THE UNIVERSITY *of* EDINBURGH
School of Physics and Astronomy

The Cosmic Dark Matter Density Field

MPhys Project Report

D. Kniazevas

Abstract

To better understand the structure formation in the Universe, the cosmic dark matter density field of a 1600 Mpc/h side-length N-body simulation with 2048^3 particles, spanning redshifts from $z = 99$ to 0 from the Legacy Project was analysed. The expected departure from the linear perturbation growth theory $\delta \propto a$ was shown to happen at around $a = 0.5$. Additionally, a great match of the data to the structure growth approximation introduced by Carroll et al. (1992) was shown. Lastly, the probabilities of the regions growing close to the Carroll approximation were calculated and presented, and the massive cluster impact on the structure growth was shown.

Supervisor: Professor S. Khochfar

Personal statement

The first three weeks of the project were spent familiarising with the access and usage of the computing cluster cuillin, Python package yt used in the analysis, EMACS, and the Legacy Project data itself. Time was spent learning SLURM manager commands, using them to get access to cuillin nodes, and running test scripts. Additionally, I learnt to use GNU Screen. However, most of the time was devoted to using yt to get familiar with the downsized project's data. Since this was my first time using this package, it took a while to get used to all the data coming in fields and learning how to manipulate it. Lastly, I was doing background reading on cosmological structure evolution in parallel to the coding tasks.

The next few weeks were used on learning how to parallelise the code with yt and how to run MPI jobs on cuillin to start obtaining data. I was struggling with the cuillin side of things for a while and the cuillin wiki was not helpful, but with the help received from Ben Morton I was able to learn how to do it. Another small stop was caused by the unfamiliarity with halo catalogues from the project. I had trouble getting the data I needed out of them at first, but after help from my supervisor and TMoX wiki it became clear. The written code also worked as expected on test data so it was run on the full simulation and started producing the overdensity histogram data.

Even though the first results were as expected, when the analysis shifted to halos, the results produced did not make complete sense. The goal was to produce random walks around halos similar to ones in previous literature, but the ones obtained looked vastly different. After realising that the overdensity was calculated in a wrong way and fixing it, the results started making more sense.

The project worked slowed down during the revision period for December exams and came to a complete halt during the winter break due to issues with cuillin. One of the RAID controllers died in early December and /disk12/ with the Legacy data was offline for over a month until the beginning of second semester. Therefore, no coding work was done during this period and instead the time was spent gathering resources for the background section of the report.

The first half of the second semester was spent running code to obtain the overdensity evolution data. The initial results made no sense as a wrong calculation was used. It took some time of getting my head around the formulas and scaling to produce what was expected, but after discussions with my supervisor, it was achieved. Following that, time was spent visualising and interpreting the data. Additionally, I drafted the background section of the report.

Towards the middle of the semester it became increasingly difficult to run the code on cuillin. It became overcrowded and often overloaded with thousands of jobs in queues. Most of my

submitted jobs spent several days waiting in the queue each time before even launching, severely slowing down any new progress. Therefore, as much as possible was tried to be achieved from the existing data. Using it, the probabilities of predicted structure growth were obtained. Furthermore, the time spent waiting for cuillin was spent working on the project report and producing the plots for it.

Lastly, the end of second semester was mostly spent on writing up the report while also doing some minor coding and analysis of probabilities and halos.

Acknowledgments

Contents

1   Introduction

A century ago the existence of an unknown form of cosmic matter was first postulated (Kapteyn 1922, Lundmark 1930, Oort 1932, Zwicky 1933). Not interacting electromagnetically, and therefore not being able to be "seen", this matter was dubbed as the dark matter. Ever since, continuous research and experiments have been poured into the effort to uncover the nature of this elusive form of matter. However, even to this day it remains one of the most puzzling questions in the current day cosmology and physics.

Though "invisible", dark matter is of great importance to various phenomena observed in the Universe, and the state and shape of the world we see today. Being five times more common than ordinary matter (Planck Collaboration 2020) and able to interact gravitationally, it is deemed responsible for the clustering of galaxies (Allen et al. 2011), the observed galaxy rotation (Corbelli & Salucci 2000), gravitational lensing effects (Natarajan et al. 2017), and the overall structure formation in the Universe (Padmanabhan 1993), among other things. Most importantly, without dark matter, no galaxies would exist as they are today.

The standard cosmological model assumes the Universe was isotropic and highly homogeneous at early times, with minute density perturbation imprinted onto it that arose due to quantum mechanical effects. As the Universe expanded, these perturbations grew by attracting matter due to gravity. Using a simple Newtonian theory, it can be predicted, how such matter seeds grew over time and what they would end up like (Mo et al. 2010). However, by applying this theory to ordinary matter, it becomes clear that the initial perturbations would have not had enough time to evolve into such dense structures like galaxies we see today. The solution to this problem unsurprisingly lies in dark matter.

Ordinary baryonic matter, like protons, neutrons, and electrons, interacts with radiation. Hence, as radiation was the dominant component of the early Universe, the charged baryons were prevented from condensing into structures by being continuously scattered by photons. It was only after around 370,000 years after the Big Bang, during a cosmological epoch known as recombination (Tanabashi et al. 2018), that neutral matter was created and photons decoupled from it, allowing ordinary matter to start condensing. However, dark matter, being unaffected by radiation, was able to collapse around the density perturbations before that. The formed dark matter gravitational potential wells would then start attracting ordinary matter later, speeding up its structure formation and allowing it to catch up.

However, the simple Newtonian theory for perturbation growth has its shortcomings and the observed structure growth swiftly starts deviating from it. Therefore, this project aims to analyse the evolution of the cosmic dark matter density field and its departure from the Newtonian theory while introducing and analysing more sophisticated models and explanations.

In this project, dark matter simulation data from the Legacy Project (Meriot et al. 2022) was used in tracing the density field. The Gaussian properties of this field were analysed and the data was compared to the Newtonian theory and additionally, the approximation to the structure growth introduced by Carroll et al. (1992). Furthermore, probabilities of matter following this Carroll approximation were computed and discussed. Additionally, the structure evolution around the most massive current day dark matter halos was observed and analysed.

2 Background

2.1 ΛCDM Universe

Many different models with different components of the Universe have been theorised and are analysed. One of the most simplest and the most adapted ones this day is known as the ΛCDM Model. This Big Bang cosmological model depicts an universe containing three major components: cosmological constant (Λ), cold dark matter (CDM), and ordinary baryonic matter.

The cosmological constant Λ arising in Einstein's field equations represents the vacuum energy of space and is used as an explanation for dark energy. It describes a constant energy density homogeneously filling all space and therefore increasing with time as the Universe and space expands. The latest measurements (Planck Collaboration 2020) indicate that 68.5% of the total energy of the present-day observable Universe is in the form of dark energy, however the exact nature of this energy remains unknown.

Meanwhile, dark matter is the second most prevalent component in the Universe. With the measured fraction of 26.6% (Planck Collaboration 2020) of the content in the present-day Universe, dark matter is several times more common than ordinary matter (4.9%) and therefore is of significant importance. As it does not interact with electromagnetic fields, it cannot reflect light and is therefore essentially invisible. However, as dark matter is affected by gravity, its existence is implied by various gravitational effects which cannot be explained by other means.

2.2 The Expanding Universe

The first key observation of the modern cosmology was the discovery of the Universe expansion which is credited to Edwin Hubble in 1929. By observing Cepheid variable stars in Andromeda (M31) galaxy, he was able to formulate a linear relationship between object's distance d and recession velocity v, known as Hubble's Law:

$$v = Hd, \tag{2.1}$$

where H is the Hubble parameter which varies with time. At the present time, its value is estimated to be $H_0 \approx 70$ km s^{-1} Mpc^{-1} and is referred to as the Hubble constant[1]. Often it is more convenient to work with a dimensionless Hubble constant which is defined as

$$h \equiv \frac{H_0}{100 \text{ km s}^{-1} \text{ Mpc}^{-1}}. \tag{2.2}$$

This expansion of the Universe is parameterised by a cosmic scale factor $a(t)$, which relates the proper observed coordinates and distances $\mathbf{r}$ to comoving coordinates and distances $\mathbf{x}$, which are perceived by observers moving together with the expansion:

$$\mathbf{r} = a(t)\mathbf{x}. \tag{2.3}$$

The comoving coordinates factor out the expansion of the Universe, giving distances that do not change in time due to the expansion of the space itself, making them a natural choice for analysing structural changes in time that are not solely due to the expansion. Additionally, by differentiating Eq. (2.3) and using Eq. (2.1), an expression for Hubble parameter in terms of the scale factor can be optained:

$$H(t) = \frac{\dot{a}(t)}{a(t)}. \tag{2.4}$$

Moreover, it is often convenient to relate the scale factor to the redshift z as

$$\frac{a(t)}{a_0} = \frac{1}{1+z}, \tag{2.5}$$

where $a_0 = a(t_0)$ is the present day scale factor (common to take $a_0 = 1$) and $z \equiv (\lambda_0 - \lambda_e)/\lambda_e$ is the relative change of emitted and observed wavelenght caused by the Universe expansion.

2.3 Dynamics of the Expansion

The equation of motion (or time evolution) of $a(t)$ can be obtained from Einstein's field equations, and is known as the Friedmann equation:

$$\left(\frac{\dot{a}}{a}\right)^2 = H^2(t) = \frac{8\pi G}{3}\rho - \frac{Kc^2}{a^2}, \tag{2.6}$$

where G is Newton's gravitational constant, c is the speed of light in vacuum, ρ is the volumetric mass density, and K is a constant that can take the values -1, 0, and +1 and determines the curvature of the Universe (with $K = -1$ representing open universe, $K = 0$ being flat, and $K = 1$ closed). Additionally, the density in this equation can be split up

---

[1]1pc (parsec) is equal to 3.26 ly (light years) or 3.0857×10^{16} m.

into a non-relativistic matter component ρ_m, radiation component ρ_r, and a vacuum energy (cosmological constant) component ρ_Λ:

$$\rho = \rho_{\mathrm{m}} + \rho_{\mathrm{r}} + \rho_\Lambda. \tag{2.7}$$

However, these density components scale differently as the Universe expands. The number density of pressureless matter will scale as a^{-3}, and in turn so will the density $\rho_{\mathrm{m}} \propto a^{-3}$. For radiation, the number density of the photons will again scale as a^{-3}, but additionally, their energy will be redshifted by a factor of a^{-1}, giving $\rho_{\mathrm{r}} \propto a^{-4}$. Lastly, the vacuum energy density is constant and does not change with a: $\rho_\Lambda = a^0 = 1$. Using these scalings, and denoting the present time t_0 densities by $\rho_{\mathrm{m},0}$, $\rho_{\mathrm{r},0}$, and $\rho_{\Lambda,0}$ respectively, the Friedmann equation (2.6) can be written as

$$\left(\frac{\dot{a}}{a}\right)^2 = H^2(t) = \frac{8\pi G}{3}\left[\rho_{\mathrm{m},0}\left(\frac{a_0}{a}\right)^3 + \rho_{\mathrm{r},0}\left(\frac{a_0}{a}\right)^4 + \rho_{\Lambda,0}\right] - \frac{Kc^2}{a^2}. \tag{2.8}$$

The density value making $K = 0$ (flat Universe) is known as critical density and can be derived from Eq. (2.6) to yield

$$\rho_{\mathrm{crit}}(t) = \frac{3H^2(t)}{8\pi G}. \tag{2.9}$$

Using this critical density, a dimensionless quantity known as the density parameter Ω is defined as a ratio of density to critical density:

$$\Omega \equiv \frac{\rho}{\rho_{\mathrm{crit}}} = \frac{8\pi G\rho}{3H^2(t)}. \tag{2.10}$$

Equivalently, the matter, radiation, and vacuum energy density parameters Ω_{m}, Ω_{r}, and Ω_Λ respectively can be defined by dividing the appropriate density by ρ_{crit}. Using these density parameters allows to rewrite the Friedmann equation (2.8) once again as

$$\left(\frac{\dot{a}}{a}\right)^2 = H_0^2\left[\Omega_{\mathrm{m},0}\left(\frac{a_0}{a}\right)^3 + \Omega_{\mathrm{r},0}\left(\frac{a_0}{a}\right)^4 + \Omega_{\Lambda,0}\right] - \frac{Kc^2}{a^2}. \tag{2.11}$$

The most relevant component of this equation for dark matter analysis is the Ω_m parameter. Therefore, it is appropriate to investigate the time when this component dominates, also known as the Matter Dominated Epoch. In this epoch the radiation density term becomes negligible, and assuming $\Lambda = 0$, Eq. (2.11) reduces to

$$\left(\frac{\dot{a}}{a}\right)^2 = H_0^2\Omega_{\mathrm{m},0}\left(\frac{a_0}{a}\right)^3 - \frac{Kc^2}{a^2}. \tag{2.12}$$

The $K = 0$ case is known as an Einstein–de Sitter (EdS) universe and it is trivial to find its solution

$$\frac{a}{a_0} = \left(\frac{3}{2}H_0 t\right)^{2/3}. \tag{2.13}$$

2.4 Dark Matter

First postulated in the early twentieth century, dark matter and its nature continues to be one of the most puzzling questions in the current day cosmology. Since then, its existence has been implied through several different observations.

A key observational evidence for dark matter is found analysing galaxy rotation curves as shown by Begeman et al. (1991), Corbelli & Salucci (2000), and others. For a spherically symmetric system, the rotation speed for circular orbits V is

$$V^2(R) = \frac{GM(<R)}{R},$$

(2.14)

which falls off as $V \propto 1/\sqrt{R}$ at large radii where the total mass inside that radius has already converged. However, for most galaxies the observed rotation curve remains flat with increasing radius, as the one in Figure 1.
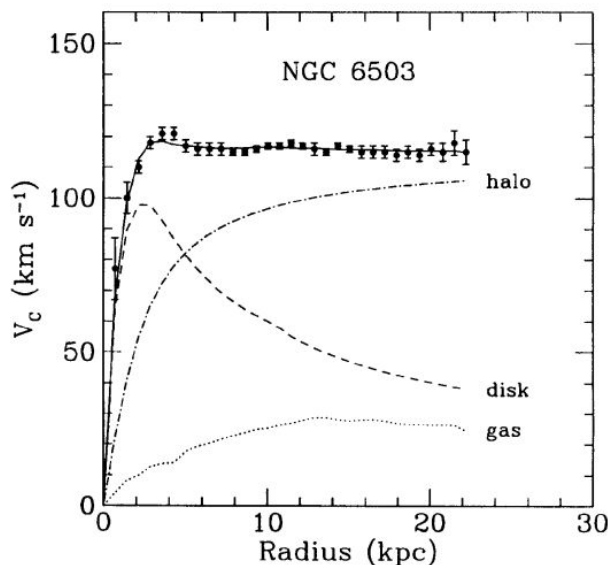

Figure 1. The rotation curve of the NGC 6503 galaxy (Begeman et al. 1991). Data points show a flat rotational trend which is explained by the combination of the disc, gas, and dark matter halo components.

It is therefore deduced that there must be a non-luminous mass distribution that is more dominant toward the outskirts of the galaxy, hence suggesting that galaxies are sitting in a massive dark matter halo. Additionally, other observations, like discrepancy in velocity dispersions in elliptical galaxies (Faber & Jackson 1976), large galaxy cluster mass measured by gravitational lensing (Allen et al. 2011), and imprints on the cosmic microwave background (Planck Collaboration 2020) also lead toward the evidence of dark matter.

Lastly, dark matter effect can be seen on the structure formation of the Universe as well. As radiation was the dominant component of the Universe at early times, it interacted heavily with ordinary matter. As a result, baryon density perturbations were disturbed and unable to fall into a denser structure. Hence, if all matter in the Universe were baryonic, it would not have been able to grow into the structures like galaxies and galaxy clusters in the time up until now (Padmanabhan 1993). However, since dark matter does not interact with radiation, its density perturbations were able to grow undisturbed, forming potential wells that sped up the infall of ordinary matter later.

Therefore, dark matter is of incredible importance to the formation of the structure seen at the current time. Analysing the growth of dark matter perturbations allows us to predict the structure evolution at later times and helps to better understand the nature of dark matter.

2.5   Structure Formation and Growth

In the standard model of Cosmology, the Universe is considered to be highly homogeneous at early times. However, tiny initial density perturbations existed arising from quantum fluctuations, which later, due to effects of gravity, would end up growing to bigger structures, such as galaxies observed today. The evolution of these perturbations can be described using Newtonian theory for the evolution of a non-relativistic fluid of density ρ and velocity $\mathbf{u}$ in a gravitational field with potential ϕ. This description is valid for particles whose mean free path is much smaller than the scale of structure we are interested in, and therefore applies to collisionless dark matter which we want to analyse.

The evolution of this non-relativistic fluid is given by the continuity equation, Euler equations, and the Poisson equation, which describe the mass conservation, fluid motion, and gravitational field respectively:

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} + \rho \nabla_{\mathbf{r}} \cdot \mathbf{u} = 0 \quad \text{(continuity)}, \tag{2.15}$$

$$\frac{\mathrm{D}\mathbf{u}}{\mathrm{D}t} = -\frac{\nabla_{\mathbf{r}}P}{\rho} - \nabla_{\mathbf{r}}\phi \quad \text{(Euler)}, \tag{2.16}$$

$$\nabla_{\mathbf{r}}^2 \phi = 4\pi G \rho \quad \text{(Poisson)}, \tag{2.17}$$

where $\mathbf{r}$ is the proper coordinate, $\partial/\partial t$ is the partial time derivative for fixed $\mathbf{r}$, and

$$\frac{\mathrm{D}}{\mathrm{D}t} \equiv \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla_{\mathbf{r}} \tag{2.18}$$

is the convective time derivative for quantity moving together with the fluid (Mo, van den Bosch & White 2010). Together with the equation of state specifying the fluid pressure P, these six equations can be solved for the six unknowns (ρ, u_x, u_y, u_z, P, and ϕ).

However, to discuss the perturbations evolving in an expanding Universe, a change of variable to comoving coordinates $\mathbf{x}$ is required as given in Eq. (2.3). This in turn gives the proper velocity $\mathbf{u}$ at point $\mathbf{x}$ as

$$\mathbf{u} = \dot{a}(t)\mathbf{x} + \mathbf{v}, \tag{2.19}$$

where the overdot denotes a derivative with respect to t, and $\mathbf{v} \equiv a\dot{\mathbf{x}}$ is the peculiar velocity which describes the motion of the fluid element as seen by the fundamental observer comoving with the background at $\mathbf{x}$.

With the variable change, the time and spatial derivatives transform as

$$\frac{\partial}{\partial t} \to \frac{\partial}{\partial t} - \frac{\dot{a}}{a}\mathbf{x} \cdot \nabla_{\mathbf{x}} \quad \text{and} \quad \nabla_{\mathbf{r}} \to \frac{1}{a}\nabla_{\mathbf{x}}. \tag{2.20}$$

Furthermore, introducing overdensity δ as the density perturbation contrast against the background:

$$\delta(\mathbf{x}, t) = \frac{\rho(\mathbf{x}, t) - \bar{\rho}(t)}{\bar{\rho}(t)} \tag{2.21}$$

and noting that the mean background density scales as $\bar{\rho} \propto a^{-3}$, allows to rewrite the fluid equations (2.15) – (2.17) in comoving coordinates as

$$\frac{\partial \delta}{\partial t} + \frac{1}{a}\nabla \cdot (1 + \delta)\mathbf{v} = 0, \tag{2.22}$$

$$\frac{\partial \mathbf{v}}{\partial t} + \frac{\dot{a}}{a}\mathbf{v} + \frac{1}{a}(\mathbf{v} \cdot \nabla)\mathbf{v} = -\frac{\nabla \Phi}{a} - \frac{\nabla P}{a\bar{\rho}(1 + \delta)}, \tag{2.23}$$

$$\nabla^2 \Phi = 4\pi G \bar{\rho} a^2 \delta, \tag{2.24}$$

where $\nabla \equiv \nabla_{\mathbf{x}}$, $\partial/\partial t$ is for fixed $\mathbf{x}$, and $\Phi \equiv \phi + a\ddot{a}x^2/2$ is the potential perturbation.

For isentropic density perturbations in a pressureless fluid these equations reduce to a second-order differential equation (Mo, van den Bosch & White 2010)

$$\frac{\partial^2 \delta}{\partial t^2} + 2\frac{\dot{a}}{a}\frac{\partial \delta}{\partial t} = 4\pi G \bar{\rho}\delta. \tag{2.25}$$

The second term on the left-hand side is proportional to Hubble parameter H and suppresses the perturbation growth due to the expansion of the Universe, therefore being known as the Hubble drag term. While the term on the righ-hand side is the gravitational source term causing the perturbations to grow.

For Einstein–de Sitter universe results in §2.3, where $\Omega_{\mathrm{m}} = \Omega = 1$,

$$\frac{a(t)}{a_0} = \left(\frac{3}{2}H_0 t\right)^{2/3} \quad \Rightarrow \quad \frac{\dot{a}}{a} = \frac{2}{3t}, \tag{2.26}$$

$$\Omega = \frac{8\pi G\bar{\rho}}{3H^2(t)} = 1 \quad \Rightarrow \quad 4\pi G\bar{\rho} = \frac{2}{3t^2}, \tag{2.27}$$

giving the linear perturbation equation

$$\frac{\partial^2 \delta}{\partial t^2} + \frac{4}{3t}\frac{\partial \delta}{\partial t} - \frac{2}{3t^2}\delta = 0. \tag{2.28}$$

Looking for solutions of form $\delta \propto t^n$, we find solutions for $n = -1$, corresponding to the decaying mode, and $n = 2/3$, the growing mode. Neglecting the decaying mode, we find

$$\delta(t) \propto t^{2/3} \propto a(t). \tag{2.29}$$

i.e. the structure growth is linear with a. This is known as the Linear Perturbation Theory.

However, in an universe with non-zero cosmological constant like ours, where $\Omega_{m,0} < 1$, perturbations are expected to grow slower due to larger expansion rate than in an EdS universe and therefore enhanced Hubble drag term (Ferreras 2019). Additionally, when the perturbations have grown sufficiently, the density in a small region around them becomes much higher than $\bar{\rho}$. In that case, the physics become significantly more complicated and the linear pressureless fluid approximation becomes invalid. The full Newtonian theory of gravity needs to be considered which is best analysed through N-body simulations rather than analytically.

2.6   The Legacy Project

The data used for this project came from the Legacy Project (Meriot et al. 2022) which consisted of a series of cosmological N-body simulations run with GAlaxies with Dark matter and Gas intEracT (GADGET) code (Springel 2005). These simulations have been designed to sample -2, -1, 0, +1, and +2 σ of the mean density value, as well as extremely high (cluster) and very low (void) density regions, and are therefore ideal in studying different environmental effects.

Ω_Λ	Ω_{m}	Ω_{b}	h	σ_8	n_{S}	Y_{He}	N_ν	w
0.7150	0.2850	0.0450	0.6950	0.8280	0.9632	0.2480	3.0400	-1.0

Table 1. The Dark energy density Ω_Λ, matter density Ω_{m}, baryon density Ω_{b}, Hubble parameter h, matter fluctuation amplitude σ_8, scalar spectral index n_{S}, primordial helium abundance Y_{He}, number of neutrino species N_ν, and dark energy equation of state parameter w values used in the Legacy Project from WMAP9 results.

The initial conditions for the simulations have been created using the MUlti-Scale Initial Conditions (MUSIC) code (Hahn & Abel 2011). The transfer function for the cosmological model needed to generate the initial conditions was made using the Code for Anisotropies in the Microwave Background (CAMB) (Lewis et al. 2000). The cosmological parameters for

the simulations were chosen to be consistent with the latest Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP9) (Hinshaw et al. 2013) results, and are as shown in Table 1.

The simulation chosen for the analysis is a box with a side-length of 1600 Mpc/h and resolution of 2048^3 particles, each of mass $M_{\mathrm{p}} = 5.43 \times 10^{10} M_\odot$, which is shown in Figure 2. This simulation was initiated at redshift $z = 99.0$ and ran to $z = 0.0$ saving all the particle data at 102 snapshots, evenly spaced in $\log a$. Additionally, 300 halo catalogs produced from $z = 20.0$ to $z = 0$ using a friend-of-friends Rockstar halo finder algorithm (Behroozi et al. 2013) were used for halo data.
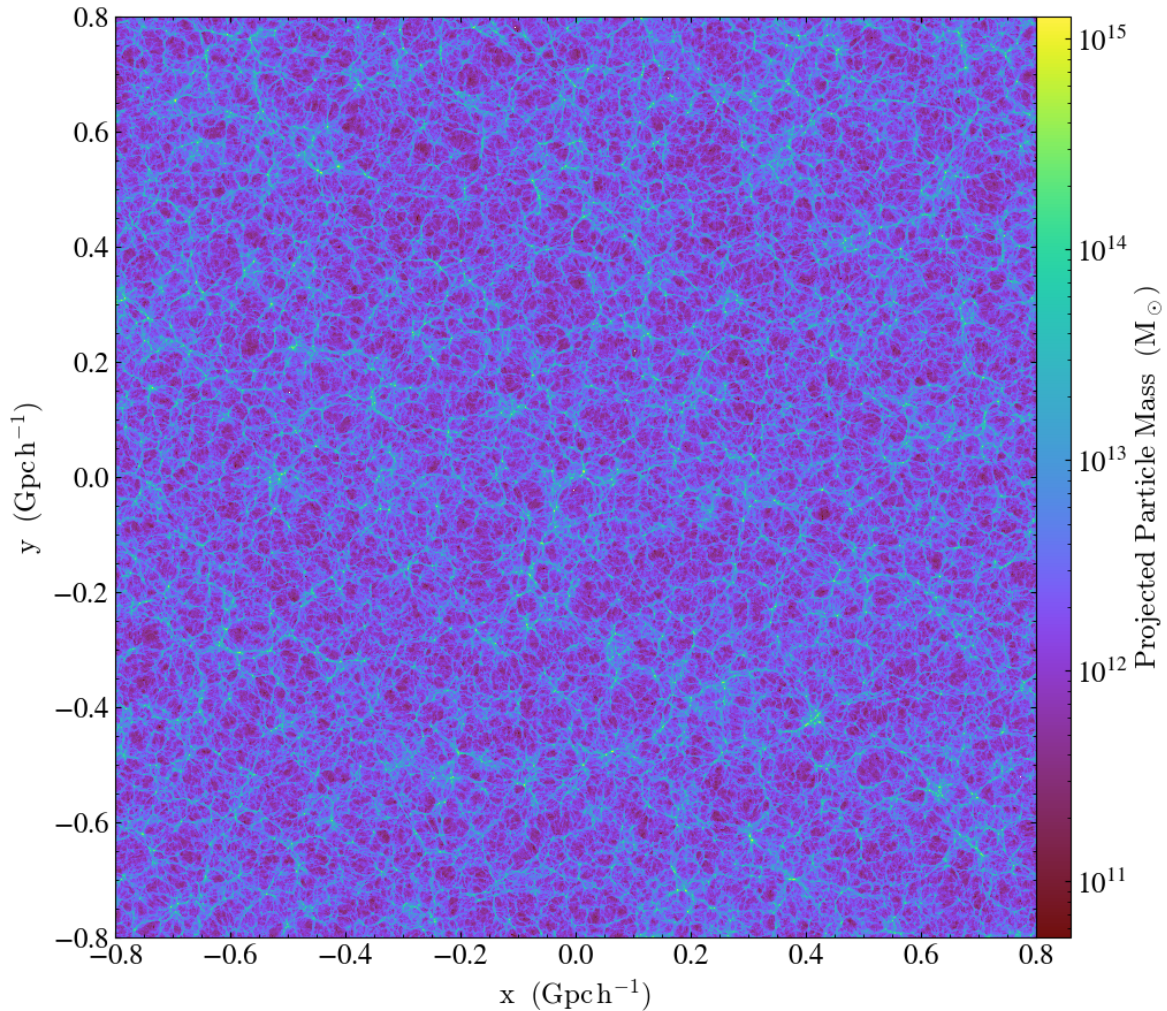

Figure 2. Snapshot 101 ($z = 0$) of the full-box Legacy simulation with 2048^3 particles (70 Mpc/h in depth). Note the cosmic web structure emerging at this time.

3 Methods

The following analysis was done with Python, using the *yt* package (Turk et al. 2011). The plots were produced using a *matplotlib* package (Hunter 2007) (except the ones visualising the simulation, which were made with *yt*). All of the coding and data analysis was performed on the University of Edinburgh Institute for Astronomy research computing cluster *cuillin* with SLURM workload manager. Message Passing Interface (MPI) was used for parallel computing with *mpi4py* package (Dalcín et al. 2005), parallelising over separate snapshot datasets using `yt.parallel_objects`. Additionally, open source *NumPy* and *SciPy* Python packages were used.

3.1 Gaussian Random Fields

To start with, the Gaussian properties of the dark matter density field were analysed. The simulation snapshots were loaded in using `yt.load()` with the base length unit of 1 Mpc/h.

Five hundred spheres of six different radii ranging between 25 Mpc/h to 50 Mpc/h with increments of 5 Mpc/h were created at random locations in the simulation box using *yt* sphere data objects (Figure 3). The coordinates for these random locations were generating using *numpy's* `random.randint` function. The density, and in turn the overdensity (Eq. 2.21), was calculated inside each sphere, using the mass of the particles inside the region obtained with *yt's* `total_mass()` quantity associated with the sphere data object. The average density $\bar{\rho}$ used in the calculation was taken as the density of the whole simulation box, scaled by a^{-3}. This was done for sixteen snapshots, spanning from $z = 0$ to $z = 99$ [2].

Histograms of these overdensities were created for each radius at each snapshot with 20 bins each. Additionally, every histogram was fitted with a Gaussian fit:

$$y = a \exp\left[-\frac{(x - \bar{x})^2}{2\sigma^2}\right] \tag{3.1}$$

for 3 parameters: amplitude a, mean $\bar{x}$, and spread σ using a non-linear least squares fitting algorithm `optimize.curve_fit` from *SciPy* package on the histogram bin mid-point values.

3.2 Overdensity Evolution

Next, the overdensity evolution at fixed random points throughout time was analysed. Once again, spheres were created in the simulation box, of 10, 30, and 50 Mpc/h radius. These spheres were placed at random locations at the same comoving coordinates in every snapshot (Figure 3). Furthermore, all 102 available snapshots were used. 1000 spheres of 10 Mpc/h radius were created, 500 of 30 Mpc/h radius, and 600 of 50 Mpc/h radius. The overdensity in these spheres was calculated the same way as described in §3.1.

---

[2]Snapshots of redshifts $z = 0, 0.25, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 99$ were used
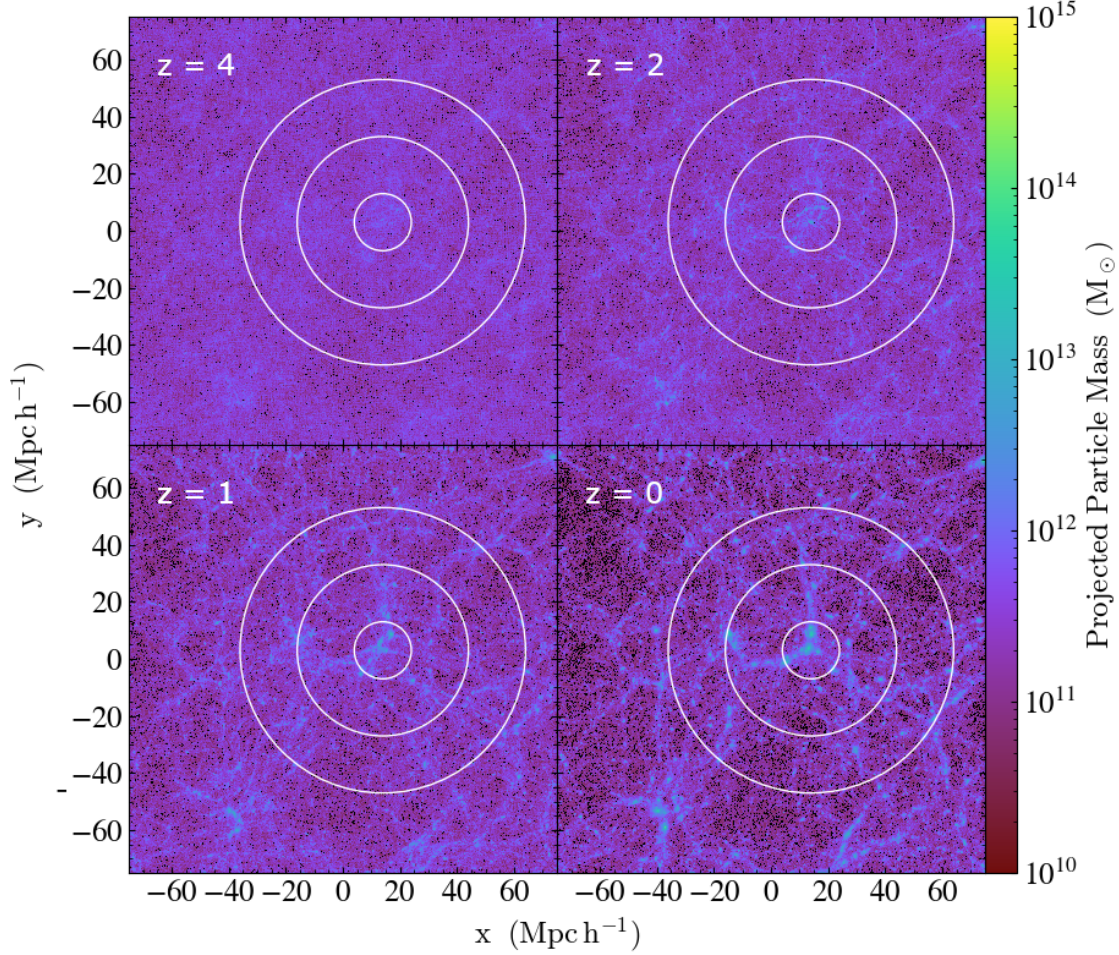
Figure 3. 10 Mpc/h, 30 Mpc/h, and 50 Mpc/h radius spheres around a point in Legacy simulation which happens to be around a massive halo near the centre of the simulation box. Note the condensing of the structure web with time.

3.3 Evolution of Massive Halos

Moving on, the analysis shifted to halos. Halo catalogue of 10^7 halos at redshift $z = 0$ was loaded in using `yt.load()` and its data was accessed with `all_data()` attribute. The centre coordinates of the five most massive halos[3] were extracted from this catalogue data.

The analysis that followed is familiar to that in §3.2: spheres (of radius 10, 30, and 50 Mpc/h) were created at the halo centre coordinates (at $z = 0$) in all 102 snapshots spanning from $z = 0$ to $z = 99$, and the overdensity inside them was calculated the same way as before. Note, that the sphere centre coordinates used were always the same comoving coordinates, equal to the halo centre coordinates at $z = 0$, therefore, following the point in space rather than halo centre itself.

---

[3]Masses of $(2.86, 2.88, 3.79, 4.05, 5.20) \times 10^{15} M_\odot$ and virial radii between 2.9 and 3.6 Mpc/h at $z = 0$.

4 Results and Discussion

4.1 Comparison to the Linear Theory

Overdensity histograms of different radius spheres at random points, produced as described in §3.1, are shown in Figures 4 and 5 for 25 Mpc/h spheres. A least squares Gaussian fit to the data is also overplotted for each histogram in orange. Figure 4 shows the histograms for overdensity in spheres of the same $R = 25$ Mpc/h radius at different redshifts of $z = 99$, 10, 5, and 0. The x-axis denoting δ is chosen to be the same for all histograms, showcasing their evolution and spread with time. Meanwhile, Figure 5 is a widened version of the $z = 99$ histogram in Figure 4. The histograms for other radius spheres were similar in shape.
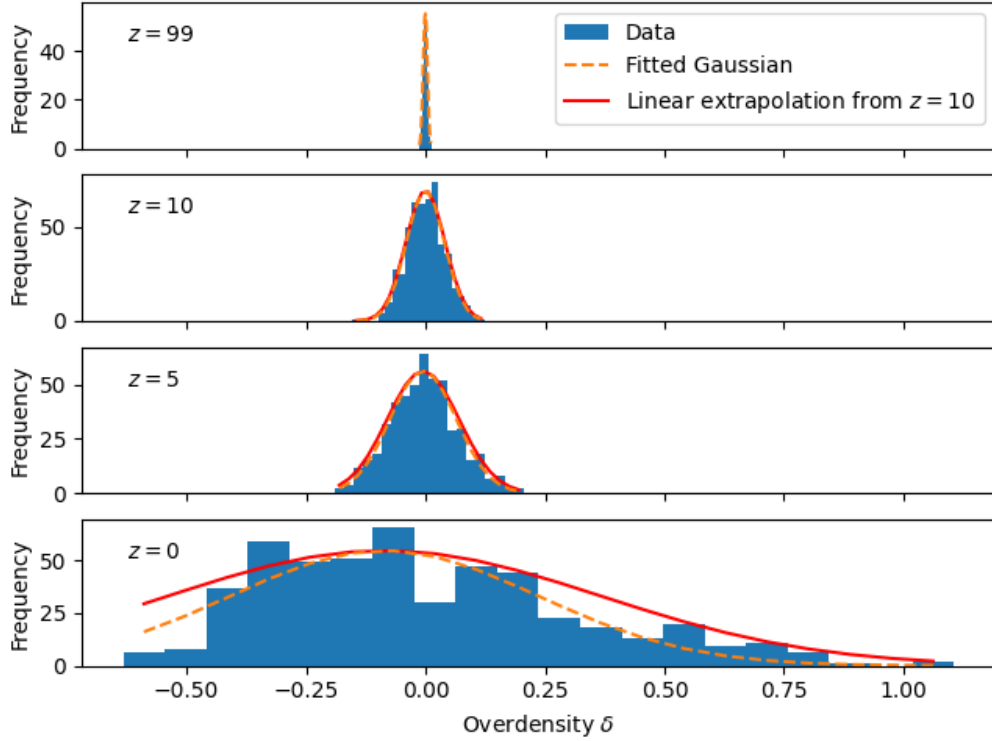

Figure 4. Histograms of overdensities calculated in 500 spheres of 25 Mpc/h radius placed at random points in the simulation at $z = 99$, 10, 5, and 0. A Gaussian fit to the data is plotted in orange. An extrapolation of the $z = 10$ Gaussian using linear theory is plotted in red for $z = 5$ and 0 histograms. x-axis denoting overdensity is the same for all plots, showcasing their evolution with redshift. Note the data at $z = 0$ having less spread than expected by the extrapolation. A widened version of $z = 99$ histogram is available in Figure 5.

Additionally, the $z = 10$ data was used to make a linear extrapolation of how δ would evolve if it scaled linearly with scale factor, $\delta \propto a(z)$, as approximated by linear theory. This was achieved by setting $\delta = ba(z)$ for some constant b and obtaining this constant using the values of the spread of the fitted Gaussian σ, and $a = 1/11$. The value obtained was $b = 0.44(2)$ and the estimated Gaussians obtained using this value can be seen plotted in red in Figure 4.
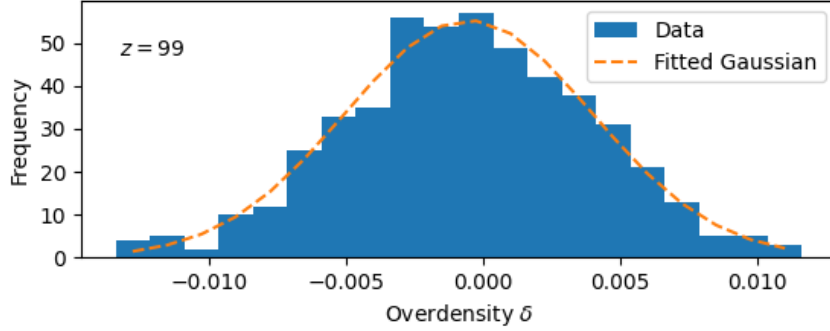
Figure 5. The histogram of overdensities calculated in 500 spheres of 25 Mpc/h radius placed at random points in the simulation at $z = 99$. A Gaussian fit to the data is plotted in orange.

From Figures 4 and 5 it can be seen that the overdensity histograms appear to be strongly Gaussian at early times (high z). Additionally, the width of the histogram and the variance of the Gaussian increase with decreasing redshift. This is explained by the Universe being much smoother at early times and developing more overdense or underdense regions (filaments and voids) at later times due to effects of gravitational structure formation as described in §2.5. However, it can also be seen, that the Gaussians (orange line) at later times (low z), particularly the present ($z = 0$), are narrower than they should be, compared to the extrapolation based on the linear theory (red line). Therefore, this supports the argument posited at the end of §2.5 that the structures grow slower than linearly in an universe with non-zero cosmological constant.

To visualise the disagreement of the linear theory with our Universe better, the evolution of σ of the fitted Gaussians was plotted in Figure 6 for the 6 different sphere sizes used. It is firstly apparent that σ is overall bigger when smaller size regions are used for overdensity calculations, as bigger regions will tend to more closely match the average background density $\bar{\rho}$ resulting in δ values closer to zero. But more importantly, it can also be seen that for each sphere size the evolution appears to be linear at first (low a and high z) but starts flattening at later times (as $a \rightarrow 1$ and $z \rightarrow 0$), deviating from the linear expectation.

However, even though providing the predicted behaviour, the methods used here come with several shortcomings. Firstly, a big number of data points is needed for accurate results and precision. This, however, soon becomes computationally expensive as hundreds and thousands of the simulation particles need to be added up to obtain the total mass inside a sphere for each new region. Secondly, the histograms may not always fit a Gaussian, as indicated by the $z = 0$ case in Figure 4, which is strongly skewed to the right. This right-skewed behavior is always present at low z and more noticeable with overdensities calculated in smaller regions (lower R). Therefore, it might not be appropriate to fit all the data with an unskewed Gaussian and use its variance for the analysis. However, this method brilliantly demonstrates the Gaussian nature of the dark matter density field at early times. Finally, an alternative approach, as detailed in §3.3 − 3.2 was considered, and an approximation to the structure growth more refined than linear theory was used in the following work.
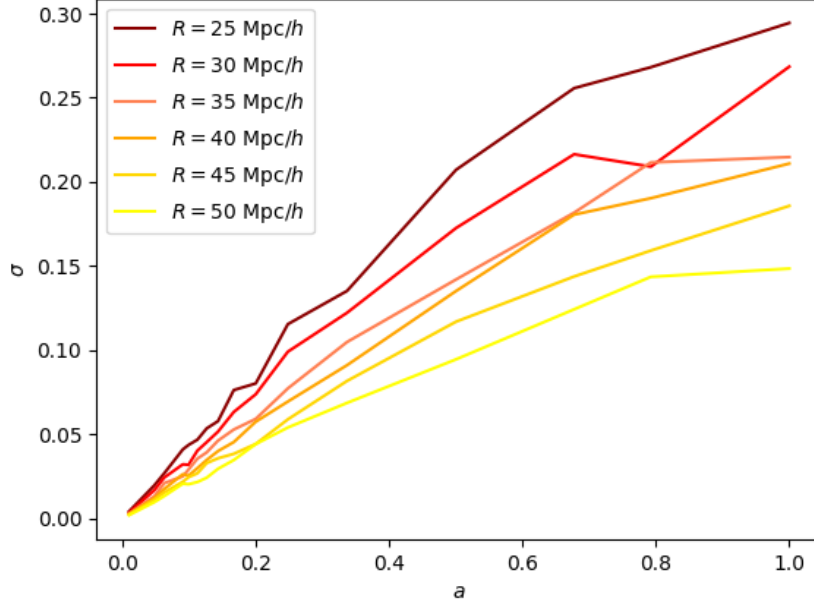
13

Figure 6. The evolution of the spread of the Gaussian fit to the data σ with cosmic scale factor a for different sphere sizes. Note the deviation from linear behaviour of evolution with $a \to 1$.

4.2 Comparison to the Carroll Approximation

This time the overdensity evolution with a was calculated in regions around fixed comoving coordinates as detailed in §3.2 and plotted for each sphere. Figure 7 shows the example for 30 Mpc/h radius spheres while same plots for 10 Mpc/h and 50 Mpc /h radius spheres are available in Appendix A.

It is notable that the evolution appears linear at low a and slows down for high a. Additionally, using regions around random locations in the simulation allowed a distributed sampling of both overdense ($\delta > 0$) and underdense ($\delta < 0$) regions.

To make the analysis easier, each overdensity evolution rate was normalised by its end-point value $\delta_0 = \delta(z = 0)$, to make all lines end at the same point at $a = 1$, and plotted as a heatmap (Figure 8). Additionally, the data was compared to a perturbation growth approximation by Carroll et al. (1992):

$$\delta \propto g(z)a(z), \tag{4.1}$$

where

$$g(z) \approx \frac{5}{2}\Omega_m(z) \left\{ \Omega_m^{4/7}(z) - \Omega_\Lambda(z) + \left[1 + \frac{\Omega_m(z)}{2}\right]\left[1 + \frac{\Omega_\Lambda(z)}{70}\right] \right\}^{-1}, \tag{4.2}$$
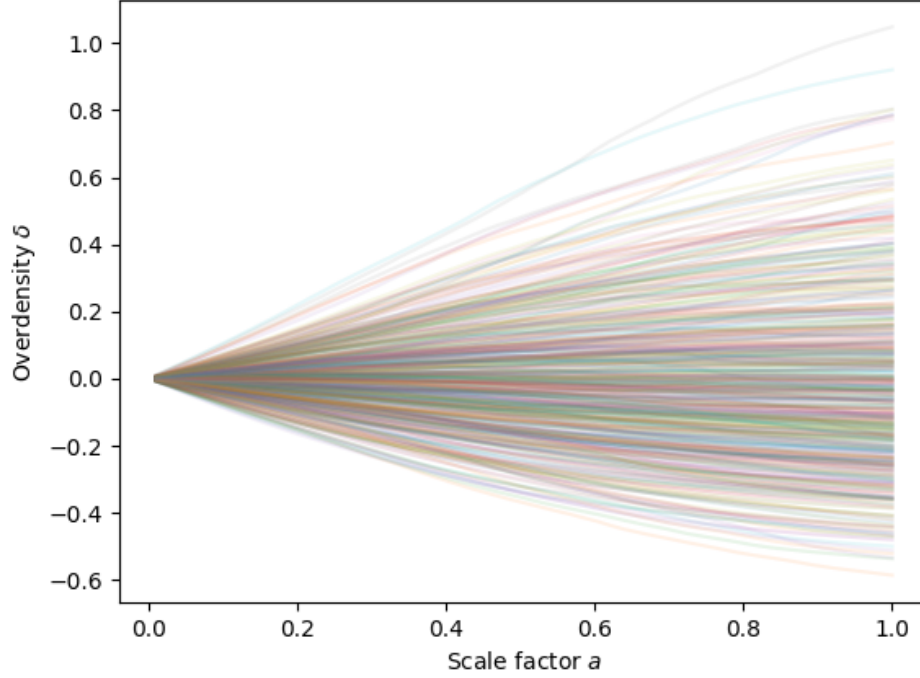
14

Figure 7. Overdensity evolution with scale factor in five-hundred 30 Mpc/h radius spheres at fixed random comoving locations.

with $\Omega_m(z)$ and $\Omega_\Lambda(z)$ obtained from Friedmann equation (2.11) using present day values $\Omega_{m,0}$ and $\Omega_{\Lambda,0}$ from Table 1. This approximated growth rate was also normalized to be equal to 1 at $a = 1$ like the data. The Carroll approximation for each heatmap is shown by a blacked dashed line. The linear theory $\delta \propto a(z)$ extrapolation was obtained as the tangent line to Carroll approximation as $a \to 0$ and is plotted in white.

The data clearly follows the Carroll approximation well and deviates from the linear prediction for high a. Overdensities obtained in bigger regions (50 Mpc/h radius spheres) follow the Carroll line closer than those using smaller regions as expected, as Figure 6 showed that overdensity in bigger regions has less spread around its mean value. It can also be seen that the departure from linear theory starts happening at approximately $a \approx 0.5$ ($z \approx 1$).

The residuals of the data were calculated as:

$$\mathrm{res} = \frac{\delta(z)/\delta_0 - g(z)a(z)}{g(z)a(z)}, \tag{4.3}$$

with $g(z)$ from Eq. 4.2 and scale factor $a(z)$. These were plotted in Figure 9 for 30 Mpc/h radius spheres. Regions less than 10% away from the Carroll approximation for all a were selected and are shown in red. Additional residual plots for 10 and 50 Mpc/h radius spheres can be found in Appendix B.
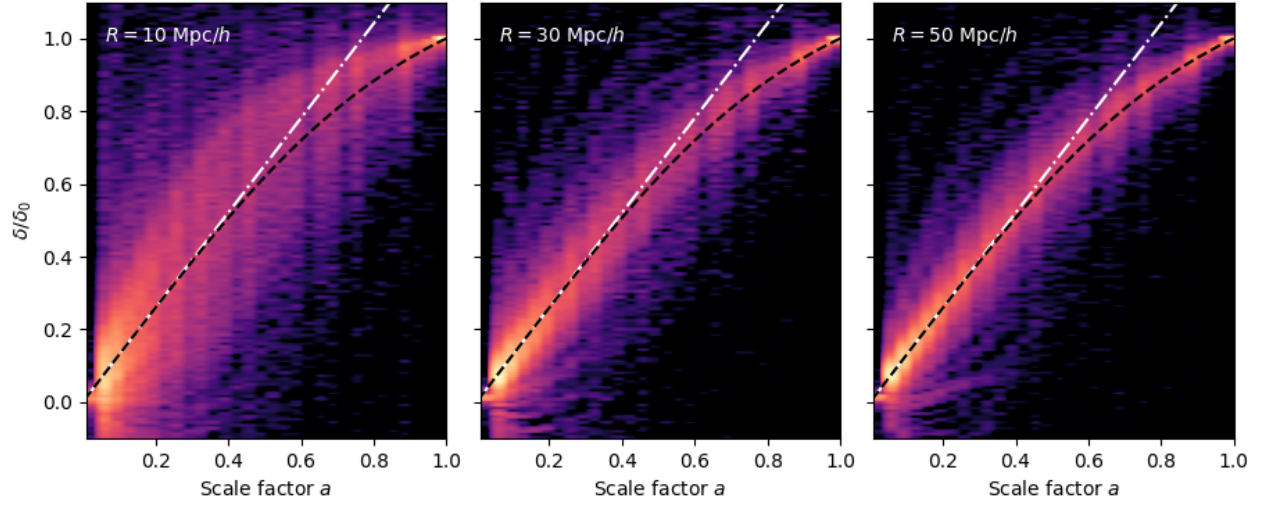
Figure 8. Heatmaps of overdensity evolution around fixed comoving random points. The left heatmap shows overdensities of 1000 spheres of 10 Mpc/h radius, the middle one is of 500 spheres of 30 Mpc/h radius, and the right heatmap is of 600 spheres of 50 Mpc/h radius. The overdensity values were normalised by the values at $a = 1$ to make them share the same end-point. The Carroll approximation (Eq. 4.1) is shown in black and the linear theory extrapolation (Eq. 2.29) is in white. Note the close matching of the data to the Carroll approximation and departure from linear theory at around $a \approx 0.5$.
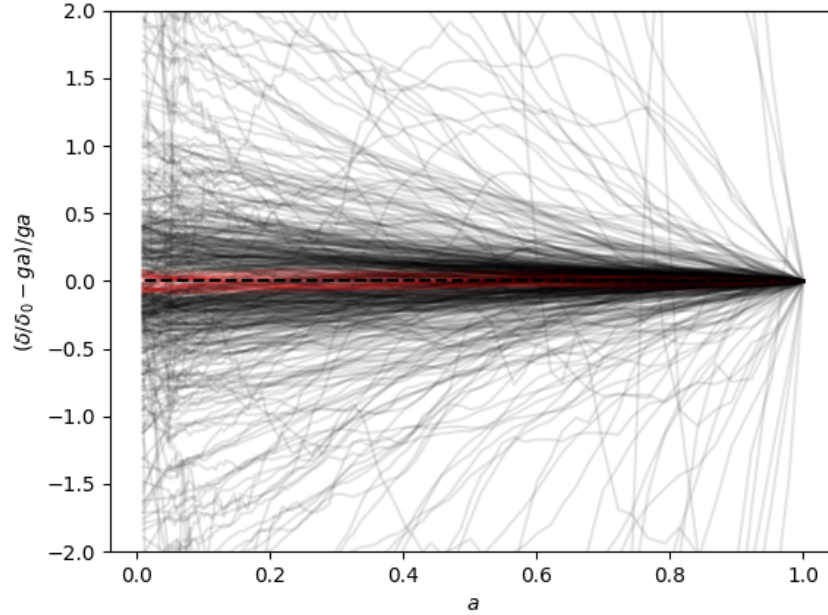

Figure 9. Plot of residuals of δ/δ_0 around the Carroll approximation (Eq. 4.1) for five-hundred 30 Mpc/h sphere regions. 39 regions matching the approximation within 10% are in red.

The residual plots show an even spread around the approximation, however, a lot of the lines are further away from it than expected. 0.5% (5 out of 1000) of the residuals for 10 Mpc/h regions lie within 10% of the approximation, and so do 7.8% (39 out of 500) of 30 Mpc/h regions, and 19% (114 out of 600) of 50 Mpc/h regions. These numbers increase to 3.1%, 25.8%, and 45.2% respectively for residuals within 20% of the approximation. Additionally, there is a lot of scatter in the lines, especially for smaller regions, with some of them crossing the prediction back and forth multiple times. It is thought that this could be caused with overdensities moving in and out of the spherical region defined at a fixed location as the time passes. Other factors, like non-linear effects caused by intersecting particle trajectories and particles forming orbits, and the cosmic web structure are also thought to be the contributors to this spread at higher a. Furthermore, most of the scatter is for $a \approx 0$ but that is expected as overdensities started out of different sizes, and their residuals are amplified by the data being divided by the Carroll approximation (Eq. 4.3) which is close to zero at that time.

4.3 Probability of Predicted Growth

Moving on, using the obtained overdensity evolution lines, the conditional probabilities for predicted structure growth were calculated. These probabilities were defined as the probability of $\delta(z')/\delta_0$ in a region being less than one (or two) σ away from the Carroll approximation at redshift z', given $\delta(z)/\delta_0$ in that same region was less than one σ away from the approximation at redshift z. These probabilities were calculated for two separate regions above and below the Carroll approximation[4].

To calculate the probabilities, first, the number of evolution lines that fell one σ away from Carroll approximation in either the positive or negative region at a given redshift z, $N(z)$, was taken. Out of those lines, only the ones that fell less than one (or two) σ away in the corresponding region at z' were chosen, with their number being $N'(z')$. The probability was then simply obtained by dividing one number by another:

$$\mathbb{P}(z, z') = \frac{N'(z')}{N(z)}. \tag{4.4}$$

The σ used here was calculated by slicing through the heatmaps (Figure 8) at a given redshift z, producing a slice plot like in Figure 10, and fitting a Gaussian function to it, obtaining its spread σ (in the same way as described in §3.1). The σ obtained for that given z slice plot was then used for every z' probabilities. The probabilities were calculated using starting regions at $z = 1, 2, 3, 6$, and for all available z'. The probabilities for positive σ regions (less than one or two σ above the Carroll approximation) are plotted in Figure 11, while the probabilities for negative σ (less than one or two σ below the Carroll approximation) are shown in Figure 12. Note that the $z = 0$ data were excluded from the plots as δ was scaled to always match 1 in present.

---

[4]Regions between $g(z)a(z)$ and $g(z)a(z) + \sigma$, and between $g(z)a(z) - \sigma$ and $g(z)a(z)$. Equivalent for 2σ.
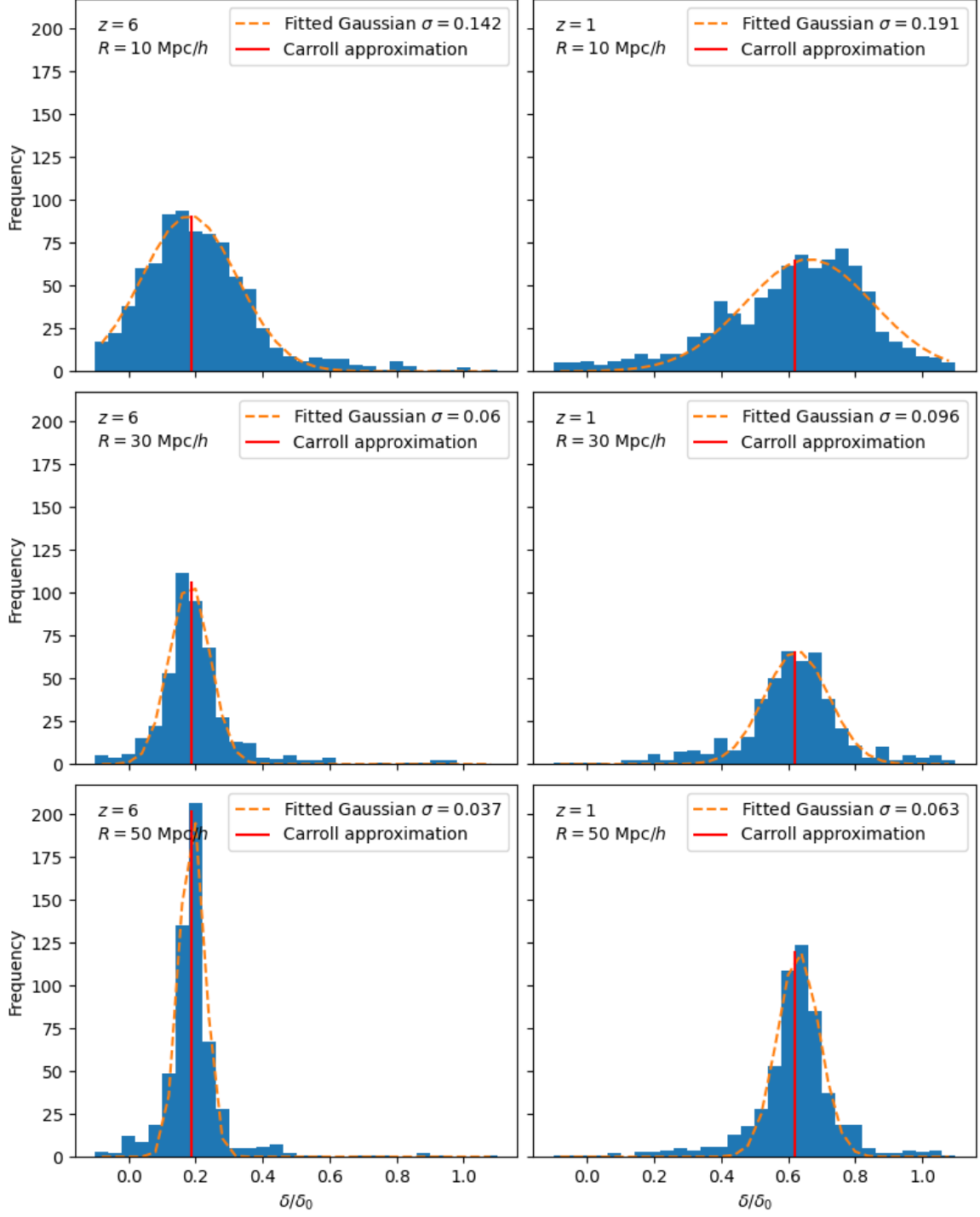
Figure 10. Slice plots of the heatmaps in Figure 8 at $z = 6$ and 1 ($a = 0.14, 0.5$) for 10 Mpc/h (top), 30 Mpc/h (middle), and 50 Mpc/h (bottom) radius spheres. The Carroll approximation for each z is marked in red and a fitted Gaussian is in orange with its σ in the legend. Note the reduce in spread for bigger sphere regions. Since there were different numbers of regions used for each sphere radius, the histograms do not have the same number of data points.
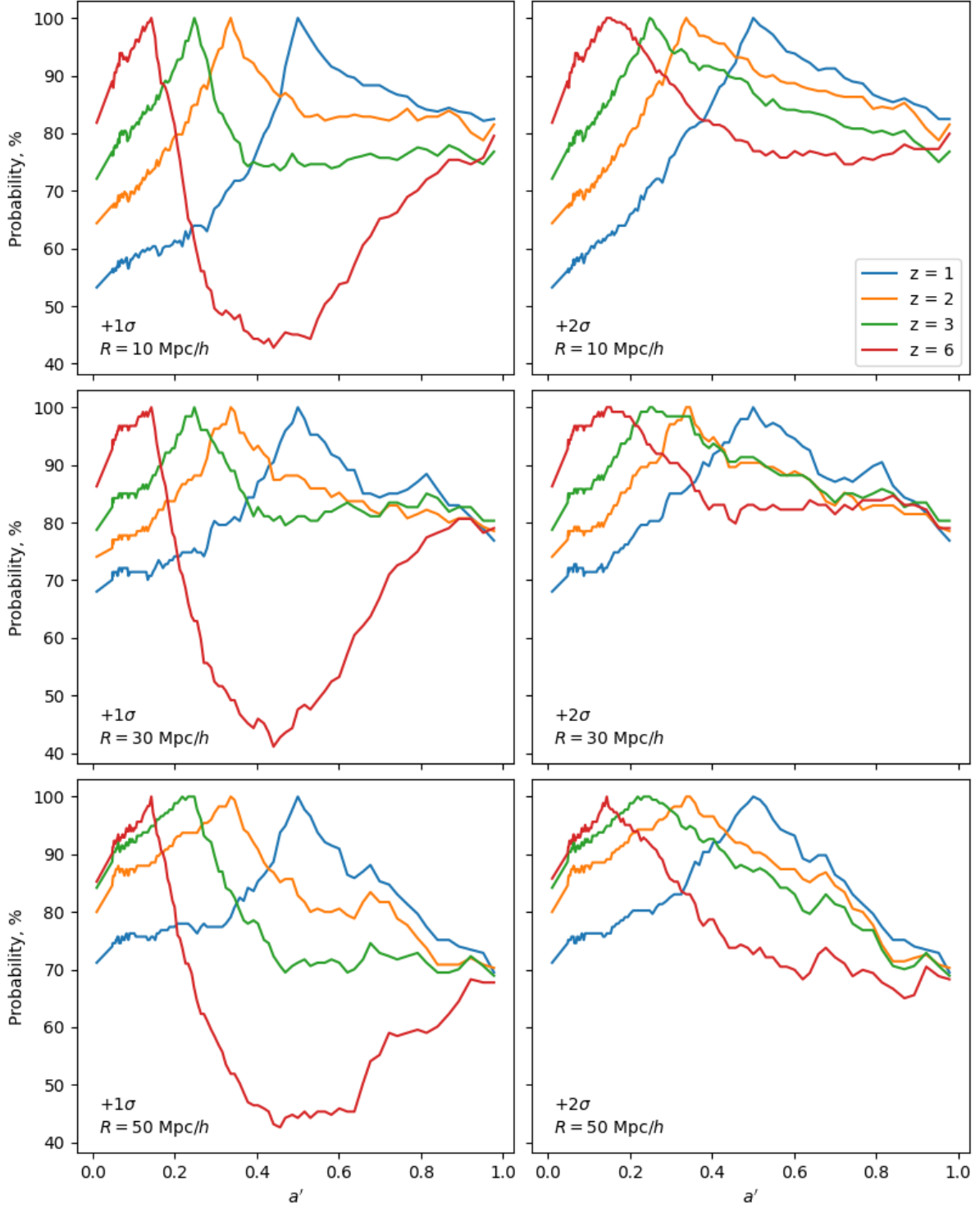
Figure 11. Probabilities of $\delta(z')/\delta_0$ being less than one (left) or two (right) σ above the Carroll approximation at z', given $\delta(z)/\delta_0$ in that same region was less than one σ above the approximation at z (Eq. 4.4). Probabilities for 10 (top), 30 (middle), and 50 (bottom) Mpc/h radius spherical regions are provided.
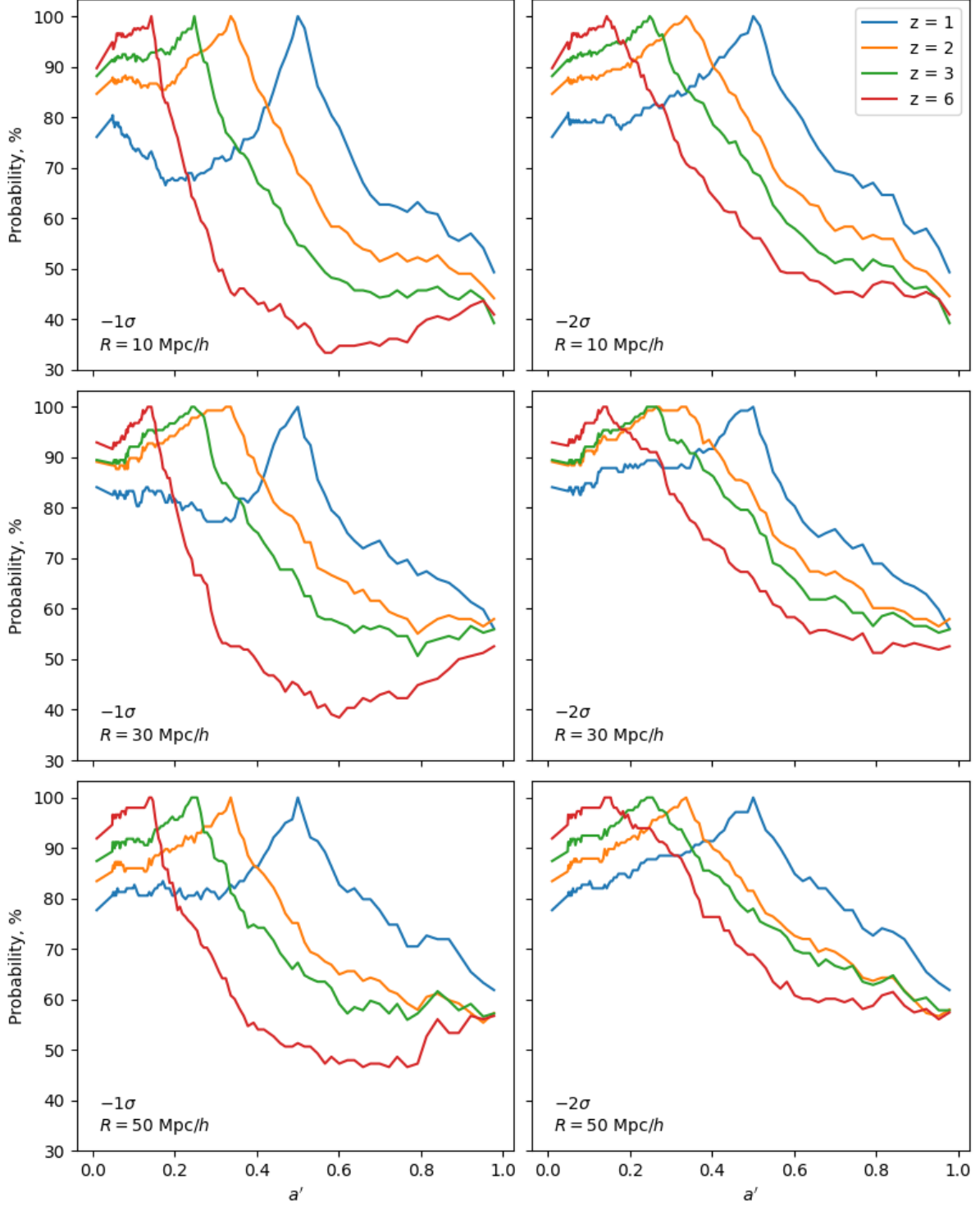
Figure 12. Probabilities of $\delta(z')/\delta_0$ being less than one (left) or two (right) σ below the Carroll approximation at z', given $\delta(z)/\delta_0$ in that same region was less than one σ below the approximation at z (Eq. 4.4). Probabilities for 10 (top), 30 (middle), and 50 (bottom) Mpc/h radius spherical regions are provided.

By definition (Eq. 4.4), these probabilities are equal to 100% at $z' = z$. They appear to decrease linearly with a for z' close to z but start converging for $a \to 1$. However, the convergence does not appear to happen at the same points for positive and negative regions. This is particularly noticeable for smaller regions, where the probabilities less than one σ above $g(z)a(z)$ approach 75–85% meanwhile the ones below approach 40–50% for 10 Mpc/h radius spheres as $a \to 1$ (Figures 11 & 12). This suggests that the regions that were evolving in a more overdense way in the past are more likely to evolve like that in the present as well, but the ones evolving bellow the Carroll prediction are more likely to cross over it with time. Furthermore, even though less noticeable, the behavior is opposite in the $a \to 0$ direction, where the probability lines interestingly start out higher (around 80–90%) for negative σ region than the positive one (50–85%).

Lastly, the evolution with scale factor of the spread σ of the Gaussians fitted to the slice plots (Figure 10) is given in Figure 13. Here, the σ values used for $z = 1, 2, 3,$ and 6 can be seen.
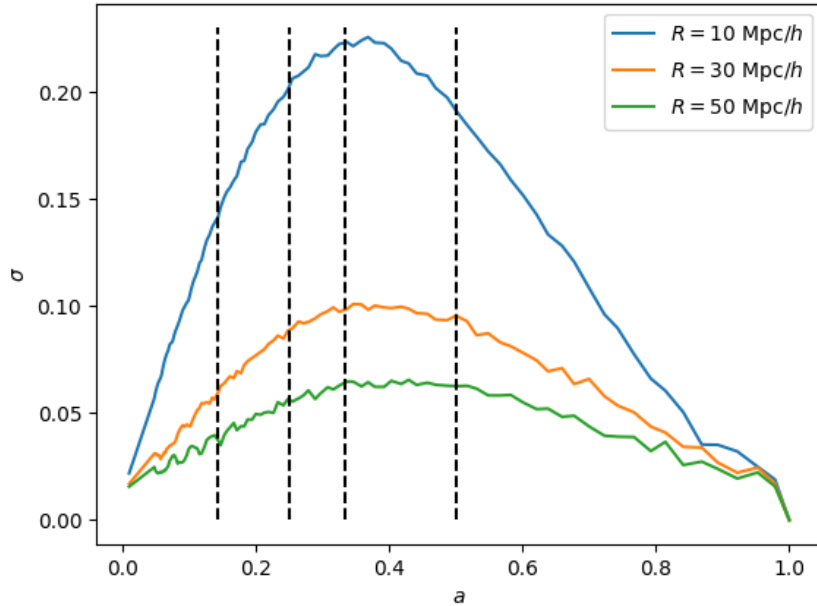

Figure 13. The spread σ of the Gaussians fitted to the slice plots in Figure 10 at different a for the three region sizes used. Dashed vertical lines indicate the σ values used in computing the probabilities given in Figures 11 and 12 for $z = 1, 2, 3, 6$ ($a = 0.5, 0.33, 0.25, 0.14$).

These σ trajectories are like expected, with $\sigma \to 0$ at both ends as $a \to 0, 1$ and higher in between. This represents the behavior seen in the heatmaps in Figure 8 and the turnover for higher a is due to the scaling of δ by δ_0.

4.4 Halo Growth

Lastly, the overdensity evolution in regions around five most massive halos (at $z = 0$) was obtained as explained in §3.3 and is shown in Figure 14 where δ was calculated in 10, 30, and 50 Mpc/h radius sphere regions.
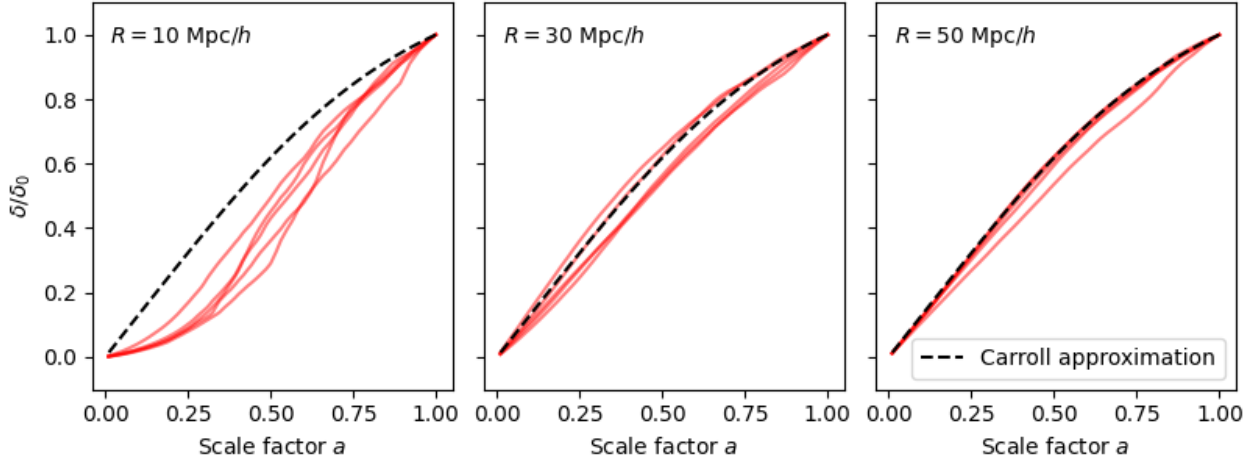

Figure 14. The overdensity evolution in 10 Mpc/h (left), 30 Mpc/h (middle), and 50 Mpc/h (right) radius spherical regions around the five most massive halos at $z = 0$. The overdensity values were normalised by the values at $a = 1$ to make them share the same end-point. The Carroll approximation (Eq. 4.1) is shown in black.

The data from these five massive halo regions appears to be consistent and they all follow a similar evolution. Most interestingly, a noticeable deviation from the Carroll's predicted evolution can be seen for 10 Mpc/h radius sphere regions. The evolution seems to accelerate for higher a, starting at around $a \approx 0.25$, and it stays linear much longer after that than for bigger regions. Therefore, there is a significant cluster impact to the structure growth in a region around them, even as big as 10 Mpc/h in radius (compared to the virial radius of these halos of around 3 Mpc/h). This region becomes strongly overdense and matter dominated early on, and the conditions inside the region start to resemble an Einstein de Sitter matter dominated universe, so the structure evolution follows the linear theory (§2.5). However, for larger regions (30 and 50 Mpc/h radius in Figure 14) the massive cluster impact diminishes and they closely follow the Carroll line.

Finally, the overdensities obtained in 30 and 50 Mpc/h radius spherical regions around 100 most massive halos at $z = 0$ were calculated and scaled. They were plotted against each other as seen in Figure 15 to further confirm that the cluster impact in regions of this size is no longer significant. However, it can also be seen that it is still non-zero as there is some spread around the diagonal, and even a couple outliers.
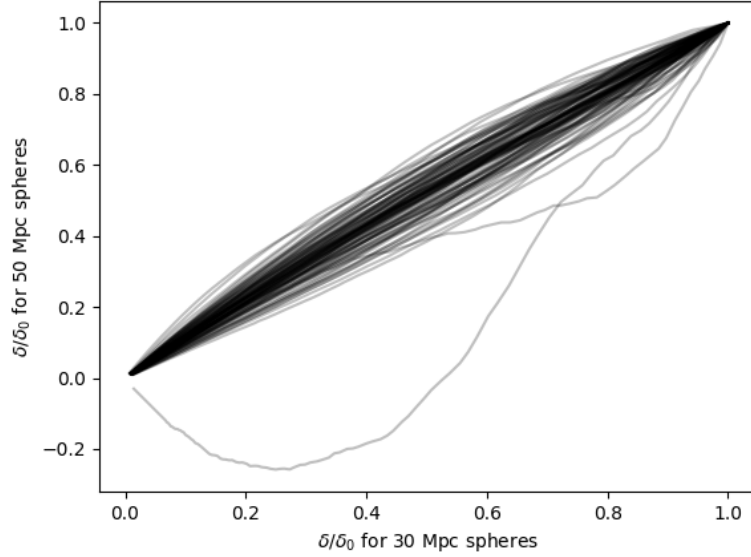
Figure 15. The evolution of δ in 50 Mpc/h radius spheres around 100 most massive halos at $z = 0$ compared to the evolution of δ in 30 Mpc/h radius spheres around the same halos. All δ values were scaled by δ_0.

5 Conclusion

In this project, the properties of the cosmic dark matter density field and the growth of its perturbations were analysed. The attention was directed toward examining the non-linear structure formation. N-body simulation from the Legacy project of a side-length 1600 Mpc/h containing 2048^3 dark matter particles, each of mass $M_{\mathrm{p}} = 5.43 \times 10^{10} M_{\odot}$, and spanning redshifts from $z = 99$ to 0, was used in the analysis. Overdensity δ was calculated in the spherical regions in the simulation box and its evolution with scale factor a was studied.

Placing the spheres at random points in the simulation created Gaussian δ histograms which started deviating from the normal distribution with higher a. Additionally, the spread σ of the histograms was shown to evolve slower than predicted by the linear perturbation growth theory at larger a, as expected for an universe with non-zero cosmological constant.

Next, δ evolution was tracked in the regions of the same comoving coordinates. The same evolution line shape, linear for small a and flattening for higher a, was observed. An approximation of a more detailed perturbation growth by Carroll et al. (1992) was introduced, which was shown to match the data well. It was also observed that the deviation from linear theory starts happening at around $a \approx 0.5$. Lastly, the conditional probabilities of δ following close to the Carroll prediction, given it was close to it at a different time, were calculated. Interestingly, they revealed that if a region was evolving above the prediction in the past, it is likely to continue to evolve above the prediction in the future as well, but if it

was below the line in the past, it is more likely to cross over it in the future.

To finish, the massive cluster impact on the structure growth was shown by looking at the δ's calculated in the regions around the five most massive halos in the simulation at $z = 0$. It revealed the cluster effect on scales of up to around 10 Mpc/h radius around the halo centre. The structure growth in this region starts accelerating at $a \approx 0.25$ and stays linear (but steeper) for longer.

Following work could involve using the data to study other non-linear structure formation models, like the ones described by Lahav et al. (1991) for example, or further analysing the massive cluster impact on the structure growth around them. Additionally, the effect of the cosmic web on the structure growth could be looked into.

References

Allen, S. W., Evrard, A. E. & Mantz, A. B. (2011), 'Cosmological Parameters from Observations of Galaxy Clusters', *Annual Review of Astronomy and Astrophysics* **49**(1), 409–470.

Begeman, K. G., Broeils, A. H. & Sanders, R. H. (1991), 'Extended rotation curves of spiral galaxies : dark haloes and modified dynamics.', *Monthly Notices of the Royal Astronomical Society* **249**, 523.

Behroozi, P. S., Wechsler, R. H. & Wu, H.-Y. (2013), 'The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores', *The Astrophysical Journal* **762**(2), 109.

Carroll, S. M., Press, W. H. & Turner, E. L. (1992), 'The cosmological constant.', *Annual Rev. Astron. Astrophys.* **30**, 499–542.

Corbelli, E. & Salucci, P. (2000), 'The extended rotation curve and the dark matter halo of M33', *Monthly Notices of the Royal Astronomical Society* **311**(2), 441–447.

Dalcín, L., Paz, R. & Storti, M. (2005), 'MPI for Python', *Journal of Parallel and Distributed Computing* **65**(9), 1108–1115.

Faber, S. M. & Jackson, R. E. (1976), 'Velocity dispersions and mass-to-light ratios for elliptical galaxies.', *The Astrophysical Journal* **204**, 668–683.

Ferreras, I. (2019), *Fundamentals of Galaxy Dynamics, Formation and Evolution*, UCL Press.

Hahn, O. & Abel, T. (2011), 'Multi-scale initial conditions for cosmological simulations', *Monthly Notices of the Royal Astronomical Society* **415**(3), 2101–2121.

Hinshaw, G., Larson, D., Komatsu, E. et al. (2013), 'Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results', *The Astrophysical Journal Supplement Series* **208**(2), 19.

Hunter, J. D. (2007), 'Matplotlib: A 2D graphics environment', *Computing in Science & Engineering* **9**(3), 90–95.

Kapteyn, J. C. (1922), 'First Attempt at a Theory of the Arrangement and Motion of the Sidereal System', *The Astrophysical Journal* **55**, 302.

Lahav, O., Lilje, P. B., Primack, J. R. & Rees, M. J. (1991), 'Dynamical effects of the cosmological constant.', *Monthly Notices of the Royal Astronomical Society* **251**, 128–136.

Lewis, A., Challinor, A. & Lasenby, A. (2000), 'Efficient computation of CMB anisotropies in closed FRW models', *The Astrophysical Journal* **538**, 473–476.

Lundmark, K. (1930), 'Über die Bestimmung der Entfernungen, Dimensionen, Massen und Dichtigkeit fur die nächstgelegenen anagalacktischen Sternsysteme.', *Meddelanden fran Lunds Astronomiska Observatorium Serie I* **125**, 1–13.

Meriot, R., Khochfar, S., Oñorbe, J. & Smith, B. (2022), 'The Cosmic Mach Number as an environment measure for the underlying dark matter density field', *Monthly Notices of the Royal Astronomical Society* **512**(1), 27–40.

Mo, H., van den Bosch, F. & White, S. (2010), *Galaxy Formation and Evolution*, Cambridge University Press.

Natarajan, P., Chadayammuri, U., Jauzac, M. et al. (2017), 'Mapping substructure in the HST Frontier Fields cluster lenses and in cosmological simulations', *Monthly Notices of the Royal Astronomical Society* **468**(2), 1962–1980.

Oort, J. H. (1932), 'The force exerted by the stellar system in the direction perpendicular to the galactic plane and some related problems', *Bulletin of the Astronomical Institutes of the Netherlands* **6**, 249.

Padmanabhan, T. (1993), *Structure Formation in the Universe*, Cambridge University Press.

Planck Collaboration (2020), 'Planck 2018 results - VI. Cosmological parameters', *Astronomy & Astrophysics* **641**, A6.

Springel, V. (2005), 'The cosmological simulation code GADGET-2', *Monthly Notices of the Royal Astronomical Society* **364**(4), 1105–1134.

Tanabashi, M., Hagiwara, K., Hikasa, K. et al. (2018), 'Review of Particle Physics', *Physical Review D* **98**(3), 030001.

Turk, M. J., Smith, B. D., Oishi, J. S., Skory, S., Skillman, S. W., Abel, T. & Norman, M. L. (2011), 'yt: A Multi-code Analysis Toolkit for Astrophysical Simulation Data', *The Astrophysical Journal Supplement Series* **192**, 9.

Zwicky, F. (1933), 'Die Rotverschiebung von extragalaktischen Nebeln', *Helvetica Physica Acta* **6**, 110–127.

Appendices

Note — Appendices are provided for completeness only and any content included in them will be disregarded for the purposes of assessment.

A Overdensity Evolution at Fixed Random Locations

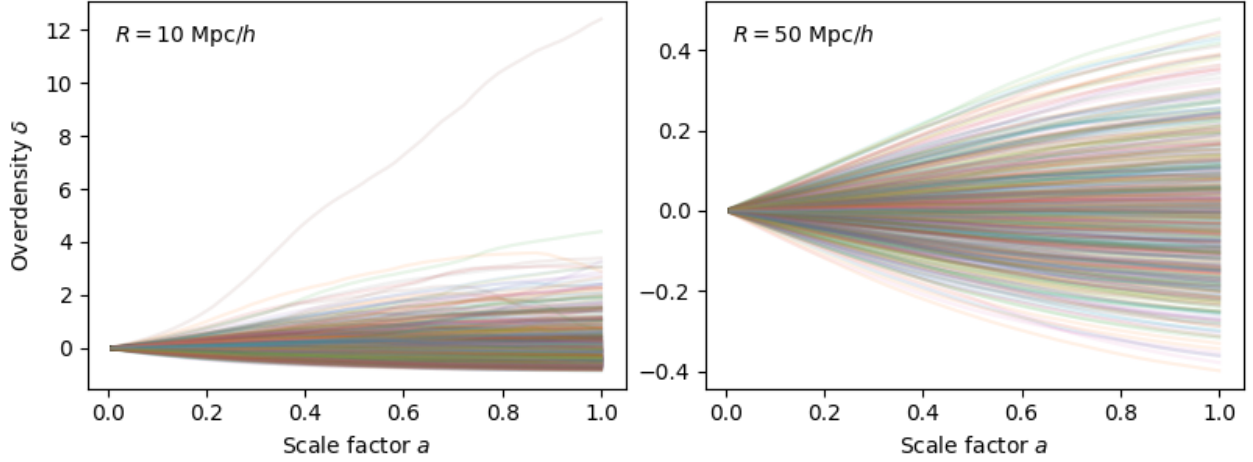The supplementary data to the results in §4.2 is presented in Figure A.1.

Figure A.1. Overdensity evolution with scale factor in one-thousand 10 Mpc/h radius spheres (left) and six-hundred 50Mpc/h radius spheres (right) at fixed random comoving locations. Note the high δ line in the left plot which appears to be from a region containing a massive halo. The same plot for five-hundred 30 Mpc/h spheres and discussion is presented in §4.2.
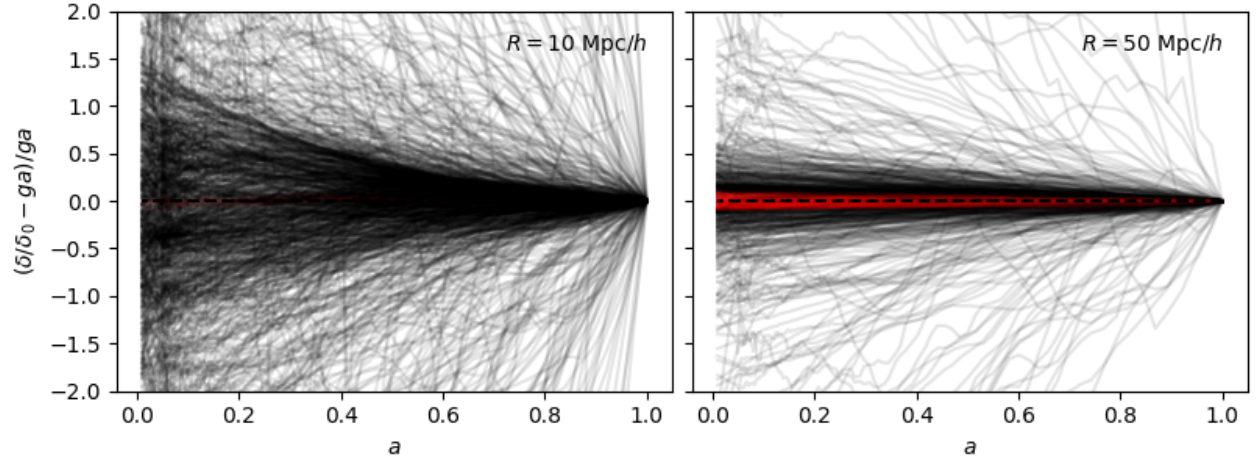
B Residuals of Overdensity Evolution

Figure B.1. Plot of residuals of δ/δ_0 around the Carroll approximation (Eq. 4.1) for one-thousand 10 Mpc/h sphere regions (left) and six-hundred 50 Mpc/h sphere regions (right). 5 10 Mpc/h and 114 50 Mpc/h regions matching the approximation within 10% are in red.