Mon. Not. R. Astron. Soc. **376,** 13–38 (2007) doi:10.1111/j.1365-2966.2006.11315.x

The Shear Testing Programme 2: Factors affecting high-precision weak-lensing analyses

Richard Massey,[1⋆] Catherine Heymans,[2] Joel Bergé,[3] Gary Bernstein,[4] Sarah Bridle,[5] Douglas Clowe,[6] Håkon Dahle, [7] Richard Ellis,[1] Thomas Erben,[8] Marco Hetterscheidt,[8] F. William High,[1,9] Christopher Hirata,[10] Henk Hoekstra,[11] Patrick Hudelot,[12] Mike Jarvis,[4] David Johnston,[13] Konrad Kuijken,[14] Vera Margoniner,[15] Rachel Mandelbaum,[16] Yannick Mellier,[17,18] Reiko Nakajima,[4] Stephane Paulin-Henriksson,[19] Molly Peeples,[1,20] Chris Roat,[15] Alexandre Refregier,[3] Jason Rhodes,[1,13] Tim Schrabback,[8] Mischa Schirmer,[21] Uroš Seljak,[16] Elisabetta Semboloni[2,17] and Ludovic Van Waerbeke[2]

[1] *California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, USA*
[2] *University of British Columbia, 6224 Agricultural Road, Vancouver, BC V6T 1Z1, Canada*
[3] *Service d'Astrophysique, CEA Saclay, F-91191 Gif sur Yvette, France*
[4] *Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*
[5] *Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT*
[6] *Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tuscon, AZ 85721, USA*
[7] *Institute of Theoretical Astrophysics, University of Oslo, PO Box 1029, Blindern, N-0315 Oslo, Norway*
[8] *Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, 53121 Bonn, Germany*
[9] *Department of Physics, Harvard University, 17 Oxford Street, Cambridge, MA 01238, USA*
[10] *Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540, USA*
[11] *University of Victoria, Elliott Building, 3800 Finnerty Road, Victoria, BC V8P 5C2, Canada*
[12] *Observatoire Midi-Pyrénées, UMR5572, 14 Avenue Edouard Belin, 31000 Toulouse, France*
[13] *Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109, USA*
[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] *Department of Physics, Jadwin Hall, Princeton University, Princeton, NJ 08544, USA*
[17] *Institut d'Astrophysique de Paris, UMR7095 CNRS, Université Pierre & Marie Curie - Paris, 98 bis bd Arago, 75014 Paris, France*
[18] *Observatoire de Paris - LERMA, 61 avenue de l'Observatoire, 75014 Paris, France*
[19] *INAF/Catania Astrophysical Observatory, via S. Sofia 78, 95123 Catania, Italy*
[20] *Department of Astronomy, Ohio State University, 140 W. 18th Avenue, Columbus, OH 43210, USA*
[21] *Isaac Newton Group of Telescopes, Calle Alvarez Abreu 70, 38700 Santa Cruz de la Palma, Spain*

Accepted 2006 November 20. Received 2006 November 17; in original form 2006 August 29

ABSTRACT

The Shear Testing Programme (STEP) is a collaborative project to improve the accuracy and reliability of weak-lensing measurement, in preparation for the next generation of wide-field surveys. We review 16 current and emerging shear-measurement methods in a common language, and assess their performance by running them (blindly) on simulated images that contain a known shear signal. We determine the common features of algorithms that most successfully recover the input parameters. A desirable goal would be the combination of their best elements into one ultimate shear-measurement method. In this analysis, we achieve previously unattained discriminatory precision via a combination of more extensive simulations and pairs of galaxy images that have been rotated with respect to each other. That removes the otherwise overwhelming noise from their intrinsic ellipticities. Finally, the robustness of our simulation approach is confirmed by testing the relative calibration of methods on real data.

Weak-lensing measurements have improved since the first STEP paper. Several methods now consistently achieve better than 2 per cent precision, and are still being developed. However,

⋆E-mail: rjm@astro.caltech.edu

we can now distinguish all methods from *perfect* performance. Our main concern continues to be the potential for a multiplicative shear calibration bias: not least because this cannot be internally calibrated with real data. We determine which galaxy populations are responsible for bias and, by adjusting the simulated observing conditions, we also investigate the effects of instrumental and atmospheric parameters. The simulated point spread functions are not allowed to vary spatially, to avoid additional confusion from interpolation errors. We have isolated several previously unrecognized aspects of galaxy shape measurement, in which focused development could provide further progress towards the sub-per cent level of precision desired for future surveys. These areas include the suitable treatment of image pixellization and galaxy morphology evolution. Ignoring the former effect affects the measurement of shear in different directions, leading to an overall underestimation of shear and hence the amplitude of the matter power spectrum. Ignoring the second effect could affect the calibration of shear estimators as a function of galaxy redshift, and the evolution of the lensing signal, which will be vital to measure parameters including the dark energy equation of state.

Key words: gravitational lensing – methods: data analysis – cosmology: observations.

1 INTRODUCTION

The observed shapes of distant galaxies become slightly distorted by the (differential) gravitational deflection of a light bundle as it passes near foreground mass structures. Such 'cosmic shear' happens regardless of the nature and state of the foreground mass. It is therefore a uniquely powerful probe of the cosmic mass distribution, dominated by dark matter. Observations of gravitational lensing are directly and simply linked to theories of structure formation that are otherwise ill-equipped to predict the distribution of light (for reviews, see Bartelmann & Schneider 2001; Wittman 2002; Refregier 2003). Measurements are *not* limited by astrophysical bias (e.g. Dekel & Lahav 1999; Weinberg et al. 2004; Gray et al. 2002; Hoekstra et al. 2002b; Smith et al. 2003), which affects optical surveys, neither by unknown physics of distant supernovae (e.g. Hillebrandt & Niemeyer 2000; James et al. 2006; Sullivan et al. 2006; Travaglio, Hillebrandt & Reinecke 2006), nor by the uncertain relations between the mass of galaxy clusters and their observable X-ray luminosity or temperature (e.g. Pierpaoli, Scott & White 2001; Viana, Nichol & Liddle 2002; Huterer & White 2003). Gravitational lensing is a purely geometric effect, requiring knowledge of only deflection angles and distances. By directly observing the growth of the mass structures over cosmic time, and by investigating the large-scale geometry of the universe, it is also an effective probe of dark energy (Hoekstra et al. 2006; Jarvis et al. 2006; Schimd et al. 2006; Semboloni et al. 2006) and can test alternative theories of gravity that move beyond general relativity (White & Kochanek 2001).

The practical use of weak lensing in cosmology effectively began with the simultaneous detection of a coherent cosmic shear signal by four independent groups (Bacon, Refregier & Ellis 2000; Kaiser, Wilson & Luppino 2000; Van Waerbeke et al. 2000; Wittman et al. 2000). Since then, the field of weak lensing has advanced dramatically. Large, dedicated surveys with ground- and space-based telescopes have recently measured the projected 2D power spectrum of the large-scale mass distribution and drawn competitive 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; Bacon et al. 2003; Refregier, Rhodes & Groth 2002; Jarvis et al. 2003; Brown et al. 2003; Hamana et al. 2003; Massey et al. 2005; Rhodes et al. 2004; Heymans et al. 2005; Van Waerbeke, Mellier & Hoekstra 2005; Hoekstra et al. 2006; Semboloni et al. 2006; Jarvis, Bernstein & Dolney 2005; Hetterscheidt et al. 2006; Schrabback et al. 2006; Dahle 2006). The results from these efforts are found to be in broad agreement and are rapidly becoming more credible, with the most recent publications presenting several different diagnostic tests to determine the levels of systematic error. Ambitious plans are being laid for dedicated telescopes both on the ground (e.g. VST-KIDS, DES, VISTA darkCAM, Pan-STARRS, LSST) and in space (e.g. DUNE, SNAP, JDEM). Indeed, future weak-lensing surveys were recently identified as the most promising route to understanding the nature of dark energy by the joint NSF–NASA–DOE Astronomy and Astrophysics Advisory Committee (AAAC) and NSF–DOE High Energy Physics Advisory Panel (HEPAP) Dark Energy Task Force.[1] The importance of weak lensing in future cosmological and astrophysical contexts seems assured.

However, the detection and measurement of weak gravitational lensing presents a technical challenge. The $\sim$1 per cent distortion induced in the observed shapes of galaxies is an order of magnitude smaller than their typical intrinsic ellipticities, and a similar factor smaller than the spurious shape distortions created by convolution with the telescope's point spread function (PSF). Correction for these effects is crucial and complex. To test the reliability of weak-lensing measurements, it has therefore been necessary since the first detections to manufacture simulated images that closely resemble real data but contain a known shear signal. Bacon et al. (2001), Erben et al. (2001) and Hoekstra et al. (2002) ran their shear-measurement methods on such images. By comparing the input and mean measured shears, they determined the calibration error inherent to each technique, and in some cases discovered (and hence corrected) a multiplicative calibration bias. This is most important because it cannot be self-calibrated from a survey itself. Other systematics can be checked for in real data via correlation of the galaxies and the PSF, or via an $E-B$ decomposition (Crittenden et al. 2002; Schneider, Van Waerbeke & Mellier 2002; Schneider & Kilbinger 2007). These early tests determined that the first successful shear-measurement methods were accurate to $\leqslant$10 per cent of the signal.

[1] http://www.nsf.gov/mps/ast/detf.jsp.

To maximize progress in this technical field, and to foster the exchange of data and theoretical knowledge within the weak-lensing community, we launched the Shear Testing Programme (STEP). In the first STEP paper (Heymans et al. 2006, STEP1), we parametrized the performance of methods in terms of their multiplicative shear calibration bias m, an additive residual shear offset c and, in some cases, a nonlinear responsivity to shear q. That analysis confirmed that the main difficulty in weak lensing lies in the calibration of the shear signal, but encouragingly showed that all the methods used on existing weak-lensing surveys achieve better than $\sim$7 per cent accuracy. Shear-measurement error is therefore not currently a dominant source of error.

Unfortunately, this accuracy will not be sufficient to realize the potential of the ambitious and much larger future surveys. STEP1 found that the most-accurate shear-measurement methods were successfully calibrated to within a few per cent, but the limited size and precision of the first STEP simulations forbade any finer analysis than this. The morphologies of galaxies in the first simulated images were also overly simplistic, in a way that did not fully test the assumptions of some shear-measurement methods that galaxies lack substructure and complex shapes.

In this second STEP paper, we include complex galaxy morphologies and conduct a more precise test of current and developing shear-measurement algorithms to the $\leqslant$0.5 per cent level. We achieve this precision through the combination of a more extensive set of simulated images and an ingenious use of galaxy pairs rotated with respect to each other (Nakajima & Bernstein 2006). This removes the otherwise dominant noise from galaxies' intrinsic ellipticities. To focus on shear-measurement errors rather than PSF inerpolation, we hold the PSF fixed across each simulated image. However, we have designed the set of images to span a wide range of simulated observing conditions and to test several potentially challenging regimes for shear measurement. The data set is sufficiently large for it to be divided into different observing conditions and for independent tests to be carried out within each. We thereby test the effects of the following parameters on shear-measurement precision.

(i) Complex galaxy morphology;
(ii) galaxy size;
(iii) galaxy magnitude;
(iv) selection effects related to galaxy ellipticity;
(v) direction of the shear signal relative to the pixel grid;
(vi) PSF size; and
(vii) PSF ellipticity.

16 different shear-measurement codes have been run on the simulated images. These can be categorized into four distinct categories. We provide a brief description of each algorithm, and outline the relative successes of each method. The STEP programme has dramatically sped the development of new shear-measurement methods (e.g. Bernstein & Jarvis 2002; Refregier & Bacon 2003; Massey & Refregier 2005; Kuijken 2006; Nakajima & Bernstein 2006; Bridle et al., in preparation), and we particularly focus on these. However, these methods necessarily remain experimental, and development continues. The results from such methods should therefore be taken as an indication of progress rather than a judgement on their ultimate potential.

This paper is organized as follows. In Section 2, we describe the simulated images. In Section 3, we review the different shear-measurement methods used by each author, translating them into a common language for ease of comparison, and categorizing them into four distinct groups. In Section 4, we compare each author's measured shear with the input signal, and split the simulations in various ways to isolate areas of potential difficulty in shear measurement. Because of the number of different methods used, this is a rather daunting process. In Section 5, we provide some perspective on the results, assessing the relative performance of the different methods, and the categories of methods. In Section 6, we derive some general conclusions and outline suggestions for future development.

2 SIMULATED IMAGES

We have used the Massey et al. (2004a) simulation package to manufacture artificial images that closely resemble deep r-band data taken in good conditions with the Suprime-Cam camera on the *Subaru* telescope. We specifically mimic the weak-lensing survey data of Miyazaki et al. (2002b). The *Subaru* telescope was built with careful consideration of weak-lensing requirements, and has reliably obtained the highest-quality weak-lensing data to date (Miyazaki et al. 2002a; Wittman 2005; Kasliwal et al., in preparation). It therefore represents the current state of the art, and will most closely match future dedicated survey instruments. The simulated images are publicly available for download from the STEP website.[2]

To aid the interpretation of our results, the simulated images incorporate several 'unrealistic' simplifications: neither the noise level, the input shear signal nor the PSF vary as a function of position. This does not adversely affect the validity of the results, as any combination of PSF size, PSF ellipticity, and shear signal can usually be found in one of the images. However, it does let us simply average the measured shear for the large number of galaxies in each image, without explicitly keeping track of either the shear or the PSF applied to each object. We have attempted to decouple our investigation of shear measurment precision from the challenge of PSF interpolation. Solutions to this separate problem are being devised elsewhere (Hoekstra 2004; Jarvis & Jain 2004).

As in STEP1, the main figure of merit throughout our analysis will be the mean shear measured within each image, $\langle\tilde{\gamma}\rangle$, and deviations of that from the known input shear $\gamma^{\rm input}$. If the mean shear can be determined without bias for any input shear (and for any PSF), all the commonly used statistics typical in cosmic shear analysis should also be unbiased (but the distribution of the shear estimates will affect their noise level).

To address the specific topics outlined in the introduction, we manufactured six sets of simulated images. These span a range of realistic observing conditions, in a carefully orchestrated way that will isolate various effects. The differences between the images are described in Table 1. Each set contains 128 7 $\times$ 7-arcmin^2 images, with a pixel scale of 0.2 arcsec. In the first simulated image of each set, the galaxies are not sheared. For the next 63 images, which all feature the same patch of sky in order to maximize sensitivity to shear calibration, the galaxies are sheared by a random amount. This amount is chosen with a flat PDF within $|\gamma^{\rm input}| < 6$ per cent. To concentrate on cosmic shear measurement rather than cluster mass reconstruction, this limit is smaller than the maximum shears used in STEP1. However, the shears are now crucially chosen from a continuous distribution and are allowed to be in any direction relative to the pixel grid. Note that we are really attempting to measure 'reduced shear' (Seitz & Schneider 1997) throughout this analysis, although there is explicitly zero convergence in the simulations. The

[2] http://www.physics.ubc.ca/$\sim$heymans/step.html.

Table 1. The six different sets of images used in the STEP2 analysis are carefully chosen to isolate and test particular aspects of weak shear measurement. Either the PSF shape, or the form of galaxies' intrinsic morphologies varies in a prescribed way between sets.

Image set	PSF description	Galaxy type
A	Typical *Subaru* PSF (∼0.6 arcsec)	shapelets
B	Typical *Subaru* PSF (∼0.6 arcsec)	pure exponential
C	Enlarged *Subaru* PSF (∼0.8 arcsec)	shapelets
D	Elliptical PSF aligned along x-axis	shapelets
E	Elliptical PSF aligned at 45°	shapelets
F	Circularly symmetric *Subaru* PSF	shapelets

input signals were not disclosed to any of the groups analysing the data.

We can predict the signal-to-noise ratio (S/N) in the shear measurement from these images. We first define a complex ellipticity for each galaxy:

$$e = e_1 + \mathrm{i}e_2 \equiv \frac{a-b}{a+b}[\cos(2\theta) + \mathrm{i}\sin(2\theta)], \tag{1}$$

where a and b are the major and minor axes, and θ is the orientation of the major axis from the x-axis. This definition is widely used because it is more convenient than a two-component parametrization involving θ. Both the real and imaginary parts are well defined (zero) for a circular object or, on average, for an unsheared population of objects. In the absence of PSF smearing and shear-measurement errors, the observed galaxy ellipticity $e^{\rm obs}$ is related to its intrinsic ellipticity $e^{\rm int}$ by

$$e^{\rm obs} = \frac{e^{\rm int} + \gamma}{1 + \gamma^* e^{\rm int}} \tag{2}$$

(Seitz & Schneider 1997), where $\gamma \equiv \gamma_1 + i\gamma_2$ is the complex shear applied to each image. With only a finite number N of galaxies, all with non-zero intrinsic ellipticity, measurement of the mean shear $\langle\tilde{\gamma}\rangle = \langle e^{\rm obs}\rangle$ is limited by an intrinsic shot noise

$$\text{SN error} \approx \langle e^{\rm int}\rangle = 0 \pm \sqrt{\frac{\left\langle \left(e_i^{\rm int}\right)^2 \right\rangle}{N}}. \tag{3}$$

In the STEP2 simulations, $\sqrt{\langle e_i^2\rangle} \sim 0.1$, about an order of magnitude larger than the shear signal.

Since the morphologies of the simulated galaxies are uncorrelated, this noise can be slowly beaten down by increasing the size of the simulations. However, to dramatically improve the efficiency of the simulations, and circumvent the meagre $1/\sqrt{N}$ behaviour, we introduce an innovation in the remaining 64 images. Following a suggestion in Nakajima & Bernstein (2006), the entire sky, including the galaxies, was artificially rotated by 90° before being sheared by the same signals and being convolved with the same PSF as before. This rotation flips the sign of galaxies' intrinsic ellipticites. To measure biases in shear-measurement methods, we can then consider matched pairs of shear estimators from the unrotated and rotated versions of each galaxy. Averaging these estimators explicitly cancels the intrinsic shape noise, leaving only measurement noise and any imperfections in shear measurement. We thus form a shear estimator for each galaxy pair

$$\tilde{\gamma} = \frac{e^{\rm obs,unrot} + e^{\rm obs,rot}}{2}. \tag{4}$$

Since $e^{\rm int,unrot} = e^{\rm int} = -e^{\rm int,rot}$, we can use equation (2) to find

$$\begin{aligned}\tilde{\gamma} &= \left(\frac{e^{\rm int}+\gamma}{1+\gamma^* e^{\rm int}} + \frac{-e^{\rm int}+\gamma}{1-\gamma^* e^{\rm int}}\right) \Big/ 2 \\ &= \frac{\gamma - \gamma^* (e^{\rm int})^2}{1-(\gamma^* e^{\rm int})^2}.\end{aligned} \tag{5}$$

Averaging this shear estimator over $N/2$ galaxy pairs now gives a shot noise error in $\langle\tilde{\gamma}\rangle$ of

$$\text{SN error} \approx \gamma \left\langle \left(e_i^{\rm int}\right)^2 \right\rangle = 0 \pm \gamma \sqrt{\frac{\left\langle \left(e_i^{\rm int}\right)^4 \right\rangle}{2N}}, \tag{6}$$

which has been significantly reduced from equation (3). In the STEP2 simulations, $\sqrt{\langle (e_i^{\rm int})^4\rangle} \sim 0.05$ and $|\gamma| < 0.06$. Nothing is lost by this approach. All 128 images can still be analysed independently – and we do pursue this approach in order to measure the total shape measurement noise in an ordinary population of galaxies.

The Massey et al. (2004a) image simulation pipeline required extensive development from previously published versions to mimic ground-based data. We will therefore now describe its three main ingredients: stars (i.e. PSF), galaxies and noise.

2.1 Stars

The simulated images are observed after convolution with a various point-spread functions (PSFs). The PSF shapes are modelled on real stars observed in Suprime-Cam images, and are shown in Fig. 1. They are modelled using shapelets (Bernstein & Jarvis 2002; Refregier 2003; Refregier & Bacon 2003; Massey & Refregier 2005), a (complete) set of orthogonal basis functions that can be used to describe the shape of any isolated object. The decomposition of an image into shapelet space acts rather like a localized Fourier transform, with images $f(\boldsymbol{x})$ being expressed in shapelet space as a set of indexed coefficients $f_{n,m}$ that weight the corresponding basis function:

$$f(\boldsymbol{x}) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} f_{n,m} \chi_{n,m}(r, \theta; \beta), \tag{7}$$

with $m \leqslant n$, and where the Gauss–Laguerre basis functions are

$$\chi_{n,m}(r,\theta;\beta) = \frac{C_{n,m}}{\beta}\left(\frac{r}{\beta}\right)^{|m|} L^{|m|}_{(n-|m|)/2}\left(\frac{r^2}{\beta^2}\right) \mathrm{e}^{-r^2/2\beta^2} \mathrm{e}^{-\mathrm{i}m\theta}, \tag{8}$$

with a normalizing constant $C_{n,m}$ and scale size β.

The PSFs can therefore take a complex form. They contain substructure, skewness and chirality. In general, the ellipticity of their isophotes varies as a function of radius. For computational efficiency, the shapelet series is truncated at an order $n_{\rm max} = 12$. The limited wings and the rapid convergence of the PSFs to zero at large radii compared to those used in STEP1 are *not* a consequence of this truncation, but a confirmation of the excellent optical qualities of Suprime-Cam.

PSF A is modelled from a fairly typical star towards the centre of a 40-min-long Suprime-Cam exposure (which, in practice, is likely to be assembled from four 10-min exposures). It has a full width at half-maximum (FWHM) of 0.6 arcsec. PSF B is identical to PSF A. PSF C is the same star, but enlarged to model slightly worse seeing, and has an FWHM of 0.8 arcsec. This is the worst that might be expected in future weak-lensing surveys, with nights during poorer conditions typically used to obtain data in additional colours. PSF D is modelled on a star at the edge of the same Suprime-Cam exposure. The phases of all its $m = 2$ shapelet coefficients were adjusted to the same value so that at all radii (and therefore with any radial

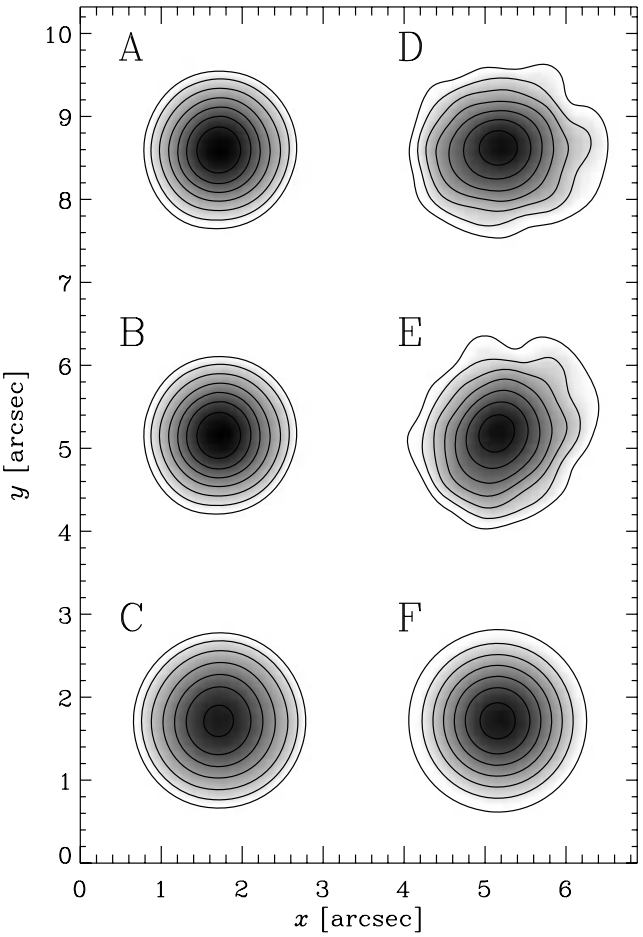

Figure 1. The PSFs used to generate the six different sets of simulated images. The colour scale is logarithmic, and the contours, which are overlaid at the the same absolute value on each PSF, are spaced logarithmically by factors of two. They are designed to target specific aspects of weak-lensing measurement that could potentially prove difficult to control. See Table 1 and the text for a description of each PSF.

weight function), its ellipticity derived from quadrupole moments points in exactly the same direction. Substructure and skewness apparent in the real *Subaru* PSF is otherwise untouched. As PSF D, the ellipticity is directed parallel to the x-axis of the pixel grid. The star is rotated by 45° to make PSF E. It is an example of extreme ellipticity, which highlights ellipticity-dependent effects. However, it might be possible to limit such ellipticity in weak-lensing surveys by improving the optical design of future telescopes or optimizing survey tiling and scheduling strategies. PSF F is a circularized version of that star, obtained by setting all its $m \neq 0$ shapelet coefficients to zero, which is equivalent to averaging the PSF over all possible orientations.

2.2 Shapelet galaxies

Most of the simulated images contain galaxy shapes also constructed from weighted combinations of the shapelet basis functions, using a version of the Massey et al. (2004a) image-simulation pipeline similarly modified to imitate ground-based data (Fig. 2). The complex and irregular galaxy morphologies that are possible using this method represent an important advance from the STEP1 analysis using the SKYMAKER image simulation package (Erben et al. 2001).

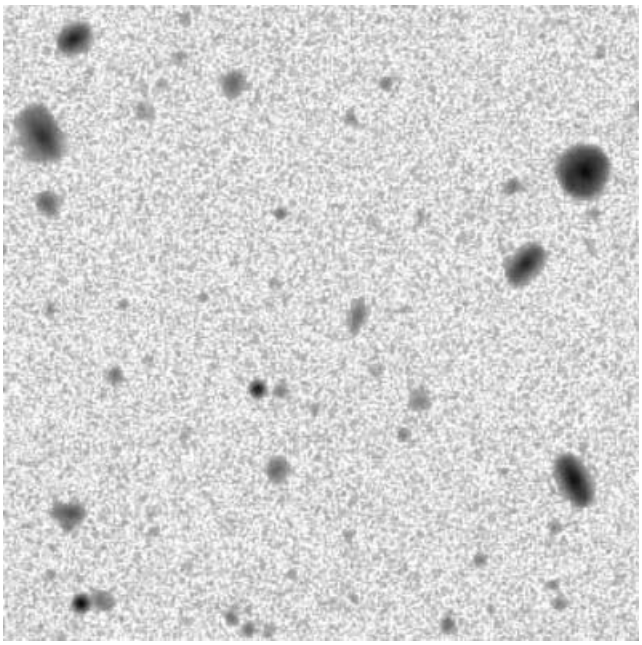

Figure 2. A 1 × 1-arcmin^2 section of a simulated image from set A, containing shapelet galaxies with complex morphologies. The colour scale is logarithmic, and the same as that in Fig. 3.

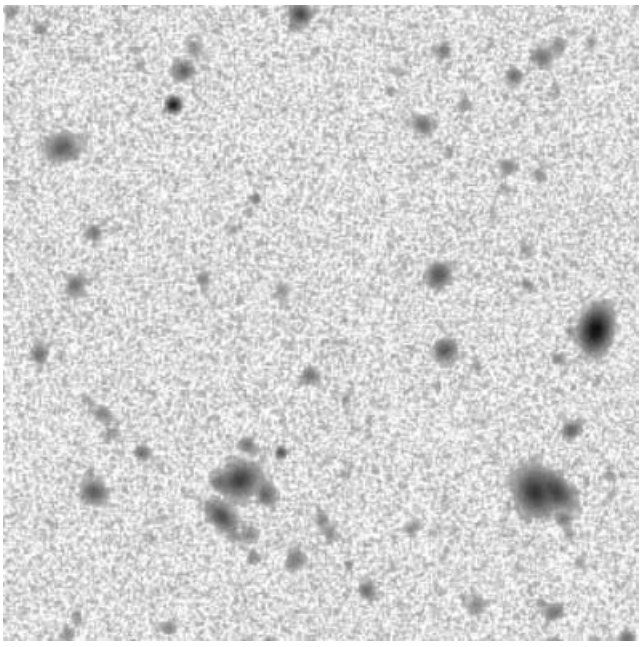

Figure 3. A 1 × 1-arcmin^2 section of a simulated image from set B, containing idealized galaxies with exponential radial profiles and simple morphologies. The colour scale is logarithmic, and the same as that in Fig. 2.

The measurement of weak lensing in STEP1 was considerably simplified by the galaxies' smooth and unperturbed isophotes. Several shear-measurement methods are based on the assumption that galaxy shapes and the PSF are concentric, elliptical, and in some cases Gaussian. In addition, the SKYMAKER galaxies have reflection symmetry about the centroid which could feasibly cause any symmetrical errors to vanish. By contrast, PSF correction and galaxy shape measurement are rendered more challenging in STEP2 by the

realistic morphologies that include spiral arms, dust lanes and small-scale substructure. Our analysis is thus designed to test the robustness of weak-lensing measurement methods.

The joint size–magnitude morphology distribution of galaxies was copied from the *Hubble Space Telescope* COSMOS survey (Scoville et al. 2007). This is a uniform, two square degree set of images taken with the *F814W* filter on the *Advanced Camera for Surveys* (ACS), to a depth of 28.7 for a point source at 5σ. It is deeper than our intended simulations, and with a much finer resolution, so provides an ideal source population. The extent of the COSMOS survey also provided sufficient real galaxies to avoid duplication in the simulations without needing to perturb shapelet coefficients, as in section 4 of Massey et al. (2004a). We simply used the shapelet models of COSMOS galaxies, randomly rotated, inverted and repositioned. The positions of galaxies in the simulations were chosen at random, without attempting to reproduce higher-order clustering.

Since the galaxy models are inevitably truncated at some level in shapelet space, and since we did not deconvolve the galaxies from the ACS PSF, the smallest simulated galaxies are intrinsically slightly rounder than those in real *Subaru* data. However, this convolution occurs before shearing and does not alter the necessary steps for shear measurement. As in real data, the simulated galaxy ellipticity and morphology distributions do vary with galaxy magnitude and size. We adopt an alternative definition of ellipticity:

$$(\varepsilon_1,\ \varepsilon_2) \equiv \frac{a^2-b^2}{a^2+b^2}[\cos(2\theta),\ \sin(2\theta)], \tag{9}$$

where a and b are the major and minor axes, and θ is the orientation of the major axis from the x-axis. Note the difference from equation (1); this version is closer to the notation used by most shear estimators. Before PSF convolution, the width of this ellipticity distribution

$$\sigma_\varepsilon^{\rm int} \equiv \left(\left(\sigma_{\varepsilon_1}^{\rm int}\right)^2 + \left(\sigma_{\varepsilon_2}^{\rm int}\right)^2\right)^{1/2} \tag{10}$$

as measured by SExtractor (Bertin & Arnouts 1996) is $\sigma_\varepsilon^{\rm int} = 0.35 \pm 0.03$ at $r = 22$ and $\sigma_\varepsilon^{\rm int} = 0.20 \pm 0.02$ at $r = 26$. Note that this ε is a different quantity than the e used in equation (3).

The galaxies were then sheared analytically in shapelet space, using equation (41) of Massey & Refregier (2005). This operation is to first order in γ. Terms of the order of γ^2 are ignored, but, for typical galaxy shapes, the coefficients by which these are multiplied are also smaller than those multiplying the first order terms. This therefore introduces only a very small error. The galaxies were then convolved with the PSF, also in shapelet space, using equation (52) of Refregier (2003). They were pixellated by analytically integrating the shapelet models within adjoining squares, using equation (34) of Massey & Refregier (2005).

2.3 Idealized galaxies

We have also manufactured one set (B) of simulated images with the same observing conditions but in which the galaxies have simple, exponential profiles and concentric, elliptical isophotes (Fig. 3). These idealized galaxies provide a contrast to the morphological sophistication of the shapelet galaxies, and an independent test of the shapelet-based shear-measurement methods. We intentionally chose a very simple form for the idealized galaxy shapes, with a sharp cusp and extended wings, to most effectively pronounce any difference to the results from galaxies with realistically complex morphologies. As before, the size–magnitude distribution of unsheared galaxies was modelled on that observed in the ACS COSMOS images. Galaxy ellipticities were assigned randomly from a Gaussian distribution. Like STEP1, we used a constant distribution of intrinsic ellipticity. This had width $\sigma_\varepsilon^{\rm int} = 0.3$ for galaxies at all magnitudes.

To add a shear signal, the random ellipticities are then perturbed at the catalogue level. Under a small shear γ_i, the ellipticity ε defined in equation (9) transforms as

$$\varepsilon_i^{\rm obs} = \varepsilon_i^{\rm int} + 2\left(\delta_{ij} - \varepsilon_i^{\rm int}\varepsilon_j^{\rm int}\right)\gamma_j + \mathcal{O}(\gamma^3) \tag{11}$$

(e.g. Rhodes, Refregier & Groth 2000), where δ_{ij} is the Kronekerdelta symbol, and the summation convention was assumed. Similarly, the mean square radius $d \equiv a^2 + b^2$ becomes

$$d'^2 = d^2\left(1 + 2\varepsilon_i^{\rm int}\gamma_i\right) + \mathcal{O}(\gamma^2). \tag{12}$$

These two expressions are valid up to first order in the shear. Note that, to this order, the flux F is unaffected by a pure shear. These results are valid for any galaxy with self-similar isophotes (as long as the moments converge).

To create a simulated galaxy image $f(\boldsymbol{x})$ with a desired ellipticity, we first specify the desired size r_0 and mean radial profile $p(r^2)$, where $r^2 = x_1^2 + x_2^2$ is the square radius and $\boldsymbol{x} = (x_1, x_2)$ are Cartesian coordinates on the sky, centred on the centroid of the galaxy. For convenience, we choose the normalization and angular scale of the generic profile such that

$$\int\int p(r^2)\,{\rm d}^2\boldsymbol{x} = \int\int r^2\,p(r^2)\,{\rm d}^2\boldsymbol{x} = 1\ . \tag{13}$$

The exponential profile used in these simulations is given by

$$p(r^2) = \frac{\sqrt{6}}{2\pi r_0}{\rm e}^{-\sqrt{6(r/r_0)^2}} \tag{14}$$

(cf. Rhodes et al. 2000 for the alternative case of a Gaussian profile). Using the conventions of equation (13) and a coordinate transformation

$$\mathbf{J} = \mathbf{R}(\theta)^{\rm T}\begin{pmatrix} a^2 & 0 \\ 0 & b^2 \end{pmatrix}\mathbf{R}(\theta) = d^2\begin{pmatrix} 1+\varepsilon_1 & \varepsilon_2 \\ \varepsilon_2 & 1-\varepsilon_1 \end{pmatrix}, \tag{15}$$

where $^{\rm T}$ denotes transpose and the rotation matrix

$$\mathbf{R}(\theta) \equiv \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \tag{16}$$

it is then easy to show that the elliptical galaxy image should have surface brightness

$$f(\boldsymbol{x}) = F\,\mathbf{J}^{-\frac{1}{2}}\,p(\boldsymbol{x}^{\rm T}\mathbf{J}^{-1}\boldsymbol{x}), \tag{17}$$

where the vertical bars denote the matrix determinant. The tails of their exponential profiles were artificially truncated at elliptical isophotes $5 \times r_0$ from the centre. To pixellate the galaxies, the value of the analytic function was computed at the centre of each pixel. The PSF was similarly pixellated, and convolution was then performed in a real space to produce the final image $I(\boldsymbol{x})$. Strictly, these operations should be reversed, and they do not commute. However, the pixels are small and the PSFs are Nyquist sampled, so the error introduced should be minimal.

2.4 Noise

A two-component noise model is then superimposed on to the images. Instrumental performance mimics that attained with a stack of four 10-min exposures with Suprime-Cam on the 8 m *Subaru* telescope (Miyazaki et al. 2002b). They are complete to $r = 25.5$, and the galaxies selected for lensing analysis are likely to have a median redshift $z_m \approx 0.9$. This is slightly deeper than most existing

weak-lensing surveys, and is towards the deep end of ground-based surveys planned for the future. The number density of useable galaxies found in these simulated images is therefore unlikely to be greatly surpassed.

The first component of 'photon counting' shot noise is first added to the true flux in every pixel. This is drawn from a Gaussian distribution with a width equal to the square root of the photon count. The images are then renormalized to units of counts per second. In the renormalized images, the rms of the Gaussian is 0.033 times the intensity in a pixel.

A second component of sky background is then added throughout each image, with an rms of 4.43 counts s^{-1}. The constant background level is assumed to be perfectly subtracted. The model *Subaru* images were combined using DRIZZLE, and the sky background noise is correlated in adjacent pixels. To mimic this effect, we smoothed the sky noise component (but not the flux in objects) by a Gaussian of FWHM 3.5 pixel. After this process, the rms of the sky noise is 1.65 counts s^{-1}. A simulated image of a completely blank patch of sky was also available to measure the covariance between pixels. The correlated noise particularly affects the detection of small, faint objects, and impedes the calculation of objects' weights from their detection S/N. It will be instructive in the future to consider which image resampling kernels and co-addition methods are optimal for shape measurement, or indeed whether we should stack the data at all. Jarvis et al. (2003) suggested measuring galaxy ellipticities on individual frames and combining these at the catalogue level. Note that faint simulated galaxies are created to the depth of the COSMOS survey, below the limiting magnitude of the simulated ground-based images, and these unresolved sources will also add slightly to the overall sky background.

3 SHEAR-MEASUREMENT METHODS

16 different shear-measurement codes have been run on the simulated images, by the authors listed in Table 2. Those that have been used elsewhere on real data, attempt to preserve as similar a pipeline as possible. Each method must first find and measure the shape of stars in each image. It must interpolate the PSF shape across the field, without assuming that it is constant. It must then find and measure the shapes of galaxies, correcting them appropriately for the effects of seeing. Note that we still consider object identification and classification to be part of a shear-measurement method, as shape biases can easily be introduced at this point (e.g. Bernstein & Jarvis 2002; Hirata & Seljak 2003); however, that task is likely to be separated in future STEP projects.

All the methods work by obtaining, for each galaxy, a two-component polarization ε_i that behaves like a generalized ellipticity. Precise definitions of polarization vary between methods, but it is important to note that easily measurable quantities do not usually change linearly with applied shear, so that $\langle\varepsilon\rangle \neq \gamma^{\rm input}$ for all values of $\gamma^{\rm input}$. To obtain an unbiased shear estimator, methods must determine how their polarizations change under an applied shear, and compute either a shear susceptibility tensor $P^{\gamma}_{ij} \equiv \delta\varepsilon_i/\delta\gamma_j$ or a shear responsivity factor $\mathcal{R}$. These are essentially interchangeable concepts, but with the word 'susceptibility' used to imply measurement from the higher-order shape moments of each galaxy (which are then often averaged or fitted across a galaxy population), and the word 'responsivity' to mean an average susceptibility for the population, measured from moments of the galaxy ellipticity distribution. In either case, this quantity can be inverted, and used to form a shear estimator

$$\tilde{\gamma} \equiv (P^{\gamma})^{-1}\,\varepsilon \tag{18}$$

or

$$\tilde{\gamma} \equiv \frac{\varepsilon}{\mathcal{R}}. \tag{19}$$

When computing the mean shear from a limited subset of galaxies, such as those in one size or magnitude bin, we will investigate two approaches to the calculation of $\mathcal{R}$. We try using the constant, global value, as has been done in published work, and we also try calculating $\mathcal{R}$ from the statistics of the smaller population. The latter is more noisy, but takes into account the evolution of galaxy morphology between samples (see Section 5.5).

In Table 3, the methods are broadly distinguished by their solutions to the two most important tasks in shear measurement. Some methods correct for the PSF at the catalogue level, by essentially subtracting the ellipticities of the PSF from that of each galaxy; others attempt to deconvolve each galaxy from the PSF, and measure the ellipticity of a reconstructed model. To obtain a polarization, some ('passive') methods measure combinations of galaxies' observed shape moments; other ('active') methods shear a model of an intrinsically circular source until it most closely resembles

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

Author	Key	Method
Bergé	JB	Shapelets (Massey & Refregier 2005)
Clowe	C1	KSB+ (same PSF model used for all galaxies)
Clowe	C2	KSB+ (PSF weight size matched to galaxies')
Hetterscheidt	MH	KSB+
Hoekstra	HH	KSB+
Jarvis	MJ	Bernstein & Jarvis (2002)
Jarvis	MJ2	Bernstein & Jarvis (2002) (new weighting scheme)
Kuijken	KK	Shapelets (Kuijken 2006)
Mandelbaum	RM	Reglens (Hirata & Seljak 2003)
Nakajima	RN	Bernstein & Jarvis (2002) (deconvolution fitting)
Paulin-Henriksson	SP	KSB+
Schirmer	MS1	KSB+ (scalar shear susceptibility)
Schirmer	MS2	KSB+ (tensor shear susceptibility)
Schrabback	TS	KSB+
Semboloni	ES1	KSB+ (shear susceptibility fitted from population)
Semboloni	ES2	KSB+ (shear susceptibility for individual galaxies)

Table 3. Broad classification scheme to distinguish different types of shear-measurement methods. The asterisks denote methods not tested in this paper. The top left-hand quadrant is red; the top right-hand quadrant is blue; the bottom left-hand quadrant is orange; and the bottom right-hand quadrant is green.

PSF correction scheme \ Shear measurement method	Passive	Active
Subtraction	KSB+ (various) Reglens (RM) RRG* K2K* Ellipto*	BJ02 (MJ, MJ2)
Deconvolution	Shapelets (JB)	Shapelets (KK) BJ02 (RN) im2shape*

the observed galaxy. We will now provide a brief description of each method, starting in the top-left quadrant of Table 3. Since the STEP programme has dramatically sped the development of new shear-measurement methods (Bernstein & Jarvis 2002; Refregier & Bacon 2003; Massey & Refregier 2005; Kuijken 2006; Nakajima & Bernstein 2006; Bridle et al., in preparation), we will particularly concentrate on the latest developments in those algorithms.

3.1 Red class methods

3.1.1 KSB+ (C1, C2, MH, HH, SP, MS1, MS2, TS, ES1 and ES2)

The shear-measurement method developed by Kaiser, Squires & Broadhurst (1995), Luppino & Kaiser (1997) and Hoekstra et al. (1998) is in widespread use by many current weak-lensing surveys. This has led to a high level of optimization of the basic method. The base IMCAT code is publicly available from the World Wide Web.[3] Many variations have been developed, and the ten implementations tested in this paper represent a cross-section of those that have been applied to real data. The details of each method are compared fully in the appendix of STEP1. The differences that STEP2 results reveal to be particularly significant are summarized again in Table 4.

The core of the method requires the measurement of the quadrupole moments of each observed galaxy image $I(\boldsymbol{x})$ weighted by a Gaussian of size $r_{\rm g}$. From these is formed a polarization

$$(\varepsilon_1,\ \varepsilon_2) \equiv \frac{\int\int I(\boldsymbol{x})\ W(\boldsymbol{x})\ r^2\big(\cos(2\theta),\ \sin(2\theta)\big)\ \mathrm{d}^2\boldsymbol{x}}{\int\int I(\boldsymbol{x})\ W(\boldsymbol{x})\ r^2\ \mathrm{d}^2\boldsymbol{x}}, \tag{20}$$

where

$$W(\boldsymbol{x}) = \mathrm{e}^{-r^2/2r_{\rm g}^2}. \tag{21}$$

The polarization is corrected for smoothing of the PSF via the smear susceptibility tensor $P^{\rm sm}$ and calibrated as shears via the shear polarizability tensor $P^{\rm sh}$: both of which involve higher-order shape moments. Using stars to denote measurements from stars (for which a smaller weight function is sometimes used) instead of galaxies, these form a shear estimator

$$\tilde{\gamma} = (P^{\gamma})^{-1}\,[\varepsilon - P^{\rm sm}(P^{\rm sm\star})^{-1}\,\varepsilon^{\star}], \tag{22}$$

[3] http://www.ifa.hawaii.edu/∼kaiser/imcat.

where

$$P^{\gamma} = P^{\rm sh} - P^{\rm sm}(P^{\rm sm\star})^{-1}\,P^{\rm sh\star}. \tag{23}$$

The tensor inversions can be performed in full, but these measurements of faint objects are particularly noisy. In practice, since the diagonal elements of P^{γ} are similar, and its off-diagonal elements are about an order of magnitude smaller, it can be approximated as a scalar quantity. Many implementations of KSB+ therefore simply divide by a shear susceptibility factor. The noise in P^{γ} is also sometimes reduced by fitting it from the entire population as a function of other observable quantities like galaxy size and magnitude. Reducing noise in any nonlinear aspect of shear measurement is vital, because the lensing signal is so much smaller than both the intrinsic ellipticity and photon shot noise, and must be obtained by linearly averaging away those sources of noise over a large population of galaxies.

Unfortunately, fundamental limitations in the mathematical formalism of KSB+ introduce further decisions that must also be resolved to approximate an ideal scenario in practical implementations. The KSB+ method makes no provision for the effects of pixellization; assumes that the PSF isophotes are concentric; and is mathematically ill-defined for non-Gaussian or non-concentric PSF and galaxy profiles. The various implementations developed by groups participating in the STEP2 analysis represent a cross-section of those choices.

Since STEP1, the TS method has incorporated a shear calibration factor of 0.91^{-1}, determined from the STEP1 results, but without knowledge of the STEP2 data. STEP2 therefore tests the robustness of this sort of calibration. As in STEP1, the C1 and C2 methods incorporate a calibration factor of 0.95^{-1} to eliminate the effect of close galaxy pairs. The C1 method uses a constant model of the PSF for all galaxies; the C2 method lets the size of the weight function $r_{\rm g}^{\star} = r_{\rm g}$ change to match each galaxy. The new SP method numerically integrates weight functions within pixels, uses the trace of P^{γ} from individual galaxies, and similar galaxy weights to the HH method. The ES1 method is based on the LV method from STEP1 but, rather than fitting the shear susceptibility from the galaxy population as a function of size and magnitude, it finds the 20 most similar galaxies in terms of those parameters, and uses their average value. This same procedure was used in the Semboloni et al. (2006) analysis of the CFHTLS deep survey. Subsequent tests on STEP1 images suggested that better results could be obtained by using individual measurements of P^{γ} from each galaxy, and ignoring the galaxy weights. These improvements have been incorporated into the new ES2 method.

One final finesse is required for methods that use weights w_i on each galaxy i that could vary between the rotated and unrotated images. For all N pairs of galaxies, we determine normalized weights

$$w_i' = \frac{N\ w_i}{\sum_{j=1}^{N} w_j} \tag{24}$$

and then calculate three estimates of the mean shear in each image

$$\langle\tilde{\gamma}^{\rm unrot}\rangle = \frac{1}{N}\sum(w^{\rm unrot\prime}e^{\rm obs,unrot}) \tag{25}$$

$$\langle\tilde{\gamma}^{\rm rot}\rangle = \frac{1}{N}\sum(w^{\rm rot\prime}e^{\rm obs,rot}) \tag{26}$$

$$\langle\tilde{\gamma}\rangle = \frac{1}{2N}\sum(w^{\rm unrot\prime}e^{\rm obs,unrot} + w^{\rm rot\prime}e^{\rm obs,rot}). \tag{27}$$

Errors on these are estimated using a bootstrap technique.

Table 4. Choices adopted by each of the shear-measurement methods that significantly affect their performance in this paper. See the appendix in STEP1 for more details about the differences between the various implementations of KSB+.

Author	Pixellization	Galaxy weighting scheme	Calculation factor	Shear susceptibility
JB	Analytic integration	None	—	Global mean shear responsivity $\mathcal{R} = 2 - \langle\varepsilon^2\rangle$
C1	Centre of pixel	min (ν, 40)	1/0.95	$\frac{1}{2}\mathrm{Tr}[P^\gamma]$, fitted as $f(r_\mathrm{g}, \varepsilon_i)$
C2	Centre of pixel	min (ν, 40)	1/0.95	$\frac{1}{2}\mathrm{Tr}[P^\gamma]$, fitted as $f(r_\mathrm{g}, \varepsilon_i)$
MH	Numerical integration	$1/(0.15 + \sigma_\varepsilon^2 + \sigma(\frac{1}{2}\mathrm{Tr}[P^\gamma(r_\mathrm{g})])^2)$	1/0.88	$\frac{1}{2}\mathrm{Tr}[P^\gamma]$, from individual galaxies
HH	Numerical integration	$1/(\sigma_\varepsilon^2 + s_\varepsilon^2/((1 - \frac{\varepsilon^2}{2})\frac{1}{2}\mathrm{Tr}[P^\gamma])^2)$	—	$(1 - \frac{\varepsilon^2}{2})\frac{1}{2}\mathrm{Tr}[P^\gamma]$, fitted as $f(r_\mathrm{g})$
MJ	Centre of pixel	$1/\sqrt{\varepsilon^2 + 2.25 s_\circ^2}$	—	Global mean shear responsivity $\mathcal{R}$
MJ2	Centre of pixel	$1/s_\circ^2$	—	Global mean shear responsivity $\mathcal{R}$
KK	Centre of pixel	$1/(0.1^2 + \sigma_{e_1}^2 + \sigma_{e_2}^2)$	—	Global mean shear responsivity $\mathcal{R} = 1 - \langle\varepsilon^2\rangle$
RM	Centre of pixel	$f(S/N)$	—	Global mean shear responsivity $\mathcal{R}$
RN	Centre of pixel	$1/\sqrt{\varepsilon^2 + 2.25 s_\circ^2}$	—	Global mean shear responsivity $\mathcal{R}$
SP	Numerical integration	$1/(0.15 + \sigma_\varepsilon^2 + \sigma(\frac{1}{2}\mathrm{Tr}[P^\gamma(r_\mathrm{g})])^2)$	—	$\frac{1}{2}\mathrm{Tr}[P^\gamma]$, Individual galaxies
MS1	Numerical integration	$1/\sigma_\varepsilon^2(r_\mathrm{g}, \mathrm{mag})$	—	$\frac{1}{2}\mathrm{Tr}[P^\gamma]$, fitted as $f(r_\mathrm{g}, \mathrm{mag})$
MS2	Numerical integration	$1/\sigma_\varepsilon^2(r_\mathrm{g}, \mathrm{mag})$	—	Full P^γ tensor, fitted as $f(r_\mathrm{g}, \mathrm{mag})$
TS	Numerical integration	None	1/0.91	$\frac{1}{2}\mathrm{Tr}[P^\gamma]$, from individual galaxies
ES1	Numerical integration	$1/(\sigma_\varepsilon^2(r_\mathrm{g}, \mathrm{mag}) + 0.44^2)$	—	$\frac{1}{2}\mathrm{Tr}[P^\gamma]$, smoothed from galaxy population $f(r_\mathrm{g}, \mathrm{mag})$
ES2	Numerical integration	None	—	$\frac{1}{2}\mathrm{Tr}[P^\gamma]$, from individual galaxies

3.1.2 Reglens (RM)

The Reglens (RM) method consists of two parts: the SDSS data-processing pipeline PHOTO (Ivezić et al. 2004), followed by the re-Gaussianization pipeline (Hirata & Seljak 2003; Mandelbaum et al. 2005). The magnitude cut was adjusted, and one additional subroutine was required for the STEP2 analysis, to properly determine the noise variance in the presence of correlated background noise. The STEP2 images are more crowded than SDSS images, leading to occasional deblending problems. Objects with failed deblending were automatically eliminated, after visual inspection indicated that nearly all of them were really several galaxies very close to each other.

PSF correction is performed via a two-step procedure that addresses KSB+'s limitation of being exact only in the limit of Gaussian PSF and galaxy profile. The PSF is first split into a Gaussian component $G(\boldsymbol{x})$ plus a small residual $\epsilon(\boldsymbol{x})$, so that the observed image

$$I = (G + \epsilon) \otimes f = G \otimes f + \epsilon \otimes f, \tag{28}$$

where $f(\boldsymbol{x})$ is the galaxy image before convolution of the PSF, and $\otimes$ signifies convolution. Assuming knowledge of f, it would be possible to find

$$I' \equiv G \otimes f = I - \epsilon \otimes f, \tag{29}$$

the galaxy image as it would appear when convolved with a perfectly Gaussian PSF. Although f is not known in practice, it is convolved with a small correction ϵ in the final equality, so equation (29) is fairly accurate even with an approximation f_0. The SDSS and STEP2 analyses used an elliptical Gaussian as f_0, with its size and ellipticity determined from the difference between the best-fitting Gaussians to the observed image and the full PSF. Possible alternatives to this approximation are discussed in Hirata & Seljak (2003).

Correction for the isotropic part of the now Gaussian PSF reqires a subtraction similar to that in KSB+ equation (22), except that Reglens directly subtracts moments of the PSF from those of the galaxy (i.e. the numerator and denominator of equation 20) before they are divided (i.e. the ratio in equation 20). Furthermore, the moments are calculated using weight functions $W_{I'}(\boldsymbol{x})$ and $W_G(\boldsymbol{x})$ that are the best-fitting elliptical Gaussians to the image and to the PSF respectively. The advantage of these adaptive weight functions is that they do not bias the shape measurement or require later correction. Correction for the anisotropic part of the Gaussian PSF is finally performed by shearing the coordinate system, including I', until G is circular.

In the absence of galaxy weights, a shear estimate for each galaxy would be computed via equation (19). The shear responsivity

$$\mathcal{R} = 2 - \sigma_\varepsilon^2 \equiv 2 - \left\langle \varepsilon_1^2 + \varepsilon_2^2 - s_{\varepsilon_1}^2 - s_{\varepsilon_2}^2 \right\rangle, \tag{30}$$

is calculated from shape distribution statistics of the entire galaxy population and the error on each polarization, s_{ε_i}, is calculated by propagating measured photon shot noise in the image. During our analysis, it became apparent that, for the RM, MJ, MJ2 and RN methods, it is necessary to recalculate $\mathcal{R}$ in each bin of galaxy size or magnitude when the catalogue is so split.

To improve the S/N, galaxies are each weighted by a factor

$$w = \frac{1}{\sigma_\varepsilon^2 + s_{\varepsilon_1}^2}. \tag{31}$$

An estimate of the mean shear in each image is then simply

$$\langle\tilde{\gamma}\rangle = \sum w \frac{\varepsilon}{\mathcal{R}} \Big/ \sum w, \tag{32}$$

with a shear responsivity (Bernstein & Jarvis 2002):

$$\mathcal{R} = \sum w \left(2 - 2k_0 - k_1\varepsilon^2\right) \Big/ \sum w, \tag{33}$$

where $k_0 = \sigma_\varepsilon^2 - w\sigma_\varepsilon^4$ and $k_1 = w^2\sigma_\varepsilon^4$.

Note that this calculation of $\mathcal{R}$ in the STEP2 images is much more uncertain than in SDSS data, because the correlated background noise in the STEP2 images is not as well understood. Consequently, this may introduce some bias into the STEP2 results that does not exist with the real data.

3.1.3 Other methods not tested in this paper

Rhodes et al. (2000, RRG) is a modification of the KSB+ method for space-based data in which the PSF is small. In this limit, $\varepsilon^\star$ becomes noisy. Like Reglens, RRG therefore deals directly with moments rather than polarizations for as long as possible, and performs the

subtraction before the division. The moments use a circular weight function, and therefore require correction for this truncation as well as the PSF. RRG uses a global shear responsivity $\mathcal{R} \approx 2 - \langle\varepsilon^2\rangle$.

Kaiser (2000, K2K) also seeks a resolution of the Gaussian PSF limitation in KSB+. The galaxy image is first convolved by an additional 're-circularizing kernel', which is a modelled version of the observed PSF that has been rotated by 90°. PSF correction and shear measurement are thereafter fairly similar to KSB. However, particular efforts are made to correct biases that arise from the use of P^γ measured after shear rather than before shear.

Ellipto (Smith et al. 2001) also uses a re-circularizing kernel to eliminate the anisotropic component of the PSF, following (Fischer & Tyson 1997). It then repeats object detection to remove PSF-dependent selection biases. Galaxy polarizations are derived from moments weighted by the best-fit elliptical Gaussian. It is a partial implementation of BJ02, discussed in the next section, and primarily differs from BJ02 by using a simpler re-circularizing kernel.

3.2 Blue class methods

3.2.1 BJ02 (MJ and MJ2)

The remaining methods are based on expansions of the galaxy and PSF shapes into Gauss–Laguerre ('shapelet') basis functions. The JB and KK methods use them with a circular basis function, as defined in equations (7) and (8), while the MJ, MJ2 and RN methods use more general elliptical versions. Shapelets are a natural extension of KSB+ to higher order. The first few shapelet basis functions are precisely the weight functions used in KSB+, with r_g reinterpreted as the shapelet scale size β. Generalized versions of the $P^{\rm sh}$ and $P^{\rm sm}$ matrices are derived in Refregier & Bacon (2003). Extending the basis set to higher order than KSB+ allows complex shapes of galaxies and PSFs to be well described, even when the ellipticity varies as a function of object radius. The shapelet basis set is mathematically well suited to shear measurement because of the simple transformation of shapelet coefficients during typical image manipulation.

The two Jarvis (MJ, MJ2) methods correct for the anisotropic component of the PSF by first convolving the image with an additional, spatially varying kernel that is effectively 5 × 5 pixels. This convolution is designed to null both the Gaussian-weighted quadrupole of the PSF and its next higher $m = 2$ shapelet coefficient (since it is the $m = 2$ components of the PSF that mostly affects the observed shapes of galaxies). For PSF ellipticities of order ∼0.1 or less, a 5 × 5 pixel kernel is sufficient to round a typical PSF up to approximately 50 pixels in diameter: much larger than the PSFs used in this study.

The shapelet basis functions are sheared, to make them elliptical, then pixellated by being evaluated at the centre of each pixel. Shapelet coefficients $f_{n,m} = 0$ are determined for each galaxy in distorted coordinate systems, and the polarizability ε is defined as -1 times the amount of distortion that makes each object appear round (i.e. $f_{2,2} = 0$). Some iteration is required to get this measurement to converge. In the distorted coordinate frame where the galaxy is round, the weight function for this coefficient is a circular Gaussian of the same size as the galaxy. Matching the shape of the weight function to that of the galaxy has the advantage that the polarizability no longer requires correction for truncation biases introduced by the weight function.

Finally, a correction for the PSF dilution (the circularizing effect of the PSF) is applied by also transforming the PSF into this coordinate system, and then using formulae proposed by Hirata & Seljak (2003):

The two methods (MJ, MJ2) differ only in the weights applied to each galaxy. The MJ method is identical to the MJ method used for the STEP1 study. It uses weights

$$w_{\rm MJ} = \frac{1}{\sqrt{e^2 + 2.25 s_\circ^2}}, \tag{34}$$

where $s_\circ$ is the uncertainty in the polarizability due to image shot noise, as measured in the coordinate system where the galaxy is round. STEP1 revealed that this optimized weight gave incorrect responsivities as the input shear became large (≈0.1). For this study, method MJ2 was therefore added, which is identical except that it uses weights that are not a function of the galaxies' polarizations

$$w_{\rm MJ2} = \frac{1}{s_\circ^2}. \tag{35}$$

These weights should be less biased for larger input shears. The MJ weight might be more appropriate for cosmic shear measurements, and the MJ2 weight for cluster lensing.

The shear responsivity $\mathcal{R}$ for the MJ2 method is the same as that in equation (33). For the ellipticity-dependent weight used by the MJ method, this is generalized to

$$\mathcal{R} \equiv \frac{\sum \left[w \left(2 - 2k_0 - k_1\varepsilon^2\right) + \varepsilon(\partial w/\partial\varepsilon)\left(1 - k_0 - k_1\varepsilon^2\right)\right]}{\sum w}, \tag{36}$$

where the summations are over the entire galaxy population, or for each size or magnitude bin. For either method, an estimate of the mean shear in each image is then

$$\langle\tilde{\gamma}\rangle = \sum w \frac{\varepsilon}{\mathcal{R}} \Big/ \sum w. \tag{37}$$

Note that, in the absence of shape noise, equation (36) reproduces the extra $(1 - \varepsilon^2/2)$ term multiplying P^γ in the HH implementation of KSB+ (see Table 4).

3.3 Orange class methods

3.3.1 Shapelets (JB)

The Bergé (JB) shear-measurement method uses a parametric shapelet model to attempt a full deconvolution of each galaxy from the PSF. Deconvolution is an ill-defined operation in general, since information is irrevocably lost during convolution. In shapelet space, however, it is easy to restrict the galaxy model to include only that range of physical scales in which information is expected to survive. Massey & Refregier (2005) described an iterative algorithm designed to optimize the scale size of the shapelets and to thus capture the maximum range of available scales for each individual galaxy. A complete software package to perform this analysis and shapelet manipulation is publicly available from the shapelets website.[4]

To model a deconvolved galaxy shape, the basis functions are first convolved with the PSF in shapelet space, then integrated analytically within pixels: thus undergoing the same processes as real photons incident on a CCD detector. The convolved basis functions are then fit to the data, with the shapelet coefficients as free parameters. Reassembling the model using *un*convolved basis functions produces a deconvolved reconstruction of each galaxy. This

[4] http://www.astro.caltech.edu/∼rjm/shapelets.

performs better than a Wiener-filtered deconvolution in a Fourier space, because shapelets have a preferred centre. The available basis functions act as a prior on the reconstruction, localizing it in a real space (and also allowing a slightly higher resolution at the central cusp than at large radii). The deconvolved model can also be rendered free of noise by ensuring that a sufficient range of scales are modelled to lower the residual $\chi^2_{\rm reduced}$ to exactly unity. Unfortunately, achieving exactly this target is hindered by the presence of correlated background noise in the STEP2 simulations. Incorporating the noise covariance matrix is theoretically trivial but infeasibly slow in practice, because of the required matrix inversion: a workable implementation has not yet been developed. Proceeding regardless, the shape of this analytic model can be directly measured (see Massey, Refregier & Bacon 2004b; Massey et al. 2007), including its *unweighted* moments. These cannot be measured directly from real data because observational noise prevents the relevant integrals from converging.

Once a deconvolved model is obtained, extraction of a shear estimator is easy. It could mimic the KSB method. However, removing the weight function (like the Gaussian in equation 20) makes the polarization itself into an unbiased shear estimator

$$\tilde{\gamma} = \frac{\int\int f(\boldsymbol{x})\, r^2\, [\cos(2\theta),\ \sin(2\theta)]\, \mathrm{d}^2\boldsymbol{x}}{\int\int f(\boldsymbol{x})\, r^2\, \mathrm{d}^2\boldsymbol{x}}. \tag{38}$$

The numerator of this expression has a shear susceptibilty equal to the denominator. However, that denominator is a scalar quantity, with explicitly zero off-diagonal elements in the susceptibility tensor, which can therefore be easily inverted. It is also a simple product of a galaxy's flux and size, both low-order quantities that can be robustly measured. The method is intended to be completely linear for as long as possible, and to introduce minimal bias for even faint objects in this final division. Since the denominator also changes during a shear, a population of galaxies acquires an overall shear responsivity factor

$$\mathcal{R} = 2 - \langle \varepsilon^2 \rangle. \tag{39}$$

The method is still under development. The shear responsivity factor has currently been calculated only from the entire galaxy population. No weighting scheme has yet been applied to the shear catalogue when calculating mean shears. Once galaxies have passed crude cuts in size, flux and flags (which indicate successful convergence of the shapelet series and of the iteration), they are all counted equally. These aspects will be improved in the future.

3.4 Green class methods

3.4.1 Shapelets (KK)

The Kuijken (KK) shear-measurement method (Kuijken 1999) assumes that each galaxy was intrinsically circular, then shears it, and smears it by the PSF, until it most closely matches the observed image. The shear required is the stored as the polariz ation ε. As described in Kuijken (2006), this approach is desirable, because it is understood precisely how a circular object changes under a shear.

This process could operate in a real space; however, the convenient properties of shapelets make the required image manipulations easier and faster in shapelet space. The pixellated image need be accessed only once, when each galaxy is initially decomposed into shapelets (without deconvolution). Models of circular sources can have arbitrary radial profiles, parametrized by shapelet coefficients with $m = 0$ and $n \leqslant 12$. This is sheared in shapelet space to first order in γ, although, in principle, this could also be increased to accommodate more highly elliptical objects. Also in shapelet space, it is smeared by a model of the PSF. Since there is only one shapelet decomposition overall, and one forward convolution for each object, the code is much faster than the Bergé (JB) method. Furthermore, the decomposition uses completely orthogonal shapelet basis functions, so the errors on shapelet coefficients are also uncorrelated at that stage. To avoid iterating the decomposition, the optimum scale size β for each object is approximated from SExtractor parameters, and the range of scales is fixed in advance. In the current implementation, the basis functions are evaluated at the centre of each pixel. Since both the PSF and the galaxy are pixellated, its effects ought to drop out. In terms of the orthogonality of the shapelet basis functions, this approach is satisfactory as long so the range of scales is small, and oscillations in the basis functions remain larger than the pixel scale (cf. Berry, Hobson & Withington 2004).

To determine the shear required to make a circular source match each real galaxy, a fit is performed using a numerical recipe Newton-Raphson algorithm, which is quadratic in shapelet coefficients, the centroid and the shear. Since the galaxies are not really all circular, in practice the global population does have a non-trivial shear susceptibility or 'responsivity' $\mathcal{R}$. For an ensemble population of galaxies, this is a scalar quantity. As can be deduced from equation (11), it involves the variance of the intrinsic polarization distribution

$$\mathcal{R} \equiv 1 - \langle e^2 \rangle. \tag{40}$$

Unlike other methods that use a shear responsivity correction, this quantity was calculated only once for the KK method, from the entire galaxy population. However, the calculation of $\langle e^2 \rangle$ properly takes into account the galaxy weights

$$\langle e^2 \rangle = \frac{\sum \left[w \left(e_1^2 + e_2^2 - s_{e_1}^2 - s_{e_2}^2 \right) \right]}{\sum w} - \left(\frac{\sum w (e_1 + e_2)}{\sum w} \right)^2, \tag{41}$$

where s_{e_i} is the noise on each polarization calculated by propagating photon shot noise, and the weight for each galaxy is

$$w = \frac{1}{\left(\sigma_e^{\rm int}\right)^2 + s_{e_1}^2 + s_{e_2}^2}. \tag{42}$$

Note that the estimates of errors on the polarizations did not take into account the fact that the background noise was correlated between adjacent pixels, and are therefore likely to be underestimated.

Shear estimates for individual galaxies are then computed similarly to equation (37), but where $\tilde{\gamma} \equiv e/\mathcal{R}$ here.

3.4.2 BJ02 (RN)

The 'deconvolution fitting method' by Nakajima (RN) implements nearly the full formalism proposed by BJ02, which is further elaborated in Nakajima & Bernstein (2006). Like MJ and MJ2, it shears the shapelet basis functions until they match the ellipticity of the galaxy. The amount of distortion that makes an object appear round (i.e. $f_{2,2} = 0$) defines the negative of its polarizability ε.

Since no PSF interpolation scheme has yet been developed, the pipeline deviates from the STEP rules by using prior knowledge that the PSF is constant across each image (but not between images). Deconvolution from the PSF is performed in a similar fashion to the JB method. The Gauss–Laguerre basis functions are convolved with the PSF to obtain a new basis set. These are evaluated at the centre of each pixel. The new basis functions are fitted directly to the observed pixel values, and should fully capture the effect of highly asymmetric PSFs or galaxies, as well as the effects of finite sampling. The fit iterates until a set of sheared Gauss–Laguerre basis

functions are obtained, in which the coefficients $f_{2,0} = f_{2,2} = 0$ and hence the deconvolved galaxy appears round. All PSF coefficients were obtained to $n \leqslant 12$, and galaxy coefficients to $n \leqslant 8$.

The weights applied to each galaxy are optimized for small shears, using the same prescription as the MJ2 method in equation (35). The shear responsivity $\mathcal{R}$ is similarly calculated using equation (36), averaged over the entire galaxy population or within size and magnitude bins as necessary.

The evolution of the RN method during the STEP2 analysis highlights the utility of even one set of STEP simulations. In the first submission, it was noted that a few outlying shear estimates in each field were destabilizing the result. These were identified as close galaxy pairs, so an algorithm was introduced to remove these, and the size and magnitude cuts were also gradually adjusted over several iterations to improve stability.

3.4.3 Other methods not tested in this paper

IM2SHAPE (Bridle et al. 2001) performs a similar PSF deconvolution, but parametrizes each galaxy and each PSF as a sum of elliptical Gaussians. The best-fitting parameters are obtained via a Markov Chain Monte Carlo sampling technique. Concentric Gaussians are usually used for the galaxies, in which case the ellipticity is then a direct measure of the shear via equations (1) and (2). For alternative galaxy models using non-concentric Gaussians, shear estimators like that of the JB method could also be adopted. The 'active' or 'passive' classification of this method is somewhat open to interpretation.

4 RESULTS

Individual authors downloaded the simulated images and ran their own shear-measurement algorithms, mimicking as closely as possible the procedure they would have followed with real data. None of the authors knew the input shears at this stage. Their galaxy catalogues were then compiled by Catherine Heymans and Richard Massey. Independently of the other authors, the mean shears in each image were compared to the input values. Galaxies in the measured catalogues were also matched to their rotated counterparts and to objects in the input catalogues, with a 1-arcsec tolerance. Except for determining false detections or stellar contamination in the measured catalogues (which were removed in the matched catalogues), no results using the input shapes are presented in this paper.

In this section, we present low-level data from the analyses, in terms of direct observables. For further discussion and interpretation of the results in terms of variables concerning global survey and instrumental performance, see Section 5. To conserve space, only a representative sample of the many results are displayed here. The rest is described in the text, in relation to the illustrative examples, and is also available from the STEP website.[2] First, we will describe the measurement of stars, then the number density of galaxies and then shears in each set of images. Finally, we will split the galaxy catalogues by objects' observed sizes and magnitudes.

4.1 PSF modelling

The first task for all shear-measurement methods is to identify stars and measure the shape of the PSF. Table 5 lists parameters of the PSF model generated by the TS implementation of KSB+. These quantities are more familiar than those derived analytically from the shapelet models, and also demonstrate the differences between measured PSF ellipticities and inputs described in Table 1. The few per cent polarizations measured for components of PSFs D and E that should be zero are typical of several other methods. These may explain the peculiar residual shear offsets described in Section 5.3.

Table 5. PSF models for the six sets of images used in the STEP2 analysis by the TS implementation of KSB+, averaged over stars in the simulated images. These quantities may be more familiar to some readers. FLUX_RADIUS is directly from SExtractor, and the ellipticities are all measured using a Gaussian weight function of rms size $r_g = 0.6$ arcsec=3 pixels.

Image set	PSF model from TS implementation of KSB+		
	FLUX_RADIUS (arcsec)	ε_1 (per cent)	ε_2 (per cent)
A	0.334	$-(0.68 \pm 0.10)$	(1.21 ± 0.07)
B	0.334	$-(0.66 \pm 0.07)$	(1.28 ± 0.05)
C	0.406	$-(0.47 \pm 0.07)$	(0.97 ± 0.06)
D	0.390	(11.49 ± 0.11)	(2.20 ± 0.14)
E	0.390	$-(2.21 \pm 0.14)$	(11.29 ± 0.16)
F	0.392	$-(0.01 \pm 0.12)$	(0.01 ± 0.01)

4.2 Galaxy number counts and the false detection rate

The methods used a variety of object-detection algorithms and catalogue selection criteria. For each method and each PSF, Table 6 lists the density of objects per square arcminute, $n_{\rm gals}$, their mean magnitude, and the percentage of false detections. Clearly, methods that are able to successfully measure the shapes of more (fainter) galaxies, while avoiding false detections, will obtain a stronger measurement of weak lensing, especially because the lensing signal grows cumulatively with galaxy redshift. The false detection and stellar contamination rate is generally low, and the effective survey depth is lowered by less than 0.1 mag for all methods after matching rotated and unrotated catalogues. Nor does matching has a significant effect on the overall mean polarization of galaxies, which is always consistent with zero both before and after matching – as might not have been the case in the presence of selection effects (Bernstein & Jarvis 2002; Hirata & Seljak 2003).

Table 6 also shows the measured dispersion of shear estimators σ_γ for each population. This statistic represents a combination of the intrinsic ellipticity of galaxies and the shape measurement/PSF correction noise introduced by each method. Lower values will produce stronger measurements of weak lensing. Since shear measurement is more difficult for smaller or fainter galaxies, and the intrinsic morphology distribution of galaxies varies as a function of magnitude in images other than set B, $n_{\rm gals}$ and σ_γ are likely to be correlated in a complicated fashion. Galaxy selection effects and weighting schemes are discussed in Sections 5.6 and 5.7.

4.3 Shear calibration bias and residual shear offset

As with STEP1, we assess the success of each method by comparing the mean shear measured in each image with the known input shears $\gamma_i^{\rm input}$. We quantify deviations from perfect shear recovery via a linear fit that incorporates a multiplicative 'calibration bias' m and an additive 'residual shear offset' c. With a perfect shear-measurement method, both of these quantities would be zero. Since the input shear is now applied in random directions, we measure two components each of m and c, which correspond to the two components of shear:

$$\begin{aligned} \langle\tilde{\gamma}_1\rangle - \gamma_1^{\rm input} &= m_1\gamma_1^{\rm input} + c_1 \\ \langle\tilde{\gamma}_2\rangle - \gamma_2^{\rm input} &= m_2\gamma_2^{\rm input} + c_2. \end{aligned} \tag{43}$$

Table 6. Number density of galaxies used by each method, and the shear-measurement noise from those galaxies. The number of galaxies per square arcminute are listed for the unmatched unrotated/rotated catalogues and after matching. The number in brackets is the percentage of stars or false detections.

Author	Image set	$n_{\rm gals}$		Mean magnitude	Per cent mag	$\sigma\gamma$	
		Original/	Matched	(original)	Decrease	Original/	Matched
JB	A	37 (0)	25	24.04	1.2	0.012	0.007
	C	28 (1)	21	23.50	1.0	0.014	0.008
C1	A	51 (2)	45	23.70	0.3	0.008	0.003
	C	46 (2)	40	23.64	0.4	0.009	0.003
C2	A	50 (2)	45	23.70	0.3	0.008	0.003
	C	45 (2)	40	23.64	0.4	0.009	0.003
MH	A	38 (0)	35	23.68	0.4	0.008	0.003
	C	33 (0)	29	23.56	0.5	0.009	0.004
HH	A	28 (0)	26	23.05	0.2	0.010	0.002
	C	24 (0)	21	22.97	0.3	0.012	0.002
MJ	A	27 (1)	24	23.30	0.3	0.009	0.003
	C	25 (0)	22	23.26	0.4	0.009	0.003
MJ2	A	27 (1)	24	22.58	0.1	0.014	0.002
	C	25 (0)	22	22.48	0.2	0.016	0.002
KK	A	32 (0)	26	23.46	0.5	0.009	0.003
	C	27 (0)	21	23.35	0.5	0.010	0.003
RM	A	36 (0)	32	23.41	0.3	0.009	0.002
	C	27 (0)	23	23.21	0.4	0.010	0.003
RN	A	22 (1)	19	23.10	0.3	0.009	0.003
	C	16 (1)	13	23.03	0.5	0.011	0.004
SP	A	27 (11)	15	23.13	0.4	0.014	0.003
	C	25 (10)	13	23.10	0.4	0.016	0.004
MS1	A	43 (1)	39	23.68	0.3	0.007	0.003
	C	37 (1)	33	23.55	0.3	0.008	0.003
MS2	A	41 (1)	36	23.46	0.1	0.010	0.004
	C	35 (1)	30	23.26	0.1	0.013	0.006
TS	A	40 (0)	36	23.74	0.5	0.008	0.004
	C	34 (0)	29	23.64	0.6	0.010	0.005
ES1	A	40 (0)	34	23.81	0.6	0.008	0.003
	C	35 (0)	30	23.71	0.7	0.008	0.003
ES2	A	40 (0)	34	23.74	0.6	0.016	0.009
	C	35 (0)	30	23.69	0.7	0.017	0.009

An illustrative example of one typical measurement of the first component of shear is shown in Fig. 4. The grey points correspond to sets of rotated and unrotated galaxies, and are explained in Section 4.4. In this example, the negative slope of the black dashed line in the bottom panel (m_1) shows that this method systematically underestimates shear by $\sim$2.5 per cent. However, the negligible y-intercept shows that the PSF was successfully corrected and no residual shear calibration (c_1) remained. The measurement of the second component of shear is not shown. Note that the range of input shear values is smaller than STEP1 and, in this weak shear régime, none of the methods exhibit the non-linear response to shear seen with the strong signals in STEP1. We therefore do not attempt to fit a quadratic function to any of the shear in *vs* shear out results.

4.4 Combining rotated and unrotated galaxies

An important advance in this second STEP project is the simultaneous analysis of galaxies that had been rotated by 90° before the application of shear and convolution with the PSF. This can largely remove noise due to scatter in galaxies' intrinsic morphology, but complicates the production of a joint shear catalogue, especially where the galaxies are given different weights in the two catalogues.

Taking the rotated and unrotated sets of images individually, we obtain two sets of mean shear estimators $\langle\tilde{\gamma}^{\rm unrot}\rangle$ and $\langle\tilde{\gamma}^{\rm rot}\rangle$, which are defined in equations (25) and (26). We typically find that $m_i^{\rm rot} \approx m_i^{\rm unrot}$ and $c_i^{\rm rot} \approx -c_i^{\rm unrot}$. Such stability to changes in image rotation is to be expected: cross-talk between ellipticity and shear directions are second order in γ according to equation (2), and the mean ellipticity is overwhelmingly dominated by the intrinsic ellipticities of a finite number of galaxies (as demonstrated by the offset between the squares and diamonds in Fig. 4). Intruigingly, for the MS1 and MS2 methods, the shear calibration bias changes significantly between the rotated and the unrotated catalogues, and when the two are matched. These methods use smaller galaxies than most, including some 10–25 per cent around or below the stellar locus on a size *vs* magnitude plane, and this effect may be caused by instabilities in the PSF correction of the smallest. As an alternative explanation, there are also second-order effects inherent in the non-linear lensing equation that involve the dot product of ellipticity and shear, which would become significant in the presence of an ellipticity-dependent selection bias. However, we do not understand

Table 7. Tabulated values of shear-calibration bias ($\times 10^{-2}$) from Fig. 5. In each entry, the top line refers to the first component of shear, and the bottom line to the second.

Author	Image set A	Image set B	Image set C	Image set D	Image set E	Image set F
JB	2.34 ± 2.89	-1.46 ± 2.46	5.08 ± 3.06	7.31 ± 3.08	3.44 ± 3.02	1.92 ± 3.14
	5.34 ± 2.91	-7.78 ± 2.05	-2.57 ± 2.87	2.12 ± 3.01	-6.85 ± 3.77	-1.25 ± 3.31
C1	-9.33 ± 1.12	-6.30 ± 1.12	-15.78 ± 1.27	-17.01 ± 1.05	-15.60 ± 1.09	-9.18 ± 1.24
	-7.44 ± 1.07	-6.06 ± 0.97	-11.69 ± 1.19	-18.11 ± 0.97	-18.90 ± 1.35	-9.22 ± 1.32
C2	-7.97 ± 1.13	-4.13 ± 1.14	-12.68 ± 1.35	-7.39 ± 1.19	-7.64 ± 1.19	-5.50 ± 1.28
	-6.05 ± 1.12	-4.43 ± 0.97	-8.31 ± 1.24	-9.16 ± 1.13	-11.99 ± 1.49	-6.50 ± 1.37
MH	3.73 ± 1.30	7.76 ± 1.46	6.26 ± 1.55	2.51 ± 1.37	0.82 ± 1.41	1.78 ± 1.56
	3.67 ± 1.24	7.70 ± 1.23	3.39 ± 1.52	4.57 ± 1.39	-2.88 ± 1.75	0.86 ± 1.63
HH	-0.05 ± 0.81	-3.57 ± 0.84	1.53 ± 0.94	-1.94 ± 0.82	-1.33 ± 0.83	-0.30 ± 0.90
	-1.88 ± 0.79	-3.33 ± 0.70	0.28 ± 0.88	-2.50 ± 0.81	-4.95 ± 1.04	-1.89 ± 0.94
MJ	-1.05 ± 1.18	-1.07 ± 1.01	-0.35 ± 1.31	-0.98 ± 1.20	-1.92 ± 1.21	-2.81 ± 1.30
	-4.07 ± 1.10	-2.11 ± 0.84	1.09 ± 1.21	-0.75 ± 1.16	-3.18 ± 1.49	-3.54 ± 1.33
MJ2	-0.74 ± 0.97	-3.99 ± 0.89	0.81 ± 1.04	-0.73 ± 0.94	-0.01 ± 0.94	-1.77 ± 0.96
	-3.04 ± 0.90	-3.20 ± 0.75	1.58 ± 0.98	-1.49 ± 0.92	-4.10 ± 1.14	-1.14 ± 1.01
KK	-1.06 ± 1.05	-0.74 ± 1.21	-6.28 ± 1.26	-3.38 ± 1.15	-3.04 ± 1.13	-2.58 ± 1.19
	-2.20 ± 1.01	-1.96 ± 1.01	-4.34 ± 1.17	-2.38 ± 1.10	-4.74 ± 1.36	-4.51 ± 1.28
RM	-1.88 ± 0.97	-4.05 ± 0.90	1.08 ± 1.14	-1.13 ± 1.04	-0.99 ± 1.04	-0.39 ± 1.14
	-3.58 ± 0.94	-3.91 ± 0.75	-0.65 ± 1.11	-3.67 ± 0.99	-6.17 ± 1.26	-4.20 ± 1.22
RN	-2.28 ± 1.27	-0.79 ± 1.16	-4.16 ± 1.57	-3.52 ± 1.33	-3.90 ± 1.35	-6.20 ± 1.46
	-4.85 ± 1.21	-3.04 ± 0.96	-6.55 ± 1.48	-5.26 ± 1.28	-7.68 ± 1.66	-6.18 ± 1.53
SP	-10.52 ± 1.25	-7.52 ± 1.40	-12.60 ± 1.49	-12.67 ± 1.55	-14.41 ± 1.34	-12.20 ± 1.44
	-3.96 ± 1.25	-3.49 ± 1.31	-6.89 ± 1.55	-5.66 ± 1.56	-9.62 ± 1.87	-6.91 ± 1.60
MS1	-15.19 ± 1.15	-13.40 ± 1.00	-22.79 ± 1.30	-11.85 ± 1.22	-15.45 ± 1.25	-13.93 ± 1.29
	-15.79 ± 1.11	-12.76 ± 0.85	-21.68 ± 1.24	-11.92 ± 1.19	-19.01 ± 1.45	-14.87 ± 1.56
MS2	-3.40 ± 1.75	-8.09 ± 1.30	-12.55 ± 2.31	-0.70 ± 2.08	-0.68 ± 1.97	-1.99 ± 2.10
	-2.94 ± 1.75	-4.18 ± 1.19	-6.55 ± 2.21	5.13 ± 2.07	-11.98 ± 2.61	-1.70 ± 2.40
TS	-1.43 ± 1.47	2.82 ± 1.57	0.26 ± 1.87	-2.76 ± 1.55	-3.69 ± 1.58	-2.04 ± 1.74
	-0.97 ± 1.38	1.88 ± 1.30	-2.54 ± 1.67	-1.11 ± 1.56	-7.81 ± 1.98	-2.60 ± 1.79
ES1	-15.51 ± 1.27	-8.11 ± 1.29	-19.03 ± 1.34	-19.09 ± 1.26	-17.31 ± 1.26	-12.45 ± 1.45
	-18.07 ± 1.21	-8.02 ± 1.06	-21.05 ± 1.19	-19.65 ± 1.17	-20.60 ± 1.60	-16.80 ± 1.51
ES2	13.66 ± 3.28	11.68 ± 3.34	-1.36 ± 3.47	3.03 ± 2.97	1.06 ± 2.85	3.00 ± 3.47
	4.61 ± 3.10	14.64 ± 2.70	-4.93 ± 3.20	3.10 ± 2.73	-3.82 ± 3.61	-7.25 ± 3.74

why this would affect only this pipeline and not others. We have not attempted to investigate this isolated effect in more detail.

We obtain a third set of parameters m_i and c_i from the matched catalogue with $\langle\tilde{\gamma}\rangle$ defined in equation (27). In general, we find that $m_i \simeq (m_i^{\rm unrot} + m_i^{\rm rot})/2$ and $c_i \simeq c_i^{\rm unrot} - c_i^{\rm rot}$, with significantly smaller errors in this matched analysis. An example of all three shear estimators for the KK method on image set F are plotted in Fig. 4. The fitted parameters for all the shear-measurement methods, on all the PSFs, are shown in Fig. 5. Parameters measured from the matched pair analysis are also tabulated in the appendix. Results from the most successful methods are averaged across all the sets of simulated images and compared directly in Fig. 6.

4.5 Analysis as a function of galaxy population

It is possible to measure the mean shear correctly from a large population of galaxies, but to underestimate the shears in some and overestimate it in others. This was frequently found to be the case in STEP2 data as a function of galaxy size or magnitude, but correlations could also be present as a function of galaxy morphological type. Anything that correlates with galaxy redshift is particularly important, and Fig. 7 shows the correlation of shear calibration bias and residual shear offset with galaxy size and magnitude for an illustrative selection of shear-measurement methods. Of course, these proxies are not absolute: the fundamental parameters of interest are the size of galaxies relative to the pixel or PSF size, and the flux of galaxies relative to the image noise level. This must be taken into account before drawing parallel conclusions on data sets from willower surveys or those taken in different observing conditions.

The results for the TS method are fairly representative of most implementations of KSB+. The calibration bias changes by 0.2–0.3 between bright and faint galaxies. The mean shear calibration bias changes between methods by merely raising or lowering this curve. The ES2 curve is least affected, with only a $\sim$5 per cent change. The shear calibration bias also generally changes as a function of galaxy size. The HH method controls this the best, no doubt due to its fitting of P^γ as a function of size only. However, this method still displays significant variation as a function of magnitude; it is not clear in Fig. 7 because the final point expands the y-axis scale. The fairly constant residual shear offset as a function of galaxy magnitude is typical; as is the dramatic improvements for bigger galaxies in the image sets D and E with highly elliptical PSFs. That demonstrates that it is a PSF-correction problem. The RM method behaves similarly to the implementations of KSB+.

Other methods exhibit more idiosyncratic behaviour. The main difference is between the KK method and the others that use a global shear responsivity $\mathcal{R}$. This was calculated only once for the KK method, from the entire galaxy population. For the other methods, it was recalculated using a subset of galaxies for each size and magnitude bin. The large trends in the shear calibration bias as a function of size and magnitude merely reflect the evolving distribution of intrinsic galaxy ellipticities. The MJ, MJ2, RM and RN methods also all look like this with a single value of $\mathcal{R}$, and the KK method would presumably be improved by this step. The JB results are atypical,

Table 8. Tabulated values of residual shear offset ($\times 10^{-4}$) from Fig. 5. In each entry, the top line refers to the first component of shear, and the bottom line to the second.

Author	Image set A	Image set B	Image set C	Image set D	Image set E	Image set F
JB	-6.8 ± 6.5	-17.2 ± 5.4	-34.5 ± 7.5	24.5 ± 7.6	83.7 ± 8.0	17.4 ± 7.3
	1.3 ± 6.6	-15.0 ± 5.5	-1.0 ± 7.5	-80.3 ± 7.6	46.5 ± 8.0	10.6 ± 7.5
C1	21.2 ± 2.5	26.7 ± 2.5	-5.2 ± 3.2	124.2 ± 2.6	64.1 ± 2.9	-11.8 ± 2.9
	21.2 ± 2.5	-5.4 ± 2.6	23.2 ± 3.1	-70.0 ± 2.5	130.2 ± 2.9	8.5 ± 2.9
C2	-3.3 ± 2.5	-1.1 ± 2.5	-21.6 ± 3.3	259.4 ± 2.9	29.7 ± 3.2	-6.2 ± 2.9
	38.3 ± 2.6	18.8 ± 2.6	39.8 ± 3.2	-36.6 ± 2.9	276.6 ± 3.2	3.6 ± 3.0
MH	10.2 ± 3.0	19.8 ± 3.3	19.6 ± 3.9	101.2 ± 3.4	91.6 ± 3.8	-4.4 ± 3.6
	5.4 ± 3.0	21.9 ± 3.3	6.7 ± 4.0	-84.2 ± 3.6	99.3 ± 3.8	6.3 ± 3.6
HH	1.6 ± 1.8	-4.8 ± 1.8	-6.1 ± 2.3	3.7 ± 2.0	75.2 ± 2.2	-2.2 ± 2.1
	-4.6 ± 1.8	-3.1 ± 1.8	-0.6 ± 2.2	-65.5 ± 2.0	-5.9 ± 2.1	9.8 ± 2.1
MJ	-11.8 ± 2.5	-9.5 ± 2.2	-6.5 ± 3.2	18.2 ± 2.9	13.8 ± 3.1	-2.2 ± 2.8
	-0.9 ± 2.6	6.0 ± 2.1	1.6 ± 3.1	-12.7 ± 2.8	16.6 ± 3.0	1.2 ± 2.8
MJ2	-10.3 ± 1.9	-4.8 ± 1.7	0.4 ± 2.3	23.9 ± 1.9	15.5 ± 2.2	-0.8 ± 1.9
	1.5 ± 1.9	3.1 ± 1.7	2.3 ± 2.2	-16.8 ± 2.0	19.7 ± 2.1	1.2 ± 1.9
KK	-2.1 ± 2.4	-5.2 ± 2.7	-14.0 ± 3.1	-71.6 ± 2.8	66.6 ± 3.0	0.1 ± 2.8
	-2.7 ± 2.4	-6.6 ± 2.7	2.1 ± 3.0	-69.5 ± 2.7	-56.9 ± 2.9	-3.9 ± 2.8
RM	22.9 ± 2.2	14.9 ± 2.0	26.5 ± 2.9	-33.5 ± 2.5	112.0 ± 2.8	0.1 ± 2.6
	-9.9 ± 2.2	-3.1 ± 1.9	-5.8 ± 2.8	-105.7 ± 2.5	-19.4 ± 2.7	2.4 ± 2.7
RN	-5.3 ± 2.8	-5.0 ± 2.5	-6.3 ± 3.8	-34.9 ± 3.1	43.1 ± 3.4	2.5 ± 3.1
	1.8 ± 2.7	-0.1 ± 2.5	8.9 ± 3.7	-33.1 ± 3.2	-26.8 ± 3.3	4.6 ± 3.2
SP	-1.1 ± 2.5	-3.4 ± 2.9	-4.5 ± 3.3	-69.9 ± 3.3	71.6 ± 3.4	5.5 ± 3.0
	-1.1 ± 2.7	-7.6 ± 3.2	-4.6 ± 3.8	-55.3 ± 3.6	-13.3 ± 3.5	4.1 ± 3.3
MS1	-5.6 ± 2.5	-22.3 ± 2.1	4.9 ± 3.2	105.1 ± 3.0	58.5 ± 3.3	-7.6 ± 3.1
	10.3 ± 2.6	23.1 ± 2.1	7.2 ± 3.2	-45.7 ± 3.0	83.8 ± 3.1	6.1 ± 3.2
MS2	-7.9 ± 3.9	-21.3 ± 2.8	3.2 ± 5.6	140.5 ± 5.0	41.5 ± 5.2	-0.4 ± 4.9
	14.4 ± 4.0	24.3 ± 3.0	19.5 ± 5.9	-28.7 ± 5.1	154.4 ± 5.4	9.2 ± 5.3
TS	-2.9 ± 3.3	-4.3 ± 3.5	2.7 ± 4.5	-46.2 ± 3.9	70.3 ± 4.3	-3.5 ± 4.0
	-3.0 ± 3.2	-1.3 ± 3.6	0.4 ± 4.4	-65.4 ± 3.9	-40.3 ± 4.2	-3.1 ± 4.0
ES1	-9.1 ± 2.8	-4.1 ± 2.9	5.7 ± 3.3	153.1 ± 3.1	54.3 ± 3.4	-5.5 ± 3.4
	4.0 ± 2.8	8.7 ± 2.8	9.9 ± 3.2	-58.7 ± 3.0	132.0 ± 3.4	0.4 ± 3.3
ES2	-11.0 ± 7.4	8.5 ± 7.4	15.0 ± 8.2	95.3 ± 7.1	96.7 ± 7.8	-10.4 ± 8.2
	-11.2 ± 7.4	-3.3 ± 7.2	5.7 ± 8.4	-92.9 ± 7.1	77.9 ± 7.7	7.7 ± 8.1

but their additional noise level represents that in all analyses lacking an optimal galaxy weighting scheme.

5 INTERPRETATION

We will now revisit the questions posed in the introduction, concerning the accuracy with which current methods can measure shear, and in which régimes that accuracy begins to deteriorate. By noting the variation of results with different PSFs, we will investigate the effects of changing atmospheric and observing conditions. We will also investigate the effects of image pixellization, galaxy morphology and morphology evolution, selection biases and weighting effects. In light of our results, we will then review the consequences for previously published measurements of cosmic shear.

The rotated pairs of galaxies provide an unprecedented level of discriminatory power, and we can now identify high level causes of shear-measurement error. Overall, both the shear calibration (multiplicative) bias and the anisotropic PSF correction (additive) errors depend on the PSF model. From this information, we can deduce that some aspects of shape measurement have been suitably controlled. We can deduce that others still provide difficulty, and it is work in these identified areas that will provide a route to the desired sub-per cent level of precision. This section describes various lessons that we have learned from our tests, in terms of high-level variables.

5.1 PSF size

Within the precision accessible by this analysis, all the methods are reassuringly tolerant to reasonable changes in observing conditions. Image set A (0.6-arcsec FWHM PSF) represents typical seeing at a good site, and image set C (0.8-arcsec FWHM PSF) the worst that might be expected for a weak-lensing survey after appropriate telescope scheduling.

Differences in the residual shear offsets between the two sets of images with different seeing are generally not significant. The few methods with a significant difference are JB, MH, KK and ES. In all four cases, the 2–3σ offset is in c_1 but not c_2. The two KSB+ methods have a positive offset, and the two shapelets methods have a negative one, but no general conclusion seems manifest.

As expected, most methods demonstrate minimal shear calibration bias with image set A, and fare slightly worse on image set C. Shear calibration bias for the JB and RN methods is stable to changes in observing conditions at the $\sim$0.5 per cent level. The MH KSB+ method achieves $\sim$1 per cent consistency, although its applied shear calibration factor is apparently a little overzealous.

No global trends emerge that are able to include all the KSB+ methods. However, for the generally most successful KSB+ implementations by MH, HH and TS, as well as the BJO2 (MJ, MJ2) methods, m is higher in image set C than in set A. These methods are all on the top row of Table 3, and correct for the PSF by subtracting combinations of shape moments. The trend is reversed in the KK deconvolution method on the bottom row, and the

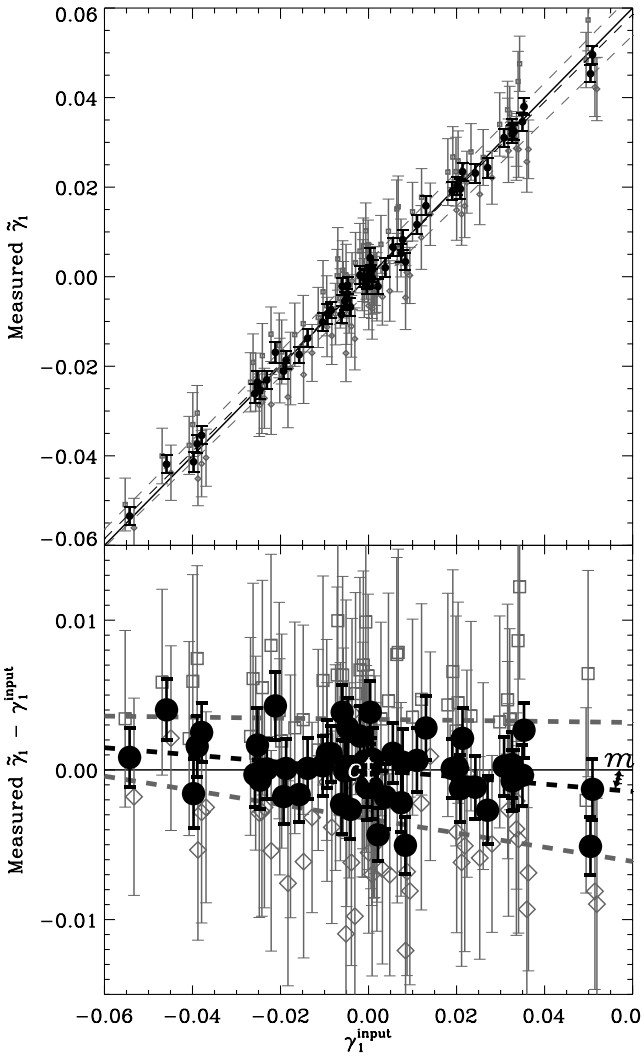

Figure 4. An example of the input *vs* measured shear for one representative method. This is for the first component of shear measured by the KK method in image set F. It is neither the best method on this image set, nor the best image set for this method, but shows behaviour that is typical of most. The grey squares and diamonds show results from independent analyses of the rotated and unrotated images; the black circles show the effect of matching pairs of otherwise identical galaxies. The bottom panel shows deviations from perfect shear recovery, which is indicated in both panels by solid lines. Linear fits to the data are shown as dashed lines. The fitted parameters m (shear-calibration bias) and c (residual shear offset) are plotted for all methods and all for all images sets in Fig. 5.

calibration bias does not vary in the JB and RN methods. These correct for the PSF via a full deconvolution. Although all implementations of KSB+ do not necessarily fit this trend, it does suggest that the isotropic component of the PSF might be being overcorrected by some moment subtraction schemes. Furthermore, as the PSF moments get larger, this oversubtraction exaggerates pixellization effects (see Section 5.3). The best PSF correction is generally attained by methods that model the full PSF and attempt to deconvolve each galaxy – but this currently works on slightly fewer galaxies (see Section 5.6).

5.2 PSF ellipticity (and skewness)

Image sets D and E demonstrate the ability of methods to correct for highly elliptical PSFs, and can be compared to image set F, which has a circularly symmetric PSF. Imperfect correction for PSF anisotropy will emerge mainly as a residual additive shear offset, c. The method that was most efficient at removing all the different strengths of PSF anisotropy to better than 0.2 per cent accuracy was MJ/MJ2, and all the PSF deconvolution methods had better than 1 per cent accuracy. The most successful KSB+ correction was the HH implementation. The residual shear offsets are smallest with large galaxies, and deteriorate only as galaxies get smaller. This behaviour is as expected if the problems are caused by imperfect PSF correction.

Many methods have a spurious residual shear offset in both components of shear, while the PSF is highly elliptical in only the ε_1 or ε_2 direction. This cross-contamination might come from the ignored off-diagonal elements of the $P^{\rm sm}$ tensor in KSB+, and is indeed slightly better controlled in MS2 (with the full tensor inversion) than in MS1. However, this cannot explain all of the effect; the off-diagonal elements *are* exactly zero for the circular PSF in image set F, and a few methods (JB, C1, RN, SP, MS1, ES2) have a significantly non-zero residual shear offset for even this set of images.

A more likely source of the contamination lies in the measurement of stellar ellipticities. The non-zero residual shear offsets with image set F probably come from shot noise in the measurement of PSF ellipticity, which is higher than the shot noise for galaxies because of the smaller number of stars. It will therefore be worthwhile to make sure that future methods gather the maximum possible amount of information about the PSF. In particular, small galaxies provide as much information about the PSF as their own shapes, and this is currently discarded. Furthermore, PSFs D and E are not only highly elliptical, but also skewed. The centre of those PSFs therefore depends strongly on the size of the weight function used. While the main direction of ellipticity is not in doubt, changing the centre of the PSF also perturbs its apparent ellipticity. The C1 method, with a fixed stellar weight function and a constant PSF model, removes stellar ellipticity more consistently that the C2 method, in which the size of the stellar weight function is altered to match each galaxy (although matching the galaxy weight function provides a better shear calibration). Methods that involve deconvolution from a full model of the PSF, or correction of PSF non-Gaussianity, and which allow the galaxy centroid to iterate during this process, do indeed seem to be able to better control PSF ellipticity and centroiding errors.

We cannot conclusively explain the cross-contamination of both shear components by a PSF strongly elongated in only one direction, but hypothesise that it is introduced by skewness and substructure in the PSF. Neither of these are addressed by the formalism of KSB+, and they are both controlled more reliably by newer methods that explicitly allow such variation. However, it is also worth noticing the remarkable success of most methods on other image sets with more typical PSF ellipticities, and remarking that this is still a small effect that will not dominate shear measurement for the near future.

Our investigation of PSF effects in the STEP2 images is confused by other competing manifestations of imperfect shear measurement, and the realism of the simulations. The combination of image pixellization (see Section 5.3), correlated galaxy sizes and magnitudes, and the evolution of intrinsic galaxy size and morphology as a function of redshift all hinder interpretation. Higher

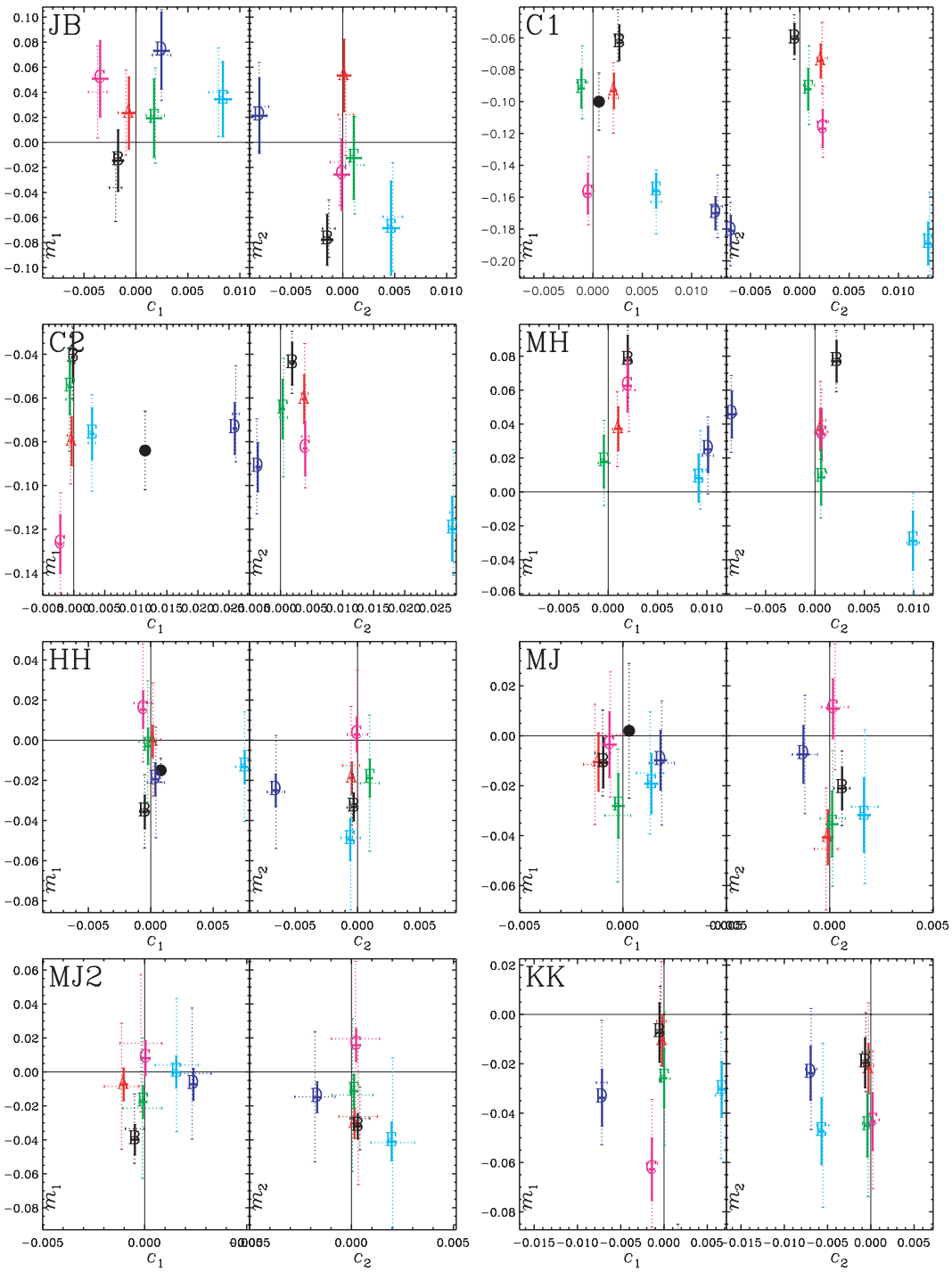

Figure 5. Fitted values of residual shear offset and shear-calibration bias for each method and for each PSF. In all cases, the left hand panel shows results for the γ_1 component of shear, and the right hand panel for the γ_2 component. The dotted lines show rms errors after a combined analysis of the rotated and unrotated galaxies, after the two catalogues have been matched (and only common detections kept). The solid lines show the reduced errors after removing intrinsic galaxy shape noise from the matched the pairs of galaxies. Note that the scales on each panel are different, but the frequency of the axis labels is preserved. The red points correspond to image set A. The black points correspond to image set B, and, where available, the filled black circles reproduce results from STEP1. The pink, dark blue, light blue and green points correspond to image sets C, D, E and F respectively.

precision tests in the future will counterintuitively require *less* realistic simulated images: for example, ones that are tailored to compare otherwise identical galaxies at fixed multiples of the PSF size.

5.3 Pixellization effects

This is the first STEP project in which the input shear has been applied in many directions, and in which the two components of shear

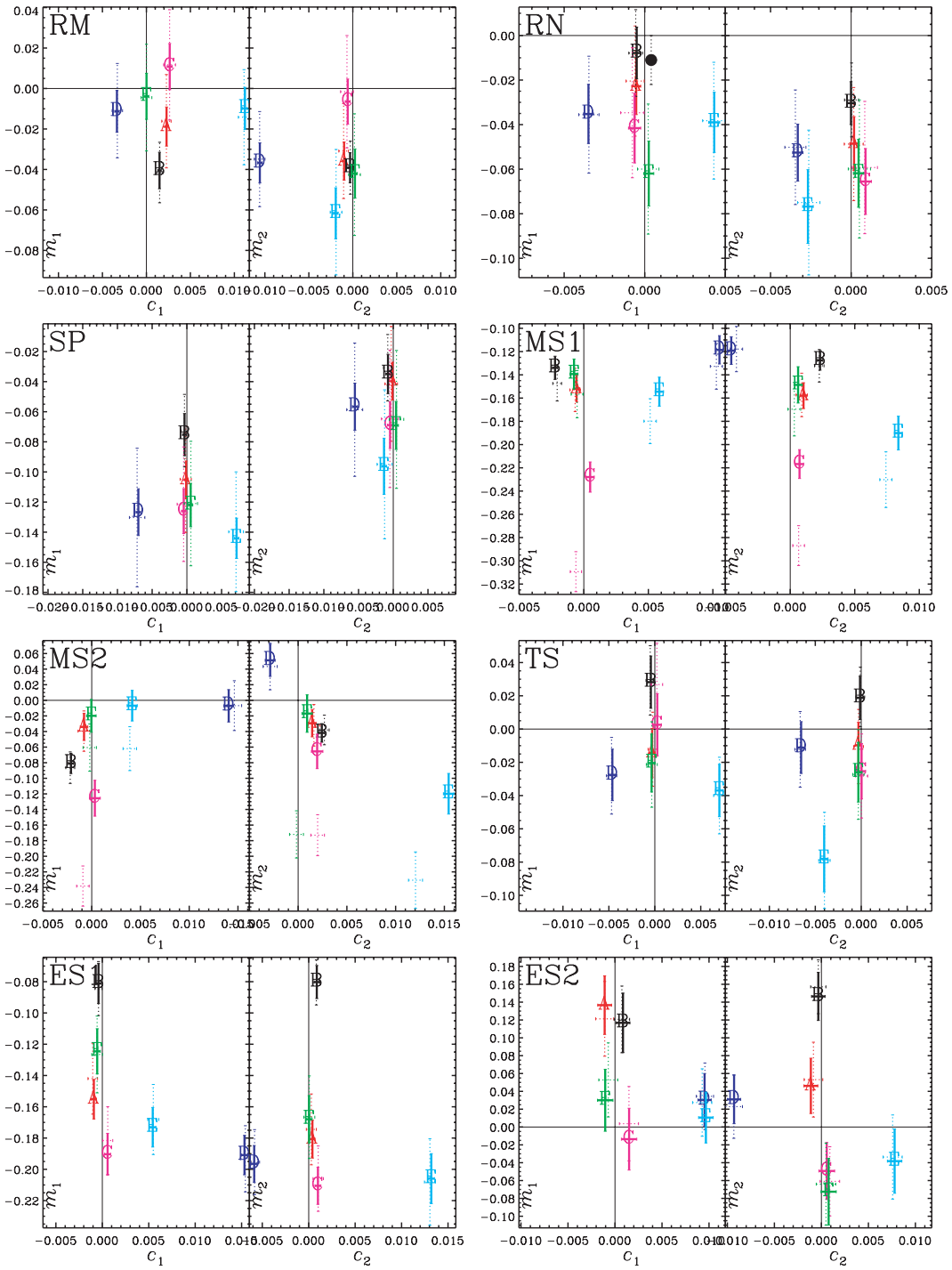

Figure 5 – *continued*

can be measured independently.s In general, residual shear offsets c are consistent between components. However, we find that the γ_1 component, aligned with the square pixel grid, is typically measured more accurately than the γ_2 component, along the diagonals. This is even observed for image set F, in which the analytic PSF is circularly symmetric. Since there is no other preferred direction, this phenomenon must therefore be an effect of pixellization. An image pixellization, which is similar (but not identical) to convolution, slightly circularizes galaxies, thereby reducing their ellipticity. Not explicitly correcting for pixellization may therefore explain both the general 1–3 per cent underestimation of γ_1, and the slightly larger underestimation of γ_2, in which direction the distance between pixels is exaggerated. For almost all methods, we consistently find that $m_1 > m_2$.

In KSB+, there is no formal mathematical framework to deal with image pixellization. Two different approaches have been adopted to approximate the integrals in equation (20) with pixellated data. The C1 and C2 implementations calculate the value of the weight functions at the centre of each pixel and then form a discrete sum; all the others numerically integrate the weight functions by subdividing

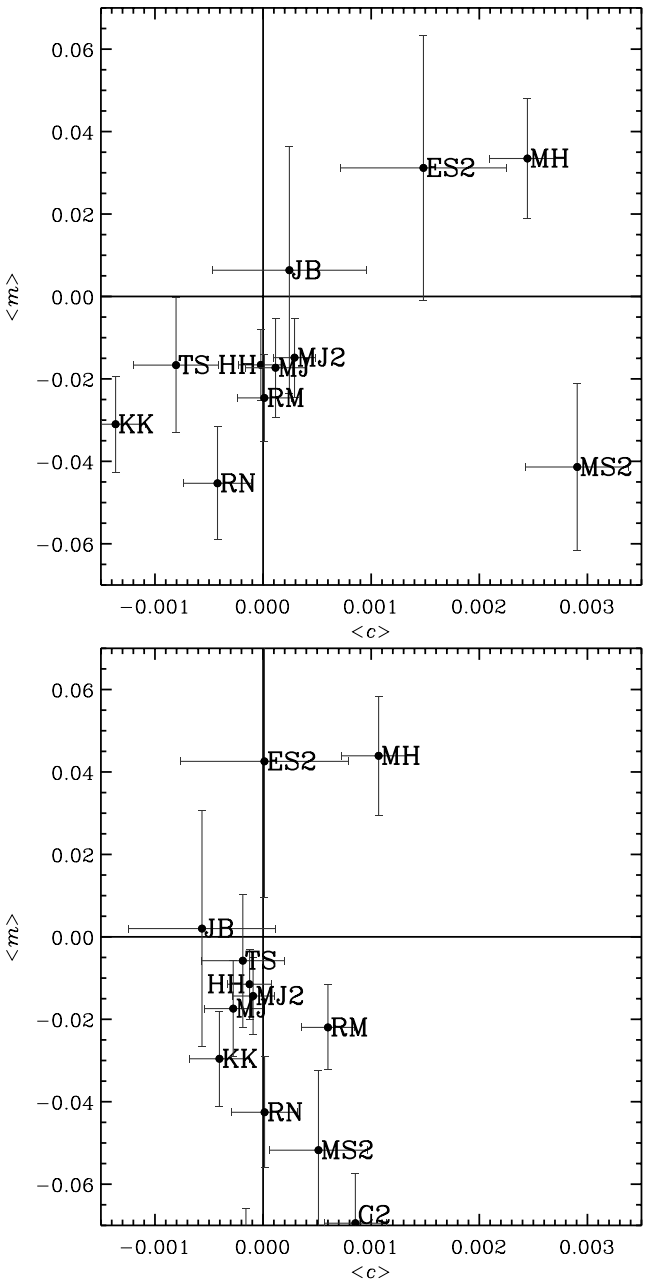

Figure 6. Comparison of shear-measurement accuracy from different methods, in terms of their mean residual shear offset $\langle c \rangle$ and mean shear-calibration bias $\langle m \rangle$. In the top panel, these parameters have been averaged over both components of shear and all six sets of images; the bottom panel includes only image sets A, B, C and F, to avoid the two highly elliptical PSFs. Note that the *entire region* of these plots lie inside the grey band that indicated good performance for methods in Fig. 3 of STEP1. The results from methods C1, SP, MS1 and ES1 are not shown here.

pixels into a number of smaller regions. Neither approach is ideal. Independent experiments by Tim Schrabback, running objects with Gaussian radial profiles though his implementation of KSB+, have shown that pixellization can cause a systematic underestimation of ε and $P^{\rm sm}$, and an overestimation of $P^{\rm sh}$. This effect can be up to $\sim$10 per cent for small objects. However, as stars and faint galaxies are similarly affected, the error on the shear estimate approximately cancels. An integration using linearly interpolated subpixels makes the measurement more stable to the subpixel position of the object centroid, but slightly increases the individual bias. Bacon et al. (2001) tested a variant of the C1 method, and found a similar $\sim$13 per cent overall calibration bias, which was used to correct subsequent measurements. With hindsight, the different calibration of γ_1 and γ_2 are also already visible in that work.

The MJ2, KK and TS methods are least affected by pixellization. This might have suggested that the extraction of a shear estimator by shearing circular objects removes the problem, were it not for the peculiar behaviour of the RN method. For this method, image sets A and C follow the usual pattern that $m_1 > m_2$, but that bias is reversed when PSF is circular (image set F and the zero-ellipticity components of PSFs D and E). The SP method is similar. Strangely, the JB method, which ostensibly tries the hardest to treat pixellization with mathematical rigour, displays the most difference between m_1 and m_2. However, this method does break a trend by not having an overall negative shear calibration bias. If this bias is indeed caused by pixellization, this method appears to have most successfully eliminated it.

Pixellization could also hinder shear measurement, and bring about the observed results, via two additional mechanisms. First, it may exaggerate astrometric errors in the PSF, and produce the consequences described in the previous section. We would be unable to distinguish these effects. Secondly, the undersampling of objects may also fundamentally prevent the measurement of their high order shape moments. All the STEP2 PSFs (and hence the galaxies) are Nyquist sampled. It would be unfortunate for lensing if Nyquist sampling were theoretically sufficient to measure astrometry, but not shapes. As it happens, for methods other than MJ, the pixellization bias is more pronounced for image set C (with poor seeing, and therefore better sampled) than on image set A (with good seeing). This suggests that the pixellization effects are *not* due to undersampling. The STEP1 simulations had the same pixel scale but worse seeing ($\sim$1-arcsec FWHM), so objects were better sampled there.

We therefore hypothesise that the circularizing effects of pixellization explain the general underestimation of shear and the differential calibration of the γ_1 and γ_2 components. Indeed, a dedicated study of simulated images with varying pixel scales by High et al. (in preparation) supports this view. They found that the shear calibration bias of the RRG method tends to zero with infinitely small pixels, grows linearly with pixel scale, and that the bias $m_2 \approx \sqrt{2} m_1$. Because of the isotropy of the Universe, this differential calibration of shear estimators ought not affect two-point cosmic shear statistics. However, it can certainly affect the reconstruction of individual cluster mass distributions, and is inherently quite disconcerting. The next STEP project will feature sets of images with varying pixel scales to investigate this effect on a wider scale. In the mean time, dealing properly with pixellization will provide a promising direction for further improvement in shear-measurement methods.

5.4 Galaxy morphology

The introduction of complex galaxy morphologies tends to hinder shear measurement with KSB+ methods. The shear calibration bias is more negative with image set A (shapelet galaxies) than with image set B (simple galaxies) for the C1, C2, MH, SP, MS1, TS and ES1 implementations. Of the implementations of KSB+, only HH and MS2 reverse this trend. This is perhaps not surprising, given the inherent limitation of KSB+ in assuming that the ellipticity of a galaxy does not change as a function of radius.

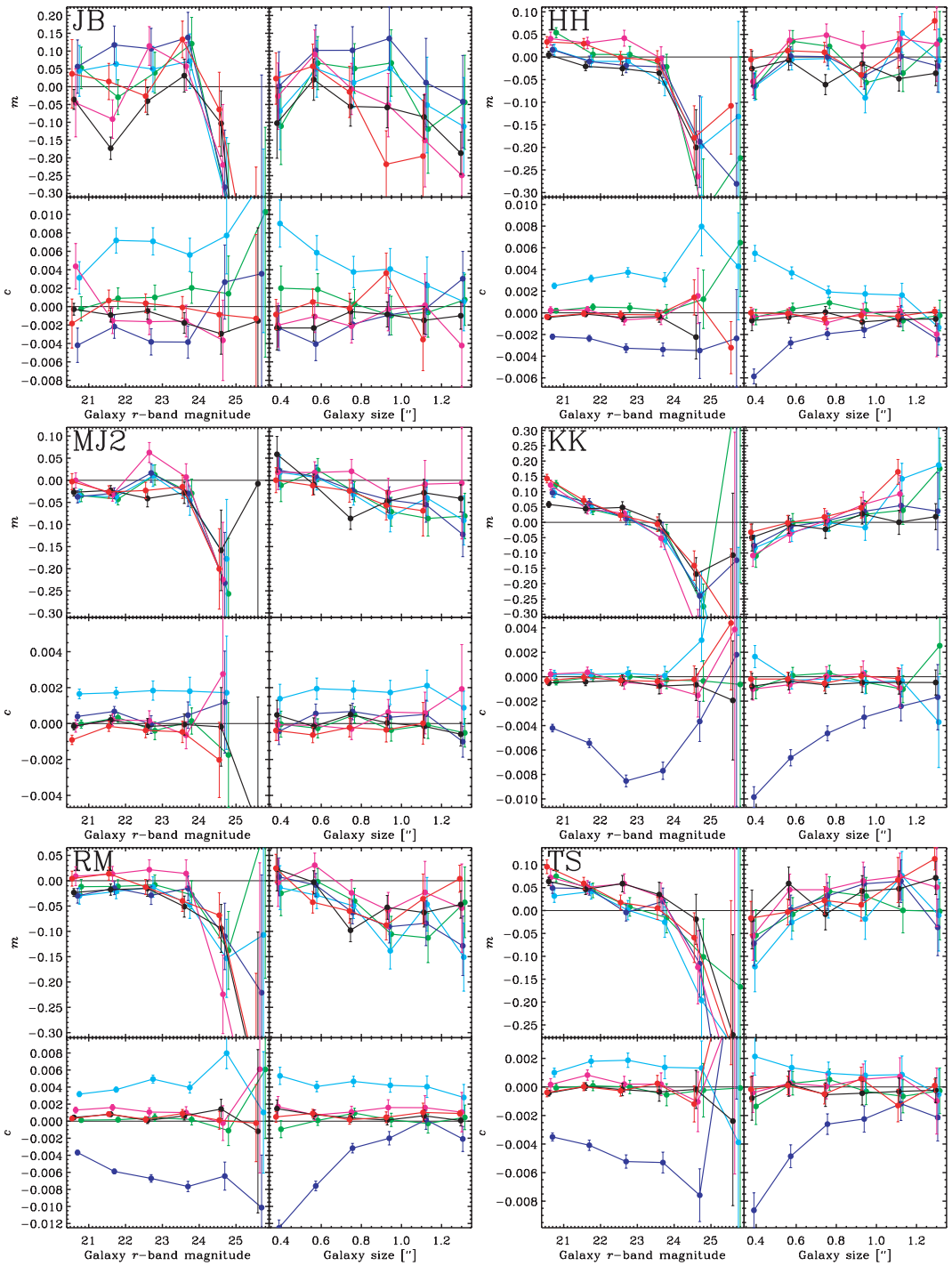

Figure 7. Variation in shear-calibration bias and residual shear offset as a function of galaxy magnitude and size, for a representative sample of methods. The input values of these are used, which do not have noise. The 'size' on the abscissae is the unweighted rms size of galaxies from equation (53) in Massey & Refregier (2005). The six coloured lines in each plot correspond to the six sets of images, coloured in the same way as in Fig. 5. In all cases, measurements of the two components of shear have been averaged.

Many of the newer methods deal with complex galaxy morphologies very successfully. Particularly KK, but also the MJ and MJ2 methods, have no significant difference in the shear calibration bias or residual shear offset measured between image sets A and B. Future ground-based shear surveys are therefore unlikely to be limited at the 0.5 per cent level by complex galaxy morphologies. Indeed, it is apparent in Fig. 2 that most of the substructure in galaxies that will be used for lensing analyis is destroyed by the atmospheric seeing. Although complex galaxy morphologies may become important at the level of a few tenths of a per cent, they do not currently pose a dominant source of error or instability in shear measurement from the ground.

One of the crucial findings of this study, however, concerns the effect of galaxy morphology *evolution*. This could potentially affect

the calibration of shear measurement as a function of galaxy redshift, and is investigated further in the next section.

In the next STEP project, which will simulate space-based observations, we will repeat our investigation of galaxy morphology by comparing three similar sets of image simulations. A galaxy substructure will be better resolved from space and, because the galaxies observed there are likely to be at a higher redshift, their intrinsic morphologies may be both more irregular and more rapidly evolving. Both of these effects will amplify any differences seen from the ground.

5.5 Shear calibration for different galaxy populations

The STEP2 results reveal that the calibration bias of some shear-measurement methods depends on the size and magnitude of galaxies. There seem to be two causes. There is often a sudden $\sim$30 per cent deterioration of performance at very faint magnitudes, due to being noise blown up during the nonlinear process of shear measurement (and exacerbated by ellipticity-dependent galaxy weighting schemes). This is even observed with many methods that are otherwise robust (e.g. HH, MJ2, RN), and may urge more caution in the use of faint galaxies at the limits of detection. There is also a gradual transition in shear calibration between bright and faint galaxies that is probably caused by evolution of the intrinsic morphology distribution as a function of redshift. The observed variation is least pronounced for image set B, in which the galaxies explicitly do not evolve.

Shear-calibration bias that changes gradually as a function of galaxy redshift has important consequences for any weak-lensing measurement. In a 2D survey, it will change the effective redshift distribution of source galaxies, with all the consequences discussed by Van Waerbeke et al. (2006). In a 3D analysis, it will affect the perceived redshift evolution of the matter power spectrum, and the apparent large-scale geometry of the universe. During the STEP2 analysis, we have developed ways to partially control this, as a function of other observables like galaxy size and magnitude. To first order, these act as suitable proxies for redshift, but the underlying causes will need to be well understood, because neither of these are redshift. Even if the mean shear in size–magnitude bins could be made correct, this does not necessarily imply that the mean shear would be correct in redshift bins. The techniques could be applied in multicolour surveys as a function of photometric redshift, but this is not perfect either, not least because of the inevitable presence of catastrophic photo-z failures.

The obvious place to start looking for shear-calibration errors is in the shear susceptibility and responsivity factors. All the KSB+ implementations allow variation in P^γ as a function of at least one of galaxy size and galaxy magnitude. However, the behaviour is neither well understood, nor stable at the desired level of precision. Massey et al. (2005) have already observed that P^γ fitted from a population ensemble varies for any given object as a function of the catalogue selection cuts. There is less variation in the shear-calibration bias of the MS1 method ($\Delta m \approx 0.1$), which fits only the trace of P^γ, than of the MS2 method ($\Delta m \approx 0.2$), which models the entire tensor – except for image set B, in which there is little variation in either. Realistic galaxy morphologies therefore do not have shear susceptibility that is a simple function of these observables; and trying to model the variation of all the components of this tensor merely adds noise. The TS implementation of KSB+, which uses P^γ from individual objects, suffers particularly from this noise, which enters into the denominator equation (18), and has at least as much sudden deterioration at faint magnitudes as other methods. However, this method is about the least affected by gradual variation in shear-calibration bias, with $\Delta m \approx 0.05$. Size galaxy size and magnitude are correlated, the variation with galaxy magnitude usually carries over to variation with galaxy size. However, the HH method has notably little variation in m as a function of galaxy size. This is presumably due to the particularly individual form of the function used to model $P^\gamma(r_g)$. Unfortunately, P^γ is not fitted as a function of galaxy magnitude, and the HH method still shows strong ($\Delta m \approx 0.1$) variation with this. The shear susceptibility in this implementation is calculated separately in three magnitude bins, and correction of the faintest galaxies therefore required an extrapolation.

Many of the other shear-measurement methods require global calibration via a responsivity $\mathcal{R}$ factor, which is determined from the distribution of galaxy ellipticities. This factor is designed to ensure that the mean shear in a population is unbiased. However, it must be calculated from precisely that population. For the KK method, it was calculated only once, from the entire catalogue. Although it estimated the overall mean shear correctly, it then underestimated the shear in small/faint galaxies, and overestimated that in large/bright galaxies. This bias was addressed for the MJ, MJ2, RM and RN methods by recalculating $\mathcal{R}$ within each size and magnitude bin. There is no particular reason why this should not, in future, be fitted and allowed to vary continuously like the shear susceptibility in KSB+ methods. The estimates of $\mathcal{R}$ in bins were more noisy, but removed the differential shear calibration (in fact, the variation as a function of galaxy magnitude was slightly overcorrected in the case of the MJ2 and RM methods).

5.6 Galaxy selection effects

There is a marked difference between the depth of the various galaxy catalogues. At one extreme, the C1/C2 catalogues are deeper, and more ambitious, than all others. At the other, the RN catalogue (and to some extent the MJ/MJ2 catalogue) is very willow. The RN method obtained extremely good results, but only from large and bright galaxies, and it would be interesting to test whether its PSF deconvolution iteration can converge with a deeper sample. The JB catalogue of individual rotated and unrotated images is deeper, but not all the galaxies at the magnitude limit converged successfully, leading to a relatively willow matched catalogue. We could conclude from this that the full deconvolution of every galaxy is an overly ambitious goal: it is a panacea for many image analysis problems, but all that we require is one shear estimator. Maximising the number density of useable galaxies will remain crucial in the near future, to overcome noise from their intrinsic ellipticities. However, there has been far less time spent developing the deconvolution methods than the moment subtraction methods, so we reserve judgement for now because of their promise of robust PSF correction. Furthermore, it is not only the methods that require complicated iterations that suffer from catalogue shortcomings: the SP catalogue includes a significant number of spurious detections (10 per cent) and stars (1 per cent). Neither of these contain any shear signal, and their presence partly explains the large, negative calibration bias of the SP method in the rotated and unrotated images (they are removed during the galaxy matching).

Most other methods use a fairly standard density of $\sim$30 galaxies per square arcminute in this simulated data. This is unlikely to be increased dramatically by any future weak-lensing observations. Since selection effects in the STEP2 analysis must be measured from the individual unrotated and rotated catalogues, rather than

the matched catalogues, the results about catalogue selection biases are hardly more profound than those of STEP1.

5.7 Galaxy weighting schemes

The weighting schemes applied to galaxies also vary significantly between methods used in this paper, and these do affect the results in the matched catalogue. Most of the methods increase the contribution to the estimated mean shear from those galaxies whose shapes are thought to be most accurately measured. Such schemes have long been used in the analysis of real 2D data, but the exact form of the weighting scheme as a function of size, magnitude and ellipticity varies widely. Even more sophisticated weighting schemes will also need to be developed for the 3D analyses essential to fully exploit future weak-lensing surveys.

In this analysis, the effectiveness of each weighting scheme can be seen in the difference between the size of error bars in the analysis of independent galaxies and of rotated/unrotated pairs of matched galaxies. In the independent analysis, the scatter includes components from intrinsic galaxy shapes and measurement noise (e.g. due to photon shot noise). The former is essentially removed by matching pairs of galaxies. If a set of error bars shrink dramatically by matching, the method was dominated by intrinsic galaxy shapes: this is an ideal situation. If the error bars change little, the measurement was dominated by measurement noise.

The weighting schemes of MJ2 and KK are very effective in this analysis: their error bars shrink by up to 75 per cent. The weighting schemes of HH, SP and MJ are similarly effective – but these methods weight ellipticities using a function of ellipticity, which may be less accurate in regimes where the mean shear is large, such as cluster mass reconstruction. Indeed, the aggressive weighting scheme of MJ was shown in STEP1 to be useful with small input shears, but introduced a non-linear shear response that became important if the shear was high. A new weighting scheme was developed for MJ2 to address this concern; however, the range of input shears in STEP2 does not provide sufficient lever arm to evaluate the potential nonlinear response of any method.

The value of a successful weighting scheme is demonstrated by the lesser performance of methods without one. The JB, TS and ES2 methods apply crude weighting schemes that are merely a step function (cut) in galaxy size and magnitude. Their error bars shrink by only 30–50 per cent during galaxy matching. Their results are also less stable to the sudden deterioration of performance seen in several methods with galaxies fainter than or smaller than a particular limit. This shortfall is easy to correct, and we urge the rapid adoption of a more sophisticated weighting scheme in those methods.

It is important to remember the limitations of the STEP simulations to optimize a galaxy weighting scheme, because of their inherent simplification that all galaxies are sheared by the same amount. In real data, the lensing signal increases cumulatively with redshift, and the distant galaxies therefore contain the most valuable signal. However, when weighting objects by the accuracy of their shape measurement, it is the contribution of these small, faint sources that is usually downweighted. It would instead be better to set weights that vary as a function of the S/N in shear signal – although the exact variation of the signal is of course unknown in advance. A statistically 'optimal' weighting scheme verified from the STEP simulations will therefore not be optimal in practice. Weighting schemes can also act like calibration biases as a function of galaxy redshift, exacerbating the problems of differential shear calibration discussed in the previous section.

5.8 Consequences for previously published measurements

The largest cosmic shear survey to date, which has been published since STEP1, comes from the Canada–France–Hawaii Telescope Legacy Survey (CFHTLS) *i*-band data. The CFHTLS wide survey (Hoekstra et al. 2006) was analysed using the HH shear-measurement method, and the CFHTLS deep survey (Semboloni et al. 2006) using the ES1 method. These methods perform very differently on the simulated images.

The HH method recovers shear in the STEP2 images with remarkable success. The seeing in the CFHTLS data is most similar to that in image set C, for which the overall shear calibration is within 1 per cent: well within the current error budget. Hoekstra et al. (2006) also featured a parallel analysis using an independent KSB+ pipeline, which agreed with the HH results, and also demonstrates the potential robustness of KSB+ at this level of precision (similar comparisons have also been performed by Massey et al. (2005) and Schrabback et al. (2006), and these also give results consistent with that work). The HH method had difficulty only with the calibration of very faint galaxies, due to its non-smooth fitting of P^γ as a function of magnitude. If a similar bias is present in the CFHTLS analysis, it will have lowered the effective redshift distribution of source galaxies, and slightly diluted the overall signal. Both of these effects would have led to an underestimation of σ_8, although only by a small amount, due to the low weight given to faint galaxies. As discussed by Van Waerbeke et al. (2006), a more significant bias (which acts in the opposite sense) arises from using the Hubble Deep Field to infer the redshift distribution of galaxies. As the survey area of the CFHTLS grows, and the statistical error bars decrease, it may be prudent for this analysis to conservatively use slightly fewer galaxies.

The ES1 method underestimates shear in the STEP2 images by 20 per cent overall, and by as much as 30 per cent for the faintest galaxies. We have verified this result retrospectively in STEP1 simulations, and also confirmed it in real images, by comparing the results of the HH and ES1 shear-measurement pipelines on the same CFHTLS deep data. Of course, the true 'input' shear is not known for real data. Fig. 8 shows the *relative* calibration of the two methods in real data, with the dashed line indicating their relative calibration in simulated image set C. This should not be interpreted as a strict prediction, since the simulation was not designed to mimic this specific survey: the simulated and real data have very different noise properties, and the only similarity between their PSFs is their size. None the less, the agreement is impressive. Fig. 9 shows a further comparison of the methods' relative calibration, in which galaxies have been split by size and magnitude. Once again, overlaying the performance of ES1 from image set C confirms the results of the STEP simulations with remarkable success. A likely source of the shear-calibration bias is in the smoothing of P^γ as a function of r_g and magnitude. Tests indicate that the shear susceptibility is more stable if it is instead fitted as a smooth function of size and magnitude, or even by using the raw values. The strong magnitude dependence is probably related to the sudden drop at small sizes. Note also that both pipelines started from scratch with the individual exposures, reducing them and stacking them independently. All the available exposures are stacked in both versions, so the two sets of images have effectively the same depth. The full data reduction pipeline of both groups is being tested, and the differences could therefore have been introduced at any stage.

Fig. 10 shows the two-point correlation functions of the matched shear catalogues (using the weights of the individual catalogues), which are normally used to constrain cosmological parameters at the

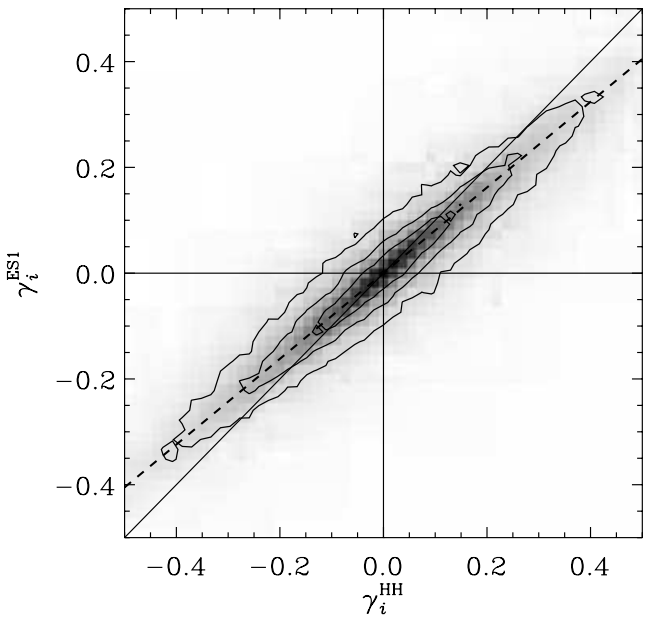

Figure 8. Comparison of shear measurement in real CFHTLS deep data, from a galaxy-by-galaxy comparison of matched catalogues from the ES1 analysis (Semboloni et al. 2006) and a reanalysis using the HH method. The relative calibration of both components of shear are indistinguishable, and are here included in the same plot. A slope of unity would imply perfect agreement. The dashed line indicates the relative calibration of the two methods in simulated image set C, which is the most closely matched to actual observing conditions. Although this should not be regarded as a strict prediction, since there are many image parameters that are not matched, its agreement with the real data is striking.

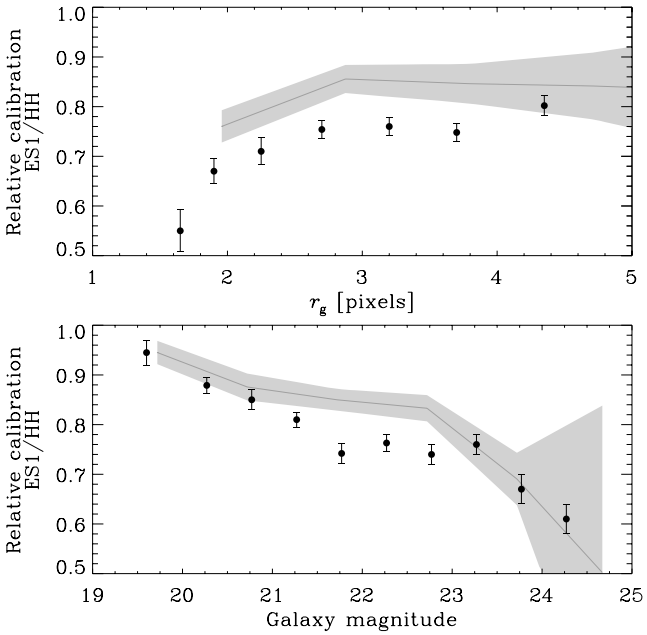

Figure 9. Comparison of shear measurement in real CFHTLS deep data, as a function of galaxy size and magnitude. The relative shear calibration of the ES1 and HH methods is obtained from the ratio of the mean shear calculated in 3×3-arcmin^2 subfields of each CFHTLS deep field. A value of unity would imply perfect agreement between the catalogues. Note that we have reconciled the different definitions of galaxy size in the simulations compared to real data by approximating $R \approx r_{\rm g}$. We have dealt with the different relationship between galaxy magnitude and S/N (cf. Section 4.5) by offsetting the magnitudes of objects in the deeper simulated data by -1. The grey band indicates the relative calibration of the two methods in simulated image set C, which is the most closely matched to the CFHTLS data.

end of a weak-lensing analysis. Although the ES1 analysis consistently measures a lower signal than the HH analysis, the discrepancy is not uniform on all scales. The relative bias is most pronounced on small scales when measuring the variance of the aperture mass statistic, and on both small and large scales for the shear variance in cells. Such variation is not seen in the galaxy-by-galaxy comparison of relative shear calibration. For example, the signal in Fig. 9 is stable to changes in the size of the area over which the shears are averaged.

We hypothesise that there may therefore be an *additional* source of bias in the ES1 CFHTLS analysis, due to PSF anisotropy residuals. Since the PSF anisotropy varies spatially, the residual would average out across the survey, and not affect the overall bias. The correlation functions were calculated using the procedure in Van Waerbeke et al. (2005), which deals with an unknown constant of integration in the calculation of $\sigma_\gamma^2(\theta)$ by forcing the B-modes of to zero on large scales. This prior on the B-modes can add spurious power to the E-modes, and could have artificially re-raised the cosmic shear signal. Indeed, the ratio of the sum of the E and B modes between analyses is flatter than that of the E modes alone. Furthermore, the star–star correlation functions (Semboloni et al. 2006) show an excess before PSF correction, on similar scales to that observed in the left-hand panel of Fig. 10.

A naïve correction for a 20 per cent shear-calibration bias in the CFHTLS deep survey (Semboloni et al. 2006) would raise the measured value of σ_8 almost proportionally. This would remain within the estimated error budget for the lensing analysis due to non-Gaussian cosmic variance (Semboloni et al. 2007), but adds tension to an existing discrepancy with the three year results from the *WMAP* (Spergel et al. 2006). In practice, a more sophisticated recalibration will probably be required. If our hypothesis of an additional systematic is correct, this would have partially cancelled the shear-calibration bias. Judging by the ratio of the observed correlation functions, the net underestimation of σ_8 could have been around 10–15 per cent. More work is needed to test this hypothesis, but it is beyond the scope of this paper. A full reanalysis of the CFHTLS survey, including the latest data, will therefore follow.

The striking confirmation of the STEP results on real data demonstrates the success of our simulation project, and highlights the vital role that artificial images will play in the exploitation of future surveys. Ideally, they ought not be relied on for simple empirical recalibration, but they will be essential to verify the performance of methods derived from first principles. The STEP images remain publicly available to test future weak-lensing analyses. Simultaneously, the complexity of our correlation function results also highlight the importance of subtleties in weak shear measurement that may arise only within the complex environment of real observational data. To fully understand such effects, we will pursue further development of the *data*STEP project[2], an ongoing comparison of the output from various shear-measurement methods on a common sample of real data.

6 CONCLUSIONS

Performance has improved since STEP1, and the STEP project continues to drive progress and innovation in shear-measurement methods. The most-accurate methods, with better than ∼2 per cent level

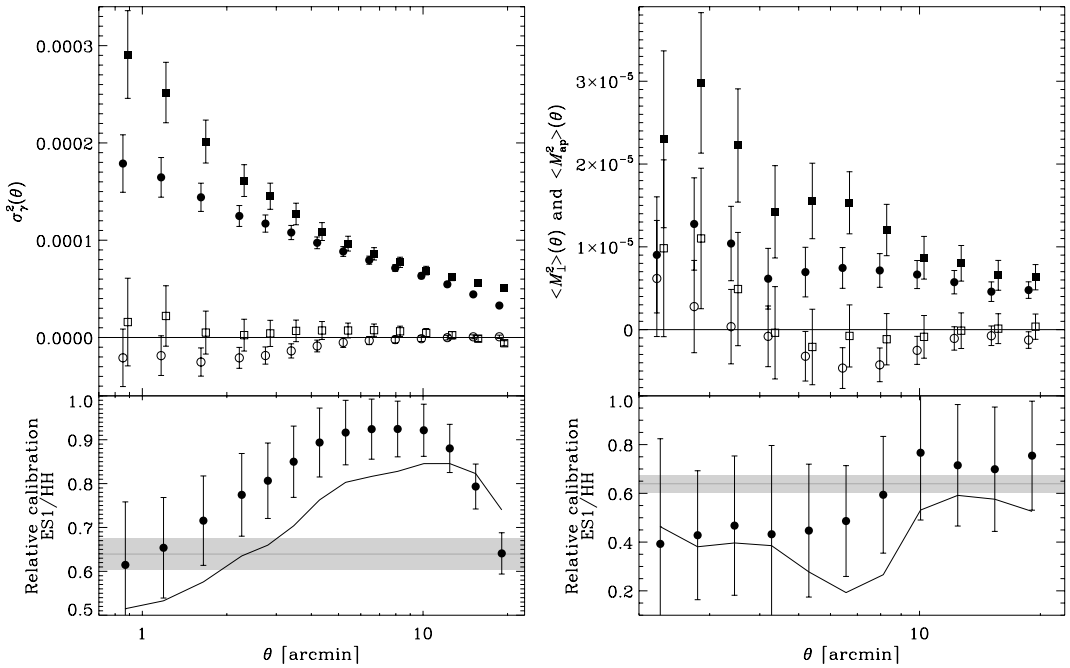

Figure 10. Comparison of shear-shear correlation functions measured from real CFHTLS deep survey data, after HH (squares) and ES1 (circles) analyses. The correlation functions are split into E and B modes in two different ways: the variance of the shear in cells is shown on the left as a function of cell radius, and the variance of the mass aperture statistic is shown on the right. In both cases, the solid points show the E mode, and the open points the B mode. The error bars show statistical errors only (i.e. no account is made for cosmic variance since the survey region is identical), but note that the difference between the two data sets is in fact more significant than indicated, because the same galaxies are used in each analysis, so noise enters only from the shape measurement process and not from variation in intrinsic galaxy ellipticities. In the lower panels, the points show the ratio of the E modes calculated from the two analyses, and the lines show the ratio of the E plus B modes. The grey bands indicate the relative calibration of the two methods in simulated image set C, which is the most closely matched to actual observing conditions.

calibration errors for most of the tested observing conditions, were the MJ2 implementation of BJ02, the TS and HH implementations of KSB+, the KK and JB implementations of shapelets and the RM implementation of Reglens. Particular advances are apparent in methods that used the results of STEP1 to tune their algorithms, which bodes well for the future of this project. For example, the introduction of a calibration factor to the TS method has proved reassuringly robust with our new, more realistic simulated images. We have also verified the STEP results on real data, finding striking confirmation of methods' relative shear calibration in the CFHTLS deep survey.

There is no one shear-measurement method that is doing everything best. With the increased precision possible in this analysis, we can now distinguish all the methods from *perfect* performance. Since absolute shear calibration cannot be directly ascertained from real data, this remains the most important issue. The calibration bias in most methods leads to a slight underestimation of shear. Both the shear-calibration (multiplicative) errors and anisotropic PSF correction (additive) errors are also found to depend on characteristics of the PSF. Technical advances in individual methods will therefore still be required. Ideally, one would attempt to take the most successful aspect of several methods and combine them. The fundamentally different approaches to the two main tasks in shear measurement make this difficult, but there is common ground (e.g. object-detection algorithms, the shapelet basis functions, and galaxy weighting schemes), so the individual lessons learned with each method may not necessarily be irreconcilable. To this end, we have developed a classification scheme for shear-measurement methods, and have described all existing methods in a common language so that their similarities and differences are apparent. Development is continuing in earnest.

We have used our improved simulations to identify various aspects of shear measurement that have been effectively solved at the current level of precision. We have also uncovered other, specific areas that remain problematic. Studying these may provide a route to the most rapid technological advances. Development needs to be focused towards

(i) pixellization;
(ii) correlated background noise;
(iii) PSF measurement; and
(iv) galaxy morphology evolution.

These four points are explained below.

This is the first STEP project in which the input shear has been applied in arbitrary directions relative to the pixel grid. That this direction affects the calibration of shear-measurement methods, even for images with a circular PSF and no other preferred direction, implies that pixellization is not fully controlled. Pixel effects may also explain the general tendency of methods to underestimate shear. Since no explicit provision is made for pixellization in many methods, this result is not surprising. This work has quantified just how much of an effect it has, and thereby emphasized the importance of a proper treatment in the future. High et al. (in preparation) are specifically investigating pixellization through tailor-made image simulations with varying pixel scales.

Although not all data sets have background noise that is significantly correlated between adjacent pixels, it is particularly apparent in natively undersampled data, for which several exposures dithered

by subpixel shifts must be co-added. The introduction of correlated background noise to the STEP2 simulations hindered several methods: during the detection of faint objects, the modelling of objects to a specified fidelity, and the weighting of individual shear estimators. Now that this issue has been raised, work is underway in the context of several of the shear-measurement methods.

Some methods seem to be having trouble with the initial measurements of the PSF from individual stars. The measurement of the shape of each star affects shear estimates from many galaxies, and is therefore of vital importance. When the PSF is highly elliptical, this work has revealed some peculiar residual shear offsets, in the directions orthogonal (at 45°) to that ellipticity. We have not yet found a satisfactory explanation for this, but speculate that it might be caused by difficulties measuring the centroid and the ellipticity of stars that have substructure, skewness, and no single, well-defined ellipticity. Methods that model the full PSF, and especially those that attempt PSF deconvolution, are less affected, but at the expense of a having smaller number density of useable galaxies for which the complicated deconvolution algorithms currently converge. We have *not* attempted to investigate the consequences of spatial PSF variation, or the errors introduced by imperfect interpolation of the PSF across a field (Massey et al. 2002; Hoekstra 2004; Jarvis & Jain 2004). Such variation poses a significant challenge in real data, the analysis of which is prone to additional systematics missing from our idealized case. However, decoupling this layer of uncertainty has simplified the interpretation of our results, allowing us to concentrate on the main issue of galaxy-shape measurement. Simultaneous investigation of shear measurement in a varying PSF will eventually require a full simulation of telescope optics and atmospheric turbulence, combined with a mock lensing analysis continued to the calculation of correlation functions, where the residual errors will become apparent.

Issues of galaxy morphology evolution become particularly important for those methods whose calibration relies on the overall distribution of galaxies' intrinsic ellipticities. High-redshift galaxies are both more elliptical and more irregular; and evolution in the ellipticity variance directly affects the shear calibration. For a 2D cosmic shear survey, even if the mean shear is correctly measured, this can bias the effective redshift distribution of source galaxies and the geometrical interpretation of the lensing signal, with all the consequences discussed in Van Waerbeke et al. (2006). For a 3D analysis, it can change the apparent redshift evolution of the signal and hence the apparent cosmological matter distribution.

The next STEP project will analyse a set of simulated space-based images. With their higher spatial resolution, we expect that variation in galaxy morphology will more profoundly affect shear measurement. We will therefore repeat the exercise of comparing the analysis of complex shapelet galaxies with more idealized objects, and also separate the galaxy populations by morphological class. The cuspy space-based PSFs will provide a different (easier) régime in which to test centering, and we will explicitly avoid PSF interpolation errors by allowing methods to assume that the PSF is constant. This should make interpretation easier. Background noise will also be left intentionally uncorrelated. However, variations in the pixel scale will be introduced, to specifically test methods' robustness to pixellization effects.

Such ongoing improvements are vital to the success of gravitational lensing as a viable probe of cosmology. Although the measurement of weak lensing is not limited by unknown physical processes, the technical aspect of galaxy shape measurement at such high precision remains computationally challenging. In this paper, we have demonstrated that simulated images can drive progress in this field, and can provide a robust test of shear measurement on real data. Previous cosmic shear measurements would have benefitted from access to STEP, and the future exploitation of dedicated surveys relies on the development of methods that are being tested here first. Both the tools and the collective will are now in place to meet this challenge. The STEP simulations remain publicly available, and the weak-lensing community is progressing to the next level of technical refinement in a spirit of open cooperation. We conclude with the hope that, by accessing the shared technical knowledge compiled by the STEP projects, all future shear-measurement methods will be able to reliably and accurately measure weak-lensing shear.

ACKNOWLEDGMENTS

Funding for the development of the shapelet image simulation pipeline and STEP telecons was provided by DoE grant #96859. Further funding for group communication was provided by CITA and NSERC. We thank the NASA Jet Propulsion Laboratory for financial and administrative support of the STEP workshop. We thank the COSMOS collaboration, particularly Anton Koekemoer and Nick Scoville, for providing the high-resolution *HST* images from which the population of galaxy morphologies was drawn. We thank Matt Ferry and Mandeep Gill for continued help in developing the shapelets image simulation pipeline. We thank CalTech ADPF staff and particularly Patrick Shopbell for help with the computing resources required to manufacture and to distribute the simulated images. CH is supported by a CITA national fellowship.

REFERENCES

Bacon D., Refregier A., Ellis R., 2000, MNRAS, 318, 625
Bacon D., Refregier A., Clowe D., Ellis R., 2001, MNRAS, 325, 1065
Bacon D., Massey R., Refregier A., Ellis R., 2003, MNRAS, 344, 673
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Bernstein G., Jarvis M., 2002, AJ, 123, 583 (BJ02)
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Berry R., Hobson M., Withington S., 2004, MNRAS, 354, 199
Bridle S., Gull S., Bardeau S., Kneib J. P., 2001, in Scientific N. W., ed., Proceedings of the Yale Cosmology Workshop
Brown M., Taylor A., Bacon D., Gray M., Dye S., Meisenheimer K., Wolf C., 2003, MNRAS, 341, 100
Crittenden R., Natarajan R., Pen U., Theuns T., 2002, ApJ, 568, 20
Dahle H., 2006, ApJ, 653, 954
Dekel A., Lahav O., 1999, ApJ, 520, 24
Erben T., Van Waerbeke L., Bertin E., Mellier Y., Schneider P., 2001, A&A, 366, 717
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
Hamana T. et al., 2003, ApJ, 597, 98
Hetterscheidt M. et al., 2006, A&A, submitted (astro-ph/0606571)
Heymans C. et al., 2005, MNRAS, 361, 160
Heymans C. et al., 2006, MNRAS, 368, 1323 (STEP1)
Hillebrandt W., Niemeyer J., 2000, ARA&A, 38, 191
Hirata C., Seljak U., 2003, MNRAS, 343, 459
Hoekstra H., 2004, MNRAS, 347, 1337
Hoekstra H., Franx M., Kuijken K., Squires G., 1998, ApJ, 504, 636
Hoekstra H., Yee H., Gladders M., 2002a, ApJ, 577, 595
Hoekstra H., van Waerbeke L., Gladders M., Mellier Y., Yee H., 2002b, ApJ, 577, 604
Hoekstra H., Yee H., Gladders M., Barrientos L. F., Hall P., Infante L., 2002c, ApJ, 572, 55
Hoekstra H. et al., 2006, ApJ, 647, 116
Huterer D., White M., 2003, ApJ, 578, L95

Ivezić Z. et al., 2004, Astronomische Nachrichten 325, 583
James B., Davis T., Schmidt B., Kim A., 2006, MNRAS, 370, 933
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., Jain B., Bernstein G., Dolney D., 2006, ApJ, 644, 71
Kaiser N., 2000, ApJ, 537, 555
Kaiser N., Squires G., Broadhurst T., 1995, ApJ, 449, 460
Kaiser N., Wilson G., Luppino G., 2000, preprint (astro-ph/0003338)
Kuijken K., 1999, A&A, 352, 355
Kuijken K., 2006, A&A, 456, 827
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., Bacon D., Refregier A., Ellis R., 2002, in Shanks T., Metcalfe N., eds, ASP Conf. Ser. Vol. 283, A New Era In Cosmology. Astron. Soc. Pac., San Francisco, p. 193
Massey R., Refregier A., Conselice C., Bacon J., 2004a, MNRAS, 348, 214
Massey R., Refregier A., Bacon D., 2004b, in Meylan M., ed., Impact of Gravitational Lensing on Cosmology
Massey R., Bacon D., Refregier A., Ellis R., 2005, MNRAS, 359, 1277
Massey R., Rowe B., Refregier A., Bacon D., Bergé J., 2007, MNRAS, submitted
Miyazaki S. et al., 2002a, PASJ, 54, 833
Miyazaki S. et al., 2002b, ApJ, 580, 97
Nakajima R., Bernstein G., 2006, AJ, submitted (astro-ph/0607062)
Pierpaoli E., Scott D., White M., 2001, MNRAS, 325, 77
Refregier A., 2003, ARA&A, 41, 645
Refregier A., Bacon D., 2003, MNRAS, 338, 48
Refregier A., Rhodes J., Groth E. J., 2002, ApJ, 572, L131
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
Schneider P., Kilbinger M., 2007, A&A, 462, 841
Schneider P., Van Waerbeke L., Mellier Y., 2002, A&A, 389, 741
Schimd C., Tereno I., Uzan J.-P. et al., 2006, A&A, in press (astro-ph/0603158)
Schrabback T. et al., 2006, A&A, submitted (astro-ph/0606611)
Scoville N. et al., 2007, ApJ, in press
Seitz C., Schneider P., 1997, A&A, 318, 687
Semboloni E. et al., 2006, A&A, 452, 51
Semboloni E., van Waerbeke L., Heymans C., Hamana T., Colombi S., White M., Mellier Y., 2007, MNRAS in press (astro-ph/0606648)
Smith D. R., Bernstein G., Fischer P., Jarvis M., 2001, ApJ, 551, 643
Smith G., Edge A., Eke V., Nichol R., Smail I., Kneib J.-P., 2003, AJ, 590, L79
Spergel D. et al., 2006, ApJ submitted (astro-ph/0603449)
Sullivan M. et al., 2006, ApJ, 648, 868
Travaglio C., Hillebrandt W., Reinecke M., 2006, A&A, 443, 1007
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
Van Waerbeke L., White M., Hoekstra H., Heymans C., 2006, Astropart. Phys., 26, 91
Viana P., Nichol R., Liddle A., 2002, ApJ, 569, 75
White M., Kochanek C., 2001, ApJ, 560, 539
Wittman D., 2002, in Courbin F., Minniti D., eds, Dark Matter and Gravitational Lensing. LNP Top. Vol., Springer-Verlag, Berlin
Wittman D., Tyson J., Kirkman D., Dell'Antonio I., Bernstein G., 2000, Nat, 405, 143
Wittman D., 2005, ApJ, 632, 5
Weinberg D., Davé R., Katz N., Hernquist L., 2004, ApJ, 601, 1

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