

The Dependence of Halo Mass on Galaxy Size at Fixed Stellar Mass and Colour using Galaxy-Galaxy Lensing

by

Darshak A. Patel

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Master of Science
in
Physics

Waterloo, Ontario, Canada, 2025

© Darshak A. Patel 2025

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Apart from the introduction, this thesis is based on the following articles:

CHAPTER 4: Burger, P. A., **Patel, D. A.**, & Hudson, M. J. (2025). Constraining cosmological parameters using density split lensing and the conditional stellar mass function. *Phys. Rev. D*, 111, 123515. doi:10.1103/3kmm-3w11

CHAPTER 2/3/5/6/7/Appendices: Based on **Patel et al.** in prep. The title of the publication is Weak Lensing with UNIONS: The Halo Mass Dependence on Galaxy Size at Fixed Stellar Mass and Colour, and will be submitted to MNRAS.

The implementation of the CSMF HOD into the `AbacusSummit` simulations, and the subsequent generation of mock galaxy catalogues were done by Dr. Pierre Burger with input from Dr. Mike Hudson and myself. I did the majority of the computational analysis presented in this thesis, with some software development aid from Dr. Pierre Burger. Dr. Mike Hudson and Dr. Pierre Burger provided valuable insight on the interpretation of results for the manuscript that is currently in preparation.

Abstract

To advance our holistic understanding of galaxy formation physics, we must examine the relationship between baryonic matter and dark matter (DM) within the universe. In this thesis, we investigate the correlation between dark matter halo mass and galaxy size at fixed stellar mass and colour. Using galaxy-galaxy lensing, we measure excess surface density (ESD) profiles for red, early-type galaxies from the Dark Energy Spectroscopic Instrument (DESI) Legacy Imaging survey, with shape measurements from the Ultraviolet Near Infrared Optical Northern Survey (UNIONS). We model the lensing signal using a conditional stellar mass function (CSMF) halo occupation distribution (HOD) calibrated on the **AbacusSummit** simulations and adopt a power-law relation between halo mass and galaxy size at fixed stellar mass: $M_h \propto r_{\text{eff}}^\eta$. To fit the HOD parameters to the observationally measured ESD vectors, we train and utilize neural-network emulators, a form of non-linear interpolators. The analysis performed in this work follows a two-step process. First, we fit the HOD parameters to the ESD vectors corresponding to three stellar mass bins, not split further into two size bins. We freeze the best-fit HOD parameters and generate the corresponding best-fit mock catalogue. Using this best-fit mock catalogue, we apply our halo mass-size model and fit to the size-split observational ESD measurements.

For central galaxies with stellar mass $10.5 \leq \log_{10}(M_*/M_\odot) < 11.2$, we find no significant correlation. At higher stellar masses, $11.2 \leq \log_{10}(M_*/M_\odot) < 11.78$, we detect a positive correlation with $\eta_{\text{cent}} = 0.51 \pm 0.14$. A linear fit as a function of the logarithm of stellar mass yields a slope of $s_{\eta_{\text{cent}}} = 0.66 \pm 0.28$, indicating that the halo mass-size correlation strengthens with increasing stellar mass. For satellite galaxies, we observe a negative correlation at low and intermediate stellar masses for the host halo mass-galaxy size relation of $\eta_{\text{sat}} = -0.19 \pm 0.05$ and $\eta_{\text{sat}} = -0.09 \pm 0.05$, respectively. The correlation is consistent with zero within the stellar mass range of $11.2 \leq \log_{10}(M_*/M_\odot) < 11.78$.

Acknowledgements

To Dr. Mike Hudson: Thank you for providing me with an opportunity to explore my passion for astrophysics research and the chance to grow in a professional setting. I greatly appreciate the immense support, guidance, and direction you have given me over the past two years. The skills and knowledge that I have gained working with you will, without a doubt, be critical for my future. Thank you!

To Dr. Pierre Burger: Thank you for being an exceptional mentor and teacher. I have truly learned so much from working alongside you and greatly appreciate all the time and effort you put into me and my work. You were right by my side, helping and encouraging me through the highs and the lows, which I will be forever grateful for.

To my fellow graduate students and friends: Martine, James, Julian, Rashaad, Sofia, Andrea, Neo, and David, thank you for providing me with a supportive and fun environment in which I truly felt like I belonged. I am grateful to you all for being there when I (and any of us) needed encouragement and space to vent. I can't wait to see what we all accomplish from here!

Finally, I would like to thank the members of my thesis committee, Dr. Michael Balogh and Dr. James Taylor, for their time and valuable insight.

Dedication

THIS THESIS IS DEDICATED TO MY PARENTS AND MY BROTHER.
THANK YOU FOR THE ENDLESS SUPPORT AND ENCOURAGEMENT.

Table of Contents

Author’s Declaration	ii
Statement of Contributions	iii
Abstract	iv
Acknowledgements	v
Dedication	vi
List of Figures	xi
List of Tables	xiv
List of Abbreviations	xv
1 Introduction	1
1.1 Cosmology and Halo Physics	1
1.1.1 Cosmological Background	1
1.1.2 Structure Formation	6
1.1.3 Halo Mass Function	8
1.1.4 Halo Density Profiles	12
1.2 Galaxy Formation & Evolution	14

1.2.1	Gas Accretion & Stellar Mass Buildup	15
1.2.2	Stellar Mass Function	16
1.2.3	Mechanisms that Alter the Evolution of Galaxies	17
1.2.4	The Sizes of Galaxies	20
1.2.5	Scaling Relations of Galaxy Observables	22
1.3	Galaxy-Halo Connection Models	24
1.4	Weak Gravitational Lensing	26
1.5	Observational Status of Halo Mass vs Galaxy Size Studies using GGL . . .	30
1.6	Objective & Outline of Thesis	31
2	Data	33
2.1	UNIONS: Shapes of Source Galaxies	33
2.2	DLIS: Lens Galaxy Catalogue	34
2.3	Lens Galaxy Sample Selection	36
2.3.1	Colour Cut	36
2.3.2	Sérsic Index Cut	37
2.3.3	Stellar Mass Bins	37
2.3.4	Splitting by Galaxy Size	38
3	$\Delta\Sigma$ Measurements	43
3.1	The $\Delta\Sigma$ Estimator	43
3.1.1	Random Subtraction	45
3.1.2	Boost Factor	45
3.2	DLIS $\Delta\Sigma$ Results	46
4	Methods & Modelling	49
4.1	AbacusSummit Simulations	50
4.2	Conditional Stellar Mass Function Halo Occupation Distribution Model . .	50

4.2.1	Subhalo Contribution	52
4.2.2	Implementation of Halo Mass-Galaxy Size Dependence	53
4.3	Building the Mocks and the Emulators	54
4.3.1	Generating the Mocks	55
4.3.2	Model ESD	55
4.3.3	Emulator Training & Accuracy	56
4.3.4	$\Delta\Sigma_{\text{tNFW}} f_{\text{sat}}$ Emulators	57
5	Results	59
5.1	Fit for Size Dependence	61
5.2	Calculation of η	62
6	Discussion	68
6.1	Comparison to Previous Works	70
6.2	Impact of Galaxy Size Uncertainty	73
6.3	Insights from the Environmental Dependence on the Size-Stellar Mass Relation	76
7	Conclusion	78
7.1	Direction for Future Works	78
7.1.1	Revisiting Satellite Placement	78
7.1.2	Satellite Galaxy: Subhalo Mass-Size Relation?	79
7.1.3	Halo Concentration	80
7.2	Summary	81
	References	83
	APPENDICES	97
A	Additional Information on Emulator Construction	98
A.1	Emulator Training Space	98
A.2	Emulator Parameter Visualization	99
A.3	Emulator Correlation Matrices and Accuracy Plots	99

B Behaviour of β	104
C Varying the s Parameter	107

List of Figures

1.1	Illustration of various commonly used Halo Mass Functions (HMFs). The curves plotted were generated using the <code>python</code> package <code>Colossus</code> by Diemer (2018).	10
1.2	Figure 1 from Robles et al. (2019), which illustrates a simulated merger tree. The x-axis represents the distance from the main branch.	11
1.3	Illustration of some commonly used halo density profiles. The curves plotted were generated using the <code>python</code> package <code>Colossus</code> by Diemer (2018). . .	14
1.4	a) Plot of average SFR as a function of redshift for fixed halo masses. b) Plot of the SMF for varying redshifts.	17
1.5	Figure 1 taken from Mandelbaum et al. (2005), which illustrates the various halo terms that contribute to central (top panel) and satellite (middle panel) galaxies to produce the overall lensing profile (bottom panel).	29
2.1	Histogram plot of one of the blinded photometric redshift distributions of Ultraviolet Near-Infrared Optical Northern Survey (UNIONS) <code>ShapePipe</code> v1.3 sources and the photometric redshift distribution of the red, Sérsic cut, declination cut, $M_{\star, \text{limit}}$ applied lens galaxy sample from DESI Legacy Imaging Survey (DLIS).	35
2.2	Stellar mass versus photo- z for a cut subsample of the original DLIS catalogue.	39
2.3	Plot of the galaxy size-stellar mass for red, Sérsic cut, stellar mass complete DLIS galaxies.	41
3.1	Plot of the full stellar mass bin (not size split) $r_p \Delta \Sigma$ profiles from the DLIS catalogue.	46
3.2	Plot of the size split subsamples $r_p \Delta \Sigma$ profiles from the DLIS catalogue for each stellar mass.	47

4.1	Plot of observational $r_p \Delta \Sigma$ for each stellar mass bin, with emulator mocks plotted in blue.	57
4.2	Accuracy plot of the $\Delta \Sigma_{\text{mocks}}$ emulator.	58
5.1	Comparison of the $r_p \Delta \Sigma$ measurements from the full stellar-mass bin DLIS with predictions from the best-fit HOD parameters obtained from the emulator and from the true mock.	60
5.2	Histogram of the halo masses of centrals and satellites in the best-fit mock split into <i>small</i> and <i>large</i> subsamples.	64
5.3	Hamiltonian Monte Carlo (HMC) posterior distributions for β_{cent} and β_{sat}	66
6.1	Summary plot highlighting the results in the literature with the results of this thesis.	69
6.2	Halo mass-size dependence results concerning satellite galaxies.	72
6.3	Results of a simulated approach to gauge the effect of observational scatter on measured $\Delta (\Delta R')$ values and its consequence on η	75
A.1	Visualization of the emulator parameter space and location of best-fit Halo Occupation Distributions (HOD) parameters.	100
A.2	Rainbow plot illustration of the effect of each HOD parameter on the predicted Excess Surface Mass Density (ESD) vector from the emulator.	101
A.3	Rainbow plot highlighting the effect of the $\log_{10} \tau$ parameter on the output of the $\Delta \Sigma_{\text{tNFW}} f_{\text{sat}}$ emulator.	102
A.4	Correlation matrix for the $\Delta \Sigma_{\text{mocks}}$ emulators.	102
A.5	Accuracy plots for the $\Delta \Sigma_{\text{tNFW}} f_{\text{sat}}$ emulators.	103
A.6	Correlation matrix for the $\Delta \Sigma_{\text{tNFW}} f_{\text{sat}}$ emulators.	103
B.1	Rainbow plot illustration of the effect of varying β_{cent} and fixing β_{sat} on the <i>small</i> galaxy and <i>large</i> galaxy $r_p \Delta \Sigma$ vectors.	105
B.2	Rainbow plot illustration of the effect of fixing $\beta_{\text{cent}} = 0$ and varying β_{sat} on the <i>small</i> galaxy and <i>large</i> galaxy $r_p \Delta \Sigma$ vectors.	105
C.1	Illustration of the impact of the s parameter on the spatial radial distribution of satellite galaxies in a halo of mass $M_h = 1 \times 10^{14} M_\odot$	109

C.2	Rainbow plot illustration of the effect that varying the satellite placement parameter, s , has on the $r_p\Delta\Sigma$ vectors from AbacusSummit	110
C.3	Comparison of the $s = 0$ and $s = -0.3$ fits to the size split ESD vectors. . .	111
C.4	Comparison of η_{cent} and η_{sat} values determined for red, early-type galaxies in the case of $s = 0$ and $s = -0.3$	112

List of Tables

2.1	Summary of the stellar mass bins and subdivided size bins used in this thesis along with the values of the number of lens galaxies per bin (N_{gals}) used for the weak lensing measurements, the mean $g - z$ colour, the mean stellar mass, the mean photometric redshift, the mean Sérsic index, the mean galaxy size, and the mean galaxy size compared to the median. . . .	42
5.1	Best-fit HOD parameter results.	60
5.2	Mean halo mass values and satellite fractions for central and satellite-type galaxies for each stellar mass bin and size subsample taken from the best-fitting mock.	63
5.3	Results for β , and η which quantify the fit to size dependence of the observational size split DLIS ESD vectors.	67

List of Abbreviations

Λ CDM Λ -Cold Dark Matter [2](#), [3](#), [7](#), [32](#)

AGN Active Galactic Nuclei [18](#)

BOSS Baryon Oscillation Spectroscopic Survey [70](#)

CFHT Canada-France-Hawai’i Telescope [33](#), [73](#)

CFHTLens Canada-France-Hawaii Telescope Lensing Survey [31](#), [34](#)

CFIS Canada-France Imaging Survey [33](#), [34](#)

CLF Conditional Luminosity Function [25](#), [26](#)

CMB Cosmic Microwave Background [3](#), [6](#)

CSMF Conditional Stellar Mass Function [26](#), [31](#), [49](#), [50](#), [52–55](#)

DESI Dark Energy Spectroscopic Instrument [36](#), [98](#)

DLIS DESI Legacy Imaging Survey [xi](#), [xii](#), [xiv](#), [31–33](#), [35–41](#), [45](#), [46](#), [49](#), [50](#), [54–60](#), [73](#), [78](#), [79](#), [99](#), [103](#), [107](#), [108](#), [110](#), [111](#)

DOF Degrees of Freedom [72](#), [73](#), [108](#), [111](#)

ESD Excess Surface Mass Density [xii–xiv](#), [27](#), [32](#), [43](#), [44](#), [46](#), [48–50](#), [52](#), [53](#), [55–62](#), [67](#), [70](#), [71](#), [78–81](#), [98–101](#), [104–108](#), [110](#), [111](#)

ETG Early-Type Galaxy [76](#), [77](#)

FoF Friends-of-Friends [50](#)

GGL Galaxy-Galaxy Lensing [26](#), [27](#), [30–32](#), [38](#), [43](#), [56](#), [79](#), [80](#)

GR General Relativity [26](#)

HAM Halo Abundance Matching [24](#)

HMC Hamiltonian Monte Carlo [xii](#), [61](#), [62](#), [65–67](#), [98](#), [99](#), [104](#), [108](#)

HMF Halo Mass Function [xi](#), [8–10](#), [16](#), [17](#), [24](#)

HOD Halo Occupation Distributions [xii](#), [xiv](#), [24–26](#), [31](#), [49–61](#), [79–81](#), [98–101](#), [107](#), [108](#)

HSC Hyper Suprime-Cam [31](#), [33](#)

LF Luminosity Function [24](#), [25](#)

LSS Large Scale Structure [20](#), [21](#), [27](#)

NFW Navarro-Frenk-White [12](#), [13](#), [70](#), [107](#), [109](#)

RMS Root Mean Square [8](#), [32](#)

SDSS Sloan Digital Sky Survey [37](#), [77](#)

SFR Star Formation Rate [16–18](#)

SHAM Subhalo Abundance Matching [24](#)

SHMR Stellar-to-Halo Mass Relation [23](#), [24](#), [46](#), [51](#), [68](#)

SMF Stellar Mass Function [16](#), [17](#), [24](#), [98](#)

SMR Size-Stellar Mass Relation [76](#), [77](#)

SNR Signal-to-Noise Ratio [30](#), [43](#), [74](#)

SOM Self-Organizing Map [34](#)

tNFW Truncated Navarro-Frenk-White [49](#), [52](#), [53](#), [57](#), [80](#)

UNIONS Ultraviolet Near-Infrared Optical Northern Survey [xi](#), [31](#), [33–35](#), [38](#), [40](#), [43–45](#), [50](#)

Chapter 1

Introduction

In this thesis, we investigate the dependence of halo mass on galaxy size at fixed stellar mass and colour. Identifying such a dependence provides critical insight into the processes driving galaxy formation and evolution. Galaxy formation is inherently complex, but significant progress has been made in connecting theoretical models, simulations, and observations. To understand galaxy formation and to model this process as accurately as possible, there are three key areas of interest: cosmology, baryonic physics, and galaxy observables. This introduction will provide background on the most relevant topics and their connection to understanding the work of this thesis.

1.1 Cosmology and Halo Physics

1.1.1 Cosmological Background

Galaxy formation and evolution can be thought of as an application of (relatively) well-known physics with cosmological initial and boundary conditions (Mo et al., 2010).

As emphasized in Mo et al. (2010), the cosmology of the Universe plays a critical role in our ability to understand the formation of galaxies, as galaxies are observed on cosmological lengths and time scales. That is to say, understanding the properties of the universe greatly impacts the conclusions we make about the evolutionary history of galaxies and the factors that alter them. Modern-era cosmology is based on the cosmological principle, which states

that, when viewed on a sufficiently large scale, the universe appears spatially homogeneous and isotropic. From these basic principles, there have been many attempts to formulate an overarching model of the universe. Presently, the widely accepted cosmological model for the universe is the [Λ-Cold Dark Matter \(ΛCDM\)](#) model, also referred to as the concordance model. This model consists of three major components. First, the presence of baryonic matter in the universe. This includes protons, neutrons, and electrons, which together form various elements and compounds that ultimately make up the visible matter seen in the universe. Second, as denoted by the *CDM* abbreviation, we postulate the existence of cold dark matter, which makes up $\sim 27\%$ ([Aghanim et al., 2020](#)) of the energy budget of the Universe. Finally, this model predicts that there exists a non-zero cosmological constant, Λ , which is linked to the dark energy within the universe. The cosmological constant also describes the accelerated expansion of the universe that we see today.

The Friedmann-Lemaître-Robertson-Walker (FLRW) metric ([Friedmann, 1922](#); [Lemaître, 1933](#); [Robertson, 1935](#); [Walker, 1937](#)) is the best prescription for a homogeneous, isotropic, expanding universe. The FLRW metric for a line element ds is,

$$ds^2 = c^2 dt^2 - a^2(t) \left[\frac{dr^2}{1 - Kr^2} + r^2 (d\vartheta^2 + \sin^2 \vartheta d\varphi^2) \right], \quad (1.1)$$

where c is the speed of light, t is cosmic time, $a(t)$ is the scale factor, K defines the curvature of the universe, and r , ϑ and φ define the coordinate system. Given this formalism, the proper distance is written as

$$l = a(t) \int_0^r \frac{dr}{\sqrt{1 - Kr^2}} = a(t) \chi(r). \quad (1.2)$$

The variable $\chi(r)$ is the comoving distance, which is the proper distance in units of the scale factor. This value varies depending on the choice of curvature for the universe.

$$\chi(r) = \begin{cases} \sin^{-1}(r) & \text{for } K = +1 \\ r & \text{for } K = 0 \\ \sinh^{-1}(r) & \text{for } K = -1 \end{cases}. \quad (1.3)$$

A positively curved, closed universe is characterized by $K = +1$; $K = 0$ describes a flat, Euclidean universe, and $K = -1$ describes a negatively curved, open universe.

The [Friedmann \(1922, 1924\)](#) equations,

$$H^2(a) = \left(\frac{\dot{a}}{a} \right)^2 = \frac{8\pi G\rho}{3} - \frac{Kc^2}{a^2} + \frac{\Lambda c^2}{3}, \quad (1.4)$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3} \left(\rho + 3\frac{p}{c^2} \right) + \frac{\Lambda c^2}{3}, \quad (1.5)$$

describe the dynamics and expansion history of the universe, assuming the components of radiation, matter, and dark energy are perfect fluids. The curvature of the universe is defined as K , ρ is the density of an ideal fluid, and p is its pressure. Given the Λ CDM model outlined above, the first Friedmann equation can be rewritten in terms of the matter, radiation, and dark energy densities

$$H(a)^2 = H_0^2 \left[\Omega_r a^{-4} + \Omega_m a^{-3} + (1 - \Omega_{\text{tot}}) a^{-2} + \Omega_\Lambda \right]. \quad (1.6)$$

The parameter, $H = \frac{\dot{a}}{a}$, is the Hubble parameter, which defines the expansion rate and is a function of the scale factor, a . The density parameters denoted by Ω are defined as

$$\Omega_x = \frac{\rho_{x,0}}{\rho_{\text{crit}}}, \quad \Omega_\Lambda = \frac{\Lambda c^2}{8\pi G \rho_{\text{crit}}}, \quad \Omega_{\text{tot}} = \Omega_m + \Omega_r + \Omega_\Lambda, \quad \rho_{\text{crit}} = \frac{3H_0^2}{8\pi G}, \quad (1.7)$$

where $\rho_{x,0}$ is the present-day density for either matter (m), radiation (r), or baryons (b) and ρ_{crit} is the critical mass density. The value of the Hubble parameter at the present day is denoted by H_0 . The total matter density parameter, Ω_m , includes both baryonic and dark matter components of the universe. The Planck 2018 results (Aghanim et al., 2020) use the Cosmic Microwave Background (CMB) power spectrum to place constraints on the base Λ CDM model. They find $H_0 = 67.27 \pm 0.60$, $\Omega_\Lambda = 0.6834 \pm 0.0084$, $\Omega_m = 0.3166 \pm 0.0084$, $\Omega_b h^2 = 0.02236 \pm 0.00015$, and $\Omega_c h^2 = 0.1202 \pm 0.0014$. H_0 is the Hubble parameter, Ω_Λ is the dark energy density, Ω_m is the matter density, Ω_b is the baryon density, Ω_c is the cold dark-matter density, and h is the dimensionless Hubble constant.

Relevant Cosmological Parameters

In this thesis, we do not aim to probe cosmological information and constrain any of the cosmological parameters. We choose to fix the relevant cosmological parameters, which we now briefly describe. The parameters are:

- H_0 and h : Respectively, these represent the Hubble constant and the dimensionless Hubble constant, where $h = H_0/100 \text{ km s}^{-1} \text{ Mpc}^{-1}$. The Hubble constant quantifies the expansion rate of the universe and was first determined by Hubble (1929). In this thesis, we set $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and $h = 0.7$.
- Ω_m : The matter density of the Universe. In this thesis, we set $\Omega_m = 0.279$.

- Ω_b : The baryon density of the Universe. In this thesis, we set $\Omega_b = 0.046$.
- n_s : The spectral index, which describes the slope of the primordial power spectrum. In this thesis, we set $n_s = 0.97$.
- σ_8 : A measure of the amplitude of matter clustering in the Universe. Colloquially, this is thought of as a measure of the *clumpiness* of the Universe. In this thesis, we set $\sigma_8 = 0.82$.

The cosmological parameters adopted here are chosen to match the base cosmology of the `AbacusSummit` simulations, which will be described in Chapter 4.

Redshifts

One important property of celestial objects is their redshift, which provides insight into the epoch of the universe in which these objects formed and the time at which the light we now observe was emitted. Generally, celestial objects moving relative to us on Earth exhibit a shift in their light spectrum, depending on whether they are moving away or toward us. Objects moving towards us are *blueshifted*, such that the observed wavelength is compressed to shorter wavelengths relative to what the emitted wavelength is. Similarly, objects moving away from us are *redshifted*, such that their observed wavelengths are stretched to higher wavelengths. As the universe continues to expand, the observed wavelengths of objects are generally always longer than those emitted. By defining the observed wavelength as λ_o and the emitted as λ_e , we define the redshift of an object by the fractional difference of the observed and emitted wavelengths,

$$z = \frac{\lambda_o - \lambda_e}{\lambda_e} . \quad (1.8)$$

The redshift can also be written in terms of the scale factor.

$$z = \frac{a(t_0)}{a(t_e)} - 1 = \frac{1}{a(t_e)} - 1 , \quad (1.9)$$

whereby convention, the scale factor at present-day $a(t_0) = 1$. Redshifts are often measured spectroscopically or photometrically. In this thesis, photometric redshifts are utilized. The process of obtaining these measurements will now be summarized.

Photometric redshifts, z_{phot} , can be estimated by fitting observed photometric data of galaxies with spectral energy distribution (SED) templates. This method consists of

the measurement of magnitudes/brightnesses of objects through various broadband filters, each of which allows light of a select wavelength range to be detected. With N broadband filters, this results in $N - 1$ independent colours that constrain the galaxy's continuum within colour space. Broadband photometry is sensitive only to the most prominent spectral features, such as the 4000Å break, the Balmer break, or the Lyman break. This is in contrast to spectroscopic techniques, which enable precise measurements of positions and intensities of individual emission and absorption features. As the redshift increases, the spectral features identified will shift to higher wavelengths, thus redshifting. By comparing the observed colours and location of the strong spectral features predicted by SED templates at different redshifts, the photometric redshift can be estimated. Photometric redshifts are quite inaccurate and still suffer from outliers even when calibrated using a small subsample of spectroscopically confirmed galaxies. Nevertheless, this technique is optimal for wide-field galaxy surveys (as a first pass) that have large catalogues of galaxies for which redshifts are required.

Distances

Measurements of distances to objects, as we will encounter later, play a crucial role in weak gravitational lensing. In particular, concepts such as comoving distance, luminosity distance, and angular diameter distance are integral in understanding the distribution of galaxies throughout the universe. To begin, we revisit the definition of the comoving distance. Often, it is more convenient to define the comoving distance in terms of redshift. Recalling that $H = \frac{\dot{a}}{a}$, and $z = \frac{1}{a(t)} - 1$, which thus leads to $dz = -(da/a^2)$, we define the following:

$$d\chi = \frac{c dt}{a(t)} = \frac{c da}{a^2 H(a)} = \frac{c dz}{H(z)}, \quad (1.10)$$

where

$$\chi = \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')}, \quad (1.11)$$

$$E(z) = \frac{H(t)}{H_0} = \sqrt{\frac{\Omega_r}{a^4} + \frac{\Omega_m}{a^3} + \frac{\Omega_K}{a^2} + \Omega_\Lambda}. \quad (1.12)$$

Another measure of distance is the luminosity distance (D_L). Objects further away from us often appear to be fainter than expected. The apparent brightness (m) is used to quantify the brightness of an object as seen by an observer on Earth. The absolute magnitude (M) is defined as the brightness an object would have if it were placed at a

distance of 10 parsecs from Earth. The difference of these two magnitudes is related to the luminosity distance.

$$\mu = m - M = 5 \log_{10} D_L + 25. \quad (1.13)$$

In a flat, Euclidean universe, the luminosity distance is related to the comoving distance

$$D_L = (1 + z) \chi. \quad (1.14)$$

The angular diameter distance (d_A) relates the physical size of an object (D) to its angular size (ϑ) given in radians.

$$d_A = \vartheta D. \quad (1.15)$$

In terms of redshift and comoving distance, it is defined as $d_A = \chi/(1 + z)$.

1.1.2 Structure Formation

Although the universe is, on the largest scales, homogeneous and isotropic, locally this is not the case, as the existence of small amplitude deviations eventually gives rise to the cosmic structure we observe today. The evolution of dark matter, in particular, can be broken down into a linear and non-linear regime of growth. The linear regime is characterized by the existence of small amplitude density fluctuations, which first grow linearly. Following a period of linear growth, a characteristic shift occurs when density fluctuations become $\delta \geq 1$, in which dark matter grows non-linearly, resulting in the formation of the first collapsed and virialized dark matter halos. This section will provide an overview of this timeline.

Structure Formation in the Linear Regime

Structure formation begins with the existence of small-amplitude density fluctuations. Inflation, a theorized period of extremely rapid growth of the universe, amplified these tiny fluctuations, allowing them to grow to cosmic scales (Guth & Pi, 1982). Observationally, in 1992, the Cosmic Background Explorer (COBE) detected anisotropies in the CMB (Smoot et al., 1992), which became compelling evidence towards characterizing the nature of these initial quantum fluctuations. Early works predicted that the initial density field should be Gaussian (Bardeen et al., 1986) and have a form similar to the Harrison-Zel'dovich spectrum (Harrison, 1970; Zel'dovich et al., 1972; Sunyaev & Zeldovich, 1970). The COBE results, and subsequent results from the Wilkinson Microwave Anisotropy Probe (Bennett et al., 2003; Hinshaw et al., 2007) and Planck (Akrami et al., 2020), are consistent with Gaussian initial perturbations and the Harrison-Zel'dovich spectrum on large scales.

The overdensity field is defined as

$$\delta(\mathbf{x}, t) = \frac{\rho(\mathbf{x}, t) - \bar{\rho}(t)}{\bar{\rho}(t)}, \quad (1.16)$$

where $\rho(\mathbf{x}, t)$ is the density in terms of comoving coordinates $\mathbf{x}$ and cosmic time t . The mean density is $\bar{\rho}(t)$. Within the regime of linear growth, overdensities are $\delta(\mathbf{x}, t) \ll 1$. Once the universe is matter-dominated (i.e., after radiation-matter equality at $z \sim 3500$), dark matter is assumed to behave like a collisionless fluid and is thus governed by the continuity, Euler, and Poisson equations. Under linear perturbation theory, all higher-order terms can be neglected, resulting in the following linearized differential equations:

$$\frac{\partial \delta}{\partial t} + \frac{1}{a} \nabla_{\mathbf{x}} \mathbf{u} = 0 \quad (1.17)$$

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{\dot{a}}{a} \mathbf{u} = \frac{-1}{a} \nabla_{\mathbf{x}} \Phi. \quad (1.18)$$

$$\nabla^2 \Phi = 4\pi G \bar{\rho} \delta \quad (1.19)$$

The comoving potential and comoving velocity are $\Phi(\mathbf{x}, t)$ and $u(\mathbf{x}, t)$, respectively. Together, these equations lead to the linear growth equation shown below

$$\frac{\partial^2 \delta}{\partial t^2} + 2 \frac{\dot{a}}{a} \frac{\partial \delta}{\partial t} = 4\pi G \bar{\rho} a^2 \delta. \quad (1.20)$$

There are two possible solutions, the growing and decaying modes, δ_+ and δ_- . For structure growth, the decaying modes are irrelevant at later times, and thus, we focus on the growing modes. An approximation for the growing mode presented by [Carroll et al. \(1992\)](#) is

$$\delta_+ \propto D(z) \propto g(z)/(1+z) \quad (1.21)$$

where $g(z)$ is

$$g(z) \approx \frac{5}{2} \Omega_{\text{m}}(z) \left(\Omega_{\text{m}}^{4/7}(z) - \Omega_{\Lambda} + [1 + \Omega_{\text{m}}(z)/2] [1 + \Omega_{\Lambda}(z)/70] \right)^{-1}. \quad (1.22)$$

For Λ CDM, a simpler approximation is given by [Peebles \(1980\)](#) of $f(\Omega_{\text{m}}) \approx \Omega_{\text{m}}^{0.6}$ where $f(\Omega_{\text{m}})$ relates the logarithmic growth rate to the matter density at redshift z ,

$$f(\Omega_{\text{m}}) = -\frac{\text{d} \ln D(z)}{\text{d} \ln(1+z)}. \quad (1.23)$$

Halo Formation in the Non-Linear Regime

Later in cosmic time, when the overdensities become $\delta \sim 1$, the standard picture of linear growth breaks down and we enter the regime of non-linear growth. It is within this regime that gravitational collapse drives the formation of dark matter halos. Dark matter halos are the first structures to collapse and virialize in the universe, followed by the formation of the first stars and galaxies. Later, through hierarchical mergers, large-scale structures such as groups and clusters emerge. The first dark matter halos (colloquially referred to as mini-halos) are predicted within simulations to have formed within a redshift range of $z \sim 20 - 50$ (Ogiya et al., 2016; Angulo et al., 2017). The connection between initial perturbations moving from linear to nonlinear growth has been developed into the spherical collapse model by Gunn & Gott (1972). The spherical collapse model assumes overdensities to be spherical (and ignores angular momentum), from which linear theory suggests that any spherical region above a linear overdensity of 1.686 will collapse into a halo. These halos grow until their overdensities are approximately 5.55 times the critical mass density ρ_{crit} . It is at this point that the halos decouple from the expansion of the universe and collapse under their own gravity (also known as turnaround), eventually becoming virialized. The *virialization* of a halo, given by the virial theorem, $2E_{\text{kinetic}} + E_{\text{potential}} = 0$, occurs when the dark matter has sufficient kinetic energy to overcome any further gravitational collapse under its own gravity.

More accurate descriptions beyond spherical collapse include secondary infall models (Fillmore & Goldreich, 1984; Bertschinger, 1985; White & Zaritsky, 1992).

1.1.3 Halo Mass Function

From the initial density perturbations, a model for the number density of dark matter halos as a function of halo mass can be determined. Press & Schechter (1974) developed such an analytical model, which is called the **HMF**. The **HMF** as defined by Press & Schechter (1974) is:

$$n(M, z) \, dM = \sqrt{\frac{2}{\pi}} \frac{\bar{\rho} \, \delta_c(z)}{M^2 \sigma(M)} \exp\left(-\frac{\delta_c^2(z)}{2\sigma^2(M)}\right) \left|\frac{d \ln \sigma(M)}{d \ln M}\right| dM, \quad (1.24)$$

where $n(M, z) \, dM$ is the number density of halos within a mass range of M to $M + dM$. δ_c is the critical linear overdensity required for collapse, which is $\simeq 1.69$ in an Einstein-de Sitter universe (Einstein & de Sitter, 1932). $\bar{\rho}$ is the mean matter density of the universe, and $\sigma(M)$ is the **Root Mean Square (RMS)** fluctuations of the density field. The

variable $\nu = \delta_c(z)/\sigma(M)$, called the ‘peak height’ parameter, describes how many standard deviations above the critical threshold, δ_c , a density fluctuation must be for a halo of mass M to form at redshift z . The [HMF](#) can be rewritten in the following form,

$$n(M, z) \, dM = \frac{\bar{\rho}}{M^2} f_{\text{PS}}(\nu) \left| \frac{d \ln \nu}{d \ln M} \right| dM, \quad (1.25)$$

where

$$f_{\text{PS}}(\nu) = \sqrt{\frac{2}{\pi}} \nu \exp(-\nu^2/2). \quad (1.26)$$

f_{PS} is the fraction of mass within halos over a mass range in units of $\ln \nu$. With the continuous improvements to N-body simulations, it became apparent that the Press-Schechter formalism could not accurately predict halo number densities on all scales. Notably, the Press-Schechter [HMF](#) was found to overpredict the number of halos at the characteristic mass M^* and underpredict halos of $M > M^*$ ([Governato et al., 1999](#); [Sheth & Tormen, 1999](#)). Building on the work of [Press & Schechter \(1974\)](#), later improvements incorporated subhalos to address the underprediction of low-mass halos using the excursion set framework ([Bond et al., 1991](#)). Additionally, [Sheth et al. \(2001\)](#) generalized the [HMF](#) formalism by not assuming spherical collapse, but rather an ellipsoidal collapse model. This formalism, although an improvement over its predecessor, has since been shown to be inaccurate at high redshifts but very accurate at $z = 0$ ([Klypin et al., 2011](#)). Efforts to determine a universal [HMF](#) model that is independent of cosmology have led to many different analytical descriptions in the literature. Figure 1.1 plots some commonly used [HMF](#) models. Overall, we see the trend that the number density of halos decreases with increasing halo mass, where around $\sim 10^{14} - 10^{15} M_\odot$, there is a steep drop-off in the number density, signifying the rarity of observing such massive halos. The existence of high-mass halos is a natural consequence of the ‘‘bottom-up’’ structure growth scenario. In this scenario, dark matter halos grow hierarchically, such that larger halos are formed by the merging of smaller progenitor halos ([Peebles, 1965](#)). The different assembly histories of these halos can be tracked using merger trees. Figure 1.2 illustrates a simulated merger tree in which, at various snapshots, the colour map shows the mass of the main branch halo and the evolution of branches representing other halos. We observe a trend in which, along certain branches, the halo becomes a subhalo of the main branch halo, indicated by the change from circle to triangle markers. In this process, the halo also loses mass before merging into the main branch halo. By tracing the assembly histories of $z = 0$ halos via merger trees, we gain critical insight into how variations in halo growth affect the baryons they contain and their ability to form galaxies.

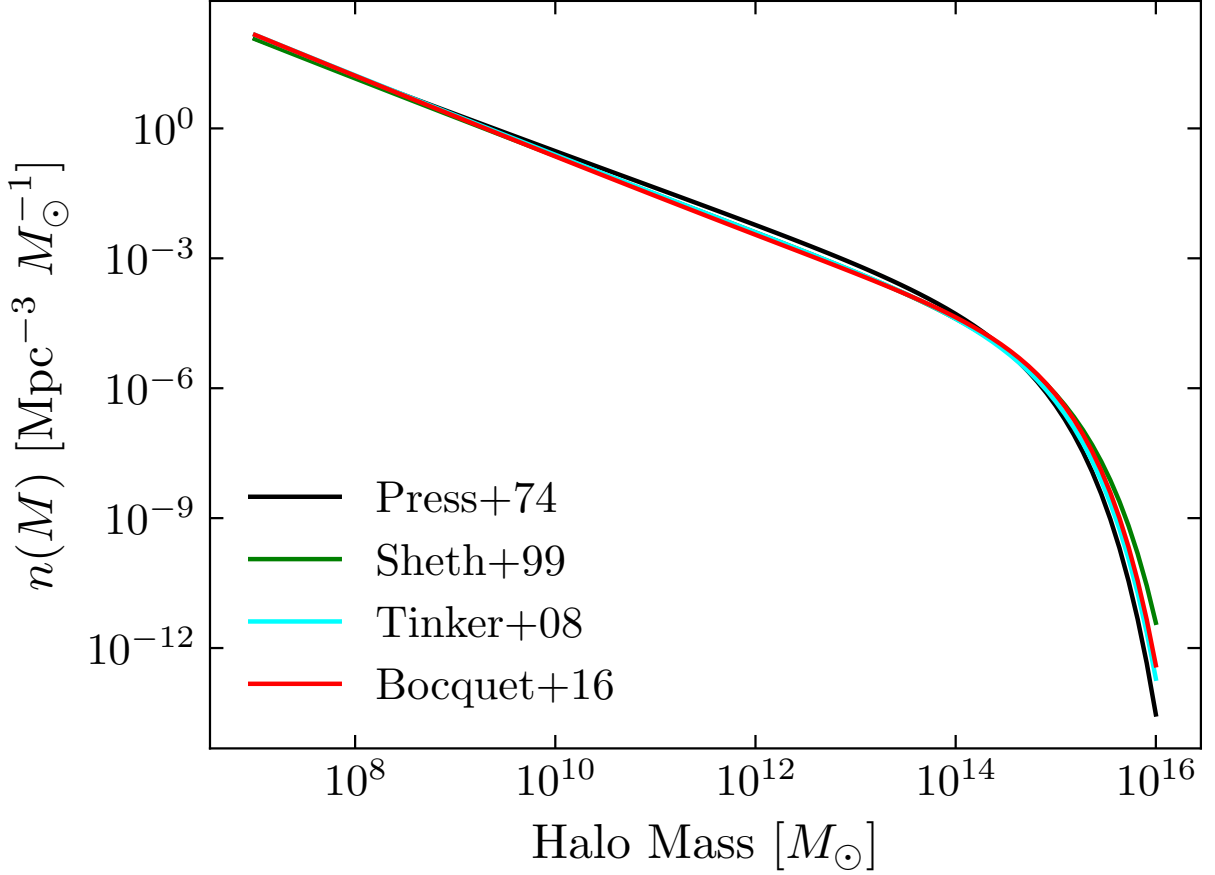


Figure 1.1: Illustration of various commonly used [HMFs](#). The curves plotted were generated using the `python` package `Colossus` by [Diemer \(2018\)](#). The Press-Schechter (black solid line) and Sheth-Tormen (solid green line) formalisms use a Friends-of-Friends algorithm to identify dark matter halos and their masses in their N-body simulations. [Tinker et al. \(2008b\)](#) (solid cyan line) utilizes spherical overdensity halo masses, whereas [Bocquet et al. \(2016\)](#) generated [HMFs](#) following either a 200m, 200c, or 500c mass definition. The adopted cosmology used to generate this plot is the standard Planck2018 cosmology ([Aghanim et al., 2020](#)).

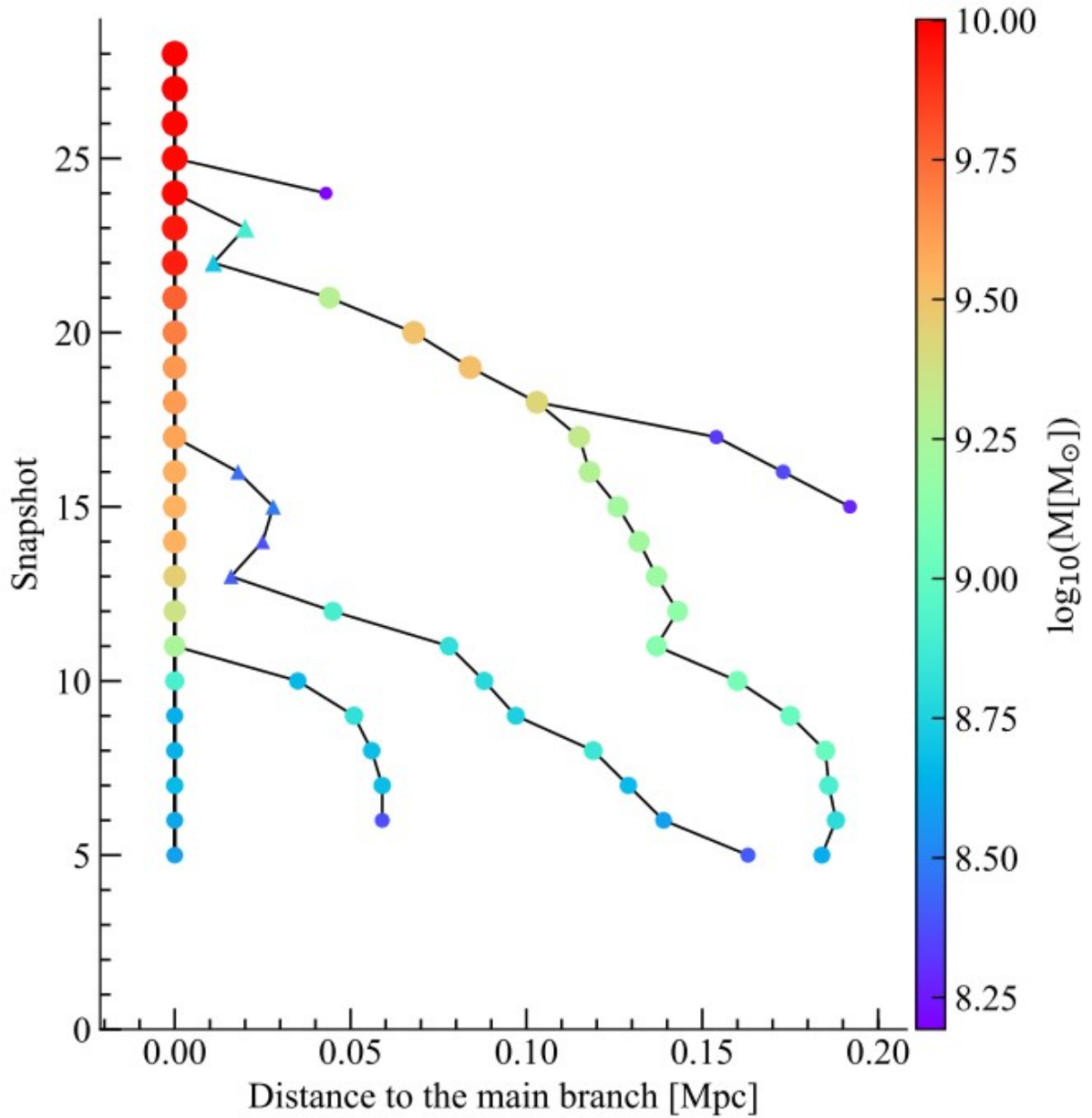


Figure 1.2: Figure 1 from [Robles et al. \(2019\)](#), which illustrates a simulated merger tree. The x-axis represents the distance from the main branch. The colour scale shows the mass of the halos considered in this merger tree. Circles represent individual halos, and triangles represent subhalos within the main branch halo.

1.1.4 Halo Density Profiles

There exist many functional forms to describe the density profiles of dark matter halos. The most common functional form is

$$\rho(r|M) = \frac{\rho_s}{(r/r_s)^\alpha (1 + r/r_s)^\beta}. \quad (1.27)$$

The parametrization for the functional form above, which has become widely accepted, is that of [Navarro et al. \(1996\)](#), where $(\alpha, \beta) = (1, 2)$. This is called the [Navarro-Frenk-White \(NFW\)](#) profile. Other halo profiles of this form include the [Hernquist \(1990\)](#) profile $(\alpha, \beta) = (1, 3)$, and the [Moore et al. \(1999b\)](#) profile $(\alpha, \beta) = (1.5, 1.5)$. ρ_s and r_s are the scale density and radius, respectively. The scale radius, is defined as

$$r_s = \frac{r_{\text{vir}}}{c}, \quad (1.28)$$

where r_{vir} is the virial radius, and c is the concentration of the halo characterized by an adopted mass-concentration relation for dark matter halos (e.g. [Bullock et al., 2001a](#); [Duffy et al., 2008](#); [Mandelbaum et al., 2008](#); [Diemer & Kravtsov, 2015](#); [Correa et al., 2015](#); [Klypin et al., 2016](#); [Ishiyama et al., 2021](#)). The concentration is a measure of how the matter within the halo is distributed. Larger values of c indicate more matter concentrated near the centre of the halo, whereas smaller values indicate a more extended distribution of the matter. The virial mass can be obtained trivially by integrating over the assumed density profile for the halo up to the virial radius

$$M_{\text{vir}} = \int_0^{R_{\text{vir}}} 4\pi r^2 \rho(r|M) dr. \quad (1.29)$$

Profiles that do not follow the same base functional form as the [NFW](#) profile do exist, such as the [Einasto \(1965\)](#) profile, which follows the same functional form as Sérsic's law,

$$\rho(r) = \rho_s \exp\left(-\frac{2}{\alpha} \left[\frac{r}{r_s} - 1\right]^\alpha\right). \quad (1.30)$$

Although less commonly used, the Einasto profile is argued to be more flexible and more accurate on smaller scales due to the shape parameter α . [Diemer \(2023\)](#) describes an improved Einasto-like profile, a truncated Einasto profile that introduces two shape parameters controlling how steep the profile becomes at two different radii. More specifically, the profile introduced by [Diemer \(2023\)](#) models the density of orbiting material (e.g., dark matter particles within a simulation) within dark matter halos. At larger radii, the model

captures the transition to the infalling regime, where material is more loosely bound, by introducing a truncation at a characteristic radius. This truncation reflects the sharp decline in the density of orbiting matter and the increasing dominance of infalling particles beyond that radius. Figure 1.3 shows how the density profiles of several models discussed in this section vary with radius. Generally, there is good agreement on what the halo density profile for an isolated halo is.

The [NFW](#) profile is a universal and widely used description of dark matter-only halos, but it is limited in that it only works well to describe regions within the halo radius. Issues begin to arise when trying to generalize the [NFW](#) profile to the dark matter outside the halo radius. On large scales, the clustering of halos leads to the contribution of 2-halo terms to the overall mass profile, which results in the standard [NFW](#) halo model being insufficient on large scales. ‘Amending’ the halo models we currently use to account for such problems is still under discussion and of interest today ([Hayashi & White, 2008](#); [Tavio et al., 2008](#); [Chen & Afshordi, 2020](#)).

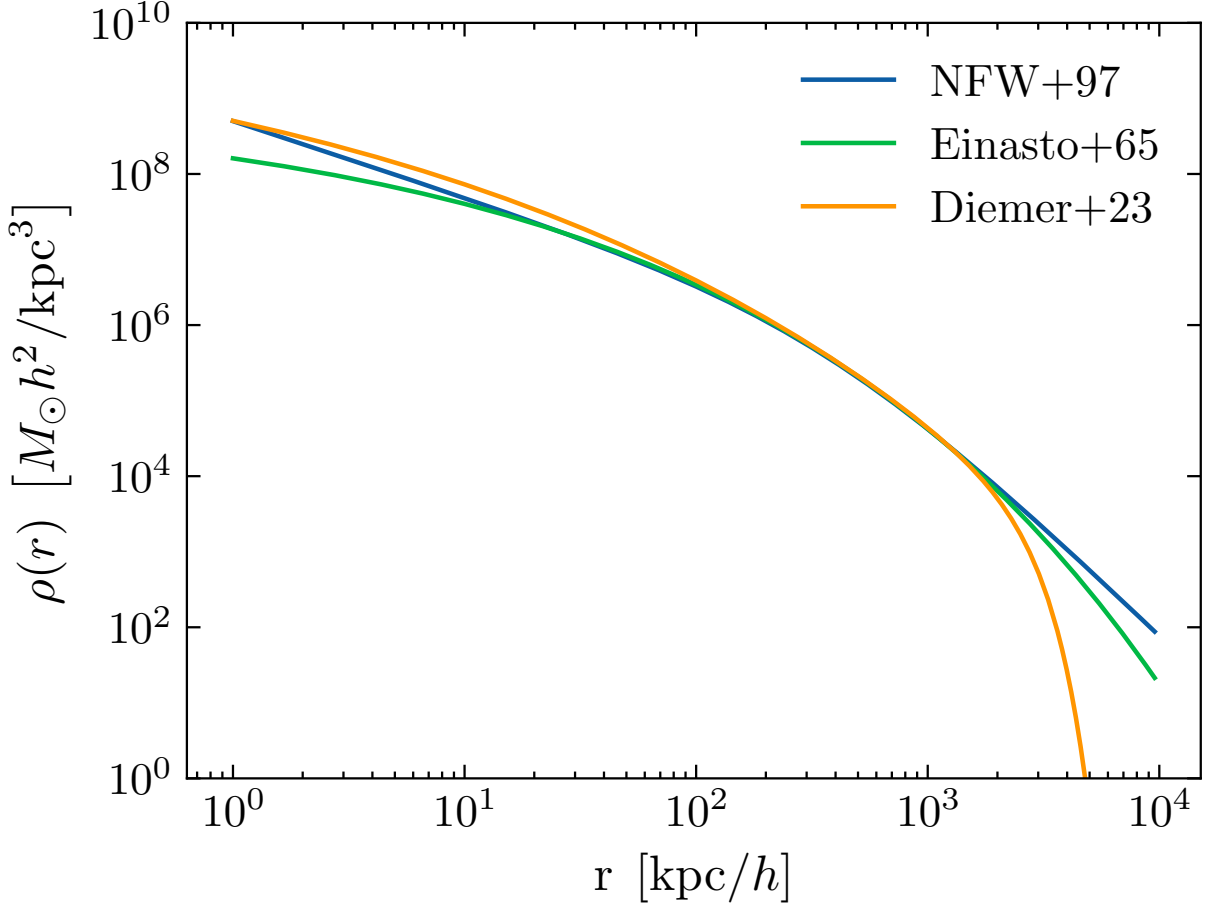


Figure 1.3: Illustration of some commonly used halo density profiles. The curves plotted were generated using the `python` package `Colossus` by [Diemer \(2018\)](#). These density profiles are for a hypothetical halo of mass $M_h = 10^{15} M_\odot$ and concentration of $c = 5$ at a redshift of $z = 0$. The adopted mass definition is the virial mass. The adopted cosmology used to generate this plot is the standard Planck2018 cosmology ([Aghanim et al., 2020](#)).

1.2 Galaxy Formation & Evolution

Based on the discussion in the previous sections, dark matter halos can be thought of as a ‘skeleton’ in which many baryonic processes occur. Baryons decouple from photons later in the early universe than the dark matter. Once free, the baryons settle within the deep potential wells of virialized dark matter halos. Tracking the evolution of baryons

is more complex than tracking the evolution of dark matter halos, primarily due to the additional interactions that baryons experience amongst other baryons. In this section, we discuss the key physical processes that drive galaxy formation and evolution, such as gas accretion/cooling, star formation, feedback, and environmental effects.

Humanity’s attempts at studying galaxies date back to the late 1700s. French astronomer Charles Messier was the first to identify and begin to catalogue galaxies (referred to at the time as “nebulae”), despite not knowing what they truly were. Debates on the nature of these objects were not solidified till accurate (for the time) measurements of the distances and radial velocities of galaxies were presented by [Hubble \(1929\)](#) that showed definitive evidence that these “nebulae” were extragalactic. Despite having catalogued and studied millions of galaxies, there still does not exist a universally agreed-upon definition for what a *galaxy* is. [Willman & Strader \(2012\)](#) proposed that “a galaxy is a gravitationally bound collection of stars whose properties cannot be explained by a combination of baryons and Newton’s laws of gravity”.

1.2.1 Gas Accretion & Stellar Mass Buildup

The visible light we observe from galaxies is a direct result of star formation from baryonic matter trapped within the potential wells of halos. Gas necessary for star formation may accrete through two methods: cold and hot accretion. The conventional picture for galaxy formation is one where gas is shock-heated to the virial temperature of the halo before cooling radiatively (e.g., Bremsstrahlung, collisional excitation/ionization, and recombination), fragmenting based on the principles of Jeans’ instability ([Jeans, 1902](#)) and eventually, forming stars ([Rees & Ostriker, 1977](#); [Silk, 1977](#)). This describes the mode of *hot accretion*, which is typical for high-mass systems. For halos where the cooling radius exceeds the virial radius, any shock-heated gas is unable to cool efficiently. The cooling radius is defined as the radius at which the cooling time of the gas is equal to the age of the universe (at a given redshift) ([Keres et al., 2005](#)). However, in halos below the critical mass threshold of $10^{11.5} - 10^{12} M_{\odot}$ ([Dekel & Birnboim, 2006](#)) stable virial shocks do not form, and gas accretes cold along filaments without ever reaching the virial temperature. This mode is referred to as *cold accretion* ([Katz et al., 2003](#); [Keres et al., 2005](#)). The mode of gas accretion is dependent on redshift. The cooling time and gas density are inversely related, $t_{\text{cool}} \propto \rho^{-1}$; thus, at high redshifts, the cooling time, t_{cool} , is shorter as the density is higher. This results in cold accretion being the dominant mode for gas accretion (at fixed virial temperature). Additionally, there is a dependence on halo mass, as larger systems have higher virial temperatures, leading to longer cooling times in which hot accretion becomes

dominant (Birnboim & Dekel, 2003; Van de Voort et al., 2011). Thus, the mode of accretion changes over time. A natural question from this is, *how do the differences in accretion modes affect galaxy properties?* Brooks et al. (2009) used hydrodynamical simulations to show that cold accretion through cold flows from filaments is the most dominant growth factor to galaxies, especially at early times, which would result in early star formation and, by construction, galaxies of higher stellar masses. Contrarily, Kocjan et al. (2024) find evidence in simulations that gas via cold accretion takes several hundred megayears longer to form stars than hot-accreted gas. Yoshida et al. (2002) compares different techniques for tracking the evolution of gas accretion and cooling and finds no discernible differences in both the total amount of cooled gas and properties of individual galaxies. These works highlight the ongoing efforts in better understanding the true impact accretion in the early universe has on the galaxies we see today.

Over billions of years, galaxies convert gas into stars, steadily increasing their stellar mass. The **Star Formation Rate (SFR)**, to first order, can be considered a proxy for quantifying how the mass of a galaxy changes over time. The **SFR** is correlated to the amount of gas available to form stars. The Kennicutt-Schmidt law (Schmidt, 1959; Kennicutt, 1989) states that the **SFR** surface density follows a power law relationship with the total gas surface density, $\Sigma_{\text{SFR}} \propto (\Sigma_{\text{gas}})^n$, where $n = 1.4 \pm 0.15$ (Kennicutt, 1998). **SFR** is a quantity that directly impacts the ability of a galaxy to accumulate stellar mass over its evolutionary history. **SFR** is strongly dependent on redshift and halo mass. In general, more massive halos at $z = 0$ have **SFRs** that peak much earlier in their formation history, allowing them to accumulate more stellar mass (Behroozi et al., 2013). Halos of mass $M_h \approx 10^{12} M_\odot$ are the most efficient at star formation (Marinoni & Hudson, 2002; Conroy & Wechsler, 2009; Behroozi et al., 2013). As seen in Figure 1.4a, a galaxy’s **SFR** is not linear over time, and many other factors, internal and external to a galaxy and halo, need to be accounted for. Some of these factors will be discussed in a subsequent section.

1.2.2 Stellar Mass Function

The **Stellar Mass Function (SMF)** is very similar to the previously discussed **HMF**. The **SMF** describes the number density of galaxies as a function of their stellar mass. Figure 1.4b, taken from Behroozi et al. (2013), shows the **SMF** and its evolution with redshift, comparing both observational data and best-fit models. We see that the **SMF** does not look like the **HMF**, which highlights the fact that baryons and dark matter do not evolve identically. We see that for stellar masses larger than $M_\star \approx 10^{11} M_\odot$, there is an exponential cut-off in the number of galaxies. Below $M_\star \approx 10^{11} M_\odot$, the **SMF** follows a power relation, from which at even lower masses, less than $M_\star \approx 10^{9.5} M_\odot$, we see a steep increase in the

number of low stellar mass galaxies (Baldry et al., 2012; Moustakas et al., 2013; Tomczak et al., 2014; Wright et al., 2017). Considering the dependence on redshift, we find that the SMF’s overall trend of decreasing number density with increasing stellar mass remains, but the transition between the power-law regime and exponential cut-off is ill-defined. We, however, have seen the trend observationally that at high redshifts the SMF shifts to lower stellar mass (e.g. Song et al., 2016).

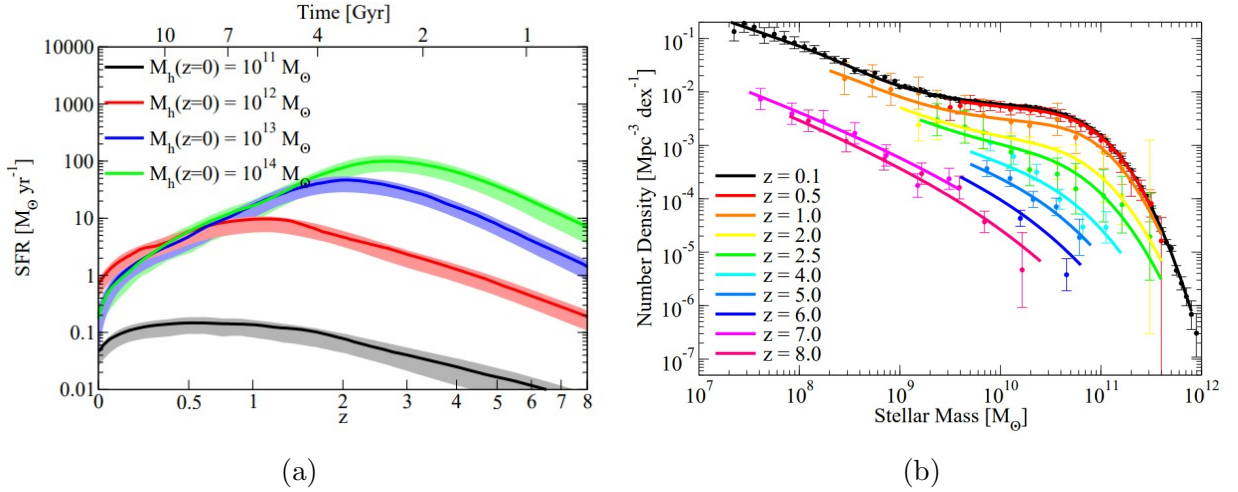


Figure 1.4: a) Plot of the average SFR for galaxies in halos of the indicated halo mass (coloured solid lines) against redshift. These halo masses are masses seen for halos at the present day, $z = 0$. The shaded region indicates the 1σ uncertainties. This is Figure 6 (right panel) from Behroozi et al. (2013). b) Plot of the SMF as a function of stellar mass for various different redshifts. The solid colour lines are the best-fitting models that have been fit to observational data (colour data points with error bars). This is Figure 3 (top panel) from Behroozi et al. (2013).

1.2.3 Mechanisms that Alter the Evolution of Galaxies

As stated in the previous section, the SMF and HMF, although similar in what they quantify, are very different in nature. Baryons and dark matter do not evolve in the same manner due to baryon-baryon interactions. These interactions can result from both internal processes that occur within a galaxy and external processes that arise from its environment. In this next section, we briefly describe key factors that must be accounted for in galaxy formation models that affect the evolution of galaxies.

Internal Mechanisms

Apart from the processes previously described (cooling, cold/hot accretion, etc.), internal to galaxies, *feedback* is one of the most dominant factors for the decrease in [SFRs](#). One form of feedback is supernovae feedback resulting from the explosion of massive stars. As massive stars run out of fuel to maintain nuclear fusion in their cores, they can no longer sustain the pressure needed to offset gravitational collapse, thus resulting in the catastrophic expulsion of their outer layers, leaving behind a dense remnant. The resulting explosion transfers energy into the surrounding medium, which can be broken down into two components, kinetic and thermal. Kinetically, supernovae drive turbulence in the interstellar medium and circumgalactic medium and can eject gas outside the galaxy (an outflow), losing the material required to form stars ([Ostriker & Shetty, 2011](#); [Oku et al., 2022](#)). This effect is most prominent in low stellar mass galaxies ([Dekel & Silk, 1986](#)). Thermally, supernovae can generate hot bubbles, which are rapidly expanding envelopes of hot ionized gas ([McKee & Ostriker, 1977](#); [Spitzer, 1990](#); [Oey, 2004](#)). The reheated gas cannot form stars at the same rate and must cool radiatively again, thus reducing the [SFR](#).

[Active Galactic Nuclei \(AGN\)](#) feedback has a large impact on galaxy evolution. [AGN](#) feedback is an energetic process that occurs as a supermassive black hole (SMBH) accretes matter, forming an accretion disc and jets that produce bubbles. This type of feedback is more prominent in massive elliptical-type galaxies, which are often centres of large groups or galaxy clusters. Similar to supernovae feedback, there are two distinct modes. The quasar mode, which occurs when a large amount of matter accretes onto the SMBH, and the radio mode, which occurs when the SMBH accretes at a slower rate. A popular model for describing the suppression of gas cooling in massive galaxies is energetic jets during the radio mode ([Silk et al., 2014](#)), thus resulting in a decrease in [SFR](#). [AGN](#) feedback also works to explain the *overcooling* problem ([Thacker & Couchman, 2000](#); [Balogh et al., 2001](#)), which describes the discrepancy between the total stellar mass of massive galaxies in simulations being larger than what is observed. In simulations, without accounting for [AGN](#) feedback, stars are able to form in excess ([Ardila et al., 2020](#)). On the contrary, there are instances of positive [AGN](#) feedback. As [AGN](#) outflows expand, they compress surrounding clouds, increasing their density, which accelerates molecular cloud formation and thus results in a short-lived starburst ([Gaibler et al., 2012](#); [Zinn et al., 2013](#); [Contini et al., 2024](#); [Clavijo-Bohórquez et al., 2024](#)).

External Mechanisms

The environment in which a galaxy forms and subsequently evolves has an immense impact on galaxy properties. It is suggested that about 50 – 70% of galaxies reside in dense environments like groups or clusters (Vollmer, 2013). This also suggests that at some point in time, many of these galaxies were field galaxies in a lower-density, isolated environment, which later fell into the deep potential well of an overall group or cluster. This process of falling into a dense environment causes a slew of dynamical processes that alter the growth and evolution of a galaxy. There have been many efforts in studying field vs. group/cluster galaxies to discern compelling evidence of differences between their galaxy properties (Moore et al., 1996; Bernardi et al., 1998; Balogh et al., 1999; Rettura et al., 2010; Saracco et al., 2017; Abdellah et al., 2025). The consensus via observational and simulation studies is that cluster/group galaxies are predominantly elliptical (early-type) galaxies that are no longer star-forming. Field galaxies are predominantly spiral, irregular-type galaxies (late-type), which show a significantly higher fraction of active star formation.

There are a plethora of gravitationally and dynamically driven processes that alter galaxies as they fall into a group/cluster and over time, while traversing a deep potential well. We briefly summarize these effects. A non-negligible number of infalling galaxies experience fast encounters, also known as “fly-bys,” with surrounding cluster/group members. These gravitationally driven fly-bys are more efficient at stripping away baryonic matter from low-mass galaxies whose potential wells are far shallower than a larger galaxy. This is called *galaxy harassment* (Moore et al., 1996, 1999a). The more gradual stripping of matter from gravitational encounters between two galaxies due to dynamical friction is called *tidal stripping* (Merritt, 1984; Cypriano et al., 2006). The tidal radius is proportional to $M_{\text{sat}}^{1/3}$ and $M_{\text{host}}^{-1/3}$, which effectively encodes a comparison between the mean density of the satellite and that of the host. In other words, tidal stripping is governed by the ratio of their densities. The difference between the mean density of the satellite within the tidal radius and the mean density of the host at the satellite’s orbital radius dictates the stripping strength. For example, satellites with significantly lower mean densities are more susceptible to tidal stripping, as with satellites residing in more massive/denser host halos. Studies have found that tidal stripping is not only dependent on a comparison of mean densities but also on the orbits of satellites and their stars (e.g. Read et al., 2006a). Another consequence of infall is the rapid removal of gas within galaxies due to the intercluster medium, in a process called *ram pressure stripping* (Abadi et al., 1999; Boselli et al., 2009; Brown et al., 2023). Ram pressure stripping works to deprive galaxies of an ongoing supply of gas necessary to sustain star formation, and so, these galaxies slowly become quiescent. This is called *galaxy starvation* or *strangulation* (Larson et al., 1980;

Balogh et al., 1999; Bekki et al., 2002). Finally, galaxy mergers and collisions driven by dynamical friction play an important role in explaining the tidal features of cluster galaxies (Toomre & Toomre, 1972) in addition to the characteristic sizes of galaxies, which we explore further in the subsequent section.

1.2.4 The Sizes of Galaxies

In this section, we explore the galaxy observable of size from a theoretical and observational standpoint. We aim to explore *how galaxies get their size?* as well as observational results concerning the evolution of galaxy size. For a theoretical review of these topics, we reference Chapter 5 of Benson (2010) and the corresponding literature within.

Theoretical

The picture of how late-type galaxies form through galactic disks is well understood. The fundamental process at play is the conservation of angular momentum. A system that is collapsing under the influence of its own gravity will cease to collapse when the rotational support, which induces a centrifugal force, can offset the gravitational potential. Hoyle (1949)¹ argued that the angular momentum imparted to protogalactic gas, which eventually forms galactic discs, arises from the same mechanism that gives dark matter halos their angular momentum—tidal torque from surrounding Large Scale Structure (LSS). This tidal torque theory was later analyzed by Peebles (1969) using linear perturbation theory, followed by Doroshkevich (1970); White (1984). To first order, the physics determining the final size of the galaxy disc also sets the characteristic size of the galaxy itself when neglecting external effects. This physics has been worked out and described in Fall & Efstathiou (1980). Given that the angular momentum distribution is known for the baryonic matter within halos, the rotationally supported disc is described by

$$\frac{j^2(M)}{R^3(M)} = \frac{\partial}{\partial R} \Phi(R), \quad (1.31)$$

where $j(M)$ is the specific angular momentum with enclosed mass M , $\Phi(R)$ is the potential of the galactic disc, and $R(M)$ is the radius of the disc enclosing mass M , which is what is solved. To summarize, the physical process is as follows:

¹Sir Fred Hoyle’s contributions to galaxy formation theory, along with documented transcripts of discussions/presentations, can be found in Efstathiou (2003).

1. An initial angular momentum distribution is imparted to the dark matter and baryonic matter due to nearby [LSS](#).
2. The dark matter and baryonic matter begin to collapse together in overdense regions. Since the baryon density at early times is low, they are unable to cool and remain coupled to the dark matter.
3. During the collapse, the baryon density begins to rise as they fall towards the centre of the halo. When the baryon density is sufficiently high, they decouple from dark matter and begin to cool and collapse under their own gravity.
4. As the baryons collapse due to radiative cooling, they accelerate. Since they started with a non-zero angular momentum distribution, the baryons cannot collapse all the way to the centre of the halo and thus settle into a rapidly rotating disc, whose vertical thickness decreases over time while its radial extent is maintained to conserve angular momentum.

The formation of early-type galaxies, such as elliptical and spheroidal galaxies, remains less clear. Two different pathways have been proposed. First, generally elliptical-type galaxies tend to have lower net angular momentum ([Fall & Romanowsky, 2013](#)), making them more pressure supported than late-type disc galaxies. This has led to the view that such galaxies assemble around dense stellar cores formed from low angular momentum gas that efficiently cooled and condensed at early times ([Naab & Ostriker, 2017](#)). Secondly, the growth and size evolution of early types are thought to be influenced primarily by minor mergers, where smaller satellite galaxies accrete onto more massive central galaxies in predominantly collisionless (dry) mergers. For a spheroidal system undergoing collisionless minor mergers, satellite stars are stripped from larger radii, leading to the formation of extended stellar envelopes in early-type galaxies, effectively increasing the galaxy size ([Naab & Ostriker, 2017](#)). It has been shown within hydrodynamical simulations that elliptical galaxies are compact at high redshifts (which is in agreement with observations; see next section) and that the accretion of stripped stellar material from minor mergers increases the effective size of the system (e.g. [Naab et al., 2009](#); [Hilz et al., 2013](#); [Bédorf & Portegies Zwart, 2013](#)).

Observational

High-redshift spectroscopic surveys have shown that early-type galaxies of redshift $z > 1 - 1.5$ are smaller in size than early-type galaxies of the same stellar mass at present-day

(e.g. Daddi et al., 2005; Van Dokkum et al., 2008; Trujillo et al., 2009). Analysis conducted by Newman et al. (2012) using the CANDELS survey revealed that, within the redshift range of $0.4 < z < 2.5$, the mean size of galaxies at a fixed stellar mass increased by a factor of 3.5 ± 0.3 with decreasing redshift. Van der Wel et al. (2008) compiles data sets of high-redshift early-type galaxies up to $z = 2.5$ and finds a size evolution with redshift of $R_{\text{eff}} \propto (1 + z)^{-1.20 \pm 0.12}$. The leading explanation for the size evolution of early-type galaxies remains the minor merger argument highlighted previously.

Given the clear evidence for a strong redshift dependence in the sizes of early-type galaxies, it is valuable to explore how galaxy size may be connected to the size of their dark matter halos. The virial radius (and halo mass) is typically defined by 200 times the critical mass density, $\rho_{\text{crit}}(z)$. The critical density is a function of redshift, $\rho_{\text{crit}}(z) \propto H(z)^2$, where at high- z the expansion rate is large and thus, $\rho_{\text{crit}}(z)$ increases rapidly. As a consequence, halos of the same mass, at high redshift, have significantly higher mean densities and are more compact. Naturally, we wonder whether galaxies that form in these compact high- z halos are also compact and smaller in size. Remarkably, Kravtsov (2013) found that over eight orders of magnitude in stellar mass, galaxies of all morphological types follow a linear relation between half-mass radius ($r_{1/2}$) and virial radius (R_{200c}) of $r_{1/2} \approx 0.015 R_{200c}$. This relation is also consistent with models that assume galaxy size is dependent on the halo angular momentum (λ), $r_{1/2} \propto \lambda R_{200c}$. These works demonstrate that there should exist an inherent difference in galaxy sizes based solely on the formation time of the halo in which they reside. This ultimately adds further complexity to our understanding of the evolution of galaxy size.

We note that halo mass and virial radius are correlated: $M_{200} \propto R_{200c}^3$. Additionally, in the next section, we will discuss that the correlation between stellar mass and galaxy size is well-established. This poses the question, *which is more fundamental?* Addressing this question is an objective of this thesis.

1.2.5 Scaling Relations of Galaxy Observables

Understanding the correlation between galaxy observables in the form of scaling relationships enables us to better model galaxy formation and evolution within simulations. Scaling relations are empirical parameterizations that connect one galaxy property to another. In this section, several key scaling relations that are most relevant to the work of this thesis will be discussed.

Comprehensive studies of galaxies, in particular, early-type, elliptical galaxies, involve measurements of effective radius (R_{eff}), average surface brightness (μ_e), and velocity disper-

sion (σ). The average surface brightness is determined by measuring the flux, or luminosity, of the observed galaxy, $\langle I_e \rangle = L/2\pi R_{\text{eff}}^2$. The correlation between these three quantities forms the fundamental plane (FP), which has a general form of

$$\log_{10}(R_{\text{eff}}) = \alpha \log_{10}(\sigma) + \beta \log_{10}\langle I_e \rangle + \gamma. \quad (1.32)$$

There have been many studies of the FP using different observational datasets and surveys, all of which present their own constraints for α, β, γ (Djorgovski & Davis, 1987; Dressler et al., 1987; Jørgensen et al., 1996; Hyde & Bernardi, 2009b; Samir et al., 2020; De Graaff et al., 2021; D’Eugenio et al., 2021; D’Onofrio & Chiosi, 2024). The FP provides insight into three impactful galaxy observables and their correlation, which further enabled studies of mass-to-light ratios (Prugniel & Simien, 1996; De Graaff et al., 2021; D’Eugenio et al., 2021) and provided a method to estimate distances to galaxies (Dressler et al., 1987; Djorgovski & Davis, 1987).

The galaxy luminosity-size relation is tight when studied observationally and within simulations. The overall trend found is that more luminous galaxies are larger in effective radius (Hyde & Bernardi, 2009a; Favole et al., 2018). Luminosity and stellar mass, M_* , are coupled quantities in that galaxies with inherently more stellar mass content are also more luminous. As such, the stellar mass-size relation, which we will explore with our data in Chapter 2, follows the same trend: more massive galaxies in terms of stellar content are larger in physical size. The redshift evolution of the stellar mass-size relation shows that the overall positive correlation remains the same when varying redshift, but at fixed stellar mass, galaxy size is shown to decrease with increasing redshift (Van der Wel et al., 2014; Roy et al., 2018; Yang et al., 2024). Yoon et al. (2017) finds evidence of environmental dependence in the stellar mass-size relation for galaxies heavier than $M_* > 10^{11.2} M_\odot$, such that these galaxies are 20% – 40% larger than similar galaxies in underdense regions. This is in contrast to Nair et al. (2010), who argues there is no environmental dependence in the stellar mass-size relation. The environmental dependence on this relationship is still debated and actively investigated within the literature.

A key scaling relation to understanding how baryonic matter and dark matter co-evolve is the **Stellar-to-Halo Mass Relation (SHMR)**. As discussed in Section 1.2.1, dark matter halos of mass $M_h = 10^{12} M_\odot$ are the most efficient at star formation. Studies of the stellar-to-halo mass ratio would be a method of verifying these findings. If halos of mass $M_h = 10^{12} M_\odot$ are truly the most efficient at accreting stellar mass, then we might expect to see the ratio of M_*/M_h to peak at an approximate halo mass of $10^{12} M_\odot$. Behroozi et al. (2010a) finds that the integrated star formation peaks at a halo mass of $M_h = 7 \times 10^{11} M_\odot$ and increases by a factor of 5 to $z = 4$. The increase in the peak from $z = 0$ to $z = 4$

shows redshift evolution of the [SHMR](#). Further evidence for the redshift evolution of the [SHMR](#), using weak gravitational lensing alone, was by [Hudson et al. \(2015\)](#). Studies of the [SHMR](#) allow us to better understand the co-evolution of stellar mass and halo mass in different environments. Observationally, we find evidence that the [SHMR](#) for satellite galaxies decreases with decreasing halo-centre distance, which provides evidence of mass loss due to stripping processes ([Niemiec et al., 2017](#); [Wang et al., 2024](#); [Kumar & More, 2024](#))². The [SHMR](#) is an important scaling relation to understand if we hope to simulate galaxy formation and evolution accurately.

1.3 Galaxy-Halo Connection Models

Models that correlate galaxy observables with halo properties (usually halo mass) are referred to as galaxy-halo connection models. Galaxy-halo connection models range from physical models, which involve hydrodynamical simulations and semi-analytic models, to empirical models such as [Halo Abundance Matching \(HAM\)](#), [Subhalo Abundance Matching \(SHAM\)](#), and the [HOD](#). In this section, we focus primarily on the latter, empirical models, as they are most relevant to this work.

The goal of [HAM](#) and [SHAM](#) models is to map galaxies onto simulated dark matter halos to reproduce observed trends. Commonly, halo mass M_h or proxies like maximum circular velocity V_{max} , describe halos, while luminosity or stellar mass characterize galaxies. These galaxy properties are matched to simulations by fitting the observed [SMF](#) or [Luminosity Function \(LF\)](#). The main difference between [HAM](#) and [SHAM](#) is that [SHAM](#) includes satellite subhalos (as [SHAMs](#) are inherently [HAMs](#), we hereafter refer only to [SHAM](#)). Both still assume that the most massive or luminous galaxies reside in the most massive halos. In reality, this relation is not strictly monotonic. More accurate [SHAM](#) models include scatter and stochasticity to account for uncertainties (e.g. [Behroozi et al., 2010b](#); [Leauthaud et al., 2011](#); [Trujillo-Gomez et al., 2011](#)). For example, [Behroozi et al. \(2010b\)](#) incorporates scatter in the [SHMR](#) for central galaxies and uncertainties in the [SMF](#) and [HMF](#). Overall, [SHAM](#) reproduce observed [SMFs](#) (e.g. [Behroozi et al., 2010b](#); [Moster et al., 2013](#); [Reddick et al., 2013](#)), [LFs](#) (e.g. [Vale & Ostriker, 2004](#); [Tasitsiomi et al., 2004](#); [Hearin et al., 2013](#)), galaxy clustering (e.g. [Conroy et al., 2006](#); [Reddick et al., 2013](#); [Masaki et al., 2023](#)), and weak lensing (e.g. [Contreras et al., 2023](#)).

²[Kumar & More \(2024\)](#) analyzes the luminosity-halo mass relation for satellite galaxies, which can be considered to be robust as luminosities can be obtained directly from flux measurements of galaxies and do not require any assumptions like what is needed to obtain stellar mass values.

HODs are another empirical method in modelling the relationship between galaxies and dark matter halos, which does so with halo occupation statistics. For both central and satellite galaxies alike, a probability distribution, denoted by $P(N|M)$, is used to describe the number of galaxies within a halo of mass M . The use of **HODs** was pioneered by many in the early 2000s (e.g. [Benson et al., 2000](#); [Peacock & Smith, 2000](#); [White et al., 2001](#); [Bullock et al., 2002](#); [Berlind & Weinberg, 2002](#); [Kravtsov et al., 2004](#); [Zheng et al., 2005](#); [Tinker et al., 2008a](#); [Reddick et al., 2013](#)). In its simplest form, **HODs** are described by up to five parameters; for example, the commonly used parametrization from [Zheng et al. \(2005\)](#) for either luminosity- or stellar mass-limited galaxy samples is

$$\langle N_{\text{central}} \rangle_M = \frac{1}{2} \left[1 + \text{erf} \left(\frac{\log M - \log M_{\min}}{\sigma_{\log M}} \right) \right], \quad (1.33)$$

for central galaxies, and,

$$\langle N_{\text{satellite}} \rangle_M = \left[\frac{M - M_0}{M'_1} \right]^\alpha, \quad (1.34)$$

for satellite galaxies when $M > M_0$. The parameter $M_{\min}$ sets the characteristic minimum mass of halos such that they can host a central galaxy, $\sigma_{\log M}$ reflects the scatter in the luminosity-halo mass relation of central galaxies, α controls the slope of the satellite occupation, M'_1 is a normalization for satellite occupation numbers, and M_0 reflects a cut-off mass below which no satellites are populated into host halos. With this model, either a 3-parameter fit of $\{M_{\min}, M'_1, \alpha\}$ or a 5-parameter fit with $\{M_{\min}, \sigma_{\log M}, M'_1, M_0, \alpha\}$ can be adopted. Using this base **HOD** prescription, extensions can be made to include other correlations, such as galaxy colour, as done by [Skibba & Sheth \(2009\)](#).

A powerful extension to **HODs** is the use of the **Conditional Luminosity Function (CLF)**. The **CLF** ([Yang et al., 2003](#)), $\Phi(L|M_h)dL$, describes the number of galaxies between luminosities of $L + dL$ and $L - dL$ that reside in a halo of mass, M_h , which can be implemented to fit to an observed **LF**. Similar to standard occupation statistics, the **CLF** is broken down into two components, one for centrals and one for satellites.

$$\Phi(L|M) = \Phi_{\text{central}}(L|M) + \Phi_{\text{satellite}}(L|M). \quad (1.35)$$

A lognormal distribution is chosen for centrals and a modified Schechter function for satellites. For example, the parametrization used by [Yang et al. \(2008\)](#) is

$$\Phi_{\text{central}}(L|M) = \frac{A}{\sqrt{2\pi}\sigma_c} \exp \left[-\frac{(\log L - \log L_c)^2}{2\sigma_c^2} \right], \quad (1.36)$$

and,

$$\Phi_{\text{satellite}}(L|M) = \phi_s^* \left(\frac{L}{L_s^*} \right)^{(\alpha_s^*+1)} \exp \left[- \left(\frac{L}{L_s^*} \right)^2 \right]. \quad (1.37)$$

This prescription has five free parameters. $\{L_c, \sigma_c, \phi_s^*, \alpha_s^*, L_s^*\}$. CLF-based HOD models have been used to model galaxy clustering/abundance and weak lensing measurements, independently (e.g. Zehavi et al., 2011) and jointly (Cacciato et al., 2012, 2013; More et al., 2013).

In this thesis, we employ a Conditional Stellar Mass Function (CSMF)-based HOD model, which is analogous to the workings of a CLF HOD, but instead, quantifies the number of galaxies between the stellar mass of $M_* + dM_*$ and $M_* - dM_*$, which reside in a halo of mass, M_h . The details of this model are described in Chapter 4. Examples of works that utilize a CSMF HOD include Cacciato et al. (2009a); Yang et al. (2009); Zu & Mandelbaum (2015); Van Uitert et al. (2016); Dvornik et al. (2023) and Burger et al. (2025).

1.4 Weak Gravitational Lensing

From the postulates of General Relativity (GR), Einstein (1911)³ theorized that light from a celestial body gets deflected when passing by another massive body due to its gravitational influence. The deflection angle, $\delta\theta$, due to a point mass, M , is

$$\delta\theta = \frac{4GM}{c^2 b}, \quad (1.38)$$

where G is the gravitational constant, c is the speed of light, and b is the impact parameter. The constant factor of 4 was predicted by Einstein’s theory of GR. The well-known 1919 expedition led by Arthur Eddington and Frank Dyson photographed stars in the Hyades cluster near the Sun’s limb during a total solar eclipse. They measured the stars’ positions to be 1.9'' offset from their expected positions, ultimately confirming Einstein’s predictions and providing strong evidence for GR. It was still not until the work of Walsh et al. (1979) that the first extragalactic direct lensing detection was observed through imaging of the binary quasar system of 0957+561 A and B.

In this section, we introduce the mathematical formalisms of weak gravitational lensing and Galaxy-Galaxy Lensing (GGL). The information presented follows Dodelson (2017),

³English translation for Einstein (1911) <https://einsteinpapers.press.princeton.edu/vol3-trans/393>

from which we focus on key topics relevant to the work of this thesis. Gravitational lensing describes the distortions in the shape and brightness of distant objects, referred to as *sources*, due to foreground objects, referred to as *lenses*⁴. The stretching of the sources (or *shear*) and the brightening (or *magnification*) is described by the distortion tensor,

$$\psi_{ij} = \begin{pmatrix} \kappa + \gamma_1 & \gamma_2 \\ \gamma_2 & \kappa - \gamma_1 \end{pmatrix}, \quad (1.39)$$

where $\gamma_{1,2}$ quantifies the two components of shear, and the magnification is quantified by the convergence, κ . Within the regime of weak gravitational lensing and [GGL](#), the shears and magnification of background sources are very subtle. Thus, convergence is often used to define the transition between the strong and weak lensing regimes. When $\kappa > 1$, the regime of strong lensing is reached, typically characterized by the formation of multiple images, arcs, or Einstein rings of background sources. When $\kappa \ll 1$, the magnification of background sources is only observable by averaging over a large sample, thus marking the weak lensing regime. Convergence is defined as a function of the critical surface mass density, Σ_{crit} , and the mass density, $\Sigma(\vec{\theta})$, within an angular separation of $\vec{\theta}$,

$$\kappa(\theta) = \frac{\Sigma(\theta)}{\Sigma_{\text{crit}}}. \quad (1.40)$$

The critical surface mass density is defined as

$$\Sigma_{\text{crit}}(z_s, z_l) = \frac{c^2}{4\pi G} \frac{D_A(z_s)}{D_A(z_l) D_A(z_l, z_s)}, \quad (1.41)$$

where $D_A(z_s)$ is the angular diameter distance to the source galaxy at redshift z_s , $D_A(z_l)$ is the angular diameter distance to the lens galaxy at redshift z_l , and $D_A(z_l, z_s)$ is the distance between the source and lens galaxy at their respective redshifts.

Physically, shear is more interpretable when considering it in its two components of tangential (γ_t) and cross ($\gamma_\times$), which are linear combinations of γ_1 and γ_2 . Their definitions are

$$\gamma_t = -\gamma_1 \cos(2\phi) - \gamma_2 \sin(2\phi), \quad (1.42)$$

and,

$$\gamma_\times = \gamma_1 \sin(2\phi) - \gamma_2 \cos(2\phi). \quad (1.43)$$

The tangential shear offers a physical quantity that is commonly plotted within weak lensing studies, which is the [ESD](#) or $\Delta\Sigma$. [ESD](#) is defined as the difference between the

⁴To first order this is true. There are additional contributions to the distortions resulting from [LSS](#).

average surface mass density within a projected radius, R , and the surface mass density at R ,

$$\Delta\Sigma(R) = \overline{\Sigma(< R)} - \Sigma(R) = \gamma_t(R)\Sigma_{\text{crit}}(z_s, z_l). \quad (1.44)$$

Measurements of $\Delta\Sigma$ can be calculated using an estimator, which will be introduced and discussed in Chapter 3.

Overall, $\Delta\Sigma$ profiles are a combination of various halo terms, all of which dominate the lensing profile at various radii; this is depicted in Figure 1.5. The top panel depicts the lensing signal from a central galaxy of halo mass: $M_h = 5 \times 10^{12} M_\odot$. Small and intermediate scales are dominated by the 1-halo term, which originates from the galaxy's host halo. On large scales, the halo-halo term, or 2-halo term, which arises from the clustering of other neighbouring halos, dominates. The middle panel shows a typical lensing signal from a satellite galaxy. On small scales, the stripped central profile is the contribution from the satellite's stripped subhalo. Towards intermediate scales, we have the 1-halo term for satellites, which is also known as the offset-group term, and once again at large radii we have the 2-halo term. The bottom panel shows the overall lensing profile when combining centrals and satellites with an assumed satellite fraction of $\alpha = 0.2$.

Generally, weak lensing surveys obtain background galaxy shape measurements through photometric imaging, it is observationally more accessible to define the ellipticity as a function of measured axis lengths, a and b , and the position angle of the major axis of the imaged galaxy, ϕ . The following formalism is taken from Bartelmann & Schneider (2001). We define the surface brightness of an imaged galaxy at angular position $\vec{\theta}$ to be $I(\vec{\theta})$. The centre of the galaxy image is defined by

$$\vec{\bar{\theta}} = \frac{\int d^2\theta q_I[I(\vec{\theta})] \vec{\theta}}{\int d^2\theta q_I[I(\vec{\theta})]}, \quad (1.45)$$

where $q_I[I(\vec{\theta})]$ is a chosen weight function. From this definition, the tensor for the second brightness moments is defined as

$$Q_{ij} = \frac{\int d^2\theta q_I[I(\vec{\theta})] (\theta_i - \bar{\theta}_i) (\theta_j - \bar{\theta}_j)}{\int d^2\theta q_I[I(\vec{\theta})]}, \quad i, j \in [1, 2]. \quad (1.46)$$

The elements of the second brightness moment tensor can be used to define the shape, or the complex ellipticity, of the imaged galaxy

$$\epsilon \equiv \frac{Q_{11} - Q_{22} + 2iQ_{12}}{Q_{11} + Q_{22} + 2(Q_{11}Q_{22} - Q_{12}^2)^{1/2}}. \quad (1.47)$$

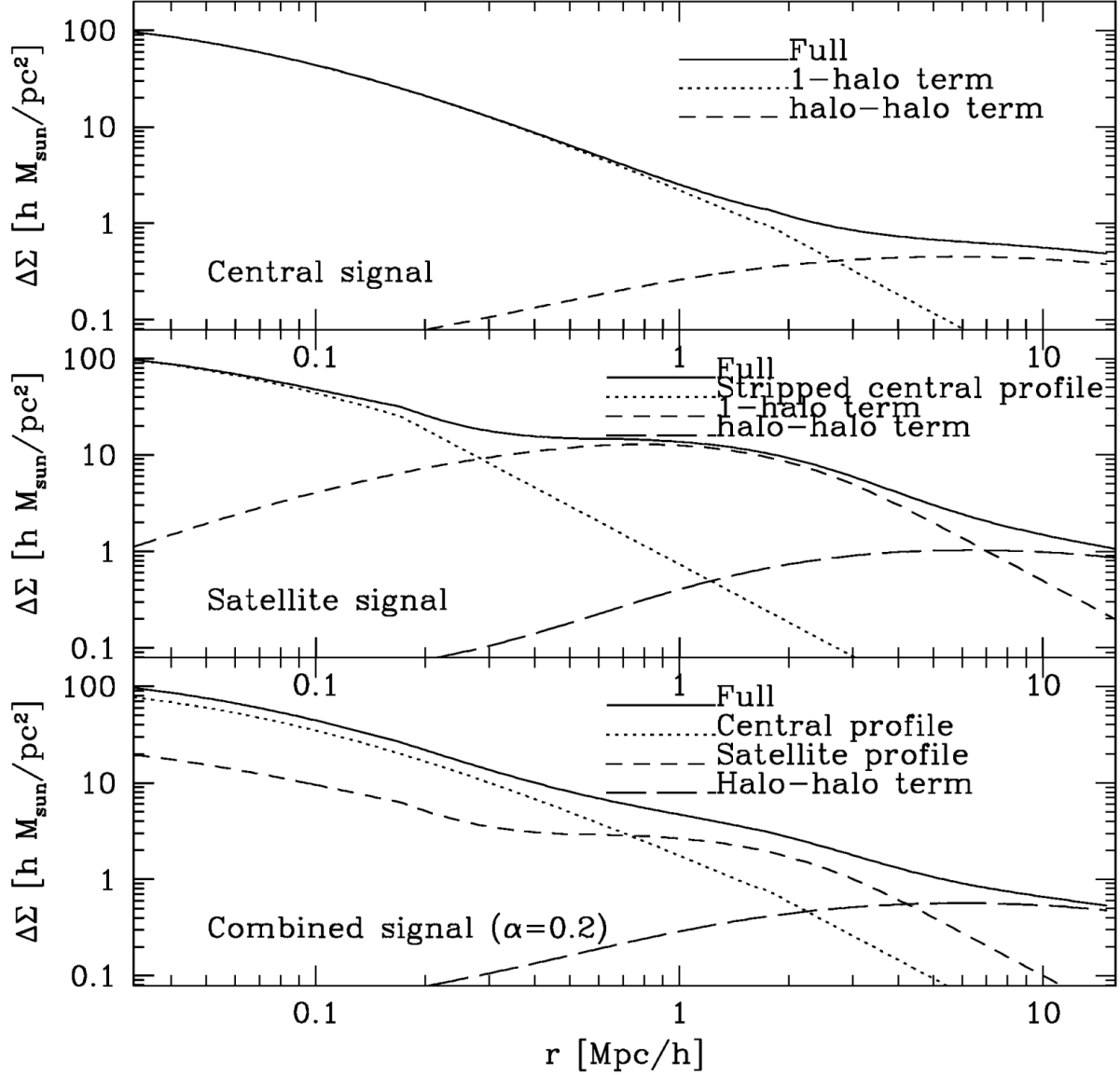


Figure 1.5: Figure 1 taken from [Mandelbaum et al. \(2005\)](#), which illustrates the various halo terms that contribute to central (top panel) and satellite (middle panel) galaxies to produce the overall lensing profile (bottom panel). For centrals, the 1-halo term and halo-halo term or 2-halo term dominate on small/intermediate and large scales, respectively. For satellites, small scales are dominated by the satellite's stripped subhalo. The satellite 1-halo term, also known as the offset-group term, dominates on intermediate scales. The 2-halo term once again dominates large scales. The halo mass for the central galaxy is $M_h = 5 \times 10^{12} M_\odot$. The satellite fraction assumed is $\alpha = 0.2$.

This can be further simplified based on the axis ratio, $r = b/a$, into

$$|\epsilon| = \frac{1-r}{1+r} = \frac{a-b}{a+b}. \quad (1.48)$$

Using the definition for the reduced shear, $g = \gamma/(1 - \kappa)$, the transformation between source and image ellipticities is

$$\epsilon = \begin{cases} \frac{\epsilon^{(s)} + g}{1 + g^* \epsilon^{(s)}} & \text{for } |g| \leq 1, \\ \frac{\epsilon^{(s)} + g \epsilon^{(s)*}}{\epsilon^{(s)*} + g^* \epsilon^{(s)}} & \text{for } |g| > 1, \end{cases} \quad (1.49)$$

where the asterisk denotes complex conjugation. In the transformation above, $\epsilon^{(s)}$ denotes the intrinsic (unlensed) ellipticity of the source galaxy, while ϵ denotes the observed ellipticity of the lensed image. In the limit of weak lensing, $\kappa \ll 1$, $|\gamma| \ll 1$, and thus $|g| \ll 1$. This means, to first order, $\epsilon \approx \epsilon^s + g$.

The dominant form of noise in weak lensing is *shape noise*. All unlensed galaxies have intrinsic shapes and intrinsic ellipticities. Fortunately, intrinsic ellipticities are random, such that the average of intrinsic shapes is 0, making it true that the average of tangential ellipticities is approximately equal to the measured tangential shear, $\langle e_t \rangle \approx g$.

As described earlier, the signal from weak lensing is a direct result of the lens galaxies and their matter distribution deflecting light from background source galaxies. The [Signal-to-Noise Ratio \(SNR\)](#) is given by

$$\left(\frac{S}{N} \right)^2 = \sum_i \frac{N_i \gamma_t^2(R_i)}{\sigma_{\text{shape}}^2}, \quad (1.50)$$

where R_i represents an annulus in projected radii, N_i is the number of galaxies within an annulus of R_i , and σ_{shape} is the shape noise. Ultimately, the power of weak lensing lies in the process of *stacking*, in which we combine the signal from many foreground lens galaxies in order to decrease shape noise, which scales as $1/\sqrt{N_i}$.

1.5 Observational Status of Halo Mass vs Galaxy Size Studies using GGL

This thesis investigates the potential dependence of halo mass on galaxy size at a fixed stellar mass and colour (red) using [GGL](#). In this section, the current observational results in

the literature concerning this correlation will be summarized. All studies discussed utilize **GGL** with different lenses and source galaxy catalogues.

Current literature on the connection between halo mass and galaxy size has yielded mixed conclusions. Early work by [Charlton et al. \(2017\)](#) (hereafter **C17**) using [Canada-France-Hawaii Telescope Lensing Survey \(CFHTLensing\)](#) ([Heymans et al., 2012](#)) reported a positive correlation, indicating that larger galaxies inhabit more massive halos. [Huang et al. \(2020\)](#) (hereafter **H20**) showed that, at fixed outer stellar mass, more extended galaxies correspond to higher halo masses. This result is consistent with the positive correlation detection from **C17**. However, [Sonnenfeld et al. \(2019\)](#) (hereafter **S19**), using [Hyper Suprime-Cam \(HSC\)](#) and BOSS data, reported no evidence for such a dependence at fixed stellar mass, and similar works by the same authors, [Sonnenfeld & Leauthaud \(2018\)](#) (hereafter **S18**) and [Sonnenfeld et al. \(2022\)](#) (hereafter **S22**), argued that the apparent positive correlation in earlier studies may arise from Eddington bias and bin-scattering effects.

There have been other studies that suggest a negative correlation between halo mass and galaxy size (e.g. [Desmond et al., 2019](#); [Taylor et al., 2020](#)). In particular, [Taylor et al. \(2020\)](#) (hereafter **T20**) performed **GGL** using the Kilo Degree Survey and the Galaxy And Mass Assembly survey and reported a negative correlation. Their galaxy sample was not limited by colour; therefore, the observed negative trend primarily reflects the dependence of galaxy size on colour.

1.6 Objective & Outline of Thesis

In this thesis, we use data from the **DLIS** ([Dey et al., 2019](#)) and the **UNIONS** ([Gwyn et al., 2025](#)) to obtain **GGL** measurements. Our primary sample is divided into three stellar-mass bins, and we further split each bin into two size bins to study size dependence. Unlike previous works, we use a **CSMF** **CSMF**-based **HOD**, which has been implemented into the **AbacusSummit** simulations ([Burger et al., 2025](#), hereafter **B25**). We also adopt a similar power-law prescription as described in equation 3 of [Charlton et al. \(2017\)](#). This is done in a two-step process. First, we fit the **CSMF** **HOD** model to the overall stellar mass-binned samples utilizing accurately trained neural-network emulators to obtain best-fit **HOD** parameters. The mock catalogue generated using the best-fit **HOD** parameters provides the most accurate representation of the **DLIS** lens galaxy sample. In the second step of this analysis, we then introduce and apply a model for the size dependence of halo mass and galaxy size on the mock catalogue generated by the best-fit **HOD** parameters.

This allows us to generate predictions and fit for the size-split observations made using the [DLIS](#) galaxy sample.

This thesis is organized into the following chapters. Chapter [2](#) details the data used in this work, notably the galaxy catalogues used to make the [GGL](#) measurements, along with any data cuts applied. Chapter [3](#) introduces the estimator used to calculate the [ESD](#) vectors from the observational data, whose results are also presented. Chapter [4](#) details the intricacies of the modelling method adopted. Chapter [5](#) presents the main results of this thesis, which we then discuss in the context of previous literature in Chapter [6](#). Finally, in Chapter [7](#) we summarize the results and present directions for future work.

The adopted cosmology for this thesis is a flat [\$\Lambda\$ CDM](#) cosmology, where the relevant parameters are the Hubble constant $H_0 = 70$ km/s/Mpc, the matter energy density $\Omega_m = 0.279$, the baryon density $\Omega_b = 0.046$, the spectral index $n_s = 0.97$, and the measure of the [RMS](#) matter fluctuations $\sigma_8 = 0.82$.

Chapter 2

Data

In this chapter, we provide a detailed description of the data products used in this thesis. We discuss the details of the source galaxy catalogue produced by [UNIONS](#), and the lens galaxy catalogue by [DLIS](#). Finally, we outline the sample selection criteria imposed to obtain the final sample of lens galaxies used for the weak lensing measurements. We begin with the [UNIONS](#) source galaxies.

2.1 UNIONS: Shapes of Source Galaxies

[UNIONS](#) is an imaging survey within the northern hemisphere, combining the data taken from three wide-field telescopes: the [Canada-France-Hawaii Telescope \(CFHT\)](#), the Panoramic Survey Telescope and Rapid Response System, and the Subaru Telescope. The main survey covers about $6,250 \text{ deg}^2$ in the *ugriz* bands. Details on the survey can be found in [Gwyn et al. \(2025\)](#).

The source galaxies, from which we require shape information, come from the [UNIONS ShapePipe v1.3](#) catalogue. This catalogue contains ~ 84 million galaxy shape measurements extracted from [Canada-France Imaging Survey \(CFIS\)](#) *r*-band data. The average seeing of the *r*-band observations is about 0.65 arcseconds, which places [UNIONS](#) at a comparable level with the [HSC](#) survey ([Mandelbaum et al., 2018](#)). Relative to the Dark Energy Survey Camera ([Zuntz et al., 2018](#)), the average seeing of [CFIS](#) is a significant improvement. In the remainder of this section, we present a summary of the shape measurements obtained using the ShapePipe pipeline using the [CFIS](#) *r*-band images. We refer readers to [Guinot, Axel et al. \(2022\)](#) and [Farrens et al. \(2022\)](#) for a comprehensive description of the methodology and [Gwyn et al. \(2025\)](#) for details of the imaging.

ShapePipe is an analysis pipeline written in `python` for weak-lensing processing. Using **SExtractor** (Bertin & Arnouts, 1996), all sources (including point sources) are extracted from stacked **CFIS** r -band images. Sources are identified and extracted from pixels whose values are greater than or equal to a 1.5σ detection threshold. Additionally, further cuts are made to account for the error in the spread model, object magnitude, the ratio between galaxy size and the size of the point spread function and the removal of low S/N objects. Galaxy shape measurements are obtained using **NGMIX** (Sheldon, 2015), which was also applied and tested on Dark Energy Survey data. Uniquely for **ShapePipe** v1.3, the multi-charged-coupled device (Liaudat et al., 2021) method is used to estimate the PSF as opposed to **PSFEx** (Bertin, 2011). Additionally, the ratio of galaxy size (T_{gal}) to PSF size (T_{PSF}) is required to be $0.5 < T_{\text{gal}}/T_{\text{PSF}} < 3$. Finally, shape measurements are calibrated using the metacalibration method by Huff & Mandelbaum (2017).

Currently, **UNIONS** does not have photo- z measurements for each source galaxy; however, an effective redshift distribution, $n(z_s)$, was estimated from colour space using the method from Wright et al. (2020). We summarize the implementation of this method to **UNIONS** data but refer the reader to Appendix A of Li et al. (2024).

First, multi-band photometry must be assigned to the **UNIONS** galaxies. Using the **UNIONS** r -band observations, the galaxies are cross-matched with **CFHTLenS** W3 (Hildebrandt et al., 2012) field galaxies to assign $ugriz$ magnitudes. The redshift distribution is then calibrated with a spectroscopic sample that originates from the DEEP2 Galaxy Redshift Survey (Newman et al., 2013), the VIMOS Very Large Telescope Deep Survey (Le Fèvre et al., 2005), and the VIMOS Public Extragalactic Redshift Survey (Scodeggio et al., 2018). **Self-Organizing Maps (SOMs)**, a type of unsupervised machine learning algorithm, are trained using the multiband photometry of the spectroscopic sample to reduce the higher-dimensional data into a 2D grid. Each **SOM** cell has a weight, w_i^{SOM} , which is defined as the ratio of the number of **UNIONS** objects (which themselves are weighted by shape weights) to the number of spectroscopic objects. These weights are then multiplied by the spectroscopic redshift distribution to estimate the redshift distribution of the sources. A blinded effective redshift distribution is shown in Figure 2.1, which is close, but not identical, to the redshift distribution used for weak lensing measurements in this thesis.

2.2 DLIS: Lens Galaxy Catalogue

The lens galaxies in our study are from the photometric redshift (photo- z) and stellar mass catalogue generated by Zou et al. (2019, 2022), for which they incorporate data up to Data

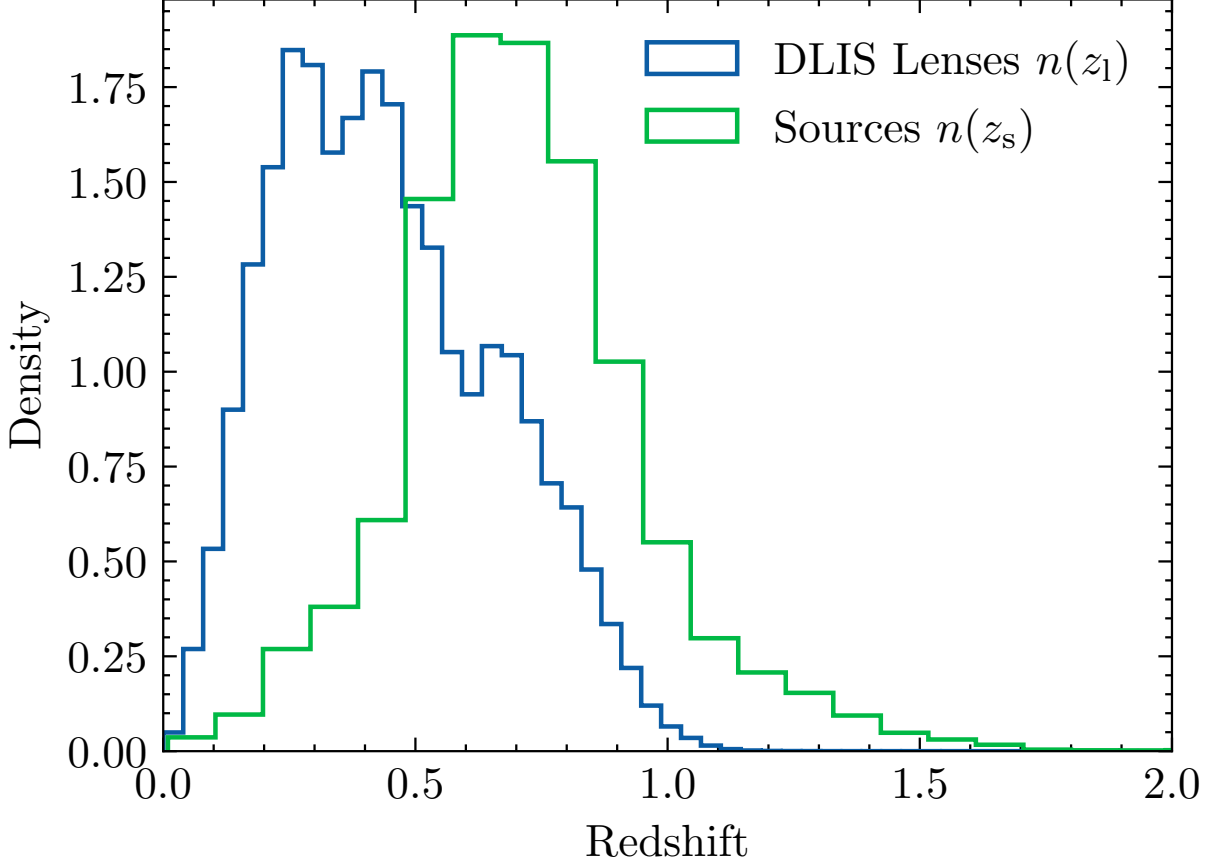


Figure 2.1: The green histogram is one of the blinded photometric redshift distributions of [UNIONS ShapePipe v1.3](#) sources. This redshift distribution is utilized only for plotting in order to preserve the blinding of the cosmological analysis. The unblinded redshift distribution, which is used for the analysis in this paper, is close, but not identical, to this blinded distribution. The blue histogram is the photometric redshift distribution of the red, Sérsic cut, declination cut, $M_{*,\text{limit}}$ applied lens galaxy sample from [DLIS](#). The details of these cuts are described in Section [2.3](#).

Release 9 of DLIS. DLIS is comprised of three independent optical surveys: the Beijing-Arizona Sky Survey (Zou et al., 2017), the Mayall z-band Legacy Survey (Silva et al., 2016), and the DECam Legacy Survey (Blum et al., 2016). Together, these imaging surveys (in optical and near-infrared) were used for target selection for the Dark Energy Spectroscopic Instrument (DESI) spectroscopic survey. The sky coverage is bounded by $-18^\circ < \delta < +84^\circ$ and $|b| > 18^\circ$ in celestial and Galactic coordinates, respectively. More details on the DLIS survey can be found in Dey et al. (2019). A total of 320,060,206 galaxies are contained within this catalogue with r -band magnitude $r < 23$. The stellar mass range for the catalogue was determined in Zou et al. (2019) to be $8.4 < \log_{10}(M_*) < 11.9$. The completeness limit is

$$\log_{10}(M_{\text{lim}}^*/M_\odot) = 1.4168 \log_{10}(2.8322 z_{\text{phot}}) + 10.5523, \quad (2.1)$$

where z_{phot} is the photometric redshift. The photo- z algorithm employed by Zou et al. (2022) is similar to the local linear regression method in Beck et al. (2016). The LEPHARE software (Arnouts et al., 1999; Ilbert et al., 2006) was used to estimate the stellar masses. The photo- z accuracy is quoted to be 0.017, with a redshift coverage of $z_{\text{phot}} < 1$. The logarithmic ($\log_{10}$) stellar mass uncertainty is ~ 0.2 dex.

2.3 Lens Galaxy Sample Selection

In this thesis, we investigate the dependence of halo mass on galaxy size (at fixed stellar mass) for a sample of red, early-type galaxies. Analysis of red, elliptical-type galaxies has been done in the literature (see Section 1.5); thus, our results would be easily comparable. In this section, the data cuts concerning colour, Sérsic index, and stellar mass, which have been applied to the overall DLIS lens galaxy catalogue, will be discussed.

2.3.1 Colour Cut

To split the DLIS catalogue of galaxies into a *blue* and *red* sample, we can do this trivially by identifying two distinct bimodal peaks in a histogram of galaxy colour. Galaxy colours can be analyzed by taking the difference in observed galaxy magnitudes in two different bands, which are ideally as far apart in terms of effective wavelength. Strateva et al. (2001) can be considered to be the first work to robustly demonstrate the ability to split a large photometric sample of galaxies into *red* and *blue* using colour separation. In their work, Strateva et al. (2001) show evidence for strong bimodality of the distribution of 147,920

galaxies from [Sloan Digital Sky Survey \(SDSS\)](#) in the $g^* - r^*$ versus $u^* - g^*$ colour-colour space.

Table 3 in [Zou et al. \(2019\)](#) outlines all the various quantities included for each lens galaxy, for which we have available absolute magnitudes in the g -band and z -band. From the distribution of $g - z$ colours, we identify the green valley at $g - z = 1.0$. To minimize contamination between the red and blue samples, we conservatively make our red cut for galaxies at $g - z \geq 1.1$.

2.3.2 Sérsic Index Cut

Sérsic profiles can be used to classify galaxy types based on their light profiles. The Sérsic profile has the form of

$$I(R) = I_0 \exp(-kR^{1/n}), \quad (2.2)$$

where I_0 is the intensity at $R = 0$, and n is the Sérsic index which controls the curvature of the profile. A de Vaucouleurs model with Sérsic index $n = 4$ ([De Vaucouleurs, 1953](#)) is typical for an elliptical galaxy, whereas an exponential model, with $n = 1$, well-describes spiral galaxies. A clear cut at $n = 4$ would result in a sample of red, early-type galaxies that are predominantly elliptical; however, it has been shown that for a sample of red galaxies, the Sérsic index ranges from ~ 1 -6 ([Caon et al., 1993](#); [Blanton & Moustakas, 2009](#)). To achieve a sample of red galaxies maximizing the number of early-type types but also accounting for this variation in the Sérsic index, we apply a cut on the Sérsic index of $n \geq 2.5$.

2.3.3 Stellar Mass Bins

To utilize galaxies from [DLIS](#), we must first determine the stellar mass completeness. Stellar mass completeness refers to the lowest stellar mass (M_*) galaxy that can be detected, such that the majority of galaxies of that same mass, or above, can be considered reliably detected. Galaxies above this limit form a statistically complete sample. This limit is redshift dependent and is thus determined for specific redshift slices to account for observational depth. Stellar mass completeness, denoted by M_{lim}^* , varies for different surveys due to differing detection thresholds and instrumental capabilities. As a result, not all of the faintest of galaxies are detected and catalogued, including faint, low-mass galaxies and high-mass galaxies with low surface brightness. Failure to account for stellar mass incompleteness can lead to statistical biases. An example of this is the [Eddington \(1913\)](#)

bias, where more high-mass galaxies scatter in or out of stellar mass bins than lower-mass, fainter galaxies, skewing measured galaxy properties within those bins.

As stated in Zou et al. (2019), the stellar mass range of galaxies is $8.4 < \log_{10}(M_*/M_\odot) < 11.9$, where the completeness limit is parameterized in Eqn. 2.1. The LEPHARE software (Arnouts et al., 1999; Ilbert et al., 2006) was used to estimate the stellar masses, which outputs M_* values in units of $\log_{10}(M_*/M_\odot)$. Figure 2.2 shows the sample of red, Sérsic cut galaxies from the DLIS catalogue with the stellar mass completeness limit drawn as the solid black line. In addition to the two aforementioned data cuts, an additional cut on declination, $\text{DEC} \geq +20^\circ$, is made as there are no UNIONS sources below a declination of $+30^\circ$. We note that the cut, $\text{DEC} \geq +35^\circ$, is applied uniquely to $\mathcal{M}_3^*$ due to survey depth being much deeper, leading to a region of characteristically higher galaxy counts when compared to the other two stellar mass bins. This is shown in Figure 13 of Zou et al. (2019). Our three volume-limited stellar mass bins are $10.5 \leq \log_{10}(M_*/M_\odot) < 10.8$, $10.8 \leq \log_{10}(M_*/M_\odot) < 11.2$, and $11.2 \leq \log_{10}(M_*/M_\odot) < 11.78$. We define and use the labels $\mathcal{M}_1^*$, $\mathcal{M}_2^*$, and $\mathcal{M}_3^*$ to reference these three stellar mass bin ranges. These bins were determined by accounting for the scattering effects of galaxy photo- z and stellar masses to mitigate any Eddington bias. C17 found no redshift evolution for any potential correlation between galaxy size and halo mass; thus, we do not further subdivide our stellar mass bins into redshift bins in this analysis.

2.3.4 Splitting by Galaxy Size

The goal of this analysis is to investigate whether the weak lensing signals differ between galaxies of different physical sizes. To accomplish this, we choose to divide our galaxies into ‘small’ and ‘large’ size subsamples. Two size bins are considered in this thesis, as opposed to more, to maintain a high signal-to-noise ratio for our GGL measurements. Additionally, if a halo mass dependence on galaxy size exists, two size bins would be sufficient to capture the overall trend. We do recognize that multiple-size bins may reveal more subtle trends, but given the current inconclusive state of the literature, the current methodology would prove insightful, leaving extensions for future work. We adopt a straightforward approach by comparing each galaxy’s size to the median size corresponding to its stellar mass, thereby subdividing galaxies into size bins. The red, Sérsic cut, stellar mass complete DLIS catalogue of galaxies is binned by stellar mass for which the median size (in kpc) is computed. A linear interpolation function is subsequently constructed by first binning the stellar mass values into 50 linear bins over the range $\log_{10}(M_*/M_\odot) \in [10, 12]$, and then computing the median galaxy size in each bin (in units of kpc). This interpolation function is employed to classify the galaxies. Galaxies with sizes above the interpolated median are

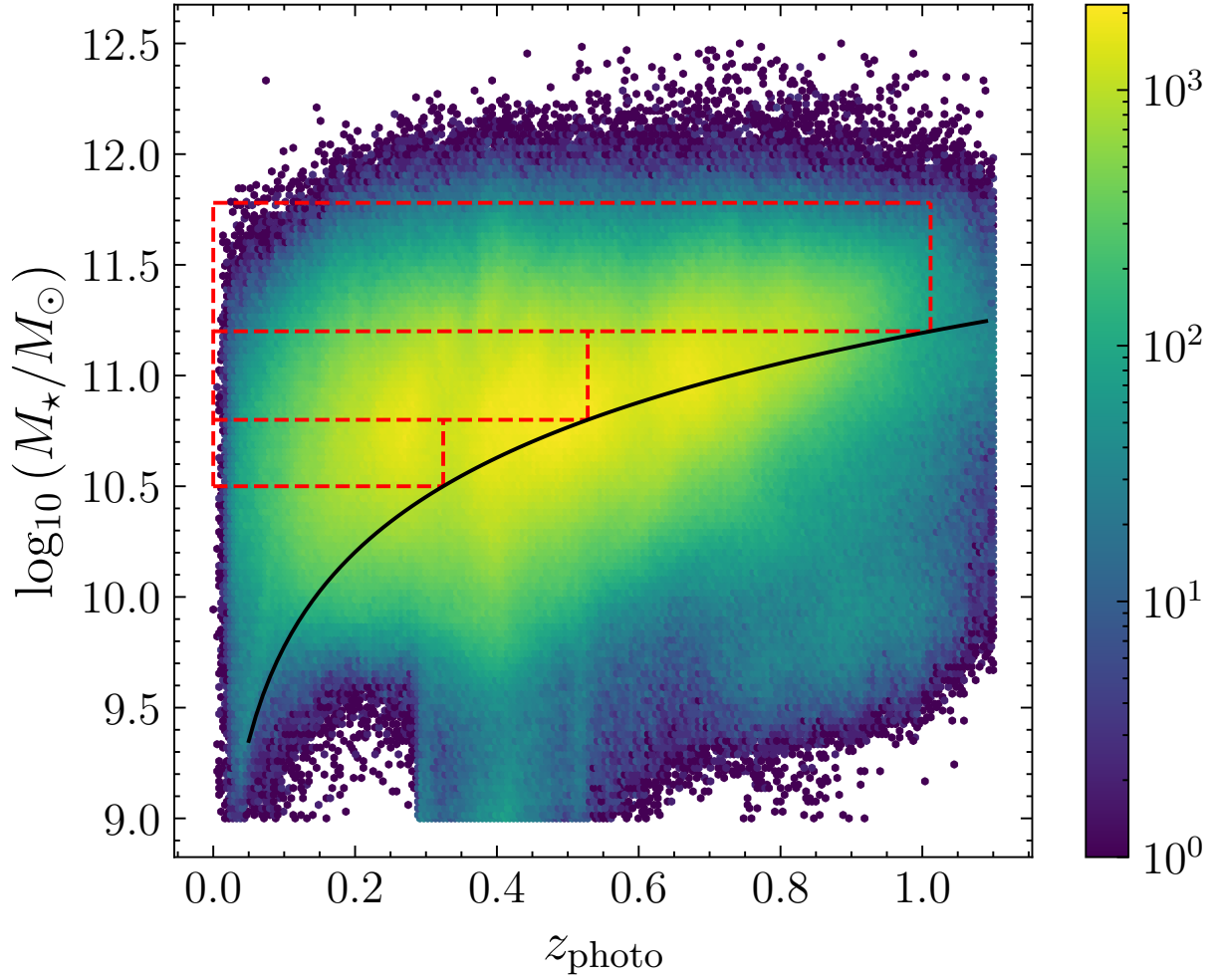


Figure 2.2: Stellar mass versus photo- z for a cut subsample of the original DLIS catalogue. The following cuts were applied: $g - z \geq 1.1$, Sérsic index (n) ≥ 2.5 , DEC $\geq 20^\circ$. The solid black line shows the stellar mass completeness limit determined in Zou et al. (2019) (see Eqn. 2.1). The galaxies to the left of the completeness line are considered in this analysis. The area enclosed by the red dashed boxes indicates the three volume-limited samples of galaxies used in this thesis.

classified as *large*, while those below are classified as *small*. Figure 2.3 plots the size (R_{eff}) of the DLIS galaxy catalogue after making the various data cuts described against stellar mass. Overplotted as a solid black line is the interpolated median size as a function of the stellar mass function.

Table 2.1 summarizes key quantities for each of our stellar mass bins and the further subdivided size bins for the lens galaxies used to make the weak lensing measurements. The tabulated quantities are the number of galaxies in each stellar mass bin and size subsample used for the weak lensing measurement denoted by (N_{gals}), the mean $g-z$ colour, ($\langle g-z \rangle$), mean stellar mass ($\log_{10}\langle M_{\star}/M_{\odot} \rangle$), mean photometric redshift ($\langle z_{\text{phot}} \rangle$), mean Sérsic index ($\langle n \rangle$), logarithm of the mean galaxy size ($\log_{10}\langle R/\text{kpc} \rangle$), and the logarithm of the mean size to median galaxy size ratio ($\log_{10}\langle \frac{R}{R_{\text{med}}(M_{\star})} \rangle$).

Figure 2.1 plots the processed DLIS galaxy sample photometric redshift distribution in blue. The green histogram is a blinded effective redshift distribution for the UNIONS ShapePipe v1.3 sources. This blinded distribution is close, but not identical to the redshift distribution used for weak lensing measurements in this thesis.

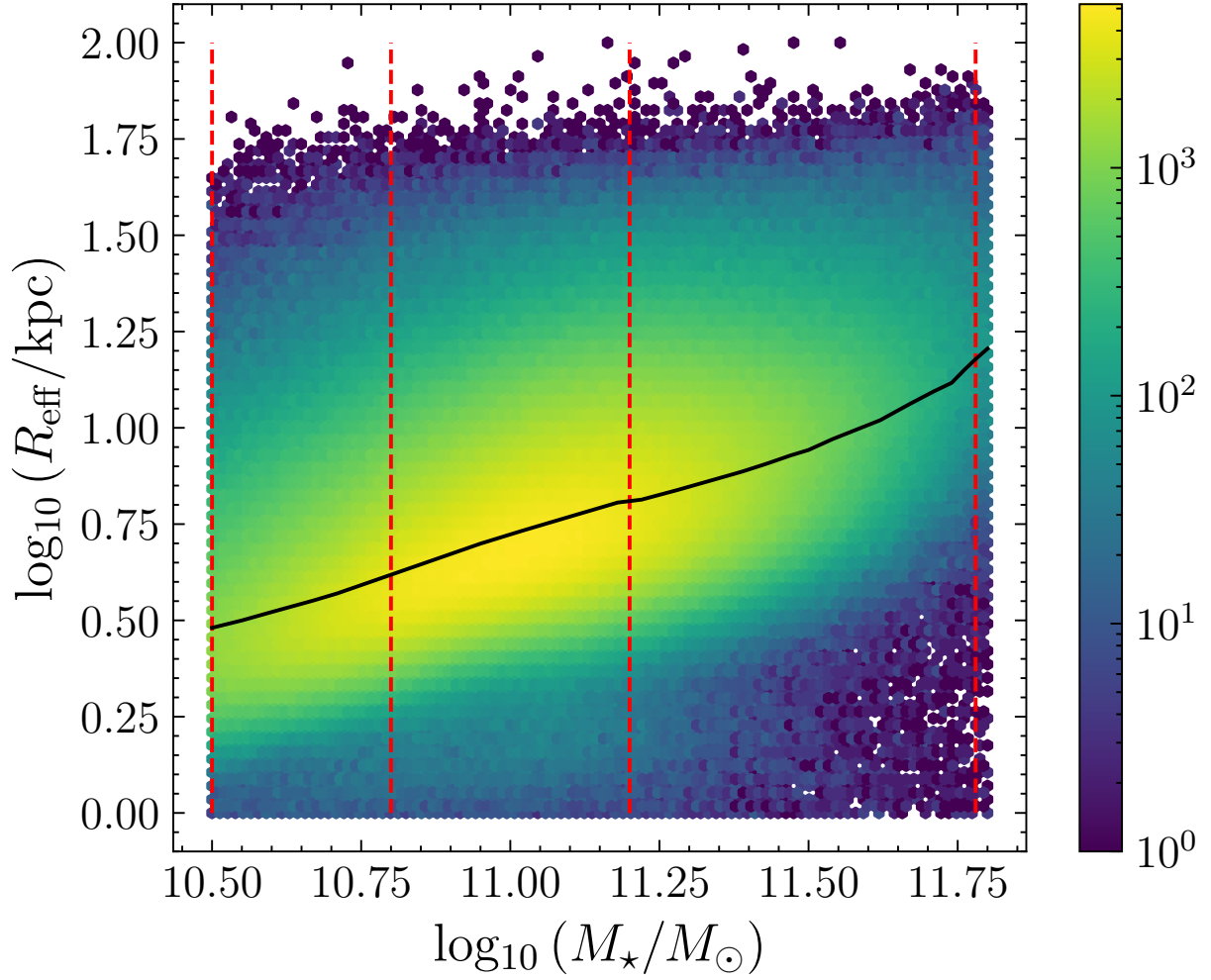


Figure 2.3: Plot of all red, Sérsic cut, stellar mass complete [DLIS](#) galaxies. From these galaxies, a median size linear interpolation function is generated and is plotted as the black solid line. Galaxies above the line are classified as *large*, and galaxies below are *small*. Vertical red dash lines indicate the three stellar mass bins considered in this work.

	N_{gals}	$\langle g - z \rangle$	$\log_{10} \langle M_*/M_{\odot} \rangle$	$\langle z_{\text{phot}} \rangle$	$\langle n \rangle$	$\log_{10} \langle R/\text{kpc} \rangle$	$\log_{10} \langle \frac{R}{R_{\text{med}}(M_*)} \rangle$
$10.5 \leq \log_{10} \left(\frac{M_*}{M_{\odot}} \right) < 10.8$	Small	282,931	1.36	10.66	0.22	4.10	0.43
	Large	270,496	1.35	10.66	0.21	4.29	0.78
$10.8 \leq \log_{10} \left(\frac{M_*}{M_{\odot}} \right) < 11.2$	Small	638,887	1.38	10.99	0.35	4.18	0.60
	Large	653,811	1.38	10.99	0.34	4.43	0.92
$11.2 \leq \log_{10} \left(\frac{M_*}{M_{\odot}} \right) < 11.78$	Small	322,919	1.45	11.38	0.63	4.15	0.75
	Large	328,054	1.43	11.38	0.54	4.51	1.10
							0.21

Table 2.1: Summary of the stellar mass bins and subdivided size bins used in this thesis along with the values of the number of lens galaxies per bin (N_{gals}) used for the weak lensing measurements, the mean $g - z$ colour, the mean stellar mass, the mean photometric redshift, the mean Sérsic index, the mean galaxy size, and the mean galaxy size compared to the median.

Chapter 3

$\Delta\Sigma$ Measurements

In Section 1.4, the basic formalisms and quantities integral to weak lensing measurements, in particular, measurements of [ESD](#) or $\Delta\Sigma$ were presented. Note that we will use [ESD](#) and $\Delta\Sigma$ interchangeably throughout this thesis. Recall that $\Delta\Sigma$ is proportional to the measured tangential shear, γ_t and the critical surface mass density, Σ_{crit} . For [GGL](#) measurements, the lensing signal from a singular lens-source pair is minuscule, and thus, averaging over a large sample of lens-source pairs is required to obtain a lensing signal with a high [SNR](#). This process of averaging, also known as *stacking*, is quite beneficial as galaxies have intrinsic shapes/ellipticities which are random. Thus, when averaging over a large enough sample of galaxies, it becomes true that $\langle e_t \rangle \approx \gamma_t$. This allows for the use of an estimator to calculate [ESD](#) profiles. We now discuss this estimator.

3.1 The $\Delta\Sigma$ Estimator

Presently, [UNIONS](#) does not have photometric redshifts available for individual source galaxies. Thus, to obtain inverse critical surface mass density values, $\Sigma_{\text{crit}}^{-1}$, we integrate over the entire effective redshift distribution, whose construction was described in Section 2.1. This integration is given by,

$$\langle \Sigma_{\text{crit}}^{-1}(z_l) \rangle = \int_{z_l}^{\infty} \tilde{\Sigma}_{\text{crit}}^{-1}(z_l, z_s) n(z_s) dz_s, \quad (3.1)$$

where $\tilde{\Sigma}_{\text{crit}}^{-1}$ is the standard definition for the critical surface mass density, but inverted,

$$\tilde{\Sigma}_{\text{crit}}^{-1}(z_l, z_s) = \frac{4\pi G}{c^2} \frac{D_A(z_l) D_A(z_l, z_s)}{D_A(z_s)}, \quad (3.2)$$

where $D_A(z_s)$ is the angular diameter distance to the source galaxy at redshift z_s , $D_A(z_l)$ is the angular diameter distance to the lens galaxy at redshift z_l , and $D_A(z_l, z_s)$ is the distance between the source and lens galaxy at their respective redshifts.

To compute $\Delta\Sigma$, we utilize the `dsigma python` package by [Lange & Huang \(2022\)](#). Since galaxies are not perfectly circular, they have an intrinsic ellipticity, or shape, which, on average, is assumed to be random. The measured (observed) tangential ellipticities (e_{obs}) of an individual galaxy are $e_{\text{obs}} = e_{\text{int}} + \gamma_t$, where $\langle e_{\text{int}} \rangle = 0$. Due to errors in shape measurements, in reality, we find that $e_{\text{obs}} = e_{\text{int}} + (1 + m)\gamma_t$, where m is the multiplicative shear bias. For large, well-observed galaxies, $m \sim 0$, but for smaller, faint galaxies, $m \neq 0$. Image simulations are used to estimate values of m , and thus, the measured observed ellipticities are scaled by $e_{\text{obs,corr}} = e_{\text{obs}} / \langle 1 + m \rangle$, resulting in $\langle e_{\text{obs,corr}} \rangle = \langle \gamma_t \rangle$.

As introduced above, shear bias is a systematic error in the measurement of galaxy shears. Results from a similar survey to [UNIONS](#), such as the Dark Energy Survey, find a mean multiplicative shear bias of $\sim 2\%$ using image simulations ([MacCrann et al., 2022](#)). Considering this at face value, there may be residual biases due to, e.g. blending of source galaxies, which needs to be calibrated via image simulations. This is ongoing work within the [UNIONS](#) collaboration. We are not aiming to constrain cosmological parameters, and any multiplicative shear bias would simply appear as a shift in the amplitude of our [ESD](#) measurements. The difference between the small and large galaxy [ESD](#) vectors would remain largely the same. Thus, multiplicative shear bias is a negligible effect for this work. Similarly, intrinsic alignments (IA) are not modelled in this work, as the analysis does not constrain cosmological parameters. The dependence of IA on galaxy properties such as morphology and size is an emerging topic of discussion. For example, early work by [Georgiou et al. \(2025\)](#) reports possible signs of a difference in intrinsic alignment amplitudes for red galaxies with Sérsic index $n < 4$ versus $n > 4$. Although the difference they find is not statistically significant, they provide direction for future work in which IA-galaxy property dependencies can be accounted for. It is at this stage, reasonable to assume that any IA effects are comparable across both red galaxy size subsamples.

Given that the use of an effective redshift distribution is required, `dsigma`, can calculate $\Delta\Sigma$ using the estimator,

$$\Delta\Sigma = \frac{\sum_{\text{ls}} w_s \langle \Sigma_{\text{crit}}^{-1}(z_l) \rangle e_t}{\sum_{\text{ls}} w_s \langle \Sigma_{\text{crit}}^{-1}(z_l) \rangle^2}, \quad (3.3)$$

where e_t is the measured tangential ellipticities of the source galaxies. The estimator presented above must be corrected to account for various effects (e.g., boost factor, random subtraction), which we describe next.

3.1.1 Random Subtraction

Systematic errors in the shape measurements of background source galaxies, as well as survey geometry and observational effects, can produce an additive bias in the inferred tangential shear γ_t . This will cause an amplification in the measured signal, which is unphysical. Random subtraction is a necessary method to remove such an excess signal. We construct a catalogue of random lens galaxies from [DLIS](#)¹, which has ~ 10 times more randoms than [DLIS](#) lens galaxies in our sample. Suppose the random lenses are truly random and not correlated to a large-scale matter field. In that case, the lensing amplitude should be 0. [Singh et al. \(2017\)](#) has shown that random subtraction also reduces the covariance of weak lensing measurements. The excess signal is subtracted from the measured signal by the following,

$$\Delta\Sigma(R) = \Delta\Sigma_{\text{raw}}(R) - \Delta\Sigma_{\text{randoms}}(R). \quad (3.4)$$

3.1.2 Boost Factor

Given the overlapping redshift distributions (see Figure 2.1) between the [UNIONS](#) sources and [DLIS](#) lens galaxies, there exist lens-source pairs in which $z_s \approx z_l$ or $z_s < z_l$. In both cases, the lens induces little to no shear on the source galaxy and would decrease the measured lensing amplitude. The boost factor correction, b , is defined as,

$$b = \frac{\sum_{\text{ls}} w_{\text{ls}}}{\sum_{\text{rs}} w_{\text{rs}}}, \quad (3.5)$$

where w_{rs} is the weight of each random-source pair. By comparing the weights of lens-source pairs to random-source pairs, we can detect any systematic overabundances of sources near lenses and not at random points. If $b \geq 1$, sources cluster around lenses strongly when compared to the random points. Additionally, lens galaxies are typically located in environments with other galaxies. Many of these clustered sources are not true background galaxies and are physically associated with the lenses. These are still included in the source catalogue, but since they are not lensed, their inclusion acts to dilute the measured lensing signal. The boost factor correction works to counteract this dilution. The correction is applied to the raw lensing measurement by $\Delta\Sigma = b\Delta\Sigma_{\text{raw}}$.

¹<https://www.legacysurvey.org/dr9/files/#random-catalogs-randoms>

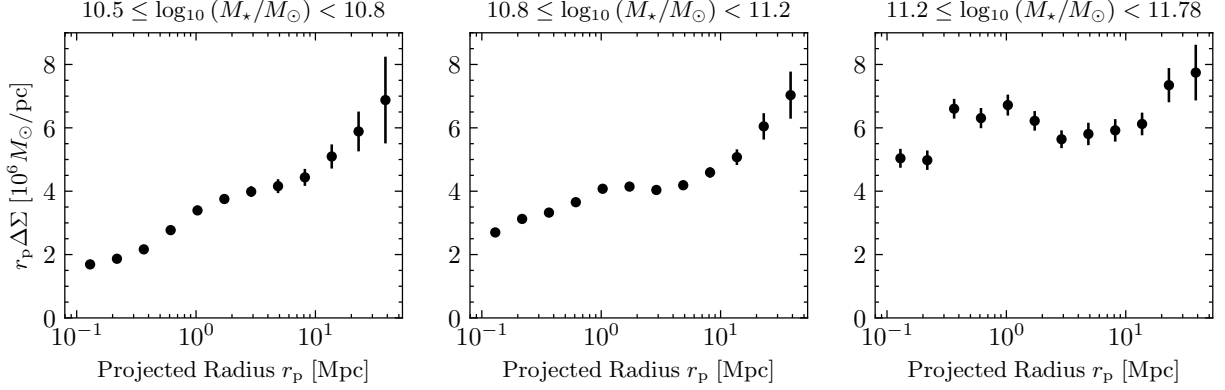


Figure 3.1: Plot of the full bin (i.e., not size split stellar mass bins) $r_p \Delta \Sigma$ profiles from the DLIS catalogue following the data cuts described in Section 2.3. Error bars for each stellar mass bin are obtained independently of each other via Jackknife resampling of the lens galaxies ($N_{\text{JK}} = 100$).

3.2 DLIS $\Delta \Sigma$ Results

Following the data cuts described above, we use the `dsigma python` package (Lange & Huang, 2022) to compute the stacked galaxy lensing profiles for each of the stellar mass bins, both with and without splitting the stellar mass bin further into small and large size subsamples. In addition to the estimator for $\Delta \Sigma$ in Eqn. 3.3, we also include a boost factor correction, and random subtraction. Covariance matrices are obtained using the jackknife resampling method, splitting the lens sample into $N_{\text{JK}} = 100$ patches. The covariance matrix is not shown as it is a diagonal matrix dominated by shape noise. We make measurements of $\Delta \Sigma$ from a projected radii of 0.1 Mpc to 50 Mpc. This choice of radii enables us to probe small scales where the 1-halo term dominates ($r_p \lesssim 0.5$ Mpc), intermediate scales where the offset-group term dominates due to contributions from a satellite’s host halo ($0.5 \text{ Mpc} \lesssim r_p \lesssim 1.5 \text{ Mpc}$) and large scales where the 2-halo term dominates ($r_p \gtrsim 1 \text{ Mpc}$). Plotted in Figures 3.1 and 3.2 are the resulting ESD vectors, with r_p values being the midpoints of our bins, which are evenly spaced in log-space. Figure 3.1 shows the ESD vectors for the full bin, which refers to the stellar mass bins that have not been further subdivided into size bins. Figure 3.2 shows the ESD vectors for our size split samples.

Comparing the ESD vectors in Figure 3.1 across all three stellar mass bins, we see the overall trend that the lensing signal increases with increasing stellar mass. This result is a natural consequence of the SHMR, in which as stellar mass increases, so does halo mass,

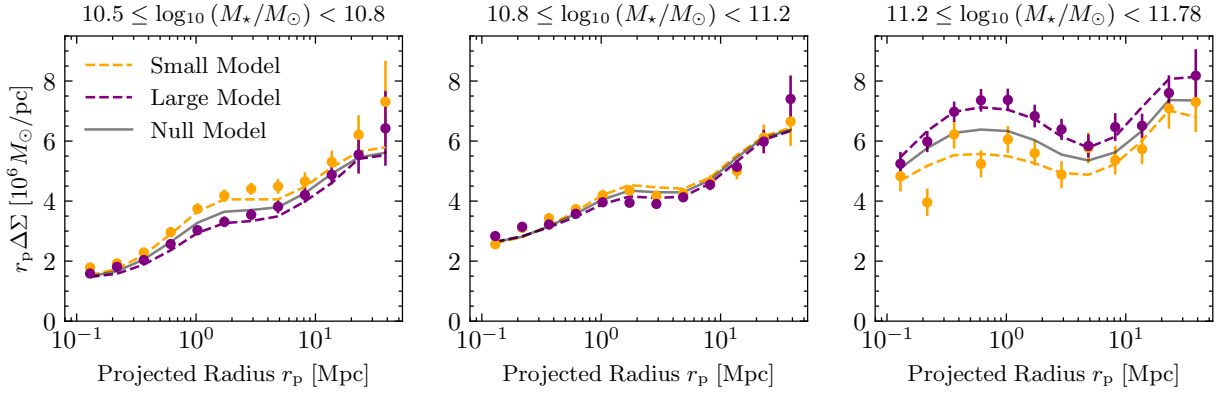


Figure 3.2: The $r_p \Delta \Sigma$ measurements for the size split subsamples for each stellar mass bin. The orange data vector is the lensing amplitude for the small galaxies whose measured size is less than the median size. The purple data vector is the lensing amplitude for the large galaxies whose measured size is greater than the median size. The measurement error bars are calculated via Jackknife resampling of the lens galaxies ($N_{\text{JK}} = 100$), and the model fitting error is propagated into the error bars shown. The error bars includes the error from the $\Delta \Sigma_{\text{mocks}}$ emulator and the $\Delta \Sigma_{\text{tNFW}} f_{\text{sat}}$ emulator. Details of the emulators can be found in Chapter 4. The orange, purple, and black lines are results from fitting our model, which is discussed in Chapter 5.

which would increase the γ_t measured for background source galaxies. This increase in γ_t is reflected as an increase in the lensing amplitude.

When we split each stellar mass bin into small and large based on physical size, we obtain the [ESD](#) vectors shown in Figure 3.2. We make note of the following trends. Firstly, in our lowest stellar mass bin, $\mathcal{M}_1^*$, we observe that at intermediate radii (0.5 Mpc to 3 Mpc), the [ESD](#) of the small galaxies exceeds that of the large galaxies. Outside of this range, however, the [ESD](#) profiles of the two samples are consistent within uncertainties. In the intermediate stellar mass bin, $\mathcal{M}_2^*$, the discrepancy is less pronounced, with the [ESD](#) profiles of small and large galaxies being consistent across nearly all scales. In contrast, in the highest stellar mass bin, $\mathcal{M}_3^*$, we observe the opposite trend: the [ESD](#) of large galaxies exceeds that of small galaxies at projected separations $r_p < 6$ Mpc. Without any knowledge of the satellite fractions of our sample, a possible interpretation of these trends might suggest a negative, null, and positive correlation between galaxy size and halo mass across the low, intermediate, and high stellar mass bins, respectively. Notably, the most significant differences between the [ESD](#) profiles occur at intermediate scales, where the signal is typically dominated by contributions from the offset-group term associated with satellite galaxies. This highlights the critical importance of accurately modelling satellite galaxy populations in our halo model.

Chapter 4

Methods & Modelling

In this chapter, we describe in detail the various aspects that make up the galaxy-halo connection model used in this work. We utilize the **AbacusSummit** simulations to generate mock lens galaxy catalogues, which are populated using a **CSMF**-based **HOD** model. Galaxy sizes are assigned to simulated galaxies by assuming a simple power-law relationship, $R_{\text{eff}} \propto M_{\text{h}}(M_{\star})^{\beta}$. Given these simulated galaxy catalogues, we compute mock **ESD** measurements, to which we analytically added a satellite galaxy subhalo contribution assuming a **Truncated Navarro-Frenk-White** (tNFW) profile. The mock **ESD** measurements are used to train the neural network emulators for each stellar mass bin. These emulators are later used to fit **HOD** parameters to the observational **DLIS** measurements. We use this simulation-based approach to model **ESD** profiles, rather than an analytical approach involving the integration of the galaxy-matter power spectra, for several reasons. First, this framework is flexible and can be readily adapted to include additional dependencies like assembly bias. Second, a simulation-based method can more effectively incorporate realistic observational scatter in redshift and stellar masses, which is not necessarily Gaussian in nature. The scattering of redshifts and stellar masses propagates throughout the halo model, and a forward-modelling approach like the one used here can capture this complexity. Finally, the implemented dependence of the halo mass and galaxy size seemed more intuitive through this simulation/forward-model approach. A key disadvantage of this approach is the stochasticity and the need for the use of emulators. These points will be expanded on within later sections of this chapter.

4.1 AbacusSummit Simulations

We leverage the high-accuracy, large, cosmological N-body simulation **AbacusSummit** (Garrison et al., 2021; Maksimova et al., 2021) to generate mock galaxy catalogues populated using a **CSMF**-based **HOD**. These mock catalogues have been well tested in previous works (e.g. Hadzhiyska et al., 2022b; Burger et al., 2024, 2025), and thus are used to calibrate emulators, which are later described and used to fit **HOD** parameters to our **DLIS** $\Delta\Sigma$ measurements. The **AbacusSummit** simulations follow the evolution of 6912^3 dark matter particles contained within a periodic box of $2\text{ Gpc}/h$. The mass resolution is $2.1 \times 10^9 M_\odot/h$. Dark matter halos are identified using the *CompaS0* algorithm (Hadzhiyska et al., 2022a), a hybrid of the **Friends-of-Friends (FoF)** and Spherical Overdensity algorithms. In this work, we fix the cosmology of the **AbacusSummit** mocks to that of the fiducial values highlighted in Section 1.6¹. The shear information is taken from a single redshift slice of $z_s = 0.65$ to measure the **ESD** using both the mock lens and source catalogues. The redshift was chosen based on the estimated median redshift of the calibration sources used to calibrate the **UNIONS** photometric redshift data.

4.2 Conditional Stellar Mass Function Halo Occupation Distribution Model

To populate lens galaxies, we make use of the halo catalogues from **AbacusSummit** and **CSMF HOD**. A **CSMF HOD** allows one to predict the number of central and satellite galaxies given a halo of mass M_h and assign a stellar mass $M_\star$ to each of the populated galaxies. Halo mass values for **AbacusSummit** are most similar to M_{200m} , which is the mass enclosed within a radius where the average density is 200 times the mean matter density of the universe. Thus, unless explicitly stated otherwise, all halo mass values discussed from this point forward should be understood as M_{200m} . The framework for this **CSMF HOD** was laid out in Burger et al. (2025, hereafter B25). The **CSMF** prescription was motivated by the conditional luminosity function described by Cacciato et al. (2009b). We now describe the galaxy-halo connection model.

The **CSMF** is the combination of both a central (c) and satellite (s) galaxy contribution,

$$\Phi(M_\star|M_h) = \Phi_c(M_\star|M_h) + \Phi_s(M_\star|M_h). \quad (4.1)$$

¹Further details on the **AbacusSummit** cosmologies can be found [here](#)

The central galaxy contribution is modelled using a log-normal distribution,

$$\Phi_c(M_\star|M_h) = \frac{1}{\sqrt{2\pi} \ln(10) \sigma_c M_\star} \exp \left[-\frac{\log_{10}(M_\star/M_c^*)^2}{2\sigma_c^2} \right] , \quad (4.2)$$

where σ_c controls the scatter between the stellar mass and halo mass of central galaxies. The satellites are modelled using a modified [Schechter \(1976\)](#) function,

$$\Phi_s(M_\star|M_h) = \frac{\phi_s}{M_s^*} \left(\frac{M_\star}{M_s^*} \right)^\alpha \exp \left[-\left(\frac{M_\star}{M_s^*} \right)^2 \right] . \quad (4.3)$$

The parameter α , which governs the power-law behaviour of satellites, is typically a function of halo mass M_h given by,

$$\alpha(M_h) = -2.0 + a_1(1 - \arctan[a_2 \log_{10}(M_\star/M_2)]) . \quad (4.4)$$

Following [B25](#) and [Dvornik et al. \(2023\)](#), we set $a_2 = 0$, thereby also removing the parameter M_2 , which governs the halo mass dependence of the faint-end slope of the satellite stellar mass function. We clarify that M_h refers to the mass of the host halo, not the subhalo of satellites. The remaining parameters, $M_c^*(M_h)$, $M_s^*(M_h)$ and $\phi_s(M_h)$ are parameterized as:

$$M_c^*(M_h) = M_0 \frac{(M_h/M_1)^{\gamma_1}}{(1 + M_h/M_1)^{\gamma_1 - \gamma_2}} , \quad (4.5)$$

$$M_s^*(M_h) = 0.56 M_c^*(M_h) , \quad (4.6)$$

and

$$\log_{10}[\phi_s(M_h)] = b_0 + b_1 \log_{10}(M_{12}) + b_2 [\log_{10}(M_{12})]^2 . \quad (4.7)$$

The normalization for the central [SHMR](#) is M_0 , the characteristic mass of the [SHMR](#) is M_1 , the low-mass slope of the [SHMR](#) is controlled by γ_1 , while γ_2 controls the high-mass slope. M_{12} is defined as $M_h/(10^{12} h^{-1} M_\odot)$. In this work, we set $b_2 = 0$ and $b_1 = 0$.

The occupation number for centrals and satellites is calculated between stellar mass values of $M_{\star,1}$ and $M_{\star,2}$ and is given by,

$$\langle N_X \rangle(M_h) = \int_{M_{\star,1}}^{M_{\star,2}} \Phi_X(M_\star|M_h) dM_\star . \quad (4.8)$$

Assembly bias is often considered within [HOD](#) models to account for correlations between galaxy observables and secondary halo properties, notably, environment-based biases

and variations in halo assembly histories. We adopt the simplest model for assembly bias, with parameters B_{cen} and B_{sat} , which modify the occupation of central and satellite galaxies depending on the local environment of the host halo. As described in [Burger et al. \(2024, 2025\)](#), modifying M_1 and a_1 emulates the effect of assembly bias. B_{cen} and B_{sat} are defined as

$$\log_{10} M_1^{\text{eff}} = \log_{10} M_1 + B_{\text{cen}}(\delta^{\text{rank}} - 0.5), \quad (4.9)$$

$$a_1^{\text{eff}} = a_1 + B_{\text{sat}}(\delta^{\text{rank}} - 0.5), \quad (4.10)$$

where $\delta^{\text{rank}} \in [0, 1]$ is assigned based on the local environment of a halo compared to the median. The local environment is defined by smoothing the matter density around the halo centres using a top-hat filter of radius $R_s = 3 h^{-1} \text{Mpc}$. This smoothing is done after removing the halo itself from the measurement. If the halo resides in an environment lighter than the median, then $\delta^{\text{rank}} < 0.5$. We deviate from [B25](#) by setting $B_{\text{sat}} = 0$ and $s = 0$, opting for the simplest [CSMF HOD](#) model possible. The parameter, s , modulates the distance from the host halo centre at which satellites are placed (see [Appendix C](#)). This leaves us with the following free parameters $\{M_0, M_1, \gamma_1, \gamma_2, \sigma_c, a_1, b_0, B_{\text{cen}}\}$.

4.2.1 Subhalo Contribution

The `AbacusSummit` simulations are a dark-matter-only (DMO) simulation whose halo catalogues do not identify and include satellite subhalos. Satellites are placed within central host halos based on the available dark matter particles. The probability for a particle to host a satellite galaxy is the same throughout, defined by

$$p_i = \frac{\langle N_s \rangle (M_h)}{N_p}, \quad (4.11)$$

where N_p is the number of particles in the host halo. As each particle has the same probability to “host” a satellite, satellites are effectively placed randomly within the host halo. We explore later in [Chapter 7](#) the effect of altering this prescription of satellite placement in the form of the s parameter. Given the absence of subhalos, to accurately model the [ESD](#) on small scales ($r_p < 0.3 \text{Mpc}$), we analytically add the contribution of the subhalo to the `AbacusSummit` [ESD](#) measurements using the `python` package `profiley`². We choose to model the subhalos using a [tNFW](#) profile. We know that as satellites fall into host halos, stripping processes affect the density profiles from outwards to inwards,

²<https://profiley.readthedocs.io/en/latest/index.html#/>

preferentially stripping the dark matter first and then the stellar content (e.g. [Smith et al., 2016](#); [Lee et al., 2018](#)), hence a [tNFW](#) can capture this effect. The [tNFW](#) density profile is modelled by

$$\rho(r) = \frac{\delta_c \rho_{\text{bg}}}{x(1+x)^2} \left(\frac{\tau^2}{\tau^2 + x^2} \right), \quad (4.12)$$

where $\tau = r_t/r_s$. The amount of truncation imposed onto the density profile is quantified by the parameter τ . Additionally, r_t is the truncation radius and r_s is the scale radius.

We make the following assumptions in our implementation of this [tNFW](#) profile. Firstly, we consider satellites as if they were in the field, and thus, they are central galaxies in their own halo, which later becomes the subhalo after infall into a denser environment, such as a group or cluster. With that assumption, we take all the central galaxies for a given stellar mass bin within our mock lens catalogue and create a linear interpolation function of the scattered stellar masses (Section 4.3 discusses the scattering performed) to median halo mass. We compute the median M_h for 50 linear bins, for a range of stellar mass values from $\log_{10}(M_*/M_\odot) \in [10, 12.5]$. With this interpolation function, we give as input the satellite galaxy stellar masses to obtain the halo mass of these galaxies prior to becoming a satellite. We then take the mean of these halo mass values as the average halo mass for the satellites for the given stellar mass bin. We use a cosmology-dependent concentration relation from the `python` package `CAMELS` ([Shao et al., 2023](#)) and the `python` package `Colossus` to make halo concentration conversions from c_{vir} to $c_{200\text{m}}$. In order to be consistent with `AbacusSummit`, the [tNFW](#) profiles are generated assuming M_h is defined as $M_{200\text{m}}$. The subhalo contribution is added to the [ESD](#) measurements, accounting for the satellite fraction for the given stellar mass bin

$$\Delta\Sigma_{\text{mock}} = \Delta\Sigma_{\text{mock,raw}} + (f_{\text{sat}}\Delta\Sigma_{\text{tNFW}}), \quad (4.13)$$

where f_{sat} is the satellite fraction of the stellar mass bin. In addition to the eight free parameters, τ , rather, $\log_{10}\tau$ is also a free parameter.

4.2.2 Implementation of Halo Mass-Galaxy Size Dependence

With the simulation-based [CSMF HOD](#) model now described, we turn our attention to implementing a halo mass-galaxy size dependence. We implement the halo mass-size correlation as an additional parameter, outside the [HOD](#) model described previously. This choice is done to isolate any correlations between halo mass and galaxy size without the complexity of [HOD](#) parameters, which may certainly introduce added degeneracies and blur the physical interpretation. [C17](#) assumed a simple power-law relationship between halo mass

and galaxy size, $M_h(M_*) \propto R_{\text{eff}}^\eta(M_*)$, where η quantifies the correlation strength. We implement a similar formalism to relate the halo mass of our mock galaxies to size so that we can split them into a small and a large sample. We define the following logarithmic ratio:

$$\Delta R' = \log_{10} \left(\frac{R_{\text{eff}}}{R_{\text{eff}}^{\text{med}}(M_*)} \right), \quad (4.14)$$

where $R_{\text{eff}}^{\text{med}}(M_*)$ is the same linear interpolation function for the median size of a galaxy at a given stellar mass M_* , which was derived from the [DLIS](#) catalogue as described in Section 2.3.4. Additionally, we define

$$\Delta M'_h = \log_{10} \left(\frac{M_h}{M_h^{\text{med}}(M_*)} \right), \quad (4.15)$$

where the halo mass of an individual mock galaxy is M_h . The parameter $M_h^{\text{med}}(M_*)$ is the median halo mass at a given stellar mass derived from the mock galaxy catalogue. The implemented halo mass-size relation is then

$$\Delta R' = \beta \Delta M'_h + \mathcal{N}(0, \sigma_{\text{extra}}). \quad (4.16)$$

The free parameter, β , controls the correlation between halo mass and size, which we will vary separately for centrals (β_{centrs}) and satellites (β_{sats}). In addition, we add Gaussian scatter with a standard deviation of σ_{extra} , which is defined as a function of β ,

$$\sigma_{\text{extra}} = \sqrt{(\sigma_{\Delta R'})^2 + \beta^2 (\sigma_{\Delta M'_h})^2}, \quad (4.17)$$

and captures the scatter of galaxy size measurements from [DLIS](#) around the mean ($\Delta R'$) and the scatter of halo mass within the [AbacusSummit](#) mock catalogues ($\Delta M'_h$). Galaxies are then split into small and large samples according to their $\Delta R'$ values. If $\Delta R' > 0$, then the galaxy is *large*, and if $\Delta R' < 0$, then the galaxy is *small*. Some literature measures values for η , given the proportionality of $M_h(M_*) \propto r_{\text{eff}}^\eta(M_*)$. We will later show, in Section 5.2, how we can easily convert our values for β into η .

4.3 Building the Mocks and the Emulators

In this section, we describe the process by which the mock catalogues were generated, the construction and training of our emulators, and the rationale behind their necessity. The goal of using the [CSMF HOD](#) is to determine a set of [HOD](#) parameters for which the

generated mock fits the [ESD](#) measurements of all three stellar mass bins of the [DLIS](#) sample (see Fig. 3.1). In practice, one could use a χ^2 minimization algorithm or a Markov chain Monte Carlo (MCMC) approach. At each step, a mock galaxy catalogue is generated from [AbacusSummit](#), the excess surface density ([ESD](#)) for the three stellar mass bins is calculated, and the results are fit to the [DLIS ESD](#) vectors to compute a χ^2 value or likelihood. While this process can achieve our objective, it is inefficient and excessively demanding in terms of resources and time. We use emulators in this work because this method is 10^3 times faster than generating mocks on the fly. Emulators are a type of non-linear interpolation model based on neural networks. Given a set of inputs ([HOD](#) parameters), emulators are trained based on the comparison of their outputs with training features ([ESD](#) measurements). Once trained, these emulators are capable of predicting [ESD](#) profiles given a new set of [HOD](#) parameters.

4.3.1 Generating the Mocks

We use the [CSMF HOD](#) described in Section 4.2 to populate the halo light cones from [AbacusSummit](#). Within these halo light cones are the 3D positions of the halos, their halo masses, and the positions of the dark matter particles contained within them. We iterate over each halo and compute the expectation value for hosting a central galaxy and satellite galaxies using Eqn. 4.8. Given these numbers, we draw from a Bernoulli distribution for centrals and a Poisson distribution for satellites. Central galaxies are placed exactly at the centre of the halos. Satellite galaxies are placed randomly within the host halo according to the positions of dark matter particles. Stellar mass values for the populated galaxies are drawn from the [CSMF](#) described in Eqn. 4.2 and Eqn. 4.3.

4.3.2 Model ESD

The [ESD](#) from the mock catalogues are obtained using [treecorr](#)³ ([Jarvis et al., 2004](#)). To capture the scattering of photo- z and stellar mass values for galaxies within [DLIS](#), we randomly scatter the redshifts and stellar masses of our mock galaxies. The normalized redshift error, $\Delta z_{\text{norm}} = (z_{\text{phot}} - z_{\text{spec}}) / (1 + z_{\text{spec}})$, has a reported standard deviation of 0.017 ([Zou et al., 2019](#)). The redshifts are scattered by

$$z_{\text{scattered}} = \mathcal{N}(0, \sigma_{z, \text{DLIS}}) (1 + z_{\text{mock}}) + z_{\text{mock}}. \quad (4.18)$$

³<https://github.com/rmjarvis/TreeCorr>

Galaxy stellar masses are scattered using

$$\log_{10} M_{\star, \text{scattered}} = \log_{10} \left(10^{M_{\star, \text{mock}}} \frac{D_L(z_{\text{scattered}})^2}{D_L(z_{\text{mock}})^2} \right), \quad (4.19)$$

where $D_L(z_{\text{scattered}})$ is the luminosity distance at the scattered redshift and $D_L(z_{\text{mock}})$ is the luminosity distance at the original mock redshift. The scattered stellar masses are used to split our sample into our three stellar mass bins. Additionally, the source catalogue contained the shear information taken from a single redshift of $z_s = 0.65$. For [GGL](#) measurements from [AbacusSummit](#), the source catalogue is downsampled to 10% randomly. The effect of this downsampling is negligible and is done to speed up the computation of $\Delta\Sigma$.

4.3.3 Emulator Training & Accuracy

To train the emulators, we generate training features, which are the measured [ESD](#) vectors from mock catalogues. Each mock catalogue is generated from a unique set of [HOD](#) parameters. We define the parameter space of the emulators in an iterative process, which we describe in detail in [Appendix A.1](#). The goal is for our emulators to take as inputs sets of [HOD](#) parameters and learn how each [HOD](#) parameter impacts the resulting [ESD](#) vector. To ensure our emulators are accurate, we apply additional ‘cleaning’ to the [ESD](#) measurements by removing any mocks with negative $\Delta\Sigma_{\text{mocks}}$ values and any noisy data vectors resulting from a low number of galaxies. In total, we have available 1998, 1998, and 1977 [ESD](#) mocks, respectively, for each stellar mass bin to train our emulators. [Figure 4.1](#) shows the mocks available for training in comparison to the [DLIS](#) $\Delta\Sigma$ measurements. We see that there is an appreciable amount of mocks scattered both above and below the [DLIS](#) measurements. This indicates that the parameter space determined through this iterative process was adequate, such that a joint fit of all three stellar mass bins should result in a set of best-fit [HOD](#) parameters. The effect of adequate sampling of this parameter space can be seen in the accuracy of the emulator (see [Figure 4.2](#)). Emulators trained on a poorly sampled space will struggle to make accurate predictions.

We randomly select 200 mocks from a sample of 500 lowest χ^2 mocks to be our test sample. The 500 mocks are, in essence, the 500 best-fitting mocks to our [DLIS](#) measurements. These 200 mocks are used to gauge the accuracy of the emulator and generate a covariance/correlation matrix for the emulator. The emulator covariance matrix is then added to the existing [DLIS](#) measurement covariance matrix for the given stellar mass bin. [Figure 4.2](#) shows the accuracy of our emulators with respect to the error on the [DLIS](#) $\Delta\Sigma$

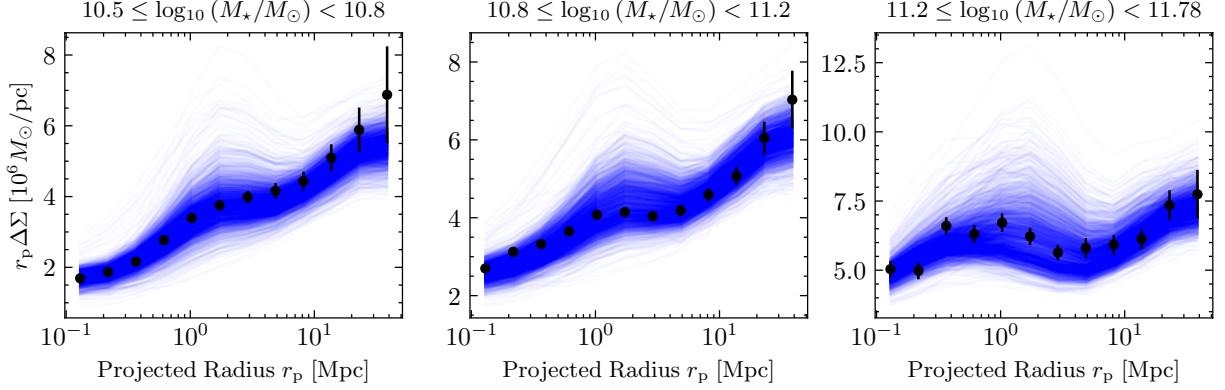


Figure 4.1: Plot of $r_p \Delta \Sigma$ comparing the available mocks for training our emulators against the [DLIS](#) $\Delta \Sigma$ measurements. Plotted in blue are the mock [ESD](#) measurements from [AbacusSummit](#) generated using the iterative process described in [Appendix A.1](#), and in black is the [DLIS](#) $r_p \Delta \Sigma$ measurement.

measurements obtained from Jackknife resampling of the lens galaxies. $\Delta \Sigma_{\text{emu}}$ is the emulator prediction of the [ESD](#) for a set of [HOD](#) parameters, $\Delta \Sigma_{\text{true}}$ is the true [ESD](#) vector, and σ_{DLIS} is the [DLIS](#) measurement error. We find our emulators are accurate to within less than 10% of the 1σ measurement error. The correlation matrices of the emulators can be found in [Appendix A.3](#). Finally, we observed that when provided with identical training and testing datasets but different initial weights and biases for the latent parameters of the neural network, multiple emulators trained for the same stellar mass bin yield slightly different predictions for a given set of test parameters. This variation arises from noise in the training data. To mitigate the impact of such fluctuations, we find that taking an ensemble average over five independently trained emulators per bin is both necessary and sufficient to ensure stable and reliable predictions. [Figure A.2](#) in [Appendix A.2](#) illustrates the behaviour of our emulators and the effect that each free parameter has on the predicted $\Delta \Sigma_{\text{mocks}}$ vectors.

4.3.4 $\Delta \Sigma_{\text{tNFW}} f_{\text{sat}}$ Emulators

The repeated calculation of the [tNFW](#) profile and its [ESD](#) profile was found to be slow, in particular due to several intermediate calculations needed to convert c_{vir} to $c_{200\text{m}}$ using [Colossus](#) ([Diemer, 2018](#)). Thus, a separate emulator, trained on the data vector of $\Delta \Sigma_{\text{tNFW}} f_{\text{sat}}$, was created. The parameter space for this emulator has no restrictions, and

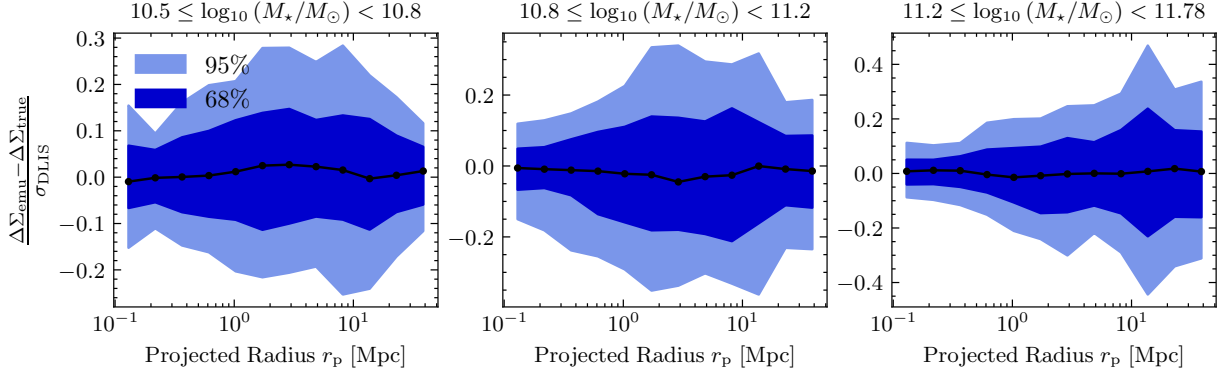


Figure 4.2: The $\Delta\Sigma_{\text{mocks}}$ emulator accuracy for the three stellar mass bins. $\Delta\Sigma_{\text{emu}}$ is the emulator prediction given a set of [HOD](#) parameters, and $\Delta\Sigma_{\text{true}}$ is the true [ESD](#) vector for the same set of [HOD](#) parameters. σ_{DLIS} is the measurement error from the [DLIS](#) [ESD](#) vectors shown in Figure [3.1](#).

so, we utilized all generated mocks, including all those generated when performing the iterative process described in [Appendix A.1](#). In total, we have 4253 mocks. For each mock, the satellite fractions and mean central halo masses for each bin were computed and saved. For each set of [HOD](#) parameters, we compute $\Delta\Sigma_{\text{tNFW}}$ for 100 random values of $\log_{10}\tau$ ranging from -5 to 5 . In total, before any additional cleaning/cuts, we have 425300 mocks for all three stellar mass bins. We remove all vectors in which $f_{\text{sat}} = 0$, leaving us with 425300, 418300, and 396100 mocks to train our emulators for each bin. The test sample consists of 20% of the total mocks randomly chosen, and the remaining 80% becomes our training sample. The accuracy plot and correlation matrix can be found in [Appendix A.3](#). We summarize that on average, the $\Delta\Sigma_{\text{tNFW}} f_{\text{sat}}$ emulators are accurate to less than 1% of the 1σ measurement error.

Chapter 5

Results

To reiterate, the goal is to derive a set of [HOD](#) parameters that produces a mock catalogue representative of our [DLIS](#) lens sample. As a reminder, the free parameters are $M_0, M_1, \gamma_1, \gamma_2, \sigma_c, a_1, b_0, B_{\text{cen}}$, and, $\log_{10} \tau$ (see Section 4.2 for definitions). With the generated emulators described in the previous sections, we ran a χ^2 minimization routine, where we fit all three stellar mass bins jointly. We then use the best-fit parameters, which fall within our training range of the emulator, to generate the ‘best-fit’ mock lens catalogue. The generated best-fit mock and its fit compared to the full-bin [DLIS](#) $\Delta\Sigma$ measurements are shown in Figure 5.1. The χ^2 values for the emulator-predicted vectors are 4.66, 17.52, and 13.90, for our three stellar mass bins, respectively. The χ^2 values for the generated best-fit mock are 11.33, 28.83, and 14.29. We fit all three stellar mass bins jointly, and so the joint χ^2 values are 36.08 and 54.45, respectively. With 36 radial bins and nine free parameters, the resulting degrees of freedom are 27. The covariance matrix used to calculate the aforementioned χ^2 values uses the [DLIS](#) measurement covariance matrix but also includes the errors for both the individual stellar mass bin $\Delta\Sigma_{\text{mock}}$ emulator and the $\Delta\Sigma_{\text{tNFW}f_{\text{sat}}}$ emulator. The fact that the χ^2 values worsen is not surprising, as we can consider the emulators to have learned the smoothed [ESD](#) vectors over an otherwise noisy parameter space. Thus, the emulator-predicted [ESD](#) vector is quite smooth when compared to the actual measurement made from generating the best-fit mock. We note that for our lowest and intermediate stellar mass bins, the differences in the χ^2 values are a result of an underprediction of the [ESD](#) vector below a radius of $r_p < 0.6$ Mpc. The subhalo contribution was expected to reduce this underestimation, as it is where we expect subhalos to affect the [ESD](#) profile. Nevertheless, the emulator predictions agree closely with the mock measurements on all other scales. The inferred satellite fractions are 15.5%, 11.4%, and 4.4% for our [DLIS](#) sample, according to the best-fit mock.

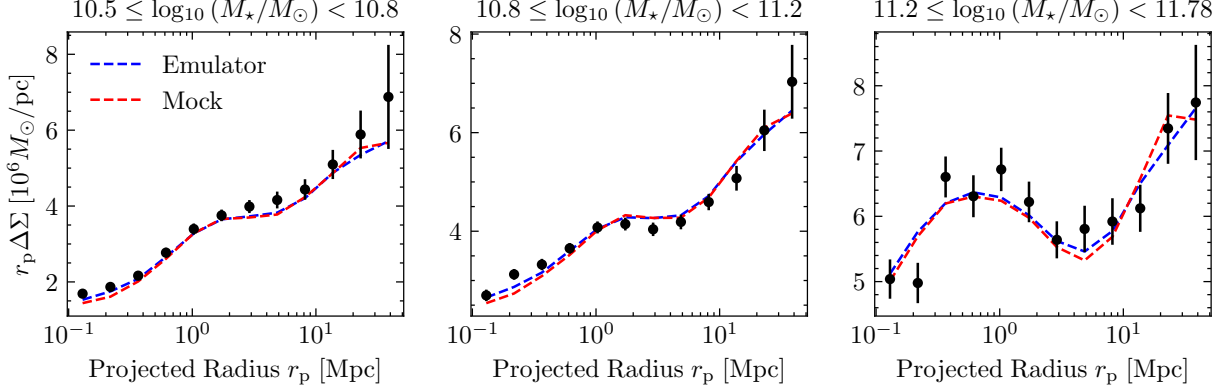


Figure 5.1: Plot of the best-fit mock against the [DLIS](#) full bin (i.e., not size split) $r_p \Delta \Sigma$. The blue dashed line is the emulator prediction for the given stellar mass bin. The red dashed line is the generated mock from [AbacusSummit](#), or in essence, the true mock given the same set of [HOD](#) parameters that the emulators predicted on. The plotted error bars are the combined error bars from the [DLIS](#) measurement covariance matrix and the covariance matrices of both the $\Delta \Sigma_{\text{mocks}}$ emulator covariance matrix and the $\Delta \Sigma_{\text{tNFW}} f_{\text{sat}}$ emulator covariance matrix. The χ^2 values for the emulator-predicted vectors are 4.66, 17.52, and 13.90. The χ^2 values for the generated best-fit mock are 11.33, 28.83, and 14.29. The joint χ^2 values are 36.08 and 54.45, respectively, for the emulator and mock prediction. With 36 radial bins and nine free parameters, there are 27 degrees of freedom.

Parameter	M_0	M_1	γ_1	γ_2	σ_c
Value	10.89 ± 0.10	11.03 ± 0.13	4.62 ± 0.44	0.12 ± 0.04	0.17 ± 0.03

Parameter	a_1	b_0	B_{cen}	$\log_{10} \tau$
Value	1.29 ± 0.28	-1.18 ± 0.18	0.11 ± 0.03	1.34 ± 0.47

Table 5.1: Best-fit [HOD](#) parameters obtained from χ^2 minimization using the emulators trained for each of the three stellar mass bins. Error bars are obtained from the inverse Hessian matrix. We caution the reader against overinterpreting these errors, as they are underestimated given that regions outside the emulator training space are explored. This results in the extrapolation of the [ESD](#) vector.

The best-fit [HOD](#) parameters found are summarized in [Table 5.1](#). The error bars stated are obtained from the inverse Hessian matrix. This matrix is an output from using a Broyden-Fletcher-Goldfarb-Shanno (BFGS) minimization algorithm. We caution the reader against overinterpreting these errors. They are underestimated because regions outside the emulator training space are explored and, in these areas, the emulator predictions are strongly prior-dominated.

5.1 Fit for Size Dependence

With our [HOD](#) parameters now frozen at the best-fit values, we now turn to quantifying the correlation between halo mass and galaxy size. Following the formalism described in [Section 4.2.2](#), we consider central and satellite galaxies individually, allowing both to have varying correlations between halo mass and galaxy size within our best-fit mock. Note that β_{cen} and β_{sat} , defined in [Eqn. 4.16](#), are not degenerate: they predict qualitatively different changes to the [ESD](#), as shown in [Appendix B](#). Because the satellites are offset from the halo centre, they have a much stronger impact on the [ESD](#) on scales of ~ 1 Mpc and above, whereas centrals affect all scales equally. We run an [HMC](#) with uniform priors for both β_{cent} and β_{sat} of -0.15 to 0.15 . [Figure 5.3](#) presents our [HMC](#) posteriors. Except for our lowest stellar mass bin of $10.5 \leq \log_{10}(M_{\star}/M_{\odot}) < 10.8$, we can constrain β_{cent} and β_{sat} . This indicates that large and small centrals in this stellar mass range inhabit halos of the same (or very similar) mass. This is shown when comparing the mean halo mass of centrals in $\mathcal{M}_1^{\star}$ (see [Table 5.2](#)), for which we find this value to be identical for small and large. Additionally, [Figure 5.2](#) shows the distribution of halo mass values for all centrals in $\mathcal{M}_1^{\star}$, revealing a distribution that is sharply peaked and noticeably narrower compared to the broader distributions in the two larger stellar mass bins. Generally, we find that as we increase in stellar mass, the value of β_{cent} becomes increasingly positive but is not tightly constrained except for our high mass bin of galaxies between $11.2 \leq \log_{10}(M_{\star}/M_{\odot}) < 11.78$. Based on [Eqn. 4.16](#), a positive β_{cent} indicates a positive correlation between halo mass and galaxy size. Physically, we interpret this as large central galaxies living in more massive halos than smaller central galaxies of the same stellar mass. For the satellites, we find that the opposite is true: for our lowest and intermediate stellar mass bins, the β_{sat} values are negative. The interpretation is that small satellite galaxies reside in a more massive host halo compared to large satellite galaxies of the same stellar mass. At intermediate radii, typically of $0.5 - 1.5$ Mpc, the offset-group halo term dominates weak lensing profiles ([Gillis et al., 2013](#)). The offset group term arises from the presence of satellite galaxies, reflecting the contribution of the dark matter halo of the host in which they reside. As

shown in Figure 3.2, there is a strong contribution from the offset-group term for the small galaxy ESD in our lowest and intermediate stellar mass bins. The offset-group term depends on both the satellite fraction and the host halo mass, which may exhibit degeneracy. Given the two-step modelling approach adopted here, we are unable to probe this potential degeneracy directly. However, we can compare the satellite fractions and host halo mass values between the *small* and *large* subsamples. Table 5.2 tabulates the satellite fraction and mean host halo mass for each size subsample within each stellar mass bin. We note that the satellite fraction in the *small* bin is slightly higher than in the *large* bin for $\mathcal{M}_1^*$ and $\mathcal{M}_2^*$. Since these satellite fractions are comparable, the observed differences in the offset-group term likely arise from the more significant differences in the host halo masses for satellites in those bins. Figure B.2 illustrates the effect of β_{sat} on the ESD vector, where we explicitly see changes to the ESD vectors around the offset-group term when populating galaxies in more massive or less massive host halos. This behaviour reinforces our earlier interpretation and results regarding the negative values of β_{sat} .

In Figure 3.2, the lensing fits to both the small and large ESD vectors using the best-fit β_{cent} and β_{sat} taken from the HMC marginal constraints are shown. Plotted as dashed lines are the best-fit lines, and plotted as a solid black line is the ESD measurement for the null cases (i.e., no correlation between halo mass and galaxy size, and so $\beta_{\text{cent}} = \beta_{\text{sat}} = 0$). The χ^2 values for these fits are summarized in Table 5.3 as individual χ^2 values for small and large and are given jointly. Overall, our joint fits are a significant improvement compared to the null case. Notably, comparing $\chi_{\text{null, joint}}^2$ to χ_{joint}^2 , we report a 3.6σ , 1.2σ , and 5.3σ significance of our fits for each stellar mass bin, respectively. These trends are further supported by the comparison of the mean halo mass values between small and large galaxy subsamples for both centrals and satellites, which are tabulated in Table 5.2. This reveals a correlation between halo mass and galaxy size for red, early-type galaxies at a fixed stellar mass.

5.2 Calculation of η

Recall that β is the derivative of how galaxy size changes with respect to halo mass. In other words, $\beta = \Delta(\Delta R') / \Delta(\Delta M'_h)$, where,

$$\Delta(\Delta R') = \Delta R'_{\text{large}} - \Delta R'_{\text{small}}, \quad (5.1)$$

and

$$\Delta(\Delta M'_h) = \Delta M'_{h, \text{large}} - \Delta M'_{h, \text{small}}. \quad (5.2)$$

		$\log_{10} \left(\frac{M_{\text{h, cent}}}{M_{\odot}} \right)$	$\log_{10} \left(\frac{M_{\text{h, sat}}}{M_{\odot}} \right)$	$N_{\text{gal, mock}}$	f_{sat}
$\mathcal{M}_1^*$	Small	11.81	13.49	1,127,029	17.0 %
	Large	11.81	13.41	1,087,564	13.8 %
$\mathcal{M}_2^*$	Small	12.21	13.71	1,433,553	12.0 %
	Large	12.23	13.67	1,414,732	10.7 %
$\mathcal{M}_3^*$	Small	12.67	14.08	179,367	4.1 %
	Large	12.80	14.12	184,309	4.7 %

Table 5.2: Mean halo mass values and satellite fractions for central and satellite-type galaxies for each stellar mass bin and size subsample taken from the best-fitting mock. Note that halo mass values for satellites are not subhalo masses but rather host halo masses in which satellites reside. The stated satellite fractions in the final column are representative of the given stellar mass and size bin. For example, $f_{\text{sat}, \mathcal{M}_1^*, \text{small}} = N_{\text{sat}, \mathcal{M}_1^*, \text{small}} / N_{\mathcal{M}_1^*, \text{small}}$, where $N_{\text{sat}, \mathcal{M}_1^*, \text{small}}$ is the number of satellite galaxies in the lowest stellar mass bin classified as *small*. The number of galaxies classified as *small* in $\mathcal{M}_1^*$ is $N_{\mathcal{M}_1^*, \text{small}}$. The weighted average of f_{sat} values given the number of galaxies in the corresponding size bins recovers the overall satellite fractions for the three stellar mass bins, which are 15.5%, 11.4%, and 4.4%, respectively.

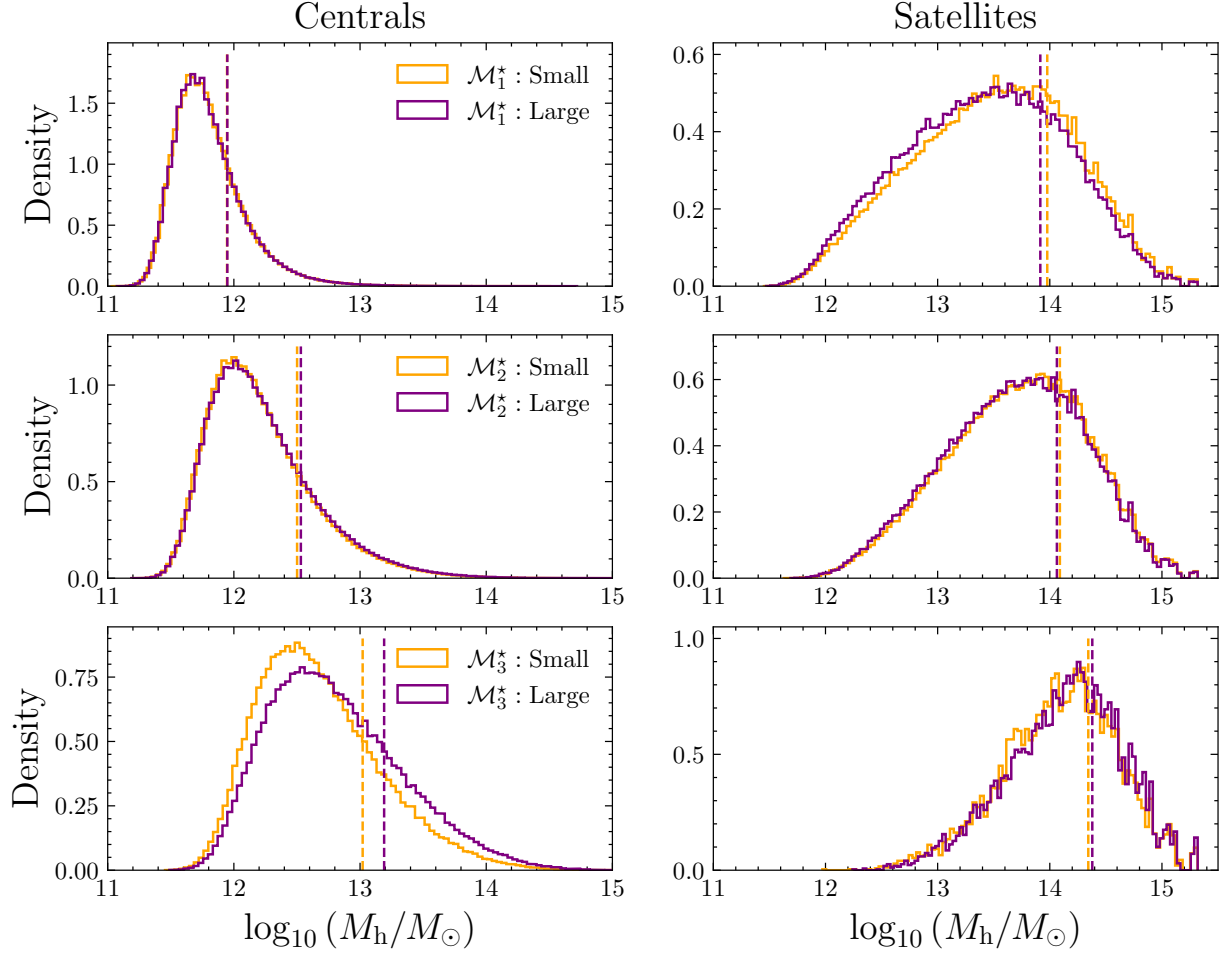


Figure 5.2: Histogram of the halo masses of centrals and satellites in the best-fit mock split into *small* and *large* subsamples. Moving from top to bottom, we increase in stellar mass bins. Orange is the *small* galaxies, and purple is the *large*. The vertical dashed orange/purple lines represent the mean halo mass for the plotted distribution.

To compare the results of this work to others in the literature (see Section 1.5), we must quantify η , where it is parametrized as

$$\eta = \Delta(\Delta M'_h) / \Delta(\Delta R') . \quad (5.3)$$

We cannot simply take the inverse of β because the process of assigning galaxy sizes introduces stochasticity and noise; see Eqn. 4.16. We instead calculate the values of η using the best-fit mock after applying the best-fitting β_{cent} and β_{sat} . For example, for central galaxies in $\mathcal{M}_3^*$, the value for $\Delta M'_h$; see Eqn. 4.15, for the large and small subsamples is 0.46 and 0.29. This gives a value of $\Delta(\Delta M'_h) = 0.17$. For $\Delta R'$, see Eqn. 4.14; the values for large and small are 0.18 and -0.15, resulting in $\Delta(\Delta R') = 0.33$. This gives a value of $\eta = 0.51$ as reported in Table 5.3. Error bars for η are determined via bootstrap resampling, which we summarize now. Given the HMC posteriors shown in Figure 5.3, we draw 1000 random unique values for β_{cent} and β_{sat} . For each sampled value, we assign sizes to galaxies within our best-fit mock according to Eqn. 4.16. The catalogue is then split into three stellar mass bins and two size bins, from which we calculate the value for η . The standard deviation of the 1000 η values is taken to be the error. We discuss and compare these results with other studies in the next section.

The main result of this thesis is the following. For red, early-type central galaxies, we find no statistically significant correlation between halo mass and galaxy size within the stellar mass ranges of $10.5 \leq \log_{10}(M_*/M_\odot) < 10.8$ and $10.8 \leq \log_{10}(M_*/M_\odot) < 11.2$. However, we find a positive correlation of $\eta_{\text{cent}, \mathcal{M}_3^*} = 0.51 \pm 0.14$ between halo mass and galaxy size for the most massive galaxies within the stellar mass range of $11.2 \leq \log_{10}(M_*/M_\odot) < 11.78$. For red, early-type satellite galaxies, we find significant evidence of a negative correlation between host halo mass and galaxy size in the lowest stellar mass bin ($\mathcal{M}_1^*$) of $\eta_{\text{sat}, \mathcal{M}_1^*} = -0.19 \pm 0.05$. For intermediate and high stellar mass satellites, our results are consistent with no correlation between host halo mass and galaxy size. A key caveat is that the satellite fraction decreases with increasing stellar mass. Within the highest stellar mass bin ($\mathcal{M}_3^*$), the satellite fraction is approximately 4%, making it difficult to place strong constraints on satellite behaviour in this regime.

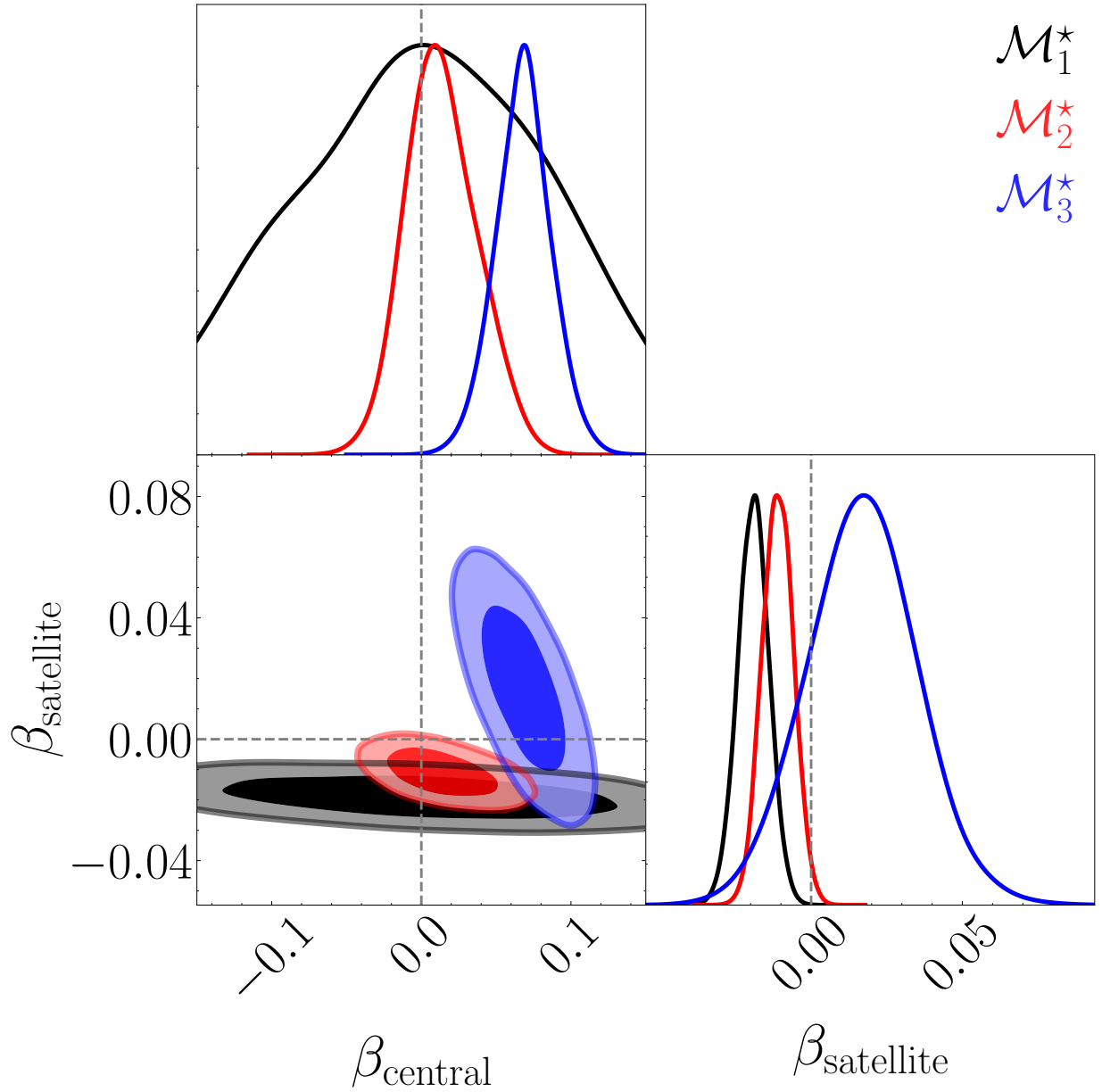


Figure 5.3: Posterior distributions of β_{cent} and β_{sat} from an [HMC](#) algorithm. The black posterior is for the lowest stellar mass bin of $10.5 \leq \log_{10}(M_*/M_\odot) < 10.8$. The red posterior is for the intermediate stellar mass bin of $10.8 \leq \log_{10}(M_*/M_\odot) < 11.2$. The blue posterior is for our largest stellar mass bin of $11.2 \leq \log_{10}(M_*/M_\odot) < 11.78$. The grey dashed line shows the position of $\beta_{\text{cent}} = 0$ and $\beta_{\text{sat}} = 0$. The marginalized 1D constraints are tabulated in Table 5.3 and used further in this analysis.

	$\chi^2_{\text{null, small}}$	$\chi^2_{\text{null, large}}$	$\chi^2_{\text{null, joint}}$	χ^2_{small}	χ^2_{large}	χ^2_{joint}
$\mathcal{M}_1^*$	20.25	10.27	30.52	9.72	5.05	14.77
$\mathcal{M}_2^*$	17.54	25.84	43.38	21.56	18.78	40.34
$\mathcal{M}_3^*$	23.94	29.16	53.10	13.88	6.93	20.81
	β_{cent}	β_{sat}	η_{cent}	η_{sat}		
$\mathcal{M}_1^*$	$0.002^{+0.088}_{-0.075}$	-0.019 ± 0.005	0.007 ± 0.24	-0.19 ± 0.05		
$\mathcal{M}_2^*$	$0.014^{+0.022}_{-0.027}$	-0.011 ± 0.005	0.090 ± 0.17	-0.09 ± 0.05		
$\mathcal{M}_3^*$	0.068 ± 0.019	0.017 ± 0.018	0.51 ± 0.14	0.1 ± 0.1		

Table 5.3: Results of the fit to size dependence. *Columns 1-6:* χ^2 values for the fits shown in Figure 3.2. The null case refers to when $\beta_{\text{cent}} = \beta_{\text{sat}} = 0$, which physically means there is no correlation between halo mass and galaxy size. For each stellar mass bin, the small and large ESD vectors are fit together. The DOF is 22, as we have 24 radial bins and two free parameters. The definitions of our stellar mass bins, $\mathcal{M}_1^*$, $\mathcal{M}_2^*$, and $\mathcal{M}_3^*$ are outlined in Section 2.3.3. *Columns 7-8:* The best-fit values for β_{cent} and β_{sat} are obtained from the HMC posteriors shown in Figure 5.3. These marginalized 1D means taken from the posteriors are taken as the best fit values and used further in this analysis. *Columns 9-10:* The values in the final two columns, η_{cent} and η_{sat} , are the measured correlations between halo mass and galaxy size for our sample of red, early-type galaxies, as inferred from our best-fit mock galaxy catalogue. $\eta = \Delta(\Delta M'_h) / \Delta(\Delta R')$ is the slope that describes the correlation between halo mass and galaxy size and is calculated for both central and satellite galaxies.

Chapter 6

Discussion

The works highlighted in Section 1.5 focus primarily on massive, red, early-type galaxies that are primarily central-type galaxies. The analysis we perform allows us to consider the effects of centrals to compare with previous studies, but also isolate and investigate satellite-type galaxies for the first time. Figure 6.1 presents results from studies discussed in Section 1.5, each of which independently determines the correlation between halo mass and galaxy size of central galaxies, for which we quantified their results as a linear slope/derivative η ; see Eqn. 5.3. The values of η found in this work are summarized in Table 5.3.

Given the detections found for central galaxies at high stellar masses and for satellites at low stellar masses, we ask, *How much does this correlation contribute to the overall scatter in the SHMR?* For central galaxies in $\mathcal{M}_3^*$ and satellite galaxies in $\mathcal{M}_1^*$, the scatter in $\Delta R'$ is 0.21 dex for both. The scatter in the halo mass, $\Delta M'_h$, accounting for the measured η value is given by $\sigma(\Delta M'_h) \approx \eta(M_\star)\sigma(\Delta R')$. The scatter in halo mass is calculated to be ~ 0.11 dex for centrals in $\mathcal{M}_3^*$, and ~ 0.04 dex for satellites in $\mathcal{M}_1^*$. The scatter in the SHMR we are interested in comparing with is the scatter in halo mass at a given stellar mass, $\sigma(M_h | M_\star)$. For high-mass central galaxies, the scatter is ≈ 0.5 dex (Kravtsov et al., 2018; Bradshaw et al., 2020). For low-mass galaxies, the scatter is ~ 0.2 dex (Zu & Mandelbaum, 2015). Given these values, it is clear that the scatter due to the halo mass-size dependence does not account for a significant amount of the scatter seen in the SHMR.

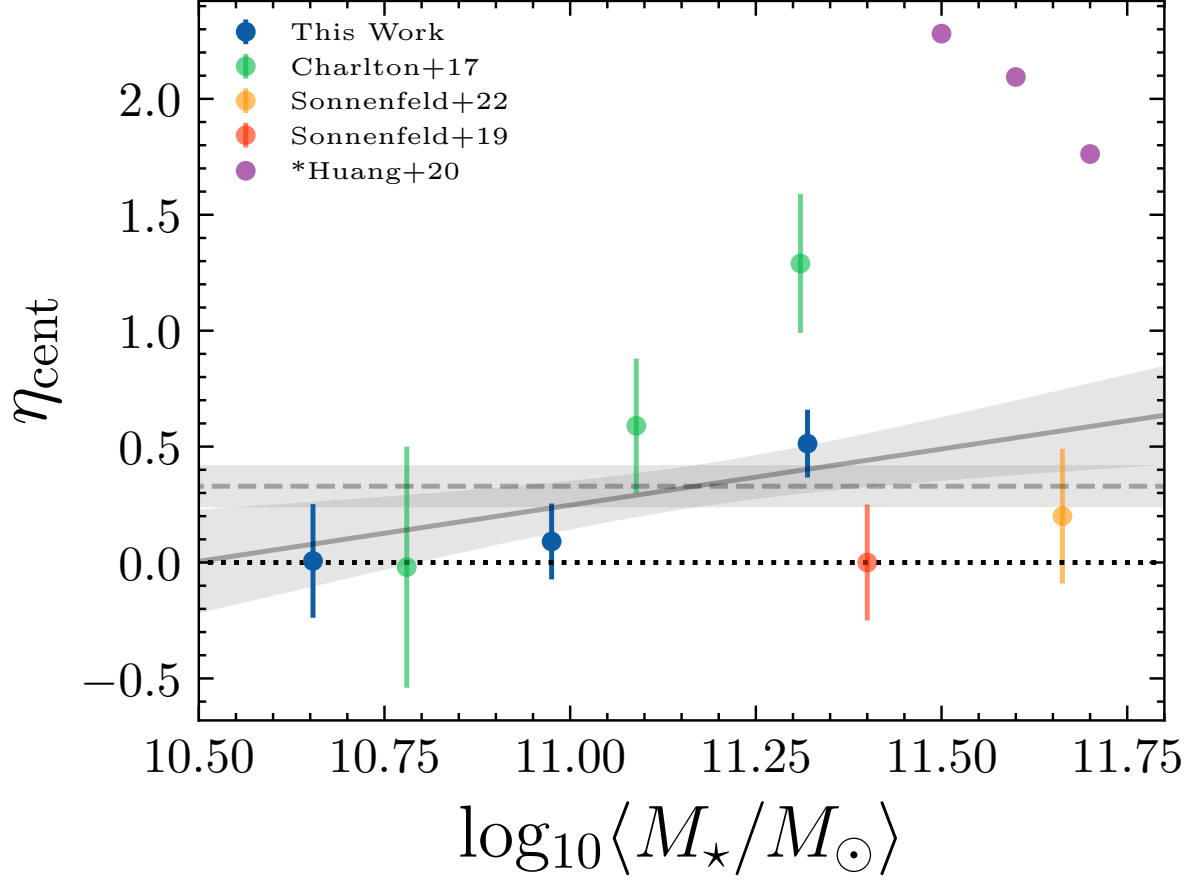


Figure 6.1: Summary plot translating results from other works into the definition of η in Eqn. 5.3. η_{cent} is the correlation of halo mass and galaxy size, strictly for red, early-type central galaxies. The dashed black line represents a fit of constant η , and the solid black line is the fit assuming a linear model of $\eta = s_{\eta} \log_{10}(M_{\star}/10^{11}M_{\odot}) + C$. The data points from H20 (in purple) were not considered when fitting the two models. The best fitting constant is $\eta = 0.33 \pm 0.08$. For the linear model, we find $s_{\eta} = 0.49 \pm 0.29$ and $C = 0.25 \pm 0.09$. The corresponding χ^2 values are 18.88 (for 8 degrees of freedom) and 16.18 (for 7 degrees of freedom), respectively, for the two fits.

6.1 Comparison to Previous Works

C17 reported an overall positive correlation of $\eta = 0.42 \pm 0.12$, averaging over their stellar mass bins and colour bins. Averaging over the stellar mass bins for red galaxies only, they report a positive correlation of $\eta = 0.53 \pm 0.15$. Although one could compute a weighted average of the slope values (η) across stellar mass bins, we choose not to do so. Each η value reflects the correlation between halo mass and galaxy size within a specific stellar mass range, and averaging would blur any potential trends or variations with stellar mass. Since our goal is to investigate how this relationship evolves at fixed stellar mass across different stellar mass bins, it is more informative to analyze the η values individually rather than combining them into a single summary statistic. The three stellar mass bins in this work are closest in value of average stellar mass ($\log_{10}\langle M_{\star}/M_{\odot} \rangle$) to C17’s three highest stellar mass bins. Our results are consistent with the overall trend reported by C17, namely that η increases with stellar mass; however, the magnitude of this increase differs. Specifically, our values for η agree with those of C17 to within 0.05σ , 1.5σ , and 2.4σ across the three stellar mass bins, respectively. A caveat to consider is the presence of systematic bias in stellar mass estimates derived from Spectral Energy Distribution (SED) fitting codes. These estimates are sensitive to assumptions and choices related to key components such as star formation history and stellar population synthesis models, leading to a typical systematic shift of approximately $0.1 - 0.2$ dex (Mobasher et al., 2015; Paulino-Afonso et al., 2022). Considering a potential $0.1 - 0.2$ dex shift of our stellar mass bins, our results are in agreement with C17. Additionally, unlike C17 who assumed and fitted a single NFW profile to their ESD measurements, we consider the contributions from additional halo terms at intermediate scales (i.g., satellite offset-group term) and at large scales (i.e., 2-halo term) within our modelling. This added complexity in the modelling of the ESD profiles could contribute to differences in η measurements between this work and theirs.

S19 followed a Bayesian hierarchical approach with HSC weak lensing data and the Baryon Oscillation Spectroscopic Survey (BOSS) CMASS sample of galaxies. They found no correlation between halo mass and galaxy size, reporting $\eta = 0.00 \pm 0.25$ using a de Vaucouleurs model when fitting for half-light radius for their galaxies. With an average $\log_{10}(M_{\star}/M_{\odot})$ for their sample of 11.4, we can compare their results to our highest stellar mass bin. Our value of $\eta_{\text{cent}, \mathcal{M}_3^{\star}} = 0.51 \pm 0.14$ is consistent with S19 within 1.8σ . S19 argued that the positive correlation detected by C17 is a result of observational scatter in the measured stellar masses not having been factored into the analysis. In this work, we account for the scattering of stellar masses in order to minimize any effect of Eddington bias. Despite that, our highest stellar mass bin result is consistent with both C17 and

S19, at the 2.4σ and 1.8σ levels, respectively. The discrepancy between these two previous studies resulted in no unified view on the matter of halo mass versus galaxy size (at high M_*). Our measurement lies between the values reported by these two works, albeit closer to the result of S19. This intermediate value may provide a more accurate characterization of the underlying correlation, given that the ESD measurements in this work are the most accurate to date on this matter.

The study by H20 investigated the connection between halo mass and the spatial distribution of stellar mass in massive central galaxies. They found that at fixed stellar mass, central galaxies residing in more massive halos exhibit more extended stellar mass distributions and contain less stellar mass within an inner aperture of 10 kpc, denoted as $M_{*,10,\text{kpc}}$. This trend is consistent with a positive correlation between halo mass and galaxy size. In Figure 9a of H20, $\log_{10}(R_{50}/\text{kpc})$ is plotted against the maximum stellar mass, M_*^{max} , and coloured by predicted halo mass. From this figure, we estimate effective η values at fixed stellar masses of 11.5, 11.6, and 11.7, which are included in Figure 6.1 (shown in purple). We do not have error estimates for these η values and are therefore unable to determine the statistical consistency (e.g., in terms of σ) between the results of H20 and S22. Nonetheless, a qualitative comparison between the η values shows a potential discrepancy in η for massive central galaxies.

Given these studies, if we consider every data point at face value, disregarding the scatter seen between η values at the same stellar mass, an interesting quantity to consider is the slope of η versus M_* , or $d\eta/d\log_{10}(M_*/M_\odot)$. We define the variable s_η as this slope. We fit a constant η and a linear model of $\eta = s_\eta \log_{10}(M_*/10^{11}M_\odot) + C$ to the data in Figure 6.1, excluding the H20 points. The best fitting constant is $\eta = 0.33 \pm 0.08$. For the linear model, we find $s_\eta = 0.49 \pm 0.29$ and $C = 0.25 \pm 0.09$. The corresponding χ^2 values are 18.88 (for 8 degrees of freedom) and 16.18 (for 7 degrees of freedom), respectively, for the two fits. Although the linear model improves the fit compared to fitting with a constant, the reduced χ^2 of both fits is much greater than 1, which can be attributed to the lack of data points and the scatter.

Using the linear and constant fits described above, we can make a comparison to the data point of S22. The average stellar mass of their galaxy sample is $\log_{10}\langle M_*/M_\odot \rangle = 11.66$. The fitted linear model predicts a value of $\eta_{\text{cent}} = 0.57 \pm 0.21$. In comparison to $\eta = 0.20 \pm 0.29$ found by S22, the predicted slope is consistent to within 1σ .

To our knowledge, no other works in the literature have investigated the correlation between host halo mass and galaxy size for red, early-type satellite galaxies. We repeat the same analysis done for η_{cent} , but for the η_{sat} values found in this work. Figure 6.2 shows the η_{sat} values found in this work, along with linear and constant model fits to the data.

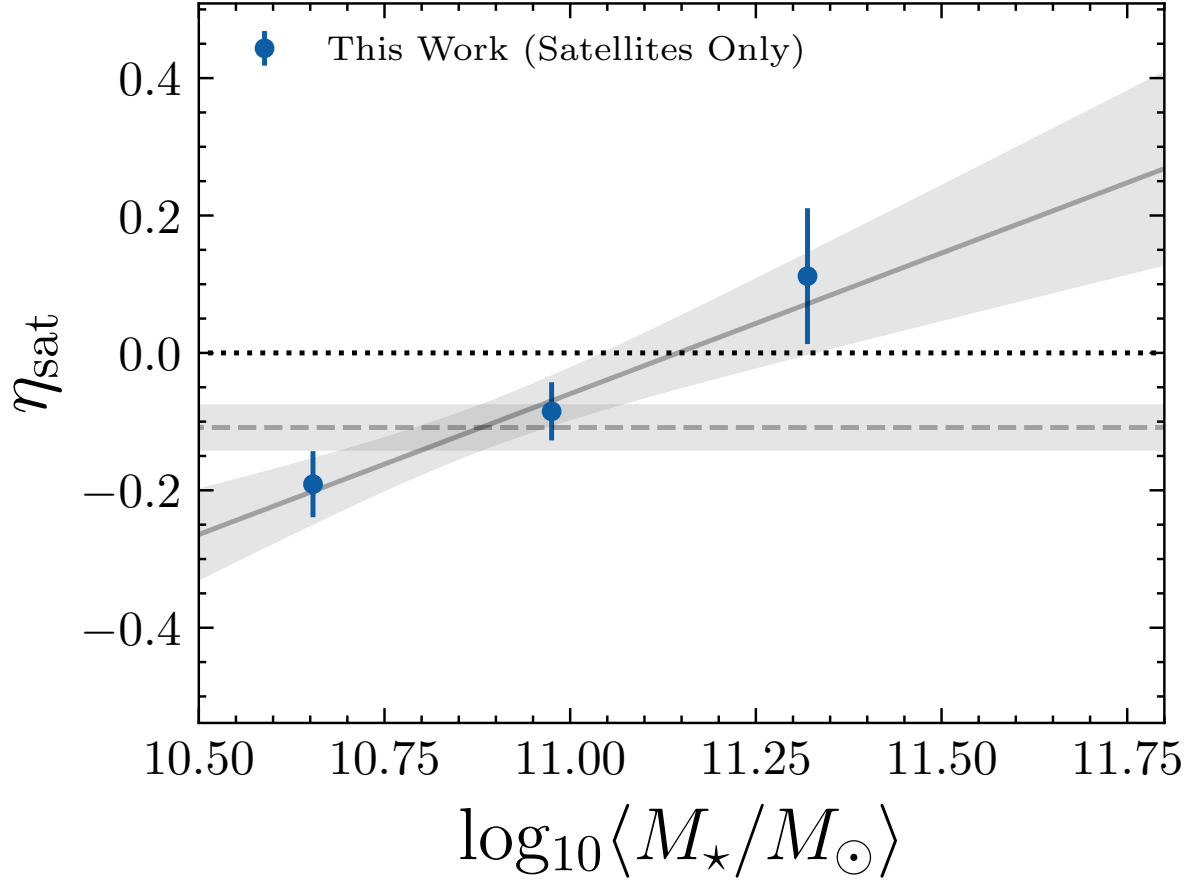


Figure 6.2: Plot of the η_{sat} values found in this work for the three stellar mass bins. Eqn. 5.3 describes the definition of η . The dashed black line represents a fit of constant η , and the solid black line is the fit assuming a linear model of $\eta = s_{\eta} \log_{10}(M_{\star}/10^{11}M_{\odot}) + C$. The best-fitting constant is $\eta_{\text{sat}} = -0.11 \pm 0.03$. The best-fit slope/intercept for the linear model is $s_{\eta_{\text{sat}}} = 0.41 \pm 0.15$ and $C = -0.059 \pm 0.035$. The χ^2 values for these two cases are 8.2 and 0.34, respectively, with [Degrees of Freedom \(DOF\)](#) of 2 and 1.

Specifically, we fit a constant η model and a linear model of $\eta = s_\eta \log_{10}(M_\star/10^{11} M_\odot) + C$. We find the best-fitting constant is $\eta_{\text{sat}} = -0.11 \pm 0.03$. For the linear model, we find a slope of $s_{\eta_{\text{sat}}} = 0.41 \pm 0.15$ and a constant of $C_{\text{sat}} = -0.059 \pm 0.035$. The χ^2 values for these two cases are 8.2 and 0.34, respectively. The [DOF](#) is 2 and 1 for the two cases.

$s_\eta > 0$ implies that the correlation between halo mass and galaxy size becomes stronger with increasing stellar mass. In practice, determining s_η offers an informative way to capture how the halo mass–size relation evolves with stellar mass, providing a potentially powerful statistic for both models and observations. Knowledge of s_η becomes particularly useful when applied to new observational or simulated galaxy catalogues. Given a sample of galaxies with known stellar masses and effective radii, and assuming a power-law relationship between halo mass and galaxy size, as done in this work, one can invert our derived η trends to estimate the typical halo masses for red, early-type central or satellite galaxies. In simulations where halo masses are known but galaxy sizes have not yet been assigned, our measured s_η can be used to inform size estimates. In this way, s_η serves as a practical tool for linking stellar mass, size, and halo mass across both data and simulations.

6.2 Impact of Galaxy Size Uncertainty

Galaxy size measurements within the [DLIS](#) catalogues are obtained from the Tractor code ([Lang et al., 2016](#)), which fits several light profiles (e.g., exponential, de Vaucouleurs, or a combination) to estimate galaxy sizes. However, no measurement uncertainties for individual galaxy sizes are provided in the [Zou et al. \(2019, 2022\)](#) catalogues. [C17](#) compared size measurements from [CFHT](#) imaging to higher-resolution HST data, and state an average error on galaxy size of about 0.15 dex. In addition to this, measurement errors in stellar mass propagate into size measurements because size estimates depend on both the galaxy’s light profile and total luminosity, which are correlated with galaxy stellar masses. Fitting a simple linear model of $\log_{10} R_{\text{eff}} = m \log_{10} M_\star$ within our three stellar mass bins, we find slopes of 0.36, 0.49, and 0.57. Assuming an average stellar mass uncertainty of 0.2 dex and an average slope of about 0.5, this propagates to an additional size uncertainty of approximately 0.1 dex. Our halo mass–size model incorporates an additional scatter term σ_{extra} (see Eqn. 4.17) when assigning galaxy sizes. On average, σ_{extra} is about 0.2 dex, which sets an upper limit on the total measurement error in R_{eff} . Defining $\sigma_{R_{\text{eff}}}(\sigma_{M_\star})$ as the measurement error due to stellar mass uncertainties and $\sigma_{R_{\text{eff}}}^{\text{obs}}$ as the direct observational size error, we write:

$$\sigma_{\text{extra}} \geq \sqrt{[\sigma_{R_{\text{eff}}}(\sigma_{M_\star})]^2 + [\sigma_{R_{\text{eff}}}^{\text{obs}}]^2}. \quad (6.1)$$

Given values of 0.15 dex and 0.1 dex for these uncertainties, the expected combined error is approximately 0.18 dex. Therefore, our model sufficiently accounts for the anticipated uncertainties in galaxy size measurements.

As described in Section 2.3.4, galaxies are split into two size bins based on the median size of a galaxy of a given stellar mass. This is done to maximize the number of galaxies in each bin and the SNR. Given that most of the galaxies lie near the interpolated median size function (see solid black line in Figure 2.3), measurement errors in galaxy size could lead to cross-contamination of galaxies. For example, we define the true and observed galaxy size as $R_{\text{eff}}^{\text{true}}$ and $R_{\text{eff}}^{\text{obs}}$, where $R_{\text{eff}}^{\text{obs}}$ contains some measurement error. If $R_{\text{eff}}^{\text{true}} < R_{\text{eff}}^{\text{median}}$ this galaxy would have been assigned into the *small* bin, but, when considering uncertainties, if $R_{\text{eff}}^{\text{obs}} > R_{\text{eff}}^{\text{median}}$ this galaxy which should have been assigned as *small* will be assigned *large*. This could lead to an underestimation of the halo mass-galaxy size dependence. To explore the effects of this cross-contamination on our measured η values, without having to re-do the observational measurements and model fitting, we consider a simulated approach to gauge the effect that observational scatter has on $\Delta(\Delta R')$ (defined in Section 5.2).

For given intrinsic (σ_{int}) and observational (σ_{obs}) scatters, we ensure that their combined value does not exceed an assumed total scatter of 0.2. Values of $\log_{10} R_{\text{eff}}^{\text{true}}$ are drawn from a lognormal distribution with mean 0.5 and sigma σ_{int} . These values are split into *small* and *large* according to the median of $\log_{10} R_{\text{eff}}^{\text{true}}$ values. For example, if $\log_{10} R_{\text{eff}}^{\text{true}} > \text{median}(\log_{10} R_{\text{eff}}^{\text{true}})$ the galaxy is classified as *large*. We compute the pre-scattered values for $\Delta(\Delta R')$. Given the value for σ_{obs} , $\log_{10} R_{\text{eff}}^{\text{true}}$ values are randomly scattered, giving $\log_{10} R_{\text{eff}}^{\text{obs}}$. We now split into *small* and *large* according to the median of $\log_{10} R_{\text{eff}}^{\text{obs}}$ values (e.g., if $\log_{10} R_{\text{eff}}^{\text{true}} > \text{median}(\log_{10} R_{\text{eff}}^{\text{obs}})$ the galaxy is *large*). We then calculate the post-scattered values for $\Delta(\Delta R')$ which are based on the true galaxy sizes. This simulation approach is repeated 10000 times for varying values of σ_{int} and σ_{obs} .

Figure 6.3 plots the post-scattered versus pre-scattered values for $\Delta(\Delta R')$, coloured by the applied σ_{obs} . We see that, as the observational error increases, values for $\Delta(\Delta R')$ approach 0 in both the pre- and post-scattered cases. For the pre-scattered case, since $\sigma_{\text{obs}} \approx 0.2$, the intrinsic scatter is 0, which leads to values of $\log_{10} R_{\text{eff}}^{\text{true}}$ that have no spread. For the post-scattered case, the sizes are effectively random given the large observational scatter, leading to a difference in $\Delta R'_{\text{large}} - \Delta R'_{\text{small}} = 0$. Previously, we argued a reasonable value for σ_{obs} of 0.18. The vertical and horizontal grey dashed lines indicate where $\sigma_{\text{obs}} \sim 0.1$. We find that the post-scattered $\Delta(\Delta R')$ value is 0.05, while the pre-scattered value is 0.12. This is a decrease of about 58%, which would lead to an overestimation of η by a factor of 2.4 (i.e., $\eta_{\text{obs}} = 2.4\eta_{\text{true}}$). Generally, we find that the post-scattered values are lower than the pre-scattered, indicating that observational scatter in R_{eff} will lead to an overall overestimation of η .

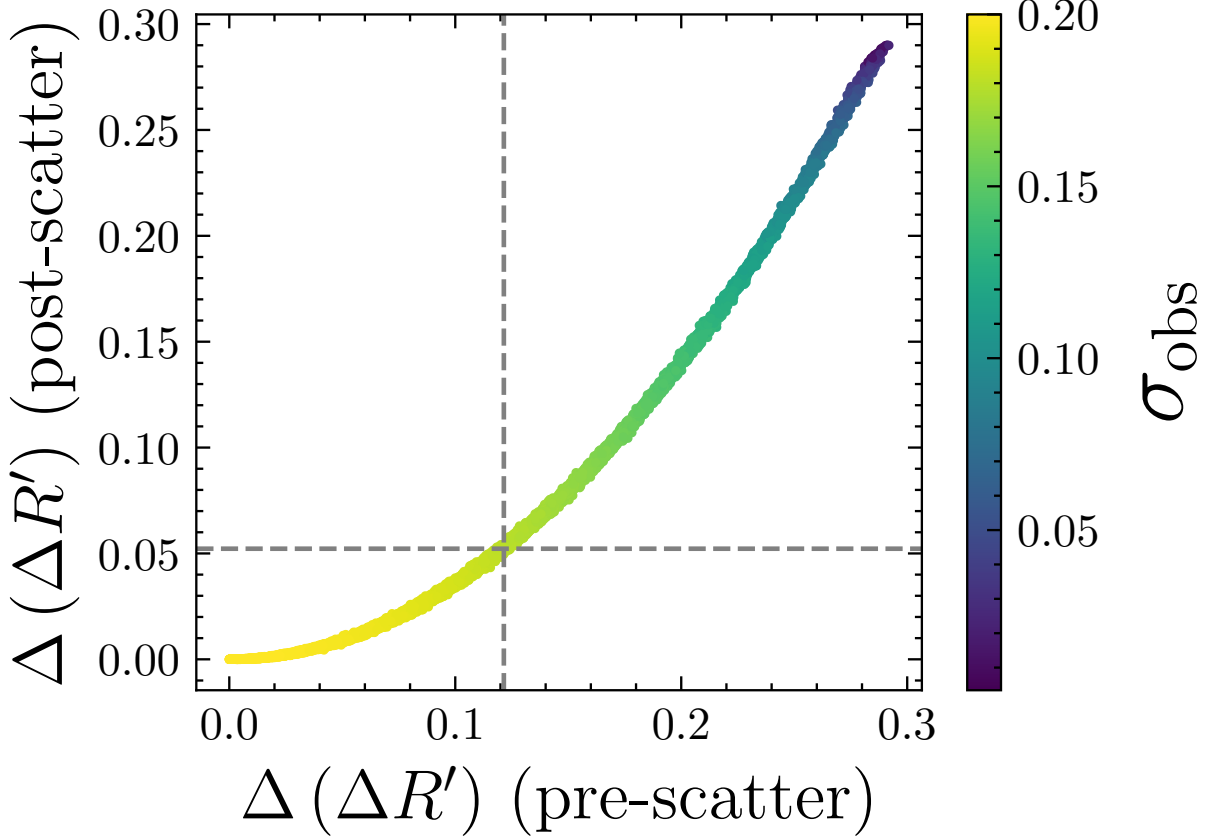


Figure 6.3: Results of a simulated approach to gauge the effect of observational scatter on measured $\Delta(\Delta R')$ values and its consequence on η . Plotted on the x-axis are the pre-scattered values drawing values for $\log_{10} R_{\text{eff}}^{\text{true}}$ from a lognormal distribution with mean 0.5 and intrinsic scatter σ_{int} . The post-scattered values are calculated after applying a random observational scatter to $\log_{10} R_{\text{eff}}^{\text{true}}$ assuming a value for the observational scatter σ_{obs} . The simulation is repeated 10,000 times for varying values of σ_{int} and σ_{obs} with the condition that $\sigma_{\text{tot}} \geq \sqrt{\sigma_{\text{int}}^2 + \sigma_{\text{obs}}^2}$, where $\sigma_{\text{tot}} = 0.2$. The vertical and horizontal dashed grey lines present where $\sigma_{\text{obs}} \sim 0.18$.

6.3 Insights from the Environmental Dependence on the Size-Stellar Mass Relation

Consistent with studies on the correlation of galaxy size and halo mass are studies of the environmental dependence of the galaxy [Size-Stellar Mass Relation \(SMR\)](#). While stellar mass is strongly correlated with galaxy size, the environment in which galaxies reside can influence size growth and evolution. This is most evident with satellite galaxies, where our results indicate that smaller satellites reside in more massive host halos than larger satellites at fixed M_* .

To date, studies on the environmental dependence of the [SMR](#) have led to conflicting results. There have been studies of varying redshifts that find evidence for no environmental dependence in the [SMR](#) (e.g. [Huertas-Company et al., 2013a,b](#); [Shankar et al., 2014](#); [Zhao et al., 2015](#); [Chen et al., 2024](#); [Figueira et al., 2024](#); [Abdullah et al., 2025](#)), and studies supporting a dependence of environment on the [SMR](#) (e.g. [Raichoor et al., 2012](#); [Cooper et al., 2012](#); [Shankar et al., 2013](#); [Lani et al., 2013](#); [Yoon et al., 2017](#); [Ghosh et al., 2024](#); [George et al., 2025](#)). For a detailed list summarizing the current state of this discussion, we refer readers to Table 3 of [Abdullah et al. \(2025\)](#). Ultimately, the main challenges faced by most of these studies stem from a complex combination of factors. These include small sample sizes for field, cluster, and group [Early-Type Galaxies \(ETGs\)](#) (which reduce statistical power), the differences in environment definitions, and the reliance on photometric redshifts and stellar mass estimates. Nevertheless, we highlight some interesting results most relevant to central and satellite galaxies.

[Yoon et al. \(2017\)](#) studied the environmental dependence of the [SMR](#) with a sample of 73,116 red [ETGs](#) with stellar masses ranging from $10.7 \leq \log_{10}(M_*/M_\odot) \leq 11.9$. The environment was defined as the surface number density of all galaxies within the 10th nearest neighbour. They find that galaxies with $\log_{10}(M_*/M_\odot) > 11.2$ are 20% – 40% larger in denser environments than similar galaxies in underdense environments. Below $\log_{10}(M_*/M_\odot) < 11.2$ this environmental dependence disappears. They use the Millennium 1 simulation ([Springel et al., 2005](#)) and detect the same environmental dependence for galaxies more massive than $\log_{10}(M_*/M_\odot) > 11.2$. These results are strikingly consistent with our results for central-type galaxies in the stellar mass range of $11.2 \leq \log_{10}(M_*/M_\odot) < 11.78$, where we find a correlation between halo mass and galaxy size of $\eta_{\text{cent}, \mathcal{M}_3^*} = 0.51 \pm 0.14$. A compelling argument for these results concerns the merger histories of the most massive galaxies. [Yoon et al. \(2017\)](#) provide simulation evidence for a strong environmental dependence on the merger histories of galaxies with $\log_{10}(M_*/M_\odot) > 11.0$. They note that the number of major and minor mergers increases

with stellar mass beyond $10^{11} M_{\odot}$, and that the number of mergers is consistently higher in denser environments. This is consistent with results from [Bernardi et al. \(2011\)](#) who identified that major mergers should dominate the assembly history of ETGs for stellar masses above $M_{\star} > 2 \times 10^{11} M_{\odot}$. Generally, mergers are significant contributors to the size evolution of massive, quiescent ETGs within simulations (e.g. [Boylan-Kolchin et al., 2006](#); [López-Sanjuan et al., 2012](#); [Oogi et al., 2016](#)) and observational studies (e.g. [Bezanson et al., 2009](#); [Van Dokkum et al., 2010](#); [Newman et al., 2012](#)), which could explain the positive correlation detected between halo mass and galaxy size for high-mass galaxies. Similar reasoning was employed by [C17](#), where they explored the effects of a toy merger model on galaxy size and found a value of $\eta \approx 1$.

On the subject of satellites, [Shankar et al. \(2014\)](#) used a spectroscopic sample of ETGs from SDSS with redshifts $z < 0.3$. Specifically, they find no evidence of environmental dependence on the SMR within their sample of galaxies; however, all the hierarchical models considered in their study predict an environmental dependence. They find that, on average, with increasing host halo mass, $M_h < 3 \times 10^{12} M_{\odot}$ to $M_h \sim 10^{15} M_{\odot}$, the median galaxy size increases by a factor of $\sim 1.5 - 3$. In their analysis of satellite galaxies, they find that their models predict no environmental dependence with satellite galaxy SMR between the stellar masses of $10^{11} M_{\odot}$ to $10^{12} M_{\odot}$. These results are consistent with our result of $\eta_{\text{sat}} = 0.1 \pm 0.1$ in the largest stellar mass bin of $11.2 \leq \log_{10}(M_{\star}/M_{\odot}) \leq 11.78$. [George et al. \(2025\)](#) analyzes a sample of $\sim 86,000$ low-mass, $M_{\star} < 10^{9.5} M_{\odot}$, quiescent galaxies (QG) and finds that at a fixed stellar mass of $\log_{10}(M_{\star}/M_{\odot}) = 10.7$, QGs are smaller in denser environments than their corresponding counterparts in the field. It is reasonable to assume that the majority of their cluster/group galaxies are satellites, which reveals an anti-correlation between galaxy size and halo mass, which goes in the same sense as our result of $\eta_{\text{sat}} = -0.19 \pm 0.05$ for satellite galaxies with stellar masses within $10.5 \leq \log_{10}(M_{\star}/M_{\odot}) \leq 10.8$. For satellite galaxies, the most prominent mechanisms driving this environmental dependence are galaxy harassment (e.g. [Moore et al., 1996, 1999a](#)) and tidal stripping (e.g. [Merritt, 1985](#); [Read et al., 2006b](#)). The influence of these mechanisms depends on the infall time of the satellites. Notably, satellite galaxies with earlier infall times into more massive clusters have been shown to quench significantly earlier than galaxies with much later infall times (e.g. [Wetzel et al., 2013](#); [Oman & Hudson, 2016](#); [Smith et al., 2019](#)). This early quenching may result in their size growth being inhibited, leading to intrinsically smaller-sized galaxies.

Chapter 7

Conclusion

In this section, we justify and discuss potential improvements to the analysis done in this work. The modelling of the offset group term could be improved, as we observed that the offset group term peaks at different radii in the [DLIS](#) and simulated [ESD](#) vectors. Additionally, the results of this work may lead to some further studies regarding the environments that satellite galaxies live in. We conclude with a summary of the main results of this thesis.

7.1 Direction for Future Works

7.1.1 Revisiting Satellite Placement

The contribution of the satellite offset group term begins to dominate at intermediate scales, typically between $0.5 \text{ Mpc} \lesssim r_p \lesssim 1.5 \text{ Mpc}$. This term is unique to satellites, as they reside within a larger host halo. At scales beyond the extent of a satellite's halo, the lensing signal begins to include contributions from the host halo itself. The [DLIS ESD](#) measurements shown in [Figure 3.2](#) indicate that, for this galaxy sample, contributions from the offset group term dominate in the range of 0.4 to 5–6 Mpc. One way to model this within simulations is to include a method that alters the placement of satellites within the host halo (i.e., places satellites closer to the host halo centre or further away). The placement of satellites will, in theory, alter the radius at which the offset group term begins to contribute and where this contribution peaks. The amplitude of the peak is determined by the mass of the host halo. The `AbacusSummit` simulations have a method to alter the

placement of satellites in the form of the s parameter. The s parameter is defined as the following:

$$p_i = \frac{\langle N_s \rangle(M)}{N_p} \left[1 + s \left(1 - \frac{2 R_i}{N_p - 1} \right) \right], \quad (7.1)$$

where N_p is the number of particles in the halo, R_i is the rank of the particles ordered by their distance to the halo centre, and $\langle N_s \rangle$ is the average number of satellites. The variable, p_i , is the probability for particle i , with rank R_i , to host a satellite. According to this parametrization, $s < 0$ results in the satellite being placed closer to the host halo centre, and $s > 0$ results in satellites being placed further away from the host halo centre (see Figure C.1). As stated in Section 4.2, we had set $s = 0$, but we now consider the effect of varying s given our current set of best-fitting HOD parameters. Details of this work can be found in Appendix C. To summarize, we find that the best-fitting s value is -0.3 , indicating that satellite galaxies in our best-fit mock lens catalogue, and by inference in the DLIS sample, tend to be located closer, on average, to their host halo centres than the dark matter particles. We caution against over-interpreting these results, as we do not have error estimates for this best-fit value. The goal of this section is to investigate whether we are sensitive to s and if we see improvements to the fitting. Nevertheless, generating new fits to the DLIS ESD vectors yields χ^2_{joint} values of 17.67, 25.60, and 19.69 for our three stellar mass bins, respectively. Comparing these values to the joint χ^2 values in Table 5.3, we find that the fit to $\mathcal{M}_1^*$ becomes worse, quoting an increase in χ^2_{joint} of 2.9. The fits to $\mathcal{M}_2^*$ and $\mathcal{M}_3^*$ see a decrease in χ^2_{joint} by 14.74 and 1.12, leading to better fits. Although the fit to $\mathcal{M}_2^*$ improves significantly, qualitatively, some mismatch remains at the scales of the offset group term, where the model slightly over-predicts the ESD signal. This is likely because the best-fitting parameters found when $s = 0$ may no longer be optimal when s is allowed to vary. Nevertheless, introducing s as a free parameter improves the model's ability to capture the offset group term. Recalculating η_{cent} and η_{sat} values showed no significant discrepancies between the values reported in Chapter 6; see Figure C.4, which demonstrates the stability of the current model without s .

7.1.2 Satellite Galaxy: Subhalo Mass-Size Relation?

In this work, we utilized the AbacusSummit simulations whose halo catalogues do not identify subhalos for satellite galaxies. Thus, halo mass values assigned to satellite galaxies are, in essence, the host halo mass. Our results for η_{sat} should be interpreted as characterizing the correlation between host halo mass and galaxy size for satellite galaxies. An interesting direction for future work is to investigate the correlation between halo mass and galaxy size for satellite subhalos, as opposed to host halos, using GGL. This, however, is quite

non-trivial. Observationally, a large sample of satellite galaxies is required to achieve a weak lensing signal with high signal-to-noise before fitting a model to estimate subhalo masses. Nevertheless, there have been GGL studies that have provided constraints for satellite galaxy subhalo masses (e.g. Li et al., 2014; Sifón et al., 2015) and the stellar-to-subhalo mass relation (e.g. Sifón et al., 2018; Kumar & More, 2024), which demonstrates its feasibility.

In this work, we adopted a tNFW profile to model and probe information about satellite galaxy subhalo truncation via the $\log_{10} \tau$ parameter. This is, however, a simplified model which does not account for the fact that truncation of satellite halos is dependent on time (i.e., infall time) and cluster centre distance (e.g. Springel et al., 2001; Gao et al., 2004; van den Bosch et al., 2005; Xie & Gao, 2015; Niemiec et al., 2019). Using a hydrodynamical simulation with a low mass resolution, such as Illustris TNG50 (Nelson et al., 2019), would be beneficial in probing the smaller scales of subhalos as opposed to AbacusSummit. Modelling satellite galaxies accurately can be more complex than central galaxies due to environmentally driven processes that they undergo, in particular, tidal stripping. We know that generally, as satellite galaxies fall into a much larger and denser environment, they experience tidal stripping. Dark matter halos are greater in extent than the stellar component of galaxies, and so, the stripping process preferentially strips the dark matter halo first, leaving the majority of the stellar content intact (Smith et al., 2016; Lee et al., 2018). As the system evolves, both dark matter and stellar matter are continuously stripped, moving the galaxy to lower halo mass, stellar mass, and physical size. This rationale was used to justify the use of a tNFW profile for analytically adding the subhalo contribution of satellites to the ESD vector. Future work focused on satellite galaxies in particular would provide insight into how the subhalo mass-galaxy size relation evolves in comparison to the host halo mass-galaxy size relation.

7.1.3 Halo Concentration

In this work, we do not consider any correlation between halo concentration and galaxy size, opting for the simplest parametrization of our HOD model. Within N-body simulations, halos that have formed earlier are more compact and concentrated (see Bullock et al., 2001b). With this in consideration, a potential reconstruction of the halo mass-size model could be one that places smaller galaxies in more concentrated halos at fixed halo mass. Notably, it has been shown that at fixed halo ($M_{200} = 10^{12.2} M_{\odot}$) and stellar mass ($M_{\star} = 10^{10.7} M_{\odot}$), the halo concentration adopted alters the measured ESD profile within the virial radius. Lower values of halo concentration tend to decrease the lensing amplitude within the virial radius. Higher concentration values have the opposite effect (see

Fig. 7 of [Charlton et al., 2017](#)). Ultimately, this would increase the ESD of the smaller galaxy relative to the larger galaxy, despite both residing in halos of the same mass. The current modelling does not distinguish between these effects; therefore, the lensing signal attributed solely to halo mass could also include differences in halo concentration. Given the simulation-based HOD framework utilized here, this work can be readily extended to include halo concentration in future analyses.

7.2 Summary

In this work, we investigated the potential correlation between dark matter halo mass and galaxy size at fixed stellar mass and colour using the weak lensing technique of galaxy-galaxy lensing. The lens galaxy sample is drawn from the DESI Legacy Imaging Survey (DLIS), while the background source galaxy shapes are taken from ShapePipe v1.3 by the Ultraviolet Near Infrared Optical Northern Survey (UNIONS). We focused on red, early-type galaxies selected with cuts on colour ($g - z \geq 1.0$) and Sèrsic index ($n \geq 2.5$), covering a stellar mass range of $10.5 \leq \log_{10}(M_*/M_\odot) < 11.78$, which we divided into three bins. Within each stellar mass bin, we further split galaxies into ‘small’ and ‘large’ based on whether their effective radius falls below or above the median size at that stellar mass. Our analysis includes over 4×10^6 lens galaxies and 84×10^6 background shape measurements. Assuming a power-law relation between halo mass and galaxy size of the form $M_h(M_*) \propto r_{\text{eff}}^{\eta}(M_*)$, we modelled the lensing signal using a conditional stellar mass function (CSMF) halo occupation distribution (HOD) framework implemented within the AbacusSummit simulations. Emulators trained on mock weak lensing measurements were used to fit HOD parameters to the observational data. Using the resulting best-fit mock catalogue, we quantified the strength of the correlation between galaxy size and halo mass encoded in the observed weak lensing signals. The conclusions from this work for red, early-type central galaxies are as follows:

1. We find no significant correlation between halo mass and galaxy size for central galaxies with stellar masses in the range $10.5 \leq \log_{10}(M_*/M_\odot) < 11.2$.
2. In our highest stellar mass bin, $11.2 \leq \log_{10}(M_*/M_\odot) < 11.78$, we detect a positive correlation of $\eta_{\text{cent}, \mathcal{M}_3^*} = 0.51 \pm 0.14$. This result helps reconcile previous tensions between [C17](#) and [S19](#), indicating a true underlying correlation that lies between these two studies.

3. Fitting a linear model of η_{cent} as a function of stellar mass $M_{\star}$ to results from literature, we find a slope of $s_{\eta_{\text{cent}}} = 0.49 \pm 0.29$, indicating that the correlation between halo mass and galaxy size strengthens with increasing stellar mass.

For red, early-type satellite galaxies, we conclude the following:

1. For satellite galaxies, we detect a negative correlation of $\eta_{\text{sat}, \mathcal{M}_1^{\star}} = -0.19 \pm 0.05$ in the lowest stellar mass bin and $\eta_{\text{sat}, \mathcal{M}_2^{\star}} = -0.09 \pm 0.05$ in the intermediate mass bin, implying that smaller satellites tend to reside in more massive host halos. This trend becomes weakly positive in the highest bin, although uncertainties at high mass prevent any strong claims.
2. The correlation between host halo mass and satellite galaxy size strengthens with increasing stellar mass, characterized by a fitted slope of $s_{\eta_{\text{sat}}} = 0.41 \pm 0.15$.

References

- Abadi M. G., Moore B., Bower R. G., 1999, [MNRAS](#), **308**, 947
- Abdellah E., Samir R. M., Awad Z., Amin M. Y., 2025, [Ap&SS](#), **370**, 7
- Abdullah M. H., Abdelhamid N. E., Samir R. M., Wilson G., 2025, [arXiv e-prints](#), p. [arXiv:2507.04212](#)
- Aghanim N., et al., 2020, [A&A](#), **641**, A6
- Akrami Y., et al., 2020, [A&A](#), **641**, A7
- Angulo R. E., Hahn O., Ludlow A. D., Bonoli S., 2017, [MNRAS](#), **471**, 4687
- Ardila F., et al., 2020, [MNRAS](#), **500**, 432
- Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Giallongo E., 1999, [MNRAS](#), **310**, 540
- Baldry I. K., et al., 2012, [MNRAS](#), **421**, 621
- Balogh M. L., Morris S. L., Yee H. K. C., Carlberg R. G., Ellingson E., 1999, [ApJ](#), **527**, 54
- Balogh M. L., Pearce F. R., Bower R. G., Kay S. T., 2001, [MNRAS](#), **326**, 1228
- Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, [ApJ](#), **304**, 15
- Bartelmann M., Schneider P., 2001, [Phys. Rep.](#), **340**, 291
- Beck R., Dobos L., Budavári T., Szalay A. S., Csabai I., 2016, [MNRAS](#), **460**, 1371
- Bédorf J., Portegies Zwart S., 2013, [MNRAS](#), **431**, 767
- Behroozi P. S., Conroy C., Wechsler R. H., 2010a, [ApJ](#), **717**, 379

- Behroozi P. S., Conroy C., Wechsler R. H., 2010b, [ApJ](#), **717**, 379
- Behroozi P. S., Wechsler R. H., Conroy C., 2013, [ApJ](#), **770**, 57
- Bekki K., Couch W. J., Shioya Y., 2002, [ApJ](#), **577**, 651
- Bennett C. L., et al., 2003, [ApJS](#), **148**, 1
- Benson A. J., 2010, [Phys. Rep.](#), **495**, 33
- Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000, [MNRAS](#), **311**, 793
- Berlind A. A., Weinberg D. H., 2002, [ApJ](#), **575**, 587
- Bernardi M., Renzini A., da Costa L. N., Wegner G., Alonso M. V., Pellegrini P. S., Rit   C., Willmer C. N. A., 1998, [ApJ](#), **508**, L143–L146
- Bernardi M., Roche N., Shankar F., Sheth R. K., 2011, [MNRAS](#), **412**, L6
- Bertin E., 2011, in Evans I. N., Accomazzi A., Mink D. J., Rots A. H., eds, *Astronomical Society of the Pacific Conference Series Vol. 442, Astronomical Data Analysis Software and Systems XX*. p. 435
- Bertin E., Arnouts S., 1996, [A&AS](#), **117**, 393
- Bertschinger E., 1985, [ApJS](#), **58**, 39
- Bezanson R., van Dokkum P. G., Tal T., Marchesini D., Kriek M., Franx M., Coppi P., 2009, [ApJ](#), **697**, 1290
- Birnboim Y., Dekel A., 2003, [MNRAS](#), **345**, 349
- Blanton M. R., Moustakas J., 2009, [ARA&A](#), **47**, 159
- Blum R. D., et al., 2016, in *AAS Meeting Abstracts #228*. p. 317.01
- Bocquet S., Saro A., Dolag K., Mohr J. J., 2016, [MNRAS](#), **456**, 2361
- Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, [ApJ](#), **379**, 440
- Boselli A., Boissier S., Cortese L., Gavazzi G., 2009, [Astronomische Nachrichten](#), **330**, 904
- Boylan-Kolchin M., Ma C.-P., Quataert E., 2006, [MNRAS](#), **369**, 1081

Bradshaw C., Leauthaud A., Hearin A., Huang S., Behroozi P., 2020, [MNRAS](#), **493**, 337

Brooks A. M., Governato F., Quinn T., Brook C. B., Wadsley J., 2009, [ApJ](#), 694, 396–410

Brown T., et al., 2023, [ApJ](#), 956, 37

Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001a, [MNRAS](#), 321, 559

Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001b, [MNRAS](#), **321**, 559

Bullock J. S., Wechsler R. H., Somerville R. S., 2002, [MNRAS](#), **329**, 246

Burger P. A., Paillas E., Hudson M. J., 2024, [Phys. Rev. D](#), 110, 083517

Burger P. A., Patel D. A., Hudson M. J., 2025, [Phys. Rev. D](#), 111, 123515

Cabezas A., et al., 2024, [arXiv e-prints](#), p. [arXiv:2402.10797](#)

Cacciato M., van den Bosch F. C., More S., Li R., Mo H. J., Yang X., 2009a, [MNRAS](#), **394**, 929

Cacciato M., van den Bosch F. C., More S., Li R., Mo H. J., Yang X., 2009b, [MNRAS](#), **394**, 929

Cacciato M., Lahav O., van den Bosch F. C., Hoekstra H., Dekel A., 2012, [MNRAS](#), **426**, 566

Cacciato M., van den Bosch F. C., More S., Mo H., Yang X., 2013, [MNRAS](#), **430**, 767

Caon N., Capaccioli M., D’Onofrio M., 1993, [MNRAS](#), **265**, 1013

Carroll S. M., Press W. H., Turner E. L., 1992, [ARA&A](#), **30**, 499

Charlton P. J. L., Hudson M. J., Balogh M. L., Khatri S., 2017, [MNRAS](#), 472, 2367–2387

Chen A. Y., Afshordi N., 2020, [Physical Review D](#), 101

Chen Z., Gu Y., Zou H., Yuan Q., 2024, [ApJ](#), **961**, 253

Clavijo-Bohórquez W. E., de Gouveia Dal Pino E. M., Melioli C., 2024, [MNRAS](#), **535**, 1696

Conroy C., Wechsler R. H., 2009, [ApJ](#), **696**, 620

Conroy C., Wechsler R. H., Kravtsov A. V., 2006, [ApJ](#), **647**, 201

Contini E., Yi S. K., Jeon S., Rhee J., 2024, [ApJS](#), **274**, 41

Contreras S., Angulo R. E., Chaves-Montero J., White S. D. M., Aricò G., 2023, [MNRAS](#), **520**, 489

Cooper M. C., et al., 2012, [MNRAS](#), **419**, 3018

Correa C. A., Wyithe J. S. B., Schaye J., Duffy A. R., 2015, [MNRAS](#), **452**, 1217

Cypriano E. S., Sodré Jr. L., Campusano L. E., Dale D. A., Hardy E., 2006, [A.J.](#), **131**, 2417

D'Eugenio F., et al., 2021, [MNRAS](#), **504**, 5098

D'Onofrio M., Chiosi C., 2024, [A&A](#), **687**, A126

Daddi E., et al., 2005, [ApJ](#), **626**, 680

De Graaff A., et al., 2021, [ApJ](#), **913**, 103

De Vaucouleurs G., 1953, [MNRAS](#), **113**, 134

Dekel A., Birnboim Y., 2006, [MNRAS](#), **368**, 2

Dekel A., Silk J., 1986, [ApJ](#), **303**, 39

Desmond H., Katz H., Lelli F., McGaugh S., 2019, [MNRAS](#)

Dey A., et al., 2019, [A.J.](#), **157**, 168

Diemer B., 2018, [ApJS](#), **239**, 35

Diemer B., 2023, [MNRAS](#), **519**, 3292

Diemer B., Kravtsov A. V., 2015, [ApJ](#), **799**, 108

Djorgovski S., Davis M., 1987, [ApJ](#), **313**, 59

Dodelson S., 2017, Gravitational Lensing. Cambridge University Press, Cambridge, United Kingdom

Doroshkevich A. G., 1970, [Astrophysics](#), **6**, 320

- Dressler A., Lynden-Bell D., Burstein D., Davies R. L., Faber S. M., Terlevich R., Wegner G., 1987, [ApJ](#), **313**, 42
- Duffy A. R., Schaye J., Kay S. T., Dalla Vecchia C., 2008, [MNRAS: Letters](#), 390, L64–L68
- Dvornik A., et al., 2023, [A&A](#), **675**, A189
- Eddington A. S., 1913, [MNRAS](#), **73**, 359
- Efstathiou G., 2003, [arXiv e-prints](#), pp astro-ph/0303623
- Einasto J., 1965, Trudy Astrofizicheskogo Instituta Alma-Ata, **5**, 87
- Einstein A., 1911, [Annalen der Physik](#), **340**, 898
- Einstein A., de Sitter W., 1932, Contributions from the Mount Wilson Observatory, **3**, 51
- Fall S. M., Efstathiou G., 1980, [MNRAS](#), **193**, 189
- Fall S. M., Romanowsky A. J., 2013, [ApJ](#), **769**, L26
- Farrens S., et al., 2022, [A&A](#), **664**, A141
- Favole G., Montero-Dorta A. D., Prada F., Rodríguez-Torres S. A., Schlegel D. J., 2018, [MNRAS](#), **480**, 1415
- Figueira M., Siudek M., Pollo A., Krywult J., Vergani D., Bolzonella M., Cucciati O., Iovino A., 2024, [A&A](#), **687**, A117
- Fillmore J. A., Goldreich P., 1984, [ApJ](#), **281**, 9
- Friedmann A., 1922, [Zeitschrift fur Physik](#), **10**, 377
- Friedmann A., 1924, [Zeitschrift fur Physik](#), **21**, 326
- Gaibler V., Khochfar S., Krause M., Silk J., 2012, [MNRAS](#), **425**, 438
- Gao L., White S. D. M., Jenkins A., Stoehr F., Springel V., 2004, [MNRAS](#), **355**, 819
- Garrison L. H., Eisenstein D. J., Ferrer D., Maksimova N. A., Pinto P. A., 2021, [MNRAS](#), **508**, 575
- George A., et al., 2025, [ApJ](#), **987**, 45

Georgiou C., Chisari N. E., Bilicki M., La Barbera F., Napolitano N. R., Roy N., Tortora C., 2025, [A&A](#), **699**, [A252](#)

Ghosh A., Urry C. M., Powell M. C., Shimakawa R., van den Bosch F. C., Nagai D., Mitra K., Connolly A. J., 2024, [ApJ](#), **971**, [142](#)

Gillis B. R., et al., 2013, [MNRAS](#), **431**, [1439](#)

Governato F., Babul A., Quinn T., Tozzi P., Baugh C. M., Katz N., Lake G., 1999, [MNRAS](#), **307**, [949](#)

Guinot, Axel et al., 2022, [A&A](#), **666**, [A162](#)

Gunn J. E., Gott III J. R., 1972, [ApJ](#), **176**, [1](#)

Guth A. H., Pi S.-Y., 1982, [Phys. Rev. Lett.](#), **49**, [1110](#)

Gwyn S., et al., 2025, [arXiv e-prints](#), p. [arXiv:2503.13783](#)

Hadzhiyska B., Eisenstein D., Bose S., Garrison L. H., Maksimova N., 2022a, [MNRAS](#), **509**, [501](#)

Hadzhiyska B., Garrison L. H., Eisenstein D., Bose S., 2022b, [MNRAS](#), **509**, [2194](#)

Harrison E. R., 1970, [Phys. Rev. D](#), **1**, [2726](#)

Hayashi E., White S. D. M., 2008, [MNRAS](#), **388**, [2](#)

Hearin A. P., Zentner A. R., Berlind A. A., Newman J. A., 2013, [MNRAS](#), **433**, [659](#)

Hernquist L., 1990, [ApJ](#), **356**, [359](#)

Heymans C., et al., 2012, [MNRAS](#), **427**, [146](#)

Hildebrandt H., et al., 2012, [MNRAS](#), **421**, [2355](#)

Hilz M., Naab T., Ostriker J. P., 2013, [MNRAS](#), **429**, [2924](#)

Hinshaw G., et al., 2007, [ApJS](#), **170**, [288](#)

Hoyle F., 1949, in *Problems of Cosmical Aerodynamics*.

Huang S., et al., 2020, [MNRAS](#), **492**, [3685](#)

- Hubble E., 1929, [Proc. Natl. Acad. Sci. U.S.A.](#), **15**, 168
- Hudson M. J., et al., 2015, [MNRAS](#), **447**, 298
- Huertas-Company M., et al., 2013a, [MNRAS](#), **428**, 1715
- Huertas-Company M., Shankar F., Mei S., Bernardi M., Aguerri J. A. L., Meert A., Vikram V., 2013b, [ApJ](#), **779**, 29
- Huff E., Mandelbaum R., 2017, [arXiv e-prints](#), p. [arXiv:1702.02600](#)
- Hyde J. B., Bernardi M., 2009a, [MNRAS](#), **394**, 1978
- Hyde J. B., Bernardi M., 2009b, [MNRAS](#), **396**, 1171
- Ilbert O., et al., 2006, [A&A](#), **457**, 841
- Ishiyama T., et al., 2021, [MNRAS](#), **506**, 4210
- Jarvis M., Bernstein G., Jain B., 2004, [MNRAS](#), **352**, 338
- Jeans J. H., 1902, [Philos. Trans. R. Soc. Lond. A](#), **199**, 1
- Jørgensen I., Franx M., Kjaergaard P., 1996, [MNRAS](#), **280**, 167
- Katz N., Keres D., Dave R., Weinberg D. H., 2003, in Rosenberg J. L., Putman M. E., eds, *Ap&SSVol. 281, The IGM/Galaxy Connection. The Distribution of Baryons at z=0*. p. 185 ([arXiv:astro-ph/0209279](#)), [doi:10.1007/978-94-010-0115-1_34](#)
- Kennicutt Jr. R. C., 1989, [ApJ](#), **344**, 685
- Kennicutt Jr. R. C., 1998, [ApJ](#), **498**, 541–552
- Keres D., Katz N., Weinberg D. H., Dave R., 2005, [MNRAS](#), **363**, 2–28
- Klypin A. A., Trujillo-Gomez S., Primack J., 2011, [ApJ](#), **740**, 102
- Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, [MNRAS](#), **457**, 4340–4359
- Kocjan Z., Cadiou C., Agertz O., Pontzen A., 2024, [MNRAS](#), **534**, 918
- Kravtsov A. V., 2013, [ApJ](#), **764**, L31
- Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, [ApJ](#), **609**, 35

- Kravtsov A. V., Vikhlinin A. A., Meshcheryakov A. V., 2018, [Astronomy Letters](#), **44**, 8
- Kumar A., More S., 2024, [arXiv e-prints](#), p. [arXiv:2409.05795](#)
- Lang D., Hogg D. W., Mykytyn D., 2016, The Tractor: Probabilistic astronomical source detection and measurement, Astrophysics Source Code Library, record ascl:1604.008 (Astrophysics Source Code Library:1604.008)
- Lange J., Huang S., 2022, dsigma: Galaxy-galaxy lensing Python package, Astrophysics Source Code Library, record ascl:2204.006
- Lani C., et al., 2013, [MNRAS](#), **435**, 207
- Larson R. B., Tinsley B. M., Caldwell C. N., 1980, [ApJ](#), **237**, 692
- Le Fèvre O., et al., 2005, [A&A](#), **439**, 845
- Leauthaud A., Tinker J., Behroozi P. S., Busha M. T., Wechsler R. H., 2011, [ApJ](#), **738**, 45
- Lee C. T., Primack J. R., Behroozi P., Rodríguez-Puebla A., Hellinger D., Dekel A., 2018, [MNRAS](#), **481**, 4038
- Lemaître G., 1933, Annales de la Société Scientifique de Bruxelles, **53**, 51
- Li R., et al., 2014, [MNRAS](#), **438**, 2864
- Li Q., et al., 2024, [ApJ](#), **969**, L25
- Liaudat T., Bonnín J., Starck J. L., Schmitz M. A., Guinot A., Kilbinger M., Gwyn S. D. J., 2021, [A&A](#), **646**, A27
- López-Sanjuan C., et al., 2012, [A&A](#), **548**, A7
- MacCrann N., et al., 2022, [MNRAS](#), **509**, 3371
- Maksimova N. A., Garrison L. H., Eisenstein D. J., Hadzhiyska B., Bose S., Satterthwaite T. P., 2021, [MNRAS](#), **508**, 4017
- Mandelbaum R., Tasitsiomi A., Seljak U., Kravtsov A. V., Wechsler R. H., 2005, [MNRAS](#), **362**, 1451
- Mandelbaum R., Seljak U., Hirata C. M., 2008, [JCAP](#), **2008**, 006
- Mandelbaum R., et al., 2018, [PASJ](#), **70**, S25

Marinoni C., Hudson M. J., 2002, [ApJ](#), **569**, 101

Masaki S., Kashino D., Ishikawa S., Lin Y.-T., 2023, [MNRAS](#), **523**, 5280

McKee C. F., Ostriker J. P., 1977, [ApJ](#), **218**, 148

Merritt D., 1984, [ApJ](#), **276**, 26

Merritt D., 1985, [ApJ](#), **289**, 18

Mo H., van den Bosch F., White S., 2010, *Galaxy Formation and Evolution*. Cambridge University Press, [doi:10.1017/CBO9780511807244](https://doi.org/10.1017/CBO9780511807244)

Mobasher B., et al., 2015, [ApJ](#), **808**, 101

Moore B., Katz N., Lake G., Dressler A., Oemler A., 1996, [Nature](#), **379**, 613

Moore B., Lake G., Quinn T., Stadel J., 1999a, [MNRAS](#), **304**, 465

Moore B., Quinn T., Governato F., Stadel J., Lake G., 1999b, [MNRAS](#), **310**, 1147

More S., van den Bosch F. C., Cacciato M., More A., Mo H., Yang X., 2013, [MNRAS](#), **430**, 747

Moster B. P., Naab T., White S. D. M., 2013, [MNRAS](#), **428**, 3121

Moustakas J., et al., 2013, [ApJ](#), **767**, 50

Naab T., Ostriker J. P., 2017, [ARA&A](#), **55**, 59

Naab T., Johansson P. H., Ostriker J. P., 2009, [ApJ](#), **699**, L178

Nair P. B., van den Bergh S., Abraham R. G., 2010, [ApJ](#), **715**, 606

Navarro J. F., Frenk C. S., White S. D. M., 1996, [ApJ](#), **462**, 563

Nelson D., et al., 2019, [Comput. Astrophys. Cosmol.](#), **6**, 2

Newman A. B., Ellis R. S., Bundy K., Treu T., 2012, [ApJ](#), **746**, 162

Newman J. A., et al., 2013, [ApJS](#), **208**, 5

Niemiec A., et al., 2017, [MNRAS](#), **471**, 1153

Niemiec A., Jullo E., Giocoli C., Limousin M., Jauzac M., 2019, [MNRAS](#), **487**, 653

- Oey M., 2004, [Ap&SS](#), 289, 269–277
- Ogiya G., Nagai D., Ishiyama T., 2016, [MNRAS](#), 461, 3385
- Oku Y., Tomida K., Nagamine K., Shimizu I., Cen R., 2022, [ApJS](#), 262, 9
- Oman K. A., Hudson M. J., 2016, [MNRAS](#), 463, 3083
- Oogi T., Habe A., Ishiyama T., 2016, [MNRAS](#), 456, 300
- Ostriker E. C., Shetty R., 2011, [ApJ](#), 731, 41
- Paulino-Afonso A., et al., 2022, [A&A](#), 662, A86
- Peacock J. A., Smith R. E., 2000, [MNRAS](#), 318, 1144
- Peebles P. J. E., 1965, [ApJ](#), 142, 1317
- Peebles P. J. E., 1969, [ApJ](#), 155, 393
- Peebles P. J. E., 1980, The large-scale structure of the universe. Princeton series in physics, Princeton University Press, Princeton, N.J
- Press W. H., Schechter P., 1974, [ApJ](#), 187, 425
- Prugniel P., Simien F., 1996, [A&A](#), 309, 749
- Raichoor A., et al., 2012, [ApJ](#), 745, 130
- Read J. I., Wilkinson M. I., Evans N. W., Gilmore G., Kleyna J. T., 2006a, [MNRAS](#), 366, 429
- Read J. I., Wilkinson M. I., Evans N. W., Gilmore G., Kleyna J. T., 2006b, [MNRAS](#), 366, 429
- Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, [ApJ](#), 771, 30
- Rees M. J., Ostriker J. P., 1977, [MNRAS](#), 179, 541
- Rettura A., et al., 2010, [ApJ](#), 709, 512
- Robertson H. P., 1935, [ApJ](#), 82, 284
- Robles S., Gómez J. S., Ramírez Rivera A., González J. A., Padilla N. D., Dujovne D., 2019, [arXiv e-prints](#), p. [arXiv:1906.09382](#)

- Roy N., et al., 2018, [MNRAS](#), 480, 1057
- Samir R. M., Takey A., Shaker A. A., 2020, [Ap&SS](#), 365, 142
- Saracco P., Gargiulo A., Ciocca F., Marchesini D., 2017, [A&A](#), 597, A122
- Schechter P., 1976, [ApJ](#), 203, 297
- Schmidt M., 1959, [ApJ](#), 129, 243
- Scodeggio M., et al., 2018, [A&A](#), 609, A84
- Shankar F., Marulli F., Bernardi M., Mei S., Meert A., Vikram V., 2013, [MNRAS](#), 428, 109
- Shankar F., et al., 2014, [MNRAS](#), 439, 3189
- Shao M. J., Anbajagane D., Chang C., 2023, [MNRAS](#), 523, 3258
- Sheldon E., 2015, NGMIX: Gaussian mixture models for 2D images, Astrophysics Source Code Library, record ascl:1508.008
- Sheth R. K., Tormen G., 1999, [MNRAS](#), 308, 119
- Sheth R. K., Mo H. J., Tormen G., 2001, [MNRAS](#), 323, 1
- Sifón C., et al., 2015, [MNRAS](#), 454, 3938
- Sifón C., Herbonnet R., Hoekstra H., van der Burg R. F. J., Viola M., 2018, [MNRAS](#), 478, 1244
- Silk J., 1977, [ApJ](#), 211, 638
- Silk J., A. D., I. D., 2014, [Proc. Int. School Phys. “Enrico Fermi”](#), 186, 137–187
- Silva D. R., et al., 2016, in AAS Meeting Abstracts #228. p. 317.02
- Singh S., Mandelbaum R., Seljak U., Slosar A., Vazquez Gonzalez J., 2017, [MNRAS](#), 471, 3827
- Skibba R. A., Sheth R. K., 2009, [MNRAS](#), 392, 1080
- Smith R., Choi H., Lee J., Rhee J., Sanchez-Janssen R., Yi S. K., 2016, [ApJ](#), 833, 109

Smith R., Pacifici C., Pasquali A., Calderón-Castillo P., 2019, [ApJ](#), **876**, 145

Smoot G. F., et al., 1992, [ApJ](#), **396**, L1

Song M., et al., 2016, [ApJ](#), **825**, 5

Sonnenfeld A., Leauthaud A., 2018, [MNRAS](#), **477**, 5460–5476

Sonnenfeld A., Wang W., Bahcall N., 2019, [A&A](#), **622**, A30

Sonnenfeld A., et al., 2022, [A&A](#), **662**, A55

Spitzer Jr. L., 1990, [ARA&A](#), **28**, 71

Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, [MNRAS](#), **328**, 726

Springel V., et al., 2005, [Nature](#), **435**, 629

Strateva I., et al., 2001, [A.J.](#), **122**, 1861

Sunyaev R. A., Zeldovich Y. B., 1970, [Ap&SS](#), **7**, 3

Tasitsiomi A., Kravtsov A. V., Wechsler R. H., Primack J. R., 2004, [ApJ](#), **614**, 533

Tavio H., Cuesta A. J., Prada F., Klypin A. A., Sanchez-Conde M. A., 2008, The dark outside: the density profile of dark matter haloes beyond the virial radius ([arXiv:0807.3027](#)), <https://arxiv.org/abs/0807.3027>

Taylor E. N., et al., 2020, [MNRAS](#), **499**, 2896

Thacker R. J., Couchman H. M. P., 2000, [ApJ](#), **545**, 728

Tinker J. L., Conroy C., Norberg P., Patiri S. G., Weinberg D. H., Warren M. S., 2008a, [ApJ](#), **686**, 53

Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008b, [ApJ](#), **688**, 709

Tomczak A. R., et al., 2014, [ApJ](#), **783**, 85

Toomre A., Toomre J., 1972, [ApJ](#), **178**, 623

Trujillo-Gomez S., Klypin A., Primack J., Romanowsky A. J., 2011, [ApJ](#), **742**, 16

- Trujillo I., Cenarro A. J., de Lorenzo-Cáceres A., Vazdekis A., de la Rosa I. G., Cava A., 2009, [ApJ](#), **692**, L118
- Vale A., Ostriker J. P., 2004, [MNRAS](#), **353**, 189
- Van Dokkum P. G., et al., 2008, [ApJ](#), **677**, L5
- Van Dokkum P. G., et al., 2010, [ApJ](#), **709**, 1018
- Van Uitert E., et al., 2016, [MNRAS](#), **459**, 3251
- Van de Voort F., Schaye J., Booth C. M., Haas M. R., Dalla Vecchia C., 2011, [MNRAS](#), **414**, 2458
- Van der Wel A., Holden B. P., Zirm A. W., Franx M., Rettura A., Illingworth G. D., Ford H. C., 2008, [ApJ](#), **688**, 48
- Van der Wel A., et al., 2014, [ApJ](#), **788**, 28
- Vollmer B., 2013, in Oswalt T. D., Keel W. C., eds, , Vol. 6, Planets, Stars and Stellar Systems. Volume 6: Extragalactic Astronomy and Cosmology. p. 207, [doi:10.1007/978-94-007-5609-0_5](#)
- Walker A. G., 1937, [Proc. London Math. Soc.](#), **42**, 90
- Walsh D., Carswell R. F., Weymann R. J., 1979, [Nature](#), **279**, 381
- Wang C., et al., 2024, [MNRAS](#), **528**, 2728
- Wetzel A. R., Tinker J. L., Conroy C., van den Bosch F. C., 2013, [MNRAS](#), **432**, 336
- White S. D. M., 1984, [ApJ](#), **286**, 38
- White S. D. M., Zaritsky D., 1992, [ApJ](#), **394**, 1
- White M., Hernquist L., Springel V., 2001, [ApJ](#), **550**, L129
- Willman B., Strader J., 2012, [A.J.](#), **144**, 76
- Wright A. H., et al., 2017, [MNRAS](#), **470**, 283
- Wright A. H., Hildebrandt H., van den Busch J. L., Heymans C., 2020, [A&A](#), **637**, A100
- Xie L., Gao L., 2015, [MNRAS](#), **454**, 1697

- Yang X., Mo H. J., van den Bosch F. C., 2003, [MNRAS](#), **339**, 1057
- Yang X., Mo H. J., van den Bosch F. C., 2008, [ApJ](#), **676**, 248
- Yang X., Mo H. J., van den Bosch F. C., 2009, [ApJ](#), **695**, 900
- Yang L., Silverman J., Oguri M., Ding X., Toba Y., Huang S., Kawinwanichakij L., 2024, [MNRAS](#), **531**, 4006
- Yoon Y., Im M., Kim J.-W., 2017, [ApJ](#), **834**, 73
- Yoshida N., Stoeckl F., Springel V., White S. D. M., 2002, [MNRAS](#), **335**, 762
- Zehavi I., et al., 2011, [ApJ](#), **736**, 59
- Zel'dovich Y. B., Rakhmatulina A. K., Syunyaev R. A., 1972, [Radiophysics and Quantum Electronics](#), **15**, 121
- Zhao D., Aragón-Salamanca A., Conselice C. J., 2015, [MNRAS](#), **453**, 4444
- Zheng Z., et al., 2005, [ApJ](#), **633**, 791
- Zinn P. C., Middelberg E., Norris R. P., Dettmar R. J., 2013, [ApJ](#), **774**, 66
- Zou H., et al., 2017, [A.J.](#), **153**, 276
- Zou H., Gao J., Zhou X., Kong X., 2019, [ApJS](#), **242**, 8
- Zou H., et al., 2022, [Research in Astronomy and Astrophysics](#), **22**, 065001
- Zu Y., Mandelbaum R., 2015, [MNRAS](#), **454**, 1161
- Zuntz J., et al., 2018, [MNRAS](#), **481**, 1149
- van den Bosch F. C., Tormen G., Giocoli C., 2005, [MNRAS](#), **359**, 1029

APPENDICES

Appendix A

Additional Information on Emulator Construction

A.1 Emulator Training Space

In this section, we outline the iterative process used to define the final parameter space from which mock galaxy catalogues and mock [ESD](#) measurements were generated. The emulators used in this work are trained on training features defined by this parameter space.

In [B25](#), the parameter space was defined according to three criteria regarding the χ^2 of the [SMF](#) of red galaxies from the [DESI](#) early data release, the satellite fractions in each stellar mass bin, and finally the availability of particles used to distribute satellites. In this work, we consider only the third criterion, which addresses a storage issue in which no more than 0.3% of the available particles are allowed to be used to ensure that any modulation of satellites and assembly bias is possible. To define our final parameter space, we followed the following iterative process. Firstly, we built a Latin-hyperspace (LHS) with uniform priors for all parameters: M_0 , M_1 , γ_1 , γ_2 , σ_c , a_1 , b_0 , B_{cen} , $\log_{10}\tau$. The criterion regarding particle usage formed the first initial cut made on the otherwise uniformly sampled parameter space. From the surviving [HOD](#) parameters, a mock galaxy catalogue was generated for each. We then calculated the [ESD](#) mock measurements from these catalogues and built three emulators, one for each stellar mass bin. With these initial emulators, a [HMC](#) routine using [Blackjax](#) ([Cabezas et al., 2024](#), ¹) was run. We saw that the emulators were split

¹<https://blackjax-devs.github.io/blackjax/>

between two regions of parameter space. One region in which $M_0 < 10$ and $M_1 < 10$, and another, more predominant region where $M_0 > 10$ and $M_1 > 10$. We found that for values of M_0 and $M_1 < 10$, the satellite fractions were not realistic. Additionally, [Dvornik et al. \(2023\)](#) find values of $M_0 > 10$ and $M_1 > 10$. Since we aim to just fit our full bin [ESD](#) vectors jointly, we generated a new parameter space by randomly sampling points from the [HMC](#) chain to ensure our emulators become more accurate around the region of interest where we expect to have a good fit. To supplement the description of this iterative process, [Figure A.1](#) shows the final parameter space that the emulators were trained on. Additionally, as shown in [Figure 4.1](#), the emulator mocks from this final parameter space scatter appreciably above and below the [DLIS ESD](#) vectors in all three stellar mass bins. This further justifies the iterative process described as acceptable.

A.2 Emulator Parameter Visualization

To better understand the influence of individual training parameters on the output [ESD](#) vector of our emulators, we visualize this effect in a rainbow plot using the $\Delta\Sigma_{\text{mocks}}$ emulator for our lowest stellar mass bin, $\mathcal{M}_1^*$, as shown in [Figure A.2](#). The black line is the [ESD](#) vector using the best-fit [HOD](#) parameters in [Table 5.1](#). The colour gradient from blue to red indicates an increasing value of the [HOD](#) parameter being varied. Similarly, [Figure A.3](#) illustrates the effect that varying $\log_{10}\tau$ has on the [ESD](#) vector using the $\Delta\Sigma_{\text{tNFW}}f_{\text{sat}}$ emulator for $\mathcal{M}_1^*$.

A.3 Emulator Correlation Matrices and Accuracy Plots

As described in [Section 4.3](#), we have 1998, 1998, 1977 [ESD](#) mocks, respectively, for each stellar mass bin that we utilized to train our emulators. We randomly select 200 mocks from a sample of 500 lowest χ^2 mocks to be our test sample. The 500 mocks are, in essence, the 500 best-fitting mocks to our [DLIS](#) measurements. In order to gauge the error on the emulators and generate a covariance matrix, we are interested in quantifying the difference between the emulator predictions versus the truth. We define Δf , which is the difference between the emulator output and the truth given a set of [HOD](#) parameters, $\Delta\Sigma_{\text{emu}} - \Delta\Sigma_{\text{true}}$. Given our test sample of 200 mocks, we have 200 Δf vectors. The resulting correlation matrix is shown in [Figure A.4](#) for the three $\Delta\Sigma_{\text{mocks}}$ emulators, one for each stellar mass bin. The construction of the $\Delta\Sigma_{\text{tNFW}}f_{\text{sat}}$ emulators follows a process similar to that of the $\Delta\Sigma_{\text{mocks}}$ emulators. The accuracy plots for these emulators are shown in [Figure A.5](#).

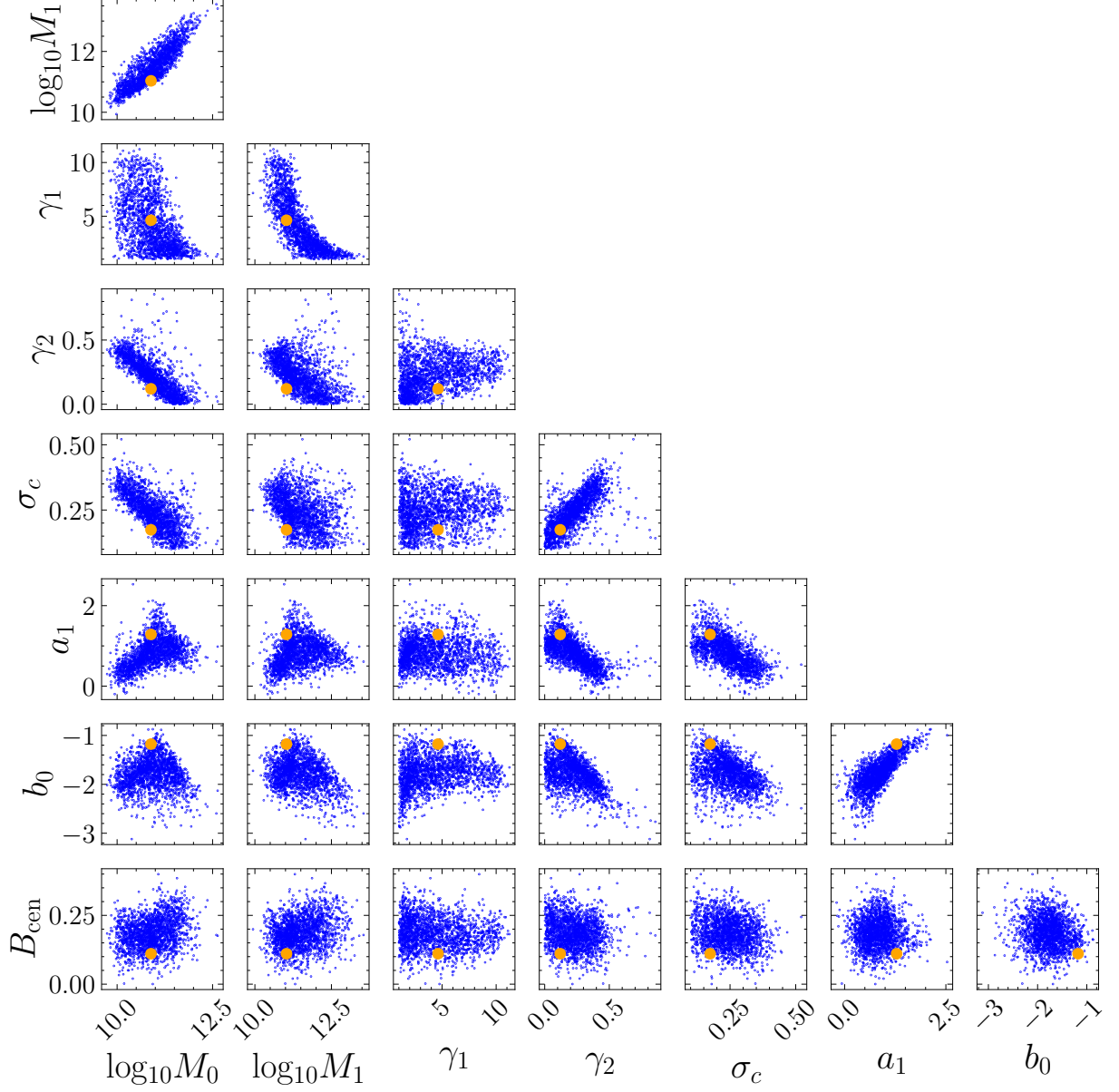


Figure A.1: Visualization of the parameter space in which the $\Delta\Sigma_{\text{mocks}}$ emulators were trained. Plotted in blue are the points in parameter space for $\mathcal{M}_3^*$ in which a mock lens catalogue was generated using **AbacusSummit**, and **ESD** measurements were calculated. The orange data point is where the best-fit **HOD** parameters, shown in Table 5.1, reside.

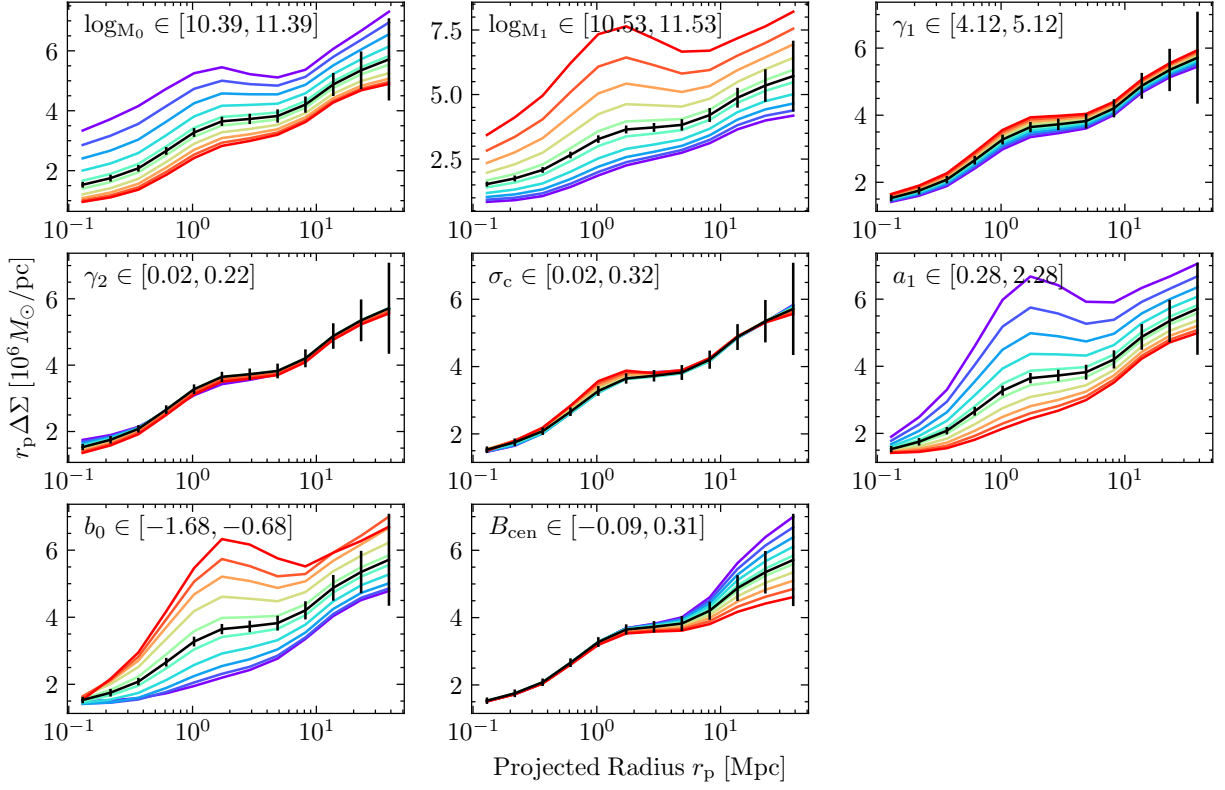


Figure A.2: Illustration of the effect of each HOD parameter on the predicted ESD vector from the emulator. The data vectors shown are from the $\Delta\Sigma_{\text{mocks}}$ emulator for $\mathcal{M}_1^*$. The black line represents the ESD vector computed using the best-fit HOD parameters listed in Table 5.1. The error bars are the sum of the measurement error and the emulator error from the $\Delta\Sigma_{\text{mocks}}$ and $\Delta\Sigma_{\text{tNFW}} f_{\text{sat}}$ emulators. The coloured lines, ranging from blue to red, show the effect of varying one HOD parameter at a time. Each parameter is varied symmetrically around its best-fit value while all other parameters are held fixed. Bluer colours are values less than the best fit, and red colours are greater than the best-fit value.

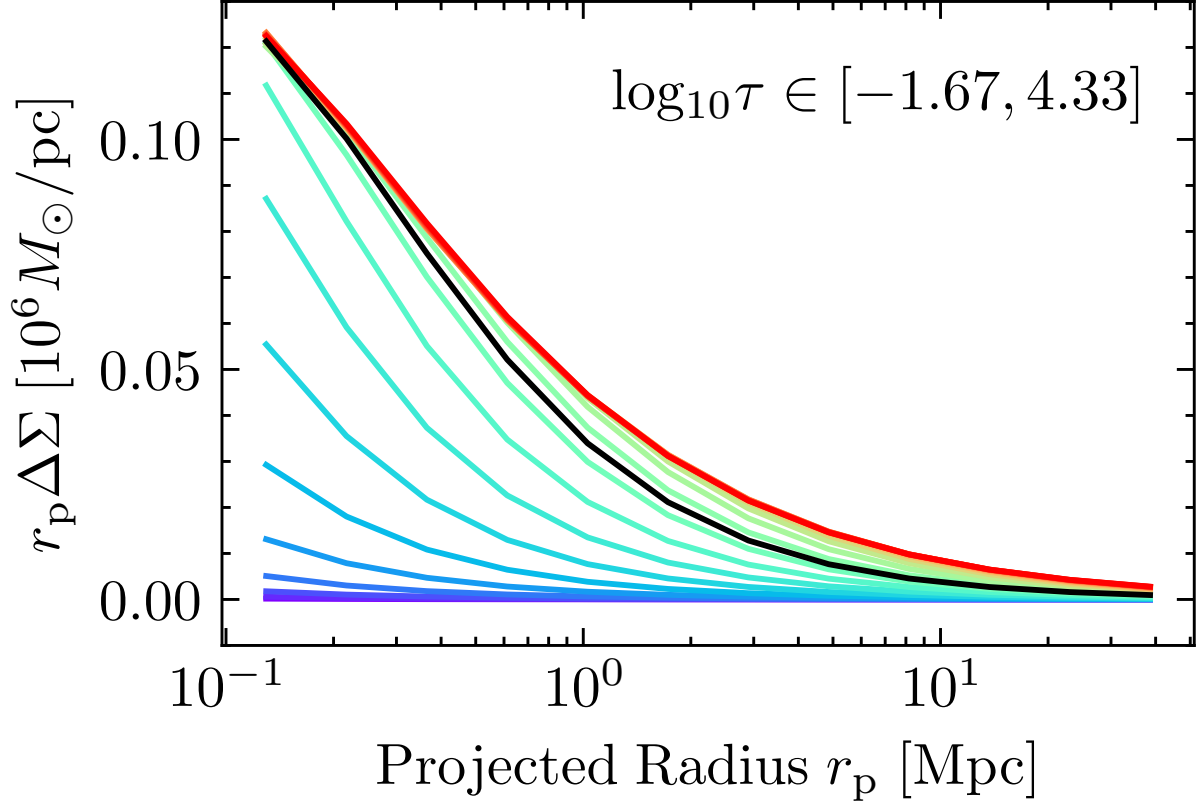


Figure A.3: Similar rainbow plot as Figure A.2, but this time showing the effect of the $\log_{10}\tau$ parameter. The predicted vectors shown are from the $\Delta\Sigma_{\text{tNFW}}f_{\text{sat}}$ emulator for $\mathcal{M}_1^*$. Error bars are not plotted, as they are significantly larger in scale than the data vectors plotted.

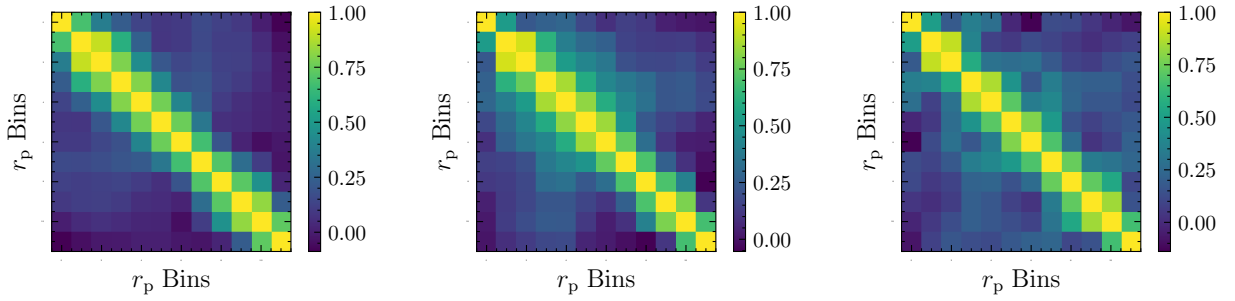


Figure A.4: Correlation matrix for the $\Delta\Sigma_{\text{mocks}}$ emulators.

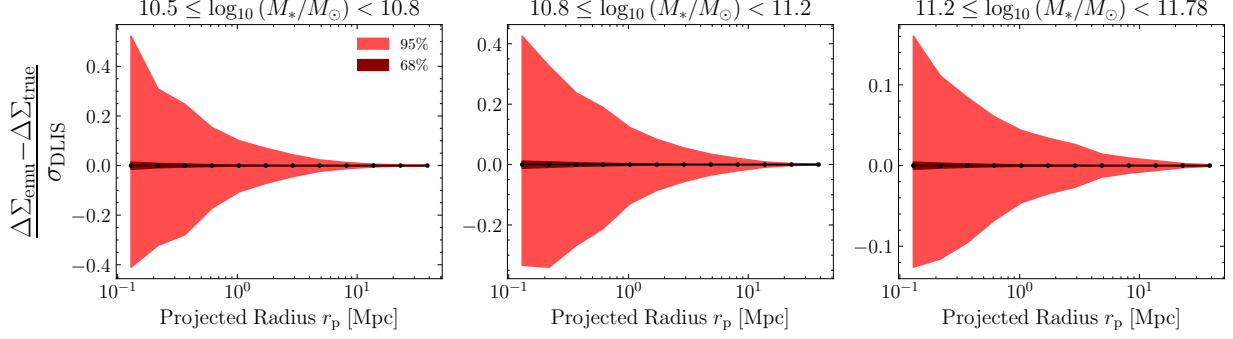


Figure A.5: Accuracy plots for the $\Delta\Sigma_{\text{tNFW}}f_{\text{sat}}$ emulators, similar to Fig. 4.2, display a clear difference between the 68% and 95% regions. The 68% band remains tightly clustered, which demonstrates that almost all emulator predictions are very close to the true values. The 95% region is much wider, which indicates that incorrect predictions are infrequent but can be quite large when they happen.

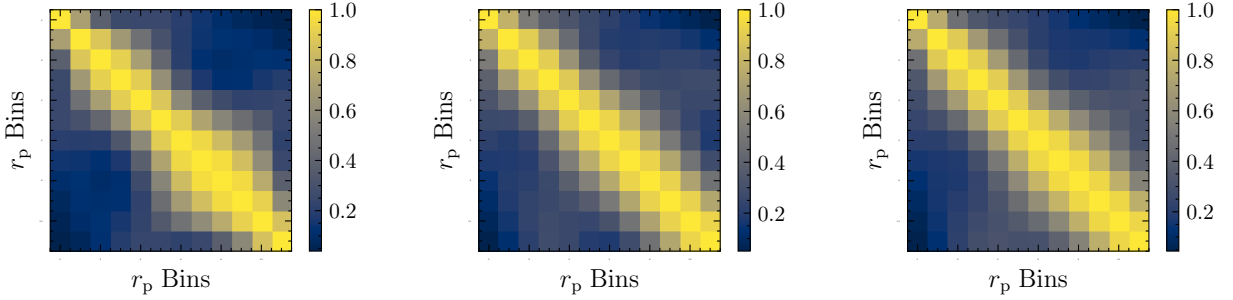


Figure A.6: Correlation matrix for the $\Delta\Sigma_{\text{tNFW}}f_{\text{sat}}$ emulators.

Given the large number of mocks available for training, these emulators are very accurate. We find they are accurate to less than 1% of the 1σ DLIS measurement error. From the entire sample of mocks available for training, 20% was set aside to be the test sample, and the remaining 80% was used for training each emulator. This resulted in 85,060, 83,660, and 79,220 Δf vectors, from which the resulting correlation matrix is shown in Figure A.6. The $\Delta\Sigma_{\text{tNFW}}f_{\text{sat}}$ emulators show much greater cross-correlations between bins further away in projected radii when compared to the $\Delta\Sigma_{\text{mocks}}$ emulator correlation matrices.

Appendix B

Behaviour of β

Figure B.1 and Figure B.2 illustrate the effect of varying β_{cent} and β_{sat} on the *small* and *large* galaxy ESD vectors, respectively. We remind the reader that $\beta_{\text{cent/satellite}} = 0$ indicates no correlation between galaxy size and halo mass, $\beta_{\text{cent/satellite}} < 0$ indicates an anti-correlation (i.e., large galaxies reside in less massive halos, and vice versa), and $\beta_{\text{cent/satellite}} > 0$ indicates a positive correlation such that large galaxies reside in more massive halos (and vice versa).

First, we discuss Figure B.1 when β_{cent} varies from -0.15 (blue) to 0.13 (red) and $\beta_{\text{sat}} = 0$. When $\beta_{\text{cent}} < 0$, the required behaviour is that smaller galaxies would populate more massive halos. This translates to the small galaxy ESD vectors having greater lensing signals than the large galaxy ESD vectors. The opposite is true when $\beta_{\text{cent}} > 0$. This is, in fact, the behaviour we see across all three stellar mass bins. More interestingly, we see that the sensitivity to β_{cent} is not the same across all three stellar mass bins. We find little sensitivity in $\mathcal{M}_1^*$, moderate sensitivity in $\mathcal{M}_2^*$, and the greatest sensitivity in $\mathcal{M}_3^*$. This trend is consistent with the HMC posteriors shown in Figure 5.3 and explains the lack of constraining power in $\mathcal{M}_1^*$. Physically, the interpretation of this can be the following. The halos in which low-mass central galaxies reside do not vary significantly in mass. Thus, when galaxies are assigned into small- and large-size subsamples, the mean halo mass, although different, is not dramatically different between the two. For low-mass galaxies, $10.5 \leq \log_{10}(M_*/M_\odot) < 10.8$, variations in central-type galaxies and their halos are not a significant factor in explaining the weak-lensing differences between the small and large galaxy subsamples. This is consistent with our quoted result for η_{cent} in $\mathcal{M}_1^*$ of 0.007 ± 0.24 (see Table 5.3).

Now, we discuss Figure B.2, where $\beta_{\text{cent}} = 0$, and β_{sat} varies from -0.15 (blue) to 0.13

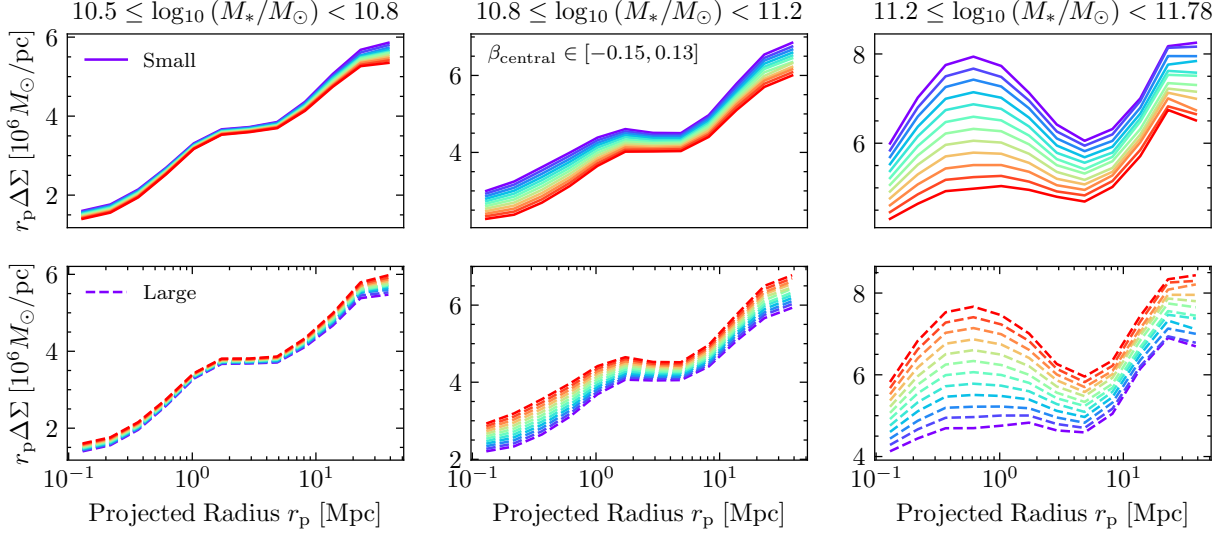


Figure B.1: Illustration of the effect of varying β_{cent} and fixing β_{sat} on the *small* galaxy and *large* galaxy $r_p \Delta \Sigma$ vectors for all three stellar mass bins. Solid coloured lines in the top row represent the *small* galaxy ESD vector. Dashed coloured lines in the bottom row represent the *large* galaxy ESD vector. β_{cent} varies from -0.15 (blue) to 0.13 (red), with $\beta_{\text{sat}} = 0$.

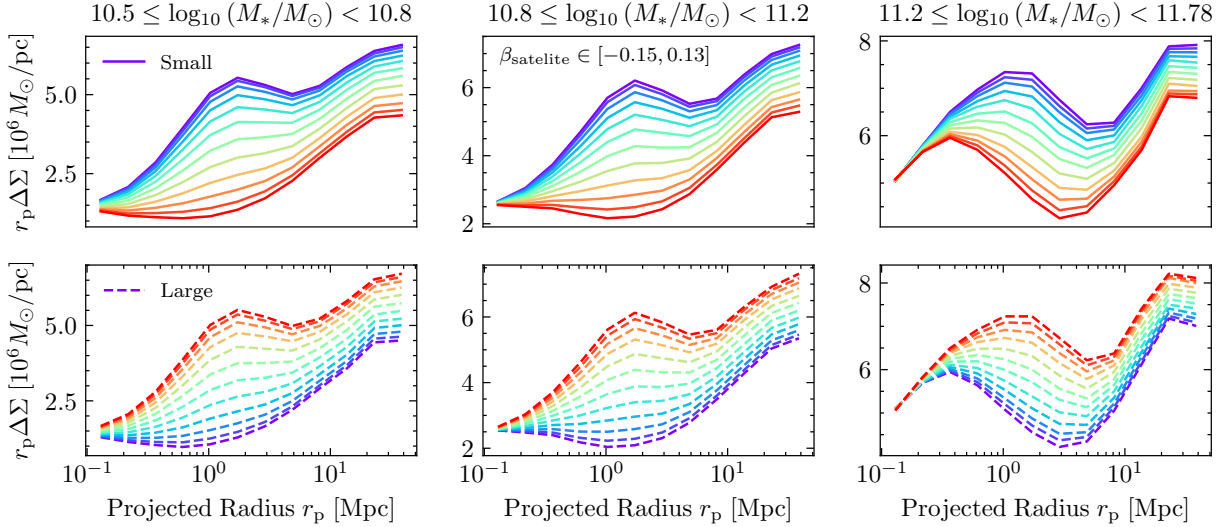


Figure B.2: Similar plot to Figure B.1, but now fixing $\beta_{\text{cent}} = 0$ and varying β_{sat} from -0.15 (blue) to 0.13 (red).

(red). Similar to the previous discussion regarding the sign of β_{cent} , the same effect on the resulting [ESD](#) vectors is shown across all three stellar mass bins when varying $\beta_{\text{sat}} < 0$ and $\beta_{\text{sat}} > 0$. Unlike β_{cent} , we are sensitive to β_{sat} in all three stellar mass bins. Notably, varying β_{sat} directly impacts the offset-group term contribution. The offset-group term is unique to satellites and contributes to the overall lensing profile at intermediate scales as the mass of the host halo begins to be included. When $\beta_{\text{sat}} = -0.15$ (purple lines), we see that the offset-group term contribution is the largest in the *small* galaxies [ESD](#) vector and negligible in the *large* [ESD](#) vector. At the other extreme, when $\beta_{\text{sat}} = 0.13$ (red lines), the effect is reversed, as expected. Interestingly, unique to our highest stellar mass bin, varying β_{sat} appears to shift the peak of the lensing signal. Taking $\beta_{\text{sat}} = -0.15$ (purple lines) for example, we see that the lensing signal for the small galaxy subsample [ESD](#) vector peaks at ~ 1 Mpc, and the large subsample peaks at ~ 0.35 Mpc. The satellite fraction for our best-fit mock in the stellar mass range of $11.2 \leq \log_{10}(M_{\star}/M_{\odot}) < 11.78$ is 4.4%; thus, the most likely explanation for this behaviour is that the satellite fraction in the large galaxy subsample is low such that there is no offset-group term contribution and the peak seen is simply the 1-halo term contribution.

Ultimately, we find that the sign of β controls the direction of the correlation, which leads to a qualitatively symmetric effect on the resulting [ESD](#) vectors for the small and large galaxy subsamples. The magnitude of β_{cent} and β_{sat} affects the [ESD](#) vectors differently and is strongly dependent on the stellar mass bin.

Appendix C

Varying the s Parameter

As previously noted, at projected separations between approximately 0.5 Mpc and 1.5 Mpc, the satellite offset group term begins to contribute to the overall [ESD](#) profile measured through galaxy-galaxy lensing. The size-split [ESD](#) measurements from the [DLIS](#) lens galaxy sample, shown in [Figure 3.2](#), reveal that this offset contribution is especially prominent in $\mathcal{M}_1^*$ and $\mathcal{M}_2^*$ but appears at larger scales than typically expected. Additionally, we find that our fits to these data vectors either slightly underpredict or overpredict the lensing signal. In particular, for $\mathcal{M}_2^*$, the radii at which the [ESD](#) profile peaks due to the offset group term appear slightly shifted relative to the fit. To explore an improvement to the modelling, we consider the s parameter within the [AbacusSummit](#) base-[HOD](#), which had previously been fixed to zero. This parameter controls the placement of satellites within their host halo, allowing them to be positioned either closer to or farther from the host halo centre. The s parameter is defined as the following:

$$p_i = \frac{\langle N_s \rangle(M)}{N_p} \left[1 + s \left(1 - \frac{2 R_i}{N_p - 1} \right) \right], \quad (\text{C.1})$$

where R_i denotes the rank of a particle based on its distance from the host halo centre (furthest has rank 0 and closest has rank $N_p - 1$, and N_p is the total number of particles in the host halo. The average number of satellites is given by $\langle N_s \rangle$. When s is negative, satellites are placed closer to the host halo centre, whereas positive values move them farther from the centre. A simulated illustration of the impact of s on the spatial radial distribution of satellites, sampled from a [NFW](#), is shown in [Figure C.1](#). For a cluster/group halo of mass $M_h = 1 \times 10^{14} M_\odot$, the concentration from [Duffy et al. \(2008\)](#) is 5.7 and the virial radius is 960 kpc. We see the intended behaviour of s , which was previously described. We note that in the $s = 0$ case, the spatial distribution is not uniform across radii. This

variation is a consequence of the halo concentration. Satellites are placed according to the locations of dark matter particles within a host halo. If the host halo is more concentrated, there are intrinsically more particles at smaller radii than larger, leading to the distribution seen in green.

To add s into our model without having to re-run the entire analysis, we generate mock galaxy catalogues using the best-fitting HOD parameters in Table 5.1, but allow s to vary. We generate a mock catalogue for 21 s values ranging from -1 to 1. For each mock catalogue, the size split ESD vector for $\mathcal{M}_1^*$, $\mathcal{M}_2^*$, and $\mathcal{M}_3^*$ galaxies is computed using the β_{cent} and β_{sat} values taken from the HMC posteriors shown in Figure 5.3. We then loop over each of the 21 ESD vectors and calculate a χ_{joint}^2 value, adding the values of χ_{small}^2 and χ_{large}^2 together. From this process, we find that the lowest χ_{joint}^2 value occurs at $s = -0.3$. We do not estimate uncertainties for this value of s , as the main goal of this investigation was to see whether we were sensitive to the s parameter and whether it improved our fits significantly. Thus, we refrain from over-interpreting this value and comparing this to results in the literature concerning the radial distribution of satellite galaxies in clusters. For a visual illustration of the effect that altering s has on the resulting $\Delta\Sigma$ vectors, refer to Figure C.2.

Since $s = 0$ is not the best-fitting value, we proceed with the methodology outlined in Chapter 5 to assess how this impacts our results for β_{cent} , β_{sat} , η_{cent} , and η_{sat} . An HMC with uniform priors on β_{cent} and β_{sat} ranging from -0.15 to 0.15 yields posteriors that are slightly shifted within the error bars of our previous results (see Figure 5.3). Additionally, the overall trends for β_{cent} and β_{sat} remain consistent with the $s = 0$ case, although the marginal constraints differ slightly. Nevertheless, using the updated values for β_{cent} and β_{sat} , we generate a new mock catalogue with $s = -0.3$, which is then split into *large* and *small* size bins. The resulting fits to the DLIS ESD measurements are shown in Figure C.3. The solid orange/purple lines correspond to the fits obtained with $s = -0.3$, while the orange and purple dashed lines show the original fits from this work with $s = 0$. The χ_{joint}^2 values for the $s = -0.3$ fits are 17.11, 31.60, and 23.66 for the three stellar mass bins, respectively. The DOF is 22. When compared to the values in Table 5.3, the inclusion of the s parameter leads to varied levels of improvement. $\mathcal{M}_2^*$ shows the most significant improvement, with the fit improving and lowering the value of χ_{joint}^2 by 8.74. The fits for $\mathcal{M}_1^*$ and $\mathcal{M}_3^*$ slightly worsen, with an increase in χ_{joint}^2 of 2.34 and 2.85, respectively. We continue our investigation by computing updated values for η_{cent} and η_{sat} , shown in Figure C.4. The blue points represent η_{cent} values, while the orange points correspond to η_{sat} . We observe slight shifts in both η_{cent} and η_{sat} when comparing results for $s = 0$ and $s = -0.3$. However, these differences are small relative to the size of the error bars and are not statistically significant. We conclude that although introducing the s parameter could

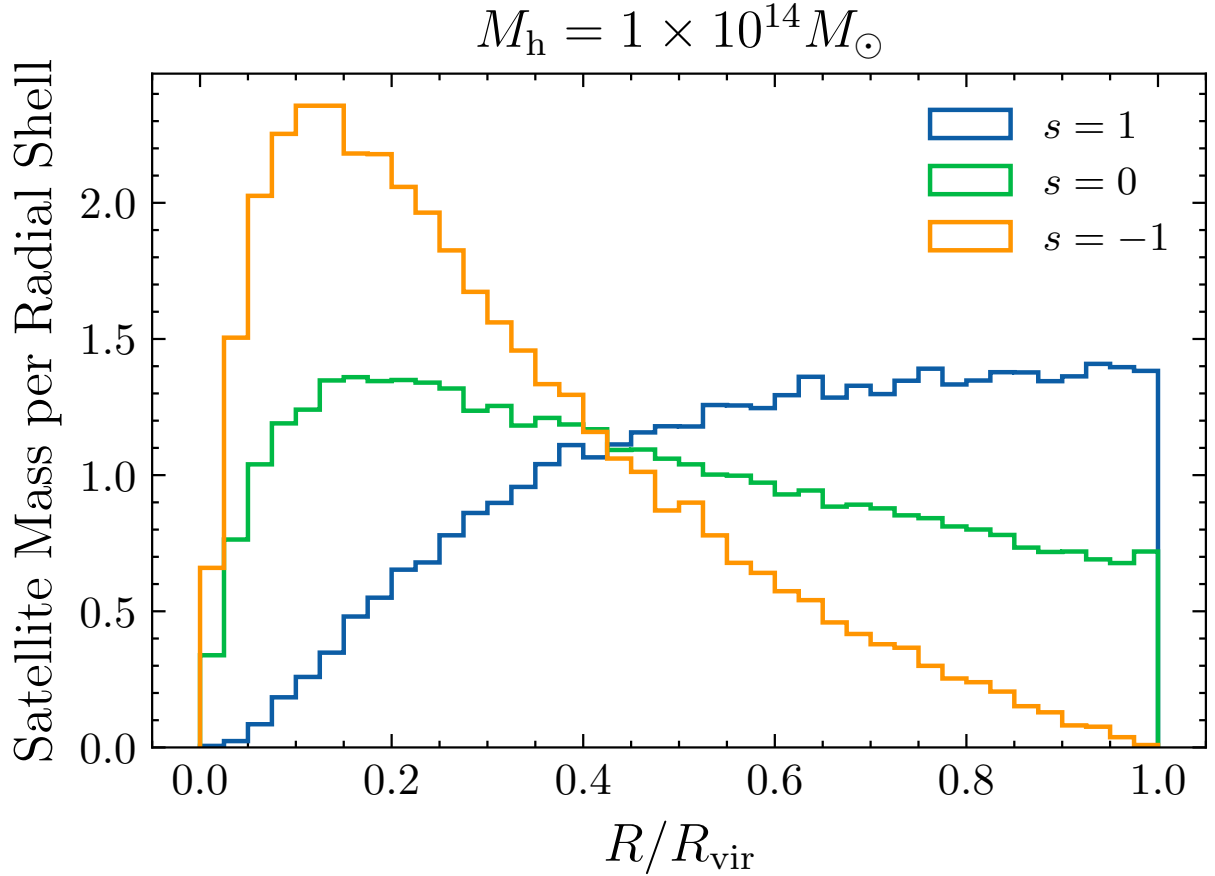


Figure C.1: Illustration of the impact of the s parameter on the spatial radial distribution of satellite galaxies in a halo of mass $M_h = 1 \times 10^{14} M_\odot$. The plotted quantity corresponds to satellite mass per radial shell, which is proportional to the density multiplied by shell volume. The concentration for this halo is 5.7 taken from the mass-concentration relation of [Duffy et al. \(2008\)](#). The virial radius is 960 kpc. Satellite positions are drawn from an NFW profile using a Monte Carlo method, with each candidate accepted or rejected according to the probability scheme in Eqn. C.1. For $s < 0$, satellites are preferentially located near the halo centre, while $s > 0$ shifts them toward the outskirts. When $s = 0$, all particles are equally likely to host satellites; in this case, the resulting spatial distribution reflects the underlying halo concentration, and because the NFW profile behaves as r^{-1} near the centre and r^{-3} at large radii

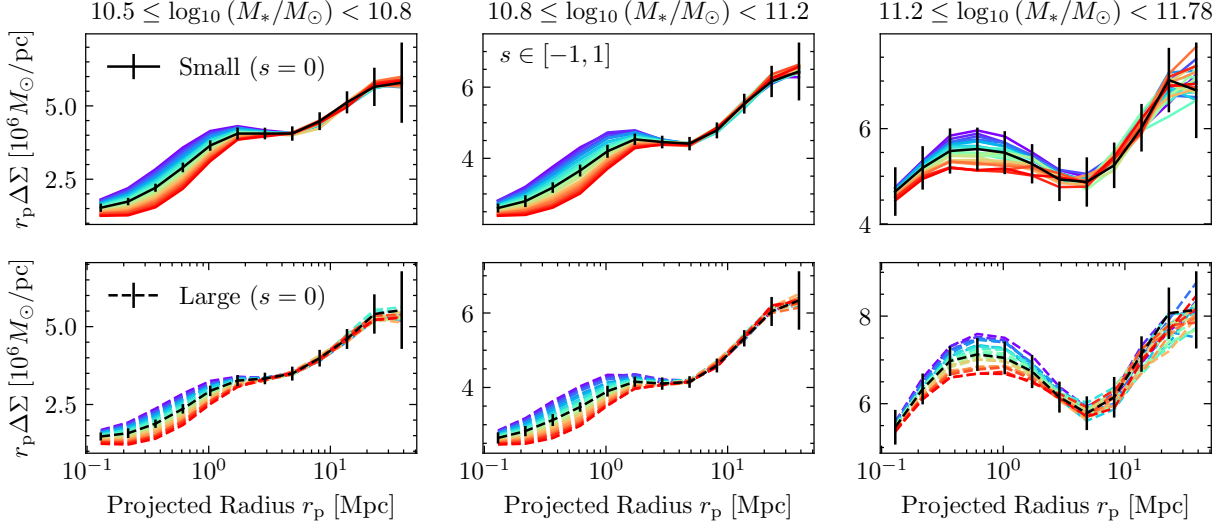


Figure C.2: Illustration of the effect that varying the satellite placement parameter, s , has on the $r_p\Delta\Sigma$ vectors from **AbacusSummit**. The top row plots the $r_p\Delta\Sigma$ vector for the *small* galaxy sample. The bottom row plots the $r_p\Delta\Sigma$ vector for the *large* galaxy sample. Moving from left to right, the columns correspond to our three stellar mass bins. Plotted as solid or dashed black lines is the **ESD** vector obtained using the best-fit HOD parameters in Table 5.1, where $s = 0$. The error bars are the sum of the measurement error and the emulator error from the $\Delta\Sigma_{\text{mocks}}$ and $\Delta\Sigma_{\text{tNFW}} f_{\text{sat}}$ emulators. Values for s are varied symmetrically around the $s = 0$, ranging from -1 to 1. Bluer colours are values < 0 , and redder colours are values > 0 .

improve the overall fit to the **DLIS ESD** measurements, excluding it as a free parameter does not impact the main conclusions of this work or affect the comparison with previous studies.

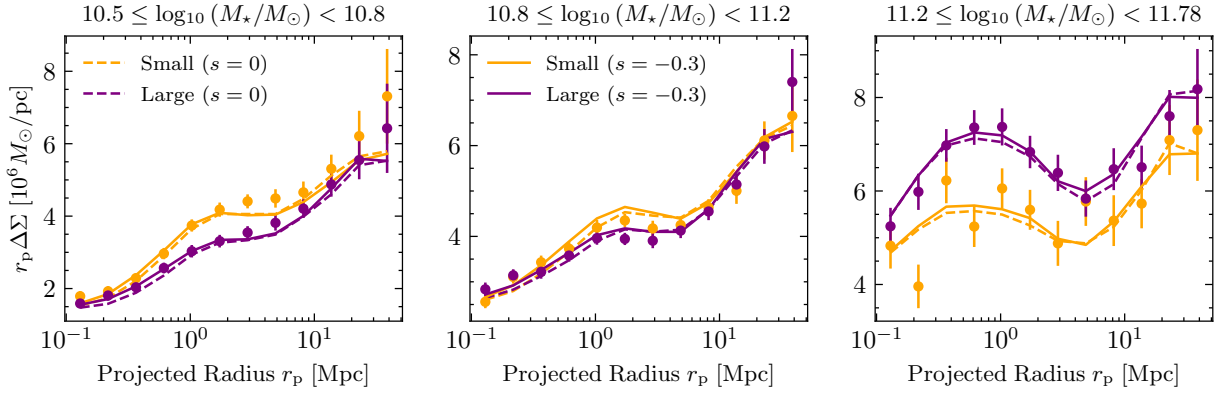


Figure C.3: Similar plot to Figure 3.2. Plotted are the $r_p \Delta \Sigma$ vectors for the three stellar mass bins. The orange data points are the *small* galaxy sample [ESD](#) vector, while the purple data points are for the *large* galaxy sample. The plotted error bars are the [DLIS](#) measurement covariance matrix obtained by Jackknife resampling with the model fitting error from the $\Delta \Sigma_{\text{mocks}}$ and $\Delta \Sigma_{\text{tNFW}} f_{\text{sat}}$ emulators added. The dashed orange and purple lines correspond to the best-fit vector for the $s = 0$ case. The solid orange and purple lines correspond to the best-fit vector, which was obtained when $s = -0.3$. The χ^2_{joint} values for the $s = -0.3$ fits are 17.1, 31.60, and 23.66. The [DOF](#) is 22. The χ^2_{joint} values for the $s = 0$ are summarized in Table 5.3.

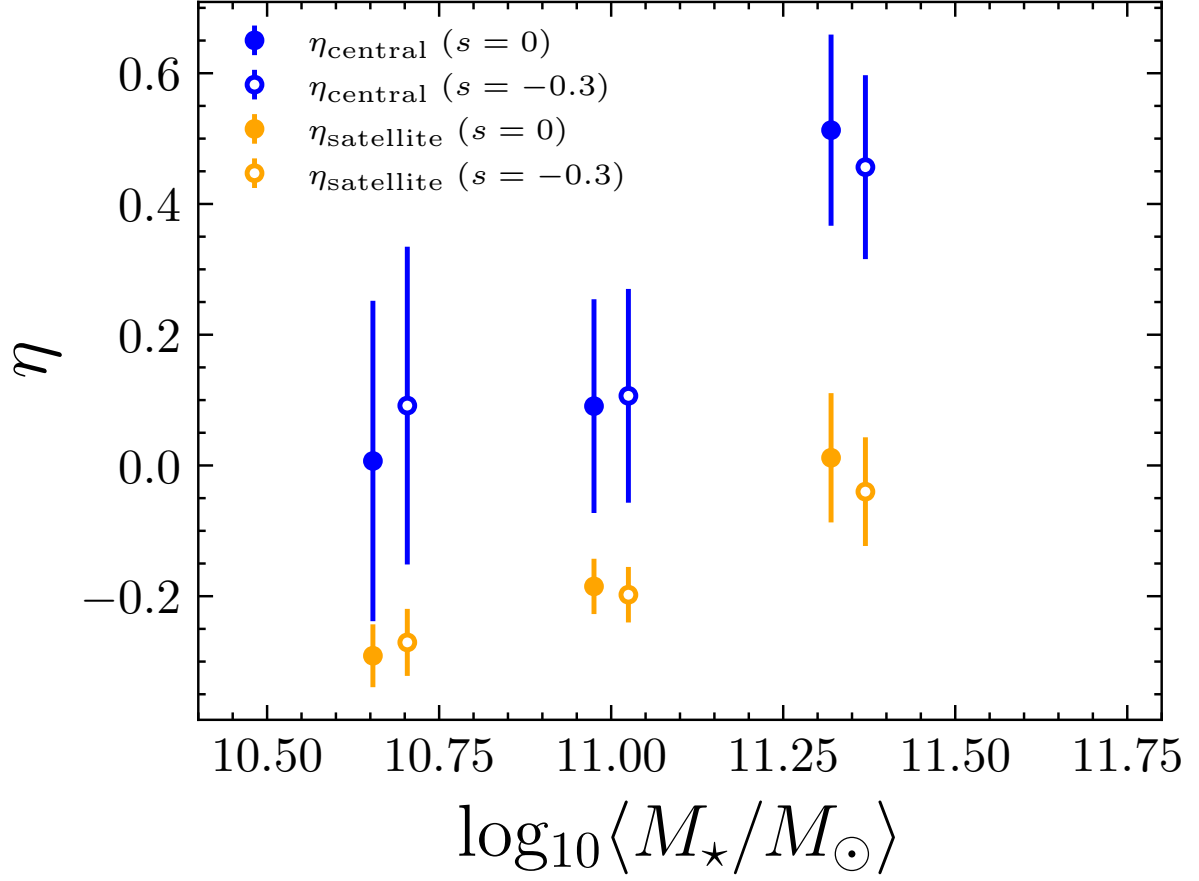


Figure C.4: Comparison of η_{cent} and η_{sat} values determined for red, early-type galaxies in the case of $s = 0$ and $s = -0.3$. η_{sat} values are shifted down by -0.1 for clarity. η_{cent} and η_{sat} values for the $s = -0.3$ case are shifted to the right by adding +0.05 to the mean $\log_{10}$ stellar mass values.