Bridging Theory and Observation: Advancing Cosmology and Astrophysics through Constrained Simulations and Machine Learning

Elena Hernández Martínez

München 2025

Bridging Theory and Observation: Advancing Cosmology and Astrophysics through Constrained Simulations and Machine Learning

Elena Hernández Martínez

Dissertation
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Elena Hernández Martínez
aus Bilbao

München, den 13.05.2025

Erstgutachter: Prof. Dr. Klaus Dolag
Zweitgutachter: Prof. Dr. Jochen Weller
Tag der mündlichen Prüfung: 27.06.2025

Zusammenfassung

Galaxienhaufen sind die größten gravitativ gebundenen und virialisierten Strukturen im Universum und bieten zahlreiche Möglichkeiten zur Untersuchung astrophysikalischer Prozesse und kosmologischer Parameter. Die präzise Gewinnung von Informationen aus Galaxienhaufen ist jedoch nach wie vor eine große Herausforderung, da ihre hochgradig nichtlineare Natur und die begrenzte Auflösung und Modellierungstiefe numerischer Simulationen eine physikalisch akkurate und sinnvolle Verbindung mit Beobachtungen erschweren. Diese Dissertation adressiert diese Problematik durch die Kombination von *constrained* kosmologischen Simulationen und simulationsbasierter Inferenz mittels maschinellen Lernens.

Im ersten Teil stellen wir eine neuartige Reihe hydrodynamischer *constrained* Simulationen vor, das SLOW-Projekt (Simulating the LOcal Web), das darauf ausgelegt ist, die großräumige Struktur des lokalen Universums im Eins-zu-eins-Vergleich realitätsnah nachzubilden. Wir zeigen, dass diese Simulationen die Positionen und Eigenschaften von über 46 beobachteten Galaxienhaufen mit hoher Genauigkeit reproduzieren. Insbesondere stimmen die Massen, Temperaturen, Röntgenleuchtkräfte und Sunyaev-Zel'dovich Effect (SZ)-Signale der simulierten Haufen sehr gut mit den Beobachtungen überein, was den Einsatz *constrained* Simulationen auf Basis von Eigengeschwindigkeiten zur Rekonstruktion realer Strukturen im lokalen Kosmos eindrucksvoll bestätigt.

Im zweiten Schritt analysieren wir die thermodynamischen Eigenschaften dieser simulierten Haufenrepliken. Dabei zeigt sich, dass die Profile auf großen und mittleren Skalen gut modelliert werden. In den Kernregionen treten jedoch deutliche Abweichungen auf, insbesondere bei Entropie- und Gasdichtewerten. Dies unterstreicht die Notwendigkeit verbesserter Auflösung und einer weiterentwickelten Modellierung der Galaxienentstehung. Darüber hinaus stellen wir einen Zusammenhang zwischen den beobachteten thermodynamischen Eigenschaften lokaler Galaxienhaufen und deren Masseaufbau-Historien in den Simulationen her, wodurch wir eine Verbindung zwischen der Entstehung eines Haufens und seiner späteren Einordnung als *Cool-Core-* oder *Non-Cool-Core*-System aufzeigen.

Abschließend demonstrieren wir, dass maschinelles Lernen, trainiert auf gemittelten thermodynamischen Profilen aus hochaufgelösten Zoom-in-Simulationen von Galaxienhaufen, sowohl kosmologische als auch astrophysikalische Parameter mit bemerkenswerter Genauigkeit rekonstruieren kann. Mithilfe der CAMELS-zoomGZ-Simulationsreihe, welche den 28-dimensionalen Parameterraum des IllustrisTNG-Modells abdeckt, erreichen wir eine bislang unerreichte Inferenzgenauigkeit. Wir zeigen, dass diese Inferenz auch bei starkem Rauschen und begrenzter radialer Abdeckung stabil bleibt. Zudem analysieren wir, wie die Inferenz basierend auf gemittel-

ten Profilen den Ansatz auf Basis integrierter Größen deutlich übertrifft. Somit zeigen wir, dass Galaxienhaufenprofile eine besonders leistungsstarke Datengrundlage für simulationsbasierte Inferenz bilden.

Insgesamt liefert diese Arbeit einen kohärenten Rahmen, der die Struktur und Entwicklung von Galaxienhaufen mit den zugrundeliegenden physikalischen Parametern verbindet. Durch die enge Verknüpfung von Simulationen, Beobachtungen und maschinellem Lernen legt diese Dissertation das Fundament für eine neue Generation präziser Kosmologie, welche den astrophysikalischen und kosmologischen Informationsgehalt von Galaxienhaufen vollständig ausschöpft.

Abstract

Galaxy clusters are the largest gravitationally bound and virialized structures in the Universe and powerful probes of both astrophysical processes and cosmological parameters. However, extracting precise information from galaxy clusters remains challenging due to their highly non-linear nature and the difficulty of connecting resolution and model-limited simulations with observations in a physically accurate and meaningful way. This thesis addresses these challenges by combining the use of constrained cosmological simulations and simulation-based inference using machine learning.

We first present a novel suite of constrained hydrodynamical simulations, the SLOW (Simulating the LOcal Web) project, designed to reproduce the large-scale structure of the local Universe on a one-to-one basis. We show how these simulations accurately replicate the positions and properties of over 46 observed galaxy clusters. We establish that the clusters' masses, temperatures, X-ray luminosities, and SZ signals are matched with high fidelity, validating the use of constrained simulations based on peculiar velocities as tools for creating replicas of real cosmic structures present in our local Universe.

In a second step, we analyze the thermodynamic properties of these simulated cluster replicas, revealing that large-scale and intermediate radial profiles are well modeled. However, core regions show strong discrepancies, specially in the entropy and gas density values. We will discuss how these point to the necessity of improvements in resolution as well as improvements in the current implementation of galaxy formation models. Additionally, we will link the observed thermodynamic features of local clusters to the clusters' mass assembly histories recovered in the simulations, establishing a connection between the formation path of clusters and their late-time cool-core/non-cool-core classification in observations.

Finally, we demonstrate that machine learning models trained on stacked thermodynamic profiles from high-resolution zoom-in simulations of galaxy clusters can recover both cosmological and astrophysical parameters with remarkable accuracy. Using the CAMELS-zoomGZ suite, which spans the 28-dimensional parameter space from the IllustrisTNG model, we achieve an unprecedented inference accuracy for cosmological parameters as well as astrophysical parameters. We will show how the inference remains robust under high noise levels and limited radial coverage. Moreover, we will explore how inference done using stacked galaxy cluster profiles outperforms the inference done with integrated quantities, positioning cluster profiles as a powerful tool for simulation-based inference.

Together, the results of this thesis provide a unified framework that connects the structure and evolution of galaxy clusters with the underlying physical parameters that govern them. By

bridging simulations, observations, and machine learning-based inference, this work aims to lay the foundation for a new generation of precision cosmology using galaxy clusters as astrophysical and cosmological laboratories.

Contents

List of Acronyms

AMR Adaptive Mesh Refinement

ANN Artificial Neural Network

AGN Active Galactic Nucleus

AGB Asymptotic Giant Branch

BAO Baryon Acoustic Oscillations

BCG Brightest Cluster Galaxy

BH Black Hole

BCG Brightest Cluster Galaxy

BH Black Hole

CCT Central Cooling Time

CGM Circumgalactic Medium

CC Cool-Core

CMB Cosmic Microwave Background

CR Constrained Realizations

CNN Convolutional Neural Network

DL Deep Learning

NN Deep Neural Network

ISM Interstellar Medium

EQS Effective Equation of State

FCN Fully Connected Network

FLRW Friedmann-Lemaître-Robertson-Walker

FoF Friends-of-Friends

GP Gaussian Process

IC Initial Condition

ICM Intracluster Medium

IMF Initial Mass Function

LSS Large Scale Structure

MHD Magnetohydrodynamics

ML Machine Learning

NN Neural Network

NCC Non-Cool-Core

SBI Simulation-Based Inference

SGD Stochastic Gradient Descent

SCC Strong Cool-Core

SGD Stochastic Gradient Descent

SPH Smoothed Particle Hydrodynamics

SMBH Supermassive Black Hole

SN Supernova

SNe II Type II Supernovae

SNe Ia Type Ia Supernovae

SZ Sunyaev-Zel'dovich Effect

tSZ thermal Sunyaev-Zel'dovich Effect

WCC Weakly Cool-Core

"Isn’t it funny how day by day nothing changes, but when you look back, everything is different?"

— commonly attributed to C.S. Lewis

Part I

Introduction

The overarching objective of cosmology is to achieve a comprehensive understanding of the Universe and its evolution. This involves tracing the development of its matter and energy components from the very beginning, nearly 14 billion years ago, to the present day. In modern cosmology, the formation of structure in the Universe is understood to result from the gravitational amplification of tiny fluctuations in a generally smooth and uniform matter distribution. These small irregularities are believed to have originated from quantum fluctuations in a scalar field, which triggered a rapid phase of accelerated expansion known as cosmic inflation (see Guth, 1997). Following this early phase, the Universe emerged from an extremely hot and dense state, and has been expanding ever since.

One of the most extensively studied pieces of evidence supporting this picture is the Cosmic Microwave Background (CMB) radiation. First detected by Penzias & Wilson (1965), the CMB is the earliest observation of our Universe, providing an image of the Universe when it was only about 380,000 years old. The CMB is characterized by its general homogeneity and isotropy, with tiny perturbations which provide the necessary seeds for the cosmic structures that we observe today. Detailed measurements of these small perturbations in the CMB have been carried out by numerous experiments (e.g., Fixsen et al., 1996; Torbet et al., 1999; Melchiorri et al., 2000; Hanany et al., 2000; Hinshaw et al., 2007; Planck Collaboration et al., 2020a). These small perturbations are believed to have evolved in the web of structures that we observe in the late-time large-scale structure of the Universe.

This arrangement of matter is often referred to as the cosmic web, and it forms a vast network connecting dense regions, filled with galaxies and stars, through elongated filaments, which are separated by large, underdense voids. The formation and evolution of this cosmic web is explained commonly with the ΛCDM (Lambda Cold Dark Matter) model. This model is widely accepted not only because it successfully accounts for a wide range of observations, but also due to its relative simplicity. It assumes that at large-scales the Universe is homogeneous and isotropic (as supported by the CMB) and the validity of general relativity. With this assumptions it is possible to describe the Universe with just six key parameters: matter density ($\Omega_{\rm m}$), baryon density ($\Omega_{\rm b}$), dark energy density (Ω_Λ), the scalar spectral index (n_s), the amplitude of matter fluctuations (σ_8), and the Hubble constant (H_0).

A key method for investigating the fundamental components of the Universe is through the study of how matter and radiation are distributed on the largest scales in the comic web, in particular at the intersections of the filaments of the web, what we call the nodes, where clusters of galaxies reside. These structures are massive gravitationally bound systems having characteristic masses ranging from 10^{14} to 10^{15} $M_\odot$ and containing hundreds or even thousands of galaxies. They consist of both baryonic and dark matter and represent the largest virialised structures in the Universe. Their formation is believed to be closely tied to the gravitational collapse of the most massive overdensities present in the primordial density field (see e.g. Press & Schechter, 1974a; Springel et al., 2005c; Tinker et al., 2008; Kravtsov & Borgani, 2012a) and thus, the number and mass distribution of these structures are highly sensitive to cosmological parameters, making them essential for constraining models of cosmic evolution (see e.g. Vikhlinin et al., 2009; Allen et al., 2011; Mantz et al., 2015; Planck Collaboration et al., 2016a).

The galaxies that reside within clusters also provide valuable insights into the processes governing structure formation. For instance, their orbits within the cluster's gravitational potential,

carry evidence of how clusters assemble through hierarchical merging (see e.g. Dressler, 1980; De Lucia & Blaizot, 2007; Kravtsov & Borgani, 2012b). At the same time, the dense environments in which these galaxies evolve subject them to processes such as tidal stripping, ram pressure and shocks (Gunn & Gott, 1972; Binney & Tremaine, 1987; Abadi et al., 1999; Markevitch et al., 2002; Roediger & Brüggen, 2007; Tonnesen & Bryan, 2012), which can significantly affect their properties and evolution.

For these reasons, galaxy clusters offer a unique setting to explore a broad range of astrophysical phenomena taking place in some of the most extreme environments in our Universe. The combination of complexity and scale that galaxy clusters offer makes them exceptional laboratories for advancing both astrophysical and cosmological research.

Providing a complete description of these systems, however, requires an accurate treatment of the highly non-linear processes involved in gravitational collapse, together with the heating and cooling mechanisms governing baryonic matter. As clusters assemble, the intracluster gas is heated to extremely high temperatures, sufficient for it to emit X-rays via thermal bremsstrahlung, through a combination of adiabatic compression and shock heating. Once confined within the gravitational potential well of the cluster, the gas tends to settle into a state of hydrostatic equilibrium.

As the intracluster gas becomes denser, radiative cooling processes become more efficient, allowing the formation of stars and fueling accretion onto supermassive black holes (SMBHs) that reside at the centers of massive galaxies within the cluster. This in turn triggers feedback mechanisms, such as energy injection from supernovae (SNe) and active galactic nuclei (AGN), which heat the intracluster gas and distribute metals throughout the cluster environment.

Understanding the highly non-linear processes involved requires sophisticated modeling methods. Semi-analytic models (SAMs) and numerical simulations are widely used to study these systems, as they allow us to follow the dynamics of cluster assembly and the influence of baryonic physics on the large-scale properties of these structures.

Through the analysis of an extended set of standard and constrained cosmological hydrodynamical simulations, this thesis aims to study the hot gas properties of galaxy clusters and their use for cosmological parameter inference.

The thesis is structured as follows:

Part II provides the theoretical and observational background necessary to contextualize the work. It reviews the standard cosmological model, the formation of large-scale structure, and the physical processes that shape the thermodynamic and observable properties of galaxy clusters.

Part III presents the methodological framework used in the thesis. This includes a description of the cosmological simulations employed, both traditional and constrained, alongside details of the machine learning architectures used for inference, and the techniques applied to construct merger trees and analyze cluster evolution.

Part IV contains the main results and is divided into three research chapters. Chapter 4 focuses on the construction and validation of the SLOW simulation suite, a set of constrained simulations designed to reproduce the observed structure of the local Universe. Chapter 5 analyzes the thermodynamic profiles of the hot gas of the simulated local clusters, comparing them to X-ray and SZ observations, and investigates the connection between core properties and cluster formation histories. Chapter 6 applies machine learning to stacked thermodynamic profiles from

zoom-in simulations to infer both cosmological and astrophysical parameters, demonstrating the power of cluster profiles as tools for simulation-based inference.

Finally, Part V summarizes the main findings of the thesis and discusses several future directions. These include high-resolution simulations with varied physics models, inference based on 2D cluster maps, and the integration of constrained zoom-in simulations for cosmological parameter inference.

Part II

Scientific Background

1 | Cosmology and Large-Scale Structure

The standard cosmological model, known as the ΛCDM model, offers a remarkably accurate description of our Universe. This chapter introduces several fundamental concepts essential to this model.

For clarity, the following conventions are adopted:

- Natural units are used where $c = \hbar = 1$.
- Greek indices (α, β) run over spacetime coordinates, while Latin indices (j, k) indicate spatial components.
- The Einstein summation convention applies throughout.
- The metric signature is $(-, +, +, +)$.
- A dot over a variable denotes differentiation with respect to physical time.
- Distances are expressed in Mpc/h and masses in $M_\odot/h$, with h defined as: $h \equiv \frac{H_0}{100 kms^{-1} Mpc^{-1}}$.
- Important constants in SI (Workman et al., 2022):
  - speed of light: $c = 299{,}792{,}458\,\mathrm{m/s}$,
  - gravitational constant: $G = 6.67430 \times 10^{-11}\,\mathrm{m^3\,kg^{-1}\,s^{-2}}$,
  - parsec: $1\,\mathrm{pc} = 3.08567758149 \times 10^{16}\,\mathrm{m}$,
  - solar mass: $M_\odot = 1.98841 \times 10^{30}\,\mathrm{kg}$.

1.1 The ΛCDM Universe

In the following, we will present the formalism related to the standard cosmological model of the Universe. The summaries covered here and in subsection 1.1.1 are based on the works by Mukhanov (2005); Carroll (2019); Baumann (2022); Dodelson & Schmidt (2024); Turner (2024).

The ΛCDM model is built on the *cosmological principle*, which is the assumption that, when viewed on sufficiently large scales, the Universe is both homogeneous and isotropic. Moreover, gravity is considered the only relevant interaction, with general relativity as the theory that governs

it. Other interactions such as electromagnetism are considered to play a role only on smaller scales compared to gravity and to affect baryonic matter but not CDM.

The connection between the curvature of spacetime and the matter-energy content of the Universe is captured by Einstein's field equations:

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{1.1}$$

where G is Newton's gravitational constant and $T_{\mu\nu}$ represents the energy-momentum tensor that encodes the distribution of matter and energy throughout the Universe. The cosmological constant Λ can be incorporated either by including a term $\Lambda g_{\mu\nu}$ on the left-hand side, or by treating it as part of $T_{\mu\nu}$.

In this context, the metric tensor $g_{\mu\nu}$ describes spacetime geometry, while $R_{\mu\nu}$ and R correspond to the Ricci tensor and Ricci scalar, respectively, with $R = R^{\mu}{}_{\mu}$. The Ricci tensor itself can be written explicitly as:

$$R_{\mu\nu} = \frac{\partial \Gamma^{\alpha}_{\mu\nu}}{\partial x^{\alpha}} - \frac{\partial \Gamma^{\alpha}_{\mu\alpha}}{\partial x^{\nu}} + \Gamma^{\alpha}_{\beta\alpha}\Gamma^{\beta}_{\mu\nu} - \Gamma^{\alpha}_{\beta\nu}\Gamma^{\beta}_{\mu\alpha}, \tag{1.2}$$

where the Christoffel symbols Γ are defined in terms of the metric components as:

$$\Gamma^{\mu}_{\alpha\beta} = \frac{1}{2}g^{\mu\lambda}\left(\frac{\partial g_{\lambda\beta}}{\partial x^{\alpha}} + \frac{\partial g_{\lambda\alpha}}{\partial x^{\beta}} - \frac{\partial g_{\alpha\beta}}{\partial x^{\lambda}}\right). \tag{1.3}$$

Equation 1.1 expresses how matter and energy influence the geometry of spacetime determined by the right hand side of the equation. The metric tensor $g_{\mu\nu}$ determines the line element:

$$ds^2 = g_{\mu\nu}dx^{\mu}dx^{\nu}. \tag{1.4}$$

Given that the cosmological principle presumes large-scale homogeneity and isotropy, the spacetime metric must reflect these symmetries, i.e. translational and rotational symmetry. The metric that satisfies these conditions is unique and is called the Friedmann-Lemaître-Robertson-Walker (FLRW) metric. Its line element is given as:

$$ds^2 = -dt^2 + a^2(t)\left[\frac{dr^2}{1 - Kr^2} + r^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right)\right], \tag{1.5}$$

where (r, θ, ϕ) are *comoving* polar coordinates, and $a(t)$ is the *scale factor* that describes the scaling of spatial distances with respect to a given distance at a chosen initial time t_0. By convention this time is chosen to be our current time, so that $a(t_0) \equiv a_0 = 1$. The scale factor relates the physical coordinates x_{phys} to the comoving coordinates q_{com} through $x_{\text{phys}} = a(t)\, q_{\text{com}}$. This means that if two objects are not subject to any net force, they will maintain fixed positions in comoving coordinates, in other words, they move along with the cosmic expansion, and their physical separation increases proportionally to $a(t)$. Finally, the parameter K indicates the spatial curvature of the Universe:

$$K = \begin{cases} 1 & \text{closed Universe (spherical geometry)} \\ 0 & \text{flat Universe (Euclidean geometry)} \\ -1 & \text{open Universe (hyperbolic geometry)} \end{cases} \tag{1.6}$$

Throughout this work, we will assume a flat Universe ($K = 0$), in concordance with theoretical predictions from inflation (Mukhanov, 2005) as well as with the latest observational data (Planck Collaboration et al., 2020a). However, for completeness, in this section we will retain K in the formalism where relevant.

Using equations 1.4 and 1.5 we can derive the components of the FLRW metric $g_{\mu\nu}$. Now, we can calculate the left hand side of Eq. 1.1. Consistent with the assumptions of homogeneity and isotropy that we applied for the metric, the Universe's matter and energy components need to also comply with these assumptions. This means that on large scales we can approximate all matter and energy contents as perfect fluids. For perfect fluids in a comoving frame the energy-momentum tensor has the form:

$$T^{\mu}_{\ \nu} = \mathrm{diag}(-\rho, p, p, p), \tag{1.7}$$

where ρ is the energy density and p the pressure. Each different matter and energy component has different energy densities and pressures. Nevertheless, the total properties are additive, which means that the total energy density and pressure are sums over the respective properties of the various fluid components in the Universe.

We use now this energy-momentum tensor combined with the FLRW metric in the Einstein's equations (1.1). Then, we arrive at the two Friedmann equations:

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{K}{a^2}, \tag{1.8}$$

$$\frac{\ddot{a}}{a} = \dot{H} + H^2 = -\frac{4\pi G}{3}(\rho + 3p), \tag{1.9}$$

where $H = \left(\frac{\dot{a}}{a}\right)$ is the Hubble parameter. Note that we included the spatial curvature K in Eq. 1.8 for completeness. We calculate the time derivative of the first Friedmann equation (1.8) and combining it with the second Friedmann equation (1.9) we get to the continuity equation, describing the local conservation of energy and momentum:

$$\dot{\rho} + 3H(\rho + p) = 0. \tag{1.10}$$

For different types of matter and energy, pressure and density are typically related via an equation of state that can expressed with the following generalized form:

$$p = w\rho, \tag{1.11}$$

If we combine this expression with the continuity equation (1.10) we arrive at an explicit scaling of density with the scale factor:

$$\rho \propto a^{-3(1+w)} \tag{1.12}$$

The parameter w characterizes the fluid, so that radiation and relativistic particles have $w_r = \frac{1}{3}$, while non-relativistic matter is treated as dust and has $w_m = 0$. The cosmological constant has

$w_\Lambda = -1$. From this, the evolution of the energy densities shows the following respective dependencies:

$$\rho_r \propto a^{-4}, \quad \rho_m \propto a^{-3}, \quad \rho_\Lambda = \text{const.} \tag{1.13}$$

These different dependencies on the scale factor determine the epochs at which different kinds of matter and energy content dominated. For instance, radiation and relativistic particles, which scale as $\rho \propto a^{-4}$, were dominant in the early Universe, when the scale factor a was still small. At intermediate times, however, matter began to take over as the dominant component. Since the cosmological constant Λ remains constant and does not depend on a, it becomes increasingly important at later times, as the densities of other components decrease with the expansion of the Universe.

For a flat Universe ($K = 0$), the first Friedman equation (1.8) at present time t_0 defines the *critical density* $\rho_{\text{crit},0}$ as:

$$\rho_{\text{crit},0} = \frac{3H_0^2}{8\pi G}, \tag{1.14}$$

Physically, the critical density represents the precise value of total energy density required to make the geometry of the Universe flat. If the actual total density exceeds this value, the Universe would have positive curvature and be closed; if it falls below, the Universe would be open with negative curvature. Thus, the critical density serves as a useful measure to compare the contribution of different components of the energy budget in the Universe, by allowing us to express the density of each component in terms of the dimensionless parameter Ω_i:

$$\Omega_{i,0} = \frac{\rho_{i,0}}{\rho_{\text{crit},0}} = \frac{8\pi G\, \rho_{i,0}}{3H_0^2}. \tag{1.15}$$

For brevity, the subscript '0' will be dropped from here onward, with Ω_i always referring to present-day values unless stated otherwise. The first Friedmann equation (1.8) can be expressed in terms of these dimensionless parameters for the previously introduced matter and energy components of the Universe:

$$\frac{H^2(a)}{H_0^2} = \Omega_{\text{r}} a^{-4} + \Omega_{\text{m}} a^{-3} + \Omega_\Lambda. \tag{1.16}$$

Here, the cosmological constant and curvature parameters are:

$$\Omega_\Lambda = \frac{\Lambda}{3H_0^2}, \quad \Omega_K = -\frac{K}{H_0^2}. \tag{1.17}$$

We now need to introduce the cosmological *redshift*, z. Let us consider a distant source that emitted light at time t_{em} with wavelength λ_{em}. If the light was observed at time t_0 with λ_0, then the redshift is defined as:

$$z = \frac{\lambda_0 - \lambda_{\text{em}}}{\lambda_{\text{em}}}. \tag{1.18}$$

As the light traveled from the source to the observer, the Universe expanded. This cosmic expansion stretches wavelengths proportionally to the scale factor, $\lambda \propto a$. The Eq. 1.18 can be expressed in terms of the scale factor for observations at present times as:

$$1 + z = \frac{1}{a(t_{\rm em})}. \tag{1.19}$$

where we considered $a_0 = 1$. The redshift is a fundamental quantity in astrophysics and cosmology and will be used across this work as a measure of distance as well as time.

Finally, combining these results, we can express the Hubble parameter for a flat universe ($\Omega_K = 0$) as a function of redshift:

$$H(z) = H_0\sqrt{\Omega_{\rm r}(1+z)^4 + \Omega_{\rm m}(1+z)^3 + \Omega_\Lambda}. \tag{1.20}$$

For a flat Universe, Ω_Λ can be simplified as:

$$\Omega_\Lambda = 1 - \Omega_{\rm m} - \Omega_{\rm r}. \tag{1.21}$$

As we mentioned previously, at low redshifts (late time Universe) the contribution of radiation to the total energy budget becomes negligible ($\Omega_{\rm r} \approx 0$), leading to a simplified form of the Hubble parameter:

$$H(z) = H_0\sqrt{\Omega_{\rm m}(1+z)^3 + 1 - \Omega_{\rm m}}, \tag{1.22}$$

which expresses the expansion of the Universe across time in terms of the current expansion rate H_0, the redshift z and the matter density content $\Omega_{\rm m}$. As we will see in future sections, the determination of the values of H_0 and $\Omega_{\rm m}$ has become a fundamental part of modern cosmology.

1.1.1 The Growth of Inhomogeneities and the Peculiar Velocity Field

Previously, we considered a Universe that is homogeneous and isotropic. However, small deviations from perfect uniformity were already present in the early Universe. The temperature anisotropies observed in the Cosmic Microwave Background (CMB) are of order $\Delta T/T \sim 10^{-5}$, corresponding to perturbations of similar magnitude in the baryonic matter density field. Due to gravitational instability, even the smallest density perturbations grow over time under the influence of gravity. Overdense regions exert a stronger gravitational pull on their surroundings, attracting more matter and further increasing their density. Meanwhile, the expansion of the Universe, driven by dark energy (or the cosmological constant in the ΛCDM model), suppresses this growth. The cosmic structures we observe today, from stars to galaxies and galaxy clusters, emerged from this ongoing competition between gravitational collapse and cosmic expansion.

In the early Universe, baryons were tightly coupled to photons. Frequent interactions with photons and the resulting radiation pressure prevented baryonic matter from collapsing under gravity. In contrast, CDM, which does not interact electromagnetically, was unaffected by radiation pressure and could begin clustering under gravity even before recombination. After recombination, once free electrons and protons combined to form neutral atoms, baryons were

able to fall into the gravitational potential wells already established by dark matter, thereby accelerating the process of structure formation.

To understand how small density fluctuations in the early universe evolved into the non-linear structures we observe today, we start by analyzing their growth using linear perturbation theory. This is a valid approximation, due to the fact that the density contrasts in the early universe were small. However, as structures grow and interactions become more complex, non-linear effects become significant, and thus more advanced methods are necessary in order to follow the evolution of structures. On scales smaller than the Hubble radius H^{-1}, and for non-relativistic matter, Newtonian gravity is enough for tracking the evolution of perturbations, as long as also cosmic expansion and the evolving background density are accounted for.

For simplicity, we treat baryons and CDM together as a single non-relativistic fluid, influenced only by gravity. This fluid is described by its density $\rho_{\rm m}(\vec{x}, t)$, negligible pressure $p_{\rm m} \ll \rho_{\rm m}$, and velocity field $\vec{v}$, with $|\vec{v}| \ll c$. For this discussion, we will refer to density and pressure simply as ρ and p, focusing solely on the matter component. Doing so, we can understand the growth of structure mathematically, starting with the definition of overdensity at some point in space $\vec{x}$ at time t:

$$\delta(\vec{x}, t) \equiv \frac{\rho(\vec{x}, t) - \bar{\rho}(t)}{\bar{\rho}(t)} \tag{1.23}$$

Initially, we will ignore the Universe's expansion to focus purely on gravitational dynamics. As has been already stablised in the introduction to this section, we will describe the density field as a perfect fluid, fully describable by its energy density ρ and pressure p. Thus, we can understand its evolution using the three main fluid equations.

The Continuity equation:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) = 0, \tag{1.24}$$

where $\vec{v}$ is velocity. This equation expresses the conservation of mass, stating that changes in mass density within a fixed volume arise solely from matter flowing into or out of that region. A positive (negative) divergence corresponds to mass flowing into (out of) the region, leading to an increase (decrease) in density. This form of the equation applies to non-relativistic matter, which is a valid approximation on the scales we are considering. However, in the relativistic case, one would need to replace ρ with $\rho + P/c^2$ in the second term to account for the contribution of pressure to the energy density.

The Euler equation:

$$\rho \frac{d\vec{v}}{dt} = -\nabla p - \rho \nabla \Phi, \tag{1.25}$$

with Φ being the gravitational potential. The Euler equation is derived from the assumption of momentum conservation and states that gravitational forces due to pressure are what induce velocity flows in the universe. It can be thought of as the fluid-dynamics equivalent of Newton's second law from classical mechanics. At the same time the gravitational potential Φ satisfies Poisson's equation:

$$\nabla^2 \Phi = 4\pi G \rho, \tag{1.26}$$

where G is the gravitational constant. This equation relates the gravitational potential to the matter and energy content of the Universe. Regions with higher matter density generate stronger gravitational fields, while regions of lower density produce weaker ones.

Together, these equations fully describe how the fluid's density, pressure, velocity, and gravitational potential evolve over time. From this equations it is possible to arrive at the general equation for the evolution of density fluctuations, which has the following form:

$$\frac{\partial^2\delta}{\partial t^2} + 2\frac{\dot{a}}{a}\frac{\partial\delta}{\partial t} = \frac{\nabla^2 p}{\rho a^2} + \frac{1}{a^2}\nabla\cdot(1+\delta)\nabla\phi \tag{1.27}$$

As perturbations are considered small, i.e., $\delta \ll 1$, we can linearize the fluid equations by neglecting higher-order terms. We also assume that the velocity field is irrotational (i.e., has no vorticity). This assumption stems from the expectation that inflation primarily generates scalar perturbations, resulting in a purely irrotational flow without vector or tensor modes (Mukhanov, 2005). Moreover, in an expanding Universe, any initial vorticity decays over time unless actively sourced. Under these two assumptions, we can write the linearized equations governing the evolution of density perturbations as follows:

$$\frac{\partial^2\delta}{\partial t^2} = 2\frac{\dot{a}}{a}\frac{\partial\delta}{\partial t} - \frac{\nabla^2 p}{\rho a^2} - 4\pi G\rho\delta, \tag{1.28}$$

$$\frac{\partial\delta}{\partial t} = -\frac{1}{a}\nabla\cdot\vec{v}. \tag{1.29}$$

Equation 1.28 is a second order differential equation. The most general solution for δ has the form:

$$\delta(\vec{x},t) = \delta(\vec{x})D_1(t) + \delta(\vec{x})D_2(t). \tag{1.30}$$

Here, D_1 represents the growing mode, while D_2 corresponds to a decaying mode. The decaying solution quickly becomes negligible, so the evolution of the perturbation is dominated entirely by the growing mode D_1. By making this approximation and substituting the result into Eq. 1.29, we can relate the velocity field to the growing mode as follows:

$$\frac{\partial\delta}{\partial t} = \delta(\vec{x})\frac{dD_1}{dt} = -\frac{1}{a}\nabla\cdot\vec{v}. \tag{1.31}$$

We define now the *linear groth rate of structure* f:

$$f \equiv \frac{d\ln D_1(a)}{d\ln a} = \frac{1}{D_1(a)}\frac{d}{d\ln(a)}D_1(a). \tag{1.32}$$

With this convention we can rewrite Eq. 1.28 as

$$\nabla\cdot\vec{v} = -aHf\delta(\vec{x},t). \tag{1.33}$$

If we integrate over all positions, we arrive at an equation for the peculiar velocity:

$$v(\vec{r}) = \frac{H_0 a f}{4\pi} \int d^3\vec{r}' \frac{\delta(\vec{r}')\,(\vec{r}' - \vec{r})}{|\vec{r}' - \vec{r}|^3} \tag{1.34}$$

Equations 1.33 and 1.34 explicitly relate the peculiar velocity field $\vec{v}(\vec{r})$ at a given position $\vec{r}$ to the gravitational influence of the surrounding density field $\delta(\vec{r}')$ at position $\vec{r}'$. The contribution of each point $\vec{r}'$ to $\vec{v}(\vec{r})$ is weighted by the term $(\vec{r}' - \vec{r})/|\vec{r}' - \vec{r}|^3$, meaning that regions closer to $\vec{r}$ exert a stronger influence on the velocity. Thus, we can also state that the peculiar velocity field $\vec{v}(\vec{r})$ is a tracer of the underlying gravitational potential field. This is a pivotal point as it lies in the foundation of the constrained simulations used in this work, which will be presented in detail in the next chapters.

We can also understand peculiar velocities as the relation between physical and comoving coordinates. The relation between physical coordinates of a point in space-time $\vec{r}(t)$ to the comoving coordinates $\vec{x}(t)$ is:

$$\vec{r} = a(t)\vec{(x)}. \tag{1.35}$$

The proper velocity is:

$$\vec{u} = \frac{d\vec{r}}{dt} = \dot{a}\vec{x} + a\vec{v}(\vec{x}, t) = \left(\frac{\dot{a}}{a}\right)\vec{r} + a\vec{v}(\vec{r}/a, t). \tag{1.36}$$

Here, $a\vec{v}$ represents the peculiar velocity, which corresponds to the time derivative of the comoving coordinate $\vec{x}$. The term $(\dot{a}/a)\vec{r} = H\vec{r}$ arises purely from the expansion of the Universe and describes the smooth Hubble flow. The second term accounts for deviations from this smooth expansion. These are the peculiar velocities, driven by the gravitational influence of local matter overdensities. In practice, most galaxies exhibit non-negligible peculiar velocities, causing them to deviate from the Hubble flow.

1.1.2 Cosmological Parameters

As explained in the previous sections, in order to describe the evolution of the universe in the ΛCDM model, we rely on a set of fundamental parameters: the density of radiation Ω_r, the density of matter Ω_m, the curvature parameter Ω_k, the contribution of the cosmological constant Ω_Λ, and the present-day value of the Hubble parameter, H_0. In addition to these, the amplitude of initial matter fluctuations is encapsulated in σ_8, which quantifies the strength of density variations on scales of 8 Mpc h^{-1}.

We have also seen that with this cosmological model, the non-relativistic matter component includes both ordinary baryons and CDM. This dual nature of the non-relativistic matter is supported by evidence coming from the early Universe (like the CMB), as well as from the later Universe, such as observations of the local Universe. Studies supporting the existence of dark matter were presented already in the 1930s, where (Zwicky, 1933) showed that visible (baryonic) mass alone could not account for the velocities of galaxies within clusters. Similarly, galactic rotation curves demonstrate the presence of unseen mass, as the baryonic contribution alone cannot explain the observed rotational speeds (Rubin et al., 1980; Bosma, 1981). Despite the

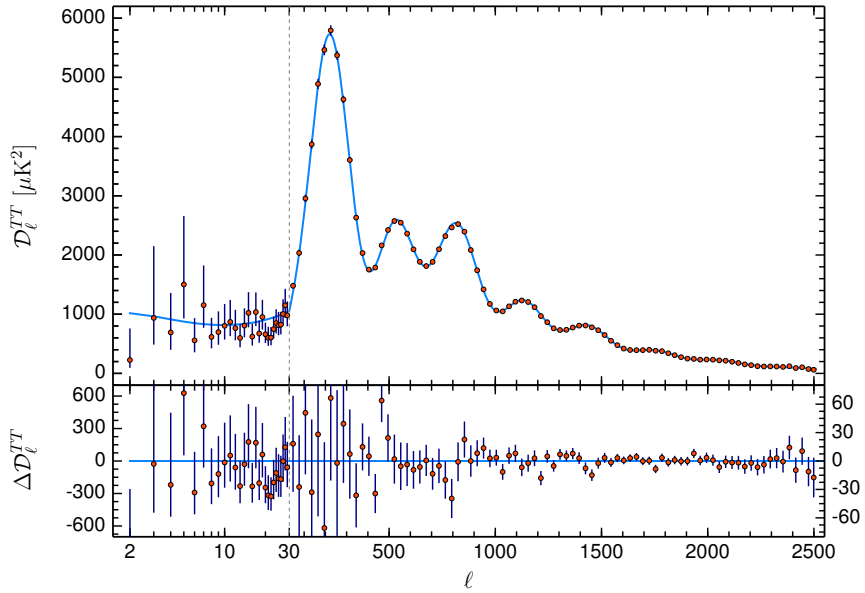


Figure 1.1: CMB temperature power spectrum as measured by the Planck Collaboration et al. (2020b). The best-fit model yields parameters $\Omega_\Lambda = 0.6889 \pm 0.0056$, $\Omega_\mathrm{m} = 0.3111 \pm 0.0056$, $\Omega_\mathrm{b}h^2 = 0.02242 \pm 0.00014$, $\Omega_\mathrm{dm}h^2 = 0.11933 \pm 0.00091$, and $\sigma_8 = 0.8102 \pm 0.0060$.

overwhelming evidence for its existence, the fundamental nature of dark matter remains elusive (for a detailed review, see Arbey & Mahmoudi, 2021). The current prevailing view, which seems to explain structure formation best, is that dark matter is "cold", i.e. moving at non-relativistic speeds, and interacts purely through gravity. Nevertheless, alternatives remain plausible and are currently being explored. Some proposed alternatives are fuzzy dark matter, consisting of ultra-light scalar fields, warm dark matter with higher velocities or even models where dark matter exhibits some degree of self-interaction, making it partially collisional. Candidates for dark matter include primordial Black Holes (BHs) and hypothetical weakly interacting massive particles (WIMPs), both of which lie beyond the Standard Model of particle physics. Modifications to general relativity itself have also been proposed but often fall short when confronted with specific observational tests (Bertone & Tait, 2018; Feng, 2010; Hu et al., 2000; Tulin & Yu, 2018; Carr & Kuhnel, 2021; Clifton et al., 2012). To fully capture the matter composition of the Universe, it is therefore essential to consider the baryonic density Ω_b and the contribution from cold dark matter Ω_DM separately.

Today, multiple techniques are employed to constrain the numerical values of these six cosmological parameters. The most comprehensive constraints come from precise measurements of the CMB anisotropies. These observations have progressively improved over time, beginning with the COBE satellite (Boggess et al., 1992), followed by WMAP (Komatsu et al., 2011), and most recently the Planck mission (Planck Collaboration et al., 2014b, 2020b). By fitting theoretical models to the temperature power spectrum of the CMB, we can derive values for all the main cosmological parameters.

The latest results from Planck Collaboration et al. (2020b) are visualized in Fig. 1.1. These

constraints are very precise and are widely adopted, for example, as inputs for cosmological simulations. These studies report a present-day Hubble parameter of $H_0 = (67.66 \pm 0.42)\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, alongside density parameters $\Omega_\Lambda = 0.6889 \pm 0.0056$, $\Omega_\mathrm{m} = 0.3111 \pm 0.0056$, $\Omega_\mathrm{b}h^2 = 0.02242 \pm 0.00014$, $\Omega_\mathrm{dm}h^2 = 0.11933 \pm 0.00091$, and $\sigma_8 = 0.8102 \pm 0.0060$. Furthermore, the curvature parameter is found to be consistent with a spatially flat Universe ($\Omega_k \approx 0$).

A different approach to measuring cosmological parameters consists of studying the Large Scale Structure (LSS) of the Universe at later times, using galaxies as tracers of the matter distribution. This requires extensive sky surveys that catalogue vast numbers of galaxies. Several such surveys are currently active. Recent constraints have been provided by the Dark Energy Spectroscopic Instrument (DESI Collaboration (2024)), the extended Baryon Oscillation Spectroscopic Survey (eBOSS) as part of SDSS (Alam et al., 2021), and the Dark Energy Survey (DES Collaboration, 2022).

In addition to galaxies, a key cosmological probe of LSS is the abundance and distribution of galaxy clusters across cosmic time. Galaxy clusters, as the most massive gravitationally bound and virialized structures in the Universe, trace the high-density peaks of the matter field and are highly sensitive to both the expansion history and the growth of cosmic structures, rather than baryonic micro-physics. Cluster counts as a function of redshift and mass provide stringent constraints on cosmological parameters, particularly the matter density Ω_m and the amplitude of matter fluctuations σ_8 (Vikhlinin et al. (2009); Mantz et al. (2015); Bocquet et al. (2019)). Observations from X-ray surveys (*Chandra*, *XMM-Newton*; Allen et al. (2008); Pratt et al. (2009)), Sunyaev-Zel'dovich Effect (SZ) surveys (*Planck*: Planck Collaboration et al. (2016a); SPT: Bleem et al. (2015); ACT: Hilton et al. (2021)), and optical surveys (e.g. DES: Abbott et al. (2022); HSC: Oguri et al. (2018)) have contributed to increasingly precise measurements of the cluster mass function and its evolution.

Alongside galaxy and cluster counts, other LSS observables such as Baryon Acoustic Oscillations (BAO) and redshift-space distortions (RSD) further enhance our understanding of the cosmic expansion and structure formation (Alam et al. (2017); Bautista et al. (2021)). Furthermore, weak gravitational lensing of galaxies, known as cosmic shear, provides a direct probe of the total matter distribution, independent of galaxy bias, offering valuable cross-checks on cosmological models (Hildebrandt et al. (2021); Amon et al. (2022)). Even the largest underdense regions in the Universe, the cosmic voids, are being explored as complementary probes sensitive to the growth of structure and properties of dark energy (Sutter et al. (2014); Hamaus et al. (2016)).

The combination of diverse LSS (late time) observables and CMB (early time) measurements aims to improve constraints on the key cosmological parameters (Kravtsov & Borgani (2012b); Allen et al. (2011)). However, important discrepancies remain. Tensions have emerged concerning the value of σ_8, and the different methods mentioned above yield also varying estimates of the Hubble constant H_0. We will explore these inconsistencies in the next section. That said, upcoming surveys such as *Euclid* (Laureijs et al. (2011)), the *Vera C. Rubin Observatory's Legacy Survey of Space and Time* (LSST; Collaboration (2009)), and the *Nancy Grace Roman Space Telescope* (Spergel et al. (2015)) are planned to expand the available data, allowing for new high-precision tests of the ΛCDM model.

1.1.3 Tensions in Cosmology

Despite the success of the ΛCDM model in describing a wide range of cosmological observations, discrepancies persist between early- and late-time probes in the inferred values of some key cosmological parameters. These tensions may point to new physics beyond the standard model or to unresolved systematics in measurements and analyses, and thus, they remain an important focus of current research. Among these tensions, the most important discrepancies are the ones surrounding the value of the H_0 and the amplitude of matter clustering quantified by σ_8 (or its derived parameter S_8). These tensions challenge the completeness of the standard ΛCDM model and have sparked intense theoretical and observational efforts, remaining an important focus of current research.

The H_0 Tension

The Hubble constant, H_0, quantifies the current expansion rate of the Universe. The measurement of its value has historically relied on two main approaches: early-Universe measurements, predominantly from the CMB observations assuming ΛCDM cosmology, and late-Universe measurements using cosmic distance ladder methods. The *Planck* satellite's analysis of the CMB anisotropies infers a value of $H_0 = (67.27 \pm 0.60)$ km/s/Mpc, under the assumption of a spatially flat ΛCDM cosmology (Planck Collaboration et al., 2020b). In contrast, direct local measurements, such as those from the SH0ES collaboration using Cepheid-calibrated Supernovae (SNe), report a significantly higher value of $H_0 = (73.04 \pm 1.04)$ km/s/Mpc (Riess et al., 2022).

This discrepancy of about 5σ represents one of the most statistically significant and persistent anomalies in modern cosmology. The tension is not confined to these two measurements alone; while early-Universe probes such as the CMB and BAO are mutually consistent, late-time, model-independent probes like distance ladder methods and strong lensing time delays also converge towards the higher H_0 value. Figure 1.2 shows an overview of the value of H_0 inferred by different methods. This convergence among independent late-Universe measurements strengthens the fact that the tension cannot be easily dismissed as a result of unknown systematics within a single method (Di Valentino et al., 2021).

Efforts to resolve the Hubble tension span a wide spectrum, including revisions to the cosmic distance ladder, improvements in the control of systematics, and the exploration of new physics beyond ΛCDM. Models involving early dark energy, modifications of gravity, or new light relics in the early Universe have been investigated as potential solutions (Knox & Millea, 2020). Some approaches propose a modification to the expansion history at early times to reconcile the observed discrepancies without disrupting the success of ΛCDM at other epochs.

The σ_8 and S_8 Tension

The parameter σ_8 characterizes the amplitude of matter density fluctuations on scales of 8 h^{-1} Mpc. A related parameter, $S_8 \equiv \sigma_8(\Omega_\mathrm{m}/0.3)^{1/2}$, is often used to combine information on σ_8 and the matter density parameter Ω_m in weak lensing analyses.

There is a persistent tension between the amplitude of matter clustering inferred from *Planck* CMB data and that measured by LSS probes at lower redshifts, such as weak gravitational lensing

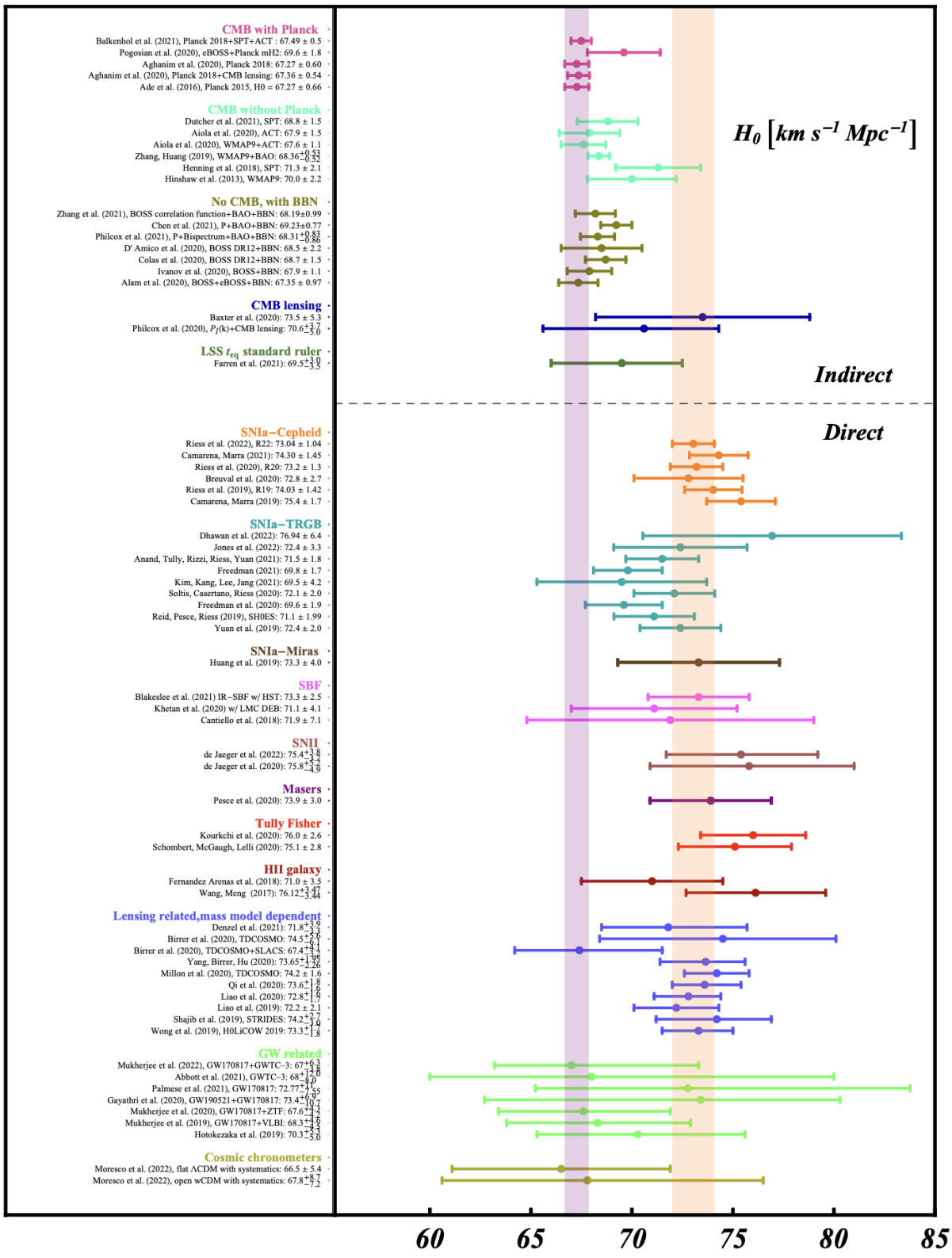


Figure 1.2: Current constraints for H_0 presented as a whisker plot showing 68% confidence levels. Credits: (Di Valentino et al., 2021)

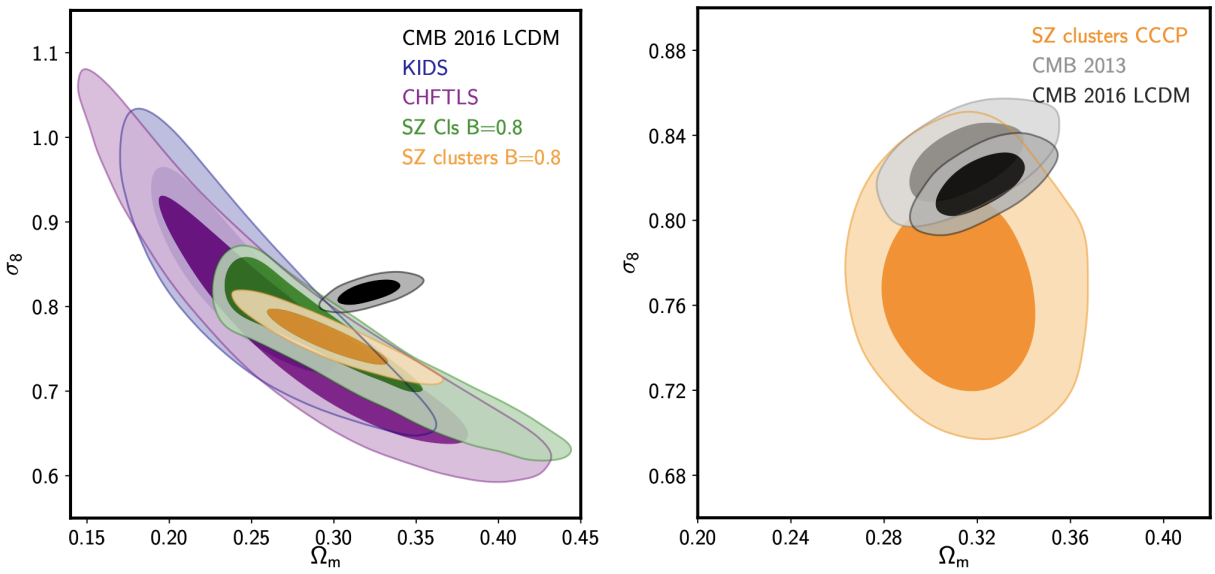


Figure 1.3: Constraints for σ_8 as presented by (Douspis et al., 2019). Left: recent large-scale constraints compared with CMB (black filled contours). Right: Comparison of SZ cluster constraints from Planck (yellow) with CCCP prior on the mass bias and Planck CMB constraints from 2013 and 2016 (grey and black, respectively). The displacement of the CMB constraints is due to the new estimate of the reionisation optical depth. Credits: (Douspis et al., 2019)

surveys (e.g. KiDS, DES, HSC, see Fig. 1.3). The CMB-inferred value tends to be higher than that from low-redshift observations, with the discrepancy reaching ~2–3σ significance Heymans & et al. (2021); DES Collaboration et al. (2022).

A critical challenge arises from the correlations between H_0 and S_8 parameters: models that aim to resolve the Hubble tension frequently exacerbate the σ_8 tension, and vice versa. For instance, introducing late-time dark energy dynamics to increase H_0 often results in a lower Ω_m to preserve the CMB-inferred value of $\Omega_m h^2$, which in turn modifies the growth history of cosmic structures and typically raises σ_8 beyond observational constraints from lensing surveys (Di Valentino et al., 2021). Similarly, early-time modifications that enhance the primordial curvature perturbations to address the H_0 tension generally lead to an increase in σ_8 (Abdalla et al., 2022).

Therefore, a viable resolution to these tensions requires models capable of simultaneously addressing both H_0 and S_8 discrepancies. Some models that are being proposed involve modifications to gravity, interactions within the dark sector, or the introduction of time-varying fundamental constants that influence both the expansion and growth histories of the Universe (Di Valentino et al., 2021; Knox & Millea, 2020).

The persistence of the H_0 and σ_8 tensions stress a potential limitations of the ΛCDM model. While the effect of systematics and of statistical fluctuations remain plausible explanations, the coherence of independent measurements points towards the possibility of new physics. Future

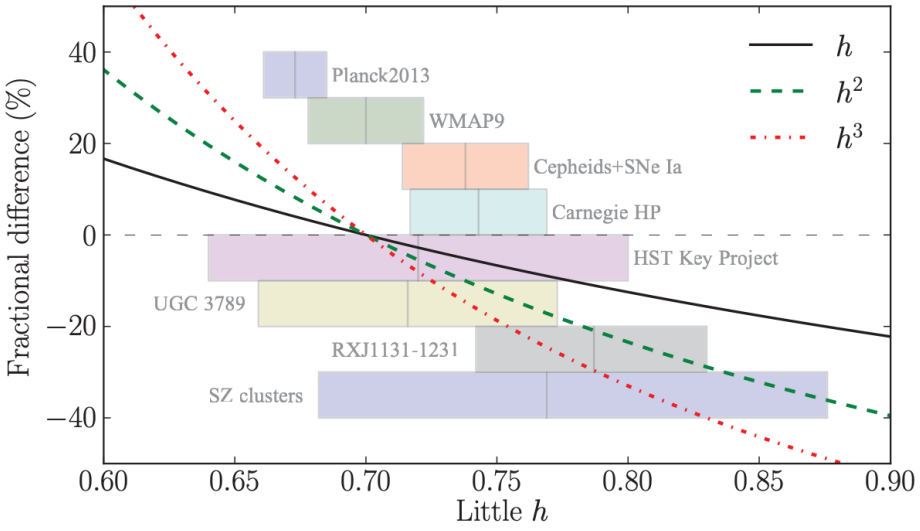


Figure 1.4: This figure illustrates how the value of a derived property can change between cosmologies with different assumptions for the Hubble constant h. It shows the fractional difference in a quantity (in percent) relative to a baseline value of $h = 0.7$, as h varies continuously from 0.60 to 0.90—a range that encompasses current estimates. The solid, dashed, and dotted lines represent how this difference evolves for properties that scale as h, h^2, and h^3, respectively. Shaded bands correspond to various measurements of h from the literature, each shown with $\pm 1\sigma$ uncertainty and vertically offset for clarity. Credits: Croton (2013).

observations from upcoming surveys and missions, such as the Vera C. Rubin Observatory, *Euclid*, and the *Nancy Grace Roman Space Telescope*, promise to enhance our ability to probe the expansion history and the growth of structure with unprecedented precision (NASA, 2025), and thus be able to better understand whether these tensions are a signal of new fundamental physics or simply reflect limitations in our current observational and theoretical frameworks.

1.1.4 Cosmological Distances

The difficulty in converging on unique values for cosmological parameters, as has been discussed in Sect. 1.1.3, affects directly the interpretation of observations of galaxies and galaxy clusters. Distant objects can only be studied through the light they emit, which travels to us over cosmological timescales. This means that measurements are constrained to our past *light cone*, rather than surfaces of constant proper time. As light propagates through an expanding universe, its properties are altered by the underlying cosmology. Therefore, the assumed cosmological parameters, particularly the Hubble constant, directly influence how we infer distances and positions of astronomical objects. This becomes especially important when comparing different observations or simulations that adopt different baseline cosmologies.

One such distance is the *luminosity distance*, d_L, which is defined to preserve the Euclidean inverse-square law for the diminution of light with distance from a point source:

$$d_L = \left(\frac{L}{4\pi l}\right)^{1/2} = \frac{a_0^2 r}{a}, \tag{1.37}$$

where L is the intrinsic luminosity of the source, l the observed flux, a the scale factor, and r the comoving coordinate distance. At low redshift, the expression for d_L reduced to:

$$d_L = \frac{c}{H_0}\left[z + \frac{1}{2}(1 - q_0)z^2 + \cdots\right]. \tag{1.38}$$

making explicit that $d_L \propto h^{-1}$.

Another important measure of distance is the *angular diameter distance*, d_A. This is constructed to preserve a geometric property of Euclidean space, namely, the variation of an object's apparent angular size with its distance from an observer. Let $D_p(t)$ denote the proper diameter of a source located at coordinate r and time t. If this diameter subtends an angle $\Delta\vartheta$, then:

$$D_p = ar\Delta\vartheta. \tag{1.39}$$

The angular diameter distance is then defined as

$$d_A = \frac{D_P}{\Delta\vartheta} = ar \tag{1.40}$$

where $D_P \propto h^{-1}$ and thus $d_A \propto h^{-1}$.

Other derived quantities, such as volumes and luminosities, inherit even stronger dependencies, scaling as h^{-3} and h^{-2}, respectively. The impact of adopting different values of h is illustrated in Fig. 1.4, which shows how properties with different h-dependencies shift when h varies between 0.60 and 0.90. For example, a property scaling as h^{-3} can differ by more than 40% across this range. These dependencies emphasize a key conceptual point: observations are inherently linked to cosmology. In practical terms, this means that most interpretations of observational data, especially at cosmological distances, necessarily rely on an assumed cosmological model. Therefore, cosmology leaves a measurable imprint on observational quantities, typically expressed with the quantity h, which must be accounted for when comparing different datasets.

1.2 Galaxy Clusters in the Cosmological Environment

As has been previously explored, in the standard model of cosmology the formation of cosmic structures is explained by the progressive amplification of small initial fluctuations in the density field, modeled as a Gaussian random field. These perturbations grow over time due to gravitational instability. In this scenario, some of the most massive structures, such as galaxy clusters, are considered the latest systems to assemble under the influence of gravity in an expanding Universe (Press & Schechter, 1974b; Peebles, 1980; Lacey & Cole, 1993; Springel et al., 2005c; Mo et al., 2010; Kravtsov & Borgani, 2012a).

The exact origin of the primordial fluctuations from which massive structures such as galaxy clusters evolve is still debated. The leading theory suggests they emerged as a consequence of

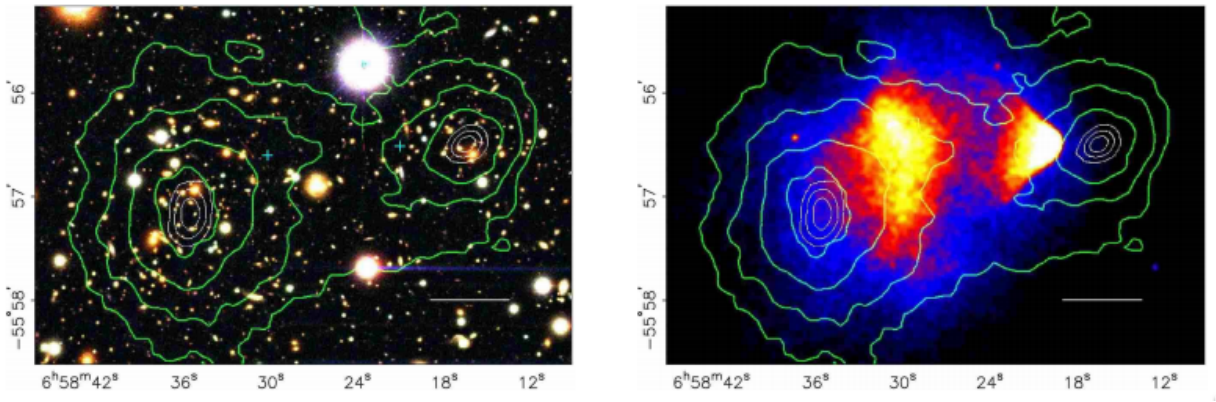


Figure 1.5: The Bullet cluster. Left panel: optical image taken by the Magellan telescope. Right panel: Chandra X-ray measurements of the hot plasma distribution (in red). The green contours denote the gravitational potential derived from weak lensing measurements. Credits: Clowe et al. (2006)

an early period of accelerated expansion, commonly referred to as inflation (Guth, 1981, 1997; Linde, 2005). During the early stages of cosmic evolution, these fluctuations remain small, characterized by a density contrast $\delta\rho/\rho$ much less than one, a regime often called linear, and thus well described by the linear theory presented in Section 1.1. As gravity continues to enhance these overdensities, their density contrast eventually approaches unity, where linear theory breaks down. At this critical or *turnaround* threshold, the overdense regions decouple from the general expansion and begin to collapse under their own gravity.

The non-linear collapse that follows is influenced by the composition of the collapsing region. The halos conforming galaxy clusters are composed of dark matter and baryonic matter (existing in various forms, including stars and gas) and Supermassive Black Holes (SMBHs). When baryonic gas collapses, shock waves are generated, resulting in a rise in entropy. If radiative cooling processes are inefficient, the system eventually settles into hydrostatic equilibrium, where the gravitational pull is counteracted by pressure gradients (see Sarazin, 1986; Voit et al., 2005, for comprehensive reviews). Conversely, if the collapsing structure is composed predominantly of collisionless dark matter (e.g., CDM), no shocks occur, but the system still undergoes a rapid phase of violent relaxation (Lynden-Bell, 1967), leading to a quasi-equilibrium state with an almost universal density profile (Navarro et al., 1997). These self-bound dark matter structures are known as dark matter halos. Their internal structure and properties have been extensively investigated using cosmological simulations (see Chapter 2 of this work), and they play a central role in our current understanding of galaxy formation (e.g., Springel et al., 2005c; Kravtsov & Borgani, 2012b; Planelles et al., 2013).

Decades of numerical studies with hydrodynamical simulations (see Kravtsov & Borgani, 2012b, for a comprehensive review), have significantly advanced our ability to model and understand the complex baryonic physics within clusters. These models increasingly succeed in reproducing the wealth of observational data collected over the years. Based on theoretical and simulation results, the main stages in the assembly of galaxy clusters can be outlined as follows:

As gravitational collapse proceeds, gas falling into the potential well of the cluster experiences shock heating and adiabatic compression, raising its temperature. The cooling of gas within clusters is regulated by several feedback mechanisms, including mechanical and radiative energy release from SN winds and Active Galactic Nucleus (AGN) jets, thermal conduction, and additional plasma processes. Encounters between galaxies, such as harassment and tidal stripping, also play a role in removing material from galaxies and redistributing it within the cluster (Merritt, 1983; Moore et al., 1996; Abadi et al., 1999; Boselli & Gavazzi, 2006; Cimatti et al., 2020, see). As a result, most galaxy clusters are characterized by the extended halo of hot gas, known as the Intracluster Medium (ICM), that roughly traces the spatial distribution of the underlying dark matter halo, enclosing the population of galaxies that orbit within the cluster's potential well. Observations of this hot, X-ray emitting plasma have provided compelling support for the existence of dark matter halos, as deeper gravitational wells are needed to explain the temperature of the ICM (Sarazin, 1986; Frenk et al., 1999; Allen et al., 2002; Voit, 2005).

One of the most clear observational confirmations of the existence of dark matter comes from the Bullet Cluster, as presented in the study by Clowe et al. (2006). This merging system exhibits a clear separation between its collisional and collisionless components: the X-ray-emitting hot gas and the dark matter. In Fig. 1.5, the optical image (left panel) shows the distribution of galaxies, while the green contours represent the bimodal structure of the gravitational potential derived from weak gravitational lensing. The right panel displays Chandra X-ray observations, showing the hot gas component, which has been slowed by hydrodynamic interactions (ram pressure and shocks) as it moves through the ambient gas of the merging system and is found near the center of the two dark matter concentrations. This configuration demonstrates that, during a merger, the collisionless dark matter passes through largely unaffected, while the gas component undergoes shock heating and slows down, eventually settling toward an equilibrium configuration. In these regions, the intracluster gas is already sufficiently dense to emit X-rays through thermal bremsstrahlung. If local conditions allow the gas to cool efficiently—especially in cluster cores—it may lose enough thermal energy to condense and form stars. The resulting stellar component, however, represents only a small fraction of the total baryon content of the cluster, accounting for roughly 12 ± 2 percent in the most massive systems (Giodini et al., 2009; Chiu et al., 2018, and references therein).

1.2.1 Galaxy Clusters as Probes of Cosmology

Galaxy clusters are highly sensitive to the general growth of cosmic structure, making them powerful cosmological probes. Thus, their abundance, spatial distribution, internal structure, as well as their baryonic content offer complementary and independent constraints on the fundamental cosmological parameters.

In Section 1.1.2, we mentioned the use of the abundance and distribution of galaxy clusters as cosmological probes. The number density of galaxy clusters as a function of mass and redshift—the cluster mass function—is a sensitive probe of the underlying cosmology. In particular, it depends on the matter density and the amplitude of primordial density fluctuations (Press & Schechter, 1974a; Vikhlinin et al., 2009b). Measurements of the evolution of the cluster mass function provide constraints on the growth rate of structure and thus on dark energy and modified

gravity scenarios (Allen et al., 2011; Kravtsov & Borgani, 2012b).

The baryonic content of clusters, particularly their gas mass fraction (f_{gas}), provides an additional cosmological probe. The ratio of baryonic to total mass in the largest, dynamically relaxed clusters is expected to be representative of the cosmic mean (White et al., 1993; Allen et al., 2008). By measuring f_{gas} as a function of redshift and assuming it is constant, one can place tight constraints on the geometry of the Universe and the nature of dark energy (Allen et al., 2004; Ettori et al., 2009; Vikhlinin et al., 2009; Mantz et al., 2010; Allen et al., 2011).

Clusters are also sensitive to modifications of gravity and the nature of dark energy through their growth history and internal structure. Tests of modified gravity theories, such as $f(R)$ models, have been conducted using cluster abundance and dynamics (Schmidt et al., 2009; Ferraro et al., 2011; Cataneo et al., 2015). Additionally, the detection and characterization of cluster outskirts and splashback radii has also been use to further understand structure formation and for constraining cosmological models (Diemer & Kravtsov, 2014; More et al., 2016; Baxter et al., 2017; Pratt et al., 2019).

Thus, it is obvious that the usefulness of galaxy clusters as cosmological probes depends strongly on how precise and accurate the measurements of their relative matter contents and total masses are. In this regard, the main challenges that galaxy clusters face arise from observational biases and systematic uncertainties.

One major source of uncertainty lies in the mass estimation techniques, which often rely on assumptions that are not universally valid. X-ray methods, for example, assume hydrostatic equilibrium between the ICM and the gravitational potential. However, non-thermal pressure support from turbulence, bulk motions, or cosmic rays can lead to an underestimation of cluster masses by up to 10–30% (Nagai et al., 2007). Similarly, gravitational lensing methods, though independent of the cluster's dynamical state, are subject to projection effects, where uncorrelated structures along the line of sight can bias mass measurements (Sereno & Ettori, 2017).

Further complications arise from triaxiality and cluster orientation. Many clusters are not spherical, and their three-dimensional shape and alignment relative to the observer can bias both lensing and X-ray-derived masses (Corless & King, 2007; Buote & Humphrey, 2012). Additionally, clumpiness in the ICM can artificially boost X-ray luminosities and inferred gas densities, further affecting mass estimates and derived scaling relations (Mathiesen et al., 1999; Nagai et al., 2007; Roncarelli et al., 2013).

Selection effects represent another critical source of bias. Flux-limited surveys tend to preferentially detect the most luminous and often more relaxed clusters, introducing a Malmquist bias that skews the sample toward objects with enhanced X-ray emission or SZ signal (Vikhlinin et al., 2009a; Mantz et al., 2010; Maughan, 2014). The Eddington bias further affects the observed mass function by scattering lower-mass clusters into higher-mass bins due to measurement uncertainties, distorting the inferred cosmological parameters (Eddington, 1913; Allen et al., 2011).

Finally, mass calibration uncertainties remain the dominant source of systematics in cluster cosmology (Allen et al., 2011; Mantz et al., 2015; Planck Collaboration et al., 2016a; de Haan et al., 2016; Pratt et al., 2019). Cross-calibration between different mass estimation techniques—hydrostatic, lensing, and SZ—along with multi-wavelength observations and improved simulations, is essential to mitigate these uncertainties.

Addressing these biases and uncertainties is an active area of research. Advances in hydrodynamical simulations, accurate modeling of the ICM, sophisticated selection functions, and multi-wavelength observations are all helping to improve the accuracy and precision of cluster-based cosmological constraints.

1.2.2 The Intracluster Medium and its Thermodynamics

The gas in galaxy clusters forming the ICM is a hot, ionized plasma that fills the deep gravitational potential wells carved out predominantly by dark matter. It constitutes the majority of the baryonic mass within clusters, typically accounting for 10–15% of the total mass, while stars contribute only a few percent. The remaining mass, around 85–90%, is in the form of dark matter (Navarro et al., 1997; Kravtsov & Borgani, 2012b). With temperatures ranging from 10^7 to 10^8 K, the ICM emits abundantly in X-rays via thermal bremsstrahlung and line emission from heavy elements, making it a rich source of information about the thermodynamic state and evolutionary history of galaxy clusters (Sarazin, 1986; Böhringer & Werner, 2010).

The ICM exists in a state of quasi-equilibrium, established by the balance between gravitational heating from accretion and mergers and various cooling and feedback processes. This state is commonly described by the virial theorem, which relates the kinetic and potential energies of the system as $2E_{\rm kin} + E_{\rm pot} - E_{\rm s} = 0$. The collisional nature of the hot plasma allows the ICM to trace the underlying gravitational potential, which itself follows a Navarro-Frenk-White (NFW) density profile (Navarro et al., 1997):

$$\rho(r) = \rho_{\rm c} \frac{\delta_c}{(r/r_s)(1 + r/r_s)^2}, \tag{1.41}$$

where $\rho_{\rm c}$ is the critical density of the Universe, r_s is the scale radius, and δ_c is related to the halo concentration. While dark matter dominates the potential, baryonic processes, such as gas cooling and AGN feedback, can modify the central regions, leading to deviations from the pure NFW form (van Daalen et al., 2011; Fedeli, 2012).

Observational studies of the ICM use multiple wavelengths to probe its various properties. X-ray observations, primarily with Chandra and XMM-Newton, provide measurements of gas density and temperature profiles by mapping surface brightness and fitting spectra. Assuming hydrostatic equilibrium, these profiles can be used to infer total cluster masses. However, hydrostatic mass estimates are affected by non-thermal pressure support from turbulence and bulk motions, introducing the hydrostatic mass bias (Nagai et al., 2007; Lau et al., 2009; Nelson et al., 2014; Biffi et al., 2016; Eckert et al., 2019; Vazza et al., 2009).

The thermal Sunyaev-Zel'dovich Effect (tSZ) provides a complementary method for probing the ICM. As CMB photons pass through the hot plasma, they are inverse Compton scattered by high-energy electrons, causing a frequency-dependent distortion of the CMB spectrum (Zeldovich & Sunyaev, 1969; Sunyaev & Zeldovich, 1972). The magnitude of this effect is quantified by the Compton-y parameter:

$$y = \frac{\sigma_T}{m_e c^2} \int P_e dl, \tag{1.42}$$

where P_e is the electron pressure integrated along the line of sight. Since the tSZ effect is independent of redshift, it offers a powerful, unbiased method for detecting galaxy clusters and measuring their integrated pressure profiles (Planck Collaboration et al., 2016a). The kinetic SZ effect, arising from bulk motions of the ICM relative to the CMB rest frame, provides additional information about internal gas velocities, though its observational signature is weaker and more challenging to isolate (Sunyaev & Zeldovich, 1980; Birkinshaw, 1999; Hand et al., 2012; Sayers et al., 2013; Battaglia et al., 2017).

Thermodynamic profiles of the ICM, specifically temperature, density, pressure, and entropy, are key to understanding both the equilibrium structure of the gas and its complex thermal history (see Voit, 2005; Nagai et al., 2007). Gas density profiles typically show central values ranging from 10^{-3} - 10^{-1} cm^{-3}, followed by a steep decline with radius, especially beyond R_{500} (e.g. Vikhlinin et al., 2006; Ettori et al., 2010). Temperature profiles exhibit a peak at intermediate radii, decreasing toward both the core and the outskirts and with values around 10^7 - $10^8 K$, which is close to the virial temperature of the halo (e.g. Vikhlinin et al., 2006; Cavagnolo et al., 2008). Entropy profiles, often parameterized as $K = k_B T n_e^{-2/3}$, rise with radius but flatten in the centers of clusters, reflecting the effects of non-gravitational heating and cooling processes (e.g. Pratt et al., 2009; Cavagnolo et al., 2009). Pressure profiles, when normalized by the characteristic pressure and radius of the halo, typically using self-similar scaling with P_{500} and R_{500}, display a universal shape and are a fundamental tool for calibrating cluster masses in cosmological studies (Arnaud et al., 2010; Planck Collaboration et al., 2013a; Ghirardini et al., 2019).

The chemical composition of the ICM, especially its metallicity, provides insight into the enrichment history of galaxy clusters. Heavy elements, primarily iron, are synthesized in stars and dispersed into the ICM by SNe and AGN-driven outflows. X-ray spectroscopy reveals that metallicity is generally higher in the inner regions of clusters—often in association with the Brightest Cluster Galaxy (BCG)—and declines with radius. However, both observations and simulations indicate that this decline flattens at large radii, where metallicity tends to approach a nearly universal floor of $\sim 0.1 - 0.3 Z_\odot$, even out to R_{200} and at higher redshifts (Mernier et al., 2018; Biffi et al., 2018). The origin of this widespread enrichment is still debated: it may result from the stripping of enriched gas from infalling satellites or from AGN-driven uplift of BCG gas. Current X-ray measurements are limited in their ability to disentangle metallicity and temperature due to spectral resolution constraints, a challenge expected to be addressed by upcoming calorimeter, based missions such as XRISM/Resolve. Overall, the observed radial metallicity behavior reflects both the star formation and feedback history, as well as the efficiency of mixing and transport processes such as turbulence (Biffi et al., 2017).

Although turbulence contributes less to the overall pressure support compared to thermal pressure, it plays an important role in regulating small-scale processes within the ICM. Turbulent motions, primarily driven by mergers, accretion flows, and AGN activity, provide non-thermal pressure support and promote mixing of gas and metals throughout the ICM (Hitomi Collaboration et al., 2016; Zhuravleva et al., 2019). Magnetic fields are also present, typically with strengths of a few microgauss (Carilli & Taylor, 2002; Bonafede et al., 2011), but their dynamical impact on large scales is generally subdominant.

1.2.3 Galaxy Cluster Cores

The cores of galaxy clusters represent regions of dense, hot gas confined within the deepest parts of the gravitational potential well. These environments are critical for understanding the thermodynamic evolution of the ICM, as they host a range of physical processes including radiative cooling, feedback from AGN, and gas dynamical interactions driven by mergers.

Early X-ray observations of galaxy clusters showed that the ICM at the cores of many clusters had such high densities that the gas was expected to cool faster than the Hubble time (e.g. Lea et al., 1973; Cowie & Binney, 1977; Fabian & Nulsen, 1977; Mathews & Bregman, 1978). These findings gave rise to the consideration of the so-called *cooling flow* scenario, in which the core gas of some clusters loses energy and cools under the influence of gravity, becoming denser as it is compressed by the weight of the overlying ICM. As a result, hotter gas from the outer regions moves inward to maintain pressure balance, producing a steady inward flow of gas known as a cooling flow and giving rise to Cool-Core (CC) clusters.

While initial X-ray measurements appeared to support the cooling flow paradigm, along with some detections of H α and ultraviolet emission, optical surveys did not confirm the high star formation rates, nor the presence of CO and molecular gas, that the model predicted (McNamara & O'Connell, 1989; Edge, 2001). The validity of the classical cooling flow scenario was further challenged when mass deposition rates inferred from ASCA spectral data diverged from those estimated using classical determination of gas density (Makishima et al., 2001, and references therein). Observations from the XMM-Newton Reflection Grating Spectrometer (RGS) revealed a shortfall in emission at lower temperatures compared to what the cooling flow model forecasts (Peterson et al., 2001; Tamura et al., 2001a,b; Kaastra et al., 2001; Xu et al., 2002; Sakelliou et al., 2002; Peterson et al., 2003; Sanders et al., 2010). Likewise, although star formation is observed in BCGs, the rates are far below those expected if the classical cooling flow model were accurate (e.g., Fraser-McKelvie et al., 2014; McDonald et al., 2018).

The mismatch between theoretical expectations and observational evidence has come to be known as the *cooling flow problem*. The traditional cooling flow framework operates under the assumption that the gas cools in the absence of any substantial heating processes. However, to reconcile the observed deviations, it is now broadly recognized that some form of energy input must counteract the cooling, thereby stabilizing the thermal state of the ICM. Any proposed heating mechanism must satisfy a number of stringent conditions. First, it should be capable of offsetting cooling across the entire core, not just the innermost regions. Second, it must function effectively across a broad range of system masses, from low-mass galaxy groups around 10^{13} $M_\odot$ to massive clusters approaching 10^{15} $M_\odot$. Lastly, it must achieve a delicate balance, suppressing cooling flows without producing clear signs of excessive heating (Fabian, 2002).

One of the first proposed mechanisms to counteract cooling in cluster cores is thermal conduction, which could, in principle, transfer heat inward from hotter outer regions due to the observed negative temperature gradients (e.g. Zakamska & Narayan, 2003). Nevertheless, Dolag et al. (2004) demonstrated that even assuming a conductivity reduced to one-third of the Spitzer value—accounting for suppression by tangled magnetic fields—has minimal impact on star formation. This is because the bulk of cooling and star formation occurs at earlier cosmic times, when the ICM is cooler and conduction is inherently less effective. Still, while thermal

conduction alone may not resolve the cooling flow issue, it could contribute to maintaining thermal balance in cluster cores and support gas mixing processes once the ICM reaches temperatures of several keV during the later stages of structure formation.

Stellar feedback has been considered as a possible contributor to heating the ICM. Nonetheless, numerical simulations indicate that, even when processes like star formation, metal enrichment, and stellar feedback are incorporated, the central regions of clusters still exhibit an overabundance of cold gas and excessive star formation (e.g. Nagai et al., 2007; Borgani & Kravtsov, 2011). Other hypotheses involve the interplay of multiple physical mechanisms, such as the combination of sound waves and thermal conduction (e.g. Ruszkowski et al., 2004) or turbulence working in tandem with conduction (e.g. Dennis & Chandran, 2005).

Despite the above proposed solutions, the leading explanation for the suppression of cooling flows in galaxy clusters is feedback from the AGN residing in the BCG. At early cosmic times ($z > 1$), this mechanism operates relatively clearly, as AGNs are typically in the quasar phase, radiating at luminosities between 10^{44} – 10^{45} erg/s, which are on par with the X-ray output of the cooling ICM, and launching jets that can extend to scales of roughly 100 kpc. This feedback process is often described as self-regulating: the cooling of hot gas promotes the condensation of cold clouds, which in turn fuel the AGN and trigger outflows that reheat the surrounding ICM (Gaspari et al., 2020).

In contrast, at later epochs, AGNs in BCGs generally shift to a radio-mode phase, with much lower radiative outputs below 10^{43} erg/s, falling short of the cooling region's X-ray luminosity (Russell et al., 2013; Fujita et al., 2014). Even so, these systems often exhibit X-ray cavities whose associated power range from 10^{42} to 10^{45} erg/s, sufficient to counterbalance the cooling losses (Rafferty et al., 2006). In the absence of high radiative efficiency, these power range is attributed to nearly fully efficient mechanical feedback, which inflates buoyant bubbles in the ICM. However, a key unresolved issue is how this energy can be redistributed uniformly throughout the cluster core.

The impact of AGN feedback on galaxy cluster cores has been extensively investigated through numerical simulations. Early efforts by Di Matteo et al. (2005) and Springel et al. (2005a) showed that AGN-driven heating could effectively lower the excessive star formation in central regions and reproduce the observed M_{BH}–σ relation. Building on this foundation, later models, such as the ones presented by Sijacki et al. (2007) and Fabjan et al. (2010), introduced more sophisticated implementations, such as a high-efficiency radio-mode feedback operating at low accretion rates. These modifications were designed to reduce star formation at low redshifts by mimicking the dominant AGN state in the present-day universe, where AGNs typically operate in radio mode and observed accretion rates are low.

However, raising the efficiency of AGN feedback is not without its drawbacks. For example, simulations by Fabjan et al. (2010) indicate that even with enhanced radio-mode feedback, the issue of overly high stellar mass fractions in massive clusters, ranging from [0.52- 18.51] $\times 10^1 4$ $M_\odot$, remains unresolved. Moreover, increasing feedback efficiency introduces new complications at the scale of galaxy groups, where it leads to decreased gas content and elevated central entropy levels.

This tendency toward overheating in group-scale halos has also been emphasized in the independent work by Gaspari et al. (2013), who argues that galaxy groups cannot simply be

treated as smaller analogues of clusters. Instead, they require more moderate, finely tuned feedback to avoid excessive heating. More recently, Bahar et al. (2024) compared simulation results to observational data from large samples of galaxy groups in the eRosita All-Sky Survey (eRASS1), finding that simulations tend to over-predict central entropy levels in these systems.

Furthermore, there is growing evidence that the core status of clusters may not be completely fixed but may have some transient nature. Rossetti et al. (2011) presents observational examples of Non-Cool-Core (NCC) clusters with residual low-entropy gas—so-called "CC remnants", suggesting that these systems may have once harbored cool cores, subsequently disrupted by mergers or AGN activity. Similarly, simulations indicate that the thermodynamic state of the core can evolve over time, with clusters transitioning between CC and NCC phases depending on the balance of heating and cooling processes (Rasia et al., 2015; Hahn et al., 2017). This transient nature of core states challenges the traditional view of the CC/NCC distribution as a static classification and stresses the importance of understanding the dynamic interplay between AGN feedback, mergers, and the cluster environment.

While AGN feedback is crucial in modulating core properties, it appears insufficient to account for the full transition from a CC to a NCC state. In fact, the energy needed for such a core change is of the order of $[1-4] \times 10^{61}$ ergs , far beyond the most energetic AGN burtst observed in the late universe (see McCarthy et al., 2008). Thus, other mechanism need to be responsible of these kind of changes. In this context, mergers between clusters are thought to be a key driver of thermodynamic evolution of cluster cores.

Numerical simulations demonstrate that mergers can heat the core gas, disrupt the cooling flow, and mix the high-entropy gas from the outer regions into the core, effectively transforming a CC into an NCC cluster (Burns et al., 2008). In fact, galaxy cluster mergers are capable of releasing gravitational binding energies of the order of 10^{64} erg (Sarazin, 2002), which would suffice to explain a CC - NCC transition. Observationally, CC clusters are rarely found among dynamically active systems with strong evidence of merge activity (Hoffer et al., 2012; Sanders et al., 2010) and NCC clusters often exhibit disturbed X-ray morphologies and evidence of recent merger activity (Rossetti & Molendi, 2010; Hudson et al., 2010). Further evidence is provided by the findings reported by Chen et al. (2007), who found a trend whereby the fraction of CC clusters decreases towards the most massive, dynamically young systems. This trend has also been reproduced qualitatively by the simulations of Burns et al. (2008) and Planelles & Quilis (2009). Mergers are thought to not only disrupt established cool cores but can also prevent their formation by delaying the re-establishment of dense, low-entropy gas in the center. However, the efficiency and timescale of this process remain areas of active study (Burns et al., 2008; Rossetti & Molendi, 2010). Clearly, both the formation history and AGN feedback play a mutually influential role when determining the state of the cluster centers.

Beyond the general challenge of understansing the mechanism controling the cooling and heating in cluster cores, a persistent issue lies in defining what truly separates CC clusters from NCC ones. Various studies adopt differing criteria to identify CC clusters—some rely on a central drop in temperature (e.g. Sanderson et al., 2006; Burns et al., 2008), short Central Cooling Time (CCT) (e.g. Bauer et al., 2005; O'Hara et al., 2006), or significant classical mass deposition rate (e.g. Chen et al., 2007). The criteria used for classifcation is nontrivial and has implications for cosmological studies since, when used as cosmological probes, clusters are often segregated

into CC/NCC subsamples. To address this, Hudson et al. (2010) examined 16 different CC diagnostics across 64 galaxy clusters from the HIFLUGCS (HIghest X-ray FLUx Galaxy Cluster Sample) catalog (Reiprich & Boehringer, 2002). Their analysis identified CCT as the parameter showing the clearest bimodal distribution, making it the most effective discriminator between CC and NCC systems.

In this work, we adopt the classification scheme established by Hudson et al. (2010), which categorizes clusters into Strong Cool-Core (SCC), Weakly Cool-Core (WCC), and NCC systems based on CCT. Specifically, clusters with CCTs below 1 $h_{71}^{-1/2}$Gyr are classified as SCC, those with cooling times between 1 $h_{71}^{-1/2}$Gyr and 7.7 $h_{71}^{-1/2}$Gyr as WCC, and clusters with cooling times above 7.7 $h_{71}^{-1/2}$Gyr as NCC. According to this classification, Hudson et al. (2010) found that SCC clusters exhibit low central entropy and a pronounced central temperature drop. In contrast, WCC clusters show moderately elevated central entropy and temperature profiles that are either flat or mildly decreasing toward the center, while NCC clusters are characterized by high central entropy, consistent with prior studies.

In summary, galaxy cluster cores are complex environments where the balance between cooling and heating processes shapes the thermodynamic properties of the ICM. The distinction between SCC, WCC, and NCC clusters reflects differences in entropy, cooling time, and feedback efficiency, influenced by both AGN activity and the cluster's formation history. Understanding these processes is essential for constructing a comprehensive picture of cluster evolution and for interpreting the thermodynamic structure observed in both relaxed and disturbed systems at our current time.

Part III

Methods

2 | Numerical Cosmology

Numerical simulations are one of the fundamental tools on which this thesis is based. They provide a powerful way to study the formation and evolution of cosmic structures across a wide range of spatial mass and time scales. By solving the complex equations governing gravity, hydrodynamics, and a host of astrophysical processes, simulations allow us to explore scenarios that are inaccessible to direct observation or analytical modeling. They offer a way to study mechanisms driving the growth of LSSs, the formation of galaxies, and the interplay between baryonic and dark matter components in the Universe across cosmic time.

Cosmological simulations can be broadly classified into three categories, each tailored to address different aspects of structure formation. Dark-matter-only simulations focus exclusively on gravitational interactions, modeling the growth and clustering of dark matter halos without considering the complex physics of baryons. The computational efficiency of these simulations has been significantly improved through the use of highly optimized and parallelized gravity solvers. Such simulations have been instrumental in establishing the LSS of the Universe and the hierarchical assembly of dark matter halos. Hydrodynamic simulations, by contrast, incorporate the baryonic component, modeling gas dynamics, star formation, feedback from SNe and active galactic nuclei (AGN), and chemical enrichment processes. While computationally more demanding, they offer a detailed and realistic picture of galaxy formation and evolution. Finally, semi-analytic models (SAMs) apply simplified, analytic prescriptions for baryonic processes onto the backbone of dark-matter-only simulations. These models are less computationally demanding, enabling rapid exploration of large parameter spaces.

In this work, we make use of cosmological hydrodynamical simulations to study the formation and evolution of galaxy clusters and their internal structure. Given the breadth and depth of the literature on cosmological simulations, this section aims to provide a general overview of the main techniques, focusing in detail on those most relevant to the work presented in this thesis. For readers seeking a more comprehensive treatment of the field, we refer to classic reviews and foundational works such as Benz (1990); Springel (2005); Dolag et al. (2005, 2008, 2009); Dehnen & Read (2011); Borgani & Kravtsov (2011); Hirschmann et al. (2014a); Springel (2016); Vogelsberger et al. (2020).

2.1 Standard Cosmological Simulations

The formation and evolution of structures in the Universe, from the smallest galaxies to the largest clusters, is a fundamentally nonlinear process, thus even if linear perturbation theory provides a solid framework for describing the growth of density fluctuations at early times, it becomes insufficient once these overdensities grow and collapse under their own gravity. Understanding this nonlinear regime requires the use of numerical simulations, which have become an essential tool in modern cosmology. Thanks to the rapid advancements in computational power over the last few decades, simulations now provide a detailed and realistic representations of the Universe across a wide range of scales.

At the core of these simulations lies gravity, which governs the large-scale evolution of matter and is primarily responsible for the hierarchical assembly of cosmic structures through the accretion and merging of dark matter halos. However, as we have seen in Section 1.2, gravity alone cannot account for the full complexity observed in galaxies and clusters, but simulations need to include the dynamics of baryonic matter and the various astrophysical phenomena that influence galaxy formation and evolution.

These processes can broadly be categorized into three main groups: gravitational dynamics, hydrodynamical interactions, and galactic-scale processes. Gravitational dynamics are typically modeled using N-body methods, which simulate the behavior of a large number of particles representing collisionless components such as dark matter and stars. These methods solve the evolution of the phase-space distribution function, which satisfies the collisionless Boltzmann equation, coupled to Poisson's equation for the gravitational potential. To make these calculations tractable for large particle numbers, several approximation techniques are commonly employed, including Particle-Mesh (PM) methods, hierarchical Tree algorithms, and hybrid Tree-PM approaches, which will be explored in detail in Sec 2.2.

Gas dynamics can be approached using two primary numerical schemes: Eulerian and Lagrangian methods. Eulerian approaches track the evolution of fluid quantities on a fixed or adaptive spatial grid. Lagrangian methods, by contrast, follow discrete fluid elements or particles as they move through space, naturally adapting resolution to regions of interest such as the dense cores of galactic halos. Both approaches have their advantages, and the choice often depends on the goals of the simulation and the specific physical processes being modeled; both will be presented in Sec 2.3.

Simulating galaxy formation presents an additional set of challenges, as it requires bridging vast differences in scale, from the megaparsec regimes relevant for galaxy clusters down to the sub-parsec scales where star formation and feedback occur. Directly resolving all relevant physical processes is computationally prohibitive. Instead, simulations rely on sub-resolution models to describe phenomena that take place below the resolution limit, such as radiative cooling, star formation, chemical enrichment, and feedback from SNe and AGN. These subgrid models aim to capture the cumulative effects of small-scale processes on the large-scale evolution of galaxies and galaxy clusters. Further details will be explored in Sect. 2.4.

2.2 N-body Solvers

2.2.1 PM code

The PM method is one of the foundational algorithms used in cosmological N-body simulations to compute the gravitational interactions between particles. In the PM scheme, the simulation volume, typically a cube with periodic boundary conditions, is divided into a uniform three-dimensional grid consisting of M^3 cells, where $N = M^3$ is the total number of mesh points. Each particle's mass is interpolated onto this grid to generate a continuous mass density field (Eastwood & Hockney, 1974; Efstathiou et al., 1985). This density field serves as the source term for solving Poisson's equation in comoving coordinates:

$$\nabla^2\phi(\vec{x},t) = 4\pi G a^2(t)\left[\rho(\vec{x},t) - \bar{\rho}(t)\right]. \tag{2.1}$$

The potential ϕ is efficiently computed by transforming the equation into Fourier space, where it becomes an algebraic relation:

$$k^2\phi_{\vec{k}} = -4\pi G\rho_{\vec{k}}. \tag{2.2}$$

This transformation allows for the use of the fast Fourier transform, which reduces the computational complexity to $O(N \log N)$, where N is the number of mesh points. Once the gravitational potential is obtained, its gradient is calculated to yield the gravitational force at each mesh point, which is then interpolated back to the particle positions to update their velocities and positions.

A key advantage of the PM method is its scalability: the computational cost for evaluating gravitational forces scales linearly with the number of particles, assuming the grid size remains fixed. However, this comes at the expense of spatial resolution. The forces are smoothed over the scale of the grid spacing, limiting the accuracy in regions where fine spatial detail is required, such as the dense cores of dark matter halos (Hockney & Eastwood, 1981; Klypin & Holtzman, 1997; Bagla, 2005).

The Particle-Particle-Particle-Mesh (P3M) algorithm attempts to address this issue by calculating the potential, performing direct summation calculations for particles that are close to each other (typically within a few mesh cell lengths), and through the PM method for the others (Hockney & Eastwood, 1981; Efstathiou et al., 1985; Bouchet et al., 1985). Another technique to increase resolution in grid-based methods involves the use of Adaptive Mesh Refinement (AMR). In this scheme, regions of higher particle density are assigned additional refinement levels, subdividing the mesh locally to improve spatial resolution (Berger & Oliger, 1984; Berger & Colella, 1989; Kravtsov et al., 1997; Teyssier, 2002; O'Shea et al., 2004; Bryan et al., 2014). This allows the simulation to allocate computational resources dynamically, refining the mesh where high resolution is needed (e.g., dense structures like halos), while keeping coarser grids in less dense regions.

2.2.2 The Tree algorithm

The Tree algorithm is a widely adopted method for computing gravitational interactions in N-body simulations, offering a significant improvement in efficiency compared to direct pairwise summation methods. Originally introduced by (Barnes & Hut, 1986), the approach strikes a balance between computational speed and force resolution, making it particularly valuable in cosmological applications where a wide range of scales must be resolved.

At its core, the Tree algorithm organizes particles into a hierarchical structure that groups them according to their spatial proximity. Instead of calculating gravitational forces directly between each pair of particles, a process that scales as $O(N^2)$ and quickly becomes intractable for large particle numbers, the Tree method reduces computational complexity to approximately $O(N \log N)$ by treating distant particle groups collectively through multipole expansions.

The algorithm begins by recursively subdividing the simulation domain into cubic regions, forming a tree structure where each node represents a spatial cell. In three dimensions, the root cell (representing the entire simulation box) is divided into eight subcells in an octree configuration. This subdivision continues until the resulting cells contain either a single particle or no particles at all. Each node of the tree stores aggregated information about the particles within its subregion, typically the total mass and the center-of-mass position.

Once the tree structure is built, the next step involves computing gravitational forces through a process known as "walking the tree." For any given particle, the algorithm traverses the tree to evaluate the gravitational influence of other groups of particles. Distant nodes can be approximated as single pseudo-particles if they satisfy an acceptance criterion based on an opening angle parameter, θ. Specifically, if the angle subtended by a node is smaller than θ, the node's multipole expansion is used directly. If not, the node is opened, and the algorithm descends further into its subcells. This hierarchical approach allows the algorithm to bypass unnecessary fine-grained calculations for distant regions while retaining high accuracy for nearby interactions. Thus, the choice of the opening angle θ governs the accuracy of the calculation. Smaller values of θ yield more accurate force calculations but increase the computational cost by forcing the algorithm to explore deeper levels of the tree. In contrast, larger θ values enhance speed at the expense of precision. We note that for strongly clustered distributions (like the ones found in evolved cosmological structures), performance can degrade due to the increased number of particles requiring direct interactions (Dolag et al., 2008).

Parallel implementations of the Tree algorithm face challenges related to data locality and communication overhead. Efficient load balancing and minimizing communication between processors are important steps for improving the performance on modern supercomputers. One popular strategy to address these challenges involves sorting particles along a space-filling Peano-Hilbert curve. This technique ensures that particles located close together in space are also close in memory and more likely to be assigned to the same processor. As a result, the Peano-Hilbert ordering minimizes communication between nodes and improves cache efficiency, making it highly suitable for parallelized Tree algorithms (Springel, 2005).

Figure 2.1 demonstrates the structure of the Peano-Hilbert curve in two and three dimensions, highlighting how the curve recursively fills space to maintain locality. Figure 2.2 illustrates the computational advantages of the Tree algorithm over direct summation. In the left panel, each

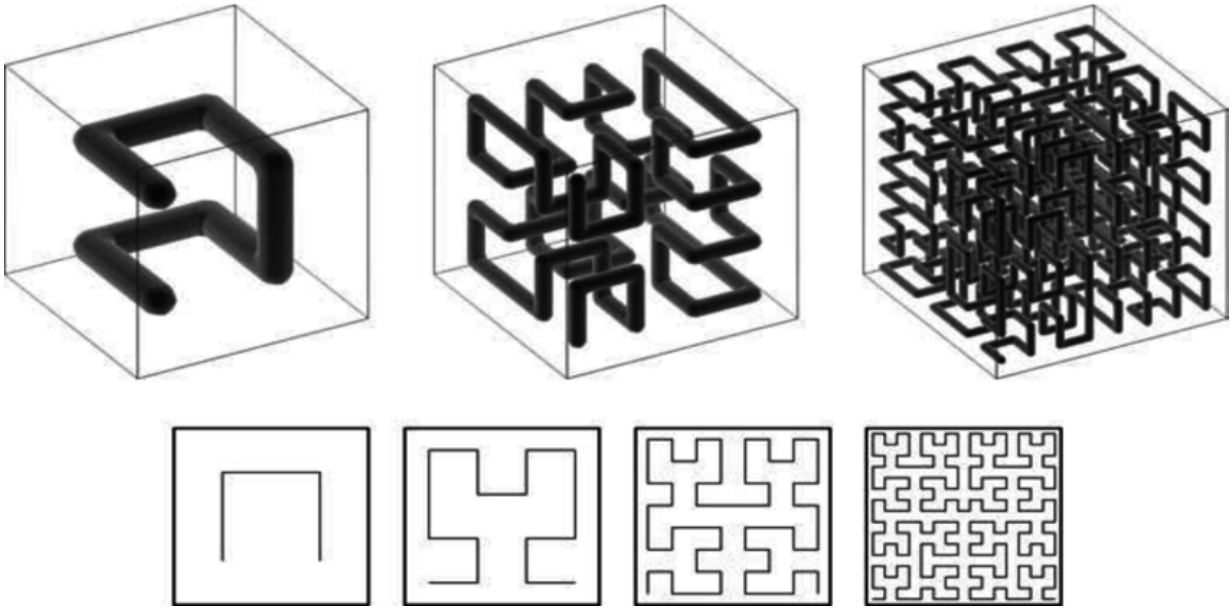


Figure 2.1: Peano–Hilbert space-filling curve illustrated in two dimensions (bottom) and three dimensions (top). Credits: Springel (2005).

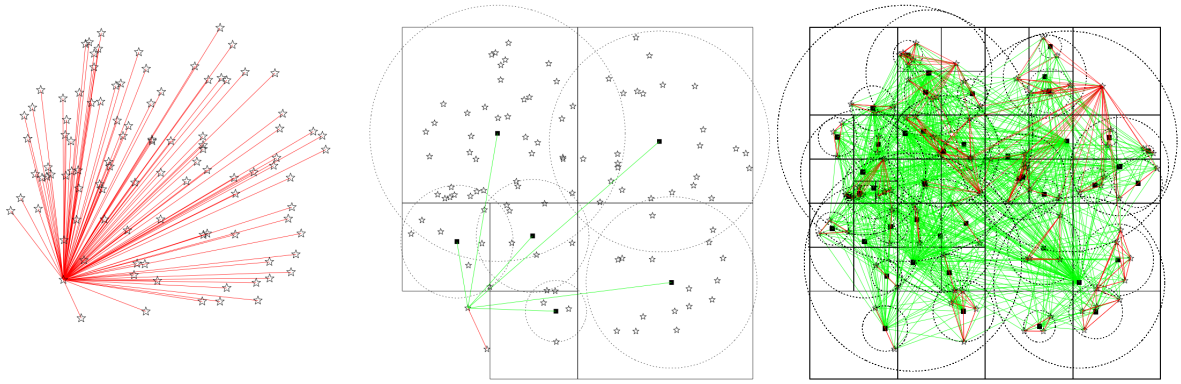


Figure 2.2: Left: Illustration of force computation for one out of 100 particles (shown as asterisks) in two dimensions for clarity, using direct summation. Each line represents an individual force calculation between pairs of particles. Middle: Approximate force calculation for the same particle using a tree algorithm. Opened cells are depicted as black squares, with their centers marked by solid squares (z) and sizes indicated by dotted circles (w). Each green line represents a cell-to-particle interaction. Right: Approximate force computation for all 100 particles using the tree method, requiring only 902 cell-particle and 306 particle-particle interactions (with opening angle $\theta = 1$ and maximum particles per cell $n_{\max} = 1$), in contrast to the 4950 direct particle-particle interactions needed with brute-force summation. Credits: Dehnen & Read (2011).

particle computes forces with every other particle individually, resulting in an $O(N^2)$ scaling. In contrast, the middle panel shows how the Tree algorithm groups distant particles into composite nodes, significantly reducing the number of force evaluations. As a result, the Tree algorithm makes it feasible to model gravitational interactions in simulations with millions or billions of particles.

Tree algorithms form the backbone of many modern cosmological simulation codes, such as GADGET and its variants (Vogelsberger et al., 2020). These codes often implement hybrid TreePM methods, combining the Tree algorithm for short-range interactions with a PM scheme for long-range forces, as will be presented in the next subsection.

2.2.3 Tree-PM code

The TreePM algorithm represents a hybrid approach that combines the PM method and the Tree algorithm. Originally introduced by Hernquist & Katz (1989) and further refined by (Xu, 1995), this technique profits from the strengths of both schemes to achieve efficient and accurate gravitational force computations over a wide range of spatial scales.

The PM method is highly effective at computing long-range gravitational forces on large scales, efficiently solving Poisson's equation on a mesh through Fourier transforms. However, its spatial resolution is fundamentally limited by the grid spacing, making it inadequate for resolving small-scale interactions where higher force accuracy is required. In contrast, the Tree algorithm excels at handling short-range forces by adaptively refining spatial resolution and directly computing interactions within localized groups of particles.

The TreePM method integrates these complementary strategies by splitting the gravitational force computation into two distinct components: a long-range part calculated using the PM scheme and a short-range part handled by the Tree algorithm. This decomposition is typically performed in Fourier space, where the gravitational potential is separated into long- and short-range contributions:

$$\phi_{\vec{k}} = \phi_{\vec{k}}^{\text{long}} + \phi_{\vec{k}}^{\text{short}}. \tag{2.3}$$

The long-range potential is computed as:

$$\phi_{\vec{k}}^{\text{long}} = \phi_{\vec{k}} \exp\left(-k^2 r_s^2\right). \tag{2.4}$$

Here, r_s denotes the spatial scale that determines the transition between long-range and short-range interactions. The exponential factor smoothly suppresses the short-range contributions in the PM calculation.

The short-range component, representing the residual forces not captured by the PM grid, is then evaluated in real space using the Tree algorithm. For distances much smaller than the box size, $r_s \ll L$, the short-range potential is given by:

$$\phi_{\text{short}}(\vec{x}) = -G \sum_i \frac{m_i}{r_i} \operatorname{erfc}\left(\frac{r_i}{2r_s}\right), \tag{2.5}$$

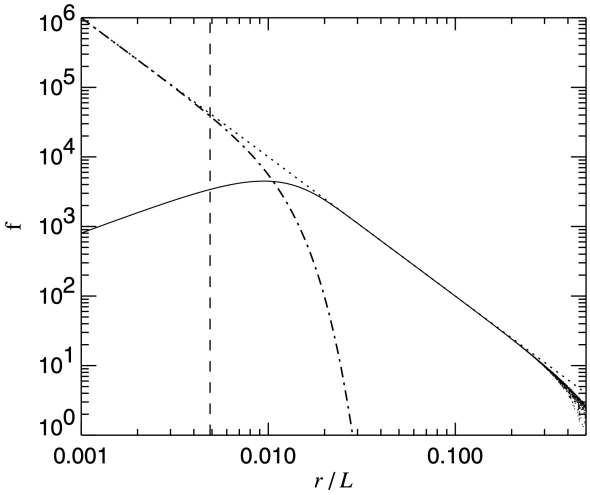


Figure 2.3: Spatial decomposition of the force in a Tree-PM, computed by placing a unit-mass particle at a random location within a periodic box and measuring the resulting forces on a set of randomly distributed test particles using the simulation code GADGET-2. Credits: Springel (2005).

where $r_i = |\vec{x} - \vec{r}_i|$ is the distance between the particle of interest and the source particle i. The complementary error function, erfc, introduces a smooth cutoff for distances beyond r_s, ensuring that the Tree code focuses computational effort where it is most needed.

By combining these two schemes, TreePM achieves a balance between accuracy and computational cost. The PM method handles the large-scale, long-range gravitational field efficiently, while the Tree algorithm refines the force computation at small scales, improving spatial and temporal resolution. The force decomposition in TreePM is illustrated in Fig. 2.3, where the gravitational force is shown to be split between long-range (dashed line) and short-range (dot-dashed line) components, with the total force (solid line) smoothly transitioning between the two regimes. This makes TreePM particularly well suited for cosmological simulations that require resolving both the LSS of the Universe and the internal properties of dense regions like galaxy clusters. In fact, the TreePM method is widely adopted in state-of-the-art simulation codes, including GADGET-2 and GADGET-3 (Springel, 2005).

A further advantage of the TreePM approach is its compatibility with parallel computing architectures. Since the PM component relies on global grid computations and the Tree calculations are localized to high-density regions, the method allows efficient domain decomposition and distributed memory parallelization. As noted in (Springel, 2005), this division of labor enhances scalability on massively parallel supercomputers, a key requirement for modern cosmological simulations. Moreover, the Tree portion of the code is typically self-contained. Once the particles within a given tree node are assigned, no further data from outside the node is required for subsequent calculations, facilitating efficient load balancing and data locality.

For the simulations presented in this thesis, we employ OPENGADGET3 (OG3 Developers et. al. in prep., Groth et al. (2023); Sala et al. (2024)) and AREPO (Springel, 2010; Weinberger et al., 2020), which use the TreePM scheme described above.

2.3 Hydrodynamical Simulations

2.3.1 Smoothed Particle Hydrodynamics

Smoothed Particle Hydrodynamics (SPH) is a mesh-free, Lagrangian method commonly used to model fluid dynamics. Originally developed by Gingold & Monaghan (1977) and Lucy (1977) for stellar applications, SPH has since become a widely adopted technique in cosmology, particularly in simulations that aim to describe the gravitational and hydrodynamical evolution of cosmic structures (see Price, 2012; Rosswog, 2015, for reviews). Its fully Lagrangian nature makes SPH particularly compatible with N-body or Tree-based gravity solvers, as both follow the trajectories of discrete particles through space and time.

In SPH, the fluid is discretized into particles that carry physical quantities such as mass, density, and internal energy. These particles move with the fluid flow, and their properties are computed by interpolating contributions from neighboring particles within a defined smoothing length h. This approach naturally adapts resolution to local density: higher resolution is achieved in dense regions where particles are more concentrated, making SPH particularly effective at following processes such as star formation and the collapse of gas clouds in galaxies.

The motion of the SPH particles follows the Lagrangian form of the hydrodynamic equations. For an ideal, non-relativistic fluid under a gravitational potential ϕ, the evolution equations can be expressed as:

$$\frac{d\rho}{dt} + \rho \nabla \cdot \vec{v} = 0, \tag{2.6}$$

$$\frac{d\vec{v}}{dt} + \frac{1}{\rho}\nabla P + \nabla\phi = 0, \tag{2.7}$$

$$\frac{du}{dt} + \frac{P}{\rho}\nabla \cdot \vec{v} = 0. \tag{2.8}$$

These represent the continuity, momentum, and energy conservation equations, respectively. The Poisson equation relates the gravitational potential to the density, and an equation of state (usually polytropic) is introduced to connect pressure P to density ρ.

To calculate each fluid quantity an estimation is performed by smoothing over neighboring particles using a kernel function $W(\vec{r} - \vec{r}', h)$. This kernel is generally symmetric, non-negative, and normalized to integrate to unity. A generic scalar field $f(\vec{x})$ is approximated by the sum:

$$\bar{f}(\vec{x}) = \sum_j m_j \frac{f_j}{\rho_j} W(\vec{x} - \vec{r}_j, h). \tag{2.9}$$

Derivatives of quantities such as pressure or velocity gradients are computed by differentiating the kernel:

$$\nabla A(\vec{x}) = \sum_j m_j \frac{A_j}{\rho_j} \nabla W(\vec{x} - \vec{r}_j, h). \tag{2.10}$$

This formulation makes SPH efficient for following the evolution of astrophysical fluids, as it depends only on local interactions within the smoothing length h.

Several kernel functions are commonly used, each with different characteristics affecting accuracy and stability. Choices include the cubic spline kernel (Monaghan & Lattanzio, 1985), the quintic spline (Morris, 1996), and higher-order Wendland kernels (Wendland, 1995; Dehnen & Aly, 2012), which are favored for reducing numerical noise and suppressing pairing instability. In practice, SPH simulations often employ a large number of neighbors to minimize discretization errors and improve stability. The effective volume of a particle is related to its smoothing length and neighbor number N_{ngb}, with the density estimated by:

$$\rho_i = \sum_{j \in \mathrm{Ngb}} m_j W(|\vec{r}_i - \vec{r}_j|, h_i). \tag{2.11}$$

Smoothing lengths are adjusted adaptively to maintain a roughly constant number of neighbors, ensuring higher resolution in regions of higher density.

The equations of motion in SPH include hydrodynamical accelerations derived from pressure gradients. In OpenGadget3, a fully conservative formulation is adopted (Springel & Hernquist, 2002), with accelerations given by:

$$\vec{a}_{\mathrm{hydro},i} = -\sum_{j \in \mathrm{Ngb}} m_j \left[f_i \frac{P_i}{\rho_i^2} \nabla_i W_{ij}(h_i) + f_j \frac{P_j}{\rho_j^2} \nabla_i W_{ij}(h_j) \right], \tag{2.12}$$

where f_i is a correction factor accounting for changes in smoothing length:

$$f_i = \left(1 + \frac{h_i}{3\rho_i} \frac{\partial \rho_i}{\partial h_i} \right)^{-1}. \tag{2.13}$$

Standard SPH implementations have historically struggled with issues such as the suppression of fluid mixing and the accurate treatment of shocks. Artificial viscosity terms are typically included to capture shocks by introducing dissipation where necessary (Morris, 1996; Agertz et al., 2007). At the same time, artificial conductivity terms (Price, 2008) help reduce spurious pressure discontinuities at contact interfaces. In OpenGadget3, these dissipation terms are implemented with spatially dependent triggers (Beck et al., 2016), minimizing their impact in regions where they are not physically required. Still, SPH struggles with some intrinsic limitations, which include difficulties in modeling fluid mixing and instabilities without additional corrective terms.

Nevertheless, its fully Lagrangian formulation, where mass elements are conserved by construction, and resolution automatically follows mass distribution, makes SPH highly efficient in cosmological simulations. Therefore, SPH remains a popular choice in large cosmological hydrodynamical simulations, such as those run with OpenGadget3 and other modern codes.

2.3.2 Mesh codes

Instead of Lagrangian methods like SPH, many cosmological simulations employ mesh-based approaches for modeling hydrodynamics. These methods are typically grounded in an Eulerian

framework, in which the discretised quantity is volume rather than mass (as in SPH). Each cell in the grid holds information about local fluid properties, such as density, velocity, and pressure, which are evolved over time by calculating fluxes between adjacent cells. Mesh-based codes have long been popular for their ability to resolve shocks and fluid instabilities with high precision (Teyssier, 2002; Agertz et al., 2007; Toro, 2009).

The most straightforward implementation of mesh-based hydrodynamics uses a static Cartesian grid. Such grids are simple to construct and easily coupled with PM gravity solvers, since mass densities are already assigned to the grid structure. However, uniform grids suffer from an intrinsic lack of adaptivity: increasing the spatial resolution in localized regions requires raising the resolution globally, which significantly inflates the computational cost, particularly in simulations with large dynamic ranges (Kravtsov et al., 1997; Bryan & Norman, 1997).

To address these shortcomings, AMR schemes have been developed (Berger & Oliger, 1984; Berger & Colella, 1989). AMR dynamically refines the grid structure by subdividing cells in regions where increased resolution is required, such as high-density environments or around shock fronts. Each refinement step typically halves the cell size, resulting in an eightfold resolution increase in three dimensions. Codes like ART (Kravtsov et al., 1997) and AP3M (Couchman, 1991) use AMR to efficiently allocate computational resources, maintaining high spatial resolution where needed while keeping coarser grids in less critical regions.

Refinement criteria can be tailored depending on the specific scientific goals, and conservation laws must be carefully enforced at the interfaces between different resolution levels to avoid introducing unphysical discontinuities or numerical artifacts.

Hydrodynamical mesh codes typically solve the equations of mass, momentum, and energy conservation using finite-volume methods. For each grid cell, fluxes are computed across cell interfaces by solving local Riemann problems, which allows for accurate modeling of discontinuities like shocks. The conservative nature of these schemes ensures that mass, momentum, and energy are preserved to a high degree of accuracy.

The update of conserved quantities over a time step involves calculating the net flux across the cell boundaries. In one dimension, the basic update formula for the conserved variables $U = (\rho, \rho v, \rho e)$ is:

$$\Delta U_i = -\frac{\Delta t}{\Delta x}\left(F_{i+1/2} - F_{i-1/2}\right). \tag{2.14}$$

This flux-based formulation, combined with high-order reconstruction schemes and slope limiters, ensures stable and accurate solutions.

One of the key strengths of mesh codes is their ability to accurately model shocks and capture fluid instabilities, such as Kelvin-Helmholtz and Rayleigh-Taylor instabilities, which are often poorly resolved in SPH schemes. They also do not require artificial viscosity or conductivity terms to handle discontinuities, as these are naturally managed by the Riemann solver (Agertz et al., 2007).

However, mesh codes face challenges when dealing with high-velocity bulk flows. In Cartesian grids, such flows can lead to numerical diffusion and errors due to the lack of Galilean invariance (O'Shea et al., 2005; Heitmann et al., 2008). On top of this, grid-alignment effects can artificially

influence fluid motion, and early refinement in AMR schemes can hinder the resolution of small halos at late times in cosmological simulations (Agertz et al., 2007; Springel, 2010).

To overcome some of these limitations, a different approach known as the moving mesh method has been developed. AREPO (Springel, 2010) is one of the most prominent implementations of this concept. It employs a moving, unstructured mesh based on a Voronoi tessellation, where mesh-generating points move with the local velocity of the fluid. This design allows the mesh to adapt continuously to the flow, ensuring that fluxes are computed in the rest frame of the interface between cells, significantly reducing numerical diffusion. Since the resolution follows the mass distribution, AREPO retains one of the key advantages of SPH, while its mesh-based nature allows for an accurate treatment of shocks and discontinuities without the need for artificial diffusion terms.

AREPO's Voronoi mesh is generated using an incremental insertion algorithm based on the Bowyer-Watson scheme (Bowyer, 1981; Watson, 1981), optimized for parallel performance. This algorithm efficiently updates the mesh at each timestep, ensuring that cells remain well-shaped and isotropic. Additionally, the moving mesh approach mitigates grid-alignment effects and maintains Galilean invariance, providing an accurate description of bulk flows and complex dynamical interactions in cosmological simulations.

2.4 Galaxy Formation Models

Galaxy formation is governed by several physical mechanisms, including gas dynamics, radiative cooling, star formation, chemical enrichment, and feedback from stars and AGNs. These processes are intrinsically multi-scale, spanning many orders of magnitude in both space and time, and are typically unresolved in cosmological simulations. As such, they must be incorporated through phenomenological subgrid models.

This section presents an overview of the two state-of-the-art galaxy formation models implemented in the simulations used throughout this thesis: the Magneticum and IllustrisTNG models. Both galaxy formation models employ sophisticated treatments of baryonic physics to capture the interplay between dark matter, gas, and stellar components, while incorporating feedback mechanisms that regulate star formation and drive galaxy evolution. Although these models differ in their numerical approaches and specific implementations, both provide valuable insights into the processes shaping galaxies and galaxy clusters across cosmic time.

2.4.1 Magneticum

The Magneticum model builds upon the core framework of the GADGET-3 code, integrating advanced subgrid models to represent key baryonic processes that are otherwise unresolved in cosmological simulations (OG3 Collaboration in prep., Beck et al. (2016); Groth et al. (2023); Sala et al. (2024)). Here we provide a summary on the implementation of the main components of the galaxy formation model:

Radiative Cooling and Heating Processes: Magneticum includes detailed models for radiative cooling and heating. The cooling rates are computed by considering both primordial and metal-line cooling contributions. Gas cooling depends on its temperature and metallicity and is calculated under the assumption of ionization equilibrium in the presence of an evolving ultraviolet background radiation field (Sutherland & Dopita, 1993; Katz et al., 1996; Wiersma et al., 2009; Hirschmann et al., 2014a).

For primordial gas (hydrogen and helium), the cooling is dominated by bremsstrahlung at high temperatures (above 10^6 K), where free electrons interact with ionized nuclei, emitting photons and losing energy. At temperatures below 10^6 K, processes such as collisional excitation and recombination dominate the cooling. Metal-line cooling is introduced using pre-computed lookup tables, allowing for an efficient and metallicity-dependent treatment (Dolag et al., 2017). The ionizing UV background is modeled following the prescription of (Madau et al., 1999), representing the contribution of quasars reionizing the intergalactic medium at $z \sim 6$. The total cooling rate $\Lambda(T, Z)$ is a function of temperature T and metallicity Z, and energy losses are integrated into the thermal evolution of gas particles.

Star Formation and the Multiphase Interstellar Medium (ISM): Star formation within Magneticum follows the sub-resolution multiphase model initially proposed by (Springel & Hernquist, 2003). Gas particles eligible for star formation are those exceeding a density threshold and undergoing cooling. These particles are treated as a composite of cold and hot gas phases, maintaining pressure equilibrium.

The conversion of cold gas into stars proceeds on a characteristic timescale t_*, which scales with gas density following a Schmidt-like relation (Schmidt, 1959):

$$\dot{\rho}_* = (1 - \beta_{\mathrm{SNe}}) \frac{\rho_c}{t_*} \,. \tag{2.15}$$

Here, ρ_c is the cold gas density, and β_{SNe} is the mass fraction of newly formed stars that end their lives as Type II Supernovae (SNe II). This fraction depends on the chosen Initial Mass Function (IMF), which in Magneticum is assumed to follow the Chabrier IMF (Chabrier, 2003).

Feedback from SNe plays a critical role in regulating star formation. A portion of the SN energy is used to drive galactic winds, which expel gas from star-forming regions, reducing cooling and suppressing further star formation. For galaxies up to $10^{10} M_\odot$ SNe help mitigate the "overcooling problem" that can otherwise lead to excessive star formation rates in massive galaxies.

Chemical Enrichment and Stellar Yields: The chemical evolution model in Magneticum tracks the production and distribution of heavy elements released by stellar populations. The primary sources of metal enrichment included are, Type Ia Supernovae (SNe Ia), SNe II and Asymptotic Giant Branch (AGB) stars.

The mass return and metal yields for each channel are implemented using the stellar lifetimes from Padovani & Matteucci (1993), and the yields are drawn from Thielemann et al. (2003) for SNe Ia, Woosley & Weaver (1995) for SNe II, and Karakas (2010) for AGB stars. The lifetime function $\tau(M)$ governs when stars of a given mass contribute to the chemical enrichment:

$$\tau(M) = \begin{cases} 10^{(1.34-\sqrt{1.79-0.22(7.76-\log M)})/0.11} - 9, & \text{if } M \leq 6.6 M_{\odot} \\ 1.2 M^{-1.85} + 0.003, & \text{otherwise .} \end{cases} \tag{2.16}$$

These metals are injected into the surrounding gas particles, enriching the interstellar and intergalactic medium over time. This process allows the simulation to model the metallicity dependence of cooling, star formation, and feedback processes.

SMBH and AGN Feedback: Magneticum includes a detailed prescription for SMBH formation and evolution. BHs are seeded in halos exceeding a mass threshold (typically $M_{\rm halo} \geq 5 \times 10^{10} h^{-1} M_{\odot}$) and grow by gas accretion and mergers. The gas accretion rate follows the Bondi-Hoyle-Lyttleton model (Bondi & Hoyle, 1944; Bondi, 1952):

$$\dot{M}_{\rm Bondi} = \alpha \frac{4\pi G^2 M_{\rm BH}^2 \rho}{(c_s^2 + v_{\rm rel}^2)^{3/2}} \,. \tag{2.17}$$

Here, α is a boost factor accounting for the unresolved structure of the ISM, c_s is the local sound speed, and $v_{\rm rel}$ is the relative velocity between the BH and the surrounding gas.

The Eddington limit provides an upper cap on the accretion rate, preventing unphysical growth:

$$\dot{M}_{\rm Edd} = \frac{4\pi G M_{\rm BH} m_p}{\epsilon_r \sigma_T c}, \tag{2.18}$$

where ϵ_r is the radiative efficiency, m_p the proton mass, σ_T the Thomson cross-section, and c the speed of light.

Feedback from AGN activity is modeled through two modes:

1. *Quasar Mode*: At high accretion rates ($\dot{M}/\dot{M}_{\rm Edd} \geq 0.05$), thermal energy is injected into the surrounding gas to suppress cooling, resulting in strong, often bursty feedback episodes.

2. *Radio Mode*: At low accretion rates ($\dot{M}/\dot{M}_{\rm Edd} < 0.05$), the same thermal feedback mechanism is used, but with an energy injection efficiency that is four times higher. The lower accretion rates result in more gradual, continuous heating that stabilizes the cooling flow over time.

The energy injection from AGN feedback is given by:

$$\dot{E}_{\rm feed} = \epsilon_f \epsilon_r \dot{M}_{\rm BH} c^2, \tag{2.19}$$

with ϵ_f representing the fraction of radiated energy thermally coupled to the gas. This dual-mode feedback regulates both star formation and BH growth, reproducing observed correlations such as the $M_{\rm BH} - \sigma_*$ relation.

2.4.2 IllustrisTNG

The IllustrisTNG model (also called TNG) builds upon the groundwork established by the original Illustris simulations (Vogelsberger et al., 2014; Genel et al., 2014). Using the moving mesh code AREPO, IllustrisTNG integrates gravitational dynamics, Magnetohydrodynamics (MHD), and a suite of subgrid physics models to simulate the formation and evolution of galaxies in a ΛCDM Universe (Nelson et al., 2018; Pillepich et al., 2018).

Star Formation and ISM Treatment: Gas above a critical hydrogen number density threshold (typically $n_{\rm H} \approx 0.1$ cm^{-3}) is modeled as a two-phase medium following the formulation of (Springel & Hernquist, 2003). This treatment includes cold clouds embedded in a hot, pressurized ambient phase, maintaining an Effective Equation of State (EQS) for star-forming regions. Stars are formed stochastically in these dense environments, in proportion to the local gas density and the empirical Kennicutt-Schmidt law (Kennicutt, 1998). A Chabrier (2003) IMF is adopted throughout.

Stellar Feedback and Galactic Winds: The TNG model introduces significant refinements in the stellar feedback processes relative to its predecessor. Winds from star-forming regions are launched kinetically, with their velocity scaling as a function of the local dark matter velocity dispersion, $\sigma_{\rm DM}$, and the Hubble parameter $H(z)$ (Oppenheimer & Davé, 2006). Specifically, the wind velocity is given by

$$v_{\rm wind} = \max\left[\kappa_w \sigma_{\rm DM}\left(\frac{H_0}{H(z)}\right)^{1/3}, v_{\rm min}\right], \tag{2.20}$$

where κ_w is a scaling factor and $v_{\rm min}$ enforces a minimum velocity to prevent unphysically slow winds in low-mass halos (Pillepich et al., 2018).

The mass loading factor, η_w, which governs the mass outflow rate relative to the star formation rate, depends on the wind energy and velocity:

$$\eta_w = \frac{2}{v_{\rm wind}^2} e_{\rm wind}(1-\tau_w), \tag{2.21}$$

where $e_{\rm wind}$ represents the energy injected into the winds, and τ_w describes the thermal fraction of this energy. To better match observed galaxy scaling relations, the model incorporates a metallicity-dependent reduction in wind efficiency, accounting for enhanced radiative losses in metal-rich gas.

Chemical Enrichment and Stellar Yields: The chemical evolution model tracks the production and distribution of nine individual elements (H, He, C, N, O, Ne, Mg, Si, Fe), plus a tenth component representing untracked metals. Stellar feedback returns mass and metals to the ISM through three primary channels: AGB stars, SNe Ia, and SNe II. The TNG model updated the yields and progenitor mass ranges used in Illustris, raising the lower mass threshold for SNe II

events from 6 $M_\odot$ to 8 $M_\odot$, lying in closer agreement with later theoretical expectations (Smartt, 2009). The injection of metals follows time-dependent delay functions, with the delay time distribution for SNe Ia adopting the formalism of (Maoz et al., 2012).

Supermassive Black Holes: BHs are seeded in halos once they exceed a mass threshold of $M_{\rm halo} = 5 \times 10^{10}\ h^{-1}\ M_\odot$, with seed BH masses set at $8 \times 10^5\ h^{-1}\ M_\odot$. Gas accretion onto BHs follows the Bondi-Hoyle-Lyttleton prescription (Bondi & Hoyle, 1944; Hoyle & Lyttleton, 1939), without the artificial α boost factor used in previous models.

Two distinct feedback modes are employed:

- High accretion rates trigger a thermal quasar-mode feedback, where energy is isotropically injected into the surrounding gas.
- Low accretion rates activate a kinetic feedback mode, releasing momentum-driven winds that suppress star formation in massive galaxies and clusters (Weinberger et al., 2017).

The kinetic mode plays a main role in preventing excessive stellar mass buildup in massive halos and addresses shortcomings from the original Illustris model, including overly large galaxy sizes and high star formation rates at low redshift.

Magnetohydrodynamics: TNG introduces a uniform seed magnetic field at high redshift. Its amplification and evolution are self-consistently modeled (Pakmor & Springel, 2013). This allows the simulation to capture the impact of magnetic fields on galaxy formation and evolution, influencing angular momentum transport, star formation, and gas dynamics (Marinacci et al., 2015).

Calibration and Observational Comparisons: The TNG model was calibrated to match a range of observations, including the galaxy stellar mass function, the cosmic star formation history, and the baryon fractions in halos. While calibration focused on global trends, the model also makes successful predictions for galaxy morphologies, color distributions, and the distribution of metals in circumgalactic and intergalactic media (Nelson et al., 2018; Pillepich et al., 2018)

2.5 Identifying Bound Structures in Simulations

Once we have a cosmological simulation, the next step is to identify the gravitationally bound objects that are contained in it. As discussed earlier, the growth of cosmic structures is highly sensitive to the underlying cosmological parameters, including the energy content of the Universe and the properties of dark energy and dark matter. To connect these theoretical predictions with observed structures like galaxies and galaxy clusters, it's essential to apply robust methods to detect and characterize bound structures within simulation data.

For this purpose, we focus on two algorithms commonly used in cosmological simulations: the Friends-of-Friends (FoF) algorithm and the Subfind algorithm (Springel, 2005). These

techniques are the backbone of the halo identification pipeline used in our simulations. For a more comprehensive discussion of other methods, we refer the reader to reviews such as the one by Knebe et al. (2013).

2.5.1 Friends-of-Friends Algorithm

The FoF method is one of the simplest and most widely used algorithms for identifying groups of particles in simulations. It links particles into groups based on their spatial proximity using a parameter called the linking length, which sets the maximum distance two particles can be apart and still be considered part of the same group. Typically, this linking length is expressed as a fraction of the mean inter-particle separation, with common values around $b \sim 0.2$ for locating virialized halos.

The algorithm begins by selecting an initial particle and identifying all neighboring particles within this linking length. These neighbors, in turn, are connected to their own neighbors, and so on, until a complete group is formed. The choice of linking length directly influences the resulting group structure: a shorter linking length typically results in smaller, denser groups, while a larger one can connect structures that might otherwise be considered distinct.

FoF was first applied in observational cosmology by Geller & Huchra (1983) to identify galaxy groups in redshift surveys. In numerical simulations, the FoF implementation described by Springel (2005) has been widely adopted. This version employs a parallel, distributed approach that is optimized for large datasets and high-performance computing architectures. The algorithm efficiently partitions the simulation domain across computational nodes, performing neighbor searches within each partition and merging groups when necessary. Iterative steps continue until no further merges occur.

However, the FoF algorithm has known limitations. It may erroneously connect structures via tenuous particle bridges or fail to recognize substructures within larger halos (Springel et al., 2001). This motivates the use of more refined algorithms like Subfind, which will be presented next.

2.5.2 The Subfind Algorithm

Subfind complements FoF by performing a more detailed analysis aimed at detecting substructures within halos. Initially developed for dark matter-only simulations (Springel et al., 2001) and later extended to hydrodynamical simulations (Dolag et al., 2009), Subfind refines the structure identification process by locating self-bound subhalos within FoF groups.

The method begins by identifying a parent halo using the FoF algorithm. For each baryonic particle (e.g., gas or stars), Subfind assigns it to the nearest dark matter particle's FoF group. Subfind then searches for overdense regions within the parent halo by analyzing the density field. It identifies local maxima and saddle points in the density distribution by descending along density gradients, effectively isolating regions that could correspond to gravitationally bound substructures.

Once candidate subhalos are located, Subfind performs an unbinding procedure: it calculates the gravitational potential for each candidate and removes particles whose total energy is positive

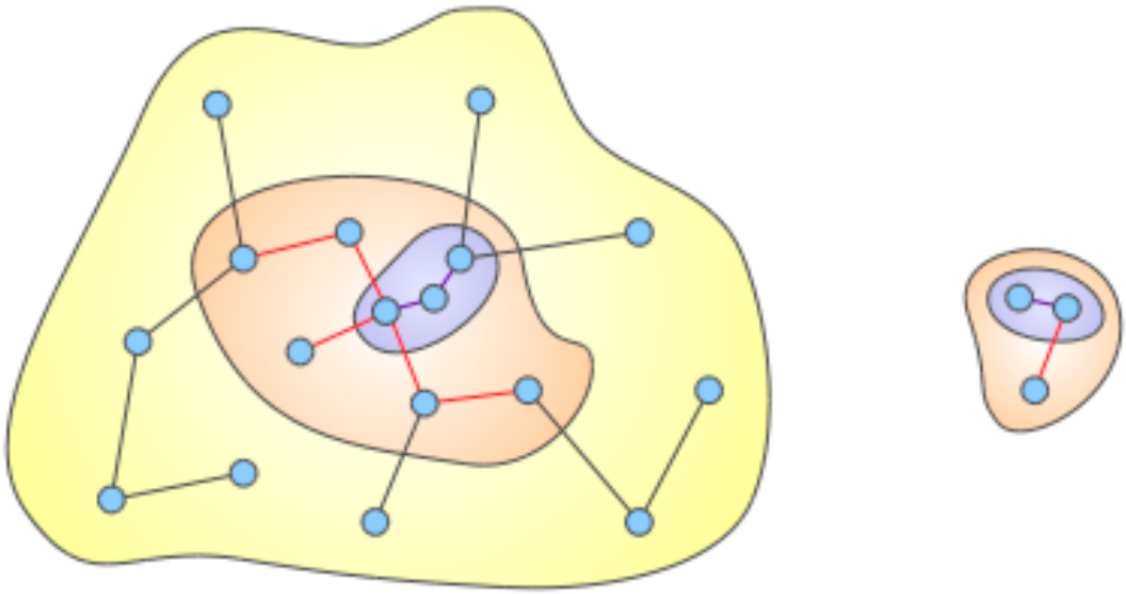


Figure 2.4: Graphical illustration of two FoF groups identified using different linking lengths. Shorter linking lengths result in smaller, more tightly bound groups.

(i.e., those not gravitationally bound). A substructure is confirmed if a minimum number of particles remain bound after this process, commonly set to 50 particles in simulations. This fixed threshold helps filter out noise and ensure a reliable identification of substructures, while remaining consistent across simulations of similar resolution.

For each genuine subhalo, Subfind identifies the center as the particle with the minimum gravitational potential. It then computes the subhalo's virial radius and mass by applying a spherical overdensity criterion, using thresholds based on the spherical collapse model (Eke et al., 1996).

An important consideration is that Subfind may assign subhalos to the FoF group of a parent halo even when they are located beyond its virial radius. Similarly, not every structure within a FoF group necessarily lies within the main halo's virial radius. Fig 2.4 illustrates examples of halo and subhalo identification in simulations, showing how different linking lengths in the FoF algorithm yield different groupings of particles.

2.6 Following a Halo Through Time: The Merger Tree

Understanding the formation and evolution of structures in the Universe requires a method to trace their assembly over cosmic time. In cosmological simulations, this is achieved by constructing merger trees: data structures that record the lineage of dark matter halos (and the galaxies they host), capturing the hierarchical growth of structure. Merger trees enable the reconstruction

of a halo's ancestry by linking its progenitors at earlier times to its descendants at later stages (Springel, 2005).

At their core, merger trees identify relationships between gravitationally bound objects, typically dark matter halos and subhalos, across different snapshots of a simulation. Each branch of the tree describes the accretion history of a halo, while the nodes represent merger events, smooth accretion, or other evolutionary processes. This structure captures not only the main progenitor, often referred to as the "main branch," but also tracks secondary halos that merge into the larger system, enabling a detailed analysis of merger histories, halo growth rates, and galaxy evolution (Boylan-Kolchin et al., 2009; Henriques et al., 2015).

2.6.1 The L-HaloTree Algorithm

In this work, we employ the L-HaloTree algorithm (Springel, 2005; Boylan-Kolchin et al., 2009), a widely used merger tree construction tool developed for the Millennium and Millennium-II simulations. L-HaloTree operates on halo catalogs identified by the Subfind algorithm (Springel et al., 2001; Dolag et al., 2009), following the evolution of these subhalos across time by establishing unique links between progenitors and descendants.

It does this by evaluating the particle content of halos in consecutive simulation snapshots. Specifically, each subhalo is represented by a set of bound particles, which are first *ranked* according to their binding energy. To trace subhalos forward in time, the algorithm assigns a descendant to each progenitor based on the *weighted overlap* of particles between snapshots. In this weighting scheme, the most bound particles, typically the top 10%, are given the highest importance, ensuring that the core of a subhalo is preferentially tracked, even if its outer layers are stripped during mergers (Springel, 2005).

A key feature of L-HaloTree is its robustness against the temporary disruption in the identification of halos. In some instances, a subhalo may momentarily disappear from a snapshot due to limitations in the halo finder or strong tidal interactions. L-HaloTree accounts for such situations by searching over multiple snapshots to reconnect subhalo trajectories and avoid losing continuity in the tree structure (Boylan-Kolchin et al., 2009).

2.6.2 Merger Tree Structure and Outputs

Each node in an L-HaloTree represents a subhalo, characterized by its mass, position, velocity, and other properties provided by the simulation. The tree structure captures several essential relationships:

1. *Descendant links*, connecting a subhalo to its descendant in the following snapshot.

2. *Progenitor links*, pointing back to the various subhalos that merge to form the descendant.

3. *Main progenitor branches*, typically defined as the branch contributing the most mass to the descendant halo.

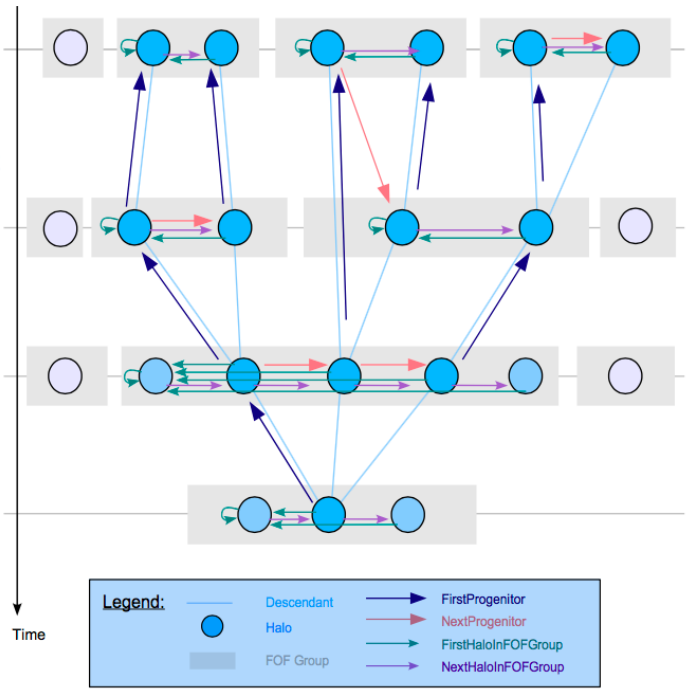


Figure 2.5: Hierarchical structure of merger trees in the L-HaloTree algorithm. Each blue circle represents a halo, and halos within the same gray box belong to a common FoF group at a given snapshot (time increases downward). Arrows illustrate the tree structure: Descendant links (light blue) connect halos to their unique descendant in the next snapshot; FirstProgenitor (dark blue) is the main (typically most massive) progenitor of a halo, while NextProgenitor (red) refers to additional progenitors that merge into the same descendant. Within FoF groups, FirstHaloInFOFGroup (green) indicates the central halo, and NextHaloInFOFGroup (purple) connects to other group members. This structure enables efficient traversal of halo lineages and group associations across time. Credit: Springel et al. (2005b)

4. *Sibling and cousin relationships*, which describe halos that are connected through shared descendants or progenitors. For example, two progenitors of the same descendant halo are considered siblings, while cousins may share a common ancestor further up the tree, providing a broader view of the merger hierarchy.

These links enable an efficient reconstruction of halo assembly histories and the identification of major and minor merger events, smooth accretion, and disruptions.

Furthermore, this hierarchical data structure eases rapid queries for progenitor and descendants of a certain halo across time. A scheme of this hierarchical structure is shown in Fig. 2.5. The flexibility of the format allows its use in a variety of downstream applications, such as SAMs of galaxy formation (e.g. Henriques et al., 2015) or the analysis of merger rates and halo mass assembly histories (e.g. Rodriguez-Gomez et al., 2015).

2.7 Constrained Cosmological Simulations

Standard cosmological simulations reproduce the LSS of the Universe and the diversity of its constituents in a statistical sense. However, they do not aim to replicate the actual observed structures of the nearby Universe, such as the Virgo cluster, the Local Group, the Shapley supercluster or the local Void, just to mention some.

Constrained cosmological simulations address this gap. Their aim is to reproduce the spatial distribution, dynamical properties, and formation histories of specific structures in the observed Universe. They provide a laboratory for studying the interplay between local cosmic structures and cosmological parameters, mitigating cosmic variance by focusing on a specific realization of the Universe, our own. This makes constrained simulations a powerful tool for astrophysical and cosmological research.

2.7.1 Fundamental Concepts and Methodologies

The main goal of constrained cosmological simulations is to produce Initial Conditions (ICs) that, when evolved forward in time, reproduce the structures observed in the local Universe. These simulations differ from conventional simulations in that they do not start from a random realization of the cosmic density field but from one shaped by observational data. Two primary methods are used to generate such ICs:

- *Backward Constrained Realizations* (Hoffman-Ribak Algorithm, Hoffman & Ribak (1991)): This approach uses linear theory and observational data, often peculiar velocities or galaxy distributions, to reconstruct the linear density and velocity fields. The algorithm allows for the generation of Gaussian random fields conditioned on these observations. Early applications of this technique used galaxy redshift surveys to constrain the density field (Mathis et al., 2002; Dolag et al., 2005), leading to simulations such as CORUSCANT (Dolag et al., 2005) and SALACIOUS (Nuza et al., 2010). Later applications using the local peculiar velocity field have led to simulations such as CLUES (Constrained Local UniversE Simulations, Sorce et al., 2016a) and SLOW (Simulating the LOcal Web, Dolag et al., 2023; Hernández-Martínez et al., 2024; Böss et al., 2023; Seidel et al., 2024) .

- *Bayesian Forward Modeling*: This approach reconstructs the ICs of the Universe by statistically inferring the initial density field that, when evolved forward through a simulation, best reproduces the observed LSS. Instead of using a fixed set of random simulations, the method explores the space of possible ICs, often by sampling from a prior on the initial Gaussian density field, and uses Bayesian inference to identify the most likely realization, given the data (see Jasche & Wandelt, 2013; Jasche & Lavaux, 2019, and references therein). This approach has inspired works such as ELUCID (Wang et al., 2016) and SIBELIUS (McAlpine et al., 2022).

The quality of the constrained simulations depends critically on the type, quantity, and quality of the observational constraints. Here, quality refers to how well the simulation reproduces

observed structures in the local Universe, such as the correct positions, masses, and dynamics of galaxy clusters or the cosmic web, given the input data. Two main types of observational data are typically used:

- *Galaxy Redshift Surveys*: Provide direct information on the spatial distribution of galaxies and, by inference, the underlying density field. However, they suffer from the fact that in order to infer the underlying density field, it is necessary to assume a bias relation between luminous and dark matter (see Desjacques et al., 2018, for a comprehensive review). Additionally, peculiar velocities along the line of sight distort redshift-based distance estimates.

- *Peculiar Velocity Surveys*: Peculiar velocities are directly related to the gravitational potential, thus tracing the underlying matter distribution (see Sect. 1.1.1 for a theoretcal explanation), without the need to assume any relation between luminous and dark matter. This feature makes them very advantageous constraints. However, they are observationally challenging, with larger measurement uncertainties and sparser datasets. Notable efforts to use peculiar velocities as constraints include the Cosmicflows program (Tully et al., 2013, 2016; Tully et al., 2023) and their implementation in the CLUES (Sorce et al., 2016a) and SLOW (Dolag et al., 2023; Hernández-Martínez et al., 2024; Böss et al., 2023; Seidel et al., 2024) simulations.

2.7.2 Historical Development

Early efforts in constrained cosmological simulations began with low-resolution dark matter-only simulations of the local environment. One of the first works was the one presented by Mathis et al. (2002), who used redshift surveys to generate constrained ICs. This simulation was later extended to include magneto-hydrodynamics (Dolag et al., 2005).

Building on these foundations, the CLUES project marked a major advancement in the field. Initiated by Yepes et al. (2009); Gottlöber et al. (2010), CLUES pioneered the use of peculiar velocity constraints rather than relying solely on galaxy redshift distributions. This approach enabled the reproduction of prominent LSS in the local Universe, including the Virgo, Perseus, and Coma clusters, within a $(160h^{-1}\mathrm{Mpc})^3$ volume. CLUES produced both dark matter-only and non-radiative hydrodynamical simulations, which provided valuable insights into the formation and evolution of the Local Group (Libeskind et al., 2010; Nuza et al., 2014; Carlesi et al., 2016).

Focusing on higher-resolution zoom-in regions around the Local Group, the HESTIA simulations (Libeskind et al., 2020) incorporated full galaxy formation physics, allowing for detailed studies of the impact of the cosmic environment on galaxy evolution within the local volume.

Subsequent developments led to CLONES (Constrained LOcal and Nesting Environment Simulations, Sorce et al. (2021)). This project expanded the constrained simulation volume to $(500h^{-1}\mathrm{Mpc})^3$ and employed improved techniques for generating constraints (Sorce, 2015; Carlesi et al., 2016b). CLONES was particularly used to match the assembly histories of individual galaxy clusters, offering a way to investigate their mass assembly history and general evolutionary paths of individual replicas of the local Universe.

More recently, the SIBELIUS project (McAlpine et al., 2022) introduced a different approach based on Bayesian forward modeling. SIBELIUS-DARK succeeded in constraining a massive volume of approximately $(550h^{-1}\mathrm{Mpc})^3$, reproducing major features of the local LSS, such as the Virgo cluster and the Local Void. Nevertheless, the number of reported identified counterparts is limited to 12, and the simulation still faces challenges in accurately reproducing properties such as the masses of individual structures.

In parallel, the SLOW (Simulating the LOcal Web) project (Dolag et al., 2023; Hernández-Martínez et al., 2024; Böss et al., 2023; Seidel et al., 2024) represents the latest generation of constrained cosmological simulations. Based on the CLONES realization and constrained with peculiar velocity data from Cosmicflows-2 (CF2, Tully et al. (2013)), SLOW covers a $(500h^{-1}\mathrm{Mpc})^3$ volume and includes full magneto-hydrodynamical simulations with galaxy formation physics. SLOW achieves unprecedented accuracy in replicating the properties of local galaxy clusters, such as Virgo, Coma, and Perseus, not only in terms of their masses and positions but also their multi-wavelength observational properties, including X-ray luminosities, temperatures, and SZ signals (Hernández-Martínez et al., 2024), marking a significant step forward in bridging the gap between simulations and observations of the local Universe. More details on this set of simulations will we covered in the next section.

2.7.3 The SLOW Project

Developed by Dolag et al. (2023) and expanded by Hernández-Martínez et al. (2024), SLOW builds upon the previous CLONES project, using peculiar velocity data from the CF2 catalogue to construct its ICs. The simulations in the SLOW project cover a comoving volume of $(500h^{-1}\mathrm{Mpc})$ each and include boxes with full hydrodynamical modeling and subgrid physics based on the Magenticum model presented in Sect. 2.4.1. It has been shown to contain more than 46 replicas of local galaxy cluster allowing for direct comparisons between simulations and multi-wavelength observations of these clusters residing in our local environment.

Initial Conditions and Methodology: The two main pillars on which constrained ICs are constructed are the prior cosmological model, like any random simulations, and the observational data. For the former, we use *Planck* cosmology (Ω_m=0.307 ; Ω_Λ=0.693 ; H_0=67.77 km s^{-1} Mpc^{-1}, and σ_8=0.829 Planck Collaboration et al., 2014b). For the latter, in our case, the observational data consist of CF2, the second catalog of the Cosmicflows project (Tully et al., 2013). This catalog contains more than 8000 distance moduli of local galaxies obtained from various distance indicators such as the Tully-Fisher relation, the fundamental plane, and SN. These distance moduli are combined with observational redshifts to determine galaxy radial peculiar velocities. The method to build the constrained ICs has been extensively discussed by, for instance, Sorce et al. (2016a). The basis for SLOW's ICs have been developed and published by Sorce (2018). We present here a summary of the main steps:

1. The first step involves removing any nonlinear motions that would affect our linear reconstruction of the linear initial fields. We are careful to keep galaxies still infalling onto

clusters outside of the latter. Galaxies are thus grouped, resulting in 5562 galaxies and groups (Sorce & Tempel, 2018).

2. We then minimize the biases inherent to any observational radial peculiar velocity catalogs. The effects of the homogeneous and inhomogeneous Malmquist biases as well as that of the lognormal error (e.g. Kaptney, 1914; Malmquist, 1920; Landy & Szalay, 1992; Tully et al., 2016) are minimized in the catalog with the method described by Sorce (2015).

3. The Wiener-Filter method (WF, Zaroubi et al., 1999) is then applied to the grouped catalog of bias-minimized radial peculiar velocity constraints to reconstruct the tridimensional cosmic displacement field.

4. The reconstructed cosmic displacement field is then used through the Reverse Zel'dovich Approximation (Doumler et al., 2013) to relocate the galaxy and group constraints to the positions of their progenitors. It ensures that structures are at the proper position at redshift zero. We also replace noisy radial peculiar velocities by their 3D WF reconstructions (Sorce et al., 2014)

5. We produce the initial density and velocity fields constrained with the grouped, bias-minimized, and relocated 3D peculiar velocities using the Constrained Realizations (CR) technique (CR, Hoffman & Ribak, 1991, 1992; van de Weygaert & Bertschinger, 1996). This method derives an estimate of the residual between the WF reconstruction and the true field using a random realization. It allows the regions where only scarce or low-quality data is available to be filled with a random Gaussian field, while converging to the constrained values where these are available. Doing so, we restore statistically the otherwise missing structures and ensure the overall compliance with the prior power spectrum. The "strength" of the observational data used as constraints determines the extent to which the ICs are likely to reproduce the observed local Universe. In those regions where the data is lacking or dominated by error, the recovered velocity field will tend towards the random realization of the prior model. The evolution of the simulation will thus be determined by the interplay between the random modes and the constraints.

6. As scales below the linear threshold (3 – 4 Mpc) are non-constrained, at least not directly[1], there is no reason to work with density and velocity fields with resolution higher than that of the galaxy groups and clusters. The resolution of the constrained density and velocity fields can then be increased by adding small-scale features using the GINNUNGAGAP software[2].

The final ICs correspond to realization number 8 of the CLONES simulations, with additional small-scale perturbations to achieve higher resolutions in different runs of SLOW.

[1]It can though, be shown that some small scales are induced by the large-scale environment. Carlesi et al. (2016a) showed that it is the case of the local Group.

[2]https://github.com/ginnungagapgroup/ginnungagap

Simulation Suite and Physics Models: The SLOW project includes a suite of simulations with varying levels of physical complexity and resolution. These range from dark matter-only runs at different resolution levels (768^3 to 6144^3 particles), to full magneto-hydrodynamical simulations incorporating cosmic rays, and hydrodynamical runs with galaxy formation physics. The latter follow prescriptions similar to those used in the Magneticum simulations (see Sect. 2.4.1), including radiative cooling, star formation, stellar feedback, chemical enrichment, and BH growth with AGN feedback (Hirschmann et al., 2014b; Dolag et al., 2016; Dolag et al., 2025, and references therein).

Reproducing the Local Universe: The SLOW project introduces several significant improvements over previous generations of constrained cosmological simulations. One of the most notable advancements lies in its use of a large simulation volume, $(500h^{-1}\mathrm{Mpc})^3$, combined with high-resolution hydrodynamical modeling that includes full galaxy formation physics. Unlike earlier efforts, such as CLUES and HESTIA, which were typically limited to smaller volumes or focused zoom-in regions, SLOW simultaneously captures the large-scale cosmic web and the detailed baryonic processes that govern galaxy and cluster formation. At current time SLOW has demonstrated high fidelity in reproducing multi-wavelength observables, including X-ray luminosities, temperatures, and SZ signals for the largest set of local Universe clusters ever replicated in a cosmological volume (Hernández-Martínez et al., 2024).

Table 2.1: Set of available SLOW simulations following different physical processes and utilizing different resolutions. All are based on the same initial constrained density and velocity field. Columns indicate: simulation name; final redshift z to which the simulation was evolved; total number of particles N; dark matter particle mass $M_{\rm DM}$; gas particle mass $M_{\rm gas}$ (if applicable); and the set of physical processes included (e.g., hydrodynamics, magnetic fields, cosmic rays, galaxy formation).

Name SLOW –	z	N	M_{DM} $[M_\odot h^{-1}]$	M_{gas} $[M_\odot h^{-1}]$	Physics
DM768^3	0	768^3	2.4×10^{10}	–	DM only
DM1576^3	0	1576^3	2.9×10^9	–	DM only
DM3072^3	0	3072^3	3.7×10^8	–	DM only
DM6144^3	–	6144^3	4.6×10^7	–	ICs only
AGN768^3	0	2×768^3	2.0×10^{10}	3.7×10^9	Cooling, SF, AGN
AGN1536^3	0	2×1536^3	2.5×10^9	4.6×10^8	Cooling, SF, AGN
AGN3072^3	2	2×3072^3	3.1×10^8	5.5×10^7	Cooling, SF, AGN
CR3072^3	0	2×3072^3	3.1×10^8	5.5×10^7	MHD, Cosmic Rays

3 | Machine Learning

Machine Learning (ML) has become a cornerstone of modern data analysis across a wide range of scientific fields, from biology and medicine to finance and astrophysics. Its strength lies in its ability to learn from data, uncovering patterns, structures, or mappings that would be difficult to identify through explicit modeling. In contrast to rule-based approaches, ML algorithms are trained to recognize relationships between inputs and outputs through exposure to example data, without the need for manual feature design or hardcoded rules (Bishop, 2006; Goodfellow et al., 2016).

The purpose of this section is to equip the reader with the essential ML background needed to follow the methods developed in later chapters, particularly Chapter 6. While ML is a vast and rapidly evolving field, we focus here on a selection of foundational concepts that are most relevant to our applications. Section 3.1 begins with an overview of the use of ML in the field of astrophysics and cosmology. Section 3.2 explores the differences between classical ML techniques and Deep Learning (DL), including the structure and function of Neural Networks (NNs), the training, and the performance metrics with which results are typically evaluated. In Section 3.3, we explore the application of ML for cosmological parameter inference, particularly exploring the CAMELS project. For readers seeking a more comprehensive treatment of the subject, we recommend the foundational texts by Bishop (2006); Goodfellow et al. (2016); VanderPlas (2016); Stevens et al. (2021); Alzubaidi et al. (2021).

3.1 The Need for Machine Learning in Cosmology

Modern cosmology stands at the intersection of theory, simulation, and observation, relying on increasingly large and complex datasets to test fundamental models of the Universe. While traditional statistical methods have long played a central role in the analysis and interpretation of cosmological data, they face significant limitations in the current era. These methods often rely on assuming a specific model, a choice that may be unjustified a priori, while typically also depending on low-dimensional summary statistics, and struggling with scalability when applied to large simulations or observational datasets. As cosmological models grow in complexity, and the data used to constrain them becomes higher dimensional and noisier, new approaches are needed to extract information efficiently and robustly.

This need is only becoming more urgent with the advent of next-generation surveys such as Euclid Laureijs et al. (2011), SKA (Braun et al., 2019), and LSST (LSST Science Collaboration

et al., 2009), which will deliver unprecedented volumes of high-dimensional data. These instruments are poised to transform cosmology into a data-intensive science, further pushing traditional methods to their limits and requiring scalable, flexible analysis techniques.

ML offers a flexible and scalable solution to many of these challenges. Unlike traditional techniques that rely on pre-defined models and assumptions, ML algorithms are capable of learning from raw or minimally processed data, capturing complex non-linear relationships without requiring explicit likelihoods. This makes ML particularly well suited for high-dimensional problems and for simulation-based workflows where no tractable likelihood exists.

Over the past decade, ML has already demonstrated its utility in a wide range of astrophysical applications. Convolutional Neural Networks (CNNs) have been used to classify galaxies by morphology, distinguish between merger types, and analyzing photometric images (Domínguez Sánchez & et al., 2018). Other ML-based models have shown promising results with weak gravitational lensing data, for example, by learning to extract lensing signals from noisy maps (Schmelzle et al., 2017), or by estimating cosmological parameters directly from simulations of LSS (Ravanbakhsh et al., 2017).

A particularly important role for ML in cosmology is in the context of *likelihood-free inference*, also known as Simulation-Based Inference (SBI). In many physical systems, the forward model (i.e., simulation) is known, but the likelihood of the data given model parameters is not. ML-based SBI methods address this by learning a mapping between simulation outputs and the parameters used to generate them, typically using NNs trained on large ensembles of simulated data (Cranmer et al., 2020; Alsing et al., 2019). This allows researchers to perform parameter estimations even in cases where the full likelihood function is unknown or intractable.

In parallel, ML is increasingly used for building *emulators*, surrogate models trained to approximate the outputs of high-fidelity simulations. These models can predict simulation results across parameter space orders of magnitude faster than running the simulations themselves (Heitmann et al., 2009; Agarwal et al., 2021). When combined with SBI techniques, emulators enable both rapid forward modeling and efficient inverse inference.

Together, these capabilities make ML a powerful and increasingly indispensable component of modern cosmological analysis. Its ability to operate on high-dimensional data, bypass likelihood assumptions, and accelerate computation opens new possibilities for interpreting observations and exploring theoretical models. As observational and simulated datasets continue to grow in size and complexity, the need for ML in cosmology will only become more pressing.

3.2 Fundamentals of Machine Learning

At the core of ML is the concept of learning a mapping: given a dataset composed of input vectors $\vec{x}$ and corresponding target values or labels $\vec{y}$, a ML algorithm is trained to approximate a function that predicts the outputs from the inputs. This process is typically referred to as the *training phase*. The function approximator, often referred to as a *model* in the ML context (e.g., a NN or decision tree), is distinct from the term "model" as used in physics or statistics, where it usually denotes a set of assumptions about the data-generating process. In contrast, ML models are data-driven and learn patterns directly from examples, without requiring an explicit physical

or probabilistic model of the system (Hastie et al., 2009).

ML is typically divided into two main paradigms: *supervised* and *unsupervised* learning. In supervised learning, each input is associated with a known output or label, allowing the algorithm to learn a direct mapping from data to target. For example, a model might be trained to identify cats in images after being shown many labelled examples. In contrast, unsupervised learning deals with unlabelled data. The model is tasked with discovering latent structure, such as clustering or statistical distribution, without predefined targets (Murphy, 2012). For instance, an unsupervised model might learn to separate images of cats and dogs purely based on patterns in the data, even if the categories are not named.

This work primarily focuses on supervised learning, which offers a robust framework for classification and regression tasks, especially when a sufficiently large and representative labelled dataset is available. Supervised models are particularly effective for processing the kind of structured outputs we obtain from simulations, such as halo properties or spatial maps (Ntampaka et al., 2019).

Training an ML model involves optimizing its internal parameters $\vec{\theta}$, which control how inputs are transformed into outputs. It is important to distinguish between *parameters*, which are learned during training, and *hyperparameters*, which are set by the user and define aspects of the model architecture (e.g., number of layers or learning rate). The performance of the final model depends critically on both (Goodfellow et al., 2016).

To evaluate model performance reliably, data is usually divided into three disjoint subsets: the *training set*, used to fit the model; the *validation set*, used to tune hyperparameters and prevent overfitting; and the *testing set*, which provides an unbiased estimate of generalization to new data (Kuhn & Johnson, 2013). Overfitting occurs when a model becomes too tailored to the training data, capturing noise or fluctuations rather than general patterns. In contrast, *underfitting* happens when the model is too simple to represent the underlying structure, resulting in poor performance even on training data (Hastie et al., 2009; Kuhn & Johnson, 2013; Géron, 2019).

Another fundamental aspect is *generalization*, i.e. the ability of the model to make accurate predictions on new, unseen inputs. This is essential in scenarios where models are deployed beyond the training data, such as in interpreting new observations or applying models to simulations generated under different conditions. Poor generalization is often a consequence of training data that is insufficient in quantity, lacks diversity, or is poorly processed and unrepresentative of the target distribution (Hastie et al., 2009; Kuhn & Johnson, 2013; Domingos, 2012; Ntampaka et al., 2019).

To ensure robust learning, input data often requires preprocessing. This may involve normalization, handling missing values, encoding categorical variables, or transforming features into more suitable representations (Géron, 2019). While these steps may seem routine, they can significantly influence training time and final performance by improving numerical stability, ensuring consistent feature scaling, and allowing models to detect relevant patterns more easily. Poor preprocessing can lead to longer convergence times, biased learning, or degraded predictive accuracy.

Although the field of ML encompasses a wide range of models and techniques, many foundational ideas are shared across architectures. In the next sections, we build on this foundation to explore more specific algorithms and their role in our astrophysical applications.

3.2.1 Deep Learning

While all ML algorithms share the objective of learning patterns from data, DL represents a specialized and particularly powerful subset of the field. DL models are constructed using Artificial Neural Network (ANN), which draw inspiration from the way biological neurons function. These models are especially effective for analyzing unstructured or high-dimensional data, such as images, audio, or spatial fields (Goodfellow et al., 2016; Géron, 2019).

The basic building block of a DL model is the *artificial neuron*, an abstraction loosely based on the *biological neuron*. In biological systems, neurons integrate electrical signals from neighboring cells through dendrites. If the cumulative signal exceeds a certain *threshold*, the neuron *activates* and transmits a signal down its axon. In artificial models, each neuron similarly takes multiple *weighted inputs*, applies a *bias*, and passes the result through an *activation function* to produce an output. The *weights* can be thought of as a kind of transformation matrix that scales and combines the input features, while the *bias* shifts the resulting value, much like adjusting the baseline or offset before applying the activation. This structure is illustrated in Fig. 3.1. The activation introduces *nonlinearity*, which is critical for learning complex mappings between inputs and outputs.

Deep Neural Network (NN) are formed by stacking multiple layers of neurons, typically including an input layer, one or more *hidden layers*, and an output layer, where *depth* refers specifically to the number of layers in the network (Goodfellow et al., 2016). A network with many neurons in a single layer is considered *wide* but not deep, whereas a deep network has multiple stacked layers, each potentially with several neurons.

Each layer in the network is designed to process and transform the representation of the data. As shown in Fig. 3.1, the input layer receives the (possibly preprocessed) input data, hidden layers extract increasingly abstract features, and the output layer returns the final prediction. This architecture enables NN to learn hierarchical feature representations that capture subtle structures in the data (LeCun et al., 2015).

Compared to traditional ML models, DL has several distinguishing features:

- **Automatic feature extraction:** Deep networks can learn useful representations from raw inputs without manual engineering.
- **Model depth and expressiveness:** With many layers, DL models can approximate highly complex functions.
- **Low human supervision:** DL architectures reduce the need for hand-crafted features or preprocessing.
- **High data and compute requirements:** Training requires large datasets and computational resources to prevent overfitting and ensure convergence.

In conclusion, DL provides a powerful framework for modeling complex data distributions. Its capacity to automatically learn layered abstractions makes it particularly effective for tasks where traditional ML models struggle. These properties are used throughout this thesis to analyze astrophysical simulations and perform parameter inference in high-dimensional parameter spaces.

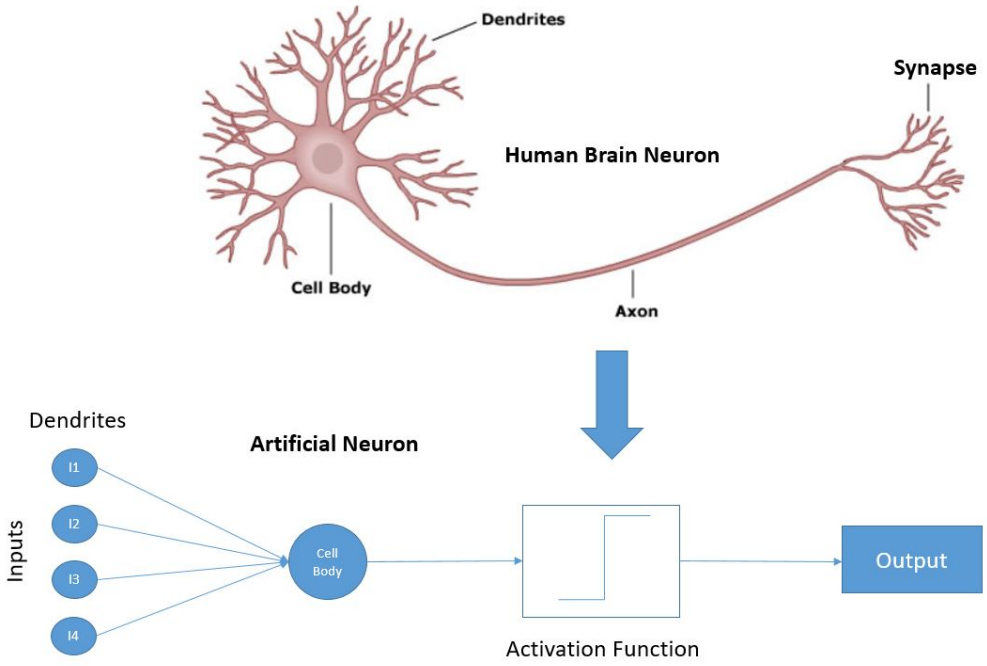


Figure 3.1: Biological neural cell (top) and artificial neuron (bottom) schemes. Artificial neurons can be modelled in terms of a simplified version of the functioning principles of biological neurons. Both neurons receive some information to process. This information is first processed by the cell body/artificial neuron. Subsequently, the information is subject to a non-linear activation layer (axon/activation function) in order to provide an answer. Credits: http://www.mplsvpn.info/2017/11/what-is-neuron-and-artificial-neuron-in.html

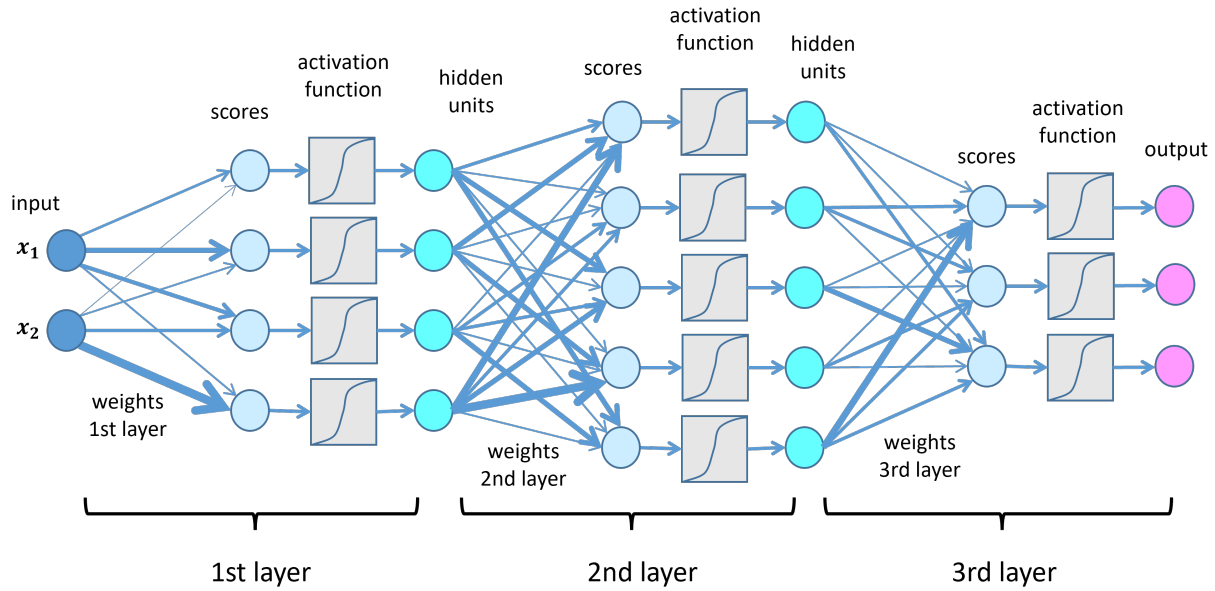


Figure 3.2: Example of a dense NN composed of an input layer, three hidden layers (labelled as 1st, 2nd, and 3rd layers), and an output layer. Each neuron in a given layer is connected to every neuron in the subsequent layer, forming a dense structure. The input features x_1 and x_2 are passed through weighted connections and transformed into intermediate values (scores), which are then processed by activation functions (e.g., sigmoid in this case) to generate the outputs of the hidden units. This process is repeated across layers, with each layer applying new weights and activation functions. The final output layer provides the prediction values. The diagram highlights the role of weights, scores, and activation functions at each step of the computation. Credits: https://lamarr-institute.org/blog/deep-neural-networks/

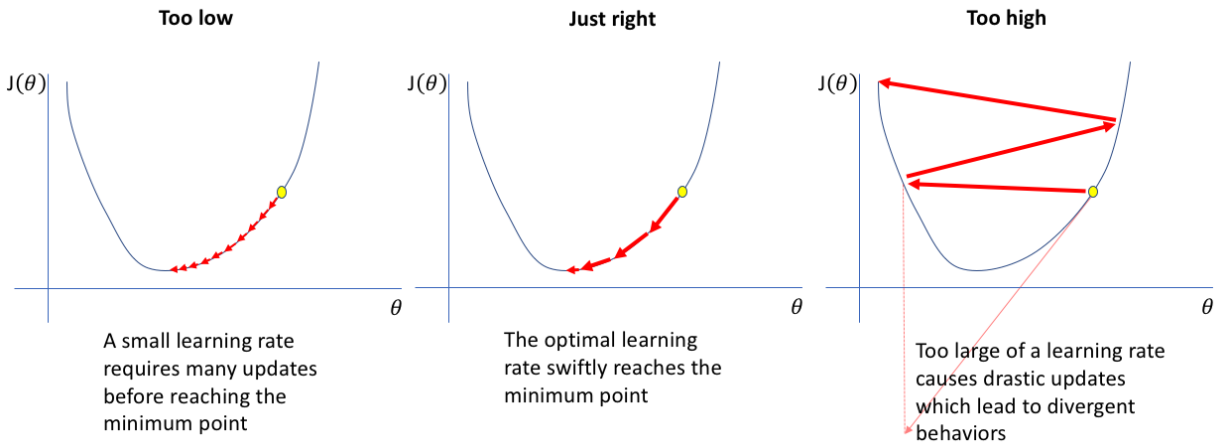


Figure 3.3: Impact of learning rate on gradient descent convergence. A learning rate that is too low results in slow progress toward the minimum (left panel), while an optimal rate achieves fast and stable convergence (central panel). In contrast, a learning rate that is too high causes overshooting and potential divergence from the minimum (right panel). Credits: https://www.jeremyjordan.me/nn-learning-rate/

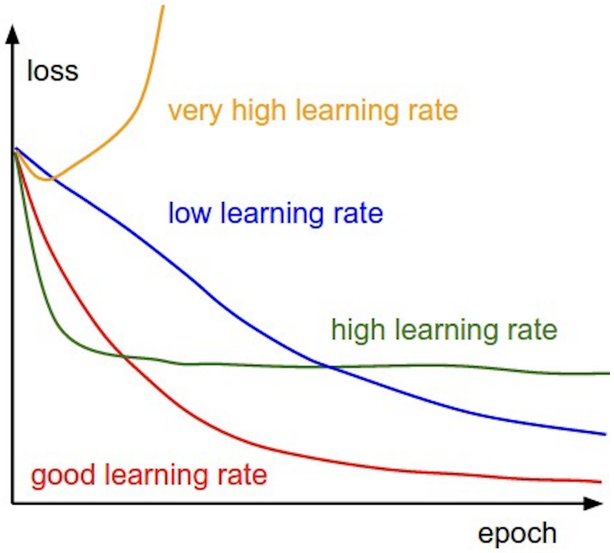


Figure 3.4: Behavior of the loss function over training epochs for different learning rates. A low learning rate results in slow convergence, while a very high learning rate causes the loss to diverge. A high learning rate may converge initially but can oscillate or stall. A well-chosen learning rate ensures fast and stable convergence of the loss. Credits: https://discuss.datasciencedojo.com/t/what-is-a-model-learning-rate-is-a-high-learning-rate-always-good/710

Dense Neural Networks

A fundamental architecture within the family of DL models is the *dense neural network*, also known as a Fully Connected Network (FCN). This type of network is defined by the full connectivity between layers: each neuron in one layer is connected to every neuron in the next. Such models are widely used due to their general-purpose nature and ability to approximate complex, nonlinear functions.

Dense networks consist of an *input layer*, one or more *hidden layers*, and an *output layer* (see Fig. 3.2 for a schematic representation). Each neuron computes a weighted sum of its inputs, adds a bias term, and applies a non-linear activation function (such as ReLU, sigmoid, or tanh). These transformations are composed layer by layer, enabling the model to learn rich hierarchical representations of the input.

Formally, the transformation applied by a dense layer can be written as:

$$\vec{h} = f(W\vec{x} + \vec{b}), \tag{3.1}$$

where $\vec{x}$ is the input vector, W is the weight matrix, $\vec{b}$ is the bias vector, and f is a nonlinear activation function applied element-wise. In a multilayer network, the output $\vec{h}$ of one layer serves as the input to the next.

One of the appealing theoretical properties of dense networks is their *universal approximation capability*: a feedforward network with a single hidden layer and a sufficient number of neurons can approximate any continuous function on a compact domain to arbitrary accuracy (Hornik et al., 1989). In practice, deeper architectures are often preferred as they allow for a more efficient representation and hierarchical decomposition of complex functions.

Despite their expressive power, dense networks can be computationally demanding. The number of trainable parameters increases rapidly with the number of neurons and layers, which can lead to overfitting if the training dataset is limited or if no regularization techniques are employed.

Emulators

In many areas of computational science, particularly in cosmology, physical models are implemented via large-scale numerical simulations. These simulations are essential for capturing the complex, nonlinear dynamics governing structure formation, galaxy evolution, and the behavior of the intergalactic medium. However, their high computational cost makes it impractical to run them repeatedly across wide ranges of parameter space. To overcome this limitation, researchers have developed *emulators*: fast surrogate models that approximate the output of simulations while requiring orders of magnitude less computational time.

An emulator is a model trained to learn the mapping between a set of input parameters (e.g., cosmological parameters, feedback strengths) and associated outputs (e.g., density fields, halo profiles, or power spectra). Once trained, the emulator can generate predictions nearly instantaneously. This enables dense sampling of the parameter space, uncertainty quantification, and a seamless integration into larger inference workflows (Heitmann et al., 2009; Cranmer et al., 2020).

The general process of emulator construction involves three main steps:

1. **Sampling parameter space:** Input parameters are selected using techniques such as Latin hypercube sampling or quasi-random sequences to ensure good coverage.
2. **Running simulations:** High-resolution simulations are performed at the sampled parameter points to provide the training data.
3. **Training the surrogate model:** A regression model, commonly a Gaussian Process (GP) or a NN, is trained to interpolate between simulation outputs based on the input parameters.

GPs are non-parametric, kernel-based models that define a distribution over functions and provide both mean predictions and associated uncertainty estimates (Rasmussen & Williams, 2006). Their probabilistic nature makes them particularly well suited to scientific applications requiring rigorous error propagation. However, GPs scale cubically with the number of training samples, which can become a bottleneck for large datasets.

To address higher-dimensional and data-intensive problems, more recent developments use DL to construct emulators. NNs, including fully connected neural networks, CNNs, and autoencoders, have been applied successfully to emulate outputs from cosmological simulations (Ravanbakhsh et al., 2017; Villaescusa-Navarro et al., 2021). These methods scale better with dataset size and can generalize across complex parameter spaces, although they may require more training data and computational resources upfront.

In this thesis, we employ a ML-based emulator named CARPoolGP (Lee et al., 2023) to generate synthetic realizations of galaxy cluster profiles. CARPoolGP is a sampling and regression technique that combines the CARPool method (Chartier et al., 2021) with GP regression to reduce the variance in emulated quantities. The key behind CARPoolGP is that, by introducing correlations between simulations at different points in parameter space, one can significantly enhance statistical efficiency. Specifically, CARPoolGP uses pairs of simulations based on the same ICs but evolved in time with different cosmological and astrophysical parameters, to create correlations and reduce the variance in the predicted/emulated quantities. Furthermore, the method supports active learning by identifying the most informative locations in parameter space for generating new simulations, thereby optimizing the use of computational resources. The interested reader can find a detailed explanation of the working principle of CarpoolGP in Section 3.3.2.

Emulators like CARPoolGP are increasingly vital tools in cosmology, offering a powerful combination of speed, accuracy, and uncertainty quantification. When integrated into likelihood-free inference pipelines, they enable efficient exploration of high-dimensional parameter spaces, making them well-suited for modern astrophysical and cosmological parameter inference (Alsing et al., 2019; Cranmer et al., 2020).

Model Training and Optimisation with `Optuna`

Training a DL model involves optimizing its parameters, weights and biases, to minimize a predefined loss function $J(\vec{\theta})$, which measures the discrepancy between predicted and true outputs. The training process relies on optimization algorithms that guide the model's parameters toward configurations that reduce this loss.

At the heart of this process is the concept of *gradient-based optimization*. To update the weights, we compute the gradient of the loss function with respect to each model parameter. These gradients inform how each parameter should be adjusted to reduce the loss. The most widely used method is *gradient descent*, in which parameters are iteratively updated by moving them in the direction of the negative gradient. However, evaluating the gradient over the entire dataset can be computationally expensive. To address this, *Stochastic Gradient Descent (SGD)* is commonly used. Instead of computing gradients on the full dataset, SGD estimates them using small subsets of the data known as *mini-batches*, allowing for a faster and more scalable training (Ruder, 2016). This technique offers a trade-off between computational efficiency and stability. A critical hyperparameter in this process is the *learning rate*, which controls the size of parameter updates. Setting it too high may prevent convergence, while too low a value slows training and risks poor local minima (see Fig. 3.3). Figure 3.4 shows the evolution of the loss function during training with optimal and sub-optimal learning rate choices.

The gradients are computed using a method known as *backpropagation*, which efficiently applies the chain rule of calculus through the network layers. In backpropagation, the gradient of the final layer is computed first and then propagated backwards through the network. This allows intermediate layers to reuse partial gradient computations, significantly improving efficiency compared to naive approaches. The number of *epochs*, or complete passes through the training dataset, also plays a key role in determining model performance.

To improve model generalization and avoid overfitting, several regularization strategies are typically employed in the training phase:

- **Weight decay**: Also known as ℓ_2-regularization, this technique penalizes large weights by adding a term proportional to the squared magnitude of the weights to the loss function. The gradient update is thus modified as:

$$\vec{\theta} \to \vec{\theta} - \nabla J(\vec{\theta}) - \lambda\vec{\theta}, \tag{3.2}$$

  where λ is the regularization strength and $\vec{\theta}$ are the model parameters.

- **Early stopping**: During training, performance is monitored on a separate validation set. If the validation loss begins to increase while the training loss continues to decrease, training is stopped. This prevents the model from overfitting to the training data and preserves its ability to generalize.

Optimizers are algorithms that determine how weights are updated during training. While basic SGD is effective, more advanced methods are often used to improve convergence and stability:

- **Momentum**: This method helps accelerate training by incorporating a running average of past gradients. By doing so, it reduces oscillations and encourages smoother convergence. The momentum term $\vec{m}_t$ is updated as:

$$\vec{m}_t \to \beta\vec{m}_{t-1} + \alpha\nabla J(\vec{\theta}), \tag{3.3}$$

  and the parameters are updated as $\vec{\theta} \to \vec{\theta} - \vec{m}_t$, where β is the momentum factor and α the learning rate.

- **Adam**: In this thesis, we make use of the Adam optimizer (Kingma & Ba, 2014), which combines the benefits of momentum and adaptive learning rates. Adam estimates the first and second moments of the gradient and uses them to adapt the step size for each parameter. This optimizer has demonstrated to perform well across a wide range of problems with minimal manual tuning.

While choosing an optimizer is crucial, selecting the right *hyperparameters*, such as learning rate, number of layers, batch size, and regularization strength, is equally important. These hyperparameters are not learned during training and must be tuned externally. Manual tuning or grid search can be prohibitively expensive, especially when the parameter space is large or training is computationally demanding. To address this challenge, we employ `Optuna`, an automated hyperparameter optimization framework (Akiba et al., 2019).

`Optuna`[1] uses a technique known as *sequential model-based optimization* to guide the search for optimal hyperparameters. It defines a *study* composed of multiple *trials*, where each trial corresponds to a model trained with a different set of hyperparameter values. These are evaluated based on a user-defined objective function, typically a validation metric such as accuracy or loss. Optuna then uses algorithms like the Tree-structured Parzen Estimator (TPE) to propose new hyperparameter combinations that are likely to perform well.

Some of the main advantages that `Optuna` presents are its computationally efficiency, as poor-performing trials are pruned early, conserving computational resources. It also offers flexibility in the exploration of parameters, allowing to control the search space using defined conditional searches. Finally, is also offers a high automation level, as once the search space and objective function are defined, Optuna performs the search with minimal human intervention.

In this thesis, `Optuna` is used to fine-tune the hyperparameters of the NNs employed in the parameter inference. By systematically optimizing these hyperparameter configurations, we improve model generalization, accelerate convergence, and ensure consistent performance across diverse datasets.

3.2.2 Performance Metrics

In the context of simulation-based inference with NNs, a number of standard metrics are commonly used to assess both the accuracy of the parameter estimates and the reliability of the associated uncertainty predictions. These metrics are typically computed on a held-out test set of N samples, where each true parameter value θ_i is compared to the predicted posterior mean μ_i and its predicted standard deviation σ_i. Four widely used metrics are described below:

1. Root Mean Squared Error (RMSE). RMSE quantifies the power of our inference by measuring the average magnitude of errors between predicted and true values. It provides a single number that describes how close the predictions are to the actual values. The RMSE is computed as:

[1] https://optuna.org

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\theta_i - \mu_i)^2} \tag{3.4}$$

Low RMSE values indicate that the predictions are close to the actual values, signifying higher accuracy fo the model. High RMSE values indicate larger errors between predictions and actual values, signifying lower accuracy. RMSE is particularly useful in contexts where the magnitude of the error is important and provides an absolute measure of fit.

2. Mean Relative Uncertainty (ϵ). This metric tells us about the precision of the model by evaluating the relative size of the standard deviation of the prediction compared to the predicted means.

$$\epsilon = \frac{1}{N}\sum_{i=1}^{N}\frac{\sigma_i}{\mu_i} \tag{3.5}$$

A lower ϵ implies higher precision, meaning the predicted values have smaller relative uncertainties. Higher values of ϵ indicate lower precision, meaning the predicted values have larger relative uncertainties. This metric helps us understand the precision of the predictions, although it does not provide information about their accuracy.

3. Pearson Correlation Coefficient (r). The correlation coefficient measures the strength and direction of the linear relationship between the predicted and actual values:

$$r(\theta, \mu) = \frac{\text{cov}(\theta, \mu)}{\sigma_\theta \sigma_\mu} \tag{3.6}$$

An r value close to 1 indicates a strong positive linear relationship between predicted and actual values, meaning the predicted values closely follow the trend of the real values. An r value close to -1 indicates a strong negative linear relationship between predicted and actual values, indicating that the predicted values follow the opposite trend of the real values. If r is close to 0, this indicates little to no linear relationship between predicted and actual values, suggesting the predicted values are scattered randomly and do not align with the trend of the real values. The correlation coefficient is useful for understanding the degree to which the predictions follow the trend of the actual data.

4. Reduced Chi-Squared (χ^2). The reduced chi-squared provides a measure of how accurate the predicted standard deviations of the network are by comparing it to the residuals, i.e., the differences between the real and expected values.

$$\chi^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\theta_i - \mu_i}{\sigma_i}\right)^2 \tag{3.7}$$

A value of $\chi^2 \approx 1$ indicates that the model fits the data well. The residuals are consistent with the expected variance, considering the uncertainties in the data. In other words, the model's predictions differ from the observed values by approximately the expected amount, given the measurement uncertainties. A value of $\chi^2 < 1$ suggests that the uncertainties in the data have been overestimated, as the deviations between the observed and predicted values are smaller than what would be anticipated given the uncertainties. If $\chi^2 > 1$, the model does not fit the data well. The deviations between the model predictions and the observed data are larger than expected based on the uncertainties. This may suggest that the uncertainties in the data have been underestimated.

These metrics collectively offer a comprehensive assessment of the model's performance, highlighting both its accuracy and precision.

3.3 Parameter Inference in Cosmology

A central objective in cosmology is to infer the underlying physical quantities that govern the evolution of the Universe, such as the matter density Ω_m, the amplitude of primordial fluctuations σ_8, or the strength of baryonic feedback, from observable data. Traditionally, this has been achieved using likelihood-based statistical methods, where one defines an explicit probability distribution for the data given the model parameters, and applies Bayes' theorem to compute the posterior (Trotta, 2008; Verde, 2010).

However, the applicability of classical inference techniques is limited in modern cosmological contexts. Forward models, such as hydrodynamical or N-body simulations, are often highly nonlinear and stochastic, and their outputs are not amenable to simple parametric likelihoods. Additionally, observables such as galaxy cluster profiles, LSS, or weak lensing maps are inherently high-dimensional and structured, making the construction of tractable likelihoods either intractable or impossible (Elsner et al., 2019; Alsing et al., 2019). Even with simplifying assumptions (e.g., Gaussian fields, linear response models), such approximations can lead to biased posteriors and loss of information.

To overcome these limitations, the field has increasingly turned to *likelihood-free inference* techniques, often implemented using ML. These methods bypass the need for explicit likelihoods by learning a direct mapping from data to parameters, or from data to posterior distributions, using large ensembles of simulations as training data. NNs, normalizing flows, and SBI architectures provide scalable and flexible alternatives capable of extracting information from complex, high-dimensional data sets (Alsing et al., 2019; Cranmer et al., 2020; Březina et al., 2022).

In the following subsection, we focus on the CAMELS project (Cosmology and Astrophysics with MachinE Learning Simulations) (Villaescusa-Navarro et al., 2021), a pioneering suite of simulations specifically designed to enable ML-based inference in cosmology. CAMELS provides a controlled environment to test, benchmark, and develop inference methods using simulated universes with varying cosmological and astrophysical parameters.

3.3.1 The CAMELS project

The CAMELS project (Villaescusa-Navarro et al., 2021) is a large-scale initiative designed to explore the use of ML techniques for cosmological inference. The central goal of CAMELS is to generate a statistically rich and physically diverse simulation dataset that allows models to learn the relationship between high-dimensional observables—such as matter density fields, halo properties, or gas profiles, and the underlying cosmological and astrophysical parameters that shape them.

CAMELS achieves this by producing thousands of hydrodynamical and N-body simulations that systematically vary both cosmological parameters (e.g., Ω_m, σ_8) and astrophysical feedback parameters (e.g., SN and AGN feedback strengths). These simulations are run using several independent codes, such as IllustrisTNG (Springel et al., 2018), SIMBA (Davé et al., 2019), Magneticum (Dolag et al., 2025), Eagle (Schaye et al., 2015), and Astrid (Bird et al., 2022), each of the models incorporating different physical prescriptions. A summary of the simulations available to date is presented in Table 3.1. By varying the input parameters across a wide domain, CAMELS enables the construction of flexible ML models that can generalize across different regimes of physical theory and numerical implementation.

To efficiently explore a multi-dimensional parameter space, the CAMELS simulations are sample across the possible cosmological and astrophysical parameters using samplings such as Latin Hypercubes or Sobol Sequences. Each simulation corresponds to a unique point in the sampled space. This diversity in simulations enables both supervised learning and SBI tasks, making CAMELS a benchmark suite for testing new methodologies.

Type	**Code**	**Subgrid model**	**Generation**	
			First	**Second**
Hydrodynamic	Arepo	IllustrisTNG	3,219	1,192
	Gizmo	SIMBA	1,171	
	MP-Gadget	Astrid	2,080	
	OpenGadget	Magneticum	77	
	Swift	EAGLE	1,052	
	Ramses		552	48
	Enzo		6	
	Gadget4-Osaka	CROCODILE	260	148
	Gizmo	Obsidian	27	
N-body	Gadget-III	–	6,136	1,072

Table 3.1: Summary of the available simulations in the CAMELS suite, classified by simulation codes, subgrid models.

Several recent studies have made use of the CAMELS simulation suite to perform parameter inference using ML, particularly focusing on field-level likelihood-free inference. For example, Shao et al. (2022) trained graph NNs on halo catalogs from N-body simulations and showed that models using positions and velocities of halos with masses $\gtrsim 10^{10} h^{-1} M_\odot$, can robustly infer cosmological parameters like Ω_m and σ_8 with high accuracy. Their models generalize well

across N-body simulations using different codes. Building on this, Shao et al. (2023) extracted a symbolic equations using graph NNs that predicts $\Omega_{\rm m}$ directly from positions and velocities of dark matter halos. de Santi et al. (2023) extended this approach to galaxy catalogs from hydrodynamic simulations, finding that models using both galaxy positions and radial velocities maintain robustness across different implementation of galaxy formation physics, when inferring $\Omega_{\rm m}$. This robustness persists even under observational systematics, as shown in de Santi et al. (2025), who incorporated realistic effects such as masking and redshift distortions. Finally, Echeverri-Rojas et al. (2023) demonstrated that NNs can recover $\Omega_{\rm m}$ from integrated galaxy properties, such as total stellar mass and metallicity, across multiple simulation codes and for redshifts $z \lesssim 3$, although fall short when infering σ_8 or parameters regulating feedback processes.

These works collectively show that while ML-based parameter inference is promising, it is highly sensitive to the choice of observables, the preprocessing of input data, and the structure of the neural architecture. In particular, the entanglement between cosmology and baryonic physics remains a major obstacle. This and the inference of all possible cosmological and astrophysical parameters are challenges we aim to address in this thesis.

3.3.2 The `CARPoolGP` Emulator

To address the high computational cost of generating large numbers of high-resolution simulations, Lee et al. (2023) introduced `CARPoolGP` (Cluster-Accelerated Regression via Pooling and GPs), a novel emulator framework designed for efficient and accurate modeling of complex simulation outputs. Developed in the context of the CAMELS-ZOOMGZ suite, CARPoolGP enables fast emulation of halo-scale properties, such as matter and gas profiles, across varying cosmological and astrophysical parameters, while accounting for both halo-to-halo variations and the effects of different physical models.

The central idea behind CARPoolGP is to pool information across pairs of simulations of the same halo. In zoom-in simulations, a given halo is re-simulated under different physical assumptions or parameter values (e.g., different feedback strengths or cosmologies), called a surrogate simulation. Rather than treating the original or base simulation and its surrogate as independent samples, CARPoolGP constructs a joint probabilistic model that learns a shared structure across both versions of the same halo. Figure 3.5 shows a schematic representation of both base a surrogate simulations across a parameter Q. Specifically, it models the outputs using a GP with a decomposition into two components: a latent, halo-specific kernel that captures shared physical structure, and an additive kernel encoding the effects of varying parameters. This pooling reduces the variance of the emulator and improves generalization, especially when training data is limited.

Formally, CARPoolGP is built within a multi-output GP regression model. For each halo, the different simulations are treated as multiple outputs of a shared underlying process. This hierarchical modeling approach enables the emulator to disentangle the influence of cosmology and feedback from intrinsic halo-specific features. Moreover, it allows the model to interpolate smoothly in parameter space, even in regions where only low-fidelity or sparse data is available. Figure 3.6 shows a simplified representation of the improved accuracy and predictive uncertainty with the CARPoolGP methodology in comparison to a standard random sampling.

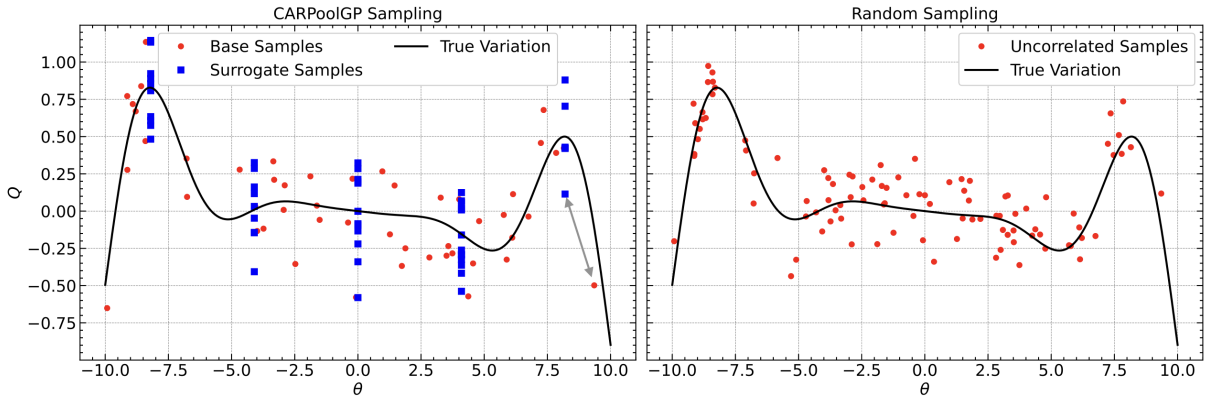


Figure 3.5: Illustrative example demonstrating the performance of CARPoolGP, as introduced by Lee et al. (2024). The left panel displays 50 base samples (red) and 50 surrogate samples (blue), where each surrogate sample is correlated with a base sample located at the nearest parameter region. A grey double-headed arrow highlights one such correlated pair. In contrast, the right panel shows 100 uncorrelated samples drawn uniformly at random across the parameter space. The black line in both panels represents the true variation of the target quantity Q.

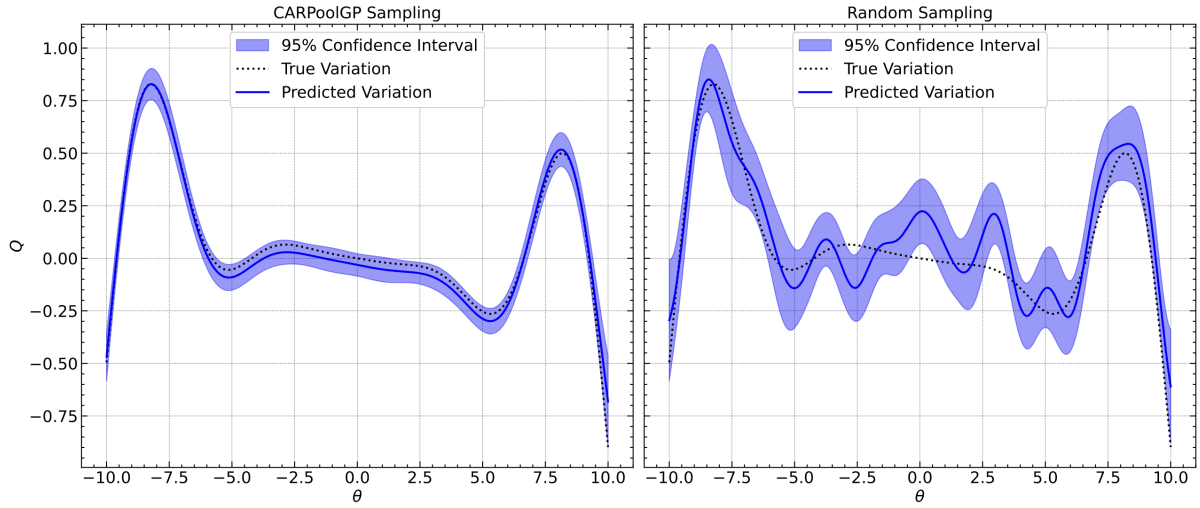


Figure 3.6: Comparison of emulation performance using CARPoolGP (left panel) and standard random sampling (right panel), both based on the same number of samples. The blue line and shaded region represent the predicted mean and 95% confidence interval, respectively, while the black dotted line shows the true underlying variation of the quantity Q. In the CARPoolGP case, base and surrogate samples are correlated across parameter islands, leading to significantly improved accuracy and reduced predictive uncertainty. Credit: Lee et al. (2024)

In this thesis, we adapt CARPoolGP to emulate thermodynamic profiles of the hot gas of galaxy clusters, such as pressure, temperature, entropy, X-ray surface brightness and electron density, across variations in astrophysical parameters. The emulator is trained on a set of high-resolution zoom-in simulations, enabling us to predict smooth and realistic profiles for unseen configurations. These emulated profiles serve as inputs for our simulation-based inference model, expanding our training set without requiring thousands of expensive simulations.

Part IV

Results

4 | Simulating the Local Web: Properties of Local Galaxy Clusters

The content of this chapter has been published by Hernández-Martínez et al. (2024) in A&A. Some adjustments have been made.

4.1 Constructing a Sample of Local Universe Galaxy Clusters

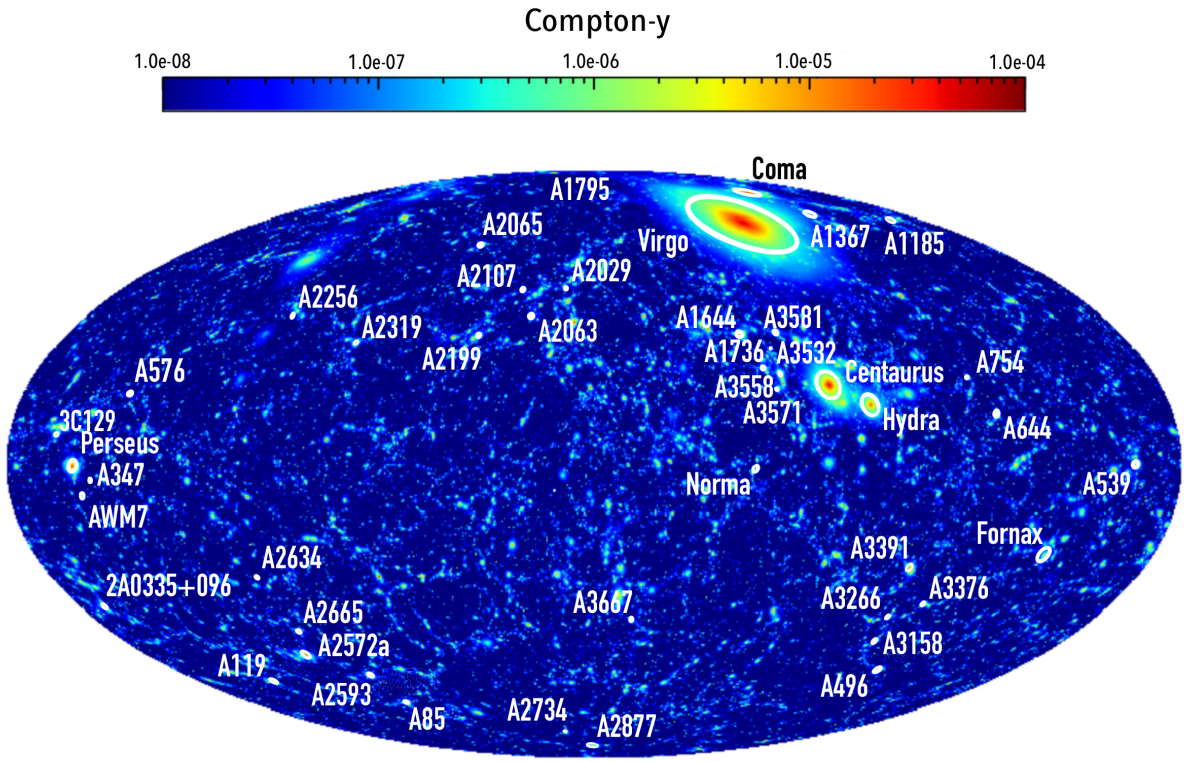
4.1.1 A General View

The local Universe encompasses a spatial region of approximately $z < 0.1$ ($R < 200$ Mpc/h). Due to its close proximity, the local Universe is the most extensively studied part of the Universe, presenting a unique laboratory to study galaxy clusters across a wide range of shapes, characteristics, and dynamical states.

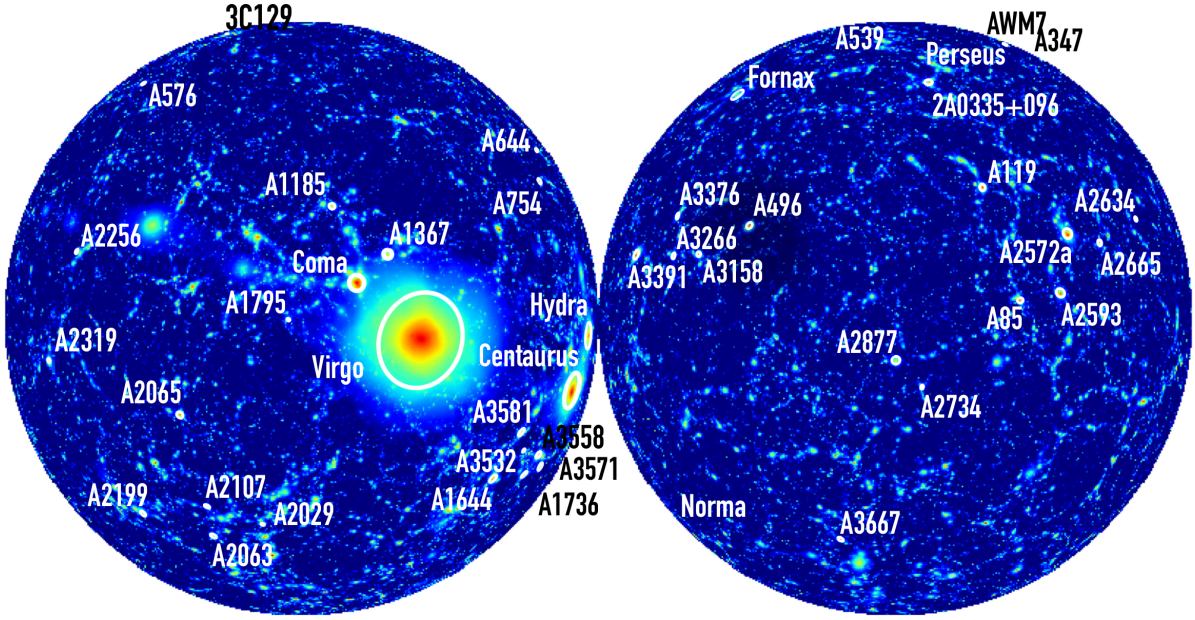
In our local volume, located approximately 15 Mpc away, the Virgo cluster stands out as the primary defining feature of our immediate cosmic neighborhood and covers the largest area in the sky. Its closeness has made it among the best observed clusters of our Universe (Binggeli & Huchra, 2000), also hosting the well-known central galaxy M87.

Moving beyond Virgo, a roster of renowned and well-observed clusters such as Centaurus, Fornax, Hydra, Norma, and Perseus take the central stage within the local volume. Perseus, in particular, has been widely observed through X-ray studies using Chandra observations (Weisskopf et al., 2000; Fabian et al., 2011), as well as ROSAT PSPC (Branduardi-Raymont et al., 1981; Schwarz et al., 1992; Allen et al., 1993), Hitomi (Aharonian et al., 2017), XMM-Newton data and Suzaku mosaics (Simionescu et al., 2012; Simionescu et al., 2013; Urban et al., 2013). These studies have shown very distinctive features that make this cluster worth studying, like its pronounced cool core, with a sharp peak in X-ray surface brightness and decreasing temperature toward the center, and strong signs of ongoing AGN feedback processes (Böhringer et al., 1993; Fabian et al., 2000a). Perseus shows also signs of a powerful merger, yet its cool core remains undestroyed. However, the motion induced by this merger seems to penetrate the cool core of the cluster, forming a large-scale sloshing inside the cluster (Simionescu et al., 2012).

At a distance of roughly 100 Mpc, we find the Coma cluster, a remarkable rich and complex structure, with an also rich history of detailed observations (Fitchett & Webster, 1987; Mellier et al., 1988; Biviano et al., 1996; Biviano, 1998; Planck Collaboration et al., 2013b). White et al.



(a)



(b)

Figure 4.1: Full-sky projections of the Compton-y signal encompassing the entire simulation volume in SLOW are shown in the following manner: (a) A full-sky projection in galactic coordinates. (b) On the left, the galactic northern sky is projected, while on the right, the galactic southern sky is shown. The zone of avoidance lies at the edge of the spheres. Circles are employed to mark the projected r_{500} values for the cross-identified clusters. Note that the cross-identification of A576 and A3571 was improved with respect to Dolag et al. (2023)

.

(1993) showed that the Coma cluster was plausibly formed by the merging of several distinct substructures which are not yet fully merged and Vikhlinin et al. (1994) showed that the extended regions of X-ray emission in the central region of Coma are associated with the subgroups NGC 4889 and 4874, two galaxies lying at the center of the cluster. It also presents signs of recent infall in the form of linear filaments to the southeast (Vikhlinin et al., 1997), extending ≈ 1 Mpc from the cluster center toward the two central galaxies aforementioned (Andrade-Santos et al., 2013; Neumann et al., 2001; Arnaud et al., 2001; Briel et al., 2001).

Beyond the previously mentioned clusters, we can find a significant number of very massive clusters extending up to a distance of over 150 h^{-1}Mpc (Courtois et al., 2013; Dupuy, A. & Courtois, H. M., 2023). Therefore, within the cosmological volume where our simulation is constrained, we account for a rich variety of galaxy clusters. This allows us to conduct detailed studies of their evolution as well as testing (hydro)dynamical and physical processes governing their formation.

4.1.2 Collection Sample of Local Galaxy Clusters

The previously mentioned examples show the immense diversity of processes and dynamical states observed in local clusters. Thus, detailed observations of the local Universe add immeasurable value to astrophysical discussions like the survivability of cool cores (e.g. McDonald et al., 2013), the effects of mergers and AGN associated feedback (e.g. Eckert et al., 2021), as well as cosmological discussions on LSS formation (e.g. Mathis et al., 2002), mass estimation based on individual clusters (e.g. Lebeau et al., 2023) and inference of cosmological parameters (e.g. Tully et al., 2023).

In this work, we collected information from a total of 221 local Universe clusters and groups from the literature. To do so, we combined the *CLASSIX* catalog of local X-ray clusters and groups (Böhringer et al., 2016), the Tully galaxy groups catalog[1] (Tully, 2015) and the SZ Cluster Database[2]. This list was then expanded with additional data from the X-rays Galaxy Clusters Database (BAX)[3]. In addition, we included other, well-known local clusters and groups from various individual observations. From this collection, we selected all clusters with SZ derived M_{500} masses[4] above 2×10^{14} $M_\odot$, with a corresponding position in the Tully galaxy group catalog. Additionally, we incorporated six renowned local clusters with an X-ray-derived M_{500} value exceeding 10^{14} $M_\odot$ as well as prominent galaxy clusters from the Tully galaxy groups catalog where additional X-ray data were available. From this, we constructed a final set of 46 clusters where we identified a counterpart candidate within the simulations. Within this selection, the inferred masses of the clusters are based on vastly different methods. Therefore, in the later analysis, we restricted direct comparisons of M_{500} to the subsample of clusters where M_{500} was based on the observed SZ signal.

[1] https://edd.ifa.hawaii.edu

[2] http://szcluster-db.ias.u-psud.fr

[3] http://bax.irap.omp.eu

[4] M_{500} is the mass within the radius R_{500} for which the cluster mean total density is 500 times the critical density at the cluster redshift

4.1.3 Finding Galaxy Cluster Replicas in SLOW

For the sample of 46 galaxy clusters, we collected data from the literature for different mass estimates and X-ray observables. The integrated Compton-y signal (Y) within R_{500} together with the corresponding M_{500} mass estimate was taken from the SZ Cluster Database, while X-ray data like temperature and luminosity in the 0.1-2.4 keV range of most of the clusters where taken directly mainly from (Ikebe et al., 2002; Shang & Scharf, 2008; Planck Collaboration et al., 2011), otherwise we took the values quoted in BAX. The inferred X-ray based M_{500} masses where extracted from (Chen et al., 2007). The Tully galaxy groups catalog presented dynamical mass estimations, $M_{\rm dyn}$, which we converted to the virial mass, by correcting it down by 12% as presented by Sorce et al. (2016b) for their simulated Virgo cluster at z = 0, and then converted the virial mass to M_{500} using the conversions presented in Ragagnin et al. (2021).

In Table 4.1, we report the observed position as sourced from the according group within the Tully galaxy catalog, together with the position of the replica candidate within our simulation. In addition, we list the relative displacement between the observed cluster position and that of the replica candidate. Table 4.3 displays the observational quantities alongside the corresponding values from the replica candidate, as obtained through the simulation, for comparison. This now allows a more thorough comparison and emphasizes the strength of comparing multiple observational signals to their counterparts from the simulation. Typically, X-ray luminosity is among the most commonly measured quantities, although it is most sensitive to inhomogeneities such as clumping and hence most dependent on the detailed treatment of cooling and star-formation processes in simulations. Therefore, it is difficult to accurately predict through simulations. In a fully virialized cluster formed solely through gravitational collapse, X-ray luminosity, temperature as well as SZ signal directly correlate with the mass of the cluster. However, in practice, such relations suffer differently from scatter due to internal structures, deviation from spherical symmetries as well as deviation from hydrostatic equilibrium. Indeed, different comparisons suffer differently from individual observational biases, as well as from the incompleteness of the simulations when it comes to reproducing the actual galaxy clusters and the treatment of relevant physical processes affecting them. Therefore, we can get a more complete picture if we make use of the entire set of observables along with the inferred masses for comparison.

To obtain the above-mentioned sample of counterpart candidates of local galaxy clusters within the SLOW simulation, we applied the following procedure:

1. Starting from our collection of observed local galaxy clusters, we first associated the cluster/group in Tully's North/South Catalog. We used the member galaxy which is closest to the position on the sky and redshift as the center and used the according supergalactic X, Y, Z coordinates. For the few observed clusters without such an association, we converted sky positions and redshifts directly into supergalactic X, Y, Z coordinates. When computing the positions of clusters within the simulation, we used the observer position as presented in (Dolag et al., 2023) as the optimized center.

2. We selected the Compton-y derived M_{500} cluster mass, or if unavailable, the X-ray derived mass or dynamical mass, depending on data availability.

3. We then identified all massive halos within a small "search radius," and with a lower mass M_{500} cut typically not less than six times the observed mass. If we fail to identify a suitable counterpart in close proximity to the observed cluster, we expand our search radius, typically not exceeding 45 Mpc/h. The resulting median distance in which we found our cluster replicas is $\approx$ 25 Mpc/h.

4. In cases where multiple candidates existed in terms of position and mass, we conducted a comparison of temperature and luminosity values, along with an assessment of the cluster's close surroundings and dynamical state, especially if it was involved in a merging process, or a stage before it. Additionally, for supercluster regions like Hydra-Centaurus or Shapley, we considered the geometry of the surrounding cluster environment, its relative position to nearby clusters, and how well the shape of the surrounding filaments matched the observed galaxy distribution (see also Seidel et al., 2024).

Although the selection process described above is guided by the availability of observational data, the combined usage of *Planck* selected clusters, the X-ray flux-limited *CLASSIX* catalog, and the 2MRS-based Tully galaxy groups catalog leads to a very complete cover of massive local galaxy clusters. Such a list of counterpart candidates is specifically designed to focus primarily on the massive clusters identified through observational proxies. A comprehensive map of the entire sky, featuring the 46 cluster replicas included in our SLOW simulation set, is provided in Fig. 4.1. We note that this sample of clusters not only contains the 13 most massive clusters but also 70% of clusters with M_{500} larger than 2×10^{14} $M_\odot$ from the *Planck* SZ cluster catalog (Planck Collaboration et al., 2014a) within a distance of $\approx$ 200 Mpc/h.

4.2 Scaling Relations for Low Redshift Clusters

It has been extensively demonstrated that the subgrid model presented in Sect. 2.4.1 yields galaxy and ICM properties in galaxy clusters that closely align with observed trends and properties (Gupta et al., 2017; Singh et al., 2020). The Magneticum simulations employing the same hydrodynamical scheme, with a slightly earlier adaptation of the galaxy formation treatment, has been compared to SZ data from *Planck* (Planck Collaboration et al., 2014b) and SPT (McDonald et al., 2014). They have successfully replicated the observable X-ray luminosity relation (Biffi et al., 2013), among other ICM characteristics. In this context, we will test the results of our SLOW simulations to demonstrate their capability to reproduce SZ data and relationships with X-ray observables scaling relations.

Scaling relations are particularly interesting as they are tightly related to the physics of cluster formation and evolution. In the idealized framework where gravity is the dominant process in cluster evolution, self-similar models predict simple scaling relations between cluster properties, such as temperature, luminosity, and Compton-y value, with total mass (Kaiser, 1986). In general, these relations are described by power laws, around which some data points scatter according to a log-normal distribution. These relations describe positive correlations, so that larger systems have on average higher values of the correlated parameter. On top of this, as these relations come from the premise of self-similarity and thus gravity domination, any deviation of the relation may

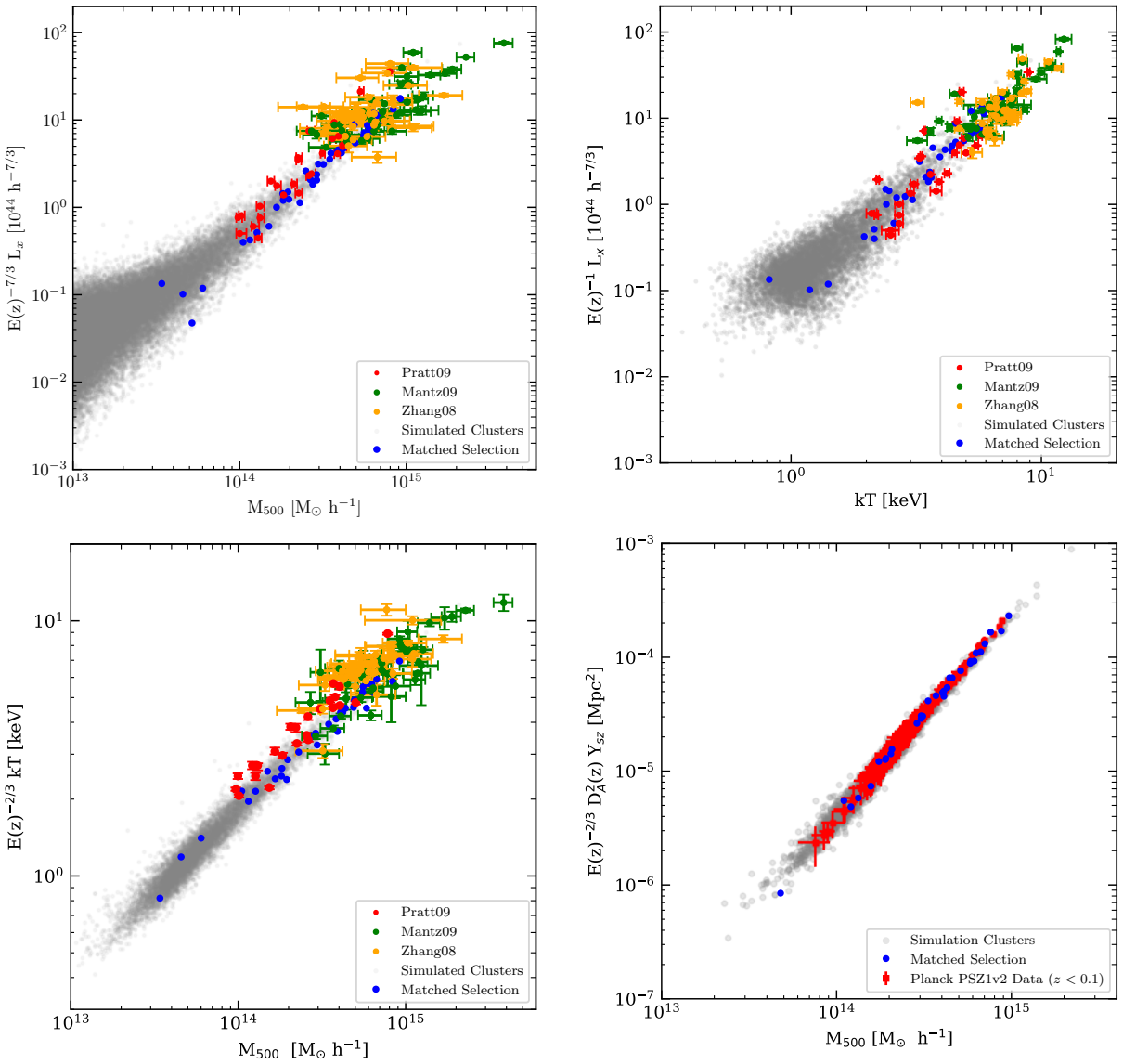


Figure 4.2: Scaling relations within R_{500} for simulated clusters at z = 0 and for observed low redshift clusters. Gray data points represent the cluster values from our simulations, estimated using SUBFIND (Springel et al., 2001; Dolag et al., 2009; Saro et al., 2006). The clusters that have been cross-identified are highlighted in blue. Luminosity-mass (L_X – M_{500}, top left), luminosity-temperature (L_X – kT, top right), temperature-mass (kT – M_{500}, bottom left) panels show observational data from Pratt et al. (2009) (depicted as red points), which encompasses clusters with redshifts ranging from 0.05 to 0.164, Zhang et al. (2008) (depicted in yellow) with clusters spanning the redshift range of 0.14 to 0.3, and Mantz et al. (2010) (depicted in green), which includes clusters with redshifts less than 0.3. The bottom right panel depicts the Compton-y – M_{500} scaling relation. For observational data in this case, we rely on Planck's catalog PSZ1v2 (Planck Collaboration et al., 2014a), considering only clusters with a redshift of less than 0.1.

be a sign of important hydrodynamical processes taking center stage in the evolutionary history of these structures. Therefore, scaling relations are an important tool in cosmology as well as in the study of the thermodynamic history of clusters and their ICM.

We tested our simulated clusters' scaling relations against the observed clusters' scaling relations reported by Zhang et al. (2008); Pratt et al. (2009); Mantz et al. (2010). The first three panels in Fig. 4.2 show the results of $L_x - M_{500}$, $L_x - T_x$ and $T_x - M_{500}$ scaling relations for the clusters selections given in Pratt et al. (2009) in a redshift range of $0.05 < z < 0.164$ (in red), Zhang et al. (2008) spanning the range $0.14 < z < 0.3$ (in yellow) and Mantz et al. (2010) containing clusters with $z < 0.3$ (in green). Even if this redshift threshold is small enough to keep evolution effects to a small level (Reichert et al., 2011), we corrected the data for possible redshift effects.

We calculated the luminosities, temperatures, and masses for our clusters in the R_{500} radius using SUBFIND (Springel et al., 2001; Dolag et al., 2009; Saro et al., 2006). For the temperature, in order to focus on the ICM gas, we performed a cutout above 10^5K. The simulated data points for all clusters in our simulation are shown in gray, while the corresponding data points for our set of 46 clusters are plotted in blue. All three figures show a great agreement between the observed clusters and the simulated clusters in our box.

Furthermore, we examined the Compton-y - M_{500} relationship for our simulated clusters. In this instance, we selected observational data from the *Planck* database for clusters with redshifts below 0.1. For the simulated values, we generated Compton-y maps for each of the clusters in our simulations employing SMAC (Dolag et al., 2005), and computed the total Compton-y emission within R_{500}. Once more, our simulations faithfully reproduce the expected scaling relation presented in the observational data. Thus, we can conclude that our improved subgrid model presented in Sect. 2.4.1 can also successfully replicate the main ICM characteristics such as $L_x - M_{500}$, $L_x - T_x$ and $T_x - M_{500}$ and $Y_{500} - M_{500}$ scaling relations.

4.3 Assessing the Fidelity of Simulated Local Galaxy Clusters in SLOW

Table 4.1 presents a comparison of the positions of local clusters and their simulated counterparts. The median separation between the observed cluster position and that of its replica in our simulation is of 25 Mpc/h, with only three exceeding a distance of 45 Mpc/h.

During the cluster identification process based on observed positions, two primary sources of errors come into play. Firstly, when using distance measurements derived from redshift for the positions of our galaxy clusters, we encounter uncertainties associated with the peculiar velocity of the cluster. To quantify this uncertainty, we referred to the study by Dolag & Sunyaev (2013, Figure 1), where the authors presented histograms of peculiar velocities of galaxy clusters in a cosmological box across different mass bins. We adopted the 1-sigma error values for each mass bin presented in Figure 1 of the mentioned paper to illustrate the typical peculiar velocity uncertainties, which have a mild dependence on mass. These resulting uncertainty values are displayed in column 8 of Table 4.1.

On the other hand, in simulations founded on peculiar velocities, uncertainties coming from

distance moduli propagate to radial velocities and subsequently extend further to uncertainties on the reconstructed density and total velocity field. Consequently, uncertainties in the observed distances (which are around a 20% of the distance value) are directly coupled with displacements of the three-dimensional positions within the evolved simulation.

In addition, a positional error originates from the creation of the constraints for the ICs, where scales below the linear threshold are typically non-constrained. However, this generally adds an error of 3 – 4 Mpc, which is significantly smaller than the two errors discussed above.

Therefore, the relative distances found for our simulated replicas, as presented in Table 4.1, are reasonably small when compared to the above-discussed uncertainties, showing that we have a high-fidelity reproduction of these observed clusters in our simulation. Note that for half of the cluster replicas, the relative distances found are already almost fully covered when only considering the uncertainty of the observed cluster position based on their expected, typical peculiar velocity.

4.3.1 Detection significance

Originally, structure identification in reconstructions of the local Universe primarily relied on tracing the large-scale linear field (Hoffman, 1993; Zaroubi et al., 1995; Fisher, 1995; Bistolas & Hoffman, 1998) More recently, reproductions based on simulations have shifted their focus toward individual, collapsed structures, leading to a manual one-to-one structure and galaxy cluster identification (e.g., Klypin et al., 2003; Dolag et al., 2004; Carlesi et al., 2016b; Sorce, 2018; Sorce et al., 2019, 2020, 2022). Recently, Pfeifer et al. (2023) extended this type of identification by associating well-defined probabilities to the cross-matched galaxy cluster pairs. Our current objective extends beyond the previous approach. We aim to identify collapsed, nonlinear structures at smaller scales within the larger-scale cosmological hydrodynamical box and assess their significance and the quality of their reproduction given the variety of multiwavelength observational data available for the individual galaxy clusters.

Therefore, once we have found counterparts for our galaxy cluster selection, we can consider the probability of such a finding compared to a random position in a very large simulation box. Here, we can take the relative distance from the observed position, the inferred mass of the galaxy cluster, or directly any observational signal that is typically used as a proxy for the cluster mass (like X-ray luminosity, temperature, etc.) directly into account.

In mathematical terms, we want to perform a hypothesis test, where we ask what the p-value of our finding is under the assumption of no difference compared to a null hypothesis, which in this case is a random simulation. This can give us a measure of how likely it is that our finding is in fact a product of the constraints in our ICs. Such a procedure is naturally only applicable for rare (e.g., relatively high-mass systems), but can be extended to smaller accompanying systems, see Seidel et al. (2024) for details. Particularly, we performed this test using the four main cluster properties that we considered throughout this study, which are M_{500}, $L\mathrm{X}_{500}$, $T\mathrm{X}_{500}$ and the $Y\mathrm{SZ}_{500}$ signal. This time, X-ray luminosities were also calculated using SMAC in the 0.1 – 2.5 keV band. We subsequently explain the procedure for M_{500}, knowing that the analogous procedure applies to the rest of the properties.

The steps are the following:

1. We select bins for M_{500}, ranging from 10^{14} $M_\odot$ to 10^{15} $M_\odot$ in steps of 10, giving a total amount of ten bins. Each of these bins will serve as a mass threshold for our cluster search.

2. We place a virtual observer at random positions in our cosmological box (a total of 1 million positions per mass bin), which is equivalent to positioning ourselves in a random box.

3. We choose each of these mass bins and use it as a threshold for the cluster property. Then we search in a radius around the random position of the first cluster, which has a value for M_{500} larger than the threshold, and we save this distance as our minimum radial distance. We perform this step using all bins as thresholds.

4. We compute a cumulative probability function for each of the mass bins, in order to calculate the one-tailored p–value for each cluster.

5. Now we study our selected clusters in the light of these cumulative probabilities. We take the relative positions of our clusters and perform a spline interpolation between the cumulative probabilities of all bins at a fixed radius. By doing this, we can estimate the exact probability of each of our clusters in terms of the M_{500} value.

6. We calculate the significance of the match as 1 – p, where p is the p-value associated with that cluster. The significance values for each cluster are listed in Table 4.1.

Figures 4.3 and 4.4 depict the cumulative distribution functions representing various bins of our four central properties. In these figures, blue points signify the observed clusters, accompanied by their associated property uncertainties. Correspondingly, the simulated counterpart values are denoted in red. Apart from the mass uncertainty, we also need to consider the distance uncertainty originating from peculiar velocities, as well as distance moduli, as discussed in Sect. 4.3. Given the uncertainty in how the distance moduli errors affect our simulations, we have limited our analysis to the peculiar velocity uncertainty. For clarity, we have graphically represented this uncertainty as a red line only for five clusters at different relative distances: Coma, Centaurus, A3532, A2107, and A496. This uncertainty applies however, to all represented clusters.

Probability values below the -1σ threshold indicate a significance level exceeding 80%. Consequently, we classify all clusters falling below this threshold as highly significant matches. It is worth noting that the significance of these clusters can exhibit notable variations when considering distance errors. In the worst-case scenario, even with distance errors increasing the relative distance, the significance for highly significant clusters remains at higher than 50%. In contrast, under the best-case scenario, where the relative distance gets reduced, their significance can increase dramatically, surpassing the 99% mark.

Some of the lines for the values representing the higher-mass bins show unusual patterns, such as uneven spacing between bins or instances where two bins overlap. This is attributed to the inherent limitations of small number statistics within our cosmological box. As we move into higher property value ranges, the occurrence of clusters with such values diminishes. Nevertheless, it is important to note that these irregularities do not have any relevant impact on our significance estimate.

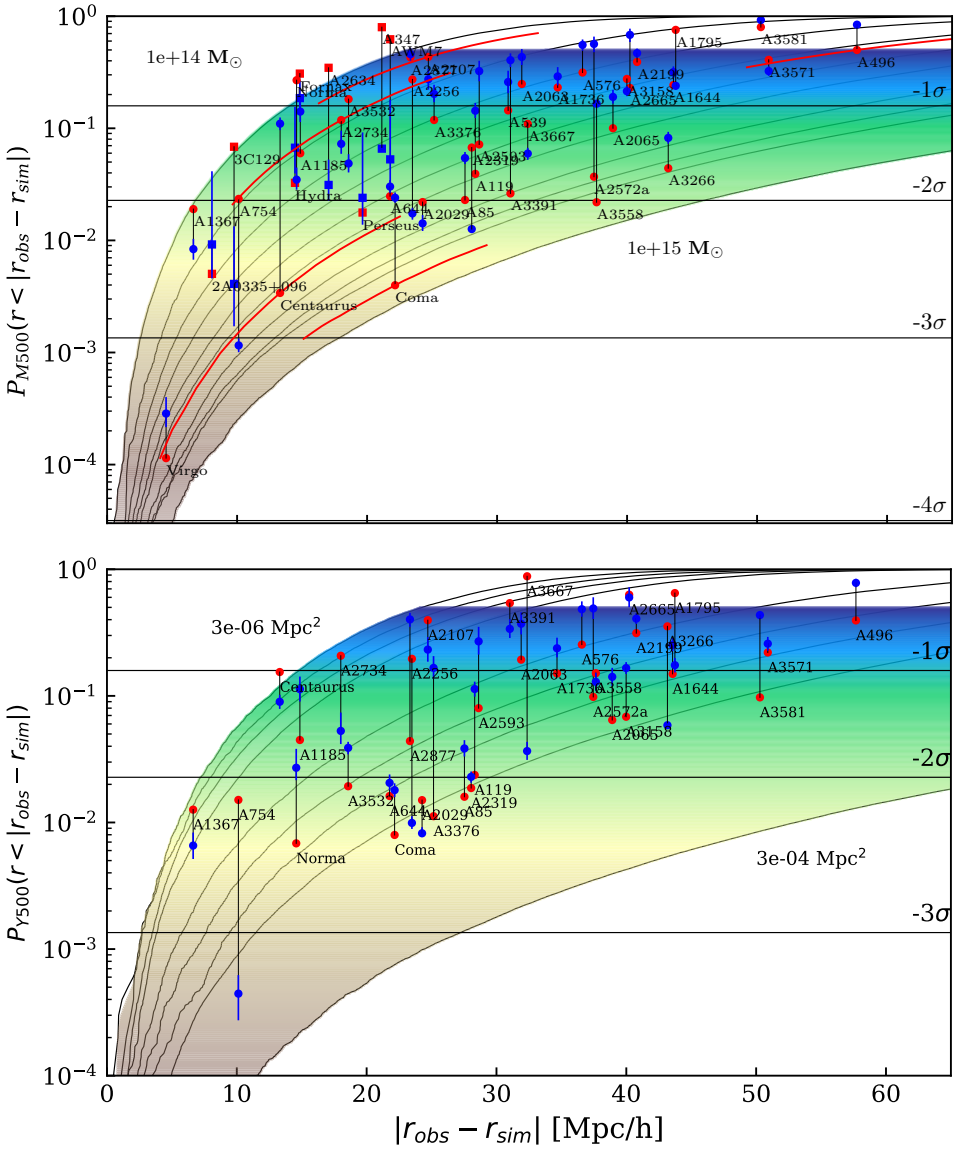


Figure 4.3: Cumulative distribution functions for M_{500} and Y_{500}, respectively. The top panel displays bins ranging from 10^{15} to 10^{14} $M_\odot$, with increments of ten, showing the random expectation of finding a cluster of a certain mass or higher within a sphere whose radius corresponds to the distance displayed on the x-axis. Simulated M_{500} values, as estimated by SUBFIND, are represented in red. The blue data points represent the observed M_{500} values obtained from clusters *Planck* signals, including their associated errors. In cases where SZ-derived M_{500} values were unavailable, X-ray-derived M_{500} values were used (indicated by square points). For cluster A347 X-ray-derived mass information was not available, thus we used the M_{500} mass derived by converting the dynamical mass estimated by Tully. The observational uncertainties in the cluster's position for Coma, A2734, AWM7, and Centaurus are represented by red lines, as an illustrative example of how such positional uncertainty can impact the results of the significance study (see Sect. 4.3.1 for a more comprehensive discussion). The bottom panel displays the cumulative distribution function for the SZ-derived signal in R_{500}. Simulated datapoint values where estimated using SMAC (Dolag et al., 2005). The bins range from 3×10^{-6} to 3×10^{-4} Mpc^2, with increments of 2.

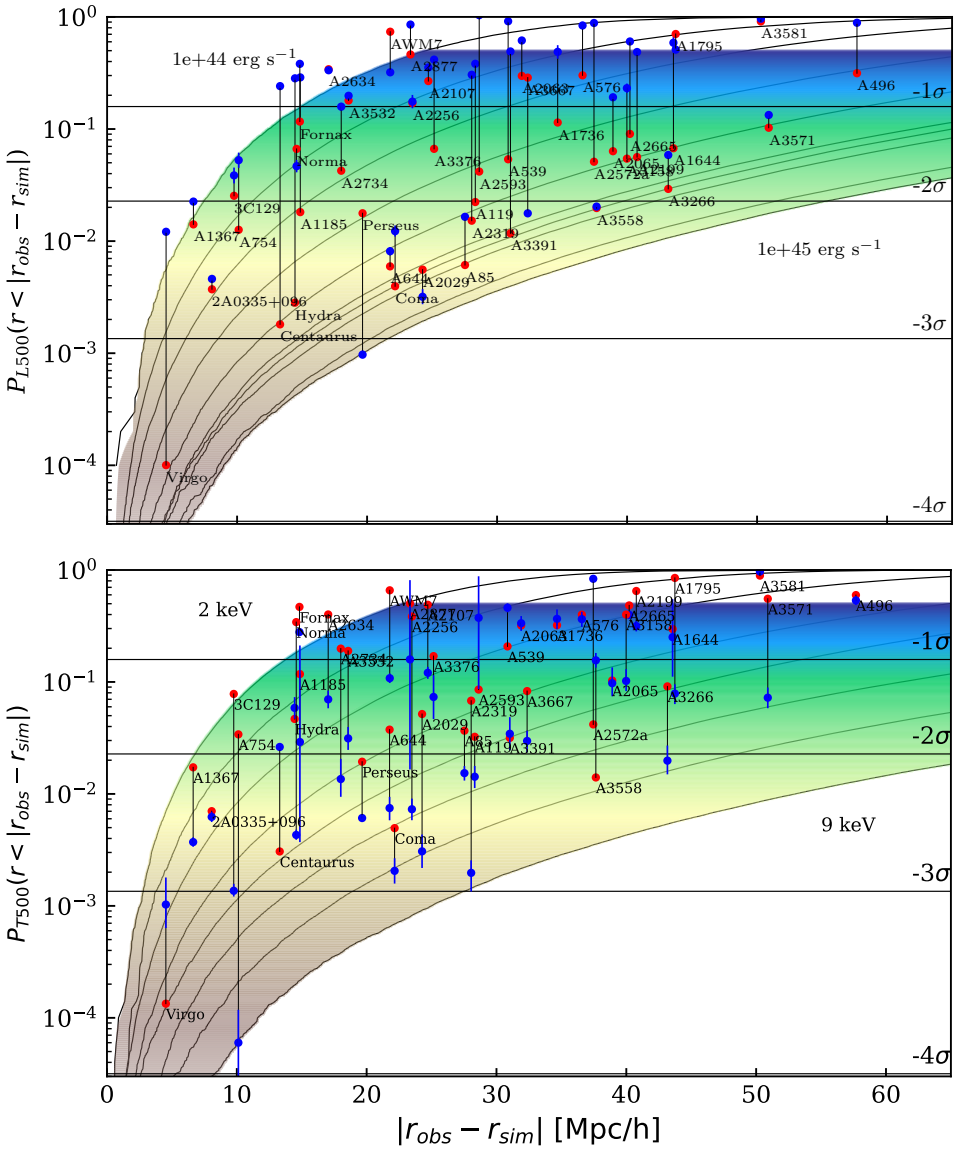


Figure 4.4: Similar to Fig. 4.3, we show cumulative distribution functions for $L_{\rm x}$ and $T_{\rm x}$ within R_{500}. For a better comparison with observations, we estimated the X-ray luminosity in the band 0.1-2.4 keV for every cluster using SMAC. In the top panel, the bin values range from $10^{44}{\rm erg}\times{\rm s}^{-1}$ to $10^{45}{\rm erg}\times{\rm s}^{-1}$, with increments of 10. Uneven spacing between higher-value bins has to do with the inherent limitations of small number statistics within our cosmological box (see Sect. 4.3.1 for a deeper discussion). The bottom panel presents the cumulative distribution function from the X-ray temperature in R_{500}. We estimated this temperature using SUBFIND. The bin values range from 2 keV to 9 keV, with increments of 1 keV.

Lastly, we would like to emphasize the exceptionally high significance values observed across all properties for several pivotal local galaxy clusters, notably including Virgo, Coma, Centaurus, Perseus, Fornax, and Norma, among many others (see Table 4.1). These significance values, hovering around 1.0, strongly indicate that the identification of these clusters is highly likely attributable to the constraints we have imposed, rather than being a random occurrence.

4.3.2 Properties' comparison

A notable advantage of our constrained local Universe simulations lies in our ability to perform direct one-to-one comparisons once we have successfully matched our clusters with their simulated counterparts. Figure 4.5 illustrates the property ratios between observed and simulated structures, in relation to their relative distances. The data points are color-coded based on their significance, with higher significance levels represented by blue, while lower significance levels are denoted by yellow.

In Fig. 4.5a, we illustrate the mass ratio of SZ-derived mass within R_{500} divided by the SUBFIND-estimated M_{500}. The dashed line in the plot represents the median value of the data, while the red-shaded region surrounding the median corresponds to the dispersion observed in the Compton-y – M_{500} scaling relations (Planck Collaboration et al., 2011, see Figure 5). The median value is 0.87 indicating that the SZ-derived mass tends to be lower than the simulated mass. This finding agrees with estimates of the hydrostatic mass bias.

The hydrostatic mass bias is relevant for gas-derived mass estimations such as X-ray or SZ effect-derived masses, as it accounts for the fractional difference between the true mass and the cluster mass inferred using a gas proxy, assuming hydrostatic equilibrium. This bias is estimated from simulations to be between 10% and 20% (Lau et al., 2009, 2013; Biffi et al., 2016; Scheck et al., 2023). In simpler terms, ICM-based mass estimates tend to underestimate the actual mass by approximately a factor of 0.8-0.9, which closely matches the shift observed in the median of our mass ratios being $(1 - b) = 0.87$.

Furthermore, a majority of the clusters closely align with the median, and approximately half of them fall within the shaded region, which corresponds to the dispersion seen in the Compton-y – M_{500} scaling relation.

We also notice that Fig. 4.5a shows a greater dispersion among clusters matched at smaller relative distances, whereas the dispersion tends to be more confined to the proximity of the shaded region for clusters matched at larger relative distances. This observation highlights a potential bias in our matching process.

Our matching process relies in its first steps on positions and masses. In the search process, sometimes we find a very massive galaxy cluster very close to the target cluster's observed position. This replica candidate has a high significance in terms of its mass as the relative distance between the observed and simulated cluster is small, pointing to the fact that its existence is very probably due to our constraints. Usually, after the complete selection procedure, that cluster will be selected as the corresponding cluster replica.

However, if we fail to identify a suitable counterpart in close proximity to the observed cluster, we expand our search radius. This broader search increases the pool of potential candidates for a match. Therefore, the probability of finding a candidate with a very similar mass to the observed

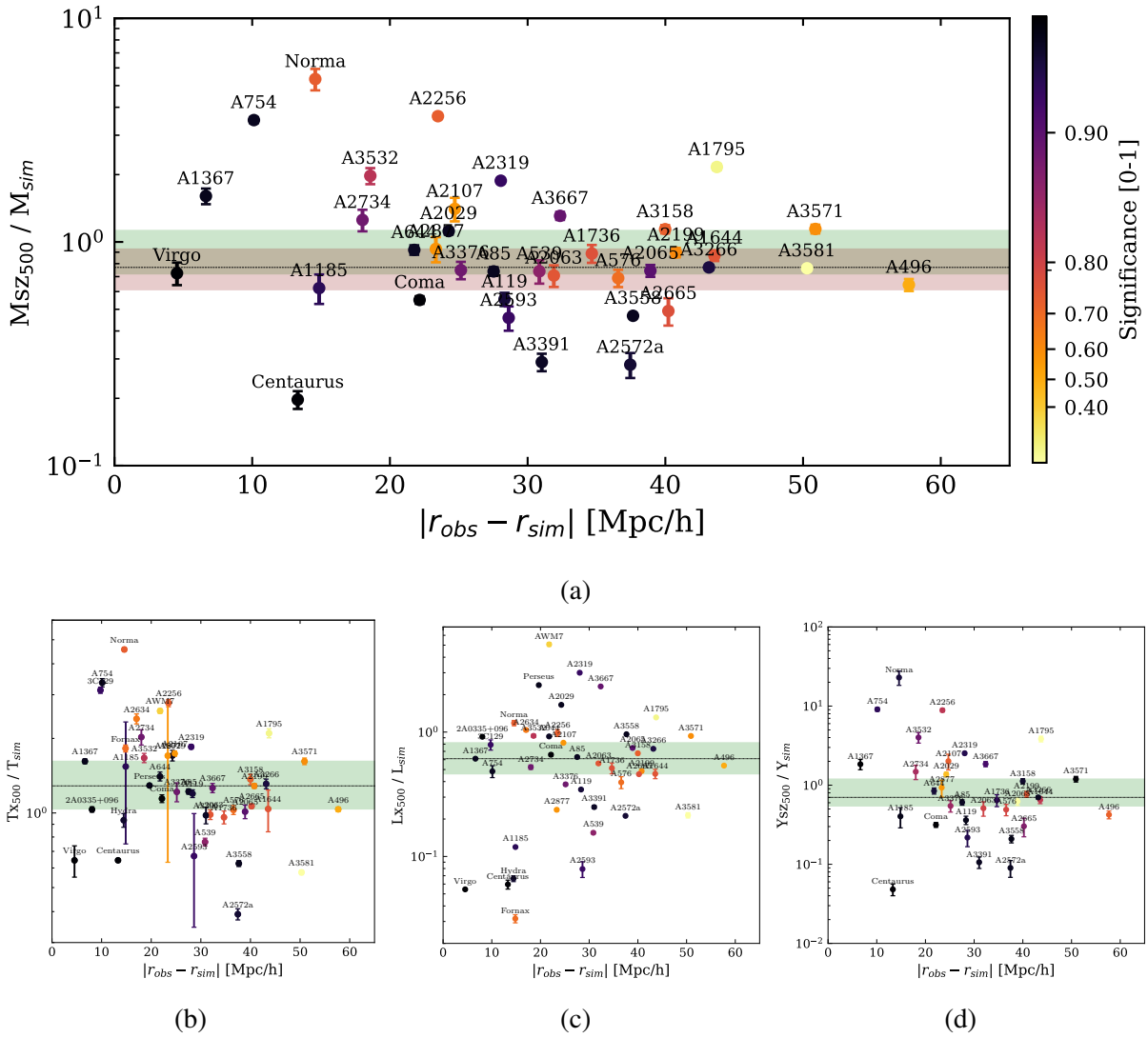


Figure 4.5: Simulated and observed quantities' ratios as a function of relative distance. (a) Ratio between the M_{500} mass, derived from the cluster SZ signal, and the SUBFIND-estimated M_{500} as a function of relative distance. The dotted line corresponds to the median of the distribution, while the shaded red region depicts the observed dispersion in the Compton-y–mass scaling relations (Planck Collaboration et al., 2011). The three bottom panels present similar ratios for (b) X-ray luminosity (L_{500}), (c) X-ray temperature (T_{500}), (d) and SZ-signal (Y_{500}). These data points are color-coded based on their mass significance, with higher significance levels indicated in black and lower significance levels denoted in yellow.

one becomes higher at these extended distances. Thus, the mass scatter becomes lower for these clusters. These trends are only slightly reflected for temperatures, luminosities, and Compton-y (see Figs. 4.5b,c,d).

In light of these results, we would like to discuss the replication quality of the subset of extensively studied clusters in the literature listed in Table 4.3.

Coma demonstrates excellent agreement in terms of mass. Its simulated mass of 9.61 $\times 10^{14}$ $M_\odot$ aligns well with the X-ray derived mass of $9.95^{+2.10}_{-2.99}$ $\times 10^{14}$ $M_\odot$, and it is just a bit higher than the SZ-derived mass. The simulated temperature of 7.06 keV also falls within the observed X-ray temperature range of 8.07 ± 0.29 keV. However, its luminosity and Compton-y values are higher by a factor of 1.5 and 3, respectively.

The Perseus cluster's simulated mass of 5.17 $\times 10^{14}$ $M_\odot$ is also in excellent agreement with its X-ray estimated mass of $6.08^{+1.55}_{-2.85}$ $M_\odot$. The simulated temperature of 4.97 keV closely matches the reported X-ray temperature of 6.42 ± 0.06 keV.

Virgo with a simulated M_{500} mass of 6.57 $\times 10^{14}$ $M_\odot$ and a virial mass of 7.12 $\times 10^{14}$ $M_\odot$, has a higher mass than the SZ estimated mass of 4.76± 0.55 $\times 10^{14}$ $M_\odot$. However, considering various studies over recent years estimating Virgo's virial mass in the range of 6.0 - 8.0 $\times 10^{14}$ $M_\odot$, like de Vaucouleurs (1960), Karachentsev et al. (2014) at 8.0 ± 2.3 $\times 10^{14}$ $M_\odot$, and Kashibadze et al. (2020) at 6.3 ± 0.9 $\times 10^{14}$ $M_\odot$, (see table 3 in Lebeau et al., 2023, for a longer list), we still view the Virgo cluster's mass in the simulation as accurately reproduced.

Hydra also agrees in its mass estimation, with 2.64 $\times 10^{14} M_\odot$ in the simulation and $1.29^{+0.44}_{-0.55}$ $\times 10^{14} M_\odot$ given by the X-ray observation. The temperature value lies also very close, from 3.37 keV in the simulation to 3.15 ± 0.05 keV from observations. It is important to note that these clusters have different dynamical states and likely underwent different formation paths, as discussed in Sect. 4.1.1

Clusters' replicas for Centaurus, AWM7, Fornax, and Norma deviate more from the observed quantities than the previously mentioned cluster replica. One key factor contributing to these differences in replication quality is the cluster's position in the sky. If we examine the positions of Centaurus, AWM7, Fornax, and Norma, we find that they are located in close proximity to the "Zone of Avoidance" (as seen in Fig. 4.1), where the extinction is severe. Consequently, the quality and quantity of observational data available for our constraints can be more limited. Perseus and Hydra, on the other hand, are better reproduced, even though they are near the Zone of Avoidance demonstrating the power of using peculiar velocities, which by tracing the potential can put constraints far beyond their actual sampling points.

Compared to other simulations of the local volume constrained with galaxy distributions such as SIBELIUS-DARK (McAlpine et al., 2022), we observe that the significant improvement in the replication of well-studied clusters like Perseus and Virgo (Sorce, 2018) can now be extended to various other local clusters. For instance, Perseus, with an observed X-ray-derived M_{500} of 6.08 $\times$ 10^{14} $M_\odot$, was previously simulated with a mass of 1.87 $\times$ 10^{15} $M_\odot$ in SIBELIUS-DARK, whereas our replication yielded a closer mass of 5.17 $\times$ 10^{14} $M_\odot$. Similarly for Virgo, its 6.57 $\times$ 10^{14} $M_\odot$ mass in our simulation lies in the range of 6 – 8 $\times$ 10^{14} $M_\odot$ reported in the previous mentioned studies by de Vaucouleurs (1960), Karachentsev et al. (2014) and Kashibadze et al. (2020), while SIBELIUS-DARK contained a replica with a lower mass of 2.7 $\times$ 10^{14} $M_\odot$ in R_{500}

and 3.5×10^{14} $M_\odot$ in R_{200}. Hydra replica is now also better reproduced in terms of mass, as we find a cluster replica with a mass of 2.6×10^{14} $M_\odot$, lying very well within the observed X-ray derived mass range of $2.50^{+0.62}_{-1.02} \times 10^{14}$ $M_\odot$, in contrast to the 3.5×10^{14} $M_\odot$ mass value reported in SIBELIUS-DARK (see Table B1 in McAlpine et al., 2022, for comparison). Indeed, when having a closer look at the quantity ratios in Fig. 4.5, we see that the general mass scatter is low, pointing to a generally good cluster reproduction.

In this context, the cases of Norma and Centaurus are particularly intriguing, as they stand out as outliers in the panels depicted in Fig. 4.5. Specifically, in the case of Norma, the simulated mass is approximately six times lower than the observed mass. As mentioned earlier, the proximity of the Norma cluster to the zone of avoidance and its association with the "Great Attractor" result in poorly constrained information about its mass and nature (Lynden-Bell et al., 1988; Woudt et al., 2007). In addition, Norma exhibits an elongated structure in both observations (as seen in Woudt et al., 2007) and our simulation. This characteristic may indicate a highly dynamic, accreting structure, further complicating the estimation of its true mass.

Norma shares its region of the sky with Hydra, Centaurus, and the Shapley clusters. Hydra and Centaurus form a cluster pair in the southern sky, positioned closer to us than Norma. While these two clusters are less massive, they are believed to play a central role in shaping the 'Great Attractor' structure (Raychaudhury, 1989). The challenge lies in their lower mass and their sky location, making it difficult to obtain high-quality observational data for accurate replication.

However, it is worth noting that the mass of Hydra is well reproduced, lying within the error region estimated from X-ray observations. The Centaurus cluster replica lies in the high mass end of the observed masses for this cluster. Indeed, its mass is 1.18 times higher than the dynamically estimated mass. Both structures if observed combined, constitute a good replica of the Centaurus - Hydra pair. When considering the observed X-ray mass for Hydra (2.5×10^{14} $M_\odot$) and the SZ and dynamical derived masses for Centaurus (1.23×10^{14} $M_\odot$ and 5.28×10^{14} $M_\odot$), the total pair mass falls within the range of 3.73×10^{14} to 7.78×10^{14} $M_\odot$. In our simulation, the combined mass of both clusters amounts to 8.85×10^{14} $M_\odot$, which lies only slightly higher than the observed range. Additionally, Fig. 4.6 depicts a density map of the Centaurus-Hydra region, with overlaid positions of observed galaxies. Galaxies directly assigned to the Hydra and Centaurus clusters are marked in red, while those in the surrounding region are in orange. This figure illustrates our ability to accurately reproduce the relative positions of Centaurus and Hydra, as well as the structure of the surrounding cosmic web. Even more, evolving the local Universe forward in time as presented by Seidel et al. (2024), shows that the Centaurus-Hydra pair merge. This indicates that although the individual clusters are less well replicated, the overall collapsing structure associated with them is well reproduced.

In conclusion, our simulations provide valuable insights into the properties of numerous key local galaxy clusters. However, there remains room for improvement, especially in accurately replicating even more clusters as well as structures situated near the zone of avoidance, where observational data is limited and replication could possibly be improved by the selection of certain, random realizations.

Improving the precision of these simulations poses several challenges: acquiring more comprehensive data, including currently sparsely sampled regions, refining the galaxy formation physics treatment within the simulations, as well as improving observational mass estimates for

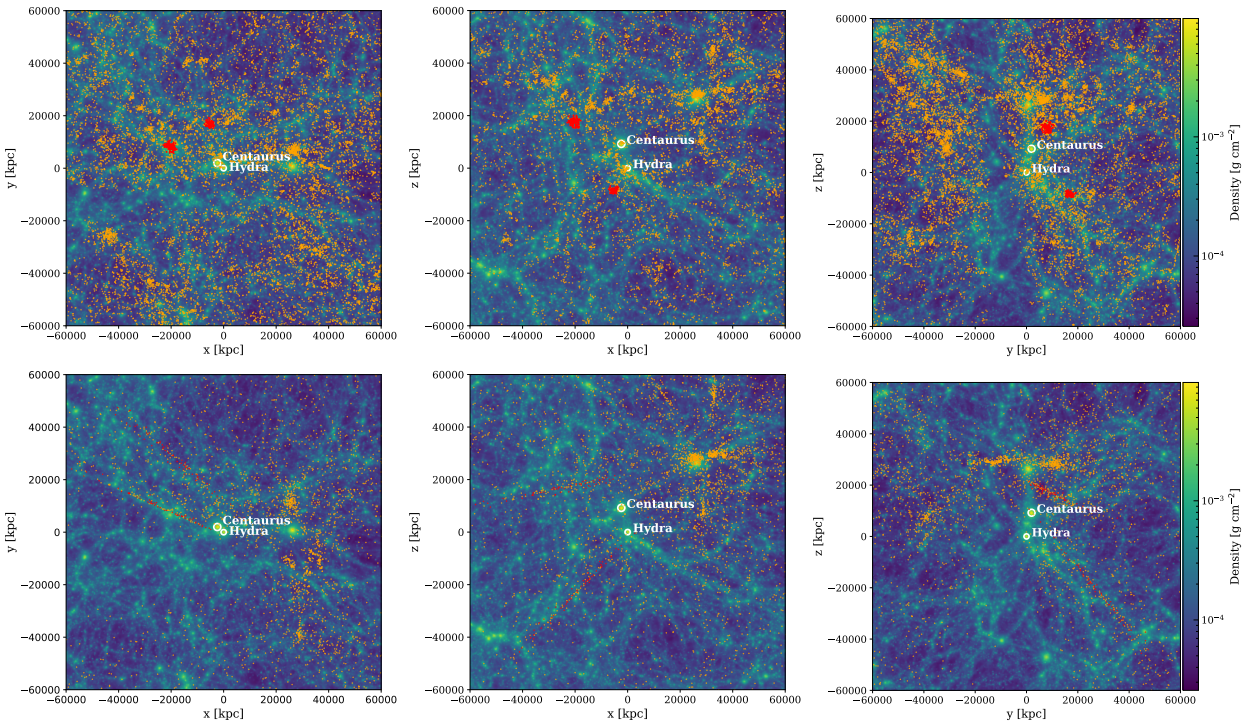


Figure 4.6: Density map of the Centaurus-Hydra region in the SLOW simulation, presented in supergalactic coordinates and created using SMAC. The locations of the Centaurus and Hydra clusters are marked with two white circles, each indicating the R_{500} radius of the respective structures. In the top panels, observed galaxies from the 2MRS1175 North Groups catalog associated with Centaurus and Hydra are highlighted in red. Additionally, galaxies listed in both the 2MRS1175 North and South Groups catalogs are shown in orange. These galaxy positions have been corrected to account for distortions such as the "fingers of god" effect. In the bottom panels, all galaxies from the CF2 Catalog in this area are displayed in orange, with those attributed to Centaurus and Hydra in red, without adjusting for the "fingers of god" correction.

better evaluation of similarity between simulated and observed clusters. Such improvements will be essential for gaining a deeper understanding of these complex sky regions.

4.3.3 Observational tracers' reliability

As previously discussed in the preceding sections, it is important to acknowledge the presence of observational errors and biases, which can at times be challenging to precisely quantify. One of our central observational parameters is mass estimations. It is noteworthy that these different mass estimation approaches yield significant discrepancies in their mass assessments for various clusters.

A potential source of bias in mass estimation appears to be the distance of the clusters from the observer. To explore this possibility, we conducted tests to assess the behavior of each mass estimation method in relation to distance. Initially, we conducted this analysis with respect to the simulated masses. Figure 4.7 illustrates the relationship between observed mass estimates and simulated masses in relation to distance. We categorized the clusters into two sets, one with closer members and the other with more distant members, each with an equal number of clusters. The median of both sets is represented by a green dashed line, while the green shaded region delineates the range encompassing 50% of the clusters above and below the median.

Masses derived from X-ray emissions (Panel 4.7b) and SZ (Panel 4.7c) exhibit minimal scatter around the median for both the closer and more distant cluster sets. The scatter shows no significant increase or decrease with cluster distance in these cases. In contrast, the scenario is quite different for dynamical masses (Panel 4.7a), where the scatter notably increases for clusters at greater distances.

Furthermore, we conducted comparative analyses among the various mass estimation methods while considering their relationship with distance, as depicted in Figs. 4.7 d, e, and f. When comparing dynamical masses against X-ray and SZ derived mass estimations, dynamical masses consistently display a high scatter, which becomes even more pronounced for clusters located farther away, mirroring the pattern observed in Fig. 4.7a. Conversely, the scatter between X-ray and SZ derived mass estimations is significantly low, as evident in panel 4.7f. Nevertheless, the scatter between different observed mass estimates is as large as the scatter between simulated and observed cluster counterparts. Taken together, this points to the fact that the level at which the constrained simulation reproduces the mass of the local galaxy clusters is at the level of the uncertainties in the observational mass estimates.

4.4 Summary and Conclusions

In this study, we have presented the first results on cluster identification in our constrained, hydrodynamical cosmological simulation of the local Universe (SLOW), which includes cooling, star formation, and the evolution of super-massive black holes. We compiled a dataset of over 221 observed galaxy clusters and groups within the local volume covered by our simulation from existing literature, collecting X-ray luminosities, temperatures, and Compton-y signals as well as inferred masses of these systems, when available. From this sample, we cross-identified 46 halos

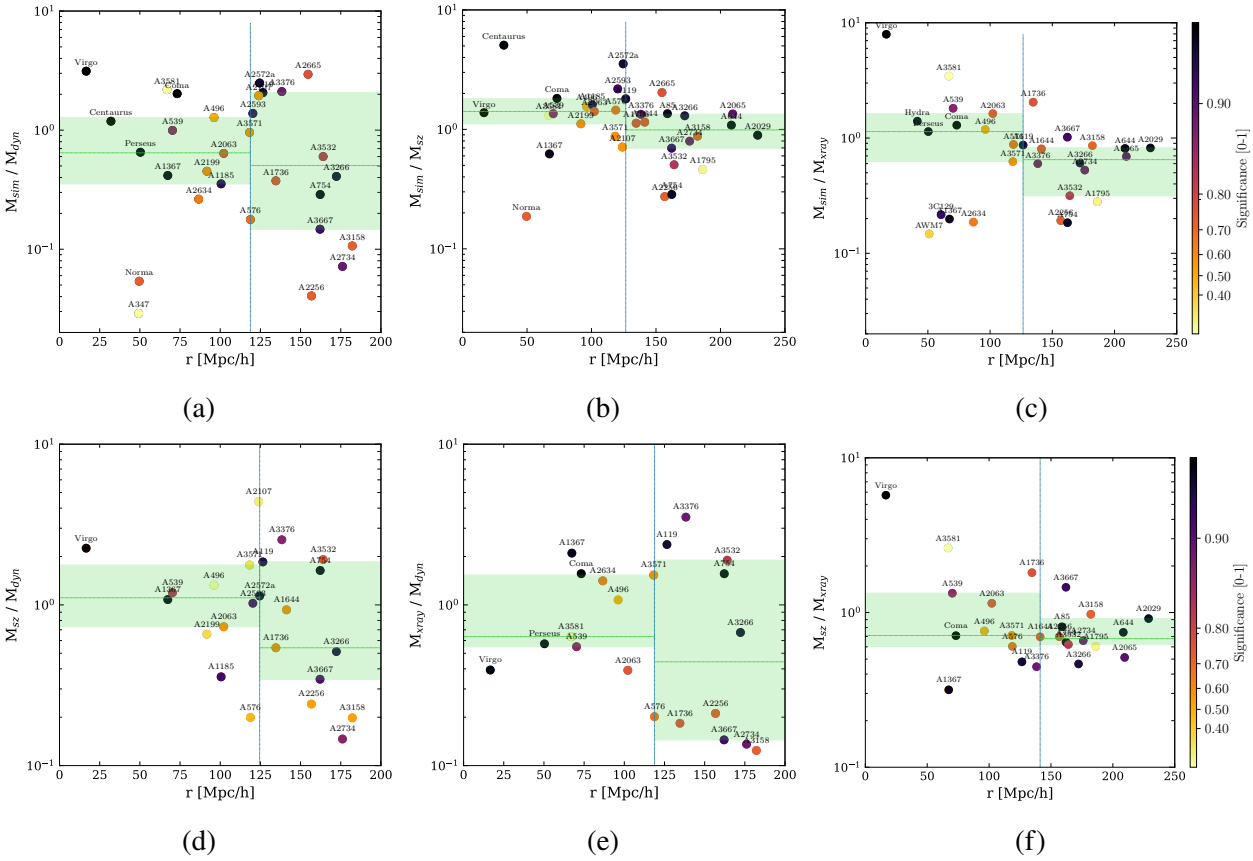


Figure 4.7: Mass ratios as a function of total distance. The top figures present the mass ratio between SUBFIND-estimated M_{500} and three different mass estimations: (a) M_{500} derived from the dynamical mass as provided in the Tully galaxy group catalog (Tully, 2015): $M_{\rm dyn}$, (b) M_{500} derived from SZ measurements as available in the Planck database: $M_{\rm SZ}$, and (c) M_{500} derived from X-ray observations: $M_{\rm xray}$. These ratios are plotted against the total distance of the cluster. In addition, the bottom figures illustrate the ratios between various observational mass estimations: (d) $M_{\rm sz}/M_{\rm dyn}$, (e) $M_{\rm xray}/M_{\rm dyn}$, and (f) $M_{\rm sz}/M_{\rm xray}$, all in relation to the total distance of the cluster. (For further details on the observed quantities, please refer to table 4.3). A vertical dashed line divides the clusters into two sets with an equal number of clusters. The green horizontal lines represent the median values for each set, while the green shaded regions indicate the range that encompasses 50% of the clusters above and below the median.

in the simulation, which are candidates to represent the according galaxy clusters within the local Universe. We then computed the significance of these associations based on different observables and compared the global properties between the observed and simulated counterparts.

1. The galaxy clusters within the simulation generally follow the observed scaling relations between mass, temperature, X-ray luminosity, and SZ signal. This has allowed us, for the first time, to evaluate the cross-identification between simulations and observations, not only on the total mass but also directly on the full bandwidth of observational signals from the ICM. Thereby, such direct comparisons avoid the various biases inherited within the different methods to determine the total mass of galaxy clusters from observational proxies.

2. We compared the positions of cross-identified galaxy clusters between observations and simulations, combined with the match of their various global properties (e.g., mass, total X-ray luminosity, temperature, or SZ signal) and computed the probability of finding a similar match in a random position within a large volume simulation. From this, we computed the significance of the match against the null hypothesis obtained from a random simulation volume.

3. This sample of clusters encompasses the 13 most massive clusters with a M_{500} exceeding $2 \times 10^{14} M_{\odot}$, as identified by the SZ signal from Planck observations within a radius of approximately 300 Mpc. This highlights the success of our constrained simulation in precisely replicating the local galaxy clusters.

4. Based on the assessment of the significance of a match, only a small fraction cannot be distinguished from a random selection. On the contrary, even 18 of the matched clusters exhibited significance values exceeding 0.8 across all studied quantities. Notably, Virgo, Coma, and Perseus achieved a significance very close to 1.0 for all quantities, with the smallest value being 0.98.

5. Compared to other constrained simulations, we report a significant improvement in the number of identified replicas of extensively-studied galaxy clusters in the local Universe, where masses now agree with their observational counterparts. Even for very prominent clusters such as Perseus, Virgo, and Hydra, we noticed that compared to simulations where the constraints are based on galaxy distributions, such as SIBELIUS-DARK, the mass is recovered better by the SLOW simulation.

6. We also investigated the impact of the observational distance uncertainty on our comparison. For the matched sample of galaxy clusters, the inferred displacement between the position in our simulation and the cluster position based on the observed redshift often reflects the uncertainties induced by the unknown peculiar velocity of the clusters. In the majority of cases, this displacement is smaller than the typical error in the distance modulus of the individual galaxies used for constructing the constraints for the ICs. This assessment further demonstrates that our constrained simulation of the local Universe reproduces a large number of galaxy clusters that can be directly compared with their observational counterpart.

7. Comparing the X-ray luminosity and temperature between observations and the simulated counterparts, the obtained width of the relative scatter is 1.7 and 1.5, respectively, for 68% of the clusters. For the SZ signal, this scatter is of a factor 2.1. Interestingly, when comparing the sample median of the M_{500} obtained from the simulations with the observed one inferred from the SZ signal, we would deduce a bias of $(1 - b) = 0.87$, which is in the range predicted by other simulations.

8. With our sample of cross-identified clusters, we compared the masses of the simulated and observed clusters using mass estimates based on different observables (this being: the SZ derived mass, dynamical masses from Tully's galaxy catalog, and X-ray signal derived masses). We find that the scatter between different observed mass estimates is as large as the scatter between simulated and observed cluster counterparts. We neither find any strong correlation of the scatter with the distance of the cluster, nor with the significance assessment, except a tendency of the observational-inferred dynamical mass to have a higher scatter at larger distances with respect to other measurements as well as with respect to the simulations. Overall, this indicates that the level at which the constrained simulation reproduces the mass of the local galaxy clusters is at the level of the uncertainties of the observational mass estimates.

Overall, the current SLOW simulation of the local Universe faithfully replicates numerous fundamental characteristics for a sizable number of galaxy clusters within our local neighborhood. This is a first step in establishing an exciting avenue for cluster and cosmological research, where such simulations enable us to investigate the formation and evolution of individual galaxy clusters. This will allow us to study the physical processes governing the formation and evolution of galaxy clusters beyond the limitations inherited when using averaged populations toward marking the transition to test the effect of individual formation histories and environments.

Table 4.1: Positions, distance uncertainties, relative distance, and significances for our selected set of galaxy clusters. The observed positions are sourced from the Tully galaxy catalog and are represented in supergalactic coordinates. In cases where Tully's positions were unavailable, we extracted coordinates from NED † and subsequently transformed them into supergalactic coordinates. These clusters are denoted by a cross symbol. The distance uncertainties were determined through the consideration of peculiar velocity uncertainties, as elaborated upon in Sect.4.3.1 for further details. Relative distance indicates the disparity between the estimated observational position and the position of the corresponding cluster in our simulation. Significances are assigned values ranging from 0 to 1, where higher numbers signify greater significance (i.e., a higher probability of deviation from a random simulation).

Name	Simulations			Observations			Distance Uncertainty	Relative Distance	Significance			
	SGX	SGY	SGZ	SGX	SGY	SGZ			Lx_{500}	Tx_{500}	Ysz_{500}	M_{500}
	[Mpc / h]						[Mpc / h]	[Mpc / h]	(0 – 1)			
A85	50.84	-143.10	25.88	46.39	-155.60	1.76	15	27.53	0.99	0.96	0.98	0.98
A119	73.61	-117.28	-24.93	57.59	-119.46	-1.67	15	28.32	0.98	0.98	0.98	0.97
A347	60.29	-2.44	-19.23	52.69	-16.38	-5.28	18	21.14	1.00	0.67	0.93	0.20
A496	-13.28	-94.52	-83.46	24.79	-51.18	-83.46	17	57.69	0.69	0.61	0.41	0.50
A539	54.47	8.98	-82.08	43.81	-18.63	-73.38	17	30.85	0.95	0.80	0.93	0.86
A576	120.30	39.98	-14.59	98.80	57.53	-38.41	17	36.57	0.70	0.60	0.74	0.69
A644	-26.48	70.51	-204.14	-9.74	83.20	-209.86	15	21.77	0.99	0.96	0.98	0.98
A754	-34.31	78.48	-144.27	-32.44	88.40	-143.56	14	10.12	0.98	0.97	0.99	0.98
A1185	24.59	112.10	-27.24	16.31	99.93	-25.16	18	14.86	0.98	0.88	0.96	0.94
A1367	-2.89	71.35	-18.73	-2.94	68.62	-12.68	17	6.64	0.99	0.98	0.99	0.98
A1644	-97.26	64.92	-5.74	-102.00	107.93	-10.76	15	43.57	0.93	0.70	0.85	0.76
A1736	-137.66	65.91	-22.87	-117.47	83.34	-0.73	17	34.66	0.88	0.67	0.85	0.77
A1795 †	6.26	201.72	29.19	-9.57	174.19	59.27	15	43.74	0.30	0.15	0.35	0.25
A2029 †	-92.61	142.37	141.38	-94.65	162.38	127.79	14	24.27	0.99	0.95	0.98	0.98
A2063	-42.47	66.58	97.18	-39.98	76.20	66.87	17	31.90	0.70	0.69	0.80	0.75
A2065 †	7.25	138.68	107.11	-9.07	161.56	134.02	15	38.91	0.94	0.89	0.94	0.90
A2107	-28.27	71.39	83.16	-18.25	92.93	89.98	17	24.71	0.73	0.51	0.60	0.0.57
A2197/99	-12.99	48.75	97.11	18.31	57.56	72.51	17	40.77	0.94	0.35	0.69	0.61
A2256	115.91	96.33	79.47	137.38	86.90	78.14	14	23.49	0.82	0.62	0.80	0.73
A2319	81.29	2.38	118.30	72.21	21.34	136.87	12	28.05	0.99	0.93	0.98	0.93
A2572a	67.40	-88.63	16.82	66.25	-80.64	54.83	17	38.86	0.96	0.97	0.90	0.96
A2593	54.51	-107.62	32.14	66.54	-93.66	54.04	17	28.62	0.96	0.91	0.92	0.93

† NED Database: https://ned.ipac.caltech.edu

* For close-by systems such as Virgo we expect the error from direct distance measurements to be substantially smaller.

Table 4.2: Table 4.1. continued.

Name	Simulations			Observations			Distance Uncertainty	Relative Distance	Significance			
	SGX	SGY	SGZ	SGX	SGY	SGZ			Lx_{500}	Tx_{500}	Ysz_{500}	M_{500}
	[Mpc / h]						[Mpc / h]	[Mpc / h]	(0 – 1)			
A2634	80.38	-52.34	32.13	64.39	-54.96	37.40	18	17.04	0.67	0.60	0.94	0.65
A2665	111.79	-123.00	44.66	75.78	-140.15	49.86	17	40.22	0.90	0.51	0.34	0.77
A2734	-20.57	-186.09	21.73	-13.99	-188.50	5.14	17	18.01	0.96	0.80	0.80	0.88
A2877 †	-8.07	-76.12	-0.78	-21.18	-65.46	-16.88	18	23.34	0.54	0.50	0.96	0.57
A3158	-49.26	-108.90	-119.25	-79.02	-135.57	-117.83	15	39.99	0.95	0.60	0.93	0.72
A3266	-59.95	-93.17	-128.12	-97.79	-105.76	-111.61	14	43.17	0.97	0.90	0.65	0.96
A3376	-36.34	-74.00	-124.29	-41.49	-50.16	-130.44	17	25.15	0.93	0.82	0.78	0.88
A3391 / 95	-53.30	-43.16	-101.67	-78.28	-51.37	-118.14	17	31.03	0.99	0.97	0.46	0.97
A3532	-156.71	96.63	-36.03	-143.78	96.68	-22.69	15	18.58	0.87	0.82	0.98	0.82
A3558	-141.91	66.10	-37.17	-130.29	78.7	-3.64	15	37.66	0.98	0.99	0.85	0.98
A3571	-142.64	54.10	-33.88	-109.39	58.57	4.39	15	50.90	0.89	0.45	0.78	0.60
A3581	-84.87	62.76	-23.92	-58.22	36.91	10.01	17	50.29	0.09	0.10	0.90	0.20
A3667	-156.83	-111.30	66.72	-133.15	-101.45	47.00	14	32.36	0.71	0.92	0.39	0.90
2A0335+096	74.52	-60.85	-58.17	70.15	-55.59	-62.44	18	8.07	0.99	0.99	0.98	1.00
3C129 †	55.55	11.33	-28.56	56.58	2.80	-23.92	18	9.76	0.97	0.92	0.99	0.93
AWM7 †	60.08	-6.71	-23.32	45.01	-11.89	-8.47	18	21.78	0.26	0.34	0.93	0.38
Centaurus/A3526	-22.82	11.21	-13.26	-34.25	14.93	-7.56	18	13.31	1.00	1.00	0.88	1.00
Fornax/AS0373	4.47	-6.59	-21.52	-1.69	-10.74	-8.69	19	14.83	0.90	0.53	0.98	0.70
Hydra/A1060†	-21.21	9.89	-19.41	-25.63	21.99	-25.97	18	14.45	1.00	0.95	0.27	0.97
Norma/A3627	-45.65	-0.38	-5.67	-50.26	-7.06	6.44	17	14.58	0.94	0.66	0.99	0.73
Perseus/A426	59.22	2.13	-24.58	49.94	-10.73	-12.98	18	19.65	0.99	0.98	1.00	0.98
Coma/A1656	-2.93	82.27	-9.14	0.48	72.79	10.59	14	22.15	1.00	1.00	1.00	1.00
Virgo	-3.60	10.36	-1.64	-3.48	14.86	-2.21	16 *	4.54	1.00	1.00	1.00	1.00

† NED Database: https://ned.ipac.caltech.edu

* For close-by systems such as Virgo we expect the error from direct distance measurements to be substantially smaller.

Table 4.3: X-ray luminosities (L_{500}) and temperatures (T_{500}) were primarily obtained from Ikebe et al. (2002). If available we also used more recent data from Shang & Scharf (2008)‡ and Planck Collaboration et al. (2011)*. For clusters not included in these studies, the corresponding data were extracted from the BAX database†. The SZ-signal (Ysz_{500}) and its derived mass (Msz_{500}) were sourced from the Planck database PSZ1v2. X-ray masses (Mx_{500}) were extracted from Chen et al. (2007), correcting for the differences in the value of H_0 assumed. Dynamical masses were retrieved from the Tully galaxy catalog and subsequently transformed to M_{500} by correcting for the 12% factor presented by Sorce et al. (2016b) and the conversion method outlined by Ragagnin et al. (2021). Simulated values for L_{500}, and Ysz_{500} were estimated using SMAC, while the simulated M_{500} and T_{500} were estimated using SUBFIND.

Name	Obs	Sim	Obs	Sim	Obs	Sim	Obs			Sim
	Lx_{500}	Lx_{500}	Tx_{500}	Tx_{500}	Ysz_{500}	Ysz_{500}	Mx_{500}	Msz_{500}	$Mdyn_{500}$	M_{500}
	[10^{44} erg s^{-1}] (band: 0.1-2.4 keV)		[KeV] (band: 0.1-2.4 keV)		[10^{-4} Mpc^2]		[$10^{14} \times M_\odot$]			
A85	5.08 ± 0.07	8.06	$6.51^{+0.16}_{-0.23}$	5.33	0.68 ± 0.05	1.12	$6.06^{+1.18}_{-2.63}$	4.90 ± 0.21	–	6.64
A119	1.66 ± 0.028 *	4.81	$6.69^{+0.24}_{-0.28}$	5.59	0.33 ± 0.04	0.92	$6.73^{+0.90}_{-1.94}$	3.34 ± 0.22	2.83	5.82
A347	–	7.86	–	0.40	–	0.004	–	–	3.19	9.16
A496	2.19 ± 0.026	4.07	4.59 ± 0.10	4.44	0.23 ± 0.03	0.54	$3.61^{+0.67}_{-1.58}$	2.74 ± 0.17	3.36	4.27
A539	0.64 ± 0.013	4.13	$3.04^{+0.11}_{-0.10}$	3.98	–	2.32	$2.01^{+0.24}_{-0.64}$	–	–	3.63
A576	1.09 ± 0.12	2.77	3.83 ± 0.16	3.72	0.14 ± 0.02	0.13	$3.46^{+2.44}_{-1.79}$	2.09 ± 0.19	17.15	3.03
A644	5.08 ± 0.09	5.51	6.54 ± 0.27	4.66	0.64 ± 0.06	0.76	$6.31^{+1.61}_{-2.86}$	4.70 ± 0.23	–	5.10
A754	2.36 ± 0.07	2.25	9.00 ± 0.35	2.68	1.17 ± 0.06	0.13	$10.39^{+3.30}_{-4.82}$	6.68 ± 0.20	6.65	1.91
A1185	0.37 †	3.11	$3.9^{+2.00}_{-1.10}$ †	2.53	0.06 ± 0.02	0.14	–	1.27 ± 0.19	5.77	2.03
A1367	0.68 ± 0.009	1.11	3.55 ± 0.08	2.19	0.1 ± 0.02	0.06	$5.56^{+0.83}_{-1.78}$	1.76 ± 0.14	2.65	1.10
A1644	2.32 ± 0.2	5.02	$4.70^{+0.90}_{-0.70}$	4.53	0.43 ± 0.04	0.66	$5.50^{+3.22}_{-3.30}$	3.83 ± 0.21	6.70	4.40
A1736	1.90 ± 0.19	3.71	$3.68^{+0.22}_{-0.17}$	3.83	0.27 ± 0.05	0.41	$1.63^{+0.46}_{-0.52}$	2.95 ± 0.27	8.89	3.32
A1795	4.02 ± 0.03	1.97	$6.17^{+0.26}_{-0.25}$	2.94	0.6 ± 0.05	0.16	$7.40^{+2.89}_{-4.11}$	4.54 ± 0.21	–	2.06
A2029	10.47 ± 0.11	6.32	$7.93^{+0.39}_{-0.36}$	4.65	1.27 ± 0.08	0.93	$7.46^{+2.47}_{-3.87}$	6.82 ± 0.24	–	6.09
A2063	1.31 ± 0.027	2.34	3.56 ± 0.16	3.61	0.13 ± 0.03	0.26	$1.77^{+0.18}_{-0.44}$	2.03 ± 0.23	4.52	2.87
A2065	3.49 ± 0.022 *	4.68	5.37 ± 0.34	5.30	0.55 ± 0.06	0.89	$8.39^{+7.18}_{-5.11}$	4.30 ± 0.26	–	5.78
A2107	1.41 †	1.72	4.00 ± 0.10 †	2.26	0.12 ± 0.03	0.06	–	1.86 ± 0.22	6.85	1.33
A2197/99	2.43 ± 0.07	5.01	4.28 ± 0.10	3.33	0.23 ± 0.02	0.30	$3.22^{+0.88}_{-1.42}$	2.78 ± 0.13	6.89	3.10
A2256	2.04 ± 0.13 ‡	2.07	$6.83^{+0.23}_{-0.21}$ ‡	2.46	1.08 ± 0.05	0.12	$9.09^{+2.56}_{-3.09}$	6.34 ± 0.18	43.03	1.74
A2319	15.78 †	5.25	$8.84^{+0.18}_{-0.14}$ †	4.78	1.90 ± 0.08	0.75	–	8.59 ± 0.22	27.20	4.57
A2572a	1.02 †	4.81	2.5 ± 0.13 †	6.40	0.15 ± 0.04	1.67	–	2.14 ± 0.27	3.06	7.57

Table 4.4: Table 4.3. continued.

Name	Obs	Sim	Obs	Sim	Obs	Sim	Obs			Sim
	Lx_{500}	Lx_{500}	Tx_{500}	Tx_{500}	Ysz_{500}	Ysz_{500}	Mx_{500}	Msz_{500}	$Mdyn_{500}$	M_{500}
	[10^{44} erg s^{-1}] (band: 0.1-2.4 keV)		[KeV] (band: 0.1-2.4 keV)		[10^{-4} Mpc^2]		[$10^{14} \times M_\odot$]			
A2593	0.33 ± 0.048 ‡	4.16	3.10 ± 1.50 ‡	4.63	0.14 ± 0.03	0.66	–	2.09 ± 0.26	3.32	4.56
A2634	0.57 ± 0.016	0.55	3.45 ± 0.16	1.44	–	0.08	$3.38^{+0.50}_{-0.75}$	–	2.40	6.27
A2665	0.499 ± 0.076 ‡	4.38	3.98 ± 0.07 ‡	3.74	0.14 ± 0.04	0.45	–	2.01 ± 0.28	1.40	4.09
A2734	1.45 ± 0.058	2.78	$5.07^{+0.36}_{-0.42}$	2.51	0.19 ± 0.04	0.13	$3.61^{+0.64}_{-1.17}$	2.38 ± 0.26	26.56	1.89
A2877	0.23 ± 0.005	0.97	$3.50^{+2.20}_{-1.10}$	2.06	0.05 ± 0.01	0.05	$5.16^{+5.05}_{-2.84}$	1.12 ± 0.14	–	1.20
A3158	3.36 ± 0.09	4.98	$5.41^{+0.26}_{-0.24}$	3.96	0.52 ± 0.04	0.46	$4.31^{+0.67}_{-1.24}$	4.20 ± 0.18	34.70	3.68
A3266	5.17 ± 0.07	7.04	$7.72^{+0.35}_{-0.28}$	5.90	1.19 ± 0.06	1.70	$14.43^{+3.57}_{-5.68}$	6.71 ± 0.18	21.42	8.72
A3376	1.27 ± 0.03	3.34	$4.43^{+0.39}_{-0.38}$	3.64	0.17 ± 0.03	0.31	$5.08^{+1.16}_{-1.49}$	2.27 ± 0.20	1.45	3.03
A3391/95	1.57 ± 0.049	6.29	$5.89^{+0.45}_{-0.33}$	6.02	0.14 ± 0.02	1.32	$4.53^{+0.55}_{-1.27}$	2.04 ± 0.18	–	7.02
A3532	1.31 ± 0.038	1.40	$4.41^{+0.19}_{-0.18}$	2.64	0.3 ± 0.05	0.73	$4.97^{+0.88}_{-2.13}$	3.09 ± 0.26	2.62	1.56
A3558	6.43 †	6.69	4.78 ± 0.13 †	7.65	0.55 ± 0.06	2.63	–	4.41 ± 0.27	37.10	9.47
A3571	4.72 ± 0.07	5.07	$6.80^{+0.21}_{-0.18}$	4.21	0.6 ± 0.05	0.50	$6.57^{+1.27}_{-2.57}$	4.67 ± 0.21	4.29	4.09
A3581	0.35 ± 0.017	1.63	$2.83^{+0.04}_{-0.02}$	3.18	0.35 ± 0.017	0.90	$0.70^{+0.14}_{-0.28}$	(1.83 ± 0.04)	1.10	2.40
A3667	5.66 ± 0.07	2.43	$6.28^{+0.27}_{-0.26}$	4.98	0.9 ± 0.07	0.49	$3.96^{+0.39}_{-0.86}$	5.77 ± 0.24	27.38	4.07
2A0335	2.58 ± 0.02	2.81	$3.64^{+0.09}_{-0.08}$	3.52	–	1.34	$2.09^{+0.82}_{-1.22}$	–	–	2.78
3C129	1.27 ± 0.12	1.61	$5.57^{+0.16}_{-0.15}$	1.78	–	0.04	$4.04^{+1.69}_{-1.75}$	–	–	0.87
AWM7	1.20 ± 0.04	0.24	$3.70^{+0.08}_{-0.04}$	1.43	–	0.05	$3.69^{+0.91}_{-1.69}$	–	–	0.54
Centaurus	0.302 ± 0.025	5.07	$3.69^{+0.05}_{-0.04}$	5.73	0.05 ± 0.01	1.09	–	1.23 ± 0.11	5.28	6.24
Fornax	0.04 ± 0.003	1.26	$1.56^{+0.05}_{-0.07}$	0.85	–	0.04	$0.97^{+0.33}_{-0.41}$	–	–	0.36
Hydra	0.32 ± 0.017	4.83	3.15 ± 0.05	3.37	–	1.45	$1.87^{+0.46}_{-0.76}$	–	–	2.61
Norma	2.04 ± 0.1	1.73	$5.62^{+0.12}_{-0.11}$	1.23	0.19 ± 0.04	0.01	–	2.55 ± 0.28	8.89	0.47
Perseus	9.33 ± 0.11	3.91	6.42 ± 0.06	4.97	–	4.34	$4.56^{+1.16}_{-2.14}$	–	7.96	5.17
Coma	4.63 ± 0.11	7.04	$8.07^{+0.29}_{-0.27}$	7.06	0.73 ± 0.05	2.31	$7.46^{+1.57}_{-2.24}$	5.29 ± 0.20	4.77	9.61
Virgo	0.254 ± 0.005 ‡	4.67	$3.67^{+0.53}_{-0.52}$ ‡	5.70	–	1.08	0.83 ± 0.01 [a]	4.76± 0.55 [b]	3.43	6.57

[a] Virgo X-ray derived mass was extracted from Simionescu et al. (2017)
[b] Virgo SZ derived mass extracted from Planck Collaboration et al. (2016b)

5 | Simulating the Local Web: Thermodynamics and Evolution of Local Galaxy Clusters

The content of this chapter is based on Hernández-Martínez et al. (in prep.) and will be submitted to A&A, with some adjustments made.

5.1 Probing Galaxy Cluster Thermodynamics with Constrained Simulations

A key component of galaxy clusters is the ICM, a hot, diffuse plasma that dominates their baryonic mass. With temperatures reaching tens of millions of kelvin, the ICM emits X-rays primarily through thermal bremsstrahlung radiation, making it observable with modern X-ray telescopes. The thermodynamic properties of the ICM, including temperature, electron density, entropy, and pressure, provide critical insights into the physical processes governing cluster formation and their applications as cosmological probes (e.g., Vikhlinin et al., 2006; Eckert et al., 2013) . Understanding these properties is essential for linking observations to theoretical models of LSS growth.

The ICM's thermodynamic profiles are shaped by the interplay of gravitational processes, such as accretion and mergers, and non-gravitational mechanisms, including radiative cooling and feedback from AGN. For instance, entropy profiles reveal the balance between gravitational heating and additional energy input from non-gravitational processes, while pressure profiles are key to connecting the thermal energy of the gas to the cluster's mass (e.g., Voit, 2005; Arnaud et al., 2010). Despite significant progress, many aspects of ICM physics remain unclear, such as the precise role of AGN feedback in regulating cluster cores and the mechanisms responsible for the observed deviations from self-similar scaling, particularly in disturbed systems (e.g., McNamara & Nulsen, 2007; Planelles et al., 2014; Le Brun et al., 2014).

The dynamical state of a cluster significantly impacts its ICM profiles. Relaxed clusters, typically characterized by symmetric profiles and cool cores, exhibit stable thermodynamic properties where cooling and heating are in balance. Conversely, disturbed clusters—often believed to be the result of recent mergers—show irregular temperature and density distributions, reflecting their turbulent formation histories. The diversity in observed cluster profiles highlights the importance

of understanding the relationship between a cluster's dynamical state and the physical processes shaping its ICM (e.g., Rasia et al., 2015; Barnes et al., 2017). These complexities pose challenges for both observational studies and simulations aiming to replicate cluster behavior across a wide range of conditions.

While statistical studies of galaxy cluster populations provide valuable insights into general trends, they often obscure the unique features of individual systems. Constrained simulations offer a complementary approach, enabling the study of specific clusters (e.g., Dolag et al., 2005; Sorce et al., 2016) by reconstructing their ICs from observational data, such as galaxy density fields (e.g., Hoffman & Ribak, 1991) or peculiar velocity measurements (e.g., Sorce et al., 2016a). This approach allows for one-to-one comparisons between simulated and observed clusters, providing a powerful tool to investigate how individual formation histories influence ICM properties. Constrained simulations thus offer a unique perspective on the diversity of cluster profiles, shedding light on the physical processes driving deviations from expected behaviors.

In this study, we investigate the thermodynamic properties,temperature, electron density, entropy, and pressure, of the ICM of galaxy clusters within the local Universe simulation SLOW (Dolag et al., 2023). We analyze these properties of the hot gas on a cluster-by-cluster basis and compare them with observational data at redshift zero. Expanding on previous work by Hernández-Martínez et al. (2024), which showed that the SLOW constrained simulations accurately reproduce the integrated properties of galaxy clusters, we demonstrate that these simulations also capture the detailed interior thermodynamic structures of galaxy clusters in our Local Universe. This ability to replicate galaxy cluster profiles, which are intrinsically linked to their formation histories, suggests that the simulations can reliably predict the evolutionary trajectories of these systems. Consequently, we establish robust one-to-one connections between the formation history of individual clusters and their observed thermodynamic states at $z = 0$. By linking the simulated cluster formation processes to their present-day observed ICM properties, we aim to deeper understand the mechanisms shaping galaxy clusters and refine galaxy formation models in a cosmological context.

5.2 Observational Thermodynamic Profiles: Data Sources, Methods, and Systematics

We compiled a robust dataset of observed deprojected profiles by gathering results from the literature (e.g., Vikhlinin et al., 2006; Arnaud et al., 2010; Eckert et al., 2013; Planck Collaboration et al., 2013a; Ghirardini et al., 2019), covering up to 12 Local Universe galaxy clusters. These profiles were originally derived from X-ray observations conducted with the *Suzaku* (Mitsuda et al., 2007), *Chandra* (Weisskopf et al., 2002), and *XMM-Newton* (Jansen et al., 2001) telescopes.

The XCOP project (Eckert et al., 2017) served as a key source, offering deprojected thermodynamic profiles such as temperature, density, pressure, and entropy for seven galaxy clusters, including A644, A1644, A1795, A2029, A2319, A3158, and A3266 (Ghirardini et al., 2019). These profiles were extracted under the assumption of spherical symmetry and were constructed by combining X-ray observations from XMM-Newton with tSZ measurements from *Planck* 2015

data release (i.e. the full intensity survey, see Ettori et al. (2019) Sec. 2 for further details).

In the case of A85 we compared our replica to the data presented by Ichinohe et al. (2015). These observed deprojected thermodynamic profiles were derived using Suzaku observations under the assumption of spherical symmetry, focusing on the direction of the infall of the southern subcluster (S subcluster). Spectral data were extracted from concentric annuli centered on the cluster and fitted employing the projection model to deproject the overlapping emission along the line of sight. These profiles were compared to theoretical models (e.g., Nagai et al., 2007; Pratt et al., 2010) to validate their accuracy and extend the analysis out to the virial radius (r_{200}). Projection corrections were also applied to refine the three-dimensional distributions.

The observed deprojected thermodynamic profiles for Abell 119, were presented in Fig. 5.2 of Hudson et al. (2010). They were extracted using radial profiles of X-ray surface brightness from Chandra observations and were deprojected assuming spherical symmetry.

The deprojected thermodynamic profiles of the Coma Cluster were obtained from multiple observational datasets, including X-ray and tSZ-effect measurements. The deprojected profiles presented in Planck Collaboration et al. (2013b, Fig. 13) were derived using a combination of SZ data from the *Planck* satellite and X-ray data from an XMM-Newton mosaic.

Additionally, we include azimuthally resolved deprojected profiles presented by Simon (2013), based on Suzaku X-ray observations, allowing for sector-dependent analyses, that included the E+NE and NW+W directions, showing regional variations in the thermodynamic properties. The deprojected profiles were obtained using the onion-shell deprojection technique, assuming spherical symmetry. Furthermore, we incorporate the median deprojected pressure profile from Mirakhor & Walker (2020), together with the shaded region representing the spread across all measured sectors.

We include observations of the Perseus cluster, one of the most X-ray luminous galaxy clusters in the local Universe. The deprojected thermodynamic profiles of Perseus were extracted from Urban et al. (2013), which uses XMM-Newton observations. As usual, the deprojection analysis was performed using an onion-shell technique.

We also include observations of the Virgo cluster, the nearest massive galaxy cluster at a distance of 16 Mpc. The deprojected thermodynamic profiles of Virgo were extracted from Urban et al. (2011), which used a mosaic of XMM-Newton observations covering the cluster out to its virial radius. The study found that Virgo has a relatively shallow gas density profile, with a power-law index of $\beta = 1.21 \pm 0.12$, indicating a less concentrated gas distribution compared to more massive, relaxed clusters, and presented an entropy profile that follows a gravitational collapse-like power law ($K \propto r^{1.1}$) within 450 kpc, but flattens at larger radii, falling below theoretical expectations. As stated by Urban et al. (2011), due to the dynamically unrelaxed nature of Virgo, significant substructure is present and deviations from hydrostatic equilibrium are expected. This unrelaxed state has been independently confirmed by studies of simulated Virgo replicas (Sorce et al., 2016; Sorce et al., 2021; Lebeau et al., 2024). Thus, the observed profiles may be affected by systematic uncertainties arising from departures from spherical symmetry, which are significantly larger than the measurement errors, making direct comparisons with theoretical models challenging. Nevertheless, being aware of these complexities gives us a strong foundation to test the robustness of simulation predictions for unrelaxed clusters.

5.3 Thermodynamic Profiling of Simulated Galaxy Clusters

The simulation allows the extraction of thermodynamic properties of the ICM, such as temperature, density, entropy, and pressure, at various radii. These profiles are computed directly from the hydrodynamical outputs of the SLOW-AGN1536[3] run, the highest resolution simulation in the suite, which includes full galaxy formation physics. We focus our study on clusters for which we have detailed observational data of their deprojected profiles.

To better represent the deprojected thermodynamic profiles extracted from observations, we divided the simulated clusters into concentric bins spanning radii from 100 kpc to five times the virial radius (where the virial radius lies tipycally at $\sim$ 1 Mpc). The lower limit of 100 kpc is set by the resolution of the simulations, ensuring reliable measurements of the ICM properties. Within each bin, we calculated the mass-weighted averages for pressure, temperature, entropy, and electron density for the hot gas component ($T > 10^5$ K), following an approach with which we aim to be consistent with methodologies typically adopted in observational analyses (Nagai et al., 2007; Lau et al., 2009; Ghirardini et al., 2019).

We adopt mass weighting consistently across all thermodynamic quantities to ensure a physically self-consistent representation of the ICM's thermal energy content. Although alternative weighting schemes, such as volume weighting for density or spectroscopic-like weighting for temperature (Mazzotta et al., 2004b), may offer closer analogs to specific observational methods, we prioritize consistency and the physical interpretability of mass-weighted quantities. In particular, mass-weighted temperatures describe the actual thermal energy per unit mass, whereas emission-weighted or spectroscopic temperatures tend to be biased toward denser, cooler gas phases that dominate X-ray emission (Roncarelli et al., 2018).

The use of mass-weighted temperatures is further justified within r_{500}, where the gas distribution is more homogeneous and the impact of multiphase structures and clumping is reduced, especially in relaxed clusters. In these regions, the mass-weighted and spectroscopic temperatures generally agree within 10 to 15% (e.g., Mazzotta et al., 2004b; Rasia et al., 2005), which includes intermediate radial ranges that constitute the main focus of our analysis.

Nevertheless, we recognize that mass-weighted temperatures may overestimate observationally derived temperatures in more complex environments. Larger discrepancies are expected in non-virialised clusters, where substructures and multiphase gas become more prominent, as well as in the innermost regions ($r \lesssim 0.1 r_{500}$) where cool cores can bias X-ray measurements towards lower values (Markevitch, 1998; Vikhlinin et al., 2005). Similarly, in the outskirts (beyond r_{500}), clumping and non-equilibrium effects introduce additional uncertainties.

An important distinction between our simulation-based profiles and observational studies is that we do not mask substructures in our analysis. X-ray observational studies typically excise or mask substructures, such as infalling clumps and merging sub-halos, to minimize contamination in the derived profiles (Vikhlinin et al., 2005; Pratt et al., 2007; Eckert et al., 2015). In contrast, by including all gas phases and substructures, our profiles capture the full complexity of the ICM but also introduce greater scatter, particularly in the outskirts and dynamically active systems. To account for this, we present both mean and median profiles. The median provides a robust characterization of the bulk ICM, being less sensitive to high-density outliers — a desirable

feature given that the distributions of thermodynamic quantities such as pressure and density are often approximately log-normal. In such skewed distributions, the median is generally more representative of the typical value than the mean, which can be significantly biased by dense clumps and shocks. The mean, on the other hand, incorporates the contribution from dense clumps and shocks, providing valuable insight into the total gas distribution and the level of gas inhomogeneity. The difference between the mean and median serves as a useful diagnostic of the cluster's dynamical state and degree of clumping (Vazza et al., 2013; Zhuravleva et al., 2013).

Despite its simplicity, this approach enables a direct comparison between simulated and observed deprojected profiles. By focusing on the simulated counterparts of observed clusters, we can directly compare absolute values and radial distances, without the need to scale by R_{500}. This strategy makes our comparison less sensitive to errors induced by the hydrostatic mass bias. When observational profiles are scaled by r_{500} or M_{500}, the derived scaling radius typically assumes hydrostatic equilibrium; if non-thermal pressure support or bulk motions are significant, this can lead to an underestimation of r_{500} and consequently shift the scaled profiles. By comparing profiles in physical units (e.g., absolute distances in kpc), we reduce sensitivity to these biases, although some implicit effects may remain through observational sample selection and analysis choices.[1]

An important consideration in our comparison between simulated and observational profiles is the cosmological model adopted in each case. The observational datasets employed in this work are based on varying cosmological assumptions, which can introduce systematic differences in key quantities such as radius, electron density, pressure, and entropy. While our simulations adopt the *Planck* cosmology with $h = 0.67$, $\Omega_M = 0.315$, and $\Omega_\Lambda = 0.685$, several observational studies assume a flat ΛCDM cosmology with $H_0 = 70$ km s^{-1} Mpc^{-1} ($h = 0.7$), $\Omega_M = 0.3$, and $\Omega_\Lambda = 0.7$. To ensure consistency, we rescale observational quantities that are sensitive to the underlying cosmology. Physical radii scale inversely with the Hubble parameter, $R \propto h^{-1}$, resulting in an increase of approximately 4.5% when converting observational radii from $h = 0.7$ to $h = 0.67$. Electron densities, derived from X-ray surface brightness profiles and dependent on the angular diameter distance, scale as $n_e \propto h^{1/2}$, decreasing by about 2.3% in this conversion. Pressure, defined as $P_e = n_e k_B T$, follows the same scaling as density, $P_e \propto h^{1/2}$, since X-ray temperatures are cosmology-independent. Entropy, given by $K \propto T n_e^{-2/3}$, rescales as $K \propto h^{-1/3}$.

We note that the XCOP sample adopts the same *Planck* cosmology as our simulations. Thus, no rescaling is necessary for the XCOP profiles, which can be directly compared to our simulated clusters. For all other observational data sets, we apply the corresponding cosmology rescalings to enable a consistent and meaningful comparison of the thermodynamic profiles and radial distances.

The constrained nature of the simulations enables a direct connection between the thermodynamic properties of the ICM and the formation histories and environments of individual galaxy clusters. This allows us to explore how cluster dynamics and large-scale environmental factors influence the diversity of ICM profiles observed at $z = 0$. By comparing simulated analogs to well-studied observed clusters, we gain valuable insights into the role of mergers, accretion,

[1]Note that the masses from the simulation, when compared to the observationally inferred ones, are compatible with a hydrostatic bias of $(1 - b) \approx 0.87$; (see Hernández-Martínez et al., 2024).

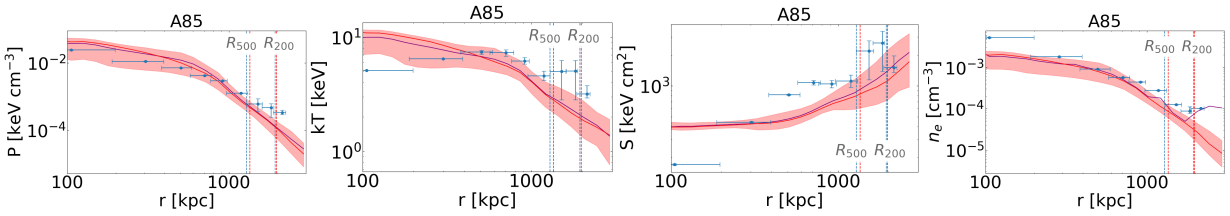


Figure 5.1: Deprojected thermodynamic profiles of the galaxy cluster Abell 85. The panels show (from left to right) the pressure, temperature, entropy, and electron density profiles as a function of radius. The red solid line represents the median of the simulated cluster, with the red shaded region indicating the σ scatter within each radial bin. The purple line represents the mean profile of the simulation. The blue data points correspond to deprojected observational data extracted using the Suzaku telescope and presented by Ichinohe et al. (2015). The vertical dashed lines indicate r_{500} and r_{200} from observations.

and surrounding structure in shaping the thermodynamic properties of the ICM and driving cluster-to-cluster variations.

5.4 Thermodynamic Profile Comparisons: Simulations vs. Observations

In the following, we present a detailed analysis on the thermodynamic state of each of the clusters analyzed and the quality of their profile reproduction.

- **Abell 0085** is a well-known SCC cluster characterized by its steep central temperature gradient, well-developed core, and relatively relaxed dynamical state (Reiprich & Boehringer, 2002). The Suzaku observations provide detailed measurements of the cluster's electron density, temperature, pressure, and entropy profiles, which reveal the thermodynamic properties of the ICM across a wide range of radii (Kawaharada et al., 2010; Ichinohe et al., 2015). As shown in Fig. 5.1, the simulation accurately reproduces the majority of the observed trends, particularly within r_{500}.

  The pressure profile of A85 is accurately replicated by the simulation. Within r_{500}, the simulation aligns closely with the observations, effectively capturing the radial decrease in pressure. This strong agreement demonstrates the gravitationally driven structure formation processes that dominate the inner regions of A85, where the cluster approaches hydrostatic equilibrium (Vikhlinin et al., 2006; Eckert et al., 2013), and the assumption of spherical symmetry is most valid in both observations and simulations. Beyond r_{500}, the simulation slightly underestimates the pressure, following a trend similar to that observed in the median electron density profile. This minor discrepancy likely stems from a combination of factors, including non-steady-state processes such as turbulence and accretion shocks, and limitations in numerical resolution, which may have a greater influence in the low-density

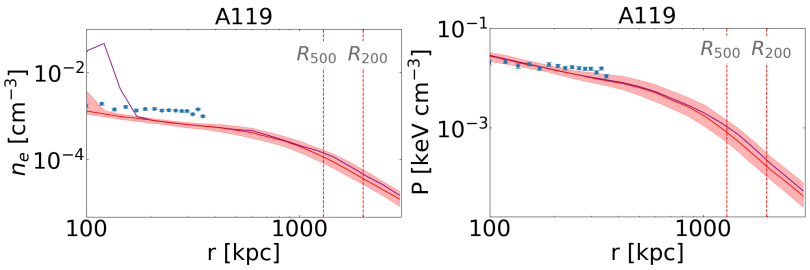


Figure 5.2: Deprojected thermodynamic profiles of the galaxy cluster Abell 119. The panels show (from left to right) the electron density and pressure profiles as a function of radius. The red solid line represents the median of the simulated cluster, with the red shaded region indicating the σ scatter within each radial bin. The purple line represents the mean profile of the simulation. The blue data points correspond to deprojected Chandra observations from the ACCEPT database. The vertical dashed lines mark characteristic radii, including r_{500} and r_{200} of the simulation. The observation profiles assume spherical symmetry and are obtained using an onion-shell deprojection technique.

outskirts. Nonetheless, these deviations remain small, and the simulation reproduces the overall pressure profiles with good accuracy.

The temperature profile is also well overall reproduced, and is well within the errors of the observations.

The electron density profile of A85 in the core and intermediate regions further demonstrates the accuracy of the simulation. Both, the mean and median predicted values closely track the decline present in observations. This demonstrates that the simulation effectively captures the overall gas distribution in the inner parts of the cluster. Beyond r_{500}, a noticeable bump in the mean values diverges from the median, coming closer to the values of the last two Suzaku observations. This increase in electron density may be linked to shocks associated with mergers involving substructures (Kempner et al., 2002; Durret et al., 2005), which appear to influence the median density values further out, particularly beyond r_{200}.

The observed entropy profile of A85 is particularly notable for its potential flattening at small radii, consistent with the presence of a cooling flow region (Hudson et al., 2010; Panagoulia et al., 2014), which is reflected in the long horizontal error bars in the observational data. A similar flattening is evident in the inner regions of the simulated cluster, appearing slightly more pronounced than in the observations. In the mid-regions, the observations exhibit a steeper rise in entropy compared to the simulation, but the profiles quickly converge as they approach r_{500}.

Overall, the reproduction of A85's deprojected ICM profiles is highly accurate, showing strong agreement with observations, suggesting that for A85 the assumptions of hydrostatic equilibrium and spherical symmetry provide a reasonable approximation.

- **Abell 0119** is considered a NCC cluster that exhibits significant elongation toward the northeast, suggesting it may be undergoing a merger (Govoni & Feretti, 2004; Hudson

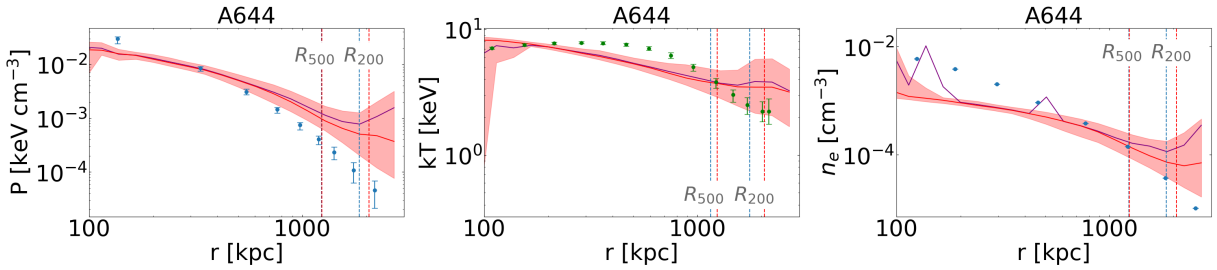


Figure 5.3: Deprojected thermodynamic profiles of the galaxy cluster Abell 644. The panels show (from left to right) the pressure, temperature, and electron density profiles as a function of radius. The red solid line represents the median of the simulated cluster, with the red shaded region indicating the 1σ scatter within each radial bin. The purple line represents the mean profile of the simulation. The blue data points correspond to deprojected XMM-Newton observations from Ettori et al. (2019), while the green points in the temperature panel represent *Planck* data as presented in the same study. The vertical dashed lines mark characteristic radii, including r_{500} and r_{200}.

et al., 2010; Parekh et al., 2015). Its X-ray peak, which does not dominate the surface brightness, is co-spatial with a central cD galaxy. The cluster is highly diffuse, lacking a prominent cool core, and features a large core region with slowly varying surface brightness, which makes deprojection highly unstable.

The electron density and pressure profiles of A119, shown in Fig. 5.2, were extracted from Cavagnolo et al. (2009) using *Chandra* observations and extend up to 500 kpc from the cluster's center. Overall, the simulation effectively captures the general trends observed in the inner regions of the ICM, demonstrating its ability to model the gas distribution and thermodynamic state of A119. However, the simulation shows a systematic deviation in slightly underestimating the density. This discrepancy may arise from lack of resolution as well as differences in the choice of the cluster center used for the deprojection (Maughan et al., 2008). The simulation employs the SUBFIND-assigned center, which corresponds to the gravitational potential center dominated by the dark matter component, whereas the observations use the X-ray peak as the center.

Due to its diffuse and highly disturbed nature, it is highly likely that the gravitationally determined center used in the simulation does not coincide with the X-ray peak center used in the observations. Furthermore, the highly asymmetrical structure of the cluster renders the assumption of spherical symmetry inappropriate for deprojection in both observations and simulations (Ventimiglia et al., 2008), adding additional uncertainty and potential sources of discrepancy. Despite these challenges, the simulation does successfully capture the overall trend of the electron density profile.

For the pressure profile, the simulation follows the observed profile very well, with all observed data points falling within the 1-sigma region of the simulated profile's median.

In summary, despite the challenges associated with profile post-processing and the asym-

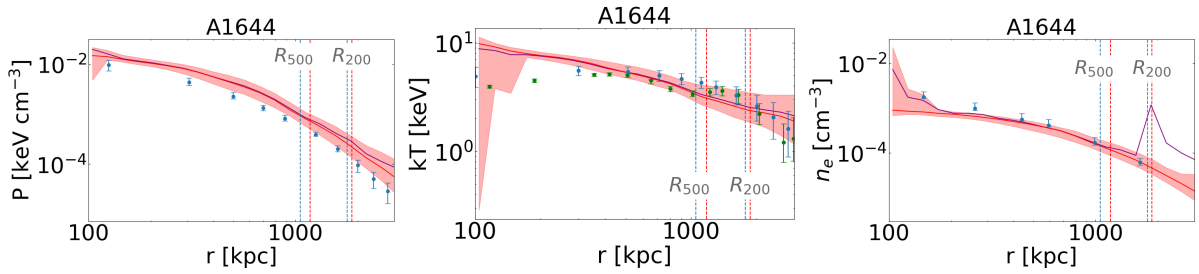


Figure 5.4: Same as Fig. 5.3 but for the galaxy cluster Abell 1644. Observational data are from Ettori et al. (2019).

metric nature of the cluster, the thermodynamic profiles of A119 are reproduced satisfactorily.

- **Abell 0644** is a dynamically disturbed galaxy cluster, often categorized as a NCC system (Buote & Tsai, 1996; Cassano et al., 2010; Rossetti et al., 2011). Previous studies have revealed significant deviations from symmetry in its ICM, likely driven by recent merger activity. Despite this, the cluster's large-scale properties remain consistent with relaxed systems (Hudson et al., 2010), presenting a unique combination of features. Observations indicate a relatively flat temperature profile in the central regions, reflecting the absence of a strong cool core (Cavagnolo et al., 2009; Lovisari & Reiprich, 2019). This combination of dynamical disturbance and large-scale quasi-relaxation makes A644 an interesting system for studying the interplay between mergers and the evolution of thermodynamic properties in the ICM.

  The thermodynamic profiles of A644 shown in Fig. 5.3, including pressure, temperature, and electron density, show a strong level of agreement between observations and simulations, particularly within r_{500}.

  The pressure profile shows that the simulation effectively captures the observed trends from the core to the outskirts of the cluster. Within r_{500}, the simulated values closely match the observations, with most data points falling within the 1σ uncertainty region of the simulated profile. In the very central region, the simulation slightly underpredicts the pressure, potentially due to the smoothing of substructures or differences in the choice of the cluster center between the simulation and observations. Beyond r_{500}, the simulation exhibits a mild bump in the pressure profile. While this feature could be influenced by unresolved small-scale substructures, other factors such as excess AGN feedback, an active dynamical phase, or intrinsic deviations from the observational counterpart may also contribute.

  Observations indicate a relatively flat temperature profile in the central regions, reflecting the absence of a strong cool core. The simulation accurately reproduces this flat trend, effectively capturing the thermodynamic behavior of the cluster within r_{500}. Beyond r_{500}, the simulation exhibits a similar bump as observed in the pressure profile, likely pointing to the presence of a merging substructure. In general, the deviation between simulation

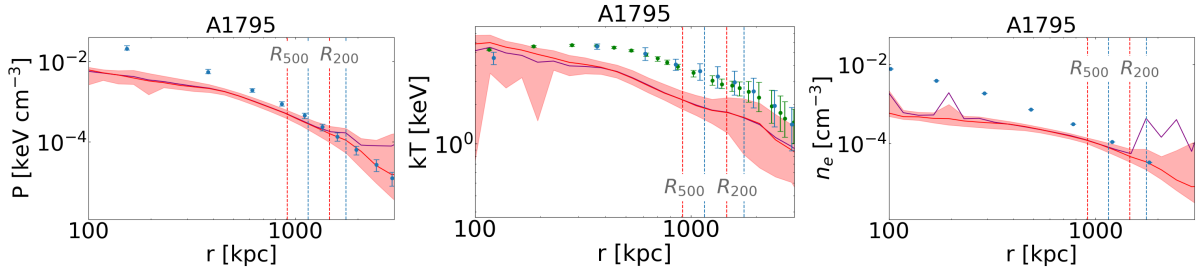


Figure 5.5: Same as Fig. 5.3 but for the galaxy cluster Abell 1795. Observational data are from Ettori et al. (2019).

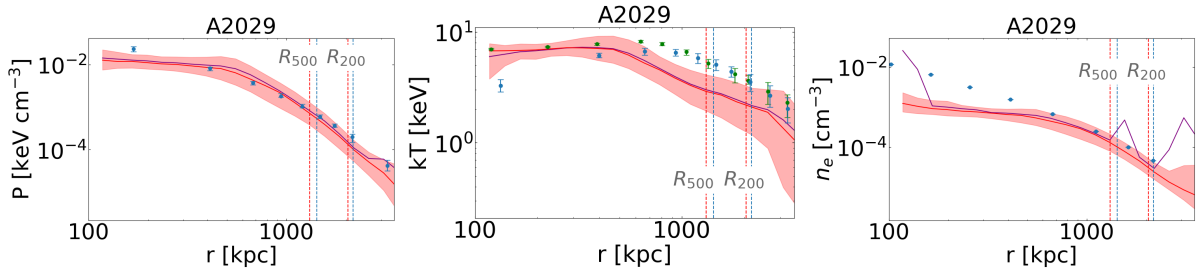


Figure 5.6: Same as Fig. 5.3 but for the galaxy cluster Abell 2029. Observational data are from Ettori et al. (2019).

and observations is of the same level as the deviations between the different observations presented in Fig. 5.3, pointing to an agreement between observations and our replica as close as the observational certainty itself. Thus, we conclude that overall, the simulation allineates well with the observed temperature profile, showing minimal deviations and demonstrating its ability to capture the thermodynamic state of the cluster's outer regions.

For the electron density, the median of the simulation lies below the observed values, but the mean exhibits noticeable bumps in the central region that come closer to the observational values. This behavior may indicate the presence of substructures in the cluster core, potentially pointing to recent merger activity also in the simulated replica.

- **Abell 1644** is a SCC cluster that exhibits clear signs of merging activity, as evidenced by the presence of a double X-ray peak (Scheck et al., 2023). Reiprich et al. (2004) analyzed XMM-Newton EPIC observations of the cluster, reporting that the flux of the northern (smaller) subclump lies below the HIFLUGCS flux limit, while the southern (larger) subclump exceeds this threshold. Further evidence suggests that the smaller subclump is being stripped as it moves through the ICM, reinforcing the picture of ongoing dynamical activity.

  The deprojected thermodynamic profiles of A1644, including pressure, temperature, and electron density, are shown in Fig. 5.4 and reveal a strong level of agreement between observations and simulations. The pressure profile demonstrates that the simulation ef-

fectively reproduces the observed trends from the core to the outskirts of the cluster. The simulated values closely match the observations, with the data points falling within the 1σ uncertainty band around the median of the simulated profile.

The temperature profile also shows good agreement between the simulation and observations, both within and beyond r_{500}. The simulated temperature captures the flat-to-declining trend observed in the core and intermediate regions. However, a slight overprediction of the central temperature is noted, potentially due to limitations in the simulation's ability to model cooling processes or the impact of localized substructures near the cluster center.

Similarly, the electron density profile reveals a close match between the simulation and observations, with predicted values lying within the observational uncertainties up to the cluster core.

- **Abell 1795.** In the left panel of Fig. 5.5, we present the deprojected radial pressure profile of the SCC cluster A1795. The red curve, indicating the median of our simulation suite, generally follows the observed data out to $r \approx r_{500}$ within the 1σ scatter (light red shaded region). Notably, in the innermost regions ($r \lesssim 200$ kpc), the observed pressure rises slightly above the median simulation curve. This offset may reflect key SCC features such as strong cooling, gas sloshing, and the presence of a cold front (Markevitch et al., 2001). These processes typically enhance the density and pressure within the cluster core, particularly when a cool filament and ongoing feedback mechanisms are present, as documented by Fabian et al. (2001), Oegerle et al. (2001) and Crawford et al. (2005). Beyond a few hundred kiloparsecs, the observations remain in concordance with the simulation, suggesting that the large-scale ICM structure is broadly captured. At larger radii ($\gtrsim r_{500}$), the separation of the mean (purple) and median (red) curves indicates some offset in the simulation, but overall consistency between the median and the observational data persists.

  The middle panel shows the deprojected temperature $kT(r)$. A1795's SCC nature manifests in a pronounced temperature drop in the core, consistent with the short CCT of $\sim 0.5\, h_{71}^{-1/2}$ Gyr found here and in previous studies (Fabian et al., 2001). Observational data (blue/green points) exhibit an inner temperature decline followed by a peak and eventual fall-off at large radii, indicative of the strong cool core and subsequent hot ICM envelope. The good agreement between the red (median simulation) curve and the observations in the intermediate radii ($r \sim 200$–500 kpc) implies that the simulations capture much of the thermal structure outside the immediate influence of the cold front. However, the presence of gas sloshing (Markevitch et al., 2001) and the prominent X-ray/Hα filament Crawford et al. (2005) in the central $\sim$ 50–100 kpc likely drives additional complexities not fully replicated in the median simulation. At larger radii, the overall shape of the temperature profile remains matched within $\pm 1\sigma$ scatter, indicating robust reproduction of global cluster properties by the simulations despite potential outliers.

  The right panel displays the deprojected electron density $n_e(r)$, which in the observations peaks sharply in the center—characteristic of a SCC environment. The simulation fails reproducing this high central densities and the steepness of the profile, most probably due to

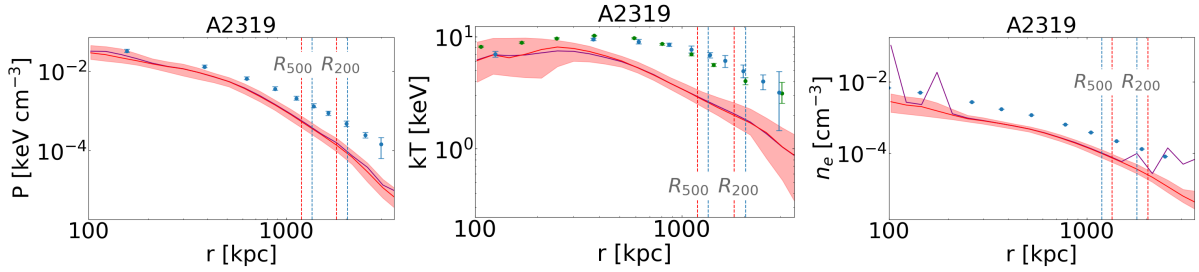


Figure 5.7: Same as Fig. 5.3 but for the galaxy cluster Abell 2319. Observational data are from Ettori et al. (2019).

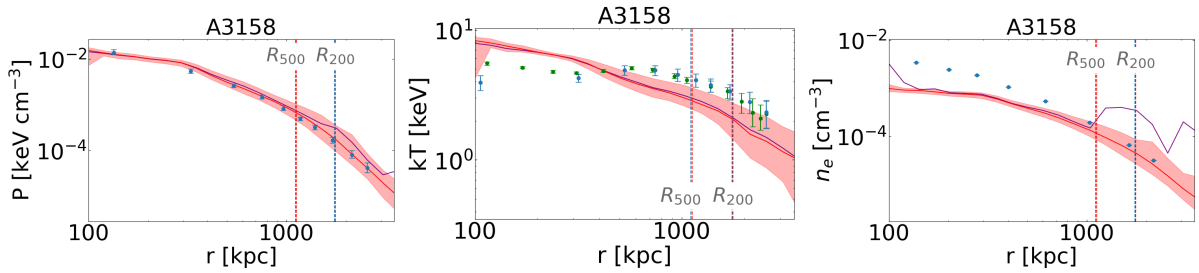


Figure 5.8: Same as Fig. 5.3 but for the galaxy cluster Abell 3158. Observational data are from Ettori et al. (2019).

resolution limitations. Even more, the estimated values for r_{500} and r_{200} are notably lower in the simulation than the ones estimated from the observations. Thus, higher resolution and possibility a better treatment of cooling and feedback would be necessary to reproduce the ICM properties of A1795.

- **Abell 2029** is a well-studied SCC cluster that features a cold front, as identified by Markevitch et al. (2003). The core has been extensively analyzed with *Chandra*, including a detailed study by Clarke et al. (2004). Thermodynamic and chemical properties reveal a dynamic and well-regulated environment. Figure 5.6 presents the deprojected thermo-

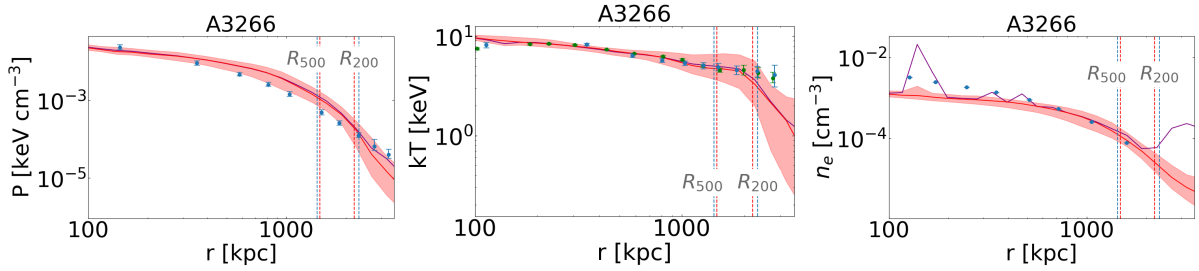


Figure 5.9: Same as Fig. 5.3 but for the galaxy cluster Abell 3266. Observational data are from Ettori et al. (2019).

dynamic profiles of A2029 from XCOP observations compared to simulated profiles. As before, the vertical grey lines indicate observational values for r_{500} and r_{200}, while the maroon lines mark the corresponding values from simulations.

The left panel of Fig. 5.6 shows the pressure profile for this cluster. The observed data start around a few 10^{-2} keV cm^{-3} in the core (at $r \sim 100$–200 kpc) and then smoothly decline by roughly two to three orders of magnitude by $r \sim 1$ Mpc. The simulation median generally tracks the measurements well, with its 1σ band covering the observations through most radii. Around r_{500} and beyond, the mean curve deviates upward from the median, hinting to some possible substructures.

The cluster's observed deprojected temperature profile shown in the middle panel of Fig. 5.6 is fairly high at small radii ($\sim$ 9–10 keV) and increases slightly to a broad peak ($\sim$ 11–13 keV) at a few hundred kpc, before declining beyond $r \sim$ 500–700 kpc. Observations (green/blue points) and the simulation median agree on the overall behavior of the profile, while the simulation peak happens slightly closer to the center ($\sim$ 500 kpc), than the observed peak at $\sim$ 600 kpc. After the peak, both simulation and observation suffer a decline, where the simulation's decline is slightly steeper. This suggests that, on average, the models replicate well A2029's hot ICM.

As in many massive clusters, A2029's electron density n_e is strongly peaked at small radii—$\sim 3 \times 10^{-2}$ cm^{-3} near 100 kpc—before falling by orders of magnitude outward, a feature the simulation finds some struggle replicating. Indeed, we find higher electron density measurements up to 700 kpc, from where simulation and observation start converging again.

In summary, A2029's profiles exhibit a characteristic "hot, dense core to cooler, lower-density outskirts" structure, which we can often find in massive galaxy clusters. While we do underestimate the electron density in the first 400 kpc of the cluster and converge from 500 kpc on, the simulation does reproduce the central and intermediate-radius regimes quite well for pressure and temperature, where we see that the overall trends and shapes are captured well along the complete profile.

- **Abell 2319** is a well-studied merger cluster, known for its large core region extending beyond 100 kpc and a prominent cold front (Feretti et al., 1997; Molendi et al., 1999; O'Hara et al., 2004). The surface brightness profile is well-represented by a β-model, reflecting the cluster's dynamic state and the impact of ongoing merger activity.

  As a dynamically active system, A2319 exhibits features such as a large core region, a prominent cold front, and substructures. These processes disrupt the equilibrium state typically assumed in many cosmological simulations, introducing complexities that are challenging to model accurately.

  In the left panel, the observed deprojected pressure profile of A2319 (blue points) decreases steadily from $\sim 10^{-2}$ keV cm^{-3} in the core to around 10^{-5} keV cm^{-3} beyond r_{500}. The median of the simulated profiles (red line) follows the observations closely through the

inner regions and starts deviating and underestimating the pressure at the intermediate and outer regions. This systematic underestimation likely arises from the inability of the simulations to fully capture the degree of gas compression and redistribution that arises during the merger. Similarly we underestimate the r_{500} and r_{200} radii for the cluster.

Turning to the middle panel, the observed temperature data points show a rise to a broad peak of $\sim$ 10–12 keV at a distance of 400 kpc. Simulated data show a peak at 300kpc and suffers a slightly steeper decline, although it does try to follow the general trend of the observational datapoints.

Finally, in the right panel, the electron density profile exhibits a steep decline with a similar slope in observations as in simulations and with a constant underestimation of around a factor 3 in the simulated reproduction.

Generally, the simulation reproduces the general trends of the thermodynamic profiles of A2319 but fails to replicate their absolute values, indicating a possible systematic bias. In the core, these biases are likely driven by an insufficient representation of gas compression and cold front dynamics, while in the outskirts, they could stem from an underestimation of shock heating and turbulence. The smoothing of substructures and the assumption of quasi-hydrostatic equilibrium in the simulations contribute to these discrepancies, as well as for the underestimation of r_{200} and r_{500}, particularly in systems like A2319 that are far from equilibrium.

- **Abell 3158** is classified as a relaxed cluster based on the velocity dispersion of its member galaxies (Łokas et al., 2006), which is consistent with the observed good agreement between simulations and observations across most of its thermodynamic profiles. The cluster's X-ray emission is elliptical in shape, and it hosts two central cD galaxies, with one located at the X-ray peak. Despite its relaxed classification, the cluster does not have a bright core, exhibiting a relatively low central electron density of approximately $5 \times 10^{-3}\,\mathrm{cm}^{-3}$.

  The simulation reproduces the observed pressure profile remarkably well. The agreement between the observations and simulation in the core and intermediate regions reflects the cluster's relaxed dynamical state, where gas is well-distributed and close to hydrostatic equilibrium.

  When it comes to the temperature profile, the simulation performs well in terms of absolute values, although the simulated replica shows a slowly decreasing trend with radius, while observations show a flatter profile. However, the core and intermediate regions ($r < r_{500}$), successfully reproduce the observed peak temperature of approximately 5.7 keV in the center.

  The observations of the electron density show a rather low density in the center; nevertheless, due to resolution limitations, this simulation is not capable of achieving central values much higher than $10^{-3}\mathrm{cm}^{-3}$, falling short in reproducing the inner electron densities of the cluster cores. Observations and simulations start converging again at $\sim r_{500}$.

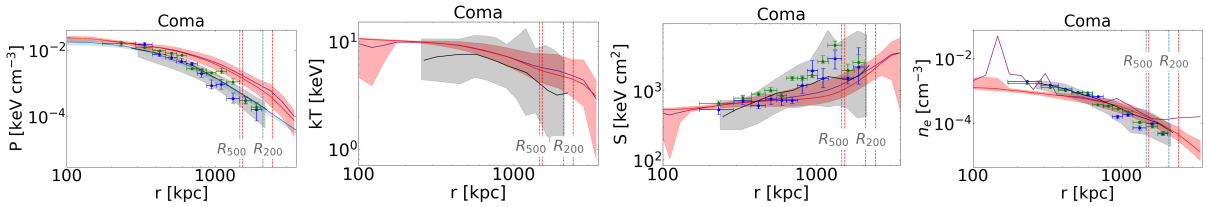


Figure 5.10: Deprojected thermodynamic profiles of the Coma cluster, comparing simulation results to observational data. The panels, from left to right, show the pressure, temperature, entropy, and electron density profiles. The red line represents the median profile of the simulated Coma cluster, while the red shaded region indicates the 1σ scatter of particles within each spherical-shell bin. The purple line represents the mean profile of the simulation. Observational data from A. Simionescu et al. (2013) are shown as blue points (E+NE direction) and green points (NW+W direction). The black line represents the median deprojected pressure profile from Mirakhor & Walker (2020), with the shaded grey region indicating the range covered by profiles in all observed sectors. In the leftmost panel, the light blue line represents the pressure profile derived from *Planck* data as presented by Planck Collaboration et al. (2013b). The vertical dashed lines indicate r_{500} and r_{200} from observations (blue) and simulations (red).

- **Abell 3266** is a well-known merging cluster (Henriksen & Tittley, 2002) The pressure profile shown in the left panel of Fig. 5.9 shows good agreement between the simulation and observations in the core and intermediate regions ($r < r_{500}$), suggesting that the simulation captures the broad thermodynamic structure of the cluster reasonably well.

  The temperature profile (middle panel of Fig. 5.9) highlights the cluster's disturbed thermodynamic state, with a relatively flat temperature distribution in the core and intermediate regions. This behavior is consistent with the turbulent mixing and heating caused by the ongoing merger, which suppresses the formation of a steep temperature gradient.

  The electron density profile shown in the right panel of Fig. 5.9 agrees well with the observed density at intermediate and outer regions but it under-predicts the density slightly in the inner parts. Interestingly, the simulated mean and median deviate strongly at r_{500}, indicating the existence of substructure possibly forming part of an ongoing merger.

  Overall, even if A3266 is a merging cluster, with presumably a disrupted inner structure, it is well reproduced in terms of its main ICM properties.

- **The Coma Cluster** is a massive, dynamically evolved, and nearly isothermal galaxy cluster, often used as a reference system for studying the thermodynamic properties of the ICM (Briel et al., 1992; Arnaud et al., 2001; Simionescu et al., 2013). Unlike SCC clusters, Coma exhibits a flat central entropy profile and lacks a strongly peaked density distribution, characteristics typical of NCC systems (Cavagnolo et al., 2009; Planelles et al., 2014). Observations indicate that the ICM is well mixed, with no evidence of a dominant central cooling flow (Matsushita et al., 2013).

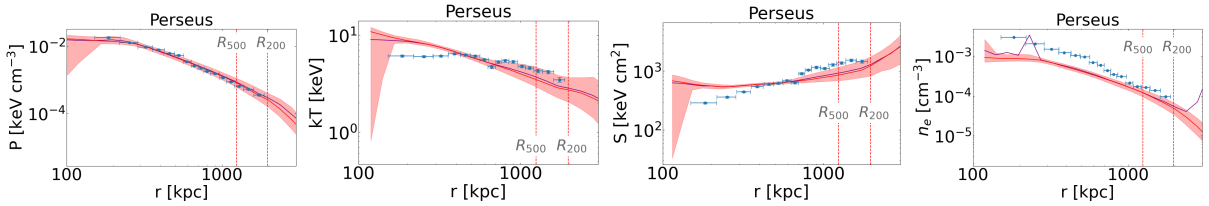


Figure 5.11: Deprojected thermodynamic profiles of the Perseus cluster compared to observational data from (Urban et al., 2013). The panels, from left to right, show the pressure, temperature, entropy, and electron density profiles. The red line represents the median profile of the simulated Perseus cluster, while the red shaded region indicates the 1σ scatter of particles within each spherical-shell bin. The purple line represents the mean profile of the simulation. The blue points correspond to deprojected observational data from Urban et al. (2013). The vertical dashed lines mark r_{500} and r_{200} of the simulations.

Figure 5.10 presents the deprojected thermodynamic profiles of Coma from multiple observational datasets compared to simulated profiles. Observational measurements include deprojected profiles from Simionescu et al. (2013), where green points represent the NW+W direction and blue points represent the E+NE direction. The black line represents the median deprojected pressure from (Mirakhor & Walker, 2020), with the shaded region indicating the range covered by deprojected profiles across all sectors. In the leftmost panel, the light blue line represents *Planck* data (Planck Collaboration et al., 2013b, Fig. 13).

The pressure profile (left panel) shows good agreement between the simulation and the observations at the core while slightly overestimating the values in the mid and outer regions. The simulated median profile lies close to the NW+W (green) deprojected data points and slightly higher than the E+NE (blue) data points.

The temperature profile (second panel) is relatively flat, consistent with the isothermal nature of the Coma Cluster (Arnaud et al., 2001). The simulation reproduces this trend very well, maintaining a nearly constant temperature over a large radial range (Mirakhor & Walker, 2020).

The entropy profile (third panel) shows that the simulation matches Simionescu et al. (2013) remarkably well, with both the simulated and observed profiles exhibiting a smooth increase with radius, characteristic of NCC clusters. The NW+W and E+NE measurements (green and blue points) follow the simulated median profile closely, confirming that the simulation captures the entropy structure of Coma accurately. Some scatter between different sectors is visible in the observational data, as also seen in the black median profile from Mirakhor & Walker (2020), but overall, the agreement between the simulation and the observations remains strong.

The electron density profile (right panel) also shows very good agreement between the simulation and observations. The simulated density profile follows the deprojected values from Simionescu et al. (2013) (blue and green points) closely at all radii. Since NCC clusters like Coma have more diffuse central gas, the simulation does not suffer from the

strong underestimation of central density that is sometimes seen in simulated CC clusters due to resolution limitations. The black line from Mirakhor & Walker (2020) provides a sector-averaged comparison, with the simulated profile staying well within its observed spread. Interestingly, the mean and the median deviate strongly in the core, showing the presence of big substructure in the center of the cluster, as expected for the Coma cluster.

Overall, the simulation provides an excellent fit to the observations from Simionescu et al. (2013) and (Mirakhor & Walker, 2020) across all thermodynamic profiles, confirming that it successfully reproduces the pressure, temperature, entropy, and electron density structure of the Coma Cluster.

- **The Perseus Cluster** is a well-studied SCC cluster and one of the most X-ray luminous clusters in the local Universe (e.g., Fabian et al., 2000b; Churazov et al., 2003). Observations from Urban et al. (2013) reveal a well-developed cool core with a steep temperature gradient and an entropy profile that flattens at small radii, consistent with a cooling flow region. The pressure and density profiles indicate a relatively relaxed structure, with no major recent disturbances within r_{500} (Zhuravleva et al., 2015).

  Figure 5.11 presents the deprojected thermodynamic profiles of Perseus from Urban et al. (2013) compared to simulated profiles. The vertical dashed line marks r_{500} and r_{200} of the simulated replica.

  The pressure profile (left panel) exhibits strong agreement between simulations and observations within r_{500}. The median simulated values closely track the observed data, and the 1σ region, remains relatively small, indicating that the pressure distribution is smooth and well-defined in the simulated ICM.

  The temperature profile (second panel) follows the observed trend well, showing a steep increase from the core, reaching a peak at intermediate radii. The simulation accurately reproduces this behavior within r_{500}, with most data points falling within the uncertainty range. However, at very small radii ($r \lesssim 200$ kpc), the simulation slightly overpredicts the temperature, which may indicate differences in cooling efficiency or AGN feedback behaviour, as it is expected to have a relevant impact at such scales (e.g., Fabian et al., 2006; McNamara & Nulsen, 2007).

  The entropy profile (third panel) shows the characteristic flattening at small radii, a feature associated with a cooling flow (Cavagnolo et al., 2009; Werner et al., 2013). The simulation captures this trend, though it slightly overestimates the entropy in the core, which could be linked to the treatment of radiative cooling and thermal conduction. At intermediate radii ($100 < r < 500$ kpc), the observed and simulated profiles align more closely, indicating that the simulation successfully models the entropy structure of the cluster.

  The electron density profile (right panel) is generally underestimated within r_{500}, but converges to observations at $r \approx r_{500}$.

- **The Virgo Cluster** The Virgo Cluster is the nearest massive galaxy cluster, located at a distance of 16 Mpc (Mei et al., 2007; Ferrarese et al., 2012). Unlike many other well-studied

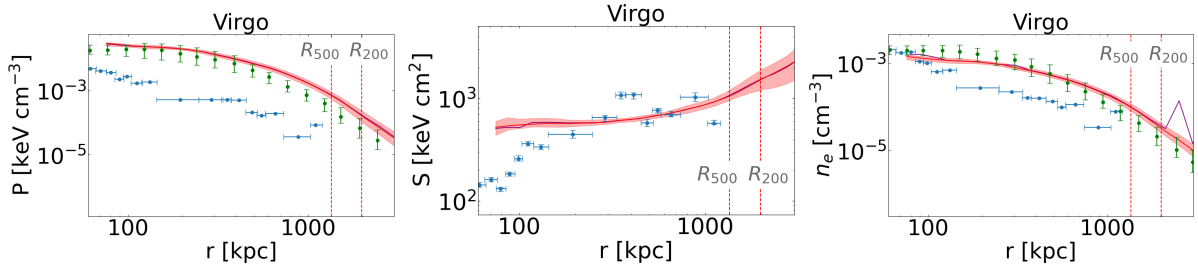


Figure 5.12: Deprojected thermodynamic profiles of the Virgo cluster, comparing simulation results to observational data. The panels, from left to right, show the pressure, entropy, and electron density profiles. The red line represents the median profile of the simulated Virgo cluster, while the red shaded region indicates the 1 σ scatter of particles within each spherical-shell bin. The purple line represents the mean profile of the simulation. Observational data from Urban et al. (2011) are shown as blue points. The green points represent deprojected profiles from a Virgo replica run using the RAMSES code presented by Lebeau et al. (2024), which employs different subgrid physics and resolution (dark matter particle mass of 10^9 $M_\odot$). The vertical dashed lines indicate r_{500} and r_{200} from the simulated replica.

clusters, Virgo is a dynamically young system, exhibiting significant substructure and signs of a late time merger and ongoing formation (e.g., Boselli et al., 2014; Sorce et al., 2016; Sorce et al., 2021). X-ray studies, such as the one performed by Urban et al. (2011) have revealed that the cluster has a relatively shallow density profile, with a power-law index of $\beta = 1.21 \pm 0.12$, which is lower than what is typically observed in more massive, relaxed clusters. The entropy profile follows a gravitational collapse-like power law ($K \propto r^{1.1}$) within 450 kpc, but beyond this radius, it flattens significantly, falling below theoretical expectations by a factor of 2–2.5 (Ettori et al., 2002; Urban et al., 2011). This behavior has been attributed to clumping in the ICM, which artificially enhances the observed density while reducing the inferred entropy (Simionescu et al., 2011; Nagai & Lau, 2011).

Due to its non-relaxed state and complex morphology, the results of the de-projection analysis, which assumes spherical symmetry, should be interpreted with caution (Mazzotta et al., 2004a). In fact, Urban et al. (2011) noted that systematic uncertainties due to deviations from spherical symmetry are significantly larger than measurement errors, complicating direct comparisons with theoretical models or simulations.

Figure 5.12 presents the deprojected thermodynamic profiles of the Virgo Cluster from Urban et al. (2011) alongside the simulated profiles. The vertical dashed lines mark r_{500} and r_{200}. Additionally, green points represent the deprojected profiles of a Virgo replica simulated with the RAMSES code as presented in Lebeau et al. (2024), which employs a different subgrid physics model and resolution (DM mass of 10^9 $M_\odot$ in the RAMSES replica, while 2.9×10^9 $M_\odot$ in the SLOW replica). Notably, both simulations (the main one and RAMSES) agree with each other, despite these differences in the implementation of code and subgrid physics.

The pressure profile (left panel) shows notable discrepancies between simulations and

observations, where observed data points systematically fall below the simulation. This may be due to departure from hydrostatic equilibrium, as turbulence and bulk motions in the Virgo ICM can redistribute pressure, leading to deviations from hydrostatic models. Additionally, projection effects in the observational de-projection analysis may contribute to the observed offset. The agreement between the two simulations suggests that this discrepancy is not simply a result of the simulation's subgrid physics or resolution. Instead, it may arise from a combination of observational systematics, Virgo's non-relaxed state, and potential shared limitations in the simulation models of ICM turbulence and heating.

The entropy profile (middle panel) shows a significant mismatch, with observed entropy values well below the simulated predictions. The simulated profile predicts a more gradual increase, while observations indicate a steeper rise followed by a flattening. This suggests that the simulation may overpredict heating processes or underestimate gas mixing, leading to systematically higher entropy values. Given the cluster's unrelaxed state, deviations from spherical symmetry may introduce further biases in the deprojected entropy values. The fact that the RAMSES simulation reproduces the same trends reinforces the robustness of the simulated entropy structure, regardless of the numerical implementation or resolution. However, the discrepancies could also indicate that the simulations fail to fully capture some of the physical processes shaping the entropy profile, such as turbulence, conduction, or small-scale gas clumping.

The electron density profile (right panel) also exhibits systematic differences. While the general decline in density is captured, the simulated cluster maintains a higher central density than the one observed. Observations indicate a more diffuse gas distribution, which could result from large-scale gas motions, substructure, or incomplete virialization. The agreement between the two simulations suggests that these discrepancies are not only an artifact of the simulation technique, but a consequence of a combination of observational de-projection assumptions, departures from hydrostatic equilibrium, and possibly limitations in how the simulations model the core structure of the Virgo cluster.

Overall, the Virgo Cluster presents significant challenges for direct comparisons between simulations and observations due to its dynamically non-relaxed nature. The large uncertainties in the de-projection analysis, as highlighted by Urban et al. (2011), suggest that some observed-simulated discrepancies may stem from the limitations of assuming spherical symmetry and hydrostatic equilibrium. However, the possibility remains that current simulations may not fully capture some aspects of Virgo's complex thermodynamic state, such as non-thermal pressure support, gas clumping, or turbulent mixing. Future work could explore non-spherical modeling approaches, alternative subgrid physics models, or dynamical state indicators to better characterize the Virgo cluster's complex ICM.

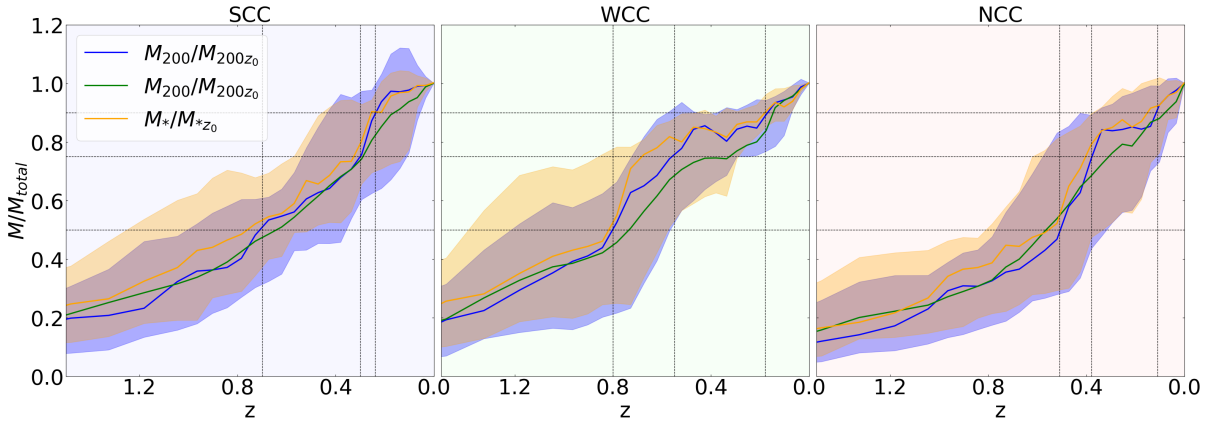


Figure 5.13: Evolution of the mass assembly history for clusters classified as SCC, WCC, and NCC based on observational criteria. The panels show the total median and mean mass growth as well as the total stellar mass growth, normalized to the final mass, as a function of redshift. The blue line represents the median $M_{200c}/M_{200,z=0}$ mass growth, while the green line shows the mean $M_{200c}/M_{200,z=0}$ mass growth. The orange line represents the median stellar mass growth, $M_\star/M_{\star,z=0}$. The shaded regions indicate the 1σ scatter for the total median mass and the median stellar mass for each classification. The vertical dashed lines mark the redshifts at which 50%, 75%, and 90% of the total mass was assembled.

5.5 The Impact of Formation History on Cluster Thermodynamics

We have now extensive evidence that the assembly history of a galaxy cluster plays a fundamental role in determining its present-day thermodynamic state (Voit et al., 2005; Burns et al., 2008; Kravtsov & Borgani, 2012; Lau et al., 2015). Current simulations, including the SLOW suite, face challenges in accurately reproducing the thermodynamic profiles of the innermost regions of galaxy clusters. In particular, they struggle to replicate the observed differences in central slopes between SCC, WCC, and NCC clusters (see e.g. Rasia et al., 2015; Planelles et al., 2017; Barnes et al., 2018; González Villalba et al., 2025). These discrepancies likely arise from an incomplete understanding of subgrid physics, particularly regarding cooling, turbulence, and feedback mechanisms. However, a key strength of the SLOW constrained simulations is their ability to faithfully reproduce the LSSs of the local Universe, ensuring that the formation histories of individual clusters are well captured (see, e.g., Sorce et al., 2016). With our simulations successfully matching the observed thermodynamic profiles at intermediate and large radii, where our lack of resolution and subgrid physics does not introduce significant biases, we can be confident that the formation pathways of these structures are realistically constrained. This enables us to investigate the connection between a cluster's assembly history and its present-day classification as SCC, WCC, or NCC. The classification of clusters follows the criteria established by Hudson et al. (2010) for the HIGFLUGCS catalog (see Sect.1.2.3 for further details).

Figure 5.13 illustrates the median evolution of total mass (M_{200}) and stellar mass (M_*),

normalized to their respective values at $z = 0$, for the clusters in our simulated volume classified by observations as SCC, WCC, and NCC clusters respectively. A clear trend emerges: current-day SCC clusters tend to finish their mass assembly earlier than their WCC and NCC counterparts. In fact, current day SCC have assembled a larger fraction of their total mass already at $z \approx 0.25$ while their WCC and NCC counterparts still grow in mass significantly after $z \approx 0.25$. The vertical markers in Fig. 5.13 indicate the redshifts at which 50%, 75%, and 90% of the total mass has been accreted. Clearly visible for SCC clusters is an extended plateau from z ≈ 0.25 to z ≈ 0. This suggests a more rapid formation history before, followed by a late-time relaxation. In contrast, NCC clusters accumulate their mass more gradually, and a significant fraction of their growth occurs more abruptly at recent times, indicating fast accretion episodes through significant mergers. WCC clusters exhibit an intermediate accretion pattern, bridging the gap between SCC and NCC populations. Their mass assembly shows subsequent increases from z≈1.0 onward, but these are less abrupt compared to the NCC case. This increased merger rate of NCC clusters is in line with recent findings supporting mergers as drivers of the transition between SCC and NCC clusters (Sarazin, 2002; Chen et al., 2007; Burns et al., 2008; Sanders et al., 2010; Rossetti & Molendi, 2010; Hudson et al., 2010; Hoffer et al., 2012; González Villalba et al., 2025).

This distinction in mass growth histories is also reflected in the evolution of the stellar mass component. The stellar mass growth curves (yellow lines in Fig. 5.13) reveal that SCC clusters build up their stellar content earlier and more efficiently compared to NCC clusters. This early stellar mass accumulation is consistent with a formation scenario where SCC clusters experience early, intense cooling and subsequent star formation within their cores. Conversely, NCC clusters show a delayed stellar mass growth, likely due to a more prolonged merger history that disrupts early cooling and extends the timescales for stellar assembly.

The delayed assembly of NCC clusters agrees with the expected more disturbed dynamical states observed at $z = 0$. Mergers inject turbulence into the ICM, suppressing the formation of a central cool core by redistributing entropy and mixing high- and low-entropy gas. This mechanism naturally explains why NCC clusters exhibit flatter entropy profiles and weaker central cooling signatures compared to SCC clusters. The enhanced scatter in the total mass accretion histories of NCC clusters further supports the idea that these systems undergo more stochastic growth, with frequent late-time mergers playing a dominant role in shaping their final structure.

WCC clusters represent an intermediate evolutionary state, exhibiting characteristics of both SCC and NCC populations. Their accretion histories suggest a mix of early mass assembly with later minor mergers that do not fully disrupt their cores. This transitional nature may indicate that WCC clusters evolve into NCC clusters over time if additional late-time mergers continue to inject energy into the ICM.

Overall, the connection between merger history and cluster classification at $z = 0$ is evident in both the total and stellar mass assembly histories. SCC clusters, which form earlier and more quiescently, retain their dense, cooling cores, while NCC clusters, shaped by extended merger activity, remain dynamically disturbed with higher entropy cores. The SLOW simulations successfully reproduce these trends, offering a detailed view of how individual formation pathways influence the present-day structure of local galaxy clusters.

5.6 Summary and Conclusions

In this study, we examined the thermodynamic properties of local galaxy clusters using the constrained cosmological simulation AGN1536[3] from the SLOW project. By performing one-to-one comparisons with observed systems, we assessed the ability of the simulation to reproduce late-time profiles in the Local Universe and explored the physical processes shaping them. The close agreement between simulated and observed deprojected profiles suggests that constrained simulations effectively capture the large-scale formation histories of individual clusters, providing a reliable framework for connecting past accretion and merger events to present-day thermodynamic states.

Our findings highlight several key aspects of cluster evolution:

1. The SLOW simulations successfully reproduce the thermodynamic structure of galaxy clusters, recovering key trends in pressure, temperature, electron density, and entropy profiles across a broad range of cluster environments. While the agreement is particularly strong for dynamically evolved systems, the simulation also matches the profiles of disturbed clusters such as Coma and A119, where the large-scale ICM properties agree well with observations.

2. Current cosmological simulations, including SLOW, face difficulties in accurately reproducing core regions, particularly in electron density and entropy profiles, which tend to be systematically underestimated or overestimated. These discrepancies arise from a combination of numerical resolution constraints and limitations in the implementation of subgrid physics, including potentially excessive AGN feedback and incomplete modeling of non-thermal pressure support. The largest deviations are found in NCC clusters, where the spherical symmetry assumption used for profile de-projection is less applicable, and merger-driven turbulence and substructure interactions introduce additional complexities that are not fully captured. However, the ability to directly compare simulated and observed clusters on a one-to-one basis reveals a new pathway to identify where current models struggle to capture key physical processes, particularly in cluster cores, and provide critical insights into the mechanisms shaping core structure and the impact of mergers, turbulence, and feedback. Identifying these mismatches serves as a new tool to identify pathways toward refining AGN and star formation models with a stronger physical foundation.

3. The overall accurate reproduction of late-time profiles indicates that the simulation captures the formation histories of clusters well. This makes it possible to connect formation history to observables that are otherwise difficult to reproduce in simulations, such as the cores of galaxy clusters. We studied the evolutionary paths of the clusters of our simulation, classifying them based on their observed cores: SCC, WCC, and NCC. We found that SCC, WCC and NCC clusters differ in their evolution, with SCC clusters forming earlier and evolving more quiescently, while WCC clusters experience extended merger activity that disrupts their cores and increases entropy accross their evolution and NCC experience

late time major mergers that make them accrete most of their mass in later times. This link between formation history and core structure provides insight into regions of the ICM where subgrid models remain incomplete, particularly in the treatment of AGN feedback, turbulence, and radiative cooling.

These results emphasize the value of constrained simulations in bridging theoretical models and observations, offering a powerful framework to investigate the formation and evolution of galaxy clusters in a cosmological context. By enabling direct comparisons between simulated and observed structures, we validate the ability of these simulations to recover the large-scale formation history of clusters while also exposing critical gaps in our understanding of cluster cores and their thermodynamic evolution. Our analysis highlights a clear connection between formation history and present-day core properties, with SCC clusters forming earlier and evolving more quiescently, WCC clusters experiencing prolonged merger activity that gradually disrupts their cores, and NCC clusters undergoing late-time major mergers that significantly reshape their structure. This evolutionary link provides a new avenue for identifying the missing physical ingredients in subgrid models, particularly in AGN feedback, turbulence, and non-thermal pressure support, neccessary to accurately simulate the formation and evolution of galaxy clusters. Moving forward, improving these models will be essential for capturing the full complexity of cluster cores and their thermodynamic states. Future high-resolution simulations, combined with increasingly precise observational data, will refine our understanding of these processes and further enhance the predictive power of constrained cosmological simulations, advancing our knowledge of galaxy cluster physics and LSS formation in the Universe.

6 | Cosmological and Astrophysical Parameter Inference from Stacked Galaxy Cluster Profiles Using CAMELS-zoomGZ

The content of this chapter has been published by Hernández-Martínez et al. (2025) in ApJ. Some adjustments have been made.

6.1 Galaxy Cluster Profiles as Cosmological Probes

Galaxy clusters serve not only as proof of the hierarchical nature of cosmic structure formation but also as invaluable laboratories for probing both astrophysical processes and fundamental cosmological parameters. Their formation and evolution are governed by the underlying cosmology, while their observable properties are shaped by complex baryonic physics such as gas cooling, star formation, and feedback from SNe and AGNs. This dual sensitivity makes galaxy clusters, and in particular their radial profiles, exceptionally powerful cosmological probes (Mandelbaum, 2018; Secco et al., 2022).

Radial profiles of galaxy clusters encapsulate a wealth of information about their thermodynamic state and composition. Properties such as the ICM gas density, temperature, metallicity, X-ray surface brightness, and the tSZ Compton-y parameter provide complementary insights into the physics of the ICM and the gravitational potential wells of clusters. For instance, X-ray observations from instruments like *Chandra* and *XMM-Newton* have mapped the temperature and density structure of clusters in detail (Vikhlinin et al., 2006; Pratt et al., 2009; Ettori et al., 2013; Zhuravleva et al., 2014), while the SZ, as measured by facilities such as the Atacama Cosmology Telescope (ACT) and the South Pole Telescope (SPT), directly probes the pressure distribution of the hot gas (Battaglia et al., 2012; Hasselfield et al., 2013; Bleem et al., 2015; Raghunathan et al., 2019; Hilton et al., 2021).

The potential of these profiles extends beyond individual cluster studies. When stacked across samples of clusters, radial profiles can mitigate the stochasticity and noise inherent to single systems, revealing underlying trends driven by cosmology and baryonic physics. This approach is particularly valuable in the era of large sky surveys, where ensemble analyses of cluster populations can make use of their statistical power to achieve precision cosmology. Upcoming missions like *eROSITA*, *Euclid*, and the Vera C. Rubin Observatory's LSST are expected to deliver

such datasets, enabling deep exploration of the mass and redshift dependence of cluster properties and their cosmological implications (Ivezić et al., 2019; Euclid Collaboration et al., 2022; Secco et al., 2022).

In this context, we employ the CAMELS-zoomGZ simulations (Lee et al., 2024) to explore the efficacy of stacked galaxy cluster profiles as cosmological probes. CAMELS-zoomGZ is a set of 768 zoom-in simulations centered on galaxy clusters and groups. It extends the CAMELS suite (Villaescusa-Navarro et al., 2021) to high-mass systems (10^{13} – $10^{14.5}\,M_{\odot}$), across the full 28-dimensional parameter space (spanning cosmological and astrophysical variables) of the IllustrisTNG model (Weinberger et al., 2017; Pillepich et al., 2018, see also Sect. 2.4.2 for further details). This expanded parameter space includes variations in key cosmological parameters such as $\Omega_{\rm m}$, H_0, and σ_8, alongside an extensive range of baryonic physics prescriptions encompassing stellar and AGN feedback, IMF slopes, and metal enrichment processes (Ni et al., 2023).

In this work, we make use this new suite of simulations to focus on extracting detailed information from simulated stacked galaxy cluster profiles. Our analysis reveals that stacked galaxy cluster profiles allow us to infer all cosmological as well as all astrophysical parameters of the model with the highest accuracy ever achieved, enhancing our ability to interpret observations from upcoming cosmological surveys.

6.2 Neural Network Architecture and Training for Parameter Inference

6.2.1 Architecture and Loss Function

In this work, we use an ML architecture to perform inference on the values of all cosmological and astrophysical parameters in the IllustrisTNG model. We do this by using stacked galaxy cluster profiles emulated applying CarpoolGP (presented in detail in Sect. 3.3.2) on the extracted profiles from the CAMELS-zoomGZ simulation set. These profiles include, temperature, metallicity, X-ray surface brightness, pressure and Compton-y. The architecture used in this work comprises multiple fully connected layer blocks, each containing a fully connected layer, followed by a LeakyReLU activation layer with a slope of 0.2 and a dropout layer, where the dropout rate is a hyperparameter. The final layer of the architecture is a fully connected layer.

Our NN is designed to compute key statistical moments, such as the mean and variance, of the marginal posterior distributions for each parameter, without making any assumption about the shape of the posterior.

The input to our models is a 1D vector containing the values of an averaged galaxy cluster profile for each radial bin. The model output $2N_{\rm params}$ values, where $N_{\rm params}$ is the number of parameters considered for inference. For each parameter i, the models return its marginal posterior mean (μ_i) and standard deviation (σ_i). This is achieved by minimizing the following loss function:

$$\begin{aligned}\mathcal{L} = &\sum_{i=1}^{28} \log\left(\sum_{j\in \text{ batch}} \left(\theta_{i,j} - \mu_{i,j}\right)^2\right) \\ + &\sum_{i=1}^{28} \log\left(\sum_{j\in \text{ batch}} \left(\left(\theta_{i,j} - \mu_{i,j}\right)^2 - \sigma_{i,j}^2\right)^2\right)\end{aligned} \tag{6.1}$$

This loss function ensures that μ_i and σ_i represent the parameter's posterior mean and standard deviation, as described in Jeffrey & Wandelt (2020) and Villaescusa-Navarro et al. (2021).

The training process uses simulated data $\{x_i, \theta_i\}$, where the parameters θ_i are sampled from prior distributions. The networks learn to map the input data x to the moments of the posterior distribution. During inference, the trained networks can rapidly evaluate the moments for newly observed data, eliminating the need for extensive sampling or density estimation, thereby streamlining the process of parameter inference in complex cosmological models.

6.2.2 Input Data

The one-dimensional vector that we will use as input can be composed of only one or several concatenated average galaxy cluster radial profiles of various types, extending up to r_{200c} for a specific point in the IllustrisTNG model parameter space. The quantities considered include radial temperature, metallicity, gas density, X-ray surface brightness, and the Compton-y profile. The first three profile types are divided into 30 bins each, while the latter two consist of 29 bins. Therefore, the input size ranges from 29 to 148 bins, depending on the combination of profiles used.

We analyze the concatenated emulated profiles to assess their impact on the inference of the 28 parameters of the IllustrisTNG model presented in Table 6.1. Additionally, we investigate the inference performance when using each profile type individually.

To investigate the effect of noise on the profiles, we generated a set of noisy profiles by adding random Gaussian noise to each bin, with noise levels of 10%, 20%, 30%, and 40% of the bin signal.

To determine where the majority of the information resides, we created a set of profiles by progressively excluding the outer regions, ranging from $0.7R_{200c}$ to $0.1R_{200c}$. Similarly, we cut out the inner regions from $0.1R_{200c}$ to $0.27R_{200c}$ to analyze the impact of the outer regions on the inference.

Additionally, we also created integrated values from the profiles to compare the inference performance of the full averaged radial profiles with that of the integrated quantities, providing insights into the relative effectiveness of each approach for different parameters.

6.2.3 Training procedure and optimization

We trained our model on the emulated galaxy cluster profiles. We examined the effect of the size of the training set on the inference and concluded that employing 300,000 averaged profiles when using one profile type yields a noticeable improvement in accuracy, but it is more convenient and

Parameter	Meaning
Ω_m	Ratio of the matter density of the universe to the critical density.
σ_8	Amplitude of matter fluctuations on a scale of 8 Mpc.
WindEnergyIn1e51erg (ASN1)	Galactic wind energy (in units of 10^{51} ergs) injected into galactic winds by SN feedback.
RadioFeedbackFactor (AAGN1)	Efficiency of energy injection from AGN in their kinetic mode into the surrounding gas.
VariableWindVelFactor (ASN2)	Controls wind velocity scaling according to local galaxy conditions, particularly setting the speed per unit of star formation.
RadioFeedbackReiorientation-Factor (AAGN2)	Determines the speed and frequency of feedback injection events in radio (kinetic) mode.
Ω_b	Ratio of the density of baryonic matter to the critical density of the universe.
H_0	Hubble parameter, i.e. current rate of expansion of the universe
n_s	Initial power spectrum spectral index.
MaxSfrTimescale (t_SFR)	The timescale over which gas is converted into stars at the star-formation density threshold.
FactorForSofterEQS	Adjustment to the EQS for gas–see Springel & Hernquist (2003)–in star-forming regions.
IMFslope	Slope of the IMF, which describes the distribution of masses for a population of stars at the time of their formation.
SNII_MinMass_Msun (SNII)	Minimum stellar mass, in solar masses, required for a star to end its life as a SNe II.
ThermalWindFraction (W1)	Proportion of the total energy from stellar feedback that is used to heat the gas and drive thermal wind (injected as thermal feedback).
VariableWindSpecMomentum (W2)	Regulates how the speed of galactic winds changes due to stellar feedback.
WindFreeTravelDensFactor (W3)	Controls wind particle travel distance by setting a density threshold, determining how far they move before recoupling , affecting galactic wind propagation through the ISM and Circumgalactic Medium (CGM).

Table 6.1: Parameters of the IllustrisTNG model.

Parameter	Meaning
MinWindVel (W4)	The minimum velocity imparted to gas particles that are ejected as part of galactic winds.
WindEnergyReductionFactor (W5)	The fraction by which the galactic wind energy budget, via the mass-loading, is reduced at higher metallicities.
WindEnergyReduction-Metallicity (W6)	Adjusts the typical metallicity at which the galactic wind energy budget will be reduced.
WindEnergyReduction-Exponent (W7)	Defines the exponent in the power-law relationship between wind energy and gas metallicity, influencing how strongly metallicity affects the wind energy.
WindDumpFactor (W8)	Represents the fraction of metals that are not fully ejected out into the galactic wind, but instead get deposited into nearby star-forming cells before the actual wind ejection.
SeedBlackHoleMass (BH1)	Initial mass of seed black holes introduced in the simulation to represent the formation of SMBHs in galaxies.
BlackHoleAccretionFactor (BH2)	Rate at which black holes accrete gas, expressed as a Bondi rate multiplier, affecting black hole growth and AGN feedback processes.
BlackHoleEddingtonFactor (BH3)	Normalization factor for the limiting Eddington accretion rate for the accretion onto SMBHs. In some cases it may allow super-Eddington accretion scenarios in the simulation.
BlackHoleFeedbackFactor (BH4)	Scales the energy released by an accreting black hole to influence its environment through thermal energy, kinetic outflows, or radiation.
BlackHoleRadiativeEfficiency (BH5)	Refers to the fraction of accreted gas mass rest energy converted into radiation, affecting AGN luminosity and its environmental feedback.
QuasarThreshold (Q1)	Eddington ratio threshold for classifying a black hole in quasar mode feedback, distinguishing AGN feedback modes by accretion rate.
QuasarThresholdPower (Q2)	Sets the power-law dependency of the threshold accretion rate, influencing the steepness of the transition between AGN feedback modes.

Table 6.2: Continuation of table 6.1.

computationally efficient to use 30,000 averaged profiles of each type when using the five profiles types altogether (see Sec. 6.4.1). These profiles were consistently divided such that 70% were used for training, and 15% were allocated for both validation and testing, respectively, with the samples randomly drawn across the parameter space.

We trained the models using the specified architecture for 1000 epochs, employing the ADAM optimizer (Kingma & Ba, 2014) to perform gradient descent, with a batch size of 256 samples. Hyperparameter optimization was conducted using the Optuna package (Akiba et al., 2019), which employs Bayesian optimization with the Tree Parzen Estimator (TPE) (Bergstra et al., 2011). The optimized hyperparameters included the number of layers, the number of neurons in the fully connected layers, the learning rate, the weight decay, and the dropout rate. This process involved at least 1000 trials, with Optuna directed to minimize the validation loss. The validation loss was computed using an early-stopping scheme to ensure only the model with the minimum validation error was saved. The selected model was then used for subsequent testing.

6.2.4 Performance Evaluation

As previously introduced in Sect. 3.2.2, our assessment of model performance relies on four complementary metrics that jointly characterize the accuracy and precision of the parameter inference: the Root Mean Squared Error (RMSE), the mean relative uncertainty (ϵ), the correlation coefficient (r), and the reduced chi-squared (χ^2).

The RMSE represents the absolute measure of the fit, providing information of the magnitude of the error in the prediction. The mean relative uncertainty, ϵ, evaluates the typical relative scale of the predicted standard deviation, serving as an indicator of model precision. The correlation coefficient, r, measures the strength and direction of the linear relationship between predicted and true values, with values close to 1 indicating predictions closely following the underlying truth, values close to 0 showing no correlation between predictions and real values and values closer to -1 indicating a strong negative linear correlation. Finally, the reduced chi-squared statistic, χ^2, provides a measure of how accurate the predicted standard deviations of the network are by comparing it to the residuals.

Together, these metrics offer a robust and comprehensive evaluation framework, enabling us to quantify both the predictive power and the uncertainty calibration of our models across the cosmological and astrophysical parameter space explored in this work.

6.3 Assessing the Power of Cluster Profiles for Cosmological Parameter Inference

In this section, we present the outcomes derived by performing inference of the IllustrisTNG model parameters using averaged galaxy cluster profiles.

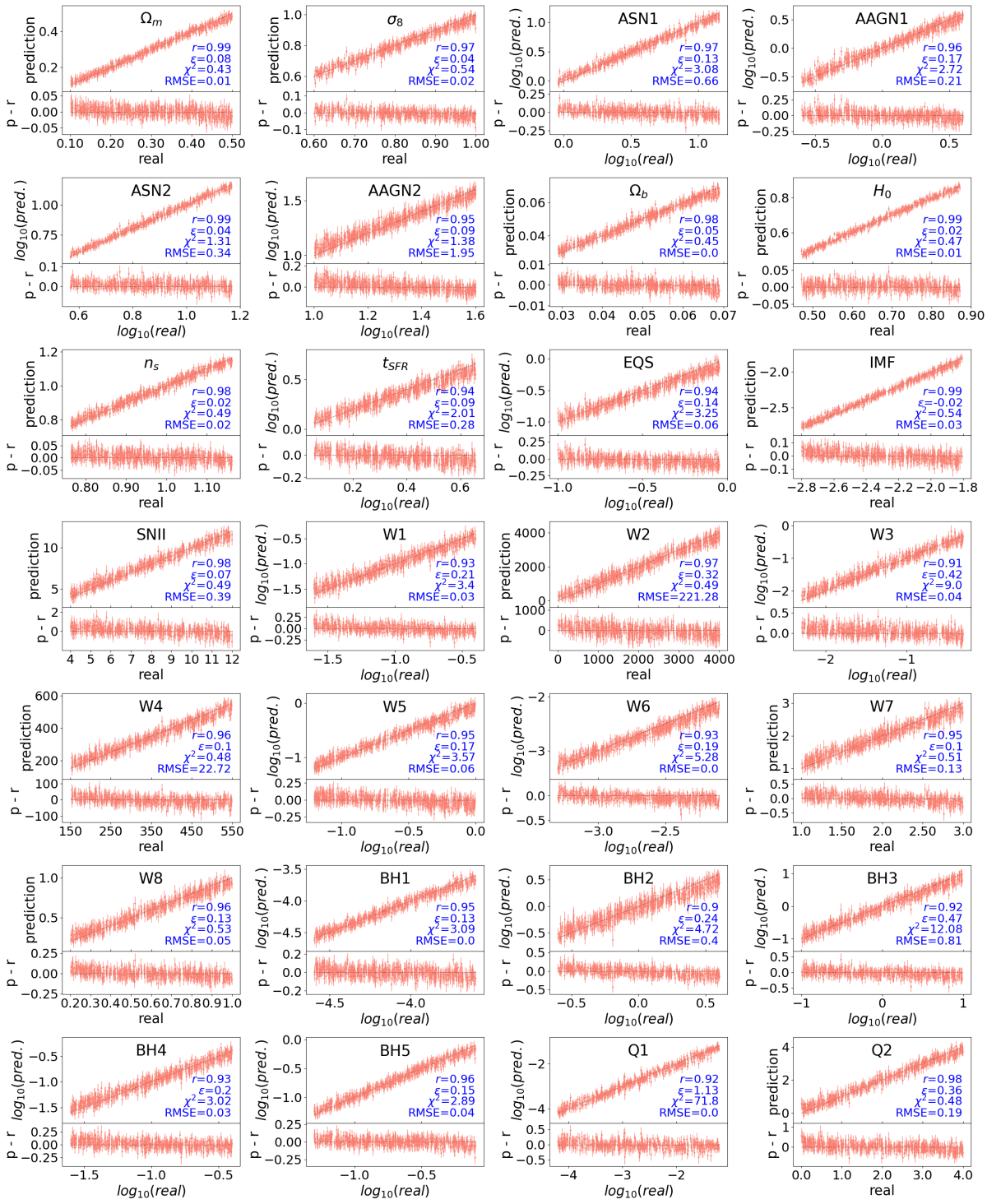


Figure 6.1: Inference results for the parameters in the IllustrisTNG model. The top panels for each parameter show the comparison between true values (x-axis) and inferred values (y-axis), with a black line representing the one-to-one correspondence. The bottom panels display the residuals (inferred minus true values) as a function of the true values to better illustrate the spread. Each subplot corresponds to a different parameter, covering a wide range of cosmological and astrophysical parameters within the model. The correlation coefficient (r), root-mean-square error (RMSE), mean relative error, and reduced chi-squared (χ^2) are reported within each top panel. The results demonstrate high accuracy and reasonable error margins across all parameters, as indicated by the strong correlation coefficients and relatively low residuals.

6.3.1 Inference with multiple cluster profiles

The primary aim of this study is to establish the theoretical framework for cosmological parameter inference using averaged or stacked galaxy cluster profiles by inferring cosmological and astrophysical parameters from simulated clusters in the IllustrisTNG model (parameters listed in Table 6.1). As a start, we establish a set of 300,000 points in the IllustrisTNG model parameter space, for which we emulate the five averaged profile quantities, at a constant mass of $10^{14} M_{\odot}$. Of these points in parameter space, 70% are selected for training, while 15% are reserved for validation and testing, respectively. For each parameter space point, a 1-dimensional vector of concatenated averaged cluster profiles serves as the input to our network.

Once the model is trained and validated with five concatenated averaged profiles using 210,000 (70%× 300,000) and 45,000 (15%× 300,000) parameter space locations for training and validation, we test it using other independent 45,000 parameter space locations. The model returns for each input vector a posterior mean and standard deviation as prediction for each parameter of the IllustrisTNG model. Fig. 6.1 shows the result of the prediction for each of the 45,000 individual points in parameter space belonging to the testing set. Fig. 6.1 displays the prediction results for each of the 45,000 test points in parameter space, with the y-axis representing the network's predicted values and the x-axis showing the actual "target" or real values. The error bars indicate the uncertainty estimated by the network for each prediction.

We employ the metrics presented in Sec. 3.2.2 to quantify the model performance, revealing high accuracy and reasonable error margins. The model notably excels in inferring cosmological parameters like $\Omega_{\rm m}$, H_0, and Ω_b, each showing a correlation coefficient of 0.99, followed closely by σ_8 and n_s with correlation coefficients of 0.98 and 0.97, respectively. All astrophysical parameters demonstrate correlation coefficients above 0.90, with some, such as the IMF slope and the ASN2, approaching 1.

To further understand the outstanding performance of our model, we examine the influence of individual parameters within the IllustrisTNG model by generating mean galaxy cluster profiles under various fiducial conditions in the IllustrisTNG parameter space, involving 27 fixed parameters and one variable parameter. This method allows for a thorough exploration of each parameter's impact and nuances on different profile types, also highlighting potential degeneracies.

Fig. 6.2 illustrates the impact of varying each cosmological parameter across its full range for each profile type. The results indicate that each parameter introduces distinct features at specific points in the profiles, which are not replicated by other parameters. This suggests that there are no significant degeneracies between the effects of different parameters on the profiles, as each produces unique, non-overlapping features. This holds true across all profile types.

Additionally, we observe that certain parameters cause greater variation than others. For example, $\Omega_{\rm m}$ and H_0 result in the most pronounced spread, followed by Ω_b, while σ_8 and n_s show a more subtle effect on the profile variation. This pattern is reflected in the correlation coefficients shown in Fig. 6.1, where $\Omega_{\rm m}$ and H_0 have the highest values, followed by Ω_b, with lower coefficients for σ_8 and n_s. This trend persists throughout the study, indicating that the former parameters are inferred more accurately than the latter.

Figs. A.1.1 - A.1.5 in Appendix A.1 showcase the variations across all profile types for all

28 parameters, both astrophysical and cosmological. We can extend the argument above to this greater set of 28 parameters, as we observe that all of them introduce unique features. Even if we do not observe degeneracies among the parameters for any profile type, certain profile types seem to be more sensitive to specific parameters. For instance, astrophysical parameters like *IMF* have a mild effect on the SBx, gas density, and Compton-y profiles, but exert the strongest influence on the metallicity profiles.

Notably, while most profile types exhibit the greatest spread with variations in cosmological parameters, the metallicity profiles are equally, if not more, affected by astrophysical parameters. This is particularly evident with parameters such as the IMF, ASN1, ASN2, and W3. The distinct responses of these profiles to the underlying parameters suggest that their combined effects are key to optimal parameter inference. As demonstrated in Figs. A.1.1 - A.1.5, there are no obvious degeneracies within the same profile type, indicating that the main challenge in parameter inference lies in unraveling the collective impact of multiple parameters—a task where machine learning techniques excel, thereby explaining the model's exceptional performance.

It is important to note, however, that we are working with stacked galaxy cluster profiles, which are noiseless and have a high number of bins (29 or 30, depending on the quantity). Additionally, the use of multiple stacked profiles helps reduce cosmic variance, contributing to the accuracy of our parameter inference.

Finally, we note that although using 300,000 parameter space locations provides the most accurate results, we find that performance nearly saturates when using five profiles with 30,000 locations (see Sec. 6.4.1). As a result, in the following sections, we will proceed with 30,000 parameter space locations when working with all five profile types, and reserve the use of 300,000 locations for cases involving only a single profile type. This balance between accuracy and computational efficiency ensures a robust parameter inference while optimizing the use of resources.

6.3.2 Inference with individual profile types

To discern the unique contributions of each profile type, we conducted independent analyses using 300,000 profiles per type while employing the same architectural framework. The upper left panel of Fig. 6.3 displays the correlation coefficients for each parameter inferred using different profile types.

Generally, some parameters, such as the cosmological parameters $\Omega_{\rm m}$, H_0, and n_s and the astrophysical parameters ASN1, ASN2, and IMF show higher correlation coefficients across most profile types, indicating more robust inference. On the other hand, most of the wind parameters (W1-W8) and black hole (BH1-BH5) parameters tend to have lower correlation coefficients, suggesting they are more challenging to infer accurately.

Each profile type is sensitive to different model parameters and with a different strength:

- **Gas density profiles:** These profiles show high correlation coefficients, particularly for cosmological parameters such as $\Omega_{\rm m}$ and Ω_b. These parameters represent the density ratios of matter and baryonic matter, respectively, to the critical density of the universe. Their strong correlations with gas density suggest that these profiles are still sensitive,

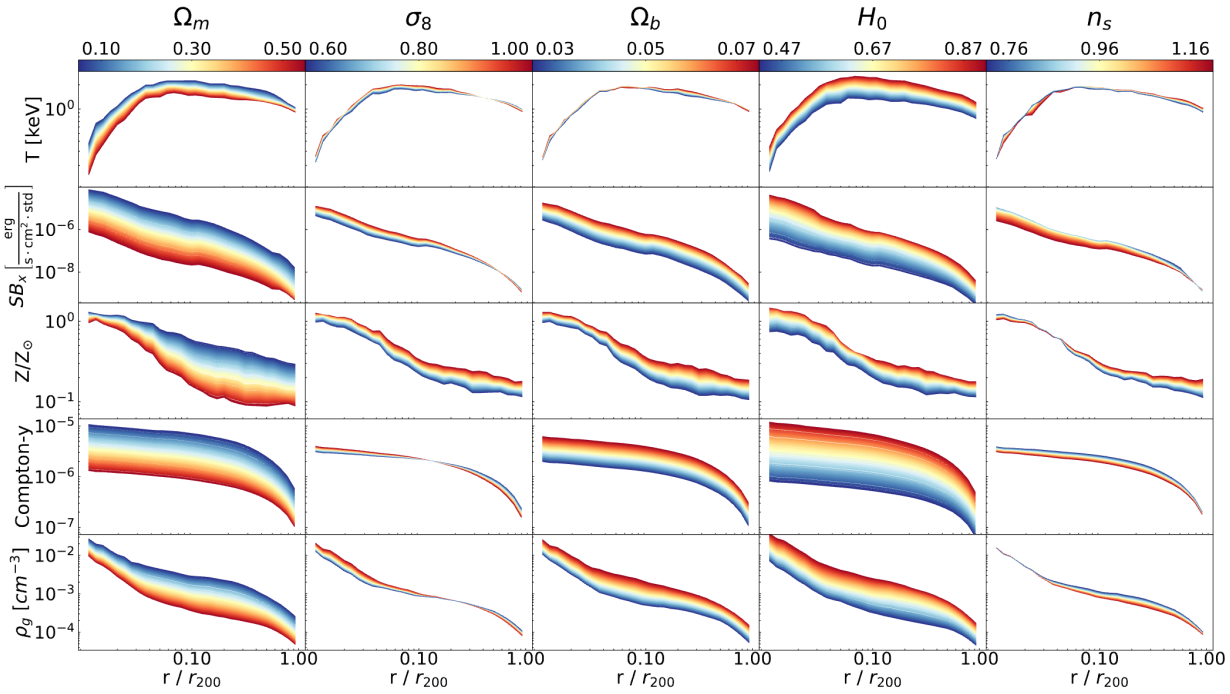


Figure 6.2: Mean galaxy cluster profiles illustrating the effects of varying each cosmological parameter individually. The rows correspond to different profiles: temperature (T), X-ray surface brightness (SB_X), normalized metallicity ($Z/Z_\odot$), Compton-y parameter, and gas density (ρ_g). Each column represents the impact of varying a single cosmological parameter (Ω_m, σ_8, Ω_b, H_0, and n_s) across the full range of their values. The profiles are shown as a function of the radial distance normalized by R_{200}. The color gradient within each panel highlights the spread introduced into the profiles by the variations in the corresponding cosmological parameter, with the color scale representing the range of parameter values. This visualization demonstrates how changes in cosmological parameters affect different physical quantities in galaxy clusters, providing insight into the sensitivity of cluster profiles to cosmological variations.

even if astrophysical phenomena are present, to small variations in the overall mass and baryonic content in clusters, which are directly influenced by these cosmological densities. A similar rationale applies to σ_8, which describes the amplitude of matter fluctuations and is best inferred by the gas density profile, which outperforms the rest of the quantities. The density profile also shows high correlation coefficients for wind parameters, black hole, and black hole feedback parameters. In the case of the wind parameters, these parameters govern aspects of wind energy, momentum, and travel distance (W1-W4) as well as energy reduction and wind particle decoupling (W5-W8). In the case of the BH parameters, they govern the initial BH mass, accretion rates, and the efficiency in energy conversion (BH1-BH5) as well as the energy injection strength and direction of the inhomogeneous feedback mode (Q1-Q2). Thus, all these parameters regulate the distribution of gas inside the cluster, leaving a distinctive imprint on the gas density profiles.

- **Temperature profiles:** The temperature profile performs well across most parameters, particularly cosmological and some feedback parameters. For instance, it outperforms all other quantities in most wind parameters. Galactic winds, driven by SNe, inject energy into the interstellar and ICM, heating the gas and altering its temperature. Temperature is a sensitive indicator of feedback processes, which are critical in regulating star formation and redistributing gas within clusters. Wind-driven feedback mechanisms inject thermal energy into the surrounding gas, raising its temperature. The temperature profile is sensitive to this, reflecting the cumulative effect of these winds over time and space within the cluster. As a result, the temperature profile becomes a key observable in understanding how winds modulate the thermal state of the gas.

- **Metallicity profiles:** The metallicity profiles show weaker performance for cosmological parameters like σ_8, and H_0. This is likely because cosmological parameters govern LSS and the overall mass distribution of the universe, which influence the gas distribution and temperature more directly than metallicity. Metallicity is more of a tracer of the cumulative effects of star formation and feedback processes over time, making it less sensitive to the direct effects of cosmological parameters. Thus, we observe that the metallicity profile tends to perform relatively better for parameters associated with stellar feedback processes (e.g., ASN1, ASN2) and star formation (e.g., IMF). This pattern can be attributed to the fact that the stellar content and the corresponding feedback in the form of SNe can enrich the ICM with metals, thereby altering the metallicity distribution. Since these processes directly influence the metallicity, the metallicity profile is more sensitive to changes in these parameters, leading to higher correlation coefficients.

  Although winds influence metal distribution, their effect on temperature is more pronounced than on metallicity. Metals are less efficiently transported by winds than thermal energy. Due to their higher atomic weight and radiative properties, metals cool more quickly and mix with the ICM in a slower and more localized fashion. This explains why changes in metallicity are more spatially confined than changes in temperature, making metallicity profiles less sensitive to wind parameters.

  We also observe that the metallicity profiles seem to be less sensitive to AGN feedback

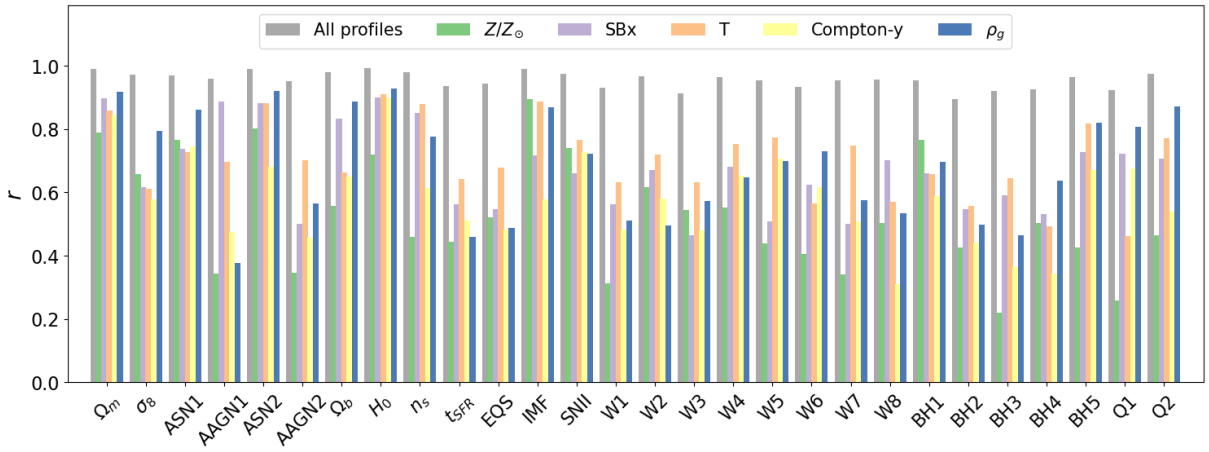


Figure 6.3: Correlation coefficient (r) values for the inference of all IllustrisTNG parameters, analyzed separately for different types of mean galaxy cluster profiles. The different colors represent the various types of profiles used: gas density (ρ_g), temperature (T), X-ray surface brightness (SBx), metallicity (Z), and Compton-y parameter (Y). Each set of bars corresponds to a different IllustrisTNG parameter, listed along the x-axis, with the correlation coefficient on the y-axis. The study uses 300,000 profiles for each type, with 70% allocated for training and the remaining 30% used for validation and testing. The plot illustrates how the choice of profile type affects the accuracy of parameter inference, with varying correlation coefficients observed across the different parameters and profile types.

and BH parameters. The metallicity reflects the long-term history of star formation and SN activity, gradually enriching the ICM with metals. This process occurs over longer timescales compared to the more episodic and intense events associated with AGN feedback. While AGN feedback can have significant short-term impacts on the local environment, such as heating and gas displacement, these effects do not necessarily result in immediate changes in metallicity. The metals might have already been distributed by earlier generations of stars and SNe, with AGN feedback playing a role more in redistributing or removing gas rather than enriching it.

- **X-Ray Surface Brightness profiles:** The surface brightness profiles generally show intermediate correlation coefficients for most parameters. It tends to perform better than Compton-y, but often underperforms compared to gas density and temperature profiles. This suggests that while surface brightness is a useful observable, it may not be sensitive to the full extent of the physical processes influencing certain parameters as effectively as other profiles.

  The surface brightness profile tends to perform relatively well for cosmological parameters such as Ω_m, σ_8, and Ω_b. Since SBx is a measure of the X-ray emission, it is sensitive to the amount and distribution of gas in clusters, which in turn is influenced by the large-scale distribution of matter and the potential well of the clusters which is governed by cosmological parameters. Regions of higher matter density will have more gas, higher

temperatures, and thus stronger X-ray emission. This makes the SBx profile a good tracer of cosmological parameters, especially those that affect the LSS of the universe, such as $\Omega_{\rm m}$ and σ_8. However, because SBx is an integrated quantity along the line of sight, it averages the X-ray emission over the entire cluster. This can smooth out local variations in gas density and temperature, especially those caused by astrophysical processes like feedback from AGNs or SNe. As a result, while SBx is sensitive to the broad effects of cosmological parameters (which impact the overall matter distribution and cluster formation), it may not be as sensitive to localized phenomena like feedback, which affect temperature and gas density more directly.

- **Compton-y profiles:** The Compton-y profiles tend to show lower correlation coefficients across most parameters compared to other profiles like gas density and temperature. This suggests that the Compton-y profile is less effective at capturing the nuances of many parameters in the IllustrisTNG model, particularly those related to feedback and small-scale processes.

  Despite its generally lower performance, the Compton-y profile does exhibit good correlation for parameters that directly influence the large-scale thermal properties of the gas, such as certain cosmological parameters ($\Omega_{\rm m}$, σ_8) and feedback parameters that affect the overall thermal energy content of the ICM.

  The Compton-y profile generally underperforms for feedback-related parameters (e.g., ASN1, ASN2, AAGN1, AAGN2) and wind parameters (W1-W8). This could be because these parameters often involve localized or dynamic processes that impact the temperature and density of the gas in ways that may not significantly alter the overall thermal pressure integrated along the line of sight. Consequently, these effects are not well captured by the Compton-y profile, which averages over these localized variations.

  Nevertheless, although the Compton-y profile appears less effective at capturing detailed, localized processes, its smoother nature may make it less sensitive to the specific subgrid models used in the simulation. As a result, it could still be very valuable when transitioning this method to future observational applications.

6.3.3 Different mass bins

Our previous sections focused on mean cluster profiles for galaxy clusters with a mass of $1 \times 10^{14} M_\odot$. Here, we expand our analysis to explore how predictions vary across different galaxy cluster masses.

To do this, we generated new sets of mean galaxy cluster profiles in 30,000 different parameter space locations, where we treated the mass of the cluster as an extra parameter so that our model has now 29 parameters (28 IllustrisTNG parameters + mass). We varied the value of the mass to cover the full spectrum of galaxy group and cluster masses in our simulations, this is from $1 \times 10^{13} M_\odot$ to $3 \times 10^{14} M_\odot$. The NN's performance in parameter inference across these different masses is illustrated in Fig. 6.4.

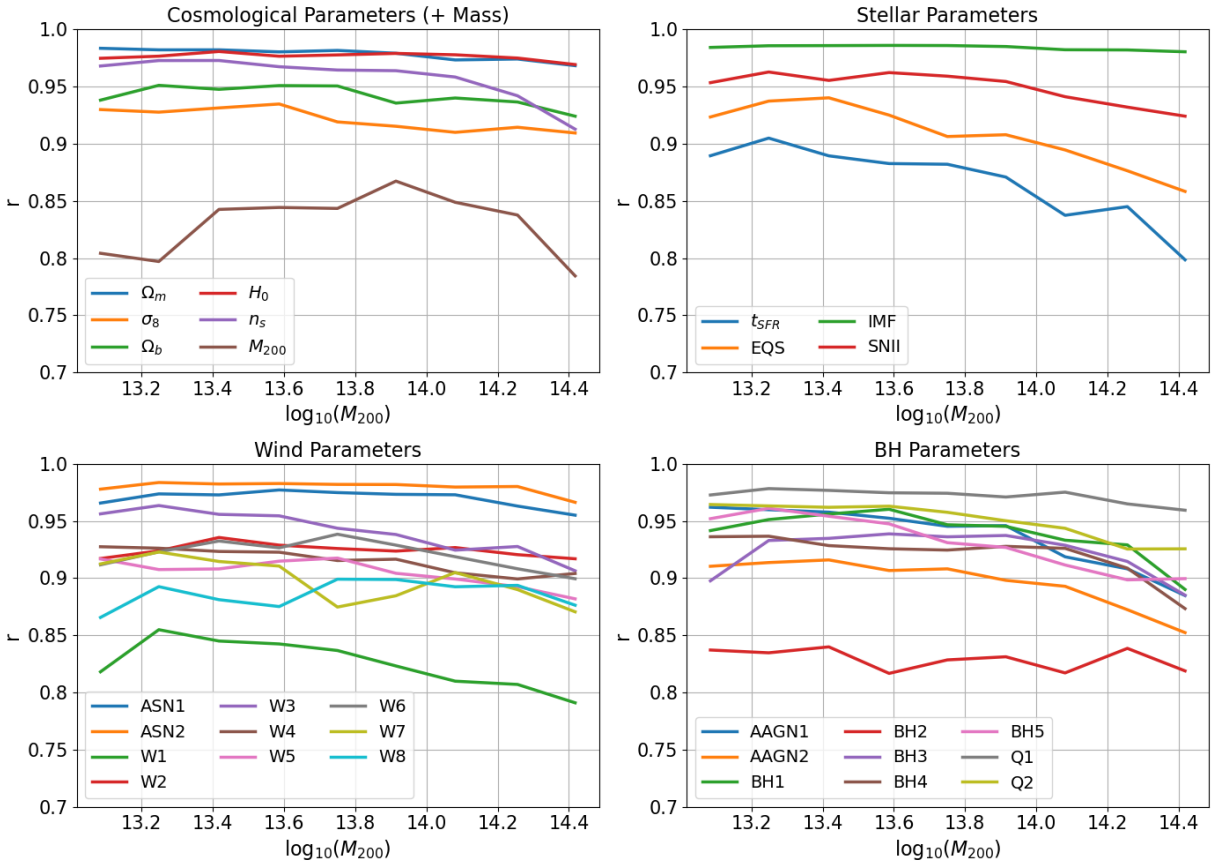


Figure 6.4: Correlation coefficients (r) for the inference of various parameters as a function of galaxy group/cluster mass. Upper left panel: Correlation coefficients for cosmological parameters and mass. Upper right panel: Correlation coefficients for star formation and feedback parameters. Lower left panel: Correlation coefficients for wind parameters. Lower right panel: Correlation coefficients for black hole (BH) and AGN parameters. The x-axis represents the mass of the galaxy groups or clusters, while the y-axis shows the correlation coefficient values. Each line represents a different parameter within each category. The minimum value of the lowest correlation coefficient is approximately 0.8 across all mass bins, indicating that the inference method is robust for galaxy groups and clusters across a mass range of 10^{13} to 3×10^{14} $M_{\odot}$. This robustness suggests reliable parameter inference across the entire mass range studied.

In the upper left panel of Fig. 6.4, the correlation coefficients for cosmological parameters ($\Omega_{\rm m}$, σ_8, Ω_b, H_0, n_s) and mass are consistently high across all mass bins. This indicates that the inference of these parameters is robust across various galaxy group and cluster masses, with minimal variation in accuracy as mass increases. Although the correlation for the mass parameter shows a slight dip, it remains above 0.8 overall, indicating stable performance.

In contrast, the upper right panel of Fig. 6.4 shows that the correlation coefficients for stellar parameters tend to decrease with increasing mass, particularly for parameters like the Maximum Star Formation Timescale (t_{SFR}) and the Factor for Softer EQS. This trend suggests that stellar feedback processes, as captured by these parameters, become more challenging to infer accurately in more massive clusters.

A similar pattern is observed for the wind parameters in the lower left panel and for the BH and AGN feedback parameters in the lower right panel, where most parameters show a slight decrease in correlation coefficients with increasing mass.

This decrease in inference accuracy with increasing mass can be attributed to the fact that, as clusters grow more massive, internal processes such as star formation, SN feedback, and AGN activity have a relatively smaller impact on overall cluster properties. Instead, gravitational dynamics and LSS become the dominant factors driving massive structures closer to self-similarity. As a result, the accuracy of inferring these parameters may decrease with mass, as cluster properties become more governed by simpler, mass-dependent scaling laws rather than detailed processes.

Nevertheless, despite the observed decreases in correlation coefficients with mass, the minimum value of the lowest correlation coefficient is approximately 0.8 across all mass bins. This indicates that the inference method remains generally robust across a wide range of galaxy groups and cluster masses, from 10^{13} to 3×10^{14} $M_\odot$. This robustness suggests that, even though certain parameters are more sensitive to mass and exhibit varying performance, mean galaxy cluster profiles still contain key information about both the processes occurring within these systems and the cosmology of the universe they inhabit, across the entire mass range studied.

6.3.4 Noisy profiles

Galaxy cluster profiles observed in astronomical data are often impacted by various sources of noise, typically quantified by the signal-to-noise ratio (S/N). It is common to perform spectral modeling in each spatial pixel to ensure a minimum S/N, usually having a value of 10 to 100 (Böhringer & Werner, 2010), which is then processed to produce either a projected or deprojected profile. To evaluate how noise affects our parameter inference, we simulate the expected observational noise by introducing Gaussian noise into our profiles. For each bin, we added a specific percentage of random Gaussian noise, creating datasets with noise levels of 10%, 20%, 30%, and 40%, corresponding to S/N ratios of 10, 5, 3.3, and 2.5, respectively. This approach allows us to study the limit of S/N = 10, above which we anticipate improved results, while also examining how inference accuracy degrades with increasingly noisy data. These modified profiles were then used for the training, validation, and testing of our model.

The effect of the introduced noise in the parameter inference is depicted in Fig. 6.5. As Gaussian noise increases from 0% to 40%, there is a general decline in the correlation coefficients

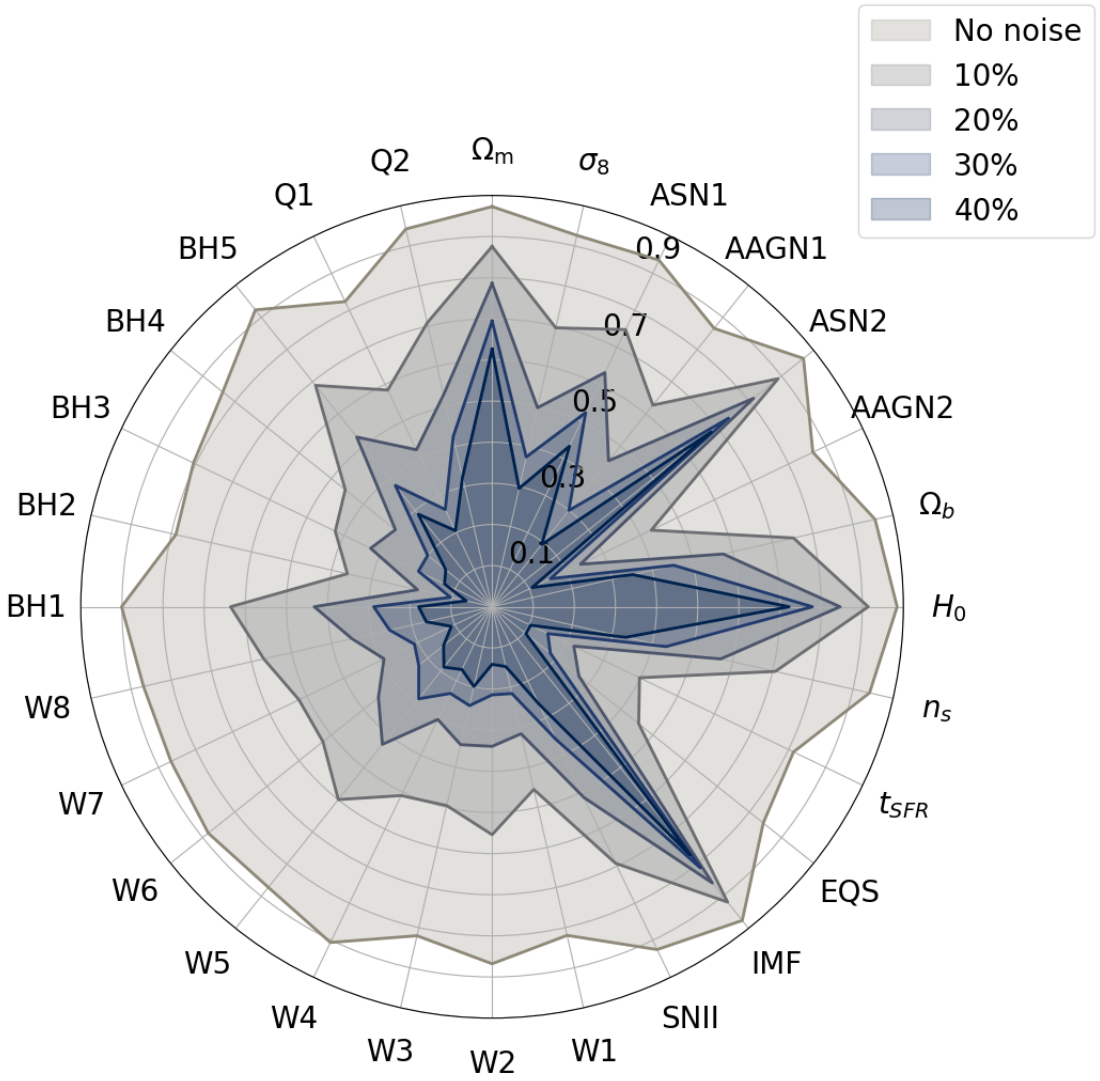


Figure 6.5: Radar chart displaying the correlation coefficients for various IllustrisTNG parameters under different levels of Gaussian noise. The radial distance from the center represents the correlation coefficient, with higher values indicating stronger correlations. Each parameter is represented by a specific symbol on the perimeter of the chart, corresponding to the different physical or cosmological parameters within the IllustrisTNG model. The different shaded regions correspond to varying noise levels, ranging from no noise to 40% Gaussian noise. As noise increases, the correlation coefficients generally decrease, indicating a reduction in the accuracy of parameter inference. The chart visually demonstrates how increasing noise impacts the reliability of inferred parameters, with the outermost regions (no noise) showing the highest correlation, and the innermost regions (40% noise) showing the lowest correlation.

for all parameters, indicating that noise consistently reduces the accuracy of parameter inference. However, the extent of this decline differs between the parameters. For example, parameters like $\Omega_{\rm m}$, Ω_b, and H_0 maintain relatively high correlations even at 40% noise, suggesting that these cosmological parameters are more robust to noise. Similarly, certain astrophysical parameters, such as ASN2 and IMF, also exhibit high correlation levels despite the added noise.

In contrast, other parameters, such as W7 among the wind parameters, and Q1 and BH2 among the black hole parameters, experience a sharper decline in correlation coefficients, indicating greater sensitivity to noise. This suggests that the inference of these parameters is more vulnerable to the degradation of input data quality.

Parameters associated with similar physical processes or feedback mechanisms tend to exhibit similar patterns of sensitivity to noise. For example, many wind parameters (W1-W8) and black hole feedback parameters (BH1-BH5) show a clustered response to increasing noise, with significant declines in correlation, suggesting that certain processes are uniformly more challenging to infer accurately in the presence of noise.

The relative ranking of parameters by their correlation coefficients remains fairly stable across different noise levels. Parameters that start with high correlations in the no-noise scenario (e.g., $\Omega_{\rm m}$, n_s) generally maintain higher correlations even as noise increases, compared to those that begin with lower correlations. This consistency implies that while noise affects all parameters, those that are inherently more robust—likely due to their strong influence on large-scale cluster properties—remain more reliable even under noisy conditions.

This analysis suggests that when noise is introduced, the results are driven less by the non-degeneracy of parameters and more by the strength of each parameter's influence on the profiles. Some parameters may only affect the profiles in specific ranges where the data is highly correlated. If this correlation is disrupted by noise, the network's performance declines. However, certain cosmological parameters like $\Omega_{\rm m}$, Ω_b, and H_0, as well as astrophysical parameters such as IMF and ASN2, have an impact on the profiles that remains robust to noise. We note that these parameters are also the ones that can be inferred well with lower training data, as shown in Fig. 6.9 of appendix 6.4.1. These parameters generate significant variation in the profiles (see Figs. A.1.1 - A.1.5), and the strength of this correlation allows them to be inferred accurately even under extreme noise conditions.

6.3.5 Inference accuracy and radial cut

To determine which regions within galaxy clusters provide the most significant information for parameter inference, we performed an analysis by progressively excluding the outer parts of the cluster. Specifically, we applied radial cuts at 70%, 50%, 27%, and 10% of the virial radius, thereby removing the outer regions beyond these thresholds. Fig. 6.6 shows the correlation coefficients obtained for each of these radial cuts compared to using the full virial radius. The correlation coefficients (r) for various IllustrisTNG parameters are plotted as a function of different radial cuts within the cluster, ranging from 1.0 R_{200} (i.e., the entire virial radius) down to 0.1 R_{200}. Each set of bars represents a different parameter, with the color gradient indicating the specific radial cut used in the analysis.

Figure 6.6 shows that as the radial cut decreases (from 1.0 R_{200} to 0.1 R_{200}), the correlation

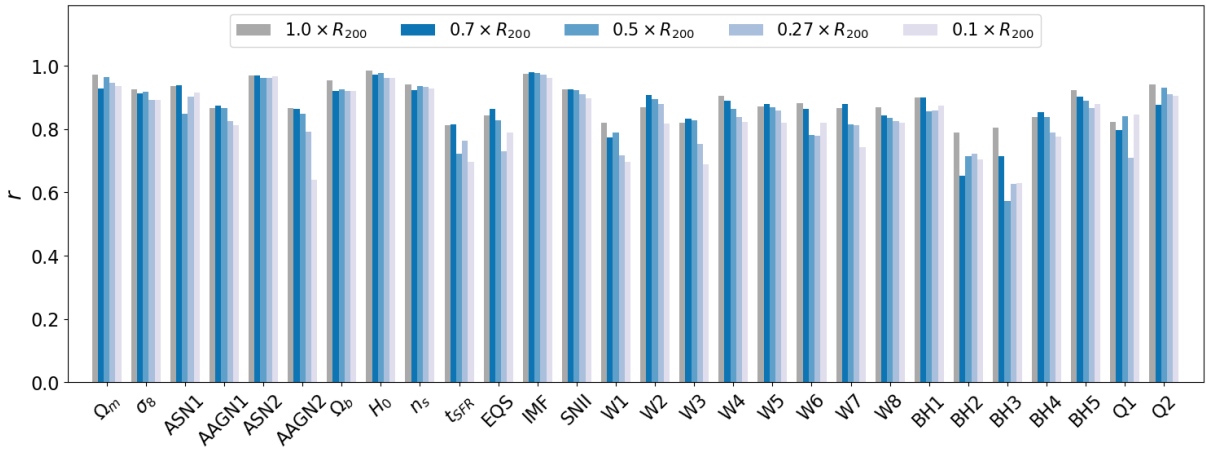


Figure 6.6: Bar chart displaying the correlation coefficients (r) for various IllustrisTNG parameters as a function of different radial cuts within the cluster, ranging from $1.0 \times R_{200}$ to $0.1 \times R_{200}$. The parameters are listed along the x-axis, with each set of bars representing a different parameter, while the color gradient indicates the radial cut used for the analysis. The results show that while correlation coefficients generally decrease with smaller radial cuts, the overall change is not very large. This suggests that a significant portion of the information required for accurate parameter inference is concentrated within the innermost $0.1 \times R_{200}$ of the cluster's radius.

coefficients generally decrease for most parameters. However, this decline is modest, suggesting that while the outer regions do contribute to parameter inference, the bulk of the relevant information is concentrated in the inner region, particularly within the innermost 0.1 R_{200}. Interestingly, for some parameters, shorter profiles yield slightly better results, which is counterintuitive since reducing the number of bins should reduce the information available to the network. We attribute this to the epistemic uncertainty in the network, arising from the inherent limitations of any model, and improvements to the architecture could potentially reduce this uncertainty. These results suggest that although parameter inference accuracy declines slightly as the radial cut decreases, strong results can still be achieved for most parameters, even when the radius is truncated to 10%.

The upper panel of Fig. 6.7 compares the correlation coefficients (r) for various IllustrisTNG parameters when using only the innermost 0.1 R_{200} region (light purple) versus the combined outer region from 0.1 to 1.0 R_{200} (dark purple). Each set of bars corresponds to a different parameter, with the correlation coefficient displayed on the y-axis. The results demonstrate that the innermost 0.1 R_{200} region generally yields higher correlation coefficients across most parameters, indicating that this central region contains the most critical information for accurate parameter inference. The outer region contributes less to the overall correlation, as reflected by the lower coefficients when this region is used in isolation.

The lower right plot offers a similar comparison, this time between the central 0.27 R_{200} region (light blue) and the outer shell from 0.27 to 1.0 R_{200} (dark blue). Similar to the previous comparison, the central 0.27 R_{200} region shows higher correlation coefficients than the outer shell. The outer shell exhibits a more pronounced drop in correlation coefficients compared to

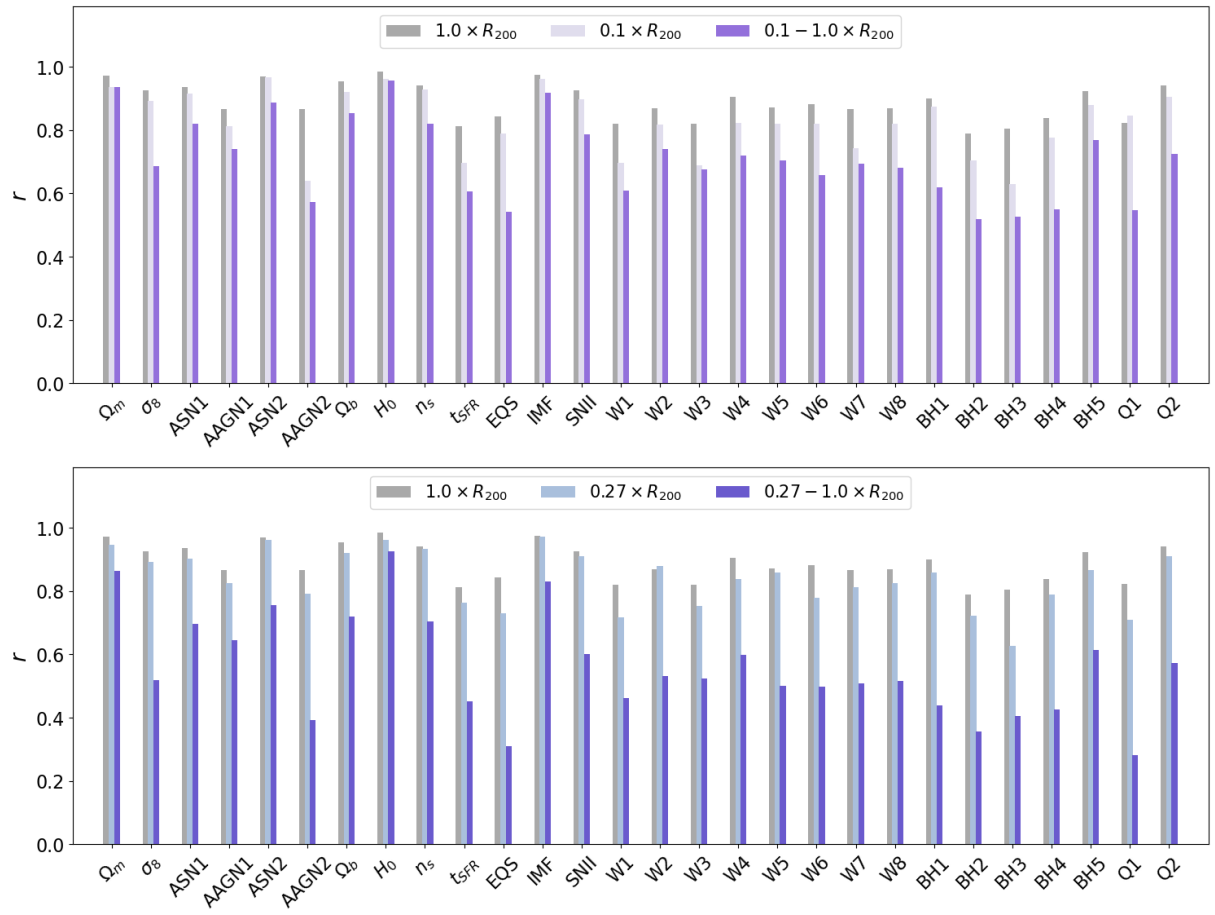


Figure 6.7: Bar charts comparing the correlation coefficients (r) for various IllustrisTNG parameters across different radial ranges within the cluster. The left panel shows the correlation coefficients for the full radial range ($1.0 \times R_{200}$, gray), the inner region ($0.1 \times R_{200}$, light purple), and the outer shell between $0.1 \times R_{200}$ and $1.0 \times R_{200}$ (dark purple). The right panel displays the correlation coefficients for the full radial range ($1.0 \times R_{200}$, gray), the central region ($0.27 \times R_{200}$, light blue), and the outer shell between $0.27 \times R_{200}$ and $1.0 \times R_{200}$ (dark blue). In both panels, each set of bars represents a different parameter, with the correlation coefficient on the y-axis. The results indicate that the central regions (up to $0.1 \times R_{200}$ or $0.27 \times R_{200}$) generally provide the highest correlation coefficients, suggesting they contain the most critical information for accurate parameter inference. The outer shells show a significant drop in correlation, especially in the right panel, highlighting the diminished contribution of the outer regions to the overall accuracy. These findings stress the importance of the inner cluster regions in the inference of the model parameters.

Figure 6.8: Radar charts showing the correlation coefficients for various IllustrisTNG parameters, with the radial axis representing the correlation coefficient values. Each chart compares the results obtained from using the profile of a specific quantity (light-shaded region) against those obtained from the integrated quantity up to R_{200} (dark-shaded region). The quantities examined are (top left) metallicity ($Z/Z_\odot$), (top center) temperature (T), (top right) surface brightness (S_X), (bottom left) Compton-y parameter, and (bottom right) gas density (ρ_g). The light-shaded regions generally encompass larger areas, indicating that using profiles provides higher correlation coefficients and, consequently, more accurate parameter inference compared to using integrated quantities.

the central region, highlighting that while it adds some information, its contribution to overall accuracy is less significant. This reinforces the conclusion that the cores of galaxy clusters are fundamental for accurate parameter determination, which is not entirely unexpected since cluster cores tend to be more affected by astrophysical processes such as cooling, strong star formation bursts, and the effects of AGN feedback.

6.3.6 Integrated quantities

Previous studies have focused on inferring some of the IllustrisTNG parameters using the integrated Y_{200c} (Lee et al., 2024). In this work, we extend the analysis to evaluate the inference capabilities of all integrated quantities and compare them to the capabilities derived from using the full cluster profile. To achieve this, we generate average galaxy cluster profiles in a specific parameter space location for 30 different masses, uniformly spaced in log space from 10^{13} to 3×10^{14} $M_\odot$. We select a total of 300,000 of these parameter space locations and create 30 averaged galaxy cluster profiles for each of them.

Our input vector, representing a single location in parameter space, consists of one profile per mass bin, with each profile containing either 30 or 29 bins, depending on the quantity, resulting in a total of 900 or 870 values. For the integrated quantities, all profiles are collapsed into a single value, so each input vector will contain 30 values, one for each mass bin.

Figure 6.8 demonstrates the correlation coefficients for the IllustrisTNG parameters, comparing the performance of using full galaxy cluster profiles (light shaded regions) versus integrated quantities up to R_{200} (dark shaded regions) for each individual quantity. Across all five radar charts, it is evident that using the full profiles generally results in higher correlation coefficients, implying more accurate parameter inference compared to the use of integrated quantities. The size of the light-shaded regions relative to the dark-shaded ones provides a visual representation of this difference.

The extent to which the profile outperforms the integrated quantity varies by parameter. Temperature profiles stand out most, as profiles achieve correlation coefficients above 0.9 for all parameters, while the integrated quantity does not surpass 0.6. This suggests that the detailed temperature profile carries a wealth of information that is lost when collapsing the profile into a single integrated quantity, highlighting the importance of spatial variations in accurately constraining model parameters. These results are followed by the gas density, metallicity, and surface brightness, which also show a better correlation coefficient for the profiles than for the integrated quantities.

For the Compton-y parameter, the difference between full profiles and integrated quantities is less pronounced. While full profiles still offer a slight advantage, the performance gap is narrower. This can be explained by the smoother nature of Compton-y profiles, which exhibit fewer localized features compared to other quantities. As Compton-y represents the overall thermal pressure integrated along the line of sight, it is a cumulative measurement. This integration leads to a smoothing effect, diminishing the sharp gradients or features observed in other profile types. As a result, the Compton-y profiles lack the detailed localized features that could be exploited for more precise parameter inference. Consequently, reducing the profile to a single integrated value leads to a smaller difference in the network's ability to infer the model parameters.

Previous studies have shown that when using Y_{200c}, ASN2 was the best-inferred parameter among AAGN1, AAGN2, ASN1, and ASN2 (Lee et al., 2024). This result is reaffirmed here for Y_{200c} and extends to the results for T_{200c} and ρ_g, further supporting the robustness of these parameters.

Overall, these findings, while consistent with prior research, demonstrate that using the full profile generally offers a clear advantage in achieving higher accuracy in parameter inference across all physical quantities and parameters. This emphasizes the value of incorporating full profiles in astrophysical analyses, particularly for parameters where spatial variations within the cluster are fundamental for accurate modeling and inference.

6.4 Further Considerations

6.4.1 Inference Accuracy and Dataset Size

In this study, CARPoolGP's primary function is to generate a comprehensive dataset of mean galaxy cluster profiles for training, validating, and testing our NN architecture. The optimal number of profiles needed to achieve accurate model performance while maintaining computational efficiency is not well-defined. To address this, we evaluated our NN's performance using all five profile types together using different parameter space points: 300 (210 for training), 3,000 (2100 for training), 30,000 (21000 for training), and 300,000 (210000 for training).

Fig. 6.9 illustrates the correlation coefficients (r) for various IllustrisTNG parameters as a function of the input size used for model training, ranging from 300 to 300,000 parameter space locations. Notably, each input size involves the use of five distinct profiles per galaxy cluster—covering gas density, temperature, X-ray surface brightness, metallicity, and Compton-y parameters. The radar chart depicts how the accuracy of parameter inference improves with increasing input size, as indicated by the outward expansion of the shaded regions corresponding to each input size.

As the input size grows, there is a clear improvement in the correlation coefficients across all parameters. The analysis shows that with just 300 parameter space locations, the average correlation coefficient starts at a modest 0.5, but increases significantly to around 0.9 with 30,000 parameter space locations, reaching a point of apparent saturation. Beyond this threshold, moving to 300,000 parameter space locations yields only marginal gains, indicating diminishing returns with larger datasets. This suggests that using 30,000 parameter space locations with five profiles each is sufficient to capture most of the relevant information for robust parameter inference.

While the increased input size benefits most parameters, the extent of improvement varies. Cosmological parameters such as $\Omega_{\rm m}$, Ω_b, and H_0 maintain relatively stable and high correlation coefficients across all input sizes, highlighting their strong influence on cluster profiles. In contrast, some astrophysical parameters, especially those linked to complex feedback mechanisms (e.g., BH and AGN parameters), show more significant gains with larger input sizes, reflecting the greater complexity and variability in their effects on the profiles.

Certain parameters, like ASN2 and IMF, which are closely related to stellar content and feedback processes within clusters, exhibit high correlation coefficients even with smaller input

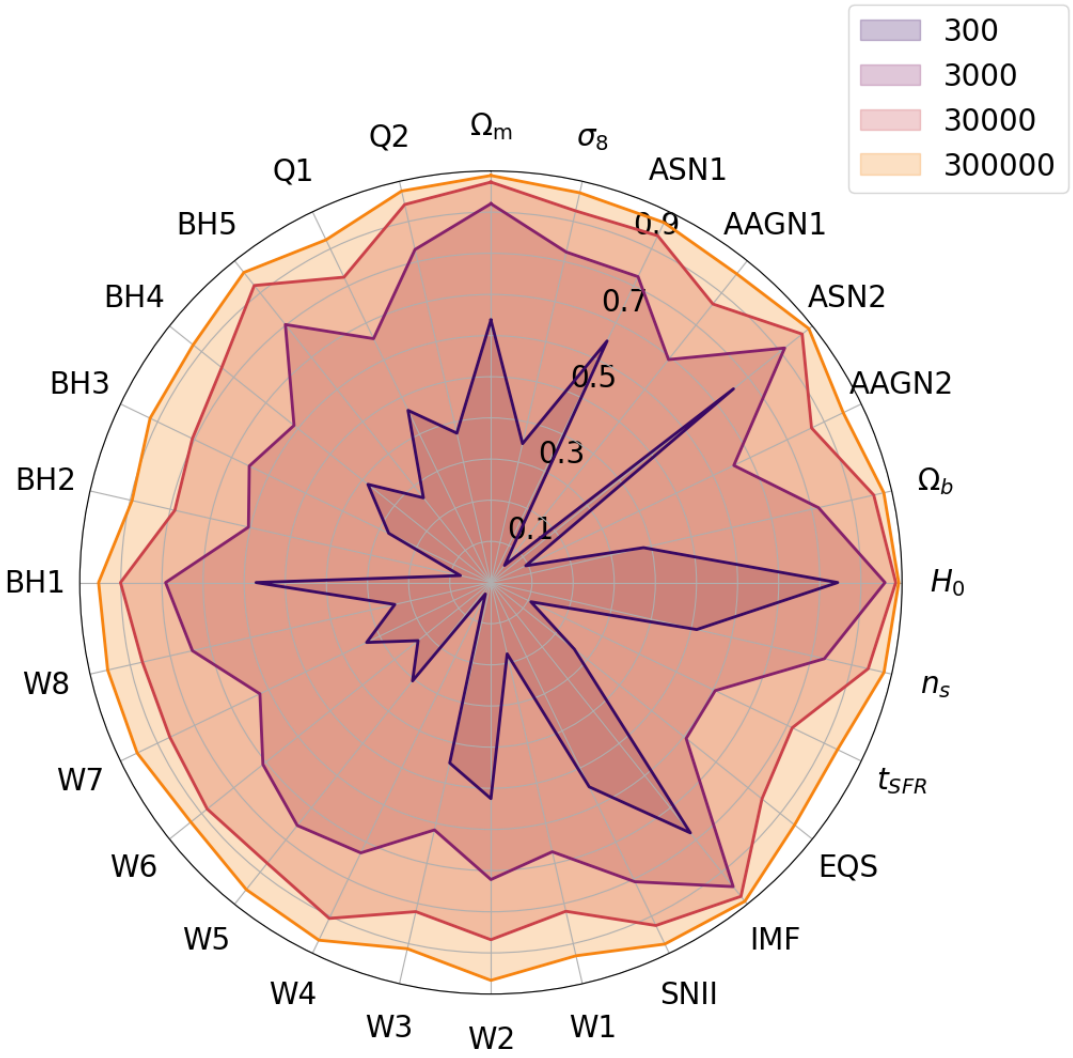


Figure 6.9: Radar chart displaying the correlation coefficients for various IllustrisTNG parameters as a function of input size. The radial axis represents the correlation coefficient values, with different symbols around the perimeter corresponding to specific IllustrisTNG parameters, including cosmological and astrophysical quantities. The shaded regions represent the results for different input sizes, ranging from 300 to 300,000 profiles. As the input size increases, the shaded regions expand outward, indicating improved correlation coefficients. Notably, the 30,000 profile input size approaches a saturation point, beyond which further increases in input size yield only marginal improvements in accuracy. This suggests that 30,000 profiles are sufficient to achieve robust parameter inference, with larger datasets providing only slight gains in accuracy across all parameters.

sizes. This robustness suggests that these parameters exert a pronounced impact on the profiles, making them easier to infer accurately with fewer data points.

The findings from Fig. 6.9 have significant implications for our NN training strategy. While using 300,000 parameter space locations may be beneficial for inference with a single profile type, the saturation point identified at around 30,000 parameter space locations suggests that training can be optimized by focusing on this input size when utilizing all five profile types together, achieving a balance between computational efficiency and inference accuracy. Consequently, we will use this number of profiles in the subsequent steps of our analysis.

6.4.2 Consistency Tests

Figure 6.10: Correlation coefficients for the inference of 28 IllustrisTNG model parameters under two different scenarios. The grey bars represent the results using full galaxy cluster profiles with 30 bins per profile type and 30,000 parameter space locations, providing sufficient information to infer all parameters with high accuracy. The red bars show the results for Test A, where the number of input values is reduced to 25 (5 bins per profile type). As expected, the NN fails to infer most parameters, accurately predicting only H_0, the IMF slope, and ASN2. Test B is represented by the orange colored bars, where only one bin for each quantity (a total of 5 bins) was used, resulting in even lower correlation coefficients for all quantities. The green bars show the results of Test C, where the tetsing was performed on corrupted data, yielding correlation coefficients close to 0 for most parameters.

To verify that our NN is not affected by any information leaks, we perform a series of tests.

Test A uses 5 bins per quantity, resulting in 25 input values for each parameter space. Since it is not feasible to infer n parameters with fewer input values ($N < n$), we expect the NN to struggle when trying to infer 28 parameters from only 25 input values. We assess the NN's performance across 30,000 parameter space locations. The results, shown in Fig. 6.10, compare the NN's performance with full galaxy cluster profiles (grey bars), which provide enough bins to infer all parameters, versus using only 25 input values (red bars). As expected, the NN fails to infer most

parameters, successfully predicting only three—H_0, the IMF slope, and ASN2—with moderate accuracy. Fig. 6.12 compares predicted and actual values, showing that while the trend is correct for these three parameters, there is more scatter and higher standard deviation compared to the results using full profiles (Fig. 6.11). With insufficient input data, the network is unable to infer the remaining parameters.

Test B reduces the input vector size further by selecting only the first bin for each quantity, resulting in just 5 input values. The brown bars in Fig. 6.10 display the correlation coefficients, showing even further degradation in the NN's performance. Fig. 6.13 compares predicted and actual values, highlighting that even for the best-inferred parameters such as H_0, the IMF slope, and ASN2, scatter and uncertainty increase significantly, while the rest of the parameters are not inferred at all.

Test C examines the impact of corrupted data. Here, the network is trained using full profiles, but during testing, one profile is corrupted by adding random noise, effectively erasing its original values. The green bars in Fig. 6.10 display the correlation coefficients for this test, where the NN clearly fails to infer any parameters from the corrupted data. Fig. 6.14 shows predicted values versus the mean for each quantity, demonstrating that the network completely misses its predictions when faced with corrupted data.

In conclusion, these tests confirm that our NN architecture does not suffer from information leaks, and the results in our study reflect physical phenomena rather than model artifacts.

6.5 Discussion and Conclusions

In this study, we investigated the inference of cosmological and astrophysical parameters using averaged galaxy cluster profiles generated from the IllustrisTNG simulations. We performed parameter inference across the 28-dimensional IllustrisTNG model parameter space. The results show that stacked galaxy cluster profiles possess strong constraining power over this multidimensional parameter space, containing enough information to accurately infer all cosmological and astrophysical parameters of the model.

1. Stacked galaxy cluster profiles allow us to infer all cosmological and astrophysical parameters of the IllustrisTNG model with high accuracy in the noiseless case with 29 to 30 bins per profile. Including gas density, temperature, X-ray surface brightness, metallicity, and Compton-y profiles, we obtain coefficients approaching 0.97 for all cosmological parameters. For the remaining quantities, correlation coefficients remain above 0.90, reflecting very high accuracy across all parameters.

2. Our analysis reveals that different profile types exhibit varying sensitivities to specific parameters. For instance, temperature profiles generally provide the highest correlation coefficients for wind and feedback parameters, likely due to their sensitivity to energy injection processes within clusters. Gas density profiles also perform well, especially for cosmological parameters, suggesting that these profiles are highly responsive to the overall matter distribution within clusters. Conversely, the Compton-y profiles tend to show lower

correlation coefficients, particularly for parameters related to localized processes, though their smoother nature may make them less susceptible to the specifics of subgrid models.

3. We observed that the inference accuracy for certain parameters diminishes as the cluster mass increases, which may be attributed to the self-similar nature of more massive clusters. These systems are less influenced by localized processes, such as star formation and feedback, leading to a reduced sensitivity to the parameters governing these processes.

4. Radial cuts within clusters reveal that, based on our results using TNG28, the majority of the information required for accurate parameter inference is concentrated within the innermost regions, particularly within $0.1\,R_{200,c}$. Interestingly, while cluster cores typically dominate the overall cluster emission, their underlying physics remains incompletely developed and unconverged across simulations. This finding highlights the importance of focusing on the central regions of clusters in future simulation efforts and observational studies to gain a deeper understanding of their physical processes and to improve their use for cosmological information extraction.

5. The addition of Gaussian noise to the profiles resulted in a general decline in correlation coefficients across all parameters, highlighting the sensitivity of the inference process to data quality. However, certain cosmological parameters, such as $\Omega_{\rm m}$, Ω_b, and H_0, along with specific astrophysical parameters like ASN2 and IMF, maintained relatively high correlation levels even under significant noise. These parameters also generate a greater spread in the profiles, suggesting that both the non-degeneracy of the parameters and the strength of their influence on the profiles are key factors when dealing with noise. It is important to note that with a typical lower S/N value of 10, the inference remains highly accurate, with correlation coefficients exceeding 0.7 for these influential parameters. Even at an extremely low S/N ratio of 2.5, the impact on these parameters is minimal, indicating that their strong effects on the profiles make them more resistant to noise.

6. When comparing the inference capabilities of integrated quantities to full cluster profiles, our results show that full profiles generally provide a more accurate inference, particularly for parameters with significant spatial variations within the cluster. The integrated quantities, while consistent across different parameters, tend to underperform compared to the full profiles. This highlights the importance of considering the full spatial information contained within cluster profiles for precise parameter inference.

Our findings indicate that stacked galaxy cluster profiles contain important information that allows us to effectively disentangle the effects of cosmology and astrophysical processes, providing accurate inference across a wide range of parameters. The robustness of our results across different mass bins, noise levels, and radial cuts suggests the potential of using stacked galaxy cluster profiles to maximize the extraction of cosmological and astrophysical information and increases their utility for cosmological parameter inference.

These results stress the value of detailed profile analysis in astrophysical and cosmological research, particularly in the context of upcoming large-scale surveys. Future work could expand

this approach to include more diverse simulation sets and even include observational data, further refining the accuracy and applicability of parameter inference in cosmology for the next-generation surveys.

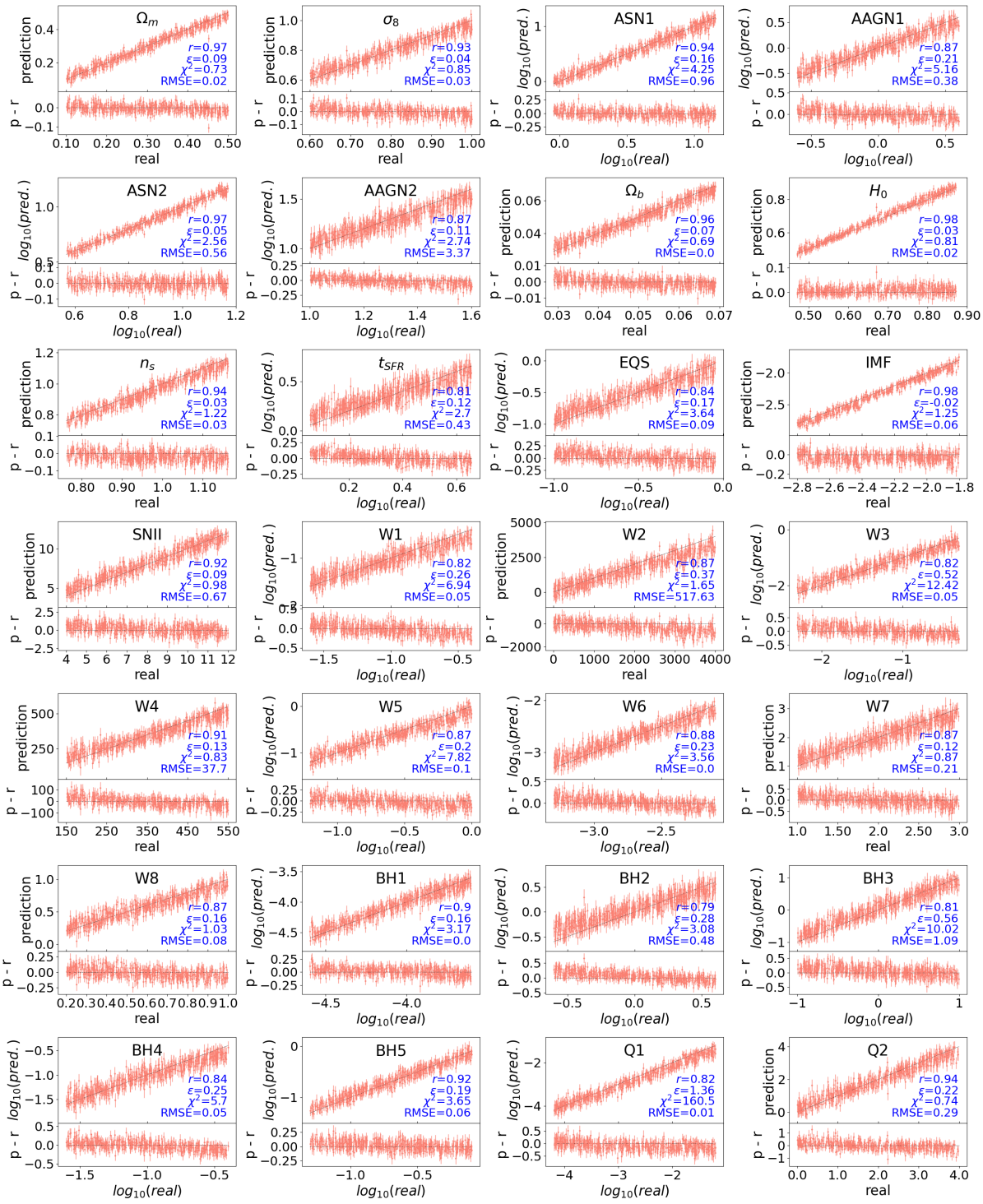


Figure 6.11: Inference results for all 28 IllustrisTNG model parameters using complete galaxy cluster profiles with full information across 30 bins. Each top panel shows the comparison between the true values (x-axis) and the predicted values (y-axis), with a black line indicating the one-to-one correspondence. The bottom of each panel displays the residuals to illustrate the spread. The correlation coefficient (r) and root-mean-square error (RMSE) mean relative error (ϵ) and reduced chi-squared (χ^2) are reported in blue for each parameter. These results demonstrate high accuracy across all parameters, with strong correlation coefficients and relatively low RMSE values. The accurate predictions underline the ability of the model to infer both cosmological and astrophysical parameters when provided with full, detailed profiles, as reflected by the close alignment between predicted and true values and the small residuals across the parameter space.

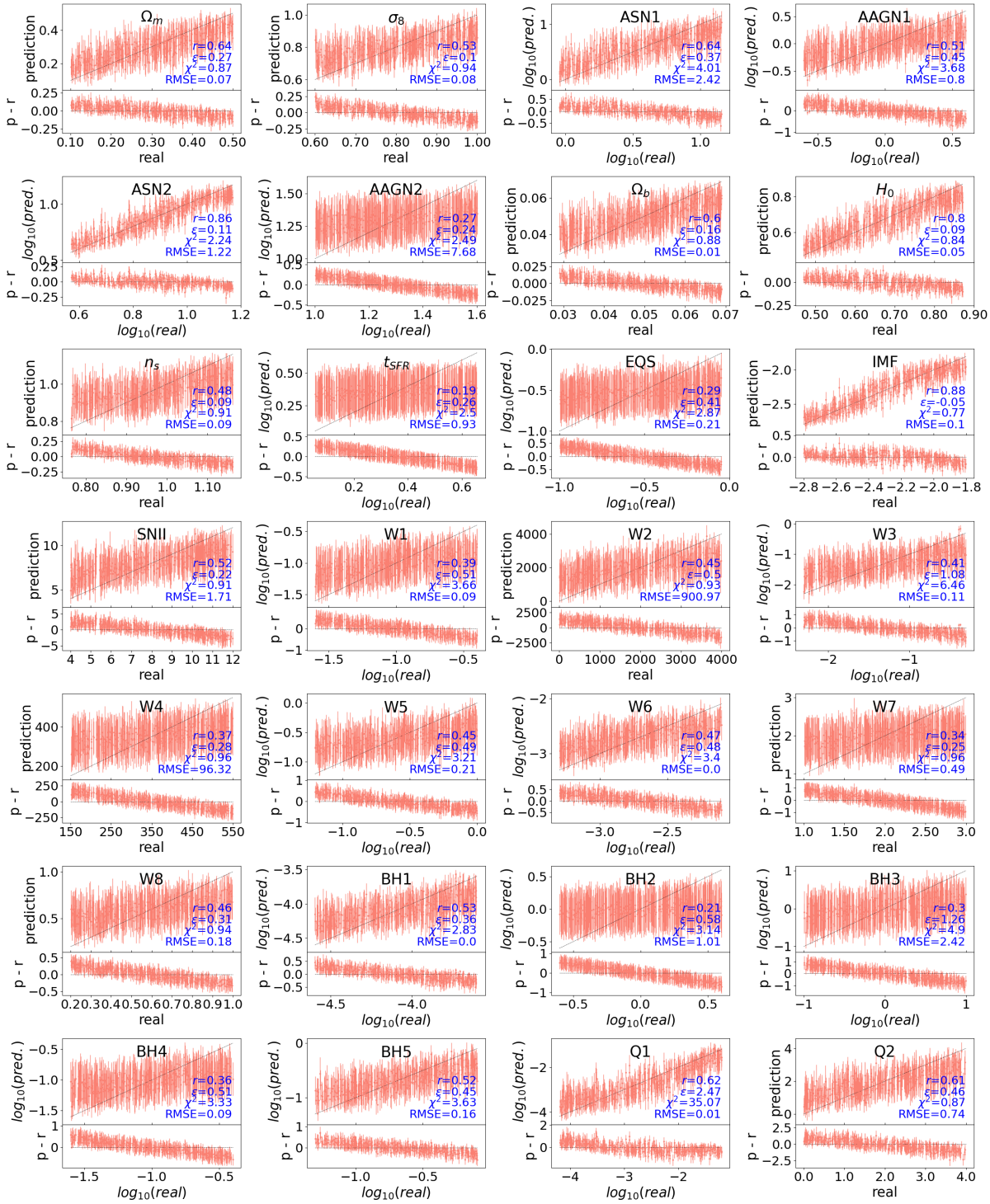


Figure 6.12: Inference results for test A. Similar to Fig. 6.11, but here the results show the decrease in inference capabilities of the network when using an input size N=25. Only ASN2 and H_0 are infered with some moderate accuracy.

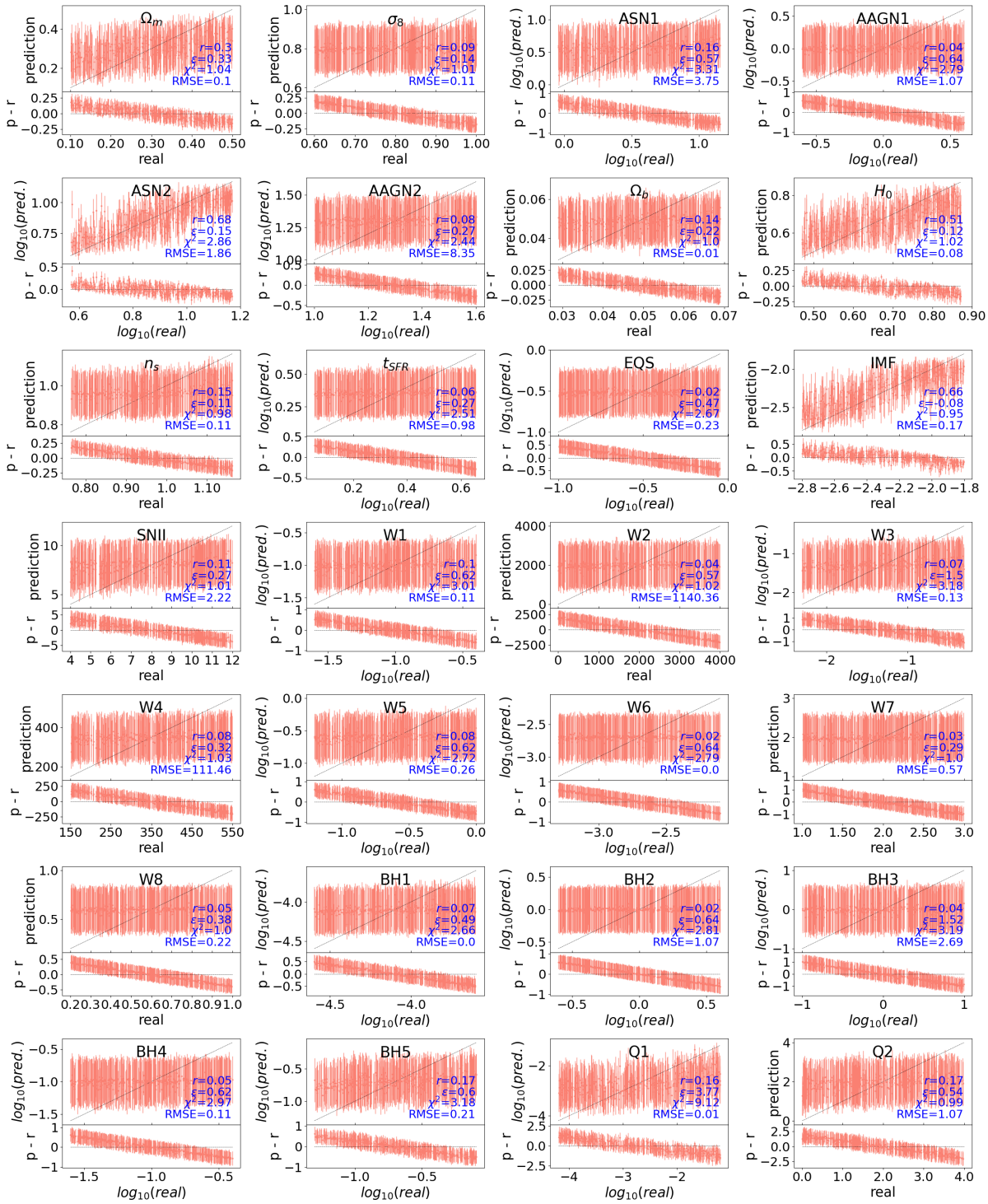


Figure 6.13: Inference results for test B. Similar to Fig. 6.11, but here the results show how the NN is not capable of inferring any parameter when using an input size of N=5.

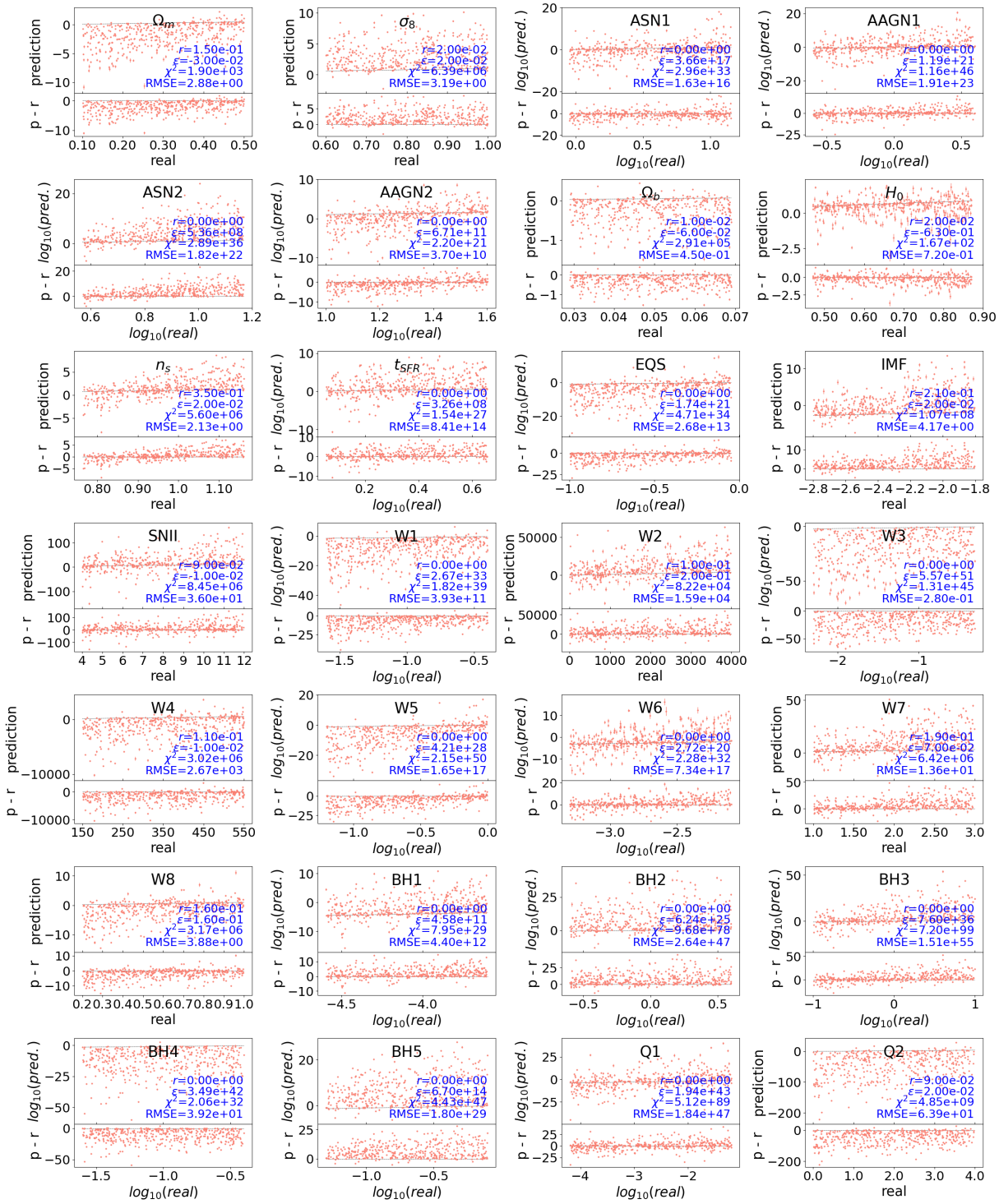


Figure 6.14: Inference results for test C. Similar to Fig. 6.11. The results show that when using corrupted data for testing, the NN is not capable of inferring any of the parameters.

Part V

Outlook & Conclusion

7 | Conclusion

In this thesis, we worked on advancing the understanding of the thermodynamic structure, assembly histories, and cosmological implications of galaxy clusters, though the use of constrained simulations and ML-based inference. Doing this, this work has established new methodologies for linking simulations to observations and has opened new pathways of using galaxy clusters as precise probes of cosmology and galaxy formation physics. The main findings and contributions of this work are summarized below.

Reproducing Observed Galaxy Clusters with Constrained Simulations: A central contribution of this work lies in demonstrating the efficacy of constrained simulations in reproducing the physical properties of individual galaxy clusters in the local Universe. As presented in Chapter 4, through a detailed comparison of 46 observed clusters with their simulated counterparts in the SLOW simulation suite, we showed that the simulations successfully match global properties such as mass, X-ray luminosity, temperature, and SZ signal, with a strong agreement for prominent systems like Virgo, Coma, and Perseus. Our study using one-to-one mass comparisons between observations and simulations supports a hydrostatic mass bias estimate of $(1-b) \simeq 0.87$ and indicates that constrained realizations can accurately reflect both the statistical and spatial diversity of real clusters. Importantly, the observed discrepancies between simulations and data are now comparable in scale to observational uncertainties, suggesting that we may be reaching the limit of current measurement precision rather than that of the simulations themselves.

Understanding Thermodynamic Structure and Formation Histories: Building on this validation, Chapter 5 of these thesis examined the thermodynamic structure and formation histories of these galaxy clusters. By analyzing pressure, temperature, entropy, and electron density profiles, we found that the local galaxy clusters are generally modeled with fidelity, particularly at large and intermediate radii. However, persistent discrepancies in cluster cores, such as flattened entropy and elevated densities, highlight the continued limitations of current subgrid models, especially those governing AGN feedback and radiative cooling. We established a physical link between the thermodynamic state of cluster cores and their mass assembly histories: clusters that formed early tend to host cool cores, while late-forming systems are more likely to be disturbed and non-cool-core. This connection provides a physically motivated explanation for the observed diversity in cluster core structures.

Performing High-Accuracy Simulation-Based Inference: The final part of this thesis, presented in Chapter 6, advances the use of simulations not only as a tool for reproducing observations, but also for inferring the underlying cosmological and astrophysical parameters driving cluster formation and evolution. Using the CAMELS-zoomGZ suite, which spans a 28-dimensional parameter space in the IllustrisTNG model, we trained neural networks on stacked galaxy cluster profiles to perform likelihood-free inference of cosmological parameters (e.g., Ω_m, H_0, σ_8) and baryonic physics parameters (e.g., IMF slope, feedback efficiencies). These models achieved correlation coefficients of 0.97 or higher for cosmological parameters and above 0.90 for most astrophysical ones, demonstrating that spatially resolved cluster profiles are powerful observables for parameter inference. Furthermore, we showed that inner cluster regions, particularly within $0.1\,R_{200}$, carry the bulk of the constraining power. Full thermodynamic profiles consistently outperformed integrated quantities up to r_{200}, underscoring the importance of preserving spatial structure in the data. Importantly, the robustness of these results was tested under realistic degradations, including radial truncation and substantial noise. Even with signal-to-noise ratios of 10 or worse, key parameters such as Ω_m, H_0, and the IMF slope could still be inferred with high confidence. This establishes galaxy cluster profiles as powerful observables for cosmological inference and lays the groundwork for future likelihood-free pipelines.

Establishing a Framework for Physically-Informed Cluster Cosmology: This thesis establishes a unified framework for using constrained simulations and machine learning to disentangle the physics and cosmology encoded in galaxy clusters. Constrained simulations offer a reliable way to reproduce individual systems and track their histories, allowing us to perform detailed one-to-one studies of the galaxy clusters in our local neighborhood. At the same time, deep learning methods enable the extraction of cosmological and astrophysical information from galaxy clusters and efficient and thus accurate parameter inference across high-dimensional cosmological/galaxy-formation models. By bridging cosmological simulations, ML-based inference, and observational validation, this work paves the way for extracting maximal information from forthcoming surveys such as eROSITA, Simons Observatory, and Euclid, and contributes toward the next-generation cluster cosmology.

8 | Outlook

The work presented in this thesis has demonstrated the power of combining constrained cosmological simulations, thermodynamic analysis of galaxy clusters, and SBI to probe both astrophysical processes and cosmological parameters. Nevertheless, several key challenges remain unresolved, and overcoming them will be essential to fully exploit the potential of galaxy clusters as precision tools in modern cosmology. This section outlines the most pressing issues identified during this research and proposes future directions to address them.

Limitations in the Modeling of Baryonic Physics: Although the constrained simulations employed in this thesis successfully reproduce the global thermodynamic properties of many local galaxy clusters, systematic discrepancies persist, particularly in the core regions. The simulated entropy and electron density profiles often deviate from observations, especially for cool-core clusters. These mismatches point to persistent resolution limitations as well as limitations in current subgrid models of baryonic physics, including the implementation of AGN feedback, radiative cooling, and star formation.

Future work will focus on the development of a comprehensive suite of high-resolution zoom-in simulations targeting the same galaxy clusters modeled in the SLOW project. By simulating each cluster with high resolution and using different galaxy formation models (e.g., IllustrisTNG, Simba, and Magneticum), it will be possible to isolate the effects of various feedback prescriptions on cluster thermodynamics. By comparing with X-ray and SZ observations from instruments such as eROSITA, Chandra, and Planck, it will be possible to evaluate and improve the physical fidelity of current subgrid models.

Understanding the Connection Between Formation History and Observable Diversity: The link between cluster formation history and present-day observables, although addressed in this thesis, could be further explored. The classification of clusters into cool-core and non-cool-core types appears to correlate with assembly history, but a detailed causal mapping is still lacking.

Future work will take advantage of merger trees and formation histories derived from zoom-in simulations to quantify how progenitor dynamics, such as early accretion events or late major mergers, shape the detailed thermodynamic properties of clusters at redshift zero. Conducting this analysis across multiple feedback models will further clarify which aspects of cluster evolution are robust and which are model-dependent, ultimately informing a more predictive theory of cluster diversity.

Toward Realistic and Spatially-Informed Inference Frameworks: The SBI framework developed in this thesis has demonstrated that thermodynamic profiles of galaxy clusters contain rich information about both cosmological and astrophysical parameters, and that machine learning models can recover these parameters with high precision under idealized conditions. However, applying these techniques directly to observational data introduces new challenges, including instrumental noise, finite resolution, and projection effects, which are not captured in standard simulations outputs.

Moreover, while radial profiles outperform integrated observables as input quantities, they still compress and average out important spatial features of the ICM. In disturbed or non-spherical systems, such as those undergoing mergers, radial statistics can obscure asymmetries, cold fronts, and substructures that carry valuable physical information.

To address these limitations, I propose constructing mock X-ray and SZ observations from simulations, accurately reproducing survey-specific characteristics such as noise levels, resolution, and angular coverage. These mocks will serve as more appropriate training data for SBI, bridging the gap between theoretical modeling and observational data.

In parallel, extending the inference to two-dimensional data will be possible by using CNNs applied directly to simulated SZ and X-ray images. This would allow us to exploit spatial correlations and morphological features that are lost through radial averaging. Additionally, the use of more expressive statistical tools, such as normalizing flows, will enable better modeling of posterior distributions, capturing parameter degeneracies and non-Gaussian features that simpler models cannot.

Together, these developments will lead to more physically grounded and observationally relevant inference, advancing the applicability of SBI methods to current and future galaxy cluster surveys.

A New Approach - Constrained Zoom-In Simulations for Cosmological Inference: A particularly promising direction for future research lies in the combination of the two main strategies used throughout this thesis: constrained initial conditions and cosmological parameter inference. To date, SBI has largely relied on random initial conditions, which neglect the specific large-scale environment of observed clusters. I propose developing a novel suite of zoom-in simulations of local galaxy clusters using constrained initial conditions, run across a range of cosmological parameters (e.g., Ω_m, H_0, σ_8). This dataset can serve as the foundation for a SBI framework that performs cosmological parameter inference directly from real observed systems in our local neighborhood. Such an approach would represent a step-change in our ability to constrain cosmology from individual galaxy clusters in the local Universe.

In conclusion, this thesis has laid the groundwork for a new generation of cluster cosmology: one in which high-resolution, physically motivated simulations are combined with flexible, data-driven inference techniques to yield meaningful constraints on both astrophysical processes and fundamental cosmological parameters. By extending these methods to more realistic conditions, incorporating spatial information, and making use of the structure of the local Universe itself, future work can decisively advance our understanding of structure formation in the cosmos.

A | Appendix

A.1 Emulated Profiles With One Parameter Variation

To better understand the impact of individual parameters of the IllustrisTNG model on the averaged galaxy cluster profiles, we performed a series of emulations where we varied one parameter at a time while keeping the remaining 27 parameters fixed at their fiducial values. This method enabled a thorough exploration of each parameter's influence on the gas density, temperature, metallicity, Compton-y, and X-ray surface brightness mean profiles, as shown in Figs. A.1.1-A.1.5. By isolating specific parameters, we gained deeper insights into their roles and identified potential degeneracies across different profile types.

We observe that cosmological parameters such as $\Omega_{\rm m}$ and H_0 result in strong spreads for all quantities, followed closely by Ω_b. Parameters such as σ_8 and n_s show a more subtle effect on the profile variation for all quantities.

Astrophysical parameters, while having a milder influence on the profiles for most quantities, show a stronger effect on metallicity, which is particularly sensitive to stellar parameters like the IMF and feedback and wind parameters such as ASN1, ASN2, and W3.

The distinct responses of these profiles to the underlying parameters suggest that their combined effects are key to optimal parameter inference, helping to explain the model's exceptional performance.

A closer examination of Figs. A.1.1-A.1.5 reveals that each parameter introduces distinct, unique features in the profiles across all quantities. This observation is key to understand the high accuracy of our inference. These non-degenerate features enable the neural network to disentangle the individual effects of each parameter, allowing for parameter inference with an accuracy of 0.97 or higher in the noiseless case. This suggests that stacked galaxy cluster profiles contain clear and unambiguous information about both the astrophysical processes occurring in groups and clusters and the cosmological details of the Universe in which these structures reside.

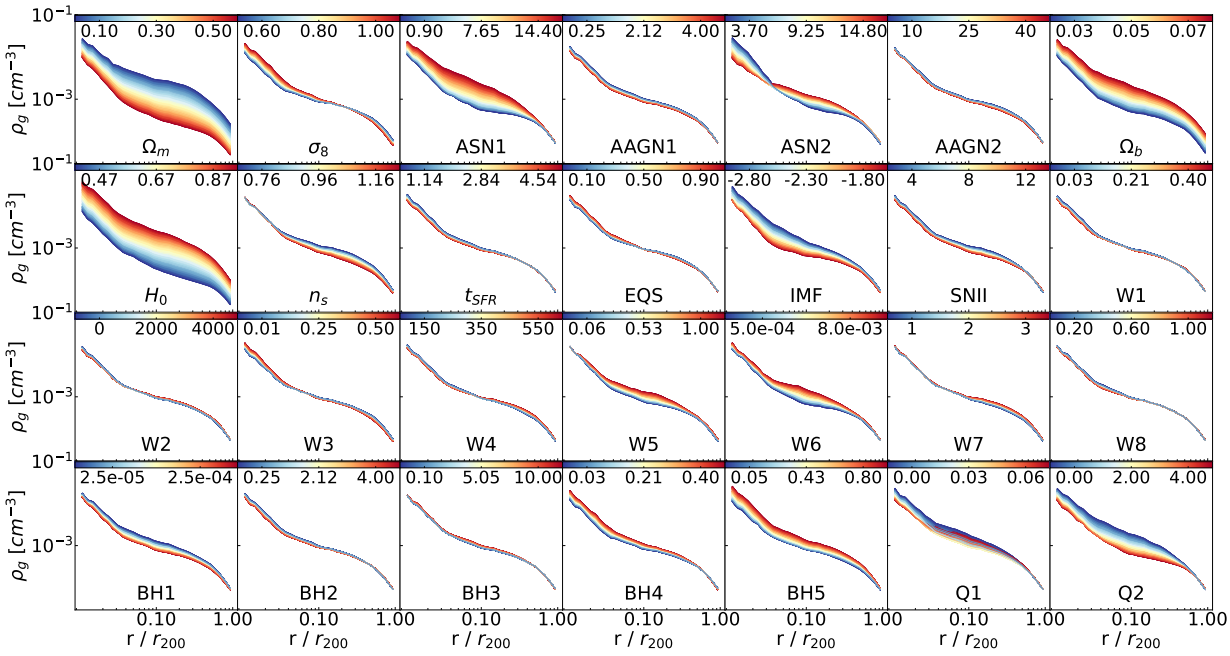


Figure A.1.1: Gas density profiles showing the impact of varying individual parameters for all parameters in the IllustrisTNG model. Each line represents the profile obtained when a single parameter is varied while keeping the others fixed at their fiducial values. The color gradient within each panel highlights the spread introduced into the profiles by the variations in the corresponding cosmological parameter, with the color scale representing the range of parameter values. This visualization demonstrates how changes in cosmological parameters affect different physical quantities in galaxy clusters, providing insight into the sensitivity of cluster profiles to cosmological variations.

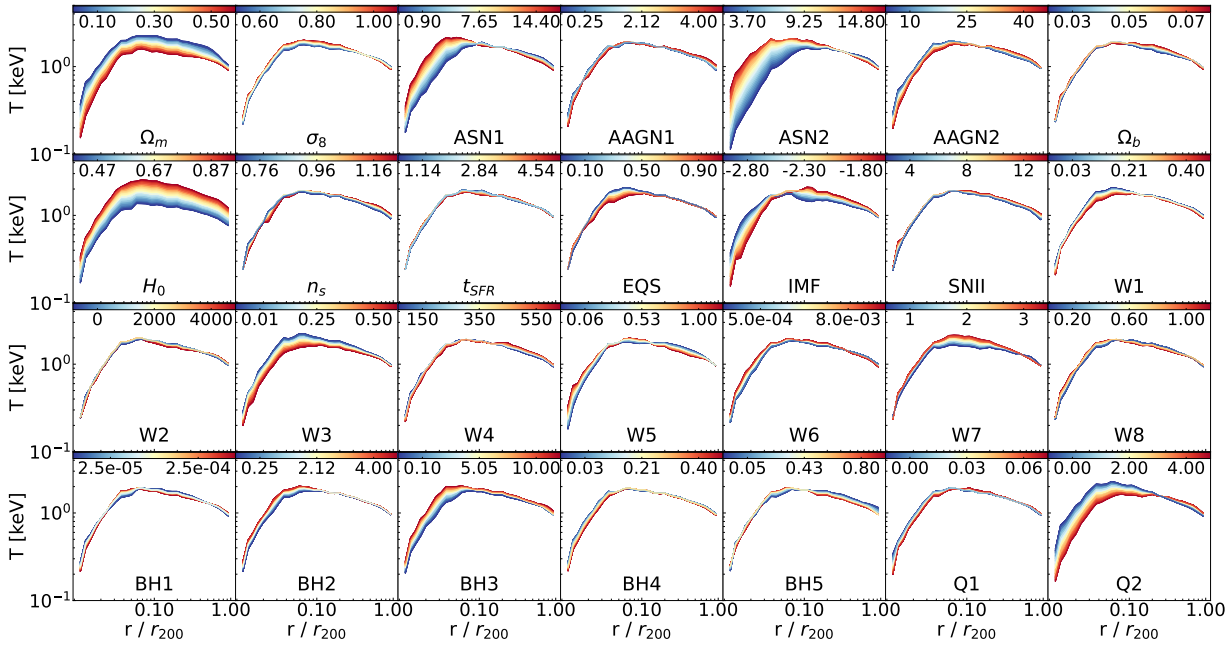


Figure A.1.2: Similar to Figure A.1.1, but for temperature profiles. Each line shows the effect of varying a single parameter while the others remain at their fiducial values, with the lines colored according to the value of the varying parameter.

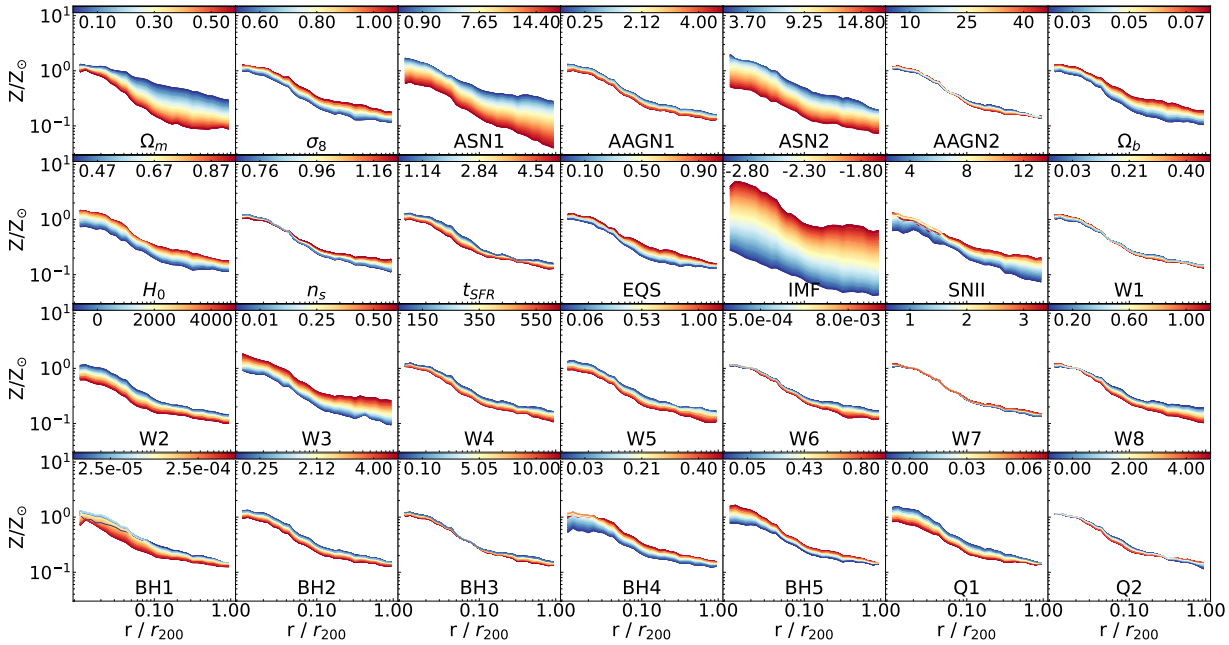


Figure A.1.3: Similar to Figure A.1.1, but for metallicity profiles. Each line shows the effect of varying a single parameter while the others remain at their fiducial values, with the lines colored according to the value of the varying parameter.

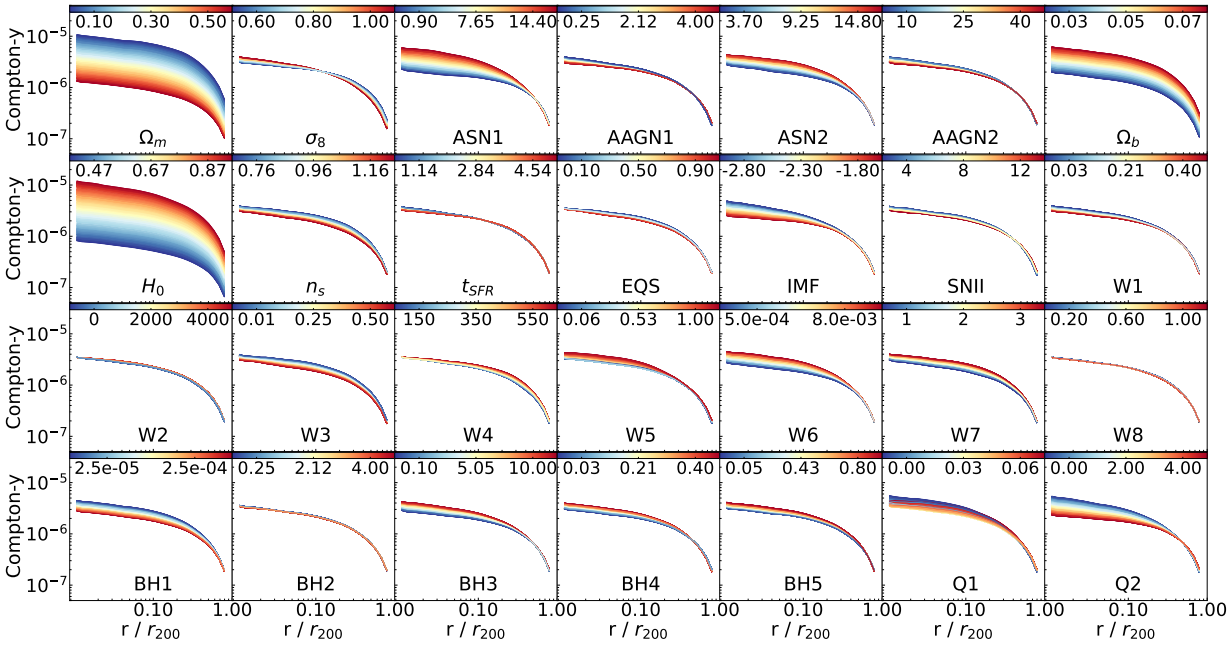


Figure A.1.4: Similar to Figure A.1.1, but for Compton-y profiles. Each line shows the effect of varying a single parameter while the others remain at their fiducial values, with the lines colored according to the value of the varying parameter.

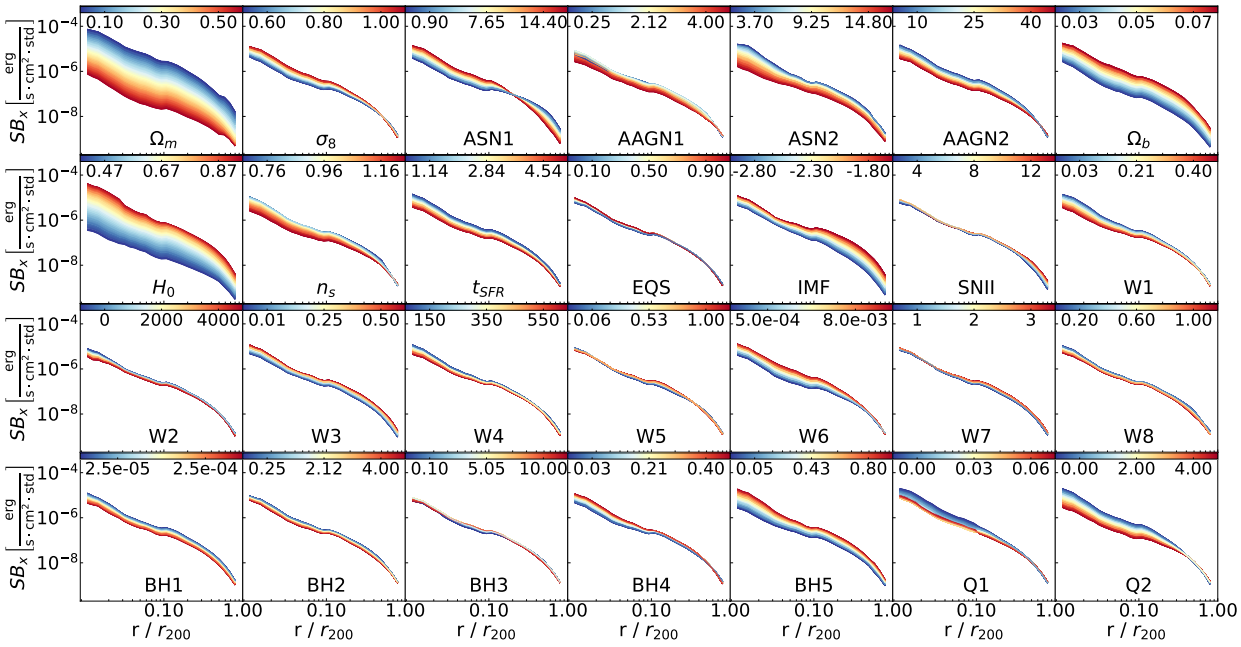


Figure A.1.5: Similar to Figure A.1.1, but for X-ray surface brightness profiles. Each line shows the effect of varying a single parameter while the others remain at their fiducial values, with the lines colored according to the value of the varying parameter.

Bibliography

Abadi, M. G., Moore, B., and Bower, R. G. (1999), *Ram pressure stripping of spiral galaxies in clusters*, Monthly Notices of the RAS, 308(4), 947

Abbott, T. M. C. et al. (2022), *Dark Energy Survey Year 3 Results: Cosmological Constraints from Galaxy Clustering and Weak Lensing*, Phys. Rev. D, 105, 023520

Abdalla, E., Abellán, G. F., Aboubrahim, A., et al. (2022), *Cosmology intertwined: A review of the particle physics, astrophysics, and cosmology associated with the cosmological tensions and anomalies*, Journal of High Energy Astrophysics, 34, 49

Agarwal, S., Villaescusa-Navarro, F., and et al. (2021), *ASTRID: Cosmology with baryons and massive neutrinos*, ApJS, 254(2), 22

Agertz, O., Moore, B., Stadel, J., et al. (2007), *Fundamental Differences between SPH and Grid Methods*, Monthly Notices of the Royal Astronomical Society, 380, 963

Aharonian, F. A., Akamatsu, H., Akimoto, F., et al. (2017), *Hitomi Constraints on the 3.5 keV Line in the Perseus Galaxy Cluster*, Astrophysical Journal, Letters, 837(1), L15

Akiba, T., Sano, S., Yanase, T., et al. (2019), *Optuna: A Next-generation Hyperparameter Optimization Framework*, in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623–2631, ACM

Alam, S. et al. (2017), *The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Cosmological analysis of the DR12 galaxy sample*, Mon. Not. R. Astron. Soc., 470, 2617

Alam, S. et al. (2021), *Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological implications from two decades of spectroscopic surveys at the Apache Point Observatory*, Phys. Rev. D, 103(8), 083533

Allen, S. W., Evrard, A. E., and Mantz, A. B. (2011), *Cosmological Parameters from Observations of Galaxy Clusters*, Annual Review of Astron and Astrophys, 49(1), 409

Allen, S. W., Fabian, A. C., Johnstone, R. M., et al. (1993), *A ROSAT PSPC observation of Abell 478 : the distribution of X-ray absorbing matter in a massive cooling flow.*, Monthly Notices of the RAS, 262, 901

Allen, S. W., Rapetti, D. A., Schmidt, R. W., et al. (2008), *Improved constraints on dark energy from Chandra X-ray observations of the largest relaxed galaxy clusters*, Monthly Notices of the Royal Astronomical Society, 383(3), 879

Allen, S. W., Schmidt, R. W., Ebeling, H., et al. (2004), *Constraints on dark energy from Chandra observations of the largest relaxed galaxy clusters*, Monthly Notices of the RAS, 353(2), 457

Allen, S. W., Schmidt, R. W., and Fabian, A. C. (2002), *Cosmological constraints from the X-ray gas mass fraction in relaxed lensing clusters observed with Chandra*, Monthly Notices of the RAS, 334(2), L11

Alsing, J., Wandelt, B., and Feeney, S. (2019), *Fast likelihood-free cosmology with neural density estimators and active learning*, Monthly Notices of the Royal Astronomical Society, 488(3), 4440

Alzubaidi, L., Zhang, J., Humaidi, A. J., et al. (2021), *Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions*, Journal of Big Data, 8(1), 1

Amon, A. et al. (2022), *Dark Energy Survey Year 3 results: Cosmology from cosmic shear and galaxy–galaxy lensing*, Phys. Rev. D, 105, 023514

Andrade-Santos, F., Nulsen, P. E. J., Kraft, R. P., et al. (2013), *DARK MATTER SUBHALOS AND THE X-RAY MORPHOLOGY OF THE COMA CLUSTER*, The Astrophysical Journal, 766(2), 107

Arbey, A. and Mahmoudi, F. (2021), *Dark matter and the early Universe: A review*, Progress in Particle and Nuclear Physics, 119, 103865

Arnaud, M., Aghanim, N., Gastaud, R., et al. (2001), *XMM-Newton observation of the Coma Galaxy cluster. The temperature structure in the central region*, Astronomy and Astrophysics, 365, L67

Arnaud, M., Pratt, G. W., Piffaretti, R., et al. (2010), *The universal galaxy cluster pressure profile from a representative sample of nearby systems (REXCESS) and the Y_{SZ} - M_{500} relation*, Astronomy and Astrophysics, 517, A92

Bagla, J. S. (2005), *Cosmological N-Body simulation: Techniques, Scope and Status*, Current Science, 88, 1088

Bahar, Y. E., Bulbul, E., Ghirardini, V., et al. (2024), *The SRG/eROSITA All-Sky Survey: Constraints on AGN Feedback in Galaxy Groups*

Barnes, D. J., Kay, S. T., Bahé, Y. M., et al. (2017), *The Cluster-EAGLE project: global properties of simulated clusters with resolved galaxies*, Monthly Notices of the RAS, 471(1), 1088

Barnes, D. J., Vogelsberger, M., Kannan, R., et al. (2018), *A census of cool-core galaxy clusters in IllustrisTNG*, Monthly Notices of the RAS, 481(2), 1809

Barnes, J. and Hut, P. (1986), *A Hierarchical O(N Log N) Force-Calculation Algorithm*, Nature, 324, 446

Battaglia, N., Bond, J. R., Pfrommer, C., and Sievers, J. L. (2012), *On the Cluster Physics of Sunyaev-Zel'dovich and X-Ray Surveys. II. Deconstructing the Thermal SZ Power Spectrum*, Astrophysical Journal, 758(2), 75

Battaglia, N., Ferraro, S., Schaan, E., and Spergel, D. N. (2017), *Future constraints on halo thermodynamics from combined Sunyaev-Zel'dovich measurements*, Journal of Cosmology and Astroparticle Physics, 2017(11), 040

Bauer, F. E., Fabian, A., Sanders, J., et al. (2005), *The prevalence of cooling cores in clusters of galaxies.*, Monthly Notices of the Royal Astronomical Society, 359(4), 1481

Baumann, D. (2022), *Cosmology*, Cambridge University Press

Bautista, J. E. et al. (2021), *The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the BAO and growth rate of structure of the luminous red galaxy sample from the final data release*, Mon. Not. R. Astron. Soc., 500, 736

Baxter, E., Chang, C., Jain, B., et al. (2017), *The Halo Boundary of Galaxy Clusters in the SDSS*, Astrophysical Journal, 841(1), 18

Beck, A. M., Murante, G., Arth, A., et al. (2016), *An Improved SPH Scheme for Cosmological Simulations*, Monthly Notices of the Royal Astronomical Society, 455, 2110

Benz, W. (1990), *Smooth particle hydrodynamics: a review*, The numerical modelling of nonlinear stellar pulsations: Problems and prospects, 269–288

Berger, M. J. and Colella, P. (1989), *Local Adaptive Mesh Refinement for Shock Hydrodynamics*, Journal of Computational Physics, 82, 64

Berger, M. J. and Oliger, J. (1984), *Adaptive Mesh Refinement for Hyperbolic Partial Differential Equations*, Journal of Computational Physics, 53, 484

Bergstra, J., Bardenet, R., Kégl, B., and Bengio, Y. (2011), in *Algorithms for Hyper-Parameter Optimization*

Bertone, G. and Tait, T. M. P. (2018), *A new era in the search for dark matter*, Nature, 562(7725), 51

Biffi, V., Borgani, S., Murante, G., et al. (2016), *On the Nature of Hydrostatic Equilibrium in Galaxy Clusters*, Astrophysical Journal, 827(2), 112

Biffi, V., Dolag, K., and Böhringer, H. (2013), *Investigating the velocity structure and X-ray observable properties of simulated galaxy clusters with PHOX*, Monthly Notices of the RAS, 428(2), 1395

Biffi, V., Mernier, F., and Medvedev, P. (2018), *Enrichment of the Hot Intracluster Medium: Numerical Simulations*, Space Science Reviews, 214(8), 123

Biffi, V., Planelles, S., Borgani, S., et al. (2017), *The History of Chemical Enrichment in the Intracluster Medium from Cosmological Simulations*, Monthly Notices of the Royal Astronomical Society, 468, 531

Binggeli, B. and Huchra, J. (2000), *Virgo Cluster*, in *Encyclopedia of Astronomy and Astrophysics*, edited by P. Murdin, 1822

Binney, J. and Tremaine, S. (1987), *Galactic dynamics*

Bird, S. et al. (2022), *Astrid: the astrophysical simulation for cosmological structure formation and reionization*, Monthly Notices of the Royal Astronomical Society, 509, 4024

Birkinshaw, M. (1999), *The Sunyaev-Zel'dovich effect*, Physics Reports, 310(2-3), 97

Bishop, C. M. (2006), *Pattern Recognition and Machine Learning*, Springer

Bistolas, V. and Hoffman, Y. (1998), *Nonlinear Constrained Realizations of the Large-Scale Structure*, Astrophysical Journal, 492(2), 439

Biviano, A. (1998), *Our best friend, the Coma cluster (a historical review)*, in *Untangling Coma Berenices: A New Vision of an Old Cluster*, edited by A. Mazure, F. Casoli, F. Durret, D. Gerbal, 1

Biviano, A., Durret, F., Gerbal, D., et al. (1996), *Unveiling hidden structures in the Coma cluster.*, Astronomy and Astrophysics, 311, 95

Bleem, L. E. et al. (2015), *Galaxy clusters discovered via the Sunyaev-Zel'dovich effect in the 2500-square-degree SPT-SZ survey*, Astrophys. J. Suppl. Ser., 216, 27

Bocquet, S. et al. (2019), *Cluster Cosmology Constraints from the 2500 square-degree SPT-SZ survey: inclusion of weak gravitational lensing data from Magellan and the Hubble Space Telescope*, Astrophys. J., 878, 55

Boggess, N. W., Mather, J. C., Weiss, R., et al. (1992), *The COBE Mission: Its Design and Performance Two Years after Launch*, The Astrophysical Journal, 397, 420

Böhringer, H., Chon, G., and Kronberg, P. P. (2016), *The Cosmic Large-Scale Structure in X-rays (CLASSIX) Cluster Survey. I. Probing galaxy cluster magnetic fields with line of sight rotation measures*, Astronomy and Astrophysics, 596, A22

Böhringer, H., Voges, W., Fabian, A. C., et al. (1993), *A ROSAT HRI study of the interaction of the X-ray emitting gas and radio lobes of NGC 1275.*, Monthly Notices of the RAS, 264, L25

Böhringer, H. and Werner, N. (2010), *X-Ray Spectroscopy of Galaxy Clusters: Studying Astrophysical Processes in the Largest Celestial Laboratories*, Astronomy and Astrophysics Review, 18, 127

Bonafede, A., Govoni, F., Feretti, L., et al. (2011), *Magnetic Field in Galaxy Clusters from Depolarization of Radio Sources.*, Memorie della Societa Astronomica Italiana, 82, 654

Bondi, H. (1952), *On spherically symmetrical accretion*, Monthly Notices of the Royal Astronomical Society, 112(2), 195

Bondi, H. and Hoyle, F. (1944), *On the mechanism of accretion by stars*, Monthly Notices of the Royal Astronomical Society, 104(5), 273

Bondi, H. and Hoyle, F. (1944), *On the mechanism of accretion by stars*, Monthly Notices of the RAS, 104, 273

Borgani, S. and Kravtsov, A. (2011), *Cosmological simulations of galaxy clusters*, Advanced Science Letters, 4(2), 204

Boselli, A. and Gavazzi, G. (2006), *Environmental Effects on Late-Type Galaxies in Nearby Clusters*, Publications of the ASP, 118(842), 517

Boselli, A., Voyer, E., Boissier, S., et al. (2014), *The GALEX Ultraviolet Virgo Cluster Survey (GUViCS). IV. The role of the cluster environment on galaxy evolution*, Astronomy and Astrophysics, 570, A69

Bosma, A. (1981), *21-cm line studies of spiral galaxies. I. Observations of the galaxies NGC 5033, 3198, 5055, 2841, and 7331.*, Astronomical Journal, 86, 1791

Böss, L. M., Dolag, K., Steinwandel, U. P., et al. (2023), *Simulating the LOcal Web (SLOW) – III: Synchrotron Emission from the Local Cosmic Web*, arXiv e-prints, arXiv:2310.13734

Bouchet, F. R., Adam, J. C., and Pellat, R. (1985), *On the way of understanding particle-in-cell simulations of gravitational clustering*, Astronomy and Astrophysics, 144(2), 413

Bowyer, A. (1981), *Computing Dirichlet Tessellations**, The Computer Journal, 24(2), 162

Boylan-Kolchin, M., Springel, V., White, S. D. M., et al. (2009), *Resolving cosmic structure formation with the Millennium-II Simulation*, Monthly Notices of the RAS, 398(3), 1150

Branduardi-Raymont, G., Fabricant, D., Feigelson, E., et al. (1981), *Soft X-ray images of the central region of the Perseus cluster.*, Astrophysical Journal, 248, 55

Braun, R., Bonaldi, A., Bourke, T., et al. (2019), *Anticipated Performance of the Square Kilometre Array – Phase 1 (SKA1)*, arXiv e-prints, arXiv:1912.12699

Briel, U. G., Henry, J. P., and Boehringer, H. (1992), *Observation of the Coma cluster of galaxies with ROSAT during the all-sky-survey.*, Astronomy and Astrophysics, 259, L31

Briel, U. G., Henry, J. P., Lumb, D. H., et al. (2001), *A mosaic of the Coma cluster of galaxies with XMM-Newton*, Astronomy and Astrophysics, 365, L60

Bryan, G. L. and Norman, M. L. (1997), *Simulating X-Ray Clusters with Adaptive Mesh Refinement*, in *Computational Astrophysics; 12th Kingston Meeting on Theoretical Astrophysics*, edited by D. A. Clarke, M. J. West, volume 12 of *Astronomical Society of the Pacific Conference Series*, 363

Bryan, G. L., Norman, M. L., O'Shea, B. W., et al. (2014), *ENZO: An Adaptive Mesh Refinement Code for Astrophysics*, The Astrophysical Journal Supplement Series, 211, 19

Buote, D. A. and Humphrey, P. J. (2012), *Spherically averaging ellipsoidal galaxy clusters in X-ray and Sunyaev-Zel'dovich studies - I. Analytical relations*, Monthly Notices of the RAS, 420(2), 1693

Buote, D. A. and Tsai, J. C. (1996), *Quantifying the Morphologies and Dynamical Evolution of Galaxy Clusters. II. Application to a Sample of ROSAT Clusters*, Astrophysical Journal, 458, 27

Burns, J. O., Hallman, E. J., Gantner, B., et al. (2008), *Why Do Only Some Galaxy Clusters Have Cool Cores?*, The Astrophysical Journal, 675(2), 1125

Březina, J., Cranmer, K., and Louppe, G. (2022), *Information maximizing neural networks for simulation-based inference*, arXiv preprint arXiv:2202.09348

Carilli, C. L. and Taylor, G. B. (2002), *Cluster Magnetic Fields*, Annual Review of Astronomy and Astrophysics, 40, 319

Carlesi, E., Hoffman, Y., Sorce, J. G., and Gottlöber, S. (2016a), *Constraining the mass of the Local Group*, Monthly Notices of the Royal Astronomical Society, 465(4), 4886

Carlesi, E., Sorce, J. G., Hoffman, Y., et al. (2016b), *Constrained local universe simulations: a local group factory*, Monthly Notices of the Royal Astronomical Society, 458(1), 900

Carlesi, E., Sorce, J. G., Hoffman, Y., et al. (2016), *Constrained Local UniversE Simulations: a Local Group factory*, Monthly Notices of the RAS, 458(1), 900

Carr, B. and Kuhnel, F. (2021), *Primordial Black Holes as Dark Matter Candidates*, arXiv e-prints, arXiv:2110.02821

Carroll, S. M. (2019), *Spacetime and geometry*, Cambridge University Press

Cassano, R., Ettori, S., Giacintucci, S., et al. (2010), *On the Connection Between Giant Radio Halos and Cluster Mergers*, Astrophysical Journal, Letters, 721(2), L82

Cataneo, M., Rapetti, D., Schmidt, F., et al. (2015), *New constraints on f(R) gravity from clusters of galaxies*, Physical Review D, 92(4), 044009

Cavagnolo, K. W., Donahue, M., Voit, G. M., and Sun, M. (2008), *Bandpass Dependence of X-Ray Temperatures in Galaxy Clusters*, The Astrophysical Journal, 682(2), 821

Cavagnolo, K. W., Donahue, M., Voit, G. M., and Sun, M. (2009), *Intracluster Medium Entropy Profiles for a Chandra Archival Sample of Galaxy Clusters*, The Astrophysical Journal Supplement Series, 182, 12

Cavagnolo, K. W., Donahue, M., Voit, G. M., and Sun, M. (2009), *Intracluster Medium Entropy Profiles for a Chandra Archival Sample of Galaxy Clusters*, Astrophysical Journal, Supplement, 182(1), 12

Chabrier, G. (2003), *Galactic Stellar and Substellar Initial Mass Function*, Publications of the ASP, 115(809), 763

Chartier, N., Wandelt, B., Akrami, Y., and Villaescusa-Navarro, F. (2021), *CARPool: fast, accurate computation of large-scale structure statistics by pairing costly and cheap cosmological simulations*, Monthly Notices of the RAS, 503(2), 1897

Chen, Y., Reiprich, T., Böhringer, H., et al. (2007), *Statistics of X-ray observables for the cooling-core and non-cooling core galaxy clusters*, Astronomy & Astrophysics, 466(3), 805

Chiu, I., Mohr, J. J., Bocquet, S., et al. (2018), *A universal gas mass fraction in galaxy clusters: evidence from the XMM-XXL and Planck data sets*, Monthly Notices of the Royal Astronomical Society, 478(3), 3072

Churazov, E., Forman, W., Jones, C., and Böhringer, H. (2003), *XMM-Newton Observations of the Perseus Cluster. I. The Temperature and Surface Brightness Structure*, Astrophysical Journal, 590(1), 225

Cimatti, A., Fraternali, F., and Nipoti, C. (2020), *Introduction to galaxy formation and evolution: from primordial gas to present-day galaxies*

Clarke, T. E., Blanton, E. L., and Sarazin, C. L. (2004), *Radio and X-Ray Interactions in the Core of Abell 2029*, in *The Riddle of Cooling Flows in Galaxies and Clusters of galaxies*, edited by T. Reiprich, J. Kempner, N. Soker, 19

Clifton, T., Ferreira, P. G., Padilla, A., and Skordis, C. (2012), *Modified gravity and cosmology*, Physics Reports, 513(1), 1

Clowe, D., Bradac, M., Gonzalez, A. H., et al. (2006), *A direct empirical proof of the existence of dark matter*, The Astrophysical Journal Letters, 648(2), L109

Collaboration, D. (2024), *The DESI Early Data Release: Science with the first year of data*, arXiv preprint

Collaboration, L. S. (2009), *LSST Science Book, Version 2.0*, arXiv preprint

Corless, V. L. and King, L. J. (2007), *A statistical study of weak lensing by triaxial dark matter haloes: consequences for parameter estimation*, Monthly Notices of the RAS, 380(1), 149

Couchman, H. M. P. (1991), *Mesh-Refined P 3M: A Fast Adaptive N-Body Algorithm*, The Astrophysical Journal, 368, L23

Courtois, H. M., Pomarède, D., Tully, R. B., et al. (2013), *Cosmography of the Local Universe*, Astronomical Journal, 146(3), 69

Cowie, L. L. and Binney, J. (1977), *Radiative regulation of gas flow within clusters of galaxies: a model for cluster X-ray sources.*, Astrophysical Journal, 215, 723

Cranmer, K., Brehmer, J., and Louppe, G. (2020), *The frontier of simulation-based inference*, Proceedings of the National Academy of Sciences, 117(48), 30055

Crawford, C. S., Sanders, J. S., and Fabian, A. C. (2005), *The giant Hα/X-ray filament in the cluster of galaxies A1795*, Monthly Notices of the RAS, 361(1), 17

Croton, D. J. (2013), *Damn You, Little h! (Or, Real-World Applications of the Hubble Constant Using Observed and Simulated Data)*, Publications of the Astronomical Society of Australia, 30, e052

Davé, R. et al. (2019), *SIMBA: Cosmological simulations with black hole growth and feedback*, Monthly Notices of the Royal Astronomical Society, 486, 2827

de Haan, T., Benson, B. A., Bleem, L. E., and et al. (2016), *Cosmological Constraints from Galaxy Clusters in the 2500 Square-degree SPT-SZ Survey*, The Astrophysical Journal, 832(1), 95

De Lucia, G. and Blaizot, J. (2007), *The hierarchical formation of the brightest cluster galaxies*, Monthly Notices of the RAS, 375(1), 2

de Santi, N. S. M., Shao, H., Villaescusa-Navarro, F., et al. (2023), *Robust Field-level Likelihood-free Inference with Galaxies*, Astrophysical Journal, 952(1), 69

de Santi, N. S. M., Villaescusa-Navarro, F., Raul Abramo, L., et al. (2025), *Field-level simulation-based inference with galaxy catalogs: the impact of systematic effects*, Journal of Cosmology and Astroparticle Physics, 2025(1), 082

de Vaucouleurs, G. (1960), *The Apparent Density of Matter in Groups and Clusters of Galaxies.*, Astrophysical Journal, 131, 585

Dehnen, W. and Aly, H. (2012), *Improving convergence in smoothed particle hydrodynamics simulations without pairing instability*, Monthly Notices of the Royal Astronomical Society, 425(2), 1068

Dehnen, W. and Read, J. I. (2011), *N-body simulations of gravitational dynamics*, European Physical Journal Plus, 126, 55

Dennis, T. J. and Chandran, B. D. (2005), *Turbulent heating of galaxy-cluster plasmas*, The Astrophysical Journal, 622(1), 205

DES Collaboration (2022), *Dark Energy Survey Year 3 Results: Cosmological Constraints from Galaxy Clustering and Weak Lensing*, Phys. Rev. D, 105, 023520

DES Collaboration, Abbott, T. M. C., Aguena, M., et al. (2022), *Dark Energy Survey Year 3 Results: Cosmological Constraints from Galaxy Clustering and Weak Lensing*, Physical Review D, 105(2), 023520

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

Di Matteo, T., Springel, V., and Hernquist, L. (2005), *Energy input from quasars regulates the growth and activity of black holes and their host galaxies*, nature, 433(7026), 604

Di Valentino, E., Mena, O., Pan, S., and et al. (2021), *In the realm of the Hubble tension—a review of solutions*, Classical and Quantum Gravity, 38(15), 153001

Diemer, B. and Kravtsov, A. V. (2014), *Dependence of the Outer Density Profiles of Halos on Their Mass Accretion Rate*, Astrophysical Journal, 789(1), 1

Dodelson, S. and Schmidt, F. (2024), *Modern cosmology*, Elsevier

Dolag, K., Borgani, S., Murante, G., and Springel, V. (2009), *Substructures in Hydrodynamical Cluster Simulations*, Monthly Notices of the Royal Astronomical Society, 399(2), 497

Dolag, K., Borgani, S., Schindler, S., et al. (2008), *Simulation techniques for cosmological simulations*, Space science reviews, 134, 229

Dolag, K., Grasso, D., Springel, V., and Tkachev, I. (2004), *Mapping Deflections of Ultrahigh Energy Cosmic Rays in Constrained Simulations of Extraglactic Magnetic Fields*, Soviet Journal of Experimental and Theoretical Physics Letters, 79, 583

Dolag, K., Grasso, D., Springel, V., and Tkachev, I. (2005), *Constrained simulations of the magnetic field in the local Universe and the propagation of ultrahigh energy cosmic rays*, Journal of Cosmology and Astroparticle Physics, 2005(1), 009

Dolag, K., Hansen, F. K., Roncarelli, M., and Moscardini, L. (2005), *The Imprints of Local Superclusters on the Sunyaev-Zel'dovich Signals and Their Detectability with Planck*, Monthly Notices of the Royal Astronomical Society, 363(1), 29

Dolag, K., Hansen, F. K., Roncarelli, M., and Moscardini, L. (2005), *The imprints of local superclusters on the Sunyaev-Zel'dovich signals and their detectability with Planck*, Monthly Notices of the RAS, 363(1), 29

Dolag, K., Jubelgas, M., Springel, V., et al. (2004), *Thermal conduction in simulated galaxy clusters*, The Astrophysical Journal, 606(2), L97

Dolag, K., Komatsu, E., and Sunyaev, R. (2016), *SZ effects in the Magneticum Pathfinder Simulation: Comparison with the Planck, SPT, and ACT results*, Monthly Notices of the Royal Astronomical Society, 463(2), 1797

Dolag, K., Mevius, E., and Remus, R.-S. (2017), *Distribution and Evolution of Metals in the Magneticum Simulations*, Galaxies, 5, 35

Dolag, K., Remus, R.-S., Valenzuela, L. M., et al. (2025), *Encyclopedia Magneticum: Scaling Relations from Cosmic Dawn to Present Day*, arXiv e-prints, arXiv:2504.01061

Dolag, K., Sorce, J. G., Pilipenko, S., et al. (2023), *Simulating the LOcal Web (SLOW): I. Anomalies in the Local Density Field*, Astronomy & Astrophysics, 677, A169

Dolag, K., Sorce, J. G., Pilipenko, S., et al. (2023), *Simulating the LOcal Web (SLOW). I. Anomalies in the local density field*, Astronomy and Astrophysics, 677, A169

Dolag, K., Sorce, J. G., Pilipenko, S., et al. (2023), *Simulating the LOcal Web (SLOW)*, Astronomy & Astrophysics, 677, A169

Dolag, K. and Sunyaev, R. (2013), *Relative velocity of dark matter and baryons in clusters of galaxies and measurements of their peculiar velocities*, Monthly Notices of the Royal Astronomical Society, 432(2), 1600

Domingos, P. (2012), *A few useful things to know about machine learning*, Communications of the ACM, 55(10), 78

Domínguez Sánchez, H. and et al. (2018), *Improving galaxy morphologies for SDSS with Deep Learning*, MNRAS, 476(3), 3661

Doumler, T., Gottlöber, S., Hoffman, Y., and Courtois, H. (2013), *Reconstructing cosmological initial conditions from galaxy peculiar velocities – III. Constrained simulations*, Monthly Notices of the Royal Astronomical Society, 430(2), 912

Douspis, M., Salvati, L., and Aghanim, N. (2019), *On the tension between Large Scale Structures and Cosmic Microwave Background*, arXiv e-prints, arXiv:1901.05289

Dressler, A. (1980), *Galaxy morphology in rich clusters: implications for the formation and evolution of galaxies.*, Astrophysical Journal, 236, 351

Dupuy, A. and Courtois, H. M. (2023), *Dynamic cosmography of the local Universe: Laniakea and five more watershed superclusters*, A&A, 678, A176

Durret, F., Lima Neto, G. B., and Forman, W. (2005), *An XMM-Newton view of the cluster of galaxies Abell 85*, Astronomy and Astrophysics, 432(3), 809

Eastwood, J. W. and Hockney, R. W. (1974), *Shaping the Force Law in Two-Dimensional Particle-Mesh Models*, Journal of Computational Physics, 16(4), 342

Echeverri-Rojas, N., Villaescusa-Navarro, F., Chawak, C., et al. (2023), *Cosmology with One Galaxy? The ASTRID Model and Robustness*, Astrophysical Journal, 954(2), 125

Eckert, D., Ettori, S., Molendi, S., et al. (2013), *The X-ray/SZ view of the virial region. II. Gas mass fraction*, Astronomy and Astrophysics, 551, A23

Eckert, D., Ettori, S., Pointecouteau, E., et al. (2017), *The XMM cluster outskirts project (X- COP)*, Astronomische Nachrichten, 338(293), 293

Eckert, D., Gaspari, M., Gastaldello, F., et al. (2021), *Feedback from Active Galactic Nuclei in Galaxy Groups*, Universe, 7(5), 142

Eckert, D., Ghirardini, V., Ettori, S., et al. (2019), *Non-Thermal Pressure Support in X-COP Galaxy Clusters*, Astronomy and Astrophysics, 621, A40

Eckert, D., Jauzac, M., Shan, H., et al. (2015), *Warm-hot baryons comprise 5-10 per cent of filaments in the cosmic web*, Nature, 528(7580), 105

Eddington, A. S. (1913), *On a formula for correcting statistics for the effects of a known error of observation*, Monthly Notices of the RAS, 73, 359

Edge, A. (2001), *The detection of molecular gas in the central galaxies of cooling flow clusters*, Monthly Notices of the Royal Astronomical Society, 328(3), 762

Efstathiou, G., Davis, M., White, S. D. M., and Frenk, C. S. (1985), *Numerical techniques for large cosmological N-body simulations*, Astrophysical Journal, Supplement, 57, 241

Eke, V. R., Cole, S., and Frenk, C. S. (1996), *Cluster evolution as a diagnostic for Omega*, Monthly Notices of the RAS, 282, 263

Elsner, F., Gilman, D., Birrer, S., and et al. (2019), *Joint likelihood-free inference of dynamical and cosmological parameters from strong lensing observations*, Journal of Cosmology and Astroparticle Physics, 2019(10), 010

Ettori, S., De Grandi, S., and Molendi, S. (2002), *Gravitating mass profiles of nearby galaxy clusters and relations with X-ray gas temperature, luminosity and mass*, Astronomy and Astrophysics, 391, 841

Ettori, S., Donnarumma, A., Pointecouteau, E., et al. (2013), *Mass Profiles of Galaxy Clusters from X-ray Analysis*, Space Science Reviews, 177(1-4), 119

Ettori, S., Gastaldello, F., Leccardi, A., et al. (2010), *Mass profiles and c-M_{DM} relation in X-ray luminous galaxy clusters*, Astronomy and Astrophysics, 524, A68

Ettori, S., Ghirardini, V., Eckert, D., et al. (2019), *Hydrostatic mass profiles in X-COP galaxy clusters*, Astronomy and Astrophysics, 621, A39

Ettori, S., Morandi, A., Tozzi, P., et al. (2009), *The cluster gas mass fraction as a cosmological probe: a revised study*, Astronomy I& Astrophysics, 501(1), 61

Euclid Collaboration, Scaramella, R., Amiaux, J., et al. (2022), *Euclid preparation. I. The Euclid Wide Survey*, Astronomy and Astrophysics, 662, A112

Fabian, A. C. (2002), *Cooling Flows in Clusters of Galaxies*, in *Lighthouses of the Universe: The Most Luminous Celestial Objects and Their Use for Cosmology: Proceedings of the MPA/ESO/MPE/USM Joint Astronomy Conference Held in Garching, Germany, 6-10 August 2001*, 24–36, Springer

Fabian, A. C., Iwasawa, K., Reynolds, C. S., and Young, A. J. (2000a), *Broad Iron Lines in Active Galactic Nuclei*, Publications of the ASP, 112(775), 1145

Fabian, A. C., Mushotzky, R. F., Nulsen, P. E. J., and Peterson, J. R. (2001), *On the soft X-ray spectrum of cooling flows*, Monthly Notices of the RAS, 321(1), L20

Fabian, A. C. and Nulsen, P. E. J. (1977), *Subsonic accretion of cooling gas in clusters of galaxies*, Monthly Notices of the Royal Astronomical Society, 180(3), 479

Fabian, A. C., Sanders, J. S., Allen, S. W., et al. (2011), *A wide Chandra view of the core of the Perseus cluster*, Monthly Notices of the Royal Astronomical Society, 418(4), 2154

Fabian, A. C., Sanders, J. S., Ettori, S., et al. (2000b), *Chandra imaging of the complex X-ray core of the Perseus cluster*, Monthly Notices of the RAS, 318(4), L65

Fabian, A. C., Sanders, J. S., Taylor, G. B., et al. (2006), *A very deep Chandra observation of the Perseus cluster: shocks, ripples and conduction*, Monthly Notices of the RAS, 366(2), 417

Fabjan, D., Borgani, S., Tornatore, L., et al. (2010), *Simulating the effect of active galactic nuclei feedback on the metal enrichment of galaxy clusters*, Monthly Notices of the Royal Astronomical Society, 401(3), 1670

Fedeli, C. (2012), *The Effects of Baryonic Cooling on the Concentration-Mass Relation*, Monthly Notices of the Royal Astronomical Society, 424, 1244

Feng, J. L. (2010), *Dark Matter Candidates from Particle Physics and Methods of Detection*, Annual Review of Astron and Astrophys, 48, 495

Feretti, L., Giovannini, G., and Böhringer, H. (1997), *The radio and X-ray properties of Abell 2319*, New Astronomy, 2(6), 501

Ferrarese, L., Côté, P., Cuillandre, J.-C., et al. (2012), *The Next Generation Virgo Cluster Survey (NGVS). I. Introduction to the Survey*, Astrophysical Journal, Supplement, 200(1), 4

Ferraro, S., Schmidt, F., and Hu, W. (2011), *Cluster abundance in f(R) gravity models*, Physical Review D, 83(6), 063503

Fisher, K. (1995), Monthly Notices of the Royal Astronomical Society

Fitchett, M. and Webster, R. (1987), *Substructure in the Coma Cluster*, Astrophysical Journal, 317, 653

Fixsen, D. J., Cheng, E. S., Gales, J. M., et al. (1996), *The Cosmic Microwave Background Spectrum from the Full COBE FIRAS Data Set*, Astrophysical Journal, 473, 576

Fraser-McKelvie, A., Brown, M. J., and Pimbblet, K. A. (2014), *The rarity of star formation in brightest cluster galaxies as measured by WISE*, Monthly Notices of the Royal Astronomical Society: Letters, 444(1), L63

Frenk, C. S., White, S. D. M., Bode, P., et al. (1999), *The Santa Barbara Cluster Comparison Project: A Comparison of Cosmological Hydrodynamics Solutions*, Astrophysical Journal, 525(2), 554

Fujita, Y., Kawakatu, N., and Shlosman, I. (2014), *AGN jet power and feedback characterised by Bondi accretion in brightest cluster galaxies*, arXiv preprint arXiv:1406.6366

Gaspari, M., Brighenti, F., and Ruszkowski, M. (2013), *Solving the cooling flow problem through mechanical AGN feedback*, Astronomische Nachrichten, 334(4-5), 394

Gaspari, M., Tombesi, F., and Cappi, M. (2020), *Linking macro-, meso- and microscales in multiphase AGN feeding and feedback*, Nature Astronomy, 4, 10

Geller, M. J. and Huchra, J. P. (1983), *Groups of galaxies. III. THe CfA survey.*, Astrophysical Journal, Supplement, 52, 61

Genel, S., Vogelsberger, M., Springel, V., et al. (2014), *Introducing the Illustris project: the evolution of galaxy populations across cosmic time*, Monthly Notices of the RAS, 445(1), 175

Ghirardini, V., Eckert, D., Ettori, S., et al. (2019), *Universal Thermodynamic Properties of the Intracluster Medium over Two Decades in Radius in the X-COP Sample*, Astronomy and Astrophysics, 621, A41

Gingold, R. A. and Monaghan, J. J. (1977), *Smoothed particle hydrodynamics: theory and application to non-spherical stars.*, Monthly Notices of the RAS, 181, 375

Giodini, S., Pierini, D., Finoguenov, A., et al. (2009), *Stellar and total baryon mass fractions in groups and clusters since redshift 1*, The Astrophysical Journal, 703(2), 982

González Villalba, J. A., Dolag, K., and Biffi, V. (2025), *How the cool-core population transitions from galaxy groups to massive clusters: A comparison of the largest Magneticum simulation with eROSITA, XMM-Newton, Chandra and LOFAR observations*, Astronomy and Astrophysics, 694, A232

Goodfellow, I., Bengio, Y., and Courville, A. (2016), *Deep Learning*, MIT Press

Gottlöber, S., Hoffman, Y., and Yepes, G. (2010), *Constrained local universe simulations (CLUES)*, in *High Performance Computing in Science and Engineering, Garching/Munich 2009: Transactions of the Fourth Joint HLRB and KONWIHR Review and Results Workshop, Dec. 8-9, 2009, Leibniz Supercomputing Centre, Garching/Munich, Germany*, 309–322, Springer

Govoni, F. and Feretti, L. (2004), *Magnetic Fields in Clusters of Galaxies*, International Journal of Modern Physics D, 13(8), 1549

Groth, F., Steinwandel, U. P., Valentini, M., and Dolag, K. (2023), *The Cosmological Simulation Code OpenGadget3 – Implementation of Meshless Finite Mass*, Monthly Notices of the Royal Astronomical Society, 526(1), 616

Gunn, J. E. and Gott, III, J. R. (1972), *On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution*, Astrophysical Journal, 176, 1

Gupta, N., Saro, A., Mohr, J. J., et al. (2017), *SZE observables, pressure profiles and centre offsets in Magneticum simulation galaxy clusters*, Monthly Notices of the RAS, 469(3), 3069

Guth, A. H. (1981), *Inflationary universe: A possible solution to the horizon and flatness problems*, Physical Review D, 23(2), 347

Guth, A. H. (1997), *The Inflationary Universe: The Quest for a New Theory of Cosmic Origins*, Addison-Wesley

Géron, A. (2019), *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, O'Reilly Media

Hahn, O., Martizzi, D., Wu, H.-Y., and Springel, V. (2017), *The Evolution of the Cool Core Phenomenon in Galaxy Clusters*, Monthly Notices of the Royal Astronomical Society, 469(3), 2936

Hamaus, N. et al. (2016), *Constraints on Cosmology and Gravity from the Dynamics of Voids*, Phys. Rev. Lett., 117, 091302

Hanany, S., Ade, P., Balbi, A., et al. (2000), *MAXIMA-1: A Measurement of the Cosmic Microwave Background Anisotropy on Angular Scales of 10'-5°*, Astrophysical Journal, Letters, 545(1), L5

Hand, N., Addison, G. E., Aubourg, E., et al. (2012), *Evidence of Galaxy Cluster Motions with the Kinematic Sunyaev-Zel'dovich Effect*, Physical Review Letters, 109(4), 041101

Hasselfield, M., Hilton, M., Marriage, T. A., et al. (2013), *The Atacama Cosmology Telescope: Sunyaev-Zel'dovich selected galaxy clusters at 148 GHz from three seasons of data*, Journal of Cosmology and Astroparticle Physics, 2013(7), 008

Hastie, T., Tibshirani, R., and Friedman, J. (2009), *The Elements of Statistical Learning*, Springer

Heitmann, K., Higdon, D., Nakhleh, C., and Habib, S. (2009), *The Coyote Universe. I. Precision determination of the nonlinear matter power spectrum*, ApJ, 705(1), 156

Heitmann, K., Lukić, Z., Fasel, P., et al. (2008), *The Cosmic Code Comparison Project*, Computational Science and Discovery, 1, 015003

Henriksen, M. J. and Tittley, E. R. (2002), *Chandra Observations of the A3266 Galaxy Cluster Merger*, Astrophysical Journal, 577(2), 701

Henriques, B. M. B., White, S. D. M., Thomas, P. A., et al. (2015), *Galaxy formation in the Planck cosmology - I. Matching the observed evolution of star formation rates, colours and stellar masses*, Monthly Notices of the RAS, 451(3), 2663

Hernández-Martínez, E., Dolag, K., Seidel, B., et al. (2024), *Simulating the LOcal Web (SLOW). II. Properties of Local Galaxy Clusters*, Astronomy & Astrophysics, 687, A253

Hernández-Martínez, E., Genel, S., Villaescusa-Navarro, F., et al. (2025), *Cosmological and Astrophysical Parameter Inference from Stacked Galaxy Cluster Profiles Using CAMELS-zoomGZ*, Astrophysical Journal, 981(2), 170

Hernquist, L. and Katz, N. (1989), *TREESPH: A Unification of SPH with the Hierarchical Tree Method*, Astrophysical Journal, Supplement, 70, 419

Heymans, C. and et al. (2021), *KiDS-1000 Cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints*, Astronomy & Astrophysics, 646, A140

Hildebrandt, H. et al. (2021), *KiDS-1000 Cosmology: Cosmic shear constraints and comparison between two point statistics*, Astron. Astrophys., 647, A124

Hilton, M. et al. (2021), *Cosmology constraints from thermal Sunyaev-Zel'dovich-selected galaxy clusters in the Advanced ACTPol survey region*, Astrophys. J. Suppl. Ser., 253, 3

Hilton, M., Sifón, C., Naess, S., et al. (2021), *The Atacama Cosmology Telescope: A Catalog of >4000 Sunyaev-Zel'dovich Galaxy Clusters*, Astrophysical Journal, Supplement, 253(1), 3

Hinshaw, G., Nolta, M. R., Bennett, C. L., et al. (2007), *Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Temperature Analysis*, Astrophysical Journal, Supplement, 170(2), 288

Hirschmann, M., Dolag, K., Saro, A., et al. (2014a), *Cosmological simulations of black hole growth: AGN luminosities and downsizing*, Monthly Notices of the Royal Astronomical Society, 442(3), 2304

Hirschmann, M., Dolag, K., Saro, A., et al. (2014b), *Cosmological simulations of black hole growth: AGN luminosities and downsizing*, Monthly Notices of the Royal Astronomical Society, 442(3), 2304

Hitomi Collaboration, Aharonian, F., Akamatsu, H., et al. (2016), *The Quiescent Intracluster Medium in the Core of the Perseus Cluster*, Nature, 535, 117

Hockney, R. W. and Eastwood, J. W. (1981), *Computer Simulation Using Particles*

Hoffer, A. S., Donahue, M., Hicks, A., and Barthelemy, R. (2012), *Infrared and ultraviolet star formation in brightest cluster galaxies in the ACCEPT sample*, The Astrophysical Journal Supplement Series, 199(1), 23

Hoffman, Y. (1993), *Reconstruction of the Large Scale Structure*

Hoffman, Y. and Ribak, E. (1991), *Constrained Realizations of Gaussian Fields: A Simple Algorithm*, Astrophysical Journal, Letters, 380, L5

Hoffman, Y. and Ribak, E. (1992), *Primordial Gaussian Perturbation Fields: Constrained Realizations*, Astrophysical Journal, 384, 448

Hornik, K., Stinchcombe, M., and White, H. (1989), *Multilayer feedforward networks are universal approximators*, Neural networks, 2(5), 359

Hoyle, F. and Lyttleton, R. A. (1939), *The effect of interstellar matter on climatic variation*, Proceedings of the Cambridge Philosophical Society, 35(3), 405

Hu, W., Barkana, R., and Gruzinov, A. (2000), *Fuzzy Cold Dark Matter: The Wave Properties of Ultralight Particles*, Physical Review Letters, 85(6), 1158

Hudson, D. S., Mittal, R., Reiprich, T. H., et al. (2010), *What is a cool-core cluster? A detailed analysis of the cores of the X-ray flux-limited HIFLUGCS cluster sample*, Astronomy I& Astrophysics, 513, A37

Ichinohe, Y., Werner, N., Simionescu, A., et al. (2015), *The growth of the galaxy cluster Abell 85: mergers, shocks, stripping and seeding of clumping*, Monthly Notices of the RAS, 448(3), 2971

Ikebe, Y., Reiprich, T. H., Böhringer, H., et al. (2002), *A new measurement of the X-ray temperature function of clusters of galaxies*, Astronomy & Astrophysics, 383(3), 773

Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. (2019), *LSST: From Science Drivers to Reference Design and Anticipated Data Products*, Astrophysical Journal, 873(2), 111

Jansen, F., Lumb, D., Altieri, B., et al. (2001), *XMM-Newton observatory. I. The spacecraft and operations*, Astronomy and Astrophysics, 365, L1

Jasche, J. and Lavaux, G. (2019), *Physical Bayesian modelling of the non-linear matter distribution: New insights into the nearby universe*, Astronomy and Astrophysics, 625, A64

Jasche, J. and Wandelt, B. D. (2013), *Bayesian physical reconstruction of initial conditions from large-scale structure surveys*, Monthly Notices of the RAS, 432(2), 894

Jeffrey, N. and Wandelt, B. D. (2020), *Solving high-dimensional parameter inference: marginal posterior densities & Moment Networks*, arXiv e-prints, arXiv:2011.05991

Kaastra, J., Ferrigno, C., Tamura, T., et al. (2001), *XMM-Newton observations of the cluster of galaxies Sérsic 159-03*, Astronomy & Astrophysics, 365(1), L99

Kaiser, N. (1986), *Evolution and clustering of rich clusters.*, Monthly Notices of the RAS, 222, 323

Karachentsev, I. D., Tully, R. B., Wu, P.-F., et al. (2014), *Infall of Nearby Galaxies into the Virgo Cluster as Traced with Hubble Space Telescope*, Astrophysical Journal, 782(1), 4

Karakas, A. I. (2010), *Updated stellar yields from asymptotic giant branch models*, Monthly Notices of the RAS, 403(3), 1413

Kashibadze, O. G., Karachentsev, I. D., and Karachentseva, V. E. (2020), *Structure and kinematics of the Virgo cluster of galaxies*, Astronomy and Astrophysics, 635, A135

Katz, N., Weinberg, D. H., and Hernquist, L. (1996), *Cosmological Simulations with TreeSPH*, Astrophysical Journal, Supplement, 105, 19

Kawaharada, M., Okabe, N., Umetsu, K., et al. (2010), *Suzaku Observation of A1689: Anisotropic Temperature and Entropy Distributions Associated with the Large-scale Structure*, Astrophysical Journal, 714(1), 423

Kempner, J. C., Sarazin, C. L., and Ricker, P. M. (2002), *Chandra Observations of A85: Merger of the South Subcluster*, Astrophysical Journal, 579(1), 236

Kennicutt, Jr., R. C. (1998), *The Global Schmidt Law in Star-forming Galaxies*, Astrophysical Journal, 498(2), 541

Kingma, D. P. and Ba, J. (2014), *Adam: A Method for Stochastic Optimization*, arXiv e-prints, arXiv:1412.6980

Klypin, A., Hoffman, Y., Kravtsov, A. V., and Gottlöber, S. (2003), *Constrained Simulations of the Real Universe: The Local Supercluster*, Astrophysical Journal, 596(1), 19

Klypin, A. and Holtzman, J. (1997), *Particle-Mesh code for cosmological simulations*, arXiv e-prints, astro–ph/9712217

Knebe, A., Pearce, F. R., Lux, H., et al. (2013), *Structure finding in cosmological simulations: the state of affairs*, Monthly Notices of the RAS, 435(2), 1618

Knox, L. and Millea, M. (2020), *Hubble constant hunter's guide*, Physics Today, 73(9), 28

Komatsu, E., Smith, K. M., Dunkley, J., et al. (2011), *Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation*, The Astrophysical Journal Supplement Series, 192, 18

Kravtsov, A. V. and Borgani, S. (2012a), *Formation of Galaxy Clusters*, Annual Review of Astronomy and Astrophysics, 50, 353

Kravtsov, A. V. and Borgani, S. (2012b), *Formation of Galaxy Clusters*, Ann. Rev. Astron. Astrophys., 50, 353

Kravtsov, A. V. and Borgani, S. (2012), *Formation of Galaxy Clusters*, Annual Review of Astron and Astrophys, 50, 353

Kravtsov, A. V., Klypin, A. A., and Khokhlov, A. M. (1997), *Adaptive Refinement Tree: A New High-Resolution N-Body Code for Cosmological Simulations*, Astrophysical Journal, Supplement, 111(1), 73

Kravtsov, A. V., Klypin, A. A., and Khokhlov, A. M. (1997), *Adaptive Refinement Tree: A New High-Resolution N-Body Code for Cosmological Simulations*, The Astrophysical Journal Supplement Series, 111, 73

Kuhn, M. and Johnson, K. (2013), *Applied Predictive Modeling*, Springer

Lacey, C. and Cole, S. (1993), *Merger rates in hierarchical models of galaxy formation*, Monthly Notices of the RAS, 262(3), 627

Landy, S. D. and Szalay, A. S. (1992), *A General Analytical Solution to the Problem of Malmquist Bias Due to Lognormal Distance Errors*, Astrophysical Journal, 391, 494

Lau, E. T., Kravtsov, A. V., and Nagai, D. (2009), *Residual Gas Motions in the Intracluster Medium and Bias in Hydrostatic Measurements of Mass Profiles of Clusters*, Astrophysical Journal, 705(2), 1129

Lau, E. T., Nagai, D., Avestruz, C., et al. (2015), *Mass Accretion and its Effects on the Self-similarity of Gas Profiles in the Outskirts of Galaxy Clusters*, Astrophysical Journal, 806(1), 68

Lau, E. T., Nagai, D., and Nelson, K. (2013), *Weighing Galaxy Clusters with Gas. I. On the Methods of Computing Hydrostatic Mass Bias*, Astrophysical Journal, 777(2), 151

Laureijs, R. et al. (2011), *Euclid Definition Study Report*, arXiv preprint

Laureijs, R., Amiaux, J., Arduini, S., et al. (2011), *Euclid Definition Study Report*, arXiv e-prints, arXiv:1110.3193

Le Brun, A. M. C., McCarthy, I. G., Schaye, J., and Ponman, T. J. (2014), *Towards a realistic population of simulated galaxy groups and clusters*, Monthly Notices of the RAS, 441(2), 1270

Lea, S. M., Silk, J., Kellogg, E., and Murray, S. (1973), *Thermal-Bremsstrahlung Interpretation of Cluster X-Ray Sources*, Astrophysical Journal, Letters, 184, L105

Lebeau, T., Sorce, J. G., Aghanim, N., et al. (2023), *Mass bias in clusters of galaxies: Projection effects on the case study of Virgo replica*, arXiv e-prints, arXiv:2310.02326

Lebeau, T., Sorce, J. G., Aghanim, N., et al. (2024), *Mass bias in clusters of galaxies: Projection effects on the case study of Virgo replica*, Astronomy and Astrophysics, 682, A157

LeCun, Y., Bengio, Y., and Hinton, G. (2015), *Deep learning*, Nature, 521(7553), 436

Lee, M., Villaescusa-Navarro, F., Anglés-Alcázar, D., et al. (2023), *Zooming by in the CARPoolGP lane: new CAMELS-TNG simulations of zoomed-in massive halos*, arXiv preprint arXiv:2410.10942

Lee, M. E., Genel, S., Wandelt, B. D., et al. (2024), *Zooming by in the CARPoolGP Lane: New CAMELS-TNG Simulations of Zoomed-in Massive Halos*, Astrophysical Journal, 968(1), 11

Libeskind, N. I., Carlesi, E., Grand, R. J., et al. (2020), *The HESTIA project: simulations of the Local Group*, Monthly Notices of the Royal Astronomical Society, 498(2), 2968

Libeskind, N. I., Yepes, G., Knebe, A., et al. (2010), *Constrained simulations of the Local Group: on the radial distribution of substructures*, Monthly Notices of the RAS, 401(3), 1889

Linde, A. D. (2005), *Particle physics and inflationary cosmology*, Contemporary Concepts in Physics, 5, Originally published in 1990 by Harwood Academic Publishers

Łokas, E. L., Wojtak, R., Gottlöber, S., et al. (2006), *Mass distribution in nearby Abell clusters*, Monthly Notices of the RAS, 367(4), 1463

Lovisari, L. and Reiprich, T. H. (2019), *The non-uniformity of galaxy cluster metallicity profiles*, Monthly Notices of the RAS, 483(1), 540

LSST Science Collaboration, Abell, P. A., Allison, J., et al. (2009), *LSST Science Book, Version 2.0*, arXiv e-prints, arXiv:0912.0201

Lucy, L. B. (1977), *A numerical approach to the testing of the fission hypothesis.*, Astronomical Journal, 82, 1013

Lynden-Bell, D. (1967), *Statistical mechanics of violent relaxation in stellar systems*, Monthly Notices of the Royal Astronomical Society, 136, 101

Lynden-Bell, D., Faber, S. M., Burstein, D., et al. (1988), *Photometry and Spectroscopy of Elliptical Galaxies. V. Galaxy Streaming toward the New Supergalactic Center*, Astrophysical Journal, 326, 19

Madau, P., Haardt, F., and Rees, M. J. (1999), *Radiative Transfer in a Clumpy Universe. III. The Nature of Cosmological Ionizing Sources*, Astrophysical Journal, 514(2), 648

Makishima, K., Ezawa, H., Fukazawa, Y., et al. (2001), *X-ray probing of the central regions of clusters of galaxies*, Publications of the Astronomical Society of Japan, 53(3), 401

Mandelbaum, R. (2018), *Weak Lensing for Precision Cosmology*, Annual Review of Astron and Astrophys, 56, 393

Mantz, A., Allen, S. W., Ebeling, H., et al. (2010), *The observed growth of massive galaxy clusters - II. X-ray scaling relations*, Monthly Notices of the Royal Astronomical Society, no–no

Mantz, A., Allen, S. W., Rapetti, D., and Ebeling, H. (2010), *The observed growth of massive galaxy clusters - I. Statistical methods and cosmological constraints*, Monthly Notices of the RAS, 406(3), 1759

Mantz, A. B. et al. (2015), *Weighing the giants–IV. Cosmology and neutrino mass*, Mon. Not. R. Astron. Soc., 446, 2205

Maoz, D., Mannucci, F., and Brandt, T. D. (2012), *The delay-time distribution of Type Ia supernovae from Sloan II*, Monthly Notices of the RAS, 426(4), 3282

Marinacci, F., Vogelsberger, M., Mocz, P., and Pakmor, R. (2015), *The large-scale properties of simulated cosmological magnetic fields*, Monthly Notices of the RAS, 453(4), 3999

Markevitch, M. (1998), *The L_X-T Relation and Temperature Function for Nearby Clusters Revisited*, Astrophysical Journal, 504(1), 27

Markevitch, M., Gonzalez, A. H., David, L., et al. (2002), *A Textbook Example of a Bow Shock in the Merging Galaxy Cluster 1E 0657-56*, Astrophysical Journal, Letters, 567(1), L27

Markevitch, M., Vikhlinin, A., and Forman, W. R. (2003), *A High Resolution Picture of the Intracluster Gas*, in *Matter and Energy in Clusters of Galaxies*, edited by S. Bowyer, C.-Y. Hwang, volume 301 of *Astronomical Society of the Pacific Conference Series*, 37

Markevitch, M., Vikhlinin, A., and Mazzotta, P. (2001), *Nonhydrostatic Gas in the Core of the Relaxed Galaxy Cluster A1795*, Astrophysical Journal, Letters, 562(2), L153

Mathews, W. G. and Bregman, J. N. (1978), *Radiative accretion flow onto giant galaxies in clusters.*, Astrophysical Journal, 224, 308

Mathiesen, B., Evrard, A. E., and Mohr, J. J. (1999), *The Effects of Clumping and Substructure on Intracluster Medium Mass Measurements*, Astrophysical Journal, Letters, 520(1), L21

Mathis, H., Lemson, G., Springel, V., et al. (2002), *Simulating the formation of the local galaxy population*, Monthly Notices of the RAS, 333(4), 739

Matsushita, K., Sato, T., Sakuma, E., and Sato, K. (2013), *Distribution of Si, Fe, and Ni in the Intracluster Medium of the Coma Cluster*, Publications of the ASJ, 65, 10

Maughan, B. J. (2014), *Observational biases in galaxy cluster samples and implications for cosmology*, Monthly Notices of the Royal Astronomical Society, 437(1), 1171

Maughan, B. J., Jones, C., Forman, W., and Van Speybroeck, L. (2008), *Images, Structural Properties, and Metal Abundances of Galaxy Clusters Observed with Chandra ACIS-I at 0.1 < z < 1.3*, Astrophysical Journal, Supplement, 174(1), 117

Mazzotta, P., Rasia, E., Borgani, S., et al. (2004a), *Spectroscopic-Like Temperature of Clusters of Galaxies and Cosmological Implications*, arXiv e-prints, astro–ph/0412536

Mazzotta, P., Rasia, E., Moscardini, L., and Tormen, G. (2004b), *Comparing the temperatures of galaxy clusters from hydrodynamical N-body simulations to Chandra and XMM-Newton observations*, Monthly Notices of the RAS, 354(1), 10

McAlpine, S., Helly, J. C., Schaller, M., et al. (2022), *SIBELIUS-DARK: a galaxy catalogue of the local volume from a constrained realization simulation*, Monthly Notices of the RAS, 512(4), 5823

McCarthy, I. G., Babul, A., Bower, R. G., and Balogh, M. L. (2008), *Towards a holistic view of the heating and cooling of the intracluster medium*, Monthly Notices of the Royal Astronomical Society, 386(3), 1309

McDonald, M., Benson, B. A., Vikhlinin, A., et al. (2014), *The Redshift Evolution of the Mean Temperature, Pressure and Entropy Profiles in 80 SPT-Selected Galaxy Clusters*, The Astrophysical Journal, 794(1), 67

McDonald, M., Benson, B. A., Vikhlinin, A., et al. (2013), *The Growth of Cool Cores and Evolution of Cooling Properties in a Sample of 83 Galaxy Clusters at 0.3 < z < 1.2 Selected from the SPT-SZ Survey*, Astrophysical Journal, 774(1), 23

McDonald, M., Gaspari, M., McNamara, B., and Tremblay, G. (2018), *Revisiting the cooling flow problem in galaxies, groups, and clusters of galaxies*, The Astrophysical Journal, 858(1), 45

McNamara, B. R. and Nulsen, P. E. J. (2007), *Heating Hot Atmospheres with Active Galactic Nuclei*, Annual Review of Astron and Astrophys, 45(1), 117

McNamara, B. R. and O'Connell, R. W. (1989), *Star formation in cooling flows in clusters of galaxies*, Astronomical Journal (ISSN 0004-6256), vol. 98, Dec. 1989, p. 2018-2043., 98, 2018

Mei, S., Blakeslee, J. P., Côté, P., et al. (2007), *The ACS Virgo Cluster Survey. XIII. SBF Distance Catalog and the Three-dimensional Structure of the Virgo Cluster*, Astrophysical Journal, 655(1), 144

Melchiorri, A., Ade, P. A. R., de Bernardis, P., et al. (2000), *A Measurement of Ω from the North American Test Flight of Boomerang*, Astrophysical Journal, Letters, 536(2), L63

Mellier, Y., Mathez, G., Mazure, A., et al. (1988), *Subclustering and evolution of the Coma cluster.*, Astronomy and Astrophysics, 199, 67

Mernier, F., Biffi, V., Yamaguchi, H., et al. (2018), *Enrichment of the Hot Intracluster Medium: Observations*, Space Science Reviews, 214(8), 129

Merritt, D. (1983), *Relaxation and tidal stripping in rich clusters of galaxies. I. Evolution of the mass distribution.*, Astrophysical Journal, 264, 24

Mirakhor, M. S. and Walker, S. A. (2020), *A complete view of the outskirts of the Coma cluster*, Monthly Notices of the RAS, 497(3), 3204

Mitsuda, K., Bautz, M., Inoue, H., et al. (2007), *The X-Ray Observatory Suzaku*, Publications of the ASJ, 59, S1

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

Molendi, S., De Grandi, S., Fusco-Femiano, R., et al. (1999), *The BEPPOSAX View of the Hot Cluster Abell 2319*, Astrophysical Journal, Letters, 525(2), L73

Monaghan, J. J. and Lattanzio, J. C. (1985), *A refined particle method for astrophysical problems*, Astronomy and Astrophysics, 149(1), 135

Moore, B., Katz, N., Lake, G., et al. (1996), *Galaxy harassment and the evolution of clusters of galaxies*, Nature, 379(6566), 613

More, S., Miyatake, H., Takada, M., et al. (2016), *Detection of the Splashback Radius and Halo Assembly Bias of Massive Galaxy Clusters*, Astrophysical Journal, 825(1), 39

Morris, J. P. (1996), *A Study of the Stability Properties of Smooth Particle Hydrodynamics*, Publications of the Astronomical Society of Australia, 13, 97

Mukhanov, V. (2005), *Physical foundations of cosmology*, Cambridge university press

Murphy, K. P. (2012), *Machine Learning: A Probabilistic Perspective*, MIT Press

Nagai, D., Kravtsov, A. V., and Vikhlinin, A. (2007), *Effects of galaxy formation on thermodynamics of the intracluster medium*, The Astrophysical Journal, 668(1), 1

Nagai, D., Kravtsov, A. V., and Vikhlinin, A. (2007), *Effects of Galaxy Formation on Thermodynamics of the Intracluster Medium*, Astrophysical Journal, 668(1), 1

Nagai, D. and Lau, E. T. (2011), *Gas Clumping in the Outskirts of ΛCDM Clusters*, Astrophysical Journal, Letters, 731(1), L10

Nagai, D., Vikhlinin, A., and Kravtsov, A. V. (2007), *Testing X-ray measurements of galaxy clusters with cosmological simulations*, The Astrophysical Journal, 655(1), 98

Nagai, D., Vikhlinin, A., and Kravtsov, A. V. (2007), *Testing X-Ray Measurements of Galaxy Clusters with Cosmological Simulations*, Astrophysical Journal, 655(1), 98

NASA (2025), *The Nancy Grace Roman Space Telescope: Science Objectives*, https://roman.gsfc.nasa.gov/science.html, Accessed: 2025-04-14

Navarro, J. F., Frenk, C. S., and White, S. D. M. (1997), *A Universal Density Profile from Hierarchical Clustering*, The Astrophysical Journal, 490(2), 493

Nelson, D., Pillepich, A., Springel, V., et al. (2018), *First Results from the IllustrisTNG Simulations: The Galaxy Colour Bimodality*, Monthly Notices of the Royal Astronomical Society, 475, 624

Nelson, K., Lau, E. T., Nagai, D., et al. (2014), *Weighing Galaxy Clusters with Gas. II. On the Origin of Hydrostatic Mass Bias in ΛCDM Galaxy Clusters*, Astrophysical Journal, 782(2), 107

Neumann, D. M., Arnaud, M., Gastaud, R., et al. (2001), *The NGC 4839 group falling into the Coma cluster observed by XMM-Newton*, Astronomy and Astrophysics, 365, L74

Ni, Y., Genel, S., Anglés-Alcázar, D., et al. (2023), *The CAMELS Project: Expanding the Galaxy Formation Model Space with New ASTRID and 28-parameter TNG and SIMBA Suites*, Astrophysical Journal, 959(2), 136

Ntampaka, M., Avestruz, C., Boada, S., et al. (2019), *The Role of Machine Learning in the Next Decade of Cosmology*, Bulletin of the AAS, 51(3), 14

Nuza, S. E., Dolag, K., and Saro, A. (2010), *Photometric and clustering properties of hydrodynamical galaxies in a cosmological volume: results at $z = 0$*, Monthly Notices of the RAS, 407(3), 1376

Nuza, S. E., Parisi, F., Scannapieco, C., et al. (2014), *The distribution of gas in the Local Group from constrained cosmological simulations: the case for Andromeda and the Milky Way galaxies*, Monthly Notices of the RAS, 441(3), 2593

Oegerle, W. R., Cowie, L., Davidsen, A., et al. (2001), *FUSE Observations of Cooling-Flow Gas in the Galaxy Clusters A1795 and A2597*, Astrophysical Journal, 560(1), 187

Oguri, M. et al. (2018), *Optically selected clusters in the Hyper Suprime-Cam Subaru Strategic Program S16A data: Catalog and weak lensing mass calibration*, Publ. Astron. Soc. Jpn., 70, S20

O'Hara, T. B., Mohr, J. J., and Guerrero, M. A. (2004), *A Chandra Study of the Effects of a Major Merger on the Structure of A2319*, Astrophysical Journal, 604(2), 604

Oppenheimer, B. D. and Davé, R. (2006), *Cosmological simulations of intergalactic medium enrichment from galactic outflows*, Monthly Notices of the RAS, 373(4), 1265

O'Shea, B. W., Bryan, G., Bordner, J., et al. (2004), *Introducing Enzo, an AMR Cosmology Application*, arXiv e-prints, astro–ph/0403044

O'Shea, B. W., Nagamine, K., Springel, V., et al. (2005), *Comparing AMR and SPH Cosmological Simulations. I. Dark Matter and Adiabatic Simulations*, The Astrophysical Journal Supplement Series, 160, 1

O'Hara, T. B., Mohr, J. J., Bialek, J. J., and Evrard, A. E. (2006), *Effects of mergers and core structure on the bulk properties of nearby galaxy clusters*, The Astrophysical Journal, 639(1), 64

Padovani, P. and Matteucci, F. (1993), *Stellar Mass Loss in Elliptical Galaxies and the Fueling of Active Galactic Nuclei*, Astrophysical Journal, 416, 26

Pakmor, R. and Springel, V. (2013), *Simulations of magnetic fields in isolated disc galaxies*, Monthly Notices of the RAS, 432(1), 176

Panagoulia, E. K., Fabian, A. C., and Sanders, J. S. (2014), *A volume-limited sample of X-ray galaxy groups and clusters - I. Radial entropy and cooling time profiles*, Monthly Notices of the RAS, 438(3), 2341

Parekh, V., van der Heyden, K., Ferrari, C., et al. (2015), *Morphology parameters: substructure identification in X-ray galaxy clusters*, Astronomy and Astrophysics, 575, A127

Peebles, P. J. E. (1980), *The large-scale structure of the universe*

Penzias, A. A. and Wilson, R. W. (1965), *A Measurement of Excess Antenna Temperature at 4080 Mc/s.*, The Astrophysical Journal, 142, 419

Peterson, J., Kahn, S., Paerels, F., et al. (2003), *High-resolution X-ray spectroscopic constraints on cooling-flow models for clusters of galaxies*, The Astrophysical Journal, 590(1), 207

Peterson, J., Paerels, F., Kaastra, J., et al. (2001), *X-ray imaging-spectroscopy of Abell 1835*, Astronomy & Astrophysics, 365(1), L104

Pfeifer, S., Valade, A., Gottlöber, S., et al. (2023), *A local universe model for constrained simulations*, Monthly Notices of the RAS, 523(4), 5985

Pillepich, A., Nelson, D., Hernquist, L., et al. (2018), *First Results from the IllustrisTNG Simulations: The Stellar Mass Content of Groups and Clusters of Galaxies*, Monthly Notices of the Royal Astronomical Society, 475, 648

Pillepich, A., Springel, V., Nelson, D., et al. (2018), *Simulating galaxy formation with the IllustrisTNG model*, Monthly Notices of the RAS, 473(3), 4077

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. (2014a), *Planck 2013 results. XXIX. The Planck catalogue of Sunyaev-Zeldovich sources*, Astronomy and Astrophysics, 571, A29

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. (2014b), *Planck 2013 results. XVI. Cosmological parameters*, Astronomy & Astrophysics, 571, A16

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. (2013a), *Planck intermediate results. V. Pressure profiles of galaxy clusters from the Sunyaev-Zeldovich effect*, Astronomy and Astrophysics, 550, A131

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. (2013b), *Planck intermediate results. X. Physics of the hot gas in the Coma cluster*, Astronomy and Astrophysics, 554, A140

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. (2011), *Planck early results. XI. Calibration of the local galaxy cluster Sunyaev-Zeldovich scaling relations*, Astronomy and Astrophysics, 536, A11

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. (2016a), *Planck 2015 results. XXIV. Cosmology from Sunyaev-Zeldovich cluster counts*, Astronomy and Astrophysics, 594, A24

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. (2016b), *Planck intermediate results. XL. The Sunyaev-Zeldovich signal from the Virgo cluster*, Astronomy and Astrophysics, 596, A101

Planck Collaboration, Aghanim, N., Akrami, Y., et al. (2020a), *Planck 2018 results. VI. Cosmological parameters*, Astronomy and Astrophysics, 641, A6

Planck Collaboration, Aghanim, N., Akrami, Y., et al. (2020b), *Planck 2018 Results. VI. Cosmological Parameters*, Astronomy & Astrophysics, 641, A6

Planelles, S., Borgani, S., Fabjan, D., et al. (2013), *Baryon census in hydrodynamical simulations of galaxy clusters*, Monthly Notices of the Royal Astronomical Society, 431(2), 1487

Planelles, S., Borgani, S., Fabjan, D., et al. (2014), *On the role of AGN feedback on the thermal and chemodynamical properties of the hot intracluster medium*, Monthly Notices of the RAS, 438(1), 195

Planelles, S., Fabjan, D., Borgani, S., et al. (2017), *Pressure of the hot gas in simulations of galaxy clusters*, Monthly Notices of the RAS, 467(4), 3827

Planelles, S. and Quilis, V. (2009), *Galaxy cluster mergers*, Monthly Notices of the Royal Astronomical Society, 399(1), 410

Pratt, G. W. et al. (2009), *Gas entropy in galaxy clusters with XMM-Newton*, Astron. Astrophys., 498, 361

Pratt, G. W., Arnaud, M., Biviano, A., et al. (2019), *The outskirts of galaxy clusters*, Space Science Reviews, 215(1), 25

Pratt, G. W., Arnaud, M., Piffaretti, R., et al. (2010), *Gas entropy in a representative sample of nearby X-ray galaxy clusters (REXCESS): relationship to gas mass fraction*, Astronomy and Astrophysics, 511, A85

Pratt, G. W., Böhringer, H., Croston, J. H., et al. (2007), *Temperature profiles of a representative sample of nearby X-ray galaxy clusters*, Astronomy and Astrophysics, 461(1), 71

Pratt, G. W., Croston, J. H., Arnaud, M., and Böhringer, H. (2009), *Galaxy cluster X-ray luminosity scaling relations from a representative local sample (REXCESS)*, Astronomy and Astrophysics, 498(2), 361

Press, W. H. and Schechter, P. (1974a), *Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation*, The Astrophysical Journal, 187, 425

Press, W. H. and Schechter, P. (1974b), *Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation*, The Astrophysical Journal, 187, 425

Price, D. J. (2008), *Modelling Discontinuities and Kelvin Helmholtz Instabilities in SPH*, Journal of Computational Physics, 227, 10040

Price, D. J. (2012), *Smoothed particle hydrodynamics and magnetohydrodynamics*, Journal of Computational Physics, 231(3), 759

Rafferty, D. A., McNamara, B., Nulsen, P., and Wise, M. (2006), *The feedback-regulated growth of black holes and bulges through gas accretion and starbursts in cluster central dominant galaxies*, The Astrophysical Journal, 652(1), 216

Ragagnin, A., Saro, A., Singh, P., and Dolag, K. (2021), *Cosmology dependence of halo masses and concentrations in hydrodynamic simulations*, Monthly Notices of the RAS, 500(4), 5056

Raghunathan, S., Patil, S., Baxter, E., et al. (2019), *Mass Calibration of Optically Selected DES Clusters Using a Measurement of CMB-cluster Lensing with SPTpol Data*, Astrophysical Journal, 872(2), 170

Rasia, E., Borgani, S., Murante, G., et al. (2015), *Cool core clusters from cosmological simulations*, The Astrophysical Journal Letters, 813(1), L17

Rasia, E., Borgani, S., Murante, G., et al. (2015), *Cool Core Clusters from Cosmological Simulations*, Astrophysical Journal, Letters, 813(1), L17

Rasia, E., Mazzotta, P., Borgani, S., et al. (2005), *Mismatch between X-Ray and Emission-weighted Temperatures in Galaxy Clusters: Cosmological Implications*, Astrophysical Journal, Letters, 618(1), L1

Rasmussen, C. E. and Williams, C. K. I. (2006), *Gaussian Processes for Machine Learning*, MIT Press

Ravanbakhsh, S., Lanusse, F., Mandell, D., and et al. (2017), *Estimating Cosmological Parameters from the Dark Matter Distribution using Deep Learning*, arXiv preprint arXiv:1711.02033

Raychaudhury, S. (1989), *The distribution of galaxies in the direction of the 'Great Attractor'*, Nature, 342, 251

Reichert, A., Böhringer, H., Fassbender, R., and Mühlegger, M. (2011), *Observational constraints on the redshift evolution of X-ray scaling relations of galaxy clusters out toz 1.5*, Astronomy &; Astrophysics, 535, A4

Reiprich, T. H. and Boehringer, H. (2002), *The mass function of an X-ray flux-limited sample of galaxy clusters*, The Astrophysical Journal, 567(2), 716

Reiprich, T. H., Sarazin, C. L., Kempner, J. C., and Tittley, E. (2004), *XMM-Newton Observation of the Merging Galaxy Cluster A1644*, Astrophysical Journal, 608(1), 179

Riess, A. G., Casertano, S., Yuan, W., and et al. (2022), *A Comprehensive Measurement of the Local Value of the Hubble Constant with 1 km/s/Mpc Uncertainty from the Hubble Space Telescope and the SH0ES Team*, The Astrophysical Journal Letters, 934(1), L7

Rodriguez-Gomez, V., Genel, S., Vogelsberger, M., et al. (2015), *The merger rate of galaxies in the Illustris simulation: a comparison with observations and semi-empirical models*, Monthly Notices of the RAS, 449(1), 49

Roediger, E. and Brüggen, M. (2007), *Ram pressure stripping of disc galaxies orbiting in clusters - I. Mass and radius of the remaining gas disc*, Monthly Notices of the RAS, 380(4), 1399

Roncarelli, M., Ettori, S., Borgani, S., et al. (2013), *Large-scale inhomogeneities of the intracluster medium: improving mass estimates using the observed azimuthal scatter*, Monthly Notices of the RAS, 432(4), 3030

Roncarelli, M., Gaspari, M., Ettori, S., et al. (2018), *Measuring turbulence and gas motions in galaxy clusters via synthetic Athena X-IFU observations*, Astronomy and Astrophysics, 618, A39

Rossetti, M., Eckert, D., Cavalleri, B. M., et al. (2011), *Cool core remnants in non-cool core clusters*, Astronomy I& Astrophysics, 532, A123

Rossetti, M., Eckert, D., Cavalleri, B. M., et al. (2011), *Back and forth from cool core to non-cool core: clues from radio halos*, Astronomy and Astrophysics, 532, A123

Rossetti, M. and Molendi, S. (2010), *The Cool-Core/Non-Cool-Core Bimodality: The Role of Mergers*, Astronomy I& Astrophysics, 510, A83

Rosswog, S. (2015), *SPH Methods in the Modelling of Compact Objects*, Living Reviews in Computational Astrophysics, 1(1), 1

Rubin, V. C., Ford, Jr., W. K., and Thonnard, N. (1980), *Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 (R=4kpc) to UGC 2885 (R=122kpc).*, Astrophysical Journal, 238, 471

Ruder, S. (2016), *An overview of gradient descent optimization algorithms*, https://arxiv.org/abs/1609.04747, arXiv:1609.04747

Russell, H., McNamara, B., Edge, A., et al. (2013), *Radiative efficiency, variability and Bondi accretion on to massive black holes: the transition from radio AGN to quasars in brightest cluster galaxies*, Monthly Notices of the Royal Astronomical Society, 432(1), 530

Ruszkowski, M., Brüggen, M., and Begelman, M. C. (2004), *Cluster heating by viscous dissipation of sound waves*, The Astrophysical Journal, 611(1), 158

Sakelliou, I., Peterson, J., Tamura, T., et al. (2002), *High resolution soft X-ray spectroscopy of M 87 with the reflection grating spectrometers on*, Astronomy & Astrophysics, 391(3), 903

Sala, L., Valentini, M., Biffi, V., and Dolag, K. (2024), *Supermassive black hole spin evolution in cosmological simulations with OPENGADGET3*, Astronomy and Astrophysics, 685, A92

Sala, L., Valentini, M., Biffi, V., and Dolag, K. (2024), *Supermassive Black Hole Spin Evolution in Cosmological Simulations with OPENGADGET3*, Astronomy and Astrophysics, 685, A92

Sanders, J., Fabian, A., Frank, K., et al. (2010), *Deep high-resolution X-ray spectra from cool-core clusters*, Monthly Notices of the Royal Astronomical Society, 402(1), 127

Sanderson, A. J., Ponman, T. J., and O'Sullivan, E. (2006), *A statistically selected Chandra sample of 20 galaxy clusters–I. Temperature and cooling time profiles*, Monthly Notices of the Royal Astronomical Society, 372(4), 1496

Sarazin, C. L. (1986), *X-Ray Emission from Clusters of Galaxies*, Reviews of Modern Physics, 58, 1

Sarazin, C. L. (2002), *The physics of cluster mergers*, Merging Processes in Galaxy Clusters, 1–38

Saro, A., Borgani, S., Tornatore, L., et al. (2006), *Properties of the galaxy population in hydrodynamical simulations of clusters*, Monthly Notices of the Royal Astronomical Society, 373(1), 397

Sayers, J., Mroczkowski, T., Zemcov, M., et al. (2013), *A Measurement of the Kinetic Sunyaev-Zel'dovich Signal Toward MACS J0717.5+3745*, Astrophysical Journal, 778(1), 52

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

Scheck, D., Sanders, J. S., Biffi, V., et al. (2023), *Hydrostatic mass profiles of galaxy clusters in the eROSITA survey*, Astronomy and Astrophysics, 670, A33

Schmelzle, J., Lucchi, A., Kacprzak, T., and et al. (2017), *Cosmological model discrimination with Deep Learning*, arXiv preprint arXiv:1707.05167

Schmidt, F., Vikhlinin, A., and Hu, W. (2009), *Cluster constraints on f(R) gravity*, Physical Review D, 80(8), 083505

Schmidt, M. (1959), *The Rate of Star Formation.*, Astrophysical Journal, 129, 243

Schwarz, H. E., Corradi, R. L. M., and Melnick, J. (1992), *A catalogue of narrow band images of planetary nebulae.*, Astronomy and Astrophysics, Supplement, 96, 23

Secco, L. F., Samuroff, S., Krause, E., et al. (2022), *Dark Energy Survey Year 3 results: Cosmology from cosmic shear and robustness to modeling uncertainty*, Physical Review D, 105(2), 023515

Seidel, B. A., Dolag, K., Remus, R. S., et al. (2024), *SLOW IV: Not all that is Close will Merge in the End. Superclusters and their Lagrangian collapse regions*, arXiv e-prints, arXiv:2412.08708

Sereno, M. and Ettori, S. (2017), *Bias and scatter in weak-lensing mass estimates of galaxy clusters*, Monthly Notices of the Royal Astronomical Society, 468(3), 3322

Shang, C. and Scharf, C. (2008), *A LOW-REDSHIFT GALAXY CLUSTER X-RAY TEMPERATURE FUNCTION INCORPORATING SUZAKU DATA*, The Astrophysical Journal, 690(1), 879

Shao, H., de Santi, N. S. M., Villaescusa-Navarro, F., et al. (2023), *A Universal Equation to Predict Ω_m from Halo and Galaxy Catalogs*, Astrophysical Journal, 956(2), 149

Shao, H., Villaescusa-Navarro, F., Villanueva-Domingo, P., et al. (2022), *Robust field-level inference with dark matter halos*, arXiv e-prints, arXiv:2209.06843

Sijacki, D., Springel, V., Di Matteo, T., and Hernquist, L. (2007), *A unified model for AGN feedback in cosmological simulations of structure formation*, Monthly Notices of the Royal Astronomical Society, 380(3), 877

Simionescu, A., Allen, S. W., Mantz, A., et al. (2011), *Baryons at the Edge of the X-ray-Brightest Galaxy Cluster*, Science, 331(6024), 1576

Simionescu, A., Werner, N., Mantz, A., et al. (2017), *Witnessing the growth of the nearest galaxy cluster: thermodynamics of the Virgo Cluster outskirts*, Monthly Notices of the RAS, 469(2), 1476

Simionescu, A., Werner, N., Urban, O., et al. (2013), *Thermodynamics of the Coma Cluster Outskirts*, Astrophysical Journal, 775(1), 4

Simionescu, A., Werner, N., Urban, O., et al. (2012), *Large-scale motions in the Perseus Galaxy Cluster*, The Astrophysical Journal, 757(2), 182

Simon, P. (2013), *Improving three-dimensional mass mapping with weak gravitational lensing using galaxy clustering*, Astronomy and Astrophysics, 560, A33

Singh, P., Saro, A., Costanzi, M., and Dolag, K. (2020), *Cosmology dependence of galaxy cluster scaling relations*, Monthly Notices of the RAS, 494(3), 3728

Smartt, S. J. (2009), *Progenitors of Core-Collapse Supernovae*, Annual Review of Astron and Astrophys, 47(1), 63

Sorce, J. G. (2015), *Minimization of biases in galaxy peculiar velocity catalogues*, Monthly Notices of the RAS, 450(3), 2644

Sorce, J. G. (2015), *Minimization of biases in galaxy peculiar velocity catalogues*, Monthly Notices of the Royal Astronomical Society, 450(3), 2644

Sorce, J. G. (2018), *Galaxy clusters in local Universe simulations without density constraints: a long uphill struggle*, Monthly Notices of the Royal Astronomical Society, 478(4), 5199

Sorce, J. G., Blaizot, J., and Dubois, Y. (2019), *Virgo: an unlikely cluster of galaxies because of its environment*, Monthly Notices of the RAS, 486(3), 3951

Sorce, J. G., Courtois, H. M., Gottlöber, S., et al. (2014), *Simulations of the Local Universe constrained by observational peculiar velocities*, Monthly Notices of the RAS, 437(4), 3586

Sorce, J. G., Dubois, Y., Blaizot, J., et al. (2021), *I–A hydrodynamical clone of the Virgo cluster of galaxies to confirm observationally driven formation scenarios*, Monthly Notices of the Royal Astronomical Society, 504(2), 2998

Sorce, J. G., Gottlöber, S., Hoffman, Y., and Yepes, G. (2016), *How did the Virgo cluster form?*, Monthly Notices of the RAS, 460(2), 2015

Sorce, J. G., Gottlöber, S., and Yepes, G. (2020), *The past history of galaxy clusters told by their present neighbours*, Monthly Notices of the RAS, 496(4), 5139

Sorce, J. G., Gottlöber, S., Yepes, G., et al. (2016a), *Cosmicflows constrained local universe simulations*, Monthly Notices of the Royal Astronomical Society, 455(2), 2078

Sorce, J. G., Gottlöber, S., Hoffman, Y., and Yepes, G. (2016b), *How did the Virgo cluster form?*, Monthly Notices of the Royal Astronomical Society, 460(2), 2015

Sorce, J. G., Ocvirk, P., Aubert, D., et al. (2022), *Reionization time of the Local Group and Local-Group-like halo pairs*, Monthly Notices of the RAS, 515(2), 2970

Sorce, J. G. and Tempel, E. (2018), *Galaxy clusters in simulations of the local Universe: a matter of constraints*, Monthly Notices of the Royal Astronomical Society, 476(4), 4362

Spergel, D. et al. (2015), *Wide-Field InfrarRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report*, arXiv preprint

Springel, V. (2005), *The cosmological simulation code GADGET-2*, Monthly notices of the royal astronomical society, 364(4), 1105

Springel, V. (2005), *The cosmological simulation code GADGET-2*, Monthly Notices of the RAS, 364(4), 1105

Springel, V. (2010), *E pur si muove: Galilean-invariant cosmological hydrodynamical simulations on a moving mesh*, Monthly Notices of the RAS, 401(2), 791

Springel, V. (2010), *E Pur Si Muove: Galiliean-invariant Cosmological Hydrodynamical Simulations on a Moving Mesh*, Monthly Notices of the Royal Astronomical Society, 401(2), 791

Springel, V. (2016), *High Performance Computing and Numerical Modelling*, Saas-Fee Advanced Course, 43, 251

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

Springel, V., Di Matteo, T., and Hernquist, L. (2005a), *Modelling feedback from stars and black holes in galaxy mergers*, Monthly Notices of the Royal Astronomical Society, 361(3), 776

Springel, V., Di Matteo, T., and Hernquist, L. (2005b), *Simulating the joint evolution of quasars, galaxies and their large-scale distribution*, Monthly Notices of the Royal Astronomical Society, 361(3), 776

Springel, V. and Hernquist, L. (2002), *Cosmological smoothed particle hydrodynamics simulations: the entropy equation*, Monthly Notices of the RAS, 333(3), 649

Springel, V. and Hernquist, L. (2003), *Cosmological Smoothed Particle Hydrodynamics Simulations: A Hybrid Multiphase Model for Star Formation*, Monthly Notices of the Royal Astronomical Society, 339, 289

Springel, V. and Hernquist, L. (2003), *Cosmological smoothed particle hydrodynamics simulations: a hybrid multiphase model for star formation*, Monthly Notices of the RAS, 339(2), 289

Springel, V., White, S. D. M., Jenkins, A., et al. (2005c), *Simulations of the Formation, Evolution and Clustering of Galaxies and Quasars*, Nature, 435, 629

Springel, V., White, S. D. M., Tormen, G., and Kauffmann, G. (2001), *Populating a cluster of galaxies - I. Results at [formmu2]z=0*, Monthly Notices of the RAS, 328(3), 726

Stevens, A. R. H., Bleem, L., Ho, S., and et al. (2021), *Challenges and Opportunities for Machine Learning in Cosmology*, Nature Astronomy, 5, 460

Sunyaev, R. A. and Zeldovich, Y. B. (1980), *The velocity of clusters of galaxies relative to the microwave background - The possibility of its measurement.*, Monthly Notices of the RAS, 190, 413

Sunyaev, R. A. and Zeldovich, Ya. B. (1972), *The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies*, Comments on Astrophysics and Space Physics, 4, 173

Sutherland, R. S. and Dopita, M. A. (1993), *Cooling Functions for Low-Density Astrophysical Plasmas*, Astrophysical Journal, Supplement, 88, 253

Sutter, P. M. et al. (2014), *A public void catalog from the SDSS DR7 galaxy redshift surveys based on the watershed transform*, Mon. Not. R. Astron. Soc., 442, 462

Tamura, T., Bleeker, J., Kaastra, J., et al. (2001a), *-observations of the cluster of galaxies Abell 496-Measurements of the elemental abundances in the intracluster medium*, Astronomy & Astrophysics, 379(1), 107

Tamura, T., Kaastra, J., Peterson, J., et al. (2001b), *X-ray spectroscopy of the cluster of galaxies Abell 1795 with XMM-Newton*, Astronomy & Astrophysics, 365(1), L87

Teyssier, R. (2002), *Cosmological Hydrodynamics with Adaptive Mesh Refinement. A New High Resolution Code Called RAMSES*, Astronomy and Astrophysics, 385, 337

Thielemann, F. K., Argast, D., Brachwitz, F., et al. (2003), *Nuclear cross sections, nuclear structure and stellar nucleosynthesis*, Nuclear Physics A, 718, 139

Tinker, J., Kravtsov, A. V., Klypin, A., et al. (2008), *Toward a Halo Mass Function for Precision Cosmology: The Limits of Universality*, Astrophysical Journal, 688(2), 709

Tonnesen, S. and Bryan, G. L. (2012), *Star formation in ram pressure stripped galactic tails*, Monthly Notices of the RAS, 422(2), 1609

Torbet, E., Devlin, M. J., Dorwart, W. B., et al. (1999), *A Measurement of the Angular Power Spectrum of the Microwave Background Made from the High Chilean Andes*, Astrophysical Journal, Letters, 521(2), L79

Toro, E. F. (2009), *Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction*, Springer, Dordrecht ; New York, third edition

Trotta, R. (2008), *Bayes in the sky: Bayesian inference and model selection in cosmology*, Contemporary Physics, 49(2), 71

Tulin, S. and Yu, H.-B. (2018), *Dark matter self-interactions and small scale structure*, Physics Reports, 730, 1

Tully, R. B. (2015), *Galaxy Groups: A 2MASS Catalog*, Astronomical Journal, 149(5), 171

Tully, R. B., Courtois, H. M., Dolphin, A. E., et al. (2013), *Cosmicflows-2: The Data*, Astronomical Journal, 146(4), 86

Tully, R. B., Courtois, H. M., and Sorce, J. G. (2016), *Cosmicflows-3*, Astronomical Journal, 152(2), 50

Tully, R. B., Howlett, C., and Pomarède, D. (2023), *Ho'oleilana: An Individual Baryon Acoustic Oscillation?*, Astrophysical Journal, 954(2), 169

Tully, R. B., Kourkchi, E., Courtois, H. M., et al. (2023), *Cosmicflows-4*, The Astrophysical Journal, 944(1), 94

Turner, R. J. (2024), *Cosmology with Peculiar Velocity Surveys*, arXiv e-prints, arXiv:2411.19484

Urban, O., Simionescu, A., Werner, N., et al. (2013), *Azimuthally resolved X-ray spectroscopy to the edge of the Perseus Cluster*, Monthly Notices of the Royal Astronomical Society, 437(4), 3939

Urban, O., Werner, N., Simionescu, A., et al. (2011), *X-ray spectroscopy of the Virgo Cluster out to the virial radius*, Monthly Notices of the RAS, 414(3), 2101

van Daalen, M. P., Schaye, J., Booth, C. M., and Dalla Vecchia, C. (2011), *The Effects of Galaxy Formation on the Matter Power Spectrum: A Challenge for Precision Cosmology*, Monthly Notices of the Royal Astronomical Society, 415, 3649

van de Weygaert, R. and Bertschinger, E. (1996), *Peak and gravity constraints in Gaussian primordial density fields: An application of the Hoffman-Ribak method*, Monthly Notices of the RAS, 281, 84

VanderPlas, J. (2016), *Python data science handbook: Essential tools for working with data*, " O'Reilly Media, Inc."

Vazza, F., Brunetti, G., Kritsuk, A., et al. (2009), *Turbulent Motions and Shocks Waves in Galaxy Clusters Simulated with Adaptive Mesh Refinement*, Astronomy & Astrophysics, 504(1), 33

Vazza, F., Eckert, D., Simionescu, A., et al. (2013), *Properties of gas clumps and gas clumping factor in the intra-cluster medium*, Monthly Notices of the RAS, 429(1), 799

Ventimiglia, D. A., Voit, G. M., Donahue, M., and Ameglio, S. (2008), *Substructure and Scatter in the Mass-Temperature Relations of Simulated Clusters*, Astrophysical Journal, 685(1), 118

Verde, L. (2010), *Statistical methods in cosmology*, arXiv preprint arXiv:1001.5217

Vikhlinin, A., Burenin, R. A., Ebeling, H., et al. (2009), *Chandra Cluster Cosmology Project III: Cosmological Parameter Constraints*, The Astrophysical Journal, 692(2), 1060

Vikhlinin, A., Burenin, R. A., Ebeling, H., et al. (2009a), *Chandra Cluster Cosmology Project. II. Samples and X-Ray Data Reduction*, Astrophysical Journal, 692(2), 1033

Vikhlinin, A., Forman, W., and Jones, C. (1994), *Mass Concentrations Associated with Extended X-Ray Sources in the Core of the Coma Cluster*, Astrophysical Journal, 435, 162

Vikhlinin, A., Forman, W., and Jones, C. (1997), *Another Collision for the Coma Cluster*, The Astrophysical Journal, 474(1), L7

Vikhlinin, A., Kravtsov, A., Forman, W., et al. (2006), *Chandra Sample of Nearby Relaxed Galaxy Clusters: Mass, Gas Fraction, and Mass-Temperature Relation*, The Astrophysical Journal, 640, 691

Vikhlinin, A., Kravtsov, A., Forman, W., et al. (2006), *Chandra Sample of Nearby Relaxed Galaxy Clusters: Mass, Gas Fraction, and Mass-Temperature Relation*, Astrophysical Journal, 640(2), 691

Vikhlinin, A., Kravtsov, A. V., Burenin, R. A., et al. (2009b), *Chandra Cluster Cosmology Project III: Cosmological Parameter Constraints*, Astrophysical Journal, 692(2), 1060

Vikhlinin, A., Markevitch, M., Murray, S. S., et al. (2005), *Chandra Temperature Profiles for a Sample of Nearby Relaxed Galaxy Clusters*, Astrophysical Journal, 628(2), 655

Villaescusa-Navarro, F., Genel, S., Angles-Alcazar, D., et al. (2021), *Robust marginalization of baryonic effects for cosmological inference at the field level*, arXiv e-prints, arXiv:2109.10360

Villaescusa-Navarro, F., Hahn, C., Genel, S., et al. (2021), *CAMELS: Cosmology and Astrophysics with MachinE Learning Simulations*, The Astrophysical Journal, 915(1), 71

Vogelsberger, M., Genel, S., Springel, V., et al. (2014), *Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe*, Monthly Notices of the RAS, 444(2), 1518

Vogelsberger, M., Marinacci, F., Torrey, P., and Puchwein, E. (2020), *Cosmological simulations of galaxy formation*, Nature Reviews Physics, 2(1), 42

Voit, G. M. (2005), *Tracing cosmic evolution with clusters of galaxies*, Reviews of Modern Physics, 77, 207

Voit, G. M. (2005), *Tracing cosmic evolution with clusters of galaxies*, Reviews of Modern Physics, 77(1), 207

Voit, G. M., Kay, S. T., and Bryan, G. L. (2005), *The Baseline Intracluster Entropy Profile from Gravitational Structure Formation*, Monthly Notices of the Royal Astronomical Society, 364, 909

Wang, H., Mo, H. J., Yang, X., et al. (2016), *ELUCID - Exploring the Local Universe with ReConstructed Initial Density Field III: Constrained Simulation in the SDSS Volume*, Astrophysical Journal, 831(2), 164

Watson, D. F. (1981), *Computing the N-Dimensional Delaunay Tessellation with Application to Voronoi Polytopes**, The Computer Journal, 24(2), 167

Weinberger, R., Springel, V., Hernquist, L., et al. (2017), *Simulating galaxy formation with black hole driven thermal and kinetic feedback*, Monthly Notices of the RAS, 465(3), 3291

Weinberger, R., Springel, V., and Pakmor, R. (2020), *The Arepo Public Code Release*, The Astrophysical Journal Supplement Series, 248(2), 32

Weisskopf, M. C., Brinkman, B., Canizares, C., et al. (2002), *An Overview of the Performance and Scientific Results from the Chandra X-Ray Observatory*, Publications of the ASP, 114(791), 1

Weisskopf, M. C., Hester, J. J., Tennant, A. F., et al. (2000), *Discovery of Spatial and Spectral Structure in the X-Ray Emission from the Crab Nebula*, The Astrophysical Journal, 536(2), L81

Wendland, H. (1995), *Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree*, Advances in computational Mathematics, 4, 389

Werner, N., Urban, O., Simionescu, A., and Allen, S. W. (2013), *A uniform metal distribution in the intergalactic medium of the Perseus cluster of galaxies*, Nature, 502(7473), 656

White, S. D. M., Navarro, J. F., Evrard, A. E., and Frenk, C. S. (1993), *Baryons and the cosmic structure*, Nature, 366, 429

White, S. D. M., Navarro, J. F., Evrard, A. E., and Frenk, C. S. (1993), *The baryon content of galaxy clusters: a challenge to cosmological orthodoxy*, Nature, 366(6454), 429

Wiersma, R. P., Schaye, J., and Smith, B. D. (2009), *The effect of photoionization on the cooling rates of enriched, astrophysical plasmas*, Monthly Notices of the Royal Astronomical Society, 393(1), 99

Woosley, S. E. and Weaver, T. A. (1995), *The Evolution and Explosion of Massive Stars. II. Explosive Hydrodynamics and Nucleosynthesis*, Astrophysical Journal, Supplement, 101, 181

Workman, R. L., Burkert, V. D., Crede, V., et al. (2022), *Review of Particle Physics*, Progress of Theoretical and Experimental Physics, 2022(8), 083C01

Woudt, P. A., Kraan-Korteweg, R. C., Lucey, J., et al. (2007), *The Norma cluster (ACO 3627) – I. A dynamical analysis of the most massive cluster in the Great Attractor*, Monthly Notices of the Royal Astronomical Society, 383(2), 445

Xu, G. (1995), *A New Parallel N-Body Gravity Solver: TPM*, The Astrophysical Journal Supplement Series, 98, 355

Xu, H., Kahn, S., Peterson, J., et al. (2002), *High-resolution observations of the elliptical galaxy NGC 4636 with the reflection grating spectrometer on board XMM-Newton*, The Astrophysical Journal, 579(2), 600

Yepes, G., Martínez-Vaquero, L., Gottlöber, S., and Hoffman, Y. (2009), *The CLUES project: Constrained Local UniversE Simulations*, in *AIP Conference Proceedings*, volume 1178, 64–75, American Institute of Physics

Zakamska, N. L. and Narayan, R. (2003), *Models of galaxy clusters with thermal conduction*, The Astrophysical Journal, 582(1), 162

Zaroubi, S., Hoffman, Y., and Dekel, A. (1999), *Wiener Reconstruction of Large-Scale Structure from Peculiar Velocities*, The Astrophysical Journal, 520(2), 413

Zaroubi, S., Hoffman, Y., Fisher, K. B., and Lahav, O. (1995), *Wiener Reconstruction of the Large-Scale Structure*, Astrophysical Journal, 449, 446

Zeldovich, Ya. B. and Sunyaev, R. A. (1969), *The Interaction of Matter and Radiation in a Hot-Model Universe*, Astrophysics and Space Science, 4, 301

Zhang, Y. Y., Finoguenov, A., Böhringer, H., et al. (2008), *LoCuSS: comparison of observed X-ray and lensing galaxy cluster scaling relations with simulations*, Astronomy and Astrophysics, 482(2), 451

Zhuravleva, I., Churazov, E., Arévalo, P., et al. (2015), *Gas density fluctuations in the Perseus Cluster: clumping factor and velocity power spectrum*, Monthly Notices of the RAS, 450(4), 4184

Zhuravleva, I., Churazov, E., Kravtsov, A., et al. (2013), *Quantifying properties of ICM inhomogeneities*, Monthly Notices of the RAS, 428(4), 3274

Zhuravleva, I., Churazov, E., Schekochihin, A. A., et al. (2014), *Turbulent heating in galaxy clusters brightest in X-rays*, Nature, 515(7525), 85

Zhuravleva, I., Churazov, E., Schekochihin, A. A., et al. (2019), *Suppressed Effective Viscosity in the Bulk Intergalactic Plasma*, Nature Astronomy, 3, 832

Zwicky, F. (1933), *Die Rotverschiebung von Extragalaktischen Nebeln*, Helvetica Physica Acta, 6, 110

Acknowledgements

First and foremost, I want to thank my PhD advisor, Prof. Dr. Klaus Dolag, for his guidance and support throughout these years. Thank you for the insightful discussions, the excellent coffee, and, above all, your flexibility in allowing me the independence to pursue the research paths I was most passionate about.

I am also deeply grateful to Ulrich Steinwandel for his mentorship, discussions, and, of course, for the best food in New York City. Thanks as well to Francisco Villaescusa-Navarro and Shy Genel, working with you both at CCA and remotely has always been a pleasure.

I would like to thank all my PhD and postdoc colleagues, both at USM and those who have since moved on, for all the scientific and not-so-scientific discussions, and for being there during both the good times and the challenging moments.

Many thanks to the administrative staff for their positive spirit and for making my life so much easier. Your help with real-world problems, especially every *Reiseantrag*, was invaluable. I can't imagine having navigated all that bureaucracy without your patience and diligence.

Beyond the professional sphere, my profound gratitude goes to Sammy, my family, and my friends for their unwavering support. This milestone would not be worth celebrating without them, and perhaps wouldn't have been possible at all! Whether through a phone call or a message, each one of them has helped and encouraged me to follow this big, starry dream.

I acknowledge support from the grant agreements ANR-21-CE31- 0019/490702358 from the French Agence Nationale de la Recherche/DFG for the LOCALIZATION project. The CAMELS project is supported by NSF grants AST-2108944, AST-2108678, and AST-21080784. The training of the NNs has been carried out using graphics processing units (GPUs) from Simons Foundation, Flatiron Institute, Center of Computational Astrophysics.

So Long, and Thanks for All the Fish!