

Revisiting the ‘Lensing is Low’ Problem With UNIONS

by

Martine Campbell

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

Waterloo, Ontario, Canada, 2025

© Martine Campbell 2025

Author's Declaration

I hereby declare that I am the sole author of this 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.

Abstract

In this thesis, we present new measurements of the galaxy–galaxy lensing (GGL) signal around Baryon Oscillation Spectroscopic Survey (BOSS) CMASS galaxies using background sources from the Ultraviolet Near-Infrared Optical Northern Survey (UNIONS). With an overlap of approximately 2650 square degrees between CMASS lenses and background source galaxies—the largest to date—we obtain precise large-scale GGL measurements. With these new measurements, we revisit the so-called ‘lensing is low’ problem, wherein galaxy–halo connection models calibrated on clustering data over-predict the GGL signal by 20–40% under cosmic microwave background (CMB)-based cosmologies. We model the galaxy–halo connection using a halo occupation distribution (HOD), and perform joint fits to both GGL and clustering signals across a wide range of scales, as well as a clustering-only fit. Similar to previous work, we find a lensing–is–low effect in the CMASS sample, although our GGL and clustering predictions are less inconsistent with each other. The best joint fits are achieved by lowering the amplitude of the matter power spectrum relative to Planck 2018, driven by the precision of our large-scale GGL measurements. Once a lower matter power spectrum amplitude is adopted, feedback is the only HOD extension that further improves the joint fit. Our feedback model redistributes matter within a halo, modifying the halo–matter cross–power spectrum. Overall, we find that two models describe our observables equally well: one where HOD and cosmological parameters are free, and one where HOD, cosmological, and feedback parameters are free. Importantly, we emphasize the role of large scales in driving the lensing–is–low effect, shifting the narrative away from a purely small-scale issue.

Acknowledgements

First and foremost, I would like to thank Dr. Michael J. Hudson for the opportunity to work on this project and to be a part of the UNIONS collaboration. Your expertise in cosmology and scientific input helped make this project successful and unique. Thank you for your support and for believing in me, especially during moments of self-doubt. I am also deeply grateful to Dr. Jack Elvin-Poole for everything: the countless hack sessions where we would debug code, practice sessions for presentations, life advice, and scientific guidance. Collaborating with both of you has been an immensely rewarding experience that I will always cherish.

Next, I would like to thank my friends Sofia, Andrea, James, Darshak, Julian, Neo, David, and Rashaad. You provided much needed distractions from the stresses of graduate student life, always made me laugh, and supported me when I needed it most. You are the greatest friends anyone could ever hope for.

I would also like to acknowledge the high school teachers who had a profound influence on me: Mr. Kure, Ms. Cerovcec, and Mr. Maurati. To Mr. Kure: thank you for your dedication and passion for teaching physics. I still remember the physics demonstrations you set up during the winter break, and how fun and engaging they were. Your enthusiasm helped spark my interest in physics. To Ms. Cerovcec: I also thank you for your dedication and passion for teaching physics. Your class taught me the remarkable engineering that goes into everyday things like highways. I greatly appreciate the time you took to explain concepts and your support during difficult times. To Mr. Maurati: thank you for making math fun and approachable. Because of you, math became less intimidating and much more exciting. You are all the gold standard of teaching, and I hope you continue to inspire students as much as you inspired me. I also extend my gratitude to my high school guidance counsellor, Mr. Eagles, for your help and unwavering support.

Pursuing graduate-level research would have been challenging without prior experience. For that, I thank my undergraduate supervisors Dr. Sarah Rauscher, Dr. Claudiu Gradinaru, and Dr. Lichen Liang. It goes without saying that your mentorship and training helped shape the researcher I am today.

I also thank the undergraduate instructors who deepened my passion for physics: Dr. Andreas Hilfinger, Dr. Marc De Benedetti, Dr. Dennis Fernandes, and Dr. Daniel Nino. Your approachability, patience, and engaging teaching style made physics both understandable and fun. When I think back to my happiest undergraduate moments, your courses always come to mind.

This acknowledgements section would be incomplete without crediting my undergraduate lab instructors, Dan Reynolds and Gideon Humphrey. Thank you for the countless hours spent discussing and debating wacky physics theories and for creating a warm, welcoming lab environment. Our conversations significantly enhanced my understanding of physics.

Finally, I would like to thank my family: mom, dad, Obada, auntie Roula, and uncle Ahmad. I would not be the person I am without you in my life, and I am immensely grateful for your love, guidance, and support.

Dedication

This thesis is dedicated to my brother and to my grandmother, who I know would be proud of this work.

Table of Contents

Author’s Declaration	ii
Abstract	iii
Acknowledgements	iv
Dedication	vi
List of Figures	x
List of Tables	xiii
List of Abbreviations	xiv
1 Introduction	1
1.1 Cosmological Framework	1
1.2 Structure Formation	5
1.3 The Halo Model	9
1.4 Halo Occupation Distributions	13
1.5 Feedback	16
1.6 Gravitational Lensing	17
1.7 Galaxy–Galaxy Lensing	20

1.8	Galaxy–Galaxy Lensing Systematics	23
1.9	Galaxy Clustering	27
1.10	The ‘Lensing is Low’ Problem	28
1.11	Outline	31
2	Data	33
2.1	UNIONS Background Source Galaxy Sample	33
2.2	BOSS Foreground Lens Galaxy Sample	34
3	Measurements	38
3.1	Excess Surface Density, $\Delta\Sigma$	38
3.2	Projected Correlation Function, w_p	42
4	Theoretical Framework	48
4.1	Theoretical Two-Point Correlation Functions	48
4.2	The DARK EMULATOR	50
4.3	Feedback	51
5	Parameter Estimation	57
6	Results and Discussion	60
6.1	HOD With Planck Cosmology	60
6.2	HOD With Trial Cosmologies	64
6.3	Comparison	74
6.4	Discussion	75
7	Conclusions	77
	References	79
	APPENDICES	86

A	Newtonian Deflection Angle Derivation	87
B	Cosmological Parameter Dependence of Measurements	91
C	$\Delta\Sigma$ and w_p Cross-Covariance Test	93
D	Additional Plots	98

List of Figures

1.1	Geometry of a lens–source system in the thin-lens approximation, wherein distances between objects are much larger than the extent of the lens mass distribution along the line of sight (LOS). Taken from Abdo, nd.	19
1.2	Convergence (κ) and shear (γ) applied to a circle. The right hand side shows an ellipse with semi-major and semi-minor axes a and b , respectively. Taken from Prat & Bacon (2026).	22
1.3	Two-dimensional correlation function, $\xi(r_p, r_\pi)$, of CMASS galaxies from Reid et al. (2014). Here, r_p is denoted by r_σ	29
2.1	Effective redshift distribution of UNIONS ShapePipe v1.4.5 galaxies (yellow) compared to the CMASS redshift distribution (blue) without any redshift cuts. Vertical black lines indicate the CMASS redshift cuts of $z = 0.43$ and $z = 0.7$. ‘PDF’ on the y-axis stands for ‘probability density function’. . . .	35
2.2	On-sky overlap between UNIONS ShapePipe v1.4.5 galaxies (orange) and CMASS galaxies in the redshift range $0.43 < z < 0.7$ (blue). We estimate the area of the overlap region to be ~ 2650 square degrees.	37
3.1	GGL signal/ESD around CMASS galaxies. The horizontal axis corresponds to the midpoints of r_p bins. Orange points show our measurements using UNIONS ShapePipe v1.4.5 source galaxies. Purple points show results from Leauthaud et al. (2017) using Canada-France-Hawai’i Telescope Stripe 82 Survey (CS82; Moraes et al., 2014) and Canada-France-Hawai’i Telescope Lensing Survey (CFHTLenS; Heymans et al., 2012; Miller et al., 2013) source galaxies. Orange points with black borders indicate measurements without the boost factor correction. Only statistical 1σ errors are shown in this plot. Note that because of the difference in overlap with CMASS galaxies, we do not expect our measurements to be the exact same as Leauthaud et al. (2017).	40

3.2	Correlation matrix computed from the $\Delta\Sigma$ covariance matrix. By construction, every matrix element must have a value between -1 and 1. Small-scale r_p bins are uncorrelated with other bins, as shape noise dominates on these scales. Large-scale bins are mildly correlated due to LSS.	43
3.3	Correlation matrix computed from the w_p covariance matrix. Small-scale r_p bins are dominated by shot noise. Large-scale bins are strongly correlated due to LSS.	47
4.1	Gas density profiles for ‘low’ (grey curves) and ‘high’ (blue curves) feedback. For the low feedback case, $\log M_c = 11.0$ and $\theta_{ej} = 0.5$. For the high feedback case, $\log M_c = 16.0$ and $\theta_{ej} = 5.0$. Larger halo masses are represented by darker curves. The following halo masses were used to generate these curves: $\log (M_h/h^{-1} M_\odot) = 12.0, 14.0, 16.0$	54
6.1	Right panel: 5-parameter HOD fit (blue curve) to the CMASS clustering signal, w_p (black points). Cosmological parameters are fixed at the Planck 2018 TT,TE,EE+lowE+lensing+bao values, and extended HOD parameters are set to the fiducial values listed in Table 5.1. Left panel: corresponding GGL prediction using the best-fit parameters from the w_p fit. Light blue curves show models drawn from the posterior distribution of the w_p fit. The 1σ error bars include both statistical and systematic uncertainties.	61
6.2	Joint fits to both the GGL and clustering signals of CMASS galaxies. The baseline model (solid red curves) varies only the 5 HOD parameters, with all other parameters fixed at their fiducial values. A ‘+’ indicates that additional parameters are varied alongside the HOD parameters. Parameters characterizing a single extension to the standard HOD are grouped together. For example, ‘off’ indicates that both off-centring parameters (p_{off} and R_{off}) are varied. Feedback parameters are represented by ‘fbk’, and cosmological parameters by ‘cosmo’.	63
6.3	HOD Posterior distributions for the HOD and HOD+cosmo models. Median values and their uncertainties are listed in Table 6.3.	66
6.4	Similar to Fig. 6.2, except the baseline model is now ‘HOD+cosmo’.	67
6.5	Off-centring posterior distributions for the HOD+cosmo+off model. Black lines indicate median values, while orange lines mark the MAP estimates. The off-centring effect vanishes if either p_{off} or R_{off} goes to zero.	68
6.6	Posterior distributions for the HOD+cosmo and HOD+cosmo+fbk models.	71

A.1	Deflection angle, $\delta\theta$, of a particle due to a foreground mass (orange point). The blue line represents the particle's trajectory.	88
C.1	Raw joint covariance matrix for $\Delta\Sigma$ and w_p . 'C' stands for clustering, and 'l' for lensing. Each block is an individual covariance matrix. Note that the colour scale is set to a range of -2 to 2	94
C.2	χ^2 distributions from the joint covariance matrix in Fig. C.1 and the same matrix but with the cross-covariance blocks (C_{cl} and C_{lc}) set to zero. . . .	96
C.3	$\Delta\chi^2$ distribution from the differences between the datasets in Fig. C.2. . .	97
D.1	Posterior distributions for the HOD+cosmo+off+fbk+ R_c model.	99
D.2	Effects of varying the parameter $\log M_{\min}$ on $\Delta\Sigma$ and w_p	100
D.3	Effects of varying the parameter $\sigma_{\log M}^2$ on $\Delta\Sigma$ and w_p	101
D.4	Effects of varying the parameter $\log M_1$ on $\Delta\Sigma$ and w_p	101
D.5	Effects of varying the parameter α on $\Delta\Sigma$ and w_p	102
D.6	Effects of varying the parameter κ on $\Delta\Sigma$ and w_p	102
D.7	Effects of varying the parameter p_{off} on $\Delta\Sigma$ and w_p	103
D.8	Effects of varying the parameter R_{off} on $\Delta\Sigma$ and w_p	103
D.9	Effects of varying the parameter R_c on $\Delta\Sigma$ and w_p	104
D.10	Effects of varying the parameter $\ln(10^{10} A_s)$ on $\Delta\Sigma$ and w_p	104
D.11	Effects of varying the parameter $\Omega_{m,0}$ on $\Delta\Sigma$ and w_p	105
D.12	Effects of varying the parameters $\log M_c$ and θ_{ej} on $\Delta\Sigma$. The fiducial value for $\log M_c$ is 14, while 1 is the fiducial value for θ_{ej}	105

List of Tables

5.1	Priors and fiducial values used for parameter estimation. ‘Extended HOD’ refers to parameters beyond the standard HOD framework that do not enter the expressions for $\langle N_c(M) \rangle$ and $\langle N_s(M) \rangle$. The symbol $\mathcal{U}(\text{low}, \text{high})$ denotes a uniform prior with lower and upper bounds given in parentheses, and $\mathcal{G}(\mu, \sigma)$ denotes a Gaussian prior with mean μ and width σ . Priors on cosmological parameters are set by DARK EMULATOR’s supporting range. Feedback is only incorporated when its corresponding parameters are allowed to vary.	59
6.1	Statistical values from best-fit models. This table only includes results from fits to both $\Delta\Sigma$ and w_p . The number of free parameters is represented by k , and the total data vector length across all observables (including $\bar{n}_g$) is 50.	70
6.2	Best-fit parameter values drawn from the MAP. For the fiducial Planck cosmology, $\sigma_8 = 0.809$ and $S_8 = 0.824$	72
6.3	Posterior constraints from MCMC chains. We report the 50th percentile as the central value, with lower and upper limits given by the differences between the 50th and 16th percentiles, and the 84th and 50th percentiles, respectively. We quote the median instead of the mean since many of the posterior distributions are skewed.	73
D.1	Fiducial parameter values used to generate the ‘rainbow plots’ in this Appendix. By default, feedback is not included.	100

List of Abbreviations

2PCF Two-Point Correlation Function [14](#), [15](#), [27](#), [28](#)

AICc Akaike Information Criterion [60](#), [62](#), [64](#), [65](#), [69](#), [70](#), [75](#), [78](#)

AM Abundance Matching [15](#)

BAO Baryon Acoustic Oscillation [5](#), [8](#)

BC Baryon Correction [16](#), [17](#), [51](#), [77](#)

BIC Bayesian Information Criterion [60](#), [62](#), [64](#), [65](#), [69](#), [70](#), [75](#), [78](#)

BOSS Baryon Oscillation Spectroscopic Survey [28](#), [30](#), [33](#), [34](#), [44](#), [75](#), [77](#)

CDM Cold Dark Matter [1–5](#), [32](#), [50](#), [58](#)

CLM Collisionless Matter [16](#), [51](#), [55](#)

CMB Cosmic Microwave Background [4](#), [5](#), [30](#), [31](#)

DES Dark Energy Survey [33](#), [39](#)

DR12 Data Release 12 [34](#), [44](#), [77](#)

EOS Equation-Of-State [2](#), [5](#), [6](#)

ESD Excess Surface Density [x](#), [21](#), [39](#), [40](#), [49](#)

FKP Feldman-Kaiser-Peacock [44](#)

FoG Fingers of God [27](#), [28](#)

GAB Galaxy Assembly Bias [12](#), [15](#), [78](#)

GGL Galaxy–Galaxy Lensing [x](#), [xi](#), [9](#), [13](#), [17](#), [18](#), [20](#), [21](#), [23–28](#), [30–34](#), [38](#), [40](#), [41](#), [48](#), [49](#), [56](#), [60–65](#), [69](#), [70](#), [75](#), [77](#), [78](#)

HAB Halo Assembly Bias [12](#)

HMF Halo Mass Function [10](#), [49](#), [56](#)

HOD Halo Occupation Distribution [xi](#), [xiii](#), [13–15](#), [28](#), [30](#), [31](#), [48](#), [50](#), [51](#), [58–63](#), [65](#), [66](#), [69](#), [70](#), [74](#), [76–78](#)

IA Intrinsic Alignment [24](#), [25](#), [41](#), [65](#), [75](#), [78](#)

LRG Luminous Red Galaxy [28](#), [34](#)

LSS Large-Scale Structure [xi](#), [23](#), [24](#), [34](#), [43](#), [47](#), [49](#)

MAP Maximum a Posteriori [xi](#), [xiii](#), [58](#), [62](#), [64](#), [65](#), [68–70](#), [72](#)

MCMC Markov Chain Monte Carlo [xiii](#), [57](#), [58](#), [73](#), [77](#)

NGC North Galactic Cap [34](#), [36](#)

RSD Redshift Space Distortion [27](#)

SDSS Sloan Digital Sky Survey [4](#), [34](#), [36](#), [44](#)

SGC South Galactic Cap [34](#), [36](#)

SHAM Subhalo Abundance Matching [15](#), [28](#)

SNe Ia Type Ia Supernovae [4](#)

SNR Signal-To-Noise Ratio [25](#), [34](#), [44](#)

SOM Self-Organizing Map [34](#), [41](#)

UNIONS Ultraviolet Near-Infrared Optical Northern Survey [x](#), [31](#), [33–35](#), [37](#), [40](#), [75](#), [77](#)

WMAP Wilkinson Microwave Anisotropy Probe [4](#), [28](#)

Chapter 1

Introduction

1.1 Cosmological Framework

Cosmology is the study of the origin, nature, and structure of the Universe. A cosmological framework specifies the components of the Universe and describes how space evolves. In this work, we adopt a flat Λ CDM cosmology, which is described here. Our primary source for this section is [Mo et al. \(2010\)](#). If the Universe is homogeneous and isotropic (has the same properties everywhere and looks the same in all directions from any given point) on large scales, its geometry is described by the Robertson-Walker (RW) metric:

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

where $a(t)$ is the scale factor, which describes how distances change with time due to cosmic expansion or contraction. The spatial curvature constant, K , specifies whether the Universe is open, flat, or closed ($K = -1, 0, 1$, respectively). Applying the RW metric to the Einstein field equation yields the Friedmann equation:

$$\left[\frac{\dot{a}(t)}{a(t)} \right]^2 = \frac{8\pi G}{3} \rho(t) - \frac{Kc^2}{a^2(t)} + \frac{\Lambda c^2}{3}. \quad (1.2)$$

Note that $\dot{a}(t) = da(t)/dt$. The Friedmann equation describes the expansion history of the Universe and encodes its energy content for any given time. Here, $\rho(t)$ is a mass density that includes contributions from matter and radiation, and Λ is the ‘cosmological

constant', originally introduced by Einstein to allow a static universe (at that time, the Universe was thought to be unchanging). Cosmologies that obey this equation are referred to as Friedmann-Robertson-Walker (FRW) cosmologies, of which the Λ CDM model is a particular case.

We now define the Hubble parameter, $H(t)$:

$$H(t) \equiv \frac{\dot{a}(t)}{a}, \quad (1.3)$$

where $H(t_0) = H_0$ is the Hubble constant. We can also express the cosmological constant in terms of an energy density, ρ_Λ , analogous to that of a vacuum:

$$\rho_\Lambda = \frac{\Lambda c^2}{8\pi G}. \quad (1.4)$$

The density above has an associated pressure, P_Λ :

$$P_\Lambda = -\rho_\Lambda c^2, \quad (1.5)$$

In the expression above, the dimensionless [Equation-Of-State \(EOS\)](#) parameter, w , is set to -1. However, w need not equal -1. Cosmologies with a constant w are referred to as w CDM cosmologies. The more exotic $w_0 w_a$ cosmology permits a time-varying w :

$$w(a) = w_0 + w_a(1 - a). \quad (1.6)$$

Note that in the expression above, a is dimensionless. The parameter ρ_Λ is called the 'dark energy density'. Dark energy refers to an unknown force that drives the accelerated expansion of the Universe. If we assume a spatially flat universe and that dark matter is 'cold' (i.e., its kinetic energy is negligible compared to its rest mass energy), then the Friedmann equation becomes:

$$H^2(t) = \frac{8\pi G}{3} \left(\rho_{m,0} \left[\frac{a_0}{a(t)} \right]^3 + \rho_{r,0} \left[\frac{a_0}{a(t)} \right]^4 + \rho_{\Lambda,0} \right), \quad (1.7)$$

where $\rho_{r,0}$ is the present-day radiation density, and $a_0 = a(t_0)$. Note that the radiation component has an extra factor of $a_0/a(t)$ due to the redshifting of photons. The present-day matter density, $\rho_{m,0}$, is given by:

$$\rho_{\text{m},0} = \rho_{\text{b},0} + \rho_{\text{CDM},0}. \quad (1.8)$$

In the expression above, $\rho_{\text{b},0}$ and $\rho_{\text{CDM},0}$ are the present-day baryon and [Cold Dark Matter \(CDM\)](#) densities, respectively. This defines the flat Λ CDM framework.

Now, we introduce the critical density, ρ_{crit} :

$$\rho_{\text{crit}}(t) \equiv \frac{3H^2(t)}{8\pi G}. \quad (1.9)$$

Using [1.9](#), we define the dimensionless density parameters:

$$\Omega_{\text{x}} = \frac{\rho_{\text{x}}}{\rho_{\text{crit}}}, \quad (1.10)$$

where $\text{x} = \text{m}, \text{r}, \text{or } \Lambda$. In terms of the dimensionless density parameters, the Friedmann equation becomes:

$$H^2(t) = H_0^2 \left(\Omega_{\text{m},0} \left[\frac{a_0}{a(t)} \right]^3 + \Omega_{\text{r},0} \left[\frac{a_0}{a(t)} \right]^4 + \Omega_{\Lambda,0} \right). \quad (1.11)$$

Before continuing, we define the cosmological redshift, z , for an expanding universe. Let λ_{e} denote the wavelength of light at the time of emission, and λ_0 the wavelength measured by an observer at the origin. The cosmological redshift is then given by:

$$1 + z \equiv \frac{\lambda_0}{\lambda_{\text{e}}} = \frac{a(t_0)}{a(t_{\text{e}})} \implies \frac{a_0}{a(z)} = 1 + z. \quad (1.12)$$

Thus, Eq. [1.11](#) can be re-parametrized in terms of z :

$$H^2(z) = H_0^2 E^2(z), \quad (1.13)$$

where

$$E(z) = \sqrt{\Omega_{\text{m},0}(1+z)^3 + \Omega_{\text{r},0}(1+z)^4 + \Omega_{\Lambda,0}}. \quad (1.14)$$

We convert redshifts to comoving distances, χ , via:

$$\chi(z) = \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')}. \quad (1.15)$$

Comoving distances factor out the expansion or contraction of the Universe, and therefore remain constant over time. They can be thought of as fixed positions on a moving grid. We express comoving distances in the typical units of h^{-1} Mpc, where $h \equiv H_0/100 \text{ km s}^{-1} \text{ Mpc}$. This convention eliminates the comoving distance dependence on a choice of H_0 .

The Λ CDM model has served as the standard cosmological framework for the past three decades. It is often referred to as the ‘concordance model’ due to its success in accurately describing various observations. Here, we highlight some of the key observations/papers that contributed to the model’s widespread acceptance. First, [Efstathiou et al. \(1990\)](#) measured the angular correlation function of galaxies from the APM Survey, along with the excess variance in galaxy counts from the Infrared Astronomical Satellite (IRAS; $z \sim 0.03$). In both cases, the ‘standard’ CDM model ($\Omega_{\text{CDM}} = 1$) under-predicted the measurements on scales where linear theory applies (see Section 1.2). By contrast, a low-density CDM model with $\Omega_{\text{CDM},0} = 0.2$ and a positive cosmological constant provided a much better fit. Complementary numerical simulations by [Davis et al. \(1985\)](#) showed that CDM clustering reproduces the filaments, voids, and superclusters observed in galaxy surveys, assuming adiabatic initial density fluctuations (see Section 1.2). Later, galaxy clustering measurements from large spectroscopic surveys such as the 2dF Galaxy Redshift Survey (2dFGRS; [Colless et al., 2001](#); [Cole et al., 2005](#)) and the Sloan Digital Sky Survey (SDSS; [York et al., 2000](#); [Tegmark et al., 2004](#)) reaffirmed the Λ CDM framework.

Another low-redshift cosmological probe is Type Ia Supernovae (SNe Ia). [Perlmutter et al. \(1999\)](#) fit the FRW magnitude-redshift relation to the apparent B-band magnitudes of SNe Ia discovered by the Supernova Cosmology Project and from the Calán/Tololo Supernova Survey. These supernovae spanned a broad redshift range, from $z < 0.1$ to $z = 0.83$. The data disfavoured both flat and open universes with $\Lambda = 0$, and instead preferred a positive cosmological constant at the 99% confidence level. Around the same time, [Riess et al. \(1998\)](#) also found that a positive cosmological constant provided a good fit to their observations of SNe Ia in the redshift range $0.16 < z < 0.97$, comparing theoretical and measured distance moduli. Numerous SNe Ia observations indicated that the Universe is undergoing accelerated expansion, as distant supernovae appeared fainter than expected in a static universe.

High-redshift data also support the Λ CDM model. Cosmic Microwave Background (CMB; $z \sim 1100$) measurements from probes such as the Cosmic Background Explorer (COBE; [Mather et al., 1990](#); [Smoot et al., 1992](#); [Boggeess et al., 1992](#)), the Wilkinson Microwave Anisotropy Probe (WMAP; [Bennett et al., 2003](#); [Hinshaw et al., 2003](#); [Spergel](#)

et al., 2003), and the Planck spacecraft (Planck Collaboration et al., 2014a,b,c) are all well-described by a Λ CDM cosmology. Different cosmologies (e.g., OCDM, SCDM) predict distinct amplitudes and shapes for the CMB power spectra, which feature peaks and troughs from acoustic oscillations in the baryon–photon fluid (Bahcall et al., 1999). A unique aspect of the CMB is that it strongly constrains the spatial curvature. The angular sizes of hot and cold spots in CMB maps would appear smaller in a negatively curved universe compared to a flat universe. Conversely, the spots would appear larger in a positively curved universe. Observations indicate a nearly flat universe, with no significant curvature. Additionally, galaxy clustering and CMB observations validate the cosmological principle—the assumption that the Universe is homogeneous and isotropic on large scales.

Although most studies agree with the Λ CDM framework, discrepancies exist in the precise values of cosmological parameters. The most prominent are the Hubble tension and the S_8 tension. Measurements of H_0 from the CMB and the local distance ladder differ by up to 6σ (Hu & Wang, 2023). CMB-based estimates place H_0 at or slightly below $70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, while low- z observations suggest a higher value. We briefly discuss the S_8 tension in Section 1.10. It is worth mentioning that recent Baryon Acoustic Oscillation (BAO) measurements from Abdul Karim et al. (2025) challenge Λ CDM by favouring a dynamical dark energy EOS, with $w_0 > -1$ and $w_a < 0$. However, the preference for a $w_0 w_a$ CDM cosmology over Λ CDM peaks at 4.2σ , and the constraints on w_0 and w_a differ between supernovae datasets. At present, Λ CDM remains the standard cosmological model.

1.2 Structure Formation

With an established cosmological framework, we now describe the current picture of structure formation, which is relevant for modelling our observables. Specifically, we discuss the linear theory evolution of structure using Mo et al. (2010) and Huterer (2023) as our main sources. The following description is valid on $\gtrsim 10 h^{-1} \text{ Mpc}$ scales. We consider the growth of density perturbations in an ideal, non-relativistic, self-gravitating fluid. These simplifying assumptions hold for perturbations whose spatial extent is smaller than the horizon. The properties of the fluid are described by the following equations:

$$\frac{\partial \rho}{\partial t} + (\mathbf{u} \cdot \nabla_{\mathbf{r}})\rho + \rho \nabla_{\mathbf{r}} \cdot \mathbf{u} = 0, \quad (1.16)$$

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla_{\mathbf{r}})\mathbf{u} = -\frac{\nabla_{\mathbf{r}} P}{\rho} - \nabla_{\mathbf{r}} \phi, \quad (1.17)$$

$$\nabla_{\mathbf{r}}^2 \phi = 4\pi G \rho. \quad (1.18)$$

The first equation is the continuity equation, which relates temporal changes in a fluid's density to the divergence of its velocity field. It states that a sink-like (source-like) velocity field corresponds to an increase (decrease) in density. The second equation is the Euler equation, and the third is the Poisson equation. Here, ρ is the fluid's density at a given position and time, $\mathbf{u}$ is the proper velocity (includes cosmic expansion), $\mathbf{r}$ is the proper coordinate, P is the pressure, and ϕ is the gravitational potential. From the Euler equation, we see that pressure and gravity compete in governing the fluid's motion.

In comoving coordinates, the governing equations become:

$$\frac{\partial \delta}{\partial t} + \frac{1}{a} \nabla \cdot [(1 + \delta) \mathbf{v}] = 0, \quad (1.19)$$

$$\frac{\partial \mathbf{v}}{\partial t} + \frac{\dot{a}}{a} \mathbf{v} + \frac{1}{a} (\mathbf{v} \cdot \nabla) \mathbf{v} = -\frac{\nabla \Phi}{a} - \frac{\nabla P}{a \bar{\rho} (1 + \delta)}, \quad (1.20)$$

$$\nabla^2 \Phi = 4\pi G \bar{\rho} a^2 \delta, \quad (1.21)$$

where the density contrast, δ , is defined as:

$$\delta(x, t) = \frac{\rho(x, t) - \bar{\rho}(t)}{\bar{\rho}(t)}. \quad (1.22)$$

The parameter $\bar{\rho}$ is the mean matter density. Note that $\nabla \equiv \nabla_{\mathbf{x}}$, where $\mathbf{x}$ is the comoving coordinate, and $\partial/\partial t$ is taken at fixed $\mathbf{x}$. Additionally, $\mathbf{v}$ is the peculiar velocity (non-Hubble flow velocity). Lastly, $\Phi \equiv \phi + a\ddot{a}x^2/2$.

Solving the governing equations requires an [EOS](#) since there are 5 equations (the Euler equation applies to each velocity component) and 6 unknowns. To simplify our work, we neglect the pressure term in the Euler equation since at late times the dynamics of the fluid are governed by gravity and cosmic expansion, with most of the gravity sourced by collisionless, pressureless dark matter. If we then assume that $|\delta| \ll 1$ and that $\mathbf{v}$ is small, the continuity and Euler equations can be linearized as follows:

$$\frac{\partial \delta}{\partial t} + \frac{1}{a} \nabla \cdot \mathbf{v} = 0, \quad (1.23)$$

$$\frac{\partial \mathbf{v}}{\partial t} + \frac{\dot{a}}{a} \mathbf{v} = -\frac{\nabla \Phi}{a}. \quad (1.24)$$

Combining Eqs. 1.21, 1.23, and 1.24 yields:

$$\frac{\partial^2 \delta}{\partial t^2} + 2\frac{\dot{a}}{a}\frac{\partial \delta}{\partial t} - 4\pi G\bar{\rho}\delta = 0. \quad (1.25)$$

The second term in the expression above is a Hubble drag term, reflecting that cosmic expansion slows perturbation growth.

Since our differential equation is linear in δ , its solutions can be expressed as a superposition of basis functions. Neglecting curvature, we adopt plane waves as our basis functions, allowing us to recast Eq. 1.25 in Fourier space:

$$\frac{d^2 \delta_{\mathbf{k}}}{dt^2} + 2\frac{\dot{a}}{a}\frac{d\delta_{\mathbf{k}}}{dt} - 4\pi G\bar{\rho}\delta_{\mathbf{k}} = 0, \quad (1.26)$$

$$\delta_{\mathbf{k}}(t) = \frac{1}{V_u} \int \delta(\mathbf{x}, t) \exp(-i\mathbf{k} \cdot \mathbf{x}) d^3\mathbf{x}, \quad (1.27)$$

$$\delta(\mathbf{x}, t) = \sum_{\mathbf{k}} \delta_{\mathbf{k}}(t) \exp(i\mathbf{k} \cdot \mathbf{x}). \quad (1.28)$$

Note that V_u is the volume of a box in which the perturbations are assumed to be periodic, and k is the wavenumber, with units of inverse length. In cosmology, it is conventional to recast equations in Fourier space as a means of simplification (e.g., operators become multiplicative factors). Working in Fourier space is especially beneficial here, as it allows us to track the evolution of independent Fourier modes.

To understand how density perturbations grow during different epochs, we consider single-component universes (e.g., $\Omega_r = 1$). The results are summarized below:

- **Radiation domination:** in this epoch, $a \propto t^{1/2} \implies \dot{a}/a \propto t^{-1}/2$, and the matter density term is negligible. Thus, the solutions to Eq. 1.26 are given by $\delta_{\mathbf{k}}(t) = A_1 + A_2 \ln(t)$, meaning that density perturbations grow only logarithmically with time.
- **Matter domination:** here, $a \propto t^{2/3} \implies \dot{a}/a \propto 2t^{-1}/3$, and $4\pi G\bar{\rho} = 3\dot{a}^2/2a^2$. Combining everything, the solutions to Eq. 1.26 are given by $\delta_{\mathbf{k}}(t) = B_1 t^{2/3} + B_2 t^{-1}$, exhibiting a growing mode and a decaying mode (with respect to time). This decaying mode quickly dies out, and is not relevant to structure formation. Since $a \propto t^{2/3}$, the growing mode scales linearly with the scale factor, enabling substantial structure formation during matter domination.

- **Dark energy domination:** in this era, the scale factor is exponentially related to time, and once again the density term is negligible. Notably, $H(t) = H = \text{const}$. The solutions to Eq. 1.26 are then given by $\delta_{\mathbf{k}}(t) = C_1 + C_2 \exp(-2Ht) \approx \text{const}$, meaning that density perturbations cease growing.

The density perturbation equation can be expressed in a dimensionless form as:

$$D(z) = \frac{\delta(z)}{\delta(z=0)}, \quad (1.29)$$

where $D(z)$ is referred to as the ‘linear growth function’. The matter power spectrum, P_{mm} , is given by:

$$\langle \delta_{\mathbf{k}} \delta_{\mathbf{k}'}^* \rangle = (2\pi)^3 \delta_{\text{D}}(\mathbf{k} - \mathbf{k}') P_{\text{mm}}(k, z), \quad (1.30)$$

where the $*$ denotes the complex conjugate, and δ_{D} is the Dirac delta function. Using linear theory, the matter power spectrum is modelled as:

$$P_{\text{mm}}(k, z) = P_{\text{i}}(k) T^2(k) D^2(z) = A_{\text{s}} \left(\frac{k}{k_{\text{piv}}} \right)^{n_{\text{s}}} T^2(k) D^2(z), \quad (1.31)$$

where P_{i} is the initial power spectrum, T the linear transfer function, A_{s} the amplitude, n_{s} the spectral index, and k_{piv} the pivot scale (typically 0.05 Mpc^{-1}). For notational simplicity, we omit the cosmological parameter dependence of the matter power spectrum. Note that the initial conditions in cosmological simulations are set by the primordial curvature power spectrum, which differs from P_{i} in its power of $n_{\text{s}} - 1$. The linear transfer function encodes scale-dependent features in P_{mm} , such as the slow rollover around the scale of radiation–matter equality (given by roughly $16[h\Omega_{\text{m},0}]^{-2} \text{ Mpc}$; [Hartlap, 2009](#)) and [BAO](#). Density perturbations grow differently during radiation domination compared to matter domination, as previously shown. Modes that entered the horizon during radiation domination grew inefficiently compared to those entering during matter domination, imprinting a characteristic bend in P_{mm} . Since density perturbations grow unimpeded during matter domination, $T(k)$ asymptotes towards 1 at small k . [BAOs](#) are acoustic oscillations in the baryon–photon fluid pre-recombination. Without dark matter, density perturbations would not grow during radiation domination due to these acoustic oscillations, contributing to inefficient structure growth. At photon–baryon decoupling, acoustic oscillations freeze out, leaving matter clustered at a characteristic scale (termed the ‘[BAO](#)’ scale).

In summary, we described the linear theory of structure formation, which is valid on large, ‘linear’ scales. We ignored relativistic effects in our derivation, and focused on sub-horizon scales. We also implicitly assumed ‘adiabatic’/isentropic initial conditions, meaning fluctuations are curvature-driven. On small scales, non-linear effects modify P_{mm} . For this reason, it is common to model P_{mm} with simulations on small scales.

1.3 The Halo Model

The evolution of density perturbations under linear theory accurately describes structure on large scales. However, this approximation breaks down on smaller scales since the density contrast exceeds unity (Mo et al., 2010). Recall that the linear approximation assumes $|\delta| \ll 1$, which is no longer valid. Consequently, scales where linear theory no longer holds are referred to as ‘non-linear’. To model structure on non-linear (and quasi-linear) scales, an alternative framework/methodology to linear theory is required. One common approach is the halo model, which stipulates that all matter is bound within halos, and that halos form via the collapse and virialization of large density perturbations (Mo et al., 2010; Asgari et al., 2023). Our predicted [Galaxy–Galaxy Lensing \(GGL\)](#) and clustering signals are based on the halo model, and we introduce its foundational equations here. The content in this section closely follows [Asgari et al. \(2023\)](#).

To begin, we define a generic field (e.g., matter, galaxy, halo) in real space:

$$\theta_{\text{u}}(\mathbf{x}) = \sum_i N_i W_{\text{u},i}(|\mathbf{x} - \mathbf{x}_i|), \quad (1.32)$$

where u denotes the field of interest, $\mathbf{x}$ is a comoving position, and $W_{\text{u},i}$ is a spherically symmetric profile associated with halo i centred at $\mathbf{x}_i$. While fields are time-dependent, we omit the temporal argument (e.g., z) for simplicity. Additionally, $N_i \in \{0, 1\}$, and it indicates whether a halo centre resides within the volume element under consideration. This condition enforces the assumption that all matter is contained within halos. In Fourier space, the field is:

$$\hat{\theta}_{\text{u}}(\mathbf{k}) = \int \theta_{\text{u}}(\mathbf{x}) \exp(-i\mathbf{k} \cdot \mathbf{x}) d^3\mathbf{x} = \sum_i \exp(-i\mathbf{k} \cdot \mathbf{x}_i) N_i \hat{W}_{\text{u},i}(k). \quad (1.33)$$

In the expression above, $\hat{W}_{\text{u},i}(k)$ is the Fourier transform of $W_{\text{u},i}(r)$, where $\mathbf{r} = \mathbf{x} - \mathbf{x}_i$.

We now assume that halo properties depend only on mass, such that $W_{u,i}(r) = W_u(M_i, r)$. This assumption becomes less reliable in the presence of assembly bias (Croton et al., 2007), which we describe at the end of this section. We also assume that halo masses are distributed according to the **Halo Mass Function (HMF)**, $n(M)$, which specifies the number density of halos per unit mass. The earliest **HMF** was derived in Press & Schechter (1974), and it is based on the spherical collapse of Gaussian density fluctuations. This **HMF** is referred to as the ‘Press-Schechter’ mass function. With these assumptions, we can write an expression for $\langle \hat{\theta}_u(\mathbf{x}) \rangle$:

$$\begin{aligned} \langle \hat{\theta}_u(\mathbf{x}) \rangle &= \left\langle \sum_i N_i W_u(M_i, |\mathbf{x} - \mathbf{x}_i|) \right\rangle = \int_0^\infty n(M) dM \int W_u(M, |\mathbf{x} - \mathbf{x}'|) d^3x', \\ &= \int_0^\infty W_u(M) n(M) dM. \end{aligned} \quad (1.34)$$

The ensemble average, including N_i , has been translated into $\int n(M) dM \Delta V_i$, where ΔV_i is the volume element. To achieve the final expression in Eq. 1.34, we have separated W_u into two components: $W_u(M, x) = W_u(M) U_u(M, x)$. The normalised profile, $U_u(M, x)$, satisfies the following condition:

$$\int U_u(M, x) d^3x = 1. \quad (1.35)$$

We now consider the correlation between two fields, θ_u and θ_v . As shown in Section 1.2, the correlation between two fields in Fourier space is directly related to the power spectrum, P_{uv} , via:

$$\langle \hat{\theta}_u(\mathbf{k}) \hat{\theta}_v^*(\mathbf{k}') \rangle = (2\pi)^3 \delta_D(\mathbf{k} - \mathbf{k}') P_{uv}(k), \quad (1.36)$$

Using Eq. 1.33, we can replace the RHS of Eq. 1.36 with:

$$\langle \hat{\theta}_u(\mathbf{k}) \hat{\theta}_v^*(\mathbf{k}') \rangle = \left\langle \sum_{i,j} e^{-i\mathbf{k} \cdot \mathbf{x}_i} e^{i\mathbf{k}' \cdot \mathbf{x}_j} N_i N_j \hat{W}_{u,i}(M, k) \hat{W}_{v,j}^*(M, k') \right\rangle. \quad (1.37)$$

The sum above can be split into two components—one where $i = j$, and one where $i \neq j$. The case $i = j$ captures correlations within a single halo, and is referred to as the 1-halo

term. In contrast, the 2-halo term ($i \neq j$) accounts for correlations between different halos. The 1-halo term is given by:

$$\langle \hat{\theta}_u(\mathbf{k}) \hat{\theta}_v^*(\mathbf{k}') \rangle = \langle e^{-i(\mathbf{k}-\mathbf{k}') \cdot \mathbf{x}_j} N_i \hat{W}_{u,i}(M, k) \hat{W}_{v,i}(M, k') \rangle. \quad (1.38)$$

Note that we set $N_i^2 = N_i$. Following the same reasoning from Eq. 1.34, the 1-halo term becomes:

$$\langle \hat{\theta}_u(\mathbf{k}) \hat{\theta}_v^*(\mathbf{k}') \rangle = \int_0^\infty n(M) dM \int e^{-i(\mathbf{k}-\mathbf{k}') \cdot \mathbf{x}_i} \hat{W}_u(M, k) \hat{W}_v(M, k') d^3x. \quad (1.39)$$

Comparing the expression above to Eq. 1.36, we can write the 1-halo power spectrum, P_{uv}^{1h} , as:

$$P_{uv}^{1h}(k) = \int_0^\infty \hat{W}_u(M, k) \hat{W}_v(M, k) n(M) dM. \quad (1.40)$$

To compute the 2-halo term, we first recast N_i and N_j as $N(M_i, \mathbf{x}_i)$ and $N(M_j, \mathbf{x}_j)$, respectively. Since we are considering correlations between distinct halos, the expectation value in Eq. 1.37 can be factorized into a product of two separate expectations, yielding:

$$\begin{aligned} \langle \hat{\theta}_u(\mathbf{k}) \hat{\theta}_v^*(\mathbf{k}') \rangle &= \int_0^\infty dM_1 \int_0^\infty n(M_1) n(M_2) \hat{W}_u(M_1, k) \hat{W}_v(M_2, k) dM_2 \times \\ &\quad \int d^3x \int d^3x' e^{-i\mathbf{k} \cdot \mathbf{x}} e^{i\mathbf{k}' \cdot \mathbf{x}'} \langle N(M_1, \mathbf{x}) N(M_2, \mathbf{x}') \rangle, \end{aligned} \quad (1.41)$$

where M_1 and M_2 are the masses of halos 1 and 2, respectively. The second factor/double integral can be recognized as the halo-halo correlation function in Fourier space:

$$\langle \hat{N}(M_1, \mathbf{k}) \hat{N}^*(M_2, \mathbf{k}') \rangle = (2\pi)^3 \delta_D(\mathbf{k} - \mathbf{k}') P_{hh}(M_1, M_2, k), \quad (1.42)$$

where P_{hh} is the power spectrum between halos. Note that at small k , the 1-halo power spectrum becomes constant, adding ‘shot noise’ to the total power spectrum. This shot noise is typically removed from P_{hh} . Thus, the 2-halo power spectrum is:

$$P_{\text{uv}}^{2\text{h}}(k) \int_0^\infty \int_0^\infty \hat{W}_{\text{u}}(M_1, k) \hat{W}_{\text{v}}(M_2, k) n(M_1) n(M_2) P_{\text{hh}}(M_1, M_2, k) dM_1 dM_2. \quad (1.43)$$

Halos are biased tracers of the underlying matter field. Leveraging this bias, the power spectrum between halos can be approximated using the matter power spectrum:

$$P_{\text{hh}}(M_1, M_2, k) = b(M_1)b(M_2)P_{\text{mm}}^{\text{lin}}(k)[1 + \beta^{\text{nl}}(M_1, M_2, k)]. \quad (1.44)$$

The parameter b is the linear halo bias, and β^{nl} captures any non-linearities. Often, the non-linear halo bias is ignored. Note that there is no analytic form for β^{nl} . In this work, the power spectrum between halos comes directly from simulations, so we do not require an expression for P_{hh} . The total power spectrum is a sum of the 1- and 2-halo power spectra:

$$P_{\text{uv}}(k) = P_{\text{uv}}^{1\text{h}}(k) + P_{\text{uv}}^{2\text{h}}(k). \quad (1.45)$$

In the derivations above, we assumed that mass is the only halo property relevant for clustering. However, this assumption may not hold due to assembly bias, which refers to the dependence of clustering on properties other than halo mass (e.g., assembly history and environment, [Croton et al., 2007](#)). Such properties are often referred to as ‘secondary’, whereas halo mass is considered a primary property. Assembly bias manifests in two forms:

- **Halo Assembly Bias (HAB)**: The dependence of halo clustering on secondary properties.
- **Galaxy Assembly Bias (GAB)**: The dependence of galaxy clustering on secondary properties.

We do not incorporate assembly bias in our models, as there does not exist an analytic description of this effect. Modelling assembly bias requires access to simulations for individual halo properties, such as the formation time: the scale factor at which a halo first obtains half of its present-day mass [Xu et al. \(2021\)](#). We briefly discuss one semi-analytic approach in Section 1.4. Note that at present, observational evidence for assembly bias is inconclusive.

1.4 Halo Occupation Distributions

To use the halo model for [GGL](#) and clustering predictions, we require some mechanism that relates galaxies to their host dark matter halos, known as the ‘galaxy–halo connection’ ([Wechsler & Tinker, 2018](#)). We adopt a [Halo Occupation Distribution \(HOD\)](#) as our galaxy-halo connection model, specifically using the vanilla [HOD](#) introduced by [Zheng et al. \(2005\)](#). [HODs](#) are empirical models that connect galaxies and dark matter halos through statistical prescriptions, effectively bypassing the complexities of galaxy formation physics. They are designed to model the distribution of luminosity/flux/mass-limited galaxy samples, with each halo treated as distinct (i.e., no subhalos). In the adopted [HOD](#), each halo can host a central galaxy and a distribution of satellite galaxies. Central galaxies reside at the centre of a dark matter halo (although this assumption can be relaxed, as we show later on), while satellite galaxies occupy the rest of the halo. Furthermore, central galaxies tend to be more massive, luminous, and older compared to satellite galaxies in hydrodynamical simulations ([Zheng et al., 2005](#)).

According to the [HOD](#), the mean number of galaxies in halos of mass M is given by the sum of the expected numbers of central galaxies and satellite galaxies:

$$\langle N_g(M) \rangle = \langle N_c(M) \rangle + \langle N_s(M) \rangle. \quad (1.46)$$

The expected number of central galaxies is parametrized as follows:

$$\langle N_c(M) \rangle = \frac{1}{2} \left(1 + \operatorname{erf} \left[\frac{\log(M/M_{\min})}{\sigma_{\log(M)}} \right] \right), \quad (1.47)$$

where $M_{\min}$ is the minimum mass of halos that can host a central galaxy above some luminosity threshold, assuming no scatter in the galaxy luminosity–halo mass relationship ($\sigma_{\log(M)} \rightarrow 0$). It is also the value where $\langle N_c(M) \rangle = 0.5$. At the low mass end, $\langle N_c(M) \rangle$ goes to 0. Conversely, $\langle N_c(M) \rangle$ asymptotes towards 1 at the high mass end. The form of $\langle N_c(M) \rangle$ comes from the assumption that central galaxy luminosity, L_c , at fixed halo mass follows a Gaussian distribution:

$$P(\log(L_c)|M) = \frac{1}{\sqrt{2\pi}\sigma_{\log(L)}} \exp \left(-\frac{[\log(L_c) - \log(\langle L_c(M) \rangle)]^2}{2\sigma_{\log(L)}^2} \right). \quad (1.48)$$

The expected number of satellite galaxies assumes the following form:

$$\langle N_s(M) \rangle = \langle N_c(M) \rangle \lambda_s(M) = \langle N_c(M) \rangle \left(\frac{M - \kappa M_{\min}}{M_1} \right)^\alpha. \quad (1.49)$$

The equation above is valid on the domain $M > \kappa M_{\min}$. If $M \leq \kappa M_{\min}$, $\langle N_s(M) \rangle = 0$. Thus, $\kappa M_{\min}$ acts as a truncation mass. α controls how steep (shallow) the slope of $\langle N_s(M) \rangle$ is at the high (low) mass end, while M_1 vertically shifts $\langle N_s(M) \rangle$ in logspace. The above parametrization was found to capture the behaviour of satellite galaxies in smoothed particle hydrodynamics (SPH) simulations (Zheng et al., 2007).

Eqs. 1.47 and 1.49 compute the mean numbers of galaxies in halos of mass M . However, they do not determine the exact number of galaxies to place in a given halo. We must also know the second moments of the occupation numbers (N_c and N_s) to model the [Two-Point Correlation Function \(2PCF\)](#) of galaxies. Hence, we need to define probability distributions for the numbers of central and satellite galaxies. We assign the following distributions: central galaxies are drawn from a Bernoulli distribution with mean $\langle N_c(M) \rangle$, and satellite galaxies are drawn from a Poisson distribution with mean $\lambda_s(M)$:

$$P(N_s) = (\lambda_s)^{N_s} \frac{\exp(-\lambda_s)}{N_s!}. \quad (1.50)$$

The assumption that the satellite galaxy occupation number follows a Poisson distribution is consistent with theoretical predictions (Zheng et al., 2007). As in Miyatake et al. (2022a), we impose a ‘central condition’ such that satellite galaxies cannot occupy a halo without a central galaxy; i.e., $P(N_s = 0 | N_c = 0) = 1$. This condition is often imposed since central galaxies typically form before satellite galaxies in simulations.

The [HOD](#) previously presented is not the only one in existence. For example, the [HOD](#) from Leauthaud et al. (2011) is quite similar to ours, although the expression for $\langle N_c(M) \rangle$ is more generalized. In going from Eq. 1.48 to Eq. 1.47, a power law is assumed for the luminosity-halo mass relationship. The [HOD](#) in Leauthaud et al. (2011) does not make this explicit assumption—instead, any functional form can be adopted. Under this [HOD](#), the expected number of satellite galaxies is modulated by the expected number of central galaxies and scaled by a power law factor. The satellite galaxy occupation statistics are also Poissonian. The ‘iHOD’ model in Zu & Mandelbaum (2015) further builds upon the framework established by Leauthaud et al. (2011), and can model volume-incomplete galaxy samples in a statistically self-consistent way using the observed stellar mass function as an input. Note that standard [HODs](#) like ours inherently assume volume completeness. This assumption can be relaxed via an incompleteness function with free parameters. The

HODs from [Tinker et al. \(2013\)](#) and [Zu & Mandelbaum \(2016\)](#) incorporate colour and star formation rate.

Going back to assembly bias, the standard HOD can be extended to incorporate secondary halo properties. Such models are referred to as ‘decorated’ HODs ([Hearin et al., 2016](#)). In other words, the occupation statistics are decorated with assembly bias. A practical, semi-analytic implementation of GAB is presented in [Xu et al. \(2021\)](#), upon which the ABACUSHOD module is built ([Yuan et al., 2022](#)). For context, this module produces mock galaxy catalogues using halos from the AbacusSummit simulations ([Maksimova et al., 2021](#); [Garrison et al., 2021](#)). The methodology is as follows: halo concentration, c , and local over-density, δ , are computed for each halo within a narrow mass bin. These properties are then ranked from 0 to 1. Using these normalised ranks, the parameters $\log(M_{\min})$ and M_1 are modified as:

$$\log(M_{\min}^{\text{mod}}) = \log(M_{\min}) + A_c(c^{\text{rank}} - 0.5) + B_c(\delta^{\text{rank}} - 0.5), \quad (1.51)$$

$$\log(M_1^{\text{mod}}) = \log(M_1) + A_s(c^{\text{rank}} - 0.5) + B_s(\delta^{\text{rank}} - 0.5), \quad (1.52)$$

where A_c , B_c , A_s , and B_s are free parameters. Local over-density has been shown to fully capture GAB, whereas halo concentration is less informative ([Xu et al., 2021](#)).

This section outlines how HODs can be used to model the galaxy–halo connection. A distinct alternative is Abundance Matching (AM; [Marinoni & Hudson, 2002](#)), or its extension to subhalos, known as Subhalo Abundance Matching (SHAM; [Vale & Ostriker, 2004](#)). In this framework, galaxies are matched with dark matter halos based on the abundance of selected properties ([Wechsler & Tinker, 2018](#)). For example, the most massive galaxy can be assigned to the most massive halo, the second most massive galaxy to the second most massive halo, and so forth. Each subhalo typically hosts a single galaxy ([Wechsler & Tinker, 2018](#)). Although AM and SHAM are formally parameter-free, some assumptions must be made to link galaxies and dark matter halos. Specifically, one must choose a galaxy property (e.g., stellar mass, luminosity) and a corresponding halo property (e.g., mass, peak circular velocity), and understand the scatter between these properties ([Wechsler & Tinker, 2018](#)). Of course, one could also take a fully numerical route by simultaneously simulating the clustering of dark matter and galaxies, from which 2PCFs can be directly computed.

1.5 Feedback

As previously mentioned, the linear perturbation theory outlined in Section 1.2 cannot describe structure on scales where non-linear physics dominates. One key non-linear effect is baryonic feedback, which includes processes such as star formation, radiative cooling, and galactic winds [Medlock et al. \(2025\)](#). In addition to dark matter, halos contain gas and stars/galaxies, all of which interact and redistribute matter. For example, active galactic nuclei (AGN) and supernovae can push gas towards the outskirts of a halo, reducing the central matter concentration ([Schneider et al., 2019](#)). If this gas becomes unbound from the halo, then power—referring to the amplitude of P_{mm} —is also suppressed on large scales ([Medlock et al., 2025](#)). That being said, the central matter concentration can increase due to the infall of baryons (i.e., during central galaxy formation) towards the halo centre ([Blumenthal et al., 1986](#)). While these feedback processes are occurring, the baryons ‘pull’ on the dark matter.

Feedback can be readily implemented in hydrodynamical cosmological simulations. However, the treatment of feedback varies across simulation suites, resulting in inconsistent predictions ([Schneider et al., 2019](#)). As an alternative, one can take the outputs of dark-matter-only N-body simulations and mimic feedback by ‘displacing’ particles according to an analytic prescription, enabling marginalization over feedback parameters. We incorporate feedback using the **Baryon Correction (BC)** model presented by [Schneider et al. \(2019\)](#) and ([Shirasaki et al., 2024](#)), which modifies the 1-halo term in the halo-matter cross-power spectrum, P_{hm} . This model is competitive with hydrodynamical simulations, and can reproduce the gas profiles from X-ray observations [Schneider et al. \(2019\)](#). We outline our procedure for modifying P_{hm} in Section 4.3. Below, we motivate the equation for the displacement of a dark/**Collisionless Matter (CLM)** mass shell due to baryons.

Early work on feedback explored how baryons impact the matter distribution within individual galaxy-halo systems, as opposed to halos containing many galaxies. Initially, dark matter mass shells were thought to *only* contract in response to the condensation of baryons towards the halo centre, forming dense central cores ([Blumenthal et al., 1986](#); [Ryden & Gunn, 1987](#); [Flores et al., 1993](#)). An analytic framework to describe this process was first established by [Zeldovich et al. \(1980\)](#) and later popularized by [Blumenthal et al. \(1986\)](#). This framework, known as the ‘adiabatic contraction’ model, relies on the adiabatic invariant for periodic orbits: $\oint pdq$, where p is the particle’s conjugate momentum and q is a generalized coordinate. Assuming the particle orbits within a spherically symmetric mass distribution, $M(r)$, the quantity $rM(r)$ remains constant, provided the mass enclosed by the particle varies slowly over time. Let M_i denote the pre-baryonification mass distribution, M_{DM} the dark matter mass distribution, and M_{b} the baryonic mass distribution.

Then, the initial and final radial positions of dark matter mass shells (relative to the halo centre) are related via:

$$r_f[M_b(r_f) + M_{\text{DM}}(r_i)] = r_i M_i(r_i),$$

$$\frac{r_f}{r_i} = \frac{M_i(r_i)}{M_{\text{DM}}(r_i) + M_b(r_f)}. \quad (1.53)$$

The expression above defines a one-to-one mapping between r_i and r_f , thereby prohibiting shell-crossing. It states that if mass is added inside a dark matter shell, the shell must respond by contracting.

The adiabatic contraction can be relaxed by introducing the nuisance parameters $\mathcal{A}$ and n as follows:

$$\frac{r_f}{r_i} = \mathcal{A} \left(\left[\frac{M_i(r_i)}{M_{\text{DM}}(r_i) + M_b(r_f)} \right]^n - 1 \right) + 1. \quad (1.54)$$

Setting $\mathcal{A}$ and n to unity yields the original expression. [Abadi et al. \(2010\)](#) found that the above parametrization with $\mathcal{A} = 0.3$ and $n = 2$ sufficiently captures the response of a dark matter halo to the formation of a central galaxy in gasdynamical cosmological simulations. In contrast, the standard adiabatic contraction model consistently overestimated the dark matter mass profiles obtained from these simulations—a discrepancy also observed by [Gnedin et al. \(2004\)](#). Furthermore, [Gustafsson et al. \(2006\)](#) reported that the adiabatic contraction model significantly over-predicts the central dark matter density. Despite its initial success in reproducing the observed shapes of spiral galaxy rotation curves ([Blumenthal et al., 1986](#)), the adiabatic contraction model has since been phased out due the reasons mentioned above. Accordingly, the [BC](#) model employed in this work adopts Eq. 1.54.

1.6 Gravitational Lensing

One of our key observables is the [GGL](#) signal. Before describing [GGL](#), we revisit the foundational principles of gravitational lensing. In this phenomenon, light from a source is coherently distorted by a lens ([Prat & Bacon, 2026](#)). The lens and source can be characterized as follows:

- **Lens:** Massive foreground object whose gravitational field bends and distorts light from a source.
- **Source:** Background object whose light is coherently distorted and magnified by the lens.

Photons travel along null geodesics. Consequently, their trajectories are perturbed by the gravitational fields of intervening masses. A key quantity in gravitational lensing is the deflection angle of light, $\delta\theta$. We derive the deflection angle under the Newtonian approximation in Appendix A.

An essential quantity in GGL is the critical surface density, Σ_{crit} . We now demonstrate how this quantity naturally emerges from the lens–source geometry shown in Fig. 1.6, under the thin-lens approximation. Note that the angle $\tilde{\alpha}$ represents the deflection angle (Dodelson, 2017). Assuming Euclidean distances, we obtain the following relationships:

$$\alpha(\theta) = \left(\frac{D_{\text{ls}}}{D_{\text{s}}} \right) \tilde{\alpha}(\theta), \quad (1.55)$$

$$\beta(\theta) = \theta - \alpha(\theta). \quad (1.56)$$

Here, $\alpha(\theta)$ is the ‘reduced’ deflection angle, and $\beta(\theta)$ is the observed angular separation between the lens and source in the absence of deflection. Additionally, D_{l} is the angular diameter distance to the lens, D_{s} the angular diameter distance to the source, and D_{ls} the angular diameter distance between the lens and source. Eq. 1.56 is referred to as the lens equation, and it has multiple solutions in the strong lensing regime (Abdo, nd). Combining Eq. A.12, Eq. 1.55, and Eq. 1.56, and setting $b = \theta D_{\text{l}}$ via the small angle approximation, the lens equation can be written as:

$$\beta(\theta) = \theta - \frac{D_{\text{ls}} 4GM}{D_{\text{l}} D_{\text{s}} c^2 \theta}. \quad (1.57)$$

In the case where the source lies directly behind the lens, $\beta = 0$, and we have:

$$\theta = \sqrt{\frac{4GM D_{\text{ls}}}{c^2 D_{\text{l}} D_{\text{s}}}} \equiv \theta_{\text{E}}, \quad (1.58)$$

where θ_{E} is the Einstein radius. An observer witnesses strong gravitational effects (e.g., multiple images, arcs) when a background object lies within the Einstein radius of a lens

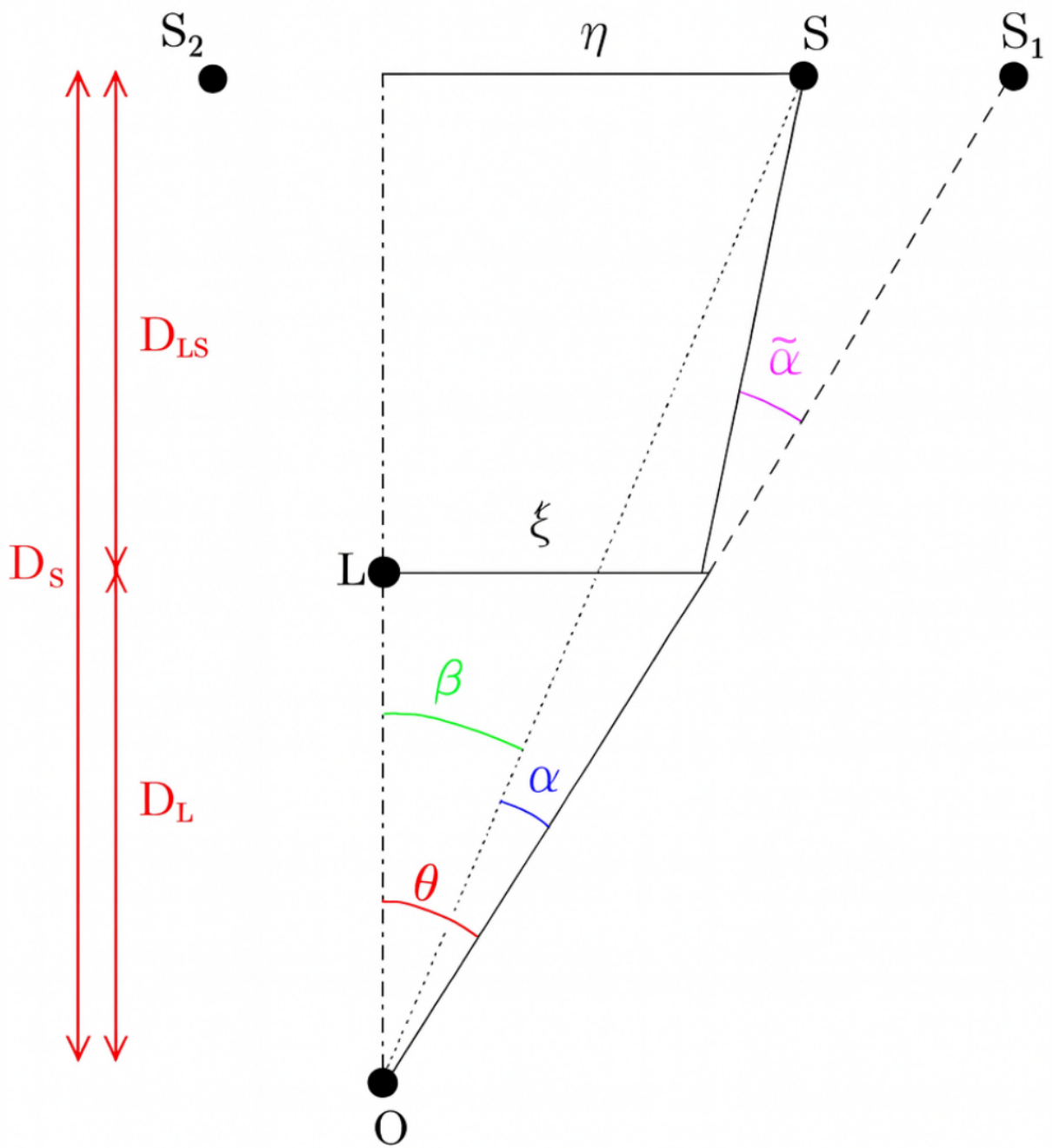


Figure 1.1: Geometry of a lens–source system in the thin-lens approximation, wherein distances between objects are much larger than the extent of the lens mass distribution along the line of sight (LOS). Taken from [Abdo, nd.](#)

(Dodelson, 2017). The most extreme lensing effect is an Einstein ring, which occurs when $\beta = 0$ (Dodelson, 2017).

So far, we have treated the lens as a point mass. We relax this assumption by treating the lens as an extended mass with a 2D surface density, Σ . We approximate the mass enclosed within a radius R using $M \approx \pi \Sigma(R) R^2$, and set $R = \theta D_l$. Plugging this into Eq. 1.58 yields:

$$\frac{\theta_E}{\theta} = \sqrt{\frac{4\pi G \Sigma D_l D_{ls}}{c^2 D_s}}. \quad (1.59)$$

Then, the critical surface density is defined as:

$$\left(\frac{\theta_E}{\theta}\right)^2 = \frac{\Sigma}{\Sigma_{\text{crit}}}, \quad (1.60)$$

$$\Sigma_{\text{crit}} = \frac{c^2 D_s}{4\pi G D_l D_{ls}}, \quad (1.61)$$

which depends on the redshift of both the lens and source. Like θ_E , Σ_{crit} serves as a threshold for strong lensing effects. Additionally, its inverse is referred to as the ‘lensing efficiency’ (Prat & Bacon, 2026). If the surface mass density of a lens exceeds Σ_{crit} for a given lens–source pair, then the source is strongly lensed (Dodelson, 2017). Eq. 1.61 reveals several key points about the lens–source configuration:

- The lens cannot be located directly in front of the observer ($z_l/D_l = 0$).
- A background source must lie at a greater redshift/distance than the lens.
- The strongest lensing effects occur when both the lens and source are distant from the observer, and are well-separated from each other. Usually, the lensing efficiency is greatest when the lens is halfway between the observer and source (Prat & Bacon, 2026).

1.7 Galaxy–Galaxy Lensing

GGL probes the projected matter distribution around foreground lens galaxies using the coherent distortion of light from background source galaxies (Singh et al., 2017). Thus,

both lenses and sources are galaxies. It is also a type of weak gravitational lensing, meaning that measurements of this phenomenon's effects are statistical in nature (Bartelmann & Schneider, 2001).

The GGL signal is formally defined as the mean Excess Surface Density (ESD) around foreground lenses:

$$\Delta\Sigma(r_p) = \overline{\Sigma(<r_p)} - \langle\Sigma(r_p)\rangle. \quad (1.62)$$

The first term in the expression above is the mean surface density around lenses inside a circle of projected radius r_p . The second term is the expected surface density at r_p .

In practice, we cannot measure the ESD around real galaxies using Eq. 1.62 since dark matter is not directly observable. We can, however, infer the ESD by measuring the shapes of background source galaxies. The mass distribution of a lens system induces a tangential shear, γ_t , in the images of background source galaxies (Prat & Bacon, 2026). This shear is directly related to the ESD via:

$$\gamma_t = \Sigma_{\text{crit}}^{-1}(z, z_s)\Delta\Sigma(r_p). \quad (1.63)$$

Images of gravitationally lensed objects are distorted in two distinct ways. Isotropic distortions are quantified by the convergence, κ (Prat & Bacon, 2026). A positive κ corresponds to an enlarged image, as shown in Fig. 1.2. On the other hand, anisotropic distortions are quantified by the shear, γ (Prat & Bacon, 2026). The shear has two components, γ_1 and γ_2 , which form a complex quantity $\gamma = \gamma_1 + i\gamma_2$ (Prat & Bacon, 2026). The real component, γ_1 , characterizes stretching along the Cartesian axes, while γ_2 characterizes stretching at 45 degrees to those axes. A positive γ_1 indicates stretching along the x-axis, whereas a positive γ_2 indicates stretching along the axis 45 degrees from the x-axis. Together, convergence and shear determine the overall magnification, μ , of a background object (Lamman et al., 2024). The magnification is given by the determinant of the matrix describing the mapping between lensed and unlensed coordinates:

$$A = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix}, \quad (1.64)$$

where $\det A = (1 - \kappa)^2 - |\gamma|^2 \approx 1 - 2\kappa$ since $\kappa, \gamma \ll 1$ in the weak-lensing regime.

We use ellipticity, ϵ , as a proxy for the shear in Eq. 1.63. Ellipticity is defined as:

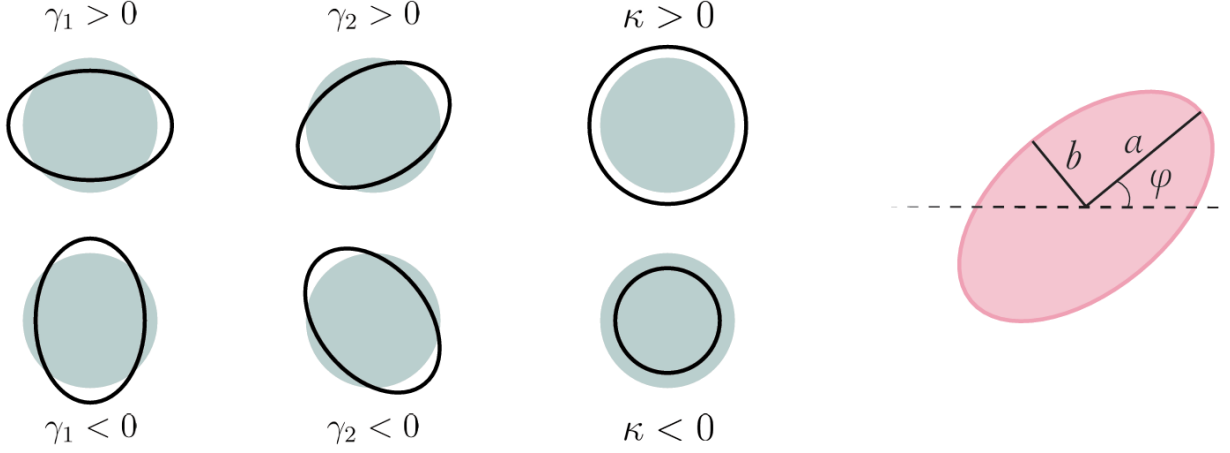


Figure 1.2: Convergence (κ) and shear (γ) applied to a circle. The right hand side shows an ellipse with semi-major and semi-minor axes a and b , respectively. Taken from [Prat & Bacon \(2026\)](#).

$$\epsilon = \frac{a - b}{a + b} e^{2i\varphi}, \quad (1.65)$$

where a and b are the semi-major and semi-minor axes, respectively. The parameter φ is the galaxy's orientation angle with respect to the x-axis, as shown in Fig. 1.2. In applying the expression above to galaxies, we implicitly assume that they are originally circular. This assumption is valid, provided one has a statistically significant number of source galaxies.

Various software packages can be used to determine galaxy shapes. For example, one can fit a Sérsic profile ([Sersic, 1968](#)) to a galaxy's surface brightness distribution. Alternatively, galaxy shapes can be modelled as a sum of Gaussians, which forms the basis of NGMIX ([Sheldon, 2015](#)). This approach/software is implemented in our weak-lensing pipeline, SHAPEPIPE ([Farrens et al., 2022](#); [Guinot et al., 2022](#)). Note that a galaxy's ellipticity can also be measured via the quadrupole moments of its surface brightness distribution ([Bartelmann & Schneider, 2001](#)). Both the fitting approach and quadrupole moments approach have their respective advantages and disadvantages. The fitting method involves modelling a profile for each galaxy and performing a χ^2 minimization, which is computationally expensive. The quadrupole moments technique requires careful handling of galaxies that are close together on the sky, particularly when their surface brightness distributions overlap.

Using Eq. 1.65, ellipticity and shear are related via:

$$\epsilon = \frac{\gamma}{1 - \kappa} \equiv g, \quad (1.66)$$

where g is referred to as the reduced shear (Prat & Bacon, 2026). Fortunately, κ is usually very small in the weak-lensing regime, so the reduced shear is often a good approximation of the true shear (Prat & Bacon, 2026). Note that the ellipticity components follow the same definitions as the shear components. Additionally, both shear and ellipticity are spin-2 quantities, meaning they are invariant under rotations of π (Lamman et al., 2024).

If ϕ is the position angle of the source galaxy relative to the lens galaxy, the tangential ellipticity can be computed using:

$$\epsilon_t = -\text{Re}[\epsilon e^{-2i\phi}], \quad (1.67)$$

$$\epsilon_t = -\epsilon_1 \cos(2\phi) - \epsilon_2 \sin(2\phi). \quad (1.68)$$

Note that the minus sign ensures that $\epsilon_t > 0$ when a source galaxy is tangent to the lens galaxy. The tangential ellipticity is analogous to ϵ_1 in our rotated reference frame. There is also a cross component, $\epsilon_\times$, which is analogous to ϵ_2 . In the absence of systematics, $\langle \epsilon_\times \rangle = 0$ (Lamman et al., 2024).

1.8 Galaxy–Galaxy Lensing Systematics

GGL measurements are inherently challenging for many reasons, including the numerous systematics involved. In this section, we outline the systematics most relevant to our analysis.

- **Lens magnification:** Intervening Large-Scale Structure (LSS) between the observer and the lens sample magnifies the lenses while also inducing a shear in background sources (Unruh et al., 2020). This magnification results in two competing effects. First, the local stretching of the sky reduces the observed number density of lenses (Prat et al., 2022). However, lenses whose fluxes previously fell just below the detection threshold may now be observed since surface brightness is conserved¹, increasing

¹Surface brightness is conserved since flux depends on distance; if an object is brought closer to an observer, its flux increases, but so does its surface area, leaving the surface brightness unchanged. Consequently, magnification increases the observed flux of objects.

the observed number density (Prat et al., 2022). This change in observed number density biases GGL measurements; an increase artificially boosts the signal, while a decrease reduces the signal. Lenses that enter the final sample due to LSS magnification are located in regions where background sources are more strongly sheared. Whether the observed lens number density increases or decreases depends on the particular sample—specifically, on the number of lenses just below the flux limit (Unruh et al., 2020).

We model the lens magnification contribution to the GGL signal following Unruh et al. (2020), where the observed tangential shear is given by:

$$\gamma_t(\theta|z_l, z_s) = \gamma_t^{\text{nomagn}}(\theta|z_l, z_s) + 2(\alpha_l(z_l) - 1)\gamma_t^{\text{LSS}}(\theta|z_l, z_s). \quad (1.69)$$

In the expression above, θ is the angular version of r_p , z_l and z_s are the lens and source redshifts, respectively, γ_t^{nomagn} denotes the tangential shear without lens magnification, γ_t^{LSS} is the tangential shear from LSS, and α_l is the slope of the lens number counts at the flux limit. Note that the factor of $2(\alpha_l(z_l) - 1)$ arises from the weak-lensing approximation $\mu \approx 1 + 2\kappa$. If $\alpha_l > 1$, the lens number counts are enhanced; conversely, they are suppressed when $\alpha_l < 1$ (Unruh et al., 2020). There is no lens magnification bias when $\alpha_l = 1$. The LSS tangential shear is modelled as:

$$\begin{aligned} \gamma_t^{\text{LSS}}(\theta|z_l, z_s) = & \frac{9H_0^3\Omega_{m,0}^2}{8\pi c^3} \int_0^\infty J_2(\ell\theta)\ell d\ell \int_0^{z_l} (1+z)^2 \frac{H_0}{H(z)} \times \\ & \frac{D_A(z, z_l)D_A(z, z_s)}{D_A(z_l)D_A(z_s)} P_{\text{mm}}\left(\frac{\ell + 1/2}{D_A(z)[1+z]}; z\right) dz, \end{aligned} \quad (1.70)$$

where J_2 is the 2nd-order Bessel function of the first kind, and $D_A(z_l) = D_l$. The expression above comes from the convergence power spectrum, which is related to P_{mm} through the Limber approximation (Limber, 1953; Kilbinger et al., 2017). The Limber approximation projects 3D power spectra into 2D.

- **Intrinsic alignments:** The observed shapes of source galaxies can be decomposed into two components: an intrinsic component, and an extrinsic component induced by gravitational lensing (Lamman et al., 2024). The intrinsic shape consists of a dominant, random component (which contributes to shape noise) and a secondary component arising from alignments with local gravitational tidal fields (Troxel & Ishak, 2015). Consequently, these alignments are referred to as ‘Intrinsic Alignments’ (IAs).

In [IA](#) notation, the galaxy shape-density correlation, $\langle \epsilon_i n_j \rangle$, is given by:

$$\langle \epsilon_i n_j \rangle = \langle G_i g_j \rangle + \langle I_i g_j \rangle + \langle G_i m_j \rangle + \langle I_i m_j \rangle, \quad (1.71)$$

where i, j are different radial bins, ϵ and n are the observed source galaxy shape and lens number density, respectively, G is the gravitationally lensed shape component, I is the tidal field intrinsic shape component, g is the lens number density in the absence of magnification, and m is the magnification contribution to the observed lens number density. Within this formalism, $\langle G_i g_j \rangle$ is the [GGL](#) signal ([Lamman et al., 2024](#)). Note that the third term in the expression above is the lens magnification contribution to the [GGL](#) signal.

The primary [IA](#) contaminant to the [GGL](#) signal is the second term ([Cross & Sánchez, 2024](#)), which is the correlation between intrinsic lens number density and intrinsic source orientation. Such correlations occur when lenses and sources are physically associated, producing radial alignments that dilute the mean tangential shear ([Toshiki, 2025](#)). Ig contamination can be mitigated through tomographic binning, ensuring that the lens and source redshift distributions are sufficiently separated ([Lamman et al., 2024](#)). This reduces the likelihood of physical associations between the two samples. Note that this strategy is not guaranteed to eliminate Im contamination ([Cross & Sánchez, 2024](#)). Also, binning must be performed carefully to avoid significantly reducing the [Signal-To-Noise Ratio](#) (SNR) of measurements. Alternatively, one could model the [IA](#) signal if the source redshifts are poorly constrained. Popular models include Tidal Alignment and Tidal Torquing (TATT), Linear Alignment (LA), and Non-linear Linear Alignment (NLA; [Cross & Sánchez, 2024](#)). However, the [IA](#) amplitude in these models is degenerate with cosmological parameters, and the predicted [IA](#) contributions vary across different models ([Bigwood et al., 2025](#)). In this work, we assume an [IA](#) amplitude of zero.

- **Source redshift uncertainty:** [GGL](#) measurements rely on a large number of source galaxies to beat down the shape noise, which stems from the intrinsic, random shapes/orientations of galaxies. Given this large number of source galaxies, spectroscopic redshifts are impractical, necessitating the use of less precise and potentially biased photometric redshifts. As a result, photometric redshifts are often calibrated using a catalogue that contains inaccurate photometric redshifts and the corresponding ‘true’ redshifts. This approach can be extended to scenarios where no reliable photometric redshifts exist, as in our analysis. However, if the mean of the effective source redshift distribution is misestimated, the resulting [GGL](#) signal will be biased. For example, an overestimated mean redshift results in an underestimated critical

surface density, suppressing the [GGL](#) signal; the opposite occurs when the mean is underestimated. This systematic can be accounted for by marginalizing over the residual bias in the mean source redshift ([Huterer et al., 2006](#)), Δz_{ph} , often requiring an informative prior. We perform an analytic marginalization and propagate the uncertainty in Δz_{ph} into our [GGL](#) covariance matrix, which is fully described in [Section 3](#).

- **Multiplicative shear calibration bias:** Galaxy shape measurement algorithms often produce biased ellipticity estimates, such that the mean measured ellipticity, $\langle \epsilon_i^{\text{meas}} \rangle$, is given by:

$$\langle \epsilon_i^{\text{meas}} \rangle = (1 + m_i) \langle \gamma_i \rangle + c_i, \quad (1.72)$$

where $i \in \{1, 2\}$, and m_i and c_i are multiplicative and additive biases, respectively. The multiplicative bias is commonly referred to as the ‘m-bias’. Note that the expression above assumes that the intrinsic ellipticity component vanishes ([Guinot et al., 2022](#)).

The SHAPEPIPE weak-lensing pipeline employs the METACALIBRATION method ([Huff & Mandelbaum, 2017](#)) to calibrate shear estimates. Unlike simulation-based approaches, which are time-consuming, this technique uses the actual image data. Essentially, artificial shears are applied to each galaxy, and the shape measurement algorithm’s response is captured in a matrix $\mathbf{R}$, defined as:

$$R_{ij} = \frac{\partial \epsilon_i^{\text{meas}}}{\partial \gamma_j} \approx \frac{\epsilon_i^+ - \epsilon_i^-}{2\Delta g_j}, \quad (1.73)$$

where $i, j \in \{1, 2\}$, and $\epsilon_i^\pm$ denotes the i th ellipticity component measured after the galaxy image is sheared by $\pm \Delta g_j$. For our shape catalogue, the matrix response is reduced to a scalar via:

$$R \approx \frac{R_{11} + R_{22}}{2}. \quad (1.74)$$

The ellipticities in our shape catalogue are divided by the weighted mean scalar response, $\langle R \rangle$. This treatment is often preferred over using individual responses, as the inverse of a noisy quantity is biased. Alternatively, one could compute the average tangential scalar response for each lens–source pair and divide the [GGL](#) signal by the weighted mean of these responses; both approaches are used ([Prat et al., 2022](#)). Any

residual multiplicative bias can be accounted for by marginalizing over a nuisance parameter Δm , which scales the predicted [GGL](#) signal by a factor of $(1 + \Delta m)$ ([Huterer et al., 2006](#)). This typically requires an informative prior on Δm . Our multiplicative bias is calibrated to the few percent level, so we do not marginalize over a Δm . We also neglect the additive bias in [Eq. 1.72](#) since it is uncorrelated with lens positions and vanishes.

1.9 Galaxy Clustering

Our other key observable is galaxy clustering, which we briefly describe in this section using [Mo et al. \(2010\)](#). The projected correlation function, or the galaxy clustering signal, quantifies the clustering of galaxies across different projected scales. When galaxies are imaged, their angular separations are readily obtained. Converting these angular separations into ‘physical’² ones requires spectroscopic redshifts, from which comoving distances can be derived via [Eq. 1.15](#). However, the observed redshift of a galaxy also contains a contribution from its peculiar velocity. At very low redshifts, the total recessional velocity, v_s , is given by:

$$v_s = v_H + v_r, \quad (1.75)$$

where v_H is the Hubble flow velocity, and v_r is the radial component of the peculiar velocity. Peculiar velocities become subdominant at higher redshifts since the Hubble flow velocity is large.

Because of peculiar velocities, the galaxy [2PCF](#) is always measured in redshift space, as opposed to configuration space. We can move from redshift space to configuration space by integrating the [2PCF](#) along the LOS, defining the projected correlation function. The upper integration limit should be large to minimize [Redshift Space Distortions \(RSDs\)](#), which are described in the next paragraph. We select an upper limit of $100 h^{-1}$ Mpc.

[RSDs](#) encompass two main effects: the Kaiser effect, and [Fingers of God \(FoG\)](#). The Kaiser effect occurs on scales larger than the typical sizes of galaxy clusters, while [FoG](#) occur on intra-cluster scales. On large scales, peculiar velocities are primarily generated by coherent infall towards over-dense regions (e.g., galaxy clusters). As a result, the redshifts of galaxies on the near side of a galaxy cluster are boosted relative to the Hubble

²Here, ‘physical’ does not mean the physical distances that change with redshift. We use this term to describe separations that have distance units rather than angular units.

flow, whereas those on the far side have reduced redshifts. This places galaxies at similar redshifts, squashing the **2PCF** along the LOS. In the most extreme case, the **2PCF** is completely flattened at the turn-around radius, where galaxies are just beginning their infall. Within galaxy clusters, galaxies exhibit large, randomized peculiar velocities. These galaxies, which are close on the sky, span a wide range of redshifts, elongating the **2PCF** along the LOS. Together, these effects introduce anisotropy in the **2PCF**, as shown in Fig. 1.3. While the Kaiser effect has a well-motivated analytic prescription (see Kaiser, 1987), **FoG** are more difficult to model.

Fortunately, galaxy clustering measurements are subject to fewer systematics than **GGL**. These are typically mitigated by assigning weights to individual galaxies, as described in Section 3.2. One systematic that requires careful treatment is fibre collisions. Our fibre collision correction methodology is outlined in Section 3.2. Here, we provide a brief overview of fibre collisions. Certain surveys use fibres to obtain spectra from individual galaxies. These fibres have a physical size, which depends on the particular survey. The corresponding angular size defines the ‘fibre collision’ scale. When two galaxies lie within the fibre collision scale, only one can be assigned a fibre in a single observation (Anderson et al., 2012). If the remaining galaxies are not allocated fibres in subsequent passes, the **2PCF** will be underestimated due to the missing pair counts. When many galaxies lie within the fibre collision scale, they form a ‘collided’ group.

1.10 The ‘Lensing is Low’ Problem

Several studies have simultaneously analysed **GGL** and clustering. Combining these two probes are especially useful for breaking degeneracies between **HOD** and cosmological parameters, enabling cosmological constraints (see Section 6.2). A notable tension between the two signals was discovered by Leauthaud et al. (2017) in the **Baryon Oscillation Spectroscopic Survey** (**BOSS**; Dawson et al., 2013) **CMASS Luminous Red Galaxy (LRG)** sample. They used standard galaxy–halo connection models (e.g., **HOD**, **SHAM**) calibrated on the clustering data to predict the **GGL** signal on small scales (up to $12 h^{-1} \text{Mpc}$), assuming either a **WMAP**- or a Planck-like cosmology. The resulting **GGL** predictions exceeded the measurements by 20–40%. This discrepancy is commonly referred to as the ‘lensing is low’ problem in the literature.

Since the original Leauthaud et al. (2017) analysis, there have been further investigations into the origin(s) of the lensing is low problem. Galaxy formation effects such as assembly bias, feedback, and satellite galaxy segregation have all been proposed as potential explanations (Leauthaud et al., 2017; Lange et al., 2019; Yuan et al., 2021; Amon et al.,

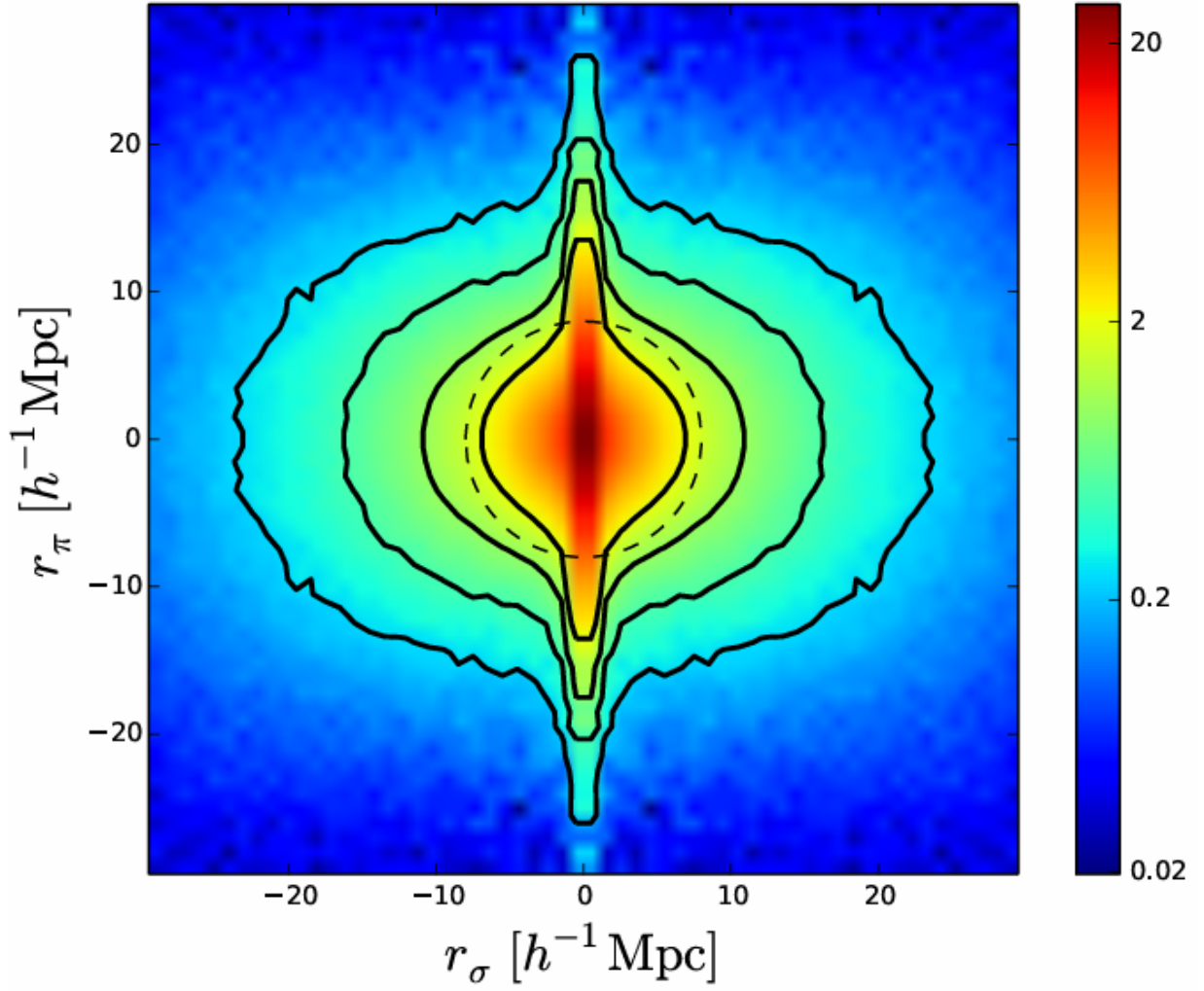


Figure 1.3: Two-dimensional correlation function, $\xi(r_p, r_\pi)$, of CMASS galaxies from [Reid et al. \(2014\)](#). Here, r_p is denoted by r_σ .

2023; Chaves-Montero et al., 2023). These effects are not captured by the standard HOD, as it bypasses galaxy formation physics by distributing galaxies in a statistical manner. Given the simplistic nature of HODs, it is easy to see how the lensing is low effect could have astrophysical origins. For assembly bias, Yuan et al. (2021) and Chaves-Montero et al. (2023) show that neglecting this effect leads to the standard HOD placing too few galaxies in low mass halos, causing an over-prediction of the GGL signal. Incorporating assembly bias through a decorated HOD shifts GGL predictions in the right direction, though it does not fully resolve the lensing-is-low tension (Lange et al., 2019; Yuan et al., 2021). Regarding feedback, if galaxies are populated into dark matter halos from dark-matter-only N-body simulations (which do not include feedback), they may not trace an accurate matter distribution since feedback redistributes matter. If feedback pushes matter towards the outskirts of a halo (which has been demonstrated in hydrodynamical simulations), the predicted GGL signal would be suppressed. However, feedback alone also cannot fully account for the lensing-is-low tension (Lange et al., 2019; Amon et al., 2023). Finally, satellite galaxy segregation refers to satellite galaxy concentration differing from halo concentration (Chaves-Montero et al., 2023). In standard HODs, satellite galaxy positions are tied to the underlying matter distribution, either by using the halo concentration to set the satellite galaxy concentration, or by placing satellite galaxies at the locations of dark matter particles. Chaves-Montero et al. (2023) showed that for simulated galaxies representative of the BOSS sample, satellite galaxy concentrations were significantly lower than halo concentrations. When satellite galaxies lie farther from the halo centre, they probe less dense regions, leading to a suppressed GGL signal. Unsurprisingly, this effect alone does not fully describe the lensing-is-low tension (Chaves-Montero et al., 2023). The same paper demonstrates that accounting for multiple galaxy formation effects simultaneously largely explains the lensing-is-low tension. These effects primarily modify GGL on small scales (with assembly bias extending to intermediate scales). However, predictions incorporating various effects converge around the 2-halo term (see Fig. 10 in Chaves-Montero et al., 2023).

Alternatively, lowering the amplitude of P_{mm} relative to that inferred from Planck has been shown to reconcile GGL and clustering (Leauthaud et al., 2017; Lange et al., 2019; Amon et al., 2023). This suppression is often quantified by the parameter $S_8 \equiv \sigma_8(\Omega_m/0.3)^{0.5}$, where σ_8 is the root-mean-square amplitude of matter density fluctuations on $8 h^{-1} \text{ Mpc}$ scales. Alleviating the tension between GGL and clustering via cosmological changes typically requires a reduction in S_8 by $2\text{--}3\sigma$ relative to the Planck 2018 constraint from primary CMB anisotropies (Planck Collaboration et al., 2020; $S_8 = 0.834 \pm 0.016$). This solution to the lensing is low problem remains debated, given that galaxy formation effects can explain much of the tension on small scales. Moreover, some recent analyses

that perform joint fits to both [GGL](#) and clustering yield S_8 values consistent with Planck 2018 (i.e., within 1σ), further complicating the situation. For example, [Zhang et al. \(2025\)](#) report $S_8 = 0.804 \pm 0.051$ using HSC Y3 data and a minimum bias model with a point-mass correction term for [GGL](#). They model the [GGL](#) signal from $2\text{--}70 h^{-1} \text{Mpc}$, and the clustering signal from $8\text{--}80 h^{-1} \text{Mpc}$. [More et al. \(2015\)](#) find $S_8 \sim 0.817 \pm 0.035$ from different CMASS subsamples using CFHTLenS data and an [HOD](#) approach. Their scales of interest span $0.1\text{--}20 h^{-1} \text{Mpc}$ for [GGL](#), and $0.85\text{--}80 h^{-1} \text{Mpc}$ for clustering. Historically, S_8 values inferred from low- z probes have tended to be lower than those constrained by the [CMB](#) ([Heymans et al., 2021](#); [Pandey et al., 2022](#)), sometimes resulting in a discrepancy known as the ‘ S_8 tension’³. Such tensions point to two possibilities: (1) there exists an unidentified systematic error in surveys, and/or (2) our physical models need revision.

To differentiate galaxy formation and cosmological effects, it is essential to precisely measure and model [GGL](#) and clustering signals across a wide range of scales, from the non-linear to linear regimes. This is made possible with weak-lensing data from the [Ultraviolet Near-Infrared Optical Northern Survey](#) ([UNIONS](#); [Gwyn et al., 2025](#)). [UNIONS](#) provides the largest on-sky overlap **to date** between CMASS and background source galaxies—approximately 2650 square degrees—enabling precise large-scale [GGL](#) measurements. Using these new measurements, we revisit the lensing is low problem and jointly model [GGL](#) and clustering with an [HOD](#) implemented by the DARK EMULATOR ([Nishimichi et al., 2019](#); [Miyatake et al., 2022a](#)). We explore various extensions to the standard [HOD](#), including feedback with an analytic prescription.

1.11 Outline

In this thesis, distances are expressed in comoving coordinates, and halo masses are reported in units of $h^{-1}M_\odot$. The structure of this thesis is as follows. In Chapter 2 we detail the galaxy catalogues used in this work. In Chapter 3 we describe our [GGL](#) and clustering measurements. In Chapter 4 we explain our theoretical [GGL](#) and clustering models. In Chapter 5 we describe our likelihood procedure for determining best-fit models. In Chapter 6 we present the results of this work, and interpret them in the context of the lensing is low problem. We finally summarize in Chapter 7, and provide additional methodological details and plots in the appendices. [GGL](#) measurements were performed assuming

³Some very recent low- z analyses are consistent with the Planck S_8 (e.g., [Bigwood et al., 2024](#); [Wright et al., 2025](#)).

a flat Λ CDM cosmology with $H_0 = 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and $\Omega_{\text{m},0} = 0.3111$ ⁴. Clustering measurements were performed with $H_0 = 67.66 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and $\Omega_{\text{m},0} = 0.3111$. For measurements, a reference cosmology must be selected to convert redshifts to distances. We emphasize that this fiducial cosmology is arbitrary, so long as predictions are rescaled accordingly (see Appendix B).

⁴We adopt $H_0 = 100 \text{ km s}^{-1} \text{ Mpc}^{-1} \implies h = 1$ since the package used to compute the GGL signal does not employ units of $h^{-1} \text{ Mpc}$ for distances.

Chapter 2

Data

This section describes the survey details of our lens and source galaxy samples. Both samples are used for [GGL](#) measurements, while the lens galaxy sample is used for clustering measurements. We use [UNIONS](#) ShapePipe v1.4.5 galaxies as our sources, and [BOSS](#) CMASS galaxies as our lenses.

2.1 UNIONS Background Source Galaxy Sample

We use the [UNIONS](#) ShapePipe ([Guinot et al., 2022](#); [Farrens et al., 2022](#)) v1.4.5 catalogue as our background source galaxy sample. This catalogue is internal to the [UNIONS](#) collaboration, and has not yet been publicly released. [UNIONS](#) is an ongoing wide-field multi-band imaging survey that uses three telescopes located in Hawai'i: the Canada-France-Hawai'i Telescope (CFHT), the Subaru Telescope, and the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS). [UNIONS](#) is projected to cover 4800 square degrees of the Northern sky, reaching magnitudes of 24.3, 25.2, 24.9, 24.3, and 24.1 in the u , r , g , i , and z bands, respectively ([Li et al., 2024](#)). The ShapePipe catalogue provides ellipticity components ϵ_1 and ϵ_2 (derived from CFIS r -band imaging), and includes over 63.2 million galaxies. These galaxies subtend an area of ~ 3200 square degrees. An earlier version of this catalogue was published by [Farrens et al. \(2022\)](#). The primary differences between that version and the one used here are as follows. (1) The current catalogue applies a Point Spread Function (PSF) size cut of $\text{HLR}_{\text{gal}}/\text{HLR}_{\text{PSF}} > 0.5$, where HLR denotes the half-light radius. (2) The galaxy weights follow the [Dark Energy Survey \(DES\)](#) weights in [Gatti et al. \(2021\)](#), which are given by $w_{s,i} = (1/\sigma_{\epsilon,i}^2)\langle R_i \rangle^2$. Here, i represents the index of

a particular (SNR , $\text{HLR}_{\text{gal}}/\text{HLR}_{\text{PSF}}$) bin, R is the scalar shear response from Section 1.8, and σ_e^2 is the variance in ellipticity.

At the time of this analysis, most **UNIONS** galaxies lack photometric redshifts. As a workaround, we instead compute an effective redshift distribution, $n(z_s)$. Here, we summarize the $n(z_s)$ construction. A more complete description is provided in Li et al. (2024). First, multi-band (*ugriz*) photometry is assigned to **UNIONS** *r*-band galaxies within the CFHTLenS W3 field. Then, spectroscopic calibration samples with CFHTLenS *ugriz* photometry are placed into a **Self-Organizing Map (SOM)**. After training the **SOM** on the calibration dataset, it is populated with the **UNIONS** galaxies that have *ugriz* photometry. Next, weights are computed for each **SOM** cell (w_{SOM}). These weights are given by the ratio of the (weighted) number of **UNIONS** objects to spectroscopic objects in a given cell. The effective redshift distribution is obtained by re-weighting the spectroscopic redshifts via:

$$n(z_s) = \int w_{\text{SOM}}(z_s) \langle R(z_s) \rangle n_{\text{spec}}(z_s) dz_s = \sum_i w_{\text{SOM}}^{(i)} \langle R \rangle^{(i)} n_{\text{spec}}^{(i)}(z_s), \quad (2.1)$$

where i denotes the i th **SOM** cell. The resulting redshift distribution is shown in Fig. 2.1.

2.2 BOSS Foreground Lens Galaxy Sample

For our foreground spectroscopic lens galaxy sample, we use the CMASS **LSS** catalogues from Data Release 12 (DR12) of the **BOSS** part of the Sloan Digital Sky Survey. These catalogues can be accessed via the Science Archive Server (SAS), in the **LSS** directories. Galaxies were imaged with the 2.5-meter Sloan Foundation Telescope at Apache Point Observatory in New Mexico. Spectroscopy was performed with the **BOSS** spectrographs mounted on the telescope. The CMASS sample consists of **LRGs** with roughly constant stellar mass in the range $\log_{10}(M_*/M_\odot) = 11.2\text{--}11.4$ (Sonnenfeld et al., 2019). Hence the term ‘CMASS’, which stands for constant mass. More details on the CMASS sample can be found in Reid et al. (2016).

We combine CMASS catalogues from the **North Galactic Cap (NGC)** and the **South Galactic Cap (SGC)** into a single catalogue. Like other works, we apply a redshift cut of $0.43 < z < 0.7$ to the entire catalogue, without further subdivision into tomographic bins. Our final catalogue contains 777,202 galaxies, 273,662 of which enter into our **GGL** calculation. The 777,202 galaxies span a total area of $\sim 10,380$ square degrees¹. Furthermore,

¹The area provided here is contiguous.

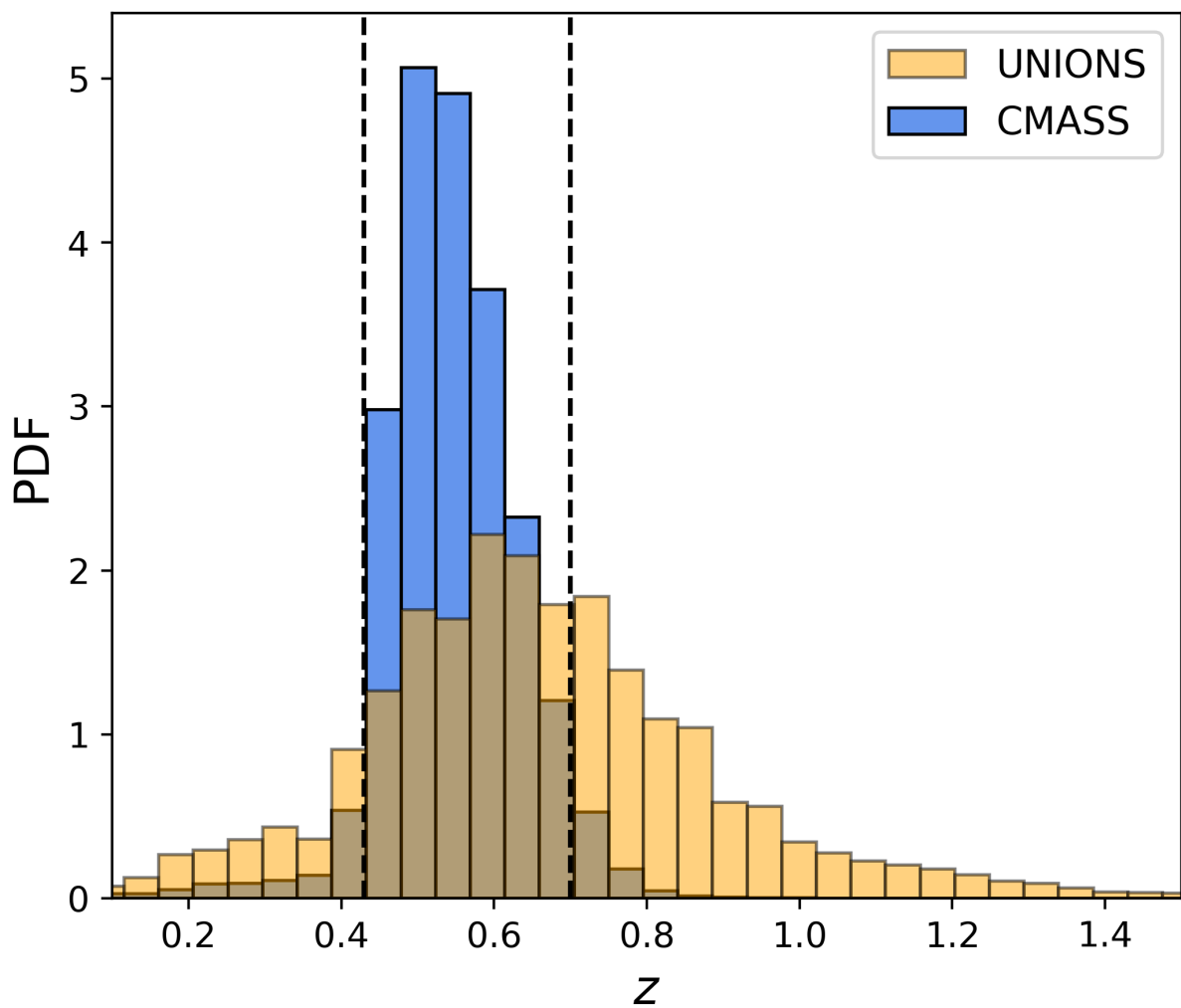


Figure 2.1: Effective redshift distribution of [UNIONS](#) ShapePipe v1.4.5 galaxies (yellow) compared to the CMASS redshift distribution (blue) without any redshift cuts. Vertical black lines indicate the CMASS redshift cuts of $z = 0.43$ and $z = 0.7$. ‘PDF’ on the y-axis stands for ‘probability density function’.

we use the [NGC](#) and [SGC](#) random catalogues provided by [SDSS](#) and merge them into one catalogue. We apply the same redshift cut as before to our random catalogue and down-sample it to contain 30 times more randoms than galaxies, enabling a faster computation of correlation functions. The angular overlap between our lens and source galaxy samples is shown in [Fig. 2.2](#).

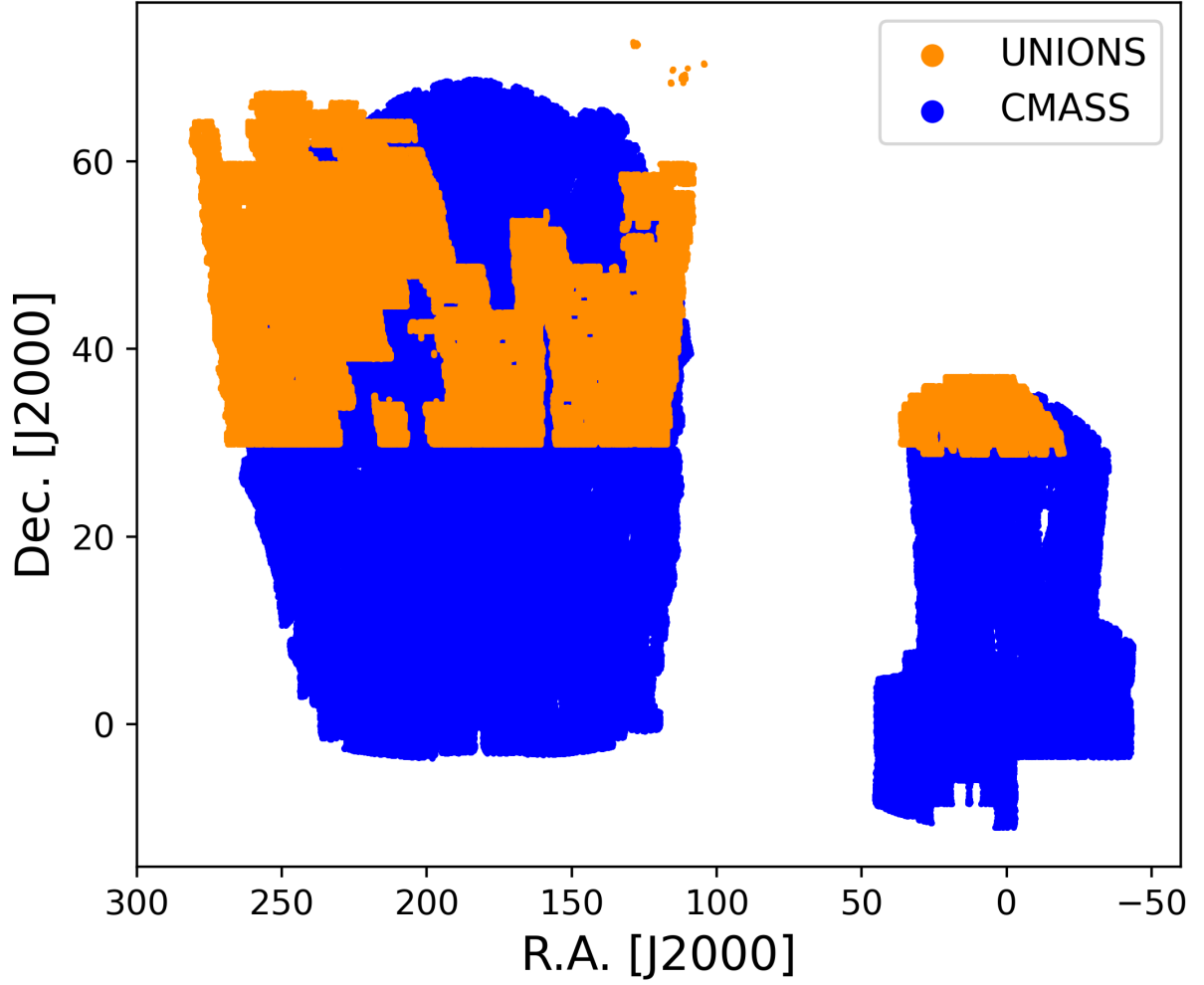


Figure 2.2: On-sky overlap between [UNIONS](#) ShapePipe v1.4.5 galaxies (orange) and CMASS galaxies in the redshift range $0.43 < z < 0.7$ (blue). We estimate the area of the overlap region to be ~ 2650 square degrees.

Chapter 3

Measurements

This section details our [GGL](#) and clustering measurements. Although the CMASS clustering signal has been extensively measured, we perform our own measurements to estimate its covariance matrix via jackknife resampling.

3.1 Excess Surface Density, $\Delta\Sigma$

We measure the [GGL](#) signal around CMASS galaxies using the Python package `DSIGMA` ([Lange & Huang, 2022](#)). As previously mentioned, the [GGL](#) signal is given by:

$$\gamma_t = \Sigma_{\text{crit}}^{-1}(z, z_s) \Delta\Sigma(r_p), \quad (3.1)$$

Recall that Σ_{crit} encodes the geometry of a lens–source galaxy pair, and is defined as:

$$\Sigma_{\text{crit}}(z, z_s) = \frac{c^2}{4\pi G} \frac{D_A(z_s)}{D_A(z) D_A(z, z_s)}, \quad (3.2)$$

This factor ensures consistent [GGL](#) measurements regardless of the chosen background source galaxies, whose distances affect the observed shear. Since our source galaxies lack photometric redshifts, we compute an effective $\Sigma_{\text{crit}}^{-1}$ for each lens galaxy by averaging over the $n(z_s)$:

$$\Sigma_{\text{crit, eff}}^{-1}(z) = \int \Sigma_{\text{crit}}^{-1}(z, z_s) n(z_s) dz_s. \quad (3.3)$$

The procedure above has been used in many analyses where photometric redshifts were either unreliable or unavailable (Sheldon et al., 2004; Mandelbaum et al., 2005). Thus, our estimator of the mean ESD around many lens galaxies is:

$$\Delta\Sigma(r_p) = \frac{\sum_{l,s} w_{\text{sys},l} w_s \Sigma_{\text{crit,eff}}^{-1}(z) \epsilon_t}{\sum_{l,s} w_{\text{sys},l} w_s \Sigma_{\text{crit,eff}}^{-2}(z)}, \quad (3.4)$$

where the sum is over all lens–source galaxy pairs in a given r_p bin, $w_{\text{sys},l}$ are lens galaxy systematic weights (see Section 3.2), and w_s are the source (DES) weights. To produce an optimal estimator, the same lens–source weights appear in both the numerator and denominator of Eq. 3.4. The expression is then divided by $\Sigma_{\text{crit}}^{-1}$ to convert ϵ_t into $\Delta\Sigma$.

We apply two corrections to the estimator in Eq. 3.4: random subtraction and the boost factor, b . In theory, the lensing signal around randoms should be consistent with zero. However, residual systematics can result in a non-zero signal. Subtracting the lensing signal around randoms from that around ‘true’ lenses can mitigate such systematics. Meanwhile, the boost factor accounts for contamination from physically associated lens–source pairs, which dilutes the lensing signal¹. This correction is particularly important in our case due to the non-negligible overlap in redshift between lens and source samples, as shown in Fig. 2.1. Under these corrections, the ESD becomes:

$$\Delta\Sigma(r_p) = b(r_p) \Delta\Sigma_{\text{true}}(r_p) - \Delta\Sigma_{\text{rand}}(r_p), \quad (3.5)$$

where b is given by:

$$b(r_p) = \frac{\sum_{l,s} w_{\text{sys},l} w_{l-s}}{\sum_{r,s} w_{\text{sys},r} w_{r-s}}. \quad (3.6)$$

In the expression above, r represents randoms, and $w_{l/r-s}$ are lens/random–source pair weights, computed as:

$$w_{l/r-s} = w_s \Sigma_{\text{crit,eff}}^{-2}(z). \quad (3.7)$$

The boost factor quantifies the degree of physical associations by computing the excess of sources near lenses relative to randoms. In the absence of physical associations, we do

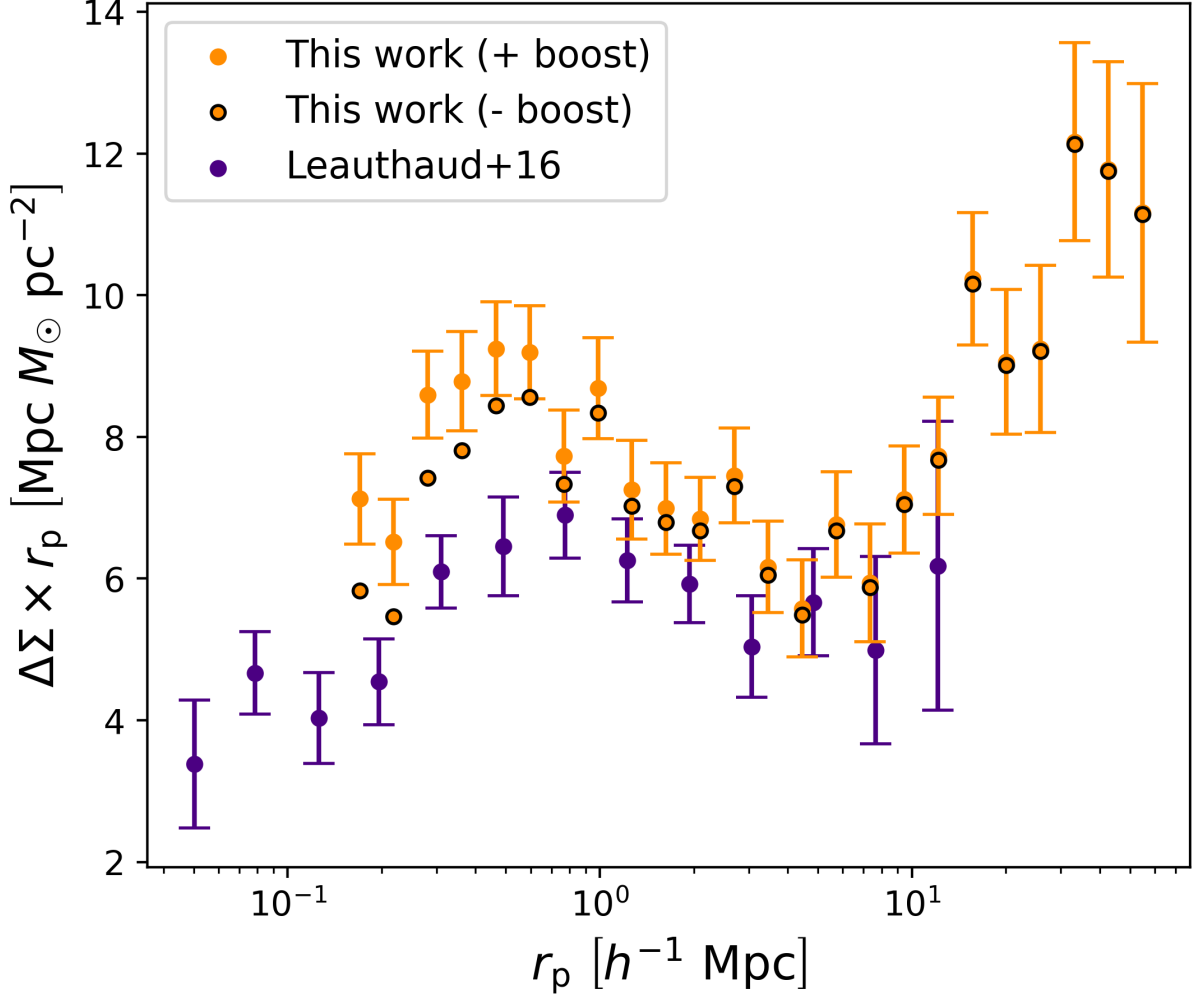


Figure 3.1: [GGL](#) signal/[ESD](#) around CMASS galaxies. The horizontal axis corresponds to the midpoints of r_p bins. Orange points show our measurements using [UNIONS](#) ShapePipe v1.4.5 source galaxies. Purple points show results from [Leauthaud et al. \(2017\)](#) using Canada-France-Hawai’i Telescope Stripe 82 Survey (CS82; [Moraes et al., 2014](#)) and Canada-France-Hawai’i Telescope Lensing Survey (CFHTLenS; [Heymans et al., 2012](#); [Miller et al., 2013](#)) source galaxies. Orange points with black borders indicate measurements without the boost factor correction. Only statistical 1σ errors are shown in this plot. Note that because of the difference in overlap with CMASS galaxies, we do not expect our measurements to be the exact same as [Leauthaud et al. \(2017\)](#).

not expect correlations between lens and source positions. Our measurement (with these corrections) of the [GGL](#) signal around CMASS galaxies is shown in Fig. 3.1.

A non-negligible systematic in this work is the uncertainty in $n(z_s)$. Internal tests show that the [SOM](#) method described in Section 2.1 overestimates the mean of $n(z_s)$ by 0.029 ± 0.016 . Following the procedure of [Bridle et al. \(2002\)](#), we analytically marginalise over the $n(z_s)$ uncertainty, thereby introducing an additional contribution to our covariance matrix. Specifically, this procedure adds positive correlations, as biases in the mean of $n(z_s)$ systematically shift the [GGL](#) signal up or down. This approach avoids adding a nuisance parameter to our theoretical model for $\Delta\Sigma$, keeping the parameter space minimal. We briefly describe the methodology below.

First, we compute $\Delta\Sigma$ using the original $n(z_s)$ shifted by $\delta z = -0.029$. This defines our fiducial $\Delta\Sigma$ measurement data vector, $\mathbf{d}_{\text{fid}}$, which is used for parameter inference. Next, we recompute $\Delta\Sigma$ with $n(z_s)$ shifted by $\pm 1\sigma$ of the [SOM](#) bias, yielding data vectors $\mathbf{d}_+$ and $\mathbf{d}_-$, respectively. We then define:

$$\Delta = \frac{\mathbf{d}_- - \mathbf{d}_+}{2}, \quad (3.8)$$

which represents the 1σ shift in $\mathbf{d}_{\text{fid}}$ due to residual $n(z_s)$ uncertainties. Following Eq. (8) in [Bridle et al. \(2002\)](#), our final covariance matrix is:

$$C_{\text{final}} = C(\mathbf{d}_{\text{fid}}) + \Delta\Delta^T, \quad (3.9)$$

where $C(\mathbf{d}_{\text{fid}})$ is the statistical covariance matrix described below. Uncertainties from this systematic are at the 5% level.

We estimate the [GGL](#) signal's statistical covariance matrix using 150 jackknife resamplings. Jackknifing is built into DSIGMA, and the underlying code constructs patches such that they cover roughly the same on-sky area. The jackknife covariance matrix is given by:

$$C_{\text{JK}}(\hat{f}_i, \hat{f}_j) = \frac{N_{\text{JK}} - 1}{N_{\text{JK}}} \sum_{k=1}^{N_{\text{JK}}} \left(\hat{f}_i^{(k)} - \bar{\hat{f}}_i \right) \left(\hat{f}_j^{(k)} - \bar{\hat{f}}_j \right), \quad (3.10)$$

¹Although related, the boost factor correction does **not** mitigate [IAs](#). This correction simply accounts for the fact that sources with redshifts close to the lenses blow up Σ_{crit} .

where $\hat{f}$ is an estimator (here, our $\Delta\Sigma$ estimator), (i, j) represents an element of the covariance matrix, N_{JK} is the number of jackknife patches, $\hat{f}_{i/j}^{(k)}$ is the estimator's evaluation when omitting the k th jackknife region, and $\overline{\hat{f}_{i/j}}$ is the mean evaluation across all realizations. We choose $N_{\text{JK}} = 150$ to obtain reliable covariance estimates at large scales, while preserving error estimates and correlation structure observed with larger N_{JK} . The $\Delta\Sigma$ jackknife covariance matrix can be seen in Fig. C.1. The correlation matrix is shown in Fig. 3.2.

Since the jackknife covariance matrix is noisy, its inverse is biased. To correct for this, we rescale the jackknife covariance matrix by the inverse Hartlap et al. (2007) factor, $\mathcal{H}^{-1}$:

$$C_{\text{JK, final}} = C_{\text{JK}} \mathcal{H}^{-1} = C_{\text{JK}} \times \frac{N_{\text{JK}} - 1}{N_{\text{JK}} - m - 2}, \quad (3.11)$$

where m is the rank of the covariance matrix (here, equivalent to the number of r_{p} points). We verified that applying the inverse Hartlap factor to a numerical covariance matrix, subsequently combined with additional covariances, yields an unbiased inverse covariance matrix. Ensuring an unbiased inverse covariance matrix is essential for parameter estimation (see Chapter 5).

We measure $\Delta\Sigma$ in 25 logarithmically spaced bins from $r_{\text{p}} = 0.15 h^{-1} \text{Mpc}$ to $r_{\text{p}} = 62 h^{-1} \text{Mpc}$. The lower bound of our r_{p} bins does not extend below $0.1 h^{-1} \text{Mpc}$, as DARK EMULATOR predictions start to become unreliable at this scale.

3.2 Projected Correlation Function, w_{p}

We measure the clustering signal of CMASS galaxies using the Python package PYCORR (Sinha & Garrison, 2019, 2020). The galaxy clustering signal, also known as the projected correlation function, is written as:

$$w_{\text{p}}(r_{\text{p}}) = \int_{-\pi_{\text{max}}}^{\pi_{\text{max}}} \xi(r_{\text{p}}, r_{\pi}) dr_{\pi}, \quad (3.12)$$

where r_{π} are separation bins parallel to the line-of-sight (LOS), and $\xi(r_{\text{p}}, r_{\pi})$ is the 2D correlation function, which measures the excess probability of finding a galaxy pair in (r_{p}, r_{π}) bins. We use the Landy-Szalay minimum-variance estimator (Landy & Szalay, 1993) to compute the 2D correlation function:

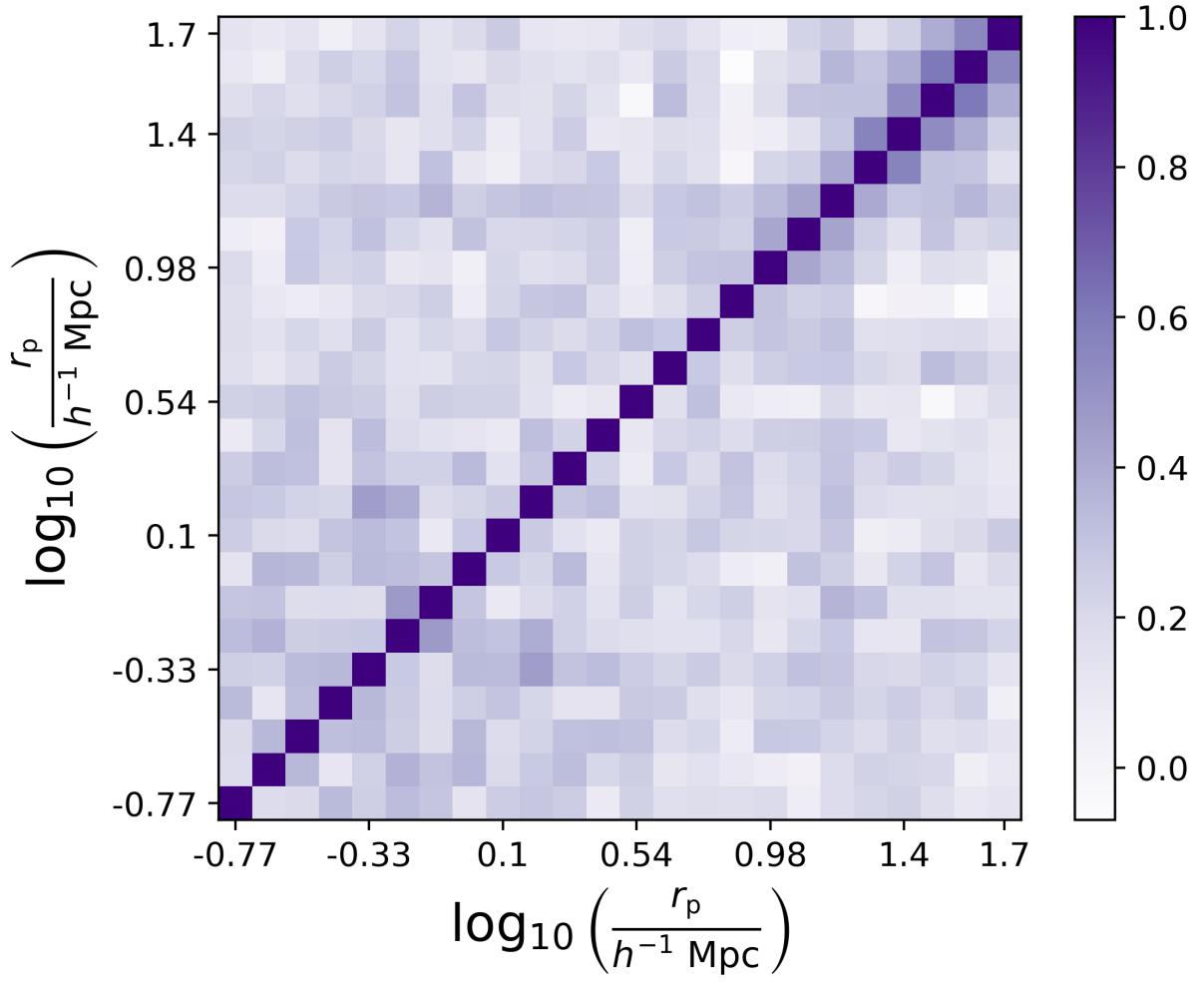


Figure 3.2: Correlation matrix computed from the $\Delta\Sigma$ covariance matrix. By construction, every matrix element must have a value between -1 and 1. Small-scale r_p bins are uncorrelated with other bins, as shape noise dominates on these scales. Large-scale bins are mildly correlated due to [LSS](#).

$$\xi(r_p, r_\pi) = \frac{DD(r_p, r_\pi) - 2DR(r_p, r_\pi) + RR(r_p, r_\pi)}{RR(r_p, r_\pi)}. \quad (3.13)$$

DD , DR , and RR denote data–data (galaxy–galaxy), data–random, and random–random (normalised) pair counts, respectively.

When performing the pair counts in Eq. 3.13, we apply the following BOSS weights to each galaxy:

$$w_{\text{tot}} = w_{\text{FKP}} w_{\text{sys}} (w_{\text{zf}} + w_{\text{cp}} - 1). \quad (3.14)$$

The [Feldman-Kaiser-Peacock](#) (FKP; [Feldman et al., 1994](#)) weights (w_{FKP}) account for spatial variations in galaxy number density and improve the clustering SNR, while systematic weights (w_{sys}) correct for seeing and stellar contamination. w_{zf} and w_{cp} are redshift failure and close-pair weights, respectively. The BOSS weights are described in more detail in [Reid et al. \(2016\)](#). The fibre collision scale for BOSS is 62", so two galaxies separated by less than this angle form a close pair. Approximately 5.8% of DR12 CMASS targets are impacted by fibre collisions ([Reid et al., 2016](#)). The redshift failure and close-pair weights were created with a ‘nearest neighbour’ up-weighting scheme, and their default values are unity. Randoms are only assigned FKP weights.

The nearest neighbour close-pair weights successfully recover the underlying clustering statistics on intermediate to large scales, but are insufficient near or below the fibre collision scale ([Bianchi & Percival, 2017](#)). So, we use ‘angular up-weights’ [Hawkins et al. \(2003\)](#) in place of the BOSS close-pair weights on small scales. The angular up-weights are given by:

$$w_{\text{pair}}(\theta) = \frac{1 + w_t(\theta)}{1 + w_s(\theta)}, \quad (3.15)$$

where $w_t(\theta)$ and $w_s(\theta)$ are the angular correlation functions of target and spectroscopic samples, respectively. The spectroscopic sample is the galaxy catalogue used in previous calculations, while the target sample contains the spectroscopic sample plus galaxies with missing redshifts due to fibre collisions. Below, we detail our methodology for constructing the target sample.

We began with the DR12 BOSS target list, and identified CMASS targets using a bit-field mask. Next, we generated ‘TARGETOBJIDs’ for each galaxy in the spectroscopic sample and in the CMASS target list, following the SDSS documentation. TARGETOBJIDs are unique identifiers for objects, combining the following fields: RUN, CAMCOL,

FIELD, ID, and RERUN. Conversion between these fields and TARGETOBJID is handled by the Python package SDSSPY ([sdsspy](#), [nd](#)). We then determined which CMASS targets were missing redshifts/were not allocated spectroscopic fibres due to being in a collision group via the ‘ICOLLIDED’ field. After, we iterated through all CMASS targets, and kept those whose IDs (a) could be found in the spectroscopic sample, and (b) were missing from the ‘specObj’ file and that had an ICOLLIDED flag. The specObj file contains spectroscopic information. By keeping targets absent from this file but marked with the ICOLLIDED flag, we recovered the missed galaxies due to fibre collisions rather than redshift failures in the spectroscopic pipeline. Finally, we applied the masks used on galaxies in the spectroscopic sample to our final CMASS targets. Applying these masks ensures consistency in the selection function between our two samples. These masks account for the survey geometry, regions around bright stars, tiling completeness, etc.

Essentially, the angular up-weighting method estimates the number of galaxy pairs lost to fibre collisions, and uses this information to up-weight affected DD pairs. A caveat of this method is that it assumes identical radial properties between missed galaxy pairs and (spectroscopically) observed galaxy pairs ([Bianchi & Percival, 2017](#)). This assumption is not entirely valid: missed close pairs are more likely to be found in target triplets rather than doublets, as triplets require more passes for all targets to be observed. Triplets are also more likely to be radially associated than doublets, meaning that the 3D separations of missed galaxy pairs differ from those of the observed galaxy pairs. However, the angular up-weighting method treats the fibre-collided missed galaxies as a random subsample of targets. Additionally, the angular up-weights approach unity on large scales, making it necessary to use another fibre collision correction scheme on these scales. When angular up-weights are used, the galaxy weights become:

$$\tilde{w}_{\text{tot}} = w_{\text{FKP}} w_{\text{sys}} w_{\text{zf}}. \quad (3.16)$$

The pair counts in Eq. 3.13 satisfy the following relationships:

$$DD(r_p, r_\pi) \propto \sum_{i \neq j} w_D^{(i)} w_D^{(j)}, \quad (3.17)$$

$$DR(r_p, r_\pi) \propto \sum_{i,j} w_D^{(i)} w_R^{(j)}, \quad (3.18)$$

$$RR(r_p, r_\pi) \propto \sum_{i \neq j} w_R^{(i)} w_R^{(j)}, \quad (3.19)$$

where w denotes the weight of a galaxy or a random in a pair. Note that the expressions above are unnormalised pair counts. With our weighting scheme, the summands in Eqs. 3.17, 3.18, and 3.19 are:

$$w_D^{(i)} w_D^{(j)} = \begin{cases} \tilde{w}_{\text{tot}}^{(i)} \tilde{w}_{\text{tot}}^{(j)} w_{\text{pair}} & (r_p \leq r_{\text{switch}}) \\ w_{\text{tot}}^{(i)} w_{\text{tot}}^{(j)} & (r_p > r_{\text{switch}}) \end{cases}, \quad (3.20)$$

$$w_D^{(i)} w_R^{(j)} = \begin{cases} \tilde{w}_{\text{tot}}^{(i)} w_{\text{FKP,rand}}^{(j)} & (r_p \leq r_{\text{switch}}) \\ w_{\text{tot}}^{(i)} w_{\text{FKP,rand}}^{(j)} & (r_p > r_{\text{switch}}) \end{cases}, \quad (3.21)$$

$$w_R^{(i)} w_R^{(j)} = w_{\text{FKP,rand}}^{(i)} w_{\text{FKP,rand}}^{(j)}. \quad (3.22)$$

The parameter r_{switch} is defined as the scale at which w_p measurements with and without angular up-weights intersect, and we find it to be $0.7 h^{-1} \text{ Mpc}$.

To estimate the clustering signal's statistical covariance matrix, we again use the jackknife resampling technique (which is implemented in PYCORR) with 150 patches. The resulting covariance matrix is rescaled by the inverse Hartlap factor, following the previous section. The w_p jackknife covariance matrix can be seen in Fig. C.1, with its correlation matrix shown in Fig. 3.3.

The final w_p covariance matrix used for parameter estimation includes additional sources of uncertainty. First, we associate a 10% uncertainty with our measurement of w_t , following Reid et al. (2014). To propagate this uncertainty, we recompute w_p with $\pm 10\%$ variations in w_t , and use the differences in w_p from the original measurement to derive error bars from the angular up-weighting method. Second, we assign a 3% uncertainty to the emulator predictions of w_p . We do not include an emulator uncertainty for $\Delta\Sigma$, as it is small compared to the statistical uncertainties. The squares of these additional errors are added to the diagonal elements of the rescaled jackknife covariance matrix. We measure w_p in 26 logarithmically spaced bins from $r_p = 0.15 h^{-1} \text{ Mpc}$ to $r_p = 80 h^{-1} \text{ Mpc}$.

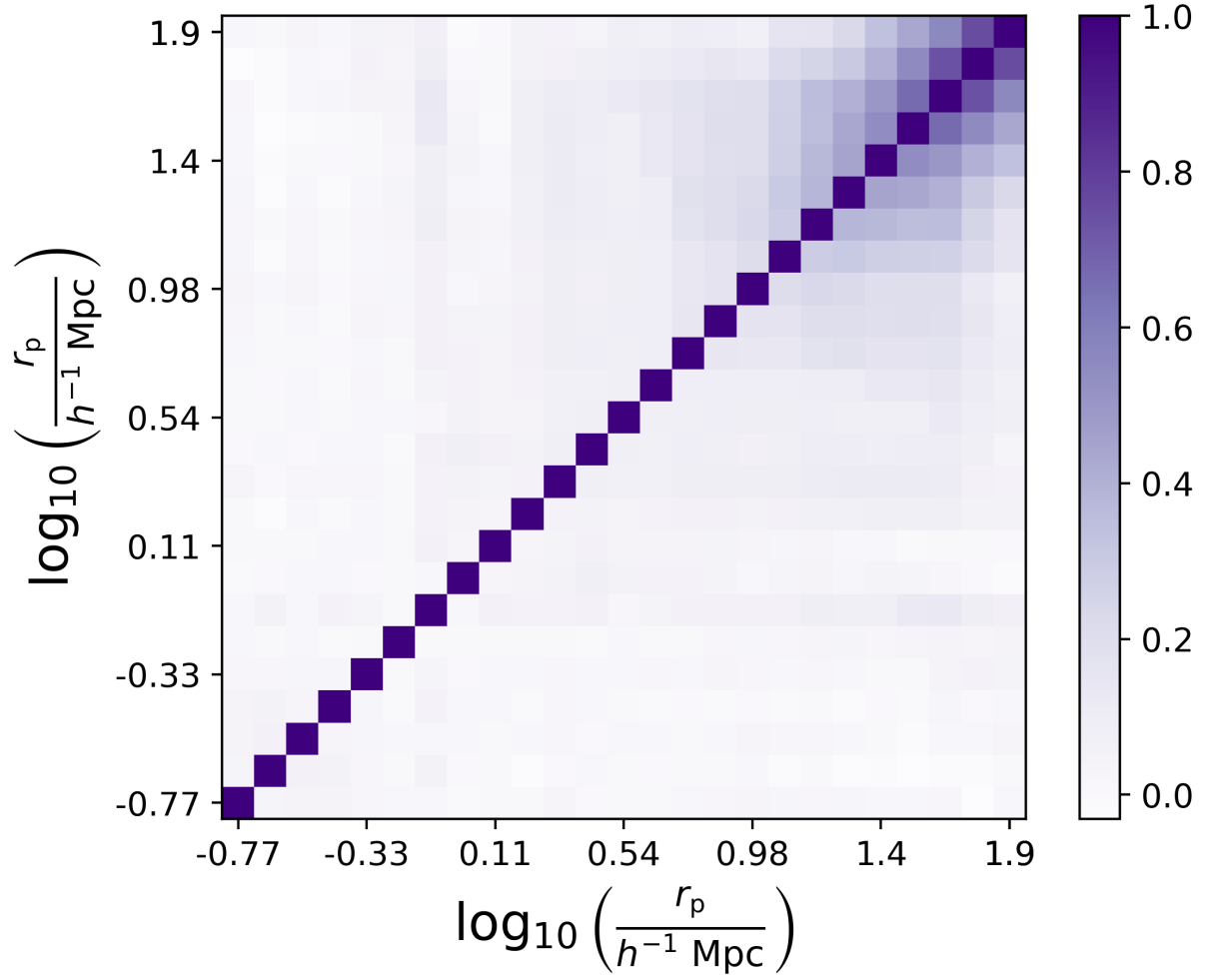


Figure 3.3: Correlation matrix computed from the w_p covariance matrix. Small-scale r_p bins are dominated by shot noise. Large-scale bins are strongly correlated due to [LSS](#).

Chapter 4

Theoretical Framework

This section outlines our theoretical expressions for the [GGL](#) and clustering signals. We also describe our extensions to the standard [HOD](#) framework, including central galaxy off-centring and feedback.

4.1 Theoretical Two-Point Correlation Functions

Using both the halo model and the [HOD](#) framework, we can derive theoretical expressions for our observables of interest: $\Delta\Sigma$ and w_p . First, we define the 3D galaxy–matter power spectrum, P_{gm} :

$$P_{\text{gm}}(k, z) = \frac{1}{\bar{n}_g} \int [\langle N_c(M) \rangle [1 - p_{\text{off}} + p_{\text{off}} \tilde{u}_{\text{off}}(k|M, z)] + \langle N_s(M) \rangle \tilde{u}_s(k|M, z)] \times P_{\text{hm}}(k, M, z) \frac{dn_h}{dM} dM, \quad (4.1)$$

where $\bar{n}_g$ is the mean number density of galaxies, p_{off} is the fraction of central galaxies that are offset from the halo centre, and $\tilde{u}_s$ and $\tilde{u}_{\text{off}}$ are the Fourier transforms of the normalised radial profiles of satellite galaxies and off-centred central galaxies, respectively. We select a Navarro–Frenk–White (NFW; [Navarro et al., 1996](#)) profile for the radial distribution of satellite galaxies, and allow the concentration to be modified by a multiplicative factor R_c ¹. From here on, we refer to R_c as the ‘satellite concentration normalisation’, with a

¹The default satellite galaxy concentration is given by the halo concentration.

fiducial value of 1. P_{hm} is the 3D halo–matter cross–power spectrum (including 1- and 2-halo terms), and dn_{h}/dM is the [HMF](#). The mean number density of galaxies is computed via:

$$\bar{n}_{\text{g}} = \int (\langle N_{\text{c}}(M) \rangle + \langle N_{\text{s}}(M) \rangle) \frac{dn_{\text{h}}}{dM} dM. \quad (4.2)$$

The Fourier transform of the normalised radial profile of off-centred central galaxies is given by:

$$\tilde{u}_{\text{off}}(k|M, z) = \exp\left(-\frac{1}{2}k^2(R_{200\text{m}}R_{\text{off}})^2\right), \quad (4.3)$$

where $R_{200\text{m}}$ follows the same definition as $R_{200\text{c}}$ in Section 4.3, but is defined with respect to the mean matter density, $\bar{\rho}_{\text{m}}$, rather than ρ_{crit} . The expression above is the Fourier transform of a Gaussian with width set by the dimensionless parameter R_{off} . Finally, we compute the [GGL](#) signal via:

$$\Delta\Sigma(r_{\text{p}}, z) = \bar{\rho}_{\text{m},0} \int_0^\infty \frac{k^2}{2\pi^2} P_{\text{gm}}(k, z) J_2(kr_{\text{p}}) dk. \quad (4.4)$$

To account for the impact of lens magnification on the [GGL](#) signal, we add a lens magnification term to the prediction above. We briefly describe the lens magnification model here. First, we compute the [LSS](#) tangential shear, $\gamma_{\text{t}}^{\text{LSS}}$, following Eq. 1.70 and using the weighted mean redshifts of lens and source galaxies. The matter power spectrum is obtained with the Python package CAMB ([Lewis et al., 2000](#); [Howlett et al., 2012](#)). The [LSS](#) tangential shear is then scaled by the weighted mean critical surface density, provided by `DSIGMA`. Finally, the [ESD](#) is multiplied by the prefactor in Eq. 1.69, where we set $\alpha_1^{\text{CMASS}} \approx 2.98$. We neglect the redshift dependence of α_1 , as it has been shown to vary slowly with redshift ([von Wietersheim-Kramsta et al., 2021](#)). The lens magnification term was computed across a grid of cosmological parameters, from which we constructed an interpolator for efficient evaluations during parameter estimation.

To compute w_{p} , we first calculate the 3D galaxy power spectrum, P_{gg} :

$$P_{\text{gg}}(k, z) = P_{\text{gg}}^{\text{1h}}(k, z) + P_{\text{gg}}^{\text{2h}}(k, z), \quad (4.5)$$

The 1- and 2-halo terms are computed as follows:

$$P_{\text{gg}}^{\text{1h}}(k, z) = \frac{1}{\bar{n}_{\text{g}}^2} \int [2\lambda_{\text{s}}(M)\tilde{u}_{\text{s}}(k|M, z)[1 - p_{\text{off}} + p_{\text{off}}\tilde{u}_{\text{off}}(k|M, z)] + \lambda_{\text{s}}^2(M)\tilde{u}_{\text{s}}^2(k|M, z)] \times \langle N_{\text{c}}(M) \rangle \frac{\text{d}n_{\text{h}}}{\text{d}M} \text{d}M, \quad (4.6)$$

$$P_{\text{gg}}^{\text{2h}}(k, z) = \frac{1}{\bar{n}_{\text{g}}^2} \int [1 - p_{\text{off}} + p_{\text{off}}\tilde{u}_{\text{off}}(k|M, z) + \lambda_{\text{s}}(M)\tilde{u}_{\text{s}}(k|M, z)] \langle N_{\text{c}}(M) \rangle \frac{\text{d}n_{\text{h}}}{\text{d}M} \text{d}M \times \int [1 - p_{\text{off}} + p_{\text{off}}\tilde{u}_{\text{off}}(k|M', z) + \lambda_{\text{s}}(M')\tilde{u}_{\text{s}}(k|M', z)] \langle N_{\text{c}}(M') \rangle \frac{\text{d}n_{\text{h}}}{\text{d}M'} \text{d}M' \times P_{\text{hh}}(k|M, M', z), \quad (4.7)$$

Now, we can convert the galaxy power spectrum into its configuration-space equivalent, ξ_{gg} :

$$\xi_{\text{gg}}(r, z) = \int_0^\infty \frac{k^2}{2\pi^2} P_{\text{gg}}(k, z) j_0(kr) \text{d}k, \quad (4.8)$$

where j_0 is the spherical Bessel function. Finally, we compute the galaxy clustering signal by integrating ξ_{gg} along the LOS:

$$w_{\text{p}}(r_{\text{p}}, z) = 2 \int_0^{\pi_{\text{max}}} \xi_{\text{gg}}\left(\sqrt{r_{\text{p}}^2 + r_{\pi}^2}, z\right) \text{d}r_{\pi}. \quad (4.9)$$

Note that all predictions are evaluated at the median CMASS redshift of 0.54. We also account for the cosmology dependence of measurements according to the equations in Appendix B.

4.2 The Dark Emulator

We implement our [HOD](#) through the DARK EMULATOR—a fast and flexible emulation tool built on the Dark Quest suite of dark matter-only N-body simulations. These simulations were specifically designed to capture halo clustering statistics. Each simulation evolves 2048^3 particles within a comoving box size of 1 or 2 h^{-1} Gpc. The emulator was trained on a set of 100 different flat w CDM cosmologies, and provides accurate predictions for P_{hm} ,

P_{hh} , and dn_{h}/dM . All other quantities, including $\Delta\Sigma$ and w_{p} , are computed internally with the analytic prescriptions previously described. DARK EMULATOR also supports extensions to the standard HOD, such as central galaxy off-centring and halo mass incompleteness. However, assembly bias is not currently implemented in the DARK EMULATOR. Users can specify the following cosmological parameters when calling the DARK EMULATOR: $\omega_{\text{b},0}$, $\omega_{\text{c},0}$, $\Omega_{\text{de},0}$, $\ln(10^{10}A_{\text{s}})$, n_{s} , and w . The parameters below are derived internally:

$$\Omega_{\text{m},0} = 1 - \Omega_{\text{de},0}, \quad (4.10)$$

$$h = \sqrt{\frac{\omega_{\text{b},0} + \omega_{\text{c},0} + \omega_{\nu,0}}{\Omega_{\text{m},0}}}, \quad (4.11)$$

where the present-day physical neutrino density, $\omega_{\nu,0}$, is by default fixed at 6.4×10^{-4} .

4.3 Feedback

To model baryonic feedback, we follow the semi-analytic approach from Schneider et al. (2019) and Shirasaki et al. (2024), commonly referred to as the BC model. This procedure expands or contracts CLM shells in response to the presence of baryons. To implement the method, we first take the DARK EMULATOR output $\xi_{\text{hm}}(r, M, z)$ and compute the average mass density profile, ρ_{hm} :

$$\rho_{\text{hm}}(r, M, z) = \bar{\rho}_{\text{m},0}[1 + \xi_{\text{hm}}(r, M, z)]. \quad (4.12)$$

Then, we truncate ρ_{hm} so that its enclosed mass converges as $r \rightarrow \infty$. DARK EMULATOR provides the total halo-matter cross-correlation, ξ_{hm} , without a decomposition into 1- and 2-halo terms. So, the truncation is somewhat arbitrary. We define the truncation radius, r_{out} , as the radius at which the logarithmic slope of ρ_{hm} is the most negative. Now, we can define a truncated average mass density profile:

$$\rho_{\text{hm}+\text{t}}(r, M, z) = \begin{cases} \rho_{\text{hm}}(r, M, z) & (r \leq r_{\text{out}}) \\ \rho_{\text{hm}}(r_{\text{out}}, M, z) \left(\frac{r}{r_{\text{out}}}\right)^{\gamma} \left(\frac{r_{\text{out}}^2 + \tau^2}{r^2 + \tau^2}\right) & (r > r_{\text{out}}) \end{cases}, \quad (4.13)$$

where γ is the dimensionless slope of ρ_{hm} at $r = r_{\text{out}}$, and we set $\tau = 3r_{\text{out}}$ as in Shirasaki et al. (2024).

Next, we define density profiles for dark matter, stars, and gas. The dark matter profile is simply Eq. 4.13 rescaled by the global mass fraction of dark matter:

$$\rho_{\text{DM}}(r, M, z) = \left(1 - \frac{\Omega_{\text{b}}}{\Omega_{\text{m}}}\right) \rho_{\text{hm+t}}(r, M, z). \quad (4.14)$$

We divide the total stellar density profile into two components: one from the central galaxy and one from satellite galaxies. The contribution from stars in the central galaxy is:

$$\rho_{\text{cga}}(r) = \frac{f_{\text{cga}} M_{\text{tot}}}{4\pi^{3/2} R_* r^2} \exp\left[-\left(\frac{r}{2R_*}\right)^2\right], \quad (4.15)$$

where f_{cga} is the fraction of stars in the central galaxy, M_{tot} is the enclosed mass of $\rho_{\text{hm+t}}$ at $r = \infty$, and R_* is the stellar half-light radius. We compute f_{cga} via:

$$f_{\text{cga}}(M_{200\text{c}}) = 0.09 \left(\frac{M_{200\text{c}}}{M_{\text{s}}}\right)^{-\eta_{\text{cga}}}, \quad (4.16)$$

where $M_{200\text{c}}$ is the mass enclosed within a radius $R_{200\text{c}}$, inside which the mean density is $200\rho_{\text{crit}}$. Parameters $M_{200\text{c}}$ and $R_{200\text{c}}$ are related as follows:

$$M_{200\text{c}} = \frac{4\pi}{3} 200 \rho_{\text{crit}}(z) R_{200\text{c}}^3 (1 + z)^{-3} = \int_0^{R_{200\text{c}}} 4\pi r^2 \rho_{\text{hm}}(r, M, z) dr. \quad (4.17)$$

Note that $R_{200\text{c}}$ is assumed to be in comoving coordinates in the formula above. We set $R_* = 0.015 R_{200\text{c}}$, $M_{\text{s}} = 2.5 \times 10^{11} h^{-1} M_{\odot}$, and $\eta_{\text{cga}} = 0.6$ as in Shirasaki et al. (2024).

One can also compute f_{sga} , the fraction of stars in satellite galaxies, with the following set of equations:

$$f_{\text{star}}(M_{200\text{c}}) = 0.09 \left(\frac{M_{200\text{c}}}{M_{\text{s}}}\right)^{\eta_{\text{star}}}, \quad (4.18)$$

$$f_{\text{sga}}(M_{200\text{c}}) = f_{\text{star}}(M_{200\text{c}}) - f_{\text{cga}}(M_{200\text{c}}), \quad (4.19)$$

where η_{star} is set to 0.32 as in [Shirasaki et al. \(2024\)](#). With f_{sga} , we can define the density profile of stars in satellite galaxies. We assume that satellite galaxies are collisionless and model their stellar distribution after the dark matter density profile:

$$\rho_{\text{sga}}(r, M, z) = f_{\text{sga}}\rho_{\text{hm+t}}(r, M, z). \quad (4.20)$$

The total stellar density profile is given by:

$$\rho_{\text{star}}(r, M, z) = \rho_{\text{cga}}(r) + \rho_{\text{sga}}(r, M, z). \quad (4.21)$$

The gas density profile assumes the following form:

$$\rho_{\text{gas}}(r) \propto \left(1 + \frac{r}{0.1R_{200\text{c}}}\right)^{-\beta_{\text{g}}} \left(1 + \left(\frac{r}{\theta_{\text{ej}}R_{200\text{c}}}\right)^{\gamma_{\text{g}}}\right)^{\frac{-(\delta_{\text{g}} - \beta_{\text{g}})}{\gamma_{\text{g}}}}. \quad (4.22)$$

Following [Shirasaki et al. \(2024\)](#), we set $\gamma_{\text{g}} = 2$ and $\delta_{\text{g}} = 7$. The parameter θ_{ej} is the maximum radius of gas ejection defined in terms of $R_{200\text{c}}$, and β_{g} is the slope of the gas density profile, parametrized as follows:

$$\beta_{\text{g}}(M_{200\text{c}}) = \frac{3(M_{200\text{c}}/M_{\text{c}})^{\mu}}{1 + (M_{200\text{c}}/M_{\text{c}})^{\mu}}, \quad (4.23)$$

where M_{c} is the halo mass at which the slope of the gas density profile becomes shallower than -3, and μ controls the slope's steepness. We set $\mu = 1.2$, following the parameter constraints in Fig. 16 from [Shirasaki et al. \(2024\)](#). The normalisation of ρ_{gas} is determined by:

$$\int_0^{\infty} 4\pi r^2 \rho_{\text{gas}}(r) dr = \left(\frac{\Omega_{\text{b}}}{\Omega_{\text{m}}} - f_{\text{star}}\right) M_{\text{tot}}. \quad (4.24)$$

We show how the feedback parameters $\log M_{\text{c}}$ and θ_{ej} shape the gas density profile in Fig. 4.1 through two configurations: a low feedback case with low values for both parameters, and a high feedback case with high values. As previously mentioned, $\log M_{\text{c}}$ governs how shallow the slope of the gas density profile becomes at a characteristic halo mass. Larger values correspond to steeper profiles across a wide range of halo masses, whereas smaller values imply weaker feedback, with only low-mass halos significantly affected. Higher values of θ_{ej} drive gas away from the halo centre, producing a more extended profile. Due

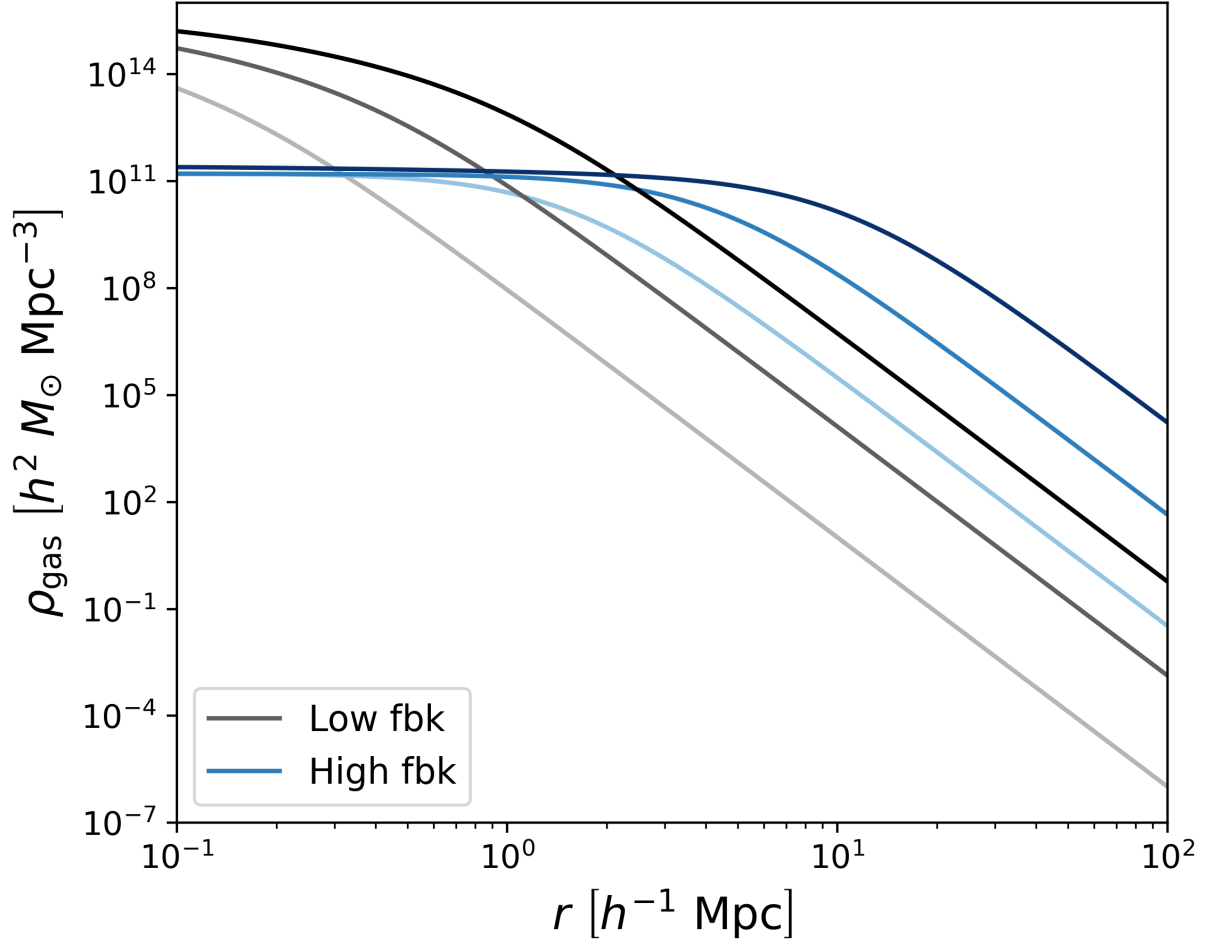


Figure 4.1: Gas density profiles for ‘low’ (grey curves) and ‘high’ (blue curves) feedback. For the low feedback case, $\log M_c = 11.0$ and $\theta_{\text{ej}} = 0.5$. For the high feedback case, $\log M_c = 16.0$ and $\theta_{\text{ej}} = 5.0$. Larger halo masses are represented by darker curves. The following halo masses were used to generate these curves: $\log(M_h/h^{-1} M_{\odot}) = 12.0, 14.0, 16.0$.

to mass conservation, the density profiles of low feedback models are greater (lower) than high feedback models on small (large) scales.

Now that we have density profiles for the different halo components, we can displace CLM spherical mass shells relative to the halo centre in response to the mass contributions from baryons. The initial and final radial positions of a CLM mass shell, r_i and r_f , are related as follows:

$$\frac{r_f}{r_i} = \mathcal{A} \left[\left(\frac{M_i}{M_f} \right)^n - 1 \right] + 1, \quad (4.25)$$

$$M_i = M_{\text{hm+t}}(r_i), \quad (4.26)$$

$$M_f = M_{\text{DM}}(r_i) + M_{\text{sga}}(r_i) + M_{\text{gas}}(r_f) + M_{\text{cga}}(r_f). \quad (4.27)$$

Note that the following parametrization of M_f yields identical results:

$$M_f = M_{\text{DM}}(r_i) + M_{\text{gas}}(r_f) + M_{\text{star}}(r_f). \quad (4.28)$$

The nuisance parameters $\mathcal{A}$ and n govern the expansion/contraction of CLM mass shells. Setting both to unity corresponds to the adiabatic contraction case introduced by Blumenthal et al. (1986). Following Shirasaki et al. (2024), we set $\mathcal{A} = 0.3$ and $n = 2$. The enclosed mass, M_x , of a given density profile (e.g., $x = \text{hm+t}$) is computed via:

$$M_x(r) = \int_0^r 4\pi \tilde{r}^2 \rho_x(\tilde{r}) d\tilde{r}. \quad (4.29)$$

One can solve for either r_i or r_f in Eq. 4.25; we went with the former. In either case, one should identify r_f , r_i pairs whose ratios satisfy the right-hand side of Eq. 4.25. Once r_i or r_f is solved for, one should evaluate the CLM enclosed mass at r_i , but associate it with r_f :

$$M_{\text{CLM}}(r_f) = M_{\text{DM}}(r_i) + M_{\text{sga}}(r_i). \quad (4.30)$$

If Eq. 4.28 was used to define M_f , then $M_{\text{CLM}}(r_f)$ is given by:

$$M_{\text{CLM}}(r_f) = M_{\text{DM}}(r_i). \quad (4.31)$$

We then compute ρ_{CLM} by taking the radial derivative of M_{CLM} :

$$\rho_{\text{CLM}}(r) = \frac{1}{4\pi r^2} \frac{d}{dr} [M_{\text{CLM}}(r)]. \quad (4.32)$$

Finally, we construct the halo-matter cross-correlation with feedback, $\xi_{\text{hm}+\text{b}}$:

$$\xi_{\text{hm}+\text{b}}(r, M, z) = \xi_{\text{hm}+\text{b}}^{\text{1h}}(r, M, z) + \xi_{\text{hm}+\text{b}}^{\text{2h}}(r, M, z), \quad (4.33)$$

$$\xi_{\text{hm}+\text{b}}^{\text{1h}}(r, M, z) = \frac{\rho_{\text{hm,dmb}}(r, M, z)}{\bar{\rho}_{\text{m},0}}, \quad (4.34)$$

$$\xi_{\text{hm}+\text{b}}^{\text{2h}}(r, M, z) = \frac{\rho_{\text{hm}}(r, M, z) - \rho_{\text{hm}+\text{t}}(r, M, z)}{\bar{\rho}_{\text{m},0}} - 1, \quad (4.35)$$

where $\rho_{\text{hm,dmb}} = \rho_{\text{CLM}} + \rho_{\text{gas}} + \rho_{\text{cga}}$, and ‘dmb’ stands for dark-matter-baryon. If ρ_{CLM} consists of only dark matter, then $\rho_{\text{hm,dmb}} = \rho_{\text{CLM}} + \rho_{\text{gas}} + \rho_{\text{star}}$. We compute the Fourier transform of $\xi_{\text{hm}+\text{b}}$ using the publicly available FFTLog code (Hamilton, 2000), and calculate the GGL signal as described in Section 4.1. All of the computations described here are performed outside of the DARK EMULATOR.

Note that the modifications discussed in this section only affect the **1-halo terms** of P_{hm} and P_{gm} , and not P_{gg} . The only thing that changes is the matter distribution, not the positions of halos and galaxies, nor the HMF. Thus, only the GGL signal is affected, while the clustering signal remains unchanged. Since GGL directly probes the underlying matter distribution, we expect it to be more sensitive to feedback. However, because feedback could also affect galaxy clustering, we vary the satellite concentration normalisation, R_{c} , in some models. Additionally, the baryonification procedure can modify the $M_{200\text{m}}$ of a given halo. When implementing the equations in Section 4.1, the halo mass is the pre-baryonification $M_{200\text{m}}$. Lastly, our feedback model is not nested within the DARK EMULATOR code; there is no meaningful parameter configuration that recovers the baseline DARK EMULATOR predictions.

Chapter 5

Parameter Estimation

We estimate the best-fit parameters using Bayesian inference. To sample the posterior distribution, $\mathcal{P}(\boldsymbol{\theta}|\mathbf{d})$, we perform a [Markov Chain Monte Carlo \(MCMC\)](#) analysis using the Python package EMCEE ([Foreman-Mackey et al., 2013](#)). According to Bayes' Theorem, $\mathcal{P}(\boldsymbol{\theta}|\mathbf{d})$ is proportional to:

$$\mathcal{P}(\boldsymbol{\theta}|\mathbf{d}) \propto \mathcal{L}(\mathbf{d}|\boldsymbol{\theta})\Pi(\boldsymbol{\theta}), \quad (5.1)$$

where $\boldsymbol{\theta}$ is a set of free parameters, $\mathcal{L}(\mathbf{d}|\boldsymbol{\theta})$ is the likelihood of the data given a specific parameter configuration, and $\Pi(\boldsymbol{\theta})$ is the prior distribution. We assume that the likelihood follows a multivariate normal distribution:

$$\ln \mathcal{L}(\mathbf{d}|\boldsymbol{\theta}) = -\frac{1}{2}\chi^2, \quad (5.2)$$

$$\chi^2 = \sum_{i,j} [d_i - t_i(\boldsymbol{\theta})](\mathbf{C}^{-1})_{ij} [d_j - t_j(\boldsymbol{\theta})], \quad (5.3)$$

where d represents an element of the data vector $\mathbf{d}$, t is a model prediction at $\boldsymbol{\theta}$, and $\mathbf{C}^{-1}$ is the inverse covariance matrix.

During parameter estimation, we perform joint fits to both $\Delta\Sigma$ and w_p . We neglect any cross-covariance between the two and compute independent likelihoods for each (see [Appendix C](#) for a justification), such that the final log likelihood is:

$$\ln \mathcal{L}(\mathbf{d}|\boldsymbol{\theta}) = -\frac{1}{2}(\chi_{\Delta\Sigma}^2 + \chi_{w_p}^2). \quad (5.4)$$

For all input model parameters, we impose flat/uniform priors. We also apply a ‘soft’ Gaussian prior to $\bar{n}_g$ with mean from Fig. 3 in [Leauthaud et al. \(2017\)](#). Our models have up to 12 free parameters, with each including the 5 [HOD](#) parameters described in Section 1.4. The degrees of freedom (dof) are calculated as: 24+25+1 - the number of free parameters. The numbers 24 and 25 come from the lengths of our $\Delta\Sigma$ and w_p data vectors, while the ‘+1’ arises from the additional $\bar{n}_g$ constraint, which is treated as an extra data point. Some models contain $\Omega_{m,0}$ and $\ln(10^{10}A_s)$ as free parameters since they directly influence P_{mm} , to which our observables are particularly sensitive. Since DARK EMULATOR takes $\Omega_{de,0}$ as an input, we vary $\Omega_{m,0}$ using the relation in Eq. 4.10. When $\Omega_{m,0}$ changes, either the baryon or [CDM](#) fraction must change—we allow the latter to vary. All other cosmological parameters are fixed at the Planck 2018 TT,TE,EE+lowE+lensing+bao values ([Planck Collaboration et al., 2020](#)). Other free parameters in our models are listed in Table 5.1. We also show how each parameter impacts our observables in Appendix D. To ensure the convergence of [MCMC](#) chains, we either require that (a) the length of a chain is at least 50 times the largest autocorrelation time estimated by EMCEE, or (b) the statistics from halves of a chain post burn-in are consistent. Lastly, we define the best-fit model as the parameter set from the [Maximum a Posteriori \(MAP\)](#).

Parameter	Type	Prior	Fiducial Value
$\log M_{\min}$	HOD	$\mathcal{U}(10.0, 15.0)$	–
$\sigma_{\log M}^2$	HOD	$\mathcal{U}(10^{-3}, 6.0)$	–
$\log M_1$	HOD	$\mathcal{U}(12.0, 17.5)$	–
α	HOD	$\mathcal{U}(10^{-3}, 3.0)$	–
κ	HOD	$\mathcal{U}(10^{-3}, 3.0)$	–
R_c	Extended HOD	$\mathcal{U}(0.1, 3.0)$	1.0
p_{off}	Extended HOD	$\mathcal{U}(0.0, 1.0)$	0.0
R_{off}	Extended HOD	$\mathcal{U}(10^{-2}, 5.0)$	–
$\log M_c$	Extended HOD	$\mathcal{U}(10.0, 17.5)$	–
θ_{ej}	Extended HOD	$\mathcal{U}(0.1, 17.0)$	–
$\ln(10^{10} A_s)$	Cosmological	$\mathcal{U}(2.4752, 3.7128)$	3.047
$\Omega_{\text{m},0}$	Cosmological	$\mathcal{U}(0.286, 0.338)$	0.3111
h	Cosmological	–	0.6766
$\Omega_{\text{b},0}$	Cosmological	–	4.897×10^{-2}
$\Omega_{\nu,0}$	Cosmological	–	1.398×10^{-3}
n_s	Cosmological	–	0.9665
w	Cosmological	–	-1
$\bar{n}_{\text{g}} [h^3 \text{Mpc}^{-3}]$	Derived	$\mathcal{G}(3.5, 0.75) \times 10^{-4}$	–

Table 5.1: Priors and fiducial values used for parameter estimation. ‘Extended HOD’ refers to parameters beyond the standard [HOD](#) framework that do not enter the expressions for $\langle N_c(M) \rangle$ and $\langle N_s(M) \rangle$. The symbol $\mathcal{U}(\text{low}, \text{high})$ denotes a uniform prior with lower and upper bounds given in parentheses, and $\mathcal{G}(\mu, \sigma)$ denotes a Gaussian prior with mean μ and width σ . Priors on cosmological parameters are set by DARK EMULATOR’s supporting range. Feedback is only incorporated when its corresponding parameters are allowed to vary.

Chapter 6

Results and Discussion

6.1 HOD With Planck Cosmology

We now present the main results of this work, summarized in Table 6.1 which reports χ^2 values, p -values, etc. Although our primary goal is to test various models through joint fits to both $\Delta\Sigma$ and w_p , we also perform a clustering-only fit¹ (Fig. 6.1), where the 5 HOD parameters from Section 1.4 are varied and cosmological parameters are fixed at the Planck 2018 TT,TE,EE+lowE+lensing+bao values. Following Leauthaud et al. (2017), we use the best-fit parameters to predict the GGL signal. Consistent with previous studies, we observe a lensing-is-low effect in the CMASS sample; many data points, including their 1σ error bars, fall below the predicted curve. On average, the predicted GGL signal is about 13% larger than our measurements. However, a 13% discrepancy is at the lower end of the values typically reported in the literature. The fit to w_p is quite good, although joint fits with other models achieve better agreement. Notably, the model struggles to fit the last two w_p data points—a trend also observed by More et al. (2015). This may be explained by the covariance matrix’s off-diagonality at large scales.

Before we compare different models, we introduce the Akaike Information Criterion (AICc) and the Bayesian Information Criterion (BIC). These two criteria facilitate model comparison/selection, balancing model simplicity and goodness of fit. For both, the preferred model amongst a set of candidates is the one with the lowest score. The AICc is defined as:

¹This fit still incorporates the galaxy abundance.

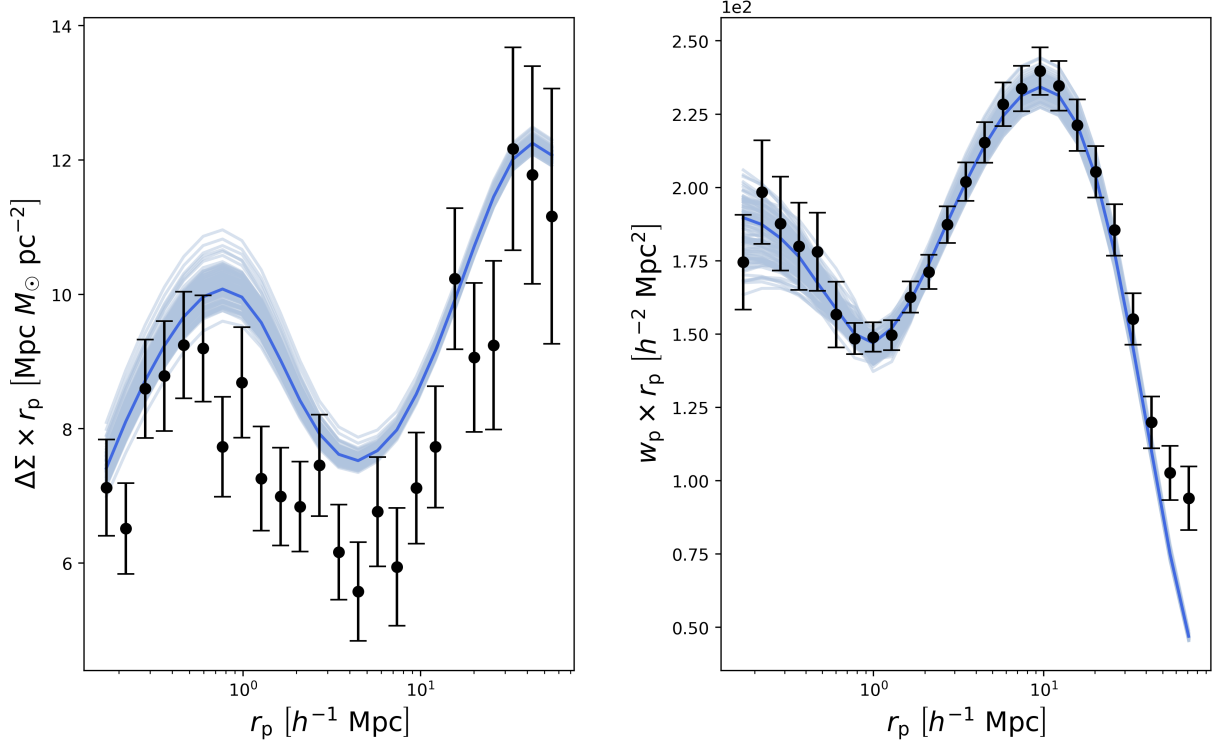


Figure 6.1: Right panel: 5-parameter [HOD](#) fit (blue curve) to the CMASS clustering signal, w_p (black points). Cosmological parameters are fixed at the Planck 2018 TT,TE,EE+lowE+lensing+bao values, and extended [HOD](#) parameters are set to the fiducial values listed in Table 5.1. Left panel: corresponding [GGL](#) prediction using the best-fit parameters from the w_p fit. Light blue curves show models drawn from the posterior distribution of the w_p fit. The 1σ error bars include both statistical and systematic uncertainties.

$$\text{AICc} = 2k - 2\ln(\mathcal{L}_{\max}) + \frac{2k^2 + 2k}{n - k - 1} = 2k + \chi_{\min}^2 + \frac{2k^2 + 2k}{n - k - 1}, \quad (6.1)$$

where k is the number of free parameters, n is the sample size ($24+25+1=50$), $\mathcal{L}_{\max}$ is the maximum likelihood, and $\chi_{\min}^2$ is the corresponding χ^2 value. We set $\chi_{\min}^2 = \chi_{\text{tot}}^2$, where χ_{tot}^2 is the total χ^2 evaluated at the [MAP](#) across all observables: $\chi_{\text{tot}}^2 = \chi_{\Delta\Sigma}^2 + \chi_{w_p}^2 + \chi_{\bar{n}_g}^2$. In the limit of large sample sizes, the [AICc](#) converges to the general Akaike Information Criterion (AIC), which is the [AICc](#) but without the 3rd term. The [BIC](#) is computed via:

$$\text{BIC} = k\ln(n) - 2\ln(\mathcal{L}_{\max}) = k\ln(n) + \chi_{\min}^2. \quad (6.2)$$

Fig. 6.2 presents the results of a joint fit to $\Delta\Sigma$ and w_p using the same setup as the clustering-only fit (HOD curves). For the next few paragraphs, we refer to this configuration as the ‘baseline’ model. Interestingly, the joint fit and fit to w_p alone yield similar [GGL](#) predictions; the maximum percent difference between the two curves is $\sim 3\%$. Consequently, $\Delta\Sigma$ remains over-predicted, indicating that the standard 5-parameter [HOD](#) at Planck-like cosmology struggles to capture the amplitude of $\Delta\Sigma$ even when [GGL](#) information is included in the fitting process. This fit yields a p -value of 0.008—the lowest in this analysis—meaning that we expect a worse fit less than 1% of the time, assuming the current model is correct. A commonly adopted threshold for an ‘acceptable’ fit is $p > 0.05$; under this criterion, the ‘HOD’ model does not provide a satisfactory description of the data, and an alternative model should be considered. The associated [AICc](#) and [BIC](#) values are 80.9 and 89.1, respectively.

We also explore individual extensions to the standard [HOD](#), with results shown in the same figure. We find that the model varying both the [HOD](#) parameters and the satellite concentration normalisation (HOD+ R_c curves) predicts a [GGL](#) signal quite similar to that of the baseline model. Consequently, the p -value (0.012) remains low, and the [AICc](#) (80.4) and [BIC](#) (90.0) hardly change (in fact, the [BIC](#) has become worse). The modest improvement in χ_{tot}^2 is achieved by lowering R_c from its fiducial value of 1 to 0.377 (see Table 6.2 for [MAP](#) values).

The baseline fit is hardly improved by introducing p_{off} and R_{off} as free parameters (HOD+off curves); the p -value remains at its original value of 0.008, and both the [AICc](#) (83.9) and [BIC](#) (94.6) have increased. The [MAP](#) estimates for p_{off} and R_{off} are 0.063 and 4.517, respectively. These values imply that $\sim 6\%$ of central galaxies are offset from their halo centres according to the spatial distribution in Eq. 4.3, where $R_{200\text{m}}R_{\text{off}} = 4.517R_{200\text{m}}$. Despite the relatively large value of R_{off} , the small fraction of affected central galaxies suggests that off-centring minimally impacts the fit.

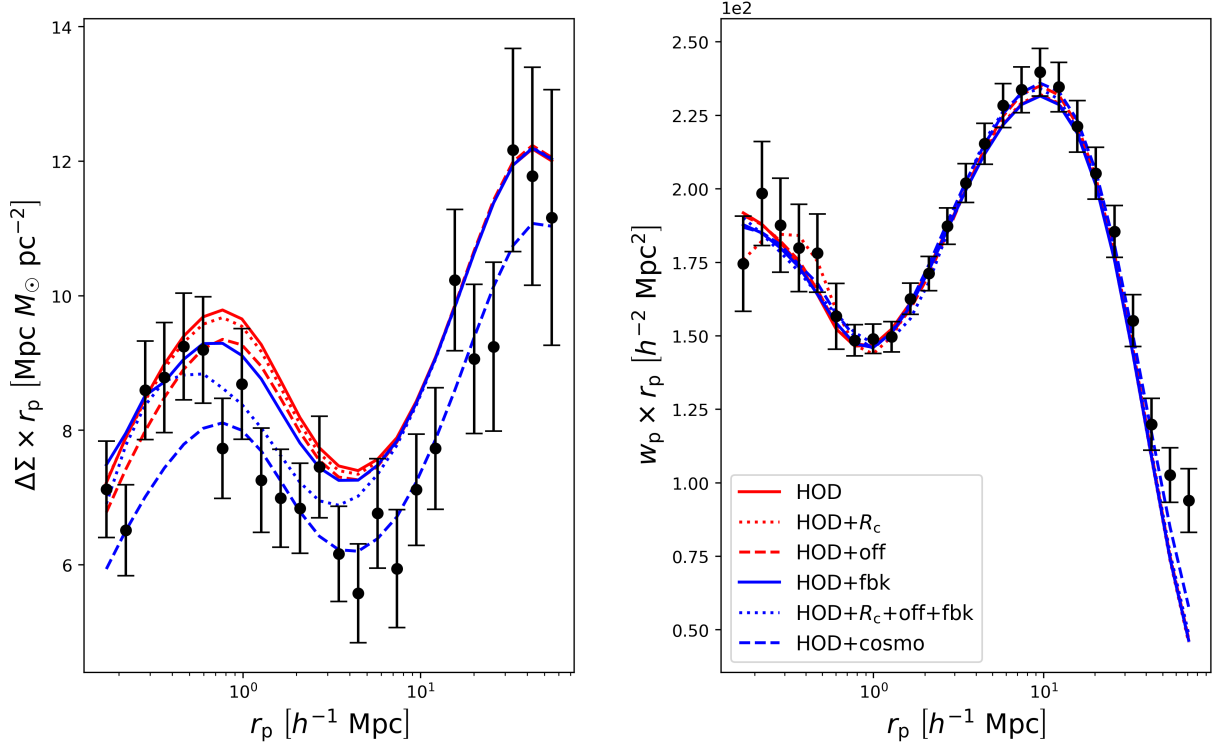


Figure 6.2: Joint fits to both the [GGL](#) and clustering signals of CMASS galaxies. The baseline model (solid red curves) varies only the 5 [HOD](#) parameters, with all other parameters fixed at their fiducial values. A ‘+’ indicates that additional parameters are varied alongside the [HOD](#) parameters. Parameters characterizing a single extension to the standard [HOD](#) are grouped together. For example, ‘off’ indicates that both off-centring parameters (p_{off} and R_{off}) are varied. Feedback parameters are represented by ‘fbk’, and cosmological parameters by ‘cosmo’.

Introducing feedback and its associated parameters ($\log M_c$ and θ_{ej}) to the baseline model increases the p -value to 0.036 (although this is still low), and decreases the [AICc](#) and [BIC](#) to 76.6 and 87.3, respectively (HOD+fbk curves). Notably, the [MAP](#) estimate for θ_{ej} is 16.918. In this scenario, gas within a halo is expelled to a maximum radius of $16.918R_{200c}$, where R_{200c} defines the halo radius². Such extreme gas ejection would require unrealistically long timescales, and would unbind the gas from its host halo. Considering both the p -value and θ_{ej} estimate, this model does not provide a plausible description of the data.

Building on all of these models, we constructed a model that simultaneously varies all parameters except those governing the cosmology (HOD+ R_c +off+fbk curves). The goal was to test whether expanding the parameter space and incorporating different physical effects produces a sufficiently flexible model, capable of capturing the amplitudes of both observables without having to change the fiducial cosmology. Here, the p -value finally reaches a marginally acceptable value of 0.065, which only increases in the following models. However, both the [AICc](#) (78.8) and [BIC](#) (92.3) have increased relative to the previous model, with the [BIC](#) exceeding its original value of 89.1. Recall that lower [AICc](#) and [BIC](#) values indicate model improvement. Interestingly, the [MAP](#) estimate for R_c shifts from 0.377 to 2.486, implying a more centrally concentrated satellite galaxy distribution. This is somewhat counteracted by the off-centring effect, with [MAP](#) estimates of $p_{off} = 0.404$ and $R_{off} = 2.471$. Here, a substantial fraction of central galaxies are offset from their halo centres, reducing the central concentration of galaxies. Feedback is also more physically realistic, with $\log M_c = 14.028$ and $\theta_{ej} = 0.380$. Although this model has 10 free parameters compared to the original 5, $\Delta\Sigma$ remains over-predicted on intermediate–large scales. In fact, all of the models discussed thus far yield identical, over-predicted [GGL](#) signals beyond $\sim 6 h^{-1}$ Mpc, regardless of their complexity.

6.2 HOD With Trial Cosmologies

When we allow freedom in the cosmology through the parameters $\ln(10^{10}A_s)$ and $\Omega_{m,0}$ (HOD+cosmo curves), we observe considerable reductions in the [AICc](#) (63.3) and [BIC](#) (74.0). Most notably, $\Delta\Sigma$ is no longer over-predicted on intermediate–large scales, although the small-scale signal is under-predicted. The [MAP](#) estimates for cosmological parameters are $\ln(10^{10}A_s) = 3.057$ and $\Omega_{m,0} = 0.286$, corresponding to $\sigma_8 = 0.761$ and $S_8 = 0.743$. In comparison, the fiducial Planck cosmology yields $\sigma_8 = 0.809$ and $S_8 = 0.824$ —considerably

²Recall that this is not the only halo radius definition. For example, DARK EMULATOR uses M_{200b} .

higher than our estimates. While our estimate for $\ln(10^{10}A_s)$ closely matches the fiducial value of 3.047, the estimate for $\Omega_{m,0}$ is substantially lower and hits the lower bound of our prior range. This low value of $\Omega_{m,0}$ drives the reduced S_8 and σ_8 values. While we provide the **MAP** estimates and constraints for cosmological parameters, we caution the reader against over-interpreting these values for two reasons. First, **IAs** may be a non-negligible contaminant of the measured **GGL** signal, especially considering the significant overlap between lens and source galaxy redshifts. Internal tests indicate that **IAs** could impact our measurements of the **GGL** signal at the 10% level. Second, we vary only a small subset of cosmological parameters, whereas a full analysis would require marginalization over the entire parameter space.

If $\Omega_{m,0}$ is reduced relative to some fiducial value, then the **GGL** signal is suppressed, while the clustering signal is enhanced (see Fig. D.11). This behaviour arises since $\Delta\Sigma \propto \Omega_{m,0}b_g\xi_{mm}$ and $w_p \propto b_g^2\xi_{mm}$. Strictly speaking, these proportionalities are only valid on linear scales. Lowering $\Omega_{m,0}$ decreases the amplitude of ξ_{mm} , which in turn increases b_g for a fixed **HOD** since galaxies become more biased tracers of the underlying matter distribution. Due to the different powers of b_g in our observables, changes in $\Omega_{m,0}$ affect $\Delta\Sigma$ and w_p in opposite directions.

Adjusting $\Omega_{m,0}$ from its fiducial value of 0.3111 to 0.286 results in a $\sim 16\%$ increase in the amplitude of w_p near its peak ($\sim 10 h^{-1} \text{Mpc}$). The clustering signal is slightly enhanced relative to the baseline prediction (resulting in a lower $\chi^2_{w_p}$), but not at the 16% level, indicating that the **HOD** parameters must have compensated to maintain consistency with w_p . Indeed, we observe non-negligible shifts in the **MAP** estimates and posterior distributions of **HOD** parameters (see Fig. 6.3). The most notable change is a considerable increase in the **MAP** estimate for $\sigma_{\log M}^2$ from 0.001 to 0.304. This shift reflects a much more softened transition from 0 to 1 central galaxy in the mean occupation distribution ($\langle N_c \rangle$), enabling more low-mass halos to host central galaxies and thereby reducing b_g (see Fig. D.3). Interestingly, the HOD+ R_c +off+fbk model yields similar **MAP** estimates for $\log M_{\min}$, $\sigma_{\log M}^2$, and $\log M_1$.

We explore the impact of progressively increasing model complexity in Fig. 6.4. Here, we adopt the HOD+cosmo model as the baseline configuration since it provides the best joint fit out of the models previously discussed. We find that freeing the off-centring parameters (HOD+cosmo+off curves) produces a slightly worse fit relative to the baseline model, resulting in higher **AICc** and **BIC** values (69.7 and 82.4, respectively). Adding free parameters to a model should either maintain or improve its performance. The small increase in χ^2_{tot} likely stems from the walkers' preference for $p_{\text{off}} = 0$ —the baseline model's fiducial value—which is difficult to sample precisely. This preference for zero off-centring is evident in the posterior distributions shown in Fig. 6.5.

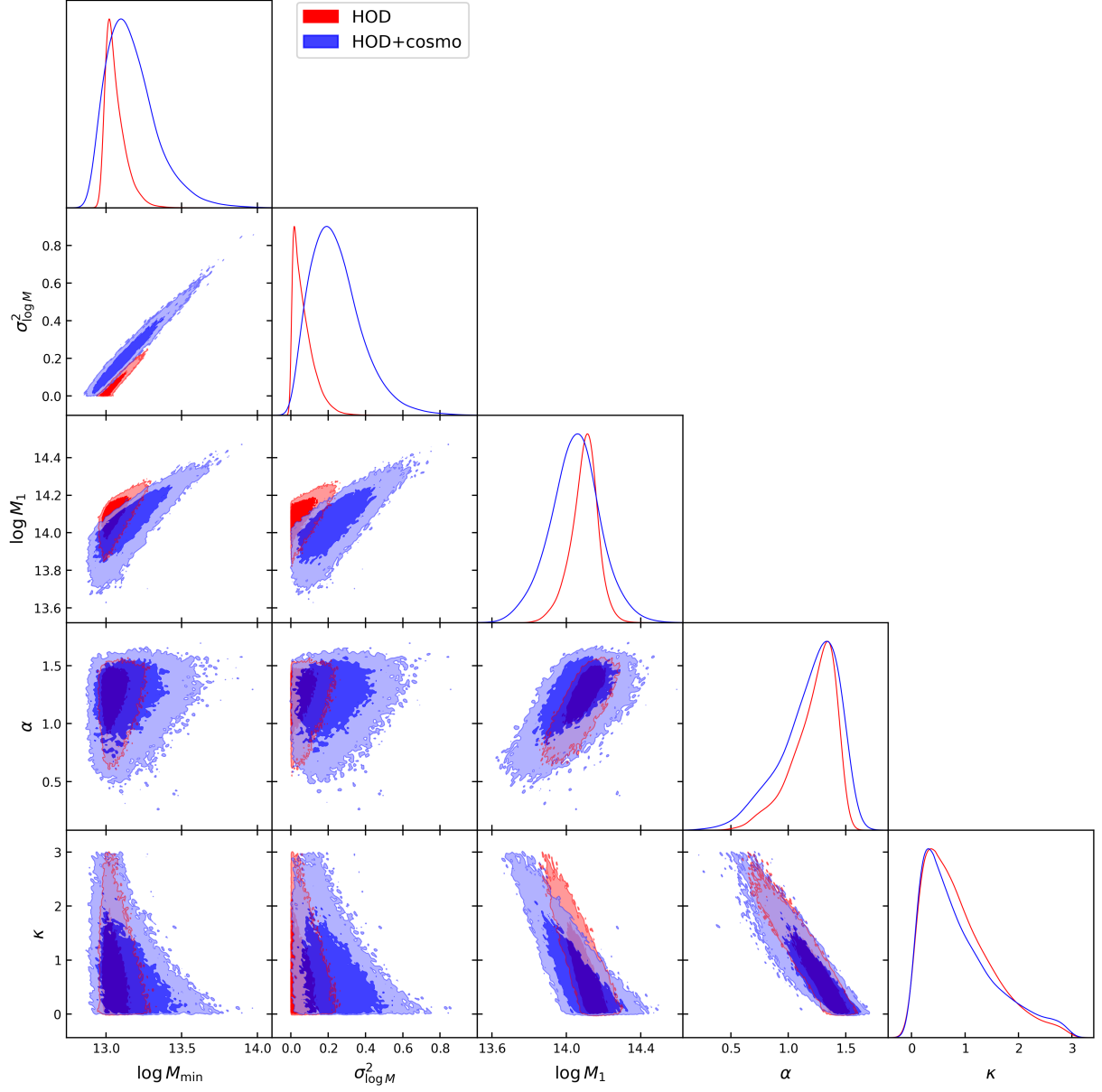


Figure 6.3: **HOD** Posterior distributions for the HOD and HOD+cosmo models. Median values and their uncertainties are listed in Table 6.3.

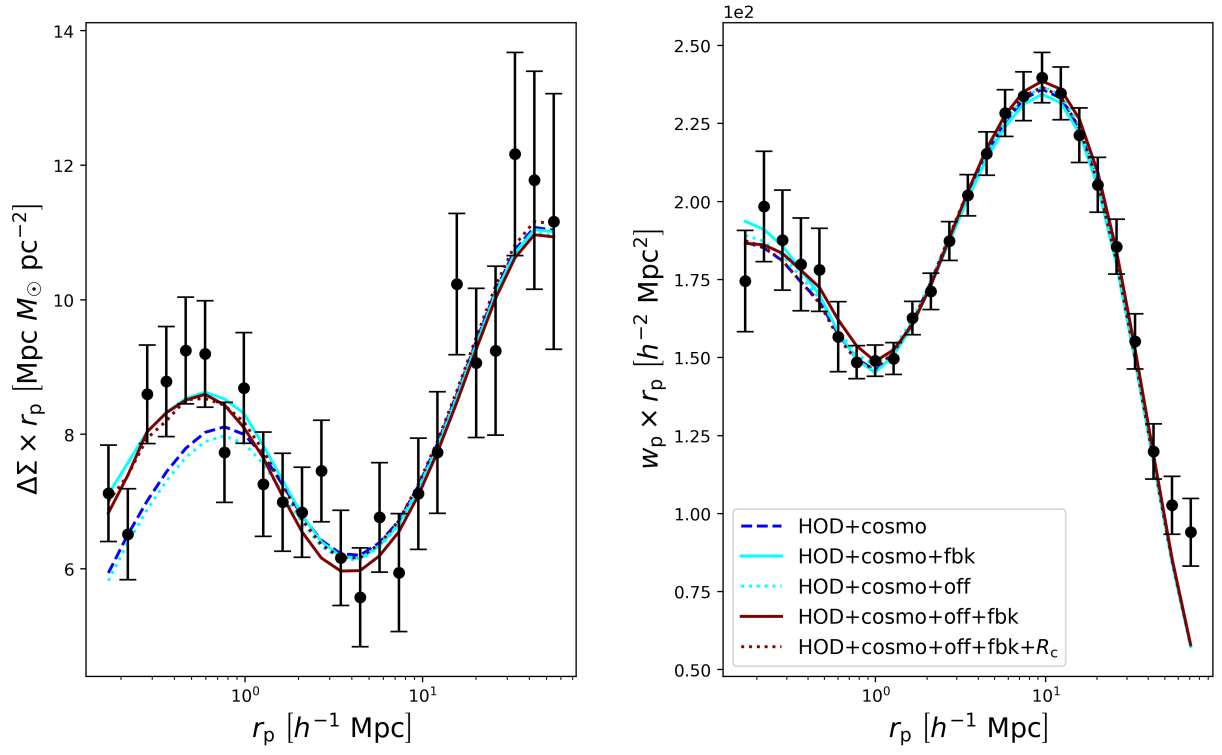


Figure 6.4: Similar to Fig. 6.2, except the baseline model is now ‘HOD+cosmo’.

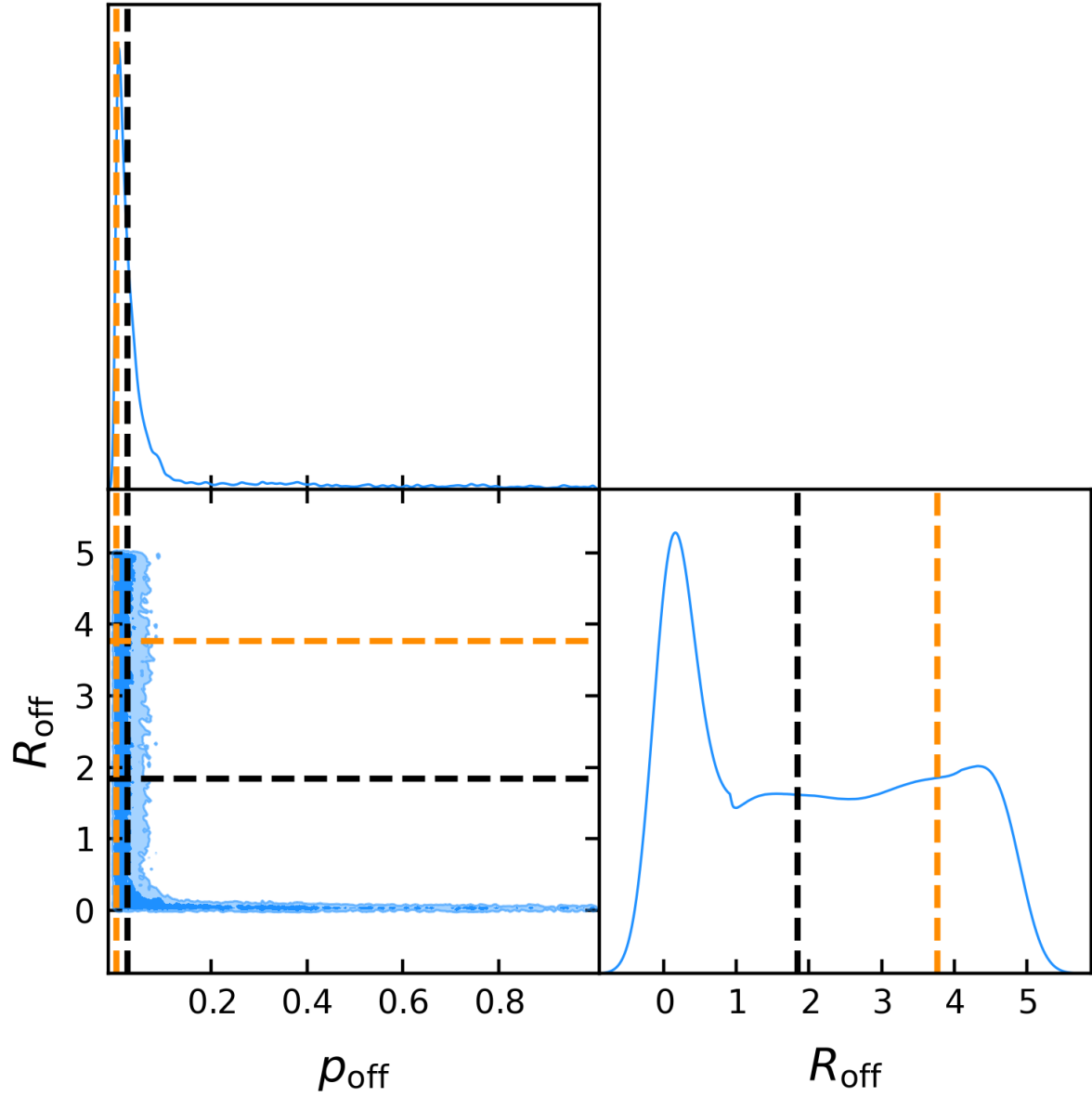


Figure 6.5: Off-centring posterior distributions for the HOD+cosmo+off model. Black lines indicate median values, while orange lines mark the [MAP](#) estimates. The off-centring effect vanishes if either p_{off} or R_{off} goes to zero.

Introducing feedback into the previous model (HOD+cosmo+off+fbk curves) yields the best fit to $\Delta\Sigma$. Notably, $\Delta\Sigma$ is much less under-predicted on small scales. Relative to the previous model, the [AICc](#) and [BIC](#) are reduced to 67.6 and 81.7, respectively. However, this reduction is not enough to outperform the baseline model. Here, the off-centring effect is no longer consistent with zero, with [MAP](#) estimates of $p_{\text{off}} = 0.078 = 7.8\%$ and $R_{\text{off}} = 3.277$. The corresponding estimates for feedback parameters are $\log M_c = 12.450$ and $\theta_{\text{ej}} = 1.318$. Due the low value of $\log M_c$ and moderate value of $\theta_{\text{ej}} = 1.318$, feedback is actually enhancing small-scale power compared to the previous fit. Recall from Fig. 4.1 that lower values of feedback parameters enhance the gas density profile on small scales.

Allowing the satellite concentration normalisation to vary (HOD+cosmo+off+fbk+ R_c curves) does not improve the previous model. In fact, χ^2_{tot} slightly increases, possibly due to scatter in the [MAP](#) estimates. Consequently, the [AICc](#) and [BIC](#) increase to 71.1 and 85.6, respectively. The [MAP](#) estimate for R_c is 0.930—close to the fiducial value of unity. Given this value of R_c , one might expect similar parameter estimates between the previous model and the current one. However, there are notable differences in the [MAP](#) estimates. For instance, p_{off} shifts from 0.078 to 0.171, and R_{off} from 3.277 to 2.031. Additionally, $\log M_c$ shifts from 12.450 to 14.255, and θ_{ej} from 1.318 to 0.474. These changes may be explained by parameter degeneracies. For example, low (high) values of p_{off} are somewhat degenerate with high (low) values of R_{off} ³. A similar relationship exists between $\log M_c$ and θ_{ej} . Indeed, the shifts in the [MAP](#) estimates follow the directions of these degeneracies. In addition to parameter degeneracies, there are degeneracies in how χ^2_{tot} can be minimized. The previous model achieves the lowest $\chi^2_{\Delta\Sigma}$ and a lower $\chi^2_{\bar{n}_g}$, whereas the current model yields the lowest $\chi^2_{w_p}$. Furthermore, the [MAP](#) estimate for R_c does not exactly match the fiducial value, so some variation in the other parameters is expected (as previously mentioned, we also expect some scatter in the [MAP](#) estimates). This model is the most complex, with 12 free parameters. As such, we show the resulting posterior distributions in Fig. D.1.

Since cosmological and feedback parameters appeared to have the largest impact on the joint fit, we constructed a model that varies these alongside the [HOD](#) parameters (HOD+cosmo+fbk curves). We find this model to be competitive with the baseline configuration (HOD+cosmo); while the [BIC](#) has slightly increased to 74.6, the [AICc](#) has decreased to 61.8. The [MAP](#) estimates for [HOD](#) parameters and their posterior distributions (see Fig. 6.6) are similar between the two models. This similarity is not surprising, given that the baseline model already provides a good fit to $\Delta\Sigma$ on intermediate–large scales, and to w_p across most scales. Our feedback model specifically modifies the small-scale

³These parameters are degenerate to some extent, as each distinctly affects the shapes of the [GGL](#) and clustering predictions (see Figures D.7 and D.8).

Model	$\chi^2_{\Delta\Sigma}$	$\chi^2_{w_p}$	χ^2_{tot}	p	AICc	BIC	k
HOD	45.5	22.7	69.5	0.008	80.9	89.1	5
HOD+ R_c	43.3	22.4	66.5	0.012	80.4	90.0	6
HOD+off	44.6	21.5	67.3	0.008	83.9	94.6	7
HOD+fbk	36.5	22.6	59.9	0.036	76.6	87.3	7
HOD+ R_c +off+fbk	30.3	22.7	53.1	0.065	78.8	92.3	10
HOD+cosmo	30.5	15.8	46.6	0.288	63.3	74.0	7
HOD+cosmo+fbk	23.1	16.2	39.3	0.499	61.8	74.6	9
HOD+cosmo+off	30.9	16.2	47.2	0.203	69.7	82.4	9
HOD+cosmo+off+fbk	21.7	16.6	38.6	0.440	67.6	81.7	11
HOD+cosmo+off+fbk+ R_c	22.3	15.6	38.7	0.394	71.1	85.6	12

Table 6.1: Statistical values from best-fit models. This table only includes results from fits to both $\Delta\Sigma$ and w_p . The number of free parameters is represented by k , and the total data vector length across all observables (including $\bar{n}_g$) is 50.

GGL signal. So, feedback parameters can fine-tune the amplitude of $\Delta\Sigma$ on small scales without impacting the fit at larger scales or the fit to w_p . The **MAP** estimates for feedback parameters are $\log M_c = 11.172$ and $\theta_{ej} = 2.938$.

Note that all ‘cosmo’ models place $\Omega_{m,0}$ at 0.286—its lowest possible value—and yield similar **MAP** estimates for $\ln(10^{10} A_s)$. Additionally, all cosmo models predict nearly identical **GGL** signals on scales larger than $10 h^{-1} \text{Mpc}$. It is also worth mentioning that $\log M_{\min} \sim 13$ and $\log M_1 \sim 14$ are robust predictions across all models, providing robust evidence that the minimum halo mass required to host a CMASS-like galaxy is $\sim 10^{13} h^{-1} M_\odot$.

In addition to the models discussed above, we analysed variants that incorporate halo mass incompleteness. For a detailed description of the incompleteness function and its parameters, we refer readers to [More et al. \(2015\)](#). We found that the incompleteness parameters were degenerate with **HOD** parameters, and that $\log M_{\min}$ estimates were unusually low. Since the incompleteness effect did not improve the joint fit, we omit related models from our discussion.

Comparing all of the different configurations, we have two preferred models: HOD+cosmo and HOD+cosmo+fbk. The HOD+cosmo model yields the lowest **BIC** score (74.0) and the second-lowest **AICc** score (63.3), while the HOD+cosmo+fbk model yields the lowest **AICc** score (61.8) and the second-lowest **BIC** score (74.6). Regardless, our observables are best fit by a model with flexible cosmology, favouring a low value of S_8 relative to Planck.

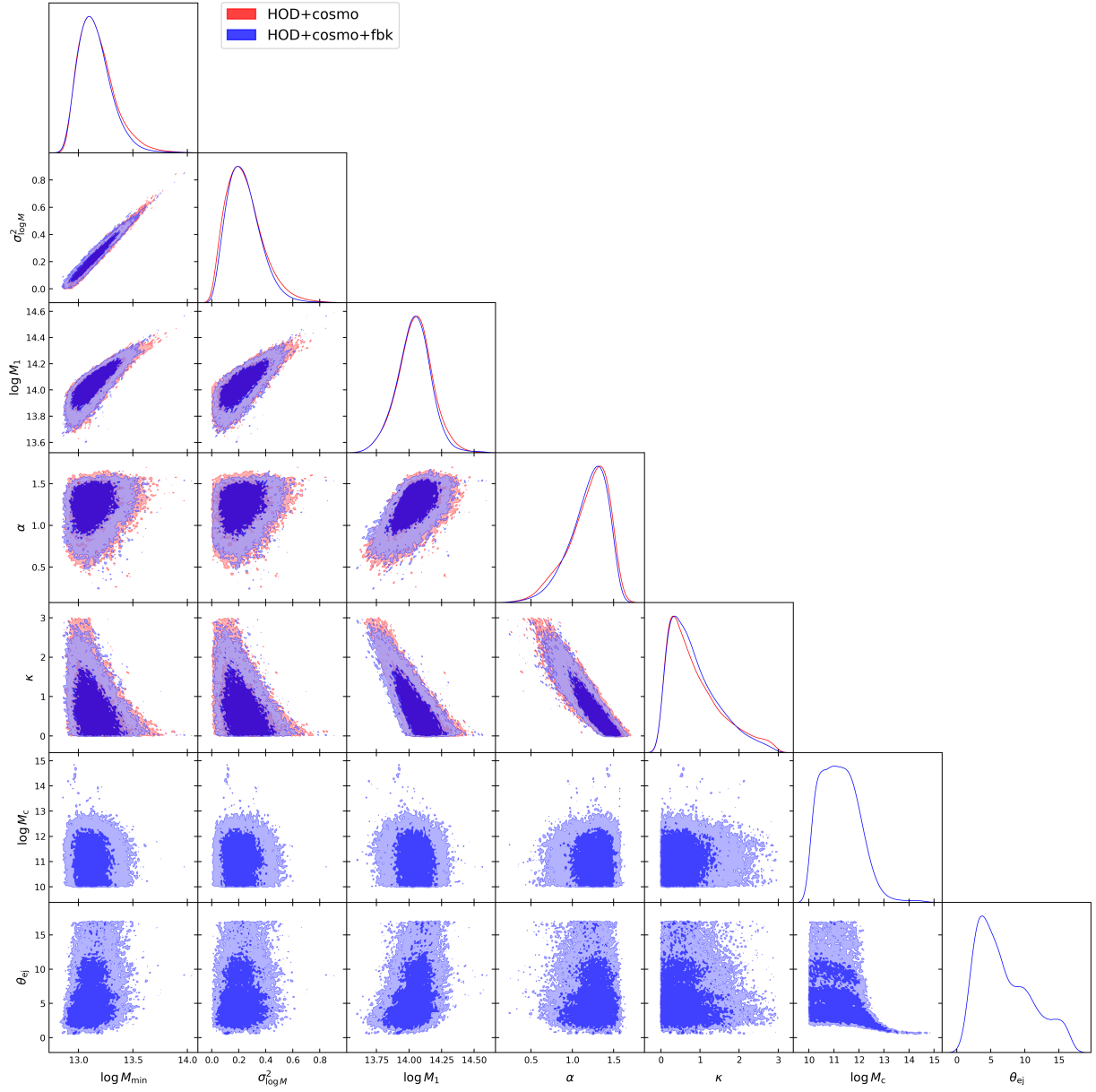


Figure 6.6: Posterior distributions for the HOD+cosmo and HOD+cosmo+fbk models.

Parameter	Model			
	HOD	HOD+ R_c	HOD+off	HOD+fbk
$\log M_{\min}$	12.983	13.004	12.997	13.016
$\sigma_{\log M}^2$	0.001	0.002	0.005	0.027
$\log M_1$	14.114	14.152	14.106	14.105
α	1.416	1.291	1.442	1.345
κ	0.107	0.092	0.032	0.467
R_c	–	0.377	–	–
p_{off}	–	–	0.063	–
R_{off}	–	–	4.517	–
$\log M_c$	–	–	–	11.432
θ_{ej}	–	–	–	16.918

Parameter	Model			
	HOD+cosmo	HOD+cosmo+fbk	HOD+cosmo+off	HOD+cosmo+off+fbk
$\log M_{\min}$	13.217	13.137	13.172	13.064
$\sigma_{\log M}^2$	0.304	0.249	0.270	0.196
$\log M_1$	14.132	14.076	14.106	13.992
α	1.359	1.341	1.411	1.358
κ	0.292	0.282	0.216	0.435
R_c	–	–	–	–
p_{off}	–	–	0.004	0.078
R_{off}	–	–	3.764	3.277
$\log M_c$	–	11.172	–	12.450
θ_{ej}	–	2.938	–	1.318
$\ln(10^{10} A_s)$	3.057	3.054	3.039	3.031
$\Omega_{\text{m},0}$	0.286	0.286	0.286	0.286
σ_8	0.761	0.761	0.756	0.752
S_8	0.743	0.743	0.738	0.734

Table 6.2: Best-fit parameter values drawn from the [MAP](#). For the fiducial Planck cosmology, $\sigma_8 = 0.809$ and $S_8 = 0.824$.

Parameter	Model			
	HOD	HOD+ R_c	HOD+off	HOD+fbk
$\log M_{\min}$	$13.047^{+0.080}_{-0.045}$	$13.048^{+0.080}_{-0.045}$	$13.070^{+0.099}_{-0.057}$	$13.102^{+0.147}_{-0.077}$
$\sigma^2_{\log M}$	$0.053^{+0.069}_{-0.039}$	$0.050^{+0.072}_{-0.037}$	$0.068^{+0.081}_{-0.050}$	$0.091^{+0.113}_{-0.065}$
$\log M_1$	$14.102^{+0.061}_{-0.078}$	$14.115^{+0.072}_{-0.010}$	$14.117^{+0.067}_{-0.070}$	$14.135^{+0.105}_{-0.091}$
α	$1.262^{+0.141}_{-0.233}$	$1.248^{+0.145}_{-0.199}$	$1.317^{+0.120}_{-0.212}$	$1.228^{+0.170}_{-0.288}$
κ	$0.729^{+0.785}_{-0.497}$	$0.643^{+0.924}_{-0.472}$	$0.525^{+0.700}_{-0.380}$	$0.774^{+0.857}_{-0.529}$
R_c	—	$0.886^{+0.864}_{-0.417}$	—	—
p_{off}	—	—	$0.056^{+0.044}_{-0.035}$	—
R_{off}	—	—	$3.084^{+1.264}_{-1.493}$	—
$\log M_c$	—	—	—	$12.218^{+4.041}_{-1.461}$
θ_{ej}	—	—	—	$13.669^{+2.355}_{-3.840}$

Parameter	Model			
	HOD+cosmo	HOD+cosmo+fbk	HOD+cosmo+off	HOD+cosmo+off+fbk
$\log M_{\min}$	$13.144^{+0.182}_{-0.136}$	$13.134^{+0.164}_{-0.130}$	$13.150^{+0.202}_{-0.135}$	$13.125^{+0.150}_{-0.121}$
$\sigma^2_{\log M}$	$0.224^{+0.153}_{-0.118}$	$0.226^{+0.135}_{-0.109}$	$0.225^{+0.165}_{-0.117}$	$0.223^{+0.123}_{-0.106}$
$\log M_1$	$14.050^{+0.126}_{-0.136}$	$14.041^{+0.119}_{-0.131}$	$14.061^{+0.134}_{-0.128}$	$14.027^{+0.110}_{-0.133}$
α	$1.239^{+0.198}_{-0.305}$	$1.230^{+0.190}_{-0.271}$	$1.270^{+0.182}_{-0.302}$	$1.230^{+0.198}_{-0.273}$
κ	$0.693^{+0.903}_{-0.480}$	$0.702^{+0.805}_{-0.477}$	$0.620^{+0.828}_{-0.444}$	$0.658^{+0.814}_{-0.469}$
R_c	—	—	—	—
p_{off}	—	—	$0.026^{+0.105}_{-0.019}$	$0.063^{+0.134}_{-0.048}$
R_{off}	—	—	$1.843^{+2.223}_{-1.789}$	$1.834^{+1.987}_{-1.699}$
$\log M_c$	—	$11.178^{+0.825}_{-0.785}$	—	$11.263^{+1.246}_{-0.884}$
θ_{ej}	—	$6.039^{+5.541}_{-3.126}$	—	$3.038^{+5.111}_{-1.831}$
$\ln(10^{10} A_s)$	$2.943^{+0.097}_{-0.151}$	$2.979^{+0.089}_{-0.091}$	$2.943^{+0.096}_{-0.176}$	$2.983^{+0.085}_{-0.089}$
$\Omega_{\text{m},0}$	$0.294^{+0.016}_{-0.006}$	$0.292^{+0.009}_{-0.004}$	$0.295^{+0.018}_{-0.007}$	$0.291^{+0.009}_{-0.004}$
σ_8	$0.737^{+0.027}_{-0.031}$	$0.743^{+0.028}_{-0.025}$	$0.739^{+0.030}_{-0.033}$	$0.744^{+0.029}_{-0.025}$
S_8	$0.733^{+0.022}_{-0.023}$	$0.736^{+0.025}_{-0.026}$	$0.734^{+0.024}_{-0.023}$	$0.737^{+0.026}_{-0.025}$
$\log M_{\min}$	—	—	—	$13.129^{+0.148}_{-0.131}$
$\sigma^2_{\log M}$	—	—	—	$0.233^{+0.130}_{-0.108}$
$\log M_1$	—	—	—	$13.983^{+0.135}_{-0.153}$
α	—	—	—	$1.251^{+0.221}_{-0.302}$
κ	—	—	—	$0.943^{+0.988}_{-0.678}$
R_c	—	—	—	$1.929^{+0.740}_{-0.889}$
p_{off}	—	—	—	$0.070^{+0.153}_{-0.051}$
R_{off}	—	—	—	$1.985^{+1.964}_{-1.711}$
$\log M_c$	—	—	—	$11.457^{+1.299}_{-0.961}$
θ_{ej}	—	—	—	$2.905^{+4.124}_{-2.046}$
$\ln(10^{10} A_s)$	—	—	—	$2.978^{+0.087}_{-0.090}$
$\Omega_{\text{m},0}$	—	—	—	$0.291^{+0.009}_{-0.004}$
σ_8	—	—	—	$0.740^{+0.035}_{-0.028}$
S_8	—	—	—	$0.730^{+0.032}_{-0.024}$

Table 6.3: Posterior constraints from **MCMC** chains. We report the 50th percentile as the central value, with lower and upper limits given by the differences between the 50th and 16th percentiles, and the 84th and 50th percentiles, respectively. We quote the median instead of the mean since many of the posterior distributions are skewed.

6.3 Comparison

To compare our ‘best’ models with results from other studies, we first compute the mean host-halo mass, $\overline{M}_h$:

$$\overline{M}_h = \frac{\int M \langle N_c(M) \rangle \frac{dn_h}{dM} dM}{\int \langle N_c(M) \rangle \frac{dn_h}{dM} dM}. \quad (6.3)$$

For the HOD+cosmo model, we obtain $\log_{10} (\overline{M}_h / h^{-1} M_\odot) = 13.351$. Comparable results are presented in [Saito et al. \(2016\)](#), who constructed mock CMASS catalogues using both a standard SHAM methodology and an extended version incorporating age matching. For the standard SHAM case, they found mean halo masses of $\log_{10} (\overline{M}_h / h^{-1} M_\odot) = 13.120$, 13.340, and 13.660 for central galaxies at redshifts $z = 0.445$, 0.565, and 0.685, respectively. The corresponding values for the age matching approach are $\log_{10} (\overline{M}_h / h^{-1} M_\odot) = 13.150$, 13.350, and 13.680. Converting our mean halo mass to the ROCKSTAR definition used in [Saito et al. \(2016\)](#) via the Python package COLOSSUS ([Diemer, 2018](#)), we get $\log_{10} (\overline{M}_h / h^{-1} M_\odot) = 13.330$, which closely matches the [Saito et al. \(2016\)](#) result at $z = 0.565$ (our fiducial redshift is $z = 0.54$). Since our feedback procedure alters the M_{200m} of a given halo, we do not compute a mean host-halo mass for the HOD+cosmo+fbk model.

We also calculate the mean satellite galaxy fraction, $\overline{f}_s$:

$$\overline{f}_s = \frac{\int \langle N_s(M) \rangle \frac{dn_h}{dM} dM}{\int \langle N_g(M) \rangle \frac{dn_h}{dM} dM} = \frac{\overline{n}_s}{\overline{n}_g}. \quad (6.4)$$

We find $\overline{f}_s = 0.085 = 8.5\%$ for the HOD+cosmo model, and $\overline{f}_s = 0.098 = 9.8\%$ for the HOD+cosmo+fbk model. Note that the feedback parameters do not affect the computation of $\overline{f}_s$. In comparison, [Saito et al. \(2016\)](#) report satellite fractions between 8.0 and 12.5% across the redshift bins previously listed. Additionally, [White et al. \(2011\)](#) find satellite fractions between 5.0 and 12.0% (including 1σ uncertainties), based on an HOD fit to w_p over approximately $0.4\text{--}23 h^{-1} \text{Mpc}$ and within the redshift range $0.4 < z < 0.7$. Lastly, [More et al. \(2015\)](#) quote satellite fractions of 2.5–11.4% (including 1σ uncertainties) between different stellar mass subsamples in the redshift range $0.47 \leq z \leq 0.59$.

Finally, we compute the large-scale effective galaxy bias, b_{eff} :

$$b_{\text{eff}} = \frac{\int b(M) \langle N_g(M) \rangle \frac{dn_h}{dM} dM}{\int \langle N_g(M) \rangle \frac{dn_h}{dM} dM}, \quad (6.5)$$

where $b(M)$ is the linear halo bias at mass M . We obtain a large-scale effective galaxy bias of 2.15 for both the HOD+cosmo and HOD+cosmo+fbk models—precisely the same central value reported by [More et al. \(2015\)](#) for their fiducial stellar mass bin ($11.1 \leq \log_{10}(M_*/h^{-2}M_\odot) \leq 12.0$). Since our feedback model only modifies ξ_{mm} (or P_{mm}) on small scales, b_{eff} is unaffected (although the scale-dependent galaxy bias would change on small scales). [White et al. \(2011\)](#) also find a large-scale galaxy bias of ~ 2 , which tends to be the value quoted in the literature.

6.4 Discussion

Our results demonstrate that the lensing-is-low effect persists on large scales, and is not confined to small scales. The precision of our large-scale $\Delta\Sigma$ measurements now demands more accurate predictions at these scales and enables us to distinguish between different cosmologies. We find that the HOD+cosmo model, one of the preferred models with the lowest [AICc](#) and [BIC](#) scores, does not provide a good fit to $\Delta\Sigma$ on small scales. However, due to its very good fit on intermediate-large scales, HOD+cosmo remains one of the best models, reinforcing the idea that the large-scale [GGL](#) signal plays a key role in the fits. This stands in contrast to [Amon et al. \(2023\)](#), who found that the large-scale data alone lacked the statistical power to distinguish between Planck and lensing cosmologies ($S_8 = 0.83$ and 0.76 , respectively). They also concluded that there was no compelling evidence for a mismatch between $\Delta\Sigma$ and w_p on large scales. Note that our $\Delta\Sigma$ error bars are smaller than those reported by [Amon et al. \(2023\)](#) for various [BOSS](#) subsamples. For example, their 1σ error on $\Delta\Sigma$ is ~ 3.6 at $r_p \sim 53.8$ for CMASS galaxies in the redshift range $0.54 < z < 0.7$, whereas our corresponding error is ~ 1.9 .

While we find that a low- S_8 cosmology largely brings $\Delta\Sigma$ and w_p into agreement, we cannot interpret our results as evidence for a low value of S_8 . This is mainly because [IAs](#), which also affect $\Delta\Sigma$ on large scales ([Prat et al., 2022](#)), have not yet been investigated/mitigated. If [IAs](#) prove to be a significant contaminant, their mitigation could considerably boost the amplitude of the measured [GGL](#) signal (or reduce the amplitude of the predicted signal). This could potentially reconcile $\Delta\Sigma$ and w_p without requiring a departure from the Planck cosmology. Ideally, [IAs](#) would be mitigated through tomography since the [IA](#) amplitude in models is degenerate with (positively correlated to) S_8 . Ensuring that the lens and source galaxies are well-separated in redshift would reduce the likelihood of physical associations, which cause [IAs](#). This strategy will be possible in the future, once photometric redshifts are available in [UNIONS](#). For now, we simply argue that a large-scale modification is required to bring $\Delta\Sigma$ and w_p into consistency under the

standard [HOD](#) framework.

Chapter 7

Conclusions

In this thesis, we revisit the ‘lensing is low’ problem introduced by [Leauthaud et al. \(2017\)](#) through a joint analysis of the [GGL](#) and clustering signals of [BOSS DR12](#) CMASS galaxies. We provide a fresh perspective on the problem by leveraging new [GGL](#) measurements from Ultraviolet Near Infrared Optical Northern Survey ([UNIONS](#)) data and analysing several extensions of the standard [HOD](#). We adopt the [HOD](#) parametrization of [Zheng et al. \(2005\)](#), implemented via the DARK EMULATOR. This model is characterized by 5 parameters describing the occupation of halos by central and satellite galaxies. We extend this framework to include central galaxy off-centring, feedback, and a modified satellite concentration normalisation. Our feedback method follows the [BC](#) models of [Schneider et al. \(2019\)](#) and [Shirasaki et al. \(2024\)](#), and modifies the 1-halo term of the halo–matter cross–power spectrum. Model parameters are constrained using [MCMC](#) sampling, where the [GGL](#) and clustering signals are fit simultaneously. We also use the observed galaxy abundance as an additional constraint on the fits. Our main results are as follows:

- We confirm a lensing–is–low effect in the CMASS sample using our new [GGL](#) measurements. Adopting the Planck 2018 TT,TE,EE+lowE+lensing+bao cosmology, the predicted [GGL](#) signal from a standard 5-parameter [HOD](#) fit to the clustering alone exceeds the amplitude of our measurements by $\sim 13\%$. Although we find a lensing–is–low effect, other works place the discrepancy between [GGL](#) predictions and measurements at higher values (20–40%).
- Our data continues to disfavour the standard [HOD](#), even when [GGL](#) is included in the fitting procedure. The maximum percent difference between the clustering-only and clustering+[GGL](#) fit is approximately 3%.

- All model variants assuming the Planck cosmology over-predict the [GGL](#) signal on intermediate–large scales.
- We can only resolve the large-scale [GGL](#) discrepancy by lowering the value of S_8 relative to Planck. Without the inclusion of feedback, such models under-predict the small-scale [GGL](#) signal.
- Using both the Akaike Information Criterion for small sample sizes ([AICc](#)) and the Bayesian Information Criterion ([BIC](#)), we find two preferred models: one where [HOD](#) and cosmological parameters are varied (HOD+cosmo), and one where [HOD](#), cosmological, and feedback parameters are varied (HOD+cosmo+fbk).
- The HOD+cosmo model yields a mean host-halo mass of $\log(\overline{M}_h/h^{-1} M_\odot) = 13.351$ and a mean satellite galaxy fraction of $\overline{f}_s = 0.085 = 8.5\%$. Meanwhile, the HOD+cosmo+fbk model yields a mean satellite galaxy fraction of $\overline{f}_s = 0.098 = 9.8\%$. Both models yield a large-scale effective galaxy bias of $b_{\text{eff}} = 2.15$.

Our results suggest that the lensing is low problem may not be explained by small-scale astrophysical effects alone, which has been the prevailing idea thus far. The high precision of our large-scale [GGL](#) measurements precludes models fixed at the Planck cosmology from alleviating the tension between [GGL](#) and clustering. Before definitive statements can be made about S_8 , the impact of [IAs](#) must be understood. In this analysis, we assume an [IA](#) amplitude of zero, though [IA](#) contamination may be non-negligible. If this is the case, mitigating [IAs](#) would boost the amplitude of the measured [GGL](#) signal (or reduce the amplitude of the predicted signal), which could bring [GGL](#) and clustering closer to consistency. This would, in turn, impact our estimate of S_8 . We leave a detailed investigation of [IAs](#) to future work, when tomography is available.

Other promising avenues include extending the analysis to the LOWZ sample (similar to CMASS but at lower redshifts), which also exhibits the lensing–is–low effect, and constraining feedback parameters with measurements of the kinetic Sunyaev–Zel’dovich (kSZ) effect (see [Bigwood et al., 2024](#)). It would also be interesting to investigate [GAB](#). However, this is not possible within the current DARK EMULATOR framework, and no purely analytic description exists for assembly bias. Such an investigation would likely require the use of other simulation work. Since [GAB](#) primarily impacts small scales ([Amon et al., 2023](#)), it is unlikely that accounting for this effect would significantly alter our primary findings. This work represents a step towards accurate modelling of [GGL](#) and clustering across non-linear and linear scales, which is crucial for disentangling the lensing–is–low effect from genuine evidence of a low S_8 value. It also highlights the necessity of extracting information from all scales.

References

- Abadi M. G., Navarro J. F., Fardal M., Babul A., Steinmetz M., 2010, [MNRAS](#), **407**, 435
- Abdo A. A., n.d., Gravitational Lensing, <https://web.pa.msu.edu/people/abdo/GravitationalLensing.pdf>
- Abdul Karim M., et al., 2025, [Phys. Rev. D](#), **112**, 083515
- Alam S., et al., 2017, [MNRAS](#), **470**, 2617
- Amon A., et al., 2023, [MNRAS](#), **518**, 477
- Anderson L., et al., 2012, [MNRAS](#), **427**, 3435
- Asgari M., Mead A. J., Heymans C., 2023, [The Open Journal of Astrophysics](#), **6**, 39
- Bahcall N. A., Ostriker J. P., Perlmutter S., Steinhardt P. J., 1999, [Science](#), **284**, 1481
- Barnes J., White S. D. M., 1984, [MNRAS](#), **211**, 753
- Bartelmann M., 2010, [Classical and Quantum Gravity](#), **27**, 233001
- Bartelmann M., Schneider P., 2001, [Phys. Rep.](#), **340**, 291
- Bennett C. L., et al., 2003, [ApJS](#), **148**, 1
- Bianchi D., Percival W. J., 2017, [MNRAS](#), **472**, 1106
- Bigwood L., et al., 2024, [MNRAS](#), **534**, 655
- Bigwood L., et al., 2025, [arXiv e-prints](#), p. [arXiv:2512.04209](#)
- Blumenthal G. R., Faber S. M., Flores R., Primack J. R., 1986, [ApJ](#), **301**, 27

Boggess N. W., et al., 1992, [ApJ](#), **397**, 420

Bridle S. L., Crittenden R., Melchiorri A., Hobson M. P., Kneissl R., Lasenby A. N., 2002, [MNRAS](#), **335**, 1193

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

Chaves-Montero J., Angulo R. E., Contreras S., 2023, [MNRAS](#), **521**, 937

Cole S., et al., 2005, [MNRAS](#), **362**, 505

Colless M., et al., 2001, [MNRAS](#), **328**, 1039

Cooray A., Sheth R., 2002, [Phys. Rep.](#), **372**, 1

Cross D. N., Sánchez C., 2024, [Phys. Rev. D](#), **110**, 123534

Croton D. J., Gao L., White S. D. M., 2007, [MNRAS](#), **374**, 1303

Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, [ApJ](#), **292**, 371

Dawson K. S., et al., 2013, [AJ](#), **145**, 10

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

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

Efstathiou G., Sutherland W. J., Maddox S. J., 1990, [Nature](#), **348**, 705

Farrens S., et al., 2022, [A&A](#), **664**, A141

Feldman H. A., Kaiser N., Peacock J. A., 1994, [ApJ](#), **426**, 23

Flores R., Primack J. R., Blumenthal G. R., Faber S. M., 1993, [ApJ](#), **412**, 443

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, [PASP](#), **125**, 306

García-Moreno S., Chaves-Montero J., 2025, [arXiv e-prints](#), p. [arXiv:2504.06770](#)

Garrison L. H., Eisenstein D. J., Ferrer D., Maksimova N. A., Pinto P. A., 2021, [MNRAS](#), **508**, 575

Gatti M., et al., 2021, [MNRAS](#), **504**, 4312

Gnedin O. Y., Kravtsov A. V., Klypin A. A., Nagai D., 2004, [ApJ](#), **616**, 16

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

Guo H., Zehavi I., Zheng Z., 2012, [ApJ](#), **756**, 127

Guo H., et al., 2015, [MNRAS](#), **446**, 578

Gustafsson M., Fairbairn M., Sommer-Larsen J., 2006, [Phys. Rev. D](#), **74**, 123522

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

Hamilton A., 2000, <https://jila.colorado.edu/~ajsh/FFTLog/index.html>

Hartlap J., 2009, <https://hdl.handle.net/20.500.11811/4088>

Hartlap J., Simon P., Schneider P., 2007, [A&A](#), **464**, 399

Hawkins E., et al., 2003, [MNRAS](#), **346**, 78

Hearin A. P., Zentner A. R., van den Bosch F. C., Campbell D., Tollerud E., 2016, [MNRAS](#), **460**, 2552

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

Heymans C., et al., 2021, [A&A](#), **646**, A140

Hinshaw G., et al., 2003, [ApJS](#), **148**, 135

Howlett C., Lewis A., Hall A., Challinor A., 2012, [J. Cosmology Astropart. Phys.](#), **2012**, 027

Hu J.-P., Wang F.-Y., 2023, [Universe](#), **9**, 94

Huff E., Mandelbaum R., 2017, [arXiv e-prints](#), p. arXiv:1702.02600

Huterer D., 2023, [A&A Rev.](#), **31**, 2

Huterer D., Takada M., Bernstein G., Jain B., 2006, [MNRAS](#), **366**, 101

Kaiser N., 1987, [MNRAS](#), **227**, 1

Kilbinger M., et al., 2017, [MNRAS](#), **472**, 2126

- Lamman C., Tsaprazi E., Shi J., Šarčević N. N., Pyne S., Legnani E., Ferreira T., 2024, [The Open Journal of Astrophysics](#), **7**, 14
- Landy S. D., Szalay A. S., 1993, [ApJ](#), **412**, 64
- Lange J., Huang S., 2022, dsigma: Galaxy-galaxy lensing Python package, Astrophysics Source Code Library, record ascl:2204.006
- Lange J. U., Yang X., Guo H., Luo W., van den Bosch F. C., 2019, [MNRAS](#), **488**, 5771
- Lange J. U., Leauthaud A., Singh S., Guo H., Zhou R., Smith T. L., Cyr-Racine F.-Y., 2021, [MNRAS](#), **502**, 2074
- Leauthaud A., Tinker J., Behroozi P. S., Busha M. T., Wechsler R. H., 2011, [ApJ](#), **738**, 45
- Leauthaud A., et al., 2017, [MNRAS](#), **467**, 3024
- Lewis A., Challinor A., Lasenby A., 2000, [ApJ](#), **538**, 473
- Li Q., et al., 2024, [ApJ](#), **969**, L25
- Limber D. N., 1953, [ApJ](#), **117**, 134
- Maksimova N. A., Garrison L. H., Eisenstein D. J., Hadzhiyska B., Bose S., Satterthwaite T. P., 2021, [MNRAS](#), **508**, 4017
- Mandelbaum R., et al., 2005, [MNRAS](#), **361**, 1287
- Marinoni C., Hudson M. J., 2002, [ApJ](#), **569**, 101
- Mather J. C., et al., 1990, [ApJ](#), **354**, L37
- Medlock I., Nagai D., Anglés-Alcázar D., Gebhardt M., 2025, [ApJ](#), **983**, 46
- Miller L., et al., 2013, [MNRAS](#), **429**, 2858
- Miyatake H., et al., 2022a, [Phys. Rev. D](#), **106**, 083519
- Miyatake H., et al., 2022b, Cosmological inference from an emulator based halo model. II. Joint analysis of galaxy-galaxy weak lensing and galaxy clustering from HSC-Y1 and SDSS ([arXiv:2111.02419](#)), [doi:10.1103/PhysRevD.106.083520](#)
- Mo H., van den Bosch F., White S., 2010, Galaxy Formation and Evolution. Cambridge University Press

- Moraes B., et al., 2014, in *Revista Mexicana de Astronomia y Astrofisica Conference Series*. pp 202–203
- More S., Miyatake H., Mandelbaum R., Takada M., Spergel D. N., Brownstein J. R., Schneider D. P., 2015, *ApJ*, **806**, 2
- Navarro J. F., Frenk C. S., White S. D. M., 1996, *ApJ*, **462**, 563
- Nishimichi T., et al., 2019, *ApJ*, **884**, 29
- Pandey S., et al., 2022, *Phys. Rev. D*, **106**, 043520
- Parejko J. K., et al., 2013, *MNRAS*, **429**, 98
- Perlmutter S., et al., 1999, *ApJ*, **517**, 565
- Planck Collaboration et al., 2014a, *A&A*, **571**, A1
- Planck Collaboration et al., 2014b, *A&A*, **571**, A15
- Planck Collaboration et al., 2014c, *A&A*, **571**, A16
- Planck Collaboration et al., 2020, *A&A*, **641**, A6
- Porredon A., et al., 2022, *Phys. Rev. D*, **106**, 103530
- Prat J., Bacon D., 2026, in *Encyclopedia of Astrophysics*, Volume 5. pp 508–537 ([arXiv:2501.07938](https://arxiv.org/abs/2501.07938)), doi:10.1016/B978-0-443-21439-4.00071-7
- Prat J., et al., 2022, *Phys. Rev. D*, **105**, 083528
- Press W. H., Schechter P., 1974, *ApJ*, **187**, 425
- Reid B. A., Seo H.-J., Leauthaud A., Tinker J. L., White M., 2014, *MNRAS*, **444**, 476
- Reid B., et al., 2016, *MNRAS*, **455**, 1553
- Riess A. G., et al., 1998, *AJ*, **116**, 1009
- Ryden B. S., Gunn J. E., 1987, *ApJ*, **318**, 15
- Saito S., et al., 2016, *MNRAS*, **460**, 1457
- Schneider A., Teyssier R., Stadel J., Chisari N. E., Le Brun A. M. C., Amara A., Refregier A., 2019, *J. Cosmology Astropart. Phys.*, **2019**, 020

- Seljak U., 2000, [MNRAS](#), **318**, 203
- Sersic J. L., 1968, Atlas de Galaxias Australes
- Sheldon E., 2015, NGMIX: Gaussian mixture models for 2D images, Astrophysics Source Code Library, record ascl:1508.008
- Sheldon E. S., et al., 2004, [AJ](#), **127**, 2544
- Shirasaki M., et al., 2024, [Phys. Rev. D](#), **110**, 103006
- Simha V., Cole S., 2013, [MNRAS](#), **436**, 1142
- Singh S., Mandelbaum R., Seljak U., Slosar A., Vazquez Gonzalez J., 2017, [MNRAS](#), **471**, 3827
- Sinha M., Garrison L. H., 2019, in Software Challenges to Exascale Computing. Second Workshop. pp 3–20 ([arXiv:1911.08275](#)), doi:10.1007/978-981-13-7729-7_1
- Sinha M., Garrison L. H., 2020, [MNRAS](#), **491**, 3022
- Smoot G. F., et al., 1992, [ApJ](#), **396**, L1
- Sonnenfeld A., Wang W., Bahcall N., 2019, [A&A](#), **622**, A30
- Spergel D. N., et al., 2003, [ApJS](#), **148**, 175
- Sugiyama S., et al., 2023, [Phys. Rev. D](#), **108**, 123521
- Tegmark M., et al., 2004, [ApJ](#), **606**, 702
- Tinker J. L., Leauthaud A., Bundy K., George M. R., Behroozi P., Massey R., Rhodes J., Wechsler R. H., 2013, [ApJ](#), **778**, 93
- Toshiki K., 2025, The Intrinsic Alignment of Galaxies: The Good, the Bad, and the Ugly, https://wwwmpa.mpa-garching.mpg.de/~komatsu/lecturenotes/Toshiki_Kurita_on_IA.pdf
- Troxel M. A., Ishak M., 2015, [Phys. Rep.](#), **558**, 1
- Unruh S., Schneider P., Hilbert S., Simon P., Martin S., Puertas J. C., 2020, [A&A](#), **638**, A96
- Vale A., Ostriker J. P., 2004, [MNRAS](#), **353**, 189

Wechsler R. H., Tinker J. L., 2018, [ARA&A](#), **56**, 435

White M., et al., 2011, [ApJ](#), **728**, 126

Wright A. H., et al., 2025, [arXiv e-prints](#), p. [arXiv:2503.19441](#)

Xu X., Zehavi I., Contreras S., 2021, [MNRAS](#), **502**, 3242

York D. G., et al., 2000, [AJ](#), **120**, 1579

Yuan S., Hadzhiyska B., Bose S., Eisenstein D. J., Guo H., 2021, [MNRAS](#), **502**, 3582

Yuan S., Garrison L. H., Hadzhiyska B., Bose S., Eisenstein D. J., 2022, [MNRAS](#), **510**, 3301

Zehavi I., et al., 2004, [ApJ](#), **608**, 16

Zeldovich Y. B., Klypin A. A., Khlopov M. Y., Chechetkin V. M., 1980, *Sov. J. Nucl. Phys.*, **31**, 664

Zhang T., et al., 2025, [arXiv e-prints](#), p. [arXiv:2507.01377](#)

Zheng Z., et al., 2005, [ApJ](#), **633**, 791

Zheng Z., Coil A. L., Zehavi I., 2007, [ApJ](#), **667**, 760

Zheng Z., Zehavi I., Eisenstein D. J., Weinberg D. H., Jing Y. P., 2009, [ApJ](#), **707**, 554

Zu Y., 2020, [arXiv e-prints](#), p. [arXiv:2010.01143](#)

Zu Y., Mandelbaum R., 2015, [MNRAS](#), **454**, 1161

Zu Y., Mandelbaum R., 2016, [MNRAS](#), **457**, 4360

sdsspy n.d., sdsspy, <https://code.google.com/archive/p/sdsspy/>

von Wietersheim-Kramsta M., et al., 2021, [MNRAS](#), **504**, 1452

APPENDICES

Appendix A

Newtonian Deflection Angle Derivation

We primarily follow [Dodelson \(2017\)](#) for this derivation. A diagram of the setup is provided in Fig. [A.1](#). We proceed by computing the deflection angle of a particle with mass m as it passes near a point mass M ($M \gg m$), which we refer to as the ‘deflector’. Placing the deflector at the origin of our coordinate system, Newton’s Second Law states:

$$m \frac{d^2 \mathbf{r}}{dt^2} = -\frac{mMG}{r^2} \hat{r}, \quad (\text{A.1})$$

where $\mathbf{r}(t)$ is a general position vector. Notably, the m drops out of our expression, suggesting that Eq. [A.1](#) can be extended to massless particles such as photons. Recasting the equation above in polar coordinates yields:

$$\left(\ddot{R} - R\dot{\phi}^2 \right) \hat{R} + \frac{1}{R} \frac{d}{dt} \left(R^2 \dot{\phi} \right) \hat{\phi} = -\frac{MG}{R^2} \hat{R}, \quad (\text{A.2})$$

where $R(t)$ is the distance between the particle and deflector, and ϕ is the angle between the line connecting the two objects and the horizontal, which passes through the deflector. Since gravity is a radial force, the term with $\hat{\phi}$ must drop out of our expression, leading to the following conservation law:

$$R^2 \dot{\phi} \equiv J_z = \text{constant}. \quad (\text{A.3})$$

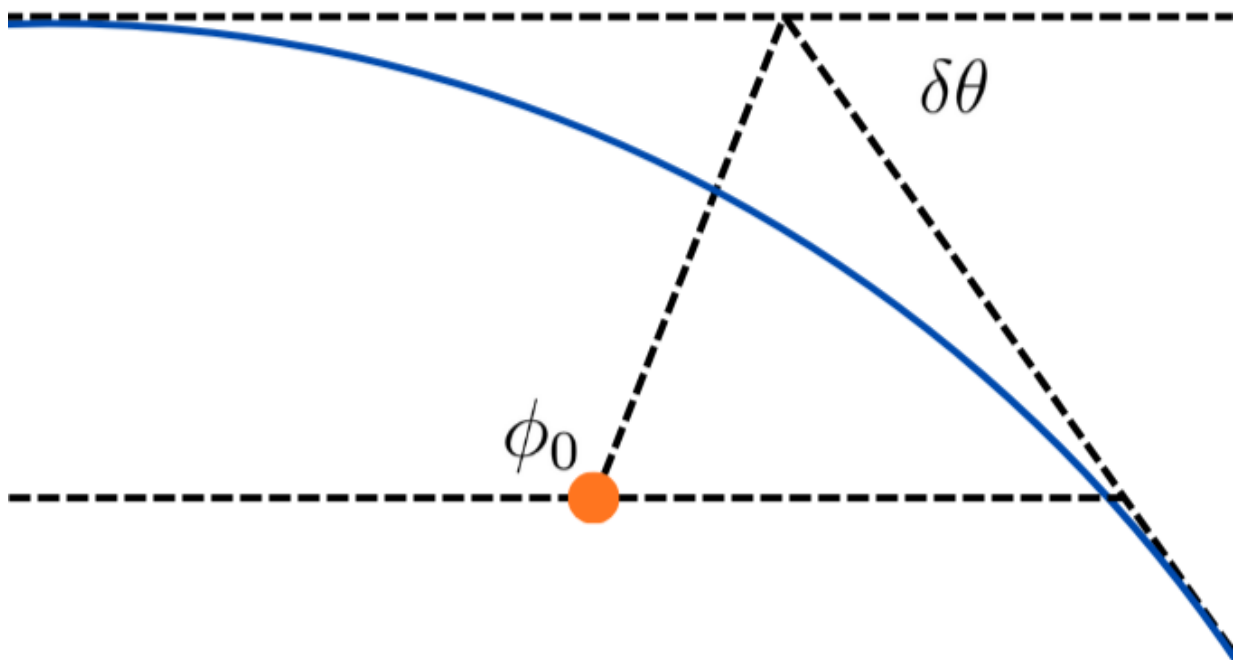


Figure A.1: Deflection angle, $\delta\theta$, of a particle due to a foreground mass (orange point). The blue line represents the particle's trajectory.

Note that J_z is the reduced angular momentum of the particle. Thus, our equation of motion becomes:

$$\ddot{R} - \frac{J_z^2}{R^3} = -\frac{MG}{R^2}. \quad (\text{A.4})$$

We now perform a change of variables from time to ϕ . Accordingly, derivatives with respect to ϕ are denoted with primes. To proceed, we use the following identity:

$$\dot{R} = \frac{dR}{dt} = \frac{dR}{d\theta} \frac{d\theta}{dt} = R' \dot{\phi}. \quad (\text{A.5})$$

Thus, the equation of motion can be written as:

$$\frac{R''}{R^2} - 2\frac{R'^2}{R^3} - \frac{1}{R} = -\frac{MG}{J_z^2}. \quad (\text{A.6})$$

As a final change of variables, we define $u \equiv 1/R$. Under this substitution, our differential equation becomes:

$$u'' + u = \frac{MG}{J_z^2}. \quad (\text{A.7})$$

The solution to the differential equation above is given by:

$$\frac{1}{R(\phi)} = A \cos(\phi - \phi_0) + \frac{MG}{J_z^2}, \quad (\text{A.8})$$

where A and ϕ_0 are determined by initial and boundary conditions.

We currently have 3 unknowns: A , ϕ_0 , and J_z . To write an expression for J_z , we introduce the impact parameter, b : the distance of closest approach between the particle and deflector in the absence of deflection. It can be shown that $J_z = bc$. To constrain the remaining parameters, we use the fact that $R \rightarrow \infty$ as $\phi \rightarrow 0$, and take the temporal derivative of [A.8](#). To a good approximation, $A = 1/b$, yielding:

$$\cos(\phi_0) = -\frac{MG}{c^2 b}. \quad (\text{A.9})$$

Since the deflection angle is expected to be small, ϕ_0 is not much larger than $\pi/2$. Thus, we can write $\phi_0 \equiv \pi/2 + \epsilon$, where $\epsilon \ll 1$. Using the approximation $\cos(\phi_0) \approx -\epsilon$, we find:

$$\phi_0 = \frac{\pi}{2} + \frac{MG}{c^2 b}. \quad (\text{A.10})$$

The parameter ϕ_0 is not the deflection angle, as can be seen in Fig. [A.1](#). Based on the geometry of the setup, the Newtonian deflection angle, $\delta\theta_{\text{Newt}}$, is given by:

$$\delta\theta_{\text{Newt}} = \frac{2MG}{c^2 b}, \quad (\text{A.11})$$

whereas the deflection angle predicted by general relativity differs by a factor of 2:

$$\delta\theta = \frac{4MG}{c^2 b}. \quad (\text{A.12})$$

The equation above paints a simple picture: particles experience greater deflection when their trajectory is closer to the deflector and the deflector is heavier.

Appendix B

Cosmological Parameter Dependence of Measurements

We measured our observables assuming a fiducial cosmology to convert redshift into distance. However, we vary $\Omega_{\text{m},0}$ in our modelling, so we must ensure that predictions and measurements correspond to the ‘same’ comoving separations. We must also consider how the amplitudes of our observables change under different cosmologies. To account for these effects, we follow the procedure in [More et al. \(2015\)](#). This procedure modifies a prediction that adopts a non-fiducial cosmology.

First, we rescale r_{p} in the following manner:

$$\tilde{r}_{\text{p}} = r_{\text{p}, \text{fid}} \left[\frac{\chi_{\text{new}}(\bar{z})}{\chi_{\text{fid}}(\bar{z})} \right], \quad (\text{B.1})$$

where $\bar{z} = 0.54$ is the median lens galaxy redshift, $r_{\text{p}, \text{fid}}$ is the r_{p} vector at which measurements were made, ‘new’ denotes the new cosmological model, and ‘fid’ denotes the fiducial cosmological model (used to make measurements). Now, we can evaluate our observables at $\tilde{r}_{\text{p}}$. However, before comparing predictions and measurements (e.g., during a likelihood analysis), we must adjust the amplitudes of our predictions:

$$\Delta\tilde{\Sigma}(r_{\text{p}}) = \Delta\Sigma(\tilde{r}_{\text{p}}) \left[\frac{\Sigma_{\text{crit}, \text{fid}}(\bar{z}, z_{\text{s}})}{\Sigma_{\text{crit}, \text{new}}(\bar{z}, z_{\text{s}})} \right], \quad (\text{B.2})$$

$$\tilde{w}_{\text{p}}(r_{\text{p}}) = w_{\text{p}}(\tilde{r}_{\text{p}}) \left[\frac{E_{\text{new}}(\bar{z})}{E_{\text{fid}}(\bar{z})} \right], \quad (\text{B.3})$$

where $\Delta\Sigma(\tilde{r}_p)$ and $w_p(\tilde{r}_p)$ represent DARK EMULATOR's prediction at $\tilde{r}_p$. The choice of z_s is somewhat arbitrary, as stated in [More et al. \(2015\)](#). Note that the fraction in Eq. B.2 differs from the expression in [More et al. \(2015\)](#), where the inverse is presented.

In addition to correcting our predictions of $\Delta\Sigma$ and w_p , we correct our prediction of $\bar{n}_g$, as it is subject to a prior:

$$\tilde{\bar{n}}_g = \bar{n}_{g,\text{new}} \frac{\chi_{\text{new}}^3(z_2) - \chi_{\text{new}}^3(z_1)}{\chi_{\text{fid}}^3(z_2) - \chi_{\text{fid}}^3(z_1)}, \quad (\text{B.4})$$

where $\bar{n}_{g,\text{new}}$ represents a prediction of $\bar{n}_g$ for redshift bin $[z_2, z_1] = [0.7, 0.43]$. DARK EMULATOR can only evaluate $\bar{n}_g$ (and all other quantities) at a fixed redshift. Therefore, we compute all quantities at $\bar{z}$. $\bar{n}_g$ remains uncorrected in the calculations of $\Delta\Sigma$ and w_p , as corrections are applied to the predicted signals at the end.

Appendix C

$\Delta\Sigma$ and w_p Cross-Covariance Test

We assume that $\Delta\Sigma$ and w_p are uncorrelated during parameter estimation. This assumption is not always valid, and the degree of correlation between the two signals depends on the extent of angular overlap between the lens and source galaxy catalogues. Here, we present our test for evaluating the assumption of negligible cross-covariance between $\Delta\Sigma$ and w_p .

We first constructed a block covariance matrix with $\Delta\Sigma$ and w_p (uncorrected) jackknife covariance matrices along the diagonal. Then, we estimated the cross-covariance blocks using jackknife resampling. We computed 150 w_p samples in the survey overlap region using the same jackknife patches from our $\Delta\Sigma$ statistical covariance estimation, from which we calculated $C_{JK}(\Delta\Sigma, w_{p,\text{overlap}})$ using Eq. 3.10. We then approximated the cross-covariance between $\Delta\Sigma$ (measured in the survey overlap region) and w_p measured across the entire CMASS footprint by rescaling our cross-covariance matrix that assumes 100% survey overlap:

$$C_{JK}(\Delta\Sigma, w_p) \approx C_{JK}(\Delta\Sigma, w_{p,\text{overlap}}) \times \frac{A_{\text{overlap}}}{A_{\text{tot}}}, \quad (\text{C.1})$$

where A denotes an angular area, and A_{tot} is the total CMASS footprint area. This expression assumes that the covariance between $\Delta\Sigma$ and w_p is zero in non-overlapping regions of the sky. We ensured that the final block covariance matrix is positive semi-definite. A visualization of the block covariance matrix is shown in Fig. C.1.

Next, we generated fake data vectors from a multivariate normal distribution using the final block covariance matrix. We then computed two sets of χ^2 values using Eq. 5.3. The

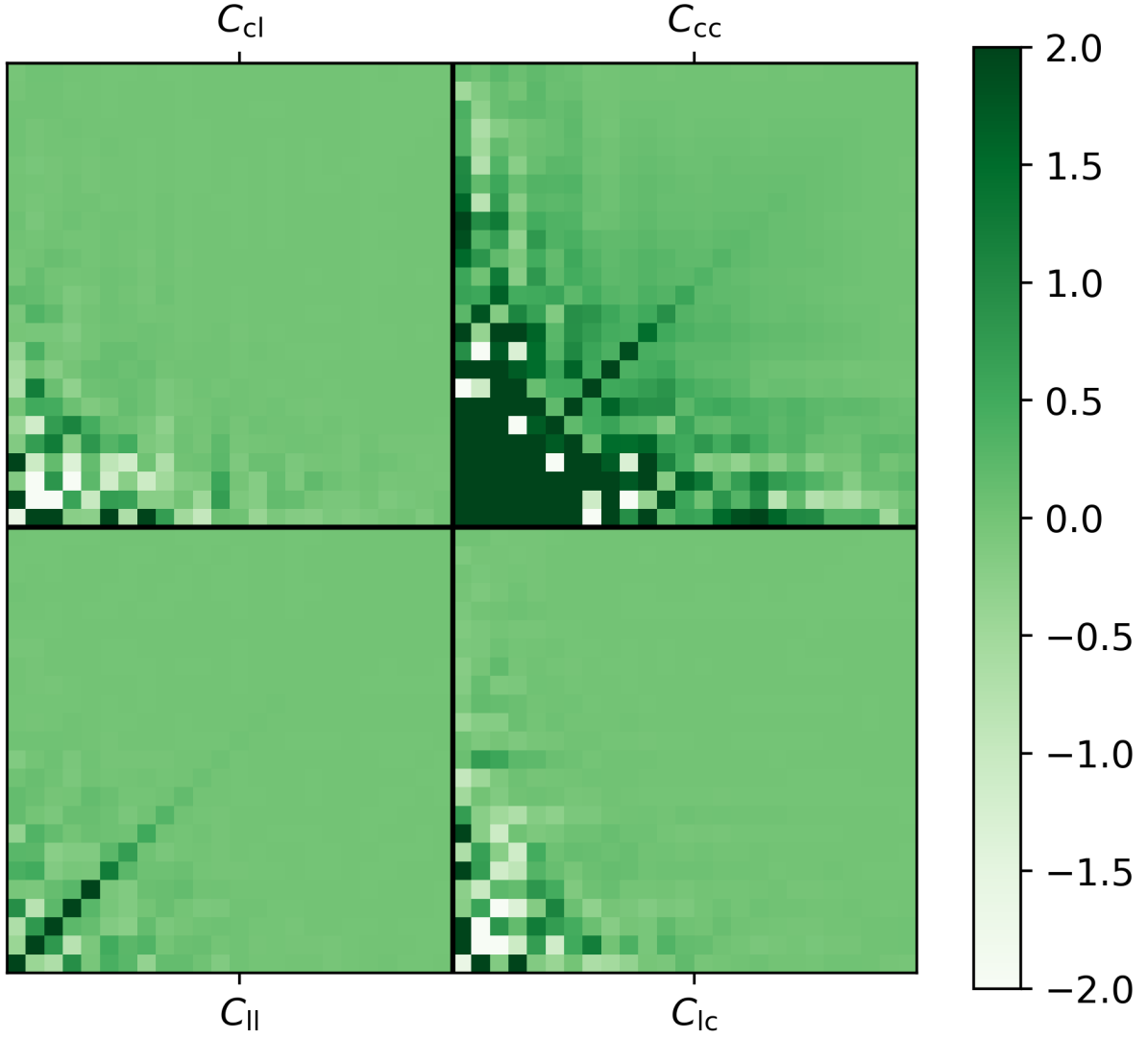


Figure C.1: Raw joint covariance matrix for $\Delta\Sigma$ and w_p . ‘C’ stands for clustering, and ‘l’ for lensing. Each block is an individual covariance matrix. Note that the colour scale is set to a range of -2 to 2 .

first set used the final block covariance matrix for $\mathbf{C}^{-1}$, while the second set used the same matrix but with the cross-covariance blocks set to zero. The resulting χ^2 distributions are shown in Fig. C.2.

The χ^2 distributions in Fig. C.2 are quite similar, and the $\Delta\chi^2$ distribution in Fig. C.3 is narrowly peaked (compared to the individual χ^2 distributions) and centred around 0. We thus concluded that ignoring the cross-covariance between our measurements of $\Delta\Sigma$ and w_p does not significantly bias χ^2 values. Note that we did not use the joint covariance matrix in our likelihood analysis, as the jackknife patches used to estimate the w_p block differ from those used for the other blocks. Properly de-biasing our joint covariance matrix would be a non-trivial task.

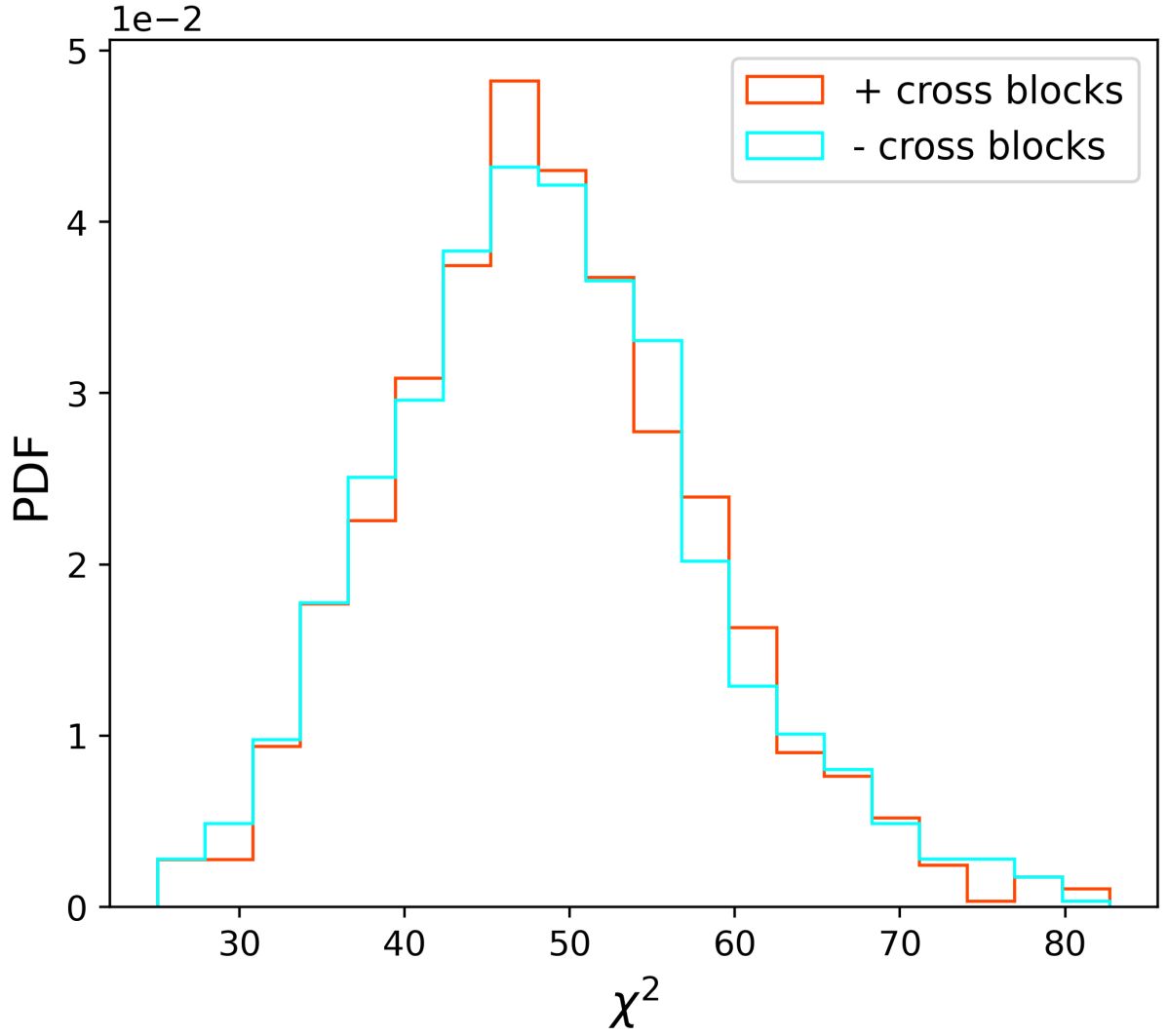


Figure C.2: χ^2 distributions from the joint covariance matrix in Fig. C.1 and the same matrix but with the cross-covariance blocks (C_{cl} and C_{lc}) set to zero.

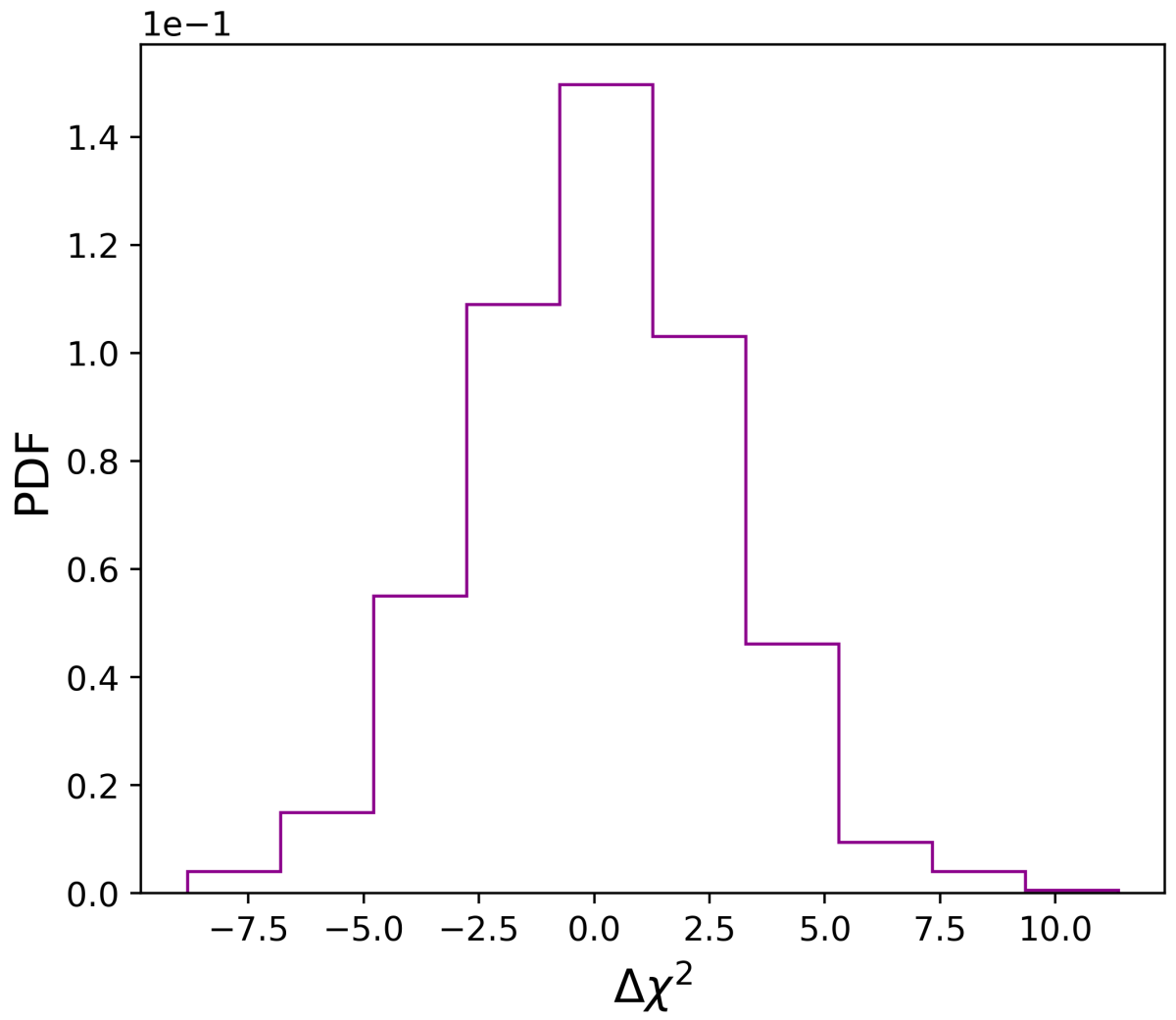


Figure C.3: $\Delta\chi^2$ distribution from the differences between the datasets in Fig. C.2.

Appendix D

Additional Plots

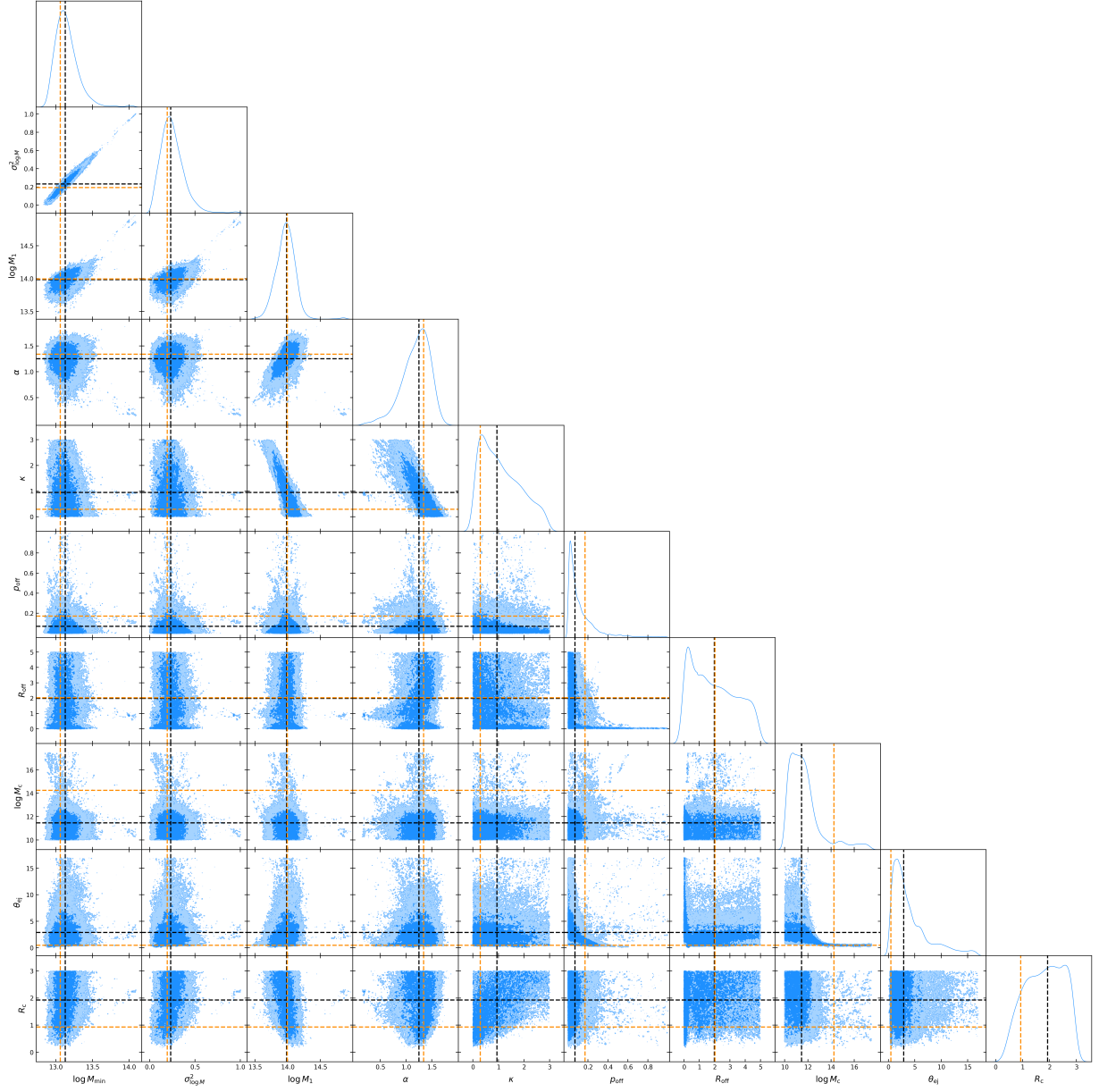


Figure D.1: Posterior distributions for the HOD+cosmo+off+fbk+ R_c model.

Parameter	Fiducial Value
$\log M_{\min}$	13.5
$\sigma_{\log M}^2$	0.5
$\log M_1$	14.0
α	1.0
κ	1.0
R_c	1.0
p_{off}	0.2
R_{off}	1.0
$\ln(10^{10} A_s)$	3.0
$\Omega_{m,0}$	0.3

Table D.1: Fiducial parameter values used to generate the ‘rainbow plots’ in this Appendix. By default, feedback is not included.

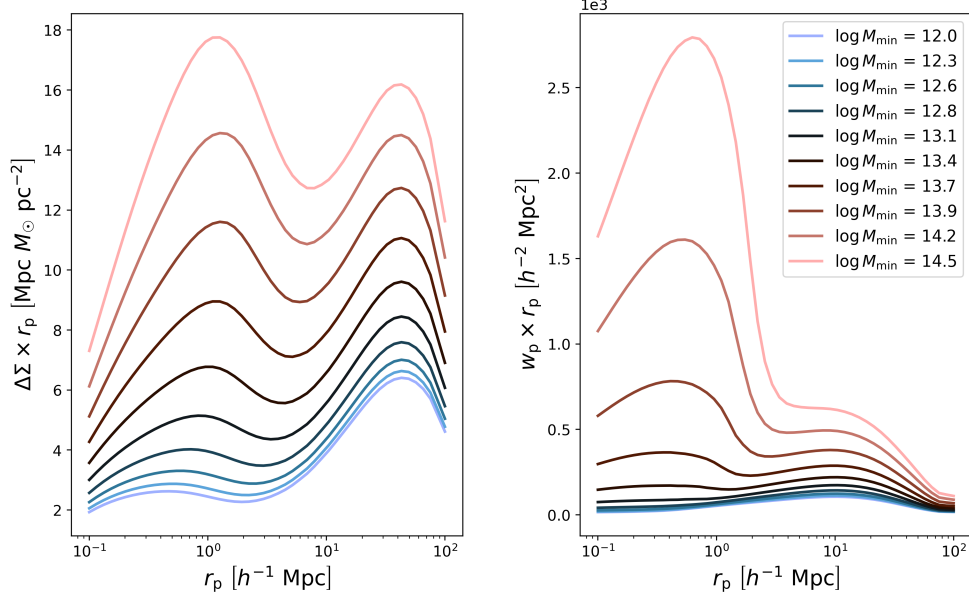


Figure D.2: Effects of varying the parameter $\log M_{\min}$ on $\Delta\Sigma$ and w_p .

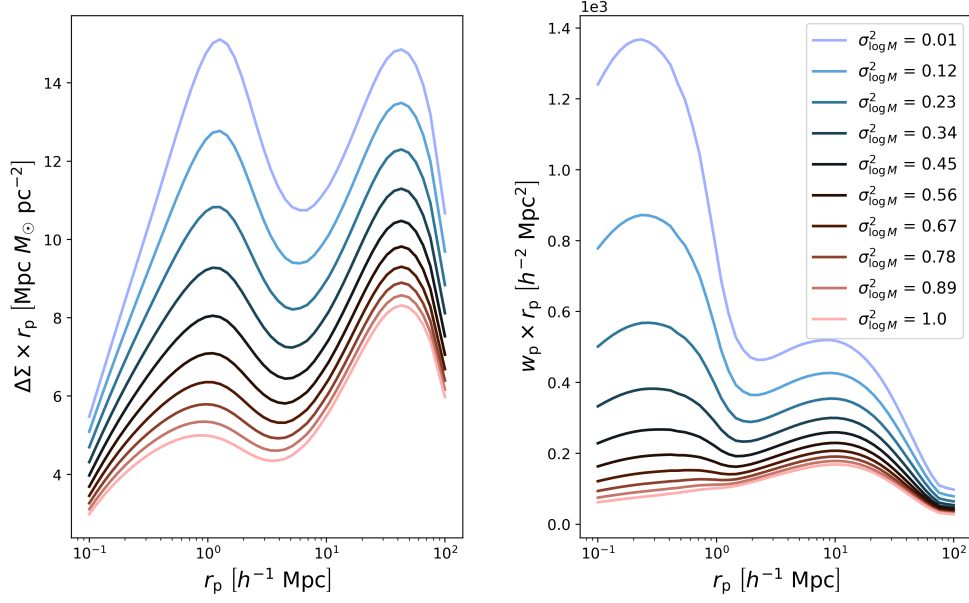


Figure D.3: Effects of varying the parameter $\sigma_{\log M}^2$ on $\Delta\Sigma$ and w_p .

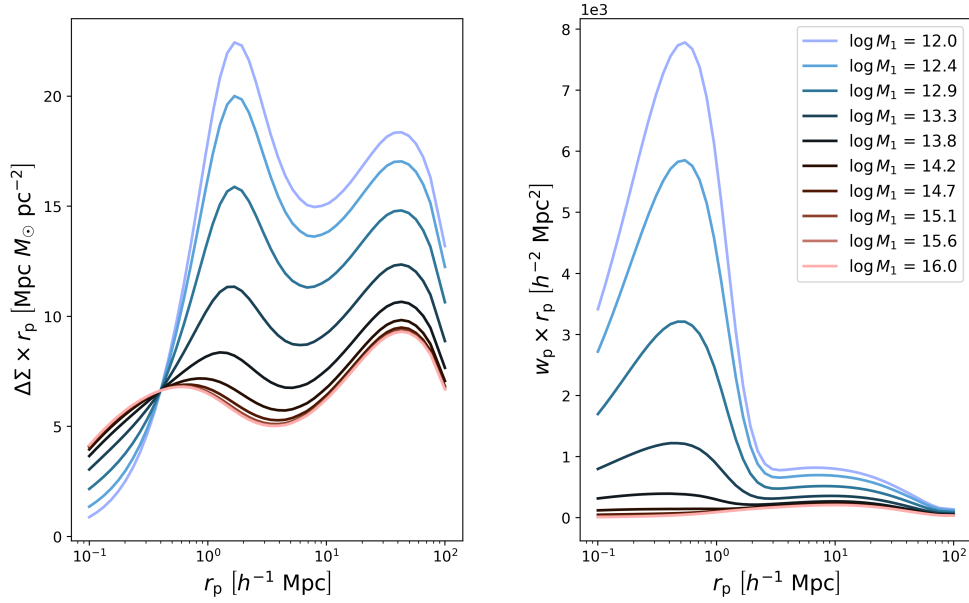


Figure D.4: Effects of varying the parameter $\log M_1$ on $\Delta\Sigma$ and w_p .

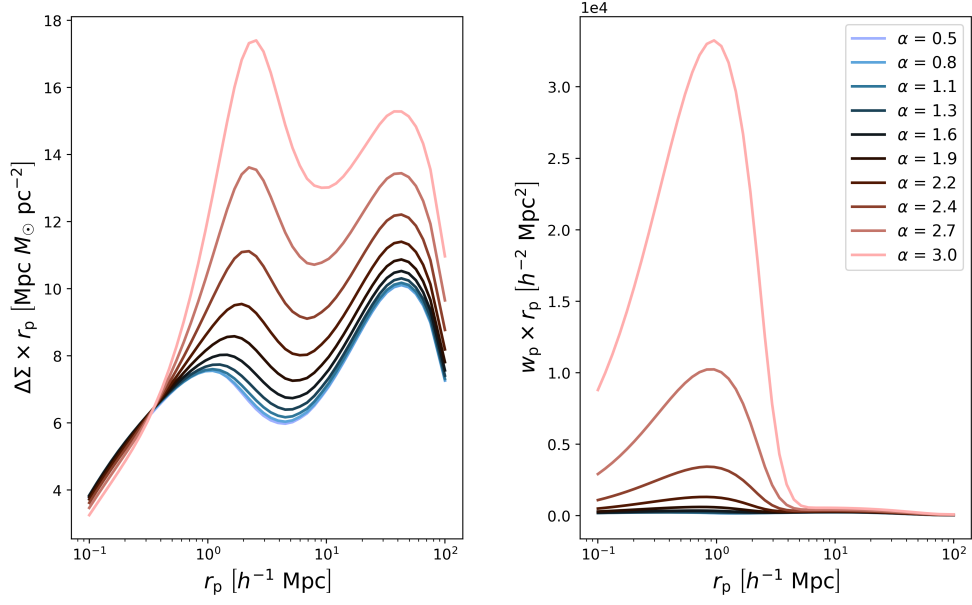


Figure D.5: Effects of varying the parameter α on $\Delta\Sigma$ and w_p .

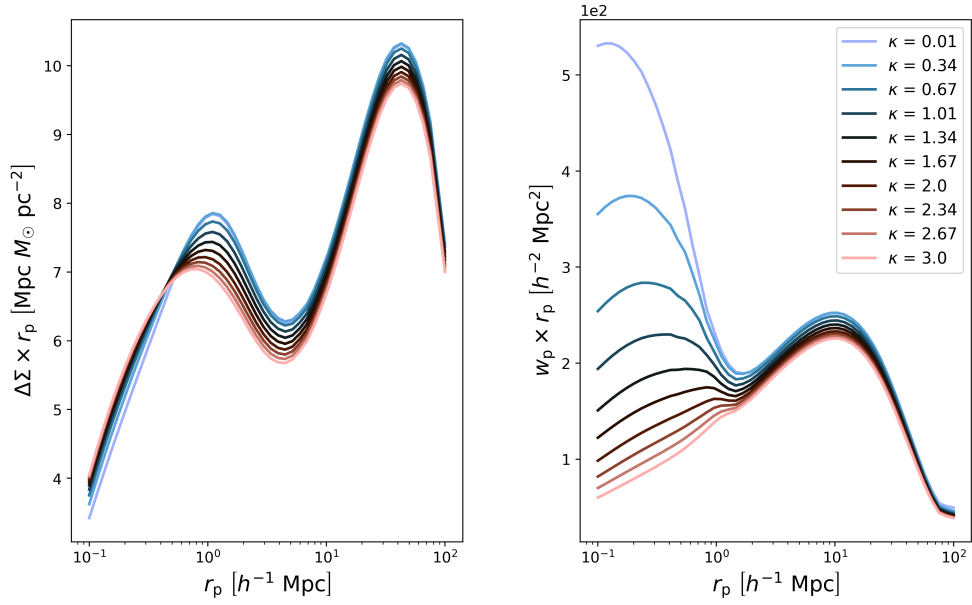


Figure D.6: Effects of varying the parameter κ on $\Delta\Sigma$ and w_p .

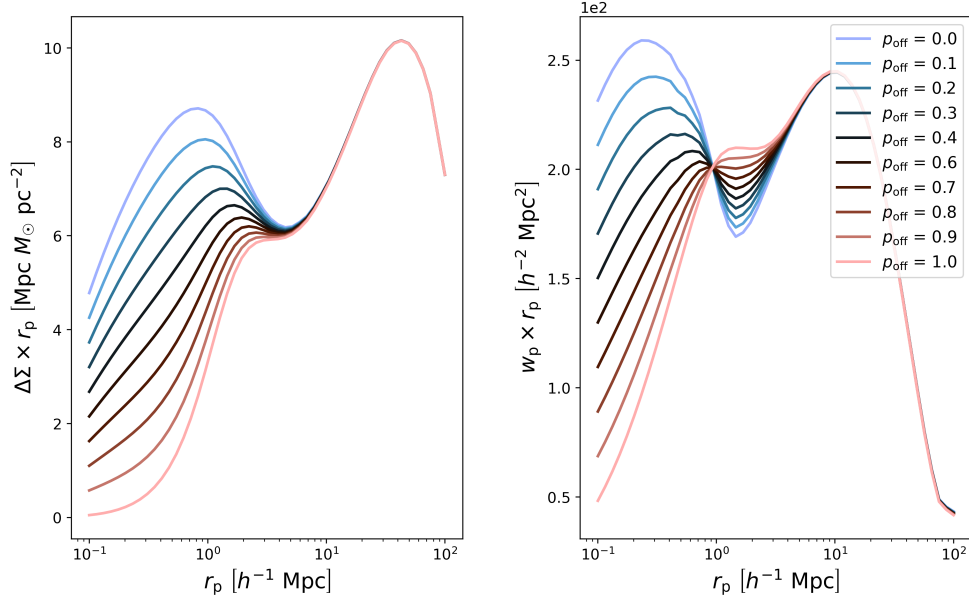


Figure D.7: Effects of varying the parameter p_{off} on $\Delta\Sigma$ and w_p .

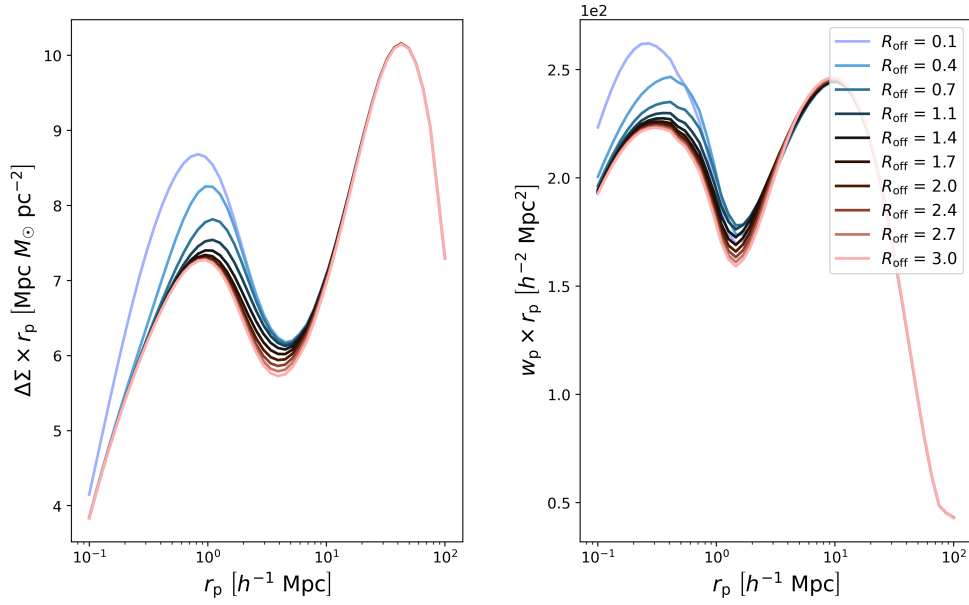


Figure D.8: Effects of varying the parameter R_{off} on $\Delta\Sigma$ and w_p .

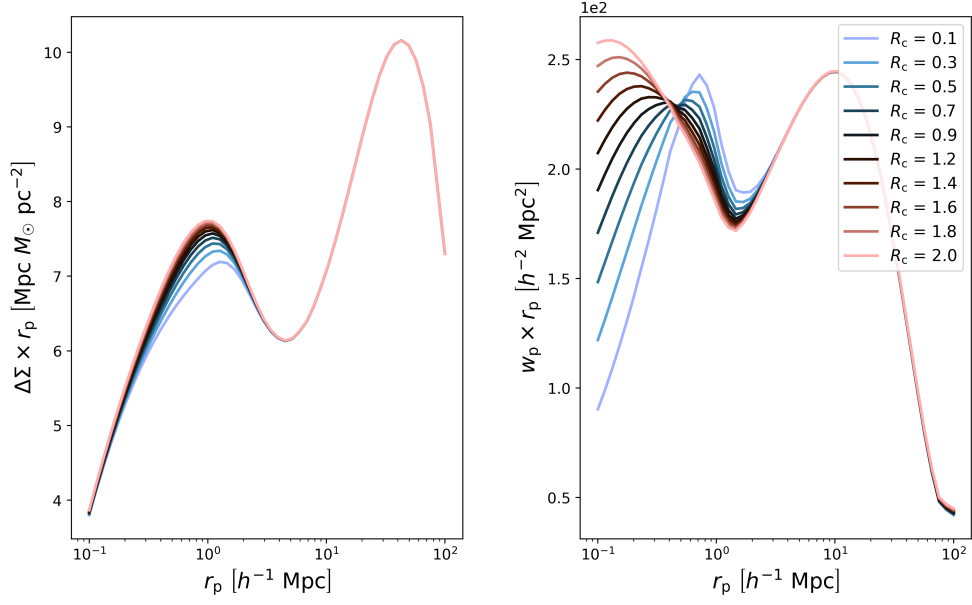


Figure D.9: Effects of varying the parameter R_c on $\Delta\Sigma$ and w_p .

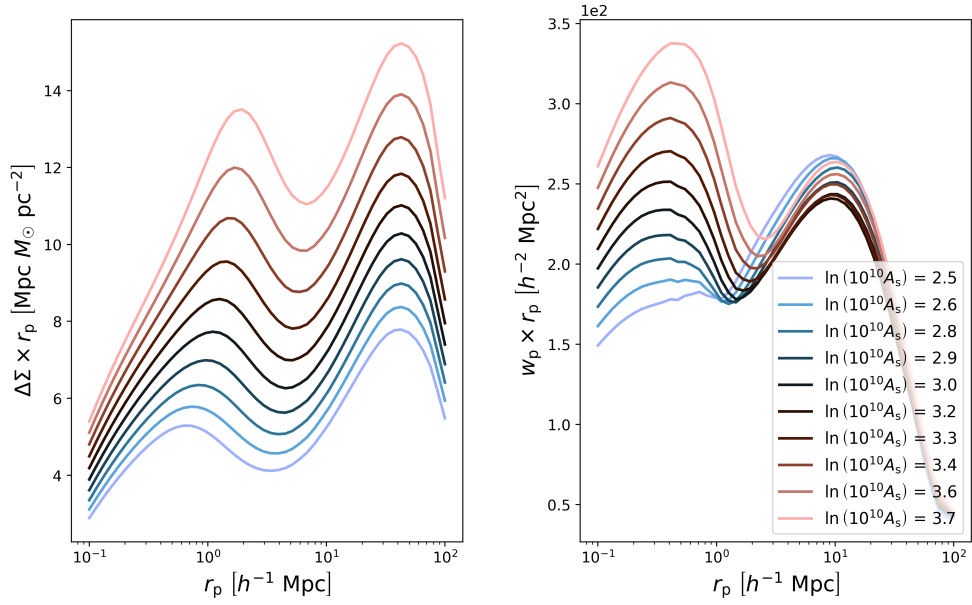


Figure D.10: Effects of varying the parameter $\ln(10^{10}A_s)$ on $\Delta\Sigma$ and w_p .

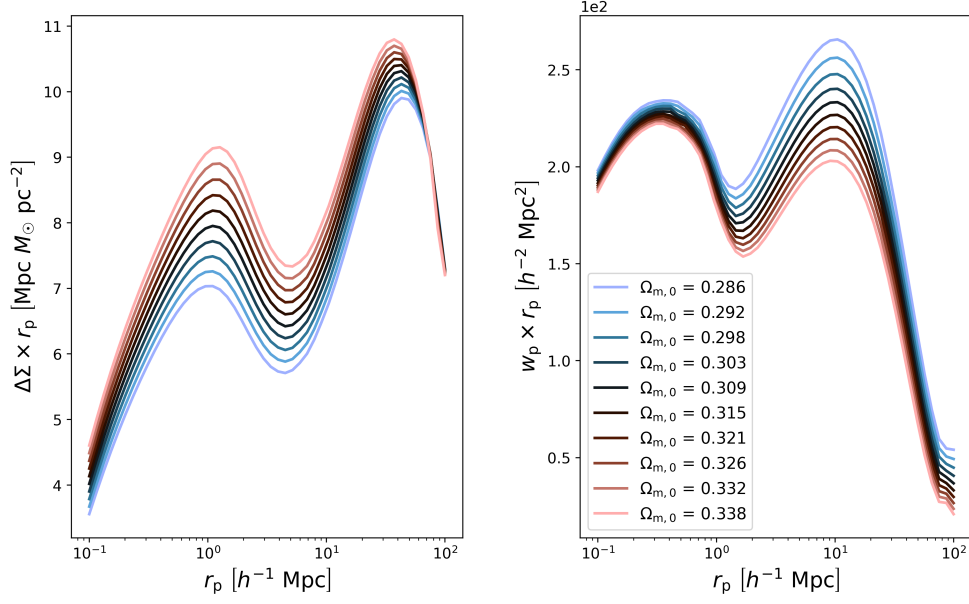


Figure D.11: Effects of varying the parameter $\Omega_{m,0}$ on $\Delta\Sigma$ and w_p .

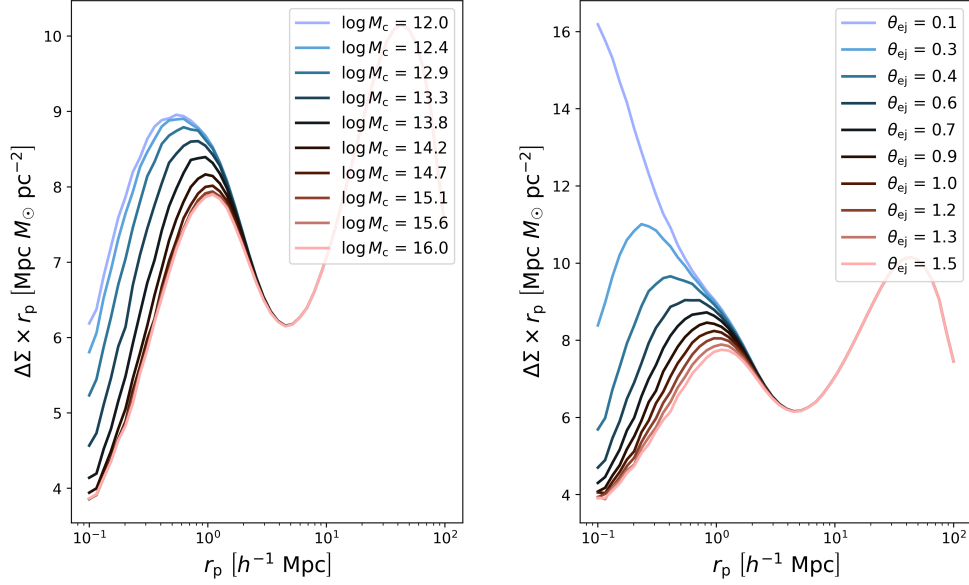


Figure D.12: Effects of varying the parameters $\log M_c$ and θ_{ej} on $\Delta\Sigma$. The fiducial value for $\log M_c$ is 14, while 1 is the fiducial value for θ_{ej} .