The effect of Baryons on the Marked Correlation Functions used for Cosmology Inference

Xiang, Zhen

The Effect of Baryons on the Marked Correlation Functions used for Cosmology Inference

THESIS

submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE
in
PHYSICS

Author :	Zhen Xiang
Student ID :	s2731126
Supervisor :	Matthieu Schaller
2^{nd} corrector :	/

Leiden, The Netherlands, March 10, 2022

The Effect of Baryons on the Marked Correlation Functions used for Cosmology Inference

Zhen Xiang

Instituut-Lorentz, Leiden University
Niels Bohrweg 2, 2333 CA Leiden, The Netherlands

March 10, 2022

Abstract

Galaxy clustering strength can be brought into play for constraining different modified cosmological models. Over the last three decades, dark matter-only simulations have been intensively used to analyse the galaxy/halo clustering compared it with hydrodynamical simulations. It is therefore crucial to understand whether hydrodynamical simulations can be used to study on this subject and which type of simulations demonstrates the higher accuracy for cosmology inference. We analyse the FLAMINGO cosmological simulations to investigate the effect of baryons by comparing dark matter-only run and full physics run. The methods for measuring clustering strength are two-point correlation function and marked correlation function using local density of halo as mark. The main result is that the marked correlation function is more sensitive when we compare different runs and it shows the difference between baryon effects is more significant than the difference between two different cosmologies. We also show that the massive halos and galaxies can cluster more than small ones and halos with high local density have higher clustering strength.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Modern galaxy survey has made extraordinary progress in recognizing the large scale structure(LSS) of the Universe. Galaxies tend to form galaxy groups, galaxy clusters and superclusters, which in turn create the cosmic web. The whole picture of LSS is a rich network of walls, filaments, nodes, and voids [1].

These observations can be explained based on the hierarchical structure formation model, which is known as Lambda cold dark matter(ΛCDM) model. A region with slightly larger initial local density will attract its surroundings to form over-dense region. When the perturbation approaches the point of over-density $\frac{\delta\rho}{\rho} \sim 1$, the expansion stops and it turns around to collapse. In ΛCDM model, baryon and collision-less dark matter are present in the over-dense region. Once the object collapses, the dark matter violently relaxes forming a dark matter halo and baryonic gas within the dark matter halo starts to form galaxy, which is the picture for a single system. When we consider more systems, on small scales, the initial density perturbations have higher amplitudes than the average. As a result, dark matter halos grow in a hierarchical way, in the sense that smaller dark matter halos form first and then merge into larger dark matter halos, which is referred as 'bottom-up' scenario [2]. On large scales, we concentrate on the its statistical properties to link to our cosmological framework.

While progress has been made in this model, there are many underlying issues that remain unsolved. For example, the CDM and the cosmological constant Λ, the main constituents of the Universe, are still unknown. Furthermore, with the development of observations of LSS, there are two major tensions between the study of the cosmic microwave background(CMB) and LSS. One of the tensions is the value of Hubble's constant between CMB and LSS measurements. The analysis of CMB from

Version of March 10, 2022– Created March 10, 2022 - 22:37

Planck results [3] favors a relatively low value of $H_0 = (67.4 \pm 0.5) km/s/Mpc$, whereas local value of Hubble constant from Hubble Space Telescope [4] is $H_0 = (73.2 \pm 1.7) km/s/Mpc$. Another tension is between the matter power spectrum measured on CMB and LSS, which is specifically the measurements on the matter density Ω_m and the matter fluctuation amplitude σ_8. Particularly, the Planck measurement of CMB anisotropies [3] infers $\Omega_m = 0.315 \pm 0.007$ and $\sigma_8 = 0.811 \pm 0,006$, while for LSS measurements namely weak gravitational lensing prefer a relatively low value, $S_8 = \sigma_8 \sqrt{\frac{\Omega_m}{0.3}} = 0.745 \pm 0.039$, which is from the Kilo Degree Survey(KiDS) [5]. There are ongoing efforts to measure the observational results in a high accuracy(eg.Euclid [6] and SKA survey [7]) and also to propose possible modified theories to resolve the tension.

As for discovering the LSS in the observation, the clustering strength of galaxies is one of most important observational constraints in cosmology, which can help understand the CDM model of galaxy formation. From the current theory, the placements of within galaxies the dark matter halos trace the underlying LSS. In large scales, the distribution of dark matter halos are mainly governed by gravity. However, on small scales, the assembly of galaxies are affected by more complex baryon physics. Galaxies are thus the biased tracer of the matter distribution [8] and it can affect the cosmological inferences when measuring the galaxy clusters.

Two-point correlation function $\eta(r)$ is the main tool to characterize the galaxy clustering, which statistically describes the spatial distribution of galaxies. From the observational study, it can be found that more massive, brighter, redder, and inactive early-type galaxies have higher clustering strength, and it is more likely to reside in the dense region [9][10][11][12][13]. This clustering measurement can provide constraint on relation between dark matter halo properties and galaxy properties, which can further constraint on galaxy formation models.

Further beyond, marked correlation functions, where galaxies are marked by some physical properties, are introduced to sensitively test how galaxy properties are affected by environment [14]. It is a high-order statistical tool which can be applied to break the degeneracies between two different cosmological models [15] and to distinguish between modified gravity and ΛCDM model [16]. Therefore, marked correlation functions nowadays are more effective tool than two-point correlation functions to help test different modified theories by comparing with observational results and further find a path to solve the tension.

The majority of LSS tests explore into the non-linear regime and hence large volume cosmological simulations are the effective tools to help un-

derstand the statistical properties of LSS, especially correlation functions mentioned before, and further to help calibrate the theoretical modelling. It has been successfully simulated a complicated web of filaments and sheets surrounding big unfilled voids [17]. Classification according to the simulated substances, it can be divided into dark matter-only simulations(N-body), and dark matter plus baryons simulations(hydrodynamical simulations). For N-body simulations, it has been widely studied and reaches to multi-trillion particle simulations [18]. The capacity of dark matter simulations to infer the matter distribution on large scales, as typically defined by the correlation function, is a great success. A good example from Millennium simulation reveals largely significant evidence that the halos of a given mass form earlier in the denser region through the marked correlation function [19].

However, using dark matter-only simulations is still crude to reproduce in detail the clustering of galaxies[20]. At present days, most simulations start turning to use cosmological hydrodynamic methods to simulate the Universe. Numerical hydrodynamic approaches, which simultaneously solve the laws of gravity, hydrodynamics, and thermodynamics for particles(or grid cells) depicting dark matter, gas, stars and supermassive blackholes(SMBHs) , are the most explicit way to decribe LSS [21]. There are three types of techniques used to solve hydrodynamical equations: Lagrangian, Eulerian or arbitrary Lagrange-Eulerian techniques. The Lagrangian methods are that particles carry the information about the fluid by summing over neighboring particles within a smoothing length. In contrast, the Eulerian methods is to partition the fluid into grid cells and then compute the physical characteristics across the cell borders. Currently, we are at the stage that differences in hydrodynamical techniques is quite small for most properties.

The baryon physics [17][21] in galaxy formation processes are typically considered in hydrodynamics simulations. The majority of these operations are carried out using effective sub-grid models.

- **Hydrodynamics and Thermal evolution**: When an area consisting of gas and dark matter that is too dense collapses, the gas starts cooling and radiating away its thermal heat. Two-body radiative processes are the dominant cooling mechanisms responsible for galaxy formation. It is the only straightforward physical process that does not require a sub-grid model to be executed.

- **Star Formation and Interstellar Medium(ISM)**: Gas has collapsed into the over-dense region, and eventually forms stars. Simulations

Version of March 10, 2022– Created March 10, 2022 - 22:37

transform this portion of the gas particles into star particles and models ISM to see how it affects star formation.

- **Star Formation feedback**: From observation, it shows that less than 10% of baryons transforms into stars. It is now widely accepted that massive stars and supernovae are involved in a variety of processes that could contribute to make star formation insufficient.
- **Supermassive Black Holes(SMBHs)**: Based on observation, simulations include models for SMBHs and seeds them in dark matter halos with large masses since their origin is not yet well-understood. These seed SMBHs grow by accreting gas or via merger.
- **Feedback from Active Galactic Nuclei(AGN)**: Most of the galaxies contain a SMBHs from observation, which shows the amount of energy released in the formation of these SMBHs must be greater than host galaxy's binding energy. This phenomena is related to AGN and has a huge impact on galaxy formation. There are two types of feedback modes: radiative mode and jet mode.
- **Radiative Transfer**: Galaxy formation can be influenced by the radiation released by the stars and AGN. Radiation alters the thermal, kinetic and chemical state of gas. From simulation, it is aimed at exploring the high redshift Universe.

Hydrodynamical simulations have numerous important advantages over other approaches(i.e.semi-analytical models). First, there is no need for additional assumption to simulate the formation history of individual halos.Particularly, without relying on simple assumptions, the interaction between dark matter and baryon components is taken into consideration [2]. Second, the mergers of dark matter halos and galaxies are evolved in a self-consistent manner. Finally, since hydrodynamical simulations has higher resolution description for baryon components, it enables to ask more detailed questions and to better compare with observation. Such as in EAGLE simulation [22], it enables to anticipate the abundance and properties of gas surrounding galaxies in great detail.

Large volume hydrodynamical simulations agree more with galaxy survey in recent years. Both in EAGLE and IllustrisTNG simulations, they reasonably simulate the stellar to halo mass ratio at relatively higher and lower redshifts [22] [23]. This ratio reaches maximum around halo masses of about $\sim 10^{12} M_{\odot}$, which is when star formation is most effective. Many properties of intracluster medium(i.e. X-ray and Sunyaev-Zeldovich scaling relation) can also be precisely simulated [24]. As for matter and galaxy

clustering, IllustrisTNG simulations study the two-point correlation function of the generated galaxies and it correlates well with the Sloan Digital Sky Survey [25]. Additionally, the galaxy scaling relations(i.e. mass-metallicity relation) can be widely reproduced by modern large volume hydrodynamical simulations[22].

Even though there are a number of achievement in modern large volume hydrodynamical simulations, the limitations are still exists. Numerical resolution is a constraint in simulations. As a result, some galaxy formation processes must still be approximated at the subgrid level, which depend on a certain number of adjustable parameters. When tried to compare to zoom-in simulations, sub-grid models are less precise. At the moment, processing power is still a major constraint, and one must choose between simulation fidelity and box size [2]. Additionally, there are also a broad collection of forecasts that change among simulations [17].

There is a concern that must be addressed here: the intensity of galaxy clustering is dependent on the various modified cosmological models. Empirically, we could use marked correlation functions, which are sensitive to the clustering of galaxies in different environments, to constrain different modified cosmological models. A large amount of studies has been present in dark matter-only simulations compared to hydrodynamical simulations. The signal can be enhanced by picking the scale maximising the difference. The main goal of this paper is to demonstrate whether the signal maximization effect in marked correlation function is also present when baryonic physics is included in the cosmological simulation model and to infer whether hydrodynamical simulations can be used for better cosmology inference.

The following is a breakdown of the thesis's structure. In chapter 2, the key aspects of the FLAMINGO simulations employed are described and how we search for galaxies and halos in simulations. 3 describes several clustering methods and their definitions. Chapter 4 explores the halo/galaxy clustering strength in different runs and compares the results among each run to examine the baryon effects. Further, the results are discussed in chapter 5. Finally, in chapter 6, we summarize our results.

Version of March 10, 2022– Created March 10, 2022 - 22:37

Chapter 2

Data

2.1 FLAMINGO Simulation

The data used for analysis is from FLAMINGO project which is the cosmological hydrodynamical simulations, which follows the ΛCDM galaxy formation model. The simulations run in a periodic cube with a side length of 200Mpc containing (particle number) particles. We have two different simulations for comparison. One is dark matter only(DMO-DES) run, which employs Newtonian gravity, and the other is full physics(referred to as REF-DES) run, which including baryon physics in the simulation. In other words, REF-DES run contain both dark matter, gas, star, black hole particles. The cosmological parameters used as initial condition are $\{\Omega_m, \Omega_b, \Omega_\Lambda, h\} = \{0.306078, 0.048600, 0.693922, 0.681\}$, which is from Dark Energy Survey year 3 (DES yr.3) results [26]. Additionally, we have another different dark matter only cosmology simulation(referred to as DMO-planck) run for comparing. The cosmological parameters are $\{\Omega_m, \Omega_b, \Omega_\Lambda, h\}$ = $\{0.314419, 0.04939, 0.685581, 0.6732\}$, which is from Planck 2018 results [3]. We summarize the cosmological parameter values of different cosmology simulations in table 2.1. The Boltzmann code CAMB is used to calculate the tranfer function [27] and the higher-order initial conditions for cosmological simulations is generated by the software package MONOFONIC MUSIC-2 [28], which also supports for massive neutrinos.

The simulations were carried out with a TreePM-SPH(Tree particle-mesh smoothed particle hydrodynamics) code SWIFT [29], where includes a new default scheme SPHENIX [30] to work well with sub-grid physics modules. The subgrid physics prescription used here is similar to BAHAMAS [31]. First, star formation technique is applied in accordance with the instructions of Schaye & Dalla Vecchia(2008) [32]. Since the simula-

tions that lack the detailed physics and resolution to simulation the multiphase interstellar medium, the critical volume density n_H exceeding a density threshold n_H^* is available for star formation, where $n_H^* = 0.1cm^{-3}$. It can reproduce the observed Kennicutt-Schmidt law by construction. As for Stellar evolution and chemical enrichment, it calculates release of 11 elements(H, He, C, N, O, Ne, Mg, Si, S, Ca and Fe) at different times by both massive and intermediate-mass stars [33]. The kinetic wind model of Dalla Vecchia & Schaye(2008) [34] is used to carry out Stellar formation feedback. The energy is dispersed as kinetic energy among the gas particles that surround newly formed stars. FLAMINGO simulations used a mass-loading factor $\eta_w = 2$ and a wind velocity $v_w = 480kms^{-1}$, equating to using around $0.25 \times 10^{51}erg$ energy from Type II supernovae.

The growth of SMBHs and the AGN feedback is performed by adopting the method of Booth & Schaye(2009) [35]. In current theoretical models, there is no precise description of how first SMBH formed. Most current cosmological simulations avoid this issue by injecting a black hole(BH) seeds into dark matter halos. In FLAMINGO simulation, BH seed particles are injected into FoF(Friends-of-Friends) groups containing at least 100 dark matter particles. Once seeded, BHs grow through Eddington-limited gas accretion and mergers. Since the simulations lack of both the precision and the physics needed to describe the multiphase interstellar medium, the Bondi-Hoyle accretion rate is frequently underestimated by a large factor α. While in Booth & Schaye(2009) model, α is a power-law function of the local density for gas over the star formation barrier. After normalization, $\alpha = 1$ for densities are equivalent to the star formation barrier, in such a way that there is no cold phase interstellar medium. From Booth & Schaye(2009), a percentage of the accreted gas's rest mass energy ϵ is utilized to heat n_{heat} nearby gas particles, by raising their temperature ΔT_{heat}. After BHs accumulate large enough energy, it will heat the n_{heat} particles by ΔT_{heat}. FLAMINGO adopted $\Delta T_{heat} = 9 \times 10^{-7}K$ and $n_{heat} = 1$ to better reproduce the intracluster medium gas mass fraction of groups and clusters. Finally, the feedback efficiency is $\epsilon \equiv \epsilon_r \epsilon_f$, where ϵ_r is the radiative efficiency and ϵ_f is the proportion of ϵ_r that relates to nearest gas. In FLAMINGO, we set $\epsilon_r = 0.1$ and $\epsilon_f = 0.15$, which has a good match with observations. We present the important subgrid physics model parameters used in FLAMINGO simulations in table 2.2.

In figure 2.1, we project the mass in x-y plane both in DMO-DES run and REF-DES run and further select a square region with side length 20 Mpc to have a better resolution. As shown in the figure, the mass distribution in REF-DES run is more diffused than DMO-DES run, which is what we expected. It is because the astrophysical processes like gas physics

Version of March 10, 2022– Created March 10, 2022 - 22:37

is taken into account in the REF-DES run. Strong shocks occur when an overly dense area of gas and dark matter collapses, increasing the gas's entropy. Finally, it will damp the total mass on small scales by 20% according to IllustrisTNG simulations [25].

Table 2.1: *cosmological parameter values for the simulations*

(1)Simulations	DMO-DES	REF-DES	DMO-Planck
(2)H_0	68.1	68.1	67.32
(3)Ω_{cdm}	0.2560	0.2560	0.2650
(4)Ω_Λ	0.6939	0.6939	0.6856
(5)Ω_b	0.0486	0.0486	0.0494
(6)n_{DM}	46656000	46656000	46656000
(7)n_{gas}	/	42440975	/
(8)Boxsize/Mpc	200	200	202.3

[1] the meaning for each columns are: (1) three different simulation runs (2) Hubble's constant (3) present-day dark matter density (4) present-day dark energy density (5) present-day baryon density (6) particle numbers of dark matter (7) particles numbers of gas (8) boxsize of the simulation

Table 2.2: *The subgrid physics model parameters used in FLAMINGO simulations*

Simulation	FLAMINGO
v_w	$480 kms^{-1}$
η_w	2
ϵ_r	0.1
ϵ_f	0.15
ΔT_{heat}	$9 \times 10^7 K$
n_{heat}	1
Accretion model	Booth & Schaye(2009) [35]
Min.FoF mass for BH seeding	100 dark matter particles

[1] v_w and η_w are the stellar feedback wind velocity and mass-loading factor. ϵ_r is the BH radiative efficiency and ϵ_f is the proportion relates to nearest gas. ΔT_{heat} is the temperature spike given to n_{heat} nearby gas particles during AGN feedback.

Version of March 10, 2022– Created March 10, 2022 - 22:37

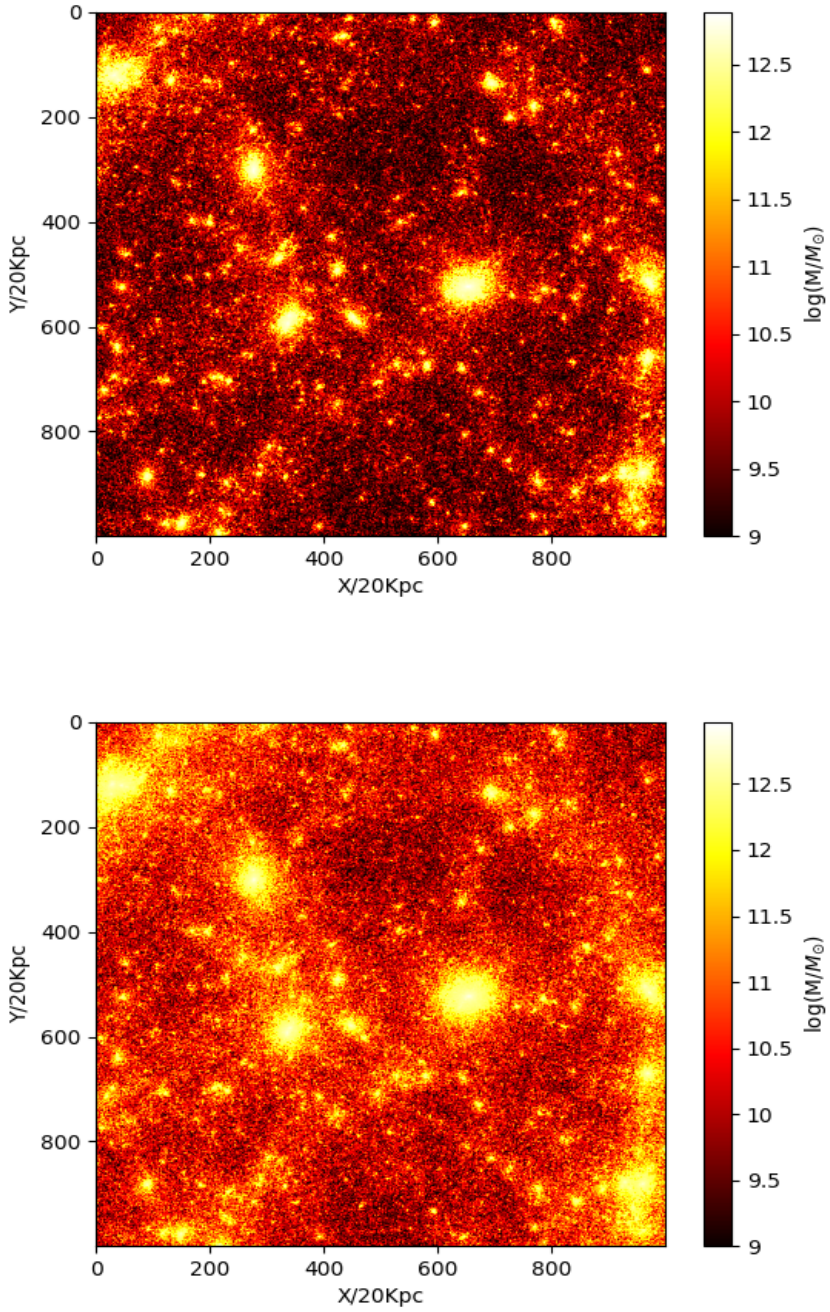

Figure 2.1: *Projected mass map of DMO-DES and REF-DES run in X-Y plane. The top panel is for DMO-DES run and bottom panel is for REF-DES run. We only show a small box of side length 20Mpc in our simulation to have a better resolution.*

Table 2.3: *outputted halo/galaxy properties*

Name	Comments
ID	Halo ID. Each halo/galaxy has the unique halo id that can used to parsed
Structuretype	For field halos are 10 which can be used to extract central halos
Mass_200crit	Overdensity mass defined by the critical density, $M_{200\rho_c}$
Xc	x coordinates of center-of-mass
Yc	y coordinates of center-of-mass
Zc	z coordinates of center-of-mass

2.2 Hunting for galaxies and halos in simulations

After obtaining the cosmological simulation's snapshots, it is important find the overdense, gravitationally self-bound regions within the simulations to understand the large scale structure and galaxy formation. For our simulations, these objects are dark matter halos and synthetic galaxies. The so-called (sub)halo finders are codes to help search for these objects. In this paper, we use VELOCIRAPTOR [36], a phase-space (sub)halo finder, to identify dark matter halos and galaxies.

VELOCIRAPTOR can also be used to get a large number of halo properties(see [36] for details), where are in the output properties profile. We list the most commonly used properties for correlation function in table 2.3. For example, the halo masses use the spherical overdensity halo masses to represent, which is

$$M_{200\rho_c} = \frac{4\pi\Delta\rho_c R_{\Delta\rho_c}}{3} \tag{2.1}$$

where ρ_c is the critical density, $\Delta = 200$, and $R_{\Delta\rho_c}$ is the radius surrounding an average density of $200\rho_c$.

Now we can use the VELOCIRAPTOR to obtain dark matter halo mass functions and galaxy stellar mass functions of FLAMINGO data showing in figure 2.2 and 2.3, which gives the number density of halos/galaxies per mass interval. For halo one, we choose $M_{200\rho_c}$ to represent. As for galaxy stellar mass, we choose the stellar mass enclosed within a radius of 30Kpc to be the galaxy stellar mass, since 30Kpc is the typical size of a galaxy. In figure 2.2 and 2.3, the halo mass function among different runs shows the similar results, since in large scales, the gravity dominates the halo and galaxy formation. It is clearly demonstrated that in each run the halo

mass function follows a power law at the low mass end from magnitude 11 to 14 and high-mass has a cut-off, which is consistent with the results of related studies[37]. Additionally, we split the total halo into central halo and subhalo, which indicates the number of central halo in each mass bin dominates the halo mass function in our simulation. Since in REF-DES run, we can easily extract the stellar mass from the halo, so we can obtain the stellar mass function pictured in figure 2.3. It follows a power-law from magnitude 10 to 11 which cuts at the high mass end, where the number density decreases nearly exponentially. It has a good agreement with the Galaxy And Mass Assembly survey(GAMA) at redshift below 0.06 [38]. However, it is worth noting that when the magnitude of halo mass is below 11 and the magnitude of stellar mass is below 11, unlike the studies, the power law trend is truncated. The reason is that the resolution of the VELOCIRAPTOR code is extremely low for low-mass dark matter halos.

On top of that, we compare the halo mass function among different runs shown in figure 2.4. In general, they follows the same power law model from magnitude 11 to 14. Nevertheless, when we look into the ratio between different runs, the ratio between REF-DES run and DMO-DES run is larger than the ratio between DMO-Planck run and DMO-DES run. It shows the level of measurement precision needed and also shows the baryon effects are larger than two different cosmologies when we consider the halo mass function.

Version of March 10, 2022– Created March 10, 2022 - 22:37

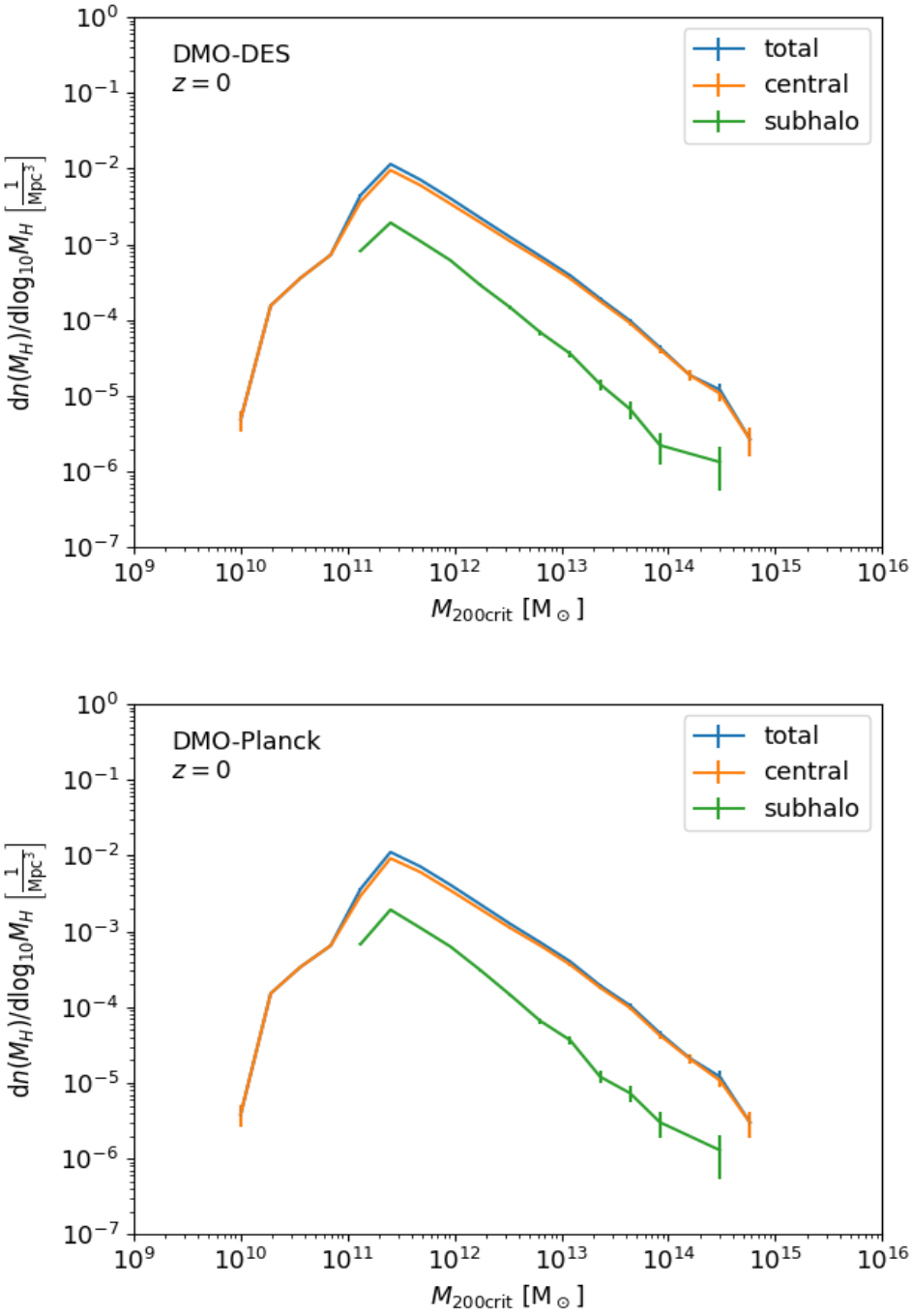

Figure 2.2: *Halo mass function of simulation data: The top panel is the halo mass function of DMO-DES data, the bottom panel is the halo mass function of DMO-Planck data.We all select the data of redshift zero which is the present time mass distribution.*

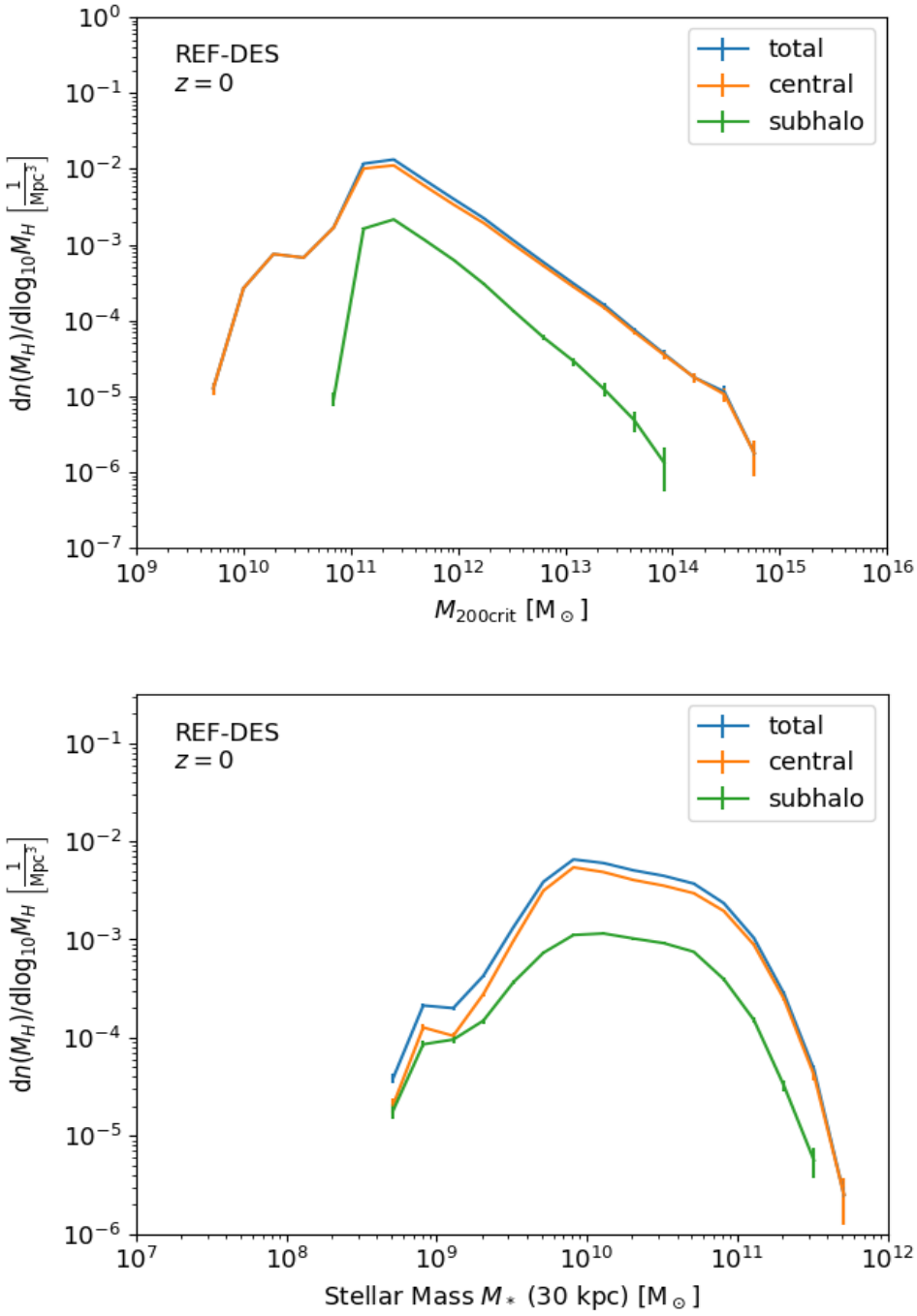

Figure 2.3: *Halo and galaxy stellar mass functions of simulation data:the top panel is the halo mass function of REF-DES data and the bottom panel is the galaxy stellar mass function of REF-DES data.We all select the data of redshift zero which is the present time mass distribution.*

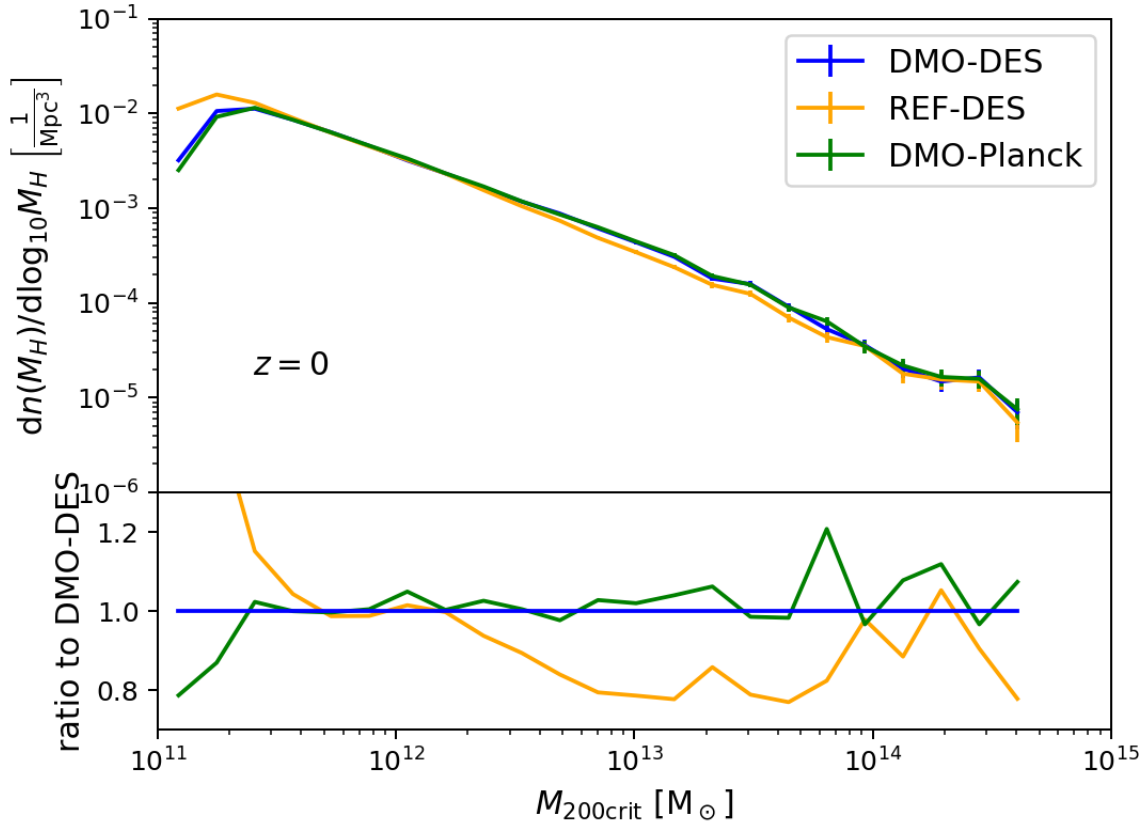

Figure 2.4: *Comparison of halo mass function among different runs at redshift 0. In the plot below, we show the ratio to DMO-DES run.*

Chapter 3

Methods

3.1 Two-point Correlation Function

The spatial distribution of galaxies is the important aspects of the galactic population. The detailed studies of the spatial pattern of galaxies are needed to reflect the overall mass map of the Universe and the intrinsic physics of galaxy formation. One thing to study the galaxies' spatial pattern is to accurately quantify the intricacy of the the cosmic web [2].

The two-point correlation function $\xi(r)$ is one of the most essential statistics for describing the geographic distribution of galaxies. It is defined as the extra likelihood above random distribution of galaxies, that is

$$dP = n\,[1 + \xi(r)]\,dV \tag{3.1}$$

Where n is the number density of galaxies, r is a given spatial separation between a pair of galaxies, and dV is the volume element [39]. On separation r, Galaxies are considered to be positively correlated if $\xi(r) > 0$. In some cases, redshifts are used to as distances to calculate the corresponding correlation function in redshift space, which is called projected correlation function. Therefore, the two-point correlation function calculated by distance is also called real-space correlation function. It follows a power law model on scales smaller than around $10h^{-1}Mpc$ given by

$$\xi(r) = (\frac{r}{r_0})^{-\gamma} \tag{3.2}$$

Where r_0 is the correlation length and γ is the slope, where r_0 is defined by $\xi(r_0) = 1$. From VIPERS survey [40], the real-space two-point correlation

function shows a power law with $r_0 = (6.97 \pm 0.37)h^{-1}Mpc$ and $\gamma = (1.86 \pm 0.06)$ in the stellar mass limit $log(M_*/M_\odot) > 10.5$ and redshift limit $0.5 < z < 0.7$.

Astrophysical studies favor estimators based on counting pairs, owing to the limitations of galaxy surveys. We introduce two common estimators used for calculating two-point correlation function. One is simple estimator,

$$\xi(r) = \frac{DD(r)}{RR(r)} - 1 \tag{3.3}$$

The other is Landy & Szalay estimator, which is broadly used, since its ability to minimize the noise from data set.

$$\xi(r) = \frac{DD(r) - 2DR(r) + RR(r)}{RR(r)} \tag{3.4}$$

Where we use normalized counts, $DD(r) = \frac{P_{DD}(r)}{\frac{(N(N-1))}{2}}$, $RR(r) = \frac{P_{RR}(r)}{\frac{(N_R(N_R-1))}{2}}$, $DR(r) = \frac{P_{DR}}{NN_R}$, with N and N_R being the total number of real data set and random set. $P_{DD}(r)$ is the number of object pairs with separations r in the actual galaxy sample, $P_{RR}(r)$ is the number to be predicted if objects were randomly scattered in space, and $P_{DR}(r)$ is the number of cross pairs of objects between two samples [41]. To eliminate noise on small scales, the number of random objects employed in the calculation is designed to be mu more than the number of real objects. In our study, we choose simple estimator for analysing, since in small scales all the estimators shows nearly same behavior, which means the difference can be negligible [41].

3.2 Marked Correlation Function

The two-point correlation function can characterize the clustering, however, each object weight equally during the measurement. We introduce the marked statistics, which can take the physical properties of the object into account. These physical properties called the marks which can be discrete or continuous values such as luminosity, colour, and stellar mass. In our work, we choose the density environment of halos as the mark to compute the marked correlation function since marked correlation functions are particularly sensitive to environmental correlations [42].

First, we need to construct the density profile for each halo. In our three dimensional box, we create a grid of size $100 \times 100 \times 100$ over the

Version of March 10, 2022– Created March 10, 2022 - 22:37

box, so that each cell's volume is round $8Mpc^3$. Then we sum up every particle's mass in each cell to obtain the total mass in each cell. Dividing the total mass of each cell by its volume, we now have a density map of the simulation. By checking each halo's center position, we assign the corresponding density of the cell to the halo which is density profile of halos. Now, we can rank order density of halos in each mass bin and select the N highest density halos and the N lowest density halos to make two-point correlation function, which can show the marked statistics.

3.3 Error Estimates

In the most perfect situation, the optimal way to analysis the error of correlation function in our simulation is that we can obtain a large number of simulations of same initial conditions. The challenge here is that it is computationally expensive, which needs a considerable amount of resources like time, processing power, memory. Therefore, there is an alternative way called internal error estimation [43], which is to utilize the single simulation data itself to calculate the error. The basic idea behind this strategy is to use an approach to perturb the original data set to produce duplicates. These duplicates follow the same distributional statistics as the data from our simulations. There are two methods that are commonly used, which are jackknife and bootstrap resampling method. In this paper, we use jackknife technique, since it has been considerably used in astronomical research [44][45].

For jackknife method, copies of the data is defined by delete each one of the N_{jk} subsamples where the original data set has been divided into. The new then containing $N_{jk} - 1$ leftover subsamples. The statistical calculation is undertaken on the data copies. In previous work, subsamples were taken by using individual objects, which can lead to incorrect errors. Therefore, we need to avoid the number of objects in each subsample to be extremely small so that copies become strongly correlated. In our work, we divide the original data set into an optimal number $N_{jk} = 10$ subsamples instead of individual objects. In such a way that each subsample has roughly hundreds of objects inside.

The covariance matrix is given by

$$C_{jk}(\xi_i, \xi_j) = \frac{N_{jk} - 1}{N_{jk}} \sum_{k=1}^{N_{jk}} (\xi_i^k(r_i) - \bar{\xi}_i)(\xi_j^k - \bar{\xi}_j) \tag{3.5}$$

Where ξ_i is the ith measurement of correlation function at bin r_i and $\bar{\xi}_i$

Version of March 10, 2022– Created March 10, 2022 - 22:37

is the average from N_{jk} copies, which is given by

$$\bar{\xi}_i = \sum_{k=1}^{N} \frac{\xi_i^k}{N_{jk}} \tag{3.6}$$

Note that the factor $N_{jk} - 1$ in equation 3.5, which takes the lack of independence between N_{jk} copies into account. Therefore, the error bar for the correlation function at the each bin is calculated by taking the square root of the diagonal components of C_{jk}.

Version of March 10, 2022– Created March 10, 2022 - 22:37

Chapter 4

Result

4.1 Two-point and Marked Correlation Function of DMO-DES run

In this section, we show the results from DMO-DES run at redshift 0. All the fit values of power-law slope γ are given in Table4.1. The two-point correlation function is shown in figure 4.1 obtained from all halos within a mass limit $12 < log(M_{halo}/M_{\odot}) < 12.5$. It indicates the two-point correlation function $\xi(r)$ of halos, which obeys a power-law model in the large scale above 1Mpc. The best-fit power-law slope is $\gamma = -1.58 \pm 0.04$. The jackknife error bars are quite narrow except those at around 1Mpc. It can be explained that the halo correlation function will show a platform feature when the scale is below 1Mpc. We can compare with the EAGLE-DMO simulation [46], although their sample and simulation cosmological parameters varies from us and there is no fit parameter in their simulation. From distance range 1Mpc to 10Mpc, it is still fit well with EAGLE-DMO simulation with a approximately power-law slope of 1.5.

We further select 2000 halos in different mass bin in DMO-DES run to measure the halos' clustering effect depending on its mass showing in figure 4.2. The mass range is $11.0 < log(M_{halo}/M_{\odot}) < 12.5$ with bin size 0.5. In each mass bin, the two-point correlation function still obeys the power-law model. It is clearly shows that as the halo mass increases, the correlation function increases within scale of 1Mpc to 30Mpc. Since the correlation function measures the clustering strength of halos, we can find more clusters in massive halos. Additionally, from the figure 4.2, on small scale around 1Mpc, the difference of two-point correlation function between each mass bin is small compared to large scales larger than 10Mpc. We can

Figure	Mass/density limit	γ
4.1	$12 < log(M_{halo}/M_{\odot}) < 12.5$	-1.58 ± 0.04
4.2	$11 < log(M_{halo}/M_{\odot}) < 11.5$	-1.44 ± 0.03
	$11.5 < log(M_{halo}/M_{\odot}) < 12$	-1.49 ± 0.03
	$12 < log(M_{halo}/M_{\odot}) < 12.5$	-1.50 ± 0.03
4.3	highest density	-1.82 ± 0.05
	lowest density	-1.40 ± 0.13

Table 4.1: *Best-fit power-law slope of all the figures in DMO-DES run*

compare this mass-dependence result with Hubble Volume ΛCDM simulation [47][48], despite it has different cosmological parameters as our DMO-DES run. It demonstrate the same trend that halo two-point correlation function increases as mass of halo increases. The scale larger than 5Mpc has good agreement for dark matter halos.

Finally, for DMO-DES run, we calculate the marked correlation function with local density mark showing in figure 4.3. We rank the total roughly 7000 central halos in mass bin of $12 < log(M_{halo}/M_{\odot}) < 12.5$ and select 2000 highest/lowest density central halos for measurement. Halo marked correlation function shows that halos with high local density cluster more than halos with low local density, which is consistent with the result in Abbas & Sheth (2005) [49]. It also presents strong difference between these two curve on small scale around 1Mpc. This difference then decreases, but still remains at large scales. It can be explained that as distance increases, the halo clustering ability correlates less on local density. Additionally, the difference between density rank can be considered as marked correlation function signal. A strong signal at a given scale suggests a higher likelihood of discovering halo pairings with bigger values for the specified attribute for both halos[39]. As shown in figure 4.3, the difference decreases, showing as distance increases, it is less likely to find halo pairs with high local density. Thus, local density can be regarded as a more straightforward indicator of environment for observation.

4.2 Two-point and Marked Correlation Function of REF-DES run

In this section, we present the results from REF-DES run at redshift 0, which includes the baryon physics in the simulation. In figure 4.4 top panel, it shows the real-space halo two-point correlation function, with

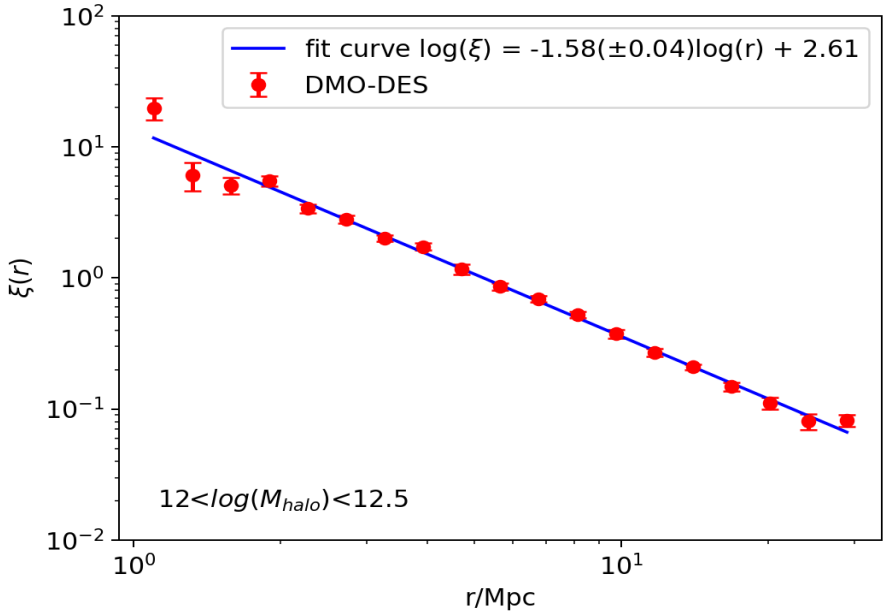

Figure 4.1: *two-point correlation function of central halos in DMO-DES run with a power-law fit by selecting mass bin between* $12 < log(M_{halo}/M_{\odot}) < 12.5$

the same selection strategy as DMO-DES run, which still holds a power-law model on the scale larger than 1Mpc. Additionally, since we add gas particles in the REF-DES run, we can obtain the real-space galaxy correlation function showing in the bottom panel of figure 4.4. The power-law fit can extend on small scale(less than 1Mpc) and the slope is $\gamma = -1.69 \pm 0.04$. We can compare this result with IllustrisTNG simulations [25], which shows the slope is $\gamma \sim -1.8$ at redshift 0. This difference may come from the selection stellar mass bin with limited sample galaxies in REF-DES run, since we only have roughly 2000 galaxies in that mass bin. The jackknife error bars are also quite large on scale smaller than 0.5Mpc.

We also investigate on mass-dependence two-point correlation function for halos and galaxies in REF-DES run, presented in figure 4.5. It clearly demonstrates that each mass bin for halos and galaxies obeys the power-law model. The galaxy correlation function can extend this effect to lower scale(less than 1Mpc) compared to halo correlation function, because there are usually numerous galaxies per halo, this is a valid guess. The halo correlation function mass dependence effect is similar to what we discussed in 4.1. As for galaxies, its stellar mass is associated with clustering. From SDSS observation [9], it is well known that the clustering

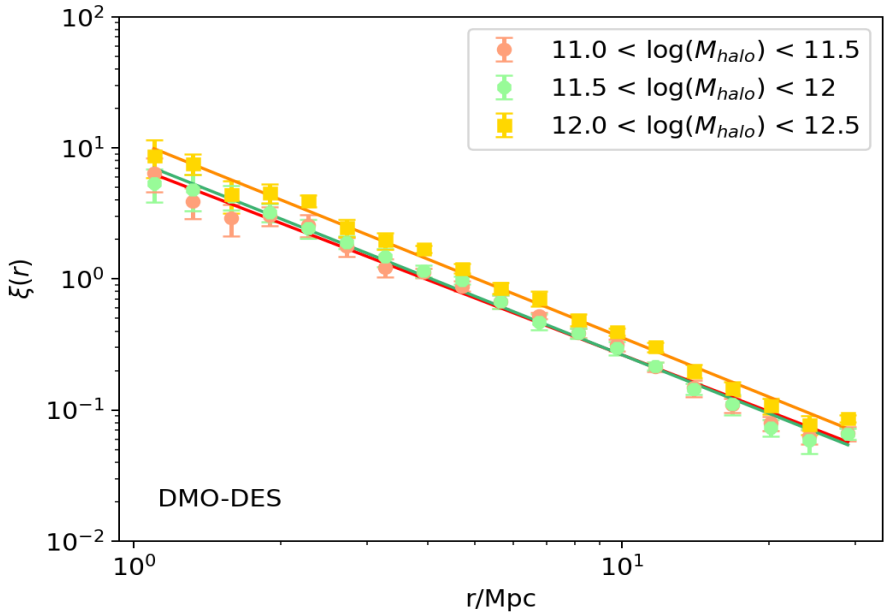

Figure 4.2: *two-point correlation function of central halos in DMO-DES run selected by different mass bin.*

strength of galaxies increases with stellar mass. From the lower panel of plot 4.5, it distinctly manifests that the correlation function increases when increasing the stellar mass of galaxies, which inferring the higher clustering strength of galaxies with higher stellar mass. We also notice that the difference between each curve is decreasing as we increase the distance scale, which means it is easier to detect this effect on smaller scales.

For REF-DES run, we finally measured the marked correlation function with local density mark presenting in figure 4.6. Comparably, we rank the total roughly 7000 halos in halo mass bin of $12 < log(M_{halo}/M_{\odot}) < 12.5$ and with same selection strategy as DMO-DES run. Figure 4.6 indicates the similar results in section 4.1, which reveals that the strong marked correlation function signal exits in small scales when baryon physics include in simulation.

4.3 Baryon effect

To further check the effect of baryon, we compare our results among different runs in figure 4.7. In the top panel, it shows the two-point halo correla-

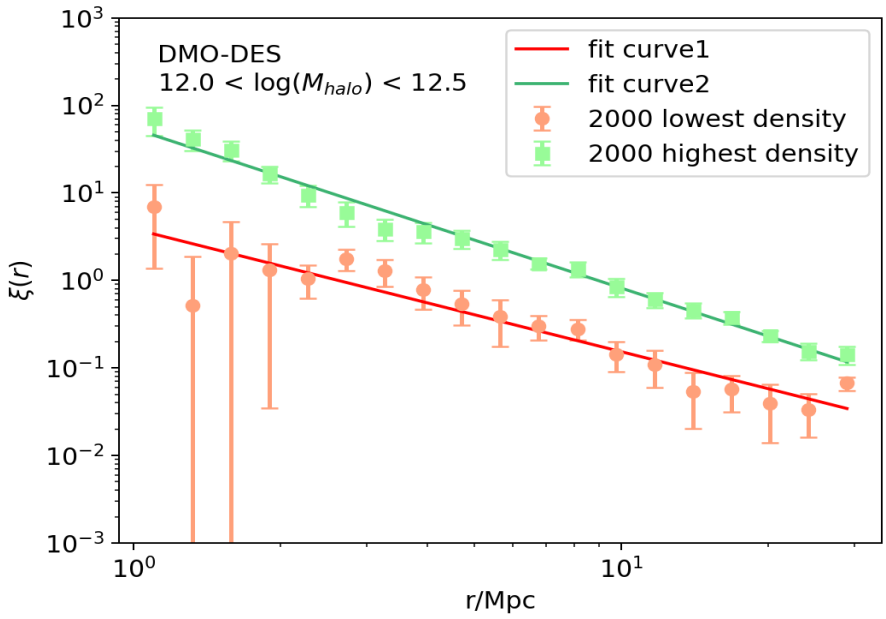

Figure 4.3: *Marked correlation function of central halos in DMO-DES run in mass bin* $12 < log(M_{halo}/M_{\odot}) < 12.5$*, choosing local density of halos as mark*

Figure	Mass/density limit	γ
4.4	$12 < log(M_{halo}/M_{\odot}) < 12.5$	-1.49 ± 0.03
	$11 < log(M_{*}/M_{\odot}) < 11.5$	-1.69 ± 0.04
4.5	$11 < log(M_{halo}/M_{\odot}) < 11.5$	-1.43 ± 0.04
	$11.5 < log(M_{halo}/M_{\odot}) < 12$	-1.44 ± 0.04
	$12 < log(M_{halo}/M_{\odot}) < 12.5$	-1.57 ± 0.03
	$10 < log(M_{*}/M_{\odot}) < 10.5$	-1.29 ± 0.06
	$10.5 < log(M_{*}/M_{\odot}) < 11$	-1.46 ± 0.05
	$11 < log(M_{*}/M_{\odot}) < 11.5$	-1.70 ± 0.05
4.6	highest density	-1.82 ± 0.04
	lowest density	-1.31 ± 0.13

Table 4.2: *Best-fit power-law slope of all the figures in REF-DES run*

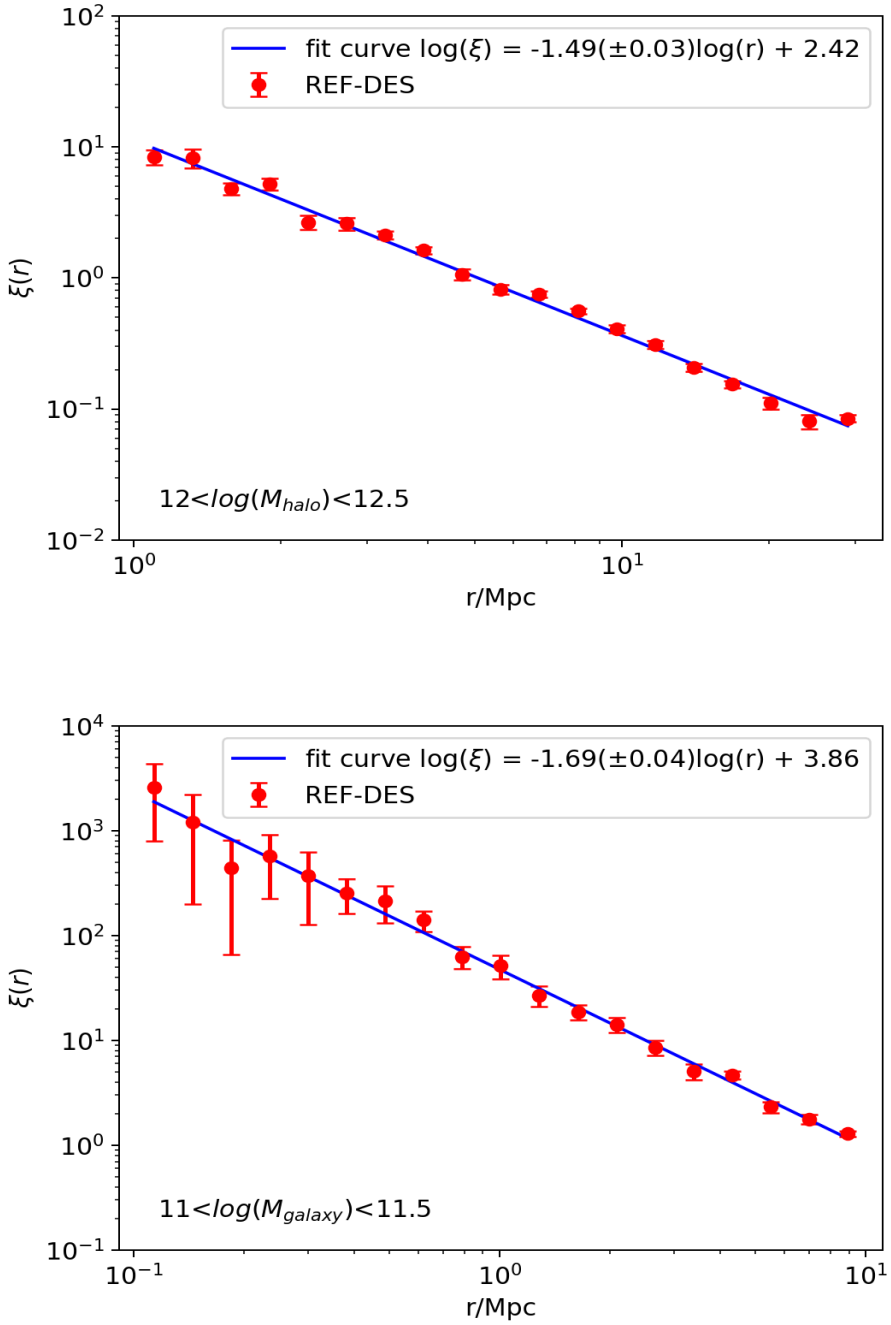

Figure 4.4: *Two-point correlation function of central halos with a power-law fit by selecting mass bin between* $12 < log(M_{halo}/M_{\odot}) < 12.5$*(top panel). Two-point correlation function of galaxies with a power-law fit by selecting mass bin between* $11 < log(M_{*}/M_{\odot}) < 11.5$*(bottom panel).*

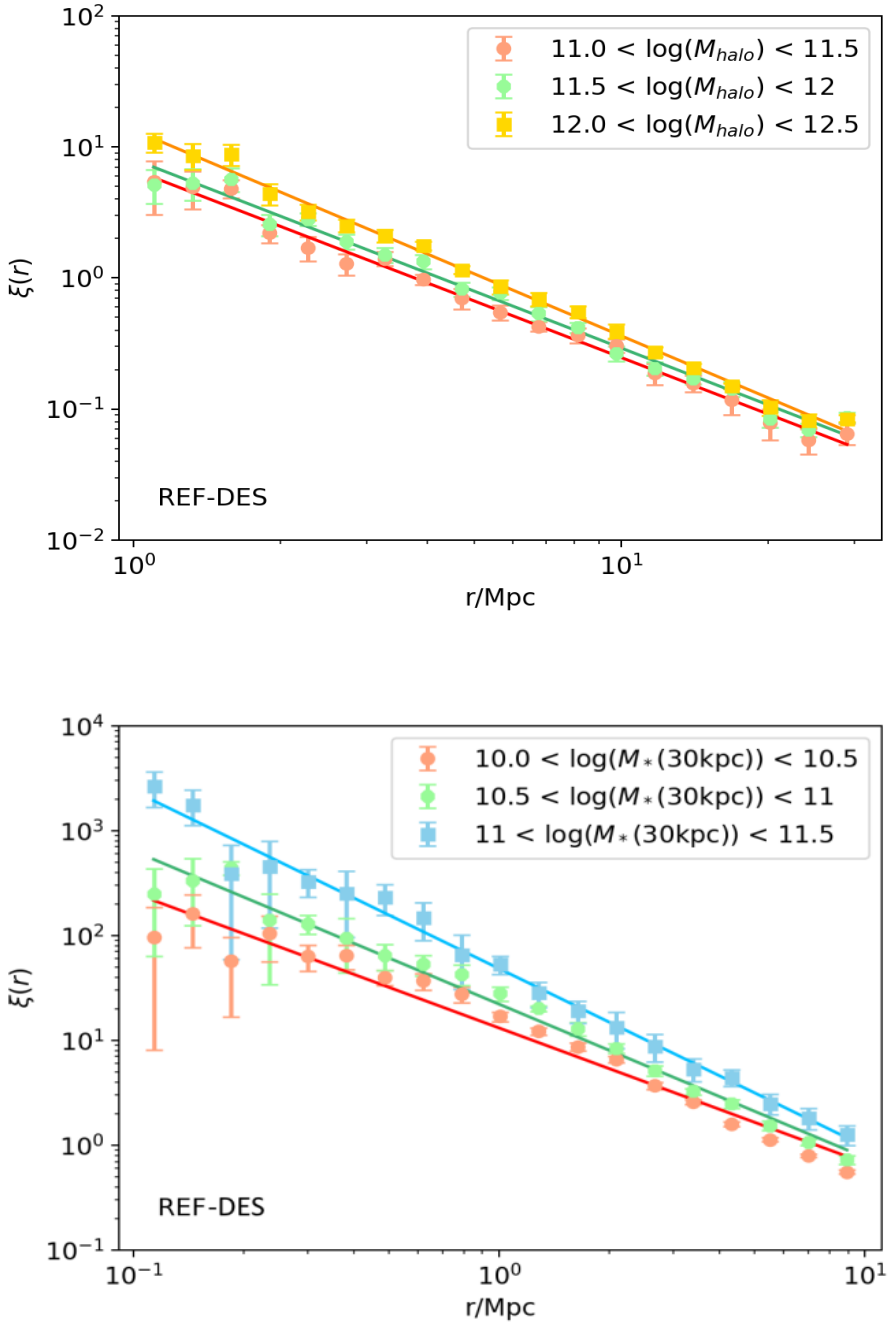

Figure 4.5: *Two-point correlation function of central halos with a power-law fit by selecting different mass bin between* $11 < log(M_{halo}/M_{\odot}) < 12.5$*(top panel). two-point correlation function of galaxies with a power-law fit by selecting different mass bin between* $10 < log(M_*/M_{\odot}) < 11.5$*(bottom panel).*

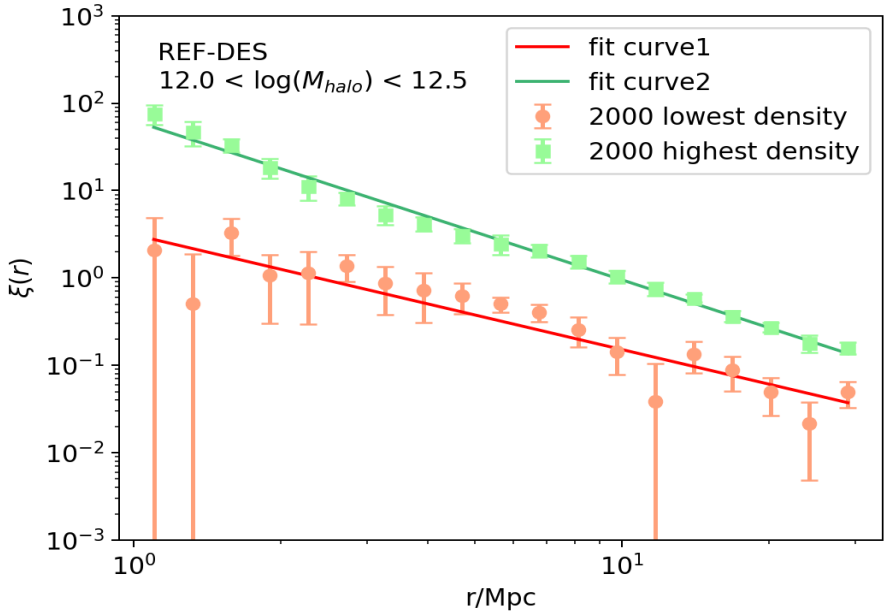

Figure 4.6: *Marked correlation function of central halos in REF-DES run in mass bin* $12 < log(M_{halo}/M_{\odot}) < 12.5$*, choosing local density of halos as mark*

tion function in the mass bin of $12 < log(M_{halo}/M_{\odot}) < 12.5$ for three runs. Specifically, for REF-DES run, we reduce the mass by 20%, which means we select the halo mass in the mass bin of $11.9 < log(M_{halo}/M_{\odot}) < 12.4$. The reason to reduce the halo mass in REF-DES run is that the same halo is less massive in REF-DES run compared to DMO-DES run as we can see in the figure 2.1. It is necessary to reduce the halo mass in REF-DES run, if we intend to select the same halo in each run for comparison. In order to check difference among each run, we show the ratio to DMO-DES below. The ratio to DMO-DES of REF-DES run is similar to DMO-Planck run which means the two-point correlation is not sensitive enough to distinguish different runs. Therefore, in the bottom panel, we demonstrates the Marked halo correlation function in the same mass bin as before and halo mass bin in REF-DES run is reduced by 20%. For marked correlation function, we select the 2000 halos with highest local density in the given mass bin since its error bar is narrow compared to the lowest density one shown in figure 4.3 and 4.6. Similarly, it still holds the power-law model among each run, so we plot the ratio to DMO-DES below as supplement. It is clearly illustrated that the ratio to DMO-DES in REF-DES run is far from 1 compared to DMO-Planck run which indicates the difference for

adding baryon effect in the simulation is more significant than two different cosmologies. Hence, the hydrodynamical simulation can give us a higher level of measurement precision used for comparing with observations and testing different cosmology models.

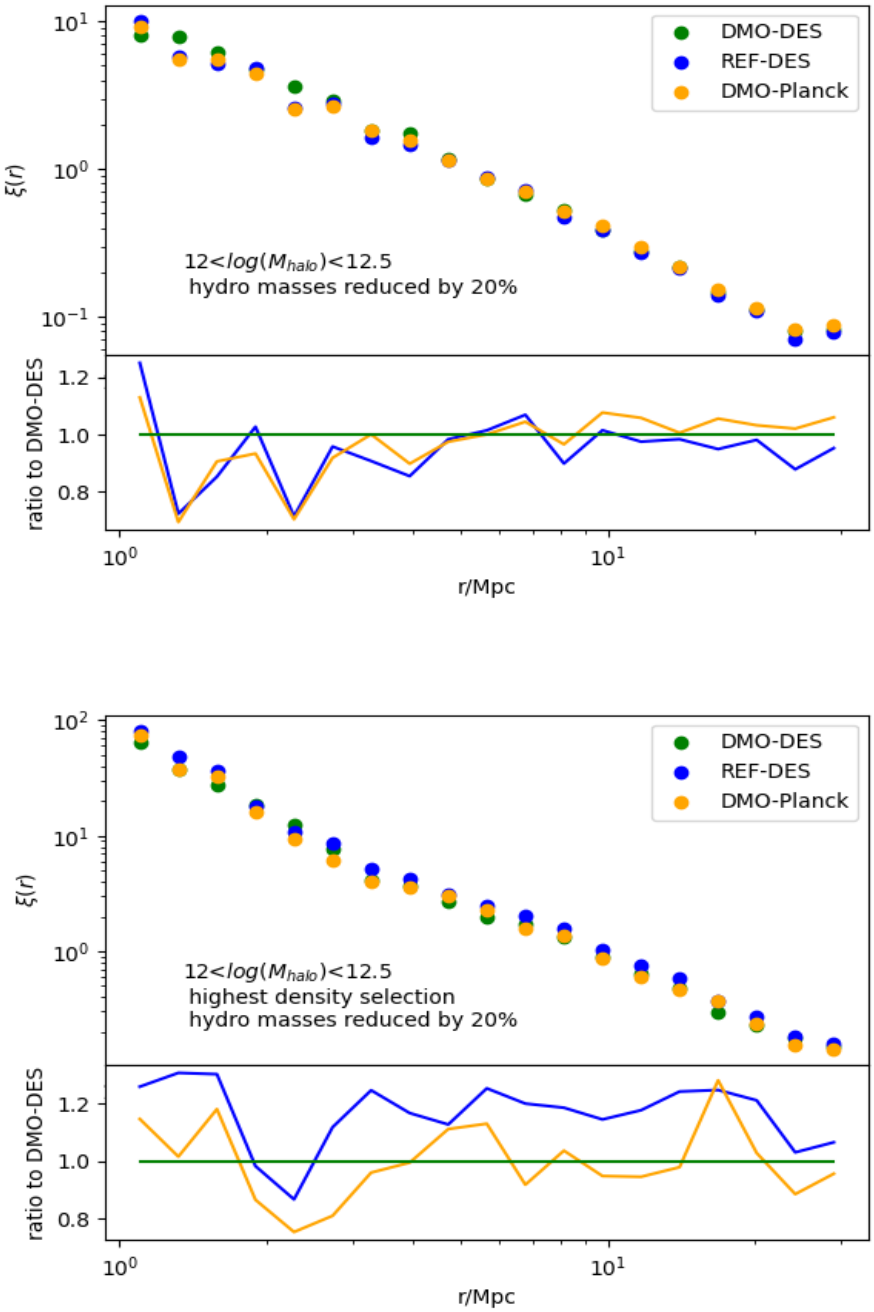

Figure 4.7: *Comparison among different runs by using correlation function and marked correlation function*

Chapter 5

Discussion

5.1 Distribution of halos is correlated with dark matter over-densities

From marked correlation function in figure 4.3 and 4.6, if we take the environmental properties into account, it shows that halos have stronger correlation with environment compared with its intrinsic properties(see figure 4.2 and 4.5). Specifically, in our study, we choose local density of each halo as mark. It demonstrates that dark matter over-densities are closely associate with the distribution of large halos. Since the galaxy stellar mass is correlated with halo mass, hence it is consistent with well-known observational results that most luminous and massive galaxies occupy the most densely populated region [50][51]. This result can be explained by the hierarchical structure formation. Small perturbations in the density field evolve into the LSS due to gravity, which forms hierarchical. In other words, small dark matter halos form first and then merge to form larger one. Galaxies properties are determined by the its host dark matter halo's mass and formation history. Further, these properties of dark matter halos are related to the initial density field, where they formed. Hence, in the over-dense region, there are more likely to find halo pairs, specifically for scale larger than 1Mpc [49].

5.2 Baryon effects

In figure 4.7, we investigate the sensitivity of DMO-DES simulation to two other different numerical models. Due to the hierarchical structure for-

mation model, marked correlation function is more sensitive to compare between different runs. From scale 1Mpc to 30Mpc, as expected, the variations for REF-DES run in marked correlation function is larger than the DMO-Planck one. The variations for REF-DES run is approximately 20 percent, likely owing to AGN feedback, which is larger than the current measurement errors for marked correlation functions. Future large scale structure survey(e.g. Euclid [6]) can perform high-sensitivity and high-resolution measurement of halo and galaxy clustering, so that the baryon effects can be distinguished when using marked correlation functions at these scales. Therefore, baryon effects need to be considered, especially when investigating on the marked correlation function of halo and galaxy clusters, to infer cosmological parameters of ΛCDM model.

Version of March 10, 2022– Created March 10, 2022 - 22:37

Chapter 6

Conclusion

In our study, we have analysed a state-of-the-art cosmological simulations called FLAMINGO to compare between dark matter-only simulations with hydrodynamical simulations by focusing on galaxy clustering strength. The simulations follow a DES yr.3 or a Planck 2018 cosmology showing in table 2.1, which were carried out with a TreePM-SPH code SWIFT. Specifically, the full physics REF-DES run considers subgrid model including star formation and its feedback, SMBHs, AGN feedback, ISM, and gas cooling. The main subgrid model parameters are shown in table 2.2.

The halos and galaxies in simulation snapshots are searched by VELOCIRAPTOR, and the used properties for correlation function are listed in table 2.3. After knowing the mass for each halo and galaxy, the halo/galaxy mass function can be obtained. In figure 2.4, by comparing the halo mass function among each run, it shows the baryon effects are more significant than two different cosmologies when compared using halo mass function.

It should be noted that we choose the local density of halo as the mark to calculate the marked correlation function, since marked correlation functions are particularly sensitive to environmental correlations. In our study, we construct the density profile for each halo by creating a grid over the box and calculating the density for each cell.

After obtaining the properties of each halo/galaxy, we compare the two-point and marked correlation function at redshift 0 among three runs. Our key results can be summarized as follows.

- The halo two-point correlation function in different runs follows a power-law model on scales larger than 1 Mpc(figure 4.1). In the mass limit $12 < log(M_{halo}/M_{\odot}) < 12.5$, the best-fit power-law slope is ~ -1.6. The galaxy two-point correlation function can extend the

power-law fit to small scales, which is below 1 Mpc(figure 4.4). In the mass limit $11 < log(M_*/M_\odot) < 11.5$, the best-fit power-law slope is ~ -1.7.

- When we connect the halo/galaxy properties to the two-point correlation function, it shows that the massive halos and galaxies cluster more than small ones(figure 4.2 and 4.5). It is consistent with the SDSS survey that the intensity of galaxies' clustering grow as stellar mass increases. From the marked correlation functions, halo with high local density have higher clustering strength(figure 4.3 and 4.6). It also reveals that the strong marked correlation function signal exists in small scales among different runs.

- Marked correlation function has a higher sensitivity when we compare different runs. The difference between baryon effects in marked correlation function is more significant than the difference between two different cosmologies, which means hydrodynamical simulations can give us a higher accuracy for breaking the degeneracy for different cosmology models.

Our measurements are the first of this kind in comparing the clustering strength between dark matter-only simulations and hydrodynamical simulations. These measurements can be a beneficial reference for clustering studies, specially surveys of the next cosmic stage, such as Euclid [6], to constrain cosmological models. However, there are several future outlooks for further developments:

- The jackknife errorbar is still non-negligible especially in galaxy two-point correlation function and marked correlation function with selecting lowest local density. One solution to narrow the errorbar is to include more objects by selecting a preferable mass bin. Moreover, in our study, the number of subsamples for jackknife method is directly selected as 10, since the number of halos/galaxies in each subsample are large than the objects in the selected maximum scale box($\sim 30Mpc$). Therefore, the number of subsamples for jackknife method can still be tested to reach higher accuracy. Additionally, in our study, we split the sample into subsamples directly by quantity. A better way for estimating errors on galaxy clustering in three dimensions is to use volume resampling [52]. Hence, for future development, it can make duplicates of data sets by picking subvolumes from the original.

- To our knowledge, in large scale structure, the relation between luminous objects(eg. galaxies, quasars and clusters) and the total matter in Universe is called bias [53]. In our study, we can measure the bias of all the stellar mass in our FLAMINGO simulations to the total matter mass comparing with observations.

- In this study, we obtained all our results using snapshots of redshift 0. In future developments, snapshots from different redshift can be used to measure the halo/galaxy clustering strength. The slope of the two-point galaxy correlation varying with redshift can be comparable to galaxy surveys. Additionally, the correlation function of different mass components(stellar mass, dark matter, black holes) can be measured to trace the evolution of different species.

- For our marked correlation function in this study, we only investigated the highest local density halo and lowest local density halo in the given mass bin. Empirically, the marked correlation function is the ratio between weighted and unweighted two-point correlation function, which can be measured for future studies. Additionally, different marks(eg.luminosity, star formation rate, and stellar mass) can be considered into the measurement to find the most reliable tracer of halo/galaxy properties.

Version of March 10, 2022– Created March 10, 2022 - 22:37

Bibliography

[1] M. Alpaslan et al., *Galaxy And Mass Assembly (GAMA): the large-scale structure of galaxies and comparison to mock universes*, Monthly Notices of the Royal Astronomical Society **438**, 177 (2014).

[2] H. Mo, F. Van den Bosch, and S. White, *Galaxy formation and evolution*, Cambridge University Press, 2010.

[3] P. Collaboration et al., *Planck 2018 results. VI. Cosmological parameters*, (2020).

[4] A. G. Riess et al., *A 2.4% determination of the local value of the Hubble constant*, The Astrophysical Journal **826**, 56 (2016).

[5] H. Hildebrandt et al., *KiDS-450: Cosmological parameter constraints from tomographic weak gravitational lensing*, Monthly Notices of the Royal Astronomical Society **465**, 1454 (2017).

[6] R. Laureijs et al., *Euclid definition study report*, arXiv preprint arXiv:1110.3193 (2011).

[7] P. E. Dewdney, P. J. Hall, R. T. Schilizzi, and T. J. L. Lazio, *The square kilometre array*, Proceedings of the IEEE **97**, 1482 (2009).

[8] N. Kaiser, *On the spatial correlations of Abell clusters*, (1984).

[9] I. Zehavi et al., *Galaxy clustering in the completed SDSS redshift survey: the dependence on color and luminosity*, The Astrophysical Journal **736**, 59 (2011).

[10] R. A. Skibba et al., *Dark matter halo models of stellar mass-dependent galaxy clustering in PRIMUS+ DEEP2 at 0.2¡ z¡ 1.2*, The Astrophysical Journal **807**, 152 (2015).

[11] R. A. Skibba et al., *PRIMUS: Galaxy Clustering as a Function of Luminosity and Color at 0.2¡ z¡ 1*, The Astrophysical Journal **784**, 128 (2014).

[12] B. Meneux et al., *The VIMOS-VLT Deep Survey-The evolution of galaxy clustering per spectral type to z 1.5*, Astronomy & Astrophysics **452**, 387 (2006).

[13] N. Mostek, A. L. Coil, M. Cooper, M. Davis, J. A. Newman, and B. J. Weiner, *The DEEP2 Galaxy Redshift Survey: Clustering Dependence on Galaxy Stellar Mass and Star Formation Rate at z 1*, The Astrophysical Journal **767**, 89 (2013).

[14] R. K. Sheth, *The halo-model description of marked statistics*, Monthly Notices of the Royal Astronomical Society **364**, 796 (2005).

[15] M. White and N. Padmanabhan, *Breaking halo occupation degeneracies with marked statistics*, Monthly Notices of the Royal Astronomical Society **395**, 2381 (2009).

[16] M. White, *A marked correlation function for constraining modified gravity models*, Journal of Cosmology and Astroparticle Physics **2016**, 057 (2016).

[17] M. Vogelsberger, F. Marinacci, P. Torrey, and E. Puchwein, *Cosmological simulations of galaxy formation*, Nature Reviews Physics **2**, 42 (2020).

[18] R. Angulo, V. Springel, S. White, A. Jenkins, C. Baugh, and C. Frenk, *Scaling relations for galaxy clusters in the Millennium-XXL simulation*, Monthly Notices of the Royal Astronomical Society **426**, 2046 (2012).

[19] G. Harker, S. Cole, J. Helly, C. Frenk, and A. Jenkins, *A marked correlation function analysis of halo formation times in the Millennium Simulation*, Monthly Notices of the Royal Astronomical Society **367**, 1039 (2006).

[20] V. Springel, C. S. Frenk, and S. D. White, *The large-scale structure of the Universe*, nature **440**, 1137 (2006).

[21] R. S. Somerville and R. Davé, *Physical models of galaxy formation in a cosmological framework*, Annual Review of Astronomy and Astrophysics **53**, 51 (2015).

[22] J. Schaye et al., *The EAGLE project: simulating the evolution and assembly of galaxies and their environments*, Monthly Notices of the Royal Astronomical Society **446**, 521 (2015).

[23] A. Pillepich et al., *First results from the IllustrisTNG simulations: the stellar mass content of groups and clusters of galaxies*, Monthly Notices of the Royal Astronomical Society **475**, 648 (2018).

[24] S. Planelles, S. Borgani, D. Fabjan, M. Killedar, G. Murante, G. Granato, C. Ragone-Figueroa, and K. Dolag, *On the role of AGN feedback on the thermal and chemodynamical properties of the hot intracluster medium*, Monthly Notices of the Royal Astronomical Society **438**, 195 (2014).

[25] V. Springel et al., *First results from the IllustrisTNG simulations: matter and galaxy clustering*, Monthly Notices of the Royal Astronomical Society **475**, 676 (2018).

[26] T. Abbott et al., *Dark energy survey year 3 results: cosmological constraints from galaxy clustering and weak lensing*, Physical Review D **105**, 023520 (2022).

[27] A. Lewis, A. Challinor, and A. Lasenby, *Efficient computation of cosmic microwave background anisotropies in closed Friedmann-Robertson-Walker models*, The Astrophysical Journal **538**, 473 (2000).

[28] O. Hahn, C. Rampf, and C. Uhlemann, *Higher order initial conditions for mixed baryon–CDM simulations*, Monthly Notices of the Royal Astronomical Society **503**, 426 (2021).

[29] M. Schaller, P. Gonnet, P. W. Draper, A. B. Chalk, R. G. Bower, J. Willis, and L. Hausammann, *SWIFT: SPH with inter-dependent fine-grained tasking*, Astrophysics Source Code Library , ascl (2018).

[30] J. Borrow, M. Schaller, R. G. Bower, and J. Schaye, *Sphenix: Smoothed Particle Hydrodynamics for the next generation of galaxy formation simulations*, Monthly Notices of the Royal Astronomical Society (2020).

[31] I. G. McCarthy, J. Schaye, S. Bird, and A. M. C. Le Brun, *The BAHAMAS project: calibrated hydrodynamical simulations for large-scale structure cosmology*, Monthly Notices of the Royal Astronomical Society , stw2792 (2016).

[32] J. Schaye and C. Dalla Vecchia, *On the relation between the Schmidt and Kennicutt–Schmidt star formation laws and its implications for numerical simulations*, Monthly Notices of the Royal Astronomical Society **383**, 1210 (2008).

[33] R. P. Wiersma, J. Schaye, T. Theuns, C. Dalla Vecchia, and L. Tornatore, *Chemical enrichment in cosmological, smoothed particle hydrodynamics simulations*, Monthly Notices of the Royal Astronomical Society **399**, 574 (2009).

[34] C. Dalla Vecchia and J. Schaye, *Simulating galactic outflows with kinetic supernova feedback*, Monthly Notices of the Royal Astronomical Society **387**, 1431 (2008).

[35] C. Booth and J. Schaye, *Cosmological simulations of the growth of supermassive black holes and feedback from active galactic nuclei: method and tests*, Monthly Notices of the Royal Astronomical Society **398**, 53 (2009).

[36] P. J. Elahi, R. Cañas, R. J. Poulton, R. J. Tobar, J. S. Willis, C. d. P. Lagos, C. Power, and A. S. Robotham, *Hunting for Galaxies and Halos in simulations with VELOCIraptor*, Publications of the Astronomical Society of Australia **36** (2019).

[37] A. Jenkins, C. Frenk, S. D. White, J. M. Colberg, S. Cole, A. E. Evrard, H. Couchman, and N. Yoshida, *The mass function of dark matter haloes*, Monthly Notices of the Royal Astronomical Society **321**, 372 (2001).

[38] I. K. Baldry et al., *Galaxy And Mass Assembly (GAMA): the galaxy stellar mass function at z¡ 0.06*, Monthly Notices of the Royal Astronomical Society **421**, 621 (2012).

[39] U. Sureshkumar et al., *Galaxy and Mass Assembly (GAMA)-Tracing galaxy environment using the marked correlation function*, Astronomy & Astrophysics **653**, A35 (2021).

[40] F. Marulli et al., *The VIMOS Public Extragalactic Redshift Survey (VIPERS)-Luminosity and stellar mass dependence of galaxy clustering at 0.5¡ z¡ 1.1*, Astronomy & Astrophysics **557**, A17 (2013).

[41] M. Kerscher, I. Szapudi, and A. S. Szalay, *A comparison of estimators for the two-point correlation function*, The Astrophysical Journal Letters **535**, L13 (2000).

[42] R. A. Skibba, R. K. Sheth, D. J. Croton, S. I. Muldrew, U. Abbas, F. R. Pearce, and G. M. Shattow, *Measures of galaxy environment–II. Rank-ordered mark correlations*, Monthly Notices of the Royal Astronomical Society **429**, 458 (2013).

[43] P. Norberg, C. M. Baugh, E. Gaztanaga, and D. J. Croton, *Statistical analysis of galaxy surveys–I. Robust error estimation for two-point clustering statistics*, Monthly Notices of the Royal Astronomical Society **396**, 19 (2009).

[44] R. Scranton et al., *Analysis of systematic effects and statistical uncertainties in angular clustering of galaxies from early Sloan Digital Sky Survey data*, The Astrophysical Journal **579**, 48 (2002).

[45] E. Gaztanaga, J. Wagg, T. Multamäki, A. Montana, and D. Hughes, *Two-point anisotropies in WMAP and the cosmic quadrupole*, Monthly Notices of the Royal Astronomical Society **346**, 47 (2003).

[46] M. C. Artale, S. E. Pedrosa, J. W. Trayford, T. Theuns, D. J. Farrow, P. Norberg, I. Zehavi, R. G. Bower, and M. Schaller, *Small-scale galaxy clustering in the EAGLE simulation*, Monthly Notices of the Royal Astronomical Society **470**, 1771 (2017).

[47] A. E. Evrard et al., *Galaxy clusters in hubble volume simulations: cosmological constraints from sky survey populations*, The Astrophysical Journal **573**, 7 (2002).

[48] Y. Suto, *Clustering of Dark Matter Halos on the Light-cone*, arXiv preprint astro-ph/0110073 (2001).

[49] U. Abbas and R. K. Sheth, *The environmental dependence of clustering in hierarchical models*, Monthly Notices of the Royal Astronomical Society **364**, 1327 (2005).

[50] D. J. Farrow et al., *Galaxy and mass assembly (GAMA): projected galaxy clustering*, Monthly Notices of the Royal Astronomical Society **454**, 2120 (2015).

[51] A. Durkalec et al., *The VIMOS Ultra Deep Survey-Luminosity and stellar mass dependence of galaxy clustering at z˜ 3*, Astronomy & Astrophysics **612**, A42 (2018).

[52] I. Zehavi et al., *Galaxy clustering in early Sloan Digital Sky Survey redshift data*, The Astrophysical Journal **571**, 172 (2002).

[53] V. Desjacques, D. Jeong, and F. Schmidt, *Large-scale galaxy bias*, Physics reports **733**, 1 (2018).