

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

Undergraduate thesis

Dimitrios Papachristopoulos¹ and Despina Hatzidimitriou²

1. Undergraduate student, ID number 1110 2018 00142
2. Supervising professor, National and Kapodistrian University of Athens



National and Kapodistrian University of Athens
School of Science
Faculty of Physics
Department of Astrophysics, Astronomy and Mechanics
April 2025

Table of contents

1 Abstract	2
1.1 English	2
1.2 Greek	2
2 Galaxies and Galaxy Morphology	7
2.1 Galaxy Classification	7
2.1.1 Dwarf Galaxies	9
3 Star Formation	10
3.1 Star-Forming Regions	10
3.2 Star Formation Rate	11
3.2.1 Estimating SFR from Spectra	11
3.3 Main Sequence Galaxies	14
3.4 Star Formation History Models	14
4 Lilly-Madau Plot and Delayed-τ model conflicts	16
4.1 Redshift and lookback time	16
4.2 Lilly-Madau Plot	17
4.3 Delayed- τ model	18
5 Computational Methods	20
5.1 Newton-Raphson	20
5.2 Markov Chain Monte Carlo	20
6 Data of the Local Cosmological Volume	22
6.1 Catalog Completeness	23
6.2 Comparing the Catalogs	24
7 Calculating the parameters	26
8 Conclusions	32
9 Appendix A: Scatter plots of the comparisons	33
References	35

1 Abstract

1.1 English

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 of galaxies as a function with a rising Star Formation Rate (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 use of Delayed- τ model for the galaxies of the Local Cosmological Volume (LCV) 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 observational data for SFR and galaxy 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, Newton-Ramphson, Markov Chain Monte Carlo (MCMC).

1.2 Greek

Οι γαλαξίες είναι τεράστια συστήματα από αστέρια, πλανήτες, αέρια και σκόνη που συγκρατούνται μαζί λόγω της βαρύτητας. Κάθε γαλαξίας αποτελεί ένα «κοσμικό νησί» με δισεκατομμύρια ή ακόμα και τρισεκατομμύρια αστέρια, πλανήτες, μαύρες τρύπες και άλλα αστρονομικά αντικείμενα. Κάθε γαλαξίας διαφέρει τόσο στο πόσο μαζικός είναι όσο και στην μορφολογία του. Συνεχώς, εντός των γαλαξιών, νεφελώματα τα οποία αποτελούνται από αέρια και σκόνη, καταρρέουν λόγω της βαρύτητας τους και παράγουν αστέρια. Η παραγωγή των άστρων επηρεάζεται από τα χαρακτηριστικά του εκάστοτε γαλαξία και μέσω της μελέτης της ιστορίας της αστρογένεσης των γαλαξιών, μπορούμε να καταλάβουμε την προέλευση των αστρικών και πλανητικών συστημάτων αλλά και την εξέλιξη τόσο των γαλαξιών όσο και του σύμπαντος.

Μέσω φασματοσκοπικών και φωτομετρικών μεθόδων μπορούμε να υπολογίσουμε των ρυθμό αστρογένεσης (Star Formation Rate, SFR) των γαλαξιών σε διάφορετικές ηλικίες τους (διαφορετικά φάσματα χρησιμοποιούνται για τον υπολογισμό διαφορετικών αστρικών πληθυσμών),

και με αυτόν τον τρόπο είναι δυνατόν να μάθουμε την ιστορία αστρογένεσης του κάθε γαλαξία (Star Formation History, SFH). Κάποια μοντέλα προσπαθούν να παραμετροποιήσουν την SFH, χρησιμοποιώντας συναρτήσεις της μορφής $SFR = f(t)$, προσπαθώντας να προσεγγίσουν την πραγματική SFH με μία ομαλή και εξιδανικευμένη εξίσωση που αγνοεί την ποικιλομορφία των γαλαξιών. Ένα από αυτά τα μοντέλα είναι το delayed- τ , το οποίο δίνεται από την συνάρτηση

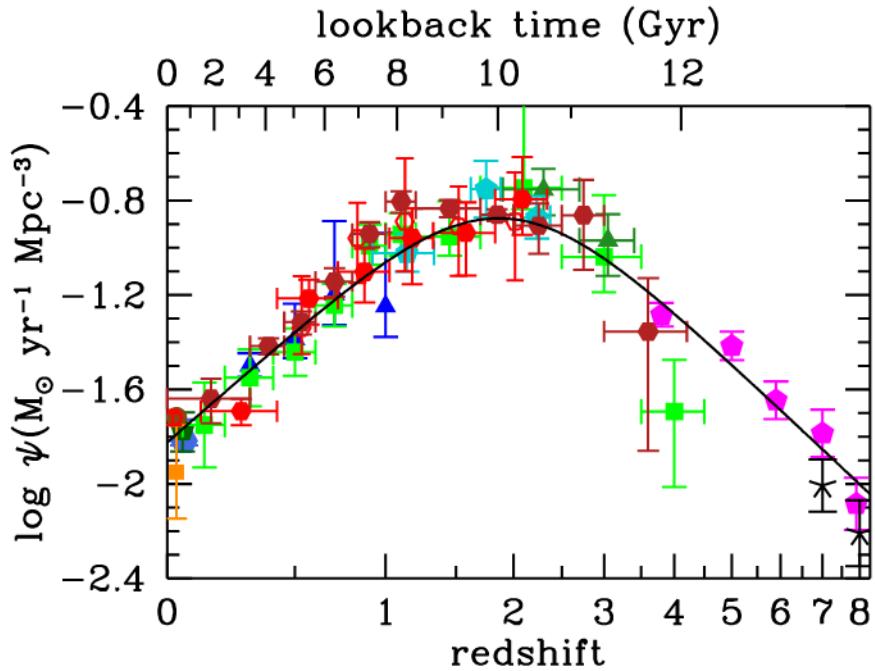
$$SFR(t_{sf}) = A_{del} \frac{t_{sf}}{\tau^2} e^{-t_{sf}/\tau}, \quad t_{sf} = T - t_{start},$$

όπου T είναι η ηλικία του σύμπαντος και t_{start} είναι ο χρόνος που ξεκίνησε η αστρογένεση, δηλαδή ο t_{sf} είναι το χρονικό διάστημα που ο γαλαξίας έχει αστρογένεση. Στην πραγματικότητα, αυτό το μοντέλο προβλέπει μια ομαλή αύξηση του SFR στην αρχή του γαλαξία, μέχρι που φτάνει στο μέγιστο όταν το t_{sf} είναι ίσο με τον χαρακτηριστικό χρόνο τ και έπειτα ο ρυθμός αστρογένεσης μειώνεται εκθετικά. Το A_{del} είναι μια σταθερά κανονικοποιησης, η οποία είναι ανεξάρτητη του χρόνου, αλλά είναι ανάλογη των αστρικών μαζών του εκάστοτε γαλαξία ($A_{del} \propto M_*$), και προκύπτει από το ολοκλήρωμα

$$\int_{t_{start}}^{t_{final}} SFR(t) dt \approx M_{*,current}$$

Η ηλικία T είναι στην πραγματικότητα η μεταβλητή της συνάρτησης, καθώς ο χρόνος που ξεκινάει η αστρογένεση είναι σταθερός. Επιπλέον, επειδή όσο πιο μακρινός είναι ένας γαλαξίας το φως χρειάζεται περισσότερο χρόνο για να φτάσει σε εμάς, όταν χρησιμοποιούμε αυτή την εξίσωση για μακρινούς γαλαξίες, πρέπει να χρησιμοποιούμε μικρότερο T από το σημερινό. Ταυτόχρονα όμως το σύμπαν διαστέλλεται και όλοι οι γαλαξίες απομακρύνονται από εμάς βάση του νόμου του Hubble $V = H_0 \times d$, όπου $H_0 \approx 69.8 \text{ km/s/Mpc}$ είναι η σταθερά το Hubble και d είναι η μεταξύ μας απόσταση. Επομένως το φως ερυθρομεταποίζεται, λόγω του φαινομένου Doppler, καθώς η πηγή του (γαλαξίας) απομακρύνεται όσο εκπέμπει, και για αυτό μπορούμε να χρησιμοποιήσουμε την κοσμική ερυθρομεταποίηση z στο μοντέλο στην θέση της ηλικίας του σύμπαντος.

Χρησιμοποιώντας παρατηρησιακά δεδομένα για το SFR γαλαξιών διάφορων z , μπορούμε να υπολογίσουμε την πυκνότητα του ρυθμού αστρογένεσης SFRD σε διαφορετικές ηλικίες του σύμπαντος. Σύμφωνα με τους Madau and Dickinson (2014), η SFRD του σύμπαντος μεταβάλλεται στον χρόνο βάση της παρακάτω γραφικής παράστασης στην οποία εμφανίζεται μια μέγιστη πυκνότητα σε $z = 1.86$.



Σχήμα 1: Διαγραμμα Lilly-Madau, που σχεδιάζει την πυκνότητα του ρυθμού αστρογένεσης του σύμπαντος, συναρτήσει του z

Οι Haslbauer, Kroupa, and Jerabkova (2023) χρησιμοποιώντας δεδομένα για το SFR και M_* γαλαξιών του Local Cosmological Volume (LV), δηλαδή για γαλαξίες με απόσταση $D < 11$ Mpc, από τον κατάλογο UNGC Karachentsev and Kaisina (2013) (1102 γαλαξίες), προσπάθησαν να αναπαραγάγουν την παραπάνω γραφική παράσταση του SFRD(z), αλλά βασισμένοι σε παραμετρικοποιημένα μοντέλα, ένα εκ των οποίων το delayed- τ . Για να μπορέσουν να το κάνουν, έπρεπε αρχικά να υπολογίσουν τις παραμέτρους του μοντέλου για τον κάθε γαλαξία. Θεωρώντας ότι όλοι οι γαλαξίες του LV ($z = 0 \Rightarrow T = T_0 = 13.8$ Gyr) ξεκίνησαν την αστρογένεση ταυτόχρονα πριν 13.6 Gyr, δηλαδή $t_{\text{start}} = 0.2$ Gyr, υπολόγισαν τα τ και A_{del} για κάθε γαλαξία. Έτσι μπόρεσαν να υπολογίσουν τις σταθερές της συνάρτησης SFR(z) για τον κάθε γαλαξία και κατ' επέκταση μπόρεσαν να υπολογίσουν την πυκνότητα

$$\text{SFRD}(z) = \frac{1}{V} \sum_i \text{SFR}_i(z)$$

Ακολουθώντας αυτή τη μέθοδο, παρατήρησαν ότι το delayed- τ παρουσιάζει ασυμφωνίες με το Lilly-Madau Plot. Συγκεκριμένα, υποστηρίζουν ότι εφαρμόζοντας το συγκεκριμένο μοντέλο για του γαλαξιών του LV, οδηγεί σε μειωμένη πυκνότητα συγκριτικά με αυτή που θα περιμέναμε, και ειδικότερα για $z = 1.86$ το μοντέλο υποτιμά την πυκνότητα κατά 2.16 ± 0.33 .

Σε αυτή την πτυχιακή, θα προσπαθήσουμε να βελτιώσουμε την παραπάνω μέθοδο υπολογισμού τους SFRD για διαφορετικά z , βασισμένοι στο ίδιο μοντέλο, αλλά χρησιμοποιώντας μεγαλύτερο δείγμα γαλαξιών και διαφορετικές υπολογιστικές μεθόδους.

Αρχικά προσπαθούμε να βελτιώσουμε τα δεδομένα με τα οποία θα δουλέψουμε. Χρησιμοποιώντας τα δεδομένα από τον κατάλογο UNGC και τον κατάλογο της HECATE (Kovlakas et al. 2021) για τους γαλαξίες του LV, ανέζανουμε το δείγμα των γαλαξιών για τα οποία έχουμε μετρήσεις για τις αστρικές μάζες και το SFR, καθώς η πυκνότητα επηρεάζεται σημαντικά από τον αριθμό του δείγματος. Αφού ενώσουμε τους δύο καταλόγους βάση των συντεταγμένων των γαλαξιών, ελέγχουμε ότι οι δύο κατάλογοι δεν έχουν σημαντικές διαφορές μεταξύ τους και συγκρίνουμε τα SFR που δίνονται από αυτούς. Καθώς ο κατάλογος UNGC δίνει SFR μετρημένο από FUV και $H\alpha$ και η HECATE δίνει από IR, η γραμμική προσαρμογή των μετρήσεων δείχνει ότι υπάρχει συσχέτιση μεταξύ τους αλλά η γραμμική σχέση δεν είναι 1 προς 1 και για αυτό χρειάζεται να μετατρέψουμε κάποιες μετρήσεις, έτσι ώστε να είναι τα μεγέθη συγκρίσιμα. Επιλέγουμε να μετατρέψουμε τα $SFR_{FUV,H\alpha}$ σε SFR_{IR} , διότι η λίστα HECATE περιέχει περισσότερους γαλαξίες από την UNGC, επομένως η μετατροπή αυτή επηρεάζει λιγότερο τα αποτελέσματα μας. Το τελικό μας δείγμα αποτελείται από 1691 γαλαξίες, για τους οποίους έχουμε προσεγγίσεις και για το SFR_{IR} αλλά και για τις αστρικές μάζες.

Στην συνέχεια, ακολουθώντας την ίδια λογική με τους Haslbauer, Kroupa, and Jerabkova (2023), προσπαθούμε να υπολογίσουμε τις σταθερές του delayed- τ μοντέλου για κάθε γαλαξία του τελικού μας δείγματος. Χρησιμοποιώντας τα δεδομένα για τις σημερινές τιμές του SFR και M_* των γαλαξιών, θα χρησιμοποιήσουμε δύο διαφορετικές μεθόδους, την μέθοδο Newton-Raphson (N-R) και την μέθοδο Markov Chain Monte Carlo (MCMC).

Η N-R, είναι μια υπολογιστική μέθοδος που προβλέπει την επίλυση γραμμικών και μη συστημάτων. Το πρόβλημα μας αποτελείται από 2 εξισώσεις, το μοντέλο

$$SFR_{current} = A_{del} \frac{t_{sf}}{\tau^2} \exp(-t_{sf}/\tau), \quad t_{sf} = T_0 - t_{start}$$

και το ολοκλήρωμα

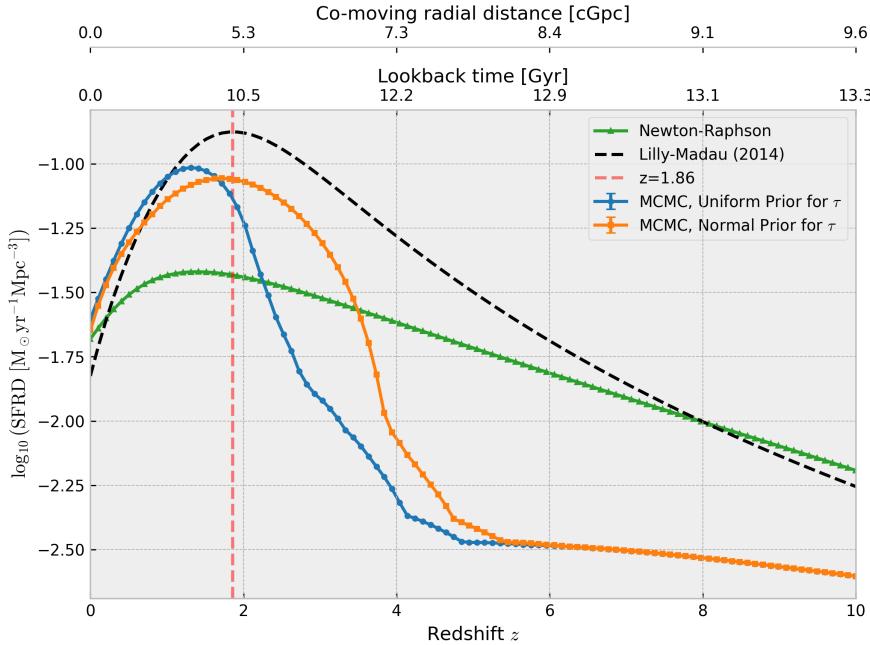
$$\int_{t_{start}}^{t_{current}} SFR(t) dt \approx M_{*,current}$$

Θεωρώντας $t_{start} = 0.2$ Gyr, το πρόβλημά έχει 2 εξισώσεις και 2 αγνώστους, το τ και το A_{del} .

Η δεύτερη μέθοδος, MCMC, βασίζεται στην τυχαία δειγματοληψία Monte Carlo και στις αλυσίδες Markov, για μπορέσει να υπολογίσει και τους τρεις αγνώστους του προβλήματος ταυτόχρονα (τ , A_{del} , t_{start}). Για να μπορέσει ο αλγόριθμος να υπολογίσει τις παραμέτρους, πρέπει να του έχουμε δώσει μια αρχική κατανομή πιθανότητας για τις μεταβλητές μας (priors), και βάση των δεδομένων μας, προσεγγίζει την πραγματική κατανομή των αγνώστων. Κρατώντας την μέση τιμή της κατανομής και την διασπορά της, έχουμε τιμές και για τις τρεις παραμέτρους του μοντέλου

για κάθε γαλαξία. Θα χρησιμοποιήσουμε 2 διαφορετικούς συνδυασμούς priors, $t_{sf} \sim \text{uniform}$ κατανομή με $\tau \sim \text{uniform}$ κατανομή, και $t_{sf} \sim \text{uniform}$ κατανομή με $\tau \sim \text{normal}$ κατανομή. Η συγκεκριμένη μέθοδος έχει το πλεονέκτημα ότι δεν χρειάζεται να περιορίσουμε σε μια συγκεκριμένη τιμή τον χρόνο που ξεκινάει η αστρογένεση, όπως χρειάζεται να κάνουμε στην N-R, αλλά επιτρέπουμε στους γαλαξίες να πάρουν διαφορετικές τιμές, δηλαδή μπορούμε με την σωστή επιλογή prior να υπολογίσουμε με μεγαλύτερη ακρίβεια το t_{sf} των γαλαξιών. Επιπλέον, δεν επιτρέπονται με αυτή τη μέθοδο αρνητικές τιμές του τ , κάτιο το οποίο είναι επιτρεπτό στην μέθοδο N-R.

Έχοντας υπολογίσει όλες τις παραμέτρους του μοντέλου με τις διαφορετικές μεθόδους, μπορούμε να πάμε να υπολογίσουμε το SFRD σε διαφορετικά z . Μέσω αυτών των διαφορετικών μεθόδων (χρήση IR αντί για FUV και H α και MCMC), καταφέραμε να δείξουμε ότι το μοντέλο μπορεί τελικά να προσεγγίσει καλύτερα την κοσμική SFRD από ότι είχε δειχθεί από την προηγούμενη βιβλιογραφία, κυρίως για $z = 1.86$ που ο λόγος από 2.16 έπεσε στο 1.56. Επίσης, δείξαμε ότι για διαφορετικά z είναι προτιμότερες διαφορετικές μέθοδοι, όπως για παράδειγμα για τα μεγαλύτερα z που το N-R προσεγγίζει καλύτερα το SFRD, υποδεικνύοντας κάποιο πρόβλημα με τα priors της μεθόδου MCMC. Τα αποτελέσματα αυτά, λοιπόν, παρόλο που δεν μπορούν να υποστηρίξουν με απόλυτη βεβαιότητα ότι το μοντέλο αυτό μπορεί να εξηγήσει πλήρως την αστρογένεση των γαλαξιών, είναι αρκετά ενθαρρυντικά τόσο για ενδεχόμενες βελτιώσεις του ίδιου του μοντέλου όσο και για τη μεθοδολογία που ακολουθήσαμε.



Σχήμα 2: Ανακατασκευή του SFRD(z) των γαλαξιών του LV

2 Galaxies and Galaxy Morphology

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.

2.1 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:

¹For more information about the Great Debate, visit https://apod.nasa.gov/debate/1920/cs_real.html

- **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.

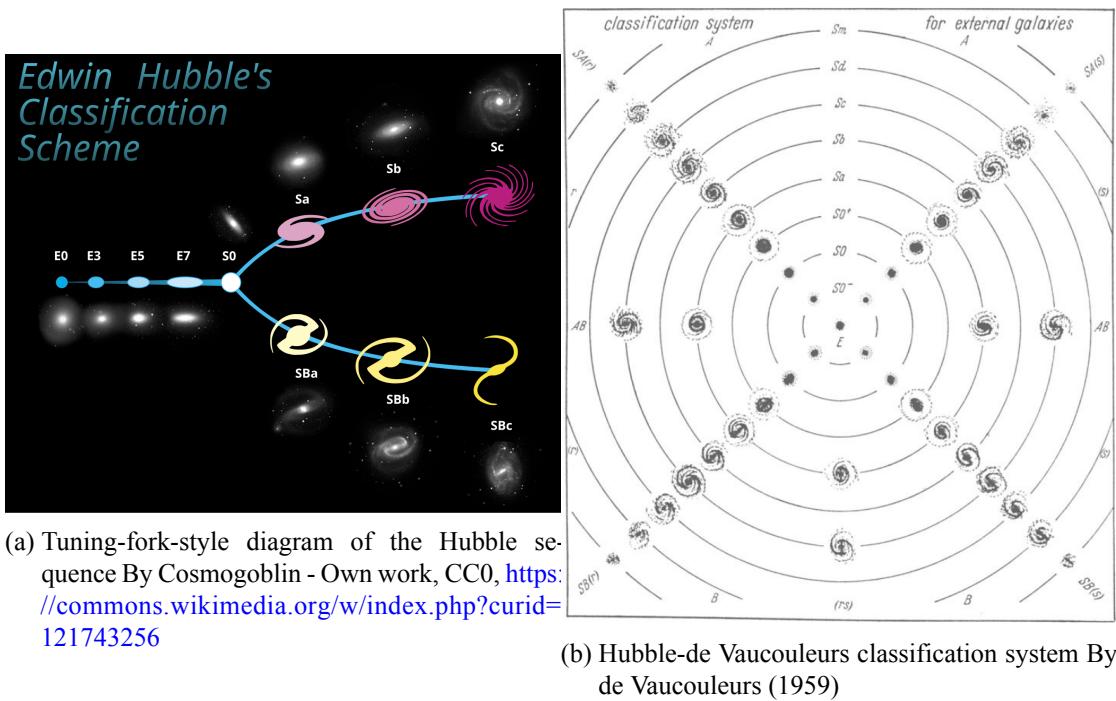


Figure 3: Diagrams visualizing the Hubble and Hubble-de Vaucouleurs morphological classification systems

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 Numerical Morphological T-Types, 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	Sc-Irr	Irr
T	-5	[-4,-3]	[-2,-1]	0	1	2	3	4	[5,6,7]	8	[9,10,11]

2.1.1 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)

3 Star Formation

ing professor 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.

3.1 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.

3.2 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:

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

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):

$$\text{SFRD} = \frac{1}{V} \sum_{i=1}^N \text{SFR}_{0,i}, \quad (2)$$

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 $V = \frac{4}{3}\pi(11 \text{ Mpc})^3 \approx 5575.3 \text{ Mpc}^3$.

3.2.1 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

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)

- **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 α ($0.6563 \mu m$) and H β ($0.4861 \mu m$). Dust can significantly affect observations. Since this estimator depends on young stars, it gives us information about the present day star formation (~ 10 Myr).
- **Far-Ultraviolet (FUV) Flux:** Mainly emitted by young, hot stars (B-type stars, ~ 100 Myr). Dust presence can also significantly affect observations. Since this estimator depends on young stars, it gives us information about the present day star formation (~ 100 Myr).
- **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 $\approx 60 \mu m$, whereas dust heated by UV-faint old or low-mass stars will emit at $\approx 100 - 150 \mu m$. 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. Since this estimator is somewhat age agnostic, but is derived from the radiatlon of heated dust, it is an estimator for the current star formation rate and not the present SFR (~ 300 Myr).

SFR for different luminosities L_i can be calculated as:

$$SFR_i = \mathcal{K}_i \times L_i \quad (3)$$

where \mathcal{K}_i is a constant specific to each L_i ($i = H\alpha, \text{IR}, \text{radio}, \text{FUV}, \text{X}$).

Since the luminosities L_{FUV} and $L_{H\alpha}$ 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 $SFR_{H\alpha}$ 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:

$$SFR_{\text{FUV}, H\alpha} = \text{mean}(SFR_{\text{FUV}}, SFR_{H\alpha}) \quad (4)$$

where L_{FUV} and $L_{H\alpha}$ 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.

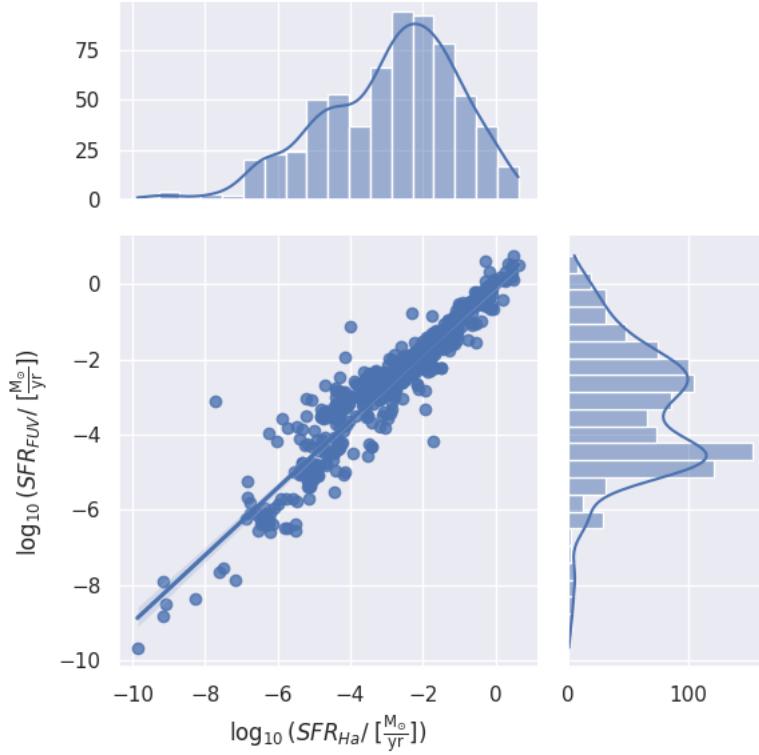


Figure 4: Plot showing the linear relation $\log \text{SFR}_{\text{FUV}} = \log \text{SFR}_{\text{H}\alpha}$, as well as their distributions, based on the data from UNGC

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

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

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

$$SFR_{\text{total}} = \text{mean}(\mathcal{K}_{\text{FUV}} \cdot L_{\text{FUV}}, \mathcal{K}_{\text{H}\alpha} \cdot L_{\text{H}\alpha}) + \mathcal{K}_{\text{IR}} \cdot L_{\text{IR}} = SFR_{\text{FUV}, \text{H}\alpha} + SFR_{\text{IR}} \quad (6)$$

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.

3.3 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)).

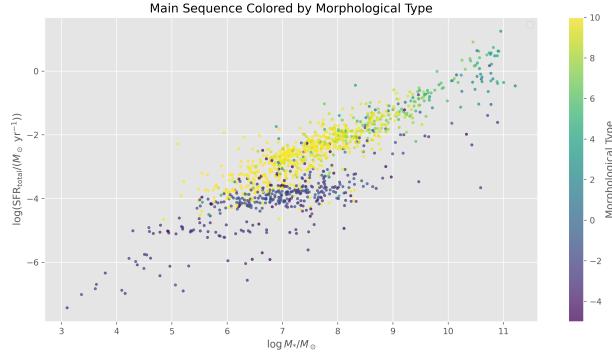


Figure 5: The Main Sequence of the galaxies within the radius of 10 Mpc, with the colorbar representing their morphological type. From the linear fit we get $\alpha(T_0) = 0.86$ and $\beta(T_0) = 9.83$. Early types have lower Stellar Masses and SFR, while spiral galaxies have more intense SF

3.4 Star Formation History 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, $t_{\text{sf}} = T_0 - t_{\text{start}}$ is the star formation time, T_0 is the age of the Universe, and t_{start} 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} .

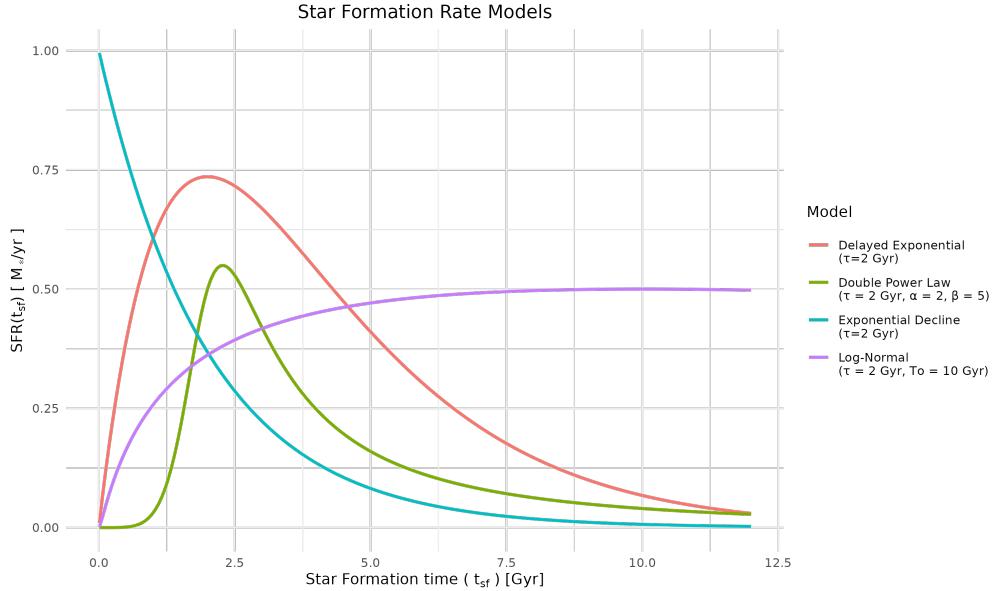


Figure 6: Star Formation Rate over the time of star formation, for different parametric models

- **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.

4 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.

4.1 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 $H_0 \approx 69.8$ km/s/Mpc is the Hubble constant and d is the distance between the two galaxies.⁴

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)

$$1 + z = \sqrt{\frac{1 + v/c}{1 - v/c}} \quad (7)$$

For small v/c we can rewrite Equation 7, as:

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

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

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

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

4.2 Lilly-Madau Plot

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):

$$\text{SFRD}(z) = 0.015 \frac{(1+z)^{2.7}}{1 + \left(\frac{1+z}{2.9}\right)^{5.6}} \left[M_\odot \text{ yr}^{-1} \text{ Mpc}^{-3} \right], \quad (8)$$

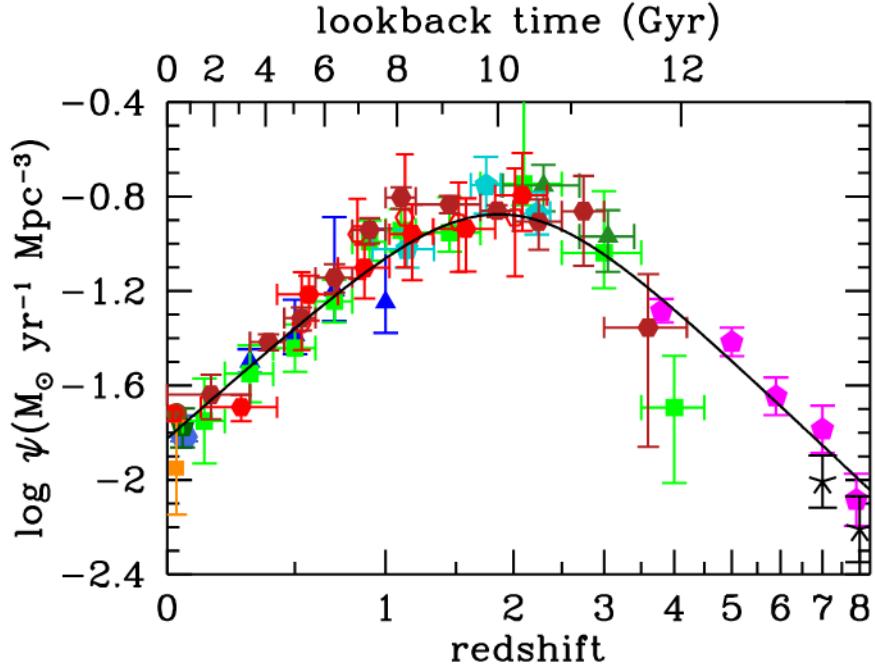


Figure 7: The history of cosmic star formation as shown in Madau and Dickinson (2014)

4.3 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:

$$\text{SFR}_0 = \text{SFR}(t_{sf}) = A_{\text{del}} \frac{t_{sf}}{\tau^2} e^{-t_{sf}/\tau}, \quad t_{sf} = T_0 - t_{\text{start}} \quad (9)$$

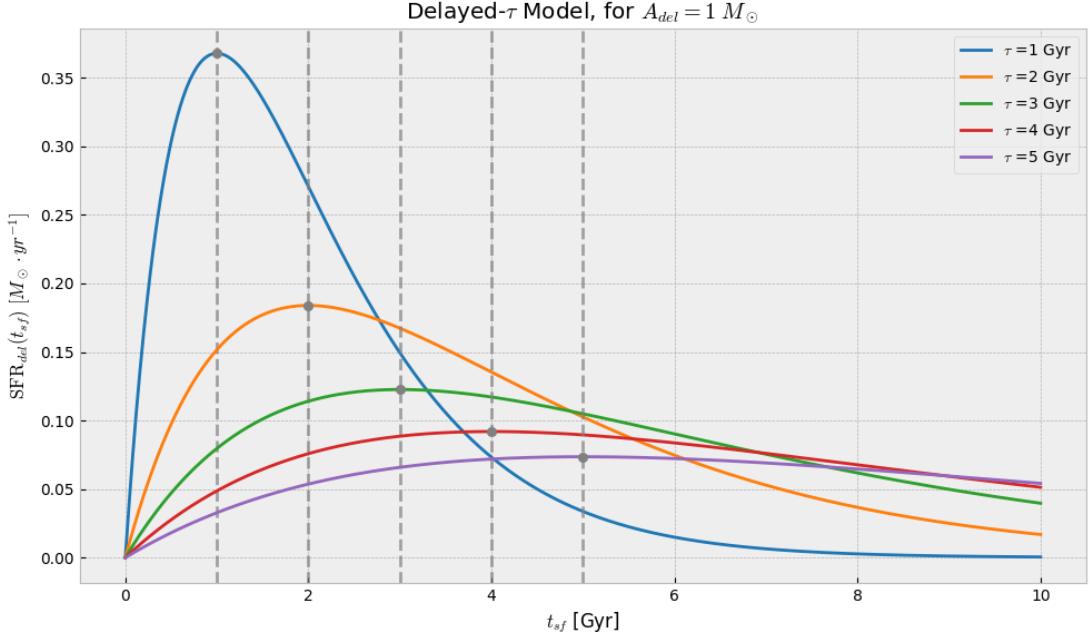


Figure 8: The evolution of the SFR of a galaxy over the time of star formation, according to the delayed- τ . The time need for the galaxy to reach the peak of its star formation is the timescale τ of the model

where the current SFR_0 of the galaxies, at the current age of the universe $T_0 = 13.8$ Gyr, 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 $SFR(t)$, we can calculate the present day averaged SFR:

$$\overline{SFR}_{\text{del}} = \frac{1}{t_{sf}} \int_{t_{\text{start}}}^{t_{\text{end}}} SFR_{\text{del}}(t) dt = \frac{A_{\text{del}}}{t_{sf}} \left[1 - \left(1 + \frac{t_{sf}}{\tau} \right) \exp \left(-\frac{t_{sf}}{\tau} \right) \right] \quad (10)$$

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

$$\overline{SFR}_{\text{del}} = \frac{\zeta M_*}{t_{sf}} \quad (11)$$

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).

5 Computational Methods

5.1 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.

5.2 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:

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

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 :

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

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

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

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 $\sum_{x^{(i)}} T(x^{(j)} \mid x^{(i-1)}) = 1$.

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

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

6 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 $z = 0 \Rightarrow t_{\text{emitted}} = T_0 = 13.8$ 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

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.

with $D \leq 11$ Mpc, we get a sample of 1321 galaxies. It offers detailed star formation rate 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 HECAFE.

6.1 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 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 HECAFE 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.
- **Star Formation Rate and Stellar Mass:** The percentage of the galaxies with measured 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 UNG.

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.

6.2 Comparing the Catalogs

Before we merge the two catalogs we need to compare the corresponding quantities of the 610 common galaxies to ensure that there are no major incompatabilities. The quantities that are comparable are:

- Distances [Mpc] (610/610 common galaxies)
- Radial Velocities [km/s] (602/610 common galaxies)
- Morphology and Geometry of the galaxies
 - Morphological type (521/610 common galaxies)
 - Inclination [degrees] (495/610 common galaxies)
 - Angular Diameter of the Major and Semi-major axis [arcmin] (579/610 common galaxies)
- K-band Luminosity [L_{\odot}] and Magnitude (172/610 common galaxies)
- B-band Magnitudes [mag] (107/610 common galaxies)
- Star Formation Rates [M_{\odot}/yr] (162/610 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 9, 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 9a, 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 2.1.1. 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.

Next we need to analyze the SFR from the two catalogs. The catalogs provide different SFR estimators, H α and FUV from UNGC and the total IR from HECATE. As already discussed in Section 3.2.1, the two estimators FUV and H α 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 10.

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.

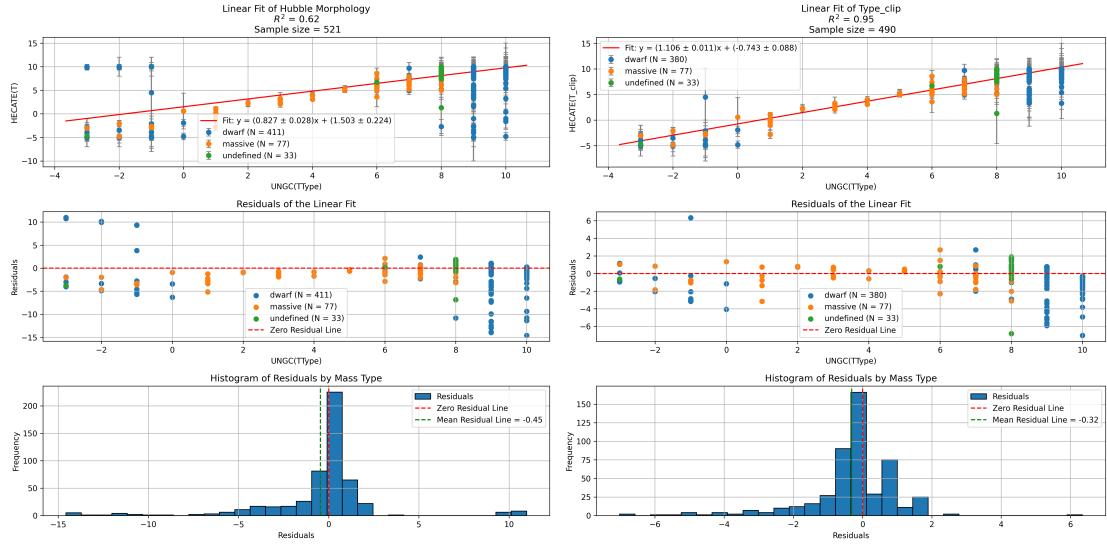


Figure 9: Scatter plots and the residuals of the Hubble Numerical morphological types show a good correlation between the two datasets, if we ignore the Transient dwarf galaxies.

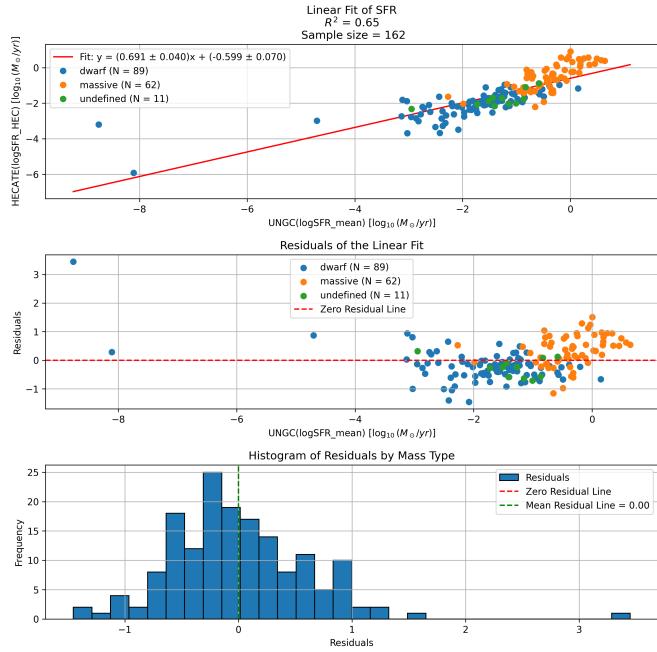


Figure 10: Scatter plot and the residuals of the mean(SFR_{FUV} , $SFR_{H\alpha}$) from the UNGC catalog and the SFR_{TIR} . The SFR estimation from the FUV and $H\alpha$ is systematically lower than the SFR_{TIR} estimator as expected. The R^2 is smaller than 80% however due to $\text{Residuals} = 0$ we can assume that the quantities are comparable.

In reality those estimators compliment each other and this is exactly why we they are not equal. The total SFR could potentially be calculated by Equation 6, 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 mean($SFR_{FUV}, SFR_{H\alpha}$) will be transformed into the expected SFR_{TIR} according to the linear regression of Figure 10.

After the merging of the two catalogs we get a final sample of 3612 galaxies out of which only 610 are included in both catalogs and were used for the comparison. 711 are unique galaxies in the UNGC and the remaining 2291 are unique in the HECATE catalogs. Additionally, 1753 galaxies have SFR data (778 from HECATE and 975 from UNGC), and 2917 galaxies have Stellar Mass data (1605 from HECATE, with the rest estimated using a mass-to-light ratio of 0.82 as provided by HECATE). 1691 galaxies have both SFR and M_* data.

7 Calculating the parameters

Using the merged catalog, we can use the SFR and M_* data to calculate the parameters for the delayed- τ model, for each individual galaxy. The parameters are going to be calculated using two different computational methods.

Firstly, using the Newton-Ramphson algorithm, the parameters τ and A_{del} can be calculated from the equations Equation 9 and Equation 10, assuming the same $t_{sf} = 13.6$ Gyr for every galaxy, meaning that all the galaxies start the star formation process around the same time, 13.6 Gyr ago, or $t_{\text{start}} = 0.2$ Gyr after the Big Bang. Solving Equation 11 and Equation 10 , for A_{del} :

$$A_{\text{del}} = \frac{\zeta M_*}{1 - (1 + x)e^{-x}}, \text{ where } x = \frac{t_{sf}}{\tau}, \quad (12)$$

and substituting in the model, the delayed- τ model takes the form:

$$\text{SFR} = \frac{\zeta}{t_{sf}} M_* \frac{x^2}{e^x - x - 1} \Leftrightarrow \text{sSFR} \frac{t_{sf}}{\zeta} = \frac{x^2}{e^x - x - 1}, \text{ where sSFR} \equiv \frac{\text{SFR}}{M_*} \left[\frac{1}{\text{yr}} \right].$$

The specific-SFR (sSFR) provides a normalized view of star formation intensity. Assuming a function $g(x)$ and its derivative $g'(x)$ for the Newton-Raphson algorithm:

$$g(x) = \frac{\zeta}{t_{sf} \cdot \text{sSFR}} \cdot x^2 - e^x + x + 1 \Leftrightarrow g'(x) = \frac{\zeta}{t_{sf} \cdot \text{sSFR}} \cdot 2x - e^x + 1$$

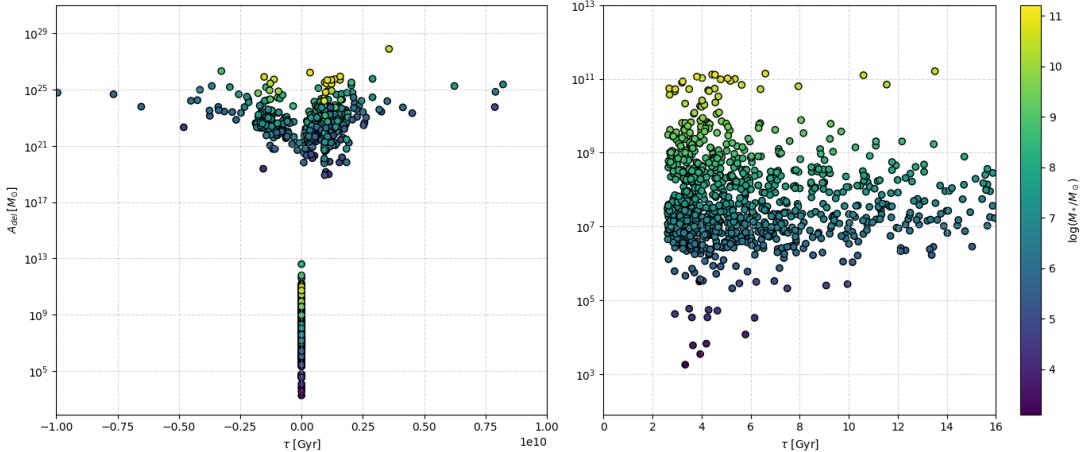


Figure 11: Parameter distribution of the delayed- τ SFH for the galaxies of the LCV. The colorbar represents the present-day stellar mass of the galaxies. The left plot shows the whole parameter range of the model, including negative timescales and A_{del} values that occur by instabilities. In the right plot, the region $\tau \in [0, 16]$ and $A_{del} \in [0, 10^{13}]$ is depicted. The majority of the galaxies have a timescale of about 3.5-4.5 Gyr as expected from Speagle et al. (2014)

the root $x_0 \Leftrightarrow g(x_0) = 0$, can be found, for each galaxy. From Equation 12 we can calculate the normalization constant of the galaxies, as well as their characteristic timescales $\tau = t_{sf} \cdot x$.

In Figure 11, two shapes are prominent. One in the range $\frac{A_{del}}{M_\odot} \in [10^{15}, 10^{29}]$ and one in the range $\frac{A_{del}}{M_\odot} \in [0, 10^{15}]$. The second range does not have any negative τ values and $\bar{\tau} = 16.2$ Gyr, while the first range has 153 negative τ values and $\bar{\tau} = -2.7 \cdot 10^9$ Gyr. It is obvious, from Figure 15c, that the first range is subject to the instabilities of the method.

From Equation 12, we expect a good linear correlation between A_{del} and M_* of a given galaxy, where the variance of the linear regression comes from $\frac{\zeta}{1-(x+1)e^{-x}}$. So, for the expected values of x , $2.7 < x < 3.4$ (Kroupa et al. 2020), and even for bigger x , there should not be a significant variance, since $\lim_{x \rightarrow \infty} \frac{\zeta}{1-(x+1)e^{-x}} = \zeta \approx 1.3$. However, for $x \rightarrow 0$, the algorithm is not able to converge since $\lim_{x \rightarrow 0} \frac{\zeta}{1-(x+1)e^{-x}} = \infty$, and the linear regression of $A_{del} - M_*$ can not be successful, as it can be seen from Figure 15c.

Since convergence can not occur for $x \ll 1$, this method should not be used for galaxies in the begining phases of their star formation ¹⁰. If the assumption, $t_{sf} = 13.6$ Gyr for all the galaxies is correct, then an inconstancy in the plot A_{del} – sSFR should only be observed, for higher values of the sSFR, since, according to the delayed- τ model, higher sSFR values indicate a rising SFR, while lower SFR values indicate that the bulk of M_* is already produced and the galaxy is in the

¹⁰For $x = 1$ is the peak of the SFR, $x < 1$ is the the rising phase and for $x > 1$ the SFR declines exponentially (Equation 9).

declining SFR phase. However, as seen from Figure 12, this inconsistency occurs in galaxies for both high and low sSFR’s, meaning that the the model has convergence issues both for slow rising SFR and quenching galaxies, indicating an intrinsic problem with this approach.

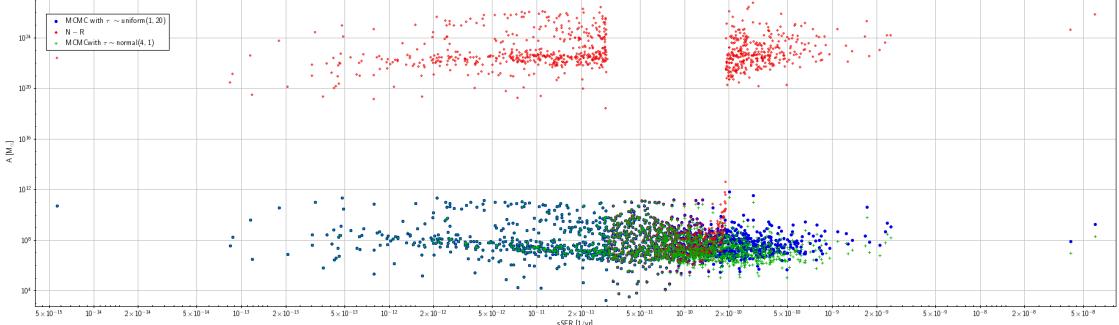


Figure 12: A_{del} – sSFR plot, where A_{del} is calculated with different methods.

The second method for the calculation of the parameters is the MCMC method. Using the logarithmic form of the model

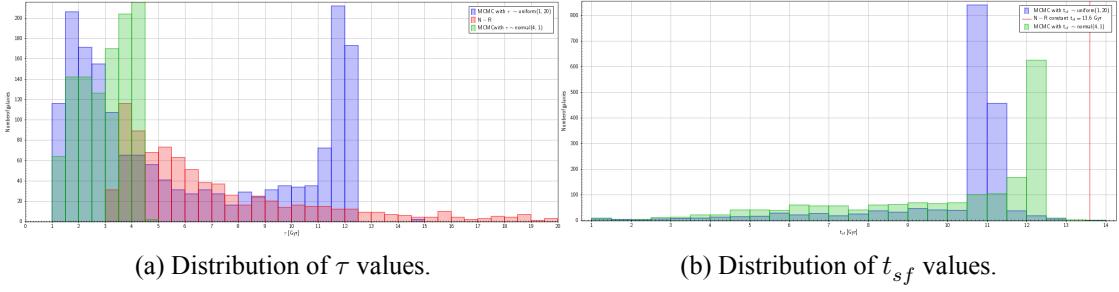
$$\log(\text{SFR}) = \log(A_{del}) + \log(x) - \log(\tau) - x \cdot \log e, \quad (13)$$

we can create a MCMC model to calculate the parameters, with the priors $t_{sf} \sim \text{uniform}(\text{lower bound} = 1 \text{ Gyr}, \text{upper bound} = 13.8 \text{ Gyr})$, $\zeta \sim \text{normal}(\text{mean}=1.3, \sigma = 0.01)$, within the limits [1, 2] and the likelihood $\log(\text{SFR}_{obs}) \sim \text{normal}(\log(\text{SFR}_{\text{MCMC}}), \sigma = 0.1)$, while taking the Stellar Masses of the galaxies as they are, without any variance, and A_{del} is calculated by Equation 12. For the τ we will consider two different priors, $\tau \sim \text{uniform}(\text{lower bound} = 1 \text{ Gyr}, \text{upper bound} = 13.8 \text{ Gyr})$ and $\tau \sim \text{normal}(\text{mean} = 4 \text{ Gyr}, \sigma = 13.8 \text{ Gyr})$. The MCMC run consists of 6 chains with 8000 iterations each (2500 iterations for the burn-in phase). Using this method, the galaxies can have different t_{sf} , while only taking positive values for their timescales.

The prior of the τ and t_{sf} follow a uniform distribution, since we do not want to introduce any unwanted bias in the value selection of the MCMC sampler. However, there is bibliography that could justify using a normal distribution for τ , for example Speagle et al. (2014) and Kroupa et al. (2020) explain that the galaxies of the Main Sequence are expected to have $3.5 \lesssim \tau/\text{Gyr} \lesssim 4.5$. The same prior will not apply for t_{sf} since it could significantly effect the results, especially of younger galaxies.

Using these priors, the metrics of the Stan software¹¹ give good results ($\hat{R} < 1.001$ and $\text{ESS} > 1000$), while the calculated values correspond to the data of the catalog. Specifically, we use the data to recalculate the SFR from the predicted parameters, using the delayed- τ model and compare the normalization constant to the stellar masses of the galaxies (Figure 15).

¹¹For more information about the usage of Stan and its metrics can be found here: <https://mc-stan.org/>



(a) Distribution of τ values. (b) Distribution of t_{sf} values.

Figure 13: Distribution of the results of τ and t_{sf} from all the methods.

In Figure 12, it can be seen that the MCMC run does not have the same inconsistencies with the N-R method, and Figure 13a shows that the MCMC method captures better both the galaxies with small timescales, and the galaxies with slow rising SFR's.

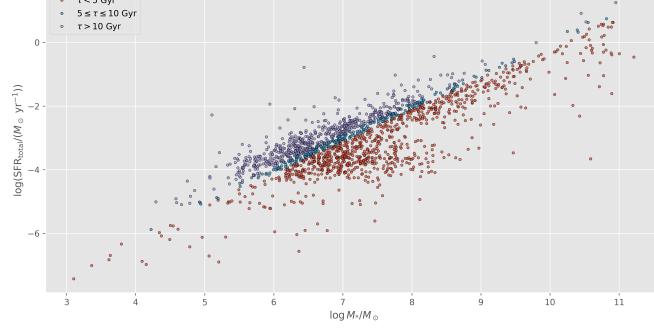


Figure 14: Main Sequence of the galaxies, color coded by the timescale τ , from the MCMC run from with the uniform prior for τ . The galaxies with slow rising SFR's are the galaxies above the center of the MS, while galaxies with τ predicted from Speagle et al. (2014) are in the lower part of the MS. In the center of the MS are the rest of the galaxies

In Figure 13a, the distribution of the τ from the MCMC run with the uniform prior, has two peaks, the first peak at ~ 2 Gyr and one at ~ 12 Gyr. As shown in Figure 14, the galaxies contributing in the two peaks are in different parts of the Main Sequence, while the MCMC run with the normal prior depicts the galaxies of the MS as they are predicted by Speagle et al. (2014), since every galaxy reaches its peak SFR at the a timescale $\tau < 5$. Additionally the MCMC run with the normal prior, predicts a peak of the t_{sf} distribution around 12-12.5 Gyr, closer to assumed values in Kroupa et al. (2020) and Haslbauer, Kroupa, and Jerabkova (2023) (Figure 13b).

Having calculated the model parameters, it is possible to reconstruct the SFH of the galaxies, and from the Equation 2, it is possible to calculate the SFRD of the LCV and recreate the Lilly-Madau Plot. Using the equations Equation 9 and Equation 2, we get the equation:

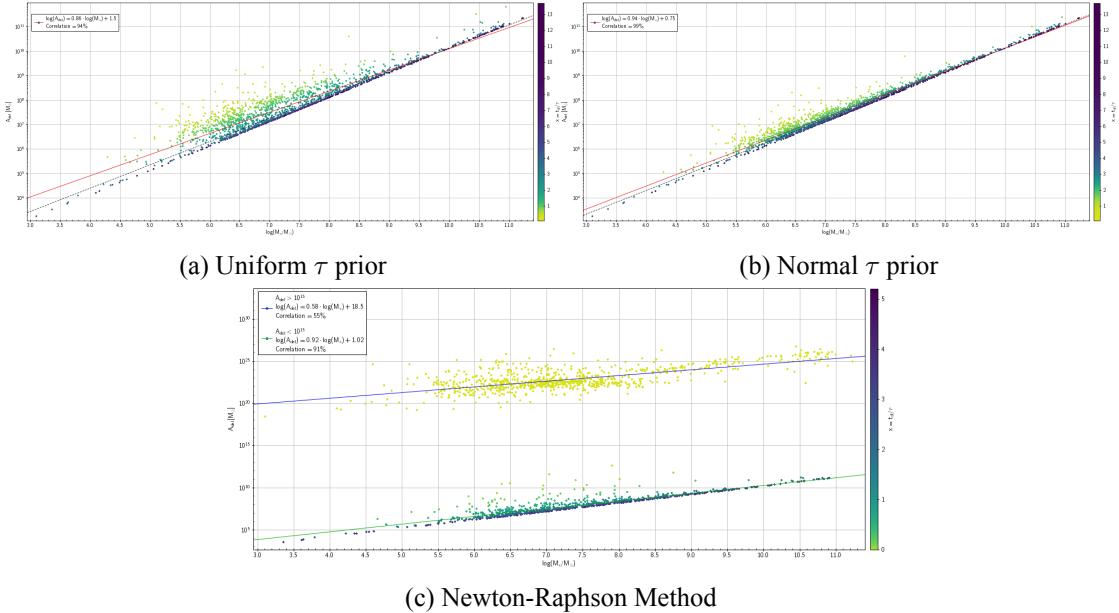
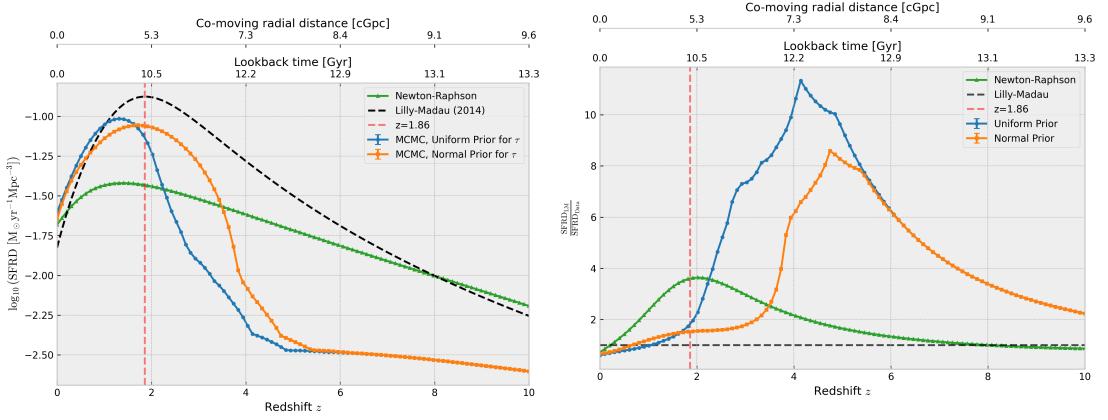


Figure 15: Linear regression for $A_{del} - M_*$ for the A_{del} values from the two MCMC runs and the Newton-Raphson method, with the colorbar representing the ratio $x = t_{sf}/\tau$. For the MCMC runs, the black dashed line represents the regression weighted by x . For the upper left plot, the regression yields a correlation of 94%, while the weighted regression is the fit $\log(A_{del}) = 0.95 \cdot \log(M_*) + 0.60$, with a correlation of 98%. For the upper right plot, the correlation is higher, at 99%, with a weighted regression of $\log(A_{del}) = 0.97 \cdot \log(M_*) + 0.44$, with a correlation of 99.5%. For the N-R plot, we do two different regressions for the two distinct regions of A_{del} , with the upper region only having a correlation of 55%, while the lower region closely relating the normalization constant with the Stellar Masses of the galaxies, as described in the MCMC plots.

$$\text{SFRD}(T) = \frac{1}{V} \sum_{i=1}^N A_{del} \frac{T - t_{\text{start}}}{\tau^2} \exp \left(-\frac{T - t_{\text{start}}}{\tau} \right), \quad (14)$$

where t_{start} , is calculated by the computed $t_{sf} = T_0 - t_{\text{start}}$, and T is the age of the universe (T_0 is the current age=13.8 Gyr). Using this equation we can recreate the parameterized SFRD.

As seen from Figure 16, the MCMC method can approximate better the SFRD of the LCV up until $z \sim 2$ with a factor less than 2. At the peak of the cosmic SFRD, at $z = 1.86$, the uniform and normal priors underestimate the SFRD by a factor of 1.82 and 1.53 respectively, while the N-R underestimates it by a factor of 3.6. In the paper, Haslbauer, Kroupa, and Jerabkova (2023), they estimate a factor of 2.16 at the peak, using the SFR's from the UNGC catalogs. However, the ratio of the MCMC grows with z , since many galaxies start forming stars at a lookback time



(a) SFRD(z) Reconstruction of the Local Cosmological Volume (b) Ratio between the cosmic SFRD measured by Madau and Dickinson (2014) and the SFRD of the LCV

Figure 16: SFRD(z) Reconstructions of the Local Cosmological Volume, based on the three different methods.

of ~ 12 Gyr, as seen in Figure 13b (the mean value of the t_{sf} of the galaxies is 10 Gyr for the uniform prior and 9.8 Gyr for the normal prior), resulting in extreme deviations from the cosmic SFRD.

8 Conclusions

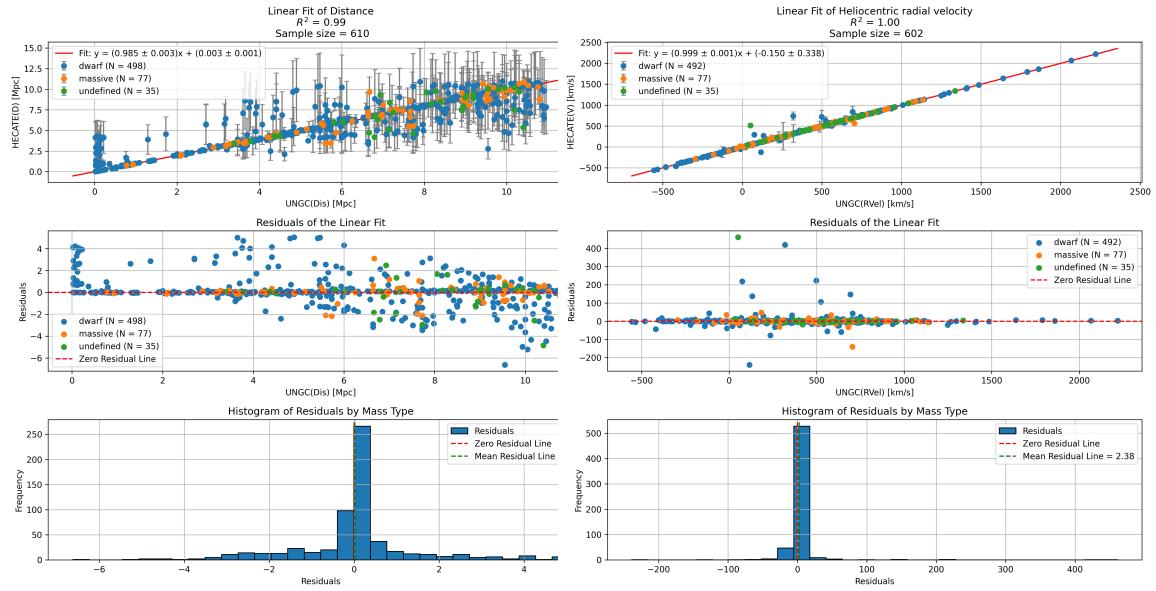
With the use of IR data for the SFR of the LCV and with the use of different algorithms, we were able to reduce the divergence of the SFRD of the LCV from the cosmic SFRD from a factor of 2.16 to 1.53 at $z = 1.86$. The N-R, can describe better the SFRD of the LCV compared to the MCMC for higher redshifts, but fails to approximate it for lower redshifts. The high divergence of the MCMC approximations, for $z > 2$ is a result of the wrong prior of the t_{sf} . Using uniform priors for the time of star formations, severely underestimates the time in which most galaxies start the star forming process, meaning that most of the galaxies stop contributing at the SFRD at earlier lookback times than they should.

A better assumption for the prior of t_{sf} , which pushes the mean $\overline{t_{start}} \approx 0.2$ Gyr, could correct the model at higher redshifts, since it would approximate the assumption of the N-R, while also keeping the lower factor of the MCMC at the peak of the Lilly-Madau plot. A possible prior for the t_{sf} could be a skewed normal distribution with a mean value $\overline{t_{sf}} \approx 13$ Gyr and with a large sigma to cover the younger galaxies of the LCV. Another possible distribution could be a log-normal distribution for the t_{start} . But since the choice of a prior can significantly affect the results and because of the lack of a bibliography for this model, for this undergraduate thesis, the choice of a uniform is the safest choice.

Based on our results, we cannot safely assume that the model works for the galaxies of the LCV. Since parameterized SFH models do not account for bursty star formation, events characteristic of starburst galaxies, incorporating additional parameters who depend on $SFR_{H\alpha}$ and SFR_{FUV} could increase the accuracy of the model. Additionally, only examining one model ignores the intrinsic characteristics of the galaxies. Haslbauer, Kroupa, and Jerabkova (2023), introduce a combined SFH model depending on the mass of the galaxies, which can reduce the difference of the predicted model and the cosmic SFRD.

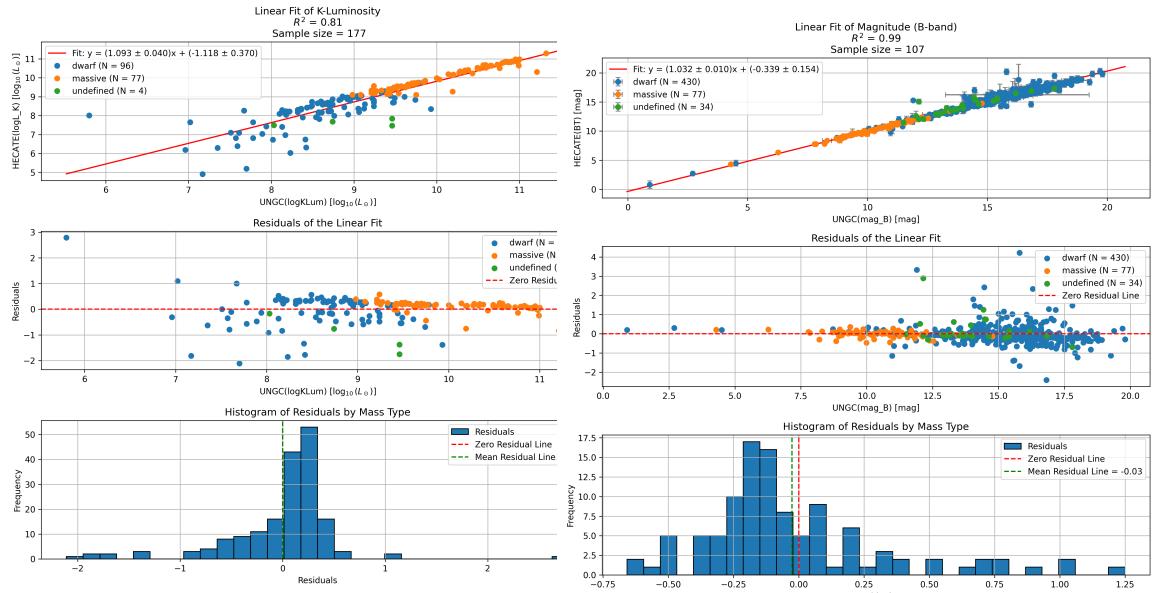
In summary, the use of the MCMC and the data from the HECATE catalog proves that the delayed- τ model a better approximation of the SFH, as originally argued by Haslbauer, Kroupa, and Jerabkova (2023). However, more fine-tuning of the model is needed to be certain if the delayed- τ model is a robust approximation of the SFH of the galaxies of the Local Cosmological Volume.

9 Appendix A: Scatter plots of the comparisons



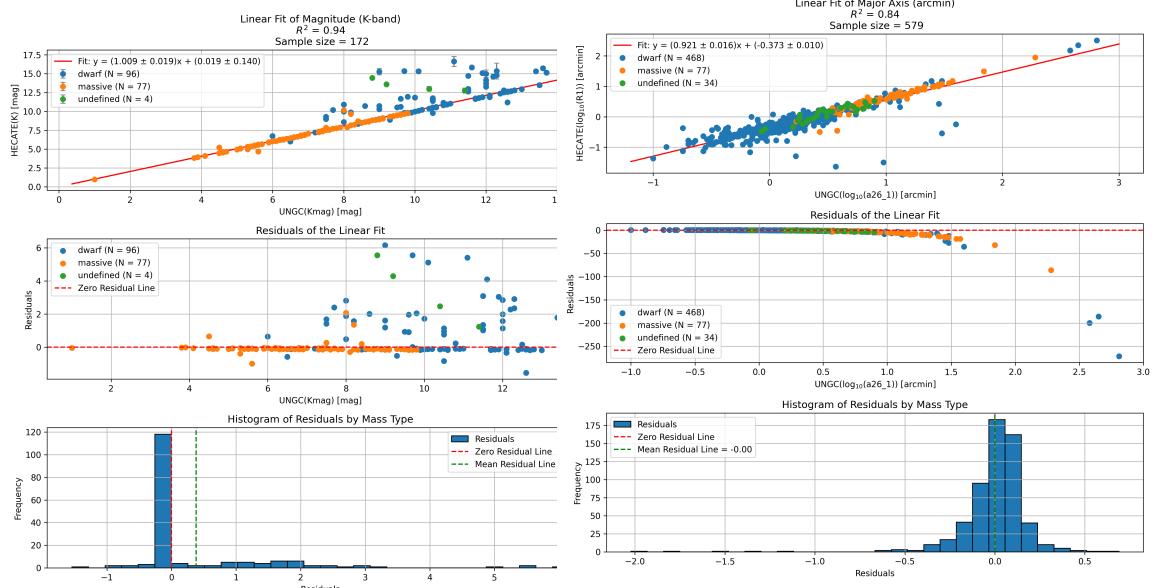
(a) Linear Regression of the Distances in Mpc

(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 Maryland.

- 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 $z \sim 0.6$.” *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>.