Sunyaev–Zel'dovich effect and X-ray scaling relations from weak lensing mass calibration of 32 South Pole Telescope selected galaxy clusters

J. P. Dietrich,[1,2]★ S. Bocquet,[3,4] T. Schrabback,[5] D. Applegate,[4] H. Hoekstra,[6] S. Grandis,[1,2] J. J. Mohr,[1,2,7] S. W. Allen,[8,9,10] M. B. Bayliss,[11] B. A. Benson,[12,13,4] L. E. Bleem,[3,4] M. Brodwin,[14] E. Bulbul,[11] R. Capasso,[1,2] I. Chiu,[1,2] T. M. Crawford,[4,13] A. H. Gonzalez,[15] T. de Haan,[16] M. Klein,[1,2] A. von der Linden,[17] A. B. Mantz,[10,8] D. P. Marrone,[18] M. McDonald,[11] S. Raghunathan,[19] D. Rapetti,[20,21] C. L. Reichardt,[19] A. Saro,[1,2,22] B. Stalder,[23] A. Stark,[24] C. Stern[1,2] and C. Stubbs[25]

Affiliations are listed at the end of the paper

ABSTRACT

Uncertainty in mass–observable scaling relations is currently the limiting factor for galaxy-cluster-based cosmology. Weak gravitational lensing can provide direct mass calibration and reduce the mass uncertainty. We present new ground-based weak lensing observations of 19 South Pole Telescope (SPT) selected clusters at redshifts $0.29 \leq z \leq 0.61$ and combine them with previously reported space-based observations of 13 galaxy clusters at redshifts $0.576 \leq z \leq 1.132$ to constrain the cluster mass scaling relations with the Sunyaev–Zel'dovich effect (SZE), the cluster gas mass $M_{\rm gas}$ and $Y_{\rm X}$, the product of $M_{\rm gas}$ and X-ray temperature. We extend a previously used framework for the analysis of scaling relations and cosmological constraints obtained from SPT-selected clusters to make use of weak lensing information. We introduce a new approach to estimate the effective average redshift distribution of background galaxies and quantify a number of systematic errors affecting the weak lensing modelling. These errors include a calibration of the bias incurred by fitting a Navarro–Frenk–White profile to the reduced shear using N-body simulations. We blind the analysis to avoid confirmation bias. We are able to limit the systematic uncertainties to 5.6% in cluster mass (68% confidence). Our constraints on the mass–X-ray observable scaling relation parameters are consistent with those obtained by earlier studies and our constraints for the mass–SZE scaling relation are consistent with the simulation-based prior used in the most recent SPT–SZ cosmology analysis. We can now replace the *external* mass calibration priors used in previous SPT–SZ cosmology studies with a direct, *internal* calibration obtained for the same clusters.

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

1 INTRODUCTION

The cluster mass function, i.e. the abundance of clusters of galaxies as a function of redshift and mass, is a sensitive cosmological probe (see Allen, Evrard & Mantz 2011 for a review). Because the cluster mass function is sensitive to both the expansion history and the history of structure formation in the Universe, cluster cosmology is in principle able to break degeneracies between cosmological parameters arising in purely geometric probes such as the primary cosmic microwave background (CMB), baryonic acoustic oscillations and Type Ia supernovae. Observable properties of galaxy clusters like X-ray luminosity and temperature, optical richness and the strength of the Sunyaev–Zel'dovich effect (SZE: Sunyaev & Zel'dovich 1970, 1972) have been shown to scale with galaxy cluster mass following *mass–observable scaling relations* (MOR). These scaling relations have intrinsic scatter around the mean relationship between the observable, which is used as a proxy for cluster mass, and the cluster mass, which has been used to parametrize the theoretical cluster mass function. Cosmological constraints from cluster mass function studies are currently limited by uncertainties in the mass–scaling relation parameters.

★ E-mail: dietrich@usm.lmu.de

Weak gravitational lensing offers the best opportunity to determine the normalization of the MOR, as it can estimate projected cluster masses with near-zero bias on average (Corless & King 2009; Becker & Kravtsov 2011; Bahé, McCarthy & King 2012). The scatter between lensing-inferred cluster masses and true halo mass, however, is large and typically exceeds the intrinsic scatter of the mass–observable relations employed for cosmological purposes. Sources of this scatter include the shape noise of lensed background galaxies, correlated and uncorrelated large-scale structure (LSS: Hoekstra 2001; Dodelson 2004; Becker & Kravtsov 2011) along the line of sight and halo triaxiality (Clowe, De Lucia & King 2004; Corless & King 2007; Meneghetti et al. 2010), the latter being the dominant source of scatter for massive galaxy clusters. Therefore, large numbers of clusters are required to achieve a competitive calibration of the normalization of mass–observable scaling relations. Several programmes making use of gravitational lensing to this end have published results (e.g. Bardeau et al. 2007; Okabe et al. 2010; Hoekstra et al. 2012, 2015; Marrone et al. 2012; Applegate et al. 2014, 2016; Gruen et al. 2014; Umetsu et al. 2014; Battaglia et al. 2016; Okabe & Smith 2016; Hilton et al. 2018), or are under way, employing data from current wide-field imaging surveys such as the Dark Energy Survey (Melchior et al. 2017) or the HyperSuprimeCam survey (Murata et al. 2017). Future surveys and missions such as the Large Synoptic Survey Telescope[1] (LSST Dark Energy Science Collaboration 2012), *Euclid*[2] (Laureijs et al. 2011) or CMB-S4[3] (Abazajian et al. 2016) will lead to much tighter constraints, while at the same time imposing much stricter requirements for the control of systematic errors.

Here we describe the weak lensing analysis of 19 intermediate-redshift clusters selected from the 2500 deg^2 SPT–SZ survey (Bleem et al. 2015), five of which have already been presented in an earlier weak lensing study (High et al. 2012). After discussing these data in Section 2, we present our weak lensing methods in Sections 3 and 4, paying particular attention to controlling systematic effects. In Section 5, we then combine our 19 clusters with 13 high-redshift clusters from the SPT–SZ survey, existing weak lensing data from the *Hubble Space Telescope* (*HST*: Schrabback et al. 2018, hereinafter S18) and X-ray data from the *Chandra* X-ray satellite for 89 clusters to perform a joint mass–observable scaling relations analysis using a newly developed framework that accounts self-consistently for selection effects and biases.

For quantities evaluated for a fixed cosmology, we assume a flat ΛCDM cosmology with $\Omega_\mathrm{m} = 0.3$, $\Omega_\Lambda = 0.7$, $H_0 = 70\,h_{70}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $h_{70} = 1$ throughout this article. When reporting cluster masses, denoted as M_Δ, we follow the convention of defining masses in terms of spherical overdensities that are a factor Δ above the critical density $\rho_\mathrm{c}(z)$ of the Universe at redshift z. Likewise, r_Δ corresponds to the radius of the sphere containing the mass $M_\Delta = (4/3)\pi r_\Delta^3 \rho_\mathrm{c}(z)$. We use standard notation for statistical distributions, i.e. the normal distribution with mean μ and covariance matrix Σ is written as $\mathcal{N}(\mu, \Sigma)$ and $\mathcal{U}(a, b)$ denotes the uniform distribution on the interval $[a, b]$.

[1] https://www.lsst.org/
[2] https://www.euclid-ec.org/
[3] https://www.cmb-s4.org/

2 DATA

2.1 Cluster sample

The South Pole Telescope (SPT: Carlstrom et al. 2011) is a 10-m telescope located at the Amundsen–Scott South Pole Station. From 2007–2011, SPT observed a contiguous 2500 deg^2 region in three bands (95, 150 and 220 GHz) to a fiducial depth of 18 μK arcmin in the 150-GHz band. Details of the survey strategy and data processing are provided elsewhere (Staniszewski et al. 2009; Vanderlinde et al. 2010; Williamson et al. 2011). Galaxy clusters in the survey were detected via their thermal SZE. The full cluster catalogue of the SPT–SZ survey was published in Bleem et al. (2015). In the SPT–SZ survey, 677 galaxy clusters were detected above signal-to-noise ratio $\xi > 4.5$ and 516 were confirmed by optical and near-infrared imaging (Bleem et al. 2015). Of these, 415 were first identified by SPT and 109 have been confirmed spectroscopically (Ruel et al. 2014; Bayliss et al. 2016). The median mass of this sample is $M_{500} \approx 3 \times 10^{14}\,\mathrm{M_\odot}$, with a median redshift of 0.55 and a maximum above 1.4 (Bleem et al. 2015). The selection function of the survey is well understood and almost flat in mass at $z > 0.25$, with a slightly higher sensitivity to lower mass systems at higher redshifts.

Cosmological constraints have been presented in de Haan et al. (2016) based on the 'cosmology subset' of the entire SPT–SZ cluster sample with redshift $z > 0.25$ and detection significance $\xi > 5$. This $\xi > 5$ threshold corresponds to a sample with 95 per cent purity from SZE selection alone. The mass calibration employed in that analysis adopted information from the cluster mass function, together with information from X-ray observable $Y_\mathrm{X} = M_\mathrm{gas} T_\mathrm{X}$ available for 82 systems. The Y_X–mass relation calibration was informed from earlier weak lensing analyses of different cluster samples (Vikhlinin et al. 2009; Hoekstra et al. 2015; Applegate et al. 2014). We limit the analysis in this article to this cosmology subset.

We obtained pointed follow-up observations of 19 clusters in the redshift range $0.29 \leq z \leq 0.61$ with the Megacam imager (McLeod et al. 2006) at the Magellan Clay telescope. In the following, we first describe these data and their analysis, before combining them with space-based *HST* weak lensing follow-up data of 13 SPT–SZ clusters in the redshift range $0.576 \leq z \leq 1.132$ (Schrabback et al. 2018). Table 1 provides an overview of our cluster sample.

2.2 Optical data

Our sample of 19 SPT clusters was observed with Megacam at the 6.5-m Magellan Clay telescope. This sample includes five galaxy cluster observations previously presented by High et al. (2012). This previous work also describes the observing strategy, data reduction and photometric and astrometric calibration in detail. We briefly summarize the observing strategy for the remaining 14 clusters. These were observed in 2011 November through g', r' and i' filters, for total exposure times of 1200, 1800 and 2400 s, respectively. In g' and r' bands, a three-point diagonal dither pattern, which covers the chip gaps, was used, while a five-point linear dither pattern was utilized for the i'-band exposures. As an exception from this strategy, SPT-CL J0240–5946 was observed in four r'-band exposures.

Care was taken to observe the r'-band images, which are used to generate the shear catalogues, in the most stable and best seeing conditions. Seeing values for all r'-band images are given in Table 2.The median seeing of our exposures is 0.79 arcsec, while

Table 1. SPT clusters with lensing data used in this article. Clusters observed with *HST* are imported from Schrabback et al. (2018) and are used in the scaling relation analysis (see Sections 5 and 6) only in combination with the Megacam cluster sample.

Cluster	α (J2000.0)	δ (J2000.0)	z	ξ	Telescope	*Chandra* data
SPT-CL J0000–5748	00:00:59.98	−57:48:23.0	0.702	8.49	*HST*	✓
SPT-CL J0102–4915	01:02:55.06	−49:15:39.6	0.870	39.91	*HST*	✓
SPT-CL J0234–5831	02:34:42.87	−58:31:17.1	0.415	14.66	Megacam	✓
SPT-CL J0240–5946	02:40:38.54	−59:46:10.9	0.400	8.84	Megacam	
SPT-CL J0254–5857	02:54:17.50	−58:57:09.3	0.438	14.13	Megacam	
SPT-CL J0307–6225	03:07:20.08	−62:25:57.8	0.579	8.46	Megacam	✓
SPT-CL J0317–5935	03:17:17.18	−59:35:06.5	0.469	6.26	Megacam	
SPT-CL J0346–5439	03:46:53.93	−54:39:01.9	0.530	9.25	Megacam	✓
SPT-CL J0348–4515	03:48:17.70	−45:15:03.5	0.358	10.12	Megacam	✓
SPT-CL J0426–5455	04:26:04.78	−54:55:10.8	0.642	8.85	Megacam	✓
SPT-CL J0509–5342	05:09:20.97	−53:42:19.2	0.461	8.50	Megacam	✓
SPT-CL J0516–5430	05:16:36.31	−54:30:39.0	0.295	12.41	Megacam	✓
SPT-CL J0533–5005	05:33:36.22	−50:05:24.4	0.881	7.08	*HST*	✓
SPT-CL J0546–5345	05:46:36.60	−53:45:45.0	1.066	10.76	*HST*	✓
SPT-CL J0551–5709	05:51:36.99	−57:09:20.4	0.423	8.21	Megacam	✓[1]
SPT-CL J0559–5249	05:59:42.02	−52:49:33.6	0.609	10.64	*HST*	✓
SPT-CL J0615–5746	06:15:51.60	−57:46:34.7	0.972	26.42	*HST*	✓
SPT-CL J2022–6323	20:22:06.25	−63:23:56.1	0.383	6.51	Megacam	
SPT-CL J2030–5638	20:30:48.87	−56:38:10.2	0.394	5.50	Megacam	
SPT-CL J2032–5627	20:32:19.37	−56:27:28.9	0.284	8.61	Megacam	
SPT-CL J2040–5725	20:40:13.75	−57:25:46.2	0.930	6.24	*HST*	✓
SPT-CL J2106–5844	21:06:04.94	−58:44:42.4	1.132	22.22	*HST*	✓
SPT-CL J2135–5726	21:35:39.92	−57:24:32.7	0.427	10.51	Megacam	✓
SPT-CL J2138–6008	21:38:01.26	−60:08:00.0	0.319	12.64	Megacam	
SPT-CL J2145–5644	21:45:52.37	−56:44:51.2	0.480	12.60	Megacam	✓
SPT-CL J2331–5051	23:31:50.59	−50:51:50.0	0.576	10.47	*HST*	✓
SPT-CL J2332–5358	23:32:25.37	−53:58:03.1	0.402	9.12	Megacam	
SPT-CL J2337–5942	23:37:24.55	−59:42:17.6	0.775	20.35	*HST*	✓
SPT-CL J2341–5119	23:41:11.78	−51:19:41.2	1.003	12.49	*HST*	✓
SPT-CL J2342–5411	23:42:45:41	−54:11:08.2	1.075	8.18	*HST*	✓
SPT-CL J2355–5055	23:55:47.95	−50:55:19.1	0.320	6.60	Megacam	✓
SPT-CL J2359–5009	23:59:41.52	−50:09:53.6	0.775	6.68	*HST*	✓

Note. [1] *Chandra* data excluded from the analysis. See Section 2.3.

the minimum and maximum values are 0.54 and 1.11 arcsec, respectively. The clusters observed with Megacam were generally the most significant SPT cluster detections that were known and visible at the time of the observing runs. An attempt was made to observe higher redshift clusters during better seeing conditions.

As in High et al. (2012), a stellar locus regression code (High et al. 2009) and cross-matching with the 2MASS catalogue (Skrutskie et al. 2006) are employed in the photometric calibration of our data. The resulting uncertainties in the absolute photometric calibration and colour measurements are ∼0.05 mag and 0.03 mag, respectively.

2.3 X-ray data

The X-ray data in this work consist of 89 galaxy clusters observed with the *Chandra* satellite and are mostly identical to the data described in de Haan et al. (2016). The reduction and analysis of these data is described in detail in McDonald et al. (2013). Changes in the data since this earlier SPT publication include the addition of eight new clusters at $z > 1$ (McDonald et al. 2017), none of which currently has weak lensing data, and the omission of SPT-CL J0551–5709. The latter cluster is part of our Megacam sample. However, after the observations were obtained it was realized that this cluster is indeed a projection of two clusters at different redshifts (Andersson et al. 2011), the SPT-selected cluster at $z = 0.423$ and the local cluster Abell S0552 with a redshift of $z = 0.09$ inferred from the cluster red sequence (High et al. 2012). We thus exclude this cluster from the X-ray analysis but not the weak lensing analysis, where the inclusion of such projections is correctly accounted for (see Sections 4.4 and 5.2.2).

Fig. 1 gives an overview of the different subsamples in this study and their (partial) overlap. All 13 clusters with *HST* weak lensing data (S18) have X-ray data, while this is the case for only 10 out of the 19 clusters observed with Megacam after the exclusion of SPT-CL J0551–5709. See also Table 1, where all clusters with lensing data are listed.

3 WEAK LENSING ANALYSIS

Weak gravitational lensing by massive foreground structures such as galaxy clusters (see Hoekstra et al. 2013 for a review of cluster lensing studies) changes the observed ellipticities of background galaxies and imprints a coherent shear pattern around the cluster centre. The azimuthally averaged tangential shear at a distance r from the cluster centre,

$$\gamma_t(r) = \frac{\langle\Sigma(<r)\rangle - \Sigma(r)}{\Sigma_{\rm crit}} \,, \tag{1}$$

depends on the mean surface mass density $\langle\Sigma(<r)\rangle$ inside and the surface mass density at this radius. This differential surface mass

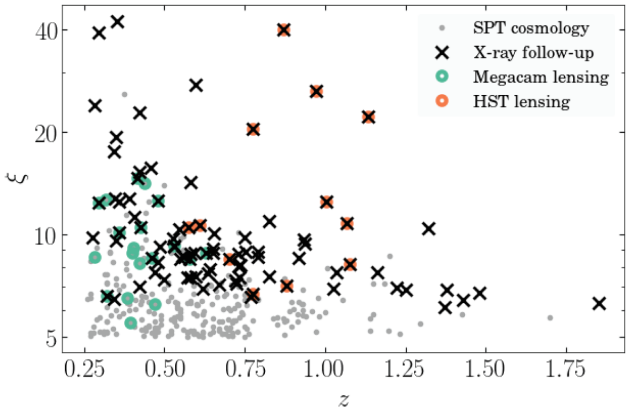


Figure 1. Overview of the SPT cosmology cluster sample, its coverage by X-ray data employed in this study and the two weak lensing subsamples used in the scaling relation analysis in Section 5. The axes show detection significance ξ plotted against cluster redshift z.

density profile is scaled by the critical surface mass density,

$$\Sigma_{\rm crit} = \frac{c^2}{4\pi G}\frac{1}{\beta D_{\rm l}}\,, \tag{2}$$

where c is the speed of light, G is the gravitational constant, $\beta = D_{\rm ls}/D_{\rm s}$ is the lensing efficiency and the D_i are angular diameter distances, where 'l' denotes the lens and 's' the source galaxy.

The observable quantity is not the shear but the reduced shear,

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

where $\kappa = \Sigma/\Sigma_{\rm crit}$ is the dimensionless surface mass density. A galaxy of intrinsic complex ellipticity $\varepsilon^{(s)}$ is sheared by the reduced gravitational shear g to have an observed (image) ellipticity (Seitz & Schneider 1997)

$$\varepsilon = \frac{\varepsilon^{(s)} + g}{1 + g^{*}\varepsilon^{(s)}} \approx \varepsilon^{(s)} + g, \qquad g \ll 1\,, \tag{4}$$

so that, because $\langle \varepsilon^s \rangle = 0$, the expectation value of ε is g.

We average the reduced shear over an ensemble of galaxies at different redshifts. Strictly speaking, redshifts for all background galaxies would be required for the correct weighting with the geometric lensing efficiency β. In the absence of such information, however, the average reduced shear can be corrected to first order using (Seitz & Schneider 1997)

$$\frac{\langle g_{\rm cor}\rangle}{\langle g_{\rm true}\rangle} = 1 + \left(\frac{\langle\beta^2\rangle}{\langle\beta\rangle^2} - 1\right)\kappa\,. \tag{5}$$

The averages $\langle\beta\rangle$ and $\langle\beta^2\rangle$ of the distribution of lensing efficiencies can be computed from the redshift distribution of lensed galaxies (see Section 3.2).

3.1 Shear catalogue creation

Our shear analysis is based on the pipeline developed for the Canadian Cluster Comparison Project (CCCP: Hoekstra et al. 2012). In this section we briefly review the main steps, but we refer the interested reader to the more detailed discussion in Hoekstra (2007) and in particular the updates discussed in Hoekstra et al. (2015), which used image simulations to calibrate the bias in the algorithm to an accuracy of 1–2 per cent.

The observed galaxy shapes are biased, because of smearing by the point-spread function (PSF): the seeing makes the galaxies appear rounder, whereas PSF anisotropy will lead to coherent alignments in the observed shapes. Noise in the images leads to additional biases (e.g. Viola, Kitching & Joachimi 2014). To obtain accurate cluster masses, it is essential that the shape measurement algorithm is able to correct for these sources of bias.

The shape measurement algorithm we use is based on the one introduced by Kaiser, Squires & Broadhurst (1995) and Luppino & Kaiser (1997), with modifications described in Hoekstra et al. (1998) and Hoekstra, Franx & Kuijken (2000). It uses the observed moments of the surface brightness distribution to correct for the PSF. However, as shown in Hoekstra et al. (2015), the measurements are still biased, predominantly because of noise. These biases can be calibrated using image simulations. Because our data cover a similar range in signal-to-noise ratio and seeing, we adopted the correction parameters found by Hoekstra et al. (2015).

Similarly to Hoekstra et al. (2012), we analyse each of the Megacam exposures and combine the measurements in the catalogue stage, to avoid the complex PSF pattern that would otherwise arise. We use SExtractor (Bertin & Arnouts 1996) to detect objects in the images and select objects with no flags raised. We use the observed half-light radius to define the width of the Gaussian weight function to measure the quadrupole (and higher) moments of the surface brightness distribution of an object.

The next step is to find an adequate model to describe the spatial variation of the PSF (both size and shape) as a function of the width of the weight function used to analyse the galaxies (see Hoekstra et al. 1998 for details). To quantify the properties of the PSF, we select a sample of bright, but unsaturated, stars based on their half-light radius and shape. The number of available stars varies from field to field and chip to chip, with a median of 16 stars per chip and 519 stars per field. As shown in Fig. 2, the PSF is anisotropic and in many cases shows a coherent tangential pattern around the central parts of the field of view. Such a pattern mimics the expected cluster lensing signal (although that should decline with radius, rather than increase as is the case for the PSF anisotropy). Therefore we have to take special care to model the PSF (also see High et al. 2012).

To capture the dominant PSF pattern, we fit a tangential pattern around the centre of the focal plane, with a radial dependence that is a polynomial in radius r up to order 4, where the order was chosen based on a visual inspection of the residuals. We also fit for the slope as a function of x and y. This model is fitted to the full field of view. Inspection of the residuals showed coherent variations on more or less the chip scale. We therefore also fit a first-order polynomial chip-by-chip in x and y to the residuals. The resulting model is used to correct for the PSF. Tests of the performance of the PSF model, described in more detail in Section 3.1.1, show that the model is adequate. We select galaxies to be objects with sizes larger than the PSF. Specifically, we require that the half-light radius $r_{\rm h}$ exceeds the half-light radius of the largest star selected from the stellar locus in a magnitude–$r_{\rm h}$ plot by a factor of 1.05. Following Hoekstra et al. (2015), we also remove galaxies with half-light radii larger than 1 arcsec, because many of these are blended objects, biasing the shape measurements.

This procedure is carried out for each exposure and bad regions are masked. The resulting catalogues (typically three per cluster) are then combined, with the shape measurements for objects that appear more than once averaged accordingly. The averaging takes into account the measurement uncertainties, thus naturally giving more weight to the better seeing data. This results in a single catalogue of galaxy shapes that is used to determine the cluster masses.

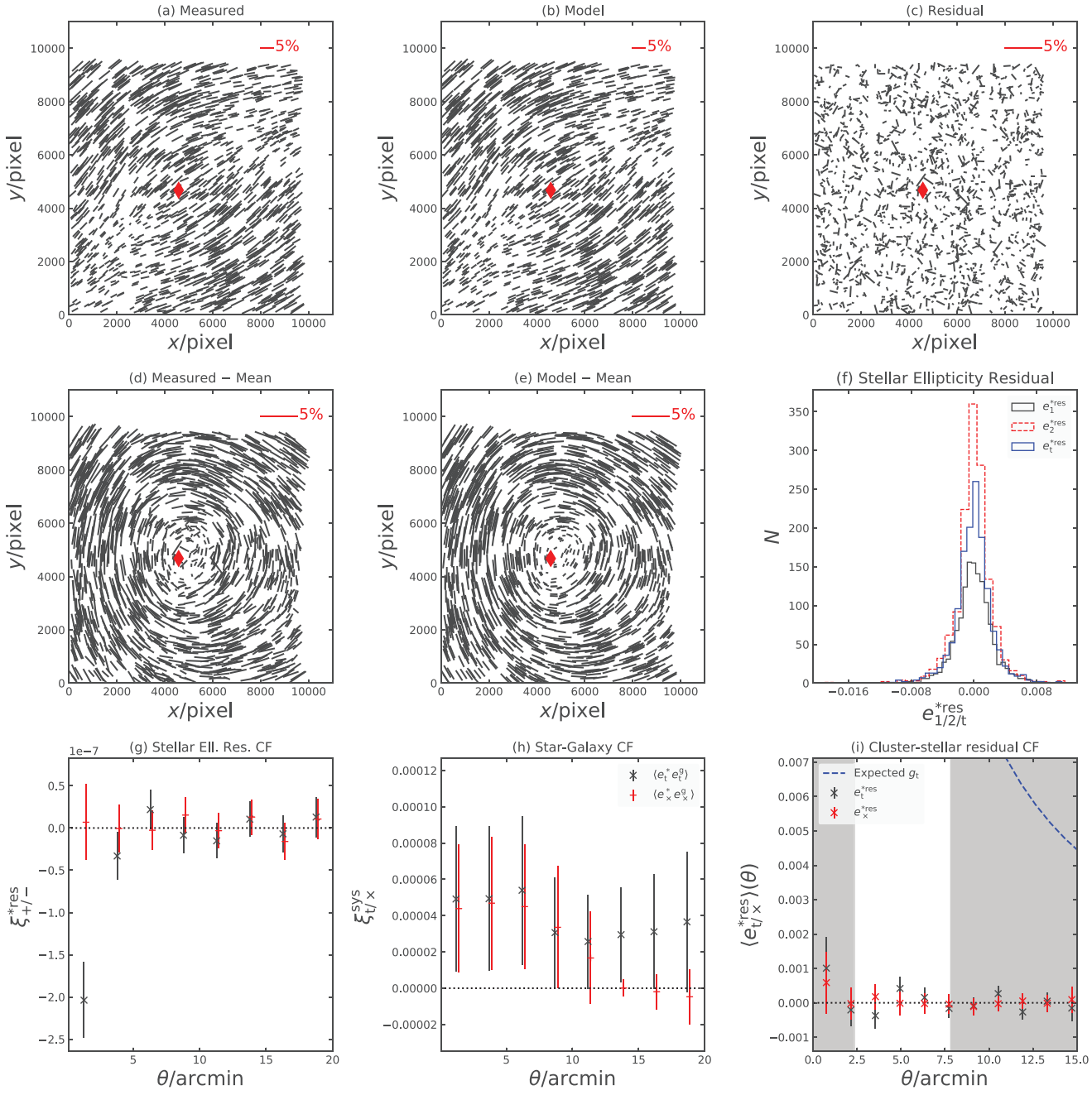


Figure 2. PSF correction diagnostic plots for the second exposure of SPT-CL J2030–5638. Red diamonds in panels (a)–(e) indicate the SZE-derived cluster centre. (a) Measured stellar ellipticity pattern; (b) model of the PSF pattern in panel (a); (c) residual between panels (a) and (b); (d) same as panel (a), with the mean ellipticity subtracted; (e) same as panel (b) with the mean ellipticity subtracted; (f) histogram of the stellar ellipticity residuals from panel (c) in the two Cartesian ellipticity components $e_{1/2}$ and the tangential ellipticity around the cluster centre e_t; (g) ellipticity correlation functions $\xi_\pm$ of the stellar ellipticity residuals; (h) ellipticity correlation functions $\xi^{sys}_{t/\times}$ between measured stellar and corrected galaxy ellipticity; (i) ellipticity–position correlation function between stellar residual tangential and cross-component ellipticity, and the cluster centre. The dashed line shows a comparison with the expected tangential shear signal based on the SZE mass estimate of the cluster. The grey shaded regions are radii that are omitted in the NFW fitting procedure.

3.1.1 Shear catalogue systematic tests

We tested the PSF correction of the shear catalogues for a range of systematic residuals to ascertain that these have negligible influence on our cluster mass estimates. These are illustrated for the extreme case of the second exposure of the cluster SPT-CL J2032–5627 in Fig. 2. This exposure shows a strong tangential alignment of the stellar ellipticity pattern almost perfectly centred on the cluster location. Left uncorrected, this PSF would lead to a spurious cluster lensing signal and thus provides a good illustration of the quality of our PSF correction. A randomly chosen, more representative example of the Megacam PSF pattern and our diagnostic plots is shown in Fig. A1 in Appendix A. Because a constant shear will average out in an azimuthal average around the cluster position, we show the PSF without its mean value across the field of view (FOV)

in panel (d) and the corresponding PSF model components in panel (e).

As a first diagnostic, we examined the distribution of the residuals of stellar ellipticity between the observed stellar ellipticity and the smooth model describing the spatial variation of the PSF across the focal plane. We verified that the means of both Cartesian ellipticity components, as well as the tangential ellipticity residuals with respect to the cluster centre, are statistically compatible with zero. A histogram showing the binned distribution of these three ellipticity components is shown in Fig. 2(f). All exposures of all fields pass this basic test.

Next we computed the correlation functions,

$$\xi_{\pm} = \langle e_{\rm t} e_{\rm t} \rangle \pm \langle e_{\times} e_{\times} \rangle, \tag{6}$$

for the residuals of stellar ellipticity, where the tangential and cross-components are defined with respect to the line connecting the stars. We used the tree code ATHENA[4] to compute this and all other correlation functions. If the PSF model represents the spatial variations of the actual PSF faithfully, no correlation should remain (Hoekstra 2004). We find that this is generally the case for almost all fields, except for the smallest bin of ξ_+, which is often slightly negative, as seen in Fig. 2(g). This indicates some overfitting of the PSF on these scales, but the values of these bins are 2–3 orders of magnitude below those of the cluster-induced gravitational shears on the angular scales of interest (cf. Fig. 2i). Moreover, this overfitting happens for individual exposures and may not be coherent across all three exposures, in which case it should approximately average out, so that its real impact is decreased even further.

A common diagnostic in cosmic shear analyses for the absence of leakage from PSF ellipticity to the shear catalogue is the star–galaxy correlation function (Bacon et al. 2003):

$$\xi^{\rm sys}_{\rm t/\times} = \frac{\langle e^*_i \gamma_i \rangle^2}{\langle e^*_i e^*_i \rangle}\,, \tag{7}$$

which can be computed for the tangential ($i = {\rm t}$) and cross-components ($i = \times$) of the uncorrected stellar ellipticities $e*$ and the observed shears of the galaxies γ. For random fields, there should not be any correlation between the stellar ellipticity and the measured shear. However, observations centred on galaxy clusters are not random fields. The cluster centre is a special location around which we expect a tangential alignment of galaxies. The absence of a significant star–galaxy correlation thus indicates that no PSF leakage into the shear catalogue occurred; its presence, however, would not be a cause for concern. Taking the covariance between spatial correlation function bins into account, we find no significant deviations of $\xi^{\rm sys}_{\rm t/\times}$ from zero.

Finally, in Fig. 2(i) we show the tangential and cross-components of the residual stellar ellipticity around the cluster centre in radial bins. A non-zero tangential component would immediately bias our cluster mass measurements, while a non-zero cross-component would render the diagnostic power of radially binned cross-shear used later worthless. We find that these ellipticity profiles are all consistent with zero mass for all exposures and fields. The occasional outlier bin is more than one order of magnitude below the expected shear signal.

The bias correction parameters derived in Hoekstra et al. (2015) and discussed in the previous section are for a circular PSF and, as shown in their appendix, in the presence of PSF anisotropy the smear polarizability is somewhat biased. We therefore boosted the smear polarizability artificially by 4 per cent for each object to correct for this bias. We find that the cluster masses estimated from the boosted catalogues, which are used in our analysis, are on average 1.1 per cent higher than those in the uncorrected catalogues, but not significantly so, because the mass scatter between boosted and unboosted catalogues is 2.5 per cent.

[4] http://www.cosmostat.org/software/athena/

3.1.2 *Blind analysis*

Attempting to measure cluster masses with gravitational lensing when other estimates of the cluster mass – such as SZE measurements – are already known presents the danger of the experimenter being influenced by confirmation bias. A number of procedures described in the following sections required careful checking of their behaviour with respect to varying input parameters. Any experimenter is faced with the challenge of deciding when the results of such tests are of sufficient quality. It is imperative that the metric of this decision does not make use of the actual mass measurement. If it did, we would be more likely to stop testing our procedures when the cluster masses seem to agree with our expectations from SZE measurements than when there is a discrepancy. To avoid such experimenter bias, the practice of 'blind analyses' has found widespread acceptance in particle and nuclear physics (Klein & Roodman 2005) and is being adopted in cosmology as well (e.g. von der Linden et al. 2014; Hildebrandt et al. 2017; DES Collaboration et al. 2018).

The analysis presented herein has been blinded so that no comparisons between the weak lensing and SZE-derived masses were made, which otherwise would have allowed premature inference of the weak lensing–X-ray observable scaling relation parameters. At the same time, we aimed to retain shear profiles that resemble those of massive clusters to test our analysis pipeline with the actual but blinded data. To ensure this, we adopt the following procedure to blind the normalization and scatter of the scaling relation. First, a random number $0.80 \leq x_{\rm l} < 0.95$ is drawn from a uniform probability distribution. Then for each cluster i a second random number f_i is drawn from the interval $[x_{\rm l}, 1)$. The shear values of each cluster are multiplied by f_i. We enforce $f_i < 1$ to avoid unphysical shears; at the same time, f_i cannot be very small, so as not to wipe out the lensing signal. The intrinsic ellipticity dispersion used in the calculation of the lensing weights (see Section 4.2) is not rescaled, i.e. the relative weighting of galaxies in any given cluster field is not changed by the blinding procedure.

3.1.3 *Changes after unblinding*

Although great care was taken to unblind the shear profiles only after the analysis was finalized, we realized that, inadvertently, we did not apply the multiplicative shear bias correction. This biased our masses low by much more than the average blinding factor turned out to be. The analysis we present in this article has the multiplicative shear bias correction applied. We stress that these correction factors were already computed at the time of unblinding and they remained unchanged by all further analysis changes.

We took the opportunity of this one very large shift in the analysis after unblinding, corresponding to a ∼20 per cent shift in mass, to make two small adjustments at the same time.

(i) We transitioned from the unboosted PSF correction catalogues to the boosted smear polarizability (see Section 3.1.1).

(ii) At the time of unblinding, the 2500 deg^2 SPT–SZ catalogue (Bleem et al. 2015) was not finalized and we used centroids, red-

shifts and estimated $M_{500}^{\rm SZ}$ from the catalogue of Andersson et al. (2011). We afterwards updated our analysis to use the quantities from the final SPT–SZ catalogue.

Both of these changes lead to shifts at the ∼1 per cent level in the absolute mass scale.

We also made changes to the scaling relation analysis scheme for our X-ray data. Theoretical considerations, as well as tests against mocks, revealed that the analysis scheme used in previous SPT cluster analyses led to a bias of the X-ray slope toward steeper values. The updated analysis method is described in Section 5.3 and was shown to produce unbiased results. We note that our constraints on the slope are dominated by the informative prior applied (see Section 5.5) and that we choose a pivot point in the scaling relation that essentially decouples the slope from the amplitude. Therefore, our final results are not much affected by this change.

Finally, while this manuscript was edited for submission, one of us realized that the blinding scheme described in the previous section only has a ∼2 per cent scatter in the mean blinding factor, while during the creation of this work we assumed it to be in the 10–15 per cent range. The mean blinding factor determines how well the true MOR normalization is hidden from us and is more important than the cluster-to-cluster blinding, which is indeed large in our method. Our erroneous assumption kept us effectively blind during the analysis. However, now that this flaw has been revealed, we strongly advocate against using this scheme and advise researchers to use a blinding scheme that first determines the mean blinding factor from a random variable with a large variance.

3.2 Background galaxy selection and critical surface mass density

The reduced shear g measured in weak lensing data is a dimensionless quantity. To connect it to the physical mass scales of our galaxy clusters, we need to determine the redshift distribution of the background galaxies, which enters in the critical surface mass density (equation 2). The three Megacam passbands in which we have data are not sufficient to estimate photometric redshifts for galaxies in our catalogues.

We used redshift-dependent colour cuts to reject likely foreground and cluster galaxies. Rather than optimising these colour cuts for every cluster, we divided the sample into four redshift slices. The polygons that define our colour cuts are illustrated in Fig. 3. These are based on the colour cuts defined in an earlier SPT weak lensing study (High et al. 2012) and were constructed in the same way. The density distribution of galaxies in the CFHTLS Deep Field 3 (Coupon et al. 2009) with $i < 25$ mag was plotted in $(g-r, r-i)$ –colour–colour space for (1) galaxies with photometric redshifts $|z_{\rm phot} - z_{\rm l}| < 0.05$ ('non-sources") and (2) all other galaxies ('sources'). Polygons were drawn by hand to reject the majority of non-source galaxies. More sophisticated approaches to select only background galaxies have been proposed, e.g. by Okabe & Smith (2016) and Medezinski et al. (2018), but the present scheme is sufficient for our purposes and its efficacy is demonstrated by the background map in Fig. 3. Additionally, we rejected all galaxies with $i < 20.5$ mag from the lensing catalogue, because such bright galaxies are very unlikely to be background galaxies.

We use an external catalogue with well-calibrated photometric redshifts to estimate the redshift distributions of the lensing catalogues. By applying the same cuts we use for the shear catalogues to the reference catalogue and by matching galaxy properties such as magnitude and size, we can draw photometric samples from the

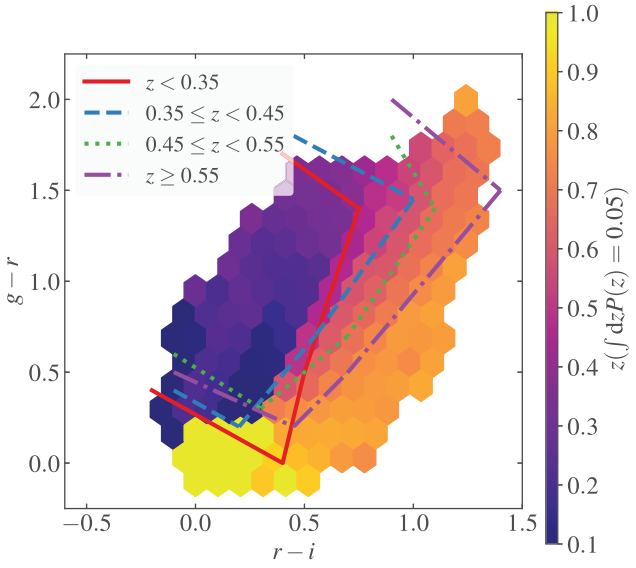


Figure 3. Colour cuts applied to the lensing catalogues to reject cluster and foreground galaxies and redshift distribution properties of the COSMOS reference catalogue (Ilbert et al. 2013). Only galaxies to the top, right and bottom of the indicated regions are kept for the lensing analysis. Different colours and line styles indicate the colour cuts applied to clusters at different redshifts. The colour-coded map indicates the fifth percentile of the photometric redshift distribution of galaxies in the COSMOS catalogue at this particular point in colour–colour space.

reference catalogue. Their photometric redshifts can then be used to determine the effective redshift distribution of our lensing catalogues.

We used a version of the COSMOS30 photometric redshift catalogue (Ilbert et al. 2013), which makes use of additional near-infrared photometry provided by the UltraVISTA survey (McCracken et al. 2012). We transformed the magnitudes in the catalogue to the CFHT system, to which our Megacam data was calibrated by using the colour terms from Capak et al. (2007):

$$\begin{aligned} g &= g^+ - 0.084(g^+ - r^+) - 0.007, \\ r &= r^+ + 0.019(g^+ - r^+) - 0.001, \\ i &= i^+ + 0.018(g^+ - r^+) - 0.005. \end{aligned} \tag{8}$$

This catalogue is complete to $i \gtrsim 24.5$ mag. Consequently, this is the limit we must adopt when performing a faint magnitude cut on the shear catalogues. Further, we impose the following constraints on galaxies in the reference catalogue:

(i) no flags set in i band;
(ii) full width at half-maximum (FWHM) > 2 pixels, to reject the stellar locus;
(iii) unsaturated in g, r and i bands;
(iv) Above 5σ detection in g, r and i bands, to reject spurious objects;
(v) same colour cuts as for the lensing catalogue;
(vi) $z < 5$, to reject objects with unrealistic photo-z estimates.

We emphasize that cuts (iii) and (iv) only remove objects from the COSMOS catalogue that cannot be present in our lensing catalogues because they are either rejected by the bright magnitude cut of the lensing catalogue or too faint to be detected in our Megacam data, where we require detection in all three passbands.

Galaxies in the shear catalogue have weights assigned to them (see Section 4.2). These are taken into account in all lensing-derived

quantities. Simply sampling from the reference catalogue, such that the samples reproduce the photometric properties of the shear catalogue without taking the lensing weights into account, could bias the computation of $\langle\beta\rangle$ and $\langle\beta^2\rangle$. The lensing weight depends chiefly on signal-to-noise ratio (SNR) and to a lesser degree on object size. We thus have to map lensing weights to the β distribution of COSMOS galaxies with the same magnitude–size distribution as in the shear catalogue.

Our version of the COSMOS catalogue (P. Capak, private communication) has a column with the object FWHM for the i-band detection image, which has not been convolved to homogenize the PSF across passbands. Assuming that atmospheric seeing causes a simple Gaussian convolution, we added the size of convolution kernels in quadrature to achieve the same seeing in the reference catalogue as the field seeings in Table 2. This is almost always possible, because the average seeing in the COSMOS field is 0.57 arcsec and thus less than the seeing in our fields, with the one exception of SPT-CL J0348–4514, which has a seeing of 0.54 arcsec. In this case, the COSMOS detection FWHM column was left unaltered.[5]

We developed an algorithm to infer (from the COSMOS catalogue) the expected β distribution for galaxies with the magnitude and size distribution of objects in the cluster-field shear catalogues, applying the lensing weights correctly. This algorithm first constructs a joint probability distribution in i'–size–lensing-weight space from the observed shear catalogue for each cluster field. Then a random deviate from this distribution is drawn and the closest match in i' magnitude and size in the COSMOS catalogue is found and the redshift of the matched COSMOS object is assigned to the random deviate. In this respect, the algorithm is similar to photo-z methods based on nearest-neighbour identification in multi-colour space (e.g. Lima et al. 2008; Cunha et al. 2009), except that we require that galaxies in the reference catalogue to follow the same magnitude and size distribution, whereas those other methods only used colour information. With a redshift (from COSMOS) and a lensing weight (from the random deviate), we can now compute weighted $\langle\beta\rangle$ and $\langle\beta^2\rangle$. In detail, the algorithm works as follows.

We construct a Gaussian kernel density estimator (KDE) of the density distribution in $i-$FWHM$-$weight space from the shear catalogue. The number of lensing galaxies with weights below a characteristic value drops sharply. This discrete feature of the density distribution and also the sharp magnitude cut at $i = 24.5$ mag are not well represented by a smooth KDE. To avoid biases at the edges of the probability density distribution, we mirror the size and magnitude distributions at their extreme values. This ensures that we have smooth distributions, which can be well described by a Gaussian KDE.

We then draw random samples from this KDE. Samples in the mirrored quadrants are flipped back into the original quadrant. For each random sample, we identify the COSMOS galaxy that minimizes the quantity

$$d = \left[\left(\frac{i_{\rm sample} - i_{\rm COSMOS}}{\sigma_i}\right)^2 + \left(\frac{{\rm FWHM}_{\rm sample} - {\rm FWHM}_{\rm COSMOS}}{\sigma_{\rm FWHM}}\right)^2\right]^{1/2} , \quad (9)$$

where the σ_x with $x \in \{i, {\rm FWHM}\}$ are the standard deviations of the i-band and FWHM distributions in the shear catalogue. This sample galaxy is assigned the weight drawn from the KDE and β and β^2 for this galaxy are computed. We verify that the samples drawn in this way from the reference catalogue are distributed consistently with the lensing catalogues by computing the Kolmogorov–Smirnov test for the marginal distributions in size and i-band magnitude.

The first two moments of the β distribution are then computed as weighted averages of β and β^2 using the lensing weights. These values are reported in Table 3.

We tested the ability of this procedure to reproduce correctly input distributions that are very different from the intrinsic COSMOS30 galaxy properties. We divided the COSMOS30 reference into two halves and created mock catalogues from one of the halves. To create the mock catalogues, we subsampled from the first half such that the magnitude distribution $P(i)$ follows the linear distribution

$$P(i) = \frac{2(i - 20.5)}{(25.0 - 20.5)^2} , \qquad 20.5 \le i \le 25.0 , \quad (10)$$

the size distribution is log-normal, $\ln P({\rm FWHM}) \sim \mathcal{N}(1, 0.0625)$, and the lensing weights are distributed according to $P(w) \sim 20 - \exp(-w)$. Just like the actual shear catalogues, these magnitude and weight distributions have sharp cut-offs to test the unbiasedness of our mirroring approach.

Following the construction of the KDE as described above, we resampled from the second half of the reference catalogues and compared the estimated values of $\langle\beta\rangle$ and $\langle\beta^2\rangle$ with the known values of the mock catalogues. We find that our resampling underestimates the true values of $\langle\beta\rangle$slightly between 0.3 per cent and 0.9 per cent as a function of redshift. At the median redshift of the cluster sample, the bias is 0.5 per cent. The values of $\langle\beta^2\rangle$ are biased low between 0.3 per cent and 0.6 per cent, with a bias of 0.5 per cent at $z = 0.42$. This level of bias is negligible for our analysis. It is caused by a slight oversampling of bright galaxies with redshifts $z < z_{\rm l}$.

We consider a number of effects contributing to uncertainties in our estimates of $\langle\beta\rangle$ and $\langle\beta^2\rangle$. First, COSMOS is a small field and variations between the LSS in COSMOS and the lines of sight to our galaxy clusters ('cosmic variance') can lead to biases. We computed $\langle\beta\rangle$ separately from all four CFHTLS fields and took the variance between these fields as our estimate for the impact of cosmic variance.[6] We find $\sigma_{\langle\beta\rangle,\,{\rm CV}} = 0.0082$. The CFHTLS photometric catalogues do not come with object size information. Computing the variance among CFHTLS fields only, rather than with the COSMOS fields also, allows us to isolate the impact of cosmic variance from the influence of object size.

Secondly, even with the high quality of the photometric redshifts of the COSMOS30 catalogue, some biases may exist, particularly at the faint, high-redshift end. To evaluate this, we matched our COSMOS catalogue with the 3D-*HST* catalogue (Momcheva et al. 2016). This catalogue contains redshifts based on spectroscopic, grism and photometric redshift estimates. We limited our comparison to objects for which the 3D-*HST* catalogue lists either spectroscopic or grism redshifts, with which the COSMOS redshifts may be compared reliably. We have 1980 objects of this type in common with their catalogue. We first computed the additional uncertainty stemming from only randomly sampling 1980 objects

[5] We ignore the wavelength mismatch between our seeing values measured in the r band and the FWHM of the COSMOS objects detected in the i band. In standard seeing theory, the difference in FWHM is only ∼4 per cent.

[6] Although we include the effect of $\langle\beta^2\rangle$ in our mass calibration, it is generally completely negligible for the radial ranges employed in this work. We therefore do not separately quote the small uncertainties in $\langle\beta^2\rangle$.

Table 2. PSF FWHM of individual cluster r-band exposures and 5σ limiting magnitude in a 2-arcsec diameter aperture of the r-band coadded image.

Cluster	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Avg.	Lim. mag.
SPT-CL J0234–5831	0.″70	0.″81	0.″74	–	0.″75	25.0
SPT-CL J0240–5946	0.″66	0.″74	0.″66	0.″73	0.″78	25.0
SPT-CL J0254–5857	0.″90	0.″87	0.″89	–	0.″89	25.0
SPT-CL J0307–6225	0.″55	0.″59	0.″65	–	0.″60	24.9
SPT-CL J0317–5935	0.″75	0.″73	0.″79	–	0.″76	25.0
SPT-CL J0346–5439	0.″71	0.″73	0.″72	–	0.″72	25.1
SPT-CL J0348–4515	0.″54	0.″54	0.″54	–	0.″54	25.2
SPT-CL J0426–5455	0.″67	0.″59	0.″58	–	0.″61	25.0
SPT-CL J0509–5342	0.″84	0.″79	0.″80	–	0.″81	25.0
SPT-CL J0516–5430	0.″69	0.″69	–	–	0.″69	24.8
SPT-CL J0551–5709	0.″79	0.″90	0.″85	–	0.″85	25.0
SPT-CL J2022–6323	0.″88	0.″89	0.″97	–	0.″91	25.1
SPT-CL J2030–5638	0.″87	0.″84	0.″80	–	0.″83	25.1
SPT-CL J2032–5627	0.″92	0.″89	0.″79	–	0.″86	24.8
SPT-CL J2135–5726	0.″88	0.″81	1.″00	–	0.″90	24.7
SPT-CL J2138–6008	0.″90	1.″11	1.″02	–	1.″01	24.5
SPT-CL J2145–5644	0.″80	0.″81	0.″82	–	0.″82	25.0
SPT-CL J2332–5358	0.″80	0.″73	0.″73	–	0.″75	25.1
SPT-CL J2355–5055	0.″66	0.″75	0.″76	–	0.″73	25.0

Table 3. Cluster redshift, source galaxy lensing efficiency and density after colour cuts.

Cluster Name	z_l	$\langle\beta\rangle$	$\langle\beta^2\rangle$	$n_{\rm gal}$ (arcmin^{-2})
SPT-CL J0234–5831	0.41	0.48	0.25	12.1
SPT-CL J0240–5946	0.40	0.50	0.27	12.3
SPT-CL J0254–5857	0.44	0.46	0.23	11.1
SPT-CL J0307–6225	0.58	0.40	0.18	7.9
SPT-CL J0317–5935	0.47	0.46	0.23	9.2
SPT-CL J0346–5439	0.53	0.40	0.18	13.1
SPT-CL J0348–4515	0.36	0.56	0.32	12.1
SPT-CL J0426–5455	0.63	0.35	0.14	8.9
SPT-CL J0509–5342	0.46	0.46	0.23	11.7
SPT-CL J0516–5430	0.29	0.60	0.37	9.3
SPT-CL J0551–5709	0.42	0.48	0.24	8.4
SPT-CL J2022–6323	0.38	0.51	0.28	7.4
SPT-CL J2030–5638	0.39	0.50	0.27	9.0
SPT-CL J2032–5627	0.28	0.60	0.37	8.4
SPT-CL J2135–5726	0.43	0.47	0.24	9.7
SPT-CL J2138–6008	0.32	0.54	0.31	4.0
SPT-CL J2145–5644	0.48	0.44	0.21	9.9
SPT-CL J2332–5358	0.40	0.51	0.27	11.5
SPT-CL J2355–5055	0.32	0.57	0.34	10.1

from the COSMOS photo-z catalogue. This is $\sigma_{\langle\beta\rangle,\,\rm sample} = 0.0013$. We then recomputed $\langle\beta\rangle$ for all clusters using only the 1980 3D-*HST* redshifts and find $\langle(\langle\beta\rangle_{\rm COSMOS} - \langle\beta\rangle_{\rm 3D\text{-}HST})/\langle\beta\rangle_{\rm COSMOS}\rangle = 0.6$ per cent, which is fully consistent with no redshift bias, up to the sampling uncertainty, where the outer average runs over all clusters. This test is reliable as long as any possible redshift bias in the COSMOS catalogue is not different for objects with or without spectroscopic or grism redshifts. At present, we have no indication of such a type-dependent bias, but also cannot confidently rule out that hitherto undiscovered biases in the COSMOS catalogue exist for faint objects.

Thirdly, we also investigate the impact of the uncertainties of the photometric calibration on our estimation of the lensing efficiency by shifting the relative and absolute photometry within the systematic calibration uncertainties. We find an additional uncertainty of $\sigma_{\langle\beta\rangle,\,\rm pc} = 0.0018$.

Finally, a more recent version of the COSMOS photo-z catalogue (Laigle et al. 2016) was published after we finalized our data vectors. We verified that this catalogue gives consistent results for $\langle\beta\rangle$ and $\langle\beta^2\rangle$, with $\Delta\langle\beta\rangle = 0.2$ per cent and $\Delta\langle\beta^2\rangle = -0.2$ per cent, and treat the difference between these catalogues as an additional source of uncertainty, $\sigma_{\langle\beta\rangle,\,\rm NC} = 0.002$.

We add all four $\sigma_{\langle\beta\rangle,\,i}$ in quadrature and arrive at a final uncertainty of $\sigma_{\langle\beta\rangle} = 0.0087$. Cluster mass scales with $M \propto \gamma^{1/\Gamma}$, where the exponent Γ depends on the clustercentric radius. For a wide range of radial fitting ranges, $\Gamma = 0.75$ (Melchior et al. 2017), and hence the systematic uncertainty in cluster mass due to uncertainty in the redshift distribution of background galaxies is 1.2 per cent. We confirmed this value by rescaling the tangential shear by a factor of 1.0087, fitting Navarro, Frenk & White (NFW) profiles (see Section 4.2) to the unscaled and scaled shear profiles and comparing the mass estimates.

Our sampling from the reference catalogue also enables us to estimate the fraction of foreground galaxies surviving our colour cuts and diluting the shear signal without biasing it. This is shown in Fig. 4. The fraction of low-redshift interlopers is below 2 per cent for clusters at redshift $z < 0.45$. At higher redshifts it jumps to $\sim$5 per cent. It is possible to optimize the colour cuts to keep the low-z contamination at $\sim$2 per cent for the $0.45 < z < 0.55$ redshift bin also, but this optimization happens at the cost of an increased contamination of the shear catalogue by cluster galaxies, as we discuss in detail in the next section.

3.3 Cluster contamination correction

Sampling from the reference catalogue – as in the preceding section – allows us to estimate the background properties and foreground contamination of the shear catalogues. However, it does not allow us to estimate the contamination by cluster galaxies remaining after the colour cuts in the shear catalogues, because cluster galaxies are a very significant overdensity in redshift space not present in the reference catalogue. Contamination of the shear catalogues by

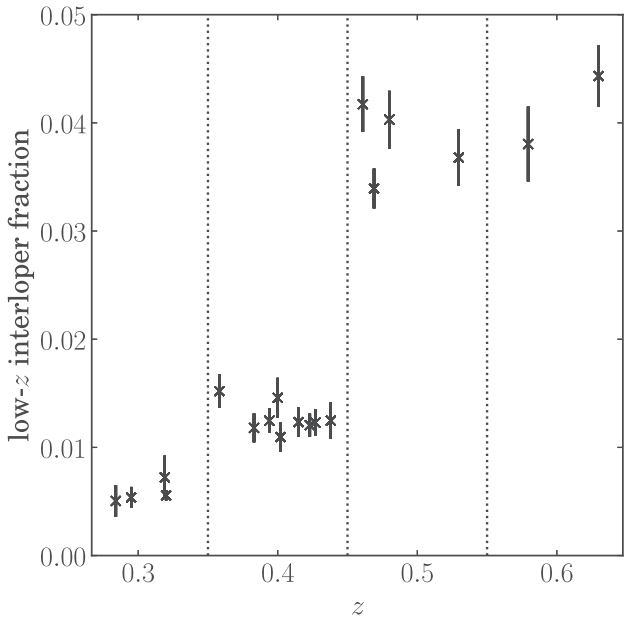


Figure 4. Fraction of low-redshift galaxies surviving our colour cuts as a function of redshift estimated by sampling from the reference catalogue. The error bars are the standard deviation of the mean number of low-z interloper galaxies. The vertical dotted lines indicate the transitions from one colour cut to another, as illustrated in Fig. 3.

cluster galaxies dilutes the shear signal, as these galaxies are not lensed and show no specific alignment (e.g. Sifón et al. 2015). Thus, they should be counted as contributing $\beta = 0$ in the estimation of the lensing efficiency.

We implement and test two different methods to estimate the contamination fraction in our cluster sample. All methods looking at radial variations of a population must carefully keep track of areas not available for observations of that population (Simet & Mandelbaum 2015; Hoekstra et al. 2015). We therefore use the mask files created for the magnification study of Chiu et al. (2016a), where details on their generation are provided. Briefly, regions covered by extended bright objects are automatically masked by SExtractor, while satellite trails and diffraction spikes are manually masked. We determine the cluster contamination fraction in radial bins and correct the bin area for masked pixels in both methods, keeping track of the area covered by bright galaxies not already included in the pixel masks. An increased incidence of blending could in principle also lead to a depletion of object detections in higher density environments. The simulations of Chiu et al. (2016b) show that this is not a problem for the choice of radial range (0.75–2.5 Mpc) considered in the present study.

3.3.1 *Number density profile*

As in Applegate et al. (2014) – based on an approach by Hoekstra (2007) – the radial profile of contaminating galaxies is modelled as

$$f_{\rm cl}(r) = \frac{n_{\rm cl}(r)}{n_{\rm cl}(r) + n_{\rm gal}} = f_{500} \exp\left(1 - \frac{r}{r_{\rm SZ,500}}\right) , \qquad (11)$$

where f_{500} is the contamination fraction at $r_{\rm SZ,\,500}$, the SZE-determined radius r_{500}, and $n_{\rm gal}$ is the number density of background galaxies. An important consideration in our case is that this approach does not rely on measuring the background number density of galaxies far away from the cluster centres, but treats it as a free parameter. The virial radius of most clusters in our sample is only slightly smaller than the FOV of Megacam, affording us no area completely free from cluster galaxies.

Table 4. Cluster contamination fractions at r_{500} show no dependence on inner radial cuts, indicating that the fits are not affected by decreasing catalogue completeness towards the cluster centre.

Method	Bin rejection	f_{500}
Applegate et al. (2014)	<2 arcmin	(4.8 ± 2.5)%
	<3 arcmin	(5.5 ± 3.1)%
	$<r_{\rm SZ,\,500}$	(4.9 ± 3.2)%
Gruen et al. (2014)	–	(2.3 ± 1.7)%

Like Applegate et al. (2014), we see an upturn in the number density in most cluster fields towards the centre. Per field measurements of cluster contamination fractions are nevertheless too noisy to be meaningful and we adopt their strategy of fitting all clusters simultaneously with a single contamination fraction f_{500} and per field $n_{\rm gal}$ values. We bin the shear catalogue radially in angular bins of width 1 arcmin from the cluster centre out to a maximum radius of 12 arcmin. We assume Poisson errors on the binned number counts. After binning, we rescale the bin locations to units of the SZE-derived r_{500} of each cluster. We emphasize that this is the only step in our analysis that depends on an SZE-derived cluster mass. Its only purpose is to correct for the range in cluster mass and any systematic covariances between the weak lensing-derived cluster masses and their SZE-based estimates introduced by this scaling are subdominant to the relatively large statistical errors in the contamination fraction.

We reject some inner bins in the fitting procedure because we do not fit the shear profiles all the way to the centre. Among other reasons, we try to minimize the impact of cluster miscentring, which would also affect the number density profiles. Another effect that could potentially be important in the inner bins but was verified to be of negligible influence in our analysis is the impact of magnification (Chiu et al. 2016b).

Table 4shows the result of performing the fit in this way and removing a varying number of bins close to the cluster centre.[7] We find no dependence of f_{500} on the inner fit radius, indicating that, over the radial ranges considered here, the catalogue completeness does not change strongly. The error bars are estimated by bootstrap resampling from the cluster sample. The estimated background galaxy number densities are reported in Table 3.

To test a possible redshift dependence of the cluster contamination, we split the sample at $z = 0.45$, where the foreground contamination shows a strong jump when we transition to a different colour cut regime. We find $f_{500}^{z>0.45} = (4.1 \pm 4.82)$ per cent and $f_{500}^{z<0.45} = (5.0 \pm 2.9)$ per cent when the fit is restricted to $r > 2$ arcmin. Both numbers are consistent with each other and the value reported in Table 4 if $r > 2$ arcmin is imposed.

Additionally, we test whether we can reduce the foreground contamination by adjusting the colour cuts without adversely affecting the cluster contamination fraction. We remove the colour cut transition at $z = 0.45$ and apply the colour cuts used for objects in the redshift range $0.35 \leq z < 0.45$ over the range $0.35 \leq z < 0.55$ instead. Indeed, this reduces the foreground contamination for the four clusters in this bin to $\lesssim$2–3 per cent. At the same time, we

[7] For most clusters in our sample, $r_{\rm SZ,\,500} > 3$ arcmin.

notice a significant steepening of the number density profiles of these four clusters, indicating an increased contamination by cluster galaxies. On the one hand, the dilution of the shear signal by foreground galaxies is taken care of reliably by setting their $\beta = 0$ in the estimation of $\langle\beta\rangle$ and $\langle\beta^2\rangle$. On the other hand, we know that the reference field cannot be a faithful representation of the galaxy density in redshift space in the presence of a massive cluster. Given the low SNR of our f_{500} measurement and the relative straightforwardness of the redshift sampling in Section 3.2, we prefer to optimize our colour cuts for rejection of cluster galaxies.

3.3.2 *Redshift-space decomposition*

An alternative method to fitting an analytical number density profile was proposed by Gruen et al. (2014). Briefly, they looked at the probability distribution of the lensing efficiency β and decomposed the observed probability distribution at a given clustercentric radius r into the cluster and field galaxy probability distribution,

$$p(\beta, r) = f_{\rm cl}(r)\, p_{\rm cl}(\beta, r) + (1 - f_{\rm cl}(r))\, p_{\rm f}(\beta)\,, \qquad (12)$$

where $f_{\rm cl}(r)$ is the radially dependent cluster contamination fraction. Once $p(\beta,\, r)$, $p_{\rm cl}(\beta,\, r)$ and $p_{\rm f}(\beta)$ are known, the contamination fraction can be found by simple χ^2 minimization. We additionally imposed the constraint that $f_{\rm cl} \in [0,\, 1]$. This method works if its two underlying assumptions are fulfilled:

(i) the redshift distribution of galaxies is constant over the image;
(ii) the cluster and field probability distributions $p_{\rm cl}$ and $p_{\rm f}$ are sufficiently independent that the full distribution function $p(\beta)$ can be written as a linear combination of the two.

It is reasonable to assume that the first condition is met in our case, because our images have a homogeneous depth per field and cover only a small solid angle. We experimentally verified that the second condition is also fulfilled by plotting $p_{\rm cl}(\beta)$ and $p_{\rm f}$. We estimated these distributions from the reference catalogue in the manner described by Gruen et al. (2014), which we summarize here.

The distributions $p(\beta,\, r)$ and $p_{\rm cl}(\beta,\, r)$ are estimated in annuli around the cluster centre. We chose nine bins of width 1 arcmin starting at the cluster centre. In each bin, for every object in the shear catalogue with magnitudes $\{g,\, r,\, i\}$ we take galaxies with $\sqrt{(\Delta g)^2 + (\Delta r)^2 + (\Delta i)^2} < 0.1$ mag from the reference catalogue. For each such sample, we compute the probability $P_{\rm cl}$ that the respective object is a cluster galaxy by assigning it the fraction of sample galaxies that have $|z - z_{\rm l}| \leq 0.06(1 + z_{\rm l})$. Also, for every sampled galaxy, we compute $\langle\beta\rangle$ from the COSMOS sample. The unweighted histogram of these $\langle\beta\rangle$ values is $p(\beta)$. The histogram weighted by the $P_{\rm cl}$ values is $p_{\rm cl}(\beta)$.

The probability distribution of β for field galaxies is estimated in a similar fashion. For each object in the shear catalogue at a large distance from the cluster – we choose $r > 10$ arcmin – samples are drawn in the same way. The probability $P_{\rm f}$ that a galaxy is a field galaxy is assigned the fraction of sample objects with $z < z_{\rm l} - 0.06(1 + z_{\rm l})$. Again, the value of $\langle\beta\rangle$ of the samples for each shear catalogue object is computed. A histogram weighted by the probabilities $P_{\rm f}$ is the distribution $p_{\rm f}(\beta)$. Following Gruen et al. (2014), the choice of $0.06(1 + z_{\rm l})$ as separation here and for computing the probability that a galaxy is a cluster galaxy is based on the 2σ uncertainty of the photometric redshifts in our reference catalogue. Varying this parameter does not influence our estimates of the contamination fraction systematically.

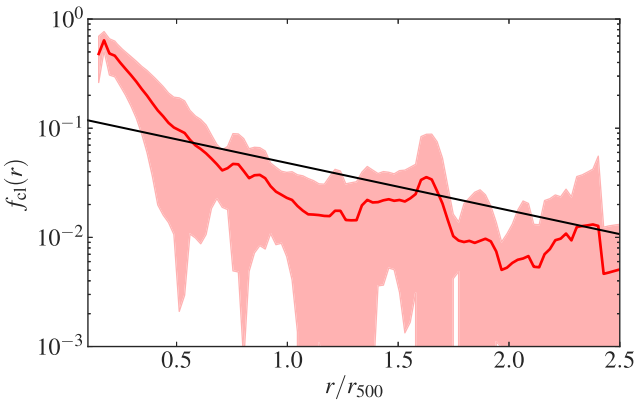


Figure 5. Contamination correction derived by the method of Gruen et al. (2014). The thick red line is a robust estimation of the mean of all clusters with its error range indicated by the shaded region. The solid black line is the exponential model derived in Section 3.3.1 for comparison.

Table 5. Impact of the cluster contamination correction on the mean cluster mass relative to no correction for various choices of inner radial fitting range. r_{500} is derived from the SZE-based mass estimate in Bleem et al. (2015).

inner radius	Applegate et al. (2014)	Gruen et al. (2014)
$0.5\, r_{500}$	3.8%	3.6%
$0.7\, r_{500}$	2.0%	0.1%
500 kpc	3.7%	3.1%
750 kpc	2.5%	1.6%

Fig. 5shows the radial contamination profile fraction derived in this way for the ensemble of all clusters. As in the case of the exponential contamination model, we found that individual cluster estimates are very noisy and there is no obvious redshift trend. Instead of correcting each cluster profile with its own noisy contamination profile $f_{\rm cl}(r)$, we estimate an average contamination profile and its error using the robust location and scale estimator of Beers, Flynn & Gebhardt (1990).

We compare the impact both methods have when they are applied over different radial ranges in the process of fitting NFW profiles (see section 4.2 of Navarro, Frenk & White 1997) to the tangential shear. We measure the relative change of mass compared with a profile fit ignoring the contamination correction. In all cases, the outer radial range considered is 2.5 Mpc and the inner radius takes the values listed in Table 5. We conclude that both methods agree to better than 2 per cent outside $0.65\, r_{500}$. As one would expect, larger corrections are necessary if one decreases the inner radius of the shear profile analysis. Nevertheless, we find that the purely empirical decomposition method is significantly steeper than the exponential model at smaller radii, indicating that the latter is not a good model and the actual contamination profile is more similar to the cored $1/r$ profile employed in Hoekstra et al. (2015). We take the difference of 0.9 per cent in mass (see the last line of Table 5, which uses the inner radius later employed in this work) between both methods considered here as an upper limit on the impact of the systematic uncertainty of the contamination correction.

We also tested for the existence of a mass-dependent trend in the mean f_{500} by splitting the cluster sample into two equal-sized bins along the detection significance ξ. The contamination fractions measured in both bins are statistically indistinguishable and fully consistent with the one determined for the whole cluster sample,

excluding any significant mass trend at the current level of uncertainty.

4 WEAK-LENSING MASS MEASUREMENTS

We present reconstructions of the projected mass density in Section 4.1 and constrain the mass of our galaxy clusters with fits to analytical shear profiles in Section 4.2. As we will discuss, these fits represent biased mass estimators, which can be calibrated with simulations. First, however, the uncorrected fits can be compared with mass estimates obtained for a subset of our clusters from the same data by High et al. (2012).

4.1 Mass reconstruction maps

Cluster mass maps are often instructive to assess the weak lensing detection of a galaxy cluster and to compare light and mass distributions. We used the finite field inversion method of Seitz & Schneider (2001) to obtain reconstructed κ-maps from the observed shear fields with a smoothing of 2 arcmin, which was selected based on the visual impression of the reconstructed maps. To compute the noise levels of the surface mass density reconstruction, we create 800 realizations of the shear catalogues with randomly rotated galaxy ellipticities while keeping the absolute value of the ellipticity and the galaxy positions fixed. The variance of these random maps is used as a noise estimator for each pixel, although pixels within the smoothing scale are of course highly correlated. Dividing the κ-map by noise maps created in this way gives SNR maps with the contours shown in the left panels of Figs B1–B19 in Appendix B.

In these figures, we compare the weak lensing significance contours with significance contours of filtered SPT–SZ maps and significance contours of the density of colour-selected red-sequence cluster galaxies. Although the SNR of the WL reconstruction is low, in most cases we find good agreement between the SPT and the WL centroid. Sizeable offsets between those are expected, due to shape noise and smoothing of an asymmetric mass distribution with a symmetric kernel (Dietrich et al. 2012), even in the absence of collisional processes separating the dark matter and gas components of a galaxy cluster (e.g. Clowe et al. 2006). The only noteworthy case in this gallery is SPT-CL J2355–5055 (Fig. B19), the field of which shows another cluster west of the SPT detection in the galaxy density contour with almost identical colours and an elongated structure extending north-east from this second cluster. These are not detected by SPT, but seem to be broadly traced, albeit at very low significance, by the mass reconstruction.

4.2 NFW profile fits

Average density profiles of galaxy clusters in cosmological simulations are known to follow a universal density profile,

$$\rho(r) = \frac{\delta_c \rho_{\rm crit}}{(c\, r/r_{200})(1 + c r/r_{200})^2} \,, \tag{13}$$

first described by Navarro et al. (1997) to a very good approximation. Here, r is the three-dimensional radius from the cluster centre, $\rho_{\rm crit}$ is the critical density of the Universe at the cluster redshift, r_{200} is the radius at which the enclosed mean density is $200\rho_{\rm crit}$, c is the concentration parameter, which determines how fast the density profile turns over from $\propto r^{-1}$ to $\propto r^{-3}$, and δ_c is a characteristic overdensity,

$$\delta_c = \frac{200}{3} \frac{c^2}{\ln(1+c) - c/(1+c)} \,. \tag{14}$$

Although the NFW profile is a very good approximation of the average density profile of galaxy clusters (e.g. Johnston et al. 2007), better fitting descriptions exist. The Einasto (1965) profile is a better description of the density profile close to the centre. At large radii ($>r_{200}$), correlated large-scale structure leads to systematic deviations from the NFW profile (Johnston et al. 2007). For the radial ranges of interest in this work, however, the original NFW profile with its well-known lensing properties (Bartelmann 1996; Wright & Brainerd 2000) is a sufficiently good description of isolated haloes. We will calibrate the impact of deviations from spherical NFW profiles using simulations (cf. Section 4.4).

We fit spherical NFW profiles to the binned tangential shear over the range 750 kpc–2.5 Mpc. Going further inwards would increase our sensitivity to miscentring (e.g. Johnston et al. 2007; Mandelbaum et al. 2010), the cluster contamination correction (see Table 5) and the mass–concentration relation, which is difficult to measure using weak lensing data alone. Going further outwards, deviations from an NFW profile become more pronounced (Becker & Kravtsov 2011), due to correlated (Johnston et al. 2007) and uncorrelated (Hoekstra 2003; Dodelson 2004) LSS. We choose the SZE peak position as cluster centre for the Megacam cluster sample and the X-ray centroid as cluster centre for the *HST* sample. We use eight linearly spaced bins over this radial range and compute weighted averages of the reduced shear in each bin:

$$\langle g_i \rangle = \frac{\sum_n w_n g_{i,n}}{\sum_n w_n} \,, \qquad i \in \{\mathrm{t}, \times\} \,, \tag{15}$$

using the lensing weights

$$w = \frac{P^{\gamma\,2}}{\sigma_\varepsilon^2 P^{\gamma\,2} + (\Delta e)^2} \,, \tag{16}$$

where P^γ is the shear polarizability (Hoekstra et al. 1998), σ_ε is the intrinsic ellipticity dispersion, which we fixed to 0.25, and Δe is the error estimate for the polarization (Hoekstra et al. 2000). The errors of the mean shear in each bin are computed as

$$\frac{1}{\sigma^2_{\langle g_i \rangle}} = \sum_n w_n \,. \tag{17}$$

We use the weighted average of the radial galaxy positions in a bin as the effective bin location. We verified that the number of radial bins and their location has no systematic influence on the recovered cluster masses for a wide range of binning schemes, if we restrict the fitting procedure to the chosen radial range of 0.75 Mpc $< r <$ 2.5 Mpc.

We correct the binned tangential shear for the remaining contamination with cluster galaxies via

$$\langle g_{\rm t,cor} \rangle (r) = \frac{\langle g_{\rm t} \rangle (r)}{(1 - f_{\rm cl}(r))} \,, \tag{18}$$

where we use the mean contamination fraction of all clusters derived from the method of Gruen et al. (2014) in Section 3.3.2. We propagate the uncertainties of this $f_{\rm cl}(r)$ profile to the reduced shear error estimates, equation (17).

When fitting the model to the observed reduced shear profile, we treat the NFW model as a one-parameter family, with M_{200} being the only free parameter, and fix the concentration parameter c to follow a mass–concentration scaling relation exactly with no intrinsic scatter. Specifically, we adopt the M–c relation of Diemer & Kravtsov (2015). This choice is justified by recent observational constraints on the M–c relation for the mass and redshift range of the Megacam cluster sample (Merten et al. 2015; Cibirka et al.

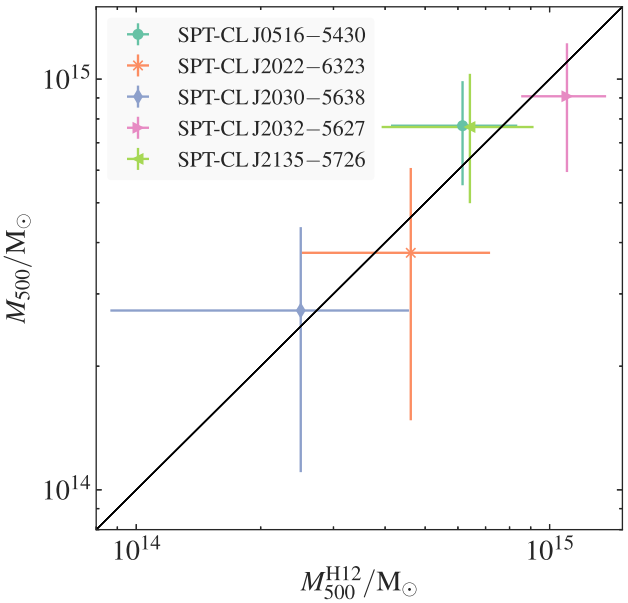


Figure 6. A comparison of the spherical NFW masses evaluated at $R_{500,\,\mathrm{WL}}$ between High et al. (2012) and this work. The black horizontal line show the one-to-one relation and is not a fit to the data points.

2017) and by measurements of the concentration in the *HST* sample itself (S18).

Observed shear profiles with best-fitting NFW models are presented in Appendix B. For all clusters, the cross-shear is consistent with zero, as expected for shear catalogues that are not significantly affected by systematics.

4.3 Comparison with earlier mass measurements

Our weak lensing analysis of Megacam data is a significant expansion of an earlier analysis of a subset of five clusters (High et al. 2012). A comparison of the mass estimates obtained is a natural part of our analysis. Although we use the same data as High et al. (2012), our analysis differs in a few key features, as described in the previous sections. Most importantly, these are (i) new shear catalogues with new PSF model and multiplicative shear bias correction, (ii) updated estimates for $\langle\beta\rangle$ and $\langle\beta^2\rangle$, (iii) improved estimation of the cluster contamination correction and (iv) different mass–concentration scaling relation.

Nevertheless, the mass estimates from this previous work and our analysis are in agreement. Fig. 6 shows a comparison of the M_{500} masses obtained from NFW fits of High et al. (2012) and our mass estimates. The weighted difference is $\langle M_{500} - M_{500}^{\mathrm{H12}}\rangle = (0.0 \pm 1.3) \times 10^{14}\,\mathrm{M}_\odot$. Given the changes in the analysis mentioned above, we consider this agreement to be coincidental. We also emphasize that these changes were made to make our mass estimates more robust and obtain better limits on the systematic uncertainties of our analysis procedures.

One example where our new methods lead to significantly different results from the one described in High et al. (2012) is the $\langle\beta\rangle$ estimation for shallow fields, the completeness of which drops sharply before our limiting magnitude of $i = 24.5$ mag. SPT-CL J2138–6008 is one such field not present in High et al. (2012), in which the cluster mass would have been overestimated by ∼14 per cent in the original analysis, leaving everything else the analysis pipeline unchanged.

4.4 Calibration of the NFW fits with simulations

As mentioned in the previous section, systematic deviations from the NFW profile and miscentring lead to biased mass estimates when fitting an NFW profile to the tangential shear. Furthermore, halo triaxiality (Clowe et al. 2004; Corless & King 2007) and projected LSS lead to additional scatter. We characterize the relation between measured weak lensing mass and true mass with a bias parameter b_{WL},

$$M_{\mathrm{WL}} = b_{\mathrm{WL}} M_{500}\,, \tag{19}$$

and scatter σ_{WL}. This scatter consists of two components: (i) a local component caused by the aforementioned deviations from a spherical NFW profile and correlated LSS, $\sigma_{\mathrm{WL,\,local}}$, assumed to be log-normal in weak lensing mass at fixed true mass; (ii) scatter caused by the projection of uncorrelated LSS, $\sigma_{\mathrm{WL,\,LSS}}$.

Our approach to calibrate b_{WL} and $\sigma_{\mathrm{WL,\,local}}$ is to create an ensemble of simulated observations that match the observational properties of a random subset of cluster fields and then apply the same measurement technique as we do to the real data. In general, we are aiming to reconstruct the probability distribution $P(M_{\mathrm{WL}}|M_{\mathrm{true}})$, which can then be included in forward probabilistic modelling of the cluster sample. However, we simplify the relation as stated above to one log-normal distribution that is the same for all observed cluster fields. Any residuals from such an oversimplification are still insignificant compared with the statistical precision of our data set.

To build our simulated observations for one observed cluster field, we start with the N-body simulations from Becker & Kravtsov (2011). These are 1-Gpc boxes with 1024^3 dark matter particles with a mass of $6.98 \times 10^{10}\,\mathrm{M}_\odot$ each. We cut out $400\,h^{-1}$ Mpc long boxes centred on the most massive 788 haloes with $M_{500,c} > 1.5 \times 10^{14}\,h^{-1}\mathrm{M}_\odot$ from the $z = 0.5$ snapshot. Particles are projected to form 2D mass maps that are then used to create shear maps via fast Fourier transforms. The observed $\langle\beta\rangle$ from a cluster observation is used to scale the shear and κ maps appropriately. Random Gaussian noise is added to the shear map to match the observed shape noise in the observations. Because in our real observations we fit a one-dimensional (1-D) profile, we select an 'observed' cluster centre for each simulation map. We assume that the displacement between the true projected centre of the simulated cluster and the 'observed' centre is randomly oriented with respect to the underlying structure, a reasonable assumption given the noise sources of SPT observations and the statistical power of this sample. Centre offsets are randomly chosen following the form specified by Song et al. (2012), a Gaussian distribution with a width dependent on the SPT beam size and the core radius of the matched filter used to detect the observed cluster. The simulated 1-D profiles are then fitted with an NFW model as in the data analysis.

We assume that $P(M_{\mathrm{WL}}|M_{\mathrm{true}})$ follows a log-normal distribution with location and scale parameters $\mu = \ln b_{\mathrm{WL}}$ and $\sigma = \sigma_{\mathrm{WL,\,local}}$, respectively. For the set of simulated fields, we find the maximum *a posteriori* location for the probability distribution:

$$P(b_{\mathrm{WL}}, \sigma_{\mathrm{WL,local}}|\mathrm{mocks}) \propto \prod_i \int P(b_{\mathrm{WL}}, \sigma_{\mathrm{WL,local}}|M_{\mathrm{WL}}) P(M_{\mathrm{WL}}|\mathrm{mock}_i)\mathrm{d}M_{\mathrm{WL}}\,. \tag{20}$$

Uninformative priors were used for the parameters of interest. Simulated observations were also created and analysed using the $z = 0.25$ snapshot from Becker & Kravtsov (2011) as well as the Millennium-XXL simulations (Angulo et al. 2012). No significant trends were seen between snapshots or simulations. We also did not see any significant trend with the observational properties of

each observed field, including the amount of shape noise or different filter core size. Our final bias ($b_{\rm WL} = 0.938 \pm 0.028$) and scatter ($\sigma_{\rm WL,\,local} = 0.214 \pm 0.040$) are then the average across the random subset of cluster fields targeted for mock-up when the mass–concentration relation of Diemer & Kravtsov (2015) is used.

4.5 Impact of the mass–concentration relation

Weak lensing data often provide poor constraints on the concentration parameter of the NFW profile. The shear signal is determined by the enclosed mass at each radius and the NFW scale radius is typically interior to the innermost radius at which the shear is measured reliably. Without observing the mass profile shape around the scale radius, via the shear profile, our analysis can only provide very weak lower bounds on the concentrations. Because most mass–concentration relations in the literature seem to agree that there is a lower bound on concentrations at $c \sim 2$–3, we only fit for the NFW mass of our clusters and enforce that they follow a mass–concentration scaling relation. Any mismatch between this relation and true galaxy clusters then introduces another source of systematic error that we need to take into account.

We can estimate the sensitivity of our analysis to uncertainty in the published mass–concentration relation by carrying out the NFW fit bias analysis of the previous section for different fixed concentrations. We find that the average mass bias at concentrations $c = 5$ and $c = 3$ is $b_{\rm WL} = 0.978$ and $b_{\rm WL} = 0.907$, respectively, implying $\mathrm{d}b_{\rm WL}/\mathrm{d}c|_{c=4} = -0.0355$. Using Gaussian error propagation on equation (19), we obtain

$$\left(\frac{\sigma_M}{M_{\rm true}}\right)^2 = \frac{1}{{b_{\rm WL}}^2}\left(\frac{\mathrm{d}b_{\rm WL}}{\mathrm{d}c}\right)^2 \sigma_c^2 \,. \tag{21}$$

Because we calibrated the bias resulting from NFW fits in Section 4.4 using our chosen M–c relation, namely the one of Diemer & Kravtsov (2015), the systematic uncertainty is not given by how well this relation describes the actual cluster sample, but by how faithfully the simulated clusters represent true clusters in the Universe. The simulations used in the previous section are dark matter only and thus the question is how much the concentrations for clusters of the mass and redshift in our sample would be impacted by baryonic effects. Duffy et al. (2010) constrain this to an upper limit of 10 per cent. Evaluating equation (21), we set $\sigma_c|_{c=4} = 0.4$ and obtain a mass uncertainty due to the mass–concentration relation of 1.5 per cent.

4.6 Impact of the miscentring model

Our baseline model for the distribution of offsets between the SZE peak position, which we use as the cluster centre in our analysis of the Megacam data, and the true cluster centre is the analytical form of Song et al. (2012) described in Section 4.4. We estimate limits on the impact on the mass calibration of this miscentring uncertainty by running the NFW bias analysis of Section 4.4 with a different miscentring model. We use a miscentring distribution adopted from the analysis of Saro et al. (2014), but based on cosmological hydrodynamical simulations with both large volume and high resolution (see e.g. Bocquet et al. 2016; Gupta et al. 2017). This includes a mock SZE signal and a simulation of the SPT cluster detection procedure, which uses the multi-frequency adaptive filter method (Melin, Bartlett & Delabrouille 2006). Briefly, a β profile with $\beta = 1$ is used as cluster template, with 12 different core radii $\theta_{\rm core}$, the same as used by SPT (Bleem et al. 2015). The highest signal-to-noise peaks within the larger of $\theta_{\rm core}$ or 1 arcmin are picked as individual cluster candidates, with the peak position as the centre. The SZE peaks identified in this way are matched to the projected halo centre, which is the most bound particle.

For this miscentring distribution and the Diemer & Kravtsov (2015) M–c relation, we find a weak lensing bias $b_{\rm WL} = 0.960 \pm 0.027$. Based on the difference from the $b_{\rm WL}$ value in our baseline analysis, we conservatively assume an uncertainty of 3 per cent in the weak lensing bias parameter.

Table 6. Overview of all known systematic error sources and their contributions to the overall systematic error budget. The different error sources are added in quadrature to obtain the total systematic error estimate (68% confidence).

Error Source	Impact on Mass	Reference
Multiplicative shear bias	2%	§ 3
PSF boost correction	2.5%	§ 3
$\langle\beta\rangle$ and $\langle\beta^2\rangle$ estimation	1.2%	§ 3
Contamination correction	0.9%	§ 3
NFW mass bias	2.8%	§ 4
M–c relation	1.5%	§ 4
Miscentring distribution	3%	§ 4
Total	5.6%	

4.7 Summary of systematic uncertainties

We now briefly summarize all contributions to our systematic uncertainty budget. An overview is presented in Table 6. Broadly, these fall into two categories: observational uncertainties and modelling uncertainties. We considered observational biases in Section 3. The first two of these four pertain to how well we can measure shear. Based on Hoekstra et al. (2015), the impact on mass of the multiplicative shear bias was estimated to be < 2 per cent. Additionally, the shear calibration of Hoekstra et al. (2015) was derived for a circular PSF. For the strongly anisotropic PSF in our data, an additional boost to the smear polarizability was suggested by Hoekstra et al. (2015) to avoid biases. Applying this correction led to an additional scatter of 2.5 per cent in mass.

The second set of observational systematics is due to uncertainties in the redshift estimates of galaxies. First, uncertainties in $\langle\beta\rangle$ and $\langle\beta^2\rangle$ come from cosmic variance of the reference field, uncertainties in the photometric redshifts of the reference field itself and uncertainties in our photometric calibration. This contributes 1.2 per cent to our systematic errors. Second, cluster galaxies evading our colour–colour cuts dilute the shear signal. We model this small signal in Section 3.3 using two approaches. We propagate the uncertainties of the model we judged to be more reliable to the *statistical* error budget and treat the difference between the two models as a source of systematic uncertainty. This difference amounts to 0.9 per cent in mass.

We considered the second category of modelling errors in Sections 4.4–4.6. We discussed three sources of modelling errors. First, a bias incurred by fitting an NFW profile following a fixed mass–concentration relation to shear profiles that could deviate from an NFW profile, e.g. from correlated LSS, miscentring and obeying a different M–c relation. We calibrate this bias factor $b_{\rm WL}$, equation (19), on N-body simulations and use its uncertainty of

2.8 per cent as the systematic error number in Table 6. Second, we used previous estimates (Duffy et al. 2010) of how much the concentration of simulated dark-matter only haloes may depart from the true cluster concentration to estimate the impact of baryonic effects on mass. This amounts to 1.5 per cent in our error budget. Finally, we studied how much uncertainties in our miscentring model affect the mass estimates. The 2.8 per cent error in $b_{\rm WL}$ quoted above is only the uncertainty of the NFW mass bias calibration for our chosen miscentring baseline model. Replacing this model with another leads to a different estimate of $b_{\rm WL}$. We take this difference of 3 per cent as the uncertainty caused by the choice of miscentring model.

Because the various sources of systematic uncertainties are not expected to be correlated, we sum them in quadrature to obtain the final systematic error budget of 5.6 per cent.

5 MASS–OBSERVABLE SCALING RELATIONS AND LIKELIHOOD FUNCTION

We use our cluster data set, containing SZE, X-ray and weak lensing measurements, to constrain the mass–observable relations for all observables. We consider two different observables for the X-ray scaling relations, the gas mass $M_{\rm gas}$ and $Y_{\rm X}$. Because both observables share the same gas mass measurements, they are not independent and we do not run any fits for both X-ray relations simultaneously; rather, we fit for either one or the other relation. In the following, we discuss all mass–observable relations, the likelihood function and our choice of priors.

5.1 SZE and X-ray scaling relations

Galaxy clusters in the SPT–SZ survey were detected via their thermal SZE in the 95- and 150-GHz maps via a multi-scale matched filter technique (Melin et al. 2006). The observable used to quantify the cluster SZE signal is ξ, the detection significance maximized over all filter scales. These filter scales are a set of 12 linearly spaced values from 0.5–2.5 arcmin and the filter scale that maximizes the detection significance is associated with the cluster core radius $\theta_{\rm c}$. Due to noise bias, ξ is a biased estimator of SNR. Therefore, an unbiased SZE significance ζ is introduced, corresponding to the signal-to-noise ratio at the true cluster position and filter scale (Vanderlinde et al. 2010). For $\xi > 2$,

$$\zeta = \sqrt{\langle\xi\rangle^2 - 3} \tag{22}$$

describes the relation between ξ and ζ, with scatter described by a Gaussian of unit width, where the average is taken over many noise realizations.

The unbiased SNR ζ can be related to cluster mass by the mass–observable scaling relation

$$\zeta = A_{\rm SZ}\left(\frac{0.7M_{500}}{3\times10^{14}\,{\rm M}_\odot h_{70}^{-1}}\right)^{B_{\rm SZ}}\left(\frac{E(z)}{E(0.6)}\right)^{C_{\rm SZ}}, \tag{23}$$

where $A_{\rm SZ}$ is the normalization, $B_{\rm SZ}$ the mass slope, $C_{\rm SZ}$ the redshift evolution and $E(z) = H(z)/H_0$. An additional parameter $\sigma_{\ln\zeta}$ describes the intrinsic scatter in ζ, which is assumed to be log-normal and constant as a function of mass and redshift.

We also relate the X-ray observables to cluster mass via mass–observable scaling relations,

$$\frac{Y_{\rm X}}{10^{14}\,{\rm M}_\odot\,{\rm keV}} = A_{Y_{\rm X}}\left(\frac{M_{500}}{5\times10^{14}\,{\rm M}_\odot\sqrt{0.7h_{70}}}\right)^{B_{Y_{\rm X}}} \times\left(\frac{E(z)}{E(0.6)}\right)^{C_{Y_{\rm X}}} \tag{24}$$

and

$$\frac{M_{\rm gas}}{5\times10^{14}\,{\rm M}_\odot} = A_{M_{\rm g}}\left(\frac{M_{500}}{5\times10^{14}\,{\rm M}_\odot}\right)^{B_{M_{\rm g}}}\left(\frac{E(z)}{E(0.6)}\right)^{C_{M_{\rm g}}}, \tag{25}$$

and assume a corresponding log-normal scatter $\sigma_{\ln Y_{\rm X}}$ ($\sigma_{\ln M_{\rm g}}$) in $Y_{\rm X}$ ($M_{\rm gas}$) at fixed mass. Note that we use the same redshift pivots as for the SZE scaling relation, but apply a slightly larger pivot point in mass, corresponding approximately to the median mass of the subsample with available X-ray observations. Also note that the parametrization of the $Y_{\rm X}$–mass relation we use here departs from the one used in previous work by the SPT collaboration (e.g. de Haan et al. 2016). We write $Y_{\rm X}$ as a function of mass, so that all mass–observable-relations (23)– (25) have the observable on the left-hand side.

5.2 Weak lensing modelling systematics

As discussed in Section 4.4, we assume a relation between the *weak lensing mass* obtained from fitting an NFW profile to our shear data and the unobservable, *true mass* $M_{\rm WL} = b_{\rm WL}M_{500}$. The normalization $b_{\rm WL}$ and the scatter about this mean relation are calibrated by taking modelling and measurement uncertainties into account; we use numerical simulations for the modelling part. As our weak lensing data set consists of two subsamples – Megacam and *HST* – with slightly different measurement and analysis schemes, we expect some systematics to be shared among the entire sample, while we expect others to affect each subsample independently.

All simulation-calibrated quantities x come with an estimate $\hat{x}$ and at least one source of uncertainty Δx on this estimate. Instead of applying a prior $\mathcal{N}(\hat{x}, (\Delta x)^2)$ on x, we write $x = \hat{x} + \delta\Delta x$ and leave δ as a free parameter in our MCMC chain with a prior $\sim\mathcal{N}(0, 1)$. We describe this in detail below for the weak lensing bias and local sources of scatter.

5.2.1 *Weak lensing bias*

We model the weak lensing bias as two independent components: mass model and measurement systematics. We calibrate the amplitude of the bias due to mass modelling against numerical simulations and model the measurement systematics such that we expect zero bias. For our likelihood analysis, we parametrize the weak lensing bias as

$$b_{{\rm WL},i} = b_{{\rm sim}_i} + \delta_{\rm WL,bias}\,\Delta b_{{\rm mass\,model}_i} + \delta_i\,\Delta b_{{\rm shearcal},\,N(z)_i}, \quad i \in \{\text{Megacam, HST}\}, \tag{26}$$

where $b_{\rm sim}$ is the mean expected bias due to the mass modelling, $\Delta b_{\rm mass\,model}$ is the uncertainty in our calibration of $b_{\rm sim}$ and $\Delta b_{{\rm shearcal},\,N(z)}$ is the quadrature sum of the uncertainties in shear calibration and in the determination of the distribution of background galaxies; $\delta_{\rm WL,\,bias}$, $\delta_{\rm Megacam}$ and δ_{HST} are free parameters in

Table 7. Parameters and priors used in the scaling relation analysis. The weak lensing parametrization is such that the fit parameters rescale the expected central values and uncertainties.

Parameter	Prior	Parameter	Prior
SZE and Y_X			
$A_{\rm SZ}$	$1/A_{\rm SZ}$	$A_{Y_{\rm X}}$	$1/A_{Y_{\rm X}}$
$B_{\rm SZ}$	$\mathcal{N}(1.63, 0.1^2)$	$B_{Y_{\rm X}}$	const.
$C_{\rm SZ}$	const.	$C_{Y_{\rm X}}$	$\mathcal{N}(0.702, 0.351^2)$
$\sigma_{\ln\zeta}$	$\mathcal{N}(0.13, 0.13^2)$	$\sigma_{\ln Y_{\rm X}}$	$\mathcal{N}(0.12, 0.08^2)$
SZE and M_{gas}			
SZE as above		$A_{M_{\rm g}}$	$1/A_{M_{\rm g}}$
		$B_{M_{\rm g}}$	const.
		$C_{M_{\rm g}}$	$\mathcal{N}(0.0, 0.2^2)$
		$\sigma_{\ln M_{\rm g}}$	$\mathcal{N}(0.12, 0.08^2)$
Weak lensing systematics			
$\delta_{\rm WL,\,bias}$	$\mathcal{N}(0, 1)$		
$\delta_{\rm Megacam}$	$\mathcal{N}(0, 1)$		
δ_{HST}	$\mathcal{N}(0, 1)$		
$\delta_{\rm WL,\,scatter}$	$\mathcal{N}(0, 1)$		
$\sigma_{\rm WL,LSS_{Megacam}}/{\rm M}_\odot$	$\mathcal{N}(9\times10^{13}, 10^{26})$		
$\sigma_{\rm WL,LSS_{HST}}/{\rm M}_\odot$	$\mathcal{N}(8\times10^{13}, 10^{26})$		
Correlated scatter			
$\rho_{\rm SZ\text{-}X}$	$\mathcal{U}(-1, 1)$		
$\rho_{\rm SZ\text{-}WL}$	$\mathcal{U}(-1, 1)$		
$\rho_{\rm WL\text{-}X}$	$\mathcal{U}(-1, 1)$		
Equation (29)	$\det(\Sigma) > 0$		
Cosmology			
$[\Omega_{\rm m}, \sigma_8]$	$\mathcal{N}\left([0.291, 0.783], \begin{pmatrix} 0.0016 & -0.0010 \\ -0.0010 & 0.0013 \end{pmatrix}\right)$		
$H_0/({\rm km\,s^{-1}\,Mpc^{-1}})$	$\mathcal{N}(73.8, 2.4^2)$		

our likelihood. With this parametrization, we put Gaussian priors of unit width centred at zero $\mathcal{N}(0,\ 1)$ on the three parameters $\delta_{\rm WL,\,bias}$, $\delta_{\rm Megacam}$ and δ_{HST}. We investigate a possible redshift dependence of $b_{{\rm sim}_i}$ and $\Delta b_{\rm massmodel}$ and find no indications for it, so we treat these terms as redshift-independent.

Due to the different observing strategies for the Megacam and *HST* samples, the mean expected biases $b_{{\rm sim}_i}$ are determined for each sample separately. The uncertainty in the mass model $\Delta b_{{\rm mass\,model}_i}$ is modelled as the quadrature sum of the uncertainty obtained from the numerical simulations, the uncertainty in the $M-c$ relation and the uncertainty due to miscentring. These uncertainties are determined in identical ways for both subsamples (although the numbers differ) and so we adopt a common fit parameter $\delta_{\rm WL,\,bias}$. This effectively correlates the uncertainties due to mass modelling between both samples. The shear calibration and determination of the distribution of background galaxies, however, is independent for each sample and we therefore adopt a fit parameter $\delta_{\rm Megacam/\mathit{HST}}$ for each sample.

5.2.2 *Weak lensing scatter*

We decompose the weak lensing scatter into two components: uncorrelated LSS modelled by a normal distribution and scatter intrinsic to the NFW modelling of the lensing halo. The latter term includes scatter due to the miscentring distribution, halo triaxiality and correlated LSS. Our motivation for this approach is twofold. First, the simulations used to calibrate the bias and scatter in Section 4.4 are not full light cones and do not capture the entirety of projected large-scale structure. Second, these simulations indicate that this local scatter is well described, at least for our purposes, by a log-normal distribution, while uncorrelated LSS leads to an additional Gaussian scatter contribution to the tangential shear. We model the latter term as Gaussian scatter on the cluster mass, although this is not entirely correct, as the relation between cluster mass and shear is non-linear (see also Hoekstra 2003). The combination of log-normal local scatter and normal non-local scatter gives us enough flexibility to model the true mass scatter, which is also neither exactly normal nor log-normal.

We calibrate the local, log-normal scatter against simulations. The Megacam and *HST* samples have different scatter properties, but these numbers are calibrated against the same simulations and therefore share the same systematics. We use

$$\sigma_{{\rm local}_i} = \sigma_{{\rm sim}_i} + \delta_{\rm WL,scatter}\,\Delta\sigma_{{\rm sim}_i}\ , \quad i \in \{\text{Megacam, HST}\}, \qquad (27)$$

where $\Delta\sigma_{{\rm sim}_i}$ is the uncertainty of the simulation calibrated scatter $\sigma_{{\rm sim}_i}$ and $\delta_{\rm WL,\,scatter}$ is a free parameter in our likelihood, on which we apply a Gaussian prior $\mathcal{N}(0,\ 1)$.

We estimate the uncorrelated LSS contribution to the weak lensing scatter in our NFW fits of the Megacam data by calculating the variance of the surface mass density inside our fit aperture following the prescription presented in Hoekstra (2001). A key difference between our work and that of Hoekstra (2001) is that Hoekstra computes the variance inside an aperture for the aperture mass statistics, while we perform NFW fits to the shear profile. The aperture mass is the radially weighted average of the mass inside a cylinder, where the weight is given by a fixed filter function chosen by the user. To adapt the prescription of Hoekstra (2001) to our case, we weigh the surface mass density power spectrum by an NFW profile rep-

resenting the average mass and redshift of the Megacam cluster sample. For a cluster with $M_{200} = 8 \times 10^{14}\,\mathrm{M}_\odot$ at $z = 0.4$, we obtain $\sigma_{\mathrm{WL,LSS_{Megacam}}} = 9 \times 10^{13}\,\mathrm{M}_\odot$.

This value is close to and slightly larger than the average value reported for the *HST* clusters, $\sigma_{\mathrm{WL,LSS_{HST}}} = 8 \times 10^{13}\,\mathrm{M}_\odot$. This may seem suprising at first, because the lensing catalogue of the *HST* is much deeper than the Megacam data and consequently integrates over more large-scale structure. The apertures employed in the lower redshift Megacam sample are, however, larger than in the *HST* sample. This more than compensates for our shallower redshift distribution.

In our analysis, we only use the mean value $\sigma_{\mathrm{WL,LSS_{HST}}} = 8 \times 10^{13}\,\mathrm{M}_\odot$ of the LSS scatter values reported in S18 as the mean of a Gaussian prior rather than an individual prior for each cluster. This reduces computational complexity and the impact on our analysis is negligible, because the various sources of scatter are (almost) fully degenerate, so that tiny deviations from reality in one scatter term are easily absorbed by another. The Gaussian prior for $\sigma_{\mathrm{Megacam}_i}$ is centred on the value computed above. Both priors have a standard deviation of $\Delta\sigma_{\mathrm{WL,LSS}_i} = 10^{13}\,\mathrm{M}_\odot$, based on the estimated scatter of $\sigma_{\mathrm{WL,LSS_{HST}}}$ (S18).

As mentioned in Section 4.4, the bias b_{WL} and scatter $\sigma_{\mathrm{WL,\,local}}$ depend on the miscentring model one adopts. In general, the centroid of the X-ray emission of the intracluster medium is expected to be a more reliable indicator of the true cluster centre than the SZE peak position based on observations with a relatively broad beam. The *HST* sample has X-ray data for all clusters and thus we choose the X-ray positions and their corresponding bias and scatter values from S18 as input to our analysis. The Megacam sample is not fully covered by *Chandra* data. For these data, we take the SZE peak position as the cluster centre.

5.3 Likelihood function and analysis pipeline

We simultaneously constrain the SZE and X-ray scaling relations (four parameters each) and the weak lensing model (six parameters) using an extension of the framework described in Bocquet et al. (2015). We summarize the main points of their likelihood function and discuss our extensions. All fit parameters are also listed in Table 7.

The translation of the weak lensing observable, i.e. the reduced shear g_t, into a physical mass scale depends on the cosmological parameters in a number of ways. First, the critical density of the Universe at the cluster redshift enters the NFW profile. Second, the translation of the angular shear profile into a radial shear profile measured in physical distances depends on the distance–redshift relation. Similarly, the distance–redshift relation enters the computation of the critical surface mass density, equation (2). Finally, while many mass–concentration relations are, strictly speaking, valid only for the cosmological parameters for which they were derived, the *M*–*c* relation of Diemer & Kravtsov (2015) we employ has an explicit cosmological dependence.

We use the observed reduced shear with cluster contamination correction applied, $g_{\mathrm{t,\,cor}}$, and redshift distribution $N(z)$ as input to the weak lensing portion of the likelihood code, which then computes $\langle\beta\rangle$, $\langle\beta^2\rangle$ and fits an NFW profile as described in Section 4.2 at every sample point of the MCMC chain. In this way, the cosmology dependence of the NFW shear profile due to the evolution of the critical density with redshift and the redshift–distance relation are taken into account.

Our cluster sample is SZE-selected. To take selection effects properly into account, for each cluster i in our sample, we evaluate the likelihood

$$P(\mathrm{X}_i, M_{\mathrm{WL}_i}|\xi_i, z_i, p) = \left[\int\int \mathrm{d}M\mathrm{d}\zeta\, P(\xi|\zeta)P(\mathrm{X}_i, M_{\mathrm{WL}_i}, \zeta|M, z_i, p)P(M|z_i, p)\right]\Big|_{\xi_i}, \tag{28}$$

where, for simplicity, we denote the X-ray observable as X, $P(M|z, p)$ is the halo mass function at redshift z and p is the vector of cosmological and scaling relation parameters. The multiplication with the halo mass function is a necessary step to account for the Eddington bias.

The term $P(\mathrm{X}, M_{\mathrm{WL}}, \zeta|M, z, p)$ contains the mass–observable relations defined in Section 5.1 as well as the intrinsic scatter about each relation. Extending the original analysis framework (Bocquet et al. 2015), we allow for correlated scatter between all observables. Namely, $\sigma_{\ln\zeta}$, $\sigma_{\ln Y_\mathrm{X}}$ and σ_{WL} are linked by correlation coefficients $\rho_{\mathrm{SZ\text{-}Y}}$, $\rho_{\mathrm{SZ\text{-}WL}}$ and $\rho_{\mathrm{WL\text{-}Y}}$, so that the intrinsic covariance matrix is

$$\Sigma_Y = \begin{pmatrix} \sigma^2_{\ln\zeta} & \sigma_{\ln\zeta}\sigma_{\ln Y_\mathrm{X}} & \sigma_{\ln\zeta}\sigma_{\mathrm{WL,local}} \\ \sigma_{\ln\zeta}\sigma_{\ln Y_\mathrm{X}} & \sigma^2_{\ln Y_\mathrm{X}} & \sigma_{\ln Y_\mathrm{X}}\sigma_{\mathrm{WL,local}} \\ \sigma_{\ln\zeta}\sigma_{\mathrm{WL,local}} & \sigma_{\ln Y_\mathrm{X}}\sigma_{\mathrm{WL,local}} & \sigma^2_{\mathrm{WL,local}} \end{pmatrix} \circ \begin{pmatrix} 1 & \rho_{\mathrm{SZ-X}} & \rho_{\mathrm{SZ-WL}} \\ \rho_{\mathrm{SZ-X}} & 1 & \rho_{\mathrm{WL-Y}} \\ \rho_{\mathrm{SZ-WL}} & \rho_{\mathrm{WL-Y}} & 1 \end{pmatrix}, \tag{29}$$

and equivalently for $\sigma_{\ln M_\mathrm{g}}$. We put flat priors allowing the full range from -1 to 1 on all three correlation coefficients, with the additional restriction that the combination of all three coefficients must be physically allowed. In practice, we compute equation (28) on a three-dimensional grid in X, M_{WL} and ζ. To make this step computationally efficient, we (i) choose an optimal range in mass for each cluster, informed by its measured SZE, X-ray and weak lensing signals and the current set of scaling parameters p, to avoid parts of the observable space with effectively zero probability, (ii) only perform this three-dimensional computation for clusters that actually have all three measurements, otherwise computing the (cheaper) two-dimensional version, and (iii) employ a fast Fourier transform convolution.

For each cluster in our sample, we compare the predicted $P(\mathrm{X}, M_{\mathrm{WL}}|\xi, z, p)$ with the actual measurement and extract the probability of consistency.

The X-ray measurements consist of a radial profile $M_{\mathrm{gas}}(r)$ and a global temperature measurement T_X, from which a $Y_\mathrm{X}(r) = M_{\mathrm{gas}}(r)T_\mathrm{X}$ profile is computed. To account for the radial dependence of the measurement and the variation of the modelled r_{500} throughout the parameter space, we define a fiducial radius r^{fid}_{500} for each cluster and evaluate the likelihood at this radius. Note that the X-ray scaling relations equations (24) and (25) predict the X-ray quantity at r_{500}. To translate this model prediction to r^{fid}_{500}, we use the fact that the X-ray profile can be well approximated by a power law in radius (see also Mantz et al. 2016). With this, the prediction at r^{fid}_{500} becomes

$$Y_\mathrm{X}(r^{\mathrm{fid}}_{500}) \equiv \left(\frac{r^{\mathrm{fid}}_{500}}{r_{500}}\right)^{\mathrm{slope}} Y_\mathrm{X}(M_{500}, z, p), \tag{30}$$

where r_{500} is derived from M_{500}. The measurement uncertainty in $Y_\mathrm{X}(r^{\mathrm{fid}}_{500})$ is captured by a log-normal distribution.

For the weak lensing data, we forward-model from M_{WL} to the observed reduced shear g_t: we convolve $P(M_{\mathrm{WL}})$ with the Gaussian LSS noise (Section 5.2.2) and then compute the reduced shear

$g_t(M_{WL}, r_j)$ for each radial bin r_j following equations (1)–(3). Finally, for each radial bin, we compute the likelihood of the measurement given $g_t(r_j)$ and multiply the likelihoods of all bins.

Ultimately, we sum the log-likelihoods for all clusters. The total likelihood function (up to an additive constant) then is

$$\ln \mathcal{L} = \sum_{i=1}^{N_{cl}} \ln P(\mathrm{X}_i, g_{t_i} | \xi_i, z_i, p) \,, \tag{31}$$

where i runs over all clusters. It is important to note that the measured cluster abundance does not enter the likelihood function or analysis in this work.

We emphasize that all sample selection effects are accounted for in our likelihood framework. Equation (31) is only evaluated for clusters that pass the SZE and redshift selection functions ($\xi > 5$ and $z > 0.25$) and the target selection of the follow-up observations (X-ray and weak lensing) is not based on these observables themselves (e.g. X-ray properties or weak lensing strength). Obviously, one must not reject follow-up observations that did not lead to a detection of the cluster. This frequently happens in the weak lensing observable due to its large scatter. Our forward modelling approach naturally deals with clusters with radial shear profile consistent with zero or less.

We use the EMCEE (Foreman-Mackey et al. 2013) implementation of the affine-invariant ensemble sampler algorithm (Goodman & Weare 2010) to evaluate the likelihood function of equation (31). We use an ensemble of 192 walkers and discard the first five autocorrelation lengths of the chain as burn-in period. We consider chains to be converged if no evolution of the mean and standard deviation is visible in trace plots and if the Gelman & Rubin (1992) criterion is $\hat{R} < 1.1$ for all parameters.

5.4 Test on mock catalogues

We test that our implementation of the calibration framework described above recovers unbiased parameter estimates using mock galaxy cluster catalogues. These are created by Poisson-sampling the halo mass function over the redshift range of the SPT–SZ cluster sample. The SZE detection significance $\ln \zeta$ and the follow-up quantities Y_X and weak lensing mass are drawn together from a multivariate normal distribution according to the fiducial scaling relation parameters, including the full intrinsic covariance matrix of equation (29). NFW shear profiles are generated from the weak lensing mass set in this way. In the mock catalogues, we select the 80 most significant clusters to have Y_X as follow-up observable. For the weak lensing follow-up, we either select the 19 most significant clusters or randomly sample from all significances. In this way, we also verify the independence of the recovered scaling relation parameters for the follow-up strategy.

We generate mock catalogues for an SPT–SZ-like 2500 deg^2 survey and for a survey 10 times the size of the actual SPT–SZ survey. For all cases, we recover the input scaling relations within 1σ uncertainty. Additionally, these mock catalogues allow for predictions about which parameters our data set will be able to constrain and choose appropriate priors for those parameters where the information content is too low to give meaningful constraints.

5.5 Choices of priors

In analysing the scaling relations described above, we aim to put informative priors only on parameters that our data cannot constrain. In addition to testing the constraining power of our data by running Monte Carlo chains with different prior choices, we also create mock realizations of the SPT + X-ray + weak lensing catalogues to ascertain that the real data behave as expected from these simulations.

The weak lensing bias b_{WL} and the overall scaling of the per cluster bias factors of the S18 samples are obviously fully degenerate with the normalizations of the scaling relations we aim to constrain. We therefore put Gaussian priors with widths corresponding to the uncertainties obtained from the calibration with simulations on them. Also, the various sources of intrinsic scatter cannot be disentangled by our analysis and we fix them using Gaussian priors.

Putting non-informative priors on the mass slopes B_{SZ}, B_{Y_X} and B_{M_g} and the redshift evolution coefficients C_{SZ}, C_{Y_X} and C_{M_g}, we learn that our data are not able to obtain meaningful constraints for these parameters. Our mock catalogues confirm that – given the current data set – we should not expect to be able to constrain these parameters. We therefore choose a Gaussian prior $B_{SZ} \sim \mathcal{N}(1.63, 0.1^2)$. The mean and central values are determined by running a full cosmological analysis of the SPT cosmology sample plus the weak lensing data sets, similarly to what was done in the recent SPT–SZ cosmology analysis (de Haan et al. 2016). Using the cluster number count data, we constrain the mass slope B_{SZ}, obtaining the values 1.63 and 0.1 for its mean and uncertainty, respectively. We choose to put flat priors on B_{Y_X}/B_{M_g} because these are constrained through their degeneracy with B_{SZ} once B_{SZ} is fixed.

We use a prior $C_{Y_X} \sim \mathcal{N}(0.70, 0.35^2)$ to encode our belief that the X-ray gas in clusters evolves (approximately) self-similarly. These values correspond to the self-similar exponent $-2/5$ in the form of the Y_X–mass relation chosen by Vikhlinin et al. (2009) and allow for 50 per cent scatter around self-similarity. We put a flat prior on C_{SZ} because it is degenerate with C_{Y_X}. Likewise, for the M_{gas} scaling relation, we assume no redshift evolution with the same uncertainty as for C_{Y_X}, i.e. we set $C_{M_g} \sim \mathcal{N}(0, 0.2^2)$.

This leaves the normalizations A_{SZ}, A_{Y_X} and A_{M_g} to be determined. Because these are the parameters we are chiefly interested in, we put non-informative priors on them. Specifically, because the scaling relations are linear in log space and the non-informative prior on the intercept of a line is flat, the non-informative prior for the normalization of a power law is proportional to $1/A_i$, $i \in \{\mathrm{SZ}, \mathrm{Y}, \mathrm{M}\}$.

Finally, we note that the scaling relation parameters are mildly cosmology-dependent. This is due to the distance–redshift relation, as well as the critical density at a given redshift being dependent on cosmology. In our analysis, we marginalize over the uncertainty of the parameters most affecting these two quantities: Ω_m, σ_8 and H_0. For the first two, our prior is a bivariate Gaussian describing the degeneracy between these parameters, based on the posterior probability distribution of the cosmology chain of de Haan et al. (2016). For the Hubble constant, we choose the Riess et al. (2011) value of $H_0 = (73.8 \pm 2.4)\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ as our prior, which was also utilized in de Haan et al. (2016). We list all priors in Table 7.

6 RESULTS AND DISCUSSION

We show parameter constraints for the Y_X and M_{gas} analyses in Figs 7 and 8, respectively, and summarize the best-fitting scaling relation parameters and their 68 per cent credible intervals in Table 8. Our key results are the normalizations of the mass–SZE and mass–X-ray scaling relations, which affect the systematic uncertainty limits of the SPT cluster cosmology results directly (Bocquet et al., in preparation; de Haan et al. 2016). The best-fitting A_{SZ} values of the Y_X and M_{gas} chain are almost identical to each other at $A_{SZ} =$

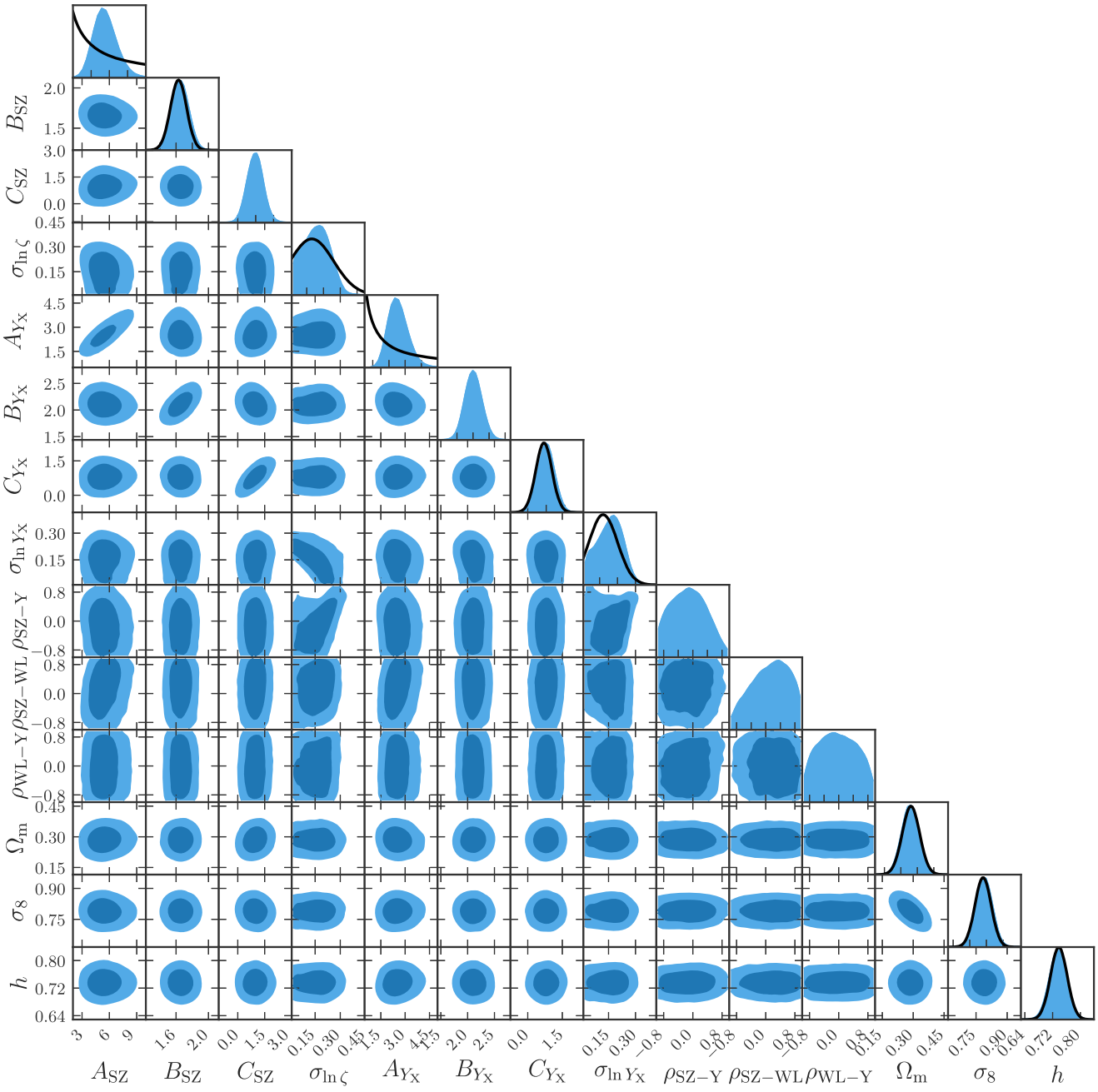


Figure 7. Parameter constraints for the SZE and $Y_{\rm X}$ scaling relation parameters. Solid black lines are the priors imposed on parameters (see Section 5.5). We show here the correlated scatter coefficients and the cosmological parameters varied within the prior ranges (see Table 7) and omit the lensing nuisance parameters due to space constraints. They are shown for the $M_{\rm gas}$ scaling relations analysis in Fig. 8 and are virtually identical to the ones omitted here.

$5.56^{+0.96}_{-1.35}$ and $A_{\rm SZ} = 5.57^{+0.90}_{-1.41}$, as one would expect, because these numbers are essentially set by the weak lensing calibration. We will discuss these results in detail below.

For the mass–SZE scaling relation, a comparison with earlier works is best illustrated by looking at the probability distribution of the mass of a typical SPT–SZ-selected cluster (Fig. 9). Our measurement of $A_{\rm SZ} = 5.56^{+0.96}_{-1.35}$ is in agreement with both the simulation-based prior of $A_{\rm SZ} = 6.01 \pm 1.80$ used in early SPT–SZ work (Vanderlinde et al. (2010), who used N-body simulations and a gas model from Shaw et al. (2009)) and the updated prior $A_{\rm SZ} = 5.38 \pm 1.61$ based on the COSMO–OWLS hydrodynamic simulations (Le Brun et al. 2014) and used in the latest SPT–SZ cluster cosmology analysis (de Haan et al. 2016).

We also compare our value of $A_{\rm SZ}$ with normalizations obtained from data in other works. Outside the SPT collaboration, Gruen et al. (2014) measured weak lensing masses of SPT and *Planck*-selected galaxy clusters using the Canada–France–Hawaii Telescope Legacy Survey and pointed follow-up observations using WFI at the 2.2-m ESO/MPG telescope. Their $A_{\rm SZ} = 6.0^{+1.9}_{-1.8}$ is in excellent agreement with ours. Gruen et al. (2014) find a slightly shallower mass slope ($B_{\rm SZ} = 1.25^{+0.36}_{-0.28}$) than we adopt from the 2500 deg^2 SPT–SZ cosmology analysis (de Haan et al. 2016) and more in line with the

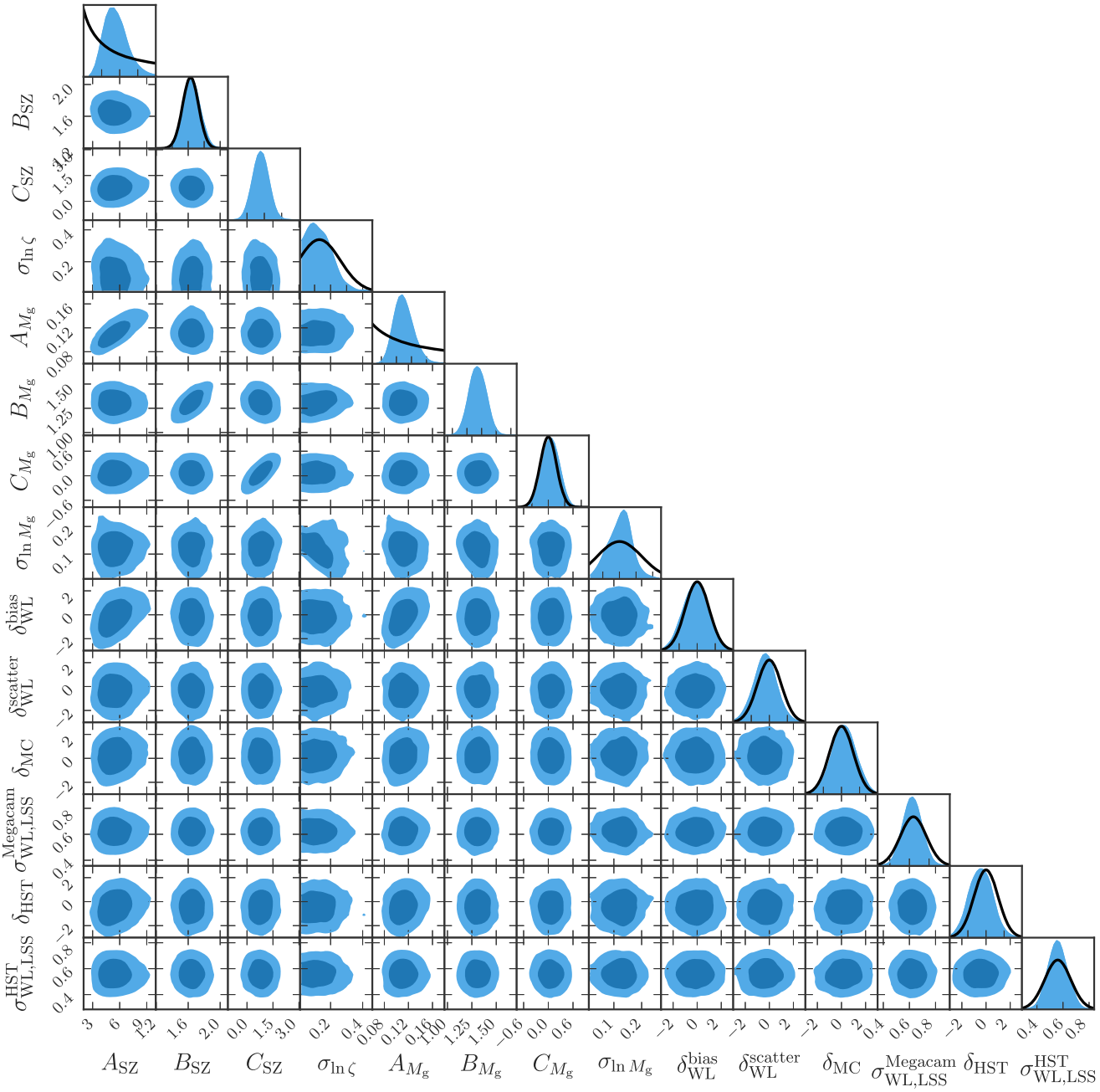


Figure 8. Same as Fig. 7 for the SZE and $M_{\rm gas}$ scaling relations. Here we show the lensing nuisance parameters omitted from Fig. 7 and omit the correlation coefficients of the scatter and the cosmological parameters instead.

expectation from simulations. Our pivot points are, however, identical, so that we can compare normalizations directly, except for a slight mismatch in $C_{\rm SZ}$, which was also held fixed in their analysis but at a value of $C_{\rm SZ} = 0.83$, which is about 1σ below our value.

Our normalization of the mass–SZE scaling relation is also in good agreement with earlier SPT work (Bocquet et al. 2015; de Haan et al. 2016). Visually, the largest disagreement is with the SPT cluster cosmology analysis of Bocquet et al. (2015), when it is combined with the first release of the *Planck* primary CMB cosmology results (Planck Collaboration et al. 2014). The combination of the velocity-dispersion-based MOR normalization constraints of Bocquet et al. (2015) with CMB data leads to a shift in $\Omega_{\rm m}$–σ_8 orthogonal to the cluster SPT cosmology constraints on these parameters. As a result the normalization of the mass–SZE MOR shifts accordingly, to account for the implied different cluster mass scale, leading to the difference seen in Fig. 9.

For a quantitative comparison, we follow Bocquet et al. (2015) to compute the significance of the difference of two distributions. We randomly draw points from two distributions and compute the difference δ between pairs of points. We use this to estimate the probability distribution P_δ of these differences and compute the likelihood that zero is within this distribution. Assuming a normal distribution, this likelihood is then converted to a significance. The lower normalization parameter $A_{\rm SZ}$, corresponding to higher cluster masses, inferred from a joint cosmological analysis of the SPT

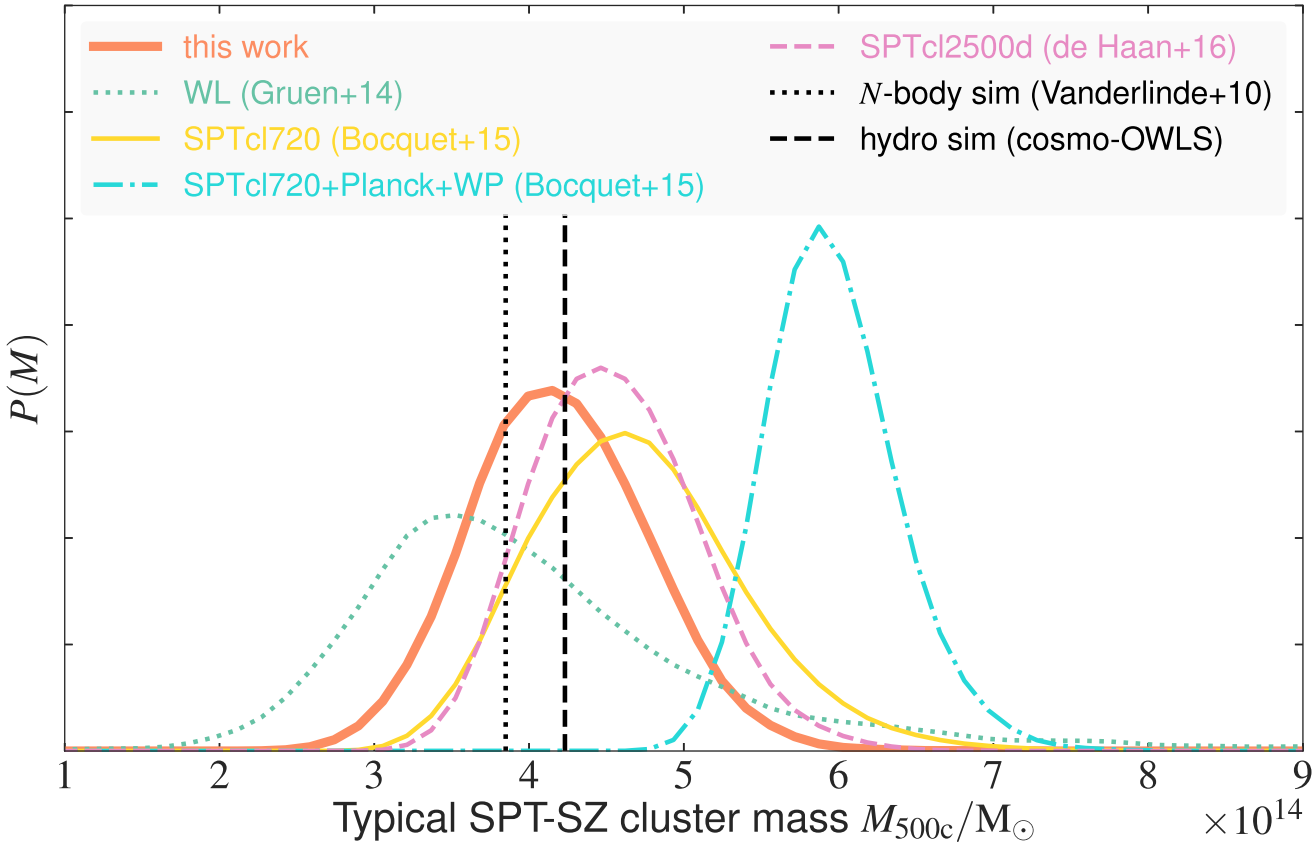


Figure 9. The probability distribution of the mass M_{500} of a typical (median) SPT cluster with $\xi = 6.5$ at $z = 0.5$ according to different mass calibration efforts. The vertical lines correspond to the predictions from simulations in Vanderlinde et al. (2010, dotted line) and the COSMO–OWLS simulation (dashed line; Le Brun et al. 2014). The mass scale in this work agrees equally well with both simulation predictions.

Table 8. Marginalized scaling relation parameter constraints for the ζ–M_{500} scaling relation and the Y_X–M_{500} scaling relation (top half) and the M_{gas}–M_{500} scaling relation (bottom half). The values reported are the mean of the posterior and the shortest 68 per cent credible interval.

Parameter	Value	Parameter	Value
A_{SZ}	$5.56^{+0.96}_{-1.35}$	A_{Y_X}	$2.57^{+0.44}_{-0.67}$
B_{SZ}	$1.656^{+0.092}_{-0.101}$	B_{Y_X}	$2.11^{+0.14}_{-0.16}$
C_{SZ}	$0.96^{+0.41}_{-0.43}$	C_{Y_X}	$0.80^{+0.33}_{-0.35}$
$\sigma_{\ln\zeta}$	$0.155^{+0.084}_{-0.079}$	$\sigma_{\ln Y_X}$	$0.154^{+0.083}_{-0.065}$
A_{SZ}	$5.57^{+0.90}_{-1.41}$	A_{M_g}	$0.112^{+0.012}_{-0.017}$
B_{SZ}	$1.648^{+0.094}_{-0.103}$	B_{M_g}	$1.310^{+0.080}_{-0.084}$
C_{SZ}	0.79 ± 0.43	C_{M_g}	$0.06^{+0.19}_{-0.20}$
$\sigma_{\ln\zeta}$	$0.131^{+0.053}_{-0.100}$	$\sigma_{\ln M_g}$	$0.120^{+0.044}_{-0.039}$

cluster sample and *Planck* CMB data sets (Bocquet et al. 2015) disagrees with our result at the 2.6σ level.

We emphasize that the change in normalization in Bocquet et al. (2015, yellow to cyan line in Fig. 9) when the underlying cosmology shifts is caused by the self-calibration of the scaling relations from cluster number counts. In this work, we adopt a cosmology with Ω_m, σ_8 and H_0 close to the results of Bocquet et al. (2015) *without* the *Planck* data added. Because we use the cluster mass function only for Eddington bias correction and not for self-calibration of the MOR, small changes in the cosmological parameters do not have any big impact on our recovered normalization A_{SZ}. In particular, changing the cosmological parameters to the ones obtained from SPT clusters with *Planck* data (Bocquet et al. 2015) changes our normalization by less than 1 per cent and does not bring it into better agreement with their lower A_{SZ} value.

Also used in the SPT–SZ cosmology analysis is the mass–Y_X scaling relation. As for the mass–ζ relation, our marginalized posterior for the normalization is in very good agreement with the prior utilized in the cosmology analysis (de Haan et al. 2016). This is an important result, as the prior was based on an *external calibration* of the normalization of the mass–Y_X scaling relation, namely the Vikhlinin et al. (2009) scaling relation updated with the weak lensing mass calibration of the Weighing the Giants (WtG) and CCCP projects (von der Linden et al. 2014; Applegate et al. 2014; Hoekstra et al. 2015). We are now able to confirm that these priors were appropriate for cosmology analysis based on an *internal calibration*. Fig. 10shows a comparison of the marginalized and joint posterior probability distributions for the normalizations A_{SZ} and A_{Y_X} in comparison with the results obtained by de Haan et al. (2016) and the priors used in this previous SPT work. Our posterior distributions are a little broader than theirs and consequently we do not yet expect that our mass calibration efforts will lead to tighter cosmological constraints with the current data set (Bocquet et al., in preparation). We note, however, that the width of the A_{SZ} posterior distributions of de Haan et al. (2016) is narrower than their prior range. This indicates that their constraint on the MOR normalization benefits from self-calibration. We do not use this self-calibration from the number counts of galaxy clusters (see Section 5.3) and thus obtain broader posterior distributions, given

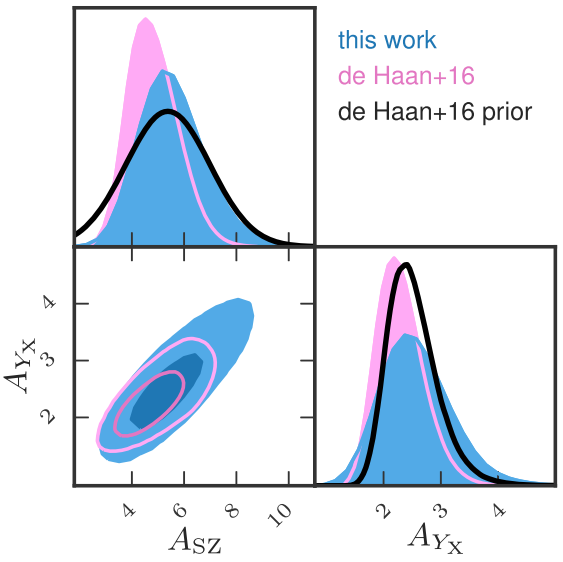


Figure 10. Comparison of our posterior probability distributions of the normalizations of the SZE and Y_X scaling relations with the posterior and priors of de Haan et al. (2016) converted to our parametrization of the mass–Y_X scaling relation (24).

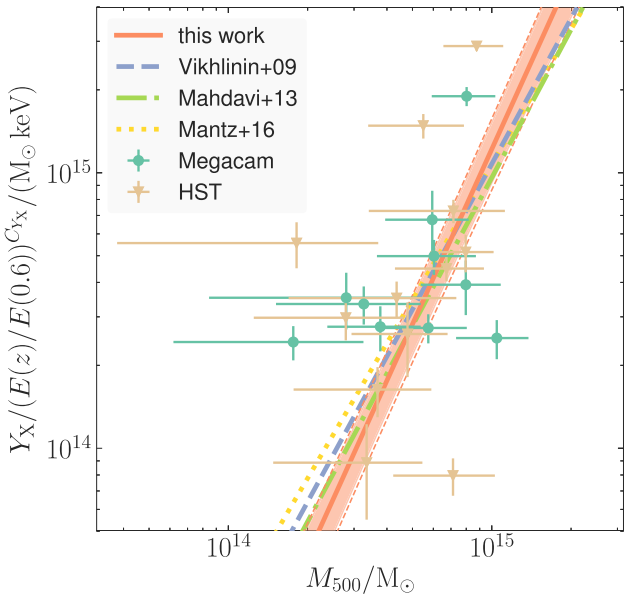


Figure 11. Y_X–weak lensing mass scaling relation. Our result is shown in red, while the Y_X mass relations of Vikhlinin et al. (2009), Mahdavi et al. (2013) and Mantz et al. (2016) are shown in blue, green and yellow, respectively, for comparison. The Y_X values are evaluated using the reference cosmology and best-fitting scaling relation parameters. We extract values for M_{WL} and Y_X as described in the text. The masses on the x-axis are debiased according to equation (19). We note that the values displayed here are only used for illustration purposes; the analysis pipeline does not use these values, but follows the forward modelling approach described in Section 5.3. The horizontal error bars take only the shape noise component into account. The vertical error bars also only show the observational error. The 19 clusters observed with Megacam are shown as green circles, while the 13 clusters observed with *HST* (Schrabback et al. 2018) are shown as light brown triangles. Shaded regions indicate the uncertainty in our best-fitting parameters. The dashed lines indicate the best-fitting intrinsic scatter added in quadrature to the parameter uncertainties. We show the intrinsic scatter only for our result; its contribution to the other scaling relations is almost exactly the same.

the still relatively small sample of SPT clusters with weak lensing information.

The mass-slope and redshift evolution parameters follow their prior probability distribution in the observable on which an informative prior was imposed. The B_{Y_X} constraint is determined purely by the B_{SZ}–B_{Y_X} degeneracy. Likewise, the C_{SZ} constraint is governed by the C_{SZ}–$C_{Y_X}(C_{M_g})$ degeneracy. The C_{SZ} values derived in this way differ by 0.3σ between the Y_X ($C_{SZ} = 0.96^{+0.41}_{-0.43}$) and M_{gas} ($C_{SZ} = 0.79 \pm 0.43$) chains. Both values are higher than the C_{SZ} prior in de Haan et al. (2016), but even the higher $C_{SZ}(Y_X)$ value deviates by only 0.5σ. All C_{SZ} posteriors of de Haan et al. (2016) agree with our values to better than 1σ.

Our modelling of the weak lensing bias and scatter (Section 5.1) introduces numerous nuisance parameters that we are not able to constrain with the data. They all follow the priors. Similarly, we are not able to distinguish between various sources of scatter in our data. As the degeneracy between $\sigma_{\ln\zeta}$ and $\sigma_{\ln Y_X}$ ($\sigma_{\ln M_g}$) shows, we are only able to put limits on the sum of their squares, i.e. the total scatter of the scaling relations.

It is expected from numerical simulations that the intrinsic scatter of the weak lensing and SZE measurements is correlated (e.g. Shirasaki, Nagai & Lau 2016). For the current data, however, we cannot constrain any of the three correlation coefficients. Furthermore, for all parameters the constraints obtained by leaving the correlation coefficients free are indistinguishable from those where we set all correlation coefficients to zero.

In Figs 11–13, we show the scaling relations (23)–(25) with marginalized uncertainties in comparison with the data points. In these plots, green circles indicate the 19 clusters followed up with Megacam, while light brown triangles are the 13 *HST* observations from S18. In all of these figures, the distributions of the Megacam and *HST* data points appear to be consistent with each other. This visual impression is confirmed by finding consistent normalizations of the scaling relations when the sample is split in redshift at $z = 0.6$ (Bocquet et al., in preparation).

Note that we do not observe the weak lensing mass M_{wl} directly and that the X-ray observables are radial profiles and not scalar quantities. In the following, we describe briefly how we extracted the quantities displayed in the figures. The X-ray measurements consist of a temperature measurement of the hot intracluster medium (ICM) and a radial gas mass profile $M_{gas}(r)$. Both quantities can be combined to give the radial $Y_X(r)$ profile. In principle, the temperature also varies radially, but this is slow enough to be approximated accurately by a global average temperature. In the case of M_{gas}, the scaling relation (25) relates the gas mass to the cluster total mass M_{500}, from which the radius r_{500} can be uniquely determined. Assuming the best-fitting scaling relation parameters for the M_{gas}–mass relation, we can now solve for M_{gas} by solving the implicit equation $M_{gas}^{data}(r) = M_{gas}^{MOR}(r)$. We can then obtain the mean and standard deviation of the recovered distribution in M_{gas} for Figs 12 and 14. The same procedure is used for Y_X and equation (24) and Fig. 11.

Our likelihood framework also never computes a weak lensing mass that fits the observed radial shear profile $g_t(r)$ best. Instead it computes how probable we are to find the observed shear profile, given the mass predicted from the scaling relations. Nevertheless, to be able to plot weak lensing masses, we perform maximum-likelihood fits to the contamination-corrected shear profiles and use their location and uncertainty when plotting weak lensing masses.

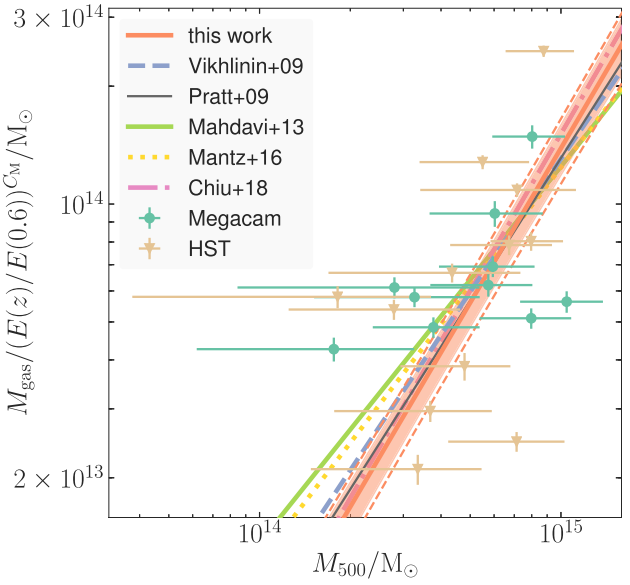


Figure 12. Same as Fig. 11, but for the $M_{\rm gas}$–mass scaling relation. We compare our best-fitting relation, where the slope was set by a galaxy cluster number count analysis (de Haan et al. 2016), with those of Vikhlinin et al. (2009), Pratt et al. (2009), Mahdavi et al. (2013), the WtG team (Mantz et al. 2016) and Chiu et al. (2018).

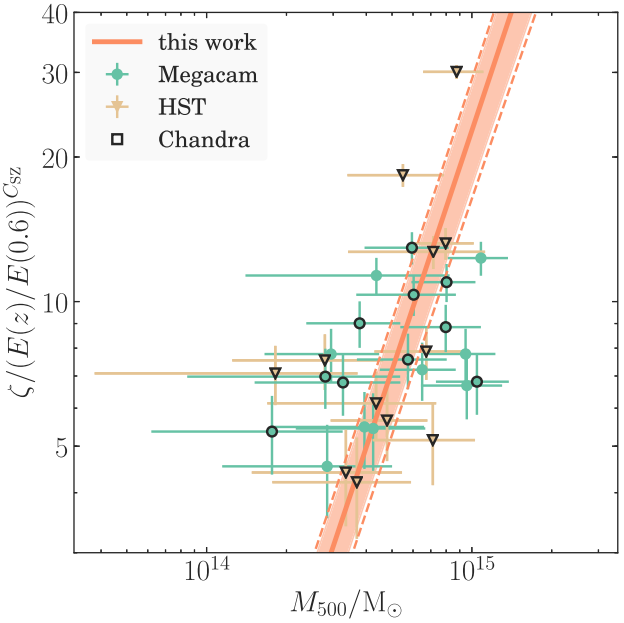


Figure 13. The ζ–mass scaling relation and estimates $\hat{\zeta}$ and M_{500} for the 32 clusters with weak lensing data. Points marked in black are clusters with *Chandra* X-ray data used in the scaling relation analysis, i.e. all clusters shown in Figs 11 and 12.

Furthermore, when we plot cluster data points in ζ–mass scaling relations we also need an estimate of ζ for each cluster. We obtain this from the observable SNR ξ via

$$\hat{\zeta} = \sqrt{\xi^2 - 3}/f_{\rm field} \ , \tag{32}$$

where $f_{\rm field}$ is a scaling factor to correct for the different depths of fields in the 2500 deg^2 SPT–SZ survey.

Figs 11–14 show the predicted scaling relations for the underlying cluster population and are not corrected for our SZE selection. This is most obvious in Fig. 14, where the two low-scatter mass proxies $\hat{\zeta}$ and $M_{\rm gas}$ are plotted against each other for a cluster population selected in ξ. The Eddington bias is clearly visible in the lower left corner of this plot from the points falling below the best-fitting line, i.e. they are preferentially scattered towards higher $\hat{\zeta}$. We remind the reader that the scaling relation analysis takes this bias into account through the shape of the mass function and the SPT cluster selection function. The scaling relation plotted in Fig. 14 is obtained by combining equations (23) and (25) into

$$\frac{M_{\rm gas}}{5 \times 10^{14}\,\mathrm{M}_\odot} = A_{M_{\rm g}} \left(\frac{6}{7}\right)^{B_{M_{\rm g}}} \left(\frac{\zeta}{A_{\rm SZ}}\right)^{B_{M_{\rm g}}/B_{\rm SZ}} \tag{33}$$

and omitting the redshift evolution terms, because they are taken care of when the plotted data are rescaled to a common redshift.

Our estimates for the normalizations of the X-ray scaling relations show good agreement with previous studies (Vikhlinin et al. 2009; Pratt et al. 2009; Mahdavi et al. 2013; Mantz et al. 2016). For the mass–$Y_{\rm X}$ relation, this holds over the entire mass range under investigation here. For the mass–$M_{\rm gas}$ relation, the sometimes significantly different slopes lead to good agreement only in the vicinity of our pivot point $M_{\rm p} = 5 \times 10^{14}\,h_{70}^{-1}\,\mathrm{M}_\odot$ and marginal discrepancies at the extreme ends of the mass range under investigation here. This is particularly obvious for the relations of Mahdavi et al. (2013), who find a slope slightly smaller than but consistent with self-similarity, and Mantz et al. (2016), whose slope is very nearly exactly self-similar. However, at our pivot $M_{500}^{\rm piv} = 5 \times 10^{14}\,\mathrm{M}_\odot$ we agree with all cited studies, within our mutual uncertainties.

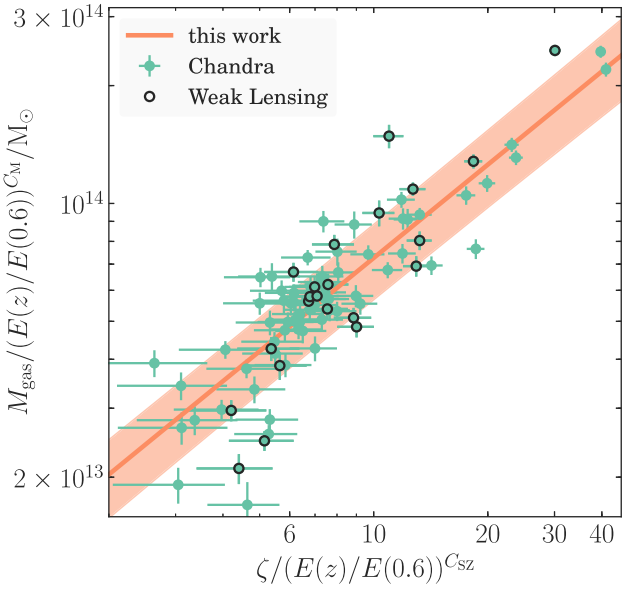


Figure 14. The derived ζ–$M_{\rm gas}$ scaling relation (equation 33). Cluster data points show the redshift evolution corrected estimate $\hat{\zeta}$ (equation 32). We show only the parameter uncertainty and not the intrinsic scatter for this relation.

We note again that we are not able to constrain the slope $B_{M_{\rm g}}$ with our present data set. Rather, our value for the slope is determined by the prior we put on $B_{\rm SZ}$ – based on the cosmology analysis of de Haan et al. (2016) – and the degeneracy between $B_{\rm SZ}$ and $B_{M_{\rm g}}$. Future weak lensing analyses of SPT-selected clusters covering a wider ξ and thus mass range will enable us to constrain the slope directly from weak lensing observations, instead of only through

self-calibration in a cosmological framework, as in de Haan et al. (2016) and Mantz et al. (2016).

In Fig. 13, we show the scaling relation between cluster mass and debiased SPT detection significance ζ . In this plot, we also highlight those clusters with *Chandra* data used in the scaling relation analysis. We find no indication that the 10 clusters from the Megacam sample without X-ray follow-up come from a different population.

Finally, we compare our mass estimate for the stack of all 19 Megacam clusters with that of a previous study using gravitational magnification instead of shear (Chiu et al. 2016b, who found a mass estimate of $M_{500} = (5.37 \pm 1.56) \times 10^{14}\,\mathrm{M}_{\odot}$). This is in very good agreement with our weighted mean mass $M_{500} = (5.96 \pm 0.61) \times 10^{14}\,\mathrm{M}_{\odot}$ for these clusters.

7 CONCLUSION

In this article we describe the observations and weak lensing analysis of 19 clusters from the 2500 deg^2 SPT–SZ survey. We pay particular attention to controlling systematic uncertainties in the weak lensing analysis and provide stringent upper limits for a large number of systematic uncertainties and avoid confirmation bias by carrying out a blind analysis. The upper limit of our total systematic error budget is 5.4 per cent (68 per cent confidence) and it is dominated by uncertainties stemming from the modelling of haloes as NFW profiles.

We used N-body simulations to calibrate our mass modelling method. The sources of systematic errors in this approach are the uncertainty in this calibration, the mass–concentration relation and the miscentring distribution. Future analyses could mitigate these either by employing a larger suite of simulations and an improved understanding of the sources of discrepancies of published M–c relations or by using other mass estimators that avoid these complications. Hoekstra et al. (2015), for example, used the aperture mass (Fahlman et al. 1994; Schneider 1996) to mitigate these problems. This, however, is done at the cost of increased statistical uncertainties, so that future studies will have to weigh carefully the cost and benefits of using either the aperture mass or an NFW modelling approach. We emphasize that in our present work we are still dominated by statistical and not systematic errors. The total uncertainty (systematic and statistical) in the mass scale is $\sigma_{M_{200}} = A_{\mathrm{SZ}}{}^{1/B_{\mathrm{SZ}}} = 8.9$ per cent.

We combined the weak lensing data of our 19 clusters with data for 13 clusters from the SPT–SZ survey at high redshift observed with *HST* (S18) and *Chandra* X-ray data to calibrate mass–observable scaling relations. We described an extension of the scal-ing relation framework of Bocquet et al. (2015) to include weak lensing information. An important feature of our method is its abil-ity to correct for Eddington bias, while at the same time not using cluster number counts to self-calibrate mass–observable relations.

The normalization of the mass–SZE relation is in good agreement with the prior used in the latest SPT cosmology analysis (de Haan et al. 2016), which is based on an external calibration of this mass–observable relation. Future SPT cosmology analyses (Bocquet et al., in preparation) will now be able to use an internal calibration of the absolute mass scale, i.e. a calibration that is performed on the same clusters used for obtaining cosmological constraints. Also, our values for the normalizations of the mass–X-ray scaling relations all agree within 1σ with those found by other authors (Vikhlinin et al. 2009; Pratt et al. 2009; Mahdavi et al. 2013; Mantz et al. 2016; Chiu et al. 2018). For example, at $Y_{\mathrm{X}} = 5 \times 10^{14}\,\mathrm{M}_{\odot}$ keV our M_{500} normalization is 2.4 per cent higher than that of Vikhlinin et al. (2009) and 6.3 per cent lower than that of de Haan et al. (2016). At $M_{\mathrm{gas}} = 6 \times 10^{13}\,\mathrm{M}_{\odot}$ we obtain M_{500} values 4.6 per cent higher than Vikhlinin et al. (2009).

At the same time, our choice to avoid self-calibration of the mass–observable scaling relation from cluster number counts limits our ability to constrain the slopes and evolution parameters of these relations with a cluster sample of the present size. We therefore chose to impose informative priors on these quantities based on the self-calibration results of the SPT–SZ cluster cosmology analysis.

We have already secured more follow-up data, including *HST* data, so that we can expect to overcome this limitation in the near future. Particularly, the planned combination of SPT–SZ data with the shear catalogues of the Dark Energy Survey survey (Zuntz et al. 2018), combined with an expanded SZE cluster sample from the SPTpol experiment (Austermann et al. 2012), should allow us to extract meaningful constraints on the slope of the mass–SZE scaling relation and lead to a more stringent estimation of mass–observable scaling relation normalizations.

ACKNOWLEDGEMENTS

We acknowledge support by the DFG Cluster of Excellence 'Origin and Structure of the Universe' and the Transregio program TR33 'The Dark Universe'. The EMCEE runs have been carried out on the computing facilities of the Computational Center for Particle and Astrophysics (C2PAP), located at the Leibniz Supercomputer Center (LRZ) in Garching. The South Pole Telescope is supported by the National Science Foundation through grant PLR-1248097. Partial support is also provided by the NSF Physics Frontier Center grant PHY-1125897 to the Kavli Institute of Cosmological Physics at the University of Chicago, the Kavli Foundation and the Gordon and Betty Moore Foundation grant GBMF 947. DA and TS acknowledge support from the German Federal Ministry of Economics and Technology (BMWi) provided through DLR under projects 50 OR 1210, 50 OR 1308, 50 OR 1407 and 50 OR 1610. Work at Argonne National Laboratory was supported under U.S. Department of Energy contract DE-AC02-06CH11357. DR is supported by a NASA Postdoctoral Program Senior Fellowship at NASA's Ames Research Center, administered by the Universities Space Research Association under contract with NASA. CR and SR acknowledge support from the Australian Research Council's Discovery Projects funding scheme (DP150103208). A. Saro is supported by the ERC-StG 'ClustersXCosmo', grant agreement 71676. We thank Peter Capak for providing the COSMOS30 + UltraVISTA photo-z catalogue ahead of publication. Based on data products from observations made with ESO Telescopes at the La Silla Paranal Observatory under ESO programme ID 179.A-2005 and on data products produced by TERAPIX and the Cambridge Astronomy Survey Unit on behalf of the UltraVISTA consortium. This research made use of ASTROPY, a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013) and TOPCAT (Taylor 2005). Figs 7, 8 and 10 were created with PYGTC (Bocquet & Carter 2016).

REFERENCES

Abazajian K. N. et al., 2016, preprint (arXiv:1610.02743)
Allen S. W., Evrard A. E., Mantz A. B., 2011, ARA&A, 49, 409
Andersson K. et al., 2011, ApJ, 738, 48
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
Applegate D. E. et al., 2016, MNRAS, 457, 1522

Astropy Collaboration et al., 2013, A&A, 558, A33
Austermann J. E. et al., 2012, in Holland W. S., Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VI, SPIE, Amsterdam. p. 84521E
Bacon D. J., Massey R. J., Refregier A. R., Ellis R. S., 2003, MNRAS, 344, 673
Bahé Y. M., McCarthy I. G., King L. J., 2012, MNRAS, 421, 1073
Bardeau S., Soucail G., Kneib J.-P., Czoske O., Ebeling H., Hudelot P., Smail I., Smith G. P., 2007, A&A, 470, 449
Bartelmann M., 1996, A&A, 313, 697
Battaglia N. et al., 2016, J. Cosmol. Astropart. Phys., 8, 013
Bayliss M. B. et al., 2016, ApJS, 227, 3
Becker M. R., Kravtsov A. V., 2011, ApJ, 740, 25
Beers T. C., Flynn K., Gebhardt K., 1990, AJ, 100, 32
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Bleem L. E. et al., 2015, ApJS, 216, 27
Bocquet S., Carter F. W., 2016, The Journal of Open Source Software, 1, 46
Bocquet S. et al., 2015, ApJ, 799, 214
Bocquet S., Saro A., Dolag K., Mohr J. J., 2016, MNRAS, 456, 2361
Capak P. et al., 2007, ApJS, 172, 99
Carlstrom J. E. et al., 2011, PASP, 123, 568
Chiu I. et al., 2016a, MNRAS, 455, 258
Chiu I. et al., 2016b, MNRAS, 457, 3050
Chiu I. et al., 2018, MNRAS, 478, 3072
Cibirka N. et al., 2017, MNRAS, 468, 1092
Clowe D., De Lucia G., King L., 2004, MNRAS, 350, 1038
Clowe D., Bradač M., Gonzalez A. H., Markevitch M., Randall S. W., Jones C., Zaritsky D., 2006, ApJ, 648, L109
Corless V. L., King L. J., 2007, MNRAS, 380, 149
Corless V. L., King L. J., 2009, MNRAS, 396, 315
Coupon J. et al., 2009, A&A, 500, 981
Cunha C. E., Lima M., Oyaizu H., Frieman J., Lin H., 2009, MNRAS, 396, 2379
de Haan T. et al., 2016, ApJ, 832, 95
DES Collaboration et al., 2018, Phys Rev D, 98, 043526
Diemer B., Kravtsov A. V., 2015, ApJ, 799, 108
Dietrich J. P., Böhnert A., Lombardi M., Hilbert S., Hartlap J., 2012, MNRAS, 419, 3547
Dodelson S., 2004, Phys. Rev. D, 70, 023008
Duffy A. R., Schaye J., Kay S. T., Dalla Vecchia C., Battye R. A., Booth C. M., 2010, MNRAS, 405, 2161
Einasto J., 1965, Trudy Astrofizicheskogo Instituta Alma-Ata, 5, 87
Fahlman G., Kaiser N., Squires G., Woods D., 1994, ApJ, 437, 56
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Gelman A., Rubin D. B., 1992, Statistical Science, 7, 457
Goodman J., Weare J., 2010, Communications in Applied Mathematics and Computational Science, 5, 65
Gruen D. et al., 2014, MNRAS, 442, 1507
Gupta N., Saro A., Mohr J. J., Dolag K., Liu J., 2017, MNRAS, 469, 3069
High F. W., Stubbs C. W., Rest A., Stalder B., Challis P., 2009, AJ, 138, 110
High F. W. et al., 2012, ApJ, 758, 68
Hildebrandt H. et al., 2017, MNRAS, 465, 1454
Hilton M. et al., 2018, ApJS, 235, 20
Hoekstra H., 2001, A&A, 370, 743
Hoekstra H., 2003, MNRAS, 339, 1155
Hoekstra H., 2004, MNRAS, 347, 1337
Hoekstra H., 2007, MNRAS, 379, 317
Hoekstra H., Franx M., Kuijken K., Squires G., 1998, ApJ, 504, 636
Hoekstra H., Franx M., Kuijken K., 2000, ApJ, 532, 88
Hoekstra H., Mahdavi A., Babul A., Bildfell C., 2012, MNRAS, 427, 1298
Hoekstra H., Bartelmann M., Dahle H., Israel H., Limousin M., Meneghetti M., 2013, Space Sci. Rev., 177, 75
Hoekstra H., Herbonnet R., Muzzin A., Babul A., Mahdavi A., Viola M., Cacciato M., 2015, MNRAS, 449, 685
Ilbert O. et al., 2013, A&A, 556, A55
Johnston D. E. et al., 2007, preprint (arXiv:0709.1159)
Kaiser N., Squires G., Broadhurst T., 1995, ApJ, 449, 460
Klein J. R., Roodman A., 2005, Ann. Rev. Nucl. Particle Sci., 55, 141
Laigle C. et al., 2016, ApJS, 224, 24
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Le Brun A. M. C., McCarthy I. G., Schaye J., Ponman T. J., 2014, MNRAS, 441, 1270
Lima M., Cunha C. E., Oyaizu H., Frieman J., Lin H., Sheldon E. S., 2008, MNRAS, 390, 118
LSST Dark Energy Science Collaboration, 2012, preprint (arXiv:1211.0310)
Luppino G. A., Kaiser N., 1997, ApJ, 475, 20
Mahdavi A., Hoekstra H., Babul A., Bildfell C., Jeltema T., Henry J. P., 2013, ApJ, 767, 116
Mandelbaum R., Seljak U., Baldauf T., Smith R. E., 2010, MNRAS, 405, 2078
Mantz A. B. et al., 2016, MNRAS, 463, 3582
Marrone D. P. et al., 2012, ApJ, 754, 119
McCracken H. J. et al., 2012, A&A, 544, A156
McDonald M. et al., 2013, ApJ, 774, 23
McDonald M. et al., 2017, ApJ, 843, 28
McLeod B., Geary J., Ordway M., Amato S., Conroy M., Gauron T., 2006, in Beletic J. E., Beletic J. W., Amico P., eds, Astrophysics and Space Science Library Vol. 336, Astrophysics and Space Science Library. p. , Springer, Dordrecht, 337
Medezinski E. et al., 2018, PASJ, 70, 30
Melchior P. et al., 2017, MNRAS, 469, 4899
Melin J.-B., Bartlett J. G., Delabrouille J., 2006, A&A, 459, 341
Meneghetti M., Rasia E., Merten J., Bellagamba F., Ettori S., Mazzotta P., Dolag K., Marri S., 2010, A&A, 514, A93
Merten J. et al., 2015, ApJ, 806, 4
Momcheva I. G. et al., 2016, ApJS, 225, 27
Murata R., Nishimichi T., Takada M., Miyatake H., Shirasaki M., More S., Takahashi R., Osato K., 2018, ApJ, 854 12
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Okabe N., Smith G. P., 2016, MNRAS, 461, 3794
Okabe N., Zhang Y.-Y., Finoguenov A., Takada M., Smith G. P., Umetsu K., Futamase T., 2010, ApJ, 721, 875
Planck Collaboration et al., 2014, A&A, 571, A16
Pratt G. W., Croston J. H., Arnaud M., Böhringer H., 2009, A&A, 498, 361
Riess A. G. et al., 2011, ApJ, 730, 119
Ruel J. et al., 2014, ApJ, 792, 45
Saro A. et al., 2014, MNRAS, 440, 2610
Schneider P., 1996, MNRAS, 283, 837
Schrabback T. et al., 2018, MNRAS, 474, 2635
Seitz C., Schneider P., 1997, A&A, 318, 687
Seitz S., Schneider P., 2001, A&A, 374, 740
Shaw L. D., Zahn O., Holder G. P., Doré O., 2009, ApJ, 702, 368
Shirasaki M., Nagai D., Lau E. T., 2016, MNRAS, 460, 3913
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., Mandelbaum R., 2015, MNRAS, 449, 1259
Skrutskie M. F. et al., 2006, AJ, 131, 1163
Song J. et al., 2012, ApJ, 761, 22
Staniszewski Z. et al., 2009, ApJ, 701, 32
Sunyaev R. A., Zel'dovich Y. B., 1970, Comments on Astrophysics and Space Physics, 2, 66
Sunyaev R. A., Zel'dovich Y. B., 1972, Comments on Astrophysics and Space Physics, 4, 173
Taylor M. B., 2005, in Shopbell P., Britton M., Ebert R., eds, ASP Conf. Ser. Vol. 347, Astronomical Data Analysis Software and Systems XIV. Astron. Soc. Pac., San Francisco, p. 29
Umetsu K. et al., 2014, ApJ, 795, 163
Vanderlinde K. et al., 2010, ApJ, 722, 1180
Vikhlinin A. et al., 2009, ApJ, 692, 1033
Viola M., Kitching T. D., Joachimi B., 2014, MNRAS, 439, 1909
von der Linden A. et al., 2014, MNRAS, 439, 2
Williamson R. et al., 2011, ApJ, 738, 139
Wright C. O., Brainerd T. G., 2000, ApJ, 534, 34
Zuntz J. et al., 2018, MNRAS, 481, 1149

APPENDIX A: PSF RESIDUAL AND SHEAR SYSTEMATIC FIGURES

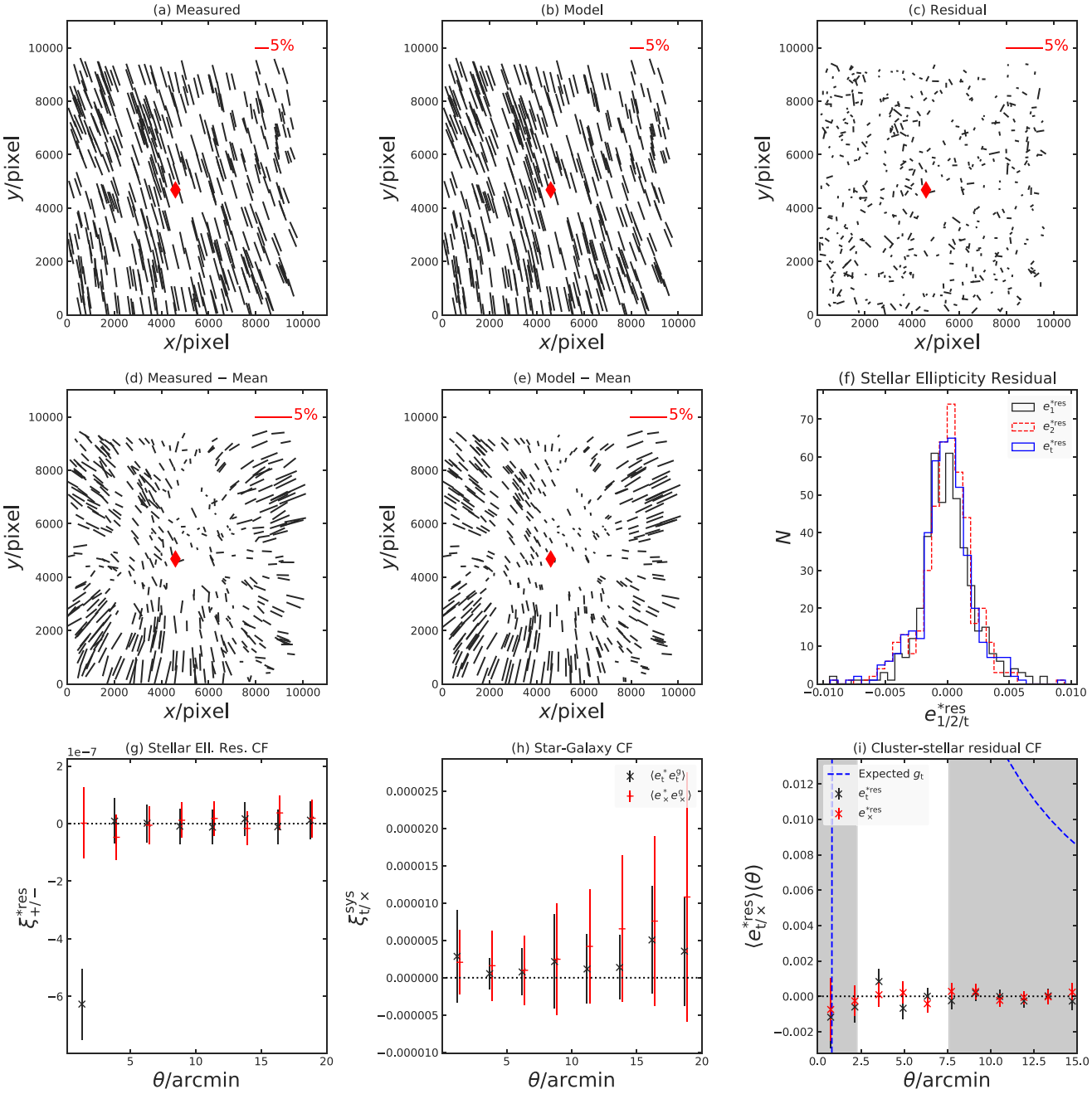
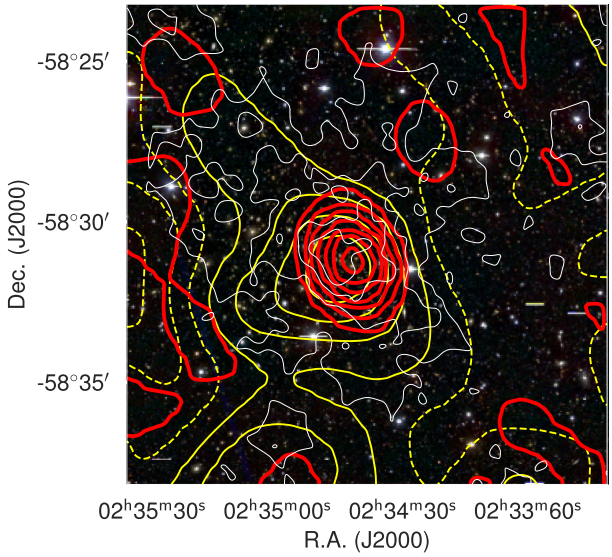
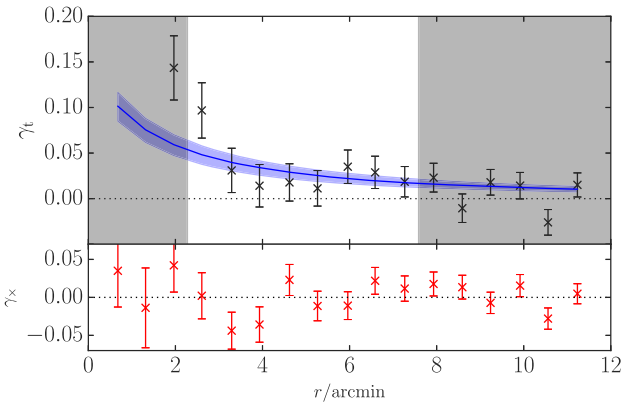


Figure A1. Same as Fig. 2 for typical, randomly chosen exposures. In this case, exposure 2 of SPT-CL J0234–5831.

APPENDIX B: MASS RECONSTRUCTIONS AND SHEAR PROFILES

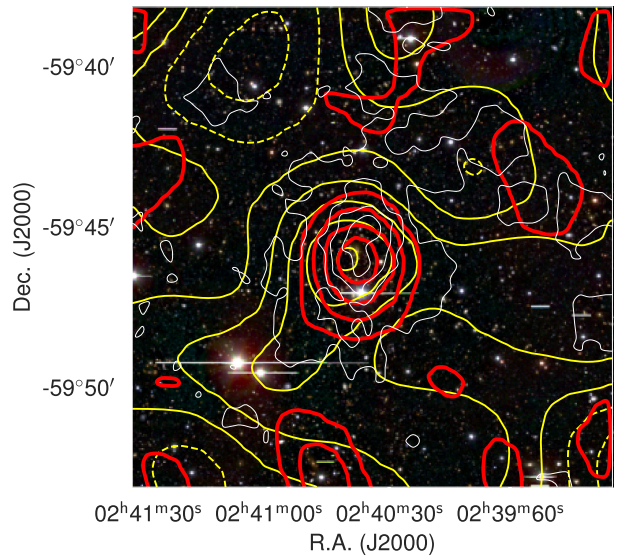


(a) Surface mass density of SPT-CL J0234−5831.

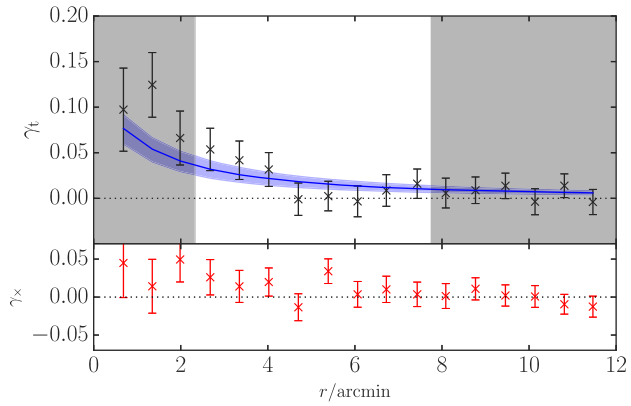


(b) Tangential shear profile of SPT-CL J0234−5831.

Figure B1. *Left panel:* significance contours of the surface mass density reconstruction in yellow (medium thick lines). These rise in steps of 1σ starting at 0 (solid lines) and decrease in steps of 1σ (dashed lines). See Section 4.1 for details on their computation. The solid red (thick) lines are the SPT SNR, also rising in steps of 1σ. The solid white (thin) lines show the SNR of the density of colour-selected red-sequence cluster galaxies. The colour image in the background is a composite of the Megacam *gri* images. *Right panel:* this panel shows the binned tangential shear around the SZ-derived cluster centre and its best-fitting NFW shear profile (see Section 4.2) in the top panel. Shaded areas were not used in the fitting procedure. The bottom panel shows the cross-shear component, which should be consistent with zero in the absence of systematic errors.

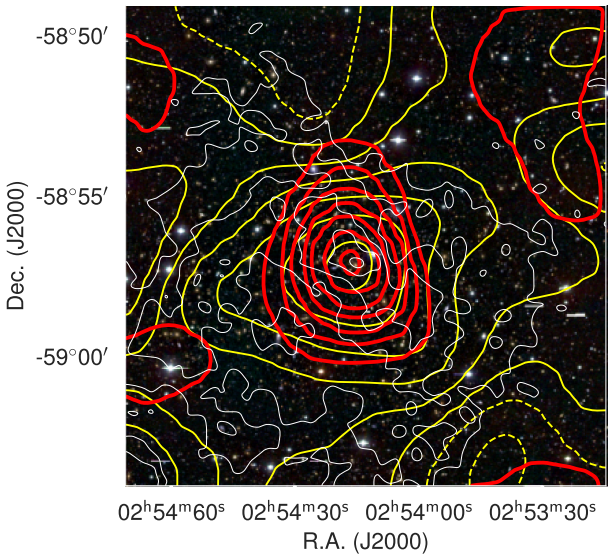


(a) Surface mass density of SPT-CL J0240−5946.

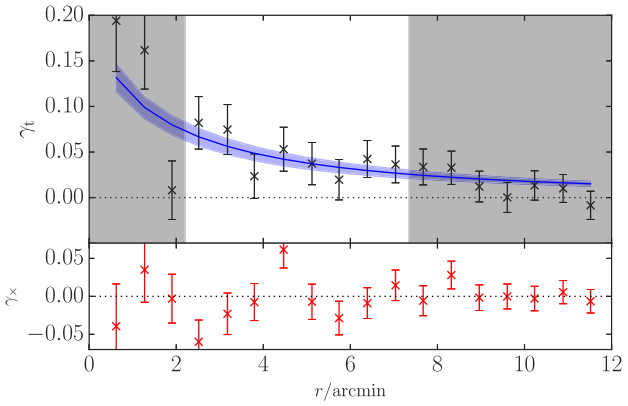


(b) Tangential shear profile of SPT-CL J0240−5946.

Figure B2. Same as Figure B1 for SPT-CL J0240–5946.

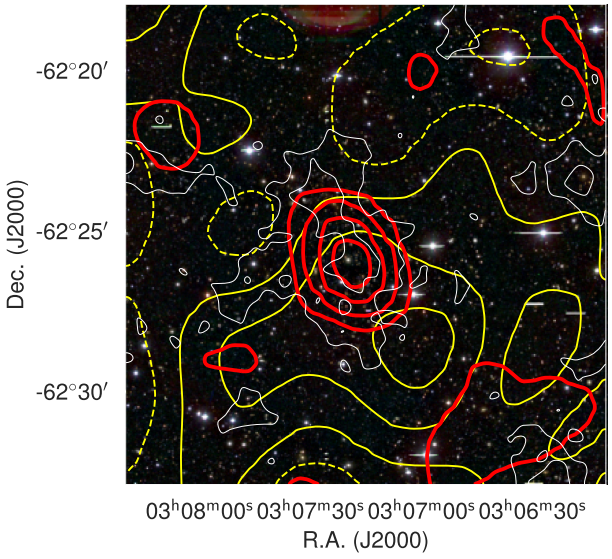


(a) Surface mass density of SPT-CL J0254−5857.

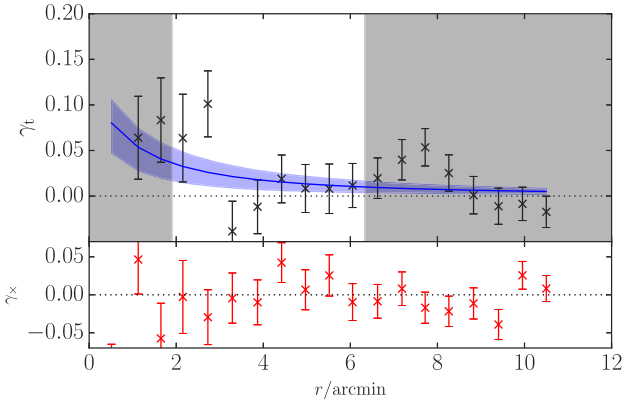


(b) Tangential shear profile of SPT-CL J0254−5857.

Figure B3. Same as Figure B1 for SPT-CL J0254–5857.

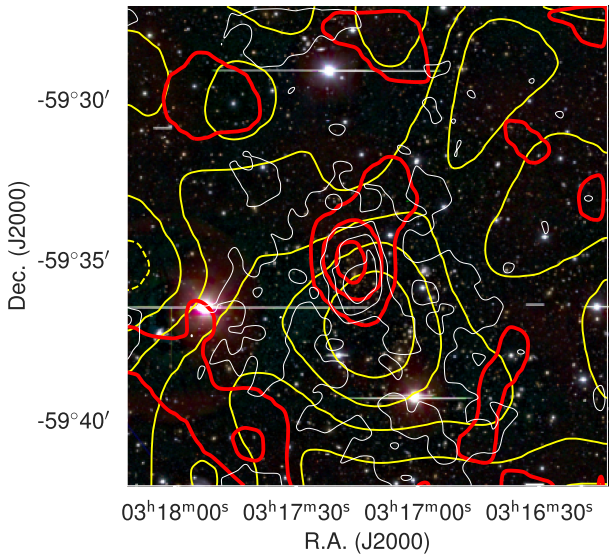


(a) Surface mass density of SPT-CL J0307−6225.

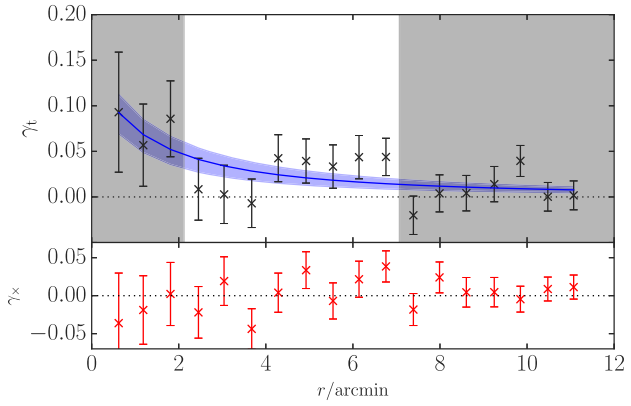


(b) Tangential shear profile of SPT-CL J0307−6225.

Figure B4. Same as Figure B1 for SPT-CL J0307–6225.

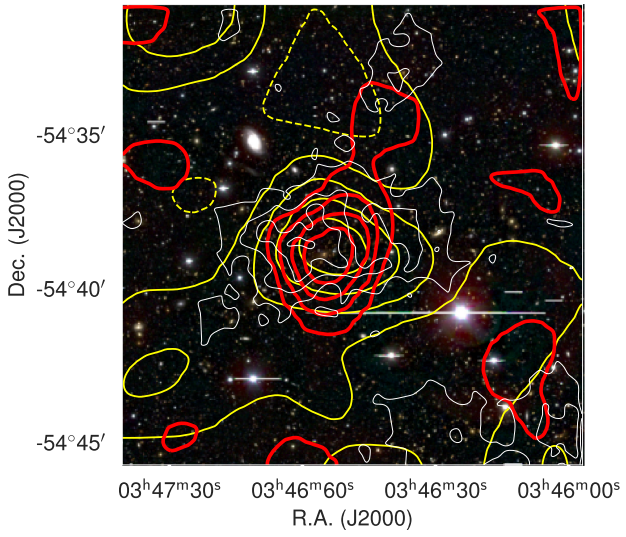


(a) Surface mass density of SPT-CL J0317−5935.

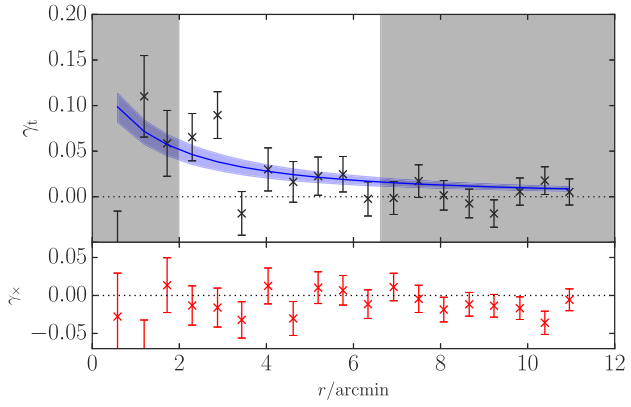


(b) Tangential shear profile of SPT-CL J0317−5935.

Figure B5. Same as Figure B1 for SPT-CL J0317–5935.

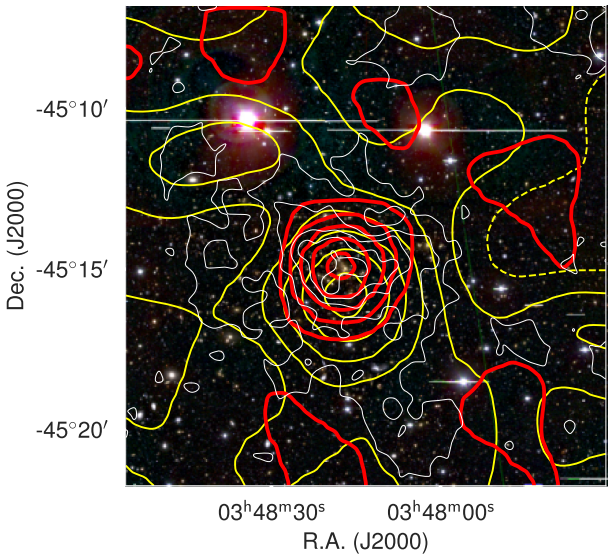


(a) Surface mass density of SPT-CL J0346−5439.

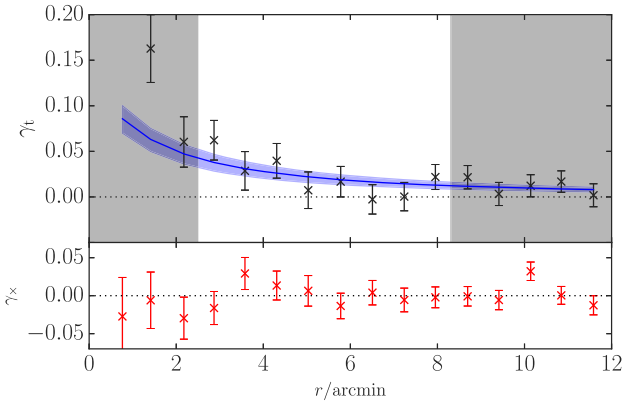


(b) Tangential shear profile of SPT-CL J0346−5439.

Figure B6. Same as Figure B1 for SPT-CL J0346–5439.

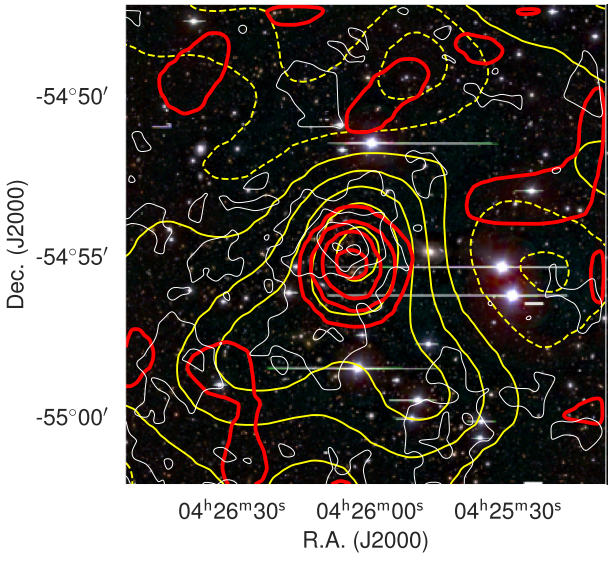


(a) Surface mass density of SPT-CL J0348−4515.

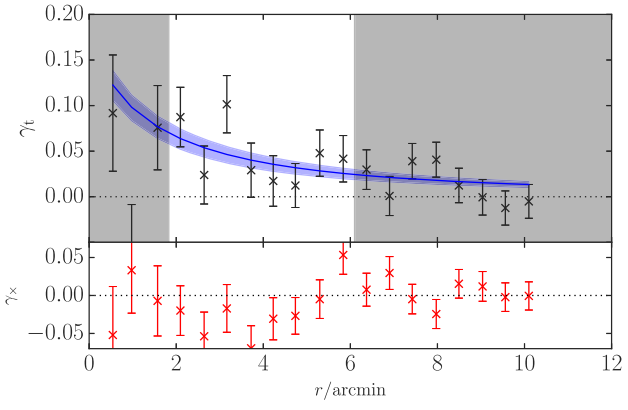


(b) Tangential shear profile of SPT-CL J0348−4515.

Figure B7. Same as Figure B1 for SPT-CL J0348–4515.

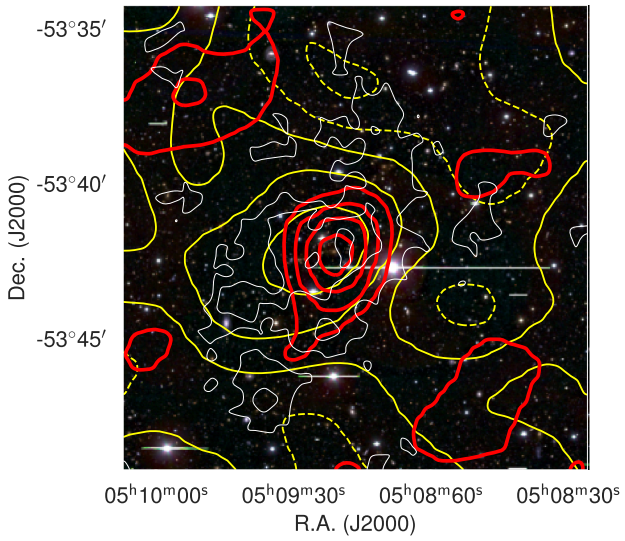


(a) Surface mass density of SPT-CL J0426−5455.

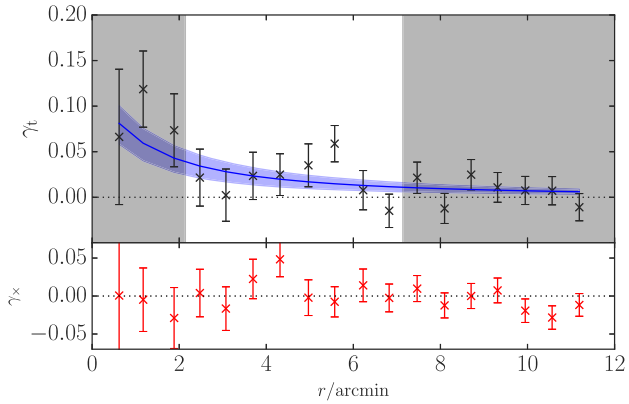


(b) Tangential shear profile of SPT-CL J0426−5455.

Figure B8. Same as Figure B1 for SPT-CL J0426–5455.

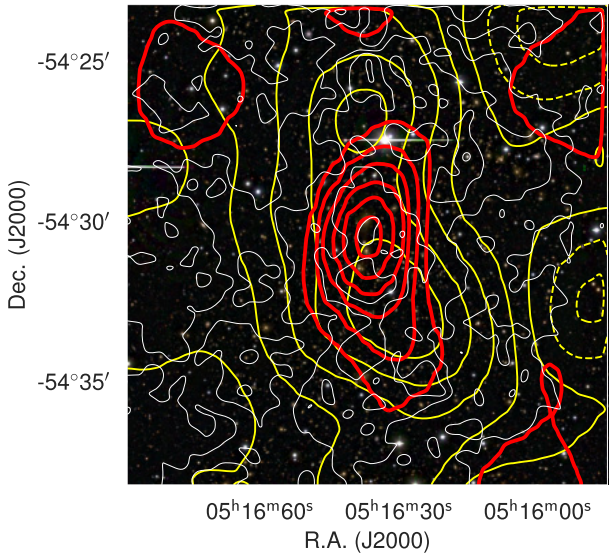


(a) Surface mass density of SPT-CL J0509−5342.

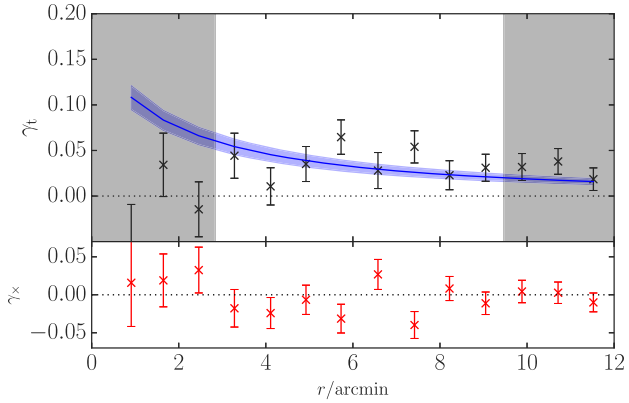


(b) Tangential shear profile of SPT-CL J0509−5342.

Figure B9. Same as Figure B1 for SPT-CL J0509–5342.

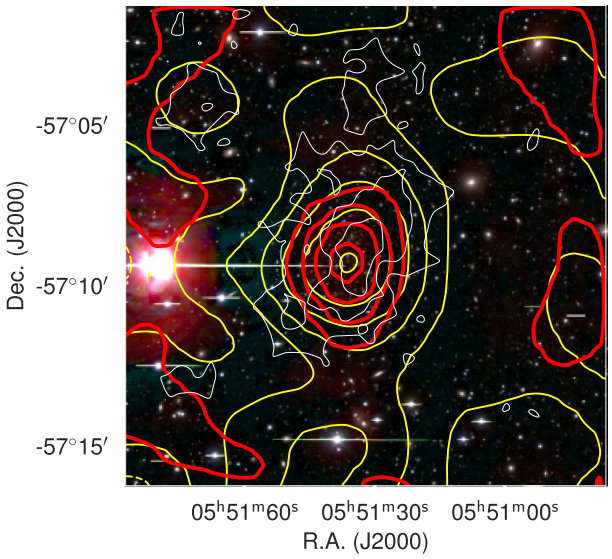


(a) Surface mass density of SPT-CL J0516−5430.

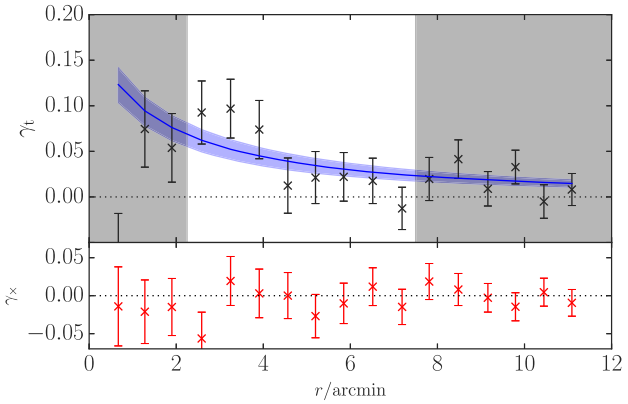


(b) Tangential shear profile of SPT-CL J0516−5430.

Figure B10. Same as Figure B1 for SPT-CL J0516–5430.

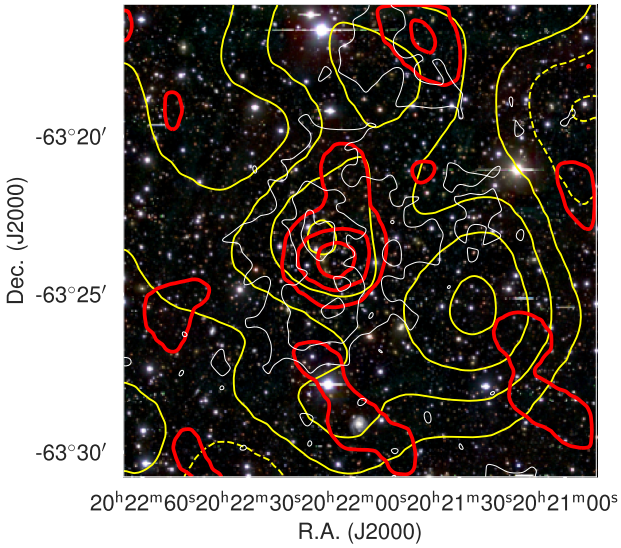


(a) Surface mass density of SPT-CL J0551−5709.

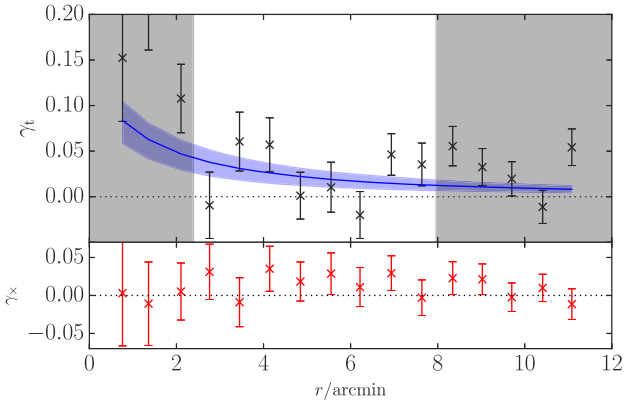


(b) Tangential shear profile of SPT-CL J0551−5709.

Figure B11. Same as Figure B1 for SPT-CL J0551–5709.

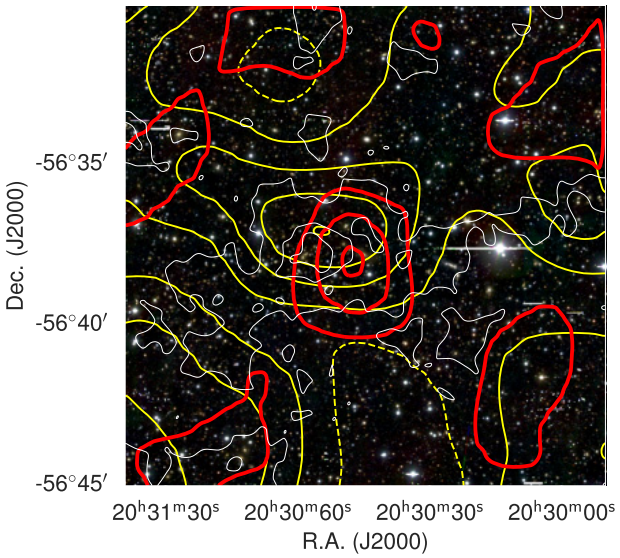


(a) Surface mass density of SPT-CL J2022−6323.

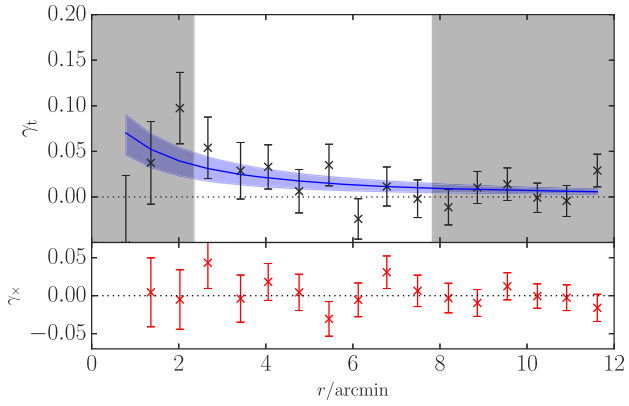


(b) Tangential shear profile of SPT-CL J2022−6323.

Figure B12. Same as Figure B1 for SPT-CL J2022–6323.

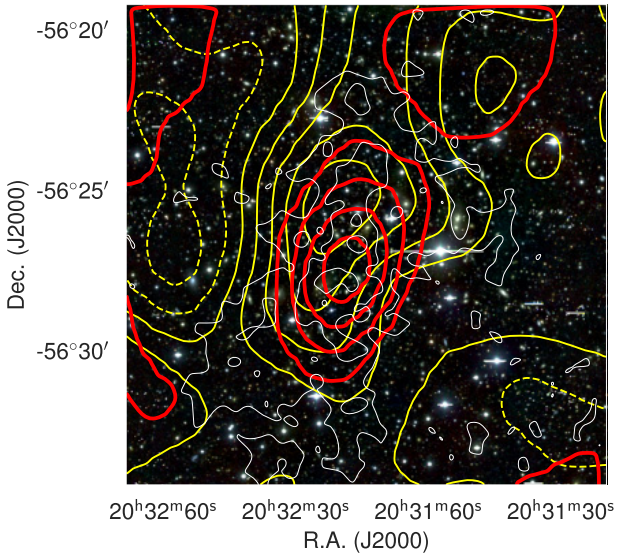


(a) Surface mass density of SPT-CL J2030−5638.

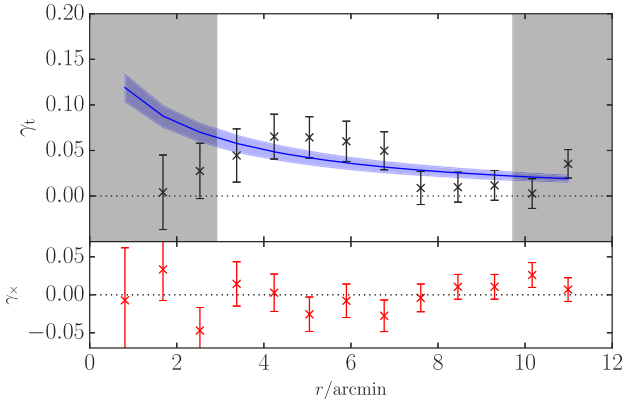


(b) Tangential shear profile of SPT-CL J2030−5638.

Figure B13. Same as Figure B1 for SPT-CL J2030–5638.

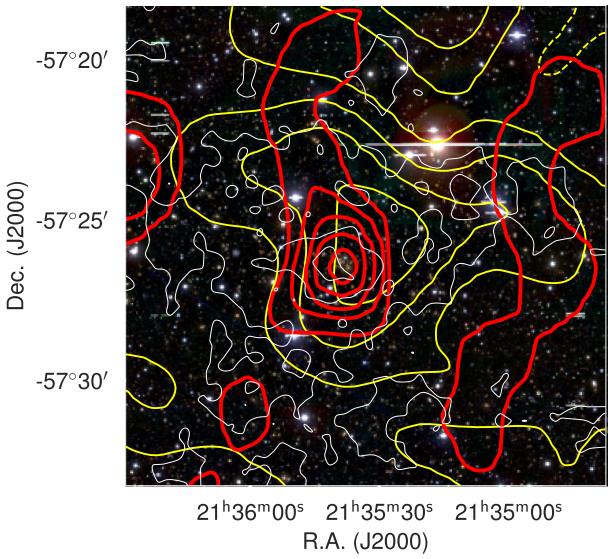


(a) Surface mass density of SPT-CL J2032−5627.

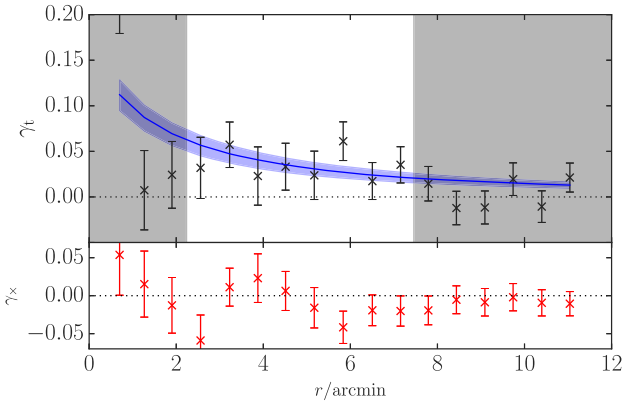


(b) Tangential shear profile of SPT-CL J2032−5627.

Figure B14. Same as Figure B1 for SPT-CL J2032–5627.

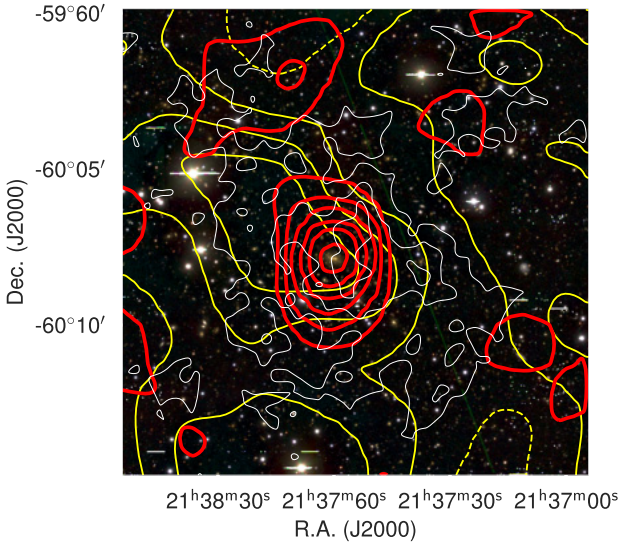


(a) Surface mass density of SPT-CL J2135−5726.

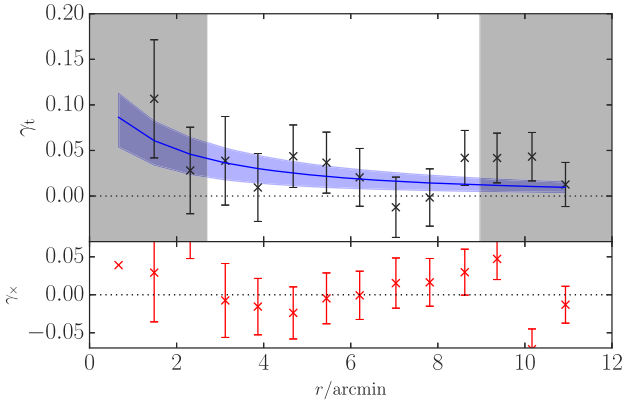


(b) Tangential shear profile of SPT-CL J2135−5726.

Figure B15. Same as Figure B1 for SPT-CL J2135–5726.

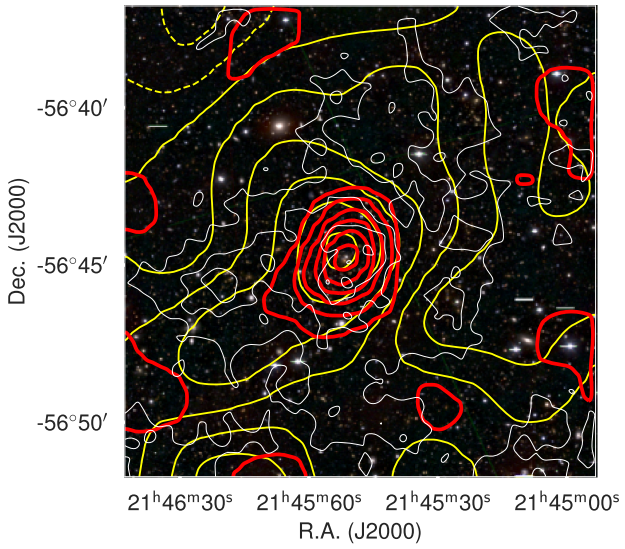


(a) Surface mass density of SPT-CL J2138−6008.

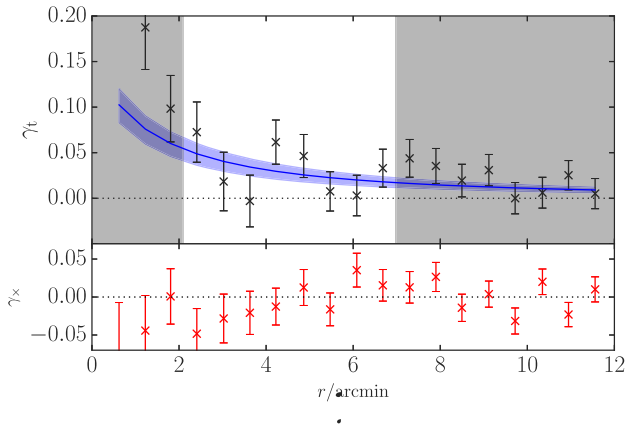


(b) Tangential shear profile of SPT-CL J2138−6008.

Figure B16. Same as Figure B1 for SPT-CL J2138–6008.

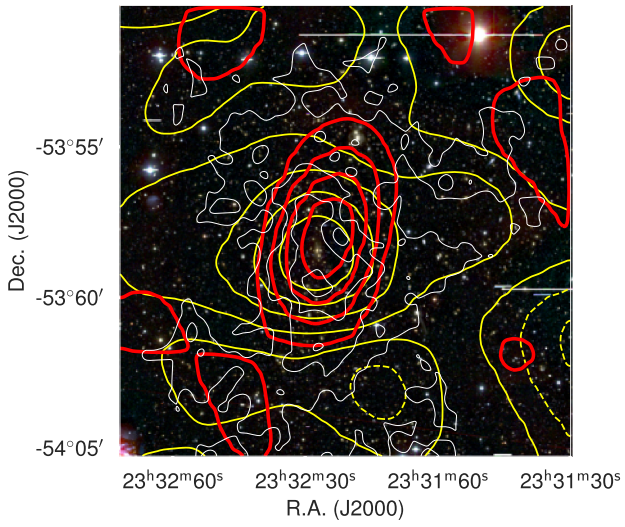


(a) Surface mass density of SPT-CL J2145−5644.

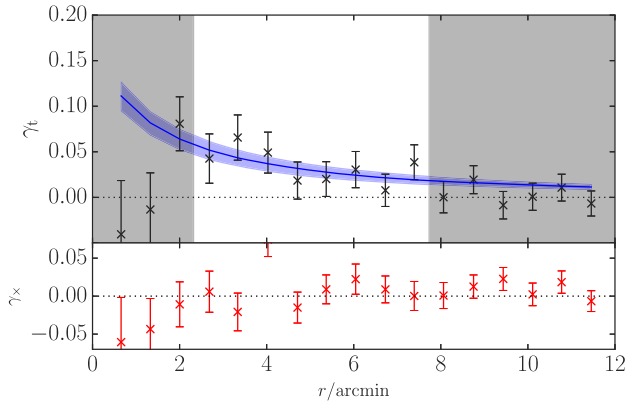


(b) Tangential shear profile of SPT-CL J2145−5644.

Figure B17. Same as Figure B1 for SPT-CL J2145–5644.

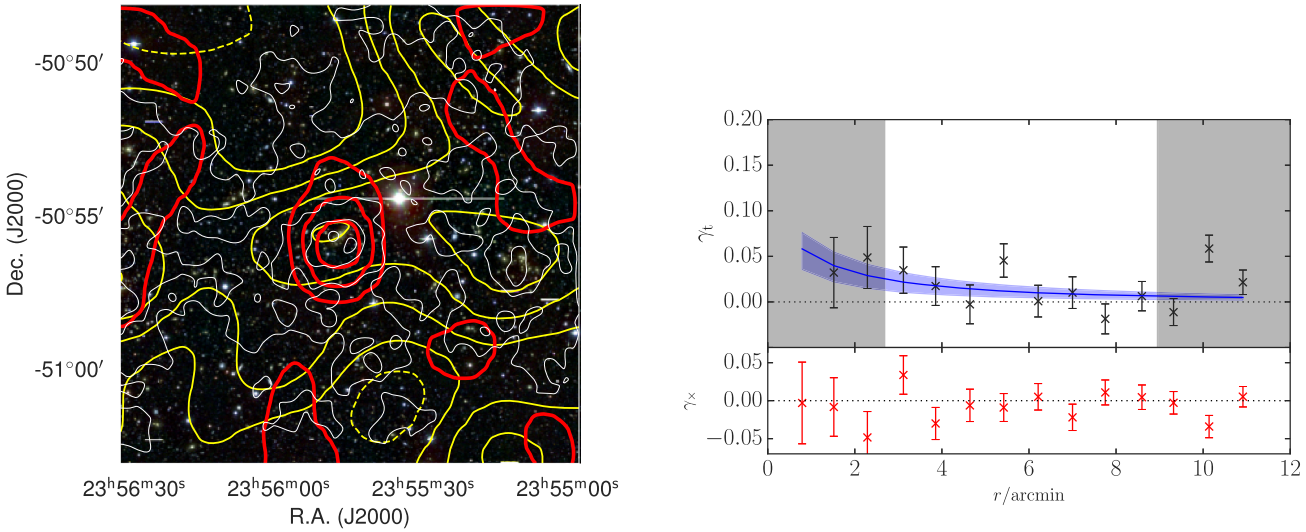


(a) Surface mass density of SPT-CL J2332−5358.



(b) Tangential shear profile of SPT-CL J2332−5358.

Figure B18. Same as Figure B1 for SPT-CL J2332–5358.



(a) Surface mass density of SPT-CL J2355−5055.

(b) Tangential shear profile of SPT-CL J2355−5055.

Figure B19. Same as Figure B1 for SPT-CL J2355–5055.

[1] *Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, D-81679 Munich, Germany*
[2] *Excellence Cluster Universe, Boltzmannstr. 2, D-85748 Garching, Germany*
[3] *Argonne National Laboratory, High-Energy Physics Division, 9700 S. Cass Avenue, Argonne, IL 60439, USA*
[4] *Kavli Institute for Cosmological Physics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA*
[5] *Argelander-Institut für Astronomie, Auf dem Hügel 71, D-53121 Bonn, Germany*
[6] *Leiden Observatory, Leiden University, Niels Bohrweg 2, 2300 CA Leiden, the Netherlands*
[7] *Max-Planck-Institut für extraterrestrische Physik, Giessenbachstr., D-85748 Garching, Germany*
[8] *Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA*
[9] *SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA*
[10] *Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 452 Lomita Mall, Stanford, CA 94305, USA*
[11] *Fermi National Accelerator Laboratory, Batavia, IL 60510-0500, USA*
[12] *Department of Astronomy and Astrophysics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA*
[13] *Department of Physics and Astronomy, University of Missouri, 5110 Rockhill Road, Kansas City, MO 64110, USA*
[14] *Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA*
[15] *Department of Astronomy, University of Florida, Gainesville, FL 32611, USA*
[16] *Department of Physics, University of California, Berkeley, CA 94720, USA*
[17] *Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY 11794, USA*
[18] *Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721, USA*
[19] *School of Physics, University of Melbourne, Parkville, VIC 3010, Australia*
[20] *Center for Astrophysics and Space Astronomy, Department of Astrophysical and Planetary Science, University of Colorado, Boulder, CO 80309, USA*
[21] *NASA Ames Research Center, Moffett Field, CA 94035, USA*
[22] *INAF–Osservatorio Astronomico di Trieste, via G. B. Tiepolo 11, I-34143 Trieste, Italy*
[23] *LSST, 950 North Cherry Avenue, Tucson, AZ 85719, USA*
[24] *Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA*
[25] *Department of Physics, Harvard University, 17 Oxford Street, Cambridge, MA 02138, USA*

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