

# Constraining the masses of high-redshift clusters with weak lensing: Revised shape calibration testing for the impact of stronger shears and increased blending

B. Hernández-Martín,<sup>1</sup> T. Schrabback<sup>1</sup>, H. Hoekstra<sup>2</sup>, N. Martinet<sup>3</sup>, J. Hlavacek-Larrondo<sup>4</sup>, L. E. Bleem<sup>5,6</sup>, M. D. Gladders<sup>6,7</sup>, B. Stalder<sup>8,9</sup>, A. A. Stark<sup>9</sup>, and M. Bayliss<sup>10</sup>

<sup>1</sup> Argelander Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, D-53121 Bonn  
e-mail: bhernandez@astro.uni-bonn.de

<sup>2</sup> Leiden Observatory, Leiden University, P.O. Box 9513, 2300 RA Leiden, The Netherlands

<sup>3</sup> Aix-Marseille Univ., CNRS, CNES, LAM, Marseille, France

<sup>4</sup> Département de Physique, Université de Montréal, Montréal, QC, Canada

<sup>5</sup> High Energy Physics Division, Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439, USA

<sup>6</sup> Kavli Institute for Cosmological Physics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA

<sup>7</sup> Department of Astronomy and Astrophysics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA

<sup>8</sup> Rubin Observatory Project Office, 950 N. Cherry Ave, Tucson, AZ 85719, USA

<sup>9</sup> Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA

<sup>10</sup> Department of Physics, University of Cincinnati, Cincinnati, OH 45221, USA

Received 28 February 2020; Accepted 27 May 2020

## ABSTRACT

Weak lensing (WL) measurements suffer from well-known shear estimation biases, which can be partially corrected for with the use of image simulations. In this work we present an analysis of simulated images that mimic Hubble Space telescope/Advance Camera for Surveys (HST/ACS) observations of high-redshift galaxy clusters, including cluster specific issues such as non-weak shear and increased blending. Our synthetic galaxies have been generated to have similar observed properties as the background-selected source samples studied in the real images. First, we used simulations with galaxies placed on a grid to determine a revised signal-to-noise-dependent ( $S/N_{\text{KSB}}$ ) correction for multiplicative shear measurement bias, and to quantify the sensitivity of our KSB+ bias calibration to mismatches of galaxy or PSF properties between the real data and the simulations. Next, we studied the impact of increased blending and light contamination from cluster and foreground galaxies, finding it to be negligible for high-redshift ( $z > 0.7$ ) clusters, whereas shear measurements can be affected at the  $\sim 1\%$  level for lower redshift clusters given their brighter member galaxies. Finally, we studied the impact of fainter neighbours and selection bias using a set of simulated images that mimic the positions and magnitudes of galaxies in Cosmic Assembly Near-IR Deep Extragalactic Legacy Survey (CANDELS) data, thereby including realistic clustering. While the initial SExtractor object detection causes a multiplicative shear selection bias of  $-0.028 \pm 0.002$ , this is reduced to  $-0.016 \pm 0.002$  by further cuts applied in our pipeline. Given the limited depth of the CANDELS data, we compared our CANDELS-based estimate for the impact of faint neighbours on the multiplicative shear measurement bias to a grid-based analysis, to which we added clustered galaxies to even fainter magnitudes based on Hubble Ultra Deep Field (HUDF) data, yielding a refined estimate of  $\sim -0.013$ . Our sensitivity analysis suggests that our pipeline is calibrated to an accuracy of  $\sim 0.015$  once all corrections are applied, which is fully sufficient for current and near-future weak lensing studies of high-redshift clusters. As an application, we used it for a refined analysis of three highly relaxed clusters from the South Pole Telescope Sunyaev-Zeldovich (SPT-SZ) survey, where we now included measurements down to the cluster core ( $r > 200$  kpc) as enabled by our work. Compared to previously employed scales ( $r > 500$  kpc), this tightens the cluster mass constraints by a factor 1.38 on average.

## 1. Introduction

Galaxy clusters are among the most massive structures in the Universe and are key to our cosmological understanding. One of the techniques used to find them is the Sunyaev-Zeldovich (SZ) effect (Sunyaev & Zeldovich 1969). In particular, the South Pole Telescope (SPT) with its 2500 deg<sup>2</sup> SPT-SZ survey (Bleem et al. 2015) has been very efficient in finding massive clusters out to the highest redshifts where they exist. This is useful to obtain robust constraints on the number of clusters as a function of redshift and mass, which provide a sensitive route to constrain cosmological parameters (e.g. Bocquet et al. 2019).

Weak lensing (WL, Bartelmann & Schneider 2001) provides a powerful tool to obtain accurate cluster mass measurements, which are required to constrain the cosmology. The images of background galaxies are distorted due to the gravitational poten-

tial of the matter along their light path. Constraining this distortion, in particular the anisotropic shear, allows us to constrain the total mass distribution inside the structures, including dark matter. In particular, weak lensing is used to directly constrain the masses of clusters and calibrate mass-observable scaling relations (e.g. Zhang et al. 2007, Dietrich et al. 2019), which can then be used to estimate the masses of other clusters from their X-ray, SZ, or optical properties. The changes in the shapes of galaxies due to WL are typically small compared to the intrinsic ellipticity of the galaxies, which leads to the so-called 'shape noise', but we can azimuthally average the shear estimates obtained from many background galaxies in order to tightly constrain the cluster shear profile in different radial bins around the cluster centre. This measurement can then be used to estimate the total mass and the mass profile of the cluster.

In this work, we study high-redshift clusters, which require deep high-resolution imaging to measure the shapes of the typically small and faint galaxies in the distant background. We follow the setup from Schrabback et al. (2018a, S18a henceforth), who study the WL signatures of 13 distant ( $z \gtrsim 0.6$ ) SPT-SZ clusters in mosaic Hubble Space Telescope (HST) observations obtained using the Advance Camera for Surveys (ACS), employing shape measurements based on the KSB+ formalism (Kaiser et al. 1995, Luppino & Kaiser 1997, Hoekstra et al. 1998).

Our main aim is to carry out the calibration of the shear estimation for weak lensing mass modelling of galaxy clusters. The difference between the real and measured shear is known as the shear bias and simulations are needed in order to calibrate it. This allows us to control the inputs and systematics, which helps us to understand how the methods behave. However, the main drawback is the need to mimic the real data as closely as possible in order to serve as proper calibration and not introduce additional biases (see Hoekstra et al. 2015, 2017). Any assumptions that simplify the simulated images should be justified and their impact studied. In particular, the presence of selection bias is often ignored, but it can have a large influence. Here, we aim to build on previous work and further understand the impact that the choices made to create our simulations have on the final results.

Schrabback et al. (2010, S10 hereafter) employed the STEP2 simulations (Massey et al. 2007) to obtain a correction for noise-related multiplicative shear measurement bias which was also applied in the cluster studies presented by S18a and Schrabback et al. (2018b, S18b henceforth). These simulations mimicked ground-based observations with a galaxy selection that did not accurately match the background-selected source samples from S18a. Because of this, S18a assumed a larger shear uncertainty to account for possible discrepancies. In this work, we create more realistic HST-like simulations in order to update the noise bias correction and focus on calibrating our shape measurement algorithm for the use in galaxy cluster studies. Most previous work on shear calibration with simulations focused on the measurement of cosmic shear (e.g. Hoekstra et al. 2017, Pujol et al. 2019, Kannawadi et al. 2019). While cosmic shear typically has tighter requirements on systematic error control, these simulations usually do not include the cluster regime of stronger shears and increased blending. In this paper we explicitly investigate these two effects.

This paper is organised as follows: in Sect. 2 we summarise the formalism of the moment-based shape measurement method employed in our analysis. In Sect. 3 we introduce our simulations and investigate how the different systematics affect the shear recovery for simulations with source galaxies placed on a grid. In Sect. 4 we study the impact of bright cluster members on our bias estimations, also as a function of their redshift. In Sect. 5 we investigate the influence of blending and selection bias by mimicking CANDELS data, for which we also estimate the signal-to-noise ratio-dependent correction used for the real data. In Sect. 6 we summarise the different bias contributions and estimate the final residual bias. Our analysis enables a robust shear recovery also in the inner parts of massive clusters, where stronger shears and increased blending occur. As a demonstration, we present an HST WL analysis of three of the most relaxed SPT-SZ clusters in Sect. 7, showing how much the mass constraints tighten by including the inner cluster regions, which have been excluded in the analysis of large samples presented by S18a and Schrabback (et al. in prep., S20 below).

In this paper all magnitudes are in the AB system. When fitting cluster shear profiles and computing cluster mass constraints

we assume a flat  $\Lambda$ CDM cosmology characterised by  $\Omega_m = 0.3$ ,  $\Omega_\Lambda = 0.7$ , and  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ .

## 2. Shape measurement method

There are different approaches to estimate the shear induced by the lensing effect. Some results are based on model-fitting (e.g. im2shape in Bridle et al. 2002, lensfit in Miller et al. 2007), whereas others employ moment-based methods, such as the Kaiser, Squires, and Broadhurst (KSB+) formalism (Kaiser et al. 1995, Luppino & Kaiser 1997, Hoekstra et al. 1998) used in this work. KSB+ is based on the computation of higher-order moments of the light distribution using a Gaussian weight function. Both kinds of methods can suffer from a bias in the recovered shear due to noise bias (Refregier et al. 2012). Extensive work has been done in the past to calibrate WL methods and correct for this and other biases. Some pioneering work which compares different methods includes the Shear TEsting Program (STEP, Heymans et al. 2006) and the GRavitational lEnsing Accuracy Testing (Bridle et al. 2010; Mandelbaum et al. 2015) and more recent work was presented for example in Hoekstra et al. (2017), Fenech Conti et al. (2017), Mandelbaum et al. (2018), Pujol et al. (2019), Euclid Collaboration: Martinet et al. (2019) and Kannawadi et al. (2019).

We employ the KSB+ formalism. It is a moment-based algorithm that determines the shapes of galaxies by performing a correction for the PSF using the stars in the field. KSB+ has been widely used for cluster studies (e.g. Hoekstra et al. 2015, von der Linden et al. 2014, S18a, S18b, Dietrich et al. 2019, Herbonnet et al. 2019). The cosmic shear community is actively developing newer methods that aim to reach the  $\sim 10^{-3}$  accuracy requirements of future cosmic shear experiments (e.g. Bernstein et al. 2016, Tewes et al. 2019). Such accuracy is however not required for current or near-future cluster studies, where an accuracy at the 1–2% level is sufficient.

The KSB+ method measures the shape of each background galaxy through the determination of the weighted quadrupole moments  $Q_{ij}$  of the light distribution in the form of

$$Q_{ij} = \frac{\int d^2\theta W(\theta) I(\theta) \theta_i \theta_j}{\int d^2\theta W(\theta) I(\theta)}, \quad (1)$$

where  $W(\theta)$  is a Gaussian weight function with scale length  $r_g$ , which we choose as the SExtractor (Bertin & Arnouts 1996) FLUX\_RADIUS, and  $I(\theta)$  is the surface brightness. The two polarisation components  $e_\alpha$  are related to the moments as

$$\begin{aligned} e_1 &= \frac{Q_{11} - Q_{22}}{Q_{11} + Q_{22}}, \\ e_2 &= \frac{2Q_{12}}{Q_{11} + Q_{22}}. \end{aligned} \quad (2)$$

The PSF impact on the ellipticity of a galaxy can be approximated as a convolution of a isotropic smearing PSF and an anisotropic kernel. We can define the PSF anisotropy kernel  $p_\mu$ , which was computed using the polarisation of stars  $e_\alpha^{*\text{obs}}$  (hence the overscript \*) as

$$p_\mu = (P^{\text{sm}*})_{\mu\alpha}^{-1} e_\alpha^{*\text{obs}}. \quad (3)$$

$P^{\text{sm}*}$  is the stellar smear polarisability tensor.

In the case of stars, following Hoekstra et al. (1998), the weight function for the moment computation in Eq. (1) is adjusted to match the  $r_g$  of the corresponding galaxy.

The PSF anisotropy-corrected polarisation can be defined as

$$e_{\alpha}^{\text{cor}} = e_{\alpha}^{\text{obs}} - P_{\alpha\beta}^{\text{sm}} p_{\beta}, \quad (4)$$

where  $P_{\alpha\beta}^{\text{sm}}$  is the smear polarisability tensor of the galaxy which describes the sensitivity to the circular smearing caused by the PSF.

We define the pre-seeing shear polarisability tensor  $P^g$  as:

$$P_{\alpha\beta}^g = P_{\alpha\beta}^{\text{sh}} - P_{\alpha\mu}^{\text{sm}} (P_{\mu\delta}^{\text{sm}*})^{-1} P_{\delta\beta}^{\text{sh}*}, \quad (5)$$

where  $P^{\text{sh}}$  is the shear polarisability tensor from Hoekstra et al. (1998), which measures the response of the galaxy ellipticity to shear in the absence of PSF effects and  $P^{\text{sh}*}$  is the stellar shear polarisability tensor.

With all these ingredients we can now define the KSB+ shear estimator as

$$\epsilon_{\alpha}^{\text{iso}} = (P_{\alpha\beta}^g)^{-1} [e_{\beta}^{\text{obs}} - P_{\beta\mu}^{\text{sm}} p_{\mu}]. \quad (6)$$

This provides an estimate for the reduced gravitational shear  $g$ . In the absence of shape measurement biases  $\langle \epsilon_{\alpha}^{\text{iso}} \rangle = g$ . In our implementation, we make the approximation  $(P^g)^{-1} = 2/\text{Tr}[P^g]$  to reduce noise, following Erben et al. (2001).

To identify the galaxies in the images we used **SExtractor**, and for the moment measurement the code **analysedac** (Erben et al. 2001). The object detection in **SExtractor** introduces a selection bias which is studied in Sect. 5.1.

Following Erben et al. (2001) we defined the KSB signal-to-noise ratio as

$$S/N_{\text{KSB}} = \frac{\sum_i W_i I_i}{\sigma \sqrt{\sum_i W_i^2}}, \quad (7)$$

where we sum over the pixels<sup>1</sup>  $i$  and  $W_i$  is the weight function evaluated in each pixel,  $I_i$  is the intensity measured in this pixel, and  $\sigma$  is the single-pixel dispersion of the sky background.

We applied similar cuts in our analysis as S10. We also required  $r_h > 1.2 r_h^*$ , where  $r_h^*$  is the measured half-light radius of the stars in the field. We only considered galaxies with  $\text{Tr } P^g/2 > 0.1$  (to exclude nearly unresolved galaxies with very large PSF corrections) and  $r_g < 10$  pixels (to ensure a sufficient coverage by the postage stamps<sup>2</sup>). We employed weights to down-weight the contributions of galaxies with noisy ellipticity estimates. Following S18a we defined weights via the magnitude-dependent RMS ellipticity, thereby avoiding biases that can occur for ellipticity-dependent weights.

Knowing the input shear ( $g^{\text{true}}$ ) in our simulations we can run our pipeline and compare it to the recovered values ( $g^{\text{obs}}$ ) assuming the following relation:

$$g_{\alpha}^{\text{obs}} = (1 + m_{\alpha}) g_{\alpha}^{\text{true}} + c_{\alpha}, \quad (8)$$

where  $\alpha$  refers to each component. In the case of cosmic shear the additive bias ( $c$ ) is important, but for cluster analyses it typically cancels out for the azimuthally averaged tangential shear

<sup>1</sup> In fact, the Erben et al. (2001) KSB+ implementation employed in our analysis uses sub-pixel interpolation to reduce the impact of pixel sampling when computing brightness moments and other relevant quantities such as  $S/N_{\text{KSB}}$ .

<sup>2</sup> S18a use an even more restrictive cut  $r_g < 7$  pixels. We find that switching between their and our cut changes the resulting multiplicative bias by less than 0.1% for our colour-selected sample of mostly high-redshift distant background galaxies.

profiles<sup>3</sup>. In the following sections, we therefore mostly concentrate the analysis on the multiplicative bias ( $m$ ). A consistency check for a possible quadratic dependence showed that this is negligible (see Sect. 3.6).

The shear measurement suffers from noise bias (Viola et al. 2014). It is caused by the presence of noise in the data and the lack of knowledge on the exact parameters of the background galaxies (e.g. centroid, size and ellipticity). This shear measurement method is particularly biased at low signal-to-noise ratios. This was partly corrected for in S10, as

$$g^{\text{cor}} = \frac{g^{\text{biased}}}{1 + m^{\text{cor}}}. \quad (9)$$

Their correction ( $m^{\text{cor}}$ ) was computed using the STEP2 simulations (Massey et al. 2007) of ground-based images and also tested for ACS-like simulations, which match the ACS resolution and detector characteristics, but do not accurately represent the background-selected source population of high-redshift cluster studies. Approximating the correction as a power law, S10 obtained:

$$m^{\text{cor}} = -0.078 \left( \frac{S/N_{\text{KSB}}}{2} \right)^{-0.38}. \quad (10)$$

This correction depends on the KSB+ signal-to-noise ratio, which is defined in Eq. (7). The effect of blends and neighbours complicates this simple picture (e.g. Hoekstra et al. 2015). Our aim in the next section is to estimate a correction which depends on the signal-to-noise ratio and has been computed using isolated galaxies. This does not fully capture all bias effects but it removes the main contribution of noise bias.

An alternative signal-to-noise ratio to  $S/N_{\text{KSB}}$  can be defined using the **SExtractor** parameters. This definition is more widely used since it does not depend of the particulars of the shape measurement implementation:

$$S/N_{\text{flux}} = \frac{\text{FLUX\_AUTO}}{\text{FLUXERR\_AUTO}}. \quad (11)$$

We use this definition primarily to apply signal-to-noise cuts to the galaxies.

### 3. Galaxies on a grid

Any shear measurement method needs to be calibrated through simulations to test for differences that might arise between the input and the recovered shear. For an accurate WL measurement, we need to understand and correct for such biases. We created customised simulations, which allowed us to test cluster-specific issues such as stronger shears, but also the impact of choices regarding the input parameters. An alternative option would be to use metacalibration (Sheldon & Huff 2017), which does not

<sup>3</sup> Additive biases can have a net impact if cluster centres are always placed at similar detector coordinates and similar PSF patterns occur with a significant net positive or negative average tangential PSF ellipticity (see e.g. Dietrich et al. 2019). For HST/ACS thermal breathing can cause substantial PSF ellipticity in individual frames. However, averaged over many exposures the mean tangential PSF ellipticity with respect to either the camera centre or the corners (which are typically used to place cluster centres, see S18a and S20) is very small (compare Fig. 8 in S10). Given the low level of additive biases for our method (see Table 2), typical cluster shears, and the accuracy requirements of our cluster studies, we therefore conclude that we can safely ignore the impact of additive biases.

use simulations, but it requires large quantities of data to avoid having very noisy results. We varied different aspects of the simulations, in order to investigate the sensitivities of our analysis to modelling details. The aim of this section is to understand the main simulation parameters that change the bias estimation, aiming to calibrate it to a tolerance level of around 1%. Here, we use a grid placement for our galaxies in order to speed up the computation, to test how the input choices alone (without the impact of neighbours) affect the bias, and for easier comparison to previous work. A more realistic approach, where we look at the effect of neighbours and selection bias is used in Sect. 4 and 5.

### 3.1. Details on the creation of the simulations

We created the simulations with the python package GALSIM<sup>4</sup> (Rowe et al. 2015) and used Sérsic profiles for the galaxy light distribution:

$$I(R) = I_e \exp(-b_n[(R/R_e)^{1/n} - 1]), \quad (12)$$

with the model half-light radius  $R_e$ , the intensity at that radius  $I_e$  and the parameter  $b_n \approx 2n - 1/3$ , with  $n$  being the so-called Sérsic index. The light profile of each galaxy was obtained from parametric fits to COSMOS galaxy images which are included in GALSIM. These provide input values for the Sérsic index and the intrinsic ellipticity. This allowed us to employ a realistic input shape distribution for the creation of our galaxies. A more in depth discussion of this choice is presented in Sect. 3.4.

We simulated galaxies as they would be observed in HST/ACS images, shearing them either by a constant shear or, for the simulations discussed in Sect. 4, a theoretical cluster shear profile. We used an ACS-like PSF obtained from the software Tiny Tim (Krist et al. 2011), using re-fitted optical parameters (Gillis et al. 2019) and including the effect of charge diffusion, which is discussed in more detail in Sect. 3.6.

We present here our ‘grid simulation’, which is used as a reference for the rest of this section when we vary the inputs and test how the results change. The simulation settings are summarised in Table 1, while the importance of these choices is investigated throughout the next subsections. In Sect. 5, we discuss a more realistic scenario where galaxies suffer from neighbour contamination.

For their creation we extended the normally studied range of shear values to the non-weak regime of  $|g| < 0.4$ , although the bias estimates reported in this work have been computed from the restricted range  $|g| < 0.2$  unless explicitly stated differently, since that is the regime used in cluster studies such as Schrabback et al. (2018a). We also note that GALSIM does not allow for the inclusion of flexion effects (Goldberg & Natarajan 2002, Goldberg & Bacon 2005, Bacon et al. 2006). This is another reason why we limited our primary analysis to the  $|g| < 0.2$  regime, where flexion only plays a minor role.

To compute the bias, each image was assigned a different value of input shear, which is constant throughout that image. We used 50 different shear values in the range  $-0.4 < g < 0.4$  with  $10^4$  galaxies per value. We created a grid of  $100 \times 100$  stamps of size  $100 \times 100$  pixels and applied a random shift with a uniform distribution from  $-0.5$  to  $0.5$  at the sub-pixel level for the galaxy position to have a small displacement with respect to the pixel centre. In order to reduce shape noise we created a second set of galaxies, which are identical except for a 90 degree rotation of the input intrinsic ellipticity (Massey et al. 2007). This makes

sure that the mean intrinsic ellipticity of the input population is 0, which reduces the number of galaxies needed, allowing us to constrain multiplicative biases to the few  $\times 10^{-3}$  level. Multiple copies of the galaxies, rotated by  $\pm 45$  degrees can also be created (e.g. Fenech Conti et al. 2017), but in our case, due to the faintness of the galaxies, the noise dominates and the inclusion of further rotated simulations does not lead to a significant improvement, hence why we chose to only implement pairs.

In this part of the analysis we only consider galaxies that provide shape estimates in both the normal and the rotated frame. By doing this, we effectively cancel the effects of selection bias, which can be quite important (Kannawadi et al. 2019). In this section, however, we concern ourselves only with the changes in the residual bias estimates due to the input choices for the creation of the simulations. Using matched rotated pairs reduces the number of galaxies needed for these estimates. Our study of selection biases will be presented in Sect. 5.1.

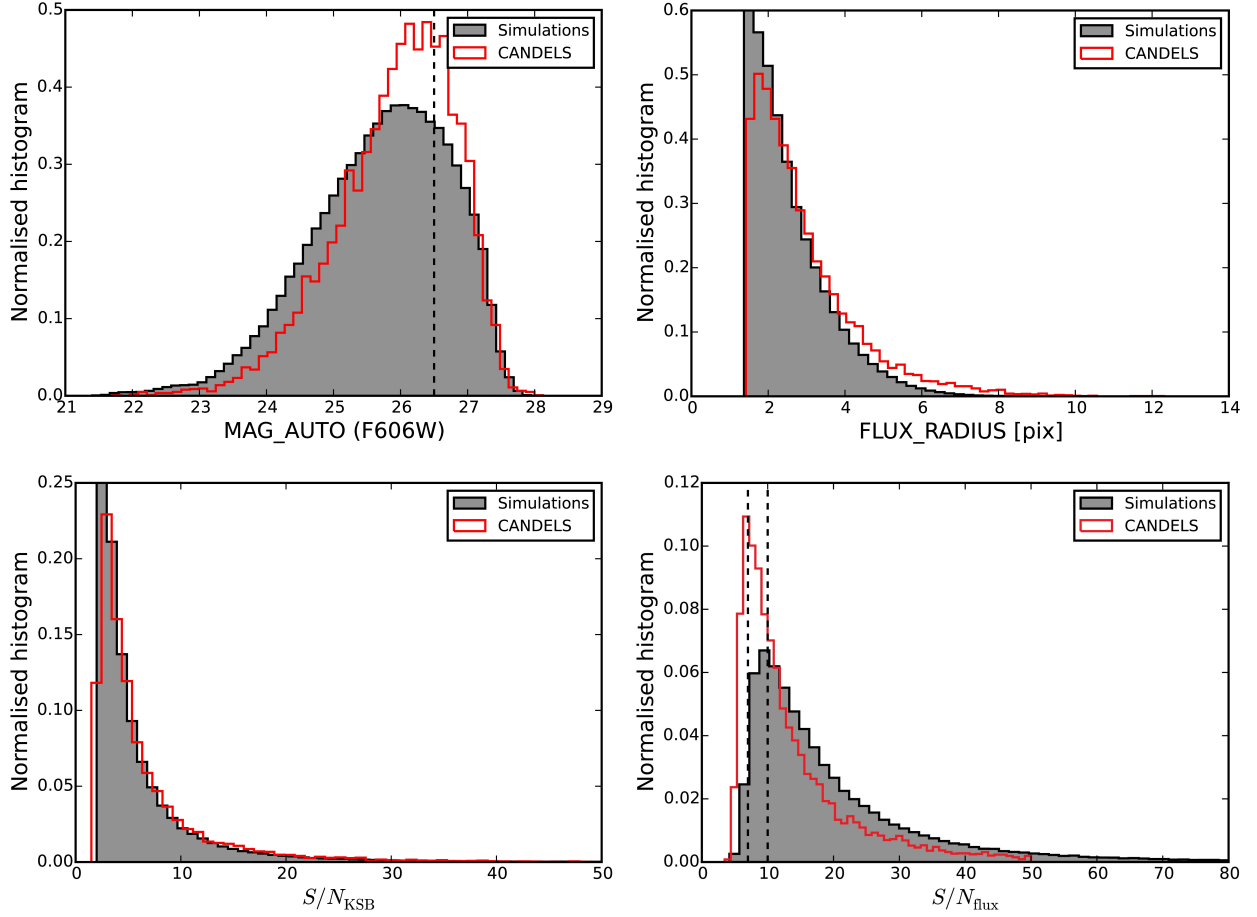
The simulations should aim to resemble the real galaxies as closely as possible to make sure that we are not introducing any artificial bias due to our choice of input parameters. In order to achieve this, we compared our measured distributions of the signal-to-noise ratio, size and magnitude with the ones obtained using data from the Cosmic Assembly Near-IR Deep Extragalactic Legacy Survey (CANDELS) fields (Grogin et al. 2011, Koekemoer et al. 2011) as analysed by S18a (see Fig. 1). In particular, we employed catalogues from S18a, which are based on ACS F606W stacks that approximately match the depth of our cluster field observations. S18a applied the same selection to these catalogues as to the cluster field observations in terms of galaxy shape parameters, magnitudes, signal-to-noise ratios, and colours. Importantly, their colour selection was tuned to provide a robust cluster member removal, selecting mostly background galaxies at  $z \gtrsim 1.4$ . By matching the measured source properties of these catalogues with our simulations we therefore make sure to adequately resemble the source properties in the cluster field WL data.

We compare the measured distributions for the galaxies in CANDELS and the simulations in Fig. 1. The resulting distributions in  $S/N_{\text{KSB}}$  and the half-light radius are well matched. The simulation contains a slightly lower fraction of faint ( $26 \leq \text{MAG\_AUTO} \leq 27$ ) galaxies with  $S/N_{\text{flux}} < 10$ . We expect that this slight mismatch is caused by incompleteness: we used the CANDELS magnitude distribution (which itself is incomplete) as an input, but at faint magnitudes our simulation analysis recovered only an incomplete fraction of these, causing the discrepancy. Employing a reweighting scheme we verified that this minor discrepancy affects our results at a negligible ( $\leq 0.3\%$ ) level only, so we can safely ignore it for our analysis. It is important to note that this issue can be reduced by using deeper input catalogues. In fact, we will do this in Sect. 5.3, where the full depth CANDELS photometric catalogue will be used as an input, rather than the single-orbit depth shape catalogue from S18a. If we were to use our shear measurement pipeline with images from different telescopes, new sets of simulations matched to the desired setup should be created to check that the results are still applicable, as we do in Appendix A for a setup that approximately resembles the VLT/HAWK-I WL data from S18b.

We computed the magnitude input distribution as well as the size distribution by measuring the properties of real CANDELS galaxies and then drawing randomly from it. The decision to obtain the size distributions from the KSB+ measured CANDELS data and not from the parametric fits to COSMOS is motivated by the difference in depth between the datasets, where the COSMOS catalogue is limited to  $F_{814W} < 25.2$ .

<sup>4</sup> <https://github.com/GalSim-developers/GalSim>





**Fig. 1.** Comparisons between the measured distributions in our simulations and the CANDELS data for the F606W magnitudes, the half-light radius, the KSB signal-to-noise ratio  $S/N_{\text{KSB}}$ , and the SExtractor signal-to-noise ratio  $S/N_{\text{flux}}$ . The vertical lines in the MAG\_AUTO distribution indicate the magnitude cuts we introduce for the bias estimation. The two vertical lines in the  $S/N_{\text{flux}}$  distribution show the two different signal-to-noise ratio cuts that we apply in this work.

**Table 1.** Summary of the input parameters used for the creation of the mock galaxies of the reference simulation.

Parameter	Type	Values
PSF model	HST/ACS-like	FWHM= 0".1
PSF ellipticity	HST/ACS-like	$ e  = 0.072$
Noise	Gaussian	$\sigma = 0.14 e^-/s$
Light distribution	Parametric fit to COSMOS*	Sersic index: 0.3 – 6
Ellipticity distribution	Parametric fit to COSMOS*	$ e  = 0 - 1$
Half-light radius	Fit to CANDELS	0".05 – 0".3
Magnitudes	Fit to CANDELS	21.7 – 28.1
Pixel scale	HST-like	0".05/pixel

\* Drawn from parametric fits to individual COSMOS galaxies.

We included uncorrelated Gaussian noise, where the level is tuned to provide a good match in the measured  $S/N_{\text{KSB}}$  distributions between CANDELS and the simulations. Our simple assumptions regarding the noise plus the slight underrepresentation of galaxies with faint measured magnitudes in the simulation (see the top panels of Fig. 1) may be the reason for the lack of low  $S/N_{\text{flux}}$  galaxies seen in the lower right panel of Fig. 1.

We drew from independent magnitude and size distributions, which means we did not fully capture the correlation between galaxy parameters. The importance of these correlations was discussed in Kannawadi et al. (2019), who put emphasis on simulating galaxies with joint distributions. This is not as important in our analysis as we do not perform tomographic cuts and always

average over our full population. Also, Kannawadi et al. (2019) show that it is especially important to account for such correlations for lower redshift sources, but that the impact becomes small for their highest redshift sources bin, which is the closest to the galaxies we are simulating. We suspect that this redshift dependence is caused by the fact that a broad range of different morphological types contribute to lower redshift source samples, while the highest-redshift sources are largely dominated by star-forming late-type galaxies.

As mentioned above, the input COSMOS catalogues used here are not as deep as our simulations. Thus, the input galaxies are brighter than the galaxies we aim to simulate. Since we modified the inputs slightly to match the CANDELS catalogues

which have approximately the same depth as our real cluster catalogues this should not be a problem.

### 3.2. Revised noise bias correction

S10 established that the multiplicative bias of our KSB+ implementation shows a strong dependence on  $S/N_{\text{KSB}}$ . This should be compensated for using a  $S/N_{\text{KSB}}$ -dependent correction in order to weaken the requirements on how well the simulations have to match the real data. However, S10 obtained their correction from the STEP2 simulations of ground-based weak lensing data (Massey et al. 2007). These differ substantially from the HST cluster data we are interested in, likely affecting the required correction. Indeed, Fig. 2 shows that our analysis of HST-like simulations with background-selected isolated galaxies yields a steeper dependence on  $S/N_{\text{KSB}}$ , which is not well described by the S10 correction. In fact, when using the correction from S10 we measure a multiplicative residual bias of  $\sim -0.0221 \pm 0.0042$ . Assuming the same power-law function form as in Eq. (10) we computed a revised correction

$$\begin{aligned} m_{1\text{corr}} &= -0.358 (S/N_{\text{KSB}})^{-1.145}, \\ m_{2\text{corr}} &= -0.357 (S/N_{\text{KSB}})^{-1.298}, \end{aligned} \quad (13)$$

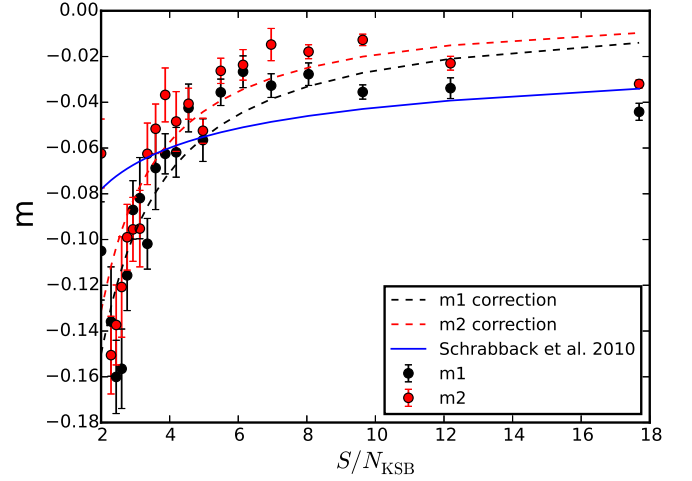
which describes the dependence reasonably well and differs slightly for the two components. We applied this revised noise bias correction in the remaining analysis, where we always applied a correction on a galaxy by galaxy basis, scaling the KSB+ shear estimates by a factor  $1/(1 + m_{\text{acorr}})$  depending on the  $S/N_{\text{KSB}}$  of the individual galaxy.

We note that not yet all relevant effects (e.g. selection bias) have been included in this analysis, which is used to derive the correction. We therefore have to add additional residual bias corrections below. We find, however, that these additional effects do not change the  $S/N_{\text{KSB}}$  dependence significantly (see e.g. Sect. 5.3), which is why we can account for them using constant ( $S/N_{\text{KSB}}$ -independent) bias offsets.

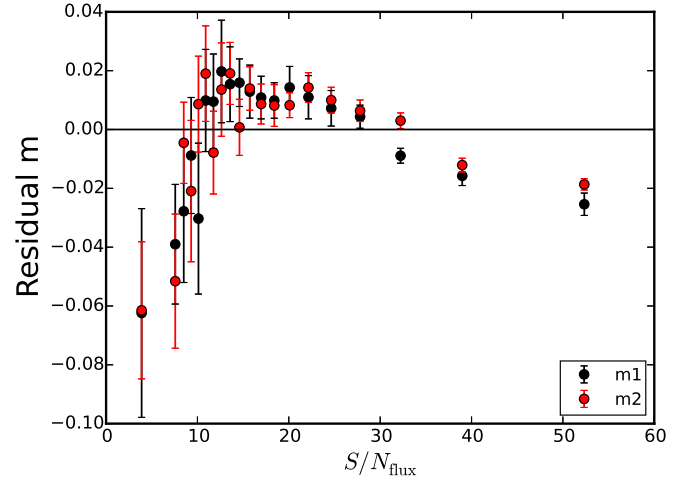
In Table 2 we present the bias estimates obtained from the grid reference simulation, computed for two different shear ranges. We used  $|g| < 0.4$  in order to study the behaviour for strong shears, but for typical datasets the  $|g| < 0.2$  regime is more relevant, so we included this estimation, which will then be used later. Nevertheless, we obtained consistent bias results for both limits. The residual multiplicative bias is smaller than 1% in both cases after applying the  $S/N_{\text{KSB}}$ -dependent corrections from Eq. (13) and a cut on  $S/N_{\text{flux}} > 10$  (which is the default  $S/N_{\text{flux}}$  cut from S18a). Errors were estimated by bootstrapping the galaxies in each image. In the following sections all bias estimates by default correspond to the  $|g| < 0.2$  setup unless explicitly stated differently. The estimates from this grid reference simulation will be used to study throughout the rest of Sect. 3 how the bias shifts as we change the input parameters.

**Table 2.** Residual multiplicative and additive bias for the simulation with galaxies placed on a grid after a  $S/N_{\text{flux}} > 10$  cut and applying the correction from Eq. (13).

Limits	$m_{1,\text{grid}}$	$c_{1,\text{grid}}$	$m_{2,\text{grid}}$	$c_{2,\text{grid}}$
		[ $\times 10^3$ ]		
$ g  < 0.4$	$-7.3 \pm 2.0$	$-1.2 \pm 0.4$	$-6.7 \pm 1.3$	$-4.3 \pm 0.3$
$ g  < 0.2$	$-8.1 \pm 3.8$	$-1.0 \pm 0.4$	$-5.1 \pm 3.3$	$-3.5 \pm 0.3$



**Fig. 2.** Dependence of the multiplicative bias on  $S/N_{\text{KSB}}$  and comparison between the  $S/N_{\text{KSB}}$  correction from S10 (shown in blue) and the new correction presented in Eq. (13). The red and black dashed lines show the correction for the two shear components. We binned the galaxies according to their  $S/N_{\text{KSB}}$  and computed the bias for each bin separately. We include no  $S/N_{\text{flux}}$  cuts here but note that the inclusion of a  $S/N_{\text{flux}} > 10$  cut changes the results only slightly.

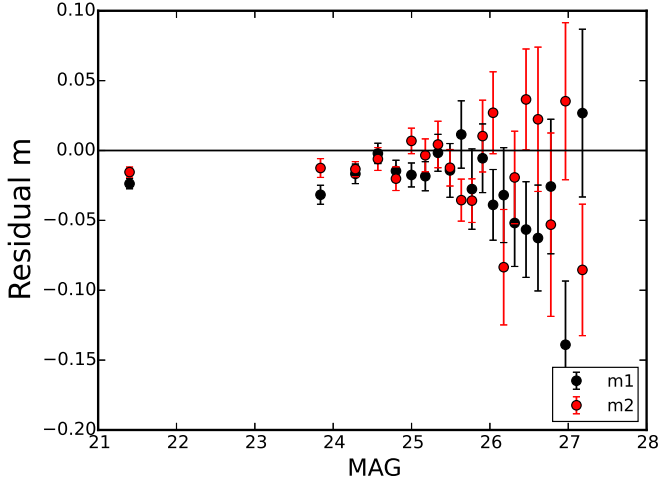


**Fig. 3.** Residual dependence of bias on the  $S/N_{\text{flux}}$ , shown here after the  $S/N_{\text{KSB}}$ -dependent correction is applied. The residual dependence is caused by a slight under- or over-corrections in different  $S/N_{\text{KSB}}$  regimes in Fig. 2, as well as differences in the definitions of the signal-to-noise ratios.

### 3.3. Residual dependence on $S/N_{\text{flux}}$ and magnitude

The use of the  $S/N_{\text{KSB}}$  correction defined in Eq. (13), should compensate for the strongest bias dependence. Here we investigate residual dependences on other parameters. In Fig. 3 we show the dependence of the residual multiplicative bias on  $S/N_{\text{flux}}$  by measuring the bias in 20  $S/N_{\text{flux}}$  bins. The residual dependence comes from the fact that the  $S/N_{\text{KSB}}$ -dependent correction over-corrects at  $5 \leq S/N_{\text{KSB}} \leq 8$ , which corresponds with the positive bias at  $10 \leq S/N_{\text{flux}} \leq 20$ . Similarly, the bias is slightly under-corrected at both very low and very high  $S/N_{\text{KSB}}$ , which can also be seen in the  $S/N_{\text{flux}}$  dependence.

In S18a, a conservative approach was used, restricting the galaxies to  $S/N_{\text{flux}} > 10$ . We tested the influence of different



**Fig. 4.** Dependence of the residual multiplicative bias on magnitude for the two ellipticity components. Here we have applied the  $S/N_{\text{KSB}}$ -dependent correction and the cut  $S/N_{\text{flux}} > 10$ . The bins are created to have the same number of galaxies, but we have fewer galaxies for brighter magnitudes, which explains the different separation of points.

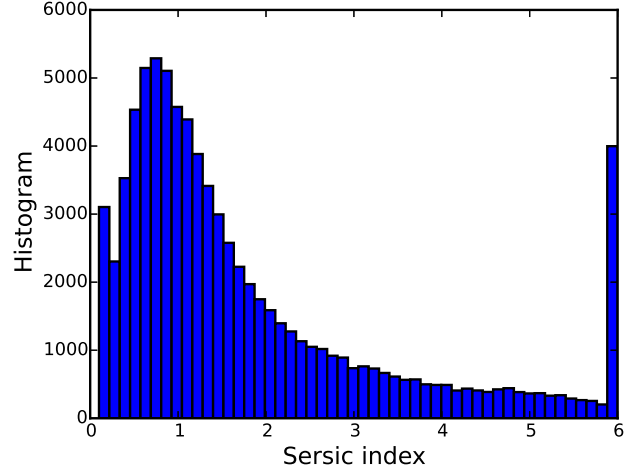
cuts by selecting the galaxies surviving each cut and performing an independent computation of the residual bias. A cut on  $S/N_{\text{flux}} > 7$  increases the bias slightly from  $-0.0073 \pm 0.0020$  to  $-0.0109 \pm 0.0018$  for  $m_1$  and from  $-0.0067 \pm 0.0013$  to  $-0.0118 \pm 0.0013$  for  $m_2$ . Since we are still in the  $\sim 1\%$  bias regime this justifies the inclusion of noisier galaxies in the mass determination of the three relaxed clusters in Sect. 7, tightening the constraints via an increased source density. By lowering our  $S/N_{\text{flux}}$  cut from 10 to 7 we increase the number of galaxies in our simulation analysis by 15%.

We also investigated the dependence of the bias on the magnitude of the galaxies, finding no significant trend after applying the  $S/N_{\text{KSB}}$ -dependent correction (see Fig. 4). This is not surprising since magnitude and signal-to-noise ratios are strongly correlated. We also investigated the residual bias as a function of the measured galaxy flux radius, finding only a weak dependence (see Sect. 5.2), which can be safely ignored given our accuracy target, but see Hoekstra et al. (2015) for a correction scheme that also depends on galaxy size.

### 3.4. Light profile

One of the critical parameters is the light distribution of the galaxies. Since we are measuring the changes in ellipticities caused by the lensing effect, the intrinsic shapes significantly affect our results. When using a Sérsic profile to describe the light distribution (see Eq. 12), larger values of  $n$  describe a more centrally-concentrated profile. These differences in concentration can affect biases in the moment measurement and the PSF correction even if galaxies with the same ellipticity are simulated.

In order to investigate the dependencies on the light profile we generated four sets of simulations with different Sérsic index distributions. These included a set of pure exponentials ( $n = 1$ ), a set of pure De Vaucouleurs ( $n = 4$ ), a set using a flat distribution (uniform Sérsic index distribution between 0.3 and 6), and the more realistic setup employed in our reference simulation. The latter uses a Sérsic index distribution (see Fig. 5) based on Sérsic profile fits to COSMOS galaxies as provided by GALSIM



**Fig. 5.** Distribution of Sérsic indices in the parametric fit to real colour-selected COSMOS galaxies. The peak in the last bin is due to the limits in the index allowed by GALSIM. Everything larger than the GALSIM maximum Sérsic index (6.2) is added to that bin.

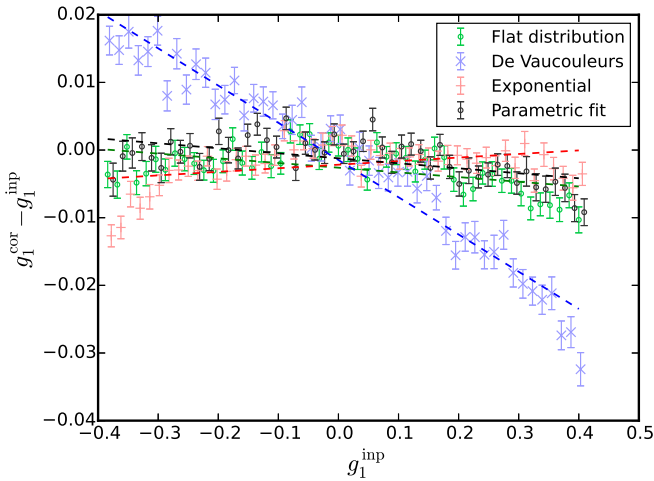
(Mandelbaum et al. 2012; Rowe et al. 2015). Importantly, we did not use their full sample of  $I_{814} < 25.2$  galaxies, but applied a  $V_{606} - I_{814} < 0.4$  colour selection to approximately match the selections applied in S18a and S20<sup>5</sup>. The resulting distribution is skewed towards lower Sérsic indices, with a median value of 1.24.

For all analyses we applied the noise bias correction calibrated on the reference simulation. Fig. 6 shows the resulting deviations between the recovered shear and the input shear as a function of the input shear. The lines indicate the linear fits according to Eq. (8), where the offsets and slopes correspond to the additive and multiplicative biases. The multiplicative bias differences are relatively small except for the simulation containing only De Vaucouleurs profiles, which exhibits a substantial residual multiplicative bias of  $\sim -5\%$ . From this we conclude that the input galaxy light profiles can play a relevant role, but that minor differences do not have a major impact.

We note that the COSMOS Sérsic index distribution was derived from a slightly brighter colour-selected galaxy sample. Accordingly, it might not exactly match the distribution of our fainter colour-selected source sample. Nevertheless we expect that the differences are small, with both populations being dominated by late-type galaxies. We used the bias difference  $|\Delta m| \approx 0.5\%$  between the fairly realistic simulation employing the parametric fits and the fairly unrealistic simulation using a flat distribution as a conservative estimate for the systematic uncertainty associated with the galaxy light profile assumptions (see Sect. 6).

When creating the mock galaxies with different light profiles we obtained slightly different distributions of the parameters shown in Fig. 1, especially for the size distribution. Performing a reweighting of the size distribution to match it to our CANDELS reference did not change the bias significantly, so the differences shown in Fig. 6 are not caused by these minor differences.

<sup>5</sup> For most of their clusters and source galaxies these studies employ a  $V_{606} - I_{814} < 0.3$  colour selection. The minor difference in the colour selection has a negligible impact on the Sérsic index distribution.



**Fig. 6.** Residual bias estimation for simulations with different input light profile distributions. The different symbols correspond to the different light profiles used to create the mock galaxies. We show the cases of a flat Sérsic index distribution in green, a purely De Vaucouleurs profile in blue, a purely exponential in red, and the more realistic case of the parametric fit to the COSMOS galaxies in grey. For the result shown here the  $S/N_{\text{KSB}}$ -dependent correction from Eq. (13) and the  $S/N_{\text{flux}} > 10$  cut have been applied. The causes of the deviation of some points at strong shear are discussed in Sect. 3.6. The indicated fits have been computed from the  $|g| < 0.2$  range, but we present the estimates for the full  $|g| < 0.4$  range.

### 3.5. Intrinsic ellipticity dispersion

All galaxies have some intrinsic ellipticity. For the reference simulations we used parametric fits to COSMOS galaxies to determine the galaxy input ellipticities, in order to have more realistic ellipticities. In this section, we test if the input intrinsic ellipticity distribution of the galaxies in the simulated images can play a significant role for the performance of the KSB+ algorithm (see e.g. Viola et al. 2014). To test this, we set up four sets of simulations with the same parameters, except for the intrinsic ellipticity. Each one had an input Gaussian ellipticity distribution for each component  $\epsilon_\alpha$ , with different RMS modulus  $\sigma(|\epsilon|)$  values ranging from 0.2 to 0.35 (computed from both ellipticity components together). The biases between different sets show only minor ( $< 0.5\%$ ) differences. These results are also consistent with the ones obtained for the reference simulation (which has a  $\sigma(|\epsilon|) = 0.28$ ), so we conclude this choice does not play a big role in the bias determination.

### 3.6. PSF shape and deviations at stronger shears

The KSB+ formalism makes the simplifying assumption that the PSF can be described as an isotropic function convolved with a small anisotropic kernel, which is not strictly met for many realistic PSFs. In fact, the Tiny Tim PSF we used has an ellipticity with  $|\epsilon| = 0.072$ . Here we therefore investigate how sensitively our bias estimates depend on the details of the PSF model that is employed in the simulation. For our simulations we used a typical PSF model created with Tiny Tim (Krist et al. 2011). We selected the PSF model parameters according to the best fit to real HST/ACS starfield images obtained by Gillis et al. (2019). In particular, we employed an ACS-like setup, with a subsampling by a factor 3 to avoid pixelation issues when convolving

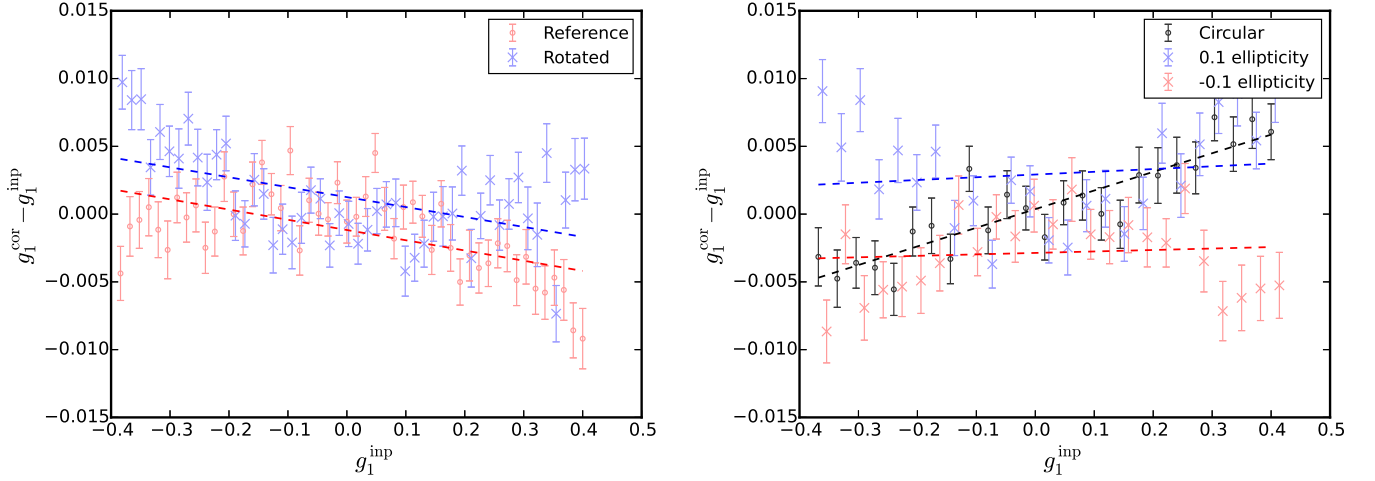
with the galaxy light profiles. The subsampled pixel scale for the PSF was  $0''.0165/\text{pixel}$ .

The ACS PSF changes due to focus variations caused by the thermal fluctuations that happen in orbit (Heymans et al. 2005; Rhodes et al. 2006; Schrabback et al. 2007). We tested the influence of this in our analysis, finding that the bias varies less than 1% within the typically expected focus ranges and spatial PSF variations across the field of view. For this reason, we did not model a position-dependent PSF and rather just took one particular model as input. As this default model we chose a typical average focus value of  $-1\mu\text{m}$ , as well as a position of  $x = 1000, y = 1000$  in chip 1, which leads to a PSF with ellipticity  $e_1 = 0.018, e_2 = 0.063$  (when measured with a weight function scale  $r_g = 2.0$  pixels).

Another effect that influences the results is charge diffusion (Krist 2003). Electrons near the edges of pixels have a chance to travel to neighbouring pixels, effectively creating a blurring effect which should be included in our models. This is implemented by using a Gaussian kernel. While testing this, we found that excluding this effect for our realistic PSF increased the residual bias to  $\sim -4\%$ . The more realistic charge diffusion-affected PSF was implemented into our simulations and compared to a Moffat profile to assess how incorrect PSF modelling affects our shape measurements. We scaled the Moffat PSF to a half-light radius of  $0''.07$ , which corresponds to the half-light radius of the ACS PSF used. The Moffat PSF changes the bias by  $\sim +0.018$ . The employed ACS PSF model is fairly realistic (see Gillis et al. 2019), while the Moffat PSF clearly only constitutes a crude approximation. We therefore consider half of this bias difference as an estimate for the systematic uncertainty associated with the realism of the PSF shape in the simulation (see Sect. 6).

For a number of consistency checks we also generated a set of simulations in which the PSF had been rotated by 90 degrees compared to the default reference simulation. For both setups the shear recovery is compared in the left panel of Fig. 7. After the  $S/N_{\text{KSB}}$ -dependent bias correction, we obtained similar parameters for the multiplicative bias ( $m_1 = -0.0076 \pm 0.0023$  and  $m_2 = -0.0069 \pm 0.013$ ) for the simulations with the rotated PSF, and a change in the sign for the additive bias ( $c_1 = 0.0012 \pm 0.0003$  and  $c_2 = 0.0045 \pm 0.0003$ ), compared to the bias in Table 2. This comparison allows us to investigate the cause for the deviations from a linear trend that can be seen at large input shears in Figs. 6 and 7. If this effect was caused by a non-linear response to shear, we would expect the deviation to be point-symmetric (with respect to the origin) for negative and positive shear values, due to the sign of the shear simply being the orientation (parallel or perpendicular) with respect to the component axis. For example, if the difference between the recovered and input shear is positive for positive shear values, then we should expect a negative difference for negative shear. However, in Fig. 7 the difference for the reference simulation is negative both for positive and negative shear values. This suggests that it is not a quadratic response to shear but rather a dependency of the additive bias on the input shear. This hypothesis is supported by the fact that the sign of the difference switches for the simulations with rotated PSF. To further test the origin of this effect, we set up three sets of simulations using a Moffat PSF with ellipticities 0.1, 0 and  $-0.1$ . As visible in the right panel of Fig. 7, the deviation from the linear behaviour depends on the ellipticity of the PSF, being non-existent for the circular PSF and having a different sign for  $-0.1$  and  $0.1$  PSF ellipticities. We conclude from these observations that our KSB+ implementation does not suffer from a non-linear shear response, but that the additive bias appears to be shear dependent in the non-weak shear regime.





**Fig. 7.** *Left:* Comparison of the residual bias from the reference simulations (red points) and a simulation in which the PSF has been rotated by 90 degrees (blue crosses). *Right:* Comparison of the residual bias obtained for a circular Moffat PSF in grey, a modified Moffat with an  $e_1 = 0.1$  ellipticity in blue, and a Moffat PSF with an  $e_1 = -0.1$  ellipticity in red. The indicated fits have been computed from the full  $|g| < 0.4$  range.

#### 4. Cluster blending

In order to increase the realism of our simulations, we took a first step by studying the impact that the presence of bright cluster members has on the bias estimation. For example, the wings of their extended light profiles might contaminate the light distribution of nearby sources, potentially leading to biased shape estimates. This is a cluster-specific issue that is not generally studied in this kind of work. Different from bright foreground galaxies, which are randomly positioned, cluster galaxies have a higher number density in the cluster cores where shears are stronger.

For this analysis, we created two sets of simulations, one containing only lensed background galaxies, and one that also contains cluster member galaxies and other foreground galaxies. For the background galaxies we used the same galaxy properties as in Sect. 3, but we placed them randomly in the image rather than on a grid. We note that this is not yet a fully realistic scenario, since the simulations are still missing the clustering of source galaxies, as well as very faint galaxies below the detection threshold. These effects are studied in Sect. 5, but are not needed for the current step, where we are investigating the impact of bright cluster and foreground galaxies. The  $S/N_{\text{KSB}}$ -dependent correction as well as the  $S/N_{\text{flux}} > 10$  cut were implemented in the analysis, although we alternatively repeated the same analysis with  $S/N_{\text{flux}} > 7$ , which changed the results only marginally (the final bias estimation is presented later in Table 5). Instead of images with constant shear we used a more realistic approach where background galaxies were sheared according to their relative position to the centre of the cluster assuming a spherically-symmetric NFW profile (Navarro et al. 1997) with the overdensity mass  $M_{200c} = 5 \times 10^{14} M_{\odot}$ <sup>6</sup>, the concentration  $c_{200c} = 4$ , the redshift of the lens  $z_{\text{lens}} = 0.3$ , and of the source  $z_{\text{source}} = 0.6$ <sup>7</sup>. We still created an identical image with a 90 degree rotation of the intrinsic ellipticity of each galaxy before the shear was ap-

plied, which is matched in the analysis to reduce shape noise, and therefore the number of simulations required to reach the desired precision on the bias estimate. For the properties of the cluster member galaxies we used catalogues from the MAGELLAN/PISCO (Stalder et al. 2014) follow-up of SPT clusters of various redshifts ( $z = 0.3 - 1.1$ , see Bleem et al. 2019), using cluster redshifts from Bocquet et al. (2019). The cluster galaxies have  $r$ -band magnitudes from  $\sim 18$  to  $\sim 24$ , which are brighter than our background galaxies, and they have a slightly lower input ellipticity dispersion  $\sigma(|e|) = 0.2$ . We created masks around very bright and extended objects, mostly cluster members, following the analysis of real clusters.<sup>8</sup> A simulated image for one of the clusters (at  $z = 0.72$ ) and an example cut-out of the background-only, the background+cluster members image, and the corresponding mask are shown in Fig. 8 (see Fig. B.1 for a simulated cluster at  $z = 0.28$ ). We created multiple images with different realisations of background galaxies and different cluster catalogues in order to stack the profiles and obtain a more significant result, independent of the particulars of each cluster.

Fig. 9 shows the mean input and recovered shear profiles for the two simulation setups (with and without bright galaxies) averaged over all clusters with  $z > 0.7$  in the top panel. The bottom panel shows the relative difference between the two recovered profiles, which corresponds to the change in the multiplicative bias caused by the presence of the bright foreground and cluster galaxies. The same relative difference is shown as a function of cluster redshift in Fig. 10, split into four distance intervals.

We generally find that adding the bright foreground and cluster members only has a minor impact on the shear recovery (see Table 3). The biggest impact is detected for lower redshift clusters ( $z < 0.7$ ) at scales 70 – 100 arcsec, amounting to a  $1.13\% \pm 0.33\%$  positive multiplicative bias. We expect that the impact decreases for higher redshift clusters given the stronger cosmological dimming of their cluster members. Indeed, computed over one broad bin between 70–165 arcsecs we find a very minor bias of  $0.48\% \pm 0.38\%$  for the simulated  $z > 0.7$  clusters, which approximately corresponds to the scales and redshifts

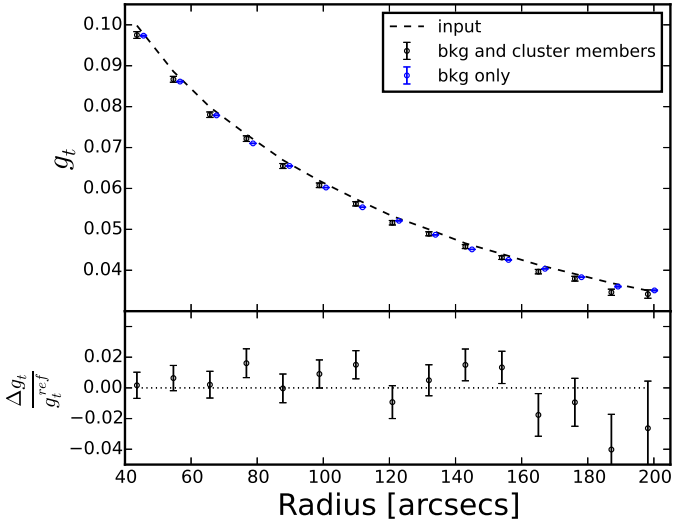
<sup>6</sup> We note that  $M_{\Delta c}$  corresponds to the mass contained within a sphere of radius  $r_{\Delta c}$ , inside of which the mean density equals to  $\Delta$  times the critical density of the Universe at the redshift of the lens.

<sup>7</sup> This configuration corresponds to typical measurements of low-redshift clusters using ground-based weak lensing data, but the resulting level of shears is in fact comparable to typical HST measurements for high-redshift clusters (compare Fig. 16).

<sup>8</sup> To create this mask, we perform a SExtractor object detection, requiring 2000 pixels with a flux that is higher than the background by  $0.5\sigma$ . We then extend the mask in all directions by 3 pixels. This will exclude areas around the brightest objects from our shear estimates.

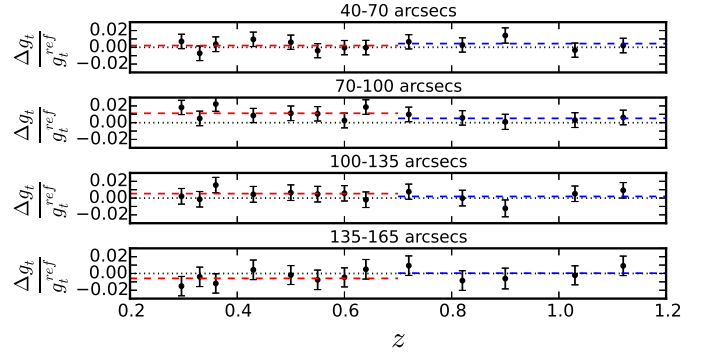


**Fig. 8.** Example image of a simulated cluster at  $z = 0.72$ . A cut-out of the full image, shown in red, can be seen in the right for the simulations with background galaxies only (top), with added cluster members (middle), and showing the mask used to remove bright objects (bottom). The full image and the cut outs span  $300'' \times 300''$  and  $100'' \times 55''$ , respectively. A similar image for a simulated low-redshift cluster is provided in Fig. B.1 in the Appendix.



**Fig. 9.** Averaged tangential shear profile of multiple simulations of background galaxies sheared following an NFW profile with (in black) and without (in blue) the presence of cluster member galaxies. Here we average the profiles of all simulated clusters at  $z > 0.7$ . The dashed line represents the average input profile. The bottom panel shows the relative difference between the tangential shears of the simulations with and without cluster member galaxies as a function of radius.

used in S18a. When applying a  $S/N_{\text{flux}} > 7$  selection, this bias shifts to  $0.25\% \pm 0.40\%$ .



**Fig. 10.** Relative difference of the measured tangential shears of the simulations with and without cluster member galaxies as a function of cluster redshift. The four panels show the values in different distance bins. The dotted black line represents the zero line for reference and the dashed line shows the mean of the points for clusters at  $z < 0.7$  (red) and  $z > 0.7$  (blue).

Given the redshift dependence of the bias caused by cluster galaxies, we decided to treat it separately, and not include the cluster galaxies in the simulations that are used in Sect. 5 to investigate the impact of nearby fainter galaxies and selection effects. We verified that the presence of bright galaxies does not lead to a significant shift in the estimates of selection bias for our method.

**Table 3.** Detailed estimate of the bias due to the presence of bright galaxies for different cluster-centric distance bins and two redshift bins. These estimates are obtained using galaxies with  $S/N_{\text{flux}} > 10$ .

Distance	Cluster redshift	Bias estimate
40''-70''	$z < 0.7$	$0.0021 \pm 0.0033$
40''-70''	$z > 0.7$	$0.0044 \pm 0.0039$
70''-100''	$z < 0.7$	$0.0113 \pm 0.0033$
70''-100''	$z > 0.7$	$0.0052 \pm 0.0039$
100''-135''	$z < 0.7$	$0.0054 \pm 0.0035$
100''-135''	$z > 0.7$	$0.0020 \pm 0.0042$
135''-165''	$z < 0.7$	$-0.0058 \pm 0.0043$
135''-165''	$z > 0.7$	$0.0003 \pm 0.0053$

## 5. Selection bias and the influence of blends and neighbours

Hoekstra et al. (2017) and Euclid Collaboration: Martinet et al. (2019) demonstrate that faint sources below the selection threshold affect shape measurements. Using a Euclid-like setup selecting galaxies with  $i < 24.5$  and accounting for realistic galaxy and clustering properties calibrated using Hubble Ultra Deep Field data (HUDF, Beckwith et al. 2006), Euclid Collaboration: Martinet et al. (2019) show that these faint sources cause an additional shape measurement multiplicative bias for our KSB+ implementation of  $\Delta m = -0.0149 \pm 0.0002$ . In Sect. 3 we placed the galaxies on a grid, which avoids any contamination by neighbours and any selection bias effect since we require matched pairs. In Sect. 4 we improved the realism of our simulation by adding bright galaxies, but the background galaxies were placed randomly and without a realistic clustering. In real images, we have a larger number of partially blended galaxies and nearby galaxies due to the fact that faint galaxies are clustered together, which was not included in the previous setups. As an example, for deep optical data with ground-based resolution, Mandelbaum et al. (2018) finds that the impact of nearby galaxies may affect the shape calibration at the  $\sim 10\%$  level. Similarly, for the Dark Energy Survey Samuroff et al. (2018) finds that neighbours can affect multiplicative biases by 3–9%. In our pipeline, to be conservative, we used masks to remove bright and extended objects and additionally apply a neighbour rejection, which reduces the impact of neighbours. If two galaxies were detected in the catalogue with a separation  $< 0''.75$  we only kept the brighter one. We expect that this, together with the fact that we were simulating high resolution data, reduces the impact of neighbours compared to previous works, but we tested it here nevertheless.

In this section we created a more realistic scenario with blended galaxies and contamination from neighbours as well as clustering of galaxies. We placed the mock galaxies following the positions and magnitudes of the galaxies in the Skelton et al. (2014) 3D-HST/CANDELS catalogues. This was done to account for a realistic clustering of the galaxies. As mentioned before, we did not include here the effect of bright galaxies from Sect. 4 given the dependence on cluster redshift. Instead, we compute the resulting net bias for the correction of the real data in Sect. 6, adding all the different contributions.

The 3D-HST/CANDELS catalogues used here are NIR detected, but the F606W number counts indicate that they do not suffer from significant incompleteness until  $V_{606} \sim 27.5$ , which is 1 magnitude fainter than the  $V_{606} = 26.5$  limit of our source sample. We therefore expect that they allow us to capture most of the impact of faint galaxies, but we revisit this issue in Sect. 5.3, using models that are based on much deeper HUDF data.

We used the same galaxy property setup as in Sect. 3. This means we obtained the Sérsic indices and ellipticity from parametric fits to COSMOS data, which provide input parameters for galaxies in the F814W band. We employed this approach due to the lack of a comprehensive catalogue of structural parameters measured in the same F606W band and we are simulating here, expecting that the impact of the minor differences between these two bands is small (following the results from Sect. 3.4). Other catalogues which have measured light profiles of galaxies, such as the one presented by van der Wel et al. (2012), provide structural parameters for CANDELS in the F160W and F125W bands, which are much more different from F606W. This is the reason why we continued using the parametric fits available in GALSIM.

We used different patches within the CANDELS fields to provide different realisations of positions and magnitudes. An example of how the simulations look when compared to the real images can be seen in Fig. 11, where we show a cutout of the GOODS North field in the real images and in our simulation. The sizes and ellipticities differ, but the positions and magnitudes are comparable. Some objects vary since we use a NIR-detected catalogue. When creating the mock images, for the galaxies that are within our colour cuts ( $V_{606} - I_{814} < 0.4$ ) and magnitude range ( $V_{606} < 26.5$ ), we stored their positions in order to later select them for our analysis. This guarantees that the galaxies used for the shear estimation are indeed similar to what is used in real images. Fainter galaxies were also included in the simulation as present in the Skelton et al. (2014) catalogue. Importantly this catalogue is deeper than the single-orbit-depth shape catalogue from S18a, which was used to define the inputs for the simulations described in Sect. 3. As a result, this reduces the incompleteness in the input catalogue, leading to a better match in the recovered distributions between the simulation analysis and the CANDELS analysis, especially in terms of  $S/N_{\text{flux}}$  (compare Figs. 1 and 13).

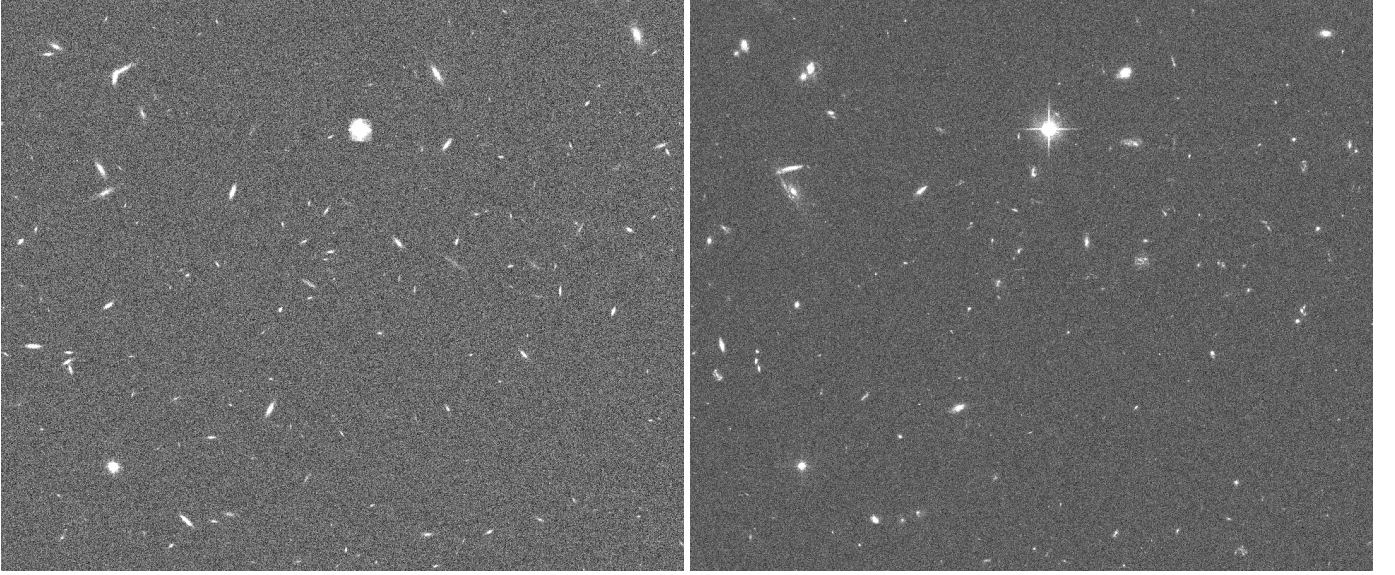
With this analysis setup we find a small  $\sim -0.7\%$  shift in the multiplicative bias due to the use of realistic positions compared to the grid-based analysis. This however still lacks the impact of selection bias and very faint  $V_{606} > 27.5$  galaxies, which are accounted for in the following subsections.

While the effect of neighbours has a relevant impact on our analysis, its impact is at a much smaller level than what was found in previous work (Mandelbaum et al. 2018) for ground-based images. This is likely due to the better resolution of the HST images, leading to a weaker impact of the blending and neighbours. In addition, our neighbour rejection may contribute to the lower impact.

### 5.1. Selection bias

Weak lensing is based on the assumption that the orientation of the intrinsic ellipticity of the galaxies is random. When we preferentially select galaxies aligned with, or orthogonally to, the shear direction we introduce a selection bias into our sample. This affects our measured shear (Heymans et al. 2006) and subsequently the mass estimation. The selection bias can come from different sources and in this section we aim to disentangle the different steps in the shape analysis and their impact in the bias determination. In the previous sections the selection bias was neglected as we required matched pairs with opposite intrinsic ellipticities, which artificially removed any preferential selection. However, selection bias can be important (e.g. Kannawadi et al. 2019). For this reason we present here a step-by-step analysis of the selection bias alone before obtaining a joint estimation of the





**Fig. 11.** Comparison between the simulated images from GOODS North (left) and the real images (right). The sizes and ellipticities are different between them, but the positions and magnitudes are matched individually.

residual bias and the selection bias in Sect. 5.4, as well as their joint signal-to-noise ratio dependence.

In order to measure the selection bias we identified the selected galaxies and computed their average intrinsic ellipticity (from the input catalogues). Any deviation from zero indicates a preferential selection of galaxies depending on their intrinsic ellipticity. We compared the average input ellipticity independently for each input shear and fit a linear relation in order to constrain the shear dependence. Unlike what we did when estimating the residual bias, we did not require detection in both the normal image and the rotated one, since this is not what happens in real images. Both normal and rotated images were still included in order to cancel shape noise for the galaxies detected in both (these galaxies do not contribute to the selection bias), tightening the constraints.

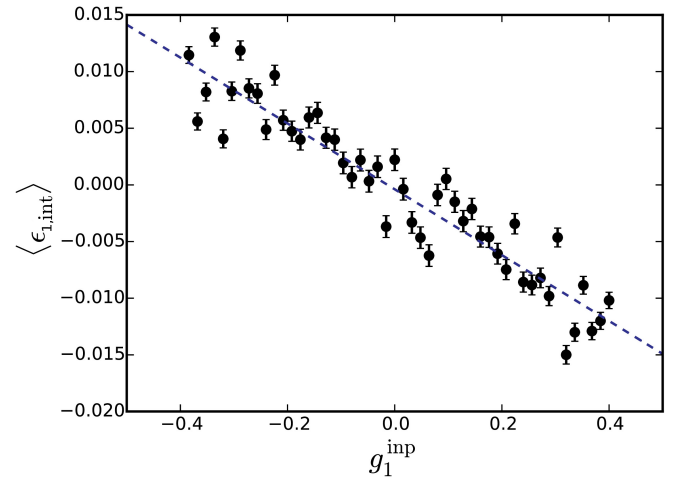
It is important to measure the selection bias using realistic positions and not with the galaxies placed on a grid since the pipeline might introduce a bias by rejecting galaxies from the analysis due to the neighbours. The selection bias that was present after every step is summarised in Table 4.

**Table 4.** Selection bias after each step in the analysis pipeline.

Step	$m_{1,\text{sel}}$	$m_{2,\text{sel}}$
SExtractor detection	$-0.0291 \pm 0.0015$	$-0.0266 \pm 0.0018$
+ $S/N_{\text{flux}} > 10$ cuts	$-0.0259 \pm 0.0020$	$-0.0259 \pm 0.0014$
+ Neighbour rejection	$-0.0211 \pm 0.0071$	$-0.0194 \pm 0.0056$
+ Final KSB+ cuts	$-0.0138 \pm 0.0020$	$-0.0174 \pm 0.0013$

### 5.1.1. SExtractor object detection

We used SExtractor to detect the objects and create a catalogue<sup>9</sup>. Fig. 12 shows the mean intrinsic ellipticity of the galaxies passing the object detection (and our colour and magnitude selection) as a function of the input shear. A linear fit to these



**Fig. 12.** Selection bias on the first component after the SExtractor detection as a function of the input shear.

points yields the selection bias, which is  $m_{1,\text{sel}} = -0.0291 \pm 0.0015$  and  $m_{2,\text{sel}} = -0.0266 \pm 0.0018$ , similar to the results from Kannawadi et al. (2019). This bias does not depend on the actual shear measurement algorithm as it comes directly from the detection software (in our case SExtractor) and therefore it is a more general issue. A main reason for this negative selection bias is likely the isotropic kernel, with which the image is convolved during the SExtractor detection stage. Thus, round galaxies are more likely to be detected. However, further factors such as the deblending may also play a role.

### 5.1.2. SExtractor S/N cut

We selected galaxies for our analysis that had  $S/N_{\text{flux}} > 10$ . This can also introduce additional selection bias. After the cut, the net selection bias was  $m_{1,\text{sel}} = -0.0259 \pm 0.0020$  and  $m_{2,\text{sel}} = -0.0259 \pm 0.0014$ . We should note that since the cuts were performed after the object detection, the selection bias is cumulative. This indicates that our  $S/N_{\text{flux}}$  selection actually leads to a

<sup>9</sup> Here the most important settings are FILTER\_NAME=gauss\_2.5\_5x5.conv, DETECT\_THRESH=1.4, DETECT\_MINAREA=8, DEBLEND\_NTHRESH=32, and DEBLEND\_MINCONT=0.01.



slight positive bias which partially corrects for the bias in the SExtractor object detection. This change is marginal, however, which indicates that the exact signal-to-noise cut we used does not have a large impact. In fact, for the alternative cut of galaxies with  $S/N_{\text{flux}} > 7$  we get  $m_{1,\text{sel}} = -0.0279 \pm 0.0010$  and  $m_{2,\text{sel}} = -0.0266 \pm 0.0010$ .

### 5.1.3. Rejection of very close neighbours

For the shape measurement, we performed a selection of objects without bright close neighbours in the detection catalogue. We rejected objects with a brighter neighbour closer than  $0''.75$ . When including this effect we obtained  $m_{1,\text{sel}} = -0.0211 \pm 0.0071$  and  $m_{2,\text{sel}} = -0.0194 \pm 0.0056$ . This means that this step also partially compensates for the original bias.

### 5.1.4. Final catalogues after KSB+ cuts

The shape measurement algorithm introduces certain cuts in order to robustly measure the shear which can also introduce artificial biases. The results indicate that there is a net selection bias in the final catalogues of  $m_{1,\text{sel}} = -0.0138 \pm 0.0020$  and  $m_{2,\text{sel}} = -0.0174 \pm 0.0013$  which is again smaller than the bias obtained in Sect. 5.1.3, suggesting that this step partially compensates for it. A small, but important selection bias is still present in our analysis. For the case when  $S/N_{\text{flux}} > 7$  we obtained a final selection bias estimate of  $m_{1,\text{sel}} = -0.0150 \pm 0.0018$  and  $m_{2,\text{sel}} = -0.0180 \pm 0.0012$ .

## 5.2. Joint correction for shape measurement and selection bias based on the CANDELS-like simulations

In this section we present a joint measurement of the multiplicative bias and the selection bias for our most realistic simulation using the CANDELS positions and magnitudes. These two effects cannot be separated perfectly, which is why it is best to constrain them jointly. Also, the selection bias may have some dependence on  $S/N_{\text{KSB}}$ , which is why we have to verify if the fit obtain in Sect. 3.2 still describes the  $S/N_{\text{KSB}}$  dependence well.

A combined analysis of both effects in 10  $S/N_{\text{KSB}}$  bins is shown in the left panel of Fig. 14. The estimates are more noisy in this case since the shape noise cancellation is less effective given that we no longer require matched pairs.

We find that the dependence of the joint bias on  $S/N_{\text{KSB}}$  is still well described by the power-law fit obtained in Sect. 3.2 if we add a constant bias offset, for which we obtained best-fit parameters

$$\begin{aligned} m_{1,\text{corr}} &= -0.358 (S/N_{\text{KSB}})^{-1.145} - 0.042, \\ m_{2,\text{corr}} &= -0.357 (S/N_{\text{KSB}})^{-1.298} - 0.039. \end{aligned} \quad (14)$$

Using this correction we obtained a residual bias of  $m_{1,\text{res}} = 0.0010 \pm 0.0040$  and  $m_{2,\text{res}} = -0.0023 \pm 0.0043$  for  $S/N_{\text{flux}} > 10$  and  $m_{1,\text{res}} = -0.0032 \pm 0.0049$  and  $m_{2,\text{res}} = -0.0047 \pm 0.0045$  for  $S/N_{\text{flux}} > 7$ . This contains the effect of the addition of faint galaxies as present in the Skelton et al. (2014) catalogues as well as the selection bias.

In the right panel of Fig. 14 we show the dependence of the bias on the FLUX\_RADIUS after the  $S/N_{\text{KSB}}$ -dependent correction from Eq. (14) is applied. We find that this dependence is very weak when both shape measurement and selection biases are taken into account, confirming that it is sufficient to apply a  $S/N_{\text{KSB}}$  dependent correction. As a cross-check we conducted a reweighting analysis, finding that the discrepancy in the

FLUX\_RADIUS distributions between the data and simulations affect the bias at a negligible  $-0.3\%$  level only.

## 5.3. Addition of faint galaxies following Euclid Collaboration: Martinet et al. (2019)

The CANDELS-like simulations provide a good approximation of the real data. However, Euclid Collaboration: Martinet et al. (2019) finds that galaxies up to 2 magnitudes fainter than the studied source sample need to be included in the simulations in order to fully account for the impact of neighbours. With the CANDELS catalogues we only included galaxies up to approximately 1 magnitude fainter. In the present analysis we simulated images in the F606W band and applied different magnitude and colour cuts, and therefore cannot directly employ the findings from Euclid Collaboration: Martinet et al. (2019). Instead, we remeasured the clustering and galaxy properties in the HUDF F606W data closely following Euclid Collaboration: Martinet et al. (2019), but using galaxies with  $24.0 < V_{606} < 26.5$  and  $V_{606} - i_{775} < 0.3^{10}$  as 'bright' sample; and surrounding sources with  $26.5 < V_{606} < 28.5$  as 'faint' sample. Based on the resulting distributions we injected faint galaxies into our grid-based simulations and remeasured the shape measurement biases. In order to study the increase in the bias due to the addition of galaxies with  $27.5 < V_{606} < 28.5$ , we computed the bias in two different setups. The first one includes faint galaxies in the magnitude range  $26.5 < V_{606} < 28.5$ , while the other one only includes faint galaxies with  $26.5 < V_{606} < 27.5$ . The shift in the bias between both setups needs to be added to the residual bias found in Sect. 5.2. This is a contribution of  $\Delta m_1 = -0.0037 \pm 0.0055$  and  $\Delta m_2 = -0.0060 \pm 0.0065$  for  $S/N_{\text{flux}} > 10$  and  $\Delta m_1 = -0.0061 \pm 0.0036$  and  $\Delta m_2 = -0.0070 \pm 0.0031$  for  $S/N_{\text{flux}} > 7$ .

As a consistency check we also verified that the setup of a grid-based simulation with added  $26.5 < V_{606} < 27.5$  HUDF-like galaxies yields shape measurement biases that are consistent with the results obtained using the CANDELS-like simulations. Also note that, in total, the full set of faint galaxies ( $26.5 < V_{606} < 28.5$ ) causes a shift in the shape measurement bias by  $\sim -1.3\%$ , as estimated with the HUDF-informed analysis.

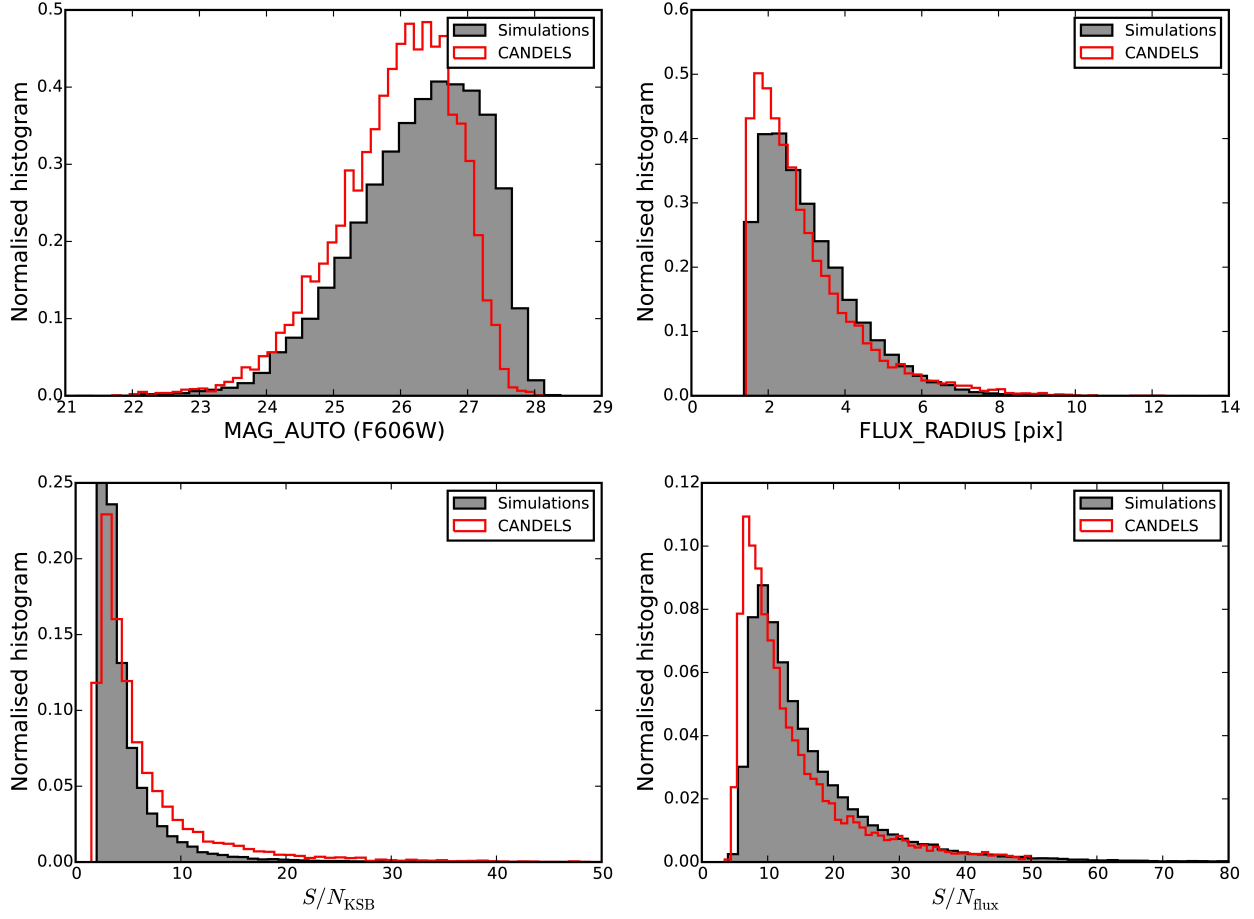
## 5.4. Selection bias due to the addition of cluster members.

Employing the simulations analysed in Sect. 4 we can estimate how much the addition of cluster members affects the selection bias. Fig. 15 shows the difference in the selection bias for the simulations with cluster members and the background-only simulations as a function of cluster-centric distance. Averaged over all scales we find that the addition of cluster members yields a slightly positive selection bias of  $0.0085 \pm 0.0024$  if all cluster redshifts are considered, and  $0.0060 \pm 0.0029$  for the clusters at  $z > 0.7$  and  $S/N_{\text{flux}} > 7$  galaxies, which we add to our total residual bias estimation in Sect. 6.

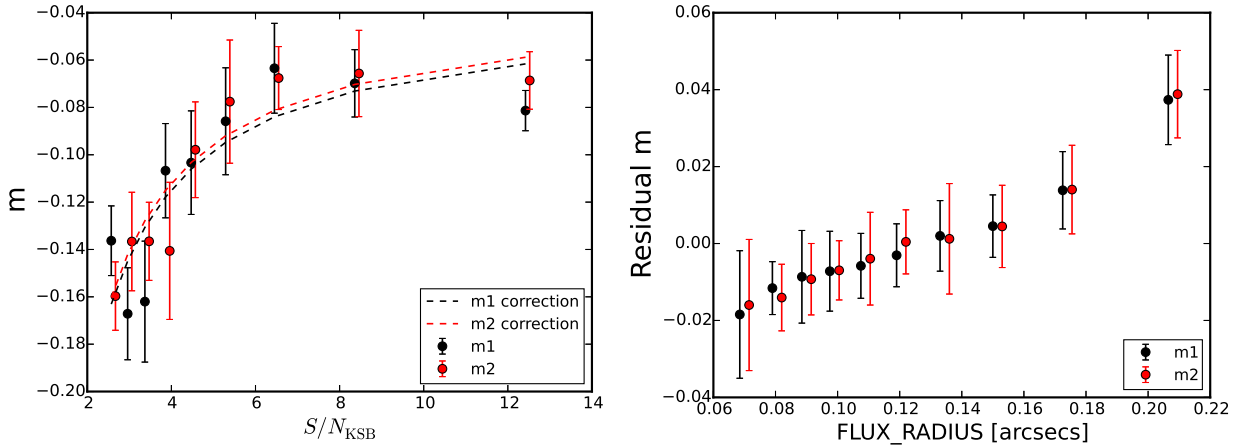
## 6. Summary of the bias estimates

The bias estimate from Sect. 5 was obtained from the CANDELS-like simulations that closely resemble the real images. However, as mentioned before, this lacks bright cluster members, as well as faint ( $V_{606} > 27.5$ ) neighbours. In this section we present a summary of the different contributions and an

<sup>10</sup> We note that this is a slightly different cut than used in the rest of this work, but this should not impact the results significantly.



**Fig. 13.** Comparisons between the measured distributions in our CANDELS-like simulations and the KSB+ CANDELS distribution for the F606W magnitudes, the half-light radius, the KSB signal-to-noise ratio  $S/N_{\text{KSB}}$ , and the SExtractor signal-to-noise ratio  $S/N_{\text{flux}}$ .



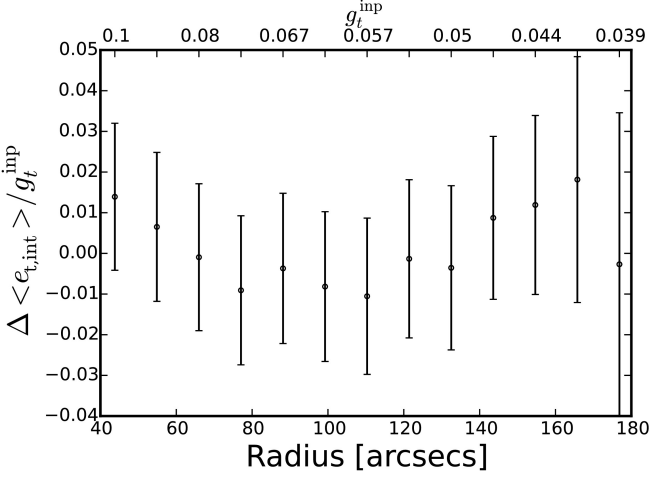
**Fig. 14.** *Left:* Dependence of the bias on  $S/N_{\text{KSB}}$  computed without correction or matching pairs from the CANDELS-like simulations, which then also includes the impact of selection bias. The dashed lines correspond to the correction from Eq. (14) for each component. *Right:* Dependence of the residual bias on the  $\text{FLUX\_RADIUS}$  estimated from the CANDELS-like simulations. Here, we have applied the  $S/N_{\text{KSB}}$ -dependent correction from Eq. (14). The second component is slightly shifted for visualisation purposes.

estimate of the final remaining biases and uncertainties after the correction in Eq. (14) is applied. The contribution of the general selection bias is not separately reported in the table, but it is included in the bias calibration and residual bias estimates from the CANDELS-like simulations (Sect. 5.2). We computed the bias in all cases for  $|g| < 0.2$  and for two different signal-to-noise ra-

tio cuts ( $S/N_{\text{flux}} > 10$  and  $S/N_{\text{flux}} > 7$ ). In order to obtain robust estimates of the uncertainties caused by limitations of our simulations we summarised the most important contributions (explained below) as 'Other modelling uncertainties' in Table 5 (all added in quadrature). These include the differences in the multiplicative bias from a flat Sérsic index distribution and the more

**Table 5.** Summary of the contribution to the bias and its uncertainties from the different effects for the two  $S/N_{\text{flux}}$  cuts used in this paper. We present a separated bias for each shear component for all effects except for the light contamination due to bright galaxies, since that is calculated from the tangential shear, which depends on both components. The modelling errors include the uncertainties related to variations in the PSF shape and the galaxy light profile distributions.

Contribution	$S/N_{\text{flux}} > 10$		$S/N_{\text{flux}} > 7$	
	$m_1$	$m_2$	$m_1$	$m_2$
Residual bias after CANDELS-like calibration	$0.0010 \pm 0.0040$	$-0.0023 \pm 0.0043$	$-0.0032 \pm 0.0049$	$-0.0047 \pm 0.0045$
Bias from extra faint $V_{606} > 27.5$ galaxies	$-0.0037 \pm 0.0055$	$-0.0060 \pm 0.0065$	$-0.0061 \pm 0.0046$	$-0.0070 \pm 0.0041$
Bright cluster galaxy contribution ( $z > 0.7$ ): Shape measurement bias	$0.0048 \pm 0.0038$		$0.0025 \pm 0.0040$	
Bright cluster galaxy contribution ( $z > 0.7$ ): Selection bias	$0.0053 \pm 0.0019$		$0.0060 \pm 0.0029$	
Other modelling uncertainties (simulation mismatch effects)	$\pm 0.0128$	$\pm 0.0130$	$\pm 0.0126$	$\pm 0.0131$
Total residual bias + uncertainty	$0.0074 \pm 0.0151$	$0.0017 \pm 0.0157$	$-0.0008 \pm 0.0151$	$-0.0032 \pm 0.0153$



**Fig. 15.** Relative difference in the selection bias estimates for the background only and the background+cluster members simulations for clusters at  $z > 0.7$  and  $S/N_{\text{flux}} > 7$  galaxies.

realistic parametric fits (see Sect. 3.4) as a conservative estimate for the uncertainties in our galaxy models. The Tiny Tim PSF model should provide a good match to the actual PSF shapes in HST data. However, as the bias shows some dependence on the PSF shape (see Sect. 3.6), we additionally included half of the bias difference from the analysis using a Moffat PSF (which is clearly not a good model for HST) as an additional error contribution. To account for the slight dependence of the multiplicative bias on the PSF ellipticity (see Sect. 3.6) we also added half of the bias difference between the setups of a circular and an elliptical ( $|e| = 0.1$ ) Moffat PSF to the systematic error budget.

The final residual bias estimate and uncertainty is  $m_1 = 0.0074 \pm 0.0151$ ,  $m_2 = 0.0017 \pm 0.0157$  for  $S/N_{\text{flux}} > 10$  galaxies and  $m_1 = -0.0008 \pm 0.0151$ ,  $m_2 = -0.0032 \pm 0.0153$  for  $S/N_{\text{flux}} > 7$  galaxies. These residual biases need to be corrected for when analysing real data, as done in Sect. 7. We find that after applying these corrections, multiplicative shape measurement biases should be accurately calibrated to the 1.5% level.

## 7. HST weak lensing measurements of three highly relaxed clusters from the South Pole Telescope Sunyaev-Zel'dovich Survey

To demonstrate the use of weak lensing measurements in the inner cluster and stronger shear regime we studied three clusters from the 2,500 deg<sup>2</sup> South Pole Telescope Sunyaev-Zel'dovich (SPT-SZ) Survey (Bleem et al. 2015). According to their X-ray properties these three clusters are among the most relaxed clus-

ters found in the SPT-SZ sample (McDonald et al. 2019). These clusters are part of a larger sample of distant SPT-SZ clusters analysed using HST weak lensing data (S18a and S20). Since general cluster samples are more strongly affected by miscentring and substructure in the inner cluster regions, S18a and S20 excluded the cluster cores from their analysis and only incorporated scales  $r > 500$  kpc when fitting radial shear profiles. As these astrophysical limitations are less severe for relaxed clusters, we can use these clusters as a test case to investigate how much constraints can be tightened by including weak lensing measurements from the cluster cores.

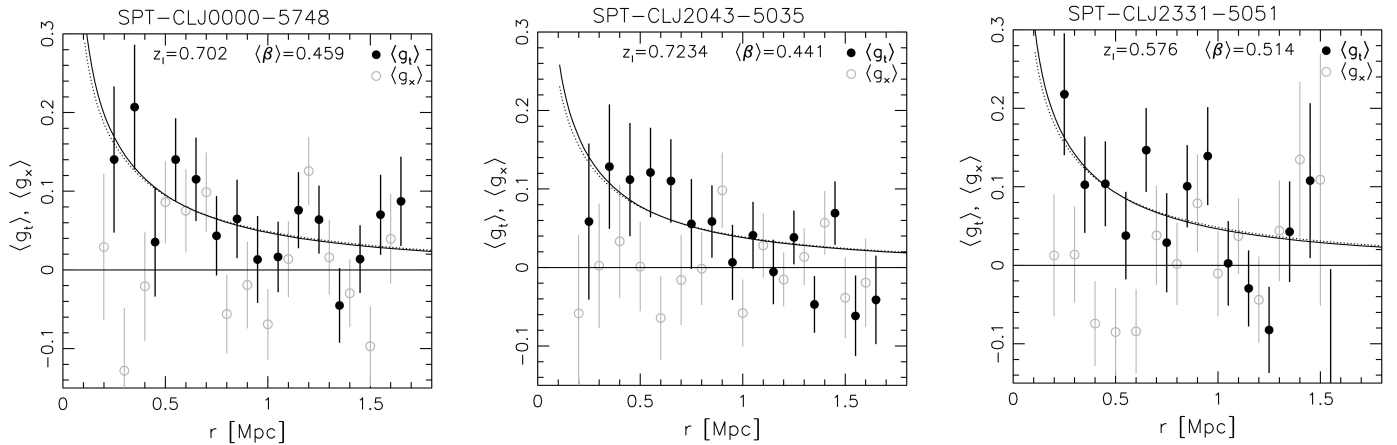
SPT-CL J0000–5748 ( $z = 0.702$ ) and SPT-CL J2331–5051 ( $z = 0.576$ ) were initially studied by S18a, who measured weak lensing shapes in  $2 \times 2$  HST/ACS mosaic  $V_{606}$  images and selected mostly background galaxies using  $V - I$  colour. For the source selection they employed HST  $I_{814}$  imaging in the cluster core and VLT  $I$ -band imaging in the cluster outskirts. S20 recently updated these measurements using a revised reference sample for the calibration of the source redshift distribution (Raihan et al. 2020) and employing our revised shear calibration as summarised in Sect. 6. S20 also incorporate deeper VLT  $I$ -band imaging for the source selection for SPT-CL J0000–5748. SPT-CL J2043–5035 ( $z = 0.723$ ) was studied in S20 using the same calibrations, employing shape measurements from mosaic ACS  $V_{606}$  images and a source selection that incorporates mosaic ACS  $I_{814}$  imaging for the full cluster field. We refer the reader to these publications for further details on the data, shape measurements, source selection, calibration of the source redshift distribution, and fitting procedure.

As primary difference to the previous work we included smaller scales ( $r > 200$  kpc)<sup>11</sup> in the analysis of the reduced shear profiles of these clusters (see Fig. 16). Following S18a and S20 we fitted NFW shear profile models (Wright & Brainerd 2000) to the tangential and cross shear

$$\begin{aligned} g_t &= -g_1 \cos 2\Phi - g_2 \sin 2\Phi, \\ g_\times &= g_1 \sin 2\Phi - g_2 \cos 2\Phi, \end{aligned} \quad (15)$$

where  $\Phi$  is the azimuthal angle with respect to the centre. The cross shear component serves as a test for systematics. For each of our targets it is consistent with zero within  $1\sigma$  when averaged over all radial bins (compare Fig. 16). We accounted for the magnitude dependence of the mean geometric lensing efficiency of the sources and the impact of weak lensing magnification on the source redshift distribution. S20 employ a fixed concentration–mass relation from Diemer & Joyce (2019, D19 henceforth). Assuming the same relation in a first step, we can estimate by how much the cluster mass constraints tighten when including the information from the inner cluster regions. We list

<sup>11</sup> This limit ensures that only reduced shear estimates in the robustly calibrated  $|g| \lesssim 0.2$  regime are included in the fit (compare Fig. 16).



**Fig. 16.** Reduced shear profiles around the X-ray centres of the three relaxed clusters analysed in this study, showing the tangential (black solid circles) and cross (grey open circles) components. The curves correspond to the best-fitting NFW models assuming the D19  $c(M)$  relation (dotted) and increased concentrations  $c_{200c}^{D19} = 1.14c_{200c}$  (solid), which are expected to more accurately represent relaxed clusters.

the mass signal-to-noise ratio  $S/N_{\text{mass,D19}} = M_{200c}/\Delta M_{200c}$  (considering only shape-noise uncertainties) for both fit ranges and all three clusters in Table 6, finding that it improves on average by a factor of 1.38 when scales  $r > 200$  kpc are used.

Table 7 lists the mass constraints obtained when including the inner cluster regions ( $r > 200$  kpc) and assuming the D19 concentration–mass relation. Here we not only list the fit uncertainties caused by shape noise, but also the additional minor noise contributions from projections of uncorrelated large-scale structure and line-of-sight variations in the redshift distribution (see S18a for details).

The D19 concentration–mass relation provides an estimate for the expected average concentration as a function of mass for an average (approximately mass-selected) cluster population, as adequate for the overall sample studied in S18a and S20. Given their relaxed nature, we would however expect that the three clusters studied here should – on average – have higher concentrations. For simulated haloes with masses comparable to the masses of our clusters, Neto et al. (2007) find that the median concentration of relaxed haloes is higher by a factor  $\times 1.14$  compared to the median concentration of the full halo population. Following S18b we therefore refitted the clusters using an increased concentration  $c_{200c} = 1.14c_{200c}^{D19}$ , where  $c_{200c}^{D19}$  is the concentration that corresponds to the best-fit mass from the initial fit when assuming the D19  $c(M)$  relation. The resulting mass constraints, which shift noticeably for  $M_{200c}$  but only little for  $M_{500c}$ , are also listed in Table 7. Our derived constraints on  $M_{500c}$  agree within  $\sim 1\sigma$  with *Chandra* X-ray estimates computed by McDonald et al. (2019) assuming the  $Y_X$ – $M$  scaling relation from Vikhlinin et al. (2009, compare Table 7).

Using simulations, S20 compute a mass modelling correction as a function of mass and redshift for approximately mass-selected cluster populations. Thus, their correction should on average be accurate for the full cluster sample studied by S20. However, since we are studying a subset consisting of particularly relaxed clusters, which should not suffer significantly from miscentring or substructure, it is more adequate to not apply these mass modelling corrections to our results. For this reason, and also given that we employed the increased concentrations applicable for relaxed clusters, our masses have lower absolute values compared to the results from S20. We expect that the mass constraints reported for these particular clusters in S20 are likely

scattered up given that their analysis is blind to cluster morphology. We note that this does not affect the population-averaged results derived by S20, as particularly disturbed clusters will likely scatter down in their analysis. In the future it might be possible to reduce this scatter by deriving morphology-dependent corrections for mass-modelling biases from hydrodynamical simulations.

Systematic mass uncertainties are small compared to the statistical uncertainties reported in Table 7. The 1.5% uncertainty in the shear calibration (see Table 5) translates to a 2.3% mass uncertainty. The mass uncertainty associated with the calibration of the source redshift distribution amounts to 4.7% (S20). While we did not apply a mass modelling correction (as discussed above), we conservatively consider the same residual (post-correction) mass modelling uncertainty as S20 to reflect the lack of a morphology-dependent mass modelling correction (5.3%). Added in quadrature, the total systematic mass uncertainty amounts to 7.4%.

## 8. Conclusions

We created sets of simulated images mimicking real HST/ACS cluster observations in order to test and better calibrate our implementation of the KSB+ algorithm which is used to measure the WL signals of clusters. As a first simplified setup we placed galaxies on a grid, using real catalogues analysed with KSB+ as comparison to determine the input parameters of the mock galaxies. With these grid simulations we studied the influence of choices for the input parameters, focussing on the galaxy properties while excluding the impact of neighbours. We found that the strongest impact comes from the distributions describing the synthetic galaxy light profiles and the details of the PSF shape. We gathered that the method yields results that are robust to  $\lesssim 1\%$  for typical source populations, but note that extreme populations such as pure De Vaucouleurs profiles yield significant bias shifts. Similarly, we find a stability of the results to  $\Delta m \sim 0.01$  when varying the PSF model from fairly realistic Tiny Tim ACS PSF models (including charge diffusion) to a Moffat PSF with similar half-light radius. Here we note that the resulting multiplicative bias is weakly dependent on the PSF ellipticity, likely reflecting limitations of the KSB+ algorithm. We checked that for the stronger shear regime ( $|g| < 0.4$ ) we do not

**Table 6.** Cluster properties and achieved weak lensing mass signal-to-noise ratios.

Cluster	$z$	$S/N_{\text{mass,D19}}^{>500\text{kpc}}$	$S/N_{\text{mass,D19}}^{>200\text{kpc}}$	$(S/N_{\text{mass,D19}}^{>200\text{kpc}})/(S/N_{\text{mass,D19}}^{>500\text{kpc}})$
SPT-CL J0000–5748	0.702	2.63	3.25	1.24
SPT-CL J2043–5035	0.723	2.15	2.61	1.22
SPT-CL J2331–5051	0.576	1.90	3.19	1.67

Note. — *Column 1:* Cluster designation. *Column 2:* Cluster redshift from Bleem et al. (2015). *Column 3:* Mass signal-to-noise ratio  $S/N_{\text{mass,D19}} = M_{200c}/\Delta M_{200c}$  (considering only shape-noise uncertainties) obtained from NFW reduced shear profile fits using scales  $0.5\text{Mpc} < r < 1.5\text{Mpc}$  and assuming the D19  $c(M)$  relation. *Column 4:* Mass signal-to-noise ratio  $S/N_{\text{mass,D19}} = M_{200c}/\Delta M_{200c}$  (considering only shape-noise uncertainties) obtained from NFW reduced shear profile fits using scales  $0.2\text{Mpc} < r < 1.5\text{Mpc}$  and assuming the D19  $c(M)$  relation. *Column 5:* Ratio of the values from columns 4 and 3.

**Table 7.** Weak lensing constraints derived for the fit range  $0.2\text{Mpc} < r < 1.5\text{Mpc}$ . Masses are in  $10^{14}M_{\odot}$  and are reported for two different overdensities  $\Delta \in \{200c, 500c\}$ . For comparison we also list *Chandra* X-ray mass estimates  $M_{500c,Y_X}$  from McDonald et al. (2019).

Cluster	Using $c(M)$ from D19			Using $c_{200c} = 1.14c_{200c}^{D19}$		
	$M_{200c}$	$M_{500c}$	$c_{200c}^{D19}$	$M_{200c}$	$M_{500c}$	$M_{500c,Y_X}$
SPT-CL J 0000–5748	$5.6^{+1.8}_{-1.7} \pm 0.7 \pm 0.3$	$3.8^{+1.3}_{-1.1} \pm 0.5 \pm 0.2$	3.72	$5.2^{+1.7}_{-1.5} \pm 0.7 \pm 0.3$	$3.7^{+1.2}_{-1.1} \pm 0.5 \pm 0.2$	$4.1^{+0.7}_{-0.6}$
SPT-CL J 2043–5035	$4.2^{+1.7}_{-1.5} \pm 0.6 \pm 0.3$	$2.8^{+1.2}_{-1.0} \pm 0.4 \pm 0.2$	3.66	$3.9^{+1.6}_{-1.4} \pm 0.7 \pm 0.3$	$2.8^{+1.1}_{-1.0} \pm 0.5 \pm 0.2$	$4.2^{+0.1}_{-0.2}$
SPT-CL J 2331–5051	$5.7^{+1.8}_{-1.7} \pm 0.8 \pm 0.4$	$3.9^{+1.3}_{-1.2} \pm 0.5 \pm 0.3$	3.69	$5.2^{+1.7}_{-1.6} \pm 0.7 \pm 0.3$	$3.6^{+1.2}_{-1.1} \pm 0.5 \pm 0.2$	$4.3^{+0.3}_{-0.4}$

Note. — *Column 1:* Cluster designation. *Column 2:*  $M_{200c}$  constraints obtained when assuming the  $c(M)$  relation from D19. *Column 3:*  $M_{500c}$  constraints obtained when assuming the  $c(M)$  relation from D19. *Column 4:* Concentration derived from the D19  $c(M)$  relation for the best-fit  $M_{200c}$ . *Column 5:*  $M_{200c}$  constraints obtained for a fixed concentration  $c_{200c} = 1.14c_{200c}^{D19}$ . *Column 6:*  $M_{500c}$  constraints obtained for a fixed concentration  $c_{200c} = 1.14c_{200c}^{D19}$ . The statistical uncertainties listed for the mass constraints correspond to shape noise (asymmetric), uncorrelated large-scale structure projections, and line-of-sight variations in the source redshift distribution. The systematic mass uncertainty amounts to 7.4%. *Column 7:* *Chandra* X-ray mass estimates computed by McDonald et al. (2019) assuming the  $Y_X$ – $M$  scaling relation from Vikhlinin et al. (2009).

find any non-linear deviations of the multiplicative bias, but we note a shear dependence of the additive bias at  $|g| > 0.2$ . We studied the  $S/N_{\text{KSB}}$  and magnitude dependence of this bias, and confirmed the need for a  $S/N_{\text{KSB}}$ -dependent correction (for which we provide updated fit parameters) in order to keep biases small at low signal-to-noise ratios. This correction reduces the bias to the  $\lesssim 0.01$  level, also when pushing our signal-to-noise cuts to  $S/N_{\text{flux}} > 7$ , which opens up the possibility to include a larger number of galaxies in the analysis. Differences in the intrinsic ellipticity distribution were shown to have negligible impact for current cluster applications.

We studied the influence of light contamination from cluster galaxies by creating two sets of simulations, one with background galaxies only and one which also includes cluster members and foreground galaxies. We generally find that the impact of such bright galaxies is small ( $\Delta m \lesssim 0.005$ ), except for clusters at lower redshifts  $z < 0.7$  and cluster centric distances  $70''$ – $100''$ , for which we obtain  $\Delta m = (0.0113 \pm 0.0033)$ . The impact of selection bias with the addition of cluster members is also low for ( $z > 0.7$ ) clusters ( $\Delta m \sim 0.006$ ).

Aiming to create more realistic scenarios which include the impact of neighbouring galaxies we generated mock CANDELS observations, which used actual galaxy positions and magnitudes to provide realistic clustering. This setup also allows us to constrain selection bias, for which we obtained a first estimate by measuring the average intrinsic input ellipticity of the detected and selected galaxies. Confirming results from Kannawadi et al. (2019), we find that the object detection in SExtractor leads to a selection bias  $m_{\text{sel}} = -0.026$ , which is however reduced to a net selection bias of  $m_{\text{sel}} = -0.014$  once additional cuts applied in our pipeline are taken into account. Since both effects do not separate perfectly, we obtained a joint correction for shape mea-

surement and selection biases from the CANDELS-like simulations (not requiring matched galaxy pairs). Here we found that the dependence of the bias on  $S/N_{\text{KSB}}$  is still well described by the calibration that was initially obtained for shape measurement bias using the grid-based simulations, if a constant bias offset is added. After applying this revised calibration the dependence of residual shape measurement bias on the galaxy flux radius is sufficiently small that it can be neglected for background-selected cluster weak lensing analyses. As final ingredient for our bias correction we obtained an estimate for the shape measurement bias caused by very faint ( $V_{606} > 27.5$ ) neighbouring galaxies, following the analysis from Euclid Collaboration: Martinet et al. (2019).

In summary, we derived a conservative updated shear calibration for our HST-like cluster weak lensing analyses, considering many different sources of bias. The small remaining residual biases can be corrected for in the analysis of real data depending on the employed  $S/N_{\text{flux}}$  cut (see Table 5). Our analysis suggests that the corrected shear estimates are accurate to 1.5%, which is fully sufficient for current cluster studies. Our results also give us confidence to extend our current shear estimates to the inner cluster regions, where the lensing effect approaches the non-weak regime. The achieved accuracy does, however, not meet the stringent requirements of next generation surveys such as Euclid (see Köhlinger et al. 2015), which will require more careful matching of the simulation parameters and likely the use of more advanced shape measurement methods. Shear calibration campaigns of next generation WL surveys will need to include the cluster regime probed in this work in their image simulations, in particular accounting for cluster galaxy blending and non-weak shears, in order to achieve the accuracy requirements for their mass estimates.

We have not included magnification or flexion effects in this work, which could further impact the results. However, the impact of flexion should be small in the more relevant  $|g| < 0.2$  regime. Magnification affects the observed distributions of faint background sources, which can impact cluster weak lensing studies by changing the source redshift distribution (S18a). The impact on shear estimation biases is however expected to be negligible at the accuracy level of our study (see also Euclid Collaboration: Martinet et al. 2019).

A set of HAWK-I-like image simulations of high- $z$  sources following the same approach showed consistent results with the use of the same  $S/N_{\text{KSB}}$ -dependent correction and a residual bias below 0.5% (see Appendix A). This indicates that the results presented here can be applied to other high- $z$  cluster weak lensing observations and not only to HST data. In particular they can be implemented in future high- $z$  cluster studies using  $K_s$ -band shape measurements from HAWK-I as pioneered by S18b.

Finally, we used the corrections derived in this paper and the confirmation that systematics are well controlled also in the non-weak regime to obtain updated mass constraints for three relaxed clusters. In particular, we demonstrated that the inclusion of smaller scales ( $r > 200$  kpc, as enabled by our work), increases the mass signal-to-noise ratio  $S/N_{\text{mass,D19}} = M_{200c}/\Delta M_{200c}$  by a factor of 1.38 on average compared to the default scale  $r > 500$  kpc used by S20.

## Acknowledgements

This work is based on observations made with the NASA/ESA *Hubble Space Telescope*, using imaging data from the SPT follow-up GO programmes 12246 (PI: C. Stubbs), 14352 (PI: J. Hlavacek-Larrondo), and 13412 (PI: Schrabback). STScI is operated by the Association of Universities for Research in Astronomy, Inc. under NASA contract NAS 5-26555. This work is also based on observations made with ESO Telescopes at the La Silla Paranal Observatory under programmes 088.A-0889 (PI: J. Mohr) and 0100.A-0217 (PI: B. Hernández-Martín). PISCO observations are supported by NSF AST-1814719. Argonne National Lab, a U.S. Department of Energy Office of Science Laboratory, is operated by UChicago Argonne LLC under contract no. DE-AC02-06CH11357.

The Bonn group acknowledges support from the German Federal Ministry of Economics and Technology (BMWi) provided through DLR under projects 50 OR 1407, 50 OR 1610, 50 OR 1803, and 50 QE 1103, as well as support provided by the Deutsche Forschungsgemeinschaft under project SCHR 1400/3-1. TS acknowledges support through ERC H2020-COMPET-2017 project #776247. HH acknowledges support from the Netherlands Organisation for Scientific Research (NWO) under grant numbers 639.043.512 and NM acknowledges support from a CNES fellowship. We want to thank Fatimah Raihan, Hannah Zohren, Diana Scognamiglio, Martin Sommer and Florian Kleinebreil for valuable discussions of this work. We also thank Bryan Gillis and collaborators for providing updated Tiny Tim parameters prior to the submission of Gillis et al. (2019). We thank Peter Schneider for useful comments on this manuscript.

## References

Bacon, D. J., Goldberg, D. M., Rowe, B. T. P., & Taylor, A. N. 2006, *MNRAS*, 365, 414  
 Bartelmann, M., & Schneider, P. 2001, *Phys. Rep.*, 340, 291  
 Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, *AJ*, 132, 1729  
 Bernstein, G. M., Armstrong, R., Krawiec, C., & March, M. C. 2016, *MNRAS*, 459, 4467

Bertin, E., & Arnouts, S. 1996, *AAP*, 117, 393  
 Bleem, L. E., Stalder, B., de Haan, T., et al. 2015, *The Astrophysical Journal Supplement Series*, 216, 27  
 Bleem, L. E., Bocquet, S., Stalder, B., et al. 2019, *ApJ*, submitted, arXiv:1910.04121  
 Bocquet, S., Dietrich, J. P., Schrabback, T., et al. 2019, *ApJ*, 878, 55  
 Bridle, S., Balan, S. T., Bethge, M., et al. 2010, *MNRAS*, 405, 2044  
 Bridle, S. L., Kneib, J. P., Bardeau, S., Gull, S. F., & Natarajan, P. 2002, in *The Shapes of Galaxies and their Dark Halos*, 38–46  
 Diemer, B., & Joyce, M. 2019, *ApJ*, 871, 168  
 Dietrich, J. P., Bocquet, S., Schrabback, T., et al. 2019, *MNRAS*, 483, 2871  
 Erben, T., Van Waerbeke, L., Bertin, E., Mellier, Y., & Schneider, P. 2001, *A&A*, 366, 717  
 Euclid Collaboration: Martinet, N., Schrabback, T., Hoekstra, H., et al. 2019, *A&A*, 627, A59  
 Fenech Conti, I., Herbonnet, R., Hoekstra, H., et al. 2017, *MNRAS*, 467, 1627  
 Gillis, B., Schrabback, T., Maggraf, O., et al. 2019, *MNRAS*, submitted  
 Goldberg, D. M., & Bacon, D. J. 2005, *ApJ*, 619, 741  
 Goldberg, D. M., & Natarajan, P. 2002, *ApJ*, 564, 65  
 Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, *ApJS*, 197, 35  
 Herbonnet, R., von der Linden, A., Allen, S. W., et al. 2019, *MNRAS*, 490, 4889  
 Heymans, C., Brown, M. L., Barden, M., et al. 2005, *MNRAS*, 361, 160  
 Heymans, C., Van Waerbeke, L., Bacon, D., et al. 2006, *MNRAS*, 368, 1323  
 Hoekstra, H., Franx, M., Kuijken, K., & Squires, G. 1998, *ApJ*, 504, 636  
 Hoekstra, H., Herbonnet, R., Muzzin, A., et al. 2015, *MNRAS*, 449, 685  
 Hoekstra, H., Viola, M., & Herbonnet, R. 2017, *MNRAS*, 468, 3295  
 Kaiser, N., Squires, G., & Broadhurst, T. 1995, *ApJ*, 449, 460  
 Kannawadi, A., Hoekstra, H., Miller, L., et al. 2019, *A&A*, 624, A92  
 Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, *ApJS*, 197, 36  
 Köhlinger, F., Hoekstra, H., & Eriksen, M. 2015, *MNRAS*, 453, 3107  
 Krist, J. 2003, *ACS WFC: HRC field dependent PSF variations due to optical and charge diffusion effects*, Tech. rep.  
 Krist, J. E., Hook, R. N., & Stoehr, F. 2011, in *Optical Modeling and Performance Predictions V*, Vol. 8127, 81270J  
 Luppino, G. A., & Kaiser, N. 1997, *ApJ*, 475, 20  
 Mandelbaum, R., Hirata, C. M., Leauthaud, A., Massey, R. J., & Rhodes, J. 2012, *MNRAS*, 420, 1518  
 Mandelbaum, R., Rowe, B., Armstrong, R., et al. 2015, *MNRAS*, 450, 2963  
 Mandelbaum, R., Lanusse, F., Leauthaud, A., et al. 2018, *MNRAS*, 481, 3170  
 Massey, R., Heymans, C., Bergé, J., et al. 2007, *MNRAS*, 376, 13  
 McDonald, M., Allen, S. W., Hlavacek-Larrondo, J., et al. 2019, *ApJ*, 870, 85  
 Miller, L., Kitching, T. D., Heymans, C., Heavens, A. F., & van Waerbeke, L. 2007, *MNRAS*, 382, 315  
 Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, *ApJ*, 490, 493  
 Neto, A. F., Gao, L., Bett, P., et al. 2007, *MNRAS*, 381, 1450  
 Pujol, A., Kilbinger, M., Sureau, F., & Bobin, J. 2019, *A&A*, 621, A2  
 Raihan, S. F., Schrabback, T., Hildebrandt, H., Applegate, D., & Mahler, G. 2020, *MNRAS*, submitted  
 Refregier, A., Kacprzak, T., Amara, A., Bridle, S., & Rowe, B. 2012, *MNRAS*, 425, 1951  
 Rhodes, J. D., Massey, R., Albert, J., et al. 2006, in *The 2005 HST Calibration Workshop: Hubble After the Transition to Two-Gyro Mode*, ed. A. M. Koekemoer, P. Goudfrooij, & L. L. Dressel, 21  
 Rowe, B. T. P., Jarvis, M., Mandelbaum, R., et al. 2015, *Astronomy and Computing*, 10, 121  
 Samuroff, S., Bridle, S. L., Zuntz, J., et al. 2018, *MNRAS*, 475, 4524  
 Schrabback, T. et al. in prep.  
 Schrabback, T., Erben, T., Simon, P., et al. 2007, *A&A*, 468, 823  
 Schrabback, T., Hartlap, J., Joachimi, B., et al. 2010, *AAP*, 516, A63  
 Schrabback, T., Applegate, D., Dietrich, J. P., et al. 2018a, *MNRAS*, 474, 2635  
 Schrabback, T., Schirmer, M., van der Burg, R. F. J., et al. 2018b, *A&A*, 610, A85  
 Sheldon, E. S., & Huff, E. M. 2017, *ApJ*, 841, 24  
 Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, *ApJS*, 214, 24  
 Stalder, B., Stark, A. A., Amato, S. M., et al. 2014, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 9147, Ground-based and Airborne Instrumentation for Astronomy V, 91473Y  
 Sunyaev, R. A., & Zeldovich, Y. B. 1969, *Nature*, 223, 721  
 Tewes, M., Kuntzer, T., Nakajima, R., et al. 2019, *A&A*, 621, A36  
 van der Wel, A., Bell, E. F., Häussler, B., et al. 2012, *ApJS*, 203, 24  
 van der Wel, A., Franx, M., van Dokkum, P. G., et al. 2014, *ApJ*, 788, 28  
 Vikhlinin, A., Burenin, R. A., Ebeling, H., et al. 2009, *ApJ*, 692, 1033  
 Viola, M., Kitching, T. D., & Joachimi, B. 2014, *MNRAS*, 439, 1909  
 von der Linden, A., Allen, M. T., Applegate, D. E., et al. 2014, *MNRAS*, 439, 2  
 Wright, C. O., & Brainerd, T. G. 2000, *ApJ*, 534, 34  
 Zhang, Y. Y., Finoguenov, A., Böhringer, H., et al. 2007, *A&A*, 467, 437

## Appendix A: VLT/HAWK-I setup

S18b demonstrate that good-seeing VLT/HAWK-I  $K_s$  images, which benefit from a sharp PSF ( $\text{FWHM} \leq 0''.4$ ), yield a similar sensitivity and efficiency for weak lensing studies of massive clusters at redshift  $0.7 \lesssim z \lesssim 1$  as HST/ACS observations when combined with a background selection in  $g_z K_s$  colour-colour space. For their HAWK-I weak lensing analysis of the cluster RCS2 J232727.7–020437 they employed the same KSB+ pipeline as used in our study. In order to provide a robust KSB+ calibration also for future HAWK-I high- $z$  cluster weak lensing studies of expanded samples, we create an additional set of simulations. Here we mimic the HAWK-I setup and employ near-infrared-measured galaxy properties, but otherwise largely follow the procedures described in this paper for the generation and analysis of HST-like image simulations.

To set up HAWK-I-like mock galaxies, we used the 3D-HST CANDELS light profile fits (van der Wel et al. 2012; van der Wel et al. 2014) as a reference, employing observations obtained in the F160W band, which is the closest available HST filter to the HAWK-I  $K_s$ -band. The Sérsic index distribution is shown in Fig. A.2. Small differences in this distribution, which for example occur if only colour-selected galaxies are used rather than the full sample change the measured bias at the  $\sim 0.5\%$  level only. This also indicates that the differences in the bias caused by the use of a slightly different band should be negligible. We use a similar setup as in Sect. 3, where the galaxies were placed on a grid in order to test the behaviour of the  $S/N_{\text{KSB}}$ -dependent correction, and a more realistic scenario, using positions and magnitudes from the 3D-HST CANDELS catalogues to estimate the influence neighbours have on the bias. Following S18b we mimic colour-selected ( $(z - K_s) > \text{MIN}[g - z, 2.5]$ ) galaxies up to  $K_s = 24.2$ . The main changes with respect to the HST-like simulations are those which are intrinsically dependent of the instrument ( $0''.106$  pixel scale,  $0''.4$  FWHM Moffat PSF).

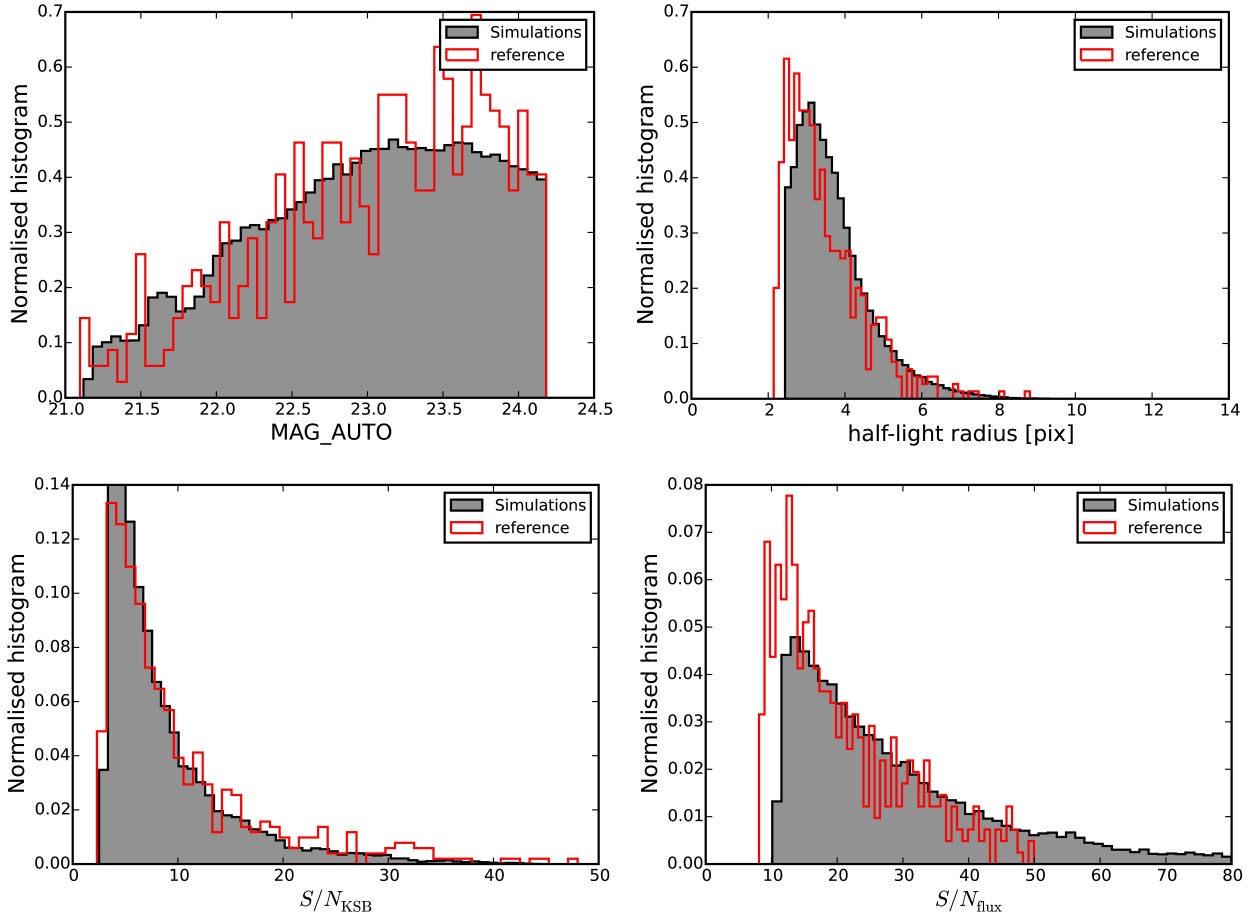
In order to understand how realistic the signal-to-noise ratios (both  $S/N_{\text{KSB}}$  and  $S/N_{\text{flux}}$ ), as well as the galaxy size and  $K_s$  magnitude distributions are, we compared the output distributions to existing HAWK-I cluster observations analysed in S18b with the same KSB+ implementation, as shown in Fig. A.1, finding good agreement.

After applying the  $S/N_{\text{KSB}}$ -dependent correction obtained in Eq. (13) and the appropriate cuts, we obtained consistent results of  $m_1 = -0.0020 \pm 0.0039$  and  $m_2 = -0.0019 \pm 0.0033$  for  $S/N_{\text{flux}} > 10$  galaxies placed on a grid. This indicates that the  $S/N_{\text{KSB}}$ -dependent correction obtained for HST/ACS images is also valid for HAWK-I cluster analyses.

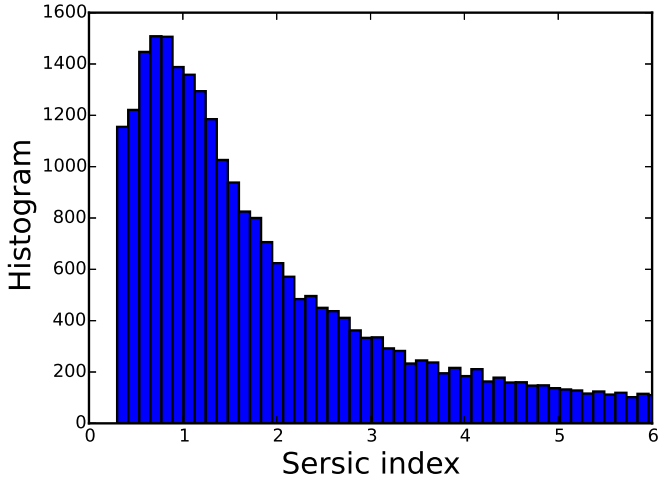
Likewise, very similar estimates are obtained using the CANDELS-like setup with realistic clustering: an estimation of the selection bias similar to Sect. 5.1 yields a value of  $m_1 = -0.0135 \pm 0.0032$  and  $m_2 = -0.0141 \pm 0.0027$ , in excellent agreement with the ACS-like analysis (compare Sect. 5.1.4). Applying the calibration for both shape measurement and selection bias from Eq. (14) to a HAWK-I-like version of the simulations described in Sect. 5.2, we find residual biases of  $m_1 = 0.0082 \pm 0.0025$  ( $0.0076 \pm 0.0030$ ) and  $m_2 = 0.0088 \pm 0.0028$  ( $0.0069 \pm 0.0032$ ) for galaxy samples with  $S/N_{\text{flux}} > 10$  ( $S/N_{\text{flux}} > 7$ ). The 3D-HST CANDELS input catalogues are only complete to  $H_{160} \approx 25$ . This is approximately 1 magnitude deeper than the main weak lensing source sample, similar to the setup from Sect. 5.2. Based on the results from Sect. 5.3 we therefore expect that even fainter galaxies lead to a further residual bias shift that is consistent with the entries in the second row of Table 5.

## Appendix B: Simulated cluster images

As a comparison, we show the same image as for Fig. 8, but for a simulated cluster at  $z = 0.28$  in Fig. B.1, featuring much brighter cluster galaxies.

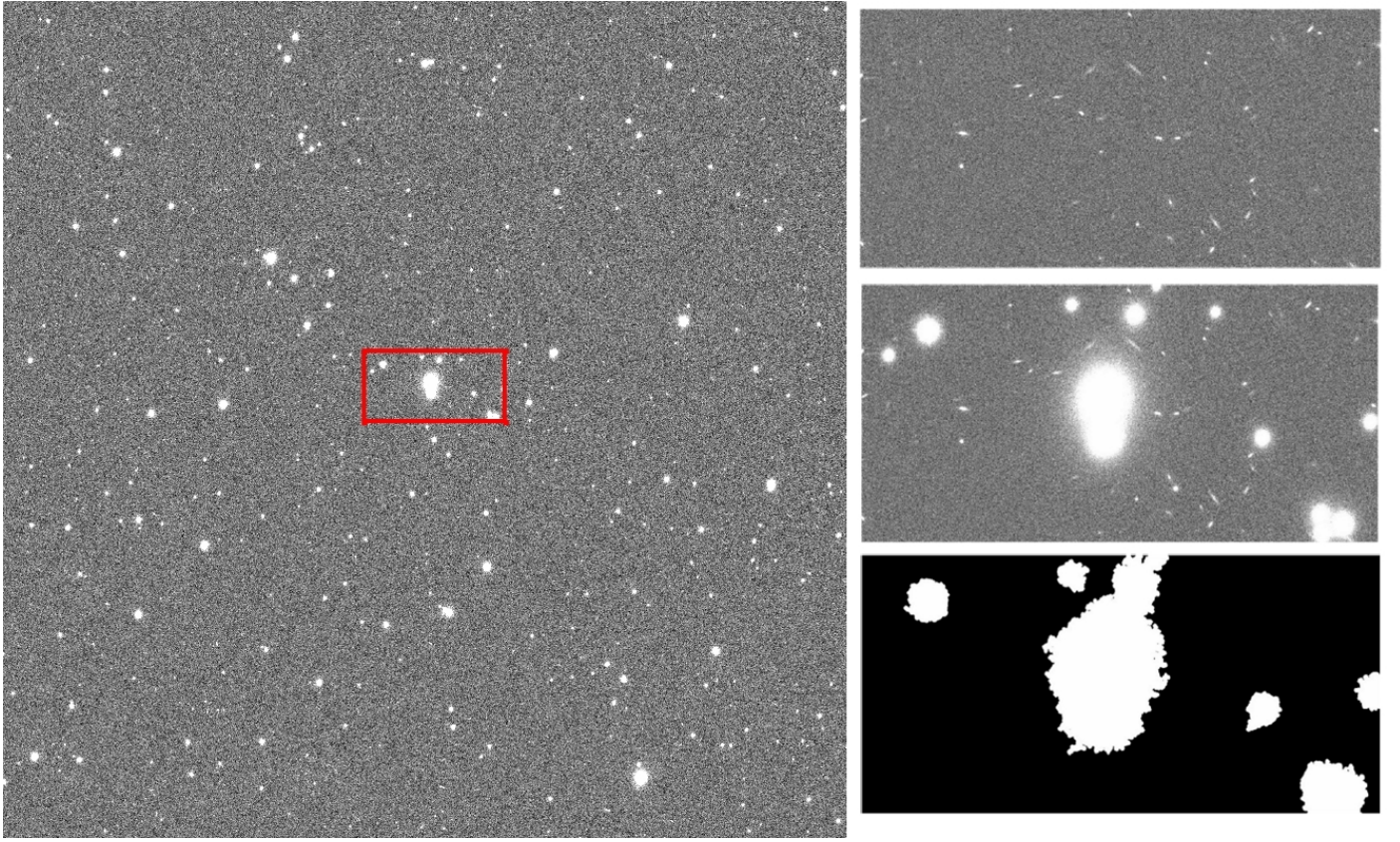


**Fig. A.1.** Comparisons between the measured distributions in the HAWK-I-like simulations and the real image distribution for the  $K_s$  magnitudes (MAG\_AUTO), the half-light radius measured by SExtractor, the KSB signal-to-noise ratio  $S/N_{\text{KSB}}$  and the SExtractor  $S/N_{\text{flux}}$ .



**Fig. A.2.** Distribution of Sérsic indices in the parametric fit to real CANDELS galaxies in the F160W filter (van der Wel et al. 2014) used as input for the HAWK-I simulations.





**Fig. B.1.** Example image of a simulated cluster at  $z = 0.28$ . A cut-out of the full image, shown in red, can be seen in the right for the simulations with background galaxies only (top), with added cluster members (middle) and showing the mask used to remove bright objects (bottom). The full image and cut outs spans  $300'' \times 300''$  and  $50'' \times 25''$ , respectively.