A multiwavelength study of massive galaxy disks across cosmic time

João Manuel Sousa Águas

Mestrado em Física
Especialização em Astrofísica e Cosmologia

Dissertação orientada por:
Doutor Fernando Buitrago
Doutor Israel Matute

2018
Acknowledgments

I would like to thank all the people that gave me the needed support and that helped me reach this stage of my life. My girlfriend and my family that have been there for me at many levels. Doutor Fernando Buitrago and Doutor Israel Matute for being my Dissertation Coordinators, from whom I learned so much and for being reliable friends. I also thank Doctor Rosalind Skelton for the helping me, by kindly providing the files of the filters used in the 3D-HST survey photometry. Doutor Alexandre Cabral, Doutor Manuel Abreu, Doutor David Castro Alves and my coworkers from LOLS (Laboratório de Óptica Lasers e Sistemas) for their support. I would like to thank Professor Rui Agostinho and Professor João Lin Yun for their help and advice. I would also like to thank Randi and Christopher Schmidt for helping me and for giving me strength. Finally, I would like to thank my teachers from Universidade de Lisboa and Universidade do Algarve for their not only for their teachings but also for their advice that helped me shape my way of thinking and the path that led me here.
Abstract

For quite some time the understanding of the origin and evolution of the Universe has been a very hot topic not only for the academic community but also for the general public. From the Cosmology, we got information about how space expanded and how the large-scale structures of the Universe came to be, how the galaxies gathered into clusters, superclusters, and filaments. Since Hubble’s discovery the Universe of the galaxies, we have gathered knowledge about how the galaxies formed, evolved and interact. Galaxies increase their masses through collision. The simulations show that the outcome of a collision depends on the mass of each of the galaxies. If the mass of both galaxies is similar, the result is an elliptical galaxy. This leaves two questions: How do massive disk galaxies appear? How is it possible that we have them in the Local Universe (z=0) ?

With this work, we aim to shed light on how the population of galaxies evolved from z=3 up to the Local Universe, paying special attention to their structural parameters like mass and size. To accomplish that we rely on data from galaxy surveys like 3D-HST/CANDELS and Huertas-Company work. On a first approach, these catalogs may give an idea of this evolution. One of our objectives is to gain experience with the software package LePHARE, as we produce our estimates of redshift, mass, along with other parameters.

Keywords: massive, galaxies, disks, evolution
Resumo

As origens e evolução do Universo sempre foi um tema de muito interesse tanto para a comunidade académica como para o público geral. Através da Cosmologia sabemos como é que o espaço se expandiu e como apareceram as estruturas de larga escala. Como as galáxias se juntaram em enxames, super-enxames e filamentos. Desde que Hubble descobriu o Universo das galáxias, que o nosso conhecimento acerca de como estas evoluem e interagem tem aumentado. Sabemos que as galáxias aumentam a sua massa através de colisões. As simulações mostram que o resultado final da colisão depende das massas de ambas galáxias. Se as massas forem similares então o resultado final será uma galáxia elíptica. Então isto deixa-nos duas questões: Como é que apareceram as galáxias disco massivas? Como é possível ainda haver galáxias disco massivas no Universo Local (z=0)?

Este trabalho tenta responder a como é que a população de galáxias evoluiu desde z=3 até ao Universo Local, no que toca aos parâmetros estruturais das mesmas. Por forma a realizar este trabalho teremos como base a informação publicada nos catálogos dos levantamentos 3D-HST, CANDELS e no trabalho de Huertas-Company. Primeiramente tentaremos utilizar a informação contida nestes catálogos para observar algumas das relações como a relação massa-tamanho. Mas temos como objetivo principal obter experiência com o programa LePHARE, por forma a podermos produzir as nossas próprias estimativas.

Palavras-chave: galáxias, discos, massivas, evolução
List of Figures

1.1 The evolution of the number density of massive galaxies over the different epochs. 1
2.1 M31, The Andromeda Galaxy .............................................. 2
2.2 The Hubble’s tuning fork ..................................................... 3
2.3 Examples of the E0, Sa, SBa a S0 galaxies ............................... 4
2.4 The Sérsic function ........................................................... 5
2.5 Galaxy mass distribution ..................................................... 6
2.6 The relation between redshift and time ................................... 8
2.7 Photometric redshift determination ......................................... 9
2.8 The comparison of a spectra with and without the effect of dust .... 10
2.9 The regions of space that light passes to leave the galaxy ............. 11
3.1 The throughput of the filters on the WFC3 and ACS cameras .......... 12
4.1 LePHARE workflow .......................................................... 15
4.2 TOPCAT main window ....................................................... 17
4.3 TOPCAT table display window. ............................................. 17
4.4 TOPCAT plot display window. ............................................. 18
4.5 Aladin display. ............................................................... 18
4.6 Comparison between LePHARE and Skelton, SHORT catalog and no minimum error. ....................................................... 21
4.7 Residual of the comparison between LePHARE and Skelton, SHORT catalog and no minimum error. ........................................ 25
4.8 Short catalog, no minimum error,ID 23837 ............................... 25
4.9 Short catalog, no minimum error,ID 41013 ............................... 26
4.10 Short catalog, no minimum error, redshift distribution ............... 26
4.11 Short catalog, no minimum error, template distribution ............. 27
4.12 Short catalog, no minimum error, reddening values distribution ... 27
4.13 Short catalog, no minimum error,χ^2 distribution .................... 28
4.14 Comparison between LePHARE and Skelton’s redshift estimates, SHORT catalog with a minimum error of 0.1. .............................. 29
4.15 Residue of the comparison between LePHARE and Skelton, SHORT catalog and no minimum error. ........................................ 29
4.16 Short catalog, with a minimum error of 0.1 ,ID 23837 ................ 30
4.17 Short catalog, no minimum error,ID 23837 ............................... 30
4.18 Short catalog, with a minimum error of 0.1 ,ID 41013 ................ 31
4.19 Short catalog, with no minimum error ,ID 41013 ...................... 31
4.20 Short catalog, with minimum error of 0.1,ID 3048 ................. 32
4.21 Short catalog, no minimum error,ID 3048 ............................... 32
4.22 Short catalog, with a minimum error of 0.1,Redshift distribution .... 33
4.23 Short catalog, with a minimum error of 0.1,Template distribution ... 33
4.24 Short catalog, with a minimum error of 0.1,Reddening distribution .... 34
4.25 Short catalog, with a minimum error of 0.1,χ^2 distribution ........ 34
4.26 Long catalog, Comparison of Photometric versus spectroscopic redshift .............................................................. 35
4.27 Long catalog, Residuals of the comparison of Photometric versus spectroscopic redshift ....................................................... 35
4.28 Long catalog, Template distribution with AUTO_ADAPT ........... 36
4.29 Long catalog, E(B-V) distribution with AUTO_ADAPT ............. 36
4.30 Long catalog, Redshift distribution with AUTO_ADAPT .............. 37
4.31 Long catalog, Comparison of photometric redshift with the manual input of the offsets ............................................................... 38
4.32 Long catalog, residue of the comparison of photometric redshift with the manual input of the offsets ................................. 38
4.33 Long catalog, Template distribution with manual offsets .................................. 38
4.34 Long catalog, E(B-V) distribution with manual offsets .................................. 39
4.35 Comparison between LePHARE and Skelton, LONG catalog fixed upper limits. 40
4.36 Residue of the comparison between LePHARE and Skelton, SHORT catalog and
no minimum error. .................................................................................................. 41
4.37 Long catalog, fixed upper limits, ID 23837 ......................................................... 41
4.38 Long catalog, with the 5σ upper limits, The SED 9870 ....................................... 41
4.39 Comparison between LePHARE and Skelton, LONG catalog with 5σ upper limits. 43
4.40 Residue of the comparison between LePHARE and Skelton, SHORT catalog and
no minimum error. .................................................................................................. 43
4.41 Long catalog, with fixed upper limits, The SED 9870 ......................................... 43
4.42 Long catalog, with the 5σ upper limits, The SED 9870 ....................................... 43
4.43 Long catalog, with fixed upper limits, The SED 13796 ....................................... 43
4.44 Long catalog, with the 5σ upper limits, The SED 13796 ....................................... 43
4.45 Long catalog, with the 5σ upper limits, Redshift distribution ............................ 43
4.46 Short catalog, with the 5σ upper limits, Template distribution .......................... 43
4.47 Short catalog, with the 5σ upper limits, Reddening distribution ........................ 43
4.48 Short catalog, with fixed upper limits, Template distribution ............................ 43
4.49 Comparison between LePHARE and Skelton, LONG catalog with the max(E(B-
V))=1.5. ................................................................................................................. 43
4.50 Long catalog, with the max(E(B-V)) = 1.5, The SED 41013 ............................ 43
4.51 Long catalog, with the max(E(B-V)) = 0.4, The SED 13572 ............................ 43
4.52 Long catalog, with the max(E(B-V)) = 1.5, The SED 13572 ............................ 43
4.53 Long catalog, with the max(E(B-V)) = 0.4, The SED 27365 ............................ 43
4.54 Long catalog, with the max(E(B-V)) = 1.5, The SED 27365 ............................ 43
4.55 Long catalog, with the max(E(B-V)) = 0.4, The SED 30084 ............................ 43
4.56 Long catalog, with the max(E(B-V)) = 1.5, The SED 30084 ............................ 43
4.57 Long catalog, with the max(E(B-V)) = 0.4, The SED 31001 ............................ 43
4.58 Long catalog, with the max(E(B-V)) = 1.5, The SED 31001 ............................ 43
4.59 The distribution of redshift estimates for 0 < E(B-V) < 0.4 and 0 < E(B-V) <
1.5 the two separate intervals and Skelton’s estimation .......................................... 43
4.60 The distribution of templates for 0 < E(B-V) < 0.4 and 0 < E(B-V) < 1.5the
two separate intervals and Skelton’s estimation ..................................................... 43
4.61 The distribution of E(B-V) for 0 < E(B-V) < 0.4 and 0 < E(B-V) < 1.5the
two separate intervals and Skelton’s estimation .................................................... 43
4.62 Comparison between LePHARE and Skelton, LONG catalog with the S/N < 5
upper limits ............................................................................................................. 43
4.63 Residuals of the comparison between LePHARE and Skelton, LONG catalog with
the S/N < 5 upper limits ......................................................................................... 43
4.64 Long catalog, with the 5σ upper limits, The SED of galaxy 41013 .................... 43
4.65 Comparison between LePHARE and Skelton, LONG catalog without IRAC 3 and 4 43
4.66 Residual of the comparison between LePHARE and Skelton, LONG catalog with-
out IRAC 3 and 4, with error bars ............................................................................. 43
5.1 The mass-size relation found in the AEGIS field ................................................. 43
5.2 The mass-size relation found in the COSMOS field ............................................ 43
5.3 The mass-size relation found in the UDS field ..................................................... 43
5.4 The mass-size relation found in the GOODS-N field .......................................... 43
5.5 The mass-size relation found in the GOODS-S field ........................................... 43
5.6 The mass-size relation taking all fields into account, in two redshift bins .......... 43
5.7 The mass-size relation taking all fields into account, separated by two redshift
bins, and two Sérsic index intervals ..................................................................... 43
List of Tables

2.1 The wavelengths of the Hydrogen emission lines at different redshifts ........ 9
3.1 Information of the CANDELS fields.(Skelton et al. 2014) .................... 12
3.2 The number of detected objects in the CANDELS fields. ....................... 13
3.3 List of the photometric data included in the individual CANDELS fields of the
   3D-HST survey(From: Skelton et al. 2014) ...................................... 14
4.1 List of tests performed to the catalog data. ...................................... 22
4.2 The BC03 template characteristics .............................................. 23
# List of Abbreviations and Acronyms

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CANDELS</td>
<td>Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey</td>
</tr>
<tr>
<td>CMB</td>
<td>Cosmic Microwave Background</td>
</tr>
<tr>
<td>CSP</td>
<td>Composite Stellar Population</td>
</tr>
<tr>
<td>FIR</td>
<td>Far Infrared</td>
</tr>
<tr>
<td>FUV</td>
<td>Far Ultraviolet</td>
</tr>
<tr>
<td>Gyr</td>
<td>Giga-year</td>
</tr>
<tr>
<td>HST</td>
<td>Hubble Space Telescope</td>
</tr>
<tr>
<td>IMF</td>
<td>Initial Mass Function</td>
</tr>
<tr>
<td>IR</td>
<td>Infrared</td>
</tr>
<tr>
<td>Mpc</td>
<td>Mega-parsec</td>
</tr>
<tr>
<td>PDF</td>
<td>Probability Distribution Function</td>
</tr>
<tr>
<td>PSF</td>
<td>Point Spread Function</td>
</tr>
<tr>
<td>QSO</td>
<td>Quasi-Stellar Object</td>
</tr>
<tr>
<td>SDSS</td>
<td>Sloan Digital Sky Survey</td>
</tr>
<tr>
<td>SED</td>
<td>Spectral Energy Distribution</td>
</tr>
<tr>
<td>SF</td>
<td>Star Formation</td>
</tr>
<tr>
<td>SFH</td>
<td>Star Formation History</td>
</tr>
<tr>
<td>SFR</td>
<td>Star Formation Rate</td>
</tr>
<tr>
<td>SNIa</td>
<td>Supernovae Type Ia</td>
</tr>
<tr>
<td>SSF</td>
<td>Specific Star Formation</td>
</tr>
<tr>
<td>SSP</td>
<td>Simple Stellar Populations</td>
</tr>
<tr>
<td>S/N</td>
<td>Signal to noise ratio</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>ΛCDM</td>
<td>Λ Cold Dark Matter</td>
</tr>
</tbody>
</table>
Contents

Acknowledgments i
Abstract ii
Resumo iii
List of Figures iv
List of Tables vii
List of Abbreviations and Acronyms viii

1 Objectives 1

2 Introduction
2.1 The Galaxies .................................................. 3
  2.1.1 Galaxy morphological classification .......................... 3
  2.1.2 The components of a galaxy .................................. 5
  2.1.3 The Sérsic Function .......................................... 5
  2.1.4 Massive galaxies ............................................. 6
2.2 The relation between distance and time .......................... 7
2.3 The effect of the expansion on the light’s wavelength .......... 8
2.4 Photometric and Spectroscopic redshifts ......................... 9
2.5 The effects of dust .......................................... 10

3 The Data 11

4 Methodology
4.1 Software packages ............................................. 15
  4.1.1 LePHARE (PHotometric Analysis for Redshift Estimate) .... 15
  4.1.2 TOPCAT (Tool for OPerations on Catalogues And Tables) .. 17
  4.1.3 Aladin (Sky Atlas) .......................................... 18
  4.1.4 Source Extractor (SExtractor) .............................. 19
4.2 Analysis ................................................. 19
  4.2.1 The data selection ........................................... 19
  4.2.2 Understanding the data ...................................... 20
  4.2.3 Test A: SHORT catalog, no minimum error .................. 24
  4.2.4 Test B: SHORT catalog with a minimum error of 0.1 .......... 29
  4.2.5 Test C: LONG catalog using AUTO_ADAPT and APPLY_SYSSHIFT 35
  4.2.6 Test D: LONG catalog with upper limits .................... 40
  4.2.7 Test E: LONG catalog with 5σ upper limits ................ 43
  4.2.8 Test F: LONG catalog with 5σ upper limits and E(B-V) maximum increased to 1.5 ... 48
  4.2.9 Test G: LONG catalog, with max(E(B-V))=1.5, considering a S/N < 5 as an upper limit .... 53
  4.2.10 Test H: LONG catalog, S/N < 5 upper limits and no IRAC bands 3 and 4 57

5 Results
5.1 Results based on catalog data .................................. 58
  5.1.1 The mass-size relation ..................................... 58
  5.1.2 The Sérsic index distribution ............................... 62
5.2 Our estimates ............................................. 63
5.2.1 Comparison of redshift .............................................. 63
5.2.2 The mass-size relation using our own determinations .............. 64

6 Conclusions ................................................................. 69

7 Future work .................................................................. 70

References ........................................................................ 71

Appendix ........................................................................... 73

A Plots with error bars ....................................................... 73
  A.1 Test A: SHORT catalog, no minimum error ...................... 73
  A.2 Test B: SHORT catalog with a minimum error of 0.1 .......... 73
  A.3 Test D: LONG catalog with upper limits ....................... 74
  A.4 Test E: LONG catalog with 5σ upper limits ................... 74
  A.5 Test F: LONG catalog with 5σ upper limits and E(B-V) maximum increased to 1.5 75
  A.6 Test G: LONG catalog, with max(E(B-V))=1.5, considering a S/N < 5 as an upper limit .................................................. 75
  A.7 Test H: LONG catalog, S/N < 5 upper limits and no IRAC bands 3 and 4 .... 76

B LePHARE ........................................................................ 77
  B.1 zphot.para ................................................................. 77

C Matlab scripts ................................................................ 81
  C.1 fluxtomag.m .............................................................. 81
  C.2 splitfilters.m ............................................................. 83
  C.3 SHORT catalog, no minimum error .............................. 84
1 Objectives

This work answers two specific needs in the Galaxy Thematic Line at the Instituto de Astrofísica e Ciências do Espaço.

1. Obtaining expertise in the photometric determination of redshifts and stellar populations using the software package LePHARE\(^1\) (Arnouts et al. 1999 and Arnouts and Ilbert 2006).

2. Attempting to shed some light on why there are so few Andromeda-like galaxies in the Local Universe.

An Andromeda-like galaxy has a disk component and a stellar mass greater than \(10^{11} M_\odot\). The number density of this type of galaxies has changed throughout the cosmic time as is shown in the work of Buitrago et al. (2013).

The most massive galaxies in the local Universe have elliptical and not disk morphologies like the Andromeda galaxy. In the present study, we attempt to understand this topic by investigating the few massive galaxies with disk morphologies at different cosmic epochs. To achieve this goal we use the software package LePHARE to produce our estimates derived from the photometry provided by the 3D-HST survey.

![Figure 1.1: The change in the number density of massive galaxies over the different epochs. The red line shows the fraction of Early-types, the blue line shows the Late-types, and the green line shows the Peculiar (Irregulars and Mergers) galaxies. The differently shaded regions mark the redshift range covered by the different surveys. (Buitrago et al. 2013)](image-url)

2 Introduction

The understanding of the formation and evolution of galaxies is a subject that has sparked the interest of not only the academic community but also of the general public. The galaxies are the objects where most of the matter (either baryonic or dark matter) is clumped, and therefore their study is imperative in order to ascertain the Universe we live in. The first observations of the characteristics of the population of elliptical galaxies showed a homogeneity regarding their old stellar populations. This was consistent with a fast formation of stars due to the collapse of massive gas clouds at an earlier epoch. This inspired Eggen, Lynden-Bell, and Sandage

\(^1\)LePHARE (PHotometric Analysis for Redshift Estimate), http://www.cfht.hawaii.edu/~arnouts/lephare.html
(1962) to propose the first model of the formation and evolution of galaxies. In this model, the galaxies formed as a result of the collapse of massive clouds of gas which would have no angular momentum. Therefore it would not explain the existence of disk galaxies. More recent theories propose a hierarchical Universe where small clumps of matter (either baryonic or dark matter) grow through mergers. During a merger, the gravitational interaction between the two bodies increases the angular momentum of the system which may result in a collapse into a disk-like object.

Simulations of major mergers involving galaxies with similar masses show that the disk components are destroyed, and that the result is a massive Early-type galaxy (Boylan-Kolchin, Ma, and Quataert 2006; Naab et al. 2007). In the minor merger scenario, we have galaxies of different masses, such as in the interaction between an already massive galaxy and its satellite galaxies. In this case, the disk may be disturbed without being destroyed (Hopkins et al. 2008). Both of these scenarios occur in our Universe, which lead to the questions of how did the Andromeda-like galaxies evolve up till our cosmic epoch? (Figure 2.1 shows the Andromeda galaxy or M31, with two satellite galaxies.)

Figure 2.1: The Andromeda galaxy also known as M31 is the closest massive galaxy. (Credit: http://hubblesite.org)

For this work, the CANDELS survey provided the structural data (effective radius and Sérsic index), and from the 3D-HST we got the photometric data, redshift and mass estimates. The LePHARE package was run on the photometric data to produce our estimates of redshift, mass, age and star formation rate, that were then compared with the results provided by 3D-HST catalog to determine and improve the quality of our estimates. Finally, we used the CANDELS structural data with our estimates of mass and redshift to observe the mass-size relation, and how it changed with redshift.

In order to help understand this work, the following subsections contain a brief introduction of the fundamental concepts. We start by defining a galaxy in section 2.1. To help identify the type of galaxies that are the object of study in this work and what their components are, section 2.1.1 introduces the morphological classification followed by section 2.1.2 that contains a brief description of the basic components of the galaxies. To have an estimate of the evolution of the morphology of the galaxies across cosmic time, we may use a function that fits the light intensity distribution and provides an estimation of the more prominent galaxy components. For this, we use the Sérsic function which is presented in section 2.1.3. In section 2.1.4 we present the Press-Schechter function, why we consider galaxies with a $M_*>10^{11}M_\odot$ as massive galaxies and also why we need data from surveys that cover large areas of the sky. In this work,
we are addressing how galaxies change with cosmic time. Therefore we need observations from
different cosmic epochs. In section 2.2 we explain the relation between cosmic time and redshift.
It is also crucial to understand how light is affected by the expansion of the Universe and the
consequences of this are described in section 2.3. Cosmic time is related to redshift, and one of
the most important aspects of this work is the estimation of the photometric redshift. Thus we
present, in section 2.4 a description of the two methods used to measure the redshift, and their
differences. In section 2.5 we give a small discussion about dust, one of the components that
has an important impact in the light that comes from the stars of a galaxy and that greatly
influences the estimates of redshift and mass.

Section 3 contains a brief description of the data. Section 4 contains the software and
methodology used to perform our estimates and to filter the data. In subsection 4.1 we introduce
the software with a brief description of its use. In subsection 4.2 we describe the data selection
(in subsection 4.2.1). In subsection 4.2.2 we describe how the photometric catalogs for our
estimates were created, based on the 3D-HST photometric data available. The tests that were
performed on the photometric data are described from subsection 4.2.3 to 4.2.10. These tests
were conducted to determine the best method to analyze the data and produce reliable estimates.
In section 5 we present our estimates and compare them to the estimates from the 3D-HST
catalog. In section 6 we present the conclusions. The future work that is still to be done is
presented in section 7.

2.1 The Galaxies

A galaxy is a system with a stellar mass greater than \(10^6 M_\odot\) that contains a dark matter halo
and may also have gas and dust. All these components are bound by gravity and maintain its
stability due to the rotation in the case of a disk-like galaxy or velocity dispersion in the case of
an elliptical galaxy.

2.1.1 Galaxy morphological classification

In the 1920s Edwin Hubble (Hubble 1926) divided the galaxies into three major groups based
on a visual classification: Early-types, Late-types, and Irregulars. Figure 2.2 shows Hubble’s
visual classification diagram, the so-called Hubble Tuning Fork diagram.

![Figure 2.2: The Hubble's tuning fork. (Picture from: https://www.wikimedia.org/)](https://www.wikimedia.org/)
In the left, we have the Elliptical galaxies which are classified from E0 to E7. The capital E stands for Elliptical, and the number characterizes the visual aspect of the galaxy, where the E0 galaxy shows a complete spherical form, and as it goes to the E7 the form changes gradually into an ellipsoid, therefore the number indicates the galaxy’s ellipticity. In the center of the fork are the S0 which are called Lenticular galaxies. They are composed of a disk and a bulge, but they have no spiral arms. The E0 up to E7 and the S0 galaxies are called Early-type. In the upper right side, there are the Spiral galaxies. In their classification, the capital S stands for Spiral. The lower case letters are the classification of the bulge to arms size ratio. a is assigned to the galaxies where the bulge is more prominent than the arms. c is assigned to galaxies where the arms are more significant than the bulge. The lower right shows the barred spiral galaxies. These galaxies have a bar in their center. So the capital letters SB stand for Spiral Barred and the lower case letter stand for the bar-to-arms ratio classification, like the spirals. The Spirals and Barred Spirals are also known as Late-type galaxies.

Figure 2.3: Examples of five types of galaxies. The E0 on the upper left. (From: http://www.astro.cornell.edu). The upper right shows a Sa. (From: http://www.nasa.gov). In the lower left it is shown an SBa. (From: http://wikimedia.org). In the lower right an S0. (From: http://www.messier-objects.com). In the center an Irregular galaxy (From: http://www.hubblesite.org)

When Hubble published his visual galaxy classification, only a few nearby galaxies had been discovered, so in the beginning, he called the Elliptical galaxies of Early-types and the Spiral and Barred Spiral galaxies Late-types because he thought that the Ellipticals would be galaxies in the first stages of the collapse that would lead to Spirals. Nowadays, we know that Early-type galaxies are "dead" galaxies where there is little gas to allow star formation. On the other hand, Late-type galaxies continue to form stars from the gas that remains in the arms. In figure 2.3 examples of the 5 types can be seen. Therefore, in that sense Early-types should be Late-types, and viceversa.
2.1.2 The components of a galaxy

A galaxy has some basic components:

- The **bulge** is a spheroidal structure that appears in the center of spiral galaxies, and that is the most important component in spheroidal galaxies. It is composed of a compact group of mostly old stars, with almost no gas. In the middle of the bulge, most galaxies usually host supermassive black holes.

- The **bar** is a component that appears at the center of the Barred Spiral galaxies. This structure contains stars, gas, and dust.

- The **arms** are structures that are connected to the bulge or the central bar of the spiral type galaxies. They are composed of a mixed population of young and old stars. They have significant amounts of gas and dust. Therefore they are star formation regions. They also have a blue color due to the presence of the O and B class stars.

- The **disk** is an almost circular and thin structure composed of gas, dust, and stars. In the Spiral and Lenticular galaxies, it is the most critical component because it contains most of the stars.

- The **halo** is the nearly spherical region which contains the galaxy at the center and around it, a low density of stars and gas.

- The **dark matter halo** is composed of dark matter, and it is more massive than the stars that it contains. The nature of the dark matter is still unknown.

Apart from the components mentioned above, there might be some asymmetric galaxy components resulting from galaxy mergers, such as shells, tidal tails, etc.

2.1.3 The Sérsic Function

The Sérsic profile provides the form of the light intensity distribution of a galaxy. Although it does not directly show what type of galaxy is being observed, it shows the weight that the bulge and disk components have on the total intensity. It follows the following equation:

\[
I(R) = I_e \exp\left\{-b_n\left(\frac{R}{R_e}\right)^{\frac{1}{n}} - 1\right\}
\]

(2.1)

Figure 2.4: This plot shows the different shapes that the Sérsic function provides, depending on the value of \(n\).

Image from: Peng et al. (2002)
$I(R)$ is the surface intensity at a distance $R$ from the center. $I_e$ is the intensity at the effective radius $R_e$, the radius that encloses half of the total light. $b_n$ is a constant that depends on $n$ and as an approximation it may be calculated as $b_n = 1.9992n - 0.3271$, when $0.5 < n < 10$ and $n$ is the Sérsic index ([Graham and Driver 2005](#)).

When $n \approx 1$, it provides the profile of a spiral galaxy. If $n \approx 4$ it approaches the de Vaucouleurs’ profile, the profile of an elliptical galaxy. Figure 2.4 shows the shape of the function depending on the Sérsic index.

2.1.4 Massive galaxies

A goal of this work is to shed light on why there are so few massive disk-like galaxies in the local universe. The galaxy mass function (Figure 2.5) shows the dependency of the galaxy number density with stellar mass. As the mass increases, the number density decreases. The number density also depends on the morphology. As it is shown below in figure 2.5, the number density has a slow decrease with the increase of mass between $10^9 M_\odot$ to $10^{10.5} M_\odot$, above around $10^{11} M_\odot$ it decreases rapidly, making the massive galaxies rare. Additionally, the fraction of galaxies changes with morphology. For masses above $10^{11} M_\odot$, there are more elliptical than disk-like galaxies. Therefore, we need data from surveys that covered vast areas to have a significant number of massive galaxies with disk component.

![Figure 2.5](image-url)

Figure 2.5: On the top, it is shown the fraction of the different galaxy types regarding their masses. The lower plot shows the number density regarding mass and morphology. ([Kelvin et al. 2014](#))

The Press-Schechter function provides a parameterization for the galaxy mass function.

$$
\Phi = \Phi^* x^\alpha e^{-x} \tag{2.2}
$$

In this equation $\Phi$ is the number density of galaxies of a given mass, $\Phi^*$ is a normalization constant, $x$ is a mass fraction $x = \frac{M}{M^*}$, $M^*$ is a constant, and $\alpha$ is a slope power law.
When the masses of the galaxies are small $M << M^*$, the equation is dominated by a power law and the number density of galaxies becomes.

$$\Phi \propto x^\alpha \quad (2.3)$$

For massive galaxies $M >> M^*$, the equation becomes dominated by the exponential part.

$$\Phi \propto e^{-x} \quad (2.4)$$

It means that for galaxies with masses greater than $M^*$ the number density decreases exponentially. This $M^*$ takes values, for most literature (see for example Bell et al. 2003), $M^* \approx 10^{11} M_\odot$. Therefore for this work, we are considering this value as the limit to consider a massive galaxy.

### 2.2 The relation between distance and time

The $\Lambda$CDM model (Ostriker and Steinhardt 1995a) and the observations of the SNIa (Riess et al. 1998), shows that the Universe has $\Lambda > 0$ and therefore $\Omega_\Lambda > 0$, this means that the scale factor is increasing at an accelerated rate. Thus the galaxies that are not bound to the observer by gravity will appear as moving away. As the speed of light is finite, the light of a galaxy will take time to reach us, so the farther that the galaxy is, the farther in time we are observing. This way by looking at galaxies at different distances, we will be looking at various cosmic epochs. Therefore distance, age, and redshift can be connected through the Friedmann equation (2.5), with the values of the cosmological constants found in the cosmic concordance (Ostriker and Steinhardt 1995b).

$$H^2(a) = \Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda \quad (2.5)$$

Where $H(a)$ is the value of the Hubble constant for a given scale factor, $H(a_0) = H_0$ is the value of the Hubble constant for the present, where $a_0 = 1$. $\Omega_r$ is the radiation density constant, $\Omega_m$ is the matter density constant which takes into account both dark matter and baryonic matter, $\Omega_k$ is the spatial curvature density, and $\Omega_\Lambda$ is the cosmological constant. On the other hand, $H(a)$ relates to the scale factor and with time:

$$H(a) = \frac{\dot{a}}{a} = \frac{1}{a} \frac{da}{dt} \quad (2.6)$$

Also, the scale factor is related to redshift $z$ through the following equation.

$$1 + z = \frac{a(t_o)}{a(t_e)} \quad (2.7)$$

Where $a(t_o) = a_0$ is the scale factor at the time of the observation, $a(t_e)$ is the scale factor at the time the light was emitted from the target galaxy, and $z$ is the redshift of the galaxy. Integrating equation (2.6) we can relate redshift with distance, and we can also calculate the age of the Universe at that redshift. In the following plot are the computed values of distance and age for the several redshift values.

Figure 2.6 shows the connection between the redshift (top axis) and the age of the Universe (right-hand axis). According to the value of the constants used to calculate the values shown on the plot, at $z=0$, (Local Universe), the age of the Universe is approximately $13.49 \text{ Gyr}$. Half of the age of the Universe is at $z \approx 0.7$.

---

2SNIa stands for type Ia Supernovae. These supernovae are considered standard “candels” due to their similar luminosities. The supernova is triggered by a white dwarf that is part of a binary system and that accretes mass from its companion. When the white dwarf reaches the Chandrasekhar limit, it explodes. This means that all type Ia supernovae have similar light intensity curves. By comparing the light curves we can determine the distance.
2.3 The effect of the expansion on the light’s wavelength

The expansion of the Universe happens in the intergalactic space far from the potential wells of the galaxies and clusters. When a photon travels through these regions, it will be affected by a change in the metric. The wavelength of the photons increases with the growth of the scale factor.

\[
a(t_{em}) = \frac{\lambda_{em}}{\lambda_0} \tag{2.8}
\]

\[
a(t_{obs}) = \frac{\lambda_{obs}}{\lambda_0} \tag{2.9}
\]

Where \( \lambda_0 \) is the wavelength of the photon when measured without the effects of the expansion, \( \lambda_{em} \) being the wavelength when the photon was emitted and \( \lambda_{obs} \) the wavelength of the photon when it was received by the observer. Replacing (2.8) and (2.9) in (2.7).

\[
z = \frac{\lambda_{obs}}{\lambda_{em}} - 1 \leftrightarrow z = \frac{\lambda_{obs} - \lambda_{em}}{\lambda_{em}} \tag{2.10}
\]

When looking at the spectrum of a galaxy, there are features that can be recognized. For instance, the emission and absorption lines that appear due to the transition of atoms and molecules that exist in the interstellar medium and the photosphere of the stars. The wavelength of these lines are well known: we know the \( \lambda_{em} \). Therefore when we measure the \( \lambda_{obs} \), the (redshift) \( z \) of the galaxy can be calculated by using equation (2.10).

The table below shows the observed wavelengths of some lines regarding different \( z \).
Table 2.1: This table displays the wavelengths of three emission lines at different redshifts. The first column indicates the line’s name, in the second column the wavelengths of the different transitions at $z = 0$ or at the rest-frame, and on the third, fourth, fifth and sixth columns are the wavelengths for the emission lines seen from objects at redshift 0.6, 1.12, 2 and 3 respectively.

Table 2.1 shows the wavelength of three known emission lines at five different redshifts, as an example of how the cosmological redshift affects the spectrum of a galaxy. As the redshift increases so do the wavelength, hence the emission lines are shifted into the red. Thus, if we need to conduct observations to determine morphology, effective radius and Star Formation (SF), we need rest-frame UV and V bands. Due to the cosmological redshift, for very distant galaxies, ($z > 2$) these observations must be done in the IR. The observations in the IR are difficult especially if they are ground-based observations, due to the temperature of the instruments and the atmosphere that creates a lot of background noise, to solve this problem, the best option is to send instrumentation into space with an excellent cooling system.

2.4 Photometric and Spectroscopic redshifts

The calculation of redshifts is one of the most important features of this work. As we have seen in Section 2.2 it relates to distance and time and therefore allows to associate it with the evolving morphological and structural parameters of the selected population of galaxies. There are two main ways of determining the redshift of an extragalactic object: the spectroscopic redshift, and the photometric redshift.

The spectroscopic redshift is determined from the alterations that the galaxy spectrum undergoes due to the changes in the Universe’s expansion. The observed spectrum of a galaxy has essential markers like the absorption and the emission lines, and the Balmer and Lyman breaks which could be easily identified. As we know their rest-frame wavelengths, by using equation 2.10 we can determine the redshift. Figure 2.7 illustrates how a spectrum of a galaxy is displaced due to redshift.

Figure 2.7: In this picture is plotted the spectra of an object at different redshifts and in the background the transmission curves of the filters. We can see that as the redshift increases the peak shifts to the higher wavelengths, leaving a smaller signal in the filters that occupy the lower wavelengths. (Picture from: http://ogrisel.github.io)
The photometric redshift comes from an iterative comparison usually by a $\chi^2$ minimization algorithm of the galaxy model fluxes that come from a set of filters with the fluxes derived from a series of galaxy spectrum templates that may be either theoretical or empirical. The best template which has previously been corrected for extinction, metallicity, age, and redshift is known as the Spectral Energy Distribution (SED) of the galaxy. The redshift applied to the template is the estimated value.

The spectroscopic redshift method relies on a direct calculation of the redshift based on the observed and rest-frame wavelengths of features that can be identified in the spectrum of the galaxy (like the emission lines $H_\alpha$, $H_\beta$ and $Ly_\alpha$). The accuracy of the estimation depends on the number of features that may be used, and the resolution of the spectrograph. The most significant drawback of this method, comes from the integration time needed to get a spectrum, from where we can measure the needed features. Also this method only allows a small number of galaxies per exposure. The photometric method although it provides a less precise estimation of redshift, as it relies on images taken with different filters, it allows getting information on all the objects in the field-of-view at the same time, reducing this way the amount of time needed to perform the observation. The drawback of this method is that the error is firmly connected to the characteristics of the filters, the number of filters used to sample the SED, and the coverage of the spectrum that is provided by the filter set. As the bandwidth of the filters increases and the number of filters decreases, the error in the redshift estimation increases. Additionally, the fewer the number of filters available, the less constrained is the galaxy SED, therefore the worst is the accuracy.

2.5 The effects of dust

The dust is a product of stellar evolution. It is composed of the metals expelled into the ISM by supernovae, coronal mass ejections, solar winds and low mass stars that are in the red giant phase. When these materials cool down, they form dust. The dust particles interact with the radiation from the stars in the galaxy, creating an attenuation by absorbing the higher energy photons, of the UV and blue region of the spectrum, and at the same time re-emitting in the IR. Therefore, the object at study displays a redder color than it would. In Figure 2.8 it is shown in blue a dust-free SED and in red the same SED but with the effects of dust. Thus, for galaxies such as Spirals, Spiral-barred and Irregulars, where star formation occurs, dust may be found in considerable amounts.

![Figure 2.8: An example two SEDs. In blue a SED free of dust and in red with dust. Comparing the two an attenuation in the shorter wavelengths could be seen followed by an increase in the emissions at longer wavelengths. From Conroy 2013](image)

We will give a small introduction about what it is known about the effects of dust with regards to galaxy light. Calzetti, Kinney, and Storchi-Bergmann (1994) analyzed the slope of

---

3In astronomy all the elements of the periodic table apart from Hydrogen and Helium are called metals.
the UV continuum\(^4\) and the \(H_\alpha/H_\beta\)\(^5\) ratio from a sample of 39 Starburst(SB) galaxies. They found that the optical depth of the Balmer lines was almost twice the estimated for the UV continuum. From this, they speculated that this was due to two components in the absorption mechanism: On one hand, the absorption due to the dust that exists in the clouds that enclose the young stars; On the other hand, the old stars would be affected by the diffuse dust in the ISM(Figure 2.9).

Meurer, Heckman, and Calzetti (1999), after comparing the UV continuum \(\beta\), with the IRX\(^6\), reached a conclusion that the SB galaxies would define a specific plane. This was known as the Meurer relation. Charlot and Fall (2000), provided a model that not only takes into account the different regions that light must transverse until it reaches the intergalactic space, but also considers a period of transition from the epoch when the star is encased inside the birth cloud to having cleaned the surrounding region (typically of \(10^7\) years). Later on with the work of Bell et al. (2002), a difference was found between the Meurer relation and the observations done to the HII regions in the LMC (Large Magellanic Cloud). Followed by the work of Kong et al. (2004) and Boquien et al. (2012), where it was seen that other parameters could change the dust attenuation contribution, like the SFH, and the geometry of the distribution. The dust attenuation is one of the important topics when studying the evolution of galaxies because it can greatly influence the observations and posterior estimates made from the data collected during the observation.

3 The Data

To perform the study of the evolution of the massive galaxy disks, we needed data from the V band rest-frame of the galaxies with a good resolution, to have accurate measurements of the structural parameters (effective radius and the estimation of the Sérsic index), estimates of the mass and redshifts. Due to the cosmological redshift, these galaxies will have their spectrum shifted into the IR. Therefore we use as primary sources of information the CANDELS and 3D-HST surveys, that are based on observations of the HST on the five CANDELS fields: AEGIS, COSMOS, GOODS-N, GOODS-S and UDS.

\(^4\)The UV continuum is the blackbody component in the UV region.
\(^5\)The ratio between the two first transitions of the Balmer series.
\(^6\)The logarithm of the ratio between dust luminosities in the IR and UV bands \(IRX \equiv \log(L_{IR}/L_{UV})\)
<table>
<thead>
<tr>
<th>Field</th>
<th>RA (h m s)</th>
<th>Dec (d m s)</th>
<th>Total Area (arcmin$^2$)</th>
<th>Science Area (arcmin$^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AEGIS</td>
<td>14 18 36.00</td>
<td>+52 39 0.00</td>
<td>201</td>
<td>192.4</td>
</tr>
<tr>
<td>COSMOS</td>
<td>10 00 31.00</td>
<td>+02 24 0.00</td>
<td>199</td>
<td>183.9</td>
</tr>
<tr>
<td>GOODS-N</td>
<td>2 35 54.98</td>
<td>+62 11 51.3</td>
<td>164</td>
<td>157.8</td>
</tr>
<tr>
<td>GOODS-S</td>
<td>03 32 30.00</td>
<td>-27 47 19.00</td>
<td>177</td>
<td>171.1</td>
</tr>
<tr>
<td>UDS</td>
<td>02 17 49.00</td>
<td>-05 12 2.00</td>
<td>201</td>
<td>191.2</td>
</tr>
</tbody>
</table>

Table 3.1: Information of the CANDELS fields. (Skelton et al. 2014)

Figure 3.1 shows the throughput of the ACS and WFC3 cameras filters used in the observations conducted with the HST that cover the spectrum from the IR to the UV.

The data from these observations was then reduced to create the two primary catalogs that our work is based upon, the CANDELS and the 3D-HST catalogs. The CANDELS catalog (Van Der Wel et al. 2013) provided the parametric fits of 109533 objects that were detected using the H band (HST filter F160W). This catalog provides single Sérsic function fits to the detected objects in all near-infrared bands provided by the HST observations: $H_{F160W}$, $J_{F125W}$ and $Y_{F105W}$. The best fits were then used to extract the structural parameters, namely the total magnitude, the half-light radius, Sérsic index, axis ratio and position angle. The other catalog used was the 3D-HST (Skelton et al. 2014), created using the data that came from the list in Table 3.3, which provided photometric data (i.e. fluxes per filter) used to estimate the photometric redshifts. The authors ran their own estimates by using the EAZY\footnote{EAZY is a code used to estimate photometric redshift by fitting SED templates to photometric data. (http://www.astro.yale.edu/eazy/)} code and also stellar population parameters using the FAST\footnote{FAST is a code used to estimate stellar population parameters by fitting photometric data to stellar population templates. (http://w.astro.berkeley.edu/ mariska/FAST.html)} code.

The 3D-HST survey is divided in six catalogs, one per field and a master catalog. The master catalog contains the photometric data and the associated 1σ errors for the observations with the HST WFC3 (Wide Field Camera 3) and ACS (Advanced Camera for Surveys) filters. It also includes photometric and spectroscopic redshift estimates\footnote{The spectroscopic redshift is not available for all galaxies due to the constraints referred in section 2.4}, data acquisition parameters and positioning information, that is the centroid location coordinates and the RA and Dec of the object for all the CANDELS fields. Table 3.2 shows the number of detections in each field.
Table 3.2: The number of detected objects in the CANDELS fields.

<table>
<thead>
<tr>
<th>Field</th>
<th>Number of objects</th>
</tr>
</thead>
<tbody>
<tr>
<td>AEGIS</td>
<td>41200</td>
</tr>
<tr>
<td>COSMOS</td>
<td>33879</td>
</tr>
<tr>
<td>GOODS-N</td>
<td>38279</td>
</tr>
<tr>
<td>GOODS-S</td>
<td>50507</td>
</tr>
<tr>
<td>UDS</td>
<td>44102</td>
</tr>
<tr>
<td>All fields</td>
<td>207967</td>
</tr>
</tbody>
</table>

About the catalogs regarding each individual field, apart from containing data from the HST observations, also contain data from other surveys that complement each field separately. The information about each field may be seen in table 3.3 where we present the telescopes, instruments and filter sets used in surveys that complement each of the fields.

As mentioned in section 2.1.3, the Sérisc index only provides information about the brightness contribution of the arms and bulge components to the overall surface brightness. This may give an indication of the galaxy type, to ensure that the galaxy has a disk component a visual-like morphological classification is needed. For this, we added the results from Huertas-Company et al. (2015) that are a morphological classification catalog for the CANDELS fields also based upon the work of Skelton et al. (2014) and Van Der Wel et al. (2013). This catalog is the result of a visual-like classification provided by a Convolutional Neural Network (ConvNets). The neural network was previously trained resorting to previous results from visual classification done in the GOODS-S field. Afterward, it was used to classify the detected objects in all the fields providing visual-like classification for $\approx 50000$ galaxies with a fraction of mis-classifications of less than 1%.

The catalog itself contains a column for ID numbers, the positions of the objects in RA and Dec, estimates of redshifts and associated errors, and the last columns contain the values of the output flags for the morphological classification of the galaxies. For each of the classified galaxies the neural network outputs a series of values that are the weight that each of the components has on the galaxy. The main flags used in this work are $f_{Sph}$, $f_{Disk}$ and $f_{Irr}$, which represent respectively the weights of the Spheroid, Disk and Irregular components. With these three flags all the galaxies can be classified. For instance a Early-type, which is a spheroid will have a bigger weight on $f_{Sph}$ and a smaller weight on the other two flags $f_{Disk}$ and $f_{Irr}$. If the object is a Late-type galaxy, the weight will be greater in the $f_{Disk}$, or in the case of an Irregular galaxy or a collision the $f_{Irr}$ will carry most of the weight. A larger contribution on the $f_{Disk}$ flag means that the galaxy has a meaningful disk component and therefore is of interest for this work. There are also the $f_{PS}$ and $f_{UC}$. The PS flag means point-spread function while UC means unclassified. With these flags, we may know if the target is a star (with a greater value with $f_{PS}$), or if $f_{UC}$ has a greater value means that the neural network could not classify the object.
<table>
<thead>
<tr>
<th>Field</th>
<th>Filters</th>
<th>Telescope/Instrument</th>
<th>Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>AEGIS</td>
<td>u, g, r, i, z</td>
<td>CFHT/MegaCam</td>
<td>CFHTLS</td>
</tr>
<tr>
<td></td>
<td>F606W, F814W</td>
<td>HST/ACS</td>
<td>CANDELS</td>
</tr>
<tr>
<td></td>
<td>J1, J2, J3, H1, H2, K</td>
<td>KPNO 4m/NEWFIRM</td>
<td>NMBS</td>
</tr>
<tr>
<td></td>
<td>J, H, Ks</td>
<td>CFHT/WIRCam</td>
<td>WIRDS</td>
</tr>
<tr>
<td></td>
<td>F140W</td>
<td>HST/WFC3</td>
<td>3D-HST</td>
</tr>
<tr>
<td></td>
<td>F125W, F160W</td>
<td>HST/WFC3</td>
<td>CANDELS</td>
</tr>
<tr>
<td></td>
<td>3.6, 4.5(\mu m)</td>
<td>Spitzer/IRAC</td>
<td>SEDS</td>
</tr>
<tr>
<td></td>
<td>5.8, 8(\mu m)</td>
<td>Spitzer/IRAC</td>
<td>EGS</td>
</tr>
<tr>
<td>COSMOS</td>
<td>u, g, r, i, z</td>
<td>CFHT/MegaCam</td>
<td>CFHTLS</td>
</tr>
<tr>
<td></td>
<td>B, V, r', i', z', 12 medium − band</td>
<td>Subaru/Suprime-Cam</td>
<td>CANDELS</td>
</tr>
<tr>
<td></td>
<td>F606W, F814W</td>
<td>HST/ACS</td>
<td>NMBS</td>
</tr>
<tr>
<td></td>
<td>J1, J2, J3, H1, H2, K</td>
<td>KPNO 4m/NEWFIRM</td>
<td>WIRDS</td>
</tr>
<tr>
<td></td>
<td>J, H, Ks</td>
<td>CFHT/WIRCam</td>
<td>UltraVISTA</td>
</tr>
<tr>
<td></td>
<td>Y, J, H, Ks</td>
<td>VISTA</td>
<td>3D-HST</td>
</tr>
<tr>
<td></td>
<td>F140W</td>
<td>HST/WFC3</td>
<td>CANDELS</td>
</tr>
<tr>
<td></td>
<td>F125W, F160W</td>
<td>HST/WFC3</td>
<td>SEDS</td>
</tr>
<tr>
<td></td>
<td>3.6, 4.5(\mu m)</td>
<td>Spitzer/IRAC</td>
<td>S-COSMOS</td>
</tr>
<tr>
<td></td>
<td>5.8, 8(\mu m)</td>
<td>Spitzer/IRAC</td>
<td></td>
</tr>
<tr>
<td>GOODS-N</td>
<td>U</td>
<td>KPNOn 300m/NEWFIRM</td>
<td>Hawaii HDFN</td>
</tr>
<tr>
<td></td>
<td>G, R_S</td>
<td>Keck/LRIS</td>
<td>GOODS</td>
</tr>
<tr>
<td></td>
<td>F435W, F606W, F775W, F850LP</td>
<td>Subaru/Suprime-Cam</td>
<td>Hawaii HDFN</td>
</tr>
<tr>
<td></td>
<td>B, V, R'_C, I'_C, z'</td>
<td>HST/WFC3</td>
<td>3D-HST</td>
</tr>
<tr>
<td></td>
<td>F140W</td>
<td>HST/WFC3</td>
<td>CANDELS</td>
</tr>
<tr>
<td></td>
<td>F125W, F160W</td>
<td>Subaru/MOIRCS</td>
<td>MODS</td>
</tr>
<tr>
<td></td>
<td>J, H, Ks</td>
<td>Spitzer/IRAC</td>
<td>SEDS</td>
</tr>
<tr>
<td></td>
<td>3.6, 4.5(\mu m)</td>
<td>Spitzer/IRAC</td>
<td>GOODS</td>
</tr>
<tr>
<td></td>
<td>5.8, 8(\mu m)</td>
<td>Spitzer/IRAC</td>
<td></td>
</tr>
<tr>
<td>GOODS-S</td>
<td>U, R</td>
<td>VLT/VIMOS</td>
<td>ESO/GOODS</td>
</tr>
<tr>
<td></td>
<td>U38, B, V, R_C, I</td>
<td>WFI 2.2m</td>
<td>GaBoDs</td>
</tr>
<tr>
<td></td>
<td>14 medium bands</td>
<td>Subaru/Suprime-Cam</td>
<td>MUSYC</td>
</tr>
<tr>
<td></td>
<td>F435W, F606W, F775W, F850LP</td>
<td>HST/ACS</td>
<td>GOODS</td>
</tr>
<tr>
<td></td>
<td>F606W, F814W, F850LP</td>
<td>HST/ACS</td>
<td>CANDELS</td>
</tr>
<tr>
<td></td>
<td>F140W</td>
<td>HST/WFC3</td>
<td>3D-HST</td>
</tr>
<tr>
<td></td>
<td>F125W, F160W</td>
<td>HST/WFC3</td>
<td>CANDELS</td>
</tr>
<tr>
<td></td>
<td>J, H, Ks</td>
<td>VLT/ISAAC</td>
<td>GOODS,FIREWORKS</td>
</tr>
<tr>
<td></td>
<td>J, Ks</td>
<td>CFHT/WIRCam</td>
<td>TENIS</td>
</tr>
<tr>
<td></td>
<td>3.6, 4.5(\mu m)</td>
<td>Spitzer/IRAC</td>
<td>SEDS</td>
</tr>
<tr>
<td></td>
<td>5.8, 8(\mu m)</td>
<td>Spitzer/IRAC</td>
<td>GOODS</td>
</tr>
<tr>
<td>UDS</td>
<td>u, g, r, i, z</td>
<td>CFHT/MegaCam</td>
<td>CFHTLS</td>
</tr>
<tr>
<td></td>
<td>F606W, F814W</td>
<td>HST/ACS</td>
<td>CANDELS</td>
</tr>
<tr>
<td></td>
<td>J1, J2, J3, H1, H2, K</td>
<td>KPNO 4m/NEWFIRM</td>
<td>NMBS</td>
</tr>
<tr>
<td></td>
<td>J, H, Ks</td>
<td>CFHT/WIRCam</td>
<td>WIRDS</td>
</tr>
<tr>
<td></td>
<td>F140W</td>
<td>HST/WFC3</td>
<td>3D-HST</td>
</tr>
<tr>
<td></td>
<td>F125W, F160W</td>
<td>HST/WFC3</td>
<td>CANDELS</td>
</tr>
<tr>
<td></td>
<td>3.6, 4.5(\mu m)</td>
<td>Spitzer/IRAC</td>
<td>SEDS</td>
</tr>
<tr>
<td></td>
<td>5.8, 8(\mu m)</td>
<td>Spitzer/IRAC</td>
<td>EGS</td>
</tr>
</tbody>
</table>

Table 3.3: List of the photometric data included in the individual CANDELS fields of the 3D-HST survey (From: Skelton et al. 2014).
4 Methodology

We present here the steps and optimizations that were taken to achieve the best estimates of redshift, mass, and other stellar population parameters, from the selected sample of galaxies. In Section 4.1 we present the software package and tools used to produce our estimates and to visualize the data and results, namely LePHARE, TOPCAT, Aladin, and SExtractor. In section 4.2 we describe how we selected the sample of galaxies and also the tests that were performed to improve the accuracy of our redshift estimates.

4.1 Software packages

4.1.1 LePHARE (PHotometric Analysis for Redshift Estimate)

LePHARE (Arnouts et al. 1999 and Arnouts and Ilbert 2006) is a software package written in Fortran that is used to estimate parameters through the comparison of the observed photometric points with templates. LePHARE is composed of several routines which are dedicated to perform certain tasks, the user has to follow a specific workflow, that is shown in figure 4.1. For more information on them I refer to Arnouts and Ilbert (2006).

![Figure 4.1: The workflow of the LePHARE package where we can see the diagram of the two phases. In the preparation phase, first the SED libraries are created, than filter sets and afterward the magnitude library. The software package compares the catalog data with the libraries to find the best fits in the z-photometric phase. Finally, in the simulated catalog phase it generates data according to the restrictions presented in the configuration files. (Arnouts and Ilbert 2006)](image)

LePHARE needs two configuration files. The zphot.para file contains the configuration of all the routines shown in the diagram above. The second configuration file, the zphot.output.para has the list of parameters that are written in the output catalog. The workflow is divided into two phases. The Preparation phase, which is divided into three steps. In the first step, we use the sedtolib routine that has its configurations in the first section of the configuration file. LePHARE has sets of observational and theoretical SEDs in subdirectories in the relative path $\$LEPHAREDIR/sed/GAL/$. These SED are selected via the configuration variable
GAL\_SED, with a range of possible ages set in the \textit{AGE\_RANGE} variable. With these configurations the \textit{sedtolib} routine creates a binary library of all the possible SEDs within the defined restrictions. In the \textbf{second step}, we create a filter set that contains the data of the transmission of the filters used in the observations. This is done by the routine \textit{filters} configured in the section \textit{Filters} of the configuration file. In the \textbf{third step} we use the routine \textit{mag\_gal} that is configured in the \textit{Theoretical magnitudes} section of the configuration file. The configuration variable \textit{GAL\_LIB\_IN} provides the path and name of the SED binary library created in the first step. The \textit{Z\_STEP} sets the redshift range applied to the SEDs, and the \textit{COSMOLOGY} variable allows to constrain the models, by eliminating models that are older than the age of the Universe. We also provide the extinction law that is applied via the \textit{EXTINC\_LAW} variable and the reddening range with the \textit{EB\_V} variable sets the range of reddening allowed. With these configurations the routine \textit{mag\_gal} adds to the templates the redshift and the reddening produced by the selected extinction law. Afterward, it convolves the filters with all the possible SEDs and outputs the results into a library of magnitudes stored in the relative path set by the variable \textit{GAL\_LIB\_OUT}.

In the \textit{z-photometric} phase we use the \textit{zphota} routine to calculate the photometric redshift and other physical parameters (namely, stellar mass, star formation rate and age). Apart from the already referred calculations, this routine may also detect systematics between the photometry from the observations and the SEDs from the templates and produce the required offsets to compensate. Solely for the estimates, the \textit{zphota} routine compares the data from the photometry catalog of the observations with the magnitude library created in the last step of the \textit{Preparation} phase, by using a $\chi^2$ minimization algorithm to find the best fit. The other physical parameters are then computed and stored in the output file \textit{zphot.out}. The best SEDs for each galaxy are also stored in the local directory with the extension \textit{.spec}. To visualize the data from these files, we used the program PAPS\textsuperscript{10} to detect and compensate the systematics, the photometric catalog must be of the type \textit{LONG}, that has two more columns that are the context and spectroscopic redshift columns. The context column contains values that encode the bands that the routine uses for the photometric estimates. The context number it calculated has follow.

\begin{equation}
context = \sum_{i=1\ldots N}^{} 2^{i-1}
\end{equation}

The spectroscopic redshift column contains the redshift of some of the galaxies of the field which happen to have spectroscopic redshifts. To detect the systematics the configuration file must have the \textit{AUTO\_ADAPT} variable set to \textit{YES}. The \textit{zphota} routine finds the best fit for the given spectroscopic redshift and computes the offsets that may be applied to the photometry to compensate for the systematics, and this results in a list of offsets. Then the \textit{AUTO\_ADAPT} is set to \textit{NO}, and the offsets are set in the \textit{APPLY\_SYSSHIFT} variable resulting in the offset being added to the values in the magnitude library when the photometric estimates are run. The output of the \textit{zphota} routine is a catalog that contains the estimates of the physical characteristics extracted from the photometry, these characteristics are selected in the \textit{zphot\_output\_para}. The name of the output catalog file is configured in the Photometric Redshift section of the \textit{zphot\_para}, by the \textit{CAT\_OUT} variable. Appendix \textbf{B.3} contains an example of the \textit{zphot\_para} file with the configurations for the EGS field.

\textsuperscript{10}PAPS (PArsing And Plotting Spectra) is a visualization program also developed by Adam Draginda at the CFHT. As it does not have a dedicated webpage, we will refer the LePHARE internal link. (http://www.cfht.hawaii.edu/\textasciitilde arnouts/LEPHARE/paps.html)
4.1.2 TOPCAT (Tool for OPerations on Catalogues And Tables)

TOPCAT\textsuperscript{11} (Taylor 2005) is a Java-based program used to manipulate, make calculations and show graphically the data from extensive catalogs related to astrophysics and cosmology. In this work TOPCAT was used to filter the data we needed from the catalogs, to calculate additional values, and to produce some of the plots presented in this document. Picture 4.2 shows TOPCAT's main window, from where all the data can be accessed.

Figure 4.2: The TOPCAT main window. On the left is the list of the loaded tables, on the right the selected tables properties, on the bottom right the connections and connection information. On the top there is the tables operations.)

This program can read and store data from different file types, from FITS to simple ASCII files with simple column separators. TOPCAT also has an internal library that allows easy and fast computation of cosmological and astrophysical parameters as well as the statistics of the data. Also, it can exchange data with other programs like Aladin and DS9 via a protocol that is shared by these programs. This feature was used to help with the visualization of the candidate objects. Picture 4.3 below shows the table display window, with the first columns of the table all\_mg\_dsk\_all\_z.fits and, above the table, it shows the available operations for this table. The subset option is one of the most used. It allows the sorting of data, for instance with it one can sort all the galaxies with a mass greater than a certain quantity. TOPCAT also allows the graphical visualization of data. The picture 4.4 below shows an example.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{topcat_window.png}
\caption{The TOPCAT table display window. Above are the dropdown menus and the icons for table operations used to create subsets of data based on rules based on mathematical relations, and below it is shown the columns and then associated data.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{topcat_table.png}
\caption{The TOPCAT table display window. Above are the dropdown menus and the icons for table operations used to create subsets of data based on rules based on mathematical relations, and below it is shown the columns and then associated data.}
\end{figure}

\textsuperscript{11}\url{http://www.star.bris.ac.uk/~mbt/topcat/}
4.1.3 Aladin (Sky Atlas)

Aladin\textsuperscript{12} (Boch and Fernique 2014, Bonnarel et al. 2000), is a Java-based program that works as a sky atlas and viewer. It can gather and compile data from several databases over the internet as well as read local files containing observed fields and assign a coordinate system to those fields in order for the user to have a match with the position of the objects of interest.

\textsuperscript{12}http://aladin.u-strasbg.fr/
This program uses a communication protocol that allows it to receive data from other programs like TOPCAT in order to display not only the image of the selected object but also the available information from local files or from internet databases. Figure 4.5 shows an example of the interaction with TOPCAT. In the center-right it shows two layers: bottom one is the image of the UDS field and the layer above marks the galaxies that are in the TOPCAT catalog called \texttt{all\_ng\_disk\_all\_z.fits}. The row on the bottom shows all the information in the catalog regarding that object.

4.1.4 Source Extractor (SExtractor)

SExtractor\footnote{https://www.astromatic.net/software/sextractor} (Bertin and Arnouts 1996) is a software package that can detect sources and extract photometric data from FITS files. SExtractor was run for the creation of the 3D-HST catalog which provides with the photometric points for our study. First it measures the background in order to establish a threshold. Afterward the pixels that compose the image are compared to the threshold to determine if it detected a source. The detections pass through a deblending process where they can be broken into several parts. The pixels from the areas containing each of the detected sources are than summed to get the photometric points for each source.

4.2 Analysis

The data of our sample of galaxies are provided mainly by two sources, the 3D-HST (Skelton et al. 2014) and CANDELS (Van Der Wel et al. 2013) catalogs. Due to the necessity of having a visual morphology selection, we also added the Huertas-Company work (Huertas-Company et al. 2016). In section 4.2.1, we describe how the information from these catalogs was used to select the galaxies from our sample. To ascertain the accuracy of our estimates we used the 3D-HST estimates as a benchmark. In section 4.2.2, we explain how we used the photometric data with LePHARE to perform our estimates and the differences between the 3D-HST and our methods. From section 4.2.3 to 4.2.10 we describe the tests that were done to ascertain the methodology.

4.2.1 The data selection

There are two base catalogs for this work. CANDELS (Van Der Wel et al. 2013) provides the structural parameters: estimates of the Sersic index, effective radius (which is the ratio of the semi-axis of the ellipse containing the galaxy), the axis ratio and the position angle. 3D-HST (Skelton et al. 2014) on the other hand provides the photometry, redshift and mass estimates. Also as ancillary data, we use the Huertas-Company et al. (2016) results to aid in the selection of the massive galaxies with the disk components. The first step was to combine these published results using TOPCAT to create a master catalog where the data could be more accessible. Afterward, the galaxies were separated in two redshift bins, $0.6 < z < 1.12$ and $2 < z < 3$ due to the displacement of the restframe V-band to the observed H-band and z-band.

As the catalogs from CANDELS and 3D-HST do not provide a morphological classification, the results of Huertas-Company et al. (2016) were added to the master catalog. These results are the output of a Convolution Neural Network used for the visual morphology classification. The ConvNet was first trained with the available visual morphological data available from the GOODS-S field. Afterward, it was used to classify the remaining 4 fields. The catalog has five flags that contain a measurement of the frequency that a classifier flags a galaxy has having a certain feature: \( f_{\text{sph}} \) corresponding to the spheroid component, \( f_{\text{disk}} \) the disk component, \( f_{\text{Irr}} \) the irregular component, \( f_{\text{PS}} \) the point spread component, \( f_{\text{UC}} \) this last tells if the neural network could identify the object or not.

Within the paper there was a suggestion of the selection criteria:
1. Pure bulges or Elliptical galaxies: \( f_{sph} > 2/3 \land f_{disk} < 2/3 \land f_{Irr} < 1/10 \)

2. Pure disks: \( f_{sph} < 2/3 \land f_{disk} > 2/3 \land f_{Irr} < 1/10 \)

3. Spheroid and Disk: \( f_{sph} > 2/3 \land f_{disk} > 2/3 \land f_{Irr} < 1/10 \)

4. Irregular disks: \( f_{sph} < 2/3 \land f_{disk} > 2/3 \land f_{Irr} > 1/10 \)

5. Irregular or Mergers: \( f_{sph} < 2/3 \land f_{disk} < 2/3 \land f_{Irr} > 1/10 \)

The galaxies with disk component were then selected resorting to the selection criteria 2, 3 and 4, which resulted in 268 galaxies in the five CANDELS fields. In order to complete the master catalog the effective radius in kpc had to be calculated. The effective radius in CANDELS was given in arcsec. In order to compute this quantity, we must first calculate the luminosity distance with the TOPCAT command.

\[
DL = \text{luminosityDistance}(z_{\text{final}}, 70, 0.3, 0.7)
\] (4.2)

Where the input parameters are as follows, the best redshift, \( H_0, \Omega_m \) and \( \Omega_\Lambda \). Afterward the effective radius of the galaxies was calculated via the expression

\[
re_{\text{kpc}} = \frac{2\pi}{360 \times 60 \times 60} \frac{DL \times 10^3}{(1 + z_{\text{final}})^2}
\] (4.3)

\( re_{\text{kpc}} \) is the output of the effective radius in kpc, \( re \) is the effective radius measured in arcseconds, \( DL \) is the luminosity distance calculated by equation 4.2, the \( 10^3 \) factor comes from the fact that the DL output is in \( Mpc \) instead of kpc and \( z_{\text{final}} \) is the best redshift estimation. The galaxy spheroid sizes depend on their projection due to their triaxial nature. Traditionally, the impact of this uncertainty has been mitigated by using the so-called circularized effective radius. It was calculated and stored in the column \( re_{\text{circ}} \).

\[
re_{\text{circ}} = re_{\text{kpc}} \sqrt{q}
\] (4.4)

\( re_{\text{circ}} \) is the circularized radius, \( re_{\text{kpc}} \) is the effective radius and \( q \) the ratio between the minor and major semi-axes. The \( \text{lmass} \) column stores the logaritm of the masses of the galaxies in units of solar masses.

\[
\text{lmass} = \log \frac{M_*}{M_\odot}
\] (4.5)

The selected 268 galaxies were afterward inspected visually to confirm that they were not image artifacts. The inspection was done resorting to the program Aladin and the mosaics of the CANDELS fields that were created from a series of images taken by HST in the \( H_{F160W} \) band (The reddest band, better suited to observe old stellar populations, as expected for massive galaxies). This final master catalog contains the estimates (namely of redshift, mass, size, Sérsic index, and other parameters) of the 3D-HST and CANDELS teams. It was produced as an exercise and also to compare to our estimates, to ascertain our accuracy in the determination of redshift and mass.

4.2.2 Understanding the data

An important goal of this dissertation was to obtain experience with LePHARE in order to produce our own estimates of redshift and mass. Therefore a considerable part of the work was understanding how LePHARE works and mainly how the photometric data of the catalogs could be used to extract the needed information. As a first approach we tried to replicate Skelton et al. (2014) results, and use it as a benchmark to compare with the estimates of the tests performed with different configurations, so that a methodology could be defined. The different
CANDELS fields have data from separate surveys which means that a different configuration file has to be made for each field. Also, LePHARE needs a file with the throughput of every filter used in each survey. The AEGIS field was the selected field to be a sandbox where all tests were conducted due to its characteristics. It is one of the largest of the CANDELS fields which increases the probability of finding the galaxies that fit our criteria, and it has a broad photometric coverage. It contains 69 galaxies that meet the requirements of being massive and that have a disk component. From these, 9 have spectroscopic redshifts that can be used to compare the quality of our photometric redshift estimation.

Before LePHARE could run with the photometry of this field, the files that contained the throughput of the filters had to be created. As it would take too much time to find all of the information for each filter, I contacted Doctor Rosalind Skelton who kindly sent a file containing all the information needed. The file that I received contained all the filters so, to make the separation, a script was written in MATLAB (The code is presented in appendix C.2). To distinguish these new files from the files of the filters that LePHARE already has, a new directory was created in $LEPHAREDIR/filt[14].

The original photometric catalogs contain fluxes and errors for each filter, but the units were not presented in the catalog documentation. It contained only the equation with the conversion to AB magnitudes. A MATLAB script was created to convert fluxes into magnitudes and also to convert the associated errors with the following equations 4.6 and 4.7.

\[
m_{AB} = 25 - 2.5 \log(f) \tag{4.6}
\]

Where \(m_{AB}\) is the measured AB magnitude value, considering the zero-point to be 25, \(f\) is the flux, and the associated error is calculated as follows

\[
e_{AB} = 2.5 \log \left(1 + \frac{e_f}{f}\right) \tag{4.7}
\]

where \(e_{AB}\) is the error associated to the calculated AB magnitude, \(e_f\) is the error of the flux measurement, and \(f\) is the flux.

A series of tests were conducted to understand the data, and to determine the methodology to be used in this work. Each test is an iteration where a single change is made to the configurations and the catalog data. Tests A and B, are run with the catalog type SHORT, containing a column with the ID of the galaxies, and then columns with the values of magnitude and associated errors measured in the different bands. The test from C to H are conducted with the LONG catalog type that contains two extra columns at the end with the context and spectroscopic redshift (when available). For the tests A, B and C, we consider discarded the negative flux, by setting the respective flux and error to -99. Therefore in these first tests, we are not considering upper limits. Test A was conducted with the default configuration, to see if the code was running correctly and if the output made sense. In test B, we set the minimum error to be 0.1 magnitudes. This was done to provide a margin for LePHARE to explore more solutions that could fit better the photometry. In test C we used the AUTO_ADAPT and APPLY_SYSSHIFT features to try to compensate for possible systematics. The tests from D to H have upper limits. An upper limit corresponds to the highest magnitude that signal can be considered a detection. Therefore if a catalog entry is an upper limit, it means that if there is a signal in that band, it is too faint to be detected. Thus LePHARE will discard all solutions are brighter than that limit, this way constraining the possible solutions. In test D we have considered all catalog entries that have a negative flux to be upper limits that are set to be the $5\sigma$ depth magnitude published with the 3D-HST catalog (Skelton et al. 2014). In test E, the upper limits values are calculated as five times the error associated with a negative flux entry. For test F we used the same upper limits as in test E, and we increased the color excess interval to 1.5. In test G all catalog entries

---

14This is the default directory that contains the directories which have the filter files from different instruments.
with a signal-to-noise \((S/N < 5)\) were set as an upper limit with a value that was five times the associated error. For test H, the IRAC 3 and 4 bands were suppressed by setting their values and associated errors to -99. Table 4.1 shows the list of tests.

<table>
<thead>
<tr>
<th>Test</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>SHORT type catalog, with default configuration</td>
</tr>
<tr>
<td>B</td>
<td>SHORT type catalog, with a minimum error of 0.1 magnitude</td>
</tr>
<tr>
<td>C</td>
<td>LONG type catalog, using AUTO_ADAPT and APPLY_SYSSHIFT</td>
</tr>
<tr>
<td>D</td>
<td>LONG type catalog, with fixed upper limits</td>
</tr>
<tr>
<td>E</td>
<td>LONG type catalog, with 5(\sigma) upper limits</td>
</tr>
<tr>
<td>F</td>
<td>LONG type catalog, with 5(\sigma) upper limits and max(E(B-V))=1.5</td>
</tr>
<tr>
<td>G</td>
<td>LONG type catalog, with max(E(B-V))=1.5 and considering upper limits from (S/N &lt; 5)</td>
</tr>
<tr>
<td>H</td>
<td>LONG type catalog, with max(E(B-V))=1.5, considering upper limits from (S/N &lt; 5) and no IRAC 3 and 4 bands</td>
</tr>
</tbody>
</table>

Table 4.1: List of tests performed to the catalog data.

The 3D-HST photometric redshift estimates were performed with the EAZY software package (Brammer, Dokkum, and Coppi 2008), using only five SED templates of different galaxies, with fixed ages, to cover the different of possibilities. While for this work, we are using the BC03 library (Bruzual and Charlot 2003), which is a theoretical SED library that contains 27 SED templates that have three metallicities with 9 star formation timescales \(\tau\), using the Chabrier initial mass function (Chabrier 2003), and no dust. The characteristics that define each template are shown in table 4.2 below.

The number of models produced to be fitted to the photometry is calculated as follows:

\[
N_{Total} = N_{Template} \times N_{Extinc.Laws} \times N_{E(B-V)} \times N_{Zsteps} \times N_{Ages} \tag{4.8}
\]

where, \(N_{Template}\) is the number of templates of the library, \(N_{Extinc.Laws}\) is the number of extinction laws applied to the templates, \(N_{E(B-V)}\) is the number of reddening values used, \(N_{Zsteps}\) is the number of redshift steps and \(N_{Ages}\) is the number of ages. The dust effects are defined in the configuration file zphot.para with the extinction law \((EXTINC.LAW)\) and the excess color \((EB.V)\). According to the configurations used in this work, we used 328860 models. As the number of models increase, the quality of the estimates may also increase. The reddening, redshift, metallicity and age are the four factors that make a SED redder, therefore as the number of models increase, the probability of having models that produce similar SEDs also increases and this produces degeneracy in the solutions which leads to unwanted solutions.
Table 4.2: The characteristics of the BC03 templates. The first column shows the template number, the second column shows the metallicity in units of the solar metallicity, and the third column contains the star formation timescale in Gyr.

<table>
<thead>
<tr>
<th>Template</th>
<th>Metallicity (Z⊙)</th>
<th>τ (Gyr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20%</td>
<td>0.1</td>
</tr>
<tr>
<td>2</td>
<td>20%</td>
<td>0.3</td>
</tr>
<tr>
<td>3</td>
<td>20%</td>
<td>1.0</td>
</tr>
<tr>
<td>4</td>
<td>20%</td>
<td>2.0</td>
</tr>
<tr>
<td>5</td>
<td>20%</td>
<td>3.0</td>
</tr>
<tr>
<td>6</td>
<td>20%</td>
<td>5.0</td>
</tr>
<tr>
<td>7</td>
<td>20%</td>
<td>10.0</td>
</tr>
<tr>
<td>8</td>
<td>20%</td>
<td>15.0</td>
</tr>
<tr>
<td>9</td>
<td>20%</td>
<td>30.0</td>
</tr>
<tr>
<td>10</td>
<td>40%</td>
<td>0.1</td>
</tr>
<tr>
<td>11</td>
<td>40%</td>
<td>0.3</td>
</tr>
<tr>
<td>12</td>
<td>40%</td>
<td>1.0</td>
</tr>
<tr>
<td>13</td>
<td>40%</td>
<td>2.0</td>
</tr>
<tr>
<td>14</td>
<td>40%</td>
<td>3.0</td>
</tr>
<tr>
<td>15</td>
<td>40%</td>
<td>5.0</td>
</tr>
<tr>
<td>16</td>
<td>40%</td>
<td>10.0</td>
</tr>
<tr>
<td>17</td>
<td>40%</td>
<td>15.0</td>
</tr>
<tr>
<td>18</td>
<td>40%</td>
<td>30.0</td>
</tr>
<tr>
<td>19</td>
<td>100%</td>
<td>0.1</td>
</tr>
<tr>
<td>20</td>
<td>100%</td>
<td>0.3</td>
</tr>
<tr>
<td>21</td>
<td>100%</td>
<td>1.0</td>
</tr>
<tr>
<td>22</td>
<td>100%</td>
<td>2.0</td>
</tr>
<tr>
<td>23</td>
<td>100%</td>
<td>3.0</td>
</tr>
<tr>
<td>24</td>
<td>100%</td>
<td>5.0</td>
</tr>
<tr>
<td>25</td>
<td>100%</td>
<td>10.0</td>
</tr>
<tr>
<td>26</td>
<td>100%</td>
<td>15.0</td>
</tr>
<tr>
<td>27</td>
<td>100%</td>
<td>30.0</td>
</tr>
</tbody>
</table>

The BC03 library (Bruzual and Charlot [2003]) is a theoretical library created from the Chabrier initial mass function (Chabrier [2003]) that provides the distribution of masses of a population of stars. For each mass we know the path that the star takes along the Hertzsprung-Russell diagram, and from templates of stellar spectra with different metallicities we know the possible spectra for each of these stars. The integration of all the spectra of these populations creates the galaxy SED. The dust component is added externally by the LePHARE routines.

By using the BC03 library, we expect to have a set of SEDs that can better represent the photometric points from the 3D-HST photometric catalog. We expect that such a high number of templates will generate degenerate solutions due to the number of free parameters (namely metallicity, star formation timescale, and reddening). Also, as the BC03 is a theoretical library, it may produce systematics due to these SEDs being approximations to what may be observed. The best way to ascertain the accuracy of our estimates is to compare our estimates of the photometric redshift with the available spectroscopic redshift. The number of galaxies with spectroscopic redshift is small, and it was only measured for galaxies at lower redshift. Therefore, in this work, we use the 3D-HST estimates of photometric redshift as a benchmark to ascertain the accuracy of our estimates. If Skelton’s method and ours have an agreement in the estimates, then the estimated values must be more accurate.

The next subsections from 4.2.3 to 4.2.10 present the results of the tests that were performed to ascertain the methodology used in this work.
4.2.3 Test A: SHORT catalog, no minimum error

In the first test, most of the configurations were kept to the default values. The configuration file can be seen in Appendix C.3. The SHORT catalog contains a ID column and then the consecutive columns of magnitudes and associated errors for each filter. The output file is `zphot.out`, is the catalog that contains the values of all the estimates done by LePHARE. After running the test, we use TOPCAT to read the output file. The first plots created were the comparison between LePHAREs and Skelton’s estimates and its residuals.

![Figure 4.6: Comparison between LePHARE and Skelton's redshift estimates.](image)

The residuals are calculated via the expression:

$$R = \frac{z_{\text{LePHARE}} - z_{\text{Skelton}}}{z_{\text{Skelton}} + 1}$$  \hspace{1cm} (4.9)

In this calculation, the difference between both estimates is divided by \((z_{\text{Skelton}} + 1)\) to compensate for the cosmological expansion. In figures 4.6 and 4.7 we can see that in the low redshift bin \((0.6 < z < 1.12)\), there is a good agreement with Skelton’s estimates, the high redshift bin \((2.0 < z < 3.0)\) presents a greater dispersion. The residuals present \(\sigma = 0.123\), where 78% of the sample is below 1\(\sigma\). In the high redshift bin, there are two extreme outliers above 4\(\sigma\) corresponding to the galaxies with the IDs 23837 and 41013.
Figure 4.7: Residual of the comparison between LePHARE and Skeltons estimates of redshift. The horizontal axis shows Skelton’s estimation, the vertical axis shows LePHARE estimation, the gray long dashed line represents no difference and the short dashed line represents the $\sigma = 0.123$ and $\bar{x} = -0.034$. The green dots represent the galaxies that have available spectroscopic redshift.

Figure 4.8: The best fits found for the outlier with the ID 23837. The black squares are the photometric points. The empty circles are the estimated photometric point for the SED. The plot on the lower right corner is the redshift’s PDF (Probability Distribution Function), where the horizontal axis is the estimated redshift, and the vertical axis is the probability.

The figures 4.8 and 4.9 show the plots of the SEDs that present the worst fits. Galaxy 23837 presents two possible solutions (see the inset for their associated parameters). The black curve represents the best SED fit that corresponds to an estimation of $z = 4.76$, while the blue curve represents the secondary SED that compares to an estimate of $z = 3.00$. Both estimates have a similar $\chi^2$ (For the primary solution, $\chi^2 = 53.4$, for the secondary solution...
Figure 4.9: The best fit found for the outlier with the ID 41013. The black squares are the photometric points. The empty circles are the estimated photometric point for the SED. The plot on the lower right corner is the redshift’s PDF (Probability Distribution Function), where the horizontal axis is the estimated redshift, and the vertical axis is the probability.

$\chi^2 = 58.6$). The less probable solution is more compatible with Skelton’s solution. Galaxy 41013 has only one solution that is above the top limit of the higher redshift bin ($2 < z < 3$). The best answer found is an SED that has a low metallicity and low reddening and therefore the fit of the photometric points LePHARE gives a high redshift as output.

The redshift distribution shows 16% galaxies between the low and high redshift bins, and 15% galaxies above the high redshift bin. Being the low redshift bin $0.6 < z < 1.12$, and the high redshift bin $2 < z < 3$. When fitting SEDs there are several components that are key to understand the solution degeneracies. Those are: age, metallicity, and extinction. As the age
increases, the luminosity in the UV and blue region of the SED diminishes due to the decrease in young stars, and in the red region increases due to the old and low mass stars. The extinction in the ISM of the galaxy contributes to a redder SED due to the dust in the ISM absorbing the shorter wavelength emissions and then by re-emitting in the IR. Therefore, as the extinction increases, the blue region of the SED diminishes comparing to the red region. The increase of the metallicity also contributes to a redder SED with the light absorption due to the metals that exist in the stellar atmospheres. This may allow several SEDs, that differ in these three parameters, to fit the photometry at different redshifts. Therefore for a given galaxy color if the model that is being fitted has a higher metallicity, extinction or age then LePHARE will present a lower redshift estimation, in order to compensate for the aforementioned parameters.

![Figure 4.11: The distribution of templates](image1)

The template distribution (in figure 4.11) presents a peak in the templates with the higher metallicity. As the metallicity increases, the SED appears redder, so the redshift estimates lower.

![Figure 4.12: Distribution of reddening values](image2)

The reddening distribution (in figure 4.12) shows that the most used value is the higher value 0.40. The increase in the reddening means that the blue part of the SED is dimmer due to the
dust in the ISM of the galaxy. An accumulation of galaxies with high reddening is an indication that the range of reddening must be enlarged.

Figure 4.13: $\chi^2$ distribution of the best fits.

When looking at the $\chi^2$ distribution (figure 4.13), we can see that there are computed values that are very high. This may be explained by the small errors associated with the photometric points.
4.2.4 Test B: SHORT catalog with a minimum error of 0.1

The previous test showed that some of the SEDs had high $\chi^2$ since some errors from the photometric catalog are too small. In this test, the only change made in the configuration file was to set a minimum error to each photometric data point. This minimum error of 0.1 mag is added in quadrature to the existing errors with the following line added to the zphot.para file. This minimum error gives the fitting code more flexibility to find other solutions.

```
ERR_SCALE 0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1 (This line adds 0.1 in quadrature with the measurements associated error)
```

Figure 4.14: Comparison between LePHARE and Skeltons redshift estimates considering a minimum error of 0.1. The gray dashed line is the one to one relation. The blue X marks represent the estimates that were performed with a minimum error of 0.1, while the red + represent the estimates performed without minimum errors. The plot with the error bars is shown in A.2.

Figure 4.15: Residue of the comparison between LePHARE and Skelton’s estimates of redshift considering a minimum error of 0.1 mag. The horizontal axis shows Skelton’s estimation, the vertical axis shows LePHARE estimation, the gray long dashed line represent no difference and the short dashed line represents the $\sigma = 0.120$ and a $\bar{x} = -0.070$. 

29
By adding the minimum error, LePHARE overestimated Skelton’s solutions. Three galaxies are overestimated above $3\sigma$ as may be seen in figure 4.15. Although galaxies 23837 and 41013 had no change in the redshift, they present fits with lower $\chi^2$ values, due to forcing a minimum error of 0.1 mag. To compensate and maintain the redshift and galaxy colors, LePHARE is showing solutions with a higher metallicity and lower extinction for these two cases. Galaxy 3038 presents a greater difference from Skelton’s estimates. The $\chi^2$ shows a better fit of an SED with lower metallicity and lower extinction which increased the redshift. Below, figures 4.16, 4.18 and 4.20 represent the SEDs produced in this test for the three galaxies with the biggest difference regarding Skelton’s estimates of redshift, while figures 4.17, 4.19 and 4.21 show the SEDs from the previous test to aid in the comparison with the current results.

**Figure 4.16:** The best fits found for the outlier with the ID 23837, with a minimum error of 0.1 mag associated with the photometric points.

**Figure 4.17:** The best fits found for the outlier with the ID 23837, with no minimum error.
Figure 4.18: The best fit found for the outlier with the ID 41013, with a minimum error of 0.1 mag associated with the photometric points.

Figure 4.19: The best fit found for the outlier with the ID 41013, with no minimum error.
Figure 4.20: The best fits found for the outlier with the ID 3048, with a minimum error of 0.1 mag associated with the photometric points.

Figure 4.21: The best fits found for the outlier with the ID 3048, with no minimum error.
The redshift distribution presents a overestimation in both redshift ranges (namely $0.6 < z < 1.12$ and $2.0 < z < 3.0$) in comparison with Skelton’s results. Between the low and high redshift bins we can find 6% of the sample, while 7% are above the high redshift bin. Also, in both redshift bins the peaks are toward the higher values.

The template distribution presents a peak in the templates of high metallicity that correspond to the numbers between 19 and 27. For the meaning of the numbers in the x-axis, please see table 4.2.
The peak in the template distribution continues to show that the higher metallicities have the greater contribution. On the other hand the reddening peak shifts to a lower value.

As expected with the minimum error of 0.1, the photometric points that previously had significant weight in the $\chi^2$ minimization are not driving the solution, and consequently, the distribution shows that the $\chi^2$ display values three orders of magnitude below the previous test.
4.2.5 Test C: LONG catalog using AUTO_ADAPT and APPLY_SYSSHIFT

In this test, we are comparing our estimates with Skelton’s work, and we are using a different library of SEDs. We are using the BC03 library mentioned previously in section 4.2.2, as these SEDs are approximations to our photometry, there may be systematics. Therefore, we ran this test to see if these features would help improve the accuracy of our estimates.

The AUTO_ADAPT and APPLY_SYSSHIFT are the two features used to calculate and correct the systematics that may exist between the templates and the photometric points. As mentioned previously in section 4.1.1, the AUTO_ADAPT feature needs a catalog of the type LONG that contains two extra columns, the context, and spectroscopic redshift columns. When running, LePHARE uses the galaxies that have spectroscopic redshift available to find SED solutions that fit the photometric points at that fixed spectroscopic redshift. With the best SEDs found, it calculates the offsets that may be added to each of the filters of the photometric catalog to compensate the systematics that may be detected. Afterward, the AUTO_ADAPT function is disabled, and the offset values are inserted into the variable APPLY_SYSSHIFT, the zphoto routine is rerun, and applies these changes to all galaxies.

The photometric catalog for the AEGIS field contains only nine galaxies with spectroscopic redshift for the redshift range $0.6 < z < 1.12$. Figures 4.26 and 4.27 show the plots of the comparison between LePHARE and Skelton’s photometric redshift estimates and the spectroscopic redshift of the 9 galaxies of the AEGIS field with the AUTO_ADAPT.

![Figure 4.26: The comparison between LePHARE and Skelton’s photometric redshift estimates versus the spectroscopic redshift.](image1)

![Figure 4.27: The residuals of the comparison between LePHARE and Skelton’s photometric redshift estimates versus the spectroscopic redshift.](image2)
Figure 4.27 shows that LePHARE’s estimates have a $\sigma$ similar to Skelton’s. Our estimates and Skelton’s show that 7 of the galaxies are below 1$\sigma$ which means that there is an agreement between our photometric and the spectroscopic redshift estimates for 78% of the sources.

For the galaxies that do not have a spectroscopic redshift, LePHARE computes the photometric redshift taking into account the calculated offsets. Figures 4.28 and 4.29 show that, for this test, the high metallicity templates were the most used and the reddening peak is the highest value allowed this indicates that the solutions were being pushed to the lower redshifts. Hence the peak in the redshift distribution for LePHARE’s estimates between redshift 1.0 and 2.0 on the plot below (figure 4.30).
The first part of the test provided a list of offsets to be applied to each independent filter, via the `APPLY_SYSSHIFT` features. In the second part of this test, the `AUTO_ADAPT` was disabled and the offsets that had been calculated in the first part were written into the configuration variable `APPLY_SYSSHIFT`. Afterward we rerun LePHARE. The plots that follow display the photometric redshift estimate comparison and the residuals (figures 4.31 and 4.32).

The LePHAREs estimates show almost all the galaxies are below $3\sigma$ of Skelton’s estimates with the exception of three galaxies that are above this value.
Figure 4.32: Comparison of Photometric redshift with the manual input of the offsets.

Figure 4.33: Template distribution with manual offsets.
The plots of the template distribution (figure 4.33) and the reddening (figure 4.34) show that the SEDs being fit are mostly high metallicity and also with a considerable peak at the maximum value of reddening. This is an indication that the reddening interval may need to be increased.

Although this method may increase the accuracy of the estimates by compensating for the systematics found between the photometry and the SED models, the comparison of the $\sigma$ of the residuals of the previous test ($\sigma = 0.120$) and the current test ($\sigma = 0.119$) shows a similar dispersion of points. Also the sample of galaxies of the AEGIS field contains only 9 galaxies that have spectroscopic redshift. Therefore each galaxy will carry much weight to the determination of the offsets. It is not advisable to use this kind of method for a such small number of galaxies with spectroscopic redshift.
4.2.6 Test D: LONG catalog with upper limits

The upper limit is a magnitude limit from which we may consider a detection. When a band has an upper limit, it means that there must be a signal in that band that is too faint to be detected. Therefore, all SED solutions with a curve that pass above that point are discarded. This constrains the SED solutions and therefore eliminates degeneracies. In the previous tests, every catalog entry that had negative fluxes were set to -99 which means that there was no data for that filter in that particular galaxy. This allowed solutions with SEDs that could be fitted passing above every point and therefore allowing degenerate and unwanted solutions. Some of the outliers show more than one solution, and the use of upper limits could help determine the more accurate solution. Some upper limits must be present in the catalog, but the documentation of the 3D-HST survey does not describe how the upper limits are written in the photometric catalog. Still, there was information about the average $5\sigma$ depth which shows the threshold magnitude to which a signal can be detected. Therefore in this test, all the catalog entries with a negative flux were marked as an upper limit with the value of the $5\sigma$ depth for that particular band.

Figure 4.35 presents the comparison between Skelton’s and LePHARE’s estimates in the test of section 4.2.4 of the current test. Figure 4.36 shows the residuals of the current test.

Figure 4.35: Comparison between LePHARE and Skelton’s estimates of redshift. The horizontal axis shows Skelton’s estimation, and the vertical axis shows LePHARE estimation, the gray dashed line is the one to one relation. The blue X marks represent the estimates that were performed without upper limits, while the red + represent the estimates performed with the fixed upper limits published with the catalog. The plot with the error bars is presented in A.3.

The galaxies that present the most significant differences from the previous test are 9870 and 13796, while 33925 and 11986 had now results that are more compatible Skelton.
Galaxies 23837 and 41013 are above $3\sigma$ and are the most extreme cases regarding the difference in the estimation of redshift.

The current solution for galaxy 23837, presents the same values of metallicity, and reddening, as in the second test (section 4.2.4, figure 4.16). The PDF plot shows that the lower redshift solutions were suppressed, leaving only the high redshift solution ($z \approx 4.759$). The elimination of the lower redshift solutions may be due to the upper limit in filter 814w ($\lambda = 804\,nm$) not allowing the fit of SEDs of lower redshifts.
Galaxy 41013 (figure 4.38), shows no change in redshift, metallicity or extinction, it also presents a single solution at a higher redshift incompatible with Skelton’s estimation. In this case, the upper limit in the R ($\lambda = 640nm$) filter may be constraining the solutions SEDs that correspond to a lower redshift. Therefore the solution found has a high redshift value.
4.2.7 Test E: LONG catalog with $5\sigma$ upper limits

In this test, we change the definition of the upper limits. The associated errors to the measurements of each filter are the standard deviation of the background. In this test, we continue to consider the catalog entries that have negative fluxes as upper limits and their value is calculated as five times the associated error. Figure 4.39 displays the comparison between LePHARE’s and Skelton’s estimates of the current and previous tests.

![Figure 4.39: Comparison between LePHARE and Skelton’s estimates of redshift. The horizontal axis shows Skelton’s estimates, and the vertical axis shows LePHARE estimation, the gray dashed line is the one to one relation. The blue X marks represent the estimates that were performed with the $5\sigma$ upper limits, while the red + represent the estimates performed with the fixed upper limits published by Skelton. The plot with the error bars is shown in A.4](image)

Figure 4.40: Residue of the comparison between LePHARE and Skelton’s estimates of redshift considering a minimum error of 0.1. The horizontal axis shows Skelton’s estimates, the vertical axis shows LePHARE estimation, the gray long dashed line represents zero, and the short dashed line represents the $\sigma = 0.121$ and a $\bar{x} = -0.075$.

Galaxies 9870 and 13796 two of the galaxies that presented the large differences between LePHARE’s and Skelton’s estimates.

43
Below, in figures 4.41 and 4.42 we present the SEDs of galaxy 9870 from the previous and current tests.

**Figure 4.41:** The SED of galaxy 9870, with fixed upper limits.

**Figure 4.42:** The SED of galaxy 9870, with the 5σ upper limits.

In the previous test (figure 4.41), galaxy 9870 presented a high redshift solution \( z \approx 3.89 \). The upper limit in the photometric point of the J1 filter \( \lambda = 1047 \text{nm} \) was constraining the solutions with magnitudes below 24.4. This point was almost two magnitudes above the neighbor photometric points, and this resulted in the elimination of solutions that better fitted the photometric points of the shorter wavelengths, this way resulting in a high redshift solution with a higher \( \chi^2 \) value \( \chi^2 = 595.2 \). In the current test, the upper limits are computed differently, resulting in a smaller difference between the J1 photometric point and the neighbor points. Thus a better solution is found, corresponding to a lower redshift solution \( z \approx 3.30 \). Even though the solution is better than the previous the, it has a \( \chi^2 = 149.3 \) due to the distance between the points and the fit at the lower wavelengths.
The changes observed for the SED estimation of galaxy 13796 have a similar explanation. In the previous test, the upper limit of filter 606w ($\lambda = 592.5 \text{nm}$) presents a magnitude of 27.3 (figure 4.43), while the other photometric points have a difference of about two magnitudes. In the case of this test (4.44), this upper limit has a lower magnitude that allows solutions of lower redshift still with a high $\chi^2$ value due to having the upper limit displaced from the other photometric points.
The redshift distribution is similar to the previous test.
The template and reddening distributions are also similar to the previous test.

The $\chi^2$ for this test has a similar distribution, with the difference of the maximum being reduced from 600 to 250.
4.2.8 Test F: LONG catalog with 5σ upper limits and E(B-V) maximum increased to 1.5

In the previous tests we used with a small interval of reddening (from 0 to 0.4 in steps of 0.1), and the reddening distribution plots showed a peak in the higher value, this meant that there could be other solutions at higher values of reddening. For a galaxy of a given color, if the reddening increases the shorter wavelengths are more attenuated, and the galaxy will appear more reddish to the observer. Therefore, by allowing a broader interval of reddening, it allows LePHARE to explore solutions at lower redshifts, this may show lower redshift solutions for the galaxies that show a higher difference comparing to Skelton’s estimates.

In the 3D-HST estimates catalog, the extinction interval is $0 < A_v < 4$, and it was necessary to increase the reddening range to approach Skelton’s estimates. The reddening and the extinction are related according to the formula:

$$E(B - V) = \frac{A_v}{R_v} \quad (4.10)$$

Where $A_v$ is the extinction, $E(B - V)$ is the reddening and $R_v$ is a correlation constant that for the Fitzpatrick extinction law (Fitzpatrick [1998]) has a value of 3.1. With equation 4.10 the maximum value was calculated to be $E(B - V) \approx 1.3$, and thus we increased the maximum value to 1.5. After running LePHARE with the same upper limits as the previous test and with the new reddening interval, the results are as follows.

![Figure 4.49: Comparison between LePHARE and Skelton’s estimates of redshift. The horizontal axis shows Skelton’s estimation, the vertical axis shows LePHARE estimation, the gray dashed line is the one to one relation. The red X marks represent the estimates that were preformed with the max(E(B-V))=0.4, while the blue + represent the estimates preformed with the max(E(B-V))=1.5. The plot with the error bars is shown in A.5.](attachment:Figure_4.49.png)

The increase of the reddening interval allows solutions of bluer SEDs to be fitted that leads to lower redshift solutions. This may explain the shift from overestimated to underestimated solutions (in comparison with Skelton’s) in the high redshift bin ($2 < z < 3$). There is still one outlier in the high redshift bin (Galaxy 41013). On the low redshift bin, there are four galaxies that increased appreciably their values, which are galaxies 13572, 27365, 30084 and 31001.
The SED of figure 4.50 shows the solutions found for the outlier galaxy 41013. Previously, this galaxy showed, in the PDF plot, the three possible solutions but with a very low probability. In this test, presents three solutions with a high probability. It is due to the increase of the reddening interval. The highest peak of the distribution relates to an SED from model 10, and no extinction, while the second solution relates to model 19 and a reddening of 1.2. As can be seen in table 4.2, this means that the solution with $z = 4.88$ corresponds to an SED with a metallicity of $0.4Z_{\odot}$ with no reddening, while the secondary solution corresponds to an SED with solar metallicity $1.0Z_{\odot}$ and a higher value of reddening. Unfortunately, LePHARE does not provide information for the third probability peak (the one at $z \approx 3$). With three redshift peaks the redshift of this galaxy is very uncertain. Below we present the SEDs of galaxies 13572, 27365, 30084, and 31001.

The SED of galaxy 41013, with the $\text{max}(E(B-V)) = 1.5$. (Figure 4.50)

The SED of galaxy 13572, with the $\text{max}(E(B-V)) = 0.4$. Model 23, E(B-V)=0.40, $z = 1.07$. (Figure 4.51)
Figure 4.52: The SED of galaxy 13572, with the $\text{max}(E(B-V)) = 1.5$. Model 11, E(B-V)=0.90, $z = 1.55$

Figure 4.53: The SED of galaxy 27365, with the $\text{max}(E(B-V)) = 0.4$. Model 27, E(B-V)=0.30, $z = 1.09$
Figure 4.54: The SED of galaxy 27365, with the $\max(E(B-V)) = 1.5$. Model 19, $E(B-V)=0.70$, $z = 1.55$

Figure 4.55: The SED of galaxy 30084, with the $\max(E(B-V)) = 0.4$. Model 23, $E(B-V)=0.30$, $z = 1.04$
Figure 4.56: The SED of galaxy 30084, with the \( \text{max}(E(B-V)) = 1.5 \). Model 18, E(B-V)=0.70, z = 1.55

Figure 4.57: The SED of galaxy 31001, with the \( \text{max}(E(B-V)) = 0.4 \). Model 23, E(B-V)=0.30, z = 1.04
The low redshift galaxies presented above do not have any upper limits that may constrain the solutions. All the solutions found depend only of the SED characteristics (namely, metallicity, reddening, and redshift). The increase in the estimation of redshift for these galaxies may be due to the increase of the reddening interval. In the previous test (Figures 4.51, 4.53, 4.55, and 4.57) the solutions that were found for these galaxies had high metallicity and the highest reddening value allowed \( E(B-V) \approx 0.4 \). In the current test (Figures 4.52, 4.54, 4.56, and 4.58), the solutions found have a medium metallicity and reddening values \( E(B-V) \approx 0.7 \). In these cases the use of lower metallicities allowed higher redshift estimates.
Figure 4.60: The distribution of templates for the two separate intervals and Skelton's estimation. The red bars present the estimates when considering $0 < E(B - V) < 1.5$, the blue bars represent the distribution when considering $0 < E(B - V) < 0.4$ and the green outlined bars represent Skelton's estimates.

The template distribution is similar. Even though with the higher reddening interval the use of top metallicity templates decreases (templates from 19 to 27). The increase of the reddening values allows other solutions to be fit. To keep the galaxy colors the increase in reddening may compensate for lower metallicities. Therefore explaining why in this test the solutions found are using lower metallicity models and higher reddening values.

Figure 4.61: The distribution of $E(B-V)$ values for the two separate intervals and Skelton's estimation. The red bars present the estimates when considering $0 < E(B - V) < 1.5$, the blue bars represent the distribution when considering $0 < E(B - V) < 0.4$ and the green outlined bars represent Skelton's estimates.

The use of higher reddening allows the fit of bluer SEDs, which may explain the shift in the redshift estimates. Seemingly, the reddening values are more correct as we do not have an accumulation of galaxies with very high values as displayed in figure 4.48.
4.2.9 Test G: LONG catalog, with max(E(B-V))=1.5, considering a $S/N < 5$ as an upper limit

In this test, the upper limits are detected according to the signal-to-noise ratio ($S/N$). The signal-to-noise ratio is a comparison between the level of a signal and the background noise. For this test, all the correlations between flux and the error that were below a ratio of 5 are considered as upper limits due to their low significance. Assuming that the error follows a normal distribution, then by setting the upper limit to 5, the probability of a signal with a $S/N > 5$ being part of the noise is very unlikely. When the script finds the ratio between the flux and the error to be below 5, then it considers it an upper limit and writes its value as five times the error. With this method, the number of detected upper limits increased, and most of them are located in the region of the UV and bluer part of the spectrum. For galaxies, as the redshift increases, the spectrum of the galaxy is shifted to the IR, leaving the bluer filters with faint signal or no signal. By setting upper limits, we may have a better constraint of the SED solutions that are found.

Below is the comparison between LePHARE and Skelton’s estimates with the previous upper limits used only in the negative fluxes (section 4.2.8) and with the $S/N < 5$ upper limits.

![Figure 4.62: Comparison between LePHARE and Skelton’s estimates of redshift. The horizontal axis shows Skelton’s estimation, and the vertical axis shows LePHARE estimation, the gray dashed line is the one to one relation. The red + marks represent the estimates that were performed with the $S/N < 5$ upper limits, while the blue X represent the estimates performed considering the negative fluxes as upper limits. The plot with the error bars is shown in A.6](image)

The changes in redshift regarding Skelton’s estimates are small in most of the galaxies. At low redshift ($0.6 < z < 1.12$), we continue to have galaxies 13572, 27365, 30084 and 31001 to display a greater difference from Skelton’s estimates, and no difference from the previous test. As this four galaxies do not have any upper limits, this is expected. Galaxy 41013, which continues to show the three solutions in the PDF, the probability of these solutions changed so that the lower redshift solution became the most probable, as can be seen below in the figure 4.64.
Figure 4.63: Residuals of the comparison between LePHARE and Skelton’s estimates of redshift. The horizontal axis shows Skelton’s estimation, and the vertical axis shows LePHARE estimation, the gray dashed line is the one to one relation. The red + marks represent the estimates that were performed with the S/N < 5 upper limits, while the blue X represent the estimates performed considering the negative fluxes as upper limits.

Figure 4.64: The SED of the galaxy 41013, with the S/N < 5 upper limits.

The two most probable solutions SEDs for this galaxy present a similar reduced $\chi^2$. The first solution uses a template of the higher metallicity and with high reddening, and therefore a lower redshift ($z \approx 2.09$). The secondary solution uses a lower metallicity template with a high reddening which leads to a higher redshift ($z \approx 3.20$). Skelton’s estimates $z \approx 2.829$. For this test, the standard deviation of the residuals is $\sigma = 0.129$. The first solution presents a redshift residual of 0.19, that is considered an outlier, while the second solution shows a redshift residual of 0.097. Thus the second solution is the most compatible with Skelton’s estimates.
4.2.10 Test H: LONG catalog, $S/N < 5$ upper limits and no IRAC bands 3 and 4

As the wavelength increases, so does the image’s Point Spread Function. In fields that have a high density of objects, fainter galaxies may contribute to the sky measurement and also to the flux of the galaxy that is being measured. Also, the BC03 library does not provide a good model beyond $\lambda = 2.0\mu m$. This led us to remove the IRAC bands 3 ($\lambda = 5759nm$) and 4 ($\lambda = 7959nm$) to see if it would improve the estimates.

![Figure 4.65](image-url)  
Figure 4.65: Comparison between LePHARE and Skelton’s estimates of redshift. The horizontal axis shows Skelton’s estimation, and the vertical axis shows LePHARE estimation, the gray dashed line is the one to one relation. The red + marks represent the estimates that were performed without IRAC 3 and four bands, while the blue X represent the estimates performed with all the photometric bands. The plot with the error bars is shown in A.7

![Figure 4.66](image-url)  
Figure 4.66: The residual of the comparison between LePHARE and Skelton’s estimates of redshift. The horizontal axis shows Skelton’s estimation, and the vertical axis shows the residual. The red + marks represent the estimates that were performed without IRAC 3 and four bands with the respective $1\sigma$ represented as a red dashed line, while the blue X represent the estimates performed with all the photometric bands, considering a reddening interval of $0 < E(B-V) < 1.5$ and upper limits detected from a $S/N < 5$, with the corresponding $1\sigma$ represented as a blue dashed line.

In the current test we may see from the residuals (figure 4.66), that in the sample of 69 galaxies, the results reveal that 55 are below $1\sigma$, which is considered an agreement between our estimates and Skelton’s estimates. Between $1\sigma$ and $2\sigma$ we have 9 galaxies, and above $2\sigma$ there are 5 galaxies. The greatest difference in the estimates comes from galaxy 30084, that is a low redshift galaxy. That in this test presents a small decrease in redshift. As the weight of the IRAC 3 and 4 was removed, it allowed the SED to be displaced toward the shorter wavelengths. The objective of the tests conducted in this work, was to find the best methodology, to produce the
most accurate estimates of redshift and mass. The results from this last test present a better agreement with Skelton’s estimates. Therefore, these are the results used for the mass-size relation.

5 Results

Initially we aimed to include all of the CANDELS fields in this work, to have a great number of galaxies, in order to obtain a better estimation of the mass-size relation. However, there was not enough time to include all the fields in this work. Therefore, we only present our estimates of redshift and mass for the AEGIS field.

The results are organized into two subsections. First (in section 5.1) we present the filtered results from the estimates of the catalogs of CANDELS and 3D-HST. This was done as an exercise to preview what we could find regarding the mass-size relation. Second (section 5.2) we present the comparison between our redshift estimates and Skelton’s, followed by our estimates of redshift, mass, and other physical parameters. At the end of this section we present the mass-size relation, resorting to our estimates of mass and redshift, and the size and Sérsic index estimates from the CANDELS catalog. In this plot we have separated the galaxies by redshift and Sérsic index, to observe the effect of the morphology and redshift in the mass-size relation.

5.1 Results based on catalog data

Here we will present the work we carried out in the beginning of this project in order to understand better the nature of massive galaxies. Results taken from the estimates of the 3D-HST and CANDELS catalogs will also be presented in order to observe the mass-size relation and the Sérsic index distribution. The mass-size relation provides a view of how the massive galaxies evolved in mass from the high redshift to the low redshift, while the Sérsic index distribution may give an indication of possible changes in the morphology.

5.1.1 The mass-size relation

We separate the massive galaxies in two redshift bins 0.6 < \( z < 1.12 \) and \( 2 < z < 3 \), to observe the mass-size relation at different epochs, and we plotted the results of the separate fields. Afterward, we take into account the galaxies from all the fields. For these plots, the x-axis corresponds to the logarithm of the stellar mass of the galaxy, the y-axis corresponds to the circularized effective radius. The grey dashed line is the mass-size relation \( (R = 2.88 \times 10^{-6} \times M_\star^{0.56}) \) for the spheroid-like galaxies \( (n > 2.5) \), in the Local Universe \( (z=0) \). The black dashed line is the mass-size relation \( (R = 0.1 \times (M_\star^{0.14}) \times ((1 + (((1/3.98 \times 10^{10}) \times M_\star)^{0.25})) \) for the disk-like galaxies in the Local Universe \( (z = 0) \) (Shen et al. 2003). The blue and red dots and their respective lines correspond to the galaxies of the low redshift bin \( (0.6 < z < 1.12) \), the high redshift bin \( (2 < z < 3) \) and the respective regression lines.
In the AEGIS field, there are 69 massive galaxies with disk components, of which 30 are in the low $z$ bin, and 39 are in the high $z$ bin.

In the COSMOS field, there are 52 massive galaxies with disk components, of which 28 are in the low $z$ bin, and 24 are in the high $z$ bin.
In the UDS field, there are 73 massive galaxies with disk components, of which 18 are in the low z bin, and 55 are in the high z bin.

In the GOODS-N field, there are 61 massive galaxies with disk components, of which 24 are in the low z bin, and 37 are in the high z bin.
In the GOODS-S field, there are 31 massive galaxies with disk components, from which 8 are in the low z bin and 23 are in the high z bin.

In all the fields together there are 286 massive galaxies with disk components, of which 108 are in the low z bin, 29 have $n < 2.5$ and 79 have $n > 2.5$. The high z bin has 178 galaxies, where 128 have $n < 2.5$ and 51 have $n > 2.5$. The separate fields contain a small number of galaxies. The scatter of points is significant, and there are not enough points to show any relation between mass and size. Also, a small group of galaxies present effective radius ($R_{e_{circ}} > 10.0 \text{ kpc}$), these measurements may be wrong. Only when all the fields together are taken into account, we can observe that for a given mass, the massive galaxies at high redshift show smaller effective radius than the massive galaxies at lower redshift (figure 5.6). From figure 5.7 we present the sample separated by redshift, and by Sérésic index interval. The inclination of the regression lines of the two Sérésic index intervals change in the two redshift bins. At high redshift the galaxies that are
more spheroid-like \((n > 2.5)\) show a steeper growth in size with the increase of mass, than the disk-like galaxies. On the other hand, at low redshift the disk-like galaxies \((n < 2.5)\) display the steeper growth than the spheroid-like. Comparing these trends with the mass-size relation for the Local Universe (displayed with the dashed grey and black lines), the mass-size relation of the disk-likes and spheroid-like galaxies at low redshift, are inverted regarding the Local Universe mass-size relation. These results may be explained by having a small number of galaxies in the sample, this gives a big statistical weight for each galaxy, and allows the outliers to have greater influence on the parameters of the regression line.

5.1.2 The Sérsic index distribution

The distribution of the values of the Sérsic index with redshift provides information about how the components of disk and bulge change with redshift and therefore with time, as it is shown below in figure 5.8. Sérsic index is a means to differentiate galaxy morphologies at high redshift when visual information becomes dubious. Usually, \(n < 2.5\) galaxies are regarded as disk-like objects, whereas \(n > 2.5\) are disk-like objects (Ravindranath et al. 2004, Buitrago et al. 2013).
This plot shows that there is a difference in the Sérsic index with redshift. At high redshift (2 < z < 3) the distribution displays two peaks between 0.5 and 1.5. This indicates that galaxies were more disk-like and therefore, the disk component dominated the luminosity. For the low redshift (0.6 < z < 1.12), the peak is at n ≈ 3 corresponding to a predominance of galaxies that are more spheroid-like whose bulge component is more dominant. Thus indicating that the components that dominate the luminosity of the galaxies change with redshift and therefore with time. Summarizing, massive galaxy undergo a transformation with redshift, moving from compact disks to large spheroids.

5.2 Our estimates

Below we present redshift and stellar population estimates for the massive galaxies within the AEGIS field. The configurations used are the same of section 4.2.10. From all the tests, these configurations provided the most compatible results with Skelton. Here we used all the templates of the BC03 library (Table 4.2). We also used the Fitzpatrick extinction law (Fitzpatrick 1998) with a reddening range between 0 and 1.5. In the photometric catalog we used all the filters, except the IRAC3 (λ = 5759 nm) and IRAC4 (λ = 7959 nm).

5.2.1 Comparison of redshift

For this work we used Skelton’s results as a benchmark. The plots that follow present the comparison between Skelton’s and LePHARE’s estimates on the AEGIS field only.

Figure 5.9: Comparison between Skelton’s and LePHARE’s estimates. The gray dashed line represents z_skelton = z_lephare.

Figures 5.9 and 5.10 show that 51 of the 69 galaxies have the residuals of the comparison with Skelton below 1σ, and thus shows agreement with Skelton’s estimates. From which 27 are in the low redshift bin (0.6 < z < 1.12) and 24 in the high redshift bin (2 < z < 3). The comparison shows a higher agreement with Skelton’s estimates in the low redshift bin rather than in the high redshift bin. One important reason for that is that the IR bands are more spaced in wavelength than the blue ones. Therefore this leads to higher uncertainties in the location of features of the spectrum, like the Balmer break, that constrain the redshift associated with the SED.
5.2.2 The mass-size relation using our own determinations

In this section we will first display the parameters that characterize the SEDs that were fitted with the photometry of the AEGIS field. The template distribution (figure 5.11) provides information of the metallicity and star formation timescales of the templates that were used to fit the photometry. The $E(B-V)$ distribution (figure 5.12) displays information about the attenuation by the dust in the ISM of the galaxy, therefore the amount of dust that the galaxy contains. Afterward, we present the parameters derived from the SEDs: redshift, estimated age and mass (respectively in figures 5.13, 5.14 and 5.15). For the mass-size relation, we estimated the mass and relied on the effective radius estimates from the CANDELS catalog. Due to the distribution of redshift from our estimated in this section we will use the intervals $0 < z < 1.12$ as the low redshift interval and $z > 1.12$ as the high redshift interval. Only in the comparisons with Skelton we use the intervals of $0.6 < z < 1.12$ and $2 < z < 3$.

The templates distribution (figure 5.11) shows that most of the solutions were found with the higher metallicity templates in either, the higher and the lower redshift intervals. It also
shows that for the high metallicity templates, the high redshift solutions have a peak toward the shorter star formation timescales (with the peak at $\tau = 0.3\, \text{Gyr}$), while for the low redshift the chosen values are higher (peak at $\tau = 2.0\, \text{Gyr}$). This may indicate that there may be a relation between the cosmic epoch and the star formation timescale. The star formation may have been faster at earlier epochs. This process is known as downsizing (e.g. Fontanot et al. 2007).

Figure 5.12: The reddening distribution of our estimates.

Figure 5.12 shows the distributions of reddening values that were used. For the lower redshift range ($0 < z < 1.12$), the peak in the reddening is $E(B - V) \approx 0.3$. While for the higher redshift range it is a higher value $E(B - V) \approx 0.7$. The reddening is related to the amount of dust that exists in the ISM. The amount of dust should increase with time due to the accumulation of metals in the successive generations of stars. Our preview is showing the opposite. This may be due to the discrete values used in the estimates. The intervals used for the reddening were large due to the computational limitations. This may have forced LePHARE to find solutions with different values of metallicity and age to compensate for the reddening intervals.

Figure 5.13: The redshift distribution of our estimates.

The redshift distribution shows two redshift ranges. At lower redshift ($0 < z < 1.2$) the values are more compacted, while it presents a wider spread of values at the higher redshift ($z > 1.4$). The dispersion at the high redshifts may be explained by the separation between the
filters in the longer wavelengths in the IR. This diminishes the accuracy in the position of the SED regarding the wavelength.

![Figure 5.14: The age distribution of our estimates.](image)

The age distribution plotted in figure 5.14 displays a mean age of 3.38 Gyr for the low redshift bin, while the mean age for the high redshift bin is 1.24 Gyr. Thus indicating that most of the massive galaxies in the sample formed at an early epoch.

![Figure 5.15: The mass distribution of our estimates.](image)

The distribution of mass for the low redshift bin displays a mean value of \( \log(M_\ast/M_\odot) = 11.08 \). While for the high redshift bin \( \log(M_\ast/M_\odot) = 10.995 \). Thus this may indicate that the galaxies may gain some stellar mass as time goes by.
Figure 5.16: The comparison of the mass estimates. The red and blue dots correspond to two redshift bins. The dashed black line represents the agreement line. The $\times$ marks the galaxies that have a redshift residual below 0.1, the + marks the galaxies that have a difference in the $\log(M_*/M_\odot)$ below 0.1.

The plot in figure 5.16 shows the comparison between Skelton’s and our estimates for the mass of the galaxies. We calculated the residuals for both redshift and mass comparisons, and we are considering all the estimates which have a residual below 0.1 as agreeing. The galaxies that are in agreement regarding redshift are marked with a $\times$ and the galaxies that are in agreement regarding mass, are marked with a +. In our sample we have 51 galaxies that are in agreement regarding only redshift, 24 in the low redshift bin ($z < 1.12$) and 27 in the high redshift bin ($z > 1.12$). There are 36 in agreement regarding only mass, 16 in the low redshift bin and 20 in the high redshift bin. In agreement with both redshift and mass, we have 33 galaxies, 15 in the low redshift bin and 18 in the high redshift bin. The galaxies that have both agreements with the redshift and mass estimates must be bonafide massive galaxies.

The galaxies that have both agreements with the redshift and mass estimates must be bonafide massive galaxies.

Figure 5.17: Mass-size relation, with estimates that have $\sigma < 0.1$. The green and blue dots correspond to two redshift bins. The dashed black line represents the agreement line. The + marks the galaxies that have a redshift residual below 0.1, the $\times$ marks the galaxies that have a difference in the $\log(M_*/M_\odot)$ below 0.1.
Figure 5.17 presents the mass-size plot using our estimates of mass and the effective radius calculated from the CANDELS catalog. To visualize where the galaxies that have an agreement in mass and redshift are located, we have added a mark. In figure 5.18 we present the mass-size plot only with the bonafide massive galaxies. As the scatter in redshift was more significant for the high redshift galaxies, we are only considering two redshift bins. The low redshift bin is defined as $z < 1.12$ and the high redshift bin as $z > 1.12$. Thus the separation in the plot below.

![Mass-size plot](image)

Figure 5.18: Our estimation of the mass-redshift relation with the galaxies that are in agreement with Skelton’s estimates regarding redshift and mass. The blue dots represent the galaxies of the low redshift bin with the associated regression line ($\log(y) = 20.514 \log(x) - 20.690$), with $R = 0.54$. The green dots represent the galaxies in the high redshift bin and the respective regression line ($\log(y) = 32.914 \log(x) - 33.870$), with $R = 0.57$. The grey dashed line is the mass-size relation ($R = 2.88 \times 10^{-6} \times M_0^{0.56}$) for the spheroid-like galaxies ($n > 2.5$), in the Local Universe ($z=0$). The black dashed line is the mass-size relation ($R = 0.1 \times (M_0^{14}) \times ((1 + ((1/3.98 \times 10^{10}) \times M_*)^{0.25})$), for the disk-like galaxies in the Local Universe ($z = 0$).

Although this population is small, the regression lines for both redshift bins show that for a given stellar mass, the galaxies at higher redshift were more compact than the ones we find in the Local Universe.
6 Conclusions

After selecting the massive galaxies from the 3D-HST and CANDELS catalogs, we could not observe the mass-size relation of the massive disk galaxies in independent fields. This may be due to the small number of galaxies per field and the dispersion being too large. When all the fields were combined, the plot showed that for a given mass, the galaxies presented a smaller effective radius at higher redshift when compared to the low redshift, like it may be seen in the work of Wel et al. 2014. The Sérsic index distribution may show an evolution of the morphology of the galaxies with redshift and therefore time. We can see in figure 5.8 that in the low redshift bin (0.6 < z < 1.12) the mean value of the Sérsic Index is around n ≈ 3 which means that the surface brightness profile is closer to a spheroid-like galaxy. The high redshift bin (2 < z < 3) presents a mean value of the Sérsic Index as n ≈ 1.8 which represent a surface brightness profile closer to a disk-like galaxy. This may indicate that the galaxies at an earlier cosmic epoch were more disk-like then they are in the Local Universe.

The tests that were performed to determine the methodology showed that the use of upper limits is an important feature to reduce the degeneracy of solutions. By considering upper limits that are S/N < 5 and by removing the IRAC bands 3 and 4, the accuracy of the solutions regarding the first test were considerably improved. Our estimates have a reduced number of galaxies in the AEGIS field, there are only 69 galaxies that may be considered massive according to our criteria (M_* > 10^{11} M_\odot). From our estimates we observed that the distribution of templates shows a peak at the high metallicity in both redshift intervals. The high values of reddening are being used in galaxies at high redshift, while at low redshift the values are lower. It would be expected to observe the higher reddening values for galaxies at low redshift instead of high redshift. As it was mentioned in section 2.5 the reddening depends on the existence of dust in the ISM, produced by several stellar mechanisms. Therefore, it is expected that the amount of dust increases with the passing of different generations of stars. This result may be due to the step between values of reddening being too large. The redshift distribution shows a considerable number of galaxies between the redshift bins used to select the galaxies from Skelton’s estimates. The age estimation distribution displays a predominance of high ages values for the galaxies at lower redshift and lower ages for the galaxies at high redshift. The distribution of masses for the low redshift bin shows a wider range of masses and a peak around 10^{11.2} M_*/M_\odot, while for the high redshift bin the range of masses is narrower and the peak is at a lower mass 10^{11} M_*/M_\odot which is consistent with the galaxies growth with redshift.

In the last plots (figures 5.16, 5.17 and 5.18) we present the comparison between Skelton’s and our mass estimates as a benchmark and to see which galaxies show an agreement between the two methods. Afterward, we display the mass-size plot with the 69 galaxies from our original sample, and we mark the galaxies that show an agreement in mass and redshift. Finally, we plot the mass-size, just with this last subset of galaxies. With this last plot, we can observe that the galaxies at higher redshift are more compact than the lower redshift galaxies.

The sample seems to be small enough for each of the galaxies to carry much statistical weight, and this way one single bad fit may change the features that the population shows. To improve the accuracy of the estimates and probably see other features of the population of massive galaxies, we need to include other fields to have a large enough sample to produce better statistics.
7 Future work

As future work, now that we have a methodology established through the work done in the AEGIS field, we need to extend the estimates for the remaining four CANDELS fields using the galaxies that were considered massive. Afterward, we will need to run LePHARE on the entire photometric catalog to extract the massive galaxies from our estimates to compare with Skelton’s estimates. This comparison of results between the two different methods could provide us with the galaxies that have better estimates. We would also need to make our estimates of the galaxy structural parameters with SExtractor and some other specific programs to fit Sérsic functions to the galaxies’ surface brightness profiles. As the 3D-HST and CANDELS catalogs do not have much information about the Local Universe, the photometry of the ALHAMBRA survey should be included in our estimates in order to be able to observe with more detail the evolution of the parameters that characterize a galaxy.
References


Arnouts, Stéphane and Olivier Ilbert (2006). “Le PHARE PHotometric Analysis for Redshift Estimations”. In:


Buitrago, Fernando et al. (2013). “Early-type galaxies have been the predominant morphological class for massive galaxies since only z ~ 1”. In: Monthly Notices of the Royal Astronomical Society. issn: 00358711. doi: 10.1093/mnras/sts124.


Skelton, Rosalind E et al. (2014). “Submitted to the astrophysical journal supplement series 3D-HST WFC3-selected photometric catalogs in the five CANDELS/3D-HST fields: Photometry, photometric redshifts and stellar masses”. In:


Van Der Wel, A et al. (2013). “STRUCTURAL PARAMETERS OF GALAXIES IN CANDELS”. In:

A Plots with error bars

A.1 Test A: SHORT catalog, no minimum error

Figure A.1: Comparison between LePHARE and Skeltons redshift estimates. The horizontal axis shows Skeltons estimation, the vertical axis shows LePHARE estimation, the gray dashed line is the one to one relation.

A.2 Test B: SHORT catalog with a minimum error of 0.1

Figure A.2: Comparison between LePHARE and Skeltons redshift estimates considering a minimum error of 0.1. The gray dashed line is the one to one relation. The blue X marks represent the estimates that were performed with a minimum error of 0.1, while the red + represent the estimates performed without minimum errors.
A.3 Test D: LONG catalog with upper limits

Figure A.3: Comparison between LePHARE and Skelton's estimates of redshift. The horizontal axis shows Skelton's estimation, and the vertical axis shows LePHARE estimation, the gray dashed line is the one to one relation. The blue X marks represent the estimates that were performed without upper limits, while the red + represent the estimates performed with the fixed upper limits published with the catalog.

A.4 Test E: LONG catalog with 5σ upper limits

Figure A.4: Comparison between LePHARE and Skelton's estimates of redshift. The horizontal axis shows Skelton's estimates, and the vertical axis shows LePHARE estimation, the gray dashed line is the one to one relation. The blue X marks represent the estimates that were performed with the 5σ upper limits, while the red + represent the estimates performed with the fixed upper limits published by Skelton.
A.5  Test F: LONG catalog with $5\sigma$ upper limits and $E(B-V)$ maximum increased to 1.5

Figure A.5: Comparison between LePHARE and Skelton’s estimates of redshift. The horizontal axis shows Skelton’s estimation, the vertical axis shows LePHARE estimation, the gray dashed line is the one to one relation. The red X marks represent the estimates that were performed with the max($E(B-V)$)=0.4, while the blue + represent the estimates performed with the max($E(B-V)$)=1.5.

A.6  Test G: LONG catalog, with max($E(B-V)$)=1.5, considering a $S/N < 5$ as an upper limit

Figure A.6: Comparison between LePHARE and Skelton’s estimates of redshift. The horizontal axis shows Skelton’s estimation, and the vertical axis shows LePHARE estimation, the gray dashed line is the one to one relation. The red + marks represent the estimates that were performed with the $S/N < 5$ upper limits, while the blue X represent the estimates performed considering the negative fluxes as upper limits.
A.7 Test H: LONG catalog, $S/N < 5$ upper limits and no IRAC bands 3 and 4

Figure A.7: Comparison between LePHARE and Skelton’s estimates of redshift. The horizontal axis shows Skelton’s estimation, and the vertical axis shows LePHARE estimation, the gray dashed line is the one to one relation. The red + marks represent the estimates that were performed without IRAC 3 and four bands, while the blue X represent the estimates performed with all the photometric bands.
B LePHARE

B.1 zphot.para

##########################################################################
# CREATION OF LIBRARIES FROM SEDs List #
# $LEPHAREDIR/source/sedtolib -t (S/Q/G) -c $LEPHAREDIR/config/zphot.para #
# help : $LEPHAREDIR/source/sedtolib -h (or -help) #
##########################################################################
#
#------ STELLAR LIBRARY (ASCII SEDs)
STAR_SED $LEPHAREDIR/sed/STAR/STAR_MOD.list # STAR list (full path)
STAR_FSCALE 3.432E-09 # Arbitrary Flux Scale
STAR_LIB LIB_STAR # Bin. STAR LIBRARY ->
# $LEPHAREWORK/lib_bin
#
#------ QSO LIBRARY (ASCII SEDs)
QSO_SED $LEPHAREDIR/sed/QSO/QSO_MOD.list # QSO list (full path)
QSO_FSCALE 1 # Arbitrary Flux Scale
QSO_LIB LIB_QSO # Bin. QSO LIBRARY ->
# $LEPHAREWORK/lib_bin
#
#------ GALAXY LIBRARY (ASCII or BINARY SEDs)
#GAL_SED $LEPHAREDIR/sed/GAL/CWW_KINNEY/CWW_MOD.list # GAL list (full path)
GAL_SED $LEPHAREDIR/sed/GAL/BC03_CHAB/BC03_MOD.list # GAL list (full path)
GAL_FSCALE 1. # Arbitrary Flux Scale
GAL_LIB LIB_CWW # Bin. GAL LIBRARY ->
# $LEPHAREWORK/lib_bin
GAL_LIB LIB_BC03_UDS
#SEL_AGE $LEPHAREDIR/sed/GAL/HYPERZ/AGE_GISSEL_ALL.dat # Age list(full path)
# (def=NONE)
AGE_RANGE 0.,13.e9 # Age Min-Max in yr
#
##########################################################################
# FILTERS #
# $LEPHAREDIR/source/filter -c $LEPHAREDIR/config/zphot.para #
# help: $LEPHAREDIR/source/filter -h (or -help) #
##########################################################################
# Filter number and context
# f300 f450 f606 f814 J H K
# 1 2 3 4 5 6 7
# 1 2 4 8 16 32 64 = 127
#
FILTER_LIST wfc3/rf160w.pb,skelt/CFHT_u.pb,skelt/SUBARU_B.pb,skelt/SUBARU_V.pb,
 wfc3/rf606w.pb,skelt/SUBARU_r.pb,skelt/SUBARU_i.pb,skelt/SUBARU_z.pb,
 wfc3/rf125w.pb,skelt/UKIDSS_J.pb,skelt/irac_1.pb,skelt/irac_2.pb,skelt/irac_3.pb,skelt/irac_4.pb
# (in $LEPHAREDIR/filt/*)
TRANS_TYPE 0 # TRANSMISSION TYPE
# 0[=def]: Energy, 1: Nb of photons
FILTER_CALIB 0
# 0[-def]: fnu=ctt
# 1 : nu.fnu=ctt
# 2 : fnu=nu
# 3 : fnu=Black Body @ T=10000K
# 4 : for MIPS (leff with nu fnu=ctt and flux with BB @ 10000K
FILTER_FILE HST_UDS.filt # output name of filter's file -> # $LEPHAREWORK/filt/
#
# THEORETICAL MAGNITUDES
# $LEPHAREDIR/source/mag_star -c $LEPHAREDIR/config/zphot.para (star only)# # help: $LEPHAREDIR/source/mag_star -h (or -help) # # $LEPHAREDIR/source/mag_gal -t (Q or G) -c $LEPHAREDIR/config/zphot.para # # (for gal. & QSO) # # help: $LEPHAREDIR/source/mag_gal -h (or -help)
#
#------- From STELLAR LIBRARY
STAR_LIB_IN LIB_STAR # Input STELLAR LIBRARY in $LEPHAREWORK/lib_bin/
STAR_LIB_OUT STAR_HST_UDS # Output STELLAR MAGN -> $LEPHAREWORK/lib_mag/
#
#------- From QSO LIBRARY
QSO_LIB_IN LIB_QSO # Input QSO LIBRARY in $LEPHAREWORK/lib_bin/
QSO_LIB_OUT QSO_HST_UDS # Output QSO MAGN -> $LEPHAREWORK/lib_mag/
#
#------- From GALAXY LIBRARY
#GAL_LIB_IN LIB_CW0 # Input GAL LIBRARY in $LEPHAREWORK/lib_bin/
GAL_LIB_IN LIB_BC03_UDS # Input GAL LIBRARY in $LEPHAREWORK/lib_bin/
#GAL_LIB_OUT CWW_HST # Output GAL LIBRARY -> $LEPHAREWORK/lib_mag/
GAL_LIB_OUT BC03_HST_UDS # Output GAL LIBRARY -> $LEPHAREWORK/lib_mag/
#
#------- MAG + Z_STEP + EXTINCTION + COSMOLOGY
MAGTYPE AB # Magnitude type (AB or VEGA)
Z_STEP 0.1,6.,0.1 # dz, zmax, dzsup(if zmax>6)
COSMOLOGY 70,0.3,0.7 # H0,om0,lb0d0 (if lb0>0->om0+lb0d0=1)
MOD_EXTINC 1,27 # model range for extinction
EXTINC_LAW LMC_Fitzpatrick.dat # ext. law (in $LEPHAREDIR/ext/*)
EB_V 0.,0.05,0.1,0.3,0.5,0.7,0.9,1.0,1.2,1.5,1.9 # E(B-V) (<50 values)
EM_LINES NO
# Z_FORM 8,7,6,5,4,3 # Zformation for each SED in GAL_LIB_IN
#
#------- ASCII OUTPUT FILES OPTION
LIB_ASCII YES # Writes output in ASCII
# in working directory
#
# PHOTOMETRIC REDSHIFTS
# $LEPHAREDIR/source/zphot -c $LEPHAREDIR/config/zphot.para#
# help: $LEPHAREDIR/source/zphot -h (or -help)
# Input Catalog Informations

**CAT_IN**: $LEPHAREDIR/candels/uds/uds.cat  # Input catalog (full path)

**INP_TYPE**: M  # Input type (F:Flux or M:MAG)

**CAT_MAG**: AB  # Input Magnitude (AB or VEGA)

**CAT_FMT**: MEME  # MEME: (Mag,Err)i

**CAT_LINES**: 1,100000  # MIN and MAX RANGE of ROWS used in input cat

**CAT_TYPE**: SHORT  # Input Format (LONG,SHORT-def)

**CAT_OUT**: $LEPHAREDIR/candels/uds/zphot.out  # Output catalog (full path)

**PARAM_OUT**: $LEPHAREDIR/candels/zphot_output.para  # Output parameter (full path)

**BD_SCALE**: 0  # Bands used for scaling

**GLB_CONTEXT**: -1  # Overwrite Context (Sum 2^n; n=0->nbd-1, 0-def:all bands)

**FORB_CONTEXT**: -1  # context for forbidden bands

**ERR_SCALE**: 0.03,0.02,0.02,0.02,0.04,0.04,0.04  # errors per band added in quadrature

**ERR_FACTOR**: 1.0  # error scaling factor 1.0 [-def]

---

### Theoretical libraries

**ZPHOTLIB**: BC03_HST_UDS,STAR_HST_UDS,QSO_HST_UDS  # Library used for Chi2 (max:3)

**ADD_EMLINES**: NO

---

#### PHOTOMETRIC REDSHIFTS OPTIONS

---

### FIR LIBRARY

**FIR_LMIN**: 7.0  # Lambda Min (micron) for FIR analysis

**FIR_CONT**: -1

**FIR_SCALE**: -1

**FIR_FREESCALE**: YES  # ALLOW FOR FREE SCALING

**FIR_SUBSTELLAR**: NO

---

### PHYSICAL LIBRARY with Stochastic models from BC07

**PHYS_LMIN**: 0

**PHYS_CONT**: -1

**PHYS_SCALE**: -1

**PHYS_NMAX**: 100000

---

### Priors

**MASS_SCALE**: 6.,16.  # Lg(Scaling) min,max [0,0-def]

**MAG_ABS**: -10.0,-26.  # Mabs_min , Mabs_max [0,0-def]

**MAG_REF**: 4  # Reference number for band used by Mag_abs

**ZFORM_MIN**: 5,5,5,5,5,5,3,1  # Min. Zformation per SED -> Age constraint

**Z_RANGE**: 0.0,99.99  # Z min-max used for the Galaxy library

**Z_RANGE**: 0.4

**EBV_RANGE**: 0.1,9  # E(B-V) MIN-MAX RANGE of E(B-V) used
# NZ_PRIOR 4,2,4 # I Band for prior on N(z)
#
#------- Fixed Z (need format LONG for input Cat)
ZFIX YES # fixed z and search best model [YES,NO-def]
#
#------- Parabolic interpolation for Zbest
Z_INTERP YES # redshift interpolation [YES,NO-def]
#
#------- Secondary peak analysis
DZ_WIN 0.5 # Window search for 2nd peaks [0->5; 0.25-def]
MIN_THRES 0.1 # Lower threshold for 2nd peaks[0->1; 0.1-def]
#
#------- Probability (in %) per redshift intervals
# PROB_INTZ 0,0.5,0.5,1.,1.,1.5 # even number
#
# ABSOLUTE MAGNITUDES COMPUTATION
#
MABS_METHOD 1 # 0[-def] : obs->Ref
    # 1 : best obs->Ref
    # 2 : fixed obs->Ref
    # 3 : mag from best SED
    # 4 : Zbin
MABS_CONTEXT -1 # CONTEXT for Band used for MABS

MABS_REF 0 # 0[-def]: filter obs chosen for Mabs :
    # ONLY USED IF MABS_METHOD=2
MABS_FILT 1,2,3,4,5,6 # Chosen filters per redshift bin (MABS_ZBIN)
    # ONLY USED IF MABS_METHOD=4
MABS_ZBIN 0,0.5,1.5,2,3,3.5,4 # Redshift bins (even number)
    # ONLY USED IF MABS_METHOD=4

OUTPUT SPECTRA
#
SPEC_OUT YES # spectrum for each object? [YES,NO-def]
CHI2_OUT NO # output file with all values : z,mod,chi2,E(B-V),...
    # BE CAREFUL can take a lot of space !!

OUTPUT PDZ ANALYSIS
PDZ_OUT uds_pdz # pdz output file name [def-NONE]
    # add automatically PDZ_OUT[.pdz/.mabsx/.mod/.zph]
PDZ_MABS_FILT 2,10,14 # MABS for REF FILTERS to be extracted
#
FAST MODE : color-space reduction
#
FAST_MODE NO # Fast computation [NO-def]
COL_NUM 3 # Number of colors used [3-def]
COL_SIGMA 3 # Enlarge of the obs. color-errors [3-def]
COL_SEL AND # Combination between used colors [AND/OR-def]
#
MAGNITUDE SHIFTS applied to libraries
#
# APPLY_SYSSHIFT 0. # Apply systematic shifts in each band

80
# used only if number of shifts matches
# with number of filters in the library
#
######### ADAPTIVE METHOD using Z spectro sample
#
# AUTO_ADAPT YES # Adapting method with spectro [NO-def]
ADAPT_BAND 4,2,4 # Reference band, band1, band2 for color
ADAPT_LIM 18,22.0 # Mag limits for spectro in Ref band [18,21.5-def]
ADAPT_POLY 1 # Number of coef in polynom (max=4) [1-def]
ADAPT_METH 3 # Fit as a function of
  # 1 : Color Model [1-def]
  # 2 : Redshift
  # 3 : Models
ADAPT_CONTEXT -1 # Context for bands used for training
  # -1[def] used context per object
ADAPT_ZBIN 0.01,6 # Redshift’s interval used for training
  # [0.001,6-Def]
ADAPT_MODBIN 1,1000 # Model’s interval used for training
  # [1,1000-Def]
ERROR_ADAPT NO # [YES,NO-def]
  # Add error in quadrature according to
  # the difference between observed
  # and predicted apparent magnitudes
#

C Matlab scripts

C.1 fluxtomag.m

% This is a script that converts flux into mAB: mAB = 25−2.5 log(f)
% This script has been written to convert the photometric catalogs of
% 3D-HST survey.
% The input file has to be in ASCII and the output will also be an
% ASCII % file.

clear all;
close all;

filename = input(’Data file: ’,’s’)
outfile = input(’Output file: ’,’s’)
disp(’Define the range of the columns that contain the filters.’)
scol = input(’The number of the starting column: ’)
ceol = input(’The number of the ending column: ’)
data = importdata(filename);

nl = size(data.data);

nl = nl(1);
% Copies the initial data columns.
for ii = 1:scol-1
    out(:,ii) = data.data(:,ii);
end

% Converts the fluxes into mAB for the filters columns.
for ii = scol:ecol
    fe = rem(ii, 2);
    for ij = 1:nl
        if (data.data(ij, ii) == -99)
            if (fe == 0)
                if (data.data(ij, ii) <= 0)
                    out(ij, ii) = -99;
                else
                    out(ij, ii) = real(25 - 2.5*log10(data.data(ij, ii)));
                end
            else
                if (out(ij, ii-1) == -99)
                    out(ij, ii) = -99;
                else
                    out(ij, ii) = real(2.5*log10(1+(data.data(ij, ii)/data.data(ij, ii-1))));
                    % out(ij, ii) = real(2.5*log10(data.data(ij, ii)/(data.data(ij, ii-1)*log(10))));
                end
            end
        else
            out(ij, ii) = data.data(ij, ii);
        end
    end
end

% Clearing the Inf values. Setting them to -99.
for ii = scol:ecol
    for ij = 1:nl
        if (out(ij, ii) == Inf | out(ij, ii) < 0)
            out(ij, ii) = -99;
        end
    end
end

% Opening the files... They have to open, it is inevitable.
fid = fopen(filename, 'r');
ofid = fopen(outfile, 'w');
hdstr = fget(s(fid);
comp=strsplit(hdstr);

% Composing and writing to the file the header.
comp2 = '# id';
sz=size(out);
sz=sz(2);
for ii=3:sz+1
    comp2 = strcat(comp2,{',', comp{ii}});
end
strout=comp2{1};
fprintf(ofid, strout);
fprintf(ofid,'\n');

% Closing all files. It's sad I know. But it has to be! But don't worry the
% files will open again next time that you run the script. :) 
fclose(fid);
fclose(ofid);

% Writing the data into the file. If not, what would be the point in
% life to
% write this script? :-|
dlmwrite(outfile, out,'-append','delimiter','

C.2 splitfilters.m

% Splitfilters is a script created to split and create a series of
% files containing the filters that Dr. Rosalind Skelton kindly
% provided. Which I am greateful. :)
close all;
clear all;
filename = input('Filter colection file: ', 's');
 fid = fopen(filename, 'r');
start = 0;
while(~feof(fid))
    tmpl = fgetl(fid);
    tmp = strsplit(tmpl);
    sz = size(tmp); sz = sz(2);
    if(sz > 4)
        if(start == 1 & fini == 1)
```
dlmwrite (outfile, data, 'delimiter', ',');
clear outfile; clear invfile; clear data;
end
fini = 0;
tmp2 = tmp{3};
sz2 = size (tmp2); sz2 = sz2 (2);
ii = sz2; jj = 1;
while (ii > 0 & tmp2 (ii) != '/')
    invfile (jj) = tmp2 (ii);
    ii = ii - 1; jj = jj + 1;
end
sz2 = size (invfile); sz2 = sz2 (2);
ii = sz2; jj = 1;
while (ii > 0)
    outfile (jj) = invfile (ii);
    ii = ii - 1; jj = jj + 1;
end
else
    if (fini == 0)
        start = 1;
        fini = 1;
        nl = 1;
    end
    vec = sscanf (tmpl, '%f');
data (nl, 1) = vec (2); data (nl, 2) = vec (3);
    nl = nl + 1;
end
C.3 SHORT catalog, no minimum error

GAL_SED $LEPHAREDIR/sed/GAL/BC03_CHAB/BC03_MOD.list
(The path to the file that contains the SED templates to be used.)

GAL_LIB LIB_BC03_EGS
(These first lines configure the galaxy SED templates to be used and the output
library where LePHARE will store all the models that fit the constraints given.)

FILTER_LIST wfc3/rf160w.pb, skelt/cfht_u.pb, skelt/cfht_g.pb, wfc3/rf606w.pb, skelt/cfht_r.pb, skelt/cfht_i.pb, wfc3/rf814w.pb, skelt/cfht_z.pb, wfc3/rf125w.pb, skelt/j1_atmos.pb, skelt/j2_atmos.pb, skelt/wirds_J.pb, wfc3/rf140w.pb, skelt/h1_atmos.pb, skelt/h2_atmos.pb, skelt/wirds_H.pb, skelt/k_atmos.pb, skelt/wirds_Ks.pb, skelt/irac_1.pb, skelt/irac_2.pb, skelt/irac_3.pb, skelt/irac_4.pb
(The FILTER_LIST contains a list of all the filters that are used in the
photometry and the sequence in which they are used.)

FILTER_FILE HST_EGS.filt
(This variable has the filename of the filter file, which is where LePHARE
stores the information of the filters to be used in the photometry.)

MAGTYPE AB (This means that the catalog will be in AB magnitudes)
```
Z_STEP 0.2,6.,0.1 (The increment in redshift is of 0.2 and the maximum is z=6. The last number 0.1 is the redshift step if it goes above z=6, which will not happen.)

COSMOLOGY 70,0.3,0.7 (The Cosmology to be used, where H0=70, $\Omega_k=0$, $\Omega_m=0.3$ and $\Omega_\Lambda=0.7$)

EXTINC_LAW LMC_Fitzpatrick.dat (This variable contains the name of the file that has the extinction data. It is using Fitzpatrick extinction law.)

EB_V 0.,0.05,0.1,0.3,0.4 (The values of E(B-V) are going to be used with the extinction law and the SED templates.)

MOD_EXTINC 1,27 (The range of templates to which is going to be applied the extinction.)

CAT_IN $LEPHAREDIR/candels/aegis/aegis_nul_short.cat (Indicates the path and filename of the photometric catalog.)

INP_TYPE M (This is the input type. It tells LePHARE what type of information is in the catalog. The M indicates that the catalog is in magnitudes.)

CAT_MAG AB (Indicates the catalog is in AB magnitudes)

CAT_FMT MEME (This variable configures the sequence in which the data is written in the catalog. In this case it is written in alternate columns of magnitudes and errors.)

CAT_TYPE SHORT (Is the type of catalog. When using the SHORT type, the catalog only has one column with the ID of the object and a sequence of columns with the magnitude/flux and associated errors.)

CAT_OUT $LEPHAREDIR/candels/aegis/zphot.out (Contains the path and filename where the estimated values are stored.)

PARA_OUT $LEPHAREDIR/candels/aegis/zphot_output.para (Contains the path and filename of the file that tells which parameters are going to be written in the output file.)

SPEC_OUT YES (Tells the program to create an output file that contains the SEDs that best fit the photometric data.)