A study of the sensitivity of shape measurements to the input parameters of weak-lensing image simulations

Hoekstra, H.; Viola, M.; Herbonnet, R.T.L.

Citation
Hoekstra, H., Viola, M., & Herbonnet, R. T. L. (2017). A study of the sensitivity of shape measurements to the input parameters of weak-lensing image simulations. *Monthly Notices Of The Royal Astronomical Society (Issn 0035-8711)*, *468*, 3295-3311. doi:10.1093/mnras/stx724

Version: Not Applicable (or Unknown)
License: Leiden University Non-exclusive license
Downloaded from: https://hdl.handle.net/1887/58962

Note: To cite this publication please use the final published version (if applicable).

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **468,** 3295–3311 (2017)
Advance Access publication 2017 March 24

doi:10.1093/mnras/stx724

A study of the sensitivity of shape measurements to the input parameters of weak-lensing image simulations

Henk Hoekstra,★ Massimo Viola and Ricardo Herbonnet
Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands

Accepted 2017 March 22. Received 2017 March 22; in original form 2016 September 8

ABSTRACT

Improvements in the accuracy of shape measurements are essential to exploit the statistical power of planned imaging surveys that aim to constrain cosmological parameters using weak lensing by large-scale structure. Although a range of tests can be performed using the measurements, the performance of the algorithm can only be quantified using simulated images. This yields, however, only meaningful results if the simulated images resemble the real observations sufficiently well. In this paper, we explore the sensitivity of the multiplicative bias to the input parameters of *Euclid*-like image simulations. We find that algorithms will need to account for the local density of sources. In particular, the impact of galaxies below the detection limit warrants further study because magnification changes their number density, resulting in correlations between the lensing signal and multiplicative bias. Although achieving sub-per cent accuracy will require further study, we estimate that sufficient archival *Hubble Space Telescope* data are available to create realistic populations of galaxies.

Key words: gravitational lensing: weak – dark energy – dark matter – cosmology: observations.

1 INTRODUCTION

In the past decades, the theoretical framework that describes the formation of cosmic structure has been tested by ever-increasing precise observations (see e.g. Planck Collaboration XIII 2016, for a comprehensive comparison of results), which are in general agreement. However, the main ingredients of this 'concordance model' are not understood at all: Dark matter and dark energy make up the bulk, 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. 2013, for an overview). To distinguish between such a multitude of ideas, dramatically better observational constraints are needed.

Of particular interest is the study of the distribution of matter as a function of redshift because it is sensitive to 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, i.e. a phenomenon called gravitational 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 dark 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: We can determine the statistical properties of the matter distribution as a function of cosmic time, which depend on the cosmological parameters (see e.g. Hoekstra & Jain 2008; Kilbinger 2015, for some recent reviews).

The typical change in ellipticity caused by gravitational lensing is about a per cent, 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 small areas of sky (e.g. Heymans et al. 2013; Jarvis et al. 2016; Jee et al. 2016; Hildebrandt et al. 2017), but future surveys will cover large 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 meaningful only 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. This will be challenging for the next generation of surveys, such as *Euclid*[1] (Laureijs et al. 2011), the *Wide-Field InfraRed Space Telescope*[2] (*WFIRST*; Spergel et al. 2015) and the Large Synoptic Survey Telescope (LSST; LSST Science Collaboration et al. 2009), which aim to measure the dark energy parameters with a precision much better than a per cent.

★ E-mail: hoekstra@strw.leidenuniv.nl

[1] http://euclid-ec.org
[2] http://wfirst.gsfc.nasa.gov

© 2017 The Authors
Published by Oxford University Press on behalf of the Royal Astronomical Society

Downloaded from https://academic.oup.com/mnras/article-abstract/468/3/3295/3089745
by Jacob Heeren user
on 10 January 2018

A detailed study of how systematic biases affect the measurements of galaxy shapes is presented in Massey et al. (2013). This work showed, not surprisingly, that the point spread function (PSF) is the dominant source, driving the desire for space-based observations (also see Paulin-Henriksson et al. 2008). Another complication is the fact the shapes are measured from noisy images. Recent studies have shown that this leads to biases in the ellipticity (e.g. Melchior & Viola 2012; Refregier et al. 2012; Miller et al. 2013). 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. 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 then immediately provides an estimate of the bias. For instance, Erben et al. (2001) and Hoekstra et al. (2002) used simulated images to examine the performance of the KSB algorithm developed by Kaiser, Squires & Broadhurst (1995). To benchmark the performance of a wider range of algorithms, the Shear TEsting Programme (STEP; Heymans et al. 2006; Massey et al. 2007) created a blind challenge: The input shear was unknown to the participants. The results showed a range in performance, even for algorithms that were, in principle, rather similar (such as various implementations of the KSB algorithm). This demonstrated the importance of how a method is actually implemented. To examine the origin of the variation in performance further, a series of challenges were carried out using highly idealized simulations. These Gravitational LEnsing Accuracy Testing (GREAT) challenges (Bridle et al. 2010; Kitching et al. 2012; Mandelbaum et al. 2015) have resulted in a steady improvement in the accuracy of the algorithms, given the metric used to compare them, while also demonstrating the importance of noise on the performance.

However, as recently shown by Hoekstra et al. (2015, hereafter H15), 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. The fidelity of the image simulations is therefore crucial, not only to quantify biases in the shape measurements, but also to correctly capture the selection of galaxies (e.g. Fenech Conti et al. 2017). The aim of this paper is a first exploration of the sensitivity of shape measurement algorithms to some of the most basic input parameters in preparation for the next generation of cosmic shear surveys, and *Euclid*, in particular. This will help define the range of parameters to consider and to measure from actual data or simulations.

In Section 2, we describe the basic principles of calibrating a shape measurement pipeline and introduce the algorithm we use. The image simulations are described in Section 3. We explore the sensitivity to the noise level in Section 4. The dependence on the properties of bright galaxies is quantified in Section 5. The impact of faint galaxies is explored in Section 6. The response to the input ellipticity distribution is studied in Section 7, and different implementations of the algorithm are examined in Section 8. The effect of stars is evaluated in Section 9.

2 THE NEED FOR A CALIBRATED ALGORITHM

The measurement of accurate shapes of small, faint galaxies from noisy data is a critical step in any weak-lensing analysis. For this reason, much effort has been focused on reducing the biases in the measurements of the ellipticity, in particular, the correction for the smearing by the PSF, which leads to rounder images (due to the size of the PSF) and preferred orientations (if the PSF is anisotropic). Moreover, all algorithms that measure shapes for individual galaxies are sensitive to the noise in the images (Viola, Kitching & Joachimi 2014). Consequently, an ideal algorithm is able to account for both the biases introduced by the PSF and the noise because both tend to vary between exposures. We note, however, that the situation is complicated further because the object selection itself may lead to bias: The significance with which galaxies are detected typically depends on their orientation with respect to the shear or the PSF (Kaiser 2000; Bernstein & Jarvis 2002; Hirata & Seljak 2003). For instance, Fenech Conti et al. (2017) find that the selection bias can be as important as the shape measurement bias. Although we do not study selection bias in this paper, it is clearly another important topic to study in future work.

The performance of shape measurement algorithms can be estimated using image simulations. Early comparisons (e.g. Heymans et al. 2006; Massey et al. 2007) included some of the complexity of real data, such as blending of galaxies. To examine differences between algorithms better, most recent studies considered idealized circumstances; for instance, the GREAT challenges (Bridle et al. 2010; Kitching et al. 2012; Mandelbaum et al. 2015) focused on isolated galaxies. However, at the accuracy required for the next generation of cosmic shear surveys, it is not sufficient to consider such idealized scenarios as a multitude of subtle effects prevent a straightforward interpretation of the inferred ellipticity in actual data.

Recently, Huff & Mandelbaum (2017) and Sheldon & Huff (2017) explored an alternative approach, where the actual survey data are modified, thus avoiding the use of synthetic data. In this method, which is similar to the one proposed by Kaiser (2000), the data are sheared by a known amount, and convolved with additional PSFs to mimic the variation in observing conditions. The response to these changes provides an estimate of the multiplicative bias for a particular (biased) shape measurement method. Huff & Mandelbaum (2017) show how this can reduce the multiplicative bias for a range of methods. Although these initial results are encouraging, it is not yet clear whether it is possible to achieve the stringent requirements for the next generation of surveys, such as *Euclid*.

Sources of bias affect the lensing results in two ways. First of all, systematics may lead to spurious correlations in the shapes of galaxies, resulting in an additional signal, i.e. it causes an additive bias c. Although correcting for the various sources of additive bias may not be easy, residual systematics can typically be identified by considering cross-correlations between the galaxy shapes and the source of bias. A well-known example is the star–galaxy correlation, which is sensitive to residuals in the correction for PSF anisotropy (see Heymans et al. 2012, for a detailed example).

Secondly, the amplitude may be biased by a factor of $(1 + \mu)$, i.e. the systematics cause a multiplicative bias. The potential of cosmic shear to constrain dark energy models relies on an accurate determination of the amplitude of the lensing signal as a function of source redshift. The amplitude of the lensing signal around galaxies as a function of (photometric) redshift can be used to test for the presence of multiplicative bias (Velander et al. 2014). Thanks to tremendous progress in cosmic microwave background (CMB) experiments, comparisons to (future) CMB lensing constraints provide an alternative observational test (e.g. Liu, Ortiz-Vazquez & Hill 2016; Schaan et al. 2016), although the precision may not be sufficient. We note, however, that these tests are compromised if

Downloaded from https://academic.oup.com/mnras/article-abstract/468/3/3295/3089745
by Jacob Heeren user
on 10 January 2018

the photometric redshifts themselves are biased. Therefore, unlike additive biases, the multiplicative bias can be reliably assessed only through image simulations.[3]

The observed shear and true shear are thus related by

$$\gamma_i^{\rm obs} = (1+\mu)\gamma_i^{\rm true} + c, \quad (1)$$

where we implicitly assumed that the biases are the same for both shear components. We do so to reduce the number of simulations we need, but note that, in practice, the biases need to be determined for both shear components separately, as we do not expect them to be the same. Moreover, μ and c may vary spatially, for instance, because the PSF properties themselves do. In this paper, we do not consider such complications.

Left unaccounted for, multiplicative and additive biases lead to systematic errors in the inferred cosmological parameters. By requiring that the systematic shifts in the parameters of interest are at least smaller than some fraction of the expected statistical uncertainties from a survey, the maximum allowed range for μ and c can be specified. In the case of *Euclid*, the design is based on the dark energy parameters $w_{\rm p}$ and $w_{\rm a}$, which describe the constant and dynamic nature of the dark energy, respectively (see e.g. Laureijs et al. 2011).

Note that the problem is not the amplitude of the bias, but rather how well the bias can be determined: A known bias can be incorporated as part of an empirical calibration step, thus reducing the 'effective' residual bias. Hence, a robust bias may be preferred over a smaller bias that is more sensitive to variations in the data or the input parameters of the simulations: The objective should be to reduce the sensitivity $|\partial\mu/\partial p|$, where p is a parameter that may affect the multiplicative bias, for instance, the noise level. This is particularly important if the parameter of interest is correlated with the lensing signal itself. For example, the sensitivity to faint, undetected galaxies results in correlations with the large-scale structure that we are trying to measure (see Section 6 for more discussion). In some cases, the uncertainty in the parameter p may be so large that it leads to an error in the bias that exceeds the requirement: The method is not calibratable.

Before we continue, it is useful to distinguish between two types of sensitivity. First of all, methods are sensitive to parameters that are required to correct for the various sources of bias; incorrect estimates of these observables lead to biased shape measurements. For example, the sensitivity to the PSF parameters scales with the ratio of the square of the PSF size over the galaxy size (Paulin-Henriksson et al. 2008; Massey et al. 2013). Similarly, the correction for noise bias (e.g. Fenech Conti et al. 2017) does not reduce the accuracy with which the noise level needs to be determined. Hence, the sensitivity of the algorithm to errors in these parameters needs to be quantified to establish requirements on how well these parameters need to be determined from the data.

The other kind of sensitivity determines how well the parameters of interest need to be captured by the image simulations. A shape measurement algorithm that is truly unbiased[4] will still be sensitive to errors in the PSF size, noise level, etc. However, these dependences can be readily determined by computing the change in bias when varying the input parameters. Hence, no image simulations are required because the use of the correct input parameters is guaranteed to yield an unbiased estimate of the shear.

On the other hand, if a method is biased, image simulations are required to determine the bias and its sensitivities to the various input parameters (but see Huff & Mandelbaum 2017, for an alternative approach). The sensitivities can be quantified using simulations where only the parameters of interest are varied. If we assume that most effects act independently, as is done in this paper, this can be done for each parameter separately. Hence, the sensitivities to input parameters can be quantified using a significantly reduced number of galaxies for which shapes need to be measured.

Once it has been established that the input parameters are sufficiently realistic, the actual bias can be determined. The resulting uncertainty in the bias should be small compared to the statistical uncertainties from the survey itself, thus defining the size of the image simulations. In this regard, there is no immediate advantage to use less biased methods, or even less sensitive methods, unless the source of bias can be eliminated. However, a lower sensitivity to a given parameter is clearly preferable because it does relax the requirements on how well it needs to be captured by the simulations.

In this paper, we examine how the multiplicative bias is affected by changes in the input parameters of the simulations and by modifications in the analysis pipeline. We consider only the multiplicative bias because it is the most constraining. It has the added benefit that we do not have to consider a suite of PSFs with different ellipticities.

The results presented in Massey et al. (2013) suggest a maximum allowed residual multiplicative bias[5] of $|\mu_{\rm tot}| < 2 \times 10^{-3}$. However, as discussed in Massey et al. (2013), a number of effects contribute to this bias, such as errors in the PSF determination and other corrections for instrumental effects. A detailed discussion of a possible breakdown is presented in Cropper et al. (2013). We note that these studies considered requirements under the assumption that systematic effects do not depend on scale. This can result in conservative limits, as was discussed in Kitching et al. (2016). None the less, in order to minimize the multiplicative bias caused by shortcomings of the image simulations, we consider an ambitious value of $|\mu_{\rm sim}| = 10^{-4}$. We note that this is not an actual allocation, but rather sets the scope of the calculations and places requirements on the knowledge of the input parameters.

To reach a statistical uncertainty of 10^{-4} for the multiplicative bias, a large number of galaxies need to be analysed. If we consider a shear of 0.01 and an intrinsic ellipticity of 0.3, then a sample of 10^{11} galaxies would be needed. This estimate, however, is too pessimistic because the uncertainty is dominated by the intrinsic ellipticity. To reduce this source of noise, pairs of images, with one rotated by 90°, can be used (see e.g. Massey et al. 2007). The use of more rotations, for example, four images rotated by 45°, suppresses shape noise more efficiently (Fenech Conti et al. 2017), but the performance is ultimately limited by the pixel noise in the images, such that, in practice, still about 10^{10} galaxies are required. To ensure that the inferred biases are robust against the uncertainties in the input parameters, we wish to explore a range of simulated data. To achieve these objectives within a reasonable amount of time and limited computational resources, we thus need to use a sufficiently fast algorithm.

2.1 Description of the shape analysis

The impact of (relatively) static sources of bias can be determined from image simulations, provided they are known and sufficiently

[3] Unless it is a priori known that the method is unbiased.

[4] Thai is, it can be proved that the algorithm is unbiased in realistic conditions.

[5] We note that our notation corresponds to that of equation (11) in Massey et al. (2013), such that $\mathcal{M} \approx 2\mu$.

Downloaded from https://academic.oup.com/mnras/article-abstract/468/3/3295/3089745
by Jacob Heeren user
on 10 January 2018

well characterized. However, the instrument configuration varies with time as does the atmosphere in the case of ground-based observations; an ideal shape measurement method should therefore accurately correct for the resulting temporal variations in the PSF in order to avoid having to create a large suite of simulations for all possible PSFs. In this paper, we do not consider varying PSFs, but assume that this can be accurately corrected for. Furthermore, we do not consider the impact of selection effects, which will be important in realistic situations, as shown by Fenech Conti et al. (2017). Instead, we focus on the sensitivity of the shape measurements to the input parameters of the image simulations.

Given an image and a model for the PSF (we assume that other sources of bias have been removed to a sufficient level of accuracy), different approaches can be used to estimate the true galaxy shape. For instance, one can fit a parametrized model of the surface brightness distribution to the data by representing galaxies by a decomposition into shapelets (Refregier & Bacon 2003). The resulting model can then be deconvolved analytically to yield an estimate of the underlying galaxy shape. However, the expansion into shapelets needs to be truncated because of noise.

To avoid the problems with direct deconvolution, forward modelling techniques have become more popular. In this case, a model image is sheared, convolved with the PSF and pixellated. The model parameters are varied until a best fit to the data is obtained. This step requires many calculations, especially if more model parameters are to be determined. Accurate priors for the parameters are required to obtain an unbiased estimate for the shear. These priors can be derived from high-quality observations, similar to what is needed as input for the image simulations. Moreover, the data themselves can be used to update the priors. Analogous to the need to truncate the shapelet expansion, the model should provide a good description of the galaxies, while having a limited number of parameters in order to avoid overfitting. A model that is too rigid will lead to model bias (e.g. Bernstein 2010), whereas a model that is too flexible tends to fit noise in the images (e.g. Kacprzak et al. 2012). We note that this can mitigated by a marginalization of the nuisance parameters in the model (Miller et al. 2013).

Unfortunately, fitting methods require many evaluations, thus increasing the computational time per galaxy significantly. We focus instead on an alternative approach: We measure the moments of the galaxy images, which are subsequently corrected for the PSF. The shapes are quantified by the polarization[6]

$$e_1 = \frac{I_{11} - I_{22}}{I_{11} + I_{22}} \quad \text{and} \quad e_2 = \frac{2I_{12}}{I_{11} + I_{22}}, \tag{2}$$

where the quadrupole moments I_{ij} are given by

$$I_{ij} = \frac{1}{I_0} \int \mathrm{d}^2\boldsymbol{x}\, x_i\, x_j\, W(\boldsymbol{x})\, f(\boldsymbol{x}), \tag{3}$$

where $f(\boldsymbol{x})$ is the observed galaxy image, $W(\boldsymbol{x})$ is a suitable weight function to suppress the noise and I_0 is the weighted monopole moment.

Matching the width of the weight function to the object size maximizes the signal-to-noise ratio (S/N) of the shape measurement. For the weight function, we adopt a Gaussian with a dispersion r_g determined by the half-light radius r_h measured by SExtractor (Bertin & Arnouts 1996). As reference, we consider $r_\mathrm{g} = r_\mathrm{h}/\sqrt{2}$, which is slightly smaller than the optimal value for a Gaussian, which would imply $r_\mathrm{g} = r_\mathrm{h}/\sqrt{2\ln 2}$. This choice does not affect our conclusions, and we explore different choices in Section 8. A further sophistication would be to try to match the shape of the weight function (e.g. Melchior et al. 2011; Okura & Futamase 2011). This optimization slows down the shape measurement algorithm and increases the sensitivity to the input ellipticity distribution. As the gain is expected to be small for the smallest galaxies, we limit our study to an axisymmetric weight function.

In practice, an estimate for the background needs to be subtracted from the observed image and contributions from nearby objects need to be suppressed. In our reference set-up, we simply mask pixels within a radius of $4r_\mathrm{g}$ around neighbouring objects, and the background is determined locally by considering an annulus with inner and outer radii of 16 and 32 pixels, respectively, from the centroid of the object. Objects located in the annulus are masked and a plane is fitted to the counts in the unmasked pixels. We prefer a local background determination over a global one because biases due to artefacts are limited to relatively small scales, and thus do not introduce coherent biases on scales that are relevant for the cosmological interpretation. In Section 8, we explore different settings, demonstrating that the background determination is an essential aspect of the algorithm performance.

The resulting weighted moments are biased because of the weight function and PSF. To undo these, we focus here on the commonly used KSB method developed by Kaiser et al. (1995) and Luppino & Kaiser (1997) with corrections provided in Hoekstra et al. (1998) and Hoekstra, Franx & Kuijken (2000), which was used in H15. The only difference with H15 is that we use SExtractor for the object detection step to speed up the analysis. We note that the KSB algorithm makes simplifying assumptions about the PSF, which are not valid for realistic cases, such as the *Euclid* PSF. However, this can be accounted for with improved moment-based methods such as DEIMOS (Melchior et al. 2011).

We stress, however, that the aim of this paper is not to find a suitable shape measurement algorithm, nor are we interested in the value of the bias. Instead, we explore the change in bias as a function of the input parameters of the image simulations. Exactly this crucial step has been largely overlooked in previous work. However, as we will see, most of the variations in bias are small and therefore not that important for current surveys. On the other hand, our results also clearly demonstrate that the situation is fundamentally different for *Euclid* because of the much more stringent requirements.

3 DESCRIPTION OF THE IMAGE SIMULATIONS

A flexible framework to create simulated images is provided by GALSIM (Rowe et al. 2015), a publicly available code that was developed for GREAT3 (Mandelbaum et al. 2014, 2015). Here we use simple parametric models for the galaxies, and the main input is a list of galaxy properties with a position, flux, half-light radius, Sérsic index and ellipticity, from which sheared galaxy images are computed. The reference galaxy parameters are described in more detail in Section 3.1.

As input to the simulations, we use a sample of galaxies for which morphological parameters were measured from resolved *F*606*W* images from the GEMS survey (Rix et al. 2004). These galaxies were modelled as single Sérsic models with GALFIT (Peng et al. 2002), and for our study, we use the measured half-light radius, apparent magnitude and Sérsic index n. For simplicity, we consider only galaxies fainter than magnitude $m = 20$.

[6] We define the ellipticity $\epsilon \equiv (a-b)/(a+b)$, with a and b being the major and minor axes, respectively. The polarization e for such a galaxy would be approximately $(a^2 - b^2)/(a^2 + b^2)$.

Downloaded from https://academic.oup.com/mnras/article-abstract/468/3/3295/3089745
by Jacob Heeren user
on 10 January 2018

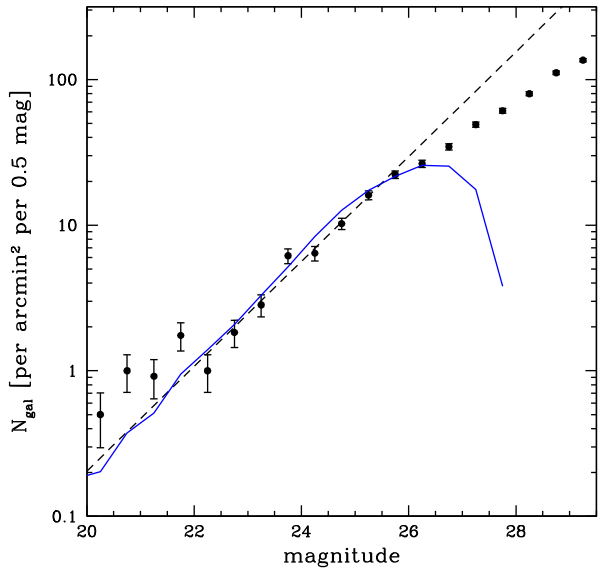

Figure 1. Number density of galaxies as a function of limiting magnitude. The dashed black line indicates our reference model, which is a power law with a slope of 0.36, which is a good description of the GEMS counts (solid blue line) for $20 < m < 25$. The points with error bars are the *F*775*W* counts from Coe et al. (2006) based on the HUDF, which suggest a flatter slope at faint magnitudes.

The solid blue line in Fig. 1 shows the number density of galaxies in the GEMS catalogue as a function of apparent magnitude. For $20 < m < 25$, the counts are well described by a power law with a slope of 0.36, indicated by the black dashed line. Comparison with the observed *F*775*W* counts from the *Hubble Ultra-Deep Field* (HUDF) by Coe et al. (2006) shows that for $m > 26$, the GEMS catalogue is incomplete. Although the HUDF counts indicate a flattening of the slope for $m > 25.5$, our reference simulations assume a simple power law with slope 0.36, but in Section 6.2, we explore the sensitivity of the results to this assumption.

The input catalogue is normalized to 36 galaxies arcmin^{-2} with $20 < m < 24.5$. We choose a nominal noise level per pixel with a dispersion of 0.8, which results in a typical number density of 47 galaxies arcmin^{-2} with a S/N larger than 10, as measured by SExtractor, and a number density of 33 galaxies arcmin^{-2} if we restrict the magnitude range to $20 < m < 24.5$. We measure shapes for all detected galaxies, but report results for galaxies with $20 < m < 24.5$ and $r_{\rm h} > 0.11$ arcsec, unless specified otherwise. For this magnitude range, the cut in half-light radius cleanly separates galaxies from stars. We note that these number densities are somewhat higher than the nominal values for *Euclid* (Laureijs et al. 2011).

We create pairs of images where the galaxies are rotated by 90° in the second image in order to reduce the noise due to the intrinsic ellipticity distribution (see e.g. Massey et al. 2007): By construction, the mean intrinsic ellipticity when both are combined is zero. We analyse the images separately, and, thus, due to noise in the images, this is no longer exactly true, especially for faint galaxies. The input shears range from −0.06 to 0.06 in steps of 0.01 (for both components), yielding 169 image pairs for each 'set'. We use the `galsim.applyShear()` function,[7] which preserves the area of the object, i.e. the galaxies are not magnified. We verified this by measuring the average sizes of the galaxies as a function of the applied shear.[8] This greatly simplifies the interpretation of our results, as magnification changes the galaxy selection as a function of shear. Although not the focus of this paper, we discuss the implications of our results on size magnification studies in Section 5.3.

Each image has a size of 10 000 × 10 000 pixels, with a pixel scale of 0.10 arcsec, corresponding to that of the *Euclid* VIS camera. Down to a limit $m = 29$ (see Section 6) for each pair of images, we include 10^6 objects, or 1.7×10^8 per set. To reduce the statistical uncertainties further, we simulate typically tens of sets created with different random seeds. The analysis of a single set, using the reference set-up of the KSB algorithm, takes approximately 60 core hours on a Dell PowerEdge R820 with Intel Xeon E5-4620 2.20-GHz processors. For example, the results presented in Fig. 7 took over 34 000 core hours.

To simulate a diffraction-limited telescope in space, we adopt a circular Airy PSF for a telescope with a diameter of 1.2 m and a PSF obscuration of 0.3 at a reference wavelength of 800 nm. This is a reasonable approximation to the *Euclid* PSF in the VIS-band. We include a small number of bright stars in the simulations, which are used to measure the PSF properties required to correct the galaxy shapes. In this paper, we do not consider the complications that arise from modelling of the PSF. In Section 9, we do explore the impact of variations in the star density on the multiplicative bias.

3.1 Input galaxy properties

The morphological properties, such as sizes, shapes and surface brightness profiles, are correlated: Fainter galaxies are, on average, smaller, whereas disc-dominated galaxies show a broader ellipticity distribution. To capture some of these correlations, we use measurements of morphological parameters (specifically magnitude, observed half-light radius and Sérsic index) from the resolved *F*606*W* images from the GEMS survey (Rix et al. 2004). The use of Sérsic profiles to describe the galaxies limits the fidelity of the simulations (see Kacprzak et al. 2014, for a study of the biases that may arise), and future work will need to examine how well morphological parameters need to be determined, including the spatial variation of galaxy colours (Semboloni et al. 2013).

Although Rix et al. (2004) also estimated ellipticities, we ignore any correlation with ellipticity, but randomly draw ellipticity values from a Rayleigh distribution given by

$$P(\epsilon;\epsilon_0) = \frac{\epsilon}{\epsilon_0^2}\,{\rm e}^{-\epsilon^2/2\epsilon_0^2}, \qquad (4)$$

where the value of ϵ_0 determines the width of the distribution, as well as the average $\langle\epsilon\rangle = \epsilon_0\sqrt{\pi/2}$. We need to truncate the distribution because the ellipticity cannot exceed unity, but also because galaxy discs have a finite thickness. We therefore set $P(\epsilon, \epsilon_0) = 0$ if $\epsilon > 0.9$.

As a reference value, we adopt $\epsilon_0 = 0.25$, which best described the data used in H15. This simplifying assumption, i.e. $P(\epsilon)$ is independent of other galaxy properties, will need to be studied in future work, as the ellipticity distributions for early- and late-type galaxies differ (e.g. van Uitert et al. 2011), which, in turn, results in dependences with the environment (Kannawadi, Mandelbaum & Lackner 2015). In practice, the ellipticity distributions for various

[7] In version 1.1, this method was deprecated.

[8] We found that the mean observed half-light radius increased by a negligible 0.2 per cent for the largest shear we consider here. However, closer investigation revealed that this change in size is solely due to a direction-dependent feature in the way the half-light radius is determined.

Downloaded from https://academic.oup.com/mnras/article-abstract/468/3/3295/3089745
by Jacob Heeren user
on 10 January 2018

subsets will need to be measured from deep observations (Viola et al. 2014).

We consider bins with a width of 0.1 in magnitude and compute the expected number of galaxies assuming a power law for the galaxy counts as described above. Down to a limiting magnitude of $m = 25.4$, we randomly draw galaxies from the corresponding magnitude bin from the GEMS catalogue. For fainter galaxies, we create duplicates with different orientations and place those postage stamps in the image. This speeds up the creation of the simulated images.

3.1.1 *Sizes of faint galaxies*

H15 showed that a robust estimate of the multiplicative bias requires the inclusion of sufficiently faint galaxies in the image simulations: For current ground-based observations, the bias converges when galaxies that are 1.5 mag fainter than the faintest galaxy used in the lensing analysis are present. As discussed in more detail in Section 6, in the case of *Euclid*, we may need to consider galaxies as faint as magnitude 29.

For galaxies brighter than $m = 26.5$, we use the observed half-light radii from Rix et al. (2004) instead of the GALFIT estimate for the effective radius because it is a more robust estimate.[9] However, for $m > 26.5$, the GEMS catalogue becomes progressively incomplete, resulting in a biased distribution of galaxy sizes. Given the small space-based PSF and the stringent requirements of the *Euclid* mission, the adopted distribution of galaxy sizes may be relevant. This is quantified in Section 6.1, but here we describe how we parametrized the size distribution of the faint galaxies.

As shown in Appendix A, the distribution of observed half-light radii for bright galaxies can be approximated by a skewed lognormal distribution. We keep the skewness fixed to a representative value of -0.58 and determine the mean and dispersion of $\log_{10} r_{\rm h}$ as a function of apparent magnitude. The results are presented as the black points in Fig. 2. For galaxies with $23 < m < 25.5$, both the mean and the dispersion of $\log_{10} r_{\rm h}$ are well described by a simple linear relation. We determine the best fit in this magnitude range (dashed lines) and use this to describe the size distribution for galaxies with $m > 26.5$.

Coe et al. (2006) provide estimates for the effective radii, $r_{\rm eff}$, from the best-fitting GALFIT model using *HST* observations of the *HST* HUDF. To allow for a direct comparison to our input parameters, we convert the values for $r_{\rm eff}$ into half-light radii using an empirical relation based on the GEMS catalogue, which provides both size estimates (see Appendix A for details). We consider only galaxies for which the effective radius was determined with a relative precision <10 per cent and present the results in the top panel of Fig. 2 (red points). The agreement with the GEMS measurements and our parametric model is good for $m < 26$. Interestingly, the results from Coe et al. (2006) suggest that the sizes of the faint galaxies may be larger than we assume here.

Similarly, we use the GEMS results to relate the dispersion in $r_{\rm eff}$ to an estimate of the scatter in $r_{\rm h}$. As shown in the bottom panel of Fig. 2, the resulting converted measurements from Coe et al. (2006) roughly match our adopted model. Although this level of agreement is adequate for the purpose of this paper, it will clearly be worthwhile to revisit the size measurements presented in Coe et al. (2006).

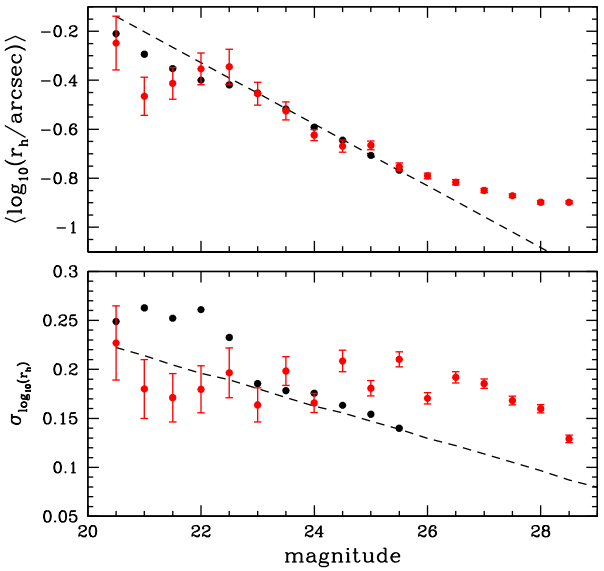

Figure 2. Top panel: mean logarithm of the observed half-light radius as a function of apparent magnitude. The black points indicate the measurements from GEMS that are used to derive our parametric model (indicated by the dashed line; see the text for details). The red points indicate the results from the HUDF, suggesting that the actual sizes of faint galaxies may be larger. Bottom panel: width of the galaxy size distribution as a function of limiting magnitude, where the black points correspond to GEMS and the red points are from the HUDF. The dashed line shows the value for the parametric model adopted for the image simulations.

4 SENSITIVITY TO NOISE

As shown in Viola et al. (2014), any ellipticity measurement is biased in the presence of noise in the image because the estimator is non-linear in terms of the pixel values: Equation (2) involves a ratio of moments. As a consequence, the observed ellipticity distribution is skewed and not centred on the true value. This needs to be accounted for when estimating the shear from an ensemble of sources because it otherwise leads to a multiplicative bias.

In the actual observations, the background level is expected to vary: in ground-based data due to the moon or changing atmospheric conditions and in space-based observations because of the zodiacal background that varies across the sky. Consequently, the shape measurement algorithm needs to be able to account for the dependence on the S/N (Miller et al. 2013; H15; Fenech Conti et al. 2017), which, in turn, implies that the statistics of the background need to be determined sufficiently well. This is typically done by measuring the distribution of unmasked pixel values where no galaxies were detected, i.e. assuming that the noise is uncorrelated. We turn to the complications posed by faint undetected galaxies in Section 6, and consider a simple background first.

To do so, we simulate 20 sets of images where we include galaxies down to a limiting magnitude of $m_{\rm lim} = 24.5$. We add Gaussian noise with a dispersion $\sigma_{\rm bg}$ (the images are created with a mean background of zero). Note that we are not attempting to simulate the actual observing process, which involves the combination of

[9] These are the values measured by SExtractor, and thus not corrected for the *Hubble Space Telescope* (*HST*) PSF. This omission, which we discovered during the refereeing process, slightly biases the sizes used in our analysis: At $m = 24.5$, the corrected sizes are 4 per cent smaller and the differences are even smaller for brighter galaxies. The change in size is larger for fainter galaxies, but this affects only the recovered biases and not the sensitivities themselves (see e.g. Fig. 8).

Downloaded from https://academic.oup.com/mnras/article-abstract/468/3/3295/3089745
by Jacob Heeren user
on 10 January 2018

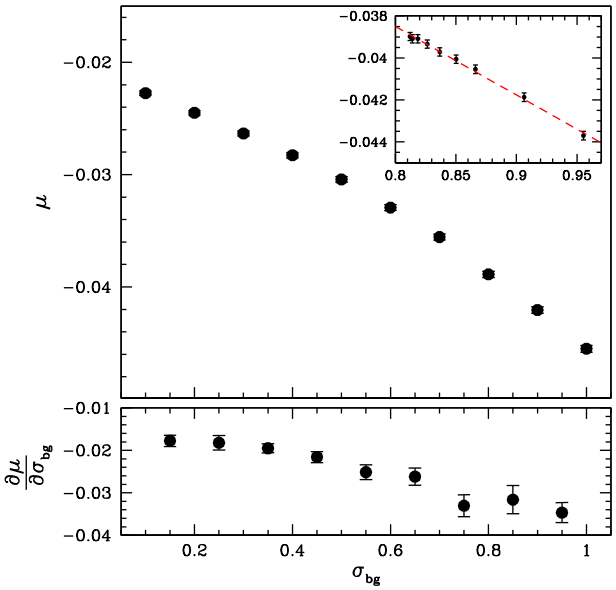

Figure 3. Multiplicative bias for galaxies with $20 < m < 24.5$ as a function of $\sigma_{\rm bg}$, the rms of the background. The error bars indicate the dispersion in the results. The bias increases with increasing noise level and the slope $\mathrm{d}\mu/\mathrm{d}\sigma_{\rm bg}$ steepens, as is evident from the bottom panel. For our implementation of the KSB algorithm, we find that the bias around $\sigma_{\rm bg} = 0.8$ (see the inset in the top panel) can be approximated by a linear relation with a slope $\partial\mu/\partial\sigma_{\rm bg} = -0.0326 \pm 0.0007$.

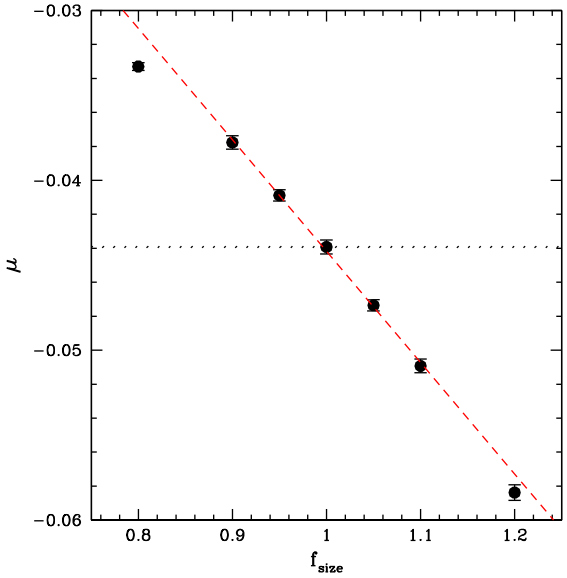

Figure 4. Change in multiplicative bias for galaxies with $20 < m < 24.5$ when the sizes of the input galaxies (we include only galaxies brighter than $m_{\rm lim} = 26$ in the simulations) are increased by a factor of $f_{\rm size}$. For reference, the hatched region indicates a variation of 10^{-4} in the value of μ. The red dashed line is the best-fitting linear relation between μ and $f_{\rm size}$.

multiple exposures. To reduce the number of simulations, the seeds for the noise realizations and galaxy properties are the same between sets. The images are analysed as described in Section 2.1. The resulting observed multiplicative bias μ as a function of the background level is presented in the top panel of Fig. 3. As expected, the bias increases when the background noise level is higher. Also note that μ does not vanish in the absence of noise, a demonstration of the fundamental limitation of using the KSB algorithm. The inset panel shows the change in bias close to the nominal background $\sigma_{\rm bg} = 0.8$, which we use for the other results presented in this paper.

The bottom panel of Fig. 3 shows the slope $\partial\mu/\partial\sigma_{\rm bg}$, which steepens as the noise level increases. When considering a small range, such as the one shown in the inset of the top panel, a constant slope is a good approximation, and around $\sigma_{\rm bg} = 0.8$, we measure a value $\partial\mu/\partial\sigma_{\rm bg} = -0.0326 \pm 0.0007$.

We can use this result to estimate how well the background rms needs to be determined, given an allocation for the bias that can be tolerated: A maximum uncertainty of $\delta\mu = 10^{-4}$ implies that $\sigma_{\rm bg}$ needs to be measured with a relative precision of approximately 0.3 per cent. If the noise is homoscedastic and Gaussian, as we assumed here, this requires about 2×10^5 blank pixels. In practice, undetected cosmic rays, galaxies, flat-field errors, etc. may also contribute to the background statistics. These are naturally included when blank pixels are used to characterize the background, provided instrumental effects that modify the statistics of the observed background do not do so over the area that contains 2×10^5 blank pixels. If we assume that half the pixels are blank, this corresponds to a square patch of 650 pixels on a side, which is much smaller than the section that is controlled by a single component of read-out electronics.

We stress that we are concerned only with coherent errors in the determination of background statistics as these lead to bias on cosmologically interesting scales. Small-scale effects that do not correlate between detectors or exposures merely increase the measurement noise slightly, which is negligible compared to the intrinsic shape noise on such small scales (see e.g. Kitching et al. 2016).

5 PROPERTIES OF BRIGHT GALAXIES

The presence of noise in the images prevents us from using unweighted moments, for which the correction for the smearing by the PSF is trivial. To relate the observed weighted moments to the true (unweighted) moments requires estimates of the higher order moments or equivalently the morphology of the galaxies (e.g. Semboloni et al. 2013). Only noisy estimates can be obtained from the data, and thus the bias depends not only on the noise level, but also on the underlying distribution of galaxy morphologies. This remains an important open topic of study, albeit not as prominent as the two properties that we will study in this section: the sizes and number densities of galaxies.

5.1 Sensitivity to galaxy size

The larger a galaxy, relative to the PSF, the easier it should be to measure its shape. Consequently, the multiplicative bias is typically a (strong) function of the galaxy size. To quantify the sensitivity to input galaxy size, we make simulated images where we change the sizes of the input galaxies by a factor of $f_{\rm size}$. We created 10 sets for each value of $f_{\rm size}$, where we include galaxies down to a limiting magnitude $m_{\rm lim} = 26$. We measure μ for galaxies with $20 < m < 24.5$ and show the results in Fig. 4. The sensitivity is substantial, with the dashed line indicating the best fit to the measurements for $0.9 \leq f_{\rm size} \leq 1.1$. Note that the slope is so steep because for our implementation μ is a strong function of galaxy size. We find a slope $\partial\mu/\partial f_{\rm size} = -0.0656 \pm 0.0010$, which would imply that $f_{\rm size}$ should be determined with a precision of 0.15 per cent if we consider a maximum tolerance of $\delta\mu = 10^{-4}$.

This is unnecessarily conservative because it would be apparent from the data themselves that the mean sizes differ from the

Downloaded from https://academic.oup.com/mnras/article-abstract/468/3/3295/3089745
by Jacob Heeren user
on 10 January 2018

simulations. This could then be corrected by improving the realism of the image simulations. For instance, Bruderer et al. (2016) propose the use of Monte Carlo control loops to adjust the simulated data such that they are statistically consistent with the observations. An alternative approach was explored in Fenech Conti et al. (2017), who resampled the output from the simulations to match the size distribution in the data. Although this provides an improved estimate for the bias for the observed sample of galaxies, local variations in galaxy sizes will not be captured.

It is therefore preferable to account for any size dependence in the algorithm through an empirical calibration. Note that such a calibration, or 'training', is not restricted to size, but can include any parameter of interest, such as brightness, surface brightness profile, local environment, etc. In this case, the image simulations are used to identify correlations between multiplicative bias and observables. For instance, Tewes et al. (2012) explored the use of supervised machine learning. Similarly, Gruen et al. (2010) used a neural network to remove residual biases. Interestingly, the run time is not determined by the application of the trained algorithm to real data, but rather by the time it takes to analyse the simulated data. Compared to typical machine learning applications, the training sample is much larger than the actual data sample because we wish to reduce the uncertainty in multiplicative biases by considering a very large volume of image simulations. Importantly, the fidelity of the machine learning step depends critically on using appropriate inputs. As we discuss in Section 6, this includes capturing the impact of galaxies below the detection limit.

Inevitably, the empirical corrections are based on observed parameters that are noisy. As a consequence, the calibration may be biased if the input properties are incorrect. Moreover, the choice of size definition matters. This was highlighted in Fenech Conti et al. (2017), who showed how selection biases in both the detection and analysis steps may result in implicit selections in ellipticity, and consequently lead to biases in the recovered shear. Hence, particular care should be taken in ensuring that consistent size definitions are used. We do not study the impact of these selection effects here, but stress that these represent an important source of bias, especially when selections are made using parameters that correlate strongly with ellipticity. Moreover, as such selection biases appear inevitable in the presence of PSF anisotropy and shear, it remains unclear whether image simulations covering a wide range of instrument states can be avoided.

5.2 Sensitivity to galaxy density

Close pairs of galaxies are another complication in real data. Most recent work on the performance of shape measurement algorithms focused on isolated galaxies: In Bridle et al. (2010), Kitching et al. (2012), Mandelbaum et al. (2015) and Jarvis et al. (2016), only postage stamps of isolated galaxies were analysed, and thus the effects of blending were not included. The image simulations we study here do include close pairs, but do not capture the full complexity expected in real data as galaxies are not positioned randomly on the sky, but are instead clustered. Hence, the local density of galaxies varies significantly, with clusters of galaxies representing the most extreme cases. For instance, for a sample of massive clusters at $z \sim 0.2$, H15 find that the number density of the brightest galaxies ($20 < m < 21$) is, on average, increased by a factor of 2 at a radius of 1 Mpc, whereas the number density of fainter galaxies ($24 < m < 25$) is increased by approximately 20 per cent.

One of the objectives of the *Euclid* mission is to use the number density of galaxy clusters as a function of mass and redshift to constrain cosmological parameters (Sartoris et al. 2016), which relies on accurate mass estimates. As shown by Köhlinger, Hoekstra & Eriksen (2015), this should be possible, provided that the multiplicative bias for this application is comparable to the required accuracy for cosmic shear studies.

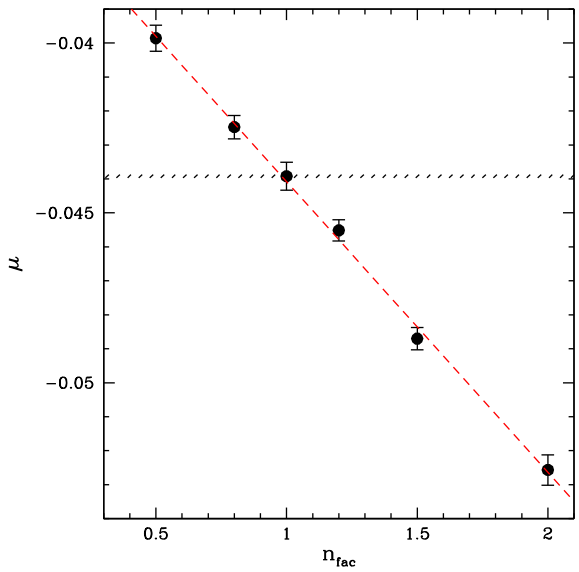

Figure 5. Multiplicative bias for galaxies with magnitude $20 < m < 24.5$ as a function of $n_{\rm fac}$, the relative increase in the number density of galaxies brighter than magnitude 26. For reference, the hatched region indicates a variation in μ of size 10^{-4}.

Given the impact of blending on shape measurements, it is useful to examine how μ depends on the number density of bright galaxies. To do so, we simulated observations with $m_{\rm lim} = 26$, while increasing the number density by a factor of $n_{\rm fac}$ compared to the reference case. We measure μ for galaxies with $20 < m < 24.5$ and show the results in Fig. 5. The red dashed line indicates the best-fitting linear relation with a slope $\partial\mu/\partial n_{\rm fac} = -0.008\,56 \pm 0.000\,17$. These results show a strong (linear) dependence on the local number density of bright galaxies. If unaccounted for, this will lead to significant biases in cluster mass estimates from *Euclid*.

The dependence on the number density of bright galaxies does not only affect cluster studies, but also complicate the cosmic-shear analysis. The large-scale structure that gives rise to the signal (with most of the contribution coming from haloes that correspond to galaxy groups) is correlated with the density of galaxies. Hence, the multiplicative bias is coupled to the cosmic-shear signal, and the image simulations should thus capture the clustering of galaxies well. This is a concern because the predicted clustering signal itself depends on cosmology.

We should therefore strive for algorithms that are minimally sensitive to neighbouring objects. Our deblending implementation is simple, and more sophisticated approaches will be studied in future work to quantify the impact of clustering. For instance, we note that Fenech Conti et al. (2017) measured a change of 2×10^{-3} in multiplicative bias for their simulations of ground-based data when the number density of galaxies was reduced by a factor of 2. We do expect fitting methods to perform better than our KSB implementation because the estimates of the moments are biased by simply masking blended objects. In contrast, fitting methods are naturally less sensitive to masked areas, but can also be adapted to fit multiple galaxies simultaneously. Whether this can fully eliminate the effect of blending should be studied further.

Downloaded from https://academic.oup.com/mnras/article-abstract/468/3/3295/3089745
by Jacob Heeren user
on 10 January 2018

As a first exploration for moment-based methods, we analysed a smaller set of image simulations using two alternative implementations (see Section 8 for more variations). First, we switched off the masking of neighbouring detected galaxies. In this case, the bias increases, and the slope $\partial\mu/\partial n_{\rm fac}$ is 30 per cent steeper. As blending may affect the local background determination, we also considered the case where the background is fixed to 0, the correct level in the simulated data. In this case, the bias is indeed reduced, but the slope is unchanged. These findings demonstrate that differences in implementation play a role, but appear unable to significantly reduce the sensitivity to blending. Something to investigate in future work is whether the impact of blending can be alleviated by interpolating over the masked regions.

Alternatively, the bias can be determined as a function of both distance to the nearest neighbour and its flux difference. For example, Fenech Conti et al. (2017) examined the additive bias as a function of galaxy separation, which is also affected by blending in the presence of an anisotropic PSF. This can be used to parametrize the residual biases caused by blending, thus reducing the sensitivity to the local density. Hence, in addition to the statistics of the local background, information about the local galaxy density should be included in the next generation of shape measurement algorithms.

Naively, an easier solution would be to remove close pairs from the analysis. Clear cases may indeed be identified and culled from the data, but some galaxies are blended to such a large degree that they are detected as single objects. As shown by Dawson et al. (2016), the latter are particularly relevant for deep ground-based observations and lead to an increase in the shape noise. Importantly, very strict criteria may result in undesirable reductions in source densities, especially in the case of deep data. The challenge is thus to find a balance between the sensitivity of the multiplicative bias due to blending and the increase in statistical uncertainties when blends are removed. Moreover, the preferential removal of sources behind overdense regions, where the lensing signal is highest, complicates the interpretation of the cosmological signal, as was shown by Hartlap et al. (2011). If ignored, the resulting shear correlation function can be biased low by a few per cent on scales of 1 arcsec. On the other hand, close pairs may also affect the fidelity of photometric redshift estimates. The impact of blending on the combination of shape and redshift determination is another open question, which requires further study using multiband image simulations.

5.3 Impact on size magnification

So far, we have focused on the measurements of galaxy shapes, but gravitational lensing also alters the sizes of galaxies and their fluxes because surface brightness is conserved. Although the measurement of the magnification signal is generally noisier than the shear, it can, in principle, be made using the same data that are used in the shear analysis. Moreover, the shear field is related to the projected surface density through a convolution (e.g. Kaiser & Squires 1993), whereas magnification, to leading order, probes the convergence field directly. This can help break parameter degeneracies, in particular, for studies of density profiles.

The change in flux modifies the number density of sources for a magnitude-limited sample, where the net effect depends on the slope of the number counts. The sources need not be resolved to measure this flux magnification signal, thus expanding the sample of potential sources to be used. The signal has been measured around clusters of galaxies (e.g. Hildebrandt et al. 2011; Ford et al. 2014; Umetsu et al. 2016) and galaxies (e.g. Hildebrandt, van Waerbeke & Erben 2009). The main challenges for flux magnification studies are the need for uniform photometry and a very clean separation of lens and source samples (see Hildebrandt 2016, for a detailed discussion on observational biases in flux magnification measurements).

On the other hand, the change in galaxy sizes, or size magnification, has not been widely used because it requires high-quality imaging. A number of results have been presented based on *HST* observations. Schmidt et al. (2012) studied a sample of galaxy groups using a combination of flux and size magnification, finding fair agreement between the shear and magnification measurements. Duncan et al. (2016) used *HST* observations of the A901/A902 supercluster and found that the magnification measurements yielded lower masses, although the statistical uncertainties are substantial.

Casaponsa et al. (2013) studied how well size magnification can be measured with LENSFIT (Miller et al. 2013) and concluded that an unbiased estimate of the convergence can be obtained, provided the source galaxies are larger than the PSF and have a S/N > 10. These constraints are similar to those for reliable shape measurements. Their image simulations, however, ignored the impact of blending, as each postage stamp contained only a single galaxy. As blending tends to bias the measured sizes, it is worthwhile to explore this using our more realistic simulations. The image simulations that we use to study the performance of shear measurements can be used to identify additive biases for magnification, as any change in mean size must be the result of a systematic. However, to quantify the systematics for size magnification studies in more detail, we simulated the impact of pure magnification, including the changes in flux. To do so, we magnified the galaxies in the input catalogue (including the mean separation between the galaxies) by a factor of $1 + \mathcal{M}_{\rm mag} \in [1.0, 1.05, 1.1]$ and analysed these images using our standard pipeline. Our lensing pipeline does not attempt to estimate PSF-corrected sizes, and instead we use the observed half-light radii to examine biases in magnification studies.

We take the simulation with $\mathcal{M}_{\rm mag} = 0$ (i.e. no magnification) and $n_{\rm fac} = 1$ as the reference, and compute

$$\mathcal{M}_{\rm mag}^{\rm obs} = \frac{\langle\tilde{r}_{\rm h}\rangle}{\langle\tilde{r}_{\rm h}\rangle_{\rm ref}} - 1, \tag{5}$$

where $\tilde{r}_{\rm h}$ is the half-light radius from which the mean PSF size was subtracted in quadrature. We found that this simple estimator scales linearly with the input magnification. Analogous to what was done to quantify the biases in the shape measurement algorithm, we define

$$\mathcal{M}_{\rm mag}^{\rm obs} = \mu_{\rm mag}\,\mathcal{M}_{\rm mag}^{\rm true} + c_{\rm mag}, \tag{6}$$

where $\mu_{\rm mag}$ and $c_{\rm mag}$ are the multiplicative and additive bias, respectively. We note that our estimate of the size may be particularly sensitive to blending, and a more detailed study is warranted. None the less, with the simulations, it is possible to highlight some of the challenges for magnification studies.

The points in the bottom panel in Fig. 6 indicate the multiplicative bias $\mu_{\rm mag}$ for galaxies with $20 < m < 24.5$ as a function of $n_{\rm fac}$, the increase in source number density relative to the reference simulation. As was the case for shape measurements, the multiplicative bias is affected by blending. If we use the median size instead in equation (5), we find similar results but with smaller biases; the red triangles in the bottom panel of Fig. 6 are offset by 0.1 to allow for a direct comparison.

The observed change in average size is a combination of the increase in size due to magnification and an increase in the number of intrinsically smaller galaxies due to flux magnification. The latter is quite relevant: If we repeat the analysis by adjusting the magnitude

Downloaded from https://academic.oup.com/mnras/article-abstract/468/3/3295/3089745
by Jacob Heeren user
on 10 January 2018

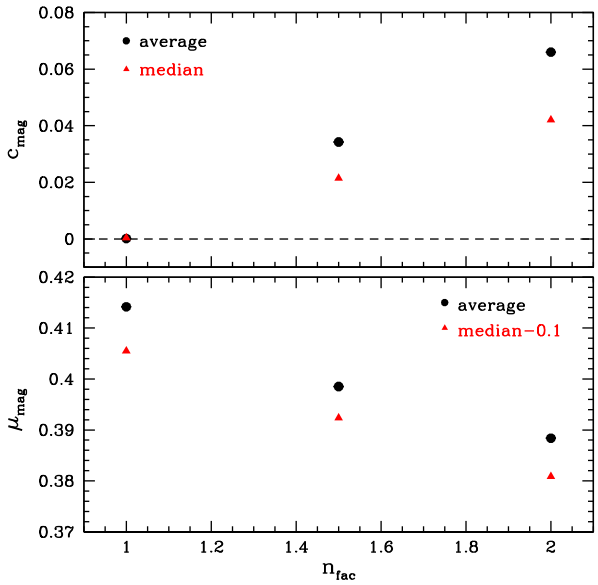

Figure 6. Multiplicative (bottom panel) and additive bias (top panel) for size magnification, as a function of the relative increase in the number density of galaxies brighter than magnitude 26. The measurements are based on the observed half-light radius of galaxies with magnitude $20 < m < 24.5$; the black points use the average size, whereas the red triangles correspond to the results when the median size is used (we subtract 0.1 from μ_{mag} in this case for easier comparison). Part of the multiplicative bias is the result of smaller galaxies entering this magnitude-limited sample due to magnification.

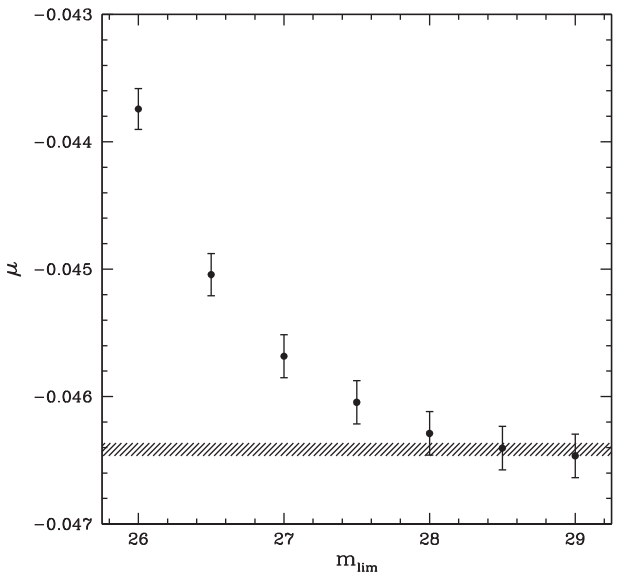

Figure 7. Multiplicative bias for galaxies with $20 < m < 24.5$ when only galaxies with magnitudes brighter than m_{lim} are included in the simulation. Because of the small PSF, even galaxies as faint as $m = 29$ affect the bias. The hatched region indicates a variation in μ of amplitude 10^{-4}.

limits to correct for the change in flux, we find $\mu_{mag} = 0.66$, and 0.81 if we use the median size. The sensitivity to the properties of galaxies below the nominal flux limit is an additional complication for magnification studies, which is essentially absent in the case of shape measurements.

More worrisome are the results presented in the top panel in Fig. 6: The additive bias c_{mag} is a strong function of the number density of bright galaxies. The bias is reduced somewhat if we use the median size, as indicated by the red triangles, and the use of more optimized size estimates may improve things further. However, as the regions of high magnification tend to correspond to regions of increased galaxy density, this substantial additive bias represents a serious complication, especially for cluster studies such as the one presented in Duncan et al. (2016) or Schmidt et al. (2012).

Our results suggest that magnification studies will also be affected by the complexity in the data. In particular, the additive bias that arises from changes in the galaxy density needs to be carefully accounted for. Although this is possible, in principle, the argument that magnification is an attractive complement to cosmic shear because it is subject to different systematics (Alsing et al. 2015) should be reconsidered: Our findings rather suggest that magnification is subject to *additional* systematics.

6 THE IMPACT OF UNDETECTED GALAXIES

The properties of sufficiently bright galaxies can be compared directly to the outputs of the simulations, and remaining trends can be quantified and accounted for, for instance, through machine learning techniques. H15, however, found that galaxies fainter than the limit of the source sample also affect the multiplicative bias of the brighter galaxies. This is partly the result of blending, but also because these galaxies act as a skewed source of background noise, affecting the local background determination. If such faint galaxies are not included in the image simulations, the multiplicative bias can be underestimated by a fair amount: H15 found that the bias doubled for their simulation of ground-based data, and saturated when the simulation included galaxies that were at least 1.5 mag fainter than the limiting magnitude of the source sample.

We therefore examine which value for m_{lim}, the limiting magnitude of the faintest galaxies included in the simulation, may be adequate for image simulations for *Euclid*. The input GEMS catalogue is incomplete for $m > 25.5$ (see Fig. 1), and we augment the catalogue by duplicating the fainter galaxies such that the input counts follow the power-law relation seen at brighter magnitudes, i.e. we adopt a slope of 0.36 over the full magnitude range (indicated by the dashed line in Fig. 1). Compared to the actual counts observed from the HUDF by Coe et al. (2006), our results represent a worst-case situation. We assign sizes following our parametric model described in Appendix A, but explore the sensitivity of the results to the size distribution in Section 6.1.

Fig. 7 shows that the multiplicative bias μ converges rather slowly as a function of m_{lim}. To reduce the number of simulations, we add fainter galaxies to the existing images, such that the bright galaxies are always in common. Hence, the variation between points is somewhat smaller than the error bars, which indicate the statistical uncertainty of a single measurement. The results are based on 81 sets of simulations for each data point. The dashed line indicates a change $|\Delta\mu| = 10^{-4}$, indicating that we may have to include galaxies as faint as $m = 29$ in the *Euclid* image simulations.

The change in bias presented in Fig. 7 is the result of two effects. Galaxies just below the detection limit affect the shape measurements mostly through blending, whereas the very faint galaxies bias the measurements by affecting the determination of the local background. We revisit this topic in Section 8, where we explore different implementations of the background determination.

Downloaded from https://academic.oup.com/mnras/article-abstract/468/3/3295/3089745
by Jacob Heeren user
on 10 January 2018

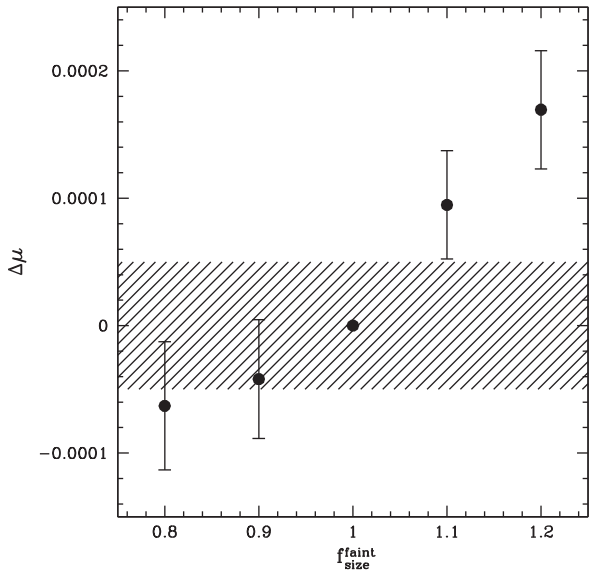

Figure 8. Change in multiplicative bias for galaxies with $20 < m < 24.5$ when the sizes of the input galaxies with $m > 27$ are increased by a factor of $f_{\rm size}^{\rm faint}$ compared to our reference distribution. The hatched region indicates $|\Delta\mu| < 5 \times 10^{-5}$.

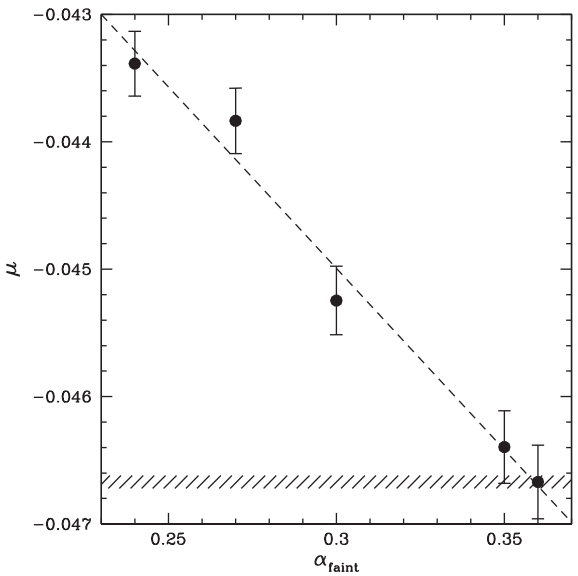

Figure 9. Multiplicative bias for galaxies with $20 < m < 24.5$ as a function of $\alpha_{\rm faint}$, the power-law slope of the galaxies counts fainter than magnitude 24.5; the reference simulation assumes $\alpha_{\rm faint} = 0.36$ for all magnitudes, whereas the HUDF counts suggest a slope of 0.24 for faint magnitudes. The hatched regions indicate a change of 10^{-4} in the estimate of μ.

6.1 Varying size distribution of faint galaxies

The effect of the faint galaxies is to add highly skewed noise to the pixels where they are located. Our parametrized model for the galaxy sizes (see Fig. 2) suggests that the faintest galaxies are small, but have sizes that are not completely negligible compared to the size of the PSF. This is different from the large PSF in ground-based data.

Size distributions for relatively bright galaxies ($m < 26$) can be determined from existing deep *HST* observations, such as the Cosmological Evolution Survey (Scoville et al. 2007) and the All-wavelength Extended Groth Strip International Survey (Davis et al. 2007). For instance, Griffith et al. (2012) present a compilation of photometric and morphological measurements. Reliable size estimates of the fainter galaxies require deeper data, which are only available for relatively small areas, such as the HUDF.

To determine whether the available data are adequate, we examine how well the mean size needs to be determined. To do so, we increase the input half-light radii of galaxies with $m > 27$ by a factor of $f_{\rm size}^{\rm faint}$ and measure the difference $\Delta\mu$ with respect to the reference simulation. To reduce the number of simulations, the positions and intrinsic ellipticities of the galaxies are the same for the different values of $f_{\rm size}^{\rm faint}$.

The results, based on 50 sets of simulations for each value of $f_{\rm size}^{\rm faint}$, are presented in Fig. 8. We find that the multiplicative bias is smaller (corresponding to positive $\Delta\mu$ because $\mu < 0$) when the faint galaxies are larger. This is expected because the galaxies are more spread out and thus introduce noise that is less skewed. If we consider an allocation $|\Delta\mu| < 5 \times 10^{-5}$, these results indicate that the mean sizes of galaxies with $m > 27$ should be determined to better than 5 per cent.

Given the width of the observed size distribution and the number of galaxies for which sizes were determined in the HUDF, we find that the mean sizes of these faint galaxies can, in principle, be constrained to better than 4 per cent. We note, however, that the analysis presented by Coe et al. (2006) will need to be revisited to ensure that biases in the mean sizes are sufficiently small.

6.2 Varying the count slope of faint galaxies

We expect the amplitude of the multiplicative bias to decrease if fewer faint galaxies are present. For our reference model, we adopted a single-power-law slope for the galaxy counts of 0.36 down to magnitude 29. The actual counts from the HUDF from Coe et al. (2006) shown in Fig. 1 suggests that the actual slope is lower; we obtain a best-fitting value of 0.237 ± 0.009 when we fit a power law to the counts of galaxies with $25 < m < 29$. The error on the slope was obtained by splitting the data into four quadrants. To quantify the sensitivity of our results to the count slope of faint galaxies, we simply change the slope of the counts for galaxies with $m > 24.5$ to a value $\alpha_{\rm faint}$.

The multiplicative bias as a function of $\alpha_{\rm faint}$ is presented in Fig. 9. The bias increases linearly with increasing slope. We find a best-fitting $\mathrm{d}\mu/\mathrm{d}\alpha_{\rm faint} = -0.0239 \pm 0.0014$, which suggests that we need to determine the mean slope with a precision of 0.004 if we wish to allocate a maximum uncertainty of $\Delta\mu = 10^{-4}$. This precision can probably not be achieved from the HUDF alone, as we expect the slope to vary due to variations in the distant large-scale structure. The impact of cosmic variance can be reduced by combining with the *Hubble Deep Fields*, the parallel observations from the Frontier Fields (although the clusters may contaminate the counts) as well as future *James Webb Space Telescope* observations. These combined observations should reduce the uncertainties to the required level.

The multiplicative bias will vary locally as a result of fluctuations in the faint galaxy counts. If these are uncorrelated with the lensing signal, the main impact is to slightly increase the noise in the cosmic-shear signal. However, the slope may be affected by gravitational lensing: As discussed in Section 5.3, magnification leads to an increase or decrease in the number density of background galaxies, depending on the slope of the number counts as a function of magnitude. The relatively flat slope we find for faint galaxies would lead to a reduction in the average number counts behind overdense regions. Hence, this introduces a correlation between the large-scale structure that causes the lensing signal and the multiplicative bias.

Downloaded from https://academic.oup.com/mnras/article-abstract/468/3/3295/3089745
by Jacob Heeren user
on 10 January 2018

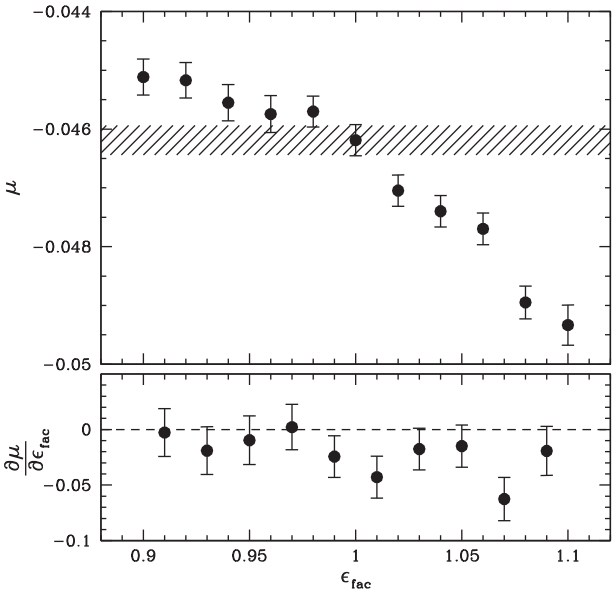

Figure 10. Top panel: multiplicative bias for galaxies with $20 < m < 24.5$ when the input ellipticities are increased by a factor of $\epsilon_{\rm fac}$. The hatched region indicates the allocated uncertainty in the multiplicative bias of 5×10^{-4}. Bottom panel: the slope $\partial\mu/\partial\epsilon_{\rm fac}$ as a function of $\epsilon_{\rm fac}$.

This is worrisome, but a detailed assessment is beyond the scope of this paper. In future work, we will explore the complications that arise from magnification using numerical simulations to create more realistic source samples.

7 VARYING THE INPUT ELLIPTICITY DISTRIBUTION

The response of a galaxy to an applied shear depends on its ellipticity: The change in shape is larger for intrinsically round objects compared to those that are intrinsically elliptical. As the shear is computed as an ensemble average of the ellipticity of a population of galaxies, the average shear responsivity depends on the ellipticity distribution. Moreover, as discussed in detail in Viola et al. (2014), the bias due to pixel noise also depends on the ellipticity distribution.

In this section, we examine the sensitivity of the multiplicative bias to the input ellipticity distribution. H15 found that the multiplicative bias for their implementation of the KSB algorithm was relatively insensitive to the value of ϵ_0 because of the use of a radial weight function (i.e. no attempt is made to optimize the weight function and match it to the galaxy shape). None the less, the recovered bias varied by 0.01 over the large range in ϵ_0 considered. Given the much more challenging constraints for *Euclid*, it is interesting to quantify how the bias depends on the ellipticity distribution used in the image simulations.

The input ellipticity distribution can be modified in a number of ways, but here we simply multiply the input ellipticity by a factor of $\epsilon_{\rm fac}$. The resulting multiplicative bias for galaxies with $20 < m < 24.5$ as a function of $\epsilon_{\rm fac}$ is presented in the top panel of Fig. 10. The hatched region indicates an uncertainty of 5×10^{-4} in the multiplicative bias due to uncertainties in the ellipticity distribution.

The bottom panel of Fig. 10 shows the slope $\partial\mu/\partial\epsilon_{\rm fac}$ as a function of $\epsilon_{\rm fac}$, which reaches a maximum absolute value of 0.05; for an allocation of $\Delta\mu = 5 \times 10^{-4}$, this implies that $\epsilon_{\rm fac}$ needs to be known with a precision of 1 per cent. This is comparable to the precision of 0.3 per cent found by Viola et al. (2014) for galaxies with S/N = 10.

Table 1. Multiplicative bias for $m_{\rm lim} = 29$ for different implementations.

Implementation	μ
Reference	$-0.046\,45 \pm 0.000\,17$
Unflagged objects	$-0.041\,46 \pm 0.000\,17$
Unmasked neighbouring objects	$-0.048\,25 \pm 0.000\,20$
Aperture of $3r_{\rm g}^{\rm ref}$	$-0.079\,72 \pm 0.000\,16$
$r_{\rm g} = 1.5r_{\rm rg}^{\rm ref}$	$-0.038\,73 \pm 0.000\,21$
$r_{\rm g} = 2r_{\rm rg}^{\rm ref}$	$-0.081\,72 \pm 0.000\,23$
$32 < r_{\rm bg} < 64$ pixels	$-0.042\,67 \pm 0.000\,17$
$64 < r_{\rm bg} < 128$ pixels	$-0.042\,46 \pm 0.000\,17$
Zero background	$-0.037\,79 \pm 0.000\,26$
SWARP background-subtracted	$-0.038\,58 \pm 0.000\,31$

Column 1: modified implementation as explained in the text; column 2: value for the multiplicative bias for galaxies with $20 < m < 24.5$ when galaxies down to $m_{\rm lim} = 29$ are included in the simulation.

Similar to the case for the observed size distribution, a mismatch between the data and the simulations can be readily identified by comparing the distributions. A simple comparison of the observed dispersion $\langle\epsilon_{\rm obs}^2\rangle^{1/2}$ for a subset of the simulations was sufficient to constrain $\epsilon_{\rm fac}$ to better than 0.5 per cent. A complication for real data is that variations in the noise level will lead to varying broadening of the observed distributions. Instead, Viola et al. (2014) advocate observing a small area of sky to a depth such that the impact of pixel noise can be neglected. In the case of *Euclid*, an area of 40 deg^2 observed with 40 times the nominal integration time is sufficient to achieve this. However, further study is needed, as Viola et al. (2014) do not account for the blending with fainter galaxies, which may act as an additional source of noise.

8 SENSITIVITY TO IMPLEMENTATION

Until now, we have focused on the sensitivity of the shape measurement algorithm to changes in the input images. Although the implementation of a simple moment-based method such as KSB is relatively straightforward, a number of choices are made, such as the values of the SExtractor deblending parameters, the width of the weight function, the background determination, etc. It is therefore useful to explore how changes in the actual implementation of a method affect the resulting multiplicative bias. This may help identify parameters that can improve the robustness. Table 1 lists the multiplicative bias values for simulations with $m_{\rm lim} = 29$ for the various implementation changes that we discuss in more detail below.

In our reference analysis, we consider all galaxies detected by SExtractor, which also provides a flag to indicate potential problems with an object. We expect that restricting the analysis to unflagged objects (flag = 0) reduces the bias, which is indeed the case. In contrast, when we do not mask the neighbouring galaxies, the bias is increased somewhat.

When we reduce the size of the postage stamp to $3r_{\rm g}^{\rm ref}$, the bias increases significantly because the small radial extent effectively truncates the estimates of the higher order moments. The results presented in Table 1 indicate that the biggest improvement is achieved by increasing the width of the weight function to 1.5 times the nominal value for $r_{\rm g}^{\rm ref}$. However, increasing the width further rapidly

Downloaded from https://academic.oup.com/mnras/article-abstract/468/3/3295/3089745
by Jacob Heeren user
on 10 January 2018

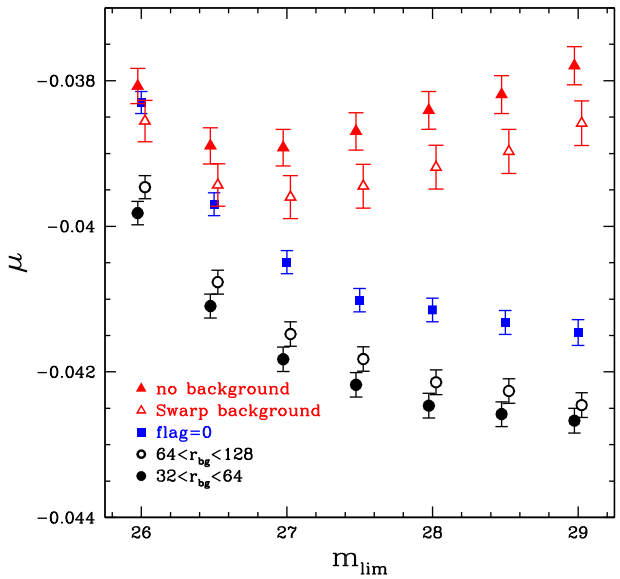

Figure 11. Multiplicative bias for galaxies with $20 < m < 24.5$ as a function of $m_{\rm lim}$, the magnitude of the faintest galaxies that are included in the simulation, for different implementations as described in the text. In all cases, faint galaxies affect the measured multiplicative bias.

increases the bias. Although most of the changes are small, they are none the less comparable or larger than the desired error budget of $\delta\mu = 10^{-4}$.

As discussed in Section 6, faint galaxies affect the local background determination. The reference analysis determines the background by fitting a plane to the quadrants of an annulus with radii in the range $16 < r_{\rm bg} < 32$ pixels. Increasing the area of this annulus should make the background estimate more robust. Indeed, we find that the bias is reduced for larger annuli, and find similar values for both cases. However, our simulated data have a constant background with an expectation value that vanishes. Hence, we can also consider the case where the background is set to 0, its true value. This yields the smallest bias of the changes considered here. To mimic a more realistic scenario, we also run SWARP[10] with `BACK_SIZE = 128` and `BACK_FILTERSIZE = 3` to determine the background for each image. This results in a slight increase of the bias, but it does perform better than our local background estimates.

Modifying the background determination can thus have a significant impact on the bias. If the sensitivity to the faint galaxies could be reduced, then this represents an interesting avenue to improve the robustness of the shape measurements against variations in the properties of galaxies that are too faint to be detected in the survey data. It is therefore interesting to repeat the analysis of the bias as a function of $m_{\rm lim}$ for different background determinations.

The results for galaxies with $20 < m < 24.5$ are presented in Fig. 11. The solid and open points indicate the results when we use larger annuli for the local background determination. The dependence on $m_{\rm lim}$ is, however, similar to that of the reference set-up. We also show results when we reduce the impact of blending by selecting only galaxies for which no SExtractor flag was raised (blue squares). The bias is indeed smaller, but it increases for larger values of $m_{\rm lim}$.

The behaviour is markedly different when we assume that we know the background level a priori (filled red triangles), which does yield the smallest bias: After initially increasing in amplitude, the multiplicative bias decreases with increasing $m_{\rm lim}$. This is also observed when we use SWARP to determine the background in a more realistic way (open red triangles). Compared to the results for the local background determination, the dependence of the bias on $m_{\rm lim}$ is in fact stronger.

To examine the origin of these different dependences, we compared the half-light radii and fluxes measured for galaxies with $20 < m < 24.5$ in simulations with $m_{\rm lim} = 26$ to measurements obtained using $m_{\rm lim} = 29$, where we matched the two catalogues. In the case of local background determination, the inclusion of faint galaxies results in a small decrease of only 0.2 per cent in the half-light radius, whereas the flux decreases by 0.3 per cent (both results are median values). Moreover, the distributions of the differences are fairly symmetric. In contrast, fixing the background to zero or using SWARP to subtract a smooth background results in a highly skewed distribution, where the measurements for most objects are unchanged. In this case, the presence of faint galaxies leads to a median increase of 2.2 per cent in size and a median increase of 3.6 per cent in flux. This is not surprising, as the faint objects can increase only the measured sizes and fluxes if the background is fixed. Hence, this is a generic result, although the details will depend on the shape measurement method.

The sensitivity to faint galaxies depends clearly on the way the background is determined. Interestingly, a global background determination may actually lead to an increased sensitivity to undetected galaxies, although we note that the number density of faint galaxies in our simulations is higher than observed in deep *HST* data (see Fig. 1). Although further study is warranted, we do advocate the use of local background determination because it also avoids coherent biases in shapes measurements that may arise from errors in the global background.

9 THE IMPACT OF STARS

Actual data will also contain stars, which may enter the galaxy catalogue or affect the shapes of galaxies through blending. H15 showed that stars contribute to multiplicative bias for ground-based observations at low Galactic latitude (see their fig. A2). Bright stars can be identified morphologically because the galaxy images are resolved, but as faint objects contribute to multiplicative bias, we are most concerned about the impact of faint stars.

To quantify the impact of blending by stars for *Euclid* observations, we need predictions for the number density of faint stars as a function of location on the sky. This requires a population synthesis model of the Milky Way that can predict the properties of the various stellar populations that make up the Galaxy. Here we consider two well-known models for the distribution of stars in the Milky Way, the Besançon model[11] (Robin et al. 2003) and v1.6 of the TRI-dimensional modeL of thE GALaxy[12] (TRILEGAL; Girardi et al. 2005). Both provide web-based interfaces that return simulated catalogues.

The synthesis models use theoretical stellar evolution tracks, prescriptions for the initial mass function (IMF), star formation history (SFH), age–metallicity relation, extinction and geometries for the Galactic components, and a stellar spectral library to predict the photometry at a given location on the sky. The models consider separate spatial distributions and SFHs for the thin and thick discs,

[10] http://www.astromatic.net/software/swarp

[11] http://model.obs-besancon.fr/

[12] http://stev.oapd.inaf.it/cgi-bin/trilegal

Downloaded from https://academic.oup.com/mnras/article-abstract/468/3/3295/3089745
by Jacob Heeren user
on 10 January 2018

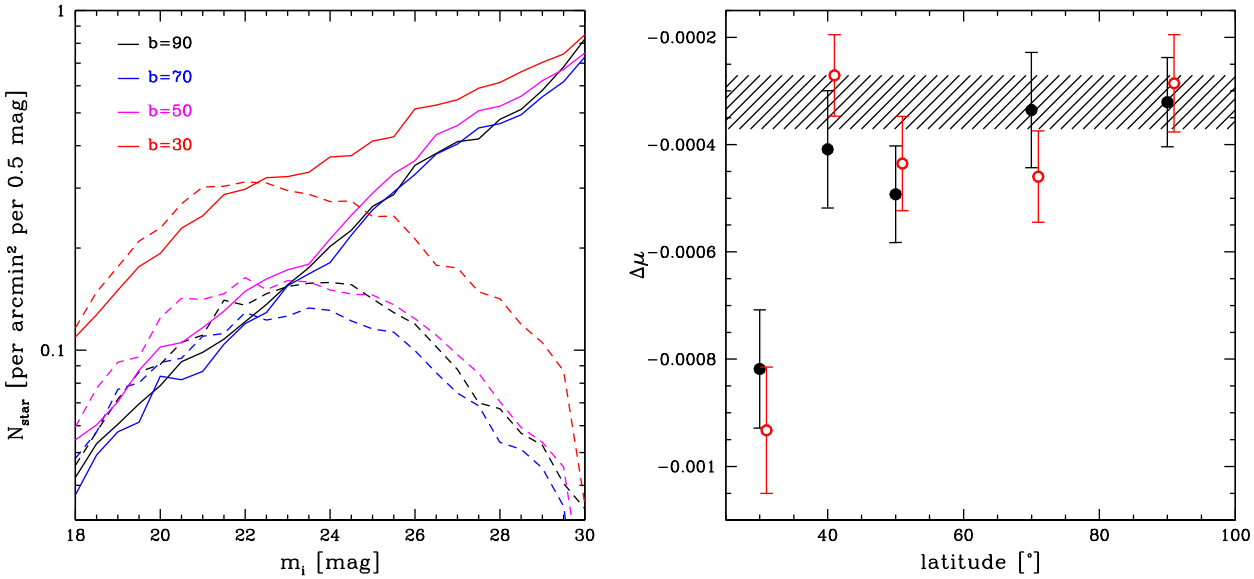

Figure 12. Left-hand panel: star counts as a function of magnitude for a Galactic longitude $l = 180°$ and latitudes b of 90°, 70°, 50° and 30° (black, blue, magenta and red lines, respectively). The solid lines indicate the results from the Besançon model (Robin et al. 2003), whereas the dashed lines correspond to the results from the TRILEGAL model (Girardi et al. 2005). For magnitude $m < 23$, the agreement between the models is fair, but we observe large differences for the faint counts. Right-hand panel: the change in multiplicative bias $\Delta\mu$ for galaxies with $20 < m < 24.5$, with respect to the reference simulations, which contain only a small number of stars for PSF modelling. The black (solid) points correspond to the TRILEGAL counts as a function of latitude for $l = 180°$, whereas the red (open) points indicate the results when the Besançon model star counts are used instead. The differences in the faint star counts do not appear to be important.

the bulge and the stellar halo, and the parameters are optimized to match certain observations, such as star counts, colours or kinematic data.

We note, however, that the assumption of a smooth stellar halo may be inadequate, as it is well known that the stellar halo contains substantial substructure as a result of minor mergers, with the Sagittarius stream (Ibata, Gilmore & Irwin 1994) being the most prominent. Although the stream can be clearly identified out to large distances using multicolour observations (e.g. Pila-Díez et al. 2015), it is less prominent in the counts themselves (see e.g. fig. 4 of Pila-Díez et al. 2014).

The Besançon and TRILEGAL models provide a good match to observed star counts at bright magnitudes. For instance, Gao, Just & Grebel (2013) present a comparison to SDSS star counts at the North Galactic Pole. At fainter magnitudes, the situation is less clear. Deep *HST* observations provide some constraints (e.g. Pirzkal et al. 2005; Stanway et al. 2008; Pirzkal et al. 2009), but also for these data, contamination by galaxies limits the selection at the faintest magnitudes, although proper motions can be used to improve the selection (e.g. Kilic et al. 2005).

The Besançon model parameters are fixed, and we consider counts in the Megacam i' filter, including stars with a maximum distance of 300 kpc. The results as a function of longitude b (for a Galactic latitude of $l = 180°$) are presented in Fig. 12 by the solid lines. The Besançon model includes a population of low-luminosity white dwarfs (see section 2.5.3 in Robin et al. 2003) that dominate the faint counts. Although these may be confused with blue extragalactic sources, we note that a comparison with Fig. 1 shows that faint galaxies outnumber the stars by two orders of magnitude.

The TRILEGAL interface allows the user to change the model parameters, such as the IMF or binary fraction, but also the geometry and SFHs of the bulge, thin disc, thick disc and stellar halo components. We use the default settings and the resulting counts as a function of Megacam i' magnitude are indicated by the dashed lines in Fig. 12. For bright stars ($m < 23$), the agreement with the Besançon predictions is fair; the large differences for the faint counts are caused by the large number of distant halo white dwarfs in the Besançon model. In this regard, we can consider the Besançon model as a worst-case scenario for the impact of stars on shape measurements.

The right-hand panel of Fig. 12 shows the resulting change in multiplicative bias (relative to the reference simulations that do not include many stars) as a function of Galactic latitude (for a Galactic longitude $l = 180°$). As for the reference case, we select galaxies requiring that the observed half-light radius $r_h > 0.11$ arcsec because the galaxies with $m < 24.5$ are larger than the stars. The black (solid) points indicate the results for the TRILEGAL star counts, whereas the red (open) points correspond to the results when using the Besançon predictions. We observe a small increase in multiplicative bias of 3×10^{-4}, which increases to 10^{-3} for low Galactic latitudes. Importantly, the biases for the two models are consistent, suggesting that the differences in faint star counts have a negligible impact.

As the number density of stars varies across the sky, so will the corresponding change in multiplicative bias. One would like to avoid having to create large suites of image simulations to quantify the impact of stars in the data. We therefore explored whether it is possible to predict the change in multiplicative bias using the observed counts of bright stars. To do so, we used TRILEGAL star counts for a range of Galactic longitudes and latitudes, and determined the resulting multiplicative bias as a function of $n_{\rm star}$, the observed number density of stars within the magnitude range $18.5 < m < 24$.

Fig. 13 shows that the bias increases linearly with $n_{\rm star}$. The residuals with respect to the best-fitting linear relation are presented in the bottom panel; with an rms of 10^{-4}, these results suggest that the bias can be predicted sufficiently well using the observed number density of stars. Note that we did not attempt to optimize the magnitude range for the bright star counts, but further refinement is limited in any case by the relative small number of simulations we created (20 sets per position).

Downloaded from https://academic.oup.com/mnras/article-abstract/468/3/3295/3089745
by Jacob Heeren user
on 10 January 2018

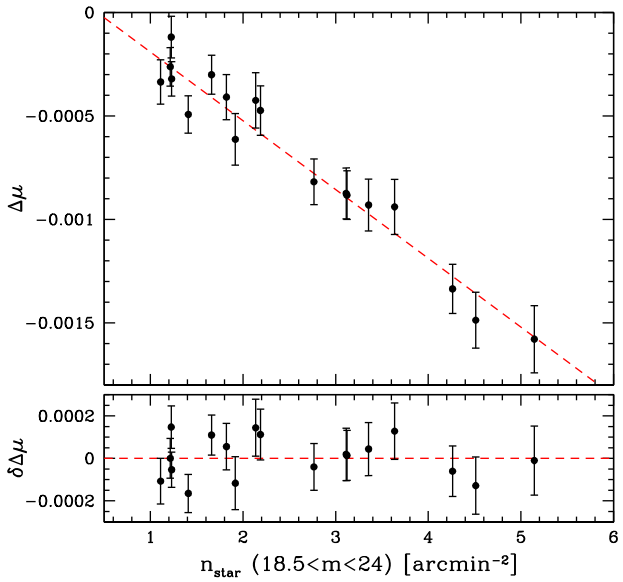

Figure 13. Top panel: change in multiplicative bias, relative to the reference simulations (which contain only a small number of stars), for galaxies with $20 < m < 24.5$ as a function of the observed density of bright stars ($18.5 < m < 24$). The red dashed line is the best-fitting linear relation. Bottom panel: the residuals $\delta\Delta\mu$ with respect to this fit, indicating that the bias can be predicted with a precision of 10^{-4} (rms).

10 CONCLUSIONS

To exploit the statistical power of future cosmic-shear surveys, the accuracy with which shapes can be measured needs to be improved. This requires a careful modelling of the PSF, which is the dominant source of bias, even for space-based observations. However, the presence of noise and neighbouring galaxies affects the accuracy of shape measurement algorithms.

The performance of shape measurement algorithms can, however, be quantified using simulated data. Moreover, the resulting biases as a function of observed properties can be used to derive an empirical calibration of the method (e.g. H15; Fenech Conti et al. 2017). For a meaningful calibration, it is essential that the simulated data are sufficiently realistic, whereas many studies have focused on idealized cases, which not only yield incorrect bias estimates, but can also lead to the development of algorithms that may not perform well on real data.

We explored the sensitivity of the multiplicative bias of a simple moment-based method to a number of real-life effects, such as noise, the presence of stars and galaxies below the detection limit. We also examined how well the sizes and intrinsic ellipticities of the galaxies need to be known. Moreover, we demonstrated how slight modifications in the algorithm can significantly change the recovered biases. Most of the changes are well below the statistical uncertainties of ongoing surveys, but need to be studied for the next generation of surveys that are an order of magnitude larger. Hence, although we focused here on the accuracy with which *Euclid* will need to determine galaxy shapes, our findings are also of relevance for LSST and *WFIRST*.

We used the publicly available code GALSIM (Rowe et al. 2015) to create the simulated data. The images are analysed with the KSB algorithm (Kaiser et al. 1995), which is fast but not well suited for the analysis of real data. However, the objective of this paper is not to calibrate this algorithm, but rather to demonstrate that shape measurements are sensitive to a number of real-life effects that have not been considered before. Although the actual sensitivities will differ between algorithms and implementations, our results are generic.

As expected, the multiplicative bias depends on the noise level in the images. The noise level will vary across the survey, in particular, due to changes in the zodiacal light, but we find that this can be measured with sufficient precision from the *Euclid* data. The bias is also a strong function of galaxy size, and changes in the input sizes thus modify the results. This can be accounted for by adjusting the input parameters by comparing the simulated data to the observations (e.g. Bruderer et al. 2016; Fenech Conti et al. 2017), or by capturing the dependence on size through an empirical correction. It is, however, important that selection biases are also accounted for (Fenech Conti et al. 2017).

The multiplicative bias also depends on the galaxy number density, which is not captured in simulations of individual galaxies. As the cosmic-shear signal is correlated with the galaxy density, it is important that shape measurement algorithms account for this. For instance, one could consider algorithms that are minimally sensitive to neighbouring objects. We find that adjustments in the implementation can indeed reduce the biases somewhat. Compared to fitting algorithms, moment-based methods are more sensitive to blends because they bias the estimates of the moments. Whether this can be reduced will require further study. We also note that strict criteria may result in undesirable reductions in source density, but can also complicate the interpretation of the signal. The number density of stars also varies across the survey, but we find that the observed star counts can be used to empirically correct the bias with adequate accuracy. Blending also affects the accuracy of measurements of the change in galaxy size due to magnification. This results in additive biases for size magnification studies that will need to be accounted for.

The importance of galaxies below the detection limit was already highlighted in H15. Given the accuracy required for *Euclid*, we find that we need to include galaxies as faint as magnitude 29. The bias is sensitive to the number density of these faint galaxies and their size distribution. Magnification of faint galaxies by the intervening large-scale structure may introduce a correlation between the lensing signal and the multiplicative bias. This requires further study, but we note that our results may be considered a worst case because the observed number density of faint galaxies is much lower than we simulate (see Fig. 1). Minimizing the sensitivity to these faint galaxies is none the less an important topic for further study.

The shape measurement also depends on the way the background level is determined. Our reference implementation uses a local measurement, which has the advantage that errors in the background due to artefacts are confined to small scales. In this case, the multiplicative bias increases as fainter galaxies are included in the simulations. We observe the opposite, but steeper, trend when we consider a global background determination.

We considered a number of complications that occur in the analysis of real data, and demonstrated that these affect the accuracy of shape measurements at a level that is relevant for future surveys such as *Euclid*. These should be considered in the selection of shape measurement algorithms. Improving the accuracy of the input parameters of the image simulations is perhaps the most important step. Although our results show that achieving sub-per cent accuracy is challenging, we estimate that existing *HST* observations, as well as future *JWST* data, are, in principle, adequate to create sufficiently realistic populations of galaxies.

Downloaded from https://academic.oup.com/mnras/article-abstract/468/3/3295/3089745
by Jacob Heeren user
on 10 January 2018

ACKNOWLEDGEMENTS

We thank the members of the *Euclid* weak-lensing science working group, in particular Mark Cropper, Tom Kitching, Lance Miller and Tim Schrabback, for useful discussions. HH also acknowledges fruitful discussions during an international team meeting on 'Cosmology with size and flux magnification' led by Alan Heavens and Hendrik Hildebrandt at the International Space Science Institute (ISSI). HH, RH and MV acknowledge support from the European Research Council FP7 grant number 279396. MV is supported by the Netherlands Organisation for Scientific Research (NWO) through grant 614.001.103.

REFERENCES

Alsing J., Kirk D., Heavens A., Jaffe A. H., 2015, MNRAS, 452, 1202
Amendola L. et al., 2013, Living Rev. Relativ., 16, 6
Bernstein G. M., 2010, MNRAS, 406, 2793
Bernstein G. M., Jarvis M., 2002, AJ, 123, 583
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Bridle S. et al., 2010, MNRAS, 405, 2044
Bruderer C., Chang C., Refregier A., Amara A., Bergé J., Gamper L., 2016, ApJ, 817, 25
Casaponsa B., Heavens A. F., Kitching T. D., Miller L., Barreiro R. B., Martínez-González E., 2013, MNRAS, 430, 2844
Coe D., Benítez N., Sánchez S. F., Jee M., Bouwens R., Ford H., 2006, AJ, 132, 926
Cropper M. et al., 2013, MNRAS, 431, 3103
Davis M. et al., 2007, ApJ, 660, L1
Dawson W. A., Schneider M. D., Tyson J. A., Jee M. J., 2016, ApJ, 816, 11
Duncan C. A. J., Heymans C., Heavens A. F., Joachimi B., 2016, MNRAS, 457, 764
Erben T., Van Waerbeke L., Bertin E., Mellier Y., Schneider P., 2001, A&A, 366, 717
Fenech Conti I., Herbonnet R., Hoekstra H., Merten J., Miller L., Viola M., 2017, MNRAS, 467, 1627
Ford J., Hildebrandt H., Van Waerbeke L., Erben T., Laigle C., Milkeraitis M., Morrison C. B., 2014, MNRAS, 439, 3755
Gao S., Just A., Grebel E. K., 2013, A&A, 549, A20
Girardi L., Groenewegen M. A. T., Hatziminaoglou E., da Costa L., 2005, A&A, 436, 895
Griffith R. L. et al., 2012, ApJS, 200, 9
Gruen D., Seitz S., Koppenhoefer J., Riffeser A., 2010, ApJ, 720, 639
Hartlap J., Hilbert S., Schneider P., Hildebrandt H., 2011, A&A, 528, A51
Heymans C. et al., 2006, MNRAS, 368, 1323
Heymans C. et al., 2012, MNRAS, 427, 146
Heymans C. et al., 2013, MNRAS, 432, 2433
Hildebrandt H., 2016, MNRAS, 455, 3943
Hildebrandt H., van Waerbeke L., Erben T., 2009, A&A, 507, 683
Hildebrandt H. et al., 2011, ApJ, 733, L30
Hildebrandt H. et al., 2017, MNRAS, 465, 1454
Hirata C., Seljak U., 2003, MNRAS, 343, 459
Hoekstra H., Jain B., 2008, Annu. Rev. Nucl. Part. Sci., 58, 99
Hoekstra H., Franx M., Kuijken K., Squires G., 1998, ApJ, 504, 636
Hoekstra H., Franx M., Kuijken K., 2000, ApJ, 532, 88
Hoekstra H., Yee H. K. C., Gladders M. D., Barrientos L. F., Hall P. B., Infante L., 2002, ApJ, 572, 55
Hoekstra H., Herbonnet R., Muzzin A., Babul A., Mahdavi A., Viola M., Cacciato M., 2015, MNRAS, 449, 685 (H15)
Huff E., Mandelbaum R., 2017, ApJ, preprint (arXiv:1702.02600)
Ibata R. A., Gilmore G., Irwin M. J., 1994, Nature, 370, 194
Jarvis M. et al., 2016, MNRAS, 460, 2245
Jee M. J., Tyson J. A., Hilbert S., Schneider M. D., Schmidt S., Wittman D., 2016, ApJ, 824, 77
Joachimi B. et al., 2015, Space Sci. Rev., 193, 1
Kacprzak T., Zuntz J., Rowe B., Bridle S., Refregier A., Amara A., Voigt L., Hirsch M., 2012, MNRAS, 427, 2711
Kacprzak T., Bridle S., Rowe B., Voigt L., Zuntz J., Hirsch M., MacCrann N., 2014, MNRAS, 441, 2528
Kaiser N., 2000, ApJ, 537, 555
Kaiser N., Squires G., 1993, ApJ, 404, 441
Kaiser N., Squires G., Broadhurst T., 1995, ApJ, 449, 460
Kannawadi A., Mandelbaum R., Lackner C., 2015, MNRAS, 449, 3597
Kilbinger M., 2015, Rep. Prog. Phys., 78, 086901
Kilic M., Mendez R. A., von Hippel T., Winget D. E., 2005, ApJ, 633, 1126
Kitching T. D. et al., 2012, MNRAS, 423, 3163
Kitching T. D., Taylor A. N., Cropper M., Hoekstra H., Hood R. K. E., Massey R., Niemi S., 2016, MNRAS, 455, 3319
Köhlinger F., Hoekstra H., Eriksen M., 2015, MNRAS, 453, 3107
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Liu J., Ortiz-Vazquez A., Hill J. C., 2016, Phys. Rev. D, 93, 103508
LSST Science Collaboration et al., 2009, preprint (arXiv:0912.0201)
Luppino G. A., Kaiser N., 1997, ApJ, 475, 20
Mandelbaum R. et al., 2014, ApJS, 212, 5
Mandelbaum R. et al., 2015, MNRAS, 450, 2963
Massey R. et al., 2007, MNRAS, 376, 13
Massey R. et al., 2013, MNRAS, 429, 661
Melchior P., Viola M., 2012, MNRAS, 424, 2757
Melchior P., Viola M., Schäfer B. M., Bartelmann M., 2011, MNRAS, 412, 1552
Miller L. et al., 2013, MNRAS, 429, 2858
Okura Y., Futamase T., 2011, ApJ, 730, 9
Paulin-Henriksson S., Amara A., Voigt L., Refregier A., Bridle S. L., 2008, A&A, 484, 67
Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2002, AJ, 124, 266
Pila-Díez B., Kuijken K., de Jong J. T. A., Hoekstra H., van der Burg R. F. J., 2014, A&A, 564, A18
Pila-Díez B., de Jong J. T. A., Kuijken K., van der Burg R. F. J., Hoekstra H., 2015, A&A, 579, A38
Pirzkal N. et al., 2005, ApJ, 622, 319
Pirzkal N. et al., 2009, ApJ, 695, 1591
Planck Collaboration XIII, 2016, A&A, 594, A13
Refregier A., Bacon D., 2003, MNRAS, 338, 48
Refregier A., Kacprzak T., Amara A., Bridle S., Rowe B., 2012, MNRAS, 425, 1951
Rix H.-W. et al., 2004, ApJS, 152, 163
Robin A. C., Reylé C., Derrière S., Picaud S., 2003, A&A, 409, 523
Rowe B. T. P. et al., 2015, Astron. Comput., 10, 121
Sartoris B. et al., 2016, MNRAS, 459, 1764
Schaan E., Krause E., Eifler T., Doré O., Miyatake H., Rhodes J., Spergel D. N., 2016, preprint (arXiv:1607.01761)
Schmidt F., Leauthaud A., Massey R., Rhodes J., George M. R., Koekemoer A. M., Finoguenov A., Tanaka M., 2012, ApJ, 744, L22
Scoville N. et al., 2007, ApJS, 172, 1
Semboloni E. et al., 2013, MNRAS, 432, 2385
Sheldon E. S., Huff E. M., 2017, preprint (arXiv:1702.02601)
Spergel D. et al., 2015, preprint (arXiv:1503.03757)
Stanway E. R., Bremer M. N., Lehnert M. D., Eldridge J. J., 2008, MNRAS, 384, 348
Tewes M., Cantale N., Courbin F., Kitching T., Meylan G., 2012, A&A, 544, A8
Troxel M. A., Ishak M., 2015, Phys. Rep., 558, 1
Umetsu K., Zitrin A., Gruen D., Merten J., Donahue M., Postman M., 2016, ApJ, 821, 116
van Uitert E., Hoekstra H., Velander M., Gilbank D. G., Gladders M. D., Yee H. K. C., 2011, A&A, 534, A14
Velander M. et al., 2014, MNRAS, 437, 2111
Viola M., Kitching T. D., Joachimi B., 2014, MNRAS, 439, 1909

APPENDIX: PARAMETRIC MODEL OF GALAXY SIZES

To describe the size distribution of faint galaxies, we extrapolate the observed distribution of half-light radii from GEMS (Rix et al. 2004)

Downloaded from https://academic.oup.com/mnras/article-abstract/468/3/3295/3089745
by Jacob Heeren user
on 10 January 2018

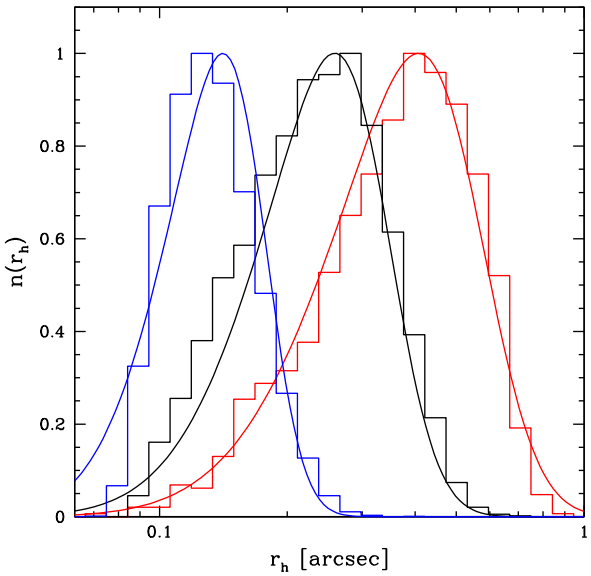

Figure A1. Histograms of the distribution of half-light radii from the GEMS catalogue (Rix et al. 2004) for three magnitude bins (with a width of 0.5 mag). From the left- to right-hand side, the central values are 26.5, 24.5 and 23.0 (blue, black and red, respectively). The smooth curves show the parametric model, which match the data fairly well.

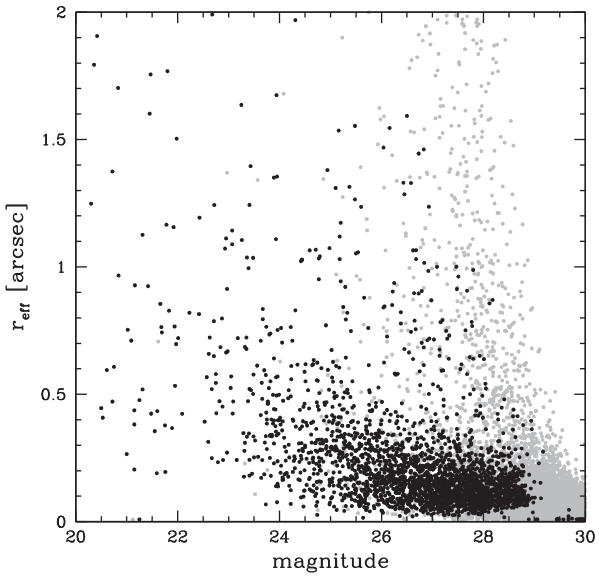

Figure A2. Plot of the best-fitting effective radius as a function of apparent magnitude from Coe et al. (2006). The black points indicate the points with a fractional uncertainty <10 per cent in the size, whereas the grey points correspond to the remaining galaxies in the catalogue.

to fainter magnitudes. In this appendix, we describe how we determined the model parameters.

The observed distribution of half-light radii is well described by a lognormal distribution, as shown in Fig. A1. The histograms show the data for three bins with a width of 0.5 mag. From the left- to right-hand side, the central values are 26.5, 24.5 and 23.0 (blue, black and red, respectively), clearly showing that the mean galaxy sizes decrease with magnitude. We note that the faintest bin in Fig. A1 is shown for reference because the GEMS catalogue is incomplete for this bin.

Instead, we derive the model parameters using bins with central values between magnitude 23 and 25.5, and assume that we can extrapolate the results to fainter magnitudes. We initially kept the skewness as a free parameter, but found that the most robust results were obtained by fixing it to a representative value of -0.58. In this magnitude range, the mean size and dispersion decrease approximately linearly with apparent magnitude m (see Fig. 2), and we obtain best fits $\langle \log_{10} r_{\rm h} \rangle = (3.086–0.145) \times m$ (with $r_{\rm h}$ in units of arcseconds) and rms $\sigma_{\log_{10} r_{\rm h}} = (0.892–0.0266) \times m$. These relations were used to compute the dashed lines in Fig. 2, which indeed match the GEMS data well. The smooth curves in Fig. A1 are the model size distributions, which match the histograms well, including the magnitude 26.5 bin, which is an extrapolation using the model. We use this simple parametric model to assign sizes to galaxies with $m > 26.5$ in our image simulations.

To compare our model to deep *HST* observations, we use measurements from the HUDF by Coe et al. (2006). Similar to Rix et al. (2004), they fit the galaxy brightness distribution using GALFIT. Fig. A2 shows the resulting best-fitting effective radii as a function of apparent magnitude. The scatter increases rapidly for galaxies with $27.5 < m < 29$, most likely due to degeneracies with other fit parameters. If we require a relative uncertainty <10 per cent in $r_{\rm eff}$, the scatter is reduced considerably, as indicated by the black points in Fig. A2.

Unfortunately, Coe et al. (2006) report only the best-fitting effective radii, whereas we use the more robust half-light radius. Instead, we assume a linear relation between $r_{\rm eff}$ and $r_{\rm h}$, and determine the parameters from the GEMS catalogue, which contains both size estimates (Rix et al. 2004): $r_{\rm h} \approx (0.066 + 0.46) \times r_{\rm eff}$. The resulting average half-light radii are indicated by the red points in Fig. 2, which indicate that our parametric model may underestimate the sizes for the faintest galaxies. Similarly, we adjust the dispersions; the red points in the bottom panel of Fig. 2 agree fairly well, although the width of the size distribution may be larger than adopted.

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

Downloaded from https://academic.oup.com/mnras/article-abstract/468/3/3295/3089745
by Jacob Heeren user
on 10 January 2018