

Extending the evolution of the stellar mass – size relation at $z \leq 2$ to low stellar mass galaxies from HFF and CANDELS

Kalina V. Nedkova^{1*}, Boris Häußler², Danilo Marchesini¹, Paola Dimauro³, Gabriel Brammer^{4,5}, Paul Eigenthaler⁶, Adina D. Feinstein^{7†}, Henry C. Ferguson⁸, Marc Huertas-Company^{9,10}, Evelyn J. Johnston¹¹, Erin Kado-Fong¹², Jeyhan S. Kartaltepe¹³, Ivo Labbé¹⁴, Daniel Lange-Vagle¹, Nicholas S. Martis¹⁵, Elizabeth J. McGrath¹⁶, Adam Muzzin¹⁷, Pascal Oesch^{4,5,18}, Yasna Ordenes-Briceño⁶, Thomas Puzia⁶, Heath V. Shipley¹⁹, Brooke D. Simmons²⁰, Rosalind E. Skelton²¹, Mauro Stefanon²², Arjen van der Wel²³, and Katherine E. Whitaker^{5,24}

¹Department of Physics & Astronomy, Tufts University, 574 Boston Avenue, MA 02155, USA

²European Southern Observatory, Alonso de Cordova 3107, Casilla 19001, Santiago, Chile

³Observatório Nacional / MCTIC, Rua General José Cristino 77, Rio de Janeiro, RJ, 20921-400, Brazil

⁴Niels Bohr Institute, University of Copenhagen, Jagtvej 128, København N, DK-2200, Denmark

⁵Cosmic Dawn Center (DAWN), Copenhagen, Denmark

⁶Institute of Astrophysics, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, 7820436 Macul, Santiago, Chile

⁷Department of Astronomy and Astrophysics, University of Chicago, 5640 S. Ellis Ave, Chicago, IL 60637, USA

⁸Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218

⁹LERMA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Universités, UPMC Univ. Paris 06, F-75014 Paris, France

¹⁰Instituto de Astrofísica de Canarias, E-38200 La Laguna, Tenerife, Spain

¹¹Núcleo de Astronomía de la Facultad de Ingeniería y Ciencias, Universidad Diego Portales, Av. Ejército Libertador 441, Santiago, Chile

¹²Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA

¹³School of Physics and Astronomy, Rochester Institute of Technology, 84 Lomb Memorial Drive, Rochester NY 14623, USA

¹⁴Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Hawthorn, Victoria 3122, Australia

¹⁵NSM: NRC Herzberg, 5071 West Saanich Rd, Victoria, BC V9E 2E7, Canada

¹⁶Department of Physics and Astronomy, Colby College, Waterville, ME 04961, USA

¹⁷Department of Physics and Astronomy, York University, 4700 Keele St., Toronto, Ontario, M3J 1P3, Canada

¹⁸Department of Astronomy, Université de Genève, Chemin Pegasi 51, 1290 Versoix, Switzerland

¹⁹Department of Physics, McGill University, 3600 Rue University, Montréal, QC H3P 1T3, Canada

²⁰Department of Physics, Lancaster University, Bailrigg, Lancaster, LA1 4YB, UK

²¹South African Astronomical Observatory, PO Box 9, Observatory, Cape Town, 7935, South Africa

²²Leiden Observatory, Leiden University, NL-2300 RA Leiden, The Netherlands

²³Sterrenkundig Observatorium, Universiteit Gent, Krijgslaan 281 S9, B-9000 Gent, Belgium

²⁴Department of Astronomy, University of Massachusetts Amherst, 710 N Pleasant Street, Amherst, MA 01003, USA

Accepted 2021 June 7. Received 2021 June 3; in original form 2021 January 7

ABSTRACT

We reliably extend the stellar mass – size relation over $0.2 \leq z \leq 2$ to low stellar mass galaxies by combining the depth of Hubble Frontier Fields (HFF) with the large volume covered by CANDELS. Galaxies are simultaneously modelled in multiple bands using the tools developed by the MegaMorph project, allowing robust size (i.e., half-light radius) estimates even for small, faint, and high redshift galaxies. We show that above $10^7 M_\odot$, star-forming galaxies are well represented by a single power law on the mass–size plane over our entire redshift range. Conversely, the stellar mass – size relation is steep for quiescent galaxies with stellar masses $\geq 10^{10.3} M_\odot$ and flattens at lower masses, regardless of whether quiescence is selected based on star-formation activity, rest-frame colours, or structural characteristics. This flattening occurs at sizes of ~ 1 kpc at $z \leq 1$. As a result, a double power law is preferred for the stellar mass – size relation of quiescent galaxies, at least above $10^7 M_\odot$. We find no strong redshift dependence in the slope of the relation of star-forming galaxies as well as of high mass quiescent galaxies. We also show that star-forming galaxies with stellar masses $\geq 10^{9.5} M_\odot$ and quiescent galaxies with stellar masses $\geq 10^{10.3} M_\odot$ have undergone significant size growth since $z \sim 2$, as expected; however, low mass galaxies have not. Finally, we supplement our data with predominantly quiescent dwarf galaxies from the core of the Fornax cluster, showing that the stellar mass–size relation is continuous below $10^7 M_\odot$, but a more complicated functional form is necessary to describe the relation.

Key words: galaxies: evolution – galaxies: structure — galaxies: high-redshift

1 INTRODUCTION

Understanding the relative importance of the different formation and evolutionary mechanisms that are responsible for shaping quiescent and star-forming galaxies at different epochs of the Universe continues to be one of the most fundamental goals of extragalactic astronomy. Signatures of the physical mechanisms that drive the formation pathways through which a galaxy evolves are imprinted on its structure, making galaxy structure and size key observational quantities. While it has now been well established that both quiescent (early-type) and star-forming (late-type) galaxies were more compact at higher redshift (e.g., Daddi et al. 2005; Trujillo et al. 2007; van Dokkum et al. 2008), it is still under debate if the size growth that galaxies have undergone since the early Universe is primarily driven by major mergers (e.g., Naab et al. 2007; McIntosh et al. 2014) minor mergers (e.g., Buitrago et al. 2008; Naab et al. 2009; Oser et al. 2012; Newman et al. 2012), feedback (e.g., Fan et al. 2008; Damjanov et al. 2009), secular processes, or other mechanisms.

Although galaxies evolve via complex, nonlinear processes, their structural parameters exhibit a number of tight scaling relations that aid us in understanding their evolution. The stellar mass – size relation is particularly interesting because both stellar mass and effective radius have been shown to correlate with the quenching process at least for galaxies with stellar masses above $10^9 M_{\odot}$ (e.g., Omand et al. 2014; Chen et al. 2020). While the exact nature of the mechanisms that quench galaxies remains unknown, it seems there are two main quenching channels: one that quenches high mass galaxies and another that quenches low mass galaxies. Peng et al. (2010b) termed these ‘mass quenching’ and ‘environment quenching’ where galaxies that undergo mass quenching are quenched by internal processes such as stellar feedback and AGN feedback. Environment quenching primarily impacts less massive satellite galaxies, which are more likely to be quenched by environmental processes such as ram pressure stripping or harassment from neighbouring objects.

Numerous models have been proposed to explain the quenching process and its relation to stellar mass and galaxy size. van der Wel et al. (2009) suggest that star-forming galaxies evolve on the stellar mass–size plane until they reach a redshift-dependent velocity dispersion threshold. Above this velocity dispersion threshold, galaxies can no longer efficiently form stars, and so, they quench. After quenching, these galaxies undergo subsequent dry, minor mergers which result in the observed trend that quenched galaxies experience a steeper growth on the stellar mass–size plane. van Dokkum et al. (2015) also describe a simple picture explaining these trends, where galaxies evolve on the mass–size plane along ‘parallel tracks’ according to $r \propto M^{0.3}$, where the growth is predominantly driven by star formation until they reach a central density, after which they quench. They then experience steeper evolutionary tracks according to $r \propto M^2$, where the growth is again primarily driven dry mergers. Recently, Chen et al. (2020) proposed another model in which the radius of star-forming galaxies indirectly dictates when they will quench. They argue that at fixed stellar mass, star-forming galaxies with larger sizes have smaller black holes due to their lower central densities. Therefore, larger galaxies must evolve to higher stellar masses in order to quench, ultimately resulting in a scenario where smaller star-forming galaxies quench at higher redshift. This toy model successfully explains several observed characteristics including the parallel tracks that van Dokkum et al. (2015) describe.

Major efforts have gone into constraining the exact behaviour of the quiescent and star-forming sequences on the mass–size plane. We now know that star-forming and quiescent galaxies follow different tracks (e.g., van der Wel et al. 2014; van Dokkum et al. 2015; Dimauro

et al. 2019) such that star-forming galaxies tend to be larger in size than quiescent galaxies across a large range of redshifts and stellar masses. van der Wel et al. (2014) have shown that the sizes of star-forming galaxies are proportional to the virial radius of their host dark matter halo, likely a result of conservation of angular momentum (Somerville et al. 2018). Quiescent galaxies, on the other hand, follow a steeper slope on the mass–size plane and are more compact than star-forming galaxies of similar masses (e.g., Cimatti et al. 2008; van der Wel et al. 2014; Dimauro et al. 2019). In recent years, Mosleh et al. (2017) and Suess et al. (2019) have argued that using half-mass, as opposed to half-light, radii results in stellar mass – size relations which are significantly shallower and suggest that galaxies, especially star-forming ones, have not grown significantly in size since the early Universe. Suess et al. (2019) therefore claim that the evolution that is observed in the stellar mass – size relation, when half-light radii are used, is primarily due to colour gradients. While these results pose a challenge to our current picture of galaxy evolution, multi-wavelength software, such as that developed as part of the MegaMorph project (Häußler et al. 2013; Vika et al. 2013), which we use in this work, are well-suited to address these questions in the future.

While these observational results have provided a valuable test of evolutionary models, a key drawback of most previous studies has been their relatively high mass limits of galaxies above $\sim 10^{9.5} M_{\odot}$. Galaxies in this high stellar mass regime are strongly gravitationally bound systems, which protects them from environmental influences (e.g., Moore et al. 1996) and therefore most likely evolve via the ‘mass quenching’ channel. In order to understand all galaxies, both high mass and low mass, it is important to constrain the stellar mass – size relation for low mass galaxies across a wide redshift range.

Significant improvements have also been made in terms of extending the stellar mass – size relation to high redshift (e.g., Allen et al. 2017; Hill et al. 2017) as well as to low stellar masses (e.g., Shen et al. 2003; Baldry et al. 2012; Lange et al. 2015; Morishita et al. 2017). Unfortunately, high redshift studies are often limited to high mass objects while studies that probe the evolution of low mass galaxies have been restricted to local galaxies (see Morishita et al. 2017, for $z \sim 0.7$). Thankfully, hydrodynamical simulations have allowed us to overcome the challenge of studying low mass systems at high redshift and have proven to be an invaluable test of our current understanding of galaxy evolution. Through hydrodynamical simulations, key scaling relations can be studied in a way that is impossible with observations alone. Furlong et al. (2017) and Genel et al. (2018) have shown that the EAGLE and IllustrisTNG simulations, respectively, are remarkably successful in reproducing the observed stellar mass – size relation. Both works found results that were consistent with van der Wel et al. (2014) and were able to evolve their galaxies in time to provide key insights on how star-forming and quiescent galaxies assemble over time. But, some tensions still remain between simulations and observed galaxy sizes, especially for small galaxies (e.g., Pillepich et al. 2018).

In order to solve these tensions, it is necessary to explore how compact, low mass galaxies evolve. In this work, we present such an analysis and extend the stellar mass – size relation to low stellar masses and to higher redshifts than previous observational studies (e.g., Trujillo et al. 2004; Oesch et al. 2010; Mosleh et al. 2012; Barro et al. 2014; van der Wel et al. 2014; Holwerda et al. 2015; Shibuya et al. 2015; Hill et al. 2017; Dimauro et al. 2019) by making use of the depth of the Hubble Frontier Fields (HFF) (Lotz et al. 2017) and the large area probed by the CANDELS images (Koekemoer et al. 2011), as well as the multi-wavelength capabilities of the software

developed as part of the MegaMorph project (Häußler et al. 2013; Vika et al. 2013).

The paper is structured as follows. In Section 2, we explain the technical setup for the galaxy profile fitting and modelling. In Section 3, we describe the galaxy sample used for this study. In Section 4, we show the stellar mass – size relation for both quiescent and star-forming galaxies, and we describe the flattening of stellar mass – size relation for quiescent galaxies in Section 5. Finally, we summarise our findings in Section 6. Throughout this paper, we use a Hubble constant of $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and cosmological density parameters $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$. We assume a Chabrier (2003) initial mass function (IMF) for all estimates of stellar mass and all magnitudes are quoted in the AB system.

2 MEGAMORPH

The tools developed as part of the MegaMorph (Measuring Galaxy Morphology) project (Häußler et al. 2013; Vika et al. 2013) allow robust size measurements as well as a characterisation of morphological properties to be reliably established, even for high redshift galaxies, because they use and model data at all available wavelengths simultaneously, effectively increasing the signal-to-noise ratio (S/N) of the data. Because of this capability, we choose to model the light profiles of all galaxies in our sample with these tools. We address key aspects of the software below, but refer the reader to Häußler et al. (2013) for details about the codes and their reliability and accuracy.

MegaMorph builds on GALAPAGOS (Barden et al. 2012) and GALFIT (Peng et al. 2010a), but both codes have been modified to allow for multi-wavelength fitting. This modification allows galaxy light profiles to be more reliably fit than with other commonly used software, such as GIM2D (Simard 1998; Simard et al. 2002), BUDDA (de Souza et al. 2004), PyMorph (Vikram et al. 2010), GALFIT (Peng et al. 2002, 2010a), and IMFIT (Erwin 2015). GALFITM, which is based on GALFIT (Peng et al. 2010a), is a two-dimensional fitting code designed to extract structural properties from galaxy images. The galaxy models allow the user to fit any number of components and functional forms. In this work, we model every object with a single Sérsic component in order to measure properties of each galaxy as a whole. GALAPAGOS-2 automates the source detection, the two-dimensional Sérsic profile modelling using GALFITM, and catalogue creation. Both codes have been adapted from their original versions to fit multiwavelength data by replacing the galaxy model parameters with wavelength-dependent functions – namely Chebyshev polynomials of the first kind (Abramowitz & Stegun 1965). While the tools developed as part of the MegaMorph project allow structural parameters to vary systematically with wavelength, user specifiable limits and parameters must be appropriately chosen. In particular, the degrees of freedom with which each parameter is allowed to vary as a function of wavelength must be specified. This is done in order to balance the advantage of multi-band fitting with constraining the model parameters to change with wavelength in a physically meaningful way. We explicitly discuss the degrees of freedom with which we allow the magnitudes, sizes, and Sérsic indices to vary below.

Spectral energy distributions (SEDs) have a complex wavelength dependence and therefore cannot be reproduced with low-order polynomials. In our fitting, we ensure that SEDs can be accurately recovered by giving the fitting function in GALFITM full freedom. While other parameters also have some wavelength dependence, full freedom is not always necessary nor, in fact, advisable, in order to make use of the advantages of multi-band fitting. For instance,

the measured size of a galaxy depends on the wavelength at which the observation is made (e.g., Evans 1994; La Barbera et al. 2010; Häußler et al. 2013; Vulcani et al. 2014), such that galaxies are often found to be much smaller when measured in redder bands when compared to bluer bands. This is because sizes measured in bluer bands are more sensitive to the younger stellar population within galaxies. These, generally speaking, tend to reside in galaxy disks, which are typically more extended than their respective bulge components. Sizes measured in redder bands instead reflect the extent of the older stars, which are generally found in the very central part of galaxies and/or their spheroidal components, i.e., bulges. As a result, sizes measured in blue bands tend to be larger than those measured at longer wavelengths. To allow for this change in size in our galaxy models, we decide to allow some variation with wavelength. As we have sufficiently many bands available, we allow the effective radius to vary as second order Chebyshev polynomial, enabling us to recover the size’s smooth wavelength dependence. Sérsic indices also have a strong wavelength dependence for similar reasons. If a galaxy consists of both a bulge and a disk component, as most spiral galaxies do, then the Sérsic index measured at longer wavelengths will reflect the light profile of the bulge, while the Sérsic index measured in the bluer bands would reflect the light profile of the disk (Vulcani et al. 2014). The Sérsic index of such a galaxy would then increase with wavelength. Because of this dependence and following the same arguments with the number of images/bands available, the Sérsic indices are also allowed to vary as second order Chebyshev polynomials. Finally, for the centre positions of the profiles, as well as axis ratios and position angles, we choose to fit constant values, with no wavelength dependence, by fitting zeroth order Chebyshev polynomials.

We further discuss the advantages of fitting our data with a multi-wavelength approach in the following section, and show fitting results from the software developed as part of the MegaMorph project in §3.7.

3 DATA

We use both CANDELS and HFF data to construct magnitude-limited samples of star-forming and quiescent galaxies over a large stellar mass and redshift range. In this section, we describe the data used in this work, from imaging/aperture photometry in §3.1 and §3.2 for the HFF and CANDELS fields, respectively, to size and stellar mass estimates in §3.4 and §3.5. We present several methods for distinguishing star-forming and quiescent galaxies in §3.6 and show galaxy light profile models from the MegaMorph tools in §3.7.

3.1 HFF

The HST Frontier Fields (HFF) program (Lotz et al. 2017) provides a unique data set that allows structural parameters to be obtained for bright objects as well as faint, high-redshift galaxies. The Frontier Fields consist of six cluster fields centred on strongly lensed galaxy clusters that were imaged in parallel with six blank fields. The HFF-DeepSpace photometric catalogues (Shingley et al. 2018) combine images from the Advanced Camera for Science (ACS) and Wide Field Camera 3 (WFC3) with K_s imaging, which was taken as part of the ‘K-band Imaging of the Frontier Fields’ (KIFF) project (Brammer et al. 2016), from the Very Large Telescope (VLT) HAWK-I and Keck-I MOSFIRE instruments. These data were combined in a consistent way to provide coverage in up to 17 filters spanning wavelengths from the UV to NIR. The HFF-DeepSpace catalogues

also contain post-cryogenic Spitzer imaging at $3.6\mu\text{m}$ and $4.5\mu\text{m}$ from the Infrared Array Camera (IRAC), as well as any available archival IRAC $5.8\mu\text{m}$ and $8.0\mu\text{m}$ data. In addition to the photometric catalogues, Shipley et al. (2018) provide catalogues of photometric redshifts and stellar population properties. We refer the reader to Shipley et al. (2018) for more details on the properties of the HFF-DeepSpace catalogues and how they were constructed, but we briefly address key aspects of the catalogue construction that are relevant for this work.

We use seven of the WFC3 bands, which are consistently available for all parallel and cluster fields (i.e., F435W, F606W, F814W, F105W, F125W, F140W, and F160W). We choose to exclude any ground-based and Spitzer imaging in order to keep the image quality consistent across all bands. The images that we use have pixel scales of $0.06''/\text{pixel}$. The depth of each field is listed in Tables 7 and 8 of Shipley et al. (2018) for the F814W and F160W bands, respectively. The HFF-DeepSpace catalogues are unique in that they model bright cluster galaxies (termed bCGs, although it should be noted that this terminology is different from the traditional use of BCG referring to the brightest cluster galaxy) together with intra-cluster light (ICL) and remove them from the images. This is done in order to obtain information about background or underlying objects, which is crucial for this work. We test if the removal of the ICL has any impact on the galaxy sizes that we obtain by comparing the size distributions of galaxies in the parallel fields to those in the cluster fields. We find that the size distributions are similar and consistent, indicating that the ICL subtraction does not introduce significant systematics.

We model the light profiles of galaxies from the ‘bCG subtracted’ images in order to gain information about objects that would otherwise be outshone by neighbouring bCGs making reliable fitting of their light profiles difficult, if not impossible. In an additional set of fits, we model the bCGs from the original images separately. This ensures that our sample consists of both the bright bCGs and any faint objects that may lie ‘behind’ them. For the parallel fields, we find no significant improvement between using the bCGs subtracted images versus the original images in terms of the number of objects recovered; however, we follow the same approach on those fields anyway, for consistency. An additional benefit of using the bCG subtracted images is that detecting the faint objects in the vicinity of these large bCGs can be drastically improved by removing those brightest objects first. Without removing them, it is simply impossible to find a SExtractor (Bertin & Arnouts 1996) setup that works on these small and faint objects, reducing our sample size by 10-15% from this effect alone.

To avoid biasing measured sizes and magnitudes, we exclude any objects from the cluster fields that may be lensed. GALAPAGOS-2, has two features to ensure this. The first is meant to flag ‘detections’ that are not real. Mostly, these are hot pixels at the edge of the field, but this also allows to ensure that galaxies are detected as one object, rather than being split up, e.g., due to internal structure. These objects are removed before starting the galaxy light profile modelling. The second allows flagging objects as ‘important, yet not real/wanted’. A prime example of this are detected diffraction spikes of stars, which need to be modelled as secondary objects (see Barden et al. 2012 for primary, secondary, and tertiary nomenclature), in order to not influence the fit results of nearby objects. However, they do not portray real objects and as such are only treated as secondary/tertiary objects, never as primary objects. They are removed from the object catalogue by the end of the fitting process. This latter feature allows us to sensibly deal with lensed objects, especially arcs.

In our sample analysed in subsequent sections, we further limit the redshift of the objects in the cluster fields to be either consistent with

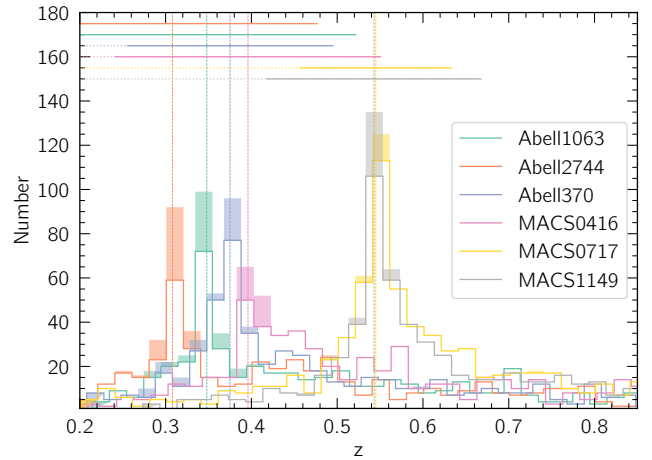


Figure 1. The redshift distribution of all six HFF cluster fields. Histograms are colour coded by cluster, with the corresponding vertical dashed lines indicating the spectroscopic redshift of each cluster. Galaxies with photometric redshifts within $3 \times \sigma_{\text{NMAD}}$ of the cluster redshift (horizontal lines) are classified as cluster members. Our final sample includes all cluster member galaxies as well as galaxies with redshifts lower than the cluster redshift and is indicated by the dotted horizontal lines. We exclude objects with redshifts higher than z_{upper} listed in Table 1 in order to avoid biasing any measurements by including lensed galaxies. The parts of the histogram that are filled in represent the redshift distribution of the bCGs.

Table 1. For each HFF cluster field, we list the spectroscopic redshift (z_{spec}), the normalised median absolute deviation (σ_{NMAD}), and the redshift limit below which we include objects (z_{upper}).

Cluster Field	z_{spec}	σ_{NMAD}	z_{upper}
Abell1063	0.348	0.043	0.522
Abell2744	0.308	0.043	0.477
Abell370	0.375	0.029	0.522
MACS0416	0.396	0.037	0.550
MACS0717	0.545	0.019	0.632
MACS1149	0.543	0.027	0.668

the cluster redshift or lower, in order to avoid any lensing effects. Motivated by Morishita et al. (2017), we use the normalised median absolute deviations (σ_{NMAD}), which have been derived by Shipley et al. (2018) for the HFF, to find cluster members. We define any galaxy with a photometric redshift that satisfies

$$3 \times \sigma_{\text{NMAD}} \geq \frac{|z - z_{\text{clu}}|}{1 + z_{\text{clu}}} \quad (1)$$

as having a redshift that is consistent with the cluster redshift. In Equation 1, z is the photometric redshift of each galaxy, z_{clu} is the spectroscopic redshift of the cluster, and σ_{NMAD} is the normalized median absolute deviation, which is obtained by comparing estimated photometric redshifts and confirmed published spectroscopic redshifts from the literature. σ_{NMAD} is different for each cluster field and is listed in Table 1.

Figure 1 shows the redshift distribution of the cluster galaxies and redshift cuts we apply in an effort to exclude galaxies whose magnitudes or sizes may be affected by lensing. For each cluster, a clear peak in the number of galaxies can be seen at the spectroscopic redshift (indicated by the vertical dotted lines) of each cluster. The

spectroscopic redshifts of the clusters are reported in Table 1. The redshift distribution of the bCGs are shown as filled histograms. As expected, the majority of bCGs have redshifts that are consistent with the spectroscopic redshifts of the cluster field they belong to. We do not restrict the redshift for the HFF parallel fields as no significant lensing effects are to be expected.

3.2 CANDELS and 3D-HST

We measure the structural parameters from the Hubble Space Telescope (HST) imaging from the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS) for the galaxies in all five CANDELS fields (i.e., GOODS-S, GOODS-N, COSMOS, EGS, and UDS) (Koekemoer et al. 2011). While the CANDELS fields are not as deep as the HFF fields, CANDELS allows us to study a much larger galaxy sample due to the area covered by this survey. Each object from the $0.03''/\text{pixel}$ CANDELS images is matched by position to galaxies in the 3D-HST photometric catalogues (Skelton et al. 2014), limiting the maximum spacial separation to $0.2''$. We note here that the $0.03''/\text{pixel}$ images provided by CANDELS that we use in this work are not the same as those used by van der Wel et al. (2012) and Dimauro et al. (2018), who derive galaxy properties from the $0.06''/\text{pixel}$ CANDELS images. This choice has no effect on the fitting parameters as we find excellent agreement in the derived magnitude, size, and Sérsic index with both van der Wel et al. (2012) and Dimauro et al. (2018).

3.3 Sample Selection

Our initial sample consists of 39685 objects from the HFF and 106663 objects from CANDELS after spatially matching objects from the images to the galaxies in the HFF-DeepSpace (Shiple et al. 2018) and 3D-HST photometric catalogues (Skelton et al. 2014), and applying the redshift limit described in §3.1 to galaxies in the HFF cluster fields. We first require that galaxies have `FLAG_GALFIT=2`, meaning that they have been successfully modelled and fit with the fitting software, `GALFITM` and `GALAPAGOS-2`, which are discussed in detail in §2. This reduces our HFF and CANDELS samples to 22090 and 105717 objects, respectively. We note that a significantly higher fraction of HFF galaxies were unsuccessfully modelled compared to CANDELS galaxies. This is because we only model HFF objects that are within the F160W footprint since we will apply a completeness cut in the H-band anyway. All CANDELS objects have been modelled, not just those with F160W coverage; however, we will apply a magnitude cut in the H-band for CANDELS as well, effectively requiring that all CANDELS objects in the final sample are within the F160W footprint, too.

For the separate bCG run, in which 330 galaxies were successfully modelled, we visually inspect the fits to ensure that they are reliable and found no major issues, apart from the typical effects of fitting two-component systems with one-component models. We do not visually inspect all objects from the ‘bCG subtracted’ images, as this would be an unfeasible task and the modelling for these ‘standard’ galaxies is much closer to the tests carried out in (Häubler et al. 2013) and other publications. We have, however, also looked at a subset of these (~ 700 galaxy models), to ensure that `GALAPAGOS-2` works as intended. For each galaxy in the HFF, we fit morphological parameters in seven bands and we require that each galaxy that is modelled is covered by at least three of those seven images. For the CANDELS fields, we have a different number of bands available for each field, with a minimum of four bands for UDS. Therefore,

requiring that galaxies are covered by at least three bands, as we do for HFF galaxies, would be a strict constraint. Hence, we require that every galaxy that is modelled in CANDELS, have sufficient data (i.e., $>30\%$ of pixels within the primary ellipse are *not* masked) in at least two images. Galaxies that do not satisfy this have unsuccessful models, and are hence removed by our first quality cut.

For galaxies which have been successfully modelled, we then require that they have a spectroscopic or a reliable photometric redshift in the range $0.2 \leq z \leq 2$. For our CANDELS sample, we also use redshifts derived from grism spectra (Momcheva et al. 2016) when available. We include objects that have `use_phot=1`, which removes stars (i.e., objects with `star_flag=1`), sources close to bright stars, and objects with $S/N < 3$ from the photometry aperture in the F160W band. These cuts leave 11438 and 71798 objects in the HFF and CANDELS samples, respectively.

To ensure that we only select objects that are bright enough to be modelled, we include only HFF galaxies that are one magnitude brighter than the 90% detection completeness limit in the F160W band when the injected mock objects are not allowed to overlap with detected objects. These completeness limits are reported in Table 8 of Shipley et al. (2018). For CANDELS, we exclude any object $>24.1\text{mag}$ in the F160W band, following the same idea. After applying these magnitude cuts, we have 8686 HFF galaxies and 27555 CANDELS galaxies. Additionally, we apply cuts based on the model parameters, which are all motivated by Häubler et al. (2013). These cuts are as follows: (i) While we allow Sérsic indices to vary between 0.2 and 12 in the fitting, we exclude objects with a Sérsic index below 0.205 and above 8 (in any band) because these objects have run into the Sérsic index constraint, are often found to be point sources, or are unreliably fit based on visual inspection. (ii) The half light radius in the modelling is restricted to be between 0.3 and 400 pixels to avoid using results from objects that are unphysical. Therefore, in the sample selection, we only include objects with effective radii between 0.305 and 395 pixels to exclude any objects that have ran into either of these constraints. For the separate bCG run, we remove the 400 pixel size limit since the bCGs can be – and are – larger than this. (iii) We require that the magnitude measured from all bands is within 5 magnitudes of the magnitude measured from our source extractor runs (`MAG_BEST` + an empirically derived offset between bands). Although this constraint is rather lenient, it is again intended to ensure that we are not using objects that have unreliable light profile fits. This ensures a sample with reliable fitting parameters of 6632 HFF galaxies and 24736 CANDELS galaxies. Finally, we limit the stellar mass of the galaxies that we use, but we provide a detailed discussion of the stellar mass cuts in §3.5.

3.4 Effective Radius

For each galaxy, we measure the half-light radius, or effective radius, along the major axis in each observed band using `GALFITM` and `GALAPAGOS-2`, which are discussed in §2. Half-light radii have been used in the literature to study galaxy sizes at least since de Vaucouleurs (1948). It has long been known, however, that using alternative radii can result in significantly different scaling relations (see e.g., Graham 2019; Trujillo et al. 2020, for recent results). In addition to using half-light radii, we have also derived the stellar mass – size relations for star-forming and quiescent galaxies using radii that contain 20% and 80% of the total light by converting the measured half-light radii into these values by analytically integrating the Sérsic profile of each object. We find that the stellar mass – size relations remain qualitatively the same in that there is still a flattening for the quiescent sample, while the star-forming relation is

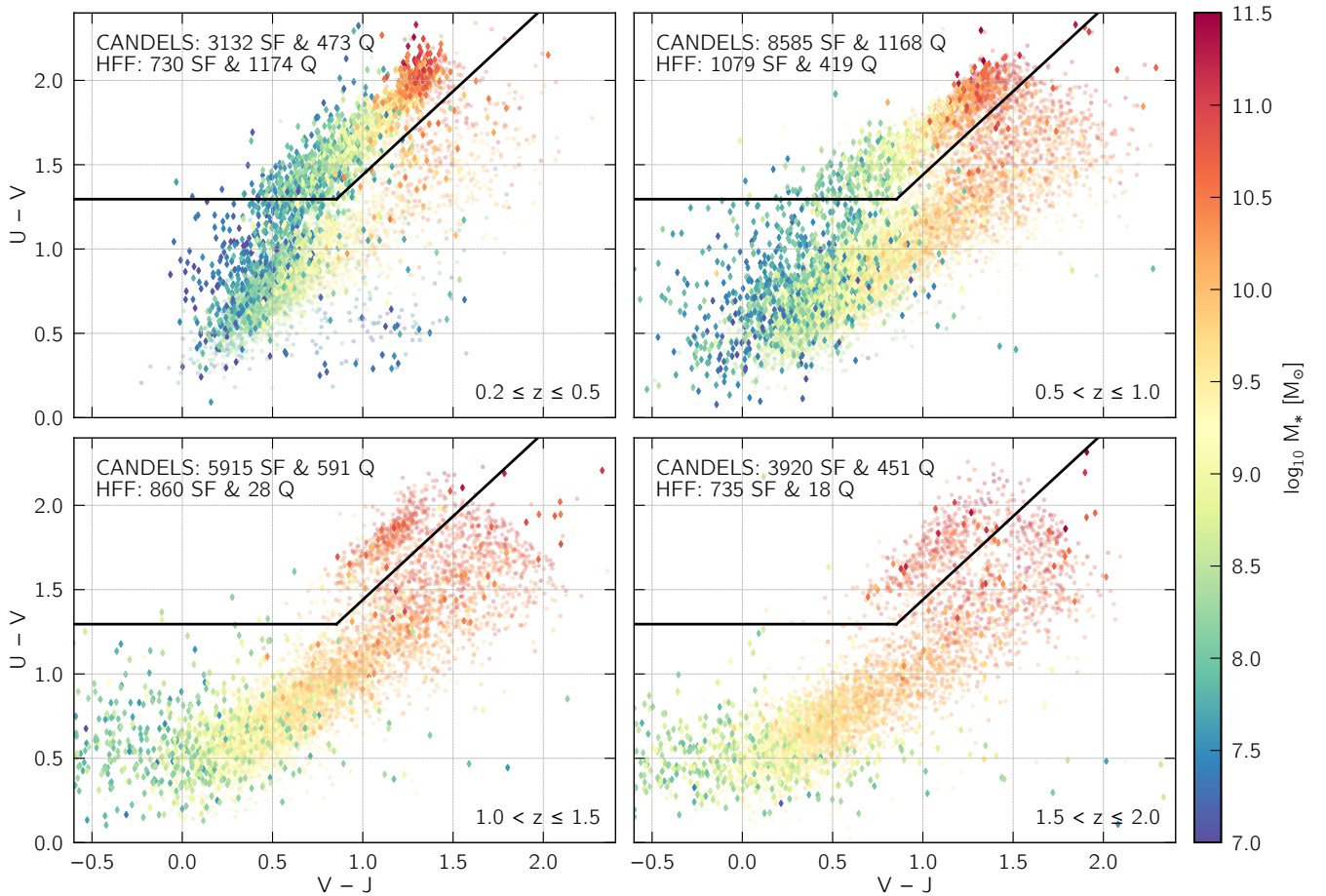


Figure 2. The UVJ diagram of the entire sample divided into four redshift bins. Objects from the HFF parallel and cluster fields are shown as diamonds and galaxies from the CANDELS fields are shown as dots. All galaxies are colour-coded by their stellar mass, which is estimated by the FAST code (Kriek et al. 2009) and corrected for the difference in the F160W magnitude from GALFITM and the photometric catalogues (see text for further details). The redshift range of each panel is indicated at the bottom right corner and the number of star-forming and quiescent objects from HFF and CANDELS is indicated in the top left.

well represented by a single power law. Apart from the expected shift in normalisation, the slopes of the curves remain largely unchanged, with one notable exception: for quiescent galaxies, the slope of the mass – size relation at the low mass end is weakly dependent on the radius definition that is used. This change, however, is small and well within the scatter of the relations and indicates a change of Sérsic index with galaxy mass, examining which is not part of this work.

Using the Chebyshev polynomials returned by GALFITM, we are able to derive the rest-frame 5000\AA size, in order to have a redshift independent measure of galaxy size, allowing a clean comparison over all redshifts. We additionally derive the rest-frame 4000\AA and 6000\AA size of each galaxy; however, we find that the sizes at the three rest-frame wavelengths are consistent. Therefore, the derived stellar mass – size relation does not depend strongly on the rest-frame wavelength that is used, at least in the $4000 - 6000\text{\AA}$ range. Throughout the remainder of this paper, we choose to use the 5000\AA size for easier comparison to the literature and because it can be obtained without extrapolating the polynomial for higher redshift galaxies.

3.5 Stellar Mass

The 3D-HST photometric catalogues (Skelton et al. 2014; Momcheva et al. 2016) contain photometric and grism redshift estimates as well as stellar population parameters determined by the FAST code, which fits stellar population synthesis models to the measured SEDs of galaxies to infer several galactic properties (Kriek et al. 2009). The HFF-DeepSpace catalogue (Shipley et al. 2018) also provides stellar masses that have been derived in a similar way. We use these FAST-derived stellar masses with their 1σ errors, which are determined by using the Monte Carlo simulation option in FAST. We test how the stellar mass changes depending on the metallicity by running FAST for the HFF data with two options: (i) fixing the metallicity of all galaxies to solar metallicity and (ii) allowing the metallicity to vary. We find no significant difference in the derived stellar mass between the two runs, therefore we choose to use the run with fixed metallicity in order to be consistent with the way the CANDELS data were modelled.

As the galaxy sizes that we measure are obtained from modelling Sérsic profiles that integrate the profile out to infinity, while the stellar masses in the HFF-DeepSpace and 3D-HST catalogues are determined from aperture photometry that miss some light at large radii, a correction must be applied to the stellar mass estimates in

order for them to be consistent with the profiles used to measure galaxy sizes. Following [van der Wel et al. \(2014\)](#), we correct for the difference between the F160W flux from the photometric catalogues and the F160W magnitude as measured with GALFITM, in the final stellar mass that we use. However, we find that this correction only has an effect on the largest galaxies, for which the corrected masses are on average a factor of 1.08 larger. There is no significant effect on the masses of the majority of objects, and the conclusions in this paper are not significantly changed by this correction.

In our final sample, we include galaxies with corrected stellar masses above $10^7 M_{\odot}$. We obtain stellar mass uncertainties from the upper and lower 68th percentiles that the FAST code ([Kriek et al. 2009](#)) returns and we exclude objects for which the stellar mass uncertainty is ≥ 2 dex. This cut is a lenient one, as it is intended to only remove galaxies for which the SED modelling is highly unreliable. After applying these last quality cuts, the final sample consists of 5043 HFF and 24235 CANDELS galaxies. A higher fraction of HFF objects are removed by the stellar mass cuts because the HFF consists of more low mass galaxies compared to CANDELS, and therefore, the $M_{*} \geq 10^7 M_{\odot}$ cut that we apply removes more HFF objects. In Figures 2 and 3, we further provide the number of star-forming and quiescent galaxies that fall into each redshift bin, noting that the HFF consists of many quiescent galaxies at $z \leq 1$, but only 28 and 18 in $1.0 < z \leq 1.5$ and $1.5 < z \leq 2.0$, respectively. This is expected as our sample does not consist of any cluster field galaxies at $z \geq 1$ as a result of the redshift limits that we impose in §3.1.

In Figure 2, we show the UVJ diagram colour-coded by stellar mass for four redshift bins spanning $0.2 \leq z \leq 2.0$. Rest-frame colours are derived using EAZY ([Brammer et al. 2008](#))¹. For galaxies in both the HFF, shown as diamonds, and in the CANDELS fields, shown as dots, the most massive galaxies have redder U–V and V–J colours, as expected from other studies. This suggests that the stellar mass estimates are reliable. We also note that we detect more low mass galaxies at low redshifts than at high redshifts, which is also expected. The UVJ boundaries that separate the star-forming and quiescent galaxies, shown as black lines in Figure 2, are exactly matched to those used by [van der Wel et al. \(2014\)](#).

For any magnitude-limited sample, the minimum stellar mass at which galaxies can be observed depends on their stellar mass-to-light ratio (M/L) and redshift. The M/L depends on the stellar populations and is therefore reflected in galaxy colours, meaning that quiescent galaxies will have higher M/L because of their older stellar populations and redder colours (e.g., [Marchesini et al. 2009](#)). Therefore, the mass-completeness limit of quiescent galaxies will be higher than that of star-forming galaxies. We do not consider a mass complete sample in this work since we do not attempt to study the number density of the galaxies in our sample and because we want to extend the stellar mass–size relation to low mass galaxies. We place a conservative magnitude limit so that the size measurements are reliable even for galaxies with stellar masses that fall below our mass completeness limit. Nevertheless, there will, naturally, be some biases, particularly against low-surface-brightness (LSB) galaxies, which could possibly comprise a large fraction of the overall low-mass galaxy population (e.g., [Wright et al. 2020](#)). To test this effect, we select 120 LSB galaxies from the Abell1063 parallel field. Determining which galaxies are labelled as LSB systems is strongly dependent on the limits of the surveys that are used (e.g., [Disney 1976](#)). For this test, we define LSB galaxies as those with effective radii larger than 20 pixels and apparent magnitudes fainter than 25.5mag in the F125W band, as these lie

well below the surface-brightness of the majority of objects in the Abell1063 parallel field. For each of the 120 LSB galaxies, we then generate 10 mock galaxies with the same magnitude and size as the original, but with randomly selected position angles and axis ratios. The mock galaxies are randomly assigned a Sérsic index of $n = 1$ or $n = 4$. We then inject these mock LSB galaxies into the F125W image and we recover 1025 ($\sim 85\%$) of the injected objects with SExtractor. The majority of the LSB galaxies that are not detected happen to be overlapping with bright foreground objects. This shows that any bias against these ‘large’ LSB galaxies, which could systematically shift a stellar mass – size relation, should be negligible.

Another class of objects that our sample could be potentially biased against are compact galaxies. Although compact galaxies are easier to detect than large objects of similar brightness because their flux is more concentrated and therefore peaks well above the background, it can be difficult to distinguish compact galaxies from point sources. In the HFF-DeepSpace ([Shipley et al. 2018](#)) and 3D-HST ([Skelton et al. 2014](#)) catalogues, compact objects are classified as stars based on the tight correlation in size and magnitude that point sources follow. Both studies show that point sources can be cleanly separated from extended sources for $\text{mag}_{F160W} \leq 25$. Objects fainter than this are assigned a different flag (i.e., `star_flag = 2`) and are hence not removed by our quality cuts discussed in §3. Because of this conservative classification of stars, it is unlikely that compact galaxies are classified as point sources, and removed. We additionally test how often the modelling failed (i.e., `FLAG_GALFIT=1`) and how often compact objects ran into fitting constraints compared to their more extended counterparts. We find that the modelling of compact objects is no more likely to fail than it is for more extended objects, but compact objects are ~ 3 times more likely to run into fitting constraints. This is expected as the fitting constraints are specifically chosen to remove point sources and any galaxies for which the modelling is unphysical. These results suggest that compact galaxies are not removed from the sample, but faint compact point sources likely are by the cuts on the model parameters, giving us confidence that the results of this study will not be strongly impacted by biases against LSB or compact objects.

3.6 Selecting star-forming and quiescent galaxies

Quiescence can be estimated based on a variety of galaxy properties including colour, star formation activity, galaxy structure, and morphology. However, great care must be taken when distinguishing quiescent galaxies from star-forming ones since quiescence does not mean that there is *no* residual star formation. Quiescent galaxies can also appear blue in colour if they have only recently stopped forming stars and have disk-like structures ([Graham & Guzmán 2003](#)). All of these caveats mean that it is likely that different galaxies are identified as quiescent depending on which criteria are used. It is therefore crucial that star-forming and quiescent galaxies are carefully separated using a robust method.

Often, galaxies are identified as quiescent based on their position on the UVJ diagram (e.g., [Labbé et al. 2005](#); [Wuyts et al. 2007](#); [Williams et al. 2009](#)). This method is powerful because the U–V colour allows galaxies with red colours, which indicates old stellar populations, to be selected while the V–J colour is used to differentiate galaxies that have old stellar populations from dusty star-forming galaxies ([Whitaker et al. 2012](#)). Unfortunately, selecting quiescent galaxies from the UVJ diagram often misses galaxies that have recently ceased their star-formation (e.g., [Marsan et al. 2015](#)). For instance, post-starburst galaxies will still be relatively blue in colour despite having very little ongoing star-formation because – while they

¹ <https://github.com/gbrammer/ezay-photoz>

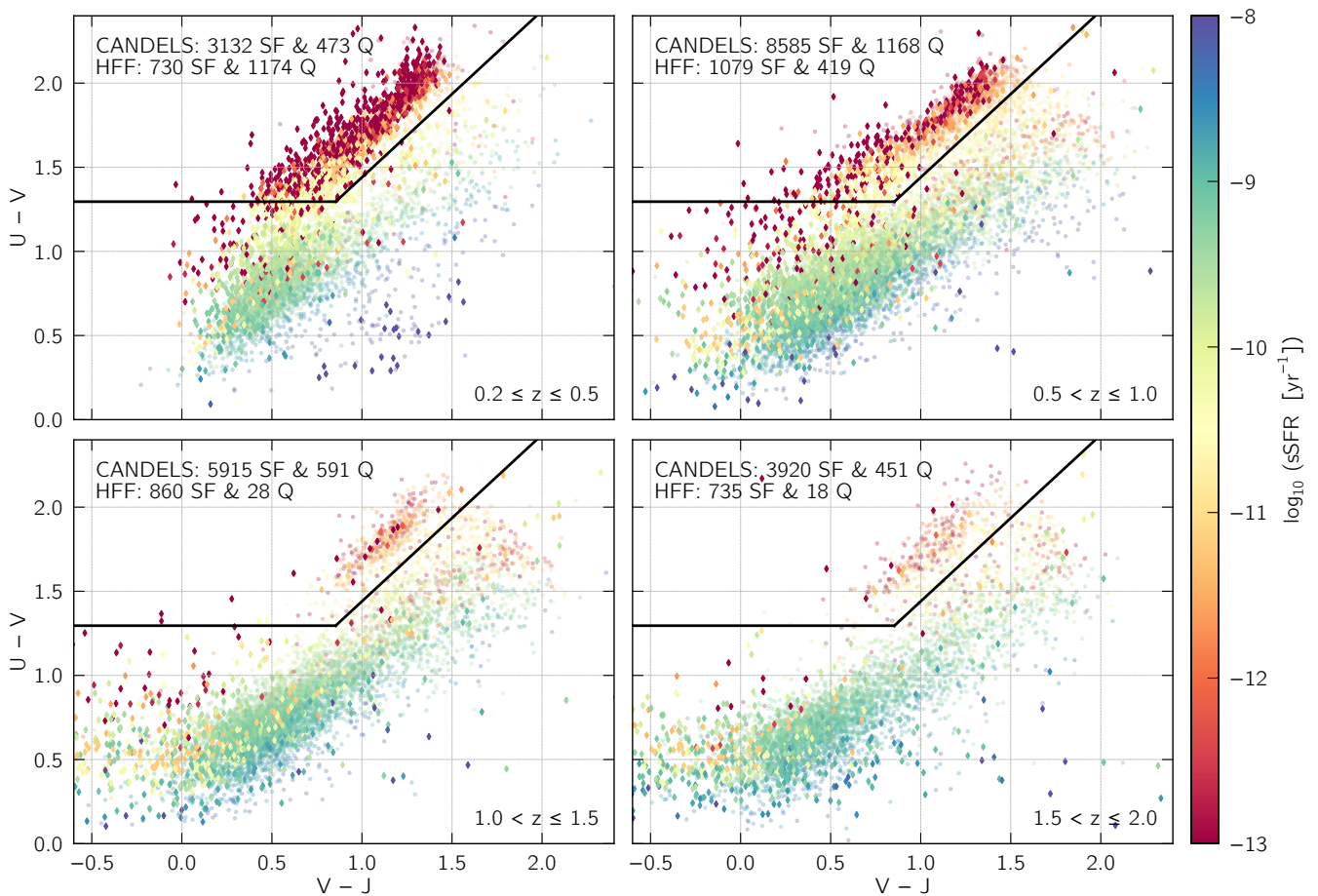


Figure 3. Same UVJ diagram shown in Figure 2, except the galaxies are colour-coded by their specific star-formation rates (sSFRs), estimated by the FAST code (Kriek et al. 2009). The redshift bin is indicated for each panel in the bottom right corner and the number of star-forming and quiescent objects from HFF and CANDELS, when galaxies are separated according to the UVJ diagram, is indicated in the top left. In general, galaxies with low measured sSFR lie within the quiescent region, as expected; however there is a small number of galaxies with low sSFR that fall outside the quiescent wedge. Upon checking, we find that most of these galaxies have uncertain sSFR measurements, although it is possible that some are post-starburst galaxies, that have recently quenched their star-formation and therefore still have bluer colours (see text).

are no longer forming stars – there are still short-lived stars within those galaxies. An added complication of using the UVJ diagram to select star-forming and quiescent galaxies is the lack of sufficiently dusty templates in EAZY (Brammer et al. 2008), resulting in the cut-off at the top right corner in each panel of Figure 2. Although properties in the HFF-DeepSpace catalogues are measured with the most up-to-date EAZY templates, the properties of the CANDELS galaxies are not corrected for this effect. Therefore, we do not apply a vertical cut in $V-J$, following, e.g., van der Wel et al. (2014) and Martis et al. (2016).

A popular alternative is selecting galaxies directly based on their star-formation activity. While this method gets around some of the challenges of selecting quiescent galaxies based on the UVJ diagram, it heavily relies on reliable SED fits, which have been shown to suffer when modelling unique galaxies. We show the UVJ diagram for our sample colour coded by the specific star-formation rate in Figure 3. There is a small fraction ($\sim 2\%$) of objects that have low specific star-formation rates but do not lie within the quiescent region on the UVJ diagram (i.e., galaxies that are colour coded red but lie outside the quiescent wedge). While some of these objects could be recently quenched galaxies that have not yet moved into the quiescent region,

we find that the majority of these objects have uncertain sSFRs. The median uncertainty on $\log(\text{sSFR})$ of all the data shown in Figure 3 is ~ 0.4 dex, while the median uncertainty of the galaxies that have low sSFR but fall outside the quiescent wedge is ~ 1.2 dex.

Quiescent galaxies and star-forming galaxies are also believed to have intrinsically different structure, making it possible to distinguish the two based on structural properties (e.g., Shen et al. 2003; Ravindranath et al. 2004). Quiescent galaxies are more generally thought to be spheroids with relatively concentrated light profiles, while star-forming galaxies are more disk-like. We therefore also show the stellar mass – size relation as a function of Sérsic index, where classical quiescent galaxies are described by a de Vaucouleurs (1948) profile (i.e., $n = 4$) and star-forming spiral galaxies are well described by an exponential light profile (i.e., $n = 1$). Of the three methods for selecting quiescence, this is the method most prone to misclassification because reliably deriving Sérsic indices, especially for non-local galaxies, is difficult, and is only a rough approximation for disk and bulge dominated objects. Another group of missed cases are quenched disk galaxies (e.g., McGrath et al. 2008; Salim et al. 2012; Carollo et al. 2016) since galaxies can quench while retaining their structure. We therefore argue that this method is the least

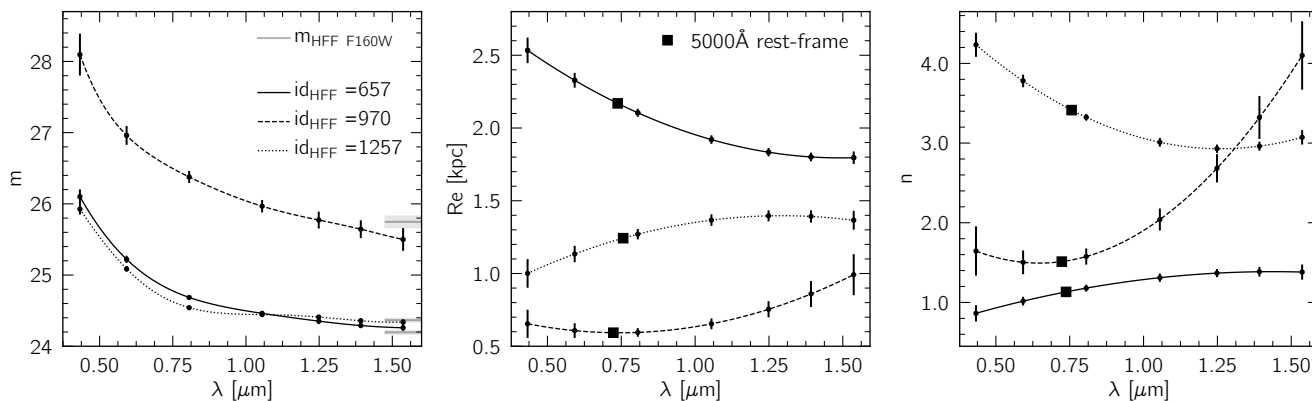


Figure 4. Example fitting results of the apparent magnitude (m), effective radius (Re), and Sérsic index (n) as a function of wavelength for three galaxies in the Abell1063 parallel field, all at roughly the same redshift. These three galaxies, are chosen to display the variety of functional forms that can be fit for different galaxy parameters. For each galaxy, the fit is shown in a different line style as indicated in the legend and the black squares indicate the 5000\AA rest-frame values that can be read from the Chebyshev polynomials. The left panel shows the apparent magnitude of each object, as derived by GALFITM, with the F160W magnitude from the HFF-DeepSpace catalogue (Shiple et al. 2018) indicated as a horizontal grey line around $1.6\mu\text{m}$, where the uncertainty is shown as a shaded grey region. Although these H-band magnitudes are measured with different methods, they are consistent, indicating that the modelling is reliable, even for faint objects, such as galaxy $id_{\text{HFF}} = 970$. In the modelling, the magnitudes are allowed to vary freely as a function of wavelength (hence, rest-frame magnitudes are not derived directly from the Chebyshev polynomial), while for the size and Sérsic index of each galaxy, which are shown in the middle and right panel, the Chebyshev polynomial fit is allowed to vary as a second order polynomial. All of the errors shown are the ones derived by GALFITM and are therefore underestimated by a factor of 2 to 2.5 (see §4). The magnitude errors are increased by a factor of 5, to illustrate that the faintest galaxy has the largest uncertainties.

reliable, but in order to understand how the selection of quiescence impacts the results of the stellar mass–size relation, we separate quiescent from star-forming galaxies based on all three criteria and compare them in Figure 6.

3.7 MEGAMORPH Galaxy Models

As discussed in §2, from the wavelength dependence of the parameters that we fit, we can identify many properties of each galaxy that might otherwise be missed with single band fitting. We illustrate the importance of modelling morphological parameters as a function of wavelength in Figure 4, where we show three example galaxies. These galaxies’ parameters show very different wavelength dependencies and have been specifically selected in order to showcase the wide variety of functional forms that we allow our fitting routine to recover. We also show the images, models, and residuals for these same galaxies in Figure 5. Although we did not consider the stellar masses nor the redshifts of the example galaxies, they are all low mass galaxies at $z \sim 0.5$. Galaxies $id_{\text{HFF}} = 657$, $id_{\text{HFF}} = 970$, and $id_{\text{HFF}} = 1257$ have corrected stellar masses of $10^{8.27} M_{\odot}$, $10^{7.83} M_{\odot}$, and $10^{8.21} M_{\odot}$, respectively. Hence, the quality of the models and residuals shown in Figure 5 are typical for low mass galaxies.

In the left panel of Figure 4, we test how apparent H-band magnitudes derived with the MegaMorph tools compare to those from the HFF-DeepSpace catalogue (Shiple et al. 2018). The GALFITM apparent magnitudes are expected to be different from those in the HFF-DeepSpace catalogue, since they are derived in different ways. The magnitudes in the HFF-DeepSpace catalogue are measured with the AUTO aperture photometry, in which the extent of a galaxy is defined and all of the flux within that area is summed. A small AUTO-to-total correction is then applied. As previously discussed in §3.5, GALFITM, on the other hand, first models galaxy profiles and then these profiles are integrated to infinity in order to obtain total-Sérsic magnitudes in each band. Despite these differences, it

can be seen that the H-band magnitudes are consistent for the three example galaxies in Figure 4, indicating that the modelling results are robust.

Given the reliability of the modelling, we now go on to investigate the properties of the three example galaxies that we can infer from the wavelength dependence of their parameters. Galaxy $id_{\text{HFF}} = 657$, shown with solid lines, is very similar to galaxy $id_{\text{HFF}} = 1257$ in terms of brightness. However, these galaxies’ sizes and Sérsic indices indicate that they are in fact very different. Perhaps the most striking property of galaxy $id_{\text{HFF}} = 657$ is its large effective radius. This galaxy has a size that is larger at short wavelengths than at longer wavelengths, indicating that this is most likely a multi-component, star-forming object, where the short wavelengths reflect the size of the disk component and the longer wavelengths reflect the size of the bulge. On the other hand, this galaxy’s Sérsic index is roughly equal to one at all wavelengths, which is indicative of a blue, disk-dominated system. The presence of a bulge is not well motivated from the Sérsic index data alone. As it is more difficult to measure reliable Sérsic indices than reliable sizes, we argue that this object is likely a two-component system. From Figure 5, this galaxy appears more extended at shorter wavelengths, which is consistent with Figure 4, where we show that the effective radius along the major axis is decreasing with wavelength. In fact, the visual impression supports that this is an edge-on disk system with a significant, large, and round, bulge component.

Galaxy $id_{\text{HFF}} = 1257$, shown with a dotted line, has parameters that are consistent with a quiescent, elliptical galaxy. This object has a large Sérsic index at all wavelengths, suggesting that this is an elliptical galaxy, which classically have deVaucouleurs profiles with $n = 4$ (e.g. Vulcani et al. 2014; Kennedy et al. 2015).

Finally, galaxy $id_{\text{HFF}} = 970$, shown as a dashed line in Figure 4 and in the central column of Figure 5, is the faintest and smallest galaxy of the three examples, and therefore has the largest measurement uncertainties. This object is slightly smaller at shorter wavelengths

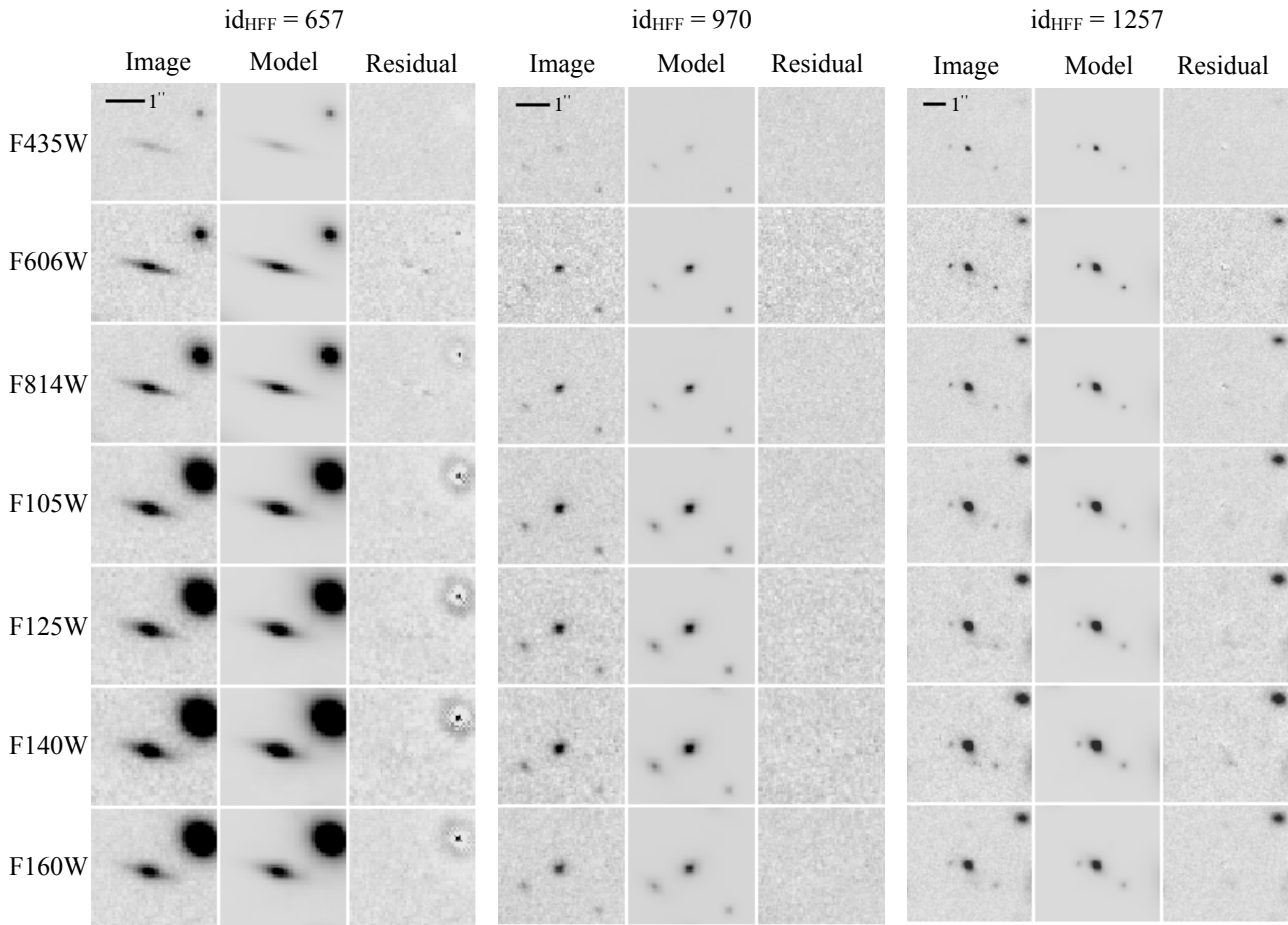


Figure 5. Fitting results for the three galaxies from the Abell1063 parallel field for which we show the parameters in Figure 4. The image, model, and residual are shown for each of the seven bands that we use for the HFF. All images are oriented such that up is North and left is East. For all examples shown here, the remaining flux in the residuals at each pixel around the primary object is less than 10% of the flux in the corresponding pixel of the image. This amount of residual flux is to be expected and is generally found when fitting smooth profiles to real galaxies.

than at longer wavelengths, which is what would be expected for a bulge-dominated galaxy. Like galaxy $\text{id}_{\text{HFF}} = 657$, this is most likely a two-component object based on the strong wavelength dependence of the Sérsic index shown in the right panel of Figure 4. The Sérsic index at the bluest wavelength is $n \approx 1.6$, while at longer wavelengths, the Sérsic index is large, again a result of the disk component being most prominent at shorter wavelength while the bulge component being more prominent at longer wavelengths.

All of these details can give us crucial information about the intrinsic properties of galaxies, which could be lost with single band fitting. In principle, in noise-less data, these details could be obtained by using single-band fits at different wavelengths and interpolating the values using the same polynomials; however, as data are not noiseless, this would only be feasible for the very brightest galaxies and not for the majority of objects we are after. Additionally, multi-wavelength fitting gives us the advantage of being able to obtain the 5000\AA rest-frame parameters from the Chebyshev polynomials directly, allowing for a fairer comparison across redshifts. We indicate the rest-frame 5000\AA sizes and Sérsic indices for the three example

galaxies shown in Figure 4 as black squares. We note that we do not derive rest-frame magnitudes in this way since we allow full freedom in order to recover the SEDs, which can lead to Runge’s phenomenon, in which the polynomial is unconstrained in between the fixed points where data is available, making estimated magnitude values unfeasible.

As can be seen from all three galaxy models in Figure 5, neighbouring objects are modelled along with the primary object. For object $\text{id}_{\text{HFF}} = 657$, there is a bright, neighbouring spheroidal galaxy in the upper right corner. The residuals of the neighbouring object are characteristic of a multi-component galaxy that has been modelled with a single Sérsic profile, exactly as we have done. Although the spheroidal neighbouring galaxy itself is not particularly well modelled in the centre, we are able to reliably recover the primary galaxy’s parameters, as can be seen from the residuals of the primary object. Another interesting feature of this figure is the bright object “next” to galaxy $\text{id}_{\text{HFF}} = 1257$, again in the upper right corner. This object is not modelled and can be clearly seen in the residual of every band. This object is too far away from object $\text{id}_{\text{HFF}} = 1257$ to contribute

to its flux, so it is masked out for the modelling. The residuals seen for these three galaxies are examples of normal residuals as these objects are not chosen to have excellent models nor particularly clean residuals.

Uncertainties in the magnitude, size, and Sérsic index measurements are produced directly by GALFITM. In Figure 4, we show the GALFITM-derived uncertainties for each parameter in each band. Unfortunately, these uncertainties have been shown to be significantly underestimated (Häussler et al. 2007). Since the 3D-HST (Skelton et al. 2014) and HFF-DeepSpace (Shiple et al. 2018) catalogues already provide reliable magnitudes and uncertainties for the objects in the CANDELS and HFF fields, respectively, we do not make an effort to obtain reliable uncertainties for the measured magnitudes but use those reported values instead. Additionally, because we do not use the uncertainties on the Sérsic index measurements for any of our results, we focus on deriving reliable uncertainties only for galaxy sizes. Häussler et al. (2007) simulated images with similar properties to the HST GEMS (Rix et al. 2004) dataset, which contained over 40,000 simulated galaxies of various magnitudes, sizes, position angles and axis ratios, and the light profiles of these galaxies were fit with GALFIT. Häussler et al. (2007) found that GALFIT substantially underestimates the true uncertainties of the fit and suggest that the uncertainties estimated by the GALFIT are not reflecting the Poisson noise of the images. They find that the GALFIT uncertainties are underestimated by a factor of ~ 10 or more. Since GALFITM uses the same assumptions to derive uncertainties as GALFIT, the errors estimated by both codes should be consistent. However, it is important to point out that the simulated galaxies in Häussler et al. (2007) were all fit and modelled in one band. Häußler et al. (2013) performed a similar analysis for multi-wavelength fitting in 9 bands for simulated galaxies with properties similar to GAMA (Galaxy And Mass Assembly; Driver et al. 2011) images. While they do not present an analysis of the error bars in that work, subsequent checks on this issue revealed that the uncertainty for the multi-wavelength fitting is only underestimated by a factor of $\sim 2 - 2.5$. Therefore, we increase all size errors estimated by GALFITM by a factor of 3 in order to be conservative. To obtain the uncertainties on the 5000\AA rest-frame size, we linearly interpolate the uncertainties on the two bands that enclose the rest-frame wavelength. Given our redshift range, the rest-frame 5000\AA always falls within our wavelength range (i.e., $< 1.6\mu\text{m}$); therefore, we do not extrapolate any rest-frame parameters or uncertainties.

4 STELLAR MASS – SIZE RELATION

The stellar mass – size relation is shown in Figure 6, where the sample is divided into four redshift bins in the same way as in Figures 2 and 3. Sizes smaller than the $\text{FWHM}_{F160W}/2$ at the maximum redshift of each bin are indicated in grey. Size measurements smaller than this limit are not as robust as larger sizes; however, further investigation of the simulations carried out in Häußler et al. (2013) reveals that galaxy sizes smaller than $\text{FWHM}/2$ can be reliably measured if the point spread function (PSF) of the image is well known, which for HST data is the case. In their simulations, the FWHM of the PSF was ~ 3.5 pixels, and galaxies were simulated down to effective radii of ~ 1 pixel, well below the $\text{FWHM}/2$ limit. No systematic effects on the measured image sizes were seen, implying that the measured sizes below this limit are not artificially shifted. Furthermore, in our sample, there are only 94 galaxies that have effective radii smaller than $\text{FWHM}_{F160W}/2$. Although we are able to measure the small sizes of these galaxies, they do not comprise a large portion of the total

sample and, therefore, would not heavily influence the stellar mass – size relation that we derive. We also note that the $\text{FWHM}_{F160W}/2$ limit is well below the galaxy size distribution of our objects. To significantly change the result of this work, a strong (factor of a few in size measurements) and systematic (all small objects would have to be fit larger) measurement error would have to be observed. No such effect was found in any galaxy regime, specifically not for small objects.

In Figure 6, we differentiate between star-forming and quiescent galaxies based on three criteria: the sSFR, UVJ diagram, and Sérsic index as discussed in §3.6. In Figure 6, the top four panels are colour coded according to the sSFR as derived from FAST. The middle four panels are colour coded according to each galaxy’s position on the UVJ diagram. Finally, the bottom four panels are colour coded by the Sérsic index following the idea that objects with high Sérsic index are predominantly passive ellipticals. As discussed in §3.6, none of these selection techniques are perfect, as they rely intrinsically on good photometric measurements, which becomes tricky especially for faint objects. Despite all of these difficulties, Figure 6 shows that there is generally good agreement between the three selection criteria. This is not necessarily expected, since the sSFRs and rest-frame colours are derived from SED models, while the Sérsic indices are obtained from modelling light profiles with GALFITM. As these quantities are independently derived, the level of consistency between these selection methods suggests that the light profile modelling *and* the SED modelling are reliable. Naturally, there are some differences between the selection methods. For instance, the flattening of the quiescent sample is less pronounced with the Sérsic cut, as there appear to be only a few low mass galaxies with high Sérsic indices; however, the flattening is visible if galaxies with Sérsic indices > 2.5 are considered to be quiescent, as is generally done in the literature (e.g., Bruce et al. 2014; Lange et al. 2015) such that flattening can be seen for the galaxies which are colour-coded as yellow points in the figure.

From Figure 6, the star-forming and quiescent sequences can be seen to occupy the same regions of the stellar mass – size plane for all three separation criteria even though individual galaxies may be classified differently by the different methods. It is also worth noting that there is a cloud of quiescent galaxies at the very low mass end (i.e. below $10^8 M_{\odot}$) that can be seen especially in the $0.2 \leq z \leq 0.5$ and $0.5 < z \leq 1.0$ redshift bins when the quiescent sample is selected based on the sSFR (top panels). This cloud of galaxies cannot be seen when we select galaxies based on the UVJ diagram or Sérsic index. Further investigation of these objects reveals that their SEDs are visually consistent with being young, blue, low-mass galaxies. However, due to the rather featureless SEDs of young, low-mass galaxies and the relatively large measurement errors in these faint objects, the derived sSFRs suffer from large uncertainties. In fact, many of these objects would be classified as star-forming if the lower 1σ sSFR were used instead of their best-fit sSFR. Combined with the uncertainties on the Sérsic indices previously discussed, we therefore choose to separate star-forming and quiescent galaxies based on their positions on the UVJ diagram for the subsequent sections of this paper because this appears to be the most robust method for our sample.

Above $\sim 10^{10} M_{\odot}$, which is the stellar mass regime that most other studies in the literature have focused on, the quiescent sequence on the stellar mass – size plane can be easily distinguished from the star-forming sequence at all redshifts. In other words, the quiescent galaxies follow a distinct trend on the mass-size plane that is different from the one that the blue, star-forming galaxies follow. This distinction can be made regardless of whether star-forming and quiescent galaxies are separated based on sSFR, position on the UVJ diagram,

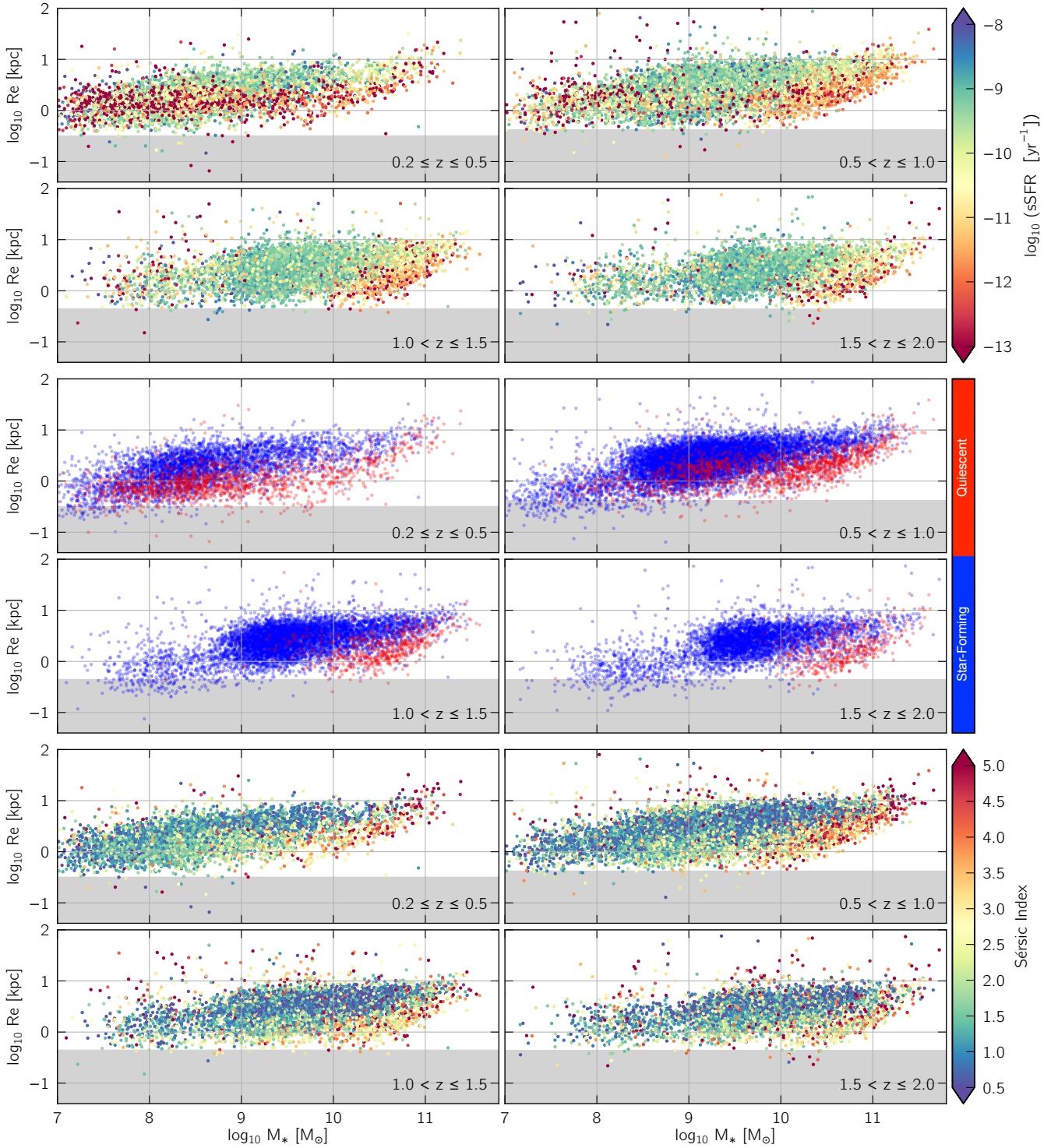


Figure 6. The stellar mass – size relations for different redshift bins for quiescent and star-forming galaxies. The sizes reported are rest-frame 5000Å sizes as derived in §3.4. The top four panels are colour-coded by specific star-formation rate (sSFR). The middle four panels are colour-coded based on each galaxy’s position on the UVJ diagram, with the quiescent galaxies shown on top of the star-forming galaxies to highlight the flattening seen at $z \leq 1$. Finally, the bottom four panels are colour-coded by Sérsic index. The colour-coding shows that there is good agreement between the three selection criteria that we test for separating star-forming and quiescent galaxies. The grey areas in each panel indicate sizes at $R_e < \text{FWHM}_{F160W}/2$ at the maximum redshift in each panel, to indicate potentially difficult sizes to measure (see text).

or Sérsic index. In this high mass regime, the quiescent galaxy sequence shows a steeper slope than the star-forming population across all redshifts, consistent with previous works (e.g., van der Wel et al. 2014; Dimauro et al. 2019; Mowla et al. 2019). Consequently, at a given stellar mass, high mass quiescent galaxies tend to be smaller in size than their star-forming counterparts, again consistent with the literature. These trends however, do not hold true for less massive objects.

Below $\sim 10^{10} M_{\odot}$, the quiescent and star-forming galaxy sequences are not quite as clearly distinguishable from each other. This could partly be an effect due to the fact that it is more difficult to separate less massive galaxies into star-forming and quiescent because it is generally harder to measure the properties of less massive objects. In spite of this, we recover similar trends for the star-forming and quiescent sequence across all selection methods, suggesting that this behaviour is real. The stellar mass – size relation of low mass quiescent galaxies appears flat and then steepens for high mass quiescent galaxies, while the star-forming galaxies continue to grow in size as they grow in mass from 10^7 to $10^{11.5} M_{\odot}$. We will quantitatively discuss these relations in the following sections. The flattening of the stellar mass – size relation for quiescent galaxies has been shown in previous works (e.g., Cappellari et al. 2013; Berg et al. 2014; Norris et al. 2014; Lange et al. 2015; Whitaker et al. 2017); however, most of these are low redshift studies or are limited to relatively high stellar mass galaxies. In this work, we extend the stellar mass – size relation to include low stellar mass galaxies while also quantitatively exploring its flattening as a function of redshift. Unfortunately, at redshifts $z > 1$, our number of quiescent galaxies with stellar masses below $10^9 M_{\odot}$ is limited. This is, in fact, why we use the multi-band tools discussed in §3.7, as they allow us to push our analysis to fainter, less massive objects than was possible in previous studies. Nevertheless, some evidence of a flattening remains at these higher redshifts.

4.1 Quiescent Mass – Size Relation

We first analyse the quiescent galaxy sample shown in Figure 6, selected using the UVJ diagram. In §4.1.1, we discuss the high mass end (i.e., $\geq 10^{10.3} M_{\odot}$) only, for easier comparison with the literature. In §4.1.2, we discuss the behaviour of the quiescent sequence over the entire stellar mass range analysed in this work (i.e., $\geq 10^7 M_{\odot}$). Finally, in §4.1.3, we show the redshift evolution of the best-fit trends and discuss what this implies for how quiescent galaxies build up their mass and grow over cosmic time.

4.1.1 The High Mass End

We begin by fitting the high mass end of the stellar mass – size relation for quiescent galaxies as this mass regime has been extensively studied in the literature (e.g., Maltby et al. 2010; van der Wel et al. 2014; Kuchner et al. 2017; Dimauro et al. 2019). These previous works have shown that the stellar mass – size relation is well represented by a single power law at the high mass end for both quiescent and star-forming galaxies. We therefore fit a power-law function of the form:

$$R_e = A \left(\frac{M_*}{5 \times 10^{10} M_{\odot}} \right)^B \quad (2)$$

following van der Wel et al. (2014), Mowla et al. (2019), and Dimauro et al. (2019) in order to make a direct comparison. In Equation 2, R_e is the 5000\AA rest-frame half-light radius along the major axis in kpc and M_* is the stellar mass in M_{\odot} and corrected as described in §3.5; A and B are best-fit parameters that describe the trend. As this

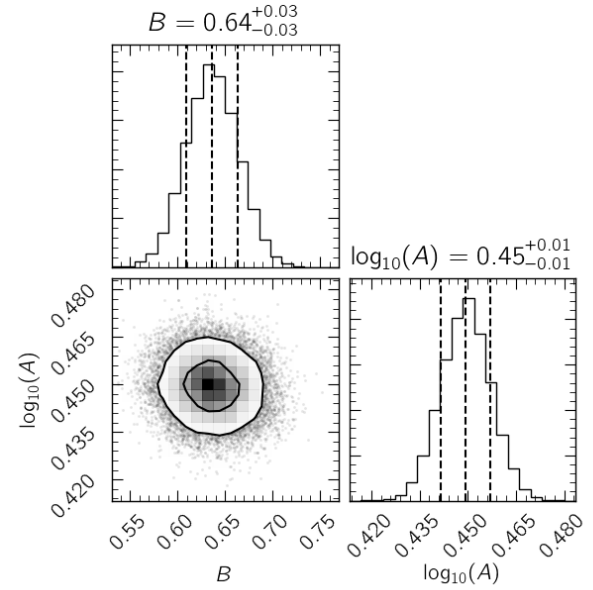


Figure 7. Corner plot for the estimated parameters for high mass quiescent galaxies in the redshift bin $0.5 < z \leq 1.0$ according to Equation 2. We choose to show the corner plot for this redshift bin because we have the largest sample size in this range and therefore, smaller uncertainties; however, the parameters exhibit similar behaviour in the other redshift bins as well. For the 1D posterior distributions, the vertical dashed lines mark the 16th, 50th, and 84th percentiles, where the 16th and 84th percentiles of each distribution are indicated as the lower and upper uncertainties of each parameter estimation, respectively. These values are reported for the quiescent fits in Table 2. For the 2D posterior distribution, the contours mark the 1σ and 2σ levels.

relation is linear in the log-log space in which we fit the data, we refer to B as the slope and $\log_{10}(A)$ as the intercept at $5 \times 10^{10} M_{\odot}$.

The best-fit parameters, A and B , of our model are determined using a Bayesian inference with a Markov Chain Monte Carlo (MCMC) approach (Foreman-Mackey et al. 2013). We assume uniform priors for both parameters and each galaxy’s contribution to the fit is weighted by its uncertainty in mass and size, which are described in §3.5 and §3.7, respectively, by taking the covariance of the two uncertainties. The parameters are fit in the $\log_{10} M_* - \log_{10} R_e$ space, as in the other works that we compare our results to. Motivated by the parameters obtained in previous studies, we allow $\log_{10}(A)$ and B to vary over $[-0.5, 2]$ and $[0, 2]$, respectively. We set 50 random walkers, and perform 10,000 MCMC iterations, which allow the parameters to converge on the best-fit value for all four fits, one for each redshift bin. The resulting corner plot (Foreman-Mackey 2016) for the fitting routine in the $0.5 < z \leq 1.0$ range is shown in Figure 7.

In Figure 8, we show the stellar mass – size relations across four redshift bins for the high mass quiescent galaxy population, where the number of objects that are fit in each redshift bin are indicated in the bottom left corner. Our best-fit models are shown in magenta where the shaded region signifies the 1σ level within which 68% of all models fall. Across the entire redshift range that we explore, we find generally good agreement with the stellar mass – size relations from van der Wel et al. (2014), Mowla et al. (2019), and Dimauro et al. (2019), which are also shown in Figure 8. Finally, we also compare our results in the lowest redshift bin to Lange et al. (2015), who fit the stellar mass – size relation with a different functional form, which we discuss further in §4.1.2.

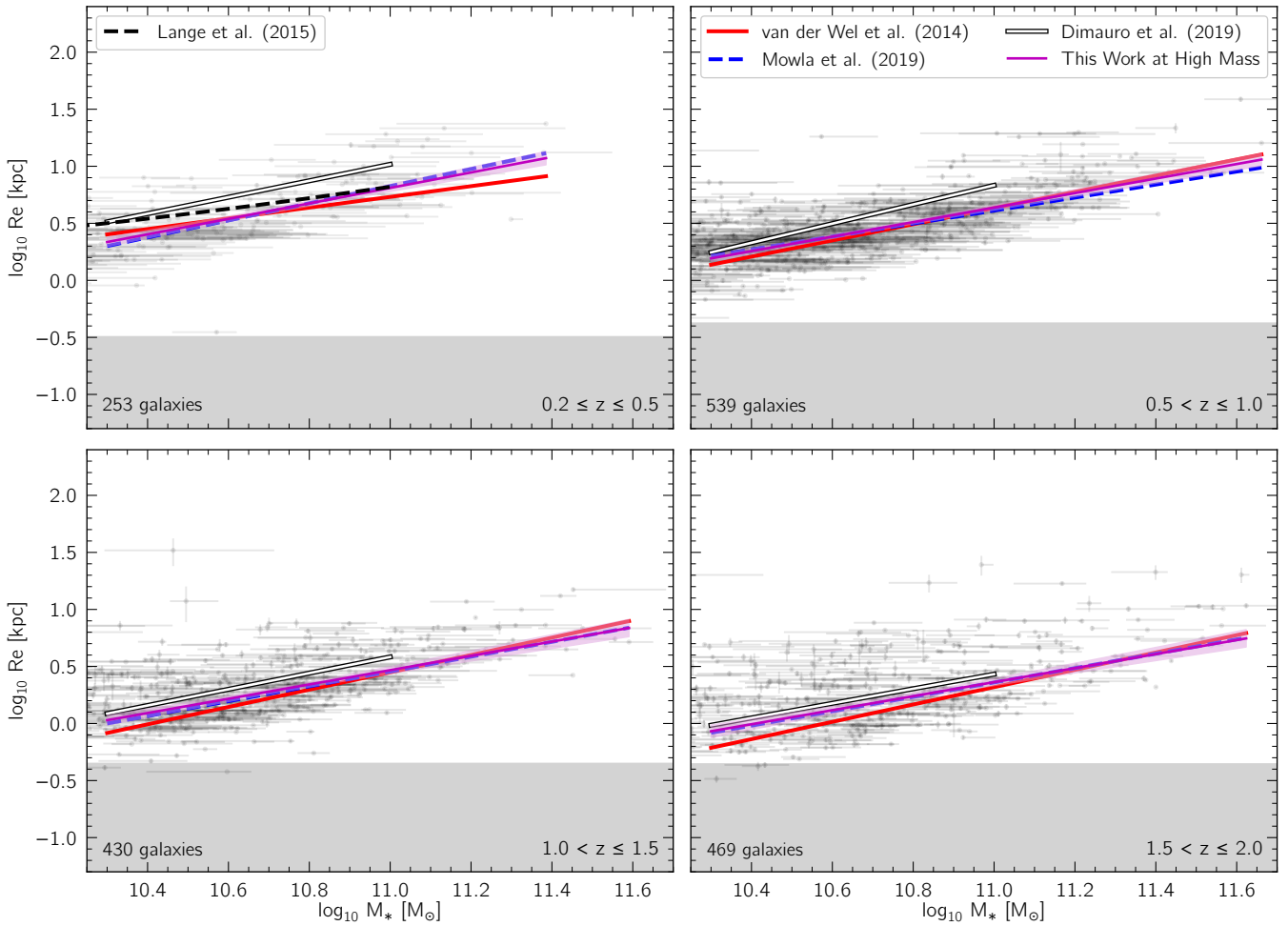


Figure 8. The stellar mass – size relations for high mass quiescent galaxies. The best-fit power law is shown in magenta and the 1σ level that includes 68% of all models is indicated by the shaded region. In order to make a fair comparison to [Dimauro et al. \(2019\)](#) (white line), [van der Wel et al. \(2014\)](#) (red line) and [Mowla et al. \(2019\)](#) (blue dashed line), we use the same lower stellar mass limit of $10^{10.3} M_{\odot}$. We additionally compare to [Lange et al. \(2015\)](#) (black dashed line in the lowest redshift bin only), who fit a different functional form to their sample, which we discuss in detail in §4.1.2, and study local galaxies with $0.01 < z < 0.1$. As in Figure 6, the grey areas indicate sizes at $R_e < \text{FWHM}_{\text{F160W}}/2$ at the maximum redshift of each panel.

4.1.2 Entire Mass Range

For the full stellar mass range that we are exploring, the stellar mass – size relation for quiescent galaxies flattens, as can be seen from Figure 6. In order to capture this flattening at the low mass end, we fit the quiescent galaxy mass – size relation with a double power-law function motivated by [Shen et al. \(2003\)](#) and [Lange et al. \(2015\)](#):

$$R_e = \gamma(M_*)^{\alpha} \left(1 + \frac{M_*}{10^{\delta}}\right)^{\beta - \alpha} \quad (3)$$

where R_e is the 5000\AA rest-frame half-light radius in kpc and M_* is the corrected stellar mass in M_{\odot} , as in Equation 2; α and β describe the slope at the low and high mass end, respectively; γ is the normalisation, or in other words, the effective radius at a stellar mass of $10^0 M_{\odot}$. Finally, 10^{δ} is the stellar mass at which the second derivative of the function is at a maximum. Since it is difficult to assign an intuitive, physical meaning to this parameter, we simply refer to it as δ . This parameter can be considered to be the distinction between high and low mass galaxies (e.g., [Shen et al. 2003](#); [Lange et al. 2015](#)); however, we find that this value does not align well with the visual transition from one slope to the other. For example, in the

highest redshift bin, the most probable 10^{δ} value is $\sim 10^{11.27} M_{\odot}$, which would imply calling even very massive galaxies ‘low mass’, ultimately making such a cut a poor choice. For quiescent galaxies, we choose to make the distinction between high mass and low mass galaxies at $10^{10.3} M_{\odot}$ since this is the stellar mass range we use in §4.1.1 following [van der Wel et al. \(2014\)](#), [Mowla et al. \(2019\)](#), and [Dimauro et al. \(2019\)](#).

The double power-law function, shown in Equation 3, is fit to the quiescent galaxy samples using an MCMC approach in a way that is similar to how the single power law is fit to the high mass quiescent galaxies. Namely, we assume uniform priors for all parameters and the galaxies are weighted by their uncertainties in mass and size. All parameters are fit in log space, and because of this we choose to fit $\log_{10}(\gamma)$ as opposed to γ . Although the same posterior probability distributions can be recovered for the parameters by allowing them to vary over a variety of ranges as long as they are large enough, we choose to let the parameters vary over the following ranges:

- The low mass slope: $\alpha \in [-2, 2]$
- The high mass slope: $\beta \in [0, 10]$
- The normalisation, i.e., $\log_{10} \text{Re at } 1 M_{\odot}$: $\log_{10}(\gamma) \in [-10, 10]$

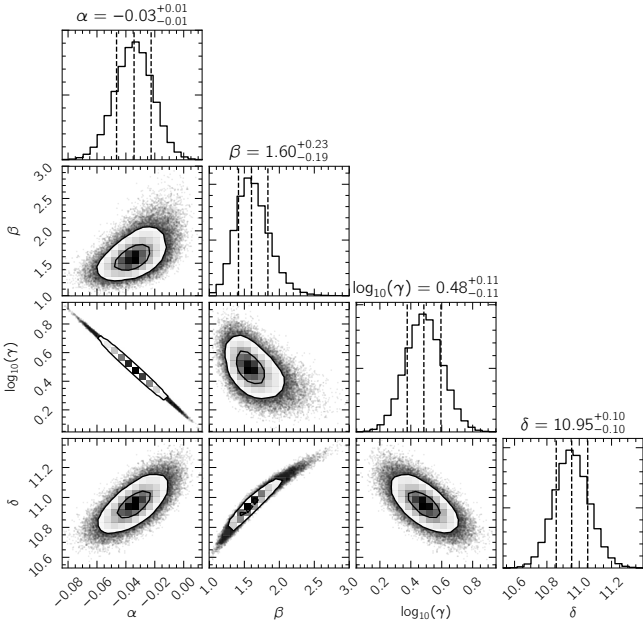


Figure 9. Corner plot for the four parameter MCMC fitting routine for the $0.5 < z \leq 1.0$ redshift bin. The vertical dashed lines in the 1D posterior distributions indicate the 16th, 50th, and 84th percentiles, where the 16th and 84th percentiles of each distribution are indicated as the lower and upper uncertainties of each parameter estimation, respectively. These results are reported for the entire redshift range used in this work in Table 2. For the 2D posterior distributions, the contours mark the 1σ and 2σ levels.

- \log_{10} of the stellar mass at which the second derivative of the function is at a maximum: $\delta \in [9, 13]$

The parameter spaces of the low mass and high mass slopes, as well as the normalisation, are derived empirically. The parameter space of 10^δ is constrained from 10^9 to 10^{13} , since this is the stellar mass range within which we expect the low mass slope to transition to the high mass slope.

For the four parameter fitting, we set 100 random walkers, and perform 30,000 MCMC iterations. The resulting corner plot (Foreman-Mackey 2016) is shown in Figure 9. For the double power law fit (Eq. 3), we see a strong anti-correlation between α – the low mass slope – and $\log_{10}(\gamma)$ – the normalisation. This indicates that if the low mass slope is a large negative value, then the resulting model would have a large size at $10^0 M_\odot$. This behaviour can be seen in Table 2, where the low mass slope is most negative for the $1.0 < z \leq 1.5$, so in turn, we see the largest normalisation in this redshift range as well. If the low mass slope is zero, then the normalisation would be the size ($\log_{10} \text{Re}$) at 10^δ . Additionally, β (the high mass slope) and δ (\log_{10} of the stellar mass at which the second derivative is at a maximum) are correlated. This behaviour explains why the four parameter model always has a steeper high mass slope than the two parameter model, as can be seen in Figure 10. If we set $10^\delta M_\odot$ as the lower stellar mass limit, instead of using $10^{10.3} M_\odot$, the resulting two parameter model has a slope that is consistent with the high mass slope that we recover with the four parameter model.

We show the stellar mass – size relation for quiescent galaxies over the full stellar mass range explored in this work in Figure 10. The best-fit model is shown in blue with the 1σ level, which includes 68% of all possible models, in light blue. We additionally show the high mass model with its corresponding 1σ level from Figure 8 in ma-

genta. Generally, we find that the ‘high mass’ and ‘all mass’ models are in good agreement since they overlap and display similar behaviour above $10^{10.3} M_\odot$. This indicates that the two independently derived models are consistent, lending confidence to our result. The best-fit model from Lange et al. (2015) is also shown in the first panel of Figure 10 as a dashed black line. Although this model has somewhat different characteristics from the model that we obtain, it is important to note that Lange et al. (2015) use GAMA data (Driver et al. 2011), so that their stellar mass – size relations are for $z \sim 0$ galaxies. Furthermore, the Lange et al. (2015) stellar mass – size relation shown here is derived in the g-band, where star-forming and quiescent galaxies are separated with a $u - r$ colour cut. Hence, it is not surprising that there is some difference between their results and ours. Nonetheless, the overall shape of the two curves is generally consistent.

From the quiescent galaxy data, it appears that a model with a positive slope at the low mass end could be better representative of the data. Given the small number of low mass quiescent galaxies at $z > 1$, the MCMC models are primarily driven by intermediate mass galaxies. A negative low mass slope would imply that quiescent galaxies with stellar masses below $\sim 10^{10} M_\odot$ are more extended than their higher mass counterparts. This result is difficult to explain physically, so we test how the model behaves if it is restricted to have a positive low mass slope. We find that the general shape of the curve is similar and the behaviour at the high mass end remains largely unchanged when the low mass slope is forced to be positive. Given that we can reproduce the general behaviour of the model with and without constraining the low mass slope, we opt to show the model in which the parameters are unconstrained, but refer explicitly to the mass range in which this model is valid and has been tested in our data. Interestingly, Genel et al. (2018) also predict an increasingly negative low mass slope with redshift when reproducing the stellar mass – size relation with Illustris.

4.1.3 Quiescent Galaxy Evolution

After galaxies quench, they continue to undergo subsequent evolution by merging with other galaxies and/or minor episodes of star-formation activity. Whereas major mergers, involving galaxies of similar mass, will lead to comparable growth in both size and mass, minor mergers can leave the overall galaxy profile shape largely unchanged, while leading to substantial size evolution. These results have been shown by a number of studies (e.g., Buitrago et al. 2008; Bezanson et al. 2009) and are also well-supported by simulations (e.g., Naab et al. 2009; Oser et al. 2012). We find that the average size evolution which galaxies experience once they quench depends strongly on their stellar mass, such that above stellar masses of $10^{10.3} M_\odot$, quenched galaxies experience a steep evolution on the mass – size plane with slopes of ~ 0.7 at $z \leq 2$. van Dokkum et al. (2015) argued that the evolution of quiescent galaxies with stellar masses above $10^{10} M_\odot$ is primarily driven by dissipationless, dry mergers. Newman et al. (2012) showed that this mechanism requires a high rate of occurrence of minor mergers to account for the observed size growth of quiescent galaxies. While multiple dry mergers can significantly increase the size of a galaxy, they do not significantly impact the total stellar mass (e.g., Bezanson et al. 2009; Hopkins et al. 2010; van Dokkum et al. 2010). This causes the quiescent sequence to follow a steep slope on the mass–size plane and our results are hence consistent with this picture at the high mass end. The evolution of the best-fit parameters with redshift in this high mass regime is shown in Figure 11. The slope, shown in the top panel, does not show a strong evolution, while the intercept, $\log_{10}(A)$, shows significant

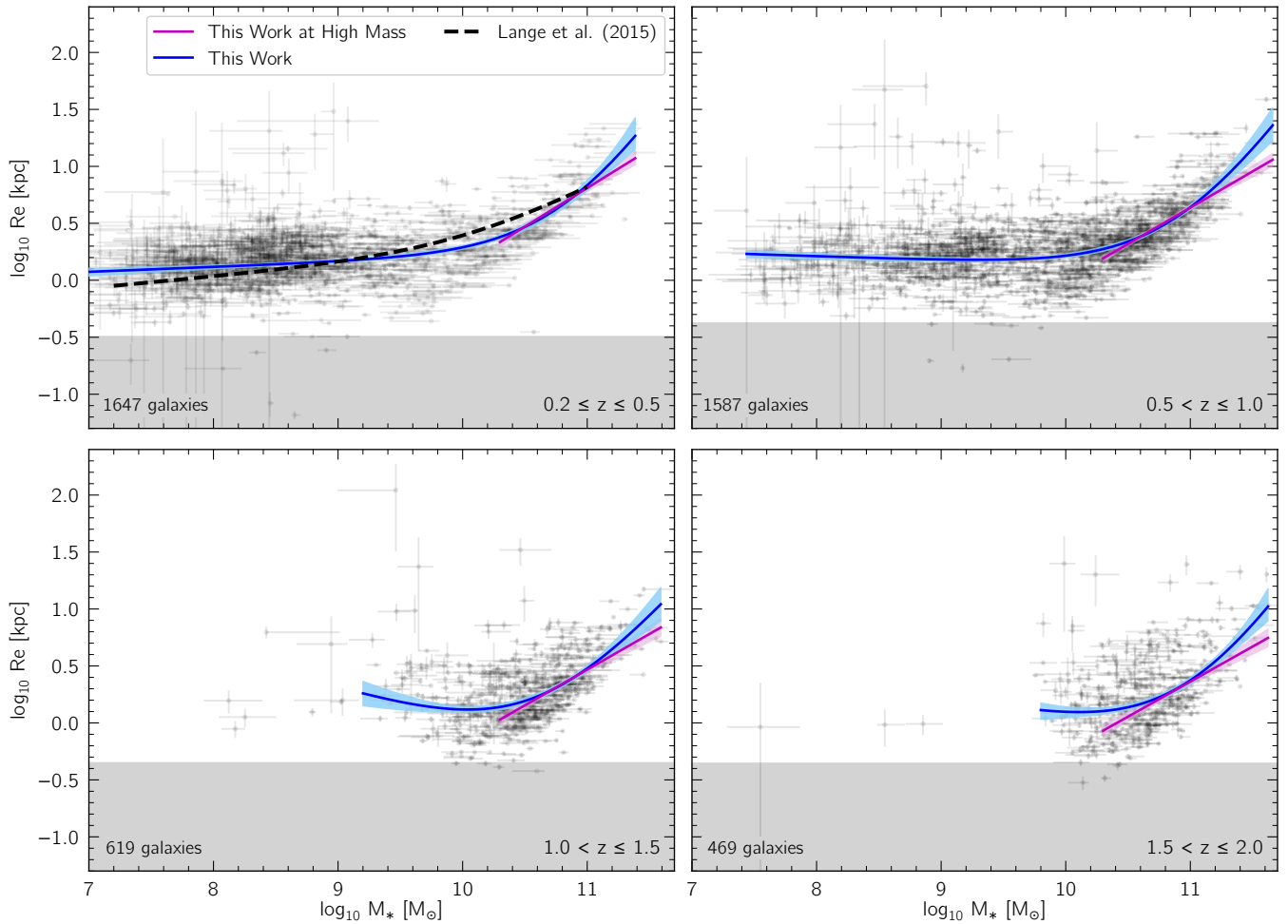


Figure 10. The stellar mass – size relation of quiescent galaxies, where the single power-law models for the high mass end are indicated in magenta and the double power-law fits are shown in blue. The number of galaxies fit with the double power-law fit are indicated in the bottom left corner of each redshift bin. As in previous figures, the grey areas indicate sizes at $R_e < \text{FWHM}_{\text{F160W}}/2$ at the maximum redshift of each panel.

evolution with redshift, representing the upwards shift of the relation over time. These results are consistent with many other studies and imply that high mass quiescent galaxies were more compact in the early Universe than they are today.

Quenched low mass galaxies, however, exhibit an almost flat relation on the stellar mass – size plane. Because of this, the quiescent galaxy sequence has to be fit with a double power law function, which is shown in Equation 3. We show the redshift evolution of the double power law fit and parameters in Figure 12. In the top panel, all models are shown, where the shading of the 1σ region indicates the redshift according to the legends. A clear evolution can be seen for the stellar mass – size relation of quiescent galaxies. In panel (a), we show the evolution of the slope at the low mass end, α , and in panel (b), we show the evolution of the normalisation, $\log_{10} \gamma$, which is anticorrelated to α . In panel (c), we show the evolution of the slope at the high mass end, β . The slope at the high mass end remains relatively constant with increasing redshift, which is consistent with the behaviour of the slope when only high mass quiescent galaxies are fit with a single power law (top panel of Figure 11). Finally, in panel (d), we show the evolution of δ . Since this parameter is correlated to the slope at the high mass end (β), we do not see a strong redshift evolution for δ .

4.2 Star-forming Mass – Size Relation

The star-forming galaxy relation on the mass–size plane is analysed in a similar fashion to the quiescent relation, such that in §4.2.1, we discuss the high mass end (i.e., $\geq 10^{9.5} M_{\odot}$) only and in §4.2.2, we discuss the trends exhibited over the entire mass range available to us. In §4.2.3, we show and discuss the redshift evolution of star-forming galaxies on the stellar mass–size plane.

4.2.1 The High Mass End

In an effort to derive consistent and fair comparisons to previous studies, we first fit high mass star-forming galaxies only. We model all star-forming galaxies with stellar masses above $10^{9.5} M_{\odot}$ to cover the same mass range as van der Wel et al. (2014) and Dimauro et al. (2019). The stellar mass – size relation of these high mass star-forming galaxies is shown in Figure 13. The best-fit power-law, using Equation 2, is shown as a magenta line with a shaded magenta region indicating the 1σ level that includes 68% of all MCMC models. These models are obtained in the same way as the high mass quiescent best-fit parameters.

At the high mass end, our results are consistent with van der

Table 2. The estimated parameters for quiescent galaxies according to the single power law shown in Eq. 2 for the high mass galaxies and the double power law shown in Eq. 3 for quiescent galaxies over $10^7 - 10^{11.5} M_{\odot}$. The high mass parameters correspond to the magenta lines in Figure 10, and the all mass parameters correspond to the blue curves. As the ‘All Mass’ parameters have posterior probabilities that are not symmetric around the median, we report the 16th and 84th percentiles as lower and upper limits. The corner plots for the $0.5 < z \leq 1.0$ redshift range can be seen in Figures 7 and 9, for the high mass and all mass parameters, respectively. We additionally provide the root-mean-square error (RMSE) for each model. In the right-most column, we report the stellar mass below which we do not place confidence in our model due to a lack of data. For the first two redshift bins, this value is the stellar mass of the lowest-mass galaxy in that bin. At higher redshift, this is visually set to a value below which we do not have a significant number of data points.

z	High Mass			α	All Mass			RMSE	Lower Limit on $\log_{10} M_{*} [M_{\odot}]$
	$\log_{10}(A)$	B	RMSE		β	$\log_{10}(\gamma)$	δ		
$0.2 \leq z \leq 0.5$	0.61 ± 0.01	0.68 ± 0.04	0.23	$0.04^{+0.01}_{-0.01}$	$1.82^{+0.44}_{-0.39}$	$-0.24^{+0.09}_{-0.09}$	$10.94^{+0.19}_{-0.17}$	0.23	7.00
$0.5 < z \leq 1.0$	0.45 ± 0.01	0.64 ± 0.03	0.24	$-0.03^{+0.01}_{-0.01}$	$1.60^{+0.23}_{-0.19}$	$0.48^{+0.11}_{-0.11}$	$10.95^{+0.10}_{-0.10}$	0.23	7.44
$1.0 < z \leq 1.5$	0.28 ± 0.01	0.63 ± 0.04	0.22	$-0.33^{+0.06}_{-0.06}$	$1.25^{+0.17}_{-0.14}$	$3.24^{+0.39}_{-0.38}$	$10.63^{+0.12}_{-0.12}$	0.25	9.20
$1.5 < z \leq 2.0$	0.18 ± 0.01	0.61 ± 0.05	0.28	$-0.17^{+0.11}_{-0.12}$	$1.84^{+0.40}_{-0.26}$	$1.76^{+0.81}_{-0.74}$	$11.09^{+0.21}_{-0.18}$	0.29	9.80

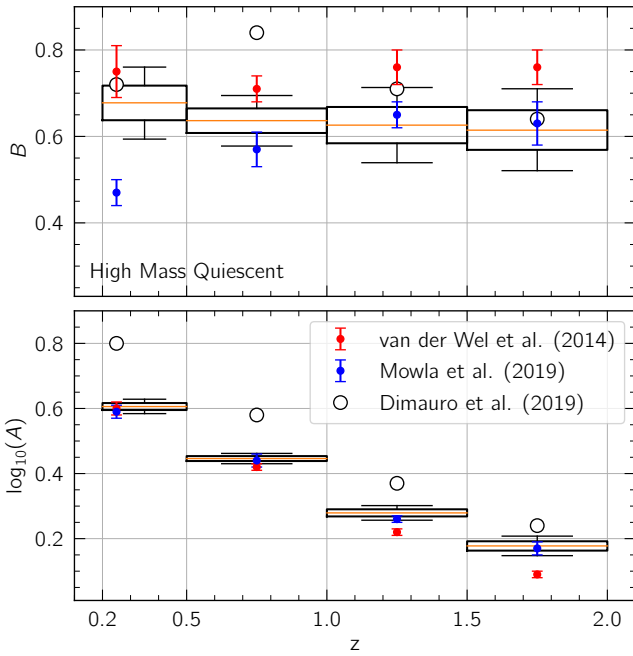


Figure 11. The evolution of the parameters B and $\log_{10}(A)$ with redshift from the single power-law fit shown in Equation 2 for high mass quiescent galaxies (i.e., $M_{*} \geq 10^{10.3} M_{\odot}$). For the posterior probability distribution in each redshift bin, we indicate the median with an orange line, the 1σ confidence level as a box, and the 2σ confidence level an error bar. The values shown here are reported in Table 3. We additionally compare these results to other high mass studies, as indicated in the legend. We find that the slope of the best-fit relation for high mass quiescent galaxies remains constant with redshift, consistent with previous works. There is a clear decrease in the intercept – $\log_{10}(A)$ – with redshift. $\log_{10}(A)$ describes the size at a fixed stellar mass of $5 \times 10^{10} M_{\odot}$ and its evolution with redshift indicates that quiescent galaxies at higher redshift were more compact, as expected.

Wel et al. (2014), shown as a red line in Figure 13. We also find very good agreement with Mowla et al. (2019), who extended the CANDELS data presented by van der Wel et al. (2014) to include very massive galaxies with stellar masses above $2 \times 10^{11} M_{\odot}$ from the COSMOS-DASH program (Momcheva et al. 2017). This is perhaps not particularly surprising as we already find consistent results with van der Wel et al. (2014) at the high mass end. However, our results

for high mass star-forming galaxies differ significantly from Dimauro et al. (2019), especially at high redshift. At $z \geq 1$, Dimauro et al. (2019) find a very shallow slope of $\beta \approx 0$ whereas we find that the slope remains roughly equal to 0.2 across the entire redshift range considered in this work (see also Fig. 15). This difference is surprising given that Dimauro et al. (2019) use the same method for deriving rest-frame sizes from CANDELS data (i.e., they use the software, GALFITM and GALAPAGOS-2, and obtain the rest-frame 5000\AA size from the Chebyshev polynomial). Likewise, they also derive their stellar masses using FAST (Kriek et al. 2009) and divide their star-forming and quiescent populations using the UVJ diagram.

In order to understand this difference, we compare the sizes measured in the F160W band for galaxies that are included in this study and in Dimauro et al. (2019). We choose to compare the measured F160W sizes as opposed to the rest-frame size in order to avoid any differences caused by different redshift estimates. We find excellent size agreement with no systematic offsets, suggesting that the MegaMorph tools are returning consistent results and are therefore not the root of this discrepancy. Dimauro et al. (2019) use redshift estimates from Dahlen et al. (2013) whereas we use redshift from the 3D-HST catalogues (Skelton et al. 2014; Momcheva et al. 2016), which have the added benefit of including redshifts derived from grism spectra. While there is scatter between the redshifts that we use and those used by Dimauro et al. (2019), we find that the redshifts are generally consistent, again, with no systematic differences.

It is important to also note that we take a simpler approach to modelling the trend exhibited by the star-forming galaxies compared to van der Wel et al. (2014) and Dimauro et al. (2019). These previous studies use a model that takes into account possible misclassifications of the star-forming and quiescent galaxies using the UVJ diagram. Although this is an important issue to consider, we recover similar results as van der Wel et al. (2014) using our simpler modelling method. This indicates that fitting the sequence with an MCMC approach, with fewer free parameters, can perform just as well and might be preferred due to its simplicity. As the galaxy sample that Dimauro et al. (2019) use is publicly available (Dimauro et al. 2018), we fit their star-forming sample with our MCMC approach. In doing this, we also obtain a slope that becomes shallower with redshift, reaching $\beta = 0.06$ at $1.5 < z < 2.0$. Given that we recover a similarly shallow slope at high redshift, we attribute the different results to a different sample selection rather than the modelling used. We discuss these differences in more detail in Appendix B.

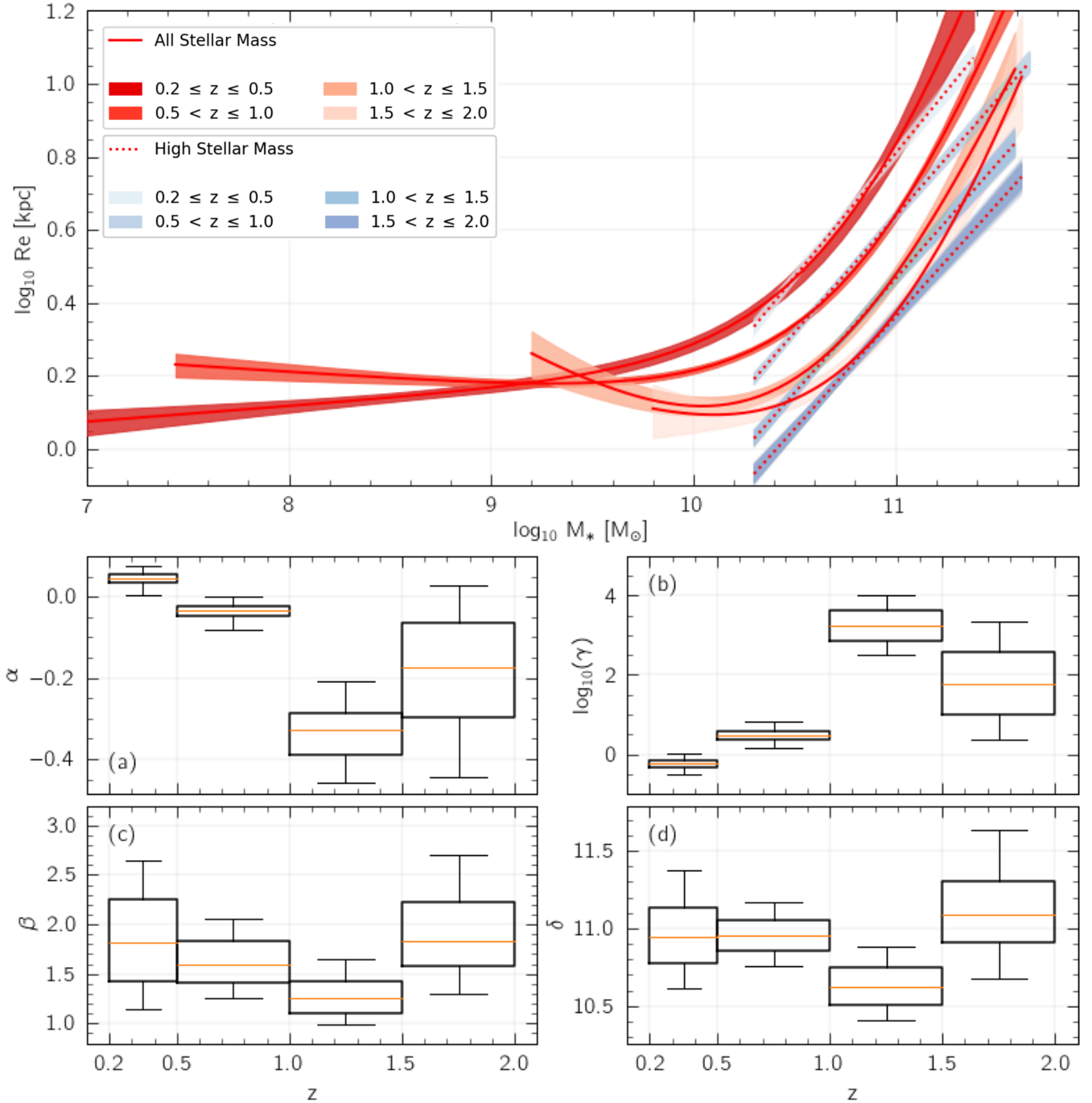


Figure 12. The evolution of the stellar mass – size relation for the quiescent sample. In the top panel, we show the quiescent best-fit trends from each of our four redshift ranges, where the concentration of the colour for the 1σ level indicates the redshift bin, as shown in the legend. The best-fit model for each redshift bin is shown as a solid red line while the high mass best-fit models are shown as dashed lines. The redshift evolution of α , β , $\log_{10} \gamma$, and δ are shown in panels (a), (b), (c), and (d), respectively. The median, 1σ and 2σ levels for each of these parameters are shown as an orange line, box, and error bar, respectively. As in Figure 9, α and $\log_{10} \gamma$ show anti-correlation while β and δ are correlated. Unsurprisingly, the uncertainties for the parameters that describe the low mass end (i.e., α and $\log_{10} \gamma$) are large in the highest two redshift bins, in which there are few low mass quiescent galaxies.

4.2.2 Entire Mass Range

Star-forming galaxies appear to follow a single power law over the full stellar mass range explored in this work. We therefore fit the star-forming galaxy population in all four redshift bins according to Equation 2, in addition to fitting it with a double power-law function,

shown in Equation 3, as we did for the quiescent sample. The single power law best-fit models are shown in Figure 14 as a blue lines, with the 1σ limits shown in light blue. We additionally show the star-forming high mass fits and the corresponding 1σ limits from Figure 13 in magenta. As can be seen from Figure 14, the blue

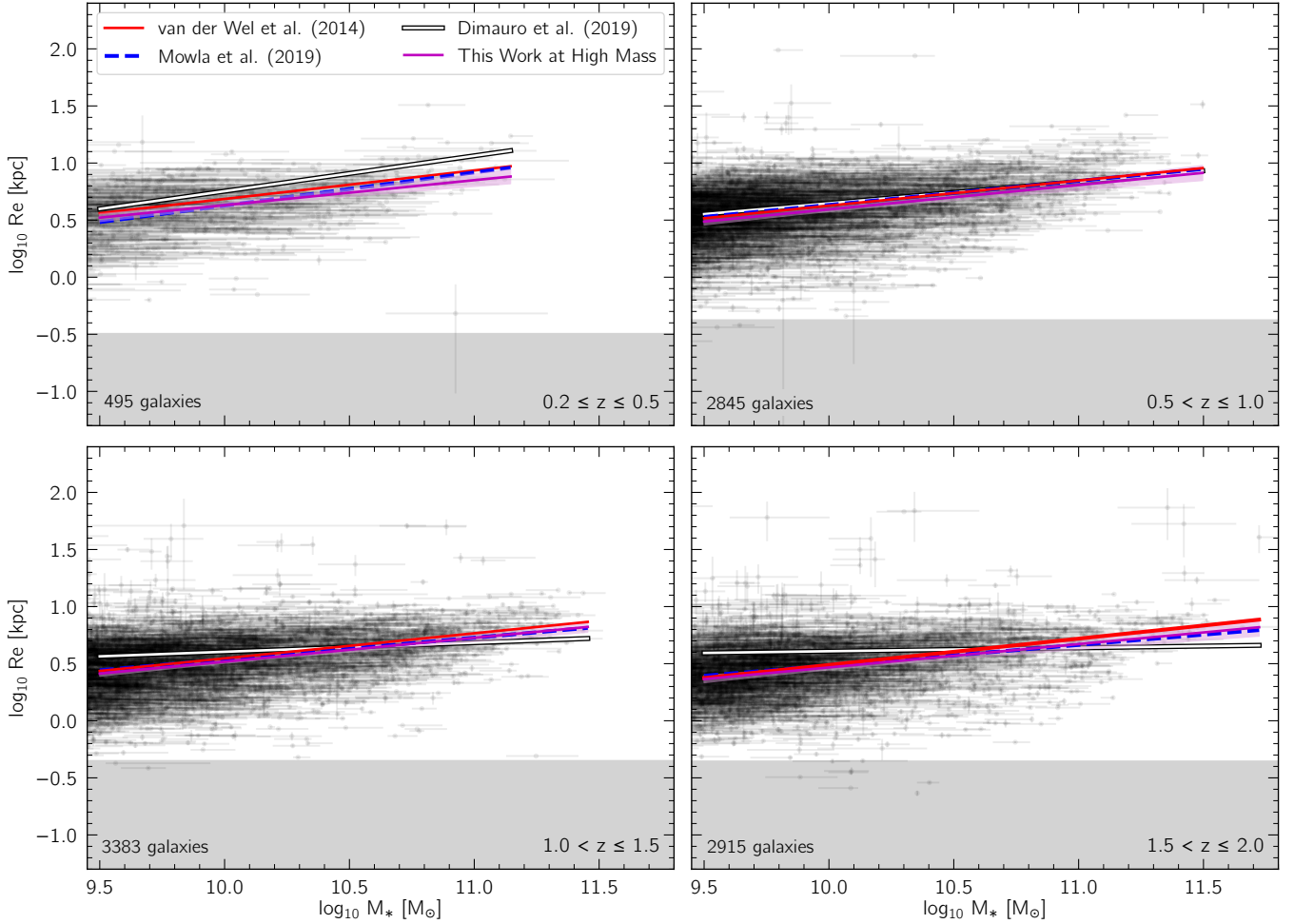


Figure 13. Same as Figure 8, but for the star-forming galaxy sample. The best-fit line to all star-forming galaxies with stellar masses above $10^{9.5} M_{\odot}$ is shown in magenta. Relations from van der Wel et al. (2014), Mowla et al. (2019), and Dimauro et al. (2019) are shown as indicated for comparison. As in previous figures, the grey areas indicate sizes at $R_e < \text{FWHM}_{F160W}/2$ at the maximum redshift of each panel. See the main text for discussion and comparison to other works.

and magenta fits either overlap entirely or are consistent within 1σ , indicating that the high mass fit is consistent with the fit over the entire stellar mass range studied in this work, across $0.2 \leq z \leq 2.0$. This also shows that the fit over our entire mass range well represents the high mass end, giving us confidence in such simple result over such a large mass range.

We fit the star-forming sample in the same way as the quiescent sample, with one notable exception – we do not constrain the parameter δ . By not constraining it, δ can occur at stellar masses much greater than the stellar mass range that is available to us, ultimately, allowing the MCMC model to essentially fit a single power law relation, even though a double power law functional form is assumed. The resulting double power law fits are shown as red dashed lines in Figure 14. In all redshift bins, the double power law is consistent with the single power law fit, highlighting that a double law fit to the star-forming population is unnecessary.

4.2.3 Star-Forming Galaxy Evolution

While they are star-forming, galaxies are constantly building up their stellar mass via new star formation. It is now well understood that

galaxies grow in size over cosmic time; however, certain physical processes, such as gas inflows to the centre of the galaxy can also cause galaxies to become more compact (Zolotov et al. 2015; Tacchella et al. 2017). This interplay between structure and star formation can be studied and better understood by constraining the stellar mass – size relation.

The average galaxy growth with cosmic time is visible in the redshift evolution of the stellar mass – size relation above $\sim 10^9 M_{\odot}$ shown in the top panel of Figure 15, where the zero point is decreasing with increasing redshift (see panel d in Fig. 15). Interestingly, this trend appears not to hold true for low mass galaxies, as shown in the top panel of Figure 15. In fact, galaxies with $\sim 10^7 - 10^8 M_{\odot}$ appear to remain in the same region of the stellar mass – size plane from $0.2 \leq z \leq 2$, suggesting that low mass galaxies are not growing much in size with cosmic time, unlike their more massive counterparts.

Our best-fit models as well as the evolution of our best-fit parameters and their uncertainties estimated by our MCMC analysis are shown in panels a-to-d in Figure 15. The top panel shows the best-fit curve for each redshift bin from Figure 14 shown in different shades of blue as indicated by the legend. Similarly, we show the high mass models (where we limit the star-forming sample to $M_{*} \geq 10^{9.5} M_{\odot}$)

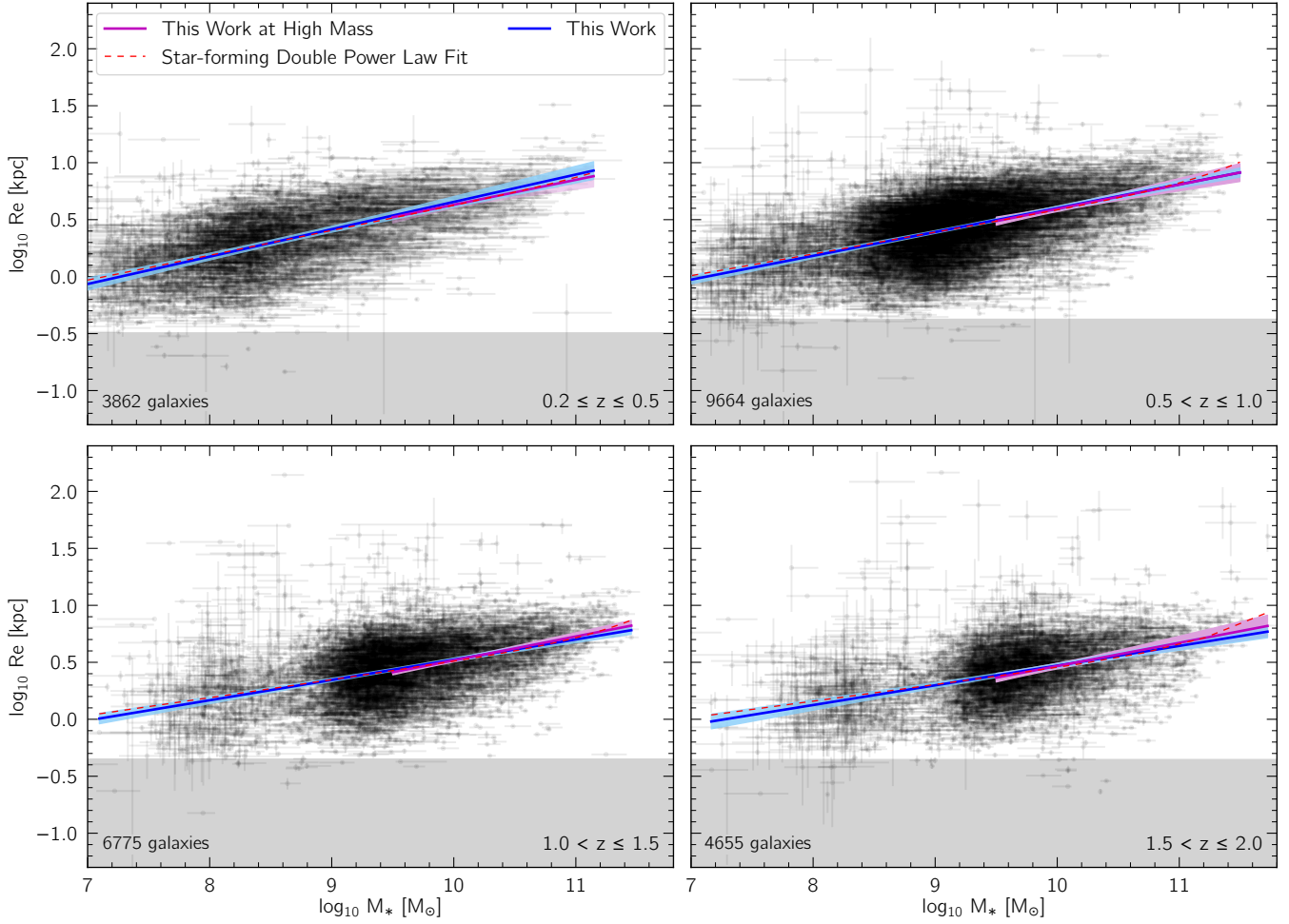


Figure 14. Same as Figure 10, but for the star-forming galaxy sample. The magenta and blue lines with the corresponding 1σ levels show the models for the high mass star-forming galaxies (i.e., $M_* \geq 10^{9.5} M_\odot$) and the entire stellar mass range (i.e., $M_* \geq 10^7 M_\odot$), respectively. The red dashed line is the best-fit double power law model, which is derived in a similar way to the double power law model for the quiescent galaxies (see text for details). The star-forming double power law fit is in excellent agreement with the all-mass fit at the low mass end and with the high mass fit at the high mass end. The blue and magenta lines from the 4-parameter and linear models are also in good agreement – they fall within 1σ of each other at all redshifts. Therefore, the double power law model is over-fitting the data and is unnecessary, but shown here as a means to indicate that a single power law fit is a good representation of the data. As in previous figures, the grey areas indicate sizes at $R_e < \text{FWHM}_{F160W}/2$ at the maximum redshift of each panel.

Table 3. The estimated parameters for star-forming galaxies according to the single power law shown in Eq. 2 for each redshift bin. We fit the high mass end only (i.e., $M_* \geq 10^{9.5} M_\odot$ for the star-forming galaxies) in order to provide direct comparisons to previous studies. The high mass parameters correspond to the magenta lines in Figure 14. The last three columns show the best-fit model parameters and the RMSE for star-forming galaxies over our entire mass range. We find good agreement between our high mass best-fit parameters and the best-fit parameters derived for our entire mass range, indicating that the models are reliable. In Figure 15, we show the evolution of these parameters with redshift.

z	High Mass			All Mass		
	$\log_{10}(A)$	B	RMSE	$\log_{10}(A)$	B	RMSE
$0.2 \leq z \leq 0.5$	0.78 ± 0.03	0.22 ± 0.03	0.21	0.82 ± 0.02	0.24 ± 0.01	0.25
$0.5 < z \leq 1.0$	0.74 ± 0.02	0.21 ± 0.02	0.23	0.75 ± 0.01	0.21 ± 0.01	0.25
$1.0 < z \leq 1.5$	0.66 ± 0.01	0.21 ± 0.01	0.23	0.65 ± 0.01	0.18 ± 0.01	0.25
$1.5 < z \leq 2.0$	0.61 ± 0.01	0.20 ± 0.02	0.24	0.59 ± 0.01	0.17 ± 0.01	0.26

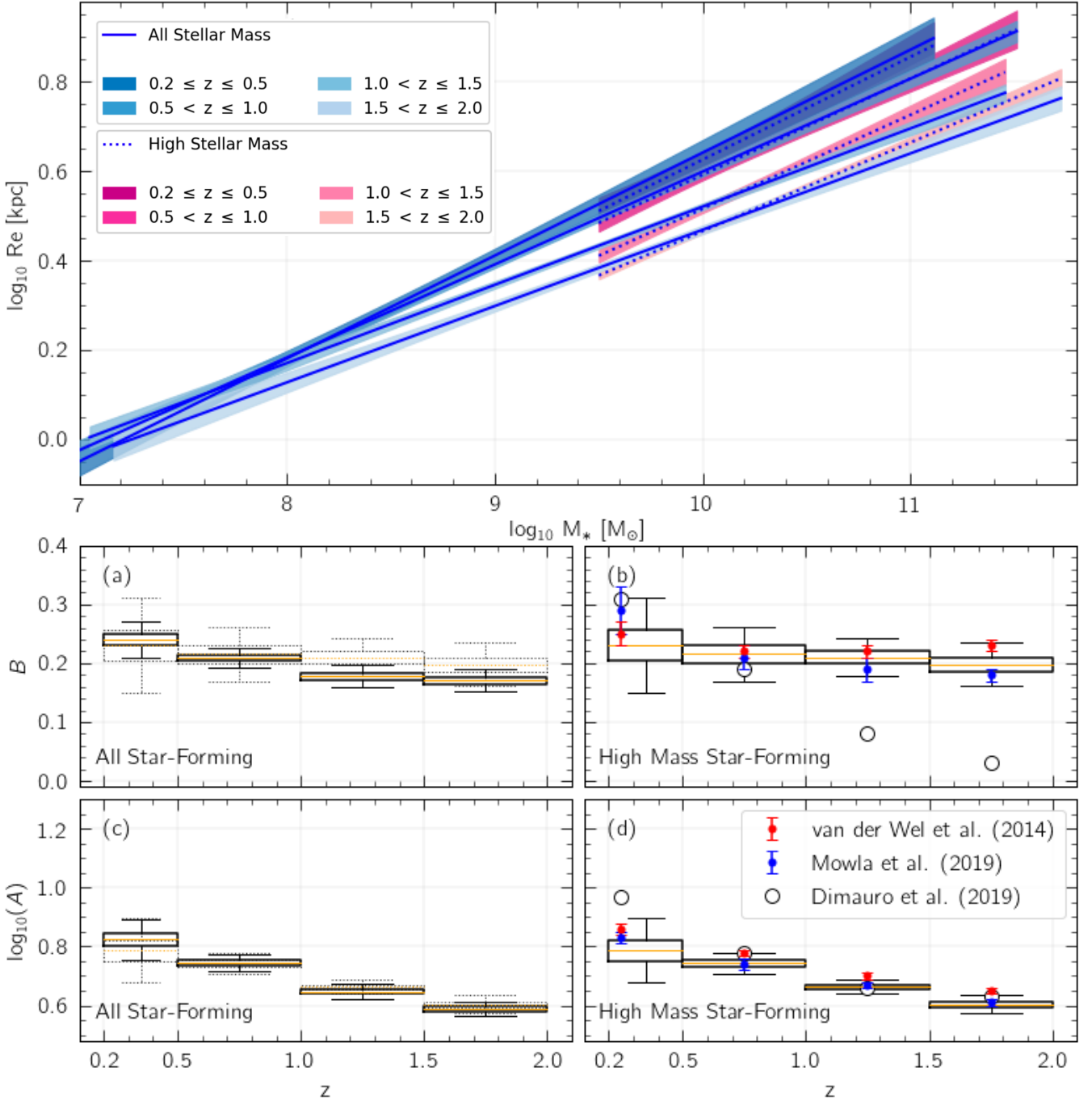


Figure 15. The evolution of the stellar mass – size relation and the best-fit parameters for the star-forming sample. The top panel shows the best-fit model for each redshift bin as a solid blue line while the high mass best-fit models are shown as dashed lines. The 1σ levels for each model are shown as blue and magenta regions, where the concentration of the colour indicates the redshift bin. The bottom four panels display the redshift dependence of the parameters B , which is the slope of the fit and $\log_{10}(A)$, which is the intercept. Panel (a) shows the evolution of B when the entire mass range is fit, and panel (b) shows the evolution when only the high mass galaxies are modelled. Likewise, panels (c) and (d) show the evolution of $\log_{10}(A)$ over the whole mass range and for the high mass range, respectively. The high mass range contains galaxies with stellar masses $10^{9.5} M_{\odot} \leq M_* \leq 10^{11.5} M_{\odot}$ in order to provide a direct comparison to van der Wel et al. (2014) and Dimauro et al. (2019). At each redshift, we show the median, the 1σ confidence level, and the 2σ confidence levels as an orange line, box, and error bars, respectively. Our results are consistent with van der Wel et al. (2014) (in red) for both the slope and normalisation, while Dimauro et al. (2019) (in white) find a much steeper evolution in the slope. Our results are also in very good agreement with those from Mowla et al. (2019) (in blue).

as dotted lines with magenta 1σ regions. The high mass fit and the fit over all mass are in excellent agreement, where the dotted line falls within 1σ region of the solid blue line. In panels (a) and (c), we show the slope and intercept as a function of redshift, respectively, when the entire mass range is used. In panels (b) and (d), we show the redshift evolution of the high mass best-fit parameters using the same lower mass limit as [van der Wel et al. \(2014\)](#), [Dimauro et al. \(2019\)](#), and [Mowla et al. \(2019\)](#). We note that the measurements for our lowest redshift bin span $0.2 \leq z \leq 0.5$, while measurements from previous studies include redshifts down to $z \sim 0$. This difference in the redshift ranges is indicated by the offset mean redshift of the studies we compare to from the mean of our lowest redshift bin at $z = 0.35$ shown in panels (b) and (d). We also show the result from panel (b) and (d) as more transparent box plots in panels (a) and (c), respectively, in an effort to compare the high mass and all mass fits. The slope, β becomes slightly steeper when the entire mass range is used, while the intercept, $\log_{10} \alpha$, is consistent for both fits.

Unsurprisingly, the normalisation, or intercept (i.e., the $\log_{10} (A)$ parameter) shows a strong evolution with redshift indicating that at fixed mass, galaxies have undergone significant size growth since $z \sim 2$. This evolution is present when we include our entire mass range as well as when we model only the high mass star-forming galaxies, as shown in panels (c) and (d). The slope, B , also shows some indication of evolution with redshift, although the evolution is not as strong and in fact appears to be absent when only the high mass star-forming sample is considered.

Interestingly, from panel (b) in Figure 15, we see that [van der Wel et al. \(2014\)](#) and [Mowla et al. \(2019\)](#) derive very different best-fit parameters for the slope of model from [Dimauro et al. \(2019\)](#). While [Dimauro et al. \(2019\)](#) find a rather steep evolution in the slope, [van der Wel et al. \(2014\)](#) and [Mowla et al. \(2019\)](#) find none. The results from our work are more consistent with [van der Wel et al. \(2014\)](#) and [Mowla et al. \(2019\)](#), as can be seen from both similarity in the power laws in Figure 14 and from the proximity of the best-fit parameters in Figure 15. As mentioned before, we discuss the difference between our own results and the results by [Dimauro et al. \(2019\)](#) in Appendix B in detail.

5 DISCUSSION

We present in this paper detailed measurements of the stellar mass–size relation of star-forming and quiescent galaxies over $0.2 \leq z \leq 2$. In this section, we provide a discussion based on the results from the previous section, covering topics from possible selection effects in §5.1 to providing possible interpretations of the results in §5.2 and comparing these findings to dwarf galaxy studies in §5.3.

5.1 Selection Effects

Before we discuss the implications of our results and their connection to low redshift studies, we first consider possible selection effects. In this work, we have combined galaxies from the cluster and parallel fields imaged as part of the HFF program with galaxies from the CANDELS fields to measure the stellar mass – size relation and its redshift evolution. As briefly noted in §3.1 and §3.2, the HFF and CANDELS data have different depths. The HFF data is deeper and therefore allows us to probe the low mass end of the stellar mass – size relation. As a result, the HFF and CANDELS data have different stellar mass distributions, with the HFF comprising the majority of low mass objects, and CANDELS providing the majority of intermediate- and high-mass objects since it samples a larger

area. Despite this, we find that the size distributions are consistent with each other and that both the HFF and CANDELS samples are consistent with the best-fit trends derived in §4. Hence, although we have combined data with different depths, we do not expect this difference to affect the stellar mass – size relations that we measure.

We have also used a combination of photometric and spectroscopic redshifts. By design, spectroscopic redshift estimates are much more reliable than photometric redshift estimates, although the typical photometric uncertainties for both HFF and CANDELS are small. [Shipley et al. \(2018\)](#) report a scatter of $\sim 0.02\text{--}0.05(1+z)$ for the HFF-DeepSpace catalogues depending on the specific HFF pointing and [Bezanson et al. \(2016\)](#) report an average scatter $\sim 0.0197 \pm 0.0003(1+z)$ for the 3D-HST catalogues. In our final sample, we have spectroscopic redshift estimates for 3441 galaxies from CANDELS and 606 galaxies from the HFF. As a result of the redshift cuts that we impose on the HFF cluster fields, discussed in §3.1, our HFF sample is devoid of galaxies with spectroscopic redshift estimates at $z \geq 1$. We find that using only objects with spectroscopic redshifts does not quantitatively change the results of this study, but using the larger photometric sample yields better statistical significance, specifically at high redshift, where spectroscopic redshift estimates are very rare.

Although we exclude any galaxies that are possibly lensed (see §3.1), [Yang et al. \(2021\)](#) recently presented the stellar mass – size relation of strongly lensed, high mass galaxies in the HFF cluster fields. After applying a lens reconstruction technique, they find that the strongly lensed galaxies lie on relations that are consistent with those derived by [van der Wel et al. \(2014\)](#). As we also find results that agree well with [van der Wel et al. \(2014\)](#), our results are also consistent with [Yang et al. \(2021\)](#) even though we have removed the galaxies that they study from our sample.

Finally, we note that while we have included galaxies from both dense and less-dense environments, only $\sim 10\%$ of the entire sample presented here comes from the HFF cluster fields. While galaxy evolution depends on the environment such that dense, cluster environments are believed to accelerate galaxy evolution (e.g., [Shankar et al. 2013](#)), the majority of galaxies presented in this work are not evolving due to cluster-specific processes, because only a small fraction of the total sample comes from dense environments. In fact, the effect of environment on the observed stellar mass – size relation is still debated. Several works have shown that galaxy sizes and masses exhibit weak, if any, environmental dependence (e.g., [Weinmann et al. 2009](#); [Maltby et al. 2010](#); [Huertas-Company et al. 2013](#); [Shankar et al. 2014](#); [Sweet et al. 2017](#)). On the other hand, [Kuchner et al. \(2017\)](#) report larger sizes for early-type galaxies and smaller sizes for massive late-type galaxies in clusters in comparison to the field at $z \sim 0.44$, consistent with higher redshift studies (e.g. [Papovich et al. 2012](#); [Delaye et al. 2014](#)). Given the low percentage of cluster galaxies that we use in this study, and that the effect of the environment is subtle, we do not expect the different environments to significantly influence the stellar mass – size relations that we recover in this work.

5.2 Interpretation of the star-forming and quiescent stellar mass–size relations

We show the stellar mass – size relation for all galaxies in our sample in Figure 6 as well as for quiescent galaxies only in Figure 10 and star-forming galaxies in Figure 14. Our results confirm previous findings that the star-forming and quiescent galaxy populations lie on distinct regions on the stellar mass–size plane (e.g., [van der Wel et al. 2014](#); [van Dokkum et al. 2015](#); [Morishita et al. 2017](#); [Dimauro](#)

et al. 2019). While the mass – size relation of star-forming galaxies is well represented by a single power-law across $10^7 M_{\odot}$ to $10^{11.5} M_{\odot}$, quiescent galaxies show a clear flattening at the low mass end. By taking the median size of galaxies with stellar masses below $10^{9.5} M_{\odot}$, we clearly observe that the flattening occurs at around 1 kpc out to $z \sim 1$, with some indication that this flattening exists out to $z = 2$ although the number of low mass quenched galaxies is limited above $z \sim 1$.

In the local Universe, it has been known that a flattening occurs in the stellar mass – size relation of quiescent galaxies for some time (e.g., Shen et al. 2003; Misgeld & Hilker 2011; Lange et al. 2015; Eigenthaler et al. 2018). At higher redshifts, Huang et al. (2017) have shown that if the half-light radius scales linearly with the dark-matter radius, as is expected from galaxy formation models, then the stellar mass – size relation of quiescent galaxies should be nonlinear, with a bend at a few $\times 10^{10} M_{\odot}$ out to at least $z \sim 3$. In contrast, for star-forming galaxies, they obtain a linear relation on the size–mass plane. Our results for both star-forming and quiescent galaxies agree very well with the findings and predictions of Huang et al. (2017). It is hence likely that the flattening in the stellar mass – size relation of quiescent galaxies that we observe is a result of the baryonic radius scaling linearly with the dark-matter radius.

In Figure 16, we directly compare the models of the quiescent galaxy relation to the models of the star-forming stellar mass – size relation. In all four redshift bins, the star-forming models are shown in blue, while the models for the quiescent galaxies are in red. The corresponding blue and red shaded regions indicate the 1σ level, which include 68% of all MCMC models. Even with the 1σ levels, we find that the quiescent and star-forming galaxies populate distinct stellar mass–size relations. For each redshift bin, we also show the scatter around each model as a dashed region. These dashed regions show the running upper and lower 68th percentiles. Up to this point, we have not addressed the scatter around our best-fit models for either the star-forming nor the quiescent samples. From Figures 10 and 14, it can be seen that there is indeed a significant amount of scatter in galaxy sizes at a given stellar mass (Ruhland et al. 2009; Somerville et al. 2018). The observed scatter in galaxy size could be due to radial stellar migration. For instance, El-Badry et al. (2016) have shown that processes such as stellar feedback can cause galaxy size to fluctuate significantly in just a few hundred million years. In Figure 16, we quantify the amount of scatter around the best-fit trends and compare this scatter to the difference in the star-forming and quiescent best-fit models. We additionally show lines of constant effective mass surface density ($\Sigma = M_*/(2\pi R_e^2)$) following (e.g., Misgeld & Hilker 2011; Eigenthaler et al. 2018). These lines indicate the median mass density within the half-light radius, such that mass assembly along these lines implies density-invariant growth. Any galaxies which lie on a stellar mass–size relation which is parallel to these lines are building up mass both inside and outside their half-light radius.

For $z > 1$, we do not place much confidence in the quiescent best-fit models at low stellar masses, as we do not have many low mass quiescent galaxies at these redshifts to constrain the model. The best-fit model is extended to $10^{9.2} M_{\odot}$ for the $1.0 < z \leq 1.5$ bin and to $10^{9.8} M_{\odot}$ for $1.5 < z \leq 2$, as shown in Table 2. We have included the extrapolated model below these stellar masses down to $10^7 M_{\odot}$, where the width of the red extrapolated region shows the 1σ confidence level. We have done this to highlight that the high redshift models are ill-constrained at low stellar masses.

For the first two redshift bins, shown at the top of Figure 16, below $\sim 10^9 M_{\odot}$, the star-forming and quiescent models intersect, such that the quiescent galaxy model has a shallower slope than the

model for star-forming galaxies. Although the models are not the same at this low mass end, the star-forming and quiescent galaxies occupy the same region of the stellar mass – size plane. The scatter around the models is larger than the difference between the models, indicating that star-forming and quiescent galaxies could be evolving through the same processes in this stellar mass regime. In the $\sim 10^9 - 10^{10.5} M_{\odot}$ stellar mass range, the star-forming and quiescent galaxies occupy very different regions of the mass – size plane, suggesting that galaxies in this mass regime are building up their mass via different mechanisms depending on whether they are actively forming stars or not. Finally, massive galaxies, with $M_* \geq 10^{10.5} M_{\odot}$, also follow different stellar mass – size relations depending on whether they are quiescent or star-forming. Massive quiescent galaxies follow a steep stellar mass – size relation, indicating that on average, quiescent galaxies are rapidly growing in size, most likely due to minor mergers, while massive star-forming galaxies appear to lie on the same linear relation as their less massive counterparts. This suggests that star-forming galaxies across $\sim 10^7 - 10^{11.5} M_{\odot}$ are primarily building up their mass and size by forming new stars, while quiescent galaxies evolve through different mechanisms depending on their stellar mass.

Indeed, comparing the stellar mass – size relations to the lines of constant Σ reveals a few interesting features that support this argument. For $0.2 \leq z \leq 2$, the stellar mass – size relation of star-forming galaxies is shallower than the lines of constant Σ . This implies that star-formation allows galaxies to assemble mass predominantly inside their effective radius. Quiescent galaxies with low to intermediate stellar masses also follow stellar mass – size relations that are shallower than the lines of constant Σ , indicating that they too are primarily assembling their mass inside the half-light radius. Massive quiescent galaxies instead lie on relations that are steeper, suggesting that they are assembling mass in their outskirts. Indeed, it is known that these massive systems primarily grow via minor mergers (e.g., Buitrago et al. 2008; Naab et al. 2009; Oser et al. 2012), which cause galaxies to assemble mass in the outer regions.

To gain more insight into the underlying processes that drive these different phases of growth, in a future paper we will present the decomposition into bulge+disk components for the galaxies in the sample presented in this work to understand how different galaxy components build up their mass (Nedkova et al., in prep).

5.3 Link to Dwarf Galaxies

It is interesting to consider why the flattening of the quiescent sequence appears to occur at an effective radius of ~ 1 kpc and whether this flattening continues to exist when the stellar mass – size relation is extended to stellar masses below $10^7 M_{\odot}$, which we can not do with our data set. In order to study this low mass end of the stellar mass – size relation, dwarf galaxies need to be considered. Many previous works have studied the connection between galaxies and dwarf galaxies. For instance, Forbes et al. (2008) have shown that elliptical galaxies and the dwarf elliptical galaxies from Binggeli & Jerjen (1998) follow a continuous trend on the size–luminosity plane which flattens to a constant size of ~ 1 kpc. More recently, Eigenthaler et al. (2018) and Ordenes-Briceño et al. (2018) have shown that there is a similar flattening for dwarf and elliptical galaxies in the Fornax cluster in the stellar mass – size and size – luminosity relations, respectively. Eigenthaler et al. (2018) investigate the properties of the low surface-brightness dwarf galaxy population in the core region ($\lesssim r_{\text{vir}}/4$) of the Fornax galaxy cluster, using data obtained as part of the Next Generation Fornax Survey (NGFS). In Figure 17, we overplot their results onto the stellar mass – size relation derived in this work.

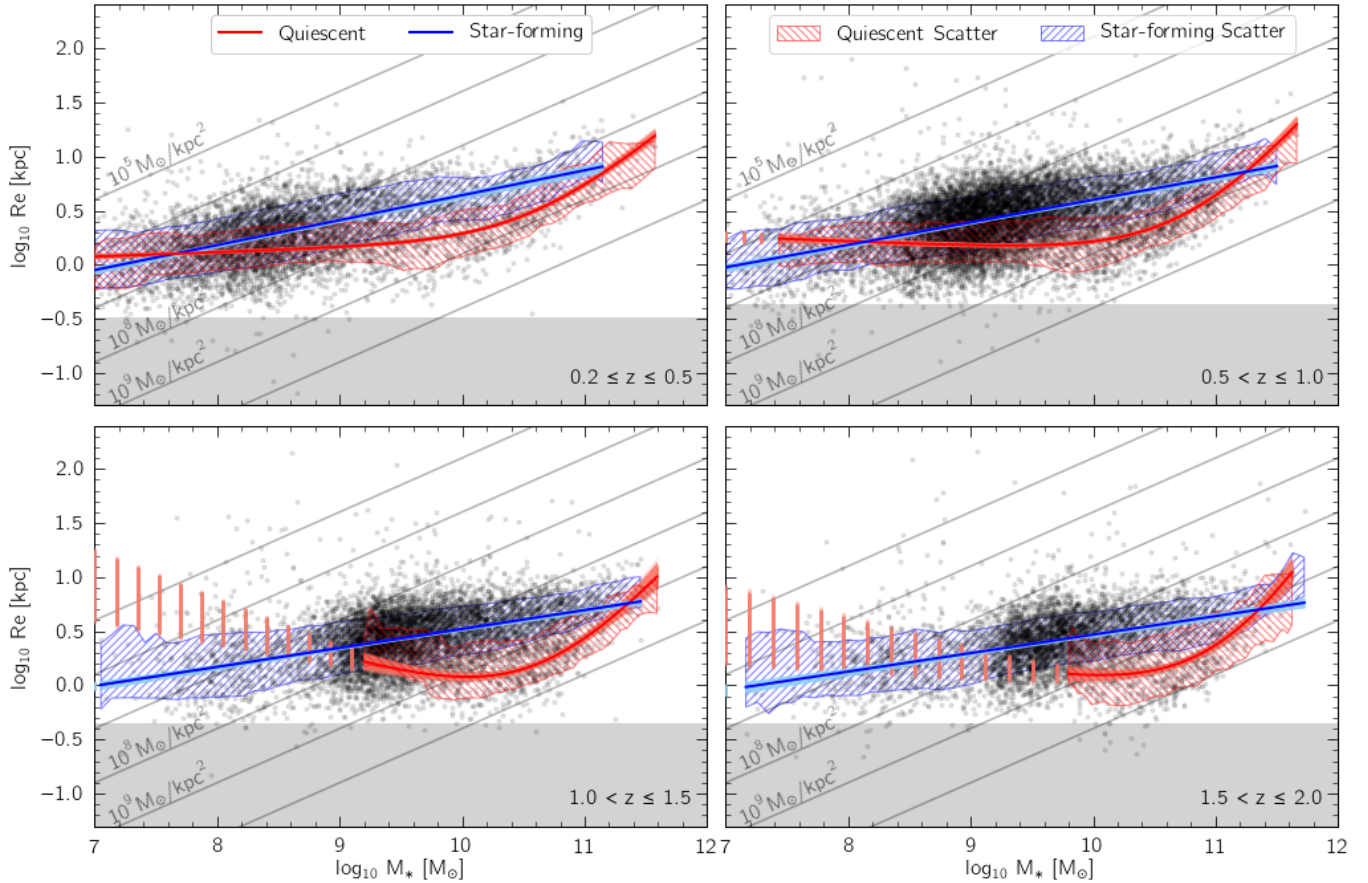


Figure 16. Comparison of the star-forming and quiescent best-fit curves from Figures 14 and 10, respectively. We show the entire sample (star-forming and quiescent) that falls into each redshift bin as black points. All models are shown as solid lines with a shaded region representing the 1σ confidence level over the stellar mass range over which the models are robust. The red and blue dashed regions indicate the scatter of the quiescent and star-forming galaxies, respectively. The stellar mass range over which we calculate the scatter is the same as the range over which we show our models. The grey indicate constant effective mass surface density (Σ). All models have also been extrapolated to $10^7 M_{\odot}$, where the extrapolation is shown as vertical lines indicating the 1σ spread of the model at low stellar masses. As in previous figures, the grey areas indicate sizes at $R_e < \text{FWHM}_{F160W}/2$ at the maximum redshift of each panel.

We note here that the effective radius is measured in the rest-frame g-band wavelength as opposed to the 5000\AA size. While this does not make a significant difference, as discussed in §3.4, we derive the stellar mass – size relation on the rest-frame g-band for consistency with Eigenthaler et al. (2018). Unfortunately, not enough information is available for the NGFS dwarf galaxies to reliably separate them into quiescent and star-forming galaxies. As this sample of dwarf galaxies is from the inner regions of the Fornax cluster, the dwarf galaxy morphology–density relation (Côté et al. 2009) would suggest that most are dwarf ellipticals (DEs), which are usually quiescent. Although the majority of the NGFS dwarf galaxies from Eigenthaler et al. (2018) are likely quiescent (e.g., Venhola et al. 2019; Carlsten et al. 2021), we compare them to both of our star-forming and quiescent mass – size relations as we cannot properly separate them using the same criteria used to identify quiescent galaxies from HFF and CANDELS.

Comparing the dwarf galaxies from Eigenthaler et al. (2018), it is interesting to see that they seem to align with *both* of our sequences. This could be coincidental, or contain interesting physics. On one hand, the dwarf galaxies seem to turn off at intermediate masses,

following the same ~ 1 kpc size that we find for the flattening of the quiescent galaxies, which might indicate a causal connection between these dwarfs and the quiescent galaxies we find. On the other hand, at lower masses, they follow closely (but not perfectly), our relation of the star-forming galaxies, indicating that they could be just a continuation of the mass–size relation of star-forming galaxies that we already find over four orders of magnitude in mass, to lower masses. As the dwarf galaxy sample is most likely a mixture of quiescent and star-forming objects, both might be true. But without pushing our analysis to lower masses and larger samples on better data, we can not distinguish the different processes just yet.

6 SUMMARY

By using a multi-wavelength modelling approach as well as the depth of the HFF survey and the large sample size of CANDELS, we have extended the evolution of the stellar mass – size relation over $0.2 \leq z \leq 2$ to low stellar mass galaxies and high redshift. We measure the stellar mass and morphological parameters, including

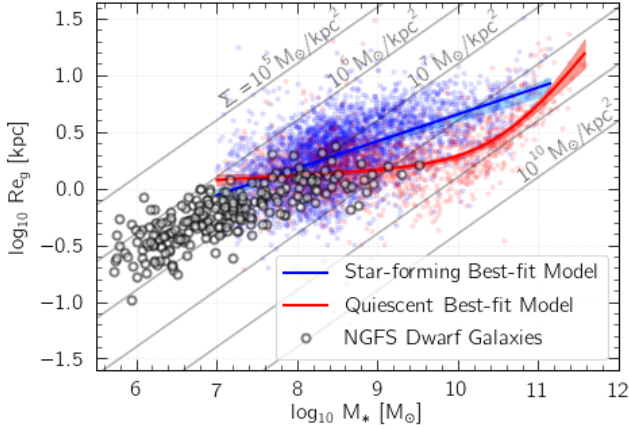


Figure 17. The stellar mass – size relation of the star-forming (shown in blue) and quiescent (shown in red) galaxies from the lowest redshift bin, i.e., $0.2 \leq z \leq 0.5$, where the half light radius has been measured in the rest-frame g-band instead of at 5000\AA . The stellar mass – size relation of dwarf galaxies from the Fornax galaxy cluster presented in Eigenthaler et al. (2018) are shown as empty black circles. Lines of constant effective mass surface density (Σ) are drawn in grey and labelled.

size and Sérsic index, in a consistent way across both HFF and CANDELS samples. From these data, we find that the star-forming and quiescent galaxies follow entirely separate sequences on the stellar mass – size plane regardless of whether star-forming and quiescent galaxies are separated based on star formation activity, colour, or Sérsic index. This result is best illustrated in Figure 6, where the quiescent sequence can be seen to flatten at lower stellar masses across all three selections. We observationally confirm that this flattening at around 1 kpc is present out to at least $z \sim 1.0$, with some signatures out to $z \sim 2$. These results confirm the findings of previous studies, but represent an important advance, since we have extended the stellar mass range over which we are measuring the stellar mass – size relation. Our main findings are as follows:

- The stellar mass – size relation of star-forming galaxies is well described by a single power law to at least $z = 2$ even when low stellar mass galaxies ($M_* \geq 10^7 M_\odot$) are included. Star-forming galaxies have a slope of $\beta \sim 0.2$ on the mass – size plane, across all redshifts studied in this work. While the slope does not show a strong evolution with redshift, the zero-point of the models increases with cosmic time, indicating that galaxies of the same mass were more compact at high redshift than their present-day counterparts. We note however, that low mass galaxies with $\sim 10^7 - 10^8 M_\odot$ occupy the same region of the stellar mass – size relation regardless of redshift, as shown in Figure 15. This suggests that these low mass star-forming galaxies are not significantly growing in size with cosmic time.

- The stellar mass – size relation of quiescent galaxies, on the other hand, is well described by a double power law at stellar masses $> 10^7 M_\odot$ due a flattening in the stellar mass – size relation that occurs at sizes of ~ 1 kpc. Dwarf galaxy studies (e.g., Eigenthaler et al. 2018; Ordenes-Briceño et al. 2018), however, suggest that this functional form might not hold at lower masses. High mass quiescent galaxies can be represented by a single power law, with a slope of ~ 0.7 over the redshift covered in this work. Our results at the high mass end are consistent with previous works as shown in Figure 8.

- We note that although the star-forming and quiescent galaxies models at the low mass end for the star-forming and quiescent sequences are different, the scatter around these relations is large at

stellar masses of $10^7 - 10^{8.5} M_\odot$. Although this increased scatter could be, and probably largely is, measurement errors on faint objects, this result hints at quiescent and star-forming galaxies evolving via the same mechanisms, or at least via mechanisms that influence galaxy size and mass growth in the same way.

- Finally, we show that, while the trends that we have recovered represent the data above $10^7 M_\odot$ well, extending the stellar mass – size relation to include dwarf galaxies can give a more comprehensive picture of galaxy formation and raises additional, interesting questions that deserve further investigations.

This study presents the stellar mass – size relation and its evolution down to $10^7 M_\odot$, but plenty of scope remains for future work. Exploring the stellar mass – relation down to even lower masses could help us better understand the connection between dwarf galaxies and giant galaxies. Additionally, while understanding how environment impacts galaxy evolution is outside the scope of this paper, the HFF presents a great data set to study this effect. Finally, with current facilities, observing low mass galaxies, particularly quiescent ones, at high redshift is challenging. Future studies with the James Webb Space Telescope, the Nancy Grace Roman Space Telescope, and Euclid present exciting possibilities for better understanding the stellar mass – size relation at high redshift.

7 ACKNOWLEDGEMENTS

We thank the anonymous referee whose comments and suggestions have significantly improved the quality of this paper. KVN gratefully acknowledges support from the Kathryn A. McCarthy Graduate Fellowship in Physics at Tufts University. KVN also thanks ESO, Chile for a seven-month scientific visit, which was instrumental in making this work possible. KVN and DM acknowledge the very generous support of the National Science Foundation under Grant Number 1513473 and by HST-AR-14302 and HST-AR-14553, provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555. KEW wishes to acknowledge funding from the Alfred P. Sloan Foundation. The Cosmic Dawn Center is funded by the Danish National Research Foundation under grant No. 140.

DATA AVAILABILITY

The 3D-HST, HFF-DeepSpace, and NGFS data that support the findings of this study are openly available. Data directly related to this publication and its figures are available on request from the corresponding author.

REFERENCES

- Abramowitz M., Stegun I. A., 1965, Handbook of mathematical functions with formulas, graphs, and mathematical tables
- Allen R. J., et al., 2017, *ApJ*, **834**, L11
- Baldry I. K., et al., 2012, *MNRAS*, **421**, 621
- Barden M., Häußler B., Peng C. Y., McIntosh D. H., Guo Y., 2012, *MNRAS*, **422**, 449
- Barro G., et al., 2014, *ApJ*, **791**, 52
- Berg T. A. M., Simard L., Mendel Trevor J., Ellison S. L., 2014, *MNRAS*, **440**, L66
- Bertin E., Arnouts S., 1996, *A&AS*, **117**, 393

- Bezanson R., van Dokkum P. G., Tal T., Marchesini D., Kriek M., Franx M., Coppi P., 2009, *ApJ*, **697**, 1290
- Bezanson R., et al., 2016, *ApJ*, **822**, 30
- Binggeli B., Jerjen H., 1998, *A&A*, **333**, 17
- Brammer G. B., van Dokkum P. G., Coppi P., 2008, *ApJ*, **686**, 1503
- Brammer G. B., et al., 2016, *ApJS*, **226**, 6
- Bruce V. A., et al., 2014, *MNRAS*, **444**, 1001
- Buitrago F., Trujillo I., Conselice C. J., Bouwens R. J., Dickinson M., Yan H., 2008, *ApJ*, **687**, L61
- Cappellari M., et al., 2013, *MNRAS*, **432**, 1862
- Carlsten S. G., Greene J. E., Greco J. P., Beaton R. L., Kado-Fong E., 2021, arXiv e-prints, p. arXiv:2105.03435
- Carollo C. M., et al., 2016, *ApJ*, **818**, 180
- Chabrier G., 2003, *PASP*, **115**, 763
- Chen Z., et al., 2020, *ApJ*, **897**, 102
- Cimatti A., et al., 2008, *A&A*, **482**, 21
- Côté S., Draginda A., Skillman E. D., Miller B. W., 2009, *AJ*, **138**, 1037
- Daddi E., et al., 2005, *ApJ*, **626**, 680
- Dahlen T., et al., 2013, *ApJ*, **775**, 93
- Damjanov I., et al., 2009, *ApJ*, **695**, 101
- de Souza R. E., Gadotti D. A., dos Anjos S., 2004, *ApJS*, **153**, 411
- de Vaucouleurs G., 1948, *Annales d'Astrophysique*, **11**, 247
- Delaye L., et al., 2014, *MNRAS*, **441**, 203
- Dimauro P., et al., 2018, *MNRAS*, **478**, 5410
- Dimauro P., et al., 2019, *MNRAS*, **489**, 4135
- Disney M. J., 1976, *Nature*, **263**, 573
- van Dokkum P. G., et al., 2008, *ApJ*, **677**, L5
- van Dokkum P. G., et al., 2010, *ApJ*, **709**, 1018
- van Dokkum P. G., et al., 2015, *ApJ*, **813**, 23
- Driver S. P., et al., 2011, *MNRAS*, **413**, 971
- Eigenthaler P., et al., 2018, *ApJ*, **855**, 142
- El-Badry K., Wetzel A., Geha M., Hopkins P. F., Kereš D., Chan T. K., Faucher-Giguère C.-A., 2016, *ApJ*, **820**, 131
- Erwin P., 2015, *ApJ*, **799**, 226
- Evans R., 1994, *MNRAS*, **266**, 511
- Fan L., Lapi A., De Zotti G., Danese L., 2008, *ApJ*, **689**, L101
- Forbes D. A., Lasky P., Graham A. W., Spitler L., 2008, *MNRAS*, **389**, 1924
- Foreman-Mackey D., 2016, *The Journal of Open Source Software*, **1**, 24
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, **125**, 306
- Furlong M., et al., 2017, *MNRAS*, **465**, 722
- Genel S., et al., 2018, *MNRAS*, **474**, 3976
- Graham A. W., 2019, *Publ. Astron. Soc. Australia*, **36**, e035
- Graham A. W., Guzmán R., 2003, *AJ*, **125**, 2936
- Häussler B., et al., 2007, *ApJS*, **172**, 615
- Häußler B., et al., 2013, *MNRAS*, **430**, 330
- Hill A. R., et al., 2017, *ApJ*, **837**, 147
- Holwerda B. W., Bouwens R., Oesch P., Smit R., Illingworth G., Labbe I., 2015, *ApJ*, **808**, 6
- Hopkins P. F., Bundy K., Hernquist L., Wuyts S., Cox T. J., 2010, *MNRAS*, **401**, 1099
- Huang K.-H., et al., 2017, *ApJ*, **838**, 6
- Huertas-Company M., Shankar F., Mei S., Bernardi M., Aguerri J. A. L., Meert A., Vikram V., 2013, *ApJ*, **779**, 29
- Kennedy R., et al., 2015, *MNRAS*, **454**, 806
- Koekemoer A. M., et al., 2011, *ApJS*, **197**, 36
- Kriek M., van Dokkum P. G., Labbé I., Franx M., Illingworth G. D., Marchesini D., Quadri R. F., 2009, *ApJ*, **700**, 221
- Kuchner U., Ziegler B., Verdugo M., Bamford S., Häußler B., 2017, *A&A*, **604**, A54
- La Barbera F., de Carvalho R. R., de La Rosa I. G., Lopes P. A. A., Kohl-Moreira J. L., Capelato H. V., 2010, *MNRAS*, **408**, 1313
- Labbé I., et al., 2005, *ApJ*, **624**, L81
- Lange R., et al., 2015, *MNRAS*, **447**, 2603
- Lotz J. M., et al., 2017, *ApJ*, **837**, 97
- Maltby D. T., et al., 2010, *MNRAS*, **402**, 282
- Marchesini D., van Dokkum P. G., Förster Schreiber N. M., Franx M., Labbé I., Wuyts S., 2009, *ApJ*, **701**, 1765
- Marsan Z. C., et al., 2015, *ApJ*, **801**, 133
- Martín N. S., et al., 2016, *ApJ*, **827**, L25
- McGrath E. J., Stockton A., Canalizo G., Iye M., Maihara T., 2008, *ApJ*, **682**, 303
- McIntosh D. H., et al., 2014, *MNRAS*, **442**, 533
- Misgeld I., Hilker M., 2011, *MNRAS*, **414**, 3699
- Momcheva I. G., et al., 2016, *ApJS*, **225**, 27
- Momcheva I. G., et al., 2017, *PASP*, **129**, 015004
- Moore B., Katz N., Lake G., 1996, *ApJ*, **457**, 455
- Morishita T., et al., 2017, *ApJ*, **835**, 254
- Mosleh M., et al., 2012, *ApJ*, **756**, L12
- Mosleh M., Tacchella S., Renzini A., Carollo C. M., Molaeinezhad A., Onodera M., Khosroshahi H. G., Lilly S., 2017, *ApJ*, **837**, 2
- Mowla L. A., et al., 2019, *ApJ*, **880**, 57
- Naab T., Johansson P. H., Ostriker J. P., Efstathiou G., 2007, *ApJ*, **658**, 710
- Naab T., Johansson P. H., Ostriker J. P., 2009, *ApJ*, **699**, L178
- Newman A. B., Ellis R. S., Bundy K., Treu T., 2012, *ApJ*, **746**, 162
- Norris M. A., et al., 2014, *MNRAS*, **443**, 1151
- Oesch P. A., et al., 2010, *ApJ*, **709**, L21
- Omand C. M. B., Balogh M. L., Poggianti B. M., 2014, *MNRAS*, **440**, 843
- Ordenes-Briceño Y., et al., 2018, *ApJ*, **859**, 52
- Oser L., Naab T., Ostriker J. P., Johansson P. H., 2012, *ApJ*, **744**, 63
- Papovich C., et al., 2012, *ApJ*, **750**, 93
- 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., 2010a, *AJ*, **139**, 2097
- Peng Y.-j., et al., 2010b, *ApJ*, **721**, 193
- Pillepich A., et al., 2018, *MNRAS*, **473**, 4077
- Ravindranath S., et al., 2004, *ApJ*, **604**, L9
- Rix H.-W., et al., 2004, *ApJS*, **152**, 163
- Ruhland C., Bell E. F., Häußler B., Taylor E. N., Barden M., McIntosh D. H., 2009, *ApJ*, **695**, 1058
- Salim S., Fang J. J., Rich R. M., Faber S. M., Thilker D. A., 2012, *ApJ*, **755**, 105
- Shankar F., Marulli F., Bernardi M., Mei S., Meert A., Vikram V., 2013, *MNRAS*, **428**, 109
- Shankar F., et al., 2014, *MNRAS*, **439**, 3189
- 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
- Shipley H. V., et al., 2018, *ApJS*, **235**, 14
- Simard L., 1998, GIM2D: an IRAF package for the Quantitative Morphology Analysis of Distant Galaxies. p. 108
- Simard L., et al., 2002, *ApJS*, **142**, 1
- Skelton R. E., et al., 2014, *ApJS*, **214**, 24
- Somerville R. S., et al., 2018, *MNRAS*, **473**, 2714
- Suess K. A., Kriek M., Price S. H., Barro G., 2019, *ApJ*, **877**, 103
- Sweet S. M., et al., 2017, *MNRAS*, **464**, 2910
- Tacchella S., Carollo C. M., Faber S. M., Cibinel A., Dekel A., Koo D. C., Renzini A., Woo J., 2017, *ApJ*, **844**, L1
- Trujillo I., et al., 2004, *ApJ*, **604**, 521
- Trujillo I., Conselice C. J., Bundy K., Cooper M. C., Eisenhardt P., Ellis R. S., 2007, *MNRAS*, **382**, 109
- Trujillo I., Chamba N., Knapen J. H., 2020, *MNRAS*, **493**, 87
- Venhola A., et al., 2019, *A&A*, **625**, A143
- Vika M., Bamford S. P., Häußler B., Rojas A. L., Borch A., Nichol R. C., 2013, *MNRAS*, **435**, 623
- Vikram V., Wadadekar Y., Kembhavi A. K., Vijayagovindan G. V., 2010, *MNRAS*, **409**, 1379
- Vulcani B., et al., 2014, *MNRAS*, **441**, 1340
- Weinmann S. M., Kauffmann G., van den Bosch F. C., Pasquali A., McIntosh D. H., Mo H., Yang X., Guo Y., 2009, *MNRAS*, **394**, 1213
- van der Wel A., Bell E. F., van den Bosch F. C., Gallazzi A., Rix H.-W., 2009, *ApJ*, **698**, 1232
- van der Wel A., et al., 2012, *ApJS*, **203**, 24
- van der Wel A., et al., 2014, *ApJ*, **788**, 28
- Whitaker K. E., van Dokkum P. G., Brammer G., Franx M., 2012, *ApJ*, **754**, L29
- Whitaker K. E., et al., 2017, *ApJ*, **838**, 19

- Williams R. J., Quadri R. F., Franx M., van Dokkum P., Labbé I., 2009, *ApJ*, **691**, 1879
- Wright A. C., Tremmel M., Brooks A. M., Munshi F., Nagai D., Sharma R. S., Quinn T. R., 2020, arXiv e-prints, p. [arXiv:2005.07634](https://arxiv.org/abs/2005.07634)
- Wuyts S., et al., 2007, *ApJ*, **655**, 51
- Yang L., Roberts-Borsani G., Treu T., Birrer S., Morishita T., Bradač M., 2021, *MNRAS*, **501**, 1028
- Zolotov A., et al., 2015, *MNRAS*, **450**, 2327

APPENDIX A: HFF STRUCTURAL CATALOGUES

Along with this work, we publish catalogues of structural properties for HFF galaxies that have been measured with GALAPAGOS-2 and GALFITM (Häußler et al. 2013). We provide a total of 12 catalogues – one for each cluster and parallel field. Although we use a combination of HFF and CANDELS galaxies for this study, we make available only HFF catalogues, as a catalogue of structural properties for CANDELS galaxies will be released in a forthcoming publication (Häußler et al., in prep.).

In this appendix, we describe the properties of the HFF structural catalogues. In Table A1, we list the columns with a description, as well as an example – namely, galaxy $\text{id}_{\text{HFF}} = 657$ from the Abell1063 parallel field that is shown in Figures 4 and 5. It is important to note that this structural catalogue is 3D catalogue in the sense that many of the columns (e.g., MAG_GALFIT_BAND, MAGERR_GALFIT_BAND, etc.) consist of seven values. The parameters that are reported for each band (i.e., those ending in ‘_BAND’) are reported for the F125W, F435W, F606W, F814W, F105W, F140W, and F160W bands, in this order. These catalogues are available through the VizieR database: <http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/MNRAS>, and all HFF data products are publicly available via the HFF-DeepSpace web page: <http://cosmos.phy.tufts.edu/~danilo/HFF/Download.html>.

APPENDIX B: COMPARISON TO DIMAURO ET AL. (2019)

As briefly discussed in §4.2.1, we find results for the high mass star-forming galaxy population that do not agree as well as expected with the stellar mass – size relations reported by Dimauro et al. (2019). In particular, the slopes derived for the star-forming galaxies are at odds especially at $z > 1$. This is particularly puzzling as the same data and largely identical software has been used to measure galaxy sizes. Here, we investigate a number of differences between our sample and the one used by Dimauro et al. (2019) to show how each impacts the derived stellar mass – size relation.

We begin by testing the size and mass distributions of the star-forming galaxies used in both works. These comparison are shown in Figure B1, where we show only the highest redshift bin as this is where the largest difference in the two samples can be seen; however, the conclusions drawn for this bin also hold across our entire redshift range. From Figure B1, it can be seen that at the high mass end (i.e., above stellar masses of $10^{9.5} M_{\odot}$), our sample predominantly consists of galaxies from the larger CANDELS fields. This is expected, as we use the HFF galaxies mostly to probe the low stellar mass range, where the number of CANDELS galaxies becomes more limited.

From the right panels of Figure B1, we note that while the number of galaxies is significantly different between the original samples (left panels), this difference is primarily caused by the stricter magnitude cut that Dimauro et al. (2019) use. We re-derive the best-fit lines for our sample when a magnitude cut of 23 in the F160W band is applied.

This fit is shown in green and is consistent with the high mass stellar mass – size relation that we derive for our sample, presented in Figure 13, and shown in Figure B1 as a black line. We therefore argue that the stellar mass – size relation that we derive at the high mass end does not significantly change when we limit the sample to brighter galaxies. This suggests that the different results between this work and Dimauro et al. (2019) are not a direct consequence of a different magnitude cut. Additionally, from Figure B1, it can be seen that when similar sample selections are applied to both samples, there are more galaxies with small sizes in our sample (vertical histogram on the right plot). The mass distributions of our sample (horizontal blue histogram on the right plot) are also skewed toward lower stellar masses compared to Dimauro et al. (2019), such that below $10^{10.3} M_{\odot}$, we have a significantly larger number of galaxies. Above this stellar mass, the mass distributions of our samples are comparable. Since we have more compact, low mass objects than Dimauro et al. (2019), this could possibly explain the different stellar mass – size relations.

To investigate this difference further, we match our sample to the Dimauro et al. (2019) sample in order to test the similarities between the galaxies that are used in both works versus those that are only used in one of the samples. The results are shown in Figure B2, where the left panels show our full sample in grey, where galaxies that are also used in Dimauro et al. (2019) are indicated with a red point. In the right panels, we show the same but for the Dimauro et al. (2019) sample (i.e., the Dimauro et al. 2019 sample is shown in grey, where galaxies that are also used in this work are indicated in red). In the histograms, we show the size and stellar mass distributions of each work, where the grey histogram corresponds to the sample from each study (and in the left panel is identical to the blue histogram in Figure B1). The red histograms are for galaxies that are used in both this work and Dimauro et al. (2019), and the hashed histogram is for the objects that are not matched, i.e., only used in one of the works.

For the star-forming galaxy population, we find that the stellar mass – size relation has a slope of ~ 0.2 at all redshifts, while Dimauro et al. (2019) derive a shallower slope, especially at high redshift. Indeed, visually from Figure B2, the Dimauro et al. (2019) sample on the right appears flatter than ours on the left. Upon modelling the Dimauro et al. (2019) sample, we also obtain a shallow slope of $B = 0.03$ at $1.5 < z \leq 2$. Although the fitting method used by both van der Wel et al. (2014) and Dimauro et al. (2019) is different from ours, we find results that are consistent with van der Wel et al. (2014) and we are able to reproduce the shallow results that Dimauro et al. (2019) find using their data. Likewise, if we fit our data with the same method as van der Wel et al. (2014) and Dimauro et al. (2019), by assuming the same functional form (i.e., Eq. 2), and computing the total likelihood for a set of six model parameters, we find very consistent parameters with those that we derive. These values are presented in Table B1, where our best-fit parameters for the high mass quiescent and star-forming galaxies from Tables 3 and 2 are shown in brackets for comparison.

In the right panel of Figure B2, we also show the galaxies that are selected as quiescent by our UVJ diagram, but identified as star-forming by Dimauro et al. (2019). These galaxies (shown as black crosses) are generally smaller and more massive compared to the overall star-forming sample, shown in grey, and therefore, including these objects in the fitting, would result in a flatter slope.

Finally, the way the uncertainties are obtained can significantly influence a model. We derive the size and mass uncertainties from GALFITM (Häußler et al. 2013) and FAST (Kriek et al. 2009), respectively, while Dimauro et al. (2019) estimate the size uncertainties from their simulations and the stellar mass uncertainties are assumed

Table A1. Columns from the released HFF structural catalogues. id_{HFF} corresponds to the id numbers from the HFF-DeepSpace catalogue (Shipley et al. 2018), while the remainder of the parameters are derived with the GALFITM and GALAPAGOS codes. All columns for which the description begins with ‘[7]’ are seven-element arrays. We provide all Re and Re error estimates in pixels as we apply size quality cuts based on the measurements in pixels. Hence, a pixel scale of $0.06''/\text{pixel}$ is needed to convert the measured Re into units of arcseconds. Position angles are defined such that 0 is ‘up’, increasing anticlockwise. Most non-integer parameter values are truncated for the example galaxy and shown here only for guidance regarding the format.

Column	Description	Example Galaxy
id_{HFF}	HFF-DeepSpace catalogue id	657
RA	SEXTRACTOR ALPHA_J2000	342.3186149
DEC	SEXTRACTOR DELTA_J2000	-44.5636569
CXX_IMAGE	SEXTRACTOR CXX_IMAGE	0.030213457
CYY_IMAGE	SEXTRACTOR CYY_IMAGE	0.11420833
CXY_IMAGE	SEXTRACTOR CXY_IMAGE	0.037952058
THETA_IMAGE	SEXTRACTOR THETA_IMAGE	-12.16
ELLIPTICITY	SEXTRACTOR ELLIPTICITY	0.53
KRON_RADIUS	SEXTRACTOR KRON_RADIUS	3.5
BACKGROUND	SEXTRACTOR BACKGROUND	9.684838E-5
FLUX_BEST	SEXTRACTOR FLUX_BEST	3.850794
FLUXERR_BEST	SEXTRACTOR FLUXERR_BEST	1.000767
MAG_BEST	SEXTRACTOR MAG_BEST	24.4761
MAGERR_BEST	SEXTRACTOR MAGERR_BEST	0.2822
FWHM_IMAGE	SEXTRACTOR FWHM_IMAGE	10.05
FLAGS	SEXTRACTOR FLAGS	3
CLASS_STAR	SEXTRACTOR CLASS_STAR	0.029
SKY_GALA_BAND	[7] GALAPAGOS-2 sky values at each band	{-1.77E-4, -7.49E-5, -2.19E-4, -1.47E-4, -1.19E-4, -2.31E-4, -1.86E-4}
SKY_SIG_BAND	[7] GALAPAGOS-2 sky values uncertainty. The sky value is fixed during the fit, but these values can be potentially used to detect ‘difficult’ fits	{2.07E-5, 1.73E-5, 2.86E-5, 7.12E-6, 1.56E-5, 2.33E-5, 1.36E-5}
SKY_FLAG_BAND	[7] GALAPAGOS-2 sky flag	{0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0}
FLAG_GALFIT	Single-Sérsic fit flag: -1 = not enough bands with data; not attempted, 0 = not attempted, 1 = fits started, but crashed, 2 = fits completed	2
NEIGH_GALFIT	Number of neighbouring profiles fit (or fixed) during the modelling	7
CHISQ_GALFIT	GALFITM χ^2 value	66011.9
CHISQ_GALFIT_PRIME	GALFITM χ^2 value within the primary ellipse	30272.7
NDOF_GALFIT	Degrees of freedom (DoF) allowed during the fit, includes number of pixels. Used to derive χ^2 and χ^2/ν	58203
NDOF_GALFIT_PRIME	same as NDOF_GALFIT but for the fit within the primary ellipse	18527
CHI2NU_GALFIT	GALFITM reduced χ^2 : χ^2/ν	1.13417
CHI2NU_GALFIT_PRIME	same as CHI2NU_GALFIT but for the fit within the primary ellipse	1.63398
use_flag	modelling parameter flag: 0 = at least one parameter has run into a fitting constraint; 1 = no parameters have run into constraints	1
X_GALFIT_DEG	DoF of x-position; not allowed to vary with wavelength	1
X_GALFIT_BAND	[7] x-position for F125W, F435W, F606W, F814W, F105W, F140W, and F160W, in this order	{57.064, 57.064, 57.064, 57.064, 57.064, 57.064, 57.064}
...		
MAG_GALFIT_DEG	DoF of apparent magnitude; full freedom has been allowed	7
MAG_GALFIT_BAND	[7] apparent magnitude measured at each band	{24.35, 26.10, 25.22, 24.68, 24.46, 24.29, 24.26}
MAGERR_GALFIT_BAND	[7] apparent magnitude uncertainties at each band	{0.007, 0.017, 0.008, 0.005, 0.005, 0.006, 0.007}
MAG_GALFIT_CHEB	[7] magnitude Chebyshev polynomial coefficients – as we allow full freedom, all values are nonzero	{24.799, -0.784, 0.338, -0.134, 0.0450, -0.002, -0.003}
MAGERR_GALFIT_CHEB	[7] uncertainties on the magnitude Chebyshev polynomial coefficients	{0.002, 0.008, 0.005, 0.003, 0.003, 0.003, 0.005}
RE_GALFIT_DEG	DoF of effective radius (Re)	3
RE_GALFIT_BAND	[7] Re [in pixels] at each band	{5.07, 7.01, 6.44, 5.82, 5.31, 4.98, 4.97}
REERR_GALFIT_BAND	[7] Re uncertainty [in pixels] at each band	{0.08, 0.24, 0.14, 0.08, 0.08, 0.08, 0.12}
RE_GALFIT_CHEB	[7] Re Chebyshev polynomial coefficients – as we use a 2nd order Chebyshev polynomial, the first 3 values are nonzero	{5.71, -1.02, 0.28, 0.0, 0.0, 0.0, 0.0}
REERR_GALFIT_CHEB	[7] uncertainties on the Re Chebyshev polynomial coefficients	{0.03, 0.13, 0.09, 0.0, 0.0, 0.0, 0.0}
N_GALFIT_DEG	DoF of Sérsic index (n)	3
N_GALFIT_BAND	[7] n at each band	{1.37, 0.86, 1.02, 1.18, 1.31, 1.39, 1.38}
NERR_GALFIT_BAND	[7] n uncertainty at each band	{0.05, 0.10, 0.05, 0.04, 0.05, 0.06, 0.09}
N_GALFIT_CHEB	[7] n Chebyshev polynomial coefficients	{1.20, 0.26, -0.08, 0.0, 0.0, 0.0, 0.0}
NERR_GALFIT_CHEB	[7] uncertainties on the n Chebyshev polynomial coefficients	{0.02, 0.06, 0.05, 0.0, 0.0, 0.0, 0.0}
Q_GALFIT_DEG	DoF of axis ratio; not allowed to vary with wavelength	1
Q_GALFIT_BAND	[7] axis ratio at each band	{0.144, 0.144, 0.144, 0.144, 0.144, 0.144, 0.144}
...		
PA_GALFIT_DEG	DoF of position angle; not allowed to vary with wavelength	1
PA_GALFIT_BAND	[7] position angle at each band	{76.21, 76.21, 76.21, 76.21, 76.21, 76.21, 76.21}
...		

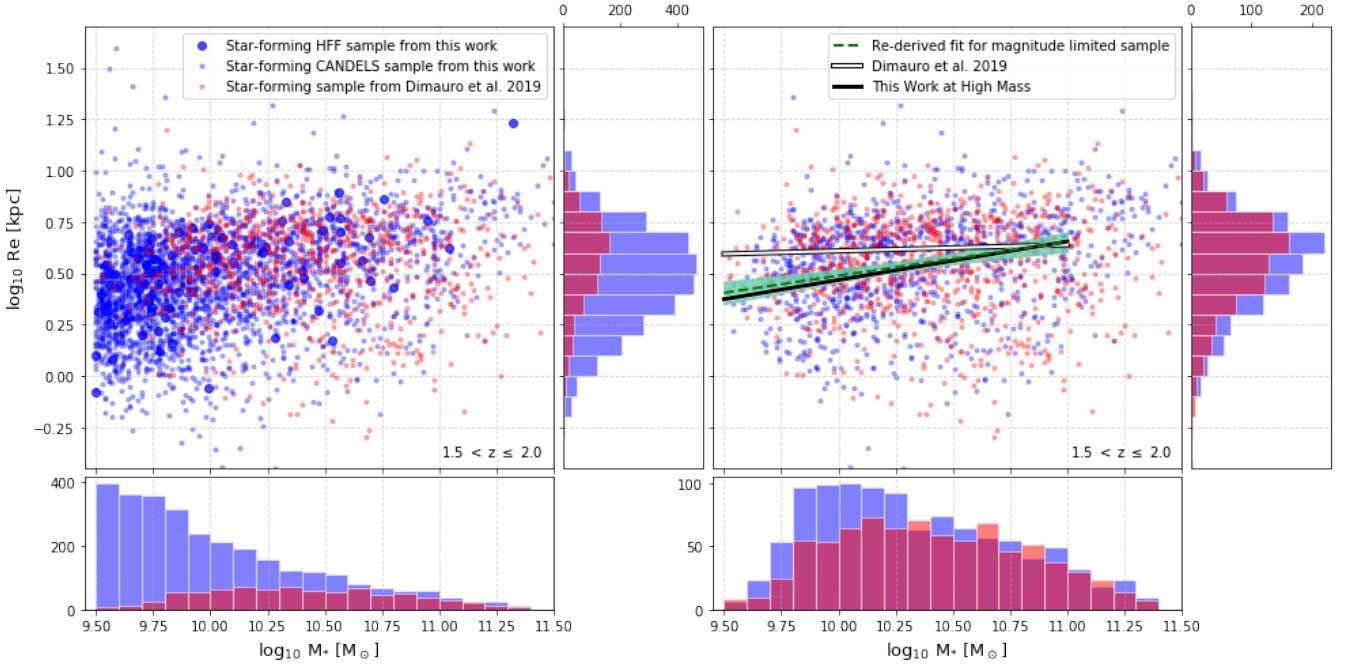


Figure B1. The stellar mass – size relation of high mass star-forming galaxies from our sample shown in blue, where the size of the points indicates whether they come from the CANDELS fields or the HFF, and the [Dimauro et al. \(2019\)](#) star-forming sample shown in red. In the left panels, we show the sample comparison using our original magnitude cut at 1 magnitude brighter than the 90% detection completeness limit. In the right panels, we show the sample comparison using the same magnitude cut for both CANDELS samples i.e., $m_{F160W} \leq 23$ following [Dimauro et al. \(2018\)](#). The size distributions of our sample (blue) and the [Dimauro et al. \(2019\)](#) sample (red) are shown as vertical histograms, while the stellar mass distributions are shown as horizontal histograms. The purple regions indicate where the two histograms overlap. In the right panel, we show the best-fit stellar mass – size relations from this work and [Dimauro et al. \(2019\)](#), as indicated in the legend. Additionally, we re-derive the relation for our sample with a magnitude cut of $m_{F160W} \leq 23$ and the result is shown in green, where the shaded green region indicates the 1σ confidence level.

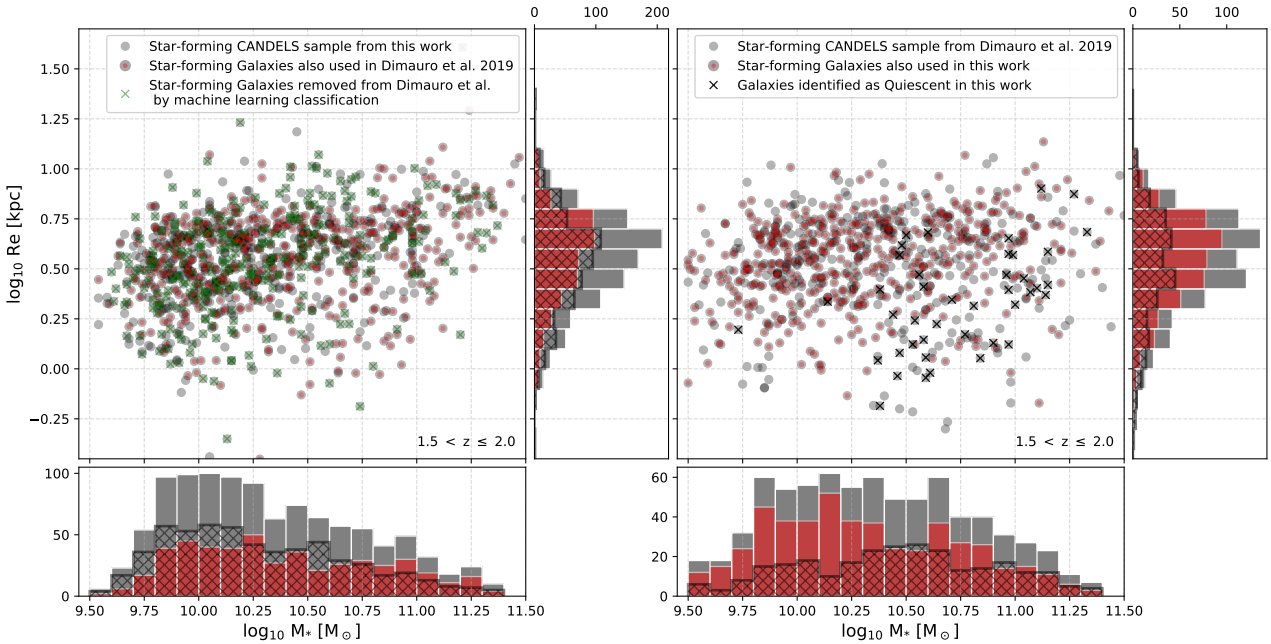


Figure B2. The stellar mass – size relation of high mass star-forming galaxies in the highest redshift bin from our sample (left panels) and the [Dimauro et al. \(2019\)](#) (right panels). Galaxies included in both samples are shown in red. In the left panel, a large fraction of the sample that we use is not used in [Dimauro et al. \(2019\)](#) as a result of the machine learning classification that they use (see [Dimauro et al. \(2018\)](#) for details). These objects are indicated in green.

Table B1. The estimated best-fit parameters for quiescent and star-forming galaxies following the same methods as [van der Wel et al. \(2014\)](#) and [Dimauro et al. \(2019\)](#) where $\sigma \log_{10}(\text{Re})$ is the scatter in size. These results are closely consistent with the results from our own fitting method, which we report in brackets, and are shown in Tables 2 and 3 for the quiescent and star-forming samples, respectively.

z	Quiescent Galaxies			Star-forming Galaxies		
	$\log_{10}(A)$	B	$\sigma \log_{10}(\text{Re})$	$\log_{10}(A)$	B	$\sigma \log_{10}(\text{Re})$
$0.2 \leq z \leq 0.5$	0.69 [0.61±0.01]	0.72 [0.68±0.04]	0.19	0.78 [0.78±0.03]	0.21 [0.22±0.03]	0.22
$0.5 < z \leq 1.0$	0.48 [0.45±0.01]	0.72 [0.64±0.03]	0.18	0.74 [0.74±0.01]	0.21 [0.21±0.02]	0.22
$1.0 < z \leq 1.5$	0.31 [0.28±0.01]	0.63 [0.63±0.04]	0.20	0.66 [0.66±0.01]	0.20 [0.21±0.01]	0.22
$1.5 < z \leq 2.0$	0.24 [0.18±0.01]	0.54 [0.61±0.05]	0.25	0.59 [0.59±0.01]	0.18 [0.20±0.02]	0.23

to be proportional to the size uncertainties. We test this idea by fitting the matched objects from our work (i.e., the red points from the left panel in Figure B2) by using the uncertainties from [Dimauro et al. \(2019\)](#). We find no significant change in the best-fit model when either set of uncertainties is used, therefore, the errors on the measurements are not driving the discrepancy between our work and [Dimauro et al. \(2019\)](#).

From all of the various tests carried out to investigate this discrepancy between our results, we are unable to say specifically which of these effects dominate the difference. However, we can rule out the fitting methods themselves as well as the weighting scheme used. From the visual impression in Figure B1, it can be seen that the sample selection of quiescent and star-forming galaxies is different. For instance, the majority of the objects in the [Dimauro et al. \(2019\)](#) sample which are *not* in our star-forming sample are massive and small, possibly indicating that they are elliptical quiescent galaxies. This leads us to the conclusion that the main difference is most likely a result of the sample selection.

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