Mon. Not. R. Astron. Soc. **000**, 1–12 (2002) Printed 1 September 2015 (MN LATEX style file v2.2)

The Cluster Shear TEsting Program: a catchy subtitle here

J. Young[1⋆] and Other.
[1] *OSU/USM*

2013

ABSTRACT

We present results from the Cluster Shear TEsting Program (CSTEP), an image analysis challenge focused on weak gravitational shear in the cluster regime. Shape measurement bias is an important source of systematic error in the measurement of cluster masses by weak gravitational lensing and, consequently, cluster cosmology. In this work the accuracy of eight shape measurement pipelines is determined from image simulations that have realistic distributions of galaxy properties and large gravitational shear. The best performing methods exhibit a multiplicative bias of a few percent that is stable as a function of redshift. Our results show that the methods tested here do not have a strong quadratic bias. We find that the shear bias due to selection effects varies widely among the methods. We model the impact of the biases found on a simulated stage III optical weak lensing cluster cosmology survey, such as the Dark Energy Survey (DES), and find it to be below the statistical uncertainty in a realistic, binned cluster sample. Current shape measurement methods are therefore suitable for upcoming cluster lensing data sets, particularly when tested and calibrated by realistic image simulations.

Key words: gravitational lensing: weak, techniques:image processing

1 INTRODUCTION

The number density and distribution in mass and redshift of galaxy clusters is a versatile cosmological probe (cf. Allen, Evrard & Mantz 2011, for a recent review). There are a number of different ways of detecting clusters, including the presence of red cluster member galaxies or the X-ray and SZ signal of hot intra-cluster gas. All of these methods provide observables which, however, must be connected to cluster mass before they can be used for cosmology. One method of achieving this is weak lensing. Large optical surveys such as the Dark Energy Survey (DES, http://www.darkenergysurvey.org/) are able to detect and measure the mass of large numbers of clusters and place strong constraints on the halo mass function if sources of systematic error can be controlled. Shape measurement pipelines have been developed to accurately measure the lensing signal, yet are known to suffer from small systematic shape measurement errors (biases). As the mass estimated for a galaxy cluster depends on the strength of the shear measured on galaxies behind the cluster, a biased weak lensing pipeline will lead to a biased determination of the cluster mass.

Previous image simulation challenges were used to determine the accuracy of shape measurement pipelines, e.g. STEP1 (Heymans et al. 2006), STEP2 (Massey et al. 2007), GREAT08 (Bridle et al. 2010), GREAT10 (Kitching et al. 2012), and GREAT3 (Mandelbaum et al. 2014). These challenges used blinded image simulations to characterize the shape measurement bias in lensing pipelines. All of these challenges have focused on different aspects of shear estimation, and have led to improvements in existing shear pipelines. The results have been used to calibrate shape measurement pipelines (e.g. Applegate et al. 2012), and determine that lensing pipelines meet accuracy requirements for scientific analyses (e.g. Bergé et al. 2008). Although previous challenges provided valuable information on many important aspects of shear pipeline estimation, they were designed to test the regime of very weak cosmic shear with $|g| \leqslant 0.06$. It is important, however, to also test the accuracy of shape measurement pipelines in the higher shear regime that occurs around the center of massive galaxy clusters, where selection effects and non-linear biases may play a more important role.

The Cluster Shear TEsting Program (CSTEP) is designed to extend image simulation challenges to the high shear regime and study aspects of shear estimation that would impact a weak lensing measurement of cluster mass for a ground based large optical survey. For this, we simulate ground-based images of galaxies with a realistic distribution of signal-to-noise ratio (SNR), size and ellipticity. Lensing

⋆ E-mail: email@address (AVR); otheremail@otheraddress (ANO)

pipelines were evaluated to determine their shape measurement bias as a function of these properties. The shape measurement bias as determined for each lensing pipeline is then used to model the systematic error in the cluster mass for a stacked weak lensing measurement and its impact on cluster cosmology.

In Section 2 we describe the structure of the challenge and the image simulations used in this project. We outline the characteristics of the shape measurement pipelines used in this analysis in Section 3. We present our findings on shape measurement bias including the impact of SNR, selection effects, quadratic bias and redshift in Section 4. The effect of shape measurement biases on stacked cluster weak lensing analyses is discussed in Section 5. Conclusions from this study are given in Section 6. Finally there is an appendix which contains a more detailed description of pipelines.

2 CLUSTER STEP CHALLENGE

Gravitational lensing is the distortion of the images of distant galaxies by the tidal gravitational field of massive structures near their line of sight. This can cause a change in the observed shape, size, and brightness of a galaxy. For massive galaxy clusters this leads to galaxies at a higher redshift than the cluster often being observed as arcs (strong lensing) near the cluster center , and to appear aligned tangentialy with decreasing amount of distortion further from the cluster center (weak lensing). Since physically unrelated galaxies have random orientations, a measurement of the average ellipticity of an ensemble of galaxies unaffected by gravitational lensing would average to zero. Measuring the mean tangential alignment of background galaxies therefore provides a way of measuring the strength of the gravitational lensing signal, and thus the mass of a galaxy cluster.

For the larger shears of the cluster regime and analyses assuming radial symmetry, both the quadratic (q) and multiplicative (m) shape bias (see eq. 1) affect the measurement to a greater extent than any constant (c) shear bias, since the latter averages to zero when measuring tangential shear over a full circle. It is important to quantify the quadratic and multiplicative bias on images with simulated shear comparable to the shear observed around large galaxy clusters. The Cluster Shear TEsting Program (CSTEP) tested eight weak lensing pipelines on images of constant shear ($|\gamma| = [0.03, 0.06, 0.09, 0.15]$) with galaxy and PSF properties that simulate the properties of data that is being observed with DES. Systematic shape measurement bias was quantified by the q, m and c determined by a fit of the quadratic equation.

$$\gamma_m = q(\gamma_t)^2 + (1+m)\gamma_t + c, \quad (1)$$

where γ_t is the true shear and γ_m is the measured shear. A linear model that has been used to quantify systematic errors for cosmic shear regimes was also used, defining the biases as (Heymans et al. 2006)

$$\gamma_m = (1+m)\gamma_t + c\,. \quad (2)$$

An example fit to the im3shape data is shown in Figure 1.

The image simulations used in the CSTEP project were designed to replicate the properties of the co-added data

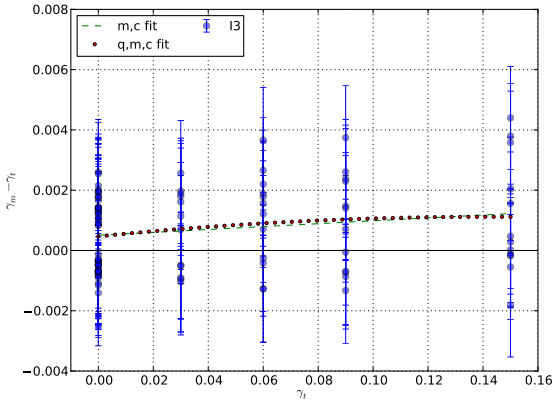

Figure 1. An example of the two types of fit used to quantify shape measurement error for the shear as measured by the im3shape pipeline (I3) on galaxy objects with Signal to Noise Ratio greater than 50. show an example with more significant difference between the two fits; use larger labels

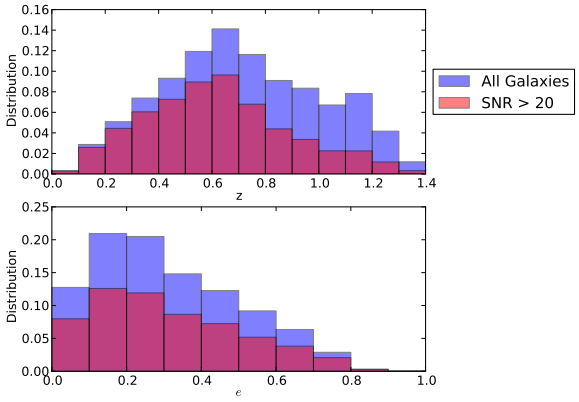

Figure 2. Distributions of various galaxy properties in the CSTEP simulation. The top and bottom panel show the distributions of redshift and intrinsic ellipticity of galaxies. The total number of galaxy objects included is 141559 and is shown in blue, the number of galaxy obejects with SNR > 20 is 84586 and is shown in magenta. use larger labels; especially the x axis labels are hard to read

of final depth that will be taken by the Dark Energy Survey. The Dark Energy survey is an optical survey of 5000 deg^2 of the southern sky, that will image around 300 million galaxies in five filters (g,r,i,z and Y) (Honscheid & DePoy 2008). To test the DES data management system and software pipelines, detailed image simulations were created (Mohr et al. 2008). Each simulated image used in CSTEP models the DES focal plane, with 62 CCDs each of which is 2K $\times$ 4K pixels (0.27"/pixel). Star and galaxy objects are rendered by drawing random samples of photons from the theoretical light profile of the source convolved by the PSF.

The CSTEP simulated images are based on an N-body simulation with a realistic mock galaxy catalog. The parent N-body simulations used by ADDGALS are the "Carmen" Las Damas (McBride et al. 2009) N-body simulation box

PSF Number	Type	Seeing
1	circular Gaussian	0.7
2	circular Gaussian	0.8
3	circular Gaussian	0.9
4	elliptical Gaussian	0.75
5	elliptical Gaussian	0.8
6	elliptical Gaussian	0.9
Shear Number	γ_1	γ_2
1	0.0	0.03
2	0.0	0.06
3	0.0	0.09
4	0.0	0.15
5	0.03	0.0
6	0.06	0.0
7	0.09	0.0
8	0.15	0.0

Table 1. Summary of the PSF and shear sets for the CSTEP simulated images.

($1h^{-3}\mathrm{Gpc}^3$) with a fiducial ΛCDM cosmology ($\Omega_M = 0.25$, $\sigma_8 = 0.8$, at $z < 1.35$). In each DES focal plane around 140,000 galaxy objects are present. Mock galaxy catalogs were created with the ADDGALS method, which reproduces observed properties of galaxies including clustering, luminosities, and colors.

The simulated galaxy images are designed to have properties similar to the galaxies that DES will observe by combining a mock catalog with empirical data from the HST/GEMS catalog for galaxies with magnitude r > 23, and from the Sloan Digital Sky Survey for galaxies with magnitude r < 23. The distributions of galaxy properties in the simulation are shown in Figure 2. The galaxy objects are created with Sersic profiles with an index that ranges from 0.5 to 5.

The CSTEP simulated images contain a constant PSF, which eliminates the technical challenge of determining the best way to model a varying PSF (cf. Kitching et al. 2013, for a simulation challenge for PSF reconstruction). For CSTEP there are two branches in the image simulation sets, those that include a circular Gaussian PSF and those that include an elliptical Gaussian PSF. The elliptical Gaussian PSF has $\epsilon = 0.03$. The simulated seeing FWHM of $0.7 - 0.9$ arcsec matches the expected seeing in final DES data for different levels of atmospheric turbulence. A summary of the PSF image properties is shown in the upper part of Table 1. For each PSF there are 8 focal plane images that contain constant shear for $\gamma = 0.0$ to $\gamma = 0.15$ in both γ_1 and γ_2, as described in the lower part of Table 1.

3 SHAPE MEASUREMENT PIPELINES

Eight weak lensing pipelines developed by members of the DES collaboration submitted results for the CSTEP simulations. They are listed in Table 2. Several of these pipelines could potentially be used to create shear catalogs for DES. The implementations represent a broad range of lensing pipeline types. They can be divided into four groups, using roughly the same criteria as in STEP2 (Massey et al. 2007). A short description of the methods is included below, while more details are provided in Appendix A.

3.1 Moment-based methods: red class

The red class methods (DE, IM, PK, KM) are based on the oldest shape measurement method KSB+ developed by Kaiser, Squires, and Broadhurst in 1995. As in STEP2, red class methods are defined as those that measure combinations of moments of each galaxy image $I(x)$ using a Gaussian weighting function. Various implementations of this class of methods have been studied in previous shape measurement challenges, and the accuracy has been shown to vary widely (Massey et al. 2007; Kitching et al. 2012). Two of the red class lensing pipeline implementations included in CSTEP have performed well on previous image simulations challenges (DE, KM). Other red class implementations that were calibrated on simulated image data have been used in recent weak lensing analyses (e.g. Gruen et al. 2013b; Applegate et al. 2012; Schrabback et al. 2010). Many of the large weak lensing cluster studies have used red class methods (Mahdavi et al. 2013; von der Linden et al. 2014), and KSB+ methods are the most common method used to measure the mass of clusters using weak lensing.

3.2 Model-fitting methods: green class

The green class methods (GM, I3) are commonly known as model fitting methods. These methods convolve various models of galaxies with different parameterizations, including their intrinsic shape, with the PSF and determine those that best fit the galaxy image. The green class methods LENSFIT and DEEPZOT were among the best performing methods in GREAT08 and GREAT10, and im3shape has been shown to perform well on the GREAT08 and GREAT10 simulations (Zuntz et al. 2013). The model-fitting method LENSFIT was the lensing pipeline for the Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS, Heymans et al. 2012).

3.3 Basis function methods: blue class

The blue class methods (MJ) are methods which model the galaxy images as a sum of orthonormal Gauss-Laguerre polynomial functions, commonly known as SHAPELETS. The MJ method competed in the STEP1, STEP2, and GREAT08 challenge.

3.4 Other methods: purple class

The purple class is used to describe a method significantly different from the three main approaches outlined above, Fourier Domain Null Testing (FDNT, Bernstein 2010). In FDNT, the surface brightness profiles of PSF and observed galaxy are transformed to Fourier space, where the latter is de-convolved by dividing by the former. Testing the Fourier representation of the galaxy for roundness after applying an inverse shear, a likelihood of shears is determined. The information used in the roundness test is limited to frequencies below those rendered infinitely noisy by the PSF. FDNT is

Method	Key	Contributor	Class
DEIMOS	DE	P. Melchior	Red
IMCAT	IM	J. Young	Red
ksbm	KM	P. Melchior	Red
PKSB	PK	D. Gruen	Red
Gaussian Mixtures	GM	E. Sheldon	Green
im3shape	I3	B. Rowe	Green
Bernstein and Jarvis (2002)	MJ	M. Jarvis	Blue
PFDNT	PF	D. Gruen	Purple

Table 2. A summary of the lensing pipelines used to analyze CSTEP simulated images.

known to perform very well on low noise simulations (Bernstein 2010) but has not been applied yet to observed data for a scientific analysis.

4 EVALUATION OF PIPELINES

In this section we evaluate the performance of shape measurement pipelines based on a set of criteria important for large cluster weak lensing studies. The success of a shape measurement pipeline depends on its ability to accurately measure shear on small faint galaxies, be unbiased when all galaxies present in an image are used to measure the shear, and have a shear bias that does not change as a function of the redshift of the galaxies included in the lensing measurement. A good shape measurement method would also be able to successfully measure shear on a high percentage of the galaxy images which it attempts to analyze. The relative importance of the above criteria in determining the best shear measurement pipeline to use may depend both on the data being evaluated and the lensing application. A more detailed examination of the performance of each individual lensing pipeline as a function of PSF ellipticity, PSF size, galaxy size, and the contribution of selection bias is included in Appendix A, along with a description of the algorithms of the shape measurement pipelines.

The measure of signal-to-noise ratio (SNR) referred to in this section was measured on the CSTEP images for each galaxy using SExtractor. It is defined as

$$\mathrm{SNR} = \frac{\mathrm{Flux}}{\mathrm{Flux\ Error}}\ , \qquad (3)$$

where the flux and flux error was determined using ... say FLUX_AUTO or FLUX_ISO here and give details on detection threshold in case it's the latter; I don't think the figure adds much since the MJ and I3 definitions of SNR are less common than this one, so I've removed it.

can you comment here or elsewhere on how you get the error bars on q,m,c? you should probably also mention how you fit for q,m,c (subtracting the mean intrinsic ellipticity of all galaxies as far as I remember, rather than doing a ring-test type thing)

4.1 Mean shape measurement bias

In this section we show the average shape measurement bias results on a representative DES lensing catalog of all sources with SNR > 20. Most previous weak lensing studies rejected all objects below some SNR threshold, but the depth chosen varies as it is determined by both the capabilities of the lensing pipeline used and the type of lensing analysis. In the CFHTLenS survey, all objects were rejected that were measured as i'(AB) < 24.7, or a SNR limit of around 10 (Miller et al. 2013). This limit was imposed for the CFHTLenS catalog as it was the depth determined at which the photometric redshifts and shape measurements were too poorly measured (Hildebrandt et al. 2012). For the CSTEP project we chose objects with SNR > 20 for the representative DES lensing catalog. This is a conservative choice, and with additional lensing pipeline development a lower SNR threshold could be chosen for the final DES lensing catalog. it would be great to compare this to the SNR cuts in Jarvis+2015, although these were of course made on a different SNR definition

The average shape measurement bias on all images in the CSTEP simulations for galaxies SNR > 20 is shown in Figure 3. Three of the lensing pipelines (I3, DE, KM) also competed in the GREAT10 challenge. The results are roughly consistent between the two challenges for I3 and KM, showing that both pipelines are robust, and demonstrate similar levels of shape measurement bias for a wide variety of sources and PSF types. Both I3 and KM are among the best performing lensing pipelines for both CSTEP and GREAT10. this leaves the reader to wonder what is the case with DE

4.2 Quadratic shear bias

In this section we discuss the importance of quadratic shape measurement bias on a representative DES lensing catalog of all sources with SNR > 20.

added this interpretation: As shown in the top panel of Fig. 3, quadratic biases of the pipelines tested in this work are at the level of or below $|q| \approx 0.1$. Even at the larges shear tested ($\gamma_t = 0.15$), this corresponds to a relatively minor bias compared to the typical level of multiplicative biases, i.e. $|q|\gamma_t^2 \ll |m|\gamma_t$. Quadratic biases are therefore a sub-dominant contribution to systematic errors in shape measurement for the pipelines tested. An exception may be the case of GM, which shows a small multiplicative but relatively large quadratic bias for the three-parameter fit.

4.3 Dependence on galaxy signal-to-noise ratio

The level of pixel noise present in galaxy images greatly impacts the ability of lensing pipelines to accurately measure shear (e.g. Melchior & Viola 2012; Okura & Futamase 2013; Refregier et al. 2012). The multiplicative bias due to pixel noise varies depending on both the source galaxy population and the lensing pipeline. There have been a number of schemes that attempt to correct for the effect of noise bias for specific lensing pipelines including Kacprzak et al. (2012), Miller et al. (2013), and Bernstein

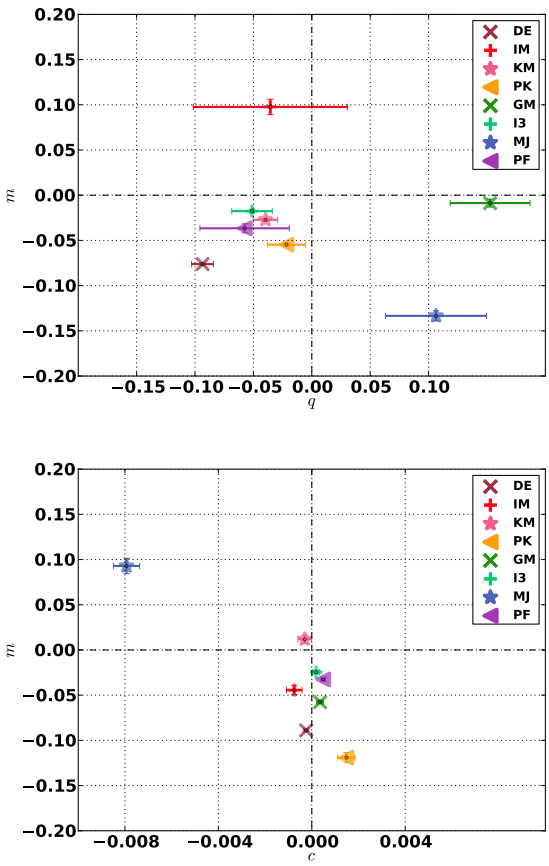

Figure 3. Average shape measurement bias as measured on all images for galaxies with SNR > 20. The top panel shows the shape measurement bias q,m as measured using a q,m,c fit. The bottom panel shows the shape measurement bias m,c as measured using a m,c fit.

& Armstrong (2014).

To make the results easier to interpret and compare to previous studies, the bias of the lensing pipelines was quantified in this section by a linear function Eqn. 2. The results of the multiplicative bias as a function of SNR are shown in Figure 4.

added this sentence of interpretation, which is still a bit on the stub side The dependence of multiplicative biases on SNR are significant for most of the pipelines tested, but strongest in the regime of low SNR (< 20). Careful calibration is therefore necessary if the shapes of these very faint galaxies are to be used in a weak lensing analysis.

4.4 Selection effects and pipeline efficiency

Lensing pipelines are not able to return a shear measurement for all galaxies in an image. If a lensing pipelines preferentially rejects objects based on their ellipticity and their orientation relative of the ellipticity of the PSF, then the assumption that the population of galaxies used to measure shear have $< \epsilon >= 0$ is no longer true. This selection bias effect has been discussed in STEP1 where it was determined that results for some lensing pipelines improved when the

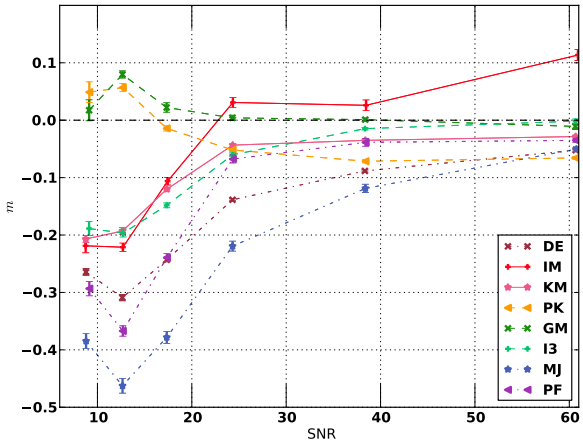

Figure 4. Multiplicative bias (m) as a function of Signal-to-Noise Ratio (SNR). The multiplicative bias was determined using a linear fit of measured to true mean shear.

non-zero mean intrinsic ellipticity of the surviving objects was corrected for.

The fraction of successful shear measurements for each pipeline is shown for sources with SNR > 20 in Table 3. The pipelines with the highest efficiency of successful measurements are PK, DE and KM, which are all moment based methods. As implemented in this challege the GM method rejects a portion of the returned shear measurements and uses a weighting system in determining the average shear. Before the flagged objects were rejected, the efficiency of GM was 0.93 before and 0.24 after rejection. These values are therefore not directly comparable to the other lensing pipelines.

Pipelines may reject objects in a ellipticity dependent way. To test for this selection bias, the γ measured is corrected for the intrinsic ellipticity of the sources as measured by each lensing pipeline. I3 and DE do not appear to have a selection bias as their accuracy is similar, before and after correction for $< \epsilon > \neq 0$. The pipeline that is most affected by selection bias is PF. Q, M and C as measured for each lensing pipeline after correcting for selection effects is included in Table **??** and Table **??**. references to tables are orphaned here

4.5 Shape measurement bias as a function of redshift

To ensure that the cosmology extracted using the halo mass function as measured on galaxy clusters using weak lensing is accurate, it is important that galaxy clusters at various redshifts have similarly low levels of shape measurement bias. The population of galaxies observed at different redshifts will vary in SNR, ellipticity distribution, size and morphology, all of which can affect the accuracy of a lensing measurement. The bias of galaxy sources with SNR > 20 as a function of redshift is shown in Figure 6. The results of bias of the lensing pipelines was quantified by fitting for M and C .

We find that MJ, DE, PK, I3 and KM show a consis-

Pipeline	efficiency
DE	0.92
PF	0.86
GM	0.24 (0.93)
MJ	0.81
PK	0.90
I3	0.79
IM	0.55
KM	0.91

Table 3. Percentage of galaxies at SNR>20 for which the shape measurement pipelines return a successful measurement.

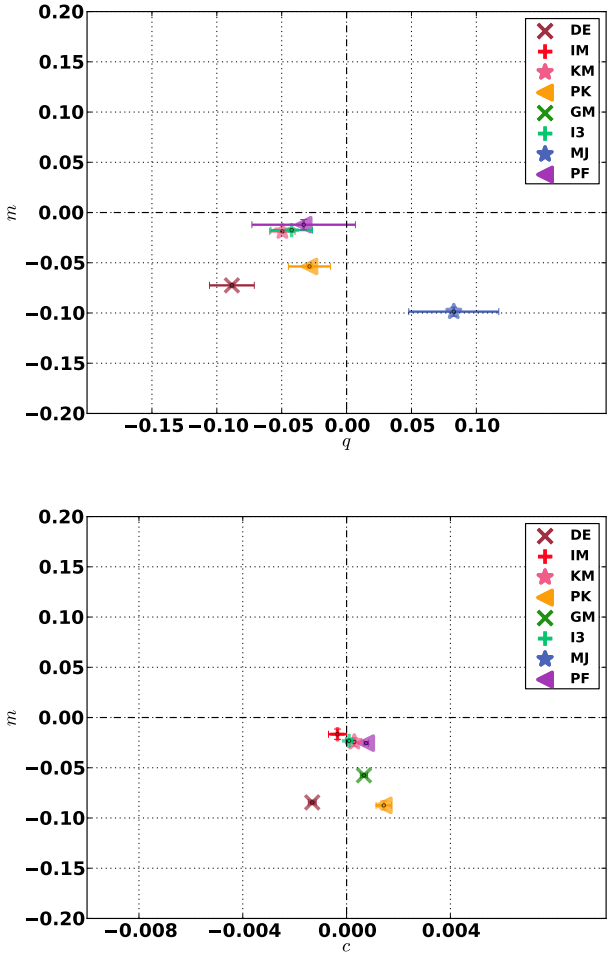

Figure 5. Average Q and M results for all pipelines for objects SNR > 20 after correcting for selection effects in the top panel and average M and C results for all pipelines for objects SNR > 20 in the bottom panel.

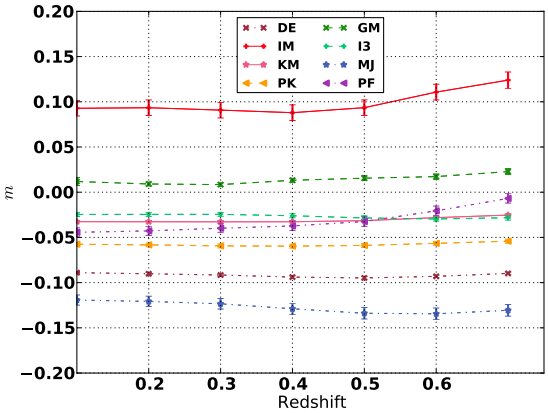

Figure 6. Multiplicative shape measurement bias (M) of pipelines as a function of redshift.

tent level of shape measurement bias as a function of redshift. Our results are consistent with an analysis that studied the shape measurement bias as a function of redshift for I3 (Kacprzak et al. 2013) on simulated COSMOS galaxies and determined that it did not not vary significantly as a function of redshift.

5 STACKED CLUSTER WEAK LENSING

An accurate measurement of the abundance of galaxy clusters within a given survey volume provides powerful constraints on cosmological parameters. While weak lensing can provide individual mass measurements of high mass clusters, lower mass clusters can only be measured on average by stacked weak lensing. Stacked weak lensing measures the mean tangential shear of background galaxies behind galaxy clusters which are binned by an observable such as richness. Stacked weak lensing of clusters in the MaxBCG catalog (Koester et al. 2007; Sheldon et al. 2009) on data from the Sloan Digital Sky Survey (York et al. 2000) has been used to derive cosmological constraints on Ω_m and σ_8 in (Zu et al. 2012; Rozo et al. 2010). Since ongoing and upcoming surveys such as DES will provide deeper imaging and better seeing over a wide area, the constraints on cosmology from stacked cluster weak lensing will substantially improve if sources of systematic error can be controlled.

There are a number of systematic effects that can bias the stacked weak lensing mass measurement, including issues with background sample selection and redshift estimation like a contamination by cluster members, mis-centering of clusters, deviations of the fitted shear profile from the truth (e.g. Becker & Kravtsov 2011), orientation bias (Dietrich et al. 2014), or a lack of full treatment of the intrinsic variation in cluster profiles (Gruen et al. 2015b). In this paper we focus on the systematic bias in the stacked weak lensing cluster mass measurement contributed by shape measurement. If lensing pipelines provide a biased measurement of the shear profile of galaxy clusters, this will bias the observed average cluster mass, and bias the derived cluster abundance function. To model the possible impact on a DES-like survey, we take the average shear profile for each stacked cluster bin,

apply a lensing bias, and then determine what cluster mass would be measured. By comparing the true and the measured cluster mass this provides an estimate of the effect of shape measurement bias for each lensing pipeline. In this paper we compare the expected statistical errors expected in a DES-like survey stacked weak lensing cluster mass measurement described in section 5.1 to the modeled bias on the mass from shape measurement errors described in section 5.2.

5.1 Stacked cluster weak lensing: statistical errors

Here we compare projected statistical errors of a DES-like survey to systematic errors introduced by shape measurement bias on a simulated stacked cluster weak lensing analysis. The distribution of cluster masses that we use is expected to be comparable to the distribution of clusters that will be observed by DES. The statistical error we model for the stacked cluster mass is dominated by shape noise in the background galaxies.

To create a simulated DES-like stacked weak lensing survey we take a halo distribution drawn according to the halo mass function how do you do this? which mass function / cosmology?. For an analysis on observed data the clusters would be binned by observables that are correlated with cluster mass such as richness, optical luminosity, and X-ray luminosity. Due to scatter in the observables, these bins are not strict mass bins, which however should not greatly affect the average shear profile. In this study we divide the cluster sample into four mass bins and six redshift bins.

The average mass, concentration, and redshift of the clusters in each bin is determined and then used to calculate the expected statistical error and an average NFW shear profile. The number of clusters in each mass bin and their average redshift is shown in Figure 7.

To estimate the average redshift of background galaxies behind each bin we use a redshift distribution of galaxies expected for a DES-like survey,

$$f(z) = z^m exp(-(z/z_*)^\beta) \, , \tag{4}$$

where $m = 2.0$, $z_* = 0.5$ and $\beta = 2.0$ as described in (Weinberg et al. 2012). The average redshift of sources selected behind each stacked weak lensing bin is shown in in Table **??** and Table B1.

The statistical mass uncertainty $\Delta \ln(M)$ for stacked weak lensing is (Weinberg et al. 2012)

$$\Delta \ln(\mathrm{M}) = \sqrt{(\Delta \ln(\mathrm{M_s}))^2 + (\sigma_{\mathrm{wl}})^2} \, , \tag{5}$$

which combines the undertainty due to shape noise ($\Delta \ln(\mathrm{M_s})$) and the scatter based on intrinsic variation in cluster shear profiles at fixed mass (σ_{wl}). From Becker & Kravtsov (2011) we take

$$\sigma_{wl} = \frac{0.3}{\sqrt{N}} \tag{6}$$

where N is the number of clusters in a given bin. To calculate the shape noise we use the equation

$$\Delta \ln(\mathrm{M_s}) = 6.0{\times}10^3 \left(\frac{\mathrm{N}}{\mathrm{N_A}}\right)^{0.5} \left(\frac{\mathrm{M}}{\mathrm{M_o}}\right)^{-0.66} \left(\frac{\overline{\mathrm{ngal}}}{\mathrm{N_o}}\right)^{-0.5} \left(\frac{\mathrm{D_A}}{0.5}\right)^{-1} \tag{7}$$

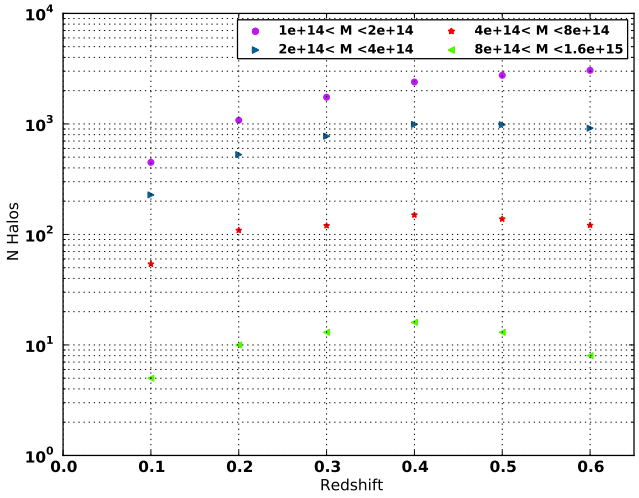

Figure 7. The number of halos in each given mass bin.

from Weinberg et al. (2012) you need to explain the symbols in this eqn. To model the concentration we expect for clusters at this redshift we assign an initial concentration c

$$c = A \left(\frac{m_{200}}{2.0 \cdot 10^{12}}\right)^B (1 + z_{\mathrm{cluster}})^C \tag{8}$$

Where $A = 7.85$, $B = -0.081$, $C = -0.71$ and m_{200} is the mass within r_{200} from (Oguri & Takada 2011). 200m or 200c? if this m-c-relation from oguri or elsewhere? the way you're citing them makes it not absolutely clear

5.2 Stacked cluster weak lensing: modeled systematic errors

Here we model the systematic errors on the stacked cluster weak lensing measurement by comparing the average mass in each bin to the mass measured by a shear profile, affected by shape measurement bias. To measure the effect of shape measurement bias, a Navarro, Frenk, & White (1997) (NFW) density profile is created for each mass bin. This density profile is used to calculate the tangential reduced shear g as described in Wright & Brainerd (2000).

The reduced shear as measured including shape bias is modeled as

$$g' = (1 + m)g + qg^2 \, . \tag{9}$$

We then fit this g' distribution to get a M_{200} and c value. need more details here ... I guess sources uniformly distributed in area over some range in radius, maximum likelihood?

interpretation of results; note that even if the systematic offset is within the statistical uncertainties in each bin, all bins combinedly being measured low is a different story... at least need to discuss that

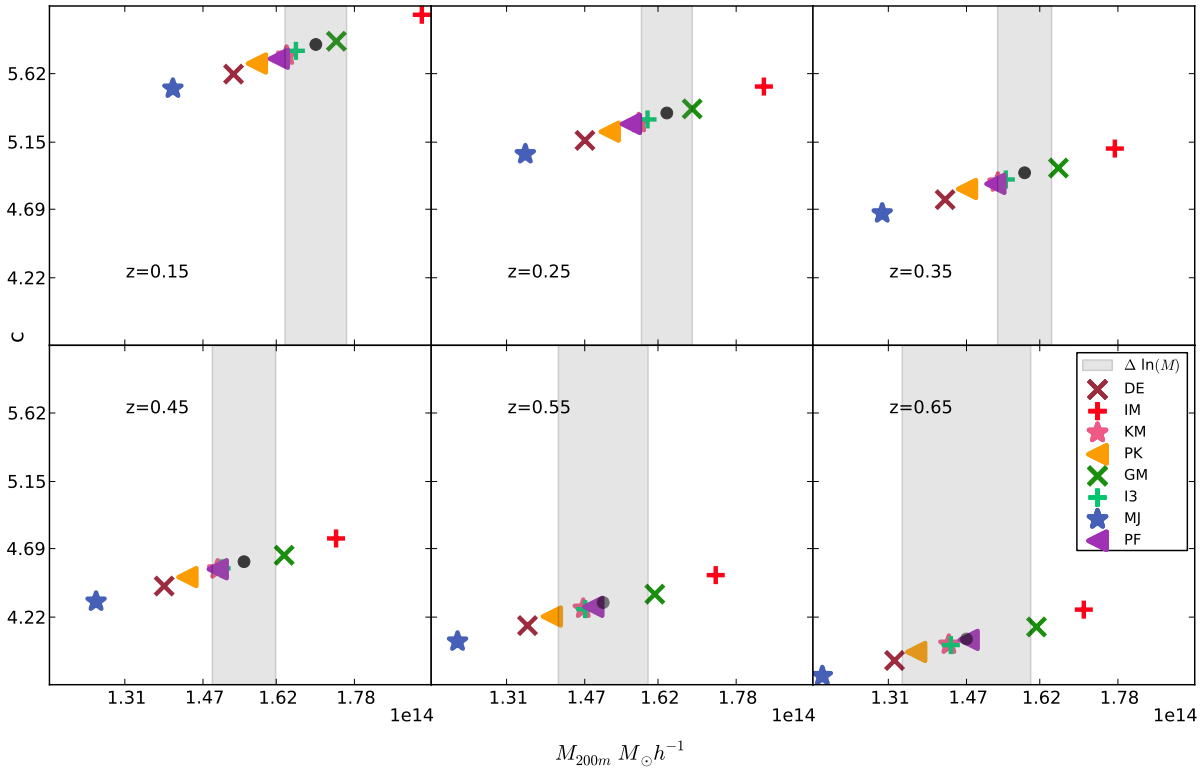

Figure 8. NFW mass and concentration measured on a reduced shear profile affected by shape measurement bias as measured on all images for galaxy objects with SNR > 20.

N	Z_{lens}	Z_{source}	C	$M_{200m}\ M_\odot h^{-1}$	$\Delta\ \ln(M)$
450	0.16	0.58	5.82	1.70	0.037
1080	0.26	0.62	5.35	1.64	0.032
1744	0.35	0.66	4.94	1.59	0.035
2396	0.45	0.73	4.60	1.55	0.042
2753	0.55	0.79	4.32	1.51	0.061
3054	0.65	0.87	4.07	1.47	0.089

Table 4. The expected statistical error due to one method of cluster stacking for a DES-like cluster distribution).

6 SUMMARY

The CSTEP project tested shape measurement pipelines in the cluster shear regime on realistic image simulations, comparable to the data collected by DES. We used these to analyzed the effectiveness and systematic errors of shear pipelines by several criteria, and tested the impact shape measurement bias has on a stacked cluster weak lensing analysis.

The results of these analyses for the eight shape measurement pipelines tested are as follows:

- Multiplicative biases at the level of a few per-cent are common among the pipelines tested. These translate to relative systematic errors in cluster mass of the same order. For the most accurate pipelines participating in CSTEP, the statistical uncertainties of stacked weak lensing cluster masses in a realistic binning are larger than these systematic effects for a DES-like survey. The combined effect on the calibration of cluster mass-observable relation may, however, still be significant. We note that calibration of shape measurement biases from dedicated simulations (cf., e.g., Jarvis et al. 2015, their Section 7.3.2) can reduce such biases considerably. Uncertainty in the multiplicative bias calibration should, however, still be included in analyses that use weak lensing constraints on cluster masses.

- We find that selection effects, i.e. the dependence of the success rate of shape measurement on the true shape

of galaxies, play a significant role for multiplicative biases. Depending on the methodology of calibration (e.g. with ring tests), this effect has not always been accounted for in the past. The selection of galaxies should be made as independent of shear as possible (e.g. with *roundified* measures of signal-to-noise ratio or size, cf. Jarvis et al. 2015).

- On a similar note, we find a sometimes strong dependence of shear bias on signal-to-noise ratio. This suggests that a bias correction that takes signal-to-noise ratio into account, rather than a global calibration factor, is a promising calibration scheme. got that from your previous summary, but can't really see it from what is in the paper: The effect of other galaxy properties, such as morphology, is less drastic.
- Quadratic biases, despite the larger distortions seen in the cluster regime, are subdominant in their effect on recovered shears and cluster masses, for most of the pipelines tested. We also do not find a strong dependence of mean biases on source redshift, which is particularly reassuring for the calibration of mass-observable relations.

This project has served as an early test of prospective shape measurement pipelines for applications in DES. The I3 and GM methods have, in updated form, been applied to DES Science Verification data (Jarvis et al. 2015) and have been shown to meet the requirements of several scientific weak lensing applications of this first DES data set (cf. Vikram et al. 2015; Chang et al. 2015; Becker et al. 2015; Gruen et al. 2015a). Despite recent advances in shape measurement techniques, larger future data sets will continue to require careful calibration and simulation-based testing of shear biases to make use of their full statistical potential.

ACKNOWLEDGMENTS

I thank.....

REFERENCES

Allen S. W., Evrard A. E., Mantz A. B., 2011, **??**jnlARA&A, 49, 409
Applegate D. E. et al., 2012, ArXiv e-prints
Becker M. R., Kravtsov A. V., 2011, ApJ, 740, 25
Becker M. R. et al., 2015, ArXiv e-prints
Bergé J. et al., 2008, MNRAS, 385, 695
Bernstein G. M., 2010, MNRAS, 406, 2793
Bernstein G. M., Armstrong R., 2014, MNRAS, 438, 1880
Bertin E., 2011, in Astronomical Society of the Pacific Conference Series, Vol. 442, Astronomical Data Analysis Software and Systems XX, Evans I. N., Accomazzi A., Mink D. J., Rots A. H., eds., p. 435
Bridle S. et al., 2010, MNRAS, 405, 2044
Chang C. et al., 2015, Physical Review Letters, 115, 051301
Dietrich J. P. et al., 2014, ArXiv e-prints
Gruen D. et al., 2013a, MNRAS, 432, 1455
Gruen D. et al., 2015a, ArXiv e-prints
Gruen D., Seitz S., Becker M. R., Friedrich O., Mana A., 2015b, MNRAS, 449, 4264
Gruen D. et al., 2013b, ArXiv e-prints
Heymans C. et al., 2006, MNRAS, 368, 1323
Heymans C. et al., 2012, MNRAS, 427, 146
Hildebrandt H. et al., 2012, MNRAS, 421, 2355
Honscheid K., DePoy D. L., 2008, ArXiv e-prints
Jarvis M. et al., 2015, ArXiv e-prints
Kacprzak T., Bridle S., Rowe B., Voigt L., Zuntz J., Hirsch M., MacCrann N., 2013, ArXiv e-prints
Kacprzak T., Zuntz J., Rowe B., Bridle S., Refregier A., Amara A., Voigt L., Hirsch M., 2012, MNRAS, 427, 2711
Kitching T. D. et al., 2012, MNRAS, 423, 3163
Kitching T. D. et al., 2013, 205, 12
Koester B. P. et al., 2007, ApJ, 660, 239
Mahdavi A., Hoekstra H., Babul A., Bildfell C., Jeltema T., Henry J. P., 2013, ApJ, 767, 116
Mandelbaum R. et al., 2014, ArXiv e-prints
Massey R. et al., 2007, MNRAS, 376, 13
McBride C., Berlind A., Scoccimarro R., Wechsler R., Busha M., Gardner J., van den Bosch F., 2009, in Bulletin of the American Astronomical Society, Vol. 41, American Astronomical Society Meeting Abstracts #213, p. 425.06
Melchior P., Viola M., 2012, MNRAS, 424, 2757
Miller L. et al., 2013, MNRAS, 429, 2858
Mohr J. J. et al., 2008, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7016, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Oguri M., Takada M., 2011, Phys. Rev. D, 83, 023008
Okura Y., Futamase T., 2013, ApJ, 771, 37
Refregier A., Kacprzak T., Amara A., Bridle S., Rowe B., 2012, MNRAS, 425, 1951
Rozo E. et al., 2010, ApJ, 708, 645
Schrabback T. et al., 2010, A&A, 516, A63
Sheldon E. S. et al., 2009, ApJ, 703, 2217
Vikram V. et al., 2015, Phys. Rev. D, 92, 022006
von der Linden A. et al., 2014, MNRAS, 443, 1973
Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2012, ArXiv e-prints
Wright C. O., Brainerd T. G., 2000, ApJ, 534, 34
York D. G. et al., 2000, AJ, 120, 1579
Zu Y., Weinberg D. H., Rozo E., Sheldon E. S., Tinker J. L., Becker M. R., 2012, ArXiv e-prints
Zuntz J., Kacprzak T., Voigt L., Hirsch M., Rowe B., Bridle S., 2013, MNRAS, 434, 1604

APPENDIX A: SHEAR MEASUREMENT PIPELINES

A1 DEIMOS

DEIMOS determines the shape of objects using the second order moments of the light profile as measured by and elliptical Gaussian weight function. In this challenge the DEIMOS pipeline eliminates objects with $\gamma > 1.0$ from the catalog.

A2 IMCAT

The KSB+ pipeline IMCAT measures the second order moments using an elliptical gaussian. The implementation of IMCAT used on the CSTEP images took an average of the P^{sm} and p^* of selected stars in the image to calculate the PSF. The average of the trace of the two matrices P^{sm} and P^{sh} were multiplied by the identity matrix. The IMCAT pipeline eliminates objects with $\gamma > 1.0$ from the catalog.

A3 ksbm

The ksbm pipeline version used here is the same as the version used to analyze the image sinulations in the GREAT10 challenge. This KSB+ implementation uses elements of the DEMIOS lensing pipeline to determine the galaxy centroid and the optimal size of the weighting function. There are no correction factors applied, but objects with $\gamma > 1.0$ are eliminated from the final catalog.

A4 PKSB

PKSB is a lensing pipeline that uses the PSFEx Bertin (2011) to create a model of the PSF and an implementation of KSB+ method for shape measurement. A modified version of this lensing pipeline that incorperates correction of shape measurement bias, based on the CSTEP simulations, has been used to measure the mass of galaxy clusters in Gruen et al. (2013b) and Gruen et al. (2013a).

A5 Gaussian Mixtures

A6 im3shape

im3shape is a modular shape measurement code that performs a maximum likelihood (ML) fit of a de Vaucouleurs bulge plus Exponential disc galaxy model to noisy images, incorporating an applied PSF (see Zuntz et al 2013/in prep.). The software is written primarily in C with supporting Python infrastructure, and is publicly available. For the ClusterSTEP data the PSF is modelled as an elliptical Moffat profile (1969), assumed to be constant across each chip. Moffat profiles were fit (using the Levenberg-Marquardt algorithm) to stellar images in the ClusterSTEP data, and the chipwise model estimated from these individual fits using an inverse variance weighted average of the resulting best-fitting parameters. For the subset of fields in which the PSF was known to be Gaussian, a β slope parameter of 1000 was fixed in the profile fitting process (the Moffat approaches the Gaussian profile for large β).

For galaxy shape measurement using parametric profiles, ML estimators are known to be biased due to the presence of noise (e.g., Refregier et al 2012; Kacprzak et al 2012). For the tests in this paper a suite of noise bias-calibrating simulations were not conducted, due to resource constraints and the challenge of producing a representative calibration suite for data with realistic distributions of size and signal-to-noise such as ClusterSTEP: the noise calibration schemes presented by Kacprzak et al (2012) and Zuntz et al (2013) were for far simpler distributions of galaxy properties. Understanding how to build such suites is an active field of research, and of great relevance to the many methods with known noise bias issues.

The performance of im3shape shear estimates is therefore expected to degrade somewhat as signal-to-noise decreases. Objects with an im3shape-determined signal-to-noise (in total flux) of lower than 10 were excluded in the final catalogue, along with catastrophic outliers in the value of the best likelihood, and objects for which any pixel value in a model-minus-data residual image was found to be greater than the peak pixel flux in the data.

A7 BJ02

A8 PFDNT

For PFDNT we prepare time image postage stamp and PSF model in the same way. Shear is estimated by applying roundness tests on the anti-sheared, deconvolved Fourier transform of the image. To this end, we use a weight function limited to frequencies where the Fourier transform of the PSF is above zero according to Bernstein (2010). We use the ellipticity estimate from the PKSB pipeline as a starting point and sample ellipticities on a hexagonal grid to find the shear estimate as the probability-weighted integral over ellipticity space.

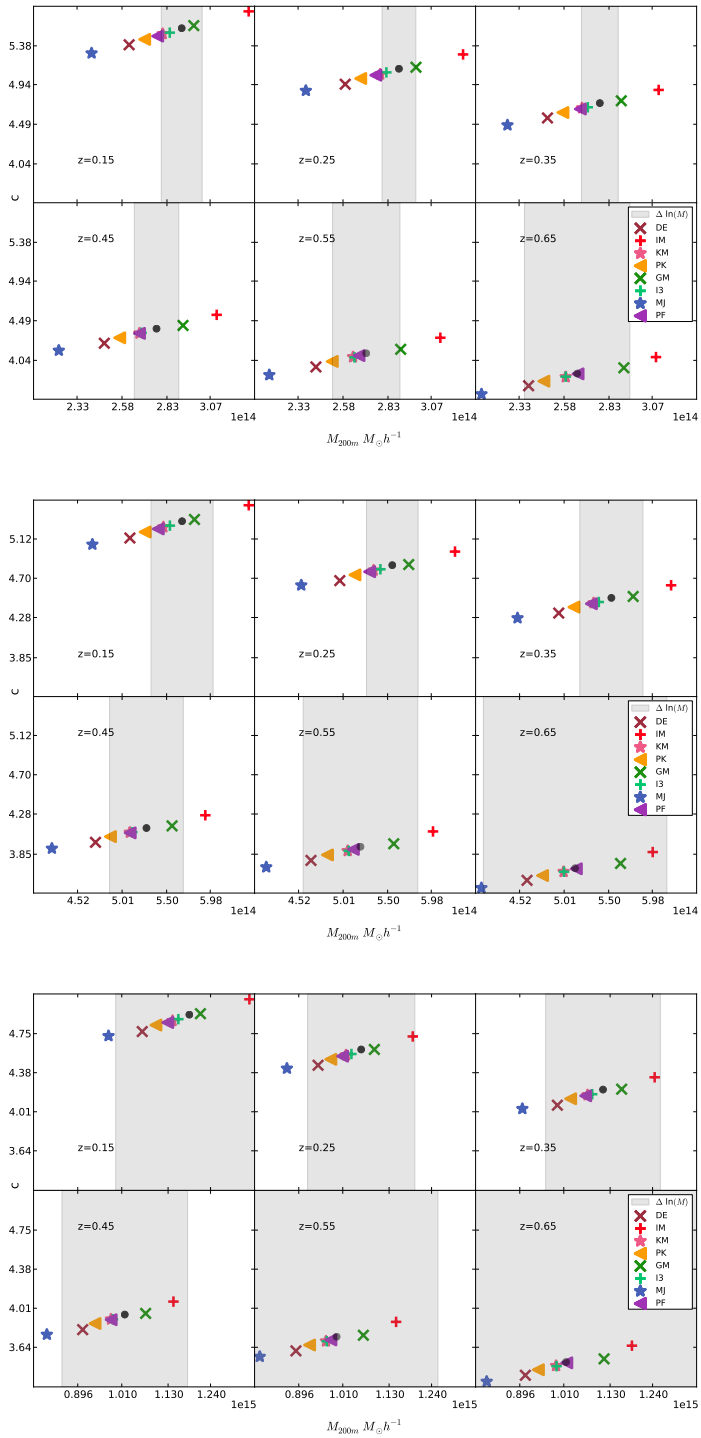

Figure B1. NFW mass and concentration measured on a reduced shear profile effected by shape measurement bias as measured on all images for galaxy objects with SNR > 20.

APPENDIX B: STACKED WEAK LENSING ERROR

This section of the appendix contains information about the expected statistical errors for a simulated cluster distribution of the stacked weak lensing signal of DES.

N	Z_{lens}	Z_{source}	C	M_{200m} $M_{\odot}h^{-1}$	Δ ln(M)
228	0.16	0.58	5.58	2.91	0.039
528	0.25	0.62	5.12	2.89	0.032
776	0.35	0.66	4.73	2.78	0.037
990	0.45	0.73	4.40	2.77	0.045
986	0.55	0.79	4.12	2.71	0.069
916	0.65	0.87	3.89	2.65	0.110
54	0.16	0.58	5.31	5.67	0.060
109	0.26	0.62	4.84	5.55	0.051
120	0.35	0.66	4.49	5.53	0.063
150	0.45	0.73	4.13	5.28	0.077
138	0.55	0.79	3.93	5.20	0.121
121	0.65	0.87	3.70	5.13	0.197
5	0.17	0.58	4.93	11.85	0.161
10	0.26	0.62	4.60	10.58	0.131
14	0.36	0.66	4.22	11.12	0.134
16	0.46	0.73	3.95	10.18	0.160
13	0.54	0.79	3.74	9.94	0.264
8	0.65	0.87	3.50	10.16	0.492

Table B1. The expected statistical error due to one method of cluster stacking for a DES like cluster distribution).