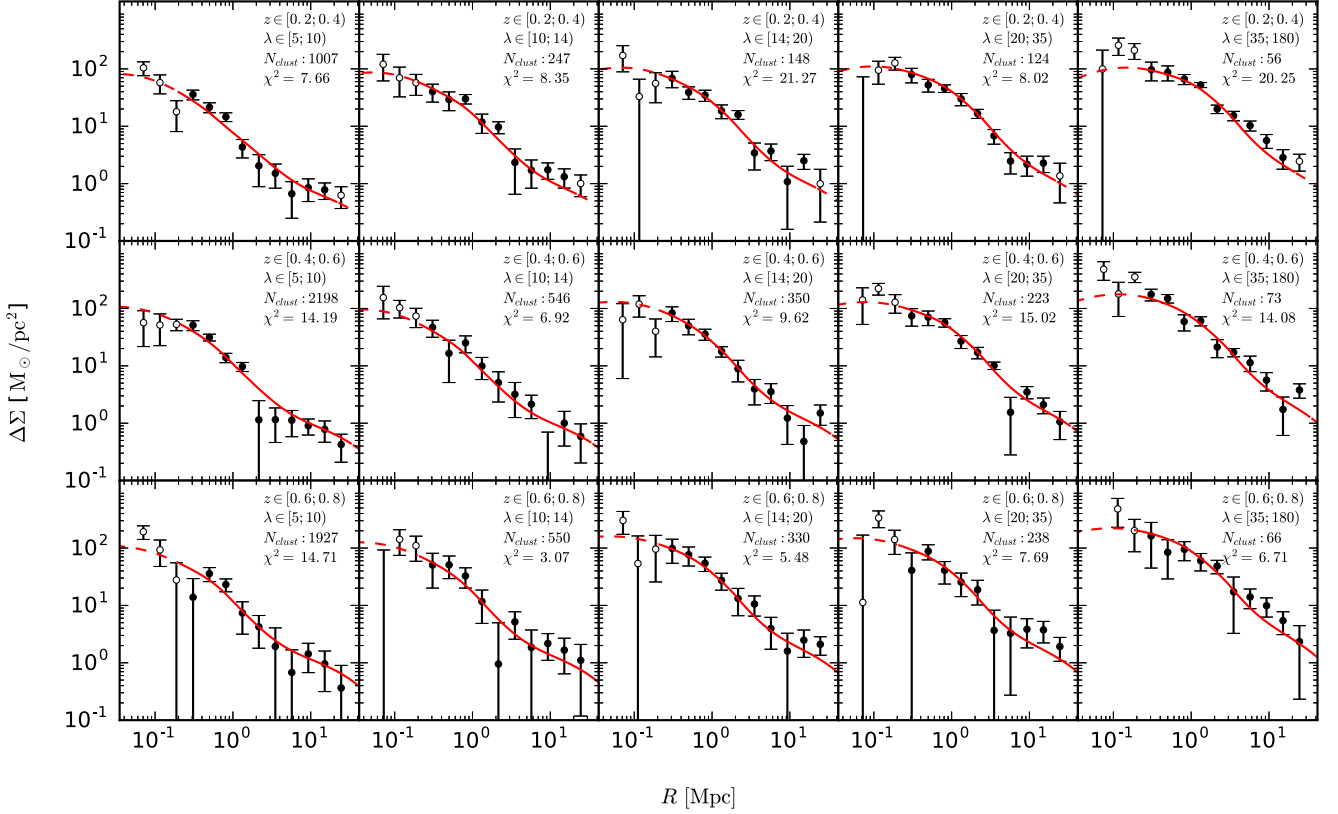
Figure 3. Mean $\widetilde{\Delta\Sigma}$ for cluster subsets split in redshift z_l (*increasing from top to bottom*) and λ (*increasing from left to right*) with errors from jackknife resampling (see Section 3.1). Our best-fitting model (*red curve*) includes dilution from cluster member galaxies (Section 4.2) and miscentring (Section 5.4); see Fig. 11 for details. Data points considered unreliable and therefore excluded from further analysis (below 200 kpc or above 1 deg) are indicated by open symbols and dashed lines. The profiles and jackknife errors are calculated after the subtraction of the random-point shear signal (see Section 4.1.2).

each lens–source pair

$$\langle \Sigma_{\text{crit}}^{-1} \rangle_{j,i} = \int \mathrm{d}z_{\text{s},i} \; p_{\text{phot}}(z_{\text{s},i}) \, \Sigma_{\text{crit}}^{-1}(z_{\text{s},i}, z_{l,j}) \tag{7}$$

that averages over the redshift probability distribution $p_{\text{phot}}(z_{\text{s},i})$ of source i, evaluated for lens j.[2] We can combine the previous equations in this section to arrive at an estimator $\widetilde{\Delta\Sigma}_{j,i}$ for each lens–source pair (j, i). The optimal estimator for the stacked density profile $\Delta\Sigma$ is then a weighted sum over these individual estimators, so that

$$\widetilde{\Delta\Sigma} = \frac{\sum_{j}^{\text{lens}} \sum_{i}^{\text{src}} w_{j,i} \, e_{j,i}^{\mathsf{T}} \left/ \langle \Sigma_{\text{crit}}^{-1} \rangle_{j,i} \right.}{\sum_{j,i} w_{j,i}}, \tag{8}$$

where the weights

$$w_{j,i} = \langle \Sigma_{\text{crit}}^{-1} \rangle_{j,i}^2 \left/ \sigma_{\gamma,i}^2 \right. \tag{9}$$

are chosen to minimize the variance of the resulting estimator (Sheldon et al. 2004).[3] The quantity σ_{γ}^2 is the uncertainty on the shear measurement, and combines the uncertainty in the shear due

to intrinsic galaxy shapes σ_{SN} with the uncertainty in the ellipticity measurement,

$$\sigma_{\gamma,i}^2 = \sigma_{\text{SN}}^2 + \frac{1}{2} \left(\mathbf{C}_{11,i} + \mathbf{C}_{22,i} \right), \tag{10}$$

where $\mathbf{C}_{11} + \mathbf{C}_{22}$ is the trace of the ellipticity subset of the covariance matrix produced by the NGMIX code. We take $\sigma_{\text{SN}} = 0.22$. Although the use of this simple σ_{γ}^2 is not optimal (Bernstein & Jarvis 2002), it only slightly increases the variance of $\Delta\Sigma$.

Our particular choice of ellipticity measurement requires division by a mean sensitivity to produce unbiased results (see Section 2.2). We therefore modify our $\Delta\Sigma$ estimator to use the mean weighted sensitivity,

$$\widetilde{\Delta\Sigma} = \frac{\sum_{j}^{\text{lens}} \sum_{i}^{\text{src}} w_{j,i} \, e_{j,i}^{\mathsf{T}} \left/ \langle \Sigma_{\text{crit}}^{-1} \rangle_{j,i} \right.}{\sum_{j,i} w_{j,i} \, s_i}. \tag{11}$$

Efficient codes to compute the estimator in equation (11) are publicly available. For this work, we used XSHEAR[4] and cross-checked its results with an independent implementation.[5] The resulting shear profiles are shown in Fig. 3. Clusters were split into three subsets $z_l = [0.2, 0.4), [0.4, 0.6), [0.6, 0.8)$ and five subsets $\lambda = [5, 10), [10, 14), [14, 20), [20, 35), [35, 180)$ with 13 logarithmically spaced

---

[2] In this work, we also estimate the cluster redshifts photometrically. However, the error on those estimates is of the order of $\Delta z \approx 0.01$ (Rykoff et al. 2016), which is negligible compared to both the width of the lens redshift bins we adopt and of the source redshift distributions. In what follows, we will therefore treat the cluster redshifts as exact.

[3] Our weight depends on the measured ellipticity noise, which correlates with the galaxy ellipticity, and thus can induce a bias in the recovered

shear. However, in Jarvis et al. (2016) we found that weighting effects are subdominant to other sources of bias for the NGMIX catalogue.

[4] https://github.com/esheldon/xshear
[5] https://github.com/pmelchior/shear-stacking

radial bins between 0.05 and 30 Mpc. The redshift splitting allows for constraints on a possible redshift evolution of the MOR, while removing only a small number of clusters at $z_l > 0.8$ whose lensing weights are low because their redshift is larger than that of most galaxies in the shear catalogue. The richness splitting was adjusted to provide roughly equal signal-to-noise ratio for each cluster subset. The number of radial bins was set so that all of them are populated, and the outer cut-off corresponds to the spatial size of the jackknife regions in our covariance estimation scheme.

3.1 Data covariance matrices

The $\Delta\Sigma(R)$ profiles that we measure from our data deviate from the true mean $\Delta\Sigma(R)$ profiles of clusters in a given richness–redshift subset by a statistical uncertainty that we characterize by a covariance matrix $\mathbf{C}_{\widetilde{\Delta\Sigma}}$. Contributions to $\mathbf{C}_{\widetilde{\Delta\Sigma}}$ include shape noise, uncorrelated large-scale structure along the line of sight (e.g. Hoekstra 2001, 2003; Hoekstra et al. 2011; Umetsu et al. 2011) and intrinsic variations of cluster profiles at fixed mass (e.g. Metzler, White & Loken 2001; Becker & Kravtsov 2011; Gruen et al. 2011, 2015).

For a stacked cluster lensing analysis in a common footprint, a single source galaxy may be within the maximum search radius of multiple galaxy clusters. Thus, the cluster $\Delta\Sigma$ measurements are not fully independent, and the covariance matrix will have significant off-diagonal terms, particularly on large scales. This is exacerbated by the spatial co-location of clusters and background galaxies (cf. Fig. 1) due to a combination of variations in survey depth, which affect the detectability of galaxies and redMaPPer clusters similarly, and cluster and galaxy clustering.

For estimating $\mathbf{C}_{\widetilde{\Delta\Sigma}}$, we therefore use a spatial jackknife scheme designed to account for the covariance of the measurements, which we expect to depend on scale. To this end, we split the source sample into $K = 40$ simply connected regions $\mathcal{R}_k$ by running a k-means algorithm on the sphere.[6] For each such $\mathcal{R}_k$ and each richness–redshift subset, we calculate equation (11) for all lenses $j \notin \mathcal{R}_k$ and denote it $\widetilde{\Delta\Sigma}_{(k)}$. We then calculate the covariance matrix according to Efron (1982),

$$\mathbf{C}_{\widetilde{\Delta\Sigma}} = \frac{K-1}{K} \sum_k^K (\widetilde{\Delta\Sigma}_{(k)} - \widetilde{\Delta\Sigma}_{(\cdot)})^T \cdot (\widetilde{\Delta\Sigma}_{(k)} - \widetilde{\Delta\Sigma}_{(\cdot)}), \qquad (12)$$

where $\widetilde{\Delta\Sigma}_{(\cdot)} = \frac{1}{K}\sum_k \widetilde{\Delta\Sigma}_{(k)}$. We note that in the above expression, $\widetilde{\Delta\Sigma}$ concatenates all redshift and richness subsets into a single data vector, enabling us to detect covariance across all redshift, richness and radial bins.

The upper left panel of Fig. 4 shows the resulting correlation matrix for one reference cluster subset. On smaller scales the diagonal is dominant, but – as expected – off-diagonal terms are present for the largest scales. We also test for cross-correlations between the profiles measured for clusters from different richness and redshift subsets, and find them to be small (cf. upper right and lower left panel of Fig. 4). We therefore will make the assumption of no cross-correlation between different cluster subsets in our likelihood analysis presented in Section 5.6. Future analyses with larger DES data sets will result in significantly reduced uncertainties of the covariance matrices, allowing us to properly include any correlation that may have remained undetected here because of the modest size of the SV data set.
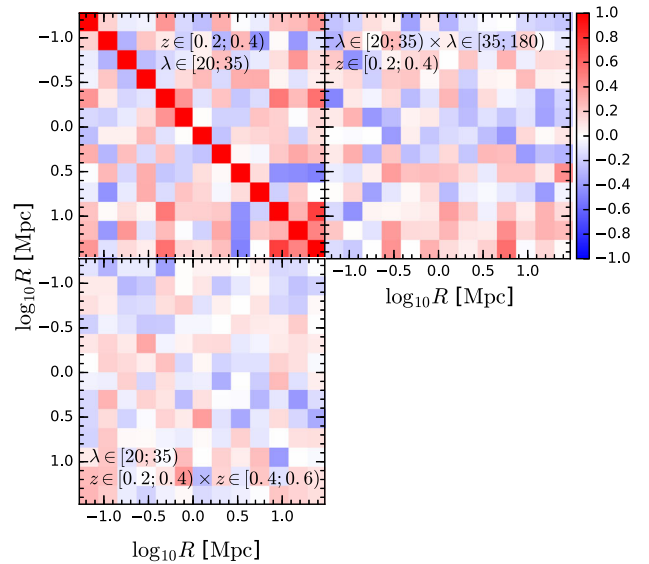
Figure 4. Correlation matrix of $\widetilde{\Delta\Sigma}$ of a single richness–redshift subset with $z \in [0.2, 0.4)$ and $\lambda \in [20, 35)$ for 40 jackknife regions *(upper left panel)*. Off-diagonal blocks show the correlation matrix between the lensing profile of the reference subset and the lensing profiles of the neighbouring redshift subset ($z \in [0.4, 0.6)$, *lower left*), and the neighbouring richness subset ($\lambda \in [35, 180)$, *upper right*).

In addition, we also performed covariance estimation with smaller jackknife regions ($K = 100$) and cluster-by-cluster jackknifing, yielding similar results on small scales. We note that jackknife schemes are prone to underestimation of the covariance on scales that exceed the size of the jackknife patches (e.g. Friedrich et al. 2016, for the case of shear autocorrelations). In the case of $K = 40$ and DES SV, this corresponds to angular scales of approximately $1°$. For this reason, we exclude $\Delta\Sigma$ measurements at $R > 1°$ from further analysis (cf. Fig. 3).

The covariance matrix estimated from the data is noisy, and its inverse is therefore a biased estimate of the true inverse covariance (e.g. Kaufman 1967; Hartlap, Simon & Schneider 2007). In addition, the use of noisy covariances leads to additional uncertainty in estimated parameters (Dodelson & Schneider 2013). The size of these effects depends on the details of the jackknife scheme, data vectors and structure of the true covariance.

To calibrate the effect, we generate random realizations of 10 uncorrelated zero-mean Gaussian random variables in 40 jackknife patches. This is close to the jackknife scheme and true covariance of the data vectors used in Section 5.7 for the estimation of masses. For each realization, we estimate the covariance matrix from equation (12), and invert it to find the best-fitting mean and its uncertainty. We compare the estimated uncertainty to the actual scatter of the best fit over a large number of realizations.

We find that the actual uncertainty of the best fit is ≈ 30 per cent larger than the one estimated with the inverse jackknife covariance. Approximately half of this excess uncertainty is corrected when applying the de-biasing factor of equation 17 in Hartlap et al. (2007) to the inverse covariance, the remaining half is consistent with the expectations from Dodelson & Schneider (2013). We correct both effects by rescaling the statistical uncertainty of mass estimates in Section 5.7 by a factor of 1.3. Future work would benefit from the use of less noisy (e.g. simulation-based) or analytical (Gruen et al. 2015) covariance matrices.

4 SYSTEMATICS

4.1 Shear systematics

The NGMIX catalogue passed an extensive set of null tests on real data and simulations (Jarvis et al. 2016), which we will briefly summarize here. We adopt the bias parametrization

$$\tilde{g} = (1 + m)g + \alpha e_{PSF} + c. \qquad (13)$$

Note that m, α and c are in principle two-component variables, but in our experience the values in both components are identical within errors. We will therefore treat them as scalars. Point spread function (PSF) size modelling errors resulted in a multiplicative bias $|m_{PSF}| < 0.01$, while inaccurate deconvolution led to PSF leakage $\alpha < 0.01$. Both of these effects are negligible compared to the remaining systematics described in this section, and are therefore ignored in our systematic error budget.

Lacking an absolute shear calibration source, we cannot test the overall shear calibration using real data. We instead adopted a set of simulations (Jarvis et al. 2016) based on the real galaxy images from the COSMOS imaging data (Lilly et al. 2007). The selection of these COSMOS galaxies does not perfectly match our selection in DES data; the COSMOS field is quite small and thus subject to cosmic variance. Thus, we cannot infer any detailed information about the shear bias in DES data; the bias in real data may be more or less than the bias we see in the simulations. Without additional information, however, we choose to model our systematics based on what we found in these simulations.

It is important to note that the IM3SHAPE catalogue was recalibrated directly from these simulations to minimize biases caused by pixel noise, whereas NGMIX did not require that step. For NGMIX, the multiplicative shear calibration error m was seen to be consistent with zero for galaxies at redshift $z \approx 1$, but as large as -0.04 for sources at $z \approx 0.3$. We believe that shear inference for the lower redshift sources in the COSMOS galaxy sample is primarily affected by 'model bias', introduced by fitting an exponential disc model to a galaxy population that has a large number of bulge-like galaxies. We note that the bulk of the cluster sample is at redshift $z_\lambda > 0.5$ (cf. Fig. 2), so that these sources, already a minority in the shape catalogues, receive significant weight only for the lowest redshift clusters. The simulation results therefore suggest that the multiplicative shear bias is controlled to $|m| \leq 0.03$.

However, a detailed comparison of the NGMIX and IM3SHAPE shape catalogues performed by Jarvis et al. (2016) found the residual systematic uncertainties to be larger: $|m| \leq 0.05$. This finding may reflect the differences between the simulations from which IM3SHAPE was calibrated, and the DES SV data, for instance in the redshift-dependent bulge fraction (Jarvis et al. 2016). Rather than adopting a top-hat prior with $|m| \leq 0.05$, we use a Gaussian prior with the same variance. Doing so preserves the total error budget, while avoiding an inappropriate sharp cut on $|m|$. The corresponding Gaussian prior (to be used in the likelihood analysis of Section 5.6) is $m = 0.00 \pm 0.03$, where the error is to be interpreted as standard deviation.

Additive errors were found to be below the cosmic-shear requirement of $|c| < 0.002$ for the SV survey area (Jarvis et al. 2016). Below we perform additional null tests, particularly useful for stacked cluster lensing analyses.

4.1.1 B modes and quadrant checks

We also perform an additional test for residuals by projecting the galaxy ellipticities on to the direction $45°$ off the tangent. This so-
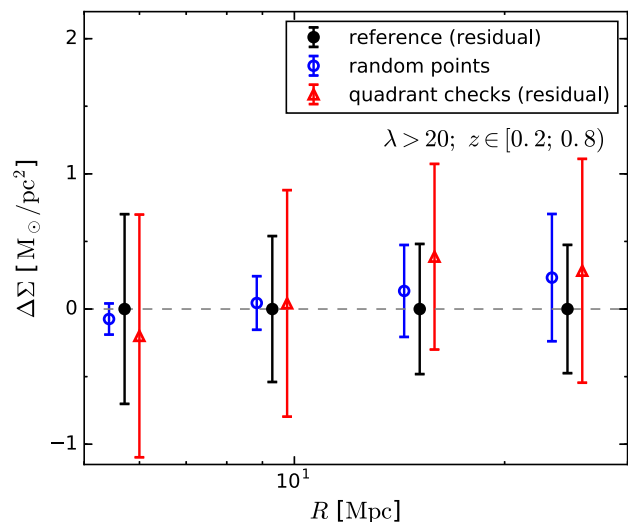
Figure 5. Corrections for additive shear systematics: the shear signal around random points, drawn such as to mimic the distribution of redMaPPer clusters in the DES SV footprint (*blue open circles*, cf. Section 4.1.2); the difference in the shear signal caused by rejecting the clusters that are excluded after the quadrant check (*red triangles*, cf. Section 4.1.1); for comparison, the errors of the $\Delta\Sigma$ profile after random-point subtraction (*black circles*).

called B mode should be zero in the mean if the signal we measure is solely from gravitational lensing. Using the jackknife covariance, we test for significant deviation of the χ^2 of B-mode shears from zero. The stacked B-mode signal of clusters with $\lambda > 20$ in all redshift bins combined yields a reduced $\chi^2/N_{dof} \approx 16/10$ for 10 radial bins, consistent with no systematic at the $p > 0.1$ level.

Additive errors of the c or α type are cancelled in $\gamma^\intercal$ when sources separated by $90°$ are averaged, which we implicitly exploit whenever sources are evenly distributed around each lens. But survey boundaries, as well as holes due to masking of bright stars and other features, result in a non-uniform distribution of source galaxy positions. The DES SV footprint is rather small, so our large maximum radius $R_{max} = 30$ Mpc around each cluster lens often intersects at least one such boundary. We identify clusters for which that is the case with the so-called quadrant check (Sheldon et al. 2004, their section 3.2.4) and show them with open circles in Fig. 1.

We show the difference between the $\widetilde{\Delta\Sigma}$ profiles before and after the quadrant-based rejection in Fig. 5 (red markers). We can see that the difference is largest at scales outside of 10 Mpc, but not statistically significant. Requiring that clusters pass the quadrant check substantially increases statistical uncertainties on large scales, where many source–lens pairs get rejected. As it does not appear to be beneficial in this work, we will therefore not demand that clusters need to pass the check and will utilize the entire cluster sample for further analysis.

4.1.2 Random-point test

Despite being not significant in the test of Section 4.1.1, additive shear systematics may still be present on all scales, which can lead to small spurious shear signals in our analysis. A simple correction can be made by measuring the tangential shear around a set of random points, which reproduce the redshift and richness distribution of clusters in the DES SV footprint, and subtracting it from the actual cluster signal.

To this end, we use the weighted random-point sample as described in Rykoff et al. (2016, their section 3.6), generated from the survey mask and redMaPPer maximum redshift maps. Since the effective survey geometry varies with cluster redshift and richness, we split these random points in the same way as our cluster sample and measure $\Delta\Sigma$ profiles around each subset.

Fig. 5 shows results from the random-point shear measurement as blue markers. We find no significant random shears even on the largest scales. To benefit from a correction of potential shear systematics below the detection limit, we will use random-point subtracted shear profiles for the rest of our analysis.

4.2 Correction for cluster members in the shear catalogue

Due to photometric uncertainties and ambiguity of the available colour information, the estimated $p(z)$ of the source galaxies is quite broad, and a source at lower redshift than the cluster will get some non-zero weight in the analysis. This is properly accounted for when calculating $\langle \Sigma_{crit}^{-1} \rangle_{j,i}$ of a source–lens pair in equation (7). However, cluster members must be treated specially, as their redshift distribution is essentially a delta function centred on the cluster redshift. This delta function is not properly accounted for in the $p(z)$, and thus not included in the integral to calculate Σ_{crit}^{-1}. Each member gets a non-zero weight, but adds zero to the mean shear.[7] Furthermore, cluster members are highly concentrated near the cluster centre, so there is a strong radially dependent bias in the inferred $\Delta\Sigma$.

As a first step for alleviating this effect, we exclude all galaxies that are likely cluster members from the shape catalogue. We run the redMaPPer algorithm for member identification (Rozo et al. 2015) and reject all galaxies consistent with being a cluster member down to the magnitude limit of the survey and out to $1.5R_\lambda$.

While this significantly reduces the contamination from cluster members, a fraction of them remains in the source catalogue. We estimate their contribution to the signal, the so-called *boost factors* (e.g. Sheldon et al. 2004, their section 4.1), for each cluster subset as a function of R, and correct the negative impact in Section 5.5. We use a method similar to the one described in Gruen et al. (2014, their section 3.1.3),[8] which method is based on the decomposition of the redshift distribution of source galaxies into a field galaxy component and a cluster member component.

Consider all lens–source pairs (j, i) in a cluster richness–redshift subset and at some projected separation R. Each source–lens pair yields an estimate of $\Delta\Sigma_{j,i} = e_{j,i}^\mathsf{T} / \langle \Sigma_{crit}^{-1} \rangle_{j,i}$ that enters the mean measured $\widetilde{\Delta\Sigma}$ with relative weight $w_{j,i}$ (cf. equations 9 and 11). Assuming that we can split the source–lens pairs into pairs with field galaxies b and pairs with cluster member galaxies m, we expand

$$\widetilde{\Delta\Sigma} = \frac{\sum_{j,i\in b} w_{j,i}\Delta\Sigma_{j,i} + \sum_{j,i\in m} w_{j,i}\Delta\Sigma_{j,i}}{\sum_{j,i\in b} w_{j,i} + \sum_{j,i\in m} w_{j,i}} \ . \tag{14}$$

Taking the expectation value and using that $\langle \Delta\Sigma_{j,i} \rangle = \Delta\Sigma$ for field galaxies and $\langle \Delta\Sigma_{j,i} \rangle = 0$ for cluster members,[7] we find

$$\langle \widetilde{\Delta\Sigma} \rangle = \frac{\sum_{j,i\in b} w_{j,i}}{\sum_{j,i\in b} w_{j,i} + \sum_{j,i\in m} w_{j,i}} \Delta\Sigma = (1 - f_{cl})\Delta\Sigma \ , \tag{15}$$

where we defined the fractional weight of cluster member galaxies

$$f_{cl}(R) = \frac{\sum_{j,i\in m} w_{j,i}}{\sum_{j,i} w_{j,i}} = 1 - \frac{\sum_{j,i\in b} w_{j,i}}{\sum_{j,i} w_{j,i}} \tag{16}$$

and the radial dependence stems from the selection of pairs (j, i) for both m and b. In principle, this allows us to correct the data for the effect of member dilution via

$$\widetilde{\Delta\Sigma}_{corr}(R) = \frac{\widetilde{\Delta\Sigma}(R)}{1 - f_{cl}(R)} \ . \tag{17}$$

In practice, rather than correct the data for this effect, we choose to dilute the predicted lensing signal to match the observational data when modelling the recovered $\Delta\Sigma$ profiles.

The remaining task is to measure f_{cl} from the weighted, estimated redshift distribution of sources,

$$p(z) = \frac{\sum_{j,i} w_{j,i} p_i(z)}{\sum_{j,i} w_{j,i}} \ . \tag{18}$$

The left-hand panel of Fig. 6 shows the $p(z)$ of sources in a set of annuli around the cluster centres. We observe a systematic increase of low-redshift sources as one moves from large cluster-centric distances to small radii, clearly illustrating the effects of cluster membership contamination of the source galaxy catalogue.

We decompose the observed photometric redshift distribution $p(z)$ as a weighted sum of a field galaxy and a member galaxy component,
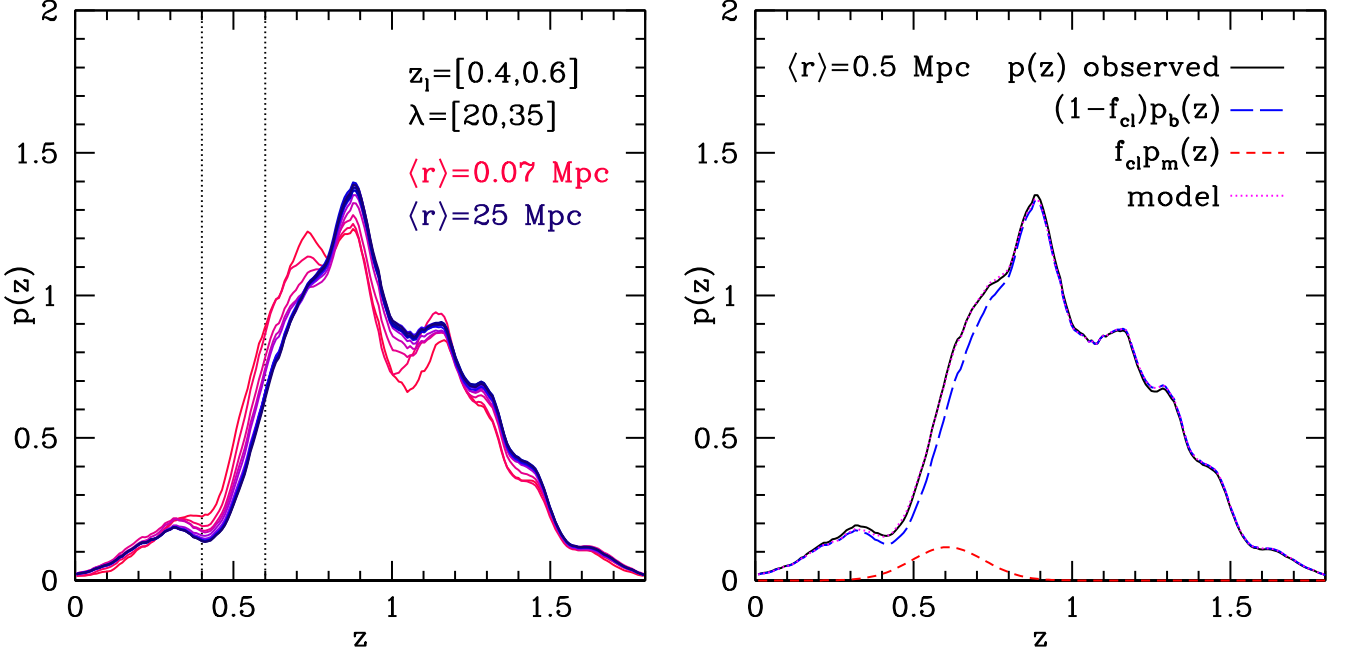
$$p(z) = \frac{\sum_{j,i\in b} w_{j,i} p_i(z) + \sum_{j,i\in m} w_{j,i} p_i(z)}{\sum_{j,i} w_{j,i}}$$
$$= (1 - f_{cl}) \, p_b(z) + f_{cl} \, p_m(z) \ , \tag{19}$$

using the stacked, weighted redshift distributions $p_b(z)$ and $p_m(z)$ of the respective samples.

We measure $p_b(z)$ from the redshift distribution of source galaxies around random points. To ensure sampling from the same distribution of survey depth as around the actual lenses, we bin and weight the random-source pairs in the same way as the lens–source pairs. The weighted cluster member redshift distribution is not well constrained by the data in each individual bin. We can, however, find robust constraints on its mean and variance with a joint fit of a Gaussian with a common mean and width to all radial bins of a cluster (sub)sample, and a free amplitude in each radial bin that corresponds to f_{cl}.

The right-hand panel of Fig. 6 shows the model for one example radial bin at $\langle R \rangle = 0.5$ Mpc. We can see that the model recovers an excess contribution (red dashed curve) from objects that are not present in the field galaxy population. We note that it is necessary to leave the mean value of the Gaussian as a free parameter: due to the skewness of the lensing weight applied to the stacked $p(z)$, which are zero at $z_s \le z_l$ and positive at higher source redshift, and the redshift prior for galaxies found in our survey, the mean of the recovered Gaussian is not expected to coincide with the cluster redshift. We have ensured that this method of decomposition yields consistent results with (a) a decomposition with free mean and width of the Gaussian in each radial bin (instead of fixed over all radial bins) and (b) a non-parametric measurement of $(1 - f_{cl})$ by the ratio of integrals over $p(z)$ and $p_b(z)$ over $|z - z_{cl}| > 0.3$.

---

[7] This is only true in absence of intrinsic alignments. Observational results on mild radial alignment of cluster members towards the halo centre vary (Hao et al. 2011; Sifón et al. 2015). If such alignment is present, it would lead to $\langle \Delta\Sigma_{j,i} \rangle_m < 0$. We consider it a higher order term and will thus neglect it in this work.

[8] We found the standard method based on correlation functions (Sheldon et al. 2004) unreliable due to the small footprint of the DES SV data set.

Figure 6. Estimation of cluster member contamination of the source sample by $p(z)$ decomposition. Left-hand panel: stacked, lensing-weighted $p(z)$ of sources around clusters in one redshift–richness bin from the innermost (red) to outermost (blue) distance bin. Right-hand panel: decomposition of observed $p(z)$ in one radial bin (black, solid line) into the sum (magenta, dotted line) of field distribution $p_b(z)$ as measured around random points (blue, long-dashed line) and Gaussian $p_m(z)$ of cluster members (red, short-dashed line).

There is an implicit assumption inherent to this method, namely that the observed $p(z)$ deviate from $p_b(z)$ only because of cluster-member contamination. This assumption may be violated for at least two reasons. First, lensing magnification changes the redshift distribution of field galaxies in a complex way that depends on the details of galaxy types, redshifts and luminosity functions characterizing the lensing source sample, and a part of this could in principle mimic cluster-member contamination. Gruen & Brimioulle (2017) investigated this effect and find the bias on $\Sigma_{\rm crit}^{-1}$ for a DES-like source population to be below the percent level.

Secondly, photo-z estimates may be affected by measurement biases caused by the presence of clusters, in particular an increased probability of blended sources in the cluster core. To investigate this possibility, we compared the results of the $p(z)$ decomposition with an extension of the method employed by Melchior et al. (2015, their section 3.4.1): we make use of the BALROG catalogue of fake objects in the DES SV footprint (Suchyta et al. 2016) to compare the lensing weights of actual DES galaxies around clusters with those of fake galaxies. When BALROG galaxies are matched to have the same properties as the galaxies in our lensed source sample, we can infer the effects of increased blended or light contamination in the dense cluster environments. Due to the lack of a photo-z catalogue for BALROG objects, we could only perform the matching to DES galaxies in the shear catalogue via proxies, for which we used the SExtractor parameters FLUX_RADIUS and MAG_AUTO in the i band. Despite this limitation, we found good agreement of the boost factors obtained from this method with the $p(z)$ decomposition presented above for the two high-redshift subsamples and a 1σ discrepancy for the lowest cluster redshift subsample, when restricted to $R \geq 200$ kpc. We will revisit this issue in forthcoming analyses with larger cluster samples, and for this work adopt the boost factors from the $p(z)$ decomposition but limit the acceptable range of the shear profiles to $R \geq 200$ kpc.

4.3 Photometric redshift systematics

Bonnett et al. (2016) assessed the impact of photometric redshift systematics on weak-lensing analyses of DES SV data, e.g. by characterizing how $\langle \Sigma_{\rm crit}^{-1} \rangle$ varies as a function of lens redshift between the various photometric redshift algorithms.

Here, we update the results of Bonnett et al. (2016) to account for the additional source weights that enter into our estimator. Specifically, consider the estimator for equation (8),

$$\widetilde{\Delta\Sigma} = \frac{\sum_{j,i} w_{j,i}\, e_{j,i}^{\mathsf{T}}\, S_{j,i}^{-1}}{\sum_{j,i} w_{j,i}}, \tag{20}$$

where we have defined $S_{j,i} \equiv \Sigma_{{\rm crit},ji}^{-1}$ as the true expectation value of $\Sigma_{\rm crit}^{-1}$ for the lens–source pair (j, i). The weights $w_{j,i}$ take the form $w_{j,i} = S_{j,i}^2 \sigma_{j,i}^{-2}$. The above estimator is appropriate when one has spectroscopic redshifts for all lens–source pairs.

We wish to determine how the estimator changes when we use a biased photometric redshift code for which $S'_{j,i} = \langle \Sigma_{\rm crit}^{-1} \rangle_{j,i} \neq S_{j,i}$. Defining $\epsilon_{j,i}$ via $S'_{j,i} \equiv S_{j,i}(1 + \epsilon_{j,i})$, we have

$$\widetilde{\Delta\Sigma'} = \frac{\sum_{j,i} \sigma_{j,i}^{-2}\, S_{j,i}\, (1 + \epsilon_{j,i})\, e_{j,i}^{\mathsf{T}}}{\sum_{j,i} \sigma_{j,i}^{-2}\, S_{j,i}^2\, (1 + \epsilon_{j,i})^2}. \tag{21}$$
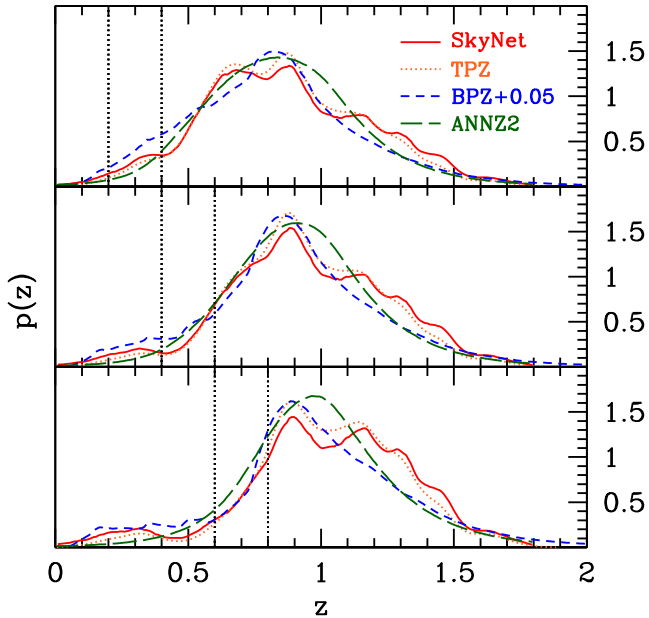
We take the expectation value of the above equation, using $\langle e_{j,i}^{\mathsf{T}} \rangle = S_{j,i}\Delta\Sigma$. Expanding to the first order in ϵ, we arrive at

$$\langle \widetilde{\Delta\Sigma'} \rangle = \Delta\Sigma\,[1 - \delta], \tag{22}$$

where we have defined

$$\delta \equiv \frac{\sum_{j,i} w_{j,i}\epsilon_{j,i}}{\sum_{j,i} w_{j,i}}. \tag{23}$$

The quantity δ has an alternative interpretation, as the difference of the true mean inverse critical surface density $\langle \Sigma_{\rm crit}^{-1} \rangle$ from its

Figure 7. Lensing-weighted, stacked $p(z)$ estimates for sources around clusters at $z \in [0.2, 0.4]$ (*top panel*, lens redshift range indicated by vertical dotted lines), $z \in [0.4, 0.6]$ (*central panel*) and $z \in [0.6, 0.8]$ (*bottom panel*) from four different photo-z codes. The $p(z)$ were estimated from sampling sources around redMaPPer random points in the $\lambda > 35$ and respective redshift subsets at a distance of ≈ 1 Mpc from the cluster, assigning the weight of equation (9). The distribution for BPZ was shifted towards larger z by 0.05.

estimate $\langle \Sigma_{\mathrm{crit}}^{-1} \rangle'$ based on a photometric redshift code,

$$\langle \Sigma_{\mathrm{crit}}^{-1} \rangle' = \frac{\sum_{j,i} w_{j,i} \, S'_{j,i}}{\sum_{j,i} w_{j,i}} = \langle \Sigma_{\mathrm{crit}}^{-1} \rangle \, [1 + \delta]. \tag{24}$$

In the absence of spectroscopic information, we adopt one of our photometric redshift codes, SKYNET, as the fiducial algorithm, and then estimate the relative offset δ at fixed lensing weights $w_{j,i}$ from equation (24). For precisely known lens redshifts, one can write

$$\langle \Sigma_{\mathrm{crit}}^{-1} \rangle' = \frac{1}{\sum_{j,i} w_{j,i}} \sum_{j,i} w_{j,i} \int \mathrm{d}z \, P(z) \, \Sigma_{\mathrm{crit}}^{-1}(z, z_{\mathrm{lens}})$$

$$= \int \mathrm{d}z \, P_{\mathrm{eff}}(z) \, \Sigma_{\mathrm{crit}}^{-1}(z, z_{\mathrm{lens}}), \tag{25}$$

where we have defined the effective source redshift distribution

$$P_{\mathrm{eff}}(z) = \frac{\sum_{j,i} w_{j,i} \, P_i(z)}{\sum_{j,i} w_{j,i}}. \tag{26}$$

The previous two equations provide a numerically convenient way to evaluate δ. Indeed, this is what was done in Bonnett et al. (2016), albeit with unit weights for all sources and a spectroscopic reference sample.

We compute the effective source redshift distributions, averaging over the lens redshifts of each of our three lens subsets, using the redMaPPer random points from the highest richness subset, in the radial bin around $R = 1 \, h^{-1}$ Mpc. The results are shown in Fig. 7. Selecting a different richness subset or radial aperture has no significant impact on these results. Table 1 lists the mean values of $\langle \Sigma_{\mathrm{crit}}^{-1} \rangle'$ for each of our photo-z codes and lens redshift bins, and the corresponding δ values. As already observed by Bonnett et al. (2016), the three machine learning codes are in much better agreement with each other than they are with BPZ, the only template-based code we consider, even after applying the global

Table 1. Systematic difference δ of lensing weighted $\langle \Sigma_{\mathrm{crit}}^{-1} \rangle$ between different photo-z codes according to equations (24) and (25). In the absence of spectroscopic redshifts for the source galaxies, we quote δ as the difference with respect to our reference pipeline SKYNET. BPZ+0.05 refers to results obtained after shifting all reported redshifts upwards by 0.05 (cf. Section 2.3). The statistical errors on $\langle \Sigma_{\mathrm{crit}}^{-1} \rangle$ are of the order of 0.01 and thus negligible compared to the systematic spread between methods.

Lens sample	Code	$\langle \Sigma_{\mathrm{crit}}^{-1} \rangle'$ $(10^{-4} \mathrm{M}_\odot^{-1} \mathrm{pc}^2)$	δ (per cent)
$z \in [0.2, 0.4]$	SKYNET	2.94	–
	ANNZ2	2.96	0.7
	TPZ	2.97	1.0
	BPZ+0.05	2.80	-4.8
$z \in [0.4, 0.6]$	SKYNET	2.74	–
	ANNZ2	2.68	2.2
	TPZ	2.75	0.4
	BPZ+0.05	2.57	-6.2
$z \in [0.6, 0.8]$	SKYNET	2.16	–
	ANNZ2	1.98	-8.3
	TPZ	2.16	0.0
	BPZ+0.05	1.97	-8.8

upwards shift of 0.05 to account for limitations in the template set (cf. Section 2.3).

For the likelihood analysis in Section 5.6, we seek to capture the spread and biases between the photo-z codes in the form of a prior. One way to set such a prior would be to consider the range of δ values obtained from the different algorithms, and then setting a flat top-hat prior over the range $\delta \in [\delta_{\mathrm{min}}, \delta_{\mathrm{max}}]$. However, we wish to allow for the possibility that δ is slightly beyond these limits by using a prior that drops off smoothly. To do so, we instead use a Gaussian prior that has the same mean and variance as the putative top-hat prior. Our final Gaussian prior is therefore

$$\delta = \frac{1}{2} (\delta_{\mathrm{max}} + \delta_{\mathrm{min}}) \pm \frac{0.577}{2} (\delta_{\mathrm{max}} - \delta_{\mathrm{min}})$$

$$= \begin{cases} -0.019 \pm 0.017 & \text{for } z \in [0.2, 0.4] \\ -0.020 \pm 0.024 & \text{for } z \in [0.4, 0.6] \\ -0.044 \pm 0.025 & \text{for } z \in [0.6, 0.8]. \end{cases} \tag{27}$$

Finally, we acknowledge that lensing magnification induces changes of the background source population that can cause additional biases specific to cluster lensing in the photo-z estimation. For a DES-like sample, resulting biases of $\Sigma_{\mathrm{crit}}^{-1}$ are of the order of 0.5 per cent and of opposite direction to the magnification-induced biases for the boost factors (Gruen & Brimioulle 2017), so we ignore both for the remainder of this work.

5 THE STACKED LENSING SIGNAL

We seek to calibrate the mass–richness relation of redMaPPer clusters. We will therefore estimate the mean cluster mass in each of our richness and redshift subsets by modelling the stacked weak-lensing signal of each cluster subset as if it is originated from a hypothetical halo of mass M. Below, we detail our model for the lensing signal as a function of the halo mass, and use numerical simulations to calibrate how the recovered mass is related to the mean cluster mass. We then consider how our model needs to be extended to account for the systematics discussed in the previous sections, as well as the additional complication caused by our limited knowledge of the true halo centre.

5.1 Surface density model

Fig. 3 shows the resulting estimates of the excess projected surface mass density $\Delta\Sigma$ defined in equation (4). The quantities $\Sigma(R)$ and $\overline{\Sigma}(<R)$ are given by

$$\overline{\Sigma}(<R) = \frac{2}{R^2} \int_0^R \mathrm{d}R' \, R' \, \overline{\Sigma}(R') \tag{28}$$

and

$$\overline{\Sigma}(R) = \int_{-\infty}^{+\infty} \mathrm{d}\chi \, \rho\left(\sqrt{R^2 + \chi^2}\right). \tag{29}$$

If the shear signal is caused by haloes of mass M, the excess three-dimensional matter density is given by

$$\rho(r) = \rho_{\mathrm{m}} \, \xi_{\mathrm{hm}}(r \,|\, M), \tag{30}$$

where $\rho_{\mathrm{m}} = \Omega_{\mathrm{m}} \, \rho_{\mathrm{c}}(1+z)^3$ is the mean matter density in physical units at the redshift of the sample, ρ_{c} is the critical density and $\xi_{\mathrm{hm}}(r|M)$ is the halo–matter correlation function evaluated for a halo of mass M.

We use the Zu et al. (2014) update to the Hayashi & White (2008) model of the halo–matter correlation function. Specifically, we set

$$\xi_{\mathrm{hm}}(r \,|\, M) = \max\{\xi_{\mathrm{1h}}(r \,|\, M), \, \xi_{\mathrm{2h}}(r \,|\, M)\}, \tag{31}$$

where we have constructed the total ξ_{hm} from the so-called 1-halo and 2-halo terms. We model the 1-halo term as a Navarro, Frenk & White (1996, hereafter NFW) density profile $\rho_{\mathrm{NFW}}(r \,|\, M)$,

$$\xi_{\mathrm{1h}}(r \,|\, M) = \frac{\rho_{\mathrm{NFW}}(r \,|\, M)}{\rho_{\mathrm{m}}} - 1, \tag{32}$$

with a statistical concentration–mass relation, for which we employ the Diemer & Kravtsov (2015) model. For the 2-halo term, we use the non-linear matter correlation function ξ_{nl} scaled by the halo bias $b(M)$ of Tinker et al. (2010) as

$$\xi_{\mathrm{2h}}(r \,|\, M) = b(M) \, \xi_{\mathrm{nl}}(r) \, . \tag{33}$$

The non-linear matter correlation function is related to the non-linear power spectrum P_{nl} as

$$\xi_{\mathrm{nl}}(r) = \frac{1}{2\pi} \int \mathrm{d}k \, k^2 P_{\mathrm{nl}}(k) \, j_0(kr) \, , \tag{34}$$

where $j_0(kr)$ is the 0-th spherical Bessel function of the first kind.

5.2 Modelling systematics

Any differences between the true $\Delta\Sigma$ profiles of cluster haloes of mean mass M and our analytical model for $\Delta\Sigma(M)$ from equation (31) can bias the recovered weak-lensing masses. For instance, while we choose the customary NFW profile to model the 1-halo term, several alternatives describe the cluster mass profiles similarly well (cf. Umetsu et al. 2014, for a recent comparison). Even within the NFW halo family, we are forced to adopt a concentration–mass relation.[9] While the consequences of these choices on mass estimates are expected to be small (cf. Hoekstra et al. 2012, their section 4.3), we none the less need to quantify them.

To do so, we measure the weak-lensing masses of dark matter haloes in numerical simulations using the same formalism

---

[9] This is a consequence of our rejection of shape measurements from $R < 200$ kpc, which could otherwise allow for data-driven constraints on the concentration parameter.
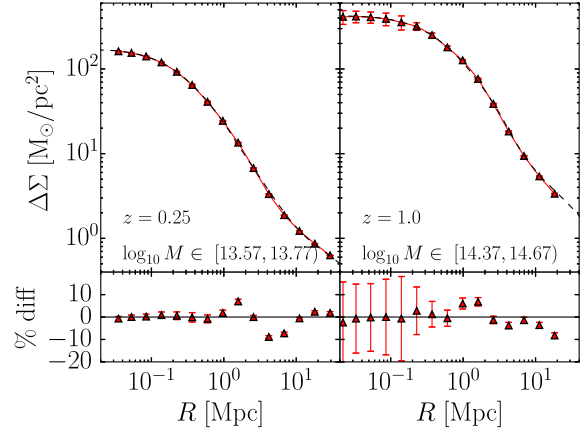

Figure 8. Simulated $\Delta\Sigma$ profile with error bars from jackknifing compared to the a priori analytical model profile from equation (31), evaluated at the average mass of all haloes in the associated subset. The two panels show the lowest redshift and lowest mass (*left*) as well as the highest redshift and highest mass subset (*right*), covering the mass and redshift range of galaxy clusters in the redMaPPer catalogue. Lower panels show the difference in percent between the analytic model and the simulated signal.

we employ with the DES data. The haloes are drawn from a *N*-body simulation of a flat ΛCDM cosmology run with the GADGET code (Springel 2005). The simulation uses 2.74 billion particles in a box that is 1050 Mpc h^{-1} on a side. The matter density is $\Omega_{\mathrm{m}} = 0.318$, implying that a $10^{13} \, h^{-1} \, \mathrm{M}_\odot$ halo is resolved with $\approx 10^3$ particles. The remaining cosmological parameters are $H_0 = 67.04$ km s^{-1} Mpc^{-1}, $\Omega_{\mathrm{b}} = 0.049$, $\tau = 0.08$, $n_{\mathrm{s}} = 0.962$ and $\sigma_8 = 0.835$. The force softening is 20 h^{-1} kpc. We discard all information below five softening lengths, and verified that the choice of extrapolation scheme for describing the correlation function below this scale does not impact our results. Haloes are identified using the ROCKSTAR halo finder (Behroozi, Wechsler & Wu 2013), using a spherical overdensity mass definition of 200 times the background density.

The numerical simulation is used to construct the synthetic weak-lensing signal of dark matter haloes drawn from four different redshift snapshots at $z = 0, 0.25, 0.5$ and 1. We split the haloes in narrow mass subsets, and compute the halo–mass correlation function with the Landy & Szalay (1993) estimator from the code TREECORR.[10] We numerically integrate the measured correlation functions to obtain the corresponding $\Delta\Sigma$ profiles as described in Section 5.1. The covariance matrix of our data points is estimated by splitting the simulation box into 64 jackknife regions.

A comparison of the simulated $\Delta\Sigma$ profiles, along with the a priori analytical model of the $\Delta\Sigma$ profile, is shown in Fig. 8. We see very small differences on small scales, which increases to ≈ 10 per cent on scales of the 1-halo to 2-halo transition. We tested five different mass bins at four different redshifts, and did not find a systematically low amplitude of the model on those scales. Moreover, we note that the Tinker et al. (2010) bias function is itself only accurate at the ± 6 per cent level, entirely consistent with the differences seen in Fig. 8. More detailed modelling (e.g. Diemer & Kravtsov 2014) can reduce these deviations at the expense of additional free parameters. Here, we opt to empirically calibrate any biases in the recovered weak-lensing masses due to these modest, highly localized model deviations.

---

[10] https://github.com/rmjarvis/TreeCorr
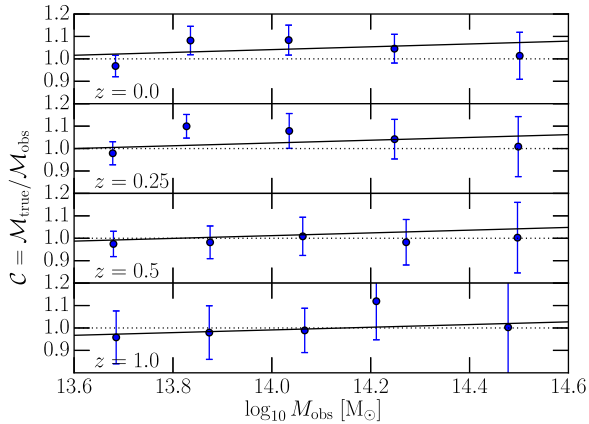
Figure 9. The mass calibration $\mathcal{C} = M_{\text{true}}/M_{\text{obs}}$ caused by adopting the analytic form of equation (31) for the $\Delta\Sigma$ profile, as a function of the observed halo mass M_{obs} for each simulated redshift bin. The solid line is the best-fitting bias model from equation (35).

When doing so, we restrict ourselves to the same radial scales employed in the weak-lensing analysis, and utilize the covariance matrices recovered from the data. Specifically, each halo stack is fit using the covariance matrix of the redMaPPer subset from Section 3.1 closest in mass and redshift to the simulated cluster stack. This ensures that the simulated data is weighted in the same way as the observed data, so that any biases are appropriately calibrated. Denoting M_{true} as the mean mass of the simulated cluster subset and M_{obs} as the result of the corresponding lensing analysis, we find small percent level biases caused by the adopted analytical form of the $\Delta\Sigma$ profile (Fig. 9). We characterize the mass bias $\mathcal{C} = M_{\text{true}}/M_{\text{obs}}$ as a function of the recovered weak-lensing mass M_{obs} and z with a power law:

$$\mathcal{C}(M_{\text{obs}}, z) = C_0 \left(\frac{1+z}{1+z_0} \right)^\alpha \left(\frac{M_{\text{obs}}}{10^{13.8} \, \text{M}_\odot} \right)^\beta \tag{35}$$

with $z_0 = 0.5$ as pivot redshift. The overall mass bias was found to be consistent with unity, with $C_0 = 1.00 \pm 0.02$, and mildly redshift and mass dependent, $\alpha = -0.071 \pm 0.080$ and $\beta = 0.026 \pm 0.029$. The redshift dependence reflects that the fit is done for each of the four snapshots. Given this mass calibration, the final estimator for the mean weak-lensing mass of a cluster stack at redshift z is

$$M'_{\text{obs}} = \mathcal{C}(M_{\text{obs}}, z) \, M_{\text{obs}}. \tag{36}$$

We calculated any residual systematic mass errors, and found them to be consistent with zero for all redshift and mass subsets.

5.3 Triaxiality and projection effects

When finding galaxy clusters in photometric data, systems that are aligned along the line of sight are typically selected with higher probability than those otherwise oriented. This break in isotropy must be accounted for when inferring the mean mass density profile in a 'stacked' analysis such as we perform here. Likewise, cluster selection can be affected by the presence of other objects along the line of sight, which increase both the observed cluster richness and the recovered weak-lensing mass. These effects have been discussed in a variety of works (White, Cohn & Smit 2010; Angulo et al. 2012; Noh & Cohn 2012; Dietrich et al. 2014), and they have competing impacts on the cluster mass–richness relation.

Projection effects tend to boost the cluster richness more than the recovered weak-lensing mass. This is easily understood: since

the weak-lensing profile $\Delta\Sigma$ is a differential measurement, any centring offset between the projected halo and the main halo has a more severe impact on the weak-lensing signal $\Delta\Sigma$ than on the cluster richness. The net effect of projections is actually to somewhat reduce the mean inferred mass at given measured richness.

Following Simet et al. (2017a), we let $\langle M \rangle_0$ be the mass of a galaxy cluster not affected by projection effects, and let p be the fraction of clusters that are affected by projections. We model a projected cluster as a sum of a primary halo that must have at least a mass $0.5\langle M \rangle_0$, and an excess mass $\epsilon\langle M \rangle_0$ where $\epsilon \in [0.0, 0.5]$. We adopt a Gaussian prior for ϵ of $\epsilon = 0.25 \pm 0.15$, so that both $\epsilon = 0$ and $\epsilon = 0.5$ are within 2σ of the central value. Simet et al. (2017a) estimated the projection rate to be $p = 10$ per cent ± 4 per cent.

We can then write the average mass of the cluster stack as

$$\langle M \rangle = (1 - p)\langle M \rangle_0 + p(0.5 + \epsilon)\langle M \rangle_0 \tag{37}$$

The mass in the absence of projection effects is simply $\langle M \rangle_0$, so to recover $\langle M \rangle_0$ from $\langle M \rangle$ we need to multiply the recovered weak-lensing masses by

$$\frac{\langle M \rangle_0}{\langle M \rangle} = \frac{1}{1 + p(\epsilon - 0.5)} = 1.02 \pm 0.02. \tag{38}$$

The numerical value above is estimated from 10^4 Monte Carlo realizations of p and ϵ within the respective priors.

Unlike projection effects, the preferential alignment of haloes along the line of sight tends to boost the recovered weak-lensing mass at fixed measured richness relative to the true halo mass. Using numerical simulations, Dietrich et al. (2014) estimated this effect leads to an overestimate of cluster masses by 4.5 per cent $\pm$ 1.5 per cent. This estimate can be understood as correlated scatter between optical richness and weak-lensing masses, which leads weak-lensing masses to overestimate cluster masses by an amount $\exp(-\beta \, r \, \sigma_{\ln M|\lambda} \, \sigma_{\ln M|M_{\text{WL}}})$, where $\beta \approx 3$ is the slope of the halo mass function, and r is the correlation coefficient between richness and weak-lensing mass. Adopting $r \in [0, 0.5]$ (Noh & Cohn 2012), $\sigma_{\ln M|\lambda} = 0.25 \pm 0.05$ (Rozo & Rykoff 2014) and $\sigma_{\ln M|M_{\text{WL}}} = 0.25 \pm 0.05$ we arrive at a correction factor 0.96 ± 0.02, in excellent agreement with the Dietrich et al. (2014) result.

Put together, these two effects modify the recovered weak-lensing masses by a multiplicative factor 0.98 ± 0.03.[11] This correction factor can be readily absorbed into the calibration correction parameter $C_0 = 1.00 \pm 0.02$ described in the previous section, resulting in $C_0 = 0.98 \pm 0.04$.

5.4 Centring correction

In our calibration strategy, we have thus far assumed that we can measure the stacked shear profile of clusters relative to the 'centre' of the halo as defined in an N-body simulation. In the simulations we used to calibrate our weak-lensing masses (Section 5.2), haloes were found using the ROCKSTAR algorithm (Behroozi et al. 2013), which identifies the halo centre as the mean position of a judiciously chosen subset of particles near the halo's density peak. This is the point relative to which our halo–matter correlation function is meant to be defined.

The density peak, corresponding to a 'centre' as defined above, is expected to host a massive galaxy, which, in many cases, can

---

[11] The balance between these two effects mildly depends on richness and redshift, while we will assume it to be constant in what follows. We will evaluate the cumulative effect of all the calibration terms in Section 5.7.

be correctly identified from photometric data. However, sometimes one may not be able to unambiguously determine which of the various cluster galaxies correspond to the density peak.

Consequently, we adopt a model in which part of the cluster sample is correctly centred and the remaining fraction $f_{\rm mis}$ is off-centred with some radial distribution $p(R_{\rm mis})$. Both $f_{\rm mis}$ and the width of $p(R_{\rm mis})$ we take as free parameters in each redshift and richness subset. Correspondingly, we model the recovered weak-lensing signal as a weighted sum of two independent contributions: a contribution $\Delta\Sigma$ from properly centred clusters, and a contribution $\Delta\Sigma_{\rm mis}$ from miscentered galaxy clusters,

$$\Delta\Sigma_{\rm model} = (1 - f_{\rm mis})\Delta\Sigma + f_{\rm mis}\Delta\Sigma_{\rm mis}. \tag{39}$$

When a cluster is miscentered by some radial offset $R_{\rm mis}$, the corresponding azimuthally averaged surface mass density is (e.g. Yang et al. 2006; Johnston et al. 2007b)

$$\Sigma_{\rm mis}(R \mid R_{\rm mis}) = \int_0^{2\pi} \frac{{\rm d}\theta}{2\pi} \Sigma\left(\sqrt{R^2 + R_{\rm mis}^2 + 2RR_{\rm mis}\cos\theta}\right). \tag{40}$$

Letting $p(R_{\rm mis})$ be the distribution of radial offsets for miscentered clusters, the corresponding $\Sigma_{\rm mis}$ profile is obtained by averaging over the ensemble,

$$\Sigma_{\rm mis}(R) = \int {\rm d}R_{\rm mis}\, p(R_{\rm mis})\, \Sigma_{\rm mis}(R \mid R_{\rm mis}). \tag{41}$$

Rykoff et al. (2016) modelled the distribution $p(R_{\rm mis})$ for the DES SV redMaPPer clusters with a Rayleigh distribution. The ansatz assumes that radial vector displacements are drawn from a 2D Gaussian with constant variance σ_R, which gives the characteristic magnitude of the resulting radial offsets. They further assumed that σ_R is a fraction of the cluster radius R_λ,

$$\sigma_{\rm R} = c_{\rm mis}\, R_\lambda. \tag{42}$$

By comparing the assigned redMaPPer cluster centres to the cluster centres estimated using high-resolution X-ray data and SZ data from the South Pole Telescope, accounting also for the uncertainty of the X-ray and SZ centres, Rykoff et al. (2016) were able to place the empirical constraint

$$\ln c_{\rm mis} = -1.13 \pm 0.22 \tag{43}$$

and found the fraction of miscentered clusters to be

$$f_{\rm mis} = 0.22 \pm 0.11. \tag{44}$$

We will adopt these values as priors in our analysis, though how best to do so is unclear. One could, for instance, let the parameters $f_{\rm mis}$ and $c_{\rm mis}$ vary between richness bins. Alternatively, one could allow for redshift dependence of these parameters, but no richness dependence, or vice versa, or both. We choose to allow the miscentring parameters $f_{\rm mis}$ and $c_{\rm mis}$ to vary independently between subsets. Should the weak-lensing data strongly favour models with varying $f_{\rm mis}$ and $c_{\rm mis}$, our chosen parametrization should allow such trends to emerge from the data. We further note that the priors, while determined for massive SPT clusters, conform well with analytic expectations for lower mass clusters. We will test in Section 5.7 whether our assumption of constant priors is justified.

5.5 Boost factor model

In Section 4.2, we noted that membership dilution biases the recovered weak-lensing profile by a factor $1 - f_{\rm cl}$. In the literature, the factor $(1 - f_{\rm cl})^{-1}$ is often referred to as a boost factor or correction factor, sometimes denoted $C(R)$, and is used to boost the recovered profile by the appropriate amount. While we choose to leave the data untouched – and therefore dilute the theoretical profile rather than boost the data – we parametrize the boost factor $\mathcal{B} \equiv (1 - f_{\rm cl})^{-1}$ when constructing a model for the cluster-member contamination:

$$\mathcal{B}(\lambda, z, R) = 1 + B_0 \left(\frac{\lambda}{\lambda_0}\right)^{C_\lambda} \left(\frac{1+z}{1+z_0}\right)^{D_z} \left(\frac{R}{R_0}\right)^{E_R} \tag{45}$$

where B_0, C_λ, D_z and E_R are parameters in the fit. We choose richness, redshift and radial pivots as $\lambda_0 = 30$, $z_0 = 0.5$ and $R_0 = 500\,{\rm kpc}$, respectively.

We also need to address that the procedure to infer the boost factors described in Section 4.2 yielded point estimates without uncertainties. We account for that by modelling the associated uncertainty on the boost factor with the assumed form

$$\sigma_{\mathcal{B}}(R) = \sigma_{1\,{\rm Mpc}}\left(\frac{1\,{\rm Mpc}}{R}\right), \tag{46}$$

where $\sigma_{1\,{\rm Mpc}}$ is an unknown parameter that represents the error at a pivot distance of $1\,{\rm Mpc}$. The $1/R$ dependence is expected for logarithmically spaced radial bins and Poissonian errors: $\sigma_b \propto 1/\sqrt{N(R)} \propto 1/R$. The corresponding log-likelihood of the measured $f_{{\rm cl},k}$ in cluster subset k given the parameters in equation (45) is then given by

$$\ln\mathcal{L}(f_{{\rm cl},k} \mid B_0, C_\lambda, D_z, E_R) = -\sum_R \frac{\left((1 - f_{{\rm cl},k}(R))^{-1} - \mathcal{B}(R)\right)^2}{2\sigma_{\mathcal{B}}^2(R)}$$
$$+ \frac{1}{2}\log\sigma_{\mathcal{B}}^2(R). \tag{47}$$

We will constrain these parameters simultaneously from all richness and redshift subsets. Because the boost factors also affect the $\Delta\Sigma$ model, we will fit it in conjunction with the lensing data to account for any possible degeneracies between their respective parameters.

5.6 The complete likelihood

As we found in Section 3.1 and Fig. 4, the shear profile measurements of individual richness–redshift subsets are nearly independent from each other. For any subset k we can thus write the log-likelihood for a measured $\widetilde{\Delta\Sigma}$ given the halo mass M (and the other nuisance parameters listed in Table 2) as

$$\ln\mathcal{L}(\Delta\Sigma_k \mid M_k, \ldots) \propto -\frac{1}{2}\mathbf{D}_k^T\, \mathbf{C}_k^{-1}\mathbf{D}_k,$$

$$\text{where } D_{k,l} \equiv \widetilde{\Delta\Sigma}_k(R_l) - \frac{A_{m,k}\,\Delta\Sigma_{\rm model}(R_l \mid M_k, f_{{\rm mis},k}, c_{{\rm mis},k})}{\mathcal{B}(R_l \mid B_0, C_\lambda, D_z, E_R; k)} \tag{48}$$

and $\widetilde{\Delta\Sigma}(R_l)$ is the measurement in the l-th radial bin from equation (11), and $\mathbf{C}$ is the corresponding covariance matrix from Section 3.1. The factor $A_m = 1 + m - \delta$ combines the effects of shear (m, Section 4.1) and photo-z (δ, Section 4.3) systematic uncertainties. Since both m and δ are assigned Gaussian priors, the width of the prior on A_m is obtained by adding the widths of the priors on m and δ in quadrature. We arrive at

$$A_m = \begin{cases} 1.019 \pm 0.034 & \text{for } z \in [0.2, 0.4] \\ 1.020 \pm 0.038 & \text{for } z \in [0.4, 0.6] \\ 1.044 \pm 0.039 & \text{for } z \in [0.6, 0.8]. \end{cases} \tag{49}$$

Note that the factor A_m alters the prediction $\Delta\Sigma_{\rm model}$ from equation (39), as opposed to correcting the data. Our approach has the benefit of preserving the covariance matrix: an alteration of the

Table 2. Parameters entering $\mathcal{L}(\Delta\Sigma)$ (equation 48) and $\mathcal{L}(f_{\rm cl})$ (equation 47). Flat priors are specified with limits in square brackets, Gaussian priors with means $\pm$ standard deviations. The posteriors are given as the mean and symmetrized 68 per cent confidence intervals. For the top half, the posteriors depend on the cluster subset, so we list the uncertainties as average confidence intervals followed by the scatter between subsets.

Parameter	Description	Prior	Posterior
$\log_{10}M$	Halo mass	[12.0, 16.0]	Table 3
$\ln c_{\rm mis}$	Miscentring offset	-1.13 ± 0.22	$-1.06 \pm 0.22 \pm 0.05$
$f_{\rm mis}$	Miscentered fraction	0.22 ± 0.11	$0.24 \pm 0.10 \pm 0.05$
A_m	Shape and photo-z bias	equation (49)	$1.026 \pm 0.037 \pm 0.013$
$\log_{10}B_0$	Boost magnitude	$[-4.0, 0.0]$	-1.399 ± 0.040
C_λ	Richness scaling	[0.25, 1.5]	0.920 ± 0.106
D_z	Redshift scaling	$[-10, 10]$	-4.00 ± 0.79
E_R	Radial scaling	$[-1.5, 1.0]$	-0.98 ± 0.09

data vector would necessarily force us to also adjust the effective covariance matrix.

The total log-likelihood for our analysis is the sum of the boost factor [equation (47)] and weak-lensing log-likelihoods:

$$\ln\mathcal{L} = \sum_{k=(\lambda,z)} \ln\mathcal{L}_k \quad \text{with}$$

$$\ln\mathcal{L}_k \equiv \ln\mathcal{L}(\Delta\Sigma_k \,|\, M_k, A_{m,k}, f_{\rm mis,k}, c_{\rm mis,k}, B_0, C_\lambda, D_z, E_R; k)$$
$$+ \ln\mathcal{L}(f_{\rm cl,k} \,|\, B_0, C_\lambda, D_z, E_R; k). \qquad (50)$$

It is thus apparent that we seek to constrain independent subset masses M_k and global boost factor parameters, the latter constrained by their effect on the $\Delta\Sigma$ profile as well as independent measurements of $f_{\rm cl}$.

5.7 Stacked cluster masses

The likelihood is sampled using the package EMCEE[12] (Foreman-Mackey et al. 2013) that allows a parallelized exploration of the parameter space. We use 20 walkers with 10 000 steps each, and discard the first 3000 steps as burn-in. We test whether the chains have converged first with an independent run of only 5000 steps per walker, which produces nearly identical results. The chains of single walkers become uncorrelated (with a correlation coefficient $|r| < 0.1$) after about 23 steps, which is much shorter than the length of each chain. The resulting number of independent draws for all walkers is $\approx$6000. We therefore believe that the likelihood has been exhaustively explored and that our inference results are robust.

The complete list of model parameters as well as their corresponding priors and posteriors are summarized in Table 2.

After determining the best-fitting masses M for each cluster subset, we apply the calibration correction described in Section 5.2 to the recorded chains. Specifically, for each point in the chain, we randomly sample the mass calibration factor $\mathcal{C}(M, z)$ from its posteriors, and replace the mass parameter value with $M'_{\rm obs}$ from equation (36). Such a postponed correction is valid because $\mathcal{C}$ is independent from other parameters in the chain, and results in an updated chain that incorporates the mass calibration and its corresponding uncertainties.

We run four variants of the final likelihood evaluation to quantify statistical and systematic uncertainties in our analysis.
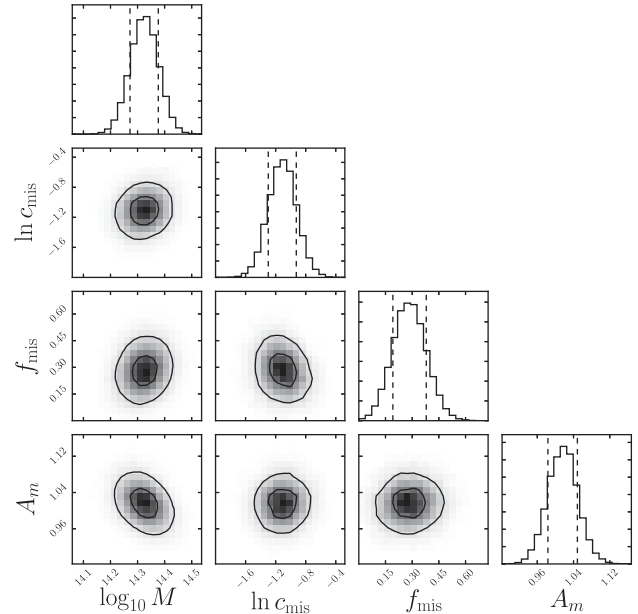
Figure 10. Parameters of $\Delta\Sigma_{\rm model}$ for the $z \in [0.2; 0.4]$, $\lambda \in [20, 35]$ subset. Contours denote the 68 per cent and 95 per cent confidence areas; dashed lines in the 1D histograms refer to 68 per cent confidence intervals.

(i) Full: all systematic parameters (modelling bias parameters, triaxiality and projection, shape and photo-z systematics, boost factors and miscentring) are allowed to vary within their respective priors. This constitutes our fiducial analysis.

(ii) FixedAm: all systematic parameters are allowed to vary within their priors, except we set $A_m = 1$ to determine the influence of only the combined shape and photo-z uncertainties.

(iii) Fixed: all systematic parameter priors are set to δ-functions at their central values to estimate the statistical uncertainties.

(iv) NonLinear: identical to Full but with a modified data vector $\Delta\Sigma \rightarrow (1-\kappa)\Delta\Sigma$, where the convergence κ was measured from the simulated mass profiles with the same subsets in mass and redshift as in Section 5.2, to approximate the non-linearity bias on the weak-shear estimator in equation (3).

The results of the Full likelihood evaluation for the parameters of $\Delta\Sigma_{\rm model}$ under the priors from Table 2 are shown in Fig. 10 for the example richness–redshift subset of $z \in [0.2, 0.4]$ and $\lambda \in [20, 35]$. The corresponding lensing data and best-fitting lensing profile are shown in Fig. 11, where we also single out the impact of miscentring and boost factors. The best-fitting models for all redMaPPer cluster subsets are overplotted on top of our weak-lensing data in Fig. 3.

As we can see from the posteriors in Table 2, the boost factors amount to a correction $B_0 \approx -4$ per cent at the pivot values of $\lambda_0 = 30$, $z_0 = 0.5$, $R_0 = 500\,{\rm kpc}$ and are consistent with linear scaling in λ and $1/R$ radial scaling, both expected from the number density of cluster member galaxies in the inner part of an NFW halo.

Posteriors on the miscentring parameters are only weakly constrained by our data. We find a weak correlation of M with $f_{\rm mis}$ (and none with $c_{\rm mis}$) in most cluster subsets, but there is no apparent trend of either miscentring parameter with richness or redshift. While tighter constraints could in principle be found by combining the lensing measurement with angular clustering of the putative
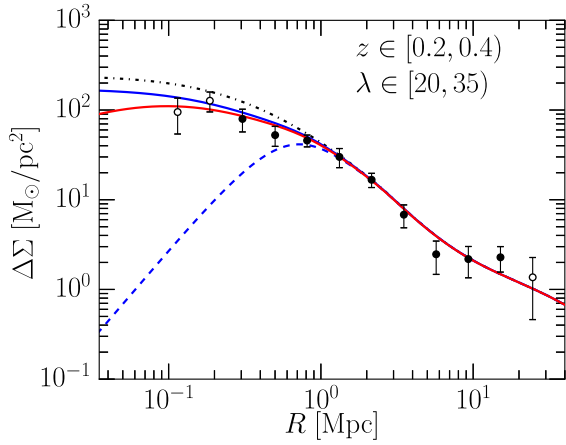
Figure 11. Fit with all components of $\Delta\Sigma_{\text{model}}$ and $\mathcal{B}$ to the cluster subset with $z \in [0.2; 0.4)$ and $\lambda \in [20, 35)$. The analytical model for the perfectly centred lensing signal $\Delta\Sigma$ from equation (31) is shown as dash–dotted black curve. The *solid blue curve* includes the effects of miscentring. The *solid red curve* additionally includes the effects of cluster member contamination from equation (45) and corresponds to the model that is fit to the data, shown as best-fitting curves in Fig. 3. We exclude data at $R < 200\,\text{kpc}$ (*open markers*) to avoid problems due to crowded field photometry and large boost-factor corrections in the cluster core. The *dashed blue curve* corresponds to the miscentered profile $\Delta\Sigma_{\text{mis}}$.

centres and member galaxies (van Uitert et al. 2016), we find that, at least for now, the uncertainties in mass due to miscentring are subdominant to statistical and other systematic uncertainties (cf. Table 5). We postpone further investigation of cluster miscentring to forthcoming works with larger DES data volumes.

To quantify statistical and systematic uncertainties of the fiducial analysis, we then perform the `Fixed` likelihood evaluation. We determine the systematic contribution to the fiducial uncertainty as the difference of uncertainties in quadrature between the `Full` and the `Fixed` run. In Table 3 we list central values and uncertainties for all subsets, and split the latter into statistical and systematic contributions.

Finally, to estimate the impact of the weak-shear assumption $g \approx \gamma$ in equation (3), we perform the `NonLinear` run, where we applied the first-order correction $\Delta\Sigma \rightarrow (1 - \kappa)\Delta\Sigma$. We find a mild overestimation of the masses from the `Full` run by ≈ 2 per cent for the highest richness subsets; all other subsets are affected at the subpercent level. Given the uncertainties in the current analysis, we will ignore this bias and its potential impact on the richness–mass relation. For future analyses, we will adopt correction schemes to suppress the non-linearity bias of the shear estimator (e.g. Seitz & Schneider 1997; Johnston et al. 2007a).

6 THE MASS–RICHNESS–REDSHIFT RELATION

We characterize the mass–richness relation of the DES SV redMaP-Per galaxy clusters as

$$\mathcal{M}(\lambda, z) \equiv \langle M \mid \lambda, z \rangle = M_0 \left(\frac{\lambda}{\lambda_0} \right)^{F_\lambda} \left(\frac{1 + z}{1 + z_0} \right)^{G_z}, \tag{51}$$

where M_0, F_λ and G_z are parameters of the model with pivot values $\lambda_0 = 30$ and $z_0 = 0.5$. We note the important distinction between M, the mass of a halo, itself a random variable, and $\mathcal{M}$, the expectation value of that random variable. For each cluster subset k, the expectation value is given by

$$\mathcal{M}_k = \frac{\sum_{j \in k} W_j \mathcal{M}(\lambda_j, z_j)}{\sum_{j \in k} W_j}. \tag{52}$$

The weights w_j of individual clusters j in subset k differ from unity for two reasons: (1) the lensing weight of each lens–source pair depends on the cluster's redshift, and (2) lower redshift clusters have more sources in any given radial bin because a fixed physical radial bin of a low-redshift cluster subtends a larger angle in the sky than the same bin does for a higher redshift cluster. We estimate the weight W_j as the sum of weights $w_{j,i}$ of all lens–source pairs around cluster j [given in equation (9)] over the radial range 0.3–3 Mpc. The choice of radial range has a subpercent impact on our results.

There is one subtle effect that remains unaccounted for in the above formula: in practice, we do not stack cluster masses. Rather, we stack the density profiles $\Delta\Sigma$. Using our analytic model for $\Delta\Sigma$, we can readily estimate the logarithmic dependence on mass of the density profile $\Delta\Sigma$ at any given radius R

$$\Gamma(R) = \frac{\mathrm{d}\ln\Delta\Sigma(R \mid M)}{\mathrm{d}\ln M}. \tag{53}$$

This logarithmic slope varies from ≈ 0.5 in the innermost regions of the density profile we utilize to ≈ 1.0 on the outskirts, with a typical value of ≈ 0.75 over a broad range of scales. For specificity, from here on we compute Γ within the radius $R_{200\text{m}}$, for which we find $\Gamma = 0.74 \pm 0.01$ within the range of masses and redshifts probed in this work. Given that we stack $\Delta\Sigma \propto M^\Gamma$, and that we then recovered the mean $\Delta\Sigma$ profile and turn it into a mass, i.e. $M \propto \Delta\Sigma^{1/\Gamma}$, we arrive at the appropriately weighted mean mass for each subset

$$\mathcal{M}_k = \left(\frac{\sum_{j \in k} W_j \mathcal{M}(\lambda_j, z_j)^\Gamma}{\sum_{j \in k} W_j} \right)^{1/\Gamma}. \tag{54}$$

For $\Gamma = 1$ this reduces to equation (52). We compare the predictions of this model to the mass M_k of subset k measured from lensing after marginalizing over all other parameters (cf. Table 3). Since we find the posterior probability density of M_k to be best described as lognormal (as expected from Stanek et al. 2010), we can express

Table 3. The calibrated marginalized posterior masses for each richness–redshift subset. Masses are given as $\log_{10}[M_{200\text{m}}]$ in units of $M_\odot$, uncertainties denote symmetrized 68 per cent confidence intervals marginalized over all other parameters. We first list the contributions from the statistical, then systematic uncertainties. In case direct matching to NFW halo results is desired, these can be obtained by reverting the calibration of equation (35).

λ	$z \in [0.2, 0.4)$	$z \in [0.4, 0.6)$	$z \in [0.6, 0.8)$
[5, 10)	$13.300 \pm 0.095 \pm 0.025$	$13.457 \pm 0.056 \pm 0.025$	$13.442 \pm 0.141 \pm 0.027$
[10, 14)	$13.758 \pm 0.115 \pm 0.011$	$13.520 \pm 0.158 \pm 0.015$	$13.637 \pm 0.152 \pm 0.033$
[14, 20)	$14.034 \pm 0.087 \pm 0.029$	$13.962 \pm 0.088 \pm 0.024$	$14.096 \pm 0.149 \pm 0.019$
[20, 35)	$14.324 \pm 0.065 \pm 0.018$	$14.297 \pm 0.080 \pm 0.018$	$14.114 \pm 0.122 \pm 0.024$
[35, 180)	$14.592 \pm 0.070 \pm 0.017$	$14.619 \pm 0.080 \pm 0.020$	$14.664 \pm 0.089 \pm 0.021$

the log-likelihood of the mass–richness relation parameters as

$$\ln \mathcal{L}(M_{\rm obs} \mid M_0, F_\lambda, G_z) \propto -\frac{1}{2}(\Delta \log M)^T \, \mathbf{C}_M^{-1} \, (\Delta \log M), \quad (55)$$

where $\mathbf{C}_M$ is the covariance matrix of the logarithm of the inferred masses M_k, which we shall derive below, and

$$\Delta \log M_k = \log M_k - \log \mathcal{M}_k. \quad (56)$$

The covariance matrix $\mathbf{C}_M$, however, does not have a trivial form. While we have found the statistical errors of lensing masses from the different richness and redshift subsets to be nearly uncorrelated (as discussed in Section 3.1), the systematic uncertainties are necessarily correlated. The primary source of calibration uncertainty was found to be model bias due to using a simple exponential disc to model all source galaxies. While the bulge fraction is certainly a function of redshift, the lensing kernel is broad enough that we expect the overall calibration error to be similar for all bins in lens redshift and richness. Likewise, photo-z systematics impact the mass measurements of all clusters. We therefore consider the recovered mass M_k and its prediction $\mathcal{M}_k$, and write uncertainties in $\log M_k$ as a sum of an uncorrelated component μ_k and a correlated component ν_k:

$$\log M_k = \log \mathcal{M}_k + \mu_k + \nu_k. \quad (57)$$

The diagonal elements of the covariance matrix are given by

$$\mathbf{C}_{M,kk} = \langle \mu_k^2 \rangle + \langle \nu_k^2 \rangle, \quad (58)$$

which is equal to the posterior uncertainties of the `Full` run in Table 3. We then evaluate the `FixedAm` run, which sets the potentially correlated parameter $A_{m,k} = 1 \, \forall k$ and thereby enforces $\nu_k = 0$, so that the recovered errors correspond to the uncorrelated statistical noise component $\langle \mu_k^2 \rangle$ only. Subtracting the two errors, we arrive at the correlated noise $\langle \nu_k^2 \rangle$.

Having solved for the correlated noise $\langle \nu_k^2 \rangle$ for each subset k, we compute the full covariance matrix $\mathbf{C}_{M,ij}$ by assuming the photo-z and shear systematic uncertainties are fully correlated between different cluster subsets. This assumption is conservative as the uncertainties are not reducible by 'averaging out' the impact of these systematics across different subsets. We thus set

$$\mathbf{C}_{M,ij} = \delta_{ij}\langle \mu_i^2 \rangle + \langle \nu_i \nu_j \rangle = \delta_{ij}\langle \mu_i^2 \rangle + \left[\langle \nu_i^2 \rangle \langle \nu_j^2 \rangle\right]^{1/2} \quad (59)$$

as the covariance matrix we employ in equation (55).

Note that with this scheme, we explicitly enforce correlated uncertainties of shear and photo-z systematics only, whereas other systematics are considered independent across subsets. Independent systematics will tend to 'average out' across bins, reducing their impact on the uncertainty in the amplitude of the scaling relation, while increasing their impact on the slope of the scaling relation. In our case, assuming non-photo-z and non-shear systematics are independent across bins reduces their effect on the amplitude of the mass–richness relation from 2 per cent on each independent subset to ≈ 1 per cent (see Table 5 for our systematics error budget).

As before, we use EMCEE to sample the likelihood of the mass–richness relation parameters as constrained from the DES SV data. Our fit is restricted to subsets with $\lambda \geq 20$ to ensure that galaxy clusters can be unambiguously identified with prominent dark matter haloes.[13] The best-fitting parameters are summarized in Table 4, and the corresponding confidence contours are shown in Fig. 12.

---

[13] The richness threshold $\lambda \geq 20$ has been adopted in all redMaPPer works since its original publication as a sufficiently conservative cut to ensure

Table 4. Parameters of the M–λ–z relation from equation (51) with their flat priors and resulting posteriors. The mass is defined as M_{200m} in units of $M_\odot$. Uncertainties denote symmetrized 68 per cent confidence intervals and are split into statistical (first) and systematic (second).

Parameter	Description	Prior	Posterior
$\log_{10} M_0$	Mass pivot	[12.0, 16.0]	$14.371 \pm 0.040 \pm 0.022$
F_λ	Richness scaling	[−10, 10]	$1.12 \pm 0.20 \pm 0.06$
G_z	Redshift scaling	[−20, 20]	$0.18 \pm 0.75 \pm 0.24$
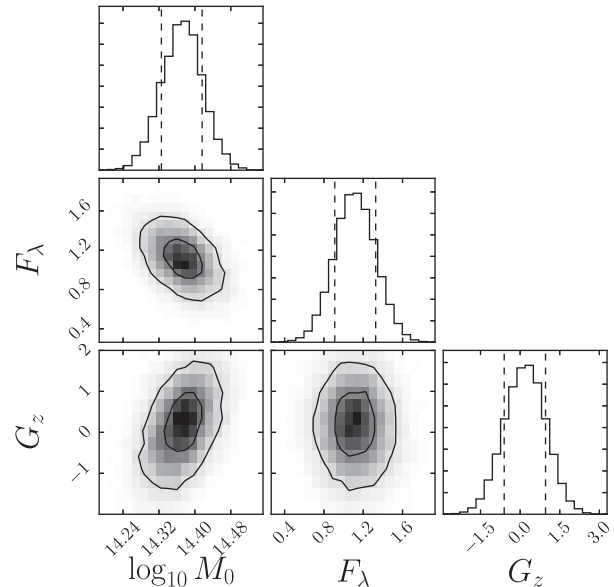
Figure 12. Parameters of the M–λ–z relation. Contours denote the 68 per cent and 95 per cent confidence areas from the `Full` run; dashed lines in the 1D histograms refer to the 68 per cent confidence intervals.

We repeat the analysis using the statistical errors from the `Fixed` run. That is, we constrain the mass–richness using statistical errors only. The central values of the resulting parameters are nearly identical to the parameters we inferred when marginalizing over all systematic uncertainties. The difference in quadrature between the two uncertainties is reported as the systematic uncertainty in the recovered mass–richness–redshift relation in Table 4.

We also carry out the analysis with an extreme value $\Gamma = 1$ in equation (54) to address the concern that our treatment does not fully capture the variation in Γ across all the scales being utilized. We find that it changes the recovered amplitude of the mass–richness relation by $\Delta \log M_0 = 0.003$, which is clearly subdominant compared to the other uncertainties in the analysis.

Our results imply that the mean mass of galaxy clusters of richness $\lambda = 30$ at redshift $z = 0.5$ is $\log \mathcal{M} = 14.371 \pm 0.040 \, (\text{stat}) \pm 0.022 \, (\text{sys})$, with a richness scaling that is slightly steeper than linear and no strong redshift evolution. This corresponds to a 10.5 per cent calibration (9.2 per cent statistical, 5.1 per cent systematic) of the amplitude of the mass–richness relation.

The current statistical uncertainties reflect the small area of the DES SV data set; we expect a similar analysis of the full survey will yield ≈ 1 per cent uncertainty after five years of operations. These

---

clean cluster samples. As reference, the typical richness uncertainty due to random projections for a cluster with $\lambda = 5$ is ± 1.5, therefore for a $\lambda = 20$ system to not be associated with at least one halo of richness $\lambda = 5$ would constitute a 10σ fluctuation.
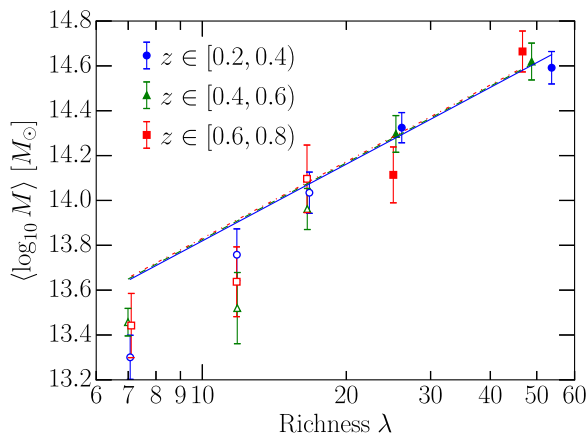
Figure 13. Best fitting model for $M(\lambda, z)$. Only subsets with richnesses $\lambda \geq 20$ (*filled markers*) are fit. Points are placed at the mean richness of their subset.

results highlight the need for substantial improvement in systematic error control within a short time frame.

Like Simet et al. (2017a), we expect the dominant systematic uncertainty in our analysis to stem from shear and photo-z systematics. We confirm this by repeating the analysis with the FixedAm run, which accounts for all systematics *except* multiplicative shear bias and photo-z systematics. As expected, the recovered posterior distributions are significantly narrower than the posteriors that include shear and photo-z systematics. We find that those two systematics alone contribute 78 per cent of the systematics-sourced variance. The remaining 22 per cent of the systematics variance is almost entirely due to projections, triaxiality and modelling systematics.

Fig. 13 compares the best-fitting mass–richness relation to the stacked cluster masses M_j. Our model presents an excellent fit for the considered range of $\lambda \geq 20$, and extrapolates well down to the $\lambda \geq 14$ cluster sample. Clusters with even lower richnesses fall significantly below our prediction, likely reflecting contamination from line-of-sight overdensities. Our results are consistent with the expectation that a richness threshold $\lambda \geq 20$ is a conservative choice for cluster scaling relation studies.

7 COMPARISON TO RESULTS IN THE LITERATURE

We now compare our calibration of the mass–richness relation to results in the literature, based on SPT Sunyaev-Zel'dovich Effect (SZE) measurements of DES redMaPPer clusters (Section 7.1) and weak lensing by SDSS redMaPPer clusters (Section 7.2). We also compare our systematic uncertainties to previous cluster lensing studies (Section 7.3).

7.1 Comparison to Saro et al. (2015)

Saro et al. (2015, hereafter S15) provided the first, indirect calibration of the richness–mass relation of DES redMaPPer clusters by cross-matching the SV cluster catalogue to the SPT cluster catalogue of Bleem et al. (2015). To do so, S15 first assumed a cosmology identical to our fiducial cosmology (flat ΛCDM with $\Omega_m = 0.3$, $h = 0.7$, with $\sigma_8 = 0.8$), then determined the best-fitting relation between the SZE signal-to-noise ratio ξ and the cluster mass by abundance matching, i.e. by comparing the predicted number of clusters above the observable threshold to a prediction based on the mass function and observable–mass relation. This allowed them to
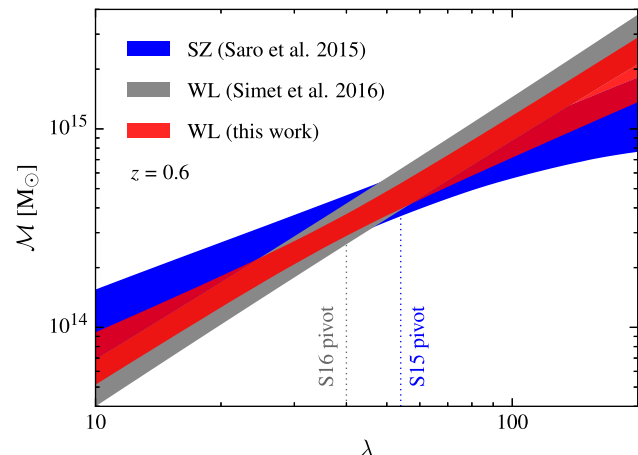
Figure 14. 68 per cent confidence intervals for the mean cluster mass M_{200m} as function of cluster richness λ at $z = 0.6$ as constrained by the analyses of Saro et al. (2015, *blue*), Simet et al. (2017a, *grey*) and this work (*red*). The Simet et al. (2017a) relation has been transported from its pivot redshift of 0.2 to 0.6 using our best-fitting redshift-evolution model.

transfer the ξ–M relation from the SPT clusters to the λ–M relation of DES redMaPPer clusters; thanks to the well-understood SPT cluster selection function.

To compare our results to S15, we directly sample the scaling relation parameters from the S15 chains. We then need to account for a key difference between our work and that of S15, namely that we constrain the mass–richness relation, while the latter constrains the richness–mass relation. One can use the approach of Evrard et al. (2014) to transform between the two. We refer the reader to that work for details, and simply note that the conversion requires a correction for the scatter in mass at fixed value of the observable. The correction itself depends on the first and second logarithmic derivatives of the halo mass function, which we compute using the Tinker et al. (2008) halo mass function evaluated at $z = 0.6$, the pivot redshift in the S15 relation.

Fig. 14 compares the recovered scaling relation from S15 to our results, after also converting the S15 masses from M_{500c} in units of $h^{-1} \, M_\odot$ to M_{200m} in units of $M_\odot$, using the method described by Hu & Kravtsov (2003) and the Bhattacharya et al. (2013) concentration–mass relation. Essentially identical results are obtained using the Diemer & Kravtsov (2015) relation. Although visually the slope of the mass–richness relation from S15 is shallower than ours, the difference is not significant: the two are consistent at the 1.3σ level. The amplitude of the scaling relations is in nearly perfect agreement: the S15 mean mass at their pivot richness of $\lambda = 54$ is $\log \langle M \,|\, \lambda = 54 \rangle = 14.67 \pm 0.11$, compared to $\log \langle M \,|\, \lambda = 54 \rangle = 14.66 \pm 0.06$ in our analysis.

7.2 Comparison to SDSS

Several studies have calibrated the mass–richness relation for SDSS redMaPPer clusters (Baxter et al. 2016; Farahi et al. 2016; Li et al. 2016; Miyatake et al. 2016; Simet et al. 2017a, S16 hereafter). All of them have been found to be consistent with each other. While the Li et al. (2016) calibration is the most precise (i.e. it has the smallest reported errors), it does not include a detailed analysis of the sources of systematic uncertainty. The measurements by Farahi et al. (2016) and Baxter et al. (2016) are useful systematics cross-checks, but are currently less precise than those from S16. Given the statistical consistency between the various methods, and the

fact that S16 is the most precise after accounting for systematic uncertainties, we focus on the comparison with S16.

S16 find that the mean mass of SDSS redMaPPer clusters of richness $\lambda = 40$ is $\log \langle M \,|\, \lambda = 40 \rangle = 14.344 \pm 0.021$ (stat) $\pm$ 0.023 (sys) at a reported pivot of $z = 0.2$. In that work, masses were measured as M_{200m} in units of $h^{-1}\,M_\odot$. At that richness and redshift, our results, converted to their units, correspond to $\log \langle M \,|\, \lambda = 40 \rangle = 14.339 \pm 0.070$, fully consistent with – but not as precise as – their result. The slopes of the mass–richness relations are also in good agreement, with a difference of 0.9σ.

We note that DES redMaPPer clusters in the SV data are somewhat more abundant than SDSS redMaPPer clusters at the same richness, although the difference does not appear to be statistically significant (Rykoff et al. 2016, their fig. 6). Nevertheless, a difference in abundance could in principle signify a systematic difference in the richness estimator λ between DES and SDSS, resulting in the same galaxy clusters being assigned different richnesses when observed in DES than when observed in SDSS. Conversely, clusters of the same observed richness λ, either from DES or from SDSS, would, on average, have somewhat different masses. The difference in cluster abundance between the two surveys could thus imply that DES SV clusters are about 7 per cent less massive than equally rich SDSS clusters.[14] Correcting for this effect, our prediction for the SDSS mass calibration would be $\log \langle M \,|\, \lambda = 40 \rangle = 14.373 \pm 0.053$, still in excellent agreement with the SDSS result. We emphasize that this effect does not affect the uncertainty in the mass calibration of the DES SV galaxy clusters; it is relevant only for the comparison of our mass calibration to that of S16. Whether the difference in number density is indeed due to a lower mass threshold in DES or merely a statistical fluctuation can be directly tested with future DES redMaPPer catalogues with larger area and partial SDSS overlap.

7.3 Comparison of systematic uncertainties to other weak-lensing cluster mass calibrations

Over the last two years, several collaborations have published results of extensive observational campaigns designed to calibrate cluster mass–observable relations. These include the Weighing the Giants (WtG; von der Linden et al. 2014a), the Canadian Cluster Comparison Project (CCCP; Hoekstra et al. 2015) and the Local Cluster Substructure Survey (LoCuSS; Okabe & Smith 2016). We also include the weak-lensing mass measurements from the Cluster Lensing and Supernova Survey with Hubble (CLASH; Postman et al. 2012; Umetsu et al. 2016) in this discussion, although the calibration of mass–observable relations is not their explicit goal. A direct comparison of our results with these works is not possible because we do not estimate individual cluster masses, and because of a lack of overlap between the clusters samples. Nevertheless, a discussion on the respective treatment of systematic uncertainties is warranted. For completeness, we also discuss the systematic error budget of S16.

We focus here on the two main sources of systematic uncertainty, multiplicative shear bias (m, cf. equation 13) and photo-z calibra-

---

[14] The ratio between the comoving space density of galaxy clusters in DES versus SDSS for clusters with richness $\lambda \geq 20$ over the redshift range $z \in$ [0.2, 0.3], where both DES and SDSS are complete, is 0.83 (Rykoff et al. 2016). If we assume that this ratio corresponds to different effective mass thresholds in the two surveys, we can use abundance matching to estimate the corresponding systematic mass difference between the two surveys.

tion (δ, cf. equation 23). For the former, each of the collaborations relied on simulations to calibrate their shear biases and to estimate the corresponding uncertainties. WtG employed calibrations from STEP2 (Massey et al. 2007) and state shear bias uncertainties of 3 per cent (Applegate et al. 2014). CLASH used STEP2 supplemented by similar, custom simulations from Oguri et al. (2012), and find 5 per cent uncertainty in m (Umetsu et al. 2016). CCCP used GALSIM (Rowe et al. 2015) to simulate analytic galaxy profiles, and quote 2 per cent shear calibration uncertainty. LoCuSS utilized two different sets of image simulations, one using the software package SKYMAKER (Bertin 2009) and another using the software package SHERA (Mandelbaum et al. 2012), and state 3 per cent uncertainty in m. S16 quote 3.5 per cent top-hat systematic uncertainty, roughly equivalent to 2.0 per cent Gaussian uncertainty, based on SHERA simulations.

The shear systematics error budget is comparable across all works. We caution, however, that our shear systematic is larger than what we would have estimated from the GREAT-DES simulations (Jarvis et al. 2016, their section 6.1) alone. Our systematic error estimate comes from the comparison of two independent source catalogues, IM3SHAPE and NGMIX, each of which we expected to have a multiplicative shear bias $|m| \leq 0.03$. Nevertheless, a detailed comparison of the two revealed a systematic uncertainty $|m| \leq 0.05$. We believe this increased systematic uncertainty reflects differences between our simulated images and our data.

There is a larger discrepancy between our work and how others have estimated photometric redshift systematics, which we find to be the second most important source of systematic uncertainty in our analysis. We emphasize that throughout this section we use the term 'photo-z systematics' to denote uncertainties in $\langle \Sigma_{\rm crit}^{-1} \rangle$ caused exclusively by biased performance of the photometric redshift estimator, e.g. due to insufficient template sets or priors and cosmic variance limiting the precision of the calibration. This systematic is distinct from cluster-member dilution, and any associated amelioration techniques such as the colour cuts employed in many studies (including WtG, LoCuSS and CLASH). We make this distinction because we find membership dilution to be a subdominant effect.

As with estimation of shear calibration systematics, our photometric redshift systematic error is estimated by comparing several independently produced photo-z catalogues. The corresponding systematic uncertainty in the weak-lensing signal ranges from 1.7 per cent for low-redshift clusters to 2.5 per cent for high-redshift clusters (cf. equation 27). This matches well the systematic uncertainty estimated directly from weighted spectroscopic validation data sets (Bonnett et al. 2016).

WtG, CCCP, LoCuSS and CLASH approach the problem of estimating photo-zs of the selected sources differently. WtG (Kelly et al. 2014) and CLASH use the template-fitting code BPZ (one of the codes we also employ) for determining photometric redshifts from five-band photometric data. Analyses with fewer imaging bands infer indirect photo-z estimates by matching the respective source samples according to magnitude (CCCP), or colour and magnitude (WtG, for their clusters with fewer than five photometric bands), or N-nearest neighbours in colour–magnitude space (LoCuSS) to a high-quality reference sample with excellent photo-z accuracy. All of these studies use the deep multiband photometry in the COSMOS field, and derive the reference photo-z estimates with the template codes LEPHARE (WtG and CLASH, with the public catalogue of Ilbert et al. 2009, LoCuSS with Ilbert et al. 2013) or EAZY (Brammer, van Dokkum & Coppi 2008, for CCCP).

The quoted systematic uncertainties in the lensing amplitudes from possible biases in the inferred redshift distribution vary from

work to work, but range from 1 per cent or less (WtG with five-band photometry, LoCuSS, CLASH) to as high as 4 per cent (WtG with two bands) and 8 per cent for high-redshift CCCP clusters. The latter two reflect cosmic variance for single colour and/or magnitude selected source galaxies. Our four-band photo-z error of ≈ 3 per cent in the mass is somewhat intermediate between the two extremes above. It is an accurate reflection of the uncertainty spanned by different photo-z codes with the four-band photometry, and in that sense it represents a *minimum* systematics floor due to algorithmic choices. It is, however, conceivable that our own photo-z error is somewhat underestimated if the variation across different photo-z codes is less than the impact of sample variance on the machine-learning photo-z codes. In Bonnett et al. (2016), we found the spread in photo-z codes to be comparable to observed uncertainties relative to $p(z)$ estimates of independent spectroscopic testing samples, we therefore believe our procedure to be fair. Nevertheless, future studies of this source of systematic uncertainty are highly desirable.

In this context, we consider the work by Nakajima et al. (2012) as an instructive baseline. Because the SDSS shape catalogue is restricted to sufficiently bright galaxies, one can construct spectroscopic galaxy samples that are exactly representative of the weak-lensing source samples. In this ideal scenario, the accuracy with which photo-z biases can be controlled is limited only by the size of these spectroscopic calibration sample, corresponding to a 3 per cent precision for SDSS clusters. In contrast to the SDSS analysis, sufficiently large, representative spectroscopic samples are currently unavailable at the depth of DES, and this lack is even more relevant at the depths of typical WtG, CCCP, LoCuSS and CLASH images. While some alleviation of this systematic naturally occurs with increasing separation of lenses and sources, it is difficult to imagine a superior scenario for photo-z bias calibration than that of Nakajima et al. (2012). This suggests that controlling photo-z systematics at the 2 per cent level or better with DES or other upcoming large-scale surveys will be difficult unless new methods for handling spectroscopic incompleteness are developed, and/or we are able to employ alternate redshift estimation techniques, e.g. the cross-correlation method of Newman (2008).

A reasonable test for the field of cluster lensing, as a whole, is the comparison of cluster masses across different works. From table 6 in Okabe & Smith (2016), we see that, depending on the precise mass definition, LoCuSS differs from CCCP by 5 per cent, and from WtG by 12 per cent. CCCP reports that their masses are in excellent agreement with WtG with a mean offset of ≈ 2 per cent, though they caution that CCCP differs from the WtG effort in how they convert shear profiles to masses. In particular, when CCCP follows the WtG prescription for estimating cluster masses, their recovered masses are lower than WtG by 8 per cent. The Umetsu et al. (2016) CLASH analysis is consistent with WtG and LoCuSS, yet finds marginally higher masses than CCCP by 16 per cent ± 10 per cent. These comparisons suggest that at present discrepancies between different groups remain at the ≈ 10 per cent level. Assuming equal parts statistical and systematic errors and ignoring possible overlaps of the cluster samples, systematic error budgets of ≈ 7 per cent appear adequate for the current state of the field, and are in reasonable agreement with corresponding estimates quoted in these works.

8 SUMMARY AND CONCLUSIONS

We measured the stacked weak-lensing signal of redMaPPer clusters in the DES SV data. The clusters were split into 15 non-overlapping richness and redshift subsets with $\lambda \geq 5$ and $0.2 \leq z \leq 0.8$, and the mean mass of each cluster stack was estimated from a model that accounts for the following:

(i) shear measurement systematics (Section 4.1),
(ii) dilution of the source sample by cluster members (Section 4.2),
(iii) source photometric redshift uncertainties (Section 4.3),
(iv) analytical modelling systematics (Section 5.2),
(v) triaxiality and projection effects (Section 5.3),
(vi) cluster miscentring (Section 5.4).

The set of masses were in turn used to determine the cluster mass–richness relation, as parametrized according to equation (51). The entire analysis was performed with a blinded shear catalogue, with a blinding factor between 0.9 and 1, which corresponds to 13 per cent uniform uncertainty on the amplitude of the mass–richness relation, comparable to a Gaussian uncertainty of about 8 per cent. By comparison, the total (statistical plus systematic) uncertainty in the amplitude of the mass–richness relation is 10.5 per cent, which implies that the blinding factor was marginally sufficient to avoid confirmation bias.

The mean cluster mass for clusters at our pivot richness of $\lambda = 30$ and pivot redshift of $z = 0.5$ is

$$M_0 = [2.35 \pm 0.22 \pm 0.12] \times 10^{14} \, \mathrm{M}_\odot. \tag{60}$$

The best-fitting slope F_λ for the mass–richness relation is

$$F_\lambda = 1.12 \pm 0.20 \pm 0.06, \tag{61}$$

while the best-fitting redshift-evolution parameter G_z is

$$G_z = 0.18 \pm 0.75 \pm 0.24. \tag{62}$$

The results are summarized in Tables 3 and 4.

We compared our inferred mass–richness relation for redMaPPer clusters to that of S15. The two works are in nearly perfect agreement with regards to the amplitude of the mass–richness relation. Our recovered slope is steeper, but the constraints from the two works are statistically consistent at 1.2σ. Since the S15 work used the abundance of SPT galaxy clusters and an $\Omega_m = 0.3$ and $\sigma_8 = 0.8$ flat ΛCDM cosmology, the excellent agreement between our work and that of S15 suggests that this cosmology is likely to provide a good fit to the abundance of redMaPPer clusters. An analysis of the abundance of redMaPPer clusters is in preparation.

We also compared our results to the weak-lensing mass calibration of SDSS redMaPPer clusters presented in S16. Our results are in excellent agreement with those of S16, both in amplitude and slope. The excellent agreement between the two works is encouraging given the difference in imaging and methods between the two works. We caution, however, that the uncertainty in the amplitude of the mass–richness relation of redMaPPer clusters at the S16 pivot redshift is ≈ 16 per cent, so the current agreement is not sufficient to test the consistency of the two works within the reported systematics uncertainties (roughly 5 per cent each at their respective pivot redshifts).

While our cluster mass calibration is currently statistics-limited, the situation will quickly change with the advent of more DES data. As Table 5 clearly shows, the systematic error budget is strongly dominated by calibration uncertainties of the multiplicative shear bias and the photo-z performance. The latter becomes increasingly important at higher redshifts.

In this work, we have assumed that photo-z and shear measurement systematics are perfectly correlated across all richness and redshift subsets. While it is clear that these systematics must be correlated at some level, it is not obvious to what degree. Our

Table 5. Systematic error budget on the amplitude of the mass–richness relation as measured with the DES SV data. The first two sources are taken as perfectly correlated between source subsets, while the next three are assumed to be independent (in brackets we list their impact if considered perfectly correlated as well). In case systematic errors are determined for the amplitude of the lensing signal $\Delta\Sigma$, we use the approximation $M \propto \Delta\Sigma^{4/3}$ (see discussion of equation 53).

Source of systematic	Amplitude uncertainty
Shear measurement	4 per cent
Photometric redshifts	3 per cent
Modelling systematics	1 per cent (2 per cent)
Cluster triaxiality	1 per cent (2 per cent)
Line-of-sight projections	1 per cent (2 per cent)
Membership dilution + miscentring	≤1 per cent
Total	**5.1 per cent** (6.1 per cent)

conservative assumption (in terms of the amplitude of the mass–richness relation) of perfect correlation means that there is no 'averaging' of these systematics across subsets. All other systematics were assumed to be uncorrelated. If we enforced perfect correlation on these systematics as well, our error budget would increase from 5.1 per cent to 6.1 per cent. In either case, the cumulative systematic uncertainty is comparable to reported values from other weak-lensing analyses (i.e. WtG, CCCP, LoCuSS, CLASH and S16).

An insight from our analysis is that systematic differences from independently produced shape and photometric redshift catalogues can reveal systematics that would otherwise go unnoticed. For the multiplicative shear calibration, the differences between the shear catalogues may partly be due to differences between DES data and the GREAT-DES simulations used to calibrate one of the pipelines. The discrepant behaviour of the photo-z estimators is less well understood, but likely due to a strong reliance on a comparison to the COSMOS field, whose small size and redshift distribution is not a representative of the DES survey.

With the rapid increase in observed DES area, future mass-calibration analyses using the same shear and photo-z methods would be systematics-limited. Efforts aimed at improving control over shear measurement and photometric redshift uncertainties are thus paramount. In the short term, shear measurement systematics can be reduced through improvements in the imaging simulations used to calibrate shear biases. More substantial gains will rely on shear estimation algorithms that are less sensitive to measurement noise and assumptions of PSF and galaxy properties (e.g. Bernstein & Armstrong 2014), or by determining the calibrations directly from observational data, without reference to external simulations.

For the long-standing issue of photo-z calibration, a concerted spectroscopic effort is now urgently needed to accurately characterize the mapping between galaxy colours and $n(z)$ (following e.g. Masters et al. 2015), as well as a thorough understanding of the limitations of current photo-z estimation schemes. Alternatively, cross-correlation photo-z methods as proposed by Newman (2008) may provide another calibration tool, but it remains to be seen whether they can achieve the ≈1 per cent level accuracy required to render this source of error subdominant. We thus anticipate that the mass calibration of galaxy clusters identified in DES Year-1 data will be limited by photo-z uncertainties. Assuming we can reduce our shear measurement uncertainties to the 1 per cent level (1.3 per cent on the mass), the next DES results should allow us to constrain the amplitude of the mass–richness relation to ≈5 per cent. Any improvements in photometric redshift uncertainties will further re-

duce this error, directly benefitting forthcoming cluster cosmology studies.

This work used simulations and computations performed using computational resources at SLAC and at NERSC.

REFERENCES

Allen S. W., Evrard A. E., Mantz A. B., 2011, ARA&A, 49, 409
Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012, MNRAS, 426, 2046
Applegate D. E. et al., 2014, MNRAS, 439, 48
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Baxter E. J., Rozo E., Jain B., Rykoff E., Wechsler R. H., 2016, MNRAS, 463, 205
Becker M. R., Kravtsov A. V., 2011, ApJ, 740, 25
Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013, ApJ, 762, 109
Benítez N., 2000, ApJ, 536, 571
Bernstein G. M., Armstrong R., 2014, MNRAS, 438, 1880
Bernstein G. M., Jarvis M., 2002, AJ, 123, 583
Bertin E., 2009, Mem. Soc. Astron. Ital., 80, 422
Bhattacharya S., Habib S., Heitmann K., Vikhlinin A., 2013, ApJ, 766, 32
Bleem L. E. et al., 2015, ApJS, 216, 27
Bocquet S. et al., 2015, ApJ, 799, 214
Bocquet S., Saro A., Dolag K., Mohr J. J., 2016, MNRAS, 456, 2361
Bonnett C. et al., 2016, Phys. Rev. D, 94, 042005
Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503
Carlstrom J. E. et al., 2011, PASP, 123, 568
Carrasco Kind M., Brunner R. J., 2013, MNRAS, 432, 1483
Coe D., Benítez N., Sánchez S. F., Jee M., Bouwens R., Ford H., 2006, AJ, 132, 926
Collister A. A., Lahav O., 2004, PASP, 116, 345
Dark Energy Survey Collaboration, 2005, preprint (arXiv:e-prints)
Dark Energy Survey Collaboration, 2016, MNRAS, 460, 1270
Diehl H. T. et al., 2014, in Peck A. B., Benn C. R., Seaman R. L., Proc. SPIE Conf. Ser., Vol. 9149, Observatory Operations: Strategies, Processes, and Systems V. SPIE, Bellingham, p. 91490V
Diemer B., Kravtsov A. V., 2014, ApJ, 789, 1
Diemer B., Kravtsov A. V., 2015, ApJ, 799, 108
Dietrich J. P. et al., 2014, MNRAS, 443, 1713
Dodelson S., Schneider M. D., 2013, Phys. Rev. D, 88, 063537
Efron B., 1982, The Jackknife, the Bootstrap and other Resampling Plans. CBMS-NSF Regional Conference Series in Applied Mathematics (SIAM).
Evrard A. E., Arnault P., Huterer D., Farahi A., 2014, MNRAS, 441, 3562
Farahi A., Evrard A. E., Rozo E., Rykoff E. S., Wechsler R. H., 2016, MNRAS, 460, 3900
Flaugher B. et al., 2015, AJ, 150, 150
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Friedrich O., Seitz S., Eifler T. F., Gruen D., 2016, MNRAS, 456, 2662
Graff P., Feroz F., Hobson M. P., Lasenby A., 2014, MNRAS, 441, 1741
Gruen D., Brimioulle F., 2017, MNRAS, 468, 769
Gruen D., Bernstein G. M., Lam T. Y., Seitz S., 2011, MNRAS, 416, 1392
Gruen D. et al., 2014, MNRAS, 442, 1507
Gruen D., Seitz S., Becker M. R., Friedrich O., Mana A., 2015, MNRAS, 449, 4264
Hao J., Kubo J. M., Feldmann R., Annis J., Johnston D. E., Lin H., McKay T. A., 2011, ApJ, 740, 39
Hartlap J., Simon P., Schneider P., 2007, A&A, 464, 399
Hasselfield M. et al., 2013, J. Cosmology Astropart. Phys., 7, 8
Hayashi E., White S. D. M., 2008, MNRAS, 388, 2
Hoekstra H., 2001, A&A, 370, 743
Hoekstra H., 2003, MNRAS, 339, 1155
Hoekstra H., Hartlap J., Hilbert S., van Uitert E., 2011, MNRAS, 412, 2095
Hoekstra H., Mahdavi A., Babul A., Bildfell C., 2012, MNRAS, 427, 1298
Hoekstra H., Herbonnet R., Muzzin A., Babul A., Mahdavi A., Viola M., Cacciato M., 2015, MNRAS, 449, 685
Hu W., Kravtsov A. V., 2003, ApJ, 584, 702

Ilbert O. et al., 2009, ApJ, 690, 1236
Ilbert O. et al., 2013, A&A, 556, A55
Jarvis M. et al., 2016, MNRAS, 460, 2245
Johnston D. E., Sheldon E. S., Tasitsiomi A., Frieman J. A., Wechsler R. H., McKay T. A., 2007a, ApJ, 656, 27
Johnston D. E. et al., 2007b, preprint (arXiv: e-prints)
Kaufman G. M., 1967, Center for Operations Research and Econometrics Discussion Paper, 6710
Kelly P. L. et al., 2014, MNRAS, 439, 28
Koester B. P. et al., 2007, ApJ, 660, 239
Landy S. D., Szalay A. S., 1993, ApJ, 412, 64
Li R. et al., 2016, MNRAS, 458, 2573
Lilly S. J. et al., 2007, ApJS, 172, 70
Mandelbaum R., Hirata C. M., Leauthaud A., Massey R. J., Rhodes J., 2012, MNRAS, 420, 1518
Mantz A. B. et al., 2015, MNRAS, 446, 2205
Massey R. et al., 2007, MNRAS, 376, 13
Masters D. et al., 2015, ApJ, 813, 53
Mehrtens N. et al., 2012, MNRAS, 423, 1024
Melchior P. et al., 2015, MNRAS, 449, 2219
Metzler C. A., White M., Loken C., 2001, ApJ, 547, 560
Miller L., Kitching T. D., Heymans C., Heavens A. F., van Waerbeke L., 2007, MNRAS, 382, 315
Miyatake H., More S., Takada M., Spergel D. N., Mandelbaum R., Rykoff E. S., Rozo E., 2016, Phys. Rev. Lett., 116, 041301
Nakajima R., Mandelbaum R., Seljak U., Cohn J. D., Reyes R., Cool R., 2012, MNRAS, 420, 3240
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563 (NFW)
Newman J. A., 2008, ApJ, 684, 88
Noh Y., Cohn J. D., 2012, MNRAS, 426, 1829
Oguri M., Bayliss M. B., Dahle H., Sharon K., Gladders M. D., Natarajan P., Hennawi J. F., Koester B. P., 2012, MNRAS, 420, 3213
Okabe N., Smith G. P., 2016, MNRAS, 461, 3794
Piffaretti R., Arnaud M., Pratt G. W., Pointecouteau E., Melin J.-B., 2011, A&A, 534, A109
Planck Collaboration XXIV, 2016, A&A, 594, A24
Planck Collaboration XXVII, 2016, A&A, 594, A27
Postman M. et al., 2012, ApJS, 199, 25
Rowe B. T. P. et al., 2015, Astron. Comput., 10, 121
Rozo E., Rykoff E. S., 2014, ApJ, 783, 80
Rozo E. et al., 2009, ApJ, 703, 601
Rozo E. et al., 2010, ApJ, 708, 645
Rozo E., Rykoff E. S., Becker M., Reddick R. M., Wechsler R. H., 2015, MNRAS, 453, 38
Rykoff E. S. et al., 2012, ApJ, 746, 178
Rykoff E. S. et al., 2014, ApJ, 785, 104
Rykoff E. S. et al., 2016, ApJS, 224, 1
Sadeh I., Abdalla F. B., Lahav O., 2016, PASP, 128, 104502
Sánchez C. et al., 2014, MNRAS, 445, 1482
Saro A. et al., 2015, MNRAS, 454, 2305 (S15)
Schmidt F., Lima M., Oyaizu H., Hu W., 2009, Phys. Rev. D, 79, 083518
Seitz C., Schneider P., 1997, A&A, 318, 687
Sheldon E. S., 2014, MNRAS, 444, L25
Sheldon E. S. et al., 2004, AJ, 127, 2544
Sifón C., Hoekstra H., Cacciato M., Viola M., Köhlinger F., van der Burg R. F. J., Sand D. J., Graham M. L., 2015, A&A, 575, A48
Simet M., McClintock T., Mandelbaum R., Rozo E., Rykoff E., Sheldon E., Wechsler R. H., 2017a, MNRAS, 466, 3103 (S16)
Simet M., Battaglia N., Mandelbaum R., Seljak U., 2017b, MNRAS, 466, 3663
Springel V., 2005, MNRAS, 364, 1105
Stanek R., Rasia E., Evrard A. E., Pearce F., Gazzola L., 2010, ApJ, 715, 1508
Suchyta E. et al., 2016, MNRAS, 457, 786
Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, ApJ, 688, 709
Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, ApJ, 724, 878

Umetsu K. et al., 2011, ApJ, 795, 163

Umetsu K. et al., 2014, ApJ, 795, 163

Umetsu K., Zitrin A., Gruen D., Merten J., Donahue M., Postman M., 2016, ApJ, 821, 116

van Uitert E., Gilbank D. G., Hoekstra H., Semboloni E., Gladders M. D., Yee H. K. C., 2016, A&A, 586, A43

von der Linden A. et al., 2014a, MNRAS, 439, 2

von der Linden A. et al., 2014b, MNRAS, 443, 1973

Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, Phys. Rep., 530, 87

White M., Cohn J. D., Smit R., 2010, MNRAS, 408, 1818

Yang X., Mo H. J., van den Bosch F. C., Jing Y. P., Weinmann S. M., Meneghetti M., 2006, MNRAS, 373, 1159

Zu Y., Weinberg D. H., Rozo E., Sheldon E. S., Tinker J. L., Becker M. R., 2014, MNRAS, 439, 1628

Zuntz J., Kacprzak T., Voigt L., Hirsch M., Rowe B., Bridle S., 2013, MNRAS, 434, 1604

[1] *Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA*

[2] *Kavli Institute for Particle Astrophysics and Cosmology, PO Box 2450, Stanford University, Stanford, CA 94305, USA*

[3] *SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA*

[4] *Department of Physics, University of Arizona, Tucson, AZ 85721, USA*

[5] *Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, D-85748 Garching, Germany*

[6] *Fakultät für Physik, Universitäts-Sternwarte, Ludwig-Maximilians Universität München, Scheinerstr. 1, D-81679 München, Germany*

[7] *Brookhaven National Laboratory, Bldg 510, Upton, NY 11973, USA*

[8] *Department of Physics, ETH Zurich, Wolfgang-Pauli-Strasse 16, CH-8093 Zurich, Switzerland*

[9] *Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA*

[10] *Fermi National Accelerator Laboratory, PO Box 500, Batavia, IL 60510, USA*

[11] *Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA*

[12] *Department of Physics and Astronomy, Pevensey Building, University of Sussex, Brighton BN1 9QH, UK*

[13] *Jodrell Bank Center for Astrophysics, School of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK*

[14] *Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*

[15] *Excellence Cluster Universe, Boltzmannstr. 2, D-85748 Garching, Germany*

[16] *Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, D-81679 Munich, Germany*

[17] *Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK*

[18] *Department of Physics, University of California, Santa Cruz, CA 95064, USA*

[19] *Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA*

[20] *Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA*

[21] *Department of Physics, The Ohio State University, Columbus, OH 43210, USA*

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

[23] *Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109, USA*

[24] *Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, Casilla 603, La Serena, Chile*

[25] *Department of Physics and Electronics, Rhodes University, PO Box 94, Grahamstown 6140, South Africa*

[26] *CNRS, UMR 7095, Institut d'Astrophysique de Paris, F-75014 Paris, France*

[27] *Institut d'Astrophysique de Paris, Sorbonne Universités, UPMC Univ. Paris 06, UMR 7095, F-75014 Paris, France*

[28] *Laboratório Interinstitucional de e-Astronomia – LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ 20921-400, Brazil*

[29] *Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ 20921-400, Brazil*

[30] *Department of Astronomy, University of Illinois, 1002 W. Green Street, Urbana, IL 61801, USA*

[31] *National Center for Supercomputing Applications, 1205 West Clark St., Urbana, IL 61801, USA*

[32] *Institut de Ciències de l'Espai, IEEC-CSIC, Campus UAB, Carrer de Can Magrans, s/n, E-08193 Bellaterra, Barcelona, Spain*

[33] *Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK*

[34] *School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK*

[35] *Department of Physics, IIT Hyderabad, Kandi, Telangana 502285, India*

[36] *Instituto de Física Teórica IFT-UAM/CSIC, Universidad Autónoma de Madrid, Cantoblanco, E-28049 Madrid, Spain*

[37] *Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA*

[38] *Astronomy Department, University of Washington, Box 351580, Seattle, WA 98195, USA*

[39] *Australian Astronomical Observatory, North Ryde, NSW 2113, Australia*

[40] *Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, CP 66318, CEP 05314-970, São Paulo, SP, Brazil*

[41] *Department of Astronomy, The Ohio State University, Columbus, OH 43210, USA*

[42] *Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA*

[43] *Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain*

[44] *Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain*

[45] *Universidade Federal do ABC, Centro de Ciências Naturais e Humanas, Av. dos Estados, 5001, Santo André, SP 09210-580, Brazil*