

# **Exploring the Unknown: Searching for Unusual Objects with the James Webb Space Telescope**

Candidate Number: 217867  
Supervisor: Prof Stephen Wilkins

Word Count: 11,315  
Last Updated: May 7, 2024



**University of Sussex**

## Abstract

This project presents an investigation into the search for unusual objects from new James Webb Space Telescope (JWST) observations. Unusual objects offer valuable insights into the nature of the universe. By identifying and studying these anomalies, new phenomena can be discovered, existing theories can be refined, and our understanding of the universe can be expanded. The data used is a catalogue from the Cosmic Evolution Early Release Science (CEERS) survey, which uses Near-Infrared Camera (NIRCam) imaging to search for high redshift galaxies. This project displays the comparison of archival observations from the Hubble Space Telescope (HST) to the recent observations made by JWST. The goal is to find objects that are fully observed by JWST but are very faint or not observed by HST. These objects will have dramatically changed or suddenly appeared in the 15 years between the HST and JWST observations. Moreover, this project outlines the analysis of inspecting the spectral energy distributions (SEDs) of the objects in the CEERS catalogue, to search for sources with unusual spectral signatures. A mathematical metric called Euclidean distance is used to compare the recorded flux at given infra-red wavelengths to the average flux of the dataset. Both parts of the project give convincing results for the presence of unusual objects using the two methods described. It is suggested that further analysis, of the objects that are selected as unusual, is necessary to create a more detailed description of the objects and to understand what makes them unusual.

# Contents

<b>Preface</b>	<b>1</b>
<b>1 Introduction</b>	<b>2</b>
1.1 Why study unusual objects? . . . . .	2
1.1.1 Quasars and Active Galactic Nuclei . . . . .	3
1.1.2 Supernova . . . . .	3
1.1.3 Hanny’s Voorwerp . . . . .	4
1.2 JWST . . . . .	5
1.2.1 NIRCam . . . . .	6
1.3 Hubble . . . . .	7
1.4 Observations . . . . .	8
1.4.1 Transients . . . . .	8
1.4.2 Strange Objects . . . . .	9
1.5 CEERS . . . . .	11
<b>2 Hubble Dropouts</b>	<b>15</b>
2.1 Data . . . . .	15
2.2 Errors . . . . .	16
2.3 Objects faint in HST but bright in JWST . . . . .	17
2.4 Thumbnails . . . . .	19
<b>3 Unusual SEDs</b>	<b>23</b>
3.1 Euclidean Distance . . . . .	23
3.2 SED Inspection . . . . .	24
<b>4 Conclusion</b>	<b>29</b>
4.1 Future Work . . . . .	31
<b>Acknowledgments</b>	<b>34</b>
<b>Bibliography</b>	<b>34</b>
<b>A Python Code</b>	<b>38</b>
A.1 Hubble Dropouts Code . . . . .	38
A.2 Unusual SEDs Code . . . . .	46
<b>B Figures</b>	<b>52</b>

# List of Figures

1.1	Light curves of the main supernova types, showing the evolution of the absolute magnitude over time. See Karttunen et al., n.d. for types Ia, Ib, II-L and II-P; Modjaz, Gutiérrez, and Arcavi, 2019 for types Ic and IIb; and Nyholm et al., 2020 for type IIIn. . . . .	4
1.2	Hanny’s Voorwerp (shown as a green gas cloud below a galaxy) (W. Keel and the Galaxy Zoo Team, 2024). . . . .	5
1.3	NIRCam filter transmission plots displaying the operating wavelength of 29 filters across a wavelength range of 0.6 to 5 microns (STScI, 2016). The 29 NIRCam filters are split into the short wavelength and long wavelength channel, with 4 filter widths: extra-wide (W2), wide (W), medium (M), and narrow (N). The wavelength range of each individual filter is shown, showing overlap between some filters. . . . .	6
1.4	HST filter transmission plots, for ACS and WFC3, displaying the operating wavelength of 16 filters across a wavelength range of 0.2 to 1.7 microns(Monna et al., 2013). . . . .	7
1.5	Galaxy NGC 7469 imaged by NIRCam, showing the typical JWST diffraction spike (ESA/Webb, 2022). . . . .	10
1.6	Part of the Extended Groth Strip (M. Davis and A. Koekemoer, 2007). . . . .	12
1.7	Comparison of the $z \simeq 9$ galaxy EGS z910 40898 in NIRCam, HST, and Spitzer imaging (Finkelstein et al., 2023). . . . .	13
1.8	A plot from Donnan et al., 2023, showing the SED from a high redshift object (CEERS-93316), identified by the CEERS survey. The thumbnails of the object at each NIRCam wavelength are presented. . . . .	13
2.1	Plot of HST (F160W) vs JWST (F150W). . . . .	15
2.2	Plot of the flux errors from the HST (F160W) and JWST (F150W) datasets. . . . .	16
2.3	Plot of the cleaned flux errors from the F160W and F150W datasets. . . . .	17
2.4	The same plot in Fig. 2.1 but zoomed into the section of the plot which shows the objects that are bright in JWST but faint in HST. . . . .	18
2.5	A histogram showing only the data points with the ratio of the flux of each object for JWST and HST greater than 10, in the log scale. . . . .	18
2.6	Thumbnails at each HST and JWST wavelength, centred on the selected object. This image shows a diffraction spike. . . . .	19
2.7	An object which is outside the edge of the imaging for F125W and F160W HST bands. . . . .	19
2.8	A series of cutout thumbnails of object 8728. The image labelling follows the nomenclature stated in Section 1.2.1, Fig. 1.3. . . . .	20
2.9	The SED plot for object 8728. . . . .	20
2.10	A series of cutout thumbnails of object 1417. . . . .	21
2.11	The SED plot for object 1417. . . . .	21

3.1	A distribution of the flux of the calculated average SED for all the objects in the CEERS catalogue. . . . .	24
3.2	A diagram showing two points in 3-dimensional space with their co-ordinates on the relevant axis. Repeated Pythagorean theorem is used to calculate the ED value (Wikipedia, 2024). . . . .	25
3.3	The SED plot for object 7272 against the theoretical average SED, along with a plot which shows the difference in flux between the average and the measured flux. . . . .	26
3.4	The SED plot for object 1684 against the theoretical average SED, along with a plot which shows the difference in flux between the average and the measured flux. . . . .	27
3.5	A plot showing the ED value of each object, labelled by their ID. . . . .	28
4.1	A plot showing data point being clustered by an unsupervised machine learning model. There are anomalies which do not fit into the main two clusters, however, may fit into their own cluster. . . . .	32
B.1	Thumbnails of objects 301 and 597, selected in Chapter 2. . . . .	52
B.2	Thumbnails of objects 3917 and 3957, selected in Chapter 2. . . . .	53
B.3	Thumbnails of objects 4325 and 4797, selected in Chapter 2. . . . .	53
B.4	Thumbnails of objects 6972 and 7977, selected in Chapter 2. . . . .	54
B.5	SEDs of objects 301 and 597, selected in Chapter 2. . . . .	55
B.6	SEDs of objects 3917 and 3957, selected in Chapter 2. . . . .	56
B.7	SEDs of objects 4325 and 4797, selected in Chapter 2. . . . .	57
B.8	SEDs of objects 6972 and 7977, selected in Chapter 2. . . . .	58
B.9	Thumbnails of objects 813, 1681, 1682, 1683, and 1684, selected in Chapter 3. . .	59
B.10	Thumbnails of objects 1685, 2199, 2200, 3892, and 4100, selected in Chapter 3. .	60
B.11	SED of object 813, selected in Chapter 3. . . . .	61
B.12	SED of object 1681, selected in Chapter 3. . . . .	62
B.13	SED of object 1682, selected in Chapter 3. . . . .	63
B.14	SED of object 1683, selected in Chapter 3. . . . .	64
B.15	SED of object 1685, selected in Chapter 3. . . . .	65
B.16	SED of object 2199, selected in Chapter 3. . . . .	66
B.17	SED of object 2200, selected in Chapter 3. . . . .	67
B.18	SED of object 3892, selected in Chapter 3. . . . .	68
B.19	SED of object 4100, selected in Chapter 3. . . . .	69

# Preface

- Chapter 1 is a review. The articles reviewed in Sections 1.1, 1.3, 1.4, and 4.1 were selected by myself, based off an independent literature review. The source material in Sections 1.2 and 1.5 was suggested by my supervisor.
- The data used in Chapters 2 and 3 was provided by my supervisor and created by the CEERS team.
- The code created for the analysis in Chapters 2 and 3 was made from scratch, by myself.
- The code to produce the thumbnails in Sections 2.4 and 3.2 was provided by my supervisor and then adapted by myself.

# Chapter 1

## Introduction

In this project, new observations from the James Webb Space Telescope are used to conduct a search for unusual objects. The data consists of objects that have been identified by the CEERS team and this project analyses these objects to highlight anything unusual about these observations. Archival data of the same objects from the Hubble Space Telescope is used as a comparison to the JWST data to see if anything has changed between the two imaging epochs.

### 1.1 Why study unusual objects?

In astrophysical research, the observation of new or unusual objects plays a pivotal role in expanding the understanding of the nature of the universe. These unusual objects include properties that differ from expected observational outcomes. They may not necessarily be phenomena or objects that we cannot explain, but rather they pique interest by being different from the known patterns.

Throughout history, the human race has stared up to the sky constantly curious with what else is out there, looking to understand the mysteries of the universe. It is very well known that even the earliest civilisations, like the ancient Greeks, with their eager observational skills and numerous philosophical questions, studied the sky in great detail. They noticed a consistent pattern of stars on the night sky but some of these stars "wandered" overtime. These wandering stars turned out to be the planets in the solar system that we are so familiar with today. This discovery changed the way astronomers and philosophers perceived the universe and lead to further great discoveries about the solar system. One of these great discoveries was made by Copernicus in 1543. His breakthrough highlighted the fact that the Sun is one of the many stars that can be seen in the vast cosmos and that each star is the centre and host to planets, like Earth. This idea that the Earth is not the centre of the universe and that planets rotated around their suns is a fundamental landmark in astronomical history.

With the advent of early telescopes, extended objects, what were then ubiquitously referred to as nebulae, were discovered. The word for "cloud" or "fog" in Latin, the objects that we now consider nebulae, are vast clouds of gas and dust that are spread out over space where stars and planetary systems are born. Due to having young, hot stars emitting copious radiation, nebulae glow brightly and are even visible with the naked eye. In the 17th century, one of the first recorded nebulae was the Orion Nebula, as it is in the own galaxy and is one of the brightest nebula in the night sky. With the advancements of astronomical technology over centuries, nebulae were studied in great detail and in the 19th century, astronomer William Herschel catalogued thousands of nebulae and star clusters (Herschel, 1864). Herschel was the first to suggest that some nebulae were not just clouds of gas and dust within the own galaxy but were, in fact, separate galaxies located far beyond the Milky Way. Subsequently, in the 20th century, Edwin Hubble observed individual stars within spiral nebulae, like the Andromeda Nebula (now known as the Andromeda Galaxy) and the Triangulum Nebula (now known as the Triangulum

Galaxy). He measured the stars' distances from Earth using a method known as Cepheid variable star pulsation and found that these spiral nebulae were actually galaxies located far beyond the Milky Way (Edwin P Hubble, 1979; Edwin Powell Hubble, 1982). The search, discovery and investigation of unusual objects have paved the way in astronomical research.

### 1.1.1 Quasars and Active Galactic Nuclei

The observation and mapping of the solar system, and other star systems, was the predominant focus for astronomy for most of recorded history. It was not until the 20th century that astronomers stumbled upon some of the most, at the time, puzzling objects in the universe: quasars and active galactic nuclei (AGN). In the 1960s, sources of intense radio waves were being discovered by the first radio telescopes. These sources appeared star-like in optical telescopes but showed very unusual properties when it came to observations at other wavelengths, predominately in radio. It was found that these objects emitted an incredible amount of energy from the cores of distant galaxies and were named "quasi-stellar radio sources" or quasars for short. The redshifts of these quasars were measured by Maarten Schmidt in the early 1960s. He found that the redshift of the quasar 3C 273 had a value of  $z = 0.158$  (Greenstein and Schmidt, 1979). This measurement placed the quasar at billions of light-years away from Earth. This showed that quasars were not nearby objects or even in nearby galaxies, but were instead some of most distant and luminous objects in the universe. The quasars challenged the existing models of galactic dynamics and introduced a rigorous study into the nature of this phenomena. It was later discovered that quasars were powered by supermassive black holes at the centres of galaxies. As matter approaches the event horizon of a supermassive black hole, it accelerates, converting its gravitational potential energy into kinetic energy. This matter, with extremely high kinetic energy, generates intense heat and radiation, leading to the emission of an incredible amount of light across the electromagnetic spectrum, known as a quasar. The first idea of black holes was theorised by Albert Einstein and Karl Schwarzschild in Einstein's general theory of relativity, published in 1915. However, it was not until the 1960s that the first evidence for black holes emerged, specifically observations of the Cygnus X-1 black hole - high-mass star binary system. Therefore, the search and observation of the unusual objects, such as quasars, helped confirm the theories and early observations around black holes, creating a new era in astrophysical research.

### 1.1.2 Supernova

Another extremely important celestial phenomena, is the violent and dramatic death of a star, a supernova. A supernova is when a massive star reaches the end of its life and, due to an imbalance in forces, the star explodes, greatly increasing the volume and luminosity of the star. The history of supernova discovery starts thousands of years ago with one of the most famous supernovae being recorded in 1054 CE by Chinese astronomers. The astronomers named the appearance of the supernova the "guest star" as it was not common to their usual night sky. The supernova was the start of the formation of the famous Crab Nebula, which is still heavily observed to this day. As supernovae are extremely bright, they offer a great way to carry out observations on stars and galaxies that may have been too dim to previously observe.

Again it was not until the 20th century, when telescopes and astronomical techniques became advanced enough to study supernovae in detail. The work of Fritz Zwicky and Walter Baade in the 1930s was one of the key contributions to the understanding of supernovae (Baade and Zwicky, 1934b; Baade and Zwicky, 1934a). They used the 100-inch Hooker telescope at the Mount Wilson Observatory in California, United States to discover and classify many supernovae. As part of this classification, supernovae were divided into two main types based on their spectra: Type I, which where supernovae whose spectra did not show hydrogen lines, and Type II, whose spectra did show hydrogen lines. This extensive research laid the groundwork for

future research into supernovae. In the decades that followed, astronomers kept refining their classification strategy and developing new ways of observation.

An important breakthrough in supernova research was when astronomers discovered that a certain subset of supernovae, Type Ia supernovae, could be used as "standard candles" for measuring large distances in the universe (Branch and Tammann, 1992). By observing the luminosity of the Type Ia supernova, as measured on Earth, astronomers could calculate the intrinsic luminosity and, therefore, through wavelength redshift, the distance to the supernova. This was due to Type Ia supernovae having a very uniform light curve, exhibiting a consistent peak luminosity. This discovery allowed for cosmologists to study large-scale structures and the evolution of the universe. With thousands of Type Ia supernovae recorded, physicists today can expand the understanding of the history of the universe to help predict its future.

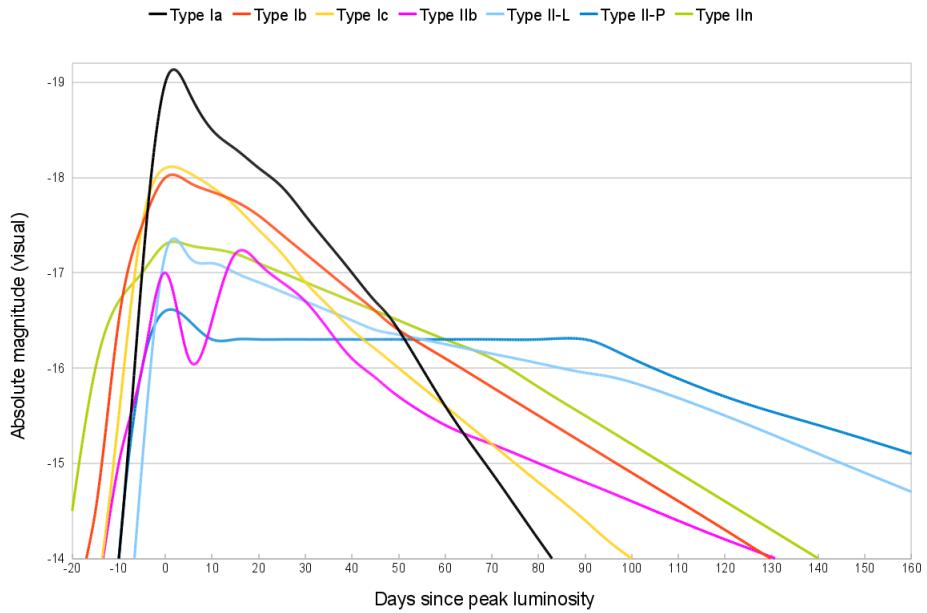


Figure 1.1: Light curves of the main supernova types, showing the evolution of the absolute magnitude over time. See Karttunen et al., n.d. for types Ia, Ib, II-L and II-P; Modjaz, Gutiérrez, and Arcavi, 2019 for types Ic and IIb; and Nyholm et al., 2020 for type IIIn.

Measuring the brightness of a supernova over time, across different wavelengths, will give the light curve of that supernova, as shown in Fig. 1.1. As a supernova is a very violent and, relative to cosmic times, a very quick process, the peak and post-peak decline can be measured over a number of days. It has been found that the way the brightness of the supernova behaves overtime is characteristic to the type of supernova. If an unusual object is identified, the brightness of the source over time can be measured to see if it shares properties of a supernova light curve.

### 1.1.3 Hanny's Voorwerp

Even today, new and unusual objects are found which help further the understanding of the universe. One example of this was the discovery of Hanny's Voorwerp by the citizen scientist, Hanny van Arkel. This object was deemed to be unusual and it confused some astronomers. It was a rare version of a quasar ionisation echo, Hanny's Voorwerp, or Hanny's Object in English. Hanny's Voorwerp was first brought to astronomical attention through observations with the Hubble Space Telescope, as part of the 'Galaxy Zoo' project (Józsa et al., 2009; Lintott et al., 2009). This project gave the public access to numerous HST images of galaxies for classification based on the shape of the galaxy.

The object, shown in Fig. 1.2, is named after Hanny, a Dutch schoolteacher who by chance discovered it in 2007. This interesting object did not follow conventional astronomical classifica-

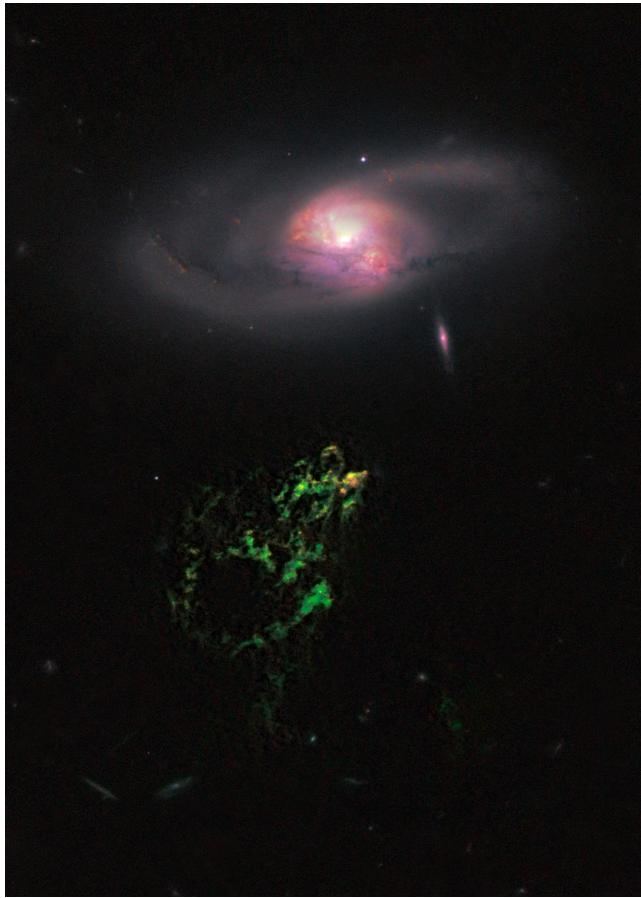


Figure 1.2: Hanny’s Voorwerp (shown as a green gas cloud below a galaxy) (W. Keel and the Galaxy Zoo Team, 2024).

tion. It shows the interaction between supermassive black holes and their surrounding gaseous environments. Hanny’s Voorwerp is characterised by its foggy green glow, a spectral signature of ionised oxygen, resulting from the radiation emitted by a nearby quasar, now dormant. The phenomenon gives information into the dynamics of quasar activity and its far-reaching consequences on the cosmic scale. Its detection highlights the importance of random discoveries in unveiling the mysteries of the universe and highlights the pivotal role of advanced telescopic instruments such as the Hubble Space Telescope in searching the galaxy.

## 1.2 JWST

As noted earlier, this project is focused on the search for unusual objects, specifically using new observations from the James Webb Space Telescope. JWST is the newest iteration of highly advanced space telescope technology. It was built to expand the understanding of the universe by observing the earliest galaxies, studying the lifetime of stars and planetary systems, pushing the boundaries of space exploration and astrophysical research. JWST started its mission on December 25th 2021, when it was sent up to the Sun-Earth Lagrange point,  $L_2$ . After reaching its halo orbit destination, JWST entered service on July 12th 2022.

Aboard JWST is the Integrated Science Instrument Module (ISIM) which includes a wide range of instruments for imaging and spectroscopy including: NIRCam, Near Infrared Spectrograph (NIRSpec), Mid-Infrared Instrument (MIRI), Fine Guidance Sensor (FGS), and Near Infrared Imager and Slitless Spectrograph (NIRISS). JWST specialises in infrared imaging, covering a wavelength range of  $0.6 - 28.5 \mu m$  (NASA, 2017).

### 1.2.1 NIRCam

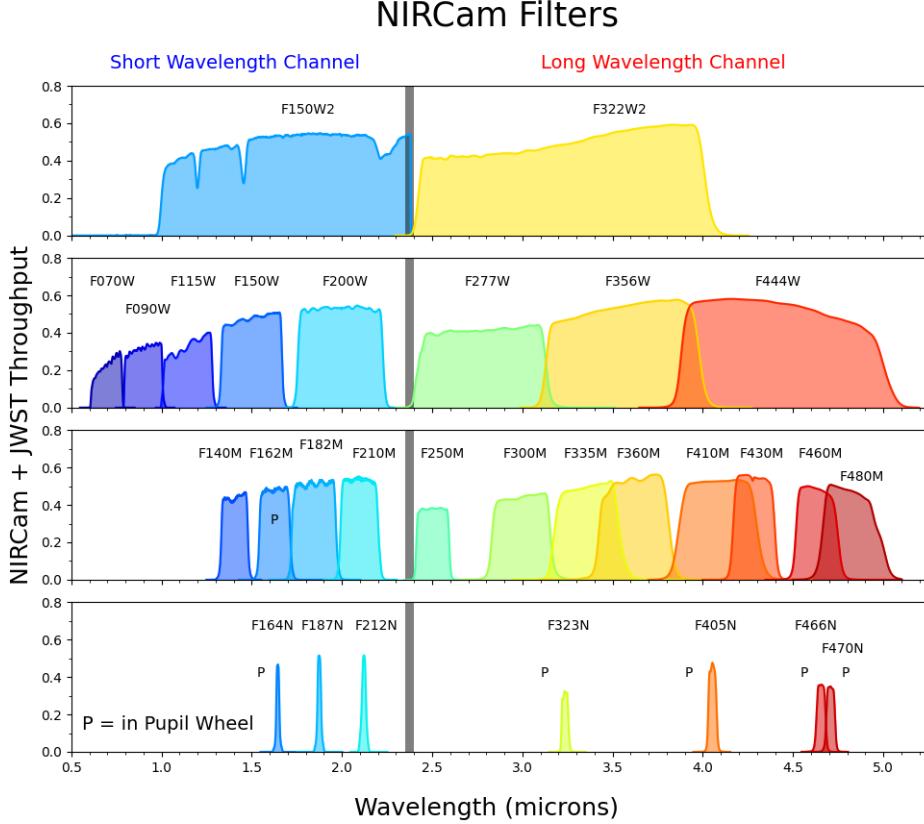


Figure 1.3: NIRCam filter transmission plots displaying the operating wavelength of 29 filters across a wavelength range of 0.6 to 5 microns (STScI, 2016). The 29 NIRCam filters are split into the short wavelength and long wavelength channel, with 4 filter widths: extra-wide (W2), wide (W), medium (M), and narrow (N). The wavelength range of each individual filter is shown, showing overlap between some filters.

One of the cameras on JWST is NIRCam, an imaging instrument operating at a wavelength range of  $0.6 - 5.0 \mu\text{m}$  (STScI, 2016). The wavelength range is split into two channels: Short Wavelength Channel ( $0.6 - 2.3 \mu\text{m}$ ) and Long Wavelength Channel ( $2.4 - 5.0 \mu\text{m}$ ). Fig. 1.3 shows all the different filters related to the wavelength range they cover. The rows are for the different widths of the filters: row 1 'W2' for extra-wide, row 2 'W' for wide, row 3 'M' for medium and row 4 'N' for narrow. To calculate the wavelength (in microns) from the filter name, divide by 100, so for example 'F115W' will be a wide band filter with central wavelength of 1.15 microns. This project will focus on 7 of the filter bands: F115W, F150W, F200W, F277W, F356W, F410M, and F444W.

NIRCam's main mission is to study the early universe, the formation and structure of stars and galaxies, and the properties of exoplanets. One of NIRCam's key features is that it has a very wide field of view. This allows it to capture large areas of the universe in a single exposure. This feature is crucial for studying large scale structures like galaxy clusters and conducting wide-field surveys to search for distant galaxies and exoplanets. The high sensitivity and wide field of view of NIRCam makes it a very powerful instrument for searching for unusual objects or detecting small changes in the universe.

## 1.3 Hubble

The Hubble Space Telescope stands as one of the most iconic and revolutionary tools in the history of astronomy. Launched into low Earth orbit in 1990, HST has provided astronomers with breathtaking images and invaluable data, significantly advancing the understanding of the cosmos. Serving as JWST's predecessor, HST paved the way for the next generation of space telescopes. Despite being primarily known for its stunning imagery, HST's scientific contributions are vast and varied, ranging from measuring the expansion rate of the universe to capturing detailed views of distant galaxies and nebulae. Its observations have led to groundbreaking discoveries in fields such as cosmology, galactic evolution, and the search for exoplanets. HST's success has set high expectations for JWST, with hopes that it will continue to push the boundaries of astronomical research and inspire generations to come.

The Hubble Space Telescope currently has 4 science instruments including the Advanced Camera for Surveys (ACS), Wide Field Camera 3 (WFC3), and the Comsic Origins Spectrograph (COS), and Space Telescope Imaging Spectrograph (STIS). The CEERS survey includes observations using both ACS and WFC3 providing coverage at both optical and near-IR (NIR) wavelengths, overlapping with the NIRCam observations.

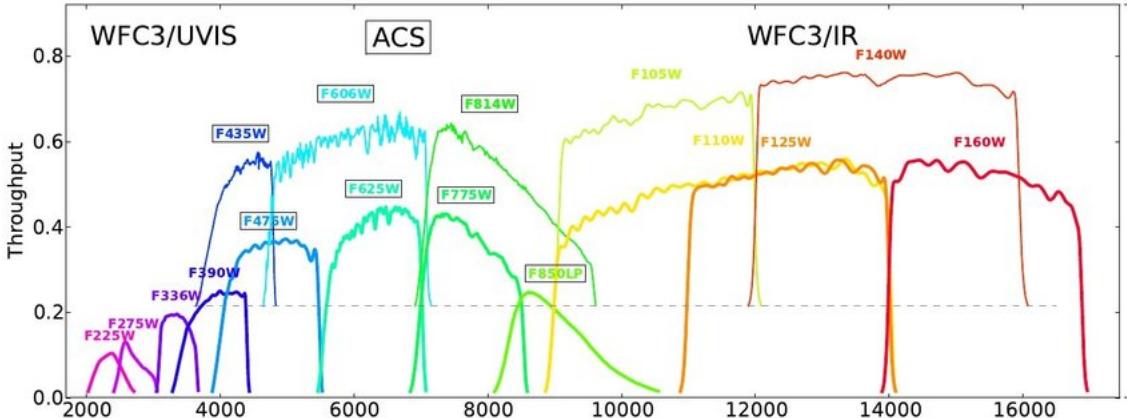


Figure 1.4: HST filter transmission plots, for ACS and WFC3, displaying the operating wavelength of 16 filters across a wavelength range of 0.2 to 1.7 microns(Monna et al., 2013).

The Advanced Camera for Surveys is an important instrument aboard the Hubble Space Telescope. The ACS, installed in 2002 during the Hubble servicing mission, is designed to capture high-resolution images across a wide field of view, enabling detailed observations of celestial objects ranging from nearby planets to distant galaxies. The Wide Field Channel (WFC), apart of the ACS, probes into the near-IR, with a wavelength range of  $0.350 - 1.050 \mu\text{m}$  (NASA/ESA, 2024a). Looking at Fig. 1.4 it can be seen that ACS overlaps with some of the NIRCam filters, meaning that previous HST observations can be compared with JWST observations to find any differences.

In addition, WFC3, installed during the final servicing mission in 2009, represents a significant upgrade to Hubble's imaging capabilities. With its two channels covering both the ultraviolet and near-IR regions, WFC3 allowed astronomers to explore celestial phenomena with unprecedented clarity and depth. Fig. 1.4 shows that WFC3 has an extended wavelength range of  $0.850 - 1.700 \mu\text{m}$  (NASA/ESA, 2024b). This wavelength range further overlaps with NIRCam and gives the ability to compare more observations of HST and JWST.

As the imaging epochs of the two telescopes is separated by around two decades, the wavelength overlap between WFC3 and NIRCam presents a powerful tool for astronomers to investigate objects that have changed over time. By combining observations from both instruments, astronomers can study these unusual objects across the near-IR spectrum. This approach allows

for a more detailed examination of celestial phenomena, such as asteroids, supernovae, and active galactic nuclei (described in Sections 1.4.1 and 1.4.2), whose characteristics may vary over time. Additionally, the optical range of HST complements these observations, enabling astronomers to easily check on these unusual objects, leveraging the unique capabilities of each instrument to gain a more comprehensive understanding of the universe.

## 1.4 Observations

### 1.4.1 Transients

As explained previously, unusual astrophysical objects can take many forms and do not have to be unusual for the same reasons. For this project, the objects are classified into two main categories: transient objects and strange objects. Transient objects are sources that will change overtime or move quickly with respect to the telescope. When comparing observations from different epochs of these transient objects, a change in the signal of these objects is to be expected. These objects include supernova, asteroids, and active galactic nuclei, to name but a few. Observing supernovae, asteroids and AGN with JWST holds great potential for advancing astrophysical research.

#### Infrared

JWST's infrared (IR) capabilities enable it to peer through the dusty clouds that often obscure the aftermath of supernova explosions (Shahbandeh et al., 2023) and the thick dust and gas surrounding AGN, providing a clearer view of their central regions. The advantage to see through dust allows astronomers to study the evolution of supernova remnants in great detail. Obscured AGN represent a significant fraction of the AGN population and are often missed by optical and X-ray surveys (Kirkpatrick, Alberts, et al., 2017). IR observation is crucial for studying the accretion disks and jets that power AGN, as well as the supermassive black holes at their cores. By observing AGN in the IR spectrum, researchers can probe the structure and dynamics of these high-energy phenomena with unprecedented detail, shedding light on the processes driving their immense energy output.

Observing AGN with JWST can provide valuable insights into the interplay between black hole accretion and galaxy formation (Goulding et al., 2023). AGN are known to influence the properties of their host galaxies through feedback processes, such as heating gas, triggering star formation, and driving galactic outflows. By studying the IR emission from AGN-hosting galaxies, researchers can trace the impact of AGN feedback on the surrounding interstellar medium and the star formation activity within these galaxies.

#### Sensitivity

The high sensitivity and resolution of JWST enables astronomers to detect and characterise a broader range of supernova types across cosmic distances and can provide critical data for characterising asteroids' physical properties, such as size, shape, and rotation rate.

While Type Ia supernovae have been extensively studied for their role in measuring cosmic distances and probing the expansion of the universe, other types, such as Type II and Type Ib/c supernovae, offer unique insights into the life cycles of massive stars and the mechanisms of their explosive deaths (Gutiérrez et al., 2017). By studying a more diverse sample of supernovae with JWST, astronomers can refine the understanding of stellar evolution and the conditions leading to different explosion mechanisms. In addition, JWST's high sensitivity allows it to capture the faint, red-shifted light from distant supernovae, extending the observational reach further back in time than ever before. This capability is crucial for tracing the evolution of supernova

rates and properties over cosmic history, shedding light on the role of these explosive events in shaping the chemical enrichment and energy balance of the universe across different epochs.

High-resolution infrared asteroid imaging from JWST can reveal detailed surface features and variations in brightness, allowing astronomers to create accurate models of asteroid shape and topography. This knowledge is essential for assessing the potential hazards posed by near-Earth asteroids and planning future asteroid exploration missions (Board, National Academies of Sciences, Medicine, et al., 2019). Furthermore, JWST's sensitivity and precision enable it to detect faint and distant asteroids, expanding the observational sample beyond the reach of ground-based telescopes (Lin and Loeb, 2014). By surveying a broader population of asteroids across different orbits and sizes, astronomers can investigate the diversity and distribution of asteroid populations in the solar system. This is crucial for studying the dynamical processes that shape the asteroid belt and influence the evolution of planetary systems.

The high sensitivity and resolution of JWST enables it to detect faint and distant AGN across cosmic distances. This capability is essential for studying the demographics and evolution of AGN populations over cosmic time. By analysing a wide range of AGN across different epochs of the universe's history, astronomers can investigate the growth of supermassive black holes, the triggering mechanisms of AGN activity, and the role of AGN in galaxy evolution (Satyapal et al., 2021; Jeon et al., 2023).

## Chemical

JWST's extensive array of imaging and spectroscopy instruments makes it excellent for observing the chemical compositions of astronomical phenomena.

By examining the spatial distribution and chemical composition of supernova remnants in diverse environments, astronomers can investigate the influence of factors such as metallicity, stellar mass, and interaction with companion stars on the outcome of stellar explosions. Also, by observing the IR signatures of elements formed in the core of a supernova explosion, such as iron, nickel, and silicon, researchers can learn about the processes driving these high-energy events and the subsequent dispersal of heavy elements into the interstellar medium.

JWST's IR capabilities offer a unique opportunity to study the composition and surface properties of asteroids across the solar system. By observing asteroids in the IR spectrum, researchers can investigate their mineralogical composition, thermal characteristics, and surface diversity (Müller et al., 2023; Gillet and Goresy, 2013; Harris and Lagerros, 2002). This information is crucial for understanding the formation and evolution of asteroids, as well as their role in the early solar system dynamics.

In addition, observing asteroids with JWST can contribute to the understanding of the origins of water and organic molecules in the solar system (Campins et al., 2010). Certain types of asteroids, known as carbonaceous or hydrated asteroids, are believed to contain significant amounts of water and organic compounds. By analysing the IR spectra of these asteroids, researchers can identify specific molecular signatures associated with water ice, hydrated minerals, and complex organic molecules. This knowledge not only sheds light on the delivery of volatiles to the early Earth, but also informs the search for habitable environments and extraterrestrial life in other planetary systems.

### 1.4.2 Strange Objects

Objects that are considered 'strange' are those, in JWST observations, which are a surprise. They are not expected, such as rare galaxies, and/or undesired, such as diffraction spikes. This can include artefacts in the imaging, where, due to the nature of the imaging apparatus, we see signal that is not actually present in the universe. An example of this is diffraction spikes, where the light from particularly bright sources spikes out, obscuring some of the image.

## Diffraction Spikes

Diffraction spikes, commonly observed in astronomical images, arise from the interaction of light with structures within the cameras. In the case of JWST, diffraction spikes come from starlight diffracted around the edges of the telescope's optical components (Rigby et al., 2023). These spikes appear as elongated patterns radiating from bright point sources, resembling the arms of a starfish. An example of a JWST diffraction spike is given in Fig. 1.5, showing the diffraction spike coming from the centre of the spiral galaxy NGC 7469.



Figure 1.5: Galaxy NGC 7469 imaged by NIRCam, showing the typical JWST diffraction spike (ESA/Webb, 2022).

Understanding and detecting diffraction spikes is crucial in astrophysical research for several reasons. Firstly, these spikes can significantly impact the accuracy of scientific measurements, particularly in studies involving precise photometry or astrometry. The presence of diffraction spikes around celestial objects can distort their apparent brightness and position, leading to incorrect conclusions about their properties and characteristics.

Moreover, the detection and removal of diffraction spikes is essential for enhancing the visual quality of astronomical images obtained with JWST. By eliminating these unwanted artefacts, researchers can obtain clearer and more accurate representations of the observed celestial objects. This, in turn, facilitates a more detailed analysis of astronomical phenomena.

Furthermore, removing diffraction spikes from images acquired by JWST enables astronomers to better identify and study faint objects located near bright sources. Without effective removal strategies, the presence of diffraction spikes can obscure or distort the features of dimmer celestial bodies, limiting the ability to study them comprehensively.

Finally, the precise characterisation and correction of diffraction spikes contribute to the overall calibration and refinement of JWST's imaging capabilities. By understanding the properties of these artefacts and implementing sophisticated algorithms for their removal, scientists can ensure the accuracy and reliability of the telescope's observational data.

## Rare Galaxies

Rare galaxies represent a unique and intriguing subset of celestial objects in the universe. These galaxies often possess unusual features, such as intense star formation activity, or exotic emission signatures, setting them apart from the majority of galaxies in the universe. Detecting and studying rare galaxies with JWST holds a great significance for advancing the understanding of galactic evolution, cosmology, and the underlying physical processes governing the formation and evolution of cosmic structures.

One compelling reason to detect rare galaxies with JWST is their potential to offer insights into the early universe (Steinhardt, Jespersen, and Linzer, 2021). Given their distinctive characteristics, rare galaxies may serve as cosmic "fossils," preserving valuable information about the conditions and events that prevailed during the early universe. The advanced IR sensitivity and imaging capabilities of JWST enable astronomers to peer deeper into the universe than ever before, allowing for the detection of faint, distant rare galaxies that would have remained beyond the reach of previous telescopes.

Rare galaxies often exhibit extreme properties that challenge existing theoretical models of galaxy formation and evolution. These galaxies may contain supermassive black holes, undergo violent mergers, or experience intense bursts of star formation, leading to the emergence of exotic phenomena such as quasars, starburst galaxies, or ultra-luminous IR galaxies (ULIRGs) (Ferrara, Pallottini, and Dayal, 2023). The high-resolution spectroscopic capabilities of JWST's instruments, coupled with its large aperture and sensitivity, facilitate detailed spectroscopic studies of rare galaxies, enabling astronomers to probe their chemical compositions, dynamics, and physical conditions with unprecedented precision.

Unusual galaxies offer a glimpse into the diversity of cosmic environments and the complex interplay between different astrophysical processes (Fudamoto, Inoue, and Sugahara, 2022). By studying the properties and distributions of rare galaxies across cosmic time, astronomers can probe the influence of environmental factors, such as galaxy interactions, mergers, and feedback from active galactic nuclei, on the formation and evolution of galactic structures. The multi-wavelength observational capabilities of JWST, spanning from the near-IR to the mid-IR, allow astronomers to investigate the complicated interactions between stars, gas, dust, and dark matter in rare galaxies, shedding light on their formation mechanisms and evolutionary pathways.

Detecting rare galaxies with JWST enables astronomers to explore uncharted regions of parameter space and discover novel astrophysical phenomena. The telescope's suite of advanced instruments, equipped with state-of-the-art detectors and spectrometers, empowers astronomers to systematically survey the sky for rare galaxies and uncover hidden gems that offer fresh insights into the rich tapestry of cosmic diversity.

New data from JWST provides the opportunity to look for all of these using a revolutionary dataset potentially opening up the possibility to find new and unusual phenomenon which shed new light on the understanding of the universe.

## 1.5 CEERS

The Cosmic Evolution Early Release Science survey is a programme which set out to take advantage of JWST to carry out extragalactic observations of the early universe. CEERS has and will provide unprecedented information about the faintest and most distant galaxies in the observable universe. Advanced near-IR imaging with HST has already revealed thousands of galaxies at  $z > 6$ , however, information about objects in the universe with  $z > 9$  is limited due to HST's capabilities (Finkelstein et al., 2023). JWST NIRCam opens the opportunity to explore galaxy formation at earlier epochs in great detail, offering near-IR sensitivity with a large increase in light-gathering power. The survey is being optimised for studying galaxies at a redshift of  $z > 9$  and galaxy assembly processes (Bagley et al., 2023). CEERS comprises a mosaic of 10 NIRCam pointings in the Cosmic Assembly Near-infrared Deep Extragalactic

Legacy Survey (CANDELS) Extended Groth Strip field. Six of these pointings were obtained alongside prime NIRSpec observations, while the remaining four were obtained in parallel with prime MIRI observations. Additionally, four of these pointings include NIRCam wide-field slitless grism spectroscopy. The light from objects of the early universe carries information about the primordial stages of the universe which could provide valuable data about the process of the galaxy evolution and the creation of cosmic structure over billions of years. Cosmologists can also use observations from the early universe to test theories and measure the cosmological factors when the universe had just started developing.

The observations will cover a patch of the sky called the Extended Groth Strip (EGS). This is small area positioned between the constellations of Ursa Major and Boötes with a length of around 70 arcminutes and width of around 10 arcminutes. The strip contains at least 50,000 galaxies, all giving clues to the galaxy formation and clustering at different stages of the universe's life (M. Davis and A. Koekemoer, 2007). In 2004, the ACS aboard Hubble carried out 63 different pointings, taking images of the EGS to collate a mosaic of images of the strip. Fig. 1.6 shows part of the EGS imaged by Hubble.

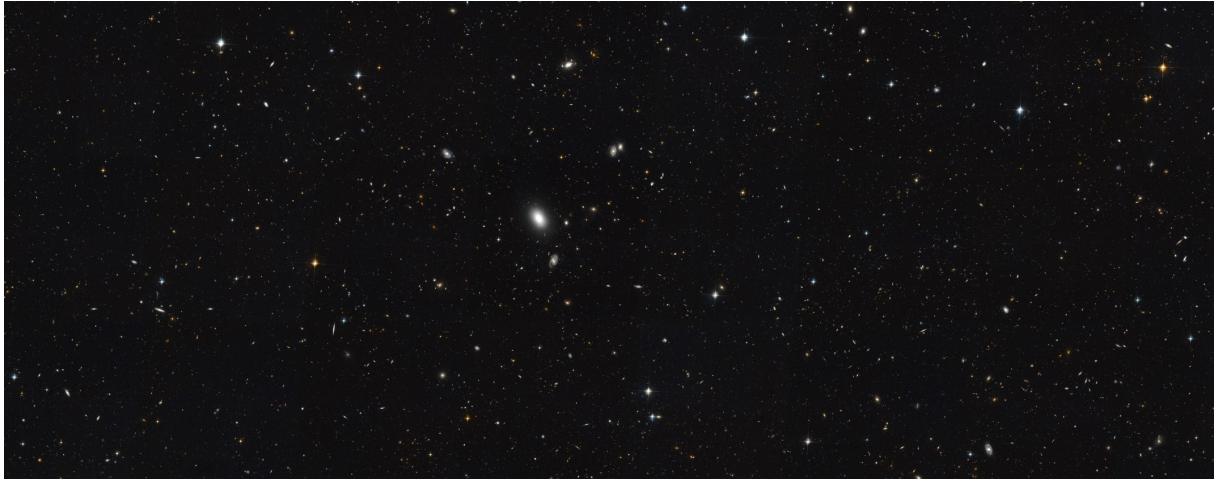


Figure 1.6: Part of the Extended Groth Strip (M. Davis and A. Koekemoer, 2007).

Eighteen years later, in 2022, JWST set out to take images of the EGS to carry on the EGS survey and use its new and improved instruments to gain more detail of the previously observed patch of sky. This gives data from not only two different time periods but also two different telescopes, with one having greater sensitivity and the ability to probe further into the earlier universe. This project will search for unusual objects by comparing the new JWST data from the CEERS programme to existing HST data, as well as investigating the JWST data in its own right.

Fig. 1.7, from Finkelstein et al., 2023, shows observations from HST, JWST and Spitzer, across a wavelength range of 0.6 - 5 microns. This figure aims to display the much improved imaging capabilities of JWST against the other telescopes. The galaxy, EGS z910 40898, was detected in HST with limited significance and hardly detected in Spitzer/IRAC observations, however, using NIRCam, the image is exceptionally well detected (signal-to-noise ratio >30) in all seven NIRCam filters. This shows the power of NIRCam's sensitivity and wide field of view to observe this high redshift galaxy in much greater detail across a wide wavelength range.

In this project, CEERS data, of observations of the EGS from the NIRCam module of JWST, is used and it contains a catalogue of 9212 objects of interest with, specifically, flux data from many different wavelength filters. In this catalogue the flux of each object, in  $nJy$ , is measured using 13 different filters. Only 7 filters are from the JWST recording period, including: F115W, F150W, F200W, F277W, F356W, F410M, F444W, as seen in Fig. 1.3. The other 6 filters are from HST, including: F606W, F814W, F105W, F125W, F140W, F160W, as seen

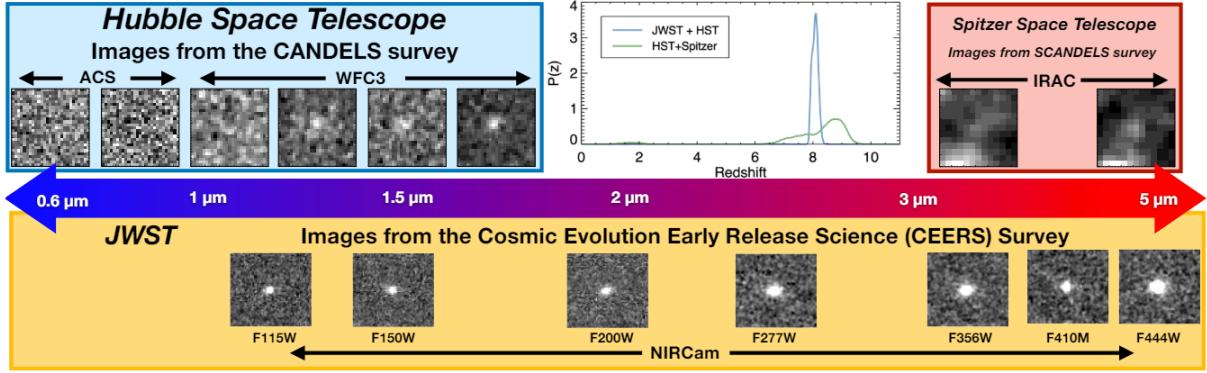


Figure 1.7: Comparison of the  $z \approx 9$  galaxy EGS z910 40898 in NIRCam, HST, and Spitzer imaging (Finkelstein et al., 2023).

in Fig.1.4. The HST data has been taken from previous observations of the EGS with ACS (F606W, F814W) and WFC3 (F105W, F125W, F140W, F160W) (Davis et al., 2007; Grogin et al., 2011; A. M. Koekemoer et al., 2011; Momcheva et al., 2016). The objects in the data catalogue had been found with the JWST imaging, the same objects were then searched for in the Hubble imaging, using coordinates from JWST observations. The data being comprised of HST data and JWST data gives the opportunity to analyse the two side by side and look for differences and discrepancies between the space telescopes. The catalogue includes multiple data sets, with information about the 9212 objects. The datasets that are of interest are all the flux datasets along with their corresponding flux error datasets. Also, there is an identification (ID) dataset which labels each object 1-9212 for easy identification.

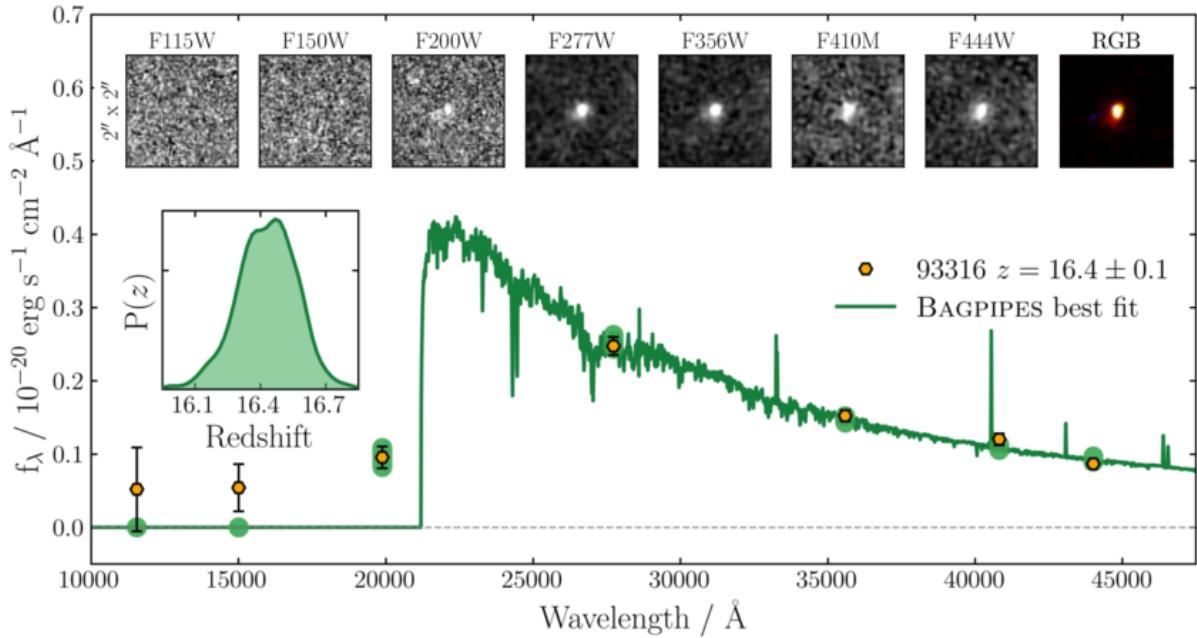


Figure 1.8: A plot from Donnan et al., 2023, showing the SED from a high redshift object (CEERS-93316), identified by the CEERS survey. The thumbnails of the object at each NIRCam wavelength are presented.

When observing and identifying objects in the universe, the spectral energy distribution of the object is analysed to look at the flux at certain wavelengths. CEERS-93316 is an object that was discovered by the CEERS team and the SED of this object is plotted in 1.8. The object

is a high redshift object, centred on  $z = 16.4$ , and is an example of the types of objects that is expected to be in the CEERS catalogue used in the project. The SED exhibits lower flux at the shorter wavelengths, due to the high redshift, with a peak in flux around F277W band. Also, as NIRCam probes deeper into the IR, the flux begins to reduce. Although not all objects will match with CEERS-93316, the SED shape of CEERS-93316 can serve as a template to validate the data integrity and reliability of the objects examined in this project.

# Chapter 2

## Hubble Dropouts

For the first category of unusual objects, a search was conducted for objects that are present in the JWST observations but missing or much fainter in the corresponding HST observations. Optimistically, such sources, if they exist, could be variable or transient phenomenon such as supernovae, AGN or solar system objects. As noted in the introduction, HST and JWST have overlapping capabilities when it comes to NIR imaging. Specifically, the Hubble/WFC3.F105W (F105W) and Hubble/WFC3.F125W (F125W) flank the JWST/NIRCam.F115W (F115W) filter, while the Hubble/WFC3.F160W (F160W) significantly overlaps with the JWST/NIRCam.F150W (F150W) filter. This allows for identification of objects which have changed in brightness at these wavelengths between the HST observations in 2005 and the more recent JWST observations. Because the catalogue of sources used NIRCam observations to detect objects, it is more likely to find sources that were fainter (or non-existent) in the HST imaging than vice versa.

### 2.1 Data

The F150W dataset from JWST and the F160W dataset from HST were selected as the flux from celestial objects should look very similar at these wavelengths. The F150W band was used as the baseline, looking for objects that were significantly present in F150W and much fainter in F160W.

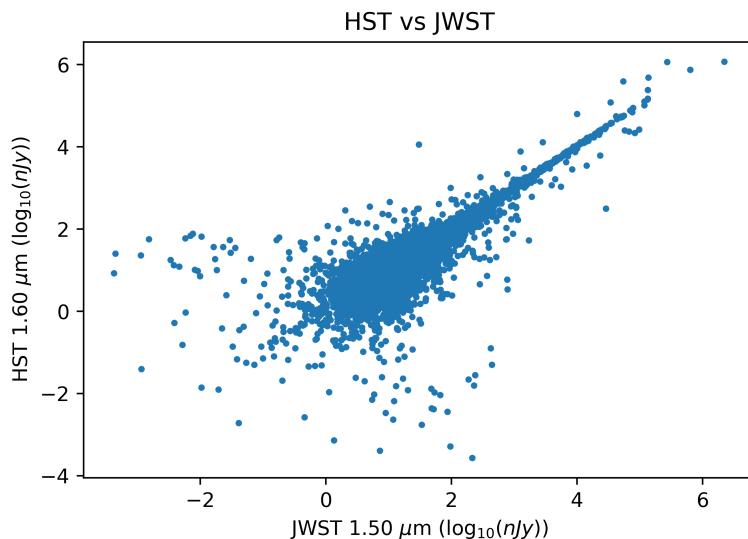


Figure 2.1: Plot of HST (F160W) vs JWST (F150W).

Fig. 2.1 shows a plot of the raw data from F160W and F150W. At the higher values of flux, the data follows a mostly linear relationship and, at the lower values of flux, the linearity begins to fall apart where the values of the object flux become mixed in with the image background noise. Furthermore, the plot shows that for bright objects (large flux) the measurement of said flux is more accurate due to the greater signal to noise. For the faint objects, the signal is comparable to the noise so the measurements are less accurate. From just this plot, there are no obvious ways to search for any anomalies within the data. To utilise this plot, the errors of the flux need to be considered.

## 2.2 Errors

With astronomy observations, the signal to noise ratio (SNR) is an vital feature of the data. The ratio is a quick and easy way to check the quality and reliability of the data sample. In this case, the signal is the measured flux of the object and the noise is the associated error. The main cause for uncertainty of the flux measurement is when the object is very faint and has similar flux to background sources and even electrical noise from the imaging instruments. As the SNR plays a crucial role in determining whether an object is bright and present in observations, the errors of the dataset need to be checked to ensure there are no problems and anomalies that will negatively influence the calculation of this ratio.

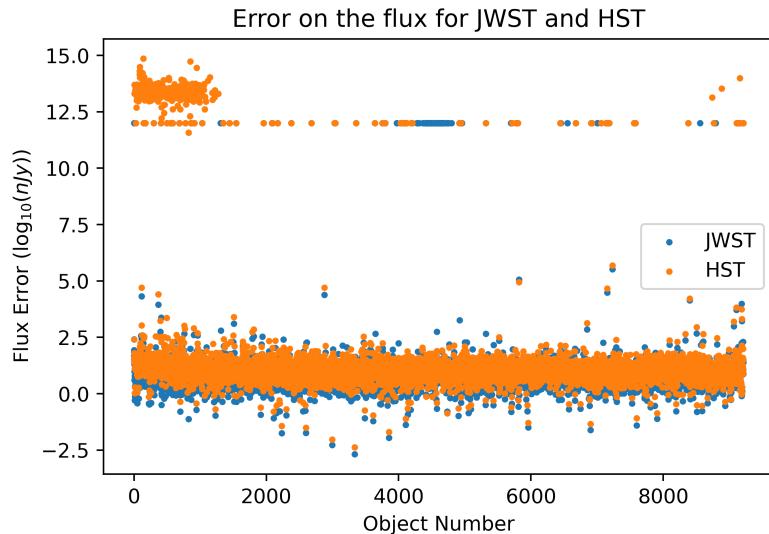


Figure 2.2: Plot of the flux errors from the HST (F160W) and JWST (F150W) datasets.

Fig. 2.2 shows the recorded error for each object from the F160W and F150W data. The plot shows a main band of the fluctuations in error but with clearly some problems, with many of the errors giving values which are larger than  $10^{10} nJy$ . These incredibly high values are most likely mistakes from the telescope imaging module or mistakes with the error calculation and processing. To give a more accurate picture of the errors on the flux, some data cleaning was done to take out these very large values in both the JWST and HST data sets.

Now that the data points with extreme error values have been omitted from the dataset, you can see in Fig. 2.3 that the error plot is giving a better description of the errors with the main band ranging from  $\sim 0 nJy$  to  $\sim 10^2 nJy$  and a maximum and minimum range of  $\sim 10^{-3} nJy$  to  $\sim 10^6 nJy$ , respectively.

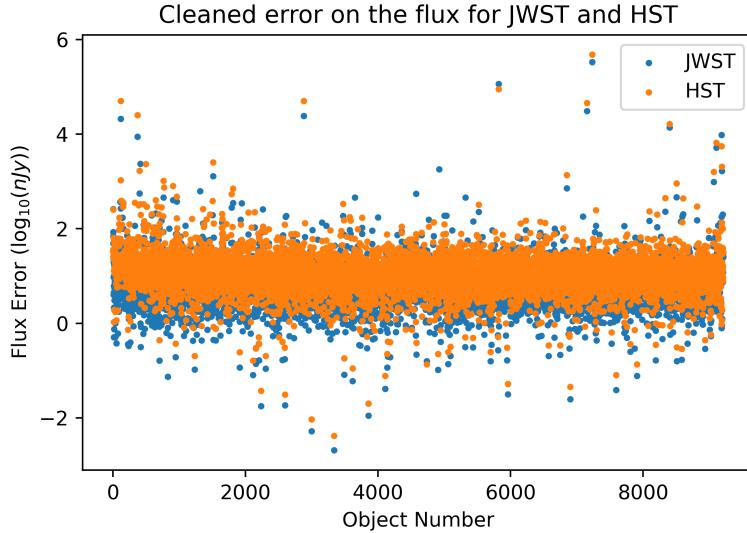


Figure 2.3: Plot of the cleaned flux errors from the F160W and F150W datasets.

### 2.3 Objects faint in HST but bright in JWST

A test was then carried out where the objects in F160W, which have a measured flux less than the corresponding error, were highlighted. This test highlights the objects with a flux value that is insignificant due to being the same or less than the recorded noise. These objects can be labelled as 'faint' or 'not present' at this wavelength. A mask was applied to all the data in Fig. 2.1 which, if less than the error, sets the flux values to the corresponding error value. This will adjust any flux values that have been recorded as extremely small.

The objects that are faint in F160W can now be seen in relation with their brightness captured by F150W. A lot of the objects follow the trend of the plot, the brightness of the object matches in HST and JWST. However, the objects of interest are the ones that appear bright in JWST but are faint in HST. These are interesting objects because why are the objects so much brighter when imaging with JWST compared to HST? As the imaging epochs are only around 15 years apart and the catalogue of sources are at high redshift, it is expected, for most sources, that there should be very little to no change in the flux when looking at F160W vs F150W.

Fig. 2.4 shows the highlighted 'faint' HST objects in orange and has zoomed in onto the part of the plot which shows the objects which should be faint in HST but bright in JWST. The other data follows with what is expected so is of no interest for this part of the project. The important part of this plot is the data points that are above 100 nJy in F150W (bright in JWST) and that are highlighted orange (faint in HST). This is a good way to quickly observe and visualise the data points of interest but a more robust method to extract the necessary points was created.

A histogram of the data was created, plotting the ratio between the flux values from JWST and HST (on a log scale). The ratio equation,

$$\text{Ratio} = \log_{10} \left( \frac{\text{HST } F150W \text{ data}}{\text{JWST } F160W \text{ data}} \right), \quad (2.1)$$

gave a histogram which had a peak around 0 (expected because in Eq. 2.1 log is taken after the ratio calculation) showing that most of the objects show similar flux in JWST and in HST. There was a slight tail to the right-hand side of the histogram. This tail showed there are some objects which have higher flux in JWST than in HST, agreeing with what we found in Fig. 2.4. The goal was to set some limits on this histogram to find the data points that were bright in

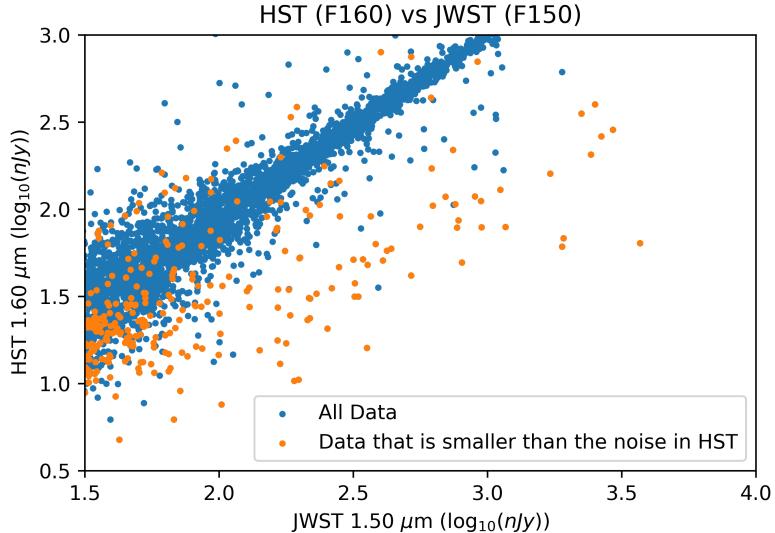


Figure 2.4: The same plot in Fig. 2.1 but zoomed into the section of the plot which shows the objects that are bright in JWST but faint in HST.

JWST and faint in HST and obtain a small enough number of data points that the individual spectral energy distributions could be evaluated for each. A limit was put in place that only included points that were at least 10 times or more brighter in F160W than in F150W.

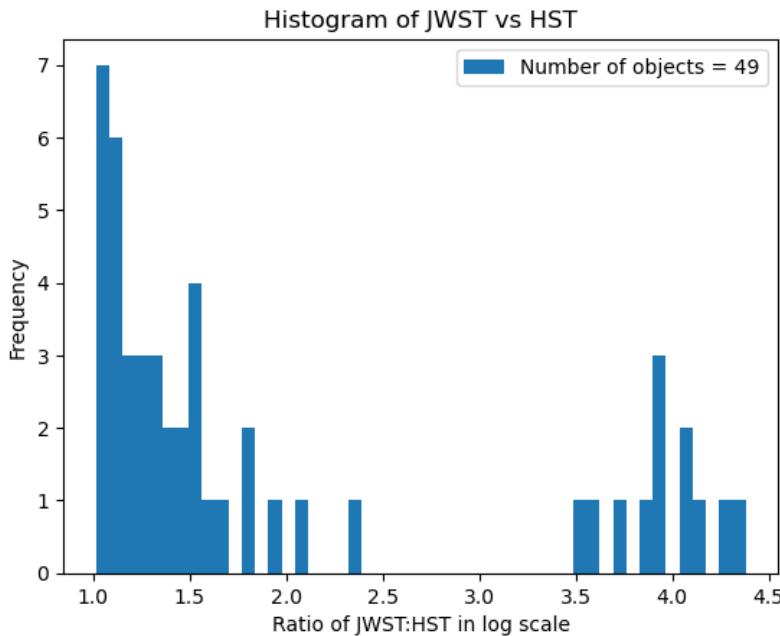


Figure 2.5: A histogram showing only the data points with the ratio of the flux of each object for JWST and HST greater than 10, in the log scale.

Fig. 2.5 shows a reduced histogram with the imposed limit given above. Only 49 objects out of the original 9212 were left and, by this method, are considered to be at least 10 times brighter in JWST than in HST. 49 was a small enough number that each data point was investigated to see why they are much brighter in JWST than in HST. Using the mask created to filter the 9212 objects down to these 49, the same mask was applied to the ID dataset to extract the IDs of these 49 objects to be used to look at the thumbnails of the images for further investigation.

## 2.4 Thumbnails

As part of the CEERS dataset, the co-ordinates of each object, in terms of x and y position, are given. Using python, a cutout of the observations, centred on the object, was created to see what the selected object looks like at a given wavelength. The F160W cutout and the F150W cutout were the main point of focus as these were the wavelengths used in the calculations above, however, the cutouts from all wavelengths were also made. Being able to look at the selected object at all wavelengths gave another check to make sure that the reason that these objects have been picked out was not due to an error in the F150W or F160W band.

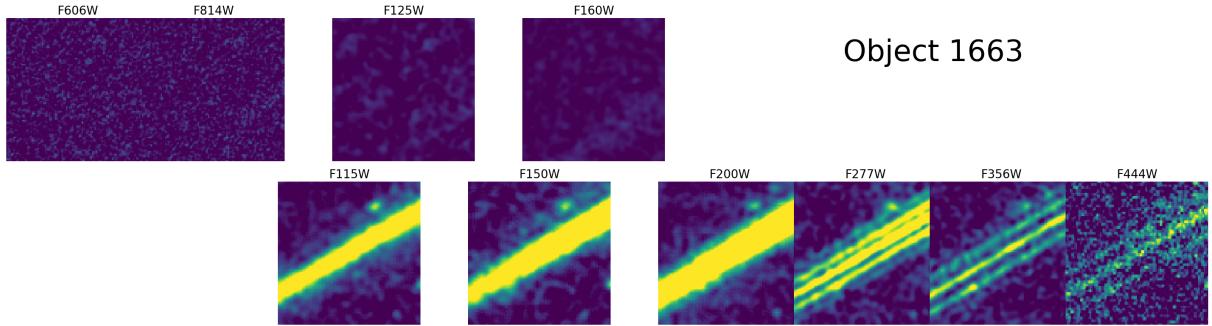


Figure 2.6: Thumbnails at each HST and JWST wavelength, centred on the selected object. This image shows a diffraction spike.

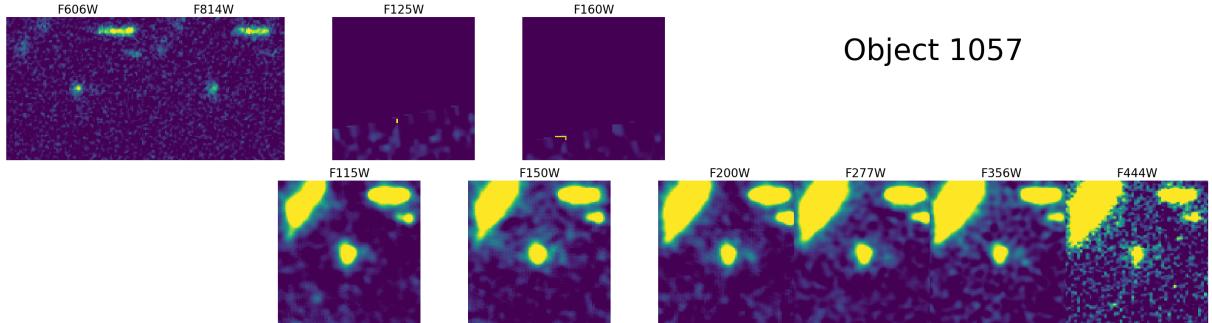


Figure 2.7: An object which is outside the edge of the imaging for F125W and F160W HST bands.

All 49 objects were inspected using visual inspection and the dataset was reduced down 10 candidates. The individual image labelling in the thumbnail plots follows the nomenclature stated in Section 1.2.1, Fig. 1.3. The other 39 objects were disregarded due to multiple reasons including diffraction spikes (Fig. 2.6) or the object was on the edge or outside, of the imaging (Fig. 2.7). It is still important to note down the IDs of the objects which have been disregarded as collecting groups of images with diffraction spikes or errors in the imaging can be important for cleaning up the dataset.

Fig. 2.8 is one of the best examples of an object that does not seem to have any problems with the imaging and shows a clear source in the F150W band and not in the F160W band. Seeing a clear difference between the image from HST than from JWST confirmed that the code, to find the objects that were not in HST but were present in JWST, performed correctly. A second way to look at the emitted flux of an object at different wavelengths is to investigate the spectral energy distribution. A SED is a plot of the energy (flux) against the wavelength. Using the SEDs of the objects, the difference in the flux between HST and JWST was observed numerically. When creating the SEDs for this project, the actual value of the flux was not as important as how each band compared with each other. Therefore, all the flux values were

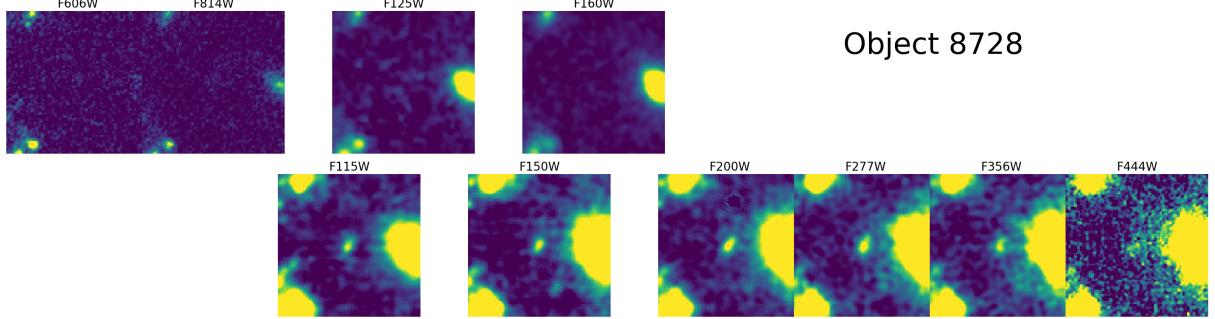


Figure 2.8: A series of cutout thumbnails of object 8728. The image labelling follows the nomenclature stated in Section 1.2.1, Fig. 1.3.

normalised to the flux of the JWST/NIRCam.F277W (F277W) JWST band as this band was used as a reference for all other NIRCam bands in Finkelstein et al., 2023.

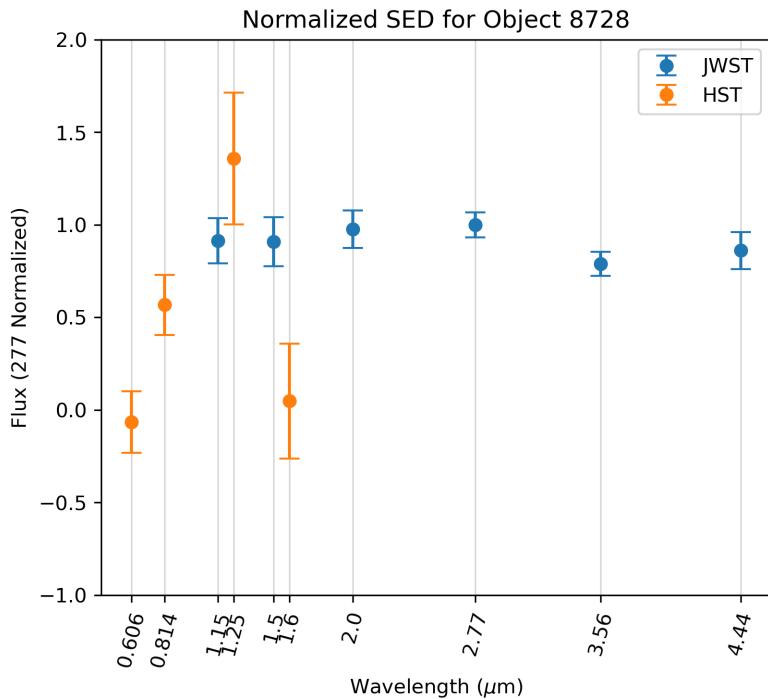


Figure 2.9: The SED plot for object 8728.

Looking at Fig. 2.9 it can be seen that the the F160W flux is definitely smaller than the recorded flux for F150W. However, the Hubble/WFC3.F814W (F814W) point and the F125W point (Hubble/WFC3.F606W (F606W) data is not significant as the flux value is negative, so can take this point as having zero flux) have similar flux to the JWST points. It is slightly suspicious that the F125W flux is very different to the F160W flux, due to these two wavelengths not being that dissimilar and being both recorded with WFC3. However, Fig. 2.8 and the flux between F160W and F150W still propose a solid argument that there is change between the HST imaging to the JWST imaging. Additionally, examining the F125W and F115W thumbnails supports the argument that there are sources brighter in JWST imaging than in HST imaging. A reason for not having complete consistency across the thumbnails and the SEDs could be that the flux data for F125W could have a wrong correction applied to it in processing or the pixel scaling on the thumbnails could be different giving false brightness.

For an example of where the F125W and F160W bands match up more closely, object 1417

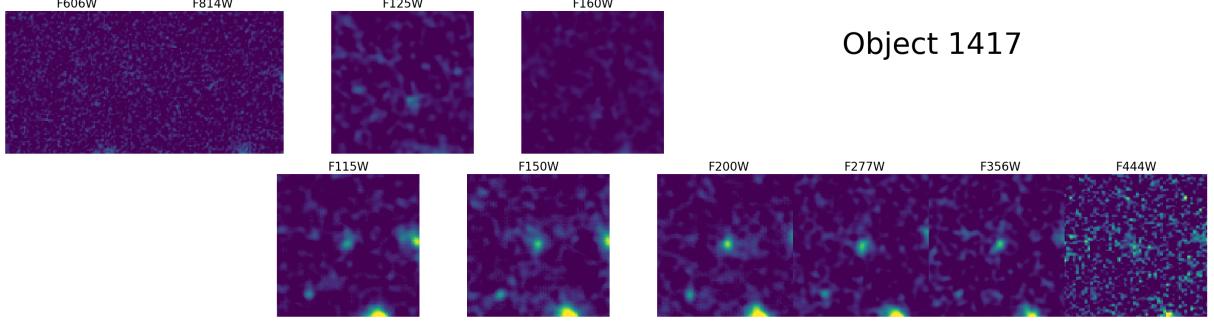


Figure 2.10: A series of cutout thumbnails of object 1417.

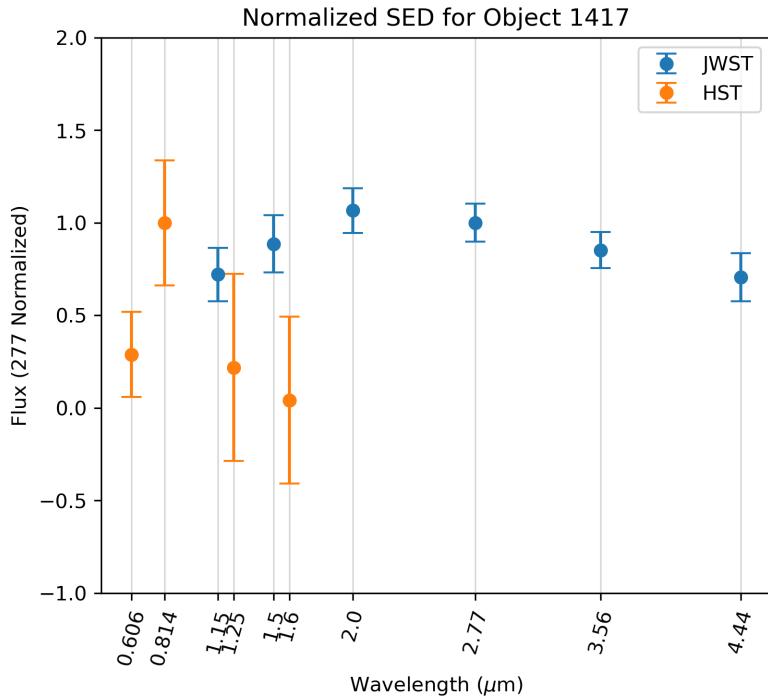


Figure 2.11: The SED plot for object 1417.

can be inspected. Fig. 2.10 shows the thumbnails for object 1417. The difference in HST vs JWST is not immediately as obvious compared to object 8728, however, it can be seen that there is a central object for JWST bands and basically nothing for the HST bands. Looking at the SED in Fig. 2.11, it can be seen that now the flux of the F125W band is much smaller and more comparable to the low F160W band. The error on F125W does match up with the flux of the F115W which does propose the difference in flux is not as significant, however, the flux in the F160W band is still significantly smaller than the flux in the F150W band.

In summary, a python script was created to search in the CEERS dataset for objects that were much brighter in JWST imaging than HST imaging. The 9212 objects were decreased to just 49 objects. These 49 objects were visually inspected to check for anomalies in the imaging such as edge effects or diffraction spikes. This further reduced the dataset to 10 possible candidates. The spectral energy distributions of each object were inspected alongside the wavelength thumbnails to confirm that the object was indeed brighter in JWST imaging than HST imaging. In all cases, the F150W band did indeed have a larger flux than the F160W, however, in some cases the value of the F150W or F160W flux did look suspiciously high or low, respectively, compared to the other bands. As suggested above, this could mean there was a small problem with the flux measurement; however, the values are not sufficiently large or small to categorise these

measurements as an anomaly. The 10 unusual objects can be further investigated through spectroscopy and more imaging at optical and infrared wavelengths to understand why they have been isolated by this analysis. This may uncover new unusual phenomena that has not yet been recorded and lead to a greater understanding of the history and structure of our universe.

# Chapter 3

## Unusual SEDs

As recognised in the introduction, objects can have a wide range of spectral energy distributions. These fluctuations in SED can result from various factors including temperature, chemical composition, and physical processes within the object. Although we have extensively measured the SEDs of galaxies, stars, and even rarer phenomena such as quasars and supernova, it is known that objects can display unique spectral signatures at higher redshifts, describing the early universe (Priddey and McMahon, 2001; Kirkpatrick, Pope, et al., 2012). In this chapter of the project, using the new data from just JWST, the SEDs of the objects in the catalogue were investigated and compared to an theoretical average SED to search for objects exhibiting unusual spectral signs.

### 3.1 Euclidean Distance

To look for objects that are different from the usual ones in the universe, the SEDs of these objects were compared to an average SED. Using the new JWST data from CEERS, an average value of the recorded flux was calculated at each JWST wavelength. The median of the flux values at each wavelength was used to calculate the average instead of the mean. This is because, when dealing with a dataset which potentially has extreme outliers, the median is much less affected by extreme values, while the mean can be heavily influenced by them. The median provides a more robust measure of central tendency, offering a better representation of the typical value in the dataset. These average values were combined to give an average SED of the most common spectral emission from the objects in the CEERS dataset. Fig. 3.1 shows this average SED, normalised by the F277W band. The SED looks as expected, comparing with Fig. 1.8, with smaller flux at the smaller and larger wavelength bands and a peak around the F277W band.

To calculate the difference in emitted flux from a select object to the theoretical average emitted flux, a mathematical metric called Euclidean distance (ED) was used (Xia et al., 2015; Liberti and Lavor, 2017; Krislock and Wolkowicz, 2012). Using just two points, A and B, with Cartesian coordinates in two dimensions, the ED,  $d$ , of these points is

$$d(A, B) = \sqrt{(A_1 - B_1)^2 + (A_2 - B_2)^2}, \quad (3.1)$$

where  $A_1$  denotes the coordinate of point A in dimension 1, and equivalent for the other variables. Fig. 3.2 is a diagram showing two points in 3-dimensional space and where the variables in Eq. 3.1 come from. Eq. 3.1 simply describes the distance between the two points in a two dimensional plane. The formula can be scaled up to N dimensions with the addition of the difference between the points in the respective dimension, as follows

$$d(A, B) = \sqrt{(A_1 - B_1)^2 + (A_2 - B_2)^2 + (A_3 - B_3)^2 + \dots + (A_N - B_N)^2}, \quad (3.2)$$

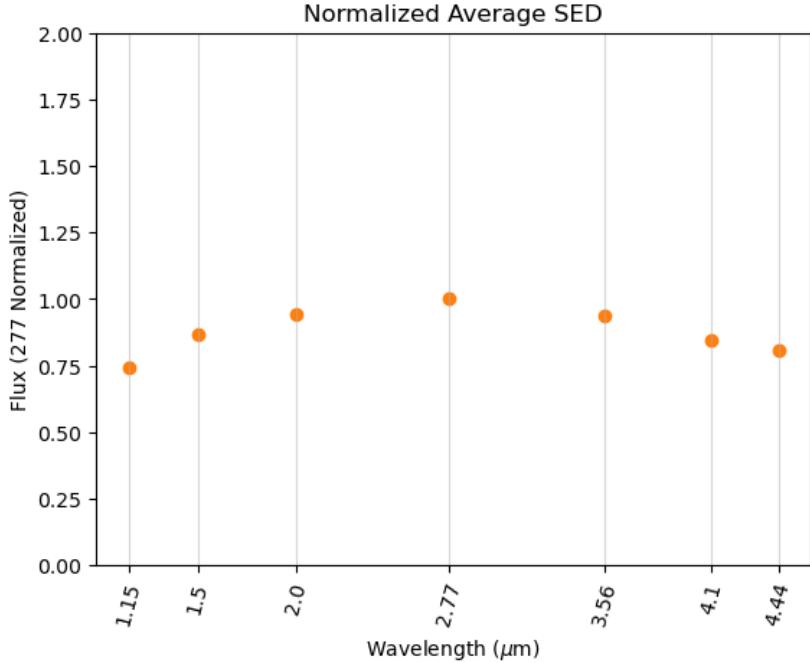


Figure 3.1: A distribution of the flux of the calculated average SED for all the objects in the CEERS catalogue.

where  $A_N$  denotes the coordinate of point A in the  $N^{th}$  dimension.

The reason why the ED metric was selected, to compare the objects in the dataset, was due to the different dimensions being represented by the different wavelengths of the dataset. So, using Eq. 3.2 as the example, point A is the theoretical average object flux and point B is the real object flux that is selected for comparison against the average object. The N dimensions relate to the 7 different JWST wavelengths:  $1.15 \mu m$ ,  $1.50 \mu m$ ,  $2.00 \mu m$ ,  $2.77 \mu m$ ,  $3.56 \mu m$ ,  $4.10 \mu m$  and  $4.44 \mu m$ . Therefore, the ED metric, calculated over the 7 dimensions, will give a value of 0 if the two points have the exact same coordinates and will positively increase with how different the two points are (up to a theoretical infinity). Also, as the mathematics of the ED are relatively simple (repeating the Pythagoras theorem) the computing power and time was low enough to quickly run through all 9212 objects in the catalogue. In the future, this would be advantageous due to the ability to scale up this test on datasets 10 or even 100 times larger than the one used in this project. All 9212 objects were run through this formula to give a list of ED values ordered from smallest to largest.

## 3.2 SED Inspection

With this metric, there is no value of the ED that decides whether something is different or not, it describes how different something is compared to something else. Therefore, instead of choosing a value of ED that was deemed to be different enough, the top 10 objects with the highest value of ED were selected for further investigation. After looking at the SEDs of the objects with the highest ED values, it was apparent that the objects with the highest ED values had errors in some of the data for specific wavelengths. Fig. 3.3 shows the object with the highest ED and, therefore, most dissimilar SED to the average one. Like in Chapter 2, the flux values have been normalised to the F277W band to just show the relative flux for visual clarity. The initial attempt showed objects which had very small values for some of the fluxes at single wavelengths, meaning the metric was thrown off for these objects. In the case of object 7272, the flux of the F277W band is dramatically smaller than the rest of the wavelength bands. This

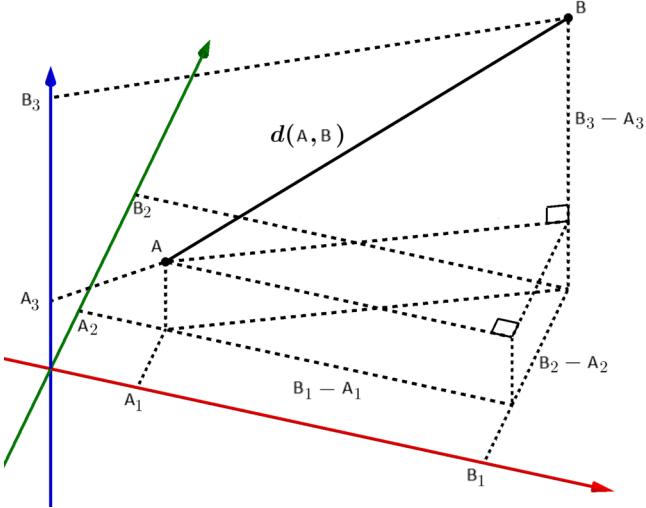


Figure 3.2: A diagram showing two points in 3-dimensional space with their co-ordinates on the relevant axis. Repeated Pythagorean theorem is used to calculate the ED value (Wikipedia, 2024).

is most likely a consequence of an error in the flux recording process and can be labelled as an anomaly. To resolve this, a constraint was applied to the dataset to only include the objects which has a SNR greater than 2 at all the JWST wavelengths. This filters out any objects with 'bad' data and just selects the objects which have 'good' data in all bands. The reduced dataset contains 9175 objects and the calculation code was run again.

Fig. 3.4 shows the SED for the objects which gave the highest value of ED and had a SNR value greater than two in all bands. When comparing Fig. 3.4 with Fig. 3.3, it can be seen that object 1684 does not have any obvious errors in each flux but still gives an SED which is different to the theoretical average. The source has significantly lower flux in 4 out of the 7 bands, meaning the object does not emit as much light at those wavelengths compared to the average. The lower plot in Fig. 3.4 gives the difference between the recorded flux and the calculated average flux. This is a quick way to numerically check how each wavelength band differs from the average.

Looking at the distribution of the ED for each object in Fig. 3.5 presents some interesting results. The ED of the objects seems to be very similar for most objects, with some having much larger values. This is what was expected, however, from this plot, it can be seen that an arbitrary threshold can be placed on the value of ED to filter out all many objects which are 'similar' to the theoretical average. This allows for a subset of objects to be created more analytically than, for example, just selecting the objects which have the top 10 largest EDs. For a threshold value of 4, there are 21 objects which have been filtered out. After checking the SEDs of these objects for any errors or anomalies, they can be pushed forward for deeper analysis into why their flux values differ from the norm.

In summary, the SEDs of objects from the new JWST observations, using the 7 wavelengths bands of NIRCam, were investigated to look for any sources which gave off unusual flux emissions. To test whether an SED is unusual or not, a mathematical metric called Euclidean distance was used to calculate the difference in flux, in multiple wavelength dimensions, from an object and theoretical average object. This theoretical average object was the construct of the average value of the measured flux by NIRCam at each wavelength. This allowed for an average SED to be created and each object in the CEERS catalogue was compared against this. Each object was assigned a ED value corresponding to by how much its SED differed from the average. Like in Chapter 2, the SEDs and thumbnails of the objects with the highest ED values were inspected.

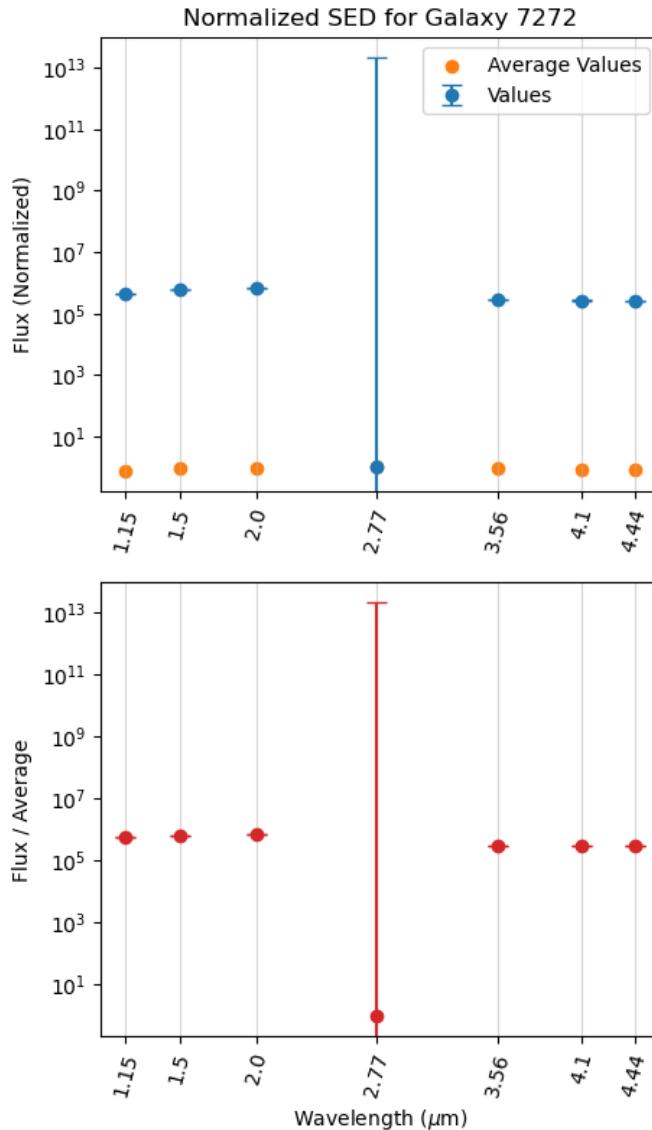


Figure 3.3: The SED plot for object 7272 against the theoretical average SED, along with a plot which shows the difference in flux between the average and the measured flux.

It was found that, due to errors in certain wavelength bands, the metric was being thrown off so a requirement of a SNR greater than 2 in all bands was put in place. An arbitrary threshold of 4 for the value of ED was chosen, from inspection of Fig. 3.5, which filtered out 21 objects which had high values of ED compared to the rest of the objects. Alongside the objects selected in Chapter 2, these objects can undergo further observation to figure out why they give off a different flux signature to the average object in the CEERS catalogue.

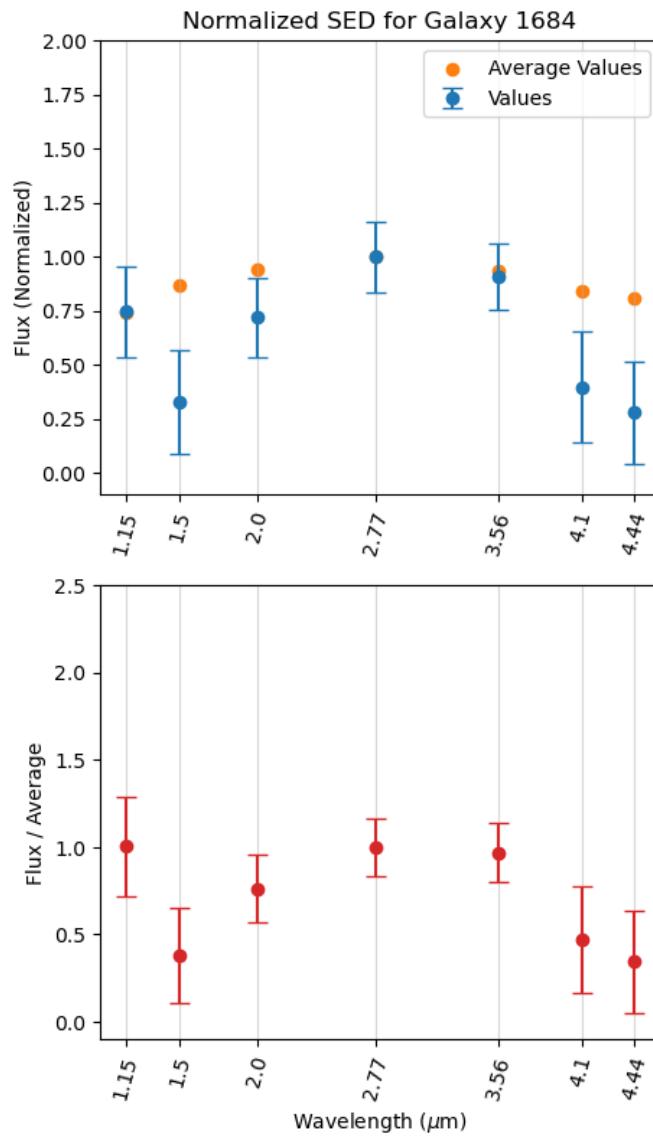


Figure 3.4: The SED plot for object 1684 against the theoretical average SED, along with a plot which shows the difference in flux between the average and the measured flux.

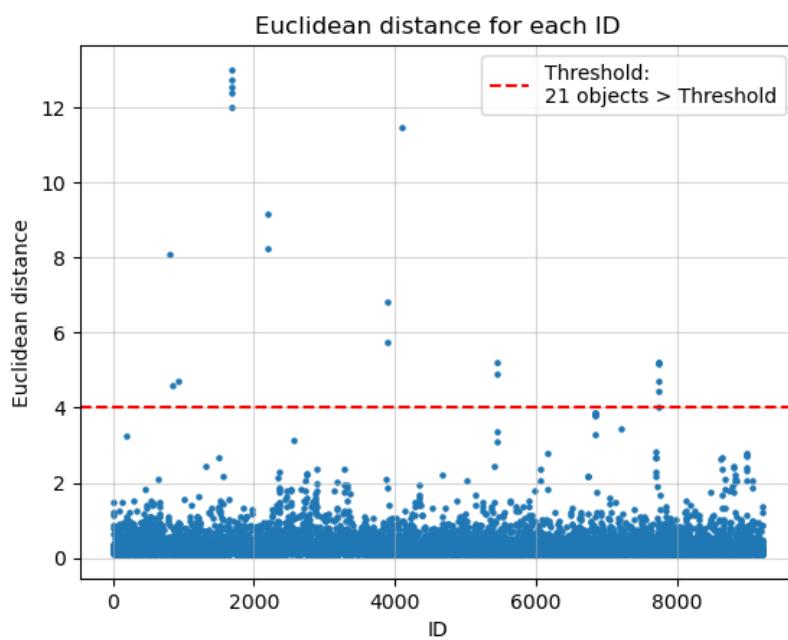


Figure 3.5: A plot showing the ED value of each object, labelled by their ID.

# Chapter 4

## Conclusion

This project set out to search for unusual objects using new JWST NIRCam imaging data from the Cosmic Evolution Early Release Science project. The search for unusual objects is crucial to the evolution of the field as it often leads to groundbreaking discoveries and expanding the understanding of the universe. CEERS is ideally suited to this search since the observations achieved unprecedented sensitivity in the rest-frame near-IR, thereby opening up new possibilities for discovery.

An innovative dataset has been used in this project to search for unusual objects using two methodologies: searching for Hubble dropouts and identifying galaxies with unusual spectral energy distributions.

### Hubble Dropouts

The first part of this project was to use the new observations from JWST to compare to archival observations from HST. As part of the catalogue created by the CEERS team, the flux of objects observed both by JWST and HST were recorded at the wavelengths covering the infrared spectrum. The HST observations came from ACS and WFC3, covering a wavelength range of  $0.606 - 1.60 \mu m$ , and the new observations from JWST came from NIRCam, covering a wavelength range of  $1.05 - 4.44 \mu m$ . One of the main reasons for comparing observations from HST to JWST, is that the HST imaging was done around 2005 and the imaging done by JWST took place in 2022. The 17-year gap between imaging epochs allows for the possibility that the sources imaged by HST may have changed or that new sources may have appeared, as measured by JWST. Another reason, is the much improved sensitivity and imaging capabilities allow JWST to detect objects that may have been just too faint for HST to detect and, also, give more detail on the objects that HST had previously detected.

The F150W and F160W bands were selected for comparison as they both fall within the IR spectrum and significantly overlap. Therefore, it is expected that if a source is detected in one band, it should also be detected in the other. The data was investigated and a need for data cleaning and filtering was identified. This included ensuring the uncertainty on the flux did not contain any values that were not physically possible and were most likely produced from errors in the data processing when creating the catalogue. Also, filtering using a SNR was carried out to establish a JWST dataset that included sources that had flux greater than the background noise. First, a visual inspection of the comparison between the F150W and F160W band was carried out to check if there were any candidates that fit the test. The results showed there were objects that appeared to be bright in the JWST imaging and faint in the HST imaging. This led to plotting a histogram of the ratio of fluxes between the two bands to analytically view the sources that had greater flux in F150W than F160W. A threshold was placed on the ratio between JWST and HST flux to obtain only the objects that were 10 times or brighter in the JWST imaging than in the HST imaging. The dataset was reduced down to just 49 objects that were selected for testing.

To confirm that the objects, selected from the initial testing, were real astronomical sources in universe, the thumbnails of the objects were inspected. The main culprits of error for this test, that can be found through the thumbnail inspection, are diffraction spikes or sources that are on the edge of the imaging. After all 49 objects were checked, only 10 were categorised to be real sources of interest. This meant that the images at all wavelengths looked clean without any defects or problems that could give false readings. Alongside the checking of the thumbnails, the SEDs of the objects were plotted to look at the numerical values of the bands (normalised to F277W). This allowed the flux values to be compared and also gave the uncertainty in the flux to see if the difference of F150W and F160W could be trusted. All 10 objects showed reasonable values of the flux at all wavelengths, however, some of the SEDs did show that the F150W or F160W band gave a value that was significantly lower or higher than the next corresponding band. It is expected that the flux of F160W is similar to flux of F125W and the flux of F150W is similar to F200W. Nonetheless, none of the flux values for F150W or F160W were too large or too small to be considered an anomaly, for now. More observations with NIRCam and spectroscopic observations with NIRSpec can be carried out to confirm the validity of the results and see if these 10 remaining objects are truly unusual.

## Unusual SEDs

The second part of the project looked at all the SEDs from the new JWST observations to see if there were any objects that had strange flux signatures. All objects in the universe give off flux over the electromagnetic spectrum. Over the history of astronomy the emitted flux or spectral energy distributions of celestial objects have been studied. Comparing the SEDs of the observed sources can categorise them and be used in identification when observing a new object. An object with a SED that does not fall into a specific category or class might present new insights into a category of phenomena or celestial bodies that have not yet been observed.

A metric called Euclidean distance was employed to decide how similar an object's SED is to an average theoretical SED. This average SED was simply made by calculating the median value of the flux at each given NIRCam wavelength. This resulted in the creation of an average SED which shows the central value of flux. It was found that using the median was the best method for calculating the average value as the mean was heavily influenced by extreme outliers. Every individual SED for all the objects in the CEERS catalogue were compared to this SED using ED and an array was created ranking all the objects by their ED value. After inspecting the SEDs of the objects with the highest ED values, it could be seen that some objects had extremely low values of flux at certain wavelengths. These were categorised as anomalies as the other bands showed reasonable results, so just the specific flux measurement was wrong. This forced the restriction of only using objects which had a SNR of greater than 2 in all bands. After repeating the analysis on the new dataset, the results looked much more convincing with no obvious errors.

The process of checking the thumbnails and SEDs of the selected objects described in Chapter 2 was repeated for this test. Due to the nature of this investigation, the thumbnails only gave information on whether the objects had image defects such as diffraction spikes or if they were on the edge of the imaging. The SED plots displayed the difference in flux to the average SED at each wavelength. The SEDs of the objects were checked to confirm that the ED metric worked as expected, giving the objects which differed the most from the average values. A distribution of the ED against the object ID was created to see if there was a value of ED that could be used as a threshold to determine whether an object is unusual or not. There was no clear value found from this distribution, however, it allowed for an arbitrary threshold value to be set. This threshold allows for a set number of objects to be labelled as unusual, however, more analysis and observations could help define a value of ED that is deemed to be a threshold for an object to be unusual.

## 4.1 Future Work

Astronomical research is notoriously open ended and, during the course of this project, limitations in the analysis and potential future directions have been identified. First, the data analysed in this project now represents only a small fraction of that available in the archive, thanks to numerous other large programmes. One immediate extension of this work would be to apply the same methodologies but over a much wider area; this would increase the likelihood of finding truly unusual objects.

For the Hubble Dropouts (chapter 2) specifically, a follow up project could involve doing the same test but for different overlapping HST and JWST wavelengths. For example, the JWST F115W and the HST F125W bands are very close in wavelength so it would be expected that the flux would be similar between the two bands such as in F150W and F160W. It would be interesting to compare the objects that appeared in F115W and not in F125W to the objects that appeared in F150W and not in F160W. If the objects were the same than this would confirm that the object is a new or definitely brighter object observed by JWST than HST. On the other hand, if there were some objects that only were selected by one of the tests, it could highlight more interesting features about this object but also highlight any errors made when conducting the test.

For the Unusual SEDs (chapter 3), the Euclidean distance (ED) is only one of many mathematical metrics that can be used to calculate the similarity between two data points (Singh, Yadav, and Rana, 2013). Measurements such as the Manhattan distance (MalkauThekar, 2013), Minkowski distance (Merigó and Casanovas, 2011), and the Mahalanobis distance (De Maesschalck, Jouan-Rimbaud, and Massart, 2000) are all similarity scores that could be calculated. The objects with the highest dissimilarity, selected by these metrics, could be compared to see if they all agree and if they do not, why not?

### Spectroscopic follow-up of unusual objects

To gain a clearer understanding of the nature of all of the unusual sources, one next step would be to obtain additional observations. These additional observations could include more imaging by NIRCam, which will now be at least 3 years later, looking at how these objects have evolved over time. Also, the observations could include using NIRSpec, for spectroscopic observations, to study features such as the chemical composition, temperature and density of the source.

For example, obtaining NIRCam observations at another epoch would allow us to confirm any source that is exhibiting significant variability. For example, a supernovae captured in the original CEERS imaging would, if observed now, appear significantly fainter. Fig. 1.1 gives information about how the brightness of all supernova changes over time. As the imaging for this project was carried out around 3 years ago, it would be expected that any supernova that had been observed would not show up in the second round of imaging. In addition, we can also seek to obtain spectroscopic follow-up allowing us to at least identify whether the object is within the solar system, Milky Way, or extragalactic in nature.

For unusual objects, a priority is to obtain deeper observations to confirm the objects unusual nature and then, ideally, spectroscopic observations and observations at different wavelengths. As mentioned previously, spectroscopic observations can provide information about certain features of a source that photometric imaging cannot. Also, observations at different wavelengths can reveal additional features not visible in a narrower wavelength range. Additionally, modelling the unusual sources using an SED fitting tool (which is used to infer the properties of galaxies or stars) can provide insights into their nature. Analysing the residuals of the best fit can refine the understanding of these unusual objects.

## Machine Learning

As an extension to this project, machine learning could be used to categorise objects based on their flux data. Machine learning can be split into supervised and unsupervised machine learning. Supervised machine learning is when a model is given some training data which contains data which is already labelled and the labels themselves. Types of clustering or categorising supervised machine learning models include: K-Nearest Neighbours (Peterson, 2009), Support Vector Machines (Steinwart and Christmann, 2008), and Decision Tree and Random Forests (Ali et al., 2012). The model will then train on this data to learn what to look for, and then categorise the data using the labels that were given. This requires human input of a dataset and labels that are already defined and can classify objects into predefined categories, allowing for accurate and targeted classification of objects. However, collecting and labelling the training data can be time-consuming and difficult. Therefore, using unsupervised machine learning, one can give the model all the new data and the model will categorise or cluster the data based off how similar the data is to each other. Types of clustering or categorising unsupervised machine learning models include: K-Means Clustering (Sinaga and Yang, 2020), Hierarchical Clustering (Murtagh and Contreras, 2012), and Gaussian Mixture Models (Reynolds et al., 2009).

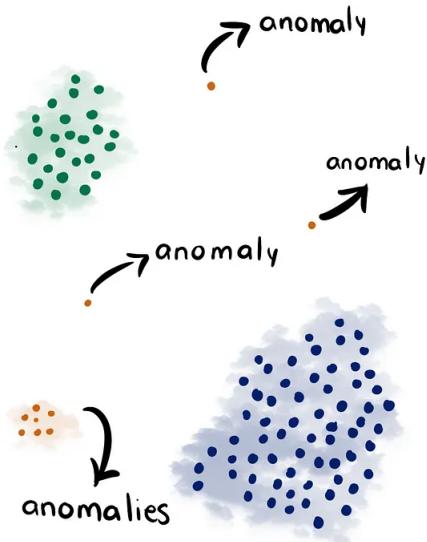


Figure 4.1: A plot showing data point being clustered by an unsupervised machine learning model. There are anomalies which do not fit into the main two clusters, however, may fit into their own cluster.

By using unsupervised machine learning in this project, a model can inspect all the new data and cluster it into groups based on their SED values. This should categorise most of the data into objects we already know (like galaxies and stars) and then any objects that are not put into these clusters can be called unusual. Fig. 4.1 shows an example of data being clustered by an unsupervised machine learning model. This shows how main clusters are formed, with the data points which are most similar to others, but there are anomalies that do not fit into these clusters. It is important to note that some of the anomalies can create their own small cluster, so even though they are anomalies, they share similar properties to one another. The advantage of this is that if all the unusual objects belong to the same cluster, once the type of one of these objects has been discovered, it is reasonable to assume that the other unusual objects are of the same type, thus not having to check all the anomalies and save time.

In summary, this project utilised new JWST imaging data from the Cosmic Evolution Early Release Science (CEERS) survey to search for unusual objects, aiming to expand the understanding of the early universe. Using two methodologies, this project focused on identifying Hubble dropouts and galaxies with unusual spectral energy distributions (SEDs). By comparing JWST and HST observations, the project identified objects exhibiting significant variability. Spectroscopic follow-up and observations at different wavelengths are crucial to confirming the unusual nature of these objects. Additionally, machine learning could be employed to categorise these objects based on their flux data, which could significantly enhance the efficiency of the analysis. Unsupervised machine learning clustering models, such as K-Means Clustering and Gaussian Mixture Models, could classify the data into familiar objects like galaxies and stars, thereby identifying any anomalies or unusual objects efficiently. These approaches could help identify and understand celestial objects that deviate from known patterns, providing invaluable insights into the structure and formation of objects in the early universe.

# Acknowledgements

Thank you to my supervisor, Prof Stephen Wilkins, for his guidance, planning the project and providing me with the data and necessary code to analyse it. Thank you to the CEERS team for collecting the data and organising it into a labelled catalogue.

# Bibliography

- Ali, Jehad et al. (2012). “Random forests and decision trees”. In: *International Journal of Computer Science Issues (IJCSI)* 9.5, p. 272 (cit. on p. 32).
- Baade, Walter and Fritz Zwicky (1934a). “Cosmic rays from super-novae”. In: *Proceedings of the National Academy of Sciences* 20.5, pp. 259–263 (cit. on p. 3).
- (1934b). “On super-novae”. In: *Proceedings of the National Academy of Sciences* 20.5, pp. 254–259 (cit. on p. 3).
- Bagley, Micaela B et al. (2023). “CEERS Epoch 1 NIRCam imaging: reduction methods and simulations enabling early JWST science results”. In: *The Astrophysical Journal Letters* 946.1, p. L12 (cit. on p. 11).
- Board, Space Studies, Engineering National Academies of Sciences, Medicine, et al. (2019). *Finding hazardous asteroids using infrared and visible wavelength telescopes*. National Academies Press (cit. on p. 9).
- Branch, David and GA Tammann (1992). “Type Ia supernovae as standard candles”. In: *Annual review of astronomy and astrophysics* 30.1, pp. 359–389 (cit. on p. 4).
- Campins, Humberto et al. (2010). “Water ice and organics on the surface of the asteroid 24 Themis”. In: *Nature* 464.7293, pp. 1320–1321 (cit. on p. 9).
- Davis, Marc et al. (2007). “The all-wavelength extended growth strip international survey (AEGIS) data sets”. In: *The Astrophysical Journal* 660.1, p. L1 (cit. on p. 13).
- De Maesschalck, Roy, Delphine Jouan-Rimbaud, and Désiré L Massart (2000). “The mahalanobis distance”. In: *Chemometrics and intelligent laboratory systems* 50.1, pp. 1–18 (cit. on p. 31).
- Donnan, CT et al. (2023). “The evolution of the galaxy UV luminosity function at redshifts z = 8–15 from deep JWST and ground-based near-infrared imaging”. In: *Monthly Notices of the Royal Astronomical Society* 518.4, pp. 6011–6040 (cit. on p. 13).
- ESA/Webb, NASA CSA (2022). *A Wreath of Star Formation in NGC 7469*. URL: <https://esawebb.org/images/potm2212a/> (visited on 05/07/2024) (cit. on pp. 4, 10).
- Ferrara, Andrea, Andrea Pallottini, and Pratika Dayal (2023). “On the stunning abundance of super-early, luminous galaxies revealed by JWST”. In: *Monthly Notices of the Royal Astronomical Society* 522.3, pp. 3986–3991 (cit. on p. 11).
- Finkelstein, Steven L et al. (2023). “CEERS key paper. I. An early look into the first 500 Myr of galaxy formation with JWST”. In: *The Astrophysical Journal Letters* 946.1, p. L13 (cit. on pp. 4, 11–13, 20).
- Fudamoto, Yoshinobu, Akio K Inoue, and Yuma Sugahara (2022). “Red Spiral Galaxies at Cosmic Noon Unveiled in the First JWST Image”. In: *The Astrophysical Journal Letters* 938.2, p. L24 (cit. on p. 11).
- Gillet, Philippe and Ahmed El Goresy (2013). “Shock events in the solar system: the message from minerals in terrestrial planets and asteroids”. In: *Annual Review of Earth and Planetary Sciences* 41, pp. 257–285 (cit. on p. 9).
- Goulding, Andy D et al. (2023). “UNCOVER: The Growth of the First Massive Black Holes from JWST/NIRSpec—Spectroscopic Redshift Confirmation of an X-Ray Luminous AGN at z = 10.1”. In: *The Astrophysical Journal Letters* 955.1, p. L24 (cit. on p. 8).

- Greenstein, Jesse L and Maarten Schmidt (1979). “The quasi-stellar radio sources 3c 48 and 3c 273”. In: *A Source Book in Astronomy and Astrophysics, 1900–1975*. Harvard University Press, pp. 811–818 (cit. on p. 3).
- Grogin, Norman A et al. (2011). “CANDELS: the cosmic assembly near-infrared deep extragalactic legacy survey”. In: *The Astrophysical Journal Supplement Series* 197.2, p. 35 (cit. on p. 13).
- Gutiérrez, Claudia P et al. (2017). “Type II supernova spectral diversity. I. Observations, sample characterization, and spectral line evolution”. In: *The Astrophysical Journal* 850.1, p. 89 (cit. on p. 8).
- Harris, Alan W and Johan SV Lagerros (2002). “Asteroids in the thermal infrared”. In: *Asteroids III* 205 (cit. on p. 9).
- Herschel, John Frederick William (1864). “Catalogue of nebulae and clusters of stars”. In: *Philosophical Transactions of the Royal Society of London*, pp. 1–137 (cit. on p. 2).
- Hubble, Edwin P (1979). *Extra-galactic nebulae*. Harvard University Press, pp. 716–724 (cit. on p. 3).
- Hubble, Edwin Powell (1982). *The realm of the nebulae*. Vol. 25. Yale University Press (cit. on p. 3).
- Jeon, Junehyoung et al. (2023). “Observability of Low-Luminosity AGN in the Early Universe with JWST”. In: *Monthly Notices of the Royal Astronomical Society*, stad1877 (cit. on p. 9).
- Józsa, GIG et al. (2009). “Revealing Hanny’s Voorwerp: radio observations of IC 2497”. In: *Astronomy & Astrophysics* 500.2, pp. L33–L36 (cit. on p. 4).
- Karttunen, Hannu et al. (n.d.). “Fundamental Astronomy [electronic resource]”. In: () (cit. on p. 4).
- Keel, W. and the Galaxy Zoo Team (2024). *Hanny’s Voorwerp*. URL: [https://en.wikipedia.org/wiki/Hanny%27s\\_Voorwerp](https://en.wikipedia.org/wiki/Hanny%27s_Voorwerp) (visited on 04/29/2024) (cit. on pp. 4, 5).
- Kirkpatrick, Allison, Stacey Alberts, et al. (2017). “The AGN–Star Formation Connection: Future Prospects with JWST”. In: *The Astrophysical Journal* 849.2, p. 111 (cit. on p. 8).
- Kirkpatrick, Allison, Alexandra Pope, et al. (2012). “GOODS-Herschel: impact of active galactic nuclei and star formation activity on infrared spectral energy distributions at high redshift”. In: *The Astrophysical Journal* 759.2, p. 139 (cit. on p. 23).
- Koekemoer, Anton M et al. (2011). “CANDELS: the cosmic assembly near-infrared deep extragalactic legacy survey—the Hubble Space Telescope observations, imaging data products, and mosaics”. In: *The Astrophysical Journal Supplement Series* 197.2, p. 36 (cit. on p. 13).
- Krislock, Nathan and Henry Wolkowicz (2012). *Euclidean distance matrices and applications*. Springer (cit. on p. 23).
- Liberti, Leo and Carlile Lavor (2017). *Euclidean distance geometry*. Vol. 3. Springer (cit. on p. 23).
- Lin, Henry W and Abraham Loeb (2014). “Finding rocky asteroids around white dwarfs by their periodic thermal emission”. In: *The Astrophysical Journal Letters* 793.2, p. L43 (cit. on p. 9).
- Lintott, Chris J et al. (2009). “Galaxy Zoo:‘Hanny’s Voorwerp’, a quasar light echo?” In: *Monthly Notices of the Royal Astronomical Society* 399.1, pp. 129–140 (cit. on p. 4).
- M. Davis, S. Faber and A. Koekemoer (2007). *Hubble Pans Across Heavens to Harvest 50,000 Evolving Galaxies*. URL: <http://hubblesite.org/newscenter/archive/releases/2007/06/image/a/> (visited on 03/26/2024) (cit. on pp. 4, 12).
- Malkauthkar, MD (2013). “Analysis of euclidean distance and manhattan distance measure in face recognition”. In: *Third International Conference on Computational Intelligence and Information Technology (CIIT 2013)*. IET, pp. 503–507 (cit. on p. 31).
- Merigó, José M and Montserrat Casanovas (2011). “A new Minkowski distance based on induced aggregation operators”. In: *International Journal of Computational Intelligence Systems* 4.2, pp. 123–133 (cit. on p. 31).

- Modjaz, Maryam, Claudia P Gutiérrez, and Iair Arcavi (2019). “New regimes in the observation of core-collapse supernovae”. In: *Nature Astronomy* 3.8, pp. 717–724 (cit. on p. 4).
- Momcheva, Ivelina G et al. (2016). “The 3D-HST Survey: Hubble Space Telescope WFC3/G141 grism spectra, redshifts, and emission line measurements for 100,000 galaxies”. In: *The Astrophysical Journal Supplement Series* 225.2, p. 27 (cit. on p. 13).
- Monna, A. et al. (Aug. 2013). “CLASH: z ~ 6 young galaxy candidate quintuply lensed by the frontier field cluster RXC J2248.7-4431”. In: *Monthly Notices of the Royal Astronomical Society* 438. DOI: [10.1093/mnras/stt2284](https://doi.org/10.1093/mnras/stt2284) (cit. on pp. 4, 7).
- Müller, Thomas G et al. (2023). “Asteroids seen by JWST-MIRI: Radiometric size, distance, and orbit constraints”. In: *Astronomy & Astrophysics* 670, A53 (cit. on p. 9).
- Murtagh, Fionn and Pedro Contreras (2012). “Algorithms for hierarchical clustering: an overview”. In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2.1, pp. 86–97 (cit. on p. 32).
- NASA (2017). *Webb Key Facts*. URL: <https://webb.nasa.gov/content/about/faqs/facts.html> (visited on 03/05/2024) (cit. on p. 5).
- NASA/ESA (2024a). *Hubble’s instruments: ACS - Advanced Camera for Surveys*. URL: <https://www.esahubble.org/about/general/instruments/acs/> (visited on 03/06/2024) (cit. on p. 7).
- (2024b). *Hubble’s Instruments: WFC3 - Wide Field Camera 3*. URL: <https://www.esahubble.org/about/general/instruments/wfc3/> (visited on 03/06/2024) (cit. on p. 7).
- Nyholm, Anders et al. (2020). “Type IIIn supernova light-curve properties measured from an untargeted survey sample”. In: *Astronomy & Astrophysics* 637, A73 (cit. on p. 4).
- Peterson, Leif E (2009). “K-nearest neighbor”. In: *Scholarpedia* 4.2, p. 1883 (cit. on p. 32).
- Priddey, Robert S and Richard G McMahon (2001). “The far-infrared–submillimetre spectral energy distribution of high-redshift quasars”. In: *Monthly Notices of the Royal Astronomical Society* 324.1, pp. L17–L22 (cit. on p. 23).
- Reynolds, Douglas A et al. (2009). “Gaussian mixture models.” In: *Encyclopedia of biometrics* 741.659-663 (cit. on p. 32).
- Rigby, Jane et al. (2023). “The science performance of JWST as characterized in commissioning”. In: *Publications of the Astronomical Society of the Pacific* 135.1046, p. 048001 (cit. on p. 10).
- Satyapal, Shobita et al. (2021). “The diagnostic potential of JWST in characterizing elusive AGNs”. In: *The Astrophysical Journal* 906.1, p. 35 (cit. on p. 9).
- Shahbandeh, Melissa et al. (2023). “JWST observations of dust reservoirs in type IIP supernovae 2004et and 2017eaw”. In: *Monthly Notices of the Royal Astronomical Society* 523.4, pp. 6048–6060 (cit. on p. 8).
- Sinaga, Kristina P and Miin-Shen Yang (2020). “Unsupervised K-means clustering algorithm”. In: *IEEE access* 8, pp. 80716–80727 (cit. on p. 32).
- Singh, Archana, Avantika Yadav, and Ajay Rana (2013). “K-means with Three different Distance Metrics”. In: *International Journal of Computer Applications* 67.10 (cit. on p. 31).
- Steinhardt, Charles L, Christian Kragh Jespersen, and Nora B Linzer (2021). “Finding High-redshift Galaxies with JWST”. In: *The Astrophysical Journal* 923.1, p. 8 (cit. on p. 11).
- Steinwart, Ingo and Andreas Christmann (2008). *Support vector machines*. Springer Science & Business Media (cit. on p. 32).
- STScI, Space Telescope Science Institute (2016). *JWST User Documentation - JWST Near Infrared Camera*. URL: <https://jwst-docs.stsci.edu/jwst-near-infrared-camera> (visited on 03/05/2024) (cit. on pp. 4, 6).
- Wikipedia (2024). *Euclidean Distance*. URL: [https://en.wikipedia.org/wiki/Euclidean\\_distance](https://en.wikipedia.org/wiki/Euclidean_distance) (visited on 04/30/2024) (cit. on pp. 5, 25).
- Xia, Shuyin et al. (2015). “Effectiveness of the Euclidean distance in high dimensional spaces”. In: *Optik* 126.24, pp. 5614–5619 (cit. on p. 23).

# Appendix A

## Python Code

### A.1 Hubble Dropouts Code

```
1  from General_Code import create_filter_dictionaries
2  import matplotlib.pyplot as plt
3  import numpy as np
4
5  plt.rcParams['figure.dpi'] = 500
6  plt.style.use('default')
7
8  # Call the function from the imported module
9  filters, filter_errors, hf, ID = create_filter_dictionaries()
10
11 # Set what filters you want
12 filter_JWST_num = 150
13 filter_HST_num = 160
14 filter_JWST = filters[filter_JWST_num]
15 filter_err_JWST = filter_errors[filter_JWST_num]
16 filter_HST = filters[filter_HST_num]
17 filter_err_HST = filter_errors[filter_HST_num]
18
19 #Calculate the ratio between the 2 filters
20 data_ratio = np.log10((filter_JWST)/(filter_HST))
21
22 #Plot the flux of the 2 filters against each other
23 plt.figure()
24 plt.title('HST vs JWST')
25 plt.scatter(np.log10(filter_JWST), np.log10(filter_HST), s=5)
26 plt.xlabel('JWST 1.50 $\mu$m ($\log_{10}(nJy)$)')
27 plt.ylabel('HST 1.60 $\mu$m ($\log_{10}(nJy)$)')
28 plt.show()
29
30 #Plot the errors from both filters
31 plt.figure()
32 plt.title('Error on the flux for JWST and HST')
33 plt.scatter(range(len(filter_err_JWST)), np.log10(filter_err_JWST), s=5, label='JWST')
34 plt.scatter(range(len(filter_err_HST)), np.log10(filter_err_HST), s=5, label='HST')
35 plt.ylabel('Flux Error ($\log_{10}(nJy)$)')
36 plt.xlabel('Object Number')
37 plt.legend()
38 plt.show()
39
#Set all the errors which are above 10^6 to 0 for cleaning
```

```

41     filter_err_JWST[filter_err_JWST > 10**6] = 0
42     filter_err_HST[filter_err_HST > 10**6] = 0
43
44 #Plot the new cleaned data
45 plt.figure()
46 plt.title('Cleaned error on the flux for JWST and HST')
47 plt.scatter(range(len(filter_err_JWST)), np.log10(filter_err_JWST), s=5, label='JWST')
48 plt.scatter(range(len(filter_err_HST)), np.log10(filter_err_HST), s=5, label='HST')
49 plt.ylabel('Flux Error ($\log_{10}(nJy)$)')
50 plt.xlabel('Object Number')
51 plt.legend()
52 plt.show()
53
54 # Create a mask for values that are below the error or 'nan'
55 mask_150 = np.logical_or(filter_JWST <= filter_err_JWST, np.isnan(filter_JWST))
56 mask_160 = np.logical_or(filter_HST <= filter_err_HST, np.isnan(filter_HST))
57
58 # Replace the corresponding values with the errors
59 filter_JWST[mask_150] = filter_err_JWST[mask_150]
60 filter_HST[mask_160] = filter_err_HST[mask_160]
61
62 #Plot the new distribution of flux between the two filters
63 plt.figure()
64 plt.title('HST (F160) vs JWST (F150)')
65 plt.scatter(np.log10(filter_JWST), np.log10(filter_HST), s=5)
66 plt.xlabel('JWST 1.50 $\mu m (\log_{10}(nJy))')
67 plt.ylabel('HST 1.60 $\mu m (\log_{10}(nJy))')
68 plt.show()
69
70 #Apply mask to data
71 plt.figure()
72 plt.title('HST (F160) vs JWST (F150)')
73 plt.scatter(np.log10(filter_JWST), np.log10(filter_HST), s=5, label='All Data')
74 plt.scatter(np.log10(filter_JWST[mask_160]), np.log10(filter_HST[mask_160]), s=5,
75             label='Data that is smaller than the noise in HST')
76 plt.xlabel('JWST 1.50 $\mu m (\log_{10}(nJy))')
77 plt.ylabel('HST 1.60 $\mu m (\log_{10}(nJy))')
78 plt.legend()
79 plt.show()
80
81 #Zoom into the part of the plot where the objects of interest are
82 plt.figure()
83 plt.title('HST (F160) vs JWST (F150)')
84 plt.scatter(np.log10(filter_JWST), np.log10(filter_HST), s=5, label='All Data')
85 plt.scatter(np.log10(filter_JWST[mask_160]), np.log10(filter_HST[mask_160]), s=5,
86             label='Data that is smaller than the noise in HST')
87 plt.xlabel('JWST 1.50 $\mu m (\log_{10}(nJy))')
88 plt.ylabel('HST 1.60 $\mu m (\log_{10}(nJy))')
89 plt.legend()
90 plt.xlim(0, 4)
91 plt.ylim(-1, 3)
92 plt.show()
93
94 #Zoom in further
95 plt.figure()
96 plt.title('HST (F160) vs JWST (F150)')
97 plt.scatter(np.log10(filter_JWST), np.log10(filter_HST), s=5, label='All Data')

```

```

97     plt.scatter(np.log10(filter_JWST[mask_160]), np.log10(filter_HST[mask_160]), s=5,
98                 ↪ label='Data that is smaller than the noise in HST')
99     plt.xlabel('JWST 1.50 $\mu$m ($\log_{10}(\mathrm{nJy})$)')
100    plt.ylabel('HST 1.60 $\mu$m ($\log_{10}(\mathrm{nJy})$)')
101    plt.legend()
102    plt.xlim(1.5, 4)
103    plt.ylim(0.5, 3)
104    plt.show()

105 #Filter by objects which are bright in JWST
106 mask_upper = (np.log10(filter_JWST) >= 1.5)
107 data_ratio_filt = data_ratio[mask_upper]

109 # Filter out 'inf' values from 'data_ratio'
110 inf_mask = ~np.isinf(data_ratio_filt)
111 data_ratio_inf = data_ratio_filt[inf_mask]

113 #Plot histogram of the ratio of the 2 filters
114 plt.figure()
115 plt.hist(data_ratio_inf, bins=1000, label = f'Number of objects = {len(data_ratio)}'
116             ↪ ')
117 plt.title('Histogram of JWST vs HST')
118 plt.xlabel('Ratio of JWST:HST in log scale')
119 plt.ylabel('Frequency')
120 plt.legend()
121 plt.show()

123 # Create a mask to filter out 'inf' values and values outside the range [1, 5]
124 data_ratio_mask = ~np.isinf(data_ratio) & (data_ratio >= 1) & (data_ratio <= 5) & (
125             ↪ np.log10(filter_JWST) >= 1.5)

127 # Apply the mask to filter the 'data_ratio' array
128 data_ratio_filt = data_ratio[data_ratio_mask]

130 #Plot a histogram with the filtered data
131 plt.figure()
132 plt.hist(data_ratio_filt, bins=len(data_ratio_filt), label = f'Number of objects =
133             ↪ {len(data_ratio_filt)})')
134 plt.title('Histogram of JWST vs HST')
135 plt.xlabel('Ratio of JWST:HST in log scale')
136 plt.ylabel('Frequency')
137 plt.legend()
138 plt.show()

139 Selected_ID = ID[data_ratio_mask]

140 print(Selected_ID)

```

```

2 import h5py
3 import numpy as np
4 import matplotlib.pyplot as plt
5 import matplotlib.cm as cm
6 from astropy.io import fits
7 from General_Code import create_filter_dictionaries

8 # Call the function from the imported module
9 filters, filter_errors, hf, ID = create_filter_dictionaries()
10

```

```

plt.style.use('default')

12
### Stephens Thumbnail Code (edited by me) ###

14
def make_cutout(data, x, y, width):
16
    """extracts a cut out from arbitrary data"""
18
    cutout = np.zeros((width, width))

20
    x = int(np.round(x, 0))
22
    y = int(np.round(y, 0))

24
    xmin = x - width // 2
    xmax = x + width // 2
26
    ymin = y - width // 2
    ymax = y + width // 2

28
    xstart = 0
30
    ystart = 0
    xend = width
32
    yend = width

34
    if xmin < 0:
        xstart = -xmin
        xmin = 0
36
    if ymin < 0:
        ystart = -ymin
        ymin = 0
40
    if xmax > data.shape[0]:
        xend -= xmax - data.shape[0]
        xmax = data.shape[0]
42
    if ymax > data.shape[1]:
        yend -= ymax - data.shape[1]
        ymax = data.shape[1]

46
    if (width % 2) != 0:
        xmax += 1
        ymax += 1

50
    data = np.array(data)
    cutout[xstart:xend,ystart:yend] = data[xmin:xmax,ymin:ymax]

54
    return cutout

56
def plot_cutout(ax, im, vmin=None, vmax=None, scaling=False, cmap=cm.magma):

58
    """plots an image cut out"""

60
    if vmin is None:
        vmin = np.min(im)
62
    if vmax is None:
        vmax = np.max(im)

64
    if scaling:
        im = scaling(im)

68
    ax.axis('off')

```

```

        ax.imshow(im, cmap=cmap, vmin=vmin, vmax=vmax, origin='lower')

70

72 if __name__ == "__main__":
    width = 100
    catalogue_file = r"C:\Users\Tom\OneDrive\Documents\A.Uni\Year 4\Final Year
    ↪ Project\Python\CEERS_NIRCam1_v0.2.h5"

76 id_list = [301, 597, 1417, 3917, 3957, 4325, 4797, 6972, 7977, 8728]

78 #4591, 44, 1057 is edge of imaging
80 #1798 possible diffraction spike
81 #476, 1663 big diffraction spike
82 #2109 very bright object
83 #5891 bright in 150

84 #id_list = [4325]

86 # id_list = [44, 92, 127, 288, 301, 449, 476, 494, 505, 517, 597, 637, 672,
87     ↪ 764, 855,
88 # 864, 975, 1057, 1067, 1073, 1417, 1510, 1615, 1645, 1651, 1654, 1663, 1798,
89 # 2042, 2043, 2044, 2093, 2109, 2294, 2471, 3917, 3957, 4325, 4422, 4591, 4797,
90 # 5891, 6253, 6972, 7336, 7977, 8549, 8605, 8728]

92 image_files = {
    'Hubble.F606W': r"C:\Users\Tom\OneDrive\Documents\A.Uni\Year 4\Final Year
    ↪ Project\Hubble Images\egs_all_acs_wfc_f606w_030mas_v1.9_nircam1_drz.
    ↪ fits",
    'Hubble.F814W': r"C:\Users\Tom\OneDrive\Documents\A.Uni\Year 4\Final Year
    ↪ Project\Hubble Images\egs_all_acs_wfc_f814w_030mas_v1.9_nircam1_drz.
    ↪ fits",
    'Hubble.F125W': r"C:\Users\Tom\OneDrive\Documents\A.Uni\Year 4\Final Year
    ↪ Project\Hubble Images\egs_all_wfc3_ir_f125w_030mas_v1.9_nircam1_drz.
    ↪ fits",
    'Hubble.F160W': r"C:\Users\Tom\OneDrive\Documents\A.Uni\Year 4\Final Year
    ↪ Project\Hubble Images\egs_all_wfc3_ir_f160w_030mas_v1.9_nircam1_drz.
    ↪ fits",
    'JWST/NIRCam.F115W': r"C:\Users\Tom\OneDrive\Documents\A.Uni\Year 4\Final
    ↪ Year Project\NIRcam Images\ceers_nircam1_f115w_sci_bkgsub_match.fits"
    ↪ ,
    'JWST/NIRCam.F150W': r"C:\Users\Tom\OneDrive\Documents\A.Uni\Year 4\Final
    ↪ Year Project\NIRcam Images\ceers_nircam1_f150w_sci_bkgsub_match.fits"
    ↪ ,
    'JWST/NIRCam.F200W': r"C:\Users\Tom\OneDrive\Documents\A.Uni\Year 4\Final
    ↪ Year Project\NIRcam Images\ceers_nircam1_f200w_sci_bkgsub_match.fits"
    ↪ ,
    'JWST/NIRCam.F277W': r"C:\Users\Tom\OneDrive\Documents\A.Uni\Year 4\Final
    ↪ Year Project\NIRcam Images\ceers_nircam1_f277w_sci_bkgsub_match.fits"
    ↪ ,
    'JWST/NIRCam.F356W': r"C:\Users\Tom\OneDrive\Documents\A.Uni\Year 4\Final
    ↪ Year Project\NIRcam Images\ceers_nircam1_f356w_sci_bkgsub_match.fits"
    ↪ ,
    'JWST/NIRCam.F444W': r"C:\Users\Tom\OneDrive\Documents\A.Uni\Year 4\Final
    ↪ Year Project\NIRcam Images\ceers_nircam1_f444w_sci_bkgsub.fits",
102
104 }

```

```

106     images = {}
107     for image_id, image_file in image_files.items():
108         hdu = fits.open(image_file)
109         images[image_id] = hdu[0].data
110
111         print(hdu[0].data.shape)
112
113     n_images = len(list(images.keys()))
114     print(n_images)
115
116     with h5py.File(catalogue_file, 'r') as hf:
117
118         for id in id_list:
119
120             # determine the x and y coordinate
121
122             s = hf['photom/ID'][()] == id
123
124             x = hf['photom/X'][s][0]
125             y = hf['photom/Y'][s][0]
126
127             x,y = y,x
128
129
130             print(hf['photom/ID'][s][0], x, y)
131
132             # Position values for each subplot
133             # positions = [(0.606, 0.8), (0.814, 0.8), (1.25, 0.8), (1.60, 0.8),
134             # (1.15, 0), (1.50, 0), (2.00, 0), (2.77, 0), (3.56, 0), (4.44, 0)]
135             #positions = [(x*2, y) for x, y in positions]
136             positions = [(0, 0.8), (0.5, 0.8), (1.2, 0.8), (1.9, 0.8),
137                           (1, 0), (1.7, 0), (2.4, 0), (2.9, 0), (3.4, 0), (3.9, 0)]
138
139             # initialise figure
140             fig, axes = plt.subplots(n_images)
141
142             # loop over axes and image_ids
143             for image_id, ax, pos in zip(images.keys(), axes, positions):
144                 print(image_id)
145                 cutout = make_cutout(images[image_id], x, y, width)
146
147                 vmin = 0
148                 vmax = 0.02
149
150                 print(vmin, vmax)
151
152                 # Adjust the position of each subplot
153                 ax.set_position([pos[0], pos[1], 0.7, 0.7])
154                 ax.axis('off')
155                 ax.imshow(cutout, vmin=vmin, vmax=vmax, origin='lower')
156                 ax.set_title(image_id.split('.')[1], fontsize=20)
157
158                 #plt.figure(figsize=(15,15))
159                 fig.suptitle(f'Object {id}', x=3.5, y=1.4, fontsize=50)
160
161                 plt.show()
162
163     ### My SED code ###

```

```

164      # Define the specific set of filters and galaxies
166      selected_filters = ['606', '814', '115', '125', '150', '160', '200', '
167          ↪ 277', '356', '444']
168
169      pivot_wavelength = {'606': 0.606, '814': 0.814, '115': 1.15, '125':
170          ↪ 1.25, '150': 1.50,
171              '160': 1.60, '200': 2.00, '277': 2.77, '356': 3.56, '
172                  ↪ 444': 4.44} # pivot wavelength in um
173
174      photom_group = hf['photom']
175
176      # Initialize empty lists to store filter numbers and normalized values
177      filter_numbers = []
178      normalized_values = []
179      normalized_error_values = []
180
181      # Filter all datasets based on SN condition and selected galaxy
182      for dataset_name in photom_group:
183          if dataset_name.startswith('F') and len(dataset_name) == 4:
184              # Extract the filter number (e.g., 'F200' -> '200')
185              filter_number = dataset_name[1:]
186
187              # Convert filter numbers to wavelengths (microns)
188              if filter_number in pivot_wavelength:
189                  wavelength = pivot_wavelength[filter_number]
190
191              if filter_number not in selected_filters:
192                  continue # Skip filters not in the selected set
193
194              # Get the values from the dataset
195              dataset_values = photom_group[dataset_name] [:]
196
197              # Filter the dataset values based on selected galaxy
198              filtered_dataset_values = dataset_values[id-1] # Indexing is 0-
199                  ↪ based
200
201              # Normalize the filtered dataset values
202              normalized_dataset_values = filtered_dataset_values / filters
203                  ↪ [277][id-1] # Indexing is 0-based
204
205              # Append the filter number and normalized values to the
206                  ↪ respective lists
207              filter_numbers.extend([wavelength])
208              normalized_values.extend([normalized_dataset_values])
209
210              if dataset_name.startswith('DF') and len(dataset_name) == 5 and
211                  ↪ dataset_name != 'DF105' and dataset_name != 'DF140':
212                  # Extract the filter number (e.g., 'F200' -> '200')
213                  filter_number = dataset_name[2:]
214
215              # Convert filter numbers to wavelengths (microns)
216              if filter_number in pivot_wavelength:
217                  wavelength = pivot_wavelength[filter_number]
218
219              if filter_number not in selected_filters:
220                  continue # Skip filters not in the selected set

```

```

216         # Extract the error dataset corresponding to each filter
217         error_dataset = photom_group[dataset_name] [:]

218         # Filter the dataset values based on selected galaxy
219         filtered_error_dataset_values = error_dataset[id-1] # Indexing is
220             ↪ 0-based

221         # Normalize the filtered dataset values
222         normalized_error_dataset_values = filtered_error_dataset_values /
223             ↪ filters[277][id-1] # Indexing is 0-based

224         # Append the filter number and normalized values to the
225             ↪ respective lists
226         normalized_error_values.extend([normalized_error_dataset_values])

227         # Desired numbers to be filtered
228         JWST_filters = [1.15, 1.5, 2.0, 2.77, 3.56, 4.44]

229         # Initialize lists to store filtered values
230         JWST_numbers = []
231         JWST_data = []
232         JWST_errors = []
233         HST_numbers = []
234         HST_data = []
235         HST_errors = []

236         # Iterate over filter_numbers and data simultaneously
237         for num, value, errors in zip(filter_numbers, normalized_values,
238             ↪ normalized_error_values):
239             if num in JWST_filters:
240                 JWST_numbers.append(num)
241                 JWST_data.append(value)
242                 JWST_errors.append(errors)
243             else:
244                 HST_numbers.append(num)
245                 HST_data.append(value)
246                 HST_errors.append(errors)

247         # JWST_yerr = [(np.log10(np.array(JWST_data) + np.array(JWST_errors)) -
248             ↪ np.log10(JWST_data)), (np.log10(JWST_data) - np.log10(np.array(
249                 ↪ JWST_data) - np.array(JWST_errors)))]
250         # HST_yerr = [(np.log10(np.array(HST_data) + np.array(HST_errors)) - np.
251             ↪ log10(HST_data)), (np.log10(HST_data) - np.log10(np.array(
252                 ↪ HST_data) - np.array(HST_errors)))]
```

# Create a scatter plot to display points for the normalized values for  
each galaxy

```

253         plt.figure(figsize=(6, 5), dpi=300)
254         # plt.errorbar(JWST_numbers, np.log10(JWST_data), yerr=JWST_yerr,
255             ↪ capsiz=5, fmt='o', label='JWST')
256         # plt.errorbar(HST_numbers, np.log10(HST_data), yerr=HST_yerr, capsiz
257             ↪ =5, fmt='o', label='HST')
258         plt.errorbar(JWST_numbers, (JWST_data), yerr=(JWST_errors), capsiz=5,
259             ↪ fmt='o', label='JWST')
260         plt.errorbar(HST_numbers, (HST_data), yerr=(HST_errors), capsiz=5, fmt=
261             ↪ 'o', label='HST')
262         #plt.axhline(0, linestyle='--', linewidth=1)
```

```

260     # Set labels for the axes and a title
261     plt.xlabel('Wavelength ($\mu m$)')
262     plt.ylabel('Flux (277 Normalized)')
263     plt.title(f'Normalized SED for Object {id}')
264     plt.ylim(-1,2)
265     #plt.yscale('log')#, nonposy='clip')
266     plt.grid(alpha=0.5, axis='x')
267     plt.xticks(filter_numbers, filter_numbers, rotation=75)
268     plt.legend()
269
270     plt.show()

```

## A.2 Unusual SEDs Code

```

from General_Code import create_filter_dictionaries
import numpy as np
import matplotlib.pyplot as plt

plt.rcParams['figure.dpi'] = 500
plt.style.use('default')

# Call the function from the imported module
filters, filter_errors, hf, ID = create_filter_dictionaries()

# Initialize empty lists to store filter numbers and normalized values
filter_numbers = []
average_values = []
average_error_values = []

photom_group = hf['photom']

pivot_wavelength = {'115': 1.15, '150': 1.50, '200': 2.00, '277': 2.77, '356':
    ↪ 3.56, '410': 4.10, '444': 4.44}

# Filter all datasets based on SN condition
for dataset_name in photom_group:
    if dataset_name.startswith('F') and len(dataset_name) == 4: # Extract the
        ↪ filter number (e.g., 'F200' → '200')
        filter_number = dataset_name[1:]
        # Get the values from the dataset
        dataset_values = photom_group[dataset_name][:]

    # Convert filter numbers to wavelengths (microns)
    if filter_number in pivot_wavelength:
        wavelength = pivot_wavelength[filter_number]

    if filter_number not in pivot_wavelength:
        continue # Skip filters not in the selected set

    # Normalize the filtered dataset values
    normalized_dataset_values = np.array(filters[int(filter_number)]) / np.
        ↪ array(filters[277])

#normalized_dataset_values = normalized_dataset_values[(
    ↪ normalized_dataset_values < 1e2) &

```

```

38                                     # (normalized_dataset_values >
40                                         ↪ -1e2)]
41             #print( filters[int(filter_number)].shape)
42
43             # Find the average value
44             normalized_average_values = np.nanmedian(normalized_dataset_values)
45             average_values.append(normalized_average_values)
46
47             #print(filter_number, normalized_average_values)
48
49             # List of desired filters
50             JWST_filters = [115, 150, 200, 277, 356, 410, 444]
51
52             # Create a list to store arrays of values
53             values_array = []
54
55             # Iterate through the keys of the dictionary and collect values from arrays with
56                 ↪ desired numbers
57             for key in sorted(filters.keys()): # Sort the keys to ensure order
58                 if key in JWST_filters:
59                     values = filters[key] / filters[277]
60                     for i, value in enumerate(values):
61                         if len(values_array) <= i:
62                             values_array.append([])
63                             values_array[i].append(value)
64
65             values_array = np.array(values_array)
66
67             # Create a list to store arrays of values
68             errors_array = []
69
70             # Iterate through the keys of the dictionary and collect values from arrays with
71                 ↪ desired numbers
72             for key in sorted(filters.keys()): # Sort the keys to ensure order
73                 if key in JWST_filters:
74                     errors = filter_errors[key] / filters[277]
75                     for i, value in enumerate(errors):
76                         if len(errors_array) <= i:
77                             errors_array.append([])
78                             errors_array[i].append(value)
79
80             errors_array = np.array(errors_array)
81
82             # Perform element-wise division
83             SN = values_array / errors_array
84
85             # Create a mask based on the condition
86             SN_mask = SN > 2
87
88             # Use the mask to filter values_array
89             filtered_values = values_array[SN_mask]
90
91             # Apply log to all values in values_array
92             log_flux_values = []
93             for values in filtered_values:
94                 log_flux_values.append(np.log10(values))
95
96             log_average_values = np.log10(average_values)

```

```

# Calculate the Euclidean distance for each array in values_array
euclidean_distances = []

for ID_val, values in zip(ID, log_flux_values):
    euclidean_distance = np.linalg.norm(log_average_values - np.array(values))
    if not np.isnan(euclidean_distance): # Filter out nan distances
        euclidean_distances.append((ID_val, euclidean_distance))

# Sort the list of tuples based on the Euclidean distances
euclidean_distances.sort(key=lambda x: x[1])

# Extract the first 100 ID values
top_ID = [item[0] for item in euclidean_distances[-10:]]
#top_100_ID.sort()
print(top_ID)

# Extracting the values
first_values = [pair[0] for pair in euclidean_distances]
second_values = [pair[1] for pair in euclidean_distances]

# Filtering the array by the second value being greater than 4
threshold = 4
filtered_distances = np.array(second_values)[np.array(second_values) > threshold]

# Plotting
plt.scatter(first_values, second_values, marker='o', s=5)
plt.axhline(y=threshold, color='r', linestyle='--', label=f'Threshold:\n{len(filtered_distances)} objects > Threshold')
plt.xlabel('ID')
plt.ylabel('Euclidean distance')
plt.title('Euclidean distance for each ID')
plt.legend()
plt.grid(alpha=0.5)
plt.show()

flux_values = []

# Iterate through the ID array and fetch the relevant flux value
for obj_id in first_values:
    flux_value = filters[277][obj_id - 1] # Assuming IDs start from 1
    flux_values.append(flux_value)

flux_values = np.array(flux_values)
second_values = np.array(second_values)

mask = (flux_values < 1e5) & (flux_values > -1e5)
flux_values = flux_values[mask]
second_values_filt = second_values[mask]

# Plotting
plt.scatter(flux_values, second_values_filt, marker='o', s=5)
plt.xlabel('Flux')
plt.ylabel('Euclidean distance')
plt.title('Euclidean distance for each ID')
plt.grid(alpha=0.5)
#plt.xscale('log')

```

```
150 plt.show()
```

```
from General_Code import create_filter_dictionaries
import matplotlib.pyplot as plt
import numpy as np

plt.rcParams['figure.dpi'] = 500
plt.style.use('default')

# Call the function from the imported module
filters, filter_errors, hf, ID = create_filter_dictionaries()

# Define the specific set of filters and galaxies
selected_filters = ['115', '150', '200', '277', '356', '410', '444']
selected_galaxies = [3892, 813, 2200, 2199, 4100, 1681, 1682, 1683, 1685, 1684]
#selected_galaxies = [7272]
#8646 most similar
#7272 most dissimilar (just the 277 band)

pivot_wavelength = {'606': 0.606, '814': 0.814, '105': 1.05, '115': 1.15, '125':
    ↪ 1.25, '140': 1.40, '150': 1.50,
    ↪ '160': 1.60, '200': 2.00, '277': 2.77, '356': 3.56, '410': 4.10,
    ↪ '444': 4.44} # pivot wavelength in um

photom_group = hf['photom']

average_values = [0.7425718368850297, 0.8683878101368937, 0.9423520053500087, 1.0,
    0.9362228270542196, 0.8439348130759242, 0.8089566333471095]

for galaxy_num in selected_galaxies:
    # Initialize empty lists to store filter numbers and normalized values
    filter_numbers = []
    normalized_values = []
    normalized_error_values = []

    # Filter all datasets based on SN condition and selected galaxy
    for dataset_name in photom_group:
        if dataset_name.startswith('F') and len(dataset_name) == 4 and dataset_name
            ↪ != 'F105' and dataset_name != 'F140':
            # Extract the filter number (e.g., 'F200' -> '200')
            filter_number = dataset_name[1:]

            # Convert filter numbers to wavelengths (microns)
            if filter_number in pivot_wavelength:
                wavelength = pivot_wavelength[filter_number]

            if filter_number not in selected_filters:
                continue # Skip filters not in the selected set

            # Get the values from the dataset
            dataset_values = photom_group[dataset_name][:]

            # Filter the dataset values based on selected galaxy
            filtered_dataset_values = dataset_values[galaxy_num - 1] # Indexing is
            ↪ 0-based

            # Normalize the filtered dataset values
            normalized_dataset_values = filtered_dataset_values / filters[277] [
```

```

        ↪ galaxy_num - 1] # Indexing is 0-based

54     # Append the filter number and normalized values to the respective lists
55     filter_numbers.extend([wavelength])
56     normalized_values.extend([normalized_dataset_values])

58     if dataset_name.startswith('DF') and len(dataset_name) == 5 and
59         ↪ dataset_name != 'DF105' and dataset_name != 'DF140':
60         # Extract the filter number (e.g., 'F200' -> '200')
61         filter_number = dataset_name[2:]

62         # Convert filter numbers to wavelengths (microns)
63         if filter_number in pivot_wavelength:
64             wavelength = pivot_wavelength[filter_number]

66         if filter_number not in selected_filters:
67             continue # Skip filters not in the selected set

68         # Extract the error dataset corresponding to each filter
69         error_dataset = photom_group[dataset_name][:]

72         # Filter the dataset values based on selected galaxy
73         filtered_error_dataset_values = error_dataset[galaxy_num - 1] # Indexing
74             ↪ is 0-based

76         # Normalize the filtered dataset values
77         normalized_error_dataset_values = filtered_error_dataset_values /
78             ↪ filters[277][galaxy_num - 1] # Indexing is 0-based

78     # Append the filter number and normalized values to the respective lists
79     normalized_error_values.append(normalized_error_dataset_values)

80 filter_numbers = np.sort(filter_numbers)

82     # Create a scatter plot to display points for the normalized values for each
83         ↪ galaxy
84     plt.figure(figsize=(5, 9))
85     plt.subplot(2, 1, 1) # Creating subplot for the first plot
86     #plt.scatter(filter_numbers, np.log(normalized_values), label='Values', s=15)
87     plt.errorbar(filter_numbers, normalized_values, yerr=normalized_error_values,
88             ↪ label='Values', capsizes=5, fmt='o')
89     plt.scatter(filter_numbers, average_values, label='Average Values', c='C1')

90     # Set labels for the axes and a title
91     plt.ylabel('Flux (Normalized)')
92     plt.title(f'Normalized SED for Galaxy {galaxy_num}')
93     plt.ylim(-0.1,2)
94     #plt.yscale('log')#, nonposy='mask')
95     plt.grid(alpha=0.5, axis='x')
96     plt.xticks(filter_numbers, filter_numbers, rotation=75)

98     plt.legend()

100 normalized_values = np.array(normalized_values)
101 average_values = np.array(average_values)

102 diff = normalized_values / average_values

```

```

106     diff_error = diff * np.sqrt((normalized_error_values/normalized_values)**2)
107     # Adding another subplot for the second plot
108     plt.subplot(2, 1, 2)
109     plt.errorbar(filter_numbers, diff, yerr=diff_error, capsize=5, fmt='o', c='C3')
110     plt.xlabel('Wavelength ($\mu m$)')
111     plt.ylabel('Flux / Average')
112     plt.ylim(-0.1,2.5)
113     #plt.yscale('log')
114     plt.grid(alpha=0.5, axis='x')
115     plt.xticks(filter_numbers, filter_numbers, rotation=75)
116
117     # Show the plot
118     #plt.tight_layout()
119     plt.show()
120
121
122     filter_num = [1.15, 1.5, 2.0, 2.77, 3.56, 4.10, 4.44]
123
124     #Average SED Plot
125     plt.figure()
126     plt.scatter(filter_num, average_values, c='C1')
127     plt.ylabel('Flux (277 Normalized)')
128     plt.title('Normalized Average SED')
129     plt.xlabel('Wavelength ($\mu m$)')
130     #plt.yscale('log')#, nonposy='mask')
131     plt.ylim(0,2)
132     plt.grid(alpha=0.5, axis='x')
133     plt.xticks(filter_numbers, filter_numbers, rotation=75)
134     plt.show()

```

## Appendix B

### Figures

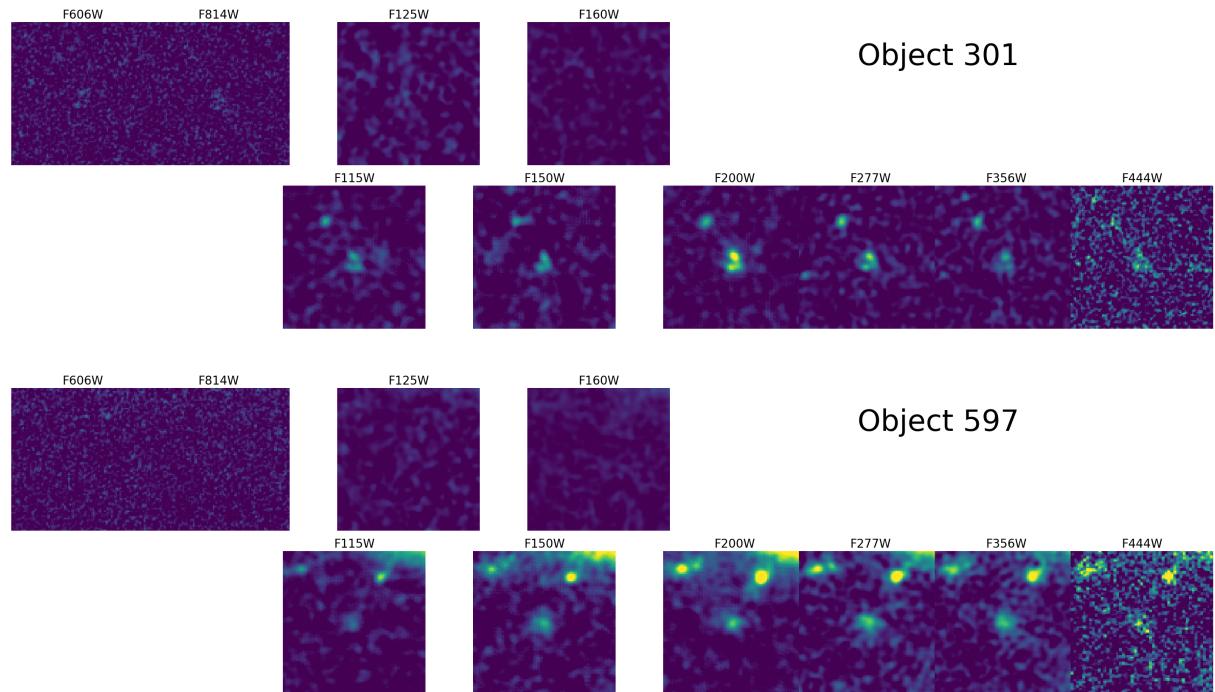


Figure B.1: Thumbnails of objects 301 and 597, selected in Chapter 2.

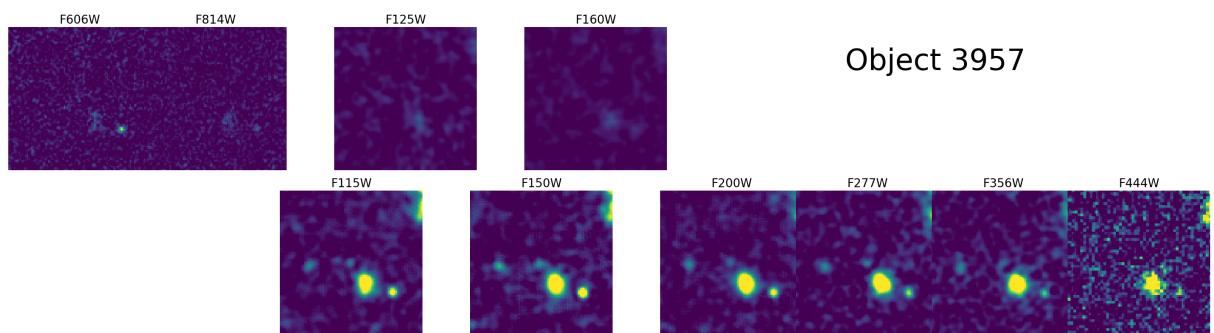
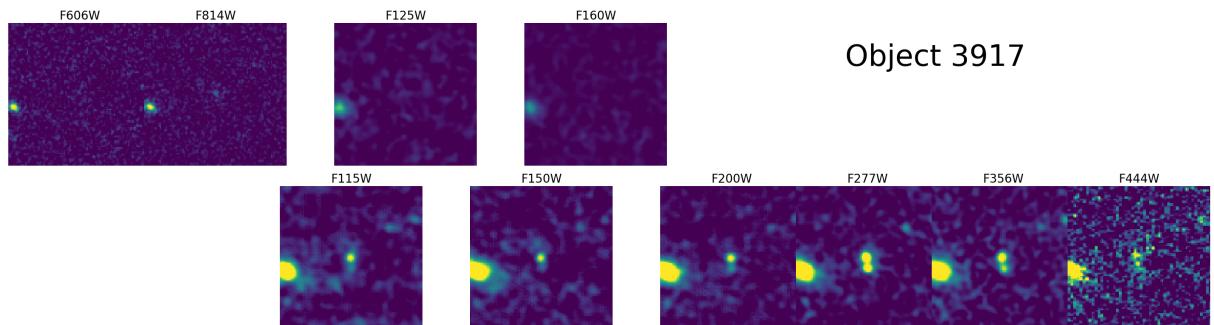


Figure B.2: Thumbnails of objects 3917 and 3957, selected in Chapter 2.

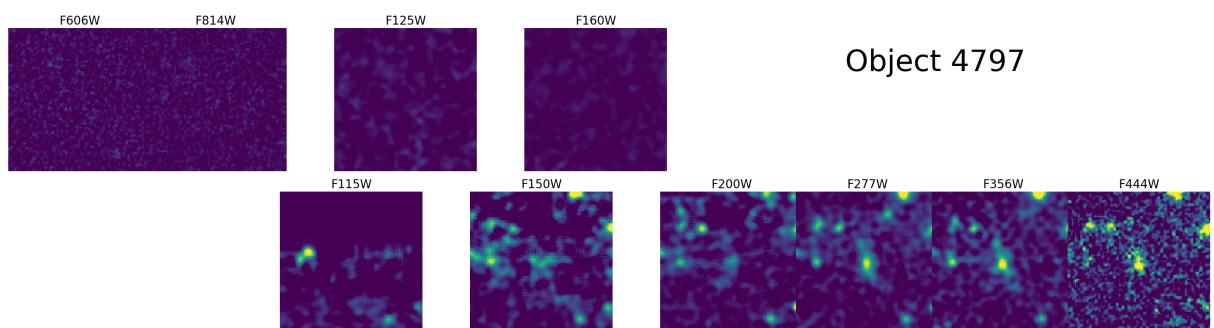
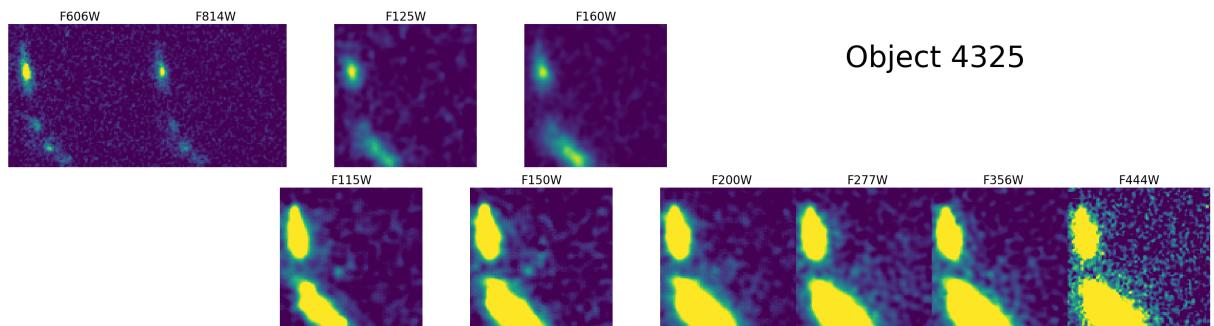


Figure B.3: Thumbnails of objects 4325 and 4797, selected in Chapter 2.

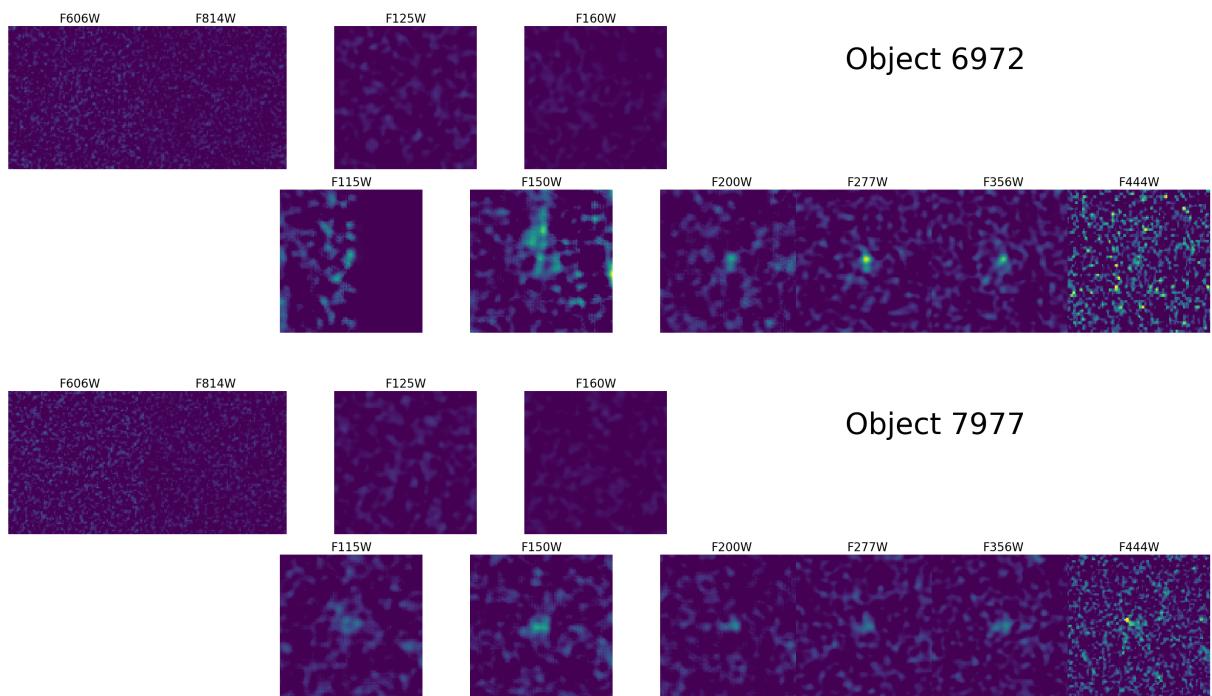


Figure B.4: Thumbnails of objects 6972 and 7977, selected in Chapter 2.

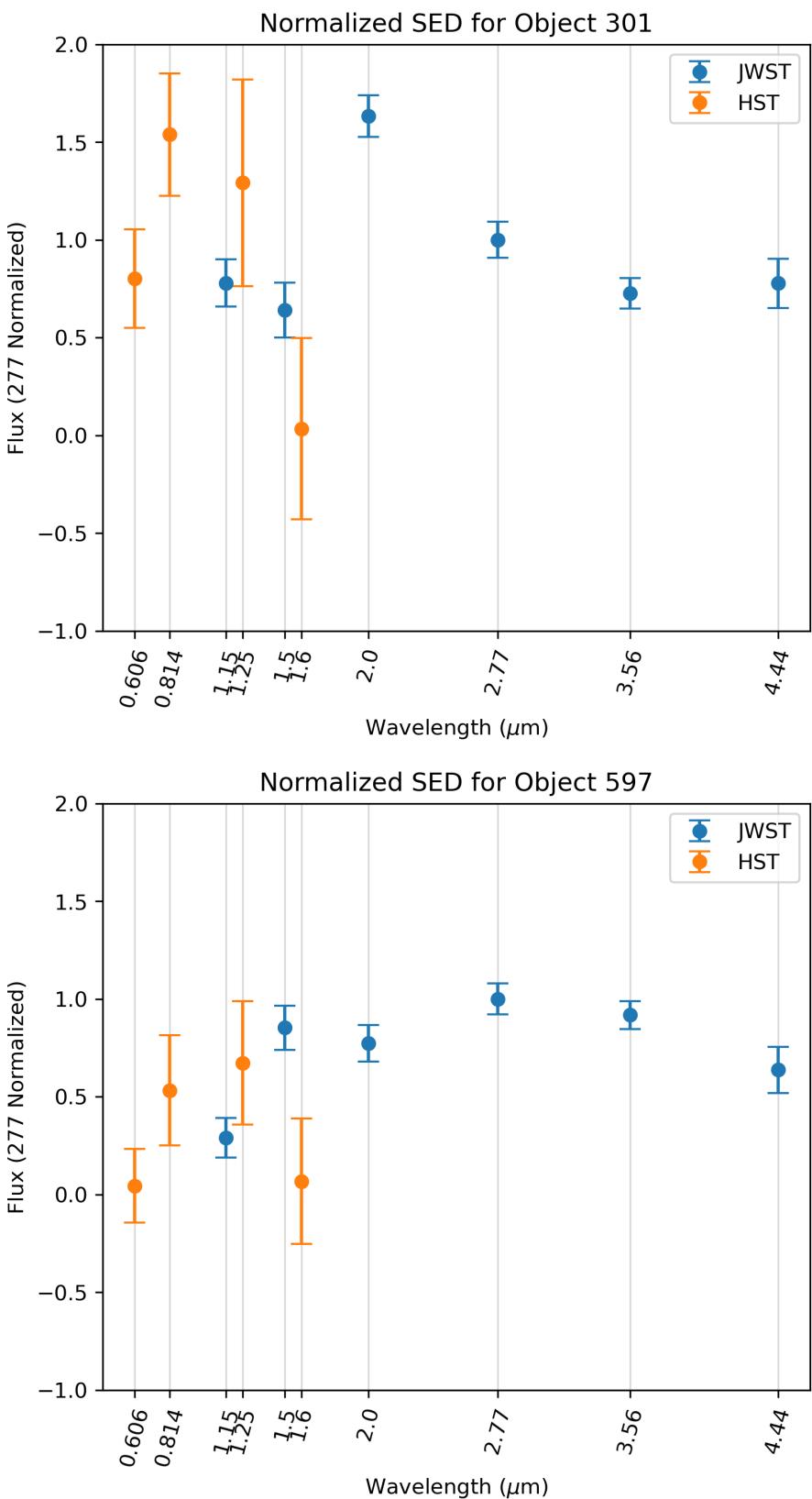


Figure B.5: SEDs of objects 301 and 597, selected in Chapter 2.

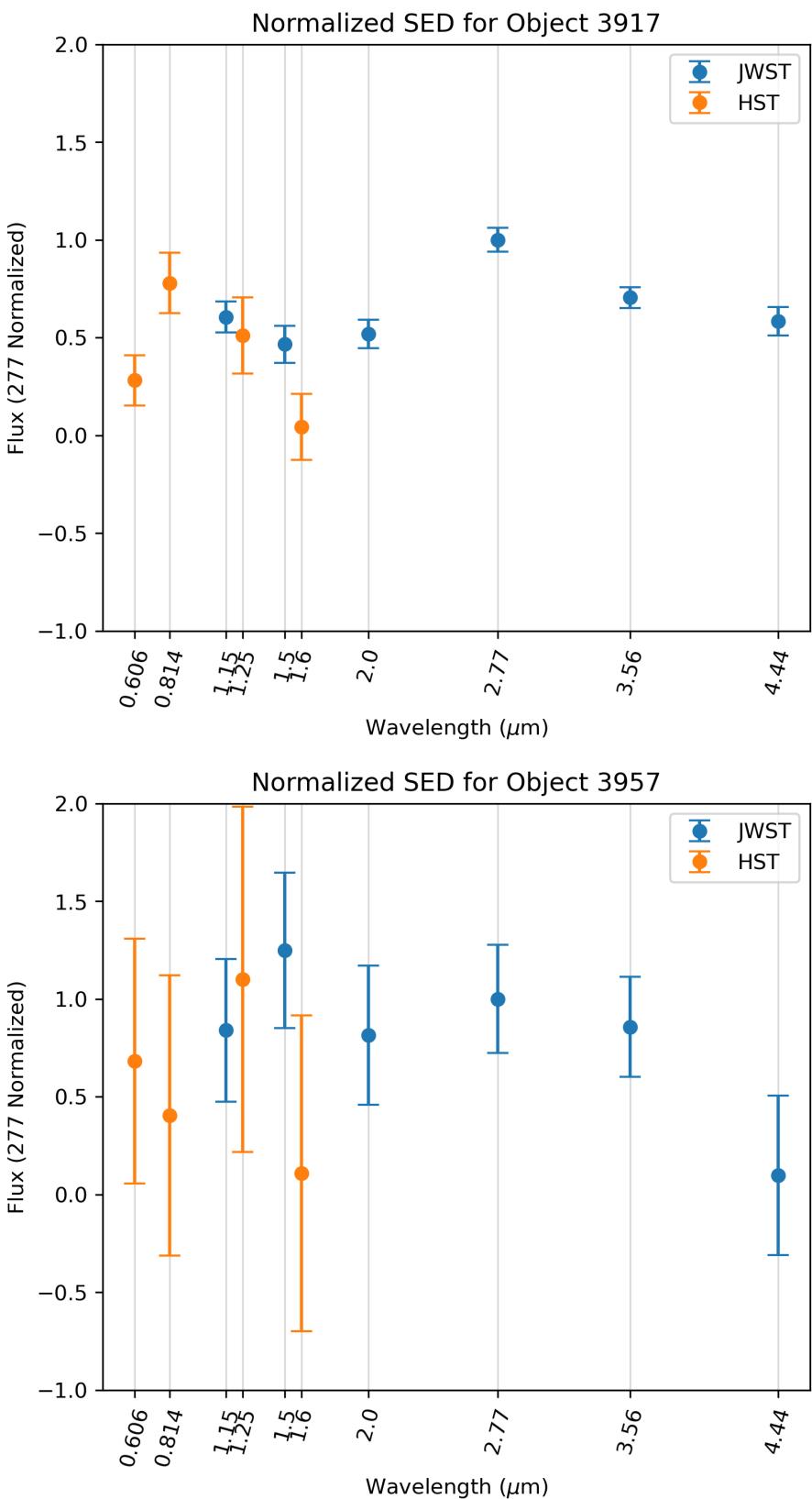


Figure B.6: SEDs of objects 3917 and 3957, selected in Chapter 2.

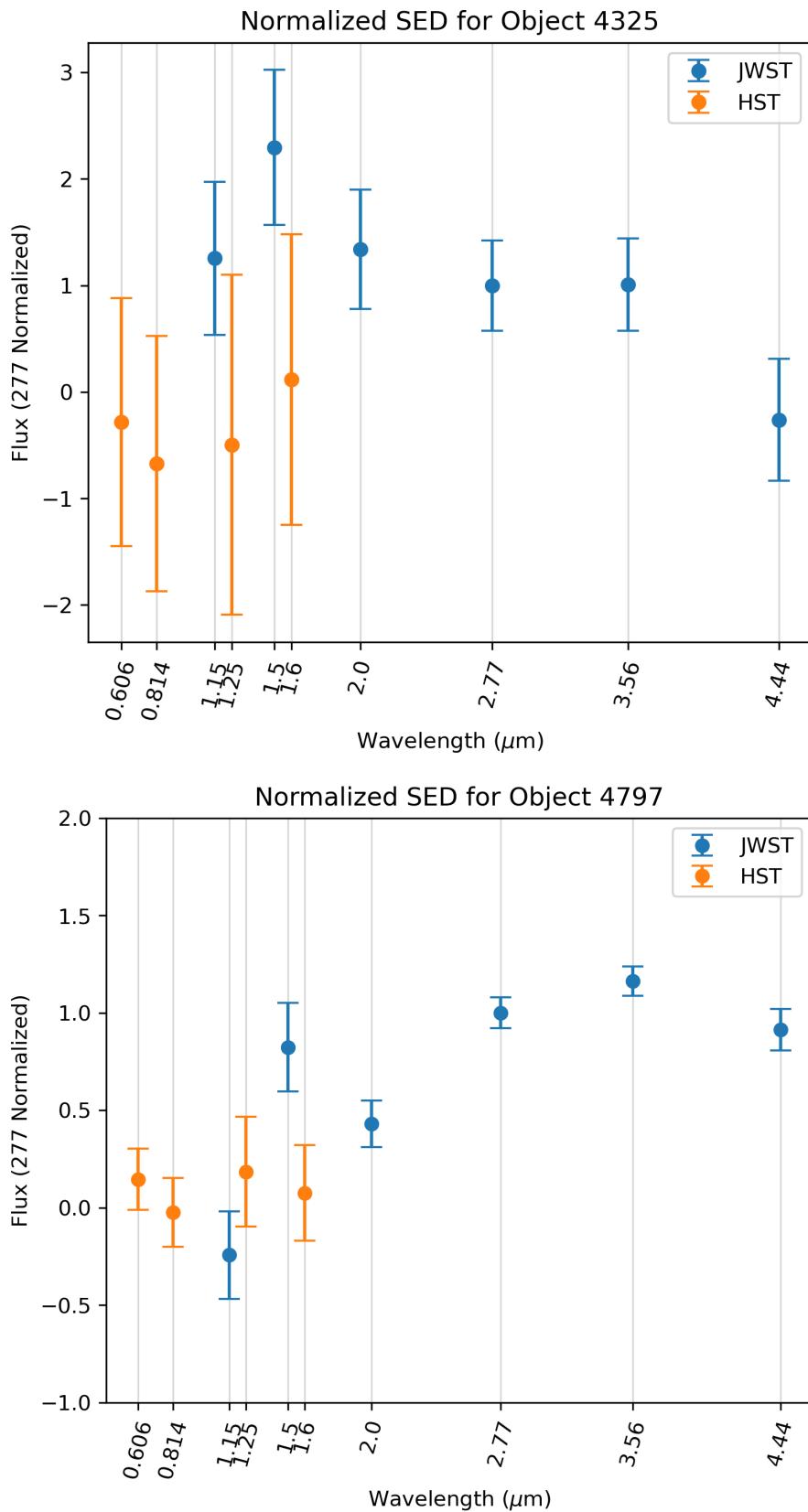


Figure B.7: SEDs of objects 4325 and 4797, selected in Chapter 2.

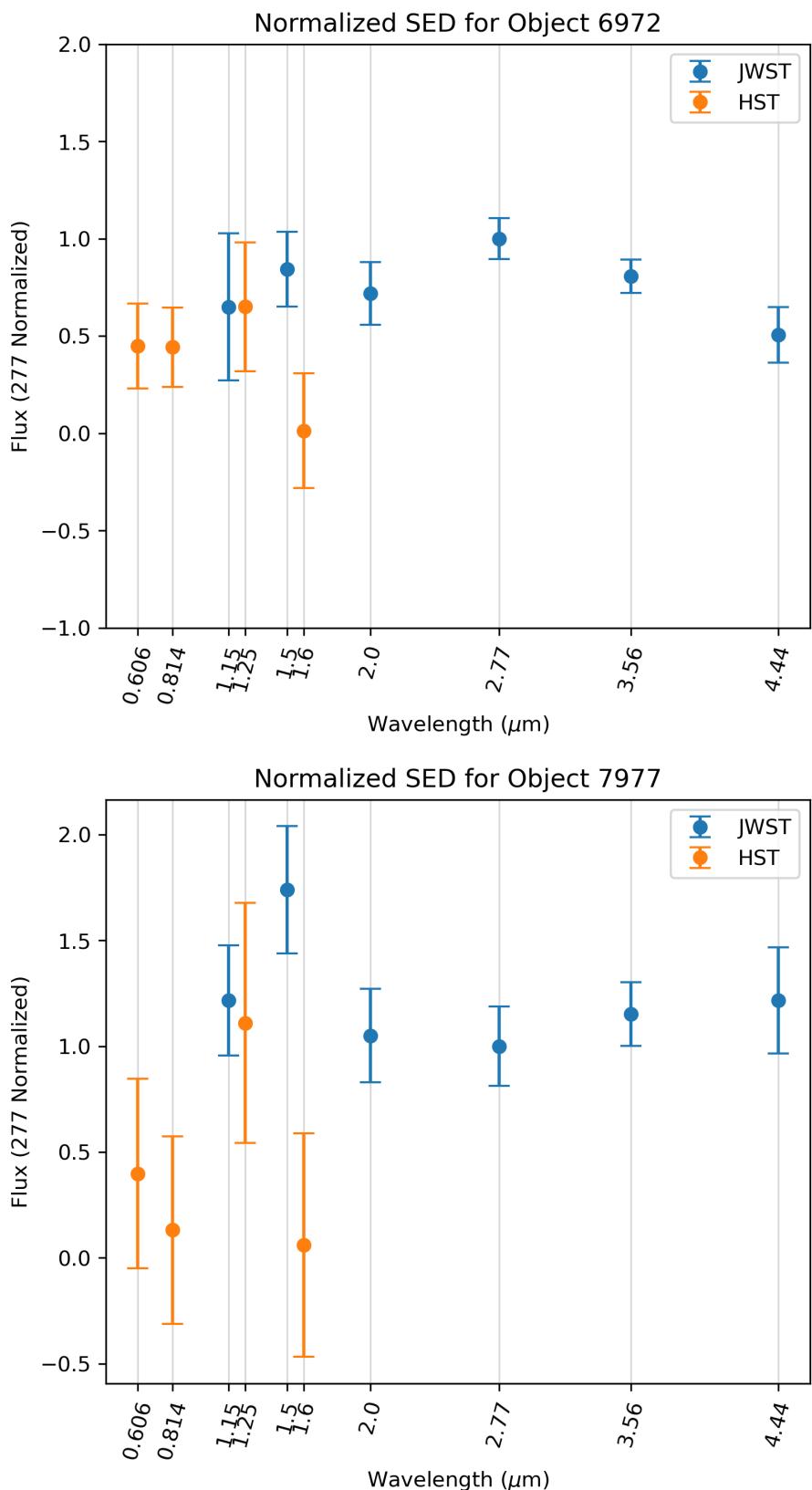
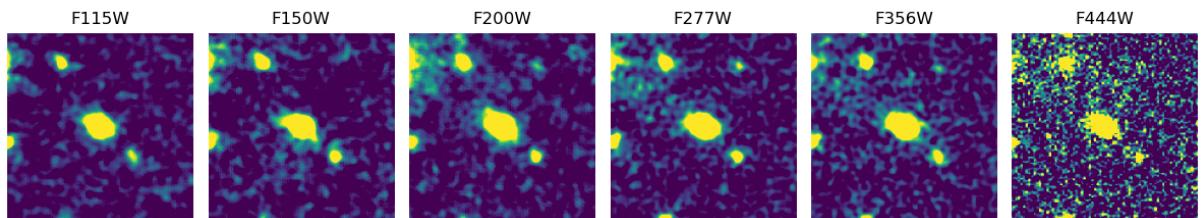
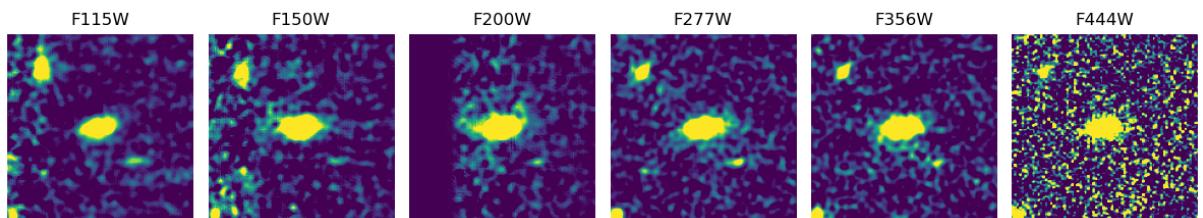


Figure B.8: SEDs of objects 6972 and 7977, selected in Chapter 2.

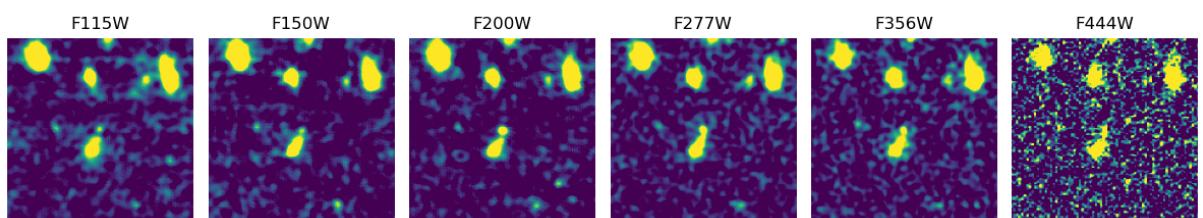
## Object 813



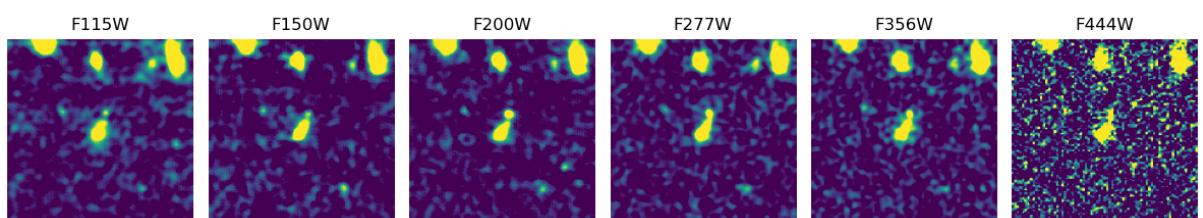
## Object 1681



## Object 1682



## Object 1683



## Object 1684

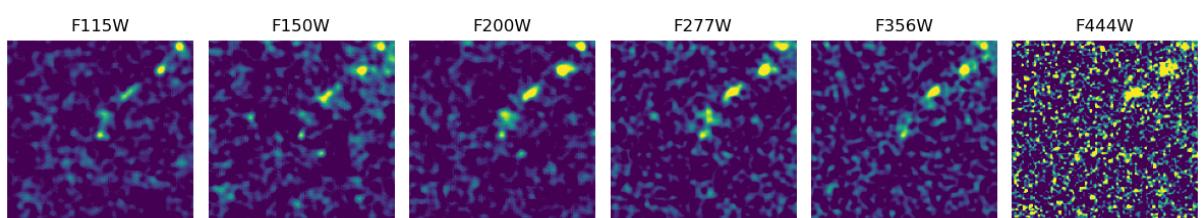
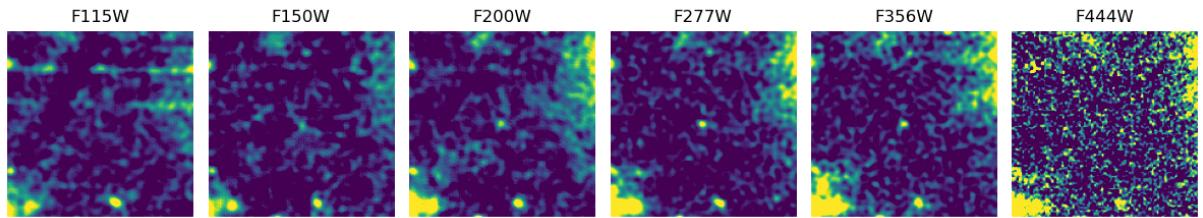
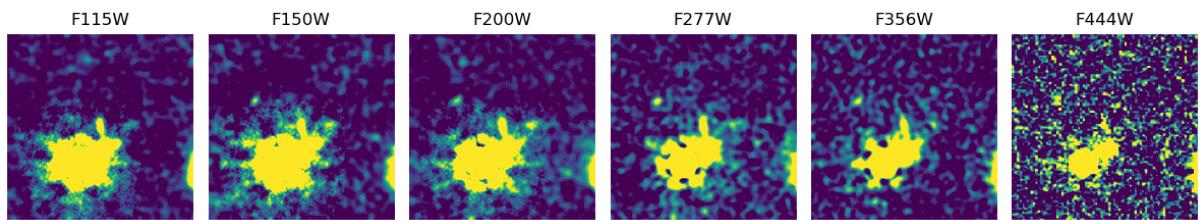


Figure B.9: Thumbnails of objects 813, 1681, 1682, 1683, and 1684, selected in Chapter 3.

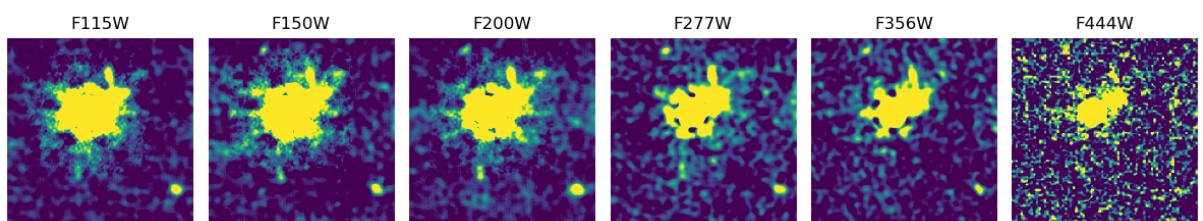
## Object 1685



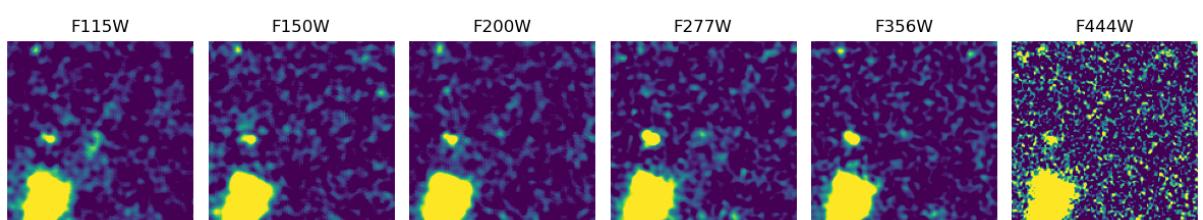
## Object 2199



## Object 2200



## Object 3892



## Object 4100

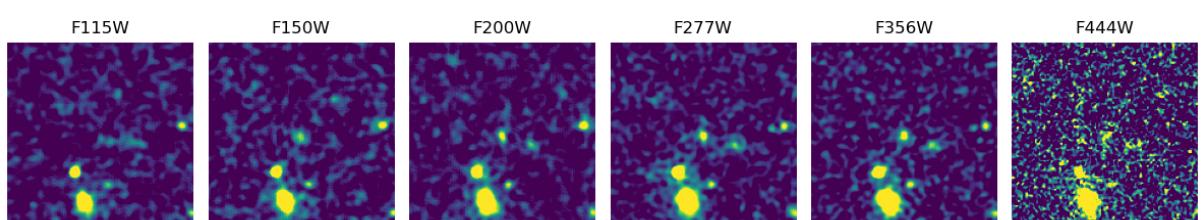


Figure B.10: Thumbnails of objects 1685, 2199, 2200, 3892, and 4100, selected in Chapter 3.

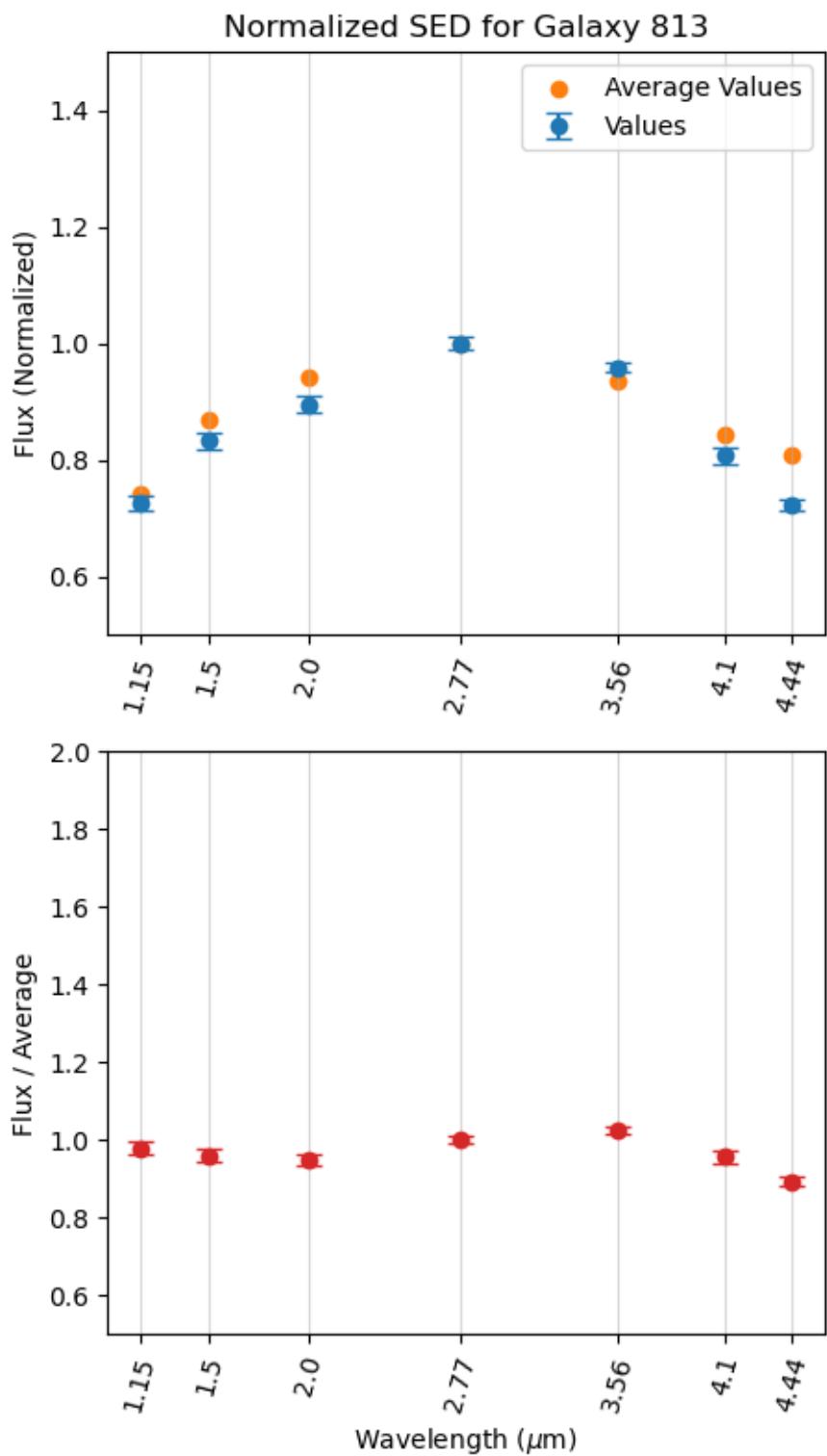


Figure B.11: SED of object 813, selected in Chapter 3.

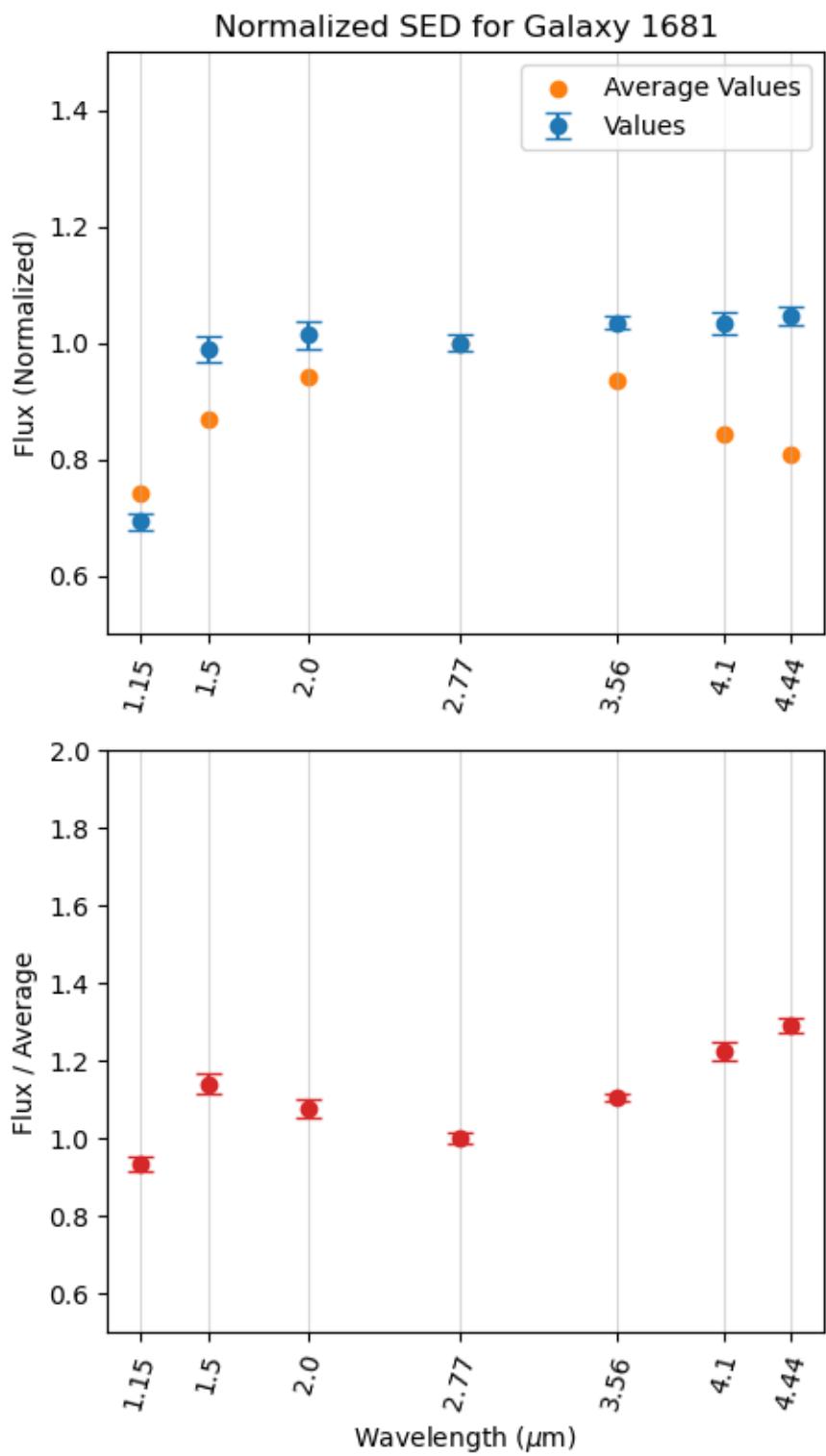


Figure B.12: SED of object 1681, selected in Chapter 3.

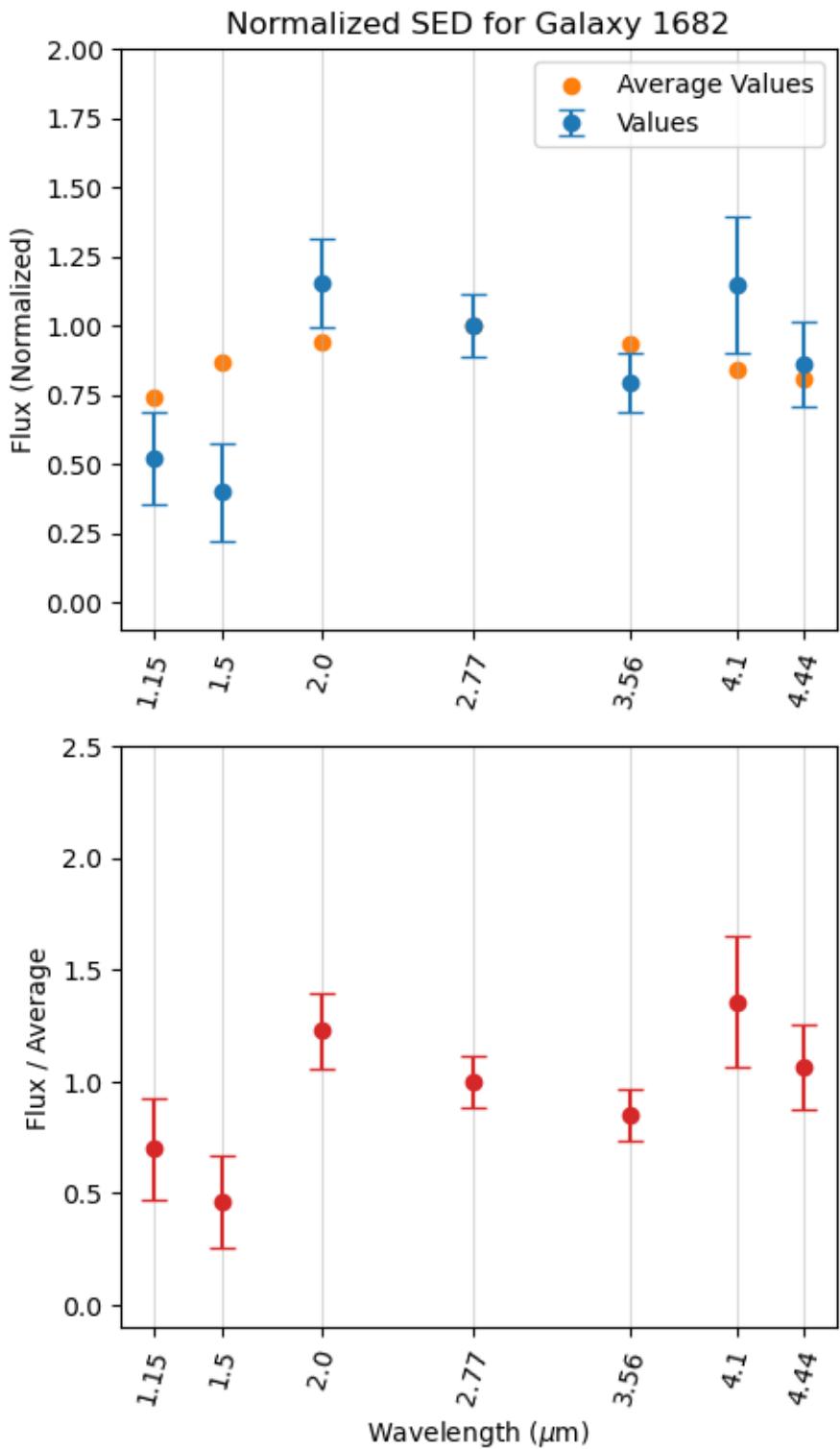


Figure B.13: SED of object 1682, selected in Chapter 3.

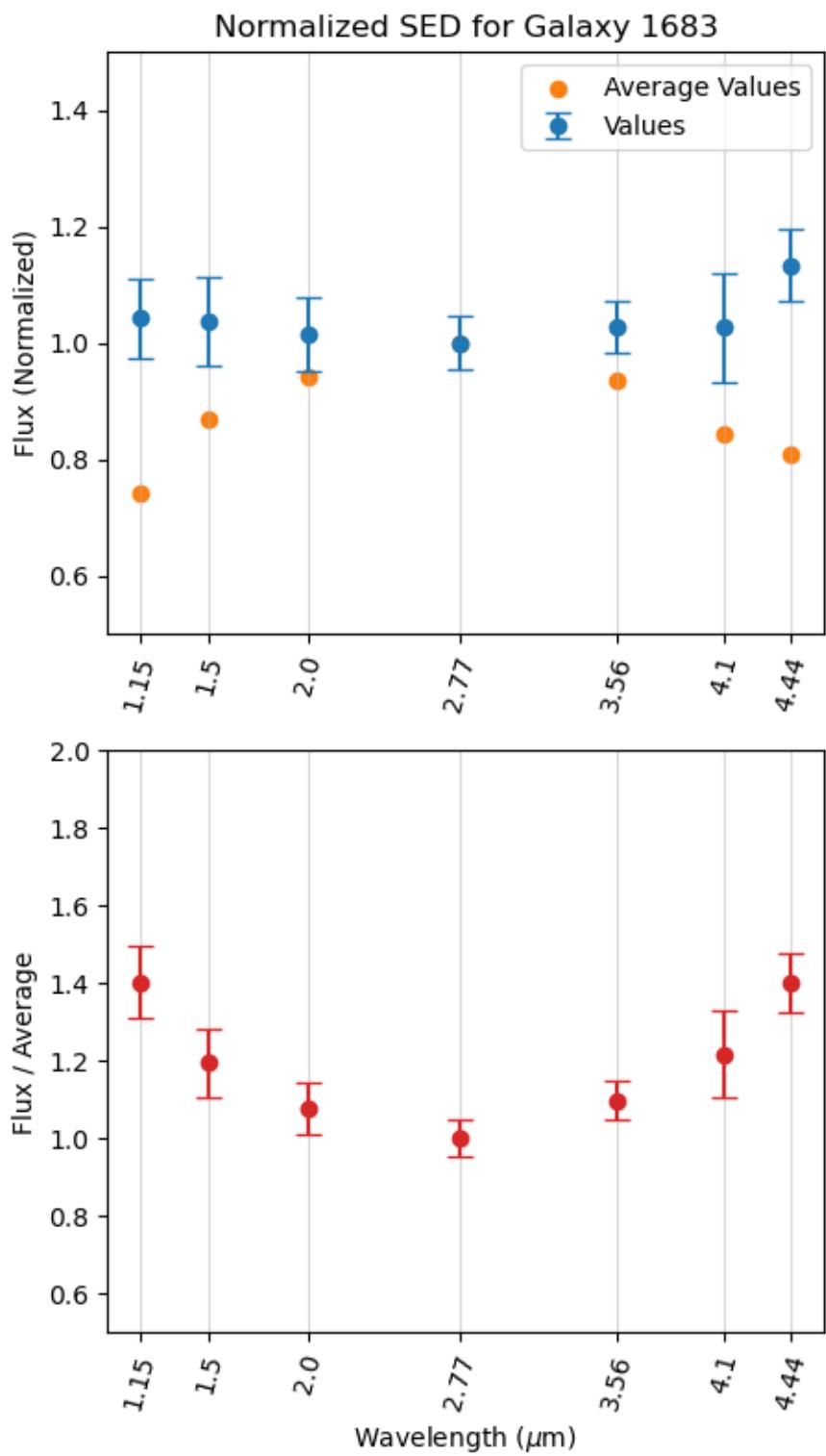


Figure B.14: SED of object 1683, selected in Chapter 3.

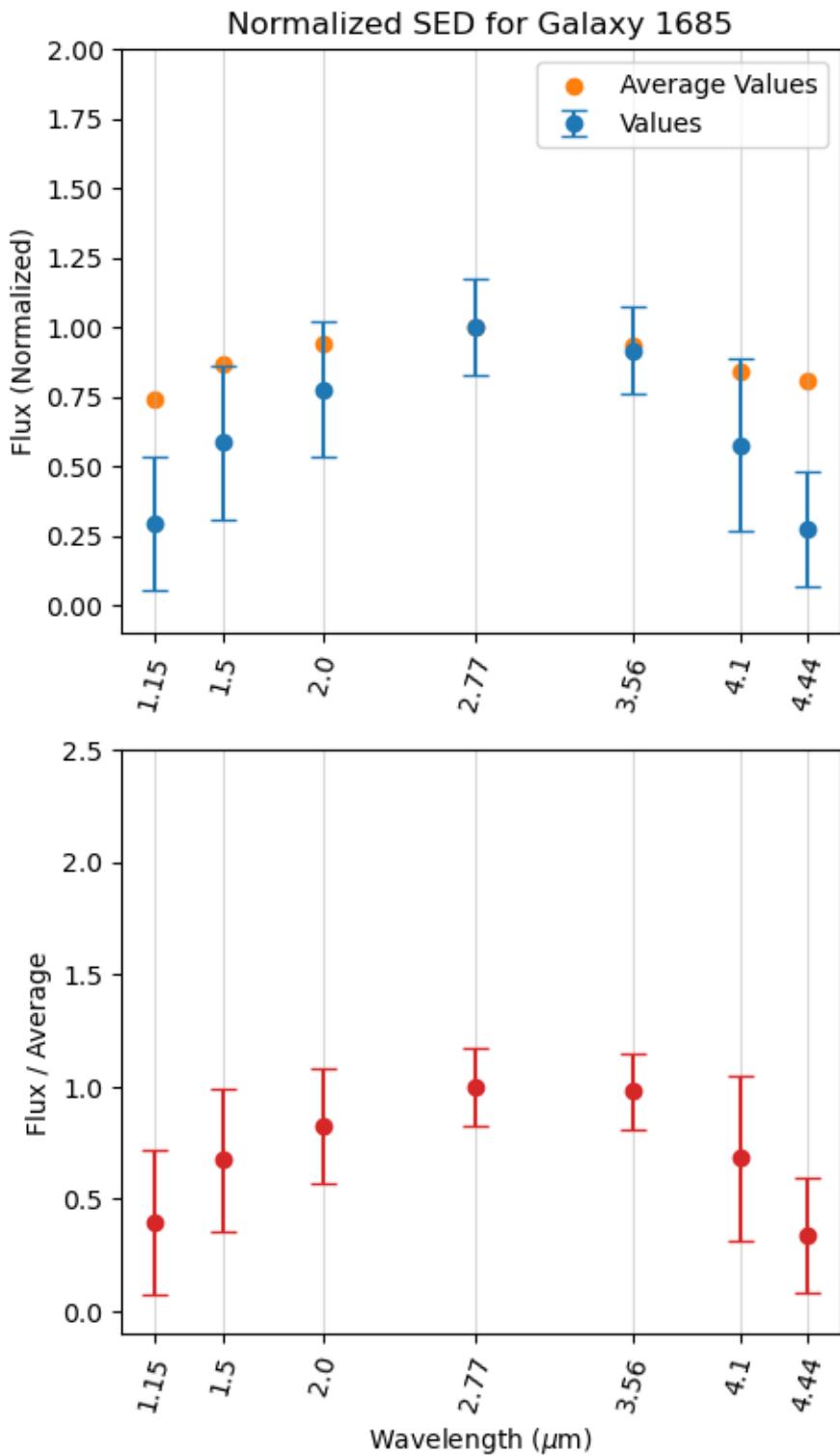


Figure B.15: SED of object 1685, selected in Chapter 3.

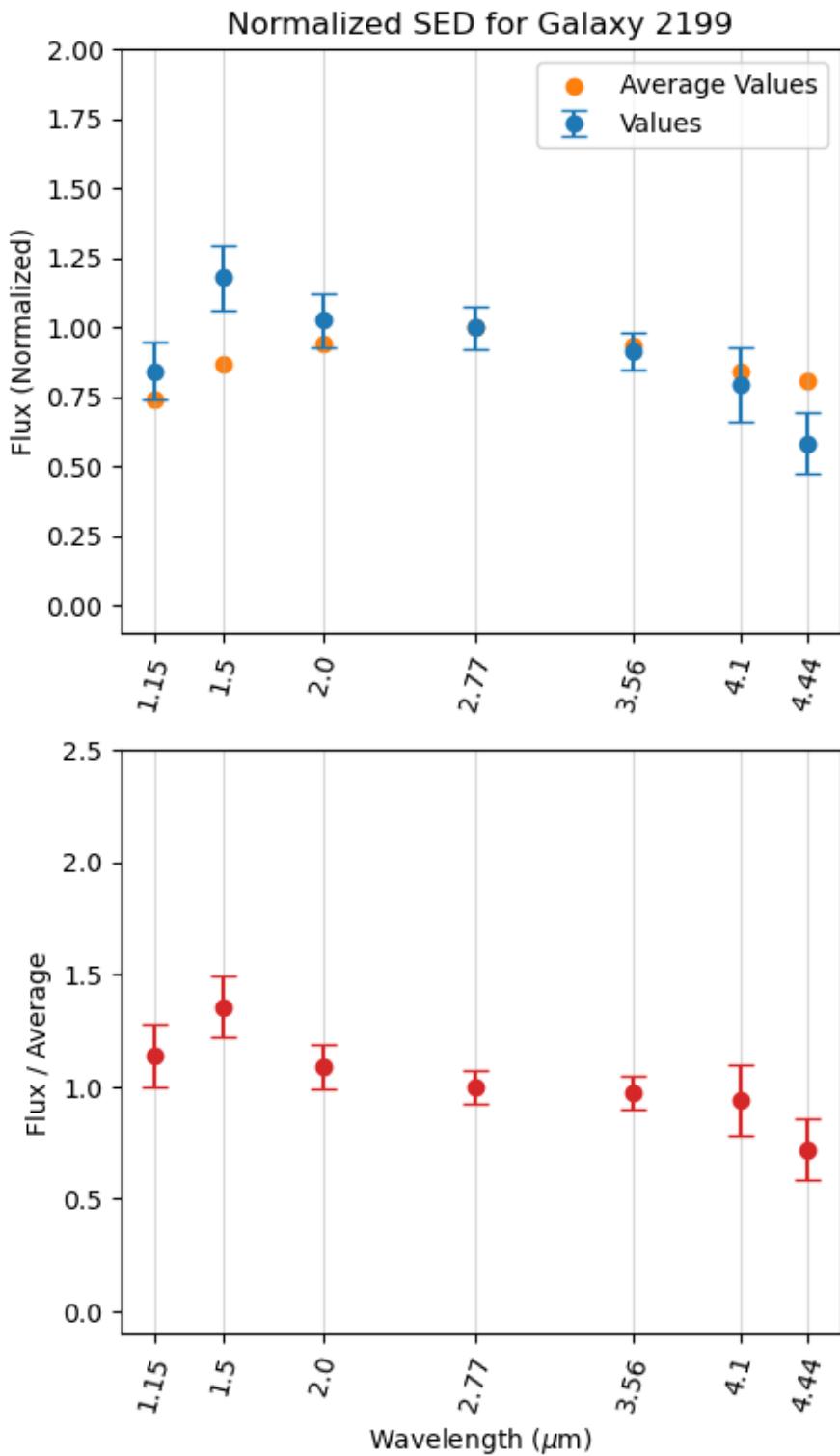


Figure B.16: SED of object 2199, selected in Chapter 3.

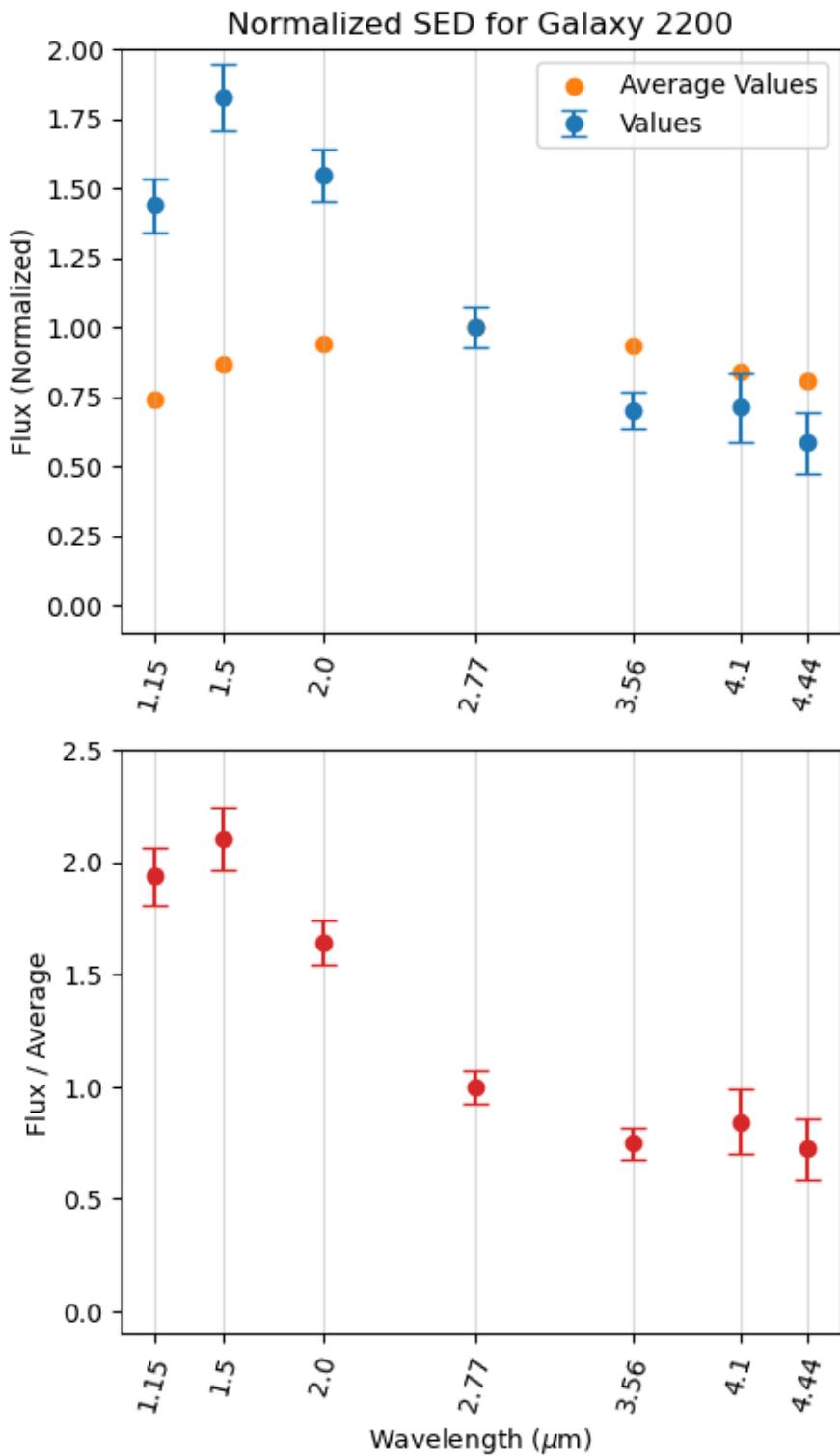


Figure B.17: SED of object 2200, selected in Chapter 3.

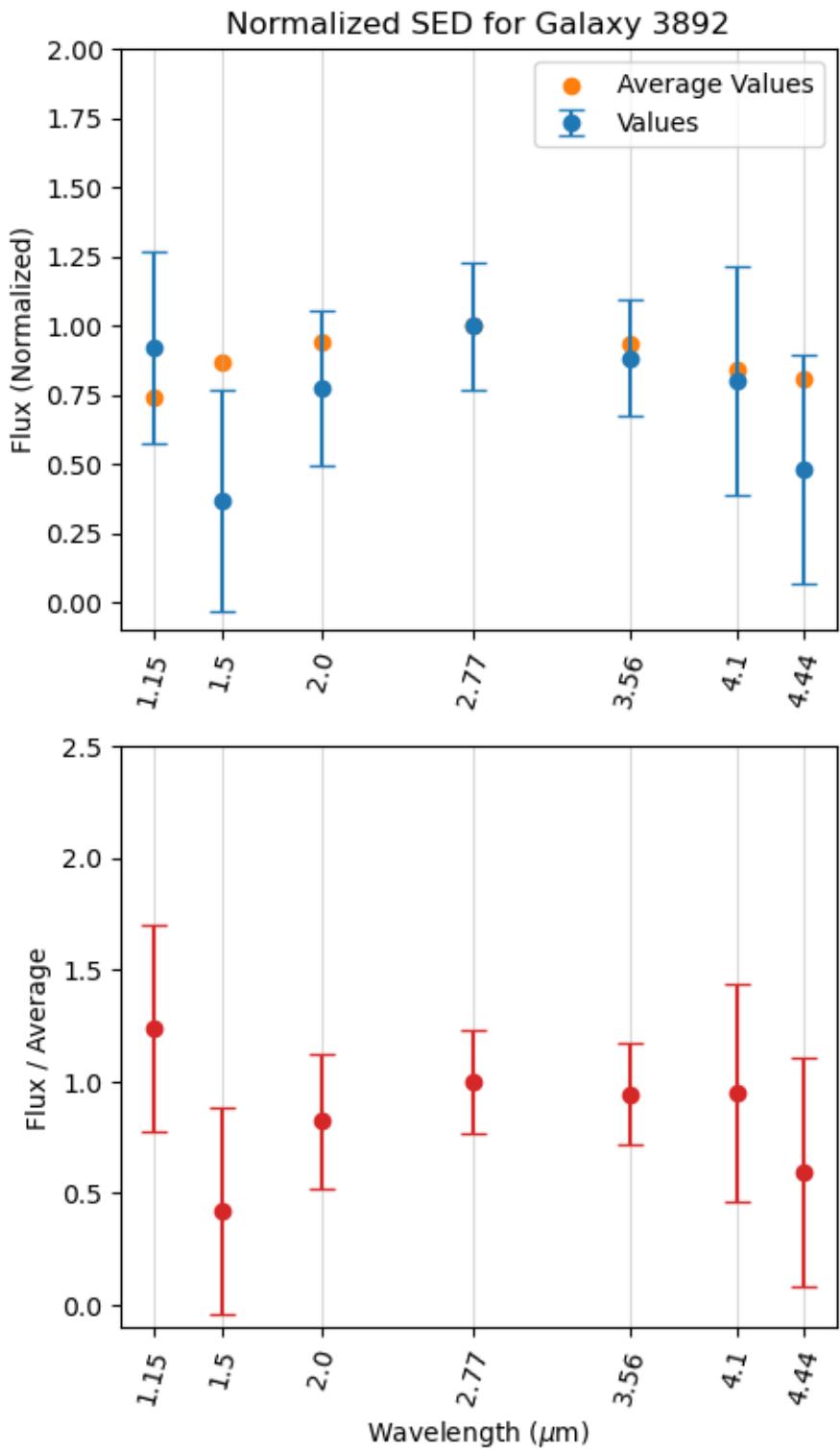


Figure B.18: SED of object 3892, selected in Chapter 3.

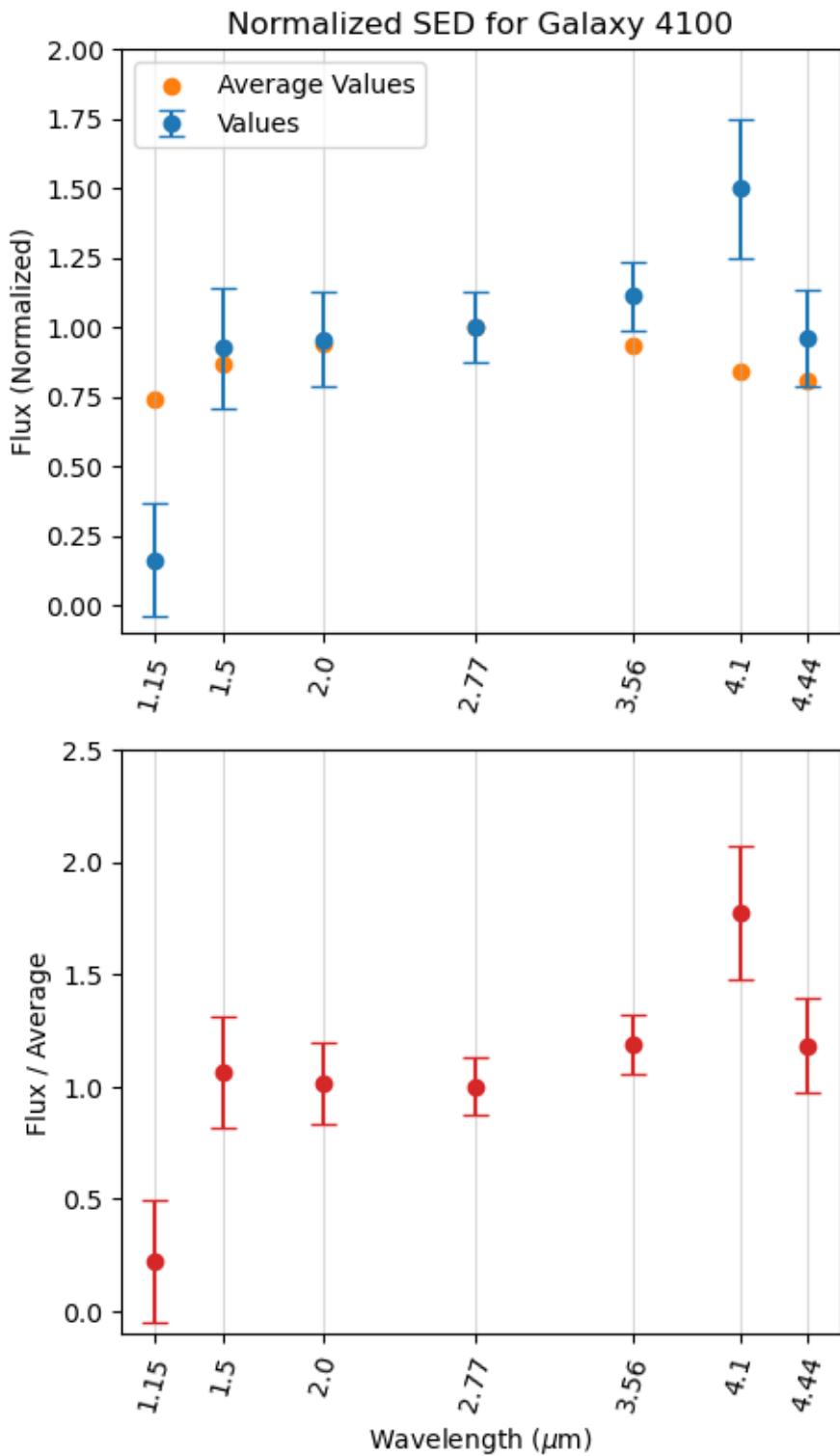


Figure B.19: SED of object 4100, selected in Chapter 3.