Universidad del País Vasco
Euskal Herriko Unibertsitatea

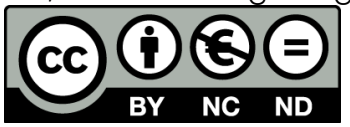
ZIENTZIA ETA TEKNOLOGIA FAKULTATEA
FACULTAD DE CIENCIA Y TECNOLOGÍA

Gradu Amaierako Lana / Trabajo Fin de Grado
Fisikako Gradua / Grado en Física

Dark matter halos and the large-scale structure of the Universe

A study of cosmological N-body simulations

Egilea/Autor:
Mikel Olaciregui Segura
Zuzendaria/Director:
Raúl Angulo de la Fuente
Raül Vera Jiménez



Leioa, 2020ko otsailaren 20a / Leioa, 20 de febrero de 2020

Contents

Introduction and objectives

Physical cosmology is a field that combines physics and astronomy to study the fundamental properties of the universe, its origin and its evolution. One of its branches deals with the large-scale structure (LSS) of the universe and its dynamics: from planetary systems and galaxies to superclusters and beyond, modern observations show that our universe is full with rich and varied structure, on every scale and in every direction.

For many years, detailed data about LSS was scarce, and as a consequence the field remained a mostly qualitative line of research. However, the last few decades have seen the rise of precision cosmology, with accurate measurements of the Cosmic Microwave Background (CMB) and systematic redshift surveys of galaxies, among many other observations. These have given cosmologists the opportunity to test their models against actual measures, and at the same time they have highlighted the need for accurate theoretical models that can explain the new observations. As a result, a robust paradigm has been developed: the standard model known as the Lambda-Cold Dark Matter (ΛCDM) cosmology, which combines Einstein's General Relativity with the most recent observational evidence.

On the one hand, measurements of the CMB have shown that the primitive universe was extremely homogeneous, with fluctuations on the order of one part in ten thousand. On the other hand, galaxy surveys indicate that at the present time the universe is dominated by compact structures, with large voids between them. These objects are called "halos" and much of their mass is in the form of cold dark matter, a mysterious substance whose presence can only be inferred from its gravitational effects. But then, how did the universe evolve from its initial homogeneity to this very different configuration?

The growth of LSS is therefore a very active area of research within physical cosmology (Springel, Frenk, and S. D. M. White 2006). For early times after the Big Bang, when the universe was very smooth, density fluctuations can be treated within linear perturbation theory. At the same time that the universe expanded from a very small initial size, initially overdense regions of space became increasingly dense, while underdensities also became even less dense. At some point, the fluctuations became so large that they cannot be treated as perturbations, and most matter clustered in collapsed structures called halos.

For the early researchers of LSS, this nonlinear stage of evolution was only accessible through analytical approximations. But in the last few decades, the steady advancement of computing technology has equipped researchers with a powerful new tool in the form of numerical simulations. By discretizing the contents of the universe into finite elements, they allow to study the dynamics and statistics of halo evolution, and to generate semi-analytic fits that attempt to predict the structure of halos depending on the values of the free parameters of the ΛCDM model.

In the near future, gravitational lensing surveys will provide abundant information about the internal structure of halos, and particularly their concentration parameter, c. Comparing these observational results to the simulated halos will allow cosmologists to further constrain the cosmological parameters and also test the viability of extensions or alternatives to the current models.

Simulations, however, are not a perfect tool – they are subject to different kinds of biases and artifacts. If a suite of simulations is to be used to develop a new fit, those numerical effects have to be carefully accounted for to ensure the robustness of the results.

This project was carried out as part of an internship at the Computational Cosmology group of the Donostia International Physics Centre (DIPC). The main objectives were:

- Studying the bibliography concerning the ΛCDM cosmology and its physical motivation.
- Understanding the physical processes that lead to the linear growth of cosmological structure, and their influence on the internal characteristics of collapsed objects.
- Learning how the nonlinear evolution of the density field is studied through state-of-the-art numerical simulations.
- Testing the results of a suite of N-body simulations in several cosmologies, with special attention to the effects of mass, redshift and cosmological parameters on halo concentration, as well as their numerical limitations.

The rest of this report is organized as follows. Chapters 1-4 follow each of the four objectives, with the first three being largely theoretical and the last being dedicated to the main results. After those, the conclusions of the project are presented, as well as the main bibliographical references. Appendices A and B contain mathematical derivations deemed too long to be included in the main text; Appendices C and D contain extra plots of some results; and Appendix E provides a list of all the symbols that are used across the report's notation. Footnotes through the text provide contextual information that may be of interest to the non-specialized reader.

Chapter 1

The smooth universe

This chapter consists of three sections. In the first one, some preliminary definitions are given that will be useful for developing the theoretical framework. Then, the metric of the flat universe is briefly derived from first principles. Finally, in the third section the physical motivation for the ΛCDM cosmological model is given, which serves as the main framework for the subsequent chapters.

1.1 An expanding universe

In General Relativity there are no inertial reference systems, and therefore the concept of "observers at rest" does not exist. Still, there is a class of observers that "move with the flow" of the universe; for instance, the mean velocity of galaxies is zero when measured in their reference frames. These observers are called ***comoving observers***.

Comoving coordinates

Let us consider an expanding universe. A more exact definition of this is given in Section 1.2, but for now it is sufficient to know this means that, even if each of two comoving observers can be said to be at a "fixed" point in space, the distance between them will increase over time. In other words, observers that feel no movement relative to the universe as a whole will still measure a relative velocity between them.

The usual way to capture this time-dependence is to define both a ***proper distance*** $\vec{r}$ and a ***comoving distance*** $\vec{x}$. The proper distance between two points is the distance measured the usual way; the comoving distance is the distance that would be measured between two comoving observers that were located at those points at a reference epoch, which is often taken to be the present day. The two are related by

$$\vec{r}(t) = a(t) \cdot \vec{x}, \tag{1.1}$$

where the dimensionless function $a(t)$ is called the ***scale factor*** and measures the expansion of the universe. It is customarily normalized to its present value: $a(t_0) = 1$, so that the comoving distance between two objects is equal to the current proper distance.

To put it another, perhaps more intuitive way: we can think of the universe as a grid where each cell is a comoving observer. The comoving distance between two points

is the number of cells between them, while the proper distance is the number of cells multiplied by the (time-dependent) size of the cells.

We shall now, for the sake of clarity, drop vector notation and work with the moduli. Differentiating with respect to time the proper distance x between two observers, we get two velocity terms:

$$\dot{\vec{x}} \equiv \frac{d}{dt}\left(a \cdot \vec{r}\right) = \dot{a}\vec{r} + a\dot{\vec{r}} = aH\vec{r} + \vec{v}_p \tag{1.2}$$

Here we have defined the ***peculiar velocity*** v_p, which is zero if the two points are both comoving; we have also defined the ***Hubble parameter***,

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

which will be important later on.

Cosmological redshift

The evolution of the scale factor has an important effect on the propagation of light. Let us consider a comoving source at some time t_e that emits radiation with wavelength λ_e, and therefore two successive wave-crests are emitted with a separation of λ_e/c, where c is the speed of light. The space between them is stretched as they travel, and a comoving observer at the present time t_0 measures a proper distance between wave-crests (i.e. a wavelength) λ_0. The comoving distance traversed by each can be obtained by integrating equation (1.1) over time; since both comoving distances are equal, we get

$$\int_{t_e}^{t_0} \frac{dt}{a(t)} = \int_{t_e+\lambda_e/c}^{t_0+\lambda_0/c} \frac{dt}{a(t)} \quad \Rightarrow \quad \int_{t_e}^{t_e+\lambda_e/c} \frac{dt}{a(t)} = \int_{t_0}^{t_0+\lambda_0/c} \frac{dt}{a(t)}. \tag{1.4}$$

The universe expands very slowly on the typical period of an electromagnetic wave, so we can assume the scale factor to be constant along each interval of integration and therefore:

$$\frac{\lambda_e}{a(t_e)} = \frac{\lambda_0}{a(t_0)} = \lambda_0 \quad \Rightarrow \quad \frac{\lambda_0}{\lambda_e} = \frac{1}{a(t_e)} \equiv z + 1, \tag{1.5}$$

where z is called the ***cosmological redshift***: the light emitted in the past arrives at us with its wavelength stretched by the expansion of space itself, in a way that is completely independent of the special-relativistic Doppler effect caused by the peculiar velocity.

If we know the functional form of $a(t)$, by measuring the redshift in the spectral lines of observed objects, we can calculate the instant at which the light was emitted, as well as our comoving distance and current proper distance to said objects. As a consequence of this, redshift is often used interchangeably as a measure of time and as a measure of distance.

1.2 The Cosmological Principle and Friedmann's equations

In Section 1.1 we have assumed the universe to be expanding. Now, we shall see a more rigorous description of that phenomenon.

General Relativity asserts that the geometry of spacetime (and the gravitational interactions that happen in it) is entirely determined by the metric, a 4x4 symmetric tensor[1] $g_{\mu\nu}$ that defines the length of a line element dx^μ:

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu. \tag{1.6}$$

The metric of the universe

We will assume, for now, that the universe is homogeneous and isotropic, i.e. it looks the same everywhere and in every direction. This is known as the ***cosmological principle*** and is a somewhat crude approximation; a weaker but more reasonable version asserts that the universe is "homogeneous and isotropic enough" at "large enough" scales.

For the sake of brevity (but at the cost of generality), we will also assume that the universe is flat, which used to not be obvious for cosmologists in the past but is now true as far as we know. Along with the cosmological principle, this implies that the geometry can only depend on time:

$$ds^2 = -dt^2 + a^2(t) \left[dx^2 + dy^2 + dz^2\right], \tag{1.7}$$

where $dx^2 + dy^2 + dz^2$ is the Euclidean metric for flat 3D space in comoving coordinates, and the speed of light $c = 1$ for convenience. Any metric that satisfies equation (1.7) is called a ***Robertson-Walker metric*** for a flat universe.

The adimensional function $a(t)$ is the ***scale factor*** first referenced in Section 1.1. It represents the ratio of the distance between two comoving points and the value of said distance at a reference time t_0, which is usually taken to be the present time; thus, it can also be thought of as describing the "size of the universe" relative to some reference epoch.

In General Relativity, the relation between the geometry of spacetime (i.e. the metric) and the energy and matter that it contains (the energy-momentum tensor) is given by Einstein's field equation:

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi G T_{\mu\nu}, \tag{1.8}$$

where G is Newton's gravitational constant, $R_{\mu\nu}$ is the Ricci curvature tensor, R is the Ricci curvature scalar, and $T_{\mu\nu}$ is the energy-momentum tensor. The Ricci tensor and Ricci scalar are constructed from the first and second derivatives of the metric; the calculation is fairly straightforward but not vital to our purposes, and thus has been relegated to Appendix A.

The Friedmann equations

In order to know the behavior of $a(t)$ we must

- Model the dynamics of the universe's matter and energy content,

[1] We will use here Einstein's notation, where Latin indices run from 1 to 3 (i.e. they only cover the spacelike dimensions) and Greek indices run from 0 to 3 (i.e. they also cover time). Indices that appear as both a subscript and a superscript are summed over.

- Use that model to construct the energy-momentum tensor, and
- Plug the tensor into Einstein's equation and solve it for the metric.

As a simple but effective approximation, we will assume matter and energy, whatever their composition, behave like a perfect fluid that is at rest in comoving coordinates, and are therefore well-described by a scalar field with homogeneous density ρ and pressure p (Carroll 2014). Then the 4-velocity vector is

$$U^\mu = (1, 0, 0, 0). \tag{1.9}$$

This means that its velocity is 0 in the spacelike dimensions, i.e. it only moves in the forward time direction. Since only the energy density and pressure terms at zero, the energy-momentum tensor is almost trivial:

$$T_{\mu\nu} = (\rho + p)U_\mu U_\nu + p g_{\mu\nu} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}. \tag{1.10}$$

It is worth noting that, since the equation is a relation between 4x4 symmetric tensors, in the most general case it is actually a system of ten nonlinear, coupled, second-order PDEs. This is a mathematical problem of formidable difficulty[2], and indeed many exact solutions bear the names of the people that discovered them.

However, the assumptions that we have made so far (the cosmological principle, the perfect fluid) make the task much less daunting. Plugging everything into Einstein's equation (1.8),

$$\begin{pmatrix} 3\left(\frac{\dot{a}}{a}\right)^2 & 0 & 0 & 0 \\ 0 & -2a\ddot{a} - \dot{a}^2 & 0 & 0 \\ 0 & 0 & -2a\ddot{a} - \dot{a}^2 & 0 \\ 0 & 0 & 0 & -2a\ddot{a} - \dot{a}^2 \end{pmatrix} = \begin{pmatrix} 8\pi G\rho & 0 & 0 & 0 \\ 0 & 8\pi G p & 0 & 0 \\ 0 & 0 & 8\pi G p & 0 \\ 0 & 0 & 0 & 8\pi G p \end{pmatrix}. \tag{1.11}$$

The left-hand side of the equation is a diagonal matrix. Therefore, the perfect fluid was not just an approximation, but a necessary condition mandated by the RW metric and, ultimately, by the cosmological principle. Any other choice of (1.9) would have introduced cross-terms into (1.10).

We can now, with a little algebra, extract the system of equations that describes the time evolution of the scale factor:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho, \tag{1.12}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p). \tag{1.13}$$

These are called the ***Friedmann equations*** and $\dot{a}/a$ is the ***Hubble parameter*** H. Its value at the present time is, by convention,

$$H_0 = 100h\frac{km}{s \cdot Mpc}, \quad \text{where } h \approx 0.6 - 0.7. \tag{1.14}$$

[2]The number of independent equations can be actually lowered to 6, using the Bianchi identities. The remaining four degrees of freedom correspond to the choice of coordinate system.

As a consequence of the uncertainty in the exact value of h, many quantities (like masses and distances) are usually quoted as a function of h.

We now have two equations for three variables: the scale factor a, the density ρ and the pressure p. In order to close the system the equations, we need an ***equation of state*** that relates ρ to p. The simplest such equation for a perfect fluid is

$$p = w\rho, \tag{1.15}$$

where w is a time-independent constant that depends on the nature of the specific perfect fluid we are dealing with.

We can also simplify the equations by explicitly writing the a-dependence of the density ρ, i.e. the evolution of the energy density of the fluid as the universe expands or contracts. There are several ways to do it; here we shall derive it from the First Law of thermodynamics. Consider a volume V of flat 3-dimensional space which contains total energy U. Then, if the scale factor a changes,

$$\begin{aligned} dU = -pdV \quad \Rightarrow \quad d(\rho a^3) = -w\rho d(a^3) \quad \Rightarrow \quad a^3 d\rho + 3\rho a^2 da = -3w\rho a^2 da \\ \Rightarrow a^3 d\rho = -3(w+1)\rho a^2 da \quad \Rightarrow \quad \frac{d\rho}{\rho} = -3(w+1)\frac{da}{a}. \end{aligned} \tag{1.16}$$

This can be easily integrated to obtain the general formula

$$\rho \propto a^{-3(1+w)}, \tag{1.17}$$

and, by introducing this dependence into the first Friedmann equation, we obtain

$$\begin{cases} a \propto t^{\frac{2}{3}(1+w)}, & \text{if} \quad w \neq -1 \\ a \propto e^{Ht}, & \text{if} \quad w = -1 \end{cases} . \tag{1.18}$$

1.3 The ΛCDM cosmology

In the previous section we have derived a general equation for the evolution of a flat universe that is valid as long as its contents can be approximated by a perfect fluid. However, knowing what the universe is made of is not a trivial task, and therefore the cosmological models have evolved substantially over the last century.

What we can see

The most obvious contents of the universe are those that can be directly observed: ***radiation*** and ***baryonic matter***. Radiation is composed of the photons that are emitted in all electromagnetic interactions. These massless particles can be treated as a quantum gas, and statistical physics gives an equation of state $p = 1/3\,\rho$. Plugging this value of w into eqs. (1.17) and (1.18) we obtain

$$\text{Radiation} \qquad w = \frac{1}{3} \qquad \Rightarrow \qquad \rho \propto a^{-4}, \qquad a \propto t^{1/2}. \tag{1.19}$$

The energy density decreases quartically: there is an a^{-3} factor that corresponds to the decrease in the number density of photons, and an additional factor of a^{-1} due

to the fact that photons lose energy as their wavelength increases with redshift. A radiation-dominated universe would grow slowly, as a power law with exponent $1/2$.

Baryonic matter is much trickier. First of all, its name is misleading: in particle physics a "baryon" refers to a composite particle made from an odd number of quarks (typically three). In cosmology, however, "baryons" refers to anything made from the charged fermions of the Standard model, i.e. all flavors of quarks plus electrons, muons and taus. Neutrinos were traditionally neglected, but they have been shown to have some relevant effects, which will be discussed in Section 1.3. The rest of this report will use the term "baryon" in the latter, cosmological sense.

These particles all interact through gravity, electromagnetism and the weak force; quarks also interact through the strong force. As a consequence, it cannot be modelized by a perfect fluid: one of the simplest equations of state, the ideal gas law $p \propto T\rho$, has non-constant w – it is temperature-dependent.

What we cannot see

However, in the 1970s the combination of several observed phenomena previously thought to be unrelated (Swart, Bertone, and Dongen 2017) led cosmologists to realize that baryons are only a small fraction of the total matter present in the universe. Most of the matter density is actually due to so-called ***dark matter*** Its nature is still speculative, but there is agreement that some of its basic properties are:

- It interacts gravitationally, i.e. it is massive, which explained some phenomena like the excess circular velocity of stars in outer regions of galaxies (respect to the apparent inner masses measured through the surface brightness method) and the anomalous velocity distributions of galaxies in clusters. More recently, this has been confirmed by gravitational lensing observations: the measured distortion to the images of background objects due to local perturbations to the metric in galaxy clusters needs much larger masses than the ones that are calculated from direct observations.
- It might interact through the weak force.
- It does not interact electromagnetically nor through the strong force, or in any case the cross-section is extremely small. That is why it is dark: it does not emit any kind of radiation, and it is also completely transparent to light.
- Therefore it is also ***collisionless*** (it is not scattered by other matter nor by itself) and ***dissipationless*** (it cannot lose kinetic energy through cyclotron radiation nor any other radiative process).

Together, these properties make its dynamics easier to model. Since pressure is driven by inter-particle collisions, a fluid composed of collisionless particles will be pressureless and have an equation of state $w = 0$.

$$\text{Dark matter} \qquad w = 0 \qquad \Rightarrow \qquad \rho \propto a^{-3}, \qquad a \propto t^{2/3}. \tag{1.20}$$

This means the energy density of dark matter decreases cubically with a: the total number of particles stays constant, but they occupy a cubically growing volume as the universe expands. A dark matter-dominated universe expands following a power law of index $2/3$, and is also known as an ***Einstein-de Sitter*** (EdS) cosmology.

The cosmological constant

There is another equation of state, which is mathematically simple but with a non-obvious physical meaning. It is the so-called ***cosmological constant***, Λ: a fluid with pressure equal in magnitude and opposite in sign to its energy density.

$$\text{Cosmological constant} \qquad w = -1 \qquad \Rightarrow \qquad \rho \propto a^0, \qquad a \propto e^{Ht}. \tag{1.21}$$

The energy density stays the same even as the universe expands; that is why it is called a constant. An universe dominated by the cosmological constant expands exponentially.

The Friedmann equations (1.12) and (1.13) can be rewritten by taking the cosmological constant out of the energy density term and writing it as a separate term:

$$\left(\frac{\dot{a}}{a}\right)^2 \equiv H^2 = \frac{8\pi G}{3}\rho + \frac{\Lambda}{3} \tag{1.22}$$

$$\frac{\ddot{a}}{a} \equiv \dot{H} + H^2 = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3} \tag{1.23}$$

The standard model

The cosmological constant was originally added by Einstein to the left-hand side[3] of (1.8) in order to have a static universe, but dropped when Hubble discovered that the universe is actually expanding. Then, in the 1990s it was discovered that the expansion is accelerating, which led again to $\Lambda > 0$. On the other hand, current evidence suggests that the universe contains a large amount of cold dark matter (CDM), i.e. dark matter with a non-relativistic velocity distribution.

These two facts, together with Einstein's General Relativity and the Friedmann Equations derived from it, conform the current standard model of cosmology: the ***ΛCDM model***. In this model, the time-dependent energy density ρ_i of each component is expressed as a fraction of $\rho_c = 3H^2/8\pi G$, the total density required for the universe to be flat:

$$\Omega_i = \frac{\rho_i}{\rho_c} = \rho_i \cdot \frac{8\pi G}{3H^2} \tag{1.24}$$

The approximate current values of these densities are $\Omega_b \approx 0.05$ for baryons, $\Omega_{\text{CDM}} \approx 0.26$ for cold dark matter and $\Omega_\Lambda \approx 0.69$ for the cosmological constant. Ω_γ for radiation is currently negligible, but due to the different evolution of each component, it is thought that the universe went initially through a radiation-dominated phase, then became matter dominated, and now its expansion is mainly driven by the cosmological constant. The ΛCDM model has proven very effective at explaining the observable properties of the universe, and several extensions have been proposed to address some of its flaws, which are briefly discussed at the end of this section. It is also a strong theoretical framework for explaining how large scale-structures grow from a near-homogeneous initial state, which is the main topic of Chapter 2.

[3] Therefore it did not have the modern physical interpretation of a negative-pressure fluid. It was instead a purely geometric factor countering the scalar curvature R, perhaps an intrinsic property of space-time itself.

We can take advantage of the fact that Friedmann's equation (1.12) is linear in H^2 and write its explicit dependence on the relative densities Ω_i and their current values $\Omega_{i,0}$. Assuming baryons can be absorbed into a CDM-dominated Ω_m, it reads:

$$\frac{\dot{a}^2(t)}{a^2(t)} \equiv H^2(t) = H_0^2 \left[\Omega_{\gamma,0} a^{-4}(t) + \Omega_{m,0} a^{-3}(t) + \Omega_{\Lambda,0}\right] \tag{1.25}$$

Extensions to the standard model

Cosmological neutrinos

Until now, we have ignored the particularities of neutrinos, which were abundant in the early universe: these particles only interact via gravity and the weak force, and have such small masses that, at high temperatures, they were relativistic and had a radiation-like ($\rho_\nu \propto a^{-4}$) behavior. However, at lower temperatures they become non-relativistic and their behaviour is more similar to that of dark matter (Zennaro 2017).

Since neutrinos do not interact electromagnetically, they decoupled from the rest of the universe at an earlier time than photons decoupled from matter. The epoch at which this happened is roughly the time when the weak force's time scale became slower than the expansion rate H. In other words, a ***Cosmic Neutrino Background*** (CνB) was created when the neutrinos' mean free path became sufficiently long and they ceased to collide with other particles.

The presence of massive neutrinos has an effect on the background metric: since they behave like radiation at early times and like matter at later times (well after decoupling), they introduce an additional term in the evolution of the scale factor (1.25). Additionally, due to their long free-streaming distances, they suppress the fluctuations of the linear density field on small scales (the meaning of this will become clearer in section 2.1).

The cosmological parameter that is commonly used to parametrize the effects of neutrinos is the sum of the three neutrino masses[4]:

$$M_\nu = m_{\nu 1} + m_{\nu 2} + m_{\nu 3} \tag{1.26}$$

Beyond the cosmological constant

The cosmological constant is actually the simplest form of what is called, in a more general sense, ***dark energy***. The exact origin of this element of nature (for lack of a better term) is currently unknown and remains a field of intense research at the intersection of cosmology and quantum field theory.

The next simplest form of dark energy is an Ω_Λ that varies in time. The equation of state is usually parametrized as

$$p = w_0(1 + w_a t)\rho, \tag{1.27}$$

where w_0 is the current equation of state of dark energy, and w_a describes its rate of change with respect to time. The cosmological constant is recovered for $w_0 = -1, w_a = 0$, and any other values lead to a different solution to Friedmann's equations.

[4] Neutrinos have three weak-force eigenstates ν_e, ν_μ, ν_τ, but each is a quantum superposition of the three mass eigenstates

Chapter 2

The linear density field

2.1 The seeds of structure

Until now we have assumed, in accordance to the cosmological principle, that the density $\rho(t)$ of the universe is constant across any spacelike slice $t = \text{const.}$, which means that its value varies only with time. While this could be an acceptable approximation at very large scales or at very early times, it is evident from observation that matter tends to cluster: it forms stars and other bodies, which form galaxies, which in turn form clusters and superclusters separated by large cosmic voids (figure 2.1).

Therefore, a description of cosmic structure requires a measure of the deviation from homogeneity and isotropy. If $\rho(\vec{x}, t)$ is the density field of the universe, we can define an average density $\bar{\rho}$ and a 'density contrast' or 'overdensity field' δ such that

$$\rho(\vec{x}, t) = \bar{\rho}(t) + \bar{\rho}(t) \cdot \delta(\vec{x}, t). \tag{2.1}$$

This adimensional scalar field δ describes the deviation from homogeneity at each point of spacetime, so by construction its mean value, averaged across space, is zero: $\langle \delta(t) \rangle = 0$.

Gaussian random fields

Before continuing further, we shall revisit some mathematical definitions. Let $\Phi(\vec{x})$ be a 3-dimensional, real-valued random scalar field. This means that, for each point $\vec{x}$, the value $\Phi(\vec{x})$ of the field is the outcome of a random process with a normalized probability distribution $p_{\vec{x}}(\Phi)$. At each point, the expected value of the field (i.e. the ensemble average over all possible realizations of the random field) is:

$$\xi_1(\vec{x}) = \langle \Phi(\vec{x}) \rangle = \int \Phi p_{\vec{x}}(\Phi) d\Phi. \tag{2.2}$$

This is also called its one-point correlation function. We can define its two-point correlation function as well,

$$\xi_2(\vec{x}_1, \vec{x}_2) = \langle \Phi(\vec{x}_1)\Phi(\vec{x}_2) \rangle = \int \int \Phi\Phi' p_{\vec{x}_1, \vec{x}_2}(\Phi, \Phi') d\Phi d\Phi', \tag{2.3}$$

which is a measure of how the similarity between the values of the field at each pair of points, averaged across all realizations.

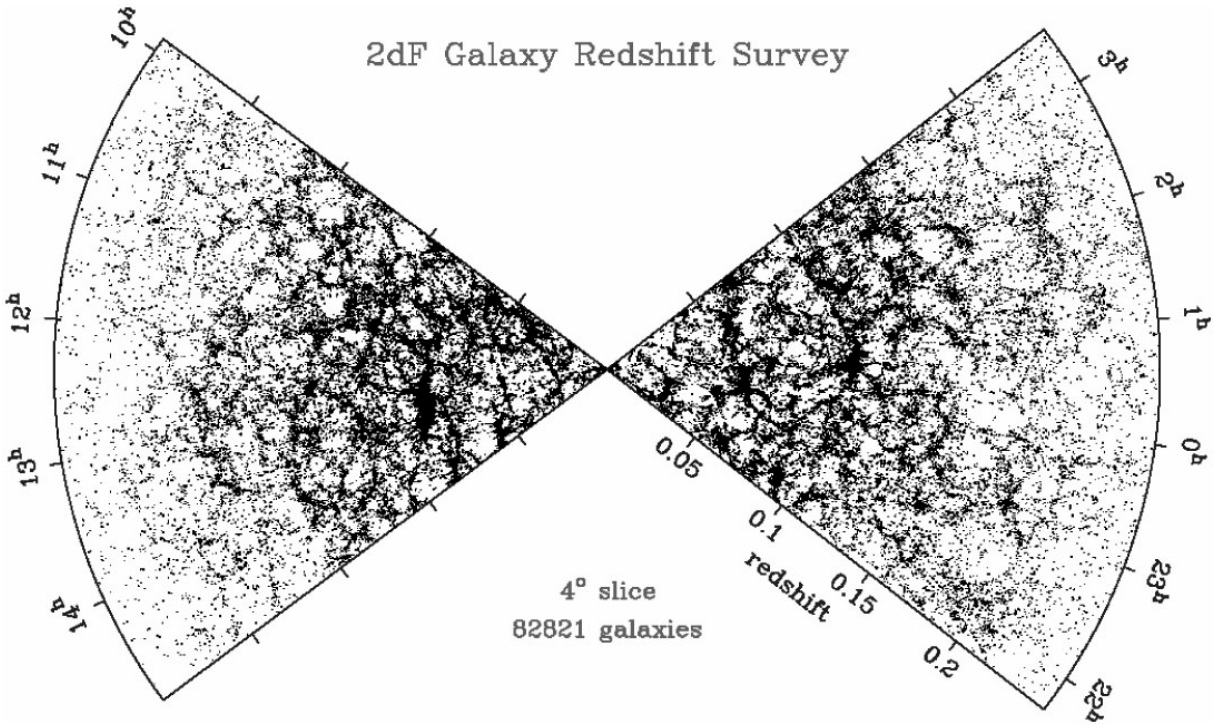


Figure 2.1: In this picture with data the 2dFGRS survey, each point represents the observed position of a galaxy. Originally from Peacock 2003.

We could define higher-order correlation functions, and any general N-variable random field would be completely described by its first N correlation functions. If the expected value of the field is zero, then $\xi_1 = 0$. If the field is statistically homogeneous (i.e. the random process p is the same for all points) then $\xi_2 = \xi_2(\vec{x}_1 - \vec{x}_2)$: it only depends on the relative position of the points, and is therefore invariant under translations. If it is also isotropic, it depends only on the distance between the points: $\xi_2 = \xi_2(|\vec{x}_1 - \vec{x}_2|) \equiv \xi_2(\Delta x)$.

There is a special class of random fields, called ***Gaussian fields***. These fields are the result of the same Gaussian process everywhere, and therefore homogeneous and isotropic. Their statistics are entirely determined just by their mean and their two-point correlation $\xi(\Delta x) \equiv \xi_2(\Delta x)$.

The cosmological power spectrum

There are good reasons to believe that the initial fluctuations (such as those seen in the CMB) were the result of a Gaussian process in the very early universe, possibly during inflation, so at those early times the overdensity δ was Gaussian distributed. This implies that its statistics were entirely determined by $\xi(\Delta x)$.

However, in the real universe we only have access to one realization of the random field, and therefore we cannot compute an ensemble average. Still, it is reasonable to assume the field is ergodic: points far enough from each other are uncorrelated, so an average over many large cells of volume V is equivalent to the ensemble average.

$$\xi(\Delta x) = \int \delta(\vec{x})\delta(\vec{x} + \Delta\vec{x})d^3x, \tag{2.4}$$

where the direction of $\Delta\vec{x}$ is irrelevant because of isotropy.

We can do a change of basis to express ξ in Fourier space:

$$\xi(r) = \int P(k) e^{ikr} d^3k \tag{2.5}$$

where $P(k)$ is the ***power spectrum*** of δ,

$$P(k) = \frac{1}{(2\pi)^3} \int \xi(r) e^{-ikr} d^3r \tag{2.6}$$

It should be noted that, even if the integrals are computed over the three dimensions of both r-space and k-space, we have already incorporated the assumption of (statistical) anisotropy and thus only the moduli of the positions and wavenumbers are used. From now on, we shall refer to each plane wave with comoving wavenumber k and amplitude $P(k)$ as a "Fourier mode" or "mode" for short.

The shape of the power spectrum determines the dominant scales of the fluctuations: a universe with strong long modes will exhibit a lot of large-scale structure like voids and filaments; if there is short-wavelength power, there will be small-scale fluctuations and it will look "noisy", whereas a suppression of those modes will make it look more "clustered" (see Fig 2.1).

Inflationary theory also gives a prediction for the primordial power spectrum: it is thought to have the shape of a power law:

$$P(k) \propto k^{n_s - 1} \tag{2.7}$$

where n_s is called the ***scalar power index*** and is an important cosmological parameter. Current measurements of the CMB give a value close to, but not quite equal, to 1. The value $n_s = 1$ is called a ***scale-free cosmology***: the early universe would have looked (statistically speaking) exactly the same at all scales, like a fractal.

The usual normalization of the power spectrum is σ_8, the amplitude of density fluctuations on (comoving) scales of $8Mpc/h$.

$$\sigma_8 \equiv P(\frac{h}{8Mpc}) = (\frac{h}{8Mpc})^{n_s - 1} \tag{2.8}$$

2.2 Fluid description

The density fluctuations of the early universe can be thought of as the seeds of all the structure seen at later times. Intuitively, it seems easy to imagine initially over-dense regions accreting mass from their surroundings and thus increasing their density further, while initially under-dense regions are hollowed out and become cosmic voids. However, it is not trivial to formulate an analytical description of this process.

Here we will use a simplified version of classical Lagrangian fluid mechanics, which consists of following each fluid element and ensuring that the balance of mass and momentum is fulfilled at every point and time. (The validity of this approach will be discussed later).

First of all, we write down the continuity equation. It reflects the conservation of mass: the rate of change in the density at some point has to be equal to the inflow or outflow of mass. In proper coordinates, it reads:

$$\frac{\partial}{\partial t} \rho(\vec{r}, t) + \vec{\nabla} \cdot [\rho(\vec{r}, t) \vec{v}(\vec{r}, t)] = 0. \tag{2.9}$$

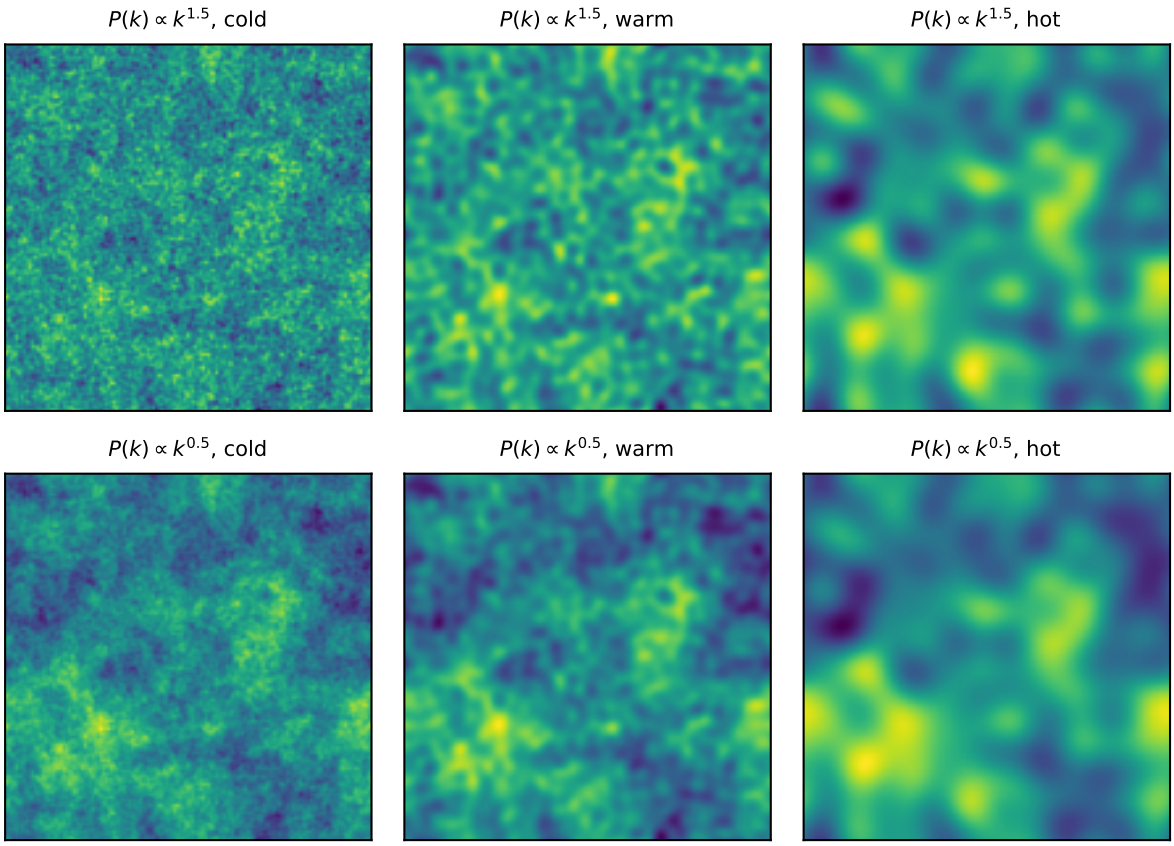


Figure 2.2: Several random fields with the same phase spectrum. Upper row: $P(k) \propto k^{1.5}$, which favours large-scale structure. Lower row: $P(k) \propto k^{0.5}$, which shows more small-scale fluctuations. Left column: an untruncated power spectrum, expected in a CDM scenario. Middle column: power spectrum suppressed on small scales, similar to a warm dark matter case. Right column: suppressed on small and medium scales, as seen with hot dark matter, like a (very massive and abundant) neutrino. All units are arbitrary; the power-law are exaggerated for visual clarity and do not correspond to realistic values of n_s.

Next we write down Euler's equation, which conveys the conservation of linear momentum:

$$\frac{\partial}{\partial t}\vec{v}(\vec{r},t) + \left[\vec{v}(\vec{r},t)\cdot\vec{\nabla}_r\right]\vec{v}(\vec{r},t) = -\frac{1}{\rho(\vec{r},t)}\vec{\nabla}_r p(\vec{r},t) - \vec{\nabla}_r\Phi(\vec{r},t), \tag{2.10}$$

where $\vec{\nabla}_r$ is the gradient with respect to coordinates $\vec{r}$, and Φ is the total gravitational potential.

The third relation is Poisson's equation, which describes the evolution of the potential:

$$\nabla^2\Phi(\vec{r},t) = 4\pi G\rho(\vec{r},t). \tag{2.11}$$

In order to solve the system, it is useful to change from the proper to the comoving coordinates:

$$\dot{\delta} + \frac{1}{a}\vec{\nabla}\left[(1+\delta)\vec{v}_p\right] = 0, \tag{2.12a}$$

$$\frac{\partial\vec{v}_p}{\partial t} + \frac{\dot{a}}{a}\vec{v}_p + \frac{1}{a}\left(\vec{v}_p\cdot\vec{\nabla}\right)\vec{v}_p = -\frac{1}{a\bar{\rho}(1+\delta)}\vec{\nabla}p - \frac{1}{a}\vec{\nabla}\Phi, \tag{2.12b}$$

$$\nabla^2\Phi = 4\pi G\bar{\rho}a^2(1+\delta). \tag{2.12c}$$

The calculation of this transformation is somewhat messy, so it is detailed in appendix B.

2.3 The linear growth of perturbations

As long as $\delta \ll 1$, we can treat cosmic structure as a perturbation on the smooth background ρ of the ΛCDM cosmology, which we covered in section 1.3. We can calculate its evolution within the framework of linear perturbation theory, which means approximating the density and velocity fields, as well as the potential, by a mean value and a small deviation. Hence if we take only the terms that are of first order in the perturbations δ, $\vec{v_p}$ and φ (potential perturbation),

$$\dot{\delta} + \frac{1}{a}\vec{\nabla}\cdot\vec{v_p} = 0, \tag{2.13a}$$

$$\dot{\vec{v_p}} + H\vec{v_p} = -\frac{c_s^2\vec{\nabla}\delta}{a\bar{\rho}} - \frac{1}{a}\vec{\nabla}\varphi, \tag{2.13b}$$

$$\nabla^2\varphi = 4\pi G\bar{\rho}a^2\delta, \tag{2.13c}$$

where the upper dot denotes a time derivative, $\vec{v_p}$ is the peculiar velocity as defined in eq. 1.2, H is the Hubble parameter as defined in eq. 1.3, and $c_s^2 = \partial p/\partial\rho$ is the squared speed of pressure waves in the fluid (for a perfect fluid, $c_s = w$).

If we take the time derivative of (2.13a) and the divergence of (2.13b), substituting (2.13c) we can combine them all into an equation for δ:

$$\ddot{\delta} + 2H\dot{\delta} = \left[\frac{c_s^2\nabla^2}{a^2\bar{\rho}} + 4\pi G\bar{\rho}\right]\delta. \tag{2.14}$$

The equation is linear, so each Fourier mode evolves independently as long as $\delta \ll 1$. The exact evolution of δ will be, in general, scale-dependent: the pressure term will damp the gravitational collapse at small scales. The modes with very long wavelengths are not causally connected, and therefore are not covered by this Newtonian formalism[1], requiring an special treatment in terms of relativistic perturbation theory.

Let us now consider a special case of particular interest. Consider an instant in the matter-dominated epoch, with all the relevant Fourier modes well inside the particle horizon and therefore satisfying (2.14). We can neglect $c_s \simeq 0$, and from (1.20) $H = 2/3t$, $\bar{\rho} = \bar{\rho}_i/t^2$, where $\bar{\rho}_i = \rho_c = 3H^2/8\pi G$ is the average density at our initial instant. Introducing a Fourier mode $D(t)\exp(ikr)$ with a comoving wavenumber k, we get an evolution equation for the amplitude $A(t)$:

$$\ddot{D}(t) + \frac{4}{3t}\dot{D}(t) = \frac{2}{3t^2}D(t) \quad \Rightarrow \quad D(t) = At^{2/3} + Bt^{-1} = Aa + Ba^{-3/2} \tag{2.15}$$

The decreasing mode vanishes over time so the amplitude of the perturbation is $D(t) \propto a$. This means in this epoch the overdensity field grows self-similarly, i.e. all modes grow at the same rate regardless of their scale. D is called the ***linear growth factor***.

[1] These modes are said to be outside the ***particle horizon***: the frontier of the sphere around us containing all the sources whose signals have had enough time to reach an observer (Davis and Lineweaver 2004)

It is customary to write the evolution of the power spectrum as the product of three factors:

$$P(k,t) = P_0(k)T^2(k)D^2(t), \tag{2.16}$$

where $P_0(k)$ is the initial shape of the power spectrum. $T(k)$ is the ***transfer function***, which describes the accumulated change in the amplitude of each mode as it traverses different regimes (pressure-dominated, etc.) until reaching the era of self-similar growth.

Chapter 3

The nonlinear density field

Once the perturbations δ reach order unity, the linear approximation breaks down for two reasons. First, the perturbation equation 2.14 is no longer valid, because the assumption $\delta \ll 1$ does not hold; therefore the Fourier modes start coupling to each other and the growth becomes k-dependent. Second, the overdensity field deviates significantly from its initially Gaussian statistics. The linear growth equations predict that the probability distribution $p_{\bar{x}}(\delta)$ becomes increasingly wide, to the point where the amplitude of some underdensities grows to a value of $\delta < 1$. That is unphysical because it would imply that these regions have a negative total density.

When the linear approximation breaks down, the overdensity field enters a phase of strongly scale-dependent, nonlinear growth. In the ΛCDM cosmology this happens during the matter-dominated phase, before the universe becomes Λ-dominated.

In this nonlinear regime, the evolution of the density field cannot be treated exactly, and the problem of structure formation must be tackled in an approximate way. There are several ways to do this: higher-order perturbation theory, the spherical collapse model, the Zel'dovich approximation, and numerical simulations (Mo, van den Bosch, and S. White 2010). The rest of this chapter covers the latter two.

3.1 The Zel'dovich approximation

In the regime where the overdensity field grows self-similarly, the linear evolution takes the form

$$\delta(\vec{x}, t) = D(a(t))\delta_i(\vec{x}), \tag{3.1}$$

where the subindex $_i$ denotes the value of a function at an initial time, and a and D are normalized so that $a_i = D(a_i) = 1$.

Since the linear growth factor satisfies the perturbation equation (2.14), the Poisson equation gives us the evolution of the gravitational potential Φ. Taking into account that $\bar{\rho} = \bar{\rho}_i a^{-3}$,

$$\begin{cases} \nabla^2\varphi = 4\pi G\bar{\rho}a^2\delta = 4\pi G\frac{\bar{\rho}_i}{a^3}a^2 D\delta_i = 4\pi G\bar{\rho}_i\frac{D}{a}\delta_i \\ \nabla^2\varphi_i = 4\pi G\frac{\bar{\rho}_i}{a_i}a_i^2\delta_i = 4\pi G\bar{\rho}_i\delta_i \end{cases} \Rightarrow \nabla^2\Phi = \frac{D}{a}\nabla^2\Phi_i \Rightarrow \Phi = \frac{D}{a}\Phi_i. \tag{3.2}$$

We can now integrate the linearized Euler equation:

$$\dot{\vec{v}}_p + H\vec{v}_p = -\frac{1}{a}\vec{\nabla}\varphi \Rightarrow \frac{d(a\vec{v}_p)}{dt} = -\vec{\nabla}\varphi \Rightarrow \vec{v}_p = -\frac{1}{a}\int \vec{\nabla}\varphi dt$$

$$\Rightarrow \vec{v}_p = -\frac{\vec{\nabla}\varphi_i}{a}\int \frac{D}{a}dt. \quad (3.3)$$

The integral term can be written explicitly, using again the fact that D satisfies the perturbation equation.

$$\frac{D}{a} = \frac{1}{4\pi G\bar{\rho}}\left[\frac{\ddot{D}}{a} + \frac{2\dot{a}\dot{D}}{a^2}\right] = \frac{1}{4\pi G\bar{\rho}_i}\left[a^2\ddot{D} + 2a\dot{a}\dot{D}\right] \Rightarrow \int \frac{D}{a} = \frac{\dot{D}}{4\pi G\bar{\rho}a}. \quad (3.4)$$

Combining eqs. (3.3) and (3.4) and integrating again with respect to time, the displacement of a particle initially at comoving position $\vec{x}_i$ is described by

$$\dot{\vec{x}} = \frac{1}{a}\vec{v}_p = \frac{\dot{D}}{4\pi G\bar{\rho}a^2}\vec{\nabla}\varphi_i(\vec{x}) \quad \Rightarrow \quad \vec{x} = \int \frac{\dot{D}}{4\pi G\bar{\rho}a^3}\varphi_i(\vec{x})dt. \quad (3.5)$$

Keeping only the first-order terms of the potential, $\varphi_i(\vec{x}) \simeq \varphi_i(\vec{x}_i)$,

$$\vec{x} = \vec{x}_i - \frac{D}{4\pi G\bar{\rho}a^3}\vec{\nabla}\varphi_i(\vec{x}_i). \quad (3.6)$$

The extrapolation in (3.6) is known as the ***Zel'dovich approximation***. In this framework the particles move in a straight line following their initial peculiar velocity. Even though this might sound simple, it is actually a powerful method for evolving perturbations in the $\delta \sim 1$ regime. In one dimension it is indeed exact, because higher-order effects cannot possibly change the direction of a particle.

Now, conservation of mass requires that

$$\bar{\rho}(1+\delta)d^3\vec{x} = \bar{\rho}_i(1+\delta_i)d^3\vec{x}_i \quad \Rightarrow \quad \frac{1}{a^3}(1+\delta)d^3\vec{x} = (1+\delta_i)d^3\vec{x}_i. \quad (3.7)$$

We can neglect $\delta_i \ll \delta$ and then $a^3/(1+\delta)$ is just the determinant of the matrix

$$\left[\frac{d^3\vec{x}}{d^3\vec{x}_i}\right] = \delta_{jk} - \frac{D}{4\pi G\bar{\rho}a^3}\frac{\partial\vec{\nabla}\varphi_i}{d\vec{x}} = \delta_{jk} - \frac{D}{4\pi G\bar{\rho}a^3}\mathcal{H}(\varphi_i), \quad (3.8)$$

where δ_{jk} is the Kronecker delta and $\mathcal{H}$ is the Hessian matrix, whose elements are the second derivatives.

The matrix (3.8) has three eigenvalues $\lambda_1(\vec{x}) \geq \lambda_2(\vec{x}) \geq \lambda_3(\vec{x})$, such that at each point the overdensity field evolves as

$$1 + \delta(\vec{x}, t) = \frac{1}{(1-\lambda_1(\vec{x})D(t))(1-\lambda_2(\vec{x})D(t))(1-\lambda_3(\vec{x})D(t))}. \quad (3.9)$$

This equation gives a good description of the non-linear evolution of an overdensity once it becomes dense enough to counteract the Hubble expansion and gravitational collapse ensues. The distribution of matter first collapses along the direction of the eigenvector associated to λ_1, forming a flat structure called a ***sheet*** or ***pancake***. Then it collapses in the second direction, forming a filament. Finally, when it collapses along the third axis, it forms a compact structure called a ***halo***.

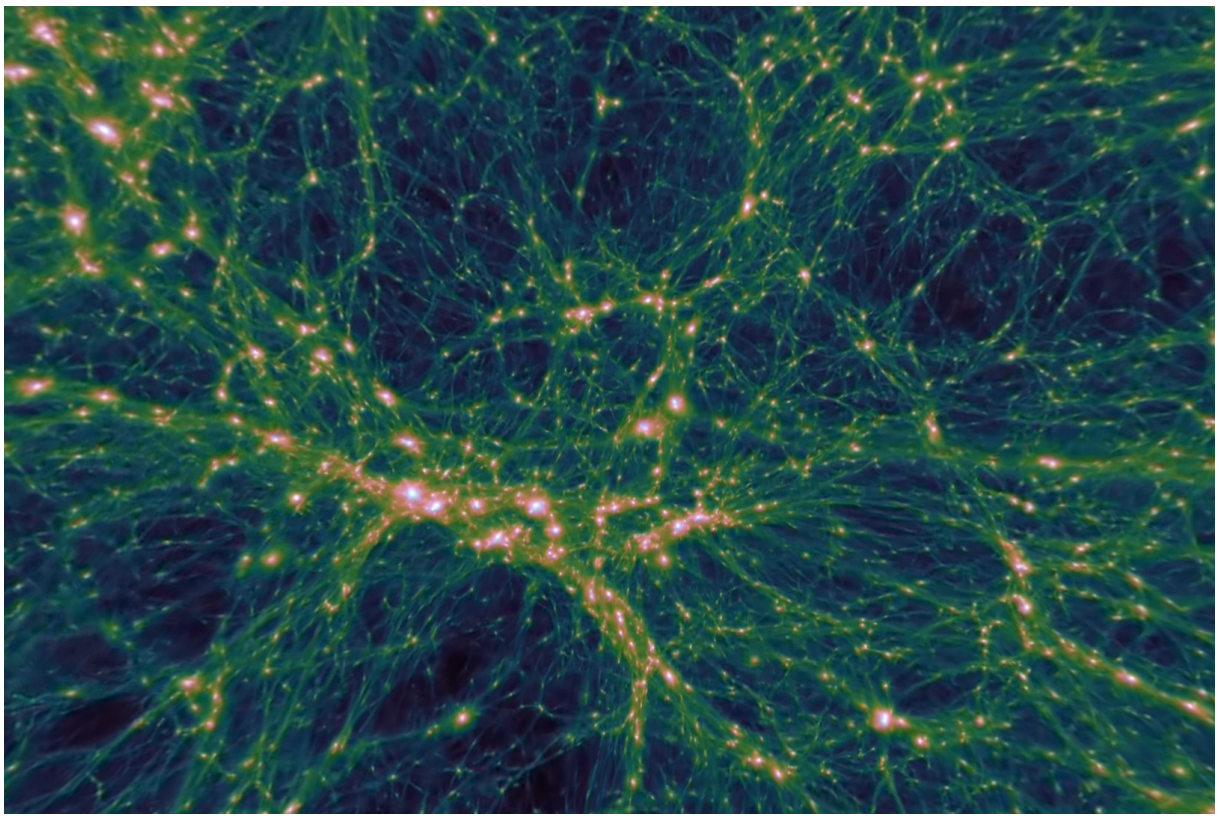


Figure 3.1: This still from a video made by Jens Stücker visualizes the result of a N-body simulation. The fully non-linear power spectrum is visible here, with small high-density halos (in white) at the intersection of medium-density filaments (in yellowish green). A Zel'dovich pancake can be seen near the centre, surrounded by cosmic voids. The full VR video can be accessed at https://www.youtube.com/watch?v=uMBvgCYiUiI

3.2 N-body simulations

The Zel'dovich approximation is only valid into the mildly non-linear regime, $\delta \sim 1$, and breaks down when the real particle trajectories start to deviate significantly from their original direction.

The most common way to follow the dynamics of gravitational collapse into the fully non-linear regime are N-body simulations (3.2). These simulations work by leveraging the great numerical power of modern computers to follow the flow of matter and its gravitational interactions.

Any cosmological N-body simulation works by making some (hopefully reasonable) simplifications of the dynamics it attempts to recreate. The most important elements that are needed are: a discretization of the density field, a set of initial conditions, and equations of motion to evolve them.

Discretizing the density field

Physical spacetime is, at least at macroscopic scale, a continuous canvas through which particles move. Computers, however, have limited storage and processing power, and therefore discretization is needed. This means slicing time into discrete steps Δt, and doing the same to either space or its contents.

In a N-body simulation, space is kept continuous[1], and its contents are divided into N finite-size particles.

In cosmological simulations involving cold dark matter, each of these particles typically has a mass $m_p = 10^6 \sim 10^{11} M_\odot/h$. These particles are a finite element version of the infinitesimal fluid elements that we used in the Lagrangian approach of section 2.2, and each can be thought as representing a set of dark matter particles that move roughly together. The smaller the particle mass, the better ***mass resolution*** the simulation is said to have.

Since the volume of space that can be simulated in a computer is only a finite portion of the entire universe, a cube-shaped box is usually used. The boundary conditions are periodic (i.e. matter that leaves the box comes in on the opposite side) and the length of the box's sides is typically $L = 100 \sim 1000 Mpc/h$. The larger L, the better ***spatial resolution***.

Setting up the initial conditions

As seen in section 2.3, density fluctuations in the early universe evolved through different stages until the universe reached the matter-dominated regime. After that, gravitational interactions became the only ones relevant to the formation of large-scale structure.

In order to keep the computational cost, simulations start at a redshift late enough that it fulfills two conditions. First, non-gravitational interactions can be neglected, at least at scales larger than galaxies. Second, gravity can be formulated in purely Newtonian terms.

As a consequence, setting up the initial conditions requires calculating the evolution of the density and velocity fields up until the starting redshift. In other words, the transfer function and linear growth rate have to be calculated, either analytically or numerically, taking into account general-relativistic effects.

The most common way to do this, shown schematically in figure 3.2, is applying the Zel'dovich approximation to an initial density field calculated from a theoretical model of the early universe.

Implementing gravity

If we start our simulation at a low enough redshift that gravity can be considered Newtonian, the mutual force between two particles will follow the classical inverse-square law

$$\vec{F} = -G\frac{m_p^2}{r^2}\hat{r}, \tag{3.10}$$

where $\vec{F}$ is the force between the two particles, r is the distance between them, and $\hat{r}$ the direction of the mutual force.

However, in this formulation particles can be subject to very high accelerations when passing very close to each other. This, combined with the discrete time steps, can lead

[1] Floating-point arithmetic imposes a limit on this continuity, because coordinates can only be stored in computer memory to a certain level of precision. This is however a very weak limitation compared to the ones imposed by discrete time-steps.

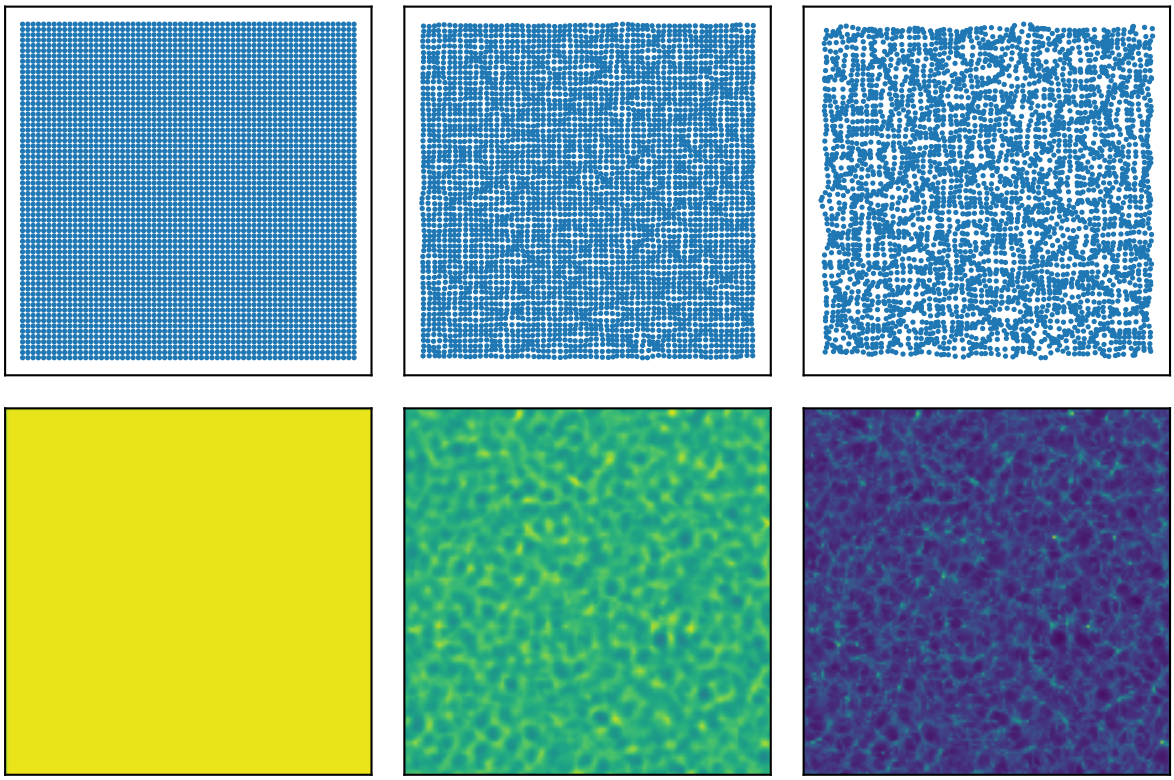


Figure 3.2: This slice of a 3-dimensional box shows schematically how the Zel'dovich approximation is used to set up the initial conditions. Upper row: the positions of the particles. Lower row: the interpolated density field (yellow is denser, blue is less dense). The first column shows the homogeneous initial conditions. The second column shows the situation at a short time after the system starts evolving. In the last column, the particles have already formed a web of filaments and halos, and the simulation is ready to start. This pattern looks similar to the luminous caustics at the bottom of a swimming pool, and indeed it is mathematically equivalent.

to inelastic two-body scattering. Such inelasticity is unphysical, because cold dark matter being collisionless and dissipationless (as seen in section 1.3) its interactions must necessarily conserve mechanical energy.

The standard workaround is introducing a ***softening length*** ϵ in eq. (3.10), such that the mutual force converges to a finite value when particles get very close to each other:

$$\vec{F} = -G\frac{m_p^2}{r^2+\epsilon^2}\hat{r} \quad \Rightarrow \quad \lim_{r\to 0} F = -G\frac{m_p^2}{\epsilon^2}. \tag{3.11}$$

This solves the problem of inelastic scattering, but also introduces potential numerical effects on the mass distribution at small scales. Therefore simulations with smaller ϵ are said to have better ***force resolution.***

There are three main ways to compute the total force acting on each particle:

- Particle-Particle (PP): every 2-body interaction is explicitly calculated using (3.11). This can be computationally costing, because it scales as N^2.
- Particle-Mesh (PM): an M-cell discretized potential field is created by summing over all particles, which scales as $M^3 \log M^3$. Then, the force acting on each parti-

cle is calculated by interpolating the local gradient of the potential. This scales as N, at the cost of being less accurate, particularly at small scales.

- Particle-Particle-Particle-Mesh (P3M): this hybrid approach consists of using the PP method for the force exerted by nearby particles, and the PM method for the effect of the rest of the mass present in the simulation. The threshold for using one method or the other determines the trade-off between numerical accuracy and computational efficiency.

3.3 Properties of halos

The NFW profile

Let $\rho(r, \theta, \phi)$ be the local density field of a halo, with $r = 0$ corresponding to the point with the lowest gravitational potential. This is in general a function that varies strongly from halo to halo, since the shapes can vary significantly from a spherical distribution. By integrating the density over the angular coordinates, we can define a spherically-averaged ***radial density profile***:

$$\rho(r) = \int \rho(r, \theta, \phi) d\Omega = \int_0^{2\pi} d\phi \int_{-\pi/2}^{\pi/2} d\theta \sin\theta \rho(r, \theta, \phi). \tag{3.12}$$

It is not obvious whether halos should share a similar profile or not. However, it was noted in an influential 1996 paper (Navarro, Frenk, and S. White 1996) that most halos in N-body simulations could be fit remarkably well to a two-parameter function, now called the ***NFW profile***:

$$\rho_{\mathrm{NFW}}(r) = \frac{\rho_0}{x\,(1+x)^2}, \qquad x \equiv \frac{r}{r_s}, \tag{3.13}$$

where ρ_0 is a characteristic density and r_s is called the scale radius.

The profile is a double power-law: it approaches the centre as $\propto r^{-1}$ and infinity as $\propto r^{-3}$, so the logarithmic slope $d\ln(\rho)/d\ln(r)$ is -1 at the inner region of the halo and -3 at the outer parts. r_s marks the point at which the logarithmic slope is exactly -2, and thus it is often denoted by r_{-2} as well.

A crucial parameter calculated from the NFW profile is the ***concentration*** of a halo:

$$c \equiv \frac{r_{200c}}{r_s}, \tag{3.14}$$

where r_{200c} is the radius at which the density decays to 200 times the critical density of the universe $\rho_c(z)$, an approximate threshold for virialization[2]. A low concentration means the density of the halo is more uniform, while a high concentration implies there is a dense central region and a much larger outer part of lower density.

[2] A many-particle system is said to have "virialized" or reached "virial equilibrium" when its total kinetic and potential energies stay constant. The ratio between both is related only to the nature of the interactions, not to the initial conditions

The characteristic density is related to the virial mass M_{200c} by

$$M_{200c} = \int_0^{r_{200c}} \rho_{\mathrm{NFW}}(r) 4\pi r^2 dr = 4\pi r_s^3 \rho_0 \int_0^c \frac{x}{(1+x)^2} dx = 4\pi r_s^3 \rho_0 \left[\frac{1}{x+1} + \ln(1+x) \right]_0^c$$
$$= 4\pi r_s^3 \rho_0 \left[\ln(1+c) - \frac{c}{1+c} \right] \quad \Rightarrow \quad \rho_0 = \frac{M_{200c}}{4\pi r_s^3} \frac{1}{g(c)}, \quad (3.15)$$

where $g(c) = \ln(1+c) - c/(1+c)$. It is also equal to $\rho(r_s)/4$.

Another useful function is the ***cumulative density*** $M(r)$, i.e. the mass enclosed inside a sphere of radius r concentric with the halo. For the NFW profile it reads

$$M(r) = 4\pi \rho_0 r_s^3 \left[\ln(1 + x/c) - \frac{x/c}{1 + x/c} \right] = M_{200c} \frac{g(x/c)}{g(c)}. \quad (3.16)$$

The c(M,z) relation

The relation between the concentration of halos and their masses and redshifts is generally difficult to predict, with a wide scatter in the distribution (Neto, Gao, Bett, et al. 2007).

In general, all models predict concentrations to increase with time (i.e. at lower redshifts) and decrease with halo mass. Qualitatively, the main reason can be explained in terms of bottom-up accretion of mass and the critical density ρ_c and its redshift-dependence (Correa, Wyithe, Schaye, et al. 2015).

At fixed mass, halos that had said mass at earlier times (i.e. higher redshift) accreted their outer shells when the average density of the universe was higher. Therefore there is more mass in regions far from the halo centre, resulting in a lower concentration.

At fixed redshift, the argument is similar: less massive halos are younger and therefore have their outer shells have been accreted in a low-density epoch. As a result, a higher fraction of mass is in the halo centre, and the concentration is higher.

In addition, the boundary defined by $200\rho_c$ expands outwards as the critical density of the universe decreases. This brings lower-density regions into the halo without large changes to the inner density profile. This effect causes the concentration to increase with time in a ***pseudo-evolution*** process, which partially offsets the previously explained dependences.

The halo mass function $n(M)$ is also expected to have a downward slope, which means that for any two masses, halos with the lower mass will be more abundant. Qualitatively speaking, this comes from the fact that the initial fluctuations were Gaussian-distributed, and therefore the large overdensities that seeded the heaviest halos were comparatively rarer.

Models of concentration

There are several models that attempt to predict the median concentration for a given mass, redshift and set of cosmological parameters; three of them are presented below. These models work by devising a formula that quantitatively parametrizes the relation between c and ρ_c, and then fitting the free parameters using numerical simulations. The influence of the cosmological parameters is included by making the appropriate changes to the formulae that are used, e.g. the definition of ρ_c.

Prada et al. (2012)

This model, presented in (Prada, Klypin, Cuesta, et al. 2012), does not use M and z directly, and instead uses $\sigma(M, z)$, the RMS fluctuation on scales[3] of M, and a new variable $x \equiv (\Omega_{\Lambda,0}/\Omega M, 0)^{1/3} a(z)$. The analytic expression is

$$c = B_0(x) A \cdot \left[\left(\frac{\sigma'}{b} \right)^c + 1 \right] \exp\left(\frac{d}{\sigma'^2} \right), \sigma' = B_1(x)\sigma(M, x), \tag{3.17}$$

where A, b, c, d are constants and

$$B_0(x) = \frac{c_{\text{min}}(x)}{c_{\text{min}}(1.393)}, c_{\text{min}}(x) \equiv c_0 + (c_1 - c_0) \left[\frac{1}{pi} \arctan(\alpha(x - x_0)) + \frac{1}{2} \right], \tag{3.18a}$$

$$B_1(x) = \frac{\sigma^{-1}_{\text{min}}(x)}{\sigma^{-1}_{\text{min}}(1.393)}, \sigma^{-1}_{\text{min}}(x) \equiv \sigma_0^{-1} + (\sigma_1^{-1} - \sigma_0^{-1}) \left[\frac{1}{pi} \arctan(\beta(x - x_1)) + \frac{1}{2} \right]. \tag{3.18b}$$

The are a total of 12 fitted parameters:

$$A = 2.881, \quad b = 1.257, \quad c = 1.022, \quad d = 0.60, \quad c_0 = 3.681, \quad c_1 = 5.033$$
$$\alpha = 6.984, \quad x_0 = 0.424, \quad \sigma_0^{-1} = 1.047, \quad \sigma_1^{-1} = 1.646, \quad \beta = 7.386, \quad x_1 = 0.526 \tag{3.19}$$

Ludlow et al. (2016)

The model presented in (A. Ludlow, Bose, Angulo, et al. 2016), takes into account the mass assembly history (MAH) of halos and models the concentration as a double power law of the peak height ν, which reflects the amplitude of the initial density fluctuation from with each halo formed:

$$c = c_0 \left(\frac{\nu}{\nu_0} \right)^{-\gamma_1} \left[1 + \left(\frac{\nu}{\nu_0} \right)^{1/\beta} \right]^{-\beta(\gamma_2 - \gamma_1)}, \quad \nu \equiv \frac{\delta_c}{\sigma(M, z)}, \tag{3.20}$$

where c_0 and ν_0 are normalization parameters; γ_1 and γ_2 are the asymptotic power law indices; and β determines the width of the intermediate region. All of these are dependent on the scale factor $a = 1/(1+z)$ through linear (and in one case, polynomial) fitted relations with a total of 13 parameters:

$$c_0 = 3.395 a^{0.215}, \qquad \nu_0 = (4.135 - 0.564 a^{-1} - 0.210 a^{-2} + 0.0557 a^{-3} - 0.00348 a^{-4}) D(z)$$
$$\gamma_1 = 0.628 a^{0.047}, \qquad \gamma_2 = 0.317 a^{0.893}, \qquad \beta = 0.307 a^{-0.54} \tag{3.21}$$

Diemer et al. (2018)

The most recent of the three models (Diemer and Joyce 2019), its analytical derivation explicitly takes into the account the pseudo-evolution of halos. It uses three variables:

[3] "Scales of M" refers to a radius $R(M)$ such that a sphere of said radius and homogeneous density ρ_c encloses a mass M.

the peak height ν, as well as the effective slope of the power spectrum n_{eff} and the effective exponent of linear growth α_{eff}:

$$n_{\text{eff}} \equiv -2\frac{d\ln\sigma(R)}{d\ln R} - 3 \qquad \alpha_{\text{eff}} \equiv -\frac{d\ln D(z)}{d\ln(1+z)} \tag{3.22}$$

where δ_c is the critical density contrast; σ is the mass fluctuation in spheres of a radius such that the enclosed mass, at the mean density ρ_m, is M, and D is the linear growth rate.

Here ν is a proxy for halo mass, n_{eff} aims to capture the influence of the power spectrum slope on concentration, and α_{eff} accounts for the effects of the expansion history. In addition to κ, the model uses another five fitted parameters (a_0, a_1, b_0, b_1, c_α).

$$c = \left(1 - c_\alpha\left(1 - \alpha_{\text{eff}}\right)\right) \cdot \tilde{G}\left[\frac{a_0\left(1 + a_1\left(n_{\text{eff}} + 3\right)\right)}{\nu} \cdot \left(1 + \frac{\nu^2}{b_0\left(1 + b_1\left(n_{\text{eff}} + 3\right)\right)}\right)\right] \tag{3.23}$$

where $\tilde{G}$ is the inverse function of

$$G(x) = \frac{x}{\left[g(x)\right]^{\frac{5+n_{\text{eff}}}{6}}} \tag{3.24}$$

and $g(x)$ is a function related to the mass profile as defined in (3.15).

Chapter 4

Testing a suite of simulations

One of the limitations of the existing models for the $c(M, z)$ relation is that they are only valid in the range of cosmologies where they were calibrated. These ranges tend to be rather limited, covering only a small bracket around the most likely value of each cosmological parameter.

As a way to improve upon this, one of the current aims of the Computational Cosmology is to develop a model that is valid in a much wider array of cosmologies. The expected way to do this is either by developing a new parametric formula or by refitting the parameters of one of the existing models.

As a first step, a suite of simulations was run by Sergio Contreras in the Atlas cluster of the DIPC Computer Centre, described in section 4.1. The practical part of the project consisted of performing several tests to study the robustness of the results and account for numerical artifacts. These tests and their results are covered in the subsequent sections of this chapter.

4.1 The `Target` simulations and `bacco` library

The `Target` simulations are a suite of 69 different runs of the `GADGET-2` N-body code developed by Volker Springel et al. This code also implements a ***friends-of-friends*** (FoF) algorithm[1] to identify halos.

The simulations cover a total of 33 cosmologies: a "main" cosmology plus four variations of each of eight different cosmological parameters. For each cosmology, two independent runs were computed with different realizations of the initial random field. The values of these parameters are quoted in table 4.1.

Additionally, three extra `ResolutionTest` simulations were run in the standard cosmology but with different resolutions. Table 4.2 shows the numerical parameters of the standard simulations, as well as those of the additional lower-resolution runs.

The analysis of the simulation data was primarily done using `bacco` (Bias And Clustering Computation, Optimized), a software library written mainly in Python and in development from 2016 onwards by Raúl Angulo et al. with funding from the ERC. It implements several classes[2] that extract information from a N-body simulation for inter-

[1] A FoF algorithm works by iteratively finding particles that are close to each other and bundling them into a halo. The process is then repeated for the remaining particles until all halos in the simulation have been identified.

[2] In object-oriented programming languages like Python, a ***class*** is a piece of code that defines a data type, along with ***methods*** to initialize an object of that type and to extract or modify its properties

	Ω_m	Ω_b	σ_8	n_s	h	M_ν	w_0	w_a
Lowest			0.730	0.920				
			0.770	0.940				
Low	0.230	0.040	0.815	0.965			-1.30	-0.30
	0.270	0.045	0.860	0.990			-1.15	-0.15
Standard	0.315	0.050	0.900	1.010	0.60	0.00	-1.00	0.00
High	0.360	0.055			0.65	0.10	-0.85	+0.15
	0.400	0.060			0.70	0.20	-0.70	+0.30
Highest					0.75	0.30		
					0.80	0.40		

Table 4.1: Values of the cosmological parameters that have been used in the simulations. For each simulation one parameter was changed to one of the non-standard values, and all others kept at their standard values. m_ν is the sum of the masses of all neutrino flavours and is in units of eV.

Resolution	L [Mpc/h]	ϵ [kpc/h]	N	$m_p\,[M_\odot/h]$
Very low	133.0	10.4	256^3	$12.27{\cdot}10^9$
Low	133.0	6.9	384^3	$3.64{\cdot}10^9$
High	133.0	5.2	512^3	$1.53{\cdot}10^9$
Very high	133.0	3.5	768^3	$0.45{\cdot}10^9$

Table 4.2: Numerical parameters used for each `ResolutionTest` resolution. All the non-standard resolutions were run with the standard cosmological parameters, as a way to discover potential numerical artifacts. The resolution of the `Target` simulations is roughly similar to the "Low" parameters seen here.

pretation by researchers. The ones that have been used in this project are:

`bacco.Simulation` It holds all the information about one N-body simulation at a certain redshift. Its main attributes are:

- `Simulation.fof` is a dictionary containing all the data about the halos that the FoF algorithm has identifies inside the simulation, such as the number of halos, their positions, and data like their masses, radii and NFW parameters.
- `Simulation.header` is another dictionary containing the numerical parameters of the simulation, such as the softening length and the particle mass.

`bacco.Cosmology` It implements a cosmological model. It provides functions to calculate $\rho_c(z)$ and other constants. It also implements the semi-analytic models, providing the predicted halo concentration for any given halo mass.

`bacco.Visualization` It provides tools for creating visually attractive plots from the positions of particles in the simulation. It has been used for figures 3.2 and 3.2.

4.2 Loading a simulation

The first task had the aim of becoming familiar with the tools provided by the `bacco` environment, as well as forming a preliminary idea of the numerical effects that have to be taken into account. A Jupyter notebook was written that did the following:

- Reading the halo mass and concentration data from the `Simulation.fof` of both runs and combining them.
- Using the `cosmology` class to calculate the theoretical concentrations according to each of the three models.
- Plotting the masses and concentration of the halos, both the measured and theoretical values, using the `matplotlib.pyplot` class.

The $c - M$ relation is plotted in figure 4.2. It shows that the mass function peaks in the $10^{12} \sim 10^{13} M_\odot/h$ range, with more massive halos being less numerous. Af fixed mass, the concentration is roughly log-normally distributed. There is also a sharp cutoff around $10^{12} M_\odot/h$, which can be attributed to insufficient mass resolution. Some halos have $c = 0$, which is `bacco`'s way of indicating that they could not be fitted to a NFW profile.

There are two conclusions we can take from this plot. First, when processing any halo data, we will have to exclude any halos with unphysical (< 1) concentration, as well as halos with masses below $10^{12} M_\odot/h$. Second, we should be careful when studying the most massive halos statistically, because the sample size will be small.

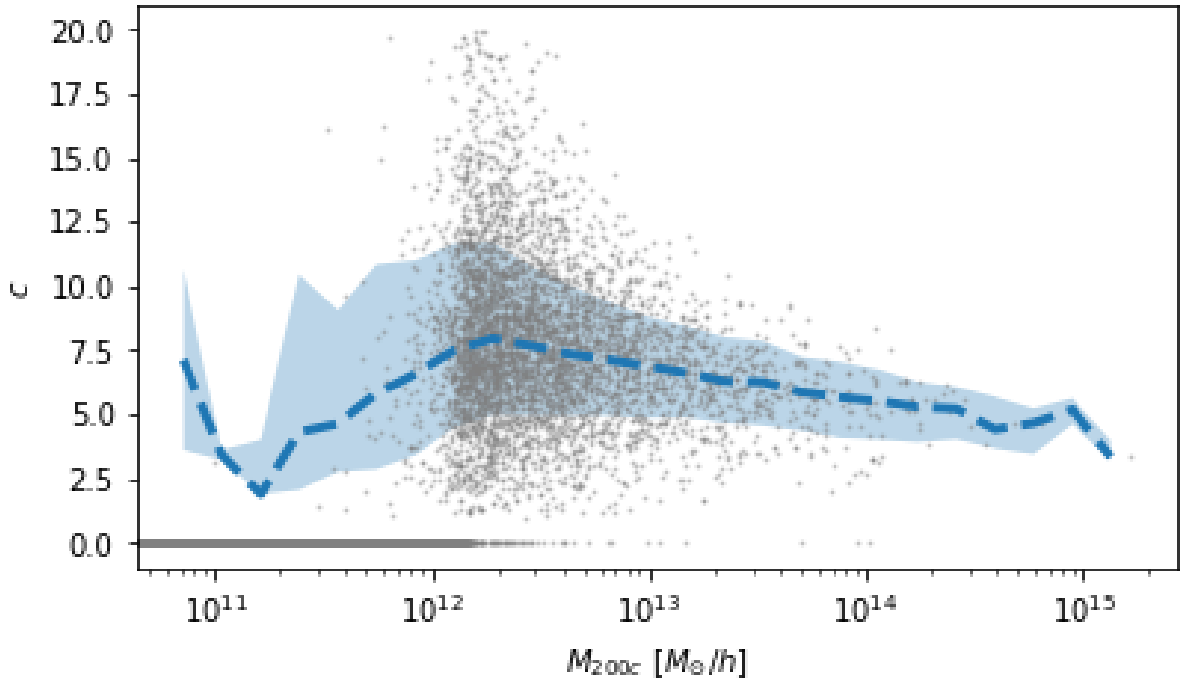


Figure 4.1: The dots are a sample of 10^5 halos chosen randomly from the simulation. The dashed line is the median concentration of the valid halos. The shaded area represents to the 68% scatter interval around the median.

4.3 The c(M,z) relation in 33 cosmologies

The first step to check the robustness of the results is to compare the concentration-mass-redshift relation with the models presented in section 3.3. To accomplish this, a Jupyter notebook was written to implement the following steps for every cosmology and redshift:

- Combine the halos from the two phase-shifted simulations, in order to have a better sample.
- Purge the halos with null concentration, i.e. those that `bacco` couldn't fit to a NFW profile.
- Sort the remaining halos by mass and dividing them into logarithmically-spaced mass bins.
- For each mass bin, compute the median concentration, as well as the 68% spread range.
- Calculate the ratio between the median concentration as measured from the simulation and as predicted by each model.

Figure 4.3 shows a sample of the plots, using two cosmological parameters: σ_8 and Ω_M. High-resolution plots for all eight parameters can be found in Appendix C.

At $z = 0$, the general picture is one of only qualitative agreement with the models. In the mass range of interest, roughly between $10^{12.2} - 10^{14.2} M_\odot/h$, the models agree with the simulation to 20% or better. That is, the for every mass bin in that range, the halo concentration predicted by the models stays between 0.8 and 1.2 times the median c measured from the simulation.

In this case, Prada et al. 2012 seems to provide the best agreement of the three, with the simulation-model ratio staying mostly constant across two orders of magnitude in mass. In contrast, both Ludlow et al. 2016 and Diemer et al. 2018 predict a steeper decrease of c with increasing mass. Ludlow et al. 2016 agrees at the lower mass end and under-predicts concentration for higher masses, while Diemer et al. 2018 over-predicts c at lower masses and agrees with the simulation at the higher end.

At a higher redshift, $z = 1$, the situation is somewhat different. In general, the model/simulation ratio stays mostly flat across mass ranges. The quality of the prediction is therefore mass-independent, but overall worse than at $z = 0$.

For this redshift Diemer et al. 2018 shows the best agreement of the three, especially in the massive neutrino cosmologies. Prada et al. 2012 and Ludlow et al. 2016 fare worse: in some cosmologies they overpredict c by at least 20% relative to the simulation. The fact that the overprediction happens at all masses suggests some kind of systematic effects.

There are two main takeaways from this test. First, the concentration predicted from the models and the simulation data only agree to a certain level, roughly 20%, particularly at higher redshift. Second, the models don't show good agreement with each other either, with more than 25% differences in some cases. As a result, none of the models can be trusted outright, and instead additional testing should be carried out in order to find the origin(s) of the discrepancies.

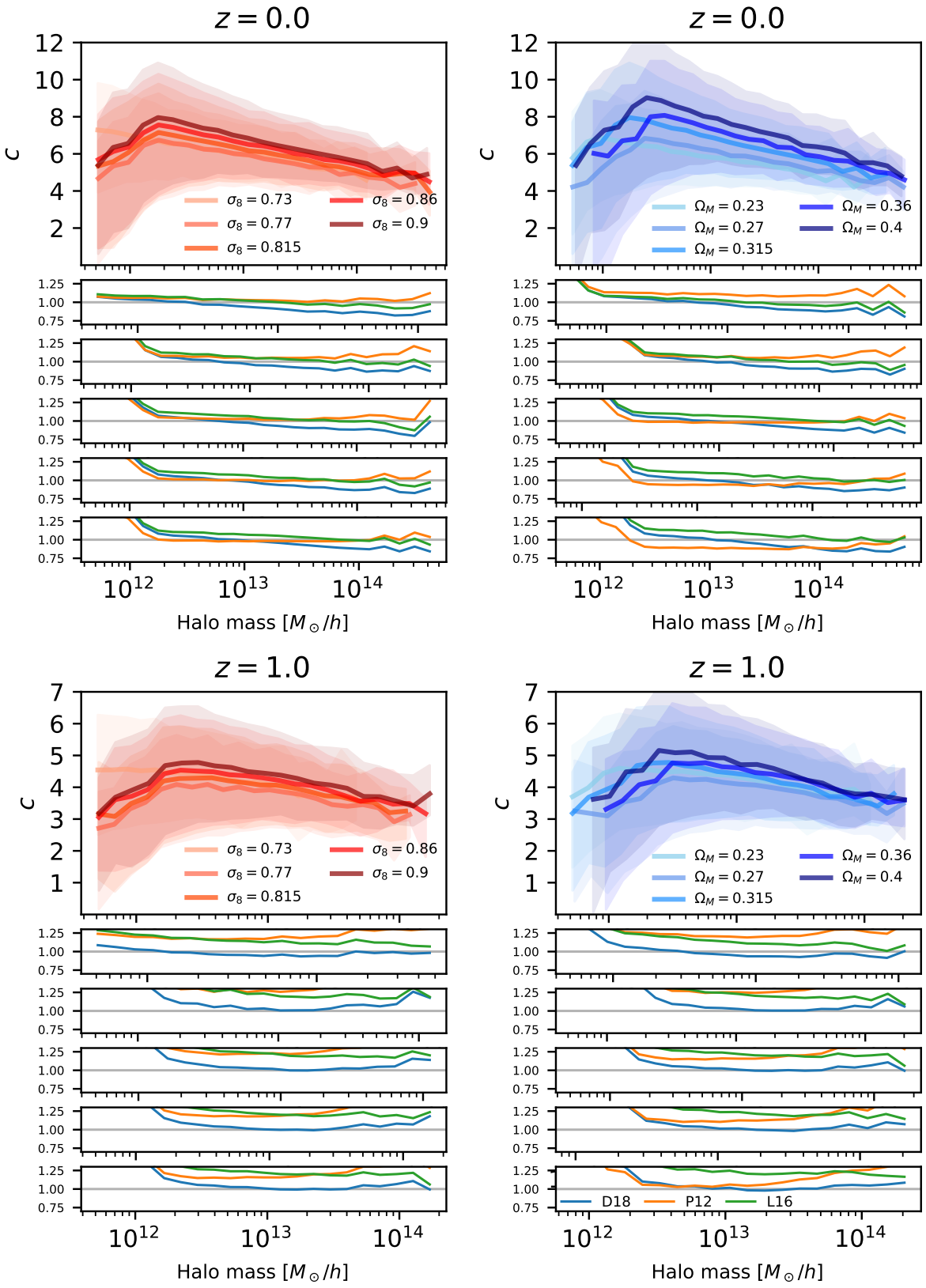


Figure 4.2: The solid lines in the main plots are the $c(M, z)$ relation in all σ_8 and n_s cosmologies. The shaded areas show the corresponding 68% scatter interval. The small panels below each plot show the ratio of the concentration expected from the models relative to the one measured in the simulation. Only halos with at least 1000 particles and mass bins with at least 25 halos have been counted.

4.4 Resolution effects on concentration

In order to study potential resolution effects on concentration, the `ResolutionTest` runs were studied with a procedure similar to that employed in the previous test. Figure 4.4 shows the median concentration and the scatter in each resolution, with each panel displaying a different redshift from $z = 0$ to $z = 3$.

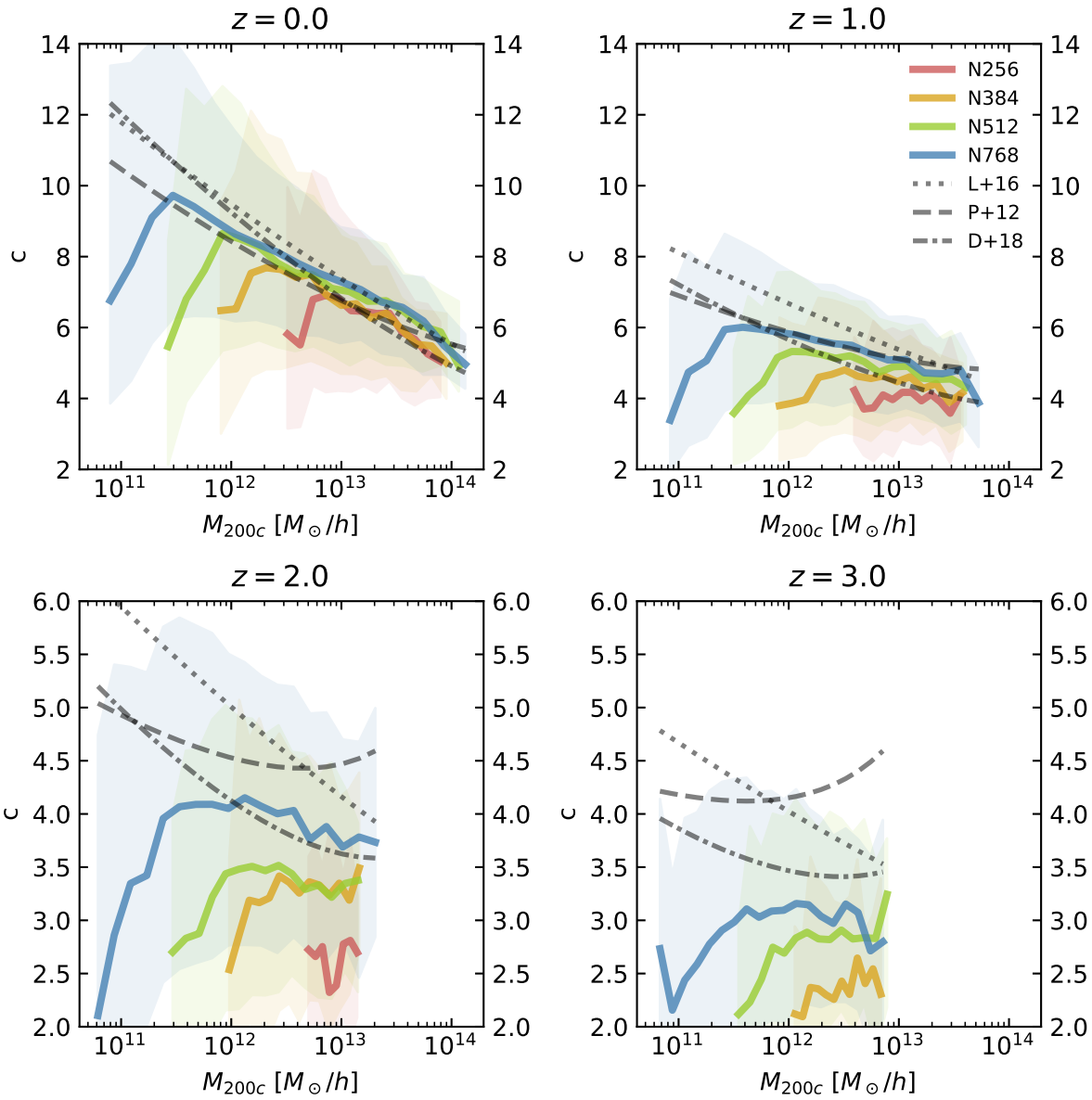


Figure 4.3: The colored lines show the median concentration, while the shaded areas represent the 68% scatter range. The grey lines show the predicted concentrations according to each model. Only mass bins with at least 25 halos are plotted, in order to ensure a reasonably large sample. This leads to the lowest-resolution simulation not appearing at the highest redshift.

It is clear from the figure that the lower-resolution simulations are not robust at high redshift, as they predict substantially lower concentrations than the high-resolution simulation. The four of them only converge to similar results at $z = 0$. In a similar way, agreement with the models is only reached at $z = 1$ or lower.

4.5 Radial density profiles

The other test involves studying the full density profiles instead of just the concentration parameter. The idea is that, if the structure of halos in the simulations does not actually conform to the NFW profile, the fitted parameters c might not be entirely meaningful.

As a way to check this potential issue, a Jupyter notebook has been written that does the following:

- Load the halo data from each simulation.
- Create three mass bins containing halos with masses in of approximately 10^{13}, $10^{13.5}$ and $10^{14} M_{\odot}/h$, respectively.
- For each mass bin, calculate the median of the radial profiles measured from the simulation.
- Plot the ratio of the theoretical profiles (NFW profiles with c predicted by each of the three models) to the median profiles.

The radial density profiles are calculated by dividing the halo into 24 concentric spherical shells, with logarithmically spaced radii. For each shell, the mass of all particles in the shell is summed and then divided by the volume of the shell to obtain the average density of the shell. The density at the inner region of halos scales as r^{-1}, whereas the volume of the shells scales as r^{-3}, resulting in the innermost shells being resolved by a smaller number of particles. Consequently, the cumulative mass function $M(r)$ has been studied instead of the radial density function $\rho(r)$, as a way to avoid statistical noise in small halo samples.

Figure 4.5 shows a selection of the results (the rest are included in appendix D). In all cosmologies at $z = 0$, the NFW cumulative mass profile $M_{\text{NFW}}(r)$ seems to hold to 10% or better, except at the inner part of halos. At $z = 1$, the agreement is similar for the cosmologies closest to the standard value, but only to 20% or better in the most extreme cases.

For halos with small masses ($10^{13} M_{\odot}/h$), 3ϵ seems like a good boundary for the region where the results appear to be robust. Below that radius, the measured mass is smaller than that expected from the models.

At higher masses, particularly the most massive halos ($10^{14} M_{\odot}/h$), 3 times the softening length does not appear to be a sufficient threshold, the region where the measured mass is lower than expected extends to around $20kpc/h$. Since there also seems to be a dependence on resolution, it seems necessary to find an alternative (possibly mass- and resolution-dependent) definition for the region in which we can trust the simulation results.

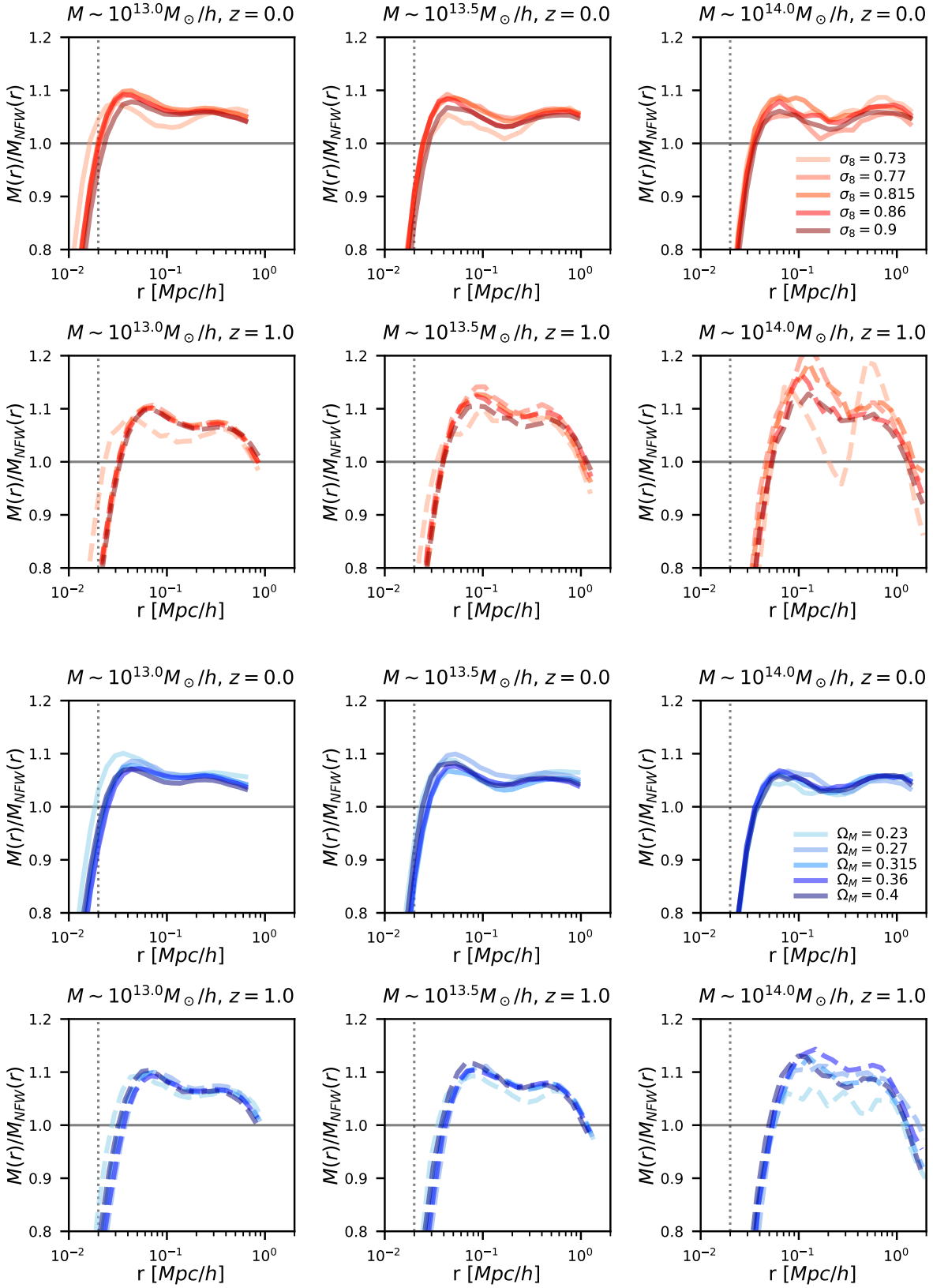


Figure 4.4: Cumulative mass profiles for halos in three mass ranges, relative to the NFW cumulative profile. The mass ranges include all halos within 0.1 dex of the quoted mass. The NFW profile has been calculated by introducing the median concentration of each sample into (3.16). Solid lines are for $z = 0$, dotted lines are for $z = 1$. The vertical line marks $r = 3\epsilon$.

4.6 Force resolution and convergence radius

A recent paper (A. D. Ludlow, Schaye, and Bower 2019) explored the idea of a ***convergence radius*** $r_{\rm conv}$ in simulated halos. This radius is the smallest distance from the centre for which the radial density profile can be considered free of numerical artifacts. It is defined as

$$t_{\rm relax}(r_{\rm conv}) \equiv \kappa t_{200}(z), \tag{4.1}$$

where $t_{\rm relax}(r)$ is the relaxation time at radius r, i.e. the time it takes a particle to "lose memory" of its initial trajectory; κ is a parameter related to the strictness of the convergence criterion; and $t_{200}(z)$ is the period of a circular orbit at the edge of the halo.

For a halo with N particles and radius R_{200c}, the convergence radius $r_{\rm conv}$ is given as a function of κ by the solution to

$$\kappa = \frac{\sqrt{N}}{4}\left[\ln\left(\frac{R_{200c}^2}{\epsilon^2}+1\right)+\frac{\epsilon^2-2R_{200c}^2}{3(\epsilon^2+R^2)}-\ln\frac{2}{2}\right]\sqrt{\frac{200\rho_c}{\bar{\rho}(r_{\rm conv})}}, \tag{4.2}$$

where $\bar{\rho}(r) \equiv M_{200c}/M(r)$ is the cumulative density function of the halo. Thus for each halo $_{\rm conv}$ is fixed by the choice of κ. Due to the functional form of (4.2), a more strict (i.e. higher) choice of κ will lead to a larger convergence radius, meaning a larger part of the halo density profile is deemed unreliable due to lack of relaxation.

In `bacco`, the absolute inner limit for calculating NFW fits is set as 3ϵ. The third round of tests thus aimed to check if $r_{\rm conv}$ imposes a more strict limit: if an unreliable part of the density profile has been included in the fit, it may lead to systematic errors in the calculation of halo concentration.

Figure 4.6 shows, for each smaller resolution, the deviation of the density profiles relative to the highest-resolution simulation.

For all halo mases and resolutions, $r_{\rm conv} > 3\epsilon$. A value of $\kappa = 5$ gives an agreement to within 5% or better, except for the N384 simulation at $z = 1$. With a higher value of κ, we could obtain an even better agreement, at the cost of discarding a larger portion of the density profile.

The most important result of this test is that the region of the halo density profile between 3ϵ and $r_{\rm conv}$ is not numerically robust. As a result, if the simulations are used in the future to calibrate a semi-analytic model, the calculation of concentrations should ignore the part of the halos that is below the convergence radius.

Going forward, this also means that the concentrations that have been used in the previous test have to be recalculated. Since those tests aimed to evaluate the agreement between the simulation and the models, some difference in the result may arise from the new fit, potentially erasing at least part of the disagreement between the models and the simulated data.

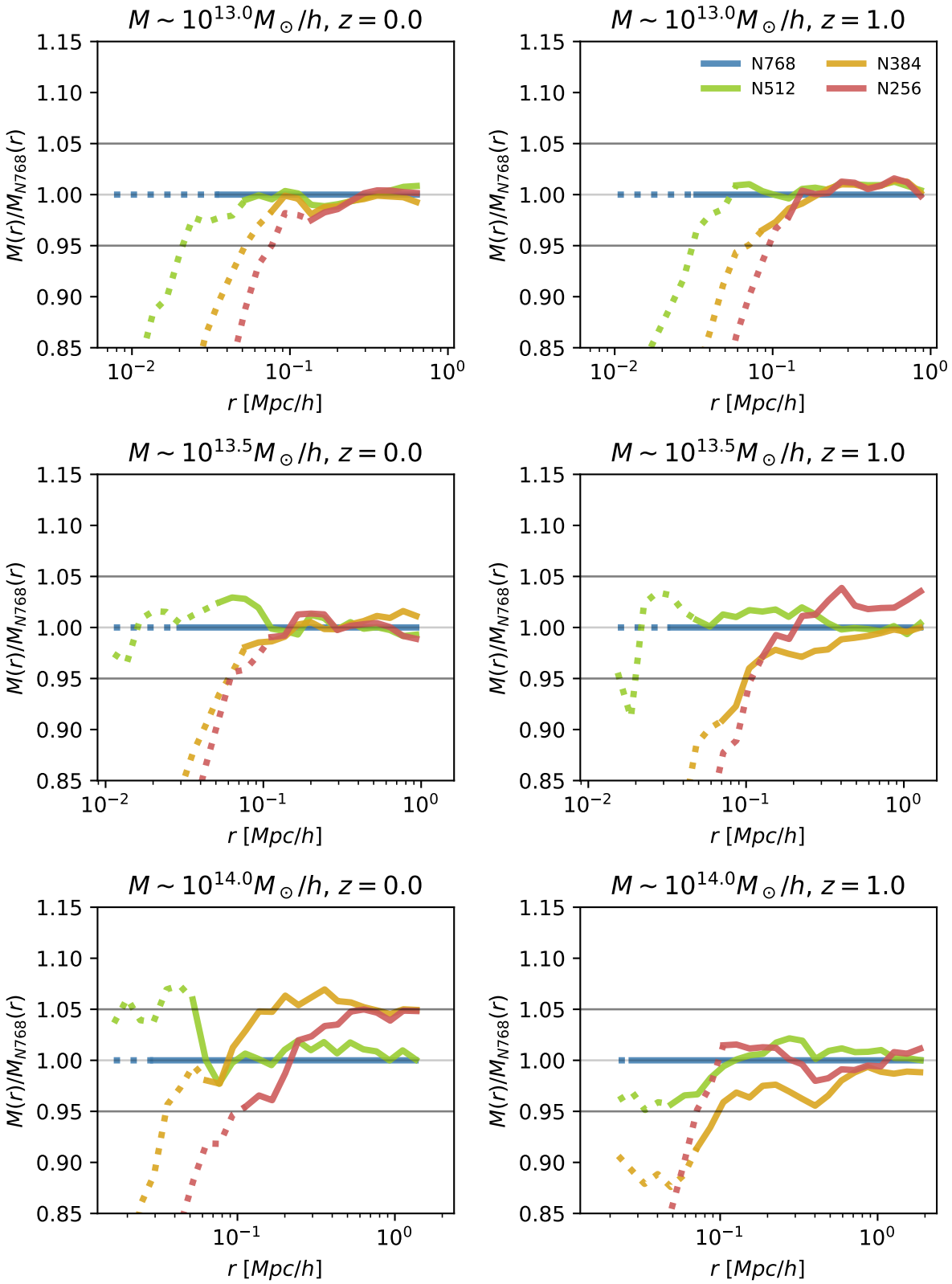


Figure 4.5: Ratio of the median cumulative mass $M(r)$ in each simulation relative to the high-resolution run. The solid segment of the colored lines represents the part at $r > r_{\text{conv}}$ for $\kappa = 5$, while the dotted segment is the part of the profile that falls below the convergence radius. The horizontal bars mark the 10% interval around the high-resolution value. Each panel has been calculated with halos within 0.1 dex of the quoted mass.

Conclusions and acknowledgments

The field of physical cosmology is notably wide and deep. One of its branches, the study of the origin and evolution of large-scale cosmic structure, has been the focus of this project. As an introduction to the research environment, a varied bibliography has been studied, from introductory texts to state-of-the-art articles from academic journals.

In this thesis, we have started by deriving from simple principles the Friedmann equations, which govern the background evolution of the Universe. Second, using Lagrangian perturbation theory, we have described the self-similar growth of the initially small density perturbations. Third, we have studied the Zel'dovich approximation for the mildly nonlinear regime, as well as N-body simulations that attempt to describe how those initial density fluctuations, after their linear growth, turn into large-scale collapsed structures in the fully nonlinear regime. We have also examined the properties of dark matter halos learned from simulations, with particular attention to the NFW density profile, the concentration parameter c and its relation to the halo mass M and redshift z.

In the practical aspect, we have studied the robustness of a suite of N-body simulations, with the aim of eventually being able to use it to calibrate an analytical model of the $c(M, z)$ relation. After several tests, we have concluded that there is a convergence radius r_{conv} which sets a lower limit on the radii at which the density profile of halos is free of numerical artifacts. In the future, the interpretation of data from this suite of simulations should take into account r_{conv} to avoid fitting invalid data and ensure the robustness of the resulting models.

I would like to thank Prof. Raúl Angulo for giving me the opportunity to work in a scientific research environment for six months, which felt much shorter, and for his always on-point suggestions. I would also like to thank the rest of the DIPC Computational Cosmology team: Dr. Sergio Contreras, for his help with setting up the necessary programming tools and for his mentorship and advice throughout my entire stay at the group; to Dr. Matteo Zennaro and (soon-to-be Dr.) Giovanni Aricò, for their eagerness to help with computational or conceptual problems of all kinds; to Dr. Jens Stücker, for kindly providing part of the source code for figures 2.1 and 3.2; and to the rest of the researchers and students, for their good humour and welcoming environment.

Special thanks as well to my advisor at UPV-EHU, Dr. Raül Vera, for his tireless pedagogical dedication and for his thorough but constructive criticism to the draft versions of this thesis. And last but not least, to all my friends and family, for showing interest in the project and carrying me over the finish line.

Bibliography

Carroll, S. (2014). ***Spacetime and Geometry: Pearson New International Edition***. Pearson Education Limited.

Correa, Camila A., J. Stuart B. Wyithe, Joop Schaye, et al. (July 2015). "The accretion history of dark matter haloes – III. A physical model for the concentration–mass relation". In: ***Monthly Notices of the RAS*** 452.2, pp. 1217–1232. DOI: 10.1093/mnras/stv1363.

Davis, Tamara M. and Charles H. Lineweaver (2004). "Expanding Confusion: Common Misconceptions of Cosmological Horizons and the Superluminal Expansion of the Universe". In: ***Publications of the Astronomical Society of Australia*** 21.1, pp. 97–109. DOI: 10.1071/as03040.

Diemer, Benedikt and Michael Joyce (Feb. 2019). "An Accurate Physical Model for Halo Concentrations". In: ***The Astrophysical Journal*** 871.2, 168, p. 168. DOI: 10.3847/1538-4357/aafad6.

Ludlow, Aaron, Sownak Bose, Raúl E. Angulo, et al. (Aug. 2016). "The mass-concentration-redshift relation of cold and warm dark matter haloes". In: ***Monthly Notices of the RAS*** 460.2, pp. 1214–1232. DOI: 10.1093/mnras/stw1046.

Ludlow, Aaron D, Joop Schaye, and Richard Bower (July 2019). "Numerical convergence of simulations of galaxy formation: the abundance and internal structure of cold dark matter haloes". In: ***Monthly Notices of the RAS*** 488.3, pp. 3663–3684. DOI: 10.1093/mnras/stz1821.

Mo, Houjun, Frank C. van den Bosch, and Simon White (2010). ***Galaxy Formation and Evolution***.

Navarro, Julio F., Carlos S. Frenk, and Simon White (May 1996). "The Structure of Cold Dark Matter Halos". In: ***The Astrophysical Journal*** 462, p. 563. DOI: 10.1086/177173.

Neto, Angelo F., Liang Gao, Philip Bett, et al. (Nov. 2007). "The statistics of Λ CDM halo concentrations". In: ***Monthly Notices of the RAS*** 381.4, pp. 1450–1462. DOI: 10.1111/j.1365-2966.2007.12381.x.

Prada, Francisco, Anatoly A. Klypin, Antonio J. Cuesta, et al. (July 2012). "Halo concentrations in the standard Λ cold dark matter cosmology". In: ***Monthly Notices of the RAS*** 423.4, pp. 3018–3030. DOI: 10.1111/j.1365-2966.2012.21007.x.

Springel, Volker, Carlos S. Frenk, and Simon D. M. White (Apr. 2006). "The large-scale structure of the Universe". In: ***Nature*** 440.7088, pp. 1137–1144. DOI: 10.1038/nature04805.

Swart, J. G. de, G. Bertone, and J. van Dongen (Mar. 2017). "How dark matter came to matter". In: ***Nature Astronomy*** 1.3. DOI: 10.1038/s41550-017-0059.

Zennaro, Matteo (Dec. 2017). "Neutrinos and the Large Scale Structure of the Universe". PhD thesis. Universita degli studi di Milano.