Dark matter fraction in massive galaxies under Self-Interacting Dark Matter cosmology

Département d'Astrophysique, CEA Paris-Saclay

Author

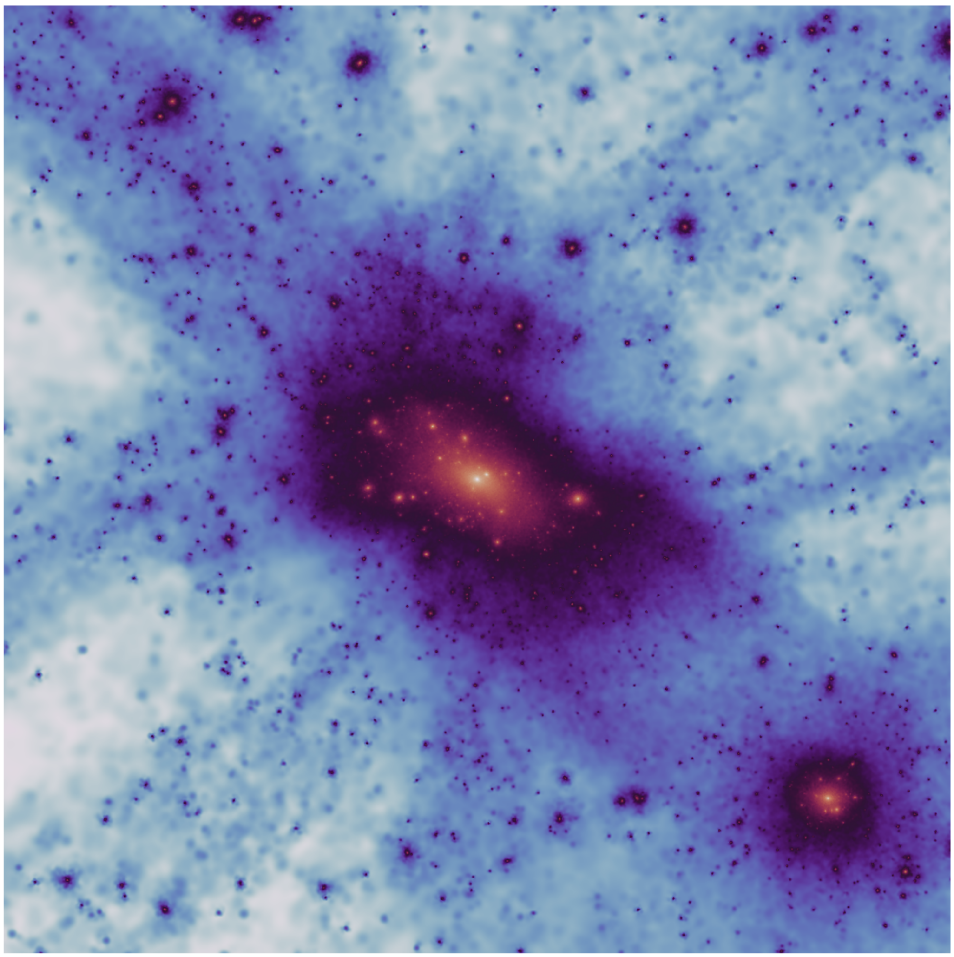
Anna Preto

Supervisor

Camila Correa

Date

March 2023 - July 2023



Central dark matter halo - TangoSIDM

Abstract

Understanding the nature of dark matter is one of the most significant challenges in modern cosmology of our century. The standard cosmological model, the Λ Cold Dark Matter (ΛCDM model), accurately predicts the physics of the universe on large scales. However, on galactic and sub-galactic scales ($\leqslant$ a few Mpc), observations seem to contradict the structure formation predicted by numerical simulations (such as Moore, 1994; Blok and McGaugh, 1997a; Moore et al., 1999; Spergel and Steinhardt, 2000; Boylan-Kolchin et al., 2011; Sales et al., 2022). This motivates the exploration of alternatives to CDM. Self-Interacting Dark Matter (hereafter SIDM), appears promising. SIDM cosmology considers that dark matter particles undergo non-gravitationnal interactions among themselves and, when properly parameterized in simulations, resolves many issues such as the diversity of the inner dark matter content in dwarf galaxies.

This internship has focused on examining the predictions of the SIDM model. The objective is to study the dark matter mass within simulated galaxies from the TangoSIDM simultions project and compare them with observations. Finally, through this observable, we aim to constrain the SIDM model and provide an estimate of the effective cross section of dark matter particles.

Plagiarism Statement

I, Anna PRETO, hereby declare on my honor:

- the results described in this report are the outcome of my own work;
- I am the author of this report;
- I have not used any sources or third-party results without clearly citing and referencing them according to the recommended bibliographic rules.

I declare that this work cannot be suspected of plagiarism.

Date: June 14, 2024
Signature:

Acknowledgments

This internship has been the most enriching experience of my academic journey. Beyond the fascinating research topic, the supervision was conducive to its successful completion. I would like to extend my deepest thanks to Dr. Camila Correa for her guidance. Her infectious passion and pedagogy, coupled with unmatched kindness, created an environment I could not have imagined more perfect for a research project that concludes five years of exciting studies. I have learned tremendously from her, both in astrophysics and in work methodology.

I also thank my partner, Téo Granger, for all the stimulating and motivating discussions that helped me surpass my limits and for all the evenings spent coding and discussing our respective research fields.

Thank you to my classmates and friends, who also completed their internship at the CEA Department of Astrophysics (Marie Ausseresse, Jean-Baptiste Billand, Clara Bonini, and Xavier Corap) for our numerous discussions and laughs, without which this experience wouldn't have had the same flavor.

As this internship concludes my academic journey, I would also like to express my broader gratitude to my mother, Marie Preto, for her support and listening throughout these years, and for her unwavering encouragement since my childhood in this passion for astrophysics.

Contents

1 Introduction **5**

I History and context . . . 5

I.1 The idea of missing mass . . . 5

I.2 In search of a candidate . . . 7

II Self-Interacting Dark Matter Model . . . 8

II.1 Why the Self-Interacting Dark Matter Model? . . . 8

III Research Project Goals . . . 11

2 Dark matter fraction **13**

I Methods . . . 13

I.1 Definition of our observables . . . 13

I.2 Catalogues construction . . . 18

II Dark matter fraction's dependencies . . . 21

II.1 Mass and compactness . . . 21

II.2 Baryonic and dark matter physics . . . 22

II.3 Galaxy types . . . 23

III Comparison with observations . . . 24

III.1 Dark matter fraction in LTGs . . . 24

III.2 Dark matter fraction in ETGs . . . 25

III.3 Lower ETGs' f_{DM} . . . 26

IV Discussion . . . 27

3 Cross section **28**

I Semi-analytic model . . . 28

I.1 Method . . . 28

I.2 Results . . . 29

I.3 Outlook and future directions . . . 32

II Machine learning approach . . . 32

II.1 Method . . . 32

II.2 Results . . . 36

III Conclusion . . . 38

A Structure of machine learning algorithms **39**

I Simple neural network . . . 39

II Complex neural network 39

Chapter 1

Introduction

If we were to, in a very general manner, outline the current astrophysical issue within which this research project fits, it would take the following form:

> What is the cross section - or non-gravitationnal interactions probability - of dark matter particles?
> *If such a property exists.*

The last sentence is not anodyne: *If such a property exists...* To grasp its significance, one must unravel a series of questions that have led the astrophysical community to consider this hypothesis. *Where does the idea of an invisible matter pervading galaxies come from? Why are astrophysicists seeking alternatives to cold dark matter? What are these alternatives and why have they emerged?* As the reader will understand, to comprehend the ins and outs of this project, these questions must be addressed to contextualize the research problem. Thus, this introduction will explore these three questions by first presenting the evidence for the existence of dark matter, then examining the various approaches that have been adopted throughout history to understand its particle nature and characteristics. Finally, it will discuss how the research project fits into this quest.

I History and context

I.1 The idea of missing mass

> *The most serious problem in extragalactic astronomy today is the notorious "missing mass". (I know that Jerry Ostriker derides the term; he tell us, quite correctly, that it is not missing mass at all, since we know from its gravitation that it must be there. But this is just a difference of taste. I prefere to call it "missing" because it is certainly missing from our understanding.)*
>
> *King (1977)*

The first evidence of the existence of dark matter dates back to the 1930s, when Fritz Zwicky, studying the velocities of galaxies in the Coma Cluster, demonstrated that the luminous mass of galaxies was much lower

than their estimated mass using the Virial theorem (Zwicky, 1933). He then postulated that undetectable matter, neither emitting nor absorbing electromagnetic radiation, existed within galaxies and introduced the innovative idea in 1937 that such objects could be detected through the gravitational lensing effect proposed a year earlier by Albert Einstein (Einstein, 1936; Zwicky, 1937).

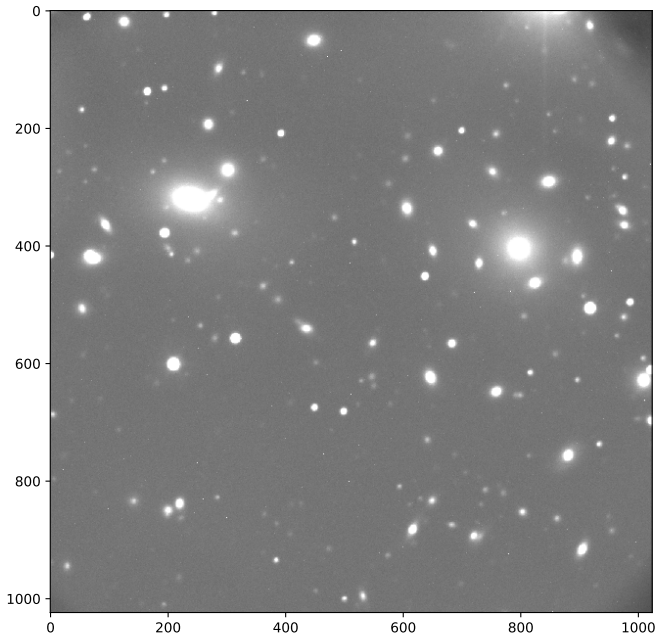


Figure 1.1: Center of the Coma Cluster. Credit: Anna Preto and Xavier Corap, Observatoire de Haute Provence (T120).

Throughout a long period extending into the 1960s, few articles attempting to address this issue were published, likely due to the historical and political context of the time. Only a handful of scientists continued to study this problem, mainly providing confirmations and refinements in measurements. Finally, in 1959, as Slipher (1917) measured the approach velocity of Andromeda towards our galaxy, Kahn and Woltjer (1959) estimated the mass of this group and highlighted the same issue as seen in clusters.

However, it was only from the mid-1970s onwards that the existence of dark matter became widely accepted, as evidenced by a significant increase in publications in this field during that period. Vera Rubin played a decisive role by providing crucial evidence for this hypothesis. In her 1970 paper (Rubin and Ford, 1970), she examined the rotation curve of stars in the Andromeda galaxy, triggering numerous subsequent studies that all led to the conclusion of massive halos of invisible matter within galaxies.

Finally, with the emergence of numerical simulations from the 1960s-1970s, the formation and gravitational stability of spiral arms in a self-gravitating disk, a analytically complex problem, could be better studied. It quickly became apparent that visible matter alone was insufficient to ensure the long-term stability of the rotating disks of spiral galaxies.

Ostriker and Peebles (1973) demonstrates that the most appropriate solution to resolve this stability problem is to consider a halo of dark matter.

The discovery of the cosmic microwave background (Penzias and Wilson, 1965) and its analysis (see for example, Smoot et al., 1992; Peebles and Yu, 1970), or even gravitational lensing observations of cluster merger (Clowe et al., 2006) also demonstrated the presence of non-baryonic matter in the Universe.

After all these observational considerations, a question arose: '*if dark matter exists, what is it made of?*".

I.2 In search of a candidate

Before considering dark matter to be of an exotic nature, a significant number of astrophysical candidates were tested, such as MACHOs (Massive AstrophysiCal Halo Objects), including brown dwarfs, black holes, or hot Jupiters (Griest, 1991). Alcock et al. (2000) demonstrates that these candidates can account for at most 20% of the halo.

Thus, the idea that dark matter consists of particles that do not interact electromagnetically was not as evident as it is today. At that time, particle physics and cosmology were very distinct fields, and it took several years for them to converge in a collaborative effort to elucidate the nature of dark matter. Moreover, Ramanath Cowsik, considered the father of astroparticle physics, confided to James Peebles in 2016 during a conversation transcribed in his book Modern Cosmology, that he himself became aware of the connection between particle physics and dark matter in 1972:

> This observational confirmation of the existence of this invisible matter led me to think that if it was capable of dominating the mass of large clusters, it must also dominate the dynamics of the Universe.

The particle physicists studying neutrinos began, initially shyly, to address the problem of missing mass, such as Gunn et al. (1978), who subtly included in their abstract that a certain type of lepton "could provide matter in galactic halos, or provide the necessary mass to bind galactic clusters", before clearly using cosmological parameter measurements to constrain the mass of resting neutrinos (for example, Gershtein and Zel'dovich, 1966). The announcement by Lubimov et al. (1980) - later invalidated - of measuring the neutrino mass (ranging from 14 eV to 46 eV), which opened up significant prospects, contributed to this collaborative momentum between the two fields. However, with the advent of simulations, neutrinos are considered to be hot dark matter and disqualified due to their disagreement with observations on the formation of large structures (Melott et al., 1983).

By the late 1980s, the idea that the majority of the mass in the Universe is composed of cold, non-baryonic particles was widely accepted by both many astrophysicists and particle physicists. Cold dark matter particles, being slow-moving, avoid early thermal decoupling and allow primordial density perturbations to grow, which is essential for the formation of galaxies and large-scale structures. This prevailing concept of cold dark matter, represented by unknown elementary particles, was accompanied by a range of potential alternative solutions, including massive neutrinos, supersymmetric particles such as neutralinos, and axions. Physicists gradually converged on a category, Weakly Interacting Massive Particles (WIMPs), which includes not only neutrinos but

also encompasses a broad spectrum of stable particles with masses ranging from MeV to TeV. For more than a decade, experimental projects have emerged to demonstrate the existence of such a particle. They fall into three types of experiments: direct detection, indirect detection, or those involving particle colliders. To date, no particle has been favored as the dark matter candidate.

In parallel with these searches, alternative theories have also been developed to explain the observations described in section I.1. Theories aiming to modify gravity, also known as Modified Newtonian Dynamics (MOND) theories, currently explain many observations on the galactic scale (Li et al., 2018; McGaugh, 2012; Milgrom, 1983) but fail at larger scales.

II Self-Interacting Dark Matter Model

II.1 Why the Self-Interacting Dark Matter Model?

Although its precise particle nature remains unclear, there is, however, a convergence towards the cold dark matter model. Nevertheless, even within this consensus, problems persist, and not all observations align with this model. On the galactic scale, three significant issues have emerged. The following section outlines the challenges faced by the CDM model and then introduces the alternative model at the center of the internship project.

Challenges Faced by the CDM Model

Early dark matter simulations predicted a significantly larger number of satellite galaxies than were observed (Klypin et al., 1999), giving rise to the "missing satellites" problem. However, the discovery of several new satellite galaxies (Simon, 2019) and the development of simulations incorporating the baryonic effects of supernova feedback and reionization seem to have resolved this issue (Font et al., 2021; Engler et al., 2021).

The "core-cusp" problem, refers to the observation that the dynamical mass of some dwarf galaxies appears to be lower than predicted by CDM simulations (Moore, 1994; Blok and McGaugh, 1997b; Oh et al., 2011; Walker and Peñarrubia, 2011). These simulations predict dark matter halos with dense central regions and steep "cusp" density profiles (Navarro et al., 1997), whereas many dwarf galaxies exhibit lower central densities with flatter "cored" density profiles (Walker et al., 2010; Boylan-Kolchin et al., 2011; Ferrero et al., 2012; Read et al., 2019).

Finally, the last issue, known as the "too-big-to-fail" problem, refers to a discrepancy between the masses of the dwarf satellite galaxy halos of the Milky Way simulated under CDM cosmology and observations (Boylan-Kolchin et al., 2011; Boylan-Kolchin et al., 2012).

These issues compel astrophysicists to explore alternative explanations to CDM. This is where our research project comes into play. Self-Interacting Dark Matter (SIDM) is a dark matter model in which particles can

interact non-gravitationally with each other. Theoretically, this model aims to address specifically the "diversity" problem. Indeed, dark matter particles, through collisions, transfer energy to the colder regions of the halo, thus reducing the central density and creating a density core rather than a cusp.

The SIDM model

The main parameter for describing interactions among dark matter particles is the cross section per unit mass σ/m_χ, as it describes the probability, and thus the frequency, of interactions (Robertson et al., 2017; Kahlhoefer et al., 2019; Kummer et al., 2019). Within the dwarf galaxies, a low cross section ($\sigma/m_\chi < 1$ cm^2/g) reproduces the results of the ΛCDM model and predicts a "cusp" in the dark matter density profile, whereas a high cross section ($\sigma/m_\chi > 1$ cm^2/g) leads to a flatter density core that is more consistent with observations. A cross section $\sigma/m_\chi > 10$ cm^2/g can reproduce both cusps and cores. The model we studied in this project is a velocity-dependent model. Figure 1.2 illustrates the cross section per unit mass of dark matter particles as a function of their velocity.

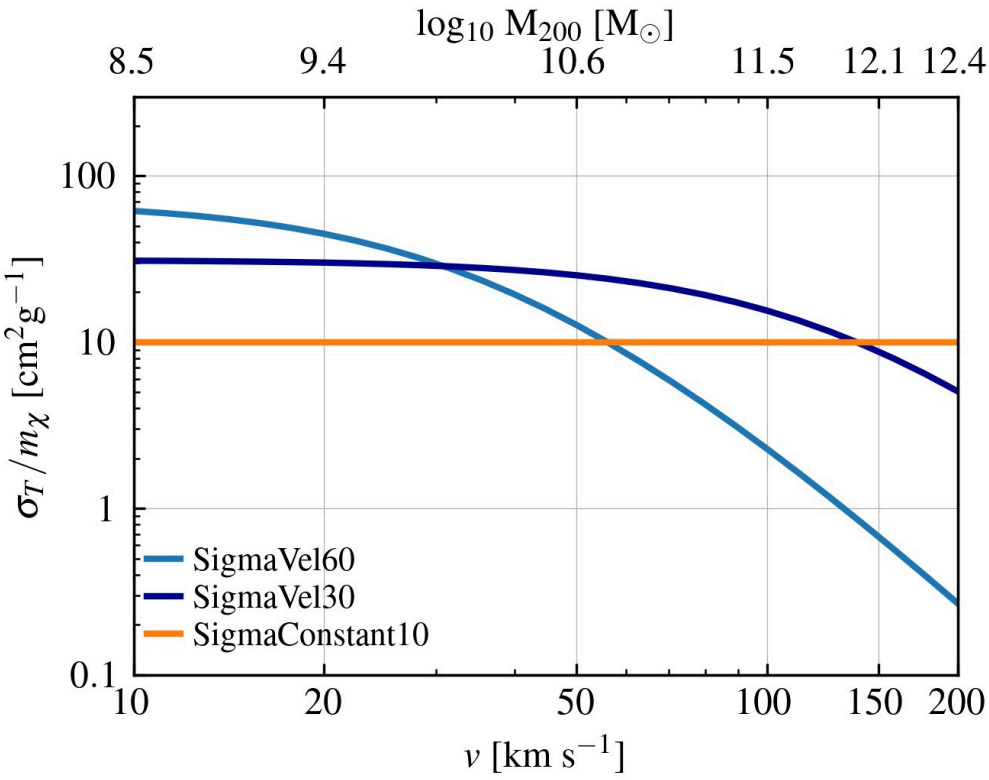


Figure 1.2: Cross section per unit mass as a function of the relative scattering velocity among dark matter particles for the SIDM models. The figure shows two velocity-dependent models, namely SigmaVel60 (light blue line) and SigmaVel30 (dark blue line) used in this work. The top x-axis indicates the typical halo mass that hosts orbits of the velocities indicated on the bottom x-axis.

This model makes dark matter to be highly collisional on small-scales while it replicates the correct predictions of the CDM model on large scales, where dark matter particles with lower velocities behave like collisionless particles.

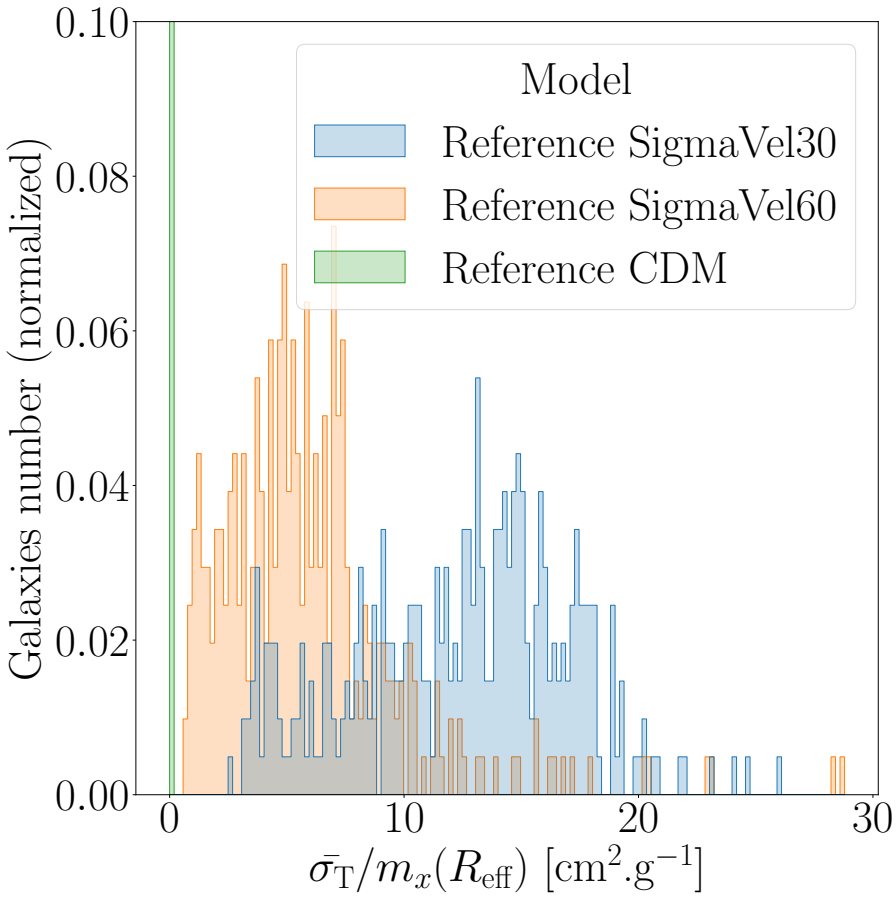


Figure 1.3: Normalized histogram of the average cross section values of the particles within the effective radius $\bar{\sigma_{\mathrm{T}}}/m_x(R_{\mathrm{eff}})$ for each model.

Figure 1.3 shows the normalized histogram of the average cross section values of the particles within the effective radius $\bar{\sigma_{\mathrm{T}}}/m_x(R_{\mathrm{eff}})$ for each model, illustrating the range of values produced by each model.

TangoSIDM

The simulation results used for this project are derived from the TangoSIDM project. The TangoSIDM project, along with its models and implementation of SIDM, is detailed in Correa et al. (2022). This section will cover key aspects of the SIDM model necessary for understanding the project's continuation.

The simulations follow the evolution of 376^3 dark matter particles and 376^3 gas particles from redshift = 127 to redshift z = 0. The dark matter particle mass is 9.70 $\times$ 10^6 $M_\odot$ and the gas initial particle mass is 1.81 $\times$ 10^6 $M_\odot$. The softening is set to 2.66 comoving kpc at early times and fixed at a physical value of 700 pc from $z = 2.8$. The adopted cosmological parameters are $\Omega_m = 0.307$, $\Omega_\Lambda = 0.693$, $h = 0.6777$, $\sigma_8 = 0.8288$, and $n_s = 0.9611$ (Planck Collaboration et al., 2014).

We studied three dark matter models: CDM and two variants of SIDM with velocity-dependent cross sections, named SigmaVel60 and SigmaVel30. The SigmaVel60 model features a cross section less than $1\,\mathrm{cm}^2 \cdot \mathrm{g}^{-1}$ at velocities above $150\,\mathrm{kpc} \cdot \mathrm{s}^{-1}$, reaching $60\,\mathrm{cm}^2 \cdot \mathrm{g}^{-1}$ at $10\,\mathrm{kpc} \cdot \mathrm{s}^{-1}$. In comparison, the SigmaVel30 model has a cross section less than $8\,\mathrm{cm}^2 \cdot \mathrm{g}^{-1}$ beyond $200\,\mathrm{kpc} \cdot \mathrm{s}^{-1}$, reaching $30\,\mathrm{cm}^2 \cdot \mathrm{g}^{-1}$ at $10\,\mathrm{kpc} \cdot \mathrm{s}^{-1}$.

These SIDM models represent extreme cases of dark matter interactions in systems comparable to the Milky Way. SigmaVel60 shows an interaction of $1 - 2\,\mathrm{cm}^2 \cdot \mathrm{g}^{-1}$ around $100\,\mathrm{kpc} \cdot \mathrm{s}^{-1}$, indicating a low interaction rate at the center of Milky Way-type halos, while SigmaVel30 reaches an interaction of $10 - 20\,\mathrm{cm}^2 \cdot \mathrm{g}^{-1}$ at $100\,\mathrm{kpc} \cdot \mathrm{s}^{-1}$, suggesting a higher interaction rate.

TangoSIDM uses two distinct models from SWIFT-EAGLE, an open-source galaxy formation model implemented in SWIFT, derived from the original EAGLE model (Schaye et al., 2014; Crain et al., 2015). The Reference model is calibrated to reproduce galactic properties within a 25 Mpc^3 volume, while the WeakStellarFB model applies parameters that significantly reduce stellar feedback.

We thus have six simulations at our disposal, exploring three dark matter cosmologies—CDM, and two velocity-dependent SIDM variants (SigmaVel60 and SigmaVel30)—under two models: Reference and WeakStellarFB.

III Research Project Goals

The main aim of this project is to understand the impact of dark matter particle collisions in the central regions of galaxies. Understanding this impact is crucial for identifying observables that will be discriminative in determining the nature of dark matter. To achieve this, it is necessary to compare the results from TangoSIDM with those from CDM simulations. The interest in having these cosmologies under two different models, Reference and WeakStellarFB, is to explore the behavior of SIDM in galaxies with different properties. For instance, Correa et al. (2024) highlights that galaxies in the WeakStellarFB model are more compact than those in the Reference model.

The observable chosen for this project is the fraction of dark matter within massive galaxies at z = 0. This property has been studied in the past, primarily within the framework of CDM cosmology (see for example Lovell et al., 2018). This observable enables immediate identification of the dominant matter type at the galaxies center. The aim is to study its evolution based on the dark matter model and galaxy properties.

Secondly, the aim is to compare these results with observations to draw conclusions about dark matter in observed galaxies, particularly to constrain the cross section per unit mass of particles if dark matter is indeed self-interacting.

To achieve these two objectives, we have implemented the following workflow:

- Develop a dataset for both observations and simulated galaxies, ensuring consistent property comparisons.
- Compare them using relevant visualizations and analyze the results.
- Explore ways to constrain the cross section of dark matter particles.

Thus, the rest of the report is divided into two parts. The first details the methods implemented to address the following issue: “how does SIDM behave and what do observations tell us about the nature of dark matter?”

Finally, the second part explains the procedures used to constrain the cross section of dark matter particles based on the properties of observed galaxies and the previous analysis.

Chapter 2

Dark matter fraction

Several observational studies have calculated the dark matter fraction in massive galaxies (Auger et al., 2009; Barnabè et al., 2011; Tortora et al., 2012; Cappellari et al., 2013; Li et al., 2024b). The dark matter fractions reported in these studies vary considerably and do not align with CDM simulations. These variations can be attributed to differences in the several Initial Mass Functions (IMF) adopted, the types of galaxies studied, the measurement techniques. Since these studies not only present observational data that diverges from the CDM model but also exhibit inconsistencies with each other, it is not appropriate to conclude the use of an alternative cosmological model to address this issue. It is essential to first rule out any systematic errors in the observational methods and ensure that equivalent quantities are being compared.

For instance, Price, S. H. et al. (2022) has shown that using different definitions of radius can significantly impact the inferred dark matter fraction, emphasizing the need to consider these biases for a proper interpretation of the results.

In the upcoming chapter, we will define our observables and thoroughly analyze the various methods used to estimate the dark matter fraction in galaxies, both in observations and simulations, while highlighting the associated biases. Subsequently, we will present the initial results regarding the behavior of SIDM in simulations. Finally, we will compare these results to observations, accounting for the identified methodological biases.

I Methods

For this project, the initial step was to extract relevant information from both the observational data and the simulation data.

I.1 Definition of our observables

The simulation data enabled us to precisely define our observables and examine the biases associated with different definitions.

Galaxy radius

Our project focuses on the dark matter fraction at the centers of galaxies, making it necessary to define the concept of galaxy radius. The observable available in many surveys is the radius containing half of the luminous mass.

This can either correspond to the projected effective radius, R_e, or the 3D radius that encloses half of the luminous mass, $R_{1/2}$. The effective radius, R_e, is defined as the radius within which half of the total light emitted by the object is included when projected onto the sky plane. In other words, if we observe the galaxy as a flat image, the effective radius is the distance from the center of this image where half of the total brightness is concentrated.

On the other hand, $R_{1/2}$ describes the actual spatial distribution of luminous mass (i.e., stellar mass) in a galaxy. It is defined as the radius of the sphere centered on the object's center of mass within which half of the total luminous mass resides. Unlike the effective radius, this measurement is three-dimensional and considers the actual distribution of stellar mass, without projection. Often in studies, $R_{1/2}$ is calculated from the effective radius and a model based on various assumptions of symmetry for deprojection.

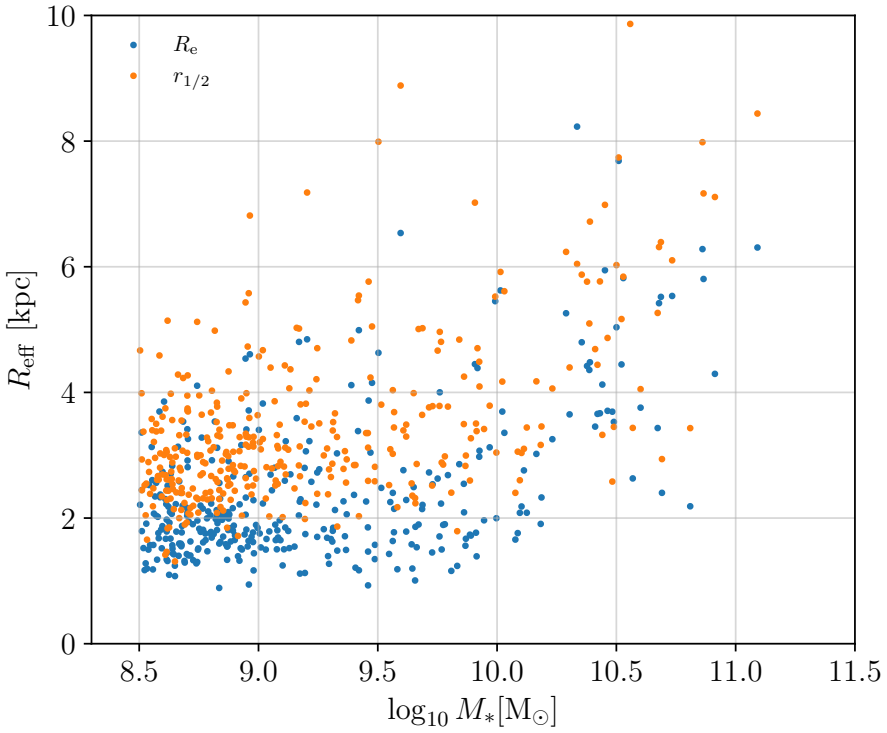


Figure 2.1: Effective radii of each galaxy as a function of stellar mass for the reference CDM model. The blue points represent the projected 2D effective radii R_e and the orange points represent the 3D radius $R_{1/2}$.

Figure 2.1 presents the effective radii of each simulated galaxy as a function of stellar mass for the reference CDM model. It illustrates the difference between the projected 2D effective radii R_e and the 3D effective radii $R_{1/2}$. The Figure shows that $R_{1/2}$ tends to be larger than R_e. This is because the stellar mass contained within R_e can often include some mass that lies outside $R_{1/2}$ in 3D dimensions. This results in an increase in the total mass considered when calculating R_e, explaining why R_e is generally smaller than $R_{1/2}$.

It's important to keep this finding in mind when interpreting observational studies to ensure comparable quantities are being compared.

Dark matter fraction

The dark matter fraction, denoted as $f_{\rm DM}$, is calculated in a very general manner as follows.

$$f_{\rm DM} = \frac{M_{\rm DM}}{M_{\rm tot}} \tag{2.1}$$

where $M_{\rm DM}$ is the dark matter mass and $M_{\rm tot}$ is the total mass.

Therefore, following the previous subsection, we have two possible definitions for the dark matter fraction at the centers of galaxies.

$$f_{\rm DM}(R_{\rm eff}) = \frac{M_{\rm DM}}{M_{\rm tot}}(R_{\rm eff}), \tag{2.2}$$

$$f_{\rm DM}(R_{1/2}) = \frac{M_{\rm DM}}{M_{\rm tot}}(R_{1/2}), \tag{2.3}$$

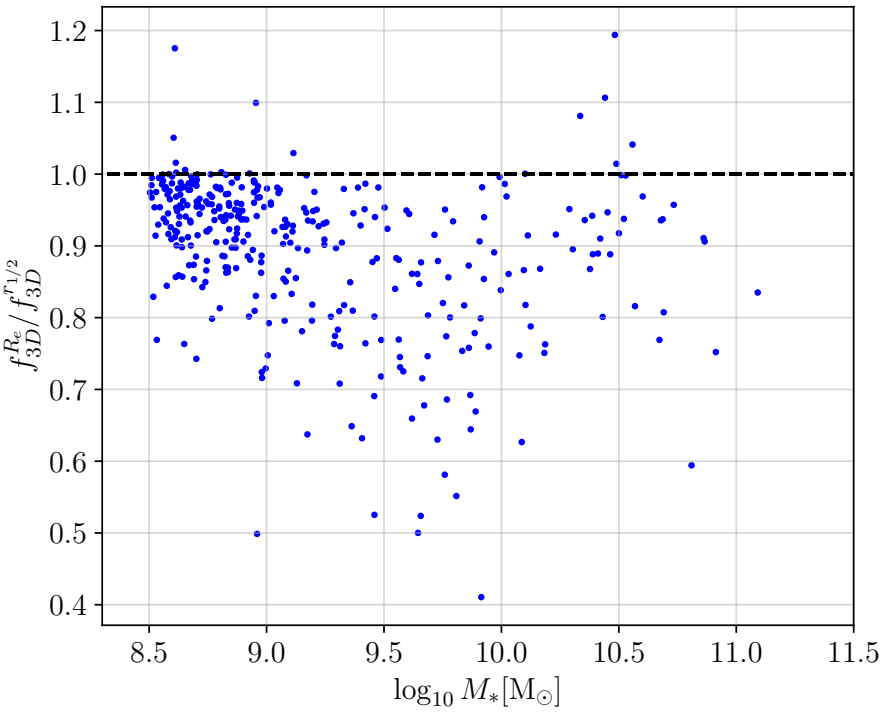


Figure 2.2: Ratio between the 3D dark matter fraction within the effective radius ($R_{\rm e}$) and the 3D dark matter fraction within the 3D half light mass radius ($R_{1/2}$) as a function of stellar mass for each galaxy in the Reference CDM model. It seems that $R_{1/2}$ contains a higher proportion of dark matter compared to $R_{\rm e}$.

Figure 2.2 presents the ratio between the 3D dark matter fraction within $R_{\rm e}$ and the 3D dark matter fraction within $R_{1/2}$ as a function of stellar mass for each galaxy in the Reference CDM model. The 3D dark matter fraction within $R_{1/2}$ appears to encompass a higher proportion of dark matter than the 3D dark matter fraction within $R_{\rm e}$. This is understandable from Figure 2.1: for the same stellar mass, $R_{\rm e}$ encompasses a smaller part of the dark matter halo than $R_{1/2}$ do, leading to smaller fractions of dark matter.

Moreover, the 3D versus 2D issue applies not only to radii but also to the dark matter fraction itself. Consequently, two additional definitions of the dark matter fraction are added to the previous two:

$$f_{\rm 2D}(r < R) = \frac{M^{\rm 2D}_{\rm DM}(r < R)}{M^{\rm 2D}_{\rm tot}(r < R)}, \tag{2.4}$$

$$f_{\rm 3D}(r < R) = \frac{M^{\rm 3D}_{\rm DM}(r < R)}{M^{\rm 3D}_{\rm tot}(r < R)}, \tag{2.5}$$

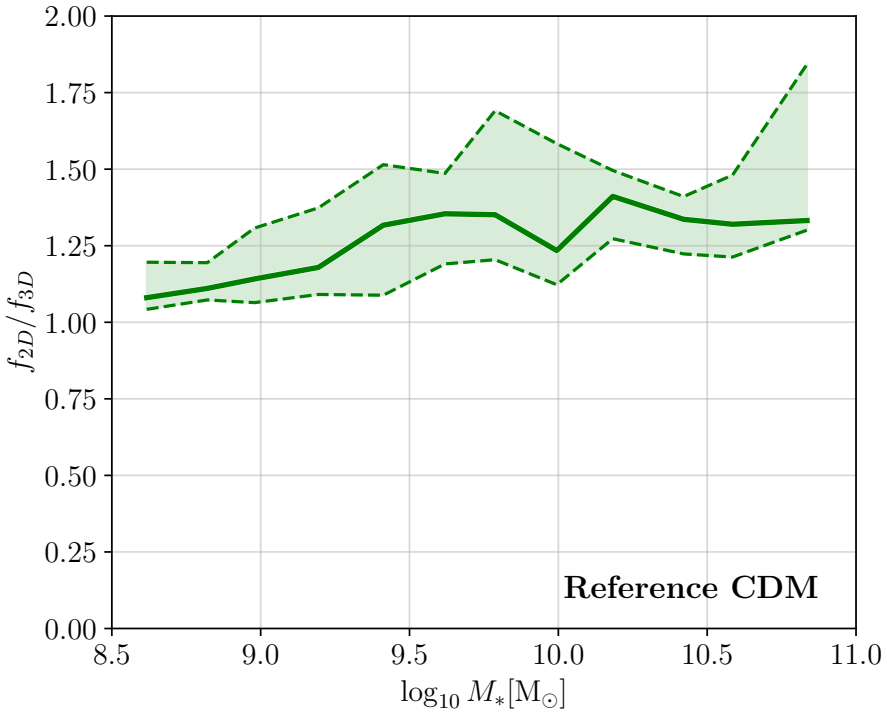


Figure 2.3: Ratio between the 2D projected and 3D dark matter fraction within $R_{\rm e}$ as a function of stellar mass for the reference CDM model. The solid green line corresponds to the median and the shaded region corresponds to the 16-84th percentiles. Considering the 2D projected dark matter fraction can lead to an overestimation of the true dark matter fraction.

Figure 2.3 shows the median ratio between the 2D projected and 3D dark matter fraction within $R_{\rm e}$ as a function of stellar mass for the Reference CDM model. The 2D projected dark matter fraction tends to overestimate the true dark matter content by 10 to 35%. Similar to the previous subsection, in 2D, we consider all mass integrated along the line of sight, without radius restriction, resulting in an overestimation of the dark matter fraction.

Early-type and Late-type galaxies

It's important to distinguish between different types of galaxies in our analysis for consistency. The terminology 'late-type galaxies' refers to spiral and irregular galaxies, while 'early-type galaxies' refers to elliptical galaxies. First, the methods used to calculate masses for early-type galaxies (ETGs) and late-type galaxies (LTGs) from observations are very different, so being aware of this distinction helps to better identify measurement biases. Second, comparing their mass values can be relevant and provide insights into the formation, history, and dynamics of these types.

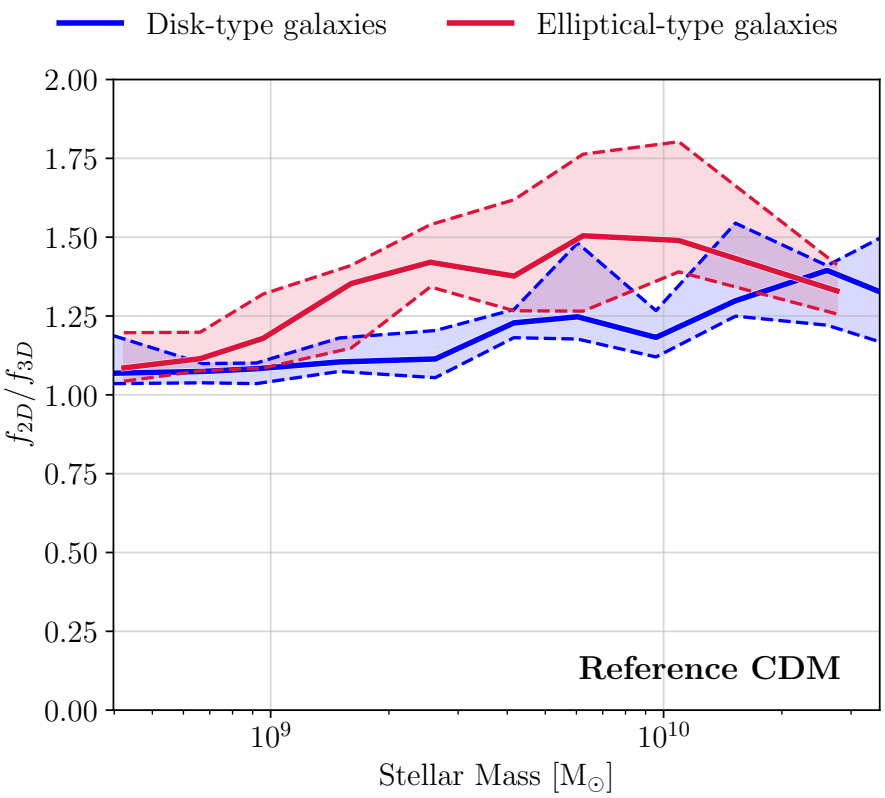


Figure 2.4: Ratio between the 2D projected and 3D dark matter fraction within R_e as a function of stellar mass for the Reference CDM model. The solid red line corresponds to the median for elliptical galaxies (or ETGs), the blue line corresponds to the median for disk galaxies (or LTGs), and the shaded region corresponds to the 16-84th percentiles. Considering the 2D projected dark matter fraction can lead to an overestimation of the true dark matter fraction, especially for elliptical galaxies.

Figure 2.4 shows the ratio between the 2D projected dark matter fraction and the 3D dark matter fraction within R_e, as a function of stellar mass, for elliptical and disk galaxies in the Reference CDM model. Our analysis indicates that the overestimation of the projected dark matter fraction is more prominent in elliptical galaxies. This can be attributed to the differing symmetries of galaxies: in elliptical galaxies, the mass along the line of sight is more significant, whereas in disk galaxies, the mass is mainly concentrated in the plane perpendicular to the line of sight. Consequently, in disk galaxies, dark matter is also more concentrated within the galaxy plane and inside R_e, leading to a ratio f_{2D}/f_{3D} closer to 1 compared to elliptical galaxies. Considering the type of galaxy helps to comprehend the distribution of dark matter in relation to baryonic mass and interpret discrepancies in data due to methodological biases related to galaxy type.

For our research, we used classifications provided by each observational dataset, while for simulation data, we distinguished between ETGs and LTGs following the approach by Correa et al. (2017). Their method is based on calculating the fraction of kinetic energy invested in ordered corotation κ. They show that for $\kappa > 0.3$, the galaxy is classified as an LTG. In our case, the value of κ for each galaxy is provided in the simulation results. In both observations and simulations, we focused on central galaxies rather than satellite galaxies.

Initial Mass Function

It is important to consider the Initial Mass Function (IMF) when studying stellar populations because it determines the distribution of stellar masses at their formation, which directly affects estimates of dark matter fraction and galaxy dynamics. The Salpeter and Chabrier IMFs are commonly used. The Salpeter IMF,

introduced in 1955, assumes a distribution where low-mass stars are much more common than higher-mass stars, meaning that small stars outnumber larger stars, influencing calculations of stellar mass. In contrast, the Chabrier IMF, introduced in 2003, adjusts this distribution by reducing the overrepresentation of very low-mass stars and increasing the proportion of intermediate-mass stars. Using a consistent IMF is crucial for studying dark matter, as different IMFs can lead to different estimates of the dark matter fraction. By using the same IMF in comparisons of observational data, potential biases are reduced, ensuring analyses are based on compatible foundations and allowing for more robust conclusions about the distribution of dark matter.

For instance, Barnabè et al. (2011) found that using stellar masses derived from a Salpeter IMF resulted in an average dark matter fraction of 31%, whereas this fraction increased to 61% when assuming a Chabrier IMF.

Simulation results are based on a Chabrier IMF, so we attempted to convert the stellar masses of our observational data that did not follow this IMF.

The next section details the development of observational catalogues, taking into account all conclusions from this section.

I.2 Catalogues construction

For creating both observed and simulated catalogues, we have chosen to define the dark matter fraction as the "3D dark matter fraction contained within the projected effective radius $R_{\rm e}$." This definition is commonly used in observational data or is the most straightforward to calculate based on the available data. Throughout this project, we will refer to this fraction as $f_{\rm DM}$ $(R_{\rm eff})$.

We compiled an observational sample by joining the catalogues of disk galaxies (or LTGs) from Pizagno et al. (2007), Reyes et al. (2011), and Lelli et al. (2016), resulting in a dataset of 440 disk galaxies with stellar masses within the range of $10^{8.7}$ $M_{\odot}$ to 2 x $10^{11.4}$ $M_{\odot}$. We compiled an observational sample by joining the catalogues of elliptical galaxies (or ETGs) from Barnabè et al. (2011) and Cappellari et al. (2013), resulting in a dataset of 92 galaxies with stellar masses within the range of $10^{9.4}$ $M_{\odot}$ to 2 x $10^{11.9}$ $M_{\odot}$. We added to our study the dataset from Li et al. (2024b) including both ETGs and LTGs and the median 3D dark matter fractions of approximately 4500 ETGs from Tortora et al. (2012).

The dynamical masses of LTGs are primarily calculated using the stars' rotational curves, while the dynamical masses M of ETGs are generally calculated using stars' velocity dispersion using :

$$M = c\frac{r_{\rm r} < \sigma^2_{\rm LOS} >}{G}, \tag{2.6}$$

where $r_{\rm r}$ is the physical radius, $< \sigma^2_{\rm LOS} >$ observed line-of-sight velocity dispersion and c is a structural constant.

SPARC dataset

The Spitzer Photometry and Accurate Rotation Curves (SPARC) database presented in Lelli et al. (2016) comprises 175 LTGs, which include spirals and irregulars. It features Spitzer photometry at 3.6 μm, allowing for the tracing of the stellar mass distribution. Additionally, the database contains high-quality HI + Hα rotation curves. We excluded negative values for size, stellar mass, and dark matter fraction from the sample, resulting in a sample of 157 LTGs.

From Lelli et al. (2016) dataset, for our analysis, we employ a constant mass-to-light ratio of $\Gamma = 0.5 M_\odot/L_\odot$, as determined by stellar population synthesis models (Schombert and McGaugh, 2014) based on a Chabrier IMF. Following Lelli et al. (2016) and utilizing their rotation curve calculations, we compute the fraction of dark matter using, with $\Gamma_{bul} = 1.4\Gamma_{\rm disk}$ and $\Gamma_{\rm disk} = 0.5$:

$$V_{\rm bar}^2 = V_{\rm gas}|V_{\rm gas}| + \Gamma_{\rm disk} V_{\rm disk}|V_{\rm disk}| + \Gamma_{bul} V_{bul}|V_{bul}|, \tag{2.7}$$

$$f_{\rm DM} = \frac{V_{\rm obs}^2(r < R_{\rm e}) - V_{\rm bar}^2(r < R_{\rm e})}{V_{\rm obs}^2(r < R_{\rm e})}, \tag{2.8}$$

Each circular velocity at the effective radius was computed by interpolating the rotational curves.

Pizagno+2007 dataset

The catalogue derived by Pizagno et al. (2007) consists of 163 spiral galaxies featuring resolved H-alpha rotation curves. It provides the effective radius and circular velocity at the effective radius of each galaxy. We excluded negative values for size and total mass, resulting in a sample of 162 spiral galaxies.

From Pizagno et al. (2007) dataset, we utilized the effective radius and circular velocity $V_{\rm circ}$ at the effective radius $R_{\rm e}$ directly from this catalogue to estimate total masses.

$$M_{\rm tot}(r < R_{\rm e}) = \frac{V_{\rm circ}^2(r < R_{\rm e}) R_{\rm e}}{G}, \tag{2.9}$$

We estimated stellar masses using the i-band magnitudes, assuming a constant I-band mass-to-light ratio of 1.2, $M_* = 1.2$ x $10^{0.4(i_\odot - i)}$ with $i_\odot = 4.11$. The mass-to-light ratio is adopted for a Chabrier IMF and it assumes the contribution of disk+bulge (Portinari et al., 2004). The effective radius for this sample is defined as the radius at 2.2 × $\rm R_{disk}$, where $\rm R_{disk}$ is the disk exponential scale length.

To be consistent with other samples, we estimated the stellar mass within the half-mass radii by dividing by 2 the total stellar masses. To convert from half-mass radii to half-light radii we follow Graaff et al. (2021), and Suess et al. (2019) who have concluded that half-light radii are larger than half-mass radii by typically 25%.

The dark matter fraction is then calculated as follows.

$$f_{\rm DM} = \frac{M_{\rm tot}(r < R_{\rm e}) - M_*(r < R_{\rm e})}{M_{\rm tot}(r < R_{\rm e})}, \tag{2.10}$$

Reyes+2011 dataset

Reyes et al. (2011) provides a thorough analysis to provide a more accurate and refined estimate of disk rotation velocities for a specific subset of 189 galaxies from the Sloan Digital Sky Survey (SDSS). This dataset contains the i-band Petrosian half-light radius, r-band Petrosian absolute magnitude (M_r), and g - r color. All values are k-corrected to z = 0 and adjusted for Galactic and internal extinction.

From Reyes et al. (2011) dataset, stellar masses were estimated following Bell et al. (2003) and converted to a Chabrier IMF. Then, we estimated the stellar mass within the half-mass radii following the same process used in Section I.2. To estimate the rotation velocity at $R_{\rm e}$ we used the arctangent model and corrections applied in Correa et al. (2024). Then, we computed $f_{\rm DM}$ using equations 2.9 and 2.10.

Barnabe+2011 dataset

Barnabè et al. (2011) provides a dataset of sixteen early-type lens galaxies from the SLACS Survey, at z = 0.08 - 0.33. They analyzed the galaxies' inner regions, i.e., within one effective radius R_e. They assumed both, a Chabrier and a Salpeter IMF. We construct our dataset using their dark matter fractions based on a Chabrier IMF. Their stellar mass values are taken from the stellar population analysis performed by Auger et al. (2009) (also under Chabrier IMF). Barnabè et al. (2011) calculated the dark matter fraction as follows.

$$f_{\rm DM} = 1 - \frac{M_*(r < R_e)}{M_{\rm tot}(r < R_e)}, \tag{2.11}$$

Tortora+2012 dataset

In Tortora et al. (2012), the median 3D dark matter fractions of approximately 4500 ETGs within a projected 2D effective radius are provided as a function of stellar masses ranging from 1.3×10^{10} to 3.2×10^{11} $M_\odot$ at redshift z from 0.05 to 0.095. The study uses spherically symmetric isotropic galaxy models and a constant mass-to-light ratio based on a Chabrier IMF.

Atlas3D dataset

Cappellari et al. (2013) sample consists of 258 ETGs (elliptical and lenticular galaxies) based on the Atlas3D Project. The stellar mass range covered by this dataset is $>= 6 \times 10^9$ $M_\odot$. The dataset provided stellar mass-to-light ratios based on a Salpeter IMF and luminosities in the r-band. Cappellari et al. (2013) employs the model "JAM (Jeans Anisotropic Modelling) with NFW dark halo" and assumes that the halo is spherical, characterized by a two-parameter double power-law NFW profile (Navarro et al., 1996). They adopt the mass-concentration M_{200}-c_{200} relation provided by Klypin et al. (2011) to make the halo profile a unique function of its mass M_{200}. This allows them to extract the galaxies' 3D dark matter fractions within a projected 2D effective radius R_e and stellar mass-to-light ratios.

We calculated the stellar masses and effective radii using mass-to-light ratios $(M_*/L)_{\rm S}$ based on a Salpeter IMF and luminosities L in the r-band from Cappellari (2013). It's important to note that they also use a Salpeter IMF for their calculation of $f_{\rm DM}$. It will be interesting to study later how converting masses from Salpeter to Chabrier affects the dark matter fraction. In order to perform this conversion as well. The following methodology is proposed based on Li et al. (2024b).

$$M_* = (M_*/L)_{\rm S} \times L - 0.25, \tag{2.12}$$

$$f_{\rm DM,C} = 1 + 10^{-0.25} \times (1 - f_{\rm DM,S}), \tag{2.13}$$

MaNGA dataset

The dataset from Li et al. (2024b) includes almost 6000 galaxies, encompassing both ETGs and LTGs from the MaNGA DynPop project. These galaxies have stellar masses ranging from 10^9 to 10^{12} solar masses ($M_\odot$). Li et al. (2024b) conducted an analysis providing information such as lower limits of 3D dark matter fraction

within a projected 2D effective radius R_e, r-band total luminosities, R_e, and a quality flag. Following their methodology, we selected galaxies with a quality flag greater than 1. From this initial selection, we further refined our criteria by requiring the difference in the mass-weighted total density slope between JAMcyl and JAMsph modeling to be smaller than three times 0.079. Additionally, we included a filtering condition: $|f_{\mathrm{DM,cyl}} - f_{\mathrm{DM,sph}}| < 0.1$, to enhance the accuracy of the dark matter fraction estimations. This dataset contains a total of 5688 galaxies.

First, we selected only central halos based on the central-satellite classification results from Yang et al. (2007). To compute stellar masses, we used r-band total luminosity L and mass-to-light ratio $(M_*/L)_S$ based on a Salpeter IMF from Lu et al. (2023). Then, we use equations 2.12 and 2.13 to convert to a Chabrier IMF. After excluding the NaN values, the catalogue now includes 2533 galaxies.

II Dark matter fraction's dependencies

After setting up our datasets and variables, we analyze the results of the simulations to understand the behavior of SIDM. In this section, we examine the main characteristics of simulated galaxies from the Reference and WeakStellarFB models within the three different dark matter cosmologies discussed in the introduction: CDM and SIDM with two different velocity-dependent cross sections.

II.1 Mass and compactness

Figure 2.5 investigates the relation between the galaxies' effective radius, stellar mass, and their inner fraction of dark matter from the WeakStellarFB (top panels) and Reference (bottom panels) models under the three dark matter cosmologies. The dark matter fraction of galaxies is represented in Figure 2.5 as a function of the effective radius, colored by the logarithm of the galaxies stellar mass.

The figure shows an anti-correlation between the fraction of dark matter and the galaxies stellar mass, so that at fixed R_{eff}, higher mass galaxies have lower fractions of dark matter in their inner region.

When we compare the Reference and WeakStellarFB models, we find that the WeakStellarFB model produces galaxies with a lower dark matter fraction than the Reference model regardless of the dark matter model. These galaxies have a radius generally smaller than 3 kpc and are more massive. At fixed stellar mass, galaxies from the Reference model have a larger effective radius than galaxies from the WeakStellarFB model. This is due to less efficient stellar feedback in the WeakStellarFB model, which produces more compact galaxies (see also Correa et al., 2024).

Figure 2.5 shows that at fixed stellar mass, the correlation between dark matter fraction and effective radius is strong. The more compact a galaxy is, the lower its dark matter fraction is. When we compare the top panels of Figure 2.5 from left to right we find that galaxies become less compact. This is probably due to collisions between dark matter particles, which produce dynamical heating in the center of galaxies, making them to grow more extended in size.

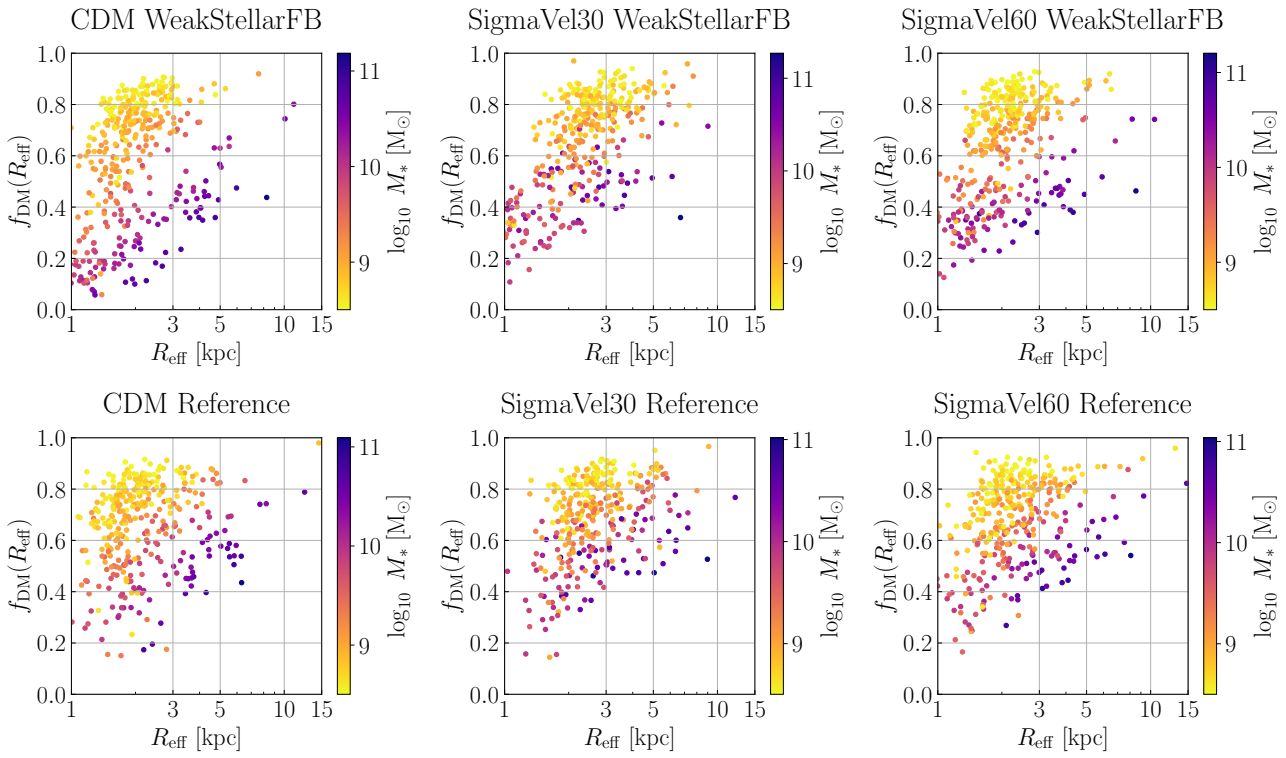


Figure 2.5: Dark matter fraction $f_{\rm DM}$ within the effective radius as a function of effective radius of galaxies colored by the logarithm of their stellar mass. The panels correspond to the WeakStellarFB models (top panels) and the Reference models (bottom panels) under CDM (left panels), SigmaVel30 (middle panels), and SigmaVel60 (right panels). The differences in mass among galaxies of same radius establish a link between galaxy compactness and the fraction of dark matter within their effective radius.

II.2 Baryonic and dark matter physics

Figure 2.6 shows the medians dark matter fractions within the effective radius from the Reference and WeakStellarFB models, for three different dark matter cosmologies. These medians are plotted as a function of the logarithm of the stellar mass, and each galaxy is represented by a dot colored according to its $R_{\rm eff}$. For every model, at a stellar mass approximatly higher than $10^{9.5}$ $M_\odot$ the medians tend to be constant. Rather than decreasing as the proportion of stellar mass increases, the fraction of dark matter remains the same. At a such stellar mass value, DM particles tend to collapse no matter the model to follow the stellar content increasing.

The galaxies in the Reference model with a stellar mass as high as $10^{9.5}$ $M_\odot$ exhibit higher dark matter fractions compared to the WeakStellarFB model under the CDM cosmology. This difference is attributed to the compactness of galaxies resulting from weak stellar feedback. The figure illustrates in the top panels that galaxies with very small effective radii (less than 1.5 kpc) tend to lower the median dark matter fraction. This is because those galaxies encompass a smaller part of the dark matter halo, leading to a notable reduction in the median dark matter fraction.

Figure 2.6 and 2.5 show that dark matter particules interactions are not sufficient to counteract the gas overcooling and size compactness in galaxies from the WeakStellarFB model as discussed in Correa et al. (2024). However, in this model the median dark matter fractions increase from 25% to 40% when moving from CDM to SIDM cosmology, despite no change in galaxy properties. SIDM particles thermalise through frequent

interactions and accumulate in the center, resulting in a higher dark matter fraction.

In the lower panels, in the reference model, SIDM particles create galaxies that are more extended. From a mass of $10^{9.7}$ $M\odot$, the median fraction of CDM is just below 0.5, while in the other two SIDM models, it is slightly above 0.5. The increased concentration of dark matter at the center, observed in the WeakStellarFB model when moving from CDM to SIDM, is less present here. The galaxies are more extended, their gravitational potential is weaker, and this leads to less concentration.

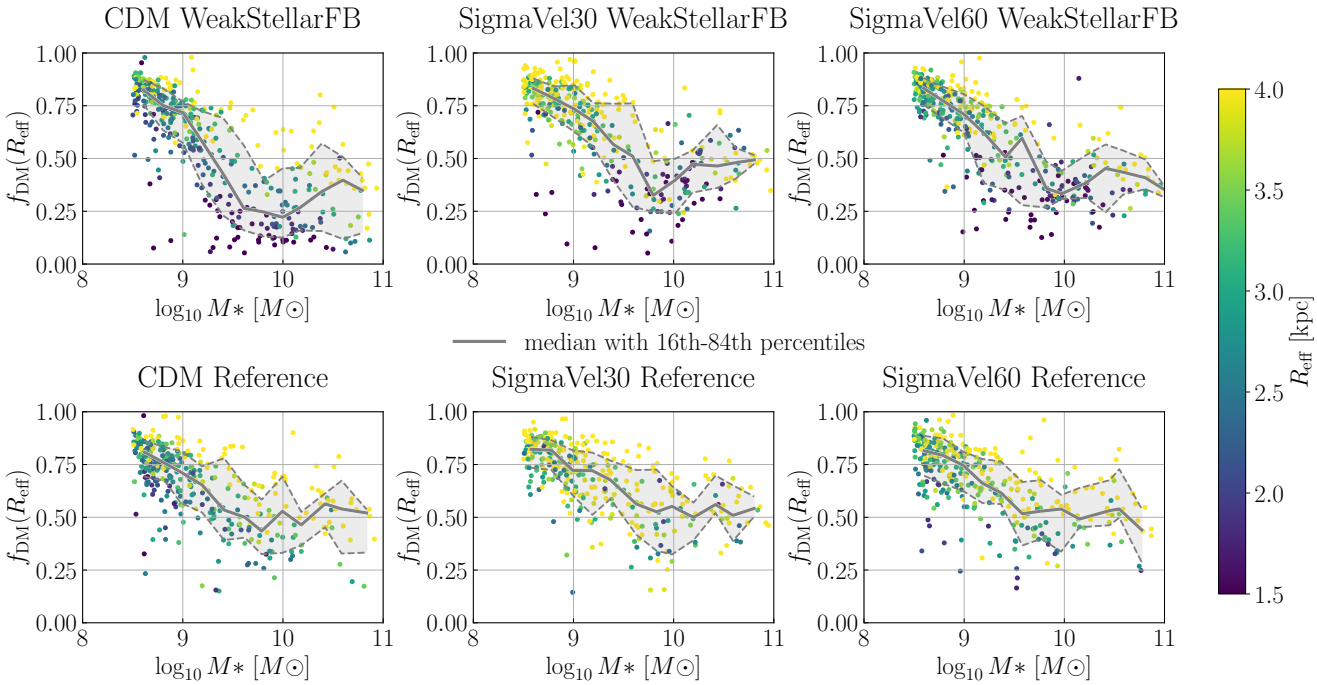


Figure 2.6: Dark matter fraction within the effective radius as a function of the logarithm of the stellar mass under CDM (left panels), SigmaVel30 (middle panels), and SigmaVel60 (right panels). The solid lines indicate the median relations, in the top panels for the WeakStellarFB model and in the bottom panels for the Reference model, while the shaded regions highlight the 16th-84th percentiles. The dots represent the data upon which these relations are based and are colored by their $R_{\rm eff}$.

II.3 Galaxy types

We investigated the impact of galaxy types on the galaxies' dark matter fraction. Figure 2.7 shows the medians of the dark matter fractions from the Reference and WeakStellarFB models for the three different dark matter cosmologies as a function of stellar masses. The solid red line represents ETGs and the solid red line represents LTGs. In the Reference model, ETGs tend to have a lower dark matter fraction in their inner region at a fixed size than LTGs. In the WeakStllarFB model, this gap is weaker. This trend is due to the compactness distribution of galaxies in each model. Galaxies in the WeakStellarFB model are generally more compact, regardless of their type. In the Reference model, ETGs are more compact than LTGs, with LTGs having an effective radius approximately 60% larger than that of ETGs.

Separating ETGs and LTGs in our sample serves two important purposes: it enables us to account for variations in galaxy compactness and brings attention to a potential issue with the methodology. This is because the observed dark matter fractions for ETGs and LTGs are obtained using very different methods.

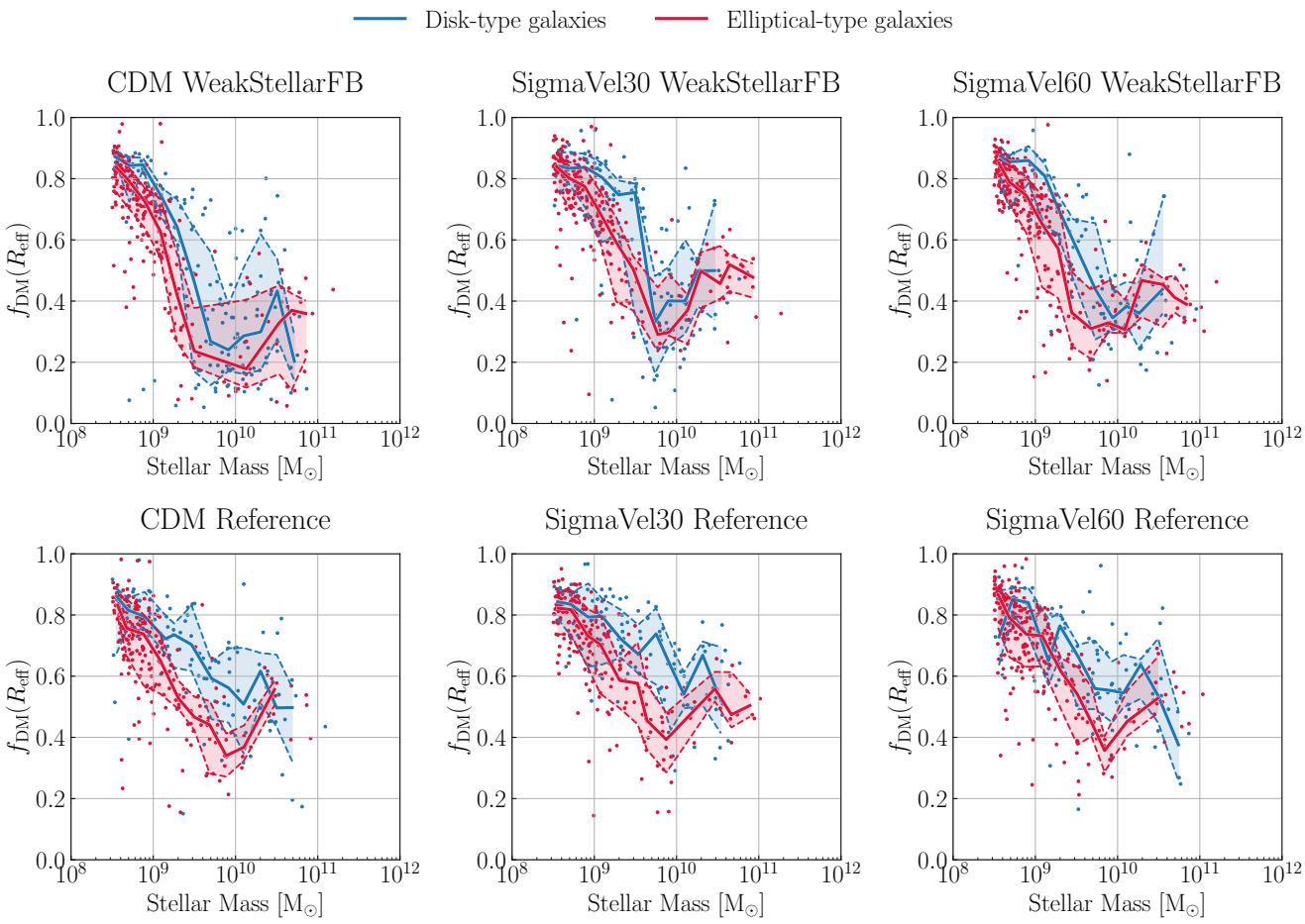


Figure 2.7: Dark matter fraction within the effective radius as a function of stellar mass for ETGs (red solid line) and LTGs (blue solid line). The panels correspond to the WeakStellarFB models (top panels) and the Reference models (bottom panels) under CDM (left panels), SigmaVel30 (middle panels), and SigmaVel60 (right panels). Solid lines represent the median, while the shaded regions highlight the 16th-84th percentiles.

III Comparison with observations

Once we have a good understanding of SIDM behavior, it is important to compare simulation results with observations to constrain our models. The following section is divided into LTGs and ETGs.

III.1 Dark matter fraction in LTGs

We compare the observational sample in the mass-$f_{\rm DM}$ plane with a sample of LTGs taken from the simulations. As seen in the section II.1, galaxies' compactness impacts their dark matter fraction so to be consistent we selected galaxies with the same range of sizes. Figure 2.8 shows the dark matter fraction of galaxies, with an effective radius ranging from 1 to 3 kpc, as a function of stellar mass for LTGs (black solid line) from the Reference (lower panels) and WeakStellarFB (upper panels) models under CDM (left panel), SigmaVel30 (middle panel) and SigmaVel60 (right panel). The solid lines represent the median relations, and the shaded

regions depict the 16-84th percentiles. The observational sample is shown in symbols, and its median relation is depicted in a solid blue line. The panels in Figure 2.8 indicate that two WeakStellarFB models are consistent with observations, since they lie within the 1 σ scatter. None of the Reference models fit the observations, suggesting that the chosen range of SIDM cross sections is inappropriate. This issue will be addressed in Chapter 3.

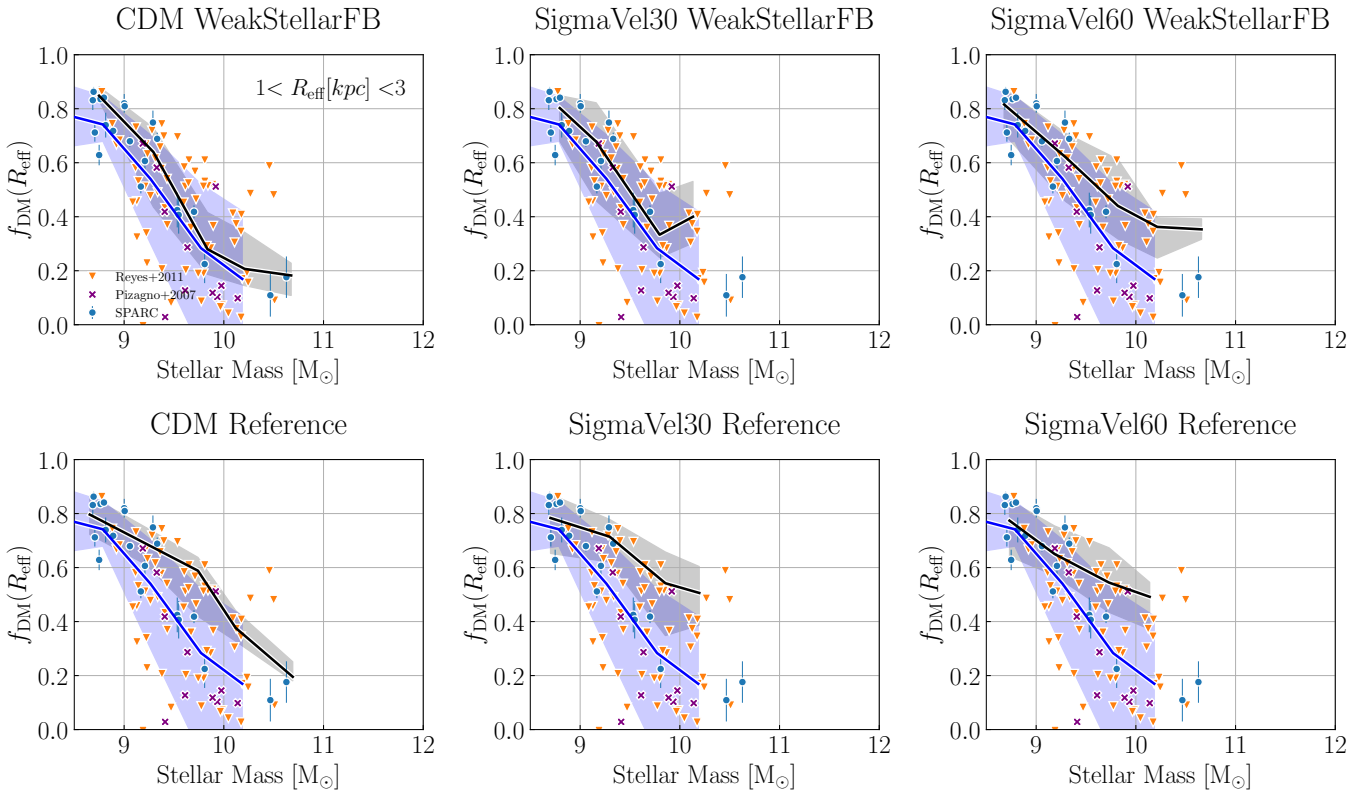


Figure 2.8: Dark matter fraction within the effective radius as a function of stellar mass. The panels correspond to the WeakStellarFB models (top panels) and the Reference models (bottom panels) under CDM (left panels), SigmaVel30 (middle panels), and SigmaVel60 (right panels). Solid lines represent the median, and the shaded regions indicate the 16th-84th percentiles. Black lines correspond to the median for simulated LTGs, while blue lines correspond to the median for observed LTGs. Each observed LTG is represented by a dot.

III.2 Dark matter fraction in ETGs

Studies with observed dark matter fractions for ETGs are too inconsistent with each other to be compared with simulated data. Figure 2.9 shows the dark matter fraction of galaxies as a function of stellar mass for ETGs from Barnabè et al. (2011), Tortora et al. (2012), Cappellari et al. (2013), and Li et al. (2024b). We kept the dark matter fraction from Tortora et al. (2012) and Li et al. (2024a) based on a Salpeter IMF. The MaNGA dataset exhibits a wide range of dark matter values and includes both ETGs and LTGs. Effective radii for galaxies are not provided by every dataset, so we cannot make a selection based on size. Only Li et al. (2024b) for the MaNGA dataset provides $R_{\rm e}$, and we noted that every galaxy with a dark matter fraction over 40% has an effective radius larger than 3 kpc. Barnabè et al. (2011) dataset diverges from the other datasets with very high dark matter fractions. Cappellari et al. (2013) and Li et al. (2024b) seem consistent with each other, while Tortora et al. (2012) has a slightly higher median dark matter fraction.

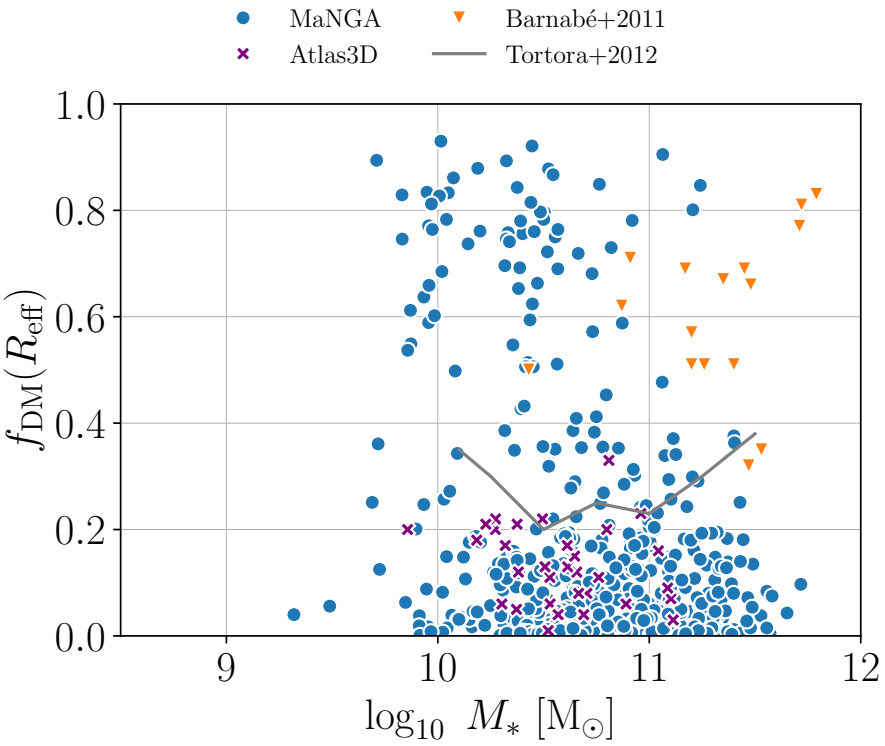


Figure 2.9: Dark matter fraction of galaxies as a function of stellar mass for ETGs from Barnabè et al. (2011), Tortora et al. (2012), Cappellari et al. (2013), and Li et al. (2024b).

III.3 Lower ETGs' $f_{\rm DM}$

We found that observed dark matter fractions in LTGs are higher than in ETGs. The LTGs sample exhibits a very wide range of dark matter fraction values. While the dark matter fractions from Cappellari et al. (2013) and Li et al. (2024b) are computed using a relatively similar stellar dynamical modeling technique, the dark matter fractions from Barnabè et al. (2011), which are nearly six times higher, are calculated using gravitational lensing. Dekel et al. (2005) suggests that this missing dark matter content can be due to projection effects. The study shows that during disk galaxy mergers, the stellar orbits in the outer regions of the resulting ellipticals become very elongated. This elongation occurs because the orbits were torn by tidal forces from their original galaxies during the first close passage, placing them on outgoing trajectories. In this context, the approximation "line-of-sight velocity dispersion equals one-third of the dispersion in the other two directions" used to calculate the velocity dispersion leads to an underestimation of the total mass and therefore an underestimation of the dark matter fraction. This missing dark matter content can also be due to the IMF used in Tortora et al. (2012) and Li et al. (2024a).

An other explanation comes from the equation used by Cappellari et al. (2013) to computed their total mass.

$$M = c\frac{r_{\rm r} < \sigma^2_{\rm LOS} >}{G}, \tag{2.14}$$

where $r_{\rm r}$ is the physical radius, $< \sigma^2_{\rm LOS} >$ observed line-of-sight velocity dispersion and c is a structural constant. Cappellari et al. (2006) found $c = 2.5$ by calibration while Wolf et al. (2010) analytically derived $c = 4$. Using $c = 4$ can lead to higher dark matter fractions.

IV Discussion

In this chapter, we examined the behavior of SIDM and identified 7 key points from our study:

- The WeakStellarFB model results in more compact galaxies.
- At around $10^{9.5}$ $M_{\odot}$, dark matter experiences such strong gravitational potential that it only aggregates regardless of the model.
- SIDM produces more extended galaxies in the Reference model but does not counteract the tendency of the WeakStellarFB model to make galaxies more compact.
- Unlike dwarf galaxies where SIDM can form a core and "empty" the galaxy center of dark matter, in massive galaxies the opposite behavior is observed, and interactions between dark matter particles lead to aggregation.
- The dark matter fraction alone is insufficient to distinguish between the Reference CDM and Reference SIDM models, as it is also necessary to consider the size of galaxies, which is influenced by the nature of dark matter in the Reference model.
- It is crucial to conduct studies on a specific type of galaxy rather than mixing them.
- ETGs studies do not yield reliable dark matter fractions in massive galaxies. Further investigations are necessary to understand and analyze biases specific to this type of galaxies.
- The ranges of values for SIDM cross sections are inconclusive at this stage.

The last point leads us to consider a method that could constrain the cross section values, not to increase the dark matter fraction, but to decrease it to align with observations. Therefore, we must move away from these extreme SIDM scenarios and find a cross section that prevents the situation where particles thermalize and aggregate. We will address this point in the next chapter.

Chapter 3

Cross section

The matter fraction is an observable that can help us constrain the cross section of dark matter particles. Therefore, two methods have been developed to constrain the SIDM model: a semi-analytic model and a machine learning model. Both methods are presented in this chapter.

I Semi-analytic model

I.1 Method

The workflow

We used a semi-analytic model to constrain the cross section. At a certain radius r_1, SIDM particles starts to behave like CDM due to the low density, causing an average dark matter particle to interact only once over the halo's entire lifetime t_{age} (Kaplinghat et al., 2016).

$$\frac{4}{\sqrt{\pi}}\rho_{\text{dm}}(r_1)v(r_1)\sigma = \frac{1}{t_{\text{age}}}, \tag{3.1}$$

On the left side of the equation, we have the scattering rate per particle. ρ_{dm} represents the density of dark matter, $\frac{4}{\sqrt{\pi}}v$ captures the average relative velocity of dark matter particles based on a Maxwellian distribution (with v the one-dimensional velocity dispersion) and σ indicates the cross section per unit mass of the dark matter particles.

We are basing our work on the semi-analytic model developed by Jiang et al. (2023). This model solves the equation 3.1 to find r_1 for a given cross section and a given CDM density profile and enables the calculation of the SIDM density profile. Our modification of the model involves using an additional observable, the dark matter fraction within R_{eff}, to determine the cross section instead of having it as an input parameter.

We determine $\rho(r_1)$ by calculating the radius r_1 where the density profile yields a dark matter fraction at R_{eff} that is closest to the one's observed galaxy. To compute this density profile, we followed Jiang et al. (2023) and computed the adiabatically contracted halo profile given a CDM halo described by an NFW profile (Navarro et al., 1996) and given a final baryon density profile described by a Hernquist profile ρ_b (Hernquist,

1990). Then, the code solves the equation 3.2 using the adiabatically contracted halo profile.

$$v^2 \nabla \ln \rho_{\rm iso} = -4\pi G(\rho_{\rm iso} + \rho_b), \tag{3.2}$$

In practice we used the public code provided by Jiang et al. (2023) to computed the NFW, Hernquist and contracted halo profiles and finally we found the dark matter particles' cross section of a given galaxy.

In summary, the developed semi-analytic model is based on the following workflow:

- Modeling the initial CDM density profile using an NFW profile (Navarro et al., 1996) given a halo concentration computed using Correa et al. (2015) and the galaxy virial mass provided by the simulated catalogues.
- Modeling the galaxy disk density profile using a Hernquist profile (Hernquist, 1990) given a galaxy baryonic mass.
- Calculating the contracted CDM density profile given the baryonic density profile and a NFW profile.
- Finding the value of r_1 by searching for the radius that allows for calculating a SIDM isothermal profile able of reproducing the observed dark matter fraction.
- Calculating the SIDM isothermal profile with the correct r_1 given a galaxy baroynic density profile and a contracted CDM profile.
- Finding the cross section of dark matter particles of a given galaxy by solving equation 3.1.

Limits

In this model, there are several limitations to consider. First, the model we have outlined does not account for the velocity-dependent aspect of the cross section. This is an area that needs improvement in the continuation of the project. For now, the cross section we tried to predict with this semi-analytic model, is the average cross section of dark patter particles within $R_{\rm eff}$. Furthermore, before determining the cross section predicted by the model, it needs to be tested. We encountered several issues during the model testing that had to be resolved. As a result, the final objective could not be fully achieved within the allotted time for the completion of this report. In the rest of this chapter, we will outline the rigorous tests performed to evaluate and refine the model.

I.2 Results

The first step was to verify whether the model could reproduce the density curves from our simulations. To achieve this, we assessed the basic model proposed by Jiang et al. (2023) to see if it could first accurately recover the dark matter fraction within $R_{\rm eff}$.

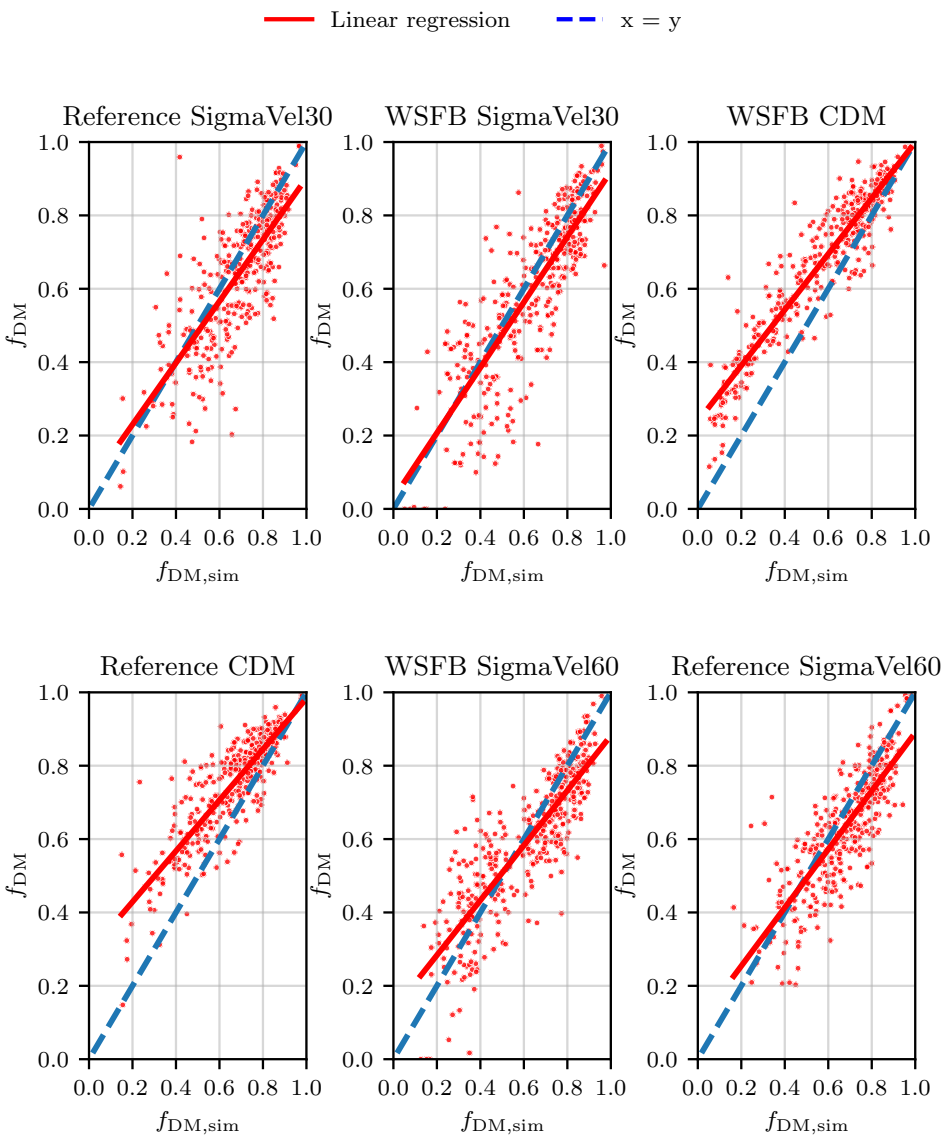


Figure 3.1: Dark matter fraction f_{DM} within the effective radius predicted by the semi-analytic model as a function of the actual dark matter fraction $f_{\mathrm{DM,sim}}$ within the effective radius for each simulated galaxy (red dots). Solid red line represents the linear regression applied to the data points for the six simulation models (listed in order from the top left panel to the bottom right: Reference SigmaVel30, WeakStellarFB SigmaVel30, WeakStellarFB CDM, Reference CDM, WeakStellarFB SigmaVel60, and Reference SigmaVel60).

Figure 3.1 show the dark matter fraction f_{DM} within the effective radius predicted by the semi-analytic model as a function of the dark matter fraction $f_{\mathrm{DM,sim}}$ for each simulated galaxy from the six simulation models. The solid red line represents the linear regression applied to the data points, while the blue line corresponds to the equation $f_{\mathrm{DM}} = f_{\mathrm{DM,sim}}$. The model takes into account the average cross section of dark matter particles within the effective radius, the stellar mass, the virial mass, the effective radius and the dark matter fraction within within the effective radius of the galaxy.

The model is unable to replicate density profiles that match the dark matter fractions of our simulated galaxies. We used three profiles estimated by the model - the Hernquist profile for the baryonic mass, the NFW profile, and the SIDM isothermal profile. We tried to identify which of these profiles was causing the problem. We checked if the stellar mass within the effective radius was the same as in our simulations. Since the issue does not stem from the Hernquist profile, it may be originating from either the NFW density profile or the SIDM density profile. Given that CDM simulation models do not use the SIDM density profile generated by

the code and this semi-analytic model also fails for this cosmology, we assume the problem primarily lies with the NFW density profile.

The NFW density profile is determined by the virial mass and the halo concentration c. Therefore, the halo concentration obtained from the formula by Correa et al. (2015) needs to be adjusted using our simulation models. This is necessary to ensure that the semi-analytic model accurately reproduces the SIDM particles cross section of our simulated galaxies. This adjustment is also crucial to provide reliable predictions for the SIDM particles cross section of observed galaxies.

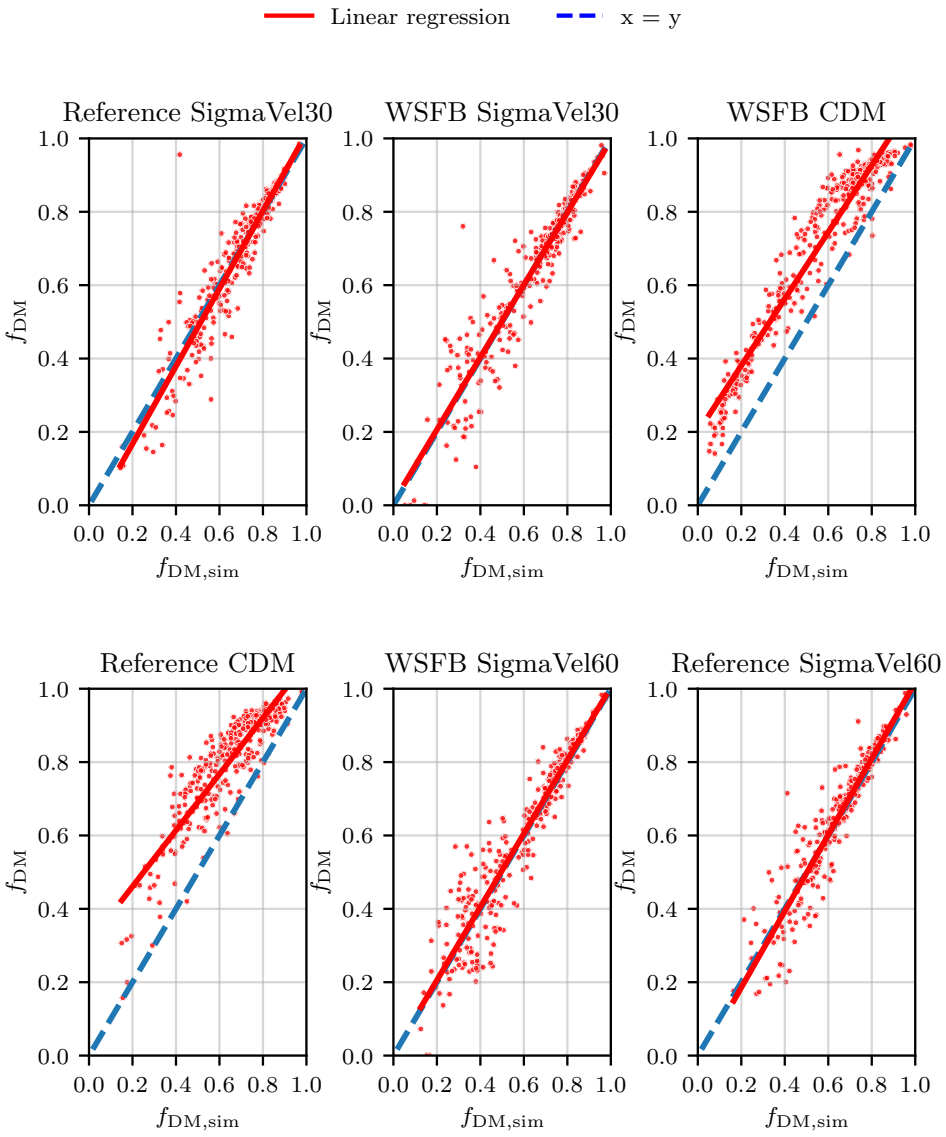


Figure 3.2: Dark matter fraction f_{DM} within the effective radius predicted by the semi-analytic model as a function of the actual dark matter fraction $f_{\mathrm{DM,\,sim}}$ for each simulated galaxy (red dots), now calibrated based on the concentration predicted by the semi-analytic model. Solid red line represents the linear regression applied to the data points for the six simulation models (listed in order from the top left panel to the bottom right: Reference SigmaVel30, WeakStellarFB SigmaVel30, WeakStellarFB CDM, Reference CDM, WeakStellarFB SigmaVel60, and Reference SigmaVel60).

We developed a method to find the concentration c that matches the observed dark matter fraction. The code is based on searching concentration c that minimizes the function $|f_{\mathrm{DM}} - f_{\mathrm{DM,sim}}|$. Figure 3.2 is identical to Figure 3.1, but the concentration used is the one obtained through the minimization process.

The calibration worked well except for CDM cosmologies. This aspect requires further exploration in the

future. Nevertheless, the calibration is effective for the most relevant models in our context in predicting the cross section of SIDM particles. We discuss potential directions in the next section.

I.3 Outlook and future directions

The next steps will involve testing the now calibrated semi-analytic model on our observational data to determine if the rest of the workflow can predict a valid cross section, and eventually match the potential errors and refine the semi-analytic model. Finally, the last step will be to find a method to generalize the calibration so that it can be applied to observations.

II Machine learning approach

In addition, we also tested a machine learning approach alongside the semi-analytic model. Machine learning focuses on developing algorithms and models capable of learning from data and making predictions or decisions based on it. Our aim is to use galaxy properties such as stellar mass, effective radius R_{eff}, dark matter fraction f_{DM} and galaxy type (early-type or late-type), to train a machine learning algorithm on simulation data. This would enable us to predict the cross section of dark matter particles in observed galaxies.

The primary goal is not to constrain the cross section values, but rather to determine whether the galaxy properties provided to the machine learning algorithms are sufficient to distinguish between different dark matter cosmologies.

II.1 Method

We had four features to train our models and one output : stellar mass, effective radius R_{eff}, dark matter fraction f_{DM} and galaxy type (early-type or late-type) and the cross section known in our simulation. The value of the cross section taken is the average cross section of dark patter particles within R_{eff}.

Visualisation

Before deploying machine learning methods, it is essential to follow several steps. We first normalized our data between. Then, visualization of our simulated data plays a crucial role in comprehending the structure of our data, which is necessary for effectively understanding and interpreting the results from the machine learning algorithms. Understanding the link between our data also helps in selecting the appropriate machine learning algorithm.

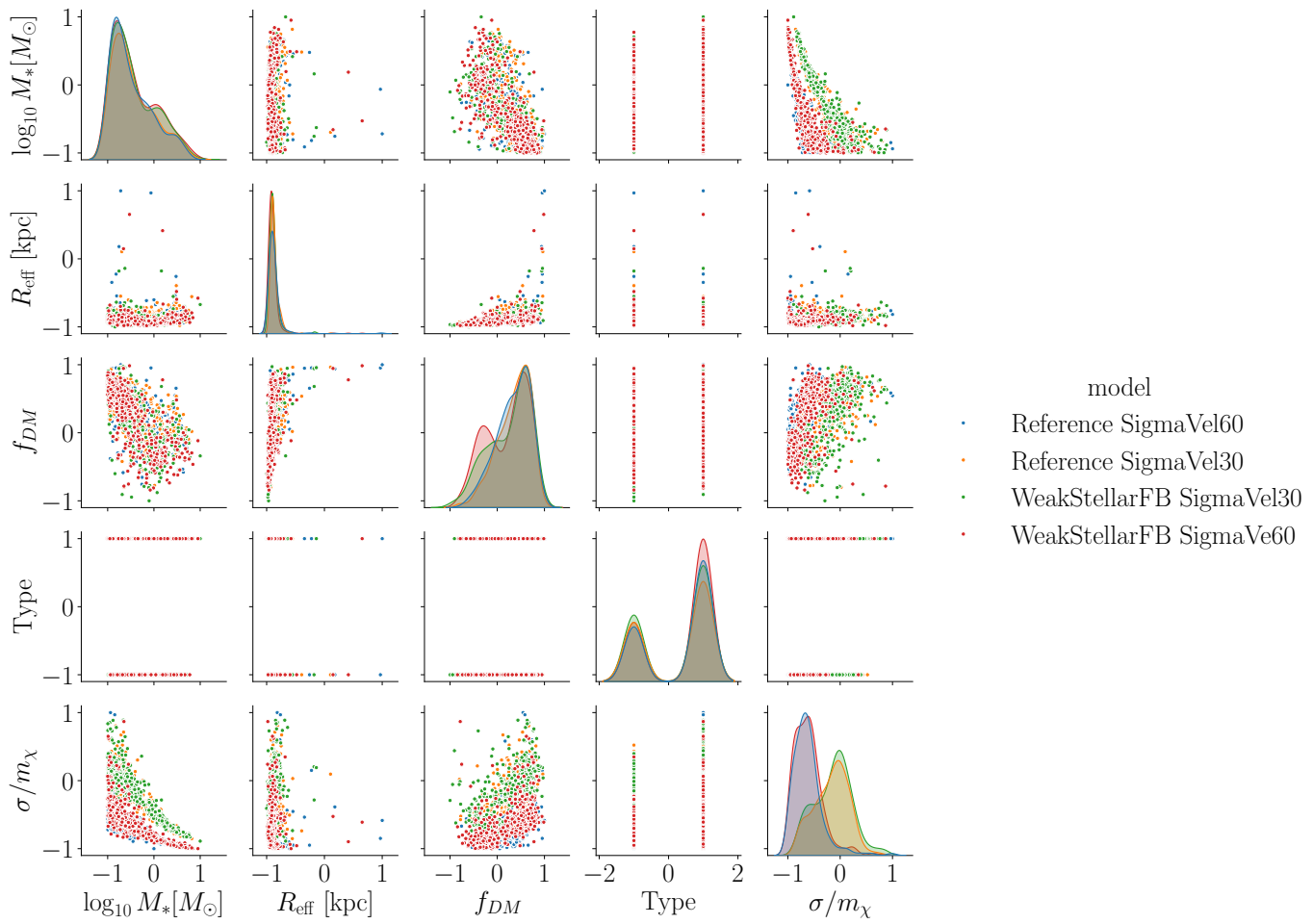


Figure 3.3: Pair plot of four training features (stellar mass, effective radius $R_{\rm eff}$, dark matter fraction within the effective radius $f_{\rm DM}$ and galaxy type) and the output (average cross section of dark matter particles within the effective radius). Blue, orange, green, and red dots correspond, respectively, to galaxies from the Reference SigmaVel60, Reference SigmaVel30, WeakStellarFB SigmaVel30, and WeakStellarFB SigmaVel60 models.

Figure 3.3 shows a pair plot of four training features and the output, σ/m_χ. Each point represent a simulated galaxy. This figure allows for a comparative analysis of how each feature is linked to the others. Each diagonal subplot shows the distribution of a single simulated variable, while the off-diagonal subplots display the correlations between pairs of simulated variables. Four models are represented: the two SIDM cosmologies under the Reference and WeakStellarFB models.

We previously highlighted that SIDM allows for an increased matter fraction in the WeakStellarFB model compared to CDM because, for galaxies with identical properties, dark matter particles thermalize through frequent interactions and accumulate in the center. However, distinguishing between SigmaVel60 and SigmaVel30 was not straightforward when only looking at dark matter fraction. In Figure 3.3, the distribution of dark matter fractions for the WeakStellarFB SigmaVel60 model shows a peak at lower dark matter fractions. This indicates that there are more galaxies with lower dark matter fractions compared to the WeakStellarFB SigmaVel30 model. Thus, the galaxies contributing to the median dark matter fraction lower than 0.5 in Figure 2.6 for the WeakStellarFB SigmaVel60 and WeakStellarFB SigmaVel30 models are more abundant in the WeakStellarFB SigmaVel60 model. This last point can be additionnal information that can help to distinguish simulation models.

When looking at the distribution of cross section values σ/m_χ, we observe that the SigmaVel60 and Sig-

maVel30 cosmologies do not form distinct ranges of values, but rather, that some particles can have cross section values that could belong to either model. This figure demonstrates that for other features, distinguishing between the models is challenging.

The cross section also appears to be linearly related to the other features, a point we will discuss further below.

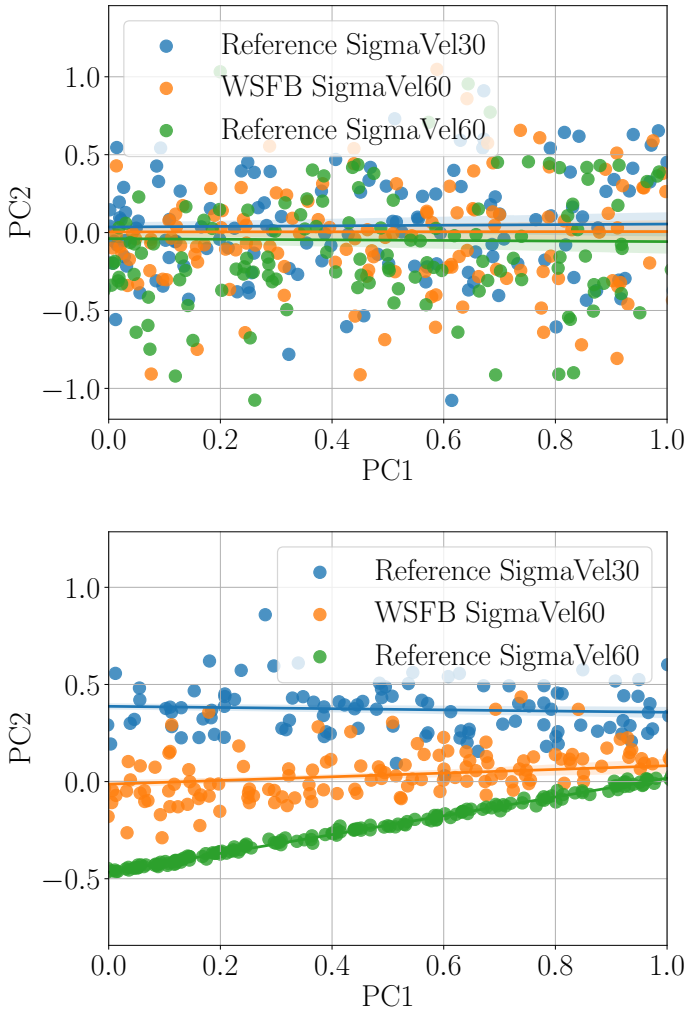


Figure 3.4: Principal Component Analysis performed using input Feature continuous variable from each galaxy: stellar mass, effective radius $R_{\rm eff}$, and dark matter fraction $f_{\rm DM}$ (upper panel) and with the addition of the cross section as an included variable (bottom panel). First Principal Component (PC1) as a function of the Second Principal Component (PC2).

Figure 3.4 show to Principal Component Analysis (PCA). PCA reduces the dimensionality of data by transforming it into a set of orthogonal components that capture the maximum variance in the original dataset. In the upper figure, PCA is performed using input Feature continuous variable from each galaxy: stellar mass, effective radius $R_{\rm eff}$, and dark matter fraction $f_{\rm DM}$. The bottom figure shows the two principal components derived from these data, with the addition of the cross section as an included variable. The top figure shows that it is difficult to distinguish the models with the features we have, and the bottom one highlights that these models produce significantly different cross section values.

Then, we tested machine learning algorithms ranging from the simplest to a much more complex one.

Machine learning algorithms tested on a single simulation model at a time

We first tested machine learning algorithms on one simulation model at a time. For training, we alternated between using either Reference SigmaVel60 or Reference SigmaVel30. Figure 3.5 shows the median error percentage in predicting a cross section value for various machine learning algorithms, ranging from the simplest to the most complex. Essentially, out of 10 tests, 5 median error results are from algorithms trained on Reference SigmaVel60, and 5 results are from Reference SigmaVel30. The averages of these median errors were calculated from those 10 tests, and the error bars show the highest and lowest scores across all model trainings.

Testing a simple linear regression model separately on both Reference SigmaVel60 and Reference SigmaVel30, we achieved a very good average median error percentage, around 15%. Moving to a more complex algorithm did not yield any significant change in the error.

Nevertheless, the error remains low, and the models are quite successful in predicting the cross section when trained on a simulation model. The good results from linear regression confirm the previous intuition regarding the correlation between cross section and other features.

Additionally, detailed information regarding the structure of our neural networks is provided in the appendix.

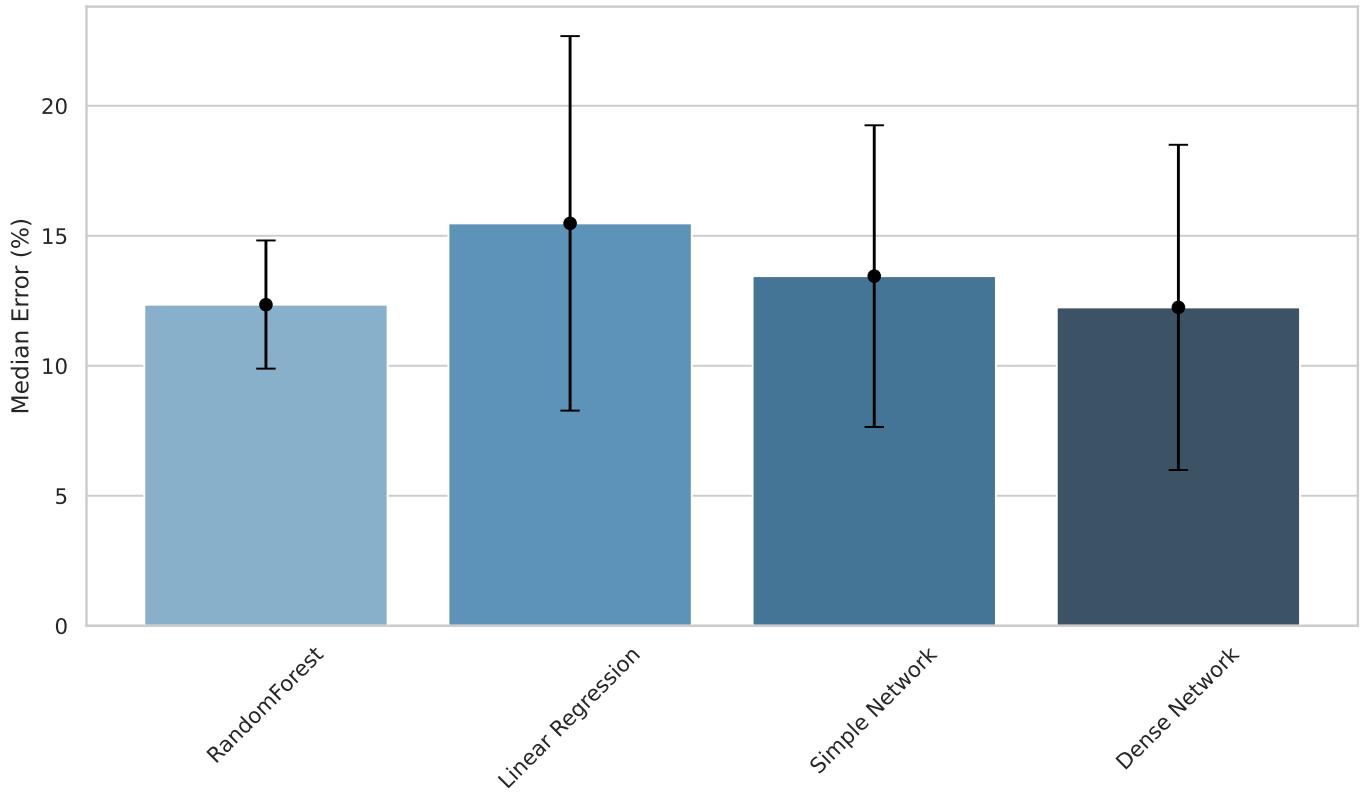


Figure 3.5: Median error percentage in predicting a cross section value for various machine learning algorithms : Random Forest, Linear Regression, Simple Neural Network, Dense Neural Network

Machine learning algorithms tested simultaneously on multiple simulation models

This time, we trained the algorithms using all the cosmologies from the Reference model. Table 3.1 shows the median errors in the predicted cross section values for each testing model using the Dense Network from 3.5 (hereafter "complex neural network"). Therefore, our neural network was trained on the Reference CDM, Reference SigmaVel60, and Reference SigmaVel30 models, and we tested the trained network on several simulation models listed in the table. We also tested the neural network on the Reference CDM model to see if it could

predict a cross section equal to 0 $cm^2 \cdot g^{-1}$. The neural network was not able to predict a cross section exactly equal to 0 $cm^2 \cdot g^{-1}$, but it consistently predicted values well below the range of the other two cosmologies. On average, the neural network predicted 0.14 $cm^2 \cdot g^{-1}$ instead of 0 $cm^2 \cdot g^{-1}$.

	Reference SigmaVel60	Reference SigmaVel30	all cosmologies
Median error	30%	10%	40%

Table 3.1: Median percentage error on cross section predictions by the complex neural network for different simulation models.

II.2 Results

With this satisfactory neural network, we decided to use it to predict the cross section values for dark matter particles in observed galaxies from the SPARC, Pizagno+2007, and Reyes+2011 datasets. We applied the same normalization as was used during the training process.

Figures 3.6 and 3.7 show the predicted values of the average cross section of dark matter particles within the effective radius of each observed galaxy as a function of stellar mass, colored respectively by the dark matter fraction and the effective radius. The error bars represent the median percentage error observed during training. They were calculated by multiplying the median percentage error of the neural network by each predicted value.

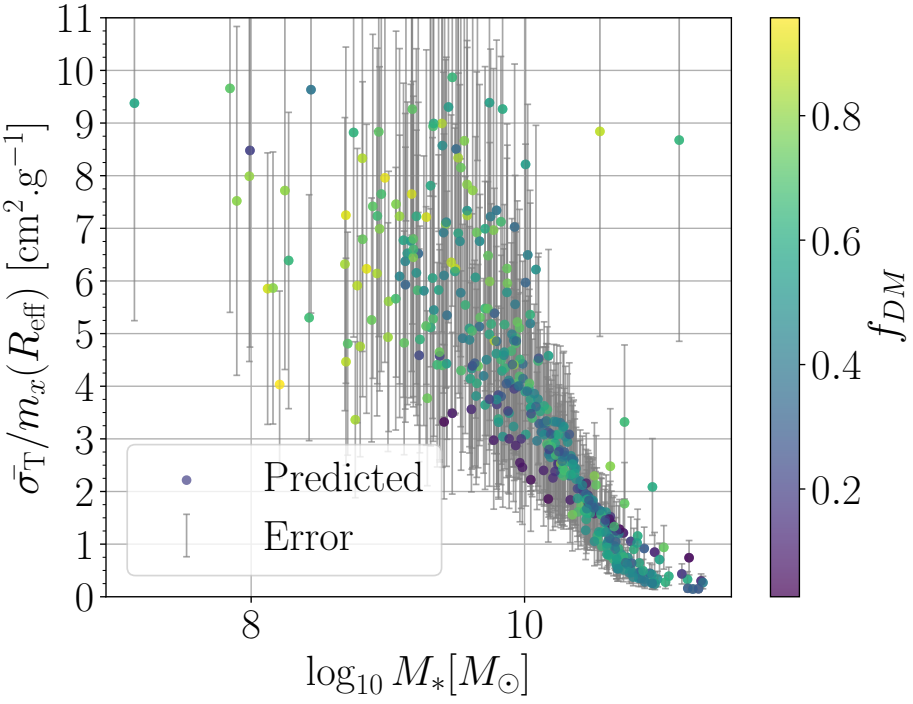


Figure 3.6: Neural network's predicted values of the average cross section of dark matter particles within the effective radius of each observed galaxy as a function of stellar mass, colored by the dark matter fraction. The error bars in grey represent the median percentage error observed during training.

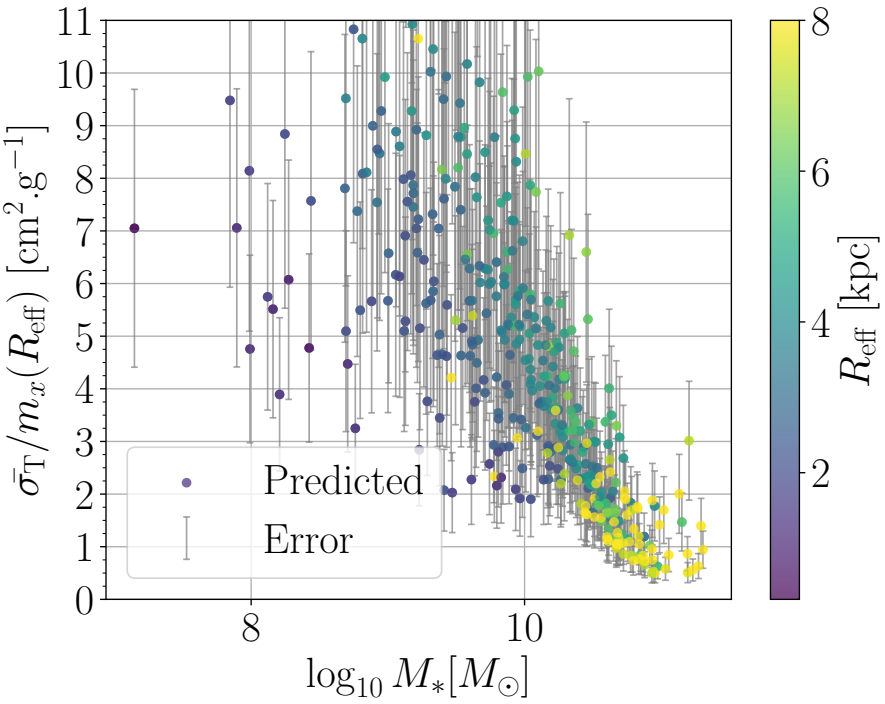


Figure 3.7: Neural network's predicted values of the average cross section of dark matter particles within the effective radius of each observed galaxy as a function of stellar mass, colored the effective radius. The error bars in grey represent the median percentage error observed during training.

In Figure 3.8, we selected all galaxies from Figure 3.7 with a dark matter fraction lower than 0.3. This result needs to be analyzed and compared to simulations in a future study.

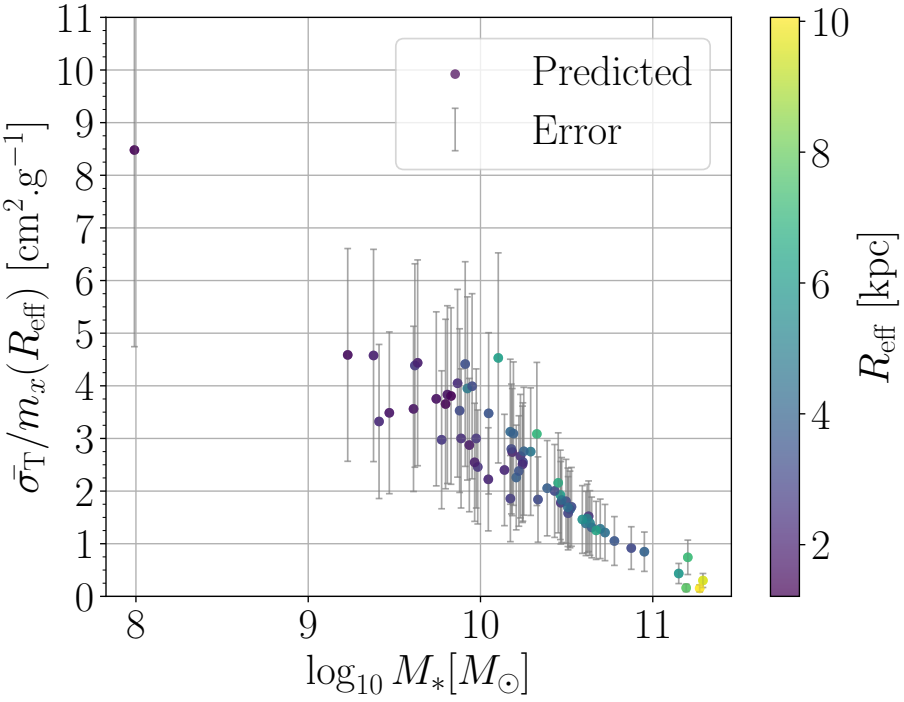


Figure 3.8: Neural network's predicted values of the average cross section of dark matter particles within the effective radius of each observed galaxy as a function of stellar mass, colored the effective radius for galaxies with a dark matter fraction lower than 0.3. The error bars in grey represent the median percentage error observed during training.

III Conclusion

In this research project, we employed several methods to address the questions "how does SIDM behave and what do observations tell us about the nature of dark matter?" and "how to constrain the cross section of dark matter particles based on the properties of observed galaxies." We developed a catalog of observed and simulated galaxies by applying a rigorous and critical examination of the methodologies used. This allowed us to identify a significant issue in the methodologies for calculating the masses of ETGs and to highlight the need for an apples-to-apples comparison.

Additionally, from the analysis of the dark matter fraction extracted from the simulated galaxies, we were able to understand different mechanisms of SIDM. Notably, at around $10^{9.5}$ $M_{\odot}$, dark matter experiences such a strong gravitational potential that it only aggregates regardless of the model. Furthermore, unlike dwarf galaxies where SIDM can form a core and "empty" the galaxy center of dark matter, in massive galaxies, the opposite behavior is observed, and interactions between dark matter particles lead to aggregation.

Finally, to constrain the cross section of dark matter particles in observed galaxies, we developed two approaches: a semi-analytic model and a machine learning approach. The results of both models need to be refined and discussed in the future. Additionally, a machine learning approach that aims not to predict the cross section value of dark matter particles but to classify our entire observed galaxies sample as belonging to either CDM or SIDM cosmologies using, potentially, a Bayesian inference algorithm could be considered.

Appendix A

Structure of machine learning algorithms

I Simple neural network

Layer	Units	Activation
Input	32	relu
Hidden	32	relu
Output	1	relu

II Complex neural network

Layer	Units	Activation
Input	40	tanh
Hidden	40	linear
Hidden	50	tanh
Hidden	30	linear
Output	1	tanh

Bibliography

[1] C. Alcock et al., *The Astrophysical Journal* 542.1 (Oct. 2000).

[2] M. W. Auger et al., *The Astrophysical Journal* 705.2 (Nov. 2009).

[3] Matteo Barnabè et al., *MNRAS* 415.3 (Aug. 2011).

[4] Eric F. Bell et al., *The Astrophysical Journal Supplement Series* 149.2 (Dec. 2003).

[5] W. J. G. de Blok and S. S. McGaugh, *Monthly Notices of the Royal Astronomical Society* 290.3 (Sept. 1997).

[6] W. J. G. de Blok and S. S. McGaugh, *MNRAS* 290.3 (Sept. 1997).

[7] Michael Boylan-Kolchin, James S. Bullock, and Manoj Kaplinghat, *MNRAS: Letters* 415.1 (July 2011).

[8] Michael Boylan-Kolchin, James S. Bullock, and Manoj Kaplinghat, *MNRAS* 422.2 (May 2012).

[9] Michele Cappellari et al., *MNRAS* 366.4 (Mar. 2006).

[10] Michele Cappellari et al., *MNRAS* 432.3 (July 2013).

[11] Douglas Clowe et al., 648.2 (Sept. 2006).

[12] Camila Correa et al. 2024.

[13] Camila A. Correa et al., *Monthly Notices of the Royal Astronomical Society* 452.2 (July 2015).

[14] Camila A. Correa et al., *MNRAS: Letters* 472.1 (Aug. 2017).

[15] Camila A. Correa et al., *MNRAS* 517.2 (Dec. 2022).

[16] Robert A. Crain et al., *MNRAS* 450.2 (Apr. 2015).

[17] A. Dekel et al., 437.7059 (Sept. 2005).

[18] Albert Einstein, *Science* 84.2188 (Dec. 1936).

[19] Christoph Engler et al., *MNRAS* 507.3 (Aug. 2021).

[20] Ismael Ferrero et al., *MNRAS* 425.4 (Oct. 2012).

[21] Andreea S Font, Ian G McCarthy, and Vasily Belokurov, *MNRAS* 505.1 (May 2021).

[22] S. S. Gershtein and Ya. B. Zel'dovich, *Soviet Journal of Experimental and Theoretical Physics Letters* 4 (Sept. 1966).

[23] Anna de Graaff et al., *MNRAS* 511.2 (Dec. 2021).

[24] Kim Griest, *The Astrophysical Journal* 366 (Jan. 1991).

[25] J. E. Gunn et al., *The Astrophysical Journal* 223 (Aug. 1978).

[26] Lars Hernquist, *The Astrophysical Journal* 356 (June 1990).

[27] Fangzhou Jiang et al., *MNRAS* 521.3 (May 2023).

[28] Felix Kahlhoefer et al., *Journal of Cosmology and Astroparticle Physics* 2019.12 (Dec. 2019).

[29] F. D. Kahn and L. Woltjer, *The Astrophysical Journal* 130 (Nov. 1959).

[30] Manoj Kaplinghat, Sean Tulin, and Hai-Bo Yu, *Phys. Rev. Lett.* 116 (4 Jan. 2016).

[31] Ivan R. King, *Evolution of Galaxies and Stellar Populations*. Ed. by Beatrice M. Tinsley and D. Campbell Larson Richard B. Gehret. Jan. 1977.

[32] Anatoly Klypin et al., *The Astrophysical Journal* 522.1 (Sept. 1999).

[33] Anatoly A. Klypin, Sebastian Trujillo-Gomez, and Joel Primack, *The Astrophysical Journal* 740.2, 102 (Oct. 2011).

[34] Janis Kummer et al., *MNRAS* 487.1 (July 2019).

[35] Federico Lelli, Stacy S. McGaugh, and James M. Schombert, *The Astronomical Journal* 152.6, 157 (Dec. 2016).

[36] Pengfei Li et al., *Astronomy Astrophysics* 615, A3 (July 2018).

[37] Shubo Li et al., *MNRAS* (Mar. 2024).

[38] Shubo Li et al., *MNRAS* 529.4 (Mar. 2024).

[39] Mark R. Lovell et al., *MNRAS* 481.2 (Dec. 2018).

[40] Shengdong Lu et al., *MNRAS* 526.1 (Sept. 2023).

[41] V. A. Lubimov et al., *Physics Letters B* 94.2 (July 1980).

[42] Stacy S. McGaugh, *The Astronomical Journal* 143.2, 40 (Feb. 2012).

[43] A. L. Melott et al., 51.10 (Sept. 1983).

[44] M. Milgrom, *The Astrophysical Journal* 270 (July 1983).

[45] B. Moore et al., *Monthly Notices of the Royal Astronomical Society* 310.4 (Dec. 1999).

[46] Ben Moore, 370.6491 (Aug. 1994).

[47] Julio F. Navarro, Carlos S. Frenk, and Simon D. M. White, *The Astrophysical Journal* 462 (May 1996).

[48] Julio F. Navarro, Carlos S. Frenk, and Simon D. M. White, *The Astrophysical Journal* 490.2 (Dec. 1997).

[49] Se-Heon Oh et al., *The Astronomical Journal* 142.1 (June 2011).

[50] J. P. Ostriker and P. J. E. Peebles, *The Astrophysical Journal* 186 (Dec. 1973).

[51] P. J. E. Peebles and J. T. Yu, *The Astrophysical Journal* 162 (Dec. 1970).

[52] A. A. Penzias and R. W. Wilson, *The Astrophysical Journal* 142 (July 1965).

[53] James Pizagno et al., *The Astronomical Journal* 134.3 (July 2007).

[54] Planck Collaboration et al., 571, A16 (Nov. 2014).

[55] L. Portinari et al., *The Astrophysical Journal* 604.2 (Apr. 2004).

[56] Price, S. H. et al., *AA* 665 (2022).

[57] J I Read, M G Walker, and P Steger, *MNRAS* 484.1 (Jan. 2019).

[58] R. Reyes et al., *MNRAS* 417.3 (Oct. 2011).

[59] Andrew Robertson, Richard Massey, and Vincent Eke, *MNRAS* 467.4 (June 2017).

[60] Vera C. Rubin and Jr. Ford W. Kent, *The Astrophysical Journal* 159 (Feb. 1970).

[61] Laura V. Sales, Andrew Wetzel, and Azadeh Fattahi, (2022).

[62] Joop Schaye et al., *MNRAS* 446.1 (Nov. 2014).

[63] James Schombert and Stacy McGaugh, *Publications of the Astronomical Society of Australia* 31 (2014).

[64] Joshua D. Simon, *Annual Review of Astronomy and Astrophysics* 57.Volume 57, 2019 (2019).

[65] V. M. Slipher, *The Observatory* 40 (Aug. 1917).

[66] G. F. Smoot et al., *The Astrophysical Journal* 396 (Sept. 1992).

[67] David N. Spergel and Paul J. Steinhardt, *Phys. Rev. Lett.* 84 (17 Apr. 2000).

[68] Katherine A. Suess et al., *The Astrophysical Journal* 885.1, L22 (Nov. 2019).

[69] C. Tortora et al., *MNRAS* 425.1 (Aug. 2012).

[70] Matthew G. Walker and Jorge Peñarrubia, *The Astrophysical Journal* 742.1 (Nov. 2011).

[71] Matthew G. Walker et al., *The Astrophysical Journal Letters* 717.2 (June 2010).

[72] Joe Wolf et al., *MNRAS* 406.2 (Aug. 2010).

[73] Xiaohu Yang et al., *The Astrophysical Journal* 671.1 (Dec. 2007).

[74] F. Zwicky, *Helvetica Physica Acta* 6 (Jan. 1933).

[75] F. Zwicky, *Phys. Rev.* 51 (8 Apr. 1937).