

# The formation of bulges, discs and two-component galaxies in the CANDELS Survey at $z < 3$

Berta Margalef-Bentabol,<sup>1</sup>★ Christopher J. Conselice,<sup>1</sup> Alice Mortlock,<sup>2</sup> Will Hartley,<sup>3</sup> Kenneth Duncan,<sup>4</sup> Harry C. Ferguson,<sup>5</sup> Avishai Dekel<sup>6</sup> and Joel R. Primack<sup>7</sup>

<sup>1</sup>University of Nottingham, School of Physics and Astronomy, Nottingham NG7 2RD, UK

<sup>2</sup>SUPA Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK

<sup>3</sup>ETH Zurich, Institute für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany

<sup>4</sup>Leiden Observatory, Leiden University, NL-2300 RA Leiden, the Netherlands

<sup>5</sup>Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA

<sup>6</sup>Racah Institute of Physics, The Hebrew University, Jerusalem 91904, Israel

<sup>7</sup>Physics Department, University of California, Santa Cruz, CA 95064, USA

Accepted 2016 June 15. Received 2016 June 3; in original form 2016 January 29

## ABSTRACT

We examine a sample of 1495 galaxies in the CANDELS fields to determine the evolution of two-component galaxies, including bulges and discs, within massive galaxies at the epoch  $1 < z < 3$  when the *Hubble* sequence forms. We fit all of our galaxies' light profiles with a single Sérsic fit, as well as with a combination of exponential and Sérsic profiles. The latter is done in order to describe a galaxy with an inner and an outer component, or bulge and disc component. We develop and use three classification methods (visual, *F*-test and the residual flux fraction) to separate our sample into one-component galaxies (disc/spheroid-like galaxies) and two-component galaxies (galaxies formed by an 'inner part' or bulge and an 'outer part' or disc). We then compare the results from using these three different ways to classify our galaxies. We find that the fraction of galaxies selected as two-component galaxies increases on average 50 per cent from the lowest mass bin to the most massive galaxies, and decreases with redshift by a factor of 4 from  $z = 1$  to 3. We find that single Sérsic 'disc-like' galaxies have the highest relative number densities at all redshifts, and that two-component galaxies have the greatest increase and become at par with Sérsic discs by  $z = 1$ . We also find that the systems we classify as two-component galaxies have an increase in the sizes of their outer components, or 'discs', by about a factor of 3 from  $z = 3$  to 1.5, while the inner components or 'bulges' stay roughly the same size. This suggests that these systems are growing from the inside out, whilst the bulges or protobulges are in place early in the history of these galaxies. This is also seen to a lesser degree in the growth of single 'disc-like' galaxies versus 'spheroid-like' galaxies over the same epoch.

**Key words:** galaxies: evolution – galaxies: high-redshift – galaxies: structure.

## 1 INTRODUCTION

Galaxy structure and morphology are important observables in order to both describe galaxies fully, as well as a critical property for understanding how galaxies form and evolve through cosmic time. We observe that, in the local Universe, most massive galaxies are classifiable into *Hubble* types, i.e. with a well-defined structure, such as spheroids or spirals. However, at higher redshift, a population of peculiar galaxies dominates in terms of number den-

sities, (e.g. Conselice, Blackburne & Papovich 2005; Mortlock et al. 2013). In particular, it is found that the majority of galaxies at  $z > 2$  are peculiar with a smaller number of spheroid-like galaxies (e.g. Mortlock et al. 2013), and with very few traditional disc galaxies. At lower redshifts, we find a gradual transition between peculiar and *Hubble*-type galaxies with an  $\sim 50 : 50$  split between peculiars and *Hubble* types at  $z \sim 1.5$  (Conselice et al. 2005; Mortlock et al. 2013; Huertas-Company et al. 2016).

Uncovering the internal processes involved in changing the morphology and structures of galaxies is therefore a useful way to understand how galaxies evolve in terms of physical processes such as star formation and merging. One of the traditional ways of doing

\*E-mail: Berta.Margalef@nottingham.ac.uk

this is to examine the effective radius and Sérsic index of galaxy populations to determine how they evolved (e.g. Ferguson et al. 2004; Daddi et al. 2005; Trujillo et al. 2006, 2007; Toft et al. 2007; Buitrago et al. 2008; van Dokkum et al. 2010; Cassata et al. 2011). For instance, Buitrago et al. (2008) and others studied the size evolution of massive galaxies showing that, while at  $z > 1$  these galaxies are extremely compact, in the local Universe, we observe that their counterparts are larger, so there must have been a growth in physical size over cosmic time at a given mass. These findings have been confirmed and expanded upon by many others since in great detail with many explanations for the evolution (e.g. Barro et al. 2013; van der Wel et al. 2014; van Dokkum et al. 2014). However, this only tells part of the story, as at the same time, these galaxies grow in size, they also become less peculiar, and develop into bulge+disc systems, something that simple Sérsic fitting cannot fully quantify.

To comprehend how galaxies make the transition to become the galaxies we observe in the nearby Universe, it is especially important to study them at high redshift ( $z > 1$ ) when they are undergoing these transformations, and to do so in a wavelength which probes the underlying stellar mass of the system. Since one of the major hallmarks of the *Hubble* sequence is the bulge and disc dichotomy, a natural next step in understanding the evolution of galaxies and their structures is to determine when and how discs and bulges and especially disc+bulge systems first formed.

These higher order structural parameters can be obtained by light decomposition, i.e. by fitting galaxy surface brightness profiles to well-known functions, such as exponential plus de Vaucouleur light profiles. However, for high-redshift galaxies, this is quite a difficult task, as galaxies are not resolved as well as they are in the local Universe. It is thus critically important to understand the effects of redshifts on our measurements of the light decomposition of these galaxies, which we also examine.

Due to the advent of the Wide Field Camera 3 (WFC3) camera on *Hubble*, we can take advantage of high-quality and high-resolution images of high- $z$  galaxies, and instead of just studying them as a whole, we can perform bulge to disc decomposition with unprecedented accuracy. In fact, there have been studies at high redshift using light decomposition in two dimensions using different codes and methods (e.g. Buitrago et al. 2008; Bruce et al. 2012; van der Wel et al. 2012; Lang et al. 2014) with a variety of results suggesting that galaxies indeed become more ‘discy’ at high redshift, i.e. high-redshift massive galaxies contain on average lower Sérsic indices at high redshift than at lower redshifts (van der Wel et al. 2011). The bulge-disc decomposition allows us to study properties of these two fundamental components separately. In these works, galaxies are typically fitted using a combination of a de Vaucouleurs and an exponential profile to describe, respectively, an assumed bulge and disc component in each galaxy.

Previously, using bulge and disc decompositions, Bruce et al. (2012) claim that at low redshift, massive galaxies are bulge-dominated. While at redshifts  $1 < z < 2$ , galaxies are a mix of bulge+disc systems, and by  $z > 2$ , they are mostly disc-dominated. Up to  $z = 3$ , there are other results showing that stellar mass correlates with the redshift at which *Hubble*-type galaxies start to dominate over peculiar (Mortlock et al. 2013). Nevertheless, it remains unclear what causes this transition and when the dominant structures of the local Universe (bulges and discs) appear as well as if these are related events. In this paper, we investigate the structures of these distant galaxies to determine when, and in what way, discs and spheroids first appear in the massive galaxy population.

We perform one and two-component light decompositions using GALFIT (Peng et al. 2002) and GALAPAGOS (Barden et al. 2012) to a

mass-selected sample of galaxies at  $1 < z < 3$ . We fit the observed two-dimensional surface brightness profiles of galaxies with several models, the first one being a single Sérsic profile (with free  $n$  and  $R_e$ ), and the second one a combination of Sérsic profile (again with free  $n$ ) and an exponential profile. The latter combination describes, respectively, a bulge and a disc. However, it is important to note that we do not assume that this dichotomy translates directly and simply to high-redshift systems, where something more complicated, or a transition phase are potentially present between peculiar systems and the classic bulge+disc systems we see in today’s Universe. By allowing the Sérsic index to vary, we are considering more general bulges, in comparison with previous work where bulges are assumed to be the classical bulge described by a Sérsic law with  $n = 4$ . In this work, we also study a larger sample of galaxies at high redshift than previous works and a wider range in masses. This can lead to a better interpretation of the role that total stellar mass plays in the evolution of bulges and discs.

By fitting the surface brightness to such models, we obtain raw structural parameters for both one- and two-dimensional fits. However, it is important to know whether an individual galaxy is better fit by a two-component profile (bulge+disc) rather than a single Sérsic profile, as in the case for pure spheroid-like galaxies and disc-like galaxies. This is a difficult task and there have been attempts using different methods: Simard et al. (2011) use the  $F$ -test probability to determine the most appropriate model, while Lang et al. (2014) use both the reduced  $\chi^2$  of the model fits and the Akaike information criterion. In this work, we study and combine three different methods: visual classification,  $F$ -test and a method based on the residual flux fraction, RFF (Hoyos et al. 2012), and explore how each method affects the results.

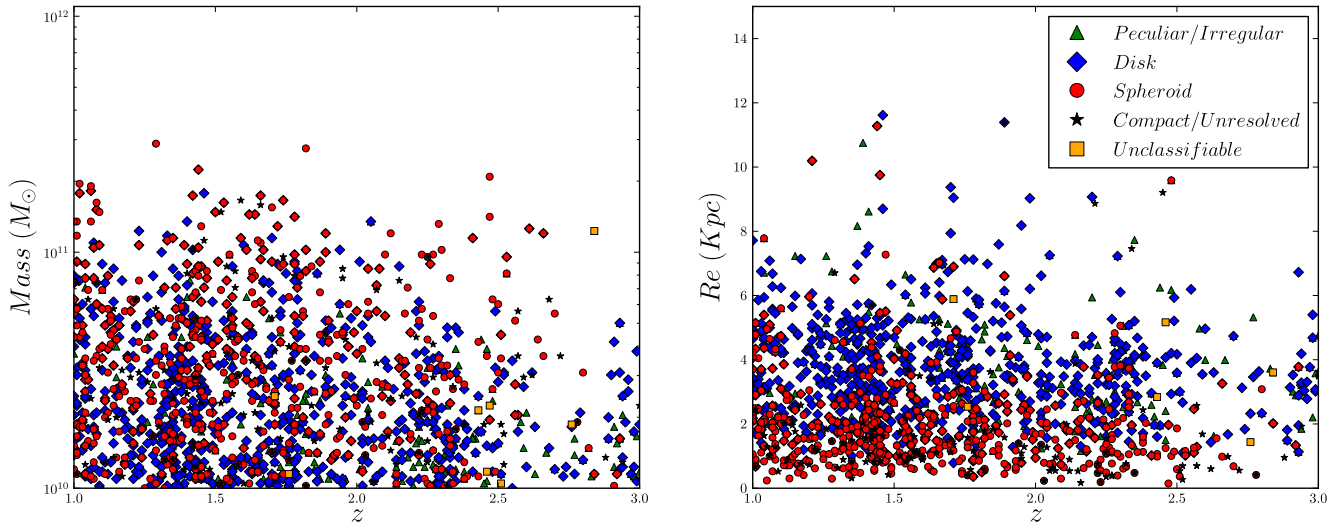
The structure of this paper is as follows. Section 2 is devoted to describing the data we use. In Section 3, we describe how the structural parameters of the galaxies in our sample are obtained, and explain the different methods used to classify them. In Section 4, the main results of the paper are gathered, and in Section 5, we discuss and summarize the results. Finally, an appendix is included with some simulations to better understand our results. Throughout this paper, we use  $AB$  magnitude units and assume the following cosmology:  $H_0 = 70 \text{ Kms}^{-1} \text{ Mpc}^{-1}$ ,  $\Omega_\lambda = 0.7$ , and  $\Omega_m = 0.3$ .

## 2 DATA

### 2.1 Imaging

For this work, we examine a sample of 1495 galaxies at redshifts  $1 < z < 3$  with stellar masses  $M_* \geq 10^{10} M_\odot$  (see Fig. 1) from the CANDELS UDS field. CANDELS (Grogin et al. 2011; Koekemoer et al. 2011) is a MultiCycle Treasury Program which images the distant Universe with both the near-infrared WFC3 and the visible-light Advanced Camera for Surveys (ACS). In total, CANDELS consists of 902 orbits with the *Hubble Space Telescope* (HST) and covers  $800 \text{ arcmin}^2$ . The survey targets five distinct fields (GOODS-N, GOODS-S, EGS, UDS and COSMOS) at two distinct depths. The deep portion of the survey is referred to as ‘CANDELS/Deep’, with exposures in GOODS-N and GOODS-S. ‘CANDELS/Wide’ is the shallow portion and images all five CANDELS fields. We have used the WFC3 data from the UDS which comprises  $4 \times 11$  tiles and covers an area of  $187 \text{ arcmin}^2$  in the F160 ( $H$ -band) filter. The  $5\sigma$  point-source depth for this filter is  $H = 27.1$  ( $AB$  mag).

The CANDELS UDS field is a subset of the larger UDS area which contains data from the  $U$ -band CHFT,  $B$ -,  $V$ -,  $R$ -,  $i$ -,  $z$ -band SXDS data and  $J$ -,  $H$ - and  $K$ -band data from UKIDSS. This includes



**Figure 1.** The stellar masses (left) and effective radius,  $R_e$ , from our one-component fit (right) versus redshift for galaxies used in this work. Colours represent the classification from Kartaltepe et al. (2015). Green triangles are peculiar or irregular galaxies, blue diamonds are disc galaxies, red circles are spheroids, black stars represent compact or unresolved sources, and yellow squares unclassifiable objects.

*F606W* and *F814W* imaging ACS, *H160*, and CANDELS *J125-band* *HST* WFC3 data, *Y* and *Ks* bands taken as part of the HAWK-I UDS and GOODS-S survey (VLT large programme ID 186.A-0898, PI: Fontana; Fontana et al. 2014). For the CANDELS UDS, the 3.6 and 4.5  $\mu\text{m}$  data are taken as part of the *Spitzer* Extended Deep Survey (SEDS; PI: Fazio; Ashby et al. 2013). SEDS is deeper than SpUDS, which is used in the UDS data set, but is only available over a  $0.17 \text{ deg}^2$  region. Therefore, SEDS is a more appropriate choice for the smaller CANDELS UDS region. For a detailed discussion of the CANDELS UDS region photometry, see Galametz et al. (2013).

In Fig. 1, we show how the stellar mass and effective radii of our galaxies are distributed with redshift, along with the morphological classification from Kartaltepe et al. (2015), where galaxies are visually classified into five main morphology classes. Such classes are based on the typical *Hubble* sequence types: discs, spheroids, irregular/peculiar, compact/unresolved and unclassifiable (more than one of these options can be selected for each galaxy). We divide our sample into star-forming and passive galaxies using the rest-frame *UVJ* colours (see Mortlock et al. 2013), where a galaxy is classified as red/passive if it satisfies the following criteria

$$\begin{cases} (U - V) > 1.3 \\ (V - J) < 1.6 \\ (U - V) > 0.88(V - J) + 0.49 \end{cases} \quad (1)$$

and as blue otherwise (see Ownsworth et al., 2016, for more details).

## 2.2 Redshifts and stellar masses

We use a combination of photometric and spectroscopic redshifts as described in Mortlock et al. (2015) and Hartley et al. (2013). The photometric redshifts and stellar masses we use are described in Mortlock et al. (2013). The photometric redshifts were computed by fitting template Spectral Energy Distributions (SEDs) to the photometric data points described in the previous section using the *EAZY* code (Brammer, van Dokkum & Coppi 2008). The photometry was fit to the linear combinations of the six default *EAZY* templates, and an additional template which is the bluest *EAZY* template with a small amount of Small Magellanic Cloud-like extinction added ( $A_V = 0.1$ ). The redshifts are retrieved from a maximum likeli-

hood analysis. For full details of the fitting procedure and resulting photometric redshifts, see Hartley et al. (2013) and Mortlock et al. (2015).

A comparison of the photometric redshifts used in this work to spectroscopic redshifts which are available in the UDS was carried out in Mortlock et al. (2015) where the spectroscopic redshifts versus the photometric redshifts for the 285 CANDELS galaxies with spectroscopic redshifts is discussed (see also Galametz et al. 2013, for details). The dispersion of  $z_{\text{photo}}$  versus  $z_{\text{spec}}$  is  $\delta z / (1 + z) = 0.026$  for the photometric redshifts, after removing the 2 per cent of catastrophic outliers. However, note that we have only a small sample of spectroscopic redshifts to compare to within the CANDELS UDS region.

Stellar masses are obtained by creating a large grid of synthetic SEDs from the stellar population of Bruzual & Charlot (2003), using a Chabrier initial mass function (Chabrier 2003). And the UDS sample is complete down to  $10^{9.5} M_{\odot}$  at  $2.5 < z < 3.2$  (see Mortlock et al. 2015), therefore, our sample of massive galaxies is mass complete. We use as a pointspreadfunction (PSF) the combination of the *TINYTIM*-simulated PSF and a stacked star empirical PSF. The reason for using this PSF is that the *TINYTIM* PSFs are better in the core region (where empirical PSFs tend to broaden), while empirical PSFs appear to fit real stars better in the wings.

## 3 METHOD

We have used *GALFIT* and *GALAPAGOS* to perform our morphological analysis on our sample. *GALFIT* is a two-dimensional fitting code used to model the surface brightness of an object with predefined functions. This program allows the user to fit any number of components and different light profiles (e.g. Sérsic, Exponential disc, Gaussian, Moffat, Nuker, etc.) The most used and useful functions to describe galaxies are the Exponential disc profile and the Sérsic profile (Sérsic 1968) for fitting, respectively, disc and bulges/spheroids.

The Sérsic profile has the following functional form given by,

$$\Sigma(R) = \Sigma_e \exp \left\{ -\kappa_n \left[ \left( \frac{R}{R_e} \right)^{1/n} - 1 \right] \right\}, \quad (2)$$

where the parameter  $R_e$  is the effective radius, such that half of the total flux is within  $R_e$ .  $\Sigma_e$  is the surface brightness at the effective radius  $R_e$ . The parameter  $n$  is the Sérsic index, and it determines the shape of the light profile. Finally,  $\kappa_n$  is a positive parameter that, for a given  $n$ , can be determined from the definition of  $R_e$  and  $\Sigma_e$ . It satisfies the equation  $\Gamma(2n) = 2\gamma(2n, \kappa_n)$ , a non-linear equation which can be solved numerically, where  $\Gamma$  is the gamma function and  $\gamma$  is the incomplete gamma function (see Graham & Driver 2005). The classic de Vaucouleurs profile that describes spheroids and massive galaxy bulges is a special case of the Sérsic profile with  $n = 4$  and  $\kappa_4 = 7.67$ . The exponential disc profile is also a special case of the Sérsic function when  $n = 1$  and  $\kappa_1 = 1.68$ . The best-fitting model is obtained by  $\chi^2$ -minimization using a Levenberg–Marquardt algorithm in GALFIT.

We carry out our fitting with GALAPAGOS and GALFIT. GALAPAGOS is a software that uses SExtractor (Bertin & Arnouts 1996) to detect and extract sources and performs an automated Sérsic profile fit using GALFIT. It is divided into four main stages: the first one detects sources by running SExtractor, the second one cuts out postage stamps for all detected objects, the third block estimates the sky background, prepares and runs GALFIT, and the last stage compiles a catalogue of all galaxies. We then fit all of our sample galaxies with both one- and two-dimensional profiles.

### 3.1 One-component model

We run GALAPAGOS on all of our  $H$ -band galaxy images to fit our sample galaxies with a single Sérsic profile, with  $n$  as a free parameter as in equation (2). GALAPAGOS creates a mask for each individual postage stamp and decides whether a neighbouring object is masked or fit simultaneously, taking into account the distance and relative brightness to the main object. It also calculates the sky value to be used in the fit. As a result we obtain, for all sources, the following parameters: position of the galaxy within the stamp ( $x, y$ ), effective radius  $R_e$ , Sérsic index  $n$ ,  $AB$ -magnitude  $m$ , axis ratio  $q$  and position angle  $PA$ . We discard any fitting with unphysical parameters: effective radius smaller than 0.5 pixels, or larger than the size of the image stamp,  $q < 0.1$ , and  $n < 0.5$  or  $n > 8$  ( $\sim 7$  per cent of the objects).

### 3.2 Two-component model

After the previous procedure, we then run GALFIT on the same postage stamps and use the sky value obtained by GALAPAGOS in Section 3.1 for the single component fit. We fit the surface brightness of the main galaxy to a Sérsic (free  $n$ ) plus an exponential profile (Sérsic profile with  $n$  fixed to  $n = 1$ ), where the total light distribution ( $\Sigma$ ) is the sum of these two models:

$$\Sigma(R) = \Sigma_e \exp \left\{ -\kappa_n \left[ \left( \frac{R}{R_e} \right)^{\frac{1}{n}} - 1 \right] \right\} + \Sigma_e \exp \left[ -1.68 \left( \frac{R}{R_e} - 1 \right) \right], \quad (3)$$

fitting simultaneously or masking neighbour objects in the same way GALAPAGOS does for the one-component model. We constrain the centre of both components to be the same. The result is a list of structural parameters for all the sample galaxies: position in the stamp ( $x, y$ ), effective radius of bulge and disc components ( $R_{eB}, R_{eD}$ ), Sérsic index of the bulge  $n_B$ ,  $AB$ -magnitude for bulge and disc ( $m_B, m_D$ ), axis ratio of bulge and disc ( $q_B, q_D$ ) and position angle of both components ( $PA_B, PA_D$ ). As in the previous model, we exclude

any fitting with unphysical parameters in any component for the effective radius, axial ratio or Sérsic index ( $R_{eB}, R_{eD}, q_B, q_D, n_B$ ).

Adding an extra component increases the degrees of freedom, hence it is more likely that the fitting gets trapped in a local minimum of  $\chi^2$  in the minimization process. To ensure that the  $\chi^2$  obtained from the fitting is the global minimum, we have run GALFIT starting with different initial values of magnitudes, effective radius and Sérsic index. For the Sérsic index, we choose alternatively as initial values  $n = 1, 2.5, 4$ . The starting values of the magnitudes of each components are: both equal to a magnitude that corresponds to half of the total flux obtained from the one-component model, one magnitude which correspond to 80 per cent of the total flux, while the other is 20 per cent and vice versa. The starting values for the effective radius are: both equal to the effective radius obtained from the one-component model, one of the components half the size of that radius, while the other is 20 per cent times larger, and vice versa. We therefore run GALFIT for the  $3^3 = 27$  possibilities. We choose the model that delivers the smallest  $\chi^2$  and does not have any unphysical parameters.

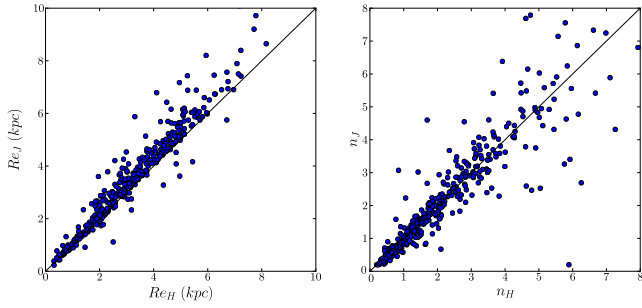
We first try to fit all the central components of our galaxies with a free  $n$  for the Sérsic profile (first term of equation 3), but in some cases ( $\sim 40$  per cent), the fitting results in an unrealistic Sérsic index (either too small or too big). In such cases, we redo the fitting with the Sérsic index fixed first at  $n = 4$  and then at  $n = 1$  in equation (3), and choose the fitting with the smallest  $\chi^2$ . In  $\sim 73$  per cent of these cases, the model prefers  $n = 1$ . There are still some objects ( $\sim 20$  per cent) that do not have any realistic result with two components, those will be directly classified as one-component galaxies (if the fitting in this case is considered good) in all methods. In the end, only about  $\sim 8$  per cent of the galaxies are not well represented with either the one or the two-component model. These galaxies are either very compact objects, or considerably faint/small, and have an average redshift of  $z = 2$ .

We later discuss in Section 4.4 how the ratio of the fluxes in the two components changes with redshift. Overall, we find that there is a fairly broad distribution of the ratio between the fluxes of the two components. Only about 6 per cent of the galaxies have a second component which is less than 10 per cent of the total flux. Otherwise, 70 per cent of the sample of two-component galaxies are disc-dominated, with  $B/T < 0.5$ .

### 3.3 Morphological $K$ -correction

We also investigate whether we should consider the effect of the morphological  $K$ -correction in our study, as the quantitative structure of galaxies changes as a function of wavelength (e.g. Taylor-Mager et al. 2007). Using the  $H$  band in the redshift range of  $1 < z < 3$  means that we are observing and comparing galaxies at a rest-frame wavelength from visible to near-IR (Conselice et al. 2011). Therefore, the difference in rest-frame wavelength is  $\leq 350$  nm at the highest and lowest redshifts. To test whether this difference can have an effect on the structure and morphology of our galaxies, we select a subsample with  $z \sim 1$  and fit their surface brightness to a single Sérsic profile in the  $J$  band. The observed rest frame in this case is  $\sim 600$  nm and by comparing with the same galaxies in the  $H$  band, we see the effect caused by a difference in rest-frame wavelength of  $\sim 200$  nm which is in a similar range to that of our whole sample of galaxies. In Fig. 2, we see that we recover the same structural parameters (effective radius and Sérsic index) whether we use the  $J$  or  $H$  band. This means that the spanning in redshift for our sample of galaxies does not affect the observed





**Figure 2.** Comparison between effective radius (left) and Sérsic index (right) obtained from fitting the surface brightness in the  $J$  band and  $H$  band with a single Sérsic profile at  $z = 1$  with a  $\Delta\lambda = 200$  nm difference.

structure and morphology. Therefore, we can continue our study without having to consider the morphological  $K$ -correction.

### 3.4 Classification

Once we have the two models for each of our galaxy profiles, we need a method to choose whether to use one- or two-component fits for each galaxy. This is critical for both determining the evolution of one-component galaxies, as well as for how multiple component galaxies form and evolve over the epoch  $1 < z < 3$ . In this paper, we investigate three different methods of deciding whether a galaxy is better ‘fit’ as a one- or two-component system, and compare the results of these methods to see how internally consistent they are. Our first method consists in visually classifying galaxies into one- or two-component systems. Our second method is based on an index called the RFF, and the final method is based on an statistical test ( $F$ -test). All of these methods are explained below.

#### 3.4.1 Visual inspection

The first method of determining whether a galaxy has one or two components consists of visually inspecting all the sample galaxies,

and their correspondent residual images from both one- and two-component best-fitting models. In Fig. 3(a), we show examples of the fitting using one-component for three different types of galaxies. For each model, we show the original image (left), the model image (middle), and the residual image (right) which is obtained by subtracting the model to the original image. Analogously, in Fig. 3(b), we show the fitting using two components.

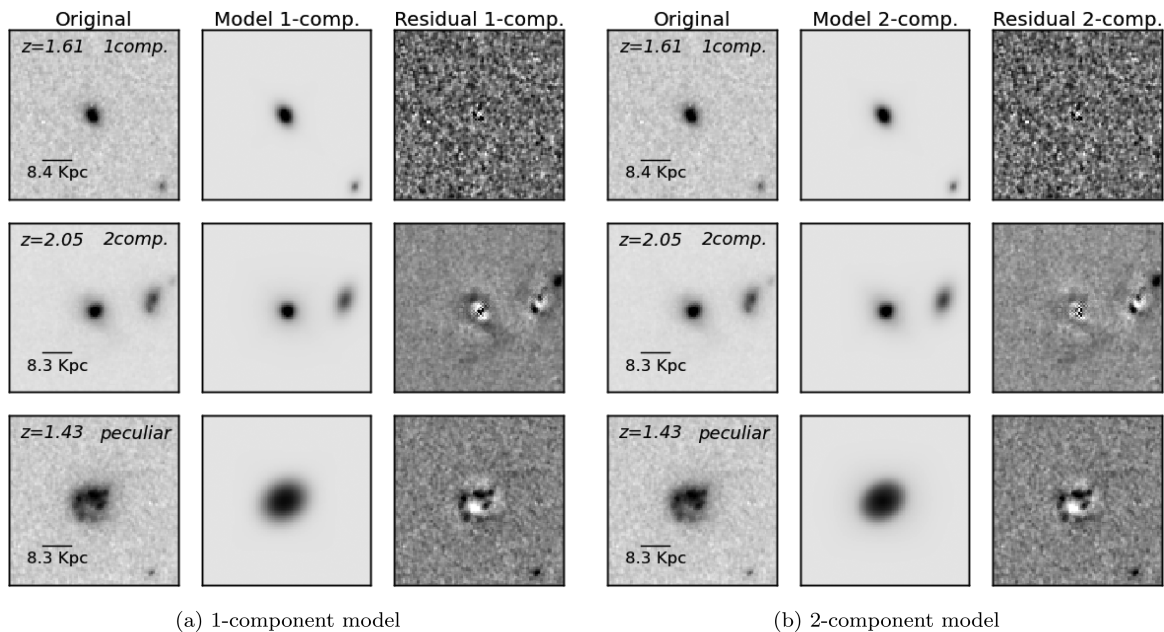
We have visually classified all the galaxies in our sample into one of three types (examples in Fig. 4) based on both the visual appearance of the galaxy and also the residuals left over from the galaxy after the best-fitting one and two-component profiles are fit.

*One-component galaxies:* these are disc-like or spheroid-like galaxies, which show no evidence of needing a second component. Indeed, a single Sérsic profile fitting is able to reproduce well the surface brightness of the galaxy as shown by the lack of structures left in the residual image.

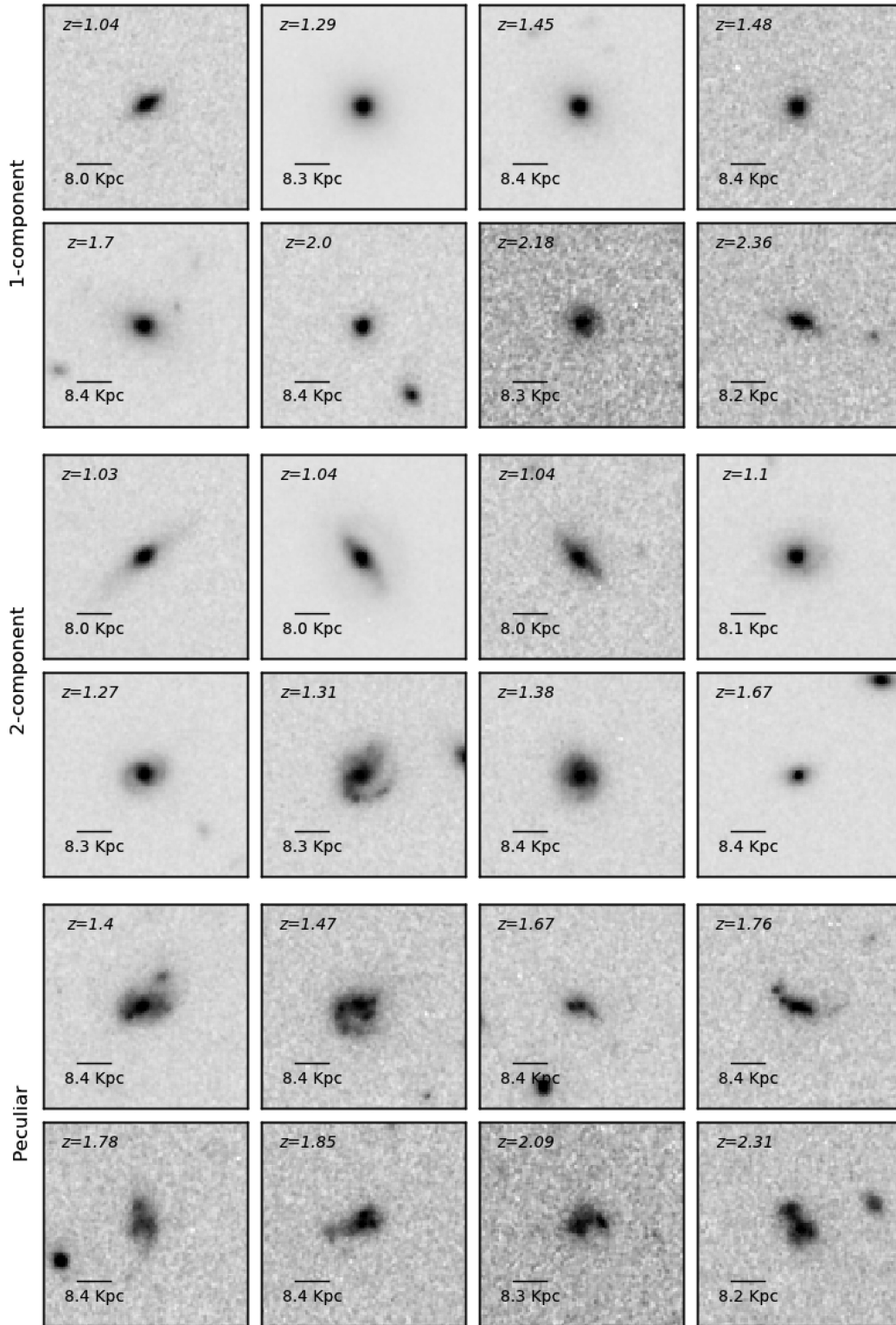
*Two-component galaxies:* these are sometimes disc galaxies with a bulge component. They are better fit with a composition of a Sérsic profile plus exponential profile. They show less residual light from the two-component models than with the single one, although a significant amount of residual can be left due to spiral arms in the disc.

*Peculiar galaxies:* these are disturbed galaxies or mergers. They show residuals from both models, and the addition of another component does not improve the fitting.

There is a very small fraction of galaxies ( $\sim 3$  per cent) that are removed from our sample: unresolved or unclassifiable (due to problems with the image) galaxies. Note that galaxies classified as one or two-component galaxies can display irregular or merger features, but unlike peculiar galaxies, they are still well represented by either the single Sérsic model or the Sérsic plus exponential model, respectively, as these features do not dominate the structure of the galaxy.



**Figure 3.** Visual classification. Left: one-component model (original image, model and residual). Right: two-component model (original image, model and residual). Top row: example of one-component best fit. Middle row: example of two-component best fit. Bottom row: example of a peculiar galaxy.



**Figure 4.** Examples of galaxies visually classify as one-component galaxies (top), two-component galaxies (middle) and peculiar galaxies (bottom). The postage stamps are  $6 \text{ arcsec} \times 6 \text{ arcsec}$  in size.

### 3.4.2 Residual flux fraction

The residual flux fraction, or RFF (Hoyos et al. 2011) is defined as the fraction of the signal contained in the residual image that cannot be explained by fluctuations of the background. Hence, the

smaller the RFF value, the better the fitting. This index is defined as follows

$$\text{RFF} = \frac{\sum_{(j,k) \in A} |I_{j,k} - I_{j,k}^{\text{GALFIT}}| - 0.8 \sum_{(j,k) \in A} \sigma_{B,j,k}}{\text{FLUX\_AUTO}}, \quad (4)$$

where  $I$  is the actual galaxy image,  $I^{\text{GALFIT}}$  is the model image created by GALFIT,  $\sigma_B$  is the background rms image and  $FLUX\_AUTO$  is the total flux of the galaxy calculated by SExtractor. Finally,  $A$  represents the area in which we calculate this index. The 0.8 factor in the numerator guarantees that for a Gaussian noise error image, the expected value of the RFF is 0 (see Hoyos et al. 2011, for details). It is important to note that the RFF diagnosis does not work well for large areas (Hoyos et al. 2011). In those cases, as the galaxy decays towards zero flux at large radius, outer areas will dominate the RFF computation, making it small even when the residual is not good at the centre. Taking this into consideration, we have decided to use the area inside the  $2.5 R_{\text{kron}}$  radius of each galaxy to calculate the RFF, where  $R_{\text{kron}}$  is the Kron radius obtained from SExtractor (e.g. Zhao, Aragón-Salamanca & Conselice 2015).

To calculate the first term of the numerator in equation (4), we sum the absolute value of the pixels inside the chosen area from the residual image (original image subtracted by the model). If there is a nearby, but different, object inside this area we do not take into account the pixels corresponding to that object when calculating the RFF. This reduces as much as possible bad fittings from nearby objects affecting the RFF of the main object. To compute the second term of the numerator, we assume

$$\sum_{(j,k) \in A} \sigma_{B,j,k} = N \langle \sigma_B \rangle, \quad (5)$$

where  $\langle \sigma_B \rangle$  is the mean value of the background sigma for the whole image, and  $N$  the number of pixels in the area we are considering in the calculation of the RFF (excluding those pixels belonging to nearby objects). We obtain the value  $\langle \sigma_B \rangle$  directly from the sky measures from SExtractor.

We compute the RFF for both the one and two-component models (denoted as  $RFF_1$  and  $RFF_2$ , respectively) for all the objects in our sample. Peculiar and spiral galaxies have similar RFF values, namely the average value of RFF for the one-component model in spirals is  $RFF_1 = 0.07$ , while for peculiars, it is  $RFF_1 = 0.08$ , making it difficult to distinguish these two populations using just the RFF. To solve this problem, we use our visual classification (3.4.1) to separate these two populations.

Spheroid-like galaxies have small  $RFF_1$  ( $RFF_1 \lesssim 0.5$ ) and  $RFF_2 \cong RFF_1$ , as they are well fit by a single Sérsic profile model. Meanwhile, galaxies that contain a bulge and a disc will generally have a larger  $RFF_1$  ( $RFF_1 \gtrsim 0.5$ ), due the spiral arms and  $RFF_2 < RFF_1$ , as the two-component fitting will be better than the single-component model. Therefore, they will occupy a different region in the plane of  $RFF_2$  versus  $RFF_1$  (see Fig. 5).

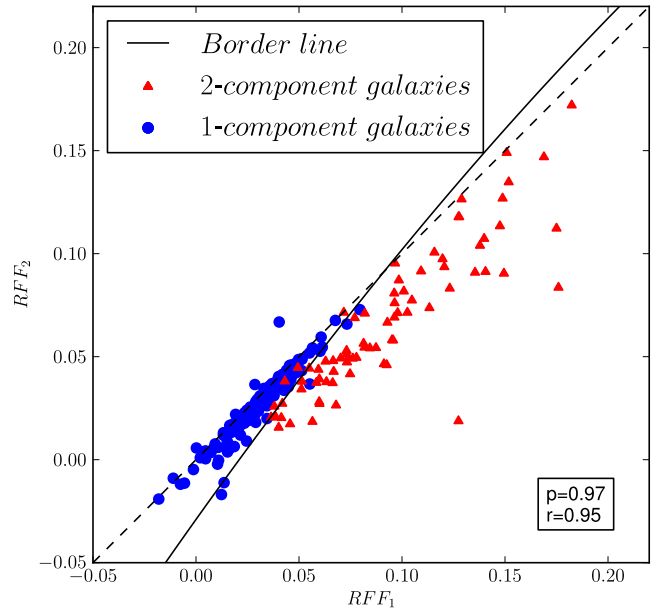
We have also used the  $F$ -score technique (Hoyos et al. 2012) to find the border in the  $RFF_2$  versus  $RFF_1$  diagram that best separates these two populations (one-component and two-component galaxies). This method consists in finding the parameters of a function (the border, that in our case will be a second-order polynomial) that maximize the  $F$ -score,  $F_\beta$  (van Rijsbergen 1979), defined as

$$F_\beta = \frac{1 + \beta^2 pr}{\beta^2 p + r}, \quad (6)$$

where  $r$  and  $p$  are the sensitivity or completeness of both populations and are given by the equations

$$r = \frac{\#\{\text{True Population}_1\}}{\#\{\text{True Population}_1\} + \#\{\text{False Population}_2\}} \quad (7)$$

$$p = \frac{\#\{\text{True Population}_2\}}{\#\{\text{True Population}_2\} + \#\{\text{False Population}_1\}}. \quad (8)$$



**Figure 5.**  $F$ -score technique using RFF measures (see Section 3.4.2 for more details). Using a training sample obtained by visual classification (red triangles: two-component galaxies, blue circles: one-component galaxies), we obtain the line (black solid line) that separates these two subsamples given by equation (9).  $p$  and  $r$  are the completeness of the two subsamples. The dashed line is the identity function.

In these definitions,  $\#\{\text{True Population}_1\}$  is the number of objects correctly classified as  $\text{Population}_1$  by the method while  $\#\{\text{False Population}_1\}$  is the number of those objects of the  $\text{Population}_2$  misclassified as belonging to  $\text{Population}_1$ . We define analogously  $\#\{\text{True Population}_2\}$  and  $\#\{\text{False Population}_2\}$ . Hence,  $r$  measures the fraction of the actual elements in  $\text{Population}_1$  correctly classified as  $\text{Population}_1$ . Finally,  $\beta$  is a control parameter, specified by the user, that determines the relative importance of  $r$  and  $p$ . We have chosen  $\beta = 1.0$  because having a complete sample of the  $\text{Population}_1$  is as important as having a complete sample of  $\text{Population}_2$ .

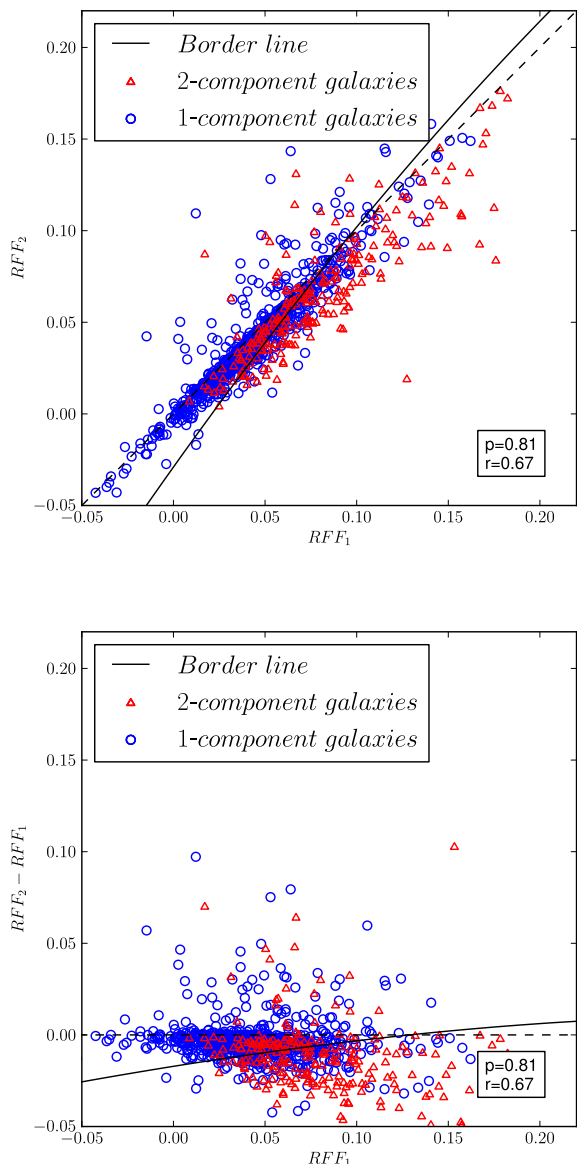
To apply this method to our sample of galaxies, we have performed the  $F$ -score technique in a training sample. This training sample is formed by the galaxies that have been visually classified most confidently as either one-component ( $\text{Population}_1$ ) or two-component systems ( $\text{Population}_2$ ). This allows us to obtain the (second-order polynomial) line that best separates these two populations, from which we can then classify the rest of the galaxies in our sample according to this criterion. It is interesting to note that, in our case, changing the value of  $\beta$  does not significantly change the border line, as both populations of the training sample are clearly separated in the  $RFF_2$  versus  $RFF_1$  plane. This maximization has been performed with the Amoeba algorithm (Press et al. 1988), using a second-order polynomial as the border line. The result of this maximization is shown in Fig. 5 and can be expressed as

$$RFF_2 = -0.023 + 1.40RFF_1 - 0.94RFF_1^2. \quad (9)$$

This line gives the following values for the completeness of the two populations, for the training sample:  $r = 0.95$ ,  $p = 0.97$ .

As mentioned earlier, once we know this line, we can plot in the  $RFF_2$  versus  $RFF_1$  plane all our objects (see Fig. 6) and classify them according to their position with respect to the equation of line (9): as one-component galaxies if they lie above the line or as two-component galaxies if they are under the line. In Fig. 6,





**Figure 6.** RFF<sub>2</sub> (top) and difference in RFF = RFF<sub>2</sub> – RFF<sub>1</sub> (bottom) versus RFF<sub>1</sub>, with the classification of all the galaxies using equation (9) (black solid line), below which galaxies are considered as having two components, and above the line, they are classified as one-component galaxies. Red triangles are galaxies visually classified as two-component galaxies and blue circles as one-component ones.

we plot the entire sample and, just for comparison with the visual classification, we have plotted in blue circles those objects that have been visually classified as one-component galaxies and in red triangles those classified as two-component galaxies. There is overall a good agreement between the two methods.

### 3.4.3 *F*-test

The *F*-test is a statistical test in which the statistic has an *F*-distribution under the null hypothesis. An *F*-distribution is formed by the ratio of two independent  $\chi^2$  variables divided by their respective degrees of freedom.

We have performed the *F*-test following the method described in Simard et al. (2011), who also use it for one- versus two-component

separation within SDSS data. In our study, we have two models: Sérsic profile (model 1) and a Sérsic+exponential profile (model 2). We consider the  $\chi^2$  for each model from the residual image, and take as degrees of freedom the number of resolution elements,  $n_{\text{res}}$ , minus the number of free parameters in the model. The number of resolution elements can be calculated as follows

$$n_{\text{res}} = \frac{n_{\text{pixels}}}{\pi \theta^2}, \quad (10)$$

where  $n_{\text{pixels}}$  is the number of unmasked object pixels used in the fitting, and  $\theta = 1.38 \text{ pix}$  is the *H*-band seeing half-width half-maximum, in units of pixels. As in the RFF calculation, we compute the  $\chi^2$  in the area inside the  $2.5 R_{\text{kron}}$  of each galaxy.

To know whether the  $\chi^2$  from model 2 is significantly smaller than the  $\chi^2$  from model 1, we have to perform a one tailed test. The hypothesis of such test can be formulated as follows.

- (i) Null hypothesis  $H_0: \chi_1^2 \leq \chi_2^2$ , (the simpler model is correct).
- (ii) Alternative hypothesis  $H_1: \chi_1^2 > \chi_2^2$ .

From the statistic of the test  $F = \chi_{\text{red},2}^2 / \chi_{\text{red},1}^2$ , the probability  $P$  of accepting the null hypothesis (i.e. the probability that the more complex model is not required) can be calculated. Following Simard et al. (2011), we set a  $1\sigma$  threshold value  $P_0 = 0.32$  below which we consider galaxies to be better fit by model 2 (Sérsic+exponential), and therefore classified as two-component galaxies. Meanwhile, those with  $P > P_0$  are classified as one-component galaxies.

Note, however, that with this method, we cannot distinguish peculiars from one-component galaxies, as in both cases, the more complex model is not required. As in the RFF method, we have used the visual classification to separate the peculiar galaxies. Those can also be separated by using the asymmetry index (e.g. Conselice et al. 2003), finding similar galaxies (Mortlock et al. 2013).

## 4 RESULTS

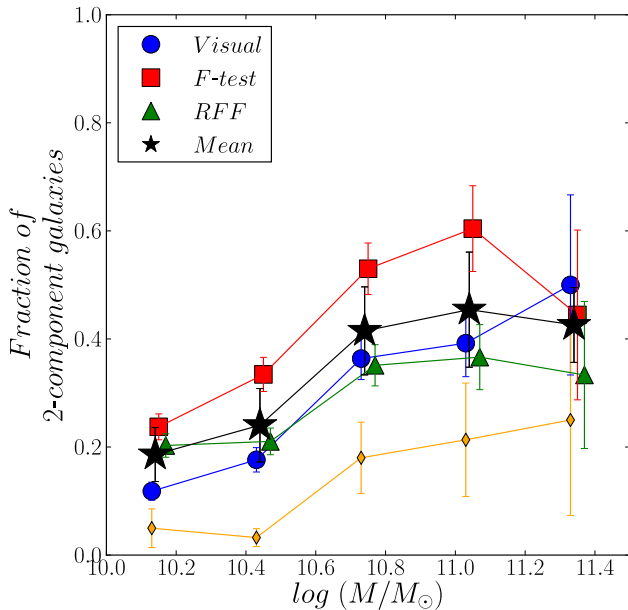
In this section, we compare how the selection of our galaxy sample into one- or two-component types varies from one method to another. For our final results, we average the properties for the three methods to take into account the strengths and weaknesses of each method. These final results include examining the fraction of one- or two-component galaxies as a function of redshift and stellar mass, as well as the evolution of the sizes of these components with redshift. In Appendix, we present simulations to test the robustness of our conclusions.

### 4.1 Method comparison and basic trends

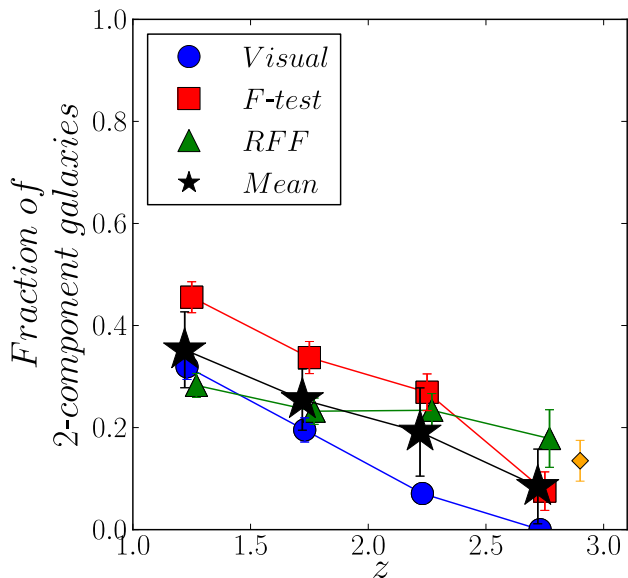
We first explore how the three methods select different galaxies as being one- or two-component systems. We demonstrate this in Figs 7 and 8 which show the fraction of galaxies selected as two-component galaxies by each method, as a function of mass and redshift, respectively, normalized by the total number of galaxies in each bin. The first thing to take away from these figures is that the agreement between the three methods is good, with the average of the three methods shown as the black stars in both figures.

In more detail, we see that the fraction of two-component galaxies increases with stellar mass (Fig. 7) by a factor of  $\sim 2$  from the lowest mass ( $19 \pm 5$  percent) to the highest mass bin ( $43 \pm 6$  percent). We explored the possibility that this trend was due to S/N instead of stellar mass, but we observed that regardless of the S/N, the trend of two-component galaxies with stellar mass was preserved, so we believe that this trend is real. This trend is also not a result of redshift effects as we show in the Appendix.





**Figure 7.** Fraction of two-component galaxies as a function of stellar mass at redshifts  $1 < z < 3$ , selected by the three selection methods (visual classification: blue circles,  $F$ -test: red squares, and RFF method: green triangles) as a function of the stellar mass. The black stars represent the mean fraction in each mass bin. The yellow diamonds are the fraction of two-component galaxies from the high-redshift simulated galaxies (see Appendix). Error bars represent the random errors.



**Figure 8.** Fraction of two-component galaxies selected by the three selection methods (visual classification: blue circles,  $F$ -test: red squares, and RFF method: green triangles) as a function of redshift. The black stars represent the mean fraction in each redshift bin. The yellow diamond is the fraction of two-component galaxies from the high-redshift simulated galaxies (see Appendix). Error bars represent the random errors.

In a similar way, we see a trend in terms of the fraction of two-component galaxies within our  $M_* > 10^{10} M_\odot$  sample at  $z < 3$  (Fig. 8). The fraction of two-component galaxies decreases with higher redshift for all the methods, from  $35 \pm 6$  per cent at  $z \sim 1$  to  $8 \pm 6$  per cent at  $z \sim 3$ . This is a significant change over a relatively quick  $\sim 2$  Gyr time-span. However, some of this evolution could be

due to redshift effects (see Appendix), so this must be a tentative conclusion at present.

Another thing to note is that in terms of the fraction of two-component galaxies evolving with redshift, the RFF method appears to be roughly constant while the other two methods decline with higher redshift. This is likely due to our galaxies having some residual light even at high redshift. This can arise from having two components at lower redshifts, where indeed the RFF agrees with the  $F$ -test and visual methods. However, at higher redshifts, what we are likely seeing is higher RFF values which are due to galaxy formation processes such as residuals from merging or star formation (e.g. Conselice et al. 2003; Conselice, Rajgor & Myers 2008; Bluck et al. 2012). These signatures would not as easily be seen in the other two methods.

#### 4.2 Number density evolution of galaxy components

Investigating the number density (number of galaxies normalized by the comoving volume) of different galaxy selections allows us to determine at which epoch different types of galaxies dominate, and how they evolve throughout the history of the Universe. We can also compare the rate at which the number density grows between different kind of systems. In this paper, we explore how the number density of galaxies best fit with either one and two components evolve in terms of their number densities during the epoch  $1 < z < 3$ .

We have split our galaxies into four categories: one-component discs or disc-like galaxies ( $n < 2.5$ ), one-component spheroids or spheroid-like galaxies ( $n > 2.5$ ), two-component galaxies, and peculiar galaxies. In Fig. 9, we plot the number density of galaxies in five redshift bins for the different types of one- and two-component galaxies (number of galaxies normalized by the comoving volume in  $\text{Mpc}^3$  corresponding to that bin). In black, we plot the mean number density of the three methods.

We fit the mean number density  $n_d(z)$  of each type of galaxy using three different functions. First, we fit a linear function

$$f_1(z) = az + b, \quad (11)$$

secondly, we also fit a power-law function

$$f_2(z) = \gamma(1+z)^\alpha, \quad (12)$$

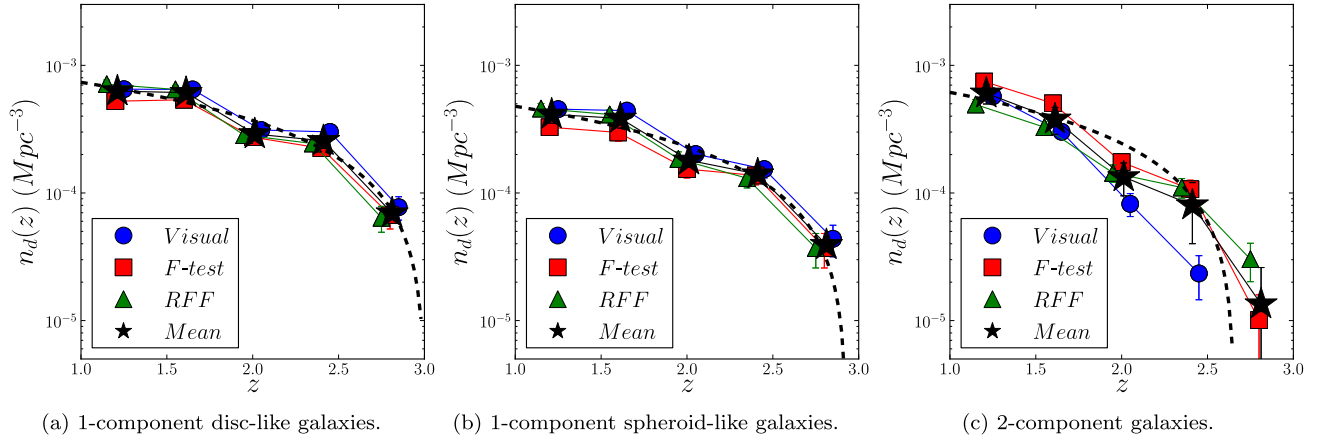
and lastly, an exponential function

$$f_3(z) = n_0 e^{(-z/z_0)}. \quad (13)$$

We show in Table 1 the result for all three fits, noting that the function that best fits the data is the linear function. To compare the number density of the different types of galaxies, we plot in Fig. 10 all four number densities, and the total number density of galaxies in the CANDELS-UDS (Mortlock et al. 2015). The dashed lines show the best fit of the linear function, i.e. equation (11).

Already there are several trends which are visible on Fig. 10. The first is the rise of the two-component galaxies. While they make up a small fraction of the galaxy population at  $z \sim 2.5$ , they rise by a factor of 30 in number density to become just as abundant as the disc-like galaxies. This reveals that this epoch of  $1 < z < 3$  is when two-component galaxies form and dominate the abundances of massive galaxies.

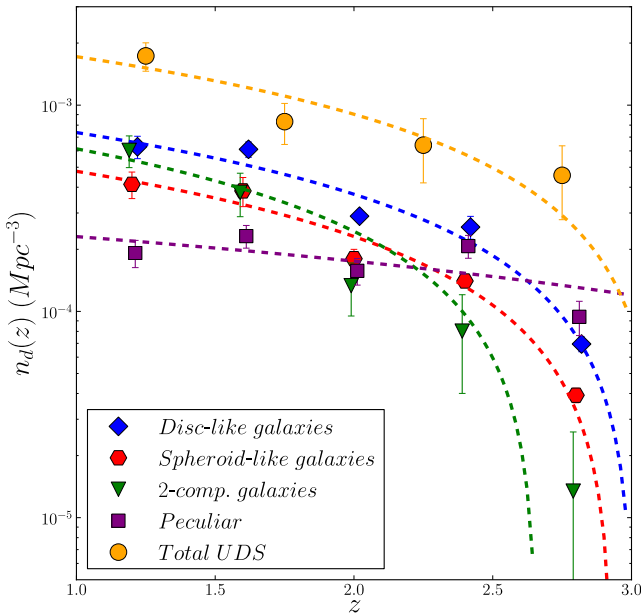
We also compare in Fig. 10 our number densities for the individual galaxy types and the total number density of all galaxies. The



**Figure 9.** Number density  $n_d(z)$  as a function of redshift for each type of galaxy (disc-like (a), spheroid-like (b) and two-component galaxies (c)) for the three selection methods (visual classification,  $F$ -test and RFF method). The points are coloured as in Fig. 7. The dashed black lines are the lines that best fit the mean values (note the log scale).

**Table 1.** Parameters of the fittings to the mean number density of the different types of galaxies and to the total density of UDS.

Function		Disc-like galaxies	Spheroid-like galaxies	2-component galaxies	Total UDS
Linear	$a$	$-0.00038 \pm 0.00004$	$-0.00025 \pm 0.00004$	$-0.00036 \pm 0.00006$	$-0.0008 \pm 0.0002$
	$b$	$0.0011 \pm 0.0001$	$0.0007 \pm 0.0001$	$0.0010 \pm 0.0001$	$0.0025 \pm 0.0004$
Power law	$\alpha$	$-2.5 \pm 0.3$	$-2.7 \pm 0.3$	$-4.2 \pm 0.8$	$-2.9 \pm 0.8$
	$\gamma$	$0.005 \pm 0.001$	$0.004 \pm 0.002$	$0.02 \pm 0.02$	$0.02 \pm 0.02$
Exponential	$z_0$	$1.1 \pm 0.1$	$1.0 \pm 0.1$	$0.6 \pm 0.1$	$1.0 \pm 0.3$
	$n_0$	$0.0021 \pm 0.0005$	$0.0015 \pm 0.0004$	$0.004 \pm 0.004$	$0.006 \pm 0.004$



**Figure 10.** Total number density evolution of the different types of galaxies in our sample (disc-like galaxies: blue diamonds, spheroid-like galaxies: red hexagons, two-component galaxies: green triangles, and peculiar galaxies: purple squares). The yellow circles are the total number density of galaxies in UDS calculated by equation (14). The dashed lines are the straight lines that better fit the mean values (in log scale).

total number density of galaxies within the UDS (Mortlock et al. 2015) is calculated as

$$\phi_{\text{Total}} = \int_{M_1}^{M_2} \phi(M; \phi^*, M^*, \alpha) dM, \quad (14)$$

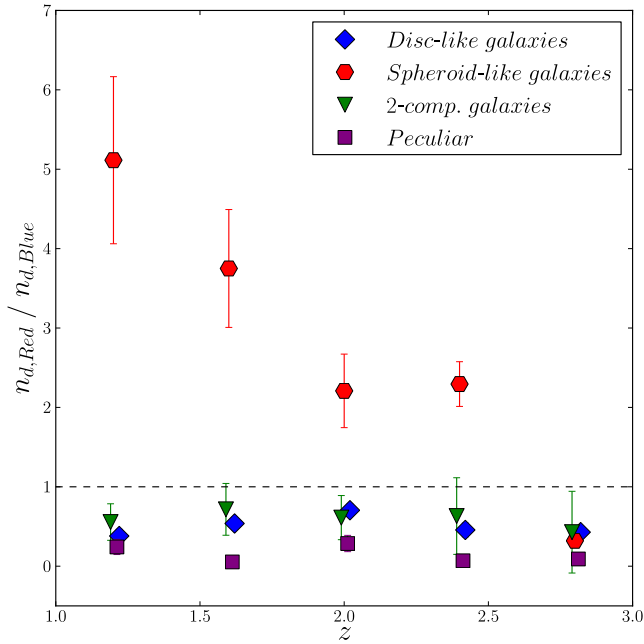
where  $M_1 = 10^{10} M_\odot$  and  $M_2 = 10^{12} M_\odot$ . The map  $\phi(M)$  is the Schechter function (Schechter 1976) given by

$$\phi(M; \phi^*, M^*, \alpha) = \phi^* \ln 10 (10^{M-M^*})^{1+\alpha} \exp(-10^{M-M^*}), \quad (15)$$

where  $\phi^*$  is the normalization of the Schechter function,  $M^*$  is the turn over mass in units of dex,  $\alpha$  is the faint end slope of the Schechter function and the variable  $M$  is the stellar mass in units of dex. The parameters  $\phi^*$ ,  $M^*$  and  $\alpha$  depend on the redshift range, and the values we use to obtain  $\phi_{\text{Total}}$  in each redshift bin are calculated in Mortlock et al. (2015).

From Fig. 10 and Table 1, we see that the total number density of one-component disc-like galaxies evolves at a similar rate to that of the two-component galaxies, but its value is about 1.5 times larger. The number density of spheroid-like galaxies increases slower than those of the other types of galaxies. We observe that, for the whole redshift range, the number density of disc-like galaxies is the highest of the four types of galaxies. The  $z \gtrsim 2$  spheroid-like galaxies have a higher number density than the two-component galaxies, but at  $z \lesssim 2$ , the number density of two-component galaxies become greater than the spheroid-like galaxies.

In Fig. 11, we compare blue and red galaxies according to the  $UVJ$  selection (equation 1) for the four different types of galaxies. Peculiar galaxies are mostly blue at all redshifts, with just a small number of them being red. For disc-like and two-component galaxies, the fraction of blue galaxies is greater than half. Interestingly, at redshift  $z \sim 3$ , most of the spheroid-like galaxies are blue, but as redshift decreases, the fraction of red spheroids rapidly increases while blue spheroids decrease. By redshift  $z \sim 1$ , the vast majority (85 per cent) of spheroid-like galaxies are red.



**Figure 11.** Ratio of the number density of red galaxies and blue galaxies for the different types of systems (disc-like galaxies: blue diamonds, spheroid-like galaxies: red hexagons, two-component galaxies: green triangles, and peculiar galaxies: purple squares).

#### 4.3 Total stellar mass density

By investigating the stellar mass density, we can determine which types of galaxies dominates the stellar mass in the Universe at which epoch. This is closely related to the number density evolution, but here we are investigating essentially whether the galaxies in each selection are more massive on average than in the other selections.

In Fig. 12, we plot the mass density  $\rho_*$  of the different types of galaxies as a function of redshift, and in black we plot the mean mass density of the three methods.

We fit the mean mass density of each type of galaxy using the same functions as in Section 4.2 (linear function, power law and exponential function). We show the result of the fits in Table 2 for all three fits. In Fig. 13, we plot the mass density for all four classes

of galaxies using the average selection, as well as the total mass density from the UDS. The dashed lines show the best fit of the linear function, i.e. equation (11).

The total mass density of galaxies in the UDS is calculated as

$$\rho_{*,\text{Total}} = \int_{M_1}^{M_2} M_* \phi(M; \phi^*, M^*, \alpha) dM, \quad (16)$$

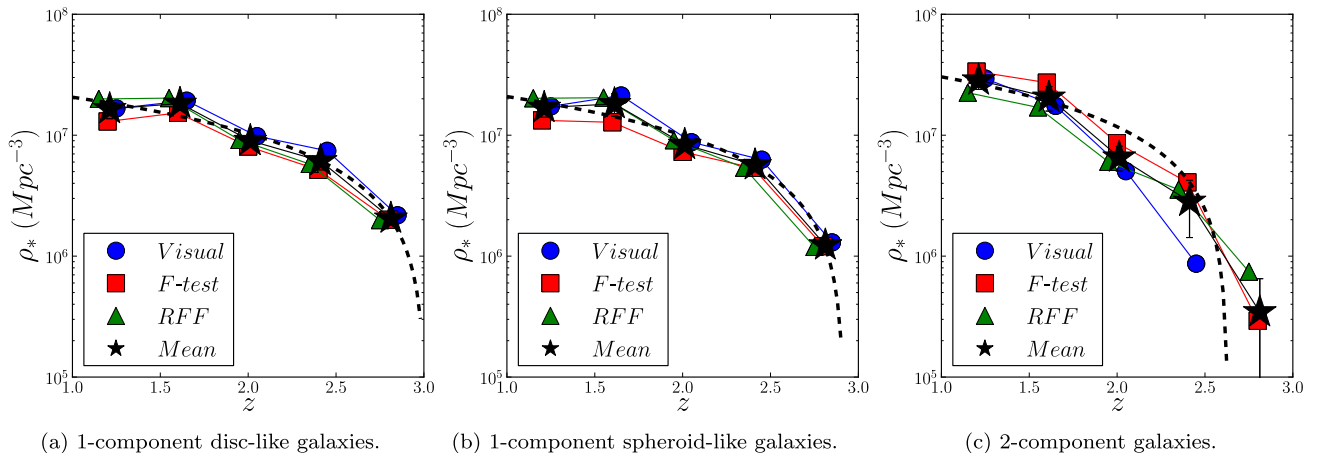
where  $M_1 = 10^{10} M_\odot$ ,  $M_2 = 10^{12} M_\odot$ , and  $\phi(M)$  is the Schechter function defined in equation (15).

The mass density of the one-component galaxies evolves at similar rates independently of the Sérsic index selection (i.e. being discs or spheroids). The mass density of two-component galaxies have the highest increase over the whole redshift range, and its contribution to the total mass density is smaller than that of one-component galaxies at  $z = 1.75-3$ , but for  $z = 1-1.75$ , the mass density of this type of galaxy becomes dominant. The two-component galaxies, in fact, dominate the mass density of massive galaxies at these lower redshifts.

Overall, we find that the mass density for two-component galaxies increases by a factor of  $\sim 100$ , which is roughly a factor of 3 higher than for its number density increase. Therefore, we see a larger effect in the integrated mass density for our galaxies than the increase in the number density. This implies that the galaxies which are driving this increase are more massive at the lower end of the redshift range around  $z \sim 1$  than at higher redshifts, relative to the one-component galaxies. This implies that the most massive galaxies preferentially become the two-component systems at lower redshifts, while the one-component systems are relatively lower mass.

#### 4.4 The size evolution of components

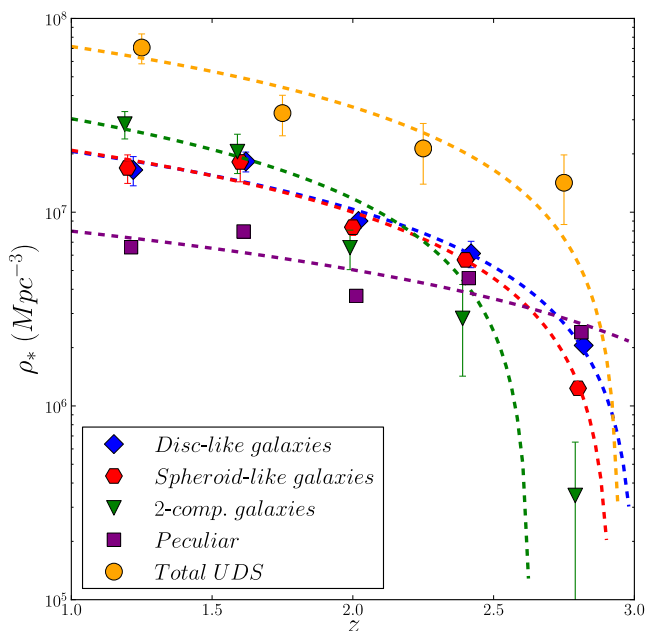
We explore the evolution in size of our galaxy sample, and their components, to determine if the inner and outer components grow together or not. As the effective radius  $R_e$  calculated from GALFIT corresponds to the major axis of an ellipse containing half of the light, in order to compare with other results, we have calculated a circularized radius  $R_{e,\text{circ}} = R_e \sqrt{b/a}$ , where  $b/a$  is the axis ratio. In Fig. 14, we plot the median circularized effective radius of the one-component galaxies, as well as the discs and bulges of two-component galaxies, as a function of redshift (for the three different methods as well as the average of the three methods).



**Figure 12.** Mass density as a function of redshift for each type of galaxy in our sample (disc-like (a), spheroid-like (b) and two-component galaxies (c)) for the three selection methods (visual classification,  $F$ -test and RFF method). The points are coloured as in Fig. 7. The dashed black line are the straight lines that best fit the mean values (black stars) (in log scale).

**Table 2.** Parameters of the fittings to the mean mass density of the different types of galaxies and to the total density of UDS.

Function		Disc-like galaxies	Spheroid-like galaxies	2-component galaxies	Total UDS
Linear	$a(10^7)$	$-1.0 \pm 0.2$	$-1.1 \pm 0.02$	$-1.8 \pm 0.03$	$-3.6 \pm 0.8$
	$b(10^8)$	$0.31 \pm 0.04$	$0.32 \pm 0.05$	$0.48 \pm 0.06$	$1.1 \pm 0.2$
Power law	$\alpha$	$-2.5 \pm 0.3$	$-2.7 \pm 0.3$	$-4.2 \pm 0.8$	$-2.9 \pm 0.8$
	$\gamma(10^8)$	$1.2 \pm 0.5$	$1.6 \pm 0.8$	$8.1 \pm 0.4$	$11 \pm 3$
Exponential	$z_0$	$1.1 \pm 0.2$	$1.0 \pm 0.1$	$0.6 \pm 0.1$	$0.8 \pm 0.3$
	$n_0(10^8)$	$0.5 \pm 0.2$	$0.7 \pm 0.2$	$2 \pm 1$	$3 \pm 1$


**Figure 13.** Mean mass density evolution of the different types of galaxies (disc-like galaxies: blue diamonds, spheroids-like galaxies: red hexagons, two-component galaxies: green triangles, and peculiar galaxies: purple squares). The yellow circles are the total mass density of galaxies in UDS calculated by equation (16).

First, we observe that there is a trend for the one-component galaxies: on average, they grow in size at lower redshifts, although the evolution appears stronger in disc-like galaxies than in spheroid-like galaxies. In the range  $1 < z < 3$ , the disc-like one-component galaxies grow on average from 1.3 to 2.1 kpc, i.e. an increase of 60 per cent. On the other hand, the one-component spheroid-like galaxies grow on average from 1.05 to 1.24 kpc, i.e. an increase of only 18 per cent. Thus, during this epoch, the disc-like systems dominate the growth. Note that our simulations of  $z = 1$  galaxies placed at  $z = 2.75$  show if anything, that the observed increase in size may be more dramatic (see Appendix). These results also show that the one-component disc-like galaxies are larger on average than one-component spheroid-like galaxies.

We find a very interesting trend when we examine the evolution of the inner ‘bulge’ component and outer ‘disc’ component of two-component galaxies. We first note that the discs in two-component galaxies are larger in size than disc-like galaxies at all redshifts. The discs of the two-component galaxies increases in size on average from 1.6 to 3.2 kpc, i.e. an increase of a factor of 2. On the other hand, we find that the bulge components of the two-component systems increases very slightly from 0.9 to 1.1 kpc on average.

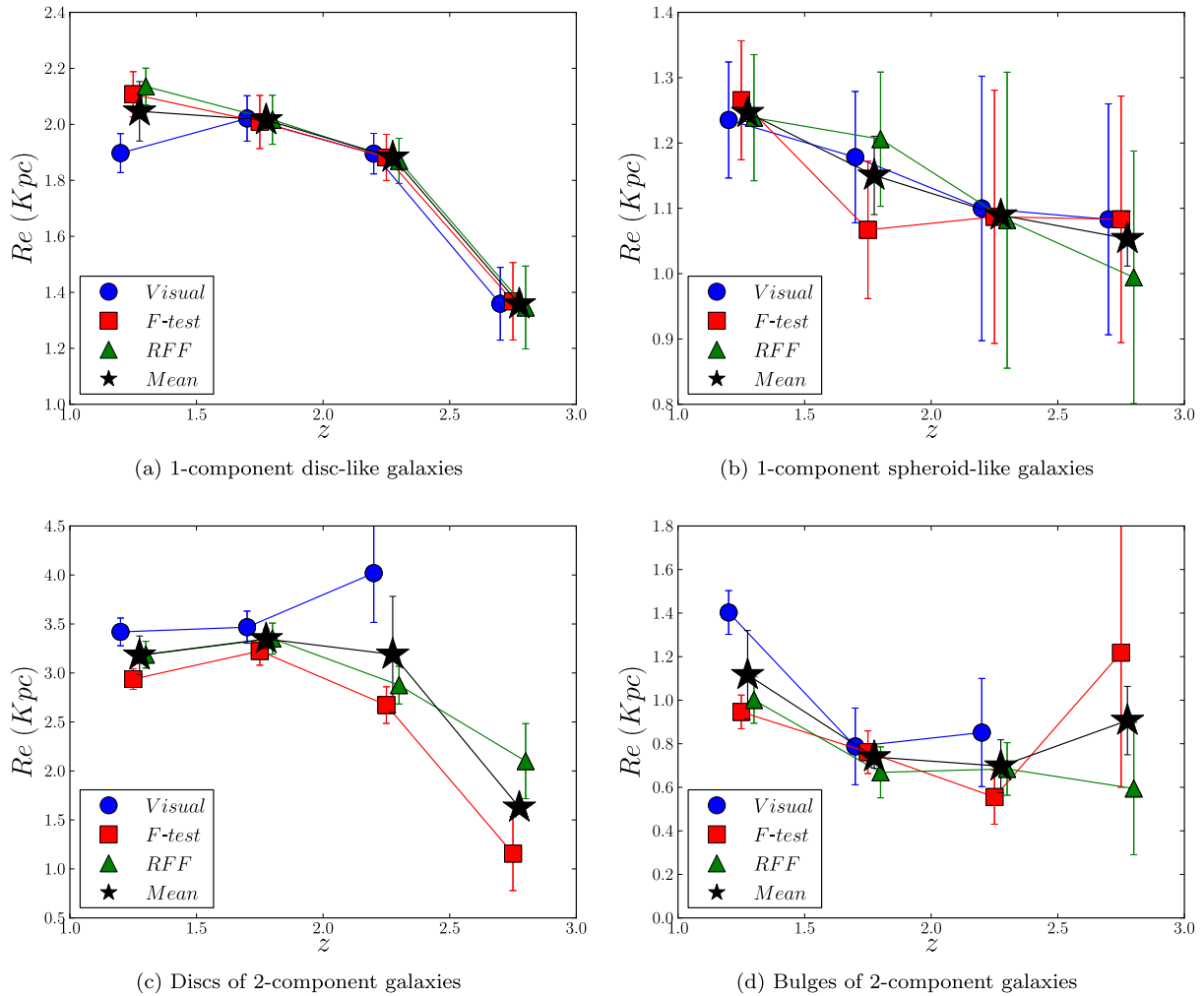
What we are likely seeing therefore is an inside-out formation of two-component galaxies such that the inner component is in place before the outer component. This is in agreement with previous studies such as van Dokkum et al. (2010) and Carrasco, Conselice & Trujillo (2010) which study the evolution of massive galaxies since  $z \sim 2$  at fixed aperture. Likewise, we find that the inner component of our sample does not grow as significantly as the outer component does. This is seen more clearly in Fig. 15, where we plot the ratio between the mean effective radius of the inner and outer component as a function of redshift. Discs seem to grow earlier, and they increase at a higher rate relative to the growth of the bulges until  $z \sim 1.5$ , when bulges appear to rapidly grow in size. We only have, however, one point showing this, which will need to be further confirmed. Nevertheless, in Fig. 16, we observe that the ratio between the flux of these two components remains fairly constant with redshift.

Overall, what we find is that the stellar masses, as traced by the light distribution (assuming a similar mass to light ratio), of these components are increasing at a similar rate, while at the same time, the sizes of the outer components are growing faster than their inner components. There are several ways to interpret this. One possibility to explain these observations is through how the additional mass is distributed in the two components. For the outer components, this new mass is added to the outer parts, increasing the size of the disc component, but within the bulges, the additional mass is still centrally concentrated. This is one way in which the mass ratio can remain constant while the size ratio increases with time. We will investigate this in more detail in a future paper (Margalef-Bentabol et al., in preparation).

It may be argued that the circularized effective radius  $R_{e,circ}$  is not the most appropriate quantity to measure the size of discs, as for flat discy objects,  $R_{e,circ}$  will depend on the inclination. In which case, the size could be better quantified by the effective radius  $R_e$  (major axis of an ellipse containing half of the light). However, we find that using  $R_e$  instead of  $R$  for our discy objects does not change our results. In the case of one-component discs, we observe the same growth of 60 per cent over the redshift range, but the sizes are about  $\sim 1.4$  times larger than the circularized radius. For discs of two-component galaxies, we observe an increase of a factor of 1.7 over the redshift range, which is still much stronger than the growth of the bulges at the same redshifts. In this case, the sizes are on average about 1.6 times greater than using the circularized values.

In Fig. 17, we plot the ratio between the sizes of the galaxy components (disc-like galaxies, spheroid-like galaxies, discs of two-component galaxies and bulges of two-component galaxies), and the sizes of galaxies within the nearby Universe at the same mass. These nearby galaxy sizes were obtained by the size–mass relation from GAMA results (Lange et al. 2015). We use the early-type relation to compare with spheroid-like galaxies, and the bulges of





**Figure 14.** Observed size evolution with redshift. Plotted are the median sizes of galaxies in our sample: disc-like (a), spheroid-like (b), discs of two-component galaxies (c) and bulges of two-component galaxies (d) for the three different selection methods (visual classification, *F*-test and RFF method). The points are coloured as in Fig. 7. In the y-axis, we plot the median sizes of the galaxies. Note that the visual classification does not select any two-component galaxies at the highest redshift bin, and therefore in the bottom plots, there is not a blue point at high redshift. The error bars on the points represent the standard deviation

two-component galaxies, and the late-type relation for comparison with disc-like galaxies and discs in two-component galaxies.

- (i) Early-type  $R_e(\text{Kpc}) = c \left( \frac{M_*}{M_\odot} \right)^\eta$
- (ii) Late-type  $R_e(\text{Kpc}) = d \left( \frac{M_*}{M_\odot} \right)^\lambda$ ,

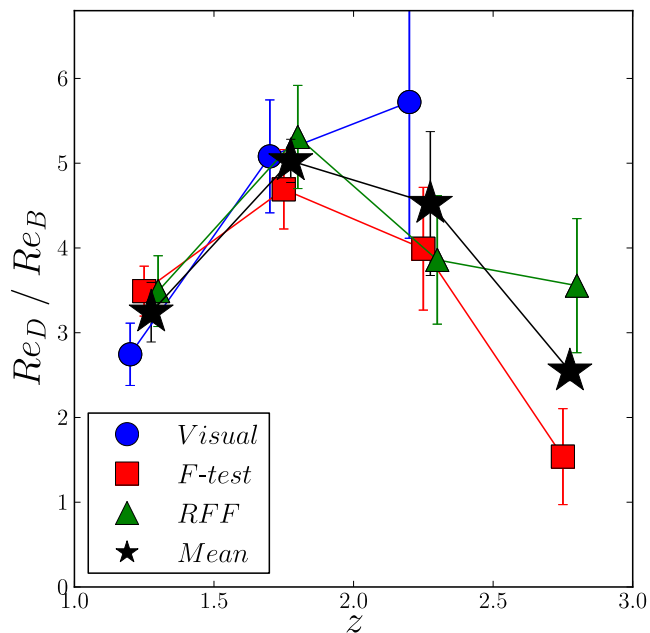
where  $c = 36.04 \times 10^{-5} \text{ Kpc}$ ,  $\eta = 0.38$ ,  $d = 25.26 \times 10^{-3} \text{ Kpc}$  and  $\lambda = 0.21$ .

We have overplotted the data from Buitrago et al. (2008) (galaxies in the same redshift range as our sample but with higher masses at  $M_* > 10^{11} M_\odot$ ) calculating the ratios in the same manner. Two-component discs seem to remain the same over redshift, while compared with their low-redshift counterparts, there may be a slight growth before  $z = 2.5$ , but they remain constant at lower redshifts. Disc-like galaxies are smaller at high redshift compared with disc-like galaxies of the same mass at the present time, in agreement with Buitrago et al. (2008). Spheroid-like galaxies also seem to have grown in size over cosmic time, but not as much as disc-like galaxies. The bulges of disc-like galaxies seem to have grown with time,

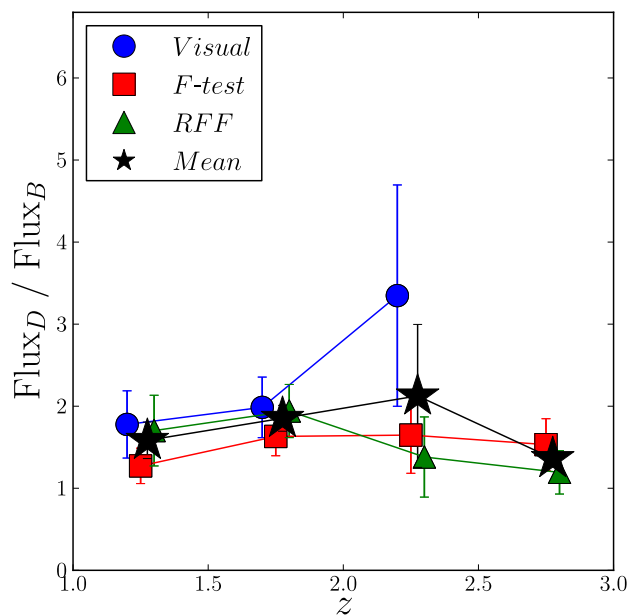
particularly after  $z = 2$ . We observe that size of spheroid-like galaxies and bulge of two-component galaxies is less than 40 per cent of that of early-type galaxies in the local Universe. This implies a significant growth from  $z = 1$  to present day for spheroid dominated galaxies. Disc-like galaxies undergo a less dramatic growth over the same epoch. This results are in agreement with previous studies, such as van der Wel et al. (2014), where they observe a growth by a factor of  $\sim 2$  since  $z = 1$  for early-type galaxies of similar masses, and only a moderate growth is seen for late-type galaxies.

## 5 SUMMARY AND CONCLUSIONS

We have carried out a detailed investigation of the light decomposition of galaxies within the CANDELS UDS field using 1495 massive galaxies with  $M_* > 10^{11} M_\odot$  at  $1 < z < 3$ . In this paper, we set out a new methodology for deciding whether a galaxy should be considered a single or two-component system using the observed *H*-band imaging, and then we examined the evolution of the individual component's sizes and mass as a function of redshift and stellar mass.

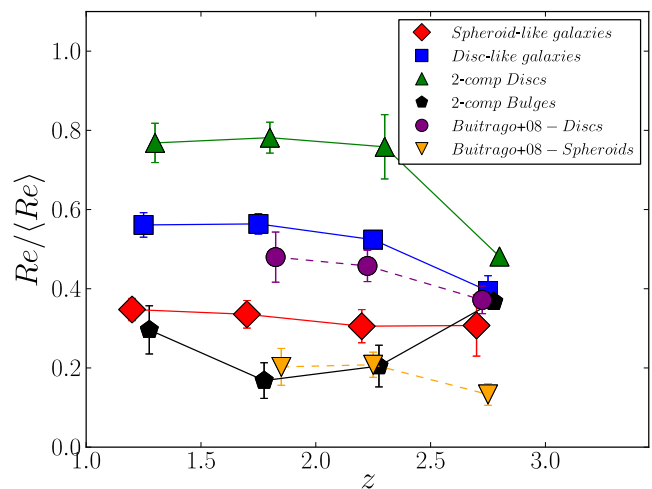


**Figure 15.** Evolution of the ratio between the sizes of discs and bulges of two-component galaxies. Note that the visual classification does not select any two-component galaxies at the highest redshift bin, and therefore there is not a blue point at high redshift.



**Figure 16.** Evolution of the ratio between the flux of discs ( $\text{Flux}_D$ ) and bulges ( $\text{Flux}_B$ ) of two-component galaxies. Note that the visual classification does not select any two-component galaxies at the highest redshift bin, and therefore there is not a blue point at the highest redshift bin.

We have used three different methods to determine if a galaxy is better fit by a one- or two-component model for their surface brightness distributions. The three methods are: visual morphology, the  $F$ -test, and by examining the RFF. We find that all three methods are largely in agreement with each other, and we calculate the mean value of the various parameters we study derived within the three methods.



**Figure 17.** Evolution of the ratio between the sizes of the galaxy components (disc-like galaxies, spheroid-like galaxies, disc of two-component galaxies and bulges of two-component galaxies), and the sizes of galaxies from the nearby Universe at the same mass ( $\langle R_e \rangle$ ).

One major result is that the fraction of two-component galaxies increases with higher stellar mass for all three methods. In fact, on average, there are  $\sim 2$  times more galaxies selected as two component for the most massive galaxies than in the lowest mass bin. We also find an evolution with redshift, such that the fraction of two-component systems decreases from about 35 per cent to 8 per cent from  $z = 1$  to 3. However, this decrease with redshift might be partially due to the degraded data at higher redshifts.

We find that disc-like galaxies have the highest relative number density at all redshifts, while spheroid-like galaxies have the lowest increase in that epoch, and by  $z \sim 2$ , two-component galaxies exceed both of them in number density. The contribution to the total density due to two-component galaxies becomes dominant at  $z < 1.75$  in spite of their being the lowest at  $z > 1.75$ . At redshift  $z \sim 3$ , the majority of spheroid-like galaxies are blue, but as redshift decreases, the number of red spheroid-like galaxies rapidly increases. The other populations of galaxies remain mostly blue for all redshifts.

We also find that for two-component galaxies, there is an increase in the sizes of their outer components, or ‘discs’ by about a factor of 2 from  $z = 3$  to 1, while the inner components or ‘bulges’ stays roughly the same size. This suggests that these systems are growing from the inside out, whilst the bulges are in place early in the history of these galaxies. This is also seen to a lesser degree in the growth of single ‘disc-like’ galaxies versus ‘spheroid-like’ galaxies over the same epoch.

We also carry out image simulations to determine how reliable our results are. We do this by reproducing how our galaxies would look at higher redshifts (for more details, see Appendix), we conclude that the decrease in size we observe within discs and bulges for two-component galaxies in Fig. 14 must be real, as the simulations show that we can accurately recover the size of the bulges in the simulated redshifted galaxies, including the smallest ones. Discs are also well recovered except for the smallest ones ( $R_{eD} < 2$  Kpc), where we recover larger discs than the originals. However, this may hint that we are indeed observing small discs at the highest redshift bin. The simulations where the  $F$ -test is used to find two-component galaxies may induce us to think that the decreasing in the fraction of two-component galaxies is due to simply redshift. However, the visual classification, which is a reliable tool to distinguish patterns, only accounts for half of the decreasing, suggesting that the observed

reduction is real (see Fig. A5). It thus seems more reasonable to think that the  $F$ -test is not as reliable using the same threshold  $P_0$  for all redshifts. This might be due to the fact that for high redshift, galaxies features are blended with the noise.

## ACKNOWLEDGEMENTS

We thank the CANDELS team for their support in making this paper possible, as well as STFC and the University of Nottingham for financial support. AM acknowledges funding from the STFC and a European Research Council Consolidator Grant (P.I. R. McLure).

## REFERENCES

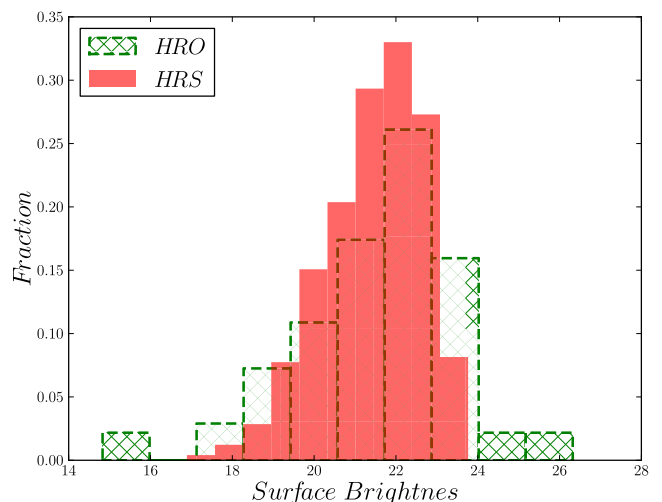
- Ashby M. L. N. et al., 2013, *ApJ*, 769, 80  
 Barden M., Jahnke K., Häußler B., 2008, *ApJS*, 175, 105  
 Barden M., Häußler B., Peng C. Y., McIntosh D. H., Guo Y., 2012, *MNRAS*, 422, 449  
 Barro G. et al., 2013, *ApJ*, 765, 104  
 Bertin E., Arnouts S., 1996, *A&AS*, 117, 393  
 Bluck A. F. L., Conselice C. J., Buitrago F., Grützbauch R., Hoyos C., Mortlock A., Bauer A. E., 2012, *ApJ*, 747, 34  
 Brammer G. B., van Dokkum P. G., Coppi P., 2008, *ApJ*, 686, 1503  
 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, *ApJ*, 687, L61  
 Carrasco E. R., Conselice C. J., Trujillo I., 2010, *MNRAS*, 405, 2253  
 Cassata P. et al., 2011, *ApJ*, 743, 96  
 Chabrier G., 2003, *PASP*, 115, 763  
 Conselice C. J., Bershadsky M. A., Dickinson M., Papovich C., 2003, *AJ*, 126, 1183  
 Conselice C. J., Blackburne J. A., Papovich C., 2005, *ApJ*, 620, 564  
 Conselice C. J., Rajgor S., Myers R., 2008, *MNRAS*, 386, 909  
 Conselice C. J. et al., 2011, *MNRAS*, 413, 80  
 Daddi E. et al., 2005, *ApJ*, 626, 680  
 Ferguson H. C. et al., 2004, *ApJ*, 600, L107  
 Fontana A. et al., 2014, *A&A*, 570, A11  
 Galametz A. et al., 2013, *ApJS*, 206, 10  
 Graham A. W., Driver S. P., 2005, *Publ. Astron. Soc. Aust.*, 22, 118  
 Grogin N. A. et al., 2011, *ApJS*, 197, 35  
 Hartley W. G. et al., 2013, *MNRAS*, 431, 3045  
 Hoyos C. et al., 2011, *MNRAS*, 411, 2439  
 Hoyos C. et al., 2012, *MNRAS*, 419, 2703  
 Huertas-Company M. et al., 2016, *MNRAS*, preprint ([arXiv:1606.04952](https://arxiv.org/abs/1606.04952))  
 Kartaltepe J. S. et al., 2015, *ApJS*, 221, 11  
 Koekemoer A. M. et al., 2011, *ApJS*, 197, 36  
 Lang P. et al., 2014, *ApJ*, 788, 11  
 Lange R. et al., 2015, *MNRAS*, 447, 2603  
 Mortlock A. et al., 2013, *MNRAS*, 433, 1185  
 Mortlock A. et al., 2015, *MNRAS*, 447, 2  
 Ownsworth J. R., Conselice C. J., Mundy C. J., Mortlock A., Hartley W. G., Duncan K., Almaini O., 2016, *MNRAS*, preprint ([arXiv:1605.05741](https://arxiv.org/abs/1605.05741))  
 Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2002, *AJ*, 124, 266  
 Press W. H., Flannery B., Teukolsky S., Vetterling W., 1988, *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge Univ. Press, Cambridge  
 Schechter P., 1976, *ApJ*, 203, 297  
 Sérsic J. L., 1968, *Atlas de galaxies australes*, preprint ([arXiv:e-prints](https://arxiv.org/abs/e-prints))  
 Simard L., Mendel J. T., Patton D. R., Ellison S. L., McConnell A. W., 2011, *ApJS*, 196, 11  
 Taylor-Mager V. A., Conselice C. J., Windhorst R. A., Jansen R. A., 2007, *ApJ*, 659, 162  
 Toft S. et al., 2007, *ApJ*, 671, 285  
 Trujillo I. et al., 2006, *ApJ*, 650, 18  
 Trujillo I., Conselice C. J., Bundy K., Cooper M. C., Eisenhardt P., Ellis R. S., 2007, *MNRAS*, 382, 109  
 van der Wel A. et al., 2011, *ApJ*, 730, 38  
 van der Wel A. et al., 2012, *ApJS*, 203, 24  
 van der Wel A. et al., 2014, *ApJ*, 788, 28  
 van Dokkum P. G. et al., 2010, *ApJ*, 709, 1018  
 van Dokkum P. G. et al., 2014, *ApJ*, 791, 45  
 van Rijsbergen C., 1979, *Information Retrieval*, 2nd edn. Butterworth-Heinemann, Newton, MA  
 Zhao D., Aragón-Salamanca A., Conselice C. J., 2015, *MNRAS*, 448, 2530

## APPENDIX

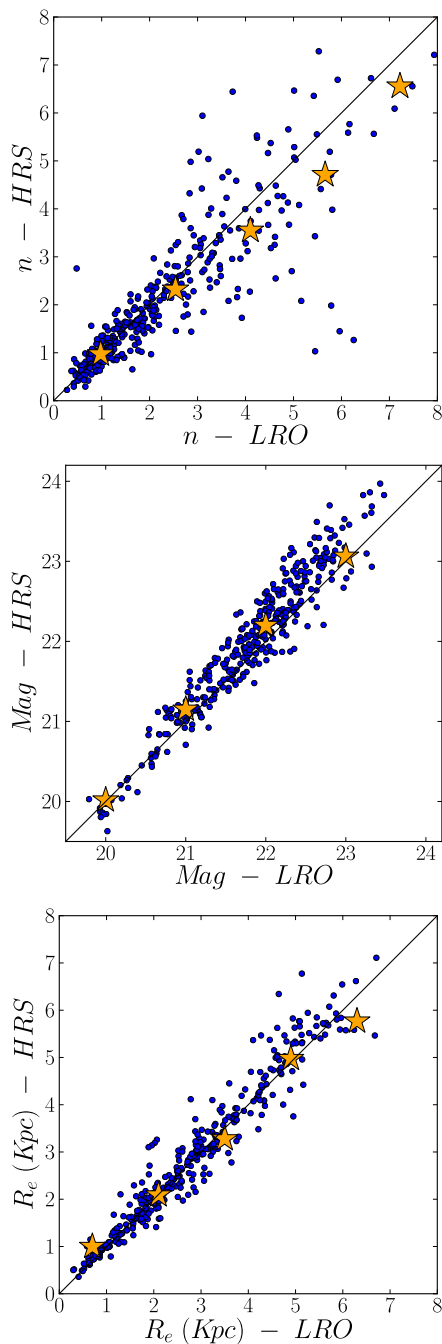
### A1 Details of the simulations

Many of the results presented in the main part of this paper originate from some mixture of redshift effects as well as real changes to the one- and two-dimensional structure of massive galaxies. It is important to separate these two effects to determine the real evolution of galaxies. As such, we artificially simulate redshifted galaxies using the *FERENGI* code from Barden, Jahnke & Häußler (2008). From our whole galaxy sample, we take a subsample of 357 objects with redshifts  $z < 1.4$  that we denote LRO (low-redshift original galaxies). For those LRO galaxies, we create new images of how we would observe the same galaxies if they were at redshift  $z = 2.75$  in the same CANDELS field, to determine the maximal effects of redshift. We denote this new sample HRS (high-redshift simulated galaxies). This procedure modifies the angular size and the surface brightness (dimming) due to cosmological effects, but also takes into account the brightness increase of high-redshift objects. We now compare the structures of the HRS galaxies with their corresponding LRO ones, as well as with the actual galaxies we examine in the main paper in the redshift range of  $2.5 < z < 3$ , that we denote HRO (high-redshift original galaxies).

To demonstrate that our simulated galaxies are a fair comparison to the actual high-redshift galaxies, we first compare the surface brightness distribution for both samples. In Fig. A1, we show the surface brightness of the HRS sample compared with the HRO sample, which shows that there is a good overlap between the two and therefore that, at least in the respect of magnitude and size distributions, these two samples can be compared.



**Figure A1.** Comparison of the surface brightness of the HRS (high-redshift simulated) galaxies (red) with the HRO (high-redshift original) galaxies at redshift  $2.5 < z < 3$  (green).



**Figure A2.** Comparison between the results from the fitting for the HRS galaxies versus the LRO sample, for the Sérsic index (top), magnitude (middle), and effective radii (bottom). The effect on the HRS magnitudes due to luminosity evolution and surface brightness dimming, introduced in the simulations, is removed to better compare with the LRO. The yellow stars represent the mean values for the HRS galaxies.

## A2 Results

Once we have our simulated subsample of HRS galaxies, we apply the same method described through the paper to determine whether each galaxy should be classified as one- or two-component system. The process for this can be summarized as follows.

(i) We fit the surface brightness with a Sérsic function with a free  $n$  for a one-component fit.

**Table A1.** Fraction of two-component galaxies selected by the three methods for different subsample of galaxies.

	LRO galaxies	HRS galaxies	HRO galaxies
Visual	0.38	0.17	0.00
$F$ -test	0.50	0.04	0.21
RFF	0.36	0.10	0.31

(ii) We fit the surface brightness with a Sérsic + Exponential function as a two-component fit.

(iii) We use our three methods to classify the galaxies as one or two components: the visual classification,  $F$ -test and RFF method.

A comparison of the Sérsic index, magnitude and effective radii between the HRS and LRO for one-component galaxies is shown in Fig. A2. We observe that these quantities behave as expected after the simulation. We recover, on average, the values of Sérsic index and effective radius, as well as the apparent magnitude. Note, for instance, that although the Sérsic index is not as well preserved for high values, the classification due to such an index is preserved, namely 80 percent of galaxies with  $n > 2.5$  for LRO still have  $n > 2.5$  after the simulation.

After applying the three methods to classify the galaxies as one or two components, in Table A1, we summarize the difference in the fraction of two-component galaxies in the HRS with respect to the LRO and the HRO.

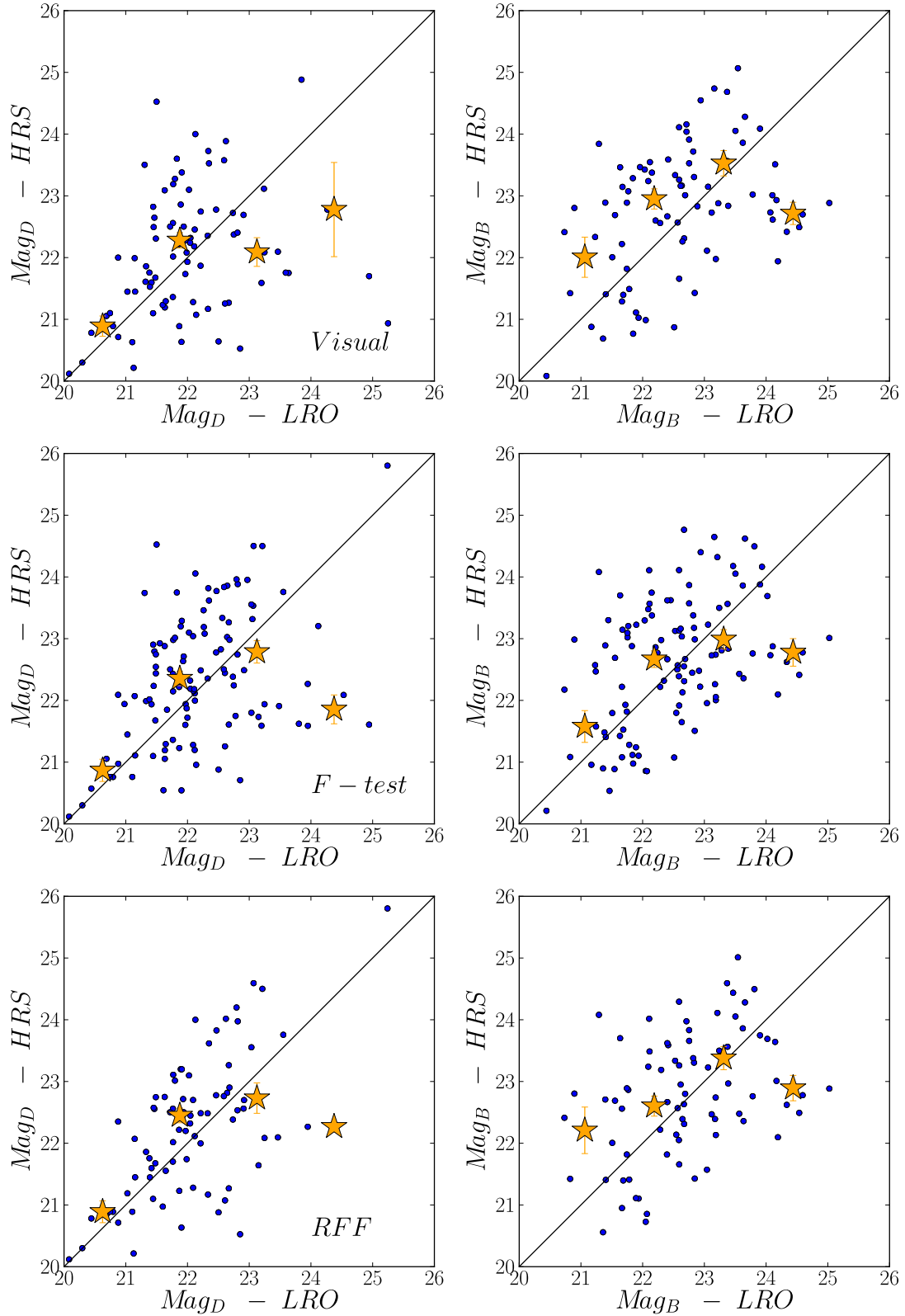
An analogous comparative of the magnitude and effective radii between the HRS and LRO for two-component galaxies is shown in Figs A3 and A4. Note, however, that while in Fig. A2 we considered the whole sample LRO versus HRS, now only the ones of LRO classified as two-component (together with the corresponding ones of HRS) are taken into account. Thus we obtain for each method two symbols for the magnitude (one for each of the two components), and two for the effective radii.

Even though there is some scatter in the plots of Fig. A4, and the errors are in some cases large, the mean values of the effective radii in HRS galaxies, for both components, are generally in agreement with the measures obtained for the LRO ones. For smaller values of the outer component in the LRO, the simulations show that we would measure larger effective radii at high redshift, which suggests that the small values of sizes we obtain for the HRO must be real.

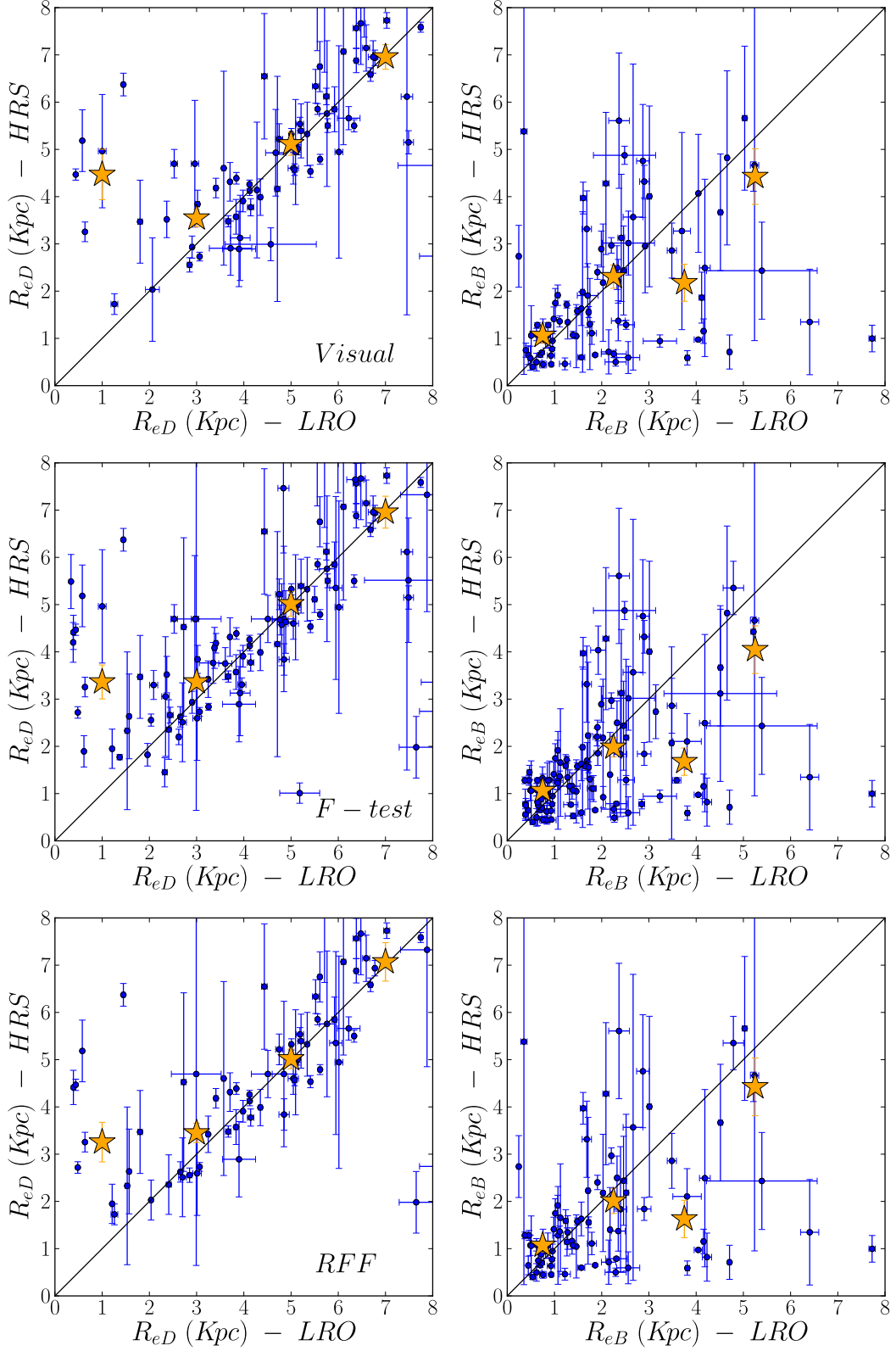
Fig. A5 also shows that the fraction of HRS two-component galaxies (yellow diamond) is similar (although slightly higher) to that of the HRO (black star at  $z \sim 2.75$ ), which might imply that most of the evolution we see in that fraction with redshift may not be real but due to resolution and depth issues at high redshift. This shows the difficulty of carrying out a study of structure at high redshift. However, if we consider only the visual method, we observe that the evolution could be real. It is interesting to note that while for the HRO the visual classification fails to select any two-component galaxies, it selects almost 20 per cent in the HRS.

In Fig. A6, we plot the fraction of two-component HRS galaxies as a function of mass. As in Fig. 7, we observe that the fraction increases with higher masses in a similar manner, but is about 2 times lower than for the original galaxies (note that in the HRO, we detect about half of the galaxies selected as two-components in the LRO sample). As we detect a larger fraction of two-component galaxies in the original sample than in the HRS galaxies, this implies that the high fraction we have found at higher masses must be real.

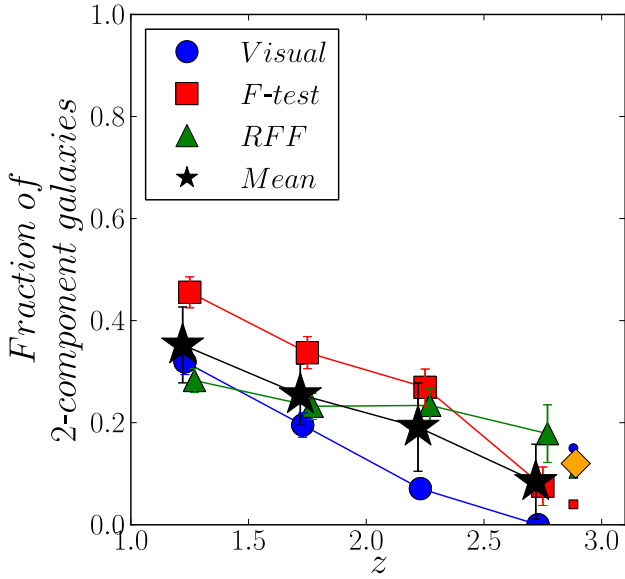




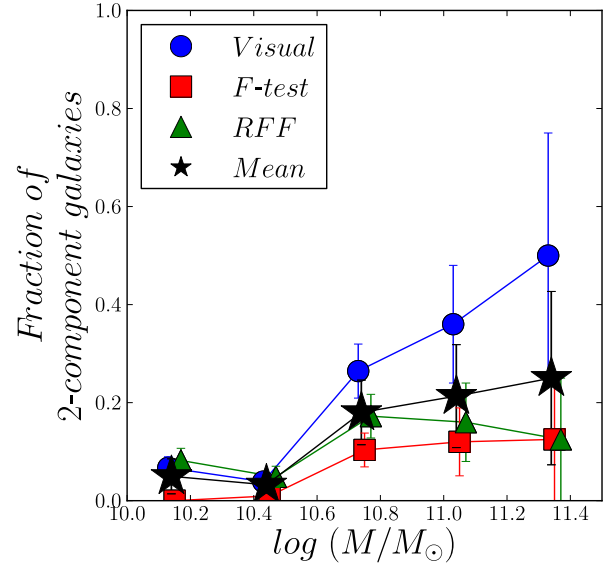
**Figure A3.** Magnitude of the disc-like outer (left) and bulge-like inner (right) component of high-redshift simulated versus low-redshift original galaxies for each method (top panel: visual classification, middle panel:  $F$ -test and bottom panel: RFF method). The effect on the HRS magnitudes due to luminosity evolution and surface brightness dimming, introduced in the simulations, is removed to better compare with the LRO. The yellow stars represent the mean value of the magnitudes for the HRS galaxies.



**Figure A4.** Effective radii of the disc-like outer (right) and bulge-like inner (left) component of high-redshift simulated versus low-redshift original galaxies for each method (top panel: visual classification, middle panel:  $F$ -test and bottom panel: RFF method). The yellow stars represent the mean value of the effective radii for the HRS galaxies.



**Figure A5.** Comparison of the fraction of two-component galaxies between observations and simulations. The bigger blue circles, red squares and green triangles represent the fractions of two-component galaxies, from the whole sample, as function of redshift. The black stars show the mean values of the three methods (same as Fig. 8). The smaller blue circle, red square and green triangle represent the fractions of two-component HRS galaxies for each method, and the yellow diamond is the mean between the Visual and RFF method.



**Figure A6.** Fraction of two-component galaxies as a function of mass for the HRS-simulated galaxies. The points are coloured as in Fig. 7.

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.