Mon. Not. R. Astron. Soc. **368,** 1323–1339 (2006) doi:10.1111/j.1365-2966.2006.10198.x

The Shear Testing Programme – I. Weak lensing analysis of simulated ground-based observations

Catherine Heymans,[1⋆] Ludovic Van Waerbeke,[2] David Bacon,[3] Joel Berge,[4] Gary Bernstein,[5] Emmanuel Bertin,[6] Sarah Bridle,[7] Michael L. Brown,[3] Douglas Clowe,[8] Håkon Dahle,[9] Thomas Erben,[10] Meghan Gray,[11] Marco Hetterscheidt,[10] Henk Hoekstra,[12] Patrick Hudelot,[13] Mike Jarvis,[5] Konrad Kuijken,[14] Vera Margoniner,[15] Richard Massey,[16] Yannick Mellier,[6,17] Reiko Nakajima,[5] Alexandre Refregier,[4] Jason Rhodes,[18] Tim Schrabback[10] and David Wittman[15]

[1]*Max-Planck-Institut für Astronomie, Königstuhl, 69117 Heidelberg, Germany*
[2]*University of British Columbia, 6224 Agricultural Rd, Vancouver, BC V6T 1Z1, Canada*
[3]*Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ*
[4] *Service d'Astrophysique, CEA Saclay, F-91191 Gif sur Yvette, France*
[5]*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*
[6]*Institut d'Astrophysique de Paris, UMR7095 CNRS, Université Pierre & Marie Curie, 98 bis boulevard Arago, 75014 Paris, France*
[7]*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT*
[8]*Steward Observatory, University of Arizona, 933 N. Cherry Ave, Tuscon, AZ 85721, USA*
[9]*Institute of Theoretical Astrophysics, University of Oslo, PO Box 1029, Blindern, N-0315 Oslo, Norway*
[10]*Institut für Astrophysik und Extraterrestrische Forschung, Universität Bonn, Auf dem Hügel 71, 53121 Bonn, Germany*
[11]*School of Physics and Astronomy, University of Nottingham, Nottingham NG7 2RD*
[12]*University of Victoria, Elliott Building, 3800 Finnerty Rd, Victoria, BC V8P 5C2, Canada*
[13]*Observatoire Midi-Pyrénées, UMR5572, 14 Avenue Edouard Belin, 31000 Toulouse, France*
[14]*Leiden Observatory, PO Box 9513, NL-2300 RA, Leiden, the Netherlands*
[15]*Department of Physics, University of California at Davis, One Shields Avenue, Davis, CA 95616, USA*
[16]*California Institute of Technology, Pasadena, CA 91125, USA*
[17]*Observatoire de Paris, LERMA 61, avenue de l'Observatoire, 75014 Paris, France*
[18]*Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109, USA*

Accepted 2006 February 15. Received 2005 December 23; in original form 2005 June 6

ABSTRACT

The Shear Testing Programme (STEP) is a collaborative project to improve the accuracy and reliability of all weak lensing measurements in preparation for the next generation of wide-field surveys. In this first STEP paper, we present the results of a blind analysis of simulated ground-based observations of relatively simple galaxy morphologies. The most successful methods are shown to achieve percent level accuracy. From the cosmic shear pipelines that have been used to constrain cosmology, we find weak lensing shear measured to an accuracy that is within the statistical errors of current weak lensing analyses, with shear measurements accurate to better than 7 per cent. The dominant source of measurement error is shown to arise from calibration uncertainties where the measured shear is over or underestimated by a constant multiplicative factor. This is of concern as calibration errors cannot be detected through standard diagnostic tests. The measured calibration errors appear to result from stellar contamination, false object detection, the shear measurement method itself, selection bias and/or the use of biased weights. Additive systematics (false detections of shear) resulting from residual point-spread function anisotropy are, in most cases, reduced to below an equivalent shear of 0.001, an order of magnitude below cosmic shear distortions on the scales probed by current surveys.

Our results provide a snapshot view of the accuracy of current ground-based weak lensing methods and a benchmark upon which we can improve. To this end we provide descriptions of each method tested and include details of the eight different implementations of the commonly

⋆E-mail: heymans@physics.ubc.ca

used Kaiser, Squires & Broadhurst method (KSB+) to aid the improvement of future KSB+ analyses.

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

1 INTRODUCTION

Gravitational lensing provides an unbiased way to study the distribution of matter in the Universe. Derived from the physics of gravity, where gravitational light deflection is dependent solely on the distribution of matter, weak gravitational lens theory describes a unique way to directly probe dark matter on large scales (see the extensive review by Bartelmann & Schneider 2001). This tool has many astronomical applications; the detection of weak shear around galaxy clusters yields an estimate of the total cluster mass (see for example Wittman et al. 2003; Margoniner et al. 2005) and enables a full mass reconstruction of low redshift clusters (see for example Dahle et al. 2002; Gray et al. 2002; Clowe, Gonzalez & Markevitch 2004); the average weak tangential shear of distant galaxies around nearby galaxies constrains the ensemble average properties of dark matter halos (see for example Hoekstra, Yee & Gladders 2004; Sheldon et al. 2004); the weak lensing of background galaxies by foreground large-scale structure directly probes the evolution of the non-linear matter power spectrum, hence providing a signal that can constrain cosmological parameters (see review by Van Waerbeke & Mellier 2003). This last application has the great promise of being able to tightly constrain the properties of dark energy with the next generation of wide-field multicolour surveys (Heavens 2003; Jain & Taylor 2003; Benabed & Van Waerbeke 2004; Bernstein & Jain 2004; Refregier et al. 2004).

Technically, weak lensing is rather challenging to detect. It requires the measurement of the weak distortion that lensing induces in the shapes of observed galaxy images. These images have been convolved with the point-spread function (PSF) distortion of the atmosphere, telescope and camera. The accuracy of any analysis therefore depends critically on the correction for instrumental distortions and atmospheric seeing. Weak lensing by large-scale structure induces percent level correlations in the observed ellipticities of galaxies, termed 'cosmic shear'. This cosmological application of weak lensing theory is therefore the most demanding technically, owing to the fact that for any weak lensing survey, the instrumental distortions are an order of magnitude larger than the underlying cosmic shear distortion that we wish to detect. We therefore focus on the demands of this particular application even though our findings will be beneficial to all weak lensing studies.

The unique qualities of weak lensing as a dark matter and dark energy probe demand that all technical challenges are met and overcome, and this desire has lead to the development of some of the most innovative methods in astronomy. The first pioneering weak lensing measurement methods by Tyson, Wenk & Valdes (1990), Bonnet & Mellier (1995) and Kaiser, Squires & Broadhurst (1995) (KSB) have improved (Luppino & Kaiser 1997; Hoekstra et al. 1998) (KSB+) and diversified (Kaiser 2000; Rhodes, Refregier & Groth 2000; Bridle et al. 2002; Bernstein & Jarvis 2002; Refregier & Bacon 2003; Massey & Refregier 2005). Novel methods to model the spatial and temporal variation of the PSF have also been designed to improve the success of the PSF correction (Hoekstra 2004; Jarvis & Jain 2004). In addition, diagnostic techniques have been developed and implemented to provide indicators for the presence of residual systematic non-lensing distortions (Crittenden et al. 2002; Schneider, Van Waerbeke & Mellier 2002; Bacon et al. 2003; Brown et al. 2003).

Rapid technical development has mirrored the growth in observational efforts with the cosmic shear analysis of several wide-field optical surveys yielding joint constraints on the matter density parameter $\Omega_{\rm m}$ and the amplitude of the matter power spectrum σ_8 (Maoli et al. 2001; Rhodes, Refregier & Groth 2001; Van Waerbeke et al. 2001; Hoekstra, Yee & Gladders 2002a; Refregier, Rhodes & Groth 2002; Bacon et al. 2003; Brown et al. 2003; Hamana et al. 2003; Jarvis et al. 2003; Rhodes et al. 2004; Heymans et al. 2005; Hoekstra et al. 2005; Jarvis, Bernstein & Dolney 2005; Massey et al. 2005; Sembolini et al. 2005; Van Waerbeke, Mellier & Hoekstra 2005) and also constraints on the dark energy equation of state parameter w (Jarvis et al. 2005; Hoekstra et al. 2005; Sembolini et al. 2005). The results from these efforts are found to be in broad agreement and are fast becoming more credible with the most recent publications presenting the results from several different diagnostic tests to determine the levels of systematic error. Table 1 lists the most recent cosmic shear results from different authors or surveys, the two-point statistics used in the cosmological parameter analysis and the statistics used to determine levels of systematic errors through an E/B mode decomposition (Crittenden et al. 2002). See Schneider et al. (2002) and Brown et al. (2003) for details about each two-point statistic and their E/B mode decomposition and Massey et al. (2005), Van Waerbeke et al. (2005) and Heymans et al. (2005) for different discussions on which statistics are best to use. For such a young field of observational research, the $\sim 2\sigma$ agreement between the results, shown in Table 1, is rather impressive. The differences between the results are, however, often cited as a reason for caution over the use of cosmic shear as a cosmological probe. For this reason the Shear Testing Programme[1] (STEP) was launched in order to improve the accuracy and reliability of all future weak lensing measurements through the rigorous testing of shear measurement pipelines, the exchange of data and the sharing of technical and theoretical knowledge within the weak lensing community.

The current differences seen in cosmic shear cosmological parameter estimates could result from a number of sources; inaccurate source redshift distributions that are required to interpret the cosmic shear signal; sampling variance; systematic errors from residual instrumental distortions; calibration biases in the shear measurement method. Contamination to cosmic shear analyses from the intrinsic galaxy alignment of nearby galaxies is currently thought to be a weak effect that is measured and mitigated in Heymans et al. (2004) (also see King & Schneider 2002, 2003; Heymans & Heavens 2003, and references therein). With the next generation of wide-field multicolour surveys many of these problems can swiftly be resolved as the multicolour photometric redshifts will provide a good estimate of the redshift distribution (see for example Brown et al. 2003)

[1] http://www.physics.ubc.ca/~heymans/STEP.html

Table 1. The most recent cosmological parameter constraints on the amplitude of the matter power spectrum σ_8 from each author or survey, for a matter density parameter $\Omega_m = 0.3$. Quoted errors on σ_8 are 1σ (68 per cent confidence) except in the case of Jarvis et al. (2005) where the errors given are 2σ (95 per cent confidence). Several different statistics have been used to constrain σ_8, as detailed, where $\langle M_{ap}{}^2 \rangle$ is the mass aperture statistic, $\langle \gamma^2 \rangle$ is the top-hat shear variance, $\xi_\pm$ are the shear correlation functions and $P^{\kappa\kappa}$ is the shear power spectrum. The statistics used to determine the level of non-lensing B-modes in each result are also listed where $\langle M_\perp{}^2 \rangle$ is the B-mode mass aperture statistic, ξ^E and ξ^B are E and B mode correlators, $P^{\beta\beta}$ is the B-mode shear power spectrum, and $P^{\kappa\beta}$ is the E/B cross power spectrum. See Schneider et al. (2002) and Brown et al. (2003) for details about each two-point statistic and their E/B mode decomposition. The shear measurement pipeline that has been used for each result is listed for reference, along with the area of the survey and the median redshift estimate of the survey z_m. Space-based surveys are denoted with an (s) in the area column.

Survey analysis	Pipeline description	σ_8	Statistic	E/B decomposition	Area (deg^2)	z_m
Hoekstra et al. (2002a)	Hoekstra et al. (1998)	$0.86^{+0.09}_{-0.13}$	$\langle M_{ap}{}^2 \rangle$	$\langle M_{ap}{}^2 \rangle \langle M_\perp^2 \rangle$	53.0	$0.54 - 0.66$
Refregier et al. (2002)	Rhodes et al. (2000)	0.94 ± 0.24	$\langle \gamma^2 \rangle$	$\langle M_{ap}{}^2 \rangle \langle M_\perp^2 \rangle$	0.36 (s)	0.9 ± 0.1
Brown et al. (2003)	Bacon, Refregier & Ellis (2000)	0.72 ± 0.09	$\xi_\pm$ $P^{\kappa\kappa}$	$P^{\kappa\kappa}$ $P^{\kappa\beta}$ $P^{\beta\beta}$	1.25	0.85 ± 0.05
Hamana et al. (2003)	Hamana et al. (2003)	$0.78^{+0.55}_{-0.25}$	$\langle M_{ap}{}^2 \rangle$	$\langle M_{ap}{}^2 \rangle \langle M_\perp^2 \rangle$	2.1	$0.6 - 1.4$
Rhodes et al. (2004)	Rhodes et al. (2000)	1.02 ± 0.16	$\langle \gamma^2 \rangle$	none	0.25 (s)	1.0 ± 0.1
Van Waerbeke et al. (2005)	Van Waerbeke et al. (2000)	0.83 ± 0.07	$\langle M_{ap}{}^2 \rangle$ ξ^E	$\langle M_{ap}{}^2 \rangle \langle M_\perp^2 \rangle$ ξ^E ξ^B	8.5	$0.8 - 1.0$
Jarvis et al. (2005)	Bernstein & Jarvis (2002)	$0.72^{+0.17}_{-0.14}$	$\langle \gamma^2 \rangle$ $\langle M_{ap}{}^2 \rangle$	$\langle M_{ap}{}^2 \rangle \langle M_\perp^2 \rangle$	75.0	0.6 ± 0.1
Massey et al. (2005)	Bacon et al. (2000)	1.02 ± 0.15	$\xi_\pm$	ξ^E ξ^B	4.5	0.8 ± 0.08
Heymans et al. (2005)	Heymans et al. (2005)	0.68 ± 0.13	$\xi_\pm$, $P^{\kappa\kappa}$	ξ^E ξ^B $P^{\kappa\kappa}$ $P^{\kappa\beta}$ $P^{\beta\beta}$	0.22 (s)	1.0 ± 0.1

and the wide areas will minimize sampling variance. In addition, all new instrumentation has been optimized to reduce the severity of instrumental distortions improving the accuracy of future PSF corrections. Implementing diagnostic statistics that decompose cosmic shear signals into their lensing E-modes and non-lensing B-modes (Crittenden et al. 2002; Schneider et al. 2002; Brown et al. 2003) immediately alerts us to the presence of systematic error within our data set. B-mode systematics can then be reduced through the modification of PSF models (Jarvis & Jain 2004; Van Waerbeke et al. 2005) or merely the selection of angular scales above or below which the systematics are removed. Calibration bias is therefore perhaps of greatest concern as, in contrast to additive PSF errors, it can only be directly detected through the cosmic shear analysis of image simulations, although see the discussion on self-calibration in Huterer et al. (2006) and Hirata et al. (2004) and Mandelbaum et al. (2005) for model-dependent estimates of shear calibration errors in the Sloan Digital Sky Survey. With the statistics currently used to place constraints on cosmological parameters, a shear calibration error contributes directly to an error in σ_8. The recent development of statistics which are fairly insensitive to shear calibration errors (Jain & Taylor 2003; Bernstein 2006) are certainly one solution to this potential problem. Also see Ishak (2005), where shear calibration uncertainties are marginalized over in the cosmological parameter estimation.

Bacon et al. (2001), Erben et al. (2001) and Hoekstra et al. (2002b) presented the first detailed cosmic shear analyses of artificial image simulations using the KSB+ method. Bacon et al. (2001) found that the KSB+ method was reliable to $\sim$5 per cent provided a calibration factor of 0.85 was included in the analysis to increase the KSB+ shear estimator. The calibration factor has since been included in the work of Bacon et al. (2003), Brown et al. (2003) and Massey et al. (2005) who implement the KSB+ pipeline tested in Bacon et al. (2001). Erben et al. (2001) found that depending on the PSF type tested and the chosen implementation of the KSB+ formula, described in Section 2.1, the KSB+ method was reliable to $\pm$10–15 per cent and did not require a calibration correction. The artificial images tested by Hoekstra et al. (2002b) included cosmic shear derived from ray-tracing simulations. They found that the input lensing signal could be recovered to better than 10 per cent of the input value. The difference between these three conclusions is important. All papers adopted the same KSB+ method, but subtle differences in their implementation resulted in the need for a calibration correction in one case but not in the others. It is therefore not sufficient to cite these papers to support the KSB+ method as every individuals' KSB+ pipeline implementation may differ slightly, introducing a discrepancy between the results.

For the cosmic shear, galaxy–galaxy lensing and cluster mass determinations published to date, $\leqslant$10 per cent errors are at worst comparable to the statistical errors and are not dominant. Much larger surveys now underway will, however, reduce statistical errors on various shear measurements to the $\sim$2 per cent level, requiring shear measurement accurate to $\sim$1 per cent. In the next decade, deep weak-lensing surveys of thousands of square degrees will require shear measurements accurate to $\sim$0.1 per cent. The technical challenges associated with measuring weak lensing shear must therefore be addressed and solved in a relatively short period of time.

Whilst KSB+ is currently the most widely used weak lensing method, promising alternative methods have been developed [Rhodes et al. 2000 (RRG); Kaiser 2000 (K2K); Smith 2000 (ELLIPTO); Bridle et al. 2002 (Im2shape); Bernstein & Jarvis 2002 (BJ02); Refregier 2003 (shapelets); Massey & Refregier 2005 (polar shapelets)] and implemented in cosmic shear analyses [see for example Wittman et al. 2001 (ELLIPTO); Jarvis et al. 2003, 2005 (BJ02); Chang, Refregier & Helfand 2004 (shapelets); Rhodes et al. 2004 (RRG)] and cluster lensing studies [see for example Dahle et al. 2002 (K2K); Bardeau et al. 2005 (Im2shape); Margoniner et al. 2005 (ELLIPTO)]. Thorough testing of these newer techniques is however somewhat lacking in the literature, although see Refregier & Bacon (2003) and Massey, Refregier & Bacon (2004a) for tests of the shapelets method.

In this paper, we present the first of the STEP initiatives; the blind[2] analysis of sheared image simulations with a variety of weak lensing measurement pipelines used by each author in their previously published work. Authors and methods are listed in Table 2. Modifications to pipelines used in published work have not been

[2]CH, LV and KK knew the input shear of the simulations.

Table 2. Table of authors and methods. The key identifies the authors in all future plots and tables.

Author	Key	Method
Bridle & Hudelot	SB	Im2shape (Bridle et al. 2002)
Brown	MB	KSB+ [Bacon et al. (2000) pipeline]
Clowe	C1 & C2	KSB+
Dahle	HD	K2K (Kaiser 2000)
Hetterscheidt	MH	KSB+ [Erben et al. (2001) pipeline]
Heymans	CH	KSB+
Hoekstra	HH	KSB+
Jarvis	MJ	Bernstein & Jarvis (2002) Rounding kernel method
Kuijken	KK	Shapelets to 12th order Kuijken (2006)
Margoniner	VM	Wittman et al. (2001)
Nakajima	RN	Bernstein & Jarvis (2002) Deconvolution fitting method
Schrabback	TS	KSB+ [Erben et al. (2001) + modifications]
Van Waerbeke	LV	KSB+

allowed in light of the results and we thus present our results openly to provide the reader with a snapshot view of how accurately we can currently measure weak lensing shear from galaxies with relatively simple morphologies. This paper will thus provide a benchmark upon which we can improve in future STEP initiatives. Note that some of the methods evaluated in this paper are experimental and/or in early stages of development, notably the methods of Kuijken (2006), the deconvolution fitting method of Nakajima (in preparation)and the Dahle implementation of K2K. The results from these particular methods should therefore not be taken as a judgment on their ultimate potential.

This paper is organized as follows. In Section 2, we review the different shear measurement methods used by each author and describe the simulated data set in Section 3. We compare each authors' measured shear with the input simulation shear in Section 4 investigating forms of calibration bias, selection bias and weight bias. Note that our discussion on the issue of source selection bias is indeed relevant for many different types of survey analysis, not only the lensing applications detailed here. We discuss our findings in Section 5 and conclude in Section 6.

2 METHODS

In the weak lensing limit the ellipticity of a galaxy is an unbiased estimate of the gravitational shear. For a perfect ellipse with axial ratio β at position angle θ, measured counter-clockwise from the x-axis, we can define the following ellipticity parameters (Bonnet & Mellier 1995):

$$\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = \frac{1-\beta}{1+\beta} \begin{pmatrix} \cos 2\theta \\ \sin 2\theta \end{pmatrix}, \tag{1}$$

and the complex ellipticity $e = e_1 + ie_2$. In the case of weak shear $|\gamma| \ll 1$, the shear $\gamma = \gamma_1 + i\gamma_2$ is directly related to the average galaxy ellipticity, $\gamma \approx \langle e \rangle$. In this section, we briefly review the different measurement methods used in this STEP analysis to estimate galaxy ellipticity in the presence of instrumental and atmospheric distortion and hence obtain an estimate of the gravitational shear γ. Common to all methods is the initial source detection stage, typically performed using the SEXTRACTOR (Bertin & Arnouts 1996) software. The peak finding tool HFINDPEAKS from the IMCAT[3] software is used as an alternative in some KSB+ methods, listed in Table A1. In order to characterize the PSF, stars are selected in all cases from a magnitude–size plot.

2.1 KSB+ method

Kaiser et al. (1995), Luppino & Kaiser (1997) and Hoekstra et al. (1998) (KSB+) prescribe a method to invert the effects of the PSF smearing and shearing, recovering a shear estimator uncontaminated by the systematic distortion of the PSF.

Objects are parameterized according to their weighted quadrupole moments

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

where I is the surface brightness of the object, θ is the angular distance from the object centre and W is a Gaussian weight function of scale length r_g, where r_g is some measurement of galaxy size. For a perfect ellipse, the weighted quadrupole moments are related to the weighted ellipticity parameters[4] ε_α by

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} = \frac{1}{Q_{11}+Q_{22}} \begin{pmatrix} Q_{11}-Q_{22} \\ 2Q_{12} \end{pmatrix}. \tag{3}$$

Kaiser et al. (1995) show that if the PSF distortion can be described as a small but highly anisotropic distortion convolved with a large circularly symmetric seeing disc, then the ellipticity of a PSF corrected galaxy is given by

$$\varepsilon_\alpha^\mathrm{cor} = \varepsilon_\alpha^\mathrm{obs} - P_{\alpha\beta}^\mathrm{sm} p_\beta, \tag{4}$$

where p is a vector that measures the PSF anisotropy, and P^sm is the smear polarisability tensor given in Hoekstra et al. (1998). $p(\boldsymbol{\theta})$ can be estimated from images of stellar objects at position $\boldsymbol{\theta}$ by noting that a star, denoted throughout this paper with *, imaged in the absence of PSF distortions has zero ellipticity: $\varepsilon_\alpha^{*\mathrm{cor}} = 0$. Hence,

$$p_\mu = (P^{\mathrm{sm}*})^{-1}_{\mu\alpha}\, \varepsilon_\alpha^{*\mathrm{obs}}. \tag{5}$$

The isotropic effect of the atmosphere and weight function can be accounted for by applying the pre-seeing shear polarisability tensor correction P^γ, as proposed by Luppino & Kaiser (1997), such that

$$\varepsilon_\alpha^\mathrm{cor} = \varepsilon_\alpha^\mathrm{s} + P_{\alpha\beta}^\gamma \gamma_\beta, \tag{6}$$

where ε^s is the intrinsic source ellipticity and γ is the pre-seeing gravitational shear. Luppino & Kaiser (1997) show that

$$P_{\alpha\beta}^\gamma = P_{\alpha\beta}^\mathrm{sh} - P_{\alpha\mu}^\mathrm{sm} (P^{\mathrm{sm}*})^{-1}_{\mu\delta} P_{\delta\beta}^{\mathrm{sh}*}, \tag{7}$$

where P^sh is the shear polarisability tensor given in Hoekstra et al. (1998) and $P^{\mathrm{sm}*}$ and $P^{\mathrm{sh}*}$ are the stellar smear and shear polarisability tensors, respectively. Combining the PSF correction, equation (4), and the P^γ seeing correction, the final KSB+ shear estimator $\hat{\gamma}$ is given by

$$\hat{\gamma}_\alpha = (P^\gamma)^{-1}_{\alpha\beta} \left[\varepsilon_\beta^\mathrm{obs} - P_{\beta\mu}^\mathrm{sm} p_\mu \right]. \tag{8}$$

This method has been used by many of the authors although different interpretations of the above formula have introduced some subtle differences between each authors' KSB+ implementation. For this reason, we provide precise descriptions of each KSB+ pipeline in the Appendix A.

[3] www.ifa.hawaii.edu/∼kaiser/imcat/

[4] The KSB+ definition of galaxy ellipticity differs from equation (1). If the weight function $W(\boldsymbol{\theta}) = 1$ in equation (2), the KSB+ ellipticity $|\varepsilon| = (1-\beta^2)/(1+\beta^2)$, where β is the axial ratio (see Bartelmann & Schneider 2001).

2.2 K2K method

One drawback of the KSB+ method is that for non-Gaussian PSF distortions, the KSB PSF correction is mathematically poorly defined. Kaiser (2000) (K2K) addresses this issue by properly accounting for the effects of a realistic PSF. It also proposes measuring shapes from images that have been convolved with a recircularizing PSF, where the recircularizing PSF is a 90° rotation of a modelled version of the PSF. Section 2.3.6 of Dahle et al. (2002) provides a condensed description of the K2K shear estimator which has been applied to the STEP simulations by Dahle (HD).

2.3 Shapelets

The shapelets formalism of Refregier (2003) allows galaxy images to be decomposed into orthogonal basis functions which transform simply under a variety of operations, in particular, shear and (de)convolution. The expansion is based on a circular Gaussian, but inclusion of higher orders allows general shapes to be described well.

Kuijken (2006) uses the shapelets formalism of Refregier (2003) to derive individual shape estimators that differ from the method of Refregier & Bacon (2003). We briefly review this method which is based on the 'constant ellipticity object' estimator of Kuijken (1999), referring the reader to Kuijken (2006) for further details. Each galaxy image is fitted as an intrinsically circular source that has been sheared and then smeared by the PSF. These operations are efficiently expressed in terms of shapelets as

$$G_{\rm model} = P \cdot (1 + \gamma_1 S_1 + \gamma_2 S_2) \cdot C, \tag{9}$$

where $G_{\rm model}$ is the model for the galaxy image, P is the known PSF convolution operator (expressed as a matrix operating on shapelet coefficients), S_i are the first-order shear operators, γ_i are the shear distortions that are fitted and C is a general circular source of arbitrary radial luminosity profile (expressed as a superposition of shapelets). Note that P is determined from stellar objects whose shapelet coefficients are interpolated separately across the field of view to the position of each observed galaxy. Fitting this model to each observed galaxy image yields a best-estimate (γ_1, γ_2) shear distortion value for each galaxy, which can then be averaged or correlated to yield shear estimators. In this paper, we use $\gamma_i = \langle \gamma_i \rangle / (1 - \langle \gamma^2 \rangle)$ as an estimate for the shear from the ensemble population. The factor in the denominator is the response of the average ellipticity of a population of elliptical sources to an overall shear (BJ02). To cope with possible centroiding errors, an arbitrary translation is included in the fit as well. The uncertainties on the pixel values of each galaxy image can be propagated into the shapelet coefficients, and to the estimates of the γ_i. This method is exact for galaxies that are intrinsically circular or elliptical. Kuijken (1999) shows that this method also works well for galaxies whose ellipticity or position angle varies with radius.

2.4 Im2shape

Im2shape (Bridle et al. 2002; Bridle et al., in preparation) fits a sum of elliptical Gaussians to each object image, taking into account unknown background and noise levels. This approach follows that suggested by Kuijken (1999).

SEXTRACTOR is used to define postage stamps containing each object[5] and galaxies and stars are selected from the size–magnitude plot from the SEXTRACTOR output. The galaxies are modelled by Im2shape using two concentric Gaussians, with six free parameters for the first Gaussian, and two additional free parameters (size and amplitude) for the second Gaussian. The noise is assumed to be uncorrelated, Gaussian and at the same level for all pixels in the postage stamp. The background level is assumed to be constant across the postage stamp. Including the noise and background levels there are 10 free galaxy parameters in total. Two Gaussians are used for the stars in all the images, except for PSF 2, for which the amplitude of the second Gaussian was found to be so small that one Gaussian was used instead. Where two Gaussians were used to fit the stars, the Gaussians were taken to have totally independent parameters, with 12 free parameters for the Gaussians, plus the noise and background levels, making 14 free parameters in total. To estimate these free parameters fast and efficiently, Im2shape makes use of the BayeSys engine (written by Skilling & Gull). This implements Markov–Chain Monte Carlo sampling (MCMC) which is used to obtain samples from the probability distribution of the unknown parameters. Estimates of the free parameters are then taken from the mean value of the parameter across the MCMC samples, and the uncertainties are taken from the standard deviation. With this data set the MCMC analysis takes ∼15 s per galaxy image on the COSMOS[6] supercomputer.

To account for the PSF a grid of 5 × 5 points was defined on each image, and the PSF at each point was estimated by taking the median parameters of the nearest five stars (note that Im2shape was run on all the stellar-like objects and cuts were then used to remove outliers). For each galaxy, the PSF shape was taken from the grid point closest to the galaxy in question. The trial galaxy parameters were then combined with the PSF parameters analytically to calculate the convolved image shape. The intensity in the centre of each pixel is calculated and this is corrected for the integration over the pixel using the curvature of the Gaussian at the centre of the pixel (for both star and galaxy shape estimation). The final ellipticity values for each galaxy (equation 1) are found from averaging over all the MCMC samples. Only galaxies with ellipticity uncertainties less than 0.25 were included in the final catalogue, as for higher ellipticity uncertainties the error estimates are less reliable resulting from the probability distribution becoming less Gaussian. To obtain an estimate of the shear from these ellipticity estimates, the ellipticities are weighted by the inverse square of the ellipticity uncertainties added in quadrature with the intrinsic ellipticity dispersion $\sigma_{\rm e}$ of the galaxies, found to be $\sigma_{\rm e} = 0.2$.

2.5 Wittman method with ELLIPTO

This method uses a recircularizing kernel to eliminate PSF anisotropy, and 'adaptive' moments (moments weighted by the best-fitting elliptical Gaussian) to characterize the ellipticity of the source galaxies. It is a partial implementation of BJ02, discussed in Section 2.6, and primarily differs from BJ02 by using a simpler recircularizing kernel.

SEXTRACTOR is used for initial object detection. SEXTRACTOR centroids and moments are then input to the ELLIPTO program (Smith 2000; Smith et al. 2001) which measures the adaptive moments. ELLIPTO also remeasures the centroid and outputs an error flag when the centroid differs from the SEXTRACTOR centroid. This typically happens with blended objects or those with nearby neighbours,

[5]The postage stamps used for this analysis were 16 × 16 pixels centered on the SEXTRACTOR position.

[6]http://www.damtp.cam.ac.uk.cosmos, SGI Altix 3700, 1.3-GHz Madison processors.

whose measured shapes may not be trustworthy in any case. Stars are selected with an automatic routine which looks for a dense locus at a constant ELLIPTO size. The selection is then visually checked. In real data, ∼5 per cent of images require manual tweaking of the star selection, although this manual stage was not required for the STEP simulations. The spatial variation of the adaptive moments is then fit with a second-order polynomial for each CCD of each exposure. This fit is then used to generate a spatially varying 3 × 3 pixel recircularizing kernel, following (Fischer & Tyson 1997). Note that a 3 × 3 kernel may be too small to properly correct a well-sampled, highly elliptical PSF; the practical limit appears to be ∼0.1 ellipticity. In those cases, the recircularization step may be applied iteratively, mimicking the effect of larger kernels. For the STEP simulations, only PSF 3 required a second iteration, but three iterations were applied to all PSFs.

After recircularization, the object detection and ELLIPTO measurements are repeated to generate the final catalogue. Note that object detection on the recircularized image in principle eliminates PSF-anisotropy-dependent selection bias. Objects are rejected from the final catalogue if: the ELLIPTO error is non-zero; measured (pre-dilution correction) scalar ellipticity >0.6 (simulations show that, with ground-based seeing, most of these are blends of unrelated objects); or size <120 per cent of the PSF size. The adaptive moments are then corrected for dilution by an isotropic PSF and a responsivity correction using the formulae of BJ02. Weighting is not applied to the data. Note that this method has been used for cluster analyses but not for any published cosmic shear results.

2.6 Bernstein and Jarvis method: BJ02

The Jarvis (MJ) and Nakajima (RN) methods each extend the ELLIPTO technique by methods detailed in BJ02. Both are based upon expansions of the galaxy and PSF shapes into a series of orthogonal two-dimensional (2D) Gaussian-based functions, the Gauss–Laguerre expansion, also known as 'polar shapelets' in Massey & Refregier (2005). Both the Jarvis (MJ) and Nakajima (RN) methods move beyond the approximation, inherent in both the ELLIPTO and KSB methods, that the PSF asymmetry can be described as a first-order perturbation to a circular PSF. The Jarvis (MJ) method applies 'rounding kernel' filters from size 3 × 3 pixels and up to the images in order to null several asymmetric Gauss–Laguerre coefficients of the PSF, not just the quadrupoles. Note that for PSF ellipticities of order ∼0.1, a 3 × 3 pixel kernel is sufficient to round out stars up to approximately 30 pixels in diameter. The galaxy shapes are next measured by the best-fitting elliptical Gaussian; formulae proposed by Hirata & Seljak (2003) are used to correct the observed shapes for the circularizing effect of the PSF.

The 'deconvolution fitting method' by Nakajima (RN) implements nearly the full formalism proposed by BJ02, which is further elaborated in Nakajima et al (in preparation): the intrinsic shapes of galaxies are modelled as Gauss–Laguerre expansions (to 8th order). These are then convolved with the PSF and fit directly to the observed pixel values in a similar fashion to Kuijken (1999). This should fully capture the effect of highly asymmetric PSFs or galaxies, as well as the effects of finite sampling. Note that both methods use the weighting scheme described in Section 5 of BJ02.

A difference between the BJ02 approaches and the Refregier & Bacon (2003) shapelets implementation is that the latter uses a circular Gaussian basis set, whereas the BJ02 method shears the basis functions until they match the ellipticity of the galaxy. This in principle eliminates the need to calculate the 'shear polarisabilities' that appear in KSB.

3 STEP SIMULATION DATA

For this analysis, we have created an artificial set of survey images using the SKYMAKER programme.[7] A detailed description of this software and the galaxy catalogue generator, STUFF,[8] can be found in Erben et al. (2001) and Bertin & Fouqué (in preparation) and we therefore only provide a brief summary here. In short, for a given cosmology and survey description, galaxies are distributed in redshift space with a luminosity and morphological-size distribution as defined by observational and semi-analytical relations. Galaxies are made of a co-axial de Vaucouleurs-type spheroid bulge and a pure oblate circular exponential thin disc (see Bertin & Arnouts 1996, for details). The intrinsic flattening q of spheroids is taken between 0.3 and 1, and within this range follows a normal distribution with $\langle q \rangle = 0.65$ and $\sigma_q = 0.18$ (Sandage, Freeman & Stokes 1970). Note that we assume the same flattening distribution for bulges and ellipticals, even if there is some controversy about this (Boroson 1981). Inclination angles i are randomly assigned following a flat distribution, as expected from uniformly random orientations with respect to the line of sight. The apparent axis ratio β is given by $\beta = \sqrt{q^2 \sin^2 i + \cos^2 i}$ for the spheroid component, and given by $\beta = \cos i$ for the thin disc. The bulge plus disc galaxy is finally assigned a random position angle θ on the sky and the bulge and disc intrinsic ellipticity parameters are then calculated from equation (1).

It has been known for some time that pure oblate circular discs, oriented with a flat distribution of inclination angles, do not provide a good match to the statistics from real disc galaxies (Binney & de Vaucouleurs 1981; Grosbol 1985; Lambas, Maddox & Loveday 1992): in particular, observations show a striking deficiency of galaxies with zero ellipticities. Although surface-brightness selection effects are not to be ignored (see for example Huizinga & van Albada 1992), there is now general agreement that this phenomenon mostly betrays intrinsic ellipticities of disc planes. The origin of these intrinsic ellipticities is not completely clear (see Binney & Merrifield 1998), and is thought to originate partly from non-axisymmetric spiral structures and/or a tri-axial potential (Rix & Zaritsky 1995). The simulations used in this analysis ignore these aspects, and the simulated galaxies are therefore intrinsically 'rounder' on average than real galaxies. This should not impact on the lensing analysis that follows, except in the cases where weighting schemes are used that take advantage of the sensitivity of intrinsically circular galaxies to measure weak lensing shear. These schemes will have an apparent signal-to-noise advantage in the current simulations, which is expected to decrease given real data.

A series of five different shears are applied to the galaxy catalogue by modifying the observed intrinsic source ellipticity to create sheared galaxies where

$$e = \frac{e^{(s)} + g}{1 + g^* e^{(s)}} \tag{10}$$

(Seitz & Schneider 1997), and g is the complex reduced shear. For this set of simulations, the convergence $\kappa = 0$, hence the reduced shear $g = \gamma/(1-\kappa) = \gamma$, where $\gamma_1 = (0.0, 0.005, 0.01, 0.05, 0.1)$, $\gamma_2 = 0.0$. Sheared bulge and disc axial ratios and position angles are then calculated from equation (1) and the model galaxy images are created. Stars are simulated assuming a constant slope of 0.3 mag^{-1} interval for the logarithm of differential stellar number counts down to and I-band magnitude $I = 25$. Model galaxy images and

[7] http://terapix.iap.fr/cplt/oldSite/soft/skymaker
[8] ftp://ftp.iap.fr/pub/from_users/bertin/stuff

Table 3. The SKYMAKER simulations are convolved with this series of uniform PSF models.

PSF ID	PSF type	Ellipticity
0	No anisotropy	0.00
1	Coma	∼0.04
2	Jitter, tracking error	∼0.08
3	Defocus	∼0.00
4	Astigmatism	∼0.00
5	Triangular (trefoil)	0.00

stellar point sources are then convolved with a series of six different optical PSFs that are listed in Table 3 and shown in Fig. 1. These PSF models were chosen to provide a realistic representation of the types of PSF distortions that are seen in ground-based observations, through ray-tracing models of the optical plane. They also include atmospheric turbulence, where the seeing scale is chosen such that when the turbulence is combined with the PSF anisotropy, all stars have FWHM of 0.9 arcsec. The ellipticity of the PSF from real data is typically of the order of 5 per cent, which is similar to the coma model PSF 1. PSF 2 which features a jitter or tracking error is very elliptical in comparison. The other PSF models test the impact of non-Gaussian PSF distortions. A uniform background with surface brightness 19.2 mag arcsec^{-2} is added to the image, chosen to match the I-band sky background at the Canada–France–Hawaii Telescope site. Poisson photon shot noise and Gaussian read-out noise is then applied.

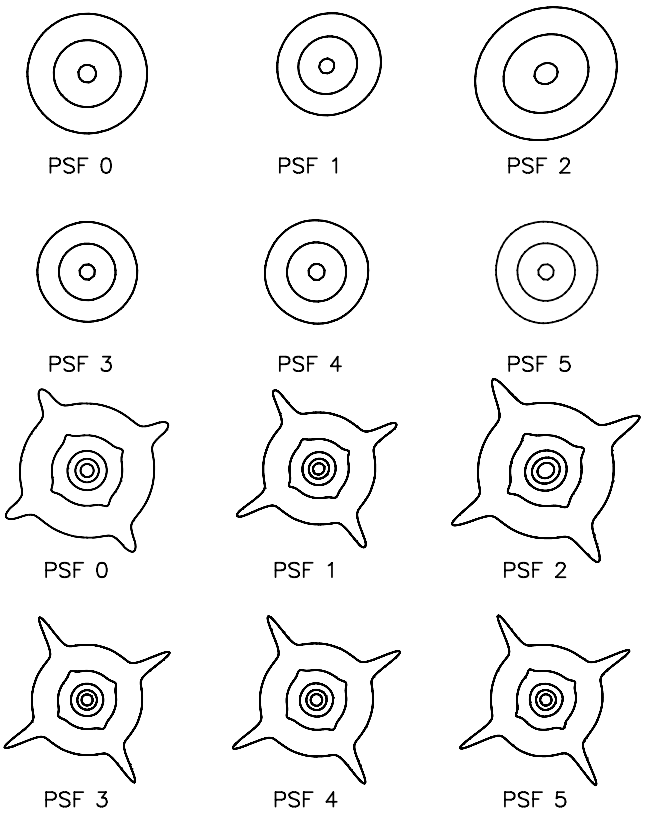

Figure 1. SKYMAKER PSF models, as described in Table 3. The upper panel shows the PSF core distortion, with contours marking 3, 25 and 90 per cent of the peak intensity. The lower panel shows the extended diffraction spikes, with contours marking 0.003, 0.03, 0.3, 3 and 25 per cent of the peak intensity.

The combination of six different PSF types and five different applied shears gives 30 different data sets where each set consists of an ensemble of 64 4096 × 4096 pixel images of pixel scale 0.206 arcsec. For computational efficiency the data in each set stems from the same base catalogue, and as the sky noise levels are the same for each data set, many of the parameters required for the SEXTRACTOR source detection software are the same for each data set. Aside from this time-saving measure of setting some of the SEXTRACTOR source detection parameters only once, prior information about the simulations have not been used in the cosmic shear analyses. Each image contains ∼15 galaxies per square arcmin resulting in low-level shot noise from the intrinsic ellipticity distribution at the 0.1 per cent level for each data set. Stellar object density is ∼10 stars per square arcmin of which roughly 150 per image were sufficiently bright for the characterization of the PSF. This density of stellar objects is slightly higher than that found with typical survey data and was chosen to aid PSF correction. It does however increase the likelihood of stellar contamination in the selected galaxy catalogue. Although the PSF is uniform across the field of view, uniformity has only been assumed in one case (RN).

The reader should note that the SKYMAKER simulations should, in principle, provide an easy test of our methods as many shear measurement methods are based on the assumption that the galaxy shape and PSF are smooth, elliptical and in some cases Gaussian. In reality the shapes of faint galaxies can be quite irregular and, particularly, in the case of space-based observations, the PSF can contain significant structure. In addition, the SKYMAKER galaxies have reflection symmetry about the centroid which could feasibly cause any symmetrical errors to vanish. We should also note that some of the authors have previously used SKYMAKER simulations to test their methods (see Erben et al. 2001; Hoekstra et al. 2002b). These issues will therefore be addressed by two future STEP publications with the blind analysis of a more realistic set of artificial images that use shapelet information to include complex galaxy morphology (Massey et al. 2004b). With these shapelet simulations, we will investigate the shear recovery from ground-based observations (Massey et al., in preparation) and space-based observations (Rhodes et al., in preparation).

4 ANALYSIS

In this section, we compare each authors' measured shear catalogues with the input to each SKYMAKER simulation. We match objects in each authors' catalogue to the input galaxy and stellar catalogue, within a tolerance of 1 arcsec. Table 4 lists several general statistics calculated from the PSF model 0 (no anisotropy) $\gamma = (0.005, 0.0)$ set which is a good representation of the STEP simulation data. The source extraction method used by each author is listed in Table 4 as well as the average number density of selected sources per square arcmin, Ngals. To minimize shot noise, we wish to maximize the number of sources without introducing false detections into the sample (note the percentage of false detections listed in the 'per cent false' column in Table 4) or contaminating the sample with stellar objects (note the percentage of stellar contamination listed in the 'per cent stars' column in Table 4). Both false objects and stars add noise which can dilute the average shear measurement. Typically the number of false detections are negligible and the stellar contamination is below 5 per cent. The notable exception is the Dahle (HD) method that suffers from strong stellar contamination for all PSF types, a problem that can easily be improved upon in future analyses. Where authors use object weights w_i in their analysis, the weighted percentage stellar contamination (per cent stars

Table 4. Table to compare the different number density of selected sources per square arcmin, Ngals and the percentage of stellar contamination (per cent stars) and false detections (per cent false) in each authors' catalogue. Each catalogue has been created using either the SEXTRACTOR and/or the HFINDPEAKS software. Where authors use object weights, the weighted percentage of stellar contamination (per cent stars′) and false detections (per cent false′) are also listed. The final columns give estimates of the signal-to-noise of the resulting shear measurement as described in the text. $\mathrm{SNR} = \gamma_i^{\mathrm{true}}/\sigma_\gamma$ is the signal-to-noise ratio of the shear measurement. $\mathrm{S/N_s}$ is the signal-to-shot-noise determined from the galaxies selected by each author. Where authors use object weights, the signal-to-weighted-shot-noise $\mathrm{S/N_s'}$ is also determined.

Author	Ngals (per arcmin^2)	Per cent stars	Per cent false	Per cent stars′	Per cent false′	Software	SNR	$\mathrm{S/N_s}$	$\mathrm{S/N_s'}$
SB	18	1.9	3.8	1.5	3.9	SEXTRACTOR	6	7	6
MB	14	7.1	0.1	–	–	HFINDPEAKS	8	10	–
C1	12	2.6	0.0	1.1	0.0	HFINDPEAKS & SEXTRACTOR	9	9	11
C2	12	2.7	0.0	1.2	0.0	HFINDPEAKS & SEXTRACTOR	9	9	11
HD	17	44.8	0.0	–	–	HFINDPEAKS	7	8	–
MH	14	3.9	0.0	2.4	0.0	SEXTRACTOR	12	11	14
CH	12	2.9	0.0	–	–	SEXTRACTOR	7	11	–
HH	16	10.8	0.0	0.1	3.6	HFINDPEAKS	8	10	11
MJ	9	0.0	3.6	0.0	1.0	SEXTRACTOR	16	8	22
KK	9	0.8	0.0	0.3	0.0	SEXTRACTOR	10	10	12
VM	13	3.8	0.0	–	–	SEXTRACTOR	10	10	–
RN	9	0.9	0.4	1.5	0.1	SEXTRACTOR	19	10	24
TS	10	1.4	0.0	0.9	0.0	SEXTRACTOR	12	11	14
LV	13	0.0	0.0	0.0	0.0	SEXTRACTOR	11	11	12

$= [\Sigma_{i=\mathrm{stars}}\, w_i/\Sigma_{i=\mathrm{all}}\, w_i] \times 100$ per cent) and weighted percentage of false object contamination (per cent false) are also listed. This shows, for example, that in the case of Hoekstra (HH), the 10 per cent stellar objects are given a very low weight and therefore do not significantly contribute to the weighted average shear measurement.

Average centroid offsets measured from each authors selected catalogues, were found to be <0.001 pixels for SEXTRACTOR based catalogues and $\sim 0.005 \pm 0.001$ pixels for HFINDPEAKS-based catalogues. Centroid accuracy is however likely to be data dependent, and S/N dependent (see Erben et al. 2001). Thus care should still be taken in determining centroids to prevent the problems described in Van Waerbeke et al. (2005) where errors in the SEXTRACTOR centroiding in one field were found to be the source of strong B-modes on large scales. Note that starting from version 2.4.3, SEXTRACTOR provides iterative, Gaussian-weighted centroid measurements XWIN_IMAGE and YWIN_IMAGE which have been shown to be even more accurate than previous SEXTRACTOR centroid measures (Bertin & Fouqué, in preparation).

For each data set, we calculate the mean (weighted) shear measured by each author, treating each of the 64 images as an independent pointing. We take the measured shear for each data set γ_i to be the mean of the measurements from the 64 images and assign an error σ_γ given by the error on the mean. The final three columns of Table 4 demonstrate the effect of weights and galaxy selection on the signal-to-noise of the measurement. The signal-to-noise of the shear measurement is defined as $\mathrm{SNR} = \gamma_i^{\mathrm{true}}/\sigma_\gamma$, where γ_i^{true} is the input shear ($\gamma_1^{\mathrm{true}} = 0.005$ for the data analysed in Table 4). The signal-to-shot-noise is defined as $\mathrm{S/N_s} = \gamma_i^{\mathrm{true}}/\sigma$, where σ is the error on the mean galaxy ellipticity e (equation 1) measured from the 64 images. Note that the shot noise σ is calculated from the known input ellipticities of galaxies selected by each author. The final column applies to authors who use weights, where the signal-to-weighted-shot-noise is defined as $\mathrm{S/N_s'} = \gamma_i^{\mathrm{true}}/\sigma'$, where σ' is the error on the mean weighted galaxy ellipticity.

Several things can be noted from the signal-to-noise calculations. Firstly, the high magnitude, as weak shear has not been measured from data with SNR >10. One must not forget however that if weak lensing shear was constant across large areas of sky, shear would have been measured with such high signal-to-noise. Secondly, we find that the signal-to-shot-noise $\mathrm{S/N_s}$ is not strongly dependent on the number of galaxies used in the analysis. We find that instead the shot noise is more dependent on the galaxies that have been selected in the analysis, but note that this statement is unlikely to apply to data where the shear varies. Taking Im2shape (SB) and BJ02 (MJ) as an example, we find ~ 2 times as many galaxies selected for the Im2shape (SB) analysis as for the BJ02 (MJ) analysis, but very similar values for the signal-to-shot-noise $\mathrm{S/N_s}$. As discussed in section 3, the distribution of galaxy ellipticities is strongly non-Gaussian with more intrinsically round galaxies than is seen in real data. The galaxy selection of Im2shape (SB) results in a smaller proportion of these intrinsically round galaxies being included in the analysis increasing the 1σ variation of the selected galaxy ellipticities. Several of the KSB+ analyses make galaxy selection based on galaxy ellipticity, removing the most elliptical galaxies, again this reduces the shot noise, independent of the number of galaxies used in the analysis. Lastly, comparing the signal-to-shot-noise $\mathrm{S/N_s}$ and the signal-to-weighted-shot-noise $\mathrm{S/N_s'}$, we see the effectiveness of some of the weighting schemes used in this analysis. The BJ02 weighting scheme (MJ, RN) puts more weight on the intrinsically round galaxies, this effective weighting scheme produces the highest signal-to-noise measurements in the STEP analysis, although see Section 5.6 for the implication of using this aggressive weighting scheme.

4.1 Calibration bias and PSF contamination

In this section, we measure the levels of multiplicative calibration bias and additive PSF contamination in each authors' shear measurement. Calibration bias will result from a poor correction for the atmospheric seeing that circularizes the images. Selection bias and weight bias are also forms of calibration bias which we investigate further in Sections 4.2 and 4.3. PSF contamination will result from a poor correction for the PSF distortion that coherently smears the image.

We calculate the mean shear γ_i for each data set as described above. For each author and PSF type, we then determine, from the

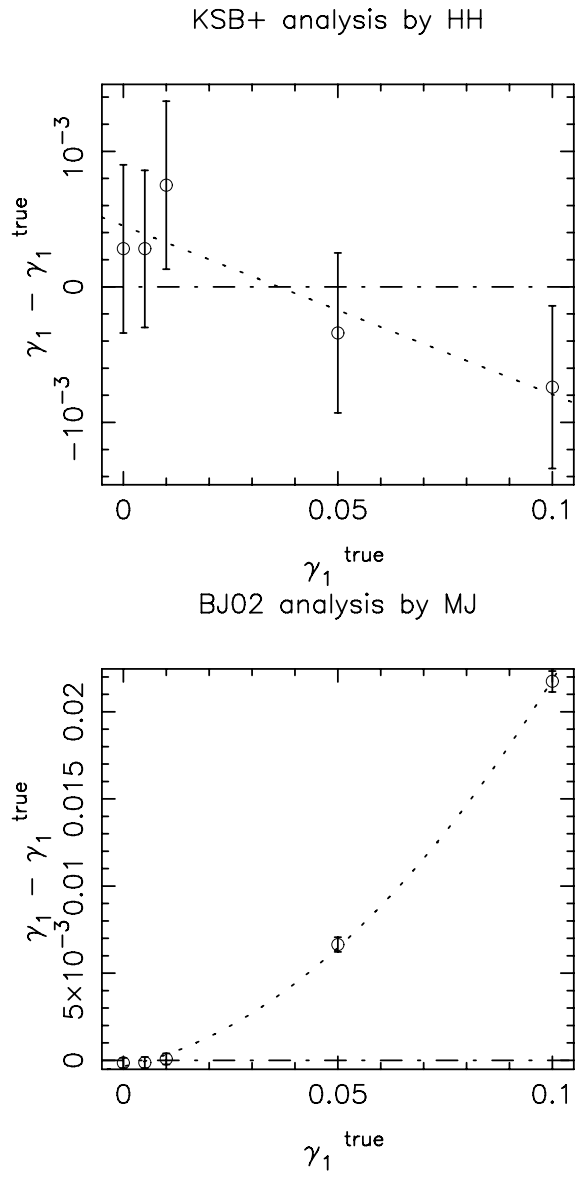

Figure 2. Examples of two analyses of PSF 3 simulations using KSB+ (HH implementation, upper panel) and BJ02 (MJ implementation, lower panel) comparing the measured shear γ_1 and input shear $\gamma_1^{\rm true}$. The best-fitting to equation (11) is shown dashed, and the optimal result (where $\gamma_1 = \gamma_1^{\rm true}$) is shown dot–dashed. Both analyses have additive errors that are consistent with shot noise (fitted y-offset parameter c) and low 1 per cent calibration errors (fitted slope parameter m). The weighting scheme used in the BJ02 analysis introduces a non-linear response to increasing input shear (fitted quadratic parameter q), reducing the shear recovery accuracy for increasing shear. The accuracy of the KSB+ analysis responds linearly to increasing input shear and so these results were refit with a linear relationship, i.e. $q = 0$.

range of sheared images, the best-fitting parameters to

$$\gamma_1 - \gamma_1^{\rm true} = q\left(\gamma_1^{\rm true}\right)^2 + m\gamma_1^{\rm true} + c_1, \qquad (11)$$

where $\gamma_1^{\rm true}$ is the external shear applied to each image. Fig. 2 shows fits to two example analyses of PSF 3 simulations using KSB+ (HH implementation) and BJ02 (MJ implementation). In the absence of calibration bias, we would expect $m = 0$. We would also expect $c_1 = 0$ in the absence of PSF systematics and shot noise, and $q = 0$ for a linear response of the method to shear. In the case where the fitted parameter q is consistent with zero, we refit with a linear relationship, as demonstrated by the KSB+ example in Fig. 2.

For all simulations the external applied shear $\gamma_2^{\rm true} = 0$ and we therefore also measure for each PSF type $c_2 = \langle\gamma_2\rangle$, averaged over the range of sheared images. In the absence of PSF systematics and shot noise, we would expect to find $c_2 = 0$. From this analysis, we found the values of m and q to be fairly stable to changes in PSF type and we therefore define a measure of calibration bias to be $\langle m\rangle$ and a measure of non-linearity to be $\langle q\rangle$ where the average is taken over the six different PSF sets. We find the value of $\langle c_i\rangle$ averaged over the six different PSF sets to be consistent with shot noise at the 0.1 per cent level for all authors, with the highest residuals seen with PSF model 1 (coma) and PSF model 2 (jitter). We therefore define σ_c as a measure of our ability to correct for all types of PSF distortions, where σ_c^2 is the variance of c_1 and c_2 as measured from the six different PSF models. As the underlying galaxy distributions are the same for each PSF this measure removes most of the contribution from shot noise, although the galaxy selection criteria will result in slightly different noise properties in the different PSF data sets. σ_c therefore provides a good estimate of the level of PSF residuals in the whole STEP analysis. A more complicated set of PSF distortions will be analysed in Massey et al. (in preparation) to address the issue of PSF-dependent bias more rigorously.

Fig. 3 shows the measures of PSF residuals σ_c and calibration bias $\langle m\rangle$ for each author, where the author key is listed in Table 2. For the non-linear cases where $q \neq 0$, denoted with a circle, the best-fitting $\langle q\rangle$ parameter is shown with respect to the right-hand scale. Results in the shaded region suffer from less than 7 per cent calibration bias. All methods which have been used in a cosmological parameter cosmic shear analysis lie within this region. With regard to PSF contamination, these results show that PSF residuals are better than 1 per cent in all cases and are typically better than 0.1 per cent. Note that for clarity the results plotted in Fig. 3 are also tabulated in Table 5.

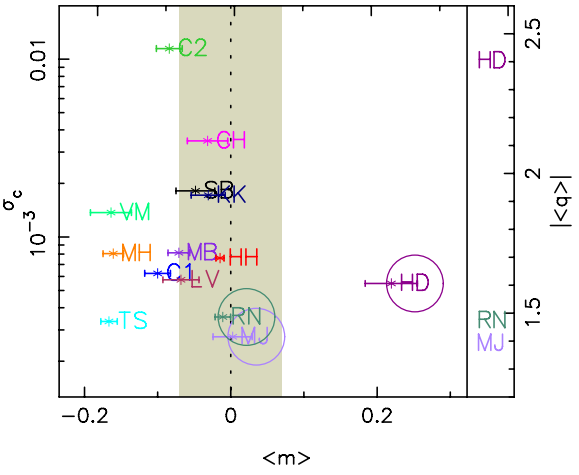

Figure 3. Measures of calibration bias $\langle m\rangle$, PSF residuals σ_c and non-linearity $\langle q\rangle$ for each author (key in Table 2), as described in the text. For the non-linear cases where $\langle q\rangle \neq 0$ (points enclosed within a large circle), $\langle q\rangle$ is shown with respect to the right-hand scale. In short, the lower the value of σ_c, the more successful the PSF correction is at removing all types of PSF distortion. The lower the absolute value of $\langle m\rangle$, the lower the level of calibration bias. The higher the q value the poorer the response of the method to stronger shear. Note that for weak shear $\gamma < 0.01$, the impact of this quadratic term is negligible. Results in the shaded region suffer from less than 7 per cent calibration bias. These results are tabulated in Table 5.

Table 5. Tabulated measures of calibration bias $\langle m \rangle$, PSF residuals σ_c and non-linearity $\langle q \rangle$ for each author (key in Table 2), as described in Section 4.1 and plotted in Fig. 3. For the non-linear cases where $\langle q \rangle \neq 0$, $\langle q \rangle$ is listed. 'Uncontaminated' calibration bias $\langle m_{\rm uncontaminated} \rangle$ is measured from object catalogues cleansed from stellar contamination and false object detections. This can be compared to the measured selection bias $\langle m_{\rm selc} \rangle$ as described in Section 4.2 and plotted in Fig. 4. Weight bias $\langle m_{\rm weight} \rangle$, described in Section 4.3, is also tabulated. For reference, the final column lists which pipelines have been used in cosmic shear analyses that have resulted in measurements of the amplitude of the matter power spectrum, σ_8, as detailed in Table 1.

Author	$\langle m \rangle$	σ_c	$\langle q \rangle$	$\langle m_{\rm uncontaminated} \rangle$	$\langle m_{\rm selc} \rangle$	$\langle m_{\rm weight} \rangle$	σ_8 analysis ?
SB	-0.048 ± 0.027	0.0018	–	-0.017 ± 0.030	0.006 ± 0.004	0.007 ± 0.002	×
MB	-0.071 ± 0.015	0.0008	–	-0.009 ± 0.021	-0.008 ± 0.002	–	√
C1	-0.100 ± 0.018	0.0006	–	-0.090 ± 0.018	-0.046 ± 0.022	0.011 ± 0.004	×
C2	-0.084 ± 0.018	0.0115	–	-0.074 ± 0.018	-0.045 ± 0.022	0.010 ± 0.003	×
HD	0.219 ± 0.036	0.0005	-2.40 ± 0.27	0.217 ± 0.028	-0.021 ± 0.006	–	×
MH	-0.161 ± 0.014	0.0008	–	-0.142 ± 0.015	-0.017 ± 0.001	0.032 ± 0.003	×
CH	-0.032 ± 0.028	0.0035	–	0.004 ± 0.027	-0.010 ± 0.003	–	√
HH	-0.015 ± 0.006	0.0008	–	0.018 ± 0.004	-0.001 ± 0.001	0.006 ± 0.001	√
MJ	0.002 ± 0.027	0.0003	1.39 ± 0.23	0.011 ± 0.027	0.005 ± 0.006	–	√
KK	-0.031 ± 0.023	0.0017	–	-0.029 ± 0.023	0.006 ± 0.003	0.020 ± 0.002	×
VM	-0.164 ± 0.028	0.0014	–	-0.116 ± 0.021	-0.015 ± 0.006	-	×
RN	-0.011 ± 0.011	0.0004	1.47 ± 0.09	0.001 ± 0.013	-0.037 ± 0.009	–	×
TS	-0.167 ± 0.011	0.0003	–	-0.158 ± 0.010	-0.045 ± 0.006	0.024 ± 0.003	×
LV	-0.068 ± 0.025	0.0006	–	-0.068 ± 0.025	-0.001 ± 0.002	0.005 ± 0.001	√

In the weak $\gamma \leqslant 0.01$ regime, the most successful method is found to be the BJ02 technique (MJ,RN) producing percent level accuracy. For stronger shear distortions, however, this methodology breaks down which can be seen from the high $\langle q \rangle$ value. This method is therefore unsuitable for low redshift cluster mass reconstructions where shear distortions of $\sim$10 per cent are not uncommon, although see the discussion in Section 5.6 for a solution to this issue of non-linearity. Over the full range of shear distortions tested, $0 < \gamma < 0.1$, the most successful method is found to be the Hoekstra implementation of the Kaiser et al. (1995) method (KSB+), producing results accurate to better than 2 per cent. All KSB+ pipelines are accurate to better than $\sim$15 per cent but the wide range of accuracy in these results that are based on the same methodology is somewhat disconcerting. It is believed that this spread results from the subtly different interpretation and implementation of the KSB+ method which we detail in the Appendix A. The results from the Dahle implementation of K2K (HD) are non-linear, suffering from calibration bias at $\sim$20 per cent level for weak shear $\gamma < 0.01$. The Wittman/Margoniner method (VM) (see Section 2.5) fares as well as the Hetterscheidt (MH) and Schrabback (TS) implementation of KSB+ with an accuracy of $\sim$15 per cent. Im2shape (Bridle et al. 2002) (SB) and the Kuijken (2006) (KK) implementation of shapelets typically fare as well as the methods used in cosmological parameter cosmic shear analyses with an accuracy of $\sim$4 per cent.

4.2 Selection bias

Selection bias is an issue that is potentially problematic for many different types of survey analysis. With weak lensing analyses, which relies on the fact that when averaging over many galaxies, the average source galaxy ellipticity $\langle e^{(\rm s)} \rangle = 0$, removing even weak selection biases is particularly important. When compiling source catalogues one should therefore consider any forms of selection bias that may alter the mean ellipticity of the galaxy population. This bias could arise at the source extraction stage if there was a preference to select galaxies oriented in the same direction as the PSF (Kaiser 2000) or galaxies that are anticorrelated with the gravitational shear (and as a result appear more circular) (Hirata & Seljak 2003). Selection criteria applied after source extraction could also bias the mean ellipticity of the population if the selection has any dependence on galaxy shape. In this section, we determine the level of selection bias by measuring the unweighted mean intrinsic source ellipticity $\langle e^{(\rm s)} \rangle$ (unlensed, equations 1 and 10) from the 'real' galaxies selected by each author for inclusion in their shear catalogue (false detections are thus excised from the catalogue at this stage). We follow a similar analysis to Section 4.1, by determining for each author and each PSF type, from the range of sheared images, the best-fitting parameters to

$$\begin{aligned} \left\langle e_1^{(\rm s)} \right\rangle_{\rm selc} &= m_{\rm selc} \gamma_1^{\rm true} + c_1^{\rm s}, \\ \left\langle e_2^{(\rm s)} \right\rangle_{\rm selc} &= c_2^{\rm s}. \end{aligned} \tag{12}$$

$\langle m_{\rm selc} \rangle$ averaged over the six different PSF data sets gives a measure of the shear-dependent selection bias and $(\sigma_c^{\rm s})^2$, the variance of $c_1^{\rm s}$ and $c_2^{\rm s}$ as measured from the six different PSF models, gives a measure of the PSF-anisotropy-dependent selection bias. We find that PSF-anisotropy-dependent selection bias is very low with $\sigma_c^{\rm s} < 0.001$ for all methods. Shear-dependent selection bias is <1 per cent in most cases with some notable exceptions in the cases of Clowe (C1 & C2), Schrabback (TS), Dahle (HD) and Nakajima (RN) as shown on the vertical axis of Fig. 4. The significant variation between the different PSF data sets of $m_{\rm selc}$ measured with the Clowe (C1 & C2) catalogues suggests that the selection criteria of this method are affected by the PSF type.

Fig. 4 also shows the value of $\langle m_{\rm uncontaminated} \rangle$ determined from equation (11) using the authors' measured shear catalogues now cleansed of false detections and stellar contamination, with author-defined object weights. With unbiased weights and an unbiased shear measurement method (where the shear is measured accurately but the source selection criteria are potentially biased), points should fall along the 1:1 line plotted. We can therefore conclude from Fig. 4 that in many cases the calibration bias seen in Section 4.1 cannot be solely attributed to selection bias. See Section 5 for a discussion on sources of selection bias. The results plotted in Fig. 4 are also tabulated in Table 5. Comparing the calibration biases measured from the original catalogues $\langle m \rangle$ in Section 4.1, and from the 'uncontaminated' catalogues $\langle m_{\rm uncontaminated} \rangle$ shows the impact of false detections and stellar contamination in each authors' catalogue. Typically the impact is low with <3 per cent changes found for the average

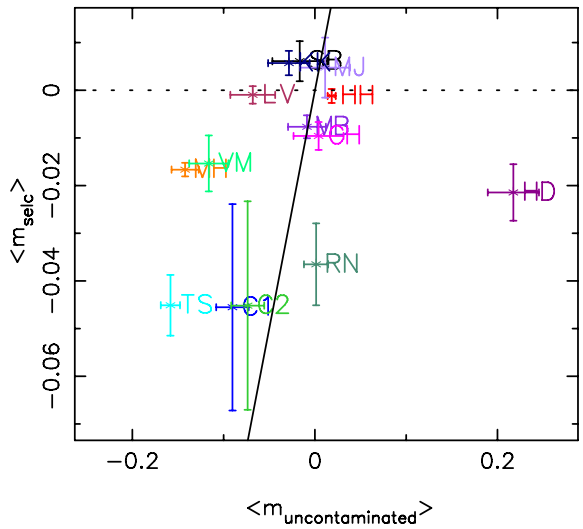

Figure 4. Measures of selection bias $\langle m_{\rm selc}\rangle$ for each author (key in Table 2), as described in the text. The lower the absolute value of $\langle m_{\rm selc}\rangle$ the lower the level of selection bias. Selection bias can be compared to the calibration bias $\langle m_{\rm uncontaminated}\rangle$ measured from catalogues cleansed of false detections and stellar contamination. Unbiased shear measurement methods, where the shear is measured accurately but the source selection criteria are potentially biased, would fall along the 1:1 line overplotted. These results are tabulated in Table 5.

measured shear of most authors. One noticeable exception is the result from the Brown (MB) pipeline, where the underestimation of the shear by ∼7 per cent is found to be predominantly caused by the diluting ∼7 per cent stellar contamination in the object catalogues.

4.3 Weight bias

In this section, we investigate the impact of the different object-dependent weighting schemes used by Bridle (SB), Clowe (C1 & C2), Hetterscheidt (MH), Hoekstra (HH), Kuijken (KK), Schrabback (TS) and Van Waerbeke (VW). All other methods use unit weights, except for the methods of Jarvis (MJ) and Nakajima (RN) which will be discussed at the end of this section. An optimal weighting scheme should reduce the noise on a measurement without biasing the results. Using the author defined weights, we compare the average unweighted and weighted mean intrinsic galaxy ellipticity, performing a similar analysis to Sections 4.1 and 4.2. For each author and PSF type, we calculate from the range of sheared images, the best-fitting parameters to

$$\left\langle e_1^{(\rm s)}\right\rangle_{\rm selc} - \left\langle e_1^{(\rm s)}\right\rangle'_{\rm selc} = m_{\rm weight}\gamma_1^{\rm true} + c_1^{\rm w}, \qquad (13)$$

where $\langle e_1^{(\rm s)}\rangle_{\rm selc}$ is an unweighted average and $\langle e_1^{(\rm s)}\rangle'_{\rm selc}$ is a weighted average. In the absence of PSF-dependent weight bias, $c_1^{\rm w}$ should be consistent with zero and we find this to be the case for all the weighting schemes tested. In the absence of shear-dependent weight bias, $m_{\rm weight}$ should be consistent with zero. All weighting schemes are found to introduce low percent level bias as shown in Table 5, where $\langle m_{\rm weight}\rangle$ is averaged over the six different PSF models. In most cases these biases are small (<2 per cent) and we can therefore conclude the cases of calibration bias seen in Section 4.1 cannot be solely attributed to weight bias. For percent level precision in future analyses the issue of weight bias will need to be considered.

The Jarvis (MJ) and Nakajima (RN) analyses make use of the ellipticity-dependent weighting formulae in BJ02, Section 5. This weighting scheme takes advantage of the $e = 0$ peak in the shape distribution of galaxies to improve the signal-to-noise of weak shear measurement. This is evidenced by the high signal-to-noise results with the Jarvis (MJ) and Nakajima (RN) methods as listed in Table 4. Shearing the galaxies does change the assigned weights, but the BJ02 formulae explicitly account for this effect via a factor called the responsivity. The non-linear response to shear seen in the results of the Jarvis (MJ) and Nakajima (RN) methods is an undesirable consequence of this weighting scheme which we discuss further in Section 5.6.

4.4 Shear measurement dependence on galaxy properties

The simulations analysed in this paper were sheared uniformly across the field-of-view. In reality however, the gravitational shear experienced by each galaxy is dependent on position and more importantly redshift. High redshift galaxies have a lower apparent magnitude and smaller angular size when compared to their lower redshift counterparts. It is therefore important that shear measurement methods are stable to changes in galaxy magnitude and size. For each author, we measure the average shear as a function of magnitude and input disc size. In general, we find that the average shear binned as a function of magnitude and disc size varies <1 per cent to the average shear measured from the full data set, and an example plot of shear measured as a function of galaxy magnitude is shown from the KSB+ implementation of HH in Fig. 5. The dot–dashed line shows the average $\gamma_1 - \gamma_1^{\rm true}$ measured from the full galaxy sample which is dominated by the faint magnitude galaxies. For this particular analysis the shear measured from bright galaxies is slightly underestimated, and the shear from faint galaxies is slightly overestimated. The reader should note however that the shear measured from each magnitude bin is $<1\sigma$ from the average for all but one case and that for weaker input shears, this effect is even less prominent.

Investigating the dependence of shear on galaxy properties, we found that some methods introduced correlations between shear and

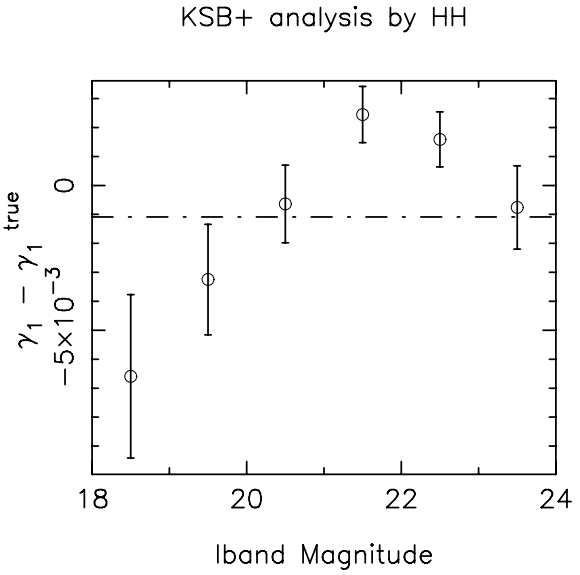

Figure 5. An example plot of the difference between measured shear γ_1 and input shear $\gamma_1^{\rm true}$ as a function of galaxy I-band magnitude. This plot is taken from the KSB+ analysis of HH using the PSF 0 simulations with an input shear $\gamma_1^{\rm true} = 0.05$. The dot–dashed line shows the average $\gamma_1 - \gamma_1^{\rm true}$ measured from the full galaxy sample.

magnitude, whilst others between shear and disc size. Interestingly however all methods revealed very different dependencies on galaxy properties that we were unable to directly parameterize. As such we cannot fully address the issue of shear measurement dependence on galaxy properties at this time. For percent level precision in future analyses this issue will certainly need to be revisited and it will be addressed further in future STEP projects using simulations with constant shear and constant galaxy magnitude.

5 DISCUSSION

In this section, we discuss some of the lessons that we have learnt from the first STEP initiative and highlight the areas where we can improve our methods in future analyses.

5.1 KSB+

The subtle differences between the eight tested KSB+ pipelines, detailed in the Appendix A, introduces an interesting spread in the KSB+ results. Using the information in the Appendix A, KSB+ users can now modify pipelines to improve their results. The different ways of implementing KSB+ and the effect of using different methods will be discussed in more detail in a future paper (Hetterscheidt et al., in preparation), but comparing methods and results makes clear which interpretations of the KSB+ method are best for ground-based data. A good example of this is the PSF correction method of Heymans (CH) and Clowe (C2) where the correction is calculated as a function of galaxy size. For ground-based data where the PSF ellipticity is fairly constant at all isophotes (although note that this was not the case with PSF 2), a PSF correction determined only at the stellar size produces a less noisy and more successful PSF correction, as shown by the success of the PSF correction by other KSB+ users. This however would not necessarily be the case for space-based data where the PSF ellipticity varies with size (see for example Heymans et al. 2005) which will be tested in a future STEP analysis of simulated space-based observations. The Schrabback (TS) method produces a more successful size-dependent PSF correction by limiting the image region about stellar objects over which the PSF correction parameter $p_\mu(r_g)$ is calculated ($\theta_{max} = 3r_g^*$, see Section A2). This measure reduces the noise on $p_\mu(r_g)$ thus improving the overall correction.

For several methods selection bias is well below the percent level from which we can conclude that current source detection methods are suitable for weak lensing analyses and that any selection bias seen with other methods has been introduced after the source extraction stage. The first clue to understanding the selection bias, we see in some cases comes from comparing $\langle m_{\rm selc}\rangle$ for the Hetterscheidt (MH) and Schrabback (TS) results in Fig. 4. These two analyses stem from the same SEXTRACTOR catalogue. The main differences between these two methods are the technique used to correct for the PSF distortion and the catalogue selection criteria where Schrabback (TS) places more conservative cuts on galaxy size defined by the FLUX_RADIUS parameter of SEXTRACTOR. Whilst there is no correlation within the simulations for intrinsic galaxy ellipticity with disc size, we find that the measured HFINDPEAKS r_g parameter and the measured SEXTRACTOR FLUX_RADIUS and FWHM parameters are somewhat correlated with galaxy ellipticity. For this reason galaxy size selection criteria based on r_g, FLUX_RADIUS or FWHM will introduce a bias. This finding is one of the first lessons learnt from this STEP initiative which can now be improved upon in future STEP analyses.

5.2 K2K

The Dahle (HD) K2K results appear noisier than other pipelines which could result from an upper significance cut in order to remove big, bright galaxies, which in real data are at low redshift unlensed galaxies. This step rejects ∼24 per cent of the objects. The method is optimized for mosaic CCD data with a high number of galaxies for each exposure, it therefore suffers somewhat from the low number of objects in each 4096 × 4096 STEP image. In addition, as a space-saving measure, images were stored in integer format, this may have introduced some extra noise in the 'recircularized' images. In considering the success of K2K applied to STEP simulations one should keep in mind that the man-hours invested in testing and fine-tuning KSB+ is at least an order of magnitude more than for any of the other methods. With the STEP simulations future tests and optimization are now feasible, the results of which will be demonstrated with the next STEP analysis of shapelet-based image simulations.

5.3 Shapelets

In the first, blind Kuijken (KK) analysis of the simulations all sources were fitted to eighth order in shapelets, which gives a good fit to the PSF-convolved sources. This, however, resulted in a systematic underestimate of the shear amplitude of some 10 per cent. Later investigation showed that even without any PSF smearing or noise, the ellipticity of an exponential disc is only derived correctly if the expansion is extended to 12th order. As this method has, to date, not been used in scientific analyses, it was decided that a re-analysis of the simulations with 12th order shapelets would be permitted. The results of the non-blind re-analysis are shown in this paper. Using the higher order shapelet terms removed the systematic underestimate for the high S/N sources. There is still a tendency for noisy sources to have their ellipticities underestimated however and this is still under investigation.

5.4 Im2shape

Im2shape uses MCMC sampling to fit elliptical Gaussians to the image. Before the STEP analysis it was believed that using too few iterations in the MCMC analysis would add noise to the ellipticities of each galaxy but would not systematically bias them. It became apparent during this STEP analysis however, that a bias is in fact introduced as the number of iterations is decreased. The number of iterations was chosen by systematically increasing the number of iterations in the analysis of a subsample of the data until the measured average shear converged.

5.5 Wittman method with ELLIPTO

A post-STEP analysis of the shape catalogue revealed that the measured galaxy shape distribution resulting from this method had rather asymmetric tails. The core of the distribution reflected the shear much more accurately than did the mean of the entire distribution. This method could thus be greatly improved by some type of weighting or robust averaging scheme. For example, a simple iterative 3σ clip reduced the 15 per cent underestimate of the strongest applied shear, where $\gamma = 0.1$, to an 8 per cent underestimate, while rejecting only 2.2 per cent of the sources. A slightly harsher clip at 2.8 σ further reduced the underestimate to 3.5 per cent, while still rejecting only 3.9 per cent of the sources. The stellar contamination rate of

3.8 per cent is presumably responsible for the remaining underestimate. Note that the real data to which this method has been applied is much deeper than the STEP simulations. The stellar contamination rate would therefore be much lower, as the galaxy counts rise more steeply with magnitude in comparison to the star counts.

Of course, one would prefer to understand the origin of the asymmetric outliers rather than simply clipping them at the end. A brief analysis shows that they are not highly correlated with the obvious variables such as photometric signal-to-noise or size relative to the PSF. Therefore, a simple inverse-variance weighting scheme would not be enough to solve the problem. The prime task for improving this method would thus be understanding the cause of this asymmetric tail and developing a mitigation scheme.

5.6 Bernstein & Jarvis method: BJ02

The ellipticity-dependent weighting scheme of BJ02 is responsible for the significant increase in the signal-to-noise of the STEP shear measurements, as shown in Table 4. It has, however, also been found to be the cause of the non-linear response of the Jarvis (MJ) and Nakajima (RN) methods to shear. After the blind testing phase, the results of which are shown in this paper, Jarvis (MJ) reran the analysis with shape-independent weights finding a linear response to the range of weak shears tested such that the non-linearity parameter, q, measured by equation (11) became consistent with zero. The signal-to-noise dropped, however, by a factor of 1.5. We can thus recommend that weak shear studies use aggressive weights which help to probe small departures of $\langle\gamma\rangle$ from zero, while studies of stronger shear regions use unweighted measurements to minimize the effects of non-linearity.

The false detections in the Nakajima (RN) analysis were investigated and found to be either double objects detected by SEXTRACTOR as a single object or diffraction spikes. Double object detections could be reduced by varying SEXTRACTOR parameters to encourage the deblending of overlapping sources. When the data is taken in several exposures an additional measure to reduce the number of false detections can be introduced. This approach, taken by Jarvis et al. (2003), demands that a source is detected in at least two of the four exposures taken of each field. The STEP simulations were single exposure images and so this procedure could not be implemented. These false detections will generally be faint and highly elliptical in the case of diffraction spikes. Thus, with the weighting scheme implemented in both the Jarvis (MJ) and Nakajima (RN) analyses, these down-weighted objects do not affect the overall average measured shear.

6 CONCLUSION

In this paper, we have presented the results of the first Shear TEsting Programme, where the accuracy of a wide range of shear measurement methods were assessed. This paper has demonstrated that, for smooth galaxy light profiles, it is currently feasible to measure weak shear at percent level accuracy using the Bernstein & Jarvis (2002) method (BJ02) and the Hoekstra implementation of the KSB+ method. It has also shown how important it is to verify shear measurement software with image simulations as subtle differences between each individuals implementation can result in discrepancy. We therefore strongly urge all weak lensing researchers to subject their pipelines to a similar analysis to ensure high accuracy and reliability in all future weak lensing studies. To this end the STEP simulations will be made available on request.

The removal of the additive PSF anisotropic distortion has been successful in all methods, reduced to an equivalent shear of $\sim$0.001 in most cases. Significant calibration bias is however seen in the results of some methods which can be explained only in part by the use of biased weights and/or selection bias. Using the simulations analysed in this paper, errors can now be pin-pointed and corrected for, and modifications will be introduced to remove sources of calibration error. For authors using the KSB+ method, detailed descriptions have been given of each pipeline tested in this analysis to aid the improvement and development of future KSB+ methods. One positive aspect of the KSB+ method is that its response to shear has been shown to be very linear. This is contrast to the BJ02 method tested in this paper, where the ellipticity-dependent weighting scheme was found to introduce a non-linear response to shear. For this reason KSB+ or an unweighted version of the BJ02 method is currently the preferred method for measuring weak shear around nearby galaxy clusters. Cosmic shear, on average, is very weak, but with the next generation of cosmic shear surveys covering large areas on the sky and thus imaging regions of both high and low shear, cosmic shear measurement also requires a method that is linear in its response to shear. Thus KSB+ or an unweighted version of the BJ02 method is currently the preferred cosmic shear measurement method. In the weakest regime of galaxy–galaxy lensing, the weighted BJ02 method measures shear at a higher signal-to-noise with a better accuracy than KSB+ and thus appears to be the most promising of the methods that have been tested in this analysis for galaxy–galaxy lensing studies.

Selection bias has been shown to be consistent with zero in some cases, from which we can conclude that current source detection methods are suitable for weak lensing analyses. Some object weighting schemes were found to be unbiased at the below percent level. The use of such schemes may however require revision in the future when low level biases become important. All the methods tested were found to exhibit rather different <1 per cent dependences on galaxy magnitude and size. For real data where shear scales with depth and hence magnitude and size, these issues will need to be addressed.

In this paper, we have provided a snapshot view of how accurately we can measure weak shear today from galaxies with relatively simple galaxy morphologies. We are unable to answer the question, what method ought I to use to measure weak lensing shear? KSB+, used with care, and BJ02 clearly fare well, but some of the methods tested here that are currently still in their development stage may still provide a better method in the future. For the cosmic shear, galaxy–galaxy lensing and cluster-mass determinations published to date, $\leqslant$7 per cent calibration errors are within statistical errors and are certainly not dominant. $\sigma_c < 0.01$ is also small enough to be subdominant in present work. We voice caution in explaining the $\sim 2\sigma$ differences in cosmological parameter estimation from cosmic shear studies by the scatter in the results that we find in this analysis. The true reason is likely to be more complex involving source redshift uncertainties, residual systematics and sampling variance in addition to the calibration errors we have found. Many of these sources of error will be significantly reduced with the next generation of surveys where the large areas surveyed will minimize sampling variance and the multicolour data will provide a photometric redshift estimate of the source redshift distribution. The now widespread use of diagnostic tools to determine levels of non-lensing residual distortions also allows for the quantification and reduction in systematic errors. Calibration errors, however, can only be directly detected through the analysis of image simulations.

This first STEP analysis has quantified the current levels of calibration error, allowing for improvement in calibration accuracy in future shear measurement methods. The upcoming next generation

of wide-field multicolour optical surveys will reduce statistical errors on various shear measurements to the $\sim$2 per cent level, requiring calibrations accurate to $\sim$1 per cent. In the next decade, deep weak-lensing surveys of thousands of square degrees will produce shear measurements that will be degraded by calibration accuracies $\gtrsim$0.1 per cent, well below even the precision of the current STEP tests. Similarly the additive errors represented by σ_c will ultimately have to be reduced to a level of $\sigma_c < \approx 10^{-3.5}$ if this spurious signal is to be below the measurement limits imposed by cosmic variance of full-sky surveys. The collective goal of the weak lensing community is now to meet these challenges.

The next STEP project will analyse a set of ground and space-based image simulations that include complex galaxy morphologies using a 'shapelet' composition (Massey et al. 2004b). Initial tests with shapelet simulations suggest that complex morphology rather complicates weak shear measurement for methods that assume Gaussian light profiles. Further STEP projects will address the issue of PSF interpolation and modelling, and the impact of using different data reduction and processing techniques (Erben et al. 2005). These future STEP projects will be as important as this first STEP analysis in order to gain more understanding and further improve the accuracy of our methods. We conclude with the hope that by using the shared technical knowledge compiled by STEP, all future shear measurement methods will be able to reliably and accurately measure weak lensing shear.

ACKNOWLEDGMENTS

We thank TERAPIX (Traitement Élémentaire, Réduction et Analyse des Pixels de megacam) at the Institut d'Astrophysique de Paris for hosting the SKYMAKER simulations. We also thank the Max-Planck-Institut für Astronomie for financial and administrative support of STEP teleconferencing, and the Jet Propulsion Laboratory for financial and administrative support of the STEP workshop. CH is supported by a CITA National fellowship and acknowledges financial support from GIF. DB and MLB are supported by PPARC fellowships. SB used the UK National Cosmology Supercomputer Centre funded by PPARC, HEFCE and Silicon Graphics/Cray Research. HD is funded by a post-doctoral fellowship from the research council of Norway. TE acknowledges support from the German Science Foundation (DFG) under contract ER 327/2-1. KK acknowledges financial support provided through the European Community's Human Potential Program under contract HPRN-CT-2002-00316, SISCO. We thank Richard Ellis for helpful discussions about the STEP project and the referee for useful comments.

REFERENCES

Bacon D., Refregier A., Ellis R., 2000, MNRAS, 318, 625
Bacon D. J., Refregier A., Clowe D., Ellis R. S., 2001, MNRAS, 325, 1065
Bacon D., Massey R., Refregier A., Ellis R., 2003, MNRAS, 344, 673
Bardeau S., Kneib J.-P., Czoske O., Soucail I., Smail I., Ebeling H., Smith G., 2005, A&A, 434, 433
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Benabed K., Van Waerbeke L., 2004, Phys. Rev. Lett., 70, 123 515
Bernstein G., 2006, ApJ, 637, 598
Bernstein G., Jain B., 2004, ApJ, 600, 17
Bernstein G. M., Jarvis M., 2002, AJ, 123, 583
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Binney J., de Vaucouleurs G., 1981, MNRAS, 194, 679
Binney J., Merrifield M., 1998, Princeton Ser. Astrophys., Galactic Astronomy. Princeton Univ. Press, Princeton, NJ
Bonnet H., Mellier Y., 1995, A&A, 303, 331
Boroson T., 1981, ApJS, 46, 177
Bridle S., Gull S., Bardeau S., Kneib J. P., 2002, in Natarajan P., ed., Proc. Yale Cosmology Workshop, The Shapes of Galaxies and Their Dark Halos. World Scientific, New Haven
Brown M., Taylor A., Bacon D., Gray M., Dye S., Meisenheimer K., Wolf C., 2003, MNRAS, 341, 100
Chang T., Refregier A., Helfand D. J., 2004, ApJ, 617, 794
Clowe D., Gonzalez A., Markevitch M., 2004, ApJ, 604, 596
Crittenden R., Natarajan R., Pen U., Theuns T., 2002, ApJ, 568, 20
Dahle H., Kaiser N., Irgens R. J., Lilje P. B., Maddox S. J., 2002, ApJS, 139, 313
Erben T., Van Waerbeke L., Bertin E., Mellier Y., Schneider P., 2001, A&A, 366, 717
Erben T. et al., 2005, Astron. Nachr., 326, 432
Fischer P., Tyson J. A., 1997, AJ, 114, 14
Gray M., Taylor A., Meisenheimer K., Dye S., Wolf C., Thommes E., 2002, ApJ, 568, 141
Grosbol P. J., 1985, A&AS, 60, 261
Hamana T. et al., 2003, ApJ, 597, 98
Heavens A., 2003, MNRAS, 343, 1327
Heymans C., Heavens A., 2003, MNRAS, 339, 711
Heymans C., Brown M., Heavens A., Meisenheimer K., Taylor A., Wolf C., 2004, MNRAS, 347, 895
Heymans C. et al., 2005, MNRAS, 361, 160
Hirata C., Seljak U., 2003, MNRAS, 343, 459
Hirata C. M. et al., 2004, MNRAS, 353, 529
Hoekstra H., 2004, MNRAS, 347, 1337
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., Gladders M., 2002a, ApJ, 577, 595
Hoekstra H., Yee H., Gladders M., Barrientos L. F., Hall P., Infante L., 2002b, ApJ, 572, 55
Hoekstra H., Yee H. K. C., Gladders M. D., 2004, ApJ, 606, 67
Hoekstra H. et al., 2005, A&A, submitted (astro-ph/0511089)
Huizinga J. E., van Albada T. S., 1992, MNRAS, 254, 677
Huterer D., Takada M., Bernstein G., Jain B., 2006, MNRAS, 366, 101
Ishak M., 2005, MNRAS, 363, 469
Jain B., Taylor A., 2003, Phys. Rev. Lett., 91, 141 302
Jarvis M., Jain B. 2004, ApJ, submitted (astro-ph/0412234)
Jarvis M., Bernstein G., Jain B., Fischer P., Smith D., Tyson J., Wittman D., 2003, ApJ, 125, 1014
Jarvis M., Bernstein B. J. G., Dolney D., 2005, ApJ, in press (astro-ph/0502243)
Kaiser N., 2000, ApJ, 537, 555
Kaiser N., Squires G., Broadhurst T., 1995, ApJ, 449, 460
King L., Schneider P., 2002, A&A, 396, 411
King L., Schneider P., 2003, A&A, 398, 23
Kuijken K., 1999, A&A, 352, 355
Kuijken K. 2006, A&A, submitted (astro-ph/0601011)
Lambas D. G., Maddox S. J., Loveday J., 1992, MNRAS, 258, 404
Luppino G. A., Kaiser N., 1997, ApJ, 475, 20
Mandelbaum R. et al., 2005, MNRAS, 361, 1287
Maoli R., Van Waerbeke L., Mellier Y., Schneider P., Jain B., Bernardeau F., Erben T., 2001, A&A, 368, 766
Margoniner V. E., Lubin L. M., Wittman D. M., Squires G. K., 2005, AJ, 129, 20
Massey R., Refregier A., 2005, MNRAS, 363, 197
Massey R., Refregier A., Bacon D., 2004a, in Mellier Y., Meylan M., eds, Proc. IAU Symp. 225, Impact of Gravitational Lensing on Cosmology. Cambridge Univ. Press, Cambridge
Massey R., Refregier A., Conselice C. J., David J., Bacon J., 2004b, MNRAS, 348, 214
Massey R., Bacon D., Refregier A., Ellis R., Brown M., 2005, MNRAS, 359, 1277
Refregier A., 2003, MNRAS, 338, 35
Refregier A., Bacon D., 2003, MNRAS, 338, 48
Refregier A., Rhodes J., Groth E. J., 2002, ApJ, 572, L131
Refregier A. et al., 2004, AJ, 127, 3102

Rhodes J., Refregier A., Groth E. J., 2000, ApJ, 536, 79
Rhodes J., Refregier A., Groth E. J., 2001, ApJ, 552, L85
Rhodes J., Refregier A., Collins N. R., Gardner J. P., Groth E. J., Hill R. S., 2004, ApJ, 605, 29
Rix H., Zaritsky D., 1995, ApJ, 447, 82
Sandage A., Freeman K. C., Stokes N. R., 1970, ApJ, 160, 831
Schneider P., Van Waerbeke L., Mellier Y., 2002, A&A, 389, 741
Seitz C., Schneider P., 1997, A&A, 318, 687
Sembolini E. et al. 2005, A&A, submitted (astro-ph/0511090)
Sheldon E. S. et al., 2004, AJ, 127, 2544
Smith D. R., 2000, PhD thesis, Univ. Michigan
Smith D. R., Bernstein G. M., Fischer P., Jarvis M., 2001, ApJ, 551, 643
Tyson J. A., Wenk R. A., Valdes F., 1990, ApJ, 349, L1
Van Waerbeke L., Mellier Y., 2003, Gravitational Lensing: A Unique Tool for Cosmology. Preprint (astro-ph/0305089)
Van Waerbeke L. et al., 2000, A&A, 358, 30
Van Waerbeke L. et al., 2001, A&A, 374, 757
Van Waerbeke L., Mellier Y., Hoekstra H., 2005, A&A, 429, 75
Wittman D., Tyson J., Margonier V. E., Cohen J. G., Dell'Antonio I., 2001, ApJ, 557, L89
Wittman D., Margoniner V. E., Tyson J. A., Cohen J. G., Becker A. C., Dell'Antonio I. P., 2003, ApJ, 597, 218

APPENDIX A: KSB+ IMPLEMENTATION

The KSB+ method, used by a large percentage of the authors, has been shown in this STEP analysis to produce remarkably different results. In this Appendix, to aid the future understanding of these differences, we detail how different authors have implemented KSB+ with their weak lensing pipelines, as summarised in Table A1.

A1 Source detection, centroids and size definitions

Most authors use the SEXTRACTOR software (Bertin & Arnouts 1996) to detect objects and define galaxy centroids. Exceptions are Hoekstra (HH) and Brown (MB) who use HFINDPEAKS from the IMCAT software. The Gaussian weight scale length r_g is then either set to the FLUX_RADIUS SEXTRACTOR parameter or the 'optimal' r_g value defined by HFINDPEAKS. Clowe (C1&2) uses both pieces of software using a version of HFINDPEAKS to determine the optimal weight scaling r_g that keeps the centroid fixed to the SEXTRACTOR co-ordinates. Hetterscheidt (MH) and Schrabback (TS) measure half light radii r_h and refine the SEXTRACTOR centroids using the iterative method described in Erben et al. (2001).

A2 Quadrupole moments and integrals

The weighted ellipticity ε (equation 3), and the smear and shear polarisability tensors $P^{\rm sm}$ and $P^{\rm sh}$ are calculated for each object using software developed from the IMCAT subroutine GETSHAPES. The continuous integral formula are calculated from the discrete pixelised data by approximating the integrals as discrete sums. The weighted ellipticity ε is calculated from the quadrupole moment which in its discrete form can be written as follows

$$Q_{ij} = \frac{\sum_{\theta_i,\theta_j=-\theta_{\rm max}}^{\theta_{\rm max}} \Delta\theta^2\, W(\theta_i,\theta_j)\, I(\theta_i,\theta_j)\,\theta_i\theta_j}{\sum_{\theta_i,\theta_j=-\theta_{\rm max}}^{\theta_{\rm max}} \Delta\theta^2\, W(\theta_i,\theta_j)\, I(\theta_i,\theta_j)}, \qquad \text{(A1)}$$

where θ is measured, in pixel units, from the source centroid. Table A1 lists each authors' chosen values for $\theta_{\rm max}$ and $\Delta\theta$. For real values of θ, the intensity $I(\theta_i, \theta_j)$, known at pixel positions, is estimated from a first-order interpolation over the four nearest pixels to (θ_i, θ_j) (denoted 'interpolation' in Table A1). The interpolation stage is by-passed by some authors by setting $\Delta\theta = 1$ pixel and approximating $I(\theta_i, \theta_j) \approx I(\mathrm{Int}[\theta_i], \mathrm{Int}[\theta_j])$ (denoted 'Approx' in Table A1), or by exchanging the value of θ, in the above formula, for its nearest integer value Int $[\theta]$ (denoted 'Integer' in Table A1). $P^{\rm sm}$ and $P^{\rm sh}$ are functions of weighted moments, up to fourth order, that include $\theta_i\theta_j$ terms. Some authors treat these second order terms in θ differently using the nearest integer values of θ (denoted 'Integer' in the $P^{\rm sh}$ and $P^{\rm sm}$ estimate column of Table A1).

A3 Anisotropic PSF modelling

Stellar objects are selected by eye from the stellar locus in a size-magnitude plane and are then used to produce a polynomial model of the PSF as a function of chip position. Hetterscheidt (MH), Heymans (CH) and Schrabback (TS) fit directly to p_μ (equation 5) which, in the case of Heymans (CH) and Schrabback (TS), is measured for varying r_g (Hoekstra et al. 1998). This is in contrast to Hetterscheidt (MH) who measures p_μ with $r_g = r_g^*$. Brown (MB), Clowe (C1& 2), Hoekstra (HH) and Van Waerbeke (LV) create models of $\varepsilon_\alpha^{*\rm obs}$, $P^{\rm sm*}$ and $P^{\rm sh*}$ separately where for Brown (MB), and the first Clowe method (C1) stellar shapes are measured with $r_g = r_g^*$. The second Clowe method (C2), the Hoekstra (HH) method and the Van Waerbeke (LV) method measures the stellar parameters for varying r_g. Note that the Van Waerbeke (LV) method fits each component of the $P^{\rm sm*}$ and $P^{\rm sh*}$ tensors. With PSF models in hand observed galaxy ellipticities are corrected according to equation (4).

A4 Isotropic P^γ correction

The application of the anisotropic PSF correction leaves an effectively isotropic distortion making objects rounder as a result of both the PSF and the Gaussian weight function used to measure the galaxy shapes. To correct for this effect and convert weighted galaxy ellipticities ε into unbiased shear estimators $\hat{\gamma}$, we use the pre-seeing shear polarisability tensor P^γ, equation (7). P^γ is calculated for each galaxy from the measured galaxy smear and shear polarisability tensors, $P^{\rm sm}$ and $P^{\rm sh}$, and a term that is dependent on stellar smear and shear polarisability tensors; $(P^{\rm sm^*})^{-1}_{\mu\delta}\, P^{\rm sh^*}_{\delta\beta}$. Brown (MB) and the first method of Clowe (C1) use the stellar smear and shear polarisability tensors measured with a Gaussian weight of scale size $r_g = r_g^*$. Hetterscheidt (MH), Heymans (CH), Hoekstra (HH), Schrabback (TS), Van Waerbeke (LV) and the second method of Clowe (C2) calculate this stellar term as a function of smoothing scale r_g. Comparing the C1 and C2 results therefore demonstrates the impact of the inclusion of scale size at this stage.

P^γ is a very noisy quantity, especially for small galaxies. This noise is reduced somewhat by treating P^γ as a scalar equal to half its trace (note that the off diagonal terms of P^γ are typically an order of magnitude smaller than the diagonal terms). None of the methods tested in this analysis uses the full P^γ tensor correction (see Erben et al. (2001) to compare the results achieved when using a tensor and scalar P^γ correction). In an effort to reduce the noise on P^γ still further, P^γ is often fit as a function of r_g, although note that this fitting process has recently been shown, with the Brown (MB) pipeline, to be dependent on which significance cuts are made when selecting galaxies (Massey et al. 2005). Table A1 details which method is used by each author. In the case of Clowe (C1&2), P^γ is also fit as a function of ε, and with the method of Van Waerbeke, P^γ is also fit as a function of magnitude.

In real data Hoekstra (HH) has previously found a clear dependence of $P^{\rm sh}$ on ε. To correct for this shape dependence the Hoekstra

Table A1. The stages implemented by different authors using the KSB+ method described in Section 2.1. Table notation; pix = pixel units; $P(r_g)$ implies that parameter P is measured as a function of scale size r_g; $P(r_g^*)$ implies that parameter P is measured at the stellar scale size r_g^*. See the Appendix text for more details.

KSB Author	Brown	Clowe	Clowe	Hetterscheidt
Key	MB	C1	C2	MH
Source Detection	HFINDPEAKS	HFIND + SEXT	HFIND + SEXT	SEXTRACTOR
PSF:	$2^{\rm nd}$order	$3^{\rm rd}$order	$3^{\rm rd}$order	$3^{\rm rd}$order
2D polynomial	to ε^* and	to ε^*, $P^{\rm sm*}$, $P^{\rm sh*}$	to ε^*, $P^{\rm sm*}$, $P^{\rm sh*}$	to p^μ (r_g^*)
Model	$P^{\rm sm*}$, $P^{\rm sh*}$		$(P^{\rm sh}/P^{\rm sm})(r_g)$	3.5σ clipping
Galaxy size r_g	from HFINDPEAKS	from HFINDPEAKS	from HFINDPEAKS	FLUX_RADIUS
Quadrupole estimate	Approx	Approx	Approx	Interpolation
$\theta_{\rm max}$ and $\Delta\theta$	Int[$4r_g$], 1 pix	Int[$3r_g$], 1 pix	Int[$3r_g$], 1 pix	$3r_g$, 0.25 pix
$P^{\rm sh}$ and $P^{\rm sm}$ estimate	Approx	Approx	Approx	Interpolation
P^γ correction	Fit of	Fit $P^\gamma_{ii}(r_g, e)$	Fit $P^\gamma_{ii}(r_g, e)$	$\frac{1}{2}{\rm Tr}[P^\gamma]$
	$\frac{1}{2}{\rm Tr}[P^\gamma](r_g)$	$(P^{\rm sh}/P^{\rm sm})(r_g^*)$	$(P^{\rm sh}/P^{\rm sm})(r_g)$	(no fit)
Weights	none	$\langle\gamma^2\rangle^{-1}(r_g, \nu)$	$\langle\gamma^2\rangle^{-1}(r_g, \nu)$	$\langle\gamma^2\rangle^{-1}(r_g, {\rm mag})$
γ correction	Calibration	Close-pair	Close-pair	
	$\gamma_{\rm cor} = \gamma/0.85$	$\gamma_{\rm cor} = \gamma/0.95$	$\gamma_{\rm cor} = \gamma/0.95$	
Ellipticity cut	$\lvert\varepsilon_{obs}\rvert \leqslant 0.5$			$\lvert\varepsilon_{obs}\rvert \leqslant 0.8$
Size cut	$r_g > r_g^*$	$r_g^* < r_g < 6$ pix	$r_g^* < r_g < 6$ pix	$r_h > r_h^*$
Significance cut	$\nu > 5$	$\nu > 10$	$\nu > 10$	
P^γ cut		$P^\gamma_{ii} \geqslant 0.15$	$P^\gamma_{ii} \geqslant 0.15$	$\frac{1}{2}{\rm Tr}[P^\gamma] > 0$
γ cut				
Other		$\lvert d\rvert < 1$ pix	$\lvert d\rvert < 1$ pix	$\lvert d\rvert < 3$ pix
		SEX class <0.8	SEX class <0.8	
		No sat/bad pix	No sat/bad pix	

KSB Author	Heymans	Hoekstra	Schrabback	Van Waerbeke
Key	CH	HH	TS	LV
Source Detection	SEXTRACTOR	HFINDPEAKS	SEXTRACTOR	SEXTRACTOR
PSF:	$2^{\rm nd}$order	$2^{\rm nd}$order	$3^{\rm rd}$order	$2^{\rm nd}$order
2D polynomial	to $p^\mu(r_g)$ and	to $\varepsilon^*(r_g)$,	to $p^\mu(r_g)$	to $\varepsilon^*(r_g)$
Model	$(P^{\rm sm*})^{-1}_{\alpha\beta}\, P^{sh*}_{\beta\gamma}(r_g)$	$P^{\rm sm*}(r_g)$ and $P^{\rm sh*}(r_g)$		$P^{\rm sm*}(r_g)$ and $P^{\rm sh*}(r_g)$
Galaxy size r_g	FLUX_RADIUS	from HFINDPEAKS	FLUX_RADIUS	FLUX_RADIUS
Quadrupole estimate	Approx	Interpolation	Interpolation	Approx
$\theta_{\rm max}$ and $\Delta\theta$	Int[$4r_g$], 1 pix		$3r_g$, 0.25 pix	Int[$4r_g$], 1 pix
$P^{\rm sh}$ and $P^{\rm sm}$ estimate	Integer	Interpolation	Interpolation	Approx
P^γ correction	$\frac{1}{2}{\rm Tr}[P^\gamma]$	$P^{\rm sh} \rightarrow (1-\varepsilon^2/2)P^{\rm sh}$	$\frac{1}{2}{\rm Tr}[P^\gamma]$	Fit in$(r_g, {\rm mag})$
	(no fit)	Fit to	(no fit)	to $\frac{1}{2}{\rm Tr}[P^\gamma]$
		$\frac{1}{2}{\rm Tr}[P^\gamma](r_g)$		
Weights	none	Hoekstra et al.	$\langle\gamma^2\rangle^{-1}(r_g, {\rm mag})$	Hoekstra et al.
		eqn A8,9		eqn A8,9
γ correction				
Ellipticity cut	$\lvert\varepsilon_{obs}\rvert \leqslant 0.5$		$\lvert\varepsilon_{cor}\rvert \leqslant 0.8$	
Size cut	$1.2r_g^* < r_g < 7$ pix	r_h selection	$r_h > 1.2r_h^*$	
Significance cut	$\nu > 10$	$\nu > 5$		$\nu > 15$
P^γ cut			$\frac{1}{2}{\rm Tr}[P^\gamma] > 0$	
γ cut	$\lvert\gamma\rvert < 2$			
Other	Close pairs		$\lvert d\rvert < 3$ pix	
	<10 pix			
	removed			

pipeline multiplies $P^{\rm sh}$ by $(1-\varepsilon^2/2)$ at the P^γ correction stage. This modification is not used in any of the other analyses.

A5 Weights

Some authors employ a weighting scheme in their analysis. Hoekstra (HH) and Van Waerbeke (LV) use weights based on the error in the ellipticity measurement. These weights are derived in Appendix A1 of Hoekstra, Franx & Kuijken (2000). Clowe (C1&2), Hetterscheidt (MH) and Schrabback (TS) use a weighting scheme based on the inverse of $\langle\gamma^2\rangle$ for all galaxies within a given amount of r_g and magnitude (TS, MH) or significance ν (C1&2) of the galaxy using a minimum of 20/50 (TS, MH/C1&2) galaxies. Note that this type of weighting applied to galaxies that have experienced a constant shear will introduce a stronger bias that when the same weights are applied to data where the shear varies.

A6 Selection criteria and calibration correction

After applying the KSB+ method to the data each author has included a set of selection criteria, listed in Table A1. These criteria are based on object significance ν, 'optimal' size r_g, half light radius r_h, observed ellipticity $\varepsilon^{\rm obs}$, corrected ellipticity $\varepsilon^{\rm cor}$, measured shear γ, SEXTRACTOR stellar class (1 = star, 0 = galaxy), measured/modeled P^{γ} and so on. The IMCAT software GETSHAPES determines the offset of the flux averaged galaxy centroid (first moment) from the given input galaxy centroid, scaled by the galaxy flux. This measure, d, is used by Clowe (C1&2) to select 'good' galaxies. A similar selection criterion is included in the methods of Hetterscheidt (MH) and Schrabback (TS), where objects are only selected if their iterative refinement of the centroid position converges and fixes the position to better than 2×10^{-3} pixels independently in x and y. IMCAT also flags up saturated and bad pixels which add noise to the quadrupole moments. Clowe (C1&2) removes galaxies with any saturated or bad pixels within $3r_g$ of the centroid.

Brown (MB) includes a calibration correction $\gamma_{\rm cor} = \gamma/0.85$ as suggested from the analysis of image simulations in Bacon et al. (2001). Clowe (C1 & C2) includes a close-pair calibration correction $\gamma_{\rm cor} = \gamma/0.95$ to account for the diluting effect of blended objects. Normally Clowe visually inspects data to remove double objects classified as a single source and sources with tidal tails in addition to optical defects such as stellar spikes and satellite trails. This is feasible with the typical amounts of data analysed in cluster lensing analyses. For wide-field cosmic shear surveys however visual inspection becomes rather time consuming. For this analysis Clowe therefore visually inspected 10 images from the simulation resulting in the rejection $\sim$5 per cent of the objects. This process was found to increase the average shear measured in the visually inspected images by $\sim$5 per cent. Thus Clowe includes a close-pair correction factor in the STEP analysis to account for this effect in the whole simulation set.

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