

Master Thesis

Galaxy Morphology with the James Webb Space Telescope at Redshifts beyond 3

by

Arpita Ganguly

June 21 2024

Supervisors

Dr. Mengyuan Xiao, Prof. Pascal Oesch

Thesis Jury: Dr. Mengyuan Xiao (advisor)
Prof. Pascal Oesch (advisor)
Dr. Miroslava Dessauges
Observatoire de Genève



**UNIVERSITÉ
DE GENÈVE**

FACULTÉ DES SCIENCES
Département d'astronomie

Acknowledgements

To Mengyuan and Pascal, thank you for your constant guidance and encouragement during this project. I am grateful for the knowledge and research experience I gained under your mentorship. It has truly been an amazing experience. I am thankful for all the insightful and candid conversations I could have with you during my final year.

To Andrea, Rui and my friend, Maxime, thank you for taking the time for our many conversations on science and coding.

Thank you to all the members of the Galaxies at Cosmic Dawn group, I loved sharing my work and getting to know all the exciting work being done by everyone during the weekly meetings.

To Mirka, thank you for agreeing to be on my jury. I hope you will enjoy reading this thesis.

To Daniel, for making my life as an international student easier. Thank you for your unwavering support.

To all my friends in the master's programme, Laurent, Sara, Maxime, Marine, Avidaan, Chloe, Antonin, Yuna and Samson, thank you for making these two years special.

To Loris, thank you for being such a pillar of support and love for me during these past two years. To my parents, Kabita and Arijit, for making my dreams come true and for all the assurance and love you gave during the good and bad times.

Abstract

The unprecedented resolution and sensitivity of the James Webb Space Telescope (JWST) in the Infrared (IR) wavelengths, has revealed a population of extremely dusty and massive star-forming galaxies (SFGs) which remained elusive to detection by the Hubble Space Telescope (HST). These sources, called Optically Faint Galaxies (OFGs) have been missed by studies of rest-frame UV-selected galaxies, and are likely to contribute significantly to the cosmic star formation rate density of the early universe. Thus studying the nature of these galaxies is crucial to understand their role in galaxy formation and evolution.

In this work, we focus on the stellar morphology using the Sérsic light profile of OFGs and try to understand the connection between stellar mass and size, and dust obscuration in OFGs within the redshift range of $3 < z < 10$. We have used three large surveys in the PRIMER and CEERS fields utilising the 4.44 m JWST-NIRCam (Near Infrared Camera) images for our analysis. We classify OFGs (and Little Red Dots, LRDs) using colour cuts based on previous studies.

Our main results revolve around the physical properties of OFGs and how they compare to galaxies in the parent SFG sample. We derive different scaling relations for these sources, namely, the mass-size relation and effective stellar mass density ($\Sigma_{Re,kpc}$)-mass relations, with a focus on the redshift range of 3 to 4 due to the abundance of OFGs in this range. We see that OFGs follow slightly different scaling relations as compared to parent SFGs. Their general location on the scaling relations is seen to be intermediate between SFGs at the same redshifts and quiescent galaxies (QGs) seen at lower redshifts ($2 < z < 3$), hinting at possible signs of evolution towards QGs. Stellar mass is found to be a clear proxy for dust attenuation. Scaled at the same stellar mass, OFGs show over 4 times higher dust attenuation and about 2 times smaller effective radii than parent SFGs. Orientation has no correlation with the dust attenuation of our sources. We also perform a PSF-Sérsic decomposition and find that 52% of the LRDs come out to be “PSF-dominated” ($> 50\%$ flux contribution from the PSF component in the light profile) while only 40% of OFGs show the same nature.

Contents

1	Introduction	4
1.1	Background of this thesis	6
1.1.1	Optically Faint/ Dark Galaxies (OFGs)	6
1.1.2	Little Red Dots (LRDs)	7
2	Observations and Sample Selection	8
2.1	JWST/NIRCam and the CEERS + PRIMER Survey	9
2.2	Selection of Optically Dark/ Faint Galaxies at $z \geq 3$	12
3	Fitting Galaxy Morphologies	14
3.1	The Sérsic Profile	15
3.2	The Point Spread Function(PSF) Profile:	16
3.3	Fitting Algorithms	17
3.3.1	GALFIT	17
3.3.2	Implementing GALFIT	18
3.4	pysersic	19
3.4.1	Forward Modelling, Bayesian Statistics & MCMC Sampling	19
3.4.2	Setting Priors in pysersic :	20
4	Running Tests	21
4.0.1	Creating fake sources	22
4.1	Results from Tests	25
4.1.1	GALFIT or pysersic ?	25
4.2	Testing the limits of pysersic	28
5	Results	29
5.1	Preliminary results from pysersic analysis	29
5.2	Physical Properties of Optically Dark/Faint Galaxies	32
5.3	Drivers of Dust Attenuation	37
5.4	OFGs: Compact progenitors of low redshift Quiescent galaxies?	38
6	Discussion	40
7	Conclusion & Future work	42
A	Appendix	45

1 Introduction

Galaxies, the building blocks of the universe, have evolved over cosmic time through processes shaped by gravity, gas dynamics, interactions and, stellar feedback. The structure formation of galaxies, galaxy clusters and larger structures arose from small early density fluctuations seen in the cosmic microwave background (CMB) which then grew through gravitational amplification to give rise to today's structures. The Λ Cold Dark Matter (Λ CDM) cosmological model is now a well-established standard model to provide the framework for understanding galaxy formation and evolution, where Λ stands for the component of the Universe responsible for its accelerated expansion, Dark Energy. It accounts for the hierarchical structure formation scenario (or bottom-up scenario) of the universe where matter density grows with time with small fluctuations in the universe through the gravitational collapse of cold dark matter. The CDM paradigm is the favoured scenario as it also preserves the fine structures observed today and reproduces the observed galaxy distribution in the universe (Frenk 1986) and is now supported by a variety of cosmological probes such as the CMB (Planck Collaboration et al. 2016), the Hubble expansion rate estimated from Type Ia supernovae (Dhawan, Jha, and Leibundgut 2018) and the halo mass function (HMF; Vikhlinin et al. 2009).

From the dawn of the discovery of galaxies, one of the fundamental aims of their study has been to characterise their physical structures and morphologies. The morphology of a galaxy tells us about the distribution of stars, gas and, dust in it. Morphological studies usually probe the stellar content of the galaxy under study. In the local universe, galaxies can be split mainly into spiral, elliptical and irregular galaxies. The Hubble tuning fork diagram is a well-known classification scheme used to categorise early-type and late-type galaxies (Figure 1).

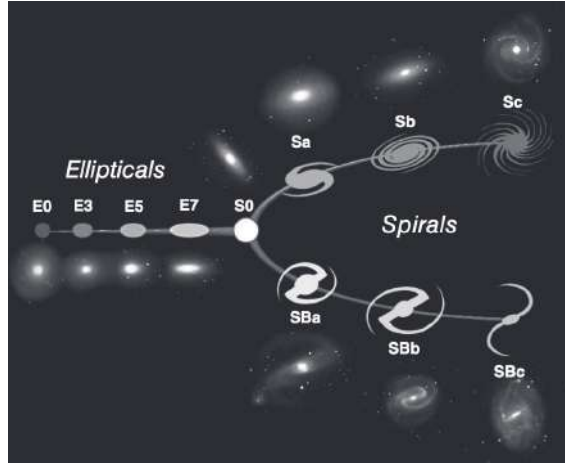


Figure 1: Edwin Hubble Tuning Fork diagram, to classify galaxy morphology of local or low redshift (z) galaxies. Credit: NASA, ESA reproduced image in “Introduction to Galaxy Formation and Evolution”, by Andrea Cimatti, Filippo Fraternali and Carlo Nipoti, 2019.

While the majority of galaxies in the local universe can be placed on the Hubble sequence, not much is known about the different types of galaxies observed at higher redshifts (z). The morphology of galaxies shows their evolutionary history and is strongly linked to physical properties such as the stellar mass, star formation

rates and past interactions or merger events. Galaxies at high redshifts are more prone to mergers in the early universe showing irregular morphology, making their visual classification difficult (Tohill, Bamford, and Conselice 2023). So, we try to probe the physical properties of high- z galaxies using their stellar morphology and place them within a larger context of galaxy build-up in the early universe. There exist scaling relations involving galaxy masses, sizes, and characteristic velocities, like the Tully-Fisher relation, which point towards some regularity in the formation and assembly of galaxies. Thus, studying such scaling relations using the stellar morphology of galaxies viewed at different snapshots of time can give us an idea about how these galaxies evolved across cosmic history. This can be achieved by

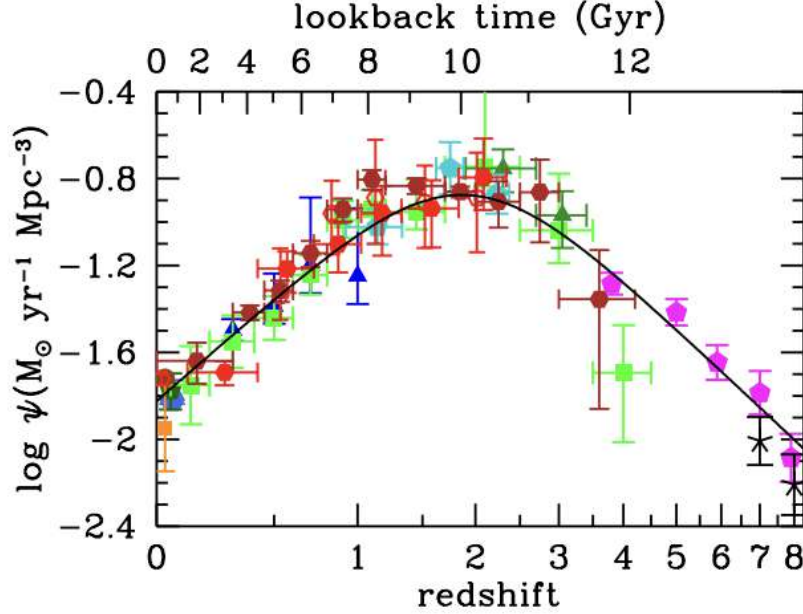


Figure 2: Cosmic Star Formation Rate Density versus lookback time in Gyr, Credit: Madau & Dickinson, 2014

conducting deep wide-field surveys through which we can peer and observe galaxies at increasing relative distances from us.

Ground and space-based observatories like the Hubble Space Telescope (HST), Spitzer Space Telescope and the Atacama Large Millimeter/submillimetre Array (ALMA) have been instrumental in advancing our understanding of the extragalactic universe. HST with its WFC3 camera, has provided us with some of the deepest imaging data in surveys like the Hubble Ultra Deep Field (HUDF), and the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS) owing to its sensitivity and resolution in the optical or rest-frame UV observations, greatly expanding our census of distant galaxies up to $z \sim 10$ (R. J. Bouwens et al. 2010; P. Oesch et al. 2016). Spitzer and ALMA revealed dusty galaxies from within the first few billion years of the universe, helping us determine their physical properties and star formation activity (Alcalde Pampliega et al. 2019, J. A. Zavala et al. 2021, Casey, Jorge A. Zavala, et al. 2021). We now have the James Webb Space Telescope (JWST) to push the observational frontiers beyond what one could have expected before.

We realise that unobscured galaxies can be easily studied using optical observatories while for dusty galaxies to be detected, we need to probe longer wavelengths like the near/ far infrared and millimeter wavelengths. Despite the presence of other infrared instruments, what makes JWST far more revolutionary is its unprecedented sensitivity and resolution in the near infrared regime of wavelengths. Thus, JWST is not only probing a much higher redshift universe to witness the dawn of galaxy formation and evolution but also allowing us to detect and inspect distant, dusty galaxies, which shine brighter at infrared wavelengths than in the optical in our observer frame.

1.1 Background of this thesis

Prior to the advent of JWST, our understanding of galaxies in the early universe ($z \geq 3$) primarily relied on samples selected based on their rest-frame ultra-violet (UV) light, as comprehensively reviewed by Madau and Dickinson 2014. While UV-based selection techniques, such as the Lyman break dropout (Madau, Ferguson, et al. 1996), to select Lyman-break galaxies (LBGs; Fu et al. 2024, R. Bouwens et al. 2020, P. A. Oesch et al. 2014) or the search for Ly- α emitters (LAEs), are very effective at identifying blue, (typically) low-mass star forming galaxies, these methods are strongly biased against red, dusty, or evolved galaxies, which typically make up most of the massive galaxy population at mid-to-high redshifts (Marrone et al. 2018). The cosmic star formation history (SFH) also relies on the most massive, dusty starburst galaxies, which are limited to a rare population and are not characteristic of the more common type of galaxies on the star formation main sequence (SFMS; eg., Magdis et al. 2010 Schreiber et al. 2015, etc.). Thus, rest-frame UV-selected samples at high redshift are likely incomplete, missing massive red galaxies that could potentially be identified with observations at longer wavelengths. This means that our comprehension of the cosmic star formation rate density of the universe before JWST (see Figure 2), had overlooked a potentially significant contribution of massive, dust-obscured star forming galaxies at high redshifts (Wang, Schreiber, and Elbaz 2019).

1.1.1 Optically Faint/ Dark Galaxies (OFGs)

While we know that there exist a small number of extremely dusty, highly star forming galaxies (SFGs), the submillimeter galaxies (SMGs) at $z > 4$, in recent years there has been increasing evidence that there may exist a much more common population of massive dusty galaxies with milder star formation rates in the early universe. They are believed to be the likely significant contributors to the massive ($M_* > 10^{10} M_\odot$) end of the SFR density of the Universe at $3 \lesssim z \lesssim 6$, (Barrufet et al. 2023; Xiao et al. 2023). Among these galaxies are the optically faint or dark galaxies (OFGs). These galaxies, often also referred to as “HST-Dark” or “H-dropout” galaxies, have remained elusive to detection by the Hubble Space Telescope, hence the name, due to attenuation of their light by dust (Manning et al. 2022, Enia et al. 2022, C. Gómez-Guijarro et al. 2022, Xiao et al. 2023). Dust grains absorb UV and optical light and emit strongly in the infrared (IR) and millimetre (mm) wavelengths, thereby making it difficult to study their detailed stellar morphologies in the HST era, see Figure 3. While some of these very dusty galaxies were already known from IRAC/Spitzer surveys, some were also missed as they were not only fainter but

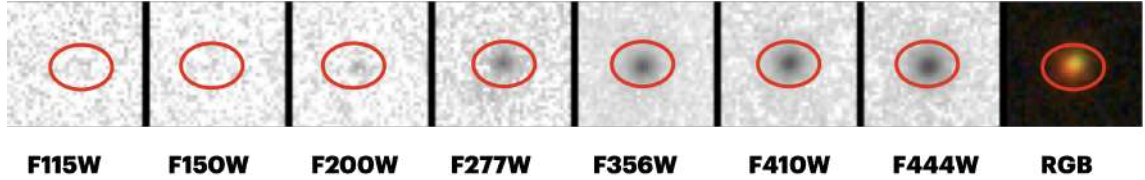


Figure 3: Example of an optically faint/dark galaxy as seen in 7 NIRCam bands F115W, F150W, F200W, F277W, F356W, F410W, F444W from left to right and RGB stacked image. Credit: Postage stamps taken using FitsMap (Hausen and Robertson 2022)

also the resolution of IRAC was about ten times worse than JWST, meaning that they could have been blended with foreground sources. With the advent of JWST, we have been able to overcome these challenges owing to its high sensitivity and resolution in IR wavelengths (Figure 3). Recent studies as in Gottumukkala et al. 2024, studied stellar mass functions (SMFs) of such massive obscured galaxies and suggested that red, optically faint galaxies represent the missed fraction of about 20% of the galaxy population at $3 < z < 4$. At $4 < z < 6$ epochs in the early universe, these galaxies overtake the SMF seen in the pre-JWST era at high stellar mass of around $\log M_*/M_\odot \sim 10.375$. They also show a rapid evolution at the high mass end of the SMF ($\log M_*/M_\odot \sim 9.5$ to 11), suggesting rapid dust obscured stellar mass assembly at these epochs. Previous studies on the morphological properties of OFGs like Carlos Gómez-Guijarro et al. 2023 used JWST data from the Cosmic Evolution Early Release Science (CEERS) survey, an early release science program in the Extended Groth Strip (EGS) field (Finkelstein, Bagley, et al. 2022). The CEERS data include JWST/NIRCam imaging over ten pointings covering a total area of 97 arcmin². They also used ancillary HST to complement shorter wavelengths. They studied the drivers of dust attenuation, particularly the mass, size and orientation of the selected OFGs. Drawing significant inspiration from this work, one of our key objectives is to also determine the main drivers of dust attenuation within our chosen sample of OFGs. What sets us apart and gives us an edge over this previous study is our utilisation of not only the CEERS field but also the PRIMER-UDS and the PRIMER-COSMOS fields (refer to section 2.1 for more details). This expanded dataset provides us with a larger pool of galaxies thus making our statistical analyses more robust. We have made a similar colour selection as done in this study but applied it to new datasets. Our statistical sample size (397 OFGs selected from preliminary sample selection) is drastically larger as compared to this study (25 OFGs selected), see section 2.2).

1.1.2 Little Red Dots (LRDs)

With the advent of JWST, many sources at high redshift started appearing like point sources. Labbe2023a studied 6 allegedly massive dusty star forming galaxies. This study stated that galaxies with stellar masses as high as $\sim 10^{11}$ solar masses had been identified out to redshifts of $z \sim 6$, i.e., approximately one billion years after the Big Bang. This implied that if the redshifts and fiducial masses were to be correct, then the mass density in the most massive galaxies would exceed the *total* previously estimated mass density (integrated down to $M_* = 10^8 M_\odot$) by a factor of ~ 2 at $z \sim 8$ and by a factor of ~ 5 at $z \sim 9$. A fundamental issue arose with

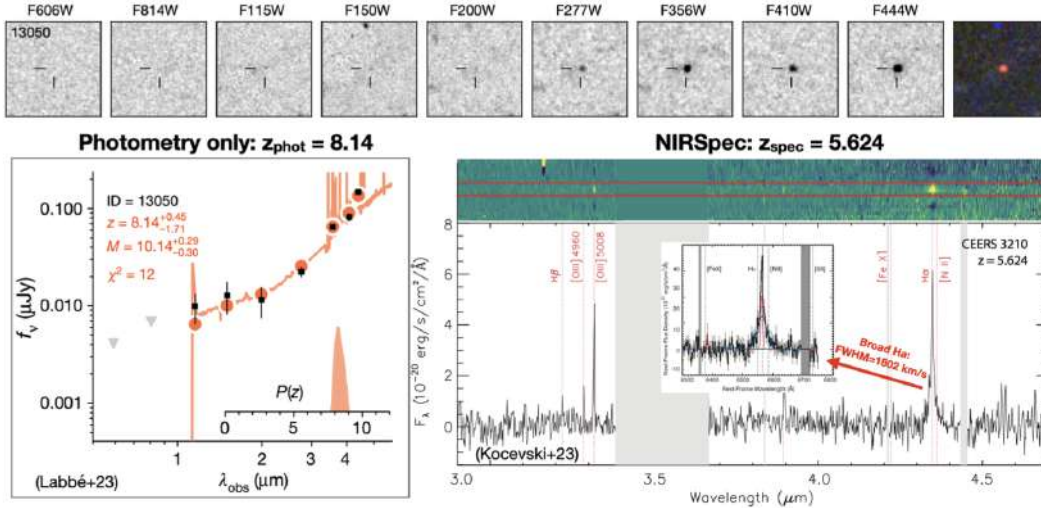


Figure 4: A galaxy with photometric data, showing ultra-massive properties with $\log(M_*/M_\odot) = 10.14^{+0.29}_{-0.30}$ at $z_{\text{phot}} = 8.14$ (Labbe+23). However, NIRSpect observations reveal this source to be a low-mass, broad line AGN at $z_{\text{spec}} = 5.624$ (Kocevski+23).

these new findings. Stellar mass densities as high as these are difficult to realise in a standard Λ CDM cosmology, as pointed out by several studies (Menci et al. 2022; Boylan-Kolchin 2023). The fiducial mass densities push against the limit set by the number of available baryons in the most massive dark matter halos thus puzzling astronomers as to how they could have gained so much mass in such a relatively short time. In addition, Matthee et al. 2024 used spectroscopic data from FRESCO and EIGER surveys, and selected broad-line AGNs (Active Galactic Nuclei) and from their morphology they all resembled LRDs. They are believed to be compact (with a small radius) galaxies with a type 1 AGN and obscured (covered by dust), which accounts for their red colour and why they are easily observed in the infrared. The spectrum of the galaxy has a ‘V’ shape, a blue continuum from the unobscured part of the AGN in the galaxy, and a red continuum from the obscured AGN (Labbe et al. 2023; Matthee et al. 2024; Barro, Pérez-González, et al. 2024). People further inspected such sources and as shown in the study by Kocevski et al. 2023, with spectroscopic follow-up, one of the sources studied in Labbe+23b has been found to be an AGN at lower redshift (Figure 4). Hence, there is a strong link between LRDs and AGNs, thus making them interesting objects to study.

The AGN are very interesting in their own right, but having said that, they end up contaminating our catalogues of galaxies as their estimated stellar mass, redshift and dust attenuation are most likely wrong since the AGN component is not included in our templates forming the Spectral Energy Distribution (SED) fits. Hence a further analysis of size and morphology is required to identify possible AGN-dominated sources among our colour-selected OFGs and LRDs.

2 Observations and Sample Selection

The primary goal of this thesis is to understand the nature of OFGs and in a larger context, their contribution to galaxy formation and evolution at high redshift. Out

of several possible scientific approaches, we try to use stellar morphology as a probe to understand the behaviour of these massive, dusty star forming galaxies (SFGs). One may ask should we study galaxy morphology? The physical appearance of a galaxy in itself contains a lot of information about its evolutionary history. Different morphological features not only help us categorise galaxies but also give us insights into various internal and external processes that may have shaped the galaxies over time. Stellar light is an indicator of the stellar mass distribution of galaxies. Thus studying the light profiles of galaxies can tell us how the mass is distributed in them, thus further giving hints about their evolution.

With this motive, we intend to fit Sérsic profiles on OFGs to characterise their stellar light distribution, and to understand their mass-size relations and effective surface mass densities in comparison to main sequence (MS) SFGs and quiescent galaxies (QGs). Since they are dusty galaxies, they are characterised by their reddening effect due to the attenuation of the stellar continuum emission by dust. This causes these galaxies to appear faint or invisible at shorter wavelengths like $1.5\ \mu\text{m}$ and very bright and red at longer wavelengths like $4.4\ \mu\text{m}$.

To accomplish our goals, we exploit the high sensitivity and resolution of JWST NIRCam in infrared wavelengths. Moreover, the nature of our research is to provide robust statistics on the physical properties of selected galaxies for which we need a large statistical sample. This can be achieved by using large sky coverage surveys. Therefore, in this section, we discuss the James Webb Space Telescope and its infrared imaging instrument, NIRCam. Specifically, we focus on three large fields: CEERS, PRIMER-COSMOS, and PRIMER-UDS. We will select a parent sample comprising of all star forming galaxies at redshifts above 3, followed by a more rigorous selection of OFGs (and LRDs). We will explore in detail the process of selecting our desired sources from the aforementioned fields, describing methods of preliminary sample selection. These methods are further refined through simulation tests, aiming to identify the optimal sample for measuring the morphological parameters of the selected galaxies.

2.1 JWST/NIRCam and the CEERS + PRIMER Survey

The James Webb Space Telescope is a revolutionary infrared space observatory, built with the aim of observing the universe between $0.6\ \mu\text{m}$ to $28.3\ \mu\text{m}$. JWST's unprecedented scientific power is a result of both the size of its primary mirror comprising a mosaic of 18 gold-plated, hexagonal mirrors, spanning 6.5 meters in diameter; and the extreme sensitivity and precision of its four scientific instruments mounted in ISIM or the heart of the telescope: Near-Infrared Camera (NIRCam; M. J. Rieke et al. 2023), Webb's primary imager that covers the infrared wavelength range 0.6 to $5\ \mu\text{m}$; Near-Infrared Spectrograph (NIRSpec; Birkmann et al. 2022), a spectrograph observing in the same wavelength range as NIRCam; Mid-Infrared Instrument (MIRI; G. H. Rieke et al. 2015), covering a wavelength range of 5 to $28.3\ \mu\text{m}$; the Near-Infrared Imager and Slitless Spectrograph/Fine Guidance Sensor (NIRISS/FGS; Doyon et al. 2023) observing between 0.6 and $5\ \mu\text{m}$.

For this work, we have used data from the high-resolution NIRCam instrument. As mentioned, it is the primary imager of JWST. NIRCam offers a selection of 29 filters tailored for imaging, categorized as extra-wide, wide, medium, and narrow

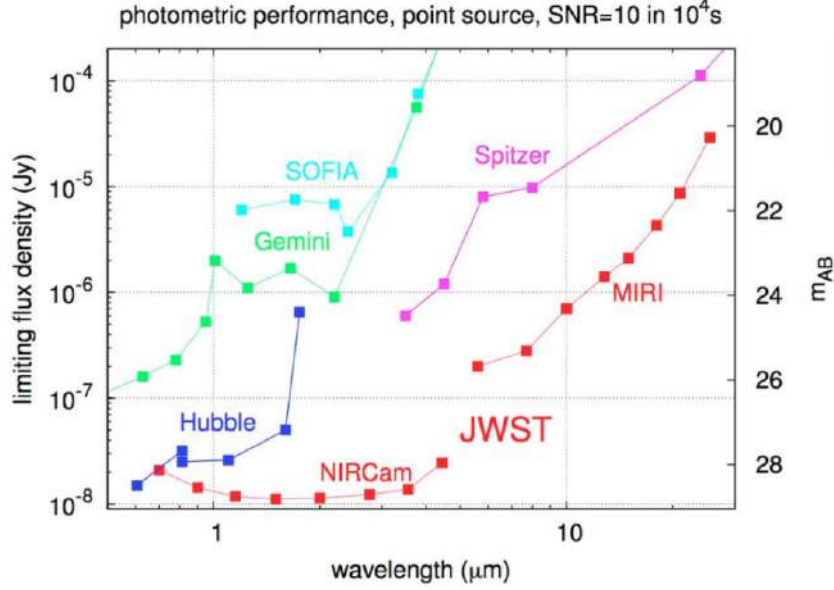


Figure 5: Limiting flux density (Jy) versus wavelength (μm) of several instruments. The limiting flux density is the minimum flux a point source can have to be detected with a signal-to-noise ratio (SNR) of 10 in a 10,000 sec integration time with JWST. This image highlights the comparison of the sensitivity of JWST in the 1- 5 μm range with HST and Spitzer. The unprecedented sensitivity of NIRCam as an infrared instrument is evident here. Credit: STScI/NIRCam Imaging

based on the wavelength range they allow to pass through them. Each wavelength channel utilizes one short and one long-wavelength filter, enabling simultaneous observation through the use of a dichroic. This division splits the wavelength range into short (0.6 - 2.3 μm) and long (2.4 - 5.0 μm) channels, facilitating simultaneous viewing with both modules. NIRCam coverage overlaps with that of HST/WFC3 and Spitzer/IRAC imagers. Figure 5 compares the sensitivity of JWST with other well-known telescopes such as HST, Spitzer, Gemini (an optical/infrared observatory in Chile), and SOFIA (an airborne infrared observatory). It shows that JWST outperforms HST and Spitzer in the 1 - 5 μm range. In addition to its unparalleled sensitivity, JWST's resolving power distinguished it from other telescopes. Figure 6 demonstrates the unprecedented resolving power of JWST/NIRCam in comparison to Spitzer/IRAC over the same patch of the sky.

The JWST imaging and data for photometric catalogues used for this work have been obtained by several surveys: (1) the Cosmic Evolution Early Release Science Survey (CEERS, program ID 2079, PI Finkelstein, Bagley et al. 2023; Finkelstein, Leung, et al. 2023) (2) by the Public Release Imaging for Extragalactic Research (PRIMER, program ID 1837, PI Dunlop, Dunlop et al. in prep.) in the UDS and COSMOS fields which partially overlaps with COSMOS-Web (program ID 1727, PIs Kartaltepe & Casey, Casey, Kartaltepe, et al. 2023). The total effective survey areas for each of the three fields are as follows: 82.88 arcmin^2 for CEERS, 127.14 arcmin^2 for PRIMER- COSMOS, and 224.49 arcmin^2 for PRIMER-UDS. The 5 sigma depths of all photometric filters from each field are given in Table 1 in the A. The NIRCam data in the CEERS survey comprises of 10 pointings, including the

following filters: F115W, F150W, F200W, F277W, F356W, F410M, and F444W (where W and M indicate a wide or medium band filter). The PRIMER survey comprises deep JWST imaging in 10 bands: F090W, F115W, F150W, F200W, F277W, F356W, F444W and F410M with NIRCcam, and F770W and F1800W with MIRI.

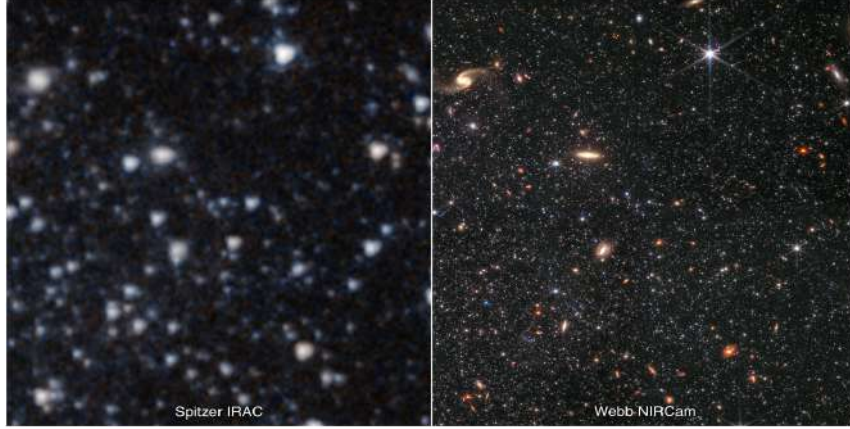


Figure 6: Comparison of image resolution between Spitzer Space Telescope’s Infrared Array Camera (Spitzer/IRAC, left) and the James Webb Space Telescope’s Near Infrared Camera (JWST/NIRCcam, right). The Spitzer image shows $3.6 \mu\text{m}$ light in cyan, $4.5 \mu\text{m}$ light in orange, (IARC1, IARC2). The JWST image shows $0.9 \mu\text{m}$ light in blue, $1.5 \mu\text{m}$ light in cyan, $2.5 \mu\text{m}$ light in yellow and $4.3 \mu\text{m}$ light in red, (filters: F090W, F150W, F250W, F430W). Credit: NASA, ESA, CSA, IPAC, Kristen McQuinn (Rutgers University)

The primary data used in this work is taken from a catalogue compiled using the public JWST imaging from various cycle 1 and 2 programs (Weibel et al. 2024). We focus on the F444W image data, i.e., the image captured via the $4.44 \mu\text{m}$ wide filter on NIRCcam. The JWST data is complemented with available HST imaging data from all the surveys that have covered these fields over the past two decades. The common tool used for photometric measurements is SourceExtractor (Bertin & Arnouts 1996). Starting with `grizli` v7.0 mosaics (Fujimoto et al. 2022), the “weight” files were combined with the exposure maps and science maps to construct the “full” weight maps, including Poisson noise from the flux of sources as well as from the background. The “full” weight maps were then converted to rms images as we know that $\text{rms} = 1/\sqrt{\text{weight}}$. PSFs were extracted directly from the mosaics in each field. The Spectral Energy Distribution (SED) fitting code `eazy` (Gabriel B. Brammer, Dokkum, and Coppi 2008) was used to broadly select galaxies at $z \geq 3$. Further SED fits were done with the Bayesian Analysis of Galaxies for Physical Inference and Parameter ESTimation tool (BAGPIPES; Carnall+18) to obtain the final redshifts and physical properties for the selected galaxies. A delayed- τ model has been used for the star formation history (SFH) with broad uniform priors in age, the time of the beginning of star formation, from 0.01 to 5 Gyr, and uniform priors in the logarithm of τ for all the galaxies in the catalogue. `BPASS – v2.2` stellar population models (Stanway and Eldridge 2018) assuming a broken power-law IMF with slopes of $\alpha_1 = -1.3$ from 0.1 to $0.5 M_\odot$ and $\alpha_2 = -2.35$ from 0.5 to $300 M_\odot$ as shown in Eldridge et al. 2017, taken from Kroupa, Tout, and Gilmore 1993, have been used. For the dust attenuation, a Calzetti dust attenuation curve with a

uniform prior on the dust extinction parameter $A_v \in (0, 5)$ and a uniform prior on the stellar metallicity $Z \in (0.1, 1)Z_\odot$ has been used (refer to Weibel et al. 2024 for more details).

The final physical properties like stellar masses, redshifts, A_v used in this work are from BAGPIPES.

The F444W band was found to be optimal for our science case because our goal is to unveil and study galaxies that become increasingly faint at shorter wavelengths. The F444W band comprises the longest wavelengths of the JWST/NIRCam bands. Moreover, since we aim to study the stellar morphology of the selected galaxies, we must use the flux in rest-frame optical/ near-infrared (NIR) wavelengths. We intend to probe redshifts beyond 3. Knowing that most of the sources selected lie between $z \gtrsim 3$ to 10, $4.44 \mu\text{m}$ corresponds to a rest-frame wavelength range of $1.11 \mu\text{m}$ (rest-frame NIR) down to 493 nm (rest-frame optical). Thus the F444W NIRCam band allows us to probe the stellar morphology using rest-frame optical light.

To begin our selection process, we first select our “parent” sample. The parent sample caters to a broad selection of all star forming Galaxies (SFGs) at $z \geq 3$ in our field images. It is selected from all the fields using the following criteria:

$$\begin{aligned} \text{SNR}(444) &> 15 \\ \text{flag}(150) &= 0, \text{flag}(444) = 0 \\ \text{flux}(444) &> 0, \text{flux}(150) > 0 \\ \text{flux}_{\text{error}}(444) &> 0, \text{flux}_{\text{error}}(150) > 0 \\ \text{junk-flag} &= 0 \end{aligned}$$

The flags and junk flags are used to mask sources of contamination like bright saturated stars, diffraction spikes, hot pixels, and so on.

After using these selection criteria we find that for the parent sample we obtain **2949** sources from the PRIMER-COSMOS field, **3792** sources from the PRIMER-UDS field, and **2950** sources from the CEERS field making a total of **9691** sources. Note that there are a total of **9426** unique sources owing to some overlap among the different fields. This sample is then further filtered for the selection of OFGs and LRDs.

2.2 Selection of Optically Dark/ Faint Galaxies at $z \geq 3$

The most prevalent way of selecting specific kinds of galaxies is to calculate the difference in magnitude between two photometric bands, also known as *colour*. We can define the relation between AB magnitude and observed flux as follows :

$$\text{mag}_{\text{AB}} = 23.9 - 2.5 \log_{10}(\text{flux in nJy}/1000) \quad (1)$$

This method has been extensively used in several studies over the last decade to identify high redshift sources. Colour selections are typically used to select galaxies with reddened stellar continuum emission (Caputi+12; Wang+16; Barrufet+23; Pérez-González+23; Xiao+23a). For our study, we have used colour cuts to select not only optically dark galaxies but also those which appear faint in the bluer bands.

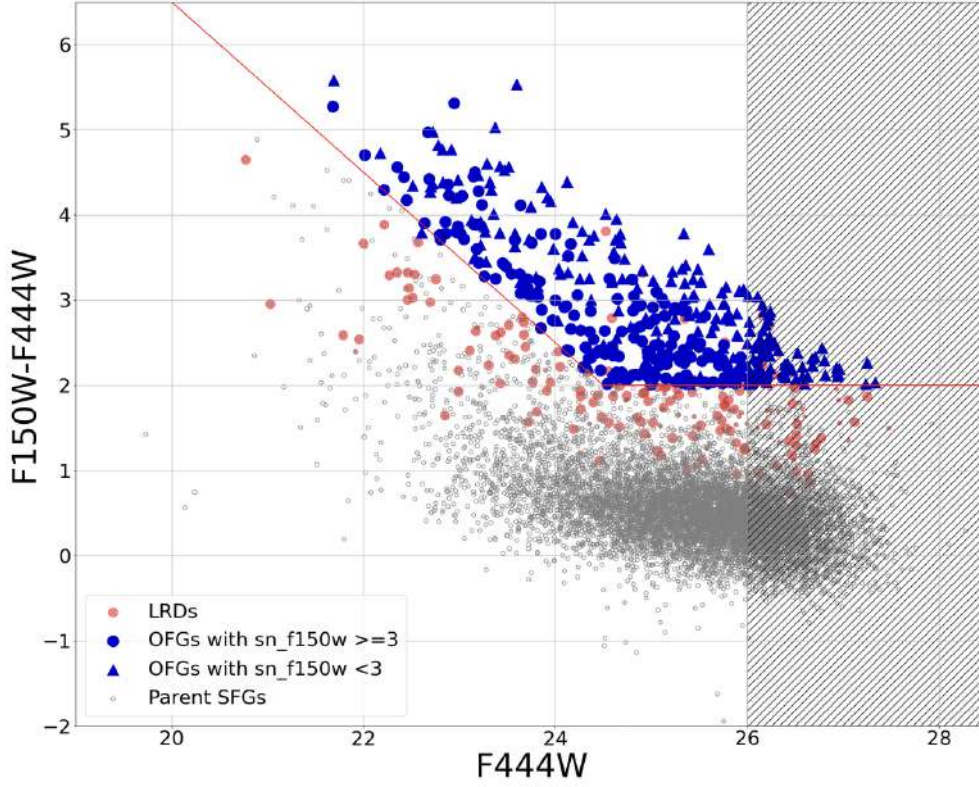


Figure 7: The summarised sample selection from the CEERS, PRIMER-COSMOS and PRIMER-UDS fields. The sources are visualised based on their “redness” (defined by the colour calculated from the difference flux observed at wavelengths of $1.50 \mu\text{m}$ and $4.44 \mu\text{m}$) versus their brightness observed in the F444W band. The grey points indicate the entire parent sample of SFGs at $z \geq 3$. The blue circles indicate OFGs very well detected in both F150W and F444W bands while the blue triangles indicate sources which have $\text{SNR}(150) < 3$ thus their 2-sigma upper limits on magnitude are used as the flux at $1.50 \mu\text{m}$. We also show the selection of LRDs (red circles) based on Labbé+23b. The shaded area denotes the AB magnitude limit of 26 in F444W band used for the final selection of galaxies based on results from simulation tests described in section 4. Thus any source fainter than F444W magnitude of 26 is not considered for morphological analysis.

Thus the OFGs are selected as follows :

$$\begin{aligned} & \mathbf{F150W > 26.5\text{mag}} \\ & \mathbf{SNR(F150W) > 3} \\ & \mathbf{F150W - F444W > 2\text{mag}} \end{aligned}$$

If $\text{SNR}(150) < 3$, the 2-sigma upper limit is used as the flux in F150W.

LRDs, although not the main focus of the thesis, are also selected from the parent sample. We select LRDs to understand their nature from the perspective of stellar morphology. The nature of LRDs is an open question in astronomy today. As mentioned in section ??, investigations have shown that some selected LRDs are AGNs. Thus, we can use morphology to check the prevalence of AGN candidates in the sample. AGN candidates would most likely have a high flux concentration at the source’s centre. Furthermore, since LRDs are also selected based on their “redness” it would be interesting to see if there are any overlapping sources with OFGs which

may lead to the contamination of the latter sample.

So, LRDs have been selected using colour cuts inspired by Labbé+23b.

$$\begin{aligned}
& \mathbf{F444W < 27.7mag} \\
& \mathbf{red1 \& red2 criteria taken from Labbe+23b:} \\
& \mathbf{red1 = (F115W - F150W < 0.8)} \\
& \mathbf{(F200W - F277W > 0.7)} \\
& \mathbf{(F200W - F356W > 1.0)} \\
& \mathbf{red2 = (F150W - F200W < 0.8)} \\
& \mathbf{(F277W - F356W > 0.7)} \\
& \mathbf{(F277W - F444W > 1.0)} \\
& \mathbf{with SNR (FxxxW) \geq 2}
\end{aligned}$$

Based on these criteria we gather a total of **397 OFGs** and **228 LRDs** from the Parent sample. A clear summary of the final sample section is shown in Figure 7.

Further, based on the recent studies in Greene+24, there may exist contamination of brown dwarfs in our sample defined by the following criteria :

$$\begin{aligned}
& \mathbf{F444W < 27.7mag} \\
& \mathbf{115 - 200 < -0.5} \\
& \mathbf{277 - 444 > 1.0)} \\
& \mathbf{with SNR (xxx) \geq 3}
\end{aligned}$$

Note that an AB magnitude limit of 26 in the F444W band is set for the final selection of galaxies based on results from simulation tests described in section 4. Thus any source fainter than F444W magnitude of 26 is not considered for morphological analysis. Moreover, the contamination of brown dwarfs will not affect our final results as we have removed the “PSF-dominated” sources, as later explained in section 5.

3 Fitting Galaxy Morphologies

To fit the stellar morphology of celestial objects (galaxies, in our case), we must model the light distribution of selected sources. Analytic functions are often used to achieve this. These functions have free parameters that can be adjusted to best match the observed data. Said parameters typically represent physical properties of the sources, such as luminosity, size, orientation and ellipticity. Playing with these parameters allows us to quantitatively describe the physical nature of galaxies. Most commonly, galaxies are modelled using Sérsic profiles, Gaussian profiles, Moffat profiles or different combinations of the same. Different light profiles, when fit well, allow us to constrain the central concentration, outer extent and possible asymmetries of galaxies, providing valuable information on their dynamics and evolution through cosmic times.

Let’s discuss the most commonly used light profiles used to characterise the structure of galaxies :

3.1 The Sérsic Profile

The Sérsic power-law is one of the most frequently used to study galaxy morphology and has the following functional form :

$$\Sigma(r) = \Sigma_e \exp \left[-\kappa \left(\frac{r}{r_e} \right)^{1/n} - 1 \right]$$

where Σ_e is the pixel surface brightness at the effective radius r_e . The effective radius r_e corresponds to the radius within which half of the total flux of the observed galaxy is contained. The parameter n , often referred to as the Sérsic Index or the concentration parameter plays an important role in defining the shape of the light profile. A large value of n results in a steep profile and an extended wing, while a small n leads to a shallow inner profile and a sharp truncation at larger radii (Figure 8). The dependent variable κ is coupled to n and hence is not a free parameter. κ can be approximated as :

$$\kappa = \frac{2n-1}{3} \quad \text{for } n > 8$$

or

$$\kappa = \frac{2n-1}{3} + \frac{4}{405n} + \frac{46}{25515n^2} + \frac{131}{1148175n^3} - \frac{2194697}{30690717750n^4} \quad \text{for } n > 0.36$$

It can be shown that κ satisfies $\gamma(2n; \kappa) = \frac{1}{2}\Gamma(2n)$, where Γ and γ are respectively the Gamma function and lower incomplete Gamma function. We mostly make use of special cases of the Sérsic profile to categorise different galaxy morphologies. They are as follows :

1. **de Vaucouleurs Profile (Sérsic Index $n=4$ ($\kappa=7.67$))**: The classic de Vaucouleurs profile is widely used to describe the light distribution of galaxy bulges, particularly elliptical galaxies. It is characterized by a steep inner profile, resembling a power-law decline in surface brightness with increasing radius. This profile implies a relatively rapid decrease in brightness as one moves away from the centre of the galaxy. The de Vaucouleurs profile is effective in capturing the dense, concentrated structure of galaxy bulges. The functional form can be written as :

$$\Sigma(r) = \Sigma_e \exp \left[-7.67 \left(\frac{r}{r_e} \right)^{1/4} - 1 \right]$$

2. **Exponential Profile (Sérsic Index $n=1$)**: The exponential profile describes a light distribution that decreases exponentially with increasing radius. It is commonly used to model the light distribution of galactic disks, including spiral galaxies. This profile implies a gradual decrease in surface brightness as one move outwards from the centre of the galaxy. The exponential profile is particularly suitable for representing the extended, disk-like structure of spiral galaxies.

$$\Sigma(r) = \Sigma_0 \exp \left(-\frac{r}{r_e} \right)$$

where Σ_0 is the central surface brightness.

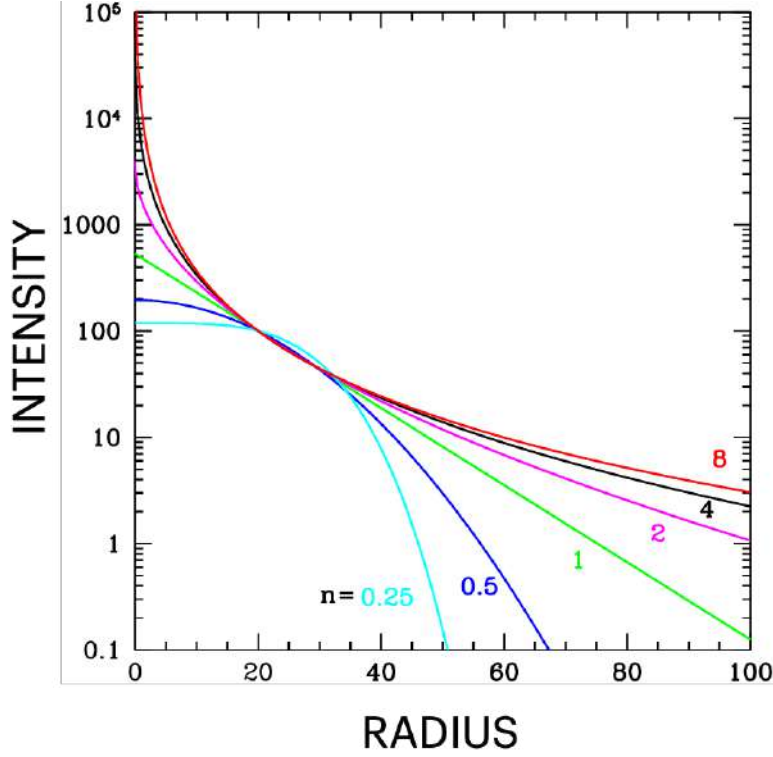


Figure 8: Sérsic profile where effective radius, (r_e) and the pixel surface brightness, Σ_e are fixed. The larger the value of n , the steeper the central slope and the more extended the outer wing of the light profile of a source. A low value of n corresponds to a flatter central core and a more abruptly truncated outer wing. (Peng et al. 2010)

3. **Gaussian Profile (Sérsic Index $n=0.5$):** In the limit where the Sérsic index approaches the value of 0.5, the Sérsic profile converges to a Gaussian function. The Gaussian profile is characterized by a symmetric, bell-shaped curve with a single peak. It is often used to model the light distribution of simple, compact objects or unresolved sources. This profile implies a smooth, symmetric distribution of light around the centre of the object.

$$\Sigma(r) = \Sigma_0 \exp\left(\frac{-r^2}{2\sigma^2}\right)$$

where σ is the standard deviation of the Gaussian, the FWHM is related to σ by, $\text{FWHM} = 2\sqrt{2\ln(2)}\sigma$

Other well-known light profiles include the **Nuker** profile, **Moffat** profile, **Modified Ferrer** profile, **King** profile, etc. However, these functions are not relevant to our work.

3.2 The Point Spread Function(PSF) Profile:

The PSF function is simply the convolution PSF image corresponding to the field in which the galaxies lie, hence there is no prescribed analytical functional form. It could be seen as similar to the Gaussian profile ($n \leq 0.5$).

For our intended morphological analysis, we first attempt to fit pure Sérsic profiles and pure PSF profiles for all the selected galaxies. We then incorporate the use

of a PSF profile alongside the Sérsic profile, resulting in a two-component model. This inclusion is motivated by one of the objectives of our research, which is to identify AGN candidates in our sample and remove their contamination from OFGs. While the dominance of the PSF component in the fitting does not guarantee that these sources are indeed AGN candidates, it provides valuable insights into what to look for during follow-up analysis.

Introducing the PSF profile alongside the Sérsic profile also offers an opportunity to gain insights into the morphological differences between OFGs and LRDs. While colour cuts, as discussed before, provide an initial means of selecting and categorizing these sources, we observe a significant overlap between OFGs and LRDs during this selection process (see Figure 17 in section 5). This overlap implies the necessity of extending our methods to include morphological selection criteria for identifying and distinguishing between these two types of sources.

By implementing morphological selection criteria based on the Sersic + PSF profile, we can discern subtle differences in the morphology of OFGs and LRDs that may not be apparent from colour-based classifications alone. For example, LRDs may exhibit more compact and centrally concentrated structures compared to OFGs, which may display more extended and diffuse distributions of light. Moreover, our sample may also contain contamination from unresolved sources, which further emphasizes the importance of incorporating the PSF profile in our analysis.

3.3 Fitting Algorithms

We make use of two well-known fitting codes to be able to assess the morphology of our sources, **GALFIT** and **pysersic**. This section describes the algorithms employed in the codes, the various user inputs needed and the output formats containing relevant fitting information that can be retrieved from the fitting codes.

3.3.1 GALFIT

GALFIT (Peng et al. 2002, 2010) has been extensively used in literature for fitting two-dimensional surface brightness profiles to galaxies and other astronomical objects. It is a least-squares fitting algorithm of the “non-linear” type. It uses a Levenberg-Marquardt algorithm to find optimal solutions of coefficients to various functions, power-laws, exponents, etc. Similar to all least-squares algorithms, **GALFIT** finds the “goodness” of a fit by determining the χ^2 and computing ways to adjust parameters in the next iteration as long as the χ^2 decreases. The “goodness” of the fit is finally indicated by normalised χ^2 i.e χ_ν^2 and can be expressed as:

$$\chi_\nu^2 = \frac{1}{N_{DOF}} \sum_{x=1}^{nx} \sum_{y=1}^{ny} \frac{(f_{data}(x, y) - f_{model}(x, y))^2}{\sigma(x, y)^2}$$

summed over all nx and ny pixels, where N_{DOF} is the degree of freedom (number of pixels in the area of image being fit - number of free parameters). The user can specify which model to fit which can comprise a combination of various kinds of light profiles, the PSF profile, and also an estimate of the sky, finally generating the model image, f_{model} . As its basic use, **GALFIT** can fit an ellipsoid model to a light profile in a 2-dimensional image. In more complex scenarios, **GALFIT** can

be used to reproduce highly detailed shapes with various characteristics including curvature, asymmetry, truncation, ring structures, spiral arms, etc. These features could be combined in a single model component or integrated with other components to create complex shapes.

3.3.2 Implementing GALFIT

As aforementioned, we exploit the multi-component feature using simple Sérsic and PSF profiles in this work. The resultant parameters from **GALFIT** with a Sérsic only profile comprise the centroid coordinates, total AB magnitude, effective radius, Sérsic index (n), axis ratio ($q = b/a$, b : semi-major axis, a : semi-minor axis), and position angle (PA) of the fitted object. The parameters retrieved using the pure PSF fit are of course only the total AB magnitude and centroid coordinates. The dual PSF+ Sérsic fit provides separate central positions and AB magnitudes for both the PSF and Sérsic components. Additionally, it returns unique parameters specific to the Sérsic model. The key difference between the pure Sérsic model and the PSF+ Sérsic decomposition lies in their scope: the former estimates Sérsic parameters for the entire object, whereas the latter computes them solely for the Sérsic component, thus accounting for the possible prevalence of a very compact component at the centre of the source whose information is stored within the PSF component of the model.

To find the contribution of the PSF and Sérsic components, we compare the magnitudes distributed between the two models in the resultant fit.

In order to run a successful fit, **GALFIT** requires a range of inputs. It needs the user to provide it with the primary image file containing the source to be analysed, an output file path where the output image block shall be saved, a PSF image, a sigma image or rms map representing the noise properties of each pixel in the input image, a mask to avoid fitting surrounding sources if any, model components and the initial guess for the parameters of the model components which must be fitted on the source. All this information is stored in an input `feedme` file. An example is shown below. For our work, the initial guesses for **GALFIT** fitting are based on the physical properties of the sources as derived using `SourceExtractor`.

```
=====
A) input.fits          # Input data image (FITS file)
B) output.fits         # Output data image block
C) sigma.fits          # Sigma image name (made from data if blank or "none")
D) psf.fits            # Input PSF image and (optional) diffusion kernel
E) 1                   # PSF fine sampling factor relative to data
F) mask.fits           # Bad pixel mask (FITS image or ASCII coord list)
G) none                # File with parameter constraints (ASCII file)
H) 1 100 1 100         # Image region to fit (xmin xmax ymin ymax)
I) 100 100             # Size of the convolution box (x y)
J) 28.9                # Magnitude photometric zeropoint
K) 0.004 0.004         # Plate scale (dx dy) [arcsec per pixel]
O) regular             # Display type (regular, curses, both)
P) 0                   # Choose: 0=optimize, 1=model, 2=imgblock, 3=subcomps
=====
# Object number: 1
```

```

0) sersic                # Object type
1) 50.0 50.0 1 1         # Position x, y
3) 26.5 1                # Integrated magnitude
4) 10.0 1                # R_e (effective radius) [pix]
5) 1.5 1                 # Sérsic index n (de Vaucouleurs n=4)
6) 0.8 1                 # Axis ratio (b/a)
7) 45.0 1                # Position angle (PA) [deg: Up=0, Left=90]
Z) 0                     # Skip this model in output image? (yes=1, no=0)
# Object number: 2
0) psf                   # Object type
1) 50 50 1 1             # Position x, y
3) 26.5 1                # Total magnitude
Z) 0                     # Skip this model in output image? (yes=1, no=0)

```

It is always beneficial to have an approximate idea about the physical parameters of the source we want to fit. To relay this information to the fitting code we must provide it with a “constraints” file. This file encapsulates the range of values that the fitting code can explore for each parameter. For our work, we give “soft” constraints to all our parameters to avoid restricting the fitting code to a conservative parameter space. Here is a list of constraints that we have used in our work,

1	n	0.5 to 5
1	re	1.5 to 100
1	q	0.01 to 1
1	mag	0 to 40
1_2	x	offset
1_2	y	offset

All the information from the best-fit results in **GALFIT** is stored in a multi-extension image cube, whose **HEADER** provides values of the best-fit model parameters, as well as a **fit.log** file summarising χ^2 or the goodness of fit for a particular run along with the best-fit parameter values. The multi-extension cube stores the input image, best-fit model for the image and the residuals (Figure 9).

3.4 pysersic

pysersic (Imad Pasha & Tim B. Miller, 2023) is a recently developed Python package which uses Bayesian statistics to find best fits for Sérsic and other light profiles on astronomical images. It uses a forward modelling approach in which parametric models are fit directly to images while accounting for the effect of the PSF.

3.4.1 Forward Modelling, Bayesian Statistics & MCMC Sampling

To accurately constrain the parameters of a model we need to have a good understanding of the physical phenomenon under scrutiny. An accurate model implies two things, the model fits the data well and the parameter of our model has a physical significance and feasibility. The approach in which we fit our model to the observed

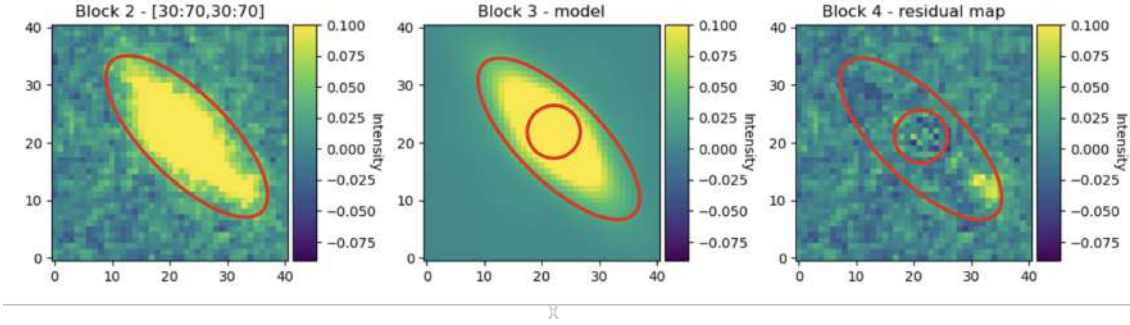


Figure 9: The image shows three main blocks of the multi-extension image cube that **GALFIT** produces as an outfit: the input image, model and residual going from left to right. The smaller red circle highlights the PSF component in the Sérsic+PSF modelling for this given source while the larger ellipse encapsulates the Sérsic component of the same. For a bright resolved source as shown, **GALFIT** is a very useful tool.

data is called forward modelling. In other words, we sample a large number of model parameters, input them into the model and compare the outcomes or predictions to observed data to determine the best set of parameters.

Forward modelling is intricately linked with the Monte Carlo Markov Chain (MCMC) sampling technique. The key idea is to begin with a point estimate and then explore the parameter space around that point. The reason for exploring around the point estimate is that the point estimate is not inherently more significant than the neighbouring parameters in the “parameter space”. The point estimate is just one value within a range of possible values. Neighbouring values could also be likely, given the data, especially in the presence of uncertainty and variability in measurements. Moreover, in many models, parameters are interdependent. Exploring around the parameter space allows us to understand these inter-dependencies and how they affect the model predictions. In Bayesian statistics, the entire posterior distribution of the parameters is considered rather than just a point estimate. This distribution reflects the uncertainty about the parameters, and exploring around the point estimate gives a more complete picture of this uncertainty.

In the Bayesian approach, we begin with a set of “priors” which indicate prior knowledge of the parameters of the model. These priors can be informative or non-informative. The former allows us to constrain the distribution of our parameters to a fixed space while the latter allows more freedom for the model to explore given that we do not possess any prior knowledge of its parameters, thus providing a more “relaxed” prior distribution. For instance, in our light profile models, we know that having a negative value of radii and sérsic index is physically impossible, thus we do not allow the model to explore negative values of these parameters while still allowing a “relaxed” range of positive values for the same.

3.4.2 Setting Priors in pysersic:

One of the most convenient features of **pysersic** is its flexibility in the way of providing priors to the model. We first initialise the **pysersic.FitSingle** object to fit a single source. We then have the option to use the **autoprior** function that measures the observed properties of the source using the **photutils** package.

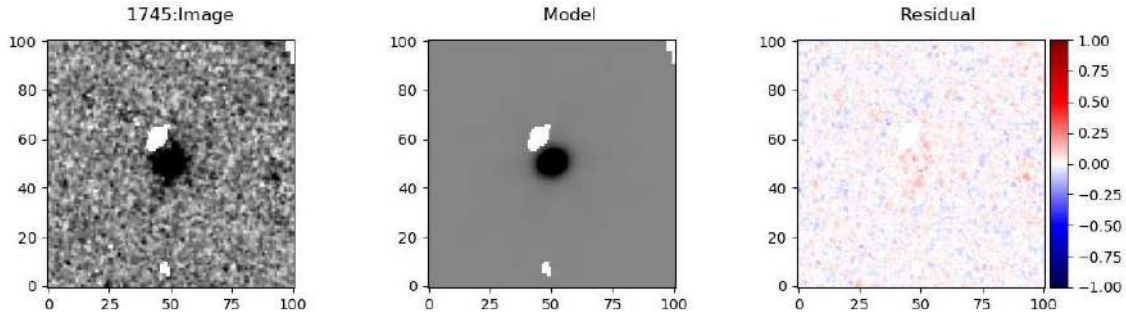


Figure 10: An example of the best-fit model constructed using Point Estimate values for a given source using **pysersic**. From left to right we see the input image, model and residual.

However, to have more control over the priors we employ the **custom_prior** object as shown :

```
custom_prior = PySersicSourcePrior(profile_type='sersic',
sky_type= 'flat', sky_guess=sky_med, sky_guess_err= 2*sky_med_unc)
```

We can then use informative or weakly informative priors depending on our science goals. For our case, we use informative, Gaussian priors for the position and flux of the source, and weakly informative Gaussian priors, i.e, a Gaussian with a large Full Width Half Maximum (FWHM) for the physical parameters that we want the model to predict, i.e., the sérsic index and effective radius and finally uniform priors on ellipticity, PSF-fraction and so on. The peak of the Gaussian for all the physical parameters is based on the initial point estimate of let's say 100 sources. Details about the priors shall be explained in section 4. The input required by **pysersic** are similar to **GALFIT**. It needs the user to provide the input image, rms map corresponding to the input image, PSF and a mask to account for contamination around the main object in the image. The outputs provided by **pysersic** can be visualised as needed by the user. We can see how the Point Estimate looks like using residual plot visualisation (Figure 10). For our work, we utilise the 16th, 50th and 84th percentiles produced for each physical parameter which corresponds to their medians and the 1σ uncertainty around these median values. We can also see graphical representations of the best-fit model in the form of a corner plot (Figure 11).

4 Running Tests

Now that we have an idea about the functioning of the fitting codes, we try to understand their behaviour and performance. Thus, before proceeding to fitting light profiles to our real galaxies, we must run a few preliminary tests on the codes. This step is essential as, to the user, the code in use may appear like a “Black Box” meaning that the inner workings are not readily apparent. By running tests or simulations, the user can gain more insight into the behaviour of the code like the specific input requirements, input and output formats, and potential sources of error. These tests help in understanding the strengths and limitations of the code. The user can assess the software's performance under different conditions,

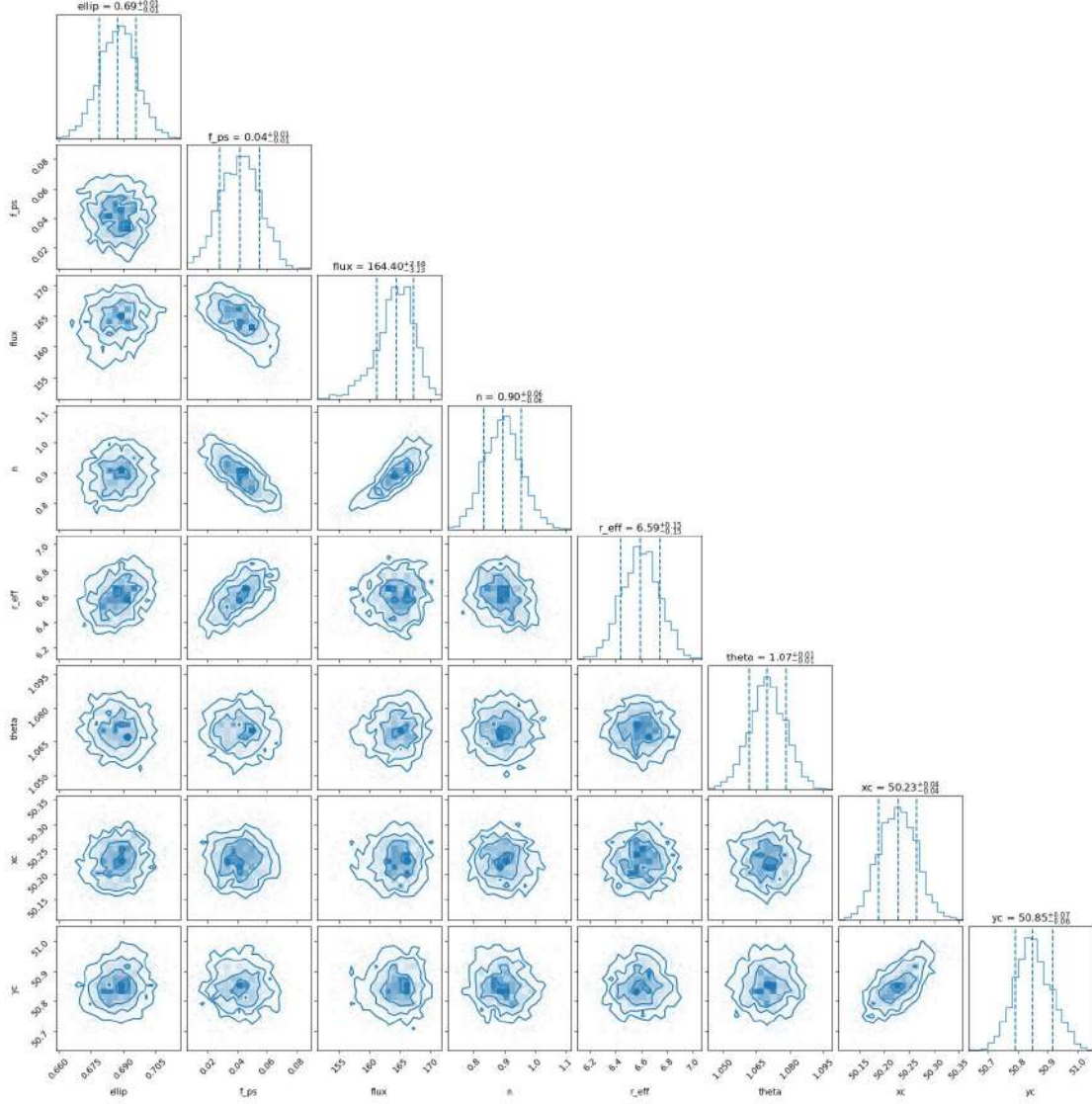


Figure 11: Here we see a corner plot visualisation for one particular fit. The histograms show the marginal distributions of each parameter, giving an idea of their spread and central tendency. The off-diagonal contour plots show the correlation between a pair of parameters. These contours also represent regions of higher density, indicating the likelihood of parameter combinations.

such as varying signal-to-noise ratios or background levels, intensities, object sizes and morphologies. We can use different parameter settings, and sampling methods and explore various modules of the fitting codes to optimise its implementation on real images and maximise the quality of the results. Understanding the dependencies and computational requirements of the code is also essential to have an idea of the time and computational resources needed for the desired task to be completed with ease.

4.0.1 Creating fake sources

The key goal of our project is to be able to determine the size and also PSF fraction of our sources by seeing the percentage of contribution to the intensity of our galaxies

from the Sérsic and PSF components. To see how robust **GALFIT** and **pysersic** are in estimating the PSF Fractions and sizes we must first create simulated objects with varying PSF fractions and fluxes (or intensities) and check if the predictions from the code can reproduce the parameters which were used to create these objects in the first place.

To do so we try to produce a wide range of objects which would mimic the selected sample of our real galaxies. For this, we need to know the range of fluxes or AB magnitudes prevalent in the real sample. The basic idea is to create fake sources with two model components, a Sersic component created using the Sersic2D modelling class in the **Astropy** package and a PSF component. From previous studies of OFGs (Carlos Gómez-Guijarro et al. 2023, Pérez-González et al. 2023), we have an estimate of the expected effective radius and sérsic indices of the galaxies. To create the simulated sources, they were assigned an effective radius of their sérsic component to 10 pixels for the first test and brought down to 3 pixels for the subsequent tests. For the sérsic index, we assign a fixed value of 1.5 to all the sources. Of course, this needs to be convolved with a PSF taken from the Observed PSF in the PRIMER-UDS field. We then use the same Observed PSF to act as the PSF component in the model.

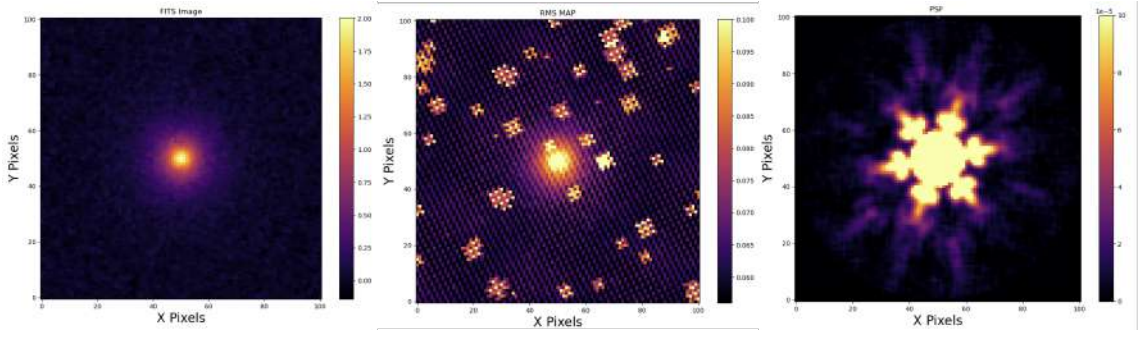
We note that the brightest selected objects lie at a magnitude of 22 while the faintest lie at a value of 28 in the F444W band. So, we have 60 different magnitudes within this range with an increment of 0.1.

Since the PSF fraction will affect the final modelled image we divide the total flux assigned for our source based on this fraction. The PSF fraction can be calculated as :

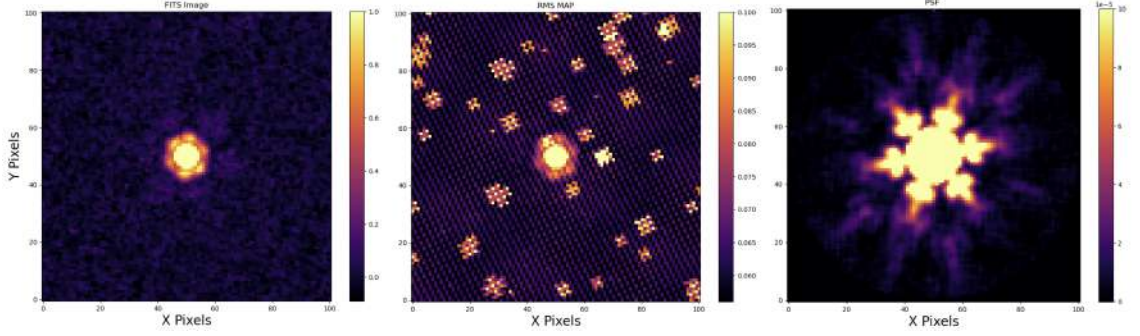
$$\text{PSF fraction} = \frac{\text{Flux}_{\text{PSF-model}}}{\text{Flux}_{\text{PSF-model}} + \text{Flux}_{\text{Sersic-model}}} \quad (2)$$

Thus we select 10 different fractions from 0 to 1.

Other parameters of these galaxies are less relevant to our work. These parameters are intended to characterise the elliptical shape of the galaxies. They are namely, the axis ratio (b/a), ellipticity ($1 - b/a$) and θ or position angle (PA), the position angle of the a axis relative to the first image axis. It is counted positive in the direction of the second axis. The axis ratio and ellipticity of a galaxy are used to describe its orientation as seen from Earth. So, an axis ratio of 0 or ellipticity of 1 means that a galaxy is “edge-on” and hence we see a highly elongated shape of the disk-edge. Conversely, if the axis ratio is 1 (ellipticity is 0), the galaxy appears circular and is “face-on”. By visual inspection of our selected OFGs and LRDs, we note that a majority of them appear circular with small sizes making it imperative that we try to test the limits of the codes by testing them on small circular sources. Thus for the ellipticities and axis ratios, we assign values of 0 and 1 respectively while for position angles, the value is set to 0. Hence we consider “face-on” orientation with no inclination with respect to the image axis, for all the simulated sources.



(a) SOURCE WITH PSF FRACTION = 0



(b) SOURCE WITH PSF FRACTION = 1

Figure 12: Example of simulated sources created for tests on **GALFIT** and **pyser-sic**. From left to right on both panels we see the fake source created, the RMS map with Poisson noise for the given source and the PSF from the PRIMER-UDS field which is used both for model creation and fitting, respectively. Here, we see the extreme cases with PSF fractions being 0 and 1, but we created fake objects with 10 different PSF fractions ranging between the extremes.

Using this setup, the resultant simulated source can be constructed as :

$$\text{Simulated Source} = \text{PSF fraction} \times \text{model}_{\text{PSF}} + (1 - \text{PSF fraction}) \times \text{model}_{\text{Sérsic}} \quad (3)$$

Note that the PSF model and Sérsic model were normalised to 1 before the implementation of equation 3.

Next, we want to be able to make these simulated sources comparable to the real images, so we use an empty sky cutout of 101×101 pixels taken from the PRIMER-UDS Field image. The empty sky is found conveniently using the segmentation maps to avoid any contaminants (neighbouring sources or fluctuations). The simulated sources are then injected one at a time into this background cutout to finally produce a 101×101 -pixel image with the simulated source at the centre with the pixel coordinates of (50, 50). We create unique RMS maps for each simulated source. The RMS cutout is created at the same RA, DEC positions as the background cutout using the exposure maps and introduces Poisson noise based on the flux of the simulated sources. This process allows us to bring our simulated images closer to reality.

As a test sample, we finally had 600 sources with varying combinations of fluxes and PSF fractions ready to test the fitting codes in hand. Examples of such simulated sources as shown in Figure 12.

4.1 Results from Tests

In this section, we discuss the various deductions made from the tests on simulated sources using **GALFIT** and **pysersic**. At the end of this section, we decide which code is more suited for our purpose and explain the set-up used for the final fitting process.

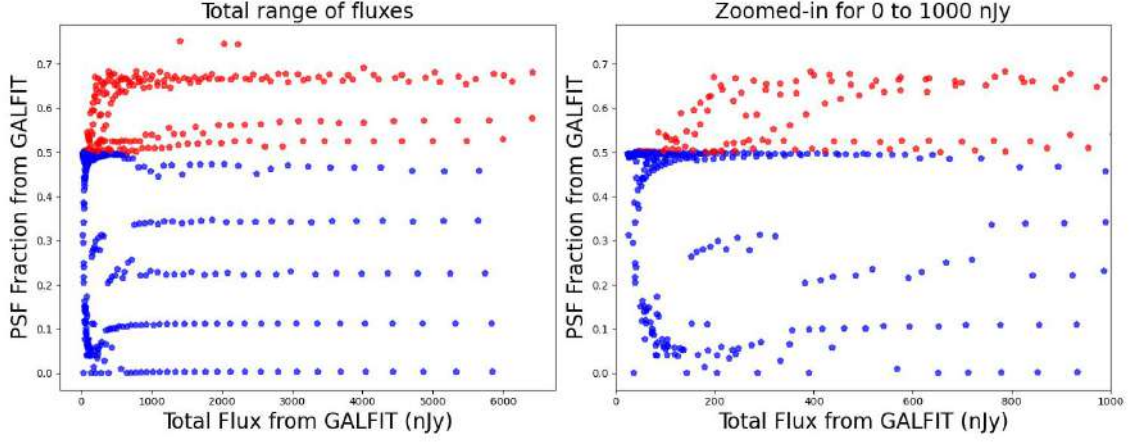
4.1.1 GALFIT or pysersic ?

We ran tests on both **GALFIT** and **pysersic** to check which one, if not both, is more robust in serving our scientific purpose.

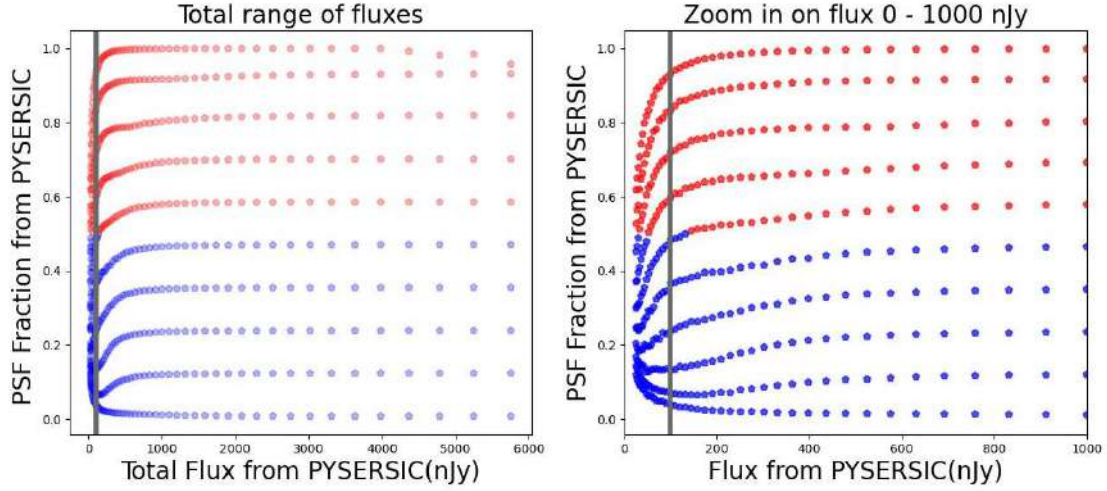
Let's check the parameter specifications we choose for the test. The fake sources for this first test were created with an effective radius of the Sérsic component being 10 pixels, along with the PSF from the PRIMER-UDS field for the PSF component of the galaxy model. The contribution of the components was based on the PSF fraction assigned to each fake source. The effective radius is one of the most important parameters that we want to assess. For the NIRCам instrument's F444W band filter, the Empirical PSF FWHM is approximately 0.14 arcseconds. Given that our images have a pixel-to-arcsecond scale of 0.04 arcsecond per pixel, we get an empirical PSF FWHM of about 3.5 pixels. In a 2D Gaussian profile, the radius that encompasses 50% of the light, or the effective radius ($R_{e,PSF}$), is half of the FWHM. Therefore, assuming that the PSF is a 2D Gaussian, to estimate the half-light radius or ($R_{e,PSF}$) we can simply use $\frac{1}{2}$ of FWHM as the reference value. Using this relation we find that the $R_{e,PSF}$ comes out to be 1.75 pixels. Knowing this is crucial as it represents the resolution limit for compact sources. Hence, when fitting the effective radius parameter, we should use this value as a reference for the lower limit. This ensures that our fitting process respects the inherent resolution constraints of the instrument making our measurements more reliable. Using this information we set the lower limit of the effective radius parameter to 1.5 pixels, slightly lower than the resolution limit, to not impose an extremely tight constraint on the size.

For the initial comparison, we set weakly informative uniform priors for both fitting codes. For **GALFIT** we allow Sérsic index, n a range of 0.5 to 5, effective radius (pixels) to range from 1.5 to 100 and q to range between 0.01 and 1. The central positions of the PSF and Sérsic components are fixed to move together.

In **pysersic**, we provide the input flux values for the simulated images and allow a small error while all the physical parameters are given the same constraints as in **GALFIT** with an additional parameter of PSF fraction to range between 0 and 1. The simulated sources are first considered to be well-resolved and extended. We provide an effective radius of 10 pixels to the Sérsic component, with the total or actual effective radius to be calculated using Equation 4. It is known that for the simulated sources, for each value of PSF fraction (0 to 1), there are 60 sources ranging from the faintest flux of 23 nJy (AB magnitude, 28) to the brightest flux of a little over 5754 nJy (AB magnitude, 22), giving us a total of 600 uniformly spaced sources to test. We see from Figure 13 that **pysersic** is much better at retrieving this same distribution as mentioned as compared to **GALFIT** which exhibits significant fluctuation in the retrieval total fluxes and PSF fractions. We do not see uniformity with **GALFIT** while the trend is maintained with **pysersic** until very faint magnitudes.



(a) PSF Fraction versus Flux retrieved from **GALFIT**

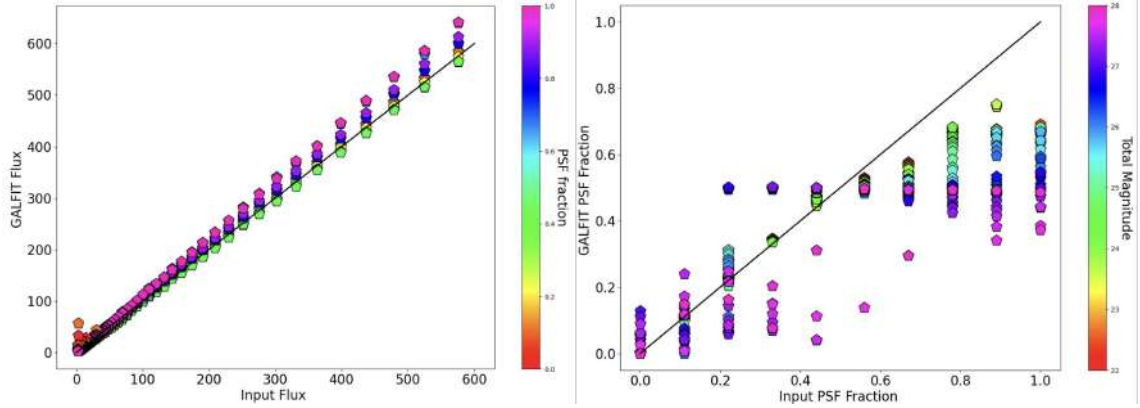


(b) PSF Fraction versus Flux retrieved from **pysersic**

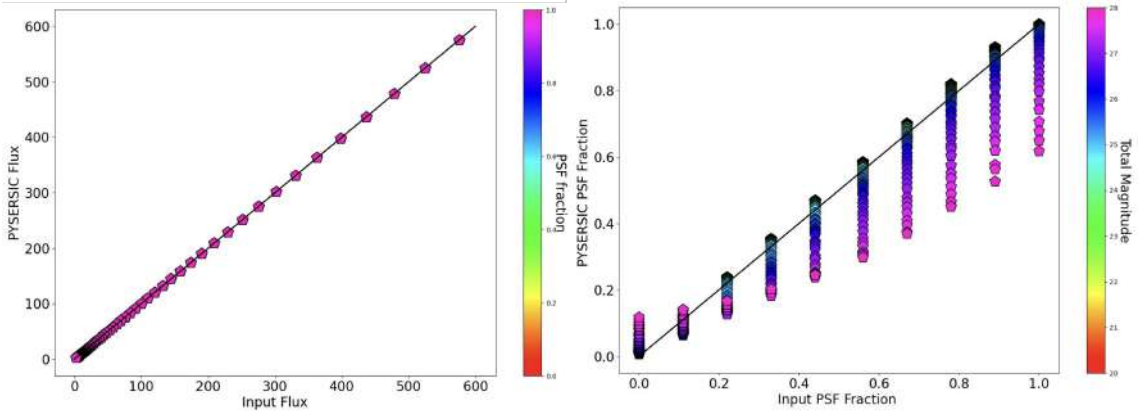
Figure 13: Comparison of retrieval of PSF fraction versus Flux from **GALFIT** and **pysersic** in the first test. In Figure 13a, no consistent trend is seen as compared to that in Figure 13b. For **pysersic** (Figure 13b) the accuracy of results only begins to break down at a magnitude of 26.4 (flux, 100 nJy) as shown by the vertical grey lines, before which the PSF-fractions are within 10% error of the input value. After this, the PSF fraction gets underestimated for highly “PSF-dominated” sources and overestimated for sources which have low input PSF fractions.

When we compare the input and output values for flux and PSF fractions, this difference becomes even clearer. The flux retrieval from **pysersic** is well constrained, while there is noticeable scatter in the results from **GALFIT**. **GALFIT**’s PSF fraction results show no consistent trend or uniformity, whereas **pysersic** provides reasonable and consistent results (Figure 14). For sources with a PSF fraction ~ 0 , it is expected that the fitting code might add a PSF component along with the Sérsic component, resulting in a slightly higher PSF fraction than the input value. The same reasoning applies to sources with a PSF fraction ~ 1 . As sources become fainter, the fitting code reaches its limits, leading to inaccuracies, i.e., PSF fraction

gets underestimated for highly PSF-dominated sources and overestimated for sources which have low input PSF fractions, see Figure 14. In the first test, the effective radius of the fake sources was considered 10 pixels where this limit was reached at 26.4 mag or 100 nJys. The consistency of results with the input values remains within an error of 10% until this flux value. Below this flux, the difference between the input and output PSF fractions begins to be significantly large. For the next test, the effective radius is considered 3 pixels for the fake sources, so there may occur a shift to this limit as we reach small sizes. So, in the following test, we aim to determine where this limit lies.



(a) Input versus Output Flux and PSF Fraction retrieved **GALFIT**



(b) Input versus Output Flux and PSF Fraction retrieved from **pysersic**

Figure 14: Comparison between Input and Output FLux and PSF Fractions yielded by **GALFIT** and **pysersic** from test on simulated sources. The diagonal black lines indicate a 1:1 ratio of X and Y axes. This line is plotted to understand the accuracy of the codes and their deviation from the exact expected values.

From visual inspection, we know that many of our sources are compact and near the resolution limit of the instrument. Also, our sources range from very bright (AB magnitude = 22) to very faint (AB magnitude = 28). Banking on the fact that **pysersic** more accurately retrieves the fluxes and reproduces well-constrained PSF fractions than **GALFIT** suggests that we should proceed with **pysersic** for our science case. **GALFIT** does prove effective for very bright and well-resolved sources, a scenario not entirely reflective of our specific research needs.

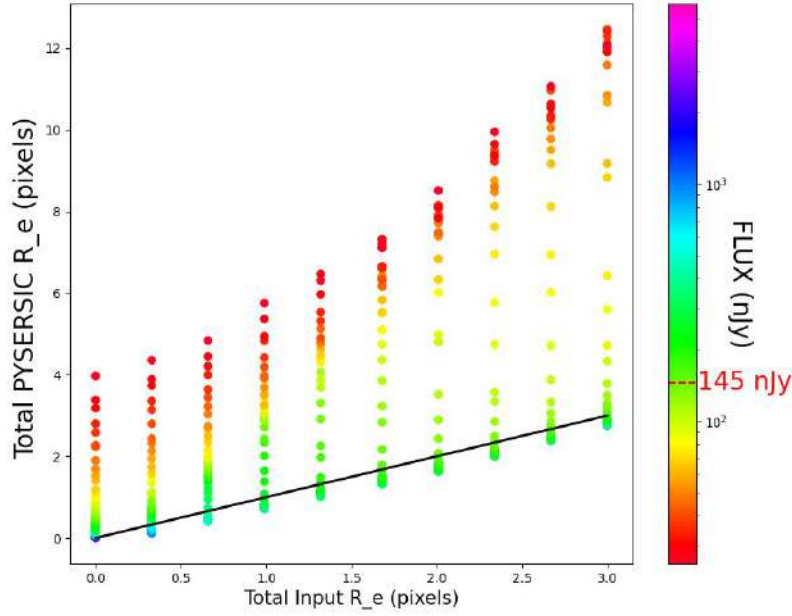


Figure 15: Comparison between Input and Output Values of Weighted Effective Radius. A threshold of 145 nJy denotes the limiting flux considered for selecting the real sources. The black line shows the 1:1 relation with the X and Y axes.

4.2 Testing the limits of pysersic

Now, with **pysersic**, we need to test how small and faint we can go to accurately determine the morphology of the sources. In the first test, we set the simulated sources to have an effective radius of the Sérsic component ($R_{e,\text{Sérsic}}$) to 10 pixels. In this second test, we will focus on sources with an $R_{e,\text{Sérsic}}$ of 3 pixels. This value is more realistic based on the point estimates for our real selection of OFGs.

In Figure 15, the comparison between input and output values of the weighted or total effective radius of tested sources is shown. This weighted effective radius is the “real” effective radius of the source, which is dependent on the PSF Fraction of the source. It can be defined as:

$$\text{Total } R_e = \text{PSF fraction} \times R_{e,\text{PSF}} + (1 - \text{PSF fraction}) \times R_{e,\text{Sérsic}} \quad (4)$$

From this test, it is evident that around 145 nJy, which corresponds to an AB magnitude of 26, the deviation from the expected value of the effective radius becomes significantly pronounced ($> 10\%$ error). Beyond this flux level, the discrepancies increase to a point where the measurements are physically unreliable. In the previous test, we wanted to know if the reduced effective size would affect this flux limit. We see that for smaller simulated sources with an effective radius of 3 pixels, the limit lies at 0.4 mag brighter flux than that for sources with a 10-pixel effective radius.

Therefore, we establish an AB magnitude of 26 as the limiting flux for our morphological analysis. This threshold is crucial for ensuring the reliability of the physical properties derived using **pysersic**. Consequently, this decision directly impacts the initial sample selection, as it filters out sources that are too faint for accurate

analysis. We implement a magnitude cut at 26 in the F444W band for our selection process: any source from the parent, OFG or LRD sample fainter than an AB magnitude of 26 will be excluded from our morphological analysis. By setting this magnitude limit, we ensure that our study focuses only on sources for which **pysersic** provides meaningful morphological parameters. This step is essential for maintaining the reliability of our results, as including fainter sources would lead to biased conclusions about their physical properties. Moreover, implementing weakly informative Gaussian priors has proven to be advantageous for obtaining more accurate constraints on the measurements of physical parameters compared to using uniform priors. By providing a more informed starting point, these priors guide the fitting process towards more plausible solutions, thereby improving the reliability of the parameter estimates for very faint or compact sources. The peaks of the weak Gaussian priors for the final fitting are based on the point estimates of the physical parameters like flux, size, and sérsic index for these sources. The final fitting priors can be seen below,

```
custom_prior = PySersicSourcePrior(profile_type='sersic_pointsource')
custom_prior.set_gaussian_prior('xc', 50., 1)
custom_prior.set_gaussian_prior('yc', 50., 1)
custom_prior.set_gaussian_prior('flux', flux_value, error_flux_value)
custom_prior.set_gaussian_prior('r_eff', 3., 3.)
custom_prior.set_uniform_prior('ellip', 0.01, 1)
custom_prior.set_uniform_prior('theta', 0, 2 * np.pi)
custom_prior.set_gaussian_prior('n', 1.5, 1.)
custom_prior.set_uniform_prior('f_ps', 0.0, 1.0)
```

After having conducted the aforementioned tests, we can move on to implementing **pysersic** on the selection of real galaxies from the OFG, LRD and Parent samples. In the next section, we shall discuss the various insights gained on the physical properties of these galaxies.

5 Results

In this section, we present the various analyses done in this thesis. We first present the preliminary results using the fitting code **pyserisc** which allows us to finalise the sample for morphological analysis. We then present the physical properties of OFGs and deduce trends. We also present the relationship of dust attenuation (A_v mag) of OFGs with said physical properties. Finally, we present results which may indicate that OFGs could be compact progenitors of quiescent galaxies at lower redshifts.

5.1 Preliminary results from **pysersic** analysis

In section 2, we discussed the initial sample selection, which resulted in **9426 Parent SFGs**, **397 OFGs** and **288 LRDs** for morphological analysis. We then applied a cut at an AB magnitude of 26 in F444W band to eliminate unreliable measurements from **pysersic**. This reduced our sample to **6661 Parent SFGs**, **319 OFGs** and **228 LRDs** at redshifts between 3 and 10.

One of the goals of this thesis is to identify possible AGN-dominated sources among

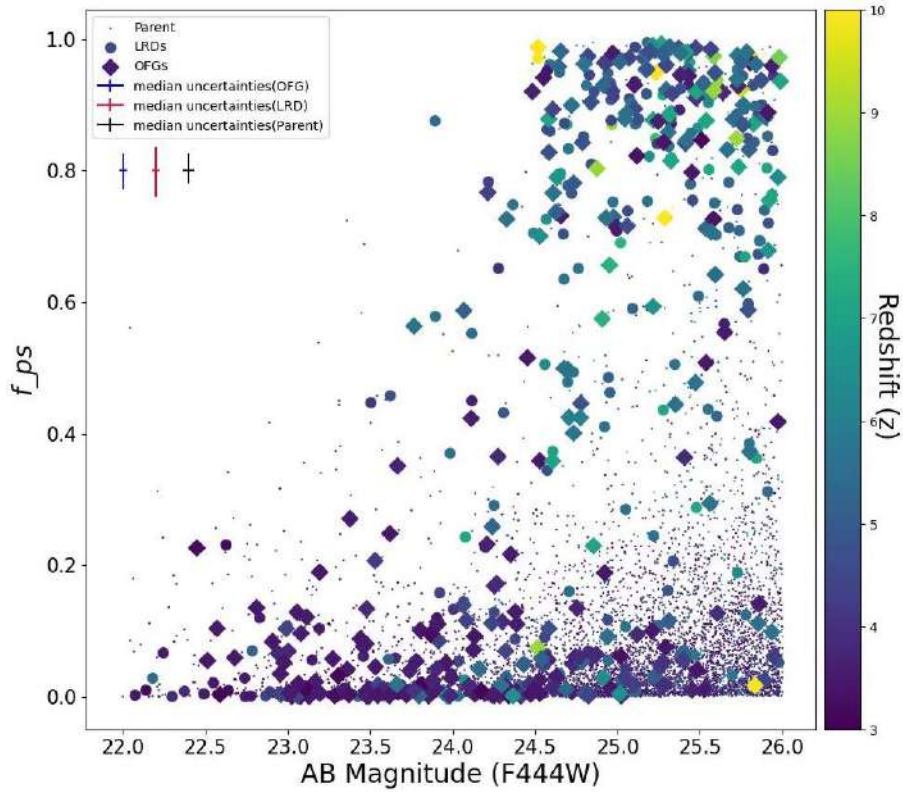


Figure 16: PSF Fractions yielded by **pysersic** versus the AB Magnitude in F444W JWST band of individual sources colour-coded by redshift. The median uncertainties of the different samples are given by errorbars on the top left corner of the plot: blue for OFGs, red for LRDs, and black for parent SFGs.

our OFG and LRD samples. For this purpose, we performed a morphological decomposition of our sources using the Sérsic profile + PSF profile model. Thus, we utilized the “**sersic – pointsource**” model in **pysersic** to determine the PSF fraction, denoted by the “**f_ps**” parameter, which indicates the flux contribution from the PSF component relative to the total flux in the best-fit model for each source. We established a threshold to classify a source as “PSF-dominated” if the flux contribution from the PSF component exceeds **50%** of the total flux. Otherwise, the source is deemed “Sérsic-dominated”. We can see the resultant “**f_ps**” parameter for individual sources in Figure 16. We observe a bimodal distribution of the “**f_ps**” output values for the OFG and LRD sample, while most of the sources in the Parent sample tend to have low “**f_ps**” values between 0 and 0.2. Most sources in the OFG and LRD sample either have a PSF fraction of > 0.75 , i.e., highly “PSF-dominated”, or a low PSF fraction close to 0, i.e., purely “Sérsic-dominated” with a few sources with intermediate PSF-fractions. At brighter magnitudes (below 24.5 mag), most sources appear “Sérsic-dominated” while at fainter magnitudes (above 24.5 mag) more sources seem to be “PSF-dominated”, for these two samples.

This bimodality in the distribution of PSF fraction is not observed in the simulations. This could be due to several factors. The simulations used circular, face-on sources with perfect Sérsic profiles to test the capabilities of **pysersic**. In reality, however, the galaxies do not necessarily exhibit a perfect Sérsic light distribution. Additionally, the sources display varying ellipticity and inclinations and are evidently

compact, not to mention the bias of high PSF fraction towards small, unresolved sources at high redshift. Some sources could also be clumpy due to possible regions of starburst. Thus, in the future, we must test the effects of orientation and inclination on the accuracy of fitting using **pysersic** to better understand the difference between simulations and reality.

Having said that, as this bimodality is not observed in the Parent sample, this gives credit to the fact that LRDs and OFGs indeed possess different morphologies than the parent SFGs.

Based on this “domination” criterion, the final distribution of the sources is illustrated in Figure 17. Out of the **319** OFGs, **193** or **60%** of the sources were identified as “Sérsic-dominated”. In the LRD sample, **111 out of 228** or **48%** of the sources were “Sérsic-dominated”, with an overlap of **32** sources between the two samples. Conversely, **126 out of 319** OFGs and **117 out of 228** LRDs were PSF-dominated, with an overlap of **69** sources. There exists a larger overlap of “PSF-dominated” sources between the OFG and LRD samples than “Sérsic-dominated” sources.

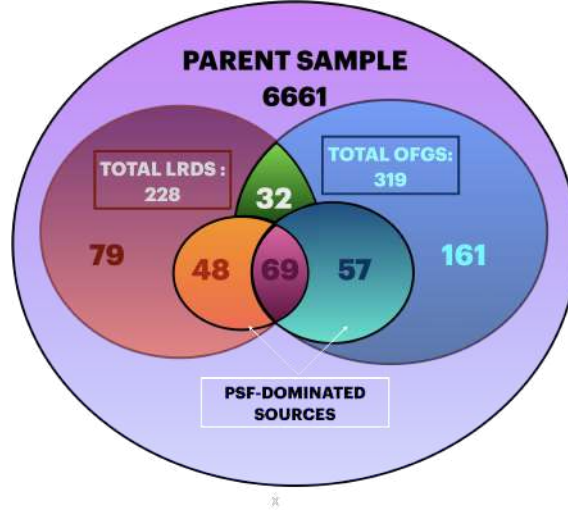


Figure 17: Venn Diagram showing the distribution of galaxies among Parent SFG, OFG and LRD samples. The smaller circles show the number of “PSF-dominated” sources within the selection. There exists an overlap between both “Sérsic-dominated” and “PSF-dominated” sources in the OFG and LRD samples.

We check the evolution of “PSF-dominated” sources with increasing redshift, see Figure 18. For both OFGs and LRDs, we observe a general increase in the number of “PSF-dominated” sources with increasing redshift. This increase could be because many of these sources could be real galaxies which are not resolved at higher redshift as they appear not only fainter but also much smaller (below the resolution limit of the instrument).

Also note that, if a source is considered PSF-dominated, we can confirm that more than 50% of the total flux is contributed by the central part of the source. As stellar light probes the stellar mass within a galaxy, this tells us that there must be a substantial concentration of mass at the centre of the source. For a large proportion of mass to be concentrated at the centre, we could suggest the presence of a very compact central galactic component. Since we probe the observed NIR or rest-frame

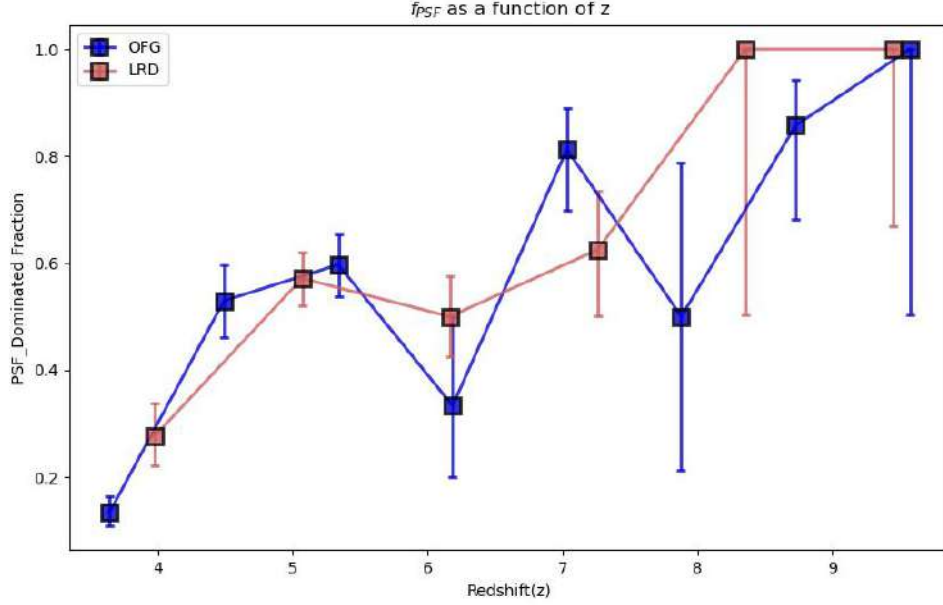


Figure 18: (Redshift evolution of Fraction of “PSF-dominated” sources in OFGs (blue), LRD(red). The error bars are calculated using `binom_conf_interval` in the `astropy.stats` package, where `k` is the number of successes (Number of “PSF-dominated” sources), `n` is the number of trials (total number of sources) with a desired confidence interval of 1σ or 68% confidence. Note that the AB magnitude cut of 26 is applicable here as well.

optical, this flux could be associated with dust-reddened AGN as described in recent literature Morishita et al. 2020; Endsley, Stark, Lyu, et al. 2023.

Knowing this fraction of “PSF-dominated” sources is beneficial as we can conduct spectroscopic follow-ups as well as high-resolution imaging and the modelling of spectral energy distributions (SEDs; e.g. (Endsley, Stark, Fan, et al. 2022; Onoue et al. 2023; Labbe et al. 2023) on these AGN/unresolved candidates to contribute to the census of high-redshift AGN (Carnall et al. 2023; Kocevski et al. 2023; Larson et al. 2023; Maiolino et al. 2023; Harikane et al. 2023).

5.2 Physical Properties of Optically Dark/Faint Galaxies

We can now explore the resultant physical properties of “Sérsic-dominated” Optically dark/Faint Galaxies (OFGs) in comparison to the parent star forming galaxies. We only focus on the sérsic parameters of the “Sersic-dominated” sources because being “PSF-dominated” implies: 1) the source is extremely compact resembling a point source, meaning that it is either unresolved or an AGN candidate or, 2) it is a “PSF-dominated” source but still has a sérsic profile associated to it, meaning a system with an AGN and an extended disk. The former scenario would mean that the resultant values of R_e [kpc], q and n are not reliable, while the latter scenario would mean that we are looking at the physical parameters of only the extended disk and not the source as a whole. Both scenarios would lead to biases in our statistical analysis.

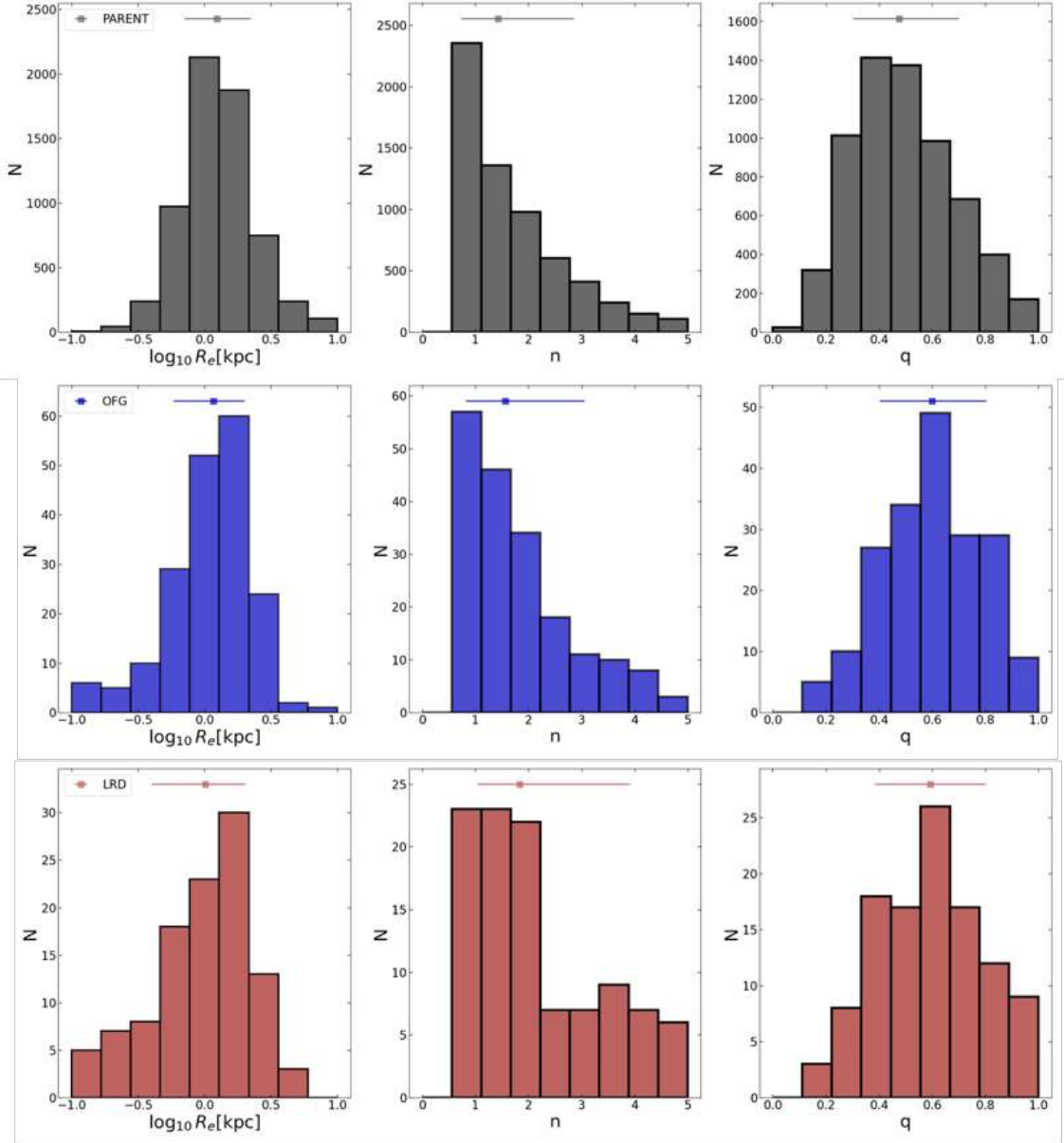


Figure 19: Histogram showing effective radius (left), sérsic index (middle), and axis ratio (right) of the parent SFG (top row, grey), OFG (middle row, blue), and LRD (bottom row, red) samples. The 50th(median), 16th and 84th percentiles of the distributions are shown in the insets above each histogram.

Sample	Effective Radius (R_e [kpc])	Sérsic Index (n)	Axis Ratio (q)
Parent	$1.24^{+0.96}_{-0.54}$	$1.43^{+1.43}_{-0.73}$	$0.48^{+0.22}_{-0.12}$
OFG	$1.16^{+0.84}_{-0.58}$	$1.56^{+1.49}_{-0.74}$	$0.59^{+0.21}_{-0.19}$
LRD	$1.06^{+0.96}_{-0.67}$	$1.83^{+2.07}_{-0.8}$	$0.59^{+0.21}_{-0.21}$

Table 1: Summary of Output Values for physical parameters of the light profile of Sérsic dominated sources in the Parent, OFG and LRD samples.

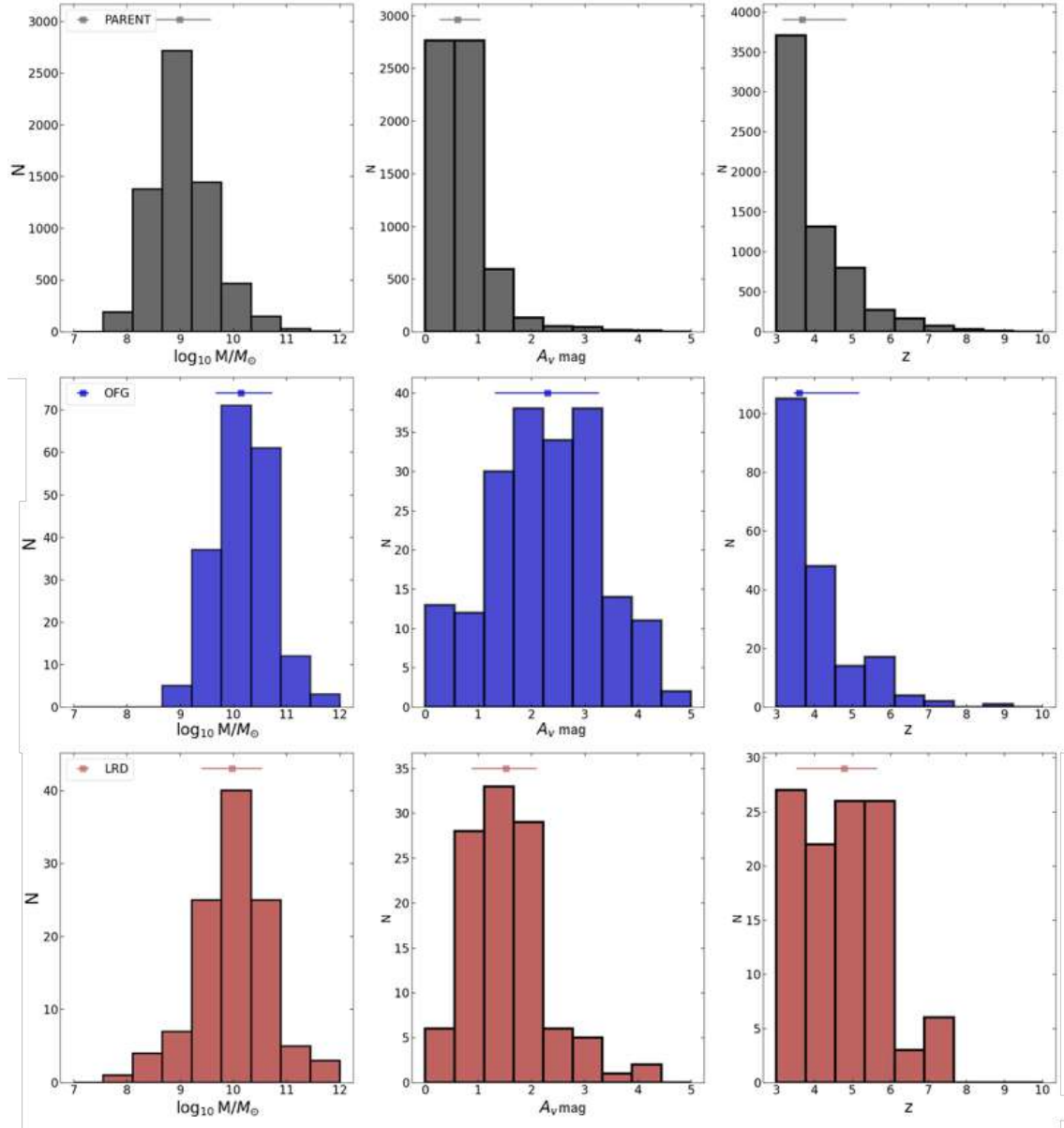


Figure 20: Histogram showing Stellar Mass (left), Dust Attenuation (middle), and Redshift (right) of the parent SFG (top row, grey), OFG (middle row, blue), and LRD (bottom row, red) samples. The 50th(median), 16th and 84th percentiles of the distributions are shown in the insets above each histogram.

Sample	Stellar Mass ($\log_{10} M_*/M_\odot$)	Dust Attenuation (A_V, mag)	Redshift (z)
Parent	$8.99^{+0.59}_{-0.46}$	$0.6^{+0.44}_{-0.32}$	$3.69^{+1.16}_{-0.52}$
OFG	$10.14^{+0.59}_{-0.47}$	$2.30^{+0.96}_{-0.99}$	$3.60^{+1.59}_{-0.15}$
LRD	$9.98^{+0.56}_{-0.58}$	$1.52^{+0.58}_{-0.64}$	$4.78^{+0.87}_{-1.25}$

Table 2: Summary of stellar-based properties of the Sérsic dominated sources in the Parent, OFG and LRD samples

Moreover, as stated in Weibel et al. 2024, since the BAGPIPES SED-fitting does not take into account the contribution from an AGN component, properties like stellar mass and photometric redshift for “PSF dominated” sources would likely be wrongly estimated.

In terms of morphological parameters, in Figure 19, we present R_e [kpc], \mathbf{q} and \mathbf{n} distributions of the “Sérsic dominated” sources in parent SFG, OFG and LRD samples. At first glance, the distributions are rather similar in the different samples. On comparing the exact values, the OFGs have effective radii peaking at 1.16 kpc while Parent SFGs peak at 1.24 kpc, and the LRDs have a peak value at 1.06 kpc. OFGs show a slightly higher \mathbf{n} of 1.56 as compared to the parent SFGs which have 1.43 as the median value. The LRDs here show an even higher \mathbf{n} of ~ 1.83 . On comparing the \mathbf{q} values among the three samples, we see that OFGs and LRDs have higher \mathbf{q} values, in general, i.e. they are more face-on or circular in appearance than the Parent SFG sample. Table 1 summarises the final values obtained from the histogram distribution.

We then look at the stellar-based properties of these galaxies. From Figure 20, we see that the OFGs are much more massive and dust attenuated than the parent sample of SFGs, with moderate to high levels of (A_v mag) between 1.31 and 3.26. The OFGs have a median dust attenuation magnitude of 2.3, a factor of 3.8 higher than the median A_v mag of 0.6 for the parent SFG sample. The parent SFGs and OFGs have a similar redshift distribution with a median at $z \sim 3.6$, with that of LRDs being shifted to slightly higher median redshifts of $z \sim 4.8$. The median stellar mass in the entire OFG sample is at $\log M/M_\odot \sim 10.14 \pm 0.5$ while the median stellar mass for the Parent SFGs lies at $\log M/M_\odot \sim 8.99 \pm 0.5$. The “Sersic-dominated” LRDs shown here exhibit lower median stellar mass and dust attenuation than OFGs. Table 2 summarises the different stellar properties as seen from the histogram distribution in Figure 20.

We then examine the distribution of stellar mass versus their size within our sample. The mass-size relation at a given redshift can be defined by:

$$\log_{10} \left(\frac{R_e}{\text{kpc}} \right) = \alpha \log_{10} \left(\frac{M_*}{M_0} \right) + \log_{10} A \quad (5)$$

where M_* is the stellar mass in solar masses, M_0 is the normalisation parameter, also in solar masses, α is the slope, and $\log_{10} A$ is the intercept or the size of the source at the stellar mass of M_0 . The best-fit result for star forming galaxies at every redshift is compared (grey, Figure 21) with the values from Allen et al.(in prep)(green). This study discusses the mass-size relations of star forming galaxies at $z \sim 3$ to 9 fitted using single component Sérsic models in 7 NIRCcam bands within the same fields as in our work. In general, our mass-size measurement is consistent within the 1σ dispersion of measurements done in Allen et al.(in prep). Subtle differences in the best fit parameters could be because we use only 1 NIRCcam band for our measurements. Additionally, we have a larger sample size and larger range of masses in the parent SFG sample as compared to this work.

In general, we observe that the slope of the mass-size relation increases with an increase in redshift for the star forming parent sample. The best-fit relation of the OFGs lies at the junction between those for the star forming and quiescent galaxies(red). The best-fit parameters for the same are found to be $\alpha = 0.0809 \pm 0.0304$ and $\log A(M_0 = 10^9) = 0.0245 \pm 0.0414$, with an intrinsic scatter of 0.17 for

the distribution of sources at $3 < z < 4$. For higher redshifts ($z > 4$), the best-fit relation for OFGs cannot be found due to a very low number of sources. A summary of the best fit relation for parent star forming galaxy sample at different redshifts is given in Table 4 in section A. As galaxies grow in size and mass over time, this location of OFGs in the mass-size relation could be an indication that OFGs are evolving into quiescent galaxies that we observe at lower redshifts. This topic will be explored further in section 5.4.

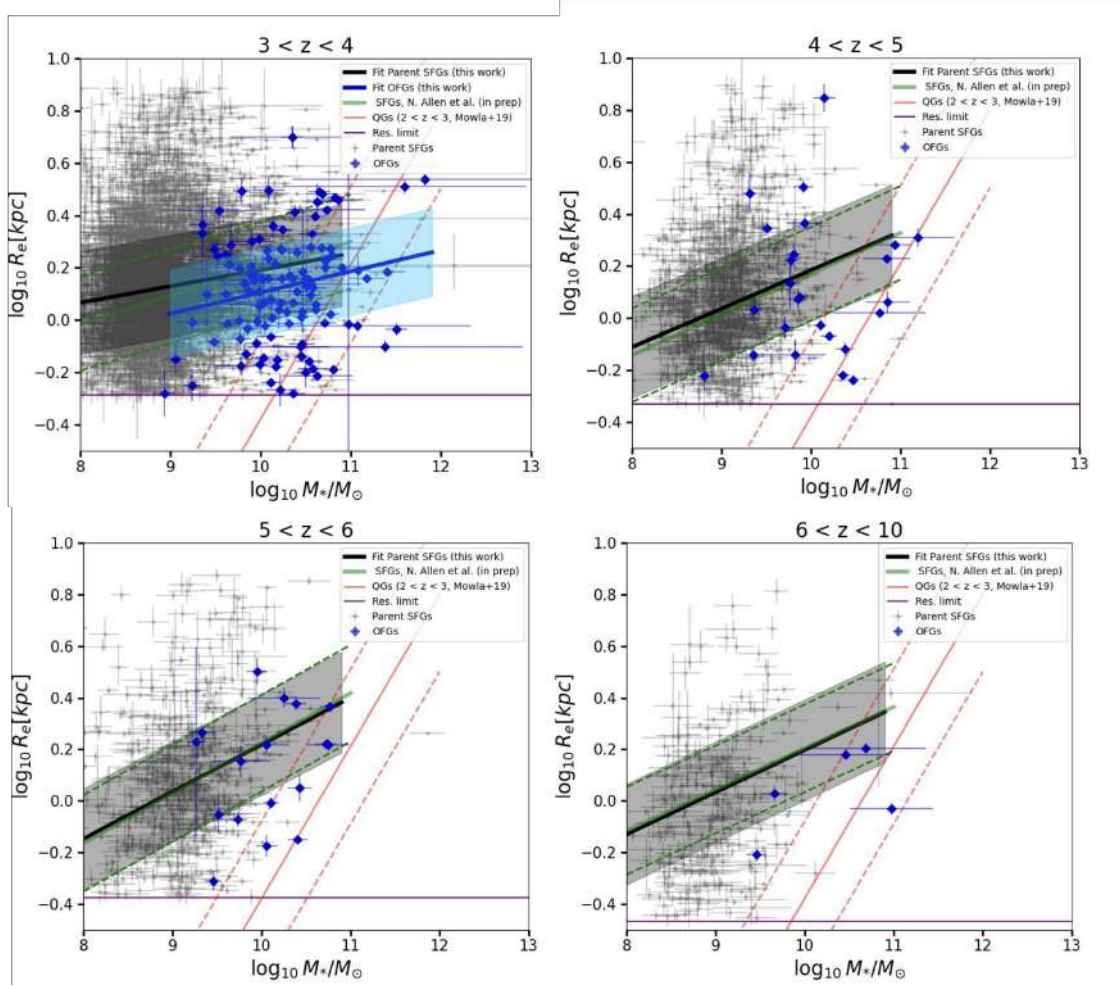


Figure 21: Rest-frame Optical Mass-Size Relation for Parent star forming Galaxies (grey and black lines) and Optically dark and faint Galaxies (blue) for redshift ranges of 3 to 4. Due to the lack of many OFGs at $z > 4$, the mass-size relation has not been plotted for those ranges. The green lines represent the mass-size relation at $3 < z < 4$ obtained from JWST Rest-frame UV/Optical sizes (Allen et al., in prep.), for reference. The red lines indicate the mass-size relation of Quiescent galaxies at $2 < z < 3$ (Mowla et al. 2019). The shaded light blue and grey regions show the 1σ scatter of the best-fit relations of OFGs and Parent SFGs, respectively.

5.3 Drivers of Dust Attenuation

“Drivers” of dust attenuation are parameters that can be seen as indicators of dust attenuation as they show correlation with A_V mag. In Figure 22 we study the relationship of stellar mass, $R_e[kpc]$ and q with dust attenuation at a redshift range of 3 to 4. Figure 22a shows A_V mag as a function of stellar mass. As we have already seen in Carlos Gómez-Guijarro et al. 2023, dust attenuation correlates with stellar mass in SFGs, (eg. Zahid et al. 2013; Pannella et al. 2015; Álvarez-Márquez et al. 2016) as well as OFGs. With more robust statistics owing to a higher number of OFGs in our selection, we can confirm that stellar mass acts as a primary proxy for dust attenuation in SFGs. In figure 22b, we compare values of $R_e[kpc]$, scaled to the same stellar mass of $10^{10}M_\odot$ with A_V mag. This approach allows us to better understand how galaxy morphology influences dust attenuation, independent of the effects of stellar mass and redshift evolution. The OFGs are scattered around smaller sizes as compared to the parent SFG sample. Moreover, we see about four times higher values in A_V mag with twice as small values in $R_e[kpc]$ for OFGs as compared to the parent SFGs. Thus, the main driver of such high A_V mag in OFGs compared to other galaxies is their compact size. We do not see any dependence of dust attenuation values with an increase in q values for either of the samples as shown in Figure 22c. The distribution seems almost random. The OFGs, however, seem more “face-on” compared to the parent sample.

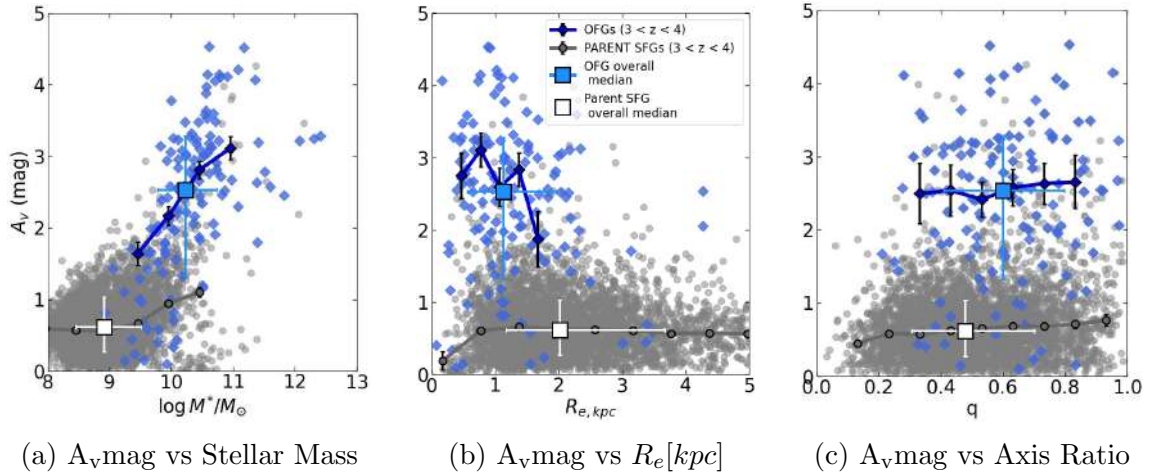


Figure 22: Comparison of Stellar Mass, Size ($R_e[kpc]$) and Axis Ratio with Dust Attenuation of OFGs and Parent SFGs at redshifts 3 to 4. Sliding medians for the two galaxy types shown in the legend are displayed with bold diamonds for OFGs, and bold circles for Parent SFGs, with the error bars representing the uncertainty of the medians. Bins with less than ten galaxies are not displayed. For Figure 22b, the $R_e[kpc]$ values have been scaled to the same stellar mass of $10^{10}M_\odot$.

We now explore if any other additional morphological parameters could be correlated with dust attenuation. We compare the effective surface mass density ($\Sigma_{Re,kpc} = \frac{M_*}{2\pi R_e^2}$), the mass enclosed within a circular area of radius equal to the effective radius of the galaxy with their dust attenuation. We divide our sample into redshift bins and compare the $\Sigma_{Re,kpc}$ after scaling all sources to a stellar mass of $\log M/M_\odot = 10$. This has been illustrated in Figure 23. We do not see any correlation of $\Sigma_{Re,kpc}$ with dust attenuation.

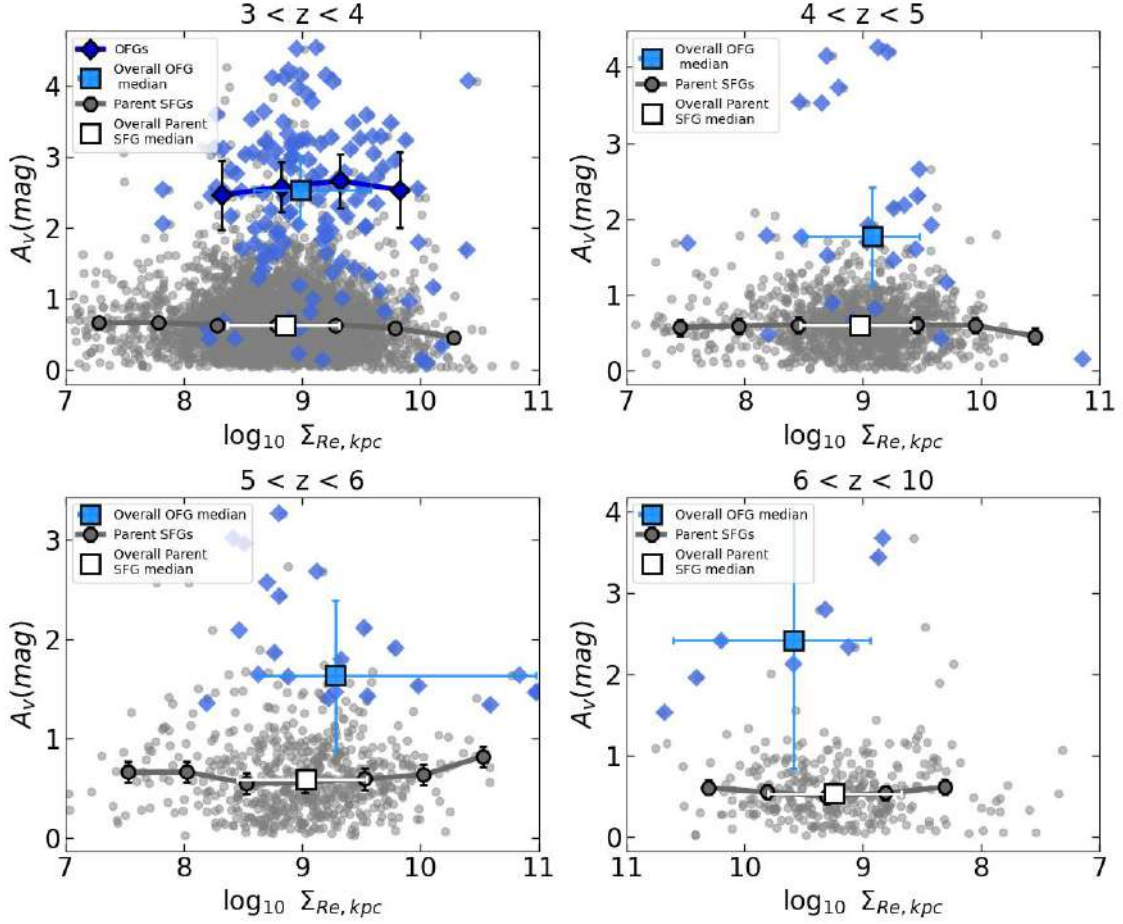


Figure 23: Dust attenuation Vs Effective stellar mass surface density ($\Sigma_{Re,kpc}$), in $M_{\odot} \text{ kpc}^{-2}$. The $\Sigma_{Re,kpc}$ has been scaled to the same stellar mass of $10^{10} M_{\odot}$ and binned by redshifts. Sliding medians for the two galaxy types shown in the legend are displayed with bold diamonds for OFGs, bold circles for Parent SFGs, with the error bars representing the uncertainty of the medians. Bin medians with less than ten galaxies are not displayed.

5.4 OFGs: Compact progenitors of low redshift Quiescent galaxies?

A common feature in the evolution of massive galaxies is the development of a dense core, indicating that the structural growth and the quenching of star formation could be traced by the stellar mass density at the centre of the galaxy (Fang et al. 2013; van Dokkum et al. 2014a; K. E. Whitaker, Bezanson, et al. 2017. Drawing inspiration from Barro, Faber, et al. 2017), we aim to calculate the stellar mass densities of our galaxy samples and compare them with established scaling relations for SFGs and quiescent galaxies as mentioned in this study.

The scaling relation is expressed as:

$$\log \Sigma_{Re,kpc} = \alpha \left[\log \left(\frac{M_{\star}}{M_{\odot}} \right) - 10.5 \right] + \log A(z), \quad (6)$$

where α represents the slope of the size–mass relation, and $\log A(z)$ is a scaling constant.

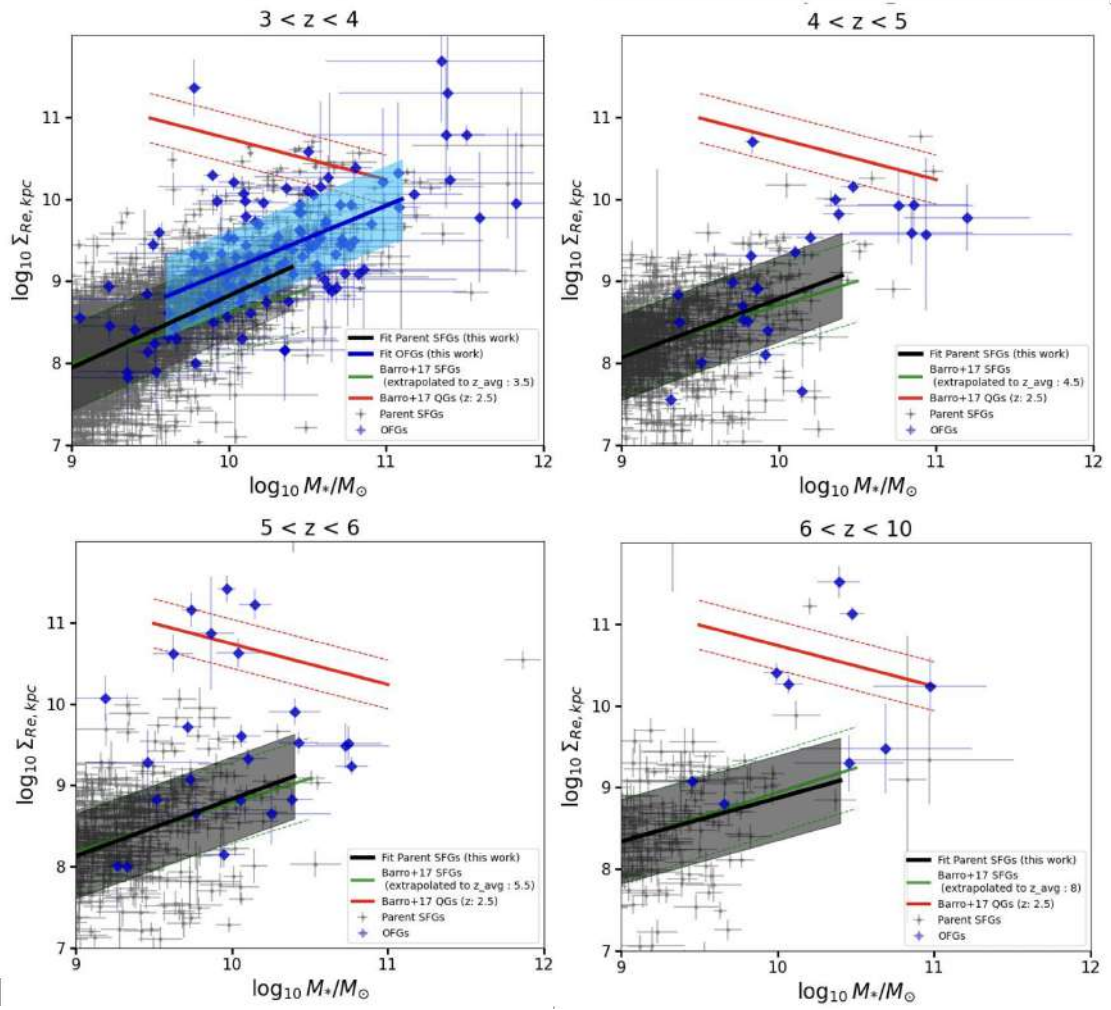


Figure 24: Redshift binned Effective stellar surface mass density $\Sigma_{Re,kpc}$ ($M_\odot \text{ kpc}^{-2}$) Vs Stellar Mass. Grey points denote SFGs while blue diamonds denote OFGs. The red solid lines represent $\Sigma_{Re,kpc}$ - Mass relation of quiescent galaxies (QGs) at redshifts of 2 to 3, and dashed red lines show the intrinsic scatter of this population, taken from Barro, Faber, et al. 2017. The green solid and dashed lines represent the redshift extrapolated best-fit lines and intrinsic scatter for star forming galaxies from the same study. The black lines represent $\Sigma_{Re,kpc}$ - Mass relation of Parent SFGs at given redshifts as shown on top of each subplot. We have also shown the $\Sigma_{Re,kpc}$ - Mass relation for OFGs ($3 < z < 4$) with the dark blue line. The shaded light blue and grey regions show the 1σ scatter of the best-fit relations of OFGs and Parent SFGs, respectively.

We aim to see if our OFGs lie between SFGs and quiescent galaxies in terms of their stellar mass densities. This would potentially indicate that OFGs serve as an intermediate population, filling a gap between parent SFGs and QGs at lower redshifts.

Figure 24, shows the distribution of the effective surface mass densities for the parent SFGs and OFGs. To recall, the effective surface mass density is described as the mass enclosed within a circular surface area of radius equal to the effective radius

of the galaxy. It can be expressed as :

$$\Sigma_{\text{Re,kpc}} = \frac{M_*}{2\pi R_e^2} \quad (7)$$

The majority of our OFGs lie within the redshift range of 3 to 4 (top-left plot, figure 24). These slopes for parent SFGs are relatively constant with time, $\alpha^{SF} \sim 0.6$ and they are consistent with the redshift extrapolated scaling relations from Barro, Faber, et al. 2017 shown in green. The slope of the best-fit relation for OFGs is similar to the parent SFGs sample but has a greater offset towards higher $\Sigma_{\text{Re,kpc}}$. The best-fit slope and zero point for the OFG ($3 < z < 4$) comes out to be $\alpha = 0.865 \pm 0.077$, $\log A = 9.229 \pm 0.023$. This is an indication that the OFGs lie at a junction between the parent SFG and quiescent galaxy ($\alpha^Q \sim -0.5$) scaling relations, thus showing initial signs of evolution of OFGs towards the quiescent phase.

6 Discussion

In this work, we explored in detail the nature of Optically Faint or Dark Galaxies. In this section, we shall try to discuss our various results while trying to connect and compare them with previously done studies.

- **Sample Selection:** We see that using our sample selection, we have successfully selected optically faint, dust-obscured, star forming galaxies. The UVJ criterion is a method to differentiate dusty SFGs and QGs using the V-J colour as a proxy for dust attenuation (Wuyts et al. 2007; G. B. Brammer et al. 2009). It relies on the sensitivity of the rest-frame V band to the effects of dust attenuation (K. Whitaker 2011). Based on their UVJ colour criteria as shown in Figure 25, very few OFGs fall within the locus of quiescent galaxies. These are likely compact SFGs in the final stages of star formation, probably starting to quench. Some of these galaxies occupy a region of the UVJ diagram (left of the red dashed line) where recently quenched galaxies lie (K. E. Whitaker, Kriek, et al. 2012). Only 5 galaxies from the OFG sample and 31 galaxies from the parent SFG sample show colours likely corresponding to a quiescent nature.
- **Physical properties of Optically Faint/ Dark Galaxies:** OFGs are seen to be much more heavily dust attenuated with a factor of 3.8 higher than the median for parent SFGs. When scaled to the same stellar mass, this factor increases to over 4. The majority of these massive dusty star forming populations lie between redshifts of 3 and 4 and have stellar masses $> 10^8 M_\odot$ with about 75 % of sources with stellar masses between $10^{9.5}$ and $10^{12} M_\odot$.

After conducting morphological analysis on a large sample of selected galaxies using `pysersic`, and trying different cut-offs for the classification of “PSF-dominated” and “Sersic-dominated” sources, we decide to set the flux cut-off for “PSF-dominated” sources at above 50 % in order to have a good statistical sample. About 40% of OFGs are deemed “PSF-dominated” while almost 52

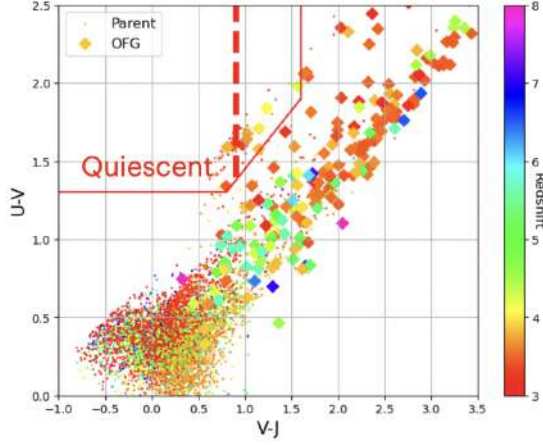


Figure 25: UVJ diagram (Williams et al, 2009) for OFGs and Parent SFGs. Very few sources lie on the quiescent locus (in red) of the UVJ selection criteria, based on their rest frame colors.

% of LRDs are found to be “PSF-dominated”. Of course, this estimation is slightly biased as we decided to discard all sources fainter than a total flux of 145 nJys, which would have bumped up these numbers otherwise. We then study the “Sérsic-dominated” sources and find that there are subtle differences in morphology between Parent SFGs and OFGs. The results on the effective radius, sérsic index and axis ratio values are consistent with the work in Carlos Gómez-Guijarro et al. 2023. OFGs have slightly higher values for all these parameters than the Parent SFG population, while most sources have small effective radii. A median sérsic index close to 2 and a large scatter at higher sérsic indices for OFGs indicates that the light profile is more centrally concentrated than an exponential disk profile, indicating signs that bulge formation might be taking effect in these galaxies. The OFGs have high axis ratios with a median of 0.5, indicating that they are mostly oriented face-on. It is important to note that, for really compact galaxies as found for OFGs, it is difficult to constrain the axis ratios too accurately as their sizes are very close to the resolution limit of the instrument.

The location of OFGs in the mass-size relation is also interesting. Their best-fit relation neither matches the slope for star forming galaxies nor does it match the steep slopes of quiescent galaxies (QGs). This is the first indicator that the nature of OFGs is intermediate between SFGs and QGs.

- **Drivers of Dust Attenuation:** We asses different morphological parameters which may or may not be used as proxies for dust attenuation. It is evident that higher stellar mass is a great indicator of higher dust attenuation in dust star forming galaxies. The correlation between stellar mass and dust attenuation has been set in place before the JWST era from various studies on galaxies at $z < 4$. Simply put, the stellar mass signifies how deep the galaxy’s potential well goes, thus telling us how well they can retain their metals and dust which could be blown off by galactic winds. As mentioned in previous studies, the mass metallicity relation is a great proof of this.(Sánchez et al. 2013; Genzel et al. 2015 etc.). The OFGs are scattered around smaller sizes as compared to the parent SFG sample. When scaled to the same stellar masses,

we see that the main driver of such high dust attenuation in OFGs compared to other galaxies is their compact size.

We further looked at the effect of effective stellar mass density ($\Sigma_{Re,kpc}$) and axis ratios on dust attenuation. At a fixed stellar mass and redshift, we saw no correlation between $\Sigma_{Re,kpc}$ and dust attenuation. The median value of $\Sigma_{Re,kpc}$ for OFGs is, however, slightly higher than that for parent SFGs. This is observed across all redshifts. Axis ratios are indicators of the orientation of galaxies. More OFGs have comparatively higher axis ratios which means they are most likely face-on disk galaxies. In previous literature, we have seen a correlation between orientation and dust attenuation (Zuckerman et al. 2021). In our study, however, we see that there is no connection between orientation and dust attenuation and the distribution is almost random which could imply that these high redshift galaxies comprise thick disks. This result is consistent with those from Carlos Gómez-Guijarro et al. 2023 and Lorenz et al. 2023.

• **OFGs: Compact progenitors of low redshift Quiescent galaxies (QGs)?:**

In this section we studied $\Sigma_{Re,kpc}$ or the effective mass density of OFGs as a function of their total stellar mass to measure their true compactness with respect to normal SFG and QG populations. Evidently, the best-fit relation of OFGs lies between SFG and QG $\Sigma_{Re,kpc} - M_*$ scaling relations. The OFG sample shows great similarity with compact SFGs as mentioned in Barro, Faber, et al. 2017. Previous studies have shown, at every redshift, QGs show higher surface mass densities and higher sérsic indices than the SFGs. This has been explained in detail by several studies (Fang et al. 2013; van Dokkum et al. 2014b, etc). These studies suggest that immediate star forming progenitors of quiescent galaxies seen at lower redshifts, undergo a phase of core growth which allows them to grow faster than those following the $\Sigma_e - M_*$ relations. Supporting the work of Barro, Faber, et al. 2017, which had a statistical sample up to a redshift of 3, we can say this behaviour of compact SFGs is seen in OFGs which exist at higher redshifts.

7 Conclusion & Future work

In this work, we investigated the nature of Optically dark and faint galaxies (OFGs), their stellar morphologies, more specifically their sizes, drivers of dust attenuation and their connection with quiescent galaxies at lower redshifts. We have used JWST/NIRCam 4.44 μm data in three large fields, namely, PRIMER-COSMOS, PRIMER-UDS, and CEERS. The main results from this work are shown below:

- The physical properties of OFGs prove to be much different from the parent SFG sample selected for statistical analysis. 75% of OFGs lie at a stellar mass range of $\log M_*/M_\odot = 9.5$ to 12 with 3.8 factors higher median A_v mag values than the parent SFGs, and a factor of 4 when scaled to the same stellar mass.
- After conducting a PSF-Sérsic morphological decomposition we find that about 40% of OFGs are deemed PSF-dominated while almost 52% of LRDs are found to be PSF-dominated. There exists a bimodal distribution of “PSF-dominated” sources and “Sérsic-dominated” sources in the color selected OFGs

and LRDs. The parent SFG sample on the other hand shows that 95% come out to be “Sérsic-dominated”.

- Among the “Sérsic-dominated” sources, the OFGs show a higher median sérsic index close to 2 with a large scatter towards higher sérsic index values, and higher axis ratios (~ 0.59) as compared to parent SFGs.
- Stellar mass is confirmed to be a proxy for dust attenuation for OFGs. The effective radius is correlated to dust attenuation. OFGs are scattered around small effective radii and show higher dust attenuation towards smaller values of effect radius, meaning that the more compact the source higher is its dust attenuation. There is no correlation with orientation, i.e, axis ratio of OFGs with their dust attenuation. The distribution seems almost random.
- OFGs show some signs of potentially being compact progenitors of QGs at lower redshifts. They not only lie in an intermediate location between SFGs and QGs in the mass-size relation but also show a higher effective stellar mass density as compared to the parent SFG sample, indicating some signs of onset of core or bulge growth.

- **Future work:**

- **Caveats of this work:** In this work, we have tried to constrain the mass-size relation for parent SFGs and OFGs. As seen from Figure 21, there exists a large scatter of values at high values of $R_e[kpc]$ for the Parent SFG sample. There may exist a bias towards higher size values. In the future, we plan to test these biases using more complex simulations and further inspections. Moreover, there exists a bimodality of PSF fraction distributions, as seen in Figure 16. Thus there also exists a bias towards very high (~ 1) or very low (~ 0) PSF fractions, which is not observed in our simulations. In reality, galaxies do not necessarily exhibit a perfect Sérsic light distribution. Thus we need to complexify our tests by incorporating a varied range of axis ratios and position angles to our simulated sources.
- We made a preliminary check to see if OFGs show any signs of being potential progenitors of quiescent galaxies observed at lower redshifts. Comparing the effective stellar mass density($\Sigma_{Re,kpc}$), we saw that OFGs show higher values of $\Sigma_{Re,kpc}$ than the parent sample, and are more massive. Scaling them to the same stellar mass we saw that the median value of the of $\Sigma_{Re,kpc}$ is higher than that of the parent sample, thus suggesting that they are indeed more massive and compact than the parent SFGs. It would be interesting to see the ΔSFR versus $\Delta \Sigma_{Re,kpc}$ defined by their offset from scaling relations at a given redshift. It is suggested to be an effective selection to identify quenched systems. If we can see that the “L-shaped” distribution is achieved as seen in Barro, Faber, et al. 2017, we can further confirm or deny if any compaction and quenching is observed in our OFGs. Knowing that they might be potential progenitors of quiescent galaxies, it could be interesting to search for ALMA counterparts of these galaxies in submillimetre wavelengths. Probing the Infrared black body emission of these sources would allow us to further

investigate the dusty nature of OFGs and understand the effect of dust on quenching mechanisms at play.

- Despite the aforementioned fitting biases, we managed to find 52% of colour-selected LRDs and 40% of OFGs to be “PSF-dominated” sources which could be candidates for AGNs or unresolved sources. Of course, these sources show unresolved sizes in their Sérsic profile fits. So an analysis of their spectral energy distributions and possible spectroscopic follow-ups could help us better understand the nature of these PSF-dominated sources.

A Appendix

Table 3: 5σ depths in HST/ACS, HST/WFC3 and JWST/NIRCam filters. Depths are given in AB magnitudes, and have been provided by A. Weibel (Weibel et al. 2024).

Telescope/Instrument	Filter	5σ depth [AB mag]
HST/ACS	F435W	28.27
	F606W	28.36
	F814W	28.19
HST/WFC3	F105W	27.96
	F125W	27.74
	F140W	26.99
	F160W	27.81
JWST/NIRCam	F115W	28.63
	F150W	28.65
	F200W	28.93
	F277W	29.17
	F356W	29.17
	F410M	28.41
	F444W	28.81

Table 4: Best-fit parameters for size mass equation $R_e[\text{kpc}] = \alpha \log_{10}(M_* - M_0) + \log_{10} A$ in four redshift bins: $3 \leq z < 4$, $4 \leq z < 5$, $5 \leq z < 6$, $6 \leq z < 10$. α is the slope, $\log_{10} A$ is the intercept and σ_{int} is the intrinsic scatter

Redshift	Parameter	Parent SFGs $M_0 = 10^9$	OFGs $M_0 = 10^9$
$3 \leq z < 4$	α	0.0624	0.0809
	σ_α	0.0055	0.0304
	$\log_{10} A$	0.1291	0.0245
	$\sigma_{\log_{10} A}$	0.0028	0.0414
	σ_{int}	0.2171	0.1672
$4 \leq z < 5$	α	0.1478	-
	σ_α	0.009	
	$\log_{10} A$	0.0365	
	$\sigma_{\log_{10} A}$	0.0041	
	σ_{int}	0.2073	
$5 \leq z < 6$	α	0.1575	-
	σ_α	0.0101	
	$\log_{10} A$	0.0234	
	$\sigma_{\log_{10} A}$	0.0045	
	σ_{int}	0.2114	
$6 \leq z < 10$	α	0.1612	-
	σ_α	0.0108	
	$\log_{10} A$	-0.0727	
	$\sigma_{\log_{10} A}$	0.0171	
	σ_{int}	0.3015	

References

- Alcalde Pampliega, Belén et al. (May 2019). “Optically Faint Massive Balmer Break Galaxies at $z \gtrsim 3$ in the CANDELS/GOODS Fields”. In: *ApJ* 876.2, 135, p. 135. DOI: 10.3847/1538-4357/ab14f2. arXiv: 1806.04152 [astro-ph.GA].
- Álvarez-Márquez, J. et al. (Mar. 2016). “Dust properties of Lyman-break galaxies at $z \sim 3$ ”. In: *A&A* 587, A122, A122. DOI: 10.1051/0004-6361/201527190. arXiv: 1512.04120 [astro-ph.GA].
- Bagley, Micaela B. et al. (Mar. 2023). “CEERS Epoch 1 NIRCам Imaging: Reduction Methods and Simulations Enabling Early JWST Science Results”. In: *ApJ L* 946.1, L12, p. L12. DOI: 10.3847/2041-8213/acbb08. arXiv: 2211.02495 [astro-ph.IM].
- Barro, Guillermo, S. M. Faber, et al. (May 2017). “Structural and Star-forming Relations since $z \sim 3$: Connecting Compact Star-forming and Quiescent Galaxies”. In: *ApJ* 840.1, 47, p. 47. DOI: 10.3847/1538-4357/aa6b05. arXiv: 1509.00469 [astro-ph.GA].
- Barro, Guillermo, Pablo G. Pérez-González, et al. (Mar. 2024). “Extremely Red Galaxies at $z = 5-9$ with MIRI and NIRSpect: Dusty Galaxies or Obscured Active Galactic Nuclei?” In: *ApJ* 963.2, 128, p. 128. DOI: 10.3847/1538-4357/ad167e. arXiv: 2305.14418 [astro-ph.GA].
- Barrufet, L. et al. (June 2023). “Unveiling the nature of infrared bright, optically dark galaxies with early JWST data”. In: *MNRAS* 522.1, pp. 449–456. DOI: 10.1093/mnras/stad947. arXiv: 2207.14733 [astro-ph.GA].
- Birkmann, Stephan et al. (Sept. 2022). *NIRSpec and NIRCам detector verification*. JWST Proposal. Cycle 1, ID. #2769.
- Bouwens, R. J. et al. (Feb. 2010). “Discovery of $z \sim 8$ Galaxies in the Hubble Ultra Deep Field from Ultra-Deep WFC3/IR Observations”. In: *ApJ L* 709.2, pp. L133–L137. DOI: 10.1088/2041-8205/709/2/L133. arXiv: 0909.1803 [astro-ph.CO].
- Bouwens, Rychard et al. (Oct. 2020). “The ALMA Spectroscopic Survey Large Program: The Infrared Excess of $z = 1.5-10$ UV-selected Galaxies and the Implied High-redshift Star Formation History”. In: *ApJ* 902.2, 112, p. 112. DOI: 10.3847/1538-4357/abb830. arXiv: 2009.10727 [astro-ph.GA].
- Boylan-Kolchin, Michael (June 2023). “Stress testing Λ CDM with high-redshift galaxy candidates”. In: *Nature Astronomy* 7, pp. 731–735. DOI: 10.1038/s41550-023-01937-7. arXiv: 2208.01611 [astro-ph.CO].
- Brammer, G. B. et al. (Nov. 2009). “The Dead Sequence: A Clear Bimodality in Galaxy Colors from $z = 0$ to $z = 2.5$ ”. In: *ApJ L* 706.1, pp. L173–L177. DOI: 10.1088/0004-637X/706/1/L173. arXiv: 0910.2227 [astro-ph.CO].
- Brammer, Gabriel B., Pieter G. van Dokkum, and Paolo Coppi (Oct. 2008). “EAZY: A Fast, Public Photometric Redshift Code”. In: *The Astrophysical Journal* 686.2, pp. 1503–1513. ISSN: 1538-4357. DOI: 10.1086/591786. URL: <http://dx.doi.org/10.1086/591786>.
- Carnall, Adam C. et al. (July 2023). “A massive quiescent galaxy at redshift 4.658”. In: *Nature* 619.7971, pp. 716–719. DOI: 10.1038/s41586-023-06158-6. arXiv: 2301.11413 [astro-ph.GA].
- Casey, Caitlin M., Jeyhan S. Kartaltepe, et al. (2023). *COSMOS-Web: An Overview of the JWST Cosmic Origins Survey*. arXiv: 2211.07865 [astro-ph.GA].

- Casey, Caitlin M., Jorge A. Zavala, et al. (Dec. 2021). “Mapping Obscuration to Reionization with ALMA (MORA): 2 mm Efficiently Selects the Highest-redshift Obscured Galaxies”. In: *ApJ* 923.2, 215, p. 215. DOI: 10.3847/1538-4357/ac2eb4. arXiv: 2110.06930 [astro-ph.GA].
- Dhawan, Suhail, Saurabh W. Jha, and Bruno Leibundgut (Jan. 2018). “Measuring the Hubble constant with Type Ia supernovae as near-infrared standard candles”. In: *Astronomy and Astrophysics* 609, A72. ISSN: 1432-0746. DOI: 10.1051/0004-6361/201731501. URL: <http://dx.doi.org/10.1051/0004-6361/201731501>.
- Doyon, Rene et al. (2023). *The Near Infrared Imager and Slitless Spectrograph for the James Webb Space Telescope – I. Instrument Overview and in-Flight Performance*. arXiv: 2306.03277 [astro-ph.IM].
- Eldridge, J. J. et al. (2017). “Binary Population and Spectral Synthesis Version 2.1: Construction, Observational Verification, and New Results”. In: *Publications of the Astronomical Society of Australia* 34. ISSN: 1448-6083. DOI: 10.1017/pasa.2017.51. URL: <http://dx.doi.org/10.1017/pasa.2017.51>.
- Endsley, Ryan, Daniel P. Stark, Xiaohui Fan, et al. (May 2022). “Radio and far-IR emission associated with a massive star-forming galaxy candidate at $z \sim 6.8$: a radio-loud AGN in the reionization era?”. In: *MNRAS* 512.3, pp. 4248–4261. DOI: 10.1093/mnras/stac737. arXiv: 2108.01084 [astro-ph.GA].
- Endsley, Ryan, Daniel P. Stark, Jianwei Lyu, et al. (Apr. 2023). “ALMA confirmation of an obscured hyperluminous radio-loud AGN at $z = 6.853$ associated with a dusty starburst in the 1.5 deg^2 COSMOS field”. In: *MNRAS* 520.3, pp. 4609–4620. DOI: 10.1093/mnras/stad266. arXiv: 2206.00018 [astro-ph.GA].
- Enia, Andrea et al. (Mar. 2022). “A New Estimate of the Cosmic Star Formation Density from a Radio-selected Sample, and the Contribution of H-dark Galaxies at $z \geq 3$ ”. In: *ApJ* 927.2, 204, p. 204. DOI: 10.3847/1538-4357/ac51ca. arXiv: 2202.00019 [astro-ph.CO].
- Fang, Jerome J. et al. (Oct. 2013). “A Link between Star Formation Quenching and Inner Stellar Mass Density in Sloan Digital Sky Survey Central Galaxies”. In: *ApJ* 776.1, 63, p. 63. DOI: 10.1088/0004-637X/776/1/63. arXiv: 1308.5224 [astro-ph.CO].
- Finkelstein, Steven L., Micaela B. Bagley, et al. (Dec. 2022). “A Long Time Ago in a Galaxy Far, Far Away: A Candidate $z \sim 12$ Galaxy in Early JWST CEERS Imaging”. In: *ApJ L* 940.2, L55, p. L55. DOI: 10.3847/2041-8213/ac966e. arXiv: 2207.12474 [astro-ph.GA].
- Finkelstein, Steven L., Gene C. K. Leung, et al. (2023). *The Complete CEERS Early Universe Galaxy Sample: A Surprisingly Slow Evolution of the Space Density of Bright Galaxies at $z \sim 8.5-14.5$* . arXiv: 2311.04279 [astro-ph.GA].
- Frenk, C. S. (Dec. 1986). “Galaxy Clustering and the Dark-Matter Problem”. In: *Philosophical Transactions of the Royal Society of London Series A* 320.1556, pp. 517–541. DOI: 10.1098/rsta.1986.0133.
- Fu, Shuqi et al. (Mar. 2024). “The Magellan M2FS Spectroscopic Survey of High-redshift Galaxies: The Brightest Lyman-break Galaxies at $z \sim 6$ ”. In: *ApJ* 963.1, 51, p. 51. DOI: 10.3847/1538-4357/ad1ce5.
- Fujimoto, S. et al. (Apr. 2022). “A dusty compact object bridging galaxies and quasars at cosmic dawn”. In: *Nature* 604.7905, pp. 261–265. DOI: 10.1038/s41586-022-04454-1. arXiv: 2204.06393 [astro-ph.GA].

- Genzel, R. et al. (Feb. 2015). “Combined CO and Dust Scaling Relations of Depletion Time and Molecular Gas Fractions with Cosmic Time, Specific Star-formation Rate, and Stellar Mass”. In: *ApJ* 800.1, 20, p. 20. DOI: 10.1088/0004-637X/800/1/20. arXiv: 1409.1171 [astro-ph.GA].
- Gómez-Guijarro, C. et al. (Mar. 2022). “GOODS-ALMA 2.0: Starbursts in the main sequence reveal compact star formation regulating galaxy evolution prequenching”. In: *A&A* 659, A196, A196. DOI: 10.1051/0004-6361/202142352. arXiv: 2201.02633 [astro-ph.GA].
- Gómez-Guijarro, Carlos et al. (Sept. 2023). “JWST CEERS probes the role of stellar mass and morphology in obscuring galaxies”. In: *A&A* 677, A34, A34. DOI: 10.1051/0004-6361/202346673. arXiv: 2304.08517 [astro-ph.GA].
- Gottumukkala, R. et al. (May 2024). “Unveiling the hidden Universe with JWST: the contribution of dust-obscured galaxies to the stellar mass function at $z = 3 - 8$ ”. In: *MNRAS* 530.1, pp. 966–983. DOI: 10.1093/mnras/stae754. arXiv: 2310.03787 [astro-ph.GA].
- Harikane, Yuichi et al. (Dec. 2023). “A JWST/NIRSpec First Census of Broad-line AGNs at $z = 4-7$: Detection of 10 Faint AGNs with $M_{BH} 10^6-10^8 M_{\odot}$ and Their Host Galaxy Properties”. In: *ApJ* 959.1, 39, p. 39. DOI: 10.3847/1538-4357/ad029e. arXiv: 2303.11946 [astro-ph.GA].
- Hausen, R. and B.E. Robertson (2022). “FitsMap: A simple, lightweight tool for displaying interactive astronomical image and catalog data”. In: *Astronomy and Computing* 39, p. 100586. ISSN: 2213-1337. DOI: <https://doi.org/10.1016/j.ascom.2022.100586>. URL: <https://www.sciencedirect.com/science/article/pii/S2213133722000257>.
- Kocevski, Dale D. et al. (Sept. 2023). “Hidden Little Monsters: Spectroscopic Identification of Low-mass, Broad-line AGNs at $z \lesssim 5$ with CEERS”. In: *ApJ L* 954.1, L4, p. L4. DOI: 10.3847/2041-8213/ace5a0. arXiv: 2302.00012 [astro-ph.GA].
- Kroupa, Pavel, Christopher A. Tout, and Gerard Gilmore (June 1993). “The Distribution of Low-Mass Stars in the Galactic Disc”. In: *MNRAS* 262, pp. 545–587. DOI: 10.1093/mnras/262.3.545.
- Labbe, Ivo et al. (June 2023). “UNCOVER: Candidate Red Active Galactic Nuclei at $z \gtrsim 7$ with JWST and ALMA”. In: *arXiv e-prints*, arXiv:2306.07320, arXiv:2306.07320. DOI: 10.48550/arXiv.2306.07320. arXiv: 2306.07320 [astro-ph.GA].
- Larson, Rebecca L. et al. (Aug. 2023). “A CEERS Discovery of an Accreting Supermassive Black Hole 570 Myr after the Big Bang: Identifying a Progenitor of Massive $z \gtrsim 6$ Quasars”. In: *ApJ L* 953.2, L29, p. L29. DOI: 10.3847/2041-8213/ace619. arXiv: 2303.08918 [astro-ph.GA].
- Lorenz, Brian et al. (July 2023). “An Updated Dust-to-Star Geometry: Dust Attenuation Does Not Depend on Inclination in $1.3 \leq z \leq 2.6$ Star-forming Galaxies from MOSDEF”. In: *ApJ* 951.1, 29, p. 29. DOI: 10.3847/1538-4357/acdd1. arXiv: 2304.08521 [astro-ph.GA].
- Madau, Piero and Mark Dickinson (Aug. 2014). “Cosmic Star-Formation History”. In: *ARA&A* 52, pp. 415–486. DOI: 10.1146/annurev-astro-081811-125615. arXiv: 1403.0007 [astro-ph.CO].
- Madau, Piero, Henry C. Ferguson, et al. (Dec. 1996). “High-redshift galaxies in the Hubble Deep Field: colour selection and star formation history to $z \sim 4$ ”. In:

- MNRAS* 283.4, pp. 1388–1404. DOI: 10.1093/mnras/283.4.1388. arXiv: astro-ph/9607172 [astro-ph].
- Magdis, G. E. et al. (Nov. 2010). “Herschel reveals a T_{dust} -unbiased selection of $z \sim 2$ ultraluminous infrared galaxies”. In: *MNRAS* 409.1, pp. 22–28. DOI: 10.1111/j.1365-2966.2010.17551.x. arXiv: 1007.4900 [astro-ph.CO].
- Maiolino, Roberto et al. (Aug. 2023). “JADES. The diverse population of infant Black Holes at 4 $\leq z \leq 11$: merging, tiny, poor, but mighty”. In: *arXiv e-prints*, arXiv:2308.01230, arXiv:2308.01230. DOI: 10.48550/arXiv.2308.01230. arXiv: 2308.01230 [astro-ph.GA].
- Manning, Sinclair M. et al. (Jan. 2022). “Characterization of Two 2 mm detected Optically Obscured Dusty Star-forming Galaxies”. In: *ApJ* 925.1, 23, p. 23. DOI: 10.3847/1538-4357/ac366a. arXiv: 2111.02428 [astro-ph.GA].
- Marrone, D. P. et al. (Jan. 2018). “Galaxy growth in a massive halo in the first billion years of cosmic history”. In: *Nature* 553.7686, pp. 51–54. DOI: 10.1038/nature24629. arXiv: 1712.03020 [astro-ph.GA].
- Matthee, Jorryt et al. (Mar. 2024). “Little Red Dots: An Abundant Population of Faint Active Galactic Nuclei at $z \sim 5$ Revealed by the EIGER and FRESCO JWST Surveys”. In: *ApJ* 963.2, 129, p. 129. DOI: 10.3847/1538-4357/ad2345. arXiv: 2306.05448 [astro-ph.GA].
- Menci, N. et al. (Oct. 2022). “High-redshift Galaxies from Early JWST Observations: Constraints on Dark Energy Models”. In: *ApJ L* 938.1, L5, p. L5. DOI: 10.3847/2041-8213/ac96e9. arXiv: 2208.11471 [astro-ph.CO].
- Morishita, T. et al. (Nov. 2020). “SuperBoRG: Exploration of Point Sources at $z \sim 8$ in HST Parallel Fields”. In: *ApJ* 904.1, 50, p. 50. DOI: 10.3847/1538-4357/abba83. arXiv: 2009.10078 [astro-ph.GA].
- Mowla, Lamiya A. et al. (July 2019). “COSMOS-DASH: The Evolution of the Galaxy Size-Mass Relation since $z \sim 3$ from New Wide-field WFC3 Imaging Combined with CANDELS/3D-HST”. In: *ApJ* 880.1, 57, p. 57. DOI: 10.3847/1538-4357/ab290a. arXiv: 1808.04379 [astro-ph.GA].
- Oesch, P. A. et al. (May 2014). “The Most Luminous $z \sim 9$ -10 Galaxy Candidates Yet Found: The Luminosity Function, Cosmic Star-formation Rate, and the First Mass Density Estimate at 500 Myr”. In: *ApJ* 786.2, 108, p. 108. DOI: 10.1088/0004-637X/786/2/108. arXiv: 1309.2280 [astro-ph.CO].
- Oesch, Pascal et al. (Aug. 2016). *A Spectroscopic Redshift for the Most Luminous Galaxy Candidate at $z \sim 10$* . HST Proposal. Cycle 22, ID. #13871.
- Onoue, Masafusa et al. (Jan. 2023). “A Candidate for the Least-massive Black Hole in the First 1.1 Billion Years of the Universe”. In: *ApJ L* 942.1, L17, p. L17. DOI: 10.3847/2041-8213/aca9d3. arXiv: 2209.07325 [astro-ph.GA].
- Pannella, M. et al. (July 2015). “GOODS-Herschel: Star Formation, Dust Attenuation, and the FIR-radio Correlation on the Main Sequence of Star-forming Galaxies up to $z \sim 4$ ”. In: *ApJ* 807.2, 141, p. 141. DOI: 10.1088/0004-637X/807/2/141. arXiv: 1407.5072 [astro-ph.GA].
- Pérez-González, Pablo G. et al. (Mar. 2023). “CEERS Key Paper. IV. A Triality in the Nature of HST-dark Galaxies”. In: *The Astrophysical Journal Letters* 946.1, p. L16. ISSN: 2041-8213. DOI: 10.3847/2041-8213/acb3a5. URL: <http://dx.doi.org/10.3847/2041-8213/acb3a5>.

- Planck Collaboration et al. (Sept. 2016). “Planck 2015 results. XIII. Cosmological parameters”. In: *A&A* 594, A13, A13. DOI: 10.1051/0004-6361/201525830. arXiv: 1502.01589 [astro-ph.CO].
- Rieke, G. H. et al. (July 2015). “The Mid-Infrared Instrument for the James Webb Space Telescope, I: Introduction”. In: *Publications of the Astronomical Society of the Pacific* 127.953, pp. 584–594. ISSN: 1538-3873. DOI: 10.1086/682252. URL: <http://dx.doi.org/10.1086/682252>.
- Rieke, Marcia J. et al. (Feb. 2023). “Performance of NIRCам on JWST in Flight”. In: *Publications of the Astronomical Society of the Pacific* 135.1044, p. 028001. ISSN: 1538-3873. DOI: 10.1088/1538-3873/acac53. URL: <http://dx.doi.org/10.1088/1538-3873/acac53>.
- Sánchez, S. F. et al. (June 2013). “Mass-metallicity relation explored with CALIFA. I. Is there a dependence on the star-formation rate?” In: *A&A* 554, A58, A58. DOI: 10.1051/0004-6361/201220669. arXiv: 1304.2158 [astro-ph.CO].
- Schreiber, C. et al. (Mar. 2015). “The Herschel view of the dominant mode of galaxy growth from $z = 4$ to the present day”. In: *A&A* 575, A74, A74. DOI: 10.1051/0004-6361/201425017. arXiv: 1409.5433 [astro-ph.GA].
- Stanway, E. R. and J. J. Eldridge (Sept. 2018). “Re-evaluating old stellar populations”. In: *MNRAS* 479.1, pp. 75–93. DOI: 10.1093/mnras/sty1353. arXiv: 1805.08784 [astro-ph.GA].
- Tohill, Clár-Bríd, Steven Bamford, and Christopher Conselice (2023). *Exploring the Morphologies of High Redshift Galaxies with Machine Learning*. arXiv: 2302.11482 [astro-ph.GA].
- van Dokkum, Pieter G. et al. (Aug. 2014a). “Dense Cores in Galaxies Out to $z = 2.5$ in SDSS, UltraVISTA, and the Five 3D-HST/CANDELS Fields”. In: *ApJ* 791.1, 45, p. 45. DOI: 10.1088/0004-637X/791/1/45. arXiv: 1404.4874 [astro-ph.GA].
- (Aug. 2014b). “Dense Cores in Galaxies Out to $z = 2.5$ in SDSS, UltraVISTA, and the Five 3D-HST/CANDELS Fields”. In: *ApJ* 791.1, 45, p. 45. DOI: 10.1088/0004-637X/791/1/45. arXiv: 1404.4874 [astro-ph.GA].
- Vikhlinin, A. et al. (Feb. 2009). “Chandra Cluster Cosmology Project III: Cosmological Parameter Constraints”. In: *ApJ* 692.2, pp. 1060–1074. DOI: 10.1088/0004-637X/692/2/1060. arXiv: 0812.2720 [astro-ph].
- Wang, T., C. Schreiber, and D. Elbaz (2019). “A dominant population of optically invisible massive galaxies in the early Universe”. In: *Nature* 572, pp. 211–214. DOI: 10.1038/s41586-019-1452-4.
- Weibel, Andrea et al. (Mar. 2024). “Galaxy Build-up in the first 1.5 Gyr of Cosmic History: Insights from the Stellar Mass Function at $z \sim 4 - 9$ from JWST NIRCам Observations”. In: *arXiv e-prints*, arXiv:2403.08872, arXiv:2403.08872. DOI: 10.48550/arXiv.2403.08872. arXiv: 2403.08872 [astro-ph.GA].
- Whitaker, K. (July 2011). “Quiescent galaxies through cosmic time from (ultra-) low resolution NIR spectroscopy”. In: *Galaxy Formation*, P41, P41.
- Whitaker, Katherine E., Rachel Bezanson, et al. (Mar. 2017). “Predicting Quiescence: The Dependence of Specific Star Formation Rate on Galaxy Size and Central Density at $0.5 < z < 2.5$ ”. In: *ApJ* 838.1, 19, p. 19. DOI: 10.3847/1538-4357/aa6258. arXiv: 1607.03107 [astro-ph.GA].
- Whitaker, Katherine E., Mariska Kriek, et al. (Feb. 2012). “A Large Population of Massive Compact Post-starburst Galaxies at $z < 1$: Implications for the Size

- Evolution and Quenching Mechanism of Quiescent Galaxies”. In: *ApJ* 745.2, 179, p. 179. DOI: 10.1088/0004-637X/745/2/179. arXiv: 1112.0313 [astro-ph.CO].
- Wuyts, Stijn et al. (Jan. 2007). “What Do We Learn from IRAC Observations of Galaxies at $2 < z < 3.5$?” In: *ApJ* 655.1, pp. 51–65. DOI: 10.1086/509708. arXiv: astro-ph/0609548 [astro-ph].
- Xiao, Mengyuan et al. (Sept. 2023). “Massive Optically Dark Galaxies Unveiled by JWST Challenge Galaxy Formation Models”. In: *arXiv e-prints*, arXiv:2309.02492, arXiv:2309.02492. DOI: 10.48550/arXiv.2309.02492. arXiv: 2309.02492 [astro-ph.GA].
- Zahid, H. J. et al. (Feb. 2013). “The Observed Relation between Stellar Mass, Dust Extinction, and Star Formation Rate in Local Galaxies”. In: *ApJ* 763.2, 92, p. 92. DOI: 10.1088/0004-637X/763/2/92. arXiv: 1211.7062 [astro-ph.CO].
- Zavala, J. A. et al. (Mar. 2021). “The Evolution of the IR Luminosity Function and Dust-obscured Star Formation over the Past 13 Billion Years”. In: *ApJ* 909.2, 165, p. 165. DOI: 10.3847/1538-4357/abdb27. arXiv: 2101.04734 [astro-ph.GA].
- Zuckerman, Leah D. et al. (Dec. 2021). “Reproducing the UVJ Color Distribution of Star-forming Galaxies at $0.5 < z < 2.5$ with a Geometric Model of Dust Attenuation”. In: *ApJ L* 922.2, L32, p. L32. DOI: 10.3847/2041-8213/ac3831. arXiv: 2109.14721 [astro-ph.GA].