Alma Mater Studiorum · Università di Bologna

SCUOLA DI SCIENZE

Corso di Laurea Magistrale in Astrofisica e Cosmologia

A new *free software* package for painting galaxies into dark matter haloes

Presentata da:

Federico Zangrandi

Relatore:

Prof. Federico Marulli

Correlatori:

Prof. Lauro Moscardini

Prof. Carlo Giocoli

Dott. Sofia Contarini

Appello IV

Anno Accademico 2018-2019

"Voglio una vita tranquilla,
senza pensare se dover moltiplicare o dividere per h . . . "

Contents

Abstract

The large-scale structure properties of extra-galactic objects provide a fundamental tool to infer information about our Universe. The scientific analysis of galaxy catalogues started in the 1920's with E. Hubble, who proposed the first classification of galaxies based on their morphological appearance. During the years, several improvements in the study of the physical galaxy properties, such as stellar mass and luminosity, have been achieved. Nowadays, statistical analyses are performed using primarily the stellar mass function (SMF) and the two-point correlation function (2PCF), at different redshifts. Galaxies are biased tracers of the underlying matter density field and their clustering allows to infer the spatial distribution of the large scale structure of the Universe, whose properties depend on the cosmological model. On the other hand, the SMF can be used to infer information about the evolution of galaxies and about their formation channels. Despite all the observational studies that have been conducted, a full and exhaustive theory for the galaxy formation and evolution is far from be reached, due to the multitude of different and competitive processes involved. A common practice to face this issue is to compare observed galaxy properties with the ones obtained from simulated galaxy samples. Many efforts have been made to construct increasingly large and detailed galaxy catalogues, but the requirement of high computational power or, alternatively, the employment of empirical approximations, makes the realisation of accurate simulations a difficult task to achieve.

In this Thesis work we present a new *free software* package based on the Halo Occupation Distribution (HOD) technique to construct galaxy catalogues populating whatever Dark Matter (DM) halo catalogues. This code has been implemented inside the CosmoBolognaLib environment, which is a set of *free software* C++/Python numerical libraries for cosmological calculations (Marulli et al. [2016]). Our code offers the possibility to use dif-

ferent methods and various parameterisations to populate the DM haloes. In particular, this code is based on the prescriptions of three different HOD models, calibrated by Zehavi et al. [2005], Zehavi et al. [2011] and Moster et al. [2010]. We test these three methods employing a DM halo catalogue extracted from the Euclid Flagship simulations (Castander et al. [2020, in preparation]). We compare the SMF of our catalogue with the observed one obtained by Panter et al. [2007], and with the SMF measured from Flagship galaxy mock catalogues. We find a good agreement between our predicted SMF and the observed one. We test then the clustering properties of our mock galaxy catalogue, measuring the 2PCF and comparing it to the 2PCF obtained from Flagship HOD galaxy catalogue. As expected, we find an increasing in the clustering as a function of the stellar mass selections applied, in agreement with theoretical expectations and observational measurements.

Sommario

Le proprietà degli oggetti extra-galattici relative alla struttura a grande scala rappresentano uno strumento fondamentale per dedurre informazioni sull'Universo. Lo studio scientifico delle galassie iniziò negli anni '20, con una prima classificazione morfologica delle galassie fornita da di E. Hubble. Nel corso del anni sono stati compiuti molti passi avanti nello studio delle proprietà fisiche delle galassie, come la misura della loro massa e luminosità. Al giorno d'oggi lo studio statistico delle galassie è svolto principalmente attraverso la funzione di massa stellare (SMF) e la funzione di correlazione a due punti (2PCF), per diversi redshift. Le galassie sono un tracciante biassato della distribuzione di massa nell'Universo e il loro clustering permette di studiare la struttura a grande scala dell'Universo, la quale dipende dal modello cosmologico. Dall'altro lato abbiamo la funzione di massa stellare, che può essere usata per dedurre informazioni sull'evoluzione e la formazione delle galassie. Nonostante tutti gli studi osservativi condotti, siamo ancora lontani dall'avere a disposizione una teoria completa ed esaustiva per la formazione delle galassie a causa della complessità dei fenomeni coinvolti, spesso tra loro competitivi. Per affrontare questa difficoltà è pratica comune confrontare le proprietà osservate delle galassie con le proprietà ottenute da galassie simulate. Nel corso del tempo sono stati ottenuti cataloghi simulati di galassie sempre più dettagliati e grandi, ma l'elevato costo in termini di tempo di calcolo e la necessità di usare approssimazioni rendono difficile la simulazione di cataloghi accurati e dettagliati.

In questo lavoro di Tesi presentiamo un nuovo codice in grado di popolare un qualsiasi catalogo di aloni di materia oscura attraverso l'uso delle funzioni di probabilità definite dal metodo detto Halo Occupation Distribution (HOD). Questo codice è stato implementato all'interno dell'ambiente delle CosmoBolognaLib (CBL), che costituiscono un insieme di librerie pubbliche progettate per effettuare calcoli cosmologici, e scritte nei linguaggi

C++/Python (Marulli et al. [2016]). Il nostro codice offre la possibilità di popolare gli aloni di materia oscura usando diversi modelli e parametrizzazioni, calibrate da Zehavi et al. [2005], Zehavi et al. [2011] e Moster et al. [2010]. Abbiamo testato i diversi modelli popolando un catalogo di aloni di materia oscura estratto dalle simulazioni Euclid Flagship (Castander et al. [2020, in preparation]). Abbiamo confrontato la funzione di massa stellare misurata dal nostro catalogo di galassie simulate con quella misurata da Panter et al. [2007] e quella ottenuta dal catalogo di galassie simulate nelle Flagship. Troviamo un buon accordo tra la funzione di massa stellare osservata e quella predetta da noi. Inoltre abbiamo testato le proprietà di clustering del nostro catalogo di galassie misurandone la funzione di correlazione a due punti, per poi confrontarla con quella ottenuta dalle galassie presenti nelle Flagship. Abbiamo ritrovato l'andamento crescente atteso per il clustering all'aumentare massa stellare, in accordo con le predizioni teoriche e le osservazioni.

Introduction

At the present day, the commonly accepted cosmological model is the Λ-cold dark matter (ΛCDM). This model is based on the theory of the General Relativity (GR) (Einstein [1916]) that is, currently, the best description of the gravitational force, which is the dominant force on large scales. In the ΛCDM model the observed accelerated expansion of the Universe is described by introducing the so-called Cosmological Constant, Λ, in the Einstein's field equations. The first evidence of the accelerated expansion of the Universe was provided by Perlmutter et al. [1998] and Riess et al. [1998] using the SuperNovae Type Ia as standard candles. The standard cosmological model assumes the existence of CDM particles, that are particles which interact only through gravity. Several indirect evidences for DM have been found during the last decades. The first that introduced the hypothesis of the existence of the DM was Fritz Zwicky (Zwicky [1933]). He measured the velocity dispersion of galaxies in the Coma cluster and, applying the virial theorem, he found that the majority of the cluster mass was not visible. Evidences for DM have been also obtained with other independent probes, e.g. exploiting the rotational curves in spiral galaxies, and using the gravitational lensing effect and the X-ray emission from hot gas in cluster of galaxies.

According to the ΛCDM model, the Universe is composed of $\sim 70\%$ of Dark Energy (DE), which is described by in the GR field equations, and $\sim 25\%$ of CDM. The remaining $\sim 5\%$ is composed of the so-called baryonic matter, that constitutes the only component of the Universe directly observable through light. The present day cosmological sources, such as galaxies and clusters of galaxies, probes the observable cosmological framework. The standard scenario for the formation of the cosmic structures is based on the assumption that the initial fluctuations in the primordial matter density field were Gaussian distributed. These fluctuations started growing driven

by gravity. The baryonic matter started its collapse only after it became neutral, thus after the so-called *recombination*. Then it fell into the DM halo potential wells and rapidly reached the same overdensity values of the DM field. When overdensities reached unity, their growth became non-linear. Except for the highly approximated spherical collapse model, there are no analytical models able to describe the non-linear regime. Therefore numerical simulations are required to describe the non-linear evolution of cosmic structures.

When the first DM collapsed structures formed, they evolved hierarchically assembling into even more larger structures, following the path described by so-called merger trees. The baryonic matter, interacting also through the electro-magnetic force, can collapse further than DM, driven by radiative cooling. The several physical processes involved during the collapse of the baryonic matter make its description even more complicated than the one of the DM structures. For this reason, there are no general and complete models yet able to fully describe the formation and evolution of galaxies. To overcome this issue, the numerical approach is needed once again. Generally, constructing simulated galaxy catalogues is not an easy task and several approaches are available. The most accurate one involves the solutions of the magneto-hydrodynamic equations. However, this set of equations is strongly non-linear and the computational time required makes this method computationally demanding. Another approach to simulate the formation and evolution of galaxies consists of parameterising the physical processes involved through approximated relations, searching for the parameter values that make the main properties of simulated galaxies close to the observed ones. The last method that we report is the most empirical one, consisting in the calibration of the probability distribution for DM haloes to contain a certain number of galaxies. This method exploits the so-called Halo Occupation Distribution (HOD) framework, and allows to construct mock galaxy catalogues, calibrating the adopted models through observations. In this Thesis work, we implement a new *free software* code for populating DM haloes with galaxies within the HOD framework.

This work is organized as follows:

- In Chapter 1 we introduce the cosmological framework within this Thesis work is developed. We introduce some basic concepts of GR and illustrate how the evolution of the Universe can be mathematically

described.

- In Chapter 2 we discuss the formation of cosmological structures, starting from density field fluctuations and coming to the formation of collapsed DM haloes. We describe also some of the main DM halo properties, such as the density radial profile and mass function.

- In Chapter 3 we introduce the classification of galaxies and some galaxy formation and evolution scenarios. We describe the stellar mass function and discuss how to use it as an observable tool used to infer information on the cosmological evolution of galaxies. Then we describe the common techniques used to measure the galaxy stellar masses. Finally, we define the main estimators for the computation of the two-point correlation function, showing how the two-point correlation function of galaxies can be exploied as a powerful tool to investigate the large-scale structures of the Universe.

- In Chapter 4 we present the three main methods to construct simulated galaxy samples. We start describing the main cosmological simulation techniques, then the semi-analytic models and, in the end, the HOD and similar methods. Finally, we describe the Flagship simulations used in this work, from which we extracted the DM halo catalogue to populate with galaxies.

- In Chapter 5 we present the new HOD code implemented in this Thesis work, discussing the parameterisations used to populate the halo catalogue with galaxies. Then we describe the galaxy properties of the obtained mock catalogue, comparing our results with some reference models.

- In Chapter 6 we summarise our work and draw our final conclusions. Finally, we analyse the future perspectives of this work, focusing on the possible improvements and applications of our code.

Chapter 1

Cosmological framework

In this Chapter we provide an overview of the Cosmological framework that we assumed in this Thesis. We summarise the mathematical tools used to describe the Universe, thought as a four dimensional differentiable manifold, with the Theory of General Relativity (GR). We introduce the Hubble-Lemaître Law and we define the concept of redshift. After that we will see how the cosmological structures can grow up in this contest.

1.1 General Relativity

The gravitational force is the predominant force that governs the evolution of the Universe as a whole. Two of the four fundamental forces of nature - Strong and Weak nuclear forces - act on very short ranges of distances, so on large scales we can neglect them. Electromagnetic forces have an effect also on large scales, but the matter in Universe is overall neutral. Thus the gravitational force plays the main role in the evolution of the Universe.

The most powerful model to describe this force is the GR, introduced in 1916 by Einstein in order to combine his former theory of Special Relativity and the force of gravity (Einstein [1916]).

In this framework, gravity is not thought anymore as a force, but as a property of the space-time itself. In GR space-time is described as a four-dimensional differentiable manifold. Every point on it is called *event* and has four coordinates, that is three space-like and one time-like. How the coordinates of different points in space-time are connected from one to the other is described by a metric tensor $g_{\mu\nu}$, which characterize the intrinsic properties of a manifold or, in other words, is the metric that determines

all the distance relations between points. Indeed the metric tensor, taking into account the intrinsic curvature of the manifold, describes the geometry of the space-time. We shall require the geometry to be locally Euclidian. That is the geometry is Euclidian in the neighborhood of any point.

We can write the displacement between two close *events* $x^i = (t, x_1, x_2, x_3)$ and $x^i + dx = (t + \mathrm{d}t, x_1 + \mathrm{d}x_1, x_2 + \mathrm{d}x_2, x_3 + \mathrm{d}x_3)$ in this way:

$$\mathrm{d}s^2 = g_{ik}\mathrm{d}x^i\mathrm{d}x^k \quad (\mu, \nu = 1, 2, 3, 4).$$

The shortest paths between any two *events* are called the geodesics, and generalise the concept of straight lines in curved spaces, and can be assessed by minimising ds^2:

$$\delta \int ds = 0 \quad ,$$

where δ represents a small variation in the path with respect to the path that minimises the distance between any two *events*. All objects move along these geodesics. The paths can be obtained by solving the so-called geodesic equation:

$$\frac{\mathrm{d}^2x^i}{\mathrm{d}s^2} + \Gamma^i_{kl}\frac{\mathrm{d}x^k}{ds}\frac{\mathrm{d}x^l}{\mathrm{d}s} = 0 \quad ,$$

where Γ are the Christoffel's symbols which contain the metric tensor

$$\Gamma^i_{kl} = \frac{1}{2}g^{im}\left[\frac{\partial g_{mk}}{\partial x^l} + \frac{\partial g_{ml}}{\partial x^k} - \frac{\partial g_{kl}}{\partial x^m}\right] \quad ,$$

and

$$g^{im}g_{mk} = \delta^i_k \quad ,$$

is the Kronecker delta, which is unity if $i = k$, and zero otherwise. According to GR, the metric tensor itself is influenced by how the matter is distributed and how it moves.

We can describe the main properties of the distribution of matter in space-time introducing the energy-momentum tensor T_{ik} , which describes the density, flux energy and momentum of matter. For a perfect fluid with pressure p and density ρ, the energy-momentum tensor can be written as:

$$T_{ik} = (p + \rho c^2)U_iU_k - pg_{ik} \quad , \tag{1.1.1}$$

where $U_i = g_{ik}U^k = g_{ik}\frac{dx^k}{ds}$ is the four-velocity of the fluid and $x^k(s)$ is the world line of a fluid element.

In differential geometry, the energy conservation law can be found by imposing that the covariant derivative is null:

$$T_i^k; k = 0 \quad , \tag{1.1.2}$$

where, conventionally, the semicolon indicates the covariant derivative, which for a tensor $\mathbf{A}$ is defined as:

$$A^{kl...}_{pq...}; j = \frac{\partial A^{kl...}_{pq}...}{x^j} + \Gamma^k_{mj} A^{ml...}_{pq...} + \Gamma^l_{nj} A^{kn...}_{pq...} + ... - \Gamma^r_{pj} A^{kl...}_{rq...} - \Gamma^s_{qj} A^{kl...}_{ps...} - ... \quad .$$

From Eq.(1.1.2), Einstein derived the Poisson's equation in the classical limit:

$$\nabla^2 \phi = 4\pi G \rho \quad ,$$

which relates the gravitational potential, ϕ, to the density, ρ, of the source of the gravitational field. This means that the metric tensor, g_{ij}, is connected to the energy-momentum tensor, T_{ij}, with an equation which contains only the first two derivatives of g_{ij}, and has zero covariant derivative.

The Riemann–Christoffel tensor, R^i_{klm}, can be used to assess whether a space is curved or flat:

$$R^i_{klm} = \frac{\partial \Gamma^i_{km}}{\partial x^l} + \frac{\partial \Gamma^i_{kl}}{\partial x^m} + \Gamma^i_{nl}\Gamma^n_{km} - \Gamma^i_{nm}\Gamma^n_{kl} \quad . \tag{1.1.3}$$

From Eq. (1.1.3) we can construct the so-called Ricci tensor and Ricci scalar, that describe the curvature:

$$R_{ik} = R^l_{ilk} \quad ,$$

$$R = g^{ik} R_{ik} \quad ,$$

from which we can define the Einstein tensor as:

$$G_{ik} \equiv R_{ik} - \frac{1}{2} g_{ik} R \quad .$$

It can be shown that:

$$G^k_{i;k} = 0 \quad .$$

The fundamental equation of GR proposed by Einstein is the following:

$$G_{ik} \equiv R_{ik} - \frac{1}{2} g_{ik} R = \frac{8\pi G}{c^4} T_{ik} \quad , \tag{1.1.4}$$

where c is the speed of light and G is the Newtonian gravitational constant. The quantity $8\pi G/c^4$ ensures to obtain the Poisson's equation in the weak

gravitational field limit. In Eq.(1.1.4) G_{ik} contains second derivatives of the metric tensor. As it was soon revealed, the solutions of Eq.(1.1.4) describe an expanding Universe, which at that time was considered physical meaningless. Einstein himself proposed an alternative form for his field equation to obtain solutions without expansion, introducing the so-called cosmological constant, Λ:

$$G_{ik} - \Lambda g_{ik} = R_{ik} - \frac{1}{2} g_{ik} R - \Lambda g_{ik} = \frac{8\pi G}{c^4} T_{ik} \quad . \tag{1.1.5}$$

The term Λ can be equivalently be interpreted as a modification of gravity, when it is put on the left-hand side of Eq.(1.1.5), or as an additional energy component, if it is on the right-hand side. Some years after the introduction of Λ, the astronomer E. Hubble observed that galaxies are receding from us with a velocity that is proportional to the distance from Earth. This observational fact is interpreted as the effect of the expansion of the Universe and made the cosmological constant useless. However, recent observational evidences, like the Supernovae type Ia(SNIa) distance diagram, showed that the Universe is actually accelerating (Perlmutter et al. [1998], Riess et al. [2019]). So the cosmological constant has been recently reintroduced, though with an opposite purpose with respect to the original one.

Eq.(1.1.5) tells us how the matter moves due to the *shape* of space-time, and that the space-time is curved due to the matter/energy content. Solving the Einstein's equation for arbitrary distributions of matter is not generally possible with analytic techniques. We have to make some $a-priori$ assumptions that simplify the form of the metric tensor.

1.2 Cosmological Principle

The cosmological principle asserts that the Universe is homogeneous and isotropic on large scales. Homogeneity means that everywhere in space the average properties of the Universe are the same, while isotropy means that the average properties of the Universe are the same in every direction. These assumptions are observationally confirmed today on sufficient large scale, $r \geq 200$ Mpc (Scrimgeour et al. [2012]). The cosmological principle was introduced by Einstein himself, based on the Ernst Mach's philosophical point of view, that can be summarised with the idea that all laws of physics are determined by the global distribution of matter in the Universe. Einstein thought that the only way to construct a cosmological theory was to

assume some global *simplicity* in the Universe, enabling symmetries in the local distribution of matter. The Cosmic Microwaves Background radiation (CMB) provides one of the strongest evidence in support of the Cosmological Principle. Indeed, the CMB temperature shows very small fluctuations in different directions. If we add to this observational evidence the so-called *Coperinican Principle*, which claims that we are not in any special place, we automatically regain also the homogeneity. So the observed isotropy together with the *Copernican Principle* implies the *Cosmological Principle*.

The *Cosmological Principle* can be taken further with the *Perfect Cosmological Principle*, which says that average properties of the Universe are the same in all places, in all directions and in all times. This stronger version of the *Cosmological Principle* in addition to the observational fact that the Universe is expanding, led to develop the so-called *steady-state* model of the Universe, which was eventually abandoned from the 1960's, due to the discover of CMB radiation.

1.3 Friedman-Lemaître-Robertson-Walker Metric

In this section, we describe the metric which satisfies the Cosmological Principle. As a consequence of the homogeneity's hypothesis, the Universe can be divided into many time slices, Σ_t, described as sub-manifolds where the proper time t is the same everywhere. In first approximation, we can treat the matter content of the Universe as a perfect fluid. A set of *comoving coordinates* x^i $(i = 1, 2, 3)$ is associated to each element of this fluid.

The metric that describes a homogeneous and isotropic space is the so-called Friedman-Lemaître-Robertson-Walker (FLRW) metric:

$$\mathrm{d}s^2 = (c\mathrm{d}t)^2 - a^2(t)\left[\frac{\mathrm{d}r^2}{1 - Kr^2} + r^2(\mathrm{d}\theta^2 + \sin^2(\theta))\mathrm{d}\phi^2\right] \quad , \tag{1.3.1}$$

where r,θ,ϕ are spherical comoving coordinates. The scale r is conventionally dimensionless, while $a(t)$, which is called the *scale factor* and is a function of the proper time, has the dimension of a length and accounts for the expansion of Universe. The term K is called the *curvature parameter* and can take only the values: -1, 0, +1. It describes the curvature and thus the geometrical properties of the Universe. At fixed time t, we are on a time slice Σ_t whose topology is determined by the value of K:

- K=0 or Flat Universe:

$$\mathrm{d}l^2 = \mathrm{d}r^2 + \left(\mathrm{d}\theta^2 + \sin^2(\theta)\mathrm{d}\phi^2\right) = \mathrm{d}r^2 + r^2\mathrm{d}\Omega^2 \quad ,$$

  this means that Σ_t is flat and has the geometrical properties of an Euclidean space;

- K= 1 or Closed Universe:

$$\mathrm{d}l^2 = \mathrm{d}\chi^2 + \sin^2(\chi)\mathrm{d}\Omega^2 = \frac{\mathrm{d}r^2}{1-r^2} + r^2\mathrm{d}\Omega^2 \quad ,$$

  where $0 \leq \chi \leq \pi$. In this case Σ_t is a three-dimensional sphere;

- $K = -1$ or Open Universe:

$$\mathrm{d}l^2 = \mathrm{d}\Psi^2 + \sinh^2(\Psi)\mathrm{d}\Omega^2 = \frac{\mathrm{d}r^2}{1+r^2} + r^2\mathrm{d}\Omega^2 \quad ,$$

  where $0 \leq \Psi < \infty$. In this case Σ_t is a three-dimensional hyperboloid.

1.4 Cosmological Distance and Hubble-Lemaître's Law

The *proper distance*, d_p, is the distance between a point P_0, that we can locate at in the origin of the coordinate system (r, θ, ϕ), and another point P. Ideally, this distance is measured by a chain of rules which connect P and P_0 at time t, so we are in a slice of time Σ_t, that is with $\mathrm{d}t = 0$. From the Friedman-Leimaître-Robertson-Walker metric, we can write:

$$d_p = \int_0^r \frac{a\ \mathrm{d}r'}{(1-Kr'^2)^{1/2}} = af(r) \quad ,$$

where the form of $f(r)$ depends on the specific value of K:

$$f(r) = \begin{cases} \sin^{-1}(r) & K = 1 \\ r & K = 0 \\ \sinh^{-1}(r) & K = -1 \end{cases} \quad .$$

The definition of the proper distance is operationally useless as it is not possible to measure simultaneously all distance elements separating any two points. The proper distance at time t is related to that at the present time t_0 by:

$$d_p(t_0) = a_0 f(r) = \frac{a_0}{a} d_p(t) \quad .$$

One can define another comoving coordinate, instead of r, as:

$$d_c = a_0 f(r) \quad ,$$

which is related to the proper distance as follows:

$$d_c = \frac{a_0}{a} d_p(t) \quad .$$

Since d_p depends on time, the proper distance between two points separated by a fixed value of d_p changes due to the expansion of the Universe. Instead, d_c is time-independent because it is rescaled by a_0 from which we can compute the velocity of the expansion by deriving the proper distance:

$$v_r = \frac{\mathrm{d}d_p}{\mathrm{d}t} = \dot{a} f(r) = \frac{\dot{a}}{a} d_p \quad . \tag{1.4.1}$$

This is known as *Hubble-Lemeîetre's Law* and it is usually written in terms of the *Hubble's parameter*, H, defined as follows:

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

This parameter at the present time, $H(t_0) \equiv H_0$, has a value around $H_0 \approx 70$ km s^{-1} Mpc^{-1}. Usually, it is convenient to define the dimensionless parameter $h \equiv H_0/100$ km s^{-1} Mpc^{-1}. Recently, it comes out that there is a tension between the values of H_0 measured from the CMB angular spectrum, $H_0 = 67.4 \pm 0.5$ kms^{-1} Mpc^{-1} (Planck Collaboration et al. [2018]), and those measured using distance ladders as Cepheids or SNIa where $H_0 = 74.03 \pm 1.42$ km s^{-1} Mpc^{-1}(Riess et al. [2019]). Indeed, local measurements of H_0 have a discrepancy of 4.2σ with respect to the value of H_0 coming from CMB. It seems unlikely that the tension is totally caused by systematic errors in the measurements, as different teams using independent methods obtained consistent results. One possible way to solve the H_0 tension is to introduce a scalar field, the so-called *Early Dark Energy (EDE)*, which acts just before the appearance of the CMB (*e.g.* Poulin et al. [2019]). But the problem is still open.

1.4.1 Redshift

A fundamental cosmological observable is the redshift in the electromagnetic spectrum of extra-galactic sources which is directly related to the expansion of the Universe. The redshift is defined as follows:

$$z \equiv \frac{\lambda_0 - \lambda_e}{\lambda_e} \quad ,$$

where λ_o is the wavelength of radiation from a source observed at the time t_0 and emitted at the time t_e. This shift is due to the expansion of the Universe. Because the light follows null geodesics ($\mathrm{d}s^2 = 0$) we have:

$$\int_{t_e}^{t_0} \frac{c\mathrm{d}t}{a(t)} = \int_0^r \frac{\mathrm{d}r}{(1 - Kr^2)^{1/2}} = f(r) \quad . \tag{1.4.2}$$

The light emitted by a source at time $t'_e = t_e + \delta t_e$ reaches the observer at time the $t'_0 = t_0 + \delta t_0$. Since r is a comoving coordinate, and both the source and observer are moving with the cosmological expansion, $f(r)$ does not change. Thus we have:

$$\int_{t'}^{t'_0} \frac{c\mathrm{d}t}{a(t)} = f(r) \quad . \tag{1.4.3}$$

If δt and δt_0 are small enough, Eqs. (1.4.2) and (1.4.3) imply

$$\frac{\delta t_0}{a_0} = \frac{\delta t}{a} \quad .$$

The time intervals δt and δt_0 are related to the frequencies of the radiation emitted and observed, so we have $\nu_e a = \nu_0 a_0$ or, equivalently:

$$\frac{a}{\lambda_e} = \frac{a_0}{\lambda_0} \quad .$$

Thus, according to the redshift definition, we obtain:

$$1 + z = \frac{a_0}{a} \quad . \tag{1.4.4}$$

The more distant is a source from us, the higher is its z. Therefore redshift measurements are used to infer comoving distances of extra-galactic sources.

1.4.2 Other distance definitions

The proper distance definition can not be used operationally to measure distances because it depends on a that changes with time. We need other

ways to measure the distances of astronomical objects. One option is the *luminosity distance*, d_L, that is defined to preserve the relation between the flux received by the observer and the luminosity emitted by a source. If L is the power emitted by a source and l is the flux observed, then:

$$d_L = \left(\frac{L}{4\pi l}\right)^{1/2} \quad .$$

The spherical surface centered on the source and passing from the observer at time t_0 is $4\pi a_0^2 r^2$. The photons that reach the orbserver are redshifted by a factor a/a_0 due to the expansion of the Universe. Also, photons emitted by a source in the small time interval δt, arrive in an interval $\delta t_0 = (a_0/a)\delta t$ due to the time dilatation. So the flux reaching the observer is:

$$l = \frac{L}{4\pi a_0^2 r^2}\left(\frac{a}{a_0}\right)^{1/2} \quad .$$

Using Eq. (1.4.4), the relation between the luminosity distance and the redshift is the following:

$$d_L = r a_0 (1+z) \quad .$$

This definition has been used to estimate the distance of standard candles, like $SNIa$. Another definition of distance commonly used in cosmology is the angular distance, d_A, which is defined to keep the angular size of an object constant in distance from an observer. Let D_p be the proper diameter of a source that subtends an angle $\Delta\theta$. From Eq. (1.3.1) we have:

$$D_p = a r \Delta\theta \quad .$$

So the angular distance becomes:

$$d_A = \frac{D_p}{\Delta\theta} = a r \quad . \tag{1.4.5}$$

The angular distance and the luminosity distance can be measured independently from each other. As shown by Etherington [1933] in the FLRW metric the relation between the two is:

$$d_A \equiv \frac{d_L}{(1+z)^2} \quad .$$

This observable relation provides a powerful probe to test the validity of the FLRW metric (*e.g.* Li et al. [2011]).

1.5 Friedman Equations

Assuming the *Cosmological Principle*, we can solve the Einstein's field equation (1.1.4), with the FLRW metric (1.3.1) and the energy-momentum tensor of a perfect fluid (1.1). Under these assumptions, the field equation has only two independent solutions, called the *Friedman equations*:

$$\ddot{a} = -\frac{4\pi G}{3}\Big(\rho + 3\frac{p}{c^2}\Big)a \quad , \tag{1.5.1}$$

$$\dot{a}^2 + Kc^2 = \frac{8\pi G}{3}\rho a^2 \quad . \tag{1.5.2}$$

Actually Eqs. (1.5.1) and (1.5.2) are not independent. One can be obtained from the other assuming that the Universe is a closed system and so its expansion is adiabatic, without loss of energy:

$$\mathrm{d}\big(\rho c^2 a^3\big) = -p\mathrm{d}a^3 \quad . \tag{1.5.3}$$

Friedman was the first that found these solutions in 1922. At that time, before the Hubble's observations of receding galaxies, cosmologists thought that the Universe was static, that is $\ddot{a} = \dot{a} = 0$, and from Eqs. (1.5.1),(1.5.2):

$$\rho = -\frac{3p}{c} \quad ,$$

that has no classical physical meanings. As already mentioned in the Section (1.1), for this reason Einstein decided to introduce the cosmological constant, Λ, in the modified field equation given by Eq. (1.1.5).

1.6 The Einstein Universe

The field equation with the cosmological constant can be written in terms of the effective energy-momentum tensor, $\tilde{T}_{ik}$:

$$R_{ik} - \frac{1}{2}g_{ik}R = \frac{8\pi G}{c^4}\tilde{T}_{ik} \quad , \tag{1.6.1}$$

with:

$$\tilde{T}_{ik} \equiv T_{ik} + \Lambda g_{ik}\frac{c^4}{8\pi G} \quad .$$

Defining the effective pressure, $\tilde{p} \equiv p - \frac{\Lambda c^4}{8\pi G}$, and effective density, $\tilde{\rho} \equiv \rho + \frac{\Lambda c^2}{8\pi G}$, the tensor $\tilde{T}_{ik}$ can be written as follows:

$$\tilde{T}_{ik} = -\tilde{p}g_{ik} + \big(\tilde{p} + \tilde{\rho}c^2\big)u_i u_k \quad . \tag{1.6.2}$$

The mathematical form of the Eq.(1.6.1) is the same as the one of Eq.(1.1.4), thus also the solutions have to be the same, with pressure $\tilde{p}$ and density $\tilde{\rho}$:

$$\begin{cases} \ddot{a} = -\frac{4\pi G}{3}\left(\tilde{\rho} + \frac{3\tilde{p}}{c^2}\right)a \\ \dot{a}^2 + Kc^2 = \frac{8\pi G}{3}\tilde{\rho}a^2 \end{cases} . \tag{1.6.3}$$

In static universes, $\ddot{a} = \dot{a} = 0$. In the *dusty* assumption, the pressure is null and Eqs.(1.6.3) give:

$$\begin{cases} \Lambda = \frac{K}{a^2} \\ \rho = \frac{Kc^2}{4\pi G a^2} \end{cases} .$$

Assuming that $p = 0$, the only way to have a positive density is to impose $K = 1$ that is a closed Universe. This model is known as the Einstein Universe that is a static Universe of matter (*dusty* Universe), with positive curvature. After the Hubble's discovery, discussed in Section 1.4, static models were abandoned. In 1990's, observations suggested that the Universe (Perlmutter et al. [1998]) is an accelerated expansion. The cosmological constant was thus reintroduced to describe the acceleration.

1.7 The de Sitter Universe

In order to understand the role of the cosmological constant, we can consider a flat ($K = 0$) empty Universe model ($p = 0$ and $\rho = 0$), where the only contribute to the energy momentum comes from Λ. Imposing $p = 0$ and $\rho = 0$, Eq. (1.6.2) gives:

$$\tilde{p} = -\tilde{\rho}c^2 \quad . \tag{1.7.1}$$

Using Eq.(1.7.1), the solution of Friedman equations (1.6.3) brings to:

$$a(t) = A \exp\left[\sqrt{\frac{\Lambda}{3}}ct\right] \quad . \tag{1.7.2}$$

This means that the cosmological constant could account for accelerated expansion ($\ddot{a} > 0$).

1.8 Friedman Models

We can approximate the content of the Universe as a perfect fluid, then it is necessary to desccribe its equation of state, which connects p and ρ. In

general, we can write:

$$p = w\rho c^2 \quad . \tag{1.8.1}$$

The value of w can be assessed by considering the speed of sound:

$$c_{sound} = \left(\frac{\partial p}{\partial \rho}|_{S=const}\right)^{1/2} \quad ,$$

where S is the entropy. The speed of sound has to be major than zero but larger than c. This consideration brings to $0 \leq w < 1$, which is known as the *Zeldovich interval*. The values of w depend on the type of fluid we are considering.

- If we consider a perfect gas, the pressure and density are related by:

  $$p = nk_B T \quad ,$$

  but being the pressure almost negligible in this case $w = 0$.

- If we consider a fluid composed by relativistic particles, for examples photons, we have:

  $$p = \frac{1}{3}\rho c^2 \quad ,$$

  so in this case $w = 1/3$.

- If we are in vacuum where only the cosmological constant is important we have Eq.(1.7.1), which implies $w = -1$.

It is useful to rewrite the adiabatic equation (1.5.3) using the equation of state (1.8.1):

$$\rho_w = \rho_{0,w}\left(\frac{a}{a_0}\right)^{-3(1+w)} \quad . \tag{1.8.2}$$

This means that:

- for a dusty Universe, that is $w = 0$, the density is:

  $$\rho_{matter} = \rho_{0,matter}\left(\frac{a}{a_0}\right)^{-3} = \rho_{0,matter}(1+z)^3 \quad ,$$

- for a radiation Universe, that is $w = 1/3$, the density is:

  $$\rho_{radiation} = \rho_{0,radiation}\left(\frac{a}{a_0}\right)^{-4} = \rho_{0,radiation}(1+z)^4 \quad ,$$

- for a vacuum Universe, with the cosmological constant, that is $w = -1$, the density is:

$$\rho_\Lambda = \rho_{0,\Lambda}(a/a_0)^0 = \rho_{0,\Lambda} \quad .$$

We conclude that the matter and radiation density components changes during expansion with different rates. So we can divide the history of the Universe into epochs based on which component was the dominant at that epoch, as shown in Figure 1.1.

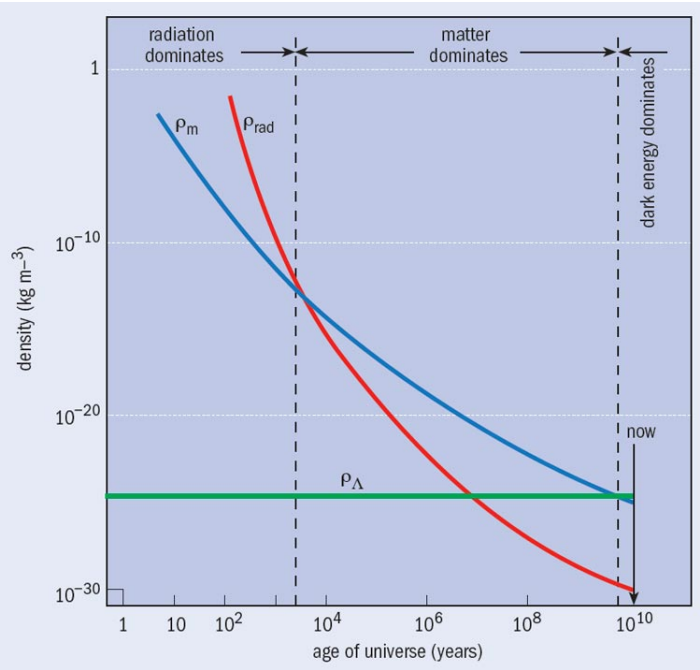

Figure 1.1: Density trends with time. We see that at the beginning the radiation has the most important role in the evolution of the Universe. While at the present time the cosmological constant dominates. Image from Freedman and Kaufmann [2007]

It is useful to introduce three new quantities

- the *critical density*:

$$\rho_{crit}(t) \equiv \frac{3H(t)^3}{8\pi G} \quad , \tag{1.8.3}$$

- the *density parameter*:

$$\Omega \equiv \frac{\rho}{\rho_{crit}} \quad , \tag{1.8.4}$$

- the *deceleration parameter*:

$$q \equiv -\frac{\ddot{a}a}{\dot{a}^2} \quad . \tag{1.8.5}$$

In order to understand the meaning of $\rho_{crit}(t)$, we rewrite the Friedman Eqs.(1.5.1) (1.5.2) as follows:

$$-\frac{K}{a(t)^2} = \frac{H(t)^2}{c^2}\left(1 - \frac{\rho(t)}{\rho(t)_{crit}}\right) \quad .$$

At the present time we have:

- K=0 $\rightarrow$ Flat Universe $\rightarrow \rho_0 = \rho_{0,crit} \rightarrow \Omega_0 = 1$.
- K=1 $\rightarrow$ Closed Universe $\rightarrow \rho_0 > \rho_{0,crit} \rightarrow \Omega_0 > 1$.
- K=-1 $\rightarrow$ Open Universe $\rightarrow \rho_0 < \rho_{0,crit} \rightarrow \Omega_0 < 1$.

With these definitions and the adiabatic condition given by Eq.(1.5.3), we can write the second Friedman equation:

$$H^2(t) = H_0^2\left(\frac{a_0}{a}\right)^2\left[1 - \sum_i \Omega_{0,w_i} + \sum_i \Omega_{0,w_i}\left(\frac{a_0}{a}\right)^{(1+3w_i)}\right] \quad , \tag{1.8.6}$$

where Ω_{0,w_i} is the density parameter of a specific component (radiation, matter, or Λ) at the present time. In Eq.(1.8.6), but for each epoch we can consider only the dominant component and neglect the others. This is true if we are sufficiently far from the equivalent points, that are the times at which two components have similar energy densities. At the present time in particular we have $\Omega_\Lambda \approx \Omega_M$.

Assuming that $p > 0$ and $\rho > 0$, from the Friedman Eqs.(1.5.1) and (1.5.2), neglecting the possibility of cosmological constant we have $\ddot{a} < 0$. From this fact and the observed expansion of the Universe ($\dot{a}_0 > 0$), we conclude that $a(t)$ is a monotonic and increasing function so at some point in the past there was $a(t) = 0$. This means that the Universe started its expansion from a point where density, pressure and temperature diverged. This is known as *Big Bang*.

1.9 The Enstein-de Sitter Model

In the Einstein-de Sitter (EdS) model the Universe is assumed to be flat and with only one dominant component, thus $\Omega_w = 1$. From Eq.(1.8.6) we can

derive the following behaviors:

$$a(t) = a_0 \left(\frac{t}{t_0}\right)^{2/3(1+w)} \quad , \tag{1.9.1}$$

$$t(a) = t_0 \left(\frac{a}{a_0}\right)^{3(1+w)/2} = t_0 (1+z)^{-3(1+w)/2} \quad , \tag{1.9.2}$$

$$H(t) = \frac{2}{3(1+w)t} = H_0 (1+z)^{3(1+w)/2} \quad , \tag{1.9.3}$$

$$\rho(t) = \frac{1}{6\pi G} \frac{1}{(1+w)^2 t^2} \quad . \tag{1.9.4}$$

In the EdS Universe we can straightforwardly evaluate the age of the Universe, that is inversely proportional to the Hubble's parameter at the present day:

$$t_0 = \frac{2}{3(1+w)} \frac{1}{H_0} \quad ,$$

that results to be about 10 billion years, with the value of H_0 commonly accepted (Planck Collaboration et al. [2018]).

1.10 Curved Universes

Now we consider mono-component curved Universes (indeed $\Omega_0 \neq 1$). The Friedman equation (1.8.6), to completeness:

$$\left(\frac{\dot{a}}{a_0}\right)^2 = H_0^2 \left[1 - \Omega_0 + \Omega_0 \left(\frac{a_0}{a}\right)^{(1+3w)}\right] \quad . \tag{1.10.1}$$

- The first term in the bracket of Eq.(1.10.1) $(1 - \Omega_0 \equiv \Omega_k)$ is in to the Universe curvature.
- The second term in the bracket of Eq.(1.10.1) $(\Omega_0 \left(\frac{a_0}{a}\right)^{(1+3w)})$ depends on the content of the Universe and decreases when a increases. For this reason the curvature is important only when $a \sim a_0$ and every kind of Universe near the Big Bang can be approximated by a flat Universe.

The two terms became equal at z^*, given by the following equation:

$$\frac{a_0}{a^*} = (1+z^*) = \left(\frac{1-\Omega_0}{\Omega_0}\right)^{1/1+3w} \quad .$$

This means that for $z > z^*$, that is closer to the Big Bang, the curvature does not matter and the Universe behaves as a flat Universe, regardless of the actual Ω value. Indeed the solutions near the Big Bang are the same of the EdS model but multiplied by a factor $\Omega_0^{1/2}$.

Open Universes

The Universes with $\Omega < 1$ are called *Open Universes*. The $a(t)$ grows with time and the first term in Eq.(1.10.1) is always positive. For $a(t) > a^*$ the second term in Eq.(1.10.1) is negligible, so we can approximate the equations as follows:

$$\dot{a} \sim a_0 H_0 (1 - \Omega_0)^{1/2} \quad ,$$

$$a \propto t \quad ,$$

$$H \propto 1/t \quad ,$$

$$q \to 0 \ .$$

In *Open Universes*, $a(t)$ always grows and the space expands towards infinity with constant expansion rate. This brings to the thermal death of the Universe.

Closed Universes

The Universes with $\Omega > 1$ are called *Closed Universes*. From Eq.(1.10.1) we find that there is a time, t_m, where $\dot{a}(t_m) = 0$. At this time the value of a is maximum while the value of the density ρ is minimum:

$$a_{max} = a_0 \left(\frac{\Omega_0}{1 - \Omega_0} \right)^{1/1+3w} \quad ,$$

$$\rho_{min} = \rho_0 \left(\frac{\Omega_0}{1 - \Omega_0} \right)^{1+3w/3(1+w)} \quad .$$

After the time t_m, Eq.(1.10.1) changes sign and the Universe stops to grow and starts to contract. At the time $2t_m$ the Universe collapses completely. These behaviors are shown in Figure 1.2, where $a(t)$ is plotted against t, showing the evolution in different Universe models.

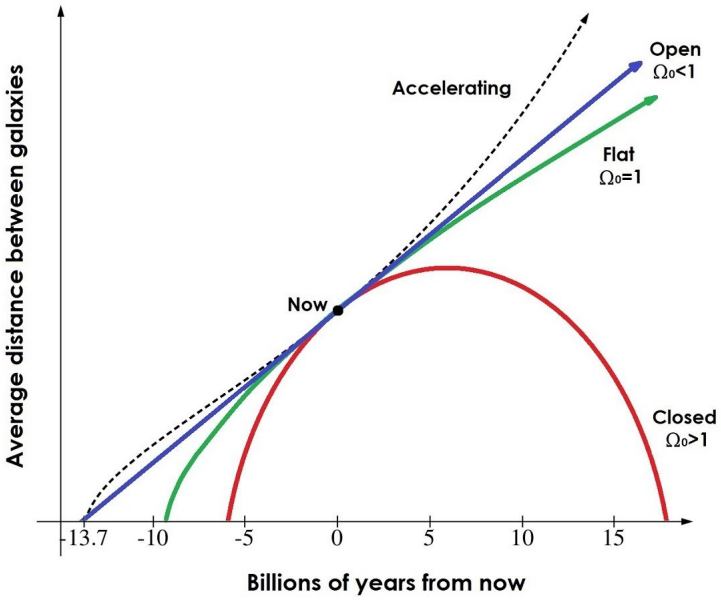

Figure 1.2: Solid lines show the evolution of the scale factor $a(t)$, or equivalently the average distance between galaxies, for different values of Ω_0. The dashed lines show a model with $\Omega_\Lambda \neq 0$. Credits to www.thestargarden.co.uk/Big-bang.html.

The Standard Cosmological Model

The standard cosmological model, called ΛCDM, is set up on the Big Bang theory. The Big Bang theory is consistent with the Hubble's law, with the primordial abundance of light elements, like Helium (Fuller et al. [1988]), and with the CMB radiation properties. Two observational issues, that are the *horizon problem* and the *flatness problem*, can be solved introducing a period of time, in the early Universe, of extremely rapid expansion, called $inflation$ (Guth [1981]).

The Universe in the ΛCDM model is flat, with $\Omega = \Omega_M + \Omega_\Lambda = 1$. The dominant component at the present time is Ω_Λ, that is about 70% of the energy component of the Universe. The remaining 30% of the energy density is composed by Ω_M with only 5% is ordinary matter, also called baryonic matter, while the remaining 25% is composed of Dark Matter (DM). In fact, several observations suggest that there must be large amount of mass in the Universe that we can not directly observe through light but only

via gravitational effects. The first astronomer that supposed the existence of this kind of matter was Zwicky in 1933 (Zwicky [1933]) measuring the velocity dispersion of galaxies in clusters. Other dynaminacal proofs for the existence of DM come from the *Rotational Curves* in disk galaxies (Rubin et al. [1978]), or the velocity dispertion of stars in elliptical galaxies.Besides dynamic measurements, other observables suggest that there is more matter than the visible one, such as for examples, the X-ray emission from Intra Cluster Medium (ICM) (Vikhlinin et al. [2006]), or the gravitational lensing (Massey et al. [2010]). The DM can be classified into two types:

- Hot Dark Matter (HDM) made of low mass relativistic particles. The best candidates are massive neutrinos.
- Cold Dark Matte (CDM) made of massive non-relativistic particles. The best candidates are Weakly Interacting Massive Particles (WIMPS).

Several particle candidates have been proposed and tested. As we will see in the next Chapter 2, the structure formation and evolution models imply that the dominant DM component must be cold.

Chapter 2

Structure Formation

The standard scenario of structure formation is based on the idea that the present-time observed structures, such as galaxies and clusters of galaxies, lie in DM haloes formed via gravitational instabilities which grew in an expanding Universe. As shown by both approximate analytic models and numerical simulations, this evolution brings to the so-called *Large scale Structure* of the Universe, also known as the *Cosmic Web* (Davis et al. [1985]). Soon after the Big Bang, the Universe starts cooling due to adiabatic expansion. The matter becomes the dominant component of the Universe from the Equivalence time ($z_{eq} \sim 10000$). The DM starts its collapse soon after the *Inflationary Epoch* because the DM particles interact with each other only through gravity, and there are no other forces that contrast their collapse. On the other hands, the electromagnetic interactions prevent the baryonic matter to collapse at that time. Baryonic matter can collapse only after the *Recombination time* ($z_{rec} \sim 1100$) because at that time the baryonic matter becomes neutral and starts falling into the gravitational wells of DM. We expect that at that time density fluctuations are of the same order of magnitude as the fluctuations of temperature in CMB, therefore:

$$\delta \equiv \frac{\delta\rho}{\rho} \sim \frac{\delta T}{T} \sim 10^{-5} \quad .$$

The aim of this Chapter is to describe how and under which conditions these perturbations grew in an expanding Universe.

2.1 Jeans Theory

The first discussion on gravitational collapse was made by James Jeans (Jeans [1902]). The basic idea is that a small overdensity,δ, a perturbation in the density field, grows attracting material from the surrounding regions and forming a bound structure. We can define a characteristic length that tells us if an overdensity can grow or not. This is called the *Jeans scale*, R_J, and its dependencies on the other physical quantities can be found in three different ways:

- Exploiting the energetic equilibrium. The kinetic energy of the gas, $E_k = \frac{1}{2}Mv^2$, contrasts the collapse, while the gravitational potential energy, $E_g = -GM\rho R^2$, tends to accumulate mass onto the overdensity. Finding the equilibrium between the two, $E_k = -E_g$, we have:

$$R_J = v\sqrt{\frac{1}{2G\rho}} \quad .$$

- Imposing the balancing of the forces involved during the collapse, thus the gravitational force, $F_g = GM/R^2$, and the pressure force, $F_p = pR^2/M = c_s^2/R$, which contrasts the collapse. Equating these two terms we obtain:

$$R_J = c_s\sqrt{\frac{1}{G\rho}} \quad ,$$

  where c_s is the speed of sound, $c_s \approx p/\rho$.

- Comparing the time scale of the free-fall, $\tau_{FF} \propto 1/\sqrt{G\rho}$, with the typical crossing-time, $\tau_{cross} \propto 2R/v$, we achieve the relation:

$$R_J \propto \frac{v}{2\sqrt{G\rho}} \quad .$$

These approximated methods to evaluate R_J lead to different results, but it is important to notice is that in all cases the Jeans's scale results proportional to the typical particle velocity of the system and inversely proportional to the square root of the density. Now we can simply associate to the *Jeans's scale* the *Jeans's mass*:

$$M_J \propto \rho R_J^3 \quad .$$

The objects with a mass larger than the *Jeans's mass* can grow through the gravitational collapse.

To investigate the Jeans instability more accurately we need to study the dynamics of a self-gravitating fluid in a static background. The fluid has to satisfy the following system of equations:

$$\begin{cases} \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) = 0 & \text{(Mass conservation)} \\ \frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \nabla)\vec{v} = -\frac{1}{\rho} - \nabla\phi & \text{(Momentum conservation)} \\ \nabla^2 \phi = 4\pi G \rho & \text{(Poisson equation)} \\ p = p(\rho, S) = p(\rho) & \text{(Equation of state)} \\ \frac{\mathrm{d}S}{\mathrm{d}t} = f = 0 & \text{(Entropy equation)} \end{cases} \tag{2.1.1}$$

Where ∇ is the *nabla* operator, ρ, p are the density and pressure of the fluid respectively, ϕ is the gravitational potential and S is the entropy of the fluid. Since we consider only adiabatic perturbations, the Entropy equation is null and the pressure is a function of density only. To find the solution of Eqs.(2.1.1) we use the *Perturbative Theory*, that is we suppose to know the *unperturbed* solutions and then we add a small perturbation to study how the solutions change. This is possible only in linear regime and for small perturbations ($\delta x \ll 1$). The unperturbed solutions are the following:

$$\begin{cases} \rho = const = \rho_b \\ \phi = const = \phi_b \\ \vec{v} = \vec{0} \\ p = const = p_b \end{cases} \tag{2.1.2}$$

where the subscript b represents background quantities. We note that these solutions are not strictly valid, indeed $\phi = const$ would imply that $\nabla^2\phi = 0$, and therefore $\rho = 0 \neq \rho_b$. Nevertheless we can neglect this issue for the moment, since it will be properly solved when the expansion of the Universe will be introduced in the model. To obtain the perturbed hydrodynamic equations we add small perturbations to the static solutions, and we search for the solutions again, omitting the second-order terms:

$$\begin{cases} \frac{\partial \delta\rho}{\partial t} + \rho_b \nabla \cdot (\rho \vec{v}) = 0 \\ \frac{\partial \delta\vec{v}}{\partial t} = \frac{c_s^2}{\rho_b} - \nabla\delta\phi \\ \nabla^2 \delta\phi = 4\pi G \delta\rho \end{cases} \tag{2.1.3}$$

Specifically we look for the solutions of Eqs.(2.1.3) as plane waves. We are interested in the density field, so we write the solutions as a function of the

overdensity $\delta \equiv \delta\rho/\rho$:

$$\delta(\vec{r}, t) = \delta_k \exp\left[i\vec{k} \cdot \vec{r} + i\omega t\right] \quad , \tag{2.1.4}$$

where δ_k is the amplitude, $\vec{k} = 2\pi\hat{k}/\lambda$ is the wave number, and ω is the angular frequency of the wave. If we use the plane wave of Eq. (2.1.4) in the perturbated Jeans's equation, Eq. (2.1.3), we find the *dispersion relation*, which relates ω to k, as follows:

$$\omega^2 = c_s^2 k^2 - 4\pi G\rho_b \quad . \tag{2.1.5}$$

Eq. (2.1.5) has two solutions, depending on the sign of ω^2. One is a *Real* solution while the other is an *Imaginary* one. We set $\omega^2 = 0$ and solve Eq. (2.1.5) for k:

$$k_j = \frac{\sqrt{4\pi G\rho_b}}{c_s} \quad .$$

Since the unit of $\vec{k}$ is the inverse of a length, we can recover the *Jeans's scale*:

$$\lambda_J = \frac{2\pi}{k_J} \equiv c_s\sqrt{\frac{\pi}{G\rho_b}} \quad .$$

which has the same dependency on c_s, ρ_b and G that we found before with the simple qualitative description. For $\lambda < \lambda_J \implies k > k_j \implies \omega^2 > 0$ we have:

$$\omega_{1,2} = \pm kc_s\left[1 - \left(\frac{\lambda}{\lambda_J}\right)^2\right]^{1/2} \quad ,$$

and the amplitude of plane wave does not change during propagation. Instead for $\lambda > \lambda_J \implies k < k_j \implies \omega^2 < 0$ and

$$\omega_{1,2} = \pm i\sqrt{4\pi G\rho_b}\left[1 - \left(\frac{\lambda}{\lambda_J}\right)^2\right]^{1/2} \quad .$$

In this case, due to the fact that $\delta = \delta_k \exp\left[i\vec{k} \cdot \vec{r} - i\omega t\right]$, the amplitude exponentially increase or decrease with time. We will consider the increasing solution because it leads to the collapsed structures. These results hold for a static Universe, but since we live in an expanding Universe we have to include the effects given by the expansion, which contrasts the structure growing.

2.2 Jeans Theory in expanding Universe

To describe the evolution of perturbations in an expanding Universe we define the *Cosmological Horizon*:

$$R_H \equiv a(t) \int_0^t \frac{c\,\mathrm{d}t'}{a(t')} \quad . \tag{2.2.1}$$

The *Cosmological Horizon* defines the fraction of the Universe in causal connection with the observer. The horizon is finite, due to limited speed of light, and grows with time.

2.2.1 Scales larger than the Cosmological Horizon

For scales larger than R_H the force of gravity is the most important and the others can be neglected. To take into account the expansion, we use a closed Universe, ($K = 1$), embedded into an EdS Universe ($K = 0$) to describe the perturbation. As discussed in Section 1.10, we know that at certain time a closed Universe starts to collapse. We write the Friedman equations for both Universes, as follows:

$$\begin{cases} H_p^2 + \frac{c^2}{a^2} = \frac{8\pi G}{3}\rho_b & \text{Perturbation} \\ H_b^2 = \frac{8\pi G}{3}\rho_b & \text{Background} \end{cases} \tag{2.2.2}$$

imposing $H_p = H_b$ we have:

$$\delta = \frac{3c^2}{8\pi G}\frac{1}{\rho_b a^2} \quad .$$

The perturbation growth depends on the dominant component of the background Universe. We can define the *Equivalence time*, t_{eq}, that is when the contribution to the mean density of the Universe due to the *Matter* and *Radiation* are equal (see Figure (1.1)). The time evolution of the density contrast before and after the equivalence is:

- $t < t_{eq};\ z > z_{eq};\ w = 1/3 \Longrightarrow \delta_{rad} \propto a^2 \propto t \quad .$
- $t > t_{eq};\ z < z_{eq};\ w = 0 \Longrightarrow \delta_M \propto a \propto t^{2/3} \quad .$

We conclude that outside the cosmological horizon the perturbation always grow.

2.2.2 Scales smaller than the Cosmological Horizon

Inside the *Cosmological Horizon* also the micro-physics is important which is described by the equations of hydrodynamics. We write them in physical units, which are connected to comoving coordinates as $\vec{r} = a\vec{x}$. The velocity is $\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = \dot{\vec{r}} = \dot{a}\vec{x} + a\dot{\vec{x}} = H\vec{r} + \vec{v} \equiv \vec{u}$, where $\vec{v}$ is the peculiar velocity, that is the object velocity with respect to the Hubble flow. The hydrodynamical equations become:

$$\begin{cases} \frac{\partial \rho}{\partial t}|_{\vec{r}} + \nabla \cdot (\rho \vec{u}) = 0 \\ \frac{\partial \vec{u}}{\partial t}|_{\vec{r}} + (\vec{u} \cdot \nabla)\vec{u} = -\frac{1}{\rho}\nabla p + \nabla \phi \\ \nabla^2 \phi = 4\pi G \rho \end{cases} \tag{2.2.3}$$

To solve these equation we apply the *Perturbation Theory* so we suppose that the background solutions are known and we add them a small perturbation. The background solution are:

$$\begin{cases} \rho = \rho_b(1 + \delta) \\ p = p_b + \delta p \\ \vec{u} = H\vec{r} + \vec{v} \\ \phi = \phi_b + \delta\phi \end{cases} \tag{2.2.4}$$

We place these solutions in Eqs.(2.2.3). Neglecting second order terms, we obtain:

$$\begin{cases} \frac{\partial \delta\rho}{\partial t} + \rho_b \nabla \cdot (\vec{v}) + 3H\delta\rho = 0 \\ \frac{\partial \vec{v}}{\partial t} + H\vec{v} + H\vec{r}\nabla \cdot \vec{v} = -\frac{1}{\rho_b}\nabla \delta p + \nabla \delta\phi \\ \nabla^2 \delta\phi = 4\pi G \delta\rho \end{cases} \tag{2.2.5}$$

or in co-moving coordinates we have:

$$\begin{cases} \frac{\partial \delta\rho}{\partial t}|_{\vec{x}} + \frac{\rho_b}{a}\nabla_x \cdot (\vec{v}) + 3H\delta\rho = 0 \\ \frac{\partial \vec{v}}{\partial t}|_{\vec{x}} + H\vec{r} = -\frac{c_s^2 \nabla_x \delta}{a} - \frac{1}{a}\nabla_x \delta\phi \\ \frac{\nabla_x^2 \delta\phi}{a^2} = 4\pi G \rho_b \delta \end{cases} \tag{2.2.6}$$

Where $\nabla_x = a\nabla$, is the relation between the differential operator ∇ in physical units and comoving coordinates. Eqs.(2.2.6) describe how perturbations evolve in an expanding Universe. We are looking for a plane wave solution. In this case the amplitude of density perturbation has a direct time

dependence. We put into Eq. (2.2.6) a plane wave solution like:

$$f(\vec{x},t) = f_k(t)\exp\left[i\vec{x}\cdot\vec{k}\right] \quad .$$

We find the following system of equations:

$$\begin{cases} \dot{\delta}_k + i\frac{\vec{k}\cdot\vec{v_k}}{a} = 0 \\ \dot{\vec{v}}_k + \frac{\dot{a}}{a}\vec{v_k} = -i\frac{\vec{k}}{a}\left[c_s^2\delta_k + \delta\phi_k\right] \\ \delta\phi_k = -\frac{4\pi G\rho_b}{k^2}\delta_k a^2 \end{cases} \tag{2.2.7}$$

It can be shown that $\vec{k}$ and $\vec{v_k}$ have the same direction, so we can combine the three equations system to obtain only one differential equation that describes the δ behavior:

$$\ddot{\delta}_k + 2H\dot{\delta}_k + \delta_k\left[\frac{k^2c_s^2}{a^2} - 4\pi G\rho_b\right] \quad , \tag{2.2.8}$$

in comoving coordinates Eq. (2.2.8) has two solutions, which depend whether the perturbation is larger or smaller than the *Jeans's length*:

- If $\lambda < \lambda_J$ or $k > k_J$, the solution propagates as a plane wave, and its amplitude does not increase or decrease with time.
- If $\lambda > \lambda_J$ or $k < k_J$, the solution has a *Real* time-dependent term thus the amplitude can change, either increasing or decreasing with time.

We are interested in the increasing solution, δ_*, of the second case because it leads to the formation of the cosmic structures. For a generic cosmology, there are no exact analytic solutions of Eq. (2.2.8), we can use the following approximate solution:

$$\delta_+ \propto -H(z)\int_\infty^z \frac{(1+z')}{a_0^2H^3(z)} \quad . \tag{2.2.9}$$

Eq. (2.2.9) holds for perturbations inside the cosmological horizon and for $t > t_{eq}$ ($z < z_{eq}$). From Eq. (2.2.9), we can see that in *Closed Universes* grow faster than perturbations in *Open Universes*. We note that Eq. (2.2.9) is still a linear solution, valid for $\delta \ll 1$.

Linear perturbation theory leads to another evidence of DM existence. The radiation pressure prevents the baryonic matter to collapse. In fact the baryonic matter can grow only after the *Recombination Time*, $z \sim 1100$. In order to reach at least $\delta \sim 1$ before $z = 0$ the starting overdensity

field of baryonic matter should have been $\delta \sim 10^{-3}$. This is in contrast to observations of CMB radiation which reveal an overdensity field of $\delta \sim 10^{-5}$. As shown by Eq. (2.2.9) in a *Closed Universe* the perturbations grow faster than in a flat one. In an Universe composed only by baryonic matter the total density parameter should be $\Omega \sim 20$ in order to justify the observed collapsed structures at the present time. This is in contrast with commonly accepted value $\Omega = 1$ measured from CMB (Sievers et al. [2003] or Planck Collaboration et al. [2018]).

From the *Recombination time* the baryonic matter falls into DM haloes, which are already grew up, thanks to their earlier decoupling. This behavior is called *baryon catch-up*, and allows baryonic overdensities to rapidly reach the values of δ_{DM}. The growth of dark matter and baryonic matter are shown in Figure 2.1. Eq. (2.2.9) can be used to define the so-called *growing*

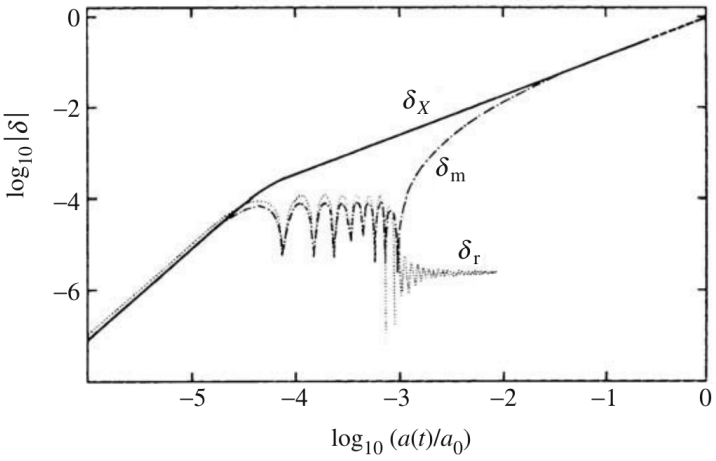

Figure 2.1: The evolution of perturbation δ_x is CDM, δ_m is baryonic component, and δ_r is the radiative component, at mass scale $M \sim 10^{15} M_\odot$. The radiation never grows and prevent the collapse of baryonic matter until recombination, after that baryonic matter falls into DM haloes. Credits to Coles and Lucchin [2003]

factor:

$$f \equiv \frac{\mathrm{d}\ln\delta_+}{\mathrm{d}\ln a} \approx \Omega_M^\gamma + \frac{\Omega_\Lambda}{70}\left(1 + \frac{\Omega_M}{2}\right) \quad , \tag{2.2.10}$$

which has a strong dependence on Ω_M, and a very weak dependence on Ω_Λ. In General Relativity $\gamma \approx 0.55$, thus measuring f is used to test the gravity theory.

2.3 Spherical collapse

The evolution of cosmic structures can be described analytically in the linear regime only when $\delta \ll 1$. The later non-linear evolution can not be described exactly with analytic models, but in simple, highly symmetric systems, such as in the case of the *spherical collapse*. In more general cases the non linear evolution model numerical simulations are used as we will see in Section 4.1.

The spherical collapse model describes the non-linear evolution of a spherical overdensity, δ, embedded in a background Universe. We suppose that at the beginning the perturbation follows the evolution of the background Universe. For simplicity at the initial time we assume an EdS background Universe. The spherical collapse model treats the overdensity as a separate Universe with its own density parameter, Ω_p. The overdensity will collaps if $\Omega_p > 1$ and the following equations hold:

$$\begin{cases} \delta(t) = \frac{3}{5}\delta_+(t_i)\left(\frac{t}{t_i}\right)^{2/3} \\ \delta_+ > \frac{3}{5}\frac{1-\Omega_b(t_i)}{\Omega_b(t_i)(1+z_i)} \end{cases} \tag{2.3.1}$$

where Ω_b is the density parameter of the background Universe and δ_i is the initial overdensity which depends on redshift. From Eq. (2.3.1) we see that for closed and flat Universess the any overdensity can collapse, while in open Universes only the overdensity that exceed a threshold value can collapse. The expansion of the perturbation is described by the Friedman equation (Eq. (1.8.6)). The overdensity follows the evolution of background Universe until gravity stops the perturbation expansion at the *turn-around point* and the overdensity starts to collapse. The turn-around time reads:

$$t_{turn} = \left[\frac{3\pi}{32G\rho(t_{turn})}\right]^{1/2} , \tag{2.3.2}$$

where $\rho(t_{turn})$ is the perturbation density. The complete collapse is achieved at $t_c \approx 2t_{turn}$. Once the overdensity is collapsed, it can reach the virialization at time $t_{vir} \approx 3t_{turn}$. We can define the ratio between the perturbation density and the background mean density, $\chi(t)$. If the background is still EdS at the collapse time we find:

$$\chi(t_c) = \frac{\rho(t_c)}{\rho_b(t_c)} \sim 180 ,$$

$$\chi(t_{vir}) = \frac{\rho(t_{vir})}{\rho_b(t_{vir})} \sim 400 .$$

These values strongly depends on cosmology. Using the linear theory, the overdensity value of collapsed structures would be:

$$\delta(t_c) \sim 1.68 \quad . \tag{2.3.3}$$

This value is not cosmology-dependent. This fact will be useful in the next Section 2.4.

2.4 The halo mass function

Let us now define the halo mass function (MF) as the number of objects with mass between M and $M + \mathrm{d}M$ per unit of comoving volume. The MF allows us to summarize the main statistical properties of the collapsed DM haloes, being related to their formation and evolution. To model the HMF, we have to integrate the probability distribution of perturbations, $P(\delta)$, which tells us the probability of having a perturbation above a given threshold. We assume this distribution to be Gaussian at the beginning of the perturbation growing, that is immediately after the *inflation*. When the growth is non-linear, the distribution can not be Gaussian anymore, and its *skewness* becomes not null. We integrate $P(> \delta_c)$, where δ_c is the critical overdensity above which the perturbation will become a collapsed structure. We can exploit the fact that the $P(\delta)$ in the linear regime, still is a Gaussian distribution. We integrate $P(> \delta_c)$ above the overdensity threshold $\delta_c \sim 1.68$, as introduced in Section 2.3.1. This value is the critical overdensity predicted by linear theory in the spherical collapse assumption, and does not depend strongly on cosmology. As shown by Press and Schechter [1974] it is possible to find an analytical form of the HMF, as follows:

$$n(M,z)\mathrm{d}M = \sqrt{\frac{2}{\pi}}\frac{\bar{\rho}_M(z)}{M^2}\frac{\delta_c}{\sigma_M}\left|\frac{\mathrm{d}\ln\sigma_M(z)}{\mathrm{d}\ln M}\right|\exp\left(-\frac{\delta_c^2}{2\sigma_M(z)}\right)\mathrm{d}M \tag{2.4.1}$$

where $\bar{\rho}_M(z)$ is the mean matter density δ_c is the threshold of the overdenty that allows the collapse, and σ_M is the root mean square of the mass fluctuations.We note that Eq. (2.4.1) holds only for structures formed through a spherical collapse. It is possible to find an analytical solution of the MF even considering an ellipsoidal collapse, as showed by Sheth and Tormen [1999] and Sheth et al. [2001]. Figure 2.2 shows the evolution of the MF with redshift. The MF can be modelled as a Schechter function, that is a

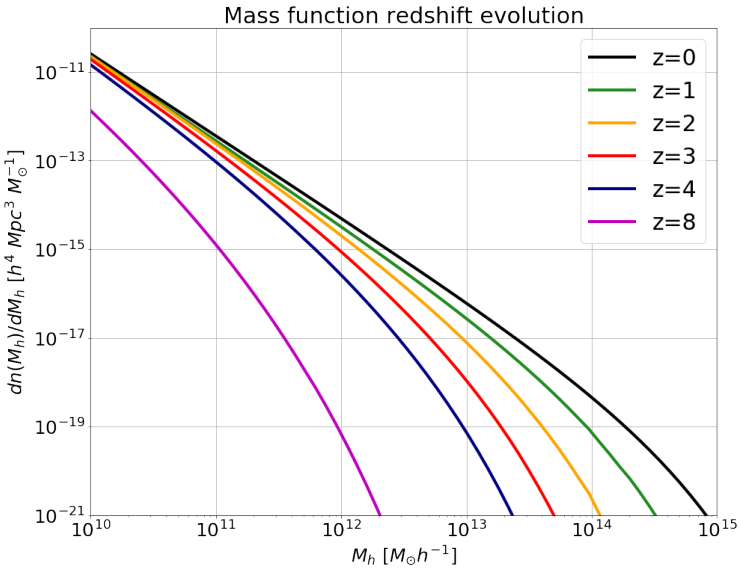

Figure 2.2: The evolution of the HMF as a function of the redshift in a ΛCDM cosmology. The density of objects decreases with the redshift and increases with time. The image is realised using the functions implemented in the public set of libraries CosmoBolognaLib.

power law with an exponential cut off at the characteristic mass, M^*, given by:

$$M^* \equiv M_0 \left(\frac{2}{\delta_c^2} \right)^{1/2\alpha} \quad , \tag{2.4.2}$$

where $\alpha = (n+3)/6$ while M_0 is the normalization. Due to the Eq. (2.4.2), the value of M^* increases with time, because it depends on n through α. At the present time its value is about $M^* \approx 10^{15} M_\odot$, which is the typical mass of cluster galaxies.

2.5 Halo properties

Once an halo is collapsed we can argue of their properties as a single object. We introduce the DM density profile inside a collapsed halo because it will be useful in Chapter 5. The DM density profile inside an halo can be described by the Navarro-Frenk-White (NFW) profile (Navarro et al. [1997]):

$$\rho_{DM}(r) = \frac{\rho_s}{\frac{r}{r_s}\left(1+\frac{r}{r_s}\right)^2} \quad , \tag{2.5.1}$$

where r_s is a scale radius, which depends on the halo mass and on the dimensions of the halo. The $\rho_s = \delta_c \rho_{crit}$ where ρ_{crit} has been defined by Eq. (1.8.3) and δ_c is the critical overdensity, which can be written using the concentration parameter, c:

$$\delta_c = \frac{200}{3} \frac{c^3}{\ln(1+c) - \frac{c}{1+c}} \quad .$$

The concentration parameter, c, can be defined as $c_{200} \equiv r_{200}/r_s$, where r_{200} is the distance from the halo center which contains DM desinty 200 times the ρ_{crit} of the Universe. Another possibility is to define the concentration parameter using the virial radius $c_{vir} \equiv r_{vir}/r_s$. The concentration parameter is a decreasing function of the halo mass and of the formation time of the halo. Indeed small haloes form at higher redshift, when the mean density of the Universe was higher, are denser. While massive objects form at relative recent time, when the mean density of the Univere is decrease.The concentration parameter depends on the halo mass, and this can be written following Bullock et al. [2001] as:

$$c_{vir} \approx \frac{9}{1+z} \left(\frac{M_{vir}}{1.310^{13} h^{-1} M_\odot} \right)^{-0.13} \quad . \tag{2.5.2}$$

The quantity r_s in Eq. (2.5.1) can be written as $r_s = \frac{r_v ir}{c_v ir}$, where r_{vir} is the radius that contains all the virial mass of the DM halo. We note that the NFW profile for $r \ll r_s$ follows $\rho_{DM} \propto r^{-1}$, while for the outermost regions $\rho_{DM} \propto r^{-3}$.

2.6 Subhalo mass function and infall mass

According to the standard scenario of structures formation, the haloes assembly through hierarchical mergers of smaller systems. We expect that inside large haloes there are many *subhaloes* (e.g. Springel et al. [2008]), which are the remnants of the haloes that survived to accretion into larger host haloes. This scenario can be represented in Figure 2.3. The theoretical study of these substructures is allowed by N-body simulation with high mass resolution,where single subhalo can be detected and its evolution can be traced.From these simulations it is also possible to study the statistical properties of subhaloes population, like the Subhaloes Mass Function (SHMF), which represent the number of substructures per halo mass. The

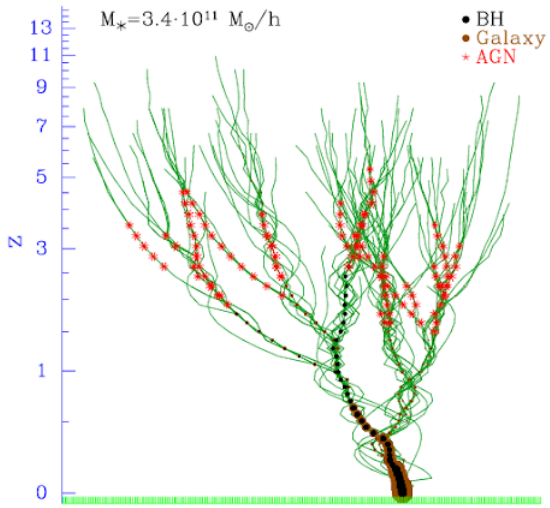

Figure 2.3: Schematic view of a merger tree, representing the growth of haloes and galaxies through mergers. The green lines track the DM haloes progenitors, while brown and black dots represent progenitor galaxies and central supermassive black holes, respectively. The dot sizes are proportional to the mass of galaxies and black holes. The presence of AGN is indicated with red stars and their sizes are proportional to the AGN bolometric luminosities. Credits to Marulli et al. [2009]

SHMF has a universal form, as discussed in Giocoli et al. [2010], which is a power law at small subhaloes masses with a strong exponential cut off. The SHMF depends on the halo mass, and has an universal form, as:

$$\frac{\mathrm{d}N}{\mathrm{d}\log(m_{sub}/M)} = A(1+z)^{1/2} m_{sub}^{\alpha} M \ln 10 \exp\left[-\beta\left(\frac{m_{sub}}{M}\right)^3\right] \quad , \qquad (2.6.1)$$

where $A = 9.3310^{-4}$ is the normalization, $\alpha = -0.9$ is the exponent of the power law at smallest subhalo masses, and $\beta = 12.2715$ determines the steepness of the exponential cut off. These parameter values, found by Giocoli et al. [2010], are used in Chapter 5. It is commonly accepted that the major part of substructures contain galaxies, the so-called satellite galaxies. Both, the subhaloes and the galaxies inside them are subject to several environmental effects, as tidal forces from host haloes, dynamical friction, impulsive encounter. Each effect tends to dissolve the subhaloes. In low

mass haloes, that on average form early, the substructures have suffered more mass loss. This bring to a shift in the normalization of the SHMF. The comparison between the SHMF at the accreation time with the present time SHMF, allow to evaluate the mass loss rate of the substructures. Figure 2.4 shows the SHMF dependens on redshift and on the host halo mass. For

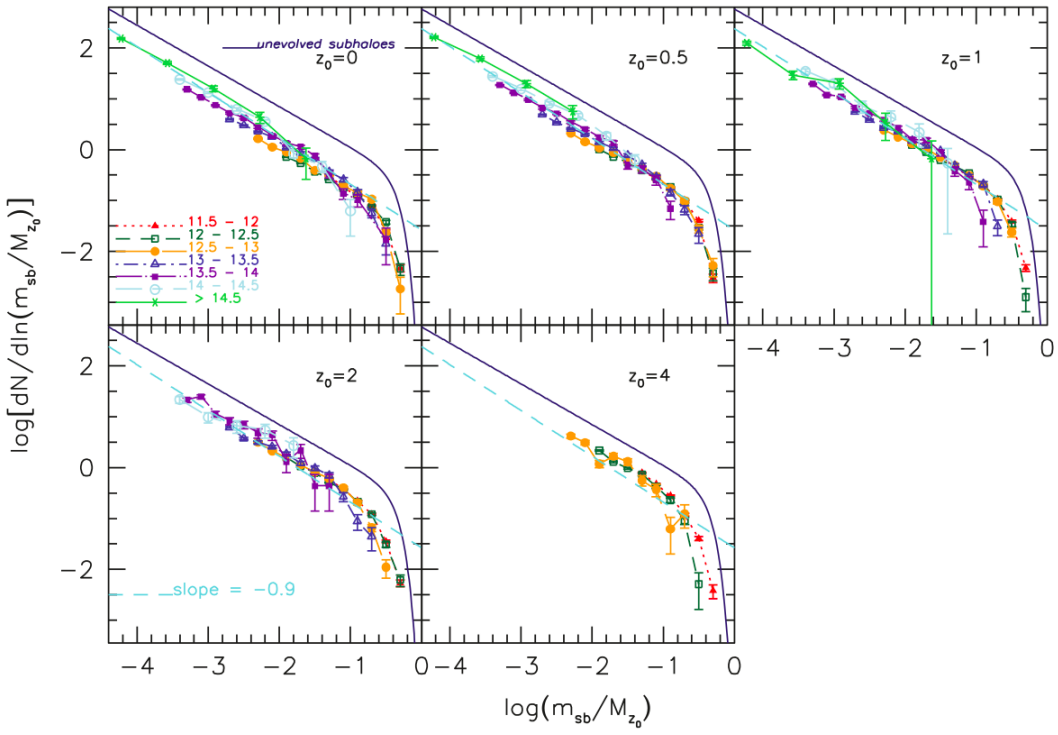

Figure 2.4: In these plot the coloured dots represent the subhalo mass functions measured in different host halo mass bin at five different redshifts. The dashed line is the best fit for evolved SHMF while solid curve represents the unevolved subhalo mass function. Credits to Giocoli et al. [2010] .

these reasons the mass of each subhalo is a decreasing function of time. We denote with m_{infall} the subhalo mass at the accreation time, that is when the subhalo felt into the greater halo. The subhaloes nearest to the host halo center are the substructures accreated earlier, and so they lost a great part of their mass. On the other hand the substructures more distant from the host halo centre had less time to be stripped. Through mergers tree extracted from high resolution N-body simulations it is possible to parameterise the

relation between infall mass and the position of the substructures inside the halo (e.g. Gao et al. [2004], Girelli et al. [2020]):

$$\frac{m_{obs}}{m_{infall}} = 0.63\left(\frac{r}{R_{200}}\right)^{2/3} \quad , \tag{2.6.2}$$

where m_{obs} indicates the subhalo mass at the present time, while R_{200} is the radius at which the density of the halo is 200 times the critical density, ρ_{crit}. It is possible to find also a relation between the substructure mass and the redshift of accreation, e.g. inverting the relation presented in Giocoli et al. [2008]:

$$m_{infall}(t) = m_{obs}\exp\left[\frac{t - t_{infall}}{\tau(z)}\right] \quad , \tag{2.6.3}$$

where t_{infall} is the infall time of the substructure and $\tau(z)$ describes the dependence of the mass loss rate with redshift. Van Den Bosch et al. [2005] proposed a functional form for $\tau(z)$, as follows:

$$\tau(z) = \tau_0\left[\frac{\Delta_V(z)}{\Delta_0}\right]^{-1/2}\left[\frac{H(z)}{H_0}\right]^{-1} \quad , \tag{2.6.4}$$

where $\tau_0 = 2$ Gyr, $H(z)$ is the Hubble parameter at redshift z and $\Delta_V = \rho/\rho_{crit}$. Let us introduce the subhalo mass fraction which represent the contribution of the substructure mass to the host halo mass. It is defined by the integral of the SHMF as:

$$f = \int_{m_{min}}^{M} m\frac{\mathrm{d}N}{\mathrm{d}m}\mathrm{d}m \quad , \tag{2.6.5}$$

where m is the subhalo mass, m_{min} is the minimum substructure mass and M is the host halo mass. As shown in Figure 2.5, the subhalo mass fraction is an increasing function of halo mass, therefore less massive haloes, on average form earlier and have less surviving substructure at the present. This because the tidal forces strip substructure mass for a longer time. If the tidal stripping act for a sufficient time the substructure is completely dissolved (e.g. Giocoli et al. [2010], Van Den Bosch et al. [2005]).

2.7 Correlation Function and Power Spectrum

As shown in Subsection 2.2.2, in linear regime we can describe the perturbation growth in Fourier space. Now we want to study the evolution of

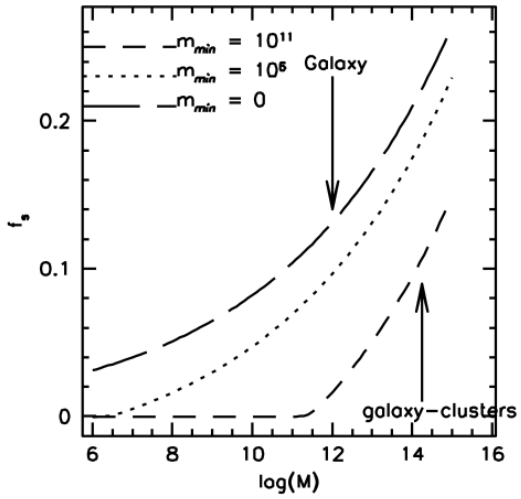

Figure 2.5: These curves describe the subhalo mass fraction, f, as function of host halo mass. The f depends on the minimum subhalo mass that we are considered. Credits to Giocoli et al. [2010]

perturbations looking at their full spectral composition. The end of the *Inflationary epoch* generates some perturbations of the metric of the space-time, that correspond to fluctuations of the gravitational potential, ϕ, from a Newtonian point of view. These fluctuations are connected to the density field through the Poisson equation Eq. (2.1.1). We suppose that these fluctuations are randomly distributed, thus also $\delta(\vec{x})$ is a stochastic field and then the *Imaginary* part of the δ_k in Fourier space is randomly distributed. This leads to a Gaussian distribution of the perturbations, both in Fourier space and in real space. The relations between real space and Fourier space, are the following:

$$\delta(\vec{x}) = \frac{1}{(2\pi)^3} \int \delta(\vec{k}) \exp\left[i\vec{k}\cdot\vec{x}\right] \mathrm{d}\vec{k} \quad ,$$

$$\delta(\vec{k}) = \int \delta(\vec{x}) \exp\left[-i\vec{k}\cdot\vec{x}\right] \mathrm{d}\vec{x} \quad .$$

with $\vec{k} = \frac{2\pi}{\vec{x}}$.
An interesting quantity is the correlation function:

$$\xi(\vec{r}) = \xi(r) \equiv \left\langle \delta(\vec{x})\delta(\vec{x}+\vec{r}) \right\rangle \quad . \tag{2.7.1}$$

This quantity measures the probability to have an overdensity δ in $\vec{(x)}+\vec{r}$, given the overdensity in $\vec{x}$, $\delta(\vec{x})$. It depends on module of $|(\vec{r})| = r$ due to

the isotropy of the Universe. We can write the correlation function using the Fourier transform of δ, as follows:

$$\xi(r) = \frac{1}{(2\pi)^3} \int P(k) \exp\left[i\vec{k} \cdot \vec{r}\right] \mathrm{d}\vec{k} \quad , \tag{2.7.2}$$

where $P(k) \propto \left\langle |\delta_k|^2 \right\rangle$ is the so-called Power Spectrum, that is the power spectrum is the Fourier transform of the correlation function.

Due to the definition of δ, the mean of its distribution is null and the first quantity of interest is the *variance*, σ^2, which is defined as follows:

$$\sigma^2 \equiv < \delta^2 > = \frac{1}{V} \int \langle \delta^2(\vec{x}) \rangle \mathrm{d}\vec{x} \quad ,$$

where V is the volume of the observed Universe. The variance can be also written using the power spectrum:

$$\sigma^2 = \frac{1}{V} \int P(k) \mathrm{d}\vec{k} = \frac{4\pi}{V} \int P(k) k^2 \mathrm{d}k \quad . \tag{2.7.3}$$

The overdensity field $\delta(\vec{x})$ is generally assessed through observable *tracers* such as galaxies and clusters of galaxies, in a finite volume. Differently to the continuum matter overdensity field, the distribution of cosmic tracers is discrete. If we count galaxies in a volume V_i, we can still define an overdensity field as:

$$\delta_{gal}(V_i) = \frac{N_{gal}(V_i) - \bar{N}_{gal}(V_i)}{\bar{N}_{gal}(V_i)} \quad ,$$

where N_{gal} is the number of galaxies in the volume V_i, and $\bar{N}_{gal}$ is the mean number of galaxies in that volume. The previous quantity is like a smoothing of the discrete distribution of galaxies. The same thing can be done with the mass, describing the mass overdensity field δ_M. The quantity that connects δ_{gal} and δ_M is called *bias factor* b:

$$\delta_{gal} = b\left(\frac{M(V_i) - \bar{M}(V_i)}{\bar{M}(V_i)}\right) = b\delta_M \quad , \tag{2.7.4}$$

The bias b describes the differences between the collapse of baryonic matter and DM. The matter density contrast, δ_M, is measured by filtering the overdensities above a certain mass, or equivalently above a certain dimension R. This can be written using a *window function* to convolve the overdensity field $\delta(\vec{x})$:

$$\delta_M(\vec{x}) = \delta(\vec{x}) \otimes W(\vec{x}, R) \quad .$$

Usually $W(\vec{x}, R)$ is taken as a Top Hat function:

$$W(\vec{x}, R) = \frac{3}{4\pi R^3}\Theta\left(1 - \frac{|x - x'|}{R}\right) \quad ,$$

where Θ is the Heaviside step function. We can thus write the mass variance using the window function as follows:

$$\sigma^2 = \frac{1}{(2\pi)^3}\int P(k)\hat{W}^2(\vec{k}, R)\mathrm{d}\vec{k} \quad ,$$

where $\hat{W}$ is the Fourier transform of the windows function. So the values of the mass overdensity and the mass variance depend on the mass filter or, equivalently, on the dimension of region considered in the windows function:

- If $R \to 0 \;\; (M \to 0) \implies \delta_M(\vec{x}) \to \delta(\vec{x}) \implies \sigma_M^2 \to \sigma^2$.
- If $R \to \infty \;\; (M \to \infty) \implies \delta_M(\vec{x}) \to 0 \implies \sigma_M^2 \to 0$.

2.8 The time evolution of the Power Spectrum

The primordial matter power spectrum can be described by a power law $P(k) = Ak^n$, with no special scale. The variance can be written as follows:

$$\sigma^2 = \frac{A}{(2\pi)^3}\int_0^\infty k^{2+n}\mathrm{d}k \quad ,$$

which converges if for $k \to 0 \implies n > -3$ and for $k \to +\infty \implies n < -3$. The power spectrum evolves in time as follows:

$$P(k, t) = P_{in}(k)\delta_+^2(t) \quad . \tag{2.8.1}$$

Combining Eq. (2.7.3) with Eq. (2.8.1), we obtain:

$$\sigma_M^2 \propto \delta_+^2(t)k^{n+3} \quad ,$$

and since $k \propto R^{-1}$ and $M \propto R^3$ we can write:

$$\sigma_M^2 \propto M^{-(3+n)/3} \quad .$$

Standard inflation models predict a *white noise* in the metric perturbation, that is no special scales are promoted at the end of inflationary epoch. From a Newtonian point of view the metric fluctuations are fluctuations of the gravitational potential, therefore:

$$\delta\phi \propto \frac{G\delta M}{R} = G\delta\rho R^2 \propto \sigma_M R^2 \propto M^{\frac{1-n}{6}} \quad .$$

In order to have *white noise* we request that $n = 1$. This means that standard inflationary models predict a primordial spectrum, called *Zeldovich power spectrum* like:

$$P_{in}(k) = Ak \quad . \tag{2.8.2}$$

Where A is the normalization of power spectrum which can be constrained from observations. According to Eq. (2.8.1), the primordial power spectrum increases of the same magnitude on every scales. Indeed, it translates rigidly increasing its normalization, as shown in Figure 2.6.

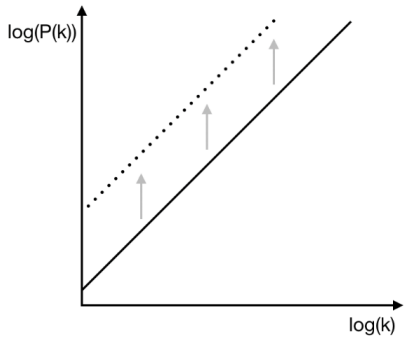

Figure 2.6: Evolution of the primordial power spectrum, consisting in constant increasing at all scales.

As explained in Section 2.2.1, all perturbations that are outside the Cosmological Horizon, R_H, always grow, while inside the Cosmological Horizon the growth depends on which is the dominant component at that Epoch. Because R_H increases with time, more perturbation scales enter into the cosmological horizon during the evolution. A perturbation stops its growing rate when enters into the Cosmological Horizon leading to a modification of power spectrum shape. To describe the power spectrum evolution, we introduce the *Transfer Function*, $T(k)$:

$$P(k, t_{eq}) = P_{in}T^2(k) = AkT^2(k) \quad ,$$

with:

$$T^2(k) \propto \begin{cases} k^{-4} & \text{for } k > k_{eq} = 2\pi/R_H(t_{eq}) \\ 1 & \text{for } k < k_{eq} = 2\pi/R_H(t_{eq}) \end{cases}$$

The peak is at k_{eq}, that is the dimension of the Cosmological Horizon in the Fourier space at the *Equivalence time*. The position of this peak strongly

depends on Ω_M. Increasing Ω_M brings forward the *Equivalence time*, moving the peak towards greater values of k. While a smaller value of Ω_M delays the *Equivalence time*, moving the peak towards smaller values of k.The power spectrum shape is shown in Figure 2.7.

Another important scale is the *free streaming* scale, k_{FS}, where $P(k_{FS}) = 0$. The *free streaming* scale can be associated to the minimum mass that can grow, that is the $Jeans's\ Mass$. The k_{FS} can be used to discriminate which is the kind of DM can bring to the observed collapsed structures. It comes out that $k_{FS_{CDM}} > k_{FS_{HDM}}$, as shown in Figure 2.7. This means

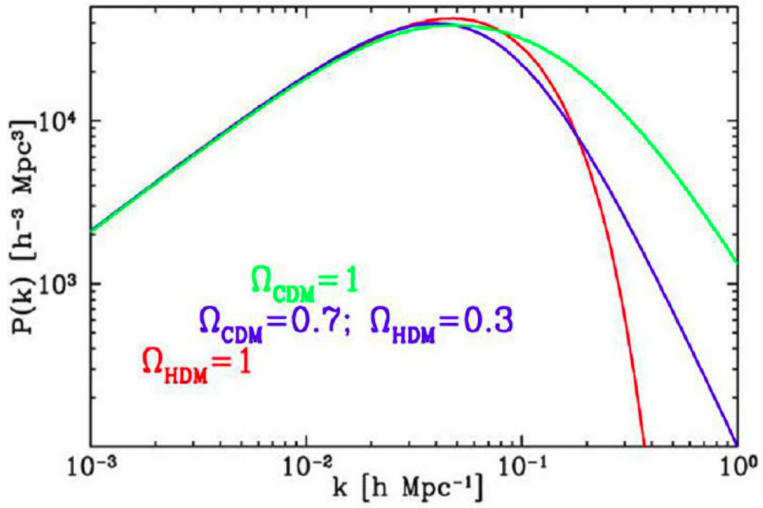

Figure 2.7: Differences in power spectrum between CDM and HDM. Note that $k_{FS_{CDM}}$ is much bigger than $k_{FS_{HDM}}$ bringing to a minimum collapsed mass much smaller for CDM scenario with respect to HDM scenario. Figure credits to Kolb and Turner [1990].

that:

$$M_J \sim \begin{cases} 10^{16} M_\odot & \text{for HDM} \\ 10^{5-6} M_\odot & \text{for CDM} \end{cases}$$

This fact leads to two different models of structure formation. For HDM the massive objects are formed first and the smaller structures are created by fragmentation of the bigger structures. This is the so-called *monolitic model* (or *Top-Down scenario*). Instead, in the CDM scenario, the smallest structures formed first, and the bigger ones are obtained by mergers. This is the so-called *hierarchical model* (or *Bottom-Up scenario*).

The values of M_J for HDM is bigger than the mass of the most massive observed structures, that is clusters of galaxies. Thus the observations supports the CDM scenario, rather than the HDM one.

Chapter 3

Galaxy formation and evolution

In the previous Chapters (1 and 2) we introduced the cosmological framework in which this Thesis work is developed. In this Chapter we will introduce some general properties of galaxies, such as colours and stellar masses, and we will discuss the standard galaxy formation and evolution models. Moreover, we will introduce the Stellar Mass Function (SMF), which is a useful summary statistics that can be exploited to infer information about the physical mechanisms that drive the evolution of galaxies. Given the wide variety of the topic, in the next Sections we will not be able to provide an exhaustive description of the field. Instead, we will focus on the subjects that are more related to this Thesis work. For a more detailed review of the topic, see Somerville and Davé [2015], Naab and Ostriker [2017], Robertson [2019]

3.1 Galaxy classification

Galaxies are the *building blocks* of the Universe. They are made of DM, stars and of the so-called Inter Stellar Medium (ISM), which is in turn composed of gas and dust. Galaxies are non-homogeneously spatially distributed, aggregating in filaments, groups and clusters, and are biased tracers of the underlying DM mass distribution of the Universe.

According to the standard structure formation scenario, galaxies can form only inside DM haloes, as mentioned in Section 2.2.2. DM constitutes

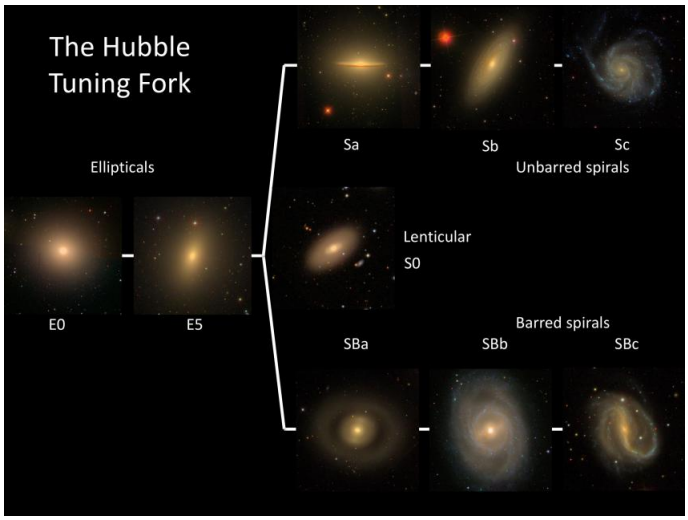

Figure 3.1: Hubble's fork diagram showing the galaxy morphological classification. On the left are placed elliptical galaxies, while going to the right spiral galaxies and barred spirals are represented. Credit: Karen Masters, Sloan Digital Sky Survey, https://ras.ac.uk/media/37.

the main contribution to the total galaxy mass. Due to the fact that galaxies reside inside DM haloes, we expect that the properties of galaxies and haloes are connected. The study of how galaxies are distributed within DM haloes is one of the main goals of this Thesis work, as will be discussed in Chapter 4.

From the first observations of the Large Scale Structure (LSS) of the Universe in the early 1920's, it was evident that galaxies can be divided into two main families: the so-called *Early Type Galaxies* (ETGs), i.e. elliptical galaxies, and the *Late Type Galaxies* (LTGs), i.e. spiral galaxies. ETGs are typically found in high density environments, that is groups and clusters of galaxies, while LTGs are often located inside the filaments of the cosmic web, or are isolated. The most evident difference between these two galaxy types is in their morphology. Edwin Hubble was the first to propose a classification based on galaxy morphology. This classification is known as the $Hubble's\ fork$ (Hubble [1926]), and is illustrated in Figure 3.1. ETGs and LTGs differ also in their stellar population.

Let us define the *colour* of galaxies as the difference in the observed magnitudes referred to two different band filters. Generally, ETGs result to

be redder than LTGs, the majority of which are blue galaxies. This colour bimodality is corroborated by several observational studies (e.g. Strateva et al. [2001] , Bell et al. [2004], Balogh et al. [2004], Weiner et al. [2005], Dekel and Birnboim [2006], Pozzetti et al. [2010]). Since ETGs show redder colours, it is commonly accepted that they are composed by quiescent stellar populations. This means that ETGs are typically not forming new stars, which implies that some mechanisms are required to inhibit star formation. On the other hand, LTGs are blue because they are still forming stars. In fact, they contain significant amount of cold gas and dense molecular clouds, which can collapse after radiative cooling.

3.2 Galaxy formation and evolution

The galaxy bimodality suggests different channels for formation and evolution of these two types of galaxies. A complete and exhaustive description of the theory of galaxy formation and evolution in the standard cosmological framework is beyond the scope of this Thesis work. In the following we will just resume the main features of the standard evolutionary scenarios.

As discussed in Section 2.2.2, from the *Recombination time* the baryonic matter falls into DM haloes, which are already grew up. The gas, falling into DM potential wells, converts kinetic energy (given by the infall velocity) to heat through shock fronts. The gas is heated to the virial temperature of the DM halo, T_{vir}.At this point, the baryonic matter can cool down and collapse due to radiative cooling. In fact, cooling is one of the fundamental ingredients for galaxy formation. Depending on the gas temperature and density, several different radiative processes can occur. These processes can be modelled through the so-called *cooling function* $\Lambda(T, \rho, x, Z, z)$, that describes the amount of energy, per unit of time, that is radiated by the gas. This function depends on gas temperature, T, gas density, ρ, chemical composition, Z, ionization fraction, f, and redshift, z. The leading parameter is the gas temperature. Indeed different emission processes mainly depend on gas temperature. For example, for $T_{vir} > 10^7 K$, the gas is completely ionized and the main emission mechanism is breemsstrahlung from free electrons. For temperature $10^4 K < T_{vir} < 10^6 K$, recombinations of electrons and ions and collisional excitation occur. Because different atoms have different bounding energies in these temperature ranges, the cooling function

strongly depends on the gas metallicity, Z. At temperature $T_{vir} < 10^4 K$, the gas is completely neutral and cooling can act only through vibrational and rotational excitations and de-excitations of the molecules. For this reason, the cooling function at low temperature is not very efficient. During baryonic collapse two competing processes occur. The fist is the gas cooling, which favours the collapse, while the second is the gas compression during the inward fall which increases gas density and temperature, preventing further collapse. To model these competitive processes, it is convenient to define the typical time scales. We define the *cooling time*, t_{cool}, as follows:

$$t_{cool} = \frac{E_{gas}}{\dot{E}} = \frac{3}{2}\frac{(n_e + n_H)k_B T}{\Lambda n_e n_H} \quad , \tag{3.2.1}$$

where n_e is the electron number density, n_H the hydrogen number density, k_B is the Boltzmann constant, T is the gas temperature, and Λ is the cooling function. This time scale describes how fast the gas with energy E_{gas} can lose all its energy. The smaller is t_{cool}, the more the gas cooling is efficient. The dynamic time scale describes the typical time that a gas sphere with density ρ needs to collapse, due to its own gravity. It can be approximated as follows:

$$t_{dyn} \propto \left(\frac{1}{G\rho}\right)^{1/2} \quad , \tag{3.2.2}$$

where G is the Newtonian universal gravitational constant. Finally, we consider the *Hubble time*, t_H, which is the inverse of the Hubble constant, and proportional to the age of the Universe, as mentioned in Chapter 1. The comparison between these time scales allows us to understand if cooling is sufficiently effective to permit the gas collapse:

- if $t_{cool} < t_{dyn}$, the cooling is efficient and the collapse proceed until $t_{cool} = t_{dyn}$.
- if $t_{dyn} < t_{cool} < t_H$, the cooling is not much efficient, though it permits the collapse. The gas can be considered in hydrostatical equilibrium during the collapse.
- if $t_{cool} > t_H$, the collapse cannot not happen, as the cooling time exceeds the entire age of the Universe.

These conditions are valid in every formation scenarios.

The efficiency of gas cooling is only one of the parameters affecting the formation of galaxies. Other physical quantity that has to be considered is the gas angular momentum, which derives from possible tidal interaction.

In the following we will describe the standard formation scenarios of disk galaxies (LTGs) and spheroidal galaxies (ETGs).

3.2.1 Disk galaxies

In 1962, the so-called *top-down* scenario of galaxy formation and evolution was proposed by Olin Eggen, Donald Lynden-Bell, and Allan Sandage (Eggen et al. [1962]). Specifically, they proposed that galaxies come from the monolitic collapse of a large cloud of gas, that during the contraction fragments into smaller cores. Then cores continue to collapse because they are still cooling.Tidal interactions between different cores generate initial angular momentum. If the gas has a large initial angular momentum during the collapse, the momentum will increase generating an highly rotating gas disk. The rotating disk continues to cool from the inner zone outward. The disk that is cooling is unstable and gas fragmentation occurs leading to the formation of stars. This scenario is the analogue of the cosmological model of structure formation with HDM (see Chapter 2).

In the alternative *bottom-up* scenario, the first formed objects are the smaller ones, with the typical mass of globular clusters. These structures are then assembled by gravity forming gradually larger structures, up to the formation of clusters of galaxies. This scenario is in accordance with the commonly accepted cosmological theory of hierarchical structure formation in ΛCDM (see Chapter 2). Again, if the initial angular momentum of the gas is high, during the collapse the gas will further increase its angular velocity forming a disk. This scenario predicts a larger number of small galaxies than the *top-down* model.

As described in Chapter 2, the hierarchical *bottom-up* scenario is consistent with the main cosmological observations at large scales (see e.g. Kauffmann et al. [1993], Reed et al. [2003]), while the *top-down* scenario has been ruled out by observations.

3.2.2 Spheroidal galaxies

In the formation and evolution of spheroidal galaxies, galaxy mergers are crucial. Mergers can be classified as major mergers or minor mergers, de-

pending on the mass ratio of the interacting galaxies. What differs between these two scenarios is the remaining stellar kinematics of the formed galaxy (Emsellem et al. [2007]).

Numerical simulations showed that major gas-rich mergers of disk galaxies (*wet mergers*) can generate a final system with a spheroidal shape, characterized by a random velocity distribution of the stars [Toomre and Toomre, 1972]. A *wet merger* is followed by an intense star formation. Indeed, if the parent merging galaxies contain a significant amount of gas, this will constitute the fuel for star formation. Because of the merging event, the gas density grows and the radiative cooling becomes more efficient. At this phase, some feedback processes can arise and slow down the star formation. These feedback mechanisms are though to be caused by SNe explosions, which heat gas through the kinetic energy of the expanding shock, preventing further star formation. Moreover, SNe can eject gas from the smallest galaxies, removing the star formation fuel. Some of the gas can fall onto the super massive black hole (SMBH) hosted at the galaxy centre. This gas is heated by friction and emits energy, forming an hot accretion disk around the SMBH. Several other emission mechanisms and physical processes are involved, such as relativistic ject, radio emission through syncrotron processes, X-ray emission from the region close to the SMBH, and the set of all these emissions bring to the formation of the so-called Active Galactic Nucleus (AGN). The emission from AGN can injects a significant amount of energy in galaxy, which might heat the gas, possibly quenching the star formation. This process is the so-called AGN feedback.

Though galaxy mergers play a crucial role in the formation and evolution of galaxies, observational evidences suggest that alternative formation scenarios for spheroidal galaxies are still possible. ETGs can born in a monolithic way, from a gas could with a small initial angular momentum, which can rapidly cool down leading to the formation of a spheroidal system [Eggen et al., 1962]. Another possible way to form a spheroidal galaxy is from violent disk instabilities [Ceverino et al., 2014]. The subsequent galaxy evolution can also be influenced by mergers, especially the more frequent gas-poor mergers (*dry mergers*). In these cases, there is a few gas that can form stars.

The environment also plays an important rule in galaxy evolution. Inside groups and clusters of galaxies, we can consider separately the galaxies

residing at the center of the host halo, the so-called central galaxies, and the satellite galaxies which are orbiting around a central galaxy. According to the hierarchical scenario, satellite galaxies were central galaxies inside their previous hosting halo, which subsequently felt into a larger potential well and became part of a more massive DM halo. Central and satellite galaxies are subject to different evolutionary processes. Central galaxies continue to receive cold gas, which tends to accumulate at the bottom of the potential well. Satellites galaxies tend to fall towards the halo centre due to dynamical friction. So we expect that central galaxies continue to grow due to accretion of cooling gas and mergers of satellite galaxies. Instead, satellite galaxies undergo through different environmental effects, such as ram-pressure stripping, tidal stripping, tidal interaction with the host halo itself and interactions with other satellite galaxies which can cause a morphological transformation, as shown e.g. in Moore et al. [1999].

Several observational evidences suggest the most massive ETGs formed at higher redshift and more rapidly than the less massive galaxies of the same type [e.g. Thomas et al., 2005, Bell et al., 2005, Cimatti, A. et al., 2006, Graves et al., 2007, Fontanot et al., 2009]. This scenario has been called *downsizing* and it is illustrated in Figure 3.2. In particular, Fig. 3.2 shows the specific star formation rate (sSFR), which is the star formation rate (SFR) divided for the galaxy stellar mass, as a function of time. The peak position provides an indication of the galaxy formation time. The more massive galaxies formed at higher lookback times (higher redshifts). The height of the curves measures the strength of the sSFR, while their width indicates the time that SFR lasts. The *downsizing* scenario may seem to be not consistent with the standard cosmological hierarchical model, which predicts that the objects form from the smallest to the biggest. However, these two scenarios can in fact co-exist, as they are referred to different phenomena. The downsizing scenario regards the stellar population, while the ΛCDM model predicts the formation and mass assembly of DM haloes.

According to the current standard scenario, the formation and evolution of ETGs can be summarized in two main phases:

- At $2 < z < 6$, ETGs formed with a rapid collapse, with a burst of star formation rapidly quenched.
- At $z < 2$, the initial star formation is followed by a later evolution of accretion of stars formed in other galaxies.

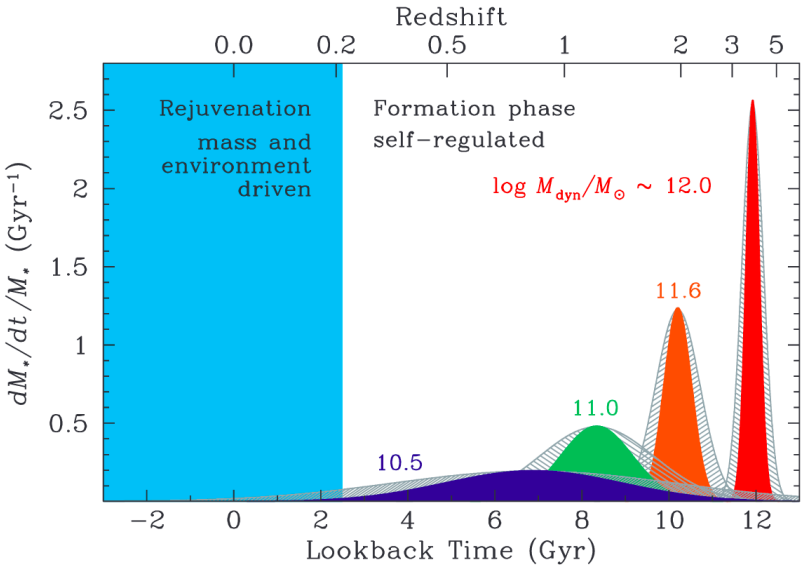

Figure 3.2: Specific star formation rate as a function of lookback time, for ETGs of different masses. The more massive galaxies formed at higher lookback time (higher redshift) with an intense SFR. The thickness of the curve shows the rapidity in the galaxy formation. The plot is divided in two regions: at $z > 0.2$ the environment is not important for the formation of galaxies, instead for lower redshift ($z < 0.2$) the environment affects the galaxy formation. Credits to Thomas et al. [2010].

In this Section, we provided a brief overview on the main physical mechanisms that drive the formation of galaxies. Figure 3.3 summarises the logic flow chart of the various possible channels of galaxy formation. Though real processes do not separate as neatly as this figure suggests, the chart provides anyway a useful summary of the main mechanisms that are currently believed to determine the formation of elliptical and disk galaxies. Figure 3.3.

3.3 The galaxy stellar mass function

The galaxy SMF is defined as the comoving number density of galaxies with stellar mass in the range $[M_*;\ M_* + dM_*]$. This statistical function provides a description of the stellar mass assembly at a given epoch, and plays a fundamental role in the study of galaxy evolution. In fact, the SMF provides information on the link between galaxies and their host DM haloes.

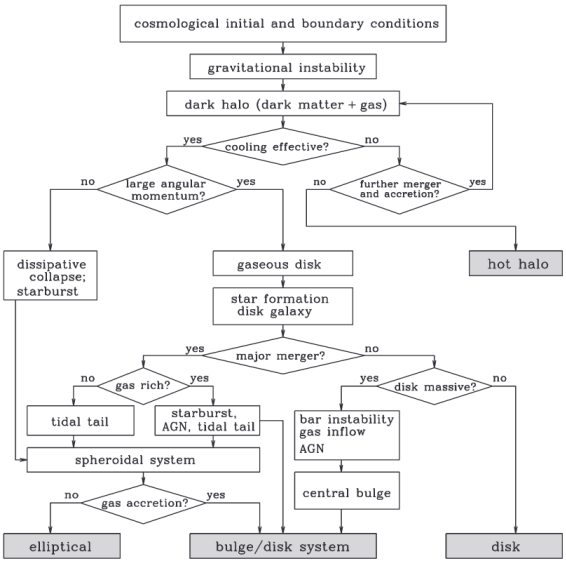

Figure 3.3: Logic flow chart for galaxy formation from Mo et al. [2010]. The initial and boundary conditions for galaxy formation are set by the cosmological framework. The paths bringing to the formation of different types of galaxies are shown.

Moreover, it provides indirect evidences on the main feedback processes, that are key mechanisms in the history of galaxy formation and evolution, as emphasised in the previous Section. Thus the SMF is a key observable, representing a benchmark for galaxy formation and evolution models [e.g. Bolzonella et al., 2010, Guo et al., 2011b, Vulcani et al., 2011, Henriques et al., 2015, Davidzon et al., 2016].

The galaxy SMF is well described by a Schechter function [Panter et al., 2004, Panter et al., 2007, Pozzetti et al., 2010, Davidzon et al., 2017, Moutard et al., 2016], that is a power law with an exponential cut off:

$$\phi(M_*)\mathrm{d}M_* = \frac{\phi_0}{M_0}\left(\frac{M_*}{M_0}\right)^{\alpha} \exp\left[-\frac{M_*}{M_0}\right]\mathrm{d}M_* \quad . \tag{3.3.1}$$

The star formation history (SFH) represents a weighted version of the SMF, as it describes the evolution of the latter as a function of time. The SMF shape is affected by the feedback processes that act on the baryonic matter, modifying the stellar mass growth of the galaxy during its evolution. In particular, the feedback mechanisms lower the star formation efficiency. The strength of this quenching depends on the galaxy stellar mass.

Figure 3.4 compares the DM MF to the galaxy SMF, showing the role of SNe and AGN feedbacks to lower the SMF at both high and low masses. The different slopes of SMF and HMF, both at low masses and at high masses, indicate that the star formation efficiency is constant in mass. At low stellar masses the quenching is thought to be due to SN explosions, which eject gas outside small galaxies (e.g. Dekel and Silk [1986], Benson et al. [2003], Lucas et al. [2020]). On the other hand, the energy injection from AGN can effectively reduce gas cooling in massive galaxies (e.g. Springel et al. [2005], Schawinski et al. [2009]).

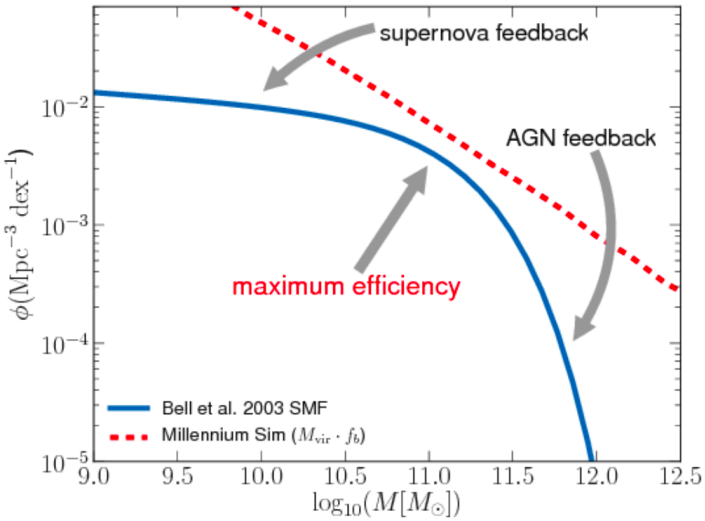

Figure 3.4: A comparison between the observed SMF (blue solid line) (Bell et al. [2003]) and the HMF of the Millennium Simulation (red dashed line). Credits to Mutch et al. [2013].

Both the stellar mass and the star formation efficiency depend on the DM mass. The peak of star formation efficiency corresponds to the value of the galaxy stellar mass (and the corresponding halo mass) where the SMF is closest to the HMF. This happens at the value corresponding to the *knee* of the Schechter function, which is approximately $M_0 \sim 10^{11}\ M_\odot$, at $z \sim 0$.

For these reasons, the redshift evolution of the SMF is fundamental to understand the connection between galaxies and DM haloes. Moreover, to probe the whole galaxy evolution history, it is necessary to combine low and high redshift measurements of the SMF. This task is difficult to access due to the fact that samples of galaxies at different redshifts are built with differ-

ent photometric filters, that causes different selection effects. For instance, the minimum stellar mass measurable increases with redshift, due to flux limit: at increasingly higher redshifts, we lose increasingly fainter galaxies. Nevertheless, reliable SMF measurements are now available up to $z \approx 7.5$ (e.g. Pozzetti et al. [2010], Grazian et al. [2015]and Leja et al. [2019]).

Figure 3.5 shows an example of SMFs measured at different redshifts, compared to the best-fit Shechter function models (Davidzon et al. [2016]). For stellar masses $M_* > 10^{10}\ M_\odot$ to $M_* > 10^{11.5}\ M_\odot$, we can see an increase in the comoving number density of galaxies, which makes the *knee* of the Schechter function sharper as the redshift lowers. Whereas the number density of small mass galaxies increase at all redshifts. This feature can be interpreted by assuming some threshold mass, M^*, that discriminates where the quenching is efficient. Specifically, when galaxies reach M^* the star formation slows down and the galaxies can not easily grow further (Davidzon et al. [2017]). This causes the build-up of galaxies, and increases the galaxy number density around the M^* value. Different theoretical models showed that the mass-dependent quenching can be explained by various internal feedback processes, such as AGN feedback or virial shocks (see e.g. Gabor et al. [2010], Ilbert et al. [2010], Ilbert et al. [2013], Manzoni et al. [2019]).

Several works studied the Stellar to Halo Mass relation (SHMR), which compares the galaxy stellar mass to the host halo mass. This can be done using different approaches, for example with the Sub-Halo Abundance Matching (SHAM) technique (see Chapter 4), which assigns a stellar mass to each halo and subhalo above a certain mass threshold, or using the Conditional Stellar Mass Function (CMF) technique, which provides the join probability to have a stellar mass conditioned by the host halo mass (e.g. Behroozi et al. [2010] Moster et al. [2010], Moster et al. [2013], Wang et al. [2013], Kravtsov et al. [2018], Behroozi et al. [2019], Girelli et al. [2020]). The SHMR can be parameterised as follows:

$$M_* = \frac{2Mk}{\left[\left(\frac{M}{M_1}\right)^{-\beta} + \left(\frac{M}{M_1}\right)^{\gamma}\right]} \quad , \tag{3.3.2}$$

where M indicates either the virial mass of the DM haloes, or the infall mass of the subhaloes (see 2.6), k is the normalisation factor, M_1 represents the characteristic halo mass at which the SHMR changes the slope, and β and γ are slopes at small halo masses and high halo masses, respectively.

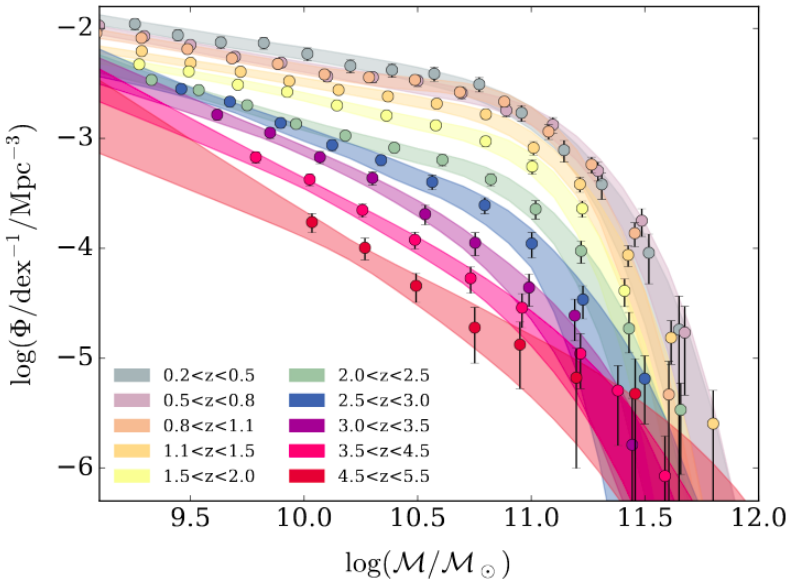

Figure 3.5: Evolution of the SMF between $z = 0.2$ and 5.5 of the COSMOS2015 galaxy sample. Filled circles show the observed data, while coloured areas represent the 1σ uncertainty on the best-fit Schechter function models. Credits to Davidzon et al. [2017].

Figure 3.6 shows the stellar mass to halo mass ratio, comparing different approaches. The peak position indicates the host halo mass where the star formation efficiency is maximum, which comes out to be approximately at $M \approx 10^{11} M_{\odot}$.

3.3.1 Stellar mass measurements

The most common method used to measure the stellar mass of galaxies consists of fitting the observed galaxy Spectral Energy Distribution (SED), which may cover different wavelengths, possibly from the UV to the infrared band in optical surveys.

The galaxy SEDs describe how the radiation energy span over a given wavelength range. SED measurements allow to infer information about the emission processes occurring inside galaxies. The main contribution of light coming from galaxies is generated by stars, and the galaxy SEDs change in time according to the stellar evolution. The theory of stellar evolution is well known, and allows to calculate the light emitted by stars knowing

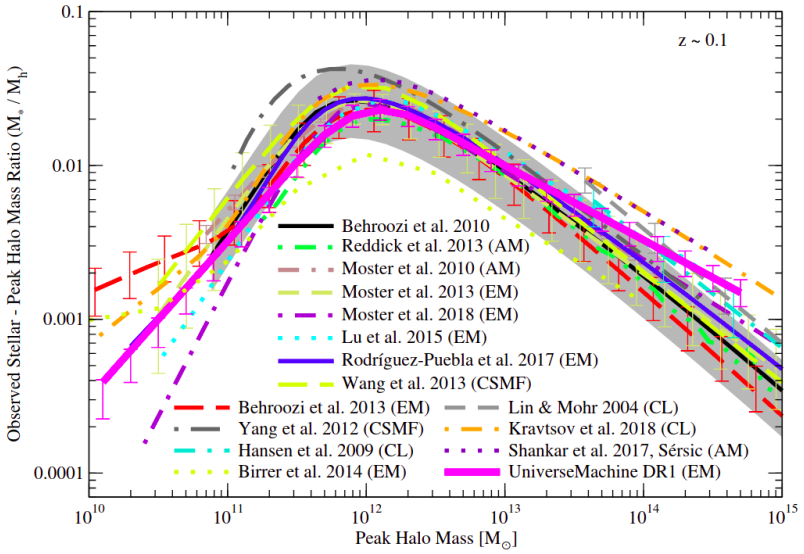

Figure 3.6: The stellar-to-halo mass ratio estimated with different approaches: Empirical Modeling (EM), Abundance Matching (AM), Conditional Stellar Mass Function (CSMF) and from X-ray measurements (CL). Credits to Behroozi et al. [2019].

their density distribution as a function of mass, chemical composition and evolutionary stage. The SEDs trace all the history of star formation and evolution inside galaxies. Hence measuring the light emitted from galaxies and knowing the stellar evolution theory, we can deduce the total galaxy stellar mass.

Galaxies are complex systems composed by multiple stellar populations. Indeed, the stellar population inside a galaxy is not homogeneous and stars present a wide distribution of ages, metallicities and masses. Moreover, the stars inside galaxies are usually not spatially resolved. Galaxies can also contain dust which absorbs light with the shortest wavelengths, reddening the starlight because the extinction is stronger for blue spectra than for red ones. All these aspects make the inference of the galaxy stellar mass a tricky task. The light coming from galaxies is the convolution of the radiation generated by different single stellar population (SSP) which are an ensemble of coeval stars. The SSPs can have different ages, metallicities and masses. In principle, we can generate synthetic spectra changing the SSP parameters and compare the resulting convolved SED with the observed

galaxy SED.

An important assumption in this method is the Initial Mass Function (IMF), which describes the distribution of stellar masses at the formation time of a SSP. A first widely-used functional form for the IMF has been proposed by Salpeter [1955], while Kroupa [2001] and Chabrier [2003] IMFs are also largely used. The IMF has to be inferred from models of star formation of first generation stars, which are not completely understood. For this reason, the IMF represents a great uncertainty in the model. Assuming a specific IMF, we can write the monochromatic integrated spectra of each SSP as follows:

$$F_\lambda(t, Z) = \int_{M_{down}}^{M_{up}} f_\lambda(M, t, Z)\Phi(M)\mathrm{d}M \quad , \tag{3.3.3}$$

where $f_\lambda(M, t, Z)$ is the monochromatic flux of a single star of mass M, age t and metallicity Z, $\Phi(M)\mathrm{d}M$ is the IMF and M_{down} and M_{up} are the smallest and biggest masses of stars of the SSP. Usually $M_{down} \approx 0.1M_\odot$ and $M_{up} \approx 100M_\odot$.

Let us define a Composite Stellar Population (CSP) as the sum of several SSPs, following Tinsley [1980]. The SED can be written as:

$$F_\lambda(t) = \int_0^t \psi(t - t')S_\lambda(t', \zeta(t - t'))\mathrm{d}t' \quad , \tag{3.3.4}$$

where $\psi(t)$ is the star formation rate, $\zeta(t)$ is the function that describes the stellar metal enrichment, while $S_\lambda(t', \zeta(t - t'))$ is the power per unit mass and wavelength emitted by a SSP with age t' and metallicity $\zeta(t - t')$.

We can generate a library of synthetic SEDs from CSP assuming a Stellar Population Synthesis Model (SPSM), which consists of assuming an IMF, *isochrones*, stellar spectra and SFH. A largely used SPSM was the one proposed by Bruzual and Charlot [2003]. Varying for example the redshift, age, metallicity and dust extinction of the CSP, we can obtain a template of simulated SEDs. These simulated SEDs can be compared to the observed ones finding the best correspondence, through e.g. a χ^2 analysis. This procedure is the so-called *SED fitting*. One of the output parameter is the stellar mass to luminosity ratio, M_*/L, from which the stellar mass of the galaxy can be inferred from the observed galaxy luminosity.

However, the SED fitting results might be affected by degeneracies, that is the best-fit model is generally not unique. Other uncertainties on the

stellar mass measurements arise from the signal-to-noise ratio of the observed data, from uncertainties on the IMF and also from uncertainties on the fitting procedure itself. Possible ways to disentangle these degeneracies and reduce the errors on the stellar mass estimates are discussed in Mobasher et al. [2015].

3.4 The two-point correlation function

As discussed in section 2.7, the 2PCF describes the probability of having the overdensity $\delta(\vec{x}+r)$, at a distance r from an other overdensity $\delta(\vec{x})$ as given by Eq.(2.7.1). To measure the 2PCF of a given set of observed sources, it is convinient to write it in a different way with respect to one given in Section 2.7. Let us consider two comoving volumes $\mathrm{d}V_1$ and $\mathrm{d}V_2$. The probability of having one object in each volume can be written as follows:

$$\mathrm{d}P_{12} = \bar{n}^2 \mathrm{d}V_1 \mathrm{d}V_2 [1 + \xi(r_{12})] \quad , \tag{3.4.1}$$

where $\bar{n}$ is the average density of object and r_{12} is the comoving distance between the two volumes. The 2PCF, $\xi(r)$, measures how much the probability of finding an object in the volume $\mathrm{d}V_2$ is influenced by the presence of an other object in the $\mathrm{d}V_1$. In other words, the 2PCF measures the difference between a random distribution and the actual distribution of the objects. In particular, for a random distribution of objects we have $\xi(r_{12}) = 0$. If $\xi(r_{12}) > 0$ the probability of finding two objects separated by r_{12} is higher than in a random distribution, and *viceversa*. From Eq.(3.4.1) we see that $\xi \geq -1$, to have positive probabilities. We can use Eq.(3.4.1) to write the joint probability of having an object in $\mathrm{d}V_2$ knowing that there is another object is $\mathrm{d}V_1$

$$\mathrm{d}P(2|1) = \bar{n} \mathrm{d}V_2 [1 + \xi(r_{12})] \quad . \tag{3.4.2}$$

Integrating Eq.(3.4.2) over the volume, we can estimate the number of objects we expect there are around another object:

$$\big\langle N(< r) \big\rangle = \frac{4\pi}{3} r^3 \bar{n} + 4\pi \int_0^r r'^2 \xi(r') \mathrm{d}r' \quad . \tag{3.4.3}$$

If the distribution was random, only the first term of Eq.(3.4.3) would be not null. The second term takes into account the deviation from a random distribution.

3.4.1 Two-point correlation function Estimators

A common method to measure the 2PCF of discrete population of objects consists in comparing the number of object pairs respect to the number of pars in random distribution of objects with the same boundaries and selection function of the data. With N_D we indicate the number of objects in the real catalogue, measured from observed data, while with N_R we indicate the number of objects present in the random catalogue. The total number of pairs, will be $N_{DD} = N_D(N_D-1)/2$ and $N_{RR} = N_R(N_R-1)/2$ respectively, while the number of cross-pairs between the two catalogues are: $N_{DR} = N_D N_R$. The so-called natural estimator of the 2PCF is the following:

$$1 + \xi(r) \approx \frac{DD(r)}{RR(r)} \quad , \tag{3.4.4}$$

where $DD(r)$ is the number of objects pairs in the pair separation range $(r - \mathrm{d}r, r + \mathrm{d}r)$, while $RR(r)$ is the number of random pairs at the same scales. More accurate estimators, in terms of variance and bias, have been investigated in the literature. The most widely used are the following:

- The Davis and Peebles [1983] estimator:

$$1 + \xi(r) = \frac{N_{RR}}{N_{DD}} \frac{DD(r)}{RR(r)} \quad . \tag{3.4.5}$$

- The Landy and Szalay [1993] (LS) estimator:

$$\xi(r) = 1 + \frac{N_{RR}}{N_{DD}} \frac{DD(r)}{RR(r)} - 2\frac{N_{RR}}{N_{DR}} \frac{DR(r)}{RR(r)} \quad . \tag{3.4.6}$$

The variance of the LS estimator depends on the number of pairs in the catalogue, as shown in Landy and Szalay [1993]. As demonstrated in Kerscher et al. [2000], the LS estimator is an optimal and unbiased estimator of 2PCF. For further details see Keihänen et al. [2019].

3.4.2 The two-point correlation function of galaxies

The 2PCF of galaxies is a fundamental tool to investigate on the Large Scale Structure (LSS) of the Universe and it allows to achieve tight constrains the cosmological parameters. However, as discussed in Chapter 2, galaxies are biased tracers of the underlying DM distribution and this has an impact on the measured 2PCF. The difference between the DM density field and

the galaxy one is described the bias factor defined by Eq. (2.7.4). Thanks to the huge amount of data coming from wide galaxy surveys, it has been largely demonstrated that the galaxy spatial distribution depends on several galaxy properties such as luminosity, colours, stellar mass, morphology and environment (Davis and Geller [1976], Giovanelli et al. [1986], Norberg et al. [2002], Zehavi et al. [2005], Meneux et al. [2008], Ross et al. [2011], Momose et al. [2020]). In general, the galaxy clustering grows with the increasing of luminosity and stellar mass. This fact can be explained by assuming that the most luminous (and massive) galaxies reside into the most massive DM haloes, which have in turn higher bias values with respect to the less massive ones (Kaiser [1987], White and Rees [1978]). It has been observed also that red galaxies are more clustered than blue ones (e.g. Zehavi et al. [2005], Zehavi et al. [2011]). In particular, these trends with stellar mass and luminosity are more evident for galaxies that are more massive than the characteristic mass M_0 in the Schechter SMF (see Eq. 3.3.1). Unlike the 2PCF of DM haloes, the galaxy 2PCF shows a steepening at small scales, $r \leq 1$ h^{-1} Mpc. This is the so-called *1-halo term*, which is caused by the countings of pairs of objects (galaxies) not only on large scales, but also inside the DM haloes. The galaxy 2PCF is affected by non-linear phenomena such as mergers, dynamical friction, cooling and feedback processes concerning the galaxy evolution. To investigate on the impact of these mechanisms on the galaxy bias we can make use of semi-analytical models and full hydrodynamic simulations (which will be described in Chapter 4), or also exploit observational measurements. After the assumption of a cosmological framework, the comparison between the observed and the predicted bias can be used to constrain the parameters of galaxy formation models.

The observed galaxy 2PCF is well represented by a power law on a wide range of scales, from about 10 kpc/h to 100 Mpc/h (Jones et al. [2005]):

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma} , \tag{3.4.7}$$

where r_0 is called *correlation scale length*, while γ is the correlation slope ($\gamma \sim 1.80$ for low-redshift galaxies, Zehavi et al. [2005]). Figure 3.7 shows the observed 2PCF of galaxies for the VIMOS Public Extragalactic Redshift Survey (VIPERS) (see e.g. Marchetti et al. [2013], de la Torre et al. [2013], Marulli et al. [2013]) for increasing cuts in the galaxy stellar mass and in different bins of redshifts. This demonstrates how the bias factor grows with

the stellar mass and with the redshift of the galaxy sample. In Figure 3.8

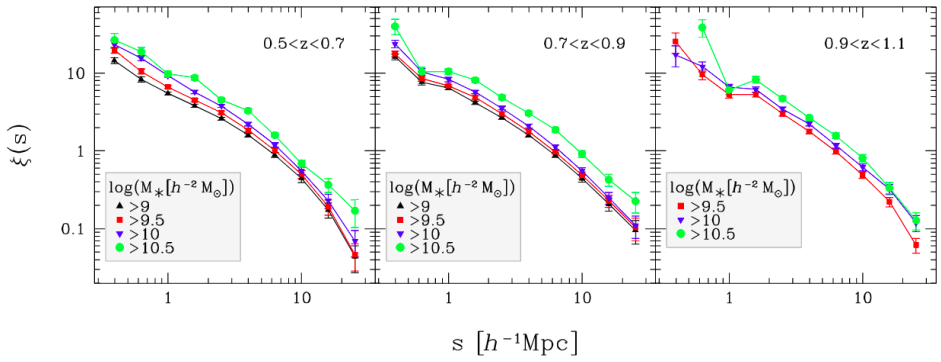

Figure 3.7: 2PCF of galaxies measured from the survey VIPERS with the different stellar mass thresholds, as shown in the label. As expected, the most massive galaxies are also more clustered. This trend holds for different redshift bins. Credits to: Marulli et al. [2013]).

are shown the correlation length, r_0 and γ best fit found in Marulli et al. [2013], as a function of galaxy stellar mass. They are monotonic increasing function that reflect the increasing of bias factor with stellar mass.

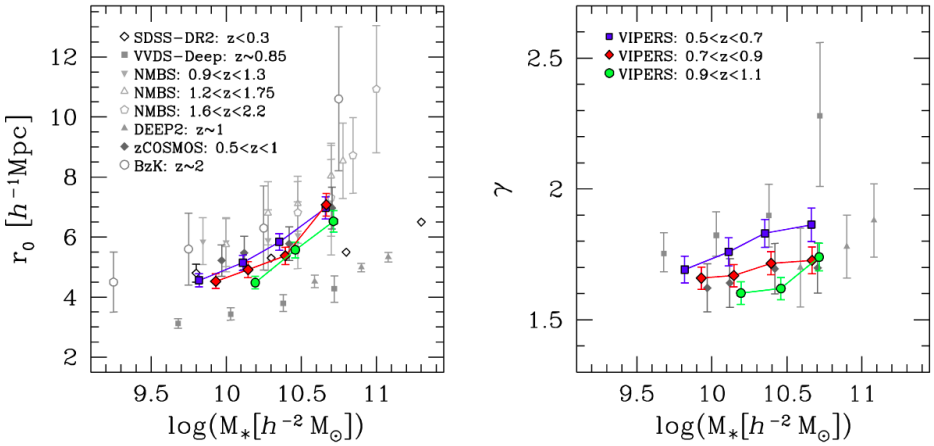

Figure 3.8: *Left panel*: best-fit of the correlation length r_0 as a function of stellar mass. *Right panel*: best-fit values of the slope γ as a function of stellar mass. The constraints obtained from VIPERS are represented with blue squared for $0.5 < z < 0.7$, with red diamonds for $0.7 < z < 0.9$ and with green circles for $0.9 < z < 1.1$. The grey symbols represent the measurements made using other catalogues. Credits to: Marulli et al. [2013]).

Chapter 4

Building simulated galaxy catalogues

As discussed in Chapter 3, the formation and evolution of galaxies are complex and, nowadays, not fully understood phenomena. They involve a variety of non-linear physical processes. For instance, the evolution of baryonic matter depends on complex physical mechanisms such as cooling, shocks, turbulence, re-heating and thermal conduction. For these reasons, it is hopeless to derive a detailed description of these phenomena using fully analytic techniques.

Three main approaches have been proposed to face this problem. The first makes use of numerical algorithms to solve the fully non-linear equations involved in the process of galaxy formation and evolution. N-body and hydrodynamic simulations belong to this class of algorithms. The second approach exploits the so-called semi-analytic models, which make use of both numerical and analytic techniques to approximate the physics involved in these processes, with a degree of approximation that depends on the complexity of the physical phenomenon analysed. The third class of algorithms, to which the Halo Occupation Distribution (HOD) and the Sub-Halo Abundance Matching (SHAM) algorithms belong, aims at reproducing the observed distribution of galaxies within DM haloes, the latter obtained from N-body simulations, using a probabilistic description calibrated on real measurements. In particular, most of these algorithms make use of the relation between the stellar masses of galaxies and the total masses of their hosting DM haloes.

In this Chapter, we will describe these three classes of algorithms, discussing the fundamental techniques that can be used to obtain galaxy mock catalogues. Then, we will introduce the cosmological simulations used as reference in this Thesis work, providing a brief overview on the main features of the catalogues employed.

4.1 Numerical Simulations

The evolution of cosmic structures can be described in the linear regime until $\delta \ll 1$, while the later non-linear evolution cannot be described exactly with analytic models. Therefore, numerical simulations are required in order to investigate the non-linear regime, and model the main properties of the LSS of the Universe and their cosmological evolution. With the increasing power of computational facilities, numerical simulations have grown in size and resolution, allowing to study extensively the formation and evolution of DM haloes and galaxies.

Two main types of numerical simulations have been used in the last few decades to predict the investigate the LSS of the Universe:

- N-body simulations: they are used to simulate the DM distribution through particles from small scales to large scales.

- Hydrodynamic simulations: they involve also hydrodynamics, and so they can be used to describe the baryonic matter density field.

4.1.1 N-body simulations

Specifically, the DM simulations describe the evolution of the non-collisional matter component, which depends only on gravity and on the initial conditions. These simulations are based on a solid theoretical background, relying on the well-known classic gravity theory equations or, alternative, on modified gravity models of cosmological interest. In N-body simulations, the following set of equations are solved to make the system evolve:

$$\begin{cases} F_i = GM_i \sum_{i \neq j} \frac{M_j}{r_{ij}^2} \quad , \\ \frac{\mathrm{d}v_i}{\mathrm{d}t} = \frac{F_i}{M_i} \quad , \\ \frac{\mathrm{d}x_i}{\mathrm{d}t} = v_i \quad . \end{cases}$$

Generally, all particles in N-body simulations have the same mass. At initial conditions, the position and velocity of each particle is known. The code evaluates the gravitational force, F_i, that acts on the i^{th} particle due to the other particles of the system. Then the force is used to evaluate the acceleration, and then the velocity of the i^{th} particle is computed. Finally, from the velocity we can obtain the new position of the i^{th} particle. This operation is repeated on all the other particles of the system a number of times, according to the available computational time. The main differences between these types of algorithms arise in the method used to calculate the force F_i. We can summarise the main strategies proposed in literature as follows:

- **Particle-Particle (PP)** The gravitational force is calculated exactly, so $\vec{F_i} \propto \sum_{i \neq j} \frac{\vec{x_i} - \vec{x_j}}{|\vec{x_i} - \vec{x_j}|^3}$. This provides an accurate description of the evolution of the system, but it is very time expensive ($t_{CPU} \propto N^2$). Moreover, we have $F \to \infty$ when $r \to 0$. Thus, we have to introduce a softening length, ϵ_{soph}, inside which the gravitational force is set to zero.

- **Particle-Mesh (PM)** The volume of the simulation is divided into a regular grid where the field quantities, such as the density or the gravitational potential, are computed. To calculate the gravitational force acting on each particle, an interpolation method is requested. From the density field estimated at each grid cell, the gravitational potential is computed using the Poisson equation (2.1.1) in Fourier space, and used to evaluate the force in real space using the Fast Fourier Transform algorithm. The PM method is less accurate than the PP method, as the cell size reduces the spatial resolution compared with the PP method, but it is much faster ($t_{CPU} \propto N \log N$).

- **Particle-Particle-Particle-Mesh (P3M)** The aim of these algorithms is to combine the accuracy of the PP method to the rapidity of the PM one. To do that, these algorithms exploit the PP method on small scales, while the PM approach is used on large scales where the required accuracy is less. It is important in this case to adopt a coherent linking for the two regimes, choosing an arbitrary critical distance to discriminate where to use the PP method or the PM one.

- **Hierarchical Tree (HT)** For each i^{th} particle, the force is calculated

precisely (PP like) for its neighbour, while distant particles are assembled into the so-called nodes. The force on the i^{th} particle produced by these nodes is calculated using the center of mass of each nodes, instead of all single particles. The nodes that are more distant, from the i^{th} particle, than a given scale, will appear *bigger*, and viceversa. Also in this case, a definition of an arbitrary distance is needed. These algorithms are both accurate and fast ($t_{CPU} \propto N \log N$).

A widely used strategy consists in running large-scale simulations at low resolution, identifying where there are underdense or overdense regions. Then zoomed simulations of the interested regions are run over a smaller volume, but at a higher resolution. Nowadays N-body simulations of a given cosmological model provide consistent results regardless of the adopted approach.

4.1.2 Hydrodynamic simulations

To simulate the evolution of the baryonic matter density field, other physical processes have to be properly modelled, in addition to gravity, like cooling, shocks, turbulence, re-heating, conduction. These phenomena can be included by making use of the exact solution of the set of Eqs. (2.1.1). There are two main approaches to the problem:

- Eulerian methods: fixed or adaptive grids are used to compute mean values of the field quantities. These methods work well to describe strong gradients, and in fact are generally used to describe shocks. However, they suffer in spatial resolution, because of the grid. The introduction of adaptive meshes can overcome the issue.

- Lagrangian methods: the fluid is modelled as an ensemble of particles. The values of the different field quantities, that act on each particle, are calculated as the weighted average of the properties of the neighborhood particles.

The choice of the most efficient method depends on the type of problem to solve. The Eulerian approach ensures a larger resolution in mass, but a lower spatial resolution, with respect to the Lagrangian method. Actually, the most recent cosmological simulations are able to solve the so-called magnetohydrodynamic equations (MHD), which take into account the magnetic and electric fields of the fluids. Eqs (2.1.1) have to be modified, adding

Maxwell's equations which describe the electromagnetic fields, obtaining the following set of magneto-hydrodynamic equations:

$$\begin{cases} \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) = 0 & \text{(continuity equation)} \\ \rho \frac{(\partial \vec{v}}{\partial t} + (\vec{v} \cdot \nabla))\vec{v} = \vec{J} \mathrm{x} \vec{B} - \nabla p - \rho \nabla \phi & \text{(Cauchy momentum equation)} \\ \vec{J} \mathrm{x} \vec{B} = \frac{(\vec{B} \cdot \nabla)\vec{B}}{\mu_0} - \nabla \left(\frac{B^2}{2\mu_0} \right) & \text{(Lorentz force)} \\ \mu_0 \vec{J} = \nabla \mathrm{x} \vec{B} & \text{(Ampere's law)} \\ \frac{\partial \vec{B}}{\partial t} = -\nabla \mathrm{x} \vec{E} & \text{(Faraday's law)} \\ \nabla \cdot \vec{B} = 0 & \text{(magnetic divergence)} \\ \frac{\mathrm{d}}{\mathrm{d}t}(\frac{p}{\rho^\gamma}) = 0 & \text{(Energy equation)} \end{cases} \tag{4.1.1}$$

where p is the fluid pressure, $\vec{B}$ and $\vec{E}$ are the magnetic and electric fields, $\vec{v}$ is the fluid velocity, $\vec{J}$ is the current density, ϕ is the gravitational potential, μ_0 is the magnetic permeability and γ is the specific heat ratio. Finally, the equations of state of the fluids have to be considered, in addition to Eqs. (4.1.1). Usually, cosmic fluids are assumed to be modelled as *perfect* fluids, so their equation of state reads $p = nk_BT$. In the set of Eqs. (4.1.1), we neglected the various dissipative terms, which would have to be considered for a more realistic modelling. This set of equations is very complex and a significant amount of computing time is necessary to solve it in realistic cases. Nevertheless, modern super-computers allow to perform huge and high resolution hydrodynamic simulations. The state-of-art in this field is represented by the Magneticum simulations (Dolag et al. in preparation), and by the TNG-Illustris simulations (Nelson et al. [2019]), which are a set of simulations whith different physical sizes, mass resolution and different complexity of the physical processes employed. As an illustrative case, Fig. 4.1 shows the distribution of gas inside the DM haloes of the TNG-Illustris simulations, which traces the so-called cosmic-web structure. The Figure is colour coded and the transparency indicates the gas density, while different colours indicate the gas temperature. However, even the latest hydrodynamic cosmological simulations do not have sufficient resolution to model, at the same time, the physical processes inside galaxies, such as the formation of stars, and the large-scale phenomena, such as the galaxy mergers. The full description of both large and small scales is nowadays impossible to achieve with a single simulation.

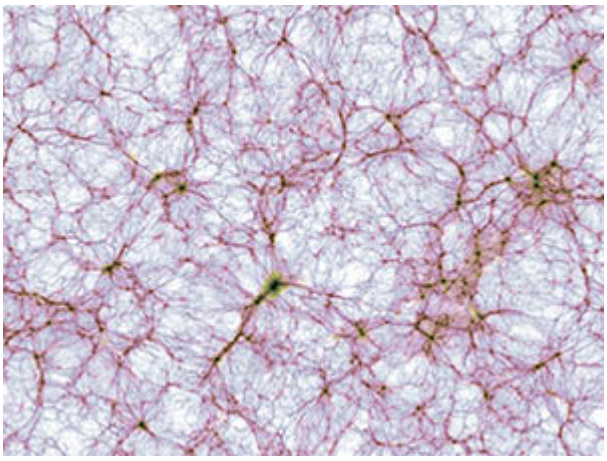

Figure 4.1: Slice of the TNG300 simulation of IllustrisTNG. The cosmic large-scale structure is shown with a projection of the baryonic density field. The image brightness indicates the projected mass density, while the colours quantify the mean projected gas temperature. Credits to https://www.tng-project.org/media/.

4.2 Semi-analytic models

Semi-Analytic Models (SAMs) are used to populate the DM haloes with galaxies. The DM haloes are obtained from DM merger trees, that can be constructed with either the extended Press–Schechter formalism (Press and Schechter [1974], Bond et al. [1991]) or through N-body simulations. These models describe the formation and evolution of galaxies inside DM haloes by exploiting some approximated analytic equations used to describe complex physical phenomena, such as gas-dynamic and radiative processes, or the formation of stars and black holes, as well as the stellar and AGN feedback. These analytic recipes are inferred both from theoretical models and observations. For example, the SFR in the disks of galaxies is observed to depend on the local surface mass density, Σ_g, of gas (Kennicutt-Schmidt law Schmidt [1959]), so that $\dot{M}_* = A\Sigma_g^{\beta}$. The A and β parameters are adjusted comparing model predictions with observations.

The analytic recipes used by SAMs are often too simplified to trace the galaxy evolution in details, and can be used to model only the basic properties of the galaxy population. Nevertheless, comparing the SAM predictions with observed data, it is possible to calibrate the free parameters of these

models, thus improving their accuracy. Furthermore, with this type of approach, the influence of different physical mechanisms can be investigated separately, which can help to isolate the different processes and to improve the accuracy of each semi-analytic recipe.

SAMs are less time consuming with respect to hydrodynamic simulations, but they are also less accurate in the description of the processes related to the baryonic matter. Indeed, the dynamics of this component and its interaction with DM particles are not followed directly. Moreover, given the large freedom in the parameterisation of the physical processes in SAMs, the inaccurate modelling of one of the involved phenomena can lead to compensate a competing process with the risk to obtain the right observable predictions for the the wrong reason.

Nevertheless, current SAMs can reproduce observed galaxy statistical properties quite accurately, making use of large cosmological volumes and spanning over a large range of galaxy masses and redshifts (e.g. Somerville et al. [2008] and Guo et al. [2011a]). For instance, SAMs have been shown to be able to well reproduce statistical properties like the luminosity functions, SMFs and SFRs of massive galaxies ($M_{star} \gtrsim 10^{10} M_{\odot}$) at high redshifts ($z \lesssim 6$) (e.g. Somerville et al. [2012], Fontanot et al. [2009]).

4.3 Halo Occupation Distribution

Among the algorithms of the third class we introduced in this Chapter, the so-called Halo Occupation Distribution (HOD) represents one of the most powerful ones to construct mock galaxy catalogues with given large-scale structure properties. We will provide a more detailed and complete description for this kind of algorithms, since it is the one that will be exploited in this Thesis work. This class of methods is nowadays widely used in the scientific community, and represents a simple and effective empirical model to populate DM haloes with galaxies. This technique is based on the distribution function $P(N|M)$, that is the conditional probability of having N galaxies, above some mass or luminosity threshold, hosted by haloes of mass M. In other words, this function describes how much the galaxies are biased with respect to the DM. The functional form of these probability distributions is theoretically justified by the predictions of hydrodynamic simulations and semi-analytic models (e.g. Zheng et al. [2005]). The func-

tion $P(N|M)$ can be divided into different components:

- The mean occupation number, $\langle N \rangle(M)$, as a function of the halo mass M.
- The distribution of the number of galaxies with respect to a mean occupation value, $P(N|\langle N \rangle)$.
- The spatial and velocity distributions of galaxies into each halo.

The galaxy sample is usually divided into central and satellite galaxies. In the following Sections, we will make use of the subscripts c and s to refer to central and satellite galaxies, respectively. The functional form of $\langle N \rangle(M)$ is different from central and satellite galaxies. The central galaxies reside at the center of their host DM haloes. Indeed, the number of central galaxies per each halo can be either 0 or 1. For this reason the functional form of the mean number of central galaxies is taken as a step function, or as a smoothed step function, in which a softened cut-off in inserted to take into account the scatter in the relation between the galaxy luminosity and the halo mass (e.g. Kravtsov et al. [2004], Zehavi et al. [2005], Zheng et al. [2005] Mor, Zehavi et al. [2011]). So, for central galaxies we can assume the following functional form:

$$\langle N_c(M) \rangle = \begin{cases} 0 \text{ if } M < M_{min} \\ 1 \text{ if } M > M_{min} \end{cases} \tag{4.3.1}$$

The transition from 0 to 1 is smoothed due to the intrinsic scatter between the galaxy properties and the host halo mass. As shown in Kravtsov et al. [2004], a general possible parameterisation is the following:

$$\langle N_c(M) \rangle = \text{erf}\left[5\left(1 - \frac{M}{M_{min}}\right)\right] \tag{4.3.2}$$

where $erf(x) = \frac{2}{\pi}\int_0^x e^{-t^2}\mathrm{d}t$ is the error function and M_{min} is the minimum mass for an halo to host a central galaxy with a luminosity or mass above a certain threshold.

The satellite galaxies are spatially distributed following the spatial distribution of DM inside the haloes. The mean number of satellite galaxies can be approximated by a power-law of the halo mass, as theoretically demonstrated in Kravtsov et al. [2004]. We can write the mean occupation number

for satellites as a power-law:

$$\langle N_s(M)\rangle = \left(\frac{M}{M_1}\right)^{\beta}, \tag{4.3.3}$$

where usually $\beta \approx 1$, and M_1 is the mass of an halo which can host at least two galaxies, a central and a satellite. We can define a parameter, α, which tells us if the probability distribution is well described by a Poissonian distribution or not:

$$\alpha^2 \equiv \frac{\langle N(N-1)\rangle}{\langle N\rangle^2} = 1 - \frac{1}{\langle N\rangle^2}\,. \tag{4.3.4}$$

Eq.(4.3.10) shows that the occupation number of galaxy has a Poisson distribution (α^2=1) at high host halo masses, and drop to zero in the low host mass regime.

Figure 4.2 shows the different contributions to the total mean occupation number of galaxies given by satellite and central galaxies, separately. The mean total number of galaxies as a function of the halo mass is represented with a blue solid line and appears as a complex form, composed by a step function, a shoulder, and an high-mass power-law tail. The mean number of satellite galaxies is shown in pink and follows a simple power-law function. The dotted line shows the mean occupation number of central galaxies and it is basically a smoothed step function, that becomes equal to 1 after a certain value of halo mass, M_{min} (in this case for $M_h \approx 10^{11} M_\odot$), which represents the minimum value of the host halo mass which can contain a central galaxy. M_{min} depends on the galaxy luminosity threshold, which is reflected in the galaxy bias dependence on the galaxy stellar mass. Indeed, the most massive galaxies reside in the most massive haloes, and the halo bias factor increases with the mass, as also shown in Chapter 2. Therefore we expect that also the galaxy bias increases with the minimum halo mass.

Typically M_{min}, α and M_1 are treated as free parameters that are constrained by observations. Since DM haloes contain either zero or one central galaxies, the functional form of $P(N_c|\langle N_c\rangle)$ is a nearest-integer distribution, with $P(1) = 1 - P(0) = \langle N_c\rangle$. The functional form of $P(N_s|\langle N_s\rangle)$ of the satellite galaxies is often assumed to follow a Poisson distribution with mean equal to $\langle N_s\rangle$.

The strongest assumption in the HOD treatment is that all the properties of the observed galaxy population depend only by the mass of the host halo.

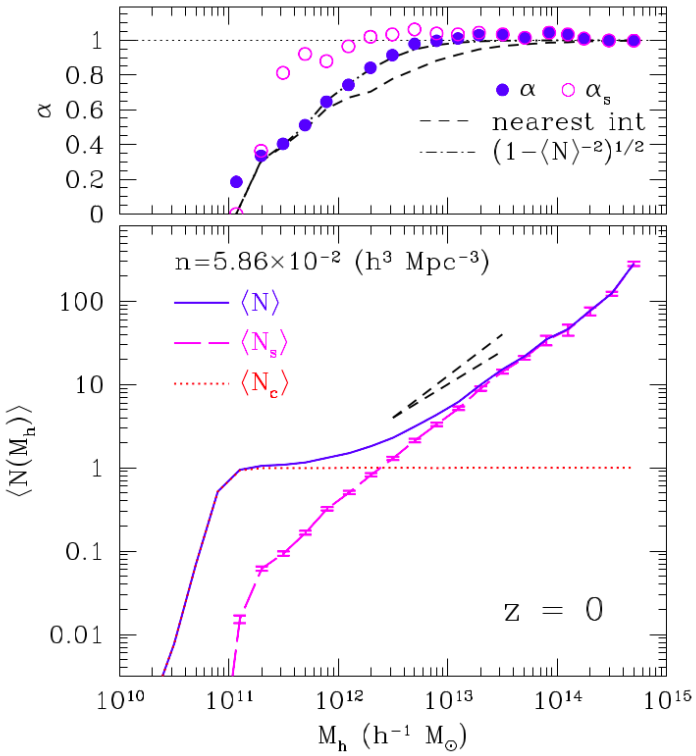

Figure 4.2: *Lower panel*: mean occupation number of galaxies as a function of halo mass. The blue solid line shows the mean total number of galaxis per halo of mass M. The pink long-dashed line represents the mean number of satellites only, while the red dotted line shows the mean number of central galaxies per host halo mass, that is a step function. The error bars show the uncertainty on the mean values. *Upper panel*: the α parameter (see Eq. 4.3.10) for satellite (pink points) and total number of structures (blue points) as a function of the halo mass. For large halo masses, $\alpha \sim 1$ indicates that the distribution of probability is Poissonian, while at small halo masses it is sub-Poissonian. Nevertheless, α_s remains close to one also at relative small masses. Credits to Kravtsov et al. [2004]

.

This rough approximation is one of the main issues related to this method. Simulations showed indeed that also other halo properties, such as halo spin, environment, and previous halo assembly history, could affect the halo clustering. This dependence of the halo clustering on other properties than the halo mass is the so-called *halo assembly bias*(Sheth and Tormen [2004], Gao et al. [2005], Angulo et al. [2008], Lazeyras et al. [2017]). However, it is still not clear if the galaxies properties are strongly influenced by the historical assembly of their host haloes or not. Moreover galaxy properties could depend on the LSS of the Universe, and they can be affected by the so-called *galaxy assembly bias* (Reed et al. [2007], Zu et al. [2008], Zentner et al. [2014], Chaves-Montero et al. [2016], Zehavi et al. [2018]). According to this assumption, the *galaxy assembly bias* can influence the HOD, since these models are parameterised using clustering measurements. It seems also that environment influences the HOD parameters: central galaxies tend to be hosted in smaller haloes when they are located in dense regions, while satellite galaxies show an higher mean occupation number. Zehavi et al. [2018] found also that M_{min} is smaller for haloes with higher formation redshift (older haloes). On the contrary, the satellites tend to be less in early-formed haloes (Artale et al. [2018], Bose et al. [2019]).

4.3.1 Conditional luminosity and mass functions

Closely related to the HOD, the Conditional Luminosity Functions (CLF) ,$\Phi(L|M)$, and Conditional Mass Functions (CMF), $\Phi(m_*|M)$, models provide a fast and powerful tool to populate DM haloes with galaxies. These statistical functions describe the mean number of galaxies with luminosity within $L \pm \mathrm{d}L/2$, or stellar mass within $M_* \pm \mathrm{d}M_*/2$, respectively, that belongs to a DM halo of mass M. Also in this, case it is possible to separate the contribution of CLF and CMF into central and satellite galaxies as follows:

$$\Phi(L|M) = \Phi_c(L|M) + \Phi_s(L|M) \quad . \tag{4.3.5}$$

For central galaxies the functional form of the conditional function is usually assumed as a log-normal distribution. From Yang et al. [2008] we have for central galaxies:

$$\Phi_c(L|M) = \frac{1}{\sqrt{2\pi}\sigma_c} \exp\left[-\frac{\left(\log L - \log \bar{L}_c\right)^2}{2\sigma_c^2} \right] \quad ,$$

where $\bar{L}_c$ is the mean luminosity of the central galaxies hosted by an halo with mass M and σ_c is the scatter in luminosity with respect to the mean value. While for CMF we can follow Moster et al. [2010], using for central galaxies the following equation:

$$\Phi_c(M_*|M) = \frac{1}{\sqrt{2\pi}\ln 10 M_* \sigma_c} \exp\left[-\frac{\log^2\left(M_*/\bar{M}_{*,c}\right)}{2\sigma_c^2}\right] \quad , \tag{4.3.6}$$

where $\bar{M}_{*,c}$ is the mean stellar mass of central galaxies hosted in an halo of mass M, and σ_c is the scatter of stellar masses with respect to the mean stellar mass. Usually, the σ_c value is assumed to be constant. Alternatively, it can be parameterised as a function of halo mass as follow (e.g. Moster et al. [2010]):

$$\sigma_c = \sigma_\infty + \sigma_1\left[1 - \frac{2}{\pi}\arctan\left(\xi \log(M/M_2)\right)\right] \quad , \tag{4.3.7}$$

where σ_∞, σ_1, ξ, M_2 depend on the host halo mass.

The functional form of CLF and CMF for satellite galaxies is usually modelled as a Schechter function, or as a modified Schechter function. Following Yang et al. [2008], we can write $\Phi_s(L|M)$ as follows:

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

and following Moster et al. [2010], we can write the CMF as:

$$\Phi_s(M_*|M) = \frac{\phi_s^*}{\bar{M}_{*,s}}\left(\frac{m_*}{m_{*,s}}\right)^{\alpha_s} \exp\left[-\left(\frac{M_*}{\bar{M}_{*,s}}\right)^2\right] \quad , \tag{4.3.8}$$

Following Moster et al. [2010] we can make further parameterisations

$$\phi_s^* = \Phi_0\left(\frac{M}{M_\odot}\right)^\lambda \quad , \tag{4.3.9}$$

$$\alpha_s = \alpha_\infty + \alpha_1\left[1 - \frac{2}{\pi}\arctan\left(\zeta \log\left(\frac{M}{M_3}\right)\right)\right] \quad , \tag{4.3.10}$$

where ϕ_s^*, L_s^*, α_s, L_s^*, $m_{*,s}$ are free parameters which depend on the halo mass, and can be fitted in order to obtain the observed luminosity function or stellar mass function. Indeed, following Moster et al. [2010] or Wang et al.

[2013], we can parameterise the relation between the mean stellar mass and the halo mass with a double power law:

$$\bar{M}_* = 2M\left(\frac{m_*}{M}\right)_0 \left[\left(\frac{M}{M_1}\right)^{-\beta} + \left(\frac{M}{M_1}\right)^{\gamma}\right]^{-1} \quad , \tag{4.3.11}$$

From the CLF and CMF, we can recover the HOD function. Indeed we can integrate the CLF and CMF, in luminosity and in stellar mass, respectively, as follows:

$$\big\langle N(>L,M)\big\rangle = \int_L^\infty \Phi(L|M)\mathrm{d}L\,,$$

which provides the average number of galaxies above the luminosity L, as function of halo mass. Integrating Eqs. (4.3.6) and (4.3.8), we obtain:

$$\big\langle N_c(>M_{*,min},M)\big\rangle = \frac{1}{2}\left[1-\mathrm{erf}\left(\frac{\log M_{*,min}/\bar{M}_{*,c}}{\sqrt{2}\sigma_c}\right)\right] \quad , \tag{4.3.12}$$

$$\big\langle N_s(>M_{*,min},M)\big\rangle = \frac{\phi_s^*}{2}\Gamma\left[\frac{\alpha_s}{2}+\frac{1}{2},\left(\frac{M_{*,min}}{\bar{M}_{*,s}}\right)^2\right] \quad , \tag{4.3.13}$$

where $\Gamma(a,x)\equiv\int_x^\infty e^{-t}t^{a-1}\mathrm{d}t$ is the incomplete gamma function. The CLF and CMF connect the LF, $\phi(L)$, and the SMF, $\phi(M_*)$ to the HMF, $\frac{\mathrm{d}N}{\mathrm{d}M}$. For instance, integrating in halo mass the product of conditional function times the host HMF, we cab obtain the SMF:

$$\phi(M_*) = \int_0^\infty \Phi(M_*|M)\frac{\mathrm{d}N}{\mathrm{d}M}\mathrm{d}M \quad . \tag{4.3.14}$$

4.3.2 Subhalo Abundance Matching

In this Section we give a brief overview on a useful method that will be also exploited in this Thesis work, the so-called Subhalo Abundance Matching (SHAM). The SHAM is another simple technique to populate DM , or subhaloes, with galaxies. The starting DM halo catalogues are typically obtained with N-body simulations with sufficient mass and spatial resolution to identified also the substructures inside DM haloes. The SHAM method is based on the assumption that there is a monotonic relation between the properties of galaxies and their hosting DM structures, which are haloes or subhaloes. The galaxy is placed at the center-of-mass of the host subhalo, while the galaxy velocity is imposed to be the same of the hosting subhalo velocity.

The simplest application of this method consists in assuming a monotonic relation between the stellar mass (or luminosity) of galaxies and the subhalo mass. In this case, to associate galaxies to DM haloes, we have to require that the cumulative number density of haloes is equal to the cumulative number density of galaxies:

$$n_{gal}(> M_*) = n_h(> M) \quad . \tag{4.3.15}$$

This simple prescription is good enough to construct mock galaxy catalogues with clustering properties consistent with observations in a wide range of redshifts ($0 < z < 5$) (Conroy et al. [2006]), though it is harder to reproduce other galaxy properties, such as star formation rate or metallicity. The SHAM approach can be improved by introducing a scattered relation, instead of using a one-to-one matching in the relation between stellar and subhalo mass (Behroozi et al. [2010], Trujillo-Gomez et al. [2011], Reddick et al. [2013], Zentner et al. [2014]).

Reddick et al. [2013] applied this approach to the SDSS, fitting simultaneously both the measured clustering and the CMF. Moreover, Simha and Cole [2013] exploited this technique to achieve constrains on cosmological parameters, which are in agreement with the ones found from other established methods.

In Chapter 5, we will apply the SHAM approach in combination with the HOD method to match subhaloes and satellite galaxies. Indeed, while the HOD has the advantage of accurately reproducing the spatial distribution properties of a galaxy population, the SHAM can provide the relation between DM haloes and subhaloes with observational properties that would otherwise require a full-hydrodynamical treatment.

4.4 Euclid Flagship mock galaxy catalogue

In this Section, we will describe the mock catalogues exploited in this Thesis work. Since we want to deal with the state-of-art of the HOD mock galaxy catalogues, we make use of one of the most recent Euclid Flagship catalogues (Castander et al. [2020, in preparation]). The Euclid Flagship mock galaxy catalogue is based on a very large simulation of two trillion DM particles, in a periodic box of $L = 3780\,\mathrm{Mpc}\,h^{-1}$ on a side, in a ΛCDM cosmological framework. The simulation contains more than two billion

galaxies distributed over the 3D space that the European Space Agency (ESA) space mission Euclid will survey in the next future. This mock catalogue has been specifically created to simulate the future observations of the Euclid space telescope, and to test the algorithms that will be used for data analysis. The DM particle simulation was developed on the Piz Daint supercomputer, hosted by the Swiss National Supercomputing Centre, by a team of scientists at the University of Zurich led by Joachim Stadel, using the Rockstar halo finder (Behroozi et al. [2012]). The cosmological parameters assumed by Flagship simulation are $\Omega_M = 0.319$, $\Omega_b = 0.049$, $\Omega_\Lambda = 0.681$, $\sigma_8 = 0.83$, $n_s = 0.96$ and $h = 0.67$.

The halo catalogue is build starting from the DM particle distribution and using a "Friends-Of-Friends" algorithm. This type of halo finder defines as haloes the groups that contain all DM particles separated by distances lower than a given linking length, $b\bar{l}$, where $\bar{l}$ is the mean interparticle separation of the DM catalogue, and b is a free parameter of the code. The obtained halo catalogue has been populated with galaxies using an HOD method, similarly to what we have done in this Thesis work, with the goal to reproduce, as much as possible, all the observables relevant for Euclid's main cosmological probes. Specifically, the HOD algorithm has been calibrated exploiting several local observational constraints, using for example the local luminosity function for the faintest galaxies (Blanton et al. [2003] and Blanton et al. [2005]) and the galaxy clustering as a function of luminosity and colour (Zehavi et al. [2011]).

Figure 4.3 shows slices of light cone of the Flagship simulations. The first slice represents the total distribution of mock galaxies inside the light cone, while the second and the third row show the light-cone observable by Euclid with the instruments, VIS, which operate in the optical band and NISP, which detects Near-Infrared light. The expected number of galaxies observable with both instruments decreases with redshift, because of the flux limit. Moreover, observing the H_α line with NISP, only the sources above a certain redshift are visible. This happens because H_α is an optical rest frame line emission, and only objects that are sufficiently far from us to be redshifted in the Near-Infrared band are observable.

In this Thesis work we have used in particular the light-cone of the Flagship catalogue version 1.8.4, which contains 2.6 billions of galaxies, and spans over 5000 deg^2, up to $z = 2.3$. The mock galaxy catalog has been

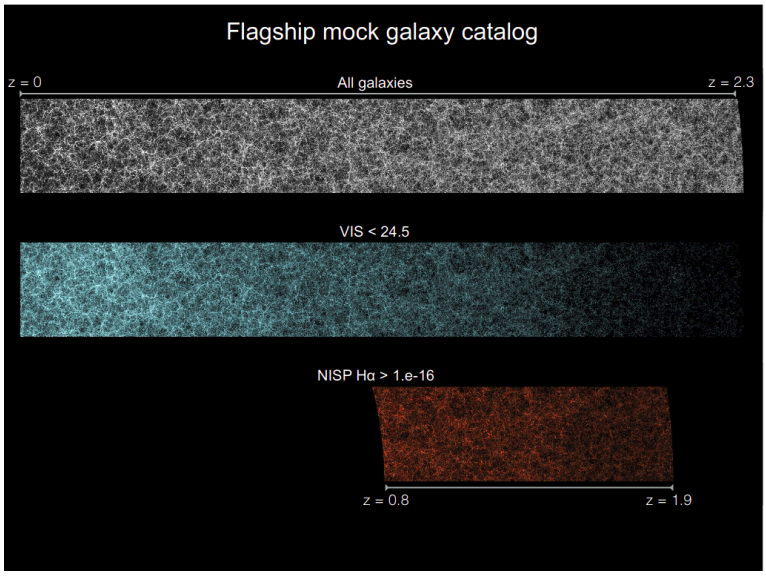

Figure 4.3: The image shows a slice, perpendicular to the observer view, of the simulated light cone of the Flagship mock galaxies. In the top slice, the complete sample of mock galaxies is shown, while the second and the third slices represent the mock galaxy that are expected to be observed by the two instruments VIS (optical) and NISP-Halpha (Near-Infrared), respectively. Credits to: https://www.euclid-ec.org/?page_id=4133.

generated at PIC using the SciPIC pipeline on top of a Big Data platform based on Apache Hadoop (Carretero et al. [2017]). Since we are interested in the local SMF, we selected from the Flagship light-cone all the galaxies and haloes up to redshift $z \sim 0.1$. The selected Flagship galaxy catalogue that will be used to verify our results contains more than 3.7 millions galaxies, while the correspondent Flagship halo catalogue counts about 1 million objects. Figure 4.4 shows the selected galaxy mock light-cone catalogue and its projections onto the three axes, with a colour-code related to the galaxy redshifts.

Figure 4.5 shows the Flagship halo MF compared to the theoretical Tinker et al. [2008] HMF, computed with the CosmoBolognaLib functions (see next Chapter).

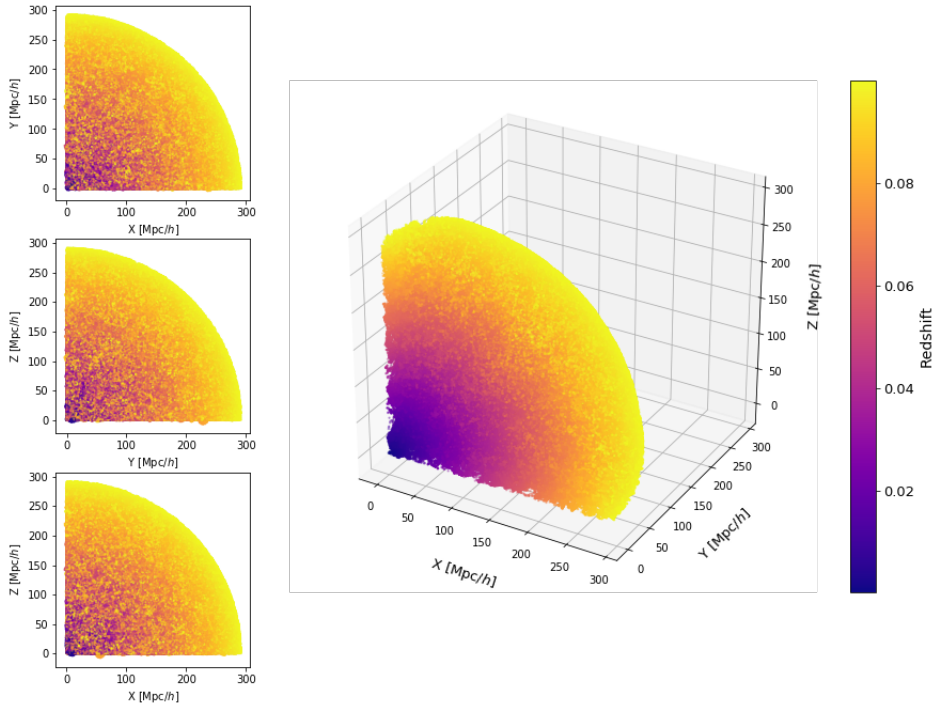

Figure 4.4: 3D representation of the galaxy mock catalogue selected from the Flagship simulations. The observed coordinates have been converted into the correspondent comoving ones, in units of Mpc/h. On the left the 2D projections on the three axes are reported. The colour map indicates the variation of the galaxy redshifts in the range $0 < z < 0.1$.

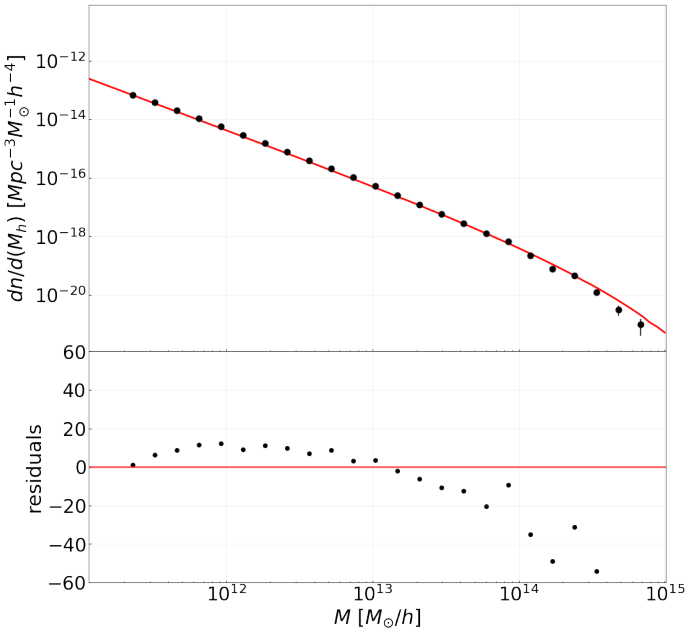

Figure 4.5: *Upper panel*: the MF of the Flagship mock halo catalogue (black points) compared to the theoretical Tinker et al. [2008] HMF. *Lower panel*: residuals computed as the difference between the Flagship and theoretical HMFs, divided by the Poissonian errors associated to the measured MF.

Chapter 5

A new Halo Occupation Distribution code

In the previous Chapters, we have introduced the general context in which this Thesis work is placed. In this Chapter, we will describe the implementation of a new HOD code into CosmoBolognaLib (hereafter CBL) (Marulli et al. [2016]), which is a large set of *free software* C++/Python libraries, that allow to carry out statistical cosmological analyses on catalogues of extra-galactic objects, such as DM haloes, galaxies, galaxy clusters and cosmic voids. The code that we developed for this Thesis work provides an empirical model of galaxy occupation, which is able to reproduce the main observable properties of galaxy catalogues, such as the galaxy stellar mass function and the two-point correlation function. This algorithm has been optimised using *OpenMP Application Program Interface* (OpenMP API) and offers high performances in terms of running time and memory usage. It will be publicly available in the next future[1], along with a full Doxygen documentation and example codes to show how to use it.

5.1 Overview of the code

As described in Chapter 4, there are different ways to construct simulated galaxy catalogues. One of them is to use hydrodynamic simulations, which

[1]The code developed in this Thesis work is currently available in the branch HOD of the DIFA git repository called ClusteringGroup/CosmoBolognaLib/, and will be publicly released in the next version of the CBL at gitlab.com/federicomarulli/CosmoBolognaLib.

start from high redshift initial conditions, and make them evolve using a full hydrodynamical treatment, including all the relevant astrophysical processes. These simulations provide the most complete description of the galaxy properties, but are complex to implement and time consuming. Moreover, they relay on the mostly uncertain assumptions made to model subgrid physical effects.

Another method consists in using SAMs to populate DM haloes, generally identified in large N-body simulations, with galaxies at high redshifts, and then evolve the baryonic component according to analytic prescriptions for all dominant physical phenomena, such as gas cooling, galaxy mergers and feedback effects, following the merging history trees back in time (e.g. Croton et al. [2006], Marulli et al. [2008] and Bonoli et al. [2009]). The main advantage of the semi-analytic approach is that it is computationally cheaper compared to the hydrodynamic simulations, while the primary disadvantage is that it involves a larger degree of approximation, since it makes use of simplified physical recipes.

The third approach described in the previous Chapter is the HOD formalism, which connects galaxies to their hosting DM haloes through the probability distribution, $P(N|M)$, that an halo of mass M contains N galaxies of a particular type (e.g. Berlind and Weinberg [2002], Kravtsov et al. [2004], Zehavi et al. [2005], Zheng et al. [2005] Zehavi et al. [2011] and Zehavi et al. [2018]). Given the complexity of the galaxy formation and evolution phenomena, it is often advantageous to rely on the HOD empirical approach. The HOD method can indeed be useful to relate observations to the assumed underlying physical structure of the Universe, and also to make predictions based on extrapolations from real data catalogues. For all of these reasons, we chose to follow this approach in this Thesis work, implementing new HOD algorithms for painting galaxies into DM haloes. The rationale behind this choice is also that HOD codes are relatively easy to implement and validate, but still provide flexible and powerful tools to construct mock galaxy catalogues with desired large-scale structure properties. Moreover, highly efficient numerical codes can be implemented within the HOD framework (Ronconi et al. [2020]).

The primary aim of this work is to implement a new HOD code within the environment of the CBL[2], to provide an easy and fast tool to construct

[2]The code implemented in this Thesis work is implemented as a new `constructor` of

mock catalogues of galaxies, given an input catalogue of DM haloes. In particular, we make use of the parameterisations suggested by Zehavi et al. [2005] (hereafter *Zehavi05*) and Zehavi et al. [2011] (hereafter *Zehavi11*), which gives the mean number of galaxies above different luminosity thresholds inside an halo of virial mass M. The main equations of standard HOD models have been discussed in Chapter 4. The specific equations of the Zehavi HOD framework for central and satellite galaxies that we implemented in this work are the following:

$$\langle N_c(M)\rangle = \begin{cases} 0 \text{ if } M < M_{min} \\ 1 \text{ if } M > M_{min} \end{cases} \tag{5.1.1}$$

$$\langle N_c(M)\rangle = \frac{1}{2}\left[1 + erf\left(\frac{\log M - \log M_{min}}{\sigma_{logM}}\right)\right] \tag{5.1.2}$$

$$\langle N_s(M)\rangle = \begin{cases} 0 & \text{if } M < M_{min} \\ \left(\frac{M-M_{min}}{M_1}\right)^{\alpha} & \text{if } M > M_{min} \end{cases} \tag{5.1.3}$$

$$\langle N_s(M)\rangle = \frac{1}{2}\left[1 + \left(\frac{M - M_0}{M_1'}\right)^{\alpha}\right] \quad . \tag{5.1.4}$$

where $\langle N(M_{min})\rangle = 0.5$ in Zehavi et al. [2011]. Another important parameter to consider in HOD models is M_1, which is the mass of the host haloes that contain at least one satellite galaxy. This definition implies that $\langle N(M_1)\rangle_{sat} = 1$. In Eq. (5.1.4) the parameter M_1' is defined differently than M_1, but is related to M_1, and also to M_0. The parameter M_0, in satellite galaxies, plays the role of M_{min} for central galaxies. Indeed M_0 indicates the cut-off for the mean number of satellites (Zheng et al. [2005]). For a complete theoretical discussion see Zheng et al. [2005].

We also implemented the HOD model proposed by Moster et al. [2010] (hereafter *Moster10*). The functional form of this HOD model is the following:

$$\langle N_c(> M_{*,min}, M)\rangle = \frac{1}{2}\left[1 - \text{erf}\left(\frac{\log M_{*,min}/\bar{M}_{*,c}}{\sqrt{2}\sigma_c}\right)\right] \quad , \tag{5.1.5}$$

where $erf(x) = \frac{2}{\pi}\int_0^x e^{-t^2}\mathrm{d}t$ is the error function, $\bar{M}_{*,c}$ is defined by Eq. (3.3.2), and represents the mean stellar mass function of the central galaxies hosted by haloes with mass M, while σ_c parameterises the scatter in the

the C++ class cbl::catalogue::Catalogue.

SHMF relation, Eq. (3.3.2). Often, $\sigma_c \sim 0.15$, but it can alternatively be expressed as a function of the host halo mass by Eq. (4.3.7). The value $M_{*,min}$ is the stellar mass threshold and represents the minimum stellar mass of the galaxy sample that we are analysing:

$$\left\langle N_s(> M_{*,min}, M)\right\rangle = \frac{\phi_s^*}{2}\Gamma\left[\frac{\alpha_s}{2}+\frac{1}{2}, \left(\frac{M_{*,min}}{\bar{M}_{*,s}}\right)^2\right], \tag{5.1.6}$$

where $\Gamma(a,x) \equiv \int_x^\infty e^{-t}t^{a-1}\mathrm{d}t$ is again the incomplete gamma function, ϕ^* and α_s are the normalisation and the exponent of the modified Schechter function, defined in Eq. (4.3.8), for the satellite galaxies. The $\bar{M}_{*,s}$ is the mean stellar mass for the satellite galaxy sample as a function of the hosting halo mass. As shown in Chapter 4, from the CMF it is possible to obtain the HOD equations by integrating the CMF in stellar mass (Eq. (5.1.5) and (5.1.6)). Thus we implemented Eqs. from (5.1.1) to (5.1.6) in order to obtain the information on the number of galaxies to assign in each halo. The most powerful model considered in this Thesis work is the one implemented by Moster et al. [2010] because it gives the mean number of galaxies with stellar mass within $M_* \pm \mathrm{d}M_*/2$ inside an halo of virial mass M, and we can choose any stellar mass threshold. On the other hand, the other two models are parameterised with specific luminosity thresholds. In this work we take into account the problems related to the different parameterisations of the models, trying to provide a self-consistent comparison between all of them. We report in Figure 5.1 an example of the mean occupation number per halo for different minimum stellar masses. To obtained these data, we applied the method labelled as *Moster10* in our HOD code. The curves represent the mean number (considering both central and satellite galaxies) of objects inside a DM halo of a specific mass M_h.

Once that the number of galaxies per halo is known, we have to assign the galaxies coordinates. Following the standard approach, we assign to central galaxies the same comoving coordinates of the host halo, while the satellite galaxies are distributed with a numerical density profile which has the same radial behaviour as the NFW profile (Eq. 2.5.1), assuming a spherical symmetry. These are relative simplified assumptions that can be improved in the future, adding for example a small scatter in the central galaxy positions with respect to the location of their host, or assuming a non-spherical symmetry in the distribution of satellites, or adding also a radial distribution of satellites conditioned by the satellite mass. As an

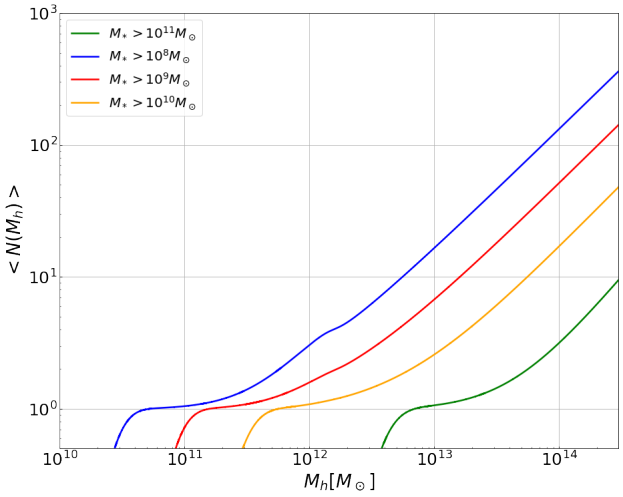

Figure 5.1: The theoretical mean occupation numbers predicted by the HOD model described by Moster et al. [2010]. These curves represent the mean number of galaxies per halo. The four colours refer to different minima in the stellar mass of galaxies. In particular, greater is the stellar mass threshold, greater is the halo mass required to host these galaxies.

illustrative example, Figure 5.2 shows the spatial distribution of galaxies inside haloes with mass $M = 10^{14} M_\odot$ populated using our code.

The stellar masses are extracted from the conditional probability distribution calibrated by Moster et al. [2010] and reported in Eqs. (4.3.6) and (4.3.8), for central and satellite galaxies, respectively. These distributions are used to extract stellar masses for each of the presented models (*Zehavi05*, *Zehavi11* and *Moster10*) exploiting the implemented CBL functions to define the distribution probability used to extract random numbers. Figure 5.3 shows the 3D view of an halo of mass $M = 10^{15} M_\odot$ populated with galaxies of stellar mass greater than $10^9 M_\odot$.

With these HOD prescriptions, we associated a population of galaxies to the DM haloes, according to their mass, and we assigned to each galaxy three comoving coordinates and a stellar mass. As described in Section 2.5, satellite galaxies usually reside in substructures, to which we can associate specific masses. To do this, we use the SHMF described in Eq. (2.6.1) from which we extract substructure masses with the same method adopted for stellar masses. Using the SHAM technique, we assign the most massive substructures to the most massive satellites. The substructures are generated

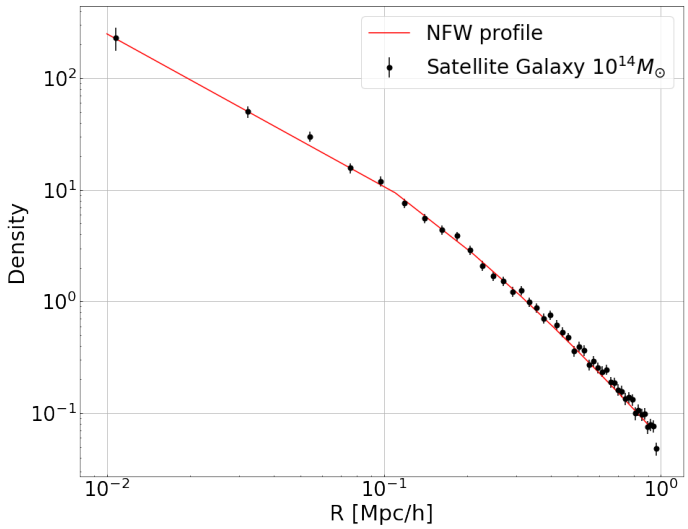

Figure 5.2: Measured number density profile of satellite galaxies generated by our code (black dots), compared to the theoretical NFW profile (red solid line). As expected, the mean density distribution of galaxies follows the NFW profile.

imposing three conditions:

- The masses are extracted from the SHMF.
- The subhalo mass fraction does not exceed the theoretical subhalo mass fraction, f, given by the integral of the SHMF (see Eq. 2.6.5).
- The number of substructures is at least equal to the number of the satellite galaxies predicted by the HOD model.

Once obtained the substructure masses, we apply the Eq. (2.6.2) to connect the subhalo infall mass to its correspondent mass at the present time, as a function of the substructure distance from the centre of its host halo. In Figure 5.4 the solid red line shows the theoretical subhalo mass fraction as a function of the halo mass (obtained from Eq. 5.4), while the blue dots are the subhalo mass fraction computed by applying our HOD code to the halo mock catalogue extracted from the Flagship simulation. The observed subhalo mass fraction is measured as the sum of each subhalo mass inside an halo of mass M, divided by the halo mass.

At the end of the procedure, we obtain a catalogue of galaxies and substructures from any input halo catalogue. The proposed HOD implementa-

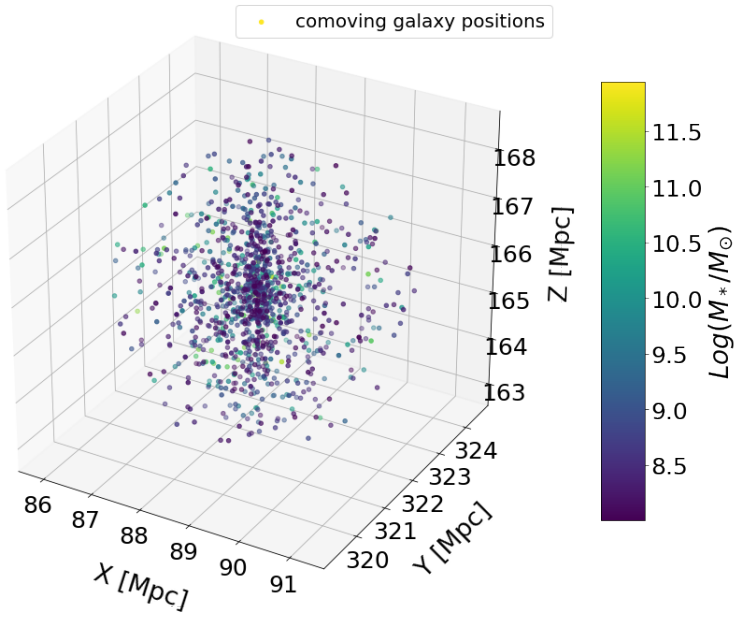

Figure 5.3: 3D representation of galaxies inside a cluster with mass $M = 10^{14} M_{\odot}$. The colormap represents the logarithmic values for the galaxy stellar masses. As expected, the low mass galaxies dominate the sample, and are more numerous near the centre of the host halo.

tion allows the users to set the parameters of the models, or to simply use the default ones that have been calibrated in literature works. The HOD parameters that can be provided in input are the followings:

- `catalogue` is the input DM halo catalogue that will be populated with galaxies[3].
- `cosmology` represents the cosmological model, which is used here to predict the sub-halo mass function[4].
- `HODType` is a `string` which selects the OD model that will be used to populate the haloes with galaxies. The HOD models currently implemented are: *Zehavi*05, *Zehavi*11 and *Moster*10.
- `threshold` is a `double` which fixes the minimum selection threshold in magnitude or stellar mass. Specifically, it is expressed in magnitude, if

[3]This is an object of class `cdl::Catalogue`.
[4]This is an object of the class `cbl::Cosmology`.

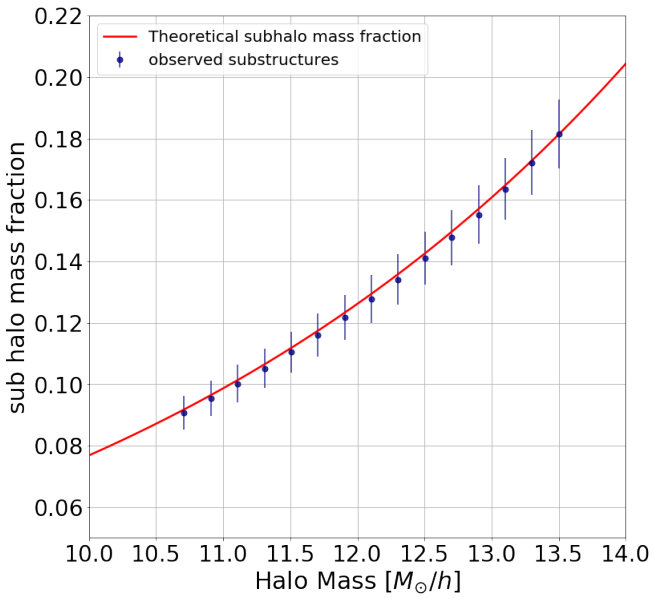

Figure 5.4: The measured subhalo mass fraction (blue dots) obtained from the DM halo catalogue populated by our code, compared to the theoretical subhalo mass fraction (red solid line). The errorbar represent the Poissonian errors associated to our data.

the selected author is *Zehavi*05 or *Zehavi*11, or in solar masses, if the selected author is *Moster*10. The possible values of this threshold with *Zehavi*05 or *Zehavi*11 are: -18, -18.5, -19, -19.5, -20, -20.5, -21., -21.5 or -22 . These values corresponds to magnitude in r-band. Otherwise, any value for the minimum stellar mass of galaxies is allowed with *Moster*10. If the authors is *Zehavi*05 or *Zehavi*11, the minimum stellar mass is automatically set by the code depending on the chosen magnitude limit.

- `substructures` is a `boolean` value. If it is `true` the code will populate the halo catalogue with galaxies and substructure, while if it is set *false* the code returns a catalogue of galaxies only.

- `parameters` is a `vector` of `double` values which contains all the parameters needed by the HOD model. The default parameter values are the ones reported in Zheng et al. [2005], Zehavi et al. [2011] and Moster et al. [2010] for *HODType* *Zehavi*05, *Zehavi*11 and *Moster*10, respectively.

For *Zehavi05* and *Zehavi11* the `parameters` values are divided in two categories: the first set of values is relative to the HOD functions and modifies the mean number of galaxies per halo, while the second one is related to the CMF and changes the stellar mass that is assigned to each galaxy.

Moster et al. [2010] proposed two different HOD parameterisations: the first neglects the intrinsic scatter between halo mass and stellar mass, while the second one takes into account the scatter in SHMR (Eq. 4.3.11). We tested both of them, and eventually we chose to rely on the latter, since it reproduces the SMF more accurately.

Each parameter of the implemented HOD algorithm can be set independently. This feature allows to include easily our HOD algorithm in a Markov Chain Monte Carlo (MCMC) analysis, to fit one or more free parameters by comparing the predicted galaxy properties with observations.

Tables 5.1 and 5.2 report the *Zehavi*05 and *Zehavi*11 default HOD parameter values for each luminosity threshold that refer to Eqs. (5.1.1), (5.1.2), (5.1.4), (5.1.3). Table 5.3 reports the *Moster*10 default parameters used for both the HOD and the CMF. They are relative to Eqs. (4.3.6), (4.3.8), (5.1.5) and (5.1.6).

M_r	$\log M_{min}$	$\log M_1$	α
-22.0	13.91	14.92	1.43
-21.5	13.27	14.60	1.94
-21.0	12.72	14.09	1.39
-20.5	12.30	13.67	1.29
-20.0	12.01	13.42	1.16
-19.5	11.76	13.15	1.13
-19.0	11.59	12.94	1.08
-18.5	11.44	12.77	1.01
-18.0	11.27	12.57	0.92

Table 5.1: List of default parameters of the *Zehavi05* HOD model, for different thresholds in r-band magnitude, M_r: $\log M_{min}$ represents the minimum mass for haloes to host a central galaxy, $\log M_1$ is the value of the minimum halo mass for hosting at least one satellite galaxy, and α is the slope of power-law used to extract the mean occupation function of satellite galaxies (Eq. 5.1.3).

The HOD code implemented in this Thesis work provides high computational performances in constructing large mock galaxy catalogues, with a running time linearly dependent on the number of CPUs employed. The computational time increases drastically when the code is used also to populate the haloes with DM sub-structures. Indeed, this task implies a high number of random extractions to satisfy the three conditions described above, which are necessary to accurately reproduce the SHMF.

5.2 Results

This Thesis work has been focused on the implementation and application of a new HOD code, with the aim of constructing samples of simulated galaxies starting from whatever DM halo catalogues. The primary requirement is that the SMF of the mock galaxies has to be in agreement with observations. On the other hand, within the HOD framework considered in this work, the clustering of galaxies is a prediction of the model, that should be compared with observations to assess the reliability of the constructed mock catalogues. To validate the HOD algorithm implemented in

M_r	$\log M_{min}$	$\sigma_{\log M}$	$\log M_0$	$\log M_1^{'}$	α
-22.0	14.06 ± 0.06	0.71 ± 0.07	13.72 ± 0.53	14.80 ± 0.08	1.35 ± 0.49
-21.5	13.38 ± 0.07	0.69 ± 0.08	13.35 ± 0.21	$14.20 \pm 0,07$	1.09 ± 0.17
-21.0	12.78 ± 0.1	0.68 ± 0.15	12.71 ± 0.26	13.76 ± 0.05	1.15 ± 0.06
-20.5	12.14 ± 0.03	0.17 ± 0.15	11.62 ± 0.72	$13.43 \pm 0,04$	1.15 ± 0.03
-20.0	11.83 ± 0.03	0.25 ± 0.11	12.35 ± 0.24	12.98 ± 0.07	1.00 ± 0.05
-19.5	11.57 ± 0.04	0.17 ± 0.13	12.33 ± 0.17	12.75 ± 0.07	0.99 ± 0.04
-19.0	11.45 ± 0.04	0.19 ± 0.13	9.77 ± 1.41	12.63 ± 0.04	1.02 ± 0.02
-18.5	11.33 ± 0.07	0.26 ± 0.21	8.99 ± 1.33	12.50 ± 0.04	1.02 ± 0.03
-18.0	11.18 ± 0.04	0.19 ± 0.17	9.81 ± 0.62	12.42 ± 0.05	1.04 ± 0.04

Table 5.2: List of default parameters of the *Zehavi11* model, for different thresholds in r-band magnitude, M_r. $\log M_{min}$ is the minimum mass for haloes to host a central galaxy, while $\sigma_{\log M}$ corresponds to the scatter applied to the smoothed step-function used for the mean occupation number of central galaxies (see Eq. 5.1.2). $\log M_0$ and $\log M_1^{'}$ are the logarithmic values of the parameters relative to satellite galaxies, and α is the slope of the power-law used to obtain the mean occupation number of satellite galaxies.

Parameter	value
$\log M_{1c}$	11.9008 ± 0.0119
$(m_c/M)_0$	0.0297 ± 0.00004
β_c	1.07557 ± 0.0097
γ_c	0.6310 ± 0.0121
$\log M_2$	11.8045 ± 0.0458
σ_∞	0.1592 ± 0.0030
σ_1	0.0460 ± 0.0029
ξ	4.2503 ± 0.9945
$\log M_{1s}$	12.0640 ± 0.0931
$(m_s/M)_0$	0.0198 ± 0.0015
β_s	0.8097 ± 0.971
γ_s	0.6910 ± 0.0390
$-\log \Phi_0$	10.8924 ± 0.4615
λ	0.8032 ± 0.0367
$\log M_3$	12.3646 ± 0.0260
$-\alpha_\infty$	1.3676 ± 0.0043
$-\alpha_1$	0.0524 ± 0.0051
ζ	9.5727 ± 6.8240

Table 5.3: List of default parameters of the *Moster10* model, relative to to Eqs. (4.3.6) and (4.3.8). $(m_c/M)_0$ and $(m_s/M)_0$ are the normalisation of SHMR (see Eq. (4.3.11)) for central and satellite galaxies, respectively. β_c, β_s, γ_c and γ_s are the slopes in SHMR (see Eq. 4.3.11) for central and satellite galaxies. β controls the slope at small stellar masses, while γ is the SHMR slope at large stellar masses. The $logM_{1c}$ and $logM_{1s}$ parameters are the logarithmic values of the characteristic mass at which the SHMR changes slopes (see Figure 3.6). The values σ_∞, σ_1, M_2 and ξ are referred to the parameterisation made by Moster et al. [2010] that describes the scatter, σ_c, of the log-normal function, Eq. (4.3.6). The remaining parameters are referred to the further parameterisation made by Moster et al. [2010], for ϕ^* and α_s (see Eqs. (4.3.9) and (4.3.10))

this work, we used it to populate the Flagship DM halo catalogue described in Chapter 4. To facilitate the comprehension of the reader we will make use, in the following Sections, of some acronyms to identify the catalogues and the models to which we want to refer. The first that we introduce is "HOD-Euclid", to indicate the galaxy catalogue extracted from the Euclid Flagship simulations.

As a first test, we compare the mean occupation number per halo of our mock catalogues to the Moster et al. [2010] results. Figure 5.5 shows how the number of simulated galaxies is distributed as a function of the host halo and for different selections in the minimum stellar mass. The measured counts are in agreement with the theoretical curves predicted by the model.

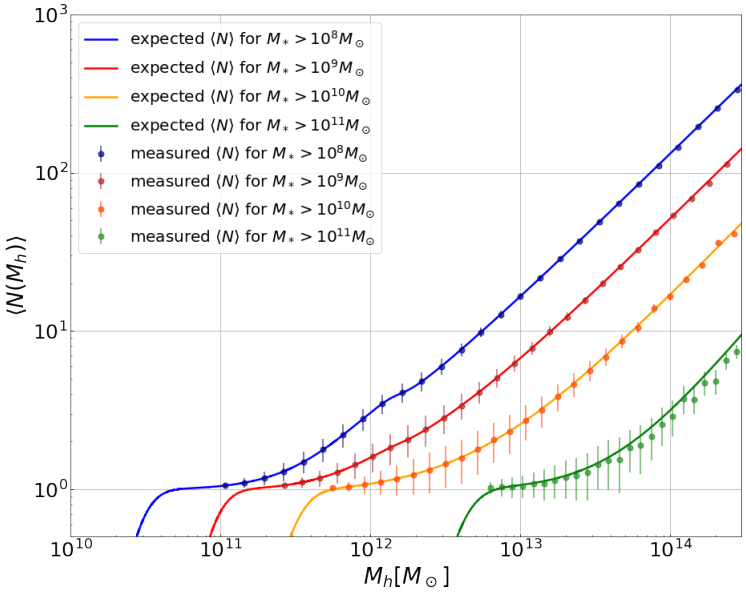

Figure 5.5: Comparison between the measured mean number of galaxies per halo and the theoretical trends by Moster et al. [2010]. The errorbars represent the Poissonian uncertainties related to the counts. The agreement between the data and the model holds for the four different stellar mass selections applied. threshold

The next and most important test is to check if the SMF of our mock galaxies is compatible to the observed one. Specifically, we assess the reliability of our HOD algorithm comparing the predicted SMF to the SDSS and the HOD-Euclid SMFs. Panter et al. [2007] measured the SMF from

the Sloan Digital Sky Survey (SDSS) DR3 (see e.g. Gil-Marín et al. [2015b], Alam et al. [2015] and Gil-Marín et al. [2015a]) and performed a fit to these data modelling the SMF as a Schechter function:

$$\phi(M_*)\mathrm{d}M_* = \frac{\phi_0}{M_0}\left(\frac{M_*}{M_0}\right)^{\alpha}\exp\left[-\frac{M_*}{M_0}\right]\mathrm{d}M_* \quad , \tag{5.2.1}$$

where $\phi_0 = (2.2 \pm 0.5) \times 10^{-3}$ Mpc^{-3}, $M_0 = 1.005 \pm 0.004 \times 10^{11} M_\odot$, $\alpha = -1.222 \pm 0.002$.

From now on we will use the abbreviation "SDSS" to refer to the best-fit to the measured SMF performed by Panter et al. [2007]. Fig. 5.6 shows the comparison between the SMF measured from the Flagship galaxy mock catalogue and the best-fit SMF model performed by Panter et al. [2007] on the SDSS data, which is assumed as the reference fiducial model. In the bottom panel we report the residuals, computed as the difference between the two SMFs, divided by the Poissonian errors associated to the Flagship measurements. The agreement between the two SMFs is good, except for the overabundance of galaxies at stellar masses in the range $10^{10} M_\odot \lesssim M_* \lesssim 2 \cdot 10^{11} M_\odot$.

Now we apply the three HOD models implemented in this work (*Zehavi05*, *Zehavi11* and *Moster10*) to the HOD-Euclid catalogue and we compare the predicted SMF to the SDSS one. We will use the acronyms "HOD-Z05", "HOD-Z11" and "HOD-M10" to denote the mock galaxy catalogues generated by applying our HOD code, using respectively the HOD models given the three by the three authors. The result is shown in Figure 5.7. All models can accurately reproduce the observed SMF, except in the range of stellar masses between $2 \cdot 10^{10} M_\odot$ and $2 \cdot 10^{11}$, as can be clearly seen in the residuals shown in the lower panel. Nevertheless, this mismatch is similar to the one shown in Fig. 5.6, thus it is related to the DM halo catalogue of the Flagship simulation, rather than on our HOD implementation. We can also note an overestimation at high stellar masses, $M_* > 10^{12} M_\odot$, that however appears to be not significant, considering the large Poissonian errors in this mass range. As expected, the HOD-Z05 and HOD-Z11 SMFs are almost indistinguishable, since the two models differs only by the function employed to extract the central galaxies (see Section 4.3).

For the SMF obtained using the *Moster*10 model, we fixed the stellar mass threshold to $M_* = 10^9$, while for both the *Zehavi*05 and *Zhevi*11 models the minimum stellar mass is $1.9 \cdot 10^9 M_\odot$. This is due to the fact that

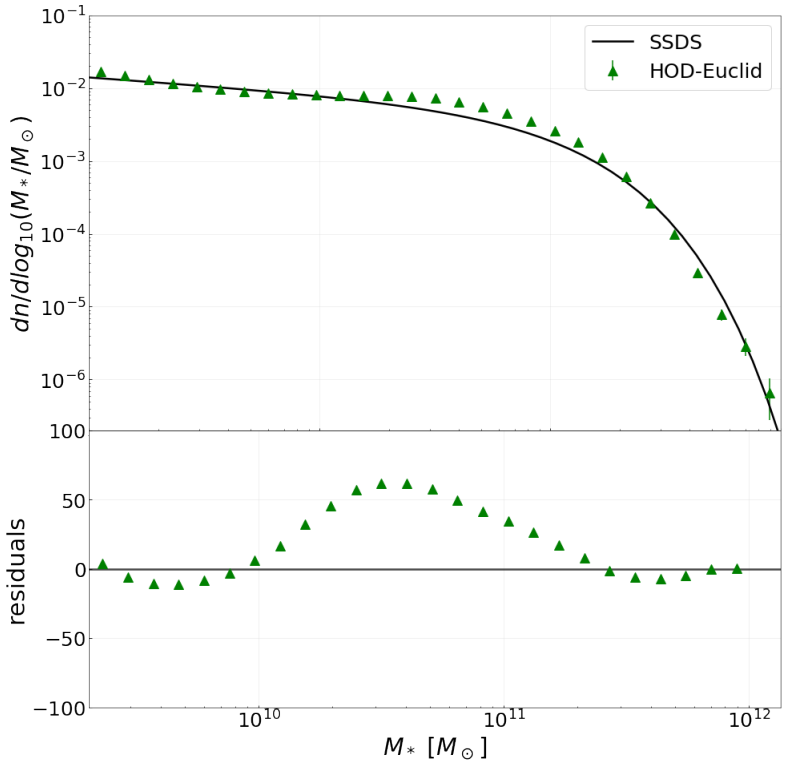

Figure 5.6: *Upper panel*: comparison between the best-fit model performed by Panter et al. [2007] on the SDSS data (black line) and the SMF of the Flagship mock galaxy catalogue (green triangles). *Lower panel*: residuals computed as the difference between the Panter et al. [2007] SMF and the Flagship SMF, divided by the Poissonian errors of the latter.

our code automatically calculates the minimum stellar mass of galaxies on the basis of the luminosity threshold, in order to make all the three models compatible. In fact, while the parameters related to the *Moster10* model are relative to stellar mass thresholds, the parameters of the *Zehavi05* and *Zehavi11* models are related to thresholds expressed in absolute magnitude. Therefore we need to define a conversion function to convert the minimum galaxy luminosity to the minimum stellar mass. To do this, we first convert the absolute magnitude, which is an input values for our constructor, into a luminosity. Then we fix a specific stellar mass to light ratio, M_*/L, to obtain the correspondent stellar mass. The conversion function adopted is probably a too simplified model and leads to an overestimation of the minimum stellar mass. A more accurate conversion between the luminosity

and stellar mass threshold is required and will be implemented in the future.

Although all the models are able to reproduce quite well the observed SMF, we decided to rely on the *Moster10* model, since it allows to reach lower stellar mass values. Moreover, with the *Moster10* model we avoid to use the conversion between the absolute magnitudes and stellar masses (required for *Zehavi05* and *Zehavi11*), that can constitute a source of uncertainty. Finally, Moster et al. [2010] calibrated the HOD parameters exactly on the Panter et al. [2007] measured SMF, thus this HOD parameterisation allows us to make a fairer comparison with the SMF that we assume as reference. Adopting the *Moster10* HOD model, we construct a galaxy catalogue that contains a number of object close to the one of the Flagship galaxy catalogue. The computational time needed to populate the Flagship halo catalogue with galaxies with a stellar mass $M_* > 10^8 M_\odot$ is less than 10 minutes. This time further decreases if the stellar mass threshold is set to higher values.

Figure 5.8 reports the final comparison between the SMF computed from HOD-M10 and the two models that we took as reference: the fit performed by Panter et al. [2007] and the measured SMF from HOD-Euclid. As we can see from the residuals reported in the lower panel, the SMF of HOD-M10 is in good agreement with both the reference SMF. Moreover, the SMF of our galaxies is in better agreement with the SDSS SMF than the one of the galaxies constructed with the Flagship HOD, especially in the range of stellar masses between $2 \cdot 10^{10} M_\odot$ and $2 \cdot 10^{11}$.

To perform a further independent test on the goodness of our code, we investigate the 2PCF computed for HOD-M10. As explained in Section 3.4.1, to measure the 2PCF a random catalogue with the same geometry of the data catalogue is required. In particular, we exploited the CBL functions to build a random catalogue shuffling the observed coordinates of the Flagship lightcone, and obtaining a catalogue with a number of random objects equal to twice the number of objects of the original one. We measure the 2PCF from 0.1 h^{-1} Mpc to 20 h^{-1} Mpc using the LS estimator described in Eq. (3.4.6). In Figure 5.9 we compare the galaxy clustering for different stellar mass selections, obtained using the Flagship simulations, with 2PCF of the DM field estimated using the Code for Anisotropies in the Microwave Background (CAMB, http://camb.info). As discussed in Section

3.4.2, we expect an increasing in the bias factor (therefore also in the amplitude of the 2PCF), with the growing of the minimum of the stellar mass. We note that the galaxies of HOD-Euclid have small bias factors, which is greater than unity only at large radii for galaxies with stellar mass major than $10^{10} M_{\odot}$. Then we compute the 2PCF using the mock galaxy catalogue generated by populating the Flagship DM haloes. Figure 5.10 shows our results. The bias factor is greater than one at radii $r > 2\ h^{-1}$ Mpc for each mass threshold, while for smaller radii the bias factor becomes less than unity. Also our galaxy 2PCF shows an increasing trend as a function of the stellar mass threshold. However, the variation with the stellar mass selections in the bias factor is clearly less important than in Flagship HOD galaxies.

Finally, we compare directly the 2PCF predicted by our HOD implementation and the one of the original Flagship galaxy catalogue (HOD-Euclid). The result is shown in Figure 5.11: the solid lines represent the 2PCF of HOD-Euclid, while the dots are our measurements. The four different colours indicate the different stellar mass selections applied. The mock galaxies generated with our code are more clustered at all scales. Moreover, the slope of the 2PCF at small scales, 1 h^{-1} Mpc, is significantly different with respect to the one of the Flagship galaxy catalogue.

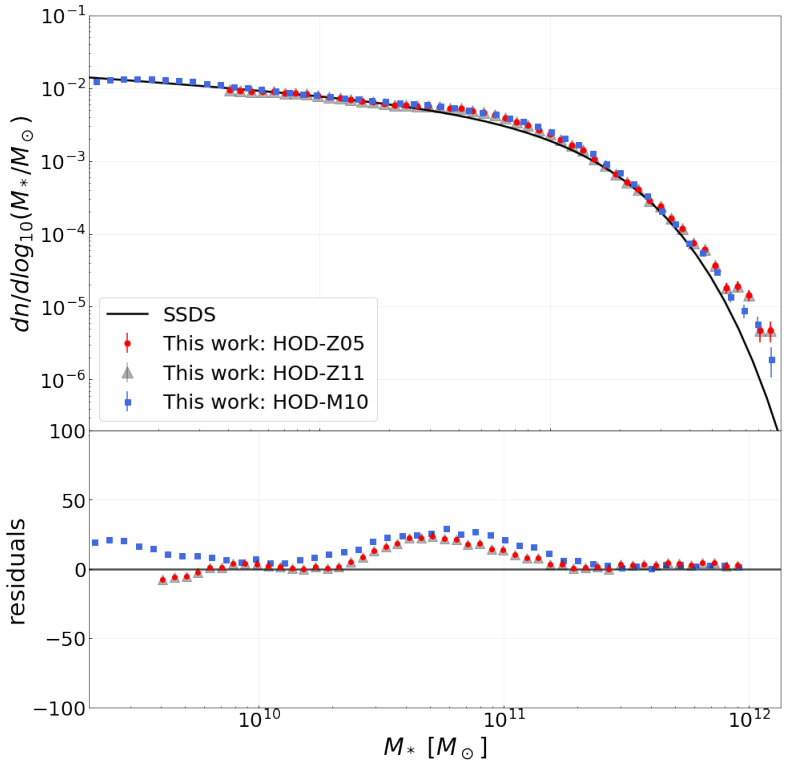

Figure 5.7: Comparison between the measured SMF obtained by applying our code to the haloes of Flagship simulations and the fit to the SMF performed by Panter et al. [2007], that we assume as the reference model (black line). *Upper panel*: the results obtained with the three HOD models implemented in our code: *Zehavi05*, *Zehavi11* and *Moster10*. The first is represented with red dots (HOD-Z05), the second with grey triangles (HOD-Z11) and the last with blue squared markers (HOD-M10). The errorbars represent the Poisson errors. The black line shows the Panter et al. [2007] result (SDSS). *Lower panel*: residuals computed as the difference between our SMFs and the Panter et al. [2007] SMF, divided by the Poisson uncertainties.

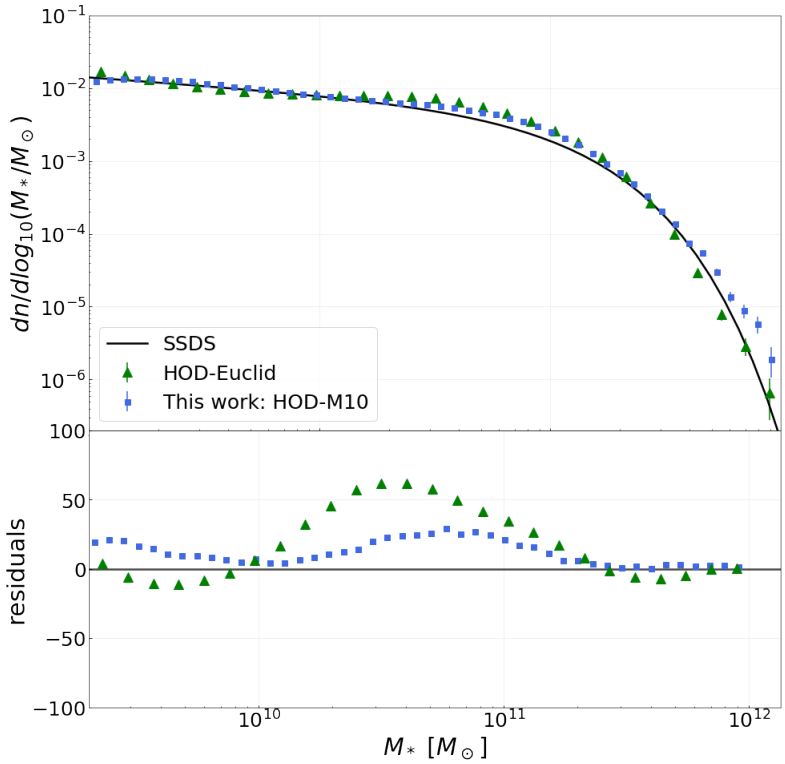

Figure 5.8: Comparison between the fit to the SDSS SMF, the SMF measured from HOD-Euclid and the SMF measured for HOD-M10. *Upper panel*: the black solid line represents the fit computed by Panter et al. [2007] (SDSS), the green triangles are the SMF measured using HOD-Euclid and the blue squared markers show the HOD-M10 one. The errorbars represent the Poissonian errors associated to the data. *Lower panel*: residuals computed as the difference between the SDSS SMF and the mock SMF data, divided by the uncertainties associated to the data.

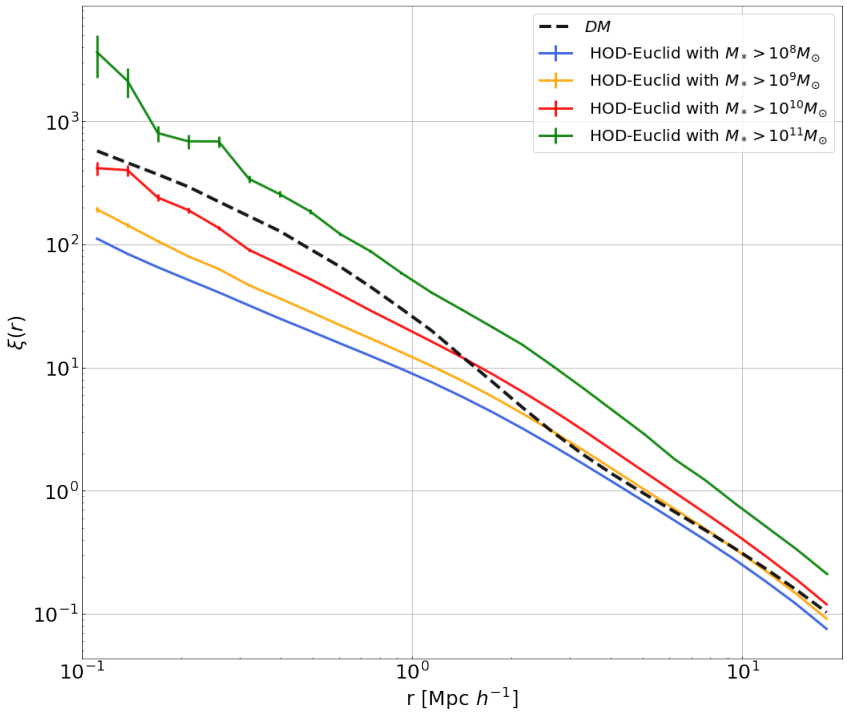

Figure 5.9: Comparison between the 2PCF measured from HOD-Euclid and the DM 2PCF calculated theoretically using CAMB. The latter is represented by a black dashed line, while the measured 2PCF is shown with coloured solid lines. The four colours indicate the different stellar mass thresholds applied. As expected, the 2PCF increases with the minimum stellar mass.

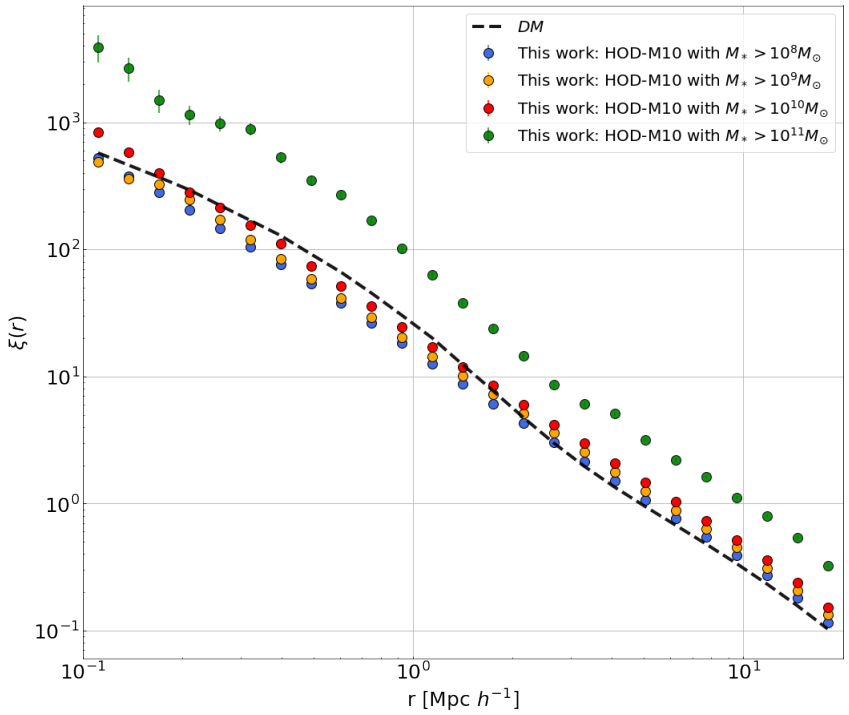

Figure 5.10: Comparison between the DM 2PCF (black line) computed with CAMB, and the 2PCF measured from HOD-M10 (colored points) at different stellar mass thresholds. It is evident an increasing in the amplitude of 2PCF with stellar mass, especially for $M_* > 10^{11} M_\odot$.

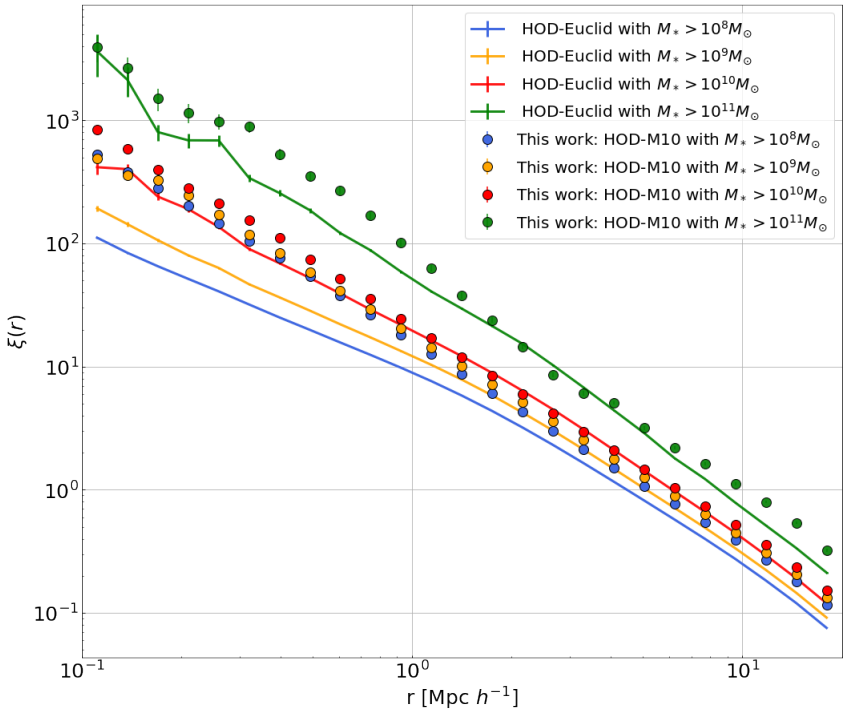

Figure 5.11: Comparison between the 2PCF obtained from HOD-Euclid and our populated galaxy catalogue (HOD-M10). We plot the HOD-Euclid 2PCF with solid coloured lines and with coloured dots the HOD-M10 2PCF, for different stellar mass thresholds. Our galaxies appear more clustered, especially at scales $r < 1/h^{-1}$ Mpc.

Chapter 6

Conclusions and future prospectives

In this work we presented a new HOD code, implemented in the environment given by the CBL, a large set of public C++/Python libraries. This HOD code allows to construct a mock galaxy catalogue starting from whatever DM halo catalogue the user chooses as input. In particular, it makes use of the HOD approach to populate DM haloes with galaxies, that aim at reproducing some observed galaxy properties, such as the spatial distribution inside DM haloes, the observed SMF and the galaxy clustering. Thanks to the exploitation of the openMP parallelisation, the code offers high performances and low running times. The major strengths of our code are the flexibility and the simplicity of his usage. Indeed, it is equipped with a detailed documentation and an explanatory example, and the full set of parameters related to the available HOD models can be set by the user. For instance, our code can be easily used to construct different mock galaxy catalogues, changing the constructor parameters and selecting the best combination that reproduces whatever observed SMF. More specifically, the code allows to select between three different HOD models, based on the parameterisations proposed by Zehavi et al. [2005], Zehavi et al. [2011] and Moster et al. [2010], and adopts the following prescriptions:

- The galaxy spatial distribution at large scales has to follow the spatial distribution of DM haloes, while inside each halo the spatial distribution follows the NFW profile.
- The number of galaxies inside each DM halo and their stellar masses

are assigned in such a way to reproduce the observed SMF.

- The substructures associated to the satellite galaxies have masses extracted from the theoretical SHMF and they reproduce the subhalo mass fraction found by Giocoli et al. [2010]

- To each substructure is also associated an infall mass, that is the mass of the substructure at the infall time, following the results of Gao et al. [2004].

In this work we applied our HOD code to the DM haloes of the Euclid Flagship simulations, selected at $z < 0.1$. To verify the reliability of our code, we compared the SMF measured from our populated galaxy catalogue (HOD-Z05, HOD-Z11 and HOD-M10) with both an observed and a simulated one. In particular, we took as reference the best-fit model provided by Panter et al. [2007] (SDSS), which measured the SMF from the SDSS, and the SMF computed using the mock galaxies extracted from the Euclid Flagship (HOD-Euclid). The SMF obtained for HOD-Z05, HOD-Z11 and HOD-M10 showed a good agreement both with the SDSS and HOD-Euclid ones. Nevertheless, we chose eventually to rely on the model provided by Moster et al. [2010], since it offers a better stellar mass resolution.

Finally, we measured the galaxy 2PCF from the HOD-M10 catalogue, and compared it to the one of the HOD-Euclid galaxies, applying different stellar mass selections. We verified that the bias of mock galaxies increases with stellar mass, as expected. Though the clustering of our mock galaxies follows the expected trend, it is not consistent with the one measured from HOD-Euclid galaxy catalogues. This issue deserves further investigations, that we postpone for a future dedicated analysis.

6.1 Future prospective

We envisage some possible short-term projects achievable for the extension of this Thesis work. Firstly, we plan to perform a deeper investigations on the discrepancies found during the comparison between both the SMF and 2PCF measured of our simulated galaxy catalogues (HOD-Z05, HOD-Z11 and HOD-M10) and our reference models (SDSS and HOD-Euclid). In particular, the results relative to the galaxy clustering are still preliminary and require further tests. Another possible task that we did not fulfill in this

Thesis is the employment of a Bayesian statistical MCMC analysis, to obtain the best-fit values of the HOD parameters needed to reproduce a specific observable. Furthermore, this work represents a first step towards a more complete implementation of an HOD code and is therefore prone to a large series of upgrades. First of all, our code is now designed to work only in the local Universe ($z \sim 0$). A new parameterisation applicable also at higher redshifts will make our code a more powerful cosmological tool. This redshift dependence can be achieved considering the redshift evolution of the CMF, following Moster et al. [2010] or Wang et al. [2013] and implementing the evolution of the parameter that governs the SHMR (Eq. 4.3.11). Another improvement that can be made to our code is to add more prescriptions and models to reproduce different galaxy properties, such as luminosity, colour and spectral type. The correct representation of the galaxy luminosity, for instance, is a goal that can be reached using the CLF reported in Eq. (4.3.5), that have been largely calibrated with different observations and measurements in the literature. Also the performances of the code can be improved in terms of computational time. For example, the Message Passing Interface (MPI) can be exploited to further optimise it. Then, as we described in Chapter 5, our code is very fast in populating DM haloes with galaxies, but the computational time increases if the substructures are generated with the task that allows to obtain the correct subhalo mass fraction. A faster procedure to compute the specific mass for each substructure, taking into account the predicted correct subhalo mass fraction, is thus required. In the end, our code can be also run to populate DM haloes generated in simulations with non-standard cosmologies, to verify if the variation of the HOD parameters allows to mimic the predictions of different cosmological models. Indeed, by comparing the HOD parameters obtained to reproduce a certain galaxy catalogue, we can identify possible degeneracies between the HOD and the cosmological parameters.

Bibliography

S. Alam, F. D. Albareti, C. Allende Prieto, F. Anders, S. F. Anderson, T. Anderton, B. H. Andrews, E. Armengaud, É. Aubourg, S. Bailey, S. Basu, J. E. Bautista, R. L. Beaton, T. C. Beers, C. F. Bender, A. A. Berlind, F. Beutler, V. Bhardwaj, J. C. Bird, D. Bizyaev, C. H. Blake, M. R. Blanton, M. Blomqvist, J. J. Bochanski, A. S. Bolton, J. Bovy, A. Shelden Bradley, W. N. Brandt, D. E. Brauer, J. Brinkmann, P. J. Brown, J. R. Brownstein, A. Burden, E. Burtin, N. G. Busca, Z. Cai, D. Capozzi, A. Carnero Rosell, M. A. Carr, R. Carrera, K. C. Chambers, W. J. Chaplin, Y.-C. Chen, C. Chiappini, S. D. Chojnowski, C.-H. Chuang, N. Clerc, J. Comparat, K. Covey, R. A. C. Croft, A. J. Cuesta, K. Cunha, L. N. da Costa, N. Da Rio, J. R. A. Davenport, K. S. Dawson, N. De Lee, T. Delubac, R. Deshpande, S. Dhital, L. Dutra-Ferreira, T. Dwelly, A. Ealet, G. L. Ebelke, E. M. Edmondson, D. J. Eisenstein, T. Ellsworth, Y. Elsworth, C. R. Epstein, M. Eracleous, S. Escoffier, M. Esposito, M. L. Evans, X. Fan, E. Fernández-Alvar, D. Feuillet, N. Filiz Ak, H. Finley, A. Finoguenov, K. Flaherty, S. W. Fleming, A. Font-Ribera, J. Foster, P. M. Frinchaboy, J. G. Galbraith-Frew, R. A. García, D. A. García-Hernández, A. E. García Pérez, P. Gaulme, J. Ge, R. Génova-Santos, A. Georgakakis, L. Ghezzi, B. A. Gillespie, L. Girardi, D. Goddard, S. G. A. Gontcho, J. I. González Hernández, E. K. Grebel, P. J. Green, J. N. Grieb, N. Grieves, J. E. Gunn, H. Guo, P. Harding, S. Hasselquist, S. L. Hawley, M. Hayden, F. R. Hearty, S. Hekker, S. Ho, D. W. Hogg, K. Holley-Bockelmann, J. A. Holtzman, K. Honscheid, D. Huber, J. Huehnerhoff, I. I. Ivans, L. Jiang, J. A. Johnson, K. Kinemuchi, D. Kirkby, F. Kitaura, M. A. Klaene, G. R. Knapp, J.-P. Kneib, X. P. Koenig, C. R. Lam, T.-W. Lan, D. Lang, P. Laurent, J.-

M. Le Goff, A. Leauthaud, K.-G. Lee, Y. S. Lee, T. C. Licquia, J. Liu, D. C. Long, M. López-Corredoira, D. Lorenzo-Oliveira, S. Lucatello, B. Lundgren, R. H. Lupton, I. Mack, Claude E., S. Mahadevan, M. A. G. Maia, S. R. Majewski, E. Malanushenko, V. Malanushenko, A. Manchado, M. Manera, Q. Mao, C. Maraston, R. C. Marchwinski, D. Margala, S. L. Martell, M. Martig, K. L. Masters, S. Mathur, C. K. McBride, P. M. McGehee, I. D. McGreer, R. G. McMahon, B. Ménard, M.-L. Menzel, A. Merloni, S. Mészáros, A. A. Miller, J. Miralda-Escudé, H. Miyatake, A. D. Montero-Dorta, S. More, E. Morganson, X. Morice-Atkinson, H. L. Morrison, B. Mosser, D. Muna, A. D. Myers, K. Nand ra, J. A. Newman, M. Neyrinck, D. C. Nguyen, R. C. Nichol, D. L. Nidever, P. Noterdaeme, S. E. Nuza, J. E. O'Connell, R. W. O'Connell, R. O'Connell, R. L. C. Ogando, M. D. Olmstead, A. E. Oravetz, D. J. Oravetz, K. Osumi, R. Owen, D. L. Padgett, N. Padmanabhan, M. Paegert, N. Palanque-Delabrouille, K. Pan, J. K. Parejko, I. Pâris, C. Park, P. Pattarakijwanich, M. Pellejero-Ibanez, J. Pepper, W. J. Percival, I. Pérez-Fournon, I. rez-Ra'fols, P. Petitjean, M. M. Pieri, M. H. Pinsonneault, G. F. Porto de Mello, F. Prada, A. Prakash, A. M. Price-Whelan, P. Protopapas, M. J. Raddick, M. Rahman, B. A. Reid, J. Rich, H.-W. Rix, A. C. Robin, C. M. Rockosi, T. S. Rodrigues, S. Rodríguez-Torres, N. A. Roe, A. J. Ross, N. P. Ross, G. Rossi, J. J. Ruan, J. A. Rubiño-Martín, E. S. Rykoff, S. Salazar-Albornoz, M. Salvato, L. Samushia, A. G. Sánchez, B. Santiago, C. Sayres, R. P. Schiavon, D. J. Schlegel, S. J. Schmidt, D. P. Schneider, M. Schultheis, A. D. Schwope, C. G. Scóccola, C. Scott, K. Sellgren, H.-J. Seo, A. Serenelli, N. Shane, Y. Shen, M. Shetrone, Y. Shu, V. Silva Aguirre, T. Sivarani, M. F. Skrutskie, A. Slosar, V. V. Smith, F. Sobreira, D. Souto, K. G. Stassun, M. Steinmetz, D. Stello, M. A. Strauss, A. Streblyanska, N. Suzuki, M. E. C. Swanson, J. C. Tan, J. Tayar, R. C. Terrien, A. R. Thakar, D. Thomas, N. Thomas, B. A. Thompson, J. L. Tinker, R. Tojeiro, N. W. Troup, M. Vargas-Magaña, J. A. Vazquez, L. Verde, M. Viel, N. P. Vogt, D. A. Wake, J. Wang, B. A. Weaver, D. H. Weinberg, B. J. Weiner, M. White, J. C. Wilson, J. P. Wisniewski, W. M. Wood-Vasey, C. Ye'che, D. G. York, N. L. Zakamska, O. Zamora, G. Zasowski, I. Zehavi, G.-B. Zhao, Z. Zheng, X. Zhou, Z. Zhou, H. Zou, and G. Zhu. The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III. , 219(1):12, July 2015. doi:

10.1088/0067-0049/219/1/12.

R. Angulo, C. Baugh, and C. Lacey. The assembly bias of dark matter haloes to higher orders. *Monthly Notices of the Royal Astronomical Society*, 387 (2):921–932, 2008.

M. C. Artale, I. Zehavi, S. Contreras, and P. Norberg. The impact of assembly bias on the halo occupation in hydrodynamical simulations. , 480 (3):3978–3992, Nov. 2018. doi: 10.1093/mnras/sty2110.

M. L. Balogh, I. K. Baldry, R. Nichol, C. Miller, R. Bower, and K. Glazebrook. The Bimodal Galaxy Color Distribution: Dependence on Luminosity and Environment. , 615(2):L101–L104, Nov 2004. doi: 10.1086/426079.

P. Behroozi, R. H. Wechsler, A. P. Hearin, and C. Conroy. Universemachine: The correlation between galaxy growth and dark matter halo assembly from z = 010. *Monthly Notices of the Royal Astronomical Society*, 488 (3):3143–3194, May 2019. ISSN 1365-2966. doi: 10.1093/mnras/stz1182. URL http://dx.doi.org/10.1093/mnras/stz1182.

P. S. Behroozi, C. Conroy, and R. H. Wechsler. A comprehensive analysis of uncertainties affecting the stellar mass-halo mass relation for 0 ¡z¡ 4. *The Astrophysical Journal*, 717(1):379–403, Jun 2010. ISSN 1538-4357. doi: 10.1088/0004-637x/717/1/379. URL http://dx.doi.org/10.1088/0004-637X/717/1/379.

P. S. Behroozi, R. H. Wechsler, and H.-Y. Wu. The rockstar phase-space temporal halo finder and the velocity offsets of cluster cores. *The Astrophysical Journal*, 762(2):109, Dec 2012. ISSN 1538-4357. doi: 10.1088/0004-637x/762/2/109. URL http://dx.doi.org/10.1088/0004-637X/762/2/109.

E. F. Bell, D. H. McIntosh, N. Katz, and M. D. Weinberg. The Optical and Near-Infrared Properties of Galaxies. I. Luminosity and Stellar Mass Functions. , 149(2):289–312, Dec 2003. doi: 10.1086/378847.

E. F. Bell, C. Wolf, K. Meisenheimer, H. Rix, A. Borch, S. Dye, M. Kleinheinrich, L. Wisotzki, and D. H. McIntosh. Nearly 5000 distant early-type galaxies in combo-17: A red sequence and its evolution sincez1. *The Astrophysical Journal*, 608(2):752–767, Jun 2004. ISSN 1538-4357. doi: 10.1086/420778. URL http://dx.doi.org/10.1086/420778.

E. F. Bell, C. Papovich, C. Wolf, E. Le Floc'h, J. A. R. Caldwell, M. Barden, E. Egami, D. H. McIntosh, K. Meisenheimer, P. G. Pérez-González, G. H. Rieke, M. J. Rieke, J. R. Rigby, and H.-W. Rix. Toward an Understanding of the Rapid Decline of the Cosmic Star Formation Rate. , 625(1):23–36, May 2005. doi: 10.1086/429552.

A. J. Benson, R. G. Bower, C. S. Frenk, C. G. Lacey, C. M. Baugh, and S. Cole. What shapes the luminosity function of galaxies? *The Astrophysical Journal*, 599(1):38–49, Dec 2003. ISSN 1538-4357. doi: 10.1086/379160. URL http://dx.doi.org/10.1086/379160.

A. A. Berlind and D. H. Weinberg. The Halo Occupation Distribution: Toward an Empirical Determination of the Relation between Galaxies and Mass. , 575(2):587–616, Aug 2002. doi: 10.1086/341469.

M. R. Blanton, D. W. Hogg, N. A. Bahcall, J. Brinkmann, M. Britton, A. J. Connolly, I. Csabai, M. Fukugita, J. Loveday, A. Meiksin, and et al. The galaxy luminosity function and luminosity density at redshiftz= 0.1. *The Astrophysical Journal*, 592(2):819–838, Aug 2003. ISSN 1538-4357. doi: 10.1086/375776. URL http://dx.doi.org/10.1086/375776.

M. R. Blanton, R. H. Lupton, D. J. Schlegel, M. A. Strauss, J. Brinkmann, M. Fukugita, and J. Loveday. The Properties and Luminosity Function of Extremely Low Luminosity Galaxies. , 631(1):208–230, Sep 2005. doi: 10.1086/431416.

M. Bolzonella, K. Kovač, L. Pozzetti, E. Zucca, O. Cucciati, S. J. Lilly, Y. Peng, A. Iovino, G. Zamorani, D. Vergani, L. A. M. Tasca, F. Lamareille, P. Oesch, K. Caputi, P. Kampczyk, S. Bardelli, C. Maier, U. Abbas, C. Knobel, M. Scodeggio, C. M. Carollo, T. Contini, J. P. Kneib, O. Le Fèvre, V. Mainieri, A. Renzini, A. Bongiorno, G. Coppa, S. de la Torre, L. de Ravel, P. Franzetti, B. Garilli, J. F. Le Borgne, V. Le Brun, M. Mignoli, R. Pelló, E. Perez-Montero, E. Ricciardelli, J. D. Silverman, M. Tanaka, L. Tresse, D. Bottini, A. Cappi, P. Cassata, A. Cimatti, L. Guzzo, A. M. Koekemoer, A. Leauthaud, D. Maccagni, C. Marinoni, H. J. McCracken, P. Memeo, B. Meneux, C. Porciani, R. Scaramella, H. Aussel, P. Capak, C. Halliday, O. Ilbert, J. Kartaltepe, M. Salvato, D. Sanders, C. Scarlata, N. Scoville, Y. Taniguchi, and D. Thompson. Tracking the impact of environment on the galaxy stellar mass function

up to z ~1 in the 10 k zCOSMOS sample. , 524:A76, Dec 2010. doi: 10.1051/0004-6361/200912801.

J. R. Bond, S. Cole, G. Efstathiou, and N. Kaiser. Excursion set mass functions for hierarchical Gaussian fluctuations. , 379:440–460, Oct. 1991. doi: 10.1086/170520.

S. Bonoli, F. Marulli, V. Springel, S. D. M. White, E. Branchini, and L. Moscardini. Modelling the cosmological co-evolution of supermassive black holes and galaxies - II. The clustering of quasars and their dark environment. , 396(1):423–438, June 2009. doi: 10.1111/j.1365-2966.2009.14701.x.

S. Bose, D. J. Eisenstein, L. Hernquist, A. Pillepich, D. Nelson, F. Marinacci, V. Springel, and M. Vogelsberger. Revealing the galaxy-halo connection in IllustrisTNG. , 490(4):5693–5711, Dec. 2019. doi: 10.1093/mnras/stz2546.

G. Bruzual and S. Charlot. Stellar population synthesis at the resolution of 2003. , 344(4):1000–1028, Oct 2003. doi: 10.1046/j.1365-8711.2003.06897.x.

J. S. Bullock, T. S. Kolatt, Y. Sigad, R. S. Somerville, A. V. Kravtsov, A. A. Klypin, J. R. Primack, and A. Dekel. Profiles of dark haloes: evolution, scatter and environment. , 321(3):559–575, Mar 2001. doi: 10.1046/j.1365-8711.2001.04068.x.

J. Carretero et al. CosmoHub and SciPIC: Massive cosmological data analysis, distribution and generation using a Big Data platform. *PoS*, EPS-HEP2017:488, 2017. doi: 10.22323/1.314.0488.

Castander et al. . 2020, in preparation.

D. Ceverino, A. Klypin, E. S. Klimek, S. Trujillo-Gomez, C. W. Churchill, J. Primack, and A. Dekel. Radiative feedback and the low efficiency of galaxy formation in low-mass haloes at high redshift. *Monthly Notices of the Royal Astronomical Society*, 442(2):1545–1559, 06 2014. ISSN 0035-8711. doi: 10.1093/mnras/stu956. URL https://doi.org/10.1093/mnras/stu956.

G. Chabrier. Galactic Stellar and Substellar Initial Mass Function. , 115 (809):763–795, Jul 2003. doi: 10.1086/376392.

J. Chaves-Montero, R. E. Angulo, J. Schaye, M. Schaller, R. A. Crain, M. Furlong, and T. Theuns. Subhalo abundance matching and assembly bias in the eagle simulation. *Monthly Notices of the Royal Astronomical Society*, 460(3):3100–3118, 2016.

Cimatti, A., Daddi, E., and Renzini, A. Mass downsizing and "top-down" assembly of early-type galaxies. *A&A*, 453(2):L29–L33, 2006. doi: 10.1051/0004-6361:20065155. URL https://doi.org/10.1051/0004-6361:20065155.

P. Coles and F. Lucchin. *Cosmology: The origin and evolution of cosmic structure*. John Wiley & Sons, 2003.

C. Conroy, R. H. Wechsler, and A. V. Kravtsov. Modeling luminosity-dependent galaxy clustering through cosmic time. *The Astrophysical Journal*, 647(1):201, 2006.

D. J. Croton, V. Springel, S. D. White, G. De Lucia, C. S. Frenk, L. Gao, A. Jenkins, G. Kauffmann, J. Navarro, and N. Yoshida. The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies. *Monthly Notices of the Royal Astronomical Society*, 365(1):11–28, 2006.

Davidzon, Ilbert, O., Laigle, C., Coupon, J., McCracken, H. J., Delvecchio, I., Masters, D., Capak, P., Hsieh, B. C., Le Fèvre, O., Tresse, L., Bethermin, M., Chang, Y.-Y., Faisst, A. L., Le Floc´h, E., Steinhardt, C., Toft, S., Aussel, H., Dubois, C., Hasinger, G., Salvato, M., Sanders, D. B., Scoville, N., and Silverman, J. D. The cosmos2015 galaxy stellar mass function - thirteen billion years of stellar mass assembly in ten snapshots. *A&A*, 605:A70, 2017. doi: 10.1051/0004-6361/201730419. URL https://doi.org/10.1051/0004-6361/201730419.

I. Davidzon, O. Cucciati, M. Bolzonella, G. De Lucia, G. Zamorani, S. Arnouts, T. Moutard, O. Ilbert, B. Garilli, M. Scodeggio, L. Guzzo, U. Abbas, C. Adami, J. Bel, D. Bottini, E. Branchini, A. Cappi, J. Coupon, S. de la Torre, C. Di Porto, A. Fritz, P. Franzetti, M. Fumana, B. R. Granett, L. Guennou, A. Iovino, J. Krywult, V. Le Brun, O. Le Fèvre, D. Maccagni, K. Małek, F. Marulli, H. J. McCracken, Y. Mellier, L. Moscardini, M. Polletta, A. Pollo, L. A. M. Tasca, R. Tojeiro,

D. Vergani, and A. Zanichelli. The VIMOS Public Extragalactic Redshift Survey (VIPERS). Environmental effects shaping the galaxy stellar mass function. , 586:A23, Feb 2016. doi: 10.1051/0004-6361/201527129.

M. Davis and M. J. Geller. Galaxy correlations as a function of morphological type. *The Astrophysical Journal*, 208:13–19, 1976.

M. Davis and P. J. E. Peebles. A survey of galaxy redshifts. V. The two-point position and velocity correlations. , 267:465–482, Apr 1983. doi: 10.1086/160884.

M. Davis, G. Efstathiou, C. S. Frenk, and S. D. White. The evolution of large-scale structure in a universe dominated by cold dark matter. *The Astrophysical Journal*, 292:371–394, 1985.

S. de la Torre, L. Guzzo, J. A. Peacock, E. Branchini, A. Iovino, B. R. Granett, U. Abbas, C. Adami, S. Arnouts, J. Bel, M. Bolzonella, D. Bottini, A. Cappi, J. Coupon, O. Cucciati, I. Davidzon, G. De Lucia, A. Fritz, P. Franzetti, M. Fumana, B. Garilli, O. Ilbert, J. Krywult, V. Le Brun, O. Le Fèvre, D. Maccagni, K. Małek, F. Marulli, H. J. McCracken, L. Moscardini, L. Paioro, W. J. Percival, M. Polletta, A. Pollo, H. Schlagenhaufer, M. Scodeggio, L. A. M. Tasca, R. Tojeiro, D. Vergani, A. Zanichelli, A. Burden, C. Di Porto, A. Marchetti, C. Marinoni, Y. Mellier, P. Monaco, R. C. Nichol, S. Phleps, M. Wolk, and G. Zamorani. The VIMOS Public Extragalactic Redshift Survey (VIPERS) . Galaxy clustering and redshift-space distortions at z 0.8 in the first data release. , 557:A54, Sept. 2013. doi: 10.1051/0004-6361/201321463.

A. Dekel and Y. Birnboim. Galaxy bimodality due to cold flows and shock heating. *Monthly Notices of the Royal Astronomical Society*, 368(1):2–20, 03 2006. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2006.10145.x. URL https://doi.org/10.1111/j.1365-2966.2006.10145.x.

A. Dekel and J. Silk. The Origin of Dwarf Galaxies, Cold Dark Matter, and Biased Galaxy Formation. , 303:39, Apr 1986. doi: 10.1086/164050.

O. J. Eggen, D. Lynden-Bell, and A. R. Sandage. Evidence from the motions of old stars that the Galaxy collapsed. , 136:748, Nov 1962. doi: 10.1086/147433.

A. Einstein. Die Grundlage der allgemeinen Relativitätstheorie. *Annalen der Physik*, 354(7):769–822, Jan 1916. doi: 10.1002/andp.19163540702.

E. Emsellem, M. Cappellari, D. Krajnović, G. Van De Ven, R. Bacon, M. Bureau, R. L. Davies, P. De Zeeuw, J. Falcón-Barroso, H. Kuntschner, et al. Fast and slow rotators: the build-up of the red sequence. *Proceedings of the International Astronomical Union*, 3(S245):11–14, 2007.

I. M. H. Etherington. On the Definition of Distance in General Relativity. *Philosophical Magazine*, 15(18):761, Jan 1933.

F. Fontanot, R. S. Somerville, L. Silva, P. Monaco, and R. Skibba. Evaluating and improving semi-analytic modelling of dust in galaxies based on radiative transfer calculations. , 392(2):553–569, Jan 2009. doi: 10.1111/j.1365-2966.2008.14126.x.

R. Freedman and W. Kaufmann. *Universe*. W. H. Freeman, 2007. ISBN 9781429281300. URL https://books.google.it/books?id=hfXHTJ6OK9oC.

G. M. Fuller, G. J. Mathews, and C. R. Alcock. Quark-hadron phase transition in the early universe: Isothermal baryon-number fluctuations and primordial nucleosynthesis. *Phys. Rev. D*, 37:1380–1400, Mar 1988. doi: 10.1103/PhysRevD.37.1380. URL https://link.aps.org/doi/10.1103/PhysRevD.37.1380.

J. M. Gabor, R. Davé, K. Finlator, and B. D. Oppenheimer. How is star formation quenched in massive galaxies? , 407(2):749–771, Sep 2010. doi: 10.1111/j.1365-2966.2010.16961.x.

L. Gao, S. D. White, A. Jenkins, F. Stoehr, and V. Springel. The subhalo populations of λcdm dark haloes. *Monthly Notices of the Royal Astronomical Society*, 355(3):819–834, 2004.

L. Gao, V. Springel, and S. D. White. The age dependence of halo clustering. *Monthly Notices of the Royal Astronomical Society: Letters*, 363(1):L66–L70, 2005.

H. Gil-Marín, J. Noreña, L. Verde, W. J. Percival, C. Wagner, M. Manera, and D. P. Schneider. The power spectrum and bispectrum of SDSS DR11

BOSS galaxies - I. Bias and gravity. , 451(1):539–580, July 2015a. doi: 10.1093/mnras/stv961.

H. Gil-Marín, L. Verde, J. Noreña, A. J. Cuesta, L. Samushia, W. J. Percival, C. Wagner, M. Manera, and D. P. Schneider. The power spectrum and bispectrum of SDSS DR11 BOSS galaxies - II. Cosmological interpretation. , 452(2):1914–1921, Sept. 2015b. doi: 10.1093/mnras/stv1359.

C. Giocoli, G. Tormen, and F. C. Van Den Bosch. The population of dark matter subhaloes: mass functions and average mass-loss rates. *Monthly Notices of the Royal Astronomical Society*, 386(4):2135–2144, 2008.

C. Giocoli, G. Tormen, R. K. Sheth, and F. C. van den Bosch. The substructure hierarchy in dark matter haloes. , 404(1):502–517, May 2010. doi: 10.1111/j.1365-2966.2010.16311.x.

C. Giocoli, G. Tormen, R. K. Sheth, and F. C. van den Bosch. The substructure hierarchy in dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 404(1):502–517, 2010.

R. Giovanelli, M. Haynes, and G. Chincarini. Morphological segregation in the pisces-perseus supercluster. *The Astrophysical Journal*, 300:77–92, 1986.

G. Girelli, L. Pozzetti, M. Bolzonella, C. Giocoli, F. Marulli, and M. Baldi. The stellar-to-halo mass relation over the past 12 Gyr. *arXiv e-prints*, art. arXiv:2001.02230, Jan 2020.

G. J. Graves, S. M. Faber, R. P. Schiavon, and R. Yan. Ages and abundances of red sequence galaxies as a function of liner emission-line strength. *The Astrophysical Journal*, 671(1):243, 2007.

A. Grazian, A. Fontana, P. Santini, J. S. Dunlop, H. C. Ferguson, M. Castellano, R. Amorin, M. L. N. Ashby, G. Barro, P. Behroozi, and et al. The galaxy stellar mass function at 3.5 z 7.5 in the candels/uds, goods-south, and hudf fields. *Astronomy Astrophysics*, 575:A96, Mar 2015. ISSN 1432-0746. doi: 10.1051/0004-6361/201424750. URL http://dx.doi.org/10.1051/0004-6361/201424750.

Q. Guo, S. Cole, C. G. Lacey, C. M. Baugh, C. S. Frenk, P. Norberg, R. Auld, I. K. Baldry, S. P. Bamford, N. Bourne, E. S. Buttiglione, A. Cava,

A. Cooray, S. Croom, A. Dariush, G. de Zotti, S. Driver, L. Dunne, S. Dye, S. Eales, J. Fritz, A. Hopkins, R. Hopwood, E. Ibar, R. J. Ivison, M. Jarvis, D. H. Jones, L. Kelvin, J. Liske, J. Loveday, S. J. Maddox, H. Parkinson, E. Pascale, J. A. Peacock, M. Pohlen, M. Prescott, E. E. Rigby, A. Robotham, G. Rodighiero, R. Sharp, D. J. B. Smith, P. Temi, and E. van Kampen. Which haloes host Herschel-ATLAS galaxies in the local Universe? , 412(4):2277–2285, Apr 2011a. doi: 10.1111/j.1365-2966.2010.18051.x.

Q. Guo, S. White, M. Boylan-Kolchin, G. De Lucia, G. Kauffmann, G. Lemson, C. Li, V. Springel, and S. Weinmann. From dwarf spheroidals to cD galaxies: simulating the galaxy population in a ΛCDM cosmology. , 413 (1):101–131, May 2011b. doi: 10.1111/j.1365-2966.2010.18114.x.

A. H. Guth. Inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev. D*, 23:347–356, Jan 1981. doi: 10.1103/PhysRevD.23.347. URL https://link.aps.org/doi/10.1103/PhysRevD.23.347.

B. M. B. Henriques, S. D. M. White, P. A. Thomas, R. Angulo, Q. Guo, G. Lemson, V. Springel, and R. Overzier. Galaxy formation in the Planck cosmology – I. Matching the observed evolution of star formation rates, colours and stellar masses. *Monthly Notices of the Royal Astronomical Society*, 451(3):2663–2680, 06 2015. ISSN 0035-8711. doi: 10.1093/mnras/stv705. URL https://doi.org/10.1093/mnras/stv705.

E. P. Hubble. Extragalactic nebulae. , 64:321–369, Dec 1926. doi: 10.1086/143018.

O. Ilbert, M. Salvato, E. Le Floc'h, H. Aussel, P. Capak, H. J. McCracken, B. Mobasher, J. Kartaltepe, N. Scoville, D. B. Sanders, S. Arnouts, K. Bundy, P. Cassata, J. P. Kneib, A. Koekemoer, O. Le Fèvre, S. Lilly, J. Surace, Y. Taniguchi, L. Tasca, D. Thompson, L. Tresse, M. Zamojski, G. Zamorani, and E. Zucca. Galaxy Stellar Mass Assembly Between 0.2 < z < 2 from the S-COSMOS Survey. , 709(2):644–663, Feb 2010. doi: 10.1088/0004-637X/709/2/644.

O. Ilbert, H. J. McCracken, O. Le Fèvre, P. Capak, J. Dunlop, A. Karim, M. A. Renzini, K. Caputi, S. Boissier, S. Arnouts, H. Aussel, J. Comparat, Q. Guo, P. Hudelot, J. Kartaltepe, J. P. Kneib, J. K. Krogager,

E. Le Floc'h, S. Lilly, Y. Mellier, B. Milvang-Jensen, T. Moutard, M. Onodera, J. Richard, M. Salvato, D. B. Sanders, N. Scoville, J. D. Silverman, Y. Taniguchi, L. Tasca, R. Thomas, S. Toft, L. Tresse, D. Vergani, M. Wolk, and A. Zirm. Mass assembly in quiescent and star-forming galaxies since z 4 from UltraVISTA. , 556:A55, Aug 2013. doi: 10.1051/0004-6361/201321100.

J. H. Jeans. The Stability of a Spherical Nebula. *Philosophical Transactions of the Royal Society of London Series A*, 199:1–53, Jan 1902. doi: 10.1098/rsta.1902.0012.

B. J. T. Jones, V. J. Martínez, E. Saar, and V. Trimble. Scaling laws in the distribution of galaxies. *Reviews of Modern Physics*, 76(4):1211–1266, Feb 2005. ISSN 1539-0756. doi: 10.1103/revmodphys.76.1211. URL http://dx.doi.org/10.1103/RevModPhys.76.1211.

N. Kaiser. Clustering in real space and in redshift space. , 227:1–21, Jul 1987. doi: 10.1093/mnras/227.1.1.

G. Kauffmann, S. D. M. White, and B. Guiderdoni. The formation and evolution of galaxies within merging dark matter haloes. , 264:201–218, Sep 1993. doi: 10.1093/mnras/264.1.201.

E. Keihänen, H. Kurki-Suonio, V. Lindholm, A. Viitanen, A. S. Suur-Uski, V. Allevato, E. Branchini, F. Marulli, P. Norberg, D. Tavagnacco, S. de la Torre, J. Valiviita, M. Viel, J. Bel, M. Frailis, and A. G. Sánchez. Estimating the galaxy two-point correlation function using a split random catalog. , 631:A73, Nov 2019. doi: 10.1051/0004-6361/201935828.

M. Kerscher, I. Szapudi, and A. S. Szalay. A comparison of estimators for the two-point correlation function. *The Astrophysical Journal*, 535(1): L13–L16, may 2000. doi: 10.1086/312702.

E. W. Kolb and M. S. Turner. *The early universe*, volume 69. 1990.

A. Kravtsov, A. Vikhlinin, and A. Meshcheryakov. Stellar mass—halo mass relation and star formation efficiency in high-mass halos. *Astronomy Letters*, 44(1):8–34, 2018.

A. V. Kravtsov, A. A. Berlind, R. H. Wechsler, A. A. Klypin, S. Gottlöber, B. o. Allgood, and J. R. Primack. The Dark Side of the Halo Occupation Distribution. , 609(1):35–49, Jul 2004. doi: 10.1086/420959.

A. V. Kravtsov, A. A. Berlind, R. H. Wechsler, A. A. Klypin, S. Gottloeber, B. Allgood, and J. R. Primack. The dark side of the halo occupation distribution. *The Astrophysical Journal*, 609(1):35, 2004.

P. Kroupa. On the variation of the initial mass function. , 322(2):231–246, Apr 2001. doi: 10.1046/j.1365-8711.2001.04022.x.

S. D. Landy and A. S. Szalay. Bias and Variance of Angular Correlation Functions. , 412:64, Jul 1993. doi: 10.1086/172900.

T. Lazeyras, M. Musso, and F. Schmidt. Large-scale assembly bias of dark matter halos. *Journal of Cosmology and Astroparticle Physics*, 2017(03): 059, 2017.

J. Leja, J. S. Speagle, B. D. Johnson, C. Conroy, P. van Dokkum, and M. Franx. A new census of the 0.2 ¡ z ¡ 3.0 universe, part i: The stellar mass function, 2019.

Z. Li, P. Wu, and H. Yu. COSMOLOGICAL-MODEL-INDEPENDENT TESTS FOR THE DISTANCE–DUALITY RELATION FROM GALAXY CLUSTERS AND TYPE ia SUPERNOVA. *The Astrophysical Journal*, 729(1):L14, feb 2011. doi: 10.1088/2041-8205/729/1/l14. URL https://doi.org/10.1088%2F2041-8205%2F729%2F1%2Fl14.

W. E. Lucas, I. A. Bonnell, and J. E. Dale. Supernova feedback and the energy deposition in molecular clouds. , Feb 2020. doi: 10.1093/mnras/staa451.

G. Manzoni, M. Scodeggio, C. M. Baugh, P. Norberg, G. De Lucia, A. Fritz, C. P. Haines, G. Zamorani, A. Gargiulo, L. Guzzo, A. Iovino, K. Malek, A. Pollo, M. Siudek, and D. Vergani. Modelling the quenching of star formation activity from the evolution of the colour-magnitude relation in VIPERS. *arXiv e-prints*, art. arXiv:1911.02445, Nov 2019.

A. Marchetti, B. R. Granett, L. Guzzo, A. Fritz, B. Garilli, M. Scodeggio, U. Abbas, C. Adami, S. Arnouts, M. Bolzonella, D. Bottini, A. Cappi, J. Coupon, O. Cucciati, G. De Lucia, S. de la Torre, P. Franzetti,

M. Fumana, O. Ilbert, A. Iovino, J. Krywult, V. Le Brun, O. Le Fevre, D. Maccagni, K. Malek, F. Marulli, H. J. McCracken, B. Meneux, L. Paioro, M. Polletta, A. Pollo, H. Schlagenhaufer, L. Tasca, R. Tojeiro, D. Vergani, A. Zanichelli, J. Bel, M. Bersanelli, J. Blaizot, E. Branchini, A. Burden, I. Davidzon, C. Di Porto, L. Guennou, C. Marinoni, Y. Mellier, L. Moscardini, R. C. Nichol, J. A. Peacock, W. J. Percival, S. Phleps, C. Schimd, M. Wolk, and G. Zamorani. The VIMOS Public Extragalactic Redshift Survey (VIPERS): spectral classification through principal component analysis. , 428(2):1424–1437, Jan. 2013. doi: 10.1093/mnras/sts132.

F. Marulli, S. Bonoli, E. Branchini, L. Moscardini, and V. Springel. Modelling the cosmological co-evolution of supermassive black holes and galaxies - I. BH scaling relations and the AGN luminosity function. , 385(4): 1846–1858, Apr. 2008. doi: 10.1111/j.1365-2966.2008.12988.x.

F. Marulli, S. Bonoli, E. Branchini, R. Gilli, L. Moscardini, and V. Springel. The spatial distribution of x–ray selected agn in the chandra deep fields: a theoretical perspective. *Monthly Notices of the Royal Astronomical Society*, 396(3):1404–1414, 2009.

F. Marulli, M. Bolzonella, E. Branchini, I. Davidzon, S. de la Torre, B. R. Granett, L. Guzzo, A. Iovino, L. Moscardini, A. Pollo, U. Abbas, C. Adami, S. Arnouts, J. Bel, D. Bottini, A. Cappi, J. Coupon, O. Cucciati, G. De Lucia, A. Fritz, P. Franzetti, M. Fumana, B. Garilli, O. Ilbert, J. Krywult, V. Le Brun, O. Le Fèvre, D. Maccagni, K. Małek, H. J. McCracken, L. Paioro, M. Polletta, H. Schlagenhaufer, M. Scodeggio, L. A. M. Tasca, R. Tojeiro, D. Vergani, A. Zanichelli, A. Burden, C. Di Porto, A. Marchetti, C. Marinoni, Y. Mellier, R. C. Nichol, J. A. Peacock, W. J. Percival, S. Phleps, M. Wolk, and G. Zamorani. The VIMOS Public Extragalactic Redshift Survey (VIPERS) . Luminosity and stellar mass dependence of galaxy clustering at 0.5 < z < 1.1. , 557: A17, Sep 2013. doi: 10.1051/0004-6361/201321476.

F. Marulli, A. Veropalumbo, and M. Moresco. CosmoBolognaLib: C++ libraries for cosmological calculations. *Astronomy and Computing*, 14: 35–42, Jan. 2016. doi: 10.1016/j.ascom.2016.01.005.

R. Massey, T. Kitching, and J. Richard. The dark matter of gravita-

tional lensing. *Reports on Progress in Physics*, 73(8):086901, jul 2010. doi: 10.1088/0034-4885/73/8/086901. URL https://doi.org/10.1088%2F0034-4885%2F73%2F8%2F086901.

B. Meneux, L. Guzzo, B. Garilli, O. Le Fevre, A. Pollo, J. Blaizot, G. De Lucia, M. Bolzonella, F. Lamareille, L. Pozzetti, et al. The vimos-vlt deep survey (vvds)-the dependence of clustering on galaxy stellar mass at z˜ 1. *Astronomy & Astrophysics*, 478(2):299–310, 2008.

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

B. Mobasher, T. Dahlen, H. C. Ferguson, V. Acquaviva, G. Barro, S. L. Finkelstein, A. Fontana, R. Gruetzbauch, S. Johnson, Y. Lu, et al. A critical assessment of stellar mass measurement methods. *The Astrophysical Journal*, 808(1):101, 2015.

R. Momose, I. Shimizu, K. Nagamine, K. Shimasaku, N. Kashikawa, and H. Kusakabe. Environmental dependence of galactic properties traced by ly $$\$alpha$$ $forestabsorption(i)$: $variationaccordingtogalaxystellarmassandstar - formationactivity. arXiv\ preprint\ arXiv:2002.07334$, 2020.

B. Moore, G. Lake, T. Quinn, and J. Stadel. On the survival and destruction of spiral galaxies in clusters. , 304(3):465–474, Apr 1999. doi: 10.1046/j.1365-8711.1999.02345.x.

B. P. Moster, R. S. Somerville, C. Maulbetsch, F. C. van den Bosch, A. V. Macciò, T. Naab, and L. Oser. Constraints on the Relationship between Stellar Mass and Halo Mass at Low and High Redshift. , 710(2):903–923, Feb 2010. doi: 10.1088/0004-637X/710/2/903.

B. P. Moster, T. Naab, and S. D. M. White. Galactic star formation and accretion histories from matching galaxies to dark matter haloes. , 428(4):3121–3138, Feb 2013. doi: 10.1093/mnras/sts261.

T. Moutard, S. Arnouts, O. Ilbert, J. Coupon, I. Davidzon, L. Guzzo, P. Hudelot, H. McCracken, L. Van Werbaeke, G. Morrison, et al. The vipers multi-lambda survey-ii. diving with massive galaxies in 22 square degrees since z= 1.5. *Astronomy & Astrophysics*, 590:A103, 2016.

S. J. Mutch, D. J. Croton, and G. B. Poole. The simplest model of galaxy formation I: A formation history model of galaxy stellar mass growth. *Mon. Not. Roy. Astron. Soc.*, 435:2445, 2013. doi: 10.1093/mnras/stt1453.

T. Naab and J. P. Ostriker. Theoretical challenges in galaxy formation. *Annual Review of Astronomy and Astrophysics*, 55(1):59–109, 2017. doi: 10.1146/annurev-astro-081913-040019. URL https://doi.org/10.1146/annurev-astro-081913-040019.

J. F. Navarro, C. S. Frenk, and S. D. M. White. A universal density profile from hierarchical clustering. *The Astrophysical Journal*, 490(2):493–508, dec 1997. doi: 10.1086/304888.

D. Nelson, V. Springel, A. Pillepich, V. Rodriguez-Gomez, P. Torrey, S. Genel, M. Vogelsberger, R. Pakmor, F. Marinacci, R. Weinberger, L. Kelley, M. Lovell, B. Diemer, and L. Hernquist. The IllustrisTNG simulations: public data release. *Computational Astrophysics and Cosmology*, 6(1):2, May 2019. doi: 10.1186/s40668-019-0028-x.

P. Norberg, C. M. Baugh, E. Hawkins, S. Maddox, D. Madgwick, O. Lahav, S. Cole, C. S. Frenk, I. Baldry, J. Bland-Hawthorn, et al. The 2df galaxy redshift survey: the dependence of galaxy clustering on luminosity and spectral type. *Monthly Notices of the Royal Astronomical Society*, 332(4):827–838, 2002.

B. Panter, A. F. Heavens, and R. Jimenez. The mass function of the stellar component of galaxies in the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society*, 355(3):764–768, 12 2004. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2004.08355.x. URL https://doi.org/10.1111/j.1365-2966.2004.08355.x.

B. Panter, R. Jimenez, A. F. Heavens, and S. Charlot. The star formation histories of galaxies in the Sloan Digital Sky Survey. , 378(4):1550–1564, Jul 2007. doi: 10.1111/j.1365-2966.2007.11909.x.

S. Perlmutter, G. Aldering, S. Deustua, S. Fabbro, G. Goldhaber, D. E. Groom, A. G. Kim, M. Y. Kim, R. A. Knop, P. Nugent, C. R. Pennypacker, A. Goobar, R. Pain, I. M. Hook, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B. J. Boyle, A. V. Filippenko,

T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, and W. J. Couch. Cosmology from type ia supernovae, 1998.

Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, J. Chluba, L. P. L. Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J. M. Delouis, E. Di Valentino, J. M. Diego, O. Doré, M. Douspis, A. Ducout, X. Dupac, S. Dusini, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, Y. Fantaye, M. Farhang, J. Fergusson, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, D. Herranz, S. R. Hildebrandt, E. Hivon, Z. Huang, A. H. Jaffe, W. C. Jones, A. Karakci, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, T. S. Kisner, L. Knox, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, P. Lemos, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Lilley, V. Lindholm, M. López-Caniego, P. M. Lubin, Y. Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M. A. Miville-Deschênes, D. Molinari, L. Montier, G. Morgante, A. Moss, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J. L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A. S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti, M. Tomasi, T. Trombetti, L. Valenziano, J. Valiviita, B. Van Tent, L. Vibert, P. Vielva, F. Villa, N. Vittorio, B. D. Wand elt, I. K. Wehus, M. White, S. D. M. White, A. Zacchei, and A. Zonca. Planck 2018

results. VI. Cosmological parameters. *arXiv e-prints*, art. arXiv:1807.06209, Jul 2018.

V. Poulin, T. L. Smith, T. Karwal, and M. Kamionkowski. Early Dark Energy can Resolve the Hubble Tension. , 122(22):221301, Jun 2019. doi: 10.1103/PhysRevLett.122.221301.

L. Pozzetti, M. Bolzonella, E. Zucca, G. Zamorani, S. Lilly, A. Renzini, M. Moresco, M. Mignoli, P. Cassata, L. Tasca, F. Lamareille, C. Maier, B. Meneux, C. Halliday, P. Oesch, D. Vergani, K. Caputi, K. Kovač, A. Cimatti, O. Cucciati, A. Iovino, Y. Peng, M. Carollo, T. Contini, J. P. Kneib, O. Le Févre, V. Mainieri, M. Scodeggio, S. Bardelli, A. Bongiorno, G. Coppa, S. de la Torre, L. de Ravel, P. Franzetti, B. Garilli, P. Kampczyk, C. Knobel, J. F. Le Borgne, V. Le Brun, R. Pellò, E. Perez Montero, E. Ricciardelli, J. D. Silverman, M. Tanaka, L. Tresse, U. Abbas, D. Bottini, A. Cappi, L. Guzzo, A. M. Koekemoer, A. Leauthaud, D. Maccagni, C. Marinoni, H. J. McCracken, P. Memeo, C. Porciani, R. Scaramella, C. Scarlata, and N. Scoville. zCOSMOS - 10k-bright spectroscopic sample. The bimodality in the galaxy stellar mass function: exploring its evolution with redshift. , 523:A13, Nov 2010. doi: 10.1051/0004-6361/200913020.

W. H. Press and P. Schechter. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. , 187:425–438, Feb 1974. doi: 10.1086/152650.

R. M. Reddick, R. H. Wechsler, J. L. Tinker, and P. S. Behroozi. The connection between galaxies and dark matter structures in the local universe. *The Astrophysical Journal*, 771(1):30, 2013.

D. Reed, J. Gardner, T. Quinn, J. Stadel, M. Fardal, G. Lake, and F. Governato. Evolution of the mass function of dark matter haloes. , 346(2): 565–572, Dec 2003. doi: 10.1046/j.1365-2966.2003.07113.x.

D. S. Reed, F. Governato, T. Quinn, J. Stadel, and G. Lake. The age dependence of galaxy clustering. *Monthly Notices of the Royal Astronomical Society*, 378(2):777–784, 2007.

A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, et al. Ob-

servational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal*, 116(3):1009, 1998.

A. G. Riess, S. Casertano, W. Yuan, L. M. Macri, and D. Scolnic. Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond ΛCDM. , 876(1):85, May 2019. doi: 10.3847/1538-4357/ab1422.

B. M. B. S. Robertson, Brant E. Galaxy formation and evolution science in the era of the large synoptic survey telescope. *Annual Review of Astronomy and Astrophysics*, 53(1):51–113, 2019. doi: 10.1038/s42254-019-0067-x. URL https://doi.org/10.1038/s42254-019-0067-x.

T. Ronconi, A. Lapi, M. Viel, and A. Sartori. ScamPy – A sub-halo clustering & abundance matching based Python interface for painting galaxies on the dark matter halo/sub-halo hierarchy. *arXiv e-prints*, art. arXiv:2002.07179, Feb. 2020.

A. J. Ross, R. Tojeiro, and W. J. Percival. Understanding the faint red galaxy population using large-scale clustering measurements from sdss dr7. *Monthly Notices of the Royal Astronomical Society*, 413(3):2078–2086, 2011.

V. C. Rubin, J. Ford, W. K., and N. Thonnard. Extended rotation curves of high-luminosity spiral galaxies. IV. Systematic dynamical properties, Sa -> Sc. , 225:L107–L111, Nov 1978. doi: 10.1086/182804.

E. E. Salpeter. The Luminosity Function and Stellar Evolution. , 121:161, Jan 1955. doi: 10.1086/145971.

K. Schawinski, S. Virani, B. Simmons, C. M. Urry, E. Treister, S. Kaviraj, and B. Kushkuley. Do Moderate-Luminosity Active Galactic Nuclei Suppress Star Formation? , 692(1):L19–L23, Feb 2009. doi: 10.1088/0004-637X/692/1/L19.

M. Schmidt. The Rate of Star Formation. , 129:243, Mar 1959. doi: 10.1086/146614.

M. I. Scrimgeour, T. Davis, C. Blake, J. B. James, G. B. Poole, L. Staveley-Smith, S. Brough, M. Colless, C. Contreras, W. Couch, S. Croom, D. Croton, M. J. Drinkwater, K. Forster, D. Gilbank, M. Gladders, K. Glazebrook, B. Jelliffe, R. J. Jurek, I.-h. Li, B. Madore, D. C. Martin, K. Pimbblet,

M. Pracy, R. Sharp, E. Wisnioski, D. Woods, T. K. Wyder, and H. K. C. Yee. The WiggleZ Dark Energy Survey: the transition to large-scale cosmic homogeneity. *Monthly Notices of the Royal Astronomical Society*, 425(1): 116–134, 09 2012. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2012.21402.x. URL https://doi.org/10.1111/j.1365-2966.2012.21402.x.

R. K. Sheth and G. Tormen. Large-scale bias and the peak background split. , 308(1):119–126, Sep 1999. doi: 10.1046/j.1365-8711.1999.02692.x.

R. K. Sheth and G. Tormen. On the environmental dependence of halo formation. *Monthly Notices of the Royal Astronomical Society*, 350(4):1385–1390, 2004.

R. K. Sheth, H. J. Mo, and G. Tormen. Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes. , 323 (1):1–12, May 2001. doi: 10.1046/j.1365-8711.2001.04006.x.

J. L. Sievers, J. R. Bond, J. K. Cartwright, C. R. Contaldi, B. S. Mason, S. T. Myers, S. Padin, T. J. Pearson, U.-L. Pen, D. Pogosyan, S. Prunet, A. C. S. Readhead, M. C. Shepherd, P. S. Udomprasert, L. Bronfman, W. L. Holzapfel, and J. May. Cosmological parameters from cosmic background imager observations and comparisons with BOOMERANG, DASI, and MAXIMA. *The Astrophysical Journal*, 591(2):599–622, jul 2003. doi: 10.1086/375510. URL https://doi.org/10.1086%2F375510.

V. Simha and S. Cole. Cosmological constraints from applying sham to rescaled cosmological simulations. *Monthly Notices of the Royal Astronomical Society*, 436(2):1142–1151, 2013.

R. S. Somerville and R. Davé. Physical models of galaxy formation in a cosmological framework. *Annual Review of Astronomy and Astrophysics*, 53(1):51–113, 2015. doi: 10.1146/annurev-astro-082812-140951. URL https://doi.org/10.1146/annurev-astro-082812-140951.

R. S. Somerville, P. F. Hopkins, T. J. Cox, B. E. Robertson, and L. Hernquist. A semi-analytic model for the co-evolution of galaxies, black holes and active galactic nuclei. , 391(2):481–506, Dec 2008. doi: 10.1111/j.1365-2966.2008.13805.x.

R. S. Somerville, R. C. Gilmore, J. R. Primack, and A. Domínguez. Galaxy properties from the ultraviolet to the far-infrared: cold dark matter models confront observations. *Monthly Notices of the Royal Astronomical Society*, 423(3):1992–2015, Apr 2012. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2012.20490.x. URL http://dx.doi.org/10.1111/j.1365-2966.2012.20490.x.

V. Springel, T. Di Matteo, and L. Hernquist. Black Holes in Galaxy Mergers: The Formation of Red Elliptical Galaxies. , 620(2):L79–L82, Feb 2005. doi: 10.1086/428772.

V. Springel, J. Wang, M. Vogelsberger, A. Ludlow, A. Jenkins, A. Helmi, J. F. Navarro, C. S. Frenk, and S. D. M. White. The Aquarius Project: the subhaloes of galactic haloes. , 391(4):1685–1711, Dec 2008. doi: 10.1111/j.1365-2966.2008.14066.x.

I. Strateva, Ž. Ivezić, G. R. Knapp, V. K. Narayanan, M. A. Strauss, J. E. Gunn, R. H. Lupton, D. Schlegel, N. A. Bahcall, J. Brinkmann, R. J. Brunner, T. Budavári, I. Csabai, F. J. Castander, M. Doi, M. Fukugita, Z. Győry, M. Hamabe, G. Hennessy, T. Ichikawa, P. Z. Kunszt, D. Q. Lamb, T. A. McKay, S. Okamura, J. Racusin, M. Sekiguchi, D. P. Schneider, K. Shimasaku, and D. York. Color Separation of Galaxy Types in the Sloan Digital Sky Survey Imaging Data. , 122(4):1861–1874, Oct 2001. doi: 10.1086/323301.

D. Thomas, C. Maraston, R. Bender, and C. Mendes de Oliveira. The Epochs of Early-Type Galaxy Formation as a Function of Environment. , 621(2):673–694, Mar 2005. doi: 10.1086/426932.

D. Thomas, C. Maraston, K. Schawinski, M. Sarzi, and J. Silk. Environment and self-regulation in galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 404(4):1775–1789, 2010.

J. Tinker, A. V. Kravtsov, A. Klypin, K. Abazajian, M. Warren, G. Yepes, S. Gottlöber, and D. E. Holz. Toward a halo mass function for precision cosmology: the limits of universality. *The Astrophysical Journal*, 688(2):709, 2008.

B. M. Tinsley. Evolution of the Stars and Gas in Galaxies. , 5:287–388, Jan 1980.

A. Toomre and J. Toomre. Galactic bridges and tails. *The Astrophysical Journal*, 178:623–666, 1972.

S. Trujillo-Gomez, A. Klypin, J. Primack, and A. J. Romanowsky. Galaxies in λcdm with halo abundance matching: Luminosity-velocity relation, baryonic mass-velocity relation, velocity function, and clustering. *The Astrophysical Journal*, 742(1):16, 2011.

F. C. Van Den Bosch, G. Tormen, and C. Giocoli. The mass function and average mass-loss rate of dark matter subhaloes. *Monthly Notices of the Royal Astronomical Society*, 359(3):1029–1040, 2005.

A. Vikhlinin, A. Kravtsov, W. Forman, C. Jones, M. Markevitch, S. S. Murray, and L. V. Speybroeck. ChandraSample of nearby relaxed galaxy clusters: Mass, gas fraction, and mass-temperature relation. *The Astrophysical Journal*, 640(2):691–709, apr 2006. doi: 10.1086/500288. URL https://doi.org/10.1086%2F500288.

B. Vulcani, B. M. Poggianti, A. Dressler, G. Fasano, T. Valentinuzzi, W. Couch, A. Moretti, L. Simard, V. Desai, D. Bettoni, M. D'Onofrio, A. Cava, and J. Varela. The evolution of early-type galaxies in clusters from z 0.8 to z 0: the ellipticity distribution and the morphological mix. , 413 (2):921–941, May 2011. doi: 10.1111/j.1365-2966.2010.18182.x.

L. Wang, D. Farrah, S. J. Oliver, A. Amblard, M. Béthermin, J. Bock, A. Conley, A. Cooray, M. Halpern, S. Heinis, E. Ibar, O. Ilbert, R. J. Ivison, G. Marsden, I. G. Roseboom, M. Rowan-Robinson, B. Schulz, A. J. Smith, M. Viero, and M. Zemcov. Connecting stellar mass and star-formation rate to dark matter halo mass out to z 2. *Monthly Notices of the Royal Astronomical Society*, 431(1):648–661, 03 2013. ISSN 0035-8711. doi: 10.1093/mnras/stt190. URL https://doi.org/10.1093/mnras/stt190.

B. J. Weiner, A. C. Phillips, S. Faber, C. N. Willmer, N. P. Vogt, L. Simard, K. Gebhardt, M. Im, D. Koo, V. L. Sarajedini, et al. The deep groth strip galaxy redshift survey. iii. redshift catalog and properties of galaxies. *The Astrophysical Journal*, 620(2):595, 2005.

S. D. M. White and M. J. Rees. Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering. *Monthly Notices of the Royal Astronomical Society*, 183(3):341–358, 07 1978. ISSN 0035-8711. doi:

10.1093/mnras/183.3.341. URL https://doi.org/10.1093/mnras/183.3.341.

X. Yang, H. J. Mo, and F. C. van den Bosch. Galaxy Groups in the SDSS DR4. II. Halo Occupation Statistics. , 676(1):248–261, Mar 2008. doi: 10.1086/528954.

I. Zehavi, Z. Zheng, D. H. Weinberg, J. A. Frieman, A. A. Berlind, M. R. Blanton, R. Scoccimarro, R. K. Sheth, M. A. Strauss, I. Kayo, Y. Suto, M. Fukugita, O. Nakamura, N. A. Bahcall, J. Brinkmann, J. E. Gunn, G. S. Hennessy, Ž. Ivezić, G. R. Knapp, J. Loveday, A. Meiksin, D. J. Schlegel, D. P. Schneider, I. Szapudi, M. Tegmark, M. S. Vogeley, D. G. York, and SDSS Collaboration. The Luminosity and Color Dependence of the Galaxy Correlation Function. , 630(1):1–27, Sep 2005. doi: 10.1086/431891.

I. Zehavi, Z. Zheng, D. H. Weinberg, M. R. Blanton, N. A. Bahcall, A. A. Berlind, J. Brinkmann, J. A. Frieman, J. E. Gunn, R. H. Lupton, R. C. Nichol, W. J. Percival, D. P. Schneider, R. A. Skibba, M. A. Strauss, M. Tegmark, and D. G. York. Galaxy Clustering in the Completed SDSS Redshift Survey: The Dependence on Color and Luminosity. , 736(1):59, Jul 2011. doi: 10.1088/0004-637X/736/1/59.

I. Zehavi, S. Contreras, N. Padilla, N. J. Smith, C. M. Baugh, and P. Norberg. The Impact of Assembly Bias on the Galaxy Content of Dark Matter Halos. , 853(1):84, Jan 2018. doi: 10.3847/1538-4357/aaa54a.

A. R. Zentner, A. P. Hearin, and F. C. van den Bosch. Galaxy assembly bias: a significant source of systematic error in the galaxy–halo relationship. *Monthly Notices of the Royal Astronomical Society*, 443(4):3044–3067, 2014.

Z. Zheng, A. A. Berlind, D. H. Weinberg, A. J. Benson, C. M. Baugh, S. Cole, R. Davé, C. S. Frenk, N. Katz, and C. G. Lacey. Theoretical Models of the Halo Occupation Distribution: Separating Central and Satellite Galaxies. , 633(2):791–809, Nov 2005. doi: 10.1086/466510.

Y. Zu, Z. Zheng, G. Zhu, and Y. Jing. Environmental effects on real-space and redshift-space galaxy clustering. *The Astrophysical Journal*, 686(1):41, 2008.

F. Zwicky. Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6:110–127, Jan 1933.