Galaxy Clustering as a Probe for Galaxy Evolution in Simulations

Nicole Thomas[1,2]

Under the Supervision of Professor Romeel Davé[1,2,3] and Co-supervision of Professor Roy Maartens[1]

[1] Department of Physics and Astronomy, University of the Western Cape, 7535, South Africa

[2] South African Astronomical Observatory, Observatory, Cape Town, 7925, South Africa

[3] Royal Observatory, Edinburgh, EH9 3HJ, Scotland, United Kingdom

keywords: Galaxies, galaxy Formation, Galaxy Evolution, Simulations, Clustering, Two Point Correlation Function, Bias, Mass Function, Luminosity Function, Star Formation, HI

A thesis submitted in fulfilment of the requirements for the degree of Magister Scientiae in the Department of Physics and Astronomy, University of the Western Cape.

17 November 2017

Abstract

Studying clustering on small scales (<10Mpc) over a large span of redshifts allows us to connect galaxies to underlying cosmic large-scale structure, and thereby provide constraints on the physical processes that drive galaxy evolution. Relatedly, studying the relative bias of galaxies and their halo occupancy quantifies how the underlying dark matter distribution is traced by baryons in galaxies. Comparing model predictions to current and future multi-wavelength galaxy surveys, enables a greater understanding of how galaxy formation processes impact the relationship between galaxies and dark matter.

In this thesis I study the clustering and bias of galaxies in cosmological galaxy formation simulations. To conduct this study, I use MUFASA – a state-of-the-art suite of cosmological hydrodynamical simulations (Davé, Thompson, and Hopkins, 2016). The number of galaxies within a probed parameter bin decreases with increasing redshift. I thus examine clustering and bias over a redshift range from $z = 0 - 2.5$ and $z = 0 - 3$ using MUFASA's fiducial $(50\mathrm{h}^{-1}\mathrm{Mpc})^3$ volume, which has 512^3 dark matter particles and 512^3 gas elements.

I find that for the full sample of galaxies in the MUFASA simulations, the amplitude of clustering decreases with redshift until $z = 2$, after which the amplitude increases with redshift. The dark matter clustering amplitude decreases with redshift, and thus with this we see an increase in bias, which is the ratio of clustering strength of baryons to that of the dark matter distribution, with increasing redshift. At a given redshift, we see an increase in clustering amplitude with stellar mass, $M_\star$, as well as with increasing brightness in r-band luminosity, M_r. However, we see a decrease in clustering amplitude with specific star formation rate, $sSFR$, as well as brightness in the Near-UV luminosities, M_{NUV}. We observe no clustering dependence on HI mass, M_{HI}, at a given redshift.

As a function of redshift, we observe a decrease in clustering amplitude for a given $M_\star$ bin. We do not see any obvious trend with M_r bin, however for $sSFR$, M_{NUV}, and M_{HI} bins we see an increase in clustering amplitude as a function of redshift. There is a general increase in bias as at a given redshift across all mentioned parameters. There seems to be no bias evolution with $M_\star$ bin however, but for all other parameters, we see a general increasing trend with redshift.

These results suggest that galaxies that are more red, quiescent, and gas-poor cluster more strongly than blue, star-forming, gas-rich galaxies. Since higher-mass dark matter halos cluster more strongly, this indicates that these halos are more likely to host groups of passive, gas-poor galaxies. Since the detailed manner in which passive and star-forming galaxies populate dark matter halos depends on detailed galaxy formation processes such as feedback from star formation and black holes, future comparisons to upcoming surveys such as MIGHTEE and VIDEO

will provide an important path towards constraining these uncertain feedback processes.

Declaration

I declare that *Galaxy Clustering as a probe for Galaxy Evolution in Simulations* is my own work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Full name **Nicole Lynn Thomas**

Date **17 November 2017**

Signed: 

Contents

List of Figures

1 Introduction

Large scale galaxy surveys and cosmological simulations have shown us the filamentary distribution of galaxies in the Universe. Due to gravity, galaxies and baryonic matter in these filamentary structures trace the underlying dark matter distribution. Within this large-scale structure we observe dense clusters of galaxies at the nodes of these filaments, with less massive groups and isolated galaxies following the filamentary arms. Different galaxy types trace this filamentary structure in different ways – we see older and redder galaxies making up the high density clustered regions, while the younger, bluer, and star-forming galaxies tend to make up the filamentary arms. This relationship may also change with redshift, as at early epochs dense regions can host very rapidly star-forming galaxies.

How baryons trace this underlying dark matter can be quantified using the galaxy Two Point Correlation Function (TPCF), which measures the excess probability of finding a pair of galaxies in a given radial bin, and comparing to that of the dark matter TPCF which measures the excess probability of finding a pair of dark matter particles in the same radial bin. Observational clustering studies have used a Halo Occupancy Distribution (HOD) formalism to predict the underlying dark matter distribution. HODs are defined as the average number of galaxies residing in a halo of a given mass, thus quantifying how galaxies populate halos as a function of M_{halo}. In cosmological hydrodynamic simulations that directly predict the evolution of galaxies, we have access to both the distribution of galaxies and dark matter, so we can directly compute the TPCF and HOD of galaxies, halos, and dark matter. Also, we can compute the bias of the galaxies relative to the dark matter. This enables predictions that can be tested against large-area galaxy survey data.

On large scales, the galaxy TPCF and their clustering properties of galaxies with respect to the underlying dark matter field provides important constraints on cosmology, particularly through Baryon Acoustic Oscillations (BAO) which are matter fluctuations in baryons at a scale of ∼150 Mpc owing to sound waves imprinted during the recombination era (Eisenstein et al., 2005). Using BAO-scale clustering measurements, cosmologists have been able to put constraints on cosmological parameters which are important to constrain current cosmological models. These parameters include the baryonic content, Ω_b, of the Universe, the amplitude of the linear power spectrum at 8 h^{-1}Mpc, σ_8, and others.

On smaller scales, studying the clustering dependence on galaxies and their intrinsic properties over time helps us understand how galaxies evolve within their environments and their host dark matter halo (Zehavi et al., 2011). While it is known that galaxy properties are correlated with environment, for instance large passive galaxies tend to live in denser regions, the detailed physical processes that drive such relationships are not fully understood. Typically, they

are related to feedback energy released by processes such as supernovae and black hole accretion, which can impact the growth and colour of galaxies. Hence by studying clustering on small scales, one can constrain these poorly understood processes of galaxy assembly.

1.1 Dark Matter and Dark Matter Halos

Only $\sim 4\%$ of the Universe as we know it today is made up of baryonic matter that we are able to directly observe with telescopes. $\sim 70\%$ is made up of Dark Energy, responsible for the accelerating expansion of the observable Universe. The remaining $\sim 25\%$ is in dark matter, observed only by it's gravitational influence on nearby baryonic particles.

The first first indirect observation of dark matter was made when Zwicky (1933) studied the motion of galaxies in the Coma cluster. He found that the orbital velocities of these galaxies were too high for the mass estimated from the brightness and number of the galaxies contained and thus inferred some invisible mass, or "dunkle materie" (dark matter). Thereafter, Rubin, Ford, and Thonnard (1980) studied the rotation curves of spiral galaxies and noticed that the amount of mass in the stellar and gaseous matter they were able to observe, did not make up for the rotation velocities and angular momentum of the spiral arms at increasing radii. Today we see the effect of dark matter not only in the rotation curves of spiral galaxies but also in gravitational lensing, the evolution of large-scale structure, as well as in other astrophysical phenomena. Owing to gravitational instability, after time they cluster enough to become dark matter halos, typically defined as a structure with mean density 200 times that of the critical density of the Universe. These halos are host to numerous galaxies - including one large central galaxy and often numerous smaller satellite galaxies. Larger dark matter halos usually host a larger number of galaxies. We can quantify this in Halo Occupancy Distributions (HODs) which are shown later on in this thesis. Because dark matter halos begin small and grow hierarchically by merging into larger structures, we expect to the see a decrease in the clustering of dark matter as we go back in time.

1.2 Galaxy Evolution

The physical processes that drive galaxy formation and evolution are still a significant puzzle in astrophysics. Many of these uncertainties lie in the nature of the mechanisms by which energy is returned to the galaxy by means of star and black hole formation that self-regulate galaxy growth (Davé, Thompson, and Hopkins, 2016). These are what is referred to as feedback processes. In blue star-forming galaxies, stellar feedback from supernovae and young stars is the primary regulator of galaxy growth. In larger galaxies, the primary feedback mechanism is due to black holes. Black hole feedback occurs mainly in massive galaxies that have large supermassive black holes, and release so much energy as to shut off, or quench, star formation. This results in the largest

galaxies living in the largest halos being red owing to a lack of young massive stars.

Many feedback processes may be triggered by the evolution of the underlying halo distribution. For instance, one hypothesis for quenching is that it becomes particularly important at a specific halo mass scale around $\sim 10^{12}\mathrm{M}_{\odot}$, a scale which may evolve with redshift. The Equilibrium Model, by constraining its parameters to match the observed evolution of the galaxy population, predicts a quenching mass that evolves with redshift as

$$M_q = (0.96 + 0.48z) \times 10^{12} M_{\odot} \quad (1)$$

(Davé, Rafieferantsoa, and Thompson, 2017). Alternatively, quenching may occur during the mergers of galaxies as driven by the mergers of their underlying dark matter halos (Hopkins et al., 2008). Hence feedback processes can impact the connection of galaxies to their host halos.

It is in this way that galaxy formation processes are related to clustering. For instance, a Universe with no or minimal black hole feedback would yield significantly more blue galaxies occupying larger dark matter halos. This would result in a large clustering amplitude for blue galaxies. Conversely, a Universe with only strong feedback we would expect much fewer blue galaxies in halos and a high clustering of red galaxies. Hence depending on the clustering strength of different galaxy types, we can constrain the strength of feedback processes that drive the growth of galaxies and their relationship to the underlying dark matter distribution. Moreover, we can provide inputs to cosmological studies that need to understand how galaxies of a given type trace halos in large-scale structure.

1.3 Galaxy Formation Simulations

In this thesis, I will study clustering using cosmological hydrodynamic simulations. In simulations such as Mufasa that we use here, a random cosmological volume is populated with a number of dark matter and gas particles at a very early time based on observed CMB fluctuations. This volume is then evolved forward in time accounting for physical processes including gravity and hydrodynamics via linear perturbation theory. The dark matter particles are evolved only by gravity since they do not interact with electromagnetic radiation. Gas is evolved with the many additional physical processes related to baryons. The end result is a prediction of the distribution of dark matter and galaxies within the modeled volume, including various observationally-accessible properties of the galaxies such as colour and HI content.

These simulations include the physical processes required to form realistic galaxies, such as star formation, supernova feedback, and quenching feedback - a still poorly understood phenomenon. It is long known that including only gravity and star formation results in the "overcooling" problem (White and

Rees, 1978) in which too many baryons form into stars. Hence feedback processes are needed to suppress star formation in order to make realistic galaxies. These simulations make use of feedback models that heuristically represent the energetic feedback from star formation and black hole accretion, since cosmological simulations cannot resolve the very small scales on which such processes occur. In this way, dark matter structure formation drives the evolution of stars and gas in galaxies, and the environment in which galaxies reside.

Cosmological simulations have played a significant role in helping to understand many of the physical processes that are still uncertain. Many advances in these fields have been made over the past decade due to the improvement of computing power and the accuracy of the physics included in the simulations. With this, simulations are now able to reproduce primary features of the observed galaxy population across cosmic time. Given this, such simulations provide a way to accurately study the clustering of galaxies, and the way in which different galaxies with different properties trace the filamentary structure of the cosmic web.

1.4 Observations of Clustering

Many observational efforts have characterised galaxy clustering both in the nearby Universe and beyond. Here we describe some such efforts whose results are relevant to this thesis, in which clustering is measured based on a specific selection of galaxy properties that we approximately mimic in our simulated galaxy samples.

Zehavi et al. (2005) and Zehavi et al. (2011) respectively use a volume limited sample of 200 000 galaxies over 2500 deg^2 extracted from the Sloan Digital Sky Survey (SDSS) and a sample extracted from the Complete SDSS Redshift Survey sample of galaxies - a sample of 700 000 galaxies covering 8000 deg^2 extending up to $z = 0.25$, to show that bright r-band galaxies have a higher clustering amplitude than that of fainter r-band galaxies. The flux-limited galaxy samples are then split into two subsamples, that of red and blue where it is found that red galaxies have a higher clustering amplitude than that of the blue galaxies of the same flux magnitude. It is also found that when the full galaxy sample is split into red and blue subsamples, within each subsample, the clustering amplitude increases with r-band brightness. Zehavi et al. (2011) also finds an increasing bias factor with increasing brightness.

Law-Smith and Eisenstein (2017) measured the colour and stellar mass dependence on clustering in the Baryon Oscillation Spectroscopic Survey component of SDSS(SDSS-III BOSS) at redshifts $0.6 < z < 0.65$ by cross correlating 66 657 spectroscopic galaxies to that of 6.6 million faint photometric galaxies. It is found that red galaxies live in denser environments than that of blue galaxies, and specifically that red galaxies cluster almost twice as much as blue

galaxies at low radii. When holding stellar mass as a fixed parameter, a clear differentiation between red and blue galaxies is found, implying that clustering is not solely determined by stellar mass. When holding colour fixed, it is found that clustering increases with stellar mass.

Papastergis et al. (2013) uses a sample of 6000 galaxies from the Arecibo Legacy Fast ALFA (ALFALFA) 21cm survey to measure the clustering properties of HI-selected galaxies. No convincing evidence is found for dependence of clustering for galaxy HI mass over $M_{HI} \approx 10^{8.5} - 10^{10.5} M_{\odot}$. The ALFALFA sample is compared to that of optically selected samples from SDSS and it is found that the HI-selected galaxies cluster more weakly than that of the relatively optically faint galaxies when there is no colour selection applied. When there is a colour selection applied, the correlation of blue galaxies is found to be indistinguishable from that of the HI-selected galaxies. SDSS with red colours are found to cluster more strongly than HI-selected galaxies.

Coil et al. (2016) studied the dependence of galaxy clustering on stellar mass and sSFR with data from the PRIMUS and DEEP2 galaxy redshift survey for redshifts $0.2 < z < 1.2$. 100 000 spectroscopic galaxies redshifts covering a total of 7.2 deg^2 is used to find that the dependence of clustering as a function of sSFR is as strong as the dependence on stellar mass. At a given sSFR bin, and for the limited mass range probed ($10^{10.5} < M_{\star} < 10^{11}$ $M_{\odot}$), there is no significant mass dependence. For the star-forming population at a given mass, galaxies above the main sequence of star formation with higher sSFR are less clustered than that of those below the main sequence of star formation with low sSFR. Within the quiescent population at a given stellar mass, galaxies with higher sSFR are less clustered than those of lower sSFR. It is shown that the clustering amplitude of galaxies increases with increasing stellar mass and decreasing sSFR, concluding that galaxies evolve not only along the main sequence, but also across it, before galaxies become quiescent. These results imply that the stellar mass to halo mass relation depends on sSFR.

Shi et al. (2016) tested the reliability of mapping the real space distribution in the SDSS DR7 for $0.01 < z < 0.12$ by using an ensemble of mock catalogues, focusing on the two point correlation function dependence on luminosity, and finding that there exists deviation from single power law and transitioning to clear 1-halo and 2-halo term dependencies.

van Daalen et al. (2016) introduced methods to allow the observed clustering of galaxies to be used along with observed luminosity and mass functions to constrain galaxy formation physics. By using large semi-analytic simulations, they had shown how the projected galaxy two point correlation function can be estimated to $\lesssim 10\%$ using a very small subsample of halo merger trees. The values found are consistent with what was previously found for the appropriate parameters. For parameters that primarily affect spatial distributions, values have been found with significantly improved constraints on galaxy formation,

which may lead to significant constraints on cosmological parameters.

These results demonstrate that the TPCF provides a robustly observable measurement of the clustering of galaxies, and thus can be used to obtain information about the physical processes which induce the evolution of galaxies and well as their environments. It can also be used to identify the relationship these galaxies have with their host dark matter halos and the underlying dark matter field.

These and other studies on galaxy clustering have told us that more massive galaxies cluster more strongly than smaller ones. The clustering amplitude of galaxies increases with increasing brightness in the r-band but decrease with sSFR. Also, red galaxies cluster almost twice as strongly as blue galaxies at low radii. There is no obvious trend in clustering amplitude for the HI content of galaxies. By comparing these observations to the galaxy samples from MUFASA, we can test whether MUFASA is producing and evolving galaxies that are viable, and thus whether MUFASA can provide robust predictions regarding the evolution of galaxy clustering in a wider variety of tracers.

1.5 Outline of Thesis

In this thesis I examine the clustering and bias of galaxies as a function of galaxy properties and redshift using MUFASA – a state-of-the-art suite of cosmological hydrodynamical simulations. I'll be studying the clustering and bias properties of galaxies as a function of stellar mass $M_\star$, specific star formation rates $sSFR$ which is the star formation rate divided by $M_\star$, r-band luminosity M_r, near-UV luminosity M_{NUV}, and HI mass M_{HI}. I will focus on how we can use this information to make predictions for future multi-wavelength surveys of which the primary science focus is galaxy evolution.

I start by outlining the simulations and analysis methods I have followed to produce the results in §2. In §3, I will describe various properties of galaxies and make comparisons to a range of galaxy property observations that show that MUFASA provides a viable sample of galaxies with which to study clustering. In §4 I describe results from the clustering analysis as a function of galaxy property and redshift. In §5 I summarise the key results and outline plans for future work.

2 Methodologies

2.1 Mufasa

Mufasa (Davé, Thompson, and Hopkins, 2016; Davé et al., 2017; Davé, Rafieferantsoa, and Thompson, 2017) is a suite of cosmological hydrodynamic simulations which use the GIZMO meshless finite mass (MFM) code. In MFM, the fluid in the simulation is evolved by means of a Riemann solver in a manner such that mass is conserved within each fluid element, and thus each fluid element can be regarded as a "particle" which we will refer to as a gas element. In Mufasa, stars are formed following a molecular gas-based star formation prescription through an approximate solver of H_2 formation that depends on the metallicity and the column density of the gas. The minimum density required for star formation is $n_H \geq 0.13cm^{-3}$ which we hereafter refer to as "ISM gas". Given an H_2 fraction, star formation follows a Schmidt (1959) Law, namely

$$\frac{dM_\star}{dt} = \epsilon_\star \frac{\rho f_{H_2}}{t_{dyn}} \tag{2}$$

where ρ is the gas density, f_{H_2} is the molecular fraction of a given element and $t_{dyn} = \frac{1}{\sqrt{G\rho}}$ is the local dynamic time. We set the star formation efficiency parameter, $\epsilon_\star = 0.02$ in accordance with observations by Kennicutt (1998). We assume a Chabrier (2003) initial mass function (IMF) which seems to be the most consistent with observed IMFs. With increasing gas density in the simulations, the Jeans mass (i.e. the mass above which gravity causes the collapse of interstellar clouds to form stars) can become smaller than that of the resolved mass of the simulation. If this happens and cooling proceeds, fragmentation which should occur on small scales is bottlenecked on large scales and produces artificial clumping. To account for this, artificial pressurisation of the ISM is implemented in such a way that the Jeans mass is always resolved above a density threshold, n_{th}.

Mufasa includes radiative cooling from primordial and metal elements using the GRACKLE package. GRACKLE-2.1 (Bryan et al., 2014) is a chemical and cooling library used to evolve primordial and metal line cooling over a cooling timescale. Non-radiative cooling is first applied over the first time step and thereafter GRACKLE is used to isochorically apply radiative cooling integrated on the cooling timescale. Mufasa tracks the evolution of 11 elements namely, H, He, and 9 metals: C, O, N, Ne, Mg, Si, S, Ca, and Fe.

Mufasa uses a sub-grid model to heuristically represent galactic outflows owing to supernova feedback. Outflow scalings in Mufasa are taken from FIRE simulations (Hopkins et al., 2014) which are much higher-resolution simulations that include supernovae feedback, radiation from massive stars, and stellar winds to self-consistently drive gas from galaxies. The outflow rate scales tightly with the stellar mass of the galaxy and is independent of redshift

with the best fit relation of the mass outflow rate at $0.25R_{vir}$ as

$$\eta = 3.55\Big(\frac{M_\star}{10^{10}M_\odot}\Big)^{-0.351} \tag{3}$$

The wind speed, v_w, is likewise taken from scalings from the FIRE simulations, normalised somewhat higher than the original prediction and including a boost factor to account for the fact that FIRE measured their velocities at a quarter of the virial radius:

$$v_w = 2\big(\frac{v_c}{200}\big)^{0.12}v_c + \Delta v_{0.25} \tag{4}$$

where $\Delta v_{0.25}$ is an extra velocity kick corresponding to the gravitational potential difference between the wind launch radius and $0.25\mathrm{R}_{vir}$.

To quench massive galaxies, a "maintenance mode" source of energetic feedback is implemented. Following the model in Gabor and Davé (2015), this is done by heating all the gas in massive halos except the gas that is self-shielded. By doing this most of the neutral and molecular hydrogen in these halos are unaffected, ensuring that cold gaseous components are only directly affected by physical processes such as stripping rather than the quenching prescription.

The results obtained in this thesis are based on a $(50\ \mathrm{h}^{-1}\mathrm{Mpc})^3$ volume which produces a population of ∼6600 resolved galaxies by $z = 0$ and reproduces the galaxy stellar mass functions from $z = 0 - 4$ to $\lesssim 1.2\sigma$ in cosmic variance. This volume is represented with 512^3 dark matter particles, each with mass resolution $\mathrm{m}_{dark} = 9.6 \times 10^7\mathrm{M}_\odot$, and 512^3 gas fluid elements with mass resolution $\mathrm{m}_{gas} = 1.82 \times 10^7\mathrm{M}_\odot$. Galaxies are identified as ≥ 32 star particle (or equivalently gas element) masses, resulting in a stellar mass resolution limit for galaxies of $\mathrm{m}_{gal} = 5.8 \times 10^8\mathrm{M}_\odot$. Mufasa assumes a cosmology consistent with Planck Collaboration et al. (2016) full likelihood constraints: $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $\Omega_b = 0.048$, $\mathrm{H}_0 = 68$ km s^{-1} Mpc^{-1}, $\sigma_8 = 0.82$, and $n_s = 0.97$ and each run starts at $z = 249$. 135 snapshots are outputted down to $z = 0$, but here we mostly consider the snapshots at $z = 0 - 3$.

2.2 Caesar-YT

We generate galaxy catalogs from Mufasa using the Caesar package. Caesar (Thompson, 2016) builds on the yt (NumFOCUS, 2012) package thus enabling it to read many different simulation types, and attach meaningful units to object attributes. It is a python framework for analyzing the outputs from the snapshots produced by cosmological simulations. The main goal is to identify halos and galaxies within the simulation, then store information about them in an easy to access manner. To identify galaxies and dark matter halos, Caesar uses a 3D Friends-of-Friends (FoF) algorithm to link neighbouring particles within a given linking length and defines these particles as groups. For Mufasa, a group made up of ≥ 32 star particles will be identified as a galaxy.

2.3 Loser

We use Line Of Sight Extinction by Ray-tracing (Loser; Davé, Rafieferantsoa, and Thompson 2017) to return a catalogue of luminosities in several bands for simulated galaxies in both extincted, including extinction by dust, and non-extincted circumstances, together with the redshift and cosmological model parameters. Single stellar populations (SSP) spectra are interpolated to the star's age and metallicity and generated using the FSPS (van Dokkum and Conroy, 2010) library. We assume a baseline extinction of $A_V = 0.1$ to avoid extinction-free lines of sight and employ a Milky Way extinction law (Cardelli, Clayton, and Mathis, 1989). Each individual stellar spectrum is extincted based on the integrated dust column to that star, and can be computed for any of 6 user-selected directions, +/-x,y,z. Colours are then determined for galaxies by summing the stellar spectra and convolving with bandpass response functions for a set of user-selected bands. For my analysis, I make use of the SDSS r-band filter centered at 6230Åas well as the GALEX NUV filter centered at 2309Å.

2.4 Clustering Measures

2.4.1 Two Point Correlation Function (TPCF)

For clustering measures, the common method observers use to calculate the TPCF is the Landy and Szalay (1993) estimator which accounts for bias and redshift distortions at large spatial scales. However, in this paper the method of clustering measure is the spatial two point correlation function, ξ_g, calculated as

$$\xi_g = \frac{DD - RR}{RR} \tag{5}$$

according to Peebles and Hauser (1974) where, for each radial bin, DD is the number of galaxy pairs and RR is the analytically calculated number of expected pairs in a homogeneously distributed sample such that if $\rho \equiv \frac{N}{V}$, then

$$RR = \rho \times \frac{4\pi}{3}[(r + \delta r)^3 - r^3] \tag{6}$$

with N = number of galaxies, V = volume of the simulation cube, r = radial separation, δr = radial bin width. This estimator is chosen to eliminate the need to generate random catalogues and analytically, and thus more accurately, compute random pairs and spare computation time. The results are comparable to the Landy and Szalay (1993) estimator in three dimensional geometries as per figure 1 in Vargas-Magaña et al. (2013).

2.4.2 Relative Bias

Relative biases compare the fluctuations in the spatial distributions of baryons to that of the underlying dark matter. It may be calculated with the following

equation:

$$b^2 = \frac{\xi_g}{\xi_{DM}} \tag{7}$$

Bias factors are presented at a separation distance of r$\sim$ 2.67 h^{-1}Mpc in accordance with Zehavi et al. (2011). This choice is made because it is outside of the extremely non-linear regime, as well as because the observations are well measured there (Zehavi et al., 2005).

2.5 Error Analysis

All error estimations provided in this thesis have been produced using jackknife errors, in which the simulation cube is split into sub-octants and analysis is done for each thereof. Jackknife resampling is commonly used to estimate the standard deviation in a smooth statistic to account for sampling bias. However, in non-smooth statistics, this method does not lead to a consistent estimate of the error (Babu, 2010). Because the volume probed by MUFASA is not sensitive to large non-smooth biases, jackknife sampling is a viable estimator of error.

For the clustering analysis, each sub-octant is swapped with another, and the analysis is redone for each pair of sub-octants. After completing this analysis for all octant pairs, we are left with $N = \frac{n^2-n}{2} + 1$ TPCFs (which includes the original TPCF) from which we can estimate the mean and variance at each separation distance as

$$\bar{\xi(r)}_i = \frac{1}{N}\sum_{j=1}^{N} \xi(r_j)_i \tag{8}$$

$$\sigma_i^2 = \frac{1}{N}\sum_{j=1}^{N} \left(\xi(r_j)_i - \bar{\xi(r)}_i\right)^2, \tag{9}$$

where i is the index of the separation distance, and j is the index of the TPCF once the jackknife resampling has been completed.

3 Large Scale Galaxy and Halo Properties

We begin by discussing MUFASA predictions for galaxies and how they populate halos, comparing to observations as appropriate. The goal is to show that MUFASA provides a viable description of the galaxy-halo connection, and thus provides a suitable platform to examine how galaxies cluster in relation to how the dark matter clusters. In the following subsections, we consider the halo occupancy distribution, mass fractions in stars and neutral hydrogen, the stellar mass, HI mass, and star formation rate functions, and the r-band and NUV luminosity function.

3.1 Halo Occupancy Distributions

The Halo Occupancy Distribution (HOD) is defined as the average number of galaxies residing in a dark matter halo with a given halo mass, M_{halo} which, by observers, is estimated from the stellar mass according to analytical determinations such as those by Moster, Naab, and White (2013). As discussed by Berlind et al. (2003), it describes how galaxies populate halos, and thus provides a way to connect the galaxy distribution to the underlying dark matter halo distribution. The HOD can be measured in observations by constructing a halo catalog, and counting the number of galaxies within each halo (Zheng, Coil, and Zehavi, 2007). In both models and observations, the HOD is typically found to have a cutoff at low halo masses set by the mass limit, and a rising power law at high halo masses with a slope near unity (Berlind et al., 2003).

We compute the HOD in MUFASA by taking each FOF halo identified using CAESAR, counting the number of galaxies within it, and taking the average value for all halos within bins of halo mass. We only consider galaxies with $M_\star > 5.8 \times 10^8 \mathrm{M}_\odot$, as this is our galaxy stellar mass resolution limit. The central galaxy is defined as the most massive galaxy within the halo and thus each halo has at most one central galaxy, but can have many satellite galaxies.

Figure 1, left panel, shows the HOD from the 50 h^{-1}Mpc MUFASA simulation at $z = 0, 1, 2$ (red, green, blue, respectively). The central galaxies' HODs are indicated by the dashed lines, while the satellites' HODs are shown by the dotted lines. Above $M_{halo} >\sim 10^{11.5}\mathrm{M}_\odot$, essentially all halos contain a central galaxy above our stellar mass resolution limit. The rising power law is driven by the satellites, which increase in number with halo mass. Fitting a power-law slope for $M_{halo} > 10^{12}\mathrm{M}_\odot$, we obtain a slope of 0.99, which is similar to the slope of 1.06 ± 0.08 obtained from SDSS data by Zheng, Coil, and Zehavi (2007). Comparing the three redshifts, we see that there is very little redshift evolution in the predicted HOD in MUFASA – galaxies populate halos in essentially the same way independent of cosmic epoch. This is also consistent with the minimal evolution seen by comparing SDSS at $z \sim 0$ with DEEP2 data at $z \sim 1$ as analysed by Zheng, Coil, and Zehavi (2007).

Figure 1, right panel, compares the HOD from our 50h^{-1}Mpc simulation with the 25h^{-1}Mpc MUFASA simulation, in order to assess resolution convergence of the HOD. Here, we only use galaxies with $M_\star > 5.8 \times 10^8 M_\odot$ in both simulations, even though the 25h^{-1}Mpc simulation resolves galaxies up to 8 times smaller, in order to directly isolate the effects of resolution. This significantly reduces the number of galaxies in the sample thus resulting in the statistical fluctuations being substantially larger particularly at z=2. However, this figure shows that the HOD is generally well-converged with numerical resolution, indicating that galaxies above our stellar mass limit are robust in terms of how they populate halos.

In summary, MUFASA produces an HOD with a near-unity power law slope, as observed, with little redshift evolution. The predicted HOD is broadly consistent with observations, showing that galaxies are populating halos in MUFASA in a manner well approximating that in the real Universe. This indicates that MUFASA provides a feasible platform to study predictions for galaxy clustering, at least for clustering within halos or so-called "one-halo clustering".

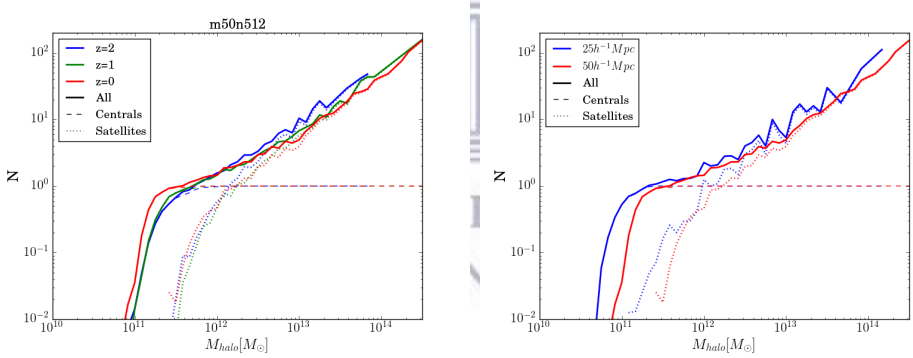

Figure 1: Halo Occupancy Distributions for 25- and 50 h^{-1}Mpc MUFASA runs. The left figures shows the fiducial (50h^{-1}Mpc)^3 run with a galaxy stellar mass resolution of $5.8 \times 10^8 M_\odot$. The figure on the right shows a comparison between the fiducial 50h^{-1}Mpc box and the higher resolution 25h^{-1}Mpc box scaled to the same mass resolution of that of the 50h^{-1}Mpc box. Errorbars have been omitted for clarity.

3.2 The Galaxy-Halo Connection

To explore the connection between galaxies and halos further, in this section we examine the galaxy-halo connection which measures how galaxies populate their host dark matter halos and how they affect each others evolution. Although the processes that govern galaxy evolution are complex, there are observational indications that galaxy properties have tight correlations including masses, luminosities, and other dynamical properties. There are also growing indications that there is a tight relation between galaxies and the dark matter

halos in which they form. There is a tight relation between the stellar mass of a galaxy and the mass of it's host dark matter halo, referred to as the stellar mass - halo mass relation (SMHM). Coming to understand these relations we can learn more about not only about the evolution of galaxies, but also the evolution and nature of the underlying dark matter distribution in the Universe.

To test the galaxy halo connection in Mufasa, we consider several galaxy properties versus that of their host halos, and compare with observations. Specifically, we consider the SMHM relation, the HI mass–halo mass relation, the HI mass fraction–halo mass relation and the sSFR–halo mass relation presented at $z = 0$. The SMHM is computed by comparing the total stellar mass of the central galaxy of a dark matter halo and plotting that against the total mass of the dark matter halo. This is similarly done for the HI mass and sSFR.

The top panel in Figure 2 shows the SMHM which describes the typical halo a galaxy of a specific stellar mass will reside in. We see that for smaller halo masses there is a steep increase in the mass of the galaxy that will occupy it up until $M_{halo} \sim 10^{11.3}\mathrm{M}_{\odot}$, thereafter this slope flattens and we see that typically, higher stellar mass galaxies occupy high mass halos. For the range of masses produced in Mufasa, comparing to the observationally-derived fitting function using abundance matching by Moster, Naab, and White (2013), we see good agreement, indicating that the way in which Mufasa populates halos is in good accord with real galaxies. This is expected because Mufasa is known to reproduce the observed stellar mass function of galaxies (Davé, Thompson, and Hopkins, 2016), and given that the halo mass function is set by cosmology, a correct stellar mass function thus implies a correct SMHM ratio.

The middle panel in Figure 2 shows the HI fraction compared to halo mass. We divide "red" and "blue" galaxies at sSFR=10^{-11} yr^{-1}. The cyan dots show the blue population in the Mufasa simulations and we see that higher HI fraction galaxies tend to be blue. We also see very few HI galaxies at high M_{halo} implying that most HI galaxies reside in smaller halos. The blue line displayed in the same panel comes from results from Saintonge et al. (2016) where the HI fraction of blue galaxies are compared to that of their stellar masses. To convert stellar mass into halo masses, the Moster, Naab, and White (2013) SMHM relation was used. There is generally good agreement, particularly in the populous low-mass regime. At the high halo masses Mufasa seems to underpredict HI, which was also seen in Davé et al. (2017) when looking at HI fraction at the highest stellar masses. This is not expected to have a large impact on the clustering statistics because there are only a few halos at the very massive end with too-low HI masses, but is nonetheless a caveat to keep in mind.

The bottom panel in Figure 2 shows the sSFR compared to halo mass. As

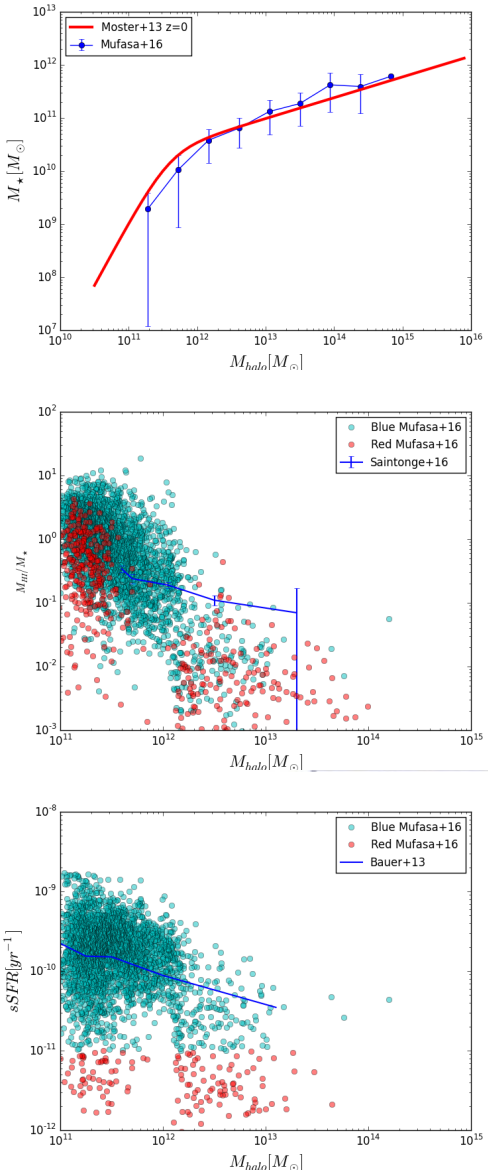

Figure 2: Central galaxy properties compared to that of their host dark matter halo. The top figure shows the SMHM relation at $z = 0$. The red line shows the analytic model by Moster, Naab, and White (2013) and the MUFASA points in blue. The middle panel shows the HI fraction per halo mass. The cyan blue dots are the blue MUFASA galaxies and the red dots are the red MUFASA galaxies. The blue line is the Saintonge et al. (2016) observed HI fraction for blue galaxies with respect to stellar mass extrapolated to halo mass using the Moster, Naab, and White (2013) relation above.
The bottom panel shows the sSFR against halo mass. Again, here the cyan dots show the blue MUFASA galaxies and the red dots show the red MUFASA galaxies. The blue line here shows the sSFR against stellar mass extrapolated to halo mass using the Moster relation above for blue galaxies by Bauer et al. (2013).

with the previous plot, the cyan dots show the blue star forming galaxies and the red dots show the non-star forming. We see that high-sSFR galaxies for their stellar mass reside in lower mass halos. A strong break at $M_{halo} \sim 10^{12}\mathrm{M}_\odot$ with many red galaxies appearing indicates the onset of MUFASA's halo mass based quenching scheme. The blue line shows data by Bauer et al. (2013) in which the sSFR is compared to that of the stellar mass, converted to halo mass as above. Note that this data only includes star-forming galaxies, so should be compared to the blue points. Once again, there is very good agreement with MUFASA and the observations.

These results imply that HI-rich galaxies live in similar conditions to that of the blue star-forming galaxies, and in this case HI-rich and blue star forming galaxies may cluster in the same way. However, a few HI-rich galaxies are part of the red quiescent group of galaxies, which may have effect on the total

clustering of HI-rich galaxies. This is an interesting test for galaxy evolution models as we expect to see a relation between gas content and star formation rate impact how galaxies populate halos. We aim to test such relationships in MUFASA, and thus constrain the underlying feedback models that establish these relations.

3.3 Mass Functions

Galaxy mass functions are a way of assessing the viability of MUFASA. Comparing these to data will indicate whether it is a useful description of the Universe, and whether the method by which MUFASA is populating halos with these properties is correct. We do not show the plots here, but refer to them in their original papers and discuss them in the context of this thesis.

3.3.1 Galaxy Stellar Mass Function

The galaxy stellar mass function (GSMF) is defined as the number density of galaxies in a given stellar mass bin. It is computed by dividing the total number of galaxies within a given stellar mass bin by the volume of the simulated box.

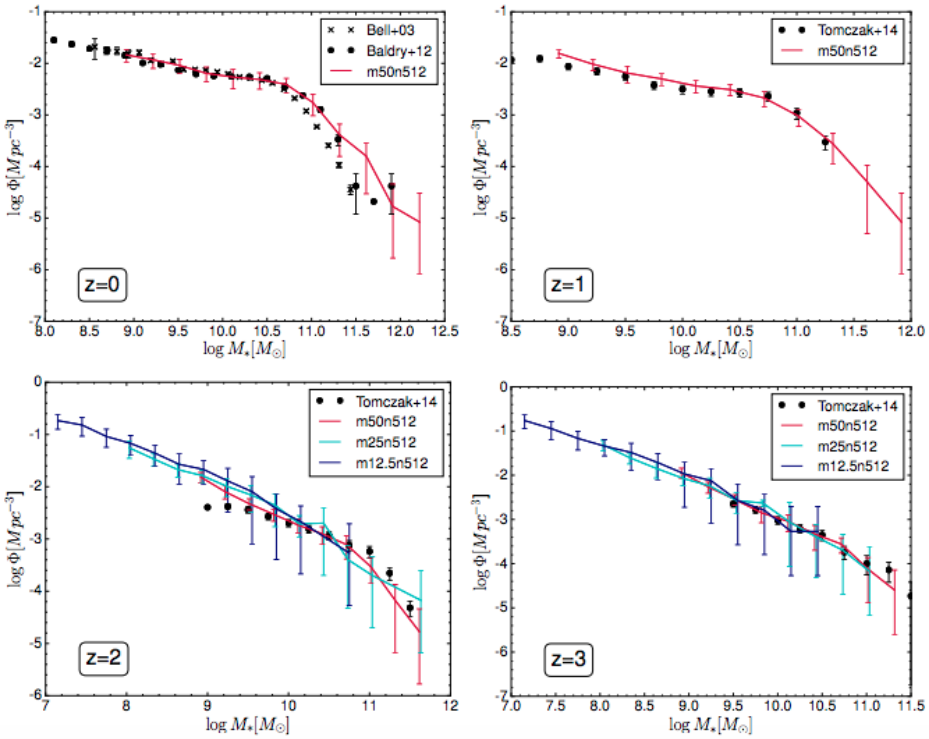

Figure 3: Galaxy stellar mass functions for MUFASA at z=0-3. Red, cyan, and blue lines show the mass functions for 50, 25, and 12.5h^{-1}Mpc cosmological volumes respectively. These are in comparison to observations (indicated by black points) by Baldry et al. (2012) at z∼0 and Tomczak et al. (2014) at z=1-3.

Figure 3 shows the evolution of the GSMF from Davé, Thompson, and Hopkins (2016) over z=0-3. in comparison to observations by, at $z = 0$, Baldry et al. (2012) and Bell et al. (2003) and Tomczak et al. (2014) for z=1 − 3.
For the redshifts probed in this thesis, MUFASA reproduces to good agreement the GSMF for $M_\star > 5.8 \times 10^8 M_\odot$ galaxies, which is the stellar mass resolution limit in the fiducial $50h^{-1}$Mpc cubic volume. From $z = 2$ we start seeing a turn-down in the mass function where quenching starts suppressing the star formation in high mass galaxies, and the turn-down becomes more prominent as galaxies evolve down to $z = 0$.

3.3.2 HI Mass Function

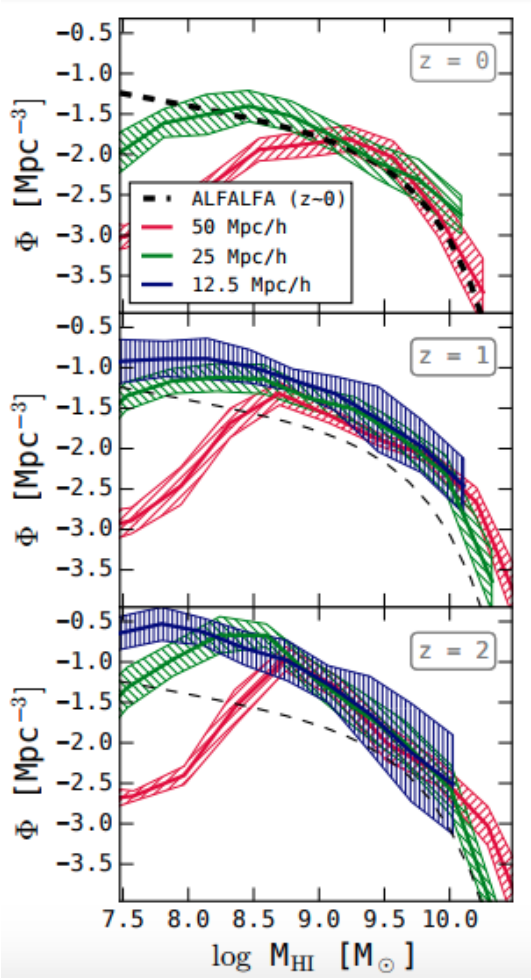

Figure 4: HI mass function for MUFASA galaxies at z=0-2. Red, green, and blue lines show the HI mass function for 50, 25, and 12.5 h^{-1}Mpc cosmological volumes respectively. These are in comparison to observations by ALFALFA at z≈0, reproduced at z=1 and 2 to depict the evolution of the HI mass function.

The same function can be reproduced for HI mass, where the HI Mass Function (HIMF) is defined as the number density of galaxies in a given HI Mass bin.
Figure 4 shows the HI mass function from Davé et al. (2017) produced by the different MUFASA volumes in comparison to HI observations from the ALFALFA survey (Haynes et al., 2011). The deviations in the different volumes are due to the different mass resolutions.
At $z = 0$, MUFASA shows a good agreement for the HIMF with the ALFALFA survey within the resolved mass range. Unfortunately, current instrumentation is not sensitive enough to characterise the HIMF at higher z, however, MUFASA predicts an increase in the number density of galaxies at a given HI mass to higher redshifts that can be tested in future surveys such as LADUMA. In the middle panel in Figure 2 we see that the HI-rich galaxies are mostly blue galaxies, however there is a significant number that are red. It is thus expected that the clustering of HI-rich galaxies follow similar trends to that of blue galaxies but not as strongly the red HI-rich population.

3.3.3 Star Formation Rate Function

The SFR function is defined as the number density of galaxies within a given SFR bin.

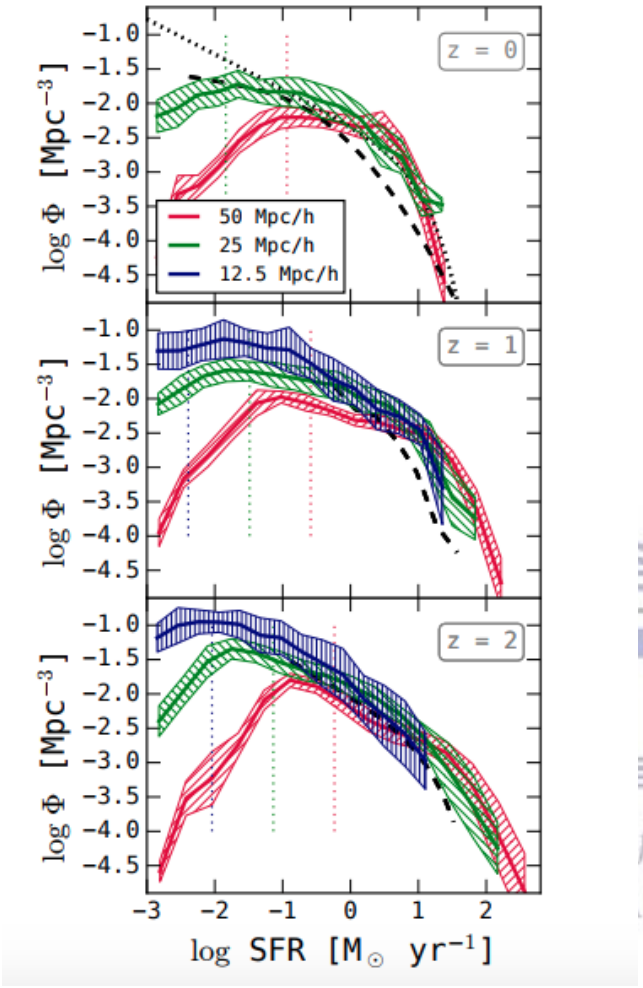

Figure 5: Star formation rate functions from Davé et al. (2017) at z=0-2 for MUFASA galaxies. Red, green, and blue lines show 50, 25, and 12.5 h^{-1}Mpc cosmological volumes respectively. Vertical lines show the SFR at the respective mass resolution limit for each volume. Observations using Hα luminosity converted to SFR functions are shown as black dashed lines by Gunawardhana et al. (2013) at z≈0, Colbert et al. (2013) at z≈1, and Mehta et al. (2015) at z≈2.

Figure 5 shows the SFR function for MUFASA galaxies from Davé et al. (2017). It shows in the upper panel, at $z = 0$, while the agreement is good over much of the range, MUFASA overproduces galaxies between $SFR \sim 1 - 10 M_\odot$. This is possibly due to Hα surveys being unable to detect the most star forming galaxies as they are highly obscured. This can be accounted for by examining Far-IR based SFR estimators, however, this introduces complexity due to the subtraction of the AGN contribution of the total flux.

This discrepancy is consistent with Figure 3 of Davé, Thompson, and Hopkins (2016) where the SFR is overpredicted at $z = 0$. We can thus conclude that MUFASA's SFR Function is broadly compatible with observations although with notable overpredictions of galaxies with a SFR similar or greater than that of the Milky Way.

For $z = 1$ the overprediction of galaxies shift to high SFRs as we're able to "detect" galaxies with $SFR \gtrsim 100 M_\odot yr^{-1}$, and we observe a weaker truncation at higher SFR compared to that of lower z. At $z = 2$, this trend continues where we see an excess at high SFR. However, the excess is not as strong as lower z and matches well to observations.

If we had produced an SFR function which had more low SFRs, we would expect to see a redder Universe and thus a more highly clustered Universe. Contrary to this, if we had a significant excess of high SFR galaxies, we would expect to see more blue and thus weaker clustering galaxies. These results give us an indication that the manner in which MUFASA produces HI masses, stellar masses, and star formation in galaxies is viable and useful when populating halos.

3.4 Luminosity Functions

In Davé, Thompson, and Hopkins (2016) and Davé et al. (2017), it was shown that MUFASA reproduces the stellar mass function, the HI mass function, and the star formation rate function well as compared to observations. This indicates that MUFASA provides a realistic sample of galaxies with intrinsic properties as observed. However, so far there has been no study of the luminosity function of galaxies in particular-bands in MUFASA. We present this here, in order to test the viability of MUFASA specifically for the bands that we will examine in our clustering analysis.

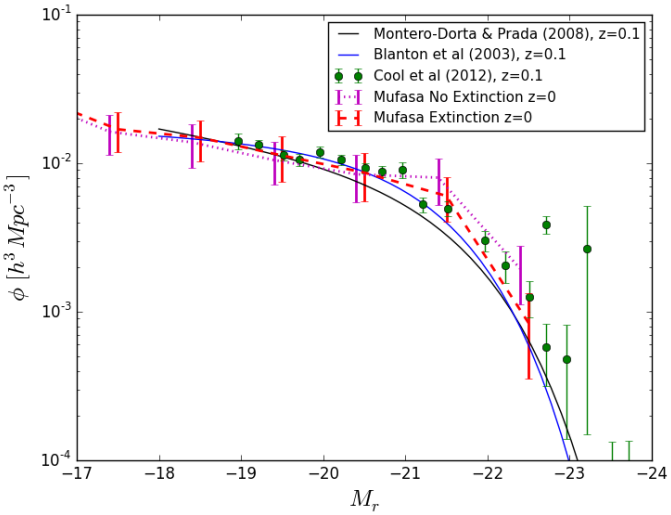

Figure 6: r-band luminosity function at $z = 0$. The red line depicts the galaxies from the simulation output with extinction, and the magenta line depicts the simulation output without extinction. The green points are observations by Cool et al. (2012). The blue and black lines are the Schechter fits predicted by Blanton et al. (2003) and Montero-Dorta and Prada (2009) respectively.

The luminosity function is defined as the number density of galaxies per given luminosity bin. It is computed by dividing the total number of galaxies within each magnitude bin by the volume of the simulated box. We use Line Of Sight Extinction by Ray-tracing (LOSER) (Davé, Rafieferantsoa, and Thompson, 2017) which returns a catalogue of luminosities in several bands for simulated

galaxies accounting for dust extinction as well as not, as mentioned in section 2.

Looking at Figure 6, we see that in the r-band, the MUFASA simulations show good agreement with that from the observations listed in the caption. There is a shallow faint-end slope, and an exponential turnover around $M_r \approx -21.5$ that represents the $L_\star$ value in a Schechter function fit for the LF in this band. We can compare directly to the data points of Cool et al. (2012), where one can see good agreement in the predicted luminosity functions as well as to the Schechter fits from other observations.

In Figure 6 we also show the difference between the galaxy sample including extinction (red) as well as without (magenta) where we see that, as expected, the extincted sample produces fewer bright r-band galaxies. This is the sample used in further analyses as with evidence from observations, since the Universe as we know it is composed of dust and galaxies are affected by extinction.

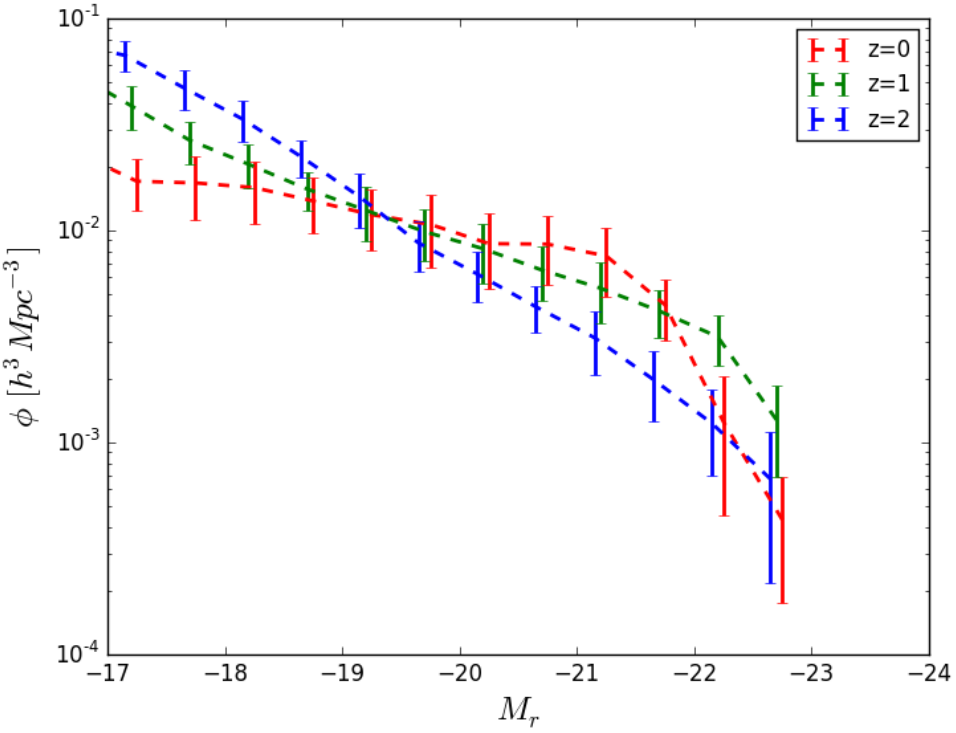

Figure 7: Evolution of the r-band luminosity function as a function of redshift for $z = 0 - 2$ (red to blue). The luminosities at each redshift take into account galaxies that have been affected by extinction.

The evolution over time of the rest-frame r-band LF is shown in Figure 7. We can see how the luminosity function evolves from a double power law or a Schechter function to that of a single power law with increasing redshift. With increasing redshift, we see an increase in the number of faint r-band galaxies, while the number of bright galaxies decrease. This is expected owing to the hierarchical growth of structure, in which small faint galaxies form first

and aggregate into larger galaxies over time. Along with this we also see an exponential cutoff in the LF starting from $z = 2$, an indication of the quenching mechanism suppressing star-formation in bright galaxies.
Relating this to expectations regarding clustering, if we were to see an excess of galaxies at the bright end of the r-band luminosity function, we would expect to see a high clustering amplitude as this would dominate the sample of galaxies produced by MUFASA. On the contrary, if we had produced more faint r-band galaxies, we would expect a decrease in the clustering amplitude.

We will also consider the GALEX NUV band in our clustering analysis. For the NUV, we can see in Figure 8 that the difference between the extincted and non-extincted galaxies is much more significant. In the case of no extinction, many more bright NUV galaxies are produced. This is understood as, in dusty galaxies, the strong UV emission is absorbed by the dust and re-emitted in the FIR resulting in fewer bright UV galaxies.

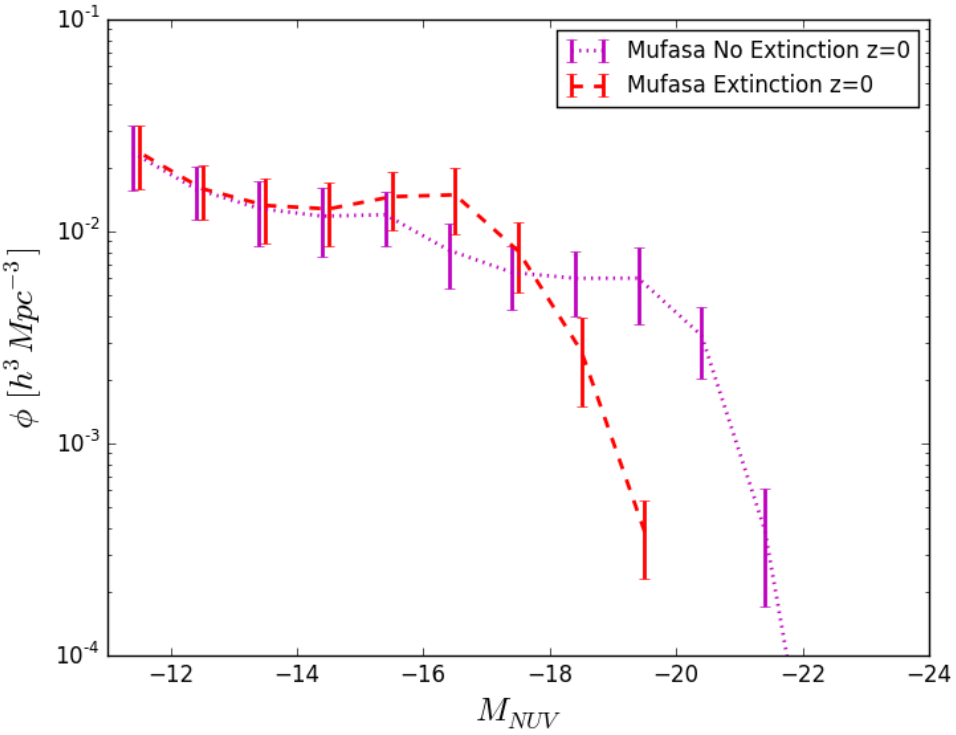

Figure 8: NUV luminosity function at $z = 0$. The red line depicts the galaxies where are affected by extinction, and the magenta line depicts where they are not.

Note that the characteristic luminosity, $L_\star$ also differs by $\approx$ 4mag. Figure 9 shows that over time the number of galaxies with $M_{NUV} < -12$ increases. Also note that the luminosity resolution for the NUV is ≈ -12 and it is for this reason we can conclude that the number of galaxies per M_{NUV} bin increases as a function of redshift, as seen in Figure 9. This implies that star formation was stronger at higher redshift or there was less dust to obscure and absorb the UV emission.

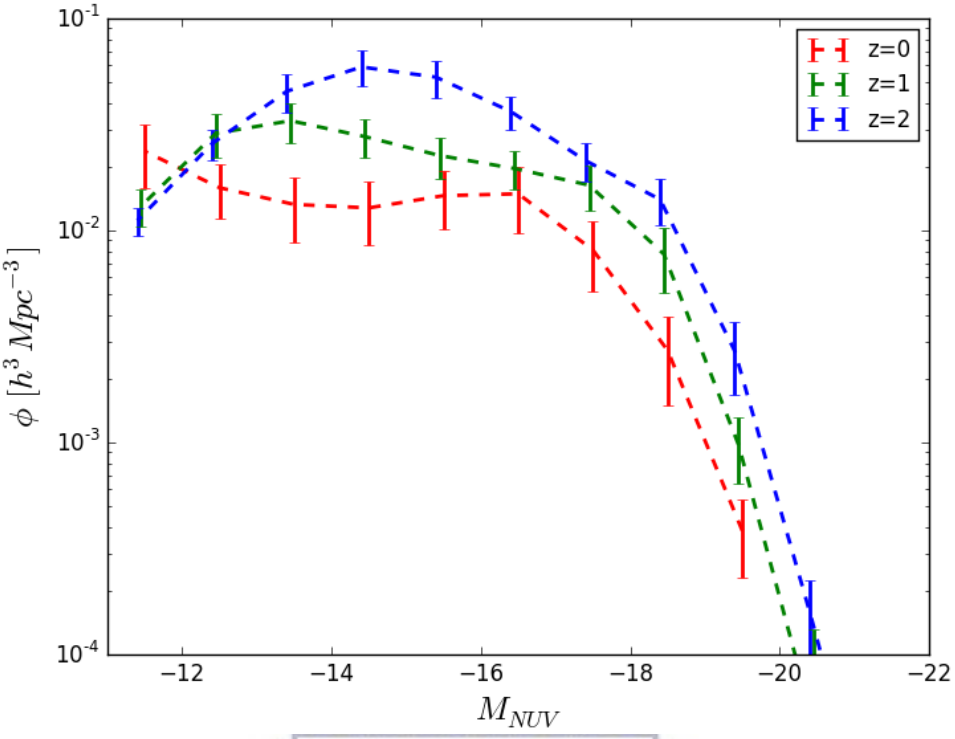

Figure 9: Evolution of the NUV luminosity function as a function of redshift for $z = 0 - 2$. The luminosities at each redshift account for dust extinction.

For NUV luminosities, we expect to see the opposite to that of the r-band in terms of clustering dependence as NUV is a significant tracer of star forming galaxies. So if we were to see an excess of galaxies at the bright end of the NUV-band luminosity function, we would expect to see a decrease in clustering amplitude whereas if we had produced more faint NUV-band galaxies, we would expect an increase in the clustering amplitude.

In summary, the results in this section show that MUFASA reproduces good agreement to the observable mass, luminosity and SFR functions, and HODs. This indicates that the galaxies produced are a suitable sample set to study clustering in this thesis.

4 Clustering Properties of Galaxies

Given that MUFASA reproduces a viable population of galaxies between $z = 0 - 3$, we can study the clustering properties of these galaxies and compare them to that of observations.

Zehavi et al. (2005) predicts that, at $z \approx 0$, the full sample of SDSS galaxies approximately follows a clustering power law in the form $(\frac{r}{r_0})^{\gamma}$ with $r_0 = 5.59 h^{-1}$Mpc and $\gamma = -1.84$. However, for galaxies with $M_r > -22$, it is predicted that $r_0 \approx 5 h^{-1}$Mpc and $\gamma \approx -1.8$. The brightest subsample presented in Zehavi et al. (2005), $-23 < M_r < -22$, has a significantly steeper $\xi_g(r)$, which creates the difference between the two trends. Hawkins et al. (2003) used the 2dF Galaxy Redshift Survey to predict a relation in the same power-law form with $r_0 = 5.05 h^{-1}$Mpc and $\gamma = -1.67$, which is somewhat different to the Zehavi SDSS prediction, likely owing to differences in selection.

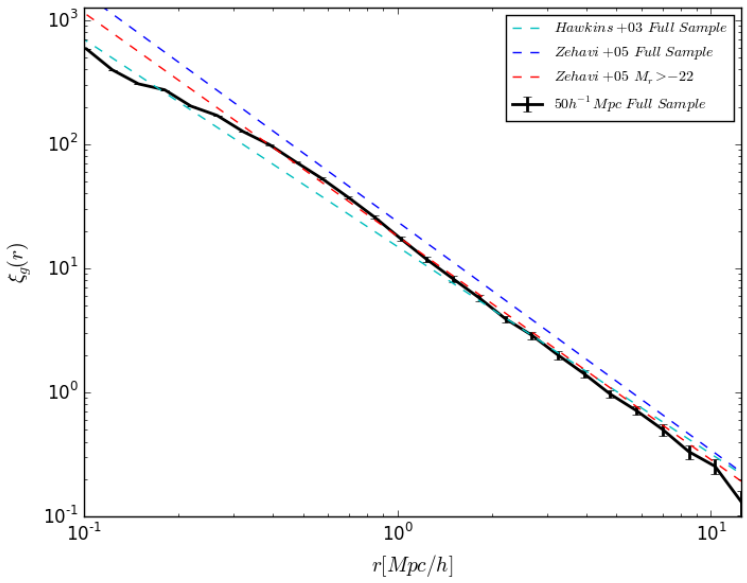

Figure 10: Spatial TPCF as a function of separation distance, r, for the MUFASA $50 h^{-1}$Mpc box in comparison with Hawkins et al. (2003) (2dFGRS) full galaxy sample (cyan dashed line) and Zehavi et al. (2005) (SDSS) full galaxy sample (blue dashed line) as well as Zehavi et al. (2005) for galaxies fainter than $M_r = -22$ (red dashed line)

Figure 10 Shows us the TPCF as a function of separation distance, r, for the full sample of MUFASA galaxies (black solid line) in comparison to that of Hawkins et al. (2003) in cyan, Zehavi et al. (2005) full galaxy sample in blue, and Zehavi et al. (2005) flux limited sample in red. When comparing the TPCFs in Figure 10, we see that by using a least squares fit, MUFASA produces a power law that is in good agreement with observations. There are however slight differences between the amplitudes which may be a result of the types of galaxies contained within each sample. SDSS covers five optical bands ($ugriz$)

and is more sensitive to bright red galaxies which thus dominates the full sample, and it is for this reason there exists such a significant difference between the two observed power-laws provided by Zehavi et al. (2005). This is not the case for 2dFGRS which covers the four bands ($ugri$) at $300 - 800$nm. The result here shows us that MUFASA produces enough galaxies up to $M_r \approx -22$ to produce a correlation function which matches well to the luminosity limited sample by Zehavi et al. (2005)

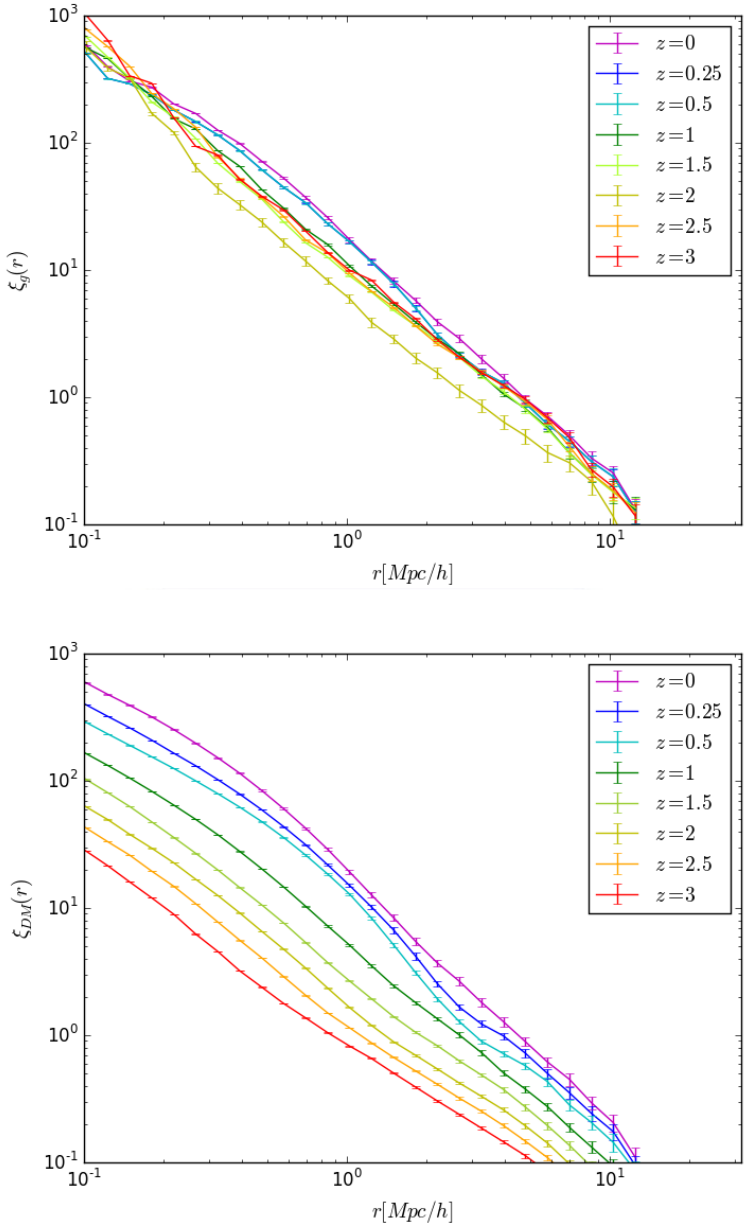

Figure 11: Evolution of galaxy TPCF (top) and dark matter TPCF (bottom) for $z = 0 - 3$ (in order of magenta to red)

For a gravity-only field, we expect to see a steady decrease in clustering ampli-

tude with redshift, which is exactly what we see in thebottom panel of Figure 11; the TPCF for the dark matter distribution, $\xi_{DM}(r)$. In the top panel of Figure 11 shows the galaxy TPCF, $\xi_g(r)$, where we see a steady decrease in clustering amplitude with redshift as well. However, at $z = 2$, we see a turnover and as redshift continues to increase, the clustering amplitude increases as well.

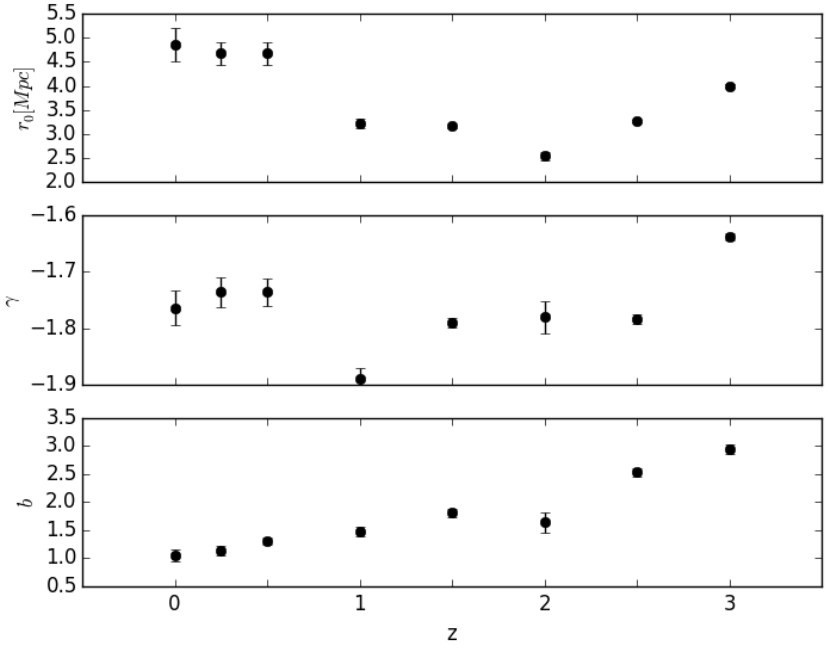

Figure 12: r_0, γ, and bias factors for full sample correlation functions as a function of redshift

By fitting a power law, we can compute the clustering lengths and slopes, r_0 and γ, for each correlation function. We can also determine the bias as a function of separation distance, r, by taking the square of the ratio of the galaxy correlation function to the dark matter correlation function as per equation 7, with this the bias factor is selected at r$\approx$ 2.67h^{-1}Mpc, and we can see how these parameters evolve as a function of time.

These parameters are shown in Figure 12, where the upper panel shows the values for the clustering parameter r_0 as a function of redshift where, due to the turn over in the galaxy correlation function, we see a turnover in the clustering length parameter as well. The middle panel shows the slope of the TPCF, γ, as a function of redshift where we see that the slope is an almost constant factor fluctuating between $-1.9 < \gamma < -1.6$, with most at $\gamma \approx -1.8$. With this, and with the steady decrease in the dark matter TPCF, in the bottom panel we see an increase in the bias factor with time.

It is possible to divide the full sample into its central and satellite counterparts and see how they evolve over time with respect to the full sample of galaxies

as well as how they evolve with bias.

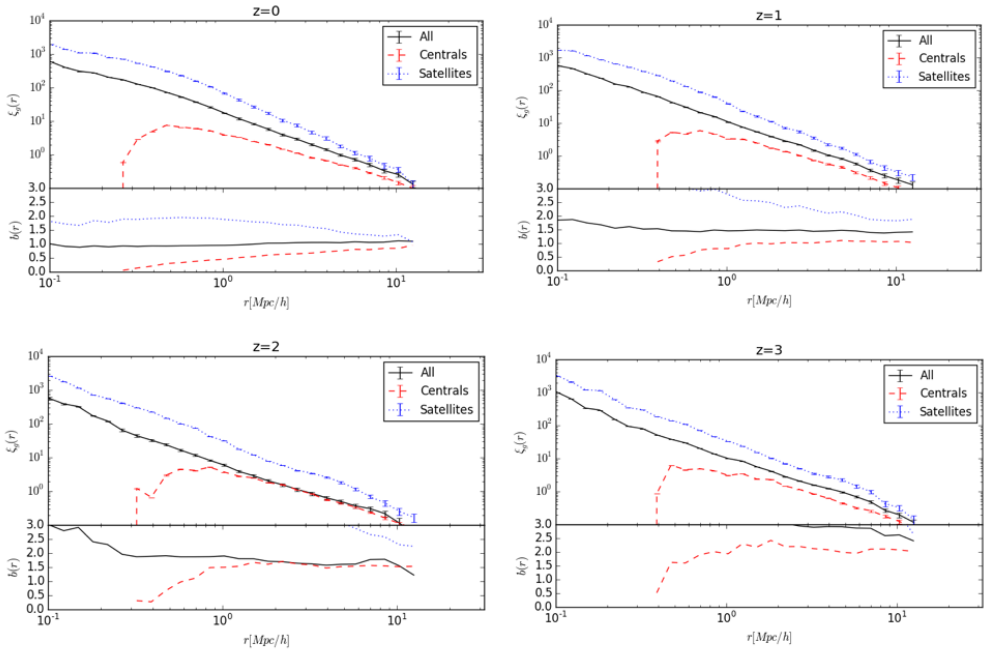

Figure 13: Spatial TPCF and bias as a function of separation distance, r, for central (blue dashed line) and satellite (red dotted line) galaxies in comparison to the full sample (black solid line) at $z = 0 - 3$.

Figure 13 shows us the TPCF as a function of separation distance, r, for the full MUFASA galaxy sample (solid black line) along with the sample split between satellites (blue dashed line) and centrals (red dotted line) for $z = 0-3$. In Figure 13 we see that the clustering dependence for satellites alone stays constantly above that of the full sample over the entire radial scale probed here. Whereas the central galaxies approach the full sample on larger separations, with $z = 2$ the centrals following the same clustering strength as the full sample and starting to deviate therefrom with increasing redshift again. This gives an indication that satellites dominate the clustering dependence at r$\lesssim 1\mathrm{h}^{-1}$Mpc, and the central galaxies for r$\gtrsim 1\mathrm{h}^{-1}$Mpc.

4.1 Stellar Mass

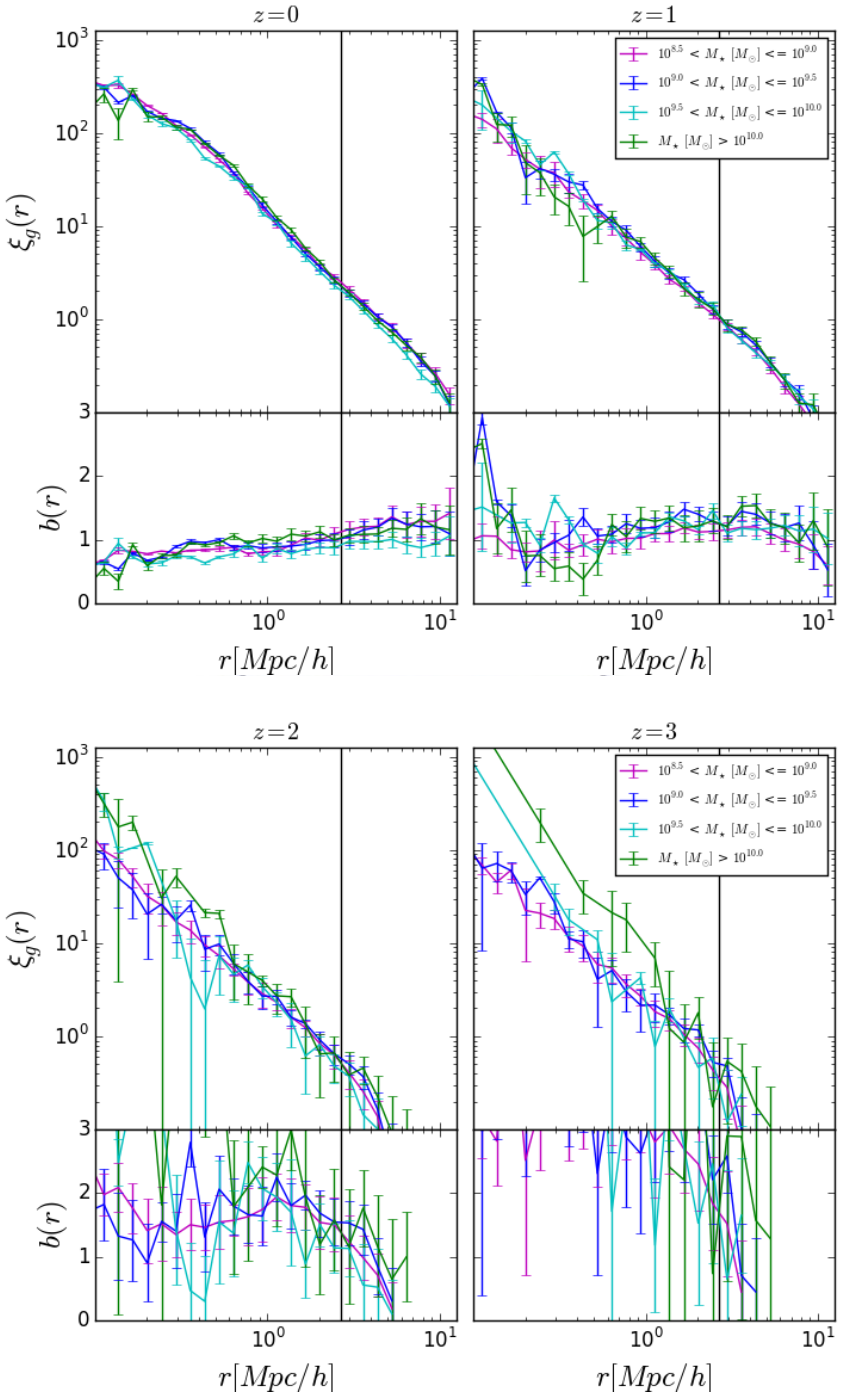

Figure 14: Clustering as a function of stellar mass for $z = 0 - 3$. Colours in the order of magenta to green indicate the stellar mass bins for $logM_\star > 8.5$. Upper panels show the galaxy TPCF and lower panels show the bias as a function of separation distance, r. Vertical black lines indicate the separation distance at which the bias factor is defined.

Studies of stellar mass, $M_\star$, have shown tight correlations with other intrinsic properties of galaxies such as luminosity, sSFR, etc. Understanding how galaxies with specific stellar masses cluster can give us an indication as to what other physical properties selected galaxies are comprised of and how they cluster and evolve in terms of those properties. Given that MUFASA reproduces a viable stellar mass function down to $z = 0$, we can examine how stellar mass effects the clustering of galaxies throughout the simulation.

Figure 14 shows the TPCF and bias as a function of separation, r, for $z = 0-3$. Each panel shows the TPCF and bias for a number of stellar mass bins in increasing orders of magnitude from magenta to green. In Figure 14, we see that at a given redshift, there is an increase in the clustering amplitude as the stellar mass of galaxies increase. At $z = 0$, in the top-left panel, this trend is very small, however as we move to higher redshifts, the separation between the clustering trends become more evident. At $z = 0$, by comparing the clustering parameter r_0, we see that the highest mass galaxies cluster about 1.2 times stronger than the lowest mass galaxies. There exists increasing scatter as we go up in redshift, specifically in the higher mass bins. This is due to the fact that there are fewer galaxies with which to compute the correlation function at these stellar masses.

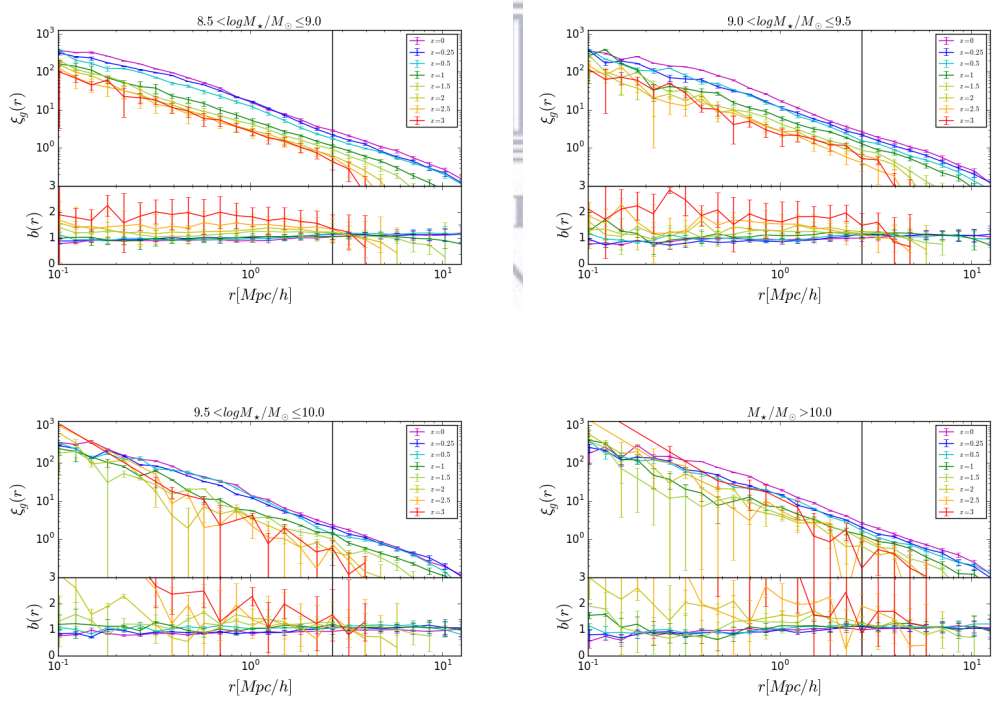

Figure 15: Clustering evolution per given $M_\star$ bin as a function of redshift. Colours in the order of magenta to red indicate increasing redshifts. Upper panels show the galaxy TPCF and lower panels show the bias as a function of separation distance, r. Vertical black lines indicate the separation distance at which the bias factor is defined.

In Figure 15, each panel shows the TPCF and bias as a function of separation distance, r, for a given stellar mass bin evolving with redshift. Colours in order from magenta to red increase as $z = 0 - 3$. We see that the clustering amplitude decreases as we go to higher redshifts, with this we see that the bias factor remains almost constant with time as seen in Figure 16; showing the evolution of the clustering parameters as in Figure 12 for each stellar mass bin and colour coded as per the legend and Figure 14, which implies that stellar mass is a strong tracer of the underlying dark matter field.

Using the PRIMUS and DEEP2 surveys, Coil et al. (2016) studied the dependence of galaxy clustering on stellar mass, in which they found that the clustering amplitude increases with stellar mass. They also showed that this amplitude decreases with redshift, which is in good agreement with the results found above. The separation at which the bias factor is defined is not provided, however they stated that the bias is determined for one-halo ($0.1 < r < 1$ h^{-1}Mpc) and two-halo ($1 < r < 10h^{-1}$Mpc) terms. We can thus compare to the two-halo scales as our separation at which we define the bias factor lies within the two-halo term. They found that the bias increases with $M_\star$ and do not evolve much over the redshifts probed. This, too, is in general agreement with the MUFASA predictions.

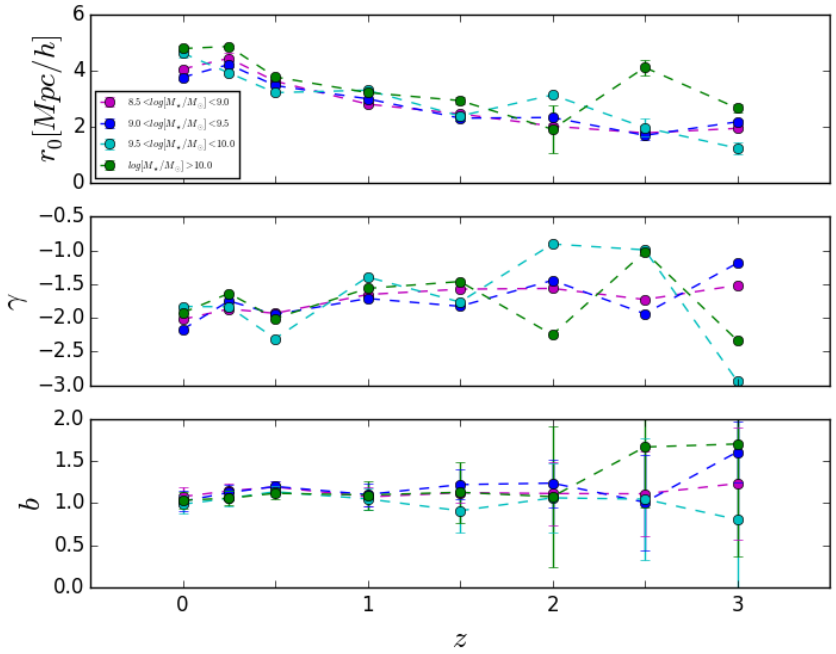

Figure 16: Evolution of clustering parameters as a function of stellar mass as in Figure 12. However, colours in the order of magenta to green indicate the stellar mass bins for $logM_\star > 8.5$ as per the legend.

4.2 r-band Luminosity

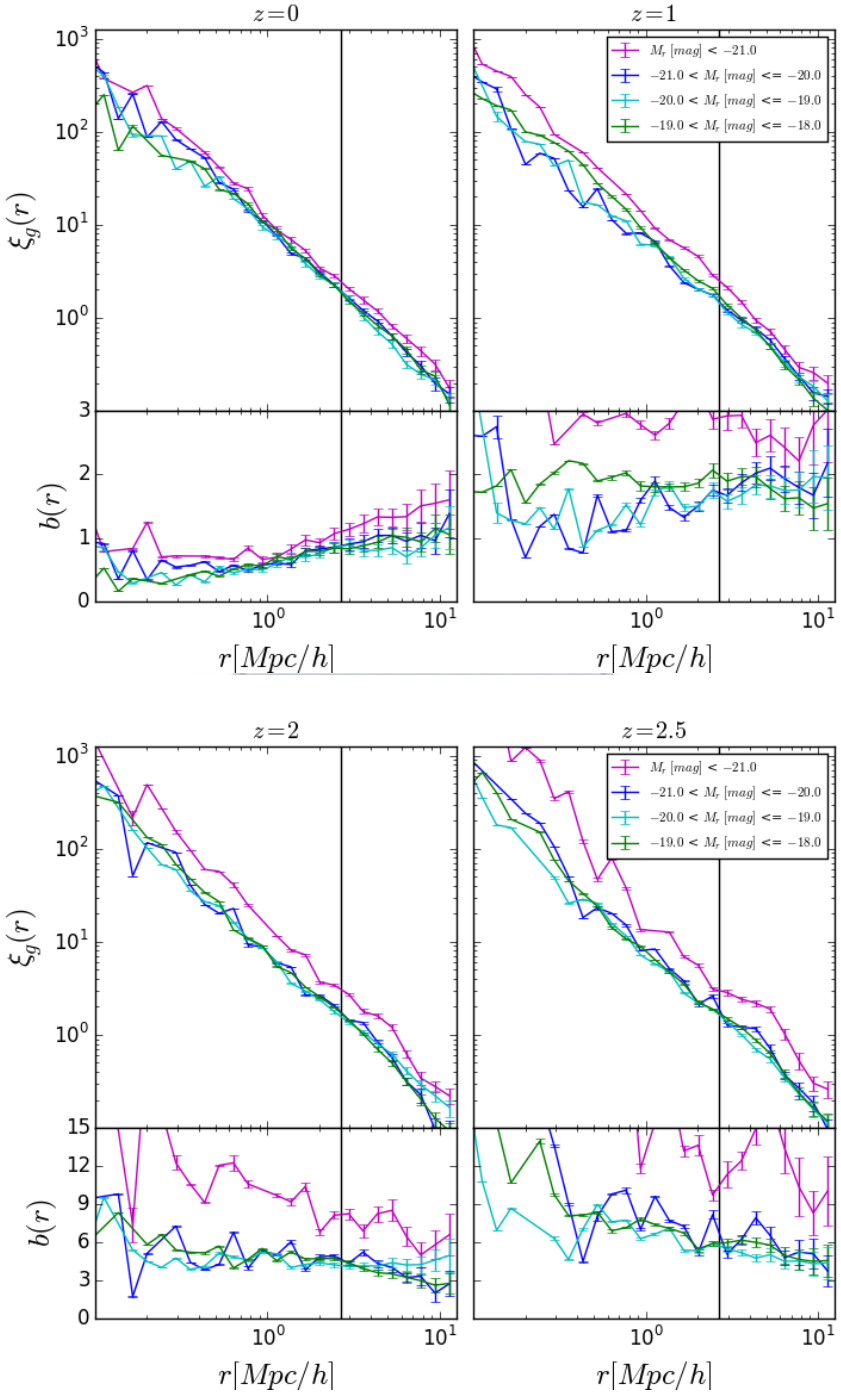

Figure 17: Clustering as a function of r-band luminosity for $z = 0-2.5$. Colours in the order of magenta to green indicate the decreasing luminosity bins. Upper panels show the galaxy TPCF and lower panels show the bias as a function of separation distance, r. Vertical black lines indicate the separation distance at which the bias factor is defined.

There are several mechanisms for the quenching of star formation in galaxies. These include stellar feedback, stripping of particularly satellite galaxies - which includes stripping from a nearby large galaxy interacting with the galactic material and stripping the gas therefrom, as well as ram pressure stripping which is the stripping of gas as the galaxy moves through the intergalactic medium. To form stars it is required that the gas within the galaxy is cold, and thus turbulent and constantly hot gas within clusters also stops star formation, leaving galaxies with only the red, old population stars. The primary mechanism for quenching star formation in massive red galaxies are black holes or AGN which are still poorly understood and are thus an interesting property to study clustering.

Several studies on the clustering dependence on r-band luminosity have been published in the past. Zehavi et al. (2011) has found that, at low redshifts, $z < 0.25$, with increasing brightness (or decreasing magnitude), the clustering amplitude for galaxies increases. With the viable reproduction of r-band galaxies in MUFASA, we can examine the clustering dependence of galaxies in the r-band and check whether MUFASA also reproduces the appropriate clustering trends as seen in observations.

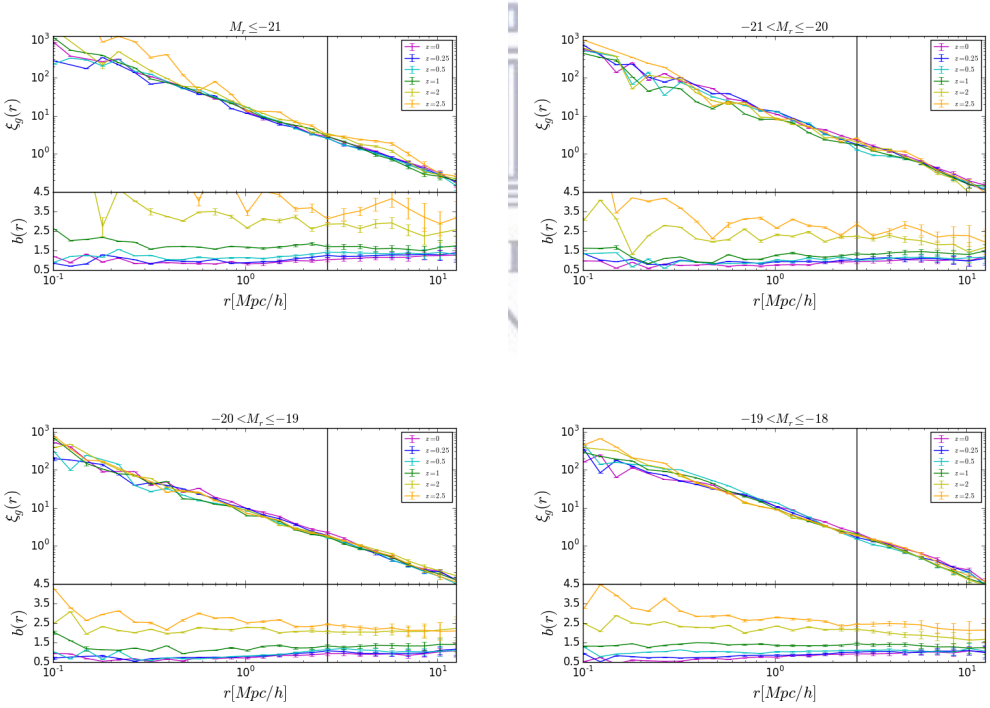

Figure 18: Clustering evolution per given M_r bin as a function of redshift. Colours in the order of magenta to yellow indicate increasing redshifts. Upper panels show the galaxy TPCF and lower panels show the bias as a function of separation distance, r. Vertical black lines indicate the separation distance at which the bias factor is defined.

Figure 17 shows the TPCF and bias as a function of separation, r, for $z = 0 - 2.5$. Each panel shows the TPCF and bias for a number of r-band magnitude bins in increasing order from magenta to green. At a given redshift in Figure 17, we can see a clear dependence of clustering amplitude on r-band luminosity. With increasing brightness in the r-band we see an increase in the clustering amplitude of the TPCF, implying that brighter, redder galaxies cluster more than fainter r-band galaxies. At $z = 0$ the brightest r-band galaxies cluster about 1.3 stronger than the faintest r-band galaxies, this factor seems to increase with redshift. As we go to higher redshifts, not only do we see this trend increasing in amplitude, but also that the dispersion between luminosity bins becoming more distinct.

In Figure 18, each panel shows the TPCF and bias as a function of separation distance, r, for a given r-band magnitude bin evolving with redshift. Colours in order from magenta to yellow increase as $z = 0 - 2.5$. For the brightest luminosity bin in Figure 18, we see an increase in clustering amplitude with redshift. However, for the fainter bins, we don't see an obvious evolution in clustering amplitude. Due to the increasing amplitude of the correlation, both with luminosity and with redshift, we see an increasing bias as a function of separation distance. There is an increase in bias factor by a factor of about 3 between $z = 0$ and $z = 2.5$. For all of the above cases, regardless of luminosity bin, we see an evolution in the bias amplitude seen in the bottom panel of Figure 19; showing the evolution of the clustering parameters as in Figure 12 for each stellar mass bin and colour coded as per the legend and Figure 17. This is due to the distinct decreasing clustering amplitude of the dark matter correlation function as a function of redshift.

Using the main sample of the Sloan Digital Sky Survey (SDSS), Zehavi et al. (2011) studied the clustering dependence of galaxies on colour and luminosity at low redshifts of $z < 0.25$. They found that the clustering amplitude increases with increasing brightness in the r-band. With this, they find that the bias factor (chosen at r$\approx 2.67 h^{-1}$Mpc) increases with increasing brightness. Unfortunately, data points for binned luminosity samples was not published and thus direct comparisons cannot be made. However, the general trends found are in agreement with the predictions from Mufasa.

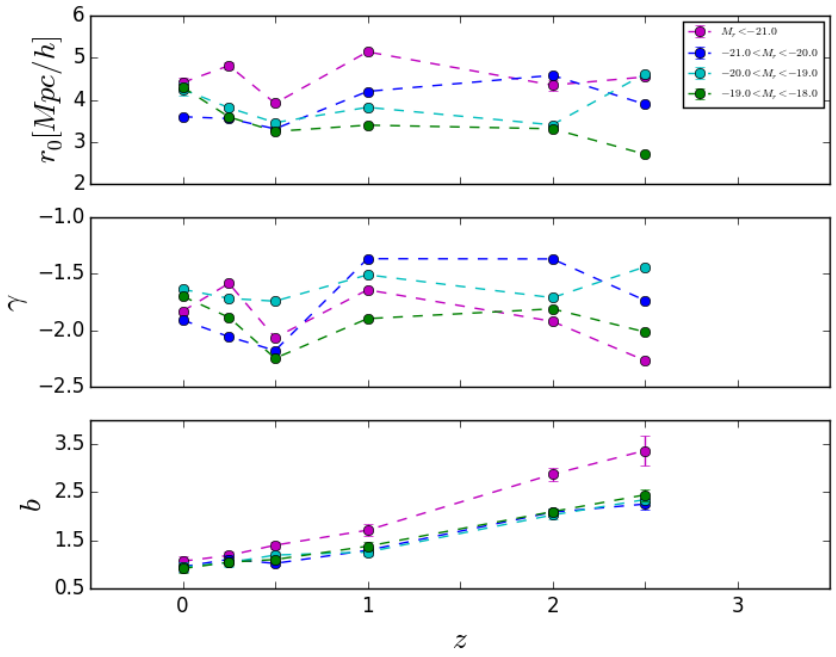

Figure 19: Evolution of clustering parameters as a function of r-band luminosity as in Figure 12. However, colours in the order of magenta to green indicate the luminosity bins for $M_r < -18$ as per the legend.

4.3 Specific Star Formation Rates

There is a tight correlation between star formation rate and the stellar mass of a given galaxy. The correlation between the two, often called the main sequence of galaxies, describes how fast galaxies grow as a function of their stellar mass. The slope of the main sequence of galaxies is referred to as the specific star formation rate, $sSFR$, which is defined as the star formation of the galaxy relative to its mass by

$$sSFR = \frac{SFR}{M_\star} \frac{[M_\odot yr^{-1}]}{[M_\odot]} \tag{10}$$

A high sSFR, or a galaxy that lies above the main sequence, implies that there are many new forming stars relative to its size and thus will appear more blue in colour, and is referred to as a starburst or highly star-forming galaxy. Galaxies with a low sSFR, or that lie below the main sequence, will appear redder due to the galaxy being made up of mostly old population stars and are often referred to quiescent, passive, or "red and dead" galaxies. Between the main sequence and quiescent galaxies lies the green valley of galaxies. It is for this reason that we expect to see a sudden change in the clustering of galaxies at a specific value for the sSFRs, this is the point at which we can differentiate between "quiescent" and "star forming", or red vs blue, galaxy populations.

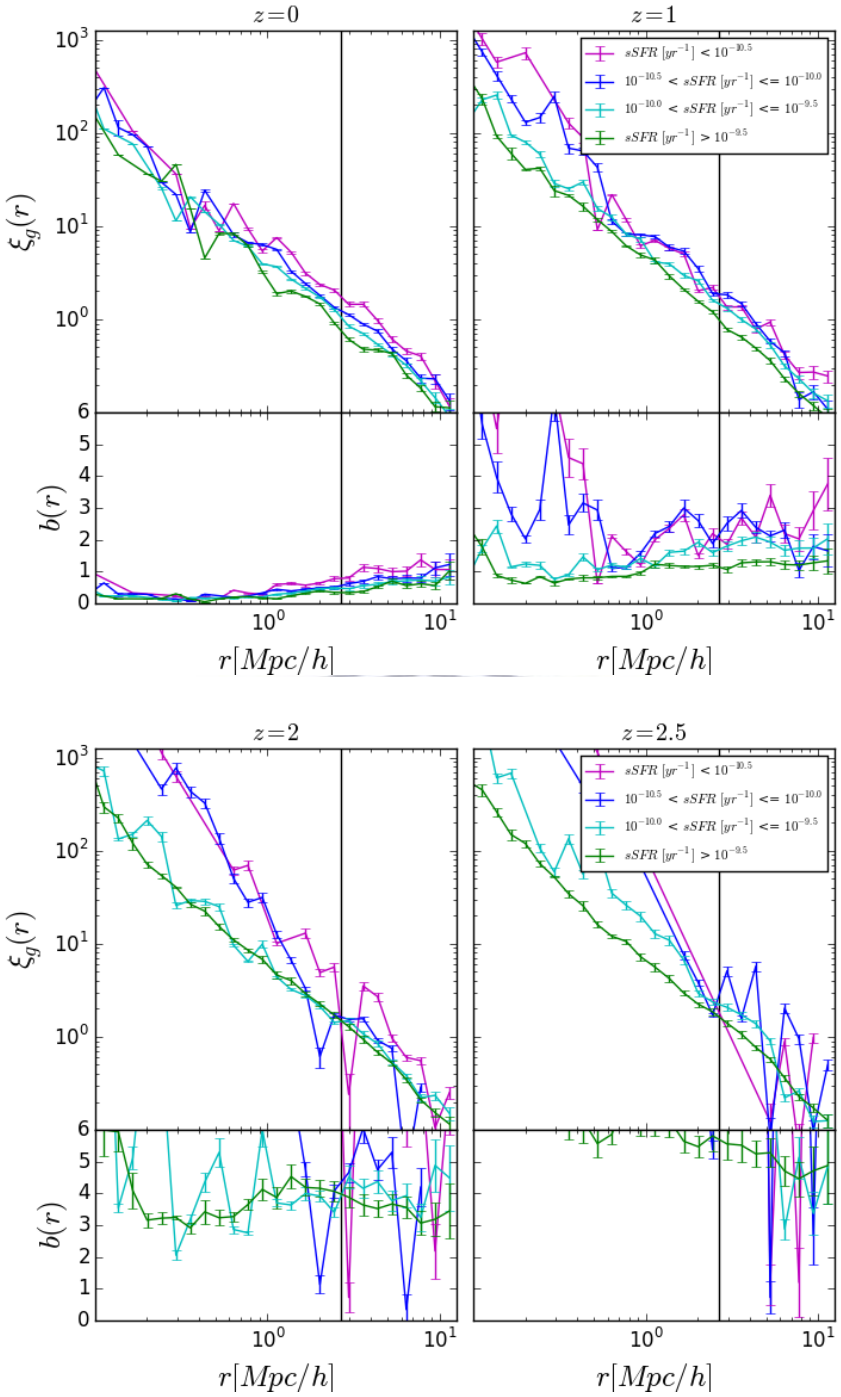

Figure 20: Clustering as a function of sSFR for $z = 0 - 2.5$. Colours in the order of magenta to green indicate the sSFR bins. Upper panels show the galaxy TPCF and lower panels show the bias as a function of separation distance, r. Vertical black lines indicate the separation distance at which the bias factor is defined.

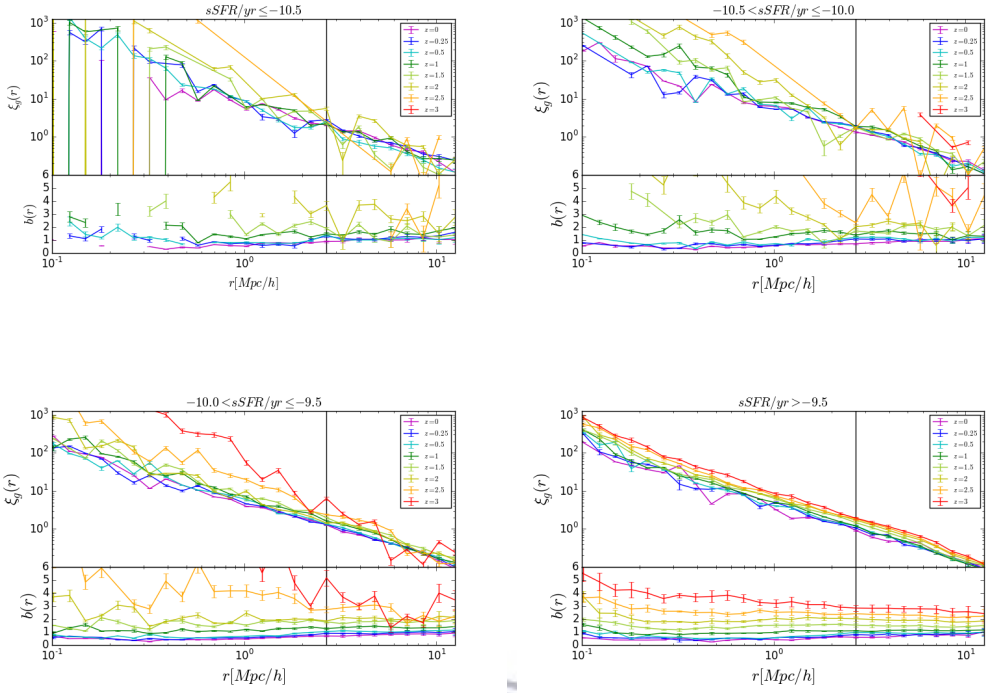

Figure 21: Clustering evolution per given sSFR bin as a function of redshift. Colours in the order of magenta to red indicate increasing redshifts for $z = 0 - 3$. Upper panels show the galaxy TPCF and lower panels show the bias as a function of separation distance, r. Vertical black lines indicate the separation distance at which the bias factor is defined.

With the viable reproduction of galaxy star formation rates in MUFASA, we can examine the clustering dependence of galaxies of sSFR and compare these to current observational studies.

Figure 20 shows the TPCF and bias as a function of separation, r, for $z = 0 - 2.5$. Each panel shows the TPCF and bias at a given redshift for a number of sSFR bins in increasing orders of magnitude from magenta to green. We see that there is an increase in clustering amplitude with decreasing sSFR. With increasing redshift, we see that the clustering amplitude increases and the dispersion between the binned samples become much greater allowing us to see the separation of quiescent and star-forming samples. This separation occurs at $log(sSFR[\mathrm{yr}^{-1}]) \sim -10$. The difference in these populations is clear at scales less than that of the clustering parameter r_0 due to the change in slope, and we thus cannot compare these parameters, however we can see a distinct difference in the slope of the clustering, γ, where the difference between the least star-forming (red) and most star-forming (blue) galaxies is a factor of about 2. We can study the separate population of galaxies in future work.

With increasing clustering amplitude, we do see an increase in bias as a function of both separation and increasing redshift. In Figure 21 we see the redshift evolution of the TPCF and bias as a function of separation distance, r, for a given sSFR bin, there is an increase in the clustering amplitude with redshift. With this, there is also an increase in bias with redshift. The dispersion between the bias also decreases with increasing sSFR. Figure 22 shows us the clustering parameters as in Figure 12 for each sSFR bin as a function of redshift and colour coded as per the legend and Figure 20.

Coil et al. (2016) studied the clustering dependence of galaxies on sSFR and found that the clustering amplitude decreases with sSFR. They also found that the bias decreases with sSFR, both of which is in agreement with predictions by Mufasa. Zehavi et al. (2011) also studied the clustering dependence of galaxies on colour which can be linked directly to sSFR, with red galaxies having lower sSFR values and blue galaxies having higher sSFR values. They had found that red galaxies tend to cluster more than blue galaxies and thus the bias increases from blue to red, which is also in agreement with the results found above.

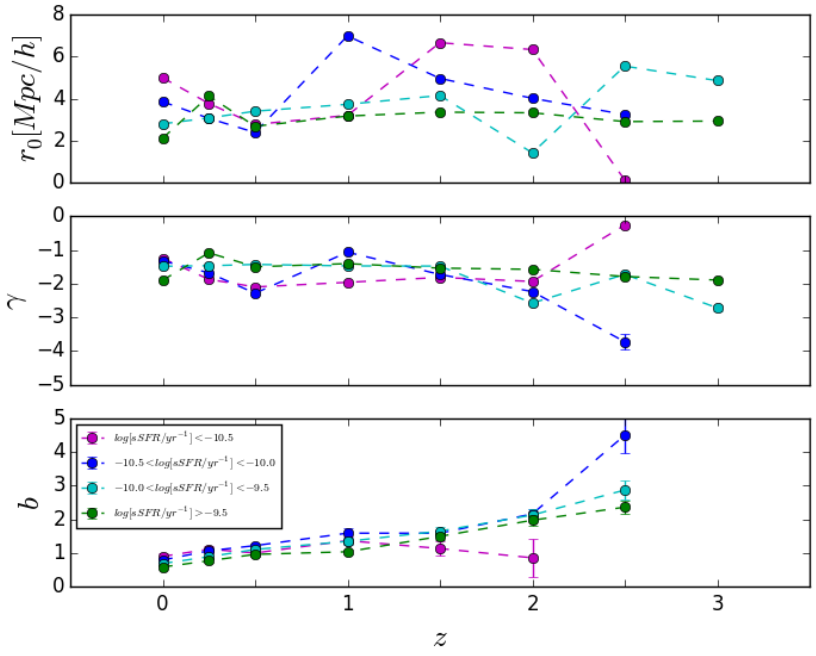

Figure 22: Evolution of clustering parameters as a function of sSFR as in Figure 12. However, colours in the order of magenta to green indicate the sSFR bins as per the legend.

4.4 NUV Luminosity

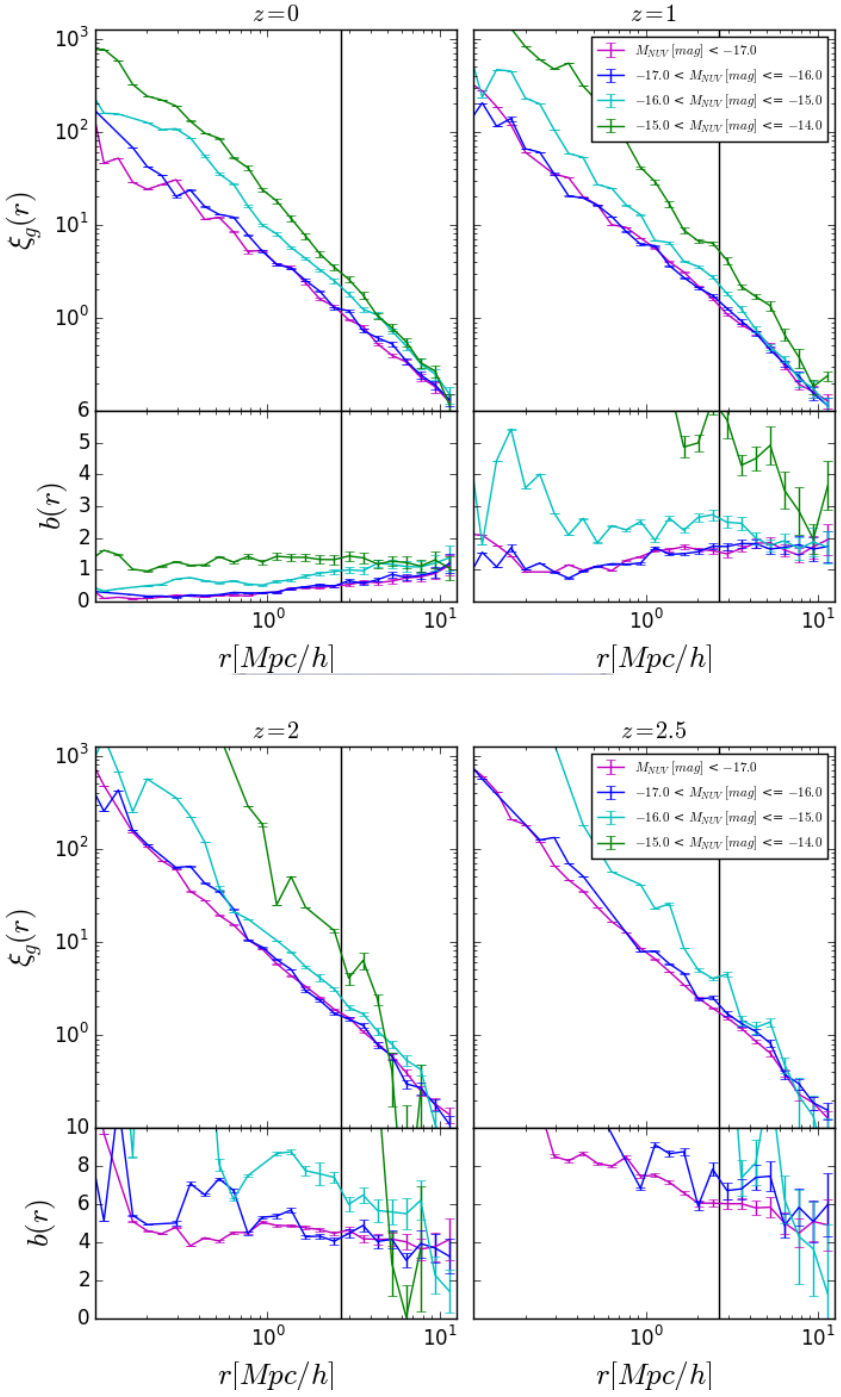

Figure 23: Clustering as a function of NUV-band luminosity for $z = 0 - 2.5$. Colours in the order of magenta to green indicate the decreasing luminosity bins. Upper panels show the galaxy TPCF and lower panels show the bias as a function of separation distance, r. Vertical black lines indicate the separation distance at which the bias factor is defined.

Young, massive stars emit strongly in the UV part of the electromagnetic spectrum, which results in a dominant feature in star-forming galaxies. For dusty galaxies, the UV is absorbed and re-emitted in the far-infrared and so the study comparing the UV and Infrared wavelengths can be used as a tool to study the dust content in the Universe as well. Because the UV is such a strong tracer of star formation rates and histories of galaxies, we expect to see the same or similar trends in the clustering dependence of galaxies on NUV luminosity to that of what we've seen in the sSFR dependence above.

Figure 23 shows the TPCF and bias as a function of separation, r, for $z = 0 - 2.5$. Each panel shows the TPCF and bias at a given redshift for a number of NUV magnitude bins in increasing order from magenta to green. At a given redshift we see an increasing clustering amplitude with decreasing NUV brightness by a factor of about 1.5 between the brightest and faintest galaxies. With this we see an increase in bias with decreasing brightness as well. This is in accordance to the similar trend found in the sSFR analysis as higher star forming galaxies i.e younger stellar population galaxies cluster less than that of low star forming galaxies.

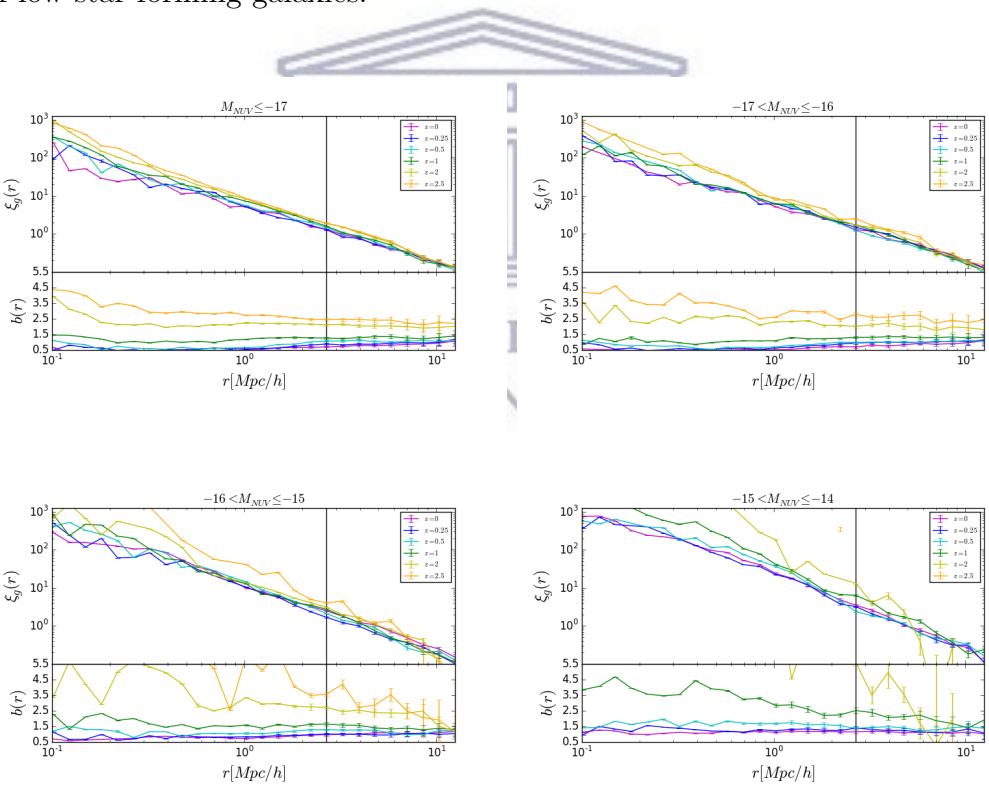

Figure 24: Clustering evolution per given M_{NUV} bin as a function of redshift. Colours in the order of magenta to red indicate increasing redshifts. Upper panels show the galaxy two point correlation function and lower panels show the bias as a function of separation distance, r. Vertical black lines indicate the separation distance at which the bias factor is defined.

In Figure 24 we see the redshift evolution of TPCF and bias as a function of separation distance, r, for a given NUV magnitude bin, and Figure 25 shows us the corresponding clustering parameters as in Figure 12 for each NUV magnitude bin as a function of redshift, colour coded as per the legend and Figure 23. For a given NUV magnitude bin in Figures 24 and 25, we see an increase in clustering amplitude with redshift, similarly to what we've seen in terms of sSFR. As redshift increases, the amplitude of the NUV clustering increases by a factor of around 1.5 by redshift 2.5, as well as creating a trend with bias that increases with redshift by a factors $\sim 1.5 - 4$.

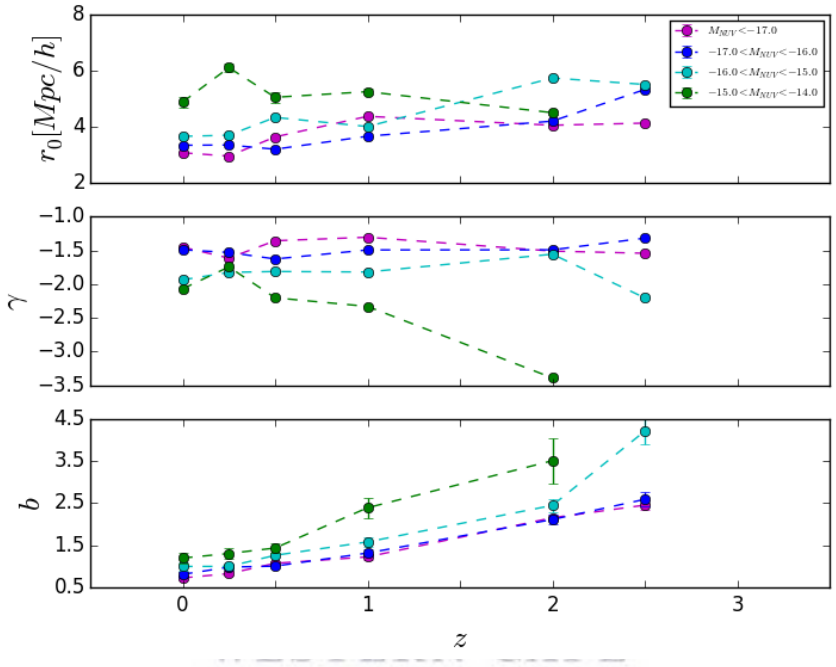

Figure 25: Evolution of clustering parameters as a function of NUV luminosity as in Figure 12. However, colours in the order of magenta to green indicate the M_{NUV} bins as per the legend.

4.5 HI Mass

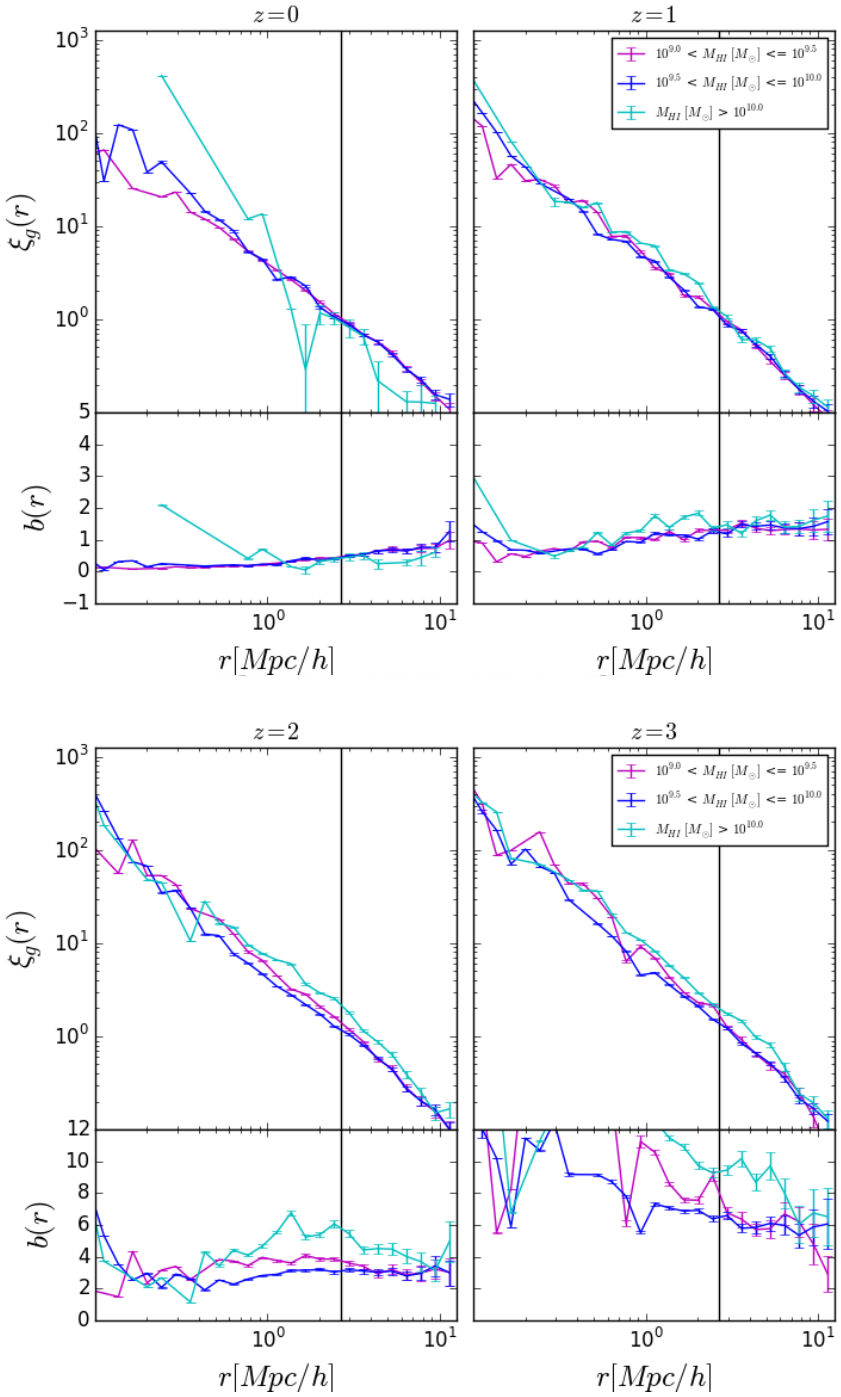

Figure 26: Clustering as a function of HI mass for $z = 0 - 3$. Colours in the order of magenta to cyan indicate the HI mass bins. Upper panels show the galaxy TPCF and lower panels show the bias as a function of separation distance, r. Vertical black lines indicate the separation distance at which the bias factor is defined.

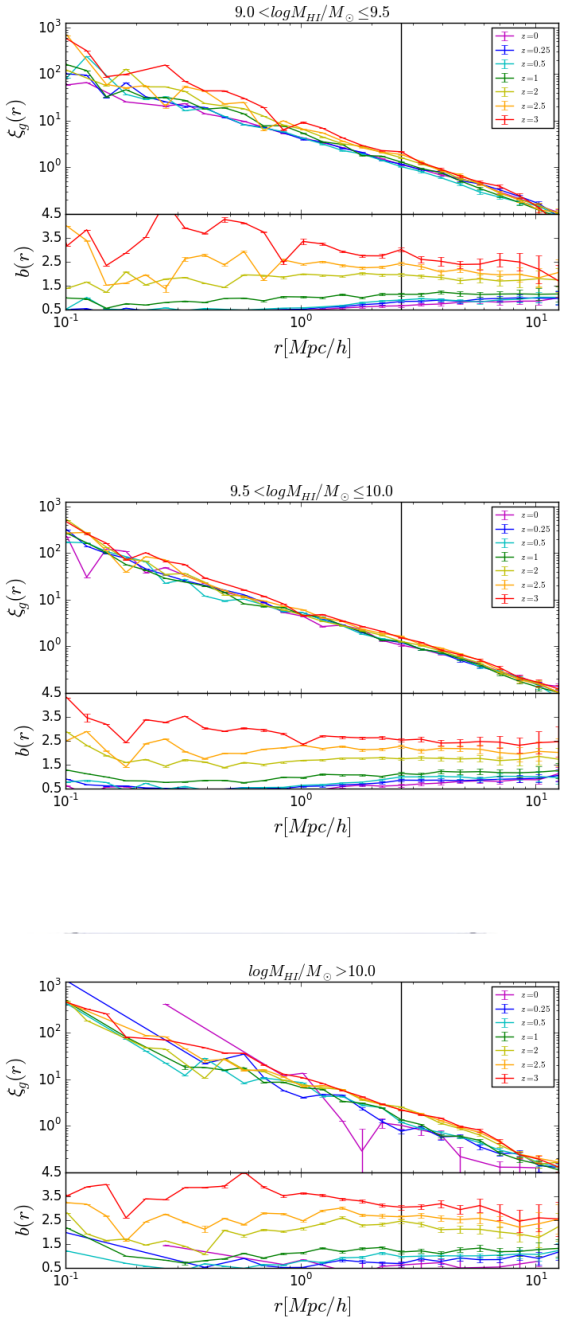

Figure 27: Clustering evolution per given M_{HI} bin as a function of redshift. Colours in the order of magenta to red indicate increasing redshifts. Upper panels show the galaxy TPCF and lower panels show the bias as a function of separation distance, r. Vertical black lines indicate the separation distance at which the bias factor is defined.

Atomic Hydrogen (HI) is particularly interesting as it provides a diffuse reservoir of potential fuel for star formation that connects the IGM to the dense molecular gas that forms stars. This then opens a new window for the process of galaxy evolution for MeerKAT surveys.

The HI atom has spin flip which emits a photon with energy corresponding to a frequency of 1.4GHz or 21cm in wavelength. This frequency lies within the radio part of the electromagnetic spectrum and it is thus that astronomers have used radio telescopes sensitive enough to detect this emission. HI science is a significant topic in current galaxy evolution fields with upcoming radio surveys such as The MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE) and Looking At the Distant Universe with the MeerKAT Array (LADUMA) which, as evident by their acronyms, make use of the MeerKAT array which is the precursor to the Square Kilometer Array (SKA) project which will be the world's largest radio telescope by completion in the 2020's.

Since MUFASA reproduces a suitable HI Mass function at $z = 0$, we are able to track HI content within galaxies and are thus able use this information to make predictions for these upcoming surveys taking us another step closer to understanding the physical processes that drive galaxy evolution. Figure 26 shows the TPCF and bias as a function of separation, r, for $z = 0 - 3$.

Each panel shows the TPCF and bias at a given redshift for a number of HI mass bins in increasing order from magenta to green. In Figure 26, at a given redshift, we don't see any obvious trend in the clustering dependence of galaxies per M_{HI} bin. Although, if we look at a given M_{HI} bin in Figure 27 which shows the redshift evolution of TPCF and bias as a function of separation distance, r, for a given HI mass bin, we see that the clustering amplitude increases with redshift and thus the bias also increases with redshift. This can also be seen in Figure 28 which shows us the clustering parameters as in Figure 12 for each HI Mass bin as a function of redshift and colour coded as per the legend and Figure 26.

Using the ALFALFA survey, Papastergis et al. (2013) studied the clustering dependence of galaxies on M_{HI} with $M_{HI} \approx 10^{8.5} - 10^{10.5}\mathrm{M}_{\odot}$ and found no obvious trends in the clustering dependence of HI alone for redshifts $0.0023 < z < 0.05$. The predictions from MUFASA show a similar conclusion for galaxies at z=0.

We will examine the cross-correlation function of HI with MUFASA in future, as was done by Papastergis et al. (2013).

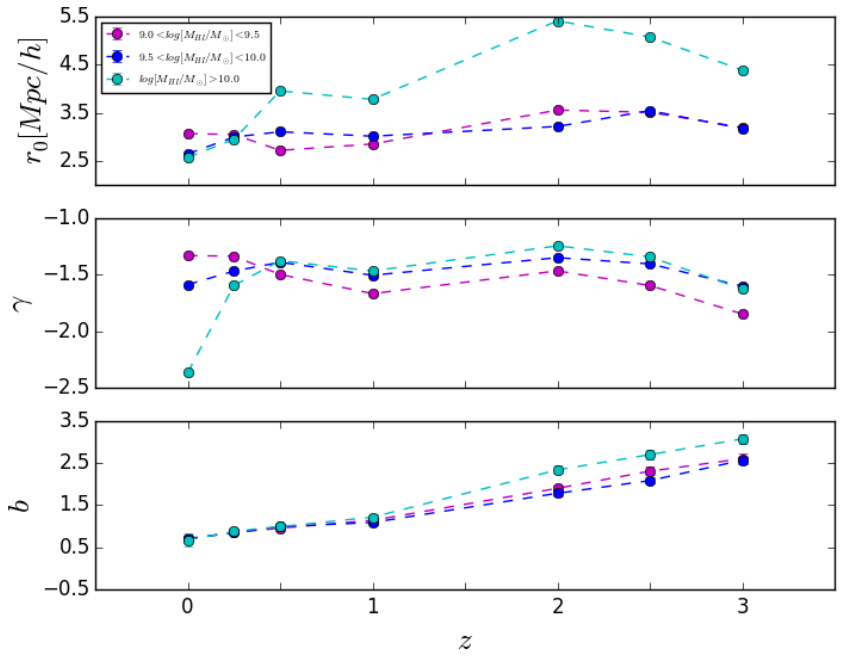

Figure 28: Evolution of clustering parameters as a function of HI Mass as in Figure 12. However, colours in the order of magenta to green indicate the M_{HI} bins as per the legend.

5 Discussion and Conclusions

Studying clustering on small scales provides constraints on the relationship between galaxies and the underlying dark matter distribution as well as describes the environment in which those galaxies live. It gives us insight into the physical processes that result in the parameters that allow for specific measures of clustering at specific scales.

We use the MUFASA cosmological hydrodynamic simulation to study small-scale (< 10 Mpc) clustering. MUFASA produces an observationally-consistent sample of galaxies by $z = 0$ with which we are able to study the clustering properties of galaxies selected through various tracers, and how these properties evolve over time.

- In agreement with Coil et al. (2016), we find that at a given redshift, even though the difference is small, higher mass galaxies cluster stronger than smaller ones. In Figures 14 and 16, at $z = 0$ the highest mass galaxies cluster about 1.2 times more strongly than the lowest mass galaxies. As we go back in time, or with increasing redshift, we see that that clustering amplitude across all $M_\star$ bins decreases along with that of the dark matter clustering implying that $M_\star$ is a highly correlated tracer of dark matter as seen in Figure 16.

- In Figures 17 and 19, we see that galaxies that are brighter in the r-band have a stronger clustering dependence than fainter r-band galaxies at a given redshift. At $z = 0$ the brightest r-band galaxies cluster about 1.3 stronger than the faintest r-band galaxies, this factor seems to increase with redshift. This was also found by Zehavi et al. (2005) and Zehavi et al. (2011). The clustering dependence of these galaxies decreases with redshift slower than that of the dark matter clustering resulting in an increasing bias with redshift, increasing by a factor of about 3 between $z = 0$ and $z = 3$. The reason for this slow increase in clustering with time is likely because bright galaxies at early times form at higher density peaks in large-scale structure, which are a more biased tracer of the underlying dark matter.

- However, the NUV-band shows the opposite trend to that of the r-band. At a given redshift in Figures 23 and 25, increase in brightness in the NUV shows a decrease in clustering with faintest galaxies clustering about 1.5 times more that brighter galaxies. As redshift increases, the amplitude of the NUV clustering increases by a factor of around 1.5 by redshift 2.5, as well as creating a trend with bias that increases with redshift by a factors $\sim 1.5 - 4$. This tells us that brighter NUV galaxies were more clustered at higher redshift and have evolved to a less dense environment today. Since NUV light is heavily extincted by dust and only present in star-forming galaxies, this indicates that star-forming

galaxies at high-z have less dust at a given level of clustering, and hence at a corresponding halo mass. This is anticipated since the metallicities and hence the dust content of high-z galaxies is lower.

- In agreement with the observations of Coil et al. (2016), we find that galaxies with higher sSFR or that are bluer in colour cluster less than that of galaxies with low sSFR or that are redder in colour. This is also evident from the results of clustering in the r- and NUV-band luminosities. In Figure 20 we see that the difference in these populations is clear at scales less than that of the clustering parameter, r_0, and we thus cannot compare these parameters, however we can see a distinct difference in the slope of the clustering, γ, where the difference between the least star-forming (red) and most star-forming (blue) galaxies is a factor of about 2 as seen in Figure 22, similarly to what was found by Zehavi et al. (2011). The evolution in the sSFR with redshift shows an increase in the clustering amplitude, similarly to that of the NUV-band. The dispersion between the two innermost bins become stronger with redshift as well, identifying the significance between the blue and red galaxy populations.

- In contrast with the clustering of galaxies selected by stellar mass, sSFR, colour, or in various bands, when selecting by the HI content of galaxies, we see no obvious trend in clustering amplitude at a given redshift. This is in agreement with observations by Papastergis et al. (2013). We do however see an evolution with redshift for the highest HI mass bin which, for the clustering length in Figure 28 increases until $z = 2$ and turns over to higher redshifts. However the other bins keep an almost constant r_0 and γ with redshift. Due to this evolution in the clustering we see a significant increase in the bias factor with time such that it increases by a factor of ~ 3 from $z = 0$ to $z = 3$.

These trends described above highlight how different multi-wavelength tracers can trace the underlying dark matter distribution in different ways. Hence looking at how the clustering evolves with specific properties can provide a stringent test of cosmological models that aim to populate galaxies into dark matter halos. For this reason, we require extensive multi-wavelength surveys to gain better constraints on galaxy formation physics.

Large-area multi-wavelength surveys are expected to improve significantly in the coming years. For instance, MIGHTEE is an upcoming MeerKAT tiered survey that will cover several fields on the sky that will total 20 deg^2 down to micro-Jansky sensitivities. The key science aims of MIGHTEE are to explore dark matter, the evolution of large-scale structure and galaxies, including AGN, star formation, and magnetic fields (Taylor and Jarvis, 2017). Because MIGHTEE will overlap with a variety of other existing and planned surveys in other-bands, we will soon able to do a multi-wavelength clustering analysis of

the galaxies residing within those fields.

With MUFASA and its follow-up simulations, we can make the matched models and predictions for SKA HI surveys such as MIGHTEE, along with the multi-wavelength counterparts, allowing for better constraints on galaxy formation physics, galaxy evolution, and dark matter. The future for clustering science is very exciting, and being able to make use of cosmological simulations provides an important complement to extract the greatest amount of information from such future surveys.

Acknowledgements

My supervisor Professor Romeel Davé for his patience, help, support, and mentorship throughout the completion of this MSc. Similarly, the support and mentorship of my co-supervisor Professor Roy Maartens. Funding for this thesis was provided by The National Astrophysics and Space Science Program (NASSP) and the South African Astronomical Observatory (SAAO).

6 References

[1] G. J Babu. Jackknife and Bootstrap. 2010. URL: https://www.iiap.res.in/astrostat/School10/LecFiles/JBabu_JackknifeBootstrap_notes.pdf.

[2] I. K. Baldry et al. "Galaxy And Mass Assembly (GAMA): the galaxy stellar mass function at z < 0.06". In: MNRAS 421.1 (2012), pp. 621–634.

[3] A. E. Bauer et al. "Galaxy And Mass Assembly (GAMA): linking star formation histories and stellar mass growth". In: MNRAS 434 (Sept. 2013), pp. 209–221. DOI: 10.1093/mnras/stt1011. arXiv: 1306.2424.

[4] Eric F. Bell et al. "The Optical and Near-Infrared Properties of Galaxies. I. Luminosity and Stellar Mass Functions". In: ApJ Supplement Series 149.2 (2003), p. 289.

[5] Andreas A. Berlind et al. "The Halo Occupation Distribution and the Physics of Galaxy Formation". In: ApJ 593.1 (2003), p. 1.

[6] Michael R. Blanton et al. "The Galaxy Luminosity Function and Luminosity Density at Redshift z = 0.1". In: ApJ 592.2 (2003), p. 819.

[7] G. L. Bryan et al. "ENZO: An Adaptive Mesh Refinement Code for Astrophysics". In: ApJs 211, 19 (Apr. 2014), p. 19. DOI: 10.1088/0067-0049/211/2/19. arXiv: 1307.2265 [astro-ph.IM].

[8] J. A. Cardelli, G. C. Clayton, and J. S. Mathis. "The relationship between infrared, optical, and ultraviolet extinction". In: ApJ 345 (Oct. 1989), pp. 245–256. DOI: 10.1086/167900.

[9] G. Chabrier. "Galactic Stellar and Substellar Initial Mass Function". In: PASP 115 (July 2003), pp. 763–795. DOI: 10.1086/376392. eprint: astro-ph/0304382.

[10] A. L. Coil et al. "PRIMUS+DEEP2: The Dependence of Galaxy Clustering on Stellar Mass and Specific Star Formation Rate at 0.2 < z < 1.2". In: ArXiv e-prints 838, 87 (Sept. 2016), p. 87. arXiv: 1609.09090.

[11] J. W. Colbert et al. "Predicting Future Space Near-IR Grism Surveys Using the WFC3 Infrared Spectroscopic Parallels Survey". In: ApJ 779, 34 (Dec. 2013), p. 34. DOI: 10.1088/0004-637X/779/1/34. arXiv: 1305.1399.

[12] R. J. Cool et al. "The Galaxy Optical Luminosity Function from the AGN and Galaxy Evolution Survey". In: ApJ 748, 10 (Mar. 2012), p. 10. arXiv: 1201.2954.

[13] Romeel Davé, Mika H. Rafieferantsoa, and Robert J. Thompson. "Mufasa: The Assembly of the Red Sequence". In: MNRAS 471.2 (2017), pp. 1671–1687.

[14] Romeel Davé, Robert Thompson, and Philip F. Hopkins. "Mufasa: Galaxy Formation Simulations with Meshless Hydrodynamics". In: MNRAS 462.3 (2016), pp. 3265–3284.

[15] Romeel Davé et al. “Mufasa: Galaxy Star Formation, Gas and Metal Properties Across Cosmic Time”. In: MNRAS 467.1 (2017), pp. 115–132.

[16] D. J. Eisenstein et al. “Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies”. In: ApJ 633 (Nov. 2005), pp. 560–574. DOI: 10.1086/466512. eprint: astro-ph/0501171.

[17] J. M. Gabor and R. Davé. “Hot gas in massive haloes drives both mass quenching and environment quenching”. In: MNRAS 447 (Feb. 2015), pp. 374–391. DOI: 10.1093/mnras/stu2399. arXiv: 1405.1043.

[18] M. L. P. Gunawardhana et al. “Galaxy And Mass Assembly: evolution of the Hα luminosity function and star formation rate density up to z < 0.35”. In: MNRAS 433 (Aug. 2013), pp. 2764–2789. DOI: 10.1093/mnras/stt890. arXiv: 1305.5308.

[19] Ed Hawkins et al. “The 2dF Galaxy Redshift Survey: correlation functions, peculiar velocities and the matter density of the Universe”. In: MNRAS 346.1 (2003), pp. 78–96.

[20] M. P. Haynes et al. “The Arecibo Legacy Fast ALFA Survey: The α.40 H I Source Catalog, Its Characteristics and Their Impact on the Derivation of the H I Mass Function”. In: AJ 142, 170 (Nov. 2011), p. 170. arXiv: 1109.0027.

[21] P. F. Hopkins et al. “A Cosmological Framework for the Co-Evolution of Quasars, Supermassive Black Holes, and Elliptical Galaxies. II. Formation of Red Ellipticals”. In: ApJ 175, 390-422 (Apr. 2008), pp. 390–422. DOI: 10.1086/524363. arXiv: 0706.1246.

[22] P. F. Hopkins et al. “Galaxies on FIRE (Feedback In Realistic Environments): stellar feedback explains cosmologically inefficient star formation”. In: MNRAS 445 (Nov. 2014), pp. 581–603. DOI: 10.1093/mnras/stu1738. arXiv: 1311.2073.

[23] R. C. Kennicutt Jr. “The Global Schmidt Law in Star-forming Galaxies”. In: ApJ 498 (May 1998), pp. 541–552. DOI: 10.1086/305588. eprint: astro-ph/9712213.

[24] S. D. Landy and A. S. Szalay. “Bias and variance of angular correlation functions”. In: ApJ 412 (July 1993), pp. 64–71. DOI: 10.1086/172900.

[25] J. Law-Smith and D. J. Eisenstein. “The Color and Stellar Mass Dependence of Small-scale Galaxy Clustering in SDSS-III BOSS”. In: ApJ 836, 87 (Feb. 2017), p. 87. arXiv: 1702.03933.

[26] V. Mehta et al. “Predicting the Redshift 2 Hα Luminosity Function Using [OIII] Emission Line Galaxies”. In: ApJ 811, 141 (Oct. 2015), p. 141. DOI: 10.1088/0004-637X/811/2/141. arXiv: 1505.07843.

[27] Antonio D. Montero-Dorta and Francisco Prada. “The SDSS DR6 luminosity functions of galaxies”. In: MNRAS 399.3 (2009), pp. 1106–1118.

[28] B. P. Moster, T. Naab, and S. D. M. White. "Galactic star formation and accretion histories from matching galaxies to dark matter haloes". In: MNRAS 428 (Feb. 2013), pp. 3121–3138. DOI: 10.1093/mnras/sts261. arXiv: 1205.5807.

[29] NumFOCUS. YT Project. 2012. URL: http://yt-project.org/.

[30] E. Papastergis et al. "The Clustering of ALFALFA Galaxies: Dependence on H I Mass, Relationship with Optical Samples, and Clues of Host Halo Properties". In: ApJ 776, 43 (Oct. 2013), p. 43. eprint: 1308.2661.

[31] P. J. E. Peebles and M. G. Hauser. "Statistical Analysis of Catalogs of Extragalactic Objects. III. The Shane-Wirtanen and Zwicky Catalogs". In: ApJ 28 (Nov. 1974), p. 19. DOI: 10.1086/190308.

[32] Planck Collaboration et al. "Planck 2015 results. XIII. Cosmological parameters". In: Aap 594, A13 (Sept. 2016), A13. DOI: 10.1051/0004-6361/201525830. arXiv: 1502.01589.

[33] V. C. Rubin, W. K. Ford Jr., and N. Thonnard. "Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/". In: ApJ 238 (June 1980), pp. 471–487. DOI: 10.1086/158003.

[34] A. Saintonge et al. "Molecular and atomic gas along and across the main sequence of star-forming galaxies". In: MNRAS 462 (Oct. 2016), pp. 1749–1756. arXiv: 1607.05289.

[35] M. Schmidt. "The Rate of Star Formation." In: ApJ 129 (Mar. 1959), p. 243. DOI: 10.1086/146614.

[36] F. Shi et al. "Mapping the Real-space Distributions of Galaxies in SDSS DR7. I. Two-point Correlation Functions". In: ApJ 833, 241 (Dec. 2016), p. 241. arXiv: 1608.02313.

[37] A Russ Taylor and Matt Jarvis. "MIGHTEE: The MeerKAT International GHz Tiered Extragalactic Exploration". In: IOP Conference Series: MS & Eng 198.1 (2017), p. 012014.

[38] Robert Thompson. CAESAR. 2016. URL: http://caesar.readthedocs.io/en/latest/.

[39] A. R. Tomczak et al. "Galaxy Stellar Mass Functions from ZFOURGE/CANDELS: An Excess of Low-mass Galaxies since z = 2 and the Rapid Buildup of Quiescent Galaxies". In: ApJ 783, 85 (Mar. 2014), p. 85. arXiv: 1309.5972.

[40] M. P. van Daalen et al. "The galaxy correlation function as a constraint on galaxy formation physics". In: MNRAS 458 (May 2016), pp. 934–949. arXiv: 1512.00008.

[41] P. G. van Dokkum and C. Conroy. "A substantial population of low-mass stars in luminous elliptical galaxies". In: nat 468 (Dec. 2010), pp. 940–942. DOI: 10.1038/nature09578. arXiv: 1009.5992.

[42] M. Vargas-Magaña et al. “An optimized correlation function estimator for galaxy surveys”. In: AAP 554, A131 (June 2013), A131. DOI: 10.1051/0004-6361/201220790. arXiv: 1211.6211.

[43] S. D. M. White and M. J. Rees. “Core condensation in heavy halos - A two-stage theory for galaxy formation and clustering”. In: MNRAS 183 (May 1978), pp. 341–358. DOI: 10.1093/mnras/183.3.341.

[44] I. Zehavi et al. “Galaxy Clustering in the Completed SDSS Redshift Survey: The Dependence on Color and Luminosity”. In: ApJ 736, 59 (July 2011), p. 59. arXiv: 1005.2413.

[45] I. Zehavi et al. “The Luminosity and Color Dependence of the Galaxy Correlation Function”. In: ApJ 630 (Sept. 2005), pp. 1–27. eprint: astro-ph/0408569.

[46] Z. Zheng, A. L. Coil, and I. Zehavi. “Galaxy Evolution from Halo Occupation Distribution Modeling of DEEP2 and SDSS Galaxy Clustering”. In: ApJ 667 (Oct. 2007), pp. 760–779. DOI: 10.1086/521074. eprint: astro-ph/0703457.

[47] F. Zwicky. “Die Rotverschiebung von extragalaktischen Nebeln”. In: Helvetica Physica Acta 6 (1933), pp. 110–127.