

Halo modelling of galaxy clustering

Henrik Brink

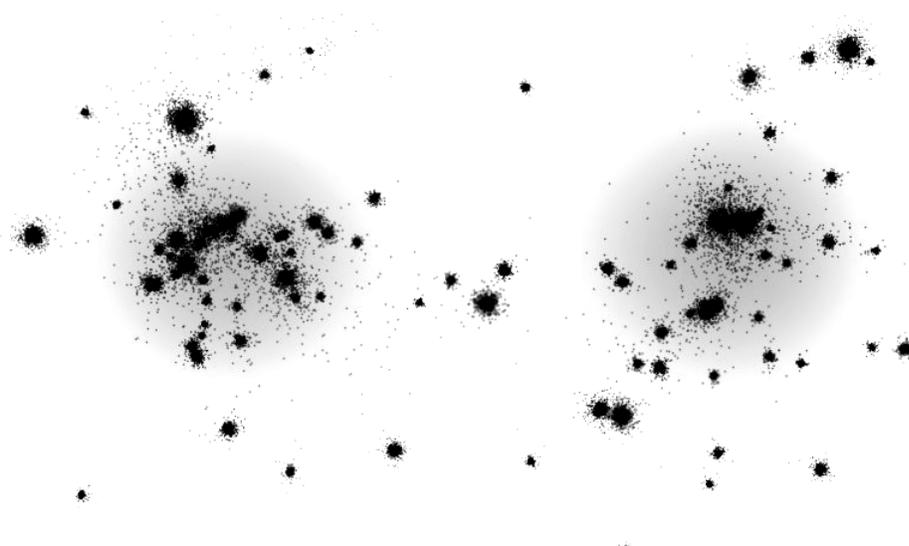
Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen

Master's thesis

Academic advisor: Berian James

Co-advisor: Steen H. Hansen

June 1, 2011



Contents

1	The evolution of structure	1
1.1	Smooth background	1
1.2	Linear evolution	2
1.2.1	Dark Matter power spectrum	2
1.2.2	Scale-dependent evolution of modes	3
1.2.3	Growth factor	6
1.2.4	Two-point correlation function	7
1.3	Non-linear evolution	8
1.3.1	Dark Matter haloes	8
1.3.2	Galaxy formation and evolution	10
1.3.3	Numerical simulations	13
2	Correlation measurements	15
2.1	COSMOS	16
2.1.1	The COSMOS field	16
2.1.2	Galaxies	16
2.1.3	Clusters	19
2.2	Cluster – galaxy cross-correlation	21
2.2.1	Estimator	21
2.2.2	Redshift effects	23
2.2.3	Computational challenges	24
2.3	Error estimation	24
2.3.1	Jackknife resampling	24
2.3.2	Moving block bootstrap	25
2.3.3	Covariance and error bars	26
2.4	Measurements	27
2.4.1	Full galaxy sample	27
2.4.2	Varying galaxy properties	30
3	Halo modelling	36
3.1	The Halo Model	36
3.1.1	Introduction	36
3.1.2	Cluster – galaxy cross-correlation	37
3.1.3	Halo Occupation Distribution	38

3.1.4	The Dark Matter profile	40
3.2	Modelling the measurements	40
3.2.1	Model parameters	41
3.2.2	Cluster ensemble	41
3.2.3	Projection	42
3.2.4	Likelihood estimation	42
3.3	Results	43
3.3.1	Full galaxy sample	44
3.3.2	Varying galaxy properties	46
3.3.3	Alternative dark matter profile	49
4	Discussion	52
4.1	Results	52
4.1.1	Halo occupation distribution	52
4.1.2	Dark matter profile	53
4.1.3	Galaxy profile	54
4.2	Future developments	55
4.2.1	Improvements	55
4.2.2	Opportunities	55
4.3	Conclusions	56
	Acknowledgements	57
A	The Standard Model of Cosmology	58
A.1	Observational basis	58
A.1.1	Hubble’s law	58
A.1.2	Cosmic Microwave Background	58
A.2	Friedmann-Robertson-Walker spacetime	60
A.3	Cosmic inventory	62
B	N-body simulations	65
B.1	N-body dynamics	65
B.2	<i>Gadget</i>	66
B.3	Analysis	67
C	PyHalo software library	68
C.1	Motivation	68
C.2	Cosmology	69
C.3	Halo	70
C.4	Modelling	70
C.5	Optimizations	71
C.5.1	Fortranization	71
C.5.2	Parallelization	72
	Bibliography	73

Abstract

In this master's thesis we present the introduction, measurement and modelling of the two-point cross-correlation between X-ray clusters and galaxies in the COSMOS field.

We start by outlining the relevant theory of structure formation from the smooth cosmological background to the collapse of dark matter and gaseous haloes from fluctuations in the density field.

Using the cluster and galaxy catalogues from COSMOS, we measure the projected correlation function while carefully taking into account the astronomical selection effects of the survey. To measure the evolution of galaxy properties with environment, we divide the galaxy catalogue into subsamples of stellar mass, color and morphology, and find good agreement with previous observations of galaxy evolution in the dense environment of clusters.

We present the halo model as a statistical framework for modelling these measurements, and introduce a new parameter, the halo tilt, that challenges the usual assumption that cluster galaxies follow the density profile of their dark matter haloes. We show that there is evidence for such a parameter across galaxy populations at a nearly constant value. We discuss these results and propose the alternative picture that the dark matter haloes are not well described by the Navarro, Frenk and White profile that is traditionally employed. We create a model for the Einasto profile and show that it outperforms the previous models, suggesting that future analysis should consider this alternative.

Chapter 1

The evolution of structure

We will use the first chapter of this project to introduce the theory of structure formation from the smooth cosmological background, through the evolution of perturbations in the early Universe to the formation of galaxies as we observe them today. This is intended to form the basis for the subsequent chapters, where we will analyse the measurements of the galaxy distribution with respect to galaxy clusters.

1.1 Smooth background

The standard model of cosmology introduces four primary sources of energy in the Universe: radiation, baryonic matter, dark matter and dark energy. These components can be treated as cosmological fluids in a smooth, expanding Universe. In the next chapter we will introduce perturbations and account for interactions between components, but it is useful to consider this idealized case first. This section is a summary of the important points from the more detailed treatment in appendix A.

In an expanding Universe, the metric of spacetime takes the form:

$$g = (-1, a(t)^2), \quad (1.1)$$

where $a(t)$ is the *scale factor* that determines the scaling of the spatial component of the metric in time, i.e. the expansion of the Universe. The evolution of the density ρ_i of component i is assumed to follow a power law in this scale factor with the slope n_i :

$$\rho_i = \rho_{i,0} a(t)^{-n_i}, \quad (1.2)$$

where $\rho_{i,0}$ is then the density at $a(t) = 1$, conventionally taken to be the present day. It is useful to define the dimensionless density parameter Ω_i that describes the density of a component today with respect to the critical density ρ_c :

$$\Omega_i = \frac{\rho_{i,0}}{\rho_c} = \frac{3H_0^2}{8\pi G}\rho_{i,0}, \quad (1.3)$$

where G is the gravitational constant and H_0 is the Hubble constant, the value of the Hubble parameter at $a(t) = 1$. Table 1.1 shows values of n_i and Ω_i for the standard model components¹.

Type	Ω_i	n_i
Baryons	0.046	2
Dark Matter	0.23	3
Radiation	8.24×10^{-5}	4
Dark Energy	0.73	0

We can now write up the evolution of the scale factor, which completely determines how the Universe evolves, using the Friedmann equation. In the notation introduced, this equation can be expressed as a sum of contributions from each of the components of the Universe:

$$\frac{\dot{a}}{a} = H_0^2[\Omega_m a^{-3} + \Omega_r a^{-4} + \Omega_\Lambda], \quad (1.4)$$

where Ω_m includes both dark and baryonic matter and we have assumed a spatially flat Universe². This equation, with the values given in table 1.1, is used throughout the project simply as *the Cosmology*.

1.2 Linear evolution

The Universe as described above is quite boring, as everything is perfectly smooth. This description is, however, a useful basis for describing reality because the fluctuations in the density field in the early phases of a nearly homogeneous Universe will be small by definition, and can thus be introduced into the system using *perturbation theory*.

1.2.1 Dark Matter power spectrum

The large-scale structure of the Universe is described by the evolution of the dark matter density field seeded in the early epochs of the Universe. It is useful to describe this field in the form of the power spectrum of dark matter, which records information of the strength of fluctuations on different scales. The power spectrum at any redshift can be written as:

$$P(k, z) = P_p(k) \times T^2(k) \times g^2(z) \quad (1.5)$$

¹The n_i values are here given by the theory, but they can also be found and verified by measurements of e.g. the CMB (see appendix A)

²Note that unless otherwise stated, we are working in units of the speed of light $c = 1$

where P_p is the *primordial* power spectrum, $T(k)$ is the *transfer function* that describes the evolution of modes according to the interactions between the different components of the Universe and $g(z)$ is the linear *growth factor* describing the change in the power spectrum independent of scale.

We first consider the primordial power spectrum. Harrison, Zel'dovich and Peebles (HZP) argued that the power spectrum of overdensities should simply be a power law in the mode (inverse scale) variable k :

$$P_p(k) \propto k^n. \quad (1.6)$$

where n is close to 1 in order to explain the large-scale structure observed today.

A theory for the physics of the very early times of our Universe is needed to explain this relation in detail, and the currently favoured one is *inflation*. Inflation introduces a period of rapid expansion of the Universe in the first moments of time where quantum fluctuations were blown up to the proportions needed to seed the overdensities that developed into the galaxy distribution we observe today. Inflation predicts a primordial power spectrum of the HZP form:

$$P_p(k) = 2\pi^2 \delta_H^2 \frac{k^n}{H_0^{n+3}}, \quad (1.7)$$

where δ_h is the primordial amplitude of fluctuations. We update eqn. 1.5 with this expression:

$$P(k, z) = 2\pi^2 \delta_H^2 \frac{k^n}{H_0^{n+3}} \times T^2(k) \times g^2(z). \quad (1.8)$$

$T(k)$ and $g(z)$ are the subjects of the coming sections.

1.2.2 Scale-dependent evolution of modes

The fluctuations introduced by inflation (or any other process with similar output) are carried to the present day by the interactions between the FRW metric (eqn. A.4) and all of the energy components of the Universe (table 1.1). Figure 1.1 illustrates these dependencies.

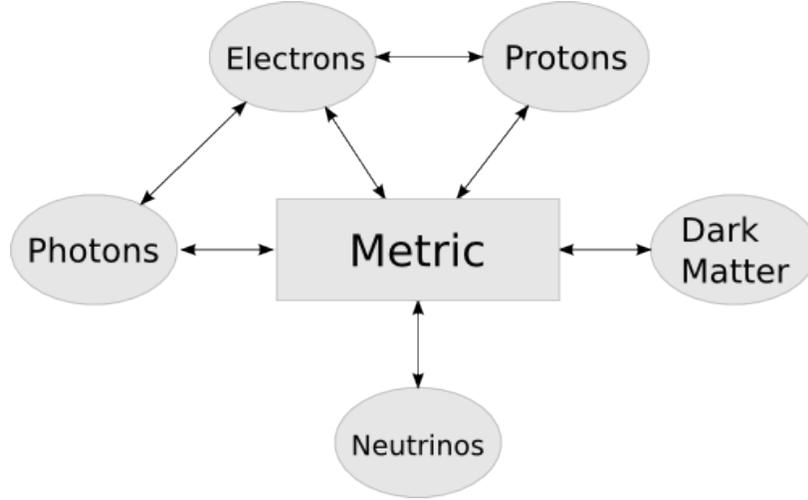


Figure 1.1: A diagram of interactions between components of the Universe. This shows that in order to follow the evolution of one species, we must follow the evolution of all other species. Copied from fig. 4.1 Dodelson [13].

The evolution of matter in the Universe is described by the codependent evolution of particles in phase space. This is formalized by the time-derivative of the phase space distribution function $f(\mathbf{x}, \mathbf{p}, t)$:

$$\frac{df(\mathbf{x}, \mathbf{p}, t)}{dt} = C(f), \quad (1.9)$$

where the right hand side $C(f)$ includes all collision terms. This is called the *Boltzmann equation* after its inventor *Ludwig Boltzmann*.

This project deals with the distribution of matter in the Universe. As the standard model of cosmology (section A) puts most of the matter in the form of collisionless cold dark matter, we will primarily consider the Boltzmann equation in this case. This means that we can neglect any collision terms and write down the Boltzmann equation using partial derivatives in variables of time t , position x^i , unit momentum \hat{p}^i and energy E :

$$\frac{df_{\text{dm}}}{dt} = \frac{\partial f_{\text{dm}}}{\partial t} + \frac{\partial f_{\text{dm}}}{\partial x^i} \frac{dx^i}{dt} + \frac{\partial f_{\text{dm}}}{\partial E} \frac{dE}{dt} + \frac{\partial f_{\text{dm}}}{\partial \hat{p}^i} \frac{d\hat{p}^i}{dt} = 0, \quad (1.10)$$

where $E = \sqrt{p^2 + m^2}$. The route to a full description for the evolution of the dark matter distribution is now a matter of introducing perturbations around the smooth Universe that we have worked with so far. The perturbations to the metric (Ψ) and the spacetime curvature (Φ) is defined such that the smooth metric $g_{\mu\nu} = (-1, a^2)$ (eqn. A.4) becomes:

$$\begin{aligned}
g_{00}(\mathbf{x}, t) &= -1 - 2\Psi(\mathbf{x}, t) \\
g_{ij}(\mathbf{x}, t) &= a^2\delta_{ij}[1 + 2\Phi(\mathbf{x}, t)].
\end{aligned}
\tag{1.11}$$

This results in an expression for the dark matter Boltzmann equation:

$$\frac{\partial f_{\text{dm}}}{\partial t} + \frac{\hat{p}^i p^i}{aE} \frac{\partial f_{\text{dm}}}{\partial x^i} - p \frac{\partial f_{\text{dm}}}{\partial E} \left[H \frac{p^2}{E} + \frac{p^2}{E} \frac{\partial \Phi}{\partial t} + \frac{\hat{p}^i p}{a} \frac{\partial \Psi}{\partial x^i} \right] = 0.
\tag{1.12}$$

We can simplify this equation by defining the density and velocity distributions:

$$n_{\text{dm}} = \int \frac{d^3 p}{(2\pi^3)} f_{\text{dm}}
\tag{1.13}$$

$$v^i = \frac{1}{n_{\text{dm}}} \int \frac{d^3 p}{(2\pi^3)} f_{\text{dm}} \frac{p \hat{p}^i}{E},
\tag{1.14}$$

in units of $\hbar = 1$. Taking moments of 1.12 and moving to Fourier space, we obtain two coupled differential equations for the density and velocity evolution of the dark matter:

$$\dot{\delta} + ikv + 3\dot{\Phi} = 0
\tag{1.15}$$

$$\dot{v} + Hv + ik\Psi = 0,
\tag{1.16}$$

with the fractional overdensity δ defined through $n_{\text{dm}} = \bar{n}_{\text{dm}}[1 + \delta(\mathbf{x}, t)]$ and the mean density \bar{n}_{dm} .³

Even though dark matter particles does not interact with other particles, they are still bound by gravitation, and we cannot treat them alone (fig. 1.1). We will need equations for the evolution of photons and baryons as well, that in turn couple to each other through collisions and the metric. We also need equations governing the evolution of the gravitational fields Φ and Ψ from the Friedmann equation in the perturbed metric. Analogous to the dark matter equations above, we write down the equations for the evolution of radiation (photons and neutrinos) and gravitational potentials in Fourier space:

$$\dot{\Theta}_{r,0} + k\Theta_{r,1} = -\dot{\Phi}
\tag{1.17}$$

$$\dot{\Theta}_{r,1} - \frac{k}{3}\Theta_{r,0} = -\frac{k}{3}\Phi
\tag{1.18}$$

$$k^2\Phi + 3H(\dot{\Phi} + H\Phi) = 4\pi G a^2[\rho_{\text{dm}}\delta + 4\rho_r\Theta_{r,0}],
\tag{1.19}$$

where $\Theta_{r,i}$ is the monopole, dipole (and so on) moment of the perturbations to the radiation distribution and we have assumed $\Phi = -\Psi$. We have left out the

³Also note that we have left out the Fourier notation of δ , v , Φ and Ψ for simplicity.

higher order moments of the radiation perturbations because they are irrelevant in the epochs where radiation contributes to the potential and affects the dark matter distribution. Baryons are neglected here because it is assumed that the baryon to dark matter density is small. This set of coupled differential equations defines the evolution from the initial conditions from inflation to the Universe at late times, where matter dominates the energy budget.

The calculation of the transfer function $T(k)$ can only be done analytically in a few special cases, and we generally have to turn to numerical computations. Alternatively, one can create empirical fitting functions based on the few analytical solutions and compared with the exact numerical results. This was first done by Bardeen et al. [3], but in this project we have used the approximation by Eisenstein and Hu [17].

1.2.3 Growth factor

The transfer function of the previous section takes the dark matter field from the initial state of the Universe through the times where matter and radiation interact through the gravitational potential. It is clear, however, that when we are deep into the matter era, and radiation has no effect on the gravitational potential, we only need to follow the equations governing the dark matter and gravitational potential. This simplifies the calculation, and we can transform the problem (eqs. 1.15) into a single second-order differential equation in the dark matter fractional overdensity δ :

$$\frac{d^2\delta}{da^2} + \left(\frac{d\ln(H)}{da} + \frac{3}{a} \right) \frac{d\delta}{da} - \frac{3\Omega_m H_0^2}{2a^5 H^2} \delta = 0. \quad (1.20)$$

The growing solution to this equation is:

$$D_1(a) = \frac{5\Omega_m H(a)}{2H_0} \int_0^a \left[\frac{H_0}{a' H(a')} \right]^3 da', \quad (1.21)$$

where the conventional form of the growth factor $g(a) = D_1/a$ is written:

$$g(z) = D_1(z)[1+z]. \quad (1.22)$$

In a universe with dark energy, this has to be solved numerically or using an approximated formula. In this project we have used the approximation by Carroll et al. [10]:

$$g(z) \approx \frac{5}{2} \Omega_m \left[\Omega_m^{4/7} - \Omega_\Lambda + (1 + \Omega_m/2)(1 + \Omega_\Lambda/70) \right]^{-1}. \quad (1.23)$$

We are now able to complete equation 1.5, and we have an expression for the linear power spectrum at any time. In most parts of this project we work with the dimensionless power spectrum $\Delta^2(k) = k^3 P(k)/(2\pi^2)$, and figure 1.2 shows the evolution of this quantity from transfer and growth functions:

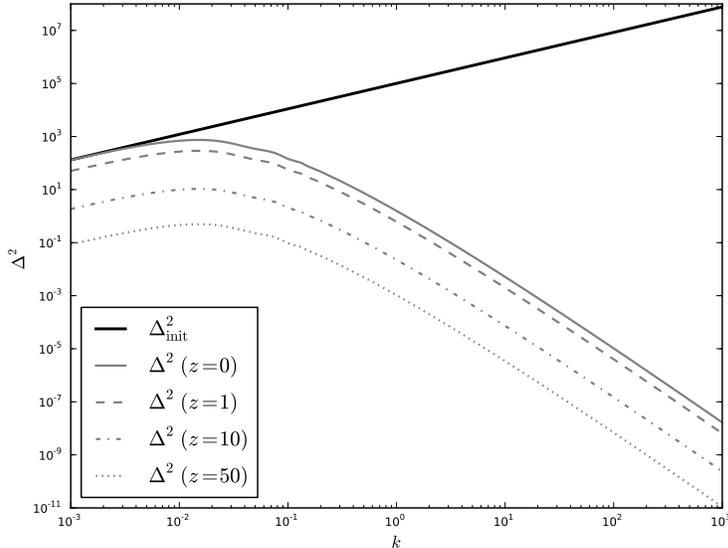


Figure 1.2: Evolution of the dimensionless dark matter power spectrum as a function of scale. The initial powerlaw shape of the power spectrum is shown in the solid black line, while the transferred power spectra is shown in gray. The solid to dotted gray lines indicate a redshift of 0, 1, 10 and 50.

1.2.4 Two-point correlation function

In equation 1.5 we introduced the linear power spectrum as the primary measure of large-scale structure. Fourier transforming the power spectrum back to real-space yields the *two-point correlation function*:

$$\xi(r) = \langle \delta(\mathbf{x})\delta(\mathbf{x} - \mathbf{x}') \rangle = \int \frac{\Delta^2(k)}{k} \frac{\sin(kr)}{kr} dk, \quad (1.24)$$

where we have once again used the dimensionless power spectrum Δ^2 . Note that the correlation function, just as the power spectrum, averages over the contributions from the different directions in \mathbf{x} , which is also the reason for the simplified Fourier transform.

The measurements and analysis done in this project will focus primarily on the correlation functions and variations hereof, as will be apparent in chapter 2. In many cases, the power spectrum and correlation functions are interchangeable, but with one subtle difference: the correlation function is defined from the real-space overdensity $\delta(\mathbf{x}) = \rho(\mathbf{x})/\bar{\rho} - 1$, and thus it measures the mean pairwise correlation with respect to the background density. It can therefore

be interpreted as the probability of finding two objects at a separation of r compared to the uniform (Poisson) probability in a volume element dV :

$$dP = n[1 + \xi(r)]dV, \quad (1.25)$$

where n is the number density of objects. In this project we will use the *cross-correlation* between two separate distributions, defined as:

$$dP = n_1 n_2 [1 + \xi_{12}(r)] dV_1 dV_2. \quad (1.26)$$

In chapter 2 we introduce estimators of the correlation function for the specific measurements that we make in this project.

1.3 Non-linear evolution

At some point in the evolution of cosmic structure, the overdensities in the dark matter field will be substantially larger than the background density, and it can no longer be described by a linear treatment of a pressureless fluid. Instead, Newtonian interactions become relevant and gravitationally bound structures start to form and decouple from the Hubble flow. In the gravitational potentials of these structures, the baryonic gas that makes up about 10% of the total matter density can start to fall in and introduce even more complex physics in the form of hydrodynamics and electromagnetism and form stars and galaxies. Additionally, these structures are not static as they can accrete nearby material and merge with other structures.

In this section we start by outlining the formation and evolution of these dark matter structures, often referred to as dark matter *haloes*, and move on to the formation of gaseous haloes and galaxies inside them.

We introduce the important subject of numerical simulations that are indispensable tools in the investigation of these matters. One would evolve the power spectrum in the linear regime using the theory of the previous section and input the result from a time where evolutions are still linear into the simulation which can then evolve the power spectrum further into the non-linear regime. It is possible, however, to go a long way describing the non-linear evolution analytically in a statistical sense, the basis of which we will need in chapter 3 for the modelling of our measurements.

1.3.1 Dark Matter haloes

The goal is to input the dark matter density field from linear theory δ_0 and find the number of virialized haloes that would form with a certain mass, a quantity usually referred to as the *mass function*. This problem was taken on by Press and Schechter [44] (PS) and their solution has been widely used since. Although the details of their results have been improved upon, it is illustrative to sketch their line of thought.

Firstly, we want to identify the regions of the linear field that are overdense enough for a halo to collapse gravitationally. For this, we define a field δ_s that smoothes the initial field with a spherical window function of radius R corresponding to a mass $M = (4\pi/3)\bar{\rho}R^3$:

$$\delta_s(\mathbf{x}) = \int \delta_0 W(\mathbf{x} + \mathbf{x}') d^3 \mathbf{x}', \quad (1.27)$$

with $W(\mathbf{x})$ defined to be 0 outside a volume $V = (4\pi/3)R^3$. The smoothed field has a variance $\sigma^2 = \langle \delta_s^2(k) \rangle$, which is the power spectrum $P(k)$ smoothed with the same window function:

$$\sigma(M) = \int \frac{\Delta^2(k)}{k} |W(kR)|^2 dk, \quad (1.28)$$

where we have again used the dimensionless power spectrum $\Delta^2(k) = k^3 P(k)/(2\pi^2)$ and the window function given in Fourier space as:

$$W(x) = [3/x^3](\sin x - x \cos x). \quad (1.29)$$

When the initial density field δ_0 is a gaussian random field, the smoothed density field δ_s will also be a gaussian random field with variance σ^2 [35]. The probability of having an overdensity higher than some critical overdensity δ_c required for a halo to collapse is then given by:

$$p(> \delta_c) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\delta_c}^{\infty} \exp\left[-\frac{\delta_s^2}{2\sigma^2}\right] d\delta_s. \quad (1.30)$$

Press and Schechter [44] now argued that the fraction $F(> M)$ of haloes with masses greater than M is equal to this probability. That is, the fraction of haloes above mass M is the same as the fraction of overdensities above a certain threshold δ_c in a field smoothed with a radius defined by M . The problem with this is that the mass located in underdensities never will be part of a structure, so Press and Schechter [44] inserted a fudge-factor of 2 to account for underdensities being part of the overdense regions, such that $F(> M) = 2p(> \delta_c)$. The density of haloes in the mass range $M \rightarrow M + dM$ is now given by $n(M)dM = (\bar{\rho}/M)(\partial F(> M)/\partial M)dM$ and reveals the Press & Schechter mass function:

$$n(M)dM = \sqrt{\frac{2}{\pi}} \frac{\bar{\rho}\delta_c}{M^2\sigma} \exp\left(-\frac{\delta_c^2}{2\sigma^2}\right) \left|\frac{d \ln \sigma}{d \ln M}\right| dM. \quad (1.31)$$

If we define the variable $\nu = \delta_c/\sigma$, the mass function can be written:

$$n(M)dM = \frac{\bar{\rho}}{M^2} f(\nu) \left|\frac{d \ln \nu}{d \ln M}\right| dM, \quad \text{where} \quad (1.32)$$

$$f(\nu) = \sqrt{\frac{2}{\pi}} \nu \exp\left(-\frac{\nu^2}{2}\right) \quad (1.33)$$

is the fraction of mass in a $\ln \nu$ halo, also known as the multiplicity function. The critical overdensity required for a halo to collapse δ_c is generally dependent on redshift and the choice of cosmology. For the standard Λ CDM cosmology introduced in the previous sections this value is found in simulations to be close to constant across redshifts at a value of $\delta_c = 1.696$ [45].

Given how the PS formalism include a certain degree of fudging, it has worked remarkably well to describe measurements and simulations in the decades since its arrival. With increasing resolution in both of these, however, the aging PS mass function is showing weaknesses. In the calculation of the mass function above, Press & Schechter used a spherical collapse model, while in the general case structures should be allowed to take any ellipsoidal form. Solving the problem in the more general case of ellipsoidal collapse, Sheth and Tormen [50] (ST) improved the mass function and introduced another expression for $f(\nu)$:

$$f(\nu) = A \left(1 + \frac{1}{a^p \nu^{2p}} \right) \left(\frac{a\nu^2}{2} \right)^{1/2} \frac{e^{-a\nu^2/2}}{\sqrt{\pi}}, \quad (1.34)$$

where $A = 0.322$, $a = 0.707$ and $p = 0.3$.

We can now use the mass function formalism to find a linear scaling to the overdensity field δ due to the gravitationally collapsed haloes with overdensity field δ_h [36]:

$$\delta_h = b_h(M)\delta, \quad (1.35)$$

where b_h is called the *linear halo bias*. It expresses that gravitationally collapsed structures are biased representations of the linearly evolved dark matter density field. Using the PS mass function, the bias takes the form:

$$b_h^{\text{PS}}(\nu) = 1 + \frac{\nu^2 - 1}{\delta_c}, \quad (1.36)$$

while for the ST case:

$$b^{\text{ST}} = 1 + \frac{a\nu^2 - 1}{\delta_c} + \frac{2p/\delta_c}{1 + (a\nu^2)^p}. \quad (1.37)$$

It is this form of the bias factor that we will employ later in chapter 3 when we model the measurements of chapter 2.

1.3.2 Galaxy formation and evolution

So far we have only been talking about dark matter, and with good reason: it makes up 90% of the mass budget of the Universe. However, we do not observe dark matter directly, but rather the luminous matter in the form of gas clouds and galaxies of stars made from the baryonic content of the Universe. We will not deal with the details of the baryonic components of haloes in the modelling of measurements in this projects, but it is nevertheless important to consider some of the processes in order to understand the galaxy populations we observe

today. We will use the heuristic arguments of this section extensively in order to discuss the measurement and analysis results later in this work.

Gas collapse

While the problem of galaxy formation is extremely complex, and still not completely understood, we can at least try to motivate the physics that drives these processes. Baryons have been tightly coupled to dark matter during the linear evolution of the Universe, but in the non-linear regime, where structures collapse under gravity, the fundamental difference between dark matter and baryons start to show. Under these circumstances, the baryonic gas can undergo shocks and heating because of its non-zero pressure, a party that dark matter is not invited to. These processes has the effect of ionizing the gas, radiating away energy and falling even deeper into the potential wells of the dark matter halo. As continued collapse increases the pressure and temperature, equilibrium between gravity and pressure is established and stars are formed. These first stars serves to ionize the gas even further with their energetic photons, and at some point most of the gas in the Universe is once again ionized. This process is therefore called *reionization*.

At this point, a lot of other processes become important [35]. Feedback from the formation of stars from interstellar clouds of gas *quenches* further star formation, supernovae drive an energetic *galactic wind* and central black holes heats the central gas serving again to quench star formation and thus the formation of galactic disks.

Mergers

An important process in the life of a young halo is *mergers*. If the haloes are of comparable mass, mergers can completely disrupt the disks of the galaxies and the end product is typically a galaxy with a larger bulge, because of the mixing of angular momenta. If the mass of the other galaxy is much lower, it is most likely accreted onto the more massive galaxy without affecting it dramatically.

We can use the formalism from section 1.3.1 to derive an analytical expression for the merger-rate of a halo of mass M . In particular, we can write the average number of haloes with a mass between M and $M_1 + dM_1$ at time t_1 that have merged to form a halo of mass M_2 at t_2 [35]:

$$n(M_1, t_1 | M_2, t_2) = \frac{M_2}{M_1} f(\nu_{12}) \left| \frac{d \ln \nu_{12}}{d \ln M_1} \right| dM_1, \quad (1.38)$$

where $f(\nu)$ is the multiplicity function (eqn. 1.33) and $\nu_{12} = (\delta_1 - \delta_2) / \sqrt{\sigma_1^2 - \sigma_2^2}$ the overdensity between t_1 and t_2 , where $\nu_i = \delta_c / D_1(t_i)$.

Evolution of galaxies in clusters

In this project we are concerned with the evolution of galaxies in dense environment such as the rich clusters of galaxies that makes up the largest gravitation-

ally bound structures in the Universe. Numerical simulations and observations show that the galaxy populations inside dense dark matter haloes are different to those in the field. Specifically, they are on average more elliptical, an observation known as the morphology-density relation [14, 24, 2, 58].

There are three processes that play a role for a galaxy in the dense environment of a cluster: (I) interactions with the gravitational potential of other members of the cluster or the cluster potential itself; (II) a gravitational process known as *dynamical friction* where the galaxy loses its angular momentum and falls towards the centre of the potential; and (III) the interaction of the baryonic matter in the galaxies with the hot X-ray emitting gas that is known to permeate clusters.

Galaxy harassment The velocity of galaxies in clusters are of the order $\sim 1000 \text{ km s}^{-1}$, much higher than the rotational velocity of the galaxies themselves. It is expected that the disruption of galaxies is related to the galaxy velocity by $\propto v^{-2}$, so cluster galaxies will typically not undergo mergers inside clusters but suffer less violent encounters [7]. A series of such events, however, can still have an effect on the galaxy.

A late-type Sc-Sd spiral galaxy, that is typically less tightly bound, will lose substantial parts of its structure in multiple fast encounters and be transformed to smaller galaxies with higher velocity dispersion, the so-called dwarf ellipticals that are seen plentiful in cluster environments. On the other hand, more regular Sa-Sb spirals that are more tightly bound, and where the orbital velocity is high enough, are able to absorb the violent interactions better and remain intact.

Dynamical friction Satellite galaxies of mass M_s undergo dynamical friction with the halo material and in the age of the Universe, they will have time to fall into the center of the halo if they are inside a radius r_{crit} [35, Chapter 12]:

$$r_{\text{crit}} \sim 0.1 r_h \left(\frac{\ln \Lambda}{10} \right)^{1/2} \left(\frac{M_s}{10^{-4} M_h} \right)^{1/2}, \quad (1.39)$$

where r_h is the halo radius and $\ln \Lambda \sim \ln M_h/M_s$. This means that more massive satellite galaxies will be subject to orbital decay at larger radii and will therefore be more likely to be cannibalized by the growing central galaxy.

Ram-pressure stripping The third process deals with the interactions between the intracluster medium (ICM) and the baryonic content of the galaxy in the interstellar medium (ISM). The consequence of these interactions is that the interacting part of the galaxy feels a wind from the cluster gas that effectively works to strip the galaxy of the gas in its outer parts. This may explain why there are a higher occurrence of S0 galaxies in cluster environments because even regular Sa-Sb spirals can be stripped of their star-forming material and take on a more elliptical morphology.

Together, this complex set of processes might transform a cluster galaxy, almost always towards more elliptical systems, explaining why we see more elliptical galaxies in dense environments, and therefore a stronger clustering of elliptical galaxies in general [14]. This heuristic argument also works for the color of the galaxy. When galaxies are disrupted in this way, they lose their star-forming gas and become older and redder systems. More massive galaxies are subject to stronger dynamical friction (eqn. 1.39) so they tend to fall towards the center of the cluster, and will also become more clustered. In practice, however, the interplay between these processes and other dynamical effects are extremely complicated and one often uses simulations to complement observations and theory.

If we want to learn something about galaxy evolution as a whole, in contrast to studying the individual systems, it is clear that we need to work with the statistics of the cluster and galaxy populations.

1.3.3 Numerical simulations

Because of the complex physics involved in nonlinear structure formation, simulations have become a very important tool in cosmology and astrophysics in general, both as a test for the theories developed and as an experiment in its own right. It complements real observations in the sense that we can get a view of the Universe at any time without the selection and redshift effects we have to struggle with in reality. But simulations also have shortcomings, and the most obvious one is resolution versus computing power. When simulating the Universe on large scale, it is hard to have an adequate resolution on galactic scales and vice versa. When simulating a large portion of the Universe, individual particles are typical $\sim 10^9 M_\odot$ per dark matter particle [52], which is a long way from the $\sim 1 \text{ TeV} \sim 10^{-54} M_\odot$ for the neutralino, a WIMP dark matter candidate. Whatever drives the Universe, it must be a hell of a supercomputer!

Even so, simulations are very powerful when used with care. The *Millennium Run* (Springel et al. [52]) is the biggest simulation of large-scale structure to date, and constitutes $\sim 10^{10}$ particles in a cube $\sim 600 \text{ Mpc}$ in length. Figure 1.3 shows a view of the dark matter density field from the Millennium simulation⁴, and we explore some of the details of creating these simulations in appendix B.

⁴<http://www.mpa-garching.mpg.de/galform/virgo/millennium/>

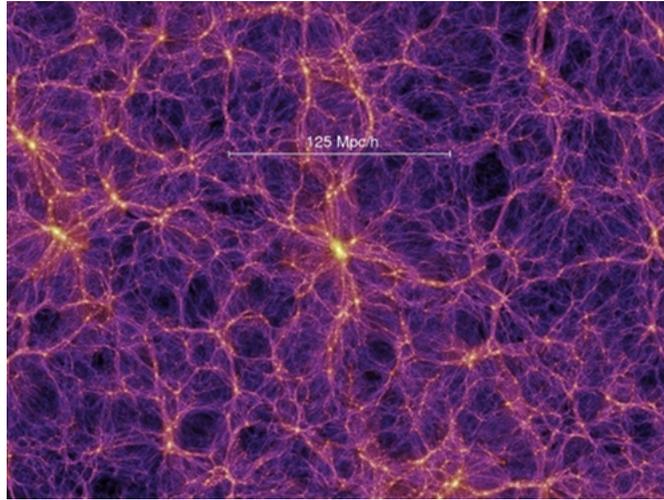


Figure 1.3: Pretty visualization of the dark matter density field in the famous Millennium simulation. We show the characteristic filament structure that has been created from the collapse of dark matter structures under gravity from the perturbations of the early Universe. Appendix B outlines the techniques used in these simulations in more detail.

Chapter 2

Correlation measurements

Having introduced the Universe in which we live, it is time to get our hands dirty and describe the data and measurements used in this project. We already introduced the concept of the two-point correlation function in section 1.2.4, and in this chapter we will describe what we want to measure and how this can be done in practice while carefully handling the challenges of working with real data.

Because of their power and simplicity, correlation statistics have been used throughout the literature to analyse measurements of the large-scale structure, usually in the form of large galaxy surveys like the *2-degree Field Galaxy Redshift Survey* (2dFGRS) [11] and *Sloan Digital Sky Survey* (SDSS) [60]. What is usually employed is the galaxy two-point auto-correlation that describes the distribution of galaxies with respect to all other galaxies (see eg. [57]). In this project we take a different approach and use the cluster–galaxy cross-correlation that measures the distribution of galaxies with respect to the rich galaxy clusters in which many of the galaxies reside. As these clusters represent the most massive dark matter haloes, this enables us to investigate how galaxies are distributed with respect to their large-scale environment. By then dividing the galaxy sample into subsamples with varying physical properties – like stellar mass, color and morphology – we can get a view of how galaxies with different properties have evolved in and outside clusters.

The *Cosmic Evolution Survey* (COSMOS) [49] makes these measurements possible by not only providing a large catalogue of galaxies to unprecedented depth, but also providing X-ray and lensing measurements of rich galaxy clusters in the same region of the sky. In this chapter we will start by introducing the COSMOS catalogues and the measurement methodology used in this project, and before presenting the final measurements, we will take a detour into the important topic of error estimation.

2.1 COSMOS

2.1.1 The COSMOS field

The COSMOS field is a 2-square degree near-equatorial patch of the sky that has been verified to have a low galactic extinction and be devoid of unwanted contamination, such as bright X-ray, UV and radio sources [49]. Observations has been carried out by many space- and ground-based telescopes, primarily the Hubble Space Telescope (HST) in the largest HST project to date, utilizing around 10% of the time in a 2 year period [48]. By choosing an equatorial field, ground-based telescopes from both the northern and southern hemispheres have access to it, thus enabling observations in many bands in order to improve the photometric redshift estimations. Some of these telescopes include the Japanese Subaru optical telescope [54] and the VLA radio telescope [47].

X-ray measurements are performed with the XMM-Newton space telescope, providing observations of the hot X-ray gas in rich galaxy clusters to unprecedented depth [21, 19]. Additionally, the Spitzer Space Telescope provides infrared observations and GALEX provides ultraviolet data [48]. Spectroscopic follow-up has been done with the VLT/VIMOS instrument for around 25,000 galaxies between redshift 0.3 and 1, providing tests and calibration of the photometric redshifts [28].

Figure 2.1 shows the masked out areas of the COSMOS field where stars from our own galaxy contaminate the observations.

2.1.2 Galaxies

The primary data product of COSMOS is the galaxy catalogue [9] built primarily from HST [48] and Subaru [54] observations. The total catalogue contains $\sim 2 \cdot 10^6$ objects to an I-band magnitude of $I_{AB} < 27$. The resulting catalogue contains, in addition to positions and redshifts, estimations of the morphology, the stellar mass, the absolute magnitudes and dust-corrected color.

These quantities are mostly estimated using a template fitting procedure where the templates of known galaxies are fitted to the data. In the catalogue used for this project, observations in more than 30 bands have gone into the template fitting leaving extremely accurate photometric redshift estimations [23].

In this project we cut the galaxies at $I_{AB} < 26$ and at the morphological index $i_M < 20$ which disregards highly irregular and starburst galaxies, leaving 59683 galaxies in total. Figure 2.2 shows the distribution of galaxies in the COSMOS field.

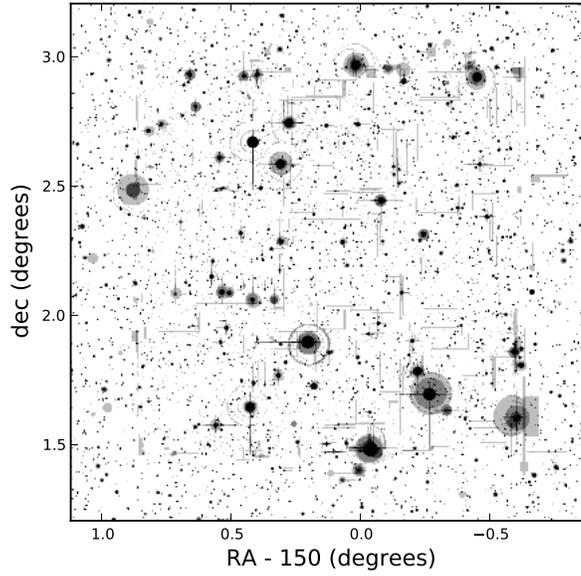


Figure 2.1: The mask of the COSMOS galaxy catalogue in 4 bands removing foreground objects from the field. Darker areas are masked in more bands.

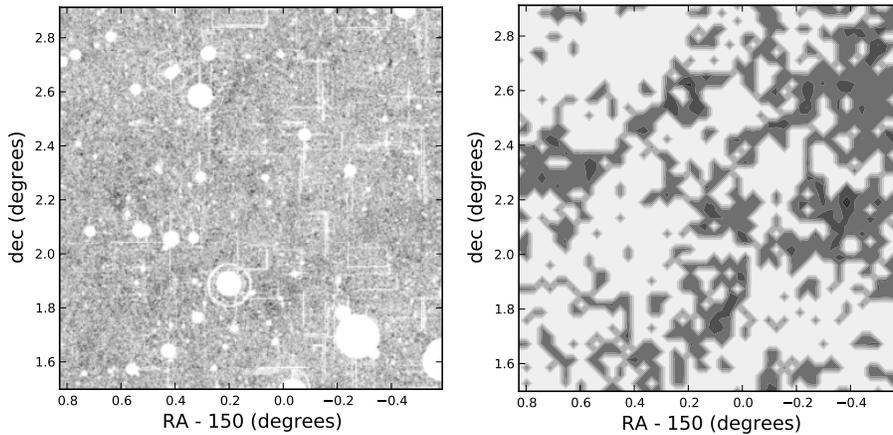


Figure 2.2: Galaxies in the COSMOS field with I-band magnitude $I_{AB} < 26$. The right panel only shows galaxies above a certain density threshold to highlight the large-scale structure, keeping in mind that some of the big voids are due to the masked-out areas.

Galaxy properties

In order to determine physical properties of the galaxies from photometry in the COSMOS passbands, Mobasher et al. [37] created a code based on spectral energy distribution (SED) fitting. The basic idea is that you start out with spectra of galaxies with known properties, which are then converted to template flux estimates in the passbands of the survey. By minimizing the χ^2 between these template fluxes and the observed fluxes, estimates for the galaxy properties are obtained. The properties fitted in this way are the redshift, the spectral type (morphology) and the extinction (E_{B-V}).

From these properties, Mobasher et al. [37] also estimates the stellar mass M_s of the COSMOS galaxies. From the measured luminosity and color, the stellar mass is found by:

$$\log\left(\frac{M_s}{M_\odot}\right) = \frac{M}{L_V} - 0.4(M_V - 4.82), \quad (2.1)$$

where M_V is the rest-frame absolute V-band magnitude and the mass to light ratio is given by the (B-V) color:

$$\frac{M}{L_V} = -0.628 + 1.305(B - V). \quad (2.2)$$

Note that the stellar mass estimate depends on the luminosity and thus redshift of the galaxy. In figure 2.3 we plot the distribution of each of the galaxy properties that is specifically used in this project.

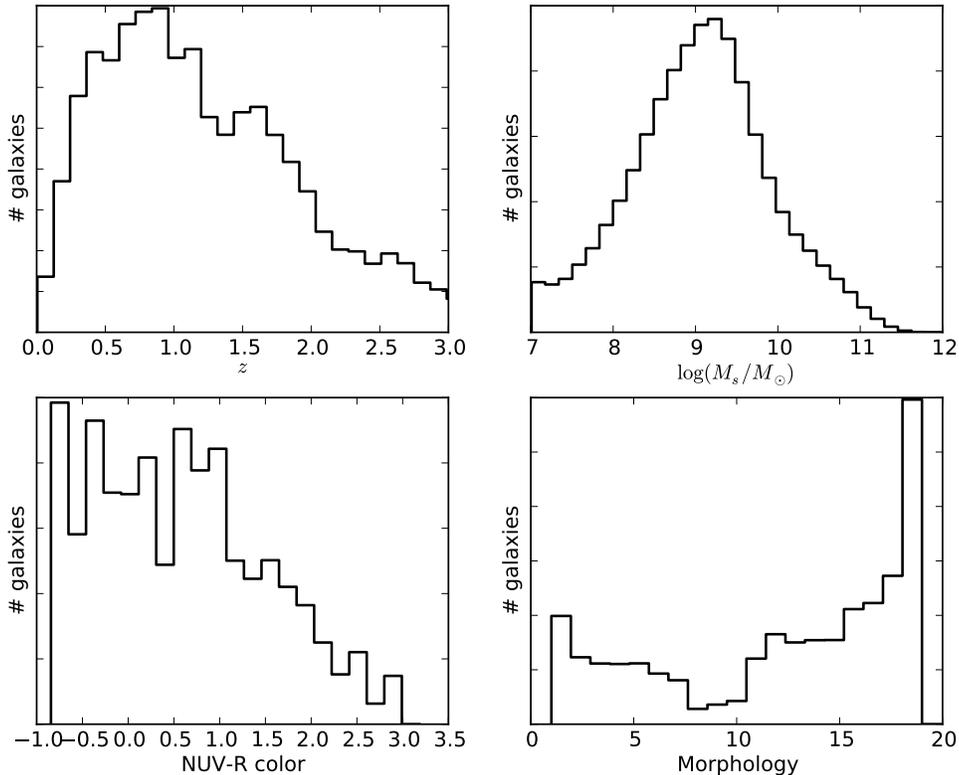


Figure 2.3: Distributions of COSMOS galaxy properties. *Top left.* Redshift distribution. *Top right.* Stellar mass in log solar masses. *Bottom left.* NUV-R dust-corrected color. *Bottom right.* Morphological index from ellipticals (1-8), Sa-Sc spirals (9-15) and Sd-Sdm spirals (16-19).

2.1.3 Clusters

The COSMOS cluster catalogue is based on the X-ray observations [21] and subsequent analysis in Finoguenov et al. [19]. In short, clusters are detected from the X-ray field under the assumption that extended regions of X-ray emission is due to the hot cluster gas. The clusters are then verified by the concentration of galaxies in the X-ray selected regions.

The clusters are plotted according to position in the field, redshift and estimated size in figure 2.4.

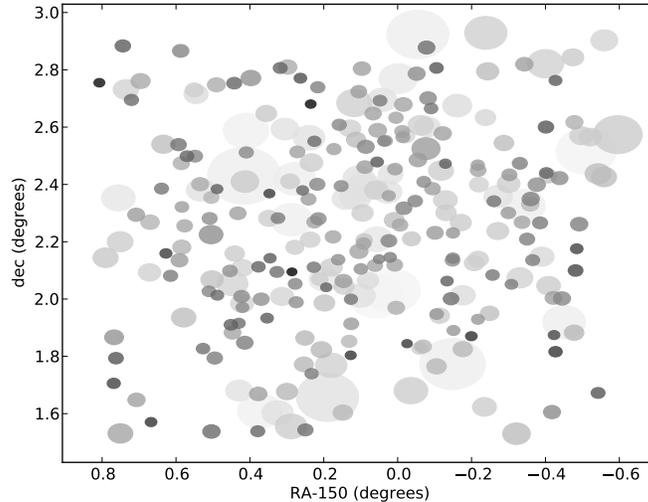


Figure 2.4: COSMOS X-ray identified rich galaxy clusters. The color reflects the redshift (lighter is deeper) and the radius of the circles is the estimated r_{200} .

Cluster properties

Cluster redshifts are determined by dividing the galaxy sample in redshift bins of ~ 0.1 and aligning the galaxy density peak with the extended X-ray emitting regions. Mass M_{500} and radius r_{500} estimates are found by the temperature-mass scaling relation:

$$M_{500} = 2.36 \times 10^{13} M_{\odot} T^{1.89} E_z^{-1}, \quad (2.3)$$

where the temperature T is defined from the K-corrected X-ray luminosity and choice of Λ CDM cosmology (see [19] for details). Because of the scaling relation scatter and uncertainty in the evolution of this relation with redshift, Finoguenov et al. [19] expect a mass estimate uncertainty of a factor of 1.4. Leauthaud et al. [26] constrain the mass-luminosity scaling relation further using weak lensing techniques.

The distribution of clusters with redshift and estimated mass is plotted in figure 2.5.

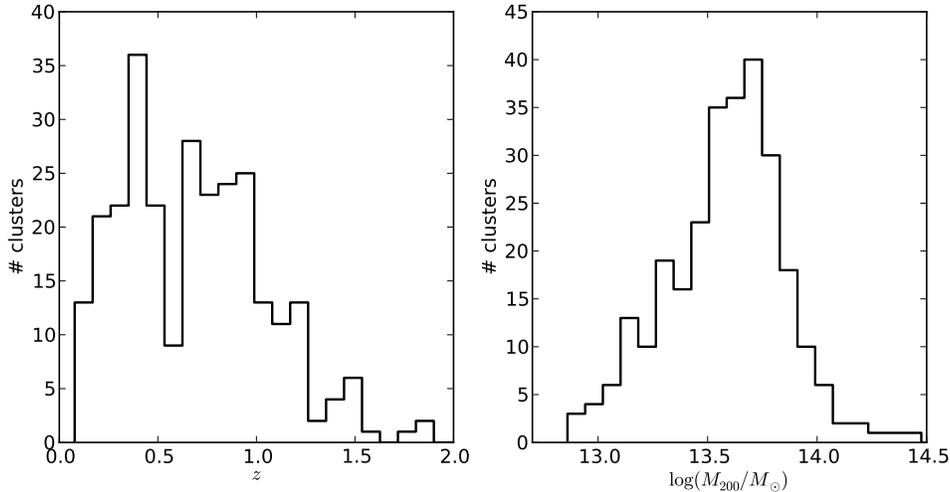


Figure 2.5: The redshift and mass distributions of COSMOS clusters.

2.2 Cluster – galaxy cross-correlation

In this project we take advantage of the fact that COSMOS provides us with both a galaxy and a cluster catalogue and we calculate the cross-correlation statistics between these two distributions. As the X-ray selected clusters represent dark matter haloes, this cross-correlation gives a better (than the galaxy auto-correlation) view of how the galaxy population depends on the underlying dark matter density field.

What makes two-point correlation functions so simple is that they in principle can be calculated simply by counting the number of pairs at certain separations. In our case of the cross-correlation, these are the pairs between any cluster and any galaxy, and we simply need to calculate the histogram of distances between these pairs and compare it to a uniform field of galaxies and clusters (see section 1.2.4). Of course this is not so simple in practice, and in addition to introducing the calculations, we will highlight three challenges that we have to overcome, namely the astronomical selection effects of the catalogues, the physical effects of working in redshift space as well as the computational difficulties in performing these calculations.

2.2.1 Estimator

To estimate the statistical quantity of the correlation function and avoid the astronomical selection effects, we need to compare the pair counts to a Poisson distributed field of galaxies. The simplest estimator is:

$$\hat{\xi} = \frac{DD}{DR} - 1, \quad (2.4)$$

where DD is the pair count histogram between data clusters and data galaxies and DR between data clusters and random galaxies, respectively. The random catalogue of galaxies need to define a uniformly distributed background of galaxies according to the selection effects of the survey, and thus needs to contain many more objects than the data catalogue.

In order to create the random catalogue of galaxies, we need to model the selection effects of the redshift distribution and a physical motivation for the expected redshift distribution can help us find a suiting model. In the small volume of local space, we see relatively few galaxies, but looking deeper the volume increases and we expect to see a polynomial increase in objects. Eventually, however, as the objects become fainter, the number will again decrease, this time with an exponential fall-off. This behaviour can be modelled using the Weibull distribution:

$$f(x; \lambda, k) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}, \quad (2.5)$$

which has a well-defined cumulative distribution from which to draw random redshifts:

$$F(x; k, \lambda) = 1 - e^{-(x/\lambda)^k}. \quad (2.6)$$

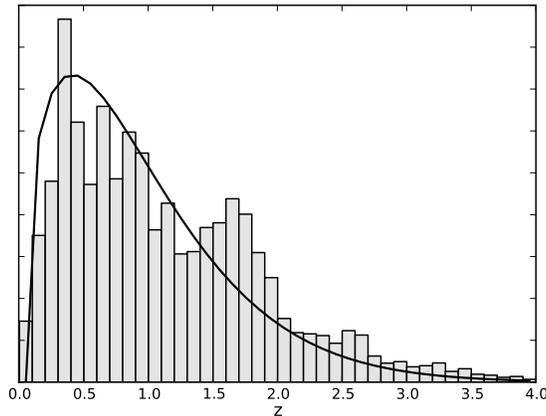


Figure 2.6: Fitting the redshift distribution of galaxies to the Weibull function. This is used to sample redshifts for the random galaxy catalogues.

In figure 2.6 we fit this distribution to the galaxies in the COSMOS catalogue. In addition to a redshift, we also need to generate an RA-dec pair for the mock

catalogue. This is a simple matter of drawing pairs from a uniform distribution inside the COSMOS field. Because we want the random catalogues to include all the non-clustering effects of the survey, we disregard RA-dec points inside the mask (fig. 2.1).

A more accurate estimator was introduced by Landy and Szalay [25] that requires mock catalogues for both the galaxy and cluster distributions, making the computations much more demanding:

$$\hat{\xi} = \frac{DD - DR - RD + RR}{RR}. \quad (2.7)$$

In practice, each of the pair count histograms needs to be normalized by the number of objects in the corresponding catalogues. The full calculation then reads:

$$\hat{\xi}_{cg}(r) = \frac{D_c D_g(r) n_c^R n_g^R}{R_c R_g(r) n_c^D n_g^D} - \frac{D_c R_g(r) n_c^R}{R_c R_g(r) n_c^D} - \frac{R_c D_g(r) n_g^R}{R_c R_g(r) n_g^D} + 1, \quad (2.8)$$

where n_i^X is the number of objects in the data/random cluster/galaxy catalogues and $X_i Y_j$ denotes the count of data D or random R objects of type clusters c or galaxies g separated by a distance of r .

2.2.2 Redshift effects

The estimated galaxy redshift is the sum of redshifts from the cosmological expansion and the peculiar velocity of the actual physical motion of the galaxies themselves¹. This means that a spherical object in real space will be distorted in various ways in redshift space because of the motion of the large-scale structures that the galaxy is part of as well as the motion of the galaxy inside non-linear collapsed haloes. These effects are of great importance when investigating the dynamics of the large-scale structure, but in this project they affect the distance (and hence correlation) estimates in unwanted ways, so we will in this project use a correlation measurement that alleviates this problem.

We start by transforming the measured distances into components along the line of sight π and perpendicular to the line of sight r_p . This transformation is defined by:

$$\pi = \frac{\mathbf{l} \cdot \mathbf{s}}{|\mathbf{s}|} \quad r_p^2 = \mathbf{l} \cdot \mathbf{l} - \pi^2 \quad (2.9)$$

where $\mathbf{l} = \mathbf{r}_c - \mathbf{r}_g$, $s = \mathbf{r}_c + \mathbf{r}_g$ and \mathbf{r}_i the position of the object in question.

The correlation function is now a function of π and r_p instead of just r , and the pair-counts form a 2-dimensional histogram in these variables. In order to get rid of the redshift effects we integrate along the line of sight, that is, we project the 2D histogram onto the r_p axis:

¹Disregarding the negligible gravitational redshifts

$$w_p(r_p) = \int_0^{\pi_{\max}} \xi(\pi, r_p) d\pi. \quad (2.10)$$

This is called the *projected* correlation function and this is the measurement that we will use throughout the rest of this project. We cut the integration at π_{\max} in order to avoid counting pairs at large line-of-sight separations.

2.2.3 Computational challenges

At last is the challenge of computing power and memory. As there is around $10^5 \cdot 10^2 = 10^7$ pair separations (of 4 bytes each), the memory requirements for calculating the histogram the usual way quickly becomes unmanageable. When working with the random catalogues with around 50 times the number of object in each, this becomes plain old impossible, requiring ~ 200 GB of RAM. To avoid this problem, we created a histogram algorithm that inserts values on the fly (requiring the boundaries to be specified).

A related problem is that of computing time. With $\sim 5 \cdot 10^{10}$ pairs in the random catalogues, this calculation can take a long time to finish. While the measurement code is mostly written in *Python*, we implement the pair counting routine in *Fortran* using the *f2py* system. Additionally, we span this easily parallelizable calculation among the CPU cores in the computer, reducing the computation time with a factor of the number of cores, neglecting any communication overhead that this introduces.

Along with other tricks, this allows us to do the random – random pair counts in a few minutes instead of many hours. For a more in-depth description of the software that we have developed as a part of this project, we refer to appendix C.

2.3 Error estimation

An important subject in this project is that of error estimation. In problems where there is no trivial way to propagate errors because the system involves too many unknown sources of error, it is common to use resampling techniques to estimate the errors from the data. There are two general classes of resampling techniques: the *jackknife* procedure, where subsets of the data are used to resample and recalculate by removing (jackknifing) parts of the data [33] and the *bootstrap* method where the samples are drawn randomly from the data with replacement [15]. For the sake of comparison, we implement both of these from the descriptions below.

2.3.1 Jackknife resampling

In the correlation measurements performed in this project, the jackknife procedure divides the RA-dec space into regions on a grid. For each recalculation of the correlation function, one cell of the grid is left out at a time. Figure 2.7

shows an example of dividing the COSMOS cluster catalogue into such a grid. For the correlation measurement the clusters in each of the cells would then be left out in turn.

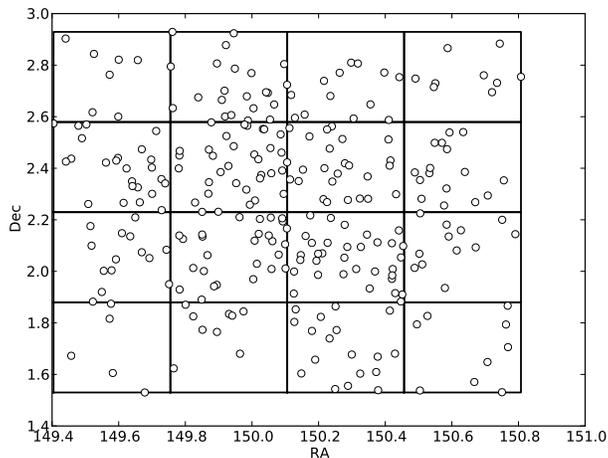


Figure 2.7: Splitting the cluster catalogue into 16 regions for the jackknife resampling procedure. The projected correlation calculation is redone with each of these regions left out in turn.

2.3.2 Moving block bootstrap

The same idea can be used for the bootstrap error estimation. In this case, we follow the so-called moving block bootstrap procedure [29] where the blocks used in the bootstrap are positioned randomly in the RA-dec plane. Figure 2.8 shows an iteration of placing these blocks. For each correlation function recalculation, a new cluster sample is constructed by the points inside the boxes. Note that one cluster might occur more than once in the constructed cluster sample and that the number of clusters vary slightly from iteration to iteration. Both of these features are known to improve the statistical accuracy of the error estimation[30].

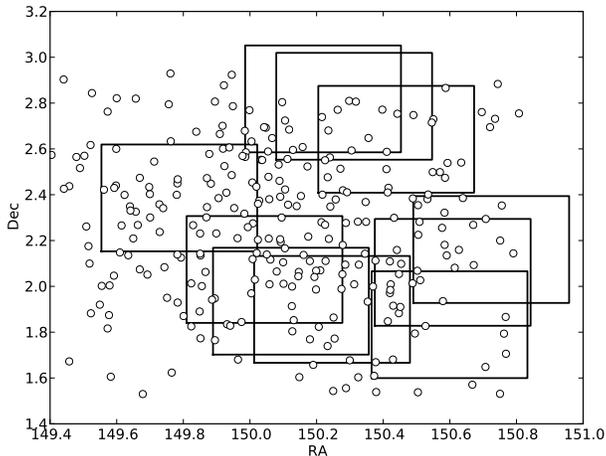


Figure 2.8: An example of a moving block run, where the correlation measurement is redone a number of times with a cluster catalogue built from blocks like these. This ensures that some clusters will be represented more than once and that the number of clusters vary slightly between runs.

2.3.3 Covariance and error bars

From the resampled measurements we create a matrix X_{ki} where k denotes the point on the correlation function (discrete r_p) and i denotes the i -th measurement of that point. Each point k will have a mean and a standard deviation given by:

$$\mu_k = \frac{1}{N} \sum_i^N X_{ki} \quad \sigma_k = \sqrt{\frac{1}{N} \sum_i^N [X_{ki} - \mu_k]^2} \quad (2.11)$$

where N is the number of samples.

This approach to error estimation does not take into account possible correlations between the points. For this reason it is common practice to calculate the full covariance matrix:

$$\Sigma_{ij} = \frac{1}{N^2} \sum_k^N [X_{ki} - \mu_i][X_{kj} - \mu_j] \quad (2.12)$$

The error bars (standard deviation) for the points are then given by the diagonal of this matrix:

$$\sigma_k = \sqrt{\Sigma_{kk}}. \quad (2.13)$$

Note that these can only serve as an indication of the errors of a point, and one must in general refer to the covariance matrix for a meaningful representation of the errors. For the jackknife method, the error estimation must be scaled by a factor $N - 1$, where N is the number of resamples, in order to account for the removal of blocks.

2.4 Measurements

In this section we present the execution of the measurement techniques described up until this point. To recap, the steps needed for each of the correlation measurements are:

1. Calculate the pair counts and obtain $\xi(\pi, r_p)$ [eqn. 2.9] via $\xi(r)$ [eqn. 2.8].
2. Calculate the projected correlation function $w_p(r_p)$ [eqn. 2.10].
3. Redo these calculation for the resampled data as described in section 2.3.3, and estimate the covariance matrix [eqn. 2.12] and the standard deviation (for error bars) [eqn. 2.13].

In the following sections we present these measurement on the full galaxy sample, on 3 stellar mass and redshift -divided subsamples, on 3 color-divided subsamples and on 3 morphologically divided subsamples, serving to quantify the evolution of clustering with these physical galaxy properties.

2.4.1 Full galaxy sample

To begin with, and to serve as an example, we present the correlation measurement for the full sample of COSMOS galaxies with I-band magnitude $I_{AB} < 26$ and morphological index below 20, that is, we only consider elliptical and early/late-type spiral galaxies.

As enumerated above, we start by calculating the $\xi(\pi, r_p)$ 2D histogram and plot the results in figure 2.9. This figure shows how the measurements are distorted in redshift space; (i) on smaller scales showing elongation in the π direction due to peculiar motion in local non-linear overdensities and (ii) on larger scales showing the squashing due to motion in the large-scale gravitational potential².

²Note that the figure is stretched in the r_p axis because we work in log-space. Working in linear space would show these effects even better

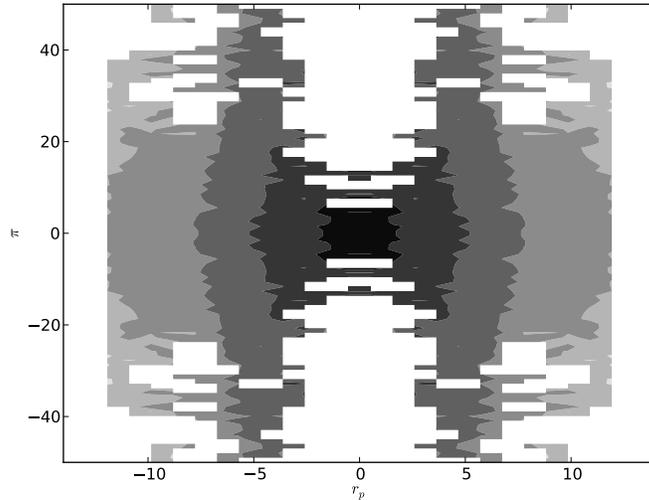


Figure 2.9: The $\xi(\pi, r_p)$ measurement of the full galaxy sample. The image has been mirrored in both axis to better demonstrate the redshift effects.

In figure 2.10 we plot the projected correlation function calculated from $\xi(\pi, r_p)$ using equation 2.10. We see clearly how the curve splits into two separate functions that represent correlations between a cluster and its galaxies and between a cluster and galaxies outside it (> 1 Mpc), respectively. In chapter 3, when we model these measurements, we will see how this separation follows naturally from the theory of dark matter haloes.

The covariance matrix is plotted in figure 2.11 and shows how the points are correlated. This particular covariance estimate has been made using the moving block bootstrap procedure of section 2.3.2, but it has been checked that the jackknife procedure of section 2.3.1 yields similar results.

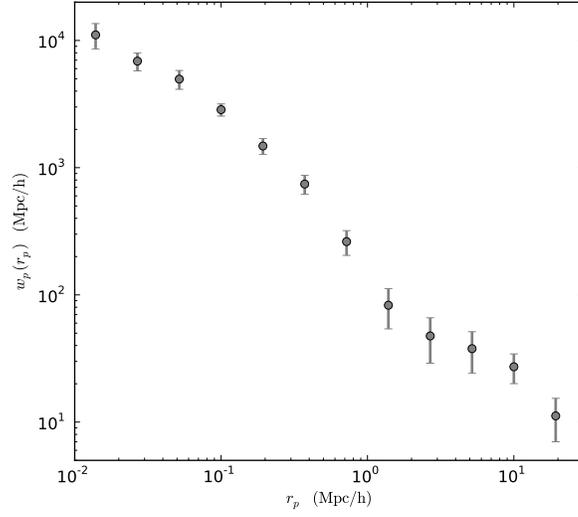


Figure 2.10: The projected correlation function for the full galaxy sample. The transition between the 1- and 2-halo terms is clearly visible around 1 Mpc h^{-1} .

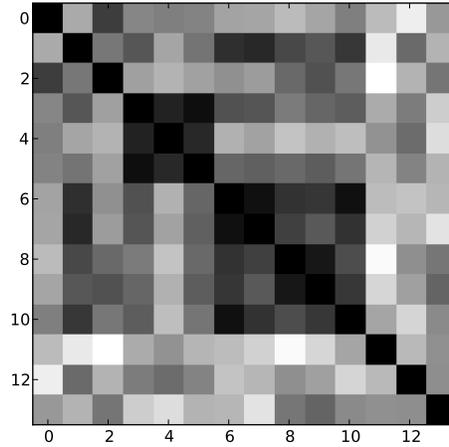


Figure 2.11: The covariance matrix for the correlation measurement of the full galaxy sample from the moving block resampling of clusters. The diagonal has been normalized to 1.

2.4.2 Varying galaxy properties

We are not only interested in the clustering of the full galaxy population, but also in the evolution of the clustering with different galaxy properties. For this task we divide the galaxies into subsamples of varying properties. Doing this, the number of pairs is reduced, so we expect the measurements to become noisier and consequently more uncertain. In addition we expect from section 1.3.2 that different galaxy populations have had different evolution histories, and that some types of galaxies are less represented in denser environments (such as clusters), which would create variations in the correlation function amplitudes across subsamples.

Evolution with stellar mass and redshift

We start by looking at the evolution with stellar mass and redshift. From section 1.3.2 we know that cluster galaxies are on average more massive than their field counterparts because of the accretion of mass onto the dark matter halo. This leads us to believe that the correlation measurements should show a higher amplitude for higher stellar mass galaxies. This simple argument is challenged if the ratio of dark matter to luminous matter is not constant with halo mass, as it would be possible for very massive galaxies to be fainter than expected. In fact, observations show that there is an increase in the mass-to-light ratio for the most massive galaxies because they will undergo ram-pressure stripping inside the cluster that quenches the star-formation [7]. For lower mass haloes, the mass-to-light ratio is also higher because their relatively weak potential could not capture the hot gas in the epoch of reionization.

In figure 2.12 we show how the galaxies are selected in the stellar mass – redshift plane. The boxes are placed so that we ensure completeness of the galaxies, i.e. no galaxies are left out because of the selection effects of the survey. In figure 2.13 we then present the results of the correlation measurement on these subsamples, with error bars from the covariance matrix diagonal as shown in section 2.3.3.

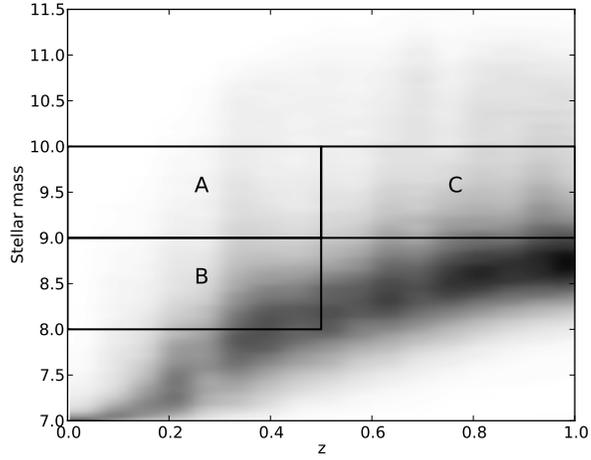


Figure 2.12: The three stellar mass – redshift subsamples of COSMOS galaxies.

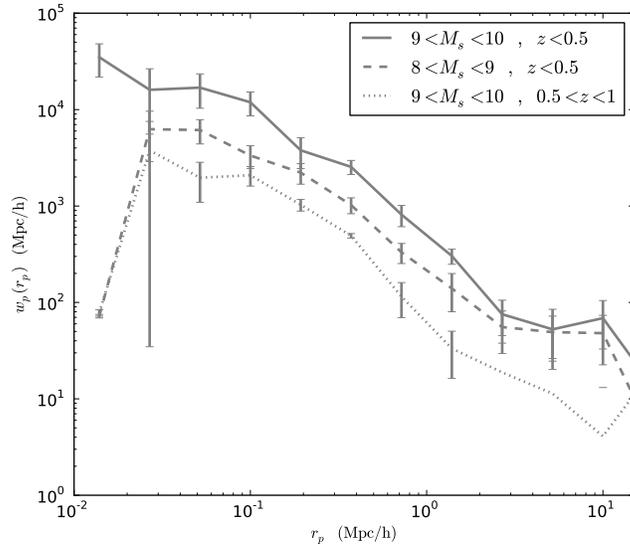


Figure 2.13: The projected correlation measurements of the three stellar mass – redshift subsamples of galaxies.

Firstly, we see a strong dependence on redshift, which is explained by the

fact that clusters will have had less time to fall in. It is also clear that the clustering is stronger for higher stellar mass galaxies, which supports the simple assumption that the evolution inside clusters is towards higher mass and more luminous galaxies. Additionally, in the central part of the correlation function, we see evidence for a stronger concentration of massive galaxies compared to the less massive. This is explained in section 1.3.2 as galactic cannibalism where satellite galaxies are slowly moving towards the center of the potential and eventually accreted onto the central galaxy.

Color

As outlined in section 1.3.2, the quenching of star-formation of galaxies in denser environments leads to an older, and hence redder, stellar population. In the same way as massive galaxies show stronger correlations in clusters, we therefore expect the correlation function of redder galaxies to have a higher amplitude.

The color selections for this measurement are shown in figure 2.14, and figure 2.15 shows the results of the correlation measurements.

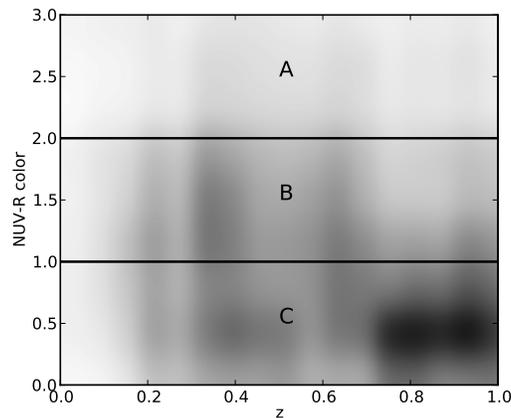


Figure 2.14: The three color-divided subsamples of COSMOS galaxies.

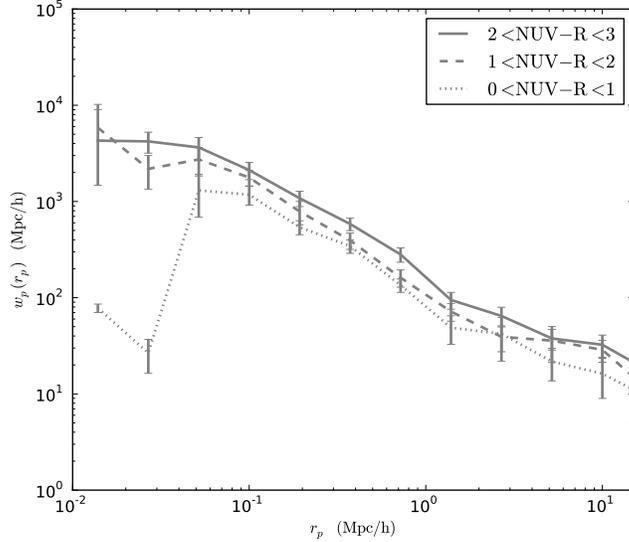


Figure 2.15: The projected correlation measurements of the three color-divided subsamples of galaxies.

Not as pronounced as in the stellar mass case, but there is evidence for evolution with color. On smaller scales we see how blue galaxies are on average less likely to reside in cluster centers, supporting the theory of accretion of old stars onto the central galaxy. Together with the results in the previous section, we can say that central galaxies are more likely to be more massive and redder. Of course, we see from section 2.1.2 that any massive galaxy is more likely to be redder, so the question is which of these properties is the primary causal connection to the evolution inside clusters. The basic idea from the previously discussed processes would in this case be that galaxies are redder because of ram-pressure stripping and subsequent quenching of star formation while they are more massive because of the accretion of mass from encounters. Because these effects are both present in denser environment supporting hot gas and a greater number density of galaxies, they often go together.

Note that for the bottom measurement, the error bars in the central part are clearly not to be trusted. They result from the calculation of errors in log-space from resamples with negative values.

Morphology

Lastly we have divided the galaxy population into bins in the morphological index going from 1-19 with elliptical to S0 having indices of 1-8, Sa to Sc spirals in 9-15 and Sd to Sdm spirals in 16-19. This is shown in figure 2.16 and

the results of the measurements are shown in figure 2.17.

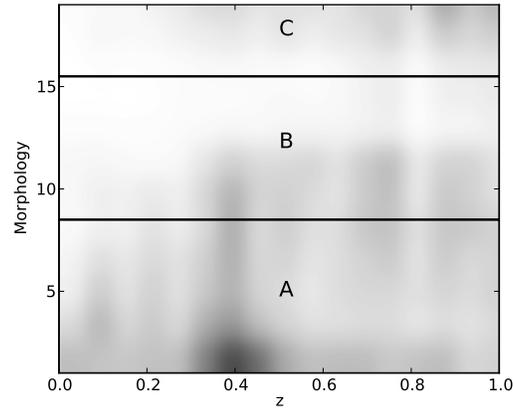


Figure 2.16: The three morphology-divided subsamples of COSMOS galaxies.

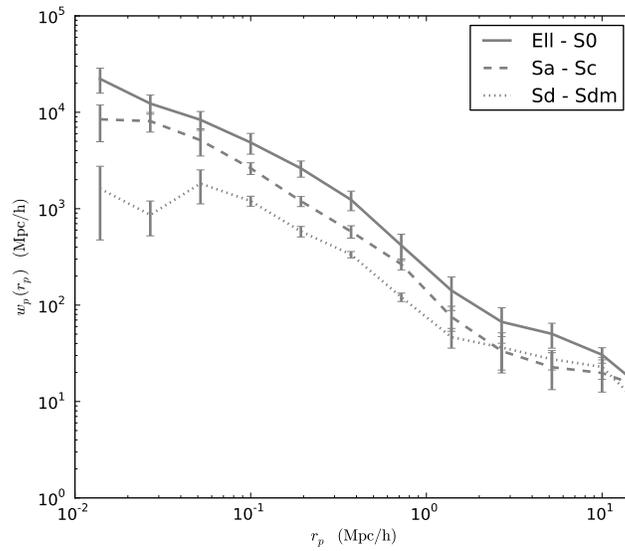


Figure 2.17: The projected correlation measurements of the three morphology-divided subsamples of galaxies.

If we follow the reasoning of the previous sections, it is obvious that more

elliptical galaxies, created from encounters between galaxies, will be more represented in cluster environments and therefore show stronger clustering. And for the same reasons that central galaxies are more massive and redder, we expect them to be ellipticals. The measurements in figure 2.17 show very clearly a trend towards higher central clustering for more elliptical galaxies.

Again, as discussed in the previous section, section 2.1.2 shows that any massive and any red galaxy is on average more elliptical. Knowing that galaxy encounters produce larger bulges because of the mixing of rotational velocities into a more random velocity dispersion, and that encounters are more likely in clusters, this correlation between stellar mass, color and morphology makes sense.

Chapter 3

Halo modelling

In the previous chapter we introduced the measurement technique and presented the measurements of the COSMOS cluster-galaxy cross-correlation function for various subsamples of the galaxy catalogue. In this chapter we review the halo model framework in section 3.1, introduce the parameters of the model and outline important aspects of applying this modelling in section 3.2 and finally we present the modelling results in section 3.3.

3.1 The Halo Model

To model the measurements in this project, we use the *halo model*, a theoretical framework for describing the distribution of matter in the Universe [12]. It builds on the ansatz that all matter is distributed in independent haloes and sub-haloes of dark matter that interact gravitationally. We will introduce this framework and how it applies to the specific case of cluster-galaxy cross-correlations. This review is based mostly on [12, 35] and builds on the description of dark matter haloes in section 1.3.1.

3.1.1 Introduction

If we assume that the halo mass M completely determines the profile of a halo, the dark matter density of a single halo can be described by:

$$\rho_{\text{halo}}(\mathbf{x}) = Mu(\mathbf{x}|M), \quad (3.1)$$

where \mathbf{x} is the position from the halo center and $u(\mathbf{x}|M)$ is the normalized halo profile given by M . Remembering that all mass is part of a halo, the total dark matter density field is then determined by summing up all the regions of space small enough to contain only one halo:

$$\rho_{\text{DM}}(\mathbf{x}) = \sum_i \mathcal{N}_i M_i u(\mathbf{x} - \mathbf{x}_i | M), \quad (3.2)$$

where $\mathcal{N}_i \in \{0, 1\}$ determines if there is a halo in the i -th region or not and $\mathbf{x} - \mathbf{x}_i$ is the distance from the center of the i -th region to \mathbf{x} .

On average, the probability of finding a halo of mass M in a volume element ΔV is given by $n(M)\Delta V$, where $n(M)$ is the number density of haloes of mass M – the mass function (section 1.3.1). The average density of the Universe is thus given by:

$$\langle \rho_{\text{DM}}(\mathbf{x}) \rangle = \int dM M n(M) \sum_i \Delta V_i u(\mathbf{x} - \mathbf{x}'|M). \quad (3.3)$$

In the limit of small ΔV , we can convert the sum to an integral over space:

$$\langle \rho_{\text{DM}}(\mathbf{x}) \rangle = \int dM M n(M) \int d^3 \mathbf{x}' u(\mathbf{x} - \mathbf{x}'|M), \quad (3.4)$$

Since $\int d^3 \mathbf{x}' u(\mathbf{x} - \mathbf{x}'|M) = 1$ per definition, the average density of the Universe $\bar{\rho}$ is given in the halo model as:

$$\bar{\rho} = \int dM M n(M). \quad (3.5)$$

3.1.2 Cluster – galaxy cross-correlation

As we've seen in section 1.2.4, the two-point correlation function $\xi(r)$ of dark matter haloes is defined from the overdensity $\delta(\mathbf{x}) \equiv \rho(\mathbf{x})/\bar{\rho} - 1$:

$$\xi_{\text{DM}}(r) \equiv \langle \delta(\mathbf{x}_1) \delta(\mathbf{x}_2) \rangle, \quad (3.6)$$

where $r \equiv |\mathbf{x}_1 - \mathbf{x}_2|$. Using the notation of the previous section, this can be expressed as:

$$\xi_{\text{DM}}(r) = \frac{1}{\bar{\rho}^2} \sum_{i,j} \langle \mathcal{N}_i M_i \mathcal{N}_j M_j u(\mathbf{x}_1 - \mathbf{x}_i | M_i) u(\mathbf{x}_2 - \mathbf{x}_j | M_j) \rangle. \quad (3.7)$$

This divides naturally into correlations from the same halo ($i = j$) and from different haloes ($i \neq j$), so that $\xi_{\text{DM}}(r) = \xi^{\text{1h}}(r) + \xi^{\text{2h}}(r)$. For the *1-halo term* we can then write:

$$\xi^{\text{1h}}(r) = \frac{1}{\bar{\rho}^2} \int dM M^2 n(M) \int d^3 \mathbf{y} u(\mathbf{y} - \mathbf{x}_1 | M) u(\mathbf{y} - \mathbf{x}_2 | M), \quad (3.8)$$

while the *2-halo term* includes contributions from each of the distributions as well as a halo-halo factor ξ_{hh} :

$$\begin{aligned} \xi^{\text{2h}}(r) &= \frac{1}{\bar{\rho}^2} \int dM_1 M_1 n(M_1) \int dM_2 M_2 n(M_2) \int d^3 \mathbf{x} \int d^3 \mathbf{x}' \\ &\times u(\mathbf{x}_1 - \mathbf{x} | M_1) u(\mathbf{x}_2 - \mathbf{x}' | M_2) \xi_{\text{hh}}(\mathbf{x} - \mathbf{x}' | M_1, M_2). \end{aligned} \quad (3.9)$$

On large scales, where the 2-halo term is important, individual haloes can be treated as points and the shape of the haloes themselves $u(\mathbf{x}|M)$ can be disregarded. As seen in section 1.3.1, haloes are biased representations of the linear dark matter field such that the halo-halo factor can be written as $\xi_{\text{hh}} = b_h(M_1)b_h(M_2)\xi_{\text{lin}}$. The 2-halo term (eqn. 3.9) then becomes:

$$\xi^{2\text{h}}(r) = \frac{1}{\bar{\rho}^2} \int dM_1 M_1 n(M_1) \int dM_2 M_2 n(M_2) b_h(M_1) b_h(M_2) \xi_{\text{lin}}. \quad (3.10)$$

This expression describes the correlation between dark matter fields and we need to convert it to an expression for the correlation between galaxy clusters and galaxies. The cluster population is determined by the cluster centers only, setting $u_{\text{cl}}(\mathbf{x}|M) = 1$. The galaxy profile $u_g(\mathbf{x}|M)$ describes the run of galaxies around cluster centers and is usually assumed to follow the dark matter profile. In addition, the weights M and $\bar{\rho}$ of the dark matter profile is exchanged for the average number of galaxies in haloes of mass M – the halo occupation distribution (HOD) $\langle N_g \rangle(M)$ – and the average density of galaxies $\bar{n}_g = \int dM n(M) \langle N_g \rangle(M)$. The 1-halo term thus takes the form:

$$\xi^{1\text{h}}(r) = \frac{1}{\bar{\rho}} \int dM M n(M) \frac{\langle N_g \rangle(M)}{\bar{n}_g} u_g(r|M). \quad (3.11)$$

Similarly, the 2-halo term becomes:

$$\xi^{2\text{h}}(r) = \xi_{\text{lin}} \left[\int dM \frac{M n(M)}{\bar{\rho}} b_c(M) \right] \left[\int dM n(M) b_c(M) \frac{\langle N_g \rangle(M)}{\bar{n}_g} \right], \quad (3.12)$$

where we refer to the square bracket factors as the *cluster bias* and *galaxy bias*, respectively.

The physical motivations of these expressions are as follows. The 1-halo term is built from the ensemble of clusters ($\int \dots dM$), all contributing to the correlation function with a galaxy profile weighted by the average number of galaxies for haloes of that mass ($\langle N \rangle M$). The 2-halo term is the result of taking the linearly evolved dark matter correlation of a chosen cosmology (section 1.2.4) and scaling it with the linear dark matter bias (section 1.3.1) and the halo occupation number.

Our model is then determined from the mass of the haloes and our choice of halo profile and occupation distribution. These choices are the subjects of the following sections.

3.1.3 Halo Occupation Distribution

The halo occupation distribution (HOD) is a central part of modelling the measurements of this project, and it is defined as the average number of galaxies that are present in haloes of a certain mass. N-body simulations has shown that dark matter haloes tend to build up an extensive envelope of substructure

from the accretion and disruption of haloes falling into the larger halo [38, 53]. These subhaloes represent galaxies in a galaxy cluster, so if we can describe the occupation of subhaloes, we can infer the occupation of galaxies and vice versa.

From the merger rate of the Press-Schechter formalism (section 1.3.1), it follows that more massive clusters will accrete more substructure and hence have a higher halo occupation. We parametrize this by the occupation slope α so the halo occupation number reads:

$$\langle N_g \rangle(M) = \left(\frac{M}{M_s} \right)^\alpha, \quad (3.13)$$

where the scale mass M_s scales the relation. At lower mass, a nascent dark matter halo may not have the potential to capture the hot gas of the reionization epoch, so we expect there to be a threshold mass M_0 above which a dark matter halo will host a single galaxy. The HOD can now be expressed as [5]:

$$\langle N_g \rangle(M) = \begin{cases} 0 & \text{if } M < M_0 \\ \left(\frac{M}{M_s} \right)^\alpha & \text{if } M \geq M_0 \end{cases}. \quad (3.14)$$

Because of the complex physics of galaxy evolution inside haloes, it is hard for Press-Schechter and similar theories to predict anything other than the general form of the HOD as written above. The parameter values (still) have to be found from simulations or fitted to observations, and we present some recent values from Tinker et al. [57] in table 3.1.3.

Parameter	Value
$\log(M_0/M_\odot)$	12.21 ± 0.11
$\log(M_s/M_\odot)$	13.46 ± 0.05
α	1.03 ± 0.05

Table 3.1: HOD parameter values for intermediate luminosity haloes from Tinker et al. [57].

As we will discuss further in section 4.1.1, the COSMOS clusters are distributed in a relatively narrow mass range, making it impossible for us to constrain M_0 and making M_s and α highly degenerate. In the actual modelling in the next sections, we therefore fix M_s to recent values and concentrate on α as the free HOD parameter. Also because of this, we should be reluctant to rely on the absolute values of α fitted to the data, but it will serve as a useful comparison of the HOD from different galaxy populations.

We also have to note that it is common to use more complicated expressions for HOD, where the transition between 0, 1 and N-galaxy haloes are modelled in much more detail (see eg. [57]). In this work we have no haloes in the mass regions relevant for these effects, so we ignore them.

3.1.4 The Dark Matter profile

The other important part of the modelling is of course the density profile of the dark matter haloes, or rather, in this project, the density profile of galaxies around cluster centers. It is usually assumed that this profile follows that of the dark matter profile, which is a good first assumption because of their gravitational connection. But since galaxies are completely different objects than dark matter particles, they don't need to follow the dark matter profile exactly, and indeed one of the goals of this project is to test parts of this assumption.

For the dark matter profile we choose the Navarro, Frenk and White profile (NFW) [39]. We have chosen this profile because of the extensive use in the literature, although recent research suggests that this profile may not be the most precise [31] (see sections 3.3.3 and 4.1.2). The NFW density profile can be written as:

$$\rho(r) = \frac{\Delta_c}{\left(\frac{r}{r_s}\right) \left(1 + \frac{r}{r_s}\right)^2}, \quad (3.15)$$

where the amplitude Δ_c and the scale radius r_s are parameters of the model. As we've seen in section 3.1.1, we assume that this profile is determined completely by the mass of the cluster, so the task at hand is to find the connections from the halo mass to these two parameters. This procedure is based on Navarro et al. [39] and Peacock and Smith [42].

From the peak parameter $\nu = \delta_c/\sigma(M)$, defined in section 1.3.1, we can find the formation redshift z_f that is empirically related to the profile amplitude Δ_c :

$$\Delta_c = 3000(1 + z_f)^3 \quad (3.16)$$

$$z_f = D_1^{-1}(1 + \nu^{-1}), \quad (3.17)$$

where D_1 is the linear growth factor $g/(1+z)$ defined in section 1.2.3.

The scale radius is found through the relation between Δ_c and the more familiar concentration parameter c that is in turn related to r_s via the virial radius r_{vir} :

$$c^{-1} = \frac{400}{3\Delta_c} + \left(\frac{110}{\Delta_c}\right)^{0.387} \quad (3.18)$$

$$c = r_{\text{vir}}/r_s, \quad (3.19)$$

where $r_{\text{vir}} \equiv r_{200} = [3(200\bar{\rho})/(4\pi M)]^{1/3}$ and $\bar{\rho} = \Omega_0 h^2 \rho_c$ the physical background density of matter in the Universe.

3.2 Modelling the measurements

From the formalism of the halo model, we now turn to the actual modelling of data. Before we present the results of the modelling in the next section,

we will in this section determine the free parameters of the model and how we implement the equation numerically. In contrast to the review in the previous section, this section and the next describes our work.

3.2.1 Model parameters

With the halo profile in place as a function of the halo mass alone, we are in a position to choose the parameters we want to investigate.

Occupation distribution

The first unknown parameter is the halo occupation slope α . As already mentioned in section 3.1.3, this parameter is highly degenerate with the other relevant HOD parameter, the satellite scale mass, M_s . We therefore keep this parameter fixed at a value inferred from recent studies of the HOD [56, 57]. We have no power to constrain the minimum mass M_0 because our cluster population does not sample the low-mass range where this effect is important.

In section 4.1.1 we discuss these subjects further.

Halo tilt

Motivated by the discussion of how galaxies evolve inside clusters in section 1.3.2, we want to introduce a parameter that quantifies possible discrepancies between the dark matter profile of the halo and the number density of galaxies (subhaloes) in the halo. Recent studies include a parameter that accounts for a possible difference in amplitude [57], but the general shape of the 1-halo term might also be different. To parametrize this difference, we introduce a new parameter that we name the *halo tilt* γ_h . This parameter describes a tilt of the 1-halo term of the correlation function:

$$\xi_g^{1h}(r) = [\xi_{DM}^{1h}(r)]^{\gamma_h}, \quad (3.20)$$

Physical motivations for such a discrepancy would be that due to the nature of the baryonic content of galaxies and the dense cluster environment, galaxies might be clustered stronger or weaker towards the halo center (see section 1.3.2).

We could have tilted the dark matter profile alone, that would in turn affect the 1-halo term. In the present choice, the amplitude of the HOD, determined from the other free parameter α , is also affected slightly by changes in γ_h . In our case, where we are not in a position to determine the absolute value of α , however, this effect is ignored (see sections 3.1.3 and 4.1.1).

3.2.2 Cluster ensemble

In section 3.1.2 we introduced analytical expressions for the cluster – galaxy cross-correlation 1- and 2-halo terms (eqs. 3.11, 3.12). These equations include integrals over the halo mass range, but since we have a finite number of clusters

in our sample, we can convert these integrals to a sum over COSMOS clusters. The 1-halo term becomes:

$$\xi^{1\text{h}}(r) = \sum_i^{n_c} \frac{\langle N_g \rangle(M_i)}{\bar{n}_g} u_g(r|M_i), \quad (3.21)$$

with n_c the total number of clusters. The 2-halo term takes the form:

$$\xi^{2\text{h}}(r) = \xi_{\text{lin}} \left[\sum_i^{n_c} b_c(M_i) \right] \left[\sum_i^{n_c} \frac{b_c(M_i) \langle N_g \rangle(M_i)}{\bar{n}_g} \right]. \quad (3.22)$$

In principle we need to specify the full mass function integrated over all masses, but the mass range of COSMOS clusters are such that we can neglect the effect of the large-scale environment on the occupation distribution [56].

3.2.3 Projection

In the measurements described in section chapter 2, we do not end up with $\xi(r)$, but rather the projected correlation function $w_p(r_p)$. In order for the model to describe the same thing, we need to transform the analytical $\xi(r)$ into the 2D $\xi(\pi, r_p)$ and project it onto the r_p axis, ie. integrating along the line-of-sight variable π :

$$w_p(r_p) = \int_0^{\pi_{\text{max}}} \xi(r) d\pi. \quad (3.23)$$

From equations 2.9 we see that:

$$r^2 \equiv \mathbf{1} \cdot \mathbf{1} = r_p^2 + \pi^2, \quad (3.24)$$

which allows us to make the substitution $r = \sqrt{r_p^2 + \pi^2}$:

$$w_p(r_p) = 2 \int_{r_p}^{r_{\text{max}}} \frac{\xi(r) r dr}{\sqrt{r^2 - r_p^2}}, \quad (3.25)$$

where $r_{\text{max}} = \sqrt{\pi_{\text{max}}^2 + r_p^2}$.

3.2.4 Likelihood estimation

The χ^2 is a common choice for expressing the *goodness of fit*:

$$\chi^2(\theta) = \sum_i \left[\frac{x_i - f(x_i|\theta)}{\sigma_i} \right]^2, \quad (3.26)$$

where θ is the parameter vector, the function $f(x_i)$ denotes the model prediction of point x_i and σ_i is the standard deviation. As explained in section 2.3.3, it is better to use the full covariance matrix in stead of just the standard deviation, and the χ^2 is generalized into the matrix equation:

$$\chi_g^2(\theta) = [\mathbf{x} - f(\mathbf{x}|\theta)]^T \Sigma^{-1} [\mathbf{x} - f(\mathbf{x}|\theta)]. \quad (3.27)$$

The fitting procedure consists of minimizing this quantity. Many minimization methods exist, but in this project we use a simple algorithm for multi-dimensional minimization called the *downhill simplex* or *Nelder-Mead* method [41]. This is a method implemented in most scientific toolboxes, and specifically we use the *fmin* function of the *scipy.optimize* package for *Python* (see appendix C).

Minimization of the χ^2 allows us to optimize the parameters of the model, but does not provide any information regarding the uncertainty of the parameters. In order to obtain this, we calculate the full (joint) likelihood distribution, a function of the model parameters (introduced in the previous section):

$$p(\alpha, \gamma_h) = \exp \left[-\frac{\chi^2(\alpha, \gamma_h)}{2} \right]. \quad (3.28)$$

This procedure yields the joint probability distribution of α and γ_h , and we can find and plot the 68%, 95% and 99% confidence limits as shown in the results section below (fig. 3.3). Furthermore, we can marginalize each of the parameters in turn to get the estimated standard error for a single parameter in order to show the error bars on a plot. This is done by summing up the values onto a single axis and again finding the value where some percentage of the data falls within.

Of course there are many methods for optimizing and finding the likelihood of variables. Because our case was rather simple, we chose a simple solution. If we wanted to optimize in a parameter space of higher dimension, we could use the more powerful *Markov Chain Monte Carlo* methods such as the *Metropolis-Hastings* algorithm [32].

The Nelder-Mead method can quite easily get stuck in local minima, so one has to be careful about this. In our case of only two free parameters, we have checked that this is not a problem. As we will see in the next section, the likelihood distribution is not bimodal in any way, so we are confident that the method finds the correct minimum.

3.3 Results

In this section we present the results of the modelling as outlined in the previous sections of this chapter. To refresh, we summarize the modelling steps here:

1. Obtain the r_p , w_p and error data from the measurements presented in section 2.4.
2. Calculate the 1- and 2-halo terms of the model from the cluster sample from equations 3.21 and 3.22.
3. Integrate over π to get the projected correlation function (eqn. 3.25).

4. Fit the model to the data and calculate χ^2 and likelihood estimates of the free parameters.

The code that enables this analysis is described further in appendix C.

3.3.1 Full galaxy sample

As with the measurements, we start out by looking at the full sample of COSMOS galaxies with $I_{AB} < 26$ and morphology < 20 . Even though we introduced the halo tilt parameter, it is illustrative to see how the halo model fits the data with only the HOD slope α as a free parameter. The result is plotted in figure 3.1.

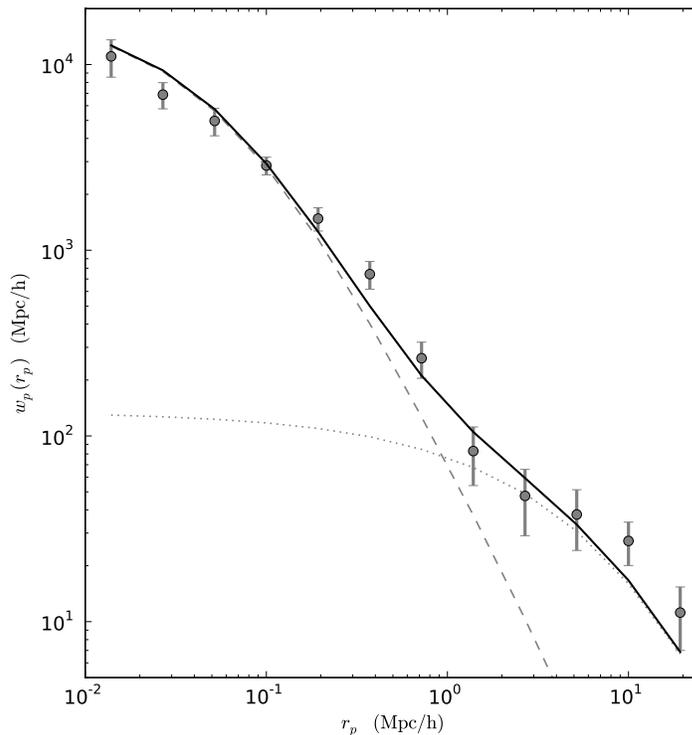


Figure 3.1: The halo model fit (solid line) to the measured projected correlation function of the full COSMOS galaxy catalogue. Dashed lines show the individual 1- and 2-halo terms. The model shows nice agreement with the data for $\alpha = 0.91 \pm 0.053$ and $\chi^2 = 2.5$.

We see here how the model is able to fit very well to the data given only a single free parameter. For reference we find $\alpha = 0.91 \pm 0.053$, in good agreement with recent studies (eg. [57], see table 3.1.3).

If we look at the figure in more detail, however, a slight discrepancy can be seen in the 1-halo regime, where the dark matter profile seems to overestimate the inner part and underestimate the outer part. This is good support for the halo tilt parameter that we have introduced in section 3.2.1, and in figure 3.2 we plot the modelling with both α and the halo tilt as free parameters.

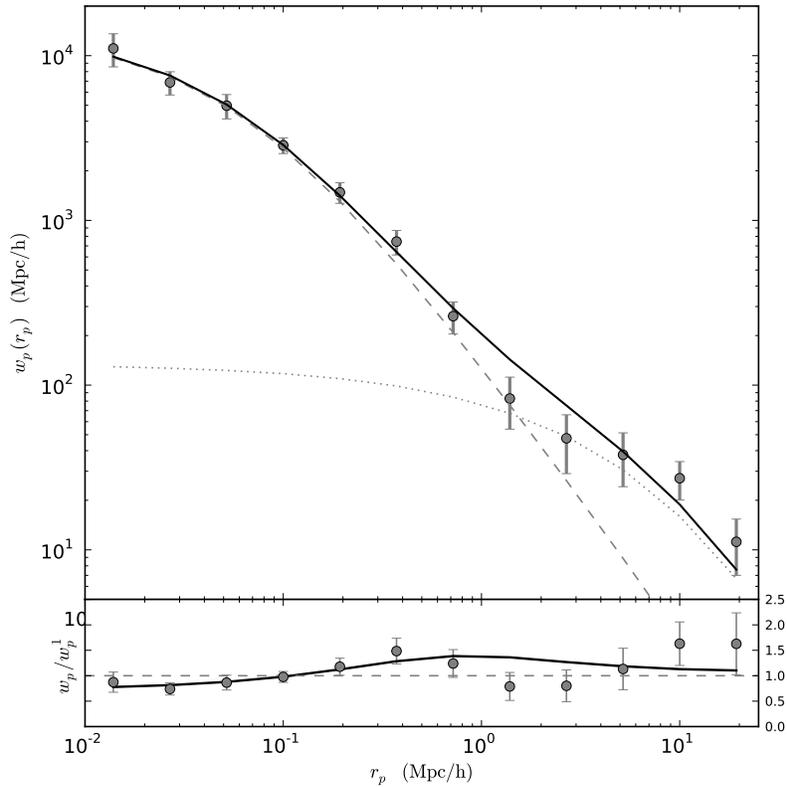


Figure 3.2: The projected correlation function (points) with the full model (solid line) and the 1- and 2-halo terms (dashed and dotted lines). The bottom panel shows the tilted (solid) to the untilted model (dashed), indicating that the data is better fitted by the tilted model in the inner part. A maximum likelihood ratio test between the two models favors the tilted model with 5.2% confidence. This test is a simple (maximum likelihood) case of the full bayesian evidence calculation.

The tilted model is a better fit to the 1-halo term of the measurements, while the more uncertain outer region is fitted more poorly due to the tilting. Other explanations for this discrepancy on intermediate scales might be that this analysis requires a scale-dependent bias factor that we ignore. We will discuss this further in chapter 4.

In figure 3.3 we plot the likelihood distribution with joint confident limits of the α and γ_h fit. We see that the parameters are anti-correlated, so that increasing one will decrease the other. This is expected because both parameters will scale the correlation function as discussed in section 3.2.1. Marginalizing this distribution in γ_h , we find that there is $\sim 5\%$ chance of no tilt ($\gamma_h = 1$) for this measurement.

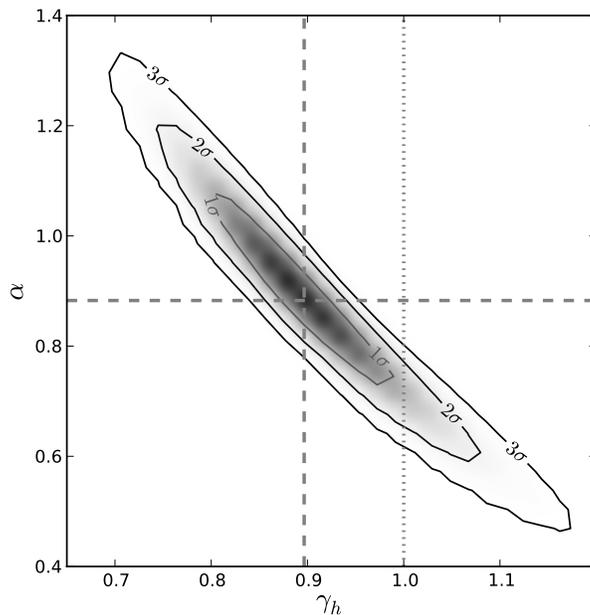


Figure 3.3: The 2D likelihood plot of the fitted α and γ_h parameters. The contours show the 68%, 95.4% and 99.73% confidence intervals.

3.3.2 Varying galaxy properties

We now turn to the various subsamples of galaxies, selected in section 2.4.2, to see how the HOD and halo tilt vary with galaxy properties. From the arguments in section 1.3.2, we believe that the correlation function should be scaled with galaxy properties, specifically we expect more massive, redder and more elliptical galaxies to be more clustered, which should scale the HOD slope ac-

cordingly. Regarding the tilt, which seemed to be favoured for the full galaxy sample above, it will be interesting to see if there is any evolution with galaxy properties.

We repeat the modelling steps with the tilted halo model and calculate the joint likelihood distribution from which we marginalise onto the parameter axis in order to show points with error bars in an $\alpha - \gamma_h$ plot.

In figure 3.4 we plot the results of the modelling on the correlation measurements of each of the galaxy subsamples introduced in section 2.4.2.

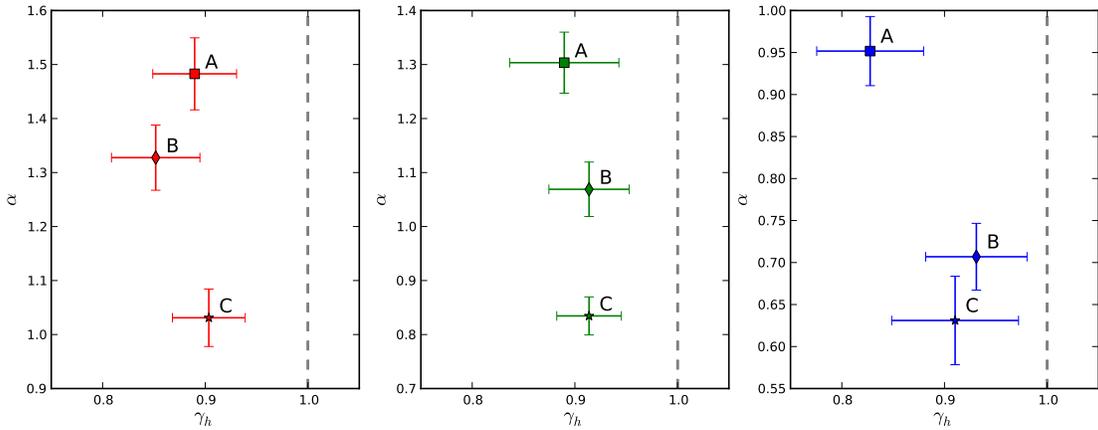


Figure 3.4: The results of the tilted modelling of the subsamples of galaxies, showing the values of the model parameters α and γ_h with marginalised uncertainties with respect to no tilt (dashed line). *Left.* Stellar mass and redshift: (A) $9 < M_s < 10$ and $0 < z < 0.5$ (B) $8 < M_s < 9$ and $0 < z < 0.5$ (C) $9 < M_s < 10$ and $0.5 < z < 1$. *Center.* Morphology: (A) Ellipticals - S0 (B) Sa - Sc spirals (C) Sd - Sdm spirals. *Right.* Color: (A) $2 < NUV - R < 3$ (B) $1 < NUV - R < 2$ (C) $0 < NUV - R < 1$.

We see in this figure the evolution in the halo occupation and tilt with galaxy properties. As expected, the more massive, redder and more elliptical galaxies have higher HOD slope α . The tilt, on the other hand, seems to stay more or less at a fixed value around 0.9. A tilt of 0.9 means that the galaxies are clustered less steeply than the dark matter in the halo. Having fewer galaxies in the next-to-central part of the halo ($r < 0.1$ Mpc/h) could be motivated physically by having a central galaxy eating up massive satellites in close orbits because of dynamical friction (eqn. 1.39). Likewise, a higher number of galaxies in the outer region ($0.1 < r < 1$ Mpc/h) could be explained by the accretion of subhaloes onto the main halo.

Table 3.3.2 lists the values for alpha and tilt for all of the measurements in this project, and figure 3.5 collects all measurements in a single plot. From this

plot we see that there might be a slight indication that the tilt evolves with α , which would mean that more clustered populations are more tilted. As we find that more massive galaxies are more clustered, this supports our thoughts on dynamical friction of massive galaxies in close orbits.

Measurement	α	γ_h	$p(\gamma_h = 1)\%$	$\#\sigma$
Full sample	0.883 ± 0.061	0.897 ± 0.061	4.57	1.7
$9 < M_s < 10, 0 < z < 0.5$	1.483 ± 0.067	0.890 ± 0.041	0.34	2.7
$8 < M_s < 9, 0 < z < 0.5$	1.328 ± 0.060	0.852 ± 0.043	0.03	3.4
$9 < M_s < 10, 0.5 < z < 1$	1.031 ± 0.053	0.903 ± 0.035	0.32	2.7
Elliptical - S0	1.303 ± 0.057	0.890 ± 0.053	1.90	2.1
Sa - Sc	1.069 ± 0.050	0.914 ± 0.039	1.38	2.2
Sd - Sdm	0.834 ± 0.035	0.914 ± 0.031	0.31	2.7
$2 < \text{NUV-R} < 3$	0.952 ± 0.041	0.828 ± 0.052	0.05	3.3
$1 < \text{NUV-R} < 2$	0.707 ± 0.040	0.931 ± 0.049	8.11	1.4
$0 < \text{NUV-R} < 1$	0.631 ± 0.053	0.910 ± 0.062	7.33	1.5

Table 3.2: Table of modelling results with values of HOD slope α and halo tilt γ_h . Rightmost columns show, assuming gaussian errors, the probability of no tilt in percent and the multiple of σ this corresponds to.

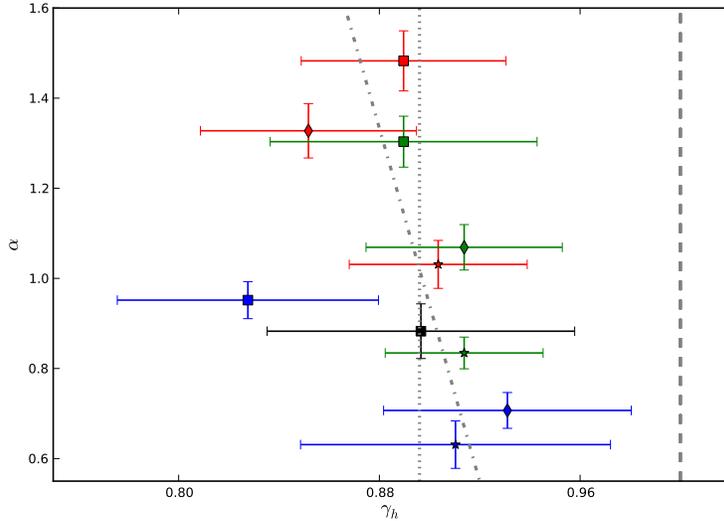


Figure 3.5: Collection of all points from fig. 3.4, including the result from the full galaxy sample (black point). The dot-dashed line shows the best fit line through the data, and we find the evolution to be just barely favored at the 5% confidence level compared to no evolution (dotted line).

3.3.3 Alternative dark matter profile

Another interpretation of the halo tilt is that the dark matter profile was wrong to begin with, ie. the NFW profile does not describe the dark matter profile correctly, so we see a discrepancy in the galaxy correlation function. Furthermore, if we look at figure 3.2, we see that tilting the 1-halo term overestimates the correlation function on intermediate scales ($r \sim 1$ Mpc/h).

This could lead us to consider a profile that is more curved than the NFW profile, but will still fall off rapidly at larger radii. Luckily, these are indeed features of the Einasto profile. This profile does seem to show better agreement with simulated dark matter haloes (see [40, 31]), so in this section we invent a preliminary model of the 1-halo term using this density profile.

In its general form, the profile reads:

$$\ln \left(\frac{\rho}{\rho_{-2}} \right) = - \left(\frac{2}{\alpha_e} \right) \left[\left(\frac{r}{r_{-2}} \right)^{\alpha_e} - 1 \right], \quad (3.29)$$

where r_{-2} is the radius at which the slope is -2 (isothermal), ρ_{-2} is the density at r_{-2} and α_e controls the curvature of the profile¹. Gao et al. [20] has

¹Note that we have added the e subscript to distinguish it from the α of the HOD

shown that this shape parameter can be connected to the halo mass through the overdensity parameter ν :

$$\alpha_e = 0.155 + 0.0095\nu^2. \quad (3.30)$$

The amplitude parameter ρ_{-2} is taken (arbitrarily) to be the same as the amplitude of the NFW model, as we are not interested in exact figures, only to see how the shape of the profile performs. This leaves us with one free parameter (r_{-2}) that we fit to the data along with the HOD slope α . The result is seen in figure 3.6.

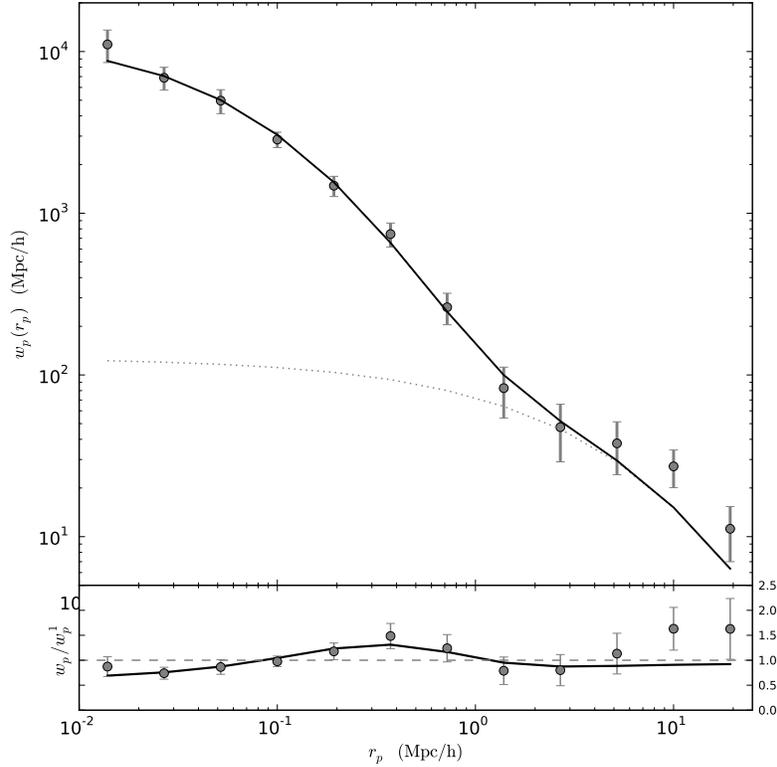


Figure 3.6: Same as fig. 3.2, but using the preliminary Einasto model with 2 free parameters. The bottom panel suggests that this model might be better than the tilted NFW model at intermediate scales. The likelihood ratio test indeed favors the Einasto model with a confidence of $\sim 16\%$ over the tilted model and $\sim 21\%$ over the untilted model.

We see clearly how this (rough) model using the Einasto profile in the 1-halo

term can improve the fit to the data compared to the normal and tilted NFW profiles. As the Einasto falls off more rapidly at larger scale, the correlation function at intermediate scales is not overestimated. In section 4.1.2 we discuss this result further.

At a reduced χ^2 of 0.12, the model might seem to be overfitting. But as we are only allowed to vary the amplitude and a single shape parameter, this seems unlikely, and leads us to believe that the error bars are overestimated. For reference, the reduced χ^2 for the untilted model, where we can only vary the amplitude, is 0.23.

Chapter 4

Discussion

We will in this last chapter go into some discussion about where the results of this project fit into the emerging view of galaxy evolution and the profiles of dark matter halos, as well as some speculations about which topics might be interesting to follow up on and develop further.

4.1 Results

4.1.1 Halo occupation distribution

The HOD is a very active research subject in the field of structure and galaxy evolution [55, 57, 27]. We have seen in section 3.1.2 how this, together with the mass function and halo profile, determines the correlation function that we can measure from galaxy surveys or simulations. We have shown in section 3.3.2 that the HOD parameters vary with the galaxy population, so that eg. the number of massive galaxies grows with halo mass (fig. 3.4) as expected from the theory of galaxy evolution in section 1.3.2.

However, the COSMOS cluster catalogue is not well suited to constrain the full profile of the occupation number because of the relatively narrow range of masses. This can be seen by comparing the mass histogram of COSMOS clusters (fig. 2.5) with a simulated HOD plot from Yan et al. [59] (fig. 4.1). Lower mass haloes with only one or a few galaxies are excluded from the detection method (section 2.1.3), as are the most massive super-structures. Halo finding algorithms can yield a large number of haloes across a greater mass range when used on large galaxy surveys, such as the Sloan Digital Sky Survey [60]. This is done recently in eg. Tinker et al. [55], Tinker et al. [57], Leauthaud et al. [27], and the large halo distribution puts better constraints on the HOD model.

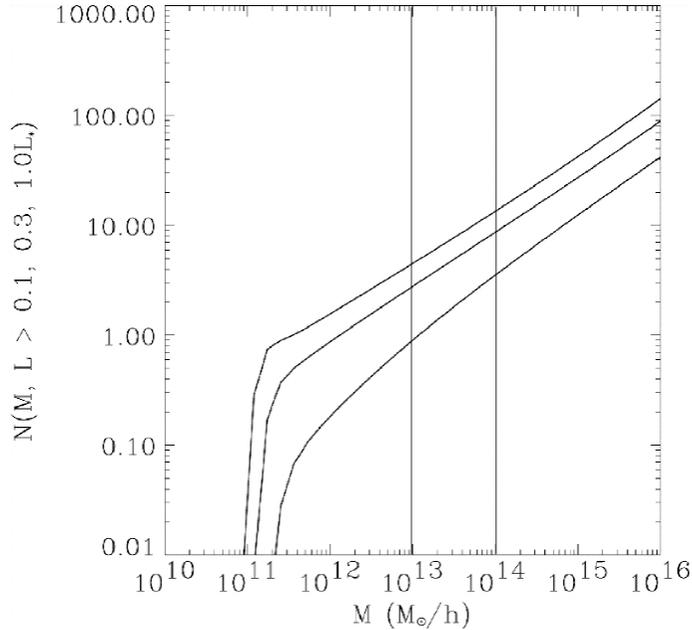


Figure 4.1: The HOD from simulations plotted for 3 different luminosity selections of galaxy clusters. Figure is taken from Yan et al. [59]. The vertical lines indicate the mass range of COSMOS clusters.

The goal of the research in the HOD is a better understanding of galaxy evolution, ultimately to the point where we can create a fully analytical model of the HOD that fit the observations from the large galaxy surveys. We have shown in this project that the HOD depend intimately on the properties of the galaxies in question, so any such analytical model must be able to account for this.

4.1.2 Dark matter profile

Another active research area of dark matter haloes is of course the density profile. For some years the NFW profile has been the de-facto profile when modelling dark matter haloes of all sizes, and much work is being done on the apparent near-universality that such systems seem to show [eg. 22]. We have briefly touched upon this topic when realizing that the discrepancy between galaxy and dark matter profiles might be explained by the assumption of NFW form in the 1-halo term, and subsequently introducing the Einasto profile that has been favoured in recent years by numerical simulations [40, 31].

In order to fully use the Einasto profile in the halo modelling, we must relate the mass of the halo to the profile properties, just as we in section 3.1.4 did for the NFW profile. In section 3.3.3 we created a preliminary model, but a more

formal treatment is required in order to promote the Einasto profile amongst halo modellers. Even with this rough model, however, we show that it might be worth considering this profile, especially if high precision modelling is required. In that spirit, a direct comparison between a number of dark matter profiles in the context of halo modelling would be very helpful.

4.1.3 Galaxy profile

In our question about relationship between the dark matter and galaxy profiles, there is a clear ambiguity between improving the dark matter profile (as discussed above) and introducing discrepancies between the dark matter and galaxy profile. Our initial choice was towards the later, which is why we introduced the halo tilt γ_h as a tilt for the entire 1-halo term in stead of just the dark matter profile ρ_{dm} . The net result for the profile shape is exactly as if we had tilted the dark matter profile itself, which is why the discussion in the previous section is still valid.

But going back to the original question, we could also say that the NFW profile might describe the dark matter, while the tilted or Einasto profiles describe the galaxies. Ie. the difference from dark matter to galaxy profile is the difference between the NFW and the best-fit Einasto profile. The reason that the former explanation might be preferable is of course that according to simulations, the Einasto profile might be a better fit to the dark matter profile as well, in which case the Einasto fit in figure 3.6 seems to suggest that the dark matter and galaxy profiles are actually very close.

If we look at it from the other angle, eg. using the galaxy transformations discussed in section 1.3.2, there seems to be many ways that the galaxy profile could differ from the dark matter density. If cluster galaxies undergo mergers, stripping and harassment, any displacements from the correct position (in the sense that they follow the dark matter profile) would have to be reverted quickly enough that it would not give us a signal when averaged over clusters. This doesn't seem like a tough constraint, but there is another process that could cause the galaxy profile to be altered, namely dynamical friction. As explained in section 1.3.2, dynamical friction causes massive galaxies to loose angular momentum to the environment, and effectively spirals the galaxy into the center of the cluster potential. The time it takes for a galaxy of mass M to move to the center the cluster because of dynamical friction can be approximated by [35]:

$$t_{\text{df}} \approx \frac{1.17}{\ln(M_h/M)} \frac{M_h}{M} \frac{1}{10H(z)}, \quad (4.1)$$

where M_h is the mass of the halo and $1/H(z)$ is the age of the Universe (Hubble time). This shows us that the galaxy must be very massive ($M_h/M > 15$) to spiral in from the edge of the halo in the age of the Universe. This tells us that it will happen very rarely, but also that it might be possible to detect in the galaxy profile for a sample of massive galaxies, perhaps at larger redshift where structures were less equilibrated. In the left panel of figure 3.4 however, where we vary exactly the galaxy mass and redshift, we see no clear evidence for this.

Figure 3.5, does actually indicates a slight evolution of the tilt with increasingly concentrated populations, but future studies in these regions might tell us more.

These novel investigations is also the subject of a forthcoming paper by Brink et al. (MNRAS), which is currently in its final stages.

4.2 Future developments

4.2.1 Improvements

The approach of this project has been to use only data in the constraining process, while the approach of others working with halo modelling often includes results from N-body simulations mimicking the survey from which the observations are obtained (eg. Tinker et al. [57]). When using simulations, they are able to improve the estimation of a simplified quantity in the model of the 2-halo term, namely the clustering bias b_c . In the derivation of the bias in section 1.3.1 we completely ignore higher order and scale-dependant effects, but there is evidence that this must be considered when trying to model correlation measurements to higher precision [8]. An improved model might use a bias expression from simulations but measuring this effect on real data is of course a goal in itself.

Another room for improvement might be in the semi-analytical relation between the halo mass and the parameters of the dark matter profile. In section 3.1.4 we derived the concentration and scaling parameters of the NFW profile from the cluster mass, but these relations are just approximations, motivated again by simulations, that might be improved. In a forthcoming paper by J. B. James et al., we investigate the relation between mass and concentration using the same data and measurements as in this project.

4.2.2 Opportunities

Because of the importance of the HOD in the halo modelling, it might be interesting to make direct measurements of the occupation number in COSMOS clusters for various cluster and galaxy populations. We still have to take care of survey selection effects that eg. cause us to find a higher fraction of massive galaxies at higher redshift simply because we do not see the less luminous galaxies. If properly handled, though, direct HOD measurements in COSMOS and other catalogues could improve the HOD models which in turn improves our understanding of galaxy evolution. One could also take a non-parametric approach and investigate how HOD distributions are scattered with galaxy properties without assuming any model. This would serve to give an unbiased view of how different galaxy populations have evolved in the dark matter haloes.

Continuing on the track of non-parametric modelling, another very important problem in cosmology is the estimation of the large-scale dark matter density field from galaxy surveys. This project shows clearly how different galaxies are distributed different according to the environment, so we might expect meth-

ods of estimating the large-scale structure to improve if more galaxy properties were considered. Improving such estimates could improve the estimation of cosmological parameter such as the matter density parameter Ω_m .

And when talking about cosmological parameter estimation, it is clear from recent papers by Tinker et al. [57], Leauthaud et al. [27] that halo modelling has the potential to improve the constraining power from galaxy surveys, one of only a handful of cosmological probes. Improving this precision, by eg. using the Einasto profile as argued in the previous sections, might have a positive effect on the parameter estimates.

4.3 Conclusions

In this project we have gone from the theoretical basis of structure formation to the measurement of the two-point cross-correlation function between clusters and galaxies in the COSMOS field. These measurements are analysed using the halo model approach, a statistical description of structure in the Universe, and the results of this modelling is discussed along with future prospects in this field.

The measurements are obtained by carefully handling the selection effects of the COSMOS survey, and we introduce the measurement for various galaxy populations based on the stellar mass, color and morphology of the galaxies. To the best of our knowledge, this has not been done with the COSMOS data using the cross-correlation with X-ray selected galaxy clusters, and therefore provides a unique opportunity to investigate the evolution of galaxies with respect to their large-scale environment.

We introduce the concept of the halo model and the how the correlation measurements can be modelled given expressions for the linear dark matter bias factor, the halo occupation distribution and the dark matter halo profile. A prime focus is on the potential discrepancy between the density profile of dark matter haloes and the galaxies residing herein. To that end, we introduce a novel parameter, the halo tilt, and investigate the evolution of this parameter with different galaxy populations. We find that there is strong evidence for the introduction of a tilt to the NFW profile, but that we cannot distinguish between the discrepancy between dark matter and galaxies and the potentially wrong choice of dark matter profile. Creating a preliminary model with the Einasto profile, also novel work, we show that the model is improved considerably. We acknowledge that a more rigorous future investigation could be important for the continued development of the halo model framework.

Lastly, the measurements and analysis of this project has been prepared in a paper, soon to be released. An outline of the code behind the project is included and we will work towards releasing this to the general public in the coming months.

Acknowledgements

A special thanks is needed for the primary supervisor of this project, Berian James. Almost before meeting, we were in the process of planning a two-months stay for me at the University of California Berkeley with which he is a joint fellow with the Dark Cosmology Centre. Needless to say, it was an amazing opportunity for me to work with expert researchers in the field of structure formation and be introduced to other exiting fields of astrophysical research. I am very grateful to the Dark Cosmology Centre and the Center for Time-Domain Informatics at UCB for funding this trip, it would not have been possible without their help.

Thanks to Berian and Steen Hansen for providing very valuable feedback in the finishing phases of this manuscript, and to the fellow master's students and crew at DARK for making it such a cool place.

Appendix A

The Standard Model of Cosmology

A.1 Observational basis

A.1.1 Hubble's law

A pivotal event in the history of cosmology, and physics in general, occurred when Hubble, among others, first presented findings of galaxies receding from Earth with velocities proportional to their distances:

$$v = H_0 d \tag{A.1}$$

, where H_0 is the *Hubble constant* of proportionality, measuring the rate at which the Universe is currently expanding. With measurements of distances to what had recently been discovered to be extra-galactic objects, Hubble showed a plot of this proportionality with large error bars due to large systematic errors in the distance measurements. His first estimate of the Hubble constant was also terribly wrong, but he nevertheless started a revolution in the scientific world view from a static Universe to an expanding one, and laid the foundations for the *Big Bang* theory that follows intuitively as the initial state of an expanding Universe. As we will see shortly, the Hubble constant is a very important concept in modern cosmology that influences the dynamics of the Universe and its contents.

A.1.2 Cosmic Microwave Background

As a great support for and at least as important as the discovery of an expanding Universe, is the discovery of the Cosmic Microwave Background radiation (CMB). The CMB was first seen more or less by accident by Penzias and Wilson in 1964 after being predicted by various independent researchers from around 1948 up to its discovery. Since then, many observations have been made from

telescopes and high-altitude balloons to measure the CMB signal across the sky, and today the CMB is the most important tool for constraining models of our Universe.

The most revolutionizing measurements came with the COBE satellite results published in 1992 [1] that showed remarkable agreement with predictions of anisotropies of the order $\sim 10^{-5}$ and a black-body spectrum with a characteristic temperature of ~ 2.7 . Recently, the WMAP satellite has greatly improved the resolution of these measurements and the next generation, the Planck mission, is already releasing their first data. Figures A.1 and A.2 shows the WMAP map of the CMB across the sky and the extremely well fitted black-body spectrum, respectively.

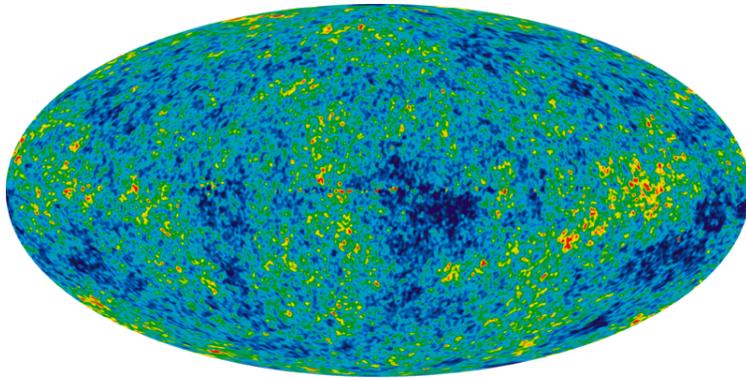


Figure A.1: The WMAP CMB temperature map relative to the mean temperature map corrected for the dipole anisotropy due to the Sun's motion relative to the CMB and with foreground contamination from the Galaxy removed. The anisotropies shown here are of the order $\sim 10^{-5}$.

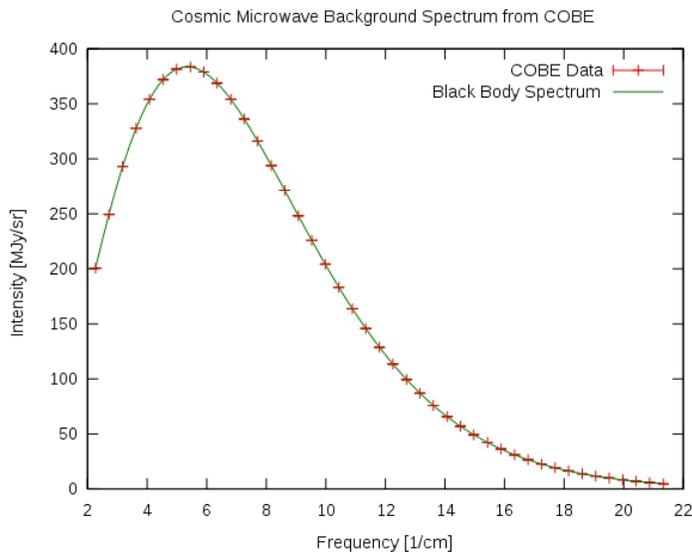


Figure A.2: The fit of the black-body spectrum (solid line) with a temperature of $T = 2.725$ to the WMAP data (crosses). The error bars are hidden behind the solid line.

A.2 Friedmann-Robertson-Walker spacetime

The framework in which we describe the Universe on large scales is *General relativity*. This theory, relating the energy contents and the 4-dimensional spacetime geometry of the Universe to what we know as gravity, was first theorized by Einstein [16] and has been confirmed to high accuracies in the course of the last century, lately by the final results of the *Gravity Probe B* satellite [18]. This relation is set out in the *Einstein field equation* (EFE)¹:

$$R_{\mu\nu} = 8\pi G(T_{\mu\nu} - \frac{1}{2}Tg_{\mu\nu}), \quad (\text{A.2})$$

where $R_{\mu\nu}$ is the *Ricci curvature tensor* describing the curvature of spacetime, G is Newton's gravitational constant, $T_{\mu\nu}$ is the *Stress-Energy tensor* describing the energy inventory (such as mass and radiation) and finally $g_{\mu\nu}$ is the metric describing the geometry of spacetime. Although this equation looks quite simple, it is most definitely not. The Ricci tensor is related to the metric in a non-linear way. But the basic idea, that I can give you a description of the contents and you can then calculate the geometry and curvature of the Universe, remains very simple and powerful, and we can usually simplify the calculation of the EFE using physical knowledge of the system we wish to describe.

¹Note that we work in units of $c = 1$

Even though its hard to believe when we look at the sky at night, the Universe is assumed to be very much the same everywhere. This idea of course leads to simplifications of the description of the geometry, encoded in the metric, and the spacetime of such a Universe is said to be isotropic and homogeneous: the Universe looks the same in all directions and is the same everywhere. Observations of the real world, however, suggest that this simplification might be a bit of a stretch; the Universe is expanding, and thus cannot be static in time. The solution to this is that the Universe is isotropic and homogeneous in space but evolving in time. The metric in such cosmologies, the description the geometry of spacetime, can be written:

$$ds^2 = -dt^2 + R^2(t)d\sigma^2, \quad (\text{A.3})$$

where $R(t)$ is the time-dependant scale-factor scaling the symmetric space metric $d\sigma^2$ in time. By assuming spherical symmetry we can rewrite $d\sigma^2$ in terms of the curvature parameter κ and the angular metric $d\Omega^2$:

$$ds^2 = -dt^2 + a^2(t) \left[\frac{dr^2}{1 - \kappa r^2} + r^2 d\Omega^2 \right], \quad (\text{A.4})$$

where $a(t)$ is the dimensionless scale factor. This is called the *Robertson-Walker* metric, an important result for working with cosmology in the context of General Relativity. With $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$, we only need a description of the energy-momentum in $T_{\mu\nu}$ and its trace T to solve the Einstein equation. For this we make another simplification, namely that everything in the Universe can be described as a perfect fluid with density ρ and pressure p . The energy-momentum tensor and its trace can be written in terms of these properties, the momentum U_μ and the metric:

$$T_{\mu\nu} = (\rho + p)U_\mu U_\nu + pg_{\mu\nu} \quad T = -\rho + 3p. \quad (\text{A.5})$$

This calls for a relationship between density and pressure, often called the *equation of state*:

$$p = \omega\rho, \quad (\text{A.6})$$

which we will define for the various components of the Universe in the next section. If we consider conservation of energy, we obtain a relation between the density of the universe and the scale factor that will be useful later:

$$\rho \propto a^{-3(1+\omega)}. \quad (\text{A.7})$$

The EFE with FRW metric and perfect fluid stress-energy tensor yields the *Friedmann equations*:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{\kappa}{a^2} \quad (\text{A.8})$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p). \quad (\text{A.9})$$

The first of these is often simply referred to as the *Friedmann equation*, which we can simplify by identifying the *Hubble parameter* as $H = \dot{a}/a$ and define the density to be a sum of the density contributions from the different types of energy content. The Friedmann equation (A.8) now reads:

$$H^2 = \frac{8\pi G}{3} \sum_i \rho_i. \quad (\text{A.10})$$

Notice that we have transformed the curvature term into an energy density analog:

$$\rho_{\text{cu}} = -3\kappa/(8\pi G a^2). \quad (\text{A.11})$$

This merely makes the equations simpler and does not mean that curvature is a kind of energy in the Universe. We don't believe that the universe is completely homogeneous and isotropic except on the largest scales. Indeed, locally the Universe is highly inhomogeneous, and this project is concerned with understanding these variations in detail. In the later sections of this chapter we will work our way down the scale-ladder, but when describing the global dynamics of the Universe, these approximations are well justified by e.g. CMB measurements.

A.3 Cosmic inventory

Having set up the Friedmann equations that govern the evolution of the Universe, it's time to put in some content. Motivated by equation A.7, we assume that the evolution of the energy density of any component can be written as a powerlaw in the scale factor a :

$$\rho_i = \rho_{i,0} a^{-n_i}, \quad (\text{A.12})$$

where $\rho_{i,0}$ is the energy density of component i at present time and n_i is the power law slope related to ω by $w_i = \frac{1}{3}n_i - 1$. The problem has then been reduced to finding values for n or w for a Universe where the components behave as perfect (smooth) fluids.

Matter, in the cosmological context, is the non-relativistic pressureless component of the energy distribution. With this realization we can guess how the energy density of matter will evolve with the scale factor, namely $\rho_m \propto a^{-3}$. This is because as the universe expands with a , the volume will grow as a^3 , and the density $\rho = \text{energy}/\text{volume}$ will decrease the same amount assuming the amount of energy is conserved. Note that "matter" here stands for both the baryonic matter that make up the visible part of the Universe and a cold dark matter component that does not interact electromagnetically.

The argument is the same for radiation, except in that case the expansion of the Universe also increases the wavelength of the photons in addition to spreading it across a larger volume. Therefore, the evolution of the energy density of radiation is $\rho_r \propto a^{-4}$. The assumption that no photons are created

is of course not true, as we see a lot of stars in a lot of galaxies emit a lot of photons. However, if we compare the estimated total energy density of starlight to that of the early Universe, we find that it is only around $\sim 10\%$ [46, chapter 5.1].

In addition to the more obvious components of energy, general relativity permits the existence of an energy density embedded in the vacuum of the Universe itself, in the form of a *vacuum energy* ρ_Λ , in cosmology often referred to a *Dark Energy*. This was first introduced by Einstein in order to obtain a solution to his equations that would allow the Universe to be static, as was rightfully believed at that time. Einstein later called this his greatest blunder, but in modern cosmology this term has found renewed interest as the *cosmological constant* because it can explain the accelerated expansion of the Universe that we are observing. The energy density of the cosmological constant is by definition constant, so there is no evolution in a and the value at all times is just the present value $\rho_\Lambda = \rho_{\Lambda,0}$.

Lastly, the evolution of the fictional curvature energy density, along with its amplitude, is given by its definition in equation A.11. Below we present a table of the values of ω and n from the motivations above:

Type	ω_i	n_i
Matter	0	3
Radiation	1/3	4
Curvature	-1/3	2
Constant	-1	0

Before we present the final version of the Friedmann equation, we will introduce the dimensionless density parameter at present time $\Omega_i = \rho_{i,0}/\rho_c$, where the current *critical density* is $\rho_c = 3H_0^2/(8\pi G)$. From equations A.10 and 1.2 we have:

$$H^2 = H_0^2[\Omega_m a^{-3} + \Omega_r a^{-4} + \Omega_\Lambda + \Omega_c a^{-2}]. \quad (\text{A.13})$$

This equation is the standard equation of cosmology in the sense that specifying H_0 and the Ω_i s determines the evolution of the scale factor of the Universe. By probing the Universe on the largest scales, this is exactly what cosmologists are trying to do today. Several experiments, from the measurement of the CMB anisotropies and the distribution of matter to the accelerated recession of supernova galaxies, have given us the values of these parameters to increasing accuracy and in figure A.3 we plot the evolution of the scale factor a with cosmic time for the best-fit values. Cosmic time, the time t of the Robertson-Walker metric (A.4), can be calculated from a by recalling that $Ha = da/dt$, multiplying both sides of A.13 with a^2 and integrating over dt :

$$H_0 t = \int_0^a [\Omega_m a^{-3} + \Omega_r a^{-4} + \Omega_\Lambda + \Omega_c a^{-2}]^{-1/2} da \quad (\text{A.14})$$

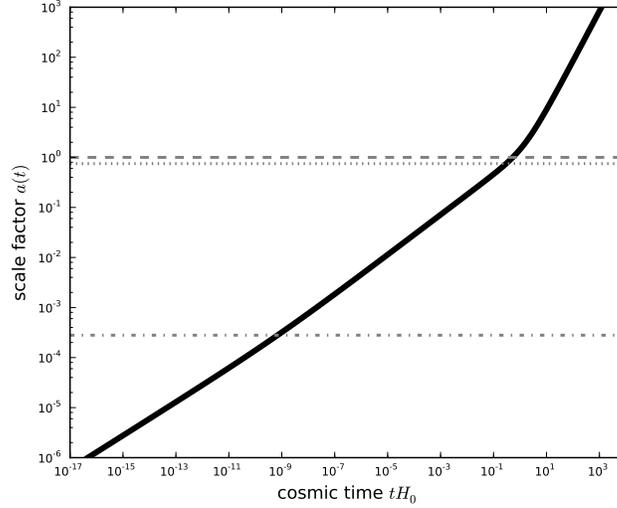


Figure A.3: The evolution of the scale factor a with cosmic time t (solid line). Values of a for present (dashed line), the matter- Λ equality (dotted) and matter-radiation equality (dash-dotted) are also plotted. Values for the parameters are $\Omega_r = 5 \cdot 10^{-5}$, $\Omega_m = 0.3$, $\Omega_\Lambda = 1.0 - \Omega_r - \Omega_m \approx 0.7$ for a flat Universe.

We see in figure A.3 that the Universe with the parameter values that we have currently measured will continue to expand in the future, driven by the cosmological constant.

Appendix B

N-body simulations

We are not using specific results from simulations in the analysis parts of this project, but many of the insights of structure formation are given or motivated by simulations. So an account of structure formation would not be complete without an outline of the most important techniques used in numerical N-body simulations, the subject of this section.

B.1 N-body dynamics

Modern simulations contain a lot of different physics to explain systems in more and more detail, but one thing sits at the center of any simulation: gravity. The basic problem that simulations try to solve is that of calculating the force on a particle due to the N-1 other particles, hence the name N-body simulations. We can also think of it as calculating the gravity between pairs of particles and summing up the net force due to Newton's law in a Universe expanding with the scale factor a [43]:

$$\frac{d\mathbf{x}}{dt} = \frac{\mathbf{v}}{a} \tag{B.1}$$

$$\frac{d\mathbf{v}}{dt} + H\mathbf{v} = \mathbf{g} \tag{B.2}$$

$$\nabla \cdot \mathbf{g} = -4\pi G a [\rho(\mathbf{x}, t) - \bar{\rho}(t)], \tag{B.3}$$

where \mathbf{x} is the position and \mathbf{v} is the velocity of the particle and a and $H = \dot{a}/a$ are given by the cosmological model. If we define the time variable $s = \int a^{-2} dt$, the first two equations simplify to a single second order differential equation:

$$\frac{d^2\mathbf{x}}{ds^2} = a\mathbf{g}. \tag{B.4}$$

Calculating the force between every particle pair is still highly ineffective, but a slew of optimizations have been developed. The first and perhaps most

obvious is that of the hierarchal tree algorithm [4] where the field is split into a recursive hierarchy of cells so that cells of size l that are a distance d from \mathbf{x} , the point where we want to calculate $g(\mathbf{x})$, is treated as a single particle if the angle is less than a small value θ :

$$\frac{l}{d} < \theta, \quad (\text{B.5})$$

where we can then vary θ to determine the precision of the calculation versus computing time.

Another popular method is the so-called particle-mesh (PM) method [34] where the gravitational potential is confined to a grid of a chosen resolution. The density field ρ is estimated at each grid by counting up the mass from nearby particles using an appropriate interpolation method. The density field is then transformed to Fourier space where it is faster to solve Poisson's equation for the gravitational potential:

$$\tilde{\phi}(\mathbf{k}, t) = -4\pi G a^2 \frac{\tilde{\rho}(\mathbf{k}, t)}{k^2} \quad (\text{B.6})$$

and back again to real space where the gravitational acceleration is calculated from eq. B.3 and in turn interpolated back on the particles using the same interpolation method as for the mass assignment.

Variations of these methods exists, like an adaptive mesh algorithm that splits the space into smaller grids when needed or even falls back to pair calculation in dense environments.

B.2 *Gadget*

One of the most popular codes for cosmological simulations is *Gadget* and more recently *Gadget-2* by Springel [51]. This code is a highly parallelized code that uses the PM method for calculation of long range interactions and the tree algorithm on short range where the PM method is most inaccurate due to the interpolation on a grid. It is the *Gadget-2* code that has been used in the Millennium simulation, and as such is highly proven.

To simulate gas dynamics, *Gadget-2* uses the smooth particle hydrodynamics (SPH) technique that solves the fluid equation for baryons in an expanding universe by smoothing the mass of baryons out with a variable smoothing length. This method fits well in with the tree and PM algorithms, as the same techniques can be applied to find particle neighbors included in the smoothing window.

Gadget-2 is a highly portable code that uses the *GSL* and *FFTW* libraries. It is parallelized using the *MPI* message parsing library, making it possible to run simulations on your laptop or the biggest supercomputers. The details of the parallelization methods are complicated, but the basic idea is to divide the 3D space of particles into a 1D vector and splitting this into an appropriate amount of samples that are then distributed onto the processors.

Of course, many other codes exist that uses other or similar techniques, but the basic arguments are mostly the same. See Aarseth [1] for a recent review of N-body/SPH techniques.

B.3 Analysis

The output of a simulation is of course the full field of particle positions and velocities, but it is of course impossible to compare this directly to theory or observations, since the theory of structure formation cannot predict the detailed structure of individual haloes and because observations are subject to selection effects of the survey and other uncertainties. We must turn to statistical comparisons of the particles, velocities or the interpolated density field.

In previous sections in this chapter we have talked about the power spectrum a great deal, and measuring the power spectrum of from the particles in the simulation output is of course an obvious choice for comparison. A related statistics is that of the two-point correlation function that is the Fourier transform of the power spectrum. Many other informative statistics exists as reviewed in Bertschinger [6].

Another analysis that is often employed on simulation output is halo finding. In section 1.3.1 we introduced the dark matter halo as the primary dark matter structures in nonlinear theory, and we would of course like to test both theory and simulations by comparing, say, the mass function from theory and from simulations. But the definition of a halo is not very well settled upon and so the result of the comparison might vary with the choice of halo-finding algorithm.

Appendix C

PyHalo software library

For this project we have developed a Python code that supports statistical measurements on large catalogues of galaxies and clusters, such as the measurement we present in chapter 2, as well as modelling of these measurements using the theory of the halo model presented in chapter 3.

We will first give a brief motivational speech about why Python is cool and then move on to describe the code we have developed. In the coming months, we hope to untangle the code from the specific problems of this project and make it publicly available. It can serve as a platform for research in cosmology or learning about numerical cosmology in general.

Many of the functions in this library has been tested using the *Cosmology Routine Library in Fortran* by Eiichiro Komatsu ¹.

C.1 Motivation

There are many reasons for using Python in scientific research, but here we will highlight two important points. A lot of research is about investigating things quickly, to get a handle on the validity of an idea, and not spending a large amount of time developing robust code. Python is a very easy language for doing this, as you can get good results with a small amount of code. Because it is very simple, you don't have to worry about uninteresting infrastructure, such as reading files or dealing with strings, something that can be a pain in more low-level languages such as C or Fortran. Of course this comes with a trade-off: speed. Low level, compiled languages are much faster to run code, but the time for writing the code is usually much higher. Luckily, if Python simply is not fast enough for you, there are several ways of improving this, as we will show in section C.5. The general idea is that you should write your code in python to get results fast. If anything is too slow, isolate it in separate function that you can then optimize.

¹<http://gyudon.as.utexas.edu/~komatsu/CRL/>

The other merit that we will highlight about Python is that it is publicly available and runs on any (and every) computer you might have. This sounds like a soft value, but it is a very important aspect. Coupled with the Py-lab environment², Python provides an environment that can be compared to those of Matlab and IDL, but these environments are proprietary and expensive. Granted, the Python environment might not feel as much as a single product, and it can be hard to navigate the different technologies as a newcomer, but this is also a strength. Because Python and its libraries are open, anyone can contribute and chances are that someone solved your problem. With increasing popularity, the PyLab environment is improving accordingly.

C.2 Cosmology

The first class of the library deals with the cosmology that you will work in. It defined the most important cosmological parameters and provides functions to work with the cosmology.

When you initialize it without arguments, you simply get the default cosmology based on the WMAP7 data. You can change any of the parameters afterwards and recalculate derived parameters. Below is an example of initializing the `Cosmology` class. Note that we assume that the `pylab` module is loaded in the variable `p` and that the everything from the `PyHalo` module is imported.

```
lcdm = Cosmology()
print lcdm.om_0 #> 0.272
lcdm.om_c = 0.25
lcdm.update()
print lcdm.om_0 #> 0.3
```

Here we initialize the cosmology, change the cold dark matter density parameter Ω_c , update the derived parameters and see the new value for the total matter density Ω_0 .

Given a scale as an initialization argument, the cosmology class will calculate other functions of interest, such as the power spectrum and two-point correlation function of matter on these scales. It automatically calculates the primordial power spectrum and evolves it to the present time using the transfer function of section 1.2. To plot the base power spectrum, you simply issue:

```
scales = p.logspace(-2,2,1000)
lcdm = Cosmology(scales)
p.loglog(lcdm.k, lcdm.ps)
```

To get the power spectrum at different redshifts, you can scale this base spectrum with the `growth(z)` function.

²A conglomerate of various Python libraries: NumPy, Scipy, Matplotlib, iPython and others.

C.3 Halo

The next class we will introduce is the `Halo` class that represents dark matter haloes. These objects require at least three arguments on initialization: the cosmology, the mass and the redshift of the halo. Given these, a halo object will be returned that includes values for important parameters of the halo, such as the radius, the overdensity parameter ν , the bias factor for objects at that redshift and mass and the dark matter density profile of the halo. So, for instance, to get the density profile of a dark matter halo:

```
scales = p.logspace(-2,2,1000)
lcdm = Cosmology(scales)
h = Halo(lcdm, 1.0e13, 1.0)
p.loglog(scales, h.profile)
```

This plots the density profile of the halo from the halo mass and redshift by scaling the power spectrum with the growth factor and going through the calculations of section 3.1.4 to get the NFW parameters that determine the profile.

The `Halo` class also has functions i.e. `to` to return the correlation function of subhaloes and galaxies based on the halo occupation distribution of section 3.1.3 and the correlation function of the linear dark matter field at that redshift and scale.

C.4 Modelling

When dealing with data you often want to test different models in order to describe the data best, constrain different parameters and generally just compare models. The PyHalo framework makes this easy by providing a base `Model` class that can be subclassed to define a testable model. For the moment it only works for the measurements that we are modelling in this projects, but plans are in order for generalizing this to other measurements.

Lets make a simple example for illustration. Suppose you have some measurements of the two-point correlation function of galaxies in a large survey, and you wish, ignorant as you are, to fit these models to a simple power law. We create a subclass of the `Model` class and define a single `model` method that takes a set of parameters, in this case the scale and slope of the power law, and returns a power law that can be fit to the data:

```
class MyPowerLaw(Model):
    def model(self, params)
        scale, slope = params
        return scale*self.r**slope
```

where we have used the scales `r`.

The data can now be fit to the model by calling the `fit` method with data, errors and initial guess in your model object initialized with `r`:

```
scales = p.logspace(-2,2,1000)
model = MyPowerLaw({'r': scales})
data, errors = make_measurements()
params = model.fit(data, errors, [1.0e4, -1.5])
print params # prints best fit parameters
```

In general you'd want to implement better models (eg. based on the `HALO` class), but it's as easy as that. For the moment, only a simple optimization algorithm is used, but more sophisticated ones are being implemented. We also support likelihood calculations that can be used to get uncertainties of the fitted parameters.

C.5 Optimizations

The last section of this appendix is devoted to the optimization of Python code in the context of scientific computing. Python has the advantage of being easy to write, but in demanding scientific applications it is not always efficient enough. In this project we have combined two different methods for speeding up the critical calculations.

C.5.1 Fortranization

The first thing one should do is to write the Python code as efficient as possible and use the NumPy methods if working with arrays or matrices. But if that is not enough, there are ways to run more low level code from Python. In this case we use Fortran, as a matter of preference, but we could just as well have used C or C++.

For running Fortran code from Python we use the F2Py library that is a part of the SciPy package. This command takes any Fortran file with subroutines and creates an importable Python module that contains these functions. In that way, you can focus on rewriting your critical calculations in Fortran in stead of the very time-consuming task of rewriting the entire code.

In our case, the critical function was the calculation of pair separations needed to calculate the correlation function. For the smallest combination of $\sim 1e5 \times 3e2$ pairs, this would take ~ 20 minutes in Python, and for the largest datasets, this could not be done. In the analogues Fortran code, the time it takes to do the pair counts of the small dataset is barely noticeable, while the big calculation takes ~ 5 minutes.

Nothing has to change in the Python code other than you have to import the Fortran module as you would any other Python module and call the function from it.

C.5.2 Parallelization

While we could reduce the time of the pair count calculation to around 5 minutes, this was still along time when we have to make a bunch of measurements in a row.

Most modern computers come with multi-core CPUs, so they can perform multiple calculations at once. Python programs does not do this by themselves, however, and the developer has to write code that supports the fairly complicated paradigm of parallel computing. This is true of my laptop, a 2.5 year old MacBook Pro, so speeding the calculation up another factor of 2 should in principle be possible.

Python comes with support for making Threads, small lines of execution that seem to run simultaneously. They don't however, because they are restricted by something called the Global Interpreter Lock that locks the Python interpreter to a single processor core. A way to go around this problem is by spawning several Python interpreters and communicate between them, and there is of course a Python library in newer version of Python that makes this easy: the `multiprocessing` module.

Using this module, it is a matter of a few lines of extra code to spawn any number of processes that executes any python function on any data. The return value of this function is a map of the results from each of the processes, and it is easy to reduce this to the usual result by eg. summing over these. The following code snippet shows an example of calculating a simple function across 5 processors:

```
calculation = lambda x: x**2
x = p.linspace(0,10,1000)
# Serial version
print sum(calculation(x))
# Parallel version
from multiprocessing import Pool
procs = Pool(5)
result = procs.map(calculation, p.array_split(x,5))
print sum(result)
```

So in this project we use Python to distribute data to a bunch of Python processes that run Fortran code and improve the time of calculation from forever to a couple of minutes. In the old way, a Fortran program could be parallelized by using MPI (or similar), writing much more code and thinking very differently about your program. Using the method presented here, you can very quickly get to a fully competitive code without growing old as you type.

Bibliography

- [1] S. Aarseth. *Gravitational N-body simulations*. Cambridge Univ Pr, 2003.
- [2] I.K. Baldry, M.L. Balogh, R.G. Bower, K. Glazebrook, R.C. Nichol, SP Bamford, and T. Budavari. Galaxy bimodality versus stellar mass and environment. *Monthly Notices of the Royal Astronomical Society*, 373(2): 469–483, 2006.
- [3] J.M. Bardeen, JR Bond, N. Kaiser, and AS Szalay. The statistics of peaks of gaussian random fields. *The Astrophysical Journal*, 304:15–61, 1986.
- [4] J. Barnes and P. Hut. A hierarchical $O(N \log N)$ force-calculation algorithm. *nature*, 324:4, 1986.
- [5] A.A. Berlind and D.H. Weinberg. The halo occupation distribution: Toward an empirical determination of the relation between galaxies and mass. *The Astrophysical Journal*, 575:587, 2002.
- [6] E. Bertschinger. Simulations of structure formation in the universe. *Annual Review of Astronomy and Astrophysics*, 36(1):599–654, 1998.
- [7] J. Binney and S. Tremaine. Galactic dynamics. *Galactic Dynamics: Second Edition, by James Binney and Scott Tremaine. ISBN 978-0-691-13026-2 (HB). Published by Princeton University Press, Princeton, NJ USA, 2008.*, 1, 2008.
- [8] M. Blanton, R. Cen, J.P. Ostriker, and M.A. Strauss. The physical origin of scale-dependent bias in cosmological simulations. *The Astrophysical Journal*, 522:590, 1999.
- [9] P. Capak, H. Aussel, M. Ajiki, HJ McCracken, B. Mobasher, N. Scoville, P. Shopbell, Y. Taniguchi, D. Thompson, S. Tribiano, et al. The first release cosmos optical and near-ir data and catalog. *The Astrophysical Journal Supplement Series*, 172:99, 2007.
- [10] S.M. Carroll, W.H. Press, E.L. Turner, et al. The cosmological constant. *Annual Review of Astronomy and Astrophysics*, 30(1):499–542, 1992.

- [11] M. Colless et al. The 2dF Galaxy Redshift Survey: spectra and redshifts. *MNRAS*, 328:1039–1063, December 2001. doi: 10.1046/j.1365-8711.2001.04902.x.
- [12] A. Cooray and R. Sheth. Halo models of large scale structure. *Phys. Rep.*, 372:1–129, December 2002. doi: 10.1016/S0370-1573(02)00276-4.
- [13] S. Dodelson. *Modern cosmology*. Academic Pr, 2003.
- [14] A. Dressler. Galaxy morphology in rich clusters-implications for the formation and evolution of galaxies. *The Astrophysical Journal*, 236:351–365, 1980.
- [15] B. Efron. Bootstrap methods: another look at the jackknife. *The annals of Statistics*, 7(1):1–26, 1979. ISSN 0090-5364.
- [16] A. Einstein. The foundation of the general theory of relativity. *Annalen Phys*, 49(769-822):31, 1916.
- [17] D.J. Eisenstein and W. Hu. Power spectra for cold dark matter and its variants. *The Astrophysical Journal*, 511:5, 1999.
- [18] C W F Everitt. Gravity probe b: Final results of a space experiment to test general relativity. *Physical Review Letters*, 2011. URL <http://prl.aps.org/accepted/L/ea070Y8dQ491d22a28828c95f660a57ac82e7d8c0>.
- [19] A. Finoguenov et al. The XMM-Newton Wide-Field Survey in the COSMOS Field: Statistical Properties of Clusters of Galaxies. *ApJS*, 172:182–195, September 2007. doi: 10.1086/516577.
- [20] L. Gao, J. F. Navarro, S. Cole, C. S. Frenk, S. D. M. White, V. Springel, A. Jenkins, and A. F. Neto. The redshift dependence of the structure of massive Λ cold dark matter haloes. *MNRAS*, 387:536–544, June 2008. doi: 10.1111/j.1365-2966.2008.13277.x.
- [21] G. Hasinger et al. The XMM-Newton Wide-Field Survey in the COSMOS Field. I. Survey Description. *ApJS*, 172:29–37, September 2007. doi: 10.1086/516576.
- [22] J. Hjorth and L. L. R. Williams. Statistical Mechanics of Collisionless Orbits. I. Origin of Central Cusps in Dark-matter Halos. *ApJ*, 722:851–855, October 2010. doi: 10.1088/0004-637X/722/1/851.
- [23] O. Ilbert, M. Salvato, P. Capak, E. Le Floch, H. Aussel, H. J. McCracken, S. Arnouts, B. Mobasher, D. B. Sanders, N. Scoville, and Y. Taniguchi. Photometric Redshifts and Stellar Mass Assembly in the 2-deg² COSMOS Field. In T. Kodama, T. Yamada, & K. Aoki, editor, *Panoramic Views of Galaxy Formation and Evolution*, volume 399 of *Astronomical Society of the Pacific Conference Series*, pages 169–+, October 2008.

- [24] G. Kauffmann, S.D.M. White, T.M. Heckman, B. Ménard, J. Brinchmann, S. Charlot, C. Tremonti, and J. Brinkmann. The environmental dependence of the relations between stellar mass, structure, star formation and nuclear activity in galaxies. *Arxiv preprint astro-ph/0402030*, 2004.
- [25] S. D. Landy and A. S. Szalay. Bias and variance of angular correlation functions. *ApJ*, 412:64–71, July 1993. doi: 10.1086/172900.
- [26] A. Leauthaud, A. Finoguenov, J.P. Kneib, J.E. Taylor, R. Massey, J. Rhodes, O. Ilbert, K. Bundy, J. Tinker, M.R. George, et al. A weak lensing study of x-ray groups in the cosmos survey: Form and evolution of the mass-luminosity relation. *The Astrophysical Journal*, 709:97, 2010.
- [27] A. Leauthaud, J. Tinker, P.S. Behroozi, M.T. Busha, and R. Wechsler. A theoretical framework for combining techniques that probe the link between galaxies and dark matter. *Arxiv preprint arXiv:1103.2077*, 2011.
- [28] S. J. Lilly et al. zCOSMOS: A Large VLT/VIMOS Redshift Survey Covering $0 < z < 3$ in the COSMOS Field. *ApJS*, 172:70–85, September 2007. doi: 10.1086/516589.
- [29] R.Y. Liu and K. Singh. Moving blocks jackknife and bootstrap capture weak dependence. *Exploring the limits of bootstrap*, 225:248, 1992.
- [30] J. M. Loh. A Valid and Fast Spatial Bootstrap for Correlation Functions. *ApJ*, 681:726–734, July 2008. doi: 10.1086/588631.
- [31] D. Merritt, A. W. Graham, B. Moore, J. Diemand, and B. Terzić. Empirical Models for Dark Matter Halos. I. Nonparametric Construction of Density Profiles and Comparison with Parametric Models. *AJ*, 132:2685–2700, December 2006. doi: 10.1086/508988.
- [32] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, et al. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087, 1953.
- [33] R.G. Miller. The jackknife-a review. *Biometrika*, 61(1):1, 1974. ISSN 0006-3444.
- [34] R.H. Miller. Gravitational n-body calculation in a discrete phase space. *Journal of Computational Physics*, 6(3):449–472, 1970.
- [35] H. Mo, F. Van den Bosch, and S. White. *Galaxy formation and evolution*. Cambridge Univ Pr, 2010.
- [36] H. J. Mo and S. D. M. White. An analytic model for the spatial clustering of dark matter haloes. *MNRAS*, 282:347–361, September 1996.

- [37] B. Mobasher, P. Capak, NZ Scoville, T. Dahlen, M. Salvato, H. Aussel, DJ Thompson, R. Feldmann, L. Tasca, O. Lefevre, et al. Photometric redshifts of galaxies in cosmos. *The Astrophysical Journal Supplement Series*, 172:117, 2007.
- [38] B. Moore, S. Ghigna, F. Governato, G. Lake, T. Quinn, J. Stadel, and P. Tozzi. Dark matter substructure within galactic halos. *The Astrophysical Journal Letters*, 524:L19, 1999.
- [39] J. F. Navarro, C. S. Frenk, and S. D. M. White. A Universal Density Profile from Hierarchical Clustering. *ApJ*, 490:493–+, December 1997. doi: 10.1086/304888.
- [40] J. F. Navarro, E. Hayashi, C. Power, A. R. Jenkins, C. S. Frenk, S. D. M. White, V. Springel, J. Stadel, and T. R. Quinn. The inner structure of Λ CDM haloes - III. Universality and asymptotic slopes. *MNRAS*, 349: 1039–1051, April 2004. doi: 10.1111/j.1365-2966.2004.07586.x.
- [41] J.A. Nelder and R. Mead. A simplex method for function minimization. *The computer journal*, 7(4):308, 1965. ISSN 0010-4620.
- [42] J. A. Peacock and R. E. Smith. Halo occupation numbers and galaxy bias. *MNRAS*, 318:1144–1156, November 2000. doi: 10.1046/j.1365-8711.2000.03779.x.
- [43] P.J.E. Peebles. *The large-scale structure of the universe*. Princeton Univ Pr, 1980.
- [44] W. H. Press and P. Schechter. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *ApJ*, 187:425–438, February 1974. doi: 10.1086/152650.
- [45] D. S. Reed, R. Bower, C. S. Frenk, A. Jenkins, and T. Theuns. The halo mass function from the dark ages through the present day. *MNRAS*, 374: 2–15, January 2007. doi: 10.1111/j.1365-2966.2006.11204.x.
- [46] B.S. Ryden. *Introduction to cosmology*. Addison-Wesley, 2003.
- [47] E. Schinnerer, V. Smolčić, CL Carilli, M. Bondi, P. Ciliegi, K. Jahnke, NZ Scoville, H. Aussel, F. Bertoldi, AW Blain, et al. The vla-cosmos survey. ii. source catalog of the large project. *The Astrophysical Journal Supplement Series*, 172:46, 2007.
- [48] N. Scoville, R. G. Abraham, H. Aussel, J. E. Barnes, A. Benson, A. W. Blain, D. Calzetti, A. Comastri, P. Capak, et al. COSMOS: Hubble Space Telescope Observations. *ApJS*, 172:38–45, September 2007. doi: 10.1086/516580.
- [49] N. Scoville et al. *ApJS*, 172:1–8, September 2007. doi: 10.1086/516585.

- [50] R. K. Sheth and G. Tormen. Large-scale bias and the peak background split. *MNRAS*, 308:119–126, September 1999. doi: 10.1046/j.1365-8711.1999.02692.x.
- [51] V. Springel. The cosmological simulation code gadget-2. *Monthly Notices of the Royal Astronomical Society*, 364(4):1105–1134, 2005.
- [52] V. Springel, S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, D. Croton, J. Helly, J. A. Peacock, S. Cole, P. Thomas, H. Couchman, A. Evrard, J. Colberg, and F. Pearce. Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 435:629–636, June 2005. doi: 10.1038/nature03597.
- [53] V. Springel, J. Wang, M. Vogelsberger, A. Ludlow, A. Jenkins, A. Helmi, J.F. Navarro, C.S. Frenk, and S.D.M. White. The aquarius project: the subhaloes of galactic haloes. *Monthly Notices of the Royal Astronomical Society*, 391(4):1685–1711, 2008.
- [54] Y. Taniguchi, N. Scoville, T. Murayama, D. B. Sanders, et al. The Cosmic Evolution Survey (COSMOS): Subaru Observations of the HST Cosmos Field. *ApJS*, 172:9–28, September 2007. doi: 10.1086/516596.
- [55] J. Tinker, A.V. Kravtsov, A. Klypin, K. Abazajian, M. Warren, G. Yepes, S. Gottl
"ober, and D.E. Holz. Toward a halo mass function for precision cosmology: the limits of universality. *The Astrophysical Journal*, 688:709, 2008.
- [56] J. L. Tinker, C. Conroy, P. Norberg, S. G. Patiri, D. H. Weinberg, and M. S. Warren. Void Statistics in Large Galaxy Redshift Surveys: Does Halo Occupation of Field Galaxies Depend on Environment? *ApJ*, 686: 53–71, October 2008. doi: 10.1086/589983.
- [57] J. L. Tinker, E. S. Sheldon, R. H. Wechsler, M. R. Becker, E. Rozo, Y. Zu, D. H. Weinberg, I. Zehavi, M. Blanton, M. Busha, and B. P. Koester. Cosmological Constraints from Galaxy Clustering and the Mass-to-Number Ratio of Galaxy Clusters. *ArXiv e-prints*, April 2011.
- [58] S.M. Weinmann, F.C. Van Den Bosch, X. Yang, and HJ Mo. Properties of galaxy groups in the sloan digital sky survey-i. the dependence of colour, star formation and morphology on halo mass. *Monthly Notices of the Royal Astronomical Society*, 366(1):2–28, 2006.
- [59] R. Yan, M. White, and A.L. Coil. Mock catalogs for the deep2 redshift survey. *The Astrophysical Journal*, 607:739, 2004.
- [60] D. G. York and other. The Sloan Digital Sky Survey: Technical Summary. *AJ*, 120:1579–1587, September 2000. doi: 10.1086/301513.