Sizes, Colour gradients and Resolved Stellar Mass Distributions for the Massive Cluster Galaxies in XMMUJ2235-2557 at z = 1.39

Jeffrey C.C. Chan[1,2*], Alessandra Beifiori[2,1], J. Trevor Mendel[1], Roberto P. Saglia[1,2],
Ralf Bender[1,2], Matteo Fossati[2,1], Audrey Galametz[1], Michael Wegner[2],
David J. Wilman[2,1], Michele Cappellari[3], Roger L. Davies[3], Ryan C. W. Houghton[3],
Laura J. Prichard[3], Ian J. Lewis[3], Ray Sharples[4] and John P. Stott[3]

[1]*Max-Planck-Institut für extraterrestrische Physik (MPE), Giessenbachstr. 1, D-85748 Garching, Germany*
[2]*Universitäts-Sternwarte, Ludwig-Maximilians-Universität, Scheinerstrasse 1, D-81679 München, Germany*
[3]*Sub-department of Astrophysics, Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK*
[4]*Centre for Advanced Instrumentation, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*

ABSTRACT
We analyse the sizes, colour gradients, and resolved stellar mass distributions for 36 massive and passive galaxies in the cluster XMMUJ2235-2557 at $z = 1.39$ using optical and near-infrared Hubble Space Telescope imaging. We derive light-weighted Sérsic fits in five HST bands ($i_{775}, z_{850}, Y_{105}, J_{125}, H_{160}$), and find that the size decreases by $\sim 20\%$ going from i_{775} to H_{160} band, consistent with recent studies. We then generate spatially resolved stellar mass maps using an empirical relationship between $M_*/L_{H_{160}}$ and $(z_{850} - H_{160})$ and use these to derive mass-weighted Sérsic fits: the mass-weighted sizes are $\sim 41\%$ smaller than their rest-frame r-band counterparts compared with an average of $\sim 12\%$ at $z \sim 0$. We attribute this evolution to the evolution in the $M_*/L_{H_{160}}$ and colour gradient. Indeed, as expected, the ratio of mass-weighted to light-weighted size is correlated with the M_*/L gradient, but is also mildly correlated with the mass surface density and mass-weighted size. The colour gradients (∇_{z-H}) are mostly negative, with a median value of ~ 0.45 mag dex^{-1}, twice the local value. The evolution is caused by an evolution in age gradients along the semi-major axis (a), with $\nabla_{age} = d\log(age)/d\log(a) \sim -0.33$, while the survival of weaker colour gradients in old, local galaxies implies that metallicity gradients are also required, with $\nabla_Z = d\log(Z)/d\log(a) \sim -0.2$. This is consistent with recent observational evidence for the inside-out growth of passive galaxies at high redshift, and favours a gradual mass growth mechanism, such as minor mergers.

Key words: galaxies: clusters: general – galaxies: elliptical, lenticular, cD – galaxy: evolution – galaxies: formation – galaxies: high-redshift – galaxies: fundamental parameters.

1 INTRODUCTION

The study of galaxy clusters at high redshift has attracted a lot of attention over the last decade, as these large structures provide a unique environment for understanding the formation and evolution of massive galaxies we see in the present day Universe. Massive galaxies in clusters especially in the cluster cores are preferentially in the red passive population, have regular early-type morphology and are mainly composed of old stars (e.g. Dressler 1980; Rosati et al. 2009; Mei et al. 2009). Nevertheless, in higher redshift clusters ($z \gtrsim 1.5$) a substantial massive population are recently found

to be still actively forming stars (e.g. Hayashi et al. 2011; Gobat et al. 2013; Strazzullo et al. 2013; Bayliss et al. 2014). The member passive galaxies reside on a well-defined sequence in colour-magnitude space, namely the red sequence which is seen in clusters up to redshift $z \sim 2$ (e.g. Kodama & Arimoto 1997; Stanford, Eisenhardt & Dickinson 1998; Gobat et al. 2011; Tanaka et al. 2013; Andreon et al. 2014). Previous works have shown that the stars in these galaxies have formed (and the star formation was quenched) early, the stellar mass is largely assembled before $z \sim 1$ (e.g. Lidman et al. 2008; Mancone et al. 2010; Strazzullo et al. 2010; Fassbender et al. 2014). These galaxies then evolve passively in the subsequent time (e.g. Andreon 2008; De Propris, Phillipps & Bremer 2013). Nonetheless, the details of how these massive passive

* E-mail: jchan@mpe.mpg.de

cluster galaxies formed and evolved, in particular the physical processes involved, remains a matter of debate.

An important component of the above question is the evolution of the structure of these passive galaxies over time. It has now been established that galaxies at high redshift are much more compact: those with stellar masses $M_* \geqslant 10^{11} M_\odot$ at $z \sim 2$ have an effective radius of only $\simeq 1$ kpc (e.g. Daddi et al. 2005; Trujillo et al. 2006a). At $z \sim 0$ such massive dense objects are believed to be relatively rare (Trujillo et al. 2009), yet the exact abundance is still under debate (Valentinuzzi et al. 2010a; Trujillo, Carrasco & Ferré-Mateu 2012; Poggianti et al. 2013). Previous studies suggest that massive passive galaxies have grown by a factor of ~ 2 in size since $z \sim 1$ (e.g. Trujillo et al. 2006b; Longhetti et al. 2007; Cimatti et al. 2008; van der Wel et al. 2008; Saglia et al. 2010; Beifiori et al. 2014), and a factor of ~ 4 since $z \sim 2$ (e.g. Trujillo et al. 2007; Buitrago et al. 2008; van Dokkum et al. 2008; Newman et al. 2012; Szomoru, Franx & van Dokkum 2012; Barro et al. 2013; van der Wel et al. 2014). This progressive growth appears to happen mainly at the outer envelopes, as several works have shown that massive ($M_* \gtrsim 1 \times 10^{11} M_\odot$) passive galaxies at high-redshift have comparable central densities to local ellipticals, suggesting the mass assemble took place mainly at outer radii over cosmic time (i.e. the "inside-out" growth scenario, Bezanson et al. 2009; van Dokkum et al. 2010; Patel et al. 2013).

To explain the observed evolution, the physical processes invoked have to result in a large growth in size but not in stellar mass, nor drastic increase in the star formation rate. Most plausible candidates are mass-loss driven adiabatic expansion ("puffing-up") (e.g. Fan et al. 2008, 2010; Ragone-Figueroa & Granato 2011) and dry mergers scenarios (e.g. Bezanson et al. 2009; Naab, Johansson & Ostriker 2009; Trujillo, Ferreras & de La Rosa 2011). In the former scenario, galaxies experience a mass loss from wind driven by active galactic nuclei (AGN) or supernovae feedback, which lead to an expansion in size due to a change in the gravitational potential. In the latter, mergers either major involving merging with another galaxy of comparable mass, or minor that involves accretion of low mass companions, have to be dry to keep the low star formation rate (Trujillo, Ferreras & de La Rosa 2011). Nevertheless, major mergers are not compatible with the observed growth in mass function in clusters as well as the observed major merger rates since $z \sim 1$ (e.g. Nipoti, Londrillo & Ciotti 2003; Bundy et al. 2009). On the other hand, minor mergers are able to produce an efficient size growth (see e.g. Trujillo, Ferreras & de La Rosa 2011; Shankar et al. 2013). The rates of minor mergers are roughly enough to account for the size evolution only up to $z \lesssim 1$ Newman et al. (2012), at z ~ 2 additional mechanisms are required (e.g. AGN feedback-driven star formation Ishibashi, Fabian & Canning 2013). In addition, the effect of continual quenched galaxies onto the red sequence as well as morphological mixing (known as the "progenitor bias") further complicates the situation (e.g van Dokkum & Franx 2001). Processes that are specific in clusters such as harassment, strangulation and ram-pressure stripping (e.g. Treu et al. 2003; Moran et al. 2007) might play an important role in quenching and morphologically transforming galaxies. Several studies have already shown that the progenitor bias has a non-negligible effect on the size evolution (e.g. Saglia et al. 2010; Valentinuzzi et al. 2010b; Carollo et al. 2013; Poggianti et al. 2013; Beifiori et al. 2014; Delaye et al. 2014; Belli, Newman & Ellis 2015; Shankar et al. 2015).

In addition to size or structural parameter measurements, colour gradients also provide valuable information for disentangling the underlying physical processes involved in the evolution of passive galaxies, and have been used as tracers of stellar popu-

lation properties and their radial variation. In local and intermediate redshift passive galaxies, colour gradients are mainly attributed to metallicity gradients (e.g. Saglia et al. 2000; La Barbera et al. 2005; Tortora et al. 2010), although also affected by age and dust (see, e.g. Vulcani et al. 2014). Measuring the colour gradients at high redshift is more challenging due to compact galaxy sizes and limitations on instrumental angular resolution. Passive galaxies at high redshift appear to show negative colour gradients, in the sense that the core is redder than the outskirts (e.g. Wuyts et al. 2010; Guo et al. 2011; Szomoru et al. 2011), implying a radial variation in the stellar mass-to-light ratio (hereafter M_*/L).

Due to M_*/L gradients within the galaxies, the size of the galaxies measured from surface brightness profiles (i.e. luminosity-weighted sizes) is not always a reliable proxy of the mass distribution, especially at high redshifts when the growth of the passive galaxies is more rapid. Hence, measuring characteristic sizes of the mass distribution (i.e. mass-weighted sizes) is preferable over the wavelength dependent luminosity-weighted sizes. Recently a number of works attempted to reconstruct stellar mass profiles taking into account the M_*/L gradients primarily using two techniques: resolved spectral energy distribution (SED) fitting (e.g. Wuyts et al. 2012; Lang et al. 2014) and the use of a scaling M_*/L - colour relation (e.g. Bell & de Jong 2001; Bell et al. 2003). In the former, stellar population modeling is performed on resolved multi-band photometry to infer spatial variations in the stellar population in 2D. While this is a powerful way to derive resolved properties, deep and high-resolution multi-band imaging are required to well constrain the SED in each resolved region, which is not available for most datasets. The latter method, demonstrated by Zibetti, Charlot & Rix (2009) and Szomoru et al. (2013), relies on a M_*/L - colour relation to determine the spatial variation of M_*/L. Although this method cannot disentangle the degeneracy between age, dust and metallicity, it provides a relatively inexpensive way to study the mass distribution of galaxies.

In this study, we analyse a sample of 36 passive galaxies in the massive cluster XMMUJ2235-2557 at $z \sim 1.39$. We focus on their light-weighted sizes (in rest-frame optical, from the near-IR *HST*/WFC3 images), resolved stellar mass distribution, as well as mass-weighted sizes and colour gradients. This paper is organised as follows. The sample and data used in this study are described in Section 2. Object selection, photometry, structural analysis, and the procedure to derive resolved stellar mass surface density maps are described in Section 3. We also examine the reliability of our derived parameters with simulated galaxies and present the results in the same section. In Section 4 we describe the local sample we used for comparison. In Section 5 we present the main results, including both light-weighted and mass-weighted structural parameters derived from the stellar mass surface density maps, colour and M_*/L gradients. The results are then compared with the local sample, and discussed in Section 6. Lastly, in Section 7 we draw our conclusions.

Throughout the paper, we assume the standard flat cosmology with $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_\Lambda = 0.7$ and $\Omega_m = 0.3$. With this cosmological model at redshift 1.39, 1 arcsec corresponds to 8.4347 kpc. Magnitudes quoted are in the AB system (Oke & Gunn 1983). The stellar masses in this paper are computed with a Chabrier (2003) initial mass function (IMF). Quoted published values are transformed to Chabrier IMF when necessary.

Table 1. HST imaging of XMMUJ2235-2557 used in this study.

Name	Filter	Rest pivot wavelength at $z = 1.39$ (Å)	Exposure time (s)
i_{775}	ACS F775W	3215.2	8150
z_{850}	ACS F850LP	3776.1	14400
Y_{105}	WFC3 F105W	4409.5	1212
J_{125}	WFC3 F125W	5217.7	1212
H_{160}	WFC3 F160W	6422.5	1212

2 DATA

2.1 Sample

The cluster XMMUJ2235-2257 was serendipitously detected in an X-ray observation of a nearby galaxy by XMM-Newton and discovered by Mullis et al. (2005). Subsequent VLT/FORS2 spectroscopy confirmed the redshift of the cluster to be $z \sim 1.39$. Rosati et al. (2009) confirmed the cluster membership of 34 galaxies. Among them 16 within the central 1 Mpc are passive. Jee et al. (2009) performed a weak-lensing analysis on the cluster and estimated the projected mass of the cluster to be $\sim 8.5 \times 10^{14}\ M_\odot$, making it one of the most massive clusters seen at high-redshift. Grützbauch et al. (2012) studied the star formation in this cluster out to a projected radius of 1.5 Mpc and found that all massive galaxies have low specific star formation rates, and galaxies in the cluster centre have lower specific star formation rates than the rest of the cluster galaxies at fixed stellar mass. For the galaxy structural properties, this cluster has been investigated by Strazzullo et al. (2010) and was also included in the cluster sample of Delaye et al. (2014) and De Propris, Bremer & Phillipps (2015).

2.2 *HST* imaging

We make use of the deep optical and IR archival imaging of the cluster XMMUJ2235-2557, obtained with *HST*/ACS WFC and *HST*/WFC3 in June 2005 (PID 10698), July 2006 (PID 10496) and April 2010 (PID 12051). The ACS data are mostly from a program designed to search for Type Ia supernovae in galaxy clusters (Dawson et al. 2009), while the WFC3 data are from a calibration program aiming at cross-calibrating the zero point of WFC3 and NICMOS. The *HST*/ACS data consists of F775W and F850LP bands (hereafter i_{775} and z_{850}), while the WFC3 data comprises four IR bands, F105W, F110W, F125W and F160W (hereafter Y_{105}, YJ_{110}, J_{125} and H_{160}). The YJ_{110} data is not used in this study as it has a shorter exposure time. The WFC3 data has a smaller field of view than the ACS data, $145'' \times 126''$. A summary of the observational setup can be found in Table 1.

Data in each band are reduced and combined using Astrodrizzle, an upgraded version of the Multidrizzle package in the `PyRAF` interface (Gonzaga et al. 2012). Relative WCS offsets between individual frames are first corrected using the `tweakreg` task before drizzling. The ACS and WFC3 images have been drizzled to pixel scales of 0.05 and 0.09 arcsec pixel^{-1} respectively. The full-width-half-maximum (FWHM) of the PSF is ~ 0.11 arcsec for the ACS data and ~ 0.18 arcsec for the WFC3 data, measured from median stacked stars. We produce weight maps using both inverse variance map (`IVM`) and error map (`ERR`) settings for different purposes. The `IVM` weight maps, which contain all background noise sources except Poisson noise of the objects, are used for object detection, while the `ERR` weight maps are used for structural analysis as the

Poisson noise of the objects is included. Due to the nature of the drizzle process, the resulting drizzled images have correlated pixel-to-pixel noise. To correct for this we follow Casertano et al. (2000) to apply a scaling factor to the weight maps. Absolute WCS calibrations of the drizzled images are derived using GAIA (Graphical Astronomy and Image analysis Tool) in the Starlink library (Berry et al. 2013) with Guide Star Catalog II (GSC-II) (Lasker et al. 2008).

3 ANALYSIS

3.1 Object detection, sample selection, and photometry

3.1.1 Method

The WFC3 H_{160} image, the reddest available band, is used for object detection with SExtractor (Bertin & Arnouts 1996). The multiband photometry is obtained with SExtractor in dual image mode with the H_{160} image as the detection image. `MAG_AUTO` magnitudes are used for galaxy magnitudes and aperture magnitudes are used for colour measurements. We use a fixed circular aperture size of $1''$ in diameter. The effective radii of most galaxies in the cluster are generally much smaller than the aperture size. Galactic extinction is corrected using the dust map of Schlegel, Finkbeiner & Davis (1998) and the recalibration E(B-V) value from Schlafly & Finkbeiner (2011).

As described in the introduction, Grützbauch et al. (2012) studied the star formation in the cluster out to a projected radius of 1.5 Mpc. We cross-match our SExtractor catalogue to theirs to identify spectroscopically confirmed cluster members from previous literature (Mullis et al. 2005; Lidman et al. 2008; Rosati et al. 2009). 12 out of 14 spectroscopically confirmed cluster members are within the WFC3 FOV and identified. Figure 1 shows the colour-magnitude diagram of the detected sources within the WFC3 FOV. We identify passive galaxies through fitting the red sequence from the colour-magnitude diagram. We measure the scatter through rectifying the $z_{850} - H_{160}$ colour with our fitted relation, then marginalise over the H_{160} magnitude to obtain a number distribution of the galaxies. The dotted lines correspond to $\pm 2\sigma$ derived from a Gaussian fit to the number distribution.

Objects that are within 2σ from the fitted red-sequence are selected as the passive sample. We trim the sample by removing point sources indicated by SExtractor (i.e. those with `class_star` $\geqslant 0.9$) and applying a magnitude cut of $H_{160} < 22.5$, which corresponds to a completeness of $\sim 95\%$ (see below). This selection results in a sample of 36 objects in the cluster XMMUJ2235-2557.

3.1.2 Quantifying the uncertainties on the photometry

Since the photometric uncertainties are folded directly into our mass estimates as well as the structural parameters measurements, a realistic estimate of the photometric uncertainties is required. Previous works have shown that SExtractor tends to underestimate the photometric uncertainties and there can be a small systematic shift between `MAG_AUTO` output and the true magnitudes (Häussler et al. 2007). Hence, we perform an extensive galaxy magnitude and colour test with a set of 50000 simulated galaxies with surface brightness profiles described by a Sérsic profile on the ACS z_{850} and WFC3 H_{160} band images. This set of galaxies is also used for assessing the completeness and accuracy of the light and mass structural parameter measurements. Details of the simulations can
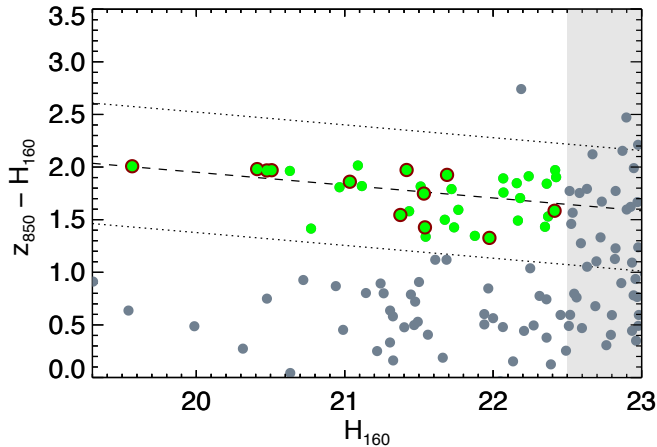
Figure 1. Colour-magnitude diagram of the cluster XMMUJ2235. WFC3 H_{160} magnitudes are `MAG_AUTO` magnitudes while the $z_{850} - H_{160}$ colour are $1''$ aperture magnitudes. The dashed line corresponds to the fitted red sequence and the dotted lines are $\pm 2\sigma$. Green circles correspond to objects that are included in our sample, which are within the dotted line and are not in the shaded area (i.e. $H_{160} < 22.5$). Objects that are spectroscopically confirmed cluster members from the catalogue of Grützbauch et al. (2012) are circled in dark red.

be found in Appendix A. Here we focus on the photometric uncertainties estimates.

The detection rate above a certain magnitude reflects the completeness of the sample at that particular magnitude cut. We find that a magnitude cut of $H_{160} < 22.5$ corresponds to a completeness of $\sim 95\%$. We then assess the accuracy of the recovered magnitudes and colours. Since the accuracies depend strongly on both input magnitude (mag_{in}) as well as the effective semi-major axis (a_e) of the galaxies, we assess the accuracy in terms of input mean surface brightness ($\Sigma = mag_{in} + 2.5 \log(2\pi a_e^2)$ in mag arcsec^{-2}) rather than input magnitudes. Below we quote the results at a mean surface brightness of 23.5 mag arcsec^{-2} in H_{160} (or 24.5 mag arcsec^{-2} in z_{850}) as a benchmark, as most objects we considered are brighter than 23.5 mag arcsec^{-2}.

For ACS z_{850}, the typical 1σ uncertainty for the `MAG_AUTO` output at mean surface brightness of 24.5 mag arcsec^{-2} is ~ 0.33 mag. For WFC3 H_{160}, at a mean surface brightness of 23.5 mag arcsec^{-2} the typical 1σ uncertainty is ~ 0.19 mag. Previous studies have shown that SExtractor `MAG_AUTO` misses a certain amount of flux especially for the faint objects (e.g. Bertin & Arnouts 1996; Labbé et al. 2003; Taylor et al. 2009a). We find a systematic shift for both filters towards low surface brightness, the shifts are on average ~ 0.42 mag for a H_{160} mean surface brightness of 23.5 mag arcsec^{-2} or ~ 0.50 mag for a z_{850} mean surface brightness of 24.5 mag arcsec^{-2}.

On the other hand, we find no systematics between the input and recovered aperture colour. Figure 2 shows the result for the $z_{850} - H_{160}$ colour from simulated galaxies. The uncertainties on colour are small i.e. ~ 0.07 mag for a H_{160} mean surface brightness of 23.5 mag arcsec^{-2}. The uncertainty in colour tends to be larger for objects with redder $z_{850} - H_{160}$ colour, solely due to the fact that the z_{850} aperture magnitude has a larger uncertainty for a redder colour.
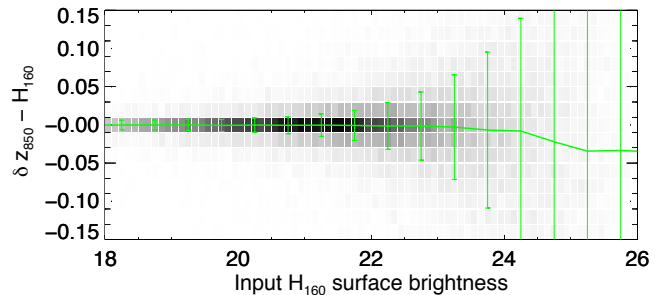

Figure 2. Differences between recovered and input aperture colour $\delta z_{850} - H_{160} = (z_{850} - H_{160})_{out} - (z_{850} - H_{160})_{in}$ as a function of input mean H_{160} surface brightness. The green line indicates the median and 1σ dispersion in different bins (0.5 mag arcsec^{-2} bin width), and the grey-shaded 2D histogram shows the number density distribution of the simulated galaxies.

3.2 Light-weighted structural parameters

3.2.1 Method

We measure the light-weighted structural parameters of the passive galaxies in five HST bands (i_{775}, z_{850}, Y_{105}, J_{125} and H_{160}) using a modified version of GALAPAGOS (Barden et al. 2012). For each object detected by SExtractor, GALAPAGOS generates a postage stamp and measures the local sky level around the object using an elliptical annulus flux growth method. This local sky level is then used by GALFIT (v.3.0.5, Peng et al. 2002), in order to model the galaxy surface brightness profile. We examine different settings of the sky estimation routine in GALAPAGOS to ensure the robustness of the results. Since the ACS and WFC3 images have a different spatial resolution, we modify GALAPAGOS to allow the use of a single detection catalogue (in our case, the H_{160} band) in all bands. The code is further adjusted to use the RMS maps derived from the `ERR` weight maps output by Astrodrizzle.

As shown in Häussler et al. (2007), contamination by neighbouring objects has to be accounted while fitting galaxy surface brightness profiles, especially in regions where the object density is high. To deal with this issue, adjacent sources are identified from the SExtractor segmentation map and are masked out or fitted simultaneously if their light profiles have a non-negligible influence to the central object. We fit a two-dimensional Sérsic profile (Sérsic 1963) to each galaxy, which can be written as

$$I(a) = I_e \exp\left[-b_n\left((\frac{a}{a_e})^{1/n} - 1\right)\right] \qquad (1)$$

where the effective intensity I_e can be described by

$$I_e = \frac{L_{tot}}{2\pi n q a_e^2 \, b_n^{-2n} \, \Gamma(2n)} \qquad (2)$$

where $\Gamma(2n)$ is the complete gamma function.

The Sérsic profile of a galaxy can be characterised by five independent parameters: the total luminosity L_{tot}, the Sérsic index n, the effective semi-major axis a_e, the axis ratio q ($= b/a$, where a and b is the major and minor axis respectively) and the position angle *P.A.*. The parameter b_n is a function of the Sérsic index ($\Gamma(2n) = 2\gamma(2n, b_n)$, where γ is the incomplete gamma function) and can only be solved numerically (Ciotti 1991). All five parameters as well as the centroid (x, y) are left to be free parameters in our fitting process with GALFIT. The constraints of each parameter for GALFIT are set to be: $0.2 < n < 8$, $0.3 < a_e < 500$ (pix), $0 < mag < 40$, $0.0001 < q < 1$, $-180° < P.A. < 180°$. The sky
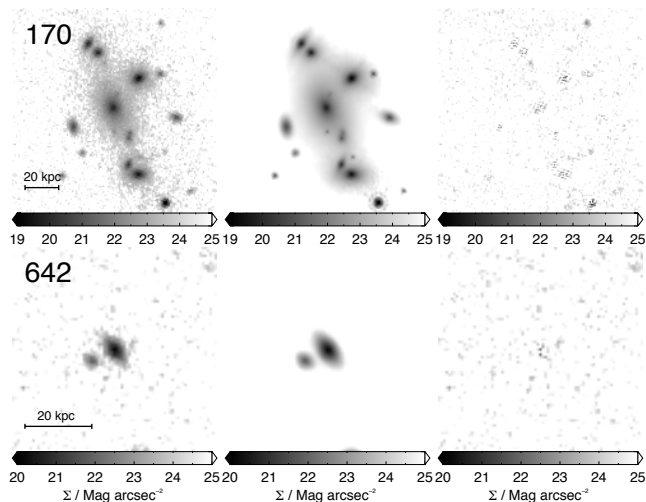
Figure 3. Examples of surface brightness profile fitting of two passive galaxies (ID 170, 642) in cluster XMMUJ2235-2557. From left to right: H_{160} galaxy image cut-out centered on the primary object, GALFIT best-fit models and residuals. The two examples are selected to demonstrate the clustering of sources. Galaxy 170, the BCG of this cluster is located in the central region of the cluster with high object density. Galaxy 642 is located in a more outer region of the cluster, yet is still affected by an extremely close neighbour. Multiple objects are fitted simultaneously as described in Section 3.2.1.

level on the other hand, is fixed to the value determined by GALA-PAGOS.

The Sérsic model is convolved with the PSF constructed from stacking bright unsaturated stars in the images. Note that we have also tried to derive a `TinyTim` PSF composite by adding PSF models using the `TinyTim` code (Krist 1995) into the raw data and drizzling them as science images. Nevertheless, we notice that the `TinyTim` drizzled PSF does not match well the empirical PSF in the outer part: a much stronger outer envelope (as well as diffraction spikes) can be seen in the empirical PSF (see also, Appendix A in Bruce et al. 2012, for a similar description). On the other hand, van der Wel et al. (2012) produced hybrid PSF models by replacing the central pixels of the median-stacked star by the `TinyTim` PSF. We do not employ this correction as we find that the median-stacked star matches the TinyTim PSF reasonably well in the inner part.

The best-fitting light-weighted parameters are listed in Table F1 in Appendix F. Two galaxies (ID 170, 642) and their best-fits are shown in Figure 3 for illustrative purposes. These two objects have been chosen to show the impact of clustering of sources in dense regions. Even in the cluster centre where there are multiple neighbouring objects, GALFIT can do a good job in determining the structural parameters by fitting multiple object simultaneously. Below we discuss the reliability and uncertainties in these light-weighted structural parameters.

3.2.2 *Reliability of the fitted structural parameters*

GALAPAGOS coupled with GALFIT performs well in most cases. However in some exceptions, it is rather tricky to obtain a good-quality fit due to various issues. We are not referring here to the global systematics and uncertainties (which are addressed in the next section), but on stability and quality control of individual fits. We find that using an inadequate number of fitting components for

the neighbouring sources (due to inadequate deblending in the SExtractor catalogue or appearance of extra structures / sources in bluer bands, e.g. z_{850} band, compare to our H_{160} detection catalogue) can lead to significant residuals that adversely affect the fit of the primary object. Similarly, since GALAPAGOS fits sources with a single Sérsic profile by default, GALFIT will likely give unphysical outputs for unresolved sources / stars in the field (with a_e hitting the lower boundary of the constraint $a_e = 0.3$ pix, or Sérsic index hitting the upper boundary $n = 8$) or even not converging in these cases, which again affects the result of the object of primary interest. Moreover, the best-fit output can vary if we use a different treatment for neighbouring sources. We notice that in a few cases the results can be very different depending upon whether neighbouring sources are masked or are fitted simultaneously.

To ensure high reliability, we perform the following checks for each galaxy: 1) We visually inspect the fits as well as the segmentation maps (output by SExtractor) in each band to ensure adjacent sources are well-fitted. Extra Sérsic components are added to poorly fitted neighbouring objects iteratively if necessary. 2) For neighbours for which GALFIT gives ill-constrained results (i.e. hitting the boundaries of the constraints), we replace the Sérsic model with a PSF model and rerun the fit, which often improves the convergence and the quality of the best-fit model. Regarding this, Barden et al. (2012) explained the need of fitting Sérsic profiles to saturated stars instead of PSF model in GALAPAGOS, since the PSF often lacks the dynamic range to capture the diffraction spikes of the bright saturated stars. In our case this is not necessary since there are only a few bright saturated stars in the field, for which we can safely mask their diffraction spikes. 3) We compare the results of masking and simultaneously fitting neighbouring objects. In most cases the two methods give results that are within 1σ. For galaxies with close neighbours (e.g. within $5\ a_e$) we prefer to fit them simultaneously as any inadequate or over-masking can result in problematic fits, judging by examining the residual map output by GALFIT. On the other hand, masking is more suitable when the neighbouring object are not axisymmetric or show certain substructures, which causes the single Sérsic fit to not reach convergence.

3.2.3 *Quantifying the uncertainties in light-weighted structural parameters*

We quantify the systematic uncertainties using the set of 50000 simulated galaxies inserted on the images. In this section we focus on the result of the test; details of the simulations can be found in Appendix A3. Note that the uncertainties quoted here are more likely to represent lower limits to the true uncertainties, as the simulated galaxies are also parametrised with a Sérsic profile.

Figure 4 shows the comparison between the input and recovered magnitudes and structural parameters for the H_{160} band. The magnitudes recovered by Sérsic profile fitting are accurate with almost no systematics and a 1σ dispersion less than 0.25 for objects having mean H_{160} surface brightness brighter than 23.5 mag arcsec^{-2}. The Sérsic index, effective radius and axis ratio measurements are generally robust for objects brighter than a mean H_{160} surface brightness of 23.5 mag arcsec^{-2}. The bias between the recovered and input Sérsic indices is less than 8% and the 1σ dispersion is lower than 30%. Effective radii have a bias less than 4% and a 1σ dispersion lower than 30% for objects brighter than H_{160} surface brightness of 23.5 mag arcsec^{-2}. For objects with high mean surface brightness (i.e. < 19 mag arcsec^{-2}) in our simulated sample, the effective radii are slightly overestimated (~ 2%) and the Sérsic indices are underestimated (~ -4%) by GALFIT. We find
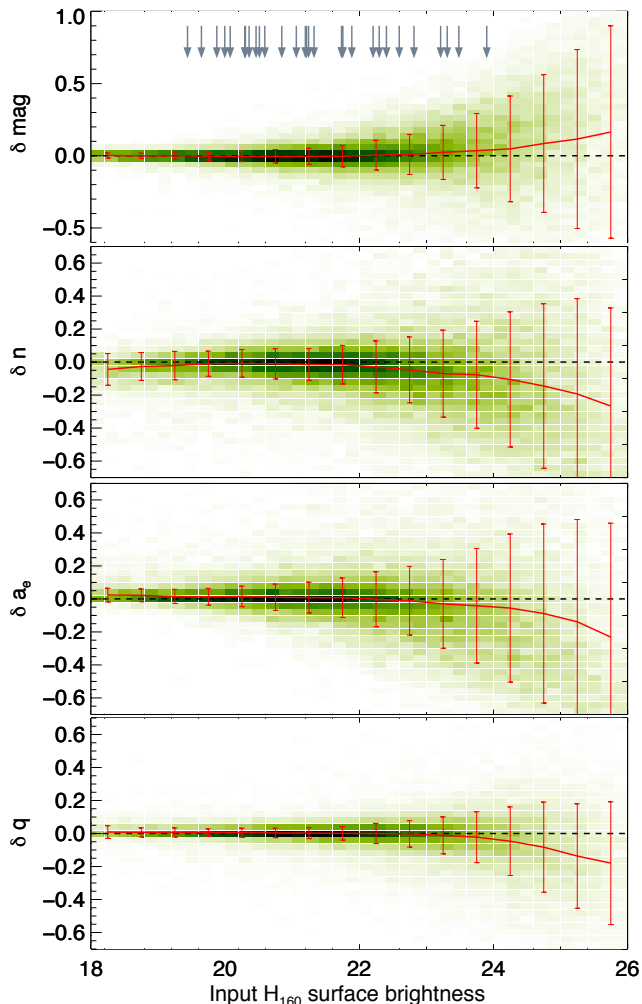
Figure 4. Differences between recovered and input structural parameters by GALFIT in function of input mean H_{160} surface brightness. From top to bottom: magnitude $\delta mag = mag_{out} - mag_{in}$, Sérsic indices $\delta n = (n_{out} - n_{in})/n_{in}$, effective semi-major axes $\delta a_e = (a_{e-out} - a_{e-in})/a_{e-in}$ and axis ratio $\delta q = (q_{out} - q_{in})/q_{in}$. Red line indicates the median and 1σ dispersion in different bins (0.5 mag arcsec^{-2} bin width) and green-shaded 2D histogram shows the number density distribution of the simulated galaxies. The grey arrows indicate the H_{160} surface brightness of the galaxies in our cluster sample.

out that this bias is due to unresolved objects in our simulations. A related discussion can be found in Appendix A3.

We have also performed the same test on a simulated background similar to the actual images (where the main difference is that the simulated background has no issue of neighbour contamination), and find that the uncertainties on the effective radius are on average $\sim 15 - 20\%$ lower compared to those derived from real images.

For each galaxy in the sample, we compute the mean H_{160} surface brightness and add the corresponding dispersion in quadrature to the error output by GALFIT.

3.3   Elliptical aperture photometry and color gradients

In addition to structural parameters, we derive $z_{850} - H_{160}$ colour profiles for the passive sample with PSF-matched elliptical annular

photometry. We first convert the 2D image in both bands into 1D radial surface brightness profiles. Morishita et al. (2015) demonstrated that deriving 1D profiles with elliptical apertures has certain advantages over circular apertures. Profiles derived with concentric circular apertures are biased to be more centrally concentrated. We perform an elliptical annular photometry on the PSF-matched z_{850} and H_{160} images at the galaxy centroid derived from GALFIT. The GALFIT best-fit axis ratios and position angles of individual galaxies (in H_{160} band) are used to derive a set of elliptical apertures for each galaxy.

Due to the proximity of objects in the cluster, it is necessary to take into account (as in 2D fitting) the effect of the neighbouring objects. The neighbouring objects are first removed from the image by subtracting their best Sérsic fit (or PSF fit in some cases) in both bands. While the fit might not be perfect, we find that this extra step can remove the majority of the flux of the neighbouring objects contributing to surface brightness profiles. For some galaxies the colour profiles show substantial change after we apply the correction.

We then measure the colour gradients of individual galaxies by fitting the logarithmic slope of their $z_{850} - H_{160}$ colour profiles along the major axis, which are defined as follows:

$$z_{850} - H_{160} = \nabla_{z_{850} - H_{160}} \times \log(a) + Z.P. \tag{3}$$

At redshift 1.39 this corresponds roughly to the rest-frame $(U - R)$ colour gradient. The depth and angular resolution of our WFC3 data allow us to derive a 1D colour profile accurately to $\sim 3 - 4\ a_e$, hence the colour gradient is fitted in the radial range of PSF half-width-half-maximum (HWHM) $< a < 3.5\ a_e$. We note that the colour gradients of most galaxies, as well as the median colour gradient, do not strongly depend on the adopted fitting radial range. Figure 5 shows the colour profiles and logarithmic gradient fits of four passive galaxies as an example. The colour profiles are in general well-described by logarithmic fits.

3.4   Stellar mass-to-light ratio – colour relation

We estimate the stellar mass-to-light ratios of the cluster galaxies in XMMUJ2235-2557 using an empirical relation between the observed $z_{850} - H_{160}$ colour and stellar mass-to-light ratio (M_*/L). At redshift 1.39, the $z_{850} - H_{160}$ colour (rest-frame $U - R$) straddles the 4000 break. Hence, this colour is sensitive to variations in the properties of the stellar population (i.e. stellar age, dust and metallicity). In addition, the effects of these variations are relatively degenerate on the colour - M_*/L plane (almost parallel to the relation, Bell & de Jong 2001; Bell et al. 2003; Szomoru et al. 2013), which makes this colour a useful proxy for the M_*/L.

We derive the relation using the NEWFIRM medium band survey (NMBS) catalogue, which combines existing ground-based and space-based UV to mid-IR data, and new near-IR medium band NEWFIRM data in the AEGIS and COSMOS fields (Whitaker et al. 2011). The entire catalogue comprises photometries in 37 (20) bands, high accuracy photometric redshifts derived with EAZY (Brammer, van Dokkum & Coppi 2008) and spectroscopic redshifts for a subset of the sample in COSMOS (AEGIS). Stellar masses and dust reddening estimates are also included in the catalogue, and are estimated by SED fitting using FAST (Kriek et al. 2009).

To derive the M_*/L-colour relation, we use the stellar masses from the NMBS catalogue in COSMOS estimated with stellar population models of Bruzual & Charlot (2003), an exponentially declining SFHs, and computed with a Chabrier (2003) IMF. We do not use the sample in AEGIS as it contains photometries with fewer
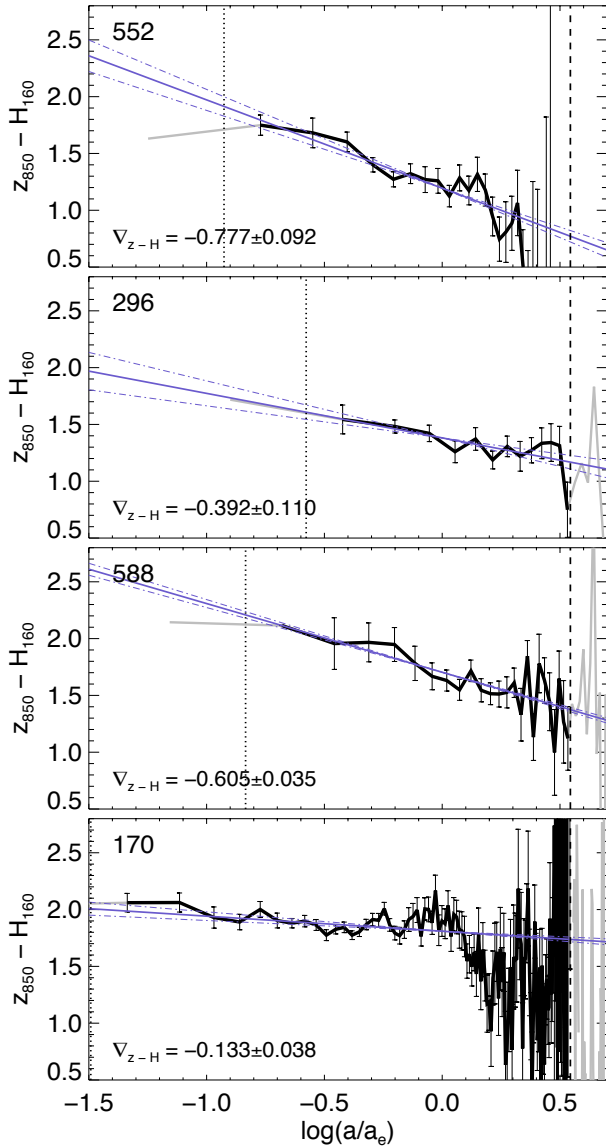
Figure 5. Examples of colour profile fitting of four passive galaxies in the cluster XMMUJ2235-2557. From top to bottom: colour profiles for galaxies ID 552, 296, 588 and 170 (with $\log(M_*/M_\odot) = 10.46, 10.54, 10.81, 11.81$) along the logarithmic major axis ($\log(a/a_e)$). The grey line in each panel is the elliptical-averaged $z_{850} - H_{160}$ colour profile. Regions that are fitted (PSF HWHM $< a < 3.5\ a_e$) are over-plotted in black. The vertical black dotted and dashed line show the minimum (PSF HWHM) and maximum radial distance for fitting (3.5 a_e). The error bars show the error on the mean of the $z_{850} - H_{160}$ colour at each distance. The blue solid line is the best logarithmic gradient fits, and the blue dotted-dashed lines are the $\pm 1\sigma$ error of the slope.

Figure 6. Relation between stellar mass-to-light ratio and z-H colour at redshift $\sim$1.39 using the public NMBS catalogue. Gray points are 718 galaxies from the NMBS catalogue that satisfy the selection criteria. Black line is the best-fit linear relation. Bottom panel shows the residuals of the relation $\delta log(M_*/L_{H_{160}})$ = data - linear fit in colour bins of 0.1.

bands. We derive the relation in the observer frame and compute the observed $z_{850} - H_{160}$ colour for all NMBS galaxies. Note that we do not adopt the typical approach to interpolate the cluster data to obtain a rest-frame colour (e.g. with InterRest, Taylor et al. 2009b) due to limited availability of bands, which would likely lead to degeneracy in choices of templates. Firstly, we rerun EAZY for all NMBS galaxies to obtain the best-fit SED template, these SEDs are then integrated with the ACS z_{850} and WFC3 H_{160} filter response for the $z_{850} - H_{160}$ colour. Similarly we obtain the luminos-
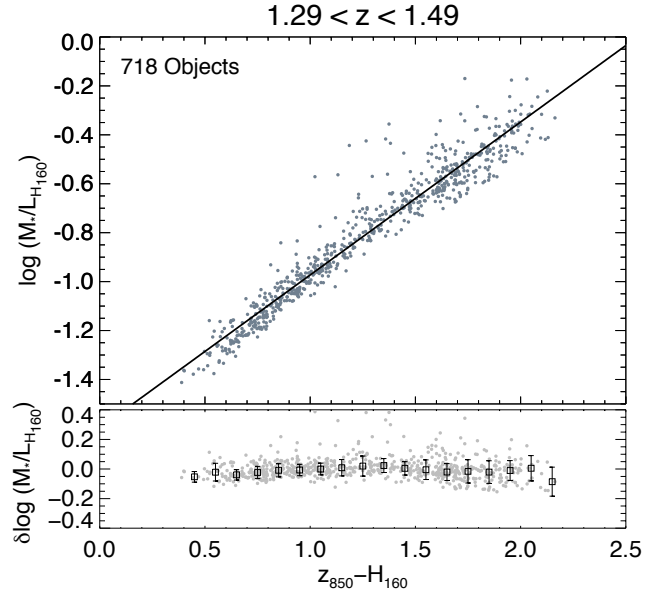
ity $L_{H_{160}}$ of each galaxy in the observed H_{160} band, from which we calculate the stellar mass-to-light ratio $M_*/L_{H_{160}}$. We select NMBS galaxies within a redshift window of 0.1 of the cluster redshift, i.e. $1.29 < z < 1.49$, and apply the magnitude cut ($H_{160} < 22.5$) and a chi-square cut ($\chi^2 < 2.0$, from template fitting in EAZY) to better match the cluster sample.

A total of 718 objects are selected by this criterion. A redshift correction is applied to these 718 galaxies to redshift their spectra to the cluster redshift (i.e. similar to k-correction in observer frame). We then measure their $z_{850} - H_{160}$ colour and $L_{H_{160}}$ in the observer frame. This redshift correction is effective in reducing the scatter of the relation, indicating that some of (but not all) the scatter is simply due to difference in redshifts. Figure 6 shows the fitted relation between $\log(M_*/L_{H_{160}})$ and $z_{850} - H_{160}$ colour. The black line is the best-fit linear relation with:

$$\log((M_*/L_{H_{160}})/(M_\odot/L_\odot)) = 0.625\,(z_{850} - H_{160}) - 1.598 \qquad (4)$$

The relation is well-defined within a colour range of $0.4 < z_{850} - H_{160} < 2.2$ (hence we choose the same range for our simulated galaxies, see Appendix A). The global scatter of the fit is ~ 0.06 dex. In the lower panel of Figure 6 we plot the residuals of the fit in colour bins of 0.1. The uncertainty in $\log(M_*/L)$ is generally < 0.1 in each bin and the bias is negligible. The remaining scatter results from redshift uncertainties and stellar population variations (age, dust and metallicity), as their effects are not exactly parallel to the relation. Note that this can lead to small systematics in measuring mass-to-light ratios and the mass-to-light ratio gradients. For example in metal-rich or old regions the mass-to-light ratio will likely be systematically slightly underestimated, and overestimated in metal-poor or young regions (Szomoru et al. 2013).

3.5   Integrated stellar masses

We estimate the integrated stellar masses (M_*) of the cluster galaxies using our M_*/L-colour relation, the $z_{850} - H_{160}$ aperture colours and the total luminosity $L_{H_{160}}$ from best-fit Sérsic models. The uncertainties in stellar mass comprise photometric uncertainties in the colour and H_{160} luminosity, as well as the scatter in the derived colour - M_*/L relation. The typical uncertainty of the masses is ~ 0.1 dex, comparable to the uncertainties obtained from SED fitting.

Previous literature computed SED mass with multi-band `MAG_AUTO` photometry obtained with SExtractor (e.g. for this cluster, Strazzullo et al. 2010; Delaye et al. 2014). Nevertheless, as we have shown in Section 3.1.2, it is known that `MAG_AUTO` can be systematically biased, due to the assumption in SExtractor that the sky background comprises only random noise without source confusion (Brown et al. 2007). Hence, more recent studies use the total luminosity from best-fit Sérsic models to correct the masses to account for the missing flux in `MAG_AUTO` (Bernardi et al. 2013; Bezanson et al. 2013). In our case, we have demonstrated that total luminosity from the best-fit Sérsic models can recover input galaxy magnitudes to a high accuracy. Hence, we scale our masses with the total luminosity $L_{H_{160}}$ from best-fit Sérsic models rather than H_{160} `MAG_AUTO` magnitudes. We also compute masses with H_{160} `MAG_AUTO`; the difference between the two is small for our sample, with $\langle M_{*,MAG\_AUTO} - M_{*,Sersic} \rangle = -0.039$ dex.

For this particular cluster, Delaye et al. (2014) estimated the galaxies masses through SED fitting with four bands (HST/ACS i_{775}, z_{850}, HAWK-I J, Ks), which also gave an uncertainty of ~ 0.1 dex in mass. The masses derived with our method are consistent with the SED masses in Delaye et al. (2014) within the uncertainties. A comparison of masses estimated using M_*/L-colour relation with masses computed using SED fitting can be found in Appendix C. The uncertainty of the absolute stellar masses is of course larger (as in the case of SED fitting), depending on the details of NMBS SED fitting and e.g. choice of IMF.

3.6   Resolved stellar mass surface density maps

We further exploit the M_*/L-colour relation to derive stellar mass surface density maps. This allows us to study the mass distribution within each galaxy, at the same time eliminating the effect of internal colour gradient which influences the light-weighted size measurements. Below we describe the main steps involved in deriving stellar mass surface density maps with the M_*/L-colour relation.

3.6.1   PSF matching

We first match the resolution of the ACS z_{850} image ($\sim 0.1''$) to the WFC3 H_{160} image ($\sim 0.18''$). PSF matching is critical in this kind of study as the measured colour has to come from the same physical projected region. We stack the unsaturated stars for each band to obtain characteristic PSFs, then generate a kernel that matches the z_{850} to H_{160} PSF using the `psfmatch` task in IRAF. The difference between the resultant z_{850} PSF and the H_{160} PSF is less than 2.5%. Details of the PSF matching can be found in Appendix B. We then apply the kernel to the ACS z_{850} image. The PSF matched z_{850} image is resampled to the same grid as the H_{160} image using the software SWarp (Bertin et al. 2002). We then generate postage stamps of each galaxy in both H_{160} and PSF matched z_{850} images for deriving resolved stellar mass surface density maps.

3.6.2   From colour to stellar mass surface density

The next step is to convert the $z_{850} - H_{160}$ colour information into mass-to-light ratios with the M_*/L-colour relation described in Section 3.4. Nevertheless, a direct pixel-to-pixel conversion is not possible for our data. The conversion requires a certain minimum signal-to-noise (S/N) level because: a) significant biases or massive uncertainties may arise if colours are not well measured. b) our relation is only calibrated within the colour range of $0.4 < z_{850} - H_{160} < 2.2$. Any low S/N colour that falls outside the calibrated range could convert to an unphysical M_*/L.

Therefore, we adopt the Voronoi binning algorithm as described by Cappellari & Copin (2003), grouping pixels to a target S/N level of 10 per bin. For each galaxy, we run the Voronoi binning algorithm on the sky-subtracted PSF-matched z_{850} band postage stamps as a reference, as it has a lower S/N compared to the H_{160} image. The same binning scheme is then applied to the sky-subtracted H_{160} image. The subtracted sky levels are determined by GALAPAGOS. The two images are then converted into magnitudes. Binned $z_{850} - H_{160}$ colour maps are obtained by subtracting the two. We then construct a binned M_*/L map by converting the colour in each bin to a mass-to-light ratio with the derived colour - M_*/L relation.

An extrapolation scheme is implemented to determine the M_*/L in regions or bins with insufficient S/N, for example in the galaxy outskirts and the sky regions. We first run an annular average to derive a 1-dimensional S/N profile in z_{850} for individual galaxies using the light-weighted galaxy centroid, axis ratio and position angle determined in Section 3.2. For the area outside the elliptical radius that has a S/N less than half of our target S/N (i.e. S/N ~ 5), we fix the M_*/L to the annular median of M_*/L bins at the last radius with sufficient S/N. We find that this extrapolation is crucial for the following structural analysis as the sky noise is preserved (see the discussion in Appendix A4).

We construct resolved stellar mass surface density maps (hereafter referred to as mass maps) by directly combining the extrapolated M_*/L map and the original (i.e. unbinned) H_{160} images. Figure 7 illustrates the procedure of deriving mass maps from the z_{850} and H_{160} images. Using the original H_{160} image instead of the binned one allows us to preserve the WFC3 spatial resolution in the mass maps. Note that in theory combining a binned (i.e. spatially discrete) M_*/L map with a smooth luminosity image would result in a discrete mass profile in low S/N region, in order words, induce an "discretization effect" in the mass maps. This effect is more severe in low S/N regions, i.e. the galaxy outskirts where the bins are larger (hence less smooth). For bright galaxies, since there are more bins with sufficient S/N and the dynamical range of the light distribution (surface brightness gradient) is much larger than the M_*/L gradient, this appears to have minimal effect and does not largely affect our result. For fainter galaxies this issue is non-negligible. To tackle this, for each galaxy we perform the above binning procedure 10 times, each with a slightly different set of initial Voronoi nodes. This ends up with a set of M_*/L maps which are then median-stacked to create the final mass map. This extra step alleviates the discretization effect.

3.7   Mass-weighted structural parameters

3.7.1   Method

We measure mass-weighted structural parameters from the resolved stellar mass surface density maps. We follow a similar procedure as with the light-weighted structural parameters, using

GALFIT to model the mass profiles with two-dimensional Sérsic profiles. All five parameters of the Sérsic profile ($M_{*,tot}$, n, a_e, q and *P.A.*) and the centroid (x,y) are left to be free parameters in the fit. We use the same GALFIT constraints as for the light-weighted structural parameters, except for allowing a larger range for the Sérsic indices: $0.2 < n < 15.0$. This is because the mass profiles are expected to be more centrally peaked compared to light profiles (Szomoru et al. 2013). As the H_{160} images are background subtracted before being converted into mass maps, the sky level (i.e. the mass level) is fixed to zero in the fitting process. The best-fitting mass-weighted parameters are given in Appendix F1.

3.7.2 *Quantifying the uncertainties in mass-weighted parameters*

We further assess the accuracy of our mass conversion procedures as well as the reliability of the mass-weighted structural parameter measurements. The details of the test are described in Appendix A4. Similar to the uncertainty in the light-weighted parameters, the uncertainties quoted here are more likely to represent lower limits to the true uncertainties.

Figure 8 shows the difference between input and recovered mass-structural parameters as a function of H_{160} surface brightness. The Sérsic index, effective radius and axis ratio measurements are generally robust for objects brighter than H_{160} surface brightness of 23.5 mag arcsec^{-2}. This is important, as it demonstrates that our mass conversion procedure does not significantly bias the result. The bias between the recovered and input Sérsic indices is less than 7% and the 1σ dispersion is lower than 40%, and effective radii have a bias less than 10% and a 1σ dispersion lower than 40%. Among the three parameters, the axis ratio can be recovered most accurately. Compared with the light uncertainties (Figure 4), the mass uncertainties in all parameters are ~ 2 times higher. Similar to the light-weighted parameters, for each galaxies we add the corresponding dispersion in quadrature to the error output by GALFIT.

We find that for a couple of objects the fits do not converge, or have resultant sizes smaller than the PSF size. To avoid biases and wrong conclusions we remove these objects that are not well-fitted from the mass parameter sample. 6 objects (out of 36) are discarded, among them one object is spectroscopically confirmed. Three of them initially have small light-weighted sizes and their fitted mass-weighted sizes become smaller than half of the PSF HWHM, which are unreliable (see the discussion in Appendix A3). Most of them are low mass galaxies (i.e. $\log(M_*/M_\odot) < 10.5$).

3.7.3 *Deviation of mass-weighted parameters - 1D vs. 2D*

Szomoru et al. (2013) derived 1D mass profiles from 1D radial surface brightness profiles and measured mass-weighted structural parameters. In theory, fitting in 1D and in 2D should give identical results, as statistically fitting an averaged smaller group of points and fitting all the points without averaging are equivalent (see, Peng 2015, for a detailed discussion.). Nevertheless, in practice deriving maps and fitting in 2D have certain advantages: a) It does not rely heavily on the Sérsic profile fitting in light. Deriving elliptical averaged profiles require a predetermined axis ratio and position angles, which, in our case, come from the light Sérsic profile fitting. This will of course fold in the uncertainties of these two parameters into the 1D profiles, which complicates the propagation of uncertainties in the mass-weighted parameters. b) In the cluster region, the object density is high and many galaxies have very close neighbours.

Hence it will be more appropriate to fit all the sources simultaneously to take into account the contribution from the neighbouring objects, rather than deriving 1D profile without deblending the neighbouring contamination. A possible way to solve this is to first subtract the best-fit 2D models of the neighbours from the 2D images before generating the 1D profiles, but of course this depends strongly on how well the neighbours can be subtracted, and still suffer from a).

4 LOCAL COMPARISON SAMPLE

In order to study the evolution of mass-weighted sizes over redshift, we compare our cluster sample at $z \sim 1.39$ to a local sample of passive galaxies from the Spheroids Panchromatic Investigation in Different Environmental Regions (SPIDER) survey (La Barbera et al. 2010b). The publicly available SPIDER sample includes 39993 passive galaxies selected from SDSS Data Release 6 (DR6), among them 5080 are in the near-infrared UKIRT Infrared Deep Sky Survey-Large Area Survey Data release (UKIDSS-LAS DR4) in the redshift range of 0.05 to 0.095. La Barbera et al. (2010b) derived structural parameters in all available bands ($grizYJHK$) with single Sérsic fitting with 2DPhot (La Barbera et al. 2008).

We use the structural parameters in g-band and r-band from the publicly available multiband structural catalogue from La Barbera et al. (2010b) to derive mass-weighted structural parameters. For the galaxy selection, we follow similar criteria as La Barbera et al. (2010b): we apply a magnitude cut at the 95% completeness magnitude ($M_r \leqslant -20.55$), a χ^2 cut from the Sérsic fit for both g-band and r-band ($\chi^2 < 2.0$), and a seeing cut at $\leqslant 1.5$". This results in a sample of 4050 objects. We compute integrated masses for the sample as in Section 3.5 with aperture $g - r$ colour. The colours are obtained from direct numeral integration of the g-band and r-band Sérsic profiles to 5 kpc instead of using GALFIT total magnitudes. Extending the integration limit to larger radius (e.g. 10 kpc) does not change largely the derived masses. With the $g - r$ colour we derive and select red-sequence galaxies within 2σ following the same method discussed in Section 3.1; we end up with a sample of 3634 objects (hereafter the SPIDER sample). On top of that we use the group catalogue from La Barbera et al. (2010c) to select a subsample of galaxies residing in high density environments. Applying a halo mass cut to the SPIDER sample of $\log(M_{200}/M_\odot) \geqslant 14$, we end up with a subsample of 627 objects (hereafter the SPIDER cluster sample), which we will use as the main comparison sample for our high-redshift cluster galaxies.

2D Sérsic model images in g-band and r-band are then generated with fitted parameters from the structural catalogue. Given the large number and relatively low object density of local galaxies compared to our high redshift cluster sample, using fitted parameters from the structural catalogue is statistically reliable and issues mentioned in Section 3.7.3 do not contribute substantially here. We construct mass maps for individual galaxies using the procedure described in Section 3.6.2 without Voronoi binning and stacking. A M_*/L-colour relation is again derived from the NMBS sample as in Section 3.4, but in g-band and r-band at $0 < z < 0.27$, a window of 0.2 in redshift around the median redshift of the SPIDER sample. A total of 1315 NMBS objects are selected. The mass maps are then fitted with GALFIT to obtain mass-weighted structural parameters.
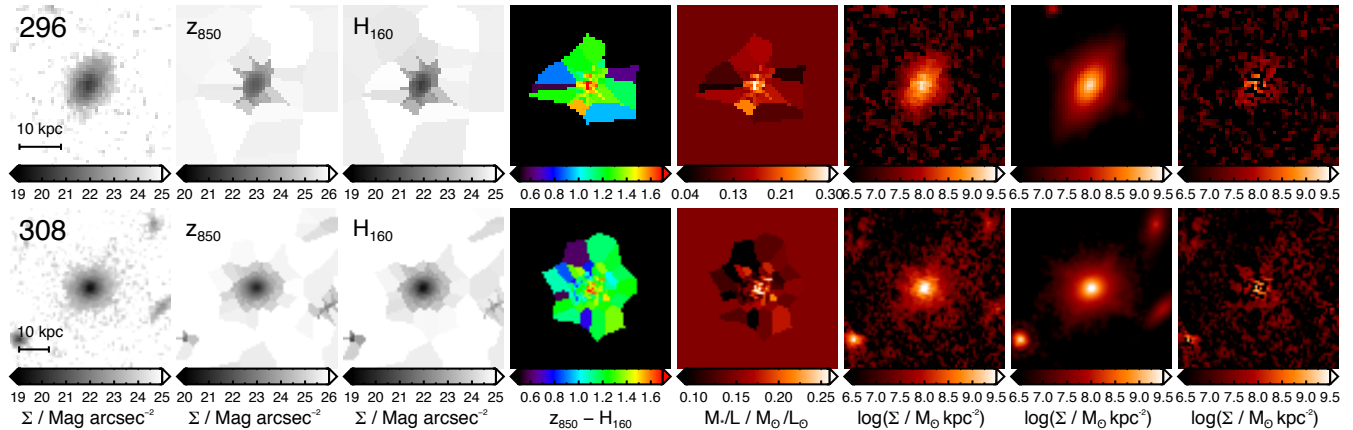
Figure 7. Examples of mass map derivation and fitting of two passive galaxies (ID 296, 308) in cluster XMMUJ2235-2557. From left to right: H_{160} galaxy image cut-outs centred on the primary object, Voronoi-binned z_{850} images, Voronoi-binned H_{160} images, $z_{850} - H_{160}$ colour maps, M_*/L, the surface mass density maps Σ_{mass}, the GALFIT best-fit models and residuals in mass. Bins that are extrapolated are masked out (shown in black) in the colour maps. The procedure is described in detail in Section 3.6.2.

5   RESULTS

5.1   Wavelength dependence of light-weighted galaxy sizes

The measured size of a galaxy depends on the observed wavelength, as different stellar populations are being traced at different wavelength (e.g., the "morphological k-correction", Papovich et al. 2003). With our multi-band measurements of light-weighted structural parameters of the cluster passive galaxies, we first investigate the wavelength dependence of galaxy sizes at this redshift. This wavelength dependence of sizes (or the size-wavelength relation) has been quantified for local passive galaxies in a number of studies (e.g. Barden et al. 2005; Hyde & Bernardi 2009; La Barbera et al. 2010b; Kelvin et al. 2012; Vulcani et al. 2014; Kennedy et al. 2015). The dependence shown by the above mentioned studies is quite strong, in the sense that galaxy sizes can decrease up to $\sim 38\%$ from g through K band in the GAMA sample (Kelvin et al. 2012), or $\sim 32\%$ across the same range in SPIDER (La Barbera et al. 2010b). Nevertheless, different authors disagree on the extent of the reduction in sizes in various datasets. For example, in a recent study Lange et al. (2015) revisited the GAMA sample with deeper NIR imaging data and found a smaller size decrease, $\sim 13\%$ from g to K_s band.

At higher redshift, study of wavelength dependence of sizes is scarce in clusters. The star formation history and age gradient may contribute significantly to the size-wavelength dependence, for example the inside-out growth scenario suggests that younger stellar population are more widespread compared to the older population in the core of passive galaxies. Various authors have shown that measured sizes in the observed optical and NIR (i.e. rest-frame UV vs rest-frame optical for high-redshift galaxies) show a difference of $\sim$20–25% (e.g. Trujillo et al. 2007; Cassata et al. 2010; Damjanov et al. 2011; Delaye et al. 2014), although some find no difference (Morishita, Ichikawa & Kajisawa 2014). The comparisons are usually done with only two bands, hence it is unclear whether this dependence can change with redshift. Recent works from CANDELS studied the wavelength dependence of sizes for 122 early-type galaxies (ETG) in the COSMOS field in three HST bands (F125W, F140W and F160W) at redshift $0 < z < 2$, and found an average gradient of $d\log(a_e)/d\log(\lambda) = -0.25$ independent of mass and redshift (van der Wel et al. 2014).

Figure 9 shows the change in size with rest-frame wavelength for our sample. Here we use the light-weighted effective semi-major axis a_e from GALFIT, as the galaxy size. We assume every galaxy in the sample is at the cluster redshift. We select 28 galaxies (out of 36) with no problematic fits in any of the five bands. The fraction of problematic fits is larger in i_{775} and Y_{105} due to shorter exposure time and lower throughput of the filter, which result in lower S/N. To facilitate comparison with the literature, the sizes in figure 9 are normalised with the median H_{160} sizes of our sample, which is approximately equal to the rest-frame r-band size. We plot the best-fitting relation for local spheroids by Kelvin et al. (2012) and the SPIDER cluster sample, normalised in the same way, for comparison. We see a smooth variation of sizes decreasing from i_{775} to H_{160} bands (rest-frame u to r). The reduction in the median size (from i_{775} to H_{160}) is $\sim 20\%$, which is consistent with the expected decrease across this wavelength range ($\sim 19\%$) following the relation of Kelvin et al. (2012) and the SPIDER cluster sample (La Barbera et al. 2010b). The average size gradient of our sample from the best-fit power law is $d\log(a_e)/d\log(\lambda) = -0.31 \pm 0.27$.

We also attempt to divide the sample in mass bins as in Lange et al. (2015) to investigate the size change with wavelength for different masses. Lange et al. (2015) showed that the size reduction decreases from $\sim 13\%$ for local passive galaxies with $\log(M_*/M_\odot) = 10.0$ to $\sim 11\%$ for those with $\log(M_*/M_\odot) = 11.0$. On the other hand, van der Wel et al. (2014) reported no discernible trends with mass in CANDELS. We split the sample in half at the median mass ($\log(M_*/M_\odot) \leqslant 10.6$ and $\log(M_*/M_\odot) > 10.6$, 14 objects per bin), plotted in grey and slate grey in Figure 9. A steeper dependence can be seen for the high mass bins ($d\log(a_e)/d\log(\lambda) = -0.57 \pm 0.28$) compared to the whole sample, while the low mass bins ($d\log(a_e)/d\log(\lambda) = -0.29 \pm 0.34$) have the same if not slightly shallower wavelength dependence within the uncertainties. This is the opposite to the finding of Lange et al. (2015). Nevertheless, the size gradients of the two bins are within 1σ, a larger sample is needed to confirm the mass dependence.
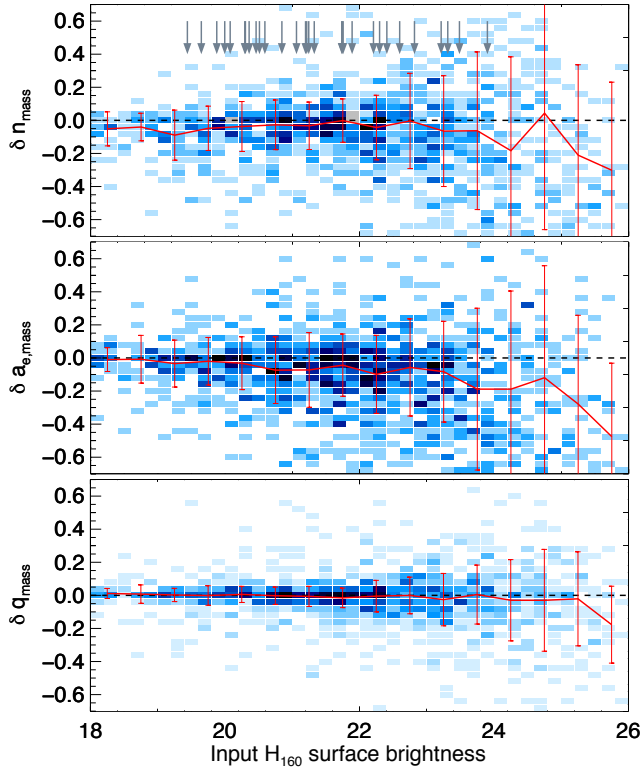
Figure 8. Differences between recovered and input mass-weighted structural parameters by GALFIT as a function of input H_{160} surface brightness. Similar to Figure 4, but for mass-weighted structural parameters. From top to bottom: Sérsic indices $\delta n = (n_{out} - n_{in})/n_{in}$, effective semi-major axes $\delta a_e = (a_{e-out} - a_{e-in})/a_{e-in}$ and axis ratio $\delta q = (q_{out} - q_{in})/q_{in}$. Red line indicates the median and 1σ dispersion in different bins (0.5 mag arcsec^{-2} bin width) and blue-shaded 2D histogram shows the number density distribution of the simulated galaxies. The grey arrows indicate the H_{160} surface brightness of the galaxies in our cluster sample.

5.2 Stellar mass – light-weighted size relation

In this section we show the stellar mass – H_{160} light-weighted size relation of the cluster XMMUJ2235-2557. The mass-size relation of this cluster (in z_{850} band, rest-frame UV) has been studied in previous literature (Strazzullo et al. 2010; Delaye et al. 2014).

In the top panel of Figure 10 we plot the mass – size relation for the passive population in the cluster selected from red sequence fitting. Circled objects are spectroscopically confirmed cluster members (Grützbauch et al. 2012). The size we use from this point onwards is the circularised effective radius (R_{e-circ}), defined as:

$$R_{e-circ} = a_e \times \sqrt{q} \tag{5}$$

where a_e is the elliptical semi-major radius and $q = b/a$ is the axis ratio from the best-fit GALFIT Sérsic profile.

The integrated stellar masses are derived from the M_*/L - colour relation and are scaled with the total GALFIT Sérsic magnitude (see Section 3.5 for details). We also plot the local mass-size relation for the SDSS passive sample by Bernardi et al. (2012) for comparison. We note that although this relation was derived for galaxies regardless of their local density, a number of studies have established that there is no obvious environmental dependence on passive galaxy sizes in the local universe (Guo et al. 2009; Weinmann et al. 2009; Taylor et al. 2010; Huertas-Company et al. 2013;
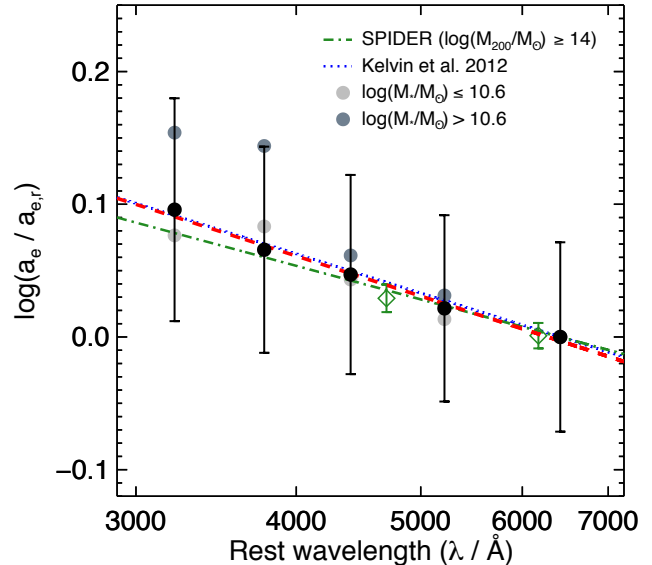

Figure 9. Size-wavelength relation of the passive galaxies in the cluster XMMUJ2235-2557. Black circles show the median sizes of the sample in each band positioned at the rest-frame pivot wavelength, normalised with the median H_{160}-band sizes (approximately rest-frame r-band, $a_{e,r}$). Error bars show the uncertainty of the median in each band, estimated as $1.253\sigma/\sqrt{N}$, where σ is the standard deviation and N is the size of the sample. The best-fit power law to the sizes in our sample is shown as a red dashed line. The green dot-dashed line is the best-fit relation for the SPIDER cluster sample (from g-band to K_s-band), while the green diamonds are the median size of the sample in g-band and r-band, normalised in the same way. The blue dotted line is the best-fit relation for local galaxies from Kelvin et al. (2012). Grey and slate grey are the median sizes for two mass bins ($\log(M_*/M_\odot) \leqslant 10.6$ and $\log(M_*/M_\odot) > 10.6$) respectively.

Cappellari 2013). We pick the single Sérsic fit relation in Bernardi et al. (2012) for consistency, which is shown to have slightly larger sizes than the two-component exponential + Sérsic fit relation.

Hyde & Bernardi (2009) first demonstrated that the mass-size relation of passive galaxies shows curvature and Bernardi et al. (2012) fitted the curvature with a second order polynomial; their best-fit values were consistent with Simard et al. (2011). As we have shown in the last section, size shows wavelength-dependence, hence care has to be taken to ensure the sizes being compared are at around the same rest-frame wavelength. The Bernardi et al. (2012) local relation is based on the Sloan r-band, while our sizes are measured in the H_{160} band at a redshift of 1.39, which roughly corresponds to the same rest-frame band. As a result, no size-correction is required as the correction to r-band is negligible.

The H_{160} band sizes of the passive galaxies in this cluster are on average $\sim 40\%$ smaller than expected from the local relation by Bernardi et al. (2012) with $\langle\log(R_{e-circ}/R_{Bernardi})\rangle = -0.21$ ($\sim 45\%$ smaller for the spectroscopic confirmed members, $\langle\log(R_{e-circ}/R_{Bernardi})\rangle = -0.25$). There are also galaxies whose sizes are $\sim 70\%$ smaller than those of their local counterparts ($\log(R_{e-circ}/R_{Bernardi}) = -0.56$). As one can see from Figure 10, the most massive object in the cluster is the BCG, which also has the largest size (~ 24 kpc) and lies on the local relation. This is consistent with findings from Stott et al. (2010, 2011), who showed that as a population, BCGs have had very little evolution in mass or size since $z \sim 1$. Tiret et al. (2011) suggested that major mergers at $z \geqslant 1.5$ are required to explain the mass growth of these extremely

massive passive galaxies. Hence below we exclude the BCG when fitting the mass-size relation. We fit the mass-size relation with a Bayesian inference approach using Markov Chain Monte Carlo (MCMC; Kelly 2007) with the following linear regression:

$$\log(R_{e-circ}/kpc) = \alpha + \beta \ (\log(M_*/M_\odot) - 10.5) + N(0,\varepsilon) \quad (6)$$

where $N(0,\varepsilon)$ is the normal distribution with mean 0 and dispersion ε. The ε represents the intrinsic random scatter of the regression.

The best-fit parameters (the intercept α, slope β and the scatter ε) for both the entire red-sequence selected sample (A) and the spectroscopically confirmed members only (B) are summarised in Table 2. For mass completeness and comparison to previous literature, we also fit only the massive objects with $\log(M_*/M_\odot) \geqslant 10.5$, the limiting mass adopted in Delaye et al. (2014) (C & D). We notice that the slope of the relation can change by more than 1σ depending on the considered mass range. We also fit the slope using the elliptical semi-major axis a_e instead of R_{e-circ}), which gives us a significantly flatter slope ($\beta = 0.35 \pm 0.15$).

Our measured slope is consistent at the 1σ level with the results of Delaye et al. (2014), who studied the mass-size relation using seven clusters at $0.89 < z < 1.2$ in the rest-frame B-band (i.e. $\beta = 0.49 \pm 0.08$ for $\log(M_*/M_\odot) > 10.5$). Papovich et al. (2012) measured the sizes of passive galaxies in a cluster at $z = 1.62$ and found that ETG with masses $\log(M_*/M_\odot) > 10.48$ have $\langle R_{e-circ} \rangle = 2.0$ kpc with the interquartile percentile range (IQR) of $1.2 - 3.3$ kpc. Sizes in XMMUJ2235-2557 are on average 40% larger ($\langle R_{e-circ} \rangle = 2.80$ kpc, IQR $= 1.45 - 4.38$ kpc).

While the fits are consistent with each other on a 1σ level, we notice that the relation in Delaye et al. (2014) for this cluster is flatter ($\beta = 0.2 \pm 0.3$) compared to both our full sample fit (A) and massive sample fit (C). This difference could be due to a combination of a) their mass-size relation is computed in the z-band while ours is in H_{160}, b) the two red sequence samples are selected differently and c) the masses computed here are scaled with total GAL-FIT Sérsic magnitude instead of `MAG_AUTO` (the relation is slightly flatter: $\beta = 0.38 \pm 0.27$ instead of 0.43 if we use the masses scaled with `MAG_AUTO`).

A caveat of the above comparison is that we have not considered the effect of progenitor bias (van Dokkum & Franx 2001). Correcting the progenitor bias (in age and morphology) has been shown to reduce the magnitude of the observed size evolution (Saglia et al. 2010; Valentinuzzi et al. 2010a; Beifiori et al. 2014). Recently, Jørgensen et al. (2014) corrected the progenitor bias by removing galaxies that are too young in the Coma cluster to be the descendants of a cluster at $z = 1.27$ and found no size evolution with redshift.

5.3   Colour gradients in the passive cluster galaxies

In Figure 11 we show the $1''$ aperture colour, colour gradients $\nabla_{z_{850}-H_{160}}$ and $\log(M_*/L)$ gradients $\nabla_{\log(M/L)}$ of the passive sample as a function of stellar mass. The $\log(M_*/L)$ gradients are derived from fitting 1D M_*/L profiles, which are derived from 1D colour profiles using the M_*/L-colour relation. Note that since the M_*/L-colour relation is essentially a one-to-one mapping, measuring the colour gradient is qualitatively equivalent to measuring the $\log(M_*/L)$ gradient.

More massive galaxies appear to have a redder $z_{850} - H_{160}$ colour, as also shown by Strazzullo et al. (2010) with HST/NICMOS data. Redder colour implies a higher median M_*/L from the M_*/L-colour relation. The passive sample has a range of
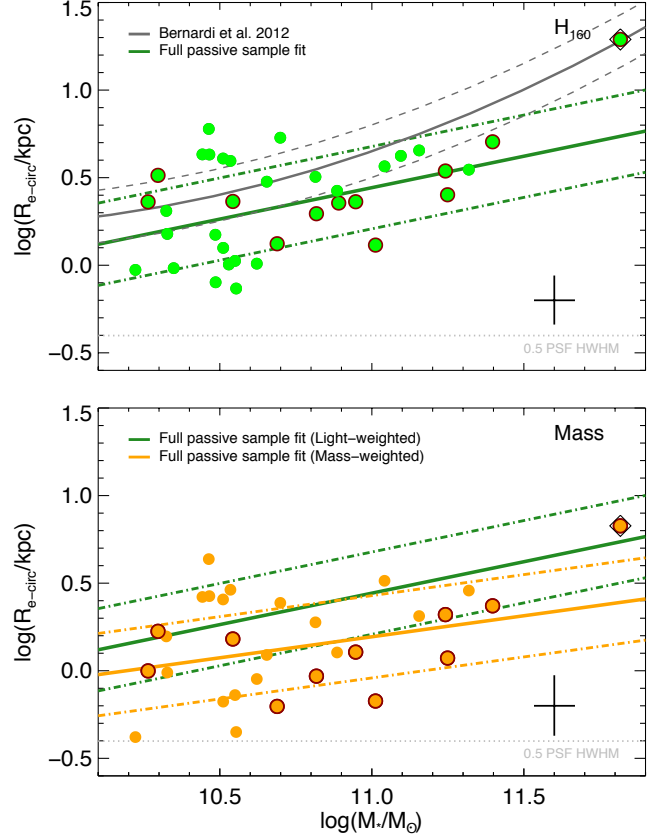

Figure 10. Stellar mass – size relations of the passive galaxies in XMMUJ2235-2557. Top: with light-weighted sizes. Green dots show the sample selected with the passive criteria described in Section 3.1. Spectroscopically confirmed objects are circled with dark red. The green line is a linear fit to the full passive sample (Case A), while the dot-dashed lines represent $\pm 1\sigma$. The dark grey line corresponds to the local r-band mass-size relation from Bernardi et al. (2012). Bottom: with mass-weighted sizes. Individual objects are shown in orange. The orange solid line corresponds to the full sample fit (Case A) for the mass– mass-weighted size relation, and the orange dot-dashed lines represent $\pm 1\sigma$. The green line is the same linear fit in the top panel for comparison. The BCG is indicated with the black diamond. The cross shows the typical uncertainty of the sizes and the median uncertainty of the integrated mass in our sample.

Table 2. Best-fit parameters of the Stellar mass – size relations

Stellar mass – light-weighted size relation

Case	Mass range	$\alpha \pm \Delta\alpha$	$\beta \pm \Delta\beta$	ε
A	$10.0 \leqslant M_* \leqslant 11.5$	0.263 ± 1.441	0.359 ± 0.135	0.235
B	$10.0 \leqslant M_* \leqslant 11.5$ (spec)	0.329 ± 2.096	0.138 ± 0.192	0.195
C	$10.5 \leqslant M_* \leqslant 11.5$	0.114 ± 1.876	0.576 ± 0.173	0.195
D	$10.5 \leqslant M_* \leqslant 11.5$ (spec)	0.149 ± 2.935	0.447 ± 0.268	0.175

Stellar mass – mass-weighted size relation

Case	Mass range	$\alpha \pm \Delta\alpha$	$\beta \pm \Delta\beta$	ε
A	$10.0 \leqslant M_* \leqslant 11.5$	0.074 ± 1.733	0.240 ± 0.162	0.235
B	$10.0 \leqslant M_* \leqslant 11.5$ (spec)	0.037 ± 2.467	0.141 ± 0.227	0.212
C	$10.5 \leqslant M_* \leqslant 11.5$	-0.043 ± 2.093	0.477 ± 0.192	0.182
D	$10.5 \leqslant M_* \leqslant 11.5$ (spec)	-0.152 ± 3.839	0.411 ± 0.350	0.209

colour from $\sim 1.2 \leqslant z_{850} - H_{160} \leqslant 2.0$, which corresponds to a range of M_*/L of $-0.37 \leqslant \log(M_*/L) \leqslant -0.85$.

Most of the galaxies have negative colour gradients. 28 out of 36 galaxies ($\sim 78\%$) show a negative gradient, and 15 out of 36 ($\sim 42\%$) have very steep gradients with $\nabla_{z_{850}-H_{160}} < -0.5$. The median colour gradient and 1σ scatter is $\langle \nabla_{z_{850}-H_{160}} \rangle = -0.45 \pm 0.43$ (error on the median 0.09) and the median $\log(M_*/L)$ gradient $\nabla_{\log(M/L)}$ is -0.27 ± 0.25 (error on the median 0.05). This is consistent with previous findings at higher redshift $1.3 < z < 2.5$ (Guo et al. 2011) which showed that passive galaxies have red cores and bluer stellar population at the outskirts. We find no strong dependence of colour gradients with stellar mass, with a median Spearman's rank correlation coefficient of $\rho \simeq 0.32, p \simeq 0.06$ computed using a bootstrapping method.

At redshift 1.39, the observed $\nabla_{z_{850}-H_{160}}$ colour gradient corresponds to rest-frame ∇_{U-R}. To ensure rest-frame ∇_{U-R} matches $\nabla_{z_{850}-H_{160}}$, we compute ∇_{U-R} from the observed $\nabla_{z_{850}-H_{160}}$ and $z_{850} - H_{160}$ colour using simple stellar population models (SSPs) of Bruzual & Charlot (2003) as a sanity check. The details of the methodology are described in Section 6.3.1. We confirm that the median gradient $\langle \nabla_{U-R} \rangle = -0.53$ is comparable to $\nabla_{z_{850}-H_{160}}$.

We overplot the average local $(U - R)$ colour gradient from Wu et al. (2005) on Figure 11 ($\nabla_{U-R} = -0.21 \pm 0.04$) for comparison. Wu et al. (2005) studied the colour gradients for a sample of 36 local field early-type galaxies from SDSS and 2MASS. Due to a lack of deep U-band imaging, the $(U - R)$ colour gradient is not available in most local galaxy surveys. In order to take into account the average age difference between field and cluster passive galaxies (e.g. Thomas et al. 2005, 2010), we evolve the gradients of Wu et al. (2005) for an additional 2 Gyr with Bruzual & Charlot (2003) SSP models (assuming an age gradient of -0.05 and a metallicity gradient of -0.2 consistent with the literature (La Barbera et al. 2005; Wu et al. 2005)). The extrapolated value ($\nabla_{U-R} = -0.20$) is very close to the one for local ETGs. The average $(U - R)$ colour gradient at $z \sim 1.39$ is found to be ~ 2 times steeper than colour gradients observed locally.

As a consistency check, we repeat the colour gradient measurements in $\nabla_{Y_{105}-H_{160}}$ ($\sim$ rest-frame $g - r$), and find consistent results with the $U - R$. We also compare them with the $g - r$ colour gradient in the SPIDER cluster sample and note that the $g - r$ gradients at $z \sim 1.39$ (median and 1σ scatter $\langle \nabla_{g-r} \rangle = -0.16 \pm -0.16$) are steeper than the local $g - r$ gradients (-0.042 ± 0.144), although with a smaller dynamic range. Details are described in Appendix E. Appendix E also explores the dependence of the local $g - r$ gradients on environment within the full SPIDER sample. We report to later sections for a discussion on the origin of colour or M_*/L gradient in these high redshift passive galaxies.

5.4 Comparison of light-weighted to mass-weighted structural parameters

In Figure 12, we compare the light-weighted sizes (R_{e-circ}) measured in H_{160} band to the mass-weighted sizes (hereafter $R_{e-circ,mass}$) measured from the mass maps. The mass-weighted sizes are $\sim 41\%$ smaller than the H_{160} light-weighted sizes, with a median difference of $\langle \log(R_{e-circ,mass}/R_{e-circ}) \rangle = -0.23$. The scatter $\sigma_{\log(R_{e-circ,mass}/R_{e-circ})}$ is ~ 0.11. In the most extreme case the mass-weighted size can be up to $\sim 60\%$ smaller than its light counterpart (excluding the cluster BCG which is $\sim 65\%$ smaller).

The general trend of mass-weighted sizes being smaller is in qualitative agreement with the study at similar redshift by Szomoru et al. (2013), who computed the mass-weighted sizes using radially
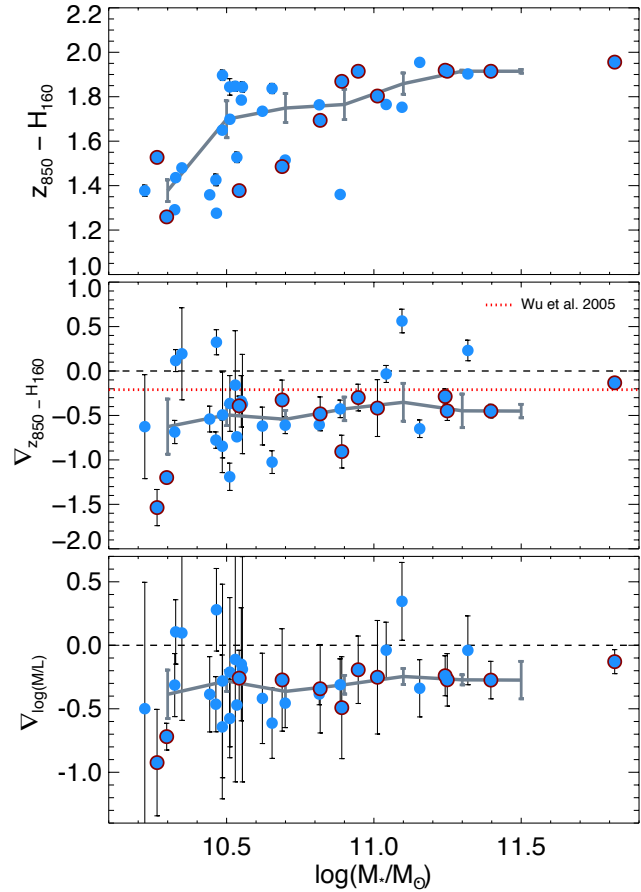

Figure 11. Colour and colour gradients in the cluster XMMUJ2235-2557. From top to bottom: $z_{850} - H_{160}$ aperture colour (1″ in diameter), colour gradient $\nabla_{z_{850}-H_{160}}$ and $\log(M_*/L)$ gradient $\nabla_{\log(M/L)}$ as a function of stellar mass. Spectroscopically confirmed objects are circled in dark red. At redshift 1.39, this roughly corresponds to rest-frame $(U - R)$ colour gradient. The black dashed line in each panel shows the reference zero level. The red dotted line shows the average local $(U - R)$ gradient from Wu et al. (2005). The grey line in each panel shows the running median and the error bars show the uncertainty of the median in each bin. When there is only one point in the bin, the uncertainty of the quantity is plotted instead.

binned 1D surface brightness profiles for passive field galaxies in CANDELS. As we will show in the discussion, this is consistent with the colour gradients in high redshift passive galaxies, in the sense that they usually have redder cores and bluer outskirts (see also, e.g. Saglia et al. 2000; Wuyts et al. 2010; Szomoru et al. 2011; Guo et al. 2011). Negative colour gradients can lead to smaller mass-weighted sizes compared to light-weighted sizes, as a higher M_*/L ratio at the centre results in a more concentrated mass distribution compared to the light distribution (hence, a smaller a_e).

5.5 Stellar mass – mass-weighted size relation

In the bottom panel of Figure 10 we show the mass – size relation with the mass-weighted sizes. In the above section we have demonstrated that the mass-weighted sizes are $\sim 41\%$ smaller than light-weighted sizes. Here we investigate how using mass-weighted sizes can affect the mass – size relation.

We fit the stellar mass – mass-weighted size relation, using equation 6. The best-fit parameters are summarised in the second
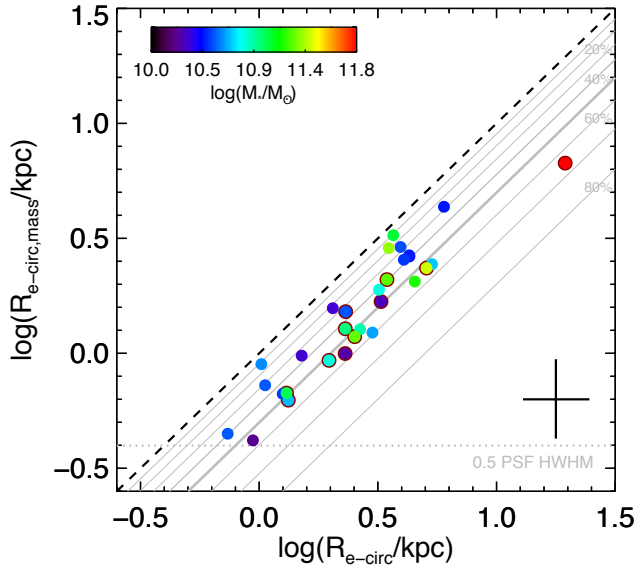
Figure 12. Comparison between mass-weighted size $R_{e-circ,mass}$ and light-weighted size R_{e-circ} of passive galaxies in the cluster XMMUJ2235-2557. Spectroscopically confirmed objects are circled in dark red. The dashed line is the one-to-one relation. Each successive grey line represents a 10% decrement to the one-to-one relation. 50% decrement are marked with a thick grey line. The galaxies are colour coded with their integrated mass. The cross shows the median uncertainty of the light-weighted sizes and mass-weighted sizes. Note that the two uncertainties are correlated to some extent.

half of Table 2. The fact that the mass-weighted sizes are smaller can be seen from the intercept of the fits. Apart from the intercept, there seems to be a slight change in the slope of the relation if mass-weighted sizes are used. The best-fitted relation for the full sample has a value $\beta = 0.240$, 34% lower than the light-weighted size – mass relation, though the two relations are consistent within 1σ.

We check that the change of slope is not due to the discarded objects. More statistics are required to confirm if there is a shallower mass dependence for mass-weighted sizes with respect to light-weighted sizes.

6 DISCUSSION

6.1 Dependence of ratio of mass-weighted to light-weighted sizes on galaxy properties

In Section 5.4 we have shown that the mass-weighted sizes are smaller than the corresponding light-weighted sizes and that the majority of the galaxies have negative colour gradients steeper than local passive galaxies. Intuitively, one might expect some correlation between the ratio of mass-weighted to light-weighted sizes ($\log(R_{e-circ,mass}/R_{e-circ})$, hereafter the size ratio) on the physical parameters that are related to the underlying stellar population, such as colour, stellar mass, and various structural parameters.

Hence, here we investigate the origin of the size ratio in our cluster by examining the correlation with a number of integrated properties. In Figure 13 we show the correlations between the size ratio with the stellar mass, colour gradient $\nabla_{z_{850}-H_{160}}$, M_*/L gradient $\nabla_{\log(M/L)}$, $z_{850} - H_{160}$ colour, light-weighted / mass-weighted Sérsic indices, sizes, mean surface brightness Σ, mean surface mass density Σ_{mass} and the mean surface mass density within a radius of

1 kpc Σ_1. Recent works have shown that Σ_1 is tightly correlated with stellar mass and is closely related to quenching of star formation (Fang et al. 2013; van Dokkum et al. 2014; Barro et al. 2015). All parameters plotted are given in Table F1. Running median and 1σ scatter are over-plotted in each panel.

We search for possible correlations with these physical parameters and compute the correlation coefficients again using the bootstrapping method. We see a mild dependence with the mass surface density with a ρ value of $-0.45, p \simeq 0.01$. There is also a weak dependence for the colour gradient and the M_*/L gradient, with $\rho \simeq 0.32, p \simeq 0.07$ and $\rho \simeq 0.38, p \simeq 0.04$. In addition, we see a weak dependence with the mass-weighted size, which has the highest ρ among the light-weighted and mass-weighted structural parameters. In Section 5.5, we suspect a difference in the mass dependence for the mass-weighted sizes with respect to the light-weighted sizes, which if genuine, implies a correlation between size ratio and the stellar mass. Nevertheless, there is no significant correlation with mass. All other correlations have a $|\rho|$ value < 0.3.

In summary, with the exception of mass surface density, most of the parameters do not show significant dependence with the size ratio. That our measured mass-weighted sizes tend to be significantly smaller than light-weighted sizes can only happen because there are gradients in mass-to-light ratio and colours, as seen previously. Therefore it is encouraging to see that there are (mildly significant) positive correlations between the ratio of sizes and the gradients in colour and M_*/L. That the correlations are not perfect illustrates the contributions of both uncertainties in the data and method, and the fact that our Sérsic fits are actually quite different from a straightforward linear 1D fit as used to derive the gradients. A more sophisticated fit to these parameters, and better correlations, require higher S/N and / or a larger sample.

6.2 Evolution of the ratio of mass-weighted to light-weighted sizes to $z \sim 0$

To investigate the evolution of the size ratio, we compare the size ratio of the cluster sample with the local size ratio computed from a sample of local passive galaxies in high-density environment selected from the SPIDER survey in Figure 14 (the SPIDER cluster sample, see Section 4 for details). We also compare with the SPIDER sample for completeness. We binned the size ratio of the SPIDER cluster sample (and the SPIDER sample) in mass bins of 0.2, in the mass range $10.2 \leqslant \log(M_*/M_\odot) \leqslant 11.6$ ($10.0 \leqslant \log(M_*/M_\odot) \leqslant 11.6$ for the SPIDER sample), to ensure there are sufficient numbers of local galaxies (> 50) in individual bins.

We find that the mass-weighed sizes in the SPIDER cluster sample are on average $\sim 12\%$ smaller than the r-band sizes with a median $\langle\log(R_{e-circ,mass}/R_{e-circ})\rangle = -0.055$ (~ 13 % for the SPIDER sample, $\langle\log(R_{e-circ,mass}/R_{e-circ})\rangle = -0.062$), consistent with previous result (Szomoru et al. 2013).

In addition, we find that there is an intriguing offset between the median size ratio of the cluster sample and the SPIDER cluster sample, with a difference of $\langle\log(R_{e-circ,mass,1.39}/R_{e-circ,1.39}) - \log(R_{e-circ,mass,0}/R_{e-circ,0})\rangle = -0.18$ (-0.17 for the SPIDER sample).

A possible issue is the effect of recently quenched galaxies on the size evolution, i.e. the progenitor bias. It has been shown to have a non-negligible effect on inferred size evolution, and is able to explain part if not all of the observed size evolution (e.g. Saglia et al. 2010; Valentinuzzi et al. 2010b; Carollo et al. 2013; Poggianti et al. 2013; Jørgensen et al. 2014; Belli, Newman & Ellis 2015;
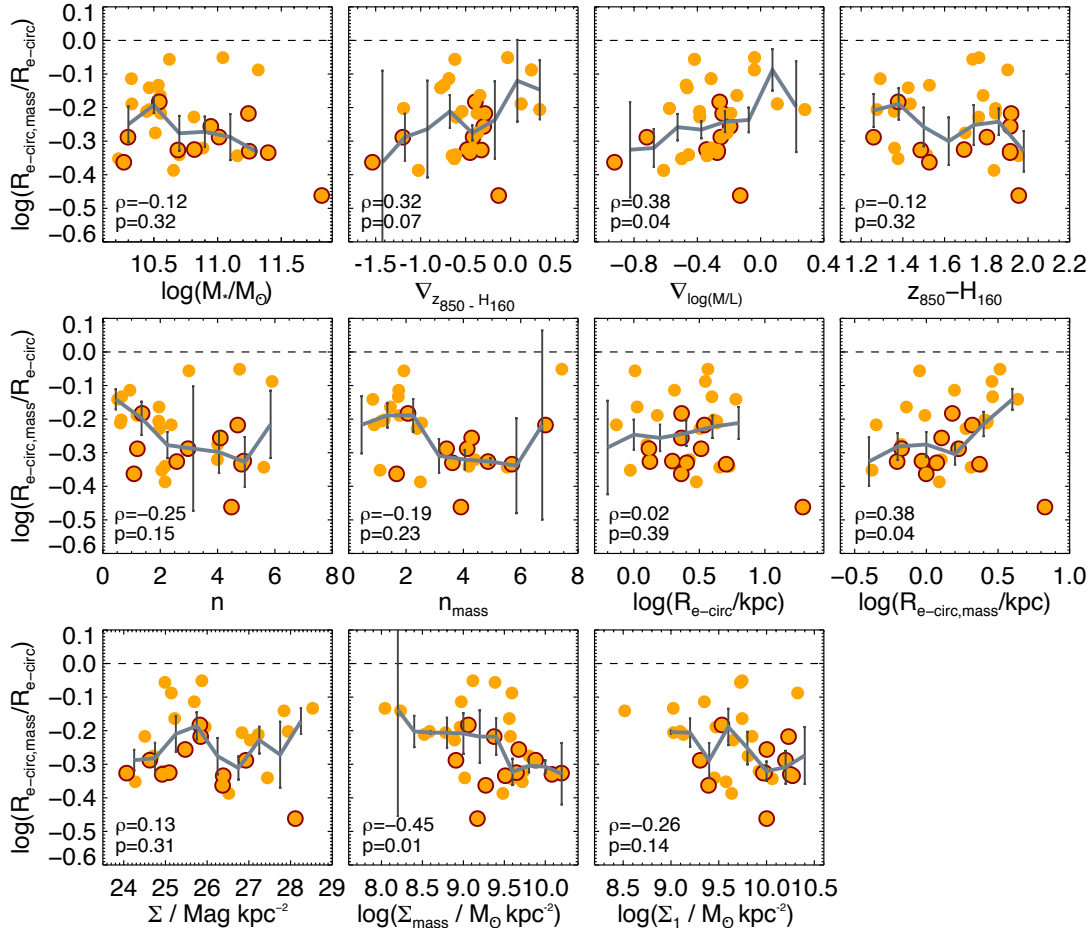
Figure 13. Dependence of ratio of mass-weighted to light-weighted sizes on different galaxy parameters. H_{160} band sizes (rest-frame r-band) are used to compute the ratio of mass-weighted to light-weighted sizes ($R_{e-circ,mass}/R_{e-circ}$). From top left to bottom right: stellar mass, colour gradient $\nabla_{z_{850}-H_{160}}$, M_*/L gradient $\nabla_{\log(M/L)}$, $z_{850}-H_{160}$ colour, light-weighted Sérsic index n, mass-weighted Sérsic index n_{mass}, light-weighted effective radius R_{e-circ}, mass-weighted effective radius $R_{e-circ,mass}$, mean surface brightness Σ, mean surface mass density Σ_{mass} and mean surface mass density within 1 kpc Σ_1. Spectroscopically confirmed objects are circled in dark red. Grey line in each panel shows the running median. The error bars show the uncertainty of the median in each bin. When there is only one point in the bin, the uncertainty of the ratio is plotted instead.

Keating et al. 2015). The effect on the evolution of the size ratio is however unclear, as the newly quenched galaxies may have a range of M_*/L gradients that depends on the quenching mechanism involved.

Using age measurements from La Barbera et al. (2010a), we try to correct the progenitor bias in the size ratio of the SPI-DER cluster sample. An age cut is applied to the SPIDER cluster sample to remove galaxies that are younger than 8.98 Gyr, the time duration from $z \sim 1.39$ to $z \sim 0$. The result is shown as a light brown line and wheat band in Figure 14 in the mass range $10.2 \leqslant \log(M_*/M_\odot) \leqslant 11.6$. Although some changes can be seen, the size ratios of the progenitor bias corrected sample are in general consistent with the SPIDER cluster sample, with median $\langle \log(R_{e-circ,mass}/R_{e-circ}) \rangle = -0.065$. The median logarithmic size ratios in each bin between the two are within ± 0.05. The offset between the median size ratio of the cluster sample and the progenitor bias corrected SPIDER cluster sample is -0.16. Hence, the progenitor bias alone does not explain the observed offset.

The smaller size ratio at $z \sim 1.39$ suggests that the M_*/L gradient is larger (i.e. steeper) in these high redshift passive cluster galaxies compared to the local ones. This implies an evolution of

M_*/L gradient with redshift, consistent with our finding that the colour gradient at high redshift cluster passive galaxies is much steeper than the local ones. In the next section, we try to explore the origin of the colour (and M_*/L) gradient and the physical processes for the evolution of these passive galaxies.

6.3 Origin and evolution of colour gradients with redshift

In Section 5.3 we have shown that the median colour gradient in our sample is ~ 2 times steeper than the measured local $(U - R)$ gradient ∇_{U-R} from Wu et al. (2005).

The origin of the colour gradients is directly related to how the stellar population in galaxies assembled and evolved. It is however challenging to segregate the impact of age or metallicity using a small sample of galaxies due to degeneracies between colour, age and metallicity. In previous studies, colour gradients are mostly interpreted as either age gradients ($\nabla_{age} = d\log(age)/d\log(a)$) at fixed metallicity or metallicity gradients ($\nabla_Z = d\log(Z)/d\log(a)$) at fixed age. Some works in clusters at $z \sim 0.4$ (Saglia et al. 2000) and local clusters (e.g. Tamura & Ohta 2003) showed however that the colour gradients may be preferentially produced by radial varia-
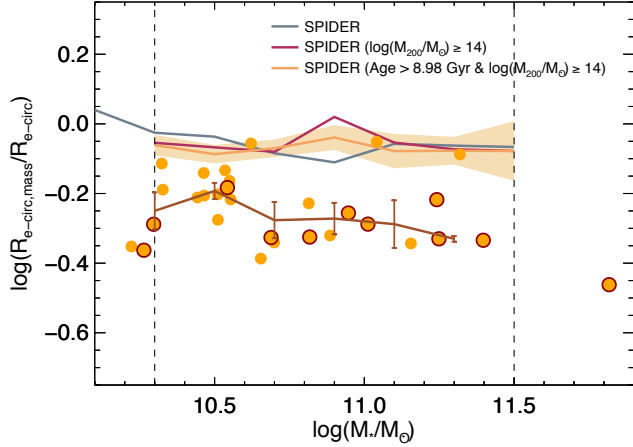
Figure 14. Comparison of the ratio of mass-weighted to light-weighted sizes vs. stellar mass at different redshift. Same as top leftmost panel of Figure 13 but includes the local size ratios from the SPIDER cluster sample. Spectroscopically confirmed objects are circled in dark red. The brown line corresponds the running median in mass bins of 0.2 with a window of width 0.3, and the error bars show the uncertainty of the median in each bin. The median size ratio of the SPIDER cluster sample (from mass range $10.2 \leqslant \log(M_*/M_\odot) \leqslant 11.6$) is plotted as a dark magenta line, while the median size ratio of the SPIDER sample (from mass range $10.0 \leqslant \log(M_*/M_\odot) \leqslant 11.6$) is plotted as a slate grey line in mass bin of 0.2. The light brown line and wheat shaded region correspond to the median size ratio and $\pm 1\sigma$ error on the median for the progenitor bias corrected SPIDER cluster sample (age > 8.98 Gyr and $\log(M_{200}/M_\odot) \geqslant 14$) from mass range $10.2 \leqslant \log(M_*/M_\odot) \leqslant 11.6$.

tion in metallicity rather than age. The age gradients in local passive galaxies are consistent with 0 (or slightly positive), while the average metallicity gradient is found to be of $\nabla_Z \approx -0.1$ to -0.3 (see also Mehlert et al. 2003; Wu et al. 2005; La Barbera & de Carvalho 2009). This result is also supported by recent studies with integral field spectroscopy (e.g Kuntschner et al. 2010; González Delgado et al. 2014; Oliva-Altamirano et al. 2015; Wilkinson et al. 2015).

Our $(U-R)$ and $(g-r)$ colour gradient measurements alone unfortunately do not allow us to break the age-metallicity degeneracy. Nevertheless, with additional colour information at $z \sim 0$, by studying the evolution of the colour gradient with redshift we can shed light on the origin of the colour gradients in our sample of early type galaxies in clusters.

6.3.1   *Methodology*

We investigate quantitatively the evolution of colour gradients by modeling them in our cluster sample under different assumptions of the radial variation of stellar population properties. Simply put, we would like to evolve the observed $z_{850} - H_{160}$ colour gradients $\nabla_{z_{850}-H_{160}}$ in the cluster sample to see under which conditions in age and metallicity gradient will they match the observed $(U-R)$ gradient at $z \sim 0$.

For simplicity, here we assume the stellar populations in the passive galaxies are coeval and chemically homogeneous in the regions we considered, hence they can be described by simple stellar populations (SSP) models. We use the models of Bruzual & Charlot (2003) (hereafter BC03) and adopt a Chabrier IMF. The BC03 distribution provides SSP models with metallicities $Z = [0.0001, 0.0004, 0.004, 0.008, 0.02, 0.05]$ from $t = 0$ to the age of the Universe in unequally spaced time steps. The results below
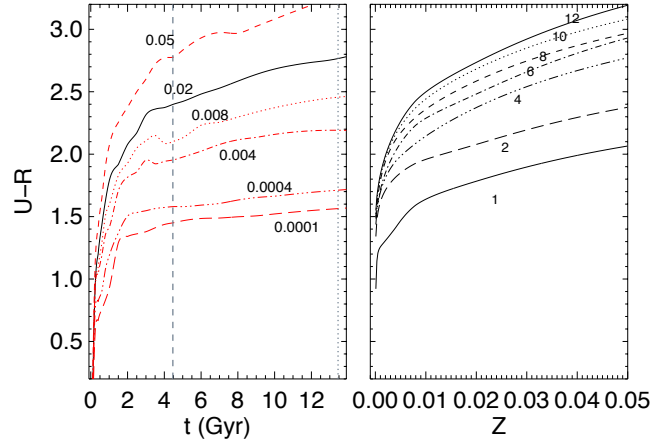

Figure 15. Rest-frame $U - R$ colour of stellar populations with different ages and metallicities. Left panel: $U - R$ colour-age relations. The black line shows the stellar populations with solar metallicity ($Z = 0.02$). The red lines show populations with different metallicities ($Z = 0.0001, 0.0004, 0.004, 0.008, 0.05$) as indicated. The grey dotted line shows the current age of the Universe (13.45 Gyr) with our choice of cosmology and the grey dashed line shows the age of Universe at redshift 1.39 (4.465 Gyr). Right panel: $U - R$ colour-metallicity relations. The black lines shows populations with different ages (in Gyr) as indicated.

do not strongly dependent on the choice of IMF, since the $U - R$ broad band optical colours under different IMFs (e.g. Chabrier vs. Salpeter) are in reasonable agreement with each other. In Appendix D we show that adopting exponentially declining τ-models for this analysis (instead of SSPs), does not change the results.

We compute the rest-frame $U - R$ colour for individual SSP models with different ages and metallicities by convolving the model SEDs with the U and R filters. The colours are then interpolated with a cubic spline to obtain an equally spaced colour grid in age and metallicity. Figure 15 shows the $U - R$ colour at different ages (left, i.e. the colour-age relations) and metallicities (right, i.e. the colour-metallicity relations) as an example. Using the same method, we compute a $z_{850} - H_{160}$ colour grid by redshifting the SSP models to $z = 1.39$.

Similar to Section 5.3, we have also repeated the above analysis using the $g - r$ colour gradients and find results which are completely consistent (Appendix E). Since the $g - r$ colour has less dynamic range in the evolution, in the following we will mainly discuss the result of the $U - R$ colour gradients.

To simplify the modeling process, we analyse the evolution of colour at two radii, $0.5a_e$ and $2a_e$, representing the inner region and outer region of the galaxy respectively. In similar studies (e.g. Saglia et al. 2000) more central regions are used instead ($0.1 - 0.2R_e$), but this is not possible at this redshift due to limited resolution. Nevertheless, our choice of radial range is sufficient for the purpose as the colour gradients are well-fitted by a linear relation in logarithmic radius (see Figure 5).

Because of the age-metallicity degeneracy, we consider several scenarios with additional assumptions in the age or metallicity gradients. In this study we explore three possibilities (cases) to interpret the colour gradient evolution:

● **Case I - Pure age-driven gradient evolution** – In this case we explore the possibility of using a single age gradient to interpret the evolution of colour gradients. The inner and outer regions are assumed to have identical metallicities (i.e. flat metallicity gradients

$\nabla_Z = 0$). Assuming a certain metallicity for the inner regions (and equivalently the outer regions), we derive ages of the stellar population of the inner and outer regions in each galaxy respectively through matching the observed $z_{850} - H_{160}$ colours to the derived $z_{850} - H_{160}$ SSP colours.

• **Case II - Pure metallicity-driven gradient evolution** – In this case we assume that the stellar population in the inner and outer region are coeval (i.e. flat age gradients $\nabla_{age} = 0$). Assuming a certain metallicity for the inner regions, we derive the inner ages in each galaxy using the same method as I. The same age is then applied to the outer regions. With ages and $z_{850} - H_{160}$ colours, metallicities in the outer regions are then derived using the colour-metallicity relations.

• **Case III - Age-driven gradient evolution with an assumed metallicity gradient** – Same as case I, but assume a fixed metallicity gradient with $\nabla_Z = -0.2$, which is the mean value observed in local passive galaxies (e.g. Tamura & Ohta 2003; Wu et al. 2005; Brough et al. 2007; Reda et al. 2007) as well as in recent simulations (e.g. Hirschmann et al. 2015). Again assuming a certain metallicity for the inner regions, we derive the inner ages from the $z_{850} - H_{160}$ colour. The metallicity in the outer regions is then computed according to the assumed gradient; outer ages are then derived with the computed metallicity.

In summary, in each case we obtain the ages and metallicities in the inner and outer region of each cluster galaxy. We then evolve the corresponding SSPs in both regions to $z = 0$, and compute the corresponding local $(U - R)$ colour gradients. We also compute the rest-frame $(U - R)$ gradient for the high-redshift sample for comparison. For each of the three cases above, we test three scenarios with different assumed metallicity for the inner regions, sub-solar, solar and super-solar ($Z = 0.008, 0.02, 0.05$ or equivalently $[Fe/H] = -0.33, 0.09, 0.56$) (Bruzual & Charlot 2003). Assuming metallicities with $Z < 0.008$ or $Z > 0.05$ is unphysical for most galaxies in the cluster sample.

Under different assumed metallicity for the inner region, occasionally the age (or metallicity for case II) determination for some galaxies results in an unphysical age (or metallicity). With our choice of cosmological parameters, the age of universe at $z = 1.39$ is 4.465 Gyr. Deduced ages that are too old (> 4.465 Gyr for $z = 1.39$ or > 13.45 Gyr for $z = 0$ within 1σ uncertainty) are rejected to avoid drawing incorrect conclusions. Galaxies that are rejected may simply be unphysical to be modelled with particular metallicity (see, for example in Figure 15, a galaxy with $U - R > 2.15$ at $z = 1.39$ will result in an unphysical age if one assumes $Z = 0.008$) or have a more complicated star formation history, which cannot be well-represented by SSPs.

6.3.2 *Case I – Pure age-driven gradient evolution*

In the top panel of Figure 16, we show the evolution of the rest-frame $(U - R)$ colour gradient from $z = 1.39$ to $z = 0$ under the assumption of pure age gradient. We show the scenario with assumed solar metallicity ($Z = 0.02$) in the inner region. We find that although the gradients evolve in the correct direction, the median gradient of the evolved sample is too shallow. The result is almost identical if we assume sub-solar or super-solar metallicity for the inner region instead, with median evolved colour gradient of $(Z, \nabla_{U-R}) = (0.008, -0.036), (0.02, -0.034), (0.05, -0.037)$. Under the assumption of sub-solar metallicity, 19 out of 36 galaxies have a physical age. On the other hand, most of the galaxies are retained if we assume a solar (33 out of 36) or super-solar metal-
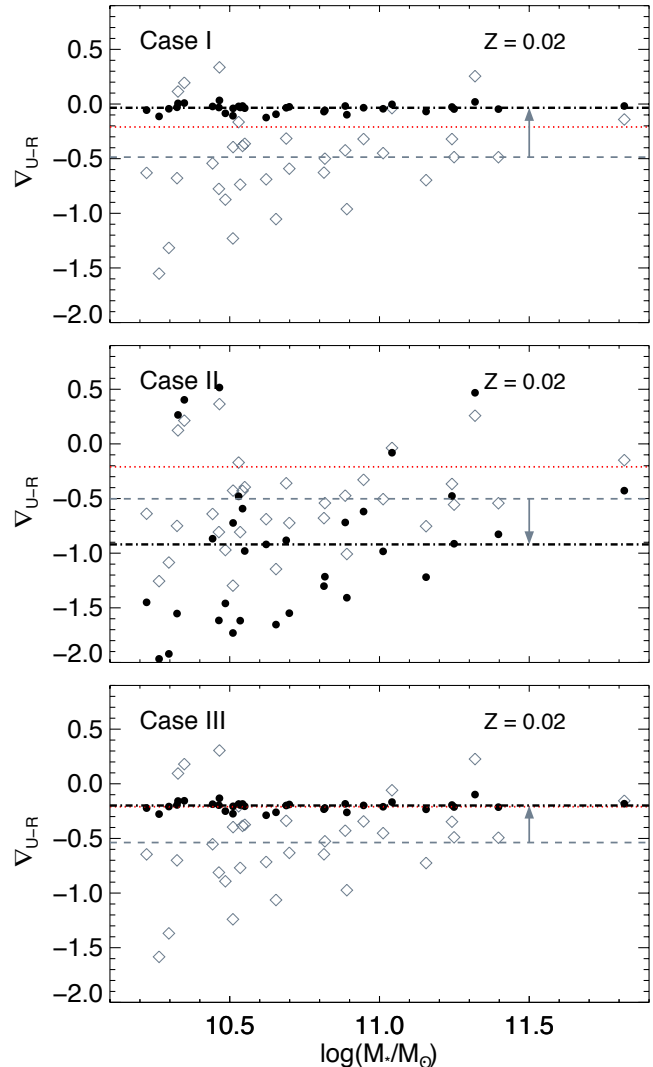

Figure 16. Evolution of colour gradient under different assumptions in the age / metallicity gradient. From top to bottom: Case I – Pure age-driven gradient evolution, Case II – Pure metallicity-driven gradient evolution and Case III – Age-driven gradient evolution with assumed metallicity gradient. Only the solar ($Z = 0.02$) metallicity scenario in each case is shown. Grey diamonds correspond to the $(U - R)$ gradient at redshift 1.39, with the median plotted as the grey dashed line. Black circles indicate the predicted $(U - R)$ gradient at redshift 0 of the same galaxy, and the black dot-dashed line indicate the median. Their masses remain unchanged as we do not consider any mass growth over the period. The grey arrow in each panel shows the direction of evolution of the median gradient. The red dotted line corresponds to the observed $(U - R)$ gradient at redshift 0 by Wu et al. (2005).

licity (36 out of 36). We conclude that in the reasonable range of metallicity that we covered, a pure age-driven gradient is not able to match the observed evolution of colour gradient.

The reason behind the rapid evolution is the flattening of the SSP colour-age relation over time. Since we assume identical metallicities for both inner and outer regions, the inner and outer region of an individual galaxy lie on the same colour-age relation in Figure 15. Take the solar metallicity $Z = 0.02$ case (black solid line) as an example, the $U - R$ colour increases sharply from 0 to 4 Gyr but flattens after, hence the $(U - R)$ gradient evolves to almost zero at redshift 0.

6.3.3   Case II – Pure metallicity-driven gradient evolution

Instead of using a flat metallicity gradient as case I, the middle panel of Figure 16 shows the evolution of the $(U - R)$ gradient under the assumption of pure metallicity-driven gradient, or in other words, a flat age gradient $\nabla_{age} = 0$. Again, we show the scenario with assumed solar metallicity $(Z = 0.02)$ in the inner region. Similar to case I, galaxies that have unphysical ages / metallicities are discarded. 21, 33 and 31 out of 36 galaxies are retained in each metallicity scenario $(Z = 0.008, 0.02, 0.05)$ respectively. From Figure 16, we can see that the median gradient of the evolved sample is even steeper compared to the one at redshift 1.39. Hence, it is clear that a pure metallicity-driven gradient fails to reproduce the observed gradient. The median gradients of the evolved sample in the three metallicity scenarios are $(Z, \nabla_{U-R}) = (0.008, -0.691), (0.02, -0.920), (0.05, -0.900)$.

The evolution can be explained using the colour-metallicity relations in the right panel of Figure 15. As the population ages, the $U - R$ colour-metallicity relation steepens (for example, from 2 Gyr to 4 Gyr), which causes the colour gradient to become more negative. The steepening stops at around ~ 4 Gyr, thus the colour gradient after then remains unchanged. For the solar metallicity scenario, the median metallicity gradient and 1σ scatter we found is $\langle \nabla_Z \rangle \sim -1.32 \pm 1.22$ at $z = 1.39$, much higher than that observed in local galaxies. Of course in reality metallicity in individual galaxies differs, but mixing galaxies with different metallicity within our metallicity range would not change this conclusion.

6.3.4   Case III – Age-driven gradient evolution with assumed metallicity gradient

The bottom panel of Figure 16 shows the evolution of the $(U - R)$ gradient with a metallicity gradient as observed in local passive galaxies: $\nabla_Z = -0.2$. Similar to case I, galaxies with unphysical ages in the inner or outer regions are rejected. 19, 33 and 36 out of 36 galaxies are retained in each metallicity scenario respectively. The solar metallicity scenario works reasonably well for the majority of the sample with evolved median gradient of $(Z, \nabla_{U-R}) = (0.02, -0.198)$, which is in close agreement with the observed value in the local universe by Wu et al. (2005). Despite a number of objects have to be discarded due to unphysical age, the median gradient as well as the individual gradients of the evolved samples in the sub-solar metallicity scenario $(Z, \nabla_{U-R}) = (0.008, -0.180)$ is also close to but slightly smaller than the observed local value. Assuming super-solar metallicity for the inner regions on the other hand, predicts gradients that are slightly too steep $(Z, \nabla_{U-R}) = (0.05, -0.232)$.

Besides the median values, the scatter in the evolved colour gradients is also in excellent agreement to the local value by Wu et al. (2005) $(\nabla_{U-R} = -0.21 \pm 0.04)$. For example, for the solar metallicity scenario the scatter reduces from 0.37 at $z \sim 1.39$ to 0.06 at $z = 0$.

6.3.5   Implications and limitations

From the above case study, we find that the presence of an age gradient is a necessary condition for the evolution of the colour gradient, and with metallicity gradient they can sufficiently reproduce the magnitude of the evolution of the colour gradient from $z = 1.39$ to $z = 0$.

An age-driven gradient evolution with a metallicity gradient close to the local value is the most probable scenario, as it can well-reproduce the observed evolution of the colour gradients over redshift in both median and scatter. Below we try to understand why this is the case. Among the three metallicity scenarios, the one with solar metallicity seems to best match the evolution of colour gradients for most galaxies. In this scenario we find a median age gradient and 1σ scatter of $\langle \nabla_{age} \rangle = -0.33 \pm 0.37$ at $z = 1.39$ (i.e. a median age difference ~ 1.4 Gyr between the inner and outer regions).

Figure 17 shows the evolution of the colour gradient in this case (case III), assuming a formation redshift of the inner regions of $z_{form} = 3.0$ and an age gradient of $\langle \nabla_{age} \rangle = -0.33$ at $z = 1.39$. The top left panel shows the change in the evolution for different formation redshifts $z_{form} = 2.0, 4.0, 5.0$, selected to be consistent with findings in recent spectroscopic studies at similar redshift (Gargiulo et al. 2012; Belli, Newman & Ellis 2015). The net evolution from $z \sim 1.39$ to $z \sim 0$ is clearly insensitive to the formation redshift. On the other hand, the path of evolution depends largely on the age gradient; from the top right panel we show that with different initial age gradient at $z = 1.39$, a large range of colour gradients at $z = 1.39$ can reach similar value at $z \sim 0$. In other words, if the colour gradients in high redshift passive galaxies are mainly due to radial variation in age, this assumption would be able to match the evolution of colour gradient for most galaxies.

This is in agreement with Gargiulo et al. (2012), who investigated the origin of the colour gradient on a sample of early-type galaxies at $0 < z < 1.9$ with spatially resolved colour and global SED fitting. They found that the colour gradients of $\sim 50\%$ of their sample can be reproduced with pure age gradients, while invoking pure metallicity gradients can only explain a small subset of their sample. In addition, extremely steep metallicity gradients are required that are only marginally comparable with those observed in the local Universe. A similar recent study by De Propris, Bremer & Phillipps (2015) studied the ratio of galaxy sizes in two bands (as a proxy of the colour gradient) in red sequence galaxies in four clusters with $< z > \sim 1.25$ (including this cluster) also found an indication of negative colour gradients, which they also attribute to due to the presence of age gradients.

Our result is also not inconsistent with studies on local and intermediate redshift passive galaxies which suggest colour gradients are mainly due to metallicity gradients (e.g. Saglia et al. 2000; Tamura & Ohta 2003). For example, Kuntschner et al. (2010) found a mean $\nabla_Z = -0.25 \pm 0.11$ and mean $\nabla_{age} = 0.02 \pm 0.13$ for galaxies with age > 8 Gyr. Nevertheless, the age gradient (or its presence) is very difficult to constrain in local passive galaxies. As Gargiulo et al. (2012) pointed out, the effect of the age difference in the inner and outer regions is much more enhanced when the stellar population is young (i.e. at high redshift). Indeed, the age gradient flattens quickly over redshift. For example, with a median age gradient of $\langle \nabla_{age} \rangle = -0.33 \pm 0.37$ at $z = 1.39$ (from our best scenario); assuming passive evolution this corresponds to a median age gradient of $\langle \nabla_{age} \rangle = -0.05 \pm 0.06$ at $z = 0$, which is consistent with a flat age gradient.

Given the assumption that the cluster passive galaxies at $z = 1.39$ have the same metallicity gradient as the local ones, it is implied that the evolution of colour gradients from $z \sim 1.4$ to 0 can be explained simply through passive evolution. This is consistent with luminosity function studies in clusters (e.g. Andreon 2008; De Propris, Phillipps & Bremer 2013).
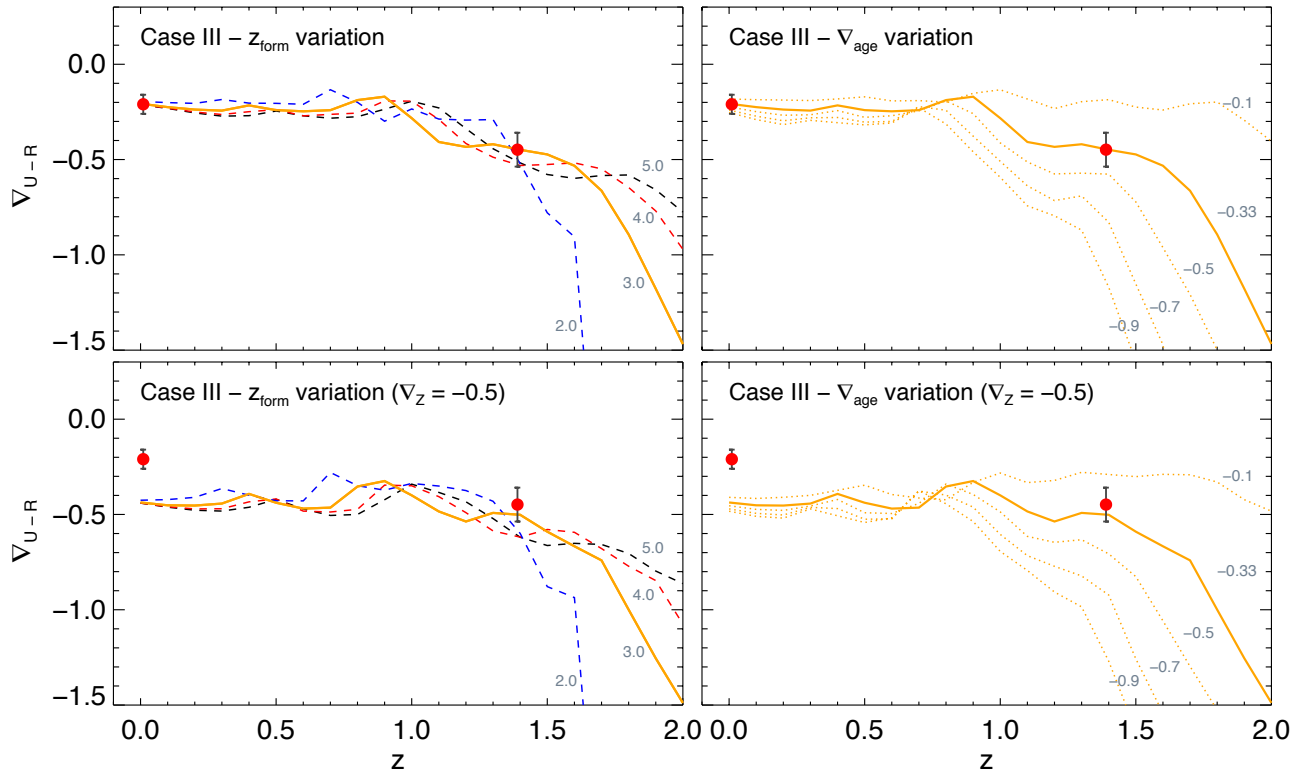
Figure 17. Evolution of colour gradient over redshift in case III (age and metallicity gradient) with assumed solar metallicity $Z = 0.02$. The inner region is assumed to have a formation redshift $z_{form} = 3.0$. The initial age gradient at $z = 1.39$ is $\nabla_{age} = -0.33$, and the assumed metallicity gradient is $\nabla_Z = -0.2$. Top left: variation in formation redshift $z_{form} = 2.0, 4.0, 5.0$ as indicated by blue, red and black dashed lines respectively. Top right: variation in age gradients at $z = 1.39$. The dotted lines show the evolution with different initial age gradient as indicated ($\nabla_{age} = -0.1, -0.5, -0.7, -0.9$). Bottom left and right: same as top left and right but with an assumed metallicity gradient of $\nabla_Z = -0.5$ as predicted by the monolithic collapse formation scenario. Red circles correspond to the median $z_{850} - H_{160}$ gradient of our sample at redshift 1.39, and the observed $(U - R)$ gradient at redshift 0 by Wu et al. (2005). The error bars show the uncertainty of the median.

6.3.6 Physical processes responsible for the evolution of size and colour gradient

While passive evolution is a very tempting conclusion, it alone cannot explain the observed size evolution in clusters over redshift. In Figure 10 we have shown that the cluster passive galaxies are $\sim 40\%$ smaller than their local counterparts. Similarly, a large number of previous studies have confirmed that the sizes of passive galaxies in high redshift clusters are smaller than those in the local Universe (e.g. Rettura et al. 2010; Strazzullo et al. 2010; Papovich et al. 2012; Strazzullo et al. 2013), although a part of this size evolution (e.g. Saglia et al. 2000) or even all (Jørgensen et al. 2014) may be due to progenitor bias. If we assume this observed size evolution is genuine, in the sense that it is not completely an effect of progenitor bias, additional physical processes must be in place over redshift to increase the size of the population but not significantly their stellar mass, and at the same time cannot severely disrupt the existing stellar population gradients.

The "puffing-up" scenario (or adiabatic expansion) is one of the candidates to explain the size evolution of passive galaxies (Fan et al. 2008, 2010). While it may work for increasing the size, it is yet unclear whether it can sufficiently explain the observed evolution of colour gradients. Further detailed investigation with an accurate model is required to test this scenario, but current models (e.g. Fan et al. 2008) are much simplified.

Minor mergers, on the other hand, seem to be a viable sce-

nario as the effects are primarily on the outer part of the galaxies. Hilz, Naab & Ostriker (2013) showed from n-body/SPH simulations that for minor (with a mass ratio 1:10) or intermediate (1:5) dry mergers, the inner region of the galaxy remains almost unchanged and the accreted mass assembles predominantly in the outer part of the galaxy. This is also seen in the cosmological simulations of Rodriguez-Gomez et al. (2015) where the accreted stars from mergers with lower mass ratio (i.e. merging with smaller galaxies) dominate at outer radius. Hence, the inner stellar populations of the galaxy can age through passive evolution without major disturbance. The negative age gradients we find here seem to be consistent with this picture given that the minor mergers are dry and the stars accreted are relatively young.

The minor merger scenario has been known to be a viable mechanism in the field. It is consistent with the observed inside-out growth as seen from the evolution of the stellar mass surface density profiles of passive field galaxies over redshift (e.g. van Dokkum et al. 2010; Patel et al. 2013). Nevertheless, traditionally this type of merger activity is believed to be suppressed in virialized clusters because of the high velocity dispersion, resulting in high relative velocities between cluster members (e.g. Conroy, Wechsler & Kravtsov 2007; Lotz et al. 2013). The exception being mergers of satellite galaxies onto the BCG due to dynamical friction (e.g Burke & Collins 2013; Burke, Hilton & Collins 2015), which contribute to the mass growth of BCG and the intracluster light (ICL).

On the other hand, merger events are thought to be very common in galaxy groups where the velocity dispersion is lower, or when the cluster is still assembling. Recent works have found that sizes of massive passive galaxies are larger in clusters compared to the field at high redshift (e.g. Cooper et al. 2012; Zirm, Toft & Tanaka 2012; Papovich et al. 2012; Lani et al. 2013; Strazzullo et al. 2013; Jørgensen & Chiboucas 2013; Delaye et al. 2014), but not in the local Universe (e.g. Huertas-Company et al. 2013; Cappellari 2013), suggesting an accelerated size evolution in high density environments. This accelerated size evolution is probably due to an enhanced rate of mergers during infall of groups, when the cluster is being assembled (Lotz et al. 2013; Delaye et al. 2014; Newman et al. 2014). While this explains the elevated sizes in cluster compared to the field at fixed redshift, it does not provide an explanation to the subsequent size evolution observed in clusters over redshift $z < 1.5$.

Despite the suppression of galaxy merging activity in clusters based on relative velocity arguments, it is clear that clusters themselves and their associated dark matter halos continue to grow by accreting galaxy groups. It is possible that some accretion can still happen to the cluster galaxy population during infall of these group-scale structures. Simulations of mergers with cluster mass halos have shown that the accreted mass resides mainly in the satellite galaxies and the ICL, but only mildly in the BCG (e.g. White et al. 2007; Brown et al. 2008). Recent simulations also demonstrate that the size of clusters members can grow significantly via major and minor mergers and the frequency of mergers is sufficient to explain the observed size growth in clusters since $z \sim 2$ (Laporte et al. 2013). In addition, the observed merger rate in the cluster galaxy population, excluding the BCG, is poorly constrained. If this is possible, this kind of gradual mass growth is able to explain at the same time the evolution of both size and colour gradient in clusters.

If we take the size evolution into account and assume the mass growth takes place predominately at the outskirts, the stellar population we considered here in the colour gradient ($a < 3.5a_e$) will correspond to the central population at $R < 1.5 - 2R_{e-circ}$ in local cluster galaxies. If the evolution is primarily merger or accretion driven as we suggest above, one would expect that the outer stellar population depends on past merger activity.

Interestingly, there has been some evidence indicating changes in stellar population properties at the outer region of local massive passive cluster galaxies (e.g. NGC 4889 in Coma cluster Coccato, Gerhard & Arnaboldi 2010) as well as in field ellipticals (e.g. Pu et al. 2010). More recent studies have extended the age and metallicity measurements to large radii ($\sim 8 - 10R_e$, for example in La Barbera et al. 2012) and revealed that the outer age or metallicity gradients at $\gtrsim 1 - 2R_e$ are distinct from those in the inner region (e.g. Greene et al. 2013; Pastorello et al. 2014; Raskutti, Greene & Murphy 2014). These changes in stellar population gradients are commonly interpreted to be result of mergers. Nevertheless, these changes can also come from recent quenched galaxies that underwent minor mergers before infalling to the cluster. A progenitor biased corrected sample is needed to address this issue.

6.3.7 Monolithic collapse model

Traditional monolithic models predict very high metallicity gradients in passive galaxies, around $\nabla_Z \sim -0.5$, and a flat if not slightly positive age gradients (e.g. Larson 1974; Carlberg 1984). The fact that we find that on average a negative age gradient is necessary to explain the evolution of the colour gradient is hence inconsistent with the monolithic collapse scenario. This conclusion is in-

dependent of the metallicity gradient we assumed; in case III we assume the metallicity gradient to be $\nabla_Z = -0.2$ as observed in local passive galaxies, but using a steeper metallicity gradient would not work. The bottom two panels in Figure 17 show the effect of assuming a metallicity gradient $\nabla_Z = -0.5$ as predicted by traditional monolithic models. While the $U - R$ colour gradient at redshift 1.39 is in reasonable agreement with the observed value, the evolved gradient is too steep (i.e. insufficient evolution) compared to the observed value at redshift 0.

We also compare our observed $z_{850} - H_{160}$ colour gradients at redshift 1.39 with recent simulations based on a revised version of the monolithic model by Pipino, D'Ercole & Matteucci (2008); Pipino et al. (2010). With semi-cosmological initial conditions, they are able to match the age gradients and metallicity gradients observed in local passive galaxies. Tortora et al. (2013) computed the colour gradients from the Pipino et al. (2010) simulations using BC03 SSP models, for which we can directly use for the comparison. Among the four models presented in their work (E1, E2, E3, E4), we find that our observed median colour gradient seems to be in reasonable agreement with models that predict steep metallicity gradient (E2, $\nabla_Z \sim -0.35$ and E4, $\nabla_Z \sim -0.45$) and nearly flat age gradient at $z = 0$. The comparison with the colour gradients in Gargiulo, Saracco & Longhetti (2011); Gargiulo et al. (2012) at redshift $1 < z < 2$ also gives a similar result. However, the local colour gradient at $z \sim 0$, as well as the $F606W - F850LP$ gradients of high-redshift galaxies in Gargiulo, Saracco & Longhetti (2011) favours the other two models (E1, E3) instead, so there is not a single model that can explain the evolution of colour gradients. Hence, our result cannot be reproduced by the revised monolithic collapse model.

6.3.8 Effect of dust obscuration

A complication that we have not considered above is the effect of dust obscuration. The colour gradient can be affected by the radial variation of dust content. For local passive galaxies, Wise & Silva (1996) pointed out that their colour gradients can be reproduced by a dust gradient, albeit with much higher dust masses than observed (Saglia et al. 2000). Hence the colour gradient in these galaxies should be driven from the variation of the stellar population. Nevertheless, the amount of dust can vary with radius; it is not uncommon to find dust at the centre, for example in Lauer et al. (2005), central dust is visible in almost half of the local passive sample.

At high redshift, measuring the radial variation of dust content is even more difficult due to the compact nature of passive galaxies and limited angular resolution. Several studies suggest that although the effect of dust cannot be completely neglected, it plays only a minor role in driving colour gradients. Belli, Newman & Ellis (2015) showed that their early-type sample at $1.0 < z < 1.6$ have little to no dust extinction (see also Mendel et al. 2015). For the radial variation of dust, Guo et al. (2011) demonstrated from spatially resolved annular SED fitting that, for their sample of six $z \sim 2$ galaxies dust partly contributes to the observed colour gradients, the inferred dust gradient and global extinction have a value of $dE(B-V)/d\log(R) \sim -0.07$ and $\langle E(B-V) \rangle \sim 0.1$. On the other hand, among the 11 early-type galaxies at $1.0 < z < 1.9$ in Gargiulo et al. (2012), half of the sample have no dust extinction (i.e. $A_V = 0$) from global SED fit and for most galaxies the main driver of the colour gradients is certainly not the radial variation of dust.

It is not possible to derive reliable dust gradients from our multi-band photometry data. Hence, for completeness we test whether the effect of dust would affect our conclusion. Assuming

dust mainly affects the central region as in local passive galaxies, we deredden the $z_{850} - H_{160}$ colour in the inner region by a certain magnitude but leaving the outer part unchanged to reduce the observed gradient, then recompute all the evolution under different assumptions. We find that on average a decrease of 0.14 mag in the $z_{850} - H_{160}$ colour at the inner region (~ 0.25 mag at $0.1a_e$) will remove the observed evolution, i.e. the difference of the observed median colour gradient at $z = 1.39$ with local passive galaxies. Assuming the extinction curve by Calzetti et al. (2000), this corresponds to a gradient of $dA_V/d\log(a) \sim -0.40$ or $dE(B-V)/d\log(a) \sim -0.10$. Hence, our conclusion is robust if the dust gradient is less steep than this value. If the genuine colour gradient is shallower at $z = 1.39$, the evolution will be best explained with a shallower age gradient (as seen from the top right panel of Figure 17), and will not change any of our conclusions. Nevertheless, although unlikely, we cannot rule out the possibility that the observed colour gradients are driven by a large amount of dust located in the central region.

7 SUMMARY AND CONCLUSIONS

We have presented the structural parameters, resolved stellar mass distribution and colour gradient of a sample of 36 passive galaxies in the red sequence of the cluster XMMUJ2235-2557 at redshift $z \sim 1.39$. With HST/ACS and WFC3 data we derive light-weighted structural parameters independently in five different bands ($i_{775}, z_{850}, Y_{105}, J_{125}, H_{160}$) through 2D Sérsic fitting. We compute 1D $z_{850} - H_{160}$ colour profiles for individual galaxies and fit logarithmic gradients $\nabla_{z_{850}-H_{160}}$ in the range of PSF HWHM $< a < 3.5a_e$ to derive colour gradients. In addition, we derive resolved stellar mass surface density maps for individual galaxies with an empirical M_*/L-colour relation and the z_{850} and H_{160} images. Mass-weighted structural parameters are derived from the resolved stellar mass surface density map. We find the following:

– From our multi-band light-weighted structural parameter measurements, the passive galaxies in this cluster show a reduction of $\sim 20\%$ in sizes from i_{775} to H_{160}, consistent with the wavelength-dependence found in local passive galaxies.
– The H_{160} band sizes in this cluster are on average $\sim 40\%$ smaller than that expected from the local mass-size relation by Bernardi et al. (2012) at the same rest-wavelength, with a median of $\langle \log(R_{e-circ}/R_{Bernardi}) \rangle = -0.21$. In the extreme cases the galaxies can be $\sim 70\%$ smaller than their local counterparts.
– The mass-weighted sizes of the galaxies are $\sim 41\%$ smaller than their own light-weighted sizes, with a median $\langle \log(R_{e-circ,mass}/R_{e-circ}) \rangle = -0.23$, in the extreme cases the mass-weighted sizes can be up to $\sim 60\%$ smaller.
– 78% of the galaxies in our sample show a negative colour gradient $\nabla_{z_{850}-H_{160}}$, with redder colours at the core and bluer colour in the outskirts. 42% have steep gradients with $\nabla_{z_{850}-H_{160}} < -0.5$. The median colour gradient is $\langle \nabla_{z_{850}-H_{160}} \rangle = -0.45$, two times steeper than the colour gradient found in local passive galaxies in previous studies.
– The ratio of mass-weighted to light-weighted size does not show any significant correlation to galaxy properties, and is only mildly correlated to M_*/L gradient, mass surface density and mass-weighted size. The mild correlation with M_*/L gradients supports our findings about smaller mass-weighted sizes compared to light-weighted sizes.
– By using the local SPIDER sample we find that the mass-weighted sizes are on average $\sim 13\%$ smaller than the rest

frame r-band light-weighted sizes, consistent with previous studies. Comparing the cluster sample to the local SPIDER cluster sample, we find an offset in the ratio of mass-weighted sizes to the H_{160} band light-weighted sizes with a median difference of $\langle \log(R_{e-circ,mass,1.39}/R_{e-circ,1.39}) - \log(R_{e-circ,mass,0}/R_{e-circ,0}) \rangle = -0.18$, which we attribute to an evolution of the M_*/L gradient over redshift. We also find that the progenitor bias cannot explain this observed offset. This also seems to be consistent with the steeper colour gradient we find in the cluster galaxies compared to those seen in local passive galaxies.

We then investigate the origin and the evolution of the observed colour gradient by modeling the colour gradients with SSPs under three different assumptions. We analyse the evolution of the rest-frame (U-R) colour gradient at two radii, $0.5a_e$ and $2a_e$, representing the inner and outer region of the galaxy respectively. We subdivide each of the assumptions summarised below into three different metallicity scenarios: we fix the metallicity of the inner regions to sub-solar, solar and super-solar $Z = (0.008, 0.02, 0.05)$:

– **Case I - Pure age-driven gradient evolution** – Evolution of colour gradients is solely due to age gradient. The inner and outer regions of the passive galaxies are assumed to have identical metallicities (i.e. flat metallicity gradients $\nabla_Z = 0$). We find that although an age gradient alone is sufficient to reproduce all the evolution in the colour gradient, it over-predicts the evolution over redshift, causing the evolved local gradients to be too shallow compared to the observation.
– **Case II - Pure metallicity-driven gradient evolution** – Evolution of colour gradients is solely due to metallicity gradient. We assume the inner and outer regions have identical ages (i.e. flat age gradients $\nabla_{age} = 0$). A metallicity gradient alone cannot explain the observation as the evolution it predicts goes into the wrong direction.
– **Case III - Age-driven gradient evolution with an assumed metallicity gradient** – Evolution of colour gradients is due to a combination of age and metallicity gradients. The galaxies are assumed to have a fixed metallicity gradient identical to that observed in local passive galaxies, $\nabla_Z \simeq -0.2$. This model works well, the solar metallicity scenario can well reproduce the observed evolution of the colour gradients from $z \sim 1.39$ to $z \sim 0$.
– We show that the above findings are still robust if any central dust reddening at $0.5a_e$ is $\lesssim 0.14$ mag, or equivalently an extinction gradient $dA_V/d\log(a) \sim 0.40$ or $dE(B-V)/d\log(a) \sim -0.10$.

Our case study indicates that the presence of an age gradient at high redshift is a necessary condition to explain the observed evolution of the colour gradients, while metallicity gradients probably dominate at $z \sim 0$. We also repeat the study using the rest-frame $g - r$ colour gradient and obtain completely consistent results. This conclusion is partially consistent with other studies (Saglia et al. 2000; Guo et al. 2011; Gargiulo et al. 2012). For the best-matching scenario (Case III with solar metallicity), the median age gradient of our cluster sample is $\langle \nabla_{age} \rangle \sim -0.33 \pm 0.37$, while the metallicity gradient we assumed is $\nabla_Z \simeq -0.2$. Given the assumption that passive galaxies at z = 1.39 have the same metallicity gradient as the local ones, the evolution of colour gradients from $z \sim 1.4$ to $z \sim 0$ can be explained by passive evolution.

This general picture is also consistent with a more gradual mass growth mechanism such as via minor mergers, in the sense

that the inner region of the galaxies remains undisturbed and the accreted younger material settles at the outskirts.

REFERENCES

Andreon S., 2008, MNRAS, 386, 1045

Andreon S., Newman A. B., Trinchieri G., Raichoor A., Ellis R. S., Treu T., 2014, A&A, 565, A120

Barden M., Häußler B., Peng C. Y., McIntosh D. H., Guo Y., 2012, MNRAS, 422, 449

Barden M. et al., 2005, ApJ, 635, 959

Barro G. et al., 2015, ArXiv e-prints

Barro G. et al., 2013, ApJ, 765, 104

Bayliss M. B. et al., 2014, ApJ, 794, 12

Beifiori A. et al., 2014, ApJ, 789, 92

Bell E. F., de Jong R. S., 2001, ApJ, 550, 212

Bell E. F., McIntosh D. H., Katz N., Weinberg M. D., 2003, ApJS, 149, 289

Belli S., Newman A. B., Ellis R. S., 2015, ApJ, 799, 206

Bernardi M., Meert A., Sheth R. K., Vikram V., Huertas-Company M., Mei S., Shankar F., 2013, MNRAS, 436, 697

Bernardi M., Meert A., Vikram V., Huertas-Company M., Mei S., Shankar F., Sheth R. K., 2012, ArXiv e-prints

Berry D., Currie M., Jenness T., Draper P., Bell G., Tilanus R., 2013, in Astronomical Society of the Pacific Conference Series, Vol. 475, Astronomical Data Analysis Software and Systems XXII, Friedel D. N., ed., p. 247

Bertin E., Arnouts S., 1996, A&AS, 117, 393

Bertin E., Mellier Y., Radovich M., Missonnier G., Didelon P., Morin B., 2002, in Astronomical Society of the Pacific Conference Series, Vol. 281, Astronomical Data Analysis Software and Systems XI, Bohlender D. A., Durand D., Handley T. H., eds., p. 228

Bezanson R., van Dokkum P. G., Tal T., Marchesini D., Kriek M., Franx M., Coppi P., 2009, ApJ, 697, 1290

Bezanson R., van Dokkum P. G., van de Sande J., Franx M., Leja J., Kriek M., 2013, ApJL, 779, L21

Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503

Brough S., Proctor R., Forbes D. A., Couch W. J., Collins C. A., Burke D. J., Mann R. G., 2007, MNRAS, 378, 1507

Brown M. J. I., Dey A., Jannuzi B. T., Brand K., Benson A. J., Brodwin M., Croton D. J., Eisenhardt P. R., 2007, ApJ, 654, 858

Brown M. J. I. et al., 2008, ApJ, 682, 937

Bruce V. A. et al., 2012, MNRAS, 427, 1666

Bruzual G., Charlot S., 2003, MNRAS, 344, 1000

Buitrago F., Trujillo I., Conselice C. J., Bouwens R. J., Dickinson M., Yan H., 2008, ApJL, 687, L61

Bundy K., Fukugita M., Ellis R. S., Targett T. A., Belli S., Kodama T., 2009, ApJ, 697, 1369

Burke C., Collins C. A., 2013, MNRAS, 434, 2856

Burke C., Hilton M., Collins C., 2015, MNRAS, 449, 2353

Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682

Cappellari M., 2013, ApJL, 778, L2

Cappellari M., Copin Y., 2003, MNRAS, 342, 345

Carlberg R. G., 1984, ApJ, 286, 403

Carollo C. M. et al., 2013, ApJ, 773, 112

Casertano S. et al., 2000, AJ, 120, 2747

Cassata P. et al., 2010, ApJL, 714, L79

Chabrier G., 2003, PASP, 115, 763

Cimatti A. et al., 2008, A&A, 482, 21

Ciotti L., 1991, A&A, 249, 99

Coccato L., Gerhard O., Arnaboldi M., 2010, MNRAS, 407, L26

Conroy C., Wechsler R. H., Kravtsov A. V., 2007, ApJ, 668, 826

Cooper M. C. et al., 2012, MNRAS, 419, 3018

Daddi E. et al., 2005, ApJ, 626, 680

Damjanov I. et al., 2011, ApJL, 739, L44

Dawson K. S. et al., 2009, AJ, 138, 1271

De Propris R., Bremer M. N., Phillipps S., 2015, MNRAS, 450, 1268

De Propris R., Phillipps S., Bremer M. N., 2013, MNRAS, 434, 3469

Delaye L. et al., 2014, MNRAS, 441, 203

Dressler A., 1980, ApJ, 236, 351

Fan L., Lapi A., Bressan A., Bernardi M., De Zotti G., Danese L., 2010, ApJ, 718, 1460

Fan L., Lapi A., De Zotti G., Danese L., 2008, ApJL, 689, L101

Fang J. J., Faber S. M., Koo D. C., Dekel A., 2013, ApJ, 776, 63

Fassbender R. et al., 2014, A&A, 568, A5

Gargiulo A., Saracco P., Longhetti M., 2011, MNRAS, 412, 1804

Gargiulo A., Saracco P., Longhetti M., La Barbera F., Tamburri S., 2012, MNRAS, 425, 2698

Gobat R. et al., 2011, A&A, 526, A133

Gobat R. et al., 2013, ApJ, 776, 9

Gonzaga S., et al., B, C, 2012, The DrizzlePac Handbook

González Delgado R. M. et al., 2014, A&A, 562, A47

Greene J. E., Murphy J. D., Graves G. J., Gunn J. E., Raskutti S., Comerford J. M., Gebhardt K., 2013, ApJ, 776, 64

Grützbauch R., Bauer A. E., Jørgensen I., Varela J., 2012, MNRAS, 423, 3652

Guo Y. et al., 2011, ApJ, 735, 18

Guo Y. et al., 2009, MNRAS, 398, 1129

Häussler B. et al., 2007, ApJS, 172, 615

Hayashi M., Kodama T., Koyama Y., Tadaki K.-I., Tanaka I., 2011, MNRAS, 415, 2670

Hilz M., Naab T., Ostriker J. P., 2013, MNRAS, 429, 2924

Hirschmann M., Naab T., Ostriker J. P., Forbes D. A., Duc P.-A., Davé R., Oser L., Karabal E., 2015, MNRAS, 449, 528

Huertas-Company M., Shankar F., Mei S., Bernardi M., Aguerri J. A. L., Meert A., Vikram V., 2013, ApJ, 779, 29

Hyde J. B., Bernardi M., 2009, MNRAS, 394, 1978

Ishibashi W., Fabian A. C., Canning R. E. A., 2013, MNRAS, 431, 2350

Jee M. J. et al., 2009, ApJ, 704, 672

Jørgensen I., Chiboucas K., 2013, AJ, 145, 77

Jørgensen I., Chiboucas K., Toft S., Bergmann M., Zirm A., Schiavon R. P., Grützbauch R., 2014, AJ, 148, 117

Keating S. K., Abraham R. G., Schiavon R., Graves G., Damjanov I., Yan R., Newman J., Simard L., 2015, ApJ, 798, 26

Kelly B. C., 2007, ApJ, 665, 1489

Kelvin L. S. et al., 2012, MNRAS, 421, 1007

Kennedy R. et al., 2015, ArXiv e-prints

Kodama T., Arimoto N., 1997, A&A, 320, 41

Kriek M., van Dokkum P. G., Labbé I., Franx M., Illingworth G. D., Marchesini D., Quadri R. F., 2009, ApJ, 700, 221

Krist J., 1995, in Astronomical Society of the Pacific Conference Series, Vol. 77, Astronomical Data Analysis Software and Systems IV, Shaw R. A., Payne H. E., Hayes J. J. E., eds., p. 349

Kuntschner H. et al., 2010, MNRAS, 408, 97

La Barbera F., de Carvalho R. R., 2009, ApJL, 699, L76

La Barbera F., De Carvalho R. R., De La Rosa I. G., Gal R. R., Swindle R., Lopes P. A. A., 2010a, AJ, 140, 1528

La Barbera F., de Carvalho R. R., de La Rosa I. G., Lopes P. A. A., Kohl-Moreira J. L., Capelato H. V., 2010b, MNRAS, 408, 1313

La Barbera F., de Carvalho R. R., Gal R. R., Busarello G., Merluzzi P., Capaccioli M., Djorgovski S. G., 2005, ApJL, 626, L19

La Barbera F., de Carvalho R. R., Kohl-Moreira J. L., Gal R. R., Soares-Santos M., Capaccioli M., Santos R., Sant'anna N., 2008, PASP, 120, 681

La Barbera F., Ferreras I., de Carvalho R. R., Bruzual G., Charlot S., Pasquali A., Merlin E., 2012, MNRAS, 426, 2300

La Barbera F., Lopes P. A. A., de Carvalho R. R., de La Rosa I. G., Berlind A. A., 2010c, MNRAS, 408, 1361

Labbé I. et al., 2003, AJ, 125, 1107

Lang P. et al., 2014, ApJ, 788, 11

Lange R. et al., 2015, MNRAS, 447, 2603

Lani C. et al., 2013, MNRAS, 435, 207

Laporte C. F. P., White S. D. M., Naab T., Gao L., 2013, MNRAS, 435, 901

Larson R. B., 1974, MNRAS, 166, 585

Lasker B. M. et al., 2008, AJ, 136, 735

Lauer T. R. et al., 2005, AJ, 129, 2138

Lidman C. et al., 2008, A&A, 489, 981

Longhetti M. et al., 2007, MNRAS, 374, 614

Lotz J. M. et al., 2013, ApJ, 773, 154

Mancone C. L., Gonzalez A. H., Brodwin M., Stanford S. A., Eisenhardt P. R. M., Stern D., Jones C., 2010, ApJ, 720, 284

Mehlert D., Thomas D., Saglia R. P., Bender R., Wegner G., 2003, A&A, 407, 423

Mei S. et al., 2009, ApJ, 690, 42

Mendel J. T. et al., 2015, ApJL, 804, L4

Moran S. M., Ellis R. S., Treu T., Smith G. P., Rich R. M., Smail I., 2007, ApJ, 671, 1503

Morishita T., Ichikawa T., Kajisawa M., 2014, ApJ, 785, 18

Morishita T., Ichikawa T., Noguchi M., Akiyama M., Patel S. G., Kajisawa M., Obata T., 2015, ArXiv e-prints

Mullis C. R., Rosati P., Lamer G., Böhringer H., Schwope A., Schuecker P., Fassbender R., 2005, ApJL, 623, L85

Naab T., Johansson P. H., Ostriker J. P., 2009, ApJL, 699, L178

Newman A. B., Ellis R. S., Andreon S., Treu T., Raichoor A., Trinchieri G., 2014, ApJ, 788, 51

Newman A. B., Ellis R. S., Bundy K., Treu T., 2012, ApJ, 746, 162

Nipoti C., Londrillo P., Ciotti L., 2003, MNRAS, 342, 501

Oke J. B., Gunn J. E., 1983, ApJ, 266, 713

Oliva-Altamirano P., Brough S., Jimmy, Kim-Vy T., Couch W. J., McDermid R. M., Lidman C., von der Linden A., Sharp R., 2015, MNRAS, 449, 3347

Papovich C. et al., 2012, ApJ, 750, 93

Papovich C., Giavalisco M., Dickinson M., Conselice C. J., Ferguson H. C., 2003, ApJ, 598, 827

Pastorello N., Forbes D. A., Foster C., Brodie J. P., Usher C., Romanowsky A. J., Strader J., Arnold J. A., 2014, MNRAS, 442, 1003

Patel S. G. et al., 2013, ApJ, 766, 15

Peng C. Y., 2015, Technical galfit faq

Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2002, AJ, 124, 266

Pipino A., D'Ercole A., Chiappini C., Matteucci F., 2010, MNRAS, 407, 1347

Pipino A., D'Ercole A., Matteucci F., 2008, A&A, 484, 679

Poggianti B. M. et al., 2013, ApJ, 762, 77

Pu S. B., Saglia R. P., Fabricius M. H., Thomas J., Bender R., Han Z., 2010, A&A, 516, A4

Ragone-Figueroa C., Granato G. L., 2011, MNRAS, 414, 3690

Raskutti S., Greene J. E., Murphy J. D., 2014, ApJ, 786, 23

Reda F. M., Proctor R. N., Forbes D. A., Hau G. K. T., Larsen S. S., 2007, MNRAS, 377, 1772

Rettura A. et al., 2010, ApJ, 709, 512

Rodriguez-Gomez V. et al., 2015, ArXiv e-prints

Rosati P. et al., 2009, A&A, 508, 583

Saglia R. P., Maraston C., Greggio L., Bender R., Ziegler B., 2000, A&A, 360, 911

Saglia R. P. et al., 2010, A&A, 524, A6

Schlafly E. F., Finkbeiner D. P., 2011, ApJ, 737, 103

Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525

Sérsic J. L., 1963, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 6, 41

Shankar F. et al., 2015, ApJ, 802, 73

Shankar F., Marulli F., Bernardi M., Mei S., Meert A., Vikram V., 2013, MNRAS, 428, 109

Simard L., Mendel J. T., Patton D. R., Ellison S. L., McConnachie A. W., 2011, ApJS, 196, 11

Stanford S. A., Eisenhardt P. R., Dickinson M., 1998, ApJ, 492, 461

Stott J. P., Collins C. A., Burke C., Hamilton-Morris V., Smith G. P., 2011, MNRAS, 414, 445

Stott J. P. et al., 2010, ApJ, 718, 23

Strazzullo V. et al., 2013, ApJ, 772, 118

Strazzullo V. et al., 2010, A&A, 524, A17

Szomoru D., Franx M., Bouwens R. J., van Dokkum P. G., Labbé I., Illingworth G. D., Trenti M., 2011, ApJL, 735, L22

Szomoru D., Franx M., van Dokkum P. G., 2012, ApJ, 749, 121

Szomoru D., Franx M., van Dokkum P. G., Trenti M., Illingworth G. D., Labbé I., Oesch P., 2013, ApJ, 763, 73

Tamura N., Ohta K., 2003, AJ, 126, 596

Tanaka M. et al., 2013, ApJ, 772, 113

Taylor E. N., Franx M., Glazebrook K., Brinchmann J., van der Wel A., van Dokkum P. G., 2010, ApJ, 720, 723

Taylor E. N. et al., 2009a, ApJ, 694, 1171

Taylor E. N. et al., 2009b, ApJS, 183, 295

Thomas D., Maraston C., Bender R., Mendes de Oliveira C., 2005, ApJ, 621, 673

Thomas D., Maraston C., Schawinski K., Sarzi M., Silk J., 2010, MNRAS, 404, 1775

Tiret O., Salucci P., Bernardi M., Maraston C., Pforr J., 2011, MNRAS, 411, 1435

Tortora C., Napolitano N. R., Cardone V. F., Capaccioli M., Jetzer P., Molinaro R., 2010, MNRAS, 407, 144

Tortora C., Pipino A., D'Ercole A., Napolitano N. R., Matteucci F., 2013, MNRAS, 435, 786

Treu T., Ellis R. S., Kneib J.-P., Dressler A., Smail I., Czoske O., Oemler A., Natarajan P., 2003, ApJ, 591, 53

Trujillo I., Carrasco E. R., Ferré-Mateu A., 2012, ApJ, 751, 45

Trujillo I., Cenarro A. J., de Lorenzo-Cáceres A., Vazdekis A., de la Rosa I. G., Cava A., 2009, ApJL, 692, L118

Trujillo I., Conselice C. J., Bundy K., Cooper M. C., Eisenhardt P., Ellis R. S., 2007, MNRAS, 382, 109

Trujillo I., Ferreras I., de La Rosa I. G., 2011, MNRAS, 415, 3903

Trujillo I. et al., 2006a, MNRAS, 373, L36

Trujillo I. et al., 2006b, ApJ, 650, 18

Valentinuzzi T. et al., 2010a, ApJ, 712, 226

Valentinuzzi T. et al., 2010b, ApJL, 721, L19

van der Wel A. et al., 2012, ApJS, 203, 24

van der Wel A. et al., 2014, ApJ, 788, 28

van der Wel A., Holden B. P., Zirm A. W., Franx M., Rettura A., Illingworth G. D., Ford H. C., 2008, ApJ, 688, 48

van Dokkum P. G. et al., 2014, ApJ, 791, 45

van Dokkum P. G., Franx M., 2001, ApJ, 553, 90

van Dokkum P. G. et al., 2008, ApJL, 677, L5

van Dokkum P. G. et al., 2010, ApJ, 709, 1018

Vulcani B. et al., 2014, MNRAS, 441, 1340

Weinmann S. M., Kauffmann G., van den Bosch F. C., Pasquali A., McIntosh D. H., Mo H., Yang X., Guo Y., 2009, MNRAS, 394, 1213

Whitaker K. E. et al., 2011, ApJ, 735, 86

White M., Zheng Z., Brown M. J. I., Dey A., Jannuzi B. T., 2007, ApJL, 655, L69

Wilkinson D. M. et al., 2015, MNRAS, 449, 328

Wise M. W., Silva D. R., 1996, ApJ, 461, 155

Wu H., Shao Z., Mo H. J., Xia X., Deng Z., 2005, ApJ, 622, 244

Wuyts S., Cox T. J., Hayward C. C., Franx M., Hernquist L., Hopkins P. F., Jonsson P., van Dokkum P. G., 2010, ApJ, 722, 1666

Wuyts S. et al., 2012, ApJ, 753, 114

Zibetti S., Charlot S., Rix H.-W., 2009, MNRAS, 400, 1181

Zirm A. W., Toft S., Tanaka M., 2012, ApJ, 744, 181

APPENDIX A: DETAILS OF THE SIMULATIONS

We perform extensive simulations with a set of 50000 simulated galaxies with surface brightness profiles described by a Sérsic profile on the ACS z_{850} and WFC3 H_{160} band images to give a more realistic estimate of the uncertainty of the photometry, light-weighted structural parameters and mass-weighted structural parameters. Here we describe the set-up and the procedures involved in these simulations.

A1 Constructing the set of simulated galaxies

The simulated galaxies (hereafter SGs) are uniformly distributed within a magnitude range of $19.0 < H_{160} < 25.0$ and a colour

range of $0.4 < z_{850} - H_{160} < 2.2$ (hence a range of z_{850} magnitudes $19.4 < z_{850} < 27.2$). This selected magnitude range is a good representation of the cluster sample, and is also the range where our M_*/L - colour relation is calibrated as we explained in Section 3.4. Each galaxy is described by a Sérsic profile with input structural parameters randomly drawn from Gaussian distributions with means and dispersions taken from the real galaxies distributions in the H_{160} band.

The input structural parameters are identical in the two bands. The means and dispersions of the Sérsic indices n, effective semi-major axis a_e and axis ratio q are $(\langle n \rangle, \sigma_n) = (3.19, 2.18)$, $(\langle a_e \rangle, \sigma_{a_e}) = (6.07, 5.16(pixel))$ and $(\langle q \rangle, \sigma_q) = (0.67, 0.20)$. The position angle *P.A.* is uniformly distributed within $0° < P.A. < 180°$. To ensure the simulated profiles are physical, we further apply the following constraints: $n > 0.2$, $a_e > 0.3$, $0.01 < q \leqslant 1$.

The SGs were then convolved with the adopted PSFs. Morishita, Ichikawa & Kajisawa (2014) pointed out that there are differences in the central part of the Sérsic profiles produced by `IRAF gallist` and `mkobjects` compared to those produced by GALFIT, which possibly originates with the PSF convolution procedure. In our case, we produce our simulated galaxies using a custom-built IDL routine[1] that over-samples the central part of the Sérsic profiles before resampling it onto a 2D grid.

To check whether the Sérsic profiles we generated are consistent with those used in GALFIT, we first fit the noise-free SGs with GALFIT and examine the residual maps to compare the Sérsic profiles. Without PSF convolution, we notice there are residuals at the centre in the residuals map output by GALFIT, although the difference is negligible ($< 0.005\%$ of the flux). Including the PSF convolution does not noticeably increase this difference.

We then inject the SGs one by one uniformly to the sky regions of both the WFC3 H_{160} images and the PSF-matched and resampled ACS z_{850} images at the same location (i.e. 50000 set of z_{850} and H_{160} images). The segmentation maps from SExtractor are used as a reference to avoid direct overlap with existing objects in the field.

A2 Photometry uncertainty test

We run SExtractor on the SG images using the same setting as for the science sample. We then assess the detection rate of the SGs in different magnitudes, as well as investigate the uncertainties of the galaxy magnitude `MAG_AUTO` and the $z_{850} - H_{160}$ colour derived from $1''$ aperture magnitudes.

A3 Light-weighted structural parameter uncertainty test

We assess the accuracy of our light-weighted structural parameter measurements by measuring the structural parameters of the SGs with the same GALAPAGOS routine.

In Section 3.1.2 we mention there is a bias in the recovered effective radius at high mean surface brightness (< 19 mag arcsec^{-2}) due to unresolved SGs in our simulations. Here we expand the discussion on this. Figure A1 shows the difference between input and recovered structural parameters by GALFIT for three subpopulations of SGs with descending range of input a_e in terms of the size of PSF. We find that the bias between input and recovered effective radii increases sharply for SGs with input $a_e < $ PSF HWHM. For

---

[1] Interactive Data Language, Exelis Visual Information Solutions, Boulder, Colourado
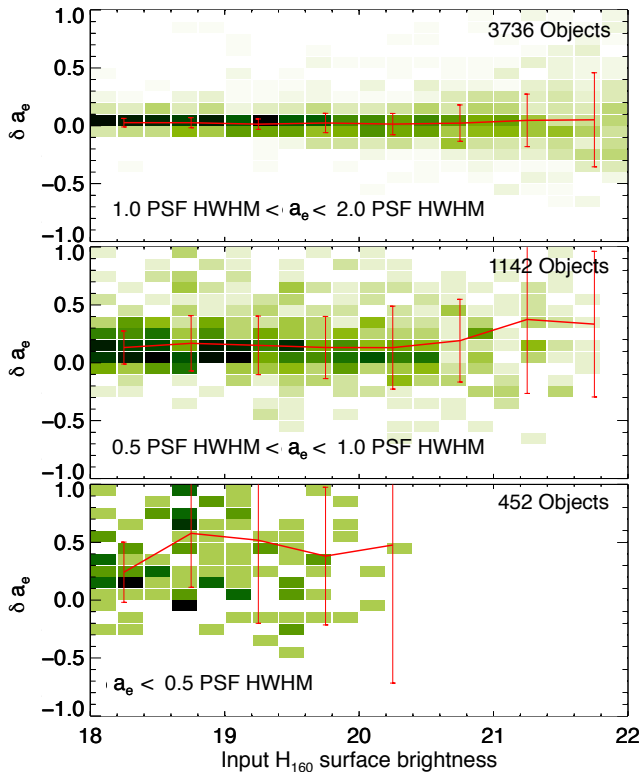
Figure A1. Differences between input and recovered effective semi-major axes by GALFIT $\delta a_e = (a_{e-out} - a_{e-in})/a_{e-in}$ in function of input mean H_{160} surface brightness. From top to bottom: simulated galaxies with different ranges of input effective radius, 1.0 PSF HWHM $< a_e <$ 2.0 PSF HWHM, 0.5 PSF HWHM $< a_e <$ 1.0 PSF HWHM and $a_e <$ 0.5 PSF HWHM. The red line indicates the median and 1σ dispersion in different bins (0.5 mag arcsec^{-2} bin width) and the green-shaded 2D histogram in each panel shows the number density distribution of the simulated galaxies.

SGs with input a_e in range of 1.0 PSF HWHM $< a_e <$ 2.0 PSF HWHM, the average bias is typically limited to $1-2\%$, while it increases to $\sim 10\%$ for SGs with 0.5 PSF HWHM $< a_e <$ 1.0 PSF HWHM. For those with $a_e <$ 0.5 PSF HWHM, the average bias rises sharply to $\sim 50\%$. We conclude that our method is unable to measure sizes reliably from galaxies with $a_e <$ 0.5 PSF HWHM. However, since the sizes of high-redshift galaxies can indeed be very small, we do not exclude this small-sized population from our set of simulated galaxies.

None of the galaxies in our sample have light-weighted sizes smaller than 0.5 PSF HWHM. Nevertheless, in the case of mass-weighted sizes, three of the objects at the low-mass end have sizes < 0.5 PSF HWHM and are hence discarded in the subsequent analyses.

A4 Mass-weighted structural parameter uncertainty test

We also perform a similar test to investigate the biases and uncertainties of our mass-weighted structural parameter measurements. We start with the z_{850} and H_{160} postage stamps output from the light uncertainty test. These images are converted into mass maps with the pipeline described in Section 3.6. The resultant stacked mass maps are then fitted with GALFIT.

As mentioned in Appendix A1, the galaxies we inject have identical initial structural parameters in both z_{850} and H_{160} bands,

there is no internal colour gradient within the set of simulated galaxies. Nevertheless, this allows us to assess the accuracy of the output mass-weighted structural parameters, as the retrieved parameters should be in theory, exactly the same as the input (light) structural parameters. As the Voronoi binning and stacking take a certain time, we perform the mass map conversion for a subsample of galaxies (~ 1500). This sample is sufficient to provide an uncertainty estimates on the mass-weighted sizes in different bins of surface brightness.

In deriving the mass maps, we implement an extrapolation scheme to determine the M_*/L in regions with insufficient signal-to-noise (described in Section 3.6.2). The low S/N or sky regions are problematic as the colours (and hence M_*/L) are not well determined. Converting mass directly on these regions will induce a huge scatter of mass in the background, which in turn have serious effects on the structural parameter measurements. Our extrapolation scheme can preserve the sky noise and at the same time provide a reasonable M_*/L estimate to these regions. We illustrate this effect in Figure A2. The top panel shows the differences between input and recovered sizes by GALFIT with extrapolation, while the bottom panel shows the differences without applying the extrapolation. In the absence of extrapolation, a huge bias can be seen in all bins of surface brightness. Sizes are more underestimated in galaxies with low surface brightness.

Previous studies use a different method to solve this issue, for example in Lang et al. (2014) who derived mass maps for galaxies in CANDELS, these low S/N regions are assigned the average M_*/L of the three nearest Voronoi bins. The data we used in this study is not as deep as the CANDELS HST imaging. We find that averaging the nearest three Voronoi bins does not work as well as our annular average extrapolation. For star-forming galaxies which have sub-structures such as star forming clumps and spiral arms (see, e.g. Wuyts et al. 2012), it might be more suitable to use the nearest neighbour extrapolation as in Lang et al. (2014). Our method, on the other hand, works well for early-type galaxies which usually have smooth(er) surface brightness profiles.

APPENDIX B: DETAILS OF PSF MATCHING

We first stack the unsaturated stars in the z_{850} and H_{160} bands to obtain characteristic PSFs respectively. The matching kernel is generated using the `psfmatch` task in IRAF. Cosine bell tapering is applied to filter the high frequency component of the input z_{850} PSF, which is presumably induced by noise, to clean the output kernel. In the `psfmatch` task, there are some free parameters that can be tweaked (e.g. kernel sizes, highest cosine bell frequencies and apodize), a systemic search is performed to find the best parameters to match the PSFs.

We assess the accuracy by comparing the fractional encircled energy of the z_{850} PSF before and after the procedure to the H_{160} PSF. The convolved z_{850} PSF matches almost perfectly to the H_{160} PSF with only tiny difference in the wing ($< 1\%$). We also reconstructed a new z_{850} PSF from the PSF matched z_{850} images to assess the result. Figure B1 shows the fractional encircled energy of the PSFs constructed from images before and after PSF matching. The ratios of their growth curves deviate by $< 2.5\%$ from unity.
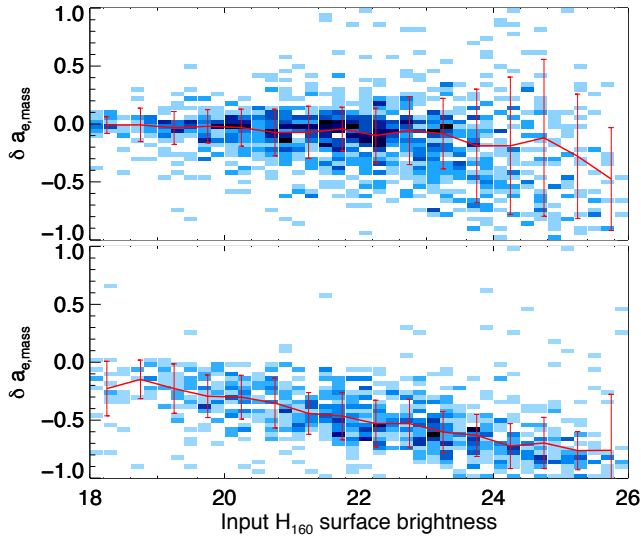
Figure A2. Differences between input and recovered effective semi-major axes by GALFIT $\delta a_e = (a_{e-out} - a_{e-in})/a_{e-in}$ in function of input mean H_{160} surface brightness. Top: sizes derived from mass maps with the extrapolation applied, same as the middle panel in Figure 8. Bottom: sizes derived from mass maps without applying the extrapolation scheme. Red line indicates the median and 1σ dispersion in different bins (0.5 mag arcsec^{-2} bin width) and blue-shaded 2D histogram in each panel shows the number density distribution of the simulated galaxies.
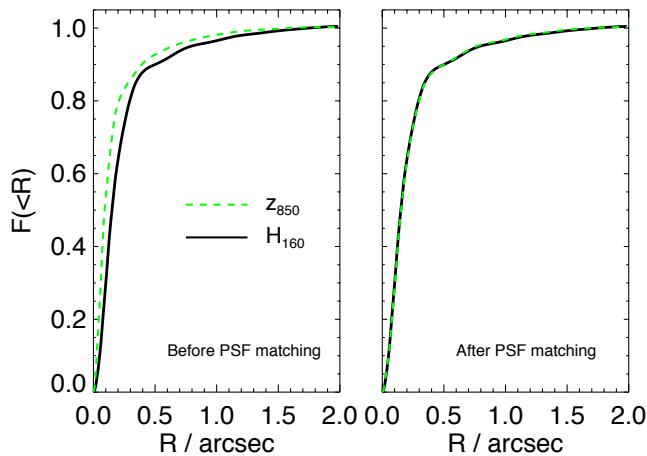

Figure B1. Fractional encircled energy of the z_{850} and H_{160} PSFs. The green dashed line corresponds to the z_{850} PSF while the black solid line corresponds to the H_{160} PSF. Left: before PSF matching. Right: after PSF matching.

APPENDIX C: COMPARISON OF MASSES DERIVED USING M_*/L - COLOUR RELATION WITH SED FITTING

In this study, the stellar masses of the galaxies are derived from an empirical M_*/L - colour relation and the total H_{160} luminosity from 2D GALFIT Sérsic fitting. Other studies usually estimate the stellar masses through spectral energy distribution fitting of multiple photometric bands (e.g. Strazzullo et al. 2010; Delaye et al. 2014). The advantage of using M_*/L - colour relation over SED fitting is that it does not require a number of photometric bands, hence is a relatively inexpensive mass indicator. The accuracy of the stellar mass estimates of then depends on how well constrained the M_*/L
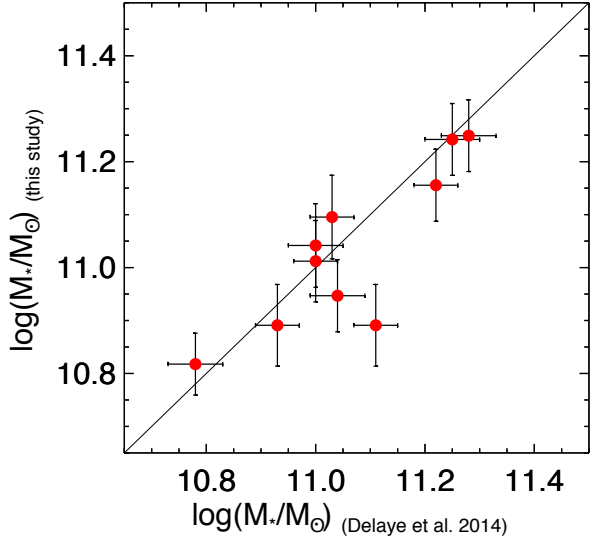

Figure C1. Comparison of masses derived using M_*/L - colour relation with SED fitting from Delaye et al. (2014). The red circles are the stellar mass estimates of 10 galaxies that are common in our sample to Delaye et al. (2014). The stellar masses in Delaye et al. (2014) are derived from SED fitting with four bands, while our masses are from M_*/L - colour relation. The solid black line is the one-to-one relation. The error bars are the 1σ uncertainties from both methods.

- colour relation is, which in turn depends on the colour used (see discussion in e.g., Bell & de Jong 2001; Bell et al. 2003). Here we assess whether our mass estimates is biased.

For this cluster, Delaye et al. (2014) estimated stellar masses of our galaxies through SED fitting with four bands (HST/ACS i_{775}, z_{850}, HAWK-I, J, K_s) with BC03 models, exponential declining SFHs and a Chabrier (2003) IMF. The setting is almost identical to the stellar masses from the NMBS catalogue we picked to construct the M_*/L - colour relation, thus can be compared directly.

Figure C1 shows a direct comparison of the mass estimated using our M_*/L - colour relation and SED fitting from Delaye et al. (2014). Our sample covers 10 out of 13 early-type galaxies in their sample. The remaining 3 galaxies are out of the field of view of our WFC3 images (but are in the FOV of the ACS z_{850} image), thus are not included in our sample. The mass estimates from the two methods are consistent with each other, with a median difference and 1σ scatter of 0.03 ± 0.09 dex. The object that deviates from the one-to-one relation the most (at $\log(M_*/M_\odot) = 11.11$) is a galaxy close to the core of cluster with a very close neighbour (ID 368), which probably affect the mass estimates in both methods. Removing this object reduces the median difference to 0.01 ± 0.07 dex. Therefore we conclude that the masses derived using M_*/L - colour relation are not biased.

As we mentioned in Section 3.5, the typical uncertainty of the mass estimates from M_*/L - colour relation is ~ 0.1 dex, which is comparable to the uncertainties obtained from SED fitting. These uncertainties are correct for the relative masses of multiple galaxies. Note that the uncertainty of the absolute stellar masses is larger as in the case of SED fitting, depending on the details of NMBS SED fitting and choice of IMF.

Table D1. The derived median age gradient at $z = 1.39$, $z = 0$ and formation redshift with exponentially declining τ-models (Case III)

τ	∇_{age} at $z = 1.39$	Evolved ∇_{age} at $z = 0$	Formation redshift z
SSP	-0.33 ± 0.37	-0.05 ± 0.06	3.0
0.2	-0.29 ± 0.29	-0.04 ± 0.06	3.5
0.4	-0.21 ± 0.22	-0.04 ± 0.04	4.0
0.6	-0.19 ± 0.19	-0.05 ± 0.04	7.0

APPENDIX D: EVOLUTION OF COLOUR GRADIENTS USING EXPONENTIALLY DECLINING TAU MODELS

In Section 6.3 we discuss the evolution of colour gradients assuming the stellar population in the inner and outer regions can be well described by simple stellar population models (SSPs). Here we present the result of using exponentially declining tau models with different τ ($\tau = [0.2, 0.4, 0.6]$) instead of SSPs.

Figure D1 shows the $U - R$ colour at different ages (left, i.e. the colour-age relations) from models with various τ using BC03, similar to Figure 15. The methodology is described in Section 6.3.1. Comparing to SSPs, due to the continual star formation the $(U - R)$ colour is bluer for a given age, which is more pronounced if τ is larger. We stop at $\tau = 0.6$ as otherwise the $(U - R)$ colour would be too blue to account for the observations. Using τ models also has the effect of reducing the colour differences between different metallicities when the galaxy is young (see the trend before the grey dashed line), while the evolution is very similar for later ages when the contribution of young stars falls off.

Using τ models instead of SSPs does not change our main conclusion. Since the evolution of the colour gradient at later times is very similar to SSPs, we find that an age-driven gradient evolution with a metallicity gradient close to the local value (case III) remains the best scenario to explain the colour gradients independent of the τ used. Figure D2 shows the best-fit scenario of the evolution of the colour gradients with different τ (case III with solar metallicity). Within the range of τ the evolution can still be well modelled. Nevertheless, due to the change in $(U - R)$ colour over time (see Figure D1), the resulting age gradient, age difference of the inner and outer population and formation redshift for the best-fit scenario varies by a certain amount with τ. With an increasing τ, a flatter but still significant age gradient (e.g. $\langle \nabla_{age} \rangle = -0.21$ for $\tau = 0.4$) is needed to explain the colour gradient. This strengthens our result that an age gradient is a necessary component in the colour gradient at high-redshift. The values of the median age gradient, the evolved age gradient at $z = 0$ and the formation redshift for the best-fit scenario with different τs can be found in Table D1.

APPENDIX E: $g - r$ COLOR GRADIENTS AND THEIR EVOLUTION WITH REDSHIFT

Here we present the measurements of the $Y_{105} - H_{160}$ colour gradient in XMMUJ2235-2557 and the $g - r$ colour gradient in the local SPIDER cluster sample. To compute the $g - r$ colour gradient from the SPIDER sample, we again make use of the structural parameters in g-band and r-band of the publicly available multiband structural catalogue from La Barbera et al. (2010c). We generate 2D Sérsic model images in the two bands with fitted parameter from the structural catalogue, then convert the 2D image in both bands into 1D radial surface brightness profiles, similar to the procedure

described in Section 3.3. This allows us to derive 1D $g - r$ colour profiles and measure the colour gradients of the galaxies in the SPIDER sample by fitting the logarithmic slope of their $g - r$ profiles along the major axis. The sample is split into low density and high density environment with a halo mass cut ($\log(M_{200}/M_{\odot}) < 14$ and $\log(M_{200}/M_{\odot}) \geqslant 14$). The detail selection is described in Section 4. We also apply the age cut (age > 8.98 Gyr) using age measurements from La Barbera et al. (2010b) to correct for the progenitor bias in the SPIDER sample. The median $g - r$ gradient and 1σ scatter in the local SPIDER cluster sample is $\nabla_{g-r} = -0.042 \pm 0.144$ (error on the median 0.008), while the median gradient in the low density sample is $\nabla_{g-r} = -0.060 \pm 0.158$ (error on the median 0.008), consistent with La Barbera et al. (2005).

We derive the $Y_{105} - H_{160}$ colour gradient in XMMUJ2235-2557 with structural parameters of the Y_{105} and H_{160} bands. This is because the above $g - r$ colour gradients are intrinsic gradients without PSF convolution, hence for better comparison and consistency we use the same method as above. Figure E1 shows the $Y_{105} - H_{160}$ colour gradients, roughly corresponds to rest-frame $g - r$. $\nabla_{Y_{105}-H_{160}}$ is less steep compared to $\nabla_{z_{850}-H_{160}}$, with a median and 1σ scatter of $\langle \nabla_{Y_{105}-H_{160}} \rangle = -0.16 \pm -0.16$ (error on the median 0.08). Hence, the $g - r$ colour gradient at $z = 1.39$ is also much steeper than the local sample.

We repeat the analysis described in Section 6.3.1 to model the $g - r$ colour gradients under different assumptions in the radial variation of stellar population properties. The $g - r$ colour is less sensitive to age variation than $(U - R)$ colour, as the g-band is on the 4000Å break. Hence, the evolution in the $g - r$ colour gradient is less pronounced than the $(U - R)$ gradients.

Figure E2 shows the evolution of the $g - r$ colour gradients (case III with solar metallicity). Despite the lack in dynamic range, the result with the $g - r$ colour gradient is completely consistent with the $(U - R)$ colour gradients, in the sense that an age-driven gradient evolution with a metallicity gradient close to local value (case III) is the best scenario to explain its evolution. A pure age gradient would predict $g - r$ gradients that are too shallow at $z = 0$, while a pure metallicity gradient would predict gradients that are too steep. In addition, we find that the derived median age gradient is in good agreement with the one derived from $(U - R)$ gradients. The evolution can be well described with an age gradient of $\nabla_{age} = -0.33$, identical to the one we found from $(U - R)$ gradients. The consistent result from $g - r$ colour gradients reinforces our conclusion that age gradient is necessary to explain the colour gradient at high-redshift.

APPENDIX F: GALAXY PROPERTIES AND BEST-FIT LIGHT WEIGHTED AND MASS-WEIGHTED STRUCTURAL PARAMETERS
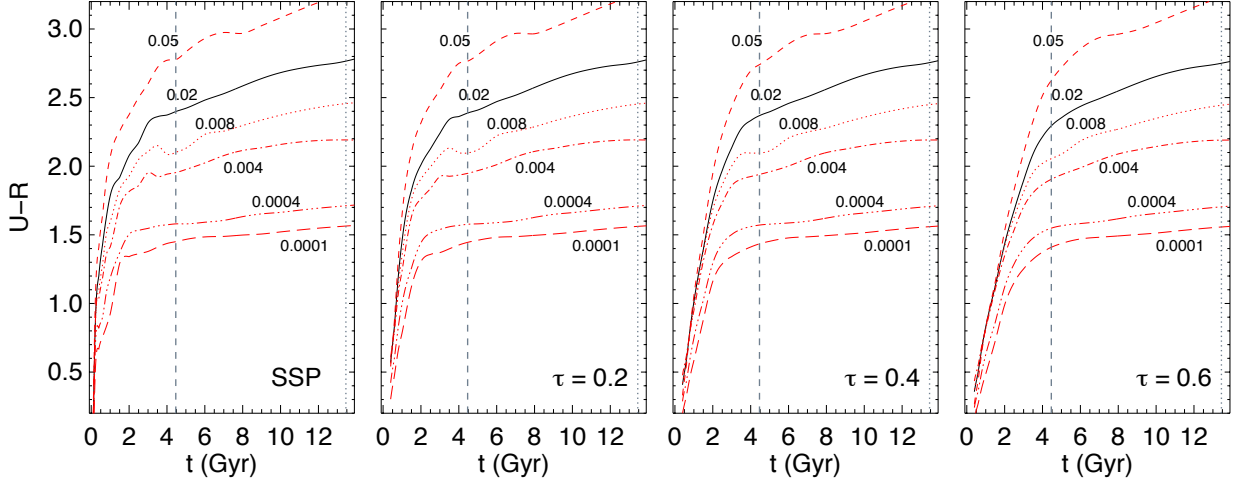
Figure D1. Rest-frame $U - R$ colour of stellar populations with different ages and metallicities with different values of τs (from left to right: SSP, $\tau = 0.2, 0.4, 0.6$). The black line shows the stellar populations with solar metallicity ($Z = 0.02$). The red lines show populations with different metallicities ($Z = 0.0001, 0.0004, 0.004, 0.008, 0.05$) as indicated. The grey dotted line shows the current age of the Universe (13.45 Gyr) with our choice of cosmology and the grey dashed line shows the age of Universe at redshift 1.39 (4.465 Gyr). Note that age refers to the time passed when the populations start forming stars.
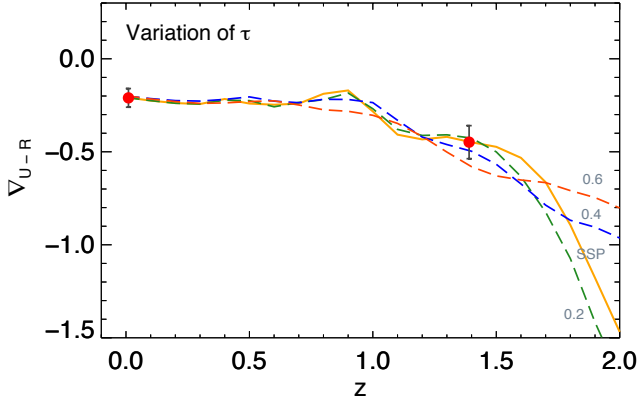

Figure D2. Evolution of colour gradient over redshift in case III (age and assumed metallicity gradient $\nabla_Z = -0.2$) using SSP and tau models with $\tau = 0.2, 0.4, 0.6$ with assumed solar metallicity $Z = 0.02$. The solid yellow line shows the result with SSP models and is identical to one in the top left panel of Figure 17. Green, blue and red lines show the result with tau models with $\tau = 0.2, 0.4, 0.6$ respectively. The derived initial age gradient at $z = 1.39$ as well as formation redshift for each τ can be found in Table D1. Red circle corresponds to the median $z_{850} - H_{160}$ gradient of our sample at redshift 1.39, and the observed $(U - R)$ gradient at redshift 0 by Wu et al. (2005). The error bars show the uncertainty of the median.
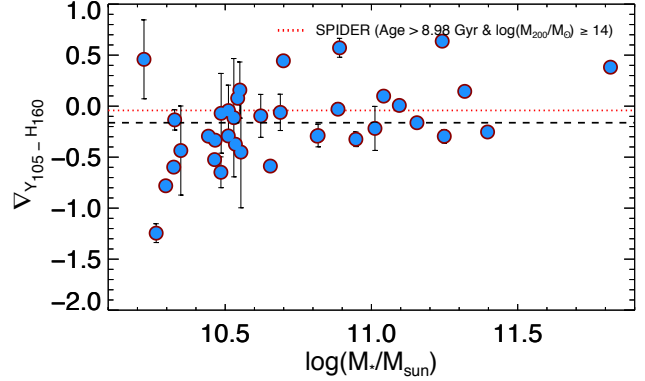

Figure E1. $Y_{105} - H_{160}$ colour gradients in the cluster XMMUJ2235-2557. At redshift 1.39, this roughly corresponds to rest-frame $g - r$ colour gradient. The red dotted line shows the median local $g - r$ gradient from the SPIDER cluster sample. The black dashed line shows the median $Y_{105} - H_{160}$ colour gradients.
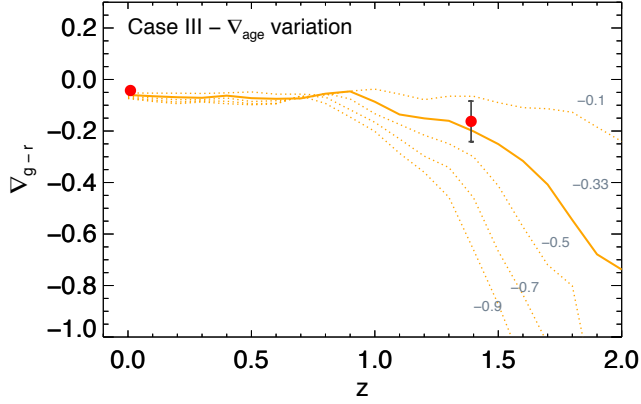
Figure E2. Evolution of $g - r$ colour gradient over redshift in case III (age and metallicity gradient) with assumed solar metallicity $Z = 0.02$. The inner region is assumed to have a formation redshift $z_{form} = 3.0$. The initial age gradient at $z = 1.39$ is $\nabla_{age} = -0.33$, and the assumed metallicity gradient is $\nabla_Z = -0.2$. The dotted lines show the evolution with different initial age gradient as indicated ($\nabla_{age} = -0.1, -0.5, -0.7, -0.9$). Red circles correspond to the median $Y_{105} - H_{160}$ gradient of our sample at redshift 1.39, and the observed local $g - r$ gradient from the SPIDER cluster sample. The error bars show the uncertainty of the median.

Table F1. Galaxy parameters

ID	R.A. (J2000)	Decl. (J2000)	$\log M_*$[a] $(M_\odot)$	$z_{850}-H_{160}$[b] (AB mag)	$\nabla_{z_{850}-H_{160}}$[c]	$\nabla_{\log(M/L)}$[c]	a_e (kpc)	n	q	$a_{e,mass}$ (kpc)	n_{mass}	q_{mass}	Σ[d] (mag kpc^{-2})	$\log(\Sigma_{mass})$[d] $(M_\odot$ kpc$^{-2})$	$\log(\Sigma_1)$[e] $(M_\odot$ kpc$^{-2})$
36	338.829468	−25.974178	11.04±0.08	1.765±0.010	−0.033±0.096	−0.038±0.220	4.12±0.41	4.77±0.48	0.79±0.03	3.66±0.85	7.43±1.11	0.79±0.05	25.87±0.09	9.12±0.20	9.74±0.03
61	338.815826	−25.971949	10.44±0.05	1.359±0.018	−0.540±0.145	−0.386±0.296	5.16±1.08	0.61±0.12	0.69±0.06	3.64±1.07	2.54±0.73	0.52±0.06	27.23±0.18	8.52±0.25	9.13±0.10
77	338.818085	−25.972015	10.51±0.08	1.844±0.037	−1.190±0.154	−0.576±0.309	5.47±1.48	0.64±0.17	0.55±0.06	3.65±1.12	1.24±0.42	0.49±0.08	27.94±0.23	8.59±0.27	9.10±0.14
148	338.833374	−25.967278	10.53±0.07	1.527±0.024	−0.740±0.017	−0.470±0.015	9.27±3.22	0.65±0.21	0.18±0.03	7.03±3.44	1.73±0.83	0.17±0.03	28.53±0.30	8.04±0.42	9.02±0.18
159	338.825348	−25.968397	10.30±0.07	1.259±0.015	−1.199±0.048	−0.719±0.105	4.09±0.68	1.20±0.19	0.64±0.04	1.97±0.47	3.42±0.68	0.73±0.07	26.93±0.15	8.91±0.21	9.31±0.06
170	338.836761	−25.961046	11.82±0.07	1.956±0.009	−0.133±0.038	−0.129±0.094	24.62±6.62	4.49±1.18	0.62±0.07	8.39±2.56	3.91±1.31	0.64±0.10	28.11±0.23	9.17±0.26	10.00±0.21
198	338.824432	−25.936956	10.22±0.05	1.377±0.025	−0.626±0.585	−0.499±0.995	1.02±0.07	2.05±0.28	0.86±0.05	0.71±0.10	1.09±0.15	0.34±0.02	24.28±0.06	9.72±0.13	9.58±0.02
220	338.845001	−25.940239	10.82±0.06	1.694±0.013	−0.482±0.190	−0.342±0.348	2.34±0.19	4.91±0.50	0.71±0.03	1.53±0.27	8.64±1.69	0.37±0.02	25.09±0.07	9.65±0.15	9.98±0.06
239	338.824738	−25.942131	10.51±0.06	1.699±0.018	−0.365±0.314	−0.212±0.587	1.37±0.10	4.00±0.43	0.84±0.03	0.90±0.15	8.38±1.72	0.55±0.04	24.71±0.06	9.80±0.14	9.81±0.04
296	338.839996	−25.957035	10.54±0.05	1.377±0.011	−0.392±0.110	−0.259±0.220	3.00±0.28	1.37±0.14	0.59±0.02	2.20±0.32	2.07±0.30	0.48±0.03	25.83±0.08	9.06±0.20	9.53±0.05
308	338.832703	−25.957813	10.89±0.05	1.360±0.006	−0.426±0.098	−0.310±0.204	2.99±0.19	3.99±0.34	0.79±0.02	1.50±0.24	4.40±0.67	0.72±0.04	24.93±0.06	9.74±0.14	9.92±0.04
343	338.840668	−25.959299	10.55±0.08	1.844±0.022	−0.371±0.558	−0.189±0.887	1.18±0.07	2.38±0.24	0.39±0.01	1.23±0.18	0.89±0.15	0.13±0.02	24.51±0.05	9.57±0.13	9.85±0.03
352	338.836304	−25.962229	11.24±0.07	1.919±0.009	−0.287±0.086	−0.240±0.157	4.58±0.45	4.70±0.47	0.57±0.02	3.42±0.78	6.86±1.02	0.37±0.02	25.84±0.08	9.38±0.20	10.23±0.06
357	338.829987	−25.959751	10.33±0.05	1.436±0.014	0.117±0.124	0.105±0.254	2.38±0.22	1.19±0.13	0.40±0.02	1.92±0.44	1.78±0.27	0.26±0.02	25.96±0.08	8.96±0.20	9.47±0.04
365	338.833984	−25.960121	10.53±0.08	1.847±0.021	−0.158±0.612	−0.111±0.964	1.38±0.10	3.98±0.47	0.53±0.02	–	–	–	24.91±0.06	–	–
368	338.837128	−25.959915	10.89±0.08	1.869±0.013	−0.906±0.184	−0.491±0.401	2.41±0.24	5.78±0.61	0.89±0.03	–	–	–	25.25±0.09	–	–
385	338.837341	−25.959743	10.62±0.08	1.735±0.013	−0.619±0.213	−0.418±0.356	1.73±0.11	3.00±0.28	0.35±0.01	1.65±0.27	1.92±0.31	0.29±0.02	24.99±0.06	9.39±0.14	9.73±0.04
407	338.836243	−25.960421	11.25±0.07	1.915±0.008	−0.448±0.108	−0.272±0.206	3.03±0.20	4.85±0.43	0.70±0.02	1.53±0.25	3.62±0.56	0.60±0.03	24.92±0.06	10.08±0.14	10.25±0.04
433	338.829346	−25.964228	10.95±0.07	1.915±0.013	−0.299±0.150	−0.193±0.266	2.76±0.23	4.10±0.40	0.70±0.02	1.72±0.35	4.29±0.67	0.55±0.03	25.47±0.07	9.68±0.18	10.00±0.04
478	338.853668	−25.943596	11.16±0.07	1.955±0.013	−0.649±0.099	−0.339±0.225	5.14±0.67	5.61±0.69	0.78±0.03	2.41±0.46	5.53±0.75	0.73±0.04	26.37±0.11	9.59±0.16	10.06±0.05
516	338.840942	−25.952995	10.32±0.07	1.292±0.011	−0.686±0.130	−0.313±0.248	2.33±0.22	0.94±0.09	0.77±0.03	1.89±0.43	1.74±0.26	0.69±0.04	25.69±0.08	8.97±0.20	9.35±0.05
534	338.840820	−25.953827	10.26±0.07	1.527±0.021	−1.536±0.203	−0.924±0.419	2.51±0.30	1.08±0.13	0.84±0.04	1.25±0.24	1.67±0.25	0.64±0.04	26.37±0.10	9.27±0.16	9.39±0.04
538	338.831543	−25.945869	10.49±0.06	1.649±0.018	−0.846±0.296	−0.643±0.566	1.65±0.14	5.37±0.63	0.82±0.03	–	–	–	25.10±0.07	–	–
552	338.838593	−25.953201	10.46±0.05	1.426±0.026	−0.777±0.092	−0.464±0.217	6.69±1.80	0.51±0.14	0.80±0.09	5.21±1.59	0.84±0.28	0.69±0.11	27.84±0.23	8.23±0.26	8.52±0.16
558	338.839447	−25.949474	11.10±0.08	1.753±0.009	0.562±0.133	0.346±0.306	4.93±0.49	4.77±0.48	0.73±0.03	–	–	–	26.11±0.09	–	–
562	338.859161	−25.945955	11.32±0.07	1.903±0.008	0.231±0.116	−0.040±0.271	3.66±0.32	5.90±0.57	0.92±0.03	2.91±0.60	9.04±1.41	0.97±0.05	25.14±0.08	9.59±0.18	10.33±0.07
571	338.857452	−25.946079	10.35±0.05	1.480±0.016	0.194±0.518	0.097±0.688	1.16±0.08	6.44±0.84	0.68±0.03	–	–	–	24.42±0.06	–	–
576	338.841461	−25.949100	11.01±0.08	1.803±0.008	−0.417±0.320	−0.252±0.447	2.18±0.11	2.97±0.24	0.36±0.01	1.47±0.21	4.13±0.57	0.21±0.01	24.62±0.04	9.88±0.13	10.20±0.03
585	338.856934	−25.949547	10.49±0.08	1.896±0.023	−0.494±0.483	−0.281±0.762	1.05±0.06	2.63±0.28	0.58±0.03	–	–	–	24.50±0.05	–	–
588	338.830658	−25.948841	10.81±0.08	1.764±0.015	−0.605±0.035	−0.384±0.084	5.42±0.91	1.98±0.31	0.35±0.02	3.65±0.86	2.38±0.47	0.27±0.03	27.04±0.15	8.89±0.21	9.65±0.07
599	338.856018	−25.947937	11.40±0.07	1.915±0.008	−0.451±0.076	−0.274±0.148	7.05±0.86	4.84±0.57	0.52±0.02	3.50±0.66	5.69±0.75	0.45±0.03	26.39±0.11	9.51±0.16	10.27±0.05
611	338.857452	−25.949520	10.47±0.07	1.276±0.012	0.324±0.141	0.280±0.325	4.70±0.80	1.95±0.31	0.83±0.05	2.77±0.66	1.14±0.23	0.93±0.09	26.84±0.15	8.78±0.21	9.03±0.10
617	338.858368	−25.948902	10.70±0.07	1.515±0.016	−0.610±0.094	−0.455±0.193	6.86±1.47	2.17±0.44	0.61±0.05	2.77±0.82	4.19±1.23	0.78±0.09	27.45±0.19	9.02±0.26	9.46±0.10
618	338.823639	−25.948795	10.65±0.08	1.837±0.023	−1.025±0.128	−0.613±0.278	3.37±0.42	2.17±0.27	0.79±0.04	1.54±0.29	2.50±0.35	0.64±0.04	26.52±0.11	9.48±0.16	9.64±0.04
637	338.844788	−25.951603	10.69±0.05	1.486±0.008	−0.325±0.223	−0.273±0.403	1.46±0.06	2.58±0.23	0.82±0.02	0.70±0.10	4.86±0.77	0.80±0.03	24.07±0.04	10.20±0.12	9.97±0.03
642	338.842316	−25.951626	10.55±0.08	1.785±0.019	−0.341±0.289	−0.149±0.445	1.71±0.14	1.96±0.20	0.38±0.02	1.25±0.25	1.45±0.23	0.34±0.02	25.22±0.07	9.56±0.18	9.74±0.03

Notes:

[a] Total stellar masses are estimated using the M_*/L-colour relation, the $z_{850}-H_{160}$ aperture colours and the total luminosity $L_{H_{160}}$ from the best-fit Sérsic models.

[b] $1''$ aperture magnitudes from SExtractor.

[c] Colour gradients $\nabla_{z_{850}-H_{160}}$ and M_*/L gradients $\nabla_{\log(M/L)}$ are defined as $d(z_{850}-H_{160})/d\log(a)$ and $d(\log(M/L))/d\log(a)$ respectively. The gradient is fitted in the radial range of PSF HWHM $< a < 3.5\,a_e$.

[d] Mean surface brightness Σ and mean surface mass density Σ_{mass} are defined as mag $+2.5\log(2\pi a_e^2)$ and $\log(M_*/2\pi a_{e,mass}^2)$.

[e] Mean surface mass density within a radius of 1 kpc Σ_1 is defined as $\log(M_*(<1\text{kpc})/\pi(1\text{kpc})^2)$, derived from integrating the fitted 2D mass profiles.