
Dark Matter Halo Density Profile Calibration

Titus N. Nde (Computing Ph.D)

Boise State University

Boise, ID 83725

titusnyarkonde@u.boisestate.edu

November 14, 2022

Abstract

In this synthesis, we review the literature on halo density calibration. Several years of research have been conducted to understand dark matter, but it remains a mystery to Cosmologists. Despite many efforts, there still remain significant gaps in the literature. The ultimate goal is to reconcile dark matter theories (which include fitting functions and simulations) with observations. This will help us obtain an accurate and thorough understanding of dark matter halo density profiles, which will in turn help us understand the formation of cosmic structures. New halo properties have been detected in simulations and observations but are still poorly understood. Analytic fitting functions such as NFW and Einasto profiles have also been shown to perform poorly under several conditions. We do not know yet if new physics can help us understand the mysterious dark matter but probing dark matter halos remains a good approach. This synthesis identifies gaps in the literature and suggests solutions to some of them. We also demonstrate how machine learning can help us calibrate density profiles and also estimate error bars on the predictions. Finally, we present computing artifacts on uncertainty quantification, available in a public GitHub repository via this link.

Keywords: Dark Matter, Halo, Density Profile, Uncertainty Quantification, Machine Learning, Neural Networks, Monte Carlo Dropout, Deep Ensembles, Markov Chain Monte Carlo, Simulation, Splashback, Dilution Effect, Miscentering, NFW Profile, Einasto Profile, DK14 Profile.

1 Introduction

A dark matter halo is a hypothesized region dominated by gravitationally bound invisible matter in which galaxies, galaxy clusters, and groups form and live. Dark matter halos are fundamental building blocks of large-scale structures and the cosmic web in general. They stretch beyond the visible region of galaxies and galaxy clusters. They are dark-matter-dominated structures whose properties help to understand the nature of the mysterious dark matter, galaxy and galaxy cluster formation and evolution, and cosmology in general. According to the Lambda Cold Dark Matter (hereafter Λ CDM) model, our universe is dominated by dark energy ($\sim 68\%$) and dark matter ($\sim 27\%$ and $\sim 85\%$ of all matter in our Universe). Baryons refer to ordinary matter — elements in the periodic table. They constitute only $\sim 5\%$ of the Universe but $\sim 15\%$ of the total matter in the Universe. Note that the Λ CDM model is referred to as the standard model of cosmology and Λ denotes dark energy. The Λ CDM cosmology is dominated by dark energy — this discovery was made in the year 1998 and has revolutionized research in astronomy.

We do not yet understand perfectly the nature of dark matter. As a matter of fact, Cosmologists know much about what it is not rather than what it is. Dark matter is attractive and accounts for structure formation in our Universe. On the other hand, dark energy is repulsive and accounts for cosmic acceleration, from which its existence was inferred.

Measuring the density profiles of Λ CDM halos is one of the interesting probes to cluster cosmology and dark matter. Density profiles offer a promising avenue to deduce dark matter characteristics. Galaxy clusters are the largest objects bound by gravity in the cosmos. Dark matter halo formation is complex but their density profiles help to explain the mass distribution in these gravitationally bound structures and give a practical way to link theory to observations from surveys (Fielder et al., 2020; Nishimichi et al., 2019; Shin et al., 2021; Diemer, 2022). We cannot directly observe cluster mass; however, it can be derived from observable properties such as

cluster richness, velocity dispersion, stellar mass, gravitational lensing, X-ray emission, and Sunyaev-Zeldovich (SZ) effect (Salcedo et al., 2019).

Gravitational lensing is the bending of light traveling from source galaxies to an observer due to the presence of a galaxy cluster acting as a lens between the observer and the source galaxies. If the observer, the lens, and the source are positioned in a straight line, then the observer will see the source as a ring around the lens. Sometimes the observer might even see multiple images of the same source galaxy or the source galaxy appearing as an arc. This is called strong gravitational lensing. However, at larger separations, there is usually a 1% distortion in the shape of the source galaxy. This is called weak gravitational lensing and it is the most predominant form of gravitational lensing in the Universe. Halos are not also observed directly but we can measure their mass distribution and density profiles via results from weak lensing (hereafter WL) measurements like the so-called concentration-mass and richness-mass relations (Henson et al., 2017; McClintock et al., 2019; Shin et al., 2021). WL mass estimation is the most robust of all mass estimation techniques available in the literature (McClintock et al., 2019). Dark-matter-only simulations typically use WL measurements. We can also derive the distribution of source galaxies from photometric redshift measurements.

Fig. 1 shows the simulation of a massive halo. Bright spots in the cosmic web are dark matter halos connected by filaments. Filaments are also dark-matter-dominated structures. Some low-mass halos live inside filaments. The dark regions are voids — empty or devoid of matter.

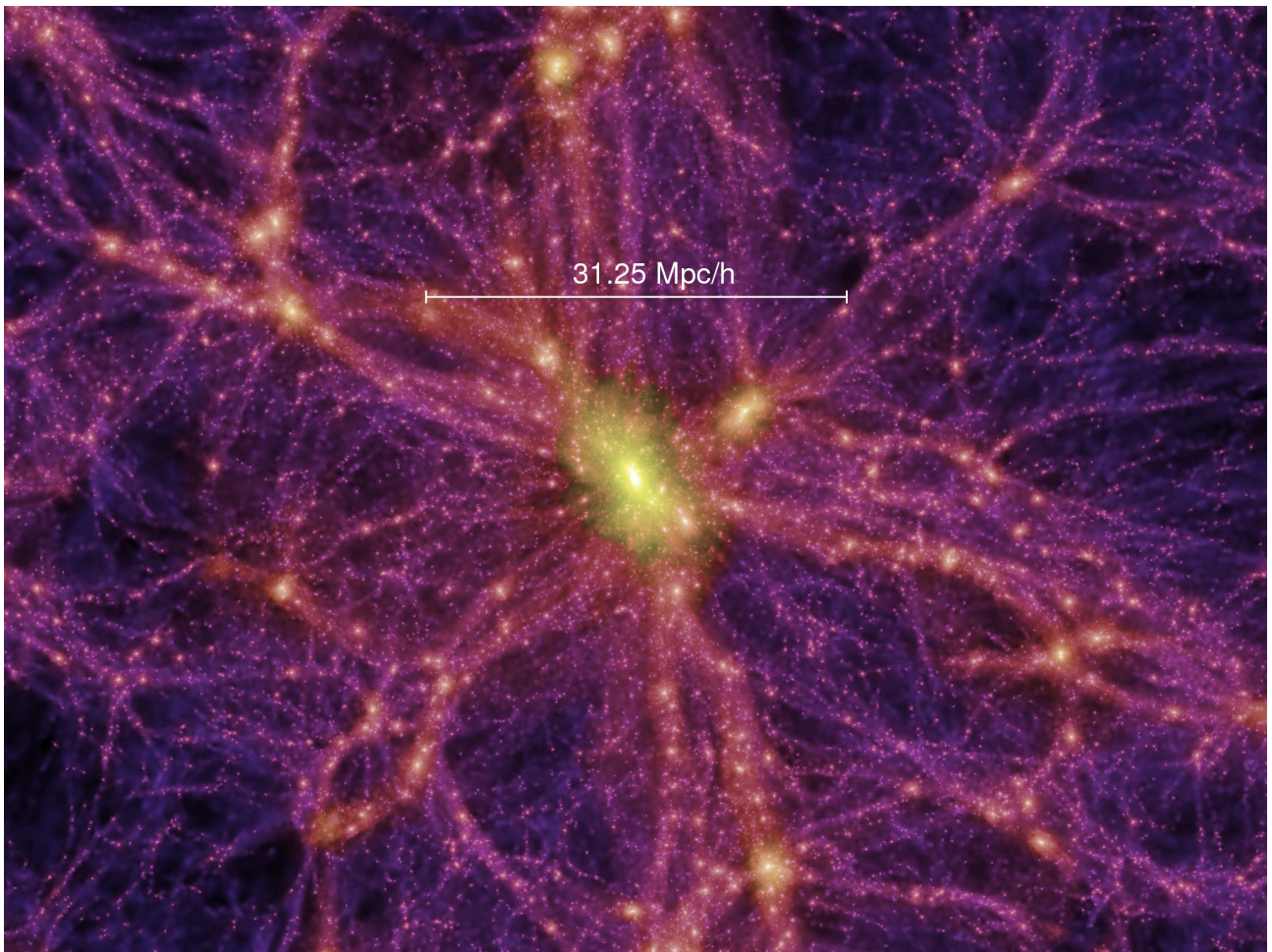


Figure 1: Image of a Λ CDM halo from a simulation. This is a zoom into a really tiny portion of the cosmic web. The distance across the image is about 400 million light-years. The redshift of the halo is 0. The brightest regions are halos connected by filaments. Filaments are also dark-matter-dominated structures. Some low-mass halos live inside filaments. The dark regions are voids — empty or devoid of matter. (Credit: The Millennium Simulation Project.)

Dark matter was inferred from gas and stellar motions in galaxies. For instance, the amount of luminous matter in spiral galaxies cannot explain the observed rotation curves of stars within them. Also, galaxies in

clusters or groups would move away from each other if clusters or groups were made of only luminous matter. This is because the galaxies are moving with very high velocities, and luminous matter alone cannot provide strong enough gravity to hold the galaxies together.

Interestingly, research findings in cosmology are evolving fast. This is because more data is being gathered from several astronomical surveys. Scientists are building sophisticated telescopes to collect extensive high-quality data. As we look further into space, the farther we can see into earlier cosmic time. This means we can observe events that took place from the very beginning of our Universe. Photons from some of these events have not yet traveled to us. Thus, we are moving from an almost purely theoretical phase to an era of vast high-quality data availability.

In Fig. 2, we plot the DK14, NFW, and Einasto profiles for a halo simulated from the Buzzard N-Body simulation (DeRose et al., 2019). We notice that all three profiles do not capture the detailed shape of the halo. Only the logarithmic slope of the DK14 profile shows the presence of a splashback feature. Also, we can only notice differences between NFW and Einasto profiles at small and large radii. The analytic models, as well as simulations, are cosmology-dependent. Differences in cosmology might result in significant differences among simulations and predictions from analytic models. It becomes necessary to understand these differences between cosmologies. Active research is being conducted in this area. Λ CDM halo density profile is one of the areas that will benefit from this. The ultimate end goal is to accurately calibrate halo density profiles as a probe to understand dark matter’s nature and characteristics. The idea is to use observational data to probe and redefine our understanding of existing theories and vice versa (Shin et al., 2021). Much research has been conducted with varying objectives with the sole aim of accurately calibrating Λ CDM halo density profiles.

Another probe to bridging the gap between theory and observation is new physics. Perhaps, a discovery of a new form(s) of dark matter might change our understanding of the Universe and its evolution. Also, we may have to redefine and modify our understanding, as well as existing theories, of gravity (such as rotation in orbits) on cosmological scales to be able to bridge this gap. Though new physics could potentially help us understand our Universe better, most cosmologists tend to adopt explanations offered by dark matter theorists because there is very strong evidence of dark matter in the Universe. Only a few cosmological studies focus on new physics in the literature. However, this remains an active area of research.

This synthesis considers the following seed papers:

- **Paper I:** Dependence of the outer profiles of halos on their mass accretion rate (Diemer & Kravtsov, 2014)
- **Paper II:** The mass and galaxy distribution around SZ-selected clusters (Shin et al., 2021)
- **Paper III:** A dynamics-based density profile for dark halos – I. Algorithm and basic results (Diemer, 2022)
- **Paper IV:** The impact of baryons on massive galaxy clusters: halo structure and cluster mass estimates (Henson et al., 2017)

The remainder of this synthesis is organized as follows. Section 2 talks about how differences in cosmologies impact theories and simulations, Section 3 introduces some important analytic models available in the literature, Section 4 talks about the current state of research on dark matter halo density calibration, Section 5 talks about uncertainty quantification and parameter inference, Section 6 discusses some important results from the literature as well as gaps and future research paths, and Section 7 introduces our computing artifacts and the machine learning models used in the artifacts. Finally, Section 8 concludes.

2 Cosmology, simulations, and terminologies

In this Section, we explore how differences in results obtained from different astronomical surveys impact cosmological simulations. We also give reasons why there is a need to resolve these significant differences.

2.1 Cosmological results

Cosmology refers to the study of the origin, nature, and evolution of the Universe. However, in this synthesis when we say cosmology, we are actually referring to the set of parameters describing the evolution and state of our Universe at the current epoch. Different astronomical surveys were conducted to derive these parameters. They help us study various properties of the evolution of our Universe by deriving cosmological parameters from observational results. Some of the landmark cosmologies usually adopted in the literature include Wilkinson

Microwave Anisotropy Probe (WMAP - Komatsu et al. (2011)) and Planck (Planck Collaboration et al., 2014). However, different astronomical surveys result in different values for this set of parameters. Sometimes these discrepancies become highly statistically significant such that we can no longer neglect them. Two salient examples are the so-called Hubble tension and the σ_8 tension. Interestingly, the same astronomical survey usually results in slightly different values for these parameters from year to year. We do not know which cosmology represents our Universe best. To settle these discrepancies we need careful control of observational systematics. This will also help to set high-accuracy constraints in analytic models, simulations, and existing theories. Table 1 describes the progression of experiments and examples of cosmologies that exist. This table shows the most commonly adopted cosmological parameters for the past two decades. The Einstein-de Sitter cosmology is ruled out but is of theoretical interest since it is scale-free.

Table 1: Cosmological parameters from different astronomical surveys. The table shows the present-day values of the various cosmological parameters describing our Universe as measured from different astronomical surveys (except the Einstein-de Sitter cosmology). The Einstein-de Sitter cosmology is a universe made of 100% matter. This universe is not realistic because structure formation is self-similar over cosmic time scales, but it can be a good proxy for the Λ CDM cosmologies under certain conditions. For details on these parameters from the different astronomical surveys, see Spergel et al. (2003); Komatsu et al. (2011); Planck Collaboration et al. (2014, 2020). The Wilkinson Microwave Anisotropy Probe (hereafter WMAP) and Planck results are based on Λ -dominated cosmologies.

Cosmology	$\Omega_{m,0}^a$	h^b	σ_8^c	n_s^d
Einstein-de Sitter	1.0000	0.7000	0.8200	1.0000
WMAP1 (2003)	0.2700 \pm 0.04	0.7200 \pm 0.05	0.9000 \pm 0.1	0.9900 \pm 0.04
WMAP7 (2011)	0.2743 \pm 0.007	0.7020 \pm 0.014	0.8160 \pm 0.024	0.9680 \pm 0.012
Planck 2018	0.3111 \pm 0.0056	0.6766 \pm 0.0042	0.8102 \pm 0.006	0.9665 \pm 0.0038

^aMatter density parameter

^bHubble constant

^cDensity fluctuation

^dScalar spectral index

2.2 Cosmological simulations

Simulations help us understand the evolution of our Universe from a theoretical perspective. N-body simulations and hydrodynamic simulations are the two major types of cosmological simulations available in the literature. However, different simulations (even those of the same kind) might result in significantly different halo properties due to their dependence on cosmology. Below we briefly describe the two types of simulations.

N-body simulations are dark-matter-only simulations. They assume dark matter halos are made of only dark matter particles. This assumption is not accurate but these simulations yield reasonably good results. N-body simulations are widely used in studies involving dark matter. The Millennium simulation (Springel et al., 2005) is an example of N-body simulation.

Unlike N-body simulations, hydrodynamic simulations account for the contribution of baryons to the mass distribution of dark matter halos. There is a systematic bias toward ignoring baryons and modeling halo density profiles as though they are dark-matter-only structures (Henson et al., 2017; Cromer et al., 2021). Hydrodynamic simulations also tend to produce more spherical and centrally concentrated halos than their dark-matter-only counterparts (Henson et al., 2017). However, the baryonic processes on cosmological scales are not very well understood. Hydrodynamic simulations are computationally expensive because simulating the baryonic effect requires much higher resolutions than gravity and solving extra differential equations.

Baryons dominate halo centers, but their dominance decreases as we move farther from the halo center. There are two major concerns here: On one hand, it is not yet well understood how dark matter is distributed near the halo center, where baryons dominate. On the other hand, it is not yet well understood how baryonic matter is distributed at large radii, where dark matter dominates (Henson et al., 2017; Shin et al., 2021; Diemer, 2022). Hydrodynamic simulations are so important because they have the potential of providing answers to some of these key cosmological quests. Some examples of hydrodynamic simulations available in the literature include the BArYons and HALoes of MAssive Systems (BAHAMAS) (McCarthy et al., 2017) and the MAssive ClusterS and Intercluster Structures (MACSIS) (Barnes et al., 2017) hydrodynamic simulations.

2.3 Halo properties and terminologies

Here we briefly explain some halo properties and terminologies to help the reader easily understand the content of this synthesis.

Redshift: This is the increase in wavelength of photons from luminous bodies as a result of the expansion of the Universe. The stretch in wavelength leads to a drop in frequency as well as photon energy. Redshift is typically used as a measure of distance to luminous bodies. Those farther away have higher redshifts.

Halo radius: This is the radius within which the halo mass is defined. For instance, R_{200m} is the radius that contains on average 200 times the mean matter density of the Universe. M_{200m} is the corresponding mass contained within the sphere of radius R_{200m} . The same definitions can be given to R_{200c} and M_{200c} in terms of the critical density of the Universe, and R_{vir} and M_{vir} in terms of the varying virial overdensity. Diemer & Kravtsov (2014) found that density profiles of the inner and outer regions are most self-similar when density profiles are rescaled by R_{200c} and R_{200m} , respectively. This means that radii defined in reference to the critical density and mean matter density of the universe better describe the inner and outer density profiles, respectively. This also suggests that we need methods to calibrate the two regions with different reference densities.

Halo boundary: This is the largest orbit of matter gravitationally bound to the halo center. Any of the radial definitions given above, R_X , can be used to define a halo boundary with corresponding mass M_X . The region within $r \lesssim R_X$ is the inner region while the region beyond that, $r \gtrsim R_X$, constitutes the outer region of the halo (Diemer, 2022).

1-halo term: This refers to the inner region, $r \lesssim R_X$, of halos. It is so-called because it refers to contributions from matter within the same host halo. This region is dominated by gravitationally bound (orbiting) matter. It is also referred to as the orbiting term.

2-halo term: This refers to the region beyond the 1-halo term, $r \gtrsim R_X$. It is used to describe significant contributions a halo receive from the mass of neighboring massive structures.

Infalling term: The outer region, $r \gtrsim R_X$, of a halo, is dominated by matter falling into it for the first time. This is called infalling matter or the infalling term. Infalling matter does not belong to a neighboring halo, unlike the 2-halo term. Infalling matter is gravitationally unbound (Diemer, 2022).

Splashback feature/region: This refers to a region of steep slope between the inner and outer regions of a halo. The splashback feature is also referred to as the transition region. This is where the orbiting term switches dominance with the infalling term as we move into the outer region. This feature is the result of a complicated interplay/interaction between the orbiting and infalling terms. It is steeper in high-mass halos than in low-mass halos. The radius at which the slope is steepest is called the **splashback radius**, often denoted as R_{sp} (Diemer & Kravtsov, 2014; Shin et al., 2021; Diemer, 2022). The splashback radius can also be used to define a physically significant halo boundary. Diemer & Kravtsov (2014) and Diemer (2022) adopted this definition for the halo boundary. See Fig. 3.

Truncation radius: This is the radius at which the slope of the density profile of the orbiting term truncates/drops sharply, often denoted as r_t . It is often used interchangeably as the splashback radius; however, Diemer (2022) reveals that the two are not exactly the same. Hence, we can only use r_t as a reasonable proxy for R_{sp} . See Fig. 2 and 3.

Scale radius: It is the radius at which the slope of the density profile is equal to -2, often denoted as r_s or r_{-2} . It is the ratio of R_X and the concentration parameter. Both NFW and Einasto profiles are scaled by only r_s (Diemer & Kravtsov, 2014; Fielder et al., 2020).

Richness: It is historically defined as the number of galaxies belonging to a cluster or halo. Modern cluster finders usually use the sum of membership probabilities of galaxies belonging to a certain galaxy cluster or halo.

Mass accretion rate: This is the rate at which a halo accumulates mass. A halo can accumulate mass via means like infalling and mergers (halos merging into one). Diemer & Kravtsov (2014) show that density profiles are sensitive to the total mass accretion rate rather than an instantaneous mass accretion via major mergers. They also show that high-mass halos have a higher mass accretion rate and vice versa. Also, halo masses pseudo-evolve due to changes in the definition of its radius (R_X) and expansion of the Universe (Diemer et al., 2013). Hence, by pseudo-evolution of mass, we mean changes in the halo mass due to the changes in the reference density with respect to which the halo mass is defined (Diemer & Kravtsov, 2014).

Concentration: This is the degree to which the halo mass is dense towards the halo center. It is defined as R_X/r_s . It is dimensionless. High-mass halos usually have a lower concentration and vice versa.

Peak height: This is also a dimensionless parameter. It is closely related to mass: more massive halos have higher peak heights and vice versa. It makes the halo density profile universal and allows for mass comparison across redshift and cosmology. With peak height, we can compare halo masses across different redshifts. Peak height is a function of mass and redshift. Diemer & Kravtsov (2014); Diemer (2022) predict halo density profiles as a function of peak height. This is one of the reasons the DK14 profile is gaining much attention.

Satellite or subhalo: This refers to a halo that lies within the R_X of another halo called the host or parent halo. Parent halos are also called isolated halos.

Critical density: This refers to the matter density that keeps the curvature of our Universe flat.

SZ effect: This is also known as the Sunyaev-Zeldovich effect. The SZ effect refers to spectral distortions in the Cosmic Microwave Background (CMB). It is unaffected by redshift. These spectral distortions are useful in detecting anisotropy in CMB and can be used to pinpoint the positions of galaxy clusters, as well as to infer information about their structure and mass distribution. See Shin et al. (2021).

3 Analytic fitting functions for density profiles

A density profile is a series of density estimates calibrated as a function of radial distance to the halo center. This Section introduces some analytic fitting equations for halo density profiles. Specifically, Subsection 3.1 gives an overview of some 3-dimensional fitting equations while Subsection 3.2 introduces the 2-dimensional density profile.

3.1 3D density profiles

Many analytic models have been developed in the literature to calibrate halo density profiles. They serve as great tools for probing and improving our understanding of the characteristics of the mysterious dark matter. The Navarro-Frenk-White (hereafter NFW) density profile developed by Navarro et al. (1996, 1997) is given as:

$$\rho(r) = \frac{\rho_{crit}\delta_{crit}}{\frac{r}{r_s} \left(1 + \frac{r}{r_s}\right)^2}, \quad (1)$$

where $\rho(r)$ is the halo density, r is the distance from the halo center, $r_s = \frac{R_{200}}{c}$ is the characteristic or scale radius, ρ_{crit} is the critical density of the Universe, and δ_{crit} is the characteristic overdensity. The scale density, ρ_s , is usually used to replace $\rho_{crit}\delta_{crit}$.

The Einasto fitting function developed by Einasto (1965) is given as:

$$\rho(r) = \rho_s \exp \left(-\frac{2}{\alpha} \left[\left(\frac{r}{r_s} \right)^\alpha - 1 \right] \right). \quad (2)$$

The steepening term, α , controls the slope and thus determines the level of decline in the Einasto profile. The bigger the value of α , the steeper the slope of the Einasto profile. It is found to be a function of the dimensionless peak height parameter in Gao et al. (2008) as $\alpha(\nu) = 0.155 + 0.0095\nu^2$.

NFW and Einasto profiles are among the most popular analytic fitting functions in the literature despite some significant drawbacks. They are particularly good for fitting the inner density profiles, $r \lesssim R_{vir}$, of halos. Beyond this radii, they poorly calibrate the density profiles (Diemer & Kravtsov, 2014; Shin et al., 2019; Diemer, 2022).

The DK14 profile developed by Diemer & Kravtsov (2014) is an extension to the Einasto profile. It introduces a new term describing the outer profile of halos. Other studies that precede Diemer & Kravtsov (2014) developed analytic models to describe the outer density profiles of halos. The idea is to prevent the analytic density profile models from approaching zero in the outer regions, $r \gtrsim R_{vir}$. This behavior is typical of the NFW profile for instance. The analytic fitting formulae proposed for describing the outer regions of halos generally tend to the mean matter density at sufficiently large radii (Diemer & Kravtsov, 2014). Prada et al. (2006) suggests adding the mean matter density of the universe to the Einasto profile to describe the outer region; meanwhile, Tavio et al. (2008) rather propose an improvement to the NFW profile. Hayashi & White (2008) also propose an approach to fit the 1-halo term by NFW or Einasto profile, and the 2-halo term by a modified form of the linear mass auto-correlation function. The maximum between the two is adopted in the transition region. This method works pretty well but there is no theoretical justification to explain why it works. On the other hand, Oguri & Hamana (2011) propose adopting the sum of the 1-halo and 2-halo terms to calibrate density profiles. However, their approach has been detected to overestimate the transition region. Despite having all these fitting formulae in the literature, the DK14 profile has gained much attention due to its flexibility in describing the variations in the outer density profiles of Λ CDM halos. The DK14 profile is given as:

$$\rho(r) = \rho_{inner} \times f_{trans} + \rho_{outer} \quad (3)$$

$$\rho(r) = \rho_s \exp\left(-\frac{2}{\alpha} \left[\left(\frac{r}{r_s}\right)^\alpha - 1\right]\right) \times \left[1 + \left(\frac{r}{r_t}\right)^\beta\right]^{-\frac{\gamma}{\beta}} + \rho_m \left[b_e \left(\frac{r}{5R_{200}}\right)^{-s_e} + 1\right]; \quad s_e > 0. \quad (4)$$

This profile tends to the Einasto profile when we omit the outer term and set γ to 0. The outer term, ρ_{outer} , is characterized by a power law plus the mean matter density of the Universe, ρ_m . r_t is the truncation radius and the f_{trans} term helps to explain how the halo profile deepens around the truncation radius. Also, at r_t the slope of the DK14 profile begins to decline more sharply than the Einasto profile. γ is the term that determines the asymptotic negative slope; it defines how the DK14 profile deepens at, $r \approx R_{200m}$, and β defines the sharpness of the change in slope (Diemer & Kravtsov, 2014). Setting $s_e > 0$ constrains the outer term, ρ_{outer} , so that the contribution from infalling matter does not surpass that of the orbiting matter in the inner region, ρ_{inner} . This is because setting $s_e > 0$ means setting an upper bound on the highest value ρ_{outer} can assume (Diemer & Kravtsov, 2014; Shin et al., 2021). According to Diemer & Kravtsov (2014), the profile at sufficiently large radii might not stick to a power law or tend to the mean matter density of the universe. For instance, it will tend to a form consistent with the matter correlation function at $r \gtrsim 9R_{vir}$. Interestingly, it is also possible to parameterize the outer density profile with infalling matter which relies much on the total mass accretion rate and the environment for fast and slowly accreting Λ CDM halos, respectively. See Diemer (2022) for more details on how to model the outer density profile of halos with infalling matter. Also, see Hayashi & White (2008) and Appendix A.1 of Diemer & Kravtsov (2014) for details on how to link the density profile of the outer regions of halos to the 2-halo term.

The DK14 fitting equation is a function of peak height or mass accretion rate. For low peak height halos, NFW and Einasto reasonably calibrate the density profiles but they are unable to capture the sharp decline in slope in the density profiles of high peak height halos around the truncation radius (Diemer & Kravtsov, 2014).

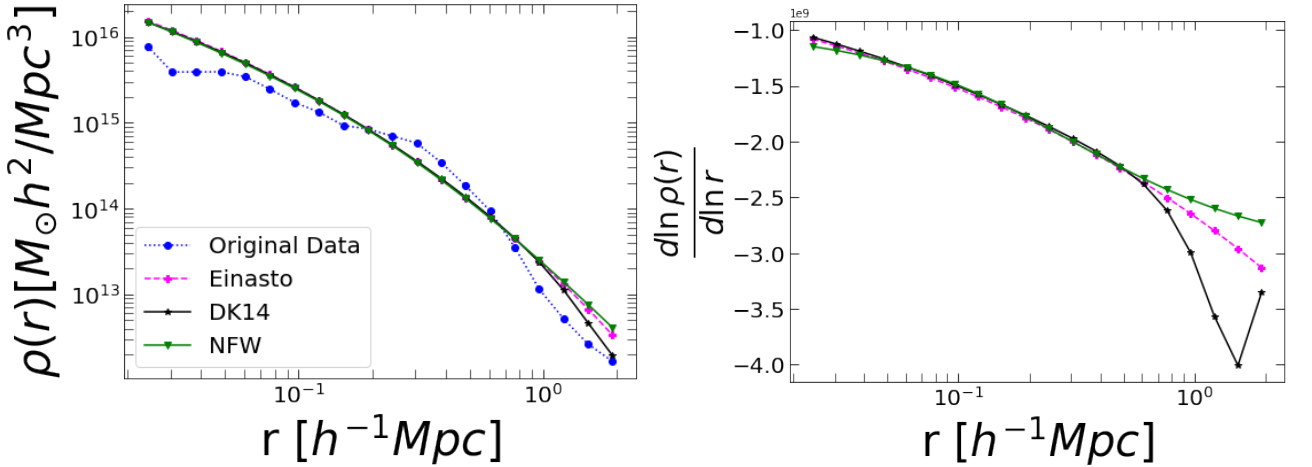


Figure 2: *Left panel:* 3-dimensional DK14, NFW, and Einasto density profiles of a halo from the Buzzard N-body simulation (DeRose et al., 2019). We notice that the analytic models are not able to capture the detailed shape of the actual halo density profile. The halo plotted above has the following properties; $M_{vir} = 3.9 \times 10^{14}$, $z = 0.39$, $c = 3.97$, $R_{vir} = 1.21$, and $\alpha = 0.24$. We adopted the default optimal parameters in COLOSSUS (Diemer, 2018), $\beta = 4$ and $\gamma = 8$, for the DK14 profile. *Right panel:* The logarithmic slope of the analytic density profiles. The clear difference between the Einasto and NFW profiles can be seen at small and large radii, and the point at which the DK14 profile begins to steepen more than the Einasto profile is the truncation radius, r_t . There is no difference between Einasto and DK14 profiles at $r \lesssim r_t$. We can only see the splashback feature from the slope of the DK14 profile.

3.2 2D density profile and gravitational lensing

We can obtain the 2D density profile by integrating the 3D density profile over the line-of-sight distance, l , to the galaxy cluster or halo. The 2D density, Σ , is so important because observationally we see the sky

in 2D. Therefore, we can only calibrate density by surface area rather than volume. 3D densities are only directly available in simulations. Moreover, the 3D density fitting equations also assume dark matter halos are spherically symmetric; though this assumption is often inaccurate (Mandelbaum et al., 2008; Chang et al., 2018; Cromer et al., 2021; Shin et al., 2021). The 2D density equation is given as follows:

$$\Sigma(R) = \int_{-l_{max}}^{l_{max}} \rho \left(r = \sqrt{R^2 + l^2} \right) dl, \quad (5)$$

where R is the projected halo radius. The surface density, Σ , is related to the so-called magnification effect. The magnification effect is a consequence of WL, which makes a galaxy or galaxy cluster appear brighter and bigger than its actual brightness and size, respectively. However, it is difficult to measure since we have to know the size of the galaxy or cluster before the magnification took place.

Therefore, the excess surface density, $\Delta\Sigma$, is even of great importance because it can be measured directly from WL using shear. Gravitational lensing shear is the systematic distortion of images of background galaxies as a result of WL by large-scale objects in the cosmos. Even though $\Delta\Sigma$ is easy to measure, it is rather noisier making it more difficult to model than Σ .

The excess surface density, $\Delta\Sigma$, is given as follows.

$$\Delta\Sigma(R) = \bar{\Sigma}(< R) - \Sigma(R), \quad (6)$$

where $\bar{\Sigma}(< R)$ and $\Sigma(R)$ are the mean surface density within a radius, R , and the projected surface density, respectively. Observationally, we can compute $\Delta\Sigma$ directly from WL tangential shear, γ_t , via the following equation.

$$\Delta\Sigma(R) = \bar{\gamma}_t(R) \Sigma_{crit}(z_l, z_s), \quad (7)$$

where z_l and z_s are the redshifts of the lens halo and the source galaxies, respectively. $\bar{\gamma}_t(R)$ represents the average shear (tangential) at a radial distance of R and $\Sigma_{crit}(z_l, z_s)$ is the critical surface density (Sheldon et al., 2004; Chang et al., 2018; McClintock et al., 2019; Varga et al., 2019; Shin et al., 2021). The tangential shear is a WL signal. It is the tangential component of the ellipticity of the source.

4 Current status of the field

This Section presents the current status of research on dark matter halo density calibration.

4.1 Observations of halo density profiles

We can observe halo density profiles via WL. On the other hand, we can also calibrate density profiles using the number of galaxies per unit area or volume. This method is associated with galaxy formation and evolution and is called the galaxy number density.

To model the observed galaxy density profiles, we can parameterize the number of galaxies per halo of a given mass. This method is a function of mass and is called the Halo Occupation Distribution (hereafter HOD). The HOD is another method for modeling galaxy number density profiles. Shin et al. (2021) used both WL and galaxy density profiles to calibrate halo density profiles and found reasonable agreement between the two profiles.

Observational data from astronomical surveys come with several systematic effects. These systematic effects contaminate the data to some extent. Therefore, there is the need to minimize the level of contamination in the data as much as possible to be able to get the best out of it. Currently, there is an active area of research in cosmology on this subject. See McClintock et al. (2019); Shin et al. (2021) for more details. Research on systematic effects is also geared towards bridging the gap between theory and observations. Bridging this gap will help us make robust predictions as well as accurately calibrate cosmological parameters. In the sub-subsections below, we briefly explain some of these systematic effects.

4.2 Dilution effect

The dilution effect is due to the uncertainties in distance measurements (photometric redshift errors). Photometric redshift is the redshift estimated using the flux (the magnitude/intensity of light) of galaxies as seen via filters. For details on how to measure photometric redshifts, see Drlica-Wagner et al. (2018).

The dilution effect is the situation where galaxies that do not belong to the source/background sample are mistakenly counted as though they belong to it. This mistake is typically associated with foreground galaxies or galaxies that are at the lens. This leads to underestimation of the WL signal because the misclassified galaxies have zero WL signal but are counted as though they have a signal. This contamination of the source sample is called the dilution effect. A good practice is to eliminate galaxies that are within a certain range of redshift, Δz , to the lens from the source catalog to reduce the extent of the contamination (McClintock et al., 2019; Shin et al., 2021). $\Delta z = 0.1$ in the case of Shin et al. (2021). We can also apply a boost factor to correct the contamination. This helps to calibrate more accurate estimates of $\Delta\Sigma$ by reducing the contamination due to the dilution effect (Chang et al., 2018; McClintock et al., 2019; Shin et al., 2021). See Chang et al. (2018); McClintock et al. (2019); Shin et al. (2021) for details on boost factor correction techniques.

4.3 Miscentering effect

Density profiles are calibrated around halo centers. Therefore, correctly identifying the halo center is very important. Usually, the location of the Brightest Cluster/Central Galaxies (BCGs) is taken to be the halo center. This assumption is usually good because the central galaxy is the result of lots of mergers, and therefore it is usually the brightest. However, BCGs are not always located at the center of galaxy clusters. A significant fraction of the clusters are sometimes wrongly centered. Therefore, we cannot ignore the effect of the miscentering on WL measurements. A typical downside of miscentering is that the density profile at small radii is severely/badly altered. A good practice is to write the estimated profile by combining the two factions (correctly and wrongly centered clusters). See Chang et al. (2018); McClintock et al. (2019); Shin et al. (2021) for more details.

A common question that comes to mind is: How do we know that a halo is miscentered if we do not know its original center? One way is to compare our cluster centers to those from X-ray measurements because X-ray cluster centers are very accurate. This is because X-ray-selected clusters have high resolution, and are less affected by systematic effects compared to optically selected clusters. See von der Linden et al. (2014); Zhang et al. (2019) for more details on how to estimate the fraction of miscentered clusters using X-ray data. Miscentering is a major systematic effect to handle.

5 Parameter inference and uncertainty quantification

We discussed in Section 2 how cosmological results differ from one astronomical survey to another and how these differences impact simulations. These differences also impact the parameters of the analytic fitting functions since they are also cosmology-dependent. Cosmologists are therefore interested in making meaningful parameter inferences.

5.1 Likelihood analysis: how do we derive parameter constraints from data?

Bayesian analysis is an important concept in statistical inference. We are usually interested in learning the distribution of the set of parameters, Θ , of a hypothesis given some data, D . Thus, we want to estimate $p(\Theta|D)$. The hypothesis is usually a model. Based on this, we can estimate the most probable values of the set of parameters, Θ , as well as their credible bounds. We can also estimate the likelihood. Generally, the likelihood can be written as $\mathcal{L}(D|\Theta) \equiv p(D|\Theta)$. It is evaluated as a function of Θ given D , which we can maximize to determine the most likely set of model parameters. Note that the likelihood function is not necessarily a probability density function because it is evaluated as a function of Θ instead of the data vector, D . For instance, the integral of the likelihood function over the domain of Θ is not expected to be equal to 1. Thus, the likelihood function is not normalized such that the area under its curve must not necessarily be 1. The idea is to estimate Θ by sampling from the posterior PDF, $p(\Theta|D)$. In cosmology the full data is usually turned into some statistics; for instance, the 2-point correlation function. The idea is to make an inference on Θ by comparing the statistic from the data, D_{data} , and that of the model $D_{model}(\Theta)$, (Boruah et al., 2022). The posterior PDF is related to the likelihood and prior via Baye's theorem as follows.

$$p(\Theta|D_{data}) = \frac{p(D_{data}|\Theta)p(\Theta)}{p(D_{data})} \propto p(D_{data}|\Theta)p(\Theta), \quad (8)$$

where $p(\Theta)$ and $p(D_{data}|\Theta)$ are the prior distribution and the likelihood function, respectively. Equation 8 tells us that the posterior PDF is proportional to the product of the prior and likelihood function. This means

we can randomly sample from the posterior PDF, $p(\Theta|D_{data})$, without necessarily knowing or estimating the evidence, $p(D_{data})$. This is possible because the evidence, acting as the normalization constant, is independent of Θ and generally costly to estimate (Gelman et al., 1995; Press et al., 2007; Verde, 2010; Foreman-Mackey et al., 2013). For a specific case of Gaussian likelihood, the log-likelihood is given as follows.

$$\ln[\mathcal{L}(D_{data}|D_{model}(\Theta))] \equiv \ln[p(D_{data}|\Theta)] = -\frac{1}{2} \underbrace{[D_{data} - D_{model}(\Theta)]^T C^{-1} [D_{data} - D_{model}(\Theta)]}_{\chi^2(\Theta)} + \text{constant}, \quad (9)$$

where C^{-1} is the inverse covariance matrix. Note that $C = \langle (D_{data} - \bar{D}_{data})(D_{data} - \bar{D}_{data})^T \rangle$. The posterior PDF defined on the parameters of the model then becomes;

$$\ln[p(\Theta|D_{data})] = \ln[\mathcal{L}(D_{data}|D_{model}(\Theta))p(\Theta)], \quad (10)$$

In cosmology, Bayesian data analysis helps to marginalize over nuisance parameters, denoted as α . We can think of nuisance parameters as extra model parameters, except that they are not cosmological parameters and are less interesting to a general audience. Also, by marginalization, we mean integrating over the set of values of a parameter and passing on its uncertainties to the final outcome (Hogg et al., 2010; Foreman-Mackey et al., 2013). The integral could be cumbersome if we have many nuisance parameters. The result of marginalization is the marginalized PDF, given below.

$$p(\Theta|D_{data}) = \int p(\Theta, \alpha|D_{data}) d\alpha, \quad (11)$$

where α is the collection of nuisance parameters.

5.2 Sampling the posterior: MCMC

The Markov Chain Monte Carlo (hereafter MCMC) method is the most commonly used technique for uncertainty quantification and parameter inference in the literature (Spergel et al., 2003; Hogg et al., 2010; Komatsu et al., 2011; Foreman-Mackey et al., 2013; Shin et al., 2021; Cromer et al., 2021). MCMC is a Bayesian inference technique. MCMC is a well-known algorithm but with lots of variations (Gelman et al., 1995; Press et al., 2007). The most popular package in cosmology is the emcee (Foreman-Mackey et al., 2013).

Because astrophysical models often cost a lot to compute and have many free parameters, MCMC is frequently utilized in this field. Even when the parameter space has a large dimension and is difficult to sample from, MCMC methods are utilized to estimate sampling approximations to a posterior PDF (Gelman et al., 1995; Press et al., 2007; Foreman-Mackey et al., 2013). The logic behind MCMC is practically important because the PDF of the data/evidence, is usually very costly to calibrate; hence, the MCMC technique allows us to sample from the posterior PDF, without necessarily knowing or estimating the evidence, $p(D_{data})$. This means we can essentially sample from any probability distribution using MCMC. This property of MCMC is crucial because using various priors results in various posteriors (Press et al., 2007; Verde, 2010). Once the MCMC samples are generated, we can compute the expectation of a function of the model parameters as follows.

$$\langle f(\Theta) \rangle = \int p(\Theta|D_{data}) f(\Theta) d\Theta \approx \frac{1}{T} \sum_{i=1}^T f(\Theta_i), \quad (12)$$

where T is the number of dependent samples drawn. The Metropolis-Hastings approach is the most popular MCMC method in the literature. The Metropolis-Hastings algorithm refers to a class of Markov chain simulation techniques that are effective for selecting samples from Bayesian posterior distributions (Gelman et al., 1995; Press et al., 2007; Verde, 2010). Below we describe how the Metropolis-Hastings algorithm works.

1. Choose a starting point, say Θ^0 from an initial distribution, $p_0(\Theta)$, such that $p(\Theta^0|D_{data}) > 0$.
2. For $t = 0, 1, 2, 3, \dots$:
 - For every step, t , randomly select Θ^* from the proposal distribution, $J_t(\Theta^*|\Theta^t)$. The proposal distribution can take any form, say, the multivariate Gaussian distribution centered at Θ^t . However, the steps are expected to exhibit some form of ergodicity in the parameter space — the steps will eventually visit every point in the parameter space of interest.

- Estimate the acceptance probability using the following:

$$\beta(\Theta^*, \Theta^t) = \min \left(1, \frac{p(\Theta^*|D_{data})/J_t(\Theta^*|\Theta^t)}{p(\Theta^t|D_{data})/J_t(\Theta^t|\Theta^*)} \right) \quad (13)$$

- Generate a random number from the uniform distribution, $r \leftarrow R \sim U[0, 1]$.
- Decision rule:

$$\Theta^{t+1} = \begin{cases} \Theta^* & \text{if } r \leq \beta(\Theta^*, \Theta^t) \\ \Theta^t & \text{otherwise} \end{cases} \quad (14)$$

- Now, adjust $t = t + 1$.
- We can then compute the transition probability as

$$p(\Theta^{t+1}|\Theta^t) = J_t(\Theta^{t+1}|\Theta^t)\beta(\Theta^{t+1}, \Theta^t) \quad \text{for all } \Theta^{t+1} \neq \Theta^t. \quad (15)$$

Generally, the distribution of the states, $\Theta^0, \Theta^1, \Theta^2, \dots, \Theta_T$, are expected to approach $p(\Theta|D_{data})$. In general, the proposal distribution and the number of iterations, T , are free parameters, which must be determined to suit the specific problem in question. Also, the number of iterations necessary to sufficiently estimate $p(\Theta|D_{data})$ depends on a number of factors, such as the relationship between the proposal and posterior distributions and the lag/order of the autocorrelation time (for discrete state space probability distributions). See Gelman et al. (1995); Press et al. (2007) for more details on this subject.

There is a special case of the Metropolis-Hastings algorithm called the Gibbs sampler, which utilizes conditional distributions rather than the full joint PDFs to generate the random draws. Details on the Gibbs sampler are captured in Gelman et al. (1995); Press et al. (2007).

5.3 An example

This Subsection describes briefly how Shin et al. (2021) used the framework described above to quantify parameter uncertainties. Their data vector, D_{data} , includes the excess surface density, $\Delta\Sigma$, and the galaxy number density, Σ_g . For D_{model} , they adopted two main models. First, the DK14 profile (discussed in Section 3), but with the outer term calibrated by a second-order perturbation theory. Second, the Halo Occupation Distribution (HOD) based analytic model. See Subsection 4.4.2 of Shin et al. (2021) and Cooray & Sheth (2002) for details on the HOD model. They adopt a Gaussian likelihood function (equation 9) and Gaussian priors on the models' parameters. The priors adopted on the set of parameters are shown in Table 1 of Shin et al. (2021).

To calculate the posterior, they adopt the affine invariant version of MCMC proposed by Goodman & Weare (2010). This algorithm is implemented via the emcee package (Foreman-Mackey et al., 2013), available in Python. An advantage of the affine invariant MCMC algorithm is that it has been shown to significantly improve the generic Metropolis-Hastings algorithm. It generates independent samples with lesser autocorrelation time. A detailed discussion of the affine invariant MCMC is beyond the scope of this synthesis, we refer the reader to the original paper by Goodman & Weare (2010).

Specifically, for mass profiles from WL, Shin et al. (2021) used emcee to randomly sample from the posterior distribution on the model parameters. For $\Delta\Sigma$, they used the 1σ (68% credible interval) interval of the posterior as the measure of uncertainty. The logarithmic slope of the 3D profile is then computed using the MCMC constraints set on the parameters of the halo model. Thus, the logarithmic slope of the 3D density is inferred from the 2D density profile, $\Delta\Sigma$, and the 1σ interval is adopted as the measure of uncertainty. The aim of the 3D profile is to estimate the location of the splashback radius. The uncertainty on the location of the splashback radius is also inferred. The 1σ credible interval on all model parameters can be found in Table 2 of Shin et al. (2021). The same strategy is used to estimate uncertainties in galaxy number density profiles, Σ_g .

6 Discussion of results, gaps, and future paths

Several papers focus on different objectives in halo density calibration; however, some issues of concern keep emerging in their analyses. In this Section, we discuss some useful results and concerns raised in the literature.

Limitation and uncertainties of fitting functions of density profiles: Fielder et al. (2020); Dutton & Macciò (2014); Chemin et al. (2011); Henson et al. (2017) compared the performance of the NFW and Einasto profiles and found Einasto superior to NFW. However, Chemin et al. (2011); Dutton & Macciò (2014); Diemer & Kravtsov (2014); Henson et al. (2017); Shin et al. (2021); Diemer (2022) have all shown that the two models poorly calibrate halo density profiles. In the case of Henson et al. (2017) the two underestimate cluster masses. Diemer (2022) attributed their poor performance to the dependence of the halo density profiles on different spatial scales (truncation radius, r_t , and scale radius, r_s). These two scales depend on different halo properties (r_s depends on the radial definition, R_X , while r_t depends on the total mass accretion rate). This means we cannot accurately model the density profiles with universal single-scale fitting functions like the NFW and Einasto density profiles. Shin et al. (2021) use a modified form of the DK14 profile with second-order perturbation theory to calibrate the outer profile but got poor calibrations from using this technique. The current form of the DK14 profile does not account for infalling matter in calibrating the outer density profiles. Diemer (2022) is in a series of three papers, and the second paper (yet to be published) in the series proposes a new fitting function as an improvement to the DK14 profile to account for factors such as the impact of infalling matter on the density profiles of the outer and the transition regions. There is also more scatter in the infalling term than in the orbiting term (Shin et al., 2021; Diemer, 2022). If we rely on analytic models, the uncertainties will be captured by the posteriors of the model parameters. The MCMC is commonly used but has been shown to be computationally expensive (Foreman-Mackey et al., 2013; Cromer et al., 2021). Henson et al. (2017) use the bootstrap resampling technique to quantify prediction uncertainty.

Contributions from subhalos and/or low-mass halos: Generally, there is a bias in WL mass estimation for low-mass halos. This is because the 2-halo term has a significant impact on the mass distribution and density profiles of low-mass halos than high-mass halos. Low-mass halos also have small statistical uncertainties but strong systematic uncertainties. Observationally, low-mass halos are hard to observe. Studies such as Dutton & Macciò (2014); Chemin et al. (2011); Fielder et al. (2020); Shin et al. (2021) did not consider low-mass halos for these reasons. Simulations also have mass limits because low-mass halos are not well-resolved and difficult to tell whether they are real clusters or not (DeRose et al., 2019). Halos of mass $\lesssim 10^{12}$ are usually considered low-mass halos. However, the mass limits differ for different studies because the mass limit in halos is dependent on the resolution. We suggest combining high-resolution, smaller-volume simulations with modest-resolution, large-volume simulations.

On the other hand, less attention is given to the contribution of subhalos to their parent halos. Subhalos account for a portion of the mass in their parent halos and make the density profiles less smooth. Fielder et al. (2020) present a possible remedy by excluding the mass associated with subhalos from the total halo mass. This practice smoothens the halo density profile and makes it easier to model. However, this still needs much attention. For instance, in the case of Fielder et al. (2020), it is sometimes difficult to tell whether a particle belongs to the host halo or the satellite halo.

Baryonic contribution: Current findings in the literature show that significant baryonic contributions to low-mass halos have been established (Henson et al., 2017). However, we do not know exactly the impact of baryons on high-mass halos. If we do find a significant baryonic impact on high-mass halos, then this will pose serious problems to dark-matter-only mass estimation methods such as WL (Becker & Kravtsov, 2011; Oguri & Hamana, 2011; Henson et al., 2017).

Resolving observational systematics: There is no unambiguous way of handling systematic effects. We need more efficient modeling strategies to fill this gap. Simulations can help in this regard. One way is to simulate the systematics and analyze the data as though there are no systematics. If the effect is minimal then those systematics do not matter a lot and vice versa. This way, we will be able to distinguish between systematic errors that count and those that do not count very much. Another way is to introduce systematics into the data and then use simulations to recover them. This will help us test the robustness of existing theories (models and simulations). Sheldon et al. (2004); Mansfield et al. (2017); Varga et al. (2019); McClintock et al. (2019); Shin et al. (2021) discuss several methods of handling observational systematics and the pros and cons of each method.

Limited sample sizes: SZ-selected data, hydrodynamic simulations, and X-ray data are limited in their sample sizes (from a statistical point of view). For instance, for hydrodynamic simulations, Henson et al. (2017) and Cromer et al. (2021) used sample sizes of 390 (for the MACSIS simulation) and 1400 clusters, respectively. Thus, it is difficult to tell if conclusions drawn from such studies are significant or just a consequence of the

limited sample sizes. Perhaps, more data could lead to different results such as the impact of baryons on high-mass halos. Despite this, it is good to state that some important results were obtained from such studies. On the other hand, optically selected clusters result in millions of data in observations and simulations (Rykoff et al., 2014, 2016; McClintock et al., 2019; Zhang et al., 2019; Diemer, 2022), but they suffer much from systematic effects.

Splashback feature: The splashback feature, which is dependent on mass accretion rate or peak height, has recently been discovered in simulations (Diemer & Kravtsov, 2014; Diemer, 2022) and observations Shin et al. (2021). However, this feature is still poorly understood (Diemer & Kravtsov, 2014; Shin et al., 2021; Diemer, 2022). Thus, there is ambiguity concerning the location of the splashback radius. Shin et al. (2021) found its location to be consistent in WL and galaxy number density profiles with 20% and 9% uncertainty, respectively. The truncation radius has also been shown to usually occur at a relatively smaller radius than the splashback radius (Diemer & Kravtsov, 2014; Diemer, 2022). Recently the splashback radius is gaining much attention in the literature because it defines a physically accurate halo boundary than the other radial definitions in the literature. It defines the largest particle orbit separating the inner and outer halo regions. Diemer & Kravtsov (2014); Diemer (2022) recommend adopting the splashback radius as a more physically reasonable halo boundary because halo density profiles are sensitive to the definition of the halo boundary. The problem is that the other radial definitions have the tendency of resulting in a halo boundary that does not contain all particles that are in orbit to the halo center. Nonetheless, we still do not know how far the outer region should extend to adequately separate the effect of infalling matter from that of the 2-halo term (Diemer, 2022).

Fig. 3 shows the inner and outer regions of a dark matter halo density profile and the location of the splashback feature. This result is obtained from observational data — from the Atacama Cosmology Telescope (ACT). The halo shown below has a mass of $2.72 \times 10^{14} h^{-1} M_{\odot}$ and the redshift, $z = 0.455$.

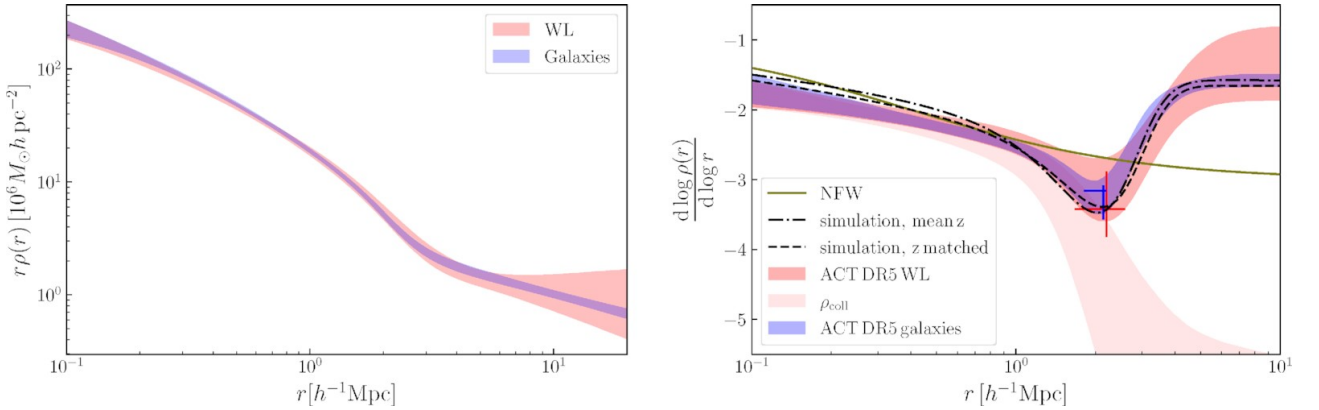


Figure 3: *Left panel:* 3D profiles calibrated from galaxy number density and WL. There is much scatter in the outer region compared to the inner region. Also, the fit of the two methods reasonably agrees with each other. The red-shaded region corresponds to the inferred 3D profile (described in Subsection 5.3) from WL and the blue-shaded region corresponds to the 3D profile, also inferred from the galaxy number density profile in a similar approach. *Right panel:* 3D logarithmic slope of the profile inferred from WL (red) and galaxy number density profiles (blue). The location of the splashback radius (along the horizontal axis) and the associated slope (along the vertical axis) are consistent between the WL estimate (red cross) and that of the galaxy number density estimate (blue cross). The crosses indicate a 1σ credible interval on the location and slope of the splashback radius. We notice from the plot that the truncation in the orbiting term, ρ_{coll} , occurs before the location of the splashback radius. The light-red shaded region is the slope of the inner density profile. The truncation in this curve provides evidence of the splashback feature. The dark curves depict the average density profiles of haloes matched by their mass from N-body simulations. The green slope is the NFW profile fitted with the same mean mass and redshift as the data. The NFW profile could not capture the shape of the splashback feature and the outer profile. (Credit: Shin et al. (2021))

7 Computing artifacts

In this Section, we present a Machine Learning (ML) approach to quantifying halo density profiles. We are building an emulator for calibrating halo density profiles. The ensemble methods described in this Section are implemented in our computing artifacts, available on GitHub.

In the GitHub repository, we present Jupyter Notebooks for dark matter halo density calibration. The data preparation methods applied to the data are available in the data preparation notebook. Another notebook is presented on ML uncertainty quantification methods. This notebook demonstrates how ML ensemble methods can help us quantify uncertainties in DM halo density predictions. The ML methods adopted include Monte Carlo dropout and deep ensembles. A description of these models is given in Subsections 7.3 and 7.4. The models are supervised and are trained on 80% of the data to predict log densities. We take the logarithm to reduce the scale/spread of the densities. The remaining 20% (consisting of about 2.2 million data points) is set aside for testing. The data set is arranged according to radial bins and halo IDs. Also, high-mass halos come before low-mass halos. Hence, the data is shuffled to prevent the ML models from learning these intrinsic patterns embedded in the data. Most of the error bars generated by the ML models capture the actual data points. This can also be seen via the chi-square measure on the uncertainties as it reveals that the error bars are realistic. The deep ensembles model has a longer training time than the Monte Carlo dropout model. Deep ensembles also have the highest adjusted R-squared value. All the details can be found in the uncertainty quantification notebook in the GitHub repository. A *README.md* file is also added to explain how to navigate the repository.

Moreover, we fit the analytic NFW, Einasto, and DK14 models to the test set and compared the predictions to that of the ML models. The ML models outperform the analytic models as measured by the Mean Squared Errors (MSE). The analytic models are fitted with the same cosmology as the Buzzard simulation. Among the analytic models, the DK14 profile has the smallest MSE

7.1 Data Preparation

The data is from the Buzzard simulation. This simulation is based on the DES project. It is the DES Synthetic Sky Catalog (DeRose et al., 2019). The data contains halo properties such as the mass, radius, redshift, scale radius, and parent ID (PID). The PID tells whether a halo is a parent or satellite halo. Each halo also has a unique halo ID. This helps to pull a halo’s records across separate data files. The data is simulated in two simulation boxes of sizes 1050 and 2600, with redshifts ranging from 0.18 to 0.32 and 0.32 to 0.68, respectively. Each halo has simulated 3D densities at 20 different radial bins, but the inner two radial bins were dropped because they have a lot of zero values. Also, the data is binned into redshift bins and halos belonging to bins in the redshift range 0.33 to 0.37 are dropped because most of the halos in this range have zero densities. This kind of contamination usually happens in the region between the two simulation boxes. We convert the data from comoving to physical units to allow for easy comparison between our density predictions and that of the COLOSSUS package (available in Python). Physical units take into account the expansion of the universe but comoving units do not.

We convert mass and redshift to peak height to allow for easy comparison across redshifts. We then binned the data into four evenly spaced peak height bins and sampled equally from each bin. The sample size in each bin is the size of the smallest peak height bin. The samples are chosen to ensure that the mean in each peak height bin is not changed significantly. We verified this using a t-test. All these are shown in the data preparation notebook on GitHub. The final data set contains 630458 halos (about 11.3 million data points since each halo has 18 different densities — corresponding to the number of radial bins). We take the ratio of the middle of the radial bins and the halo radius to make the halo profile more self-similar in peak height bins. This ratio is used as one of the predictors in the ML models.

7.2 Machine learning uncertainty quantification

Deep neural networks are popular due to their predictive capabilities, but quantifying the reliability of their predictions still remains an active area of research. Uncertainty quantification techniques help to measure a model’s confidence in its predictions. We expect the measure of uncertainty to be low when the model’s level of confidence about its prediction is high and vice versa. In regression-based ensemble methods, it is a common practice to use the mean and variance/spread as the prediction and the measure of uncertainty, respectively. For classification problems, one way is to allow the models in the ensemble to vote and the class with the majority votes wins. Entropy and max-probability are the most popular uncertainty measures for classification ensembles (Durasov et al., 2020). Deep ensembles, Monte Carlo dropout, and Bayesian neural networks are

among the most popular uncertainty quantification techniques. However, this paper uses only deep ensembles and Monte Carlo dropout.

The Bayesian neural network offers a framework that is mathematically robust in quantifying model uncertainties but it is rather computationally expensive compared to the models adopted in this synthesis. Monte Carlo dropout was introduced as an approximate Bayesian method. Deep ensembles, on the other hand, was introduced as a conceptually simple alternative to the Bayesian neural network (Caldeira & Nord, 2020; Lakshminarayanan et al., 2016; Durasov et al., 2020).

Generally, we expect the aggregate prediction from the ensemble methods to be significantly better than that of any single model or network in isolation. To the author’s best knowledge, this paper is novel such that no paper has adopted deep learning models in predicting dark matter halo density profiles.

7.3 Deep ensembles (DE)

The basic idea is to simply retrain the same neural network architecture many times with different random initialization in each training. These random initialization impact the training process and returns several independent models. These models have different sets of weights and biases making them independent; in effect, DE fits a finite number of independent models. It is possible to add an extra source of randomness by using the bagging or bootstrap aggregation technique. This is simply training each randomly initialized neural network with different subsets of the training data by sampling with replacement from it, but this method has been found to have deteriorating performance (Caldeira & Nord, 2020; Lakshminarayanan et al., 2016; Durasov et al., 2020). DE is one of the best approaches to quantify model uncertainty; however, it is rather computationally costly to train and evaluate and expensive in terms of memory usage. Thus, DE is not a suitable candidate for memory-constrained applications. In practice, DE is also known to have significantly better performance than Monte Carlo dropout in uncertainty quantification tasks with regards to prediction accuracy and error bars; this is attributed to the fact the weight configurations in Monte Carlo dropout results in insufficiently diverse predictions than that of DE thereby underestimating the resulting error bars for the Monte Carlo dropout model (Durasov et al., 2020). DE is illustrated in Fig. 4.

7.4 Monte Carlo dropout (MCD)

Originally, MCD (or simply dropout) is a regularization technique but this method now serves as a deep learning ensemble technique. The basic idea is to randomly drop a percentage of neurons at each layer for every pass during training (Hinton et al., 2012). However, Gal & Ghahramani (2015) shows that using dropout at inference time is a useful way to quantify prediction uncertainties. They introduce the theoretical and mathematical framework for MCD as approximate Bayesian inference in deep Gaussian processes. Thus, training a dropout network mathematically converges to Bayesian inference. This approximation serves as a useful alternative to Bayesian models. A major factor to consider in MCD is the percentage of neurons to drop. Gal et al. (2017) focuses on finding the optimal dropout percentages at each layer during training. Srivastava et al. (2014) provides several experimental results and gave some useful heuristics about the dropout technique. They include;

- using a very high dropout rate causes the model to under-learn while a very small dropout rate has little to no effect on the model; hence, a reasonable dropout rate should be in the range 20%-50% with 20% being a reasonable start
- use a reasonably large network
- use dropout on all units/layers
- use large learning rates with decay and high momentum

MCD is a popular and less expensive alternative to DE but it is rather less reliable as it turns to underestimate the uncertainties. Unlike DE, MCD uses a correlated set of infinite models. This is because we train a single model, randomly switch off a fraction of nodes and use the resulting models to make predictions. This process can be repeated infinite times. The resulting models are not fully connected neural networks. The models share a fraction of neurons (weights and biases). Thus, MCD trains just a single stochastic network but achieves randomness by randomly eliminating a different set of neurons several times. Doing this during testing results in correlated models. To this end, it becomes possible to make multiple predictions for every data point (Durasov et al., 2020). MCD is illustrated in Fig. 5.

8 Conclusion

In this synthesis, we discussed halo density profiles and their usefulness in cluster cosmology. Vital information useful for understanding cluster cosmology have recently been detected in observations and simulations. Some of these signals are detected around R_{vir} and beyond. Prominent among them is the splashback feature and truncation of the orbiting term. This is becoming a milestone in studies involving dark matter; however, these new discoveries have also created many unanswered questions. Recent studies such as Diemer & Kravtsov (2014); Shin et al. (2021); Diemer (2022) are dedicated to understanding these Λ CDM halo properties using various simulated and observational data.

Several astronomical surveys are still being conducted with the aim of obtaining high-quality cosmology data. Notable among them include the recently launched James Web Space Telescope, the upcoming Vera Rubin Observatory in Chile (expected in 2023), and NASA's Nancy Grace Roman Space Telescope (expected in 2027). All of these are geared toward better understanding our Universe. Some of the previous astronomical surveys include the Dark Energy Survey (DES) (The Dark Energy Survey Collaboration, 2005), the Atacama Cosmology Telescope (ACT) survey (Hilton et al., 2021), and the Sloan Digital Sky Survey (SDSS) (Aihara et al., 2011).

This synthesis demonstrates two ML approaches for calibrating halo density profiles. The methods adopted are capable of estimating prediction uncertainties. We also present computing artifacts available in a public GitHub repository via this link.

Acknowledgements

Thanks to my supervisor, Dr. Wu, for the great support in preparing for my comprehensive exam. I also express my profound gratitude to Dr. Andrius Tamosiunas, Gladys Muthoni Kamau, and the Committee Members (Dr. Mead, Dr. Jackson, Dr. Ekstrand, and Dr. Wright) for their contribution to this synthesis paper.

References

- Aihara H., et al., 2011, *ApJS*, 193, 29
- Barnes D. J., Kay S. T., Henson M. A., McCarthy I. G., Schaye J., Jenkins A., 2017, *MNRAS*, 465, 213
- Becker M. R., Kravtsov A. V., 2011, *ApJ*, 740, 25
- Boruah S. S., Eifler T., Miranda V., M S. K. P., 2022, arXiv e-prints, p. arXiv:2203.06124
- Caldeira J., Nord B., 2020, arXiv e-prints, p. arXiv:2004.10710
- Chang C., et al., 2018, *ApJ*, 864, 83
- Chemin L., de Blok W. J. G., Mamon G. A., 2011, *The Astronomical Journal*, 142, 109
- Cooray A., Sheth R., 2002, *Phys. Rep.*, 372, 1
- Cromer D., Battaglia N., Miyatake H., Simet M., 2021, arXiv e-prints, p. arXiv:2104.06925
- DeRose J., et al., 2019, arXiv e-prints, p. arXiv:1901.02401
- Diemer B., 2018, *The Astrophysical Journal Supplement Series*, 239, 35
- Diemer B., 2022, *MNRAS*, 513, 573
- Diemer B., Kravtsov A. V., 2014, *The Astrophysical Journal*, 789, 1
- Diemer B., More S., Kravtsov A. V., 2013, *ApJ*, 766, 25
- Drlica-Wagner A., et al., 2018, *ApJS*, 235, 33
- Durasov N., Bagautdinov T., Baque P., Fua P., 2020, arXiv e-prints, p. arXiv:2012.08334
- Dutton A. A., Macciò A. V., 2014, *Monthly Notices of the Royal Astronomical Society*, 441, 3359
- Einasto J., 1965, *Trudy Astrofizicheskogo Instituta Alma-Ata*, 5, 87
- Fielder C. E., Mao Y.-Y., Zentner A. R., Newman J. A., Wu H.-Y., Wechsler R. H., 2020, *MNRAS*, 499, 2426
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
- Gal Y., Ghahramani Z., 2015, arXiv e-prints, p. arXiv:1506.02142
- Gal Y., Hron J., Kendall A., 2017, *Concrete Dropout*, doi:10.48550/ARXIV.1705.07832, <https://arxiv.org/abs/1705.07832>
- Gao L., Navarro J. F., Cole S., Frenk C. S., White S. D. M., Springel V., Jenkins A., Neto A. F., 2008, *MNRAS*, 387, 536
- Gelman A., Carlin J. B., Stern H. S., Rubin D. B., 1995, *Bayesian Data Analysis*, 1 edn. Chapman and Hall, <https://www.amazon.com/Bayesian-Analysis-Chapman-Statistical-Science/dp/0412039915>
- Goodman J., Weare J., 2010, *Communications in Applied Mathematics and Computational Science*, 5, 65
- Hayashi E., White S. D. M., 2008, *Monthly Notices of the Royal Astronomical Society*, 388, 2
- Henson M. A., Barnes D. J., Kay S. T., McCarthy I. G., Schaye J., 2017, *MNRAS*, 465, 3361
- Hilton M., et al., 2021, *ApJS*, 253, 3
- Hinton G. E., Srivastava N., Krizhevsky A., Sutskever I., Salakhutdinov R. R., 2012, arXiv e-prints, p. arXiv:1207.0580
- Hogg D. W., Bovy J., Lang D., 2010, arXiv e-prints, p. arXiv:1008.4686
- Komatsu E., et al., 2011, *ApJS*, 192, 18

- Lakshminarayanan B., Pritzel A., Blundell C., 2016, arXiv e-prints, p. arXiv:1612.01474
- Mandelbaum R., Seljak U., Hirata C. M., 2008, *J. Cosmology Astropart. Phys.*, 2008, 006
- Mansfield P., Kravtsov A. V., Diemer B., 2017, *ApJ*, 841, 34
- McCarthy I. G., Schaye J., Bird S., Le Brun A. M. C., 2017, *MNRAS*, 465, 2936
- McClintock T., et al., 2019, *MNRAS*, 482, 1352
- Navarro J. F., Frenk C. S., White S. D. M., 1996, *The Astrophysical Journal*, 462, 563
- Navarro J. F., Frenk C. S., White S. D. M., 1997, *The Astrophysical Journal*, 490, 493
- Nishimichi T., et al., 2019, *The Astrophysical Journal*, 884, 29
- Oguri M., Hamana T., 2011, *Monthly Notices of the Royal Astronomical Society*, 414, 1851
- Planck Collaboration et al., 2014, *A&A*, 571, A16
- Planck Collaboration et al., 2020, *A&A*, 641, A6
- Prada F., Klypin A. A., Simonneau E., Betancort-Rijo J., Patiri S., Gottlober S., Sanchez-Conde M. A., 2006, *The Astrophysical Journal*, 645, 1001
- Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 2007, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, 3 edn. Cambridge University Press, http://www.amazon.com/Numerical-Recipes-3rd-Scientific-Computing/dp/0521880688/ref=sr_1_1?ie=UTF8&s=books&qid=1280322496&sr=8-1
- Rykoff E. S., et al., 2014, *ApJ*, 785, 104
- Rykoff E. S., et al., 2016, *ApJS*, 224, 1
- Salcedo A. N., Wibking B. D., Weinberg D. H., Wu H.-Y., Ferrer D., Eisenstein D., Pinto P., 2019, *Monthly Notices of the Royal Astronomical Society*, 491, 3061
- Sheldon E. S., et al., 2004, *AJ*, 127, 2544
- Shin T., et al., 2019, *MNRAS*, 487, 2900
- Shin T., et al., 2021, *Monthly Notices of the Royal Astronomical Society*, 507, 5758
- Spergel D. N., et al., 2003, *ApJS*, 148, 175
- Springel V., et al., 2005, *Nature*, 435, 629
- Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R., 2014, *J. Mach. Learn. Res.*, 15, 1929–1958
- Tavio H., Cuesta A. J., Prada F., Klypin A. A., Sanchez-Conde M. A., 2008, arXiv e-prints, p. arXiv:0807.3027
- The Dark Energy Survey Collaboration 2005, arXiv e-prints, pp astro-ph/0510346
- Varga T. N., et al., 2019, *MNRAS*, 489, 2511
- Verde L., 2010, in Wolschin G., ed., , Vol. 800, *Lecture Notes in Physics*, Berlin Springer Verlag. pp 147–177, doi:10.1007/978-3-642-10598-2_4
- Zhang Y., et al., 2019, *MNRAS*, 487, 2578
- von der Linden A., et al., 2014, *MNRAS*, 443, 1973

Fig. 4 illustrates the DE method. We train a finite set of independent fully connected deep learning models and use them to predict each data point.

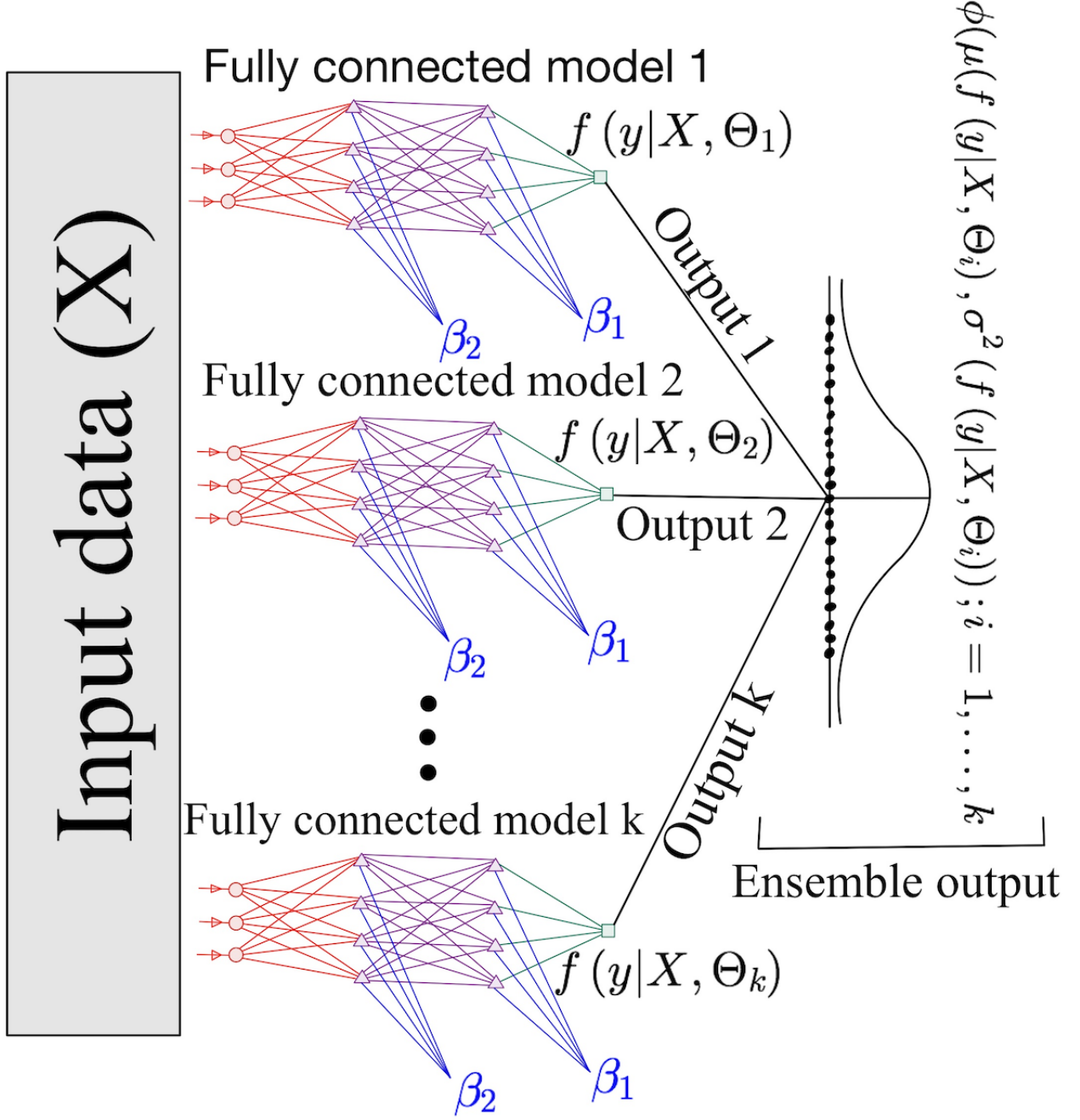


Figure 4: Illustration of DE. The same neural network architecture is retrained several times with different initialization. This produces independent models with completely non-overlapping (disjoint) parameters. All the models will be used to make predictions. The number of times we train the neural network architecture with different random initialization is the number of independent models we will have. We can only make a finite number of predictions up to the number of models trained. We can form a distribution from all the predictions for a certain observation from which we obtain the aggregate prediction of the ensemble. All models, $f(y|X, \Theta_i); i = 1, 2, \dots, k$, are independent; hence, they do not share any fraction of their parameters/weights, $\Theta_i \neq \Theta_j, \forall i \neq j; i, j = 1, 2, \dots, k$.

Fig. 4 illustrates the MCD method. We train a single fully connected deep learning model; however, we can use it to make an infinite number of predictions for each data point.

Single trained fully connected model

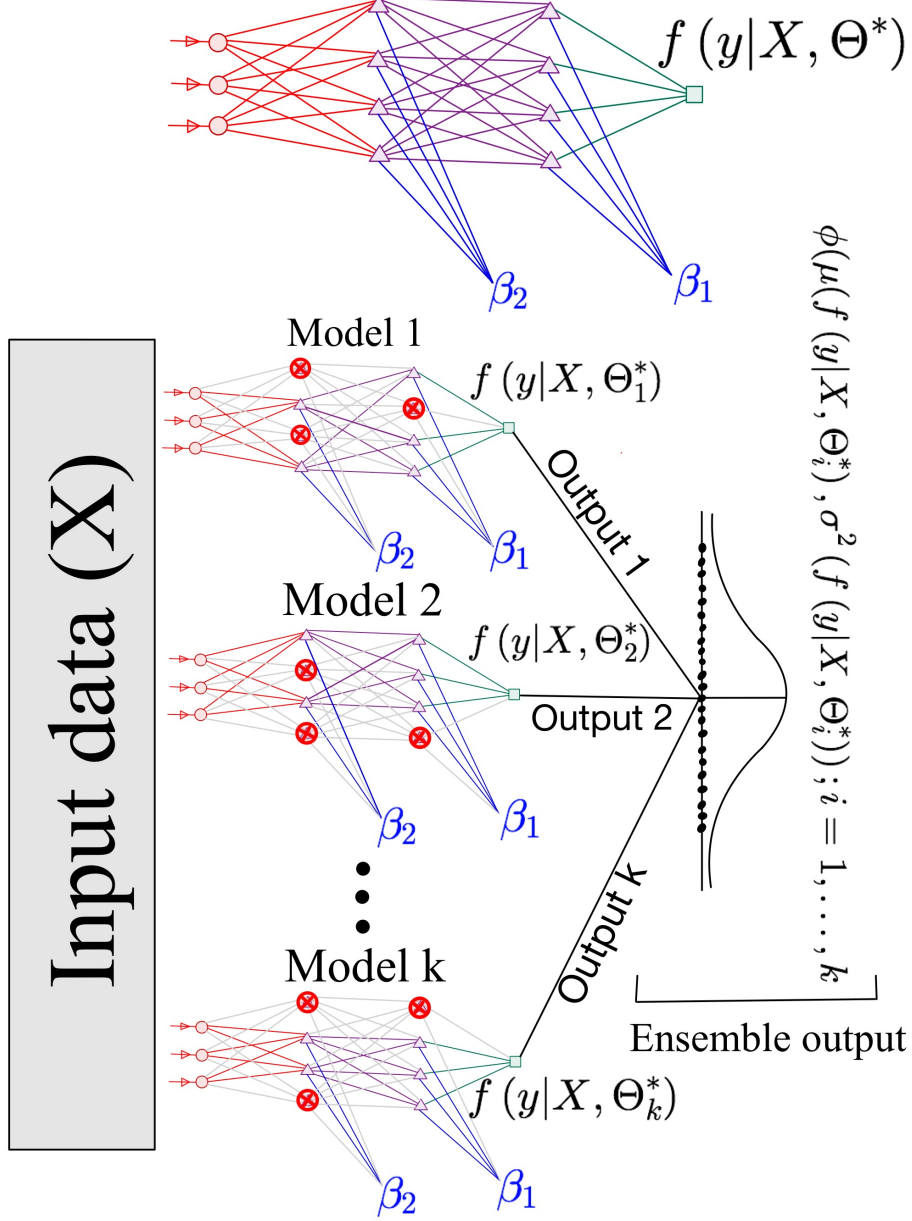


Figure 5: Illustration of MCD. The Figure shows a fully connected neural network model (top) and networks with some nodes switched off (bottom). In this diagram, the red nodes were randomly switched off. Hence, they do not receive any information from the nodes in the previous layer(s). They do not transmit any information to the nodes in the subsequent layer(s) as well. Their contribution is eliminated from the network and this is represented by the gray lines entering or going out from these switched-off nodes. Here, we train a single neural network architecture and then randomly switch off a fraction of the nodes several times and use it to make predictions each time. Thus, we can use the single-trained neural network model to make an infinite number of predictions. We can form a distribution from all the predictions for a certain observation from which we obtain the aggregate prediction of the ensemble. MCD results in a correlated set of models. Thus, every two models, $f(y|X, \Theta_i^*), f(y|X, \Theta_j^*)$; $i \neq j$; $i, j = 1, 2, \dots, k$, share a fraction of the set of their parameters/weights, Θ_i^*, Θ_j^* ; $i \neq j$; $i, j = 1, 2, \dots, k$.