

Accounting for object detection bias in weak gravitational lensing studies

Henk Hoekstra¹, Arun Kannawadi^{2,1}, and Thomas D. Kitching³

¹ Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands
e-mail: hoekstra@strw.leidenuniv.nl

² Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08544, USA

³ Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK

Received 22 July 2020 / Accepted 5 November 2020

ABSTRACT

Weak lensing by large-scale structure is a powerful probe of cosmology if the apparent alignments in the shapes of distant galaxies can be accurately measured. Most studies have therefore focused on improving the fidelity of the shape measurements themselves, but the preceding step of object detection has been largely ignored. In this paper, we study the impact of object detection for a *Euclid*-like survey and show that it leads to biases that exceed requirements for the next generation of cosmic shear surveys. In realistic scenarios, the blending of galaxies is an important source of detection bias. We find that METADETECTION is able to account for blending, leading to average multiplicative biases that meet requirements for Stage IV surveys, provided a sufficiently accurate model for the point spread function is available. Further work is needed to estimate the performance for actual surveys. Combined with sufficiently realistic image simulations, this provides a viable way forward towards accurate shear estimates for Stage IV surveys.

Key words. gravitational lensing: weak – large-scale structure of Universe

1. Introduction

Over the past decades, the theoretical framework that describes the formation of cosmic structures has been tested by ever more precise observations (see e.g., [Planck Collaboration XIII 2016](#), for a comprehensive comparison of results). Although there is discussion about small differences between cosmological parameter estimates (e.g., [Riess et al. 2019](#); [Joudaki et al. 2020](#)), the general agreement is remarkable given the difficulties in obtaining these results. Importantly, the main ingredients of this ‘concordance model’ are not understood at all: dark matter and dark energy make up the bulk of the mass-energy content of the Universe, with a ‘mere frosting’ of baryonic matter. Although a cosmological constant is an excellent fit to the current data, its unnaturally small value is by no means satisfactory. Consequently, many alternative explanations have been suggested, including modifications of the theory of general relativity (see e.g., [Amendola et al. 2018](#), for an overview). In order to distinguish between such a multitude of ideas, dramatically better observational constraints are needed.

The study of the distribution of matter as a function of redshift is of particular interest, because it is sensitive to the growth of structure, modified gravity, and the expansion history. The practical complication that most of the matter is made up of dark matter can be overcome by measuring the correlations in the ellipticities of distant galaxies that are the result of the differential deflection of light rays by intervening structures, a phenomenon called gravitational lensing. In the case that only single images of distant galaxies are distorted by the gravitational lensing effect, this is known as weak lensing. The amplitude of the distortion provides us with a direct measurement of the gravitational tidal field, which in turn can be used to ‘map’ the distribution of matter directly. This makes weak lensing by large-scale

structure, or cosmic shear, one of the most powerful probes to study dark energy and the growth of structure: the statistical properties of the matter distribution can be determined as a function of cosmic time. These measurements can be compared to models of structure formation, which depend on the cosmological parameters (see e.g., [Kilbinger 2015](#), for a recent review).

The typical change in the observed ellipticity of a distant galaxy caused by gravitational lensing (known as shear) is about a percent, much smaller than the intrinsic ellipticities of galaxies. This source of statistical uncertainty can be overcome by averaging over large numbers of galaxies, although intrinsic alignments complicate this simple picture (see e.g., [Joachimi et al. 2015](#); [Troxel & Ishak 2015](#), for reviews). The cosmological lensing signal has now been measured using ground-based observations of relatively modest areas of the sky (see e.g., [Troxel et al. 2018](#); [Hildebrandt et al. 2020](#); [Hamana et al. 2020](#), for some recent results from Stage III surveys) but future surveys will cover much larger fractions of the extragalactic sky, increasing the source samples accordingly.

The change in ellipticity is also smaller than the typical biases introduced by instrumental effects. Consequently, averaging the shape measurements of large ensembles of galaxies is only meaningful if these sources of bias can be corrected for to a level that renders them sub-dominant to the statistical uncertainties afforded by the survey (see [Mandelbaum 2018](#), for a detailed review on weak lensing systematics). This will be particularly challenging for the next generation of surveys (Stage IV), such as the ones carried out by *Euclid*¹ ([Laureijs et al. 2011](#)) and the *Nancy Grace Roman* Space Telescope² ([Spergel et al. 2015](#)) from space, and the Legacy Survey of Space and Time

¹ <https://www.euclid-ec.org/>

² <https://www.stsci.edu/roman>

by the Rubin Observatory³ (LSST Science Collaboration 2009) from the ground.

The point spread function (PSF) is the dominant source of bias in the measurements of galaxy shapes, driving the desire for space-based observations (Paulin-Henriksson et al. 2008; Massey et al. 2013). Another complication is the fact the shapes are measured from noisy images, which can lead to biases in the ellipticity (e.g., Melchior & Viola 2012; Refregier et al. 2012; Miller et al. 2013; Viola et al. 2014). Given a survey design, our current understanding of these biases, and our ability to correct for them, requirements can be placed on the instrument performance, but also on the accuracy of the shape measurement algorithm. For instance, Cropper et al. (2013) present a detailed breakdown for *Euclid*, which forms the basis for some of the numbers used in this paper.

Fortunately, the impact of the various sources of bias can be studied by applying the shape measurement algorithm to simulated data, where the galaxy images are sheared by a known amount. Comparison with the recovered values provides an estimate of the biases. For example, Erben et al. (2001) and Hoekstra et al. (2002) used simulated images to examine the performance of the KSB algorithm developed by Kaiser et al. (1995). Comparing a range of methods, the Shear TEsting Programme (STEP; Heymans et al. 2006; Massey et al. 2007) demonstrated the importance of how a method is actually implemented. To examine the origin of the variation in performance further, the GRavitational lEnsing Accuracy Testing (GREAT) challenges (Bridle et al. 2010; Kitching et al. 2012; Mandelbaum et al. 2015) used idealised simulations to demonstrate the importance of noise on the performance.

However, as recently shown by Hoekstra et al. (2015), the actual performance of the algorithms depends crucially on the input of the simulations, such as the distribution of galaxy ellipticities and the inclusion of faint galaxies. This was studied in more detail in Hoekstra et al. (2017, H17 hereafter) for a *Euclid*-like survey. These studies showed that the fidelity of the image simulations is crucial for an accurate estimate of the overall shear bias, which depends on the bias in the shape measurements and the selection of galaxies. H17 did not consider both contributions separately, but recent studies (e.g., Fenech Conti et al. 2017; Kannawadi et al. 2019) have shown that biases are already introduced in the first step of the analysis: the detection of objects. This source of bias has been largely ignored until Fenech Conti et al. (2017) showed that it can be as important as the shape measurement bias in ground-based surveys. More recently, Hernández-Martín et al. (2020) showed that detection bias is also relevant for lensing studies using *Hubble* Space Telescope data.

Consequently, even if the shapes of the detected galaxies are somehow measured perfectly, the shear will be biased. Such a detection bias is expected because the significance with a galaxy is detected typically depends on its orientation with respect to the shear (Hirata & Seljak 2003) or the PSF (Kaiser 2000; Bernstein & Jarvis 2002). In this paper we study detection bias using image simulations, similar to those used in H17. We explore how well the bias can be quantified and which parameters are most relevant. We find that the blending of galaxies is the dominant source of detection bias. Such blends are absent from studies that measure shear biases using isolated galaxies (or when placed on a grid). To reduce shape noise, studies typically use pairs of simulated galaxies where a second galaxy is rotated by 90° (or quartets, rotated by 45°). However, if one then

requires that both galaxies are detected, as in Pujol et al. (2019), the detection bias is also removed. Although this is a viable approach to reduce the number of simulated images to quantify the bias introduced by the shape measurement algorithm, it is important to realise that the resulting bias cannot be applied to the actual data, but needs to be adjusted to account for detection bias.

A further complication arises from the fact that it may not be possible to determine the shape for every detected galaxy. Hence the shape measurement step introduces additional selections, as does assigning weights to capture the fidelity of the shape measurement. Finally, to improve constraints on cosmological parameters, the source samples are split into multiple tomographic bins, using photometric redshifts. The reliance on reliable multi-band photometry introduces further selections. Those selection biases will depend on both the shape measurement algorithm and the way samples are selected.

The setup we use in this paper is very similar to the one used in H17, and in Sect. 2 we briefly describe the simulation setup, highlighting some of the changes we implemented. We study detection bias and its dependence of the SEXTRACTOR setup and the PSF in Sect. 3. Similar to H17, we explore the sensitivity to changes in the simulation input in Sect. 4. In Sect. 5 we quantify the performance of METACALIBRATION (Huff & Mandelbaum 2017; Sheldon & Huff 2017) as a way to avoid image simulations for the calibration of the shape measurement step. We also examine the usefulness of its extension, the so-called METADETECTION approach (Sheldon et al. 2020), which aims to avoid selection biases altogether in Sect. 6. We discuss the implications of our results for future surveys in Sect. 7.

2. Simulation setup

The simulated images were created using the publicly available software package GALSIM⁴ (Rowe et al. 2015). This suite of routines was originally developed for GREAT3 (Mandelbaum et al. 2014, 2015), but it has become the de facto standard for image simulations in the weak lensing community. As was done in H17, the galaxies are described by Sérsic profiles, with half-light radii, apparent magnitudes and Sérsic indices n drawn from a catalogue of morphological parameters measured from resolved *F606W* images from the GEMS survey (Rix et al. 2004). We only considered galaxies fainter than magnitude $m = 20$ and used the morphological parameters from the GEMS catalogue for galaxies down to $m = 25.4$, and normalised the counts to 36 galaxies arcmin⁻² with $20 < m < 24.5$.

As shown in H17, it is important to include galaxies down to $m_{\text{lim}} \approx 29$, and we followed the same procedure, except that we used a flatter count slope at fainter magnitudes: we adopted a power law slope of $\alpha_{\text{faint}} = 0.24$ (instead of $\alpha_{\text{faint}} = 0.36$ using by H17), which matches the observed counts better. The intrinsic ellipticities were drawn from a Rayleigh distribution with scale parameter $\epsilon_0 = 0.25$, so that the mean source ellipticity is $\langle |\epsilon^s| \rangle \approx 0.31$. We assumed that the intrinsic ellipticities ϵ^s do not correlate with the morphological parameters, but note that Kannawadi et al. (2019) have shown that this is not the case in reality. We refer the interested reader to H17 for more details on the input catalogue.

In our baseline simulations we placed galaxies randomly, but with random sub-pixel offsets. We created pairs of images, where the galaxies were placed at the same location, but rotated by 90°

³ <https://www.lsst.org>

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

Table 1. Relevant SEXTRACTOR setup parameters.

DETECT_MINAREA	6
DETECT_THRESH	1.5
FILTER_NAME	gauss_3.0_7x7.conv
DEBLEND_NTHRESH	32
DEBLEND_MINCONT	0.005
CLEAN_PARAM	1.0
BACK_SIZE	600
BACK_FILTERSIZE	3

Notes. Column 1: keyword in SEXTRACTOR configuration file; Col. 2: value of the parameter.

in the rotated case. We applied the same shear to all the galaxies in such a pair by changing the true (simulated) ellipticity using (Seitz & Schneider 1997):

$$\epsilon^{\text{obs}} = \frac{\epsilon^{\text{s}} + \gamma}{1 + \gamma^* \epsilon^{\text{s}}}, \quad (1)$$

where ϵ^{s} is the intrinsic complex ellipticity, γ is the complex shear that is applied⁵, and the asterisk indicates the complex conjugate. If both galaxies of a pair are averaged, $\langle \epsilon^{\text{s}} \rangle = 0$ and the observed ellipticity is an unbiased estimate of the shear. Hence, a non-zero detection bias implies that one of the two galaxies in a pair is not detected in a shear-dependent fashion.

In our baseline setup, the galaxies were placed at random positions, thus ignoring the impact of clustering. This was studied in more detail in Euclid Collaboration (2019) who found that faint satellite galaxies that cluster around their host galaxy do affect the bias estimates. Moreover applying a shear to a particular configuration of galaxies also changes their positions. We ignored this in our baseline simulations, but we found that shearing the positions as well as the galaxy images barely changed the results (see Sect. 3.1 and Table 2 for more details). Finally, we also created images where the galaxies were placed on a grid, so that they are about 9'' apart, thus eliminating any blending. This provided a useful reference to compare our baseline results against.

To allow for a more direct comparison to the results presented in H17, unless specified otherwise, we used the same setup for the telescope parameters, and used a circular Airy PSF for a telescope with a diameter of 1.2 m and an obscuration of 0.3 at a reference wavelength of 800 nm, which is a reasonable approximation to the Euclid PSF in the VIS-band (Cropper et al. 2018). The individual images are 4000 pixels on a side, with a pixel size of 0.1'' per pixel. The noise level is the same as used in H17, corresponding to a surface brightness of 27.7 mag arcsec⁻². This mimics the depth of four coadded exposures, and yields a typical number density of 47 galaxies arcmin⁻² with a signal-to-noise ratio larger than 10, as measured by SEXTRACTOR, and a number density of 33 galaxies arcmin⁻² if we restrict the magnitude range to 20 < m < 24.5.

2.1. Analysis setup

We used SEXTRACTOR (Bertin & Arnouts 1996) to detect objects in the simulated images. Our baseline setup uses the (relevant) parameter values listed in Table 1, which are fairly

⁵ The actual observable is the reduced shear $g \equiv \gamma/(1 - \kappa)$, where κ is the convergence, and g should be used in Eq. (1). However, we only consider the shear in this paper, so that $g = \gamma$ throughout.

standard. To detect an object, at least DETECT_MINAREA adjacent pixels need to be above the threshold, which is specified by DETECT_THRESH times the noise level. We let SEXTRACTOR determine the background level, although we could have specified a global value of zero. We explored various background determination settings, and found that they did not change our results. We discuss the purpose of some of these parameters and their impact on detection bias in more detail in Sect. 3.4 and Appendix A.

For reference, we also repeated the shape measurements using the KSB algorithm employed in H17, where we note that the results differ because of a number of changes in the pipeline that were implemented. As already discussed in Sect. 2 we changed the power law slope of the counts of faint galaxies, which shifts the bias as indicated by Fig. 9 in H17. We also improved the modelling of the PSF parameters: the pixel size of 0.1'' is relatively large compared to the FWHM of the PSF of a 1.2m diffraction limited telescope. In H17 the correction for the PSF was based on parameters that were estimated directly from the poorly sampled images. Although this does not impact their main conclusions, it does change the actual biases. Here we used measurements of the PSF shear and smear polarisabilities (Kaiser et al. 1995; Hoekstra et al. 1998) that were measured from 4× oversampled images. Moreover, we increased the width of the weight function by a factor $1/\sqrt{\ln(2)} \approx 1.2$, which also changes the shear bias⁶.

2.2. Detection and photometry performance

Figure 1 shows the fraction of simulated galaxies that were detected by SEXTRACTOR as a function of the input magnitude, m_{input} . To obtain this result we matched the input catalogue to the SEXTRACTOR output and selected those objects that were detected within a radius of 3 pixels from the input coordinate. The black line shows the results for our baseline simulation, whereas the red line shows the fraction of detected objects if the galaxies are placed on a grid about 9'' apart. In the latter case the sample of detected galaxies is complete down to $m_{\text{input}} = 23.5$, after which the completeness starts decreasing. The sample of galaxies detected in the baseline simulation is incomplete at all magnitudes, although 98% of the galaxies are detected down to $m_{\text{input}} = 23.5$. The increased incompleteness is caused by blending, because the results for galaxies that have a nearest neighbour with $m_{\text{input}} < 26$ that is at least 5'' away (blue line) resemble that of the grid-based images. If we instead select galaxies with a nearest neighbour with $m_{\text{input}} < 26$ within 2'', the incompleteness increases (light blue line).

This basic result shows that the detection of galaxies is affected by the presence of neighbouring galaxies. Before we proceed to explore the impact on shape measurements, we briefly examine the impact on the recovered magnitudes. The black line in Fig. 2 shows the distribution of Δm , the difference between m_{AUTO} , the magnitude reported by SEXTRACTOR as MAG_AUTO, and the input magnitude m_{input} , for galaxies with 20 < m_{AUTO} < 24.5 in the baseline simulations. The results show a clear tail towards negative Δm , which is what we expect for blended objects. This is confirmed if we consider the distributions for ‘isolated’ galaxies (blue; nearest neighbour >5'' away) and ‘blended’ galaxies (light blue; nearest neighbour <2''

⁶ We use the observed value of the half-light radius FLUX_RADIUS as measured by SEXTRACTOR to define the width of the weight function. For a Gaussian profile the corresponding dispersion $\sigma = \text{FLUX_RADIUS}/\sqrt{2 \ln 2}$.

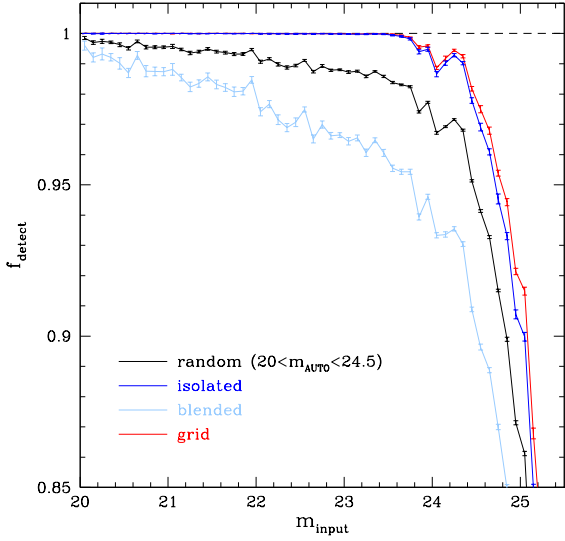


Fig. 1. Fraction of the simulated galaxies that are detected by SEXTRACTOR as a function of the input magnitude, m_{input} . The black line corresponds to the reference case where galaxies are placed randomly in the images. The blue line shows results for ‘isolated’ galaxies, with a nearest neighbour more than $5''$ away, whereas the light blue line is for galaxies with a nearest neighbour within $2''$. In the latter case the fraction of detected galaxies is considerably lower, whereas the results for the ‘isolated’ galaxies approaches that of the simulations where galaxies are placed on a grid about $9''$ apart (red lines). The error bars indicate the scatter in the results, and the lines connect the points.

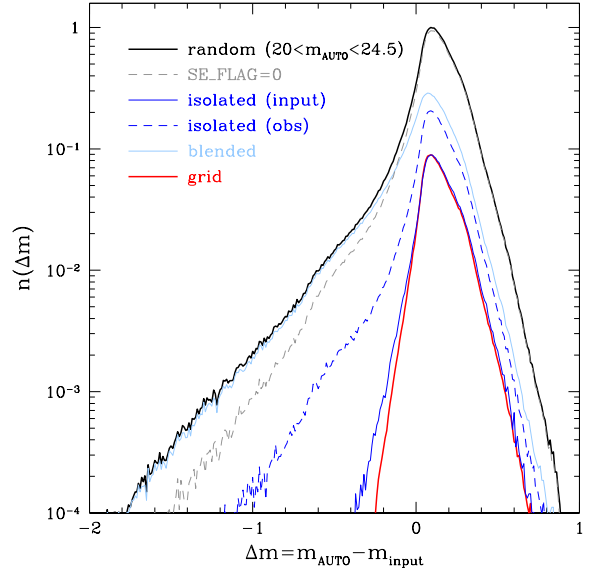


Fig. 2. Distribution of Δm , the difference between m_{AUTO} , the magnitude reported by SEXTRACTOR, and the input magnitude m_{input} for detected galaxies with $20 < m_{\text{AUTO}} < 24.5$ for our baseline setup (solid black line; galaxies placed randomly). The distribution of ‘isolated’ galaxies (solid blue line) matches that of the grid-based results (red line), whereas the tail towards negative Δm matches that of ‘blended’ galaxies. The light grey dashed line shows that many of the objects flagged by SEXTRACTOR are indeed blends, but that many remain undetected. Blends even occur for objects that have no detected neighbour within $5''$ (dashed blue line).

away): the distribution of isolated galaxies roughly matches that of the grid-based simulation (red line; normalisation matched to the blue curve), whereas the distribution of the blended galaxies, comprising 36% of the galaxies, matches the tail for $\Delta m < -0.5$.

The fraction of isolated galaxies is small, only 7.5% of the galaxies match the criterion. In practice, however, SEXTRACTOR will miss nearby neighbours if they are too close. If we use the distance to the nearest *detected* galaxy for the isolation criterion instead, we find that the fraction of apparently isolated galaxies is almost 19%; the dashed blue line in Fig. 2 shows the corresponding distribution, indicating the increased fraction of blends. Finally, SEXTRACTOR raises a flag for objects that it finds to be blended. The light grey dashed line in Fig. 2 shows that it can indeed eliminate some of the blended objects, but many remain. Undetected blends are likely to bias the photometric redshifts, coupling these to biases in the shape measurements, but exploring this further is beyond the scope of this paper.

Finally we note that the distributions do not peak around $\Delta m = 0$, but that $\langle \Delta m \rangle = 0.14$ for $20 < m_{\text{AUTO}} < 24.5$ in the grid-based simulations. The amount of missing flux does depend somewhat on the brightness, increasing from $\langle \Delta m \rangle = 0.056$ for the brightest galaxies ($m_{\text{AUTO}} = 20$) to $\langle \Delta m \rangle = 0.166$ for the faintest ones ($m_{\text{AUTO}} = 24.5$). It also depends somewhat on the source ellipticity, which partly explains the asymmetry towards positive values of Δm . Although the dependence of Δm on ellipticity is modest, it implies that a simple magnitude cut may lead to changes in the ellipticity distributions of the detected galaxies, potentially complicating the link between shape measurements and photometric redshift determinations further.

3. Detection bias

The measurement of the weak gravitational lensing signal relies on accurate estimates of the shapes of distant galaxies, which

are both faint and small. The images are corrupted by noise and instrumental effects. It is essential to remove, or at least account for, these sources of bias. For this reason most effort has focused on undoing the biases in the shape measurement step itself, but the preceding step, the detection (and selection) of galaxies that are used in the analysis, has received much less attention.

As shown already in Hirata & Seljak (2003), we do expect the detection of objects to introduce a bias. Gravitational lensing conserves the surface brightness, and as a result a galaxy with an intrinsic orientation perpendicular to the shear will appear rounder at the same surface brightness level. Since SEXTRACTOR uses a surface brightness threshold and a circular kernel for the detection, such a galaxy is more likely to be detected, resulting in the average shear to be biased low. The detection and selection biases are typically much smaller than the shape measurement biases, but they can no longer be ignored for Stage IV surveys (Albrecht et al. 2006), and require more detailed study (as shown by Fenech Conti et al. 2017; Kannawadi et al. 2019, they are already relevant for Stage III surveys).

We discuss both detection and selection biases. The former refers to the very first step in the analysis, resulting in a sample of objects for which a shape measurement can be attempted. The subsequent shape measurement may not always be successful, or different weights may be assigned to the measurement, which leads to selection biases. Similarly the desire to divide the galaxies into tomographic bins introduces selection biases that need to be accounted for. We emphasise that these biases occur even if the shape measurement itself is unbiased.

To mimic a perfect shape measurement, we follow Fenech Conti et al. (2017) and compute the true measured ellipticity based on the input complex ellipticity ϵ^s and applied complex shear γ as given by Eq. (1). For each galaxy detected by

Table 2. Average multiplicative and additive biases for galaxies with $20 < m_{\text{AUTO}} < 24.5$.

Setup	μ_1	μ_2	$c_1 [\times 10^5]$	$c_2 [\times 10^5]$
Baseline				
SEXTRACTOR	-0.01061 ± 0.00013	-0.01053 ± 0.00013	-0.73 ± 0.51	-0.73 ± 0.49
SEXTRACTOR – no background	-0.01043 ± 0.00018	-0.00974 ± 0.00018	-0.68 ± 0.73	0.36 ± 0.71
SEXTRACTOR – FLAG = 0	-0.01214 ± 0.00023	-0.01178 ± 0.00022	-0.71 ± 0.93	0.70 ± 0.92
KSB detection	-0.01917 ± 0.00023	-0.01872 ± 0.00023	-0.09 ± 0.88	1.17 ± 0.88
KSB selection	-0.01866 ± 0.00022	-0.01819 ± 0.00022	-0.08 ± 0.86	1.12 ± 0.86
KSB shapes	-0.08915 ± 0.00031	-0.08757 ± 0.00032	-3.34 ± 1.20	0.74 ± 1.21
Sheared image				
SEXTRACTOR	-0.01062 ± 0.00019	-0.01043 ± 0.00018	-0.84 ± 0.71	-0.18 ± 0.73
KSB detection	-0.01928 ± 0.00022	-0.01901 ± 0.00023	-0.19 ± 0.88	-1.18 ± 0.87
KSB selection	-0.01877 ± 0.00022	-0.01847 ± 0.00022	-0.16 ± 0.86	-1.18 ± 0.85
KSB shapes	-0.08879 ± 0.00031	-0.08766 ± 0.00033	-2.14 ± 1.21	-1.34 ± 1.22
Grid				
SEXTRACTOR	-0.00694 ± 0.00024	-0.00697 ± 0.00024	-1.53 ± 0.9	0.85 ± 0.91
KSB detection	-0.01325 ± 0.00028	-0.01257 ± 0.00029	-0.81 ± 1.15	-2.83 ± 1.12
KSB selection	-0.01281 ± 0.00027	-0.01209 ± 0.00029	-0.79 ± 1.11	-2.77 ± 1.08
KSB shapes	-0.05357 ± 0.00030	-0.05251 ± 0.00032	-1.14 ± 1.25	-1.80 ± 1.25

Notes. In the ‘baseline’ case the galaxies are placed randomly and their images are sheared. For the ‘sheared image’ results the full scene is sheared instead, thus altering the positions. Galaxies are placed on a regular grid, about $9''$ apart for the ‘grid’ results. The rows labelled ‘SEXTRACTOR’ report the detection bias. The biases for objects with a KSB shape measurement are labelled ‘KSB detection’, and as ‘KSB selection’ when the weighting scheme is included. The results using the actual KSB shape measurements are reported as ‘KSB shapes’. All the KSB measurements also include the SEXTRACTOR detection bias. The reported uncertainties may differ for similar setups, because fewer simulations were analysed.

SEXTRACTOR, we find the nearest input galaxy. For the analysis we consider only galaxies with observed magnitudes $m_{\text{AUTO}} < 25$, but the input catalogue includes many more galaxies that are fainter. As most of those are not detectable individually (see e.g., Fig. 1), we only consider the nearest object with $m_{\text{input}} < 26$ from the input catalogue. We define a mismatch if the separation is more than 3 pixels, which is the case for 0.2% of the objects with $m_{\text{AUTO}} < 25$. The fraction is larger for fainter objects (e.g., 1.4% for detections with $25 < m_{\text{AUTO}} < 26$) suggesting that some of these are just noise peaks. However, we note that such misidentifications do not bias our shear estimate, but rather introduce noise in our measurement because the shape noise is not cancelled in this case⁷. Even though the impact of these mismatches on the results is negligible, we omit them from our analysis.

More important are the cases where the object is blended with a neighbouring one, which can also lead to a shift in the location of the detection. In 0.4% of the detections with $m_{\text{AUTO}} < 25$ we identify a brighter object in the input catalogue within a radius of 3 pixels. As the galaxies are placed randomly, these are mere chance projections, which is consistent with the observed distribution of separations. In these cases we assign the input properties of the brighter object, because a shape measurement algorithm would be more sensitive to its surface brightness distribution.

We then proceed to compute the shear biases by comparing the average ellipticity of the detected galaxies to the input shear γ_i^{true} (where the index $i \in \{1, 2\}$ corresponds to the real or imaginary part of the shear, respectively). The former is an estimate of the shear, as can be seen by averaging Eq. (1): $\langle \epsilon_i^{\text{obs}} \rangle = \gamma_i^{\text{obs}}$. As is common, we assume that the observed shear and true shear

are related as:

$$\gamma_i^{\text{obs}} = (1 + \mu_i)\gamma_i^{\text{true}} + c_i, \quad (2)$$

where μ_i is the multiplicative shear bias, and c_i is the additive shear bias. The values for μ_i are expected to be very similar (Kitching et al. 2019). We determine both components separately, and if they are consistent we refer to μ as the average of the two components. Finally, we note that because we create pairs of images where the galaxies are rotated by 90° the presence of a bias means that one of the two images is not detected, or assigned a magnitude such that it is not included, and that the probability of detection depends on the applied shear itself.

3.1. Detection bias estimates

Figure 1 shows that the presence of neighbouring galaxies affects the ability of SEXTRACTOR to detect galaxies. We now proceed to explore whether this results in a bias in the shear. Unless specified otherwise we report biases for galaxies with $20 < m_{\text{AUTO}} < 24.5$, which was adopted by H17 as a good approximation for the range used by *Euclid*. This allows for a direct comparison to their results for the overall shear bias, although we note that our analysis differs somewhat (see Sect. 2.1 for details). We present results for different setups in Table 2.

For our baseline setup, where galaxies were placed randomly, we measured $\mu_1^{\text{det}} = -0.01061 \pm 0.00013$ and $\mu_2^{\text{det}} = -0.01053 \pm 0.00013$, where the uncertainties reflect the finite number of images that were analysed. We did not detect a significant additive bias, but the detection bias is significant for our *Euclid*-like setup, especially if we contrast this with the overall requirement that $|\mu| < 2 \times 10^{-3}$ (Cropper et al. 2013). Both multiplicative shear biases agree ($\langle \mu_1^{\text{det}} - \mu_2^{\text{det}} \rangle = (-0.9 \pm 1.8) \times 10^{-4}$), which is why we show the average of both components in most figures. In Table 2 we also present the detection bias when we fix the background to its true value (i.e. zero; reported as

⁷ In our case, a noise peak is still associated with an input galaxy, resulting in imperfect shape noise cancellation only. In contrast, including noise peaks in an actual cosmic shear analysis does lower the signal. In practice, however, requiring robust photometric redshifts using multi-band observations will remove most, if not all, of these.

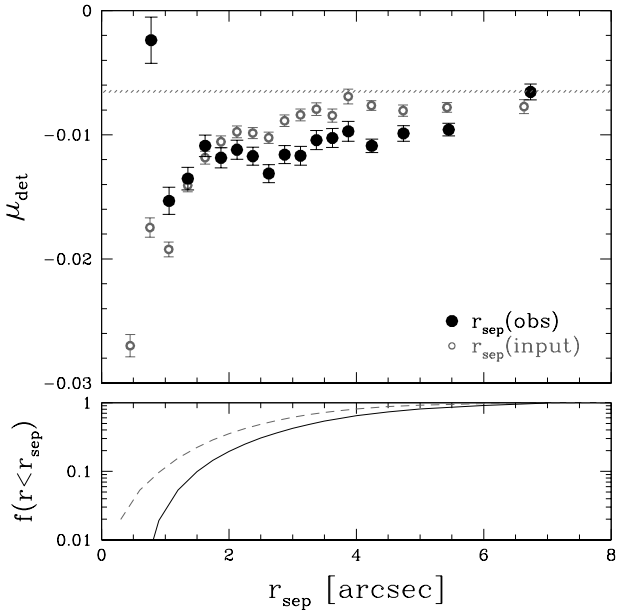


Fig. 3. *Top panel:* detection bias for galaxies with $20 < m_{\text{AUTO}} < 24.5$ as a function of r_{sep} , the distance to the nearest object detected by SEXTRACTOR (black points). The open grey points show the detection bias as a function of the nearest neighbour in the input catalogue brighter than $m_{\text{input}} = 26$ (grey open points). For reference, the hatched region indicates the detection bias for the grid-based simulations. *Bottom panel:* fraction of galaxies that have a neighbour within a distance $< r_{\text{sep}}$ in the input catalogue (grey dashed line) or detection catalogue (black line). For small separations many of the true blends are not recognised as such.

‘no background’). The changes in multiplicative shear bias are small, but significant⁸: $\Delta\mu_1 = 0.00037 \pm 0.00011$ and $\Delta\mu_2 = 0.00062 \pm 0.00011$. In the baseline setup we did not shear the full scene, but only sheared the galaxy images. In reality the shearing also alters the positions, which in turn might affect the results as the separations between neighbouring objects change slightly. If we shear the full image instead, the difference with respect to the baseline case where we only shear the galaxy images is $\Delta\mu = (0.52 \pm 1.59) \times 10^{-4}$, an insignificant difference. Similarly the additive biases are consistent with the baseline results. To obtain this estimate we used the fact that the galaxy images are the same for both setups (though not their positions), but that the background noise realisation is slightly different.

The black points in Fig. 3 show the detection bias for galaxies with $20 < m_{\text{AUTO}} < 24.5$ as a function of r_{sep} , the distance to the nearest object *detected* by SEXTRACTOR. For large separation, the bias approaches the average bias we measured for the grid-based simulations (indicated by the hatched region), but is typically larger. This is because not all blends are identified as such. For reference, we also show the bias as a function of the distance to the nearest neighbour in the input catalogue brighter than $m_{\text{input}} = 26$ (grey open points). The amplitude of the bias changes rapidly for galaxies with $r_{\text{sep}} < 1''$, and such galaxies are probably best omitted from the analysis. The bottom panel in Fig. 3 shows that this applies to about 10% of the galaxies. In reality this number will be higher because of clustering (Euclid Collaboration 2019).

⁸ The measurements for different scenarios are based on the same images, and are therefore correlated. We account for this by computing the difference first and reporting its statistics. As a result, the difference may be determined more precisely than the bias itself.

As indicated by Fig. 2, selecting objects with SEXTRACTOR FLAG=0 reduced the occurrence of blends, and we expect the detection bias to be reduced (see Fig. 3). Instead we find that the bias increased by about 13%, implying that the flagging of blended objects is actually done in a shear dependent fashion.

These results indicate that blending is a significant source of detection bias that depends significantly on the local galaxy density. We note, however, that the bias does not vanish for large separations, but rather converges to the bias we obtained for our grid-based simulations, indicated by the hatched horizontal region (and reported in Table 2), if we select galaxies based on the distance to the nearest neighbour in the input catalogue. In the more realistic case (open grey points), where we separate galaxies based on the distance to the nearest detected galaxy, the bias is even larger because many blends remain undetected.

To investigate this further, we show μ_{det} as a function of magnitude in Fig. 4. The left panel, where we show results as a function of the input magnitude, m_{input} , is the shear detection bias equivalent of Fig. 1. In this case the shape noise cancellation results in small uncertainties, because galaxies are included in the correct magnitude bin by design. The shear bias arises because the probability of detecting faint galaxies is affected by the orientation of the galaxy with respect to the applied shear: galaxies that are aligned perpendicular to the shear are more likely to be detected. The bias is negligible for bright galaxies, and thus can be reduced by increasing the depth of the observations, something we explore further in Sect. 3.2.

Similar to Fig. 1, we find that the bias for isolated galaxies ($r_{\text{sep}}^{\text{in}} > 5''$) matches that of the grid-based images, whereas the bias is larger for blended galaxies ($r_{\text{sep}}^{\text{in}} < 2''$). Comparison of the biases reported in Table 2 suggests that both blending and the shear-dependent detection probability are important. The bias at bright magnitudes is caused by blending, whereas for fainter galaxies the detection probability itself depends on the orientation with respect to the applied shear.

In reality the situation is complicated by the fact that the observed magnitudes are affected by blending, the applied shear, and measurement uncertainties, all of which spread the biases over a wider range in magnitudes and lead to larger uncertainties owing to imperfect shape noise cancellation. Consequently, the error bars in the right panel of Fig. 4 are increased, and the detection bias affects a larger range in magnitude. In particular, as shown by the asymmetric distribution of magnitude errors in Fig. 2, blending scatters objects towards a brighter magnitude bin. Such blends are not always identified, and can thus introduce significant biases even for apparently bright galaxies. For instance, the bias for the bright blended galaxies is far beyond the axis limits of the chart. We also caution that the results for the brightest magnitude bin suffer from extreme Eddington bias, because our input catalogue does not include galaxies brighter than $m = 20$.

Figure 5 shows the multiplicative detection bias as a function of the input half-light radius (r_{eff}) for the baseline (black) and grid-based (red) simulations. For both cases we observe a strong dependence with galaxy size, which is the combined result from the underlying distribution of fluxes and the correlation between size and brightness. After all, brighter galaxies are more likely to be detected, whilst for a given flux a smaller galaxy is detected with a higher significance. The latter drives the increase in detection bias with increasing r_{eff} , but as the mean brightness increases with increasing size, the probability of detection increases once more. Comparison of the bias as a function of r_{eff} for isolated galaxies with the grid-based results show that they agree well. Hence, the difference between

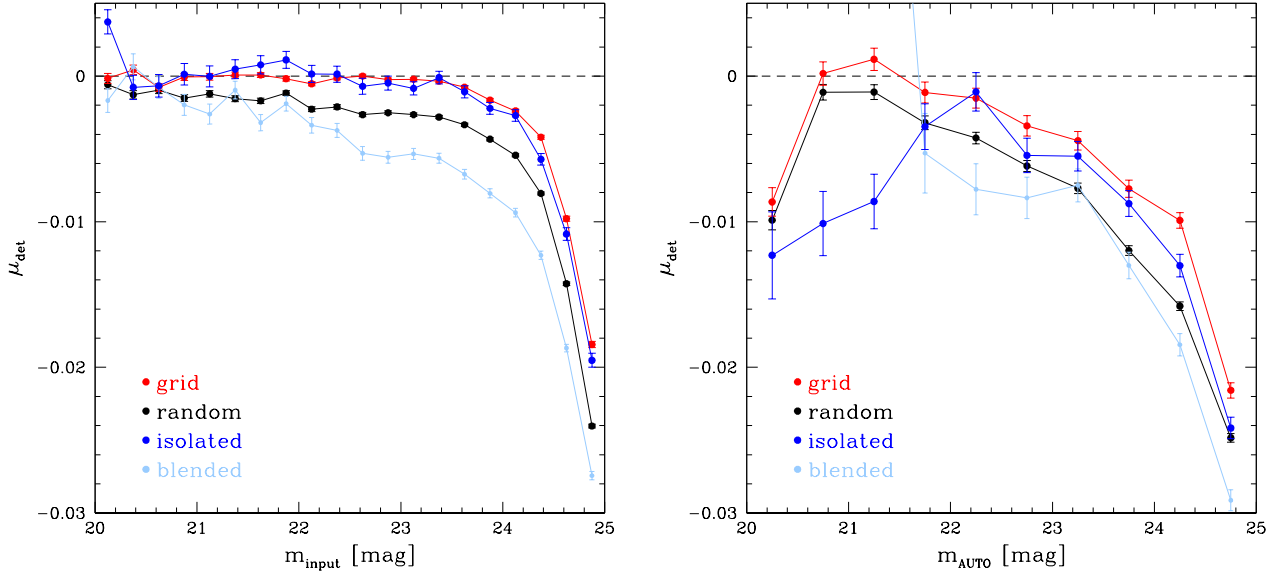


Fig. 4. *Left panel:* multiplicative detection bias μ_{det} as a function of the input apparent magnitude when galaxies are placed on a grid (red points) or placed randomly (black points). The blue points show the results for isolated galaxies where the nearest neighbour is more than $5''$ away, whereas the light blue points show the detection bias for galaxies with a neighbour within $2''$ (blended). *Right panel:* multiplicative detection bias as a function of observed properties. The classification into isolated and blended galaxies is based on the nearest *detected* galaxy in this case. The lines connect the points to show the behaviour for the different samples more clearly. The bias for the bright blended galaxies is beyond the axis limits of the chart.

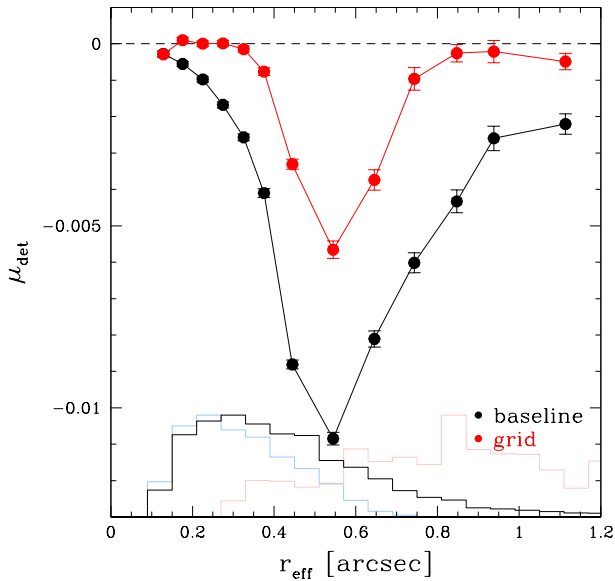


Fig. 5. Multiplicative detection bias μ_{det} as a function of the input half-light radius, r_{eff} , for galaxies with $20 < m_{\text{AUTO}} < 24.5$. The black and red lines correspond to the baseline and grid-based cases, respectively. The histograms show the distributions of galaxy sizes (black: all galaxies; red: $m_{\text{AUTO}} < 21$; blue: $24 < m_{\text{AUTO}} < 24.5$). The observed behaviour is the result of the change in size as a function of brightness.

the grid-based and baseline simulations is caused by blending, which affects galaxies of all sizes.

In contrast to what was done in Fig. 4, we do not show the bias as a function of FLUX_RADIUS, the half-light radius determined by SEXTRACTOR, because it correlates with ellipticity. Consequently, a split by FLUX_RADIUS is an implicit selection in ellipticity, resulting in large biases. If one wants to split the source sample by a particular observable, it is important to verify that it does not correlate with input ellipticity. This may not

be fully feasible in practice, but at least one should aim to minimise the dependence. Interestingly, we find that MAG_AUTO only weakly correlates with the input ellipticity. This suggests that splitting the sample into tomographic bins based on magnitude and colour may not increase the selection bias much, although further study would be required to quantify this.

3.2. Dependence on noise level

Figure 4 shows that the detection bias is negligible for bright, isolated galaxies. Hence, we expect that the detection bias can be reduced by obtaining deeper data. The results in Fig. 6 show that this is indeed the case: it shows the multiplicative detection bias when the noise level in the image is multiplied by f_{noise} (where $f_{\text{noise}} = 1$ corresponds to the baseline case). The black (red) points show the results for the baseline (grid) simulations for galaxies with $20 < m_{\text{AUTO}} < 24.5$. These are well fit by a second order polynomial (solid lines).

The average increase in detection bias of $\mu_{\text{base}} - \mu_{\text{grid}} = -0.0035$ is caused by blending and increases only weakly with increasing noise level. Moreover, even for low noise levels blending leads to a floor in the detection bias that is about ~ -0.004 . Interestingly, the bias does not completely vanish in the grid-based simulations at low noise levels. This is the result of our galaxy selection, which is based on the magnitude estimates by SEXTRACTOR. If we instead select the galaxies based on their true (but unobservable) magnitudes, the bias quickly vanishes (light red points and red dashed line). This implies that the estimate of m_{AUTO} depends slightly on the shear. For the baseline case (light grey points) the bias plateaus to $\mu_{\text{det}} = -0.0024$ as a result of blending.

These results show that the detection bias is a combination of blending and the sample selection (in our case a magnitude cut). Although we find that it may be possible to reduce detection bias somewhat using deeper observations, blending quickly becomes a limiting factor, even in space-based data.

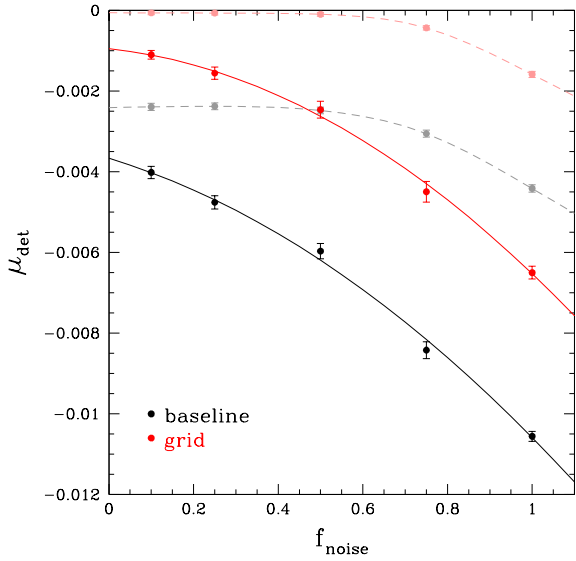


Fig. 6. Multiplicative detection bias μ_{det} as a function of the background noise level, which is multiplied by a factor f_{noise} with respect to the baseline case. The black and red lines correspond to the baseline and grid-based cases, respectively. The solid lines show results for galaxies with $20 < m_{\text{AUTO}} < 24.5$, whereas the (light-coloured) dashed lines indicate the bias if we select using the input magnitudes, $20 < m_{\text{input}} < 24.5$. In the latter case the bias vanishes for the grid-based case as the noise level is low, but for the baseline case the bias plateaus to $\mu_{\text{det}} = -0.0024$ as a result of blending.

3.3. KSB biases

In Table 2 we also present measurements for the shear biases for the KSB algorithm (Kaiser et al. 1995; Hoekstra et al. 1998), because we made a number of changes in both the simulations and the measurement setup since H17 (see Sect. 2). With this modified setup we measured a total shear bias of $\mu_1^{\text{KSB}} = -0.08915 \pm 0.00031$ and $\mu_2^{\text{KSB}} = -0.08757 \pm 0.00032$. The results also suggest that a small additive bias was introduced, although more simulations would be needed to confirm the result. The detection bias is about 9 times smaller than the total shear bias, which explains the focus of previous studies on shear bias.

We also report the biases introduced by the steps in the shape measurement analysis following the initial SEXTRACTOR detection. The ability of the KSB algorithm to measure a shape also depends on the shear, resulting in an increase in the detection bias. For the ‘KSB detection’ bias we used the true shapes, but only for those galaxies where a shape was measured. The results in Table 2 show that the bias doubles for all image setups. The bias is reduced somewhat if we weight the true ellipticities with the KSB weights (‘KSB selection’).

Although the shape measurement bias itself is dominant, the detection bias is not negligible. As the detection bias is most readily quantified using image simulations, like the one we use here, we need to quantify the sensitivity of the detection bias to the simulation setup, similar to what was done by H17 for the overall shear bias. We return to this in Sect. 4, but first examine the sensitivity to the SEXTRACTOR setup and PSF anisotropy.

3.4. Sensitivity to detection setup

Table 1 lists the main parameters that play a role in the object detection. These can be grouped into three categories. The first three pertain to the detection itself, the next three affect the

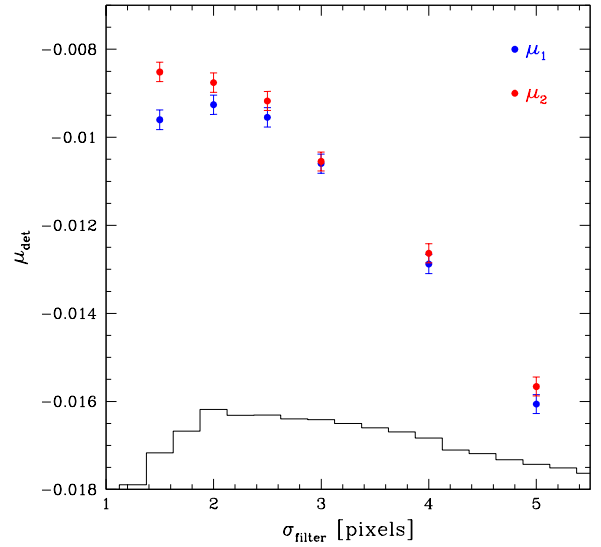


Fig. 7. Multiplicative detection bias μ_{det} as a function of the width of the filter used in the detection step for galaxies with $20 < m_{\text{AUTO}} < 24.5$. The blue (red) points correspond to μ_1 (μ_2). The histogram shows the distribution of corresponding sizes based on the half-light radius of the galaxies, suggesting that a width of 2–3 pixels is best. The bias increases quickly for larger values of σ_{filter} .

behaviour for blended objects, and the last two are relevant for the background estimation. As already mentioned, the background parameters do not play an important role for our study. Also the choices for DETECT_MINAREA and DETECT_THRESH do not affect our findings for galaxies with $20 < m_{\text{AUTO}} < 24.5$ (provided they are not modified significantly), but the choice of the filter that is used to detect objects is relevant. To detect objects in the presence of noise, the images are convolved with a suitable kernel before searching for peaks. The optimal filter has a profile that matches the object of interest. For this reason Kaiser et al. (1995) developed a hierarchical peak finder, which employs a series of filters, but is slower. SEXTRACTOR is run with a single filter, specified by the keyword FILTER_NAME. Here we run it using the various predefined round Gaussian filters, defined by their dispersion σ_{filter} .

The results are presented in Fig. 7 for galaxies with $20 < m_{\text{AUTO}} < 24.5$, where we show the multiplicative detection biases for the two shear components separately. They show a similar behaviour with filter width σ_{filter} , but we observe a small offset, which is more significant for smaller filter sizes. The histogram shows the distribution of corresponding sizes based on the half-light radii of the galaxies, suggesting that using a Gaussian filter with a width of 2–3 pixels is best. The bias increases quickly for larger values of σ_{filter} .

Figures 1 and 4 show that the presence of neighbouring objects affects the detection and introduces detection bias. We explore how changes in the parameters that affect the deblending of objects (DEBLEND_MINCONT, DEBLEND_NTHRESH, and CLEAN_PARAM) in Appendix A. We find that the detection biases for the default parameters are close to optimal, and that even substantial variations have only a minimal impact. Hence, the observed detection biases are not the result of a poorly chosen setup of SEXTRACTOR.

3.5. Sensitivity to PSF anisotropy

Thus far we focused only on the multiplicative detection bias that arises because the probability of detecting a galaxy depends on

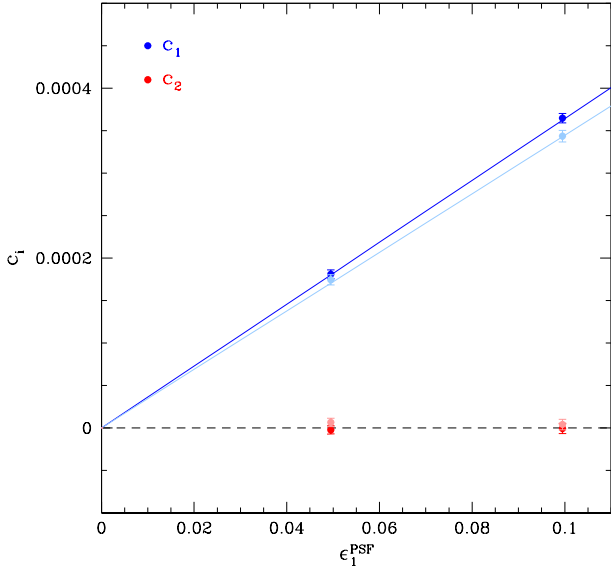


Fig. 8. Additive bias c_1 (blue) and c_2 (red) as a function of the PSF ellipticity ϵ_1^{PSF} . The bright colours correspond to the baseline case where galaxies are placed randomly, whereas the light coloured points were obtained by placing galaxies on a grid. In the former case the additive detection bias is about 5.6% higher, but in both cases galaxies are preferentially detected when their orientation is aligned with the PSF. We do not observe a significant c_2 (red points), nor a change in multiplicative bias (not shown).

its orientation with respect to the shear (Hirata & Seljak 2003). However, we expect the PSF to be anisotropic due to optical aberrations that are practically unavoidable, especially for a wide field imager. Such PSF anisotropy also introduces a preferred direction. In this case surface brightness is not conserved, and a galaxy with an intrinsic orientation parallel to the PSF ellipticity direction will have a higher peak brightness compared to a galaxy oriented orthogonal to the PSF anisotropy. As a consequence, we expect to preferentially detect galaxies that are aligned with the PSF anisotropy, leading to a positive additive bias (Kaiser 2000; Bernstein & Jarvis 2002).

To study this, we created simulated images where the PSF was made elliptical in the ϵ_1 direction and ran SEXTRACTOR to quantify the additive and multiplicative shear biases. Figure 8 shows the resulting additive bias c_i . We find that c_2 is consistent with zero (red and light red points), but we find that the object detection introduces a significant additive shear bias c_1 , both when galaxies are placed on a grid (light blue points) or placed randomly (blue points); the bias in the latter case is only 5.6% higher.

As expected, the bias has the same sign as the PSF anisotropy, demonstrating that SEXTRACTOR preferentially selects objects that are aligned with the PSF (this was also observed in Kannawadi et al. 2019). Although the amplitude is small, only 0.4% of the original PSF ellipticity, this bias cannot be ignored if the PSF is anisotropic. For instance, Cropper et al. (2013) argue that $|c| < 5 \times 10^{-4}$ is required, which is reached for $\epsilon^{\text{PSF}} = 0.137$. PSF anisotropy is therefore a non-negligible source of additive detection bias, which will vary spatially because we expect the PSF ellipticity to change across the field-of-view.

We also examined the change in multiplicative shear bias as a function of ϵ^{PSF} and we found no significant trend. This is worth noting, because we show in Appendix B that sources of additive bias tend to introduce multiplicative biases of similar amplitude,

but opposite sign in shape measurements. This connection can be used to empirically estimate the level of multiplicative bias for (residual) systematic effects that cause additive bias. In contrast, the lack of a change in multiplicative detection bias in the case of an anisotropic PSF shows that detection bias is fundamentally different from the shape measurement process itself.

4. Realism of the simulations

The blending of galaxies is a significant source of shear bias, and for a reliable estimate of the bias it is therefore critical to capture this in the simulated data. Studying the performance of galaxies on a grid may help in the comparison of methods, or the training of machine learning approaches (Gruen et al. 2010; Tewes et al. 2019; Pujol et al. 2020), but the actual estimate relies on realistic simulations. In this Section we explore how the detection bias depends on the properties of the simulated galaxies, such as their size and ellipticity distributions.

The realism is, however, not limited to the properties of the detected galaxies, because the performance of the shape measurements is also influenced by the presence of galaxies below the detection limit. This was first demonstrated by Hoekstra et al. (2015) for ground-based observations. Similarly, H17 showed that for the *Euclid*-like data we consider here, the multiplicative shear bias depends on m_{lim} , the apparent magnitude of the faintest galaxies that are included in the image simulation. They found that galaxies as faint as $m_{\text{lim}} = 29$ can modify the multiplicative bias for the KSB algorithm.

The impact of very faint galaxies was studied in more detail in *Euclid* Collaboration (2019), who found that the dependency with m_{lim} also depends on the shape measurement algorithm, and how it deals with blending. In our KSB setup the nearby objects are crudely masked, but no attempt is made to correct the surface brightness profile, thus biasing the estimates of the moments. Model fitting methods will generally do better in this regard, in line with the findings of *Euclid* Collaboration (2019). The clustering of galaxies results in a higher level of blending around brighter galaxies, and consequently, *Euclid* Collaboration (2019) showed that the clustering of the faint galaxies increases the overall bias further. We do not consider this additional complication here, but note its importance when one aims to calibrate a shear measurement algorithm to be applied to actual data.

These studies only considered the final shear bias, but in Fig. 9 we show how the SEXTRACTOR detection bias depends on m_{lim} . Our results show that the detection bias is much less sensitive to the inclusion of faint galaxies, especially when compared to the KSB shear estimates (indicated by the light grey points and dashed line). The dotted line indicates the change in bias when we select galaxies based on m_{input} . This shows that the bias partly arises from faint galaxies, for which the detection bias is larger (see Fig. 4), scattering into the sample of sources used in the analysis (defined as $20 < m_{\text{AUTO}} < 24.5$). Nonetheless, the convergence is only achieved for $m_{\text{lim}} = 27$, still 2.5 mag fainter than the magnitude limit of the sample of sources that we consider here.

4.1. Sensitivity to galaxy number density

H17 (their Fig. 5) showed that the KSB shear bias increases if the number density of the simulated galaxies is increased by a factor n_{fac} (also see Table 2). Consequently the bias will be larger near clusters and groups of galaxies, thus coupling the shear bias to the large-scale structure, which will need to be accounted for as shown by Hartlap et al. (2011). An increase in detection bias

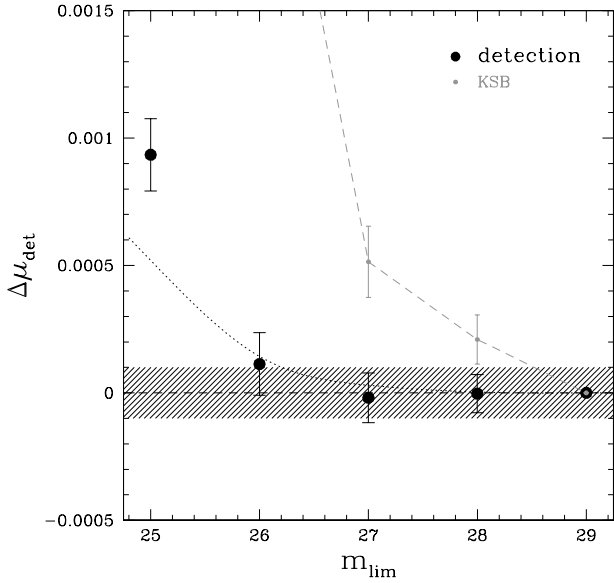


Fig. 9. Change in multiplicative detection bias $\Delta\mu_{\text{det}}$ (with respect to $\mu(m_{\text{lim}} = 29)$) for galaxies with $20 < m_{\text{AUTO}} < 24.5$ as a function of m_{lim} , the magnitude of the faintest galaxies that are included in the simulation (black points). The dotted line shows the change in bias if we select galaxies based on their input magnitude ($20 < m_{\text{input}} < 24.5$). The change in multiplicative bias for the KSB algorithm is indicated by the light grey points. The hatched region indicates a tolerance of 10^{-4} .

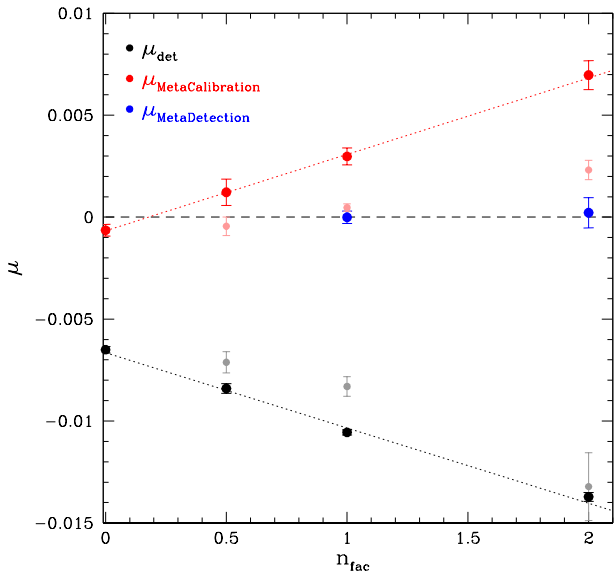


Fig. 10. Multiplicative bias as a function of n_{fac} , the relative increase in galaxy number density with respect to the baseline simulation. The grid-based results correspond to $n_{\text{fac}} = 0$. The black points show how the detection bias increases with n_{fac} . The red (blue) points correspond to the METACALIBRATION (METADETECTION) results discussed in Sect. 5 (Sect. 6). The light coloured points show the biases for relatively isolated galaxies (distance to nearest galaxy in the input catalogue larger than $2''$).

will play a role, because Fig. 3 shows that it depends on the distance to the nearest galaxy. As the density increases, the mean separation decreases and the bias increases accordingly.

We quantify the sensitivity of the detection bias to the galaxy number density in Fig. 10. The black points show μ_{det} as a function of n_{fac} , where $n_{\text{fac}} = 0$ corresponds to the grid-based simulations (no blending) and $n_{\text{fac}} = 1$ is our baseline case. For

reference, a value of $n_{\text{fac}} = 2$ roughly corresponds to the galaxy density in the innermost regions of a massive cluster of galaxies (see e.g., Fig. 11 in Hoekstra et al. 2015). We find that the detection bias increases linearly with increasing galaxy density, with $\partial\mu_{\text{det}}/\partial n_{\text{fac}} = -0.00369 \pm 0.00013$. As blending is a likely cause we repeat the measurements for a sample of relatively isolated galaxies (i.e. no neighbour brighter than $m = 26$ in the input catalogue within $2''$) and show the results as light grey points in Fig. 10. The slope is almost halved, but not fully eliminated.

The spatial variation in n_{fac} caused by the clustering of galaxies will lead to spatial variations in the multiplicative bias across the survey. Provided these variations are small, the impact on the cosmological signal is expected to be negligible, as shown in Kitching et al. (2019). However, it is important that the galaxy number density in the simulations matches the average value in the survey, because a mismatch results in an overall shift in the shear bias. We discuss the area of high-quality data that is needed to achieve this in Appendix C.

4.2. Sensitivity to morphology

The detection bias depends on the morphology of the galaxies, because the size affects the signal-to-noise ratio and the incidence of blending. Moreover, the bias depends on the intrinsic ellipticity: the detection bias vanishes if $\epsilon^s = 0$, whereas we observe a significant detection bias for our reference setup. Such dependencies on morphology are of particular concern, because they vary with redshift (Kannawadi et al. 2015), and can link shear biases to the lensing signal as the morphology depends on the galaxy density: early type galaxies are generally larger and rounder, and occupy higher density regions. Moreover, their photometric redshifts are typically more precise thanks to their more pronounced 4000 \AA break, coupling the shear measurements to the binning of galaxies into tomographic bins. These connections highlight the need for simulations that capture the full process of photometric redshift and shear estimation simultaneously. This is, however, left for future study.

To explore the impact of uncertainties in the morphology further we analysed images where the input sizes were increased by a factor f_{size} and where the input ellipticities were increased by a factor ϵ_{fac} , similar to what was done in H17 (see their Figs. 4 and 10). The black points in Fig. 11 show the change in bias as a function of these parameters. The left panel of Fig. 11 shows that the detection bias increases linearly with increasing input galaxy sizes, with a slope $\partial\mu_{\text{det}}/\partial f_{\text{size}} = -0.0211 \pm 0.0006$. We expect the bias to be smaller if the galaxies are smaller, because the galaxies will be detected with a higher signal-to-noise ratio for a given magnitude, whilst blending is reduced. Although this dependence is rather steep, the distribution of galaxy sizes is fairly well established, and mismatches between the simulations and the data can be accounted for empirically (see the discussion in H17).

The sensitivity to the input ellipticity distribution is more worrisome, because it is generally more difficult to infer from existing high-quality *Hubble* Space Telescope observations. We found $\partial\mu_{\text{det}}/\partial\epsilon_{\text{fac}} = -0.02502 \pm 0.00056$, which is about half the value that H17 measured for the full KSB bias. This suggests that a significant part of the sensitivity to the input ellipticity distribution is determined by the detection bias. Deeper observations may help improve empirical constraints on the ellipticity distribution (Viola et al. 2014), but the measurements still require an accurate algorithm to measure shapes. Moreover, the results presented in Sect. 3.2 suggest that blending limits the gain of such deeper observations. This requires further study, because

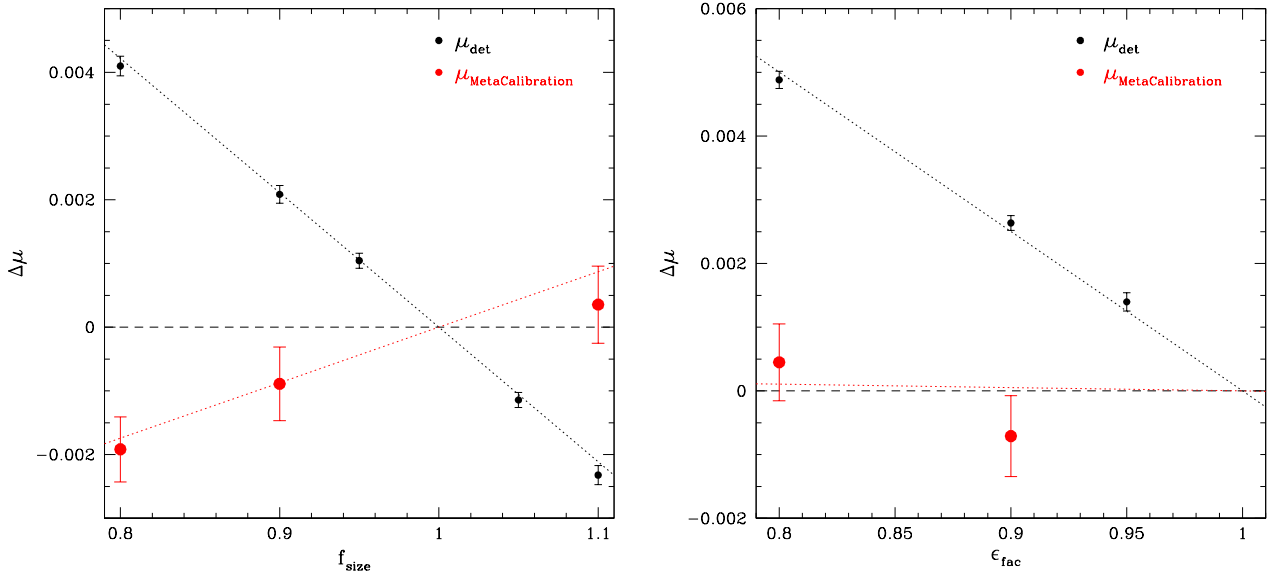


Fig. 11. *Left panel:* change in multiplicative shear detection bias $\Delta\mu$ as a function of f_{size} , the relative change in input galaxy size (black points). *Right panel:* change in multiplicative detection bias if the input ellipticities are multiplied by a factor ϵ_{fac} . The dotted lines show the best fit linear model. The red points in both panels correspond to the post-METACALIBRATION results discussed in Sect. 5.

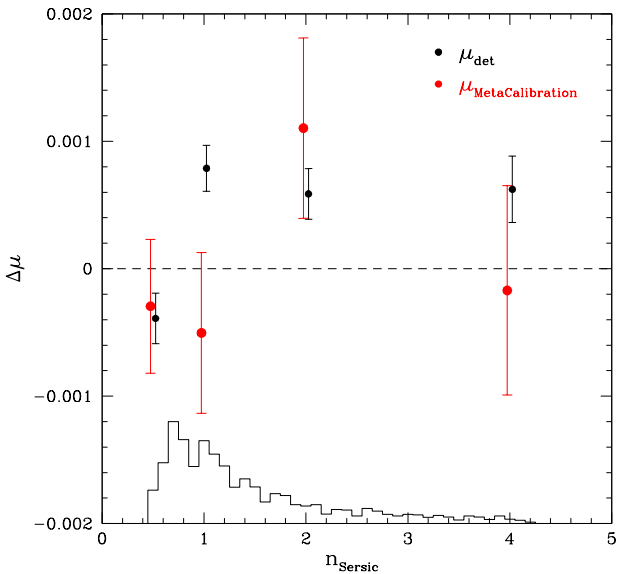


Fig. 12. Change in multiplicative shear bias $\Delta\mu$ as a function Sérsic index. The histogram indicates the distribution of Sérsic indices in the baseline simulations. The black points show the change in detection bias. The red points show the METACALIBRATION results.

Kannawadi et al. (2019) showed that the ellipticity distribution correlates with galaxy size and changes with redshift, whilst ellipticity gradients will complicate matters further.

The left panel of Fig. 11 shows that the detection bias is reduced if galaxies are smaller, as such galaxies are easier to detect, whilst blending is reduced. We therefore expect the radial surface brightness profile to influence the bias as well. We explored two modifications, namely the sensitivity to changes in the Sérsic index, n_{Sersic} , and r_{trunc} , the radius where the profile is truncated.

The black points in Fig. 12 show the change in detection bias when we keep the effective radii, fluxes and ellipticities of the galaxies the same, but fix the Sérsic indices to a single value.

Larger values for n_{Sersic} result in profiles that are more centrally peaked, reducing the detection bias. Indeed, the bias is reduced slightly with respect to the baseline case for $n_{\text{Sersic}} \geq 1$. The histogram shows the baseline distribution of n_{Sersic} , which peaks at values < 1 .

Throughout this paper we assume that the surface brightness profiles of galaxies are described by a Sérsic-profile, which are truncated at $r_{\text{trunc}} = 3.5$ effective radii for reasons of computational speed. This ignores much of the variety in galaxy morphology, where the bulge and disk components may have different ellipticities and orientations. Moreover, spiral structure complicates matters further. Better modelling of the morphologies of galaxies using deep, high-quality data will help addressing this specific problem. A less explored question, however, is the surface brightness profile at large radii. Tal & van Dokkum (2011) stacked the images of a large sample of luminous red galaxies and found that a Sérsic-profile describes the data well out to more than 7 effective radii. In contrast, detailed studies of edge-on spiral galaxies indicate that the disks are truncated around four disk scale lengths on average (Kregel et al. 2002).

We therefore created images where we truncated the profile at different values for r_{trunc} (in units of the effective radius r_{eff}). The black points in Fig. 13 show the change in SEXTRACTOR detection bias, relative to the case of $r_{\text{trunc}} = 10$. The change is small for $r_{\text{trunc}} > 3.5$, indicating that it is important to accurately capture the surface brightness out to these radii.

The results presented in this section highlight the importance of capturing the morphological diversity of galaxies with sufficient accuracy. This seems quite feasible in the case of detection bias alone, but we expect the actual shear bias to be affected more. This is evidence from Fig. 9, where the KSB bias is sensitive to very faint galaxies, whereas the SEXTRACTOR detection bias converges at $m_{\text{lim}} = 27$ already. Similarly, H17 found steeper dependencies for many parameters. The key question is therefore whether image simulations can be made sufficiently realistic to capture the redshift-dependent morphologies of galaxies for Stage IV surveys. As this appears to be challenging, we explore next a different approach that uses the survey data to calibrate the shear estimate instead.

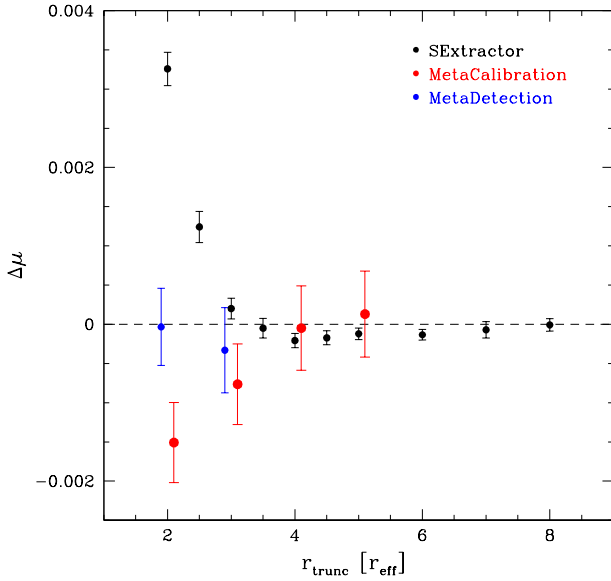


Fig. 13. Change in multiplicative shear bias $\Delta\mu$ as a function of r_{trunc} , the radius where the galaxy profile is truncated in the simulated images in units of the input half-light radius, r_{eff} . The black points show the change in SExtractor detection bias. The red (blue) points show the METACALIBRATION (METADETECTION) results discussed in Sect. 5 (Sect. 6).

5. MetaCalibration

A different approach is to use the observations themselves to determine the response of an ensemble of galaxies to a shear. Huff & Mandelbaum (2017) worked out how one can estimate the shear bias by shearing the images, whilst taking the PSF and noise into account. They refer to this data-driven approach as METACALIBRATION, and in this section we explore its potential to calibrate the multiplicative bias for our *Euclid*-like simulations. In principle METACALIBRATION can also be used to correct for PSF anisotropy, but in the following we only consider the calibration of multiplicative shear bias, which allows us to limit the study to our round Airy PSF.

The only assumption of METACALIBRATION is that we can construct a sheared version, $I^{\text{sh}}(\mathbf{x}|\boldsymbol{\gamma})$ of the true image using the observed image $I(\mathbf{x})$ via Eq. (5) of Huff & Mandelbaum (2017):

$$I^{\text{sh}}(\mathbf{x}|\boldsymbol{\gamma}) = P(\mathbf{x}) * [\hat{s}_{\boldsymbol{\gamma}}\{P(\mathbf{x})^{-1} * I(\mathbf{x})\}]. \quad (3)$$

where $\hat{s}_{\boldsymbol{\gamma}}$ is the shear operator (Bernstein & Jarvis 2002), $P(\mathbf{x})$ is the PSF, $I(\mathbf{x})$ the observed image, and ‘*’ indicates convolution. Hence, the observed image is first deconvolved ($P(\mathbf{x})^{-1} * I(\mathbf{x})$), then sheared by $\hat{s}_{\boldsymbol{\gamma}}$, and finally re-convolved by the PSF. This procedure, in its simplest form, only requires an accurate model of the PSF.

In practice, noise in the data complicates the deconvolution step, and a slightly larger PSF is needed to suppress the noise. The modified PSF, $P^{\text{meta}}(\mathbf{x})$, to use in the reconvolution step in Eq. (3) is (Huff & Mandelbaum 2017)

$$P^{\text{meta}}(\mathbf{x}) = P(\mathbf{x}/(1 + 2|\boldsymbol{\gamma}|)). \quad (4)$$

These steps implicitly assume that the images are well sampled, so that the image manipulations are not compromised. However, in the case of both *Euclid* and the *Roman* Space Telescope, the pixels are large compared to the PSF size. In our calculations we do assume that we can construct a well-sampled model of the PSF, but the images of the smallest galaxies might still be affected by undersampling. Kannawadi et al.

(2021) explore ways to mitigate this, but we note that image simulations can also be used to correct for the biases that may be introduced.

Another complication is that the shearing of the images leads to anisotropic correlated noise, which needs to be accounted for. One possibility is to determine the resulting bias using image simulations, but Sheldon & Huff (2017, SH17 hereafter) show that this problem can also be mitigated by adding anisotropic noise. The latter approach does lead to a slight increase in the overall noise level, but as shape noise typically dominates, this is only a minor concern. There are other complications that are particularly relevant for space-based observations, such as the wavelength-dependence of the PSF, which we discuss in more detail in Sect. 7.

If we use Eq. (3) to apply a small shear $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)$ to a galaxy image, and measure its shape $\mathbf{e} = (e_1, e_2)$ we can relate the resulting shape to the original value $\mathbf{e}_{\boldsymbol{\gamma}=\mathbf{0}}$, because

$$\mathbf{e} \approx \mathbf{e}|_{\boldsymbol{\gamma}=\mathbf{0}} + \left. \frac{\partial \mathbf{e}}{\partial \boldsymbol{\gamma}} \right|_{\boldsymbol{\gamma}=\mathbf{0}} \boldsymbol{\gamma} \equiv \mathbf{e}|_{\boldsymbol{\gamma}=\mathbf{0}} + \mathbf{R}^{\boldsymbol{\gamma}} \boldsymbol{\gamma}, \quad (5)$$

where $\mathbf{R}^{\boldsymbol{\gamma}}$ is the 2×2 shear response tensor. We can estimate its elements by measuring the shapes of the galaxies in the sheared images and computing

$$\mathbf{R}^{\boldsymbol{\gamma}}_{ij} = \frac{e_i^+ - e_i^-}{\Delta\gamma_j} \quad (6)$$

where the subscripts indicate the two shear components, and the superscript the sign of the applied shear, so that ‘+’ means the image was sheared by $+\gamma_j$, etc; hence, $\Delta\gamma_j = 2\gamma_j$.

This expression is true for any shape measurement, and it allows us to estimate the shear, $\hat{\boldsymbol{\gamma}}$ for an ensemble of galaxies (as $\langle \mathbf{e} \rangle|_{\boldsymbol{\gamma}=\mathbf{0}} \approx \mathbf{0}$)

$$\hat{\boldsymbol{\gamma}} \approx \langle \mathbf{R}^{\boldsymbol{\gamma}} \rangle^{-1} \langle \mathbf{e} \rangle = \langle \mathbf{R}^{\boldsymbol{\gamma}} \rangle^{-1} \langle \mathbf{R}^{\boldsymbol{\gamma}} \boldsymbol{\gamma} \rangle, \quad (7)$$

where the shape measurements are obtained from the image that is convolved with $P^{\text{meta}}(\mathbf{x})$. Hence, we average the estimates for the shapes and the shear responses, rather than using estimates per galaxy. The reason is that the estimates for $\mathbf{R}^{\boldsymbol{\gamma}}$ are very noisy for individual galaxies, and averaging reduces biases in the shear estimate, which requires the inverse of $\mathbf{R}^{\boldsymbol{\gamma}}$. To reduce the noise even further we average the estimates for $\mathbf{R}^{\boldsymbol{\gamma}}$ for a particular selection of galaxies over many images (typically 3300). We verified that $\mathbf{R}^{\boldsymbol{\gamma}}$ does not change as a function of shear in the simulated images. Moreover, we found that the off-diagonal elements vanish and we therefore assume that $\mathbf{R}^{\boldsymbol{\gamma}}$ is diagonal in the remainder of this paper.

Equation (7) shows that the resulting shear estimate for the ensemble of galaxies is actually weighted by $\mathbf{R}^{\boldsymbol{\gamma}}$, and hence one would like to use a shape measurement algorithm so that $\mathbf{R}^{\boldsymbol{\gamma}} \approx \mathbf{I}$. This is, however, not an immediate concern for our study, because the PSF is isotropic and the same shear is applied to all the simulated galaxies.

In principle it should not matter what shape measurement we use, because any intrinsic bias in the estimator will be accounted for by METACALIBRATION. We therefore simply use the polarisation χ ,

$$\chi_1 = \frac{Q_{11} - Q_{22}}{Q_{11} + Q_{22}}, \quad \text{and} \quad \chi_2 = \frac{2Q_{12}}{Q_{11} + Q_{22}}, \quad (8)$$

where the weighted quadrupole moments Q_{ij} are defined as

$$Q_{ij} = \int d^2\mathbf{x} x_i x_j W(\mathbf{x}) I(\mathbf{x}), \quad (9)$$

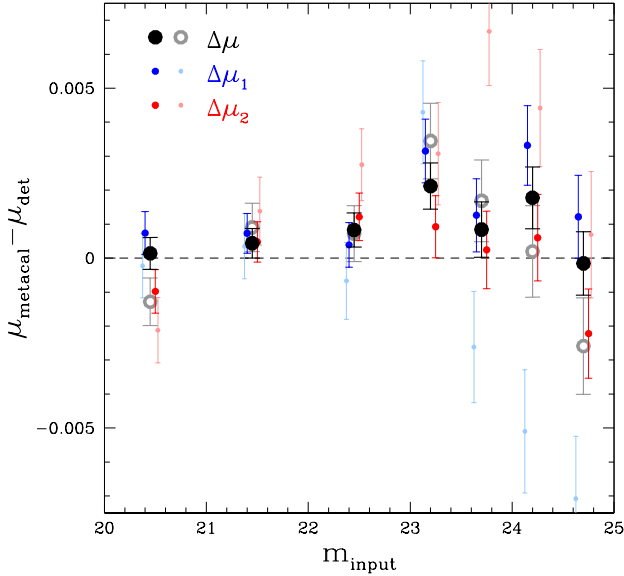


Fig. 14. Difference between the multiplicative bias after METACALIBRATION, μ_{metacal} and detection bias, μ_{det} , as a function of the input magnitude m_{input} for the grid-based simulations. The bright (light) colours show the results when we apply a shear of ± 0.02 (± 0.01) in the metacalibration step. The solid black (open grey) points show the average bias, and the blue (red) points indicate $\Delta\mu_1$ ($\Delta\mu_2$). Using a larger shear results in smaller uncertainties and a better agreement between the two shear components.

and $W(\mathbf{x})$ is the weight function, for which we use a Gaussian with a fixed value for the dispersion of σ_w . Hence, we do not try to optimise the width of the weight function to each object, nor do we try to correct for blending of objects.

The use of a fixed value for σ_w has the advantage that the measurement does not depend on the observed size of the object, which will differ for the different sheared versions of the images as it correlates with the shear, so that \mathbf{R}^γ fully captures the shear response in the absence of detection bias. We adopt $\sigma_w = 2$ pixels (i.e. $\sigma_w = 0''.2$) as our baseline, which is a reasonable value to use for the galaxies in our simulations, as suggested by Fig. 7. Moreover, as shown in our companion paper (Kannawadi et al. 2021), this weight function is wide enough to avoid aliasing bias.

We describe and test our METACALIBRATION setup using the grid-based simulations in Sect. 5.1. We study the performance on our more realistic baseline simulations in Sect. 5.2, which enable us to quantify the impact of blending. We also explore the sensitivity of the post-METACALIBRATION bias to changes in the galaxy number density and morphology. The prospects of METADETECTION (Sheldon et al. 2020) are examined in Sect. 6.

5.1. Grid-based simulations

SH17 presented a practical implementation of METACALIBRATION⁹, and we use the default setup here. Although the image manipulations can be done on postage stamps, we instead process the full simulated images. This naturally allows us to quantify selection biases as described in SH17 and Sheldon et al. (2020). However, in this section we ignore the impact of selection bias.

We used the METACALIBRATION implementation in GALSIM to create the five images needed to compute the shear response for the grid-based simulated images. To do so, we have to choose

⁹ <https://github.com/esheldon/ngmix>

the value of $\Delta\gamma$ to use. Applying a larger shear has the benefit of increasing the precision with which the shear response can be measured, but if the value is too large, higher order terms may become relevant. This was explored in SH17 who found that for $\Delta\gamma < 0.04$ the changes are negligible. We therefore consider two values for $\Delta\gamma$, namely 0.02 and 0.04 and match the resulting shape measurements to the SEXTRACTOR catalogue.

As reported in Table 2 we observe a significant multiplicative bias for both shear components, which agree with each other. If METACALIBRATION yields an unbiased shear estimate, the measured multiplicative bias, μ_{meta} , however, should recover the SEXTRACTOR detection bias, μ_{det} . Indeed, we find that the bias that can be attributed to the shape measurements part is much smaller, with $\mu_1^{\text{meta}} - \mu_1^{\text{det}} = 0.00165 \pm 0.00046$ and $\mu_2^{\text{meta}} - \mu_2^{\text{det}} = -0.00086 \pm 0.00043$ for galaxies with $20 < m_{\text{AUTO}} < 24.5$, comparable to the requirements derived in Cropper et al. (2013).

To explore the performance of METACALIBRATION further, we show $\mu_{\text{meta}} - \mu_{\text{det}}$ as a function of m_{input} in Fig. 14. The use of the input magnitude ensures efficient shape noise cancellation. Comparison of the black ($\Delta\gamma = 0.04$) and open grey ($\Delta\gamma = 0.02$) points shows that the overall performance is similar, but that using a larger shear does indeed result in smaller uncertainties. Importantly, when we consider the two shear components separately, we find that they differ for $\Delta\gamma = 0.02$ (light coloured points) when $m_{\text{input}} > 23.5$, whereas $\Delta\gamma = 0.04$ (bright points) yields consistent values for μ_1 and μ_2 . Sampling may play a role here (see e.g., Kannawadi et al. 2021), but as the differences vanish when we apply the larger shear, we adopt this as our baseline.

5.2. Baseline results

We now proceed to use the setup with $\Delta\gamma = 0.04$ and $\sigma_w = 2$ pixels to examine the performance of METACALIBRATION on the simulations where galaxies are positioned randomly. Moreover, we explore the possibility to account for the selection bias using the procedure outlined in SH17. Although our use of a fixed weight function avoids introducing a weight bias¹⁰, the selection bias introduced by SEXTRACTOR remains.

SH17 show how the selection bias can be included in METACALIBRATION, by noting it introduces an ellipticity¹¹ dependent weighting, $S(\mathbf{e})$, of an underlying ellipticity distribution $P(\mathbf{e})$. Hence the ensemble averaged mean ellipticity can be expressed as

$$\langle \mathbf{e} \rangle^S = \int d\mathbf{e} S(\mathbf{e}) P(\mathbf{e}) \mathbf{e}, \quad (10)$$

where we assume that $\int d\mathbf{e} S(\mathbf{e}) P(\mathbf{e}) = 1$. We can express the ensemble averaged version of Eq. (5) as

$$\begin{aligned} \langle \mathbf{R} \rangle &= \int d\mathbf{e} \left. \frac{\partial [S(\mathbf{e}) P(\mathbf{e}) \mathbf{e}]}{\partial \gamma} \right|_{\gamma=0} \\ &= \int d\mathbf{e} \left[S(\mathbf{e}) \left. \frac{\partial [P(\mathbf{e}) \mathbf{e}]}{\partial \gamma} \right|_{\gamma=0} + P(\mathbf{e}) \mathbf{e} \left. \frac{\partial S(\mathbf{e})}{\partial \gamma} \right|_{\gamma=0} \right] \\ &\equiv \langle \mathbf{R}^\gamma \rangle + \langle \mathbf{R}^S \rangle. \end{aligned} \quad (11)$$

¹⁰ The size estimated from the best-fit Gaussian is different for the two image rotations after a shear has been applied. Using the observed size would thus couple the weight function to the shear itself, leading to a bias.

¹¹ We use ellipticity here as a synonym for shape, but note that the discussion is independent of the estimator employed.

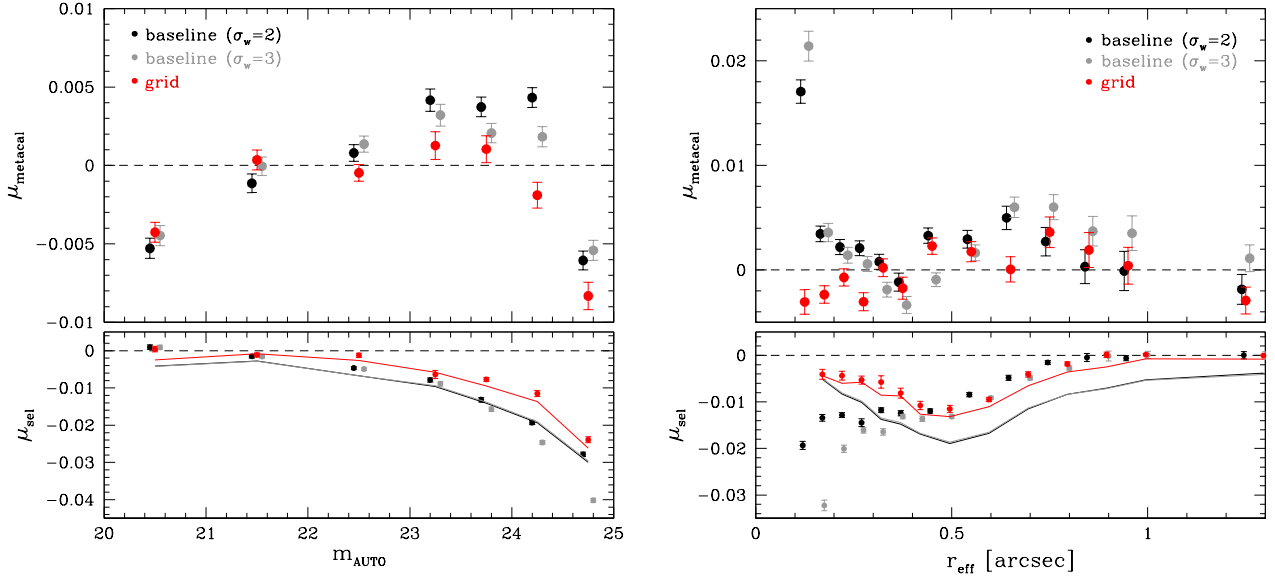


Fig. 15. *Left panel:* multiplicative bias after full METACALIBRATION as a function of m_{AUTO} for the baseline simulations (black for $\sigma_w = 2$ pixels; lightgrey for $\sigma_w = 3$ pixels) and the grid-based simulations (red points for $\sigma_w = 2$ pixels). *Right panel:* multiplicative bias after full METACALIBRATION as a function of the input half-light radius (r_{eff}) for galaxies with m_{AUTO} . *Bottom panels:* estimated selection bias from full METACALIBRATION (points). The solid lines show the corresponding direct measurements of the selection bias (cf. Figs. 4 and 5).

Table 3. Average biases after METACALIBRATION for galaxies with $20 < m_{\text{AUTO}} < 24.5$.

Setup	μ_1	μ_2	$c_1 [\times 10^5]$	$c_2 [\times 10^5]$
Baseline $\sigma_w = 2$ pixels				
METACALIBRATION (\mathbf{R}^γ only)	-0.00935 ± 0.00042	-0.00880 ± 0.00038	-0.73 ± 1.60	-1.24 ± 1.45
METACALIBRATION (full)	0.00274 ± 0.00042	0.00289 ± 0.00038	-0.74 ± 1.62	-1.26 ± 1.47
METACALIBRATION (full, $r_{\text{sep, in}} > 2''$)	0.00000 ± 0.00025	0.00095 ± 0.00026	-0.18 ± 0.96	0.93 ± 0.98
METACALIBRATION (full, $r_{\text{sep, det}} > 2''$)	-0.00694 ± 0.00031	-0.00601 ± 0.00030	-1.38 ± 1.18	0.74 ± 1.14
METADETECTION	-0.00081 ± 0.00040	0.00037 ± 0.00041	-2.90 ± 1.54	0.43 ± 1.57
Baseline $\sigma_w = 3$ pixels				
METACALIBRATION (\mathbf{R}^γ only)	-0.01259 ± 0.00042	-0.01223 ± 0.00037	-3.01 ± 1.60	0.14 ± 1.40
METACALIBRATION (full)	0.00166 ± 0.00042	0.00164 ± 0.00037	-3.05 ± 1.62	0.14 ± 1.42
METACALIBRATION (full, $r_{\text{sep}}^{\text{in}} > 2''$)	0.00021 ± 0.00027	0.00237 ± 0.00025	-0.96 ± 1.05	1.19 ± 0.96
METACALIBRATION (full, $r_{\text{sep}}^{\text{det}} > 2''$)	-0.00714 ± 0.00029	-0.00525 ± 0.00027	-2.16 ± 1.11	1.02 ± 1.03
Grid $\sigma_w = 2$ pixels				
METACALIBRATION (\mathbf{R}^γ only)	-0.00852 ± 0.00041	-0.00699 ± 0.00037	0.33 ± 1.57	1.31 ± 1.41
METACALIBRATION (full)	-0.00107 ± 0.00041	-0.00036 ± 0.00037	0.33 ± 1.58	1.31 ± 1.42

Notes. In the ‘baseline’ case the galaxies are placed randomly and their images are sheared. Galaxies are placed on a regular grid, about $9''$ apart for the ‘grid’ results. See text for details on the various selections.

If there is no selection bias, that is $S(\mathbf{e}) = 1$, the second term in Eq. (11) vanishes and we can identify the first term with \mathbf{R}^γ . The second term quantifies the response of the shear estimate to the selection bias. As discussed in SH17, \mathbf{R}^S can be estimated by measuring the mean ellipticity from the unsheared image, but selecting the measurements from the sheared images. The METACALIBRATION estimate of the selection bias is then

$$\mu_i^{\text{sel}} = \frac{\mathbf{R}_{ii}^\gamma + \mathbf{R}_{ii}^S}{\mathbf{R}_{ii}^\gamma}. \quad (12)$$

The left panel in Fig 15 shows the resulting multiplicative bias after METACALIBRATION when we account for the selection bias as a function of the observed apparent magnitude. The residuals for the grid-based results (red points) are very small, except for the galaxies with $m_{\text{AUTO}} > 24.5$. The bottom panel

shows the METACALIBRATION estimates for the selection bias, which agrees well with the actual bias that we infer from comparison to the input catalogue.

We report the mean biases for the two shear components in Table 3 for galaxies with $20 < m_{\text{AUTO}} < 24.5$. For the grid-based simulations we find $\langle \mu \rangle = 0.00064 \pm 0.00029$, well within requirements for Stage IV surveys. For the baseline case (black points) the results are similar, but we do observe a significant residual bias $\langle \mu \rangle = 0.00288 \pm 0.00029$, driven by galaxies with $m_{\text{AUTO}} > 23$. For reference we also repeated the measurements using a wider weight function with $\sigma_w = 3$ pixels, and we obtain similar results (see Table 3).

The right panel of Fig. 15 shows the post-METACALIBRATION bias as a function of the input galaxy size. For the grid based simulations (red points) the bias is flat

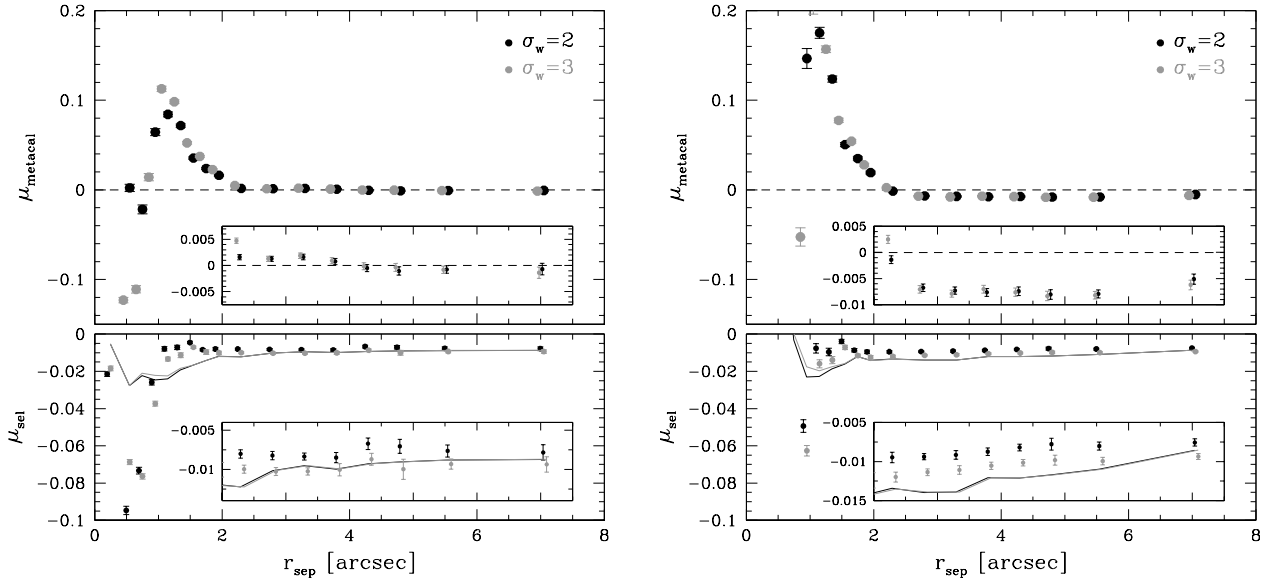


Fig. 16. *Left panel:* multiplicative bias after full METACALIBRATION for galaxies with $20 < m_{\text{AUTO}} < 24.5$ as a function of the separation to the nearest galaxy with $m_{\text{input}} < 26$ in the input catalogue (*top*) and selection bias (*bottom*) for a weight function with $\sigma_w = 2$ pixels (black) and $\sigma_w = 3$ pixels (grey). *Right panel:* idem, but now as a function of distance to the nearest detected galaxy. The insets in the panels zoom in on the results for separations larger than $2''$. The solid lines in the *bottom panels* show the corresponding direct measurements of the selection bias.

as a function of size, thus effectively correcting for the detection bias (shown in the bottom panel, as well as Fig. 5). The biases are also small for the baseline case, with both weight functions yielding consistent results. Only for the smallest galaxies do we observe a significant bias, which is not seen when galaxies are placed on a grid. This rules out sampling as the cause, but rather points to blending. Indeed, if we limit the comparison to isolated galaxies (no neighbour within $2''$ in the input catalogue with $m < 26$), the results are similar to the grid-based simulations.

This suggests that METACALIBRATION cannot fully account for the shear bias that is introduced by blending. Also the clear difference between the observed and inferred selection bias for the baseline case suggests that this is not correctly estimated (the agreement is much better for the grid simulations, shown in red). To explore this further we computed the post-METACALIBRATION bias as a function of separation to the nearest galaxy in the input catalogue ($m < 26$) and show the results in the left panel of Fig. 16.

Both choices for σ_w yield very similar results, except for very small values for r_{sep} where the larger weight function suffers more from blending, resulting in somewhat larger net biases. In both cases the bias rises quickly for separations $r_{\text{sep}} < 2''$ and becomes highly negative for $r_{\text{sep}} < 1''$, suggesting that it may be wise to exclude such galaxies from the cosmic shear analysis, if possible. As the bottom panel in Fig. 3 shows, this implies a 30% reduction in the galaxy number density, so that one may want to allow for larger residual biases, although the gain may still be limited because undetected blends also tend to increase the shape noise (Dawson et al. 2016).

The inset in the top panel shows that for $r_{\text{sep}} > 2''$ the bias is small: we find a mean bias $\langle \mu \rangle = 0.00047 \pm 0.00019$, whereas the bias for the full sample is $\langle \mu \rangle = 0.00288 \pm 0.00029$ (see Table 3 for more results). This confirms that METACALIBRATION can provide (nearly) unbiased shear estimates for isolated galaxies. Unfortunately, in practice we do not know whether or not a galaxy is blended, and the right panel of Fig. 16 shows the results for a more realistic scenario.

The bias as a function of distance to the nearest detected galaxy shows a similar dependence for small separations as in the left panel, but the biases peak at larger values. Both weight functions yield consistent biases, even though the estimated selection biases differ (bottom panel). More importantly, for $r_{\text{sep}} > 2''$ the bias no longer vanishes. Many of the blends are not identified as such, resulting in a bias of $\langle \mu \rangle = -0.00649 \pm 0.00022$ for apparently isolated galaxies. This is maybe not too surprising, because the SEXTRACTOR detection bias for apparently isolated galaxies (open grey points in Fig. 3) did not converge to the value when galaxies are placed on a grid.

As this is perhaps the cleanest sample of sources that could be identified in a survey, our results imply that an algorithm that can provide unbiased shear estimates under ideal circumstances will still be significantly biased in reality. This also has implications for machine learning approaches (Gruen et al. 2010; Tewes et al. 2019; Pujol et al. 2020), which will have to be trained on simulations that include realistic blending.

Our findings suggest that, while METACALIBRATION is able to account for selection bias for isolated galaxies, blending limits the performance in more realistic scenarios. The image simulations can, however, be used to account for these residual biases, provided the simulations capture the complexities of real data. We therefore explore the sensitivity of the post-METACALIBRATION bias to changes in the simulation inputs, similar to what we did for the SEXTRACTOR detection bias.

The red points in Fig. 10 show that the sensitivity to the galaxy density, captured by n_{fac} , has changed sign compared to the SEXTRACTOR detection bias, but the amplitude of the trend is similar with $\partial \mu_{\text{meta}} / \partial n_{\text{fac}} = 0.00375 \pm 0.00033$, suggesting that it remains important to use the correct galaxy density in the simulations. The changes in multiplicative bias as a function of f_{size} and ϵ_{fac} are shown as red points in Fig. 11. Indeed, we find that the sensitivities to these morphological parameters are reduced significantly compared to the SEXTRACTOR detection bias, with $\partial \mu_{\text{meta}} / \partial f_{\text{size}} = 0.0087 \pm 0.0022$ and $\partial \mu_{\text{meta}} / \partial \epsilon_{\text{fac}} = 0.0005 \pm 0.0027$. Similarly we find no clear change in bias if we replace the Sérsic index by a single value (red points in Fig. 12).

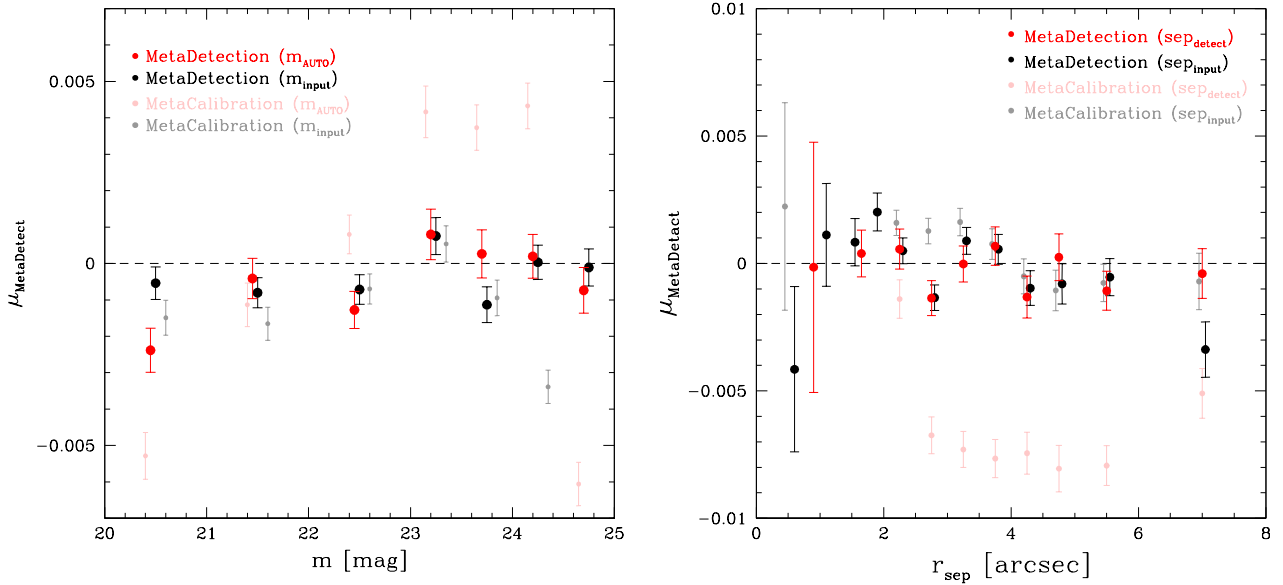


Fig. 17. *Left panel:* multiplicative bias after METADETECTION as a function of magnitude, with galaxies selected by the input magnitude (black) or the observed magnitude (red). *Right panel:* multiplicative bias for galaxies with $20 < m_{\text{AUTO}} < 24.5$ as a function of r_{sep} , the distance to the nearest neighbour in the input catalogue (black points) and the distance to the nearest detected galaxy (red points). The light coloured points indicate the corresponding results for METACALIBRATION. In the case of METADETECTION the biases show no trend with magnitude or distance to the nearest galaxy, and are consistent with zero.

The sensitivity to the truncation of the surface brightness profile is enhanced, as indicated by the red points in Fig. 13, but the bias converges for $r_{\text{trunc}} > 4$.

6. MetaDetection

The results presented in Table 3 and Fig. 16 show that undetected blending is a significant source of bias, even for space-based Stage IV surveys. High quality, deep observations can help improve the fidelity of the image simulations that are used to quantify this residual bias, and our results indicate that the sensitivity to the simulation inputs are relatively small, but it would be better if this could be avoided in principle.

Sheldon et al. (2020) proposed an alternative implementation of the METACALIBRATION approach where one effectively bypasses the steps to estimate \mathbf{R}^y and \mathbf{R}^S . This approach, called METADETECTION, uses the same sheared images, but both the detection and the shape analysis are performed on these images. By avoiding the use of the unshaped image as a reference, the detection biases should vanish. The downside, however, is the lack of such a reference image, which complicates the labelling of galaxies that is needed to associate them with a tomographic redshift bin.

We applied METADETECTION to our simulated images and found that the resulting average bias for galaxies with $20 < m_{\text{AUTO}} < 24.5$ is very small: $\langle \mu \rangle = 0.00001 \pm 0.00030$ (we report the results for the individual shear components in Table 3). We note that we have not quantified how this result changes when we shear the scene when creating the images (see Sect. 3.1), but the results presented in Sheldon et al. (2020) suggest that this difference should be small for the much smaller *Euclid* PSF.

The left panel in Fig. 17 shows the bias as a function of observed magnitude (m_{AUTO} ; red points) and input magnitude (m_{input} ; black points). The average biases are small and do not depend on magnitude, even for galaxies as faint as $m = 25$. For reference, we indicate the corresponding METACALIBRATION

results by the light coloured points. This is encouraging, because one could imagine estimating photometric redshifts for the galaxies in each of the five METADETECTION catalogues, which could subsequently be used to assign them to tomographic bins. How to incorporate this into a full cosmic shear analysis is beyond the scope of this paper, but it is clearly worthwhile to explore further.

The potential of METADETECTION is confirmed further by the right panel of Fig. 17, where we show the bias as a function of distance to the nearest neighbour in the input catalogue (black) and the nearest detected neighbour (red). The improvement with respect to the METACALIBRATION case, indicated by the light red coloured points, is evident: METADETECTION is able to account for the blending of galaxies, resulting in residual biases that meet the stringent requirements for Stage IV surveys (Cropper et al. 2013). Moreover, the blue points in Figs. 10 and 13 show that the bias after METADETECTION no longer depends on the galaxy density n_{fac} or the truncation radius r_{trunc} .

7. Discussion

Our results show that the detection of galaxies results in a significant source of bias for weak lensing surveys (also see Fenech Conti et al. 2017; Kannawadi et al. 2019; Hernández-Martín et al. 2020). Although both survey characteristics and galaxy morphologies play a role, it is clear that undetected blending is the main concern for Stage IV surveys. In particular, we used METACALIBRATION as a proxy for a perfect shape measurement algorithm, and showed that this problem persists also in this case. Nonetheless, the reduced sensitivity to the simulation setup indicates that image simulations can provide accurate estimates of residual biases. Such simulations may be needed regardless, because METACALIBRATION cannot account for all sources of bias (Huff & Mandelbaum 2017).

As discussed in Huff & Mandelbaum (2017), the image manipulation step assumes that the image is linearly related to the true surface brightness distribution. A wide range of

instrumental effects limit the accuracy of this assumption. Some of these can be partially corrected for during the image processing, but the impact of their residuals should also be assessed using sufficiently realistic image simulations. A particular concern for *Euclid* and the *Roman* Space Telescope is the fact that the pixel scale is relatively large compared to the PSF. This is not a problem per se for the PSF itself, as a well-sampled model might be inferred from the data, but galaxies with small observed sizes might still be affected. However, [Kannawadi et al. \(2021\)](#), with a larger fraction of small galaxies in their input catalogue, show that the bias due to undersampling is effectively mitigated when using a weight function with $\sigma_w \geq 0''.15$. This is consistent with the absence of any significant residual biases in this work, and therefore, undersampling of small galaxies need not be a major concern for *Euclid*.

In the case of *Euclid*, charge-transfer inefficiency and the presence of cosmic rays also bias the shape measurements. Also blending and contamination by stars affects the shear bias ([Hoekstra et al. 2017](#)), whereas spatial variations in the colours of galaxies lead to colour-gradient biases (e.g., [Semboloni et al. 2013](#); [Er et al. 2018](#)). These biases are also present after METADETECTION, which does provide unbiased shear estimates for our simulated *Euclid*-like simulations.

Our results suggest that METACALIBRATION and METADETECTION, combined with sufficiently realistic image simulations, provide viable ways forward towards accurate shear estimates for Stage IV surveys. Many practical complications remain, and we briefly review some of these here. We start by examining the computational needs: *Euclid* aims to measure the shapes of over two billion galaxies, which places constraints on the time it takes to measure a galaxy shape. We apply the METACALIBRATION-step to the full images, and run the object detection algorithm on the METACALIBRATION images, using the output for the unsheared image as our new detection catalogue. The computational needs are driven by the image manipulation steps, which take about 150 s for each 4000×4000 pixel image on a single core of an Intel Xeon Gold 5115 2.4 GHz CPU in a Dell R840 server (equipped with 80 cores). This includes some I/O because we save the five images to disk for SExtractor, which in principle can be avoided. The five SExtractor calls take on average 16 s and the shape measurements themselves take a total of 7 s for the baseline case. This amounts to a total processing time of about 0.06 s per galaxy on a single core. In this paper, we use METACALIBRATION to correct for the convolution with an isotropic PSF, but it can be extended to correct for PSF anisotropy ([SH17](#)). This requires four more images to be created, which approximately doubles the runtime of the METACALIBRATION-step.

The memory needs of our current setup are substantial when creating the METACALIBRATION images, requiring about 18 Gb per core. This prevented us from using all available cores. In practice the analysis will have to be performed on much smaller postage stamps, because the PSF will vary across the field-of-view. In fact, in the case of *Euclid*, the PSF P to use is the SED-weighted one. Although the PSF in this case varies from object to object, it can be uniquely estimated from unresolved multi-band data ([Eriksen & Hoekstra 2018](#)). As it is important that the effects of blending can be captured, the postage stamp should be at least $8'' \times 8''$. This estimate is based on the fact that the METACALIBRATION bias converges for $r_{\text{sep}} > 3''$ (see the right panel of [Fig. 16](#)). [Eriksen & Hoekstra \(2018\)](#) show that the effective *Euclid*-PSF size varies by at most

about 2%, which suggests that using a single PSF for such a postage stamp would still capture the bias caused by blending.

For our galaxy number density, a postage stamp of $8'' \times 8''$ means that the total number of pixels that needs to be manipulated increases by about 30%. Given the reduced memory needs this would actually allow more cores to be used by a typical server. Including the correction for PSF anisotropy, we thus estimate that analysing 2×10^9 galaxies would take about 70 days on our benchmark server with 80 cores. We note that this is a bare minimum, because one may want to analyse the individual exposures instead. Nonetheless, these estimates suggest that it may be possible to apply METACALIBRATION to Stage IV data sets. Alternatively, METACALIBRATION or METADETECTION can be applied to subsets of data to provide bias estimates for machine learning algorithms. Once trained these can estimate shapes very quickly (e.g., [Pujol et al. 2020](#)).

Finally we note that METACALIBRATION and METADETECTION allow us to obtain unbiased shear estimates, but intrinsic alignments of galaxies prevent a straightforward interpretation of the lensing signal (see, e.g., [Joachimi et al. 2015](#), for a review). Direct observational constraints on the intrinsic alignment signal rely on accurate ellipticity measurements ([Georgiou et al. 2019a](#)). Moreover, the strength of the alignment signal depends on the shape measurement itself (e.g., [Georgiou et al. 2019b](#)). Hence care has to be taken when using physically motivated priors for the intrinsic alignment signal ([Johnston et al. 2019](#); [Fortuna et al. 2021](#)) in a cosmological analysis when the shear estimates are based on an intrinsically biased shape estimator, like the one we adopted here.

8. Conclusions

Accurate measurements of the shapes of galaxies are a key ingredient for weak gravitational lensing studies. As a consequence improving the fidelity of the shape measurement algorithms has received much attention. Application of these algorithms to simulated data have played an important role in improving the performance. It has also become clear that it is important that the simulated data resemble the observations closely (see e.g., the discussion in [Kannawadi et al. 2019](#)). [H17](#) presented a detailed study for a simulated *Euclid*-like data set, highlighting the challenges in ensuring sufficient realism.

In this paper, we use *Euclid*-like image simulations, similar to the ones studied in [H17](#) to examine another important source of bias, which is present even if the shapes estimates are perfect. Detection bias arises because the probability with which an object is detected (or selected) in an image depends on the shear. This has been known for quite a while (e.g., [Hirata & Seljak 2003](#)), but its contribution to shear bias has been largely ignored until recently. We find that the bias is generally smaller than instrumental bias, but it does lead to multiplicative biases in the shear that exceed requirements for the next generation of cosmic shear surveys.

To quantify the size of the bias, we used SExtractor ([Bertin & Arnouts 1996](#)) to detect objects. We matched the resulting catalogues to the input catalogue from which we took the true ellipticities. This mimics the performance of an ideal shape measurement algorithm. As reported in [Table 2](#) we found that the average shear is underestimated by about 1%; five times larger than can be tolerated for Stage IV surveys ([Cropper et al. 2013](#)). This result is robust against changes in the settings of the detection algorithm. A smaller detection bias, which only affects the faintest galaxies, is observed when we place galaxies on a grid. This is caused because galaxies that are oriented

perpendicular to the shear are detected preferentially. In the case of an anisotropic PSF, we found a small positive additive detection bias because galaxies that align with the PSF are detected with a higher significance.

The larger bias in our baseline simulation, where galaxies were placed randomly, is caused by the blending of sources, with biases exceeding 2% for separations less than 1". Deeper observations can reduce the detection bias, but blending introduces a floor that still exceeds requirements. Following H17, we also explored how the detection bias depends on the simulation inputs. We found that the detection bias increases linearly with galaxy density, the result of the higher occurrence of blending. The bias is also reduced when the galaxies are smaller or rounder. We observe a slight dependence of the surface brightness profile (quantified by the Sérsic-index n). It is, however, important that the galaxy profiles are not artificially truncated before four effective radii.

Although the detection bias is far less sensitive to variations in the simulation parameters compared to the KSB algorithm studied in H17, the realism of the simulations, in particular ensuring that the variety in galaxy morphologies is adequately captured, remains a concern. We therefore explored the performance of an alternative approach that uses the data to determine the response of an ensemble of galaxies to a shear. This so-called METACALIBRATION was recently developed by Huff & Mandelbaum (2017) and SH17 and showed promise for isolated galaxies. The problem of blending was investigated in more detail by Sheldon et al. (2020) who showed that a variation of METACALIBRATION, dubbed METADETECTION, can be used to address this.

We found that METACALIBRATION provides a (near) perfect shear estimate in the absence of detection bias. Importantly, the choice of shape measurement algorithm is irrelevant and we opted for weighted quadrupole moments with a fixed width for the Gaussian weight function. For isolated galaxies the performance of METACALIBRATION is only limited by the accuracy of the PSF model (which we assume to be perfect) and biases introduced by the pixelisation of the images (which are also negligible in our case). For the grid-based images we obtained a mean multiplicative bias of $\langle\mu\rangle = 0.00064 \pm 0.00029$, well within requirements for Stage IV surveys. However, blending will limit the actual performance and for our baseline case we measured a significant bias of $\langle\mu\rangle = 0.00288 \pm 0.00029$. We showed that this is caused by blended objects, many of which cannot be identified as such. In fact selecting galaxies that appear isolated (no detected neighbour within 2") leads to a larger net bias of $\langle\mu\rangle = -0.00649 \pm 0.00022$. The post-METACALIBRATION bias is less sensitive to changes in the input galaxy sizes or the ellipticity distribution, but does still depend on the galaxy number density. Nonetheless, these findings suggest that image simulations can be used to account for the residual biases in METACALIBRATION. Such simulations are needed anyway to determine the biases caused by instrumental effects. Moreover, simulations will be essential to understand the correlation between shear bias and biases in photometric redshifts that blending should introduce.

METADETECTION uses the same sheared images as METACALIBRATION, but both the detection and shape analysis are performed on these images. The resulting multiplicative bias for galaxies with $20 < m_{\text{AUTO}} < 24.5$ is very small: $\langle\mu\rangle = 0.00001 \pm 0.00030$. Moreover, the bias does not depend on magnitude or distance to the nearest neighbour, indicating the blending does not bias the mean shear. The lack of a reference catalogue, which otherwise would re-introduce the selection bias, may lead to practical complications. However, it

may be possible to assign photometric redshifts to the different METADETECTION catalogues and define tomographic redshifts for each catalogue. Alternatively, the METADETECTION estimates for various selections of source can act as reference values for machine learning approaches. More work is needed to examine the practical implementation of both METACALIBRATION and METADETECTION, but our results suggest that these, combined with sufficiently realistic image simulations, provide a viable way forward towards accurate shear estimates for Stage IV surveys.

Acknowledgements. The authors are grateful to Erin Sheldon and the developers of GALSIM for making their software packages publicly available. They also thank Tim Schrabbach and an anonymous referee for useful comments that helped improve the paper. HH and AK acknowledge support from the Netherlands Organisation for Scientific Research (NWO) through grant 639.043.512. HH and TDK acknowledge support from the EU Horizon 2020 research and innovation programme under grant agreement 776247.

References

- Albrecht, A., Bernstein, G., Cahn, R., et al. 2006, ArXiv e-prints [arXiv:astro-ph/0609591]
- Amendola, L., Appleby, S., Avgoustidis, A., et al. 2018, *Liv. Rev. Relativ.*, 21, 2
- Bernstein, G. M., & Jarvis, M. 2002, *AJ*, 123, 583
- Bertin, E., & Arnouts, S. 1996, *A&AS*, 117, 393
- Bridle, S., Balan, S. T., Bethge, M., et al. 2010, *MNRAS*, 405, 2044
- Carretero, J., Castander, F. J., Gaztañaga, E., Crocce, M., & Fosalba, P. 2015, *MNRAS*, 447, 646
- Crocce, M., Castander, F. J., Gaztañaga, E., Fosalba, P., & Carretero, J. 2015, *MNRAS*, 453, 1513
- Cropper, M., Hoekstra, H., Kitching, T., et al. 2013, *MNRAS*, 431, 3103
- Cropper, M., Pottinger, S., Azzollini, R., et al. 2018, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Proc. SPIE, 10698, 1069828
- Dawson, W. A., Schneider, M. D., Tyson, J. A., & Jee, M. J. 2016, *ApJ*, 816, 11
- Er, X., Hoekstra, H., Schrabbach, T., et al. 2018, *MNRAS*, 476, 5645
- Erben, T., Van Waerbeke, L., Bertin, E., Mellier, Y., & Schneider, P. 2001, *A&A*, 366, 717
- Eriksen, M., & Hoekstra, H. 2018, *MNRAS*, 477, 3433
- Euclid Collaboration (Martinet, N., et al.) 2019, *A&A*, 627, A59
- Fenech Conti, I., Herbonnet, R., Hoekstra, H., et al. 2017, *MNRAS*, 467, 1627
- Fortuna, M. C., Hoekstra, H., Joachimi, B., et al. 2021, *MNRAS*, 501, 2983
- Fosalba, P., Crocce, M., Gaztañaga, E., & Castander, F. J. 2015, *MNRAS*, 448, 2987
- Georgiou, C., Johnston, H., Hoekstra, H., et al. 2019a, *A&A*, 622, A90
- Georgiou, C., Chisari, N. E., Fortuna, M. C., et al. 2019b, *A&A*, 628, A31
- Gruen, D., Seitz, S., Koppenhoefer, J., & Riffeser, A. 2010, *ApJ*, 720, 639
- Hamana, T., Shirasaki, M., Miyazaki, S., et al. 2020, *PASJ*, 72, 16
- Hartlap, J., Hiltbert, S., Schneider, P., & Hildebrandt, H. 2011, *A&A*, 528, A51
- Herbonnet, R., Sifón, C., Hoekstra, H., et al. 2020, *MNRAS*, 497, 4684
- Hernández-Martín, B., Schrabbach, T., Hoekstra, H., et al. 2020, *A&A*, 640, A117
- Heymans, C., Van Waerbeke, L., Bacon, D., et al. 2006, *MNRAS*, 368, 1323
- Hildebrandt, H., Köhlinger, F., van den Busch, J. L., et al. 2020, *A&A*, 633, A69
- Hirata, C., & Seljak, U. 2003, *MNRAS*, 343, 459
- Hoekstra, H., Franx, M., Kuijken, K., & Squires, G. 1998, *ApJ*, 504, 636
- Hoekstra, H., Yee, H. K. C., Gladders, M. D., et al. 2002, *ApJ*, 572, 55
- Hoekstra, H., Donahue, M., Conselice, C. J., McNamara, B. R., & Voit, G. M. 2011, *ApJ*, 726, 48
- Hoekstra, H., Herbonnet, R., Muzzin, A., et al. 2015, *MNRAS*, 449, 685
- Hoekstra, H., Viola, M., & Herbonnet, R. 2017, *MNRAS*, 468, 3295
- Huff, E., & Mandelbaum, R. 2017, ArXiv e-prints [arXiv:1702.02600]
- Joachimi, B., Cacciato, M., Kitching, T. D., et al. 2015, *Space Sci. Rev.*, 193, 1
- Johnston, H., Georgiou, C., Joachimi, B., et al. 2019, *A&A*, 624, A30
- Joudaki, S., Hildebrandt, H., Traykova, D., et al. 2020, *A&A*, 638, L1
- Kaiser, N. 2000, *ApJ*, 537, 555
- Kaiser, N., Squires, G., & Broadhurst, T. 1995, *ApJ*, 449, 460
- Kannawadi, A., Mandelbaum, R., & Lackner, C. 2015, *MNRAS*, 449, 3597
- Kannawadi, A., Hoekstra, H., Miller, L., et al. 2019, *A&A*, 624, A92
- Kannawadi, A., Rosenberg, E., Hoekstra, H., et al. 2021, *MNRAS*, in press, [arXiv:2010.04164]
- Kilbinger, M. 2015, *Rep. Progr. Phys.*, 78, 086901
- Kitching, T. D., Balan, S. T., Bridle, S., et al. 2012, *MNRAS*, 423, 3163

- Kitching, T. D., Paykari, P., Hoekstra, H., & Cropper, M. 2019, [Open J. Astrophys.](#), **2**, 5
- Kregel, M., van der Kruit, P. C., & de Grijs, R. 2002, [MNRAS](#), **334**, 646
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, ArXiv e-prints [arXiv:1110.3193]
- LSST Science Collaboration (Abell, P. A., et al.) 2009, ArXiv e-prints [arXiv:0912.0201]
- Mandelbaum, R. 2018, [ARA&A](#), **56**, 393
- Mandelbaum, R., Rowe, B., Bosch, J., et al. 2014, [ApJS](#), **212**, 5
- Mandelbaum, R., Rowe, B., Armstrong, R., et al. 2015, [MNRAS](#), **450**, 2963
- Massey, R., Heymans, C., Bergé, J., et al. 2007, [MNRAS](#), **376**, 13
- Massey, R., Hoekstra, H., Kitching, T., et al. 2013, [MNRAS](#), **429**, 661
- Melchior, P., & Viola, M. 2012, [MNRAS](#), **424**, 2757
- Miller, L., Heymans, C., Kitching, T. D., et al. 2013, [MNRAS](#), **429**, 2858
- Paulin-Henriksson, S., Amara, A., Voigt, L., Refregier, A., & Bridle, S. L. 2008, [A&A](#), **484**, 67
- Planck Collaboration XIII. 2016, [A&A](#), **594**, A13
- Pujol, A., Kilbinger, M., Sureau, F., & Bobin, J. 2019, [A&A](#), **621**, A2
- Pujol, A., Bobin, J., Sureau, F., Guinot, A., & Kilbinger, M. 2020, [A&A](#), **643**, A158
- Refregier, A., Kacprzak, T., Amara, A., Bridle, S., & Rowe, B. 2012, [MNRAS](#), **425**, 1951
- Riess, A. G., Casertano, S., Yuan, W., Macri, L. M., & Scolnic, D. 2019, [ApJ](#), **876**, 85
- Rix, H.-W., Barden, M., Beckwith, S. V. W., et al. 2004, [ApJS](#), **152**, 163
- Rowe, B. T. P., Jarvis, M., Mandelbaum, R., et al. 2015, [Astron. Comput.](#), **10**, 121
- Seitz, C., & Schneider, P. 1997, [A&A](#), **318**, 687
- Semboloni, E., Hoekstra, H., Huang, Z., et al. 2013, [MNRAS](#), **432**, 2385
- Sheldon, E. S., & Huff, E. M. 2017, [ApJ](#), **841**, 24
- Sheldon, E. S., Becker, M. R., MacCrann, N., & Jarvis, M. 2020, [ApJ](#), **902**, 138
- Spergel, D., Gehrels, N., Baltay, C., et al. 2015, ArXiv e-prints [arXiv:1503.03757]
- Tal, T., & van Dokkum, P. G. 2011, [ApJ](#), **731**, 89
- Tewes, M., Kuntzer, T., Nakajima, R., et al. 2019, [A&A](#), **621**, A36
- Toyozumi, H., & Ashley, M. C. B. 2005, [PASA](#), **22**, 257
- Troxel, M. A., & Ishak, M. 2015, [Phys. Rep.](#), **558**, 1
- Troxel, M. A., MacCrann, N., Zuntz, J., et al. 2018, [Phys. Rev. D](#), **98**, 043528
- Viola, M., Kitching, T. D., & Joachimi, B. 2014, [MNRAS](#), **439**, 1909

Appendix A: Sensitivity to detection setup

Several parameters influence the deblending of objects by SEXTRACTOR, and we examine their impact on the detection bias here. Compared to the choice of filter function in the detection step (see Fig. 7), the changes in bias are smaller, but as Fig. A.1 shows, they can still change by as much as 10^{-3} in the most extreme cases. The changes are negligible, however, when the parameters remain close to their default values.

As described in detail in Bertin & Arnouts (1996) SEXTRACTOR uses multi-thresholding to separate objects that were extracted as a single object during the detection step. The pixels that make up an extracted object are thresholded by DEBLEND_NTHRESH levels that are spaced exponentially between the extraction threshold and the peak value; a low value reduces the effectiveness of the deblending step. A tree model of the surface brightness is created (see Fig. 2 in Bertin & Arnouts 1996) and the model works its way down to the trunk, deciding at each junction whether or not to split the object into separate ones. This decision is governed by the value of DEBLEND_MINCONT, which is the minimum fraction of the flux that needs to be

contained in the deblended source; hence a high value of this parameter means that only sources of similar brightness are deblended.

The left and middle panels in Fig. A.1 show the change in μ_{det} when we use different values for DEBLEND_MINCONT and DEBLEND_NTHRESH, respectively. We see that the bias increases by about 10^{-3} if the deblending is minimised. The biases barely change if we vary the parameters about the baseline settings (indicated by the vertical grey dashed lines).

Noise in the images may result in the outer regions of sources to be broken up into smaller pieces. Such inadvertent ‘deblending’ is undone by cleaning the catalogue. For each object, SEXTRACTOR estimates the contribution from neighbouring galaxies to the mean surface brightness assuming a Gaussian extrapolation of their profile, and subtracts this from the object in question. If it is still above the detection threshold, the object is accepted. The width of Gaussian used to extrapolate the flux from nearby galaxies can be changed from its default estimate by CLEAN_PARAM. The right panel in Fig. A.1 shows that that little cleaning (values less than 1) rapidly increases the detection bias, whereas more aggressive cleaning has little impact.

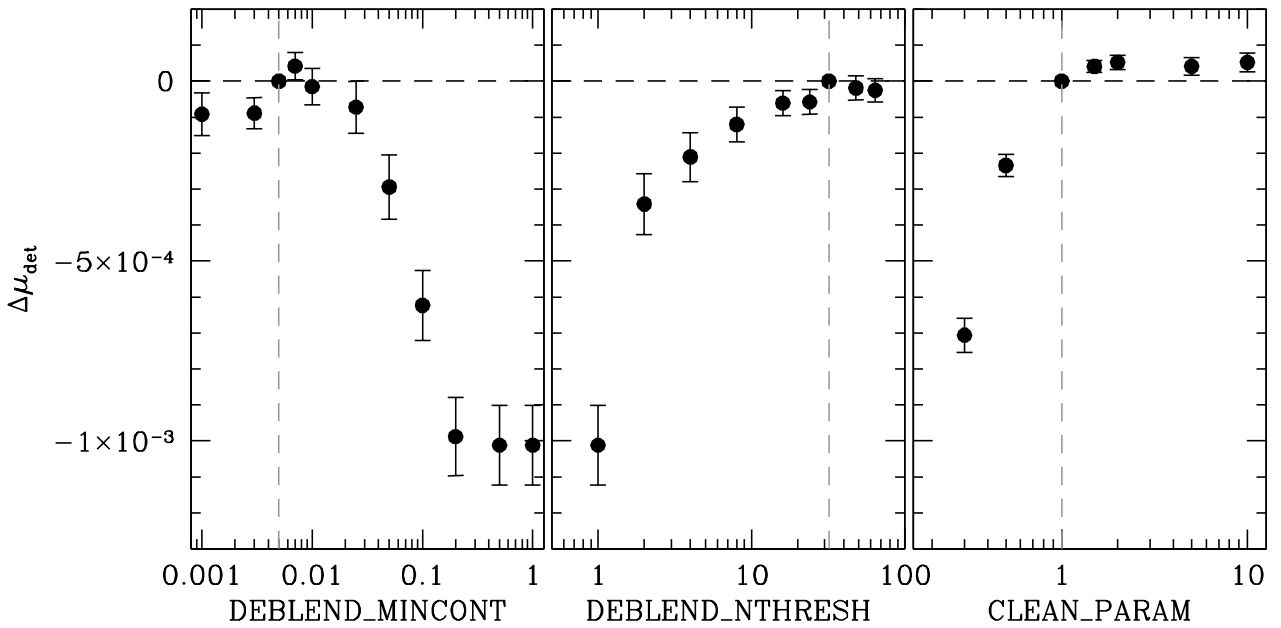


Fig. A.1. Change in multiplicative shear bias $\Delta\mu$ as a function of the SEXTRACTOR parameters that affect the deblending of objects. The vertical grey dashed line indicates the baseline value (also see Table 1). These default values result in detection biases that are close to optimal.

Appendix B: Relation between additive and multiplicative bias

If one is concerned about a particular instrumental effect that might introduce additive bias, one can simply average the shear estimates in the appropriate coordinate system (e.g., the one defined by the detector), because the cosmological signal should vanish if enough data are included. For instance, [Hoekstra et al. \(2011\)](#) used this approach to remove the additive bias caused by charge transfer inefficiency (CTI) in *Hubble* Space Telescope observations, and [Hildebrandt et al. \(2020\)](#) use this to account for an additive bias that arises from the shape measurement algorithm (as shown in [Kannawadi et al. 2019](#)). However, this empirical approach ignores the fact that such systematics may cause multiplicative bias as well, as we show here.

To do so we express the observed shape of an object in terms of the unweighted quadrupole moments Q_{ij} of its surface brightness distribution (e.g., [Massey et al. 2013](#)). These can be combined into the polarisation, which has two components χ_i defined as:

$$\chi_1 = \frac{Q_{11} - Q_{22}}{Q_{11} + Q_{22}}, \text{ and } \chi_2 = \frac{2Q_{12}}{Q_{11} + Q_{22}}. \quad (\text{B.1})$$

If we now consider a (residual) effect that changes the observed quadrupole $Q'_{11} = Q_{11} + \delta Q_{11}$, while leaving the other moments unchanged¹², the observed polarisation is:

$$\langle \chi_1^{\text{obs}} \rangle \approx \langle \chi_1^{\text{true}} \rangle \left(1 - \frac{\delta Q_{11}}{Q_{11} + Q_{22}} \right) + \frac{\delta Q_{11}}{Q_{11} + Q_{22}}, \quad (\text{B.2})$$

and

$$\langle \chi_2^{\text{obs}} \rangle \approx \langle \chi_2^{\text{true}} \rangle \left(1 - \frac{\delta Q_{11}}{Q_{11} + Q_{22}} \right). \quad (\text{B.3})$$

The last term in Eq. (B.2) corresponds to the additive bias c_1 , whereas *both* polarisation components are biased low by a factor $(1 + \mu)$. In this simple case we find that $\mu_1 = \mu_2 = \mu = -c_1$. Hence, instrumental effects that introduce additive shear bias by modifying the recorded images, generally also cause a multiplicative bias that is similar in amplitude, affects both shear components, but has the opposite sign.

To verify this result, we created images where we mimic the effect of charge trailing, which in reality might be caused by dielectric absorption in the read-out electronics ([Toyozumi & Ashley 2005](#)). Rather than computing the actual change in bias voltage, we simply assumed that the amount of charge that is added to the next pixel in the i th column is given by a power law, so that

$$f(i, j) = f(i, j) + f_{\text{trail}} f(i-1, j)^{0.4},$$

where the value for the power law slope is inspired by what is observed in OmegaCAM data ([Hoekstra et al., in prep.](#)), and f_{trail} is the amplitude. To create the images we added a realistic background level, computed the trailed image, added this to the original image and finally subtracted the background again. We analysed the resulting images as before.

Figure B.1 shows the resulting additive and multiplicative bias as a function of f_{trail} , where we note that the applied values are unrealistically large to ensure a signal that could be measured using METACALIBRATION. The light coloured points show the detection biases that we observe, which are negligible, with

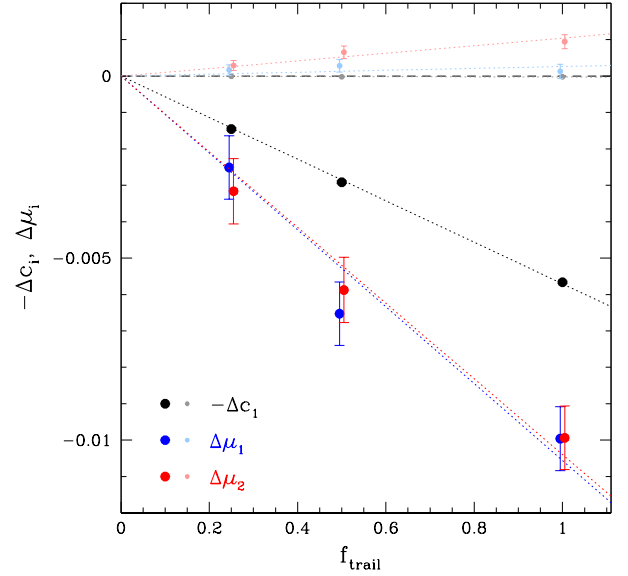


Fig. B.1. Comparison of the change in multiplicative bias and additive bias when some of the charge is trailed during the readout process. The amount of trailing is determined by the value of f_{trail} . The light coloured points show the (small) detection biases, whereas the bright coloured points show the $-\Delta c_1$ (black), $\Delta \mu_1$ (blue) and $\Delta \mu_2$ (red). The amplitude of the additive bias is about half of the multiplicative bias, but has the opposite sign, as predicted.

the exception of μ_2^{det} . The trend of μ_2^{det} with f_{trail} is largest if we consider only isolated galaxies. The trailing changes both the centroid and the flux of the galaxy, both of which will change the multiplicative bias somewhat, but it is not obvious why this does not affect μ_1^{det} . Nonetheless the detection biases are small, even for this extreme level of charge trailing.

More relevant for the discussion here are the bright coloured points, which show the biases after METACALIBRATION. As predicted, the changes in μ_1 (blue) and μ_2 (red) are consistent, and the sign of Δc_1 is opposite from $\Delta \mu$. The change is about half of what is predicted by Eq. (B.2), but we note that it is no longer applicable when weighted moments are used to measure the shapes, and corrections of $\mathcal{O}(1)$ are expected. In fact, tests with an elliptical Airy PSF, for which the unweighted quadrupole moments do not converge, show that the additive PSF biases are smaller by as much as a factor 4. This lower sensitivity to PSF anisotropy implies that allocating a residual bias of $|c_{\text{PSF}}| < 1.5 \times 10^{-4}$ corresponds to a tolerable error in the PSF ellipticity of $|\Delta \epsilon_{\text{PSF}}| < 5.8 \times 10^{-4}$ instead of $< 2 \times 10^{-4}$ adopted by [Cropper et al. \(2013\)](#), but we caution that the sensitivity depends on the PSF profile. For instance when we used PSF models that included various aberrations, we typically found larger residual biases. In all cases, however, the sensitivity was lower compared to the estimate based on unweighted moments.

An empirical correction for additive bias should therefore be considered with these limitations in mind. Although our results for this particular case show that the multiplicative bias is still within a factor 2 of our naive prediction, it cannot replace a proper physical modelling and calibration of instrumental effects. One important reason to understand the cause of any residual additive bias is that it may not be cleanly separable from other biases; this would complicate estimating the impact on the multiplicative bias. Nonetheless, provided that the biases are small to begin with, we expect that the multiplicative biases will be similar in amplitude to the additive bias. This is still

¹² For the purpose of this derivation we are free to choose a convenient coordinate system.

helpful, because [Kitching et al. \(2019\)](#) showed that the impact of such small scale-dependent multiplicative biases is reduced further when we consider the power spectrum estimates used in cosmological analyses.

Appendix C: Uncertainty in input galaxy number density

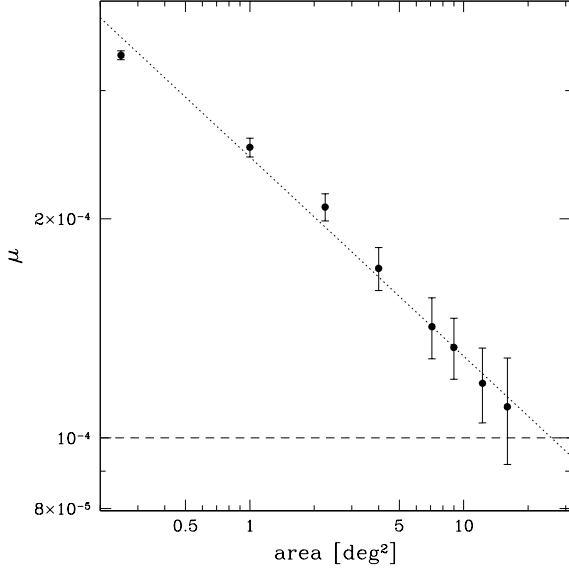


Fig. C.1. Multiplicative bias that arises from uncertainties in the average galaxy number density, as a function from the area used to determine the average density, assuming the sensitivity of μ with n_{fac} after METACALIBRATION (dashed red line in Fig. 10).

As shown in Fig. 10 the multiplicative bias depends on the number density of galaxies in the simulated images. Moreover, the results indicate that METACALIBRATION is unable to fully remove such a dependency, but in fact introduces a weak positive dependence as $\partial\mu_{\text{meta}}/\partial n_{\text{fac}} = (0.0039 \pm 0.0005)$. Here we examine what area needs to be observed so that the uncertainty in the observed value of n_{fac} leads to a bias in the multiplicative bias of $|\delta\mu| < 10^{-4}$.

To estimate the expected variation in galaxy density as a function of angular scale we used the second data release of the Marenstrum Institut de Ciències de l’Espai (MICE) grand challenge galaxy and halo light-cone simulation¹³. The mock galaxy

catalogue was obtained from a large N -body simulation, from which a light-cone was constructed (see [Fosalba et al. 2015](#), for details). The simulation was populated with galaxies using a hybrid halo occupation distribution and abundance matching technique described in [Crocco et al. \(2015\)](#) and [Carretero et al. \(2015\)](#).

The second data release includes a mock galaxy catalogue that is complete for current Stage III surveys ($m_i < 24$), but restricted to $z < 1.4$, resulting in an average number density of about 26 galaxies arcmin⁻² brighter than $m_{\text{VIS}} = 24.5$ in the *Euclid*-VIS band. Although the catalogue thus lacks high redshift galaxies, it is sufficient for our purposes because the spatial variations are larger at lower redshifts where a fixed angular scale probes a smaller volume. We retrieved 9 patches, each $10 \times 10^\circ$, to determine the dispersion in galaxy counts when we subdivide these data into smaller areas. The relative variation is a direct estimate for the dispersion in n_{fac} , which in turn can be converted into an estimate of the uncertainty in the multiplicative bias.

If we wish that the contribution to the uncertainty in the multiplicative bias due to the uncertainty in the mean galaxy density is $< 10^{-4}$, the observed sensitivity of the bias after METACALIBRATION implies that we need to know n_{fac} with a relative precision of about 2.6%. If we consider the variation in galaxy counts in a patch of 1 deg² the MICE simulations yield a dispersion of 0.064, which agrees remarkably well with observed estimates of the variation in galaxy counts by [Herbonnet et al. \(2020\)](#) on similar angular scales and depths. In contrast, the approximately 0.25 deg² covered by GEMS ([Rix et al. 2004](#)) would introduce on average a multiplicative bias of about 3×10^{-4} , taking up a significant part of the overall budget specified in [Cropper et al. \(2013\)](#).

Figure C.1 shows how the multiplicative bias μ depends on the observed area of sky used to estimate the mean density of galaxies in the image simulations. The error bars correspond to the dispersion in the measured counts in the patches. The results are well described by a power law with a slope of -0.27 ; the fit indicates that to achieve a bias $< 10^{-4}$ we need to measure the galaxy counts in an area of about 30 deg². This estimate may be somewhat optimistic because we did not consider the impact of small clustering in our image simulations. Of course the actual survey data can be used to validate the realism of the simulation, but in practice deeper observations of smaller areas are more useful as input to the image simulations. Hence, once can interpret these results as the minimum area for which deeper observations help to improve the fidelity of the image simulations.

¹³ MICECATv2 is publicly available at <https://cosmohub.pic.es/home>