The impact of cosmic variance on simulating weak lensing surveys

Arun Kannawadi[1][*], Rachel Mandelbaum[1], Claire Lackner[2],

[1] *McWilliams Center for Cosmology, Carnegie Mellon University, Pittsburgh, PA 15217, USA*
[2] *Kavli Institute for the Physics and Mathematics of the Universe (WPI), Todai Institutes for Advanced Study,*
the University of Tokyo, Kashiwa, Japan.

7 October 2014

ABSTRACT
According to our current understanding, galaxy shapes and morphologies should depend on various factors such as the local environment. Realistic image simulations for calibration of weak lensing analysis methods that use training samples from the Hubble Space Telescope can therefore be affected by these trends, due to the limited volume of the universe that has been surveyed by Hubble. We will show how redshift slices in a volume-limited subsample of COSMOS can be classified as overdense or underdense (or neither), and how the statistical properties of various morphological parameters such as ellipticity, Sérsic n , Bulge-to-Total ratio and color differ in these bins. This study requires a careful distinction between environment effects from large-scale structure, which we do not wish to include in simulations, and general trends in the galaxy population with redshift. We conclude with some guidance for how upcoming surveys can use COSMOS data as the basis for weak lensing simulations without having their conclusions overly affected by cosmic variance. I have some comments on this abtract, but prefer to do all revision of the abstract at the end once the paper is finalized.

Key words: Gravitational lensing: weak — Cosmology: Large-scale structure of Universe — Galaxies: evolution.

1 INTRODUCTION

Weak gravitational lensing, the deflection of light by mass, is one of the cleanest ways to study the nature of dark energy by tracking the growth of structure in the Universe as a function of time (e.g., Bartelmann & Schneider 2001; Albrecht et al. 2006; Weinberg et al. 2013). As light from background sources passes by matter (including dark matter) on its way to us, the apparent shapes of the background galaxies get distorted, and the galaxies get slightly magnified as well. Because of its sensitivity to dark matter and dark energy, major surveys such as the Hyper Suprime-Cam (HSC; Miyazaki et al. 2006), Dark Energy Survey (DES; The Dark Energy Survey Collaboration 2005), the KIlo-Degree Survey (KIDS; de Jong et al. 2013), the Panoramic Survey Telescope and Rapid Response System (PanSTARRS; Kaiser et al. 2010), the Large Synoptic Survey Telescope (LSST; LSST Science Collaboration et al. 2009), Euclid[1] (Laureijs et al. 2011), and Wide-Field Infrared Survey Telescope (WFIRST; Green et al. 2012) are

planned for the next two decades to gather enormous quantities of weak lensing data that will lead to precise constraints on the growth of structure with time, and therefore cosmological parameters.

For the upcoming surveys to achieve their promise, their systematic error budgets must be below their statistical error budgets. Systematic error budgets for weak lensing surveys typically include astrophysical effects, such as intrinsic alignments of galaxy shapes with large scale density fields (e.g., Troxel & Ishak 2014) and the effect of baryons on the matter power spectrum (e.g., van Daalen et al. 2011), as well as observational uncertainties such as the ability to robustly infer shears from galaxy observed shapes or photometric redshifts from their observed colors. Given the expected sub-per cent errors on upcoming surveys, systematic errors must be reduced from their typical level in the current state-of-the-art measurements that typically achieve ~ 5 per cent statistical errors at best (e.g., Schrabback et al. 2010; Heymans et al. 2013; Jee et al. 2013; Mandelbaum et al. 2013).

One method that is commonly used to test for the presence of systematic errors in the shear estimation process is image simulation, where we can cleanly test whether our methods of shear estimation recover the ground truth.

---

[*] akannawa@andrew.cmu.edu
[1] http://sci.esa.int/euclid/, http://www.euclid-ec.org

This is a valuable test, considering the numerous sources of additive and multiplicative bias such as a mismatch between galaxy model assumptions and actual galaxy light profiles (e.g., Voigt & Bridle 2010; Melchior et al. 2010), biases due to the effects of pixel noise on the shear estimates (Kacprzak et al. 2012; Melchior & Viola 2012; Refregier et al. 2012), and ellipticity gradients (Bernstein 2010). These biases often differ for galaxies with different morphologies (e.g., disks vs. ellipticals), sizes, S/N, and shape (Bridle et al. 2010; Kitching et al. 2012). A general requirement for simulations used to test shear recovery is that they should be as realistic as possible.

Realistic simulations may use samples based on images from the Hubble Space Telescope (*HST*). Software packages like GALSIM[2] (Rowe et al. 2014) can generate images of galaxies from the *HST* as they would appear with an additional lensing shear and viewed by some lower resolution telescope. Examples of training samples from the *HST* include the COSMOS survey (used by the GREAT3 challenge, Mandelbaum et al. 2014) or the Ultra Deep Field (UDF, used by Jee et al. 2013). These two examples serve as the extremes in the *HST* samples used as the basis for image simulation, with COSMOS being shallower but representing the current widest contiguous area surveyed by the *HST*, and the latter being extremely deep but narrow.

For a variety of physical reasons, some of which are still not fully understood, the shape and morphology of galaxies depends on their local environment (e.g., Carollo et al. 2014; De Propris et al. 2014). Hence, local overdensities or underdensities observed in these *HST* fields may (given the small size of the field) cause the properties of the galaxy population in redshift slices to be atypical depending on the environment in that slice. This would have the undesired consequence of including variation in galaxy properties due to the COSMOS (or other) survey cosmic variance in the simulated galaxy sample in that redshift slice, rather than only including ensemble effects that would appear in a large cosmological volume, such as true redshift evolution of galaxy properties. Our goal is to quantify the degree to which the morphology-density correlations in COSMOS cause noticeable changes in the galaxy populations in narrow redshift slices at a level that could result in difficulty using the sample to derive redshift-dependent shear calibrations. Upcoming surveys will study lensing as a function of redshift and therefore need to simulate galaxy samples at different redshifts in order to assess the shear calibration at each redshift.

The paper is structured as follows: in Sec. 2, we describe the data that we use for this study. In Sec. 3, we describe our methods for deriving the relevant galaxy properties like environment, morphology, and shape. Using these ingredients, we present our results in Sec. 4 and discuss their implications in Sec. 5.

2   DATA

The COSMOS survey (Scoville et al. 2007; Koekemoer et al. 2007; Leauthaud et al. 2007) is a flux-limited, narrow deep field survey covering a contiguous area of 1.64 deg^2 of sky, with images taken using the Advanced Camera for Surveys (ACS) Wide Field Channel (WFC) in the Hubble Space Telescope (HST). We use the COSMOS survey to define a parent sample of galaxy images to be used for making image simulations, following the approach taken to this problem in Mandelbaum et al. (2012, 2014).

We apply the following set of initial cuts to the COSMOS data, the first two of which are motivated and explained in more detail by Leauthaud et al. (2007):

(i) `MU_CLASS=1`: This criterion uses a comparison between the peak surface brightness and the background level to achieve a robust star/galaxy separation, with galaxies having `MU_CLASS=1`.

(ii) `CLEAN=1`: Objects near bright stars or those containing saturated pixels were removed; the rest pass this cut on `CLEAN`.

(iii) `GOOD_ZPHOT_SOURCE =1`: This cut requires that photometric redshifts be reliable and good enough to draw conclusions about the population (see Mandelbaum et al. 2012 for details).

High resolution images taken through the wide F814W filter (broad I) for all galaxies passing the above cuts were used to create a collection of postage stamp images for the GREAT3 challenge (Mandelbaum et al. 2014), using the procedure described in Mandelbaum et al. (2012). Each galaxy postage stamp image has a corresponding PSF image that can be used by GALSIM or other software to remove the effects of the *HST* PSF before simulating the galaxy image as it would appear at lower resolution.

To better characterize the galaxy population, parametric models were fit to the light profiles of these galaxies. These were carried out using the method described in Lackner & Gunn (2012), and include Sérsic profile fits and 2 component bulge + disk fits described in detail in Mandelbaum et al. (2014) and briefly in Sec. 3.3 of this work.

In addition to the ACS/WFC (F814W) imaging, the COSMOS field has also been imaged by Subaru Suprime-Cam, the Canada-French Hawaii Telescope (CFHT) and KPNO/CTIO, yielding many bands of imaging data from which to determine high-fidelity photometric redshifts. Photometric redshifts were determined by Ilbert et al. (2009). The accuracy of photometric redshifts for $m_{F814W} \leqslant 22.5$ is $\sigma_{\Delta z} = 0.007(1 + z)$; for $m_{F814W} \leqslant 24$, $\sigma_{\Delta z} = 0.012(1 + z)$. The photometric redshift values become noisier beyond $z \sim 1.2$, and the fits to the galaxy light profiles are also somewhat noisy once we go beyond $m_{F814W} \sim 23.5$. For this reason, we will exclude all galaxies that have F814W magnitude fainter than 23.5. However, we will use the $m_{F814W} \leqslant 25.2$ sample that was generated for the GREAT3 challenge to estimate the completeness, which is useful when generating a volume-limited sample (Sec. 3.2). We first use the $z < 1$ flux-limited sample to fit parametric redshift distribution models (Sec. 3.1), and then restrict ourselves to $z \leqslant 1$ sample for all further analysis.

Stellar mass estimates were obtained (Leauthaud et al. 2010) using the Bayesian code described in Bundy et al. (2006). This process involves constructing a grid of models that vary in age, star formation history, dust content and metallicity (always assuming a Chabrier IMF; Chabrier

---

[2] https://github.com/GalSim-developers/GalSim

2003), to which the observed galaxy spectral energy distributions (SEDs) and photometric redshift are compared. At each grid point, the probability that the SED fits the model is calculated, and by marginalizing over the nuisance parameters in the grid, the stellar mass probability distribution is obtained. The median of this distribution is taken as the stellar mass estimate.

3   METHODS

In order to study the variation in the intrinsic ellipticity distribution and various morphological indicators with the galaxy environment, there are three main steps to be carried out:

(i) Identify overdense and underdense environments in our survey from the redshift distribution of galaxies (Sec. 3.1);

(ii) volume-limit the sample such that Malmquist bias is minimized before comparing galaxies in different redshift slices (Sec. 3.2); and

(iii) estimate the galaxy axis ratios and other morphological indicators such as Sérsic index and bulge-to-total ratios (Sec. 3.3).

In this section we will describe how these steps were carried out.

3.1   Finding overdensities

It is important to keep in mind when considering the environment estimation that our goal is not to create a full 3D mapping of the density field within the COSMOS region (a task that was already addressed by Kovač et al. 2010 using the zCOSMOS spectroscopic sample). Instead, we make a coarse, 1D division of the COSMOS survey into redshift slices, just as would be done when making galaxy redshift slices as input to a weak lensing survey simulation. For each redshift slice, we can then check whether the environment is overdense or underdense on average. Our approach will tend to wash out some real trends from a 3D study, but is appropriate given our scientific goal of testing effects of the environment on weak lensing simulations based on the COSMOS survey.

For our (flux-limited) sample of galaxies, up to $z = 1.0$, we fit parametric models to the histograms of photometric redshifts in order to assign values of overdensity. We choose our bins to be 0.05 wide starting from $z = 0.3$, where the bin width is selected to be somewhat larger than the photometric redshift error but narrow enough that we can still identify rather than averaging over real cosmological structures. We neglect the lowest redshifts which have negligible cosmological volume and where the galaxy population tends to be intrinsically bright and large enough that a non-negligible fraction is lost due to the cuts we impose (Sec. 2).

The parametric redshift distributions that we use are

$$p_1(z) \propto z^{a-1} \exp(-bz) \qquad (1)$$

and

$$p_2(z) \propto z^2 \exp\left(-\left(\frac{z}{z_0}\right)^{1.5}\right), \qquad (2)$$
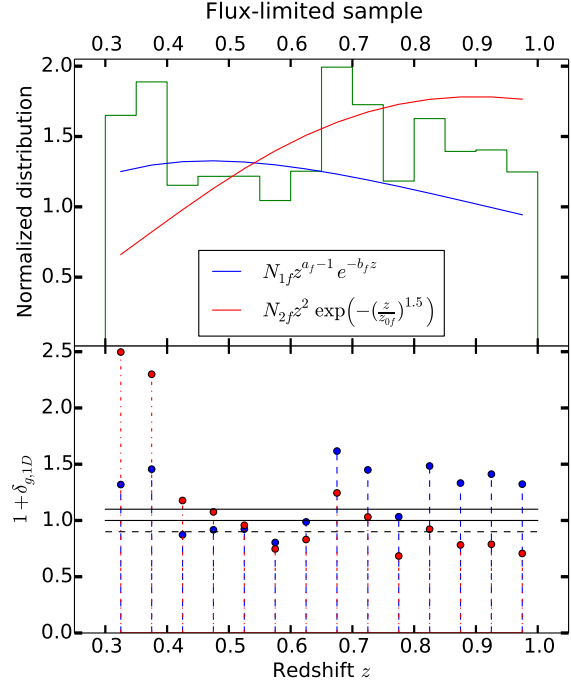
Figure 1. Upper panel: Redshift distribution of flux-limited ($m_{\mathrm{F814W}} \leqslant 23.5$) sample with photometric redshift bins that are 0.05 wide. Two analytical functions with best fit parameters are plotted over it. Lower panel: Plot of $(1 + \delta_{g,\mathrm{1D}}) = N/N_{\mathrm{mod}}$ with each functional form as the model for each redshift bin. Give parameter values. Set minor ticks. Change x axis label to "Photometric redshift (z)". Use thicker line width for the actual histogram. Change legend to say "$p_1(z)$ (Eq. 1)" and "$p_2(z)$ (Eq. 2)" rather than giving the actual equation. Also, why are two of the horizontal lines in the bottom panel solid and one dashed? What decides the style for each line? Finally, I think it makes sense to plot δ rather than $1 + \delta$.

the latter of which was first presented by Baugh & Efstathiou (1993). Here a, b, and z_0 are free parameters that are to be determined. The normalization constants depend not only on the parameters but also on the lower and the upper limit of the redshifts considered, where we fix the normalization to ensure that the predicted number of galaxies in the range used ($0.3 < z < 1$) is equal to the actual number. Fig. 1 shows the photometric redshift histogram together with the best-fitting parametric distributions. The results for the 2nd redshift distribution do not look that great. Does it help if we go to $z < 1.25$? I thought the photo-z were okay in that regime, and it's only beyond 1.25 that they get really scary. We could use up to 1.25 for determining the redshift distribution, even if we're only trying to find overdensities for $z < 1$.

The estimated overdensity in a redshift bin is defined by comparing the observed galaxy counts in the bin with the counts that are predicted in that bin by one of the models in Eqs. (1) and (2):

$$\delta_{g,\mathrm{1D}} = \frac{(N - N_{\mathrm{mod}})}{N_{\mathrm{mod}}}, \qquad (3)$$

where

$$N_{\mathrm{mod}} = \int_{z_{\min}}^{z_{\max}} p(z)\,\mathrm{d}z \qquad (4)$$

is determined by integrating the redshift distribution within the limits of that redshift slice. Note that $\delta_{g,\mathrm{1D}}$ is dependent on our choice of model redshift distribution, and should have a mean value of 0 over the entire redshift range.

Our decision criterion for identifying overdense and underdense redshift slices involves leaving a 10 per cent margin around an overdensity of zero; i.e., if $|\delta_{g,\mathrm{1D}}| < 0.1$, that is considered "neutral" (neither overdense nor underdense on average). We can then label each redshift slice as either overdense, underdense, or neutral as follows: We label a redshift bin as overdense if at least one model gives a value of $\delta_{g,\mathrm{1D}} > 0.1$ while the other gives $\delta_{g,\mathrm{1D}} > -0.1$ (neutral or overdense), and vice versa for the underdense regions. We label a redshift bin as neutral if both models give $\delta_{g,\mathrm{1D}}$ within the neutral region, *or* if use of one model redshift distribution results in the conclusion that the bin is overdense while the over leads to the conclusion that it is underdense. We thus identify the regions $z = 0.30 - 0.40$, $0.65 - 0.75$, and $0.80 - 0.85$ as overdense; $z = 0.55 - 0.65$ and $0.75 - 0.80$ as underdense; and $z = 0.40 - 0.55$ as unclassified.

We have adopted this purely 1D environment classification for reasons explained at the beginning of this section. However, as a sanity check we can compare it with a more rigorous study that includes information about structure in the plane of the sky. Kovač et al. (2010) used a sample of $\sim$10 000 zCOSMOS spectroscopic galaxies with $I_{AB} < 22.5$ to reconstruct the three dimensional overdensity field up to $z \sim 1$. We find that our classification of overdensities and underdensities agrees with this work, except for our two highest redshift bins. We believe that this disagreement is due to the errors in our photometric redshifts, with the overdensity reported by Kovač et al. (2010) in the $z = 0.875 - 1$ range leaking into our $z = 0.80 - 0.85$ slice.
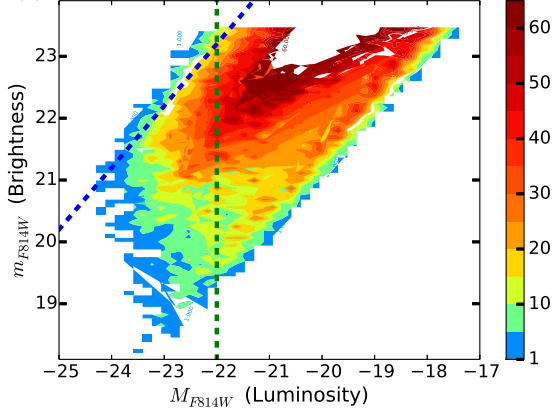
Figure 2. 2D histogram of galaxies in apparent magnitude (m_{F814W}) and absolute magnitude (M_I). What redshift range?
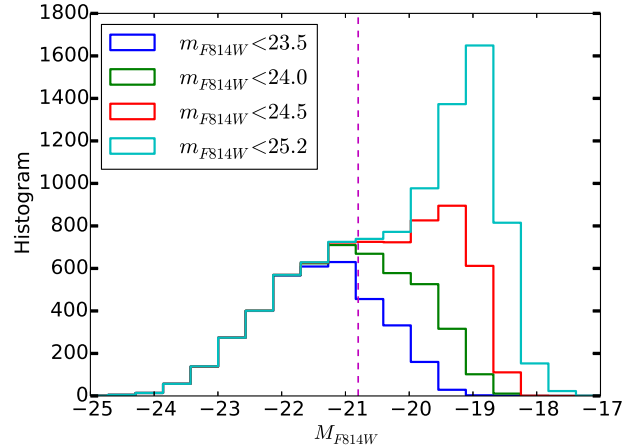
Figure 3. Distribution of absolute magnitude M_I for various flux-limited samples are plotted together. The vertical line corresponds to our luminosity cut of -20.8, brighter than which the $m_{\mathrm{F814W}} < 23.5$ sample includes > 95.3 per cent of the galaxies in the $m_{\mathrm{F814W}} < 25.2$ sample. Is this plot and the numbers in the caption for $0.8 < z < 0.85$ only? This is currently implied by the text. If that is true, then please fix the figure and caption to say this. If that is wrong, then please fix the text.

3.2   Volume limiting

COSMOS is a flux-limited survey and is therefore affected by Malmquist bias, with the galaxy samples at higher redshift being intrinsically brighter on average. Our analysis involves comparing galaxies in different redshift slices to identity significant differences in morphology that arise due to morphology-density correlations. Such an analysis would be very difficult with a flux-limited sample because there would be some variation in morphology with redshift just due to the intrinsic change in the sample properties. For a fair comparison, we must restrict ourselves to galaxies that are bright enough that they would be observed at all redshifts that we consider, which is achieved by volume-limiting the sample. We consider three different ways of carrying out this process, which results in three different galaxy sample selections, all of which we will use in the remainder of the analysis.

Our first approach is to generate a volume-limited sample that is complete up to $z = 1$, by applying a cut on luminosity such that only galaxies instrinsically brighter than a certain threshold (determined in detail below) is considered. This threshold is set on the k-corrected I-band absolute magnitudes (M_I) from the COSMOS PSF-matched photometry catalog. The joint distribution of m_{F814W} and M_I is shown in Fig. 2. Since the parent sample contains fainter galaxies and is quite complete to $m_{\mathrm{F814W}} = 25.2$, we compare the M_I distribution of the $m_{\mathrm{F814W}} = 23.5$ sample with flux-limited samples that have fainter flux limits, to see where the $m_{\mathrm{F814W}} < 23.5$ sample that we want to use for our tests is no longer complete. At $M_I \sim -22.0$, the $m_{\mathrm{F814W}} < 23.5$ sample is beginning to lose galaxies in the $0.9 < z < 1.0$ redshift bin due to the flux limit. However, the $0.85 < z < 1$ redshift bin was found to be only moderately overdense, so we choose to disregard this region for the rest of the analysis, and instead restrict to $z < 0.85$, which is advantageous because it allows us to choose a some-

what fainter intrinsic luminosity limit for the analysis. We relax our luminosity cut so that the sample is volume-limited *not* until $z = 1$ but until $z = 0.85$. We impose the cut at $M_I = -20.8$, which gives 95.3 per cent completeness in the $0.8 < z \leqslant 0.85$ bin (see Fig. 3). The resulting sample, which has 11 169 galaxies, will be called sample $\mathcal{S}1$ in the remainder of this work.

However, previous studies (e.g., Wolf et al. 2003; Giallongo et al. 2005; Willmer et al. 2006; Faber et al. 2007b) have shown that galaxy intrinsic luminosities evolve with redshift. Thus, we should also let the luminosity cut that we apply to volume-limit the sample evolve with redshift. Unfortunately, the majority of the published work on evolution of the luminosity function uses B and V band data, and it is not apparent that the results should be the same in a redder passband like I. We use the results from Faber et al. (2007b) for the evolution of B-band magnitudes from the DEEP2 and COMBO-17 surveys, which is $\Delta M_B^* \sim -1.23$ mag per unit redshift (with the sign indicating that galaxies were intrinsically brighter in the past), for a combined sample of blue and red galaxies. Typically, estimates of evolution in the much redder K band are less than the estimates of evolution in B and V bands. (Please include a citation for this statement.) Assuming that the evolution is a smooth function of the wavelength, the evolution in I-band should be in between B and K band. Therefore, by considering no evolution (a lower limit) as in our $\mathcal{S}1$, and a second sample $\mathcal{S}2$ constructed using the B-band evolution (as an upper bound on the I-band evolution), we can assume that these two samples bracket reality, which should be somewhere in between.

Thus, $\mathcal{S}2$ is constructed by letting the luminosity cut evolve, starting from $M_I = -20.8$ (same as in $\mathcal{S}1$) for the $0.8 < z \leqslant 0.85$ bin. The cut values for the other bins are defined by allowing 1.23 magnitudes of evolution to fainter magnitudes as a function of redshift (evaluated using the bin centers). Because of the sign of redshift evolution, $\mathcal{S}2$ includes more galaxies.

Finally, we can circumvent the problem of redshift evolution of the luminosity by imposing cuts on stellar mass instead. In Fig. 4, we show the stellar mass function (SMF) of our sample for various F814W flux limits. Tomczak et al. (2014) report the SMFs for the ZFOURGE survey, which includes COSMOS. They considered for this work a single stellar population (What do you mean by a single stellar population?) following a Chabrier IMF (Chabrier 2003). We plot their SMF for *all* galaxies in Fig. 4 for reference. Their SMF is higher than ours since they reach a K_s-band 5σ depth of 24.9. As for M_I, we compare the stellar mass function in the $m_{\rm F814W} \leqslant 23.5$ sample with that in the $m_{\rm F814W} \leqslant 25.2$ sample. The sample with $\log(M/M_\odot) > 10.15$ is ~ 95 per cent complete in the redshift bin $0.75 \leqslant z < 0.85$ and has 10 341 galaxies in total across all redshifts. Thus, we construct a third volume-limited sample $\mathcal{S}3$ by imposing the stellar mass cut mentioned above. The number of galaxies in redshift slices are tabulated in Table 1 for all three ways discussed in this section for obtaining a volume-limited sample. The stellar-mass limited sample is the smallest, most likely because when converting from flux to stellar mass, the stellar mass-to-light ratios vary strongly with galaxy type, so red galaxies with high M_*/L simply have too low a flux

Redshift	Environment	$\mathcal{S}1$	$\mathcal{S}2$	$\mathcal{S}3$
0.3-0.4	Overdense	1726	2505	1260
0.4-0.475	Neutral	988	1317	710
0.475-0.55	Neutral	1410	1788	902
0.55-0.65	Underdense	1797	2193	1183
0.65-0.75	Overdense	4059	4476	2593
0.75-0.8	Underdense	1159	1196	675
0.8-0.85	Overdense	2428	2428	1630

Table 1. List of different redshift bins, their environmental classification and the number of galaxies per redshift bin for volume-limited samples constructed in three different ways: using a hard luminosity cut ($\mathcal{S}1$), using a redshift-dependent luminosity cut ($\mathcal{S}2$) and using stellar-mass cuts ($\mathcal{S}3$).

compared to the blue galaxies at the same M_*, and are not observed.

There is one subtlety in our method used for estimating completeness. We have used the full $m_{\rm F814W} \leqslant 23.5$ sample for identifying overdensities and for the completeness calculations that motivated our definitions of volume-limited samples. (Do the numbers in the table include the full sample or just the subsample with postage stamps?) However, in the rest of this paper, we analyse only those galaxies for which there are postage stamp images used to create weak lensing simulations, in part because this is the sample for which fits to Sérsic profiles were carried out, which is a requirement for our morphology analysis. 12 per cent of the galaxies that pass our cuts do not have an associated postage stamp image. Postage stamps may not exist because, given the size of the galaxy, the size of the postage stamp we want to draw around it (including some blank space) intersects the edge of the CCD. If all galaxies were the same size, this would be a purely random effect, but in fact bigger galaxies are more likely to get excluded by this cut. It is commonly the case that galaxies that are nearby and intrinsically very bright do not have postage stamps associated with them, an effect that is dominant at lower redshifts (and is part of our reason for excluding $z < 0.3$). Our completeness calculation is done at high redshifts, and thus we believe that our conclusions are not affected by this bias.

The functional forms for the (flux-limited) redshift distribution that we used in Sec. 3.1 are not physically motivated. We fit them again to the redshift distribution of a volume-limited sample ($\mathcal{S}1$). Fig. 5 shows that when we do this, the values of $\delta_{g,\rm 1D}$ for the $z = 0.40 - 0.55$ bin increase and are within the $[-0.1, 0.1]$ range that we have defined as neutral. This is the reason that in Sec. 3.1 we classified them as neutral as opposed to underdense. We will see in Sec. 4 that they are more similar to overdense regions as opposed to underdense regions. The other redshift slices seem to exhibit a consistent behavior in Fig. 5 as in Fig. 1.

3.3 Describing galaxy morphology and shape

If galaxies have elliptical isophotes, its shape and size could be defined by the axis ratio and the area enclosed by a boundary isophote. However, in real galaxy images, the boundary may not be well defined and the shape may not be well approximated by an ellipse. More importantly, the effect of smearing due to the point-spread function (PSF) is
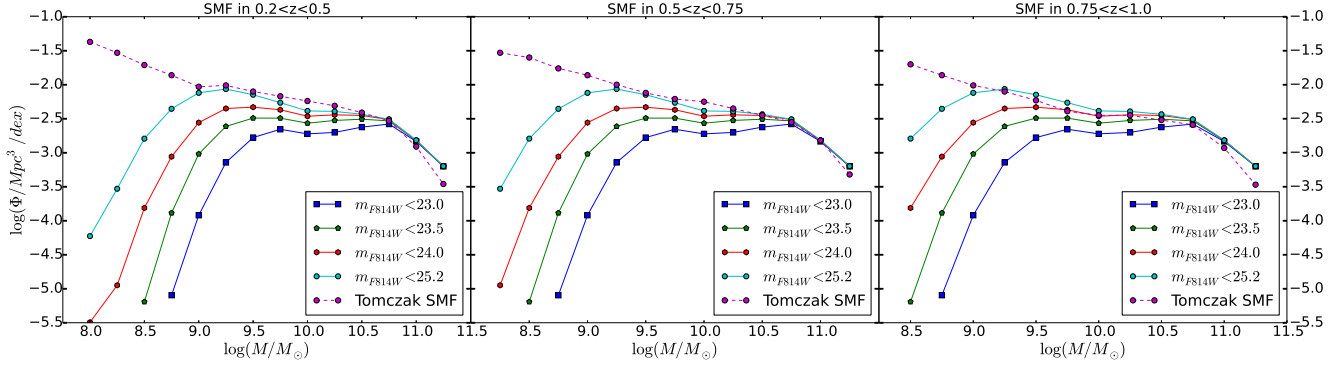
Figure 4. Stellar mass distribution for various flux-limited samples are shown in three redshift ranges as separate panels. The redshift bins have been chosen to facilitate the comparison with a study of the SMF in Tomczak et al. (2014). At high mass, the distributions are the same for various flux limits, indicating that the samples are complete in that mass range. The curves begin to deviate at low masses due to incompleteness coming from the flux limit. The point at which the deviation exceeds our threshold determines where the mass cutoff should be to volume-limit the sample.
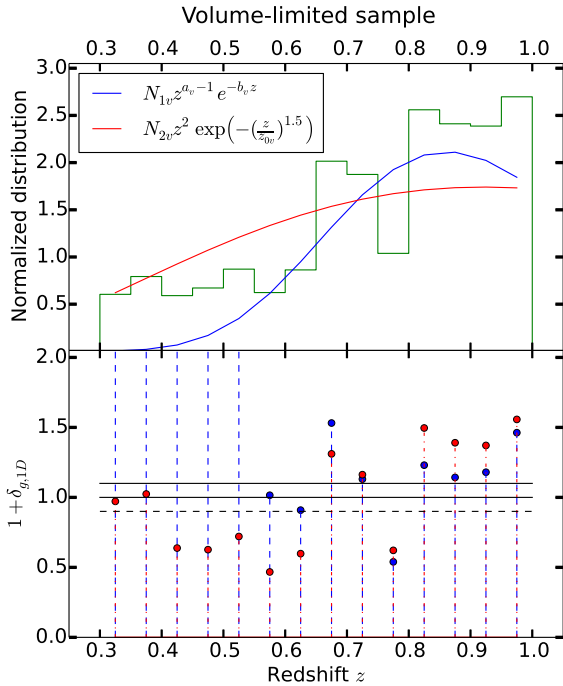
Figure 5. Upper panel: Redshift distribution of volume-limited ($M_I < -20.8$) $\mathcal{S}1$ sample with photometric redshift bins that are 0.05 wide. Two analytical functions with best fit parameters are plotted over it. Lower panel: Plot of $(1 + \delta_{g,1D}) = N/N_{\mathrm{mod}}$ with each functional form as the model for each redshift bin. I am quite concerned that something has gone wrong with this fit (in the sense that there may be a bug). Why is the redshift distribution shown in blue forced to zero at $z = 0.3$? It should go to zero at $z = 0$, not 0.3. It is clearly not normalized to the same total number as the redshift histogram, either. In terms of formatting, please implement the same changes that I suggested for the flux-limited sample redshift distribution plot.

larger. Thus, we are in need of more sophisticated methods to measure the ellipticities.

One method to estimate the ellipticities is to estimate the axis-ratios by fitting parametric models to each image. The models that we fit to the images are

(i) a Sérsic profile given by the expression

$$S = \Sigma_{1/2} \exp\left(-k(R/R_{\mathrm{eff}})^{1/n} - 1\right), \qquad (5)$$

(ii) two Sérsic component fits: de Vaucoulers bulge ($n = 4$) + exponential disc profile ($n = 1$),

where

$$R^2 = ((x - x_0)\cos\Phi + (y - y_0)\sin\Phi)^2$$
$$+ ((y - y_0)\cos\Phi - (x - x_0)\sin\Phi)^2/q^2,$$

R_{eff} is the half-light radius of the profile, $\Sigma_{1/2}$ is the surface brightness at $R = R_{\mathrm{eff}}$, (x_0, y_0) is the centroid of the image, Φ is the profile rotation angle, n is the Sérsic index, k is a n-dependent normalization factor and q is the axis ratio of the elliptical isophotes. Thus, the Sérsic profile has 7 free parameters. The bulge+disk model has 10 free parameters since the Sérsic indices are fixed as 1 and 4 and both the profiles are required to have the same centroid (x_0, y_0). Best-fit parameters were found, by minimizing the weighted sum of the difference between the image and the PSF-convolved model using Levenberg-Marquardt minimization, `mpfit2fun` in IDL (Markwardt 2009). More details about the fit can be found in Lackner & Gunn (2012).

The quantities that we would use from the single Sérsic profile fits are the Sérsic index and axis ratio and from the bulge+disk fits will be the Bulge-to-Total ratio given by $\Sigma_{1/2}(n = 4)/(\Sigma_{1/2}(n = 4) + \Sigma_{1/2}(n = 1))$. I think it would be useful to take the entire sample that we use for science, and show the overall distributions of axis ratio, distortion (one curve for Claire's and one curve for re-Gaussianization), Sérsic index, and Bulge-to-Total. This would be a nearly full-page four panel figure, but I think it's worthwhile to illustrate the nature of the sample. For example, it will clearly show why we cannot use the distributions of Sérsic index and Bulge-to-Total, because of the hard cutoffs.

An alternative method involves correcting the observed image for the PSF and computing the covariances from which the ellipticity of the galaxy can be obtained. The PSF correction scheme used on the observed images is

that of 're-Gaussianization' method described in Sec. 2.4 of Hirata & Seljak (2003) (hereafter HS03). This method models the true PSF $g(\mathbf{x})$ as a Gaussian $G(\mathbf{x})$ and the residual $\epsilon(\mathbf{x}) = g(\mathbf{x}) - G(\mathbf{x})$ is assumed to be small. Thus, the Gaussian-convolved intrinsic image of f, is $I' = G \otimes f = I - \epsilon \otimes f$, where I is the observed image. The crucial idea here is that, when ϵ is small, we get a reasonably accurate estimate of I' even if we use an approximate form for f. The form assumed for f is that of a Gaussian with covariance $M_f^{(0)} = M_I - M_g$, where M_I and M_g are the adaptive covariances of the measured object and PSF respectively, described in Sec. 2.1 of HS03, which is in turn based on Bernstein & Jarvis (2002). Once we obtain the covariance matrices of the instrinsic image f, one can compute the ellipticities of the galaxies, which we will refer to as 'ellipticities based on moments'.

4 RESULTS

Having identified the overdense and underdense regions in a volume-limited sample (Sec. 3.1, Sec. 3.2), we will now see in this section if the morphological parameters of the galaxies, listed in Sec. 3.3, have any dependence on the environment that they reside in. We have 3 different ways of volume-limiting our sample

(i) no redshift evolution of luminosity cut ($\mathcal{S}1$),
(ii) use B-band evolution rate for I-band luminosities ($\mathcal{S}2$),
(iii) impose stellar mass cuts instead of luminosity ($\mathcal{S}3$)

and we will present our results in each of the 3 cases.

4.1 Axis-ratios

We can understand the influence of environment on the ellipticities of the galaxies simply by comparing the distribution of the axis ratios for the overdense and underdense samples. Then, we proceed to consider the root mean square ellipticity, a statistic that can characterize the shape of the population/sample of galaxies.

4.1.1 Comparing distributions

We use two statistical tests namely the Kolmogorov-Smirnov test and Anderson-Darling test to compare distributions. The Anderson-Darling test is carried out using the `adk` package[3] in `R`.

We first compare the distribution of the axis ratio in *all* overdense bins and *all* underdense bins in Fig. 6. How was the volume-limiting carried out for this? You should say the answer to this question for all figures. The caption mentions it. I'm afraid that's not enough. Some people are going to read linearly and you want the info in the main text, too. Unless otherwise mentioned, the axis ratios refer to the values obtained using the method of Lackner & Gunn (2012). The cumulative distinction functions are also plotted alongside in order to be able to visualize the 'distance' between the distributions. Referring to the first two rows in Table 2,

[3] `http://www.inside-r.org/packages/cran/adk/docs/adk.test`

Redshift bins	$\mathcal{S}1$	$\mathcal{S}2$	$\mathcal{S}3$
All overdense vs.	1.1×10^{-4}	2.6×10^{-5}	1.9×10^{-6}
All underdense	1×10^{-5}	$< 1 \times 10^{-5}$	$< 1 \times 10^{-5}$
$z : [0.65, 0.75]$ (OD) vs.	0.613	0.431	0.231
$z : [0.80, 0.85]$ (OD)	0.494	0.237	0.130
$z : [0.65, 0.75]$ (OD) vs.	5.8×10^{-4}	1.5×10^{-5}	3.5×10^{-6}
$z : [0.55, 0.65]$ (UD)	9.8×10^{-4}	$< 1 \times 10^{-5}$	$< 1 \times 10^{-5}$

Table 2. *p*-values from the Kolmogorov-Smirnov (top) and Anderson-Darling (bottom) obtained by comparing the distributions of axis ratios are given for three cases: *all* overdense (OD) vs. *all* underdense (UD), two overdense bins, not very separated in redshift and a pair of adjacent overdense and underdense bins. $\mathcal{S}1$, $\mathcal{S}2$, $\mathcal{S}3$ refer to the volume-limited sample in three different ways. The Anderson-Darling *p*-values are correct only upto 5 decimal places.

Redshift bins	$\mathcal{S}1$	$\mathcal{S}2$	$\mathcal{S}3$
All overdense vs.	5.6×10^{-4}	1.0×10^{-4}	3.3×10^{-6}
All underdense	3×10^{-5}	1×10^{-5}	$< 1 \times 10^{-5}$
$z : [0.65, 0.75]$ (OD) vs.	0.9563	0.7476	0.5359
$z : [0.80, 0.85]$ (OD)	0.5162	0.3352	0.2290
$z : [0.65, 0.75]$ (OD) vs	6.0×10^{-3}	2.5×10^{-4}	2.4×10^4
$z : [0.55, 0.65]$ (UD)	1.2×10^{-2}	2.5×10^{-4}	5×10^{-5}

Table 3. *p*-values from the Kolmogorov-Smirnov (top) and Anderson-Darling (bottom) obtained by comparing the ellipticities computed using second moments are given for three cases: *all* overdense (OD) vs. *all* underdense (UD), two overdense bins, not very separated in redshift and a pair of adjacent overdense and underdense bins. $\mathcal{S}1$, $\mathcal{S}2$, $\mathcal{S}3$ refer to the volume-limited sample in three different ways. Should I write only two significant digits for higher *p*-values? The Anderson-Darling *p*-values are correct only upto 5 decimal places.

we see that the p-values from both the KS and AD tests are well below 0.05, so we reject the 'null hypothesis' that the overdense and underdense regions have same axis ratio distributions at 95% significance level.

One might imagine that the disagreement between the distributions is, at least partly, due to the fact that the overdense and underdense sample have different redshift distributions. To show that that is not the case, we will compare distributions between two overdense / underdense redshift bins, where we expect to find similarity, and between an overdense and underdense regions, where we expect the distributions to differ. Figures 7 and 8 are examples showing that the distributions are indeed similar when the environments are similar and different when the environments are different, confirming our expectation. Similar conclusion can be arrived at by comparing other redshift bins.

Such a trend is also observed in axis ratios based on second moments. If Q_{ij} represents the generic matrix element of the matrix of second moments, then define a complex number, sometimes known as the *distortion*, as

$$e = e_1 + ie_2 = \frac{Q_{11} - Q_{22} + 2iQ_{12}}{Q_{11} + Q_{22}}. \tag{6}$$
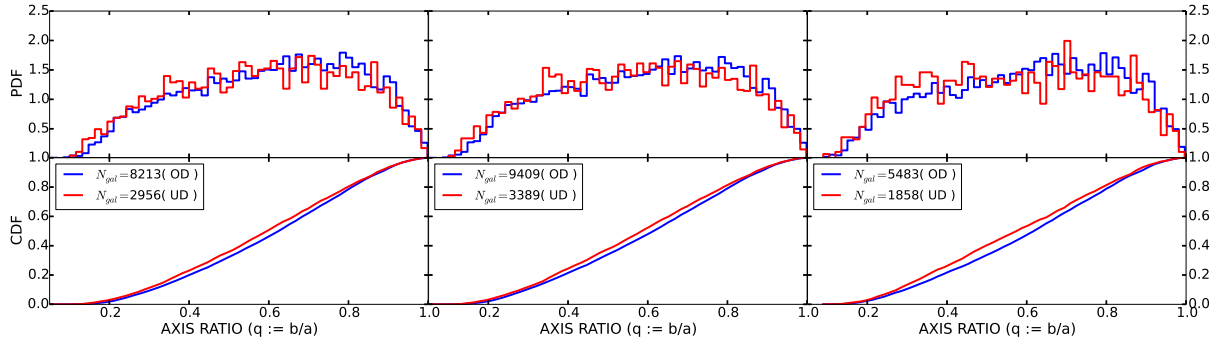
Figure 6. The distributions of axis ratios of galaxies in *all* overdense (OD) and *all* underdense (UD) regions in the case of luminosity-selected sample (left), luminosity-selected with B-band evolution taken into account (center) and stellar-mass-selected sample (right). The upper panels show the histogram and the botom panels show the empirical cumulative distribution function (ECDF). p-values are computed using these CDFs using the Kolmogorov-Smirnov and Anderson-Darling tests and are given in Table. 2. The difference between the CDFs turn out to be statistically significant. This figure and the ones below appear to be right-justified (pushed up against the right margin) instead of centered. Can you look into fixing that? Also, the := is not very standard notation, I recommend just =. Finally, the spacing in the legends is funny; I would prefer to see things like "8213 (OD)" rather than the current "8213(OD)". The former is also consistent with the next figure.
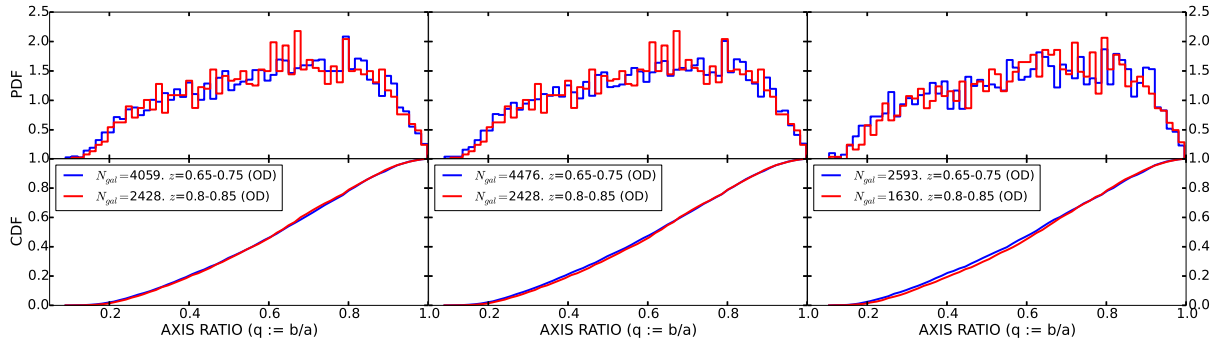
Figure 7. Comparison in similar environments: Axis ratios of galaxies in two overdense redshift bins, $z = 0.65 - 0.75$ and $z = 0.80 - 0.85$ are compared. p-values from the KS and AD test are given in Table. 2

Then, one definition of ellipticity is the magnitude of this complex number which is $\sqrt{e_1^2 + e_2^2}$.

After neglecting a small fraction ($\sim 0.0065\%$) of galaxies (36 out of 55991 galaxies between $0.3 \leqslant z < 0.85$ with $\log(M/M_\odot) > 10.15$) for which the moments do not converge, we make comparisons to the above between the distribution of ellipticities computed from the second moments and the results are tabulated in Table 3. Once again, we see that the distributions are consistent with each other when the environment is similar and are inconsistent when the environments are constrasting.

4.1.2 RMS ellipticites

For the luminosity-selected sample without any evolution, RMS ellipticities of galaxies in each redshift bin are shown in Figure 9. As one can see, the underdense regions have significantly higher values for RMS ellipticities when compared to the overdense regions. Note in particular that we've been able to capture the narrow overdense bin $0.75 \leqslant z < 0.80$. There is (almost) no redshift dependence in the figure.

The dependence on the local environment seems to make sense. Overdense regions typically have young, spiral galaxies which have high q and hence low RMS ellipticity. On the contrary, the underdense regions typically have old, elliptical galaxies and thus low q and high RMS ellipticity.

From Figs. 1 and 5, the region $0.4 \leqslant z < 0.55$ show signs of being marginally underdense but have low RMS ellipticities too that agree with the rest of the overdense regions.

When the B-band luminosity evolution is taken into account in selecting the sample, a systematic increase in the ellipticity at lower redshifts can be observed. We plot these alongside with stellar-mass selected samples, where a similar trend is found, in Fig. 10. Also, the RMS values of the ellipticites calculated from the second moments within each redshift bin are given in Fig. 11.

4.2 Other morphological parameters

For other morphological parameters such as the Sérsic index and Bulge-to-Total ratio, we do not compare the distributions themselves directly. For Sérsic profile fits, it gets tricky since we truncate the Sérsic index at 6. For Bulge-
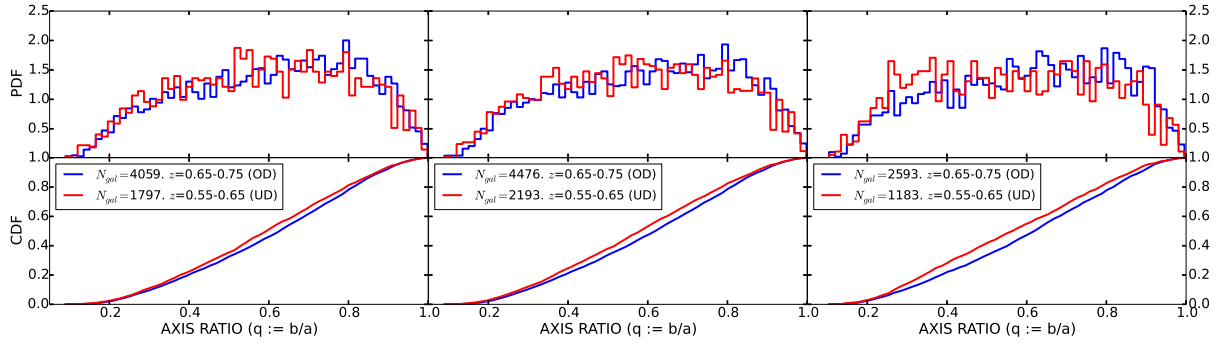
Figure 8. Comparison in contrasting environments: Axis ratios of galaxies in an underdense redshift bin, $z = 0.55 - 0.75$ are compared with those in an overdense bin, $z = 0.65 - 0.75$. p-values from the KS and AD test are given in Table. 2
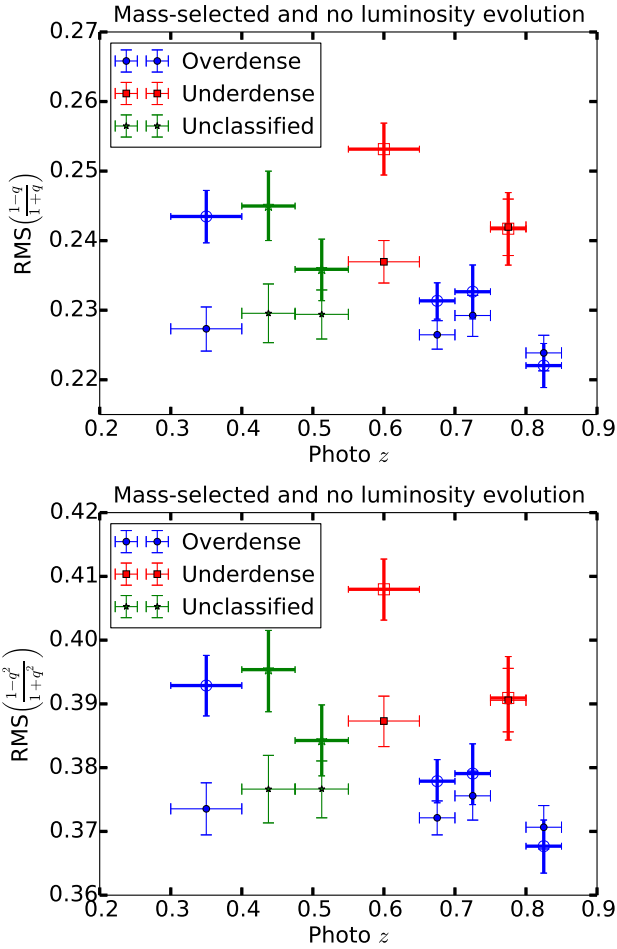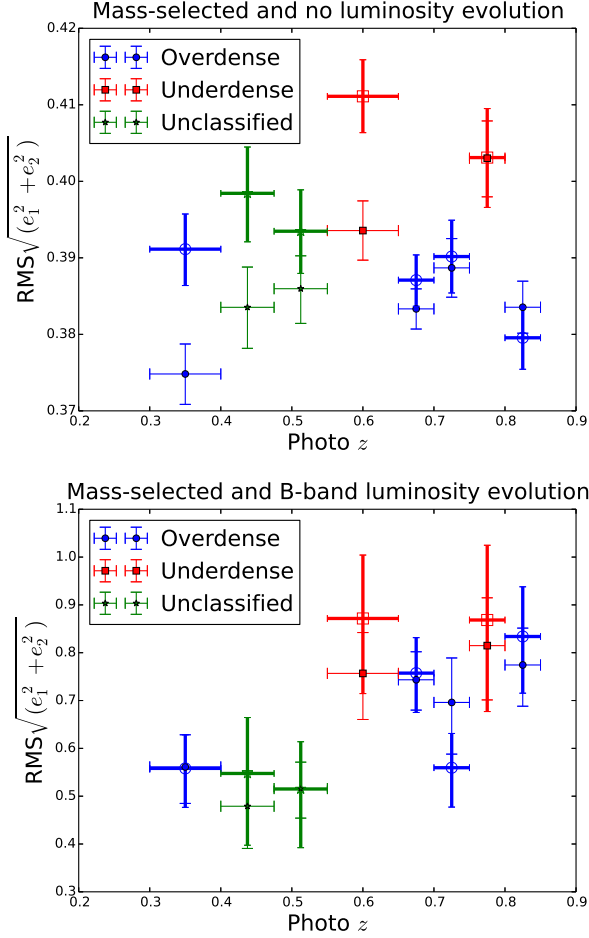

Figure 9. Left: RMS ellipticities with ellipticity defined as $\frac{1-q}{1+q}$.

Right: RMS ellipticities with ellipticity defined as $\frac{1-q^2}{1+q^2}$.

The horizontal errorbars simply correspond to the binwidth while the vertical ones are 1σ errorbars obtained by bootstrapping. I'm thinking these are so similar we just need to show one figure, and can say the results for the other in words. The same goes for the next figure.
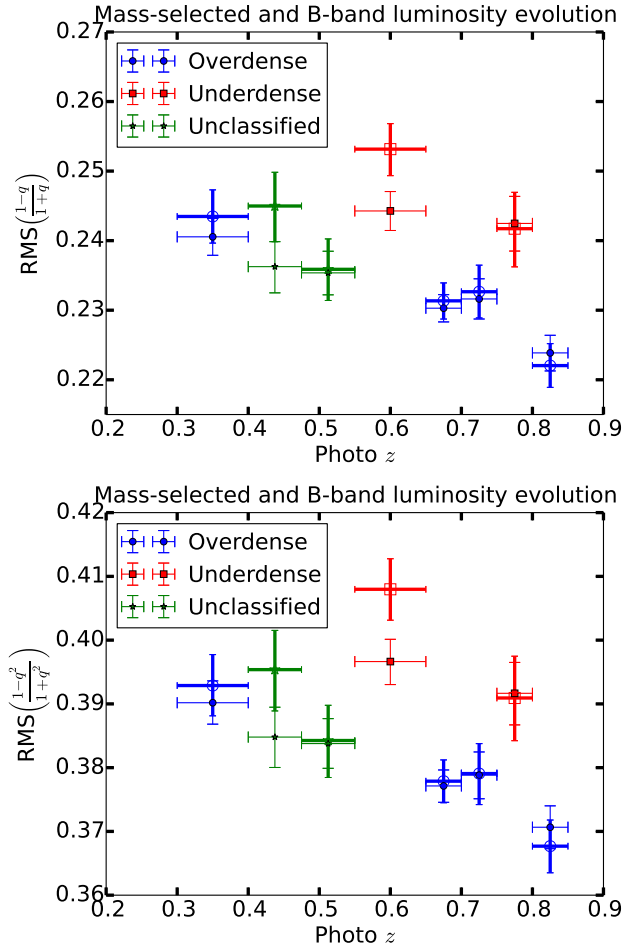
Figure 10. Left: RMS ellipticities with ellipticity defined as $\frac{1-q}{1+q}$.

Right: RMS ellipticities with ellipticity defined as $\frac{1-q^2}{1+q^2}$.

The horizontal errorbars simply correspond to the binwidth while the vertical ones are 1σ errorbars obtained by bootstrapping. The solid points correspond to the sample where the luminosity cut evolves by -1.23 per unit redshift ($\mathcal{S}2$) and the unfilled points correspond to the sample obtained from stellar mass cuts ($\mathcal{S}3$).

to-Total ratios, we force the value to be between 0.05 and 0.95. We will understand the dependence of these quantities

on environment by computing the median values in different redshift bins. We choose median over mean since it is more

Mass-selected and no luminosity evolution

Mass-selected and B-band luminosity evolution

Figure 11. RMS ellipticies with ellipticity defined as $\sqrt{e_1^2 + e_2^2}$, where e_1 and e_2 are the real and imaginary components of the distortion defined in Eq. 6. The horizontal errorbars simply correspond to the binwidth while the vertical ones are 1σ errorbars obtained by bootstrapping. The solid points correspond to the luminosity selected sample ($\mathcal{S}1$ sample on the left and $\mathcal{S}2$ sample on the right) and the unfilled points correspond to the sample obtained from stellar mass cuts ($\mathcal{S}3$).

robust to the truncation effects. Fig. 12 show the median of the Sérsic index with and without taking into account of the luminosity evolution. Median values obtained using stellar mass selected samples are plotted alongside for reference. We observe that the overdense regions tend to have higher Sérsic index than the adjacent underdense ones. Fig. 13 show the variation of median Bulge-to-Total ratio with redshift. We see that the bulge component is more in overdense regions than in the underdense regions. These two observations are consistent with each other since higher Sérsic index implies higher bulge component which are typical in young galaxies found in overdense regions.

In Fig. 12, we see that the Sérsic indices of mass selected sample ($\mathcal{S}3$) are systematically greater than those of the luminosity selected samples ($\mathcal{S}1,\mathcal{S}2$). This is consistent with the Bulge-to-Total values being higher in $\mathcal{S}3$ than in $\mathcal{S}1$ or $\mathcal{S}2$ in Fig. 13. This is not just an edge effect but can be seen from the distributions themselves that the mass se-

lected sample doesn't include as many disk-like galaxies as in luminosity selected samples.

Mass-selected and B-band luminosity evolution

Mass-selected and no luminosity evolution
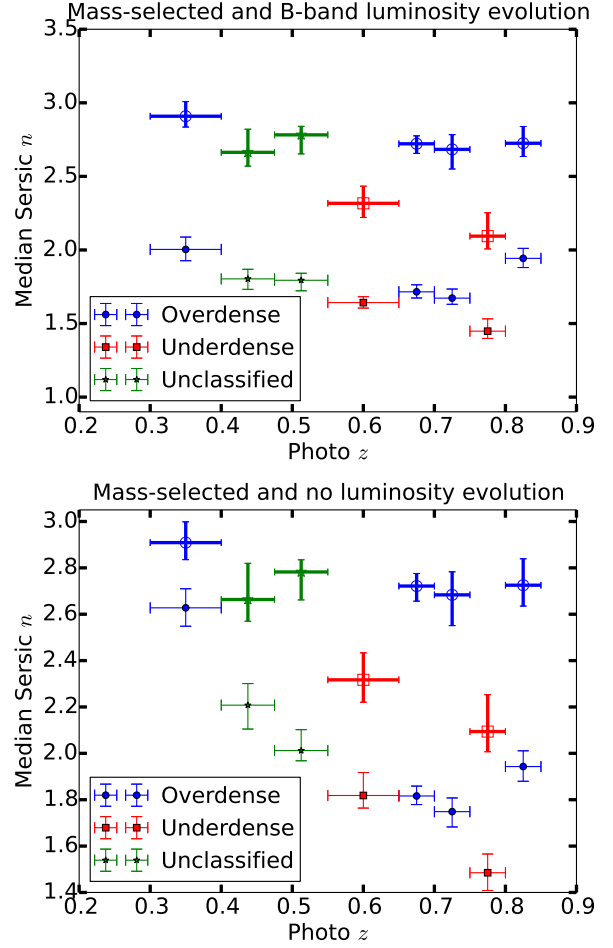
Figure 12. Median values of the Sérsic indices for volume-limited samples $\mathcal{S}1$ and $\mathcal{S}3$ are plotted (filled centers and thin errorbars) in left and right panels respectively for each redshift bin. Median values for the $\mathcal{S}2$ sample are plotted in both the panels (open centers and thick errorbars) in both the panels. The horizontal errorbars simply correspond to the binwidth while the vertical ones are 1σ errorbars obtained by bootstrapping. Only need to have a legend on one panel, since the legends on both panels are the same. Same comment applies to next figure.

5   CONCLUSIONS

In this study, we have shown that morphological parameters of galaxies depend on the local environments along the line of sight in a manner than affect Weak Lensing simulations. In studies of Weak Lensing, one simulates galaxy images in a redshift slice by learning from the images in the same redshift bin of a training sample like COSMOS, which we have used for our analysis here. The survey volume is broken up into multiple redshift slices and are classified as 'overdense', 'underdense' or 'neutral' according to their local environment. This is done by comparing the 1-D redshift distribution to some of the parametric models available in
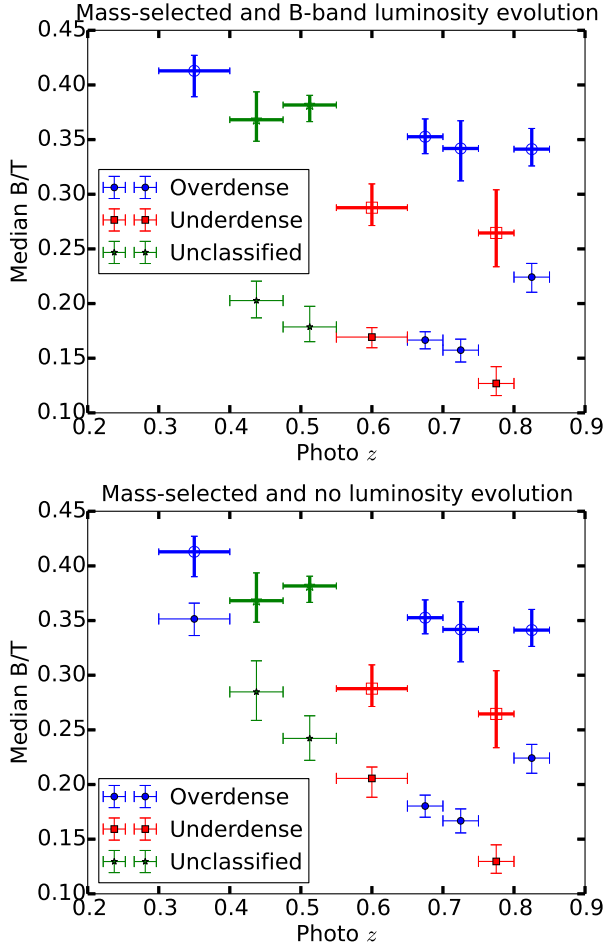
Figure 13. Median values of the Bulge-to-Total ratios for volume-limited samples $\mathcal{S}1$ and $\mathcal{S}3$ are plotted (filled centers and thin errorbars) in left and right panels respectively for each redshift bin. Median values for the $\mathcal{S}2$ sample are plotted in both the panels (open centers and thick errorbars) in both the panels. The horizontal errorbars simply correspond to the binwidth while the vertical ones are 1σ errorbars obtained by bootstrapping. These look so similar that I am questioning the need to have both. Perhaps remove one and mention that they are nearly identical in the text.

the literature. The incompleteness in the sample is minimized by constructing a volume limited sample either by imposing stellar mass cuts ($\mathcal{S}3$) or by imposing a luminosity cut. Further, depending on whether we have the cut fixed or evolve with redshift, we get two more volume limited samples $\mathcal{S}1$ and $\mathcal{S}2$ respectively. Morphological parameters are obtained by fitting a single Sérsic profile to the images or by a two component (bulge+disk) fit. The morphological quantities of interest include axis ratios or equivalently the ellipticities, Sérsic indices and Bulge-to-Total ratios.

For all the three volume-limited samples, we compare the distributions of the axis-ratios of the galaxies in overdense and underdense regions and conclude that the distributions are different after performing statistical tests on them. From the axis ratios, we compute ellipticities and find that the root mean squared value of the ellipticities of galaxies in a redshift bin vary significantly between the overdense and underdense regions. Such a behavior is also observed

when ellipticities are computed using second moments instead of a parametric model fitting. Sérsic index and Bulge-to-Total ratios also exhibit similar trends with redshift based on the local environment.

Our result has serious implications for realistic image simulations for Weak Lensing. It suggests that the training sample is affected by cosmic variance so as to possibly bias the conclusions from the simulations. This is particularly a problem with narrow surveys where a single galaxy cluster or a void could affect the environment significantly. Thus, we are forced to use wider redshift bins, much larger than the uncertainties in the redshifts, so as to be able to wash out the environmental dependence. However, future surveys such as the Large Synoptic Survey Telescope(LSST; LSST Science Collaboration et al. 2009) will image larger areas of sky mitigating the dependence of local environment in image simulations.

REFERENCES

Albrecht A. et al., 2006, ArXiv Astrophysics e-prints
Bartelmann M., Schneider P., 2001, Phys.Rep., 340, 291
Baugh C. M., Efstathiou G., 1993, MNRAS, 265, 145
Bernstein G. M., 2010, MNRAS, 406, 2793
Bernstein G. M., Jarvis M., 2002, AJ, 123, 583
Bridle S. et al., 2010, MNRAS, 405, 2044
Bundy K. et al., 2006, ApJ, 651, 120
Carollo C. M. et al., 2014, ArXiv e-prints
Chabrier G., 2003, PASP, 115, 763
de Jong J. T. A. et al., 2013, The Messenger, 154, 44
De Propris R. et al., 2014, MNRAS, 444, 2200
Faber S. M. et al., 2007a, ApJ, 665, 265
Faber S. M. et al., 2007b, ApJ, 665, 265
Giallongo E., Salimbeni S., Menci N., Zamorani G., Fontana A., Dickinson M., Cristiani S., Pozzetti L., 2005, ApJ, 622, 116
Green J. et al., 2012, ArXiv e-prints
Heymans C. et al., 2013, MNRAS, 432, 2433
Hirata C., Seljak U., 2003, MNRAS, 343, 459
Ilbert O. et al., 2009, ApJ, 690, 1236
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. et al., 2010, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7733, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series
Kitching T. D. et al., 2012, MNRAS, 423, 3163

Koekemoer A. M. et al., 2007, ApJS, 172, 196

Kovač K. et al., 2010, ApJ, 708, 505

Lackner C. N., Gunn J. E., 2012, MNRAS, 421, 2277

Laureijs R. et al., 2011, ArXiv e-prints

Leauthaud A. et al., 2010, ApJ, 709, 97

Leauthaud A. et al., 2007, ApJS, 172, 219

LSST Science Collaboration et al., 2009, ArXiv e-prints

Mandelbaum R., Hirata C. M., Leauthaud A., Massey R. J., Rhodes J., 2012, MNRAS, 420, 1518

Mandelbaum R. et al., 2014, ApJS, 212, 5

Mandelbaum R., Slosar A., Baldauf T., Seljak U., Hirata C. M., Nakajima R., Reyes R., Smith R. E., 2013, MN-RAS, 432, 1544

Markwardt C. B., 2009, in Astronomical Society of the Pacific Conference Series, Vol. 411, Astronomical Data Analysis Software and Systems XVIII, Bohlender D. A., Durand D., Dowler P., eds., p. 251

Melchior P., Böhnert A., Lombardi M., Bartelmann M., 2010, A&A, 510, A75

Melchior P., Viola M., 2012, MNRAS, 424, 2757

Miyazaki S. et al., 2006, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 6269, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series

Refregier A., Kacprzak T., Amara A., Bridle S., Rowe B., 2012, MNRAS, 425, 1951

Rowe B. et al., 2014, ArXiv e-prints

Schrabback T. et al., 2010, A&A, 516, A63

Scoville N. et al., 2007, ApJS, 172, 1

The Dark Energy Survey Collaboration, 2005, ArXiv Astrophysics e-prints

Tomczak A. R. et al., 2014, ApJ, 783, 85

Troxel M. A., Ishak M., 2014, ArXiv e-prints

van Daalen M. P., Schaye J., Booth C. M., Dalla Vecchia C., 2011, MNRAS, 415, 3649

Voigt L. M., Bridle S. L., 2010, MNRAS, 404, 458

Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, Phys.Rep., 530, 87

Willmer C. N. A. et al., 2006, ApJ, 647, 853

Wolf C., Meisenheimer K., Rix H.-W., Borch A., Dye S., Kleinheinrich M., 2003, A&A, 401, 73