Flexion measurement in simulations of Hubble Space Telescope data

Barnaby Rowe, David Bacon, Richard Massey, Catherine Heymans, Boris Häussler, Andy Taylor, Jason Rhodes, Yannick Mellier

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **435,** 822–844 (2013)
Advance Access publication 2013 August 19

doi:10.1093/mnras/stt1353

Flexion measurement in simulations of *Hubble Space Telescope* data

Barnaby Rowe,[1,2,3,4]★ David Bacon,[5] Richard Massey,[6] Catherine Heymans,[7] Boris Häußler,[8] Andy Taylor,[7] Jason Rhodes[2,3] and Yannick Mellier[4]

[1]*Department of Physics and Astronomy, University College London, Gower Place, London WC1E 6BT, UK*
[2]*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA*
[3]*California Institute of Technology, 1200 East California Boulevard, Pasadena CA 91106, USA*
[4]*Institut d'Astrophysique de Paris, UMR7095 CNRS, Université Pierre et Marie Curie – Paris 6, 98 bis, Boulevard Arago, F-75014 Paris, France*
[5]*Institute of Cosmology and Gravitation, University of Portsmouth, Mercantile House, Hampshire Terrace, Portsmouth, Hants PO1 2EG, UK*
[6]*Institute for Computational Cosmology, Durham University, South Road, Durham DH1 3LE, UK*
[7]*SUPA†, Institute for Astronomy, University of Edinburgh, The Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK*
[8]*School of Physics and Astronomy, The University of Nottingham, University Park, Nottingham NG7 2RD, UK*

Accepted 2013 July 19. Received 2013 July 17; in original form 2012 November 5

ABSTRACT

We present a simulation analysis of weak gravitational lensing flexion and shear measurement using shapelet decomposition, and identify differences between flexion and shear measurement noise in deep survey data. Taking models of galaxies from the *Hubble Space Telescope* Ultra Deep Field (HUDF) and applying a correction for the HUDF point spread function, we generate lensed simulations of deep, optical imaging data from *Hubble's* Advanced Camera for Surveys, with realistic galaxy morphologies. We find that flexion and shear estimates differ in our measurement pipeline: whereas intrinsic galaxy shape is typically the dominant contribution to noise in shear estimates, pixel noise due to finite photon counts and detector read noise is a major contributor to uncertainty in flexion estimates, across a broad range of galaxy signal-to-noise. This pixel noise also increases more rapidly as galaxy signal-to-noise decreases than is found for shear estimates. We provide simple power-law fitting functions for this behaviour, for both flexion and shear, allowing the effect to be properly accounted for in future forecasts for flexion measurement. Using the simulations, we also quantify the systematic biases of our shapelet flexion and shear measurement pipeline for deep *Hubble* data sets such as Galaxy Evolution from Morphology and SEDs, Space Telescope A901/902 Galaxy Evolution Survey or Cosmic Evolution Survey. Flexion measurement biases are found to be significant but consistent with previous studies.

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

1 INTRODUCTION

The study of weak gravitational lensing has progressed to become one of the most important techniques in observational cosmology (see e.g. Schneider 2006; Hoekstra & Jain 2008 for a review). As it does not depend upon the microscopic composition of the mass by which it is caused, gravitational lensing requires no assumptions to be made regarding baryon-dark matter physics. It can thus be used to make direct observations of the matter distribution on large scales (e.g. Hoekstra et al. 2006; Benjamin et al. 2007; Massey et al. 2007c; Fu et al. 2008; Schrabback et al. 2010; Huff et al. 2011; Heymans et al. 2012b, 2013; Benjamin et al. 2013; Jee et al. 2013; Kilbinger et al. 2013; Van Waerbeke et al. 2013).

Weak lensing studies have typically measured the small but coherent distortions in the ellipticities of distant galaxies, due to the shear γ, and have used these measurements to constrain the distribution of the intervening matter field. The weak lensing description has been extended to higher order via the flexion formalism (see Goldberg & Bacon 2005; Irwin & Shmakova 2005; Bacon et al. 2006, hereafter B06; Irwin & Shmakova 2006; Irwin, Shmakova & Anderson 2007), which describes the slight arcing of galaxy shapes. Despite being a weaker effect than shear, it has been hoped that the flexion signal may yet be profitably measured – galaxies typically display less intrinsic curvature than intrinsic ellipticity, and so the contribution to noise from intrinsic shape is reduced for measurements of flexion. Studies have suggested (e.g. B06; Leonard et al. 2007; Okura, Umetsu & Futamase 2007, 2008; Leonard, King & Goldberg 2011) that flexion measurements can provide extra constraints upon galaxy–galaxy lensing results and cluster mass reconstructions.

★ E-mail: browe@star.ucl.ac.uk
† Scottish Universities Physics Alliance.

For all weak lensing analyses, the correct treatment of systematic errors is vital: image distortions and shape bias due to convolution with an anisotropic point spread function (PSF), as possessed by all optical instruments, are commonly an order of magnitude greater than the gravitational signal we wish to measure. The optimal, unbiased estimation of weak lensing signals from real data is the subject of much ongoing research, involving a variety of different approaches towards the accurate inference about galaxy shapes, accounting for telescope optics, detector effects and noise. Many of the current methods used to correct for the effects of the PSF are based on the scheme proposed by Kaiser, Squires & Broadhurst (1995), Luppino & Kaiser (1997) and Hoekstra et al. (1998), commonly referred to as KSB or KSB+. The use of these techniques has proved to be both successful and widespread, to date.

Despite its practical success, there are certain elements of the KSB treatment that are unsatisfactory: Kaiser (2000) provides one compelling discussion of these potential limitations. This has prompted efforts to develop alternative weak lensing methods (e.g. Kaiser 2000; Rhodes, Refregier & Groth 2000; Bernstein & Jarvis 2002; Refregier 2003; Refregier & Bacon 2003; Hirata et al. 2004; Massey & Refregier 2005; Kuijken 2006; Melchior, Meneghetti & Bartelmann 2007; Nakajima & Bernstein 2007; Miller et al. 2007, 2013; Kitching et al. 2008; Bernstein 2010; Melchior et al. 2011; Viola, Melchior & Bartelmann 2011; Kacprzak et al. 2012; Zuntz et al. 2013). The Shear TEsting Program (STEP; see Heymans et al. 2006; Massey et al. 2007a) and GRavitational lEnsing Accuracy Testing (GREAT) challenge series (Bridle et al. 2009, 2010; Kitching et al. 2011, 2012) have compared a wide range of current weak shear estimation methods, using blind-tests on simulated lensing data.

The *shapelet* approach, proposed by Bernstein & Jarvis (2002) and Refregier (2003), is one such alternative to KSB methods. Shapelets expresses galaxy images as a sum of simple basis functions – Gauss–Laguerre or Gauss–Hermite polynomials – that behave well under deconvolution with a modelled PSF. In addition, the first method for the practical estimation of the flexion signal was built within the shapelets framework (Goldberg & Bacon 2005; B06). Further work by Massey et al. (2007b), hereafter referred to as M07, investigated shear and flexion measurement within the *polar* shapelets formalism of Massey & Refregier (2005); results suggested that polar shapelets provided an apparently natural framework for estimating both quantities. Velander, Kuijken & Schrabback (2011) used shapelets (in an implementation closely related to that of Kuijken 2006) to constrain flexion in *Hubble Space Telescope* (*HST*) data, although following a somewhat different modelling strategy to that suggested in M07. An alternative method for measuring flexion, referred to as Higher Order Lensing Image Characteristics (HOLICS) has also been suggested, and indeed employed in cluster modelling from real data (Goldberg & Leonard 2007; Okura et al. 2007, 2008; Leonard et al. 2007, 2011). Showing signs that it may provide less noisy measurements than shapelets (Leonard et al. 2007, 2011), HOLICS is conceptually a generalization of KSB methods to higher order image moments. The correction for an anisotropic PSF within HOLICS is, however, of significant complexity (Okura et al. 2008).

In this paper, we present an analysis of simulations of space-based lensing data, such as that taken using the *HST* Advanced Camera for Surveys (ACS; see, e.g. Hartig et al. 2003; Rhodes et al. 2007). Several wide-area imaging surveys that may be used for weak lensing exist for this instrument, including the Galaxy Evolution from Morphology & SEDS survey (GEMS; see, e.g. Rix et al. 2004; Heymans et al. 2005), the Cosmic Evolution Survey (COSMOS; see, e.g. Scoville et al. 2007; Leauthaud et al. 2007) and the Space Telescope A901/902 Galaxy Evolution Survey (STAGES; see, e.g. Gray et al. 2009). The imaging data from these surveys share important characteristics as regards lensing measurement, having a small but non-Gaussian PSF and significant correlation in the noise between pixels due to the dithering and drizzling strategies employed (see e.g. Leauthaud et al. 2007). In addition, there exist a wealth of galaxy cluster imaging data in the ACS archive that are also of interest for shear and flexion lensing analysis. Mindful of these factors, we construct realistic simulations of ACS lensing data using real sky noise taken from blank areas in the GEMS survey data, and with a PSF that matches the radial profile of the GEMS PSF. Within these simulations we apply known input shears and flexions, and use the resulting measurements to calculate the necessary calibration for the shapelet measurement of shear and flexion from space.

Our paper is organized as follows. Sections 2 and 3 begin with a brief description of the flexion formalism and a summary of our adopted shapelet measurement method. In Section 4, we describe how we make shapelet models of *Hubble* Ultra Deep Field (HUDF) data, including both stars and galaxies. These are used to generate simulations of lensed ACS data, which we describe in Section 5. In Section 6, we test how well we can measure flexion on this simulated data. Finally, we discuss our findings in Section 7.

2 WEAK SHEAR AND FLEXION FORMALISM

To begin, we review the flexion formalism developed by Goldberg & Bacon (2005) and B06, examining briefly how weak flexion is defined in relation to weak shear. We restrict the discussion to lensing measurements in the weak regime, so we do not consider the *reduced* shear or flexion (see Schneider & Seitz 1995; Schneider & Er 2008).

Gravitational lensing conserves surface brightness, so the effects of lensing may be described in terms of coordinate transformations between the lensed and unlensed sky coordinate plane. In general, the relationship between these coordinates is non-linear and is described by the lens equation (e.g. Bartelmann & Schneider 2001; Schneider 2006).

However, if we may assume that changes in the lens properties of a system occur only on angular scales that are large compared to the angular size of the lensed light source, then the lens equation may be locally linearized. In what follows, lensed and unlensed coordinates are denoted by $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$, respectively, and we define position relative to a galaxy centroid as $\boldsymbol{\Delta\theta} = \boldsymbol{\theta} - \boldsymbol{\theta}_c$ and $\boldsymbol{\Delta\theta}' = \boldsymbol{\theta}' - \boldsymbol{\theta}'_c$, where $\boldsymbol{\theta}_c$ and $\boldsymbol{\theta}'_c$ are the coordinate centroids of the galaxy in the image and source plane, respectively. Approximately linearizing the lens equation around this centroid, we write

$$\Delta\theta'_i \simeq A_{ij}\Delta\theta_j, \tag{1}$$

where A_{ij} is the Jacobian matrix of the transformation given by the lens equation. This matrix may be written as

$$A_{ij}(\boldsymbol{\theta}) \equiv \frac{\partial\boldsymbol{\theta}'_i}{\partial\boldsymbol{\theta}_j} = \delta_{ij} - \partial_i\partial_j\psi(\boldsymbol{\theta}) = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix}, \tag{2}$$

where $\psi(\theta)$ is the lensing potential, a two-dimensional projection of the gravitational potential along the line of sight (e.g. Schneider 2006). These equations define the two components of weak shear γ_1 and γ_2, and the convergence κ, which is a measure of the projected matter density.

However, the assumption of gradual variation in lens properties across the sky is not always justified, especially in very dense regions or those with significant dark matter substructure along the line of sight. In these situations, the lens transformation is more accurately described by

$$\Delta\theta_i' \simeq A_{ij}\Delta\theta_j + \frac{1}{2}D_{ijk}\Delta\theta_j\Delta\theta_k, \tag{3}$$

which is simply the expansion of (1) to second order, where

$$D_{ijk} \equiv \frac{\partial^2\theta_i'}{\partial\theta_j\partial\theta_k} = \partial_k A_{ij} = -\partial_i\partial_j\partial_k\psi. \tag{4}$$

Equation (3) describes the lensing distortions known as *flexion*, which skew galaxy light distributions and lead to weak lensing arcs, and which may be described upon an image $I(\boldsymbol{\theta})$ using the conservation of surface brightness under lensing

$$I(\boldsymbol{\theta}) = I^{(s)}(\boldsymbol{\theta}') = I^{(s)}(\boldsymbol{\theta}_c' + \boldsymbol{\Delta\theta}'). \tag{5}$$

Employing equation (3) in this surface brightness transformation takes weak lensing one order closer to the fully generalized non-linear treatment. By expanding the surface brightness as a Taylor series and substituting (3), Goldberg & Bacon (2005) showed that the lensed surface brightness of a galaxy may be approximated as

$$I(\boldsymbol{\theta}) \simeq I^{(s)}(\boldsymbol{\theta}) + \left[(A-I)_{ij}\Delta\theta_j + \frac{1}{2}D_{ijk}\Delta\theta_j\Delta\theta_k\right]\partial_i I^{(s)}(\boldsymbol{\theta}), \tag{6}$$

an expression which is useful in the construction of weak shear and flexion estimators using shapelets (Refregier 2003; B06; M07).

For its clarity and convenience, we will often employ the complex notation introduced in B06. The complex gradient operator on the sky plane is defined as

$$\partial \equiv \partial_1 + \mathrm{i}\partial_2. \tag{7}$$

It is shown in B06 that using this notation, the convergence and complex shear $\gamma = \gamma_1 + \mathrm{i}\gamma_2$ may be written as

$$\kappa = \frac{1}{2}\partial^*\partial\psi, \tag{8}$$

$$\gamma = \frac{1}{2}\partial\partial\psi = |\gamma|\mathrm{e}^{2\mathrm{i}\phi}, \tag{9}$$

which neatly encapsulates the spin-2 rotational symmetry properties of the shear. Taking a further complex gradient, we may define two more complex fields:

$$\mathcal{F} = \frac{1}{2}\partial^*\partial\partial\psi = |\mathcal{F}|\mathrm{e}^{\mathrm{i}\phi}, \tag{10}$$

$$\mathcal{G} = \frac{1}{2}\partial\partial\partial\psi = |\mathcal{G}|\mathrm{e}^{3\mathrm{i}\phi}, \tag{11}$$

referred to as the first flexion (spin-1) and second flexion (spin-3), respectively. Using equation (4) we may write D_{ijk} entirely in terms of the components of $\mathcal{F}$ and $\mathcal{G}$, as follows:

$$D_{ij1} = -\frac{1}{2}\begin{pmatrix} 3\mathcal{F}_1 + \mathcal{G}_1 & \mathcal{F}_2 + \mathcal{G}_2 \\ \mathcal{F}_2 + \mathcal{G}_2 & \mathcal{F}_1 - \mathcal{G}_1 \end{pmatrix},$$

$$D_{ij2} = -\frac{1}{2}\begin{pmatrix} \mathcal{F}_2 + \mathcal{G}_2 & \mathcal{F}_1 - \mathcal{G}_1 \\ \mathcal{F}_1 - \mathcal{G}_1 & 3\mathcal{F}_2 - \mathcal{G}_2 \end{pmatrix}. \tag{12}$$

Equations (2), (6) and (12) allow the formulation of practical estimators for γ, $\mathcal{F}$ and $\mathcal{G}$. In the following section, we outline how shapelets may be used to make estimators of shear and flexion.

3 SHAPELET MODELLING

3.1 Shapelet basis sets

The underlying concept of the shapelet approach, as introduced by Refregier (2003) and Bernstein & Jarvis (2002), is the expression of an object's surface brightness as a sum of orthonormal, two-dimensional basis functions:

$$f(\boldsymbol{\theta}) = \sum_{n_1=0}^{\infty}\sum_{n_2=0}^{\infty} f_{n_1,n_2} B_{n_1,n_2}(\boldsymbol{\theta};\beta). \tag{13}$$

The choice of basis functions is free in general, but the Cartesian shapelet basis set is defined by the basis function

$$B_{n_1,n_2}(\boldsymbol{\theta};\beta) = \frac{H_{n_1}(\theta_1/\beta)H_{n_2}(\theta_2/\beta)\mathrm{e}^{-|\theta|^2/2\beta^2}}{2^{(n_1n_2)}\beta\sqrt{\pi n_1 n_2}}, \tag{14}$$

where $H_{n_i}(x)$ is a Hermite polynomial of order n_i, and the important free quantity β is the scale size of the shapelet basis set. We refer to the sum of the two parameters n_1 and n_2 as the order of the shapelet basis function, and will generally truncate shapelet models to some limiting order $n_{\max}$ such that $n_1 + n_2 \leq n_{\max}$.

The formalism of polar shapelets, introduced by Massey & Refregier (2005), is closely related to that of Cartesian shapelets. Instead of the basis set defined by equations (13) and (14), polar shapelets express the object surface brightness $f(\boldsymbol{\theta})$ as

$$f(\boldsymbol{\theta}) = f(\theta,\phi) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} f_{n,m} P_{n,m}(\theta,\phi;\beta), \tag{15}$$

where θ is the modulus of the complex sky position vector $\theta_1 + \mathrm{i}\theta_2$, and $\phi = \arctan(\theta_2/\theta_1)$. The polar shapelet basis functions, which we label $P_{n,m}$, are defined by Massey & Refregier (2005) as

$$P_{n,m}(\theta,\phi;\beta) = \frac{(-1)^{(n-|m|)/2}}{\beta^{|m|+1}}\left\{\frac{[(n-|m|)/2]!}{\pi[(n-|m|)/2]!}\right\}^{1/2} \times \theta^{|m|}L^{|m|}_{(n-|m|)/2}\left(\frac{\theta^2}{\beta^2}\right)\mathrm{e}^{-\theta^2/2\beta^2}\mathrm{e}^{-\mathrm{i}m\phi} \tag{16}$$

using the following definition of the associated Laguerre polynomials (see e.g. Arfken & Weber 2005):

$$L_p^q(x) \equiv \frac{x^{-q}\mathrm{e}^x}{p!}\frac{\mathrm{d}^p}{\mathrm{d}x^p}\left(x^{p+q}\mathrm{e}^{-x}\right). \tag{17}$$

Each separate member of the basis set is uniquely described using the two integers n and m, with $n > 0$ and $|m| \leq n$.

Both the Cartesian and polar shapelet basis set have relatively simple behaviour under convolution (Refregier 2003) and deconvolution (Refregier & Bacon 2003); this makes them particularly suited for correcting images for the effects of an instrumental PSF. We now describe the methodology required for this correction.

3.2 Image deconvolution using shapelets

Within the shapelet framework, there are two possible methods with which to correct galaxy images for the effects of the optical system. Both approaches begin with the construction of a shapelet model of the PSF $g(\boldsymbol{\theta})$ at the location of each galaxy: this model should be as accurate as possible and will include any variation of the PSF across the image plane of the instrument (e.g. Hoekstra 2004; Jarvis & Jain 2004; Rowe 2010; Heymans et al. 2012a; Kitching et al. 2013). The model may also need to include some treatment of time-dependent effects (see e.g. Heymans et al. 2005; Rhodes et al. 2007; Schrabback et al. 2007).

The deconvolution method used in this work is that proposed by Massey & Refregier (2005), which is implemented within the SHAPELETS software package made available by these authors.[1] In this approach, the deconvolved shapelet coefficients f_{n_1,n_2} are estimated by 'forward' convolving the shapelet basis functions with the PSF model in advance, creating a new basis set which we label

$$D_{n_1,n_2}(\boldsymbol{\theta};\beta) = g(\boldsymbol{\theta}) * B_{n_1,n_2}(\boldsymbol{\theta};\beta), \tag{18}$$

with an equivalent expression for the case of the polar shapelet basis functions $P_{n,m}(\boldsymbol{\theta};\beta)$. Fitting the data $h(\boldsymbol{\theta})$ with this new basis set D_{n_1,n_2}, one returns a deconvolved shapelet model as follows:

$$\begin{aligned} h(\boldsymbol{\theta}) &= g(\boldsymbol{\theta}) * f(\boldsymbol{\theta}) \\ &= g(\boldsymbol{\theta}) * \left[\sum_{n_1=0}^{\infty}\sum_{n_2=0}^{\infty} f_{n_1,n_2} B_{n_1,n_2}(\boldsymbol{\theta};\beta)\right] \\ &= \sum_{n_1=0}^{\infty}\sum_{n_2=0}^{\infty} f_{n_1,n_2}\left[g(\boldsymbol{\theta}) * B_{n_1,n_2}(\boldsymbol{\theta};\beta)\right] \\ &= \sum_{n_1=0}^{\infty}\sum_{n_2=0}^{\infty} f_{n_1,n_2} D_{n_1,n_2}(\boldsymbol{\theta};\beta). \end{aligned} \tag{19}$$

As can be seen by comparison with equation (13), the returned shapelet coefficients f_{n_1,n_2} will reconstruct the deconvolved image when they are used with the original basis set $B_{n_1,n_2}(\boldsymbol{\theta};\beta)$.

There are obvious caveats to this approach, particularly that the convolved basis set $D_{n_1,n_2}(\boldsymbol{\theta};\beta)$ will in general no longer be orthogonal. However, errors due to this fact are small wherever the scale size of the galaxy image is larger than that of the PSF (Massey & Refregier 2005). The alternative shapelet deconvolution approach is that described by Refregier & Bacon (2003) and developed in some detail by Melchior et al. (2009), involving the inversion of a 'PSF matrix' that describes the transformation between the shapelet model coefficients of f and h. This PSF matrix is large and may be sparse, even despite efficient truncation (Refregier 2003; Refregier & Bacon 2003), causing the latter authors and Massey & Refregier (2005) to argue against its inversion as a slow and potentially unstable process. However, Melchior et al. (2009) find this not to be necessarily the case for sufficiently small and lightly structured PSF models, and so argue in favour of a modified inversion scheme for computational efficiency.

For ACS data, we find (see Appendix A3) that a simple, low $n_{\max}$ model is insufficient to describe the PSF. This makes matrix inversion schemes problematic, and thus for this *HST*-focused analysis we will test a shape measurement pipeline based upon the software of Massey & Refregier (2005), using the forward deconvolution method described above.

The estimation of deconvolved galaxy images using shapelet modelling is important at two stages in this paper. In Section 6, shapelet models will provide estimates of the weak lensing signal in samples of simulated images. First, however, the technique is used to estimate the deconvolved shapes of a sample of galaxies from which simulated images will be created. Shapelet models are convenient for creating such simulations: they are trivially rotated and inverted, and as they provide an estimate for the full surface brightness profile of a galaxy they allow a full range of distortions (including flexion) to be simulated. In the following section, we describe the construction of a shapelet galaxy weak lensing simulation using a set of deep ACS galaxy images.

[1] http://www.astro.caltech.edu/~rjm/shapelets/

4 SHAPELET MODELS OF THE HUDF

In order to simulate lensing data, we require a set of real galaxy images which are used in turn to create a set of galaxy shapelet models which we will refer to as the 'starter set'. The real images are obtained from the publicly available *Hubble* Ultra Deep Field (HUDF; see Beckwith et al. 2006 for a detailed description), a multicolour galaxy survey image composed of a single ACS pointing, with a 10^6 s total integration time over four broad spectral bands: $F435W$ (B_{435}), $V606W$ (V_{606}), $F775W$ (i_{775}) and $F850LP$ (z_{850}). The highest redshift objects are only visible in the i_{775} and z_{850} bands, but the V_{606} filter provides good sensitivity to objects at redshifts lower than $\simeq 4$ (Beckwith et al. 2006). As this is also the ACS filter used in the GEMS and STAGES lensing studies (due to its rich source number density, see Heymans et al. 2005, 2008), we choose the V_{606} filter image for the construction of the shapelet starter set, as representative of typical lensing source galaxies. A detailed description of the modelling of galaxies in the HUDF can be found in Appendix A. Given the input parameters, postage stamps, and model of the HUDF PSF described in this appendix, the SHAPELETS software was then used to find a best-fitting model of each HUDF galaxy. A total of 3867 of the 4128 input HUDF galaxies (94 per cent) were successfully modelled.

On examination of these 6 per cent of catastrophic modelling failures, it was found that they were caused either by:

(i) Galaxy images lying too close to the edge of the HUDF or other objects in the field (being thus automatically rejected and flagged in the output shapelet catalogue); or

(ii) A population at faint magnitudes which appear to suffer from either star–galaxy confusion or general modelling degeneracy.

The great majority of this second type of catastrophic failures were for objects with $V_{606} > 27$, which is currently beyond the realm of plausible lensing measurement with even deep imaging surveys such as GEMS or COSMOS. Without a clear means of further reducing this failure rate, it is tolerated given the fact that it predominantly affects objects beyond the detection limit of the simulations (See Section 5.5).

An additional cut was imposed on the sample, based upon the value of the reduced χ^2 output for each galaxy model. A histogram of this output statistic is shown in Fig. 1 for the HUDF galaxies. A cut of reduced $\chi^2 < 5$, beyond which the chances of a good fit are vanishingly small for the high degree-of-freedom shapelet

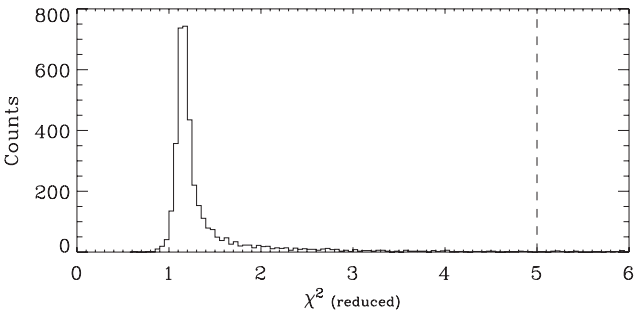

Figure 1. Histogram of reduced χ^2 values for the HUDF-deconvolved galaxy models; the dashed line is the cut introduced to exclude very poorly modelled galaxies.

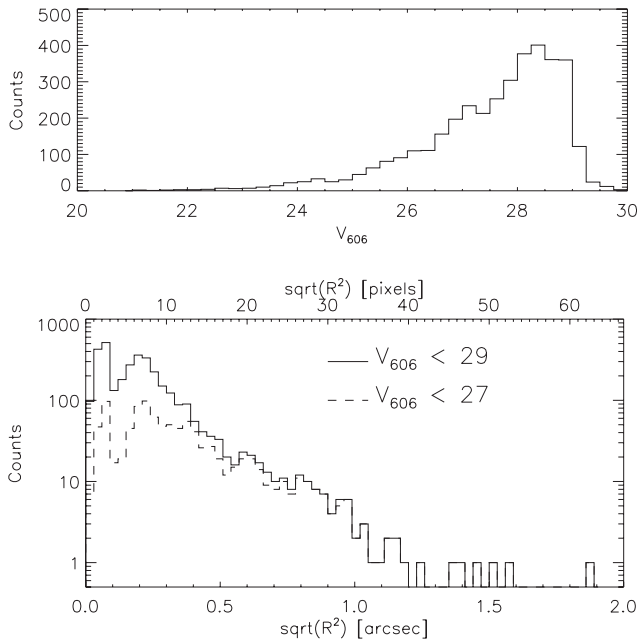

Figure 2. Upper panel: distribution of V_{606} magnitudes of shapelet galaxy models in the HUDF starter set. Lower panel: distribution of $\sqrt{R^2}$ (see equation C5 for a definition) in the HUDF starter set for two different magnitude cuts. The $V_{606} < 27$ cut approximately corresponds to the limiting magnitude at which weak lensing measurements may be made in ACS surveys such as GEMS (Heymans et al. 2005) and COSMOS (Leauthaud et al. 2007).

models used, removed a further 153 very poorly modelled galaxies not identified by shapelet flags or other indicators of total modelling failure. The remaining 3714 PSF-corrected galaxy models form the galaxy starter set that are now used to construct simulated lensing survey images with realistic galaxy morphologies. This corresponds to a high density of approximately 410 galaxies arcmin^{-2}, giving considerable freedom to realistically simulate shallower *HST* data than the HUDF (an important caveat is the limited size of the sample itself, which we discuss in the following section). Fig. 2 shows the distribution of V_{606} magnitudes and an rms size measure in the model starter set (this size is based on the shapelet-derived quantity R^2: see equation C5 in the appendices). The distribution of sizes illustrates the compact nature of galaxies in deep ACS images, also illustrated in Fig. A1.

5 SIMULATIONS OF ACS DATA

5.1 Source galaxy images

The galaxy models in the starter set described in the previous section are the fundamental data used to generate our simulations. However, there is additional processing required before models may be used to accurately simulate flexion and shear measurement.

A first problem is that the starter set itself contains the signature of gravitational shear and flexion due to the matter structure along the line of sight of the HUDF, along with any residual, uncorrected HUDF PSF anisotropy. These signals are largely eradicated by the random rotation and inversion of galaxy models. These two transformations can be performed by simple analytic manipulations of the galaxy shapelet models in the polar shapelets formalism of Massey & Refregier (2005), and so this process is fast and accurate.

A second problem is that the starter galaxy set represents a limited sample of galaxy morphologies. This may be alleviated to an extent by introducing small random perturbations to the shapelet models of the galaxy images (e.g. Massey et al. 2004; Meneghetti et al. 2008). We follow Meneghetti et al. (2008) and add an independent Gaussian random variable $N(0, \sigma_{n_1,n_2})$ to the real and imaginary part of each shapelet coefficient f_{n_1,n_2}. We choose $\sigma_{n_1,n_2} = 0.1 \times |f_{n_1,n_2}|$, which we found ensured that galaxy shapes are randomized in a way that did not introduce a high incidence of noticeably unphysical features such as negative flux (see Massey et al. 2004 for a discussion of over-randomized galaxy models).

Randomly rotated, inverted and perturbed galaxy models are then a suitable population of source galaxies, and can be assigned to locations in the final output images. Unlike the GREAT08 or GREAT10 simulations (Bridle et al. 2010; Kitching et al. 2012), we choose to assign galaxy images to random locations upon the output science tile, rather than on a regular grid. Although this complicates the analysis it has the benefit that the impact of object detection and deblending is included in our pipeline tests. These effects are important as they likely contribute to noise on shear and flexion estimates, one of the primary topics of investigation in this work, enhancing the value of any noise forecasts based on the results.

5.2 Applying shear and flexion: shapelet transformations versus raytracing

In the previous section, we arrived at a set of shapelet models suitable for use as source galaxies. The desired weak lensing distortion, a coordinate transformation governed by the lens equation, must now be applied to these galaxies. There are two approximate means by which this coordinate transformation may be applied in practice:

(i) In 'shapelet-space': shear and flexion distortions may be represented by an infinite sum of transformations upon the shapelet coefficient values in the galaxy model itself (Refregier 2003; Massey & Refregier 2005; M07). Applying successive shapelet transformations therefore allows the lensing distortion to be represented with arbitrary precision. The lensed shapelet model may then be analytically integrated across square pixels as described by Massey & Refregier (2005).

(ii) In 'real-space': the shear and flexion distortions may be represented directly using the lens equation $I(\boldsymbol{\theta}) = I^{(s)}(\boldsymbol{\theta}')$ to raytrace from the regular grid of image pixels back to an irregular sampling of points on the source plane. This is possible as each shapelet model encodes the surface brightness $I^{(s)}$ at *all* points in the source plane.

Previous shapelet-based simulation methods have preferred the former (Massey et al. 2007a; Meneghetti et al. 2008) for weak lensing transformations. For our simulations, designed to mimic space-based flexion and shear observations as closely as possible, we choose the latter real-space method for reasons that we now describe.

In the shapelet-space approach, it is necessary to apply multiple shapelet transformations to each galaxy model, each operation capturing successive terms in a series expansion representation of the true distortion, to ensure that the final distortion approximates a real shear or flexion accurately. Doing this is important if subsequently we wish to test shapelet measurement methods fairly. As shown by equation 41 of Massey & Refregier (2005), performing

a shapelet shear transformation accurate to first order in γ increases the order $n_{\max}$ of the model from $n_{\max}$ to $n_{\max}+2$, resulting in $2n_{\max}+5$ extra coefficients overall. At third order in γ, the minimum that should be considered for precision work where the applied shear γ may approach 0.1, the order of the model is increased to $n_{\max}+6$, an extra $6n_{\max}+27$ coefficients in total.

The situation is worse for flexion. M07 showed that a first-order approximation to gravitational flexion increases the order of a shapelet model to $n_{\max}+3$. Performing a shapelet flexion transformation that is accurate to third order therefore increases $n_{\max}$ by 9, meaning that a modest $n_{\max}=12$ shapelet galaxy requires a far more complex $n_{\max}=21$ model once flexed. The SHAPELETS software suffers extreme memory demands and reductions in speed beyond $n_{\max}=20$, and so this method of introducing accurate distortions becomes prohibitively slow (see Appendix A4). Unfortunately, re-truncating the shapelet model back to a more manageable $n_{\max}$ is not a simple solution to these problems. This degrades the congruence between the exact and shapelet-approximated lensing transformations in a way that is difficult to describe, as it varies with both the distortion applied and the surface brightness distribution of each individual object.

Also of concern is the action of the lensing transformations upon the β scale parameter. Lensing distortions magnify images, which for the linear-order shear and convergence transformations can be simply represented by a small change in β. With flexion a fundamental complication with β arises: as shown by Schneider & Er (2008), the determinant of the Jacobian matrix varies as a function of position for flexion-order lensing. This varying transformation in the area element cannot be reproduced via a single rescaling of the β scale parameter.

This potentially increases the difficulty of exactly reproducing non-linear flexion transformations using the shapelet formalism. When constructing simulations for flexion measurement we must be extremely careful that we are accurately describing the distortion if we wish to construct a fair test of current and future methods. The real-space, raytracing option listed above is therefore adopted for applying both flexion and shear in the simulations. As this scheme requires numerical approximation of the integration of model surface brightness over pixels, it was necessary to test the degree of approximation necessary for accurately describing shear and flexion. Details of these tests, and their results, can be found in Appendix B1. It is found that shear and flexion can be added to our simulation galaxies with sufficient accuracy for current purposes. We turn now to the problem of convolution with an instrumental PSF.

5.3 Applying the PSF

Once flexion and shear has been applied to simulation galaxies, the next stage is to convolve these correctly lensed objects with a simulation kernel that mimics a realistic target PSF. The PSF we choose for the simulations in this study is based on ACS observations of stars from V_{606} observations in the GEMS survey (Rix et al. 2004; Caldwell et al. 2008), specifically the 909 stellar objects selected as described by Heymans et al. (2005). Shapelet decompositions of these objects are created using $n_{\max}=20$, $\beta=1.80$ pixels, and stacked to make an inverse-variance-weighted average GEMS PSF model (the same procedure as described in Appendix A3 for modelling the HUDF PSF). This model can be seen as the leftmost panel in Fig. 3.

As described by Massey & Refregier (2005), any image described using polar shapelets may be easily 'circularized' (i.e. circularly averaged) by setting the model coefficients $f_{n,m}=0$ for all $m\neq 0$. The circularized GEMS PSF generated in this way can be seen in the central panel of Fig. 3 (imaged at a resolution of 0.015 arcsec pixel^{-1}), and in the rightmost panel (at the final output resolution of the simulations, 0.03 arcsec pixel^{-1}). We choose to use the circularized GEMS PSF for the simulations in this paper as its symmetry simplifies the interpretation of lensing measurement results, while still incorporating a radial profile characteristic of space-based data such as that from ACS. Diffraction spikes, such as caused by support struts for the secondary mirror, will be lacking from this adopted PSF model, but the extent to which these are successfully characterized by shapelets PSF models is unclear even when not circularizing, as in Fig. A2. The radial profile of this circularized PSF model is shown in Fig. 4. Since shapelet models are no longer being used to describe galaxies after shear or flexion is applied, the convolution must be performed numerically using a pixelized image of this PSF. As a shapelet model PSF such as that in Fig. 3 is not *formally* band-limited (see e.g. Marks 2009), this therefore requires another empirical investigation into the effects of finite sampling.

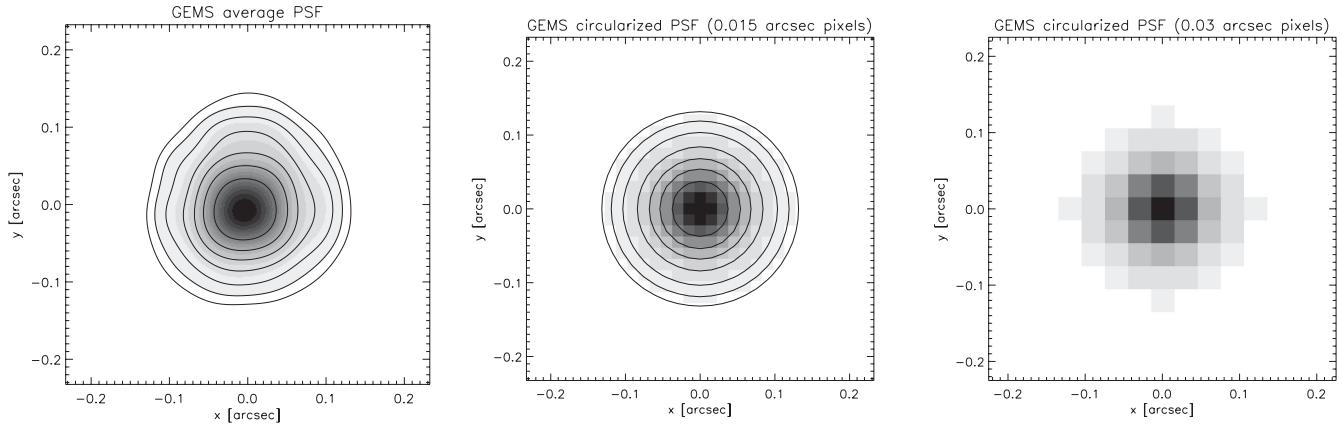

Figure 3. Left-hand panel: PSF pattern created from the weighted average of 909 shapelet models of selected stars in the GEMS V_{606} science tiles. Central panel: circularized version of this same GEMS PSF made by setting to zero all the polar shapelet coefficients $f_{n,m}$ with $m\neq 0$, and shown at the upsampled 0.015 arcsec pixel^{-1} resolution used for performing the real-space convolution. Right-hand panel: the circularized GEMS PSF, shown at the final ACS 0.03 arcsec pixel^{-1} resolution for reference. In all panels, the grey-scale is linear in surface brightness whereas the contours are logarithmic.

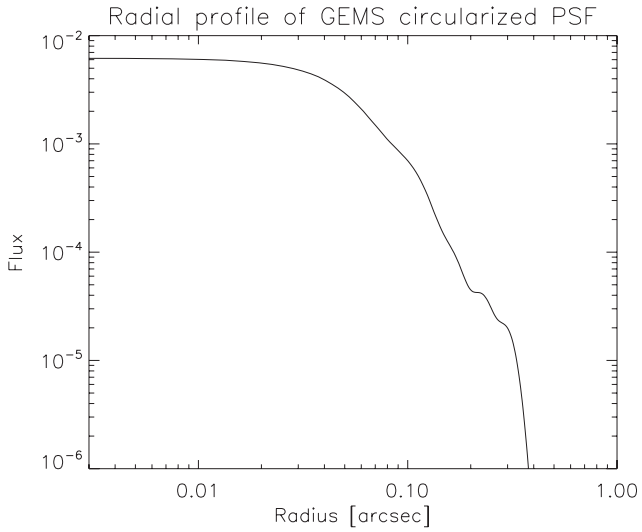

Figure 4. Radial profile of the shapelet model describing the circularized GEMS PSF as depicted in the central and right-hand panels of Fig. 3.

That this investigation is numerically feasible also illustrates a further advantage of not performing the calculation using shapelet transformations. As shown by Melchior et al. (2009), an exact shapelet treatment of convolution results in a convolved image of order $n_{\max}$ given by

$$n_{\max,\mathrm{convolved}} = n_{\max,\mathrm{galaxy}} + n_{\max,\mathrm{PSF}}. \quad (20)$$

The PSF in ACS images requires $n_{\max,\mathrm{PSF}} \gtrsim 20$, which makes describing perfect convolution on even modestly sized galaxy shapelet models extremely expensive. Re-truncating the convolved model to the original $n_{\max}$ spoils the exactness of the shapelet convolution treatment in a way that varies depending on the shape of each galaxy. In contrast, convolutions on images, performed using the fast Fourier transform (FFT, a fast algorithm for performing discrete Fourier transforms; see e.g. Press et al. 1992), can be performed to great accuracy in a fraction of the time. We test the accuracy that may be achieved using this technique in Appendix B2, and find that, as in the case of the raytracing representation of shear and flexion upon shapelet galaxy models, a desired level of accuracy can be achieved without great additional numerical cost.

5.4 Design strategy for distorted, convolved images for the ACS simulations

The results of Appendices B1 and B2 now allow informed decisions to be made regarding the construction of the ACS simulations to test the measurement of flexion and shear with shapelets. Current shear measurement methods are able to measure shear at per cent-level accuracy (Bridle et al. 2010; Kitching et al. 2012), and it is unlikely that flexion measurement will approach this capability within the near term. A conservative requirement is therefore that the treatment of distortion and convolution in the simulations should be accurate at the 10^{-3} level in terms of the impact on γ, $\mathcal{F}$ and $\mathcal{G}$ in simulated galaxies. This also matches stated requirements on the estimation of shear for an all-sky, space-based survey mission (Amara & Réfrégier 2008).

We therefore adopt an upsampling ratio of $r = 2$ when applying both the lensing image distortions (shear and flexion) and for subsequent convolution with the circularized PSF of Fig. 3, corresponding to an absolute resolution of 0.015 arcsec pixel^{-1}. It is computationally convenient that both distortion and convolution occur at this same resolution. The results of Appendices B1 and B2 suggest that errors in the pertinent moments of simulated galaxy images will be thus controlled to better than 0.1 per cent on average.

Table 1. Flexion and shear input signal values for the 270 simulated ACS tiles, where the first flexion input is given by $\mathcal{F}_{\mathrm{input}} = |\mathcal{F}_{\mathrm{input}}|\mathrm{e}^{\mathrm{i}\phi}$, the shear by $\gamma_{\mathrm{input}} = |\gamma_{\mathrm{input}}|\mathrm{e}^{2\mathrm{i}\phi}$ and the third flexion input by $\mathcal{G}_{\mathrm{input}} = |\mathcal{G}_{\mathrm{input}}|\mathrm{e}^{3\mathrm{i}\phi}$ (see Section 5.4).

	Set	$\lvert\mathcal{F}_{\mathrm{input}}\rvert$, $\lvert\mathcal{G}_{\mathrm{input}}\rvert$ arcsec^{-1}	$\lvert\gamma_{\mathrm{input}}\rvert$	ϕ	Tiles
	1	0.05	0.1	0°	1–10
‘high’	2	0.05	0.1	30°	11–20
	3	0.05	0.1	45°	21–30
	4	0.01	0.05	0°	31–40
‘mid’	5	0.01	0.05	30°	41–50
	6	0.01	0.05	45°	51–60
	7	0.005	0.01	0°	61–70
‘low’	8	0.005	0.01	30°	71–80
	9	0.005	0.01	45°	81–90

A remaining issue is the level of input shear and flexion distortions which should be applied to the simulation images. For shear, we require successful recovery of the signal due to correlated large-scale structure, but also for galaxy–galaxy lensing and the stronger shears expected around cluster lenses. For flexion, it is unlikely that the cosmological signal is measurable in the near future, but the galaxy–galaxy signal may be of interest (e.g. B06) and there are certainly applications in the field of cluster reconstruction (e.g. Leonard et al. 2007, 2011).

For the suite of simulations, we choose to split our input signals into three broad groups exploring a range of distortion signal strengths, labelled as ‘high’, ‘mid’ and ‘low’. These designations refer to the magnitude of the input gravitational signal applied, chosen to be $|\gamma_{\mathrm{input}}| = 0.1, 0.05, 0.01$, respectively, for the shear simulations.

For the flexion simulations, it was decided to bring the ‘mid’ and ‘low’ samples closer together (to concentrate on exploring sensitivity to galaxy–galaxy flexion) and extend the high signal somewhat further to investigate measurement of values which may be found in cluster studies (B06; Leonard et al. 2007, 2011). We choose $|\mathcal{F}_{\mathrm{input}}|, |\mathcal{G}_{\mathrm{input}}| = 0.05, 0.01, 0.005$ arcsec^{-1} for ‘high’, ‘mid’ and ‘low’, respectively.

In order to explore any anisotropy in signal recovery (due, e.g., to alignment with pixel axes), we split each of the three sets into a further three subsets by the angle of orientation of the input signal with respect to the image x-axis. Orientations of $\phi = 0°, 30°$ and $45°$ were chosen for the input signals $\mathcal{F}_{\mathrm{input}} = |\mathcal{F}_{\mathrm{input}}|\mathrm{e}^{\mathrm{i}\phi}$, $\gamma_{\mathrm{input}} = |\gamma_{\mathrm{input}}|\mathrm{e}^{2\mathrm{i}\phi}$ and $\mathcal{G}_{\mathrm{input}} = |\mathcal{G}_{\mathrm{input}}|\mathrm{e}^{3\mathrm{i}\phi}$, giving a total of $3 \times 3 = 9$ subsets overall for each of the three lensing distortions. These values and choices are summarized in Table 1. We note that these values span a predominantly positive range of values in the components of γ, $\mathcal{F}$ and $\mathcal{G}$: this asymmetry in the applied signal is benign due to the circular symmetry of the adopted simulation PSF (Fig. 3). 10 survey tiles were then simulated for each image set described above, each with a 3.53×3.63 arcmin^2 sky coverage area (an ACS pointing) and

at an output resolution of 0.03 arcsec pixel^{-1}. These were created by:

(i) Randomly selecting with replacement from the HUDF starter set described in Section 4.

(ii) Randomly perturbing, rotating and inverting each starter set galaxy model as described in Section 5.1.

(iii) Applying a lensing distortion as prescribed by Table 1 using the raytracing method presented in Section 5.2 with an upsampling ratio of $r = 2$.

(iv) Convolving each distorted image with the circularized GEMS PSF of Fig. 3. This convolution was performed at the image level using FFTs as described in Appendix B2, again using $r = 2$ for the upsampling ratio.

(v) Placing each simulated galaxy image at a random position in the tile.

In this way, 3×90 simulated ACS tiles were created for each of γ, $\mathcal{F}$ and $\mathcal{G}$. We note that it was decided not to include simulation tiles containing *both* shear and flexion input signals simultaneously. The possibility of cross-contamination between these signals is interesting (see Viola, Melchior & Bartelmann 2012), but in this initial study, we concentrate on examining each signal individually.

In order to take advantage of the reduction in shape noise that can be achieved by combining lensing measurements from appropriately rotated galaxy images (Massey et al. 2007a), we generate a further set of 'rotated' simulation images. These are identical to those described above except for an additional rotation of 180°, 90° and 60° (for $\mathcal{F}$, γ and $\mathcal{G}$, respectively), given to the starter set galaxies immediately after step (ii) above.

This allows the distortion measurements for matched pairs of rotated images to be averaged, cancelling the leading order impact of noise from the intrinsic galaxy shape upon shear and flexion measurement. This also allows the relative impact of galaxy shape noise and image pixel noise to be compared, which is of particular interest for flexion. The total number of simulated ACS tiles in each suite is therefore 2×90, leading to a full suite of $6 \times 90 = 540$ simulated ACS pointings (∼100 Gb of data in total).

5.5 Correlated noise

The final ingredient in the creation of simulated images is the addition of realistic measurement noise, due to diffuse background light and finite photon number counts. An important consideration for shape measurement is the impact of spatially *correlated* noise. This is present due to the standard practice of combining multiple, dithered ACS exposures to generate single 'science' images using software such as MULTIDRIZZLE (Fruchter & Hook 2002; Koekemoer et al. 2002), and is generic even for more carefully optimized linear combination schemes (e.g. Rowe, Hirata & Rhodes 2011). Such dithered images were used as the basis for weak lensing measurement in each of GEMS, COSMOS and STAGES (Rix et al. 2004; Heymans et al. 2005, 2008; Leauthaud et al. 2007; Caldwell et al. 2008; Gray et al. 2009).

We add realistic correlated noise to the simulation tiles described in Section 5.4 in a novel manner, using a 'noise mosaic' image constructed as part of the imaging reduction of the GEMS survey (Rix et al. 2004). This composite image, which is the size of a single ACS pointing, is a mosaic of multiple blank sky regions in the GEMS V_{606} science images described by Caldwell et al. (2008). Each 0.03 arcsec pixel^{-1} GEMS science image was generated from a dithered combination of three 0.05 arcsec pixel^{-1} exposures, each of total duration 2160 s. The composite noise image therefore reflects precisely the noise-correlating effects of the GEMS dither and drizzle strategy, which is typical of the reduction strategies employed for high-resolution, space-based imaging data. The galaxy image magnitude zero-points were set to match the GEMS data, resulting therefore in simulated survey images with a detection threshold around $V_{606} \sim 27$, matching the GEMS survey characteristics quite closely (Caldwell et al. 2008). The fact that the images include many more simulated objects, buried in the noise down to the HUDF starter set cutoff magnitude of $V_{606} < 29$, is another realistic feature of the simulations.

As galaxies are placed at random locations in each successive simulation tile, it is possible to re-use this noise image each time. Only a small fraction of galaxy images within tiles will consist of galaxy models that, by chance, have been repeatedly placed into the same location in the ACS pointing-sized noise mosaic ($\simeq 3.53 \times 3.63$ arcmin^2 in total area), so the level of unwanted pixel noise repetition in the simulation images will be low. In these simulations, we do not add an additional Poisson noise term to image pixels to account for variation in variance due to varying flux. However, as the overwhelming majority of simulated galaxies in the catalogues are faint relative to the background, this can be shown to represent a very small correction to the images, arguing that the effect can be safely neglected in light of other uncertainties in the modelling overall.

One final concern in the use of the GEMS noise mosaic might be if there were to be found to be some preferred direction in the noise pattern. Visual inspection of the noise map did not suggest any such artefact, and it will be seen in Section 6 that there is no evidence for an *additive* bias in shear or flexion results, or a variation in multiplicative '*m*' biases (Heymans et al. 2006) with polar angle with respect to the pixel grid. Global anisotropies or preferred directions in the noise map might be expected to cause such effects in the presence of a circularly symmetric PSF such as that adopted (Fig. 3), and the fact that we are unable to detect them within simulation uncertainties suggests that they exist at a sufficiently low level as to not affect the conclusions of this study.

It should be noted that by reusing the GEMS noise mosaic for each tile in this way, we are applying the same noise field to both galaxy images in the rotated and un-rotated galaxy pairs, as in these pairs the simulated galaxies share the same location. However, this offers an opportunity to separate the effects of pixel noise in flexion measurements from those of shape noise without diluting the former. The price paid for this opportunity is a reduction in the overall statistical power, by a factor of $\sim\sqrt{2}$ in the simplest estimate, with which measurement bias parameters may be constrained.

5.6 Summary and comparison to previous flexion simulations

After the addition of the noise image, the 540 ACS simulation tiles are complete. We now briefly summarize the differences between these simulations and those of Velander et al. (2011), in case such a comparison is of utility to the interested reader. A primary difference is that the simulations in this study employ complex morphologies in the shapelet galaxy models, constructed from HUDF imaging data, compared to the simpler parametric forms as used in Velander et al. (2011). Our simulations also employ a continuous distribution of object sizes and signal-to-noise ratio (SNR), taken directly from the UDF sample, rather than fixing galaxy sizes and SNR at fixed values of interest. Whereas the Velander et al. (2011) PSF is a Moffat profile motivated by ground-based PSFs (and made elliptical for some sets of simulated images), we instead employ a circular PSF with a radial profile taken from shapelet fits to GEMS stellar images. Deblending,

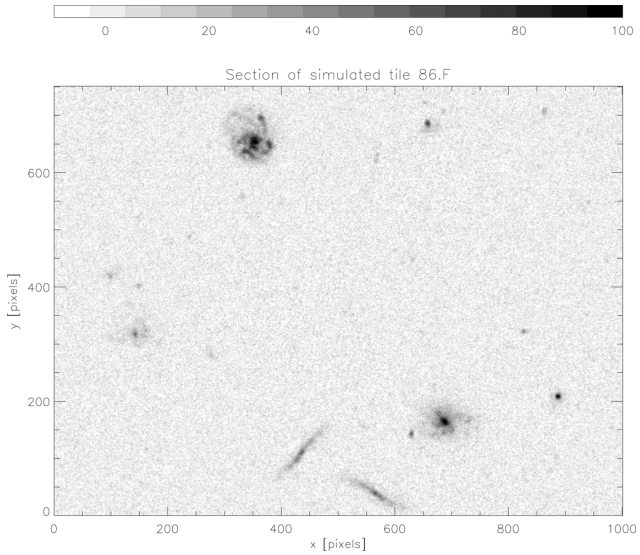

Figure 5. Illustrative 30 arcsec × 22.5 arcsec section of the 86th tile (chosen at random) in the simulated, convolved, $\mathcal{F}$-lensed images. The grey-scale is linear in flux.

and realistic noise associated with deblending, is a feature in our simulations, whereas Velander et al. (2011) placed galaxies on a regular grid. Finally, our simulations employ correlated noise taken from observations in the GEMS survey data (Caldwell et al. 2008).

In Fig. 5, we show a small section from one of these tiles, illustrating some of the realistic diversity of galaxy morphology depicted in the simulation images. All of these simulation images are available by request from the authors. We now turn to a description of the measurement of shear and flexion in these images using shapelets, allowing a calibration of the shear and flexion signal recovery using this method.

6 TESTING A SHAPELET FLEXION PIPELINE

We now describe measurements of flexion and shear made from the simulations described in Sections 4 and 5, and investigate the recovery of flexion and shear as a function of image properties such as SNR and galaxy size.

6.1 Object detection and shapelet decomposition

To test and calibrate shapelet measurements of shear and flexion, we treat the 540 ACS simulation tiles as if they were new telescope data (with properties such as galaxy positions and shapes unknown). The first step is therefore to detect objects in the images from peaks in the surface brightness distribution, and we employ the same techniques as described in Appendix A1 to create a catalogue of galaxy objects in the HUDF. We detect galaxy objects in the simulations using the SExtractor software with the parameter choices given in Table A1. An initial cut of $V_{606} < 27$ is then applied to the catalogues, along with requiring FLUX_RADIUS > 2.4 and FLUX_AUTO/FLUXERR_AUTO > 10 (see Bertin & Arnouts 1996 for descriptions of these SExtractor output parameters). These cuts are motivated by the choices made in Heymans et al. (2005), and result in catalogues containing approximately 64 galaxies arcmin^{-2}. This figure agrees well with galaxy densities found in surveys at similar depth to that simulated here, such as GEMS, COSMOS and STAGES (e.g. Heymans et al. 2005; Leauthaud et al. 2007).

Postage stamp images of each detected galaxy object are then created as described in Appendix A2 and shapelet models of the galaxies are created as described in Appendix A4, except for two important differences that we now describe. The first and most important difference is in the PSF: for this we use the shapelet model of the circularized GEMS PSF (Fig. 3) described in Appendix B2. Therefore, our calibration of flexion and shear measurement does not include the potential effects of a poorly modelled PSF, such as might be present working with real data. While unrealistic, this simplification will allows any measurement biases to be interpreted cleanly rather than being subject to external factors such as poor PSF modelling. The problem of building accurate PSF models is important enough to be addressed in its own right, and this is increasingly reflected in the literature (e.g. Hoekstra 2004; Jarvis & Jain 2004; Jarvis, Schechter & Jain 2008; Paulin-Henriksson et al. 2008; Paulin-Henriksson, Refregier & Amara 2009; Rowe 2010; Kitching et al. 2011, 2013).

The second important difference is in the choice of the NEIGHBOUR input parameter to the SHAPELETS software, setting NEIGHBOUR $= 1$. As described in Appendix A2, this causes pixels in the masked areas of each postage stamp (i.e. those associated with nearby objects, or cosmic rays, bad pixels, etc., in real data) to be given zero weighting at the shapelet modelling stage. Having made these changes to the input settings, the modelling provides a catalogue of shapelet coefficient values for each galaxy in the rotated and unrotated simulations.

The galaxies in the rotated and unrotated simulations are then matched, treating as pairs all galaxies with centroids separated by less than 0.15 arcsec in centroid (five 0.03 arcsec pixels) and 0.25 in V_{606} magnitude as estimated by SExtractor. This was found to produce a total of ∼51 000 matched galaxy pairs for each of the shear and flexion simulation suites, reducing the galaxy density to approximately 45 galaxies arcmin^{-2}. More stringent pair exclusion criteria were seen to cause significant reductions in the numbers of matched pairs. Opting for more tolerant criteria produced slightly greater numbers of galaxies (∼1–3 per cent), but not in a manner that substantially altered final results. The magnitude cut had greatest impact of the pair matching criteria, suggesting that the allocation of objects and flux in the deblending process was dominating over centroid errors in the contribution to the loss of matching pairs. The fraction of galaxies lost to catastrophic failures in shapelet modelling was small (0.4 per cent).

The difficulty of matching galaxies suggests that significant numbers of objects are being affected by noise in the determination of their properties at the SExtractor detection and deblending stage; we found ∼20–30 per cent of pairs could not be matched in this simulation, although it should be stressed that, in general, deblending errors will depend on PSF shape, noise correlation, other aspects of data quality, galaxy population morphology and SNR. In these simulations, the rate of successful matching depended upon SNR in particular, with the brighter galaxies being matched better than the faint. However, deblending does represent a realistic additional source of noise for shear and flexion measurements that is not due to the intrinsic shape of galaxies, and so it is an effect of interest for these simulations and for the Caldwell et al. (2008) two-stage object detection strategy in the adopted pipeline.

Unfortunately, this effect sets a ceiling on the number of objects that can be successfully matched after simulating an end-to-end pipeline in this way. While including deblending in the simulation test contributes realistic noise to the end measurements, it also

reduces the constraining power of the simulations as a whole when significant numbers of objects are lost. To isolate the effects, it may be preferable in future work to generate galaxies on grids as done in GREAT08/GREAT10 (Bridle et al. 2009, 2010; Kitching et al. 2011, 2012) and compare results. We will discuss the issue of deblending further in Section 7 in the context of the results of the pipeline tests.

For the pairwise matched sample, flexion and shear estimates were then generated for each galaxy from the catalogue of shapelet models, and the estimators chosen will now be described.

6.2 Estimating flexion and shear

To estimate shear and flexion from shapelet models of galaxies, we adopt an approach similar to that described by M07, and compare the relative values of the shapelet coefficients $f_{n,m}$ of best-fitting shapelet models to derive estimators of lensing distortions. In Appendix C, we describe the generation of flexion and shear estimators from catalogues of shapelet coefficient models for a population of galaxies with a realistic distribution of fluxes and sizes. We note that the multiple-decade variation in these two properties means that small modifications to the estimators proposed in M07 must be adopted. Described in detail in Appendix C, we label these estimators $\tilde{\gamma}$, $\tilde{\mathcal{F}}$ and $\tilde{\mathcal{G}}$ for the shear, $\mathcal{F}$ flexion and $\mathcal{G}$ flexion, respectively.

We also note that in order to use these estimators, it was made a condition that the shapelet model reach sufficient $n_{\max}$ as to contain non-zero values for all the shapelet coefficients required by equations (C1), (C4) and (C9) when estimating γ, $\mathcal{F}$ or $\mathcal{G}$, respectively. This is to avoid bias from artificially setting coefficient values to zero when this is not the most appropriate prior expectation.

Overall, this approach differs from that of Velander et al. (2011), who instead take circular profiles and estimate the shear and flexion required to distort these objects to match the data, in a manner similar to the shear-only estimators of Kuijken (2006) and Bernstein & Jarvis (2002). Both methods are similar in that they rely to some extent upon shapelet models being an accurate description of the underlying surface brightness distribution of galaxies to avoid what has been identified as 'underfitting bias' in galaxy shape estimation (e.g. Voigt & Bridle 2010; Bernstein 2010). However, in our simulation tests shapelet models have been used to provide the underlying galaxy light profiles, and perfect knowledge of the PSF is also available.

In principle, this allows the direct probing of potential biases due to the use of noisy data, such as imperfect deblending with SExtractor, or the biased response of non-linear estimators under noise itself (so-called noise bias, e.g. Bernstein & Jarvis 2002; Hirata et al. 2004; Kacprzak et al. 2012; Melchior & Viola 2012; Refregier et al. 2012). However, in practice, the shapelet method of Massey & Refregier (2005) truncates models once they become consistent with the noise in the image, iterating the modelling parameters β and $n_{\max}$ until this is achieved, while always seeking the model with the lowest $n_{\max}$ that meets this criterion. Galaxy models will therefore in general still be truncated, leading to the distinct possibility of underfitting biases still being present in lensing estimates from the shapelet models constructed in Section 6.1.

Using the method described in Appendix C, estimators of shear and flexion are constructed for each of the galaxies with successful detections and shapelet decompositions described in Section 6.1. In the following sections, we compare these estimates to the input gravitational signal, and explore how the properties of these estimates vary with galaxy properties such as SNR and size.

6.3 Flexion and shear estimator results

In Fig. 6, we plot $(\tilde{\gamma} - \gamma_{\rm input})$ versus $\gamma_{\rm input}$ (top panel), $(\tilde{\mathcal{F}} - \mathcal{F}_{\rm input})$ versus $\mathcal{F}_{\rm input}$ (middle panel) and $(\tilde{\mathcal{G}} - \mathcal{G}_{\rm input})$ versus $\mathcal{G}_{\rm input}$ (bottom panel) for our simulated galaxies, with estimators for rotated and unrotated pairs of galaxies being mean-averaged to reduce noise as described by Massey et al. (2007a). Results were not found to differ between components (i.e. results for γ_1 were consistent with those for γ_2, etc.) and so the plots show both real and imaginary parts for each signal. Results were binned according to input signal via the image sets described in Section 5.4, and the points for each bin represent the *median* of the averaged rotated and unrotated estimates in each case. For shear, the median results were consistent with results derived from the arithmetic mean; for flexion, the mean was found to be very noisy due to the distribution of flexion estimators (see Section 6.5), and so the median was preferred as the comparison

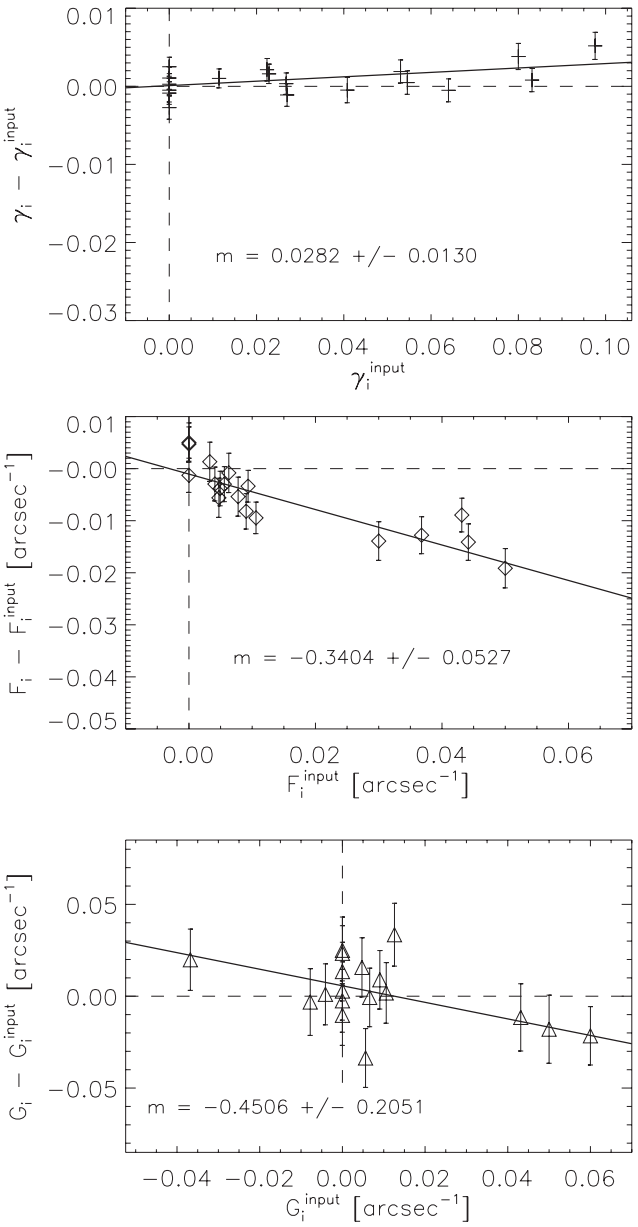

Figure 6. Lensing measurement calibration results from the full set of simulated galaxies in matched, rotated pairs, for $\tilde{\gamma}$ (top panel), $\tilde{\mathcal{F}}$ (central panel) and $\tilde{\mathcal{G}}$ (bottom panel).

statistic in both cases. The estimated median of each variate in a bivariate distribution, taken along each of its two dimensions independently, is a good estimator of the central tendency in both variates provided that correlations between them are linear. We found no evidence of non-linear correlations between the estimates of $\mathcal{F}_1$ and $\mathcal{F}_2$, indeed no evidence of correlations at all, and so do not consider the use of the median (rather than, e.g. convex hull stripping; Velander et al. 2011) to be a source of bias in this analysis. The uncertainties plotted show the standard error on the median in each bin (Lupton 1993).

We fit a linear relation to the results of Fig. 6, deriving best-fitting slope m and offset c bias parameters as used in the STEP project (Heymans et al. 2006):

$$\tilde{\gamma} - \gamma_{\rm input} = m\gamma_{\rm input} + c, \tag{21}$$

with similar expressions for the two flexion estimators. The best-fitting values, and uncertainties, are given in Fig. 6. The real and imaginary parts were again found to give consistent results, and so the best-fitting m and c describe input versus output for both components. We found that the c values were consistent with zero in all cases. This is expected for the purely circular PSF chosen for these simulations, and for the square grid which imparts no preferred sign for any of the shears or flexion distortions applied. However, it provides a useful null test of the algorithms adopted.

The value of the multiplicative bias m was found to be relatively small in the case of shear, $m = 0.028 \pm 0.013$. These results are comparable to those obtained with a number of shear estimation methods in the GREAT challenges (Bridle et al. 2010; Kitching et al. 2012), although there are a number of differences between these simulations and those used in GREAT08 and GREAT10 (e.g. correlated noise; the distribution of galaxy sizes and SNR; overlapping objects; a purely circular PSF). For flexion, we detect stronger multiplicative biases, finding $m = -0.340 \pm 0.053$ for the median of the $\mathcal{F}$ estimators, and $m = -0.45 \pm 0.21$ for the median of the $\mathcal{G}$ estimators. As discussed above in Section 6.2, the shapelet approach adopted may be affected by underfitting bias (or model bias) caused by shapelet truncation, or bias due purely to noise, or both.

The values of the biases seen are comparable to those identified by Velander et al. (2011), for galaxies based on analytic profiles and a different distribution of sizes and SNR values. We now discuss the variation of our measured estimator biases as a function of these properties, and discuss some reasons for the possible presence of this bias.

6.4 Dependence of bias on noise and apparent galaxy size

In Fig. 7, we plot the dependence of the multiplicative bias m (see equation 21) upon observed galaxy SNR for pair-matched shear and flexion estimates from the simulations. Here, SNR is defined in terms of the SExtractor output parameters FLUX_AUTO and FLUXERR_AUTO as

$$\mathrm{SNR} = \frac{\mathrm{FLUX\_AUTO}}{\mathrm{FLUXERR\_AUTO}} \times 1/\sqrt{0.316}, \tag{22}$$

where the scaling factor $1/\sqrt{0.316}$ is taken from Leauthaud et al. (2007) and adjusts SNR in drizzled *HST* images to account for correlated noise, adding the assumption that excess Poisson variance due to object flux above the background is negligible (it is not included in our simulations). This is only an approximate correction, based on a simplified model (Casertano et al. 2000) and an assumption that the COSMOS drizzling approach closely resembles that used in the GEMS noise mosaic (it does, although there is a small difference in the kernel used for the Mk II GEMS reduction; see Caldwell et al. 2008). However, it does help take correlated noise into account at a level of accuracy that is appropriate given that differences between definitions of SNR can also introduce factor $\sim$2 discrepancies.

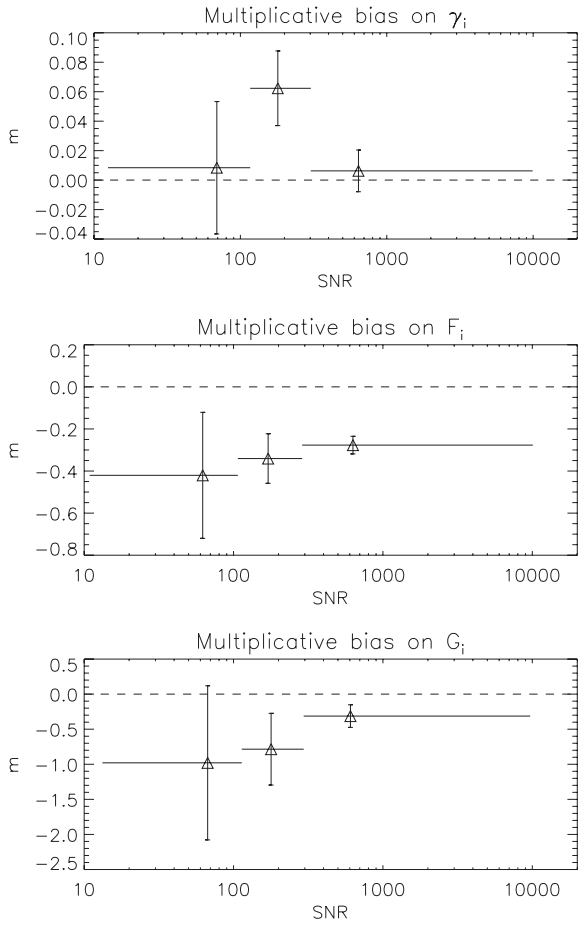

Figure 7. Variation of multiplicative bias m in shear and flexion estimation versus 'observed' SNR for the simulation galaxies (see equation 21). SNR bins were chosen to give equal numbers of galaxy in each bin: the increase in errors for low-SNR objects is due to the increasing scatter in individual estimates. Solid, horizontal lines through points show the extent of each bin.

The SNR ranges for the bins were chosen to give equal numbers of galaxies in each bin; for reference, the bin maxima and minima are given in Fig. 9 and indicated in Fig. 7 by solid horizontal lines. It is difficult to discern strong support for overall trends – results for the lower SNR bins are noisy compared to those in higher SNR bins, particularly in the case of the flexion – although the higher SNR bin consistently gives arguably better results. Overall, flexion results are broadly consistent with Velander et al. (2011), within large errors, hinting that the impact of more complex galaxy morphology (a key difference between these simulations and those of Velander et al. 2011) upon flexion estimation bias is not great compared to other properties of the observational data. In addition to the SNR of galaxy objects, the size is an important observable quantity for measuring flexion. Unlike shear, flexion has dimensions of inverse angle, and larger objects provide greater leverage for measuring flexion. In Fig. 8, we plot m for the simulated galaxy sample binned into three bins reflecting angular size, chosen to given equal numbers of galaxies in each bin. The size estimate adopted for this binning is $\sqrt{R^2}$ for the best-fitting shapelet models, plotted in arcsec, where

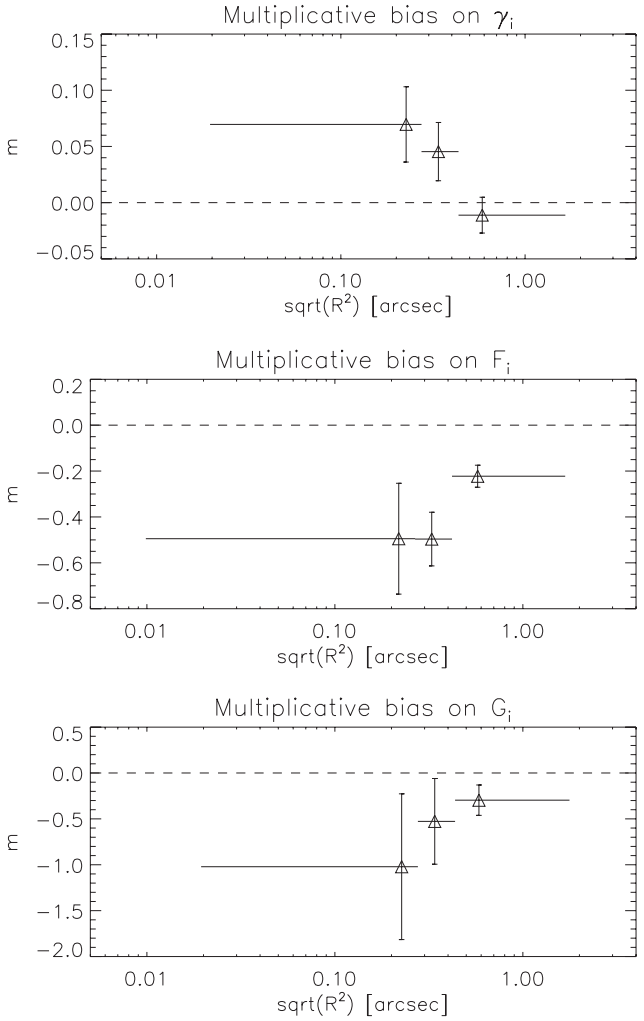

Figure 8. Variation of multiplicative bias m in shear and flexion estimation versus observed size for the simulation galaxies parametrized by $\sqrt{R^2}$ (Massey & Refregier 2005). In a similar manner to the SNR binning, size bins were chosen to give equal numbers of galaxy in each bin: the increase in errors for smaller objects is therefore due to the increasing scatter in individual estimates. Solid, horizontal lines through points show the extent of each bin.

the size measure R^2 derived from shapelet model coefficients is described by Massey & Refregier (2005) and given in equation (C5). It is equivalent to an unweighted, integrated second-moment size over the shapelet model.

Results are qualitatively similar to those found for the SNR tests, which is not altogether surprising as we expect a positive correlation between size and SNR among the simulated galaxy population. There is some evidence that estimators are better for larger apparent sizes; the shear in particular now shows a marked trend towards improvement for larger objects. But the noise on estimates of m for flexion estimators become large for smaller objects, and so for flexion clear trends are difficult to discern.

The large uncertainty on flexion estimates, despite the cancellation of intrinsic shape noise at leading order, also provides a possible explanation for the clear detection of significant m biases. Noise bias (sometimes referred to as noise rectification bias, e.g. Bernstein & Jarvis 2002; Hirata et al. 2004; Kacprzak et al. 2012; Melchior & Viola 2012; Refregier et al. 2012) is an inherent property of statistical estimators derived from non-linear combinations of random deviates, in this case the values of noisy pixels. For the shapelet technique used in this paper, the linearity of the decomposition is broken by the presence of a convolving PSF (see Massey & Refregier 2005). Flexion estimators, being sensitive to higher order (and therefore noisier) shape moments in galaxy images, are plausibly more sensitive to noise biases when compared to shear estimators derived from shapelets, as well as to underfitting biases caused by shapelet model truncation (see Section 6.2). Another reason for a greater susceptibility to noise bias might be the need to apply a centroid shift correction to in the estimation of $\mathcal{F}$, although this would not be able to account for the observed bias in $\mathcal{G}$ estimators.

6.5 The distribution of flexion and shear measurements

The results of Section 6.4 (Figs 7 and 8) show a significant increase in the uncertainty of estimates of m as SNR (and the related property $\sqrt{R^2}$) decrease, despite the number of objects being the same in each bin. This effect is strongest in the measurements of m for $\mathcal{F}$ and $\mathcal{G}$. This can be explained by considering the two separable contributions to measurement uncertainty: the scatter in the intrinsic shapes of galaxies prior to lensing, and the noise in shape estimates due to noise in pixels. The latter will increase as SNR decreases, and may also be a cause of bias as well as increased uncertainty in individual shape estimates (Melchior & Viola 2012; Refregier et al. 2012). In Fig. 9, we plot the distributions of pair-matched shear and flexion estimates as a function of SNR. These paired combinations cancel the leading order contribution of intrinsic galaxy shape to the estimation of each signal, so that the remaining scatter is that solely that due to the differing pixel noise applied to each simulated galaxy. The uncertainty of shear estimates increases as SNR decreases, but the effect is far stronger for the estimators of flexion.

Fig. 9 also illustrates the non-Gaussian *distribution* of the uncertainty in shapelet shear and flexion estimators due to noise in image pixels. For flexion in particular, this distribution is highly non-Gaussian, but this is also noticeably true for the shear estimates: the distributions show a sharp peak in the region of central tendency, accompanied by broad wings. However, as the intrinsic ellipticity in galaxy images is typically $\sigma_e \simeq 0.3$, the shear measurement uncertainty in the left-hand panel of Fig. 9 only becomes a very significant additional contribution for the lowest SNR galaxies.

For the first flexion $\mathcal{F}$, estimation of the uncertainty due to intrinsic galaxy shapes has proved to be difficult. A value of $\sigma_{\mathcal{F}} \simeq 0.04\,\mathrm{arcsec}^{-1}$ was measured by (Goldberg & Bacon 2005) for a population of galaxies that was brighter, and mostly larger (within the uncertainties of PSF correction for poorly resolved objects) than those in the current simulation set. However, even incorporating scalings with galaxy size and population, it seems likely that measurement error will represent a significant contribution to overall uncertainty for a broad range of interest in SNR. Indeed, the size of the measurement uncertainty itself suggests that it may be quite challenging to determine reliable estimates of $\sigma_{\mathcal{F}}$ for deeper galaxy populations than those explored in Goldberg & Bacon (2005): it has proved to be so in practice. That there has been no detection of gravitational $\mathcal{G}$ (despite a number of $\mathcal{F}$ detections) is also plausibly attributable to the extreme measurement scatter seen in the right-hand panel of Fig. 9, and while the intrinsic $\sigma_{\mathcal{G}}$ is yet to be reliably estimated for real galaxies this measurement uncertainty is likely to be a significant contribution to noise in future. Flexion forecasts which fail to account for the extra uncertainty due to pixel noise

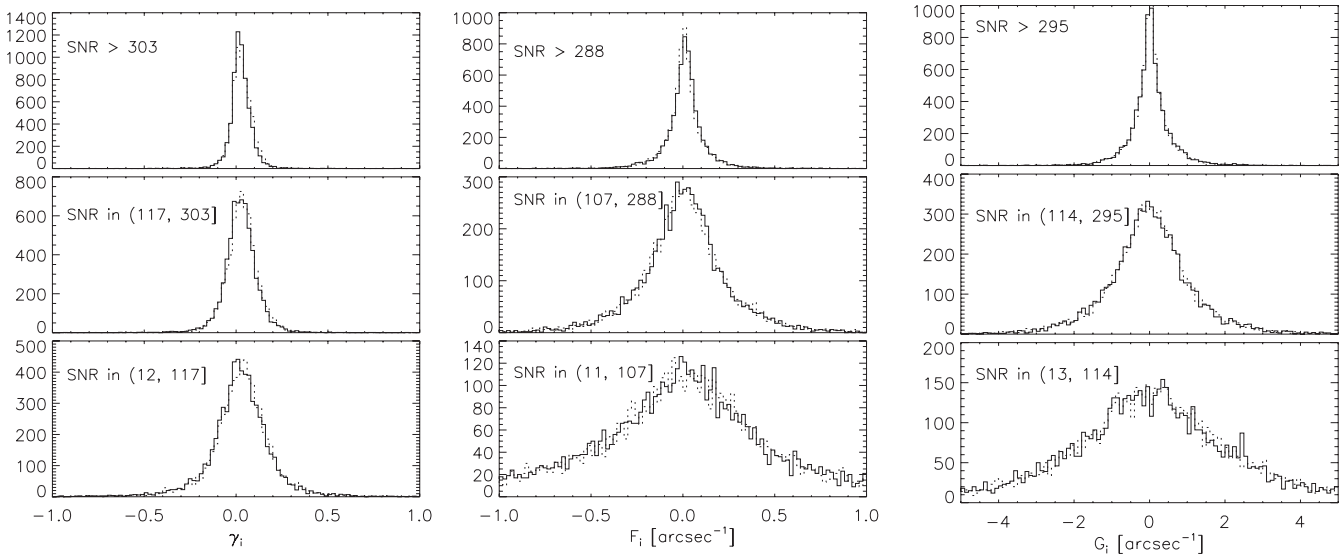

Figure 9. Distributions of matched measured $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ (solid and dashed lines, respectively; left-hand panel), $\tilde{\mathcal{F}}_1$ and $\tilde{\mathcal{F}}_2$ (solid and dashed lines, respectively; centre panel), and $\tilde{\mathcal{G}}_1$ and $\tilde{\mathcal{G}}_2$ (solid and dashed lines, respectively; right-hand panel) from the mean of matched pairs of galaxies in the rotated and unrotated simulations (see Section 5.4). These paired combinations will cancel the leading order contribution of intrinsic galaxy shape to the estimation of each signal: what remains will be dominated by uncertainty due to pixel noise and deblending.

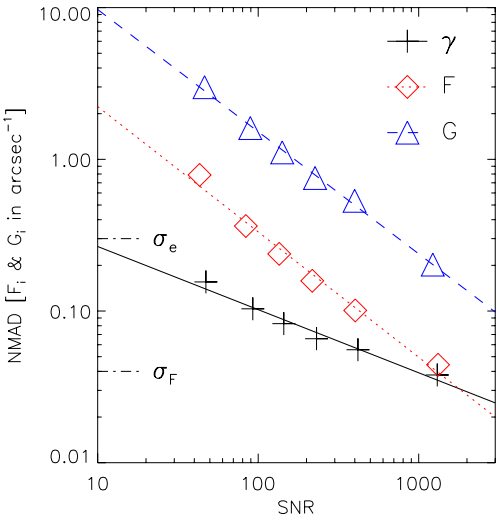

Figure 10. NMAD for the distributions of shear and flexion measurement error in Fig. 9, plotted as a function of SNR for the binned galaxy sample. The best-fitting power-law models are plotted over the data points as straight lines (model parameters listed in Table 2).

will provide overoptimistic estimates of future prospects for flexion measurement.

To provide a simple description of the shear and flexion uncertainty due to pixel noise as a function of SNR, we calculate the normalized median absolute deviation (NMAD) for shear and flexion measurement error distributions in six bins of SNR. The NMAD is useful as a robust estimator of dispersion; for distributions such as those in Fig. 9, the population standard deviation itself may be poorly defined, and the sample standard deviation may be extremely noisy and sensitive to object cuts and outlier removal. We plot the NMAD as a function of SNR in Fig. 10, along with the best-fitting power-law description of these data

$$\mathrm{NMAD} = A_{100}(\mathrm{SNR}/100)^{b}. \quad (23)$$

Table 2. Best-fitting parameters for the power-law model of equation (23), describing the NMAD of the shapelet estimator shear and flexion measurement error as a function of SNR in this study. These models are plotted over the data in Fig. 10.

Lensing measurable	A_{100}	b
γ	0.10	−0.42
$\mathcal{F}$	0.33	−0.83
$\mathcal{G}$	1.52	−0.81

Here, SNR = 100 has been chosen as the reference scaling value since this lies roughly in the middle range of simulated galaxies. The best-fitting parameters are given in Table 2. The difference in the power-law slope between the flexion and shear results, and the similarity of the slopes for the $\mathcal{F}$ and $\mathcal{G}$ estimators, is interesting: it suggests a common origin for the increased flexion measurement noise in each case, despite differences in the amplitude of the effect. A definitive theoretical explanation for differences between the shear and flexion slopes b is unclear in the presence of multiple factors (noise on pixels, deblending, centroid uncertainties), but the difference is a clear effect in these simulated data, for these estimators. One possibility (suggested by anonymous referee) is that since fainter galaxies are smaller, and flexion is a dimensional quantity with units of inverse angle, the noise will necessarily increase more steeply than shear as galaxy SNR decreases. We discuss this scenario, along with alternatives and possible tests, in Section 7.

These results highlight the importance of considering measurement noise when discussing flexion estimation in practice. We illustrate the significance of this contribution to uncertainty by adding a marker, showing the Goldberg & Bacon (2005) intrinsic flexion estimate, to Fig. 10 (it should be stressed that the galaxy population from which this value was derived is not that we are simulating here, and so this point is merely illustrative; the intrinsic $\mathcal{G}$ is still

unmeasured, as has been mentioned above).[2] In these tests, flexion estimation therefore appears to operate in a somewhat different regime to that of shear estimation, where the measurement contribution to uncertainty is dominated by the intrinsic ellipticity $\sigma_e \simeq 0.3$ of galaxies. This is true throughout the range of SNR $\gtrsim 10$ where shear estimators have been shown have some success in controlling bias (Bridle et al. 2010; Kitching et al. 2012). This fact has allowed forecasts for shear surveys to proceed using the intrinsic variance alone as a reasonable approximation to the uncertainty in shear estimators from all sources. The results of this study suggest that this should not be done for flexion, where noise due to the finite numbers of photons arriving at the detector is significant.

7 DISCUSSION

We have undertaken a detailed investigation into the problem of estimating flexion and shear from noisy galaxy images with substantial variation in underlying galaxy morphology. The simulated galaxy models also display a realistic distribution of sizes and apparent fluxes, drawn directly from the galaxy sample in the HUDF, but with additional noise realistic for a wider area (e.g. GEMS; COSMOS) survey.

We have found evidence of a qualitative difference between flexion and shear measurement uncertainties. Whereas noise in shear estimates is dominated by intrinsic galaxy ellipticity at typical survey image depths, the corollary appears not to be true for flexion. Instead, there is large uncertainty due to a combination of: (i) noise at the pixel level (due to read noise and finite photon number counts); and (ii) (related) uncertainties in deblending. Furthermore this dispersion in flexion estimates increases steeply for fainter galaxy images. It will be important to account for this fact when generating forecasts of the flexion information content in future surveys: predictions based solely on the scatter in intrinsic galaxy flexion $\sigma_{\mathcal{F}}$ alone will be too optimistic. Existing forecasts such as those of B06 that use $\sigma_{\mathcal{F}} \simeq 0.04$ arcsec^{-1} (Goldberg & Bacon 2005) may only correspond to predictions in the limit of large image SNR, and are only appropriate for the population of galaxies for which the figure of $\sigma_{\mathcal{F}}$ was measured. Other predictions, some of which have used values as low as 0.03 arcsec^{-1} for the intrinsic flexion dispersion, will be yet more optimistic.

To provide some aid to more realistic forecasts in future, we fit a simple power-law model to the dispersion of flexion measurements due to noise on pixels and deblending, as a function of source galaxy image SNR (Section 6.5). It was found that noise in flexion estimates varies significantly more strongly with galaxy SNR than was found for shear estimates, nearly a factor of 2 in power-law slope. As the majority of galaxy images in any wide-area survey are likely to be faint, it will be important to consider this effect when forecasting what may be learned from flexion in practice.

As well as considering the noise in flexion measurements, systematic biases in shapelet shear and flexion estimators were also investigated. Such tests had only previously been performed with galaxy simulations with less morphological richness in the underlying models (Velander et al. 2011). The best-performing flexion estimator in our shapelet pipeline showed comparable performance to the shapelet estimator tested by Velander et al. (2011). Because of the direct use of the HUDF galaxy size and SNR distribution in the simulations, exploration of the dependence of biases on these properties is less clean, requiring broad bins. We found little strong evidence of clear systematic trends within the errors for the flexion results, except that flexion measurement appeared to perform somewhat better for large, high-SNR galaxies. Shear results for the shapelet pipeline compare well alongside the better performing methods in the most recent shear measurement simulation challenges, GREAT08 and GREAT10 (Bridle et al. 2009, 2010; Kitching et al. 2011, 2012).

In the future, theoretical tools from the increasing body of work invested in understanding the effects of pixel noise in shear measurement (e.g. Bernstein & Jarvis 2002; Hirata et al. 2004; Kacprzak et al. 2012; Refregier et al. 2012; Melchior & Viola 2012; Miller et al. 2013) will be useful in forming a deeper understanding of noise and bias in flexion estimators. In particular, it may be possible to understand the apparent qualitative differences between flexion and shear estimators under pixel noise.

As mentioned previously in Section 6.4, one possible source of difference between estimators of $\mathcal{F}$ and estimators of γ and $\mathcal{G}$ is the need to apply a simultaneous correction for the induced centroid shift (see Appendix C). This centroid term is itself noisy. Furthermore, because it consistently appears in the *denominator* for flexion estimators (e.g. Okura et al. 2008) the shift correction could potentially increase rectification biases, as well as introduce large wings to the distribution of estimates of $\mathcal{F}$. In future work, it would be interesting to explore the uncertainties in $\mathcal{F}$ flexion estimates in synthetic test cases where galaxies model fits are forced to use the fixed, true (i.e. pre-flexion) centroids. An additional study of interest in a related topic would be an analysis in which rotated pairs of model galaxies were fitted about the same centroid: although unrealistic in practice and untenable as a remedy to the issue, such a study might give an insight into the nature of $\mathcal{F}$ flexion measurement noise.

However, the fact that the centroid shift affects only estimates of $\mathcal{F}$ and not $\mathcal{G}$ suggests that it cannot provide a full explanation for the steep power-law slope seen in the dispersion of measurements of flexion as a function of SNR (see Fig. 10 and Table 2). The estimates of this slope are similar for both $\mathcal{F}$ and $\mathcal{G}$, and markedly different from those of shear, suggesting that there may be a common origin for these noise properties. One possibility for this difference, suggested by anonymous referee, is the dimensionality of flexion as briefly mentioned in Section 6.5. Fainter galaxies tend to be smaller (although the slope of this relationship depends on the galaxy population), and flexion is a dimensional quantity with units of inverse angle. An interesting experiment would be to return to tests such as those of Velander et al. (2011) or Bridle et al. (2010), with simpler parametric galaxy profiles and grid positions for galaxy centroids (to remove noise due to deblending). These would allow a controlled exploration of noise in flexion measurement, for example as a function of galaxy size at constant SNR, or as a function of SNR at constant galaxy size.

Recent examples of such tests in the case of shear (GREAT08 and GREAT10; Bridle et al. 2010; Kitching et al. 2012) provided a compelling demonstration of the ubiquitous impact of noise biases across many measurement methods. This was something that had been pointed out as potential issue for certain shear measurement methods prior to the GREAT challenges (e.g. Bernstein & Jarvis 2002; Hirata et al. 2004). However, the demonstration of clear experimental dependences on SNR excited a recent surge in interest in the problem, leading to a greatly improved understanding of how

[2] Goldberg & Leonard (2007) found $\sigma_{a|\mathcal{F}|} \simeq 0.03$ arcsec^{-1} where a is the size of the galaxy, indicating expected intrinsic flexion dispersions in the range 0.03–0.3 arcsec^{-1} for the HUDF starter set described in Section 4. However, it is unclear how much of this may have been intrinsic shape, versus noise, dispersion.

noise biases are a consistent presence in all but the most carefully constructed shear measurement methods (Melchior & Viola 2012; Kacprzak et al. 2012; Refregier et al. 2012; Miller et al. 2013). Similar work may provide an understanding of why the dispersion on flexion measurements due to noise on image pixels increases so steeply as a function of galaxy population. It would also be interesting to compare results for a HOLICS-type method, for which there are indications of lesser uncertainty in flexion estimates relative to a shapelet treatment (Leonard et al. 2011, although we note that some details of the implementation of shapelet estimates in this work differed markedly from this analysis).

Some fraction of the dispersion in the flexion estimates presented here is likely to be due to deblending. Deblending is an inevitably non-linear process, which introduces a source of noise that is not intrinsic to the individual galaxy shape, and will contribute most greatly at low SNR. The simple gridded tests pursued by Bridle et al. (2010), Velander et al. (2011) and Kitching et al. (2012), while powerful for exploring many aspects of shape measurement, are not able to elucidate noise due to deblending. The random placement of galaxy models and the presence of many HUDF starter set galaxy models too faint to be detected versus the GEMS-like noise levels mean that these simulations are able to mimic some important aspects of the deblending problem.

It was in order to capture a realistic contribution from this affect that the adopted SExtractor parameters used in the pipeline (see Table A1) were based on the GEMS-optimized choices made by Caldwell et al. (2008). In fact, the two-pass strategy employed by these authors is a relatively sophisticated attempt to tackle deblending: many more recent (e.g. COSMOS: Leauthaud et al. 2007; CFHTLenS: Erben et al. 2013) and upcoming surveys rely on a single-pass strategy alone, and will be still more susceptible to the impact of deblending.

However, as discussed in Section 6.1, a fraction of galaxies could not be matched with their pairwise rotated partner due to noise on galaxy magnitudes and centroids from deblending. Matching criteria were made quite stringent to ensure a high purity of legitimate pairs, at the cost of the loss of objects from the sample (particularly at low SNR where matching was most difficult by far). Despite these precautions, it is a cause for concern if deblending causes the misidentification of paired sources and the breakdown of the pairwise cancellation of the intrinsic galaxy shape contribution to dispersion in flexion measurements. This effect could, potentially, lead to a contamination of the intrinsic shape dispersion into the pairwise rotated pairs designed to be free from this additional source of noise. However, while these effects are plausibly a contaminant for the lowest SNR results, the amplitude and slope found at higher SNR are unlikely to be significantly contaminated. This can be seen from the shear results of the lower-left panel of Fig. 9, for which the full width at half-maximum (FWHM) of the distribution is $\sim$0.3 (corresponding to a standard deviation of $\sigma \simeq 0.3/2.35 = 0.128$ in a Gaussian approximation to the distribution). If this were fully contaminated by miscancelling pairs, we would expect to see a Gaussian of $\sigma \simeq 0.3$, and given that these additional shape noise terms add approximately in quadrature it can be seen that the overall level of contamination cannot be large even at low SNR. At higher SNR, given the greater rate of successful matching and the width of the shear distributions in Fig. 9, the contamination is negligible.

The results of this study, originally designed to merely calibrate a flexion measurement pipeline for an analysis of ACS data (in a similar manner to Velander et al. 2011), warrant further investigation. The dispersion of flexion measurements due to pixel noise, and the related issue of deblending, has been demonstrated to be an extremely important contribution to flexion measurement and to depend steeply on galaxy SNR. Understanding the fundamental source of this strong dependence upon galaxy SNR (explanations include centroiding effects, noise rectification or the additional angular dimensionality of flexion) would be a fascinating topic for further investigation with custom-designed tests rather than simulations seeking to represent all aspects of real data. The clear difficulties of flexion measurement in practice call for a better understanding of such issues if the potential of flexion as a probe of small-scale power in matter structure is to be realized.

ACKNOWLEDGEMENTS

The authors would like to thank Dave Goldberg for interesting discussions in the course of the work that led to this paper, and the anonymous referee for many helpful suggestions and improvements. BR and CH acknowledge support from the European Research Council in the form of a Starting Grant with numbers 240672 (BR) and 240185 (CH). JR and BR were supported by the Jet Propulsion Laboratory, which is run by California Institute of Technology under a contract for NASA. This work was supported in part by the National Science Foundation under Grant No. PHY-1066293 and the hospitality of the Aspen Center for Physics.

REFERENCES

Amara A., Réfrégier A., 2008, MNRAS, 391, 228
Arfken G. B., Weber H. J., 2005, Mathematical Methods for Physicists, 6th edn. Elsevier, Amsterdam
Bacon D. J., Goldberg D. M., Rowe B. T. P., Taylor A. N., 2006, MNRAS, 365, 414 (B06)
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Beckwith S. V. W. et al., 2006, AJ, 132, 1729
Benjamin J. et al., 2007, MNRAS, 381, 702
Benjamin J. et al., 2013, MNRAS, 431, 1547
Bergé J., 2006, Shapelets Manual Volume 1, available online at: http://www.astro.caltech.edu/jberge/shapelets/manual/
Bernstein G. M., 2010, MNRAS, 406, 2793
Bernstein G. M., Jarvis M., 2002, AJ, 123, 583
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Bridle S. et al., 2009, Ann. Appl. Stat., 3, 6
Bridle S. et al., 2010, MNRAS, 405, 2044
Caldwell J. A. R. et al., 2008, ApJS, 174, 136
Casertano S. et al., 2000, AJ, 120, 2747
Erben T. et al., 2013, MNRAS, 433, 2545
Fruchter A. S., Hook R. N., 2002, PASP, 114, 144
Fu L. et al., 2008, A&A, 479, 9
Goldberg D. M., Bacon D. J., 2005, ApJ, 619, 741
Goldberg D. M., Leonard A., 2007, ApJ, 660, 1003
Gray M. E. et al., 2009, MNRAS, 393, 1275
Hartig G. F., Krist J. E., Martel A. R., Ford H. C., Illingworth G. D., 2003, in Blades J. C., Siegmund O. H. W., eds, Proc. SPIE Conf. Ser. Vol. 4854,On-Orbit Alignment and Imaging Performance of the HST Advanced Camera for Surveys. SPIE, Bellingham, p. 532
Heymans C. et al., 2005, MNRAS, 361, 160
Heymans C. et al., 2006, MNRAS, 368, 1323
Heymans C. et al., 2008, MNRAS, 385, 1431
Heymans C., Rowe B., Hoekstra H., Miller L., Erben T., Kitching T., van Waerbeke L., 2012a, MNRAS, 421, 381
Heymans C. et al., 2012b, MNRAS, 427, 146
Heymans C. et al., 2013, MNRAS, 432, 2433
Hirata C. M. et al., 2004, MNRAS, 353, 529
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. et al., 2006, ApJ, 647, 116
Huff E. M., Eifler T., Hirata C. M., Mandelbaum R., Schlegel D., Seljak U., 2011, preprint (astro-ph/1112.3143)
Irwin J., Shmakova M., 2005, New Astron. Rev., 49, 83
Irwin J., Shmakova M., 2006, ApJ, 645, 17
Irwin J., Shmakova M., Anderson J., 2007, ApJ, 671, 1182
Jarvis M., Jain B., 2004, preprint (astro-ph/0412234)
Jarvis M., Schechter P., Jain B., 2008, preprint (astro-ph/0810.0027)
Jee M. J., Tyson J. A., Schneider M. D., Wittman D., Schmidt S., Hilbert S., 2013, ApJ, 765, 74
Kacprzak T., Zuntz J., Rowe B., Bridle S., Refregier A., Amara A., Voigt L., Hirsch M., 2012, MNRAS, 427, 2711
Kaiser N., 2000, ApJ, 537, 555
Kaiser N., Squires G., Broadhurst T., 1995, ApJ, 449, 460
Kilbinger M. et al., 2013, MNRAS, 430, 2200
Kitching T. D., Miller L., Heymans C. E., van Waerbeke L., Heavens A. F., 2008, MNRAS, 390, 149
Kitching T. D. et al., 2011, Ann. Appl. Stat., 5, 2231
Kitching T. D. et al., 2012, MNRAS, 423, 3163
Kitching T. D. et al., 2013, ApJS, 205, 12
Koekemoer A. M., Fruchter A. S., Hook R. N., Hack W., 2002, in Arribas S., Koekemoer A., Whitmore B., eds, Proc. 2002 HST Calibration Workshop, Hubble after the Installation of the ACS and the NICMOS Cooling System. Space Telescope Science Institute, Baltimore, MD, p. 337
Kuijken K., 2006, A&A, 456, 827
Leauthaud A. et al., 2007, ApJS, 172, 219
Leonard A., Goldberg D. M., Haaga J. L., Massey R., 2007, ApJ, 666, 51
Leonard A., King L. J., Goldberg D. M., 2011, MNRAS, 413, 789
Luppino G. A., Kaiser N., 1997, ApJ, 475, 20
Lupton R., 1993, Statistics in Theory and Practice. Princeton Univ. Press, Princeton, NJ
Mandelbaum R., Hirata C. M., Leauthaud A., Massey R. J., Rhodes J., 2012, MNRAS, 420, 1518
Marks R. J., 2009, Handbook of Fourier Analysis and Its Applications, 1st edn. Oxford Univ. Press
Massey R., Refregier A., 2005, MNRAS, 363, 197
Massey R., Refregier A., Conselice C. J., Bacon D. J., 2004, MNRAS, 348, 214
Massey R. et al., 2007a, MNRAS, 376, 13
Massey R., Rowe B., Refregier A., Bacon D. J., Bergé J., 2007b, MNRAS, 380, 229 (M07)
Massey R. et al., 2007c, Nat, 445, 286
Melchior P., Viola M., 2012, MNRAS, 424, 2757
Melchior P., Meneghetti M., Bartelmann M., 2007, A&A, 463, 1215
Melchior P., Andrae R., Maturi M., Bartelmann M., 2009, A&A, 493, 727
Melchior P., Viola M., Schäfer B. M., Bartelmann M., 2011, MNRAS, 412, 1552
Meneghetti M. et al., 2008, A&A, 482, 403
Miller L., Kitching T. D., Heymans C., Heavens A. F., van Waerbeke L., 2007, MNRAS, 382, 315
Miller L. et al., 2013, MNRAS, 429, 2858
Nakajima R., Bernstein G., 2007, AJ, 133, 1763
Okura Y., Umetsu K., Futamase T., 2007, ApJ, 660, 995
Okura Y., Umetsu K., Futamase T., 2008, ApJ, 680, 1
Paulin-Henriksson S., Amara A., Voigt L., Refregier A., Bridle S. L., 2008, A&A, 484, 67
Paulin-Henriksson S., Refregier A., Amara A., 2009, A&A, 500, 647
Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, Numerical Recipes in FORTRAN. The Art of Scientific Computing, 2nd edn. Cambridge Univ. Press
Refregier A., 2003, MNRAS, 338, 35
Refregier A., Bacon D., 2003, MNRAS, 338, 48
Refregier A., Kacprzak T., Amara A., Bridle S., Rowe B., 2012, MNRAS, 425, 1951
Rhodes J., Refregier A., Groth E. J., 2000, ApJ, 536, 79
Rhodes J. D. et al., 2007, ApJS, 172, 203
Rix H.-W. et al., 2004, ApJS, 152, 163
Rowe B., 2010, MNRAS, 404, 350
Rowe B., Hirata C., Rhodes J., 2011, ApJ, 741, 46
Schneider P., 2006, in Meylan G., Jetzer P., North P., Schneider P., Kochanek C. S., Wambsganss J., eds, Saas-Fee Advanced Course 33: Gravitational Lensing: Strong, Weak and Micro. Springer, Berlin, p. 269
Schneider P., Er X., 2008, A&A, 485, 363
Schneider P., Seitz C., 1995, A&A, 294, 411
Schrabback T. et al., 2007, A&A, 468, 823
Schrabback T. et al., 2010, A&A, 516, A63
Scoville N. et al., 2007, ApJS, 172, 1
Van Waerbeke L. et al., 2013, MNRAS, 433, 3373
Velander M., Kuijken K., Schrabback T., 2011, MNRAS, 412, 2665
Viola M., Melchior P., Bartelmann M., 2011, MNRAS, 410, 2156
Viola M., Melchior P., Bartelmann M., 2012, MNRAS, 419, 2215
Voigt L. M., Bridle S. L., 2010, MNRAS, 404, 458
Zuntz J., Kacprzak T., Voigt L., Hirsch M., Rowe B., Bridle S., 2013, preprint (astro-ph/1302.0183)

APPENDIX A: SHAPELET GALAXY MODELS FROM THE HUDF

We model the HUDF galaxies using version 2.2 of the SHAPELETS software package, presented by Massey & Refregier (2005) and described in practical detail by Bergé (2006). The modelling of these galaxies, including PSF deconvolution, is a multistage process: catalogue creation, postage stamp image creation and PSF modelling all precede the construction of the deconvolved shapelet catalogue for the starter set. These processes are now described in turn.

A1 Star and galaxy selection

The starting point for building accurate models of galaxy images is an object catalogue. A primary science goal of the Beckwith et al. (2006) analysis was a catalogue of full, multicolour photometry in the HUDF, but we require only reliable object detection in the V_{606} in order to build galaxy models in the same filter. The Massey & Refregier (2005) SHAPELETS software package requires certain input parameters for each object that were not *all* provided in the catalogues of Beckwith et al. (2006). We chose to construct our own catalogue from just the V_{606} data, using the SExtractor software (Bertin & Arnouts 1996; version 2.5.2) in single-image mode. Using this software successfully requires a number of choices regarding configuration parameters, and we now describe the strategy adopted for object detection in the HUDF V_{606} data. We note that this same strategy will also be used in Section 6.1 as part of the lensing measurement pipeline being tested, but for simulated imaging data at a much shallower depth when compared to the HUDF data used to provide galaxy models. We adopt a two-pass SExtractor deblending strategy when constructing a source catalogue (see e.g. Rix et al. 2004; Leauthaud et al. 2007; Caldwell et al. 2008). Two catalogues are created, one with a low detection threshold so as to pick out as many faint objects as possible, and one using a more conservative detection strategy so as to limit the overdeblending of bright objects; these catalogues will be referred to as 'hot' and 'cold', respectively. These are created using exactly the same input parameter values as used by Caldwell et al. (2008) to create the hot and cold samples of objects in the GEMS V_{606} science tiles, summarized in Table A1.

All cold detections are then combined with non-overlapping objects in the hot catalogue. We define a hot object as overlapping if its centroid lies within an ellipse of semimajor axis $5.5 \times a$ and semiminor axis $5.5 \times b$, where a and b are the SExtractor-output semimajor and semiminor axes, respectively, and this larger ellipse

Table A1. SExtractor configuration parameters used to detect galaxy objects in both the HUDF (Section 4) and in the simulated ACS images (Section 6); these values are the same as used for the GEMS survey galaxies in the two-pass strategy of Rix et al. (2004) and Caldwell et al. (2008).

Configuration parameter	Cold sample	Hot sample
DETECT_THRESH	2.30	1.4
DETECT_MINAREA	100	37
DEBLEND_MINCONT	0.065	0.060
DEBLEND_NTHRESH	64	32
BACK_SIZE	214	214
BACK_FILTERSIZE	5	5

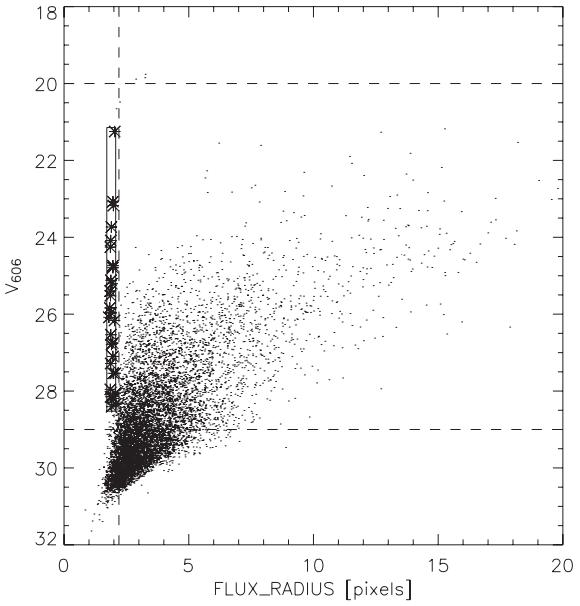

Figure A1. Size–magnitude diagram for SExtractor-selected objects in the HUDF V_{606} science image, showing the stellar locus and 30 selected stars (star-shaped points). The dashed lines show the size and magnitude limits used to define the galaxy sample chosen for shapelet modelling for inclusion in the simulation starter set.

is aligned with that defined by SExtractor. This factor of 5.5 was found, by visual inspection of the HUDF and segmentation maps output by SExtractor, to provide a suitable compromise between cold object overdeblending and the erroneous removal of hot objects. Finally, a mask is applied so as to exclude detections from the boundary regions of the CCD image. The combined hot/cold catalogue then contains a total of 8203 objects, corresponding to $\sim$900 detections arcmin^{-2}.

To select stars, we use the fact that the FLUX_RADIUS parameter found by SExtractor is typically constant, irrespective of flux, for stellar images. This radial profile of the PSF will typically vary little across a given single ACS tile (the same is therefore also true of the object FWHM). This allows the stars to be easily identified via a straight locus on a size–magnitude diagram, such as that showing all the masked HUDF-selected objects in Fig. A1. The locus chosen in the figure gives a total of 30 stellar images from which to build a model of the HUDF PSF, avoiding the confused region at greater magnitude and saturated images at lower magnitude. This low number of stars is to be expected: the HUDF was specifically chosen as a direction out of the plane of the Milky Way containing as few stars as possible. These 30 stars will be decomposed into shapelet models of the HUDF PSF in Appendix A3.

To isolate galaxies for shapelet modelling in Appendix A4, those to be included in the simulation starter set, we cut for objects with $20.0 < V_{606} < 29.0$ and $2.2 <$ FLUX_RADIUS < 150.0 (in pixels). These cuts can also be seen in Fig. A1 (barring the large-radius cut at FLUX_RADIUS $= 150.0$). There are then 4128 galaxies from the original masked sample that make these cuts, corresponding to approximately 460 galaxies arcmin^{-2}. It is this extremely deep sample that will be used for generating simulated galaxy images, although of course many of these objects will be lost in noise when simulating shallower data than the HUDF.

A2 Postage stamp image extraction

The decomposition of stars and galaxy images into shapelet models must be preceded by the creation of 'postage stamp' images of each object in the catalogue, cropped around the object in question, and masked for neighbours. A postage stamp is also made containing a map of the noise and sky background in the same vicinity.

For each object, the SHAPELETS software draws circular postage stamps centred on the SExtractor-measured centroid. The radius of this circular image is the integer number of pixels closest to a value $r_{\rm PS}$, defined in terms of the user-specified SHAPELETS input parameter NFWHM as $r_{\rm PS} = \mathrm{NFWHM} \times a + 4$, where a is the SExtractor-output semimajor axis of the object. The name of the parameter NFWHM appears to be from an earlier incarnation of the software in which $r_{\rm PS}$ was defined in terms of the FWHM. We choose NFWHM $= 6$ in this analysis, the default value of 5 being found to be often insufficiently large to allow the shapelet modelling of the extended outer profiles of deep, space-based images.

The initial version of the SHAPELETS software used in this analysis flagged as a modelling failure any objects for which the effective outer boundary of the shapelet model, defined as a locus of radius $\theta_{\rm max} = \beta\sqrt{n_{\rm max} + 1}$ (see Massey & Refregier 2005) around the object centroid, extended beyond the edge of the postage stamp. Small postage stamps led to an unacceptable number of model failures due to this extension of light profiles beyond the postage stamp boundaries, but it was found that drawing overly large postage stamps (NFWHM ≥ 7) around every object was computationally prohibitive. An algorithm for iteratively redrawing the postage stamp in the event of such model failure provided an efficient solution to this problem, and is now part of the SHAPELETS software, documented and available for download online at the web location given in footnote 1. In the iterative prescription used for this analysis, model failures due to postage stamp outgrowth are resubmitted using a new postage stamp that is increased in size by a factor REDRAW_FACTOR $= 1.3$. This process is repeated up to a maximum of MAX_N_REDRAWS $= 5$ times, after which a catastrophic failure is flagged. In tests, these parameter choices were found to give a better compromise between modelling success rates and computation time than other values tried: the number of floating point operations required to generate images approximately varies $\propto r_{\rm PS}^2$, so REDRAW_FACTOR and MAX_N_REDRAWS cannot be made arbitrarily large without computational cost.

After the drawing of a postage stamp around each object of interest, a fundamental consideration in the shapelet approach is then the masking of other, nearby objects: failure to do this well will often result in the partial modelling of nearby objects as part of the object of interest. Unlike in the KSB approach, image pixels at a distance from the object centroid are not explicitly down-weighted

increasing the importance of careful masking. The SHAPELETS software therefore constructs over each object an elliptical mask defined with semimajor axis MASK_NEIGH $\times$ a and semiminor axis MASK_NEIGH $\times$ b, aligned with the SExtractor-defined object ellipse. The factor MASK_NEIGH $= 4$ was found to give better results than the shapelet default value of 2.75: in the default setting there were often portions of the outer galaxy light profile that were unmasked and clearly visible. These caused an enhanced rate of catastrophic modelling failure for the central galaxy of interest in such postage stamps. The larger value MASK_NEIGH $= 4$ provided a significant reduction in such failures without generating unacceptable numbers of cases where the nearby mask obscures the object of interest in the postage stamp.

An inverse-variance noise weight map and an estimate of the sky background level are then made via analysis of blank sky pixels in the postage stamp, i.e. those unmasked by the central object or a neighbour. The sky background can be subtracted from the image postage stamp by fitting a choice of surfaces to blank sky pixels (Bergé 2006). For the HUDF V_{606} image, only a very small amount of residual sky background variation was found and the removal of a simple constant sky level from each postage stamp was sufficient, achieved by setting the SHAPELETS input parameter SKY $= 1$. In order to make the noise map, we estimate the root-mean-square (rms) blank sky pixel value to provide a constant, inverse-variance weight.

There is a further choice in how noise values are assigned for pixels corresponding to masked neighbours, a choice controlled by the input parameter NEIGHBOUR. For the default value NEIGHBOUR $= 0$, these pixels are assigned zero values in the inverse-variance weight map, and are therefore not considered in the shapelet modelling. For NEIGHBOUR $= 1$, the pixels are assigned the same weight as elsewhere in the noise map and set to the background level in the science image. Although it is arguably better not to include these masked pixels in any fit for shape inference purposes, for the purposes of building a simulation galaxy starter set from HUDF images it was found that NEIGHBOUR $= 0$ sometimes caused large negative-flux patches in shapelet models for unconstrained regions of the image. So as to build as physically representative a starter set as possible, we therefore set NEIGHBOUR $= 1$ for the modelling of HUDF objects, but retain NEIGHBOUR $= 0$ when later testing the shapelet lensing analysis on the derived, noisier, simulated galaxies.

Finally, after these choices for the construction of the postage stamps, the trimmed, masked, sky-subtracted science images and inverse-variance noise weight maps are then ready to be supplied to the shapelet decomposition and modelling routines as described in Massey & Refregier (2005). We now describe the use of this software to model first the stars in the HUDF, and then the deconvolved, high-resolution galaxy images that will be used in the flexion and shear simulations.

A3 Modelling the HUDF PSF

Modelling the PSF in the HUDF is important. A good PSF model allows an approximate deconvolution of the PSF from galaxy images, desirable for generating a starter set of models which accurately reflects real galaxy properties for space-based data (see e.g. Mandelbaum et al. 2012, who demonstrate a novel approach to simulating convolution-corrected galaxy images). An alternative approach is simply to model the ACS PSF-*convolved* galaxy images of the HUDF and use these as the starter set: such an approach would be acceptable if the final use of the starter set was in simulating observations with far larger PSF sizes, such as for ground-based data (this was the approach taken in STEP2; Massey et al. 2007a). However, when simulating ACS data such an approach would lead to an unrealistic size distribution for the small, fainter objects that are the most important carriers of weak lensing information. Fainter galaxies, being not much larger than the $\simeq$0.1 arcsec PSF typical in ACS images, would be noticeably oversmoothed, too large, and with an unrepresentative radial profile. We therefore correct for this image blurring as much as possible.

Taking the 30 stellar objects selected as described in Appendix A1 (see also Fig. A1), we first create masked image and noise postage stamps as described in Appendix A2. We then construct shapelet models of each star, choosing fixed values of $\beta = 1.80$ and $n_{\max} = 20$. The fact that these values are fixed, and not allowed to vary as under the amoeba-driven optimization described by Massey & Refregier (2005) is important: we wish to combine shapelet models to create an average PSF for the field, so a fixed $n_{\max}$ and β are practical as they allow simple linear co-addition of models. The SHAPELETS software also outputs the diagonal entries of the covariance matrix for all shapelet coefficients, and so these are used to combine the models for each star in an inverse-variance weighted average, giving the resulting PSF model seen in Fig. A2.

That this model incorporates no spatial variation across the HUDF field-of-view is not a great concern for the purposes of creating a realistic starter set for lensing simulations. Although the bulk effect of image convolution will be largely corrected for, it does mean that some residual ellipticity and flexion will remain in the starter set images due to the residual anisotropy variation in the HUDF PSF. These faint distortions will therefore be retained in the starter galaxy models, but as will be seen in Section 5 these starter models are then randomly rotated, inverted and further distorted before being used in simulations. Within these simulations they are then lensed, reconvolved with a new PSF and significantly noise-degraded. Distortions due to variation in the original HUDF PSF will not have a significant, coherent impact on the final galaxy images at the level of measurement possible for simulations of this size.

A4 Modelling the HUDF galaxies

In constructing deconvolved (i.e. corrected for PSF convolution) shapelet models of the HUDF galaxies, the Massey & Refregier (2005) software takes as its inputs the object catalogue constructed as described in Appendix A1, the accompanying postage stamp images and noise maps, and the shapelet model of the PSF. The best-fitting shapelet models output by the code will make up the starter set that is used to create simulated galaxies.

However, as was necessary for the modelling of the HUDF PSF, choices must be made for input parameter values that govern how these best-fitting shapelet models are selected. The most important of these are summarized in Table A2, and now described.

Unlike when modelling the PSF, for constructing galaxy models we allow $n_{\max}$ and β to vary. These quantities are chosen by an amoeba-driven optimization process (described by Massey & Refregier 2005; we note that other implementations of the shapelet method do not necessarily allow $n_{\max}$ to vary, e.g. Kuijken 2006). However, the SHAPELETS software does require the choice of a limiting maximum value of $n_{\max} =$ N_MAX, and a lower starting point N_MIN from which to begin the amoeba search.

The choice of N_MAX is largely motivated by computing resource constraints. In general, a significant fraction of overall processing time is spent modelling a very small subset of

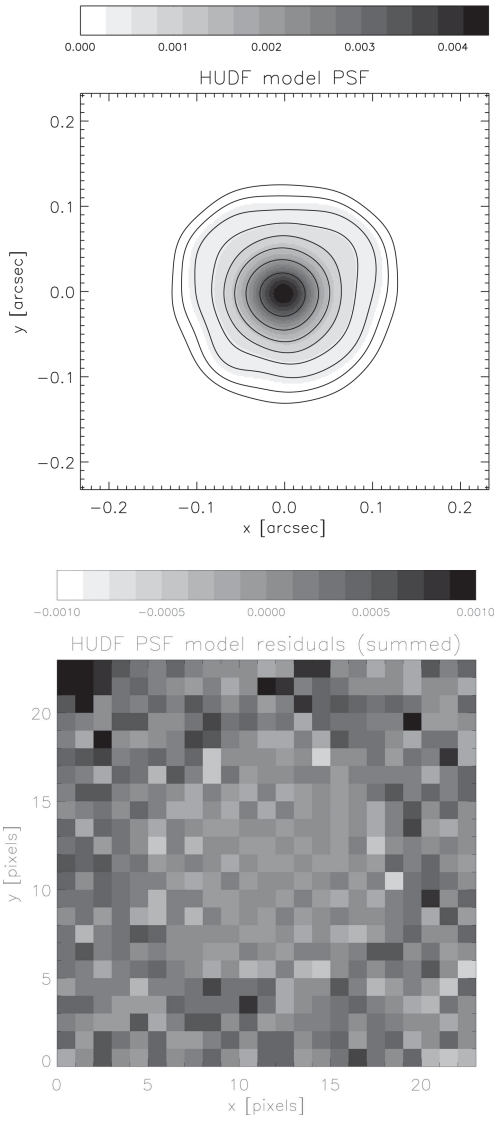

Figure A2. Upper panel: PSF pattern created from the weighted average of 30 shapelet models of selected stars in the HUDF V_{606} science image. The grey-scale is linear in surface brightness whereas the contours are logarithmic. Lower panel: mean residuals from shapelet model fits to the HUDF stars. A flux excess beyond the central region of the postage stamp is visible. There are two primary contributions to this excess in the residuals: the inability of the shapelet model to fully represent the extended wings of the HUDF PSF, and light from nearby objects not being fitted in the modelling.

large/bright galaxies with complex structure. The time taken and memory required to model a given object increases roughly as $n_{\max}^4$ (Massey & Refregier 2005). Above $n_{\max} = 20$, the calculation of pre-multipliers for the shapelet basis functions (see equation 16) causes numerical overflow in the unsigned, 64-bit integers the SHAPELETS software uses for the calculation and tabulation of factorial terms in the numerator and denominator of $P_{n,m}$. For $n_{\max} > 20$, the SHAPELETS software therefore calculates these numbers as double-precision floating point numbers, as and when they are required. This leads to significant processing overheads, albeit ones which might plausibly be averted with some changes to the internal processing design of the SHAPELETS software itself.

Table A2. SHAPELETS software input parameters. Upper section: input parameter settings that influence the construction of postage stamps. Lower section: parameters that control the shapelet modelling and deconvolution. No default value is indicated for parameters added as a modification to the publicly available software.

SHAPELETS input parameter	Chosen value	Default value
NFWHM	6	5
REDRAW_FACTOR	1.3	–
MAX_N_REDRAWS	5	–
MASK_NEIGH	4.0	2.75
NEIGHBOUR	1^a, 0^b	0
SKY	1	0
N_MIN (minimum $n_{\max}$)	0	2
N_MAX (maximum $n_{\max}$)	20	20
THETA_MIN_GEOM	0.5^a, 1.0^b	0.2

[a]Values chosen for modelling HUDF galaxies
[b]Values chosen for modelling the ACS simulated galaxies

However, the largest, brightest and best resolved galaxies are also the most likely to be local and only *very* weakly lensed in real data, and this reduces the motivation for dedicating significant amounts of development effort and processing time to producing high-order shapelet models of this population. We therefore choose N_MAX=20, which nonetheless allows a high degree of galaxy substructural complexity to be realized (e.g. Fig. 5). We choose N_MIN =0 so as to give the SHAPELETS code freedom to model galaxies with a simple circular Gaussian if there is no strong statistical support for a more complex model.

The other input parameter chosen to differ from the default choice was THETA_MIN_GEOM = 0.5. This parameter sets a lower limit on the quantity defined by Massey & Refregier (2005) as $\theta_{\min} = \beta/\sqrt{n_{\max}+1}$, which is the minimum geometric scale on which the shapelet model varies. The requirement upon the shapelet modelling amoeba of only accepting models with $\theta_{\min} \geq 0.5$, also suggested by Melchior et al. (2007), prevents excessive subpixel modelling that can lead to unphysical 'ringing' in shapelet models of galaxy images.

The best-fitting galaxy models resulting from these chosen input parameters, postage stamps and HUDF PSF model are described in Section 4.

APPENDIX B: TESTING IMAGE OPERATIONS

B1 Weak lensing distortions

When performing real-space raytracing as described in Section 5.2, we use the lens equation to map from the image pixel positions back to a non-regular grid of positions in the source plane. At these points, we sample the surface brightness from the shapelet model of the source galaxy $I^{(s)}$. Ignoring the overall deflection of galaxy images (equivalent to setting $\boldsymbol{\theta}_c = \boldsymbol{\theta}'_c = 0$), and using equation (3), we may write

$$I(\boldsymbol{\theta}) = I^{(s)}(\boldsymbol{\theta}') = I^{(s)}(A_{i,j}\theta_j + D_{ijk}\theta_j\theta_k/2). \quad \text{(B1)}$$

In practical terms, the raytracing scheme is simple: we assign $I(\boldsymbol{\theta})$ for each desired image pixel position $\boldsymbol{\theta}$ by taking the shapelet model value of $I^{(s)}$ at the position $\boldsymbol{\theta}' = \boldsymbol{\theta}'(\boldsymbol{\theta}, \gamma, \mathcal{F}, \mathcal{G})$.

Square pixels in the image plane do not map to square pixels in the source plane, and when flexion is included with shear, the

mapped pixel boundaries become curved in the source plane. This means that it is no longer possible to use the results of Massey & Refregier (2005) to perform the exact flux integral of $I(\boldsymbol{\theta})$ across each pixel. In the simplest approximation, one can adopt the pixel-centre surface brightness to estimate flux but more accurate results can be achieved by *upsampling*, i.e. by creating a higher resolution image and then summing values at the high-resolution subpixel locations to estimate the true integral. This allows approximation of the exact shear and flexion transformation to an accuracy that depends on the degree of upsampling adopted.

We can only upsample by a finite, preferably integral, factor. In order to understand what factor is necessary in our simulations, we now test its impact upon fundamental lensing measures (e.g. image moments) for noise-free ACS galaxy models.

We first extracted a random sample of 1000 galaxies from the HUDF starter set. For each galaxy, a control postage stamp image was constructed by integrating across ACS imaging survey-size, post-drizzling pixels (0.03 arcsec, as used in the HUDF, COSMOS, GEMS and STAGES final science images; see Beckwith et al. 2006; Leauthaud et al. 2007; Caldwell et al. 2008; Gray et al. 2009) using the exact Massey & Refregier (2005) analytic results. We refer to this 'true' pixelized control image as $I^{(t)}$. Simple, unweighted moments for such images may be defined as follows:

$$S = \int \mathrm{d}^2\theta I(\boldsymbol{\theta}), \tag{B2}$$

$$q_{ij} = \frac{1}{S}\int \mathrm{d}^2\theta \Delta\theta_i \Delta\theta_j I(\boldsymbol{\theta}), \tag{B3}$$

$$q_{ijk} = \frac{1}{S}\int \mathrm{d}^2\theta \Delta\theta_i \Delta\theta_j \Delta\theta_k I(\boldsymbol{\theta}), \tag{B4}$$

$$q_{ijkl} = \frac{1}{S}\int \mathrm{d}^2\theta \Delta\theta_i \Delta\theta_j \Delta\theta_k \Delta\theta_l I(\boldsymbol{\theta}), \tag{B5}$$

where $\boldsymbol{\Delta\theta} = \boldsymbol{\theta} - \boldsymbol{\theta}_c$ as before. In the noise-free case, the unweighted moments can be used to construct complex polarization measures that provide estimators of shear and flexion (e.g. Kaiser et al. 1995; Bartelmann & Schneider 2001; Okura et al. 2007).

We construct the unweighted polarization

$$e = e_1 + \mathrm{i}e_2 = \frac{q_{11} - q_{22} + 2\mathrm{i}q_{12}}{q_{11} + q_{22}}, \tag{B6}$$

(see e.g. Kaiser et al. 1995; Bartelmann & Schneider 2001) and, following Okura et al. (2007), we define

$$f = f_1 + \mathrm{i}f_2 = \frac{q_{111} + q_{122} + \mathrm{i}(q_{112} + q_{222})}{q_{1111} + 2q_{1122} + q_{2222}}, \tag{B7}$$

$$g = g_1 + \mathrm{i}g_2 = \frac{q_{111} - 3q_{122} + \mathrm{i}(3q_{112} - q_{222})}{q_{1111} + 2q_{1122} + q_{2222}} \tag{B8}$$

as equivalent measures for $\mathcal{F}$ and $\mathcal{G}$, respectively. We require that any finite degree of upsampling must cause fractional errors in e, f and g that are significantly smaller than those we expect due to noise in the final simulation results.

The first step in the test is to calculate control values $e^{(t)}, f^{(t)}$ and $g^{(t)}$ using the moments as defined in equations (B2)–(B8) for each control image $I^{(t)}$ from the sample of HUDF starter galaxies. We then make a series of ray-traced estimates I_r of $I^{(t)}$ by performing a numerical integral of the flux across the upsampled pixels. We define the upsampling ratio r as the ratio between the linear scale of output pixels and that of the subpixels in the upsampled image: $r = 1$ is equivalent to simply taking the central pixel value; $r = 2$ is equivalent to four equally spaced subpixels; etc. For each galaxy I_r, we measure a corresponding e, f and g, and calculate the fractional deviation in each component:

$$\frac{\delta e_i}{e_i} = \frac{e_i - e_i^{(t)}}{e_i}, \tag{B9}$$

with equivalent expressions for f and g, for $r = 1, \ldots, 7$. In Fig. B1, we plot the median of these fractional deviations across the sample of galaxies, as a function of r.

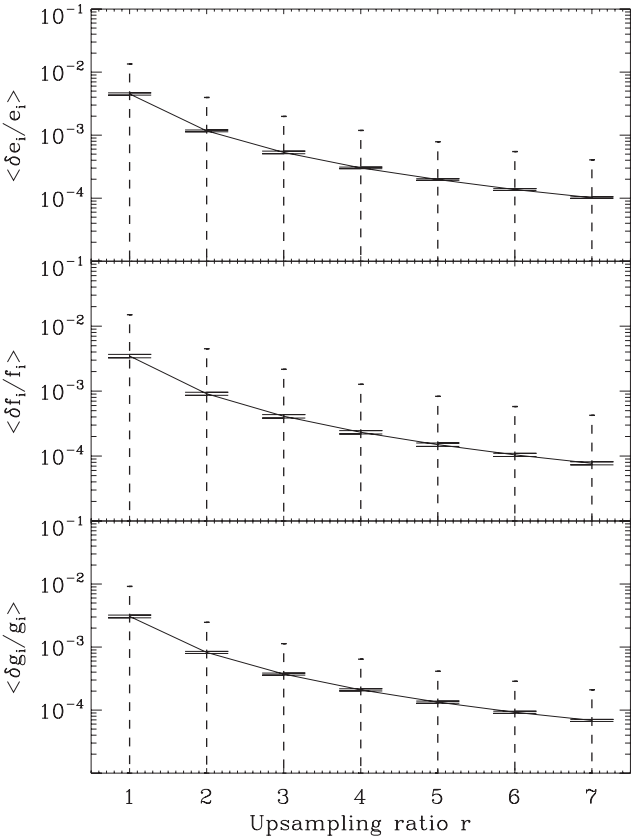

Figure B1. Testing the impact of upsampling ratio for generating images of lensed (by shear and flexion) galaxies using multiple 'ray-traced' image samples. Plot of median fractional error in simple unweighted e (top panel), f (centre panel) and g (bottom panel) estimators, for a sample of 1000 galaxies randomly selected from the starter set, with increasing upsampling ratios r as described in Appendix B1. Each δe_i, δf_i, etc. is calculated by comparing the upsampling-derived value relative to the exact (analytically integrated) value. The wide solid error bars on each point give the standard error on the median (Lupton 1993), whereas the dashed error bars illustrate the typical range of the effect as described by the NMAD of the fractional error.

As a measure of the typical range for the error due to finite upsampling, in Fig. B1 we also plot as dashed line error bars, the NMAD of the fractional error. The NMAD is a robust measure of the width of a distribution of some variable x, defined as

$$\mathrm{NMAD}(x) \simeq 1.4826 \times \mathrm{MAD}(x) \tag{B10}$$

$$= 1.4826 \times \mathrm{Median}\left(|x - \mathrm{Median}\,(x)|\right), \tag{B11}$$

where the MAD is simply the median absolute deviation, and the constant scale factor 1.4826 ensures that the NMAD of a

Gaussian distribution is approximately equal to its standard deviation, NMAD $\simeq \sigma$. Fig. B1 shows the NMAD for $\delta e_i/e_i$, $\delta f_i/f_i$ and $\delta g_i/g_i$ at each value of r, to give an idea of the typical range in fractional errors for these quantities due to the use of finite element upsampling techniques. It should be noted that median statistical measures were used because the property of interest, which in this case is specifically the fractional uncertainty on an ellipticity or flexion measurement, leads to large outlier values when there are small values of e_i, f_i or g_i in the denominator. Adopting median statistics downweights these noisy outliers to reduce overall statistical uncertainty on the central tendency. As a check on the results, simple means were also used and resulted in a consistent (but rather noisier) trend.

The results of Fig. B1 are encouraging and as expected: there is clear convergence towards smaller and smaller fractional errors with increasing r. Practically, the results imply that for the HUDF galaxies and an output pixel scale of 0.03 arcsec, we need only upsample galaxies by a factor of 2 in order to give output images that are typically accurate in e, f and g to within a factor of $\sim 10^{-3}$ of the correct value (a target value motivated by the results of, e.g., Amara & Réfrégier 2008). Interestingly, Fig. B1 shows that good accuracy may be achieved for e, f and g, even when $r = 1$, for which the error in the idealized measures is often less than 1 per cent of the underlying value.

B2 Convolution

To test the accuracy of numerical convolution, we use the methods developed in Appendix B1. We again select a random sample of 1000 HUDF galaxies from the starter set. For each galaxy, we construct a set of seven images, upsampled by a linear factor $r = 1, \ldots, 7$, and perform a convolution via FFT using an image of the circularized GEMS PSF (Fig. 3) upsampled by the same factor. The convolved image is then summed back to the final resolution of 0.03 arcsec pixel^{-1}, and then each of e, f and g are measured.

Because there is no practical way of generating the 'true' convolved image in this manner as a control (as discussed before the exact shapelet treatment is computationally unfeasible), we instead take the high-upsampling case of $r = 11$ as the reference point. The fractional deviation from this $r = 11$ is then calculated for each galaxy and each r, and in Fig. B2 we plot the resulting median values and range as in Appendix B1. As was found for the shear and flexion upsampling tests, the results of Fig. B2 are encouraging: an upsampling ratio of only $r = 2$ allows a systematic fractional error in the estimates of e, f and g that is typically 0.1 per cent or less. Interestingly, the effect is in the opposite direction to that induced when performing raytracing lensing transformations. The effect of a finite sampling approximation to convolution is to cause a net reduction in these moments of a galaxy image, whereas a finite sampling approximation to the lens mapping artificially exaggerates these moments. The overall net effect of the two approximations will thus, on average, be typically smaller than that described by either of Figs B1 or B2 in isolation.

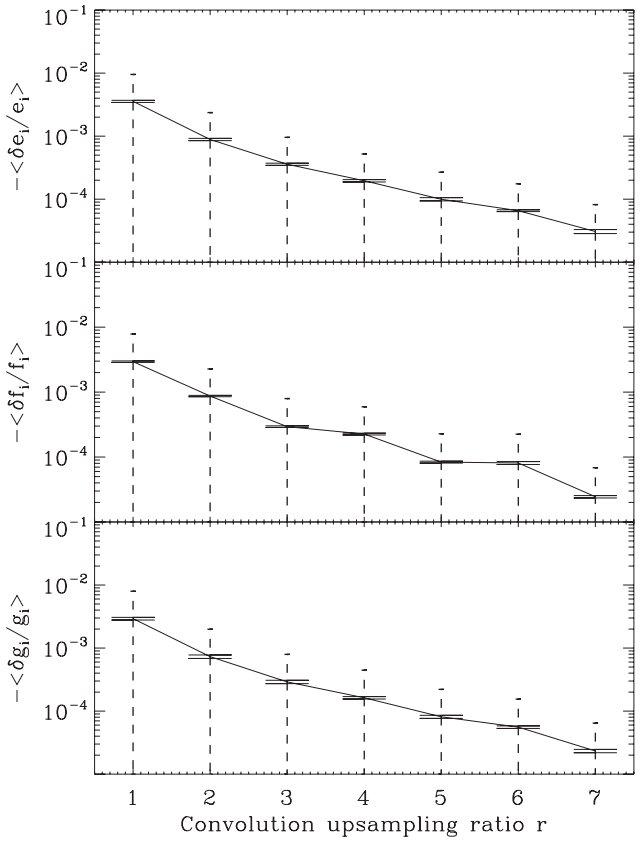

Figure B2. Testing the impact of upsampling ratio in numerical convolution via discrete Fourier transforms, using FFT. Plot of median fractional error in simple unweighted e (top panel), f (centre panel) and g (bottom panel) estimators, for a sample of 1000 galaxies randomly selected from the starter set, after performing a convolution using discrete Fourier transforms at an upsampling ratio r as described in Appendix B2. Each δe_i, δf_i, etc., is calculated by comparing the ovesampling-derived value relative to the value found for $r = 11$. The wide solid error bars on each point give the standard error on the median (Lupton 1993), whereas the dashed error bars illustrate the typical range of the effect as described by the NMAD of the fractional error.

APPENDIX C: SHAPELET ESTIMATORS OF FLEXION AND SHEAR

A variety of polar shapelet estimators for shear and flexion are described by M07. In theory, the number of lensing estimators that may be constructed using shapelets is only limited by the number of shapelet models available, $n_{\max}$. However, the majority of shapelet models will be highly truncated to somewhat low $n_{\max}$ in practice. Estimators which make extensive use of higher order information prove problematic for many galaxy images, particular those which rely upon the convergence (in the sense of converging to a limit) of sums over shapelet coefficients (see M07).

In this paper, we limit our investigations of shear and flexion measurement primarily to the most simple, and therefore widely available in noisy galaxies, shapelet estimators. We now describe construction of useful estimators based on the results of M07, taking into account important practical realities such as the wide dynamic range of galaxy fluxes and sizes.

C1 Normalizing shapelet coefficients by galaxy size and flux

The simplest shear estimator is that which employs the polar shapelet coefficient $f_{2,2}$ as the galaxy polarization estimator (see M07, section 3, for the precise usage of this term). This quantity can in fact be straightforwardly shown to correspond to a Gaussian-weighted quadrupole moment of the shapelet model, with a weighting radius of β.

The raw form of the simplest well-motivated estimator that can be constructed from $f_{2,2}$ is as follows:

$$\tilde{\gamma}^{\rm Gauss} = \sqrt{2}\frac{f_{2,2}}{\langle f_{0,0} - f_{4,0}\rangle}, \tag{C1}$$

where the expectation notation used in the denominator denotes an estimate of the ensemble average over all galaxies in the source sample (Refregier & Bacon 2003; M07).

However, we have found that this estimator suffers an important disadvantage: both the numerator and denominator share the dimensionality of the shapelet coefficients themselves, scaling with both object flux and inverse scale size β (Massey & Refregier 2005). This means that estimators of shear (and similarly constructed flexion estimators) vary strongly in magnitude as a function of the object size and flux relative to those of the ensemble average.

It is therefore desirable to generalize the estimator of equation (C1) to create a new estimator we label $\tilde{\gamma}_\nu$, defined as

$$\tilde{\gamma}_\nu = \sqrt{2}\frac{\nu f_{2,2}}{\langle \nu(f_{0,0} - f_{4,0})\rangle}, \tag{C2}$$

where the parameter ν is suitably chosen to be a property of the galaxy image that helps normalize terms in the numerator and denominator, helping to lessen the impact of the wide dynamic ranges of flux and scale length in galaxy samples. This is analogous to the adoption of a normalizing factor $1/S$ in the definition of the image moments q_{ij}, q_{ijk} and q_{ijkl} in Appendix B1. It should be stated from the outset that if ν is an estimated property of the image itself it may add both uncertainty and bias to the shear estimator. We will discuss how this danger weighs against the benefits of such a normalization later in Appendix C3.

The simplest possible polar shapelet flexion estimators can be constructed from similar combinations of shapelet coefficients. The first-order $\mathcal{F}$ estimator can be expressed as

$$\tilde{\mathcal{F}}^{\rm Gauss} = \frac{4}{3\beta}\frac{f_{1,1}}{\langle d_{1,1}\rangle}, \tag{C3}$$

where

$$d_{1,1} = \left(1 - \frac{R^2}{\beta^2}\right) f_{0,0} + \frac{R^2}{\beta^2} f_{2,0} + \frac{5\sqrt{2}R^2}{6\beta^2}\varepsilon^* f_{2,2} - f_{4,0}. \tag{C4}$$

The shapelet model parameters ε and R^2 are defined in Massey & Refregier (2005) as

$$R^2 = \frac{\sqrt{16\pi}\beta^3}{F}\sum_n^{\rm even}(n+1) f_{n,0}, \tag{C5}$$

$$\varepsilon = \frac{\sqrt{16\pi}\beta^3}{R^2 F}\sum_n^{\rm even}[n(n+2)]^{1/2} f_{n,2}, \tag{C6}$$

where F denotes the shapelet model flux,

$$F = \beta\sqrt{4\pi}\sum_n^{\rm even} f_{n,0}, \tag{C7}$$

and where all sums are over all even n coefficients $f_{n,0}$ and $f_{n,2}$ in the model ($m = 0$ and $m = 2$ coefficients do not exist for odd n). We note that ε and R^2 appear only in those terms necessary to correctly account for the flexion centroid shift at linear order in the applied $\mathcal{F}$ (see Goldberg & Bacon 2005; M07). It should also be noted that $d_{1,1}$ in equation (C4) contains a term proportional to $\varepsilon^* \times f_{n,2}$, for $n = 2$; in M07 these were argued to vanish due to rotational symmetry in the source plane. We argue that this term should not be omitted since the definition of ε above shows the inclusion of $f_{2,2}$ as the first coefficient in the sum. For most galaxies, lower-n components of the shapelet model will dominate those with higher-n, so that as a first approximation $\varepsilon^* \times f_{2,2} \sim |f_{2,2}|^2$. This does not cancel due to rotational symmetry across a population of galaxies.

For the first-order $\mathcal{G}$ estimator, we have the expression

$$\tilde{\mathcal{G}}^{\rm Gauss} = \frac{4\sqrt{6}}{3\beta}\frac{f_{3,3}}{\langle d_{3,3}\rangle}, \tag{C8}$$

where

$$d_{3,3} = \left(f_{0,0} + f_{2,0} - f_{4,0} - f_{6,0}\right). \tag{C9}$$

Analogously to the case of $\tilde{\gamma}^{\rm Gauss}$, these estimators of both $\mathcal{F}$ and $\mathcal{G}$ are directly related to Gaussian-weighted octupole moments (e.g. Okura et al. 2007) of the shapelet model galaxy.

As for shear, the large dynamic range in the shapelet coefficients due to varying galaxy size and flux suggests it may be useful to rescale numerator and denominator with a generalizing factor ν, defining

$$\tilde{\mathcal{F}}_\nu = \frac{4}{3\beta}\frac{\nu f_{1,1}}{\langle \nu d_{1,1}\rangle}, \tag{C10}$$

$$\tilde{\mathcal{G}}_\nu = \frac{4\sqrt{6}}{3\beta}\frac{\nu f_{3,3}}{\langle \nu d_{3,3}\rangle}. \tag{C11}$$

We now discuss options for a suitable choice of this normalization parameter ν.

C2 Choosing a suitable ν

Adopting $\nu = 1$ recovers the M07 $\tilde{\gamma}^{\rm Gauss}$ estimator. Natural alternative choices for ν are combinations of shapelet model parameters that make the numerator and denominator of equation (C2) dimensionless. Two of the simplest and most easily motivated potential combinations are

$$\nu = \beta/F; \tag{C12}$$

$$\nu = 1/f_{0,0}; \tag{C13}$$

(cf. the dimensionless shapelet basis functions introduced in Refregier 2003.) In Fig. C1, we plot histograms of the values of the denominators of the shear and flexion estimators defined by equations (C2), (C10) and (C11), from shapelet fits to the simulation galaxies in this study. We plot results for the two choices of ν expressed by equations (C13) and (C12). The shape of these histograms appears to show a consistent trend: using $\nu = 1/f_{0,0}$ results in distributions of denominator values that have a more clearly defined central tendency than $\nu = \beta/F$. This fact argues in favour of the adoption of $1/f_{0,0}$ to normalize to a dimensionless combination of shapelet coefficients.

To explore the distributions further, we define the following measure of skewness as a combination of the sample mean, arithmetic median and standard deviation, commonly known as Pearson's second skewness coefficient:

$$\Gamma(x) = 3\left[{\rm Mean}(x) - {\rm Median}(x)\right]/\sigma(x). \tag{C14}$$

Estimates of Γ for the distributions of Fig. C1, with uncertainties, are given in Table C1. We see that the skewness in the $\nu = \beta/F$ distributions is consistently greater than that in the distribution of denominator values when choosing $\nu = 1/f_{0,0}$. Once again, this

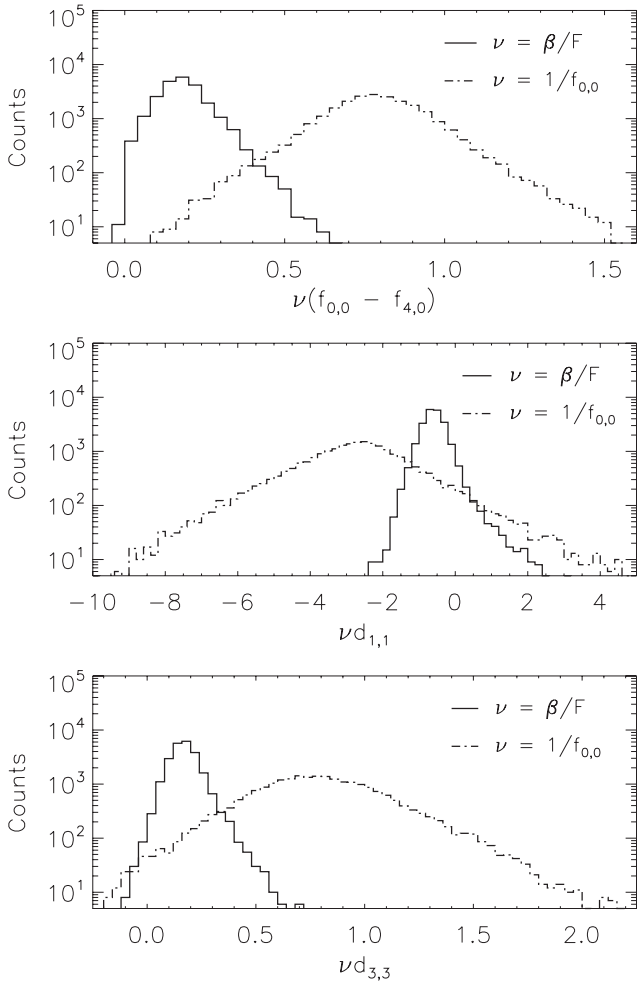

Figure C1. Histograms showing distributions of the values averaged in the denominators of equations (C2), (C10) and (C11) – upper, middle and lower panels, respectively – for two choices of normalization parameter ν. Solid line: choosing $\nu = \beta/F$. Dot–dashed line: choosing $\nu = 1/f_{0,0}$.

Table C1. Estimates of the skewness Γ as defined by equation (C14) for the distributions of Fig. C1.

	Skewness $\Gamma(x)$		
ν	$x = \nu(f_{0,0} - f_{0.4})$	$x = \nu d_{1,1}$	$x = \nu d_{3,3}$
β/F	0.264 ± 0.002	0.173 ± 0.002	0.198 ± 0.001
$1/f_{0,0}$	0.0275 ± 0.0002	− 0.0447 ± 0.0004	0.153 ± 0.001

argues in favour of the adoption of $\nu = 1/f_{0,0}$ as a measurable quantity with which to make the numerators and denominators of our shear and flexion estimators dimensionless. No other combination of observables was seen to provide better performance, either in terms of showing a clear location for the central tendency of the denominator distributions shown in Fig. C1, or in terms of reducing the required level of systematic bias calibration as shown in Fig. 6. To estimate shear and flexion from shapelet fits to our simulated images, we therefore adopt $\tilde{\gamma} = \tilde{\gamma}_\nu$, $\tilde{\mathcal{F}} = \tilde{\mathcal{F}}_\nu$ and $\tilde{\mathcal{G}} = \tilde{\mathcal{G}}_\nu$, where $\nu = 1/f_{0,0}$.

C3 Potential issues

The quantity ν, which is estimated for each object individually, is a function of random variables and therefore a random variable itself. Its inclusion in the estimators of equations (C2), (C10) and (C11) is therefore a cause of both additional uncertainty in estimators of galaxy shape and potential bias (see e.g. Melchior & Viola 2012).

These undesirable properties must be weighed against the practical advantage of the technique in providing dimensionless combinations of shapelet coefficients for shape estimation. Given the large dynamic range in both apparent galaxy sizes and fluxes in extragalactic data, and thus in the raw values of best-fitting shapelet coefficients $f_{n,m}$, this advantage is considerable. Fig. C1 shows that such combinations can provide distributions with a desirably compact support, and the same is true for the denominators of the shear and flexion estimators of equations (C2), (C10) and (C11).

To provide robust estimators of shear and flexion in the presence of such a large dynamic range in $f_{n,m}$ values would probably require the splitting of the galaxy sample into cells of apparent size and flux in a manner similar to that described by Kaiser (2000). Such a scheme carries its own biases, due to the action of shear and flexion to carry galaxies between adjacent cells. Given the degree of statistical uncertainty in both shear and flexion measurements from space-based data sets in the near and medium term, the motivation for constructing such dimensional cell-based estimators was not strong. However, such a scheme, if correctly implemented, might help reduce the calibration factors found in Section 6.3.

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