Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **429,** 661–678 (2013) doi:10.1093/mnras/sts371

Origins of weak lensing systematics, and requirements on future instrumentation (or knowledge of instrumentation)

Richard Massey,[1⋆] Henk Hoekstra,[2] Thomas Kitching,[3] Jason Rhodes,[4,5] Mark Cropper,[6] Jérôme Amiaux,[7] David Harvey,[3] Yannick Mellier,[7,8] Massimo Meneghetti,[9] Lance Miller,[10] Stéphane Paulin-Henriksson,[7] Sandrine Pires,[7] Roberto Scaramella[11] and Tim Schrabback[12,13]

[1]*Institute for Computational Cosmology, Durham University, South Road, Durham DH1 3LE*
[2]*Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands*
[3]*University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ*
[4]*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA*
[5]*California Institute of Technology, 1201 E California Blvd., Pasadena, CA 91125, USA*
[6]*Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT*
[7]*Service d'Astrophysique, CEA Saclay, F-91191 Gif sur Yvette, France*
[8]*Institut d'Astrophysique de Paris, UMR7095 CNRS, Université Pierre et Marie Curie, 98 bis Boulevard Arago, F-75014 Paris, France*
[9]*INAF, Osservatorio Astronomico di Bologna, via Ranzani 1, I-40127 Bologna, Italy*
[10]*Department of Physics, University of Oxford, The Denys Wilkinson Building, Keble Road, Oxford OX1 3RH*
[11]*INAF, Osservatorio Astronomico di Roma, via Frascati 33, I-00040 Monteporzio Catone, Italy*
[12]*Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA*
[13]*Argelander-Institut für Astronomie, Auf dem Hügel 71, D-53121 Bonn, Germany*

Accepted 2012 November 5. Received 2012 October 25; in original form 2012 September 20

ABSTRACT

The first half of this paper explores the origin of systematic biases in the measurement of weak gravitational lensing. Compared to previous work, we expand the investigation of point spread function instability and fold in for the first time the effects of non-idealities in electronic imaging detectors and imperfect galaxy shape measurement algorithms. Together, these now explain the additive $\mathcal{A}(\ell)$ and multiplicative $\mathcal{M}(\ell)$ systematics typically reported in current lensing measurements. We find that overall performance is driven by a product of a telescope/camera's *absolute performance*, and our *knowledge about its performance*.

The second half of this paper propagates any residual shear measurement biases through to their effect on cosmological parameter constraints. Fully exploiting the statistical power of Stage IV weak lensing surveys will require additive biases $\overline{\mathcal{A}} \lesssim 1.8 \times 10^{-12}$ and multiplicative biases $\overline{\mathcal{M}} \lesssim 4.0 \times 10^{-3}$. These can be allocated between individual budgets in hardware, calibration data and software, using results from the first half of the paper.

If instrumentation is stable and well calibrated, we find extant shear measurement software from Gravitational Lensing Accuracy Testing 2010 (GREAT10) already meet requirements on galaxies detected at signal-to-noise ratio $= 40$. Averaging over a population of galaxies with a realistic distribution of sizes, it also meets requirements for a 2D cosmic shear analysis from space. If used on fainter galaxies or for 3D cosmic shear tomography, existing algorithms would need calibration on simulations to avoid introducing bias at a level similar to the statistical error. Requirements on hardware and calibration data are discussed in more detail in a companion paper. Our analysis is intentionally general, but is specifically being used to drive the hardware and ground segment performance budget for the design of the European Space Agency's recently selected *Euclid* mission.

Key words: gravitational lensing: weak – instrumentation: detectors – methods: data analysis – space vehicles: instruments – telescopes – cosmological parameters.

⋆E-mail: r.j.massey@durham.ac.uk

1 INTRODUCTION

Statistical measurements of weak gravitational lensing in a large sample of galaxies offer a direct way to probe the dark sector of the Universe (see reviews by Hoekstra & Jain 2008; Massey, Kitching & Richard 2010b). Gravitational lensing is the deflection of light from distant galaxies during its journey to us, by an amount that depends on the intervening distribution of matter (including dark matter) and the geometry of space–time (which is currently governed by dark energy). The deflection of light produces slight shear distortions in the galaxies' apparent shapes, and adjacent galaxies appear to line up in characteristic patterns across the sky.

Galaxy ellipticities are typically distorted only a few per cent by weak gravitational lensing. Detecting this tiny signal is difficult because the image shapes are also changed an order of magnitude more by convolution with the point spread function (PSF) of the telescope, detector and atmosphere, as well as by distortion in the camera. These other effects must be modelled and corrected; even subtle residual contributions can significantly bias cosmological measurements.

In the first half of this paper we explore three types of error that affect galaxy shape measurement.

(i) Inaccuracies in the model of the convolutional PSF, from which observed galaxy shapes must be deconvolved (this builds upon work by Paulin-Henriksson et al. 2008).

(ii) Inaccuracies in correction for any effect that cannot be treated as a deconvolution. This includes detector effects such as charge transfer inefficiency in CCDs or interpixel capacitance in HgCdTe devices, which perturb pixel values in a non-linear fashion.

(iii) Inaccuracies in the measurement of galaxy shapes. Minimizing noise, particularly in faint galaxies, forces measurement methods to apply pixel weights which must subsequently be undone.

We propagate these measurement errors through a tomographic cosmic shear analysis (theory developed by Hu 1999; Jain & Taylor 2003; Bernstein & Jain 2004 and measurements obtained by Kitching et al. 2007; Massey et al. 2007a; Schrabback et al. 2010) to determine the bias they induce upon constraints on the dark energy equation of state parameter w (Song & Knox 2004; Ishak 2005; Simpson & Bridle 2005).

In the second half of this paper, we establish requirements on additive and multiplicative cosmic shear systematics to meet future scientific goals. We also use our earlier results to consider how residual additive biases can be empirically identified and removed, and assess the impact of residual multiplicative biases that cannot be self-consistently identified within a data set. Our analysis is intentionally performed with a scope sufficiently general to cover any future Stage IV weak gravitational lensing survey. It is particularly motivated by, and drives the hardware and ground segment performance budget for the design of the European Space Agency's recently selected *Euclid* mission (Laureijs et al. 2011). This work generalizes the conclusions of Amara, Réfrégier & Paulin-Henriksson (2010). During the final preparation of this paper, Chang et al. (2012) posted to the arXiv an analysis of future prospects for the Large Synoptic Survey Telescope (LSST). There is some overlap in ambition, but complementary methodology. Like Chang et al. we employ a bottom-up approach in this paper, propagating various instrumental imperfections through to errors on cosmological parameters. However, rather than simulating the detailed performance of a baseline telescope model, we work analytically to build a general framework for propagating general system performance. In a companion paper, Cropper et al. (2012), this allows us to perform a top-down, systems engineering analysis: starting from the science requirements and flowing down to requirements on subsystem performances. Using the understanding from this paper, the total error budget and mitigation can be sensibly allocated between individual budgets in hardware, calibration data and software performance.

This paper is organized as follows. In Section 2, we define the basic galaxy shape and cosmological quantities of interest that would be measured in a weak lensing experiment with no (or idealized) errors. In Section 3, we explore the various types of error that can be introduced during realistic galaxy shape measurement and may prevent recovery of the true signal. Our underlying approach builds upon the work of Paulin-Henriksson et al. (2008) – however the mathematical expressions rapidly lengthen when we introduce more sources of error. For clarity, we therefore choose to evolve the formalism in three stages, one for each source of error. In Section 4, we derive requirements on shear measurement biases for a cosmic shear survey seeking to measure dark energy. In Section 5, we determine whether those requirements are met by extant shear measurement software described in the literature. We do this at fixed galaxy fluxes and, using our results from the first half of the paper, averaging over the full population of galaxies that will be seen by a survey. We conclude in Section 6.

2 IDEALIZED WEAK LENSING MEASUREMENT

2.1 Perfect shear measurement

Many techniques have been developed to precisely measure the shapes of galaxies (see Bridle et al. 2010; Kitching et al. 2011b). For the sake of a concrete example, we shall consider the generic class of methods based upon galaxies' quadrupole moments. In a method based on unweighted quadrupole moments (see Bartelmann & Schneider 2001), the shape of any localized object in a 2D image $I(r, \theta)$ can be quantified via its size,

$$R^2 \equiv \frac{\iint I(r,\theta)\, r^2\, r\, \mathrm{d}r\, \mathrm{d}\theta}{\iint I(r,\theta)\, r\, \mathrm{d}r\, \mathrm{d}\theta}, \tag{1}$$

and complex ellipticity,

$$\boldsymbol{\epsilon} = \epsilon_1 + \mathrm{i}\epsilon_2 \equiv \frac{\iint I\, r^2 \mathrm{e}^{2\mathrm{i}\theta}\, r\, \mathrm{d}r\, \mathrm{d}\theta}{\iint I\, r^2\, r\, \mathrm{d}r\, \mathrm{d}\theta}. \tag{2}$$

Gravitational lensing magnifies and shears a galaxy of intrinsic size $R_{\rm int}$ and ellipticity $\boldsymbol{\epsilon}_{\rm int}$ into one of size $R_{\rm gal} > 0$ and ellipticity:

$$\boldsymbol{\epsilon}_{\rm gal} = \boldsymbol{\epsilon}_{\rm int} + P_\gamma \boldsymbol{\gamma}, \tag{3}$$

where the shear 'polarizability' $P_\gamma \equiv \partial\boldsymbol{\epsilon}_{\rm int}/\partial\boldsymbol{\gamma} \approx 1.86$ is the amount by which the ellipticity of a galaxy changes during gravitational lensing[1] (Kaiser, Squires & Broadhurst 1995; Luppino & Kaiser 1997). When this galaxy is imaged by any camera, it is

[1] Strictly, P_γ depends on galaxy morphology and is also a 2×2 tensor acting separately on the real and imaginary components of ellipticity. On average, however, it is very close to the identity tensor times a real scalar $2 - \langle|\boldsymbol{\epsilon}_{\rm int}|^2\rangle$ (Rhodes, Réfrégier & Groth 2000). Leauthaud et al. (2007) show that $\langle|\boldsymbol{\epsilon}_{\rm int}|^2\rangle$ is consistent with a constant value of 2×0.26^2 for galaxies to at least redshift $z = 2.6$. We greatly simplify subsequent analysis by assuming scalar $P_\gamma \approx 1.86$.

convolved with a PSF (of size $R_{\rm PSF}$ and ellipticity $\boldsymbol{\varepsilon}_{\rm PSF}$), producing[2] an observed source of larger size,

$$R^2_{\rm obs} = R^2_{\rm gal} + R^2_{\rm PSF}, \tag{4}$$

and perturbed ellipticity,

$$\boldsymbol{\varepsilon}_{\rm obs} = \boldsymbol{\varepsilon}_{\rm gal} + \frac{R^2_{\rm PSF}}{R^2_{\rm gal} + R^2_{\rm PSF}} \left(\boldsymbol{\varepsilon}_{\rm PSF} - \boldsymbol{\varepsilon}_{\rm gal}\right). \tag{5}$$

Weak lensing analyses observe the shape of each galaxy then try to correct it for (or deconvolve it from) the PSF, to recover the galaxy's true ellipticity. The system PSF can be measured from stars also within the field of view. Rearranging equations (4) and (5) so that only observable quantities appear on the right-hand side, the galaxy's ellipticity is

$$\boldsymbol{\varepsilon}_{\rm gal} = \frac{\boldsymbol{\varepsilon}_{\rm obs} R^2_{\rm obs} - \boldsymbol{\varepsilon}_{\rm PSF} R^2_{\rm PSF}}{R^2_{\rm obs} - R^2_{\rm PSF}}. \tag{6}$$

The intrinsic ellipticity of individual galaxies is uninteresting so, to isolate the cosmologically relevant information, ellipticity is normalized into a shear estimator:

$$\widehat{\boldsymbol{\gamma}} \equiv (P_\gamma)^{-1}\, \boldsymbol{\varepsilon}_{\rm gal}. \tag{7}$$

This ensures that, averaging over a large number of galaxies,

$$\langle\widehat{\boldsymbol{\gamma}}\rangle = (P_\gamma)^{-1}\langle\boldsymbol{\varepsilon}_{\rm int}\rangle + (P_\gamma)^{-1} P_\gamma\, \langle\boldsymbol{\gamma}\rangle, \tag{8}$$

and we recover $\langle\widehat{\boldsymbol{\gamma}}\rangle = \langle\boldsymbol{\gamma}\rangle$ so long as the intrinsic galaxy ellipticities are random and hence $\langle\boldsymbol{\varepsilon}_{\rm int}\rangle = 0$ (but see Catelan, Kamionkowski & Blandford 2001; Crittenden et al. 2001; Natarajan et al. 2001; Joachimi & Schneider 2008; Schneider & Bridle 2010; Kirk et al. 2012, for instances of 'intrinsic alignments' when this does not hold).

The average shear is zero so, to compare to theoretical models, the measured shears are then combined into two-point correlation functions:

$$\xi_+(\theta, z_A, z_B) \equiv \left\langle \boldsymbol{\gamma}_A \boldsymbol{\gamma}^*_B \right\rangle (\theta, z_A, z_B), \tag{9}$$

$$\xi_-(\theta, z_A, z_B) \equiv {\rm Re}(\langle \boldsymbol{\gamma}_A \boldsymbol{\gamma}_B\rangle(\theta, z_A, z_B)), \tag{10}$$

where the angle brackets indicate averaging over all pairs of galaxies A and B in a survey that are at redshifts z_A and z_B and separated on the sky by an angle θ, or within bins around those values (Crittenden et al. 2001; Bartelmann & Schneider 2001). The correlation functions trace a cosmological, 'cosmic shear' signal at $\theta > 0$. Results are often expressed in terms of the shear power spectrum $C(\ell)$, the Fourier transform of a weighted sum of $\xi_\pm(\theta)$.

If galaxy shapes are autocorrelated with themselves, a zero-lag term $\sigma^2_{\boldsymbol{\gamma}}\delta(\theta = 0)$ is added. This is included by Paulin-Henriksson et al. (2008, equation 11) but we disregard it because it can be readily avoided by excluding such galaxy pairs in practice. If the $\sigma^2_{\boldsymbol{\gamma}}$ autocorrelation term were not removed from an analysis, it would be white noise, independent of scale in the Fourier transform. This must be marginalized over as an unknown constant of integration, subtracted from measurements or added to theoretical models.

[2] Relationships (4) and (5) are exact using unweighted moments, but hold for some other methods only if both the galaxy and the PSF are approximately Gaussian. We shall return to this issue in Section 3.3.

2.2 Parametric shear measurement bias

Deviations from perfect shear measurement are commonly parametrized following Huterer et al. (2006) and the Shear TEsting Programme (STEP; Heymans et al. 2006; Massey et al. 2007b) as

$$\widehat{\boldsymbol{\gamma}} = (1 + m)\boldsymbol{\gamma} + \boldsymbol{c}. \tag{11}$$

We shall henceforth represent all real-world, imperfect measurements using a hat.

2.2.1 *Constant shear measurement bias*

We first consider shear measurements that have small additive bias $\boldsymbol{c}$ with constant mean $\langle\boldsymbol{c}\rangle$ and random noise σ_c, plus small multiplicative bias m with constant mean $\langle m\rangle$ and random noise σ_m. Pairs of these shear measurements can be folded through the calculation of a correlation function (9) to produce

$$\widehat{\xi_+}(\theta, z_A, z_B) \equiv \left\langle \widehat{\boldsymbol{\gamma}}_A \widehat{\boldsymbol{\gamma}}^*_B \right\rangle \tag{12}$$

$$= \langle(1 + m)(1 + m)\rangle\, \xi_+ + \langle|\boldsymbol{c}|^2\rangle \tag{13}$$

plus cross terms only in the presence of shear-dependent selection effects (see e.g. Jain, Jarvis & Bernstein 2006, and the discussion in Appendix A).

Taking the Fourier transform to yields a power spectrum:[3]

$$\widehat{C}(\ell, z_A, z_B) = (1 + \mathcal{M})\, C(\ell, z_A, z_B) + \mathcal{A}, \tag{14}$$

where

$$\mathcal{A} = 0, \tag{15}$$

$$\mathcal{M} = 2\langle m\rangle + \langle m^2\rangle \;=\; 2\langle m\rangle + \langle m\rangle^2 + \sigma^2_m\, \delta(0) \tag{16}$$

$$\approx 2\langle m\rangle. \tag{17}$$

We have expanded the mean squared error $\langle m^2\rangle$ term but note that $\langle m\rangle^2 \ll \langle m\rangle$ and that $\mathcal{M}$ terms arise only from the correlation of galaxies with other galaxies. Thus a constant multiplicative bias in shear measurements leads to a similarly constant multiplicative bias in a measurement of the shear power spectrum. If the autocorrelation terms discussed in Section 2.1 are included, equation (15) gains an additional white noise term,

$$\sigma^2_{\boldsymbol{\gamma}} = [1 + 2\langle m\rangle + \langle m^2\rangle]{P_\gamma}^{-2}\sigma^2_{\boldsymbol{\varepsilon}_{\rm gal}} + \sigma^2_{\boldsymbol{c}}. \tag{18}$$

This notation is also discussed in Kitching et al. (2012a).

2.2.2 *Spatially/temporally varying shear measurement bias*

We next consider shear measurement biases in which $\boldsymbol{c}$ and m vary from galaxy to galaxy, and their deviations from a mean value can

[3] Amara & Réfrégier (2008, equation 13) rearrange equation (12) as a Taylor series expansion of the measured correlation function,

$$\widehat{C}(\ell) \equiv C(\ell) + \{\mathcal{A}_0 + \mathcal{A}_1\, C(\ell) + \cdots\},$$

and label everything inside the curly brackets as different types of 'additive error' $C^{\rm sys}_\ell$. Simple multiplicative biases easily arise, so we instead find it helpful to keep terms $\mathcal{A}$ and $\mathcal{M}$ separate. Because of the shape of the Λ cold dark matter (ΛCDM) cosmological power spectrum, they are nearly orthogonal and have quite different implications (see Section 4). We therefore restrict our notation for $\mathcal{A}$ to refer solely to pure additive terms.

be correlated in patterns across a survey. In general, $\mathcal{A}$ and $\mathcal{M}$ become functions of scale, orientation on the sky, and redshift:

$$\widehat{C}(\boldsymbol{\ell}, z_A, z_B) = \sum_{\boldsymbol{\ell}'}(1 + \mathcal{M}(\boldsymbol{\ell}, \boldsymbol{\ell}', z_A, z_B))C(\boldsymbol{\ell}', z_A, z_B) + \mathcal{A}(\boldsymbol{\ell}, z_A, z_B) \quad (19)$$

(see Appendix A of Kitching et al. 2012a). The additive systematics $\mathcal{A}$ now include a contribution from the spatially varying additive shear measurement bias. The $\mathcal{M}$ matrices mix power from different scales, as well as physical E-mode and non-physical B-mode signals, where $C = C_E + \mathrm{i}C_B$. Anisotropic errors could arise from PSF terms in off-axis cameras, from non-square pixels, from some detector effects, or in ground-based surveys where gravity loading and the prevailing wind can impose preferred directions. In this paper, we shall only consider the simpler situation in which the systematic errors are isotropic on average within a survey. In this case, the matrices are diagonal, so $\mathcal{A}$ and $\mathcal{M}$ become functions of only scale and redshift:

$$\widehat{C}(\ell, z_A, z_B) = (1 + \mathcal{M}(\ell, z_A, z_B))C(\ell, z_A, z_B) + \mathcal{A}(\ell, z_A, z_B). \quad (20)$$

Using the notation $\sigma^2[x]$ to represent the covariance about the mean of error δx in pairs of galaxies separated by $\theta > 0$, we find

$$\mathcal{A}(\ell, z_A, z_B) = \sigma^2[|\boldsymbol{c}|](\ell, z_A, z_B), \quad (21)$$

$$\mathcal{M}(\ell, z_A, z_B) \approx \sigma^2[m](\ell, z_A, z_B) + 2\langle m\rangle(z_A, z_B). \quad (22)$$

Paulin-Henriksson et al. (2008) miss the second half of equation (22) because they ignore bias terms when expanding mean squared errors. This was reasonable for purely additive systematics, as spatially constant terms disappear during a Fourier transform; but in this case we judge that the bias term is likely to be the most problematic. Instead, we find that cosmic shear biases arise from a combination of (a) absolute biases in shear measurement and (b) uncertainty in or lack of knowledge about shear measurements. This dichotomy will emerge as a general result throughout Section 3.

3 REALISTIC WEAK LENSING MEASUREMENT ERRORS

3.1 Imperfect PSF correction

Errors in shear measurement can arise from several sources. For example, our model of the PSF will inevitably be imperfect because it is obtained from noisy stars and must be interpolated to the position and colour of each galaxy (e.g. Hoekstra 2004; Mandelbaum et al. 2005; Massey et al. 2005; Jain et al. 2006; Paulin-Henriksson, Réfrégier & Amara 2009; Cypriano et al. 2010).

Via a first-order Taylor series expansion of equation (6), model errors in the PSF size $\delta(R^2_{\rm PSF})$ and ellipticity $\delta\boldsymbol{\epsilon}_{\rm PSF}$ propagate into an imperfect estimate of the galaxy ellipticity:

$$\widehat{\boldsymbol{\epsilon}_{\rm gal}} \approx \boldsymbol{\epsilon}_{\rm gal} + \frac{\partial\boldsymbol{\epsilon}_{\rm gal}}{\partial\left(R^2_{\rm PSF}\right)}\delta\left(R^2_{\rm PSF}\right) + \frac{\partial\boldsymbol{\epsilon}_{\rm gal}}{\partial\boldsymbol{\epsilon}_{\rm PSF}}\delta\boldsymbol{\epsilon}_{\rm PSF}. \quad (23)$$

The partial derivatives of (6) are

$$\frac{\partial\boldsymbol{\epsilon}_{\rm gal}}{\partial\left(R^2_{\rm PSF}\right)} = \frac{R^2_{\rm obs}}{\left(R^2_{\rm obs} - R^2_{\rm PSF}\right)^2}(\boldsymbol{\epsilon}_{\rm obs} - \boldsymbol{\epsilon}_{\rm PSF}) = \frac{\boldsymbol{\epsilon}_{\rm gal} - \boldsymbol{\epsilon}_{\rm PSF}}{R^2_{\rm gal}}, \quad (24)$$

$$\frac{\partial\boldsymbol{\epsilon}_{\rm gal}}{\partial\boldsymbol{\epsilon}_{\rm PSF}} = -\frac{R^2_{\rm PSF}}{R^2_{\rm obs} - R^2_{\rm PSF}} = -\frac{R^2_{\rm PSF}}{R^2_{\rm gal}}, \quad (25)$$

and the derivative with respect to the other real/imaginary component of the PSF ellipticity is zero. In equations (24) and (25), the first equality is expressed in terms that are observable in an image, and the second equality reflects fundamental source properties. Inserting the latter into (23) yields

$$\widehat{\boldsymbol{\epsilon}_{\rm gal}} \approx \left\{1 + \frac{\delta\left(R^2_{\rm PSF}\right)}{R^2_{\rm gal}}\right\}\boldsymbol{\epsilon}_{\rm gal} - \left\{\frac{R^2_{\rm PSF}}{R^2_{\rm gal}}\delta\boldsymbol{\epsilon}_{\rm PSF} + \frac{\delta\left(R^2_{\rm PSF}\right)}{R^2_{\rm gal}}\boldsymbol{\epsilon}_{\rm PSF}\right\}. \quad (26)$$

Arranged thus (cf. Paulin-Henriksson et al. 2008, equation 8), the last two terms display an elegant symmetry: the product of the PSF size and our knowledge of its ellipticity, then its absolute ellipticity and our knowledge of its size. The STEP parameters can be easily read off from this expression. Note that if the PSF ellipticity is known perfectly ($\delta\boldsymbol{\epsilon}_{\rm PSF} = \mathbf{0}$), $\boldsymbol{c} = -m\boldsymbol{\epsilon}_{\rm PSF}/P_\gamma$ and the two are related.

When folding this imperfect shear estimator through the calculation of a correlation function (9), to multiply out some angle brackets we follow Paulin-Henriksson et al. (2008) in assuming that inaccuracies in the model of the PSF shape are independent of the shape of the PSF and the size of galaxies to which it is applied. If that does not hold, the angle brackets cannot be separated and some cross-terms can be introduced that are computed in Appendix A. An additional assumption that Paulin-Henriksson et al. (2008) and we make is that the size of the PSF is roughly constant across the survey, such that $\left\langle R^2_{\rm PSF}R^2_{\rm PSF}\right\rangle(\ell) \approx \left\langle R^4_{\rm PSF}\right\rangle$. In exact correspondence to the various terms of equations (21) and (22), we find

$$\mathcal{A}(\ell, z_A, z_B) = \frac{1}{{P_\gamma}^2}\left\langle\frac{R^4_{\rm PSF}}{R^4_{\rm gal}}\right\rangle\left\langle|\delta\boldsymbol{\epsilon}_{\rm PSF}|^2\right\rangle(\ell, z_A, z_B) + \frac{\left\langle|\boldsymbol{\epsilon}_{\rm PSF}|^2\right\rangle}{{P_\gamma}^2}\left\langle\frac{R^4_{\rm PSF}}{R^4_{\rm gal}}\right\rangle\frac{\left\langle\left|\delta\left(R^2_{\rm PSF}\right)\right|^2\right\rangle}{\left\langle R^4_{\rm PSF}\right\rangle}(\ell, z_A, z_B), \quad (27)$$

$$\mathcal{M}(\ell, z_A, z_B) = \left\langle\frac{R^4_{\rm PSF}}{R^4_{\rm gal}}\right\rangle\frac{\left\langle\left|\delta\left(R^2_{\rm PSF}\right)\right|^2\right\rangle}{\left\langle R^4_{\rm PSF}\right\rangle}(\ell, z_A, z_B) + 2\left\langle\frac{R^2_{\rm PSF}}{R^2_{\rm gal}}\right\rangle\frac{\left\langle\delta\left(R^2_{\rm PSF}\right)\right\rangle}{\left\langle R^2_{\rm PSF}\right\rangle}(z_A, z_B). \quad (28)$$

In combination, this reproduces equation (11) of Paulin-Henriksson et al. (2008), except for the autocorrelation term now intentionally omitted from (27) (see Section 2.1) and the linear term now appended to (28) (see Section 2.2.2).

We shall expand the (scale-dependent) mean squared error terms that reflect a measurement error, like $\langle|\delta\boldsymbol{\epsilon}_{\rm PSF}|^2\rangle$, into a bias $\langle|\delta\boldsymbol{\epsilon}_{\rm PSF}|\rangle^2$, plus a covariance about the mean $\sigma^2[|\boldsymbol{\epsilon}_{\rm PSF}|]$. For the sake of legibility, we do not likewise expand the mean squared error terms on instrument performance, such as $\langle|\boldsymbol{\epsilon}_{\rm PSF}|^2\rangle$, but the split is implicit. For legibility, we also omit the notation showing functional dependence on scale and redshift, but note that all bias terms are functions of (z_A, z_B) and all covariances are functions of (ℓ, z_A, z_B). Indeed, since $\langle R_{\rm gal}\rangle$ scales with redshift, every term really will vary

as a function of redshift. To second order in δ, we find

$$\mathcal{A} = \frac{1}{{P_\gamma}^2}\left\langle\frac{R^4_{\rm PSF}}{R^4_{\rm gal}}\right\rangle\sigma^2[|\boldsymbol{\varepsilon}_{\rm PSF}|] + \frac{\langle|\boldsymbol{\varepsilon}_{\rm PSF}|^2\rangle}{{P_\gamma}^2}\left\langle\frac{R^4_{\rm PSF}}{R^4_{\rm gal}}\right\rangle\left(\frac{\left\langle\delta\left(R^2_{\rm PSF}\right)\right\rangle^2}{\left\langle R^4_{\rm PSF}\right\rangle} + \frac{\sigma^2\left[R^2_{\rm PSF}\right]}{R^4_{\rm PSF}}\right), \quad (29)$$

where the spatially constant term in the first line disappears as a delta function at $\ell = 1$ (or the fundamental mode of the survey) in Fourier space; a similar cross term involving the (implicit) bias on $\boldsymbol{\varepsilon}_{\rm PSF}$ is also zero in the second line. Note that all ellipticities have two components that add in quadrature. Ignoring a bias term in $\mathcal{M}$ proportional to the square of one already present (therefore negligible if the bias is small), we also find

$$\mathcal{M} = 2\left\langle\frac{R^2_{\rm PSF}}{R^2_{\rm gal}}\right\rangle\frac{\left\langle\delta\left(R^2_{\rm PSF}\right)\right\rangle}{\left\langle R^2_{\rm PSF}\right\rangle} + \left\langle\frac{R^4_{\rm PSF}}{R^4_{\rm gal}}\right\rangle\frac{\sigma^2\left[R^2_{\rm PSF}\right]}{\left\langle R^4_{\rm PSF}\right\rangle}. \quad (30)$$

We shall explore concrete values for the terms in equations (29) and (30) in Section 4.5. For now, notice how the systematics are driven mainly by the size of the PSF – to the fourth power, which is why cosmic shear measurements are generally easier from above the Earth's atmosphere. However, $\delta\left(R^2_{\rm PSF}\right)$ terms (proportional to only the second power) arise if the PSF is wavelength dependent and measured from stars that are a different colour to galaxies (Cypriano et al. 2010). This effect is worse for diffraction-limited space-based observations than ground-based imaging, where the PSF is determined primarily by atmospheric turbulence. It would likely be spatially constant (and therefore disappear from $\mathcal{A}$ at least), except that chromatic aberration may exacerbate it on a characteristic scale related to the size of a telescope's field of view (Plazas & Bernstein 2012). It is anyway a function of redshift.

Equation (29) in particular shows that *overall performance is driven by the product of instrument stability and knowledge* about that instrument. This quantifies the tradeoffs discussed by Amara et al. (2010). To obtain reliable cosmological measurements, we first need high-quality instrumentation to deliver a system PSF that is

(i) small ($R_{\rm PSF}$),
(ii) nearly circular (the bias component of $\langle|\boldsymbol{\varepsilon}_{\rm PSF}|^2\rangle$) and
(iii) stable (the variance component of $\langle|\boldsymbol{\varepsilon}_{\rm PSF}|^2\rangle$; we have already assumed that its size is constant).

It is then equally important to

(i) understand and accurately model that PSF.

The $\langle\delta\rangle$ terms reflect a calibration bias in the PSF model (e.g. in its colour), and are likely to spatially constant. The $\sigma^2[\,]$ terms reflect a lack of knowledge (e.g. from sparse sampling of a spatially/temporally varying PSF pattern), and are likely to vary as a function of scale in such a way that they are largest around the mean distance between stars, the size of the telescope's field of view or (reflecting the intrinsic variation in the PSF pattern) turbulence cells in the atmosphere.

3.2 Imperfect correction for detector effects

As well as convolution with a PSF (which in practice can include all optical and electronic effects that act linearly on pixel values), astronomical images can also be degraded in more complicated ways. This can include global detector non-linearity, in which the number of counts in each pixel is a non-linear function of the incident flux, or non-local effects such as charge transfer inefficiency in CCDs (Janesick 2001; Bristow 2003) and interpixel capacitance or persistence in HgCdTe devices (Barron et al. 2007; McCullough 2008; Seshadri et al. 2008).

These operations cannot be treated mathematically as a convolution, so the correction procedure outlined in Section 3.1 does not apply. We therefore introduce a new category of non-convolutive (NC) perturbations in galaxy size $R_{\rm NC}$ and ellipticity $\boldsymbol{\varepsilon}_{\rm NC}$. The details of these may depend on the flux and size of the galaxy, but we take a generic approach (which can hold for small, faint galaxies) in which the observed quantities become

$$R_{\rm obs} \equiv \sqrt{\left(R^2_{\rm gal} + R^2_{\rm PSF}\right)} + R_{\rm NC} \quad (31)$$

and

$$\boldsymbol{\varepsilon}_{\rm obs} \equiv \boldsymbol{\varepsilon}_{\rm gal} + \frac{R^2_{\rm PSF}}{R^2_{\rm gal} + R^2_{\rm PSF}}(\boldsymbol{\varepsilon}_{\rm PSF} - \boldsymbol{\varepsilon}_{\rm gal}) + \boldsymbol{\varepsilon}_{\rm NC}. \quad (32)$$

Note that we have not explicitly included non-convolution effects on stellar images from which the PSF is modelled. The images of bright stars will also be degraded, and the budgets for $\delta R_{\rm PSF}$ and $\boldsymbol{\delta\varepsilon}_{\rm PSF}$ should allow for this. However, many of the most serious non-linear effects operate in the sense that the degradation of bright sources is much less than that of faint sources (Massey et al. 2010a; Hoekstra et al. 2011). In this case, the perturbations on galaxies $R_{\rm NC}$ and $\boldsymbol{\varepsilon}_{\rm NC}$ will dominate, in the budget for galaxy shape measurement rather than the (separable) budget for PSF modelling. A weak lensing analysis then seeks to recover

$$\boldsymbol{\varepsilon}_{\rm gal} = \left(1 + \frac{R^2_{\rm PSF}}{R^2_{\rm gal}}\right)(\boldsymbol{\varepsilon}_{\rm obs} - \boldsymbol{\varepsilon}_{\rm NC}) - \frac{R^2_{\rm PSF}}{R^2_{\rm gal}}\boldsymbol{\varepsilon}_{\rm PSF} \quad (33)$$

$$= \frac{(\boldsymbol{\varepsilon}_{\rm obs} - \boldsymbol{\varepsilon}_{\rm NC})(R_{\rm obs} - R_{\rm NC})^2 - \boldsymbol{\varepsilon}_{\rm PSF}R^2_{\rm PSF}}{(R_{\rm obs} - R_{\rm NC})^2 - R^2_{\rm PSF}}, \quad (34)$$

where we have taken care on the second line to include only observable quantities on the right-hand side.

In practice, any correction scheme will inevitably have inaccuracies $\delta R_{\rm NC}$ and $\boldsymbol{\delta\varepsilon}_{\rm NC}$, so only an imperfect estimation is possible of $\boldsymbol{\varepsilon}_{\rm gal}$. Again we expand the shape observables as a first-order Taylor series:

$$\widehat{\boldsymbol{\varepsilon}_{\rm gal}} \approx \boldsymbol{\varepsilon}_{\rm gal} + \frac{\partial\boldsymbol{\varepsilon}_{\rm gal}}{\partial\left(R^2_{\rm PSF}\right)}\delta\left(R^2_{\rm PSF}\right) + \frac{\partial\boldsymbol{\varepsilon}_{\rm gal}}{\partial\boldsymbol{\varepsilon}_{\rm PSF}}\boldsymbol{\delta\varepsilon}_{\rm PSF} + \frac{\partial\boldsymbol{\varepsilon}_{\rm gal}}{\partial(R_{\rm NC})}\delta(R_{\rm NC}) + \frac{\partial\boldsymbol{\varepsilon}_{\rm gal}}{\partial\boldsymbol{\varepsilon}_{\rm NC}}\boldsymbol{\delta\varepsilon}_{\rm NC}, \quad (35)$$

where the first two partial derivatives remain unchanged as (the second form of) equations (24) and (25), and

$$\frac{\partial\boldsymbol{\varepsilon}_{\rm gal}}{\partial R_{\rm NC}} = \frac{2R^2_{\rm PSF}(R_{\rm obs} - R_{\rm NC})(\boldsymbol{\varepsilon}_{\rm obs} - \boldsymbol{\varepsilon}_{\rm NC} - \boldsymbol{\varepsilon}_{\rm PSF})}{\left[(R_{\rm obs} - R_{\rm NC})^2 - R^2_{\rm PSF})\right]^2} = \frac{2R^2_{\rm PSF}\left(\boldsymbol{\varepsilon}_{\rm gal} - \boldsymbol{\varepsilon}_{\rm PSF}\right)}{R^2_{\rm gal}\sqrt{\left(R^2_{\rm gal} + R^2_{\rm PSF}\right)}}, \quad (36)$$

$$\frac{\partial\boldsymbol{\varepsilon}_{\rm gal}}{\partial\boldsymbol{\varepsilon}_{\rm NC}} = \frac{-(R_{\rm obs} - R_{\rm NC})^2}{(R_{\rm obs} - R_{\rm NC})^2 - R^2_{\rm PSF}} = -\frac{R^2_{\rm gal} + R^2_{\rm PSF}}{R^2_{\rm gal}}. \quad (37)$$

Thus

$$\widehat{\boldsymbol{\epsilon}_{\rm gal}} \approx \boldsymbol{\epsilon}_{\rm gal}\left\{1+\frac{R^2_{\rm PSF}}{R^2_{\rm gal}}\left(\frac{\delta\left(R^2_{\rm PSF}\right)}{R^2_{\rm PSF}}+\frac{2\delta R_{\rm NC}}{\sqrt{R^2_{\rm gal}+R^2_{\rm PSF}}}\right)\right\} - \frac{R^2_{\rm PSF}}{R^2_{\rm gal}}\left\{\boldsymbol{\delta\epsilon}_{\rm PSF}+\frac{R^2_{\rm gal}+R^2_{\rm PSF}}{R^2_{\rm PSF}}\boldsymbol{\delta\epsilon}_{\rm NC} + \left(\frac{\delta\left(R^2_{\rm PSF}\right)}{R^2_{\rm PSF}}+\frac{2\delta R_{\rm NC}}{\sqrt{R^2_{\rm gal}+R^2_{\rm PSF}}}\right)\boldsymbol{\epsilon}_{\rm PSF}\right\}. \quad (38)$$

This imperfect ellipticity measurement translates into additive cosmic shear systematics

$$\mathcal{A} = \frac{1}{P_\gamma^2}\left\langle\frac{R^4_{\rm PSF}}{R^4_{\rm gal}}\right\rangle\sigma^2[|\boldsymbol{\epsilon}_{\rm PSF}|] + \frac{1}{P_\gamma^2}\left\langle\left(1+\frac{R^2_{\rm PSF}}{R^2_{\rm gal}}\right)^2\right\rangle\sigma^2[|\boldsymbol{\epsilon}_{\rm NC}|] + \frac{\langle|\boldsymbol{\epsilon}_{\rm PSF}|^2\rangle}{P_\gamma^2}\left\langle\frac{R^4_{\rm PSF}}{R^4_{\rm gal}}\right\rangle\left(\frac{\left\langle\delta\left(R^2_{\rm PSF}\right)\right\rangle^2}{\left\langle R^4_{\rm PSF}\right\rangle}+\frac{\sigma^2\left[R^2_{\rm PSF}\right]}{R^4_{\rm PSF}}\right) + 4\frac{\langle|\boldsymbol{\epsilon}_{\rm PSF}|^2\rangle}{P_\gamma^2}\left\langle\frac{R^4_{\rm PSF}}{R^4_{\rm gal}}\right\rangle\left(\frac{\langle\delta R_{\rm NC}\rangle^2}{\left\langle R^2_{\rm NC}\right\rangle}+\frac{\sigma^2[R_{\rm NC}]}{R^2_{\rm NC}}\right). \quad (39)$$

Mixing thus emerges between corrections for convolution and non-convolution effects. In the second term for example, imperfections $\boldsymbol{\delta\epsilon}_{\rm NC}$ in the correction for detector effects are enhanced during subsequent deconvolution.

The multiplicative cosmic shear systematics

$$\mathcal{M} = 2\left\langle\frac{R^2_{\rm PSF}}{R^2_{\rm gal}}\right\rangle\left(\frac{\left\langle\delta\left(R^2_{\rm PSF}\right)\right\rangle}{\left\langle R^2_{\rm PSF}\right\rangle}+2\frac{\langle\delta(R_{\rm NC})\rangle}{\langle R_{\rm obs}\rangle}\right) + \left\langle\frac{R^4_{\rm PSF}}{R^4_{\rm gal}}\right\rangle\left(\frac{\sigma^2\left[R^2_{\rm PSF}\right]}{\left\langle R^4_{\rm PSF}\right\rangle}+4\frac{\sigma^2[R_{\rm NC}]}{\left\langle R^2_{\rm obs}\right\rangle}\right) \quad (40)$$

plus bias terms proportional to the square of those already present (therefore negligible if the bias is small). The second term $\langle\delta R_{\rm NC}\rangle$ reflects overall uncertainty in the model of non-convolution effects, such as the density and characteristic release time of charge traps in CCDs. These quantities may be stable over long periods of time, but the error may vary as a function of object flux (hence redshift) if, in this case, the CCD well-filling model is inaccurate. The fourth term $\sigma^2[\delta R_{\rm NC}]$ reflects unaccounted variation of an effect at different positions within a detector. Depending on survey tiling strategies, NC terms are likely to be largest on physical scales corresponding to linear multiples of the chip size (see Cropper et al. 2012).

3.3 Imperfect shape measurement methods

In the previous sections we examined the impact of errors in the measurements of the PSF and detector effects, but we implicitly assumed that the observed galaxy moments are unbiased. In practice, the unweighted size $R_{\rm obs}$ and shape $\boldsymbol{\epsilon}_{\rm obs}$ of a faint galaxy may be subject to errors $\delta R_{\rm obs}$, $\boldsymbol{\delta\epsilon}_{\rm obs}$ for a whole variety of reasons including miscentring, background gradients/structure, pixellization and simply noise. We therefore need to consider also the impact of imperfections in the measurements of the galaxies. This leads to new contributions to the observed ellipticity:

$$\widehat{\boldsymbol{\gamma}} \equiv (\widehat{P_\gamma})^{-1}\,\widehat{\boldsymbol{\epsilon}_{\rm gal}} \quad (41)$$

$$\approx \frac{\boldsymbol{\epsilon}_{\rm gal}}{P_\gamma}+\frac{1}{P_\gamma}\frac{\partial\boldsymbol{\epsilon}_{\rm gal}}{\partial\left(R^2_{\rm PSF}\right)}\delta\left(R^2_{\rm PSF}\right)+\frac{1}{P_\gamma}\frac{\partial\boldsymbol{\epsilon}_{\rm gal}}{\partial\boldsymbol{\epsilon}_{\rm PSF}}\delta\boldsymbol{\epsilon}_{\rm PSF} + \frac{1}{P_\gamma}\frac{\partial\boldsymbol{\epsilon}_{\rm gal}}{\partial(R_{\rm NC})}\delta(R_{\rm NC})+\frac{1}{P_\gamma}\frac{\partial\boldsymbol{\epsilon}_{\rm gal}}{\partial\boldsymbol{\epsilon}_{\rm NC}}\delta\boldsymbol{\epsilon}_{\rm NC} + \frac{1}{P_\gamma}\frac{\partial\boldsymbol{\epsilon}_{\rm gal}}{\partial\left(R^2_{\rm obs}\right)}\delta\left(R^2_{\rm obs}\right)+\frac{1}{P_\gamma}\frac{\partial\boldsymbol{\epsilon}_{\rm gal}}{\partial\boldsymbol{\epsilon}_{\rm obs}}\delta\boldsymbol{\epsilon}_{\rm obs}-\frac{\delta P_\gamma}{{P_\gamma}^2}\boldsymbol{\epsilon}_{\rm gal}. \quad (42)$$

The new partial derivatives of (34) are

$$\frac{\partial\boldsymbol{\epsilon}_{\rm gal}}{\partial\left(R^2_{\rm obs}\right)} = -\frac{R^2_{\rm PSF}}{R^2_{\rm gal}}\frac{(\boldsymbol{\epsilon}_{\rm gal}-\boldsymbol{\epsilon}_{\rm PSF})}{R_{\rm obs}(R_{\rm obs}-R_{\rm NC})}, \quad (43)$$

$$\frac{\partial\boldsymbol{\epsilon}_{\rm gal}}{\partial\boldsymbol{\epsilon}_{\rm obs}} = \frac{R^2_{\rm gal}+R^2_{\rm PSF}}{R^2_{\rm gal}}. \quad (44)$$

Alternatively, note that $\partial\boldsymbol{\epsilon}_{\rm gal}/\partial R_{\rm obs} = -\partial\boldsymbol{\epsilon}_{\rm gal}/\partial R_{\rm NC}$. Including observational error, we thus find the shear measurement $\widehat{\boldsymbol{\gamma}}$ has biases given by STEP parameters (equation 11):

$$\boldsymbol{c} = \frac{1}{P_\gamma}\frac{R^2_{\rm PSF}}{R^2_{\rm gal}}\left\{\frac{R^2_{\rm gal}+R^2_{\rm PSF}}{R^2_{\rm PSF}}(\boldsymbol{\delta\epsilon}_{\rm obs}-\boldsymbol{\delta\epsilon}_{\rm NC})-\boldsymbol{\delta\epsilon}_{\rm PSF} - \left(\frac{\delta\left(R^2_{\rm PSF}\right)}{R^2_{\rm PSF}}+\frac{2\,\delta R_{\rm NC}}{R_{\rm obs}-R_{\rm NC}}+\frac{\delta\left(R^2_{\rm obs}\right)}{R_{\rm obs}(R_{\rm obs}-R_{\rm NC})}\right)\boldsymbol{\epsilon}_{\rm PSF}\right\}, \quad (45)$$

$$m = \frac{R^2_{\rm PSF}}{R^2_{\rm gal}}\left\{\frac{\delta\left(R^2_{\rm PSF}\right)}{R^2_{\rm PSF}}+\frac{2\,\delta R_{\rm NC}}{R_{\rm obs}-R_{\rm NC}}-\frac{\delta\left(R^2_{\rm obs}\right)}{R_{\rm obs}(R_{\rm obs}-R_{\rm NC})}\right\} - \frac{\delta P_\gamma}{P_\gamma}. \quad (46)$$

We have so far considered only shape measurement using unweighted moments. This approach greatly simplifies the calculations, but potentially limits the applicability to real data. This is because the presence of any noise in an image formally leads to infinite noise in the measurements of unweighted moments. It may be feasible to measure (close to) unweighted moments in the special cases of very bright stars, or of repeated detector effects, by stacking data to suppress the noise.

It is never possible in practice to measure directly the unweighted moments of faint galaxies, and one has to use weighted moments instead. The optimal weight function to use is the one that maximizes the signal-to-noise ratio, which in turn implies that the weight function closely resembles the galaxy profile. This is naturally done by methods that fit parametric shape models to the data (e.g. Bridle et al. 2001; Miller et al. 2007; Kitching et al. 2008). Moment-based methods (e.g. Kaiser et al. 1995; Rhodes et al. 2000, hereafter KSB and RRG) instead construct sizes $R_{\rm gal\,w}$ and ellipticities $\boldsymbol{\epsilon}_{\rm gal\,w}$ from quadrupole moments weighted by a radial Gaussian function, the size of which is matched to the object. There are no

simple expressions that relate $R_{\rm gal}$ and $\boldsymbol{\varepsilon}_{\rm gal}$ in terms of observed weighted moments, equivalent to the unweighted versions (31) and (32). Derivations using weighted moments are complicated and involve mixing of higher order moments (Kaiser 2000; KSB; RRG; Réfrégier 2003; Melchior et al. 2011). For any individual galaxy, however, it is possible[4] to define without loss of generality various P' quantities to form a shear estimator from weighted moments:

$$\widehat{\gamma}_{\rm w} \equiv \frac{1}{P_\gamma}\left(1+\frac{1}{P'_R}\frac{R^2_{\rm PSF\,w}}{R^2_{\rm gal\,w}}\right)\left(\frac{\boldsymbol{\varepsilon}_{\rm obs\,w}}{P'_{\boldsymbol{\varepsilon}_{\rm obs}}}-\frac{\boldsymbol{\varepsilon}_{\rm NC\,w}}{P'_{\boldsymbol{\varepsilon}_{\rm NC}}}\right) - \frac{1}{P_\gamma}\frac{1}{P'_R}\frac{R^2_{\rm PSF\,w}}{R^2_{\rm gal\,w}}\left(\frac{\boldsymbol{\varepsilon}_{\rm PSF\,w}}{P'_{\boldsymbol{\varepsilon}_{\rm PSF}}}\right), \quad (47)$$

where we have intentionally arranged terms to resemble equation (33). For individual galaxies, especially those with complex intrinsic shapes, it can be that $\widehat{\gamma}_{\rm w} \neq \widehat{\gamma}$, as long as averaged over a large population of galaxies, $\langle\widehat{\gamma}_{\rm w}\rangle = \langle\widehat{\gamma}\rangle$.

The P' quantities fulfil two roles, and can even be expressed as the product of discrete quantities:

$$P'_x = W_x P_x. \quad (48)$$

Both components of P'_x are tensors, but they are nearly diagonal, so for simplicity we shall treat them as scalars. The first component W_x compensates for the weight function's changes to moments, e.g.

$$\frac{R^2_{\rm PSF\,w}}{R^2_{\rm gal\,w}} \equiv W_R\,\frac{R^2_{\rm PSF}}{R^2_{\rm gal}}, \quad (49)$$

$$\boldsymbol{\varepsilon}_{\rm PSF\,w} \equiv W_{\boldsymbol{\varepsilon}_{\rm PSF}}\,\boldsymbol{\varepsilon}_{\rm PSF}, \quad (50)$$

etc. Numerical values of this component depend upon the shape measurement method, but for small galaxies $W_R \sim 1$ as it governs a ratio of similar quantities and $W_\varepsilon \sim 1/2$ (for all the ellipticities). The second component P_x encodes the way in which the effective PSF is altered by the weight function, and its numerical values depend upon the PSF properties. For a Gaussian PSF, all P values are exactly equal to 1. This approximately holds for a smooth (e.g. ground-based) PSF or a small PSF (or a large galaxy). For an Airy PSF, the outer diffraction wings are damped by the weight function,[5] leading to large differences between weighted and unweighted quantities. For large galaxies, the weight function will be extended and the suppression is small. For small galaxies, size estimates are most affected, and we find $P_R \sim 2$: the net effect of the weight function is equivalent to reducing the PSF size. Ellipticities are less affected, with $P_{{\rm w}_\varepsilon} \sim 1$ in any observing regime. This depends weakly on the intrinsic ellipticity and size but, since we shall generally consider limiting cases of small/faint galaxies, we shall henceforth treat these factors as constants.

We now re-evaluate the additive and multiplicative biases, accounting for the use of weighted moments. This could involve replacing all mentions of observable sizes and shapes by their weighted equivalents. However, for comparison with our earlier results, and to eventually express engineering requirements on instrumentation, it is more convenient to continue to use unweighted quantities. Substituting equations (49) and (50) into (47), we find

$$\widehat{\gamma}_{\rm w} \equiv \frac{1}{P_\gamma}\left(1+\frac{1}{P_R}\frac{R^2_{\rm PSF}}{R^2_{\rm gal}}\right)\left(\frac{\boldsymbol{\varepsilon}_{\rm obs}}{P_{\boldsymbol{\varepsilon}_{\rm obs}}}-\frac{\boldsymbol{\varepsilon}_{\rm NC}}{P_{\boldsymbol{\varepsilon}_{\rm NC}}}\right) - \frac{1}{P_\gamma}\frac{1}{P_R}\frac{R^2_{\rm PSF}}{R^2_{\rm gal}}\left(\frac{\boldsymbol{\varepsilon}_{\rm PSF}}{P_{\boldsymbol{\varepsilon}_{\rm PSF}}}\right). \quad (51)$$

This expression clearly demonstrates how weighted moments can naturally suppress bias. However, this advantage comes at a price. The evaluation of the P factors requires knowledge of higher order shape moments, which can be well known for bright stars but are especially noisy for faint galaxies. The absolute values of $P_{\boldsymbol{\varepsilon}_{\rm PSF}}$, $P_{\boldsymbol{\varepsilon}_{\rm NC}}$ and $P_{\boldsymbol{\varepsilon}_{\rm obs}}$ adjust the balance between different contributions to the bias, but errors in those quantities are functionally identical to errors in $\boldsymbol{\varepsilon}_{\rm PSF}$, $\boldsymbol{\varepsilon}_{\rm NC}$ and $\boldsymbol{\varepsilon}_{\rm obs}$, which we have already considered. Observational errors in P_R propagate into a new source of bias, via

$$\begin{aligned}\widehat{\boldsymbol{\gamma}}_{\rm w} \approx\ & \boldsymbol{\gamma}_{\rm w} + \frac{\partial\,\boldsymbol{\gamma}_{\rm w}}{\partial\left(R^2_{\rm PSF}\right)}\delta\left(R^2_{\rm PSF}\right) + \frac{\partial\,\boldsymbol{\gamma}_{\rm w}}{\partial\boldsymbol{\varepsilon}_{\rm PSF}}\delta\boldsymbol{\varepsilon}_{\rm PSF} \\ & + \frac{\partial\,\boldsymbol{\gamma}_{\rm w}}{\partial(R_{\rm NC})}\delta(R_{\rm NC}) + \frac{\partial\,\boldsymbol{\gamma}_{\rm w}}{\partial\boldsymbol{\varepsilon}_{\rm NC}}\delta\boldsymbol{\varepsilon}_{\rm NC} \\ & + \frac{\partial\,\boldsymbol{\gamma}_{\rm w}}{\partial\left(R^2_{\rm obs}\right)}\delta\left(R^2_{\rm obs}\right) + \frac{\partial\,\boldsymbol{\gamma}_{\rm w}}{\partial\boldsymbol{\varepsilon}_{\rm obs}}\delta\boldsymbol{\varepsilon}_{\rm obs} \\ & + \frac{\partial\,\boldsymbol{\gamma}_{\rm w}}{\partial P_R}\delta(P_R) - \frac{\delta P_\gamma}{P_\gamma}\boldsymbol{\gamma}_{\rm w}, \end{aligned} \quad (52)$$

where the derivatives of $\boldsymbol{\gamma}_{\rm w}$ gain pre-factors of $1/P_R$ or $1/P_R P_\varepsilon$ compared to those of $\boldsymbol{\gamma}$ and

$$\frac{\partial\,\boldsymbol{\gamma}_{\rm w}}{\partial P_R} = -\frac{1}{P_R}\left(\frac{R^2_{\rm gal}}{P_R R^2_{\rm gal}+R^2_{\rm PSF}}\right)\left(\boldsymbol{\gamma}_{\rm w}+\frac{\boldsymbol{\varepsilon}_{\rm PSF}}{P_\gamma P_{\boldsymbol{\varepsilon}_{\rm PSF}}}\right). \quad (53)$$

If the size of the PSF depends upon wavelength, this term introduces a sensitivity to spatial variations in the colour of a galaxy (whereby the PSF is different in the bulge and the disc; Voigt et al. 2012; Semboloni et al., in preparation). This is because multiple galaxy profiles result in galaxies with identical observed moments, so the estimate for P_R becomes biased. Similar biases in P_R also arise in parametric fitting methods if the model does not reflect galaxies' true morphological characteristics (Voigt & Bridle 2010), suffers from aliasing (Bernstein 2010) or is non-linear (Réfrégier et al. 2012). In this paper we do not distinguish between these individual origins, but consider all such effects part of a general method bias.

We conclude that a shear estimator $\widehat{\boldsymbol{\gamma}}_{\rm w}$ constructed from weighted moments has STEP biases:

$$\boldsymbol{c} = \frac{1}{P_\gamma P_R}\frac{R^2_{\rm PSF}}{R^2_{\rm gal}}\left\{\left(\frac{P_R R^2_{\rm gal}+R^2_{\rm PSF}}{R^2_{\rm PSF}}\right)\left(\frac{\delta\boldsymbol{\varepsilon}_{\rm obs}}{P_{\boldsymbol{\varepsilon}_{\rm obs}}}-\frac{\delta\boldsymbol{\varepsilon}_{\rm NC}}{P_{\boldsymbol{\varepsilon}_{\rm NC}}}\right)\right.$$

[4] For example, the shear estimator in KSB (in the absence of non-convolutive effects) is

$$\widehat{\gamma}_{\rm w} = \left(P^\gamma_{\rm KSB}\right)^{-1}\left[\boldsymbol{\varepsilon}_{\rm obs\,w} - P^{\rm sm}\left(P^{\rm sm}_{\rm PSF}\right)^{-1}\boldsymbol{\varepsilon}_{\rm PSF\,w}\right],$$

where

$$P^\gamma_{\rm KSB} = P^{\rm sh} - P^{\rm sm}\left(P^{\rm sm}_{\rm PSF}\right)^{-1}P^{\rm sh}_{\rm PSF}.$$

The interpretation of such quantities is method specific. If $\widehat{\boldsymbol{\gamma}}_{\rm w} \sim (P_\gamma)^{-1}(P'_{\boldsymbol{\varepsilon}_{\rm obs}})^{-1}\boldsymbol{\varepsilon}_{\rm obs\,w}$, the middle factor can be interpreted as part of either the polarizability (e.g. KSB; Massey et al. 2007c, use higher order moments to construct $(P_\gamma P'_{\boldsymbol{\varepsilon}_{\rm obs}})^{-1}$), or as part of the ellipticity (e.g. Kaiser 2000; RRG, use higher order moments to convert weighted ellipticities to unweighted ellipticities $(P'_{\boldsymbol{\varepsilon}_{\rm obs}})^{-1}\boldsymbol{\varepsilon}_{\rm gal\,w}$).

[5] Consider the pathological example of a PSF consisting of a core plus a ring at large radius. The ring lowers the perceived flux of a galaxy, but has no effect on its size or shape as determined from weighted moments.

$$- \frac{\boldsymbol{\delta\varepsilon}_{\rm PSF}}{P_{\boldsymbol{\varepsilon}_{\rm PSF}}} - \frac{\boldsymbol{\varepsilon}_{\rm PSF}}{P_{\boldsymbol{\varepsilon}_{\rm PSF}}} \left(\frac{\delta\left(R^2_{\rm PSF}\right)}{R^2_{\rm PSF}} + \frac{2\,\delta R_{\rm NC}}{R_{\rm obs} - R_{\rm NC}} + \frac{\delta\left(R^2_{\rm obs}\right)}{R_{\rm obs}(R_{\rm obs} - R_{\rm NC})} + \frac{P_R R^4_{\rm gal}}{R^2_{\rm PSF}(P_R R^2_{\rm gal} + R^2_{\rm PSF})} \frac{\delta P_R}{P_R} \right) \Bigg\}, \quad (54)$$

$$m = \frac{1}{P_R} \frac{R^2_{\rm PSF}}{R^2_{\rm gal}} \left(\frac{\delta\left(R^2_{\rm PSF}\right)}{R^2_{\rm PSF}} + \frac{2\,\delta R_{\rm NC}}{R_{\rm obs} - R_{\rm NC}} + \mu \right), \quad (55)$$

where

$$\mu \equiv - \frac{\delta\left(R^2_{\rm obs}\right)}{R_{\rm obs}(R_{\rm obs} - R_{\rm NC})} - P_R \frac{R^2_{\rm gal}}{R^2_{\rm PSF}} \left\{ \frac{\delta P_\gamma}{P_\gamma} + \left(\frac{R^2_{\rm gal}}{P_R R^2_{\rm gal} + R^2_{\rm PSF}} \right) \frac{\delta P_R}{P_R} \right\} \quad (56)$$

is the component of bias due to the galaxy shape measurement method. The STEP parameter $\boldsymbol{q}$, which flags an (incorrect) quadratic response to shear, could be produced by measurement errors that depend on intrinsic ellipticity such as $\boldsymbol{\delta\varepsilon}_{\rm obs}(\boldsymbol{\varepsilon}_{\rm gal})$. Averaged over a galaxy population, these are functionally identical to errors δP_γ.

Observational errors are likely isotropic, i.e. $\langle \boldsymbol{\delta\varepsilon}_{\rm obs} \rangle = 0$, and spatially constant, i.e. in the absence of galaxy–galaxy autocorrelations $\sigma^2[R^2_{\rm obs}] = \sigma^2[|\boldsymbol{\varepsilon}_{\rm obs}|] = \sigma^2[P_\gamma] = \sigma^2[P_R] = 0$. This general case thus has additive cosmic shear systematics:

$$\begin{aligned} \mathcal{A} = {} & \frac{1}{P_R^2 P_\gamma^2} \left\langle \frac{R^4_{\rm PSF}}{R^4_{\rm gal}} \right\rangle \frac{\sigma^2[|\boldsymbol{\varepsilon}_{\rm PSF}|]}{P^2_{\boldsymbol{\varepsilon}_{\rm PSF}}} \\ & + \frac{1}{P_R^2 P_\gamma^2} \left\langle \left(P_R^2 + \frac{R^2_{\rm PSF}}{R^2_{\rm gal}} \right)^2 \right\rangle \frac{\sigma^2[|\boldsymbol{\varepsilon}_{\rm NC}|]}{P^2_{\boldsymbol{\varepsilon}_{\rm NC}}} \\ & + \frac{\langle |\boldsymbol{\varepsilon}_{\rm PSF}|^2 \rangle}{P_R^2 {P_\gamma}^2 P^2_{\boldsymbol{\varepsilon}_{\rm PSF}}} \left\langle \frac{R^4_{\rm PSF}}{R^4_{\rm gal}} \right\rangle \left(\frac{\left\langle \delta\left(R^2_{\rm PSF}\right) \right\rangle^2}{\left\langle R^4_{\rm PSF} \right\rangle} + \frac{\sigma^2\left[R^2_{\rm PSF}\right]}{R^4_{\rm PSF}} \right) \\ & + \frac{4\,\langle |\boldsymbol{\varepsilon}_{\rm PSF}|^2 \rangle}{P_R^2 {P_\gamma}^2 P^2_{\boldsymbol{\varepsilon}_{\rm PSF}}} \left\langle \frac{R^4_{\rm PSF}}{R^4_{\rm gal}} \right\rangle \left(\frac{\langle \delta R_{\rm NC} \rangle^2}{\left\langle R^2_{\rm NC} \right\rangle} + \frac{\sigma^2[R_{\rm NC}]}{R^2_{\rm NC}} \right) \\ & + \frac{\langle |\boldsymbol{\varepsilon}_{\rm PSF}|^2 \rangle}{P_R^2 {P_\gamma}^2 P^2_{\boldsymbol{\varepsilon}_{\rm PSF}}} \left\langle \frac{R^4_{\rm PSF}}{R^4_{\rm gal}} \right\rangle \alpha^2, \end{aligned} \quad (57)$$

where

$$\alpha^2 \equiv \frac{\left\langle \delta\left(R^2_{\rm obs}\right) \right\rangle^2}{\left\langle R^4_{\rm obs} \right\rangle} + \left\langle \frac{R^4_{\rm gal}}{R^4_{\rm PSF}} \right\rangle \left\langle \left(\frac{P_R R^2_{\rm gal}}{P_R R^2_{\rm gal} + R^2_{\rm PSF}} \right)^2 \right\rangle \frac{\langle \delta P_R \rangle^2}{\left\langle P_R^2 \right\rangle}. \quad (58)$$

The first term in α^2 could arise due to pixellization effects, but this will be zero for resolved imaging and deviations could be measured only by changing the plate scale in a camera. Note that if autocorrelations of galaxy shapes with themselves are included in the correlation function analysis (see Section 2.1), the additive cosmic shear systematics gain an extra white noise term $\sigma^2_\gamma(z_A, z_B)$ as in equation (18). The multiplicative cosmic shear systematics become

$$\mathcal{M} = \frac{2}{P_R} \left\langle \frac{R^2_{\rm PSF}}{R^2_{\rm gal}} \right\rangle \left(\frac{\left\langle \delta\left(R^2_{\rm PSF}\right) \right\rangle}{\left\langle R^2_{\rm PSF} \right\rangle} + 2 \frac{\langle \delta(R_{\rm NC}) \rangle}{\langle R_{\rm obs} \rangle} + \langle \mu \rangle \right) + \frac{1}{P_R^2} \left\langle \frac{R^4_{\rm PSF}}{R^4_{\rm gal}} \right\rangle \left(\frac{\sigma^2\left[R^2_{\rm PSF}\right]}{\left\langle R^4_{\rm PSF} \right\rangle} + 4 \frac{\sigma^2[R_{\rm NC}]}{\left\langle R^2_{\rm obs} \right\rangle} \right). \quad (59)$$

4 REQUIREMENTS TO MEET FUTURE SCIENCE GOALS

4.1 How much systematic bias is tolerable?

Total experimental error from any measurement is always a combination of systematic and statistical errors. Systematic bias (i.e. the deviation of a measured value from the truth) can be reduced by e.g. stabilizing a telescope or raising it above the Earth's atmosphere. Statistical error (i.e. the confidence interval allocated to a measured value) is limited by the finite number of measurements within a survey, and can be reduced by e.g. increasing survey volume. The diagrams inset within Fig. 1 illustrate how an unrecognized systematic bias shifts measurements, which are drawn from a statistical likelihood distribution around the offset value.

Classical astronomical survey design optimizes an observation that is limited by systematic biases inherent to a technique or its interpretation. This limitation drives surveys wider, deeper or to higher resolution, until their statistical error becomes smaller than the systematic bias. However, several surveys planned for the next decade have scientific goals that require them to image the entire sky outside the plane of the Milky Way. Further increasing survey area is impossible, and increasing survey depth can be prohibitively expensive: especially for space-based surveys, where mission cost jumps in step functions with mirror size (bigger launch vehicles become necessary) or survey duration (additional redundancy of components). For these surveys, the statistical error is fixed and the classical trade-off is inverted; the key question becomes *how much systematic bias is tolerable?* We shall answer this quantitatively by considering the probability with which an experiment's reported measurement of a particular parameter could have been obtained by an unbiased experiment. This is the overlap integral between the likelihood distribution reported by an experiment, and the likelihood distribution that would have been reported by an unbiased experiment (i.e. the same distribution, recentred around the parameter's true value).[6]

Throughout this section, we shall consider statistical errors described by a Gaussian of width σ. There are two ways in which a systematic bias can be described. Following frequent use in the literature, and as illustrated in the upper inset panel of Fig. 1, we shall first consider an experiment with an exact amount of bias, b. The lowest curve in Fig. 1 shows the probability that a reported measurement could have been sampled from the unbiased (re-centred) likelihood. This is simply the (cross-hatched) overlap integral under

[6] This is a frequentist argument based on p-value like statistics; a Bayesian methodology, in which evidence ratios for a target model with and without systematics could also be considered. We shall also only consider the one-dimensional bias on a single parameter at a time (cf. Dodelson, Shapiro & White 2006; Shapiro 2009; Shapiro et al. 2010).

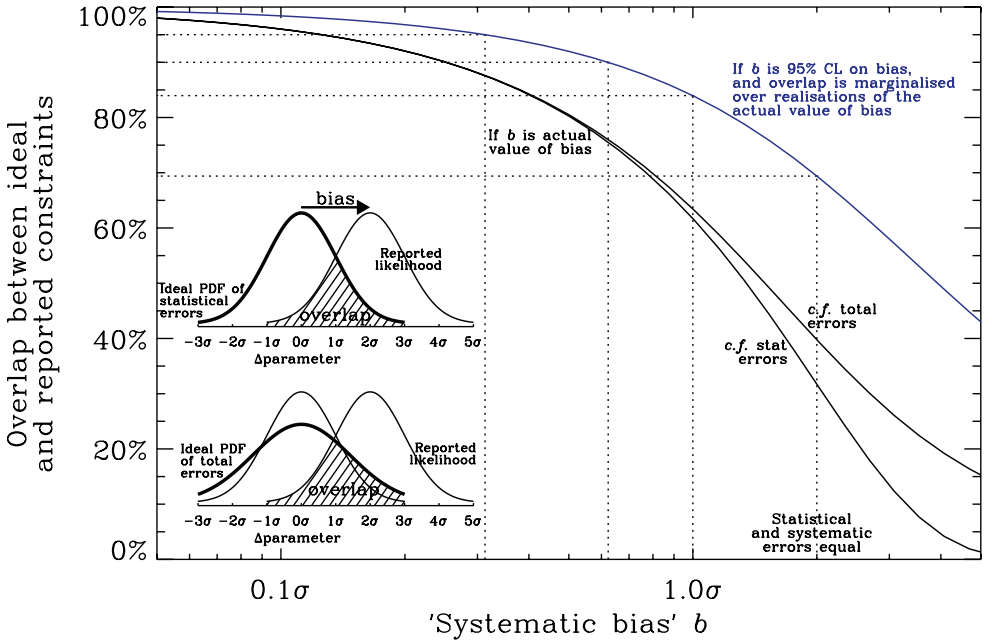

Figure 1. The effect of a systematic bias in an experiment, as a function of statistical error σ, assuming all likelihood distributions are Gaussian. The y-axis is the chance that a reported experimental result is drawn from the reported likelihood distribution, centred around the true value. The two lower (black) curves show this chance in the presence of some exact systematic bias b, for the reported statistical errors (lower) or total errors (upper). The top (blue) curve shows this chance if b is instead a 95 per cent CL on the (unknown) true bias. In this case, the true bias *could* be zero, so the overlap with the ideal likelihood distribution is always greater. Using this latter definition, we require $b < 0.31\sigma$ for a 95 per cent overlap with the ideal PDF.

two Gaussians with variance $\sigma_1 = \sigma_2 = \sigma$ and mean $\mu_1 = 0$, $\mu_2 = b$:

$$p^{\rm stat}_{\rm overlap}(b) = \int_{-\infty}^{\infty} \min\left\{ \frac{{\rm e}^{\frac{-x^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}, \frac{{\rm e}^{-\frac{(x-b)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} \right\} {\rm d}x \tag{60}$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \left(\int_{-\infty}^{\frac{|b|}{2}} {\rm e}^{\frac{-(x-|b|)^2}{2\sigma^2}} {\rm d}x + \int_{\frac{|b|}{2}}^{\infty} {\rm e}^{-\frac{x^2}{2\sigma^2}} {\rm d}x \right) \tag{61}$$

$$= 1 - {\rm erf}\left(\frac{1}{2\sqrt{2}} \frac{|b|}{\sigma} \right). \tag{62}$$

For the overlap to be at least 95 per cent (90 per cent), the absolute value of bias $|b|$ must be less than 0.13σ (0.25σ). If bias is allowed to be as large as the 1σ statistical error, the overlap integral is only 62 per cent, which is undesirable. One effect slightly improves this: as illustrated in the lower inset diagram, reported error bars will be enlarged to account for an estimate of the systematic bias. The middle curve in Fig. 1 shows what happens if the achieved level of bias were treated as a 95 per cent confidence limit (CL) on a Gaussian systematic error budget, i.e. $\sigma_{\rm b} = b/2$. In this case, the overlap integral becomes

$$p^{\rm total}_{\rm overlap}(b) = \int_{-\infty}^{\infty} \min\left\{ \frac{{\rm e}^{\frac{-x^2}{2(\sigma^2+\sigma_{\rm b}^2)}}}{\sqrt{2\pi(\sigma^2+\sigma_{\rm b}^2)}}, \frac{{\rm e}^{-\frac{(x-b)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} \right\} {\rm d}x, \tag{63}$$

although this does not significantly affect p.

However, any systematic bias that is known exactly would already have been subtracted from a measurement! We shall now re-interpret b as the 95 per cent CL on the absolute value of an *unknown* bias. A Gaussian distribution of possible biases with mean zero and width $\sigma_{\rm b} = b/2$ sometimes creates small or even zero bias, so the overlap of reported and ideal measurements is greater. Marginalizing over this distribution, the top curve in Fig. 1 shows

$$p^{\rm marginalized}_{\rm overlap}(b) = \frac{1}{\sqrt{2\pi\sigma_{\rm b}^2}} \int_{-\infty}^{\infty} {\rm e}^{\frac{-b'^2}{2\sigma_{\rm b}^2}} p^{\rm total}_{\rm overlap}(b')\, {\rm d}b'. \tag{64}$$

Achieving a 95 per cent (90 per cent) probability that a reported result could have been drawn from the likelihood distribution re-centred on the true value now requires $|b| < 0.31\sigma$ (0.62σ). Only 69 per cent overlap arises if the systematic and statistical error budgets are equal ($\sigma_{\rm b} = \sigma$). We shall henceforth require uncertain biases to have a 95 per cent CL that is less than 31 per cent of the 1σ expected statistical error.

4.2 Propagation of shear measurement errors to biases on cosmological parameters

We now propagate hypothetical shear measurement errors $\mathcal{A}(\ell, z_A, z_B)$ and $\mathcal{M}(\ell, z_A, z_B)$ from Section 3 into biases on derived cosmological parameters, via the Fisher matrix bias formalism (Taylor & Watts 2001; see also Amara & Réfrégier 2007; Kitching et al. 2009b). In particular, we concentrate on measurements of the dark energy equation of state parameter w or its derivative w_a (Chevalier & Polarski 2001; Linder 2003), and marginalize over other parameters. By requiring that the 95 per cent CL on bias is less than 31 per cent of the statistical errors afforded by Poisson noise in a finite survey volume (see Section 4.1), we obtain requirements on the accuracy with which the PSF must be modelled, detector effects must be corrected and galaxy shapes must be measured. This is more stringent than the work of Amara & Réfrégier (2008), who required bias less than 100 per cent of statistical error.

We assume a baseline 15 000 deg^2 cosmic shear survey resolving 30 galaxies per square arcminute with median redshift of 1.0 and split into 10 tomographic redshift bins. This matches the configuration of the proposed *Euclid* mission (Laureijs et al. 2011), and is likely to be similar to any proposed Stage IV survey: for example, LSST proposes to survey 18 000 deg^2 with an effective density of 40 galaxies per square arcminute (Jee & Tyson 2011; Bradshaw et al. 2012). Das et al. (2011) describe the effect of perturbing the parameters of the baseline survey in a similar analysis.

We use the ICOSMO Fisher matrix software (Kitching et al. 2009b; Réfrégier et al. 2011) to calculate the concordance ΛCDM cosmic shear power spectrum $C(\ell, z_A, z_B)$ in a top-hat basis set (200 bins) spanning scales $10 < \ell < 5000$ and every pair of redshift bins. Henceforth, ℓ, z_A and z_B refer to the median values of the population of galaxies within these bins. We assume the Limber approximation, and neglect any power spectrum due to intrinsic alignments. Using only weak lensing measurements, such an experiment can measure w with a 1σ, one-parameter statistical error of 0.065, and w_a with a statistical error of 0.41.

4.3 Constant additive and multiplicative shear measurement bias

To first explore the consequences of the simplest possible systematic errors, we first impose upon each measurement of $C(\ell)$ a constant additive shear measurement bias $\mathcal{A}$ (or σ_c^2) and a constant multiplicative shear measurement bias $\mathcal{M}$ (or m). This simultaneity of multiplicative and additive biases has not been explored before, with previous studies in the literature considering the imposition of only one type of systematic at a time. Note that although σ_c^2 is positive by definition, and m is almost always negative in practice (e.g. Bridle et al. 2010), we explore positive and negative values in both cases because *if their values are known, they would be removed from data* (or added to models). The *only* important parameter is the residual after this process, i.e. the accuracy to which $\mathcal{A}$ and $\mathcal{M}$ are known. By definition, this residual is equally likely to be either positive or negative.

We find that there is a degeneracy between the two types of bias, in terms of the way they influence constraints on the dark energy equation of state parameter w (Fig. 2, top panel). Indeed, if $\mathcal{A}$ and $\mathcal{M}$ have the same sign, they can cancel each other out to produce no net bias on w. However, the tuning of this cancellation is specific to the parameter being measured: the degeneracy is completely different for measurements of w_a, Ω_m or σ_8.

Given our first order expansion, it is not surprising that the surface of Fig. 2 is approximately fitted by a plane $b/\sigma \approx -0.093 - 3.9m + 3.3 \times 10^{10}\sigma_c^2$. Thus, if the signs of $\mathcal{A}$ and $\mathcal{M}$ are not known a priori, guaranteeing $|b| < 0.31\sigma$ requires

$$|m| + 8.6 \times 10^9 \left|\sigma_c^2\right| \lesssim 0.10. \tag{65}$$

Whilst surprisingly large constant m can be acceptable for measurements of w (since $\mathcal{M}C(\ell)$ does then not resemble $\partial C(\ell)/\partial w$), we again note that this is not true for measurements of other cosmological parameters. Most importantly, we note the necessity for joint requirements on $\mathcal{A}$ and $\mathcal{M}$. Whenever requirements are placed on $\mathcal{A}$ when assuming $\mathcal{M} \equiv 0$ or vice versa, one degenerate error budget is being spent twice. The two requirements should be halved and, since the bias surface is well fit by a plane, the two requirements can be linearly traded against each other. This degeneracy has not been taken into account by earlier work.

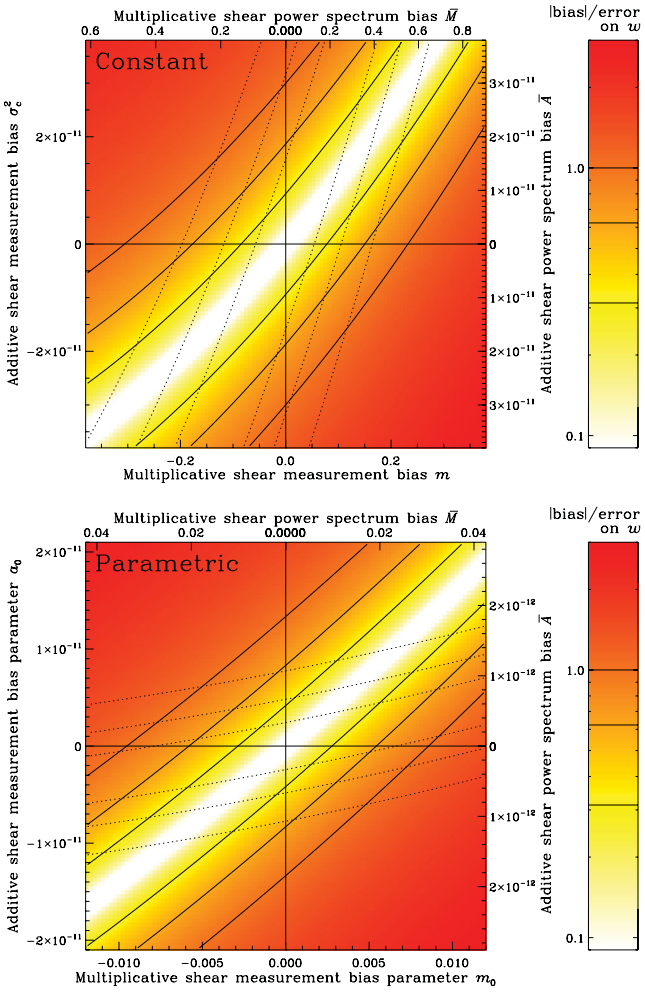

Figure 2. The (absolute value of) bias on measurements of the dark energy equation of state parameters w (colour and solid contours) and w_a (dotted contours) from weak lensing surveys with multiplicative and additive shear measurement systematics. Bias is shown as a multiple of the expected statistical error σ, and contours are drawn at the same values as in Fig. 1. Top panel: constant systematics m and σ_c^2. Bottom panel: one realization of variable systematics $\mathcal{M}(z)$ and $\mathcal{A}(\ell)$, as described in the text (note the change of scale).

4.4 Simple forms of additive and multiplicative shear measurement bias

As discussed in Section 3, systematics often affect some physical scales more than others, and it is typically more difficult to measure the shapes of distant (small, faint) galaxies than nearby (big, bright) ones. One feasibly more realistic functional form for non-constant additive systematics is

$$\mathcal{A}(\ell) = a_0 \left(1 + \frac{\ell}{\ell_0}\right)^{\beta_2-\beta_1} \left(\frac{\ell}{\ell_0}\right)^{\beta_1}, \tag{66}$$

where $\ell_0 = 1000$, $\beta_1 = -1.5$, $\beta_2 = -3$ (equation 29 of Amara et al. 2010). A feasible functional form for multiplicative systematics is

$$\mathcal{M}(z_A, z_B) = m(z_A) + m(z_B) + m(z_A)\,m(z_B), \tag{67}$$

where

$$m(z) = m_0 \frac{2}{\pi}(1+z)^{\beta_\mathrm{m}} \tan^{-1}(\alpha_\mathrm{m}(z - z_\mathrm{T})), \tag{68}$$

with $\alpha_m = 10$, $\beta_m = 1.5$ and a transition in sign at $z_T = 1$ (equation 20 of Amara & Réfrégier 2008).

The bias surface for this parametrization (Fig. 2, bottom panel) is well fit by a plane $b/\sigma \approx 0.031 + 110m_0 - 7.5 \times 10^{10} a_0$. This means that, while the error budget,

$$|m_0| + 6.7 \times 10^8 |a_0| \lesssim 2.8 \times 10^{-3}, \tag{69}$$

must again be split between additive and multiplicative systematics, the allocations can still be traded linearly against each other. Note that absolute requirements on parametric variables a_0 and m_0 are tighter than those on σ_c^2 and m partly because the unnormalized functions are much lower than unity, and partly because $\mathcal{A}$ and $\mathcal{M}C$ are now more similar to $\partial C/\partial w$.

4.5 General forms of additive and multiplicative shear measurement bias

Since the real scale dependence of systematics will remain unknown for any survey (even after its completion), we now use a Monte Carlo approach (cf. Kitching et al. 2009a) to explore *all possible* functional forms of $\mathcal{A}$ and $\mathcal{M}$. We explore this very high dimensional parameter space separately for each type of bias, but remember the caveat about duplicated error budgets and the necessity/ability to trade between requirements on each. In general, requirements will emerge upon the functional forms of $\mathcal{A}$ and $\mathcal{M}$. For tractability, we collapse each function to a single number:

$$\overline{\mathcal{A}} \equiv \frac{\sum_{z\,\mathrm{bins}} \frac{1}{2\pi} \int_{\ell_{\min}}^{\ell_{\max}} |\mathcal{A}(\ell, z_A, z_B)| \ell^2 \mathrm{d}\ln\ell}{\sum_{z\,\mathrm{bins}} \frac{1}{2\pi} \int_{\ell_{\min}}^{\ell_{\max}} \ell^2 \mathrm{d}\ln\ell}, \tag{70}$$

$$\overline{\mathcal{M}} \equiv \frac{\sum_{z\,\mathrm{bins}} \frac{1}{2\pi} \int_{\ell_{\min}}^{\ell_{\max}} |\mathcal{M}(\ell, z_A, z_B)| \ell^2 \mathrm{d}\ln\ell}{\sum_{z\,\mathrm{bins}} \frac{1}{2\pi} \int_{\ell_{\min}}^{\ell_{\max}} \ell^2 \mathrm{d}\ln\ell}. \tag{71}$$

Thus we generalize σ_{sys}^2 in Amara & Réfrégier (2008) to 3D correlation functions, and include a renormalization, by way of the denominator, that reduces sensitivity to changes in the adopted ℓ-range. Values of these performance indicators are shown on the right and upper axes of Fig. 2. For our baseline survey, the denominator in (70) and (71) is $55 \times 9.0 \times 10^5$. For the shorter ℓ-range used by Gravitational Lensing Accuracy Testing 2010 (GREAT10), the denominator is 1.8×10^5. Other possible choices for the weighting inside the integral, and the slightly different approach required for practical calculations in GREAT10, are discussed in Appendix B.

To span the space of possible systematics functions, we generate 100 000 random realizations of $\mathcal{A}(\ell, z_A, z_B)$; for now, we set $\mathcal{M} \equiv 0$ assuming conservatively that generic systematics contribute equally to all scales and redshift bins, we generate random systematics by drawing the value of $\sigma_c(\ell, z)$ in each ℓ and z bin from a Gaussian PDF centred about 0. The width of the Gaussian remains fixed as a function of ℓ and z, and we repeat this process several times with increasingly wide Gaussians (spanning a range that includes current performance and future requirements). We then smooth σ_c with a 2D boxcar of width 50 (of 200) ℓ bins and 3 (of 10) z bins, and construct $\mathcal{A}(\ell, z_A, z_B) \equiv \sigma_c(\ell, z_A)\sigma_c(\ell, z_B)$. The smoothing reflects the typically continuous form of systematic effects; it is important here because (unrealistic) realizations of systematics that are uncorrelated between adjacent bins cause less bias in cosmological parameters. The precise amount of smoothing (particularly in the ℓ direction) affects requirements on $\overline{\mathcal{A}}$ by around 15 per cent of the nominal value. While this precision is adequate for current planning purposes, detailed analysis in the future will require more accurately constrained forms of $\mathcal{A}$ and $\mathcal{M}$ to be propagated.

We propagate our random realizations of biases on the cosmic shear power spectrum into biases on w using the Fisher matrix bias formalism as before. The largest biases are generated when the shape of $\mathcal{A}(\ell, z_A, z_B)$ is close to that of $\partial C(\ell, z_A, z_B)/\partial w$. To ensure that the bias on w is less than 31 per cent of the statistical error for 95 per cent of the random realizations, we require

$$\overline{\mathcal{A}} \lesssim 1.8 \times 10^{-12} \tag{72}$$

(see Fig. 3), including a factor of 1/2 for a non-zero budget on $\mathcal{M}$. This general requirement is a factor of only ∼3 tighter than the requirement if $\mathcal{A}$ is constant (see the upper panel of Fig. 2 or the dotted line in Fig. 3a), demonstrating how bad constant additive systematics can be. Conversely, it is a factor of ∼3 looser than if $\mathcal{A}$ is restricted to the family of curves parametrized by equation (66)

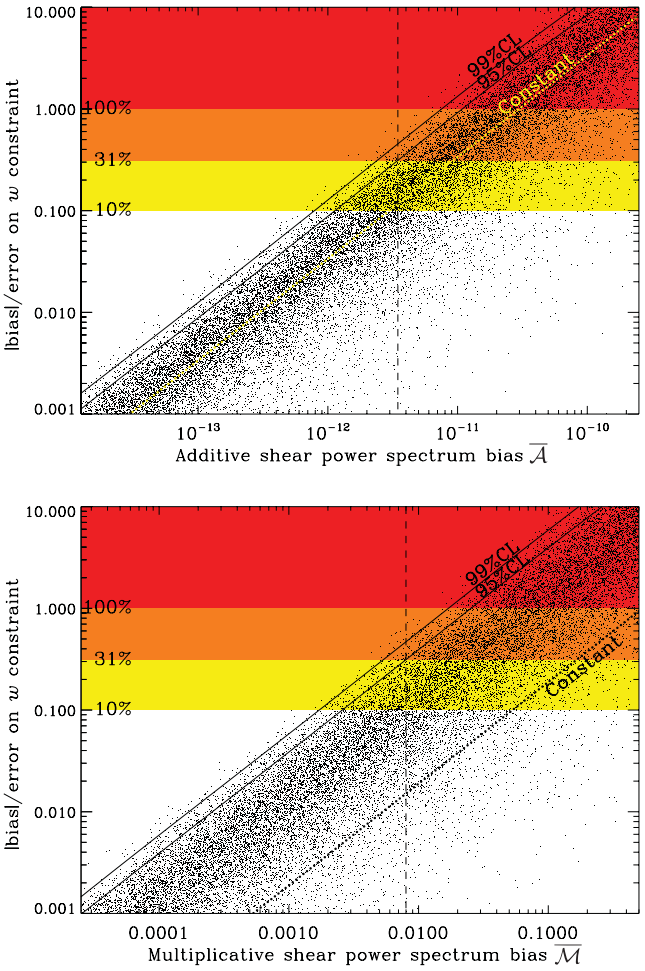

Figure 3. The bias on measurements of the dark energy equation of state parameter w from weak lensing surveys with (top panel) additive and (bottom panel) multiplicative shear measurement systematics. Each data point shows a random realization of systematics with a unique dependence upon angular scale and redshift (for clarity, only one in three are plotted). The dotted diagonal lines show the bias on cosmological parameters if the shear measurement systematics are constant. The solid diagonal curves show limiting values that include 95 and 99 per cent of random realizations with a given value of $\overline{\mathcal{A}}$ or $\overline{\mathcal{M}}$. An all-sky 3D cosmic shear survey will only be deemed successful if the measurement bias is $\lesssim 31$ per cent of the statistical measurement error. At 95 per cent CL, this will require (vertical dashed line) shear measurement better than $\overline{\mathcal{A}} \lesssim 3.5 \times 10^{-12}$ if $\mathcal{M} \equiv 0$, and $\overline{\mathcal{M}} \lesssim 8.0 \times 10^{-3}$ if $\mathcal{A} \equiv 0$.

(see lower panel of Fig. 2), which was a pathological case in the worst 1 per cent of random configurations.

For the smallest resolved galaxies $R_{\rm obs} \approx 1.25R_{\rm PSF}$ (i.e. $R_{\rm gal} = 0.75R_{\rm PSF}$), in the regime of the most elliptical PSF typically obtained from astronomical instruments $|\boldsymbol{\epsilon}_{\rm PSF}| \approx 0.1$, and with an Airy PSF such that $P_R \sim 2$, equations (57) and (72) together become

$$\begin{aligned}\overline{\mathcal{A}} \approx\ & 0.79\sigma^2[|\boldsymbol{\delta\epsilon}_{\rm PSF}|] + 5.2\sigma^2[|\boldsymbol{\delta\epsilon}_{\rm NC}|] \\ &+ 0.0023\left\{\frac{\left(\delta\left(R^2_{\rm PSF}\right)^2\right)}{R^4_{\rm PSF}} + \frac{\sigma^2\left[R^2_{\rm PSF}\right]}{R^4_{\rm PSF}}\right\} \\ &+ 0.0091\left\{\frac{(\delta R_{\rm NC})^2}{R^2_{\rm NC}} + \frac{\sigma^2[R_{\rm NC}]}{R^2_{\rm PSF}}\right\} \\ &+ 0.0023\frac{\left(\delta\left(R^2_{\rm obs}\right)^2\right)}{R^4_{\rm obs}} \qquad \lesssim 1.8\times10^{-12}. \end{aligned} \tag{73}$$

Note that $\boldsymbol{\delta\epsilon}_{\rm PSF}$ at least is likely to have two components that each contribute to the total bias.

We then generate 100 000 random multiplicative shear measurement biases $m(\ell, z)$ in the same way and with the same smoothing. We propagate these into multiplicative cosmic shear systematics $\mathcal{M}(\ell, z_A, z_B)$ via equation (67), and hence into biases on w. To ensure measurement bias is less than 31 per cent of statistical errors for 95 per cent of the Monte Carlo realizations, we require

$$\overline{\mathcal{M}} \lesssim 4.0\times10^{-3} \tag{74}$$

(see Fig. 3b), including a factor of 1/2 for a finite error budget on $\mathcal{A}$. This is a factor of ∼20 tighter than the requirements if $\mathcal{M}$ is constant (see the lower panel of Fig. 2 or the dotted line in Fig. 3b), demonstrating again that a constant multiplicative shear measurement bias has surprisingly little effect on w constraints (note that it does strongly affect constraints on $\Omega_{\rm m}$ and σ_8). The amount by which the random systematics are smoothed (particularly in the z direction) affects requirements on $\overline{\mathcal{M}}$ by around 10 per cent of the nominal value. For the smallest resolved galaxies, equations (59) and (74) become

$$\overline{\mathcal{M}} \approx 1.8\frac{\left\langle\delta\left(R^2_{\rm PSF}\right)\right\rangle}{\left\langle R^2_{\rm PSF}\right\rangle} + 3.6\left\langle\frac{\delta R_{\rm NC}}{R_{\rm obs}}\right\rangle + 2\langle\mu\rangle \lesssim 4.0\times10^{-3} \tag{75}$$

plus redundant variance terms that are already constrained more tightly by equation (73), so which we drop here.

A top-down analysis can now allocate error budgets to each of the components of $\overline{\mathcal{A}}$ and $\overline{\mathcal{M}}$, as expanded in equations (73) and (75). In the absence of other information, a natural choice would perhaps allocate budgets in, perhaps in inverse proportion to the coefficient by which they affect the overall science. Cropper et al. (2012) provide one such breakdown of these error budgets that is feasible in a dedicated space mission.

4.6 Comparison to other work

Our calculations differ from those of Amara & Réfrégier (2008), Chang et al. (2012) and Cardone et al. (in preparation) by using a form-filling approach to consider any possible ℓ-dependence of systematics, rather than just parametric forms. Amara & Réfrégier (2008) also assumed only a 2D cosmic shear analysis, with a slightly lower redshift distribution of source galaxies, and considered power spectrum measurements up to scales $\ell < 20\,000$; we exclude such non-linear scales because poorly understood effects of baryonic physics are likely to make them difficult to interpret (Kitching, Heavens & Miller 2011a; Semboloni et al. 2011). The denominator we introduced in equations (70) and (71) keeps our new $\overline{\mathcal{A}}$ and $\overline{\mathcal{M}}$ performance indicators independent (within a few per cent) of this choice of ℓ-range. However, if future understanding of small-scale baryonic effects could indeed extend cosmic shear measurements to $\ell = 20\,000$, statistical errors σ would shrink by ∼10 per cent. Exploiting this new information would require correspondingly smaller shear measurement biases.

We can mimic the 2D notation of Amara & Réfrégier (2008) by multiplying our 3D requirement (72) by its denominator, dividing it by the number of (in our case 55) redshift bin pairs that we considered, and including small corrections. This process plus small corrections for a few other differences (ℓ-range, z-distribution, $|b|/\sigma < 1$, 100 per cent CL) yields a pseudo-2D requirement a factor of ∼2 looser than their $\sigma^2_{\rm sys} \lesssim 10^{-7}$ per redshift bin. That difference presumably arises from the details of the redshift slicing, and we shall not consider it further.

5 CAN THE REQUIREMENTS BE MET?

5.1 Current best shear measurement performance

The performance of shape measurement algorithms can be tested on simulated astronomical images that contain a known shear signal. Blind competitions include the community-wide STEP (Heymans et al. 2006; Massey et al. 2007b) and GREAT (Bridle et al. 2010; Kitching et al. 2011b) programmes; these have been and are continuing to be supplemented by efforts by individual groups targeted towards specific surveys. Assessed using the GREAT metric Q, these programmes have yielded a steady improvement by a factor of ∼3.5 yr^{-1} over the past decade (Kitching et al. 2012b).

GREAT10 is the most recent blind competition, and the first to employ variable shear simulations, which are required to test scale-dependent issues. The best methods entered into GREAT10 achieve $\overline{\mathcal{A}} \sim 2.7\times10^{-12}$ and $\overline{\mathcal{M}} \sim 3.1\times10^{-3}$ on bulge+disc galaxies at detection signal-to-noise ratio (S/N) = 40 (table 4 of Kitching et al. 2012a, in which these values are expressed as $\sqrt{\overline{\mathcal{A}}}$ and $\overline{\mathcal{M}}/2$, but see Appendix B for a discussion of slight differences in approach). For these fairly bright galaxies, current performance surpasses the requirement on $\overline{\mathcal{M}}$ and the requirements on $\overline{\mathcal{A}}$ and $\overline{\mathcal{M}}$ can be traded against each other to also be met in combination. Note, however, that this shape measurement inaccuracy uses all but 1 per cent of the entire error budget. GREAT10 assumed a spatially/temporally varying PSF,[7] but that it was perfectly known, and that non-convolution effects could be perfectly corrected. Further development in shape measurement will be necessary if part of the error budget is to be set aside for e.g. PSF or charge transfer inefficiency modelling errors.

Faint galaxies are harder to measure, but must be included to reach Stage IV surveys' statistical goals on cosmological parameter estimation. At detection S/N = 20, the best methods now achieve $\overline{\mathcal{A}} \sim 2.1\times10^{-11}$ and $\overline{\mathcal{M}} \sim 5.6\times10^{-3}$; at detection S/N = 10 they

[7] The GREAT10 simulations used ground-based PSF morphologies, but STEP3 (see http://www.roe.ac.uk/ heymans/step/step3_results.html) concluded that the only factor affecting shear measurement performance was the ratio of the PSF size to the pixel size. STEP3 was a space-based equivalent of STEP2, run as another public, blind competition. Its results were never published because they were essentially identical to those from STEP2. The main conclusion was that equivalent shear measurement performance could obtained for small galaxies from space as for similarly resolved larger galaxies from the ground, irrespective of PSF morphology.

achieve $\overline{\mathcal{A}} \sim 7.4 \times 10^{-11}$ and $\overline{\mathcal{M}} \sim 1.1 \times 10^{-2}$. If all galaxies were this faint, exploiting them (consuming all of the available error budget) would exceed requirements in $\sqrt{\overline{\mathcal{A}}}$ by a factor of 3.5–6.5 and in $\overline{\mathcal{M}}$ by a factor of 1.4–2.8. If an analysis were to proceed using extant shear measurement methods, accounting for residual systematic biases would necessarily enlarge the reported error bars – if all galaxies were at detection S/N = 10, 95 per cent of realizations of bias would simultaneously satisfy $|b|/\sigma < 3.7$ for $\overline{\mathcal{A}}$ and $|b|/\sigma < 1$ for $\overline{\mathcal{M}}$ (see Fig. 3).

Shape measurement algorithms can be improved either by fundamental progress or by calibration on accurate simulated images. Extrapolating the current rate of fundamental development (Kitching et al. 2012b) suggests that, with even minimal continued development, the required algorithmic performance will be surpassed, and substantial margin will be achieved, well before the need to analyse Stage IV surveys. Indeed, noise bias (Kacprzak et al. 2012; Melchior & Viola 2012) was unaccounted for by all GREAT10 methods, but appears in faint galaxies at a level consistent with its being the dominant source of bias (Réfrégier et al. 2012). Proper treatment of noise bias will therefore improve performance for faint galaxies. Several additional improvements have also been suggested (e.g. Bernstein 2010; Viola, Melchior & Bartelmann 2011). For the first time, methods are thus emerging with sufficient accuracy to reliably and fully exploit the statistical potential of Stage IV cosmic shear surveys. Simulations could then be used solely as external verification tests of data analysis pipelines. Dedicated simulation efforts are continuing inside the teams of all weak lensing surveys,[8] and the GREAT3 programme (Mandelbaum et al., in preparation) is currently being designed by a worldwide collaboration of the weak lensing community.

5.2 Empirical diagnosis of residual additive systematics

Although the greatest improvement is formally required in additive cosmic shear measurement biases, they are potentially the least troublesome. Many additive systematics can be internally diagnosed within a shear catalogue, and those that do arise can potentially even be calibrated out at the catalogue level. This procedure has a long heritage in *Hubble Space Telescope* (*HST*) analyses (e.g. Hoekstra et al. 1998, 2011; Rhodes et al. 2004, 2007; Miralles et al. 2005; Jee et al. 2007, 2011; Schrabback et al. 2010).

5.2.1 *Calibrating PSF model errors*

The best way to internally diagnose PSF modelling errors $\delta R_{\rm PSF}$ and $\boldsymbol{\delta\varepsilon}_{\rm PSF}$ is to bootstrap real stellar shapes. The PSF model can be constructed from all but a few of the available stars, then interpolated to the positions and colours of the remaining stars as well as the galaxies. Any offset between the predicted and measured values will be a sum of $\delta R_{\rm PSF} + \delta R_{\rm obs}$ and $\boldsymbol{\delta\varepsilon}_{\rm PSF} + \boldsymbol{\delta\varepsilon}_{\rm obs}$, but the observational contributions should average to zero over a large population of stars. The number of degrees of freedom in PSF variation due to thermomechanical instability (Jarvis & Jain 2004; Rhodes et al. 2007; Schrabback et al. 2010), atmospheric turbulence (Jarvis, Schechter & Jain 2008) or changing gravity load (Iye et al. 2004) can also be usefully compared to engineering predictions from ray tracing through optics models (Krist 1995; Hook & Stoehr 2008).

[8] See www.darkenergysurvey.org, www.lsst.com, www.euclid-ec.org, http://www.naoj.org/Projects/HSC/, Rhodes et al. (2012).

A vital test of successful PSF deconvolution is obtained from the correlation of measured shears with the PSF ellipticity. No residual signature of the system's PSF should find its way into the galaxy shape catalogue, so these should be uncorrelated. However, in a flawed shear measurement, taking unweighted $\widehat{\boldsymbol{\varepsilon}_{\rm PSF}} \equiv \boldsymbol{\varepsilon}_{\rm PSF} + \boldsymbol{\delta\varepsilon}_{\rm PSF}$ and $\widehat{\boldsymbol{\gamma}}_{\rm w}$ from equations (54) and (55), we obtain

$$\begin{aligned}\left\langle \widehat{\boldsymbol{\gamma}}_{\rm w} \cdot \widehat{\boldsymbol{\varepsilon}_{\rm PSF}} \right\rangle = {} & \frac{\left\langle |\boldsymbol{\varepsilon}_{\rm PSF}|^2 \right\rangle}{P_\gamma P_R P_{\boldsymbol{\varepsilon}_{\rm PSF}}} \frac{R^2_{\rm PSF}}{R^2_{\rm gal}} \left(\frac{\delta\left(R^2_{\rm PSF}\right)}{R^2_{\rm PSF}} + \frac{2\,\delta R_{\rm NC}}{R_{\rm obs} - R_{\rm NC}} \right. \\ & \left. + \frac{\delta\left(R^2_{\rm obs}\right)}{R_{\rm obs}(R_{\rm obs} - R_{\rm NC})} + \frac{P_R R^4_{\rm gal}}{R^2_{\rm PSF}\left(P_R R^2_{\rm gal} + R^2_{\rm PSF}\right)} \frac{\delta P_R}{P_R} \right) \\ & - \frac{1}{P_\gamma P_R} \frac{R^2_{\rm PSF}}{R^2_{\rm gal}} \left(\frac{\langle \boldsymbol{\delta\varepsilon}_{\rm PSF} \rangle^2 + \sigma^2[|\boldsymbol{\varepsilon}_{\rm PSF}|]}{P_{\boldsymbol{\varepsilon}_{\rm PSF}}} \right) \\ & + \langle \boldsymbol{\gamma}_{\rm w} \cdot \boldsymbol{\delta\varepsilon}_{\rm PSF} \rangle \end{aligned} \tag{76}$$

plus many more terms of order $\mathcal{O}(\delta^2)$, including some proportional to equations (A4) and (A5). While it would be difficult to identify and then calibrate out any individual contribution from this mixed observable, it can be used as an invaluable *post facto* check that other techniques have successfully removed almost all of the additive cosmic shear systematics.

5.2.2 *Calibrating residual detector effects*

Non-convolution detector effects can accumulate in space-based instruments over time, as radiation damages the hardware. Thus any long-term, monotonic drift in the mean $\langle R_{\rm obs} \rangle$ or $\langle \boldsymbol{\varepsilon}_{\rm obs} \rangle$ within each exposure – or, even better, within a calibration field that can be returned to – indicates a non-zero $\delta R_{\rm NC}$ or $\boldsymbol{\delta\varepsilon}_{\rm NC}$.

Many detector effects also exhibit a characteristic dependence upon chip position. This is most notable for charge transfer inefficiency in CCDs, where the image degradation increases linearly with distance y from the readout register (Massey et al. 2010a), where $y_{\rm max}$ is the size of the CCD. In this case, correlating shear measurements with chip position, or fitting shear measurements as a function of chip position, measures non-zero:

$$\begin{aligned}\langle \boldsymbol{\gamma}_{\rm w} \rangle|_{y_{\rm max}} = {} & \frac{1}{P_\gamma P_R P_{\boldsymbol{\varepsilon}_{\rm NC}}} \frac{P_R R^2_{\rm gal} + R^2_{\rm PSF}}{R^2_{\rm gal}} \boldsymbol{\delta\varepsilon}_{\rm NC}|_{y_{\rm max}} \\ & - \frac{2\,\langle \boldsymbol{\varepsilon}_{\rm PSF} \rangle}{P_\gamma P_R P_{\boldsymbol{\varepsilon}_{\rm PSF}}} \frac{R^2_{\rm PSF}}{R^2_{\rm gal}} \frac{\delta R_{\rm NC}}{R_{\rm obs}}\bigg|_{y_{\rm max}} \\ & - \frac{1}{P_\gamma P_R P_{\boldsymbol{\varepsilon}_{\rm PSF}}} \frac{R^2_{\rm PSF}}{R^2_{\rm gal}} \left\{ \langle \boldsymbol{\delta\varepsilon}_{\rm PSF} \rangle \right. \\ & + \langle \boldsymbol{\varepsilon}_{\rm PSF} \rangle \left(\frac{\delta\left(R^2_{\rm PSF}\right)}{R^2_{\rm PSF}} + \frac{\delta\left(R^2_{\rm obs}\right)}{R_{\rm obs}(R_{\rm obs} - R_{\rm NC})} \right. \\ & \left. \left. + \frac{P_R R^4_{\rm gal}}{R^2_{\rm PSF}(P_R R^2_{\rm gal} + R^2_{\rm PSF})} \frac{\delta P_R}{P_R} \right) \right\}, \end{aligned} \tag{77}$$

where we assume $\boldsymbol{\delta\varepsilon}_{\rm NC}|_{y_{\rm max}}$ and $\delta R_{\rm NC}|_{y_{\rm max}}$ are constant over a sufficiently long time period to gather statistically significant measurements. If $\langle \boldsymbol{\varepsilon}_{\rm PSF} \rangle = 0$ and all other (PSF, observational) errors were zero, this would be a direct test of $\boldsymbol{\delta\varepsilon}_{\rm NC}$. However, the reality that (77) contains terms mixed with residual PSF modelling errors has made analysis of *HST* data challenging. Only by first verifying

the PSF model with tests from Section 5.2.1 (Rhodes et al. 2007; Schrabback et al. 2010; Hoekstra et al. 2011) were able to subtract this measurement of $\boldsymbol{\delta\epsilon}_{\rm NC}$ from a shear catalogue, following equation (34). However, such an empirical, catalogue-level correction should be seen as a last resort because it addresses neither $\delta R_{\rm NC}$ nor the mixing between sources of error whereby an imperfect removal of additive systematics can introduces an (undiagnosable) multiplicative cosmic shear systematic. A much more robust technique, demonstrated by Leauthaud et al. (2010), is to apply a physically motivated correction scheme at the pixel level as the first process during data reduction (e.g. Anderson & Bedin 2010; Massey 2010). The performance of this technique can again be tested via equation (77), and improved by iteration.

5.3 Impact of residual multiplicative systematics

Multiplicative cosmic shear measurement biases are potentially the most troublesome, because there is no known cosmology-independent way to accurately diagnose residual multiplicative bias internally within a data set (except that it may leak weakly into a small unphysical *B*-mode signal, Vale 2006, but so do many things). Analyses must either rely upon theoretical calculations of the shear calibration, or test a measurement pipeline on simulated images that contain a known signal and rely upon the veracity of those simulations. Since multiplicative systematic errors are thus more problematic than additive errors, and because the requirements on them are similarly hard to meet, we shall investigate them more carefully.

Rather than considering galaxies all of the same size and detection S/N, we shall now consider a realistic, full population of galaxies. Some galaxies are bigger and brighter than others, and it will be easier to measure their shapes. The form of equation (55) suggests that multiplicative shape measurement biases predominantly depend upon the relative size of the PSF and the surveyed galaxies

$$m \approx m_0 + \frac{m_1}{P_R}\left(\frac{R^2_{\rm PSF}}{R^2_{\rm gal}}\right). \tag{78}$$

This characteristically quadratic performance was indeed apparent in many of the methods tested in STEP2[9] (Massey et al. 2007b, top right-hand panels of fig. 7). Similar behaviour is suggested in GREAT08 (Bridle et al. 2010, fig. C3) and is explicitly fitted in GREAT10 (Kitching et al. 2012a, appendix B5) as

$$m \approx m_0 + \alpha_{R^2_{\rm PSF}} \frac{\langle R^2_{\rm gal}\rangle}{\langle R^2_{\rm PSF}\rangle}\frac{R^2_{\rm PSF}}{\langle R^2_{\rm gal}\rangle}, \tag{79}$$

where $\langle R^2_{\rm PSF}\rangle = 3.4^2$ pixels^2, $\langle R^2_{\rm gal}\rangle \approx 3.55^2$ pixels^2 (averaging the contribution of the bulges and discs), and the best methods achieve $\alpha_{R^2_{\rm PSF}} \approx 0.005$ (Kitching et al. 2012a, fig. 5). Note that GREAT10's fiducial PSF had a Moffat profile, for which $P_R \sim 1$. Diffraction-limited surveys with $P_R \sim 2$ will likely achieve better performance although, since that was only tested in a subset of the GREAT10 data whose results were dominated by method bias, we shall conservatively assume only the performance explicitly demonstrated.

We showed in Section 4.3 that constraints on the nature of dark energy are largely insensitive to a constant multiplicative bias m_0. The achieved value of m_1 is thus likely to be the driving requirement for success. We shall baseline a currently achievable performance of $m_0 \approx 0$ and $m_1 \approx 0.006$. We shall then fold through the observed distribution of galaxies sizes to consider the prospects of two generic regimes proposed for future surveys:[10] a space-based mission with a PSF full width at half-maximum (FWHM) of 0.2 arcsec and a ground-based telescope with a FWHM seeing of 0.7 arcsec.

5.3.1 Two-dimensional cosmic shear

To quantify the typical size of galaxies in the Universe as a function of magnitude, we measure the sizes of galaxies in $i_{775\rm W}$-band observations of the *HST* Ultra Deep Field (UDF; Beckwith, Somerville & Stiavelli 2003) (Fig. 4a). To compute the approximate intrinsic size of the galaxies, we assume that their profiles are Gaussian (with a FWHM equal to their measured FWHM), and that the Advanced Camera for Surveys (ACS) PSF has a FWHM of 0.1 arcsec. Fainter galaxies are smaller (Fig. 4b) but, down to $i_{775\rm W} \lesssim 26$, most are intrinsically larger than the ACS PSF.

Many more galaxies are resolved ($R_{\rm obs} > 1.25R_{\rm PSF}$) by the hypothetical space-based mission than the hypothetical ground-based survey (Fig. 4c). Crucially, most galaxies in space-based observations are not only resolved but *very well* resolved. Following (78), this naturally leads to a better shear measurement bias (Fig. 4 d). For a full, realistic population of source galaxies in a 2D cosmic shear survey from space, current shear measurement performance satisfies requirement (74), in the absence of PSF variation or detector effects. Any subsequent improvement will provide increased margin for imperfect PSF and detector models.

Ground-based surveys face two problems. First, a greater improvement in shape measurement techniques is required for them to reach their full potential than space-based surveys (Fig. 4 d). This is simply because of the difficulty *resolving* galaxies from the ground, without even taking into account the much harder task of modelling the PSF due to a turbulent atmosphere and more variable physical conditions. Secondly, even in extremely deep images covering the entire sky, not enough galaxies are resolved ($R_{\rm obs} > 1.25R_{\rm PSF}$) to obtain statistical measurement errors on w competitive with other techniques (Fig. 4c). More galaxies could be included in an analysis by lowering the size cut,[11] for example to $R_{\rm obs} > 1.1R_{\rm PSF}$. Increasing the density of galaxies reduces statistical measurement error, but at a cost of even more rapidly increasing systematic bias, such that current methods do not meet requirements.

5.3.2 Three-dimensional cosmic shear

Three-dimensional cosmic shear analysis requires measurements of both shear and redshift for each galaxy, and for the shears to be measured without (even relative) bias as a function of redshift (Kitching et al. 2011a). To estimate this bias in a real population of galaxies, we use photometric redshift estimates for $20 < i_{775W} <$

[9] In STEP2, methods that applied an overall 'calibration factor' from analysis of independent simulated images (e.g. TS and several not plotted) appear to have achieved $\langle m\rangle \approx 0$ by adjusting m_0 such that $m(\langle R_{\rm gal}\rangle) = 0$ for galaxies of average size.

[10] A survey's effective $R_{\rm PSF}$ may be a complicated function of the system PSF at different times. Some state-of-the-art shear measurement algorithms downweight the contribution from exposures with poor seeing. This improves the effective $R_{\rm PSF}$, at a cost of decreased imaging depth.

[11] It is far more effective to add small galaxies than faint ones, especially for a ground-based survey, because faint galaxies are also so much smaller. In practice, our crude step-function cut is also usually replaced by a smoothly varying weight function (Hoekstra, Franx & Kuijken 2000); the result of this will lie between the two extremes we have considered.

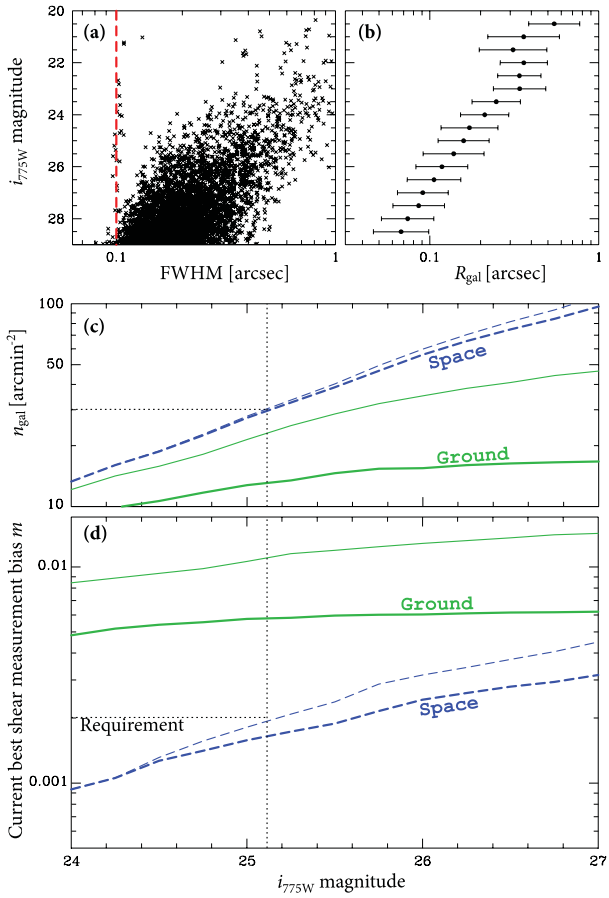

Figure 4. Prospects for 2D weak gravitational lensing surveys. Panel (a): the observed size R_{obs} and i-band magnitude of objects in the UDF. The vertical dashed line indicates the size of the ACS PSF. Panel (b): galaxies' average intrinsic size R_{gal} as a function of magnitude, under the assumption that the galaxies and PSF have Gaussian profiles. The error bars indicate the dispersion in R_{gal}. Panel (c): the cumulative number density of resolved galaxies as a function of (limiting) magnitude, with sizes $R_{obs} > 1.25R_{PSF}$ (thick lines) or $R_{obs} > 1.1R_{PSF}$ (thin lines). The dashed lines correspond to a space-based mission with a FWHM = 0.2 arcsec for the PSF. Note that *Euclid*'s wide-band observations to magnitude 24.5 correspond roughly to $i_{775W} \approx 25.2$ (vertical dotted line). The solid curves are for a typical ground-based PSF with FWHM = 0.7 arcsec. Panel (d): predicted shear measurement bias for the best current methods, averaged over the population of resolved galaxies. Requirement (74) is shown as a horizontal dotted line, assuming $\overline{\mathcal{M}} \approx 2m$ (17).

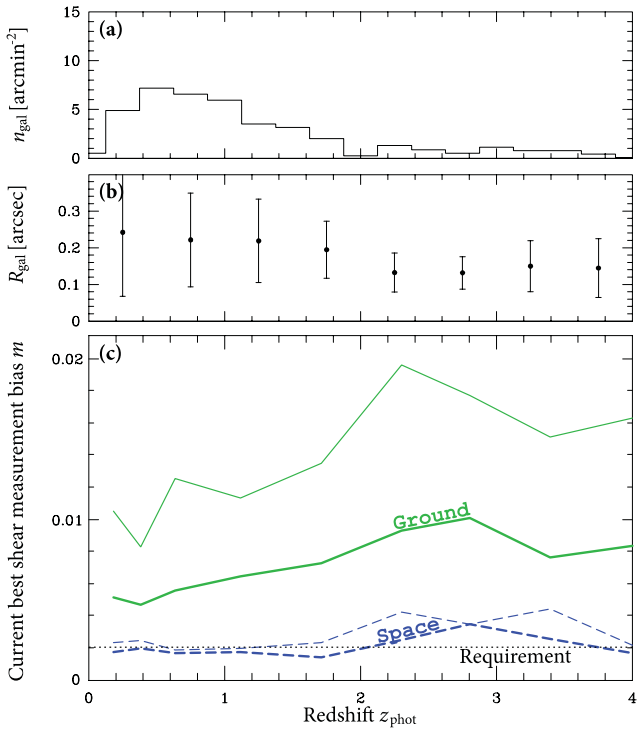

Figure 5. Prospects for 3D weak gravitational lensing surveys. Panel (a): the density of galaxies as a function of photometric redshift z_{phot} for galaxies with $20 < i_{775W} < 26.5$ (results do not depend strongly on the choice of limiting magnitude). Panel (b): average galaxy size $\langle R_{gal} \rangle$ as a function of redshift, under the assumption that the galaxies have Gaussian profiles. The error bars indicate the dispersion in R_{gal}. Panel (c): multiplicative shear calibration bias m as a function of redshift for galaxies with sizes $R_{obs} > 1.25R_{PSF}$ (thick lines) or $R_{obs} > 1.1R_{PSF}$ (thin lines). The dashed line corresponds to a space-based mission with a PSF of FHWM = 0.2 arcsec, and the solid curve is for typical ground-based seeing with FHWM = 0.7 arcsec. Requirement (74) is shown as a dotted line, assuming $\overline{\mathcal{M}} \approx 2m$ (17).

26.5 galaxies in the *HST* UDF by Coe et al. (2006). The distribution of best-fitting redshifts peaks around $z \sim 0.5$ but also samples a long tail out to $z \sim 3$ (Fig. 5a). Beyond redshift $z \sim 3$, the scarcity of UDF galaxies makes our statistics unstable.

The mean and rms apparent size of galaxies decrease noticeably above $z \sim 1.5$–2 (Fig. 5b). Multiplicative shear measurement bias will therefore get slightly worse at high redshift (Fig. 5c). For a space-based survey, meeting requirements in every redshift bin will demand algorithms with multiplicative biases a factor of 1.8–2.2 better than current methods (which could come from calibration on very accurate simulated images). Note that this analysis is completely independent of that in Section 5.1. That their conclusions are so consistent lends support to both methodologies.

Ground-based observations are more profoundly affected by the decrease in galaxy size at $z \gtrsim 1.5$. Very deep images will help, because some fraction of systematics is doubtless due to noise bias (Réfrégier et al. 2012). However, a dramatic improvement in shear measurement methods will be required for ground-based surveys to span the high redshifts needed to probe the growth of structure. As before, this argument is based purely on the small size of galaxies compared to a ground-based PSF, and does not take into account additional challenge of modelling the more variable ground-based PSF.

6 DISCUSSION AND CONCLUSIONS

We have derived expressions showing how various sources of error in galaxy shape measurement propagate into additive biases $\mathcal{A}$ (equation 57) and multiplicative biases $\mathcal{M}$ (equation 59) on cosmic shear results. Additive biases include a contribution from misestimation of a telescope's PSF shape, and multiplicative biases include misestimation of the PSF size. This agrees with the behaviour generically seen in empirical tests of shear measurement methods. For the first time, we have also propagated into cosmic shear results the consequences of imperfect correction for non-linear detector effects, and imperfect image processing algorithms.

We have ascertained the maximum level of additive biases $\mathcal{A}(\ell, z)$ (equation 72) and multiplicative biases $\mathcal{M}(\ell, z)$ (equation 74) that can be tolerated by a next generation cosmic shear survey attempting to constrain the dark energy equation of state parameter w to within 0.065 (68 per cent CL). Cosmic shear measurements of w

are surprisingly insensitive to a *constant* multiplicative bias. To explore more generic scale- and redshift-dependent systematic biases, we have used a form-filling technique; based upon the 95 per cent CL averaged equally over all possible functional forms, we define convenient requirements on mean $\overline{\mathcal{A}}$ and $\overline{\mathcal{M}}$. Cropper et al. (2012) distribute this overall requirement into budgets on the individuals sources of error (PSF knowledge, detector knowledge, accuracy of shape measurement algorithms) in an allocation that is suitable for a real space mission.

We compare our requirements on galaxy shape measurement software to the performance seen recently in the public, blind GREAT10 challenge. Extant shear measurement methods meet both requirements for a Stage IV weak lensing surveys, for bright galaxies at detection S/N = 40 or for a 2D cosmic shear survey from space in which the contributions from a large population of galaxies are combined. This will generally not provide sufficient galaxies to meet Stage IV surveys' goals for the statistical errors on cosmological parameters. This also assumes that the telescope and instrument hardware can be well modelled; a modest improvement will create margin for imperfect modelling and correction of the system PSF or detector effects.

Fully exploiting the statistical potential of Stage IV weak lensing surveys will require shear measurement software that works more accurately than current algorithms on faint galaxies. Current algorithms could introduce systematic biases of the same order of magnitude as the statistical errors, and the total reported confidence limits would need to be enlarged by a factor of $\sim\sqrt{2}$ to account for this effect. To be sure of avoiding this problem, if all galaxies were detected at S/N = 20–10 and all of them were used, additive biases must be reduced by a factor of 3.6–6.5. However, many tests can be used to identify and remove portions of a shear catalogue with additive biases; we have used our new formalism to show exactly what each test is sensitive to. Using an entire, realistic population of faint galaxies would also need a reduction in multiplicative bias by a factor of 1.4–2.8. Averaging over a realistic galaxy population extending to $z \gtrsim 1.5$, a space-based 3D cosmic shear analysis will need an improvement in multiplicative bias by a factor of 1.8–2.2. No internal tests can identify multiplicative biases, so the greatest development effort should be spent to minimize these.

Several new ideas for image analysis techniques are being discussed in the literature, and ongoing simulation programmes show potential. The past decade has seen steady improvement in shape measurement algorithms; extrapolating even minimal continued development suggests that the required algorithmic performance will be met well before the need to analyse Stage IV surveys. Importantly, it will be at least three to five times easier to meet requirements for high-resolution space-based rather than ground-based surveys, because multiplicative biases depend (theoretically and empirically) on the inverse square of the S/N and the square of the PSF size. This conclusion that ground-based surveys will require much better shear measurement methods than space-based surveys arises solely because they do not resolve galaxies well, and does not even take into account the additional challenge of modelling atmospheric turbulence or more rapidly changing physical conditions.

ACKNOWLEDGMENTS

RM and TK are supported by Royal Society University Research Fellowships. HH is supported by the Netherlands Organization for Scientific Research through VIDI grants and acknowledges support from the Netherlands Research School for Astronomy (NOVA). RM and HH also acknowledge support from ERC International Reintegration Grants. This work was done in part at JPL, run under a contract for NASA by Caltech. TS is supported by the NSF through grant AST-0444059-001, and by the Smithsonian Astrophysics Observatory through grant GO0-11147A. YM acknowledges support from CNES and CNRS/INSU.

Facilities: Hubble Space Telescope HST-GO-9978 (P.I. S. Beckwith).

REFERENCES

Amara A., Réfrégier A., 2007, MNRAS, 381, 1018
Amara A., Réfrégier A., 2008, MNRAS, 391, 228
Amara A., Réfrégier A., Paulin-Henriksson S., 2010, MNRAS, 404, 926
Anderson J., Bedin L., 2010, PASP, 122, 1035
Barron N. et al., 2007, PASP, 119, 466
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Beckwith S., Somerville R., Stiavelli M., 2003, STScI Newsletter, 20, 1
Bernstein G., 2010, MNRAS, 406, 2793
Bernstein G., Jain B., 2004, ApJ, 600, 17
Bradshaw A. et al., 2012, Am. Astron. Soc. Meeting Abstr., 219, 156.01
Bridle S. L., Kneib J.-P., Bardeau S., Gull S. F., 2002, in Natarajan P., ed., The Shapes of Galaxies and their Dark Halos, Proceedings of the Yale Cosmology Workshop, held 28-30 May 2001 in New Haven, Connecticut, USA. World Scientific Publishing Co. Pte. Ltd., p. 38
Bridle S. et al., 2009, Ann. Appl. Stat., 3, 6
Bridle S. et al., 2010, MNRAS, 405, 2044
Bristow P., 2003, Instrument Science Report CE-STIS-2003-001
Catelan P., Kamionkowski M., Blandford M., 2001, MNRAS, 320, 7
Chang C. et al., 2012, MNRAS, in press (arXiv:1206.1378)
Chevallier M., Polarski D., 2001, Int. J. Modern Phys. D, 10, 213
Coe D., Benítez N., Sánchez S. F., Jee M. J., Bouwens R., Ford H., 2006, AJ, 132, 926
Crittenden R., Natarajan P., Pen U., Theuns T., 2001, ApJ, 559, 552
Cropper M. et al., 2012, MNRAS, submitted
Cypriano E., Amara A., Voigt L., Bridle S., Abdalla F., Réfrégier A., Seiffert M., Rhodes J., 2010, MNRAS, 405, 494
Das S., de Putter R., Linder E., Nakajima R., 2012, JCAP, 11, 11
Dodelson S., Shapiro C., White M., 2006, Phys. Rev. D, 73, 3009
Heymans C. et al., 2006, MNRAS, 368, 1323
Hirata C., Seljak U., 2003, MNRAS, 343, 459
Hoekstra H., 2004, MNRAS, 347, 1337
Hoekstra H., Jain B., 2008, Annu. Rev. Nucl. Part. Sci., 58, 99
Hoekstra H., Franx M., Kuijken K., Squires G., 1998, ApJ, 504, 636
Hoekstra H., Franx M., Kuijken K., 2000, ApJ, 532, 88
Hoekstra H., Donahue M., Conselice C., McNamara B., Voit G., 2011, ApJ, 726, 48
Hook R., Stoehr F., 2008, Instrument Science Report WFC3 2008-014
Huterer D., Takada M., Bernstein G., Jain B., 2006, MNRAS, 366, 101
Hu W., 1999, ApJ, 522, L21
Ishak M., 2005, MNRAS, 363, 469
Iye M. et al., 2004, Publ. Astron. Soc. Jap., 381, 397
Jain B., Taylor A., 2003, Phys. Rev. Lett., 91, 1302
Jain B., Jarvis M., Bernstein G., 2006, J. Cosmol. Astropart. Phys., 0602, 001
Janesick J., 2001, Scientific Charge Coupled Devices. SPIE, Bellingham
Jarvis M., Jain B., 2004, ApJ, submitted (astro-ph/0412234)
Jarvis M., Schechter P., Jain B., 2008, PASP, submitted (arXiv:0810.0027)
Jee M., Tyson A., 2011, PASP, 123, 596
Jee M., Blakeslee J., Sirianni M., Martel A., White R., Ford H., 2007, PASP, 119, 1403
Jee M. et al., 2011, ApJ, 737, 59
Joachimi B., Schneider P., 2008, A&A, 488, 829
Kacprzak T., Zuntz J., Rowe B., Bridle S., Réfrégier A., Amara A., Voigt L., Hirsch M., 2012, MNRAS, 427, 4
Kaiser N., 2000, ApJ, 537, 555
Kaiser N., Squires G., Broadhurst T., 1995, ApJ, 449, 460 (KSB)
Kirk D., Rassat A., Host O., Bridle S., 2012, MNRAS, 424, 1647

Kitching T., Heavens A., Taylor A., Brown M., Meisenheimer K., Wolf C., Gray M., Bacon D., 2007, MNRAS, 376, 771
Kitching T., Miller L., Heymans C., van Waerbeke L., Heavens A., 2008, MNRAS, 390, 149
Kitching T., Amara A., Abdalla F., Joachimi B., Réfrégier A., 2009a, MNRAS, 399, 2107
Kitching T., Amara A., Rassat A., Réfrégier A., 2009b, preprint (arXiv:0901.3143)
Kitching T., Heavens A., Miller L., 2011a, MNRAS, 413, 2923
Kitching T. et al., 2011b, Ann. Appl. Stat., 5, 2231
Kitching T. et al., 2012a, MNRAS, 423, 3236
Kitching T. et al., 2012b, New Astron. Rev.
Krist J., 1995, in Shaw R. A., Payne H. E., Hayes J. J. E., eds, ASP Conf. Ser. Vol. 77, Astronomical Data Analysis Software and Systems IV. Astron. Soc. Pac., San Francisco, p. 349
Laureijs R. et al., 2011, Euclid Definition Study Report (http://sci.esa.int/science-e/www/object/index.cfm?fobjectid=48983)
Leauthaud A. et al., 2007, ApJS, 172, 219
Leauthaud A. et al., 2010, ApJ, 709, 97
Linder E., 2003, Phys. Rev. Lett., 90, 91301
Luppino G., Kaiser N., 1997, ApJ, 475, 20
McCullough P., 2008, Instrument Science Report WFC3 2008-26
Mandelbaum R. et al., 2005, MNRAS, 361, 1287
Massey R., 2010, MNRAS, 409, L109
Massey R., Réfregier A., Bacon D., Ellis R., Brown M., 2005, MNRAS, 359, 1277
Massey R. et al., 2007a, ApJS, 172, 239
Massey R. et al., 2007b, MNRAS, 376, 13
Massey R., Rowe B., Réfregier A., Bacon D., Bergé J., 2007c, MNRAS, 380, 229
Massey R., Stoughton C., Leauthaud A., Rhodes J., Koekemoer A., Ellis R., Shaghoulian E., 2010a, MNRAS, 401, 371
Massey R., Kitching T., Richard J., 2010b, Rep. Progress Phys., 73, 086901
Melchior P., Viola M., 2012, MNRAS, 424, 2757
Melchior P., Viola M., Schafer B., Bartelmann M., 2011, MNRAS, 412, 1552
Miller L., Kitching T. D., Heymans C., Heavens A. F., van Waerbeke L., 2007, MNRAS, 382, 315
Miralles J.-M., Erben T., Hämmerle H., Schneider P., Freudling W., Pirzkal N., Fosbury R., 2005, A&A
Natarajan P., Crittenden R., Pen U.-L., Theuns T., 2001, Publ. Astron. Soc. Aust., 18, 19
Paulin-Henriksson S., Amara A., Voigt L., Réfrégier A., Bridle S., 2008, A&A, 484, 67
Paulin-Henriksson S., Réfrégier A., Amara A., 2009, A&A, 500, 647
Plazas A., Bernstein J., 2012, PASP, in press
Réfrégier A., 2003, MNRAS, 338, 35
Réfrégier A., Amara A., Kitching T., Rassat A., 2011, A&A, 528, 33
Réfrégier A., Kacprzak T., Amara A., Bridle S., Rowe B., 2012, MNRAS, 425, 3
Rhodes J., Réfrégier A., Groth E. J., 2000, ApJ, 536, 79 (RRG)
Rhodes J., Réfrégier A., Collins N., Gardner J., Groth E., Hill R., 2004, ApJ, 605, 29
Rhodes J. et al., 2007, ApJS, 172, 203
Rhodes J., Leauthaud A., Stoughton C., Massey R., Dawson K., Kolbe W., Roe N., 2010, PASP, 122, 439
Rhodes J. et al., 2012, Astropart. Phys., 38, 31 (arXiv:1205.2957)
Schneider M., Bridle S., 2010, MNRAS, 402, 2127
Schrabback T. et al., 2010, A&A, 516, 63
Semboloni E., Hoekstra H., Schaye J., van Daalen M., McCarthy I., 2011, MNRAS, 417, 2020
Semboloni E. et al., 2012, MNRAS, submitted (arXiv:1211.5025)
Seshadri S., Cole D., Hancock B., Smith R., 2008, Proc. SPIE, 7021, p. 3
Shapiro C., 2009, ApJ, 696, 775
Shapiro C., Dodelson S., Hoyle B., Samushia L., Flaugher B., 2010, Phys. Rev. D, 82, 3520
Simpson F., Bridle S., 2005, Phys. Rev. D, 71, 3501
Song Y., Knox L., 2004, Phys. Rev. D, 70, 3510
Taylor A., Watts P., 2001, MNRAS, 328, 1027
Vale C., 2006, PhD thesis, Univ. California Berkeley
Viola M., Melchior P., Bartelmann M., 2011, MNRAS, 410, 2156
Voigt L., Bridle S., 2010, MNRAS, 404, 458
Voigt L., Bridle S., Amara A., Cropper M., Kitching T., Massey R., Rhodes J., Schrabback T., 2012, MNRAS, 421, 1385

APPENDIX A: REMAINING CROSS TERMS

In Section 3, we ignored several cross terms in earlier calculations of the additive cosmic shear systematic $\mathcal{A}$ because we expect their contributions to be subdominant as long as the PSF model, detector characterization and shape measurement method are working properly. However, tests for the presence of these terms in real data could be a useful, cosmology-independent way to verify that the pipeline is meeting requirements. We shall now discuss four noteworthy order $\mathcal{O}(\delta)$ terms that potentially add to $\mathcal{A}$. These are

$$-\left\langle \boldsymbol{\epsilon}_{\rm gal}\cdot\boldsymbol{\epsilon}_{\rm PSF}\right\rangle \left\langle \frac{R^2_{\rm PSF}}{R^2_{\rm gal}}\right\rangle \times\left(\frac{\delta\left(R^2_{\rm PSF}\right)}{R^2_{\rm PSF}}+\frac{2\delta R_{\rm NC}}{R_{\rm obs}-R_{\rm NC}}+\frac{\delta\left(R^2_{\rm obs}\right)}{R^2_{\rm gal}+R^2_{\rm PSF}}\right) \tag{A1}$$

in the presence of the selection bias discussed by Hirata & Seljak (2003), whereby galaxies are more likely to be detected if their intrinsic shapes are similar to that of the PSF;

$$-\left\langle \boldsymbol{\epsilon}_{\rm gal}\cdot\boldsymbol{\delta\epsilon}_{\rm obs}\right\rangle \left\langle \frac{R^2_{\rm gal}+R^2_{\rm PSF}}{R^2_{\rm PSF}}\right\rangle \tag{A2}$$

if, for example, a faulty shape measurement method systematically truncates the isophotes of elliptical galaxies and

$$+\left\langle \boldsymbol{\epsilon}_{\rm gal}\cdot\boldsymbol{\delta\epsilon}_{\rm NC}\right\rangle \left\langle \frac{R^2_{\rm gal}+R^2_{\rm PSF}}{R^2_{\rm PSF}}\right\rangle \tag{A3}$$

with charge transfer inefficiency, for which $\boldsymbol{\delta\epsilon}_{\rm NC}$ depends on $\boldsymbol{\epsilon}_{\rm gal}$ (Rhodes et al. 2010); and

$$-\left\langle \boldsymbol{\epsilon}_{\rm gal}\cdot\boldsymbol{\delta\epsilon}_{\rm PSF}\right\rangle \left\langle \frac{R^2_{\rm PSF}}{R^2_{\rm gal}}\right\rangle \tag{A4}$$

if some small galaxies (which have been sheared, so correlate with their neighbours) are accidentally confused with stars and allowed to contribute towards the PSF model. Of all these, the first two terms of (A1) are likely to be the most problematic: the first because stars and galaxies have different colours, so a PSF model naïvely obtained from stars will be systematically too large, and the secondly because model inaccuracies in non-linear correction will likely dominate variations in the effect across the detector.

There are also several terms of order $\mathcal{O}(\delta^2)$. Two that may feasibly have non-zero coefficients are

$$+\left\langle \boldsymbol{\delta\epsilon}_{\rm PSF}\cdot\frac{\delta R^2_{\rm PSF}}{R^2_{\rm PSF}}\boldsymbol{\epsilon}_{\rm PSF}\right\rangle \left\langle \frac{R^4_{\rm PSF}}{R^4_{\rm gal}}\right\rangle \tag{A5}$$

if the PSF modelling errors depend upon the ellipticity of a complex PSF whose shape changes as a function of radius; and

$$+\left\langle \boldsymbol{\delta\epsilon}_{\rm PSF}\cdot\boldsymbol{\delta\epsilon}_{\rm NC}\right\rangle \left\langle \frac{R^2_{\rm PSF}\left(R^2_{\rm PSF}+R^2_{\rm gal}\right)}{R^4_{\rm gal}}\right\rangle \tag{A6}$$

if residuals from the correction of non-linear detector effects also contaminate the bright stars from which the PSF is modelled.

Finally, we also ignored cross terms like $\langle m\boldsymbol{\gamma}\boldsymbol{c}\rangle$ in the correlation functions. Ideally, $\langle m\boldsymbol{\gamma}\boldsymbol{c}\rangle = \langle m\boldsymbol{c}\rangle\,\langle\boldsymbol{\gamma}\rangle$ and $\langle\boldsymbol{\gamma}\rangle = 0$, but this latter equality does not hold in the presence of Hirata & Seljak (2003) selection biases. Furthermore, we have shown that m and $\boldsymbol{c}$ are both correlated with $\delta R_{\rm PSF}$ and therefore with each other, so the prefactor may be considerable. This sort of combination could give rise to a whole new slew of potential intrinsic-intrinsic, intrinsic-$\boldsymbol{c}$, intrinsic-m, etc. systematics. We shall explore these in future work.

APPENDIX B: PERFORMANCE INDICATORS USED IN GREAT10

In equations (70) and (71), we introduced performance indicators $\overline{\mathcal{A}}$ and $\overline{\mathcal{M}}$, based upon integrals over a range of scales. For consistency with earlier work (Amara & Réfrégier 2008), we chose to weight the scales by $\ell^2 \mathrm{d}\ln\ell$, but different choices could have been made. Integration with respect to $\mathrm{d}\ell$ typically raises the numerical value of $\overline{\mathcal{A}}$ by $\sim$10 per cent and $\mathcal{M}$ by $\sim$3 per cent. A similar loosening would also need to be applied to the numerical value of the requirements, and this is a negligible change. Including weighting by $C(\ell)\,\mathrm{d}\ell$ inside the integrals in (71) rescales the performance indicator and requirement so that they have a numerical value similar to $\overline{\mathcal{A}}$. However, this would mean losing intuition from previous studies and make the requirements formally cosmology dependent. Furthermore, since the shape of the $C(\ell)$ weight approximately recovers that of $\ell^2 \mathrm{d}\ln\ell$, changes to numerical values are even smaller than the previous option.

Practical considerations forced the measurements in GREAT10 (Kitching et al. 2012a) to use a different range in ℓ and a different weight function. It is important to consider the effect of this, because we use the GREAT10 results as an indication of current best performance. The GREAT10 analysis measured $\widehat{C}(\ell)$ at linearly separated values $\ell = \{233, 415, 600, 789, 977, 1162, 1350, 1538\}$, then found the least-squares fitting function $(1+\mathcal{M})C(\ell)+\mathcal{A}$ with constant $\mathcal{A} = \overline{\mathcal{A}}_{\rm G}$ and $\mathcal{M} = \overline{\mathcal{M}}_{\rm G}$. This process thus minimizes

$$\chi^2(\mathcal{A},\mathcal{M}) \equiv \sum_\ell (\widehat{C}(\ell) - (1+\mathcal{M})C(\ell) - \mathcal{A})^2. \tag{B1}$$

Therefore

$$\frac{\partial\chi^2}{\partial\mathcal{A}} = 2\sum_\ell (\widehat{C}(\ell) - (1+\mathcal{M})C(\ell) - \mathcal{A}) = 0 \tag{B2}$$

so, if $\mathcal{M} = 0$,

$$\sum_\ell \mathcal{A} = \sum_\ell (\widehat{C}(\ell) - C(\ell)). \tag{B3}$$

Approximating the discrete sums with constant $\Delta\ell$ as continuous integrals, and remembering that $\mathcal{A} = \overline{\mathcal{A}}_{\rm G}$ is constant so can be extracted from the integrals,

$$\overline{\mathcal{A}}_{\rm G} = \frac{\frac{1}{2\pi}\int_{\ell_{\rm min}}^{\ell_{\rm max}} (\widehat{C}(\ell) - C(\ell))\mathrm{d}\ell}{\frac{1}{2\pi}\int_{\ell_{\rm min}}^{\ell_{\rm max}} \mathrm{d}\ell}. \tag{B4}$$

This is similar to equation (70), although a version in which the various ℓ scales are weighted differently. The different weighting changes our conclusions by less than 10 per cent, so we ignore this small perturbation.

Least-squares fitting also guarantees that

$$\frac{\partial\chi^2}{\partial\mathcal{M}} = -2\sum_\ell C(\ell)(\widehat{C}(\ell) - (1+\mathcal{M})C(\ell) - \mathcal{A}) = 0 \tag{B5}$$

so, if $\mathcal{A} = 0$,

$$\overline{\mathcal{M}}_{\rm G} = \frac{\sum C(\ell)(\widehat{C}(\ell) - C(\ell))}{\sum (C(\ell))^2} \tag{B6}$$

$$= \frac{\frac{1}{2\pi}\int_{\ell_{\rm min}}^{\ell_{\rm max}} C(\ell)(\widehat{C}(\ell) - C(\ell))\,\mathrm{d}\ell}{\frac{1}{2\pi}\int_{\ell_{\rm min}}^{\ell_{\rm max}} (C(\ell))^2\,\mathrm{d}\ell}. \tag{B7}$$

This again is merely a differently weighted version of equation (71), with negligible effect upon our conclusions.

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