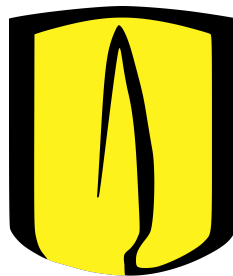


Observational evidence of star formation stochasticity in the CALIFA dataset

Nicolás Romero Díaz

201127499

Advisor: Jaime Forero-Romero



Contents

1	Introduction	2
2	The CALIFA survey	4
2.1	The CALIFA survey	4
2.2	The CALIFA datacubes	5
3	Computational tools	6
3.1	The fitsio library	6
3.2	The Pipe3D analysis pipeline	7
3.2.1	Analysis of the stellar population	7
3.2.2	Analysis of strong emission lines	8
4	Results	9
5	Discussion	16
5.1	Measuring stochasticity	16
5.1.1	The Balmer decrement	16
5.1.2	Galactic extinction	17
5.2	Comparison with theoretical results	18
6	Conclusions	20
7	Appendix	21
7.1	List of galaxies	22

Chapter 1

Introduction

Star formation processes are the mechanisms that turn galactic gas clouds into stars. These processes are probabilistic by nature and are mostly parametrized by two components. The initial mass function (IMF) and the cluster mass function (CMF). The IMF outlines the relative abundance of stars of a certain mass to be formed, and is represented by $\phi(m)$. For the most part, the relevant characteristics of the IMF are described by two pieces of information. (i) The mass interval, describing the available range of mass possible for the stars being generated ($m_{min} - m_{max}$); (ii) the relative abundance of stars with different stellar masses, taken as $\phi(m) \propto m^{-\gamma}$. When $\gamma > 0$ less massive stars are more likely to be formed.

The second parameter arises from the fact that star formation is mainly a cluster phenomenon, characterized by the cluster mass function, $\psi(M_{ecl})$. This CMF describes the number of clusters in a region where stars are being produced, with M_{ecl} the cluster mass.

The sampling of these two parameters of star formation is intrinsically related with the star formation rate (SFR) of a particular system. This correlation arises from the fact that less massive stars will dominate the bulk of star formation. Due to this, regions of low SFR will generally not produce enough stars to account for the less probable massive stars. Fumagali et. al (2011) [1] have proposed that in these areas, stochastic processes have a greater influence over the statistical nature of both the IMF and CMF. This statistical foundation of the mass functions has several distinct effects that influence different spectral properties of gas clouds surrounding star forming localities. In particular, stochasticity appears to cause the ratio of H_α flux to Far Ultra Violet (FUV) photons to fluctuate around a mean value [1].

Using the public SLUG code (Stochastically Light Up Galaxies), Forero-

Romero and Dijkstra [2] have found that this fluctuation around a mean value also takes place when comparing the ratio of Lyman alpha (Ly_α) flux to the FUV photons. This relation is a well known spectral quantity called the equivalent width (EW), which characterizes the Ly_α emission line strength. The interest in the EW and its fluctuation arises from the fact that the Ly_α line is a powerful indicator of stellar populations and intergalactic dust and can be a determining factor for classification of galaxies [2]. In another work, Mas-Ribas et.al [3] also observed fluctuation of the Ly_α luminosity when performing stochastic IMF sampling.

In order to search for this stochastic effects, we will be using data published by the Calar Alto Legacy Integral Field Area Survey (CALIFA), reported in detail by Sánchez et. al. [4, 5]. A compendious description of the CALIFA data will be presented in chapter two, as well as an overall description of the survey.

The minutiae of the handling of CALIFA data is detailed in chapter three, as well as a discussion of the different computational tools used to navigate this data. This chapter will also including a general outline of the Pipe3D analysis pipeline developed by the CALIFA team of Sánchez et. al. [6]

Representative results are included in chapter four.

In chapter five we present a discussion of the results obtained. The detection and quantification of stochasticity is also detailed, as well as how spectral data can be affected by phenomena in the interstellar medium (e.g. galactic extinction, interstellar reddening). Finally, we present our conclusions in the last chapter, a catalog of images, full results and a table consisting of all galaxies researched for this work.

Chapter 2

The CALIFA survey

In this chapter we present an overview of the CALIFA survey, including several generalities concerning the CALIFA data. In particular, we discuss the galaxies observed and the selection criteria of the galaxies which are relevant to this work. In addition, a brief description of the two different configurations available for CALIFA data is presented.

2.1 The CALIFA survey

The CALIFA survey is a comprehensive wide field integral field unit (IFU) survey carried out in the Calar Alto observatory in Almería, Spain. It is operated jointly with the Max Planck Insititut für Astronomie and the Instituto de Astrofísica de Andalucía. Observations began in June of 2010, with the survey collecting data from over six hundred galaxies in two configurations, the V500 and the V1200 set ups [4] described in the following section.

The reasoning for selecting CALIFA survey data is simple. As seen in figure ??, CALIFA data does not consists of a single spectra, instead, multiple optic fibers were pointed at a single galaxy [4, 5]. For previous surveys having a single measure of a spectra meant that for an entire galaxy there was only one reported value of total SFR. Since stochasticity theoretically appears in low SFR regions, the ability to focus on distinct regions that have lower SFR opens the possibility of searching for this stochastic effects.

In this work, we will be using “face on” galaxies from the CALIFA catalog published in [6]. These type of galaxies are optimal for a complete scanning and allow us to measure the spectra of individual regions within a galaxy. This “face on” condition implies that a manual selection of the datacubes in [6] must be done in order to select galaxies whose galactic planes are

perpendicular to our field of vision.

2.2 The CALIFA datacubes

The CALIFA datacubes consist of a three dimensional array, where the x and y coordinates indicate the right ascension and declination of the target, and the z axis contains information about flux intensity and wavelength of that pixel along with the error associated with those fluxes, a mask to cover bad pixels and the covariance weight of the error propagation [4, 5].

The V500 and V1200 configurations cover different wavelength intervals and have different resolution, with the V500 datacube having a resolution of $\lambda/\Delta\lambda \sim 850$ and covering emission lines between $3745 - 7500 \text{ \AA}$, while the V1200 configuration has a higher resolution of $\lambda/\Delta\lambda \sim 1650$ and compiles wavelengths from $3700 - 4800 \text{ \AA}$. [4, 5]

In this work, we will be using the lower resolution V500 datacubes published in [6] which have been subjected to the Pipe3D analysis pipeline explained in section 3.2 and detailed in [6]. These datacubes contain a great deal of information, including that of intermediate steps of the fitting procedure.

Chapter 3

Computational tools

In this chapter, we will discuss the details of the computational tools pertinent to this work. We will review the different libraries and procedures that have been applied in order to organize and analyze the different CALIFA datacubes. Specifically, we will discuss the structure of said datacubes, how the `fitsio` library allows us to manipulate `.FITS` files and how the Pipe3D pipeline is applied to both the spectra of the stellar population of a galaxy as well as the spectra of giant gas clouds in a galaxy in order to obtain an optimum fit of the different spectra present in the datacubes.

3.1 The `fitsio` library

The `FITS` format is widely used for spectroscopic data since it can contain all pertinent information while easily maintaining a correspondence with the geographical structure of the galaxy. This means that all observed data for a spaxel is easily accessed with its corresponding x and y coordinates.

In order to successfully extract, manage and analyze the `.FITS` datacubes we will use the `fitsio`¹ library, which allows us to navigate the three dimensional data matrices with ease. This library includes several commands that allows us to take slices from the three dimensional datacubes and thus handle data for each galaxy as if it were regular two dimensional array in Python. This library also permits us to call relevant information by means of keywords and data columns.

¹<https://github.com/esheldon/fitsio>

3.2 The Pipe3D analysis pipeline

In the case of this study, we are interested in fluctuations of the EW of the OII line emitted by gas clouds that surround star forming regions. The OII EW is defined as the ratio between the intensity of the OII emission line and the continuum radiation in its vicinity. This means that we need a robust fitting procedure in order to characterize both this emission line spectra and continuum radiation since they will greatly affect our data of the EW.

The Pipe3D pipeline proposed by Sánchez et.al [6] consist of a series of steps followed in order to organize and successfully provide a fit for the spectra in the CALIFA datacubes. Since the original data consists of a superposition of spectra of gas clouds in interstellar environments as well as the underlying star population, these two spectra have to be treated using several procedures. It is also important to make a distinction between weak and strong emission lines since fitting procedures are different for each. Since we will be studying H_α , H_β and OII lines, we will not include the fitting procedure of weak emission lines.

3.2.1 Analysis of the stellar population

We are able to study the spectra of the stellar populaion using an algorithm known as “CS-binning algorithm”, which consists of dividing the datacube into two dimensional cells called “spatial bins” and subsequently taking the mean value of the pixels included, masking pixels with bad values. Once we have the averaged spectra, Pipe3D is able to fit the observed data via an iterative procedure [6]. When applying this procedure, there appears an associated error for these new binned spectra since bad pixel data can adversely modify the resultant dataproducuct. Pipe3D provides the associated error of each spaxel as part of the original datacube. The ultimate goal of this analysis is to be able to obtain an optimum depiction of the stellar population and subtract it from the datacubes, leaving the remaining data of the spectra associated exclusively with the emission lines of the gas clouds in interstellar regions of the galaxy.

After this binning procedure, several maps are created:

- Row Stack Spectra (RSS)
- A position table for each galaxies binned cube, corresponding to the indices of the spatial bins from brightest to dimmest.
- Two intensity maps at the wavelength range of the V-band, with one

map corresponding to the intensity before the binning procedure and the other one to the binned data.

Once this results are obtained, the stellar population is fitted with a Simple Stellar Population (SSP) template and thus the systemic velocity, the velocity dispersion and dust attenuation can be derived.

The final step consists of fitting the strong emission lines and subtracting them from the datacube obtained at this stage. Then, the procedure is repeated until a certain condition of χ^2 is met. The fitting of strong emission lines is presented in the following section.

3.2.2 Analysis of strong emission lines

A Gaussian fit, though not sophisticated, is valid for fitting strong emission lines, since we will not be taking into account gas-rich major merger galaxies, AGN nuclei nor galaxies that intercept in our observational foreground. Pipe3D divides the observed wavelengths into four groups:

- $[O]_{II}\lambda 3727$ (3700 – 3750) Å
- $H_\beta, [O]_{III}\lambda 4959$ and $[O]_{III}\lambda 5707$ (4800 – 5050) Å
- $[N]_{II}\lambda 6548, H_\alpha$ and $[N]_{II}\lambda 6583$ (6530 – 6630) Å
- $[S]_{II}\lambda 6717, [S]_{II}\lambda 6731$ (6680 – 6770) Å

Before assigning a gaussian function to a particular emission line, a first guess must be done of its kinematic properties based on the expected H_α wavelength of the radiation, taking into account that particular galaxies redshift and by making a parabolic approximation to the centroid of the emission line.

Finally, the emission line is fitted by using a small range of systematic velocities centered around the initial guess. This procedure results in a set of maps of an *emission line pure cube* which includes different parameters for each emission line [6].

Chapter 4

Results

For representative results of the analysis applied we will present four galaxies: (i) IC4566; (ii) NGC0001; (iii) NGC0036 (iv) UGC09476 represented in figures 4.1 to 4.4.

We begin by visualizing each galaxy's OII, H_α and H_β intensities. Then, we plot the H_α/H_β ratio that Fumagalli checked for stochasticity in [1]. Taking the logarithm in base ten gives us a clearer view of smaller variations in the values of this ratio. Since stochasticity would theoretically alter the values of this quotient, we will present the $\log H_\alpha$ vs. $\log H_\alpha/H_\beta$ distribution in more detail in figure 4.5.

Making a histogram from the data in figure 4.5, we obtain figure 4.6. It is important to separate the H_α data in three regions. A bright region, a mid-range region and a faint region. The bright region is plotted in red, the mid-range spectra is green and the faint spectra in blue. The distinction arises from the fact that stochastic effects appear in simulations when there is low SFR and therefore have a dimmer spectra. The segmented vertical line indicates the theoretical value for the H_α/H_β ratio, which is 2.85^1 and whose value is theoretically affected by stochasticity.

We also check the distribution for the OII EW. Once again, we separate our EW data into bright H_α and faint H_α . Results for the galaxies are presented in figure 4.7. The green histogram now indicating the bright region and blue for the faint region.

After identifying the low SFR regions in figure 4.6 we perform a fit of this probability density function. The resulting values of the free parameters obtained are then presented in table 1.

¹Osterbrock, Astrophysics of Planetary Nebulae and Active Galactic Nuclei, University Science Books, 1989

IC4566

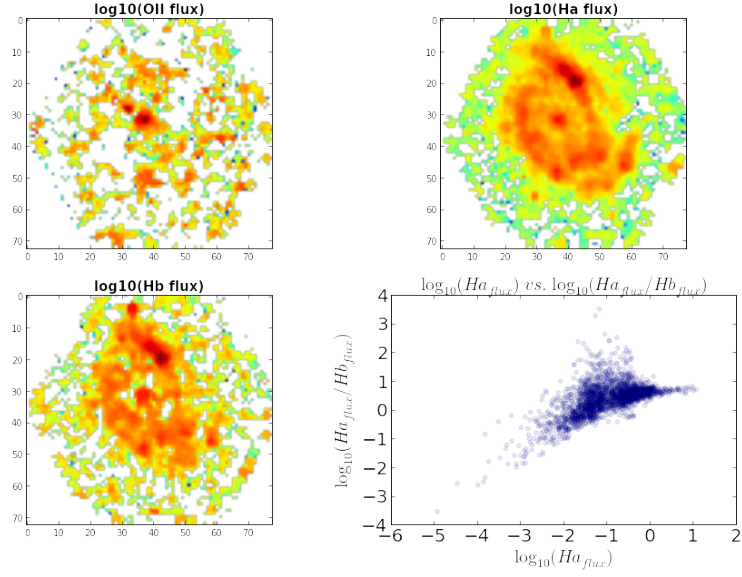


Figure 4.1: a) OII flux, b) H α flux, c) H β flux, d) $\log(H\alpha)$ vs. $\log(H\alpha/H\beta)$

NGC0001

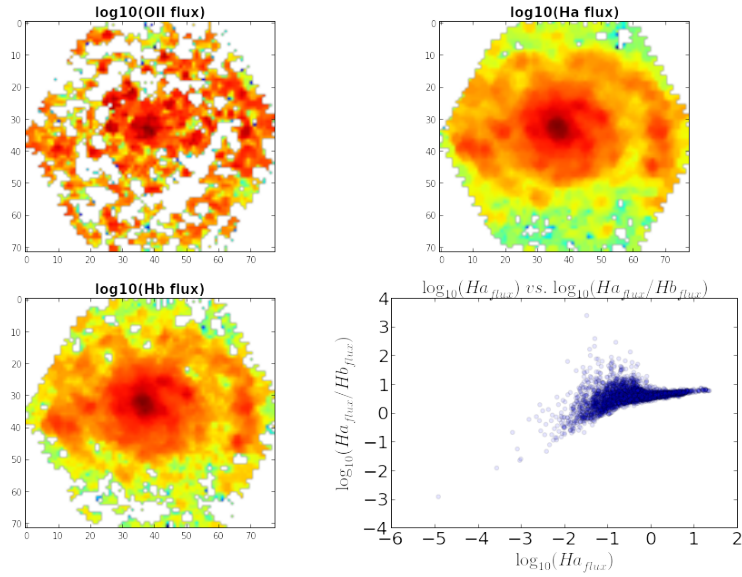


Figure 4.2: a) OII flux, b) H α flux, c) H β flux, d) $\log(H\alpha)$ vs. $\log(H\alpha/H\beta)$

NGC0036

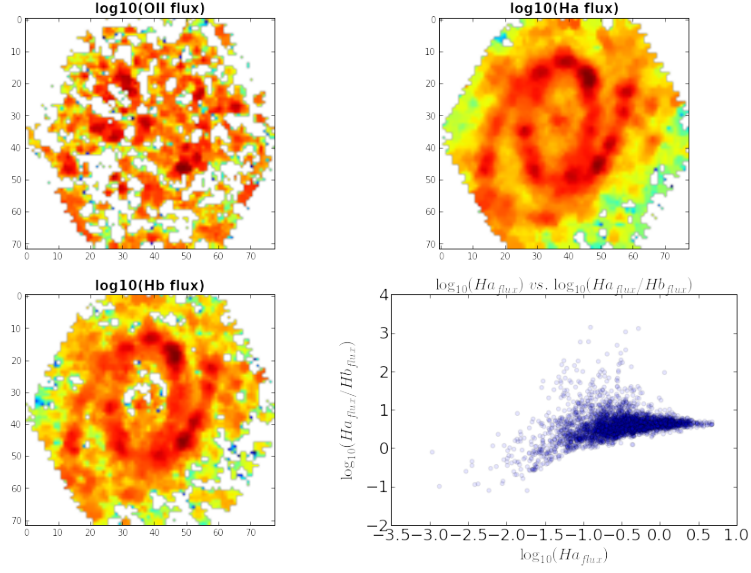


Figure 4.3: a) OII flux, b) H_α flux, c) H_β flux, d) $\log(\text{H}_\alpha)$ vs. $\log(\text{H}_\alpha/\text{H}_\beta)$

UGC09476

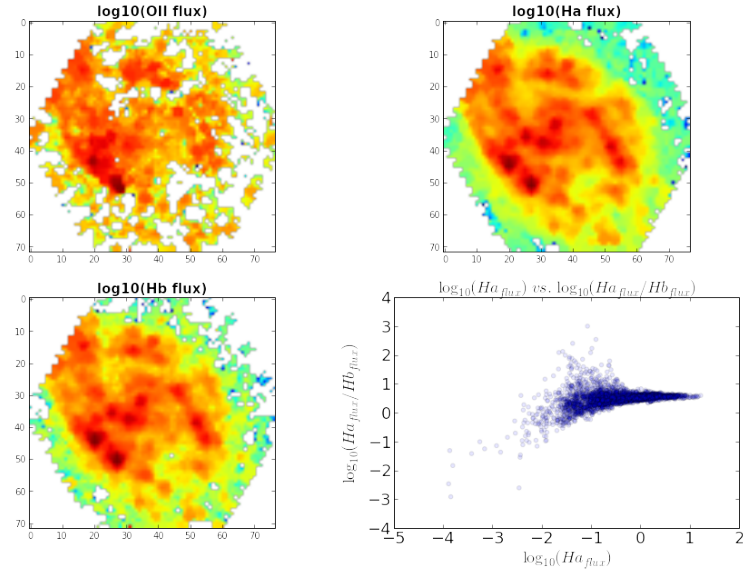


Figure 4.4: a) OII flux, b) H_α flux, c) H_β flux, d) $\log(\text{H}_\alpha)$ vs. $\log(\text{H}_\alpha/\text{H}_\beta)$

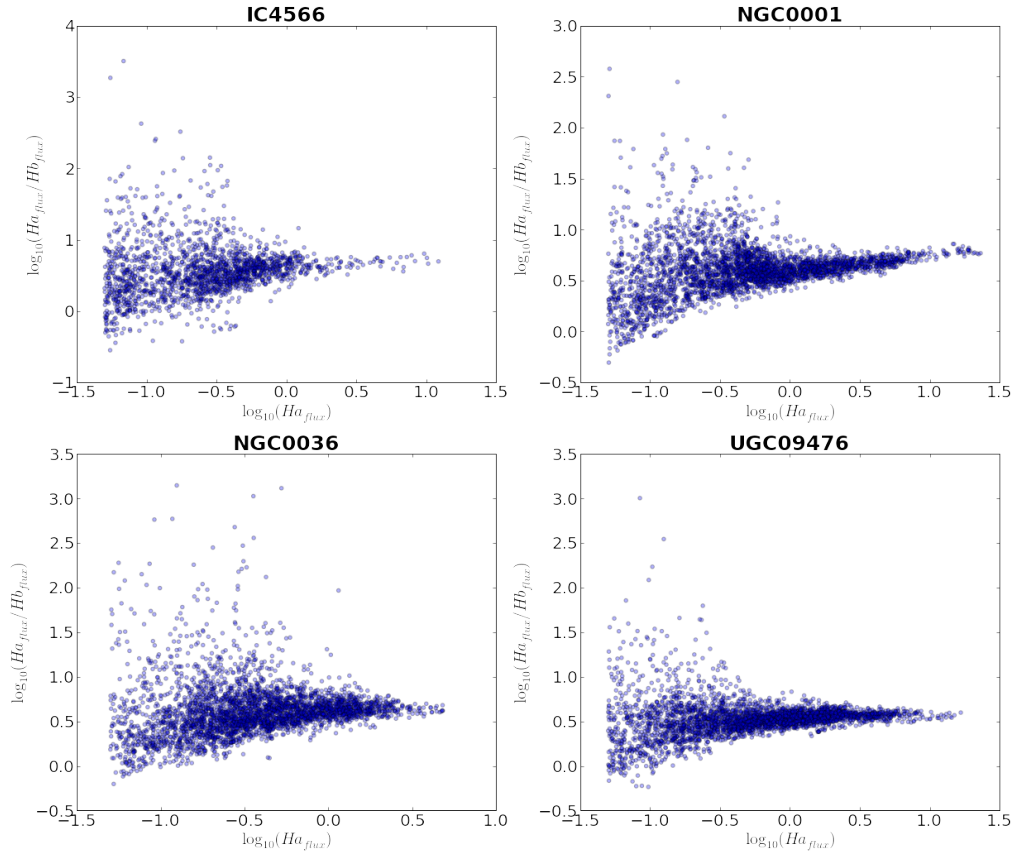


Figure 4.5: Distribution of $\log H_{\alpha}/H_{\beta}$ values for our representative galaxies

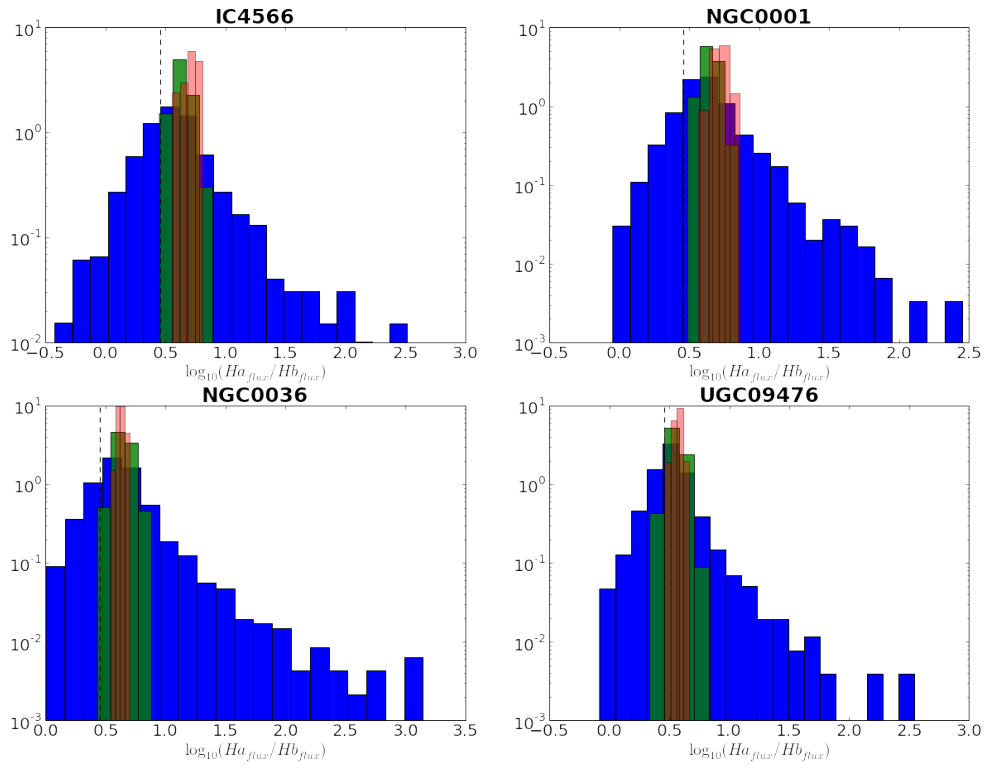


Figure 4.6: Distribution of $\log H_{\alpha}/H_{\beta}$ values for our representative galaxies plotted as histograms. The blue, green and red histograms correspond to the faint, mid range and bright regions of the H_{α} spectra respectively. The segmented line lies on the theoretical value of the ratio.

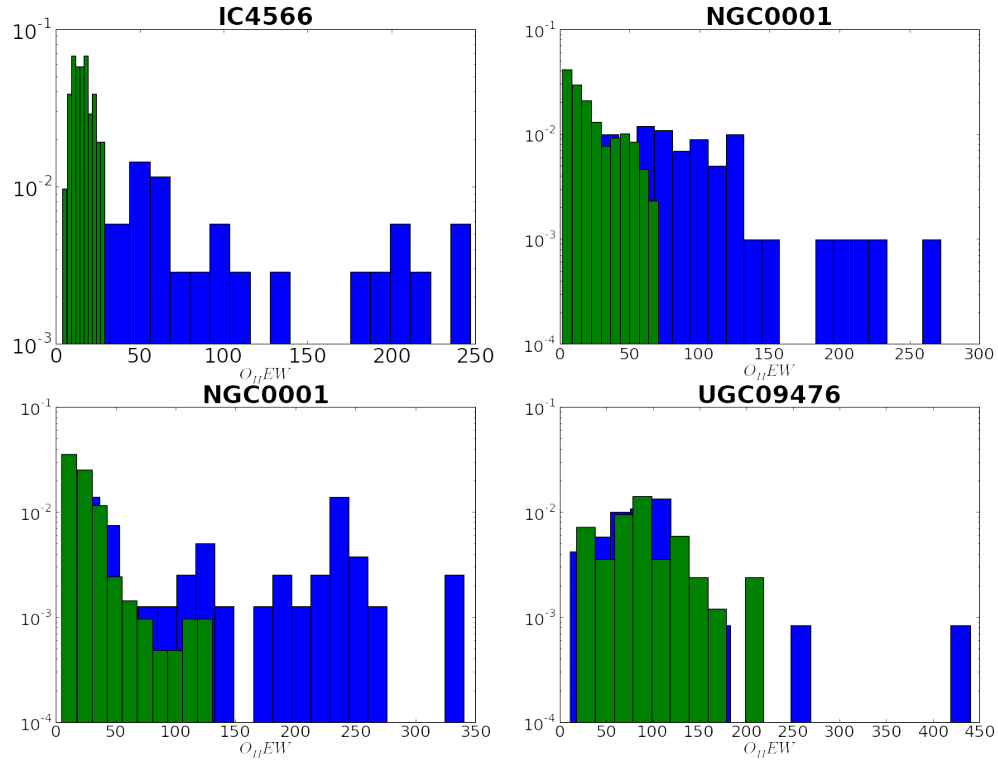


Figure 4.7: Distribution of the OII EW. The blue and green histograms correspond to the faint and bright regions of the H_{α} spectra respectively.

Table 1 Results of the estimation of the free parameters.

P_0	\mathcal{M}_0	α	γ
Me	faltan	los	resultados

Chapter 5

Discussion

5.1 Measuring stochasticity

According to [1], the H_α/H_β ratio fluctuates if stochasticity is present. Looking at the trimmed data in chapter four, figure 4.5, we can see that at a lower value of H_α , the H_α/H_β quotient starts becoming more disperse. This is consistent with the computational prediction in [1] since the H_α intensity is an indicator of star formation productivity, with higher H_α intensity corresponding to a higher SFR and a lower intensity to a lower SFR.

Once we visualize this distributions in figure 4.6, we can also recognize a distribution around the theoretical value of this ratio. In fact, the bright region of the H_α/H_β spectrum presents a more narrow distribution while the dimmer area presents a broader distribution.

In [2], Forero-Romero expands upon Fumagalli's [1] conclusions and proposes that this fluctuation should be present on other emission line ratios. We find looking at the histograms in figures 4.7 that distributions are also present in the OII EW. These distributions appear to behave in the same manner as the H_α/H_β ratio, with higher values presenting a lesser variation and lower values being more broadly scattered.

Even though we have found evidence of fluctuation of H_α/H_β and EW, we must be careful to consider this as evidence of stochasticity since there are other phenomena that can cause similar effects. Some of this phenomena are discussed below.

5.1.1 The Balmer decrement

The Balmer decrement, also known as interstellar reddening, is a phenomenon that occurs when electromagnetic radiation passes through a dust

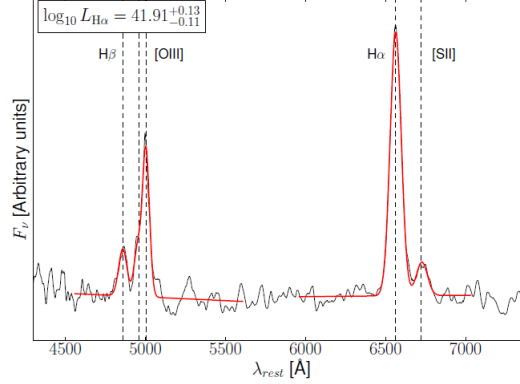


Figure 5.1: fit

cloud in the interstellar medium. These dust particles are micron sized, which causes a greater absorption of bluer light as it is more likely to collide with the particle. This means that when we observe this light it will appear redder.

When we analyze the H_α/H_β ratio this absorption must be taken into account, since the H_β emission line has a shorter wavelength than H_α and will more likely collide with dust particles and affect our measured value of H_α/H_β . This attenuation of H_β means that when checking for the value of H_α/H_β we will observe a greater number since H_β has become dimmer in comparison to H_α . This absorption can be seen in figure (FIGURA).

It is important to consider interstellar reddening when discussing our data since it shows that not all fluctuations of the H_α/H_β are due to stochasticity. Despite this, we can see from the histograms in figures (FIGURAS) that this effect does not fully explain the distributions observed, as it only causes the H_α/H_β ratio to move to the right. We can observe that in our data there are values of H_α/H_β that lie to the left of the theoretical value and are not explained by this phenomenon.

5.1.2 Galactic extinction

Galactic extinction is another effect due to the presence of dust in the interstellar medium.

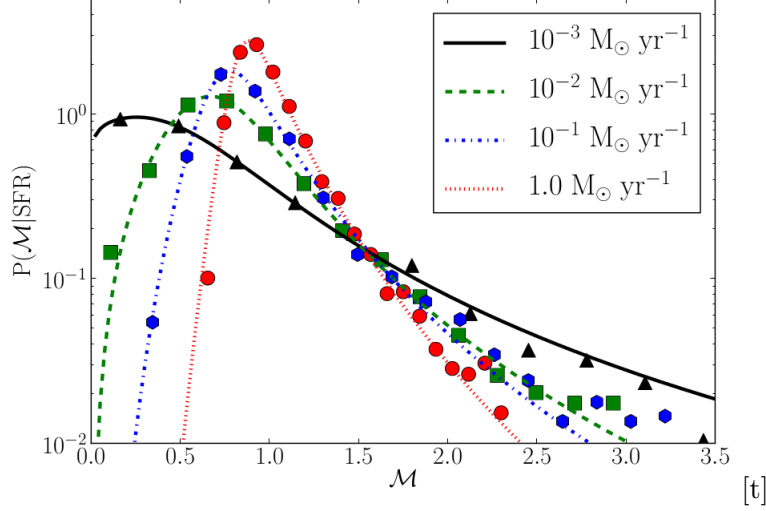


Figure 5.2: Visualization of equation 5.2 for multiple values of SFR, taken from [2]

5.2 Comparison with theoretical results

In order to simplify our comparison, it is suitable to adopt the same conventions as [2], where the parameter \mathcal{M} is defined as the variation of measured EW to expected EW.

$$\mathcal{M} \equiv \frac{\text{EW}}{\text{EW}_0} \quad (5.1)$$

The authors describe the fitting of this curve as a SFR dependent probability density function (PDF) of the form:

$$P(\mathcal{M}|\text{SFR}) = P_0 \left[\left(\frac{\mathcal{M}}{\mathcal{M}_0} \right)^{-\alpha} + \left(\frac{\mathcal{M}}{\mathcal{M}_0} \right)^{\gamma} \right]^{-1} \quad (5.2)$$

where the star formation rate is present by the constants P_0 , \mathcal{M}_0 , α , γ . This double power law is valid for values of $P(\mathcal{M}|\text{SFR}) > 10^{-2}$ due to the limited number of points available to perform a thorough fit of the data. This PDF is visualized in figure 5.2 for multiple star formation rates. In this figure, it is clear that as SFR increases, the peak of the function becomes more pronounced. Ideally, if no dust effects or stochasticity are present, this plot would consist of a delta function centered around $\mathcal{M} = 1$. It is important to observe that the dust effects can explain the values of \mathcal{M} found in the $\mathcal{M} > 1$ region, but we can see that in both the computational results and

Table 2 The set of the first six parameters obtained in computational simulations in [2]

P_0	\mathcal{M}_0	α	γ
1.15 ± 0.05	0.75 ± 0.03	0.13 ± 0.03	2.67 ± 0.10
1.44 ± 0.04	0.71 ± 0.02	0.44 ± 0.03	2.73 ± 0.09
1.66 ± 0.07	0.71 ± 0.03	0.74 ± 0.06	2.82 ± 0.16
1.86 ± 0.03	0.76 ± 0.01	1.03 ± 0.03	3.49 ± 0.09
1.97 ± 0.01	0.80 ± 0.01	1.21 ± 0.02	3.83 ± 0.05
2.37 ± 0.06	0.77 ± 0.01	1.81 ± 0.10	4.01 ± 0.10

in our data obtained, there are values that lie in the $\mathcal{M} < 1$ region which are not due to interstellar dust effects.

We can now contrast our fit obtained by our analysis in table 1 with the values of the free parameters estimated by [2]. The set of the first six parameters derived by Forero-Romero and Dijkstra are displayed in table 2.

Chapter 6

Conclusions

?

Chapter 7

Appendix

7.1 List of galaxies

Name			
UGC00005	NGC2880	UGC09067	NGC6478
NGC7819	NGC2906	NGC5520	NGC6497
IC1528	NGC2916	NGC5614	UGC11262
UGC00036	UGC05108	NGC5630	NGC6941
NGC0001	UGC05358	NGC5720	NGC6978
NGC0036	UGC05359	UGC09476	UGC11649
UGC00312	NGC3106	NGC5784	NGC7311
NGC0171	NGC3994	NGC5888	NGC7321
NGC0180	NGC4003	NGC5930	UGC12127
NGC0192	UGC07012	IC4566	UGC12185
NGC0237	NGC4185	NGC6004	UGC12224
NGC0477	NGC4210	NGC6020	NGC7549
IC1683	IC0776	NGC6021	NGC7563
NGC0496	NGC4470	NGC6032	NGC7562
NGC0774	UGC08107	NGC6063	NGC7625
NGC0776	UGC08231	IC1199	NGC7631
NGC1349	UGC08234	NGC6125	NGC7653
NGC1645	NGC5000	NGC6132	NGC7716
UGC03253	NGC5205	NGC6146	NGC7738
NGC2253	NGC5216	NGC6154	UGC12816
NGC2347	UGC08733	NGC6173	NGC7489
UGC03995	UGC08781	UGC10693	UGC12864
NGC2449	NGC5378	UGC10695	NGC7800
NGC2730	NGC5406	IC1256	NGC5947