

Analysis of observational data, to calculate the Delayed-Tau model parameters

Dimitrios Papachristopoulos

The Star Formation History (SFH) of a galaxy can offer many insights not only for the evolution and the future of the galaxy, but also for the evolution of the Universe. This is why there are various theoretical models trying to describe the SFH of galaxies. One of those models is the Delayed-Tau model, which approximates the Star Formation Rates (SFR) of galaxies as a function with a rising SFR at the beginning, until it reaches a peak at a time τ , unique for each galaxy, and then it drops at an exponential rate.

Haslbauer, Kroupa, and Jerabkova (2023) argue that the Delayed- τ model is opposed to the Lilly-Madau plot ((Madau and Dickinson 2014)), which plots the observed SFR's of galaxies with the corresponding redshifts (z) and calculates a cosmic SFR peak at $z \approx 2$. The way they calculated this inconsistency is by using observatory data for SFR and Stellar Masses from the UNGC catalog (Karachentsev, Makarov, and Kaisina (2013), Karachentsev and Kaisina (2013)) for calculating the parameters (the timescale τ and the normalization constant A_{del}) of the model. This calculation for the galaxies of the Local Cosmological Volume (LV), allows the investigation of the SFR throughout the life of each galaxy and so we can find the expected time of peak of the SFR.

In this thesis project, we will try to calculate the same parameters, by using a bigger sample size and the method Markov Chain Monte Carlo, to examine if the inconsistencies of the model derive from the results of the previous analysis, or if it is an intrinsic problem of the model

Keywords: Galaxies, Galaxy Evolution, Star Formation History (SFH), Star Formation Rate (SFR), Delayed- τ , Local Cosmological Volume, Lilly-Madau Plot, Redshift, Markov Chain Monte Carlo (MCMC).

Table of contents

1 Galaxy Morphology and Star-Forming Regions	2
1.1 Galaxy Classification	2
1.1.1 Dwarf Galaxies	5
1.2 Star-Forming Regions	5
2 Star Formation History (SFH)	6
2.1 Star Formation Rate	6
2.1.1 Estimating SFR from Spectra	7
2.2 Main Sequence Galaxies	9
2.2.1 Star Formation History Models	10
3 Lilly-Madau Plot and Delayed-τ model conflicts	11
3.1 Redshift and lookback time	11
3.2 Lilly-Madau Plot	13
3.3 Delayed- τ model	14
4 Computational Methods	15
4.1 Newton-Raphson	15
4.2 Markov Chain Monte Carlo	15
5 Data of the Local Cosmological Volume	17
5.1 Catalog Completeness	18
5.2 Comparing the Catalogs	19
6 Calculating the parameters	22
6.1 According to pre-existing bibliography	22
6.1.1 Problems with the method	22
6.2 MCMC	22
7 Appendix A: Scatter plots of the comparisons	23
References	25

1 Galaxy Morphology and Star-Forming Regions

This thesis will focus on how a specific parametric model tries to explain the Star-Formation Histories of galaxies, but to do that, we first need to understand what a galaxy is, what it is made of, the mechanisms of its evolution, and how we can distinguish them.

The study of galaxies is a very active field of astronomy since it is a relatively young discipline. Until 1920, astronomers who observed spiral nebulae were not certain what they were. In 1921, two papers were published: one argued that the Milky Way constituted the whole Universe and that the spiral nebulae were part of it, while the second argued that each spiral nebula was,

in fact, a distinct “island universe” (Shapley and Curtis 1921). This debate became known as the Great Debate¹.

In 1925, Edwin Hubble put an end to the debate by showing that the distances to the spiral nebulae were far too great compared to other objects within the Milky Way. The method he used involved observing Cepheid variable stars in these nebulae. By applying the period-luminosity relationship, which had been discovered by Henrietta Swan Leavitt (Leavitt 1907), Hubble was able to determine their distances. This groundbreaking discovery confirmed that spiral nebulae were indeed separate galaxies, marking the beginning of extragalactic astronomy and revolutionizing our understanding of the Universe.

Today we know that galaxies are large-scale structures containing, Stars, gas and dust, Stellar and planetary systems and Stellar remnants (white dwarves, neutron stars and black holes). Those structures are held together by their gravity, having Stellar Masses more than $\sim 10^5 M_\odot$, and an average diameter of ~ 4 kpc. They are extremely diverse systems, and each galaxy differs in mass, size, brightness, stellar populations and morphology. This is exactly why we have created systems to classify them.

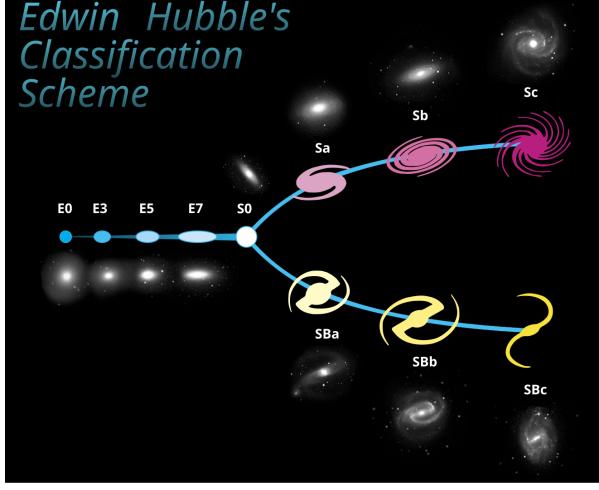
1.1 Galaxy Classification

One of the most common methods of classification is the *Hubble classification* (Hubble 1925) , which categorizes galaxies based on their morphology.

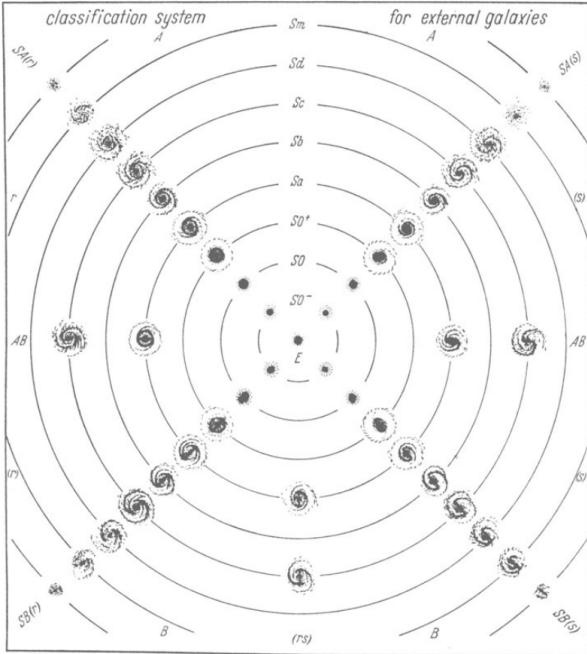
- **Elliptical Galaxies (E):** Ellipsoidal shapes with smooth brightness profiles, containing older stars and minimal interstellar matter. Their eccentricity takes values from 0 to 0.7, so we can categorize them even further, from E0 to E7
- **Lenticular Galaxies (S0):** Intermediate between elliptical and spiral galaxies, featuring a central bulge and disk but lacking significant spiral structure.
- **Spiral Galaxies:** Characterized by a central bulge and spiral arms. Depending on the arm tightness we can categorize them as *a,b,c,d* from tight to looser, and depending if they have a bar or not they are subdivided into:
 - **Unbarred Spirals (S):** No central bar; classified as Sa, Sb, Sc, etc., based on arm tightness and bulge size.
 - **Barred Spirals (SB):** Feature a central bar; denoted as SBa, SBb, SBc, etc.
- **Irregular Galaxies (Irr):** Lack regular structure, often rich in gas and dust with high star formation rates.

¹The chain is homogeneous if $T\forall x^{(i)}$ remains invariant (the probability of moving from $x^{(i-1)}$ to $x^{(i)}$ does not depend on the iteration number i .) and $\sum_{x^{(i)}} T(x^{(j)} | x^{(i-1)}) = 1$.

Edwin Hubble's Classification Scheme



(a) Tuning-fork-style diagram of the Hubble sequence By Cosmogoblin - Own work, CC0, <https://commons.wikimedia.org/w/index.php?curid=121743256>



(b) Hubble-de Vaucouleurs classification system By de Vaucouleurs (1959)

Figure 1: Diagrams visualizing the Hubble and Hubble-de Vaucouleurs morphological classification systems

Although Hubble's original scheme was revolutionary, **de Vaucouleurs** recognized that not all galaxies fit neatly into its categories. In response, he introduced a more nuanced classification system (de Vaucouleurs 1959) that:

- **Accounts for Rings:** Some galaxies feature ring-like structures around their bulge or bar. To denote this, de Vaucouleurs added (**R**) to the classification (for example, (R)SBa).
- **Differentiates Bar Strength:** Since bars can vary from subtle to dominant, he proposed **SA** (unbarred), **SAB** (weakly barred), and **SB** (strongly barred).
- **Incorporates Numerical Types (T-Types):** To capture subtle transitions along the morphological sequence, de Vaucouleurs assigned a numerical index (TTT) ranging from -6 (pure compact ellipticals) to +10 (extreme irregulars). Intermediate values (e.g., -1 for S0, 2 for Sab, 5 for Sc, etc.) let astronomers pinpoint galaxies that don't fit cleanly into the original categories.

This expanded framework also embraces **early-type** and **late-type** galaxies as part of a continuous evolutionary sequence:

- **Early-type galaxies** (E and S0):
 - Smooth appearance.
 - Predominantly older stellar populations.
 - Minimal amounts of gas and dust.
 - Assigned negative T-values ($-6 \leq T < 0$).
- **Late-type galaxies** (Spirals and Irregulars):
 - Rich in gas and dust.
 - Significant ongoing star formation.
 - Smaller bulges and more open arms from Sa/SBa to Sc/SBc.
 - Assigned positive T-values , extending to +11 for extreme irregulars.

By offering extra designations for bar strength, ring features, and transitional morphologies, de Vaucouleurs' system paints a more complete picture of galaxies and how they evolve. It allows researchers to quantify where a galaxy lies along the continuum, rather than forcing it into a single rigid label.

Table 1: This table was based on the L^AT_EX files of (“RC3 - Third Reference Catalog of Bright Galaxies” n.d.)

Hubble	E	E-S0	S0	S0/a	Sa	Sa-b	Sb	Sb-c	Sc	Irr	Sc-Irr
T	-5	[−4,−3]	[−2,−1]	0	1	2	3	4	[5,6,7]	8	[9,10,11]

1.1.1 Dwarf Galaxies

Despite the usefulness of de Vaucouleurs’ numerical T-type system , we run into practical issues when classifying dwarf galaxies. These galaxies need a more detailed classification system, since the T-Type system can denote galaxies, whose physical properties drastically differ, with by the same number. For instance, dwarf spheroidals and normal ellipticals both end up with a T-value below zero ($T < 0$), even though their properties are vastly different. Then there are “transient” dwarf galaxies (Tr), whose features bridge spheroidal (Sph)(Sph) and irregular (Ir)(Ir) types. Inaccuracies in classification can cause these hybrids to “jump” from one extreme of the T-scale to the other with only a small error in morphology.

To address this, van den Bergh suggested dividing dwarf galaxies by luminosity class, which prevents them from leaping between extreme categories. This refined approach better reflects the intrinsic diversity of dwarf systems, ensuring that subtle morphological differences are more accurately captured and reducing the risk of placing galaxies at the wrong end of the classification scale (Karachentsev, Makarov, and Kaisina 2013)

1.2 Star-Forming Regions

One of the main ingredients of the galaxies are large scale molecular clouds, rich in hydrogen, with masses of order $10^5 M_{\odot}$, typical dimensions of ~ 10 parsec, temperatures of 10-100 K and densities of 10-300 molecules/cm³ (Pols 2009). Stars are created inside these clouds when a perturbation disturbs them, and thus their pressure equilibrium, and they start to collapse into smaller clouds under their self-gravity.

The collapse leads to the formation of protostars, which eventually become main-sequence stars. The presence of dust within these clouds is crucial, as it shields the interior regions from ultraviolet radiation, allowing the gas to cool to temperatures below 100 K, facilitating star formation .

The location of the star forming regions within a galaxy depend on the morphology. In spiral galaxies, star formation mainly occurs along the disk, where the molecular clouds are dense due to the compression caused by spiral density waves. The compression not only initiates the

collapse of the clouds and thus the star formation, but also feeds the arms with gas, which sustains the star formation.

In elliptical galaxies, on the other hand, star formation is minimal because they lack the cold gas reservoirs needed for new stars to form. Irregular galaxies, with their chaotic structures, often have patchy but vigorous star-forming regions, as they retain significant amounts of gas.

A special case are the starburst galaxies, which have extremely active star-forming regions, and seem to convert the gas into stars extremely fast (even 100 times faster than the Milky Way). These starbursts often concentrate their intense star-formation activity in compact regions about 1 kpc in size (typically in galaxy nuclei). Due to their high star-formation rates, starbursts host large numbers of young stars.

2 Star Formation History (SFH)

The SFH of a galaxy describes the evolution of its star formation rate over time. By selecting an appropriate model for SFH, we can analyze stellar production, predict periods of active or quiescent star formation, and determine when SFR stabilizes.

Understanding SFH models is crucial for interpreting internal and external processes affecting galaxies and identifying conditions for intense star formation in their early stages.

2.1 Star Formation Rate

The star formation rate (SFR) is defined as the total gas mass of a galaxy converted into stars over a specific time interval. It is typically expressed in solar masses per year ($M_{\odot} \cdot \text{yr}^{-1}$).

The SFR varies significantly over time, and its integration over time provides the total stellar mass formed during the galaxy's history of star formation. Specifically:

$$\int_0^{t_{sf}} \text{SFR}(t) dt = \zeta M_*(t_{sf}), \quad t_{sf} = \text{Time of Star Formation}, \quad (1)$$

where ζ accounts for mass loss during the Star Formation and is approximately $\zeta \approx 1.3$ (Kroupa et al. (2020)).

It is also important to define the Star Formation Density (SFRD):

$$\text{SFRD} = \frac{1}{V} \sum_{i=1}^N \text{SFR}_{o,i},$$

where V is the comoving volume ² of a region of the universe, N is the number of galaxies of the regions and $\text{SFR}_{0,i}$ is the current SFR of each galaxy i .

For the Local Cosmological Volume, which we will focus on, the volume is the sphere with its center at the Milky Way and a radius of 11 Mpc $V = \frac{4}{3}\pi(11 \text{ Mpc})^3 \approx 5575.3 \text{ Mpc}^3$.

2.1.1 Estimating SFR from Spectra

SFR can be estimated using various photometric or spectroscopic methods based on the luminosity of at least one spectral band or the intensity of a spectral line. Different luminosities and intensities trace distinct emission mechanisms, offering insights into a galaxy's radiation sources. Below are common methods:³

- **H α Emission:** Young, hot, massive stars (O-type stars, ~ 10 Myr, $\sim 20 M_{\odot}$) produce a number of ionizing photons, which they ionize the surrounding hydrogen rich gas. The hydrogen undergoes recombination cascades which produce Balmer emission lines of $H\alpha$ ($0.6563 \mu\text{m}$) and $H\beta$ ($0.4861 \mu\text{m}$). Dust can significantly affect observations.
- **Far-Ultraviolet (FUV) Flux:** Mainly emitted by young, hot stars (B-type stars, ~ 100 Myr). Dust presence can also significantly affect observations.
- **Infrared (IR) Flux:** The stars in a galaxy can heat up the dust in different ways, which then emits radiation in different parts of the IR spectrum. For example, young and massive, short-lived stars, emit UV radiation which then the heated dust emits in a wavelength of $\approx 60 \mu\text{m}$, whereas dust heated by UV-faint old or low-mass stars will emit at $\approx 100 - 150 \mu\text{m}$. As a result, the total IR emission is age-agnostic and provides a more accurate approximation of the SFR because it accounts for contributions from both young and old stellar populations.
- **Radio Continuum Emission:** Strongly correlated with IR. Its origin is complex, involving synchrotron radiation from relativistic electrons and thermal Bremsstrahlung from hot gas.
- **X-Ray Emission:** In star-forming galaxies, X-rays arise from high-mass binary systems (neutron star or black hole with massive stellar companion) and hot gas from supernovae, correlating with SFR up to redshift $z \sim 4$. X-rays are dust-insensitive, enabling accurate high-redshift observations.

SFR for different luminosities L_i can be calculated as:

$$\text{SFR}_i = \mathcal{K}_i \times L_i \quad (2)$$

²Comoving volume factors out the expansion of the universe, it remains a constant measure of “space” over time, making it possible to compare star formation at different epochs

³Calzetti (2012), Mushotzky (2017)

where \mathcal{K}_i is a constant specific to each L_i ($i = \text{H}\alpha, \text{IR}, \text{radio}, \text{FUV}, \text{X}$). In our analysis, we lack radio and X-ray data.

Since the luminosities L_{FUV} and $L_{\text{H}\alpha}$ originate from young stars and are highly sensitive to dust, we either directly observe stars unaffected by dust or use correction models to account for dust absorption. It is crucial to ensure that these models neither underestimate nor overestimate the luminosities by overlooking or double-counting the same sources.

Additionally, because these luminosities are emitted by similar stellar populations, we can reasonably expect the SFR_{FUV} and $SFR_{\text{H}\alpha}$ to be approximately equal. As shown in the data from (Karachentsev and Kaisina 2013) and supported by (Kroupa et al. 2020), a suitable approach for estimating the total SFR from FUV and H α observations is to calculate their average:

$$SFR_{\text{FUV},H\alpha} = \text{mean}(SFR_{\text{FUV}}, SFR_{\text{H}\alpha}) \quad (3)$$

where L_{FUV} and $L_{\text{H}\alpha}$ are corrected for dust attenuation.

According to (Madau and Dickinson 2014), this method often underestimates the SFR, since different galaxy populations may systematically follow distinct absorption mechanisms depending on their characteristics.

Since SFR_{FUV} , based on the uncorrected L_{FUV} , represents the emission from unobstructed stellar populations, and SFR_{TIR} accounts for dust-reprocessed light, a more accurate way to calculate the total SFR of a galaxy is:

$$SFR_{\text{total}} = \mathcal{K}_{\text{FUV}} \cdot L_{\text{FUV}} + \mathcal{K}_{\text{IR}} \cdot L_{\text{IR}} \quad (4)$$

Following the same reasoning as the previous formula, the total SFR can be expressed as:

$$SFR_{\text{total}} = \text{mean}(\mathcal{K}_{\text{FUV}} \cdot L_{\text{FUV}}, \mathcal{K}_{\text{H}\alpha} \cdot L_{\text{H}\alpha}) + \mathcal{K}_{\text{IR}} \cdot L_{\text{IR}} = SFR_{\text{FUV},H\alpha} + SFR_{\text{IR}} \quad (5)$$

where L_{FUV} and $L_{\text{H}\alpha}$ are not corrected for dust absorption. However, since we do not have enough galaxies with both traces, we will use a different method of calculating the total SFR, which we will discuss later.

2.2 Main Sequence Galaxies

The SFR and stellar mass of a galaxy are tightly correlated by the relationship:

$$\log(\text{SFR}) = \alpha \log(M_*) + \beta$$

where $\alpha(t)$ and $\beta(t)$ depend on time and redshift z (Speagle et al. (2014)):

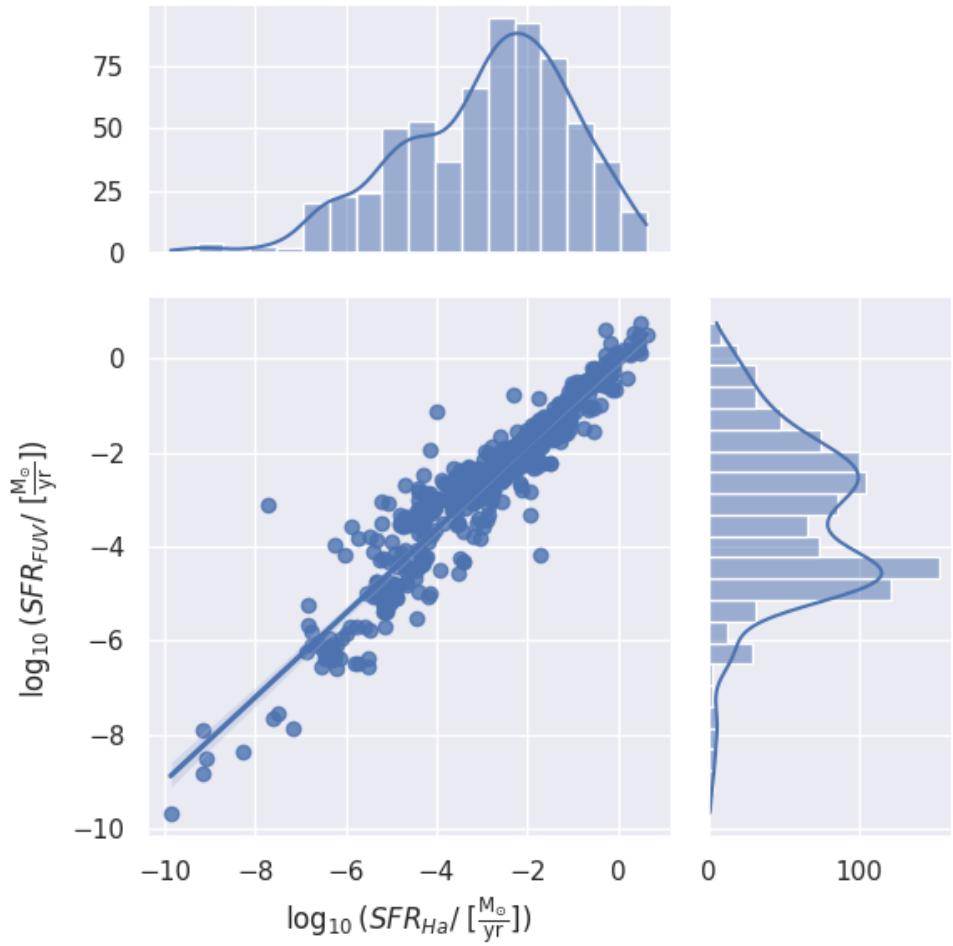


Figure 2: Plot showing the linear relation $\log_{10} \text{SFR}_{\text{FUV}} = \log_{10} \text{SFR}_{\text{H}\alpha}$, as well as their distributions, based on the data from UNGC

2.2.1 Star Formation History Models

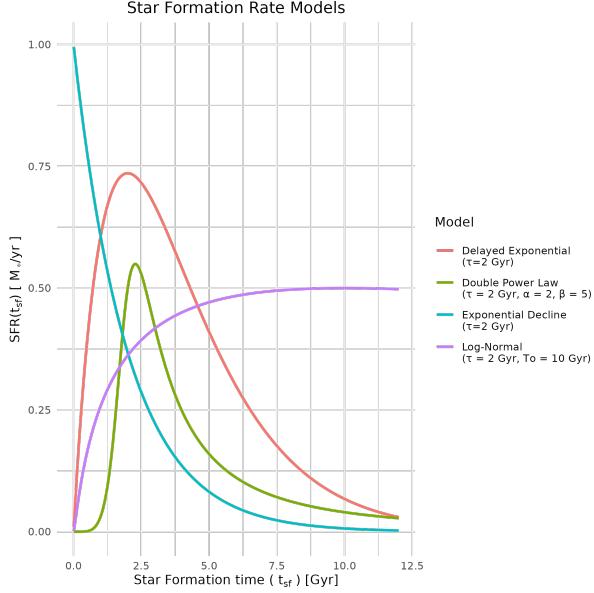


Figure 3: Star Formation Rate over the time of star formation, for different parametric models

Parameterized SFH models are commonly used, offering simplicity through a few parameters (Carnall et al. 2019):

- **Exponential Decline (Tau Model):** The star formation rate (SFR) decreases exponentially over time, following the equation:

$$\text{SFR}(t) \propto e^{-t_{\text{sf}}/\tau}$$

where τ is the timescale, $t_{\text{sf}} = t - T_0$ is the star formation time, t is the age of the Universe, and T_0 is the time when star formation began.

- **Delayed Exponential (Delayed Tau Model):** This model provides a more complex representation where the SFR initially increases, reaches a peak, and then declines exponentially over time. The equation for this model is:

$$\text{SFR}(t) \propto t_{\text{sf}} e^{-t_{\text{sf}}/\tau}$$

This accounts for an initial growth phase followed by a decline. In this case, τ represents the time it takes for the galaxy to reach SFR_{max} .

- **Log-Normal Distribution Model:** The SFR follows a normalized log-normal distribution, which can accurately model the star formation rate density ($\text{SFRD} = \text{SFR}/M_*$) in individual galaxies. The general form of the equation is:

$$\text{SFR}(t) \propto \frac{1}{\tau} \exp \left(-\frac{(\ln(t) - T_0)^2}{2\tau^2} \right)$$

where τ and T_0 are free parameters of the distribution that lack physical significance, as the SFR does not necessarily peak at $t = e^{T_0}$.

- **Double Power Law:** This model describes a scenario where the SFR rises and then falls sharply, useful for modeling galaxies experiencing rapid changes in star formation. The equation is:

$$\text{SFR}(t) \propto \left[\left(\frac{t}{\tau} \right)^\alpha + \left(\frac{t}{\tau} \right)^\beta \right]^{-1}$$

where τ is the timescale and α, β are exponents that govern the rise and fall of the SFR.

Additionally, there are non-parametric models, which do not follow a specific functional form to describe the star formation of a galaxy. These models are more flexible in adapting to galaxies with more complex star formation patterns.

3 Lilly-Madau Plot and Delayed- τ model conflicts

The Lilly-Madau plot is one of the most important plots in the field of galaxy evolution. It describes how the SFRD of the universe evolved over time, with observational data. But to understand it, we first need to understand how the observed age of the Universe and the redshifts of galaxies are related.

3.1 Redshift and lookback time

According to Hubble–Lemaître law, all the galaxies are moving away from each other, at a speed proportional to their distance, due to the expansion of the universe.

$V = H_0 \times d$, where $H_0 \approx 69.8 \text{ km/s/Mpc}$ is the Hubble constant and d is the distance between the two galaxies.⁴

⁴We also have non-Hubble motions $V = H_0 \times d + V_0$, where V_0 is the peculiar velocity and it could, for example, be due to galaxy cluster dynamics. For the current explanation we are going to ignore it. This way the radial velocity v is equal to V

Since we have galaxies with relative motions emitting light waves, we can observe the Doppler effect. Specifically, since the galaxies are moving away from each other, and thus from us also, we observe radiation with longer wavelenghts.

Redshift (z) is the doppler shift resulting from radial motion:

$$z = \frac{\lambda_{\text{observed}}}{\lambda_{\text{emitted}}} - 1$$

In special relativity, z is related to radial velocity v by (Hogg 2000)

$$1 + z = \sqrt{\frac{1 + v/c}{1 - v/c}} \quad (6)$$

For small v/c we can rewrite Equation 6, as:

$$z \approx \frac{V}{c} = \frac{H_0 \times d}{c}$$

But, because light takes time to cover the distance d between two galaxies, when the light finally reaches us, we will see the observed galaxy, as it was when the light was emitted, and not how it is at this moment. If we substitute time that it took the light to reach us over the distance, then we arrive at the relation (Longair 1998):

$$t_{\text{emitted}} \propto z^{-3/2}$$

The lookback time is the difference between the current age of the Universe and the age of the Universe when the light was emitted

$$t_L = T_0 - t_{\text{emitted}}$$

3.2 Lilly-Madau Plot

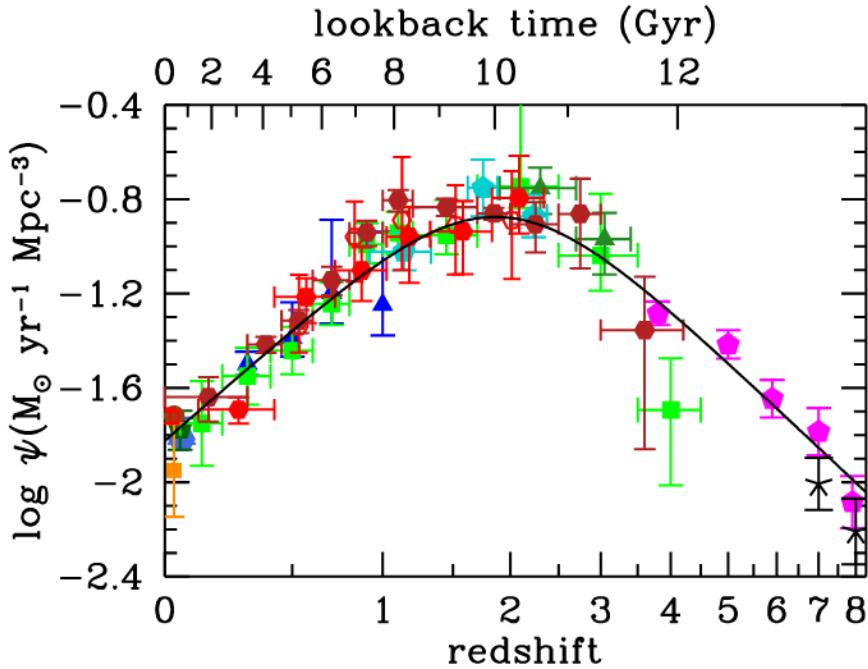


Figure 4: The history of cosmic star formation as shown in Madau and Dickinson (2014)

Using observational data from UV and IR traces for the SFR for galaxies of different redshifts (z) Madau and Dickinson (2014) were able to calculate the cosmic SFRD, at different ages of the universe, and reveals three periods of the universe.

- **Rising Era** ($\text{SFRD} \propto (1+z)^{-2.9}$, $3 \lesssim z \lesssim 8$): The SFRD is relatively low at very high redshifts (e.g., $z > 6$) and then rises as we move closer to $z \sim 2$. This indicates that early galaxies were ramping up their star-forming activity.
- **Peak Star Formation** ($z \sim 2$): Often dubbed “cosmic noon,” this epoch exhibits the highest SFRD in cosmic history when the Universe was ~ 3.5 Gyr old. Galaxies are vigorously converting gas into stars, and many of the most massive galaxy systems today formed much of their stellar content during this phase.
- **Decline to Present Day** ($\text{SFRD} \propto (1+z)^{2.7}$, $z \lesssim 2$): After $z \sim 2$, the global SFRD drops by about an order of magnitude towards the present day ($z = 0$).

The best-fit function of the comoving SFRD(z):

$$SFRD(z) = 0.015 \frac{(1+z)^{2.7}}{1 + \left(\frac{1+z}{2.9}\right)^{5.6}} \quad [M_{\odot} \text{ yr}^{-1} \text{ Mpc}^{-3}],$$

3.3 Delayed- τ model

The delayed τ model is widely used for describing an initial starburst followed by a gradual decline in SFR. This places galaxies on the main sequence. It is particularly effective for massive galaxies (Haslbauer, Kroupa, and Jerabkova (2023)). However, it assumes smooth SFR evolution and may overestimate peak SFR in high-redshift galaxies.

Using the delayed- τ model, we compute τ , t_{sf} , and normalization constant A_{del} with:

$$\text{SFR}_0 = \text{SFR}(t_{sf}) = A_{\text{del}} \frac{t_{sf}}{\tau^2} e^{-t_{sf}/\tau} \quad (7)$$

where SFR_0 is given in the catalogs. According to these model the timescale, at which the SFR peaks, is around $3.5 \lesssim \tau \lesssim 4.5$ (Speagle et al. 2014). If we intergrate the $\text{SFR}(t)$, we can calculate the present day averaged SFR:

$$\overline{\text{SFR}_{\text{del}}} = \frac{1}{t_{sf}} \int_{t_{\text{start}}}^{t_{\text{end}}} \text{SFR}_{\text{del}}(t) dt = \frac{A_{\text{del}}}{t_{sf}} \left[1 - \left(1 + \frac{t_{sf}}{\tau} \right) \exp \left(-\frac{t_{sf}}{\tau} \right) \right] \quad (8)$$

Where $\overline{\text{SFR}}$ can be calculated by the stellar masses of the galaxy of interest:

$$\overline{\text{SFR}_{\text{del}}} = \frac{\zeta M_*}{t_{sf}} \quad (9)$$

and ζ accounts for the mass-loss through stellar evolution, and is $\zeta \approx 1.3$ (Kroupa et al. 2020).

Using these two equations Haslbauer, Kroupa, and Jerabkova (2023) calculated the parameters τ , A_{del} of the model, for a constant t_{sf} for all the galaxies, and thus they were able to approximate Cosmic SFRD of the LV. They found that the SFHs implied from the LV galaxies systematically underestimate the SFRDs at $z \lesssim 3$ and overestimates the SFRDs at $z \gtrsim 6$ as in comparison to those of MD14. In particular, the SFRD at the peak of the Lilly-Madau plot is 2.16 ± 0.32 lower in the case of the delayed- τ . However, since they assume that all the galaxies begin at the same time and their sample size is not representative of the observed Universe, better calculations could potentially fix these discrepancies. However, these parametrization of the SFH could also be inadequate to describe galaxy evolution, especially for galaxies with $M_* < 10^{10} M_{\odot}$, as discussed in Haslbauer, Kroupa, and Jerabkova (2023).

4 Computational Methods

4.1 Newton-Raphson

The Newton-Raphson method is one of the most common numerical methods for estimating the root of a real-valued function. This method is based on the concept of linear approximation, where the function is locally approximated by its tangent line. It starts with an initial guess of the root and iteratively refines the result using a formula that involves derivative of the function.

Let x_n be the n-th approximation of the root and $f(x)$, $f'(x)$ the function and the derivative, respectively. Then the refined approximation of the root x_{n+1} is given by the formula

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n \in \mathbb{N}$$

The iteration stops when $f(x_n) = 0$, or after it reaches a certain number of iterations. If the initial estimation x_0 is close to the actual root, then the accuracy of the approximation increases rapidly with each iteration. If, however, the initial guess is poor, the model may fail to converge, which occurs when

$$\left| \frac{f(x_n)f''(x_n)}{f'(x_n)^2} \right| > 1.$$

Therefore, while the Newton-Raphson method is a powerful tool in numerical analysis, due to its efficiency and ease of use, it is reliant on a good initial guess which is not always possible.

4.2 Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) is a computer–driven sampling method which combines two properties: *Monte–Carlo* and *Markov chain*

A *Markov Chain* is a type of stochastic process that represents a sequence of events where the likelihood of each event depends solely on the state reached in the preceding event. In other words the future of each state depends only on the current state and not the past (Andrieu et al. 2003).

$$\Pr(X_{n+1} = x \mid X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \Pr(X_{n+1} = x \mid X_n = x_n),$$

if both conditional probabilities are well defined, that is, if $\Pr(X_1 = x_1, \dots, X_n = x_n) > 0$.

A great example of such a chain is the board game “Snake and Ladders”. The dice has a uniform probability distribution across 6 stages (integers 1 to 6). The players have a position on the board, but their next position on the board is only based on the current position and the random roll of the dice. The one dimension random walk algorithm is also a Markov Chain.

The *Monte Carlo* method is a computational method that uses random values from a given probability distribution to approximate the features of the given distribution (van Ravenzwaaij, Cassey, and Brown 2018). For example, if we want to calculate the mean of a normal distribution we can draw a large number of random samples from the distribution, and calculate the sample mean of those, instead of calculating it directly.

MCMC combines the properties of Markov chains and Monte Carlo, allowing the approximation of aspects of posterior distributions that cannot be directly calculated.

Let $x^{(i)}$ be a homogeneous Markov chain⁵ inside a high-dimensional space X of the set of possible configurations of a system and the space on which the posterior is defined, $x^{(i)} \in X = x_1, x_2, \dots$.

The chain is described by the formula:

$$p(x^{(i)} | x^{(i-1)}, \dots, x^{(1)}) = T(x^{(i)} | x^{(i-1)}),$$

where T is the transition matrix and p is the probability, following Bayes’ rule $p(x|D) \propto p(D|x) \cdot p(x)$ (D indicates the data). We chose T in such a way that its stationary distribution is the posterior p :

$$p(x^{(i)}) = \sum_{x^{(i-1)}} p(x^{(i-1)}) T(x^{(i)} | x^{(i-1)}).$$

T should be *irreducible*, so any state of the chain has a positive probability of visiting all the other states, meaning that T cannot be broken into smaller tables, and *aperiodic*, so the chain does not get trapped in cycles.

In continuous state spaces, the transition matrix T becomes an integral kernel K and $p(x)$ becomes the corresponding eigenfunction

$$p(x^{(i)}) = \int p(x^{(i-1)}) K(x^{(i)} | x^{(i-1)}) dx^{(i-1)}.$$

K is the conditional density of $x^{(i)}$, and it is chosen by the corresponding MCMC algorithm, such as the Metropolis-Hastings algorithm.

⁵The chain is homogeneous if $T \forall x^{(i)}$ remains invariant (the probability of moving from $x^{(i-1)}$ to $x^{(i)}$ does not depend on the iteration number i) and $\sum_{x^{(i)}} T(x^{(j)} | x^{(i-1)}) = 1$.

In practice, this means that given a prior distribution for the parameters of the examined model (in our case we assume priors for the parameters t_{sf} , τ , A_{del} from the delayed- τ model) and the likelihood based on observed data (we have experimental data for the SFR and M_*), a MCMC algorithm can calculate the posterior probability of the model and thus approximate the distribution of the parameters. From these distributions, their mean values are the values of the parameters and their variance is the error.

To calculate these, the algorithm firstly initializes some chains in the parameter space (t_{sf} , τ , A_{del}) and begins sampling the posterior distribution based on the values of the parameters. The first iterations are discarded, since if the sampling begins in a low probability region it can significantly affect the variance of results (burn-in period)(Andrieu et al. 2003). Eventually, the chains should converge. To examine this we use the diagnostics $\hat{R} < 1.01$ and ESS>100. The diagnostics indicate if different chains have converged to the same posterior region and if the sample of draws autocorrelates, respectively.⁶

5 Data of the Local Cosmological Volume

In our current analysis, we will use two different catalogs of galaxies, containing galaxies within the Local Cosmological Volume (LV), which is defined as a sphere with a radius of 11 Mpc centred around the Milky Way, to remain consistent with the analysis of Kroupa et al. (2020) and Haslbauer, Kroupa, and Jerabkova (2023). The reason we focus on the LV and thus only at small distances is because we want to understand the SFR of the galaxies as they are today and since the redshift depends on distance, we can assume that the redshift of our galaxy sample is $z = 0 \Rightarrow t_{\text{emitted}} = T_0 = 13.8$ Gyr.

The observational data of galaxies located inside the LV are extracted from the catalogs:

- **HECATE** (Kovlakas et al. 2021): The Heraklion Extragalactic Catalogue (HECATE) is an all-sky value-added galaxy catalog containing 204,733 galaxies with redshifts up to 0.047 ($D \leq 200$ Mpc), from which 2901 are in the LV. It incorporates data from multiple surveys and databases, providing information on positions, sizes, distances, morphological classifications, star formation rates, stellar masses, metallicities, and nuclear activity classifications. HECATE is well-suited for multi-wavelength and multi-messenger astrophysics, including demographic studies and transient event follow-up.
- **UNGC Catalog** (Karachentsev, Makarov, and Kaisina 2013; Karachentsev and Kaisina 2013): The Updated Nearby Galaxy Catalog (UNGC)⁷ provides a distance-limited sample of galaxies, within 11 Mpc or radial velocities of $V < 600 \text{ km} \cdot \text{s}^{-1}$. Selecting galaxies with $D \leq 11$ Mpc, we get a sample of 1321 galaxies. It offers detailed star formation rate

⁶For more information about the metrics used by the Stan software you can visit https://mc-stan.org/docs/2_36/reference-manual/analysis.html#effective-sample-size.section

⁷The UNGC tables can be downloaded using the following link: <https://www.sao.ru/lv/lvgdb/introduction.php>. Here, we use the latest update from 17.01.2025.

estimates derived from H α and far-ultraviolet (FUV) fluxes, making it particularly valuable for studying nearby galaxies and local star formation processes. The UNGC catalog focuses on galaxies with well-determined distances and includes detailed environmental and morphological data.

We will join the two catalogs, based on the coordinates of the galaxies, so we could assure a bigger and more complete sample size for our analysis. But first we need to understand them and examine if there are any inconsistencies between UNGC and HECATE.

5.1 Catalog Completeness

The Completeness of a catalog refers to the extent to which a catalog includes all the galaxies within a specified region of the sky, distance, or luminosity range. A complete catalog accurately represents the true population of galaxies, which eliminates the bias of the sample. No catalog could truly be 100% complete, but by understanding the shortcomings of the catalog, we can account for the possible biases that can occur in our analysis.

The *HECATE* catalog is designed to be a more comprehensive resource, covering a broader range of distances and luminosities.

- **Luminosity and Distance Dependence:** The HECATE catalog is more complete for brighter galaxies (with luminosities above $10^{9.5} L_{B,\odot}$) and distances within 33 Mpc, which includes the LV. For galaxies with luminosities around $10^{10} L_{B,\odot}$, the catalog remains complete up to distances of 100 Mpc.
- **Luminosity Density:** The catalog's completeness is also estimated based on the luminosity density in the B-band. Within 30 Mpc, the catalog is over 100% complete due to the high density of galaxies around the Milky Way. However, at distances around 170 Mpc, the completeness drops to approximately 50%, indicating that many galaxies are missing.
- **Star Formation Rate and Stellar Mass:** The completeness of SFR is around 50% for distances between 30 and 150 Mpc, primarily due to limitations in the WISE survey coverage. Similarly, the completeness of stellar mass estimates is comparable to that of the B-band luminosity, with overdensities at short distances and a sharp cutoff at larger distances.

The *UNGC* catalog focuses on galaxies within the LV, specifically those with radial velocities less than 600 km/s and distances within 11 Mpc. It is around 40%–60% complete for galaxies with B-band magnitude $M_B^c < -11^m$. In the Local Group ($D \leq 10$ Mpc) only half the galaxies have $M_B^c > -11^m$ and it is expected that for distances ~ 9 Mpc more than half of the ultra-faint galaxies are missing. Specifically, it is estimated that around 10^3 – 10^4 galaxies of the LV are not included in UNGC.

The incompleteness for ultra-faint galaxies is primarily due to their low surface brightness, making them difficult to detect beyond the Local Group. This limitation is particularly significant for studies of dwarf galaxies and the faint end of the galaxy luminosity function.

5.2 Comparing the Catalogs

Before we merge the two catalogs we need to compare the corresponding quantities of the 288 common galaxies to ensure that there are no major incompatabilities. The quantities that are comparable are:

- Distances [Mpc] (288/288 common galaxies)
- Radial Velocities [km/s] (286/288 common galaxies)
- Morphology and Geometry of the galaxies
 - Morphological type (229/288 common galaxies)
 - Inclination [degrees] (209/288 common galaxies)
 - Angular Diameter of the Major and Semi-major axis [arcmin] (261/288 common galaxies)
- K-band Luminosity [L_{\odot}] and Magnitude (70/288 common galaxies)
- B-band Magnitudes [mag] (244/288 common galaxies)
- Star Formation Rates [M_{\odot}/yr] (73/288 common galaxies)

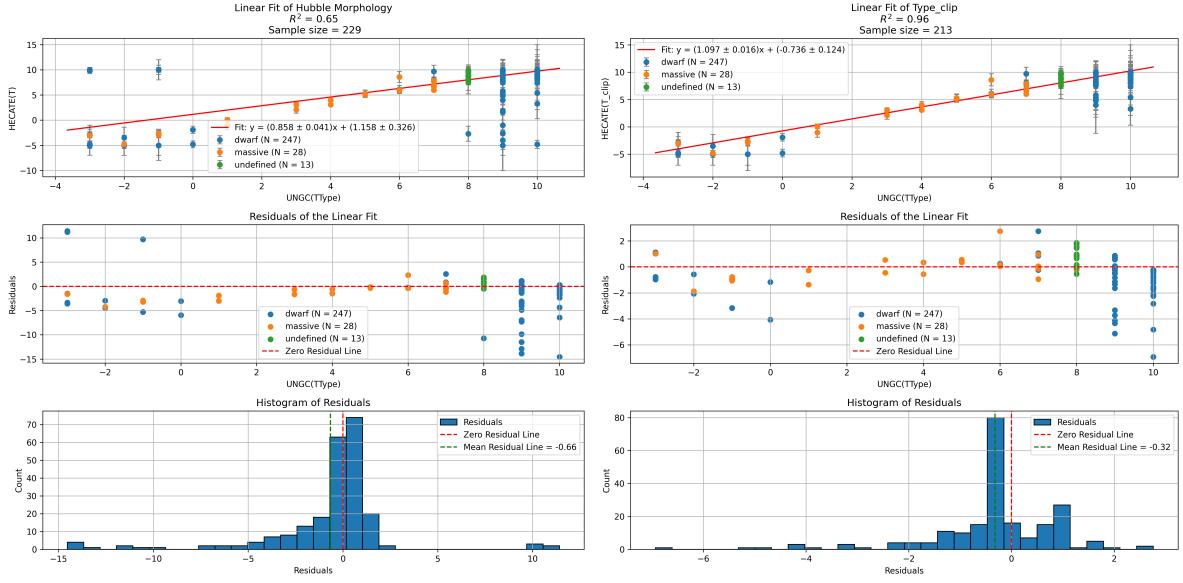
For the comparison of our data we will use linear regression, accounting for the errors of the observed data, where R^2 and the slope of the fitted line should be close to 1⁸. We will also plot the residuals ($y_{\text{Regression}} - y_{\text{observed}}$) and if the mean value $\overline{\text{Residuals}} \approx 0$, then the linear regression is considered succesfull.

All the results are shown in Section 7, except the comparisons of the Morphological Types and of the SFRs, that need more detailed analysis.⁹

To begin with the morphological types of the galaxies, we first need to address potential inconsistencies in the classifications of the two catalogs. As we can see in Figure 5a, some galaxies “jump” from one side of the classification to the other side, depending on the catalog. These galaxies are dwarf galaxies, so we can safely assume that they are Tr type galaxies, as discussed in Section 1.1.1. For the rest of the analysis we will adopt the Numerical Morphological Type from the HECATE catalog, since it includes a larger number of galaxies, ensuring better data compatibility. Additionally, the HECATE catalog boasts an accuracy of 0.1 and provides errors, while the UNGC provides a rigid categorization of the galaxies with no errors.

⁸ R^2 measures the proportion of variance explained by the linear model. For a model to be acceptable the R^2 should be higher than 80%, or close to it if $\overline{\text{Residuals}} \approx 0$

⁹The names of the x and y axes are written as they are given in the corresponding catalog, except the ones in decimal logarithmic scale, that are notated appropriately.



(a) Before the clipping of the transient types

(b) After the clipping of the transient types

Figure 5: Scatter plots and the residuals of the Hubble Numerical morphological types show a good correlation between the two datasets, if we ignore the Transient dwarf galaxies.

Next we need to analyze the SFR from the two catalogs. The catalogs provide different SFR estimators, $H\alpha$ and FUV from UNGC and the total IR from HECA. As already discussed in Section 2.1.1, the two estimators FUV and $H\alpha$ can be combined using their average ($\text{mean}(SFR_{FUV}, SFR_{H\alpha})$), however both estimators and their average can potentially underestimate the SFR due to dust attenuation (SFR_{TIR} is not affected by this), and this is exactly what is observed in Figure 6.

In reality those estimators compliment each other and this is exactly why they are not equal. The total SFR could potentially be calculated by Equation 5, however due to the lack of samples with both traces the calculation of SFR_{total} is not possible without introducing biases to our data. Due to the underestimation from the estimators of UNGC, the main estimator used in our analysis will be the TIR estimator and the $\text{mean}(SFR_{FUV}, SFR_{H\alpha})$ will be transformed into the expected SFR_{TIR} according to the linear regression of Figure 6.

After the merging of the two catalogs we get a final sample of 3934 galaxies out of which only 288 are included in both catalogs and where used for the comparison. 1033 are unique galaxies in the UNGC and the remaining 2613 are unique in the HECA catalog. Additionally, 1842 galaxies have SFR data (778 from HECA and 1137 from UNGC), and 3042 galaxies have Stellar Mass data (1605 from HECA, with the rest estimated using a mass-to-light ratio of 0.82 as provided by HECA). 1761 galaxies have both SFR and M_* data.

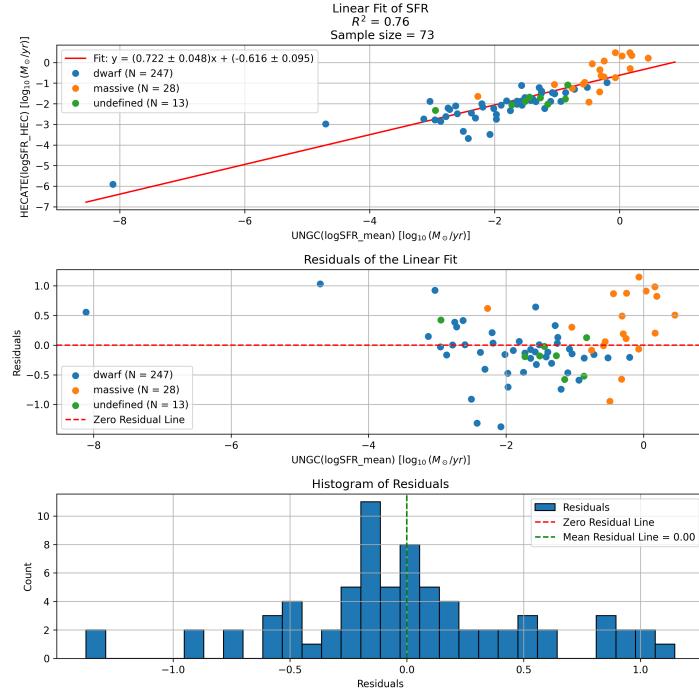


Figure 6: Scatter plot and the residuals of the mean(SFR_{FUV} , $SFR_{H\alpha}$) from the UNGC catalog and the SFR_{TIR} . The SFR estimation from the FUV and $H\alpha$ is systematically lower than the SFR_{TIR} estimator as expected. The R^2 is smaller than 80% however due to $\overline{\text{Residuals}} = 0$ we can assume that the quantities are comparable.

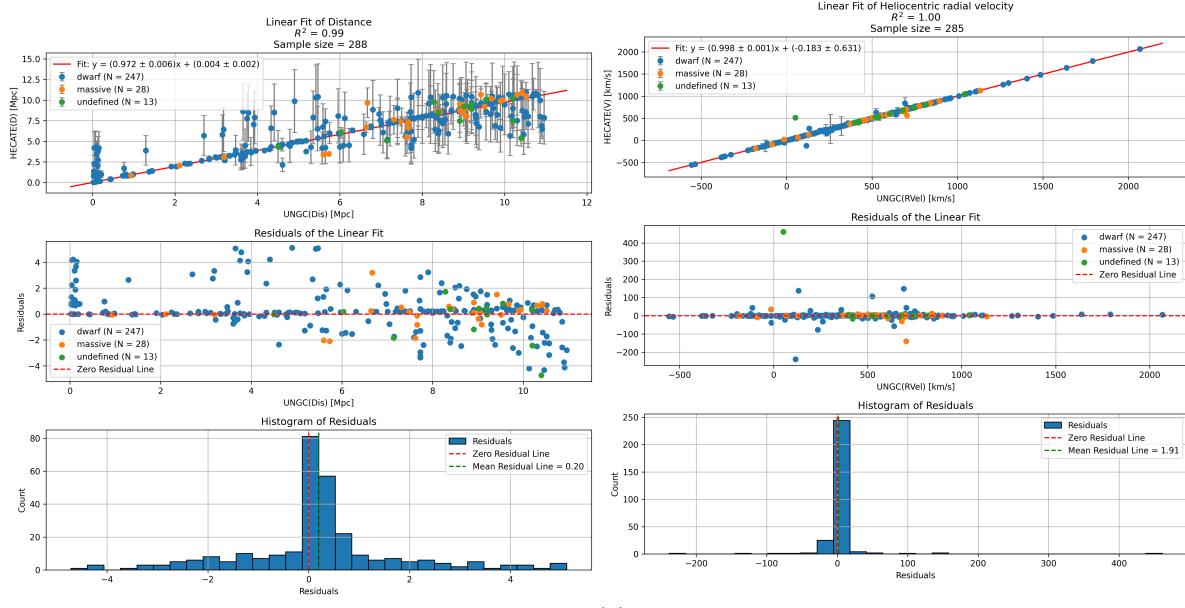
6 Calculating the parameters

6.1 According to pre-existing bibliography

6.1.1 Problems with the method

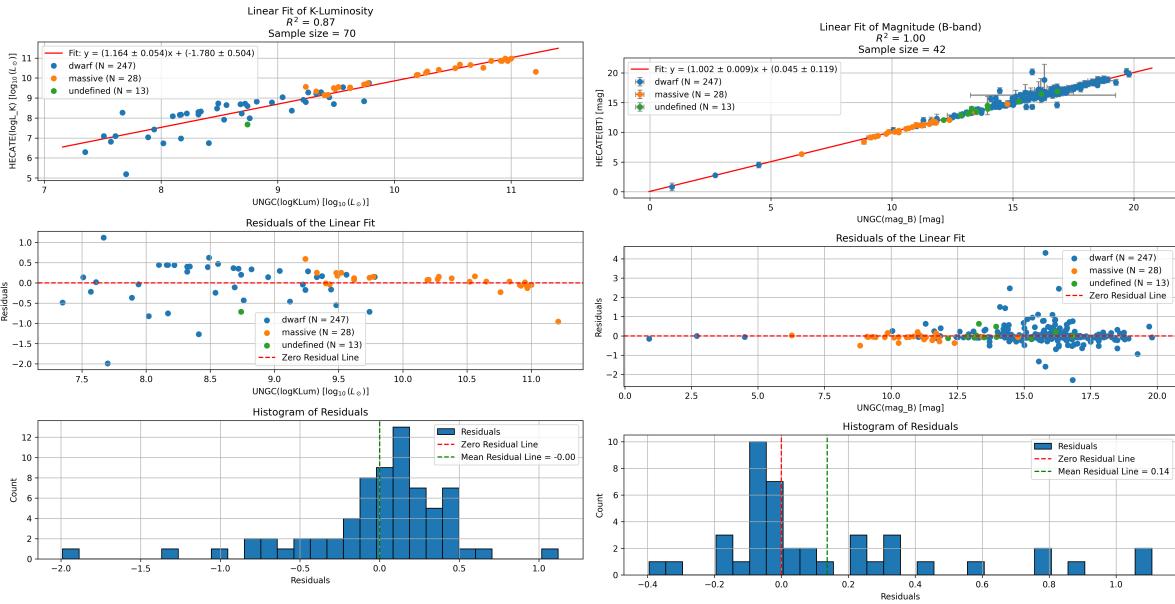
6.2 MCMC

7 Appendix A: Scatter plots of the comparisons



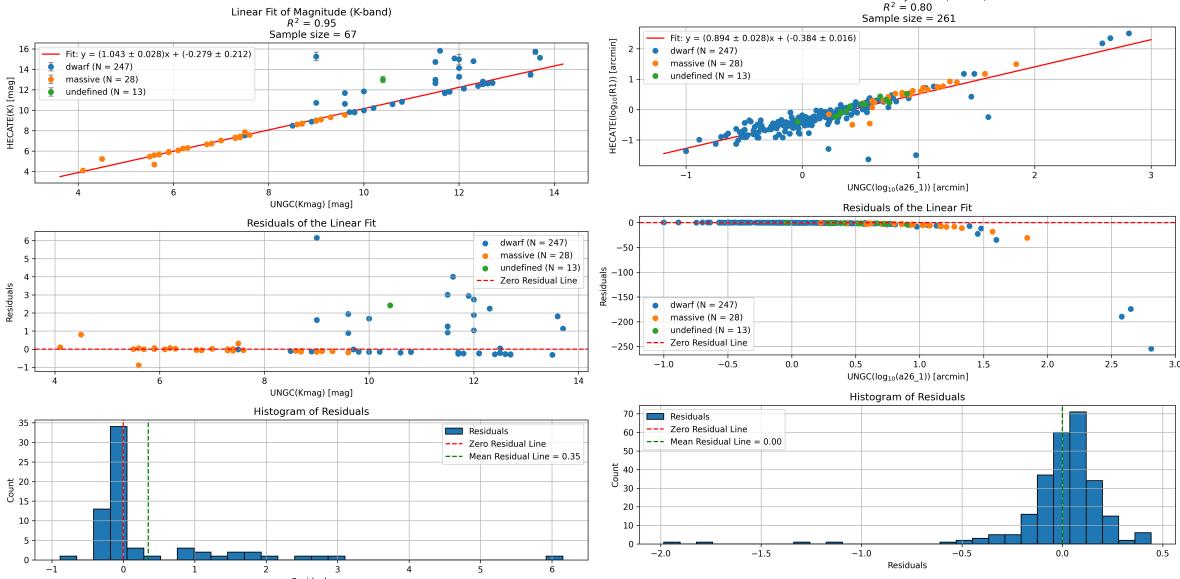
(a) Linear Regression of the Distances in Mpc

(a) Linear Regression of the Radial Velocities in km/s



(a) Linear Regression of the K-band luminosity in L_{\odot}

(b) Linear Regression of the B-Band Magnitude



(c) Linear Regression of the K-Band Magnitude

(d) Linear Regression of the Angular Diameter of the Major and Semi-major axis in arcminutes

References

- Andrieu, Christophe, Nando de Freitas, Arnaud Doucet, and Michael I. Jordan. 2003. “An Introduction to MCMC for Machine Learning.” *Machine Learning* 50 (1): 5–43. <https://doi.org/10.1023/A:1020281327116>.
- Calzetti, Daniela. 2012. “Star Formation Rate Indicators.” arXiv. <https://arxiv.org/abs/1208.2997>.
- Carnall, Adam C., Joel Leja, Benjamin D. Johnson, Ross J. McLure, James S. Dunlop, and Charlie Conroy. 2019. “How to Measure Galaxy Star Formation Histories. I. Parametric Models.” *The Astrophysical Journal* 873 (1): 44. <https://doi.org/10.3847/1538-4357/ab04a2>.
- de Vaucouleurs, Gerard. 1959. “Classification and Morphology of External Galaxies.” *Handbuch Der Physik* 53 (January): 275. https://doi.org/10.1007/978-3-642-45932-0_7.
- Haslbauer, Moritz, Pavel Kroupa, and Tereza Jerabkova. 2023. “The Cosmological Star Formation History from the Local Cosmological Volume of Galaxies and Constraints on the Matter Homogeneity.” *Monthly Notices of the Royal Astronomical Society* 524 (3): 3252–62. <https://doi.org/10.1093/mnras/stad1986>.
- Hogg, David W. 2000. “Distance Measures in Cosmology.” arXiv. <https://doi.org/10.48550/arXiv.astro-ph/9905116>.
- Hubble, E. P. 1925. “Cepheids in Spiral Nebulae.” *The Observatory* 48 (May): 139–42.
- Karachentsev, Igor D., and Elena I. Kaisina. 2013. “STAR FORMATION PROPERTIES IN THE LOCAL VOLUME GALAXIES VIA H α AND FAR-ULTRAVIOLET FLUXES.” *The Astronomical Journal* 146 (3): 46. <https://doi.org/10.1088/0004-6256/146/3/46>.
- Karachentsev, Igor D., Dmitry I. Makarov, and Elena I. Kaisina. 2013. “UPDATED NEARBY GALAXY CATALOG.” *The Astronomical Journal* 145 (4): 101. <https://doi.org/10.1088/0004-6256/145/4/101>.
- Kovlakas, K., A. Zezas, J. J. Andrews, A. Basu-Zych, T. Fragos, A. Hornschemeier, K. Kouroumpatzakis, B. Lehmer, and A. Ptak. 2021. “The Heraklion Extragalactic Catalogue (HECATE): A Value-Added Galaxy Catalogue for Multimessenger Astrophysics.” *Monthly Notices of the Royal Astronomical Society* 506 (September): 1896–1915. <https://doi.org/10.1093/mnras/stab1799>.
- Kroupa, P, M Haslbauer, I Banik, S T Nagesh, and J Pflamm-Altenburg. 2020. “Constraints on the Star Formation Histories of Galaxies in the Local Cosmological Volume.” *Monthly Notices of the Royal Astronomical Society* 497 (1): 37–43. <https://doi.org/10.1093/mnras/staa1851>.
- Leavitt, Henrietta S. 1907. “1777 Variables in the Magellanic Clouds.” *Annals of Harvard College Observatory* 60: 87–108.3.
- Longair, Malcolm S. 1998. *Galaxy Formation*. Springer Science & Business Media.
- Madau, Piero, and Mark Dickinson. 2014. “Cosmic Star Formation History.” *Annual Review of Astronomy and Astrophysics* 52 (1): 415–86. <https://doi.org/10.1146/annurev-astro-081811-125615>.
- Mushotzky, Richard. 2017. “ASTR620 Galaxies - Fall 2017.” Lectures. University of Mary-

land.

- Pols, Onno. 2009. *Stellar Structure and Evolution*. Utrecht University.
- “RC3 - Third Reference Catalog of Bright Galaxies.” n.d. <https://heasarc.gsfc.nasa.gov/w3browse/all/rc3.html>. Accessed December 28, 2024.
- Shapley, Harlow, and Heber D. Curtis. 1921. “The Scale of the Universe.” *Bulletin of the National Research Council* 2 (11): 171–217.
- Speagle, Joshua S., Charles L. Steinhardt, Peter L. Capak, and John D. Silverman. 2014. “A Highly Consistent Framework for the Evolution of the Star-Forming ”Main Sequence” from $z \sim 0\text{--}6$.” *The Astrophysical Journal Supplement Series* 214 (2): 15. <https://doi.org/10.1088/0067-0049/214/2/15>.
- van Ravenzwaaij, Don, Pete Cassey, and Scott D. Brown. 2018. “A Simple Introduction to Markov Chain Monte–Carlo Sampling.” *Psychonomic Bulletin & Review* 25 (1): 143–54. <https://doi.org/10.3758/s13423-016-1015-8>.