Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **465,** 2717–2733 (2017) doi:10.1093/mnras/stw2933
Advance Access publication 2016 November 14

The structural and size evolution of star-forming galaxies over the last 11 Gyr

Ana Paulino-Afonso,[1,2]★ David Sobral,[3,4] Fernando Buitrago[1,2] and José Afonso[1,2]

[1]*Instituto de Astrofísica e Ciências do Espaço, Universidade de Lisboa, OAL, Tapada da Ajuda, PT1349-018 Lisboa, Portugal*
[2]*Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Edifício C8, Campo Grande, PT1749-016 Lisboa, Portugal*
[3]*Department of Physics, Lancaster University, Lancaster LA1 4YB, UK*
[4]*Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands*

Accepted 2016 November 10. Received 2016 November 10; in original form 2016 July 5

ABSTRACT

We present new results on the evolution of rest-frame blue/UV sizes and Sérsic indices of Hα-selected star-forming galaxies over the last 11 Gyr. We investigate how the perceived evolution can be affected by a range of biases and systematics such as cosmological dimming and resolution effects. We use GALFIT and an artificial redshifting technique, which includes the luminosity evolution of Hα-selected galaxies, to quantify the change on the measured structural parameters with redshift. We find typical sizes of 2–3 kpc and Sérsic indices of $n \sim 1.2$, close to pure exponential discs all the way from $z = 2.23$ to $z = 0.4$. At $z = 0$, we find typical sizes of 4–5 kpc. Our results show that, when using GALFIT, cosmological dimming has a negligible impact on the derived effective radius for galaxies with <10 kpc, but we find an $\sim$20 per cent bias on the estimate of the median Sérsic indices, rendering galaxies more disc-like. Star-forming galaxies have grown on average by a factor of 2–3 in the last 11 Gyr with $r_e \propto (1 + z)^{-0.75}$. By exploring the evolution of the stellar mass–size relation, we find evidence for a stronger size evolution of the most massive star-forming galaxies since $z \sim 2$, as they grow faster towards $z \sim 0$ when compared to the lower stellar mass counterparts. As we are tracing the rest-frame blue/UV, we are likely witnessing the growth of discs where star formation is ongoing in galaxies while their profiles remain close to exponential discs, $n \lesssim 1.5$, across the same period.

Key words: galaxies: evolution – galaxies: star formation – galaxies: structure.

1 INTRODUCTION

Ever since the first classification schemes based on the visual appearance of galaxies were created (e.g. Hubble 1926; de Vaucouleurs 1959), the ways we study galaxy morphology have evolved dramatically. On one hand, we have improved on the quantification of the light distribution in galaxies either using parametric surface brightness profiles (e.g. Caon, Capaccioli & D'Onofrio 1993; Simard 1998; Peng et al. 2002, 2010) or non-parametric approaches (e.g. Abraham et al. 1994; Conselice 2003; Lotz, Primack & Madau 2004; Law et al. 2007; Freeman et al. 2013). On the other hand, there was the need to create new classifications as galaxies become more and more irregular towards higher redshift (e.g. Delgado-Serrano et al. 2010; Huertas-Company et al. 2015).

Although the observed morphology may not be directly linked with intrinsic properties of the stellar populations and dust/gas content of galaxies (see e.g. Conselice 2014, and references therein), early studies have shown that rest-frame optical morphology correlates with colour and star formation activity (e.g. Holmberg 1958), and there is a marked difference in the prevalence of different morphological populations in different environments (e.g. Dressler 1980). These effects are seen both in the local (e.g. Ball, Loveday & Brunner 2008; Bamford et al. 2009) and in the higher redshift Universe (e.g. Pérez-González et al. 2008; Viero et al. 2012; Bassett et al. 2013). Additionally, there was significant work regarding correlations between the shape of a galaxy and other physical properties such as stellar populations, mass and star formation (e.g. Roberts & Haynes 1994; Blanton et al. 2003; Conselice 2003; Wuyts et al. 2011; Whitaker et al. 2015). Processes such as galaxy mergers, *in situ* star formation and accretion of intergalactic gas can be revealed by a detailed structural analysis of galaxy samples.

The peak of star formation in the Universe occurred at $z \sim 2$ ($\sim$11 Gyr ago, e.g. Madau & Dickinson 2014, and references therein), and it is also since this peak of activity that most of the structures (disc and spheroidal galaxies) that we observe today have been formed (e.g. Buitrago et al. 2013; Mortlock et al. 2013). However, to understand how the baryonic structures grew between different cosmological epochs, it is not only necessary to study morphology, but also couple that with kinematic information.

★ E-mail: aafonso@oal.ul.pt

In the recent years, it has been possible to study the interplay between galaxy morphology and kinematics by making use of the new available Integral Field Unit (IFU) instruments. These allowed for large galaxy surveys either in the local Universe [e.g. CALIFA, Sánchez et al. (2012); SAMI, Bryant et al. (2015); MaNGA Bundy et al. (2015)] and at high redshift [e.g. KROSS, Stott et al. (2016); KMOS^{3D}, Wisnioski et al. (2015)] and added valuable information that will provide key insights on the physics that drives galaxy evolution.

Despite the large potential for progress, to connect observed properties across a large span of time we need to account for biases and systematics, which can arise either due to selection or instrumental and/or cosmological effects. To overcome these problems, we need both large and homogeneous surveys at various cosmic epochs to minimize the impact of cosmic variance and to probe a wide range of galaxy properties and environments. This is now possible when using surveys based on well-known and calibrated physical properties over a wide range of redshifts, such as Hα in narrow-band surveys (see e.g. Moorwood et al. 2000; Geach et al. 2008; Villar et al. 2008; Sobral et al. 2009, 2011; Ly et al. 2011) and in spectroscopic/grism surveys (see e.g. McCarthy et al. 1999; Yan et al. 1999; Hopkins, Connolly & Szalay 2000; Shim et al. 2009; Atek et al. 2010; van Dokkum et al. 2011).

None the less, even with an ideal, homogeneous sample, we still need to account for all effects that have a dependence on redshift as they could mimic and/or influence evolutionary trends that we observe. One of the strongest effects that impact the study of galaxy morphology is the surface brightness dimming (Tolman 1930). Between $z \sim 0$ and $z \sim 2$, this effect impacts the observed fluxes by two orders of magnitude. The first attempts to describe the impact of the surface brightness dimming on how galaxies would be observed at high redshift if they were as we see them today were conducted by Weedman & Huenemoerder (1985). Later on, studies on the impact on galaxy visual morphology (e.g. Giavalisco et al. 1996; Hibbard & Vacca 1997) and on the morphology quantifiers (e.g. Trujillo et al. 2007; Barden, Jahnke & Häußler 2008; Petty et al. 2009; Weinzirl et al. 2011; Mosleh, Williams & Franx 2013) were also carried out and find no systematics and errors $\lesssim$15 per cent on the derived sizes at $z \sim 1$. None the less, these studies are often limited to small samples and comparison between two distinct epochs (one local and one at high redshift).

There are numerous studies reporting on size and structural evolution of galaxies (e.g. Ferguson et al. 2004; Ravindranath et al. 2004; Trujillo et al. 2006; Buitrago et al. 2008; Cimatti et al. 2008; Franx et al. 2008; van Dokkum et al. 2010; Mosleh et al. 2011; Wuyts et al. 2011; Ichikawa, Kajisawa & Akhlaghi 2012; Böhm & Ziegler 2016). For star-forming galaxies (SFGs) in the range $0.5 < z \lesssim 3$, studies find moderate (e.g. $r_e \propto (1+z)^{\alpha}$; Buitrago et al. 2008; Morishita, Ichikawa & Kajisawa 2014; van der Wel et al. 2014, with $\alpha = -0.82, -0.75, -0.57$, respectively) to negligible size evolution (Stott et al. 2013) and light profiles close to exponential discs (e.g. Morishita, Ichikawa & Kajisawa 2014; Shibuya, Ouchi & Harikane 2015). For $2 < z < 4.5$, it was shown that the trend on size evolution (measured from rest-frame UV) depends on the method used to derive galaxy sizes (Ribeiro et al. 2016). And, at even higher redshifts ($4 < z < 8$), Curtis-Lake et al. (2016), using FUV rest-frame galaxy sizes, show that the derived evolution depends on the statistical estimators one uses. These evolutionary trends of galaxy growth and the relation of sizes with stellar masses are also found in large-scale cosmological simulations (e.g. Genel et al. 2014; Furlong et al. 2015). However, we do not know yet if such differences can be explained by different selection methods for the definition of the SFGs samples (e.g. Oteo et al. 2015).

Although there are existing studies on the morphologies of SFGs (e.g. Morishita et al. 2014; van der Wel et al. 2014; Shibuya et al. 2015) and on the quantification of systematic differences of structural properties at low and high redshift (e.g. Barden et al. 2008; Petty et al. 2009; Weinzirl et al. 2011; Mosleh et al. 2013), it is still unclear what the role of potential biases and systematics may be. In this paper, we take advantage of a unique Hα selection, along with the largest IFU samples, to make further progress on the current open questions.

This paper is organized as follows. In Section 2, we describe the samples of SFGs that will be used throughout the paper. We present our methodology to simulate galaxies at high redshift and to study their structural parameters in Sections 3 and 4. The results obtained for our low- and high-redshift SFGs and simulations are reported in Section 5. We discuss the implications of our results in the context of galaxy evolution in the last $\sim$11 Gyr in Section 6. Finally, in Section 7, we summarize our conclusions. Magnitudes are given in the AB system (Oke & Gunn 1983). All the results assume a Λcold dark matter cosmological model with $H_0 = 70.0\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$ and a Chabrier (2003) initial mass function.

2 GALAXY SAMPLES

The main sample of SFGs that we use in this paper comes from the High-Z Emission Line Survey (HiZELS, Sobral et al. 2013a). Being the largest Hα narrow-band survey at high redshift, it provided targets to be observed with IFU instruments such as VLT/SINFONI (Swinbank et al. 2012a,b) and VLT/KMOS (Sobral et al. 2013b; Stott et al. 2014, 2016). Other samples are selected from the currently ongoing IFU surveys in the local Universe (see Section 2.1). However, local samples differ from a simple Hα selection as done in HiZELS. To ensure that we are studying comparable populations, we apply simple sample selection based on stellar mass and Hα luminosity:

$$\log_{10}(M_\star/\mathrm{M}_\odot) > 9 \quad \wedge \quad L_{\mathrm{H}\alpha} > 0.1 L^*_{\mathrm{H}\alpha}(z), \tag{1}$$

where the luminosity cut is taken from the equation for the redshift evolution of $L^*_{\mathrm{H}\alpha}$ derived by Sobral et al. (2013a)

$$\log_{10}(L^*_{\mathrm{H}\alpha}(z)) = 0.45z + 41.87. \tag{2}$$

We choose to focus our study by selecting samples through their Hα luminosities since such samples should be representative of the full star-forming population (e.g. Oteo et al. 2015). Note that for the local Universe samples, with the exception of the CALIFA survey, we only have available Hα luminosities measured inside a 3 arcsec fibre. We thus apply aperture corrections following e.g. Garn & Best (2010) by computing the flux ratio of the total and fibre magnitudes in the Sloan Digital Sky Survey (SDSS) *r*-band (the filter that contains Hα) and applying that correction to the observed fibre-based Hα luminosity. These correction factors vary from $\sim$1.5 up to $\sim$40. Our samples and the selection criteria are shown in Fig. 1. For a quick summary, we refer the reader to Table 1.

2.1 The low-redshift samples

2.1.1 The CALIFA sample

The Calar Alto Legacy Integral Field spectroscopy Area (CALIFA; Sánchez et al. 2012) survey is a programme conducted using the Potsdam Multi-Aperture Spectrophotometer with PMAS fibre

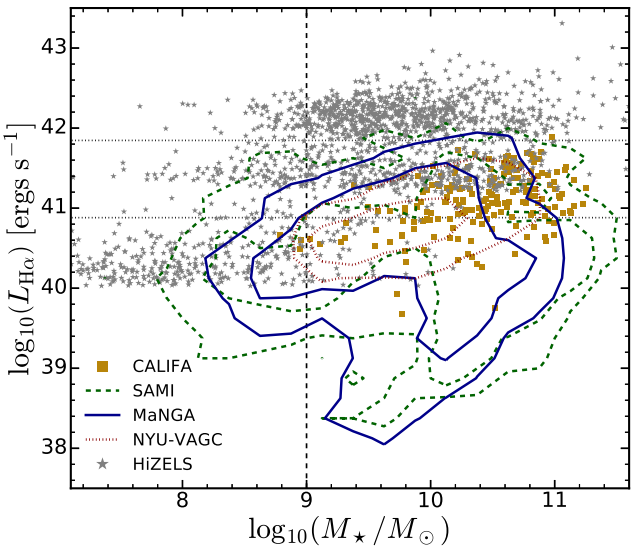

Figure 1. The stellar masses and aperture corrected observed Hα luminosities of the parent samples used in this paper. The contour lines for SAMI, MaNGA and NYU-VAGC show the limits that contain 68 per cent and 95 per cent of the sample for two-dimensional histograms with 0.25 width bins in both stellar mass and Hα luminosity. The vertical dashed line shows the stellar mass cut used for the final sample selection. We see that most our samples are above the stellar mass limit imposed in our selection and that the local redshift samples overlap. The horizontal dotted lines show the Hα limits at $z = 0.01$ (lower line) and at $z = 2.23$ (upper line). This shows that the Hα luminosity selection as a great impact on the final samples that we study.

PAcK (PMAS/PPAK) integral field spectrophotometer mounted on the 3.5 m Calar Alto telescope. The survey aims at observing ∼600 galaxies in the local Universe (redshift range $0.005 < z < 0.03$), which are selected from the SDSS as a sample limited in apparent diameter (45 arcsec $< D <$ 80 arcsec). This constraint assures that galaxies are observed within the large field of view (∼1 arcmin^2) with a large covering fraction and high spectral resolution.

We use the reported values by Catalán-Torrecilla et al. (2015) for the Hα luminosities of this sample, which are available for 270 galaxies. By taking the selection criteria defined in equation (1), we reach a final sample of 137 CALIFA galaxies (see Fig. 1).

2.1.2 *The SAMI target sample*

The Sydney-AAO Multi-object Integral field spectrograph (SAMI; Bryant et al. 2015) Galaxy survey proposes to target 3400 galaxies with the SAMI instrument mounted on the Anglo-Australian Telescope (AAT). This survey targets galaxies in the redshift range $0.004 < z < 0.095$, SDSS magnitudes $r_{\rm pet} < 19.4$, from low (10^7 M$_\odot$) to high stellar mass (10^{12} M$_\odot$), both isolated, in groups or members of clusters with halo masses of ∼10^{15} M$_\odot$. Most of the targets (with the exception of few cluster objects) have available SDSS coverage and are selected from the Galaxy and Mass Assembly survey (GAMA; Driver et al. 2009).

We use the GAMA Data Release 2 (Liske et al. 2015) to obtain Hα luminosities for the SAMI sample. By taking the selection criteria defined in equation (1) we reach a final sample of 422 SAMI galaxies (see Fig. 1). We note that the SAMI sample is stellar mass complete for our mass limit only at $z < 0.5$. We include higher redshift galaxies but do not expect it to have a great impact on our derived results.

2.1.3 *A MaNGA-like sample*

The Mapping Nearby Galaxies at Apache point observatory (MaNGA; Bundy et al. 2015) survey is part of the SDSS-IV programme and aims to study kinematics and internal composition of a sample of ∼10 000 galaxies. It will do so with fibre-bundle IFUs with diameters ranging from 12 arcsec to 32 arcsec and will provide spectral information in the full optical range. The MaNGA sample is derived from an extended version of the NASA Sloan-Atlas (NSA), based on the SDSS DR7 Main Galaxy Sample (Abazajian et al. 2009) with the additions and improvements detailed by Blanton et al. (2011).[1] It will observe galaxies at redshifts $0.01 < z < 0.15$ with stellar masses above ∼10^9 M$_\odot$ and will make use of redshift and *i*-band luminosity to achieve a homogeneous radial coverage (see fig. 8 of Bundy et al. 2015), flat stellar mass distributions and a diversity of environments. An additional selection on colour space will enhance the targeting of rarer galaxies (green valley, low-mass red and massive blue galaxies).

Using the available data from NSA, we attempt to mimic the MaNGA selection by applying the selection bands as described in Bundy et al. (2015). We use the published version of the NSA table (restricting our galaxies to $z < 0.055$) and pre-select all galaxies that fall inside the selection bands of fig. 8 by Bundy et al. (2015). We randomly select ∼6000 galaxies (with uniform sampling up to $1.5r_{\rm e}$, primary selection criteria) and ∼2500 galaxies (with uniform sampling up to $2.5r_{\rm e}$, secondary selection criteria) to roughly match the sample numbers of MaNGA (Bundy et al. 2015). We neglect the colour enhanced selection. By taking the selection criteria defined in equation (1), we reach a final sample of 1536 MaNGA galaxies (see Fig. 1).

2.1.4 *The NYU-VAGC sample*

This sample is based on the New York University Value Added Galaxy Catalog (NYU-VAGC; Blanton et al. 2005a). A subset of the catalogue, which was constructed as a volume limited sample with well-defined selection criteria (see Blanton et al. 2005b), was chosen as a control sample so that we may inspect if the IFU samples are biased against a magnitude-selected sample. To complement the information, we matched the catalogue with the Max Planck for Astronomy & Johns Hopkins University Data Release 7 catalogues (MPA-JHU DR7; Kauffmann et al. 2003; Tremonti et al. 2004; Salim et al. 2007), which have spectroscopic information for SDSS DR7 galaxies (Abazajian et al. 2009).

We first do a pre-selection of SFGs with stellar masses $9 < \log_{10}(M_\star/{\rm M}_\odot) < 12$ and Hα luminosities greater than $L_{\rm H\alpha} > 10^{39.5}$ erg s^{-1} to exclude both low and high stellar masses and match the HiZELS detection limits. This results in a total of ∼13 000 galaxies. From this sample, we have randomly selected 10 per cent (1285) of all galaxies. This selection was performed by randomly picking galaxies with a probability matched to the magnitude, radius and Sérsic indices distribution available from the NYU-VAGC catalogue. With this method, we guarantee that we probe the full morphological parameter space using this subset.

We then restrict our sample to 412 NYU-VAGC galaxies (see Fig. 1) with aperture-corrected Hα luminosities and stellar masses matching our sample selection criteria defined in equation (1).

[1] http://www.nsatlas.org

Table 1. Samples studied in this paper. The median redshift, median stellar mass and median observed aperture corrected (for SAMI, MaNGA and NYU-VAGC) Hα luminosity (in erg s^{-1}) for all samples. The last column shows the ratio of the median observed aperture corrected (for SAMI, MaNGA and NYU-VAGC) Hα luminosity to the $L^*_{\rm H\alpha}$ at the median redshift. The value in brackets shows the median values for each subsample after applying the selection criteria defined in Section 2.

	Sample	$\tilde{z}$	$N_{\rm tot}[N_{\rm subsample}]$	$\log_{10}(M_\star/{\rm M}_\odot)$	$\log_{10}(L_{\rm H\alpha})$	$L_{\rm H\alpha}/L^*_{\rm H\alpha}(z)$
Local	CALIFA	0.015 [0.016]	541 [137]	10.44 [10.53]	41.08 [41.22]	0.16 [0.22]
	SAMI	0.039 [0.049]	2349 [422]	10.06 [10.26]	40.38 [41.22]	0.03 [0.21]
	MaNGA	0.030 [0.037]	8492 [1536]	9.85 [9.91]	40.25 [41.14]	0.02 [0.18]
	NYU-VAGC	0.041 [0.041]	1285 [412]	9.73 [10.08]	40.74 [41.05]	0.07 [0.14]
HiZELS	NB921	0.400	460 [33]	8.25 [9.83]	40.39 [41.31]	0.02 [0.18]
	NBJ	0.840	425 [309]	9.69 [9.96]	41.43 [41.45]	0.15 [0.16]
	NBH	1.470	313 [250]	9.65 [9.89]	42.11 [42.12]	0.38 [0.38]
	NBK	2.230	572 [526]	9.71 [9.74]	42.19 [42.19]	0.21 [0.21]

2.2 The high-redshift Universe

By using four narrow-band filters in the z, J, H and K-bands, HiZELS (Sobral et al. 2013a) has detected thousands of Hα emitters in four distinct redshifts intervals centred at 0.4, 0.84, 1.47 and 2.23. Such large samples, selected in a uniform way, allow one to probe galactic evolution across these slices of our Universe, which map the decline since the peak of star formation activity. To ensure that the selection is effective in picking up Hα emitters, the observations were conducted on well-studied extragalactic fields (UDS, COSMOS, SA22, ELAIS N1, Boötes and Lockman Hole), where the ancillary broad-band data helps at pinpointing Hα emitters by means of colour selections and photometric redshifts to allow for the control of the contamination rates (Sobral et al. 2013a).

In this paper, we make use of the list of emitters that are found in the COSMOS field (Sobral et al. 2013a) for which we have extensive coverage of *F*814*W* imaging obtained with the *Hubble Space Telescope* (*HST*; Scoville et al. 2007). After applying the selection cuts defined in equation (1), we reach final samples with 33 galaxies at $z = 0.4$, 309 galaxies at $z = 0.84$, 250 galaxies at $z = 1.47$ and 526 galaxies at $z = 2.23$ (see Table 1).

3 ARTIFICIAL REDSHIFTING GALAXIES

To understand how the perceived structural parameters have changed across cosmic time, we study how our local galaxies would look like if they were at high redshift. We explore and evaluate the effects of cosmological dimming on various properties of galaxies (e.g. size/shape measurements) taking into account that these are prone to resolution and bandpass issues.

By artificially redshifting galaxy images, we aim to address the biases and systematics of the parent sample and to build a coherent vision of the galaxies' structural evolution. To do that, we translate the core algorithm of FERENGI (Barden et al. 2008) into PYTHON and include a more general treatment of the intrinsic luminosity evolution of galaxies as a function of redshift. This treatment ensures that we are using an artificially redshifted sample that closely matches, in brightness, the sample that we have at higher redshifts. The core of the algorithm is shown as a diagram in Fig. 2 and is briefly summarized in the following steps:

(i) We re-scale our image (preserving their total flux) to match the pixel-scale of the high-redshift observations while preserving the physical scale of the galaxy.

(ii) We apply a flux correction to the image that is the combination of two factors: the dimming factor, which scales as the inverse of the luminosity distance to the galaxy, and the luminosity evolution factor, which account for the average evolution in brightness across different redshifts and is taken from Sobral et al. (2013a). Since we are studying rest-frame blue/NUV, we assume light scales with SFR/Hα.

(iii) We compute a transformation point spread function (PSF) following the prescription described by Barden et al. (2008) applied to the PSFs of the used instruments.

(iv) We convolve the re-scaled image with the PSF computed in the previous step and then we place the image on an empty region of the target high-redshift survey.

Note that we study rest-frame blue/NUV light for a sample of galaxies that are star-forming by selection. So, we expect that at these wavelengths the light will be dominated by young stars and we do not expect great differences across the rest-frame wavelengths that we cover (2500–5800 Å).

As an example, we show the final result of this algorithm applied in four different galaxies redshifted into four different redshifts, as shown in Fig. 3. We show that most of the fine substructures at low redshift are suppressed. The galaxies are still visible, albeit at a low surface brightness level, across all redshifts.

4 PARAMETRIC MORPHOLOGICAL ANALYSIS

One of the most common ways of characterizing the structure of galaxies is to fit a Sérsic (1968) profile to the surface brightness distribution of galaxy images (e.g. Davies et al. 1988; Caon, Capaccioli & D'Onofrio 1993; Andredakis, Peletier & Balcells 1995; Moriondo, Giovanardi & Hunt 1998; Simard 1998; Khosroshahi, Wadadekar & Kembhavi 2000; Graham 2001; Möllenhoff & Heidt 2001; Trujillo, Graham & Caon 2001; Peng et al. 2002; Blanton et al. 2003; Trujillo et al. 2007; Wuyts et al. 2011; van der Wel et al. 2014). This profile is characterized by the functional form

$$I(r) = I_{\rm e} \exp[-\kappa(r/r_{\rm e})^{1/n} + \kappa],$$

where the Sérsic index n describes the shape of the light profile, $r_{\rm e}$ is the effective radius of the profile, $I_{\rm e}$ is the surface brightness at radius $r = r_{\rm e}$ and κ is a parameter coupled to n such that half of the total flux is enclosed within $r_{\rm e}$. An index of $n = 1$ corresponds to a typical exponential disc galaxy, whereas $n = 4$ corresponds to the classical de Vaucouleurs profile associated with elliptical galaxies.

To conduct this parametric analysis, we make use of GALFIT (Peng et al. 2002, 2010), which is a public algorithm designed to perform a detailed two-dimensional decomposition of galaxies using mathematical models.

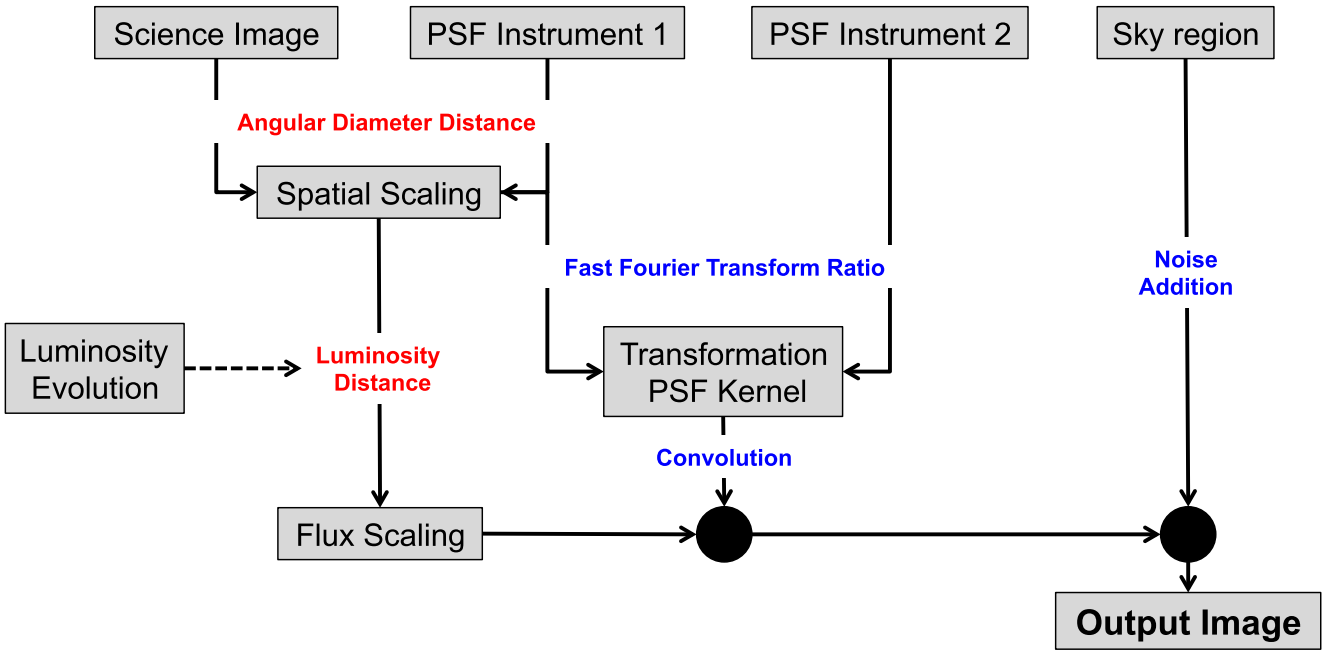

Figure 2. Schematic diagram of the algorithm for artificially redshift galaxies. Each step is described in Section 3.

On 2D images, each Sérsic model has potentially seven free parameters: the position of the centre, given by x_c and y_c, the total magnitude of the model, m_{tot}, the effective radius, r_e, the Sérsic index, n, the axial ratio of the ellipse, b/a, and the position angle, θ_{PA}, which refers to the angle between the major axis of the ellipse and the vertical axis and it has the sole purpose of rotating the model to match the galaxy's image.

Since GALFIT requires an initial set of parameters from which to start evaluating the model, it is necessary to provide it with realistic guesses. For that reason, we use the source extraction software SExtractor (Bertin & Arnouts 1996), which allows one to directly obtain a set of parameters that will serve as input to GALFIT and to find the model that best fits the data faster and with reliable values.

We use 10 arcsec × 10 arcsec cutouts of the *HST* ACS *F*814*W* (Koekemoer et al. 2007; Scoville et al. 2007) centred on each HiZELS galaxy. The cutouts are downloaded from the COSMOS *HST* archive. These images have a typical PSF FWHM (full width at half-maximum) of ∼0.09 arcsec, a pixel scale of 0.03 arcsec pixel^{-1} and a limiting point-source depth AB($F814W$) = 27.2 (5σ). For the low-redshift samples, we use images from SDSS DR9 (Ahn et al. 2012) *g*-band imaging data. These are characterized by a median PSF FWHM of ∼1.3 arcsec, a pixel scale of 0.396 arcsec pixel^{-1} and a 95 per cent completeness for point sources at a magnitude of $g = 22.2$. For each SDSS image, the cutout size is proportional to the extent of the galaxy we want to fit in order to accommodate a reasonable amount of sky area (at least 50 per cent of the total region). This is done so that GALFIT can simultaneously fit the residual sky emission.

We masked galaxy neighbours by use of the segmentation map output by SExtractor. All sources that fall outside the galaxy of interest are masked. To account for the instrumental PSF, in the case of SDSS data we use interpolated PSF models (Lupton et al. 2001) reconstructed with *sdss_psfrec.pro*.[2] For *HST* *F*814*W* images, the PSF was one of the TINYTIM (Krist 1995) models as described by Rhodes et al. (2006, 2007).

Irregular, complex and/or sources detected at low S/N are excluded from the final sample as GALFIT failed to converge on meaningful structural parameters (see Figs A1 and A2 for a comparison between those and the full sample). These cases amount from 8 per cent (in the NYU-VAGC sample) up to 40 per cent (in the $z \sim 2.23$ HiZELS sample) of our samples, which are distributed in both stellar masses and Hα luminosities in the same way as the full sample. Thus, this rate of failure does not introduce any bias against stellar masses or Hα luminosities (see Appendix A).

5 RESULTS

5.1 The impact of surface brightness dimming

We focus our attention on the impact of surface brightness dimming on two structural parameters that are often used to describe the global morphology of galaxies: the effective radius (r_e) and the Sérsic index (n) (see e.g. Morishita et al. 2014; van der Wel et al. 2014; Shibuya et al. 2015).

We investigate the ratio of fractional error for r_e for all samples described in Section 2.1 and at the four redshifts slices observed with HiZELS. We show (in Fig. 4) that the effective radius is, on average, recovered with success (within an error margin of ∼10 per cent) for galaxies with $r_{e,z\approx0} < 10$ kpc. For larger galaxies, we reach a saturation value, i.e. galaxies with $r_{e,z\approx0} > 10$ kpc are recovered, on average, with $r_e \sim 10$–20 kpc (thus, the ratio declines for larger radii). This effect is seen on the CALIFA and MaNGA subsamples, which are the ones where we have galaxies in this range of sizes. It is also noticeable that for the smaller galaxies we have a higher chance of recovering the value observed at $z \approx 0$.

We also investigate the impact on the derived Sérsic indices. We show in Fig. 5 that we recover the value of the Sérsic index at a smaller value (by ∼5–20 per cent) than the one that is observed at $z \approx 0$. This effect is larger at larger $n_{z\approx0}$ and it shows in all samples

[2] http://www.sdss.org/dr12/algorithms/read_psf/

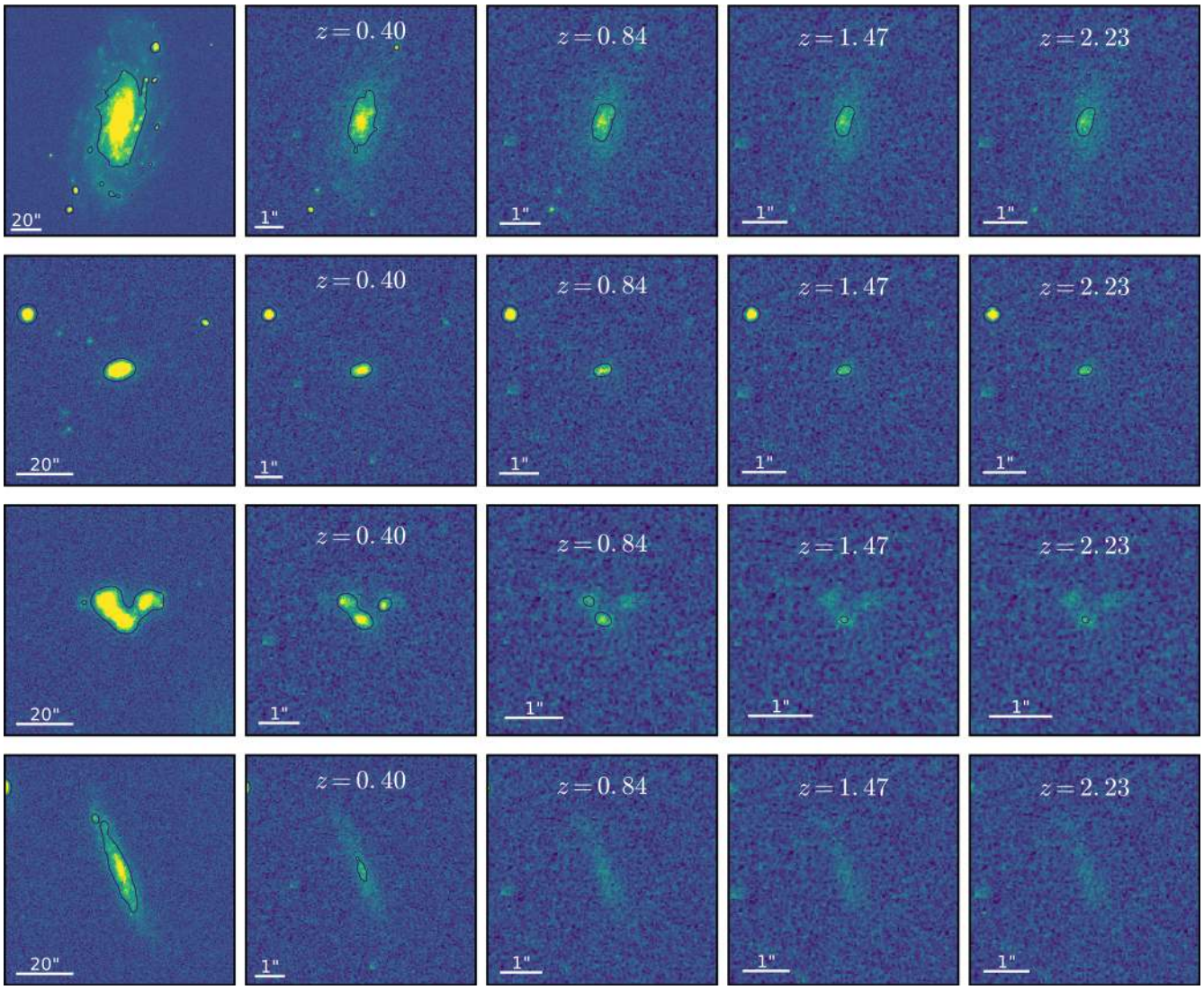

Figure 3. Artificial redshifted examples for four galaxies from the samples used in this paper. The leftmost panel is the original SDSS *g*-band image. The next four panels show the effect of artificial redshifting the galaxy to four different redshifts, increasing from left to right. The contours show the 3σ level detection in each image. In each image, the colour map ranges from -3σ to 10σ. We observe that the fainter/external regions of the galaxies observed in the local Universe completely disappear as we move towards higher redshifts. All images have the same physical scale. The observed scale is shown for each panel to compare angular sizes at different redshifts.

at all redshifts. There does not seem to be any systematic offset between different redshifts.

We show in Appendix B the impact of dimming on the recovered total magnitudes (corrected for the luminosity evolution) and the axial ratio of the profiles of individual galaxies. Our results are consistent with those exploring artificial redshifting up to $z \sim 1$ (Barden et al. 2008; Mosleh et al. 2013) and also for artificial redshifting of massive galaxies from $z \sim 0$ to $z \sim 2.5$ (Weinzirl et al. 2011).

The success of the recovery of the structural parameters for the highest redshift galaxies is directly linked to the counterbalance of the luminosity evolution and the cosmological surface brightness dimming. We have tested a scenario where galaxies undergo no luminosity evolution and the recovery of galaxies (at the largest radius) is severely affected. This would imply that we would not be able to observe large galaxies at higher redshifts and thus, our distributions would be skewed towards smaller sizes potentially mimicking an evolution scenario.

That impact is explored in Table 2 where we compare the median values of the distributions in r_e and n at each redshift compared to the observed median value at $z \approx 0$. As somewhat expected from the comparison of individual galaxies, we observe no systematic trend induced by the cosmological dimming on the recovered median values. For r_e, we see that we slightly underestimate sizes for galaxies in the CALIFA sample (∼90 per cent of the original). For the other samples, we get the same median value within $\lesssim$10 per cent at all redshifts. As for the value of n, we systematically underestimate the median value at a level of 11–27 per cent of the value observed at $z \approx 0$ with all artificially redshifted samples. This value does not seem to change as a function of redshift.

We discuss the dependence of the fractional error of sizes and Sérsic indexes on the input magnitude in Appendix B. The reader is also referred to Appendix B for a discussion on the possible explanations for a systematic overestimation of sizes due to dimming.

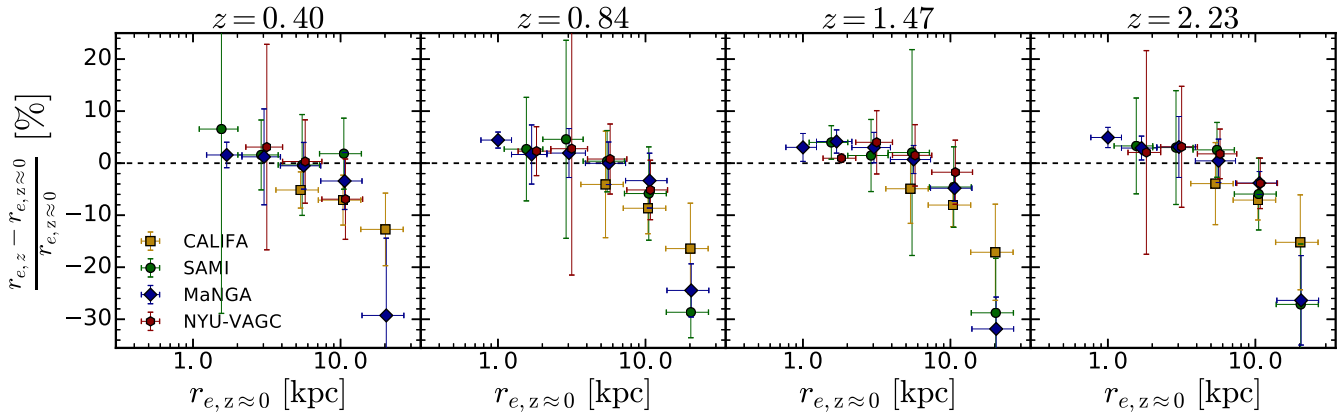

Figure 4. The fraction difference between the recovered and input effective radius $(r_{e,z} - r_{e,mz\approx0})/r_{e,z\approx0}$, for each of the local subsamples and each redshift (one per column). Each symbol represents the median values in bins (with a minimum of 10 galaxies) of width of the horizontal error bar. The vertical error bar shows the error on the median value. The black dashed line pinpoints a fractional difference of 0, i.e. when the recovered effective radius is the same as input. We find the difference to be very small (<10 per cent) up to 10 kpc. Above that value, the size of galaxies is more severely underestimated (∼10–30 per cent). Small horizontal offsets were applied to improve the readability of the plot.

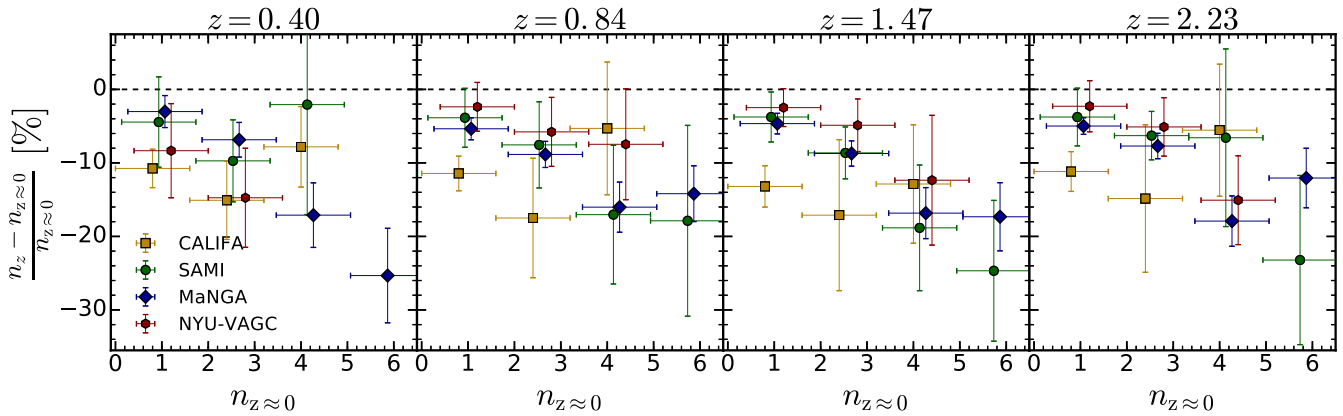

Figure 5. The fraction difference between the recovered and input Sérsic index $(n_z - n_{z\approx0})/n_{z\approx0}$, for each of the local subsamples and each redshift (one per column). Each symbol represents the median values in bins (with a minimum of 10 galaxies) of width of the horizontal error bar. The vertical error bar shows the error on the median value. The black dashed line pinpoints a fractional difference of 0, i.e. when the recovered effective radius is the same as input. We find that the value of the Sérsic index is always underestimated (∼2–30 per cent) in all redshifts and for all four surveys. Small horizontal offsets were applied to improve the readability of the plot.

Table 2. Ratio of the median values, r_e and n, for each simulated redshift to the observed values at $z \approx 0$. This dimensionless value quantifies the over/underestimation of each parameter as a function of the simulated redshift. The average values derived from the four samples are used for correction.

	$\frac{r_{e,z}}{r_{e,z\approx0}} \mid \frac{n_z}{n_{z\approx0}}$			
	$z = 0.40$	$z = 0.84$	$z = 1.47$	$z = 2.23$
CALIFA	0.885 | 0.805	0.918 | 0.828	0.863 | 0.829	0.918 | 0.839
SAMI	1.006 | 0.869	1.166 | 0.763	1.096 | 0.812	1.052 | 0.789
MaNGA	1.031 | 0.856	1.037 | 0.848	1.039 | 0.851	1.037 | 0.832
NYU-VAGC	1.107 | 0.729	1.011 | 0.841	0.989 | 0.832	1.001 | 0.887
Average	1.007 | 0.815	1.033 | 0.820	0.997 | 0.831	1.002 | 0.837

5.2 The structure and sizes of galaxies

We plot the histograms of sizes (r_e) and Sérsic indices (n) in Figs 6 and 7 and summarize the median values in Table 3. For the local samples, we see that galaxies have, on average, $r_e \sim 9$ kpc for the CALIFA sample and $r_e \sim 4$–5 kpc for the other three samples. We want to stress that the CALIFA sample is diameter selected to match the diameter of the IFU field of view and that is what is driving the larger sizes with respect to the other samples, which have lesser to no constraints on galaxy size in their selection. As for the values of n, we see that the distributions are similar, but peaking at increasing values of n for NYU-VAGC, MaNGA, SAMI and CALIFA, respectively. Again, the CALIFA sample shows an increased fraction of galaxies with a high value of n, which is likely due to the morphological constraint to cover the entire Hubble diagram.

As we examine the high-redshift galaxies from HiZELS (Fig. 7), we see that galaxies have increasingly smaller sizes as we move towards higher redshifts with median values ranging from $r_e \sim 2$ kpc at $z = 2.23$ to $r_e \sim 4$ kpc at $z = 0.40$. As for the values of the Sérsic indices, we see that it is rather stable at these redshifts with median values around $n \sim 1.1$, close to exponential disc profiles characteristic of SFGs in the local Universe. We do not observe an enhancement of the fraction of higher n galaxies at any redshift probed with HiZELS. We note that this is not driven by observational biases.

There are some cases of SFGs with large sizes and/or Sérsic indices seen in Fig. 7. However, these generally come from

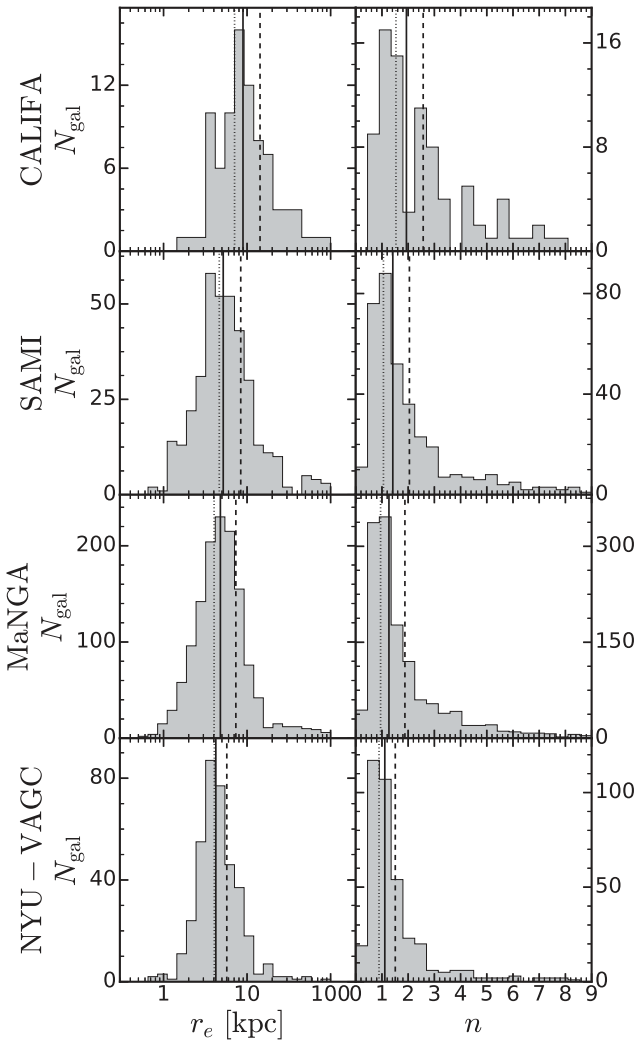

Figure 6. Distributions of sizes (r_e) and Sérsic indices (n) of the galaxies observed at $z \approx 0$ for the four local samples defined in Section 2. The vertical solid, dashed and dotted lines indicate the median, mean and modal values, respectively.

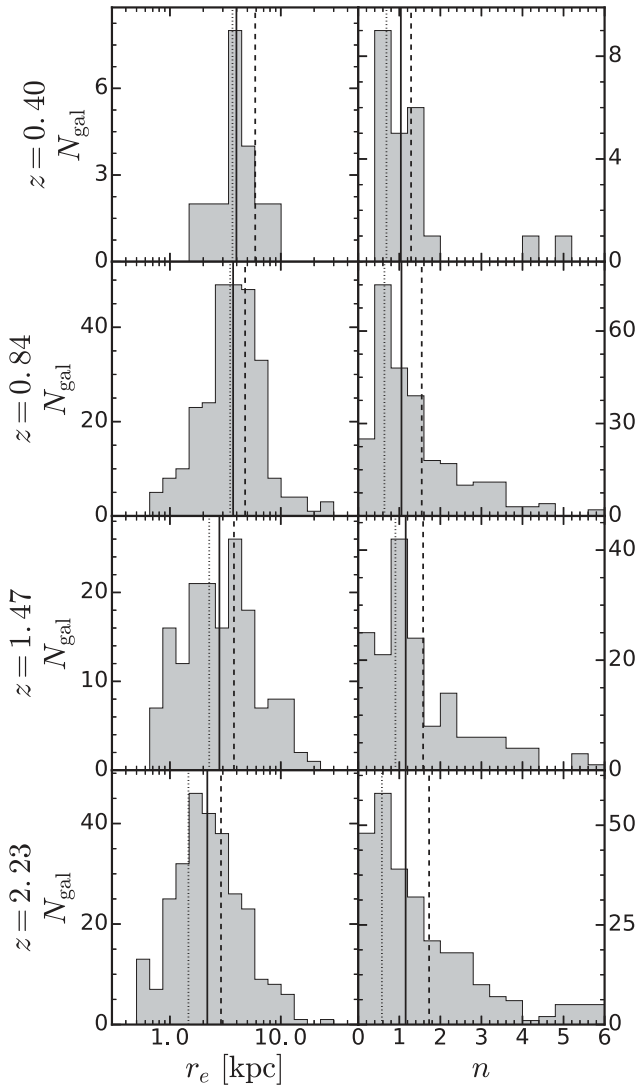

Figure 7. Distributions of sizes (r_e) and Sérsic indices (n) of the galaxies observed at high redshift from HiZELS. The vertical solid, dashed and dotted lines indicate the median, mean and modal values, respectively. The distributions are based on the observed values without any correction.

poorly constrained fits, which involve large errors on both r_e and n.

5.3 The evolution of the structure and sizes of galaxies

After obtaining the structural parameters of our samples, we can see how they compare in terms of their evolution. In Fig. 8, we show how the measured sizes and Sérsic indices depend on redshift. We fit the equation

$$X \propto (1+z)^{\alpha_X} \tag{3}$$

for $X = r_e, n$ to quantify the rate of change of these quantities across cosmic time. We compute these quantities by considering, or not, the impact of dimming in terms of the overestimated/underestimated average ratios shown in Table 2. These results are summarized in Table 4.

From an overall perspective on the Hα-selected sample, the scenario of size growth is compatible with mild evolution ($\alpha_{r_e} \approx -0.7 \pm 0.11$) between redshifts 2.2 and 0.4. We see in Table 4 that considering the entire redshift range (by including the local samples in the fit) increases slightly the value of the slope to $\alpha_{r_e} \approx -0.75 \pm 0.20$, which is fully consistent with the result obtained using only the higher redshift samples from HiZELS. We can also see that including the dimming correction has virtually no impact on the derived value of α_{r_e}, as expected from what we see in Table 2. These slopes deviate more than 3σ from the no evolution scenario ($\alpha = 0$), and thus it is very unlikely that galaxies experience no growth across this period.

We also investigate any potential evolution of the median value of the Sérsic index of galaxies where we get slightly different solutions if we look only at the higher redshift samples (decreasing the Sérsic index with time, $\alpha_n \approx 0.15 \pm 0.04$) or include all the samples in the fit (increasing the Sérsic index with time, $\alpha_n \approx -0.25 \pm 0.21$). However, we note that the slopes are very shallow and point to almost no evolution. These two slightly contrasting scenarios are appeased when we include the impact of dimming in the derived value of n derived in Table 2. With that correction we obtain $\alpha_n \approx -0.07 \pm 0.19$ for all samples and $\alpha_n \approx 0.12 \pm 0.04$ when only considering the HiZELS samples. Considering the full range in redshifts, our derived evolution encompasses the value $\alpha = 0$

Table 3. Median, mean and modal values for the effective radius (r_e) and Sérsic index (n) for all samples described in Section 2. The errors on the median denote the 16th (lower bound) and 84th (upper bound) percentiles of the variables distribution.

	r_e(kpc)			n		
Sample	Median	Mean	Mode	Median	Mean	Mode
CALIFA	$8.94^{+11.13}_{-4.60}$	14.30 ± 1.83	7.08	$1.93^{+2.34}_{-0.85}$	2.57 ± 0.19	1.53
SAMI	$5.18^{+5.54}_{-2.51}$	8.40 ± 0.62	4.65	$1.41^{+1.89}_{-0.67}$	2.05 ± 0.09	1.06
MaNGA	$4.79^{+3.94}_{-2.24}$	7.34 ± 0.30	4.03	$1.27^{+1.83}_{-0.56}$	1.88 ± 0.05	0.95
NYU-VAGC	$4.23^{+3.30}_{-1.50}$	5.72 ± 0.34	4.05	$1.11^{+1.07}_{-0.45}$	1.51 ± 0.07	0.89
$z = 0.40$	$3.98^{+2.34}_{-1.30}$	5.88 ± 1.49	3.67	$1.04^{+0.49}_{-0.47}$	1.29 ± 0.23	0.69
$z = 0.84$	$3.72^{+2.45}_{-1.87}$	4.77 ± 0.35	3.50	$1.05^{+1.49}_{-0.55}$	1.55 ± 0.08	0.64
$z = 1.47$	$2.79^{+2.89}_{-1.55}$	3.78 ± 0.25	2.26	$1.16^{+1.62}_{-0.72}$	1.58 ± 0.11	0.91
$z = 2.23$	$2.17^{+2.35}_{-1.11}$	2.89 ± 0.16	1.47	$1.15^{+1.92}_{-0.76}$	1.73 ± 0.10	0.58

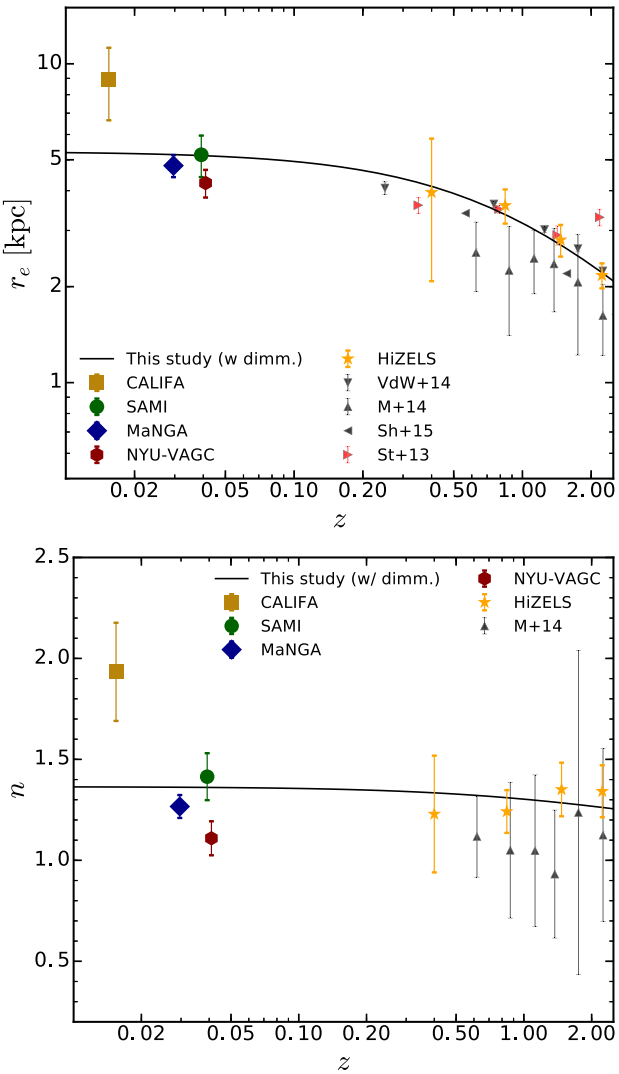

Figure 8. Structural evolution of SFGs as parametrized by Sérsic profiles. Top: the evolution of galaxies' median (observed) effective radius as a function of the median redshift. Bottom: same as top but for the galaxies' median Sérsic index. The coloured points show the values (after dimming correction for HiZELS) derived from this study. The triangles show reported values from Stott et al. (2013), van der Wel et al. (2014), Morishita et al. (2014), Shibuya et al. (2015). The points from Stott et al. (2013) have been horizontally offset for viewing purposes. The solid line shows the fit to equation (3) after dimming correction.

Table 4. Median values of α_X after fitting through all samples and for fitting only the HiZELS high-z sample.

Variable	Correction	All samples	High-z only
α_{r_e}	W/o dimm.	-0.74 ± 0.20	-0.71 ± 0.11
	W/ dimm.	-0.75 ± 0.20	-0.70 ± 0.09
α_n	W/o dimm.	-0.25 ± 0.21	0.15 ± 0.04
	W/ dimm.	-0.07 ± 0.19	0.12 ± 0.04

within the 1σ error. This supports a scenario where SFGs maintain their surface brightness distribution close to pure exponential discs at all times.

5.4 The size–mass relation across cosmic time

The other aspect to retain from the morphological analysis is the relation of the structural parameters with the stellar mass, which is displayed in Figs 9 and 10. There, we show the median values in bins of stellar mass and their dispersion on the stellar mass–size relation alongside with the individual measurements for all galaxies. In order to parametrize this relation, we use the functional form

$$r_e \propto M_\star^{\alpha_{M_\star}}, \quad (4)$$

which is overplotted as a red solid line in those figures. The slopes of this relation are summarized in Table 5.

To fit the stellar mass–size relation, we perform a Markov chain Monte Carlo simulation (using the EMCEE package; Foreman-Mackey et al. 2013) where all individual galaxies are considered for the fit. The first guess is computed from a simple fit to the binned points. We use a total of 50 chains with initial guesses randomly deviated from the simple fit. Each chain then runs for 2000 steps and we exclude the first 500 steps for each chain to erase the influence of the initial first guesses. The reported errors on the slope are the width of the posterior distribution.

We show that for the local samples, apart from CALIFA, there is a more evident correlation of sizes with stellar masses as that measured from the HiZELS samples. Again, for the CALIFA sample, the marked difference against the other local Universe samples is related to the size constraint imposed for the parent selection of CALIFA. This results in a flatter stellar mass–size relation as galaxies were chosen to have similar sizes within the sample and stresses the importance of sample selection whenever we study a galaxy population. Due to this selection effect, we disregard the CALIFA sample when establishing comparisons between local and high-redshift morphologies. None the less, despite the apparent

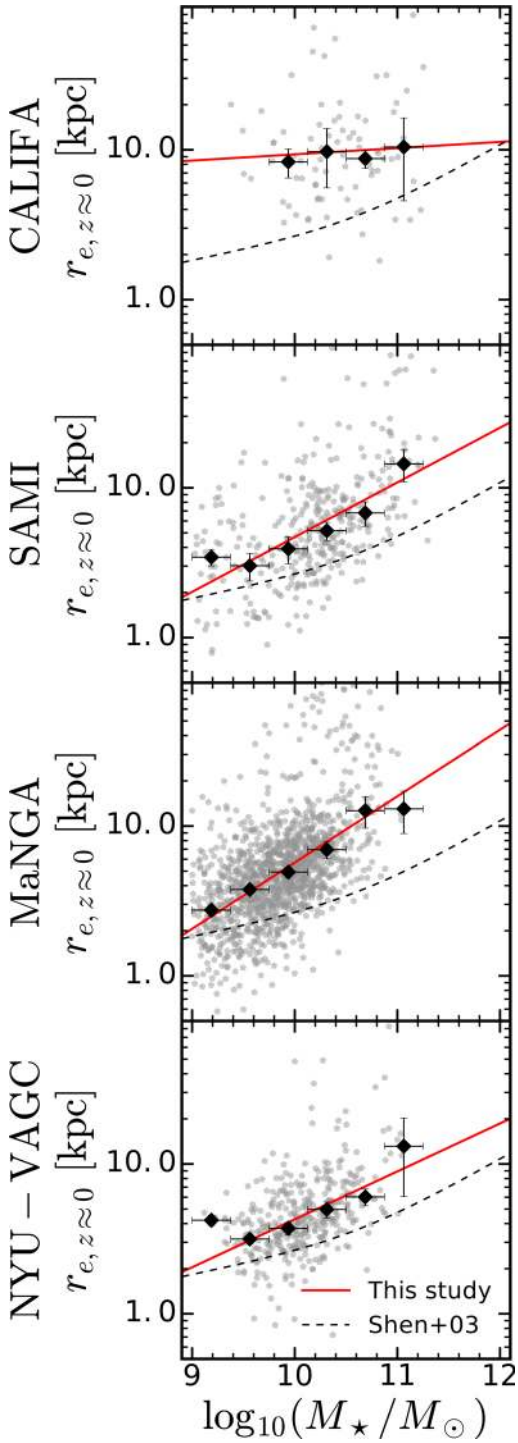

Figure 9. The stellar mass–size relation for the local Universe samples described in Section 2. Each grey pentagon represents an individual galaxy, and the large black diamonds are the median values in a stellar mass bin of width indicated by the horizontal error bar. The vertical error bar denotes the error on the median. The red solid line shows the best fit of equation (4) to the data (see also Table 5). We include the fits from Shen et al. (2003) as a solid dashed line. We note that the reason for the differences between our fits and those by Shen et al. (2003) is mainly due to the use of different size estimator. We use the major axis effective radius and they use the circularized effective radius. When using the latter in our data, we find better agreement. See Section 6 for more details.

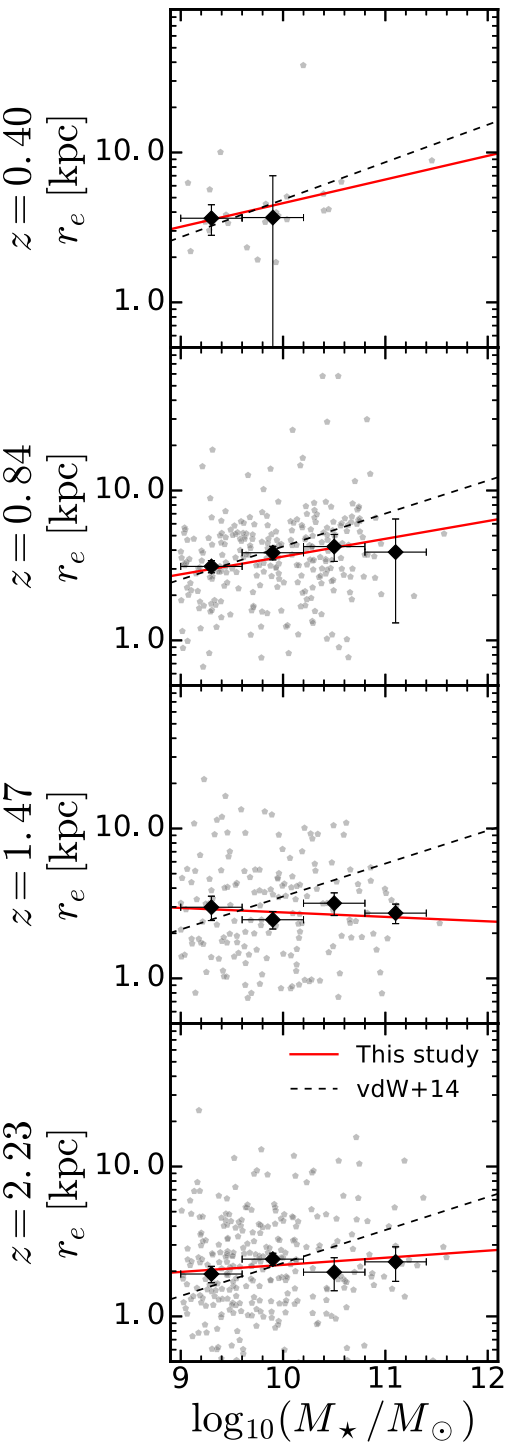

Figure 10. The stellar mass–size relation for four redshifts probed by HiZELS. Each grey pentagon represents an individual galaxy, and the large black diamonds are the median values in a stellar mass bin of width indicated by the horizontal error bar. The vertical error bar denotes the error on the median. The red solid line shows the best fit of equation (4) to the data (see also Table 5). We include the fits from van der Wel et al. (2014) as a solid dashed line (from top to bottom, the redshift bins from that paper are $0.0 < z < 0.5$, $0.5 < z < 1.0$, $1.0 < z < 1.5$, $2 < z < 2.5$).

Table 5. Slope of the stellar mass–size relation for all the samples described in Section 2.

Sample	$\alpha_{M_\star}$
CALIFA	0.05 ± 0.17
SAMI	0.37 ± 0.07
MaNGA	0.44 ± 0.05
NYU-VAGC	0.32 ± 0.09
$z = 0.40$	0.15 ± 0.36
$z = 0.84$	0.11 ± 0.11
$z = 1.47$	-0.03 ± 0.14
$z = 2.23$	0.05 ± 0.11

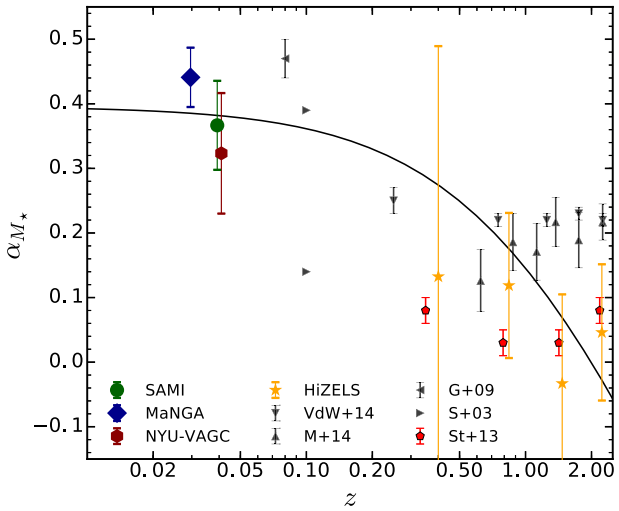

Figure 11. The value of the slope of the stellar mass–size relation as a function of the median redshift of each sample. Larger full coloured symbols are for the samples studied in this paper. The semitransparent triangles are from other reports from the literature: Shen et al. (2003); Guo et al. (2009); van der Wel et al. (2014); Morishita et al. (2014). Shen et al. (2003) symbols correspond to the low (0.14 at $<10^{10.6}$ $M_\odot$) and high (0.39 at $>10^{10.6}$ $M_\odot$) stellar mass end. Values from Stott et al. (2013) are shown as red pentagons and a horizontally shift was applied for viewing purposes. The solid line shows the best fit through our data points of $\alpha_{M_\star} = A \log_{10}(1+z) + B$, with $A = 0.8 \pm 0.2$ and $B = 0.40 \pm 0.05$.

inversion of the relation for $z = 1.47$, we observe that on average the most massive galaxies are larger than their lower mass counterparts.

We compare our best-fitting relations with those found in the literature for SFGs and find that in the local Universe the CALIFA sample is the one that deviates the most from what is found in SDSS by Shen et al. (2003) and Guo et al. (2009). This deviation from CALIFA is expected since it is a size-selected sample and thus a flatter correlation with stellar mass is unsurprising. As we move towards higher redshifts, we observe that the HiZELS sample changes its slope to shallower values. This is in contrast with the apparent constancy in the stellar mass–size relation slopes found by van der Wel et al. (2014) and Morishita et al. (2014). We note that their selection of SFGs is based on a colour-colour diagram, and not on Hα. On top of that, we measure our relation on bluer wavelengths than those studies that are based on rest-frame optical data on smaller areas of the sky.

We go further in the investigation of the evolution of this stellar mass–size relation across cosmic time and plot the values of the slope as a function of redshift in Fig. 11. Again, if one excludes CALIFA sample (due to its selection bias), the relation becomes steeper as we move towards the local Universe. We note also that the steepening appears to occur mostly between $z = 0.4$ and $z \approx 0$, as when we look only at the HiZELS samples, the rate of change on the value of $\alpha_{M_\star}$ is smaller. We have attempted to improve the robustness of our fits by relaxing the Hα luminosity cut down to $0.01 L^*_{\mathrm{H}\alpha}$ for $z = 0.4$. We have 95 galaxies in this scenario and derive a slope of $\alpha_M = 0.17 \pm 0.26$, consistent with what we have but with a slightly smaller error.

By using the fits derived in equation (4), it is possible to derive the effective radius at a given stellar mass for all the samples we are studying in this paper. We have computed the sizes of galaxies at $\log_{10} M_\star/\mathrm{M}_\odot = 10$ and 10.5 and found that they follow a similar trend as displayed in Fig. 8 and Table 3. The evolution one would derive from this quantity is similar to that of using the median sizes for the galaxy population. Additionally, one can see that the more massive galaxies tend to grow faster with cosmic time as the differences between consecutive redshift slices is larger at $\log_{10} M_\star/\mathrm{M}_\odot = 10.5$ than at $\log_{10} M_\star/\mathrm{M}_\odot = 10$.

5.5 Impact of cosmological dimming on the stellar mass–size relation

We attempt to quantify the impact that dimming may have on the derived stellar mass–size relations by exploring the dependence of the fractional error on the effective radius on the stellar mass of the input galaxy. In Fig. B5, we see that there is no strong correlation between the two quantities. However, we do find that galaxies at the highest masses ($\log_{10}(M_\star/\mathrm{M}_\odot) \gtrsim 10.5$) are more likely to have their sizes underestimated at a level of $\lesssim$20 per cent. We do not find any dependence on the effect of dimming with simulated redshift slice. These two results imply that it is unlikely that the shallow slopes that we observe for our stellar mass–size relations are caused by the cosmological dimming.

Thus, we believe that the differences we observe are due to our selection based on Hα and due to the weaker correlation between rest-frame UV light and stellar mass, which is measured mostly from rest-frame near-infrared (NIR).

6 DISCUSSION

In order to infer the true evolution of galaxies, we need to account for any observational bias that may affect our observed results. Regarding morphology, and its evolution with redshift, the strongest bias could come from cosmological surface brightness dimming. The impact of this dimming was already extensively explored out to $z \sim$ 1–2.5 by Barden et al. (2008) (see also Trujillo et al. 2007; Franx et al. 2008; Weinzirl et al. 2011; Mosleh et al. 2013), where they find that to have little impact on GALFIT-based measurements in this redshift range. Nevertheless, due to the strong dependence of dimming on redshift and the luminosity evolution of galaxies with redshift, any evolutionary trend must be taken with care. For instance, Ichikawa et al. (2012) found a small evolution on galaxy sizes in optical rest frame, but consistent with the expected effect from cosmological dimming. We show in Section 5 that indeed the cosmological dimming has a small impact on the derived sizes and it is more important on the derived Sérsic indices.

As for the local samples, we find galaxies that are two to three times larger, depending on the local sample we consider, when compared to other studies based on late-type galaxies in SDSS. For example, Shen et al. (2003) found values of the half-light Petrosian radius $r_{50} \sim 2.5$ (3.2) kpc at stellar masses of $\log_{10}(M_\star/\mathrm{M}_\odot) =$ 10 (10.5). Guo et al. (2009) studied central galaxies in SDSS, and found that late-type galaxies have $\sqrt{b/a} r_\mathrm{e} \sim 2.45$ (4.78) kpc at stellar masses of $\log_{10}(M_\star/\mathrm{M}_\odot) = 10$ (10.5).

The differences that we find in the typical sizes of galaxies are a reflection of differences in the overall stellar mass–size relation. Shen et al. (2003) fitted a double power law and found slopes of $\alpha_{M_\star} = 0.14$ for $\log_{10}(M_\star/\mathrm{M}_\odot) < 10.6$ and $\alpha_{M_\star} = 0.39$ for $\log_{10}(M_\star/\mathrm{M}_\odot) > 10.6$. Note that we use the major axis effective radius in all our plots to be comparable to those reported by the literature at high redshift, namely the values from Stott et al. (2013). However, the values reported by Shen et al. (2003) refer to the circularized effective radius. When using this size estimate, we get much better agreement between our results and those reported by Shen et al. (2003), especially on the control NYU-VAGC sample. We do, however, still find a bias that IFU samples have slightly larger galaxies at high masses. This small bias may be perceived as an attempt to maximize the covering factor of IFU instruments. We computed the absolute difference between our best-fitting model and the quoted best fit by Shen et al. (2003) in the stellar mass interval $9 < \log(M_\star/\mathrm{M}_\odot) < 11.5$ and found that difference to be a factor of 2–3 smaller when using the circularized effective radius.

Guo et al. (2009) find a steeper slope of the stellar mass–size relation with $\alpha_{M_\star} = 0.47 \pm 0.03$, which is slightly higher but statistically compatible with our observed slopes for the local samples. We note that our SFGs are lying in between the slopes that are found but have larger sizes at the same stellar masses. However, we stress that the selection of late-type galaxies in SDSS is different than what we have applied (see Section 2). They are based in either visual inspection (Guo et al. 2009) and structural separation (using the concentration $c < 2.86$ and $n < 2.5$, Shen et al. 2003). We also exclude low-mass galaxies from the fit ($\log(M_\star/\mathrm{M}_\odot) < 9$), which influence the stellar mass–size relation by Shen et al. (2003). We are also missing very massive galaxies ($\log(M_\star/\mathrm{M}_\odot) \gtrsim 11$), which could influence our slopes to shallower values if they populate a lower size than what is predicted from our fits. We note however that we agree with the results from the literature at an $\sim 2\sigma$ level and that our shallower slopes could be driven by our smaller sample size or that we are measuring sizes in the blue/NUV rest-frame region.

As for the size evolution, our derived trends are consistent with the slopes found by van der Wel et al. (2014), $\alpha_{r_e} = -0.75$ and slightly steeper than what was found by Morishita et al. (2014), $\alpha_{r_e} = -0.57$. Both these studies target SFGs selected from the *UVJ* diagram (Williams et al. 2009) at redshifts $z < 3.0$. They also focus on the stellar mass–size relation and find a consistent slope of $\alpha_{M_\star} = 0.22$ (van der Wel et al. 2014) and $\alpha_{M_\star} \sim 0.2$ (Morishita et al. 2014) at all redshifts. This slopes are within the errors of the slope found at $z < 0.84$, but we find consistently shallower slopes at higher redshifts and a possible hint of evolution with increasingly shallow slopes as we move to higher redshifts.

These shallow slopes are more consistent with those derived by Stott et al. (2013), using the same HiZELS sample but measuring sizes in rest-frame ground-based optical images. They do not find any evolution at the same redshifts and find remarkably constant sizes with $r_e(M_\star = 10^{10}\,\mathrm{M}_\odot) \sim 3.6$ kpc in the same redshifts we probe here. The small change in sizes found by Stott et al. (2013) may seem contradictory to the evolution we find, which is mostly anchored on the larger difference in sizes observed at $z = 2.23$. We believe that the different findings may be caused by a different sample selection but more importantly by the different resolution and rest-frame bands that we use. Stott et al. (2013) use *K*-band imaging data (which cover the region 6800–15 700 Å rest-frame), which can be less prone to a morphological *k*-correction than the use of *F*814*W* (which covers the region 2500–5800 Å rest-frame). This is especially true for galaxies with a strong D_n4000 Å break and for galaxies with spatially disparate young and old stellar populations. However, for strong SFGs, the break is expected to be small (e.g. fig. 2 of Li et al. 2015, and references therein). Additionally, it has been shown that for galaxies with dominant star-forming population the morphological *k*-correction is small (e.g. Conselice et al. 2000; Windhorst et al. 2002; Taylor-Mager et al. 2007). None the less, it is possible that we are seeing a different path of evolution for the young star-forming regions when compared to the older underlying stellar population. This points to a scenario where star formation happens from inside-out and we are witnessing the star formation activity extending to the outer regions of the galactic disc (see e.g. Hagen et al. 2016).

van der Wel et al. (2014) find that late-type galaxies are smaller at longer wavelengths and that this trend is stronger for the most massive galaxies. This trend is weaker as we move towards higher redshifts. When applied to the HiZELS sample, where the provided corrections can be applied, these two effects combined could contribute to shallower stellar mass–size relation slopes as the most massive, lower redshift galaxies would be the ones with the largest decrease in size. However, we already find shallower slopes than those reported by other studies, and these effects would only contribute to a more consistent, no-evolution scenario of the slope of the stellar mass–size relation, which would then be consistent with what is found by Stott et al. (2013).

We can also compare our results to other typical selections of high-redshift galaxies, namely those based on the Lyman-break galaxies (LBGs; Steidel et al. 1996). For example, Ferguson et al. (2004), Mosleh et al. (2011) and Shibuya et al. (2015) find stronger evolution of these populations with slopes $\alpha_{r_e} < -1$. Mosleh et al. (2011) also find steeper slopes for galaxies with UV bright selections. This hints at the fact that despite tracing part of a star-forming population, the LBG selection misses the global picture of SFGs that should be the dominant population at high redshifts.

There are also studies using galaxies selected by their photometric redshift (e.g. Trujillo et al. 2006, 2007; Buitrago et al. 2008; Franx et al. 2008) and they find a global size evolution scaling with $\alpha_{r_e} = -0.40, -0.82, \sim -0.59$ at $0 < z \lesssim 3$, respectively, and also a stronger size evolution going from low- to high-mass galaxies. However, they have not specifically targeted the star-forming population at these redshifts.

7 CONCLUSIONS

We present the morphological characterization of SFGs selected through their Hα luminosity and compare their evolution across the last 11 Gyr of the Universe ($z < 2.23$). We also compare the correlation of the structural parameters with the stellar mass and derive evolutionary trends for galaxy sizes (parametrized as the effective radius) and Sérsic indices including the effect of cosmological dimming in the analysis. Our main results are as follows.

(i) Cosmological surface brightness dimming (when using GALFIT) has a negligible impact ($\lesssim$10 per cent) on the derived effective radius for all galaxies with $r_e < 10$ kpc. We show that it impacts the derivation of the Sérsic index, where we find a systematic underestimation of $\sim$20 per cent at the higher redshifts in the artificially redshifted samples compared to the ones observed at $z \approx 0$. This underestimation does not change the fact that the observed galaxies have surface brightness profiles similar to exponential discs.

(ii) Effective radii of SFGs show moderate evolution ($\alpha_{r_e} \approx 0.7 \pm 0.2$) from $z = 2.23$ down to $z \approx 0$. They have a range of galaxy sizes that grow from $\sim$2 kpc at $z = 2.23$ up to 4–9 kpc at $z \approx 0$.

(iii) We find that SFGs have typically disc-like profiles with a median value of $n \sim 1.2$. Their Sérsic indices show negligible evolution (α_n is consistent with 0) across the same period with a median value of $1 < n < 1.5$, close to exponential disc profiles.

(iv) The stellar mass–size relation becomes steeper as we move towards lower redshifts. This hints at a stronger size evolution of the most massive SFGs when compared to the lower mass counterparts.

We use four different samples in the local Universe, three of them based on ongoing IFU surveys and one selected from SDSS as a control sample, to compare local to high-redshift morphologies. Due to its diameter-selected sample, the SFGs in the CALIFA survey are the most biased against a dynamic range in galaxy sizes. This limits the interpretations of the results if it is to be used as a local counterpart to the high-redshift samples being gathered with NIR IFU instruments. As for the SAMI and MaNGA samples, they seem to provide a representative morphological range of the local Universe, when compared to our NYU-VAGC control sample and, therefore, are more suitable for such comparisons.

Our results reveal that cosmological dimming plays a negligible role in the derivation of evolutionary trends on galaxy morphology for SFGs (and when using GALFIT). We show that SFGs grow in size, as seen from blue to UV rest-frame regions, by a factor of 2–3 since $z \sim 2.23$ while their profile shapes remain the same (close to exponential discs). Interestingly, this growth is not observed in the same sample as seen from red-NIR regions (Stott et al. 2013) and the observed differences are not due to sampling issues. This can also be linked to a scenario of inside-out star formation as seen by Nelson et al. (2016). Although selection effects may play a role, it is possible that we are witnessing two distinct evolution paths for active star-forming regions and the underlying older stellar population across these redshifts. Alternatively, investigating new non-parametric size measurements, which fully account for cosmological surface brightness dimming (Ribeiro et al. 2016), might provide new hints at size evolution trends. Moreover, our results put into perspective the galaxy morphologies of ongoing local IFU surveys and serve as a reference for future comparisons of local and high-redshift IFU galaxy surveys.

ACKNOWLEDGEMENTS

We thank the anonymous referee for the thorough review, positives comments, and constructive remarks which greatly improved the quality and clarity of this manuscript. We also thank J. Stott for sharing their size measurements that were used in the making of this paper.

This work was supported by Fundação para a Ciência e a Tecnologia (FCT) through the research grant UID/FIS/04434/2013. APA, PhD::SPACE fellow, acknowledges support from the FCT through the fellowship PD/BD/52706/2014. DS acknowledges financial support from the Netherlands Organization for Scientific research (NWO) through a Veni fellowship and from FCT through an FCT Investigator Starting Grant and Start-up Grant (IF/01154/2012/CP0189/CT0010).

Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation and the U.S. Department of Energy Office of Science. The SDSS-III website is http://www.sdss3.org/.

SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofisica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington and Yale University.

This work greatly benefited from observations obtained using the Wide Field CAMera (WFCAM) on the 3.8m United Kingdom Infrared Telescope (UKIRT), as part of the High-redshift(Z) Emission Line Survey (HiZELS; U/CMP/3 and U/10B/07; Best et al. 2013), and observations obtained with Suprime-Cam on the Subaru telescope (S10B-144S). The HiZELS sample used in this work is publicly available from Sobral et al. (2013a).

This work was only possible by the use of the following PYTHON packages: NUMPY and SCIPY (Jones et al. 2001; Van Der Walt, Colbert & Varoquaux 2011); MATPLOTLIB (Hunter 2007) and ASTROPY (Astropy Collaboration et al. 2013).

REFERENCES

Abazajian K. N. et al., 2009, ApJS, 182, 543
Abraham R. G., Valdes F., Yee H. K. C., van den Bergh S., 1994, ApJ, 432, 75
Ahn C. P. et al., 2012, ApJS, 203, 21
Andredakis Y. C., Peletier R. F., Balcells M., 1995, MNRAS, 275, 874
Astropy Collaboration et al., 2013, A&A, 558, A33
Atek H. et al., 2010, ApJ, 723, 104
Ball N. M., Loveday J., Brunner R. J., 2008, MNRAS, 383, 907
Bamford S. P. et al., 2009, MNRAS, 393, 1324
Barden M., Jahnke K., Häußler B., 2008, ApJS, 175, 105
Bassett R. et al., 2013, ApJ, 770, 58
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Best P. et al., 2013, in Adamson A., Davies J., Robson I., eds, HiZELS: The High Redshift Emission Line Survey with UKIRT, Vol. 37. Springer Science+Business Media, Dordrecht, p. 235
Blanton M. R. et al., 2003, ApJ, 594, 186
Blanton M. R. et al., 2005a, AJ, 129, 2562
Blanton M. R., Lupton R. H., Schlegel D. J., Strauss M. A., Brinkmann J., Fukugita M., Loveday J., 2005b, ApJ, 631, 208
Blanton M. R., Kazin E., Muna D., Weaver B. A., Price-Whelan A., 2011, AJ, 142, 31
Böhm A., Ziegler B. L., 2016, A&A, 592, A64
Bryant J. J. et al., 2015, MNRAS, 447, 2857
Buitrago F., Trujillo I., Conselice C. J., Bouwens R. J., Dickinson M., Yan H., 2008, ApJ, 687, L61
Buitrago F., Trujillo I., Conselice C. J., Häußler B., 2013, MNRAS, 428, 1460
Bundy K. et al., 2015, ApJ, 798, 7
Caon N., Capaccioli M., D'Onofrio M., 1993, MNRAS, 265, 1013
Catalán-Torrecilla C. et al., 2015, A&A, 584, A87
Chabrier G., 2003, ApJ, 586, L133
Cimatti A. et al., 2008, A&A, 482, 21
Conselice C. J., 2003, ApJS, 147, 1
Conselice C. J., 2014, ARA&A, 52, 291
Conselice C. J., Gallagher J. S., Calzetti D., Homeier N., Kinney A., 2000, AJ, 119, 79
Curtis-Lake E. et al., 2016, MNRAS, 457, 440
Davies J. I., Phillipps S., Cawson M. G. M., Disney M. J., Kibblewhite E. J., 1988, MNRAS, 232, 239
de Vaucouleurs G., 1959, Handbuch Phys., 53, 275
Delgado-Serrano R., Hammer F., Yang Y. B., Puech M., Flores H., Rodrigues M., 2010, A&A, 509, A78

Dressler A., 1980, ApJ, 236, 351
Driver S. P. et al., 2009, Astron. Geophys., 50, 5.12
Ferguson H. C. et al., 2004, ApJ, 600, L107
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Franx M., van Dokkum P. G., Schreiber N. M. F., Wuyts S., Labbé I., Toft S., 2008, ApJ, 688, 770
Freeman P. E., Izbicki R., Lee A. B., Newman J. A., Conselice C. J., Koekemoer A. M., Lotz J. M., Mozena M., 2013, MNRAS,
Furlong M. et al., 2015, preprint (arXiv:1510.05645)
Garn T., Best P. N., 2010, MNRAS, 409, 421
Geach J. E., Smail I., Best P. N., Kurk J., Casali M., Ivison R. J., Coppin K., 2008, MNRAS, 388, 1473
Genel S. et al., 2014, MNRAS, 445, 175
Giavalisco M., Livio M., Bohlin R. C., Macchetto F. D., Stecher T. P., 1996, AJ, 112, 369
Graham A. W., 2001, AJ, 121, 820
Guo Y. et al., 2009, MNRAS, 398, 1129
Hagen A. et al., 2016, preprint (arXiv:1610.01163)
Hibbard J. E., Vacca W. D., 1997, AJ, 114, 1741
Holmberg E., 1958, Meddelanden fran Lunds Astronomiska Observatorium II, 136, 1
Hopkins A. M., Connolly A. J., Szalay A. S., 2000, AJ, 120, 2843
Hubble E. P., 1926, ApJ, 64, 321
Huertas-Company M. et al., 2015, ApJ, 809, 95
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Ichikawa T., Kajisawa M., Akhlaghi M., 2012, MNRAS, 422, 1014
Jones E., Oliphant T., Peterson P. et al., 2001, SciPy: Open source scientific tools for Python. Available at: http://www.scipy.org/
Kauffmann G. et al., 2003, MNRAS, 346, 1055
Khosroshahi H. G., Wadadekar Y., Kembhavi A., 2000, ApJ, 533, 162
Koekemoer A. M. et al., 2007, ApJS, 172, 196
Krist J., 1995, in Shaw R. A., Payne H. E., Hayes J. J. E., eds, ASP Conf. Ser. Vol. 77, Astronomical Data Analysis Software and Systems IV. Astron. Soc. Pac., San Francisco, p. 349
Law D. R., Steidel C. C., Erb D. K., Pettini M., Reddy N. A., Shapley A. E., Adelberger K. L., Simenc D. J., 2007, ApJ, 656, 1
Li C. et al., 2015, ApJ, 804, 125
Liske J. et al., 2015, MNRAS, 452, 2087
Lotz J. M., Primack J., Madau P., 2004, AJ, 128, 163
Lupton R., Gunn J. E., Ivezić Z., Knapp G. R., Kent S., 2001, in Harnden F. R., Jr, Primini F. A., Payne H. E., eds, ASP Conf. Ser. Vol. 238, Astronomical Data Analysis Software and Systems X. Astron. Soc. Pac., San Francisco, p. 269
Ly C., Lee J. C., Dale D. A., Momcheva I., Salim S., Staudaher S., Moore C. A., Finn R., 2011, ApJ, 726, 109
McCarthy P. J. et al., 1999, ApJ, 520, 548
Madau P., Dickinson M., 2014, ARA&A, 52, 415
Möllenhoff C., Heidt J., 2001, A&A, 368, 16
Moorwood A. F. M., van der Werf P. P., Cuby J. G., Oliva E., 2000, A&A, 362, 9
Moriondo G., Giovanardi C., Hunt L. K., 1998, A&AS, 130, 81
Morishita T., Ichikawa T., Kajisawa M., 2014, ApJ, 785, 18
Mortlock A. et al., 2013, MNRAS, 433, 1185
Mosleh M., Williams R. J., Franx M., Kriek M., 2011, ApJ, 727, 5
Mosleh M., Williams R. J., Franx M., 2013, ApJ, 777, 117
Nelson E. J. et al., 2016, ApJ, 828, 27
Oke J. B., Gunn J. E., 1983, ApJ, 266, 713
Oteo I., Sobral D., Ivison R. J., Smail I., Best P. N., Cepa J., Pérez-García A. M., 2015, MNRAS, 452, 2018
Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2002, AJ, 124, 266
Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2010, AJ, 139, 2097
Pérez-González P. G., Trujillo I., Barro G., Gallego J., Zamorano J., Conselice C. J., 2008, ApJ, 687, 50
Petty S. M., de Mello D. F., Gallagher J. S., III, Gardner J. P., Lotz J. M., Mountain C. M., Smith L. J., 2009, AJ, 138, 362
Ravindranath S. et al., 2004, ApJ, 604, L9
Rhodes J. D., Massey R., Albert J., Taylor J. E., Koekemoer A. M., Leauthaud A., 2006, in Koekemoer A. M., Goudfrooij P., Dressel L. L., eds, The 2005 HST Calibration Workshop: Hubble After the Transition to Two-Gyro Mode. NASA, Washington, DC, p. 21
Rhodes J. D. et al., 2007, ApJS, 172, 203
Ribeiro B. et al., 2016, A&A, 593, A22
Roberts M. S., Haynes M. P., 1994, ARA&A, 32, 115
Salim S. et al., 2007, ApJS, 173, 267
Sánchez S. F. et al., 2012, A&A, 538, A8
Scoville N. et al., 2007, ApJS, 172, 38
Sérsic J. L., 1968, Atlas de galaxias australes. Observatorio Astronomico, Cordoba
Shen S., Mo H. J., White S. D. M., Blanton M. R., Kauffmann G., Voges W., Brinkmann J., Csabai I., 2003, MNRAS, 343, 978
Shibuya T., Ouchi M., Harikane Y., 2015, ApJS, 219, 15
Shim H., Colbert J., Teplitz H., Henry A., Malkan M., McCarthy P., Yan L., 2009, ApJ, 696, 785
Simard L., 1998, in Albrecht R., Hook R. N., Bushouse H. A., eds, ASP Conf. Ser. Vol. 145, Astronomical Data Analysis Software and Systems VII. Astron. Soc. Pac., San Francisco, p. 108
Sobral D. et al., 2009, MNRAS, 398, 75
Sobral D., Best P. N., Smail I., Geach J. E., Cirasuolo M., Garn T., Dalton G. B., 2011, MNRAS, 411, 675
Sobral D., Smail I., Best P. N., Geach J. E., Matsuda Y., Stott J. P., Cirasuolo M., Kurk J., 2013a, MNRAS, 428, 1128
Sobral D. et al., 2013b, ApJ, 779, 139
Steidel C. C., Giavalisco M., Pettini M., Dickinson M., Adelberger K. L., 1996, ApJ, 462, L17
Stott J. P., Sobral D., Smail I., Bower R., Best P. N., Geach J. E., 2013, MNRAS, 430, 1158
Stott J. P. et al., 2014, MNRAS, 443, 2695
Stott J. P. et al., 2016, MNRAS, 457, 1888
Swinbank A. M., Sobral D., Smail I., Geach J. E., Best P. N., McCarthy I. G., Crain R. A., Theuns T., 2012a, MNRAS, 426, 935
Swinbank A. M., Smail I., Sobral D., Theuns T., Best P. N., Geach J. E., 2012b, ApJ, 760, 130
Taylor-Mager V. A., Conselice C. J., Windhorst R. A., Jansen R. A., 2007, ApJ, 659, 162
Tolman R. C., 1930, Proc. Natl. Acad. Sci., 16, 511
Tremonti C. A. et al., 2004, ApJ, 613, 898
Trujillo I., Graham A. W., Caon N., 2001, MNRAS, 326, 869
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 Walt S., Colbert S. C., Varoquaux G., 2011, Comput. Sci. Eng., 13, 22
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., 2011, ApJ, 743, L15
Viero M. P. et al., 2012, MNRAS, 421, 2161
Villar V., Gallego J., Pérez-González P. G., Pascual S., Noeske K., Koo D. C., Barro G., Zamorano J., 2008, ApJ, 677, 169
Weedman D. W., Huenemoerder D. P., 1985, ApJ, 291, 72
Weinzirl T. et al., 2011, ApJ, 743, 87
Whitaker K. E. et al., 2015, ApJ, 811, L12
Williams R. J., Quadri R. F., Franx M., van Dokkum P., Labbé I., 2009, ApJ, 691, 1879
Windhorst R. A. et al., 2002, ApJS, 143, 113
Wisnioski E. et al., 2015, ApJ, 799, 209
Wuyts S. et al., 2011, ApJ, 742, 96
Yan L., McCarthy P. J., Freudling W., Teplitz H. I., Malumuth E. M., Weymann R. J., Malkan M. A., 1999, ApJ, 519, L47

APPENDIX A: FAILURE RATE OF CONVERGENCE OF GALFIT

We summarize in Figs A1 and A2 the impact of the GALFIT failures to converge on the final stellar mass and Hα luminosities distributions, respectively. It is evident that there is no bin in either stellar mass or

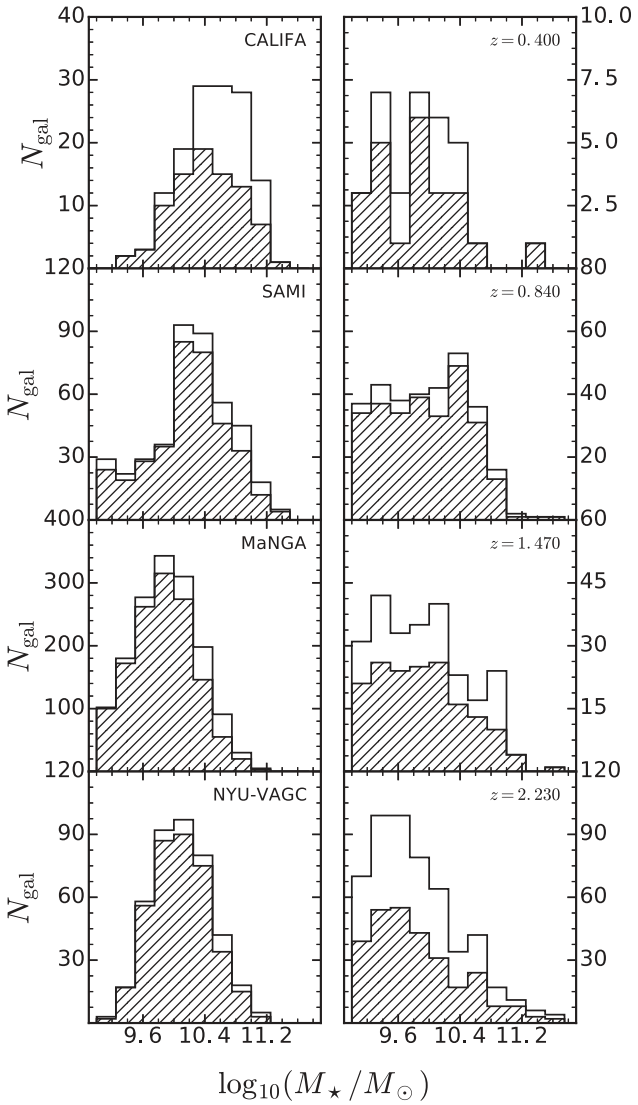

Figure A1. The distribution of galaxies for which GALFIT converged (as diagonally hatched histogram) compared to the total sample (solid black line) that was analysed as a function of their stellar mass.

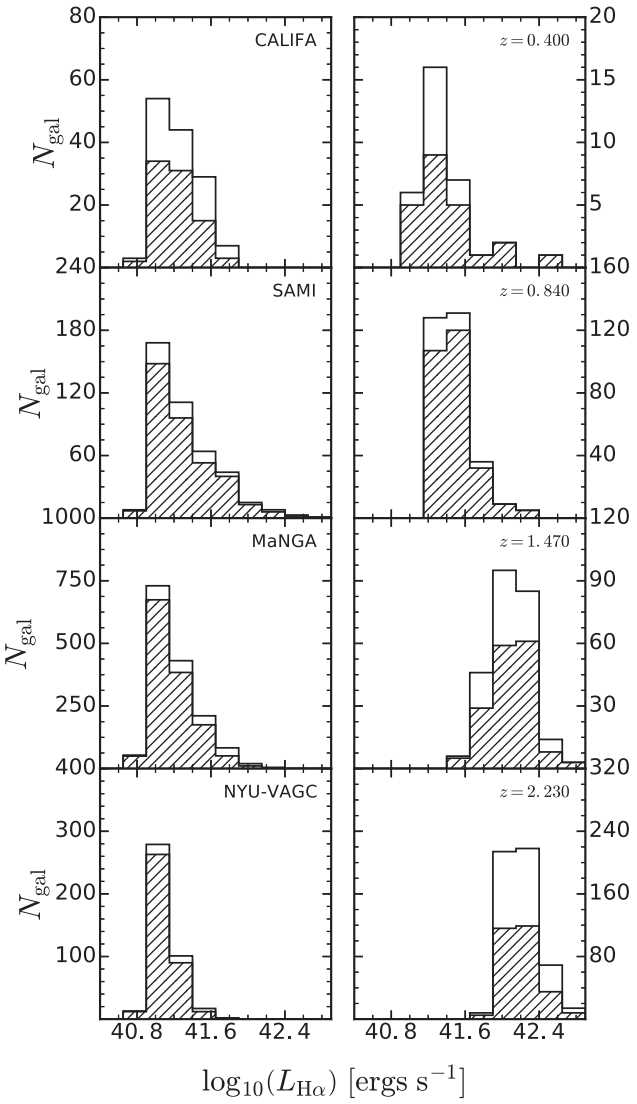

Figure A2. The distribution of galaxies for which GALFIT converged (as diagonally hatched histogram) compared to the total sample (solid black line) that was analysed as a function of their Hα luminosity.

Hα luminosity that is preferentially affected by a large failure rate when compared to other bins. This means that excluding galaxies for which GALFIT failed to converge from our final samples, from which we derive median structural parameters, does not introduce any additional bias in both stellar masses or Hα luminosities.

APPENDIX B: FURTHER REMARKS ON THE IMPACT OF COSMOLOGICAL DIMMING

We show in Fig. B1 the impact of the artificial redshift on the derived total magnitude of each galaxy. We compare absolute magnitudes and not the direct result from GALFIT, the apparent magnitudes, because it is the only way to compare magnitudes across different redshifts. We note that for comparison of the two quantities we correct the output absolute magnitude for the luminosity evolution that we impose for each redshift following the fit by Sobral et al. (2013a). We show that the impact is close to zero (<1 per cent), which makes it the more stable parameter against cosmological dimming. As for the axial ratio of galaxies (Fig. B2) we recover, on average, the observed value at $z \approx 0$, within an error of $\lesssim$5 per cent.

We observe a slight trend of overestimation of the axial ratio at smaller $(b/a)_{z\approx 0}$ and underestimation of the axial ratio at higher $(b/a)_{z\approx 0}$.

We investigate further the impact of cosmological dimming by comparing our fractional errors on the effective radius and Sérsic index to the input galaxy magnitude. In Fig. B3, we show that the brightest local galaxies are the ones most likely to have their sizes underestimated. The largest galaxies are the ones for which our sizes are not being recovered accurately as shown in Fig. 4, and they are also more likely to be the brightest galaxies in our sample, which helps to explain partly our results. The size overestimation that we observe on the faint end is possibly explained due to the likelihood of local galaxies having a bulge+disc structure which can prevent a single Sérsic profile to estimate the total extent of the galaxy. As we move the galaxy to higher redshifts, the substructures tend to no longer be resolved by the instrument PSF and GALFIT can more successfully measure sizes for the entire galaxy, hence estimating a larger size than the one measured at low redshift for the same galaxy. On the other hand, if the bulge-to-total light ratio is large, it may imply that we completely loose the disc component of the galaxy

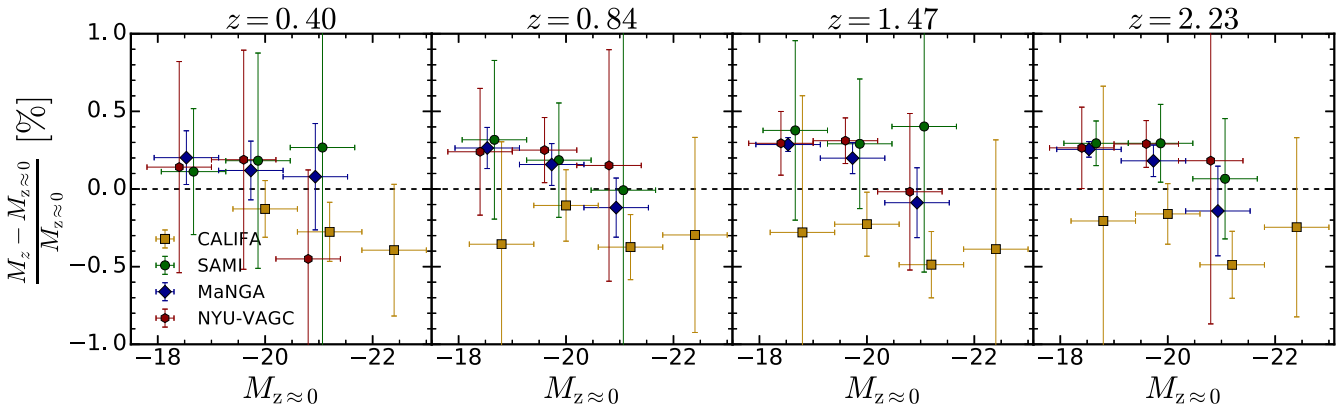

Figure B1. Variable difference $(M_z - M_{z\approx0})/M_{z\approx0}$, for each of the local subsamples and each redshift (one per column). Each symbol represents the median values in bins (with a minimum of 10 galaxies) of width of the horizontal error bar. The vertical error bar shows the error on the median value $\sigma/\sqrt{N_{\rm gal,bin}}$. The black dashed line pinpoints the accurate recovery. The recovered absolute magnitude is corrected for the luminosity evolution term of equation $\log_{10}[L(z)] = 0.45z + 41.87$. The impact on this parameter is tiny ($\lesssim$1 per cent) at all redshifts. Small horizontal offsets were applied to improve the readability of the plot.

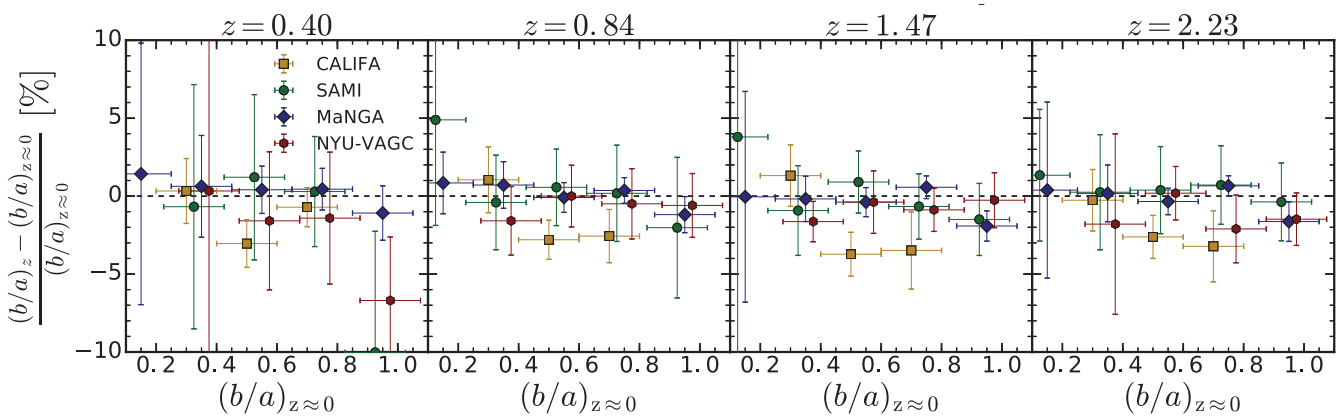

Figure B2. Variable difference $((b/a)_z - (b/a)_{z\approx0})/(b/a)_{z\approx0}$, for each of the local subsamples and each redshift (one per column). Each symbol represents the median values in bins (with a minimum of 10 galaxies) of width of the horizontal error bar. The vertical error bar shows the error on the median value $\sigma/\sqrt{N_{\rm gal,bin}}$. The black dashed line pinpoints the accurate recovery. The impact on this parameter is small ($\lesssim$5 per cent) at all redshifts, with a slight tendency to overestimate at smaller (b/a) and underestimate at higher (b/a). Small horizontal offsets were applied to improve the readability of the plot.

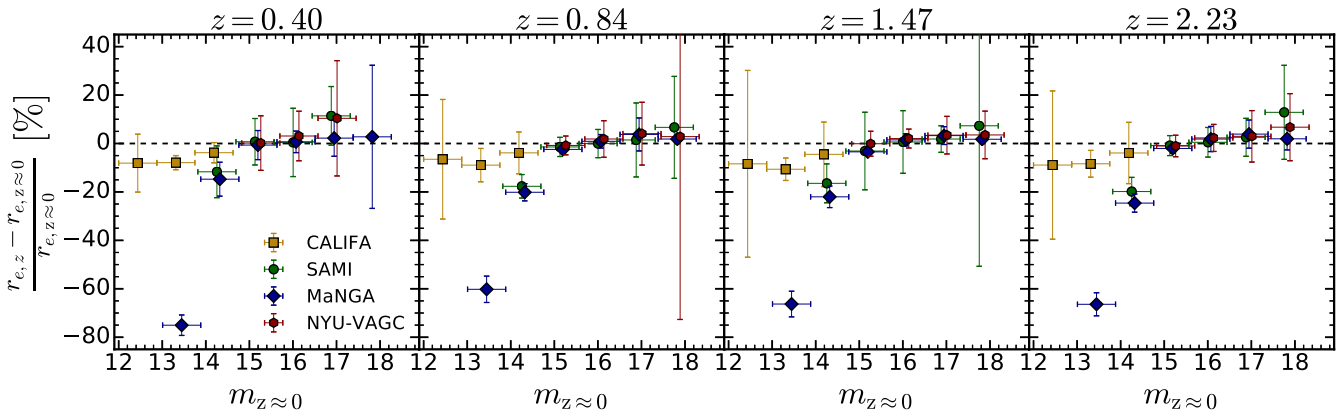

Figure B3. The fraction difference between the recovered and input effective radius $(r_{{\rm e},z} - r_{{\rm e},z\approx0})/r_{{\rm e},z\approx0}$, for each of the local subsamples and each redshift (one per column) as a function of input magnitude. Each symbol represents the median values in bins (with a minimum of 10 galaxies) of width of the horizontal error bar. The vertical error bar shows the error on the median value $\sigma/\sqrt{N_{\rm gal,bin}}$. The black dashed line pinpoints the accurate recovery. Small horizontal offsets were applied to improve the readability of the plot.

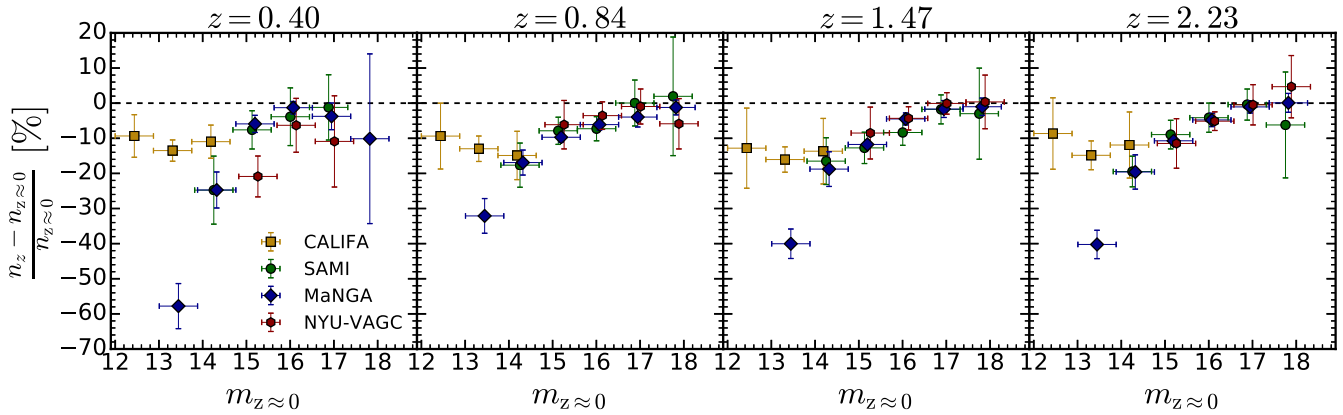

Figure B4. The fraction difference between the recovered and input Sérsic index $(n_z - n_{z\approx 0})/n_{z\approx 0}$, for each of the local subsamples and each redshift (one per column) as a function of input magnitude. Each symbol represents the median values in bins (with a minimum of 10 galaxies) of width of the horizontal error bar. The vertical error bar shows the error on the median value $\sigma/\sqrt{N_{\rm gal,bin}}$. The black dashed line pinpoints the accurate recovery. Small horizontal offsets were applied to improve the readability of the plot.

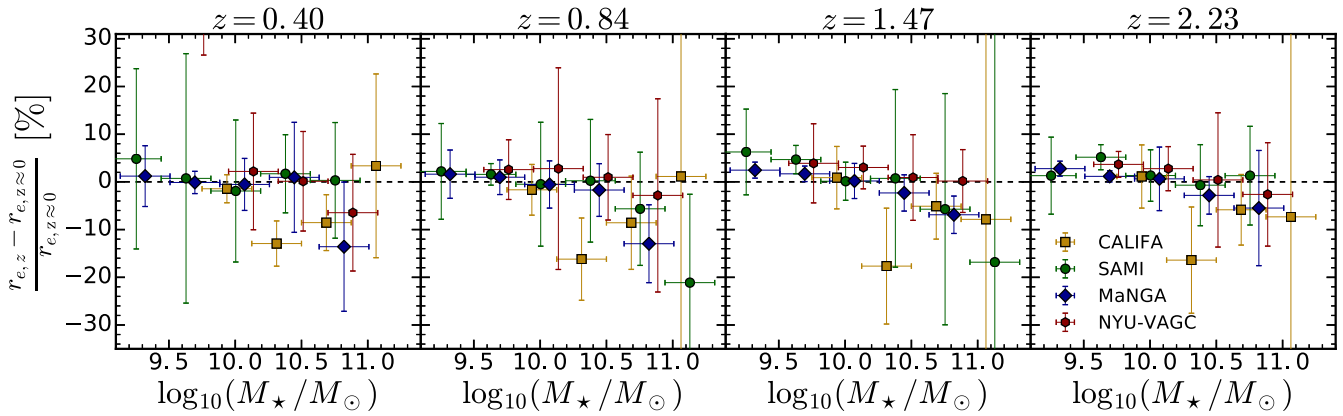

Figure B5. The fraction difference between the recovered and input effective radius $(r_{{\rm e},z} - r_{{\rm e},z\approx 0})/r_{{\rm e},z\approx 0}$, for each of the local subsamples and each redshift (one per column) as a function of input stellar mass. Each symbol represents the median values in bins (with a minimum of 10 galaxies) of width of the horizontal error bar. The vertical error bar shows the error on the median value. The black dashed line pinpoints a fractional difference of 0, i.e. when the recovered effective radius is the same as input. We do not observe a strong trend of fractional error deviation as a function of stellar mass. None the less, the most massive galaxies are expected to have their sizes underestimated at a level of $\lesssim$20 per cent. Small horizontal offsets were applied to improve the readability of the plot. In the $z = 0.40$ panel, there is one point from the SAMI survey at $(r_{{\rm e},z} - r_{{\rm e},z\approx 0})/r_{{\rm e},z\approx 0} \sim 40$ per cent for $\log_{10}(M_\star/{\rm M}_\odot) \sim 9.75$ that was excluded for visualization purposes.

and end up underestimating the size of the galaxy. This latter effect is expected to happen on the brighter galaxies since those are the ones we expected to have more likely experienced at least one major merger, which induces the formation of a prominent central bulge.

Since both effective parameters are linked through the same equation, we expect that a failure to reproduce the original effective radius leads in turn to a large error on the Sérsic index of the corresponding profile. And since the Sérsic index is the most unstable parameter of the profile, we are likely witnessing in Fig. B4 a simple consequence of the results shown before for the effective radius.

We have further separated the sample in two axial ratio bins (below and above $b/a = 0.5$) and re-did Figs B4 and B5. The results we find in these case are qualitatively the same and so we conclude that the axial ratio has no major impact on our ability to recover sizes and Sérsic indexes.

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