On galaxy cluster modelling in the context of cosmological analyses

Stefano Gallo

université PARIS-SACLAY

On galaxy cluster modelling in the context of cosmological analyses

Etude du modèle des amas de galaxies dans un contexte cosmologique

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 127, Astronomie & Astrophysique d'Île de France (AAIF)
Spécialité de doctorat: Astronomie et Astrophysique
Graduate School : Physique. Référent : Faculté des sciences d'Orsay

Thèse préparée dans l'unité de recherche **Institut d'Astrophysique Spatiale (Université Paris Saclay, CNRS)**,
sous la direction de **Marian DOUSPIS**, Astronome,
la co-direction de **Nabila AGHANIM**, directrice de recherche

Thèse soutenue à Paris-Saclay, le 23 Septembre 2024, par

Stefano GALLO

Composition du jury

Membres du jury avec voix délibérative

David ELBAZ Directeur de Recherche, Université Paris-Saclay, AIM - CEA	Président
Marco DE PETRIS Maître de conférences (HDR), "Sapienza" University of Rome, Department of Physics	Rapporteur & Examinateur
Alexandro SARO Maître de conférences (HDR), University of Trieste, Department of Physics	Rapporteur & Examinateur
Maret EINASTO Professeure, University of Tartu, Faculty of Science and Technology, Tartu Observatory	Examinatrice

THESE DE DOCTORAT
NNT : 2024UPASP075

université PARIS-SACLAY

ÉCOLE DOCTORALE
Astronomie
et Astrophysique
d'Île-de-France (AAIF)

Titre: Etude du modèle des amas de galaxies dans un contexte cosmologique
Mots clés: Cosmologie - Structures à grande échelle - Amas de galaxies - Méthodes statistiques

Résumé: Les amas de galaxies sont les objets gravitationnellement liés les plus massifs de l'Univers. Ils se forment à partir des pics les plus élevés des champs de densité primordiaux et sont situés aux noeuds d'un réseau filamentaire complexe appelé "toile cosmique". Le nombre d'amas en fonction de la masse et du redshift est une sonde puissante pour contraindre les paramètres du modèle cosmologique. La comparaison entre l'observation des amas et les prédictions théoriques nécessite une modélisation précise de la population d'amas, qui doit tenir compte des caractéristiques observables des amas, ainsi que de leurs relations physiques avec des quantités cachées telles que la masse totale. L'utilisation d'un modèle imprécis peut entraîner des contraintes biaisées sur les paramètres cosmologiques. Avec la nouvelle génération de grands relevés d'amas, qui réduira considérablement les incertitudes statistiques des analyses cosmologiques, il devient crucial d'identifier et de réduire toutes les sources possibles de biais associés à une modélisation inexacte. Il est donc important d'améliorer notre compréhension des processus physiques ayant un impact sur les amas de galaxies, et de tester les impacts des hypothèses de modélisation sur les analyses cosmologiques. Dans cette thèse, j'ai abordé ces questions en me concentrant sur deux aspects: la caractérisation de la distribution de la matière dans les environnements d'amas, au-delà de l'hypothèse commune de symétrie sphérique; et l'impact de l'hypothèse d'un modèle d'amas inexact dans le processus de détection, et son influence sur l'analyse cosmologique. Au sujet de la distribution de la matière dans et autour des amas, j'ai réalisé trois études, en me concentrant en particulier sur la composante gazeuse. Tout d'abord, j'ai étudié statistiquement la distribution azimutale de la matière dans un ensemble d'amas simulés, en quantifiant l'écart par rapport à la symétrie sphérique. J'ai montré que les caractéristiques azimutales du gaz sont fortement corrélées avec celles de la matière noire et avec les propriétés structurelles et physiques de l'amas, ainsi qu'avec le nombre de filaments connectés à l'amas. Deuxièmement, j'ai analysé la détectabilité des structures filamentaires dans la périphérie des amas de galaxies en utilisant des méthodes statistiques, basées sur des observations de rayons X et de galaxies de l'amas Abell 2744. Pour la première fois dans une analyse aveugle des seuls rayons X, j'ai identifié trois filaments connectées à l'amas. À partir de la distribution tridimensionnelle des galaxies, j'ai identifié deux structures filamentaires supplémentaires le long de la ligne de visée, à l'avant et à l'arrière de l'amas. Troisièmement, j'ai entraîné un modèle génératif pour produire des images d'amas de galaxies avec des morphologies réalistes, en évitant les coûts de calcul élevés des simulations cosmologiques. J'ai montré que les images produites par ce modèle présentent des morphologies anisotropes à grande échelle, offrant un réalisme amélioré par rapport aux images à symétrie sphérique générées de manière analytique. À petite échelle, les images générées par le modèle sont plus lisses, plus sphériques et légèrement moins concentrées que les images d'entraînement, en moyenne. Cela peut empêcher leur utilisation à la place de simulations à haute résolution, mais elles peuvent être utiles pour améliorer le réalisme dans des applications à faible résolution. Dans la seconde approche, j'ai étudié l'effet du modèle d'amas dans la détection d'amas de galaxies avec la méthode du matched multi-filter dans le contexte de la mission *Planck* , en étudiant le cas où la population réelle d'amas diffère du modèle supposé dans le template de détection. J'ai montré que la forme du profil de l'amas a un fort impact sur la fonction de complétude, alors que l'effet des morphologies d'amas non sphériques est modéré, et que ces impacts affectent les contraintes cosmologiques, les déplaçant jusqu'à $\sim 1\sigma$.

Title: On galaxy cluster modelling in the context of cosmological analyses
Keywords: Cosmology - Large scale structures - Galaxy Clusters - Statistical methods

Abstract: Galaxy clusters are the most massive gravitationally bound objects in the Universe. They form from the highest peaks in the primordial density fields, and are located at the nodes of a complex filamentary network called the cosmic web. The number of clusters as a function of mass and redshift, known as cluster number counts, has emerged as a powerful probe to constrain the parameters of the cosmological model. Comparing cluster observation with theoretical predictions requires accurate modelling of the cluster population, which needs to account for the clusters' observable characteristics, as well as their physical relationships with hidden quantities like the total mass. The use of an inaccurate cluster model can result in biased constraints on the cosmological parameters. With the new generation of large cluster surveys, which will significantly reduce the statistical uncertainties of cosmological analyses with galaxy clusters, it becomes crucial to identify and reduce all possible sources of biases associated with inaccurate modelling. It is therefore important to improve our understanding of the physical processes impacting galaxy clusters, and to test the possible impacts of simplifying modelling assumptions on the cosmological analyses. In this Thesis, I approached these issues focusing on two aspects: the characterisation of the matter distribution in cluster environments, beyond the common spherical symmetry assumption; and the impact of assuming an inaccurate cluster model in the cluster detection process, and its influence on the cosmological analysis.

Concerning the matter distribution in and around clusters, I performed three studies, focusing in particular on the gas component. First, I investigated statistically the azimuthal distribution of matter in a set of simulated clusters, quantifying the departure from spherical symmetry. I showed that the gas azimuthal features are strongly correlated with the dark matter ones and with the cluster's structural and physical properties, as well as the number of filaments connected to the cluster. Second, I conducted a case study on the detectability of filamentary structures in the outskirts of galaxy clusters using statistical methods, based on X-ray and galaxy observations of the cluster Abell 2744. I combined the results of two techniques: the aperture multipole decomposition and the T-REx filament finder. For the first time in a blind analysis of X-rays alone, I identified three filamentary structures connected to the cluster. From the three-dimensional distribution of galaxies, I identified two additional filamentary structures along the line of sight, in the front and in the back of the cluster. Third, I trained a generative model to produce images of galaxy clusters with realistic morphologies avoiding the high computational costs of cosmological simulations. I showed that the images produced by this model exhibit anisotropic large-scale morphologies, offering improved realism over spherically symmetric analytic generated images. At small scales, the model-generated images appear smoother, more spherical and slightly less concentrated than training images, on average. This may prevent the use of model-generated images in place of high-resolution simulations, but they may be useful to improve realism in low-resolution applications.

In the second approach, I studied the effect of the cluster model in the detection of galaxy clusters with the matched multi-filter method in the context of the *Planck* mission, studying the case in which the real cluster population differs from the model assumed in the detection template. I showed that the shape of the cluster profile has a strong impact on the completeness function, while the effect of non-spherical cluster morphologies is moderate, and that these impacts affect the cosmological constraints, potentially shifting them by up to $\sim 1\sigma$.

Résumé en français

Les amas de galaxies sont les structures gravitationnellement liés les plus massives de l'Univers. Ils se forment à partir des plus hauts sommets du champ de densité primordial, aux noeuds d'un réseau filamentaire complexe appelé la toile cosmique (Bond et al. 1996). Ils se développent hiérarchiquement sous l'effet de la gravité, en accrétant la matière de leur environnement et des filaments qui leur sont connectés, et en fusionnant avec d'autres groupes et amas de galaxies. L'une des principales prédictions des théories de la formation des structures est le nombre de structures liées en fonction de la masse et du décalage vers le rouge (redshift), connu sous le nom de *fonction de masse* (halo mass function, par exemple Press & Schechter 1974; Tinker et al. 2008; Despali et al. 2016). Cette fonction dépend des paramètres cosmologiques, auxquels elle est particulièrement sensible dans sa queue de masse élevée, qui correspond à la gamme de masse typique des amas de galaxies. C'est pourquoi, en comparant le nombre observé d'amas de galaxies avec celui prédit par la fonction de masse, il est possible d'imposer des contraintes sur les valeurs des paramètres cosmologiques. Les amas de galaxies constituent donc une puissante sonde cosmologique (par exemple Rozo et al. 2010; Planck Collaboration et al. 2014d, 2016f; Bocquet et al. 2019; Costanzi et al. 2019; Ghirardini et al. 2024).

Comparer l'observation des amas avec les prédictions théoriques est une tâche difficile, qui nécessite une modélisation précise de la population des amas. Un tel modèle d'amas doit rendre compte avec précision des caractéristiques observées des amas de galaxies à différentes longueurs d'onde, et il doit modéliser les relations physiques sous-jacentes entre les observables et les quantités cachées, telles que la masse totale. Cela se fait généralement à l'aide d'hypothèses simplificatrices sur les propriétés des amas, telles que la symétrie sphérique et l'équilibre hydrostatique. Comme l'utilisation d'un modèle d'amas inexact peut conduire à des contraintes biaisées sur les paramètres cosmologiques (e.g. Salvati et al. 2020), la validité des modèles d'amas actuels et leurs impacts possibles sur les analyses cosmologiques avec les amas de galaxies doivent être évalués. En même temps, une meilleure compréhension des processus physiques complexes ayant un impact sur les amas de galaxies (tels que l'accrétion, les chocs, la turbulence, les processus de rétroaction astrophysique, pour n'en citer que quelques-uns) est importante pour améliorer la modélisation des amas. Ces aspects sont particulièrement pertinents à la lumière de la nouvelle génération de grands relevés d'amas (e.g. Ade et al. 2019; Euclid Collaboration et al. 2019), qui permettront d'augmenter d'un ordre de grandeur la taille des échantillons d'amas, réduisant ainsi de manière significative les incertitudes statistiques des analyses cosmologiques.

Dans cette thèse, j'ai abordé ces questions en me concentrant sur deux aspects : d'une part, j'ai travaillé sur la caractérisation de la distribution de la matière dans les environnements des amas de galaxies, au-delà de l'hypothèse commune de symétrie sphérique ; d'autre part, j'ai étudié l'impact de l'hypothèse d'un modèle d'amas inexact dans le processus de détection des amas de galaxies, et l'influence que cela peut avoir sur les contraintes cosmologiques.

Pour la plupart des applications cosmologiques, les amas de galaxies sont approximés comme des objets isolés à symétrie sphérique. En réalité, les amas sont en général non sphériques (Limousin et al. 2013), et sont connectés à des filaments cosmiques, qui déterminent une accrétion de matière

de manière anisotrope (e.g. Gouin et al. 2021; Rost et al. 2021). J'ai réalisé trois études, présentées dans la partie II de cette thèse, dans lesquelles j'ai étudié la distribution de la matière à l'intérieur et autour des amas, en me concentrant en particulier sur la distribution de la composante gazeuse.

Premièrement, j'ai étudié statistiquement la distribution de la matière dans les régions d'amas, en relation avec les propriétés structurelles et physiques des amas, en utilisant un ensemble d'amas simulés extraits de la simulation IllustrisTNG (Nelson et al. 2019a). J'ai étudié la distribution azimutale du gaz et de la matière noire dans et autour des amas et quantifié leur écart à la symétrie sphérique. Pour cela, j'ai utilisé la technique de décomposition multipolaire dans une ouverture (Schneider & Bartelmann 1997; Gouin et al. 2017, 2020), en particulier les rapports multipolaires β_m (Buote & Tsai 1995; Gouin et al. 2022), qui quantifient l'importance relative de l'ordre multipolaire m par rapport à l'ordre 0, qui représente la symétrie circulaire. J'ai d'abord comparé les caractéristiques azimutales des distributions de gaz et de matière noire (tracées par les rapports multipolaires) à l'intérieur des amas avec les propriétés structurelles des amas (décalage du centre, fraction de masse de la sous-structure et ellipticité), trouvant de fortes corrélations entre les deux ensembles de paramètres. Cela confirme la capacité de la décomposition multipolaire à retracer différentes propriétés structurelles dans un cadre cohérent, et montre la capacité du gaz à retracer efficacement la distribution de la matière noire. Ensuite, j'ai établi des corrélations entre le niveau global d'asymétrie des amas de galaxies, tracé par le paramètre β, et les propriétés physiques non observables des amas, telles que la masse totale de l'amas, l'état dynamique, le taux d'accrétion et le redshift de formation. Dans les périphéries des amas, j'ai montré que l'asymétrie des distributions de gaz et de matière noire retrace la présence de filaments cosmiques connectés à l'amas. De ces résultats, je conclus que la distribution de la matière dans et autour des amas de galaxies est influencée par les différentes propriétés des amas, et qu'elle porte donc des informations statistiques sur l'état des amas et leur histoire d'accrétion.

Deuxièmement, je me suis concentré sur les observations d'un seul amas, Abell 2744 (A2744), comme étude de cas pour la détectabilité des structures filamentaires dans les périphéries des amas de galaxies avec des méthodes statistiques. La détection de filaments dans la périphérie des amas représente un défi considérable, bien qu'il s'agisse d'un aspect crucial de l'étude des propriétés de la matière entrante et de ses interactions complexes avec l'amas. Pour identifier des structures filamentaires dans la périphérie de A2744, j'ai analysé l'émission de rayons X et la distribution des galaxies, en combinant deux techniques : la décomposition multipolaire et le détecteur de filaments T-REx (que j'ai optimisé pour l'utilisation sur des données de rayons X Bonnaire et al. 2020, 2022). Pour la première fois dans une analyse aveugle des seuls rayons X, j'ai identifié trois structures filamentaires connectées à l'amas, dans les directions nord-ouest, est et sud. Les deux premières structures ont également été identifiées dans l'analyse de la distribution des galaxies, avec les deux méthodes, alors que la troisième n'a été clairement identifiée que par T-REx . En appliquant l'algorithme T-REx à la distribution tridimensionnelle des galaxies spectroscopiques, j'ai identifié deux structures filamentaires supplémentaires le long de la ligne de visée, à l'avant et à l'arrière de l'amas. J'ai donc prouvé qu'il était possible de détecter les filaments cosmiques liés aux amas de galaxies d'une manière qui pourrait être automatisée en vue d'une utilisation dans les grands relevés de rayons X. (e.g. XRISM Science Team 2020; CHEX-MATE Collaboration et al. 2021, eROSITA Bulbul et al. 2024).

Troisièmement, j'ai développé un modèle génératif pour produire des images réalistes d'amas de galaxies d'une masse donnée. L'intérêt d'un tel modèle réside dans sa capacité à générer de grands

échantillons d'images d'amas sans avoir à recourir à des simulations cosmologiques coûteuses en temps de calcul. Comme architecture du modèle, j'ai utilisé un autoencodeur variationnel conditionnel basé sur des réseaux de neurones convolutifs (Kingma & Welling 2013; Sohn et al. 2015). Je l'ai entraîné sur des images du paramètre Compton-y d'amas de galaxies, créées à partir de la distribution de gaz des amas de galaxies simulée dans IllustrisTNG. J'ai montré que les images produites par ce modèle présentent des morphologies anisotropes à grande échelle, offrant un réalisme amélioré par rapport aux images analytiques à symétrie sphérique. À petite échelle, les images générées par le modèle semblent plus lisses, plus sphériques et légèrement moins concentrées que les images d'apprentissage, en moyenne. Ceci est confirmé par l'analyse de deux estimateurs morphologiques: le paramètre β, qui décrit le niveau d'anisotropie des images, et le paramètre de concentration, qui évalue la distribution radiale du signal. Ces estimateurs ont également montré que la morphologie des images générées par le modèle est corrélée à la masse de l'amas d'une manière légèrement différente par rapport aux images d'apprentissage. La relation entre la masse de l'amas et le total du signal y intégré dans les images d'entraînement est bien reproduite par les images générées pour les masses faibles et intermédiaires, mais montre des déviations pour les masses élevées. J'ai conclu que la qualité actuelle des images générées par le modèle empêche leur utilisation à la place des simulations à haute résolution, mais elles présentent néanmoins un meilleur réalisme dans leurs caractéristiques morphologiques par rapport aux images à symétrie sphérique générées analytiquement, de sorte qu'elles peuvent être utiles dans les applications à basse résolution.

Dans la partie III de la thèse, je me suis concentré sur l'étude de l'impact des hypothèses de modélisation sur les analyses cosmologiques avec les amas de galaxies. En particulier, j'ai étudié l'effet du modèle d'amas dans la détection des amas de galaxies avec la méthode du matched multi-filter (dans son application spécifique sur les données *Planck*), et son impact sur l'analyse cosmologique à travers la fonction de sélection. J'ai calculé la fonction de complétude des amas de galaxies dans *Planck* en utilisant une approche Monte Carlo, en étudiant le cas où la population réelle d'amas a des propriétés qui diffèrent du modèle supposé dans le template de détection. J'ai créé des images de l'effet Sunyeav-Zel'dovich (SZ Sunyaev & Zeldovich 1970) des amas de galaxie, en utilisant à la fois des amas simulés à partir de la simulation IllustrisTNG, et générés analytiquement à partir de différents profils de pression observés. J'ai injecté ces images dans les cartes *Planck* réelles, et j'ai exécuté l'algorithme de détection avec un modèle des amas fixe pour évaluer les changements de performance, mesurés par la fonction de complétude. J'ai montré que la forme du profil des amas a un impact important sur la complétude: les amas dont le profil est plus abrupt que celui du modèle produisent des fonctions de complétude plus élevées, alors que les profils plus plats conduisent à une complétude plus faible. En étudiant l'impact de la morphologie des amas, j'ai constaté que l'écart par rapport à la symétrie sphérique a un impact modéré sur la complétude, qui tend à augmenter avec la taille de l'amas. J'ai ensuite étudié l'impact sur l'analyse cosmologique *Planck* avec amas de galaxies, du changement de la fonction de complétude, obtenue à partir des amas de profils différents. J'ai trouvé que les contraintes sur les paramètres cosmologiques sont affectées par ce changement, étant décalées d'environ 1σ dans mes tests. Cette étude m'a permis de conclure que la fonction de complétude des amas de galaxies est affectée par un modèle de détection imprécis, et que les incertitudes sur la complétude peuvent, à leur tour, avoir un impact sur les résultats cosmologiques. Par conséquent, ces incertitudes doivent être correctement propagées dans la modélisation du nombre d'amas pour effectuer des analyses

cosmologiques robustes avec les amas de galaxies.

En conclusion, dans cette thèse, j'ai souligné l'importance d'une modélisation précise et réaliste des amas de galaxies dans le contexte des analyses cosmologiques. J'ai montré l'importance et la valeur de considérer les amas de galaxies et leurs périphéries au-delà de la symétrie sphérique, pour découvrir leur dépendance vis-à-vis de leur environnement et améliorer la compréhension des processus physiques importants de l'accrétion de matière. En outre, j'ai montré que les imprécisions dans la modélisation du modèle utilisé pour détecter les amas de galaxies peuvent conduire à des biais dans les contraintes des paramètres cosmologiques, et donc que les incertitudes liées à l'efficacité du processus de détection doivent être incluses dans le pipeline des analyses cosmologiques.

… salimmo sù, el primo e io secondo,
tanto ch'i' vidi de le cose belle
che porta 'l ciel, per un pertugio tondo.
E quindi uscimmo a riveder le stelle.

Dante Alighieri, Inferno - XXXIV

Homme libre, toujours tu chériras la mer !
La mer est ton miroir ; tu contemples ton âme
Dans le déroulement infini de sa lame …

Charles Baudelaire, XIV - L'homme et la mer

Acknowledgements

This Thesis would not have been possible without the help and support of many, many people during the last three years. It was a challenging, but also exciting and enriching experience. I have an immense feeling of gratitude for sharing parts of this path with each and every one of you.

I would like to thank my PhD jury for their interest in my work and their constructive feedback on the manuscript, as well as the valuable discussion during the defence.

A huge thank you goes naturally to my thesis supervisors: Marian and Nabila. You both taught me so much, from a scientific point of view but also from a human one, helping me gain confidence in the value of my work and not get discouraged when things get more difficult. Thank you also for encouraging (and sometimes pushing) me to do the things I normally do less enthusiastically (like writing or being organised). Marian, thank you for your constant support through the numerous turns of my PhD path, for the great help when dealing with unexpected results and when the way forward seemed unclear. I greatly enjoyed our discussions, which were always insightful and often helped me get unstuck by shifting my perspective. Nabila, thank you first of all for welcoming me when I was a confused master student and making me discover how cool galaxy clusters are. Thank you for your enthusiasm and energy, for your strong ideas and ideals, and for your relentless effort to make the team a place where no one feels left behind, but welcomed. Thank you especially for the many discussions, scientific and not, that we had, which helped me grow as a scientist and as a person.

I would also like to thank all the people at IAS, present and past, that have made these three years a wonderful experience. A big thank you to the cosmology team, which became a bit like a second family during these years, with its permanent people, Nabila, Marian, Laura, Bruno, Jenny, Mathieu, Adélie, Julien, Hervé; postdocs, Céline, Hideki, Joseph, Gabriel, Toni, Lisa; and obviously the PhDs, Raphaël, Danilo, Thomas, Hubert, Valentin, Marion, Xavier, Théo, Gaspard, Anaïs, Hugo, Apolline, who rarely failed to make my days better. A big thanks goes to all the *précaires du labo*, thank you for creating an amazing environment, full of occasions of exchange and conviviality, in which I immediately felt welcomed despite being a bit of an "outsider". Thank you for teaching me French, in an immersive environment with a very hands-on approach ;). Thank you also for teaching me how to play Tarot, a knowledge I will do my best to spread in my future workplaces. I'm lacking the words to tell you all how much being a part of this lab meant for me in these three years. I treasure all the memories of the lunches, evenings, SnBs, CosmOuts, days out, conferences and workshops, and I feel really grateful for the time I spent with each one of you.

Another big thanks goes to all my gen friends in Paris: Peter, Paolo, Étienne, Giordano, Thiago, Jerome, JB, Pietro, Elodie, Alex, Adrien, Cyril, Maria Pia, Alex, Viviane, Marco, Carl, and many others. You all made me always feel welcomed and at home, especially in the beginning when I felt a bit lost in this new world I could always rely on you to go out for a beer or a walk. Thank you for the many

picnics we shared on the banks of the Seine, the evenings at La Felicità eating pizza or at someone's house playing board games.

Ancora un altro grazie va a tutti gli amici in giro per il mondo, in particolare a quelli della Balotta: Arturo, Francesco, Matteo, Enrico, Alessandro, Fabiola, Francesco, Giuseppe, Anna ed Elia. Mi avete accompagnato da quando ero uno sbarbatello alle prime armi, e so di poter sempre contare sulla vostra amicizia. Se sono arrivato qui è anche grazie a voi.

Un grazie molto speciale va alla mia famiglia. Mamma e Papà, grazie per tutto l'amore e il sostegno che mi date in quantità infinita, e per spingermi sempre a dare il meglio di me in ogni situazione. Roberto e Francesca, grazie per esserci sempre, per supportarmi e sopportarmi, e per farmi scoprire sempre nuovi orizzonti per crescere insieme. Grazie ai nonni, gli zii e i cugini, sono grato di avervi come famiglia.

The most special of thanks is for Sofia: you've seen all my highs and lows, and you manage to be a beam of light that lightens all my days. I cannot imagine how I could've gone through this PhD without your constant help and support.

Contents

Part I

Introduction

1 - Introduction

Modern cosmology is a fairly young science, whose starting point can be traced back to the first discovery of other galaxies outside our own (Hubble 1925, 1926), and the observation that these galaxies appear to be moving away from us, with velocities proportional to their distances (Hubble 1929). The theoretical explanation for this observation was provided, in the framework of General Relativity (Einstein 1916) by the Friedmann-Lemaitre-Robertson-Walker metric (FLRW Friedmann 1922; Lemaître 1931; Robertson 1935; Walker 1937), which describes a homogeneous, isotropic and expanding Universe. In the following century, many more observations have been performed to probe the Universe's content, structure and history, along with theoretical advancements that incorporated the new observations in a coherent framework (see for example Coles & Lucchin 2002).

The current standard model of cosmology is the so-called ΛCDM model, named after the two main components in the model's energy density budget today: dark energy (modelled as a cosmological constant Λ), which constitutes about 68% of the energy density in the Universe today, and dark matter (in its "cold" version, hence the acronym CDM: cold dark matter), the main matter component in the Universe, making up roughly 27% of the energy density budget (Planck Collaboration et al. 2020a). The remaining $\sim 5\%$ is made of ordinary matter, called baryons in cosmology[1], which is the only component that can be directly observed. The standard cosmological model, in its simplest form, is fully characterised by six parameters (Planck Collaboration et al. 2014e): two of them represent the density of baryons and dark matter ($\Omega_b h^2$ and $\Omega_{CDM} h^2$)[2], two describe the power spectrum of the primordial density fluctuations (A_s and n_s, see Sect. 2.1.1), one refers to the scale of the acoustic oscillations in the early Universe (θ_*) and the last is related to the reionisation of the Universe after the formation of the first stars (τ). All other quantities in the model can be derived from these parameters. The ΛCDM model is able to describe with remarkable accuracy a multitude of processes observed at different scales and different epochs across the Universe's history, from the cosmic microwave background (CMB) to the present-time large scale structures.

One of the main goals of observational cosmology is to use different observations as probes to test the cosmological model and to provide indepen-

[1] With a certain abuse of terminology compared to the particle physics definition.

[2] The density parameters Ω_i represent the fraction of the density of the component i over the critical density of the Universe. The reduced Hubble constant $h = H_0/(100\,\mathrm{km\,s^{-1}Mpc^{-1}})$ represents the rate of the expansion of the Universe today.

dent constraints on its parameters. Currently, the tightest constraints on the ΛCDM parameters are obtained from observations of the CMB anisotropies (Planck Collaboration et al. 2014e, 2016e, 2020a), but many other probes provide competitive constraints on the cosmological parameters, and offer complementary views that can help break degeneracies between parameters when analysed together. Such probes include, for example, the observation of galaxies, via the two-point correlation function or the weak lensing signal (e.g. Eisenstein et al. 2005; Parkinson et al. 2012; Hildebrandt et al. 2017; Hikage et al. 2019; Hamana et al. 2020; Alam et al. 2021; Abbott et al. 2022), expansion rate measurements (e.g. Abbott et al. 2017; Freedman et al. 2019; Birrer et al. 2020; Riess et al. 2021), and galaxy clusters (e.g. Rozo et al. 2010; Mantz et al. 2010; Pierre et al. 2011; Mantz et al. 2015b; Planck Collaboration et al. 2014d, 2016f; de Haan et al. 2016; Böhringer et al. 2017; Pacaud et al. 2018; Costanzi et al. 2019, 2021; Bocquet et al. 2019; Abbott et al. 2020; Bocquet et al. 2024; Ghirardini et al. 2024). For each of these probes, the extraction of cosmological information from the observations requires a detailed modelling of the data, from the physical processes involved in their generation to the instrumental effects during the observations, which often involves several simplifying assumptions. These assumptions need to be validated by extensive testing, and the potential biases and uncertainties that can propagate to the cosmological constraints need to be carefully assessed.

In this Thesis, I focused on galaxy clusters as a probe for cosmology. Galaxy clusters, the most massive gravitationally bound objects in the Universe, have emerged in the last decade as a powerful cosmological probe. Their formation and evolution from the highest peaks in the primordial density fluctuations is tightly related to the overall formation of the Universe's large-scale structures, and depends on the details of the cosmological model (see Chapter 2). Galaxy clusters are complex astrophysical objects, whose observed properties are influenced by several physical processes related to matter accretion, heat transfer and feedback processes, turbulence, shocks, and many others. Therefore, in cosmological contexts, galaxy clusters are modelled starting from some simplifying assumptions, such as spherical symmetry and hydrostatic equilibrium. In light of the new and upcoming cluster surveys (e.g. Euclid Collaboration et al. 2019; Ade et al. 2019; Bulbul et al. 2024), which will provide cluster catalogues with much higher statistical power, the validity of these assumptions and their potential impact on cosmological analyses needs to be verified (e.g. Salvati et al. 2020). At the same time, a deeper understanding of the physical processes impacting galaxy clusters is needed to improve the modelling of the cluster population.

The work presented in this Thesis is inscribed in the efforts to improve the galaxy cluster modelling for cosmology. In particular, I worked on two main aspects, which constitute Part II and Part III of this Thesis. The first focuses

on the study of the matter distribution inside and around galaxy clusters, going beyond spherical symmetry. Indeed, galaxy clusters are not spherically symmetric in general (Limousin et al. 2013), and their position at the nodes of the cosmic web leads to anisotropic accretion from cosmic filaments (e.g. Gouin et al. 2021; Rost et al. 2021). Therefore, in my work, I investigated in particular the distribution of gas in galaxy clusters' environments: i) analysing statistically its departure from spherical symmetry, and how this can be related to the clusters' physical and structural properties (using simulations); ii) applying statistical methods to identify filamentary structures in observed cluster outskirts, with the perspective of studying the impact of these connected structures on the cluster properties, and the physical processes of matter accretion; iii) constructing a generative model of realistic cluster images, that aims to facilitate the production of large samples of clusters with realistic morphologies, without the use of computationally expensive cosmological simulations. The second main aspect of this Thesis is centred on the investigation of the impact of modelling assumptions on cosmological analyses with galaxy clusters. In particular, I studied how the assumption of a cluster model in the matched multi-filter detection method impacts the detection probability of clusters that differ from this model, and how this impact propagates into the cosmological analysis down to the cosmological parameter constraints (Chapter 11).

The Thesis is organised as follows: Part I is dedicated to introducing galaxy clusters as cosmological probes: their formation as part of the Universe's large scale structures (Chapter 2), their observations (Chapter 3), and their use in cosmological analyses (Chapter 4). Chapter 5 expands on the role of approximations in cluster modelling, and presents the themes of the subsequent parts of the thesis in greater detail. In Part II, Chapter 6 presents an overview of the matter asymmetry estimation in and around galaxy cluster, along with the two methods I used: aperture multipole moments and T-REx filament finder. Chapters 7-9 present the three projects on matter distribution in cluster environments mentioned above. In Part III, Chapter 10 explains galaxy cluster detection with the *Planck* satellite, while Chapter 11 contains the work I did on the completeness function, in the case in which the real cluster population differs from the cluster model used in the detection.

2 - The large scale structures of the Universe

The distribution of matter observed in the Universe today is highly inhomogeneous on scales up to hundreds of Mpc. It consists of a hierarchy of objects spanning tens of orders of magnitude in mass, from stars to clusters of galaxies, organised on large scales in a complex network of filaments, nodes, sheets, and voids known as the cosmic web (Bond et al. 1996).

The formation of the large scale structures of the Universe is a fundamental process in cosmology, which originates from the density fluctuations in the primordial Universe and continues to this day, therefore tracing the whole history of the Universe. Galaxy clusters, being the most massive gravitationally bound objects in the Universe, with masses of the order of $\sim 10^{14} - 10^{15}\, M_{\odot}$, lie at the endpoint of this process. Their probability of formation depends on the details of the whole structure formation history, and thus ultimately on the cosmological parameters.

In this Chapter, I describe the theory of structure formation, from the primordial density fluctuations to the formation of bound objects in the cosmic web. For more details, see the following reviews (Coles & Lucchin 2002; Allen et al. 2011; Kravtsov & Borgani 2012; Planelles et al. 2015; Huterer 2023), on which this Chapter is based.

2.1 . The density field and its fluctuations

In the current standard scenario of structure formation, objects in the Universe form hierarchically via gravitational collapse around the peaks in the fluctuations of the primordial density field. These perturbations in the density field are generated during inflation, as quantum fluctuations are stretched to macroscopic scales by the rapid expansion, and form the seeds of all the structures in the Universe today. The density perturbations are usually described by the density contrast field:

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \overline{\rho}}{\overline{\rho}}\,, \tag{2.1}$$

where $\rho(\mathbf{x})$ is the matter density field at position $\mathbf{x}$, and $\overline{\rho}$ is the mean density of the Universe. Regions with positive δ are called overdensities, and those with negative δ underdensities.

In the following, we describe the initial properties of the field δ and sketch the theory of its evolution in the linear regime, that is, when $\delta \ll 1$.

2.1.1 . Statistical description of primordial fluctuations

As we said, the properties of the primordial density perturbation field are set during inflation, and thus depend on the details of the inflationary model. Most models predict $\delta(\mathbf{x})$ to be very close to a homogeneous and isotropic Gaussian random field (e.g. Guth & Pi 1982), so it is customary to take this as an assumption in the description of structure formation. This picture is also confirmed by the observations of the cosmic microwave background (CMB), which shows small perturbations (of the order of 10^{-5}, Planck Collaboration et al. 2014a) consistent with a realisation of a Gaussian random field (Planck Collaboration et al. 2020b).

The statistical properties of a uniform, isotropic Gaussian random field can be fully described by specifying its *power spectrum*, $P(k)$, which quantifies the power of fluctuations at wavenumber k. The power spectrum is defined as the Fourier transform of the two-point correlation function:

$$\langle \delta_{\mathbf{k}} \delta^*_{\mathbf{k}'} \rangle = (2\pi)^3 \, \delta^{(3)}(\mathbf{k} - \mathbf{k}') \, P(k) \,, \tag{2.2}$$

where $\delta_{\mathbf{k}} \propto \int \delta(\mathbf{x}) \exp(-i\mathbf{k}\cdot\mathbf{x}) d^3\mathbf{x}$ is the Fourier-space density contrast, the angular brackets signify ensemble average, and $\delta^{(3)}$ is the Dirac delta function. It is usually assumed to be in the form of a power law,

$$P(k) = A_s \left(\frac{k}{k_*} \right)^{n_s} , \tag{2.3}$$

with two parameters (k_* is a pivot scale): the normalisation A_s, and the spectral index n_s, which is expected to be close to unity. These two parameters are sufficient to describe the properties of the perturbations field, and are thus part of the parameters that define the cosmological model. The best constraints on their values are currently given by *Planck* CMB measurements (Planck Collaboration et al. 2020a).

A related quantity is the variance of the density contrast, smoothed at a certain scale R:

$$\sigma^2(R) = \frac{1}{2\pi^2} \int P(k) \, |\tilde{W}(kR)|^2 \, k^2 \, dk \tag{2.4}$$

where $\tilde{W}(kR)$ is the Fourier transform of a spherical top-hat window function[1]. Evaluating this $\sigma(R)$ at scale $R = 8\,h^{-1}\mathrm{Mpc}$ gives the often-quoted σ_8, which is used as an alternative quantity for the normalisation of the matter power spectrum in cosmological studies with large scale structures. Considering that, on average, we can associate a mass scale, $M = \frac{4\pi}{3}\bar{\rho}R^3$, to the spatial scale R, we can use the two interchangeably. In this way, we can consider the variance of fluctuation σ as a function of the mass scale M, which is used to compute the halo mass function, as we describe in Sect. 2.3.

[1] The analytic expression of $\tilde{W}$ is: $\tilde{W}(y) = \frac{3(\sin y - y\cos y)}{y^3}$.

2.1.2 . Linear evolution

The primordial density fluctuations are not static, but evolve under the effect of gravity. Overdense regions tend to attract matter from their surroundings, thus becoming denser, while at the same time the underdense regions become less dense. In this process, the initially small density fluctuations tend to grow, eventually forming gravitationally bound structures. This process is known as gravitational instability, and was first studied by Jeans (1902).

In the linear regime (i.e. when $\delta \ll 1$) the evolution of the perturbation field can be followed analytically. Considering the matter content of the Universe as a non-relativistic fluid, we can describe it with the continuity, Euler, and Poisson equations. By expressing these equations in terms of the density perturbations, δ, and keeping only the linear terms, we obtain:

$$\begin{aligned} &\frac{\partial \delta}{\partial t} + \frac{1}{a}\,\nabla \cdot \delta = 0 \\ &\frac{\partial v}{\partial t} + Hv = -\frac{1}{a}\nabla\phi - \frac{1}{a}\,c_s^2\,\nabla\delta \\ &\nabla^2\phi = 4\pi G\,\overline{\rho}\,a^2\delta \end{aligned} \tag{2.5}$$

where $a = (1+z)^{-1}$ is the scale factor, which describes cosmic expansion, v and c_s are the fluid's peculiar velocity and sound speed, respectively, $\phi(\mathbf{x})$ is the gravitational potential, and $H = \dot{a}/a$ is the Hubble parameter. Combining the equations 2.5, and expressing the result in Fourier space, we get an equation describing the evolution of the density perturbations in the linear regime:

$$\ddot{\delta}_{\mathbf{k}} + 2H\dot{\delta}_{\mathbf{k}} + \left[\frac{c_s^2 k^2}{a^2} - 4\pi G\overline{\rho}\right]\delta_{\mathbf{k}} = 0\ . \tag{2.6}$$

We see from this equation that, for scales larger than the Jeans scale, $k \ll k_J = \sqrt{4\pi G\overline{\rho}}(a/c_s)$, the "gravity" term in the square bracket (right) dominates over the "pressure" term (left). We can therefore neglect the latter term, which is the only one depending on the wavenumber k, so that the δ evolution is independent of the Fourier mode. Finally, in this regime, the solution for the density contrast can be written as:

$$\delta(\mathbf{x}, a) = \delta_+(\mathbf{x}, a_i)\,D_+(a) + \delta_-(\mathbf{x}, a_i)\,D_-(a) \tag{2.7}$$

where $D_\pm(a)$ are the growing and decaying modes of the perturbation field $\delta(\mathbf{x}, a)$, and $\delta_\pm(\mathbf{x}, a_i)$ the corresponding initial spatial distribution. We only consider the growing modes and call the function $D_+(a)$ the linear growth factor. The growth factor depends on the cosmology, so studying the evolution of cosmic perturbations can give information on the cosmological parameters; in a ΛCDM cosmology, for example, it is given by (Heath 1977):

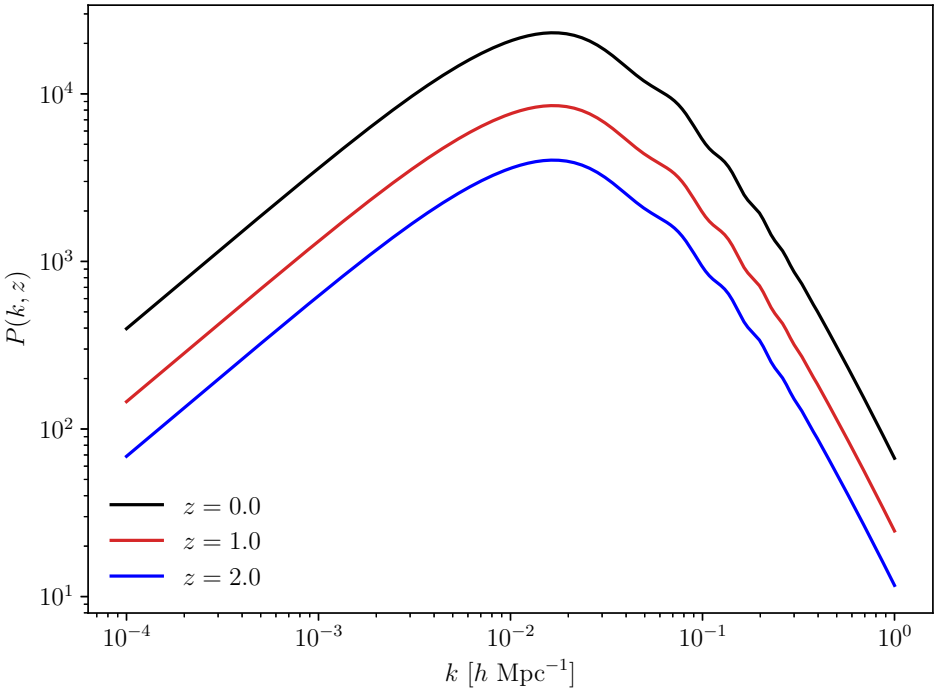

Figure 2.1: Linear matter power spectrum at different redshifts.

$$D_+(a) = \frac{5\Omega_m}{2}\, E(a) \int_0^a \frac{da'}{(a'\, E(a'))^3}\ , \tag{2.8}$$

where $E(a)$ is the normalised expansion rate,

$$E(a) \equiv \frac{H(a)}{H_0} = \sqrt{\Omega_r\, a^{-4} + \Omega_m\, a^{-3} + \Omega_k\, a^{-2} + \Omega_\Lambda}\ . \tag{2.9}$$

Because $\delta(a) \propto D_+(a)$ it is easy to see that the matter power spectrum evolves with the scale factor as $P(k,a) \propto D_+^2(a)$. However, the matter power spectrum is also influenced by different processes in the early Universe (before recombination, $z \sim 1100$), such as the growth suppression during the epoch of radiation domination and the interactions between baryons and photons. These processes influence the shape of the primordial power spectrum, $P_{\rm prim}(k)$, and are collectively described by the transfer function $T(k)$. The full matter power spectrum at late times can thus be written as:

$$P(k,a) = D_+^2(a)\, T^2(k)\, P_{\rm prim}(k)\ . \tag{2.10}$$

In Fig. 2.1 we see the linear matter power spectrum evolved at different redshifts. We see that the main effect of the transfer function is to suppress the fluctuations at small scales, as well as imprint acoustic oscillations on the

power spectrum. We also see that the growth function increases the amplitude of fluctuations on all scales, independently of k.

While of great importance, these results are only valid in the linear regime, when $\delta \ll 1$. When the fluctuations grow and reach $\delta \sim 1$, the linear theory breaks down and cannot be used to study the subsequent growth of structures, such as galaxy clusters, which are typically observed with overdensities $\delta \gtrsim 100$. In this highly non-linear regime, approximate solutions or numerical simulations are needed to follow the structure formation.

2.2 . Non-linear evolution

To study the formation and growth of the cosmic structures that we observe in the Universe we need to go beyond the linear regime, and tackle the full non-linear evolution of density perturbations.

It is possible to solve exactly the non-linear collapse only in the simplest case of an isolated, spherical perturbation with uniform density (the so-called spherical "top hat"). For more realistic cases, one can get some insights using the Zel'dovich approximation (Zel'dovich 1970) or its extensions (e.g. higher orders in Lagrangian perturbation theory, see Buchert 1994), which follows the perturbations in the mildly non-linear regime. Otherwise, the problem can be solved numerically, by performing cosmological simulations.

2.2.1 . Spherical "Top hat" collapse

We start by discussing the case of a single, uniform, spherical overdensity in an expanding background Universe. In its extreme simplicity, this model is nonetheless able to offer important insights into the gravitational collapse process.

We call the initial radius of the perturbation R_i, and the initial overdensity δ_i, so that the total mass of the overdense region is $M = \frac{4\pi}{3}(1+\delta_i)\overline{\rho}R_i^3$. We also assume, for simplicity, a background Universe dominated by matter and a zero initial peculiar velocity at the edge of the perturbation (so that R_i is expanding with the background). In this context, we can treat the perturbation as a separate Universe, which evolves following the Friedmann equations. Solving the equations, one finds that the radius of the overdensity, in the first period, increases slower than the background, decelerating until it reaches its maximum extension, $R_{\rm ta}$. At this moment, called turn-around, the perturbation stops expanding and begins the collapse. During the collapse, some of the kinetic energy is converted into heat (i.e. random motion), either by the pressure or by slight departures from spherical symmetry. Because of this, the collapse eventually stops, and the perturbation reaches an equilibrium state when the virial theorem holds, that is, when $2E_{\rm kin} + E_{\rm grav} = 0$ ($E_{\rm kin}$ and $E_{\rm grav}$ are the kinetic and gravitational energies, respectively).

At virialisation, one finds that the density is eight times the density at turn-around, $\rho_{\rm vir} = 8\rho_{\rm ta}$. Comparing it with the background density, we get that the density contrast of the collapsed structure is:

$$\Delta_{\rm vir} \simeq 178 \,. \tag{2.11}$$

Even though this value depends on the chosen background cosmology (in the case where $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$ it is $\delta_{\rm vir} \sim 100$), the idea of choosing a specific spherical overdensity to define the boundary and mass of virialised objects has become quite popular. Indeed, a commonly used prescription to define the radius of a galaxy cluster is to find the sphere centred on the cluster within which the average density is Δ times a reference density (often the critical density, $\rho_{\rm crit}(z)$, or the mean density, $\rho_{\rm mean}(z)$). Two common choices are $\Delta = 200, 500$, which define the radii $R_{200,500({\rm c,m})}$ and the masses $M_{200,500({\rm c,m})}$ (where the subscript $\rm c$ or $\rm m$ refer to the chosen reference density). The first choice ($\Delta = 200$) is mostly used in the context of cosmological simulations, while $\Delta = 500$ is more common in observations, in particular in X-rays. Throughout this Thesis, I used both values of Δ, but always took the critical density as reference (the subscript $_c$ is thus omitted in the rest of the Thesis).

One can also extrapolate the linear theory prediction and compute the linear overdensity at the time of virialisation:

$$\delta_c \equiv \delta(t_{\rm vir}) \simeq 1.69 \,. \tag{2.12}$$

The value of δ_c can be thought of as the threshold in the linearly evolved overdensity field above which a fluctuation is considered virialised. This is the basic concept at the heart of the halo mass function calculations (see Sect. 2.3).

2.2.2 . The Zel'dovich approximation

Another way to go beyond the range of validity of the linear theory presented in Sect. 2.1.2 is by using the Zel'dovich approximation (Zel'dovich 1970). This approach is essentially a first-order Lagrangian perturbation theory (as opposed to the first-order Eulerian theory of Sect. 2.1.2), and allows one to follow the evolution of density fluctuations to higher values of δ, up to the mildly non-linear regime ($\delta \sim 1$).

The idea of the Zel'dovich approximation is to consider the displacement of matter "particles" from their initial Lagrangian coordinate $\mathbf{q}$:

$$\mathbf{x}(\mathbf{q}, t) = \mathbf{q} - \frac{D_+(t)}{4\pi G\bar{\rho}a^3} \nabla\phi_i(\mathbf{q}) \,, \tag{2.13}$$

where ϕ_i is the initial gravitational potential and follows the Poisson equation $\nabla^2\phi = 4\pi G\bar{\rho}a^2\delta_i$. Using this displacement field to compute the density

contrast, one finds:

$$1+\delta(\mathbf{x},t) = \frac{1}{[1-\lambda_1(\mathbf{q})\,D_+(t)][1-\lambda_2(\mathbf{q})\,D_+(t)][1-\lambda_3(\mathbf{q})\,D_+(t)]} \\ \simeq (\lambda_1+\lambda_2+\lambda_3)\,D_+(t) \tag{2.14}$$

where $\lambda_1 > \lambda_2 > \lambda_3$ are the local eigenvalues of the deformation tensor of the displacement field. The different dependence of the perturbation growth on the local structure of the matter field in this description offers great insights into the formation of cosmic large-scale structures ($\gtrsim 10$ Mpc). Indeed, the perturbations tend to contract in the direction of the positive eigenvalues. So, if all eigenvalues are positive, the collapse happens in all directions, forming (generally) a triaxial object. If two eigenvalues are positive, we have a two-dimensional contraction that leads to a filament-like structure, extended along the direction of the non-positive eigenvalue. Instead, with only one positive eigenvalue, we end up with a two-dimensional sheet-like structure, called a wall. Finally, if all three eigenvalues are negative, the matter in the region tends to escape from it, forming an underdensity called a void. All these structures are not isolated, but related to each other, so that walls tend to form around voids, filaments at the intersection of walls, and nodes at the ends of filaments. Matter in the Universe flows continuously through this hierarchy of structures, the cosmic web (Bond et al. 1996), reaching higher and higher densities and eventually accreting into the nodes.It is precisely in the nodes of the cosmic web that galaxy clusters, the largest gravitationally bound objects in the Universe, form and grow, reaching masses up to a few $10^{15} M_\odot$.

Naturally, with the Zel'dovich approximation one can only follow the beginning of this process. Indeed, this approach fails once the particles come too close to each other and the close-range gravitational forces deviate their paths from the one dictated by the initial gravitational potential. For the full non-linear description of structure formation, one must resort to numerical simulations.

2.2.3 . Cosmological simulations

Cosmological numerical simulations are of utmost importance in the study of the Universe's structure formation, because they are the only reliable way to follow the density perturbations in the fully non-linear regime, and thus get predictions that can be compared with observations (for a review, see e.g. Borgani & Kravtsov 2011; Vogelsberger et al. 2020). For this reason, a considerable amount of work and resources is dedicated to continuously improving our simulations, in terms of accuracy and computational cost (e.g. White 1976; Aarseth et al. 1979; Efstathiou 1979; Evrard 1988; Katz & White 1993; Springel et al. 2005; Schaye et al. 2010, 2015; Dubois et al. 2014; Dolag et al. 2016; Mc-

Carthy et al. 2017; Cui et al. 2018; Nelson et al. 2019a; Villaescusa-Navarro et al. 2021; Schaye et al. 2023).

One main discriminant between cosmological simulations is whether they include or not the treatment of baryonic physics. Those without baryons are known as *dark matter only*, or *N-body* simulations, while those including baryons are called *hydrodynamical* simulations. In N-body simulations (which were the first to be performed historically), a set of N particles is evolved under the sole effect of gravity from some high-redshift initial conditions computed using linear theory. These simulations, being comparatively less expensive than their hydrodynamical counterparts, allow for large volumes to be probed, and are thus used to study the matter distribution and its statistical properties on a large range of scales. Some examples are the study of the matter power spectrum in the full non-linear regime (e.g. Boylan-Kolchin et al. 2009; Springel et al. 2018a; Villaescusa-Navarro et al. 2020), the study of the number of collapsed objects as a function of mass and redshift (i.e. the mass function, discussed in Sect. 2.3), and, on smaller scales, the internal structure of dark matter halos, which led to the observation of a nearly universal density profile, usually described with a Navarro-Frenk-White (NFW, Navarro et al. 1997) or an Einasto (Einasto 1965) profile.

In hydrodynamical simulations, the baryonic matter component is included, and the hydrodynamical interactions are implemented on top of the gravitational forces. The crucial advantage of hydrodynamical simulations is that they provide predictions on the gas and galaxy distributions and properties, that can be tested against observations. Hydrodynamics in cosmological simulations treats baryons as an ideal gas, following Euler equations. But, baryonic physics includes a wealth of processes beyond the simple hydrodynamics that are crucial for the formation of cosmic structures, such as star and black hole formation, and their feedback effects on the gas (like supernova explosions or accretion in supermassive black holes, which can impact), or the radiative cooling of the gas itself, just to name a few (for more details, see Vogelsberger et al. 2020). These processes have characteristic scales that are too small to be simulated explicitly, so they are modelled through effective relations, so-called sub-resolution or sub-grid models, which are calibrated on observations. Different simulation codes have different implementations of these sub-grid models, and different numerical treatment of hydrodynamics (some examples of popular codes are: RAMSES, Teyssier 2002; GADGET, Springel 2005; AREPO Springel 2010). Hydrodynamical simulations have therefore a much higher computational cost compared to N-body ones, but are fundamental for understanding the complex interplay between many different physical phenomena affecting the scales below ~ 10 Mpc, that is, all the scales concerning collapsed objects.

2.3 . The mass function

A crucial prediction of the theories of structure formation is the *halo mass function*, that is, the number density of collapsed objects ("halos") as a function of mass and redshift.

The first analytical approach to compute the mass function was performed by Press & Schechter (1974). The main idea of this approach is to consider that any region with linearly evolved overdensity exceeding the threshold $\delta_c = 1.69$ (set by the spherical collapse model) will collapse into a bound object. To estimate the mass of the resulting collapsed object, we can use the density contrast smoothed on the mass scale M, δ_M. Assuming a Gaussian distribution for the initial density fluctuations, the probability that a particular region collapses into a structure of mass M is:

$$p(M,z) = \frac{1}{\sqrt{2\pi}\sigma(M,z)} \int_{\delta_c}^{\infty} \exp\left(-\frac{\delta_M^2}{2\sigma^2(M,z)}\right) d\delta_M \tag{2.15}$$

where $\sigma(M,z)$ is the variance of fluctuation computed in Sect. 2.1.1, as a function of mass, evolved at redshift z with the linear theory. We obtain the mass function by differentiating this probability with respect to the mass and multiplying by $\overline{\rho}/M$:

$$\frac{dn(M,z)}{dM} = \sqrt{\frac{2}{\pi}} \frac{\overline{\rho}}{M^2} \frac{\delta_c}{\sigma(M,z)} \left|\frac{d\log\sigma(M,z)}{d\log M}\right| \exp\left(-\frac{\delta_c^2}{2\sigma^2(M,z)}\right) . \tag{2.16}$$

Despite its simplicity, the Press-Schechter (PS) mass function, shown in Fig. 2.2, is a pretty good first-order description of halo abundance. From it, we can see that, at low mass, the number of halos evolves as a power law. Higher-mass objects, though, are much less likely to collapse; thus the mass function exhibits an exponential drop above $\sim 10^{14}\,M_\odot$, which is the typical mass range of galaxy clusters. This sharp drop in the mass function means that the expected number of massive objects depends strongly on the cosmological parameters, especially on the density of matter and the amplitude of matter fluctuations. Therefore, galaxy clusters can be a powerful probe of cosmology (as further discussed in Chapter 4).

However, for cosmological analyses with galaxy clusters, the PS mass function is not accurate enough. Instead, different studies proposed mass functions calibrated on numerical simulations, using a functional form inspired by the PS approach (e.g. Tinker et al. 2008; Despali et al. 2016; Murray et al. 2013; Asgari et al. 2023).

Finally, to compare the observed cluster number counts with the theory, we need the predicted number of objects within a given survey area, $\Delta\Omega$, in given bins of mass and redshift (identified as m and l, respectively):

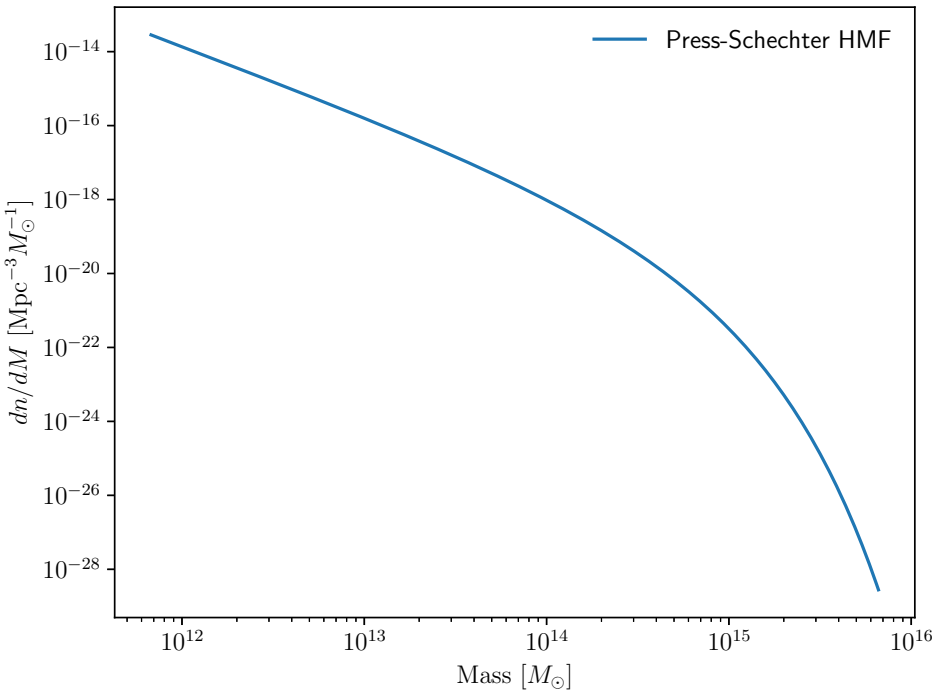

Figure 2.2: Halo mass function from Press & Schechter (1974).

$$N(M_m, z_l) = \frac{\Delta\Omega}{4\pi} \int_{z_l}^{z_{l+1}} dz \frac{dV}{dz d\Omega} \int_{M_m}^{M_{m+1}} dM \, \frac{dn}{dM} \,, \tag{2.17}$$

which depends on cosmology through the mass function and the volume element $dV/dzd\Omega$. This is a key prediction for studying cosmology with galaxy clusters.

3 - Observations of galaxy clusters

In the previous Chapter, I described the formation of the large-scale structures of the Universe, from the small primordial density fluctuations to the present time collapsed halos in the cosmic web. The most massive of these gravitationally bound objects are galaxy clusters, laying at the nodes of the cosmic web. Galaxy clusters are multi-component objects, composed mainly of dark matter (DM) for approximately 80% of their mass. The rest of the cluster mass is in the form of baryons. Of these baryons, only a relatively small fraction is in galaxies (about $4 - 5\%$ of the total), while the majority is in the form of diffuse gas. For the most part, this gas is in a hot plasma phase, with temperatures of the order of $10^7 - 10^8\,\mathrm{K}$, called the Intra-Cluster Medium (ICM) (see e.g. Allen et al. 2011; Kravtsov & Borgani 2012, and references therein).

Due to their multi-component nature, galaxy clusters are observable at different wavelengths across the electromagnetic spectrum. In optical and near-infrared, galaxies are visible through their stars' light. Therefore, they were historically the first observable used to construct galaxy cluster catalogues, observed as galaxy overdensities (e.g. Abell 1958; Zwicky & Kowal 1968). Galaxies in clusters are predominantly concentrated in a narrow region in colour-magnitude diagrams, known as the red sequence (e.g. Bower et al. 1992). This property has been used to detect clusters in photometric surveys, identifying up to tens of thousands of clusters (e.g. Koester et al. 2007; Rykoff et al. 2014; Costanzi et al. 2019; Ansarinejad et al. 2024). Spectroscopic observations of galaxies in clusters are also of primary importance for obtaining precise redshift estimates (e.g. Clerc et al. 2016; Zaznobin et al. 2023). Large galaxy surveys can also be used to observe galaxy clusters indirectly, via gravitational lensing effect. Indeed, when the light from background galaxies passes through a cluster's gravitational field, the galaxy images will appear distorted around the cluster (weak lensing), and, in particular cases, multiple images of the same galaxy will be produced (strong lensing). Gravitational lensing is sensitive to the cluster's total matter distribution, and thus can be used to infer the total cluster mass (for a review, see e.g. Bartelmann 2010; Umetsu 2020). Studying the weak lensing effect over extended portions of the sky, it is also possible to detect galaxy clusters as peaks in the weak lensing maps (Schneider 1996; Maturi et al. 2010; Shan et al. 2012; Oguri et al. 2021).

At X-ray wavelengths, the hot ICM emits thermal bremsstrahlung, due to the free-free interactions of electrons and ions in the plasma, and atomic emission lines. This emission, which scales with the square of the density

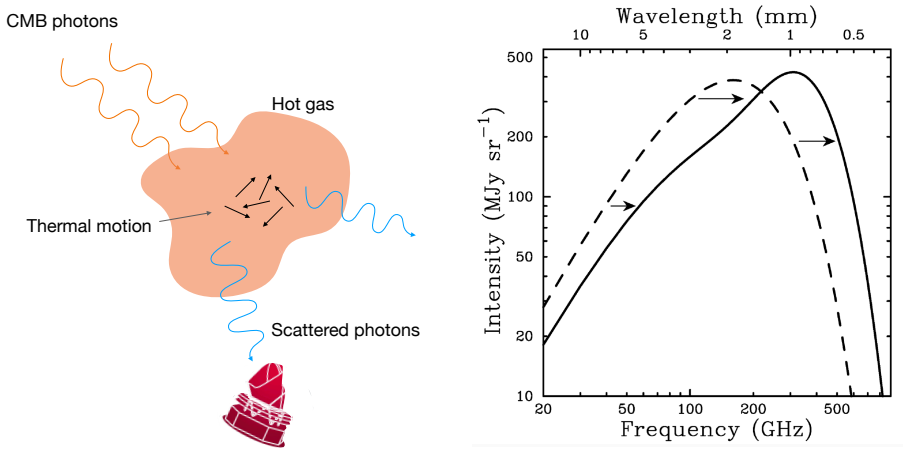

Figure 3.1: *Left*: Schematic representation of the thermal Sunyaev-Zel'dovich effect. *Right*: Distorted CMB spectrum due to thermal SZ effect (solid line), compared to the undistorted one (dashed line). For illustration purposes, the SZ effect shown here is about 1000 times stronger than that of a typical massive galaxy cluster. Figure taken from Carlstrom et al. (2002).

$\sim n_e^2$, makes galaxy clusters appear as bright, extended sources in X-ray observations (Clerc & Finoguenov 2023), allowing the construction of large, X-ray-selected cluster catalogues (e.g. Piffaretti et al. 2011; Pierre et al. 2011, 2016; Finoguenov et al. 2020; Bulbul et al. 2024). Complementary to large surveys, deep pointed observations (e.g. Eckert et al. 2017; CHEX-MATE Collaboration et al. 2021) and spectral analyses of X-rays reveal the gas distribution and its thermodynamical properties in galaxy clusters (like density, temperature, pressure and entropy), which can be used to reconstruct the cluster total mass (e.g. Ghirardini et al. 2019).

At mm wavelengths, the cluster gas is visible through the Sunyaev-Zel'dovich effect (SZ, Sunyaev & Zeldovich 1970, 1972). The SZ effect consists of the inverse Compton scattering of CMB photons off of energetic electrons in the ICM, which produces a local distortion of the CMB spectrum with a characteristic spectral shape (a representation of the SZ effect is shown in Fig. 3.1). For a thermal population of electrons, the SZ effect is proportional to the Compton-y parameter, proportional to the integral along the line of sight of the electron pressure:

$$y = \frac{k_B \sigma_T}{m_e c^2} \int n_e \, T_e \, dl \propto \int P_e \, dl \, , \tag{3.1}$$

where n_e, T_e, and P_e are the electron number density, temperature, and pres-

sure, respectively, $m_e c^2$ is the electron rest mass energy, k_B is the Boltzmann constant and σ_T is the Thomson cross-section (for more details, see Itoh et al. 1998; Birkinshaw 1999). Due to its characteristic spectral shape, the SZ effect from galaxy clusters can be detected in CMB observations (Planck Collaboration et al. 2011, 2014c, 2016d; Bleem et al. 2015; Hilton et al. 2021; Bleem et al. 2024, see also Chapter 10).

3.1 . The cluster mass problem

The total mass of galaxy clusters is a fundamental quantity for their use as cosmological probes, as shown in Sect. 2.3. However, the total mass is not directly accessible, and most observations do not probe the total matter distribution, with the exception of gravitational lensing.

Nevertheless, different methods can be used to obtain cluster mass estimates, given some simplifying assumptions, most notably spherical symmetry and hydrostatic and dynamical equilibrium. For example, from observations of galaxies moving in the cluster's gravitational potential, the cluster total mass (within a sphere of radius R) can be computed from the galaxy density and velocity dispersion profiles (Binney & Tremaine 1987):

$$M_{\rm opt}(<R) = -\frac{r\sigma^2_{v,\rm rad}(R)}{G}\left(\frac{d\ln\rho_{\rm gal}(R)}{d\ln r} + \frac{d\ln\sigma^2_{v,\rm rad}(R)}{d\ln r} - 2\frac{\sigma^2_{v,\rm tang}(R)}{\sigma^2_{v,\rm rad}(R)} + 2\right), \tag{3.2}$$

where $\sigma^2_{v,\rm rad}$ and $\sigma^2_{v,\rm tang}$ are, respectively, the radial and tangential components of the galaxy velocity dispersion, and $\rho_{\rm gal}$ is the galaxy density.

In X-rays, the total mass can be estimated from the gas density profile, $\rho_{\rm gas}(R)$ (obtained from the surface brightness), and temperature profile, $T(r)$ (from X-ray spectroscopy), using the hydrostatic equilibrium equation (Sarazin 1988):

$$M_{\rm X-ray}(<R) = -\frac{rk_BT(R)}{G\mu m_p}\left(\frac{d\ln\rho_{\rm gas}(R)}{d\ln r} + \frac{d\ln T(R)}{d\ln r}\right), \tag{3.3}$$

where k_B is Boltzmann's constant, m_p is the proton mass, and μ is the mean molecular weight.

Alternatively, the use of X-ray spectroscopic data can be avoided combining the X-ray-derived density profile with the gas pressure profile from SZ observations, $P_{\rm gas}(R)$, again using the hydrostatic equilibrium equation:

$$M_{\rm SZ}(<R) = -\frac{r^2}{G\rho_{\rm gas}(R)}\frac{d\ln P(R)}{d\ln r}. \tag{3.4}$$

Finally, in weak lensing studies, the distortion of background galaxies (represented by the complex shear, $\gamma(\boldsymbol{\theta})$, at the angular position $\boldsymbol{\theta}$ on the sky), is

related to the convergence, $\kappa(\boldsymbol{\theta})$, in turn proportional to the projected mass density, via (Pratt et al. 2019):

$$\gamma(\boldsymbol{\theta}) = \frac{1}{\pi}\int d^2\boldsymbol{\theta}'\, D(\boldsymbol{\theta}-\boldsymbol{\theta}')\kappa(\boldsymbol{\theta}')\,, \quad \text{where} \quad D(\boldsymbol{\theta}) = \frac{\theta_2^2-\theta_1^2-2i\theta_1\theta_2}{|\boldsymbol{\theta}|^4}\,, \tag{3.5}$$

and the cluster mass is then estimated fitting a model of the three-dimensional mass distribution to the two-dimensional data.

Each of these methods is based on a set of underlying assumptions, and a significant amount of effort is being dedicated to explore the possible biases of these assumptions on the total mass estimation (see Pratt et al. 2019, for a review). Furthermore, these methods require deep, high-resolution observations in order to obtain precise measurements of the distribution of the various quantities needed. This is not always possible, and typically in cluster samples for cosmology only a few cluster have their total mass estimated in this way.

Therefore, to obtain mass estimates for large numbers of clusters, it is usual to rely on observational proxies correlated with the cluster mass. These correlations are expressed as statistical *scaling relations*, which model the average relationship between the mass and the observable, and the scatter around this average. The theoretical motivation for the existence of such scaling relations is provided by the self-similar model (Kaiser 1986). In its simplest version, this model assumes that clusters of different masses are simply scaled versions of each other (thus self-similar), and are spherically symmetric and in hydrostatic equilibrium. With these assumptions, the cluster thermodynamical properties are directly proportional to its total mass (and a function of redshift) to a given power, of the form:

$$\mathcal{O} \propto M_\Delta^\alpha f(z)^\beta \tag{3.6}$$

where $f(z)$ is usually $f(z) \equiv E(z)$ for gas-related observables, and $f(z) \equiv 1+z$ for galaxy-related observables. One example of observational proxy, typical for X-ray and SZ observations (e.g. Kravtsov et al. 2006; Vikhlinin et al. 2009; Arnaud et al. 2010), is the Y parameter, $Y \propto M_{\rm gas}T$, for which the self-similar relation is $Y \propto E(z)^{2/3}M_\Delta^{5/3}$ (for a more detailed review, see Kravtsov & Borgani 2012). These self-similar scalings, derived from a rather simple model, provide a fairly good first-order description of the scaling relations, and are used as a starting point for more accurate calibrations.

Indeed, for cosmological studies, it is common practice to calibrate the scaling relations using a limited number of clusters with reliable mass estimates to fit the relation with the observable, eventually allowing for variations in the exponents to account for departures from self-similarity (e.g. Kravtsov et al. 2006; Vikhlinin et al. 2009; Andreon 2015; Planck Collaboration et al.

2014d, 2016f; Doubrawa et al. 2023; Ghirardini et al. 2024). This is a fundamental step for cosmological analyses, as shown in Chapter 4, so accurate mass estimates for the calibration sample are crucial, along with precise knowledge of the statistical properties of the sample (i.e. the selection function, see Chapter 10) to ensure that it is representative of the general cluster population.

4 - Cosmology with galaxy clusters

Galaxy clusters, as described in previous chapters, are the most massive bound objects in the Universe, and carry information about the history of structure formation. As such, they are powerful cosmological probes, and many of their properties can be used to constrain the cosmological model. The most common way to probe cosmology with galaxy cluster is by studying their abundance as a function of mass and redshift, as mentioned in Chapter 2, namely their *number counts* (e.g. Rozo et al. 2010; Mantz et al. 2010; Pierre et al. 2011; Mantz et al. 2015b; Planck Collaboration et al. 2014d, 2016f; de Haan et al. 2016; Böhringer et al. 2017; Pacaud et al. 2018; Costanzi et al. 2019, 2021; Bocquet et al. 2019; Abbott et al. 2020; Bocquet et al. 2024; Ghirardini et al. 2024). This is also the method used in this Thesis (see Chapter 11), and it is described in more detail in the rest of this Chapter.

Besides number counts, other properties of galaxy clusters can be used as cosmological probes. For example, their spatial position provide a measure of their clustering properties by studying their 2-point correlation function (e.g. Borgani et al. 1999; Marulli et al. 2018; Fumagalli et al. 2024). Similar to this is the study of the SZ power spectrum, which includes the contribution from unresolved clusters (e.g. Komatsu & Seljak 2002; Planck Collaboration et al. 2016g; Ruppin et al. 2019). Another example is the combination of cluster observations in SZ and X-rays, due to their different dependence on the angular diameter distance d_A, and therefore on cosmology (e.g. Silk & White 1978; Uzan et al. 2004; Wan et al. 2021). The properties of matter inside galaxy cluster also provide interesting windows to probe the cosmological model. One example is the gas mass fraction, defined as $f_{\rm gas} = M_{\rm gas}/M_{\rm tot}$, which can be related to the cosmic baryon fraction, Ω_b/Ω_m (e.g. White et al. 1993; Evrard 1997; Allen et al. 2008; Mantz et al. 2022; Wicker et al. 2023). The cluster mass distribution is yet another property that bears cosmological dependencies, since the way matter accumulates depends on the accretion history, and thus on the broader structure formation history. This relation can be probed via the sparsity, a non-parametric way to quantify the cluster mass concentration (Balmès et al. 2014; Corasaniti et al. 2018, 2021).

4.1 . Cosmology with cluster number counts

Constraining cosmology with galaxy cluster number counts means essentially finding the set of parameters $\boldsymbol{\theta}$, defining a cosmological model, whose predicted number of halos (Eq. 2.17) matches best with the number of observed clusters. This is usually done in a Bayesian framework, which enables

the calculation of the probability of the model parameters given the observations $p(\boldsymbol{\theta}|\mathcal{O})$. This probability, called the posterior, is given by Bayes' theorem (Bayes & Price 1763):

$$p(\boldsymbol{\theta}|\mathcal{O}) = \frac{p(\mathcal{O}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})}{p(\mathcal{O})} \tag{4.1}$$

where $p(\mathcal{O}|\boldsymbol{\theta})$, called the likelihood, is the probability of the data given the parameters, and must incorporate in a statistical description all the ingredients needed to link the model parameters with the observations. A more detailed description of the likelihood function for cosmological analyses with cluster number counts is given in the next section.

The term $p(\boldsymbol{\theta})$ is called prior, and represents the initial knowledge on the model parameters $\boldsymbol{\theta}$, before looking at the data. Such prior knowledge, coming either from previous measurements or theoretical considerations, places constraints on the parameter space and effectively restricts the possible values of some parameters. This is helpful when the data is less sensitive to some of the model parameters. In the case of galaxy clusters, this is true for example for the reduced Hubble constant, h, the baryon density, $\Omega_b h^2$, and the spectral index of primordial fluctuations, n_s, so it is common to apply Gaussian priors on these parameters, coming from observations with other probes (see e.g. Planck Collaboration et al. 2014d, 2016f; Bocquet et al. 2024). In the absence of prior knowledge, it is usual to use so-called uninformative priors, such as uniform priors, Jeffreys (1946) priors, or Jaynes (1968) priors.

The denominator of Eq. 4.1, $p(\mathcal{O})$, is called the evidence and, being independent of the model parameters, it acts as a normalisation factor.

Likelihood

The likelihood function must be carefully constructed to model as closely as possible the observed number counts, in order to avoid biasing the cosmological results with inaccurate modelling. Although the details of the likelihood implementation may vary between different studies (for example in the treatment of multi-wavelength observations, or using a binned or an unbinned likelihood, Rozo et al. 2010; Mantz et al. 2010, 2015b; Planck Collaboration et al. 2014d, 2016f; Bocquet et al. 2019, 2023, 2024; Zubeldia & Bolliet 2024), the essence of the likelihood and its main ingredients remain the same. In this Section, I present these common ingredients, following an approach based on the analysis of Planck Collaboration et al. (2016f), which I used in Chapter 11.

The observed numbers of galaxy clusters in bins of redshift and signal-to-noise ratio (S/N), $N_{i,j}$, are modelled as independent Poisson random variables, so that the log-likelihood takes the form:

$$\ln \mathcal{L} = \sum_{i,j} \left[N_{i,j} \ln \overline{N}_{i,j} - \overline{N}_{i,j} - \ln N_{i,j}! \right] , \tag{4.2}$$

where $\overline{N}_{i,j}$ is the expected number of clusters in bin (i, j), as a function of the parameters $\boldsymbol{\theta}$. This function encapsulates all the modelling from theory to observations and can be decomposed into three main ingredients: the mass function, the mass-observable scaling relations, and the selection function. The mass function, as we discussed in Sect. 2.3, gives the theoretical expectation of the number of clusters as a function of cluster mass and redshift (Eq. 2.17),

$$\frac{dN}{dMdzd\Omega} = \frac{dN}{dVdM}\frac{dV}{dzd\Omega} \,, \tag{4.3}$$

which depends on the cosmological parameters.

The second element is the scaling relations (see Sect. 3.1), which relate the total cluster mass with the chosen observable mass proxy. These are statistical relations, usually modelled as a log-normal distribution where the mean is given by a relation of the type of Eq. 3.6, and the variance represents the intrinsic scatter of the scaling relations:

$$p(\mathcal{O}|M, z, \boldsymbol{\theta}) = \mathcal{N}\big(\ln \overline{\mathcal{O}}(M, z, \boldsymbol{\theta}), \sigma_{\ln \mathcal{O}}\big) \,. \tag{4.4}$$

Finally, the selection function relates the "true" cluster observables to the actual quantity measured, thus it takes into account the observational details of the survey and detection strategy that produced the cluster catalogue. It also incorporates the effect of the selection threshold, usually on the S/N, that defines the catalogue. In this sense, the selection function connects the "true" population of clusters with the objects detected by the survey and included in the considered cluster catalogue. It can be divided into two separate functions: the purity, which is the probability that a given detection corresponds to a real cluster; and the completeness, the probability that a cluster with given "true" observables, at location $\mathbf{x}$ on the sky, is detected by the survey and gets included in the cluster catalogue. The purity can be accounted for by adding a term estimating the number of false detections as a function of the S/N to the $\overline{N}_{i,j}$. The completeness function can be expressed as:

$$\chi(\mathcal{O}, \mathbf{x}) = \int_q^\infty d\xi \, p(\xi|\mathcal{O}, \mathbf{x}) \,, \tag{4.5}$$

where ξ is the variable describing the S/N and q is the threshold above which a cluster detection is included in the catalogue. If we assume pure Gaussian noise in the measurement of $\mathcal{O}$, $p(\xi|\mathcal{O}, \mathbf{x}) \propto \exp\big[-(\xi - \overline{\xi}(\mathcal{O}, \mathbf{x}))^2/2\big]$, then the completeness becomes

$$\chi(\mathcal{O}, \mathbf{x}) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\overline{\xi}(\mathcal{O}, \mathbf{x}) - q}{\sqrt{2}}\right)\right] \,. \tag{4.6}$$

I will present a discussion of the completeness function beyond the Gaussian noise assumption in Chapter 11, in the special case of the *Planck* SZ cluster survey (Planck Collaboration et al. 2016d).

Putting everything together, we can write the expected number of clusters $\overline{N}_{i,j}$ as:

$$\begin{aligned} \overline{N}_{i,j} &= \int_{z_i}^{z_{i+1}} dz \int_{\xi_j}^{\xi_{j+1}} d\xi \left[\frac{dN}{dz d\xi} + \frac{dN_{\rm false}}{d\xi} \right] , \\ \frac{dN}{dz d\xi} &= \int_{\Delta\Omega_{\rm survey}} d\Omega \int dM \, \frac{dN}{dM dz d\Omega} \, p(\mathcal{O}|M, z, \boldsymbol{\theta}) \, p(\xi|\mathcal{O}, \mathbf{x}) \, . \end{aligned} \tag{4.7}$$

Inserting these expressions in Eq. 4.2 we get the complete form of the likelihood function.

5 - Galaxy clusters beyond first approximations

Up to now, I outlined the theoretical framework of cosmic structure formation, and how it predicts the number of collapsed objects as a function of mass and redshift. I then showed how comparing this prediction with the observed number of galaxy clusters we can place constraints on the value of some cosmological parameters, and I highlighted the three main ingredients of such cosmological analyses, namely: a theoretical mass function, a set of scaling relations, and a selection function.

Each of these parts needs to be precisely and accurately determined, and the underlying assumptions need to be tested if we want to obtain tight and unbiased constraints on cosmology. This is particularly relevant in light of the current and upcoming surveys, such as eROSITA (Bulbul et al. 2024) in X-rays, Euclid (Euclid Collaboration et al. 2024) in optical, and CMB-S4 (Abazajian et al. 2022) at mm wavelengths. These surveys will provide improvements of almost an order of magnitude in the total number of detected clusters and extend the redshift range of the detections, thus reducing significantly the statistical errors.

It is then apparent the need for an in-depth investigation of all the possible sources of systematic errors or bias. This is an ongoing effort in the community, with a considerable amount of work aimed at improving each step of the analysis. On the side of the mass function, different studies in recent years tackled the question of its universality, including redshift or cosmology dependencies in the parameters of the mass function or in the halo identification (e.g. Tinker et al. 2008; Bhattacharya et al. 2011; Despali et al. 2016; Del Popolo et al. 2017). In recent years, some studies used machine learning methods to emulate the mass function (McClintock et al. 2019; Bocquet et al. 2020). These models are trained on suites of cosmological N-body simulations with varying cosmology, and should therefore automatically account for cosmology and redshift dependence, while improving the accuracy compared to analytical fits.

On the other hand, the scaling relations and the selection function depend on how accurately galaxy clusters are modelled, and how justified are the simplifying assumptions made in the model. One of the main sources of uncertainties is for example the calibration of the scaling relations. The departure of real clusters from simple assumptions such as spherical symmetry and hydrostatic equilibrium not only introduces scatter in the scaling relation, but may also bias the mean mass-observable relation if the observable depends on other cluster properties other than mass and reshift. The calibration result might also depend on the sample of clusters used, as shown for example in Planck Collaboration et al. (2016f). A cluster model is also assumed

in many detection techniques, as we will detail below, and this can in principle influence the detection efficiency for clusters that differ from the model, thus producing catalogues that are not completely representative of the true population.

Many physical processes can introduce deviations from simple assumptions about clusters. For example, the shape of clusters is in general not spherical (e.g. Binggeli 1982; Limousin et al. 2013; Bonnet et al. 2022), due to the mergers and anisotropic matter accretion that characterise their formation history (e.g. Reiprich et al. 2013; Walker et al. 2019; Walker & Lau 2022; Rost et al. 2021; Gouin et al. 2021, 2022). A cluster's formation history also influences its dynamical state, with more disturbed clusters being further away from hydrostatic equilibrium (e.g. Markevitch et al. 2002; Clowe et al. 2006; Biffi et al. 2016; Campitiello et al. 2022; Cerini et al. 2023). Other sources of deviations from hydrostatic equilibrium include non-thermal pressure, coming for example from turbulence (Rasia et al. 2004; Pearce et al. 2020; Gianfagna et al. 2022, e.g.), magnetic fields (e.g. Dolag & Schindler 2000) and cosmic rays (e.g. Brunetti & Jones 2014; Böss et al. 2023).

From these and other examples stems the interest in studying the properties of clusters as complex astrophysical objects, to improve their models, and at the same time in quantifying the impact of each modelling choice on the cosmological analysis pipeline. This Thesis is part of this process. I focused in particular on two aspects: on the study of cluster properties, I worked on the characterisation of the departure from spherical symmetry, with a special eye for gas distribution in cluster outskirts; on the other side, I studied the impact of the cluster model on the detection process, and in particular on the selection function, which directly enters the cosmological analysis.

5.1 . The shape of clusters and the cosmic web

As mentioned in previous Chapters, galaxy clusters are not isolated objects, but are formed at the nodes of the cosmic web, the end-point of cosmic filaments. They form from the collapse of the highest peaks in the primordial density field, and grow hierarchically through the continuous accretion of matter funnelled by filaments and successive mergers with other galaxy groups and clusters. All these processes are inherently anisotropic and shape the morphology of galaxy clusters away from spherical symmetry. Furthermore, in the hierarchical formation scenario, massive structures are formed more recently, and thus have less time to relax and are in general more disturbed. Putting everything together, it is clear why galaxy clusters are found to be non-spherical in general, both in simulation and observations (Limousin et al. 2013, and references therein). The shape of clusters is related to their mass (e.g. Allgood et al. 2006; Despali et al. 2014; Vega-Ferrero et al. 2017;

Gouin et al. 2020, 2021), and it can be used to test the cosmological model (Sereno et al. 2018). Despite this evidence, spherical symmetry is still a common assumption in galaxy cluster studies for cosmology. The impact of departure from spherical symmetry needs therefore to be assessed at each stage of the pipeline. For example, it has been shown that the asphericity of clusters can impact their mass estimation (e.g. Lee et al. 2018; Lebeau et al. 2024).

The outskirts of galaxy clusters are even more impacted by the anisotropic infall from filaments, and are the places where most physical processes related to matter accretion take place. These include shocks, turbulence, bulk motions, galaxy preprocessing, inhomogeneous gas distribution and the appearance of non-equilibrium electrons (see, for a review Reiprich et al. 2013; Walker et al. 2019; Walker & Lau 2022). Understanding these processes and their impact on the whole cluster is crucial for the interpretation of high-resolution observations, and to avoid biases on the mass estimation.

In this Thesis, I focus on the distribution of gas in galaxy clusters, in particular in their outskirts. In Chapter 7, I studied the azimuthal symmetries of gas distribution, from a sample of simulated clusters, at different distances from the cluster centre, and related them to different cluster properties (published in Gouin et al. 2022). In Gallo et al. (2024a), I used two different statistical techniques to characterise the distribution of gas in the outskirts of the cluster Abell 2744, observed in X-rays, and detect the signatures of connected filaments. This work is presented in Chapter 8. Finally, I worked on a generative machine learning model to produce realistic synthetic images of the SZ signal from galaxy clusters, using simulated clusters as training data. This work is shown in Chapter 9.

5.2 . Detection of galaxy clusters

Detecting a galaxy cluster means recognising in the data a specific signature, that can be unequivocally associated with the presence of a cluster at that position. Choosing such a signature is the first step in building a detection technique, and depends on the kind of data one is planning to analyse, as well as the prior knowledge of the properties of clusters and, by extension, on the cluster model one is assuming. Such a signature, combined with the detection strategy, must be specific enough to clusters so that other astrophysical signals or noise fluctuations are not misidentified as clusters (i.e. the method must have high *purity*). At the same time, it must be flexible enough to account for cluster-to-cluster variations, so that all clusters with a given signal are detected with equal probability (i.e. the *completeness* is a function only of the signal-to-noise ratio). These requirements on the detection method ensure that the resulting cluster sample is pure and representative of the full cluster population.

Many different approaches have been developed for the detection of galaxy clusters, for observations at different wavelengths. All of these rely, to some extent, on assumptions about the cluster population.

In optical surveys, some examples include the use of the cluster red sequence (i.e. the observation that galaxy clusters tend to have a population of passive, elliptical, red galaxies Gladders & Yee 2000; Rykoff et al. 2014), percolation algorithms (e.g. Dalton et al. 1997), wavelet filtering (Eisenhardt et al. 2008; Gonzalez 2014), and matched filtering techniques (Postman et al. 1996; Bellagamba et al. 2018). In X-ray surveys, the detection of clusters is more complex. For example, in eROSITA (Brunner et al. 2022; Bulbul et al. 2024), the detection proceeds in several steps, first identifying all sources brighter than the background with a sliding-cell algorithm, and then fitting them with a source model to characterise their properties and select only the extended sources.

At mm-wavelengths, the characteristic spectral shape of the SZ effect provides a clear signature of the presence of hot ionised gas, heated in the deep gravitational potential well of clusters. Combining this well-described spectral shape with a model of the spatial distribution of SZ signal in clusters, the matched multi-filter technique (MMF, Herranz et al. 2002; Melin et al. 2006) has been used to detect galaxy clusters from *Planck* (Planck Collaboration et al. 2011, 2014c, 2016d), the South Pole Telescope (SPT, Williamson et al. 2011; Bleem et al. 2015, 2024), and the Atacama Cosmology Telescope (ACT, Hilton et al. 2021) data.

In this Thesis, I focus on this last method, and in particular its application to the *Planck* data. I studied the case where the cluster population present in the data differs from the model assumed in the detection template, and what are its effects on the completeness function. This work is detailed in Chapter 11, based on Gallo et al. (2024b).

Part II

Galaxy clusters beyond spherical symmetry

6 - Probing the anisotropies in matter distribution: overview and methods

The theme of this second Part of the Thesis is to probe the shape and distribution of matter in and around galaxy clusters, with a special focus on the gas component.

Inside clusters, better statistical knowledge of their shape and departure from spherical symmetry is crucial in order to improve our modelling of clusters and avoid biases in their mass estimation. Furthermore, anisotropies in the gas distribution are expected to be associated with disturbed dynamical states, and thus a departure from hydrostatic equilibrium. For this reason, gas anisotropies can be used to estimate the clusters' dynamical state.

Several techniques have been developed to assess the different ways the gas distribution in galaxy clusters departs from spherical symmetry. These include: the asymmetry parameter (Schade et al. 1995), the centroid shift (Mohr et al. 1993; O'Hara et al. 2006), the axial ratio (Mohr et al. 1995), the light concentration (Santos et al. 2008), the Gaussian fit parameter (Cialone et al. 2018), the Zernike polynomials decomposition (Capalbo et al. 2021), the azimuthal scatter (Vazza et al. 2011), wavelet analysis (Pierre & Starck 1998), minkowski functionals (Beisbart et al. 2001). In order to enhance the effectiveness of the classification process and minimise the potential biases, many studies used combinations of these and other estimators (e.g. Rasia et al. 2013; Mantz et al. 2015a; Lovisari et al. 2017; Andrade-Santos et al. 2017; De Luca et al. 2021; Ghirardini et al. 2022; Campitiello et al. 2022).

In this Thesis, I focus on the aperture multipole moments decomposition (Schneider & Bartelmann 1997) and the associated multipolar ratios (Buote & Tsai 1995; Gouin et al. 2022) to study the azimuthal distribution of gas inside clusters. This technique has been used before to characterise the shape of galaxy clusters and their outskirts in both observations and simulation, from weak lensing maps (Schneider & Bartelmann 1997; Dietrich et al. 2005; Mead et al. 2010; Gouin et al. 2017) and galaxy distribution (Gouin et al. 2020). The details of the method are presented in Sect. 6.1.

In cluster outskirts, the gas distribution is even more anisotropic due to the presence of cosmic filaments that connect to the cluster. Quantifying the impact of filaments on the cluster gas distribution and properties is particularly interesting to investigate the effects of the various accretion physics processes (e.g. shocks, turbulence, etc.; see e.g. Walker et al. 2019) on the cluster.

To study the transition between filaments and clusters, it is first fundamental to detect the filamentary structures connected to the cluster outskirts.

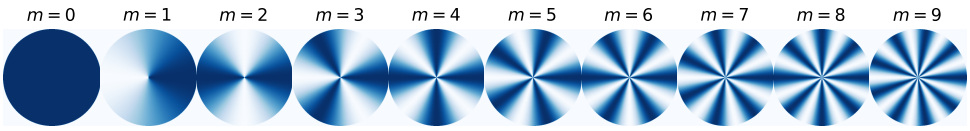

Figure 6.1: Illustration of the different angular symmetries associated to the different multipole orders m.

Nonetheless, this is a difficult task, especially for the gas component. This is due to the lower density and temperature of gas in filaments compared to clusters, which makes their signal fainter both in X-rays and Sunyaev-Zel'dovich (SZ) effect, and thus requires high sensitivity and low, stable noise.

Observations of gas in filaments are so far limited to stacked observations (e.g. de Graaff et al. 2019; Tanimura et al. 2019, 2020, 2022), or in very particular cases, between cluster pairs or around single clusters (e.g. Werner et al. 2008; Planck Collaboration et al. 2013b; Eckert et al. 2015; Bulbul et al. 2016; Bonjean et al. 2018; Veronica et al. 2024).

On the other hand, many different algorithms have been developed to automatically identify cosmic filaments, mostly based on dark matter or galaxy distributions, using many different approaches (see Libeskind et al. 2018, for a detailed review), for example topology-based methods (Aragón-Calvo et al. 2010a; Sousbie 2011), hessian-based methods (Hahn et al. 2007; Cautun et al. 2013), geometry-based methods (Tempel et al. 2016; Pereyra et al. 2020). Among the latter group is the T-REx filament finder (Bonnaire et al. 2020, 2022), which I used in my thesis work. This algorithm identifies filamentary structures in a discrete set of data points, using a graph structure to trace the ridges of the point distribution. T-REx has already been applied successfully on simulations to trace the cosmic web (Bonnaire et al. 2020; Gouin et al. 2021; Rowntree et al. 2024) and on 2D and 3D observed galaxy distributions (Aghanim et al. 2024). In the work presented in Chapter 8, I applied T-REx for the first time on an X-ray image, to identify filaments in the outskirts of a galaxy cluster. A brief presentation of the T-REx algorithm and its most relevant free parameters is given in Sect. 6.2.

6.1 . Aperture multipole moments

The aperture multipole decomposition is a method of analysing the azimuthal behaviour of a two-dimensional field within an aperture defined by the annulus $\Delta R = [R_{\rm min}, R_{\rm max}]$, where $R_{\rm min}$ and $R_{\rm max}$ are concentric radii. The field is decomposed into harmonic orders, each of which is associated with a particular symmetry, as shown in Fig. 6.1 (as a rule of thumb, each order m is associated with m regions of "high signal"). This technique is well suited to the task of distinguishing the contribution of the different angular scales to

the field considered. By focusing on the lower (larger) multipoles, one obtains information on the large-scale (small-scale) behaviour of the field.

For a generic 2D field $\Sigma(R, \phi)$ (where ϕ is the azimuth angle), the multipole of order m in the aperture ΔR can be computed as:

$$Q_m(\Delta R) = \int_{\Delta R} \Sigma(R, \phi)\, e^{im\phi}\, R\, dR\, d\phi\,. \tag{6.1}$$

Given the integration in the radial aperture in Eq. 6.1, this decomposition is sensitive only to azimuthal patterns of the 2D field inside the aperture. Therefore, it is particularly adapted for fields that exhibit little radial dependence.

The advantage of using such a decomposition is that different multipolar orders are related to specific features of the field. For example, the dipole, $m = 1$, is a tracer of the asymmetry along an axis, the quadrupole, $m = 2$, traces the elongation, and can be related to the ellipticity (see e.g. Gouin et al. 2020, as well as Chap. 7), the octapole, $m = 3$, has been related to bimodal distributions (e.g. Buote & Tsai 1995; Rasia et al. 2013; Campitiello et al. 2022), and so on. With this method, all these pieces of information are obtained in the same framework, and their different contributions can be easily compared. Another method with similar characteristics is the Zernike polynomials decomposition, which uses a different basis and includes a radial dependence, thus each order is defined by two parameters.

In order to compare the results of the multipole decomposition among different apertures or different fields, the values of Q_m are not the most adapted quantities, as they depend on the normalization of the field. In fact, it is the relative power between different orders that gives information about the field structure. Therefore, we consider the multipolar ratios β_m (Buote & Tsai 1995; Gouin et al. 2022), defined as the ratio between the modulus of the multipole of order m and that of order zero:

$$\beta_m = \frac{|Q_m|}{|Q_0|} \tag{6.2}$$

where Q_0 is the total amount of signal in the aperture, $Q_0 = \int_{\Delta R} \Sigma\, dA$. These ratios have two advantages over the use of Q_m: they are normalised, making them comparable among different apertures; at the same time, they quantify the relative amount of power of the multipole m with respect to the circular symmetry, thus probing the level of asymmetry of the considered distribution. For these reasons, they have been used in the study of the morphology of galaxy clusters (e.g. Buote & Tsai 1995; Rasia et al. 2013; Campitiello et al. 2022).

The multipole moments Q_m are complex numbers, and thus contain information not only on the power of the multipole of order m in the aperture, but also on its phase. Therefore, by re-summing the harmonic signatures of the

multipole orders according to their power and phase, it is possible to reconstruct the azimuthal structure of the original field up to a chosen order, that is, down to a chosen angular scale. In this way, it is possible to discriminate the large-scale contributions to the field from the small-scale ones, in particular keeping the former and neglecting the latter. Writing the multipoles as $Q_m = |Q_m| e^{i\varphi_m}$, the reconstructed map up to order $m_{\rm max,rec}$ is given by:

$$\Sigma_{\rm reconst}(\theta) = \sum_{m=0}^{m_{\rm max,rec}} |Q_m| \, \cos\left(m\,\theta - \varphi_m\right). \tag{6.3}$$

The choice of the maximum multipolar order to use in the reconstruction, $m_{\rm max,rec}$, is particularly relevant for the reconstructed map, as it determines the smallest angular scale included in the map.

6.2 . T-REx filament finder

The Tree-based Ridge Extractor (T-REx , Bonnaire et al. 2020, 2022, to which we refer the readers for details) is an algorithm developed to identify one-dimensional filamentary structures in a discrete set of points embedded in higher dimension (both in 2- and 3-dimensions). This method has a natural application for detecting cosmic filaments from a galaxy distribution.

The T-REx algorithm assumes a tree topology connecting the centroids of a Gaussian Mixture Model (GMM), initially centred on the data points. The model is then regularised to approximate the probability density function from which the data points are drawn. In this way, the links between the centroids trace the ridges of the filaments in the data distribution.

The main steps of the algorithm are the following. The first step is to build the minimum spanning tree (MST) over the data. The MST is a graph with a tree topology that connects all the points in the dataset, while minimising the total length of its branches. Then, a pruning step is performed on the MST by removing iteratively all end-point branches. This step is crucial to denoise the tree, removing all the small-scale branches at the extremities of the graph while retaining the relevant structures formed by long chains of connected centroids. The amount of pruning can be adjusted, according to the specifics of the dataset, via the free parameter l, which sets the number of pruning iterations. Then, the resulting pruned tree is used as prior for the GMM. This means that the nodes of the tree become the centroids of the Gaussians of the GMM. Finally, the GMM is regularised over the full dataset using an adapted version of the expectation-maximisation (EM) algorithm, where the tree connections between the centroids are enforced by an additional constrain in the optimisation, which at the end offers a smoothed representation of the tree that approximates the underlying distribution of the data. This topological constrain is regulated by the free parameter λ, which sets the strength

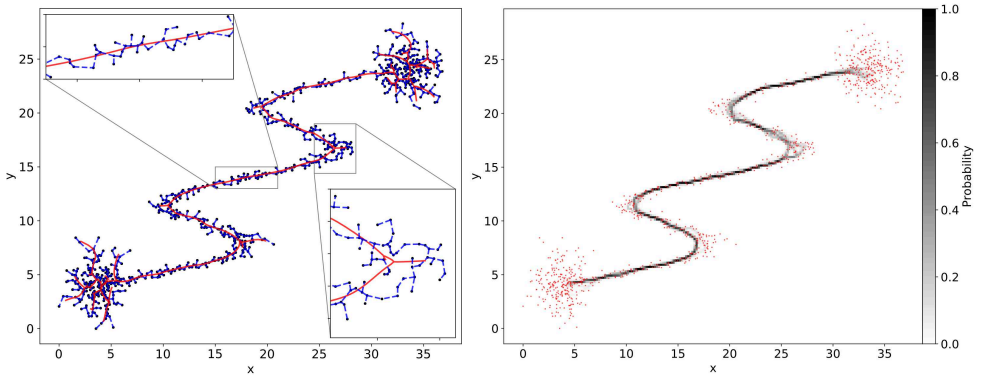

Figure 6.2: Visualisation of T-REx algorithm. Left: Result of T-REx algorithm over data points of a toy dataset. Black dots are data points, dashed blue line is the MST computed over the data, and red solid lines are edges of the regularised tree. Right: Probability map of the filamentary structures of the toy dataset. This was obtained by repeating 200 times the optimisation procedure on randomly selected subsets of the data. Figures taken from Bonnaire et al. (2020).

of the optimisation term that tries to reduce the length of the tree branches. Tuning the model parameters permits to adapt the method to vastly different data characteristics, as explained in Sect. 8.2.2, where I applied T-REx to an X-ray image for the first time. Nonetheless, the model proves to be pretty robust to small changes in the parameter values, which facilitates its application, avoiding costly fine-tuning. A more detailed discussion of the role of the parameters can be found in Bonnaire et al. (2020).

To obtain a robust representation of the filament detection together with a measure of its uncertainty, the procedure described above is repeated several times using a bootstrap approach. Each time, a subset of the data is chosen randomly, the MST is computed on the new subset and pruned. Then, the resulting GMM is optimised on the subset. This produces a set of regularised graphs, one for each random subset. From these, we can build a probability map of the filamentary structures, in which each pixel contains the fraction of times that position is crossed by a tree branch.

There are two tunable parameters in the bootstrap procedure: the number of iterations, B, and the size of the data subsample N_B. Out of the two parameters, N_B is the most relevant (as reported by Bonnaire et al. 2020). Generally, a low number of points with respect to the complete dataset produces a greater variability in the possible tree structures, while using subsets of similar size to the original data constrains the results more. On the other hand, the number of iterations B ceases to impact the results after a minimum level that depends on N_B. Therefore, it is sufficient to choose a high

enough number to ensure the stability of the probability map.

A visualisation of the various steps of the algorithm is shown in Fig. 6.2, where T-REx is applied on a toy dataset. In the left panel, we see the minimum spanning tree built on the dataset, and the regularised GMM tracing the ridge of the filamentary structure. In the right panel, the T-REx probability map is shown. Notice that in this map the filament is identified with high probability, while in the “cluster” regions at the edges no clear structure is identified.

7 - Gas distribution from clusters to filaments in IllustrisTNG

The content of this Chapter is mostly based on Gouin et al. (2022).

In this Chapter, I analysed the azimuthal distribution of gas inside and around galaxy clusters, to investigate the relation between the matter distribution and different structural and physical properties of clusters.

To do this, I used a sample of simulated clusters from the IllustrisTNG simulation suite (Nelson et al. 2019a). First, I probed the ability of the multipole decomposition to distinguish the features of different cluster structures, and their contributions to the total distribution; in particular, I tested the relation between different azimuthal orders and clusters' structural properties, such as the halo elliptical shape, the miscentering of the matter distribution, and the amount of substructures. In cluster outskirts, I also studied how the gas component traces cosmic filaments connected to clusters. Finally, I investigated whether the cluster dynamics and its accretion history impact the departure from spherical symmetry.

Based on the work of Martizzi et al. (2019), which showed that different cosmic web environments are populated by different gas phases, and Galárraga-Espinosa et al. (2021), I separate the gas in different phases to highlight their different distribution in cluster environments, thus distinguishing which phase traces which structure preferentially, in particular the cluster shape and the large-scale filamentary structures.

7.1 . Simulated cluster sample

For this work, I used a sample of galaxy clusters and their environments within $5 \times R_{200}$[1], extracted from the IllustrisTNG simulation.

IllustrisTNG (Nelson et al. 2019b; Pillepich et al. 2018; Springel et al. 2018b; Nelson et al. 2018a; Naiman et al. 2018; Marinacci et al. 2018) is a suite of large cosmological magneto-hydrodynamical simulations, run with the moving-mesh code AREPO Springel (2010); Weinberger et al. (2020). These simulations follow the coupled evolution of dark matter (DM), gas, stars, and black holes from redshift $z = 127$ to $z = 0$; using the cosmological parameters from the Planck 2015 results Planck Collaboration et al. (2016e). Of these simulations, we used the one with the largest volume, TNG300, at redshift $z = 0$. This

[1] R_Δ is the radius of a sphere centred on the halo within which the average density is Δ times the critical density $\rho_{\rm crit}(z)$; M_Δ is the mass contained inside R_Δ, $M_\Delta = (4\pi/3)\,\Delta\rho_{\rm crit}R_\Delta^3$.

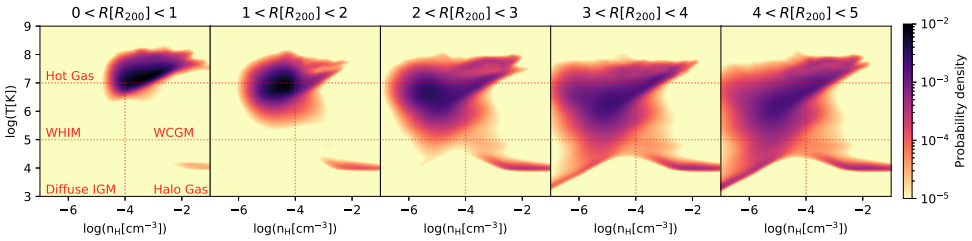

Figure 7.1: Stacked temperature-density diagrams for all gas cells around galaxy clusters and groups in IllustrisTNG, considering different radial apertures from cluster central regions $R[R_{200}] < 1$ up to $4 < R[R_{200}] < 5$.

simulation has a cubic box of length ~ 300 Mpc and mass resolution of about $7.6 \times 10^6\, M_\odot/h$ and $4.0 \times 10^7 M_\odot/h$ for gas and dark matter (DM), respectively. The large simulation box and high resolution make it ideal for accurately describing matter distribution around galaxy clusters up to their large-scale environments at $z = 0$.

The sample of clusters is selected from the catalogue of halos provided with the simulation, which are identified using a Friends-of-Friends (FoF) algorithm (Davis et al. 1985) on the DM particles. From this catalogue, we selected all halos with masses $M_{200} > 5 \times 10^{13} M_\odot/h$ that are more distant than $5\, R_{200}$ from the simulation box edges. This results in a total of 415 halos, that constitute the cluster sample. Starting from the cluster centres listed in the catalogue, we extracted from the simulation box all the DM and gas particles within a sphere of radius $5 \times R_{200}$, on which I performed the analysis.

7.1.1 . Gas phases

Gas in the Universe can exhibit remarkably different thermodynamical properties, depending on the region and on the processes it is subject to. For this reason, it is common to divide the gas into different phases, based on its distribution in the temperature-density diagram. This plane is usually divided into five regions, each identifying a phase, related to different environments and physical processes. These are: the halo gas, the diffuse intergalactic medium (Diffuse IGM), the warm-hot intergalactic medium (WHIM), the warm circumgalactic medium (WCGM), and the hot gas (the details on the definitions of each phase can be found in Martizzi et al. 2019).

Figure 7.1 shows the stacked gas distribution in temperature-density diagrams of the 415 galaxy cluster environments in five bins of cluster-centric distances from 0 to $5\ R_{200}$, illustrating how the gas is distributed among the different phases as a function of the distance from the cluster. Inside clusters ($R < 1 R_{200}$), gas is predominantly in the form of hot plasma, with high temperatures above $10^7 K$. As the distance from the cluster centres increases from 1 to $3\ R_{200}$, we can see that the majority of the gas is found at lower temperatures (in the range $10^5 < T[K] < 10^7$) and lower densities (in the range

$n_H < 10^4 \mathrm{cm}^{-3}$). At these distances, the hot gas that characterises the centre of clusters gives way to a more diffuse warm phase: the warm-hot intergalactic medium (WHIM). At distances greater than $3R_{200}$ from cluster centres, the gas is distributed across a range of phases, as indicated by temperature-density diagrams. These include cold diffuse (IGM), cold dense (halo gas), warm diffuse (WHIM), warm dense (WCGM), and hot gas. This temperature-density distribution is comparable to the overall distribution of cosmic gas in the universe at $z = 0$ (see Figure 2 of Galárraga-Espinosa et al. 2021, which considered all gas cells in the simulation box). Therefore, it suggests that beyond radial distances larger than $3\ R_{200}$ the influence of cluster environments is no longer significant. An actual quantification of the radial distribution of gas around clusters, based on mass fraction profile and different cluster mass bins, is discussed in Gouin et al. (2022).

Given the different distribution of gas phases depending on the distance from the clusters, in this work I used the hot gas as tracer of the gas distribution inside clusters, and the WHIM to probe the gas filamentary structures in cluster outskirts.

7.1.2 . Multipolar ratios as proxies for the azimuthal distribution

To probe the distribution of matter in and around galaxy clusters, I decided to use the multipolar ratios β_m described in Sect. 6.1. These ratios represent the relative importance of the different multipolar orders in a 2-dimensional distribution compared to the circular symmetry, and together they trace the total amount of asymmetry of the distribution. The use of a 2-dimensional proxy instead of a 3-dimensional one has the advantage of being more readily applicable to observations; on the other hand, its use on simulated data, for which the full 3D information is available, is useful to assess its sensitivity to projection effects, in particular when compared to the 3D-based structural properties described in Sect. 7.1.3.

As explained in the previous Section, I analysed the distribution of two gas phases, hot gas and WHIM, and compared them to the DM distribution, used as reference for the total matter distribution. To compute the β_m ratios, I first projected the gas and DM selected around each cluster (within $5R_{200}$ from the cluster centre) along the three axes (x, y and z) of the simulation box. In this way, I obtained 1245 projected cluster maps for each component (DM, hot gas, WHIM). From these maps, I computed the β_m using Eqs. 6.1 and 6.2, in different apertures (inside and outside R_{200}), for the multipolar orders from $m = 1$ to $m = 9$. This range of multipoles (and, consequently, angular scales) has been shown by Gouin et al. (2020) to capture well the relevant anisotropies in the matter field around galaxy clusters.

In Fig. 7.2, I show the mean β_m values for DM, in different radial apertures of size $0.5 \times R_{200}$, up to $4\ R_{200}$. In the top panel, we see that the average im-

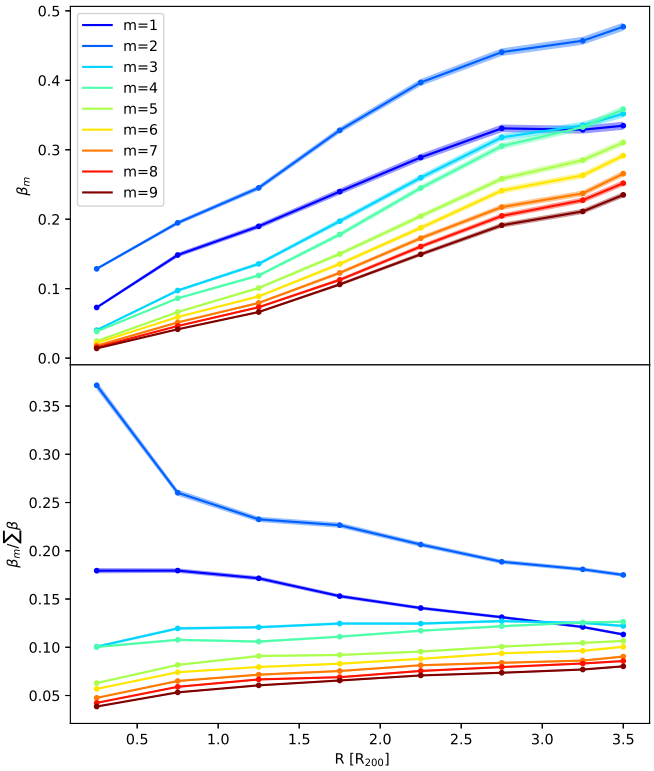

Figure 7.2: Top panel: Mean evolution of β_m parameter for $m = 1$ to $m = 9$ as a function of the cluster radial distance. Bottom panel: Mean evolution of β_m parameter normalised by the sum of all order contributions $\sum_{i=1}^{9} \beta_i$ from $m = 1$ to $m = 9$, as a function of the cluster radial distance.

portance of all the multipole orders compared to the circular symmetry grows with the distance from the cluster, which indicates that the matter distribution in cluster outskirts is in general less isotropic than in their cores (in agreement with previous observations and simulations, see e.g. Eckert et al. 2012; Gouin et al. 2020). In the bottom panel, the β_m normalised by the sum of all orders shows the relative contribution of each order to the total asphericity of the distribution. We see that the dominant order at all radii is the quadrupole (consistent with Vallés-Pérez et al. 2020), although its relative contribution to the decomposition gets lower with increasing radii. After the quadrupole, the most significant orders are the $m = 1, 3, 4$, whose contributions are always higher than 10%. Finally, the higher orders, $m > 4$, contribute less than 10% at all radii. Therefore, we see that structures with large angular scales, identified by the low multipolar orders, dominate the non-spherical part of the matter distribution. I thus focused primarily on the first 4 multipoles in the rest of the analysis.

7.1.3 . Physical and structural properties

For each cluster in the sample, several physical and structural properties were computed. More details of their computation can be found in Gouin et al. (2021). Here, I summarise the definitions and computation procedures of the different parameters.

Cluster structural properties

In order to probe the ability of the multipolar ratios to trace the anisotropic matter distribution in galaxy clusters, I compared them to three structural parameters: the centre offset $R_{\rm off}$, the subhalo fraction $f_{\rm sub}$ and the halo ellipticity ϵ.

The centre offset measures the distance between the centre of mass of the cluster, computed using all the particles in the cluster FoF halo, and the position of the density peak, identified as the particle with the lowest potential energy, normalised by the cluster radius: $R_{\rm off} = |r_{\rm cm} - r_{\rm c}|/R_{\rm vir}$.

The subhalo mass fraction quantifies the amount of mass of the cluster that is contained in clumpy subgroups. It is defined as the ratio between the mass contained in the subhalos and the total mass of the cluster, $f_{\rm sub} = \sum M_{\rm sub}/M_{\rm tot}$. The subhalos are identified using the Subfind algorithm Springel et al. (2001) and listed, together with their masses, in a catalogue provided with the simulation.

Finally, the ellipticity of the DM distribution was used as a proxy for the cluster shape. The ellipticity was computed for both the 3D and the projected 2D distributions. Following Suto et al. (2016), the DM distribution was fitted with an ellipsoid, optimised using the eigenvalues of the mass tensor and fixing the mass enclosed by the ellipsoid to be equal to M_{200}. From the axes of the ellipsoid, the ellipticity is computed as:

$$\begin{aligned} \epsilon_{2D} &= \frac{c-a}{2(a+c)}\,, \\ \epsilon_{3D} &= \frac{c-a}{2(a+b+c)}, \end{aligned} \tag{7.1}$$

where a and c are the major and minor axes, respectively, and in the 3D case b is the intermediate axis.

Mass assembly history

To explore the mass assembly history of clusters, the time evolution of cluster mass $M_{200}(z)$ has been computed utilizing the merger tree of subhalos generated with the SubLink algorithm (Rodriguez-Gomez et al. 2015). Two distinct metrics have been estimated to characterise the mass assembly history of clusters: the formation redshift $z_{\rm form}$ and the mass accretion rate Γ.

These parameters offer complementary insights into cluster mass assembly history, by capturing both the accretion phase and the formation period of an object based on its mass growth.

The accretion rate is defined as the ratio of the halo mass at $z = 0$ to the mass of its main progenitor at a specific redshift z (as defined by Diemer et al. 2013):

$$\Gamma \equiv \frac{\Delta \log(M_{200\mathrm{m}})}{\Delta \log(a)}. \tag{7.2}$$

This parameter quantifies the accretion phase of a halo over two time intervals, specifically chosen to be $z = 0$ and $z = 0.5$ (which aligns with the expected relaxation timescales of halos, according to Power et al. 2012; Diemer & Kravtsov 2014; More et al. 2015).

The formation redshift, on the other hand, is the redshift at which the mass of the main progenitor halo is exactly half of its present-day mass:

$$\frac{M_{200}(z_{\mathrm{form}})}{M_{200}(z = 0)} = \frac{1}{2}\,, \tag{7.3}$$

as defined by Cole & Lacey (1996).

Dynamical state

Different parameters have been shown to be, at least partially, indicators of the dynamical state of galaxy clusters. In this study, the clusters' dynamical state is represented by the relaxedness parameter χ_{DS}, as defined in Haggar et al. (2020). It is computed by combining three parameters: two aforementioned structural properties, R_{off} and f_{sub}, and the virial ratio, $\eta = 2T/|W|$, which quantifies the departure from virial equilibrium (T and W are the kinetic and the gravitational potential energy, respectively). The relaxedness is thus defined as:

$$\chi_{\mathrm{DS}} = \sqrt{\frac{3}{\left(\frac{R_{\mathrm{off}}}{0.07}\right)^2 + \left(\frac{f_{\mathrm{sub}}}{0.1}\right)^2 + \left(\frac{\eta - 1}{0.15}\right)^2}}. \tag{7.4}$$

Clusters with $\chi_{\mathrm{DS}} \geq 1$ are considered dynamically relaxed, while clusters where $\chi_{\mathrm{DS}} < 1$ are considered disturbed (Kuchner et al. 2020).

Connectivity

A parameter of fundamental interest in the study of the outskirts of a galaxy cluster is the number of cosmic filaments connected to it. This is called connectivity, and it is defined as the number of filaments intersecting a sphere of radius $1.5 \times R_{200}$ around the cluster (Darragh Ford et al. 2019; Sarron et al. 2019; Kraljic et al. 2020).

In this study, the filamentary structures of the cosmic web are extracted from the galaxy distribution in the whole simulation box, using the T-REx algorithm (presented in Sect. 6.2). The galaxies were selected from the subhalo catalogue, applying a lower limit on the stellar mass of $M_* \geq 10^9\, M_\odot$ (following Galárraga-Espinosa et al. 2020). The T-REx algorithm was then run on these galaxies, setting the values of the free parameters to $\lambda = 1$ and $l = 5$ (close to the ones suggested in Bonnaire et al. 2020, the first controls the smoothness of the graph, the second the denoising procedure). Finally, each cluster was associated with the closest node of the T-REx graph, and the connectivity was computed by counting the filaments at a distance of $1.5\, R_{200}$ from the node.

7.2 . Azimuthal gas distribution in relation to cluster properties

In this Section, I present the results of the comparison between the azimuthal behaviour of the gas and DM distributions in and around galaxy clusters, traced by the multipolar ratios β_m, and the structural and physical cluster properties defined in Sect. 7.1.3. In this way, I explore the ability of the gas distribution to, on the one hand, trace the shape and structure of clusters and, on the other hand, encode information about their dynamical state and accretion history.

Most of the cluster properties considered are difficult or even impossible to measure observationally, thus finding correlations between these and structural properties of the observable gas distribution can inform us of the hidden features of observed clusters.

In this section, a general colour and style code is used in the plots: (i) Hot gas, WHIM, and dark matter are respectively in red, orange and blue/black colors, (ii) the mean profiles are shown by solid lines, and the errorbars are the errors on the mean computed by bootstrap re-sampling, (iii) the number of objects used to compute the average in each bin (of x-axis) is written on the top of the figures in gray, (iv) the Spearman's Rank correlation coefficient $\rho_{\rm sp}(X, \beta_m)$ is written on the figure, with X the cluster property and β_m the multipolar ratio at the multipole order m as defined below. The *p*-value of the correlation coefficient remains lower than 10^{-3} for each plots.

7.2.1 . Azimuthal symmetries as proxies of cluster structural properties

I focus here on the azimuthal distribution of hot gas and DM inside clusters, and compare them to the structural properties defined in Sect. 7.1.3.

In particular, I stress the fact that different multipolar ratios, β_m, trace the contribution of different azimuthal structures to the total distribution of the matter component considered, as exemplified in Fig. 6.1. Therefore, I considered different multipolar orders in relation to different structural properties.

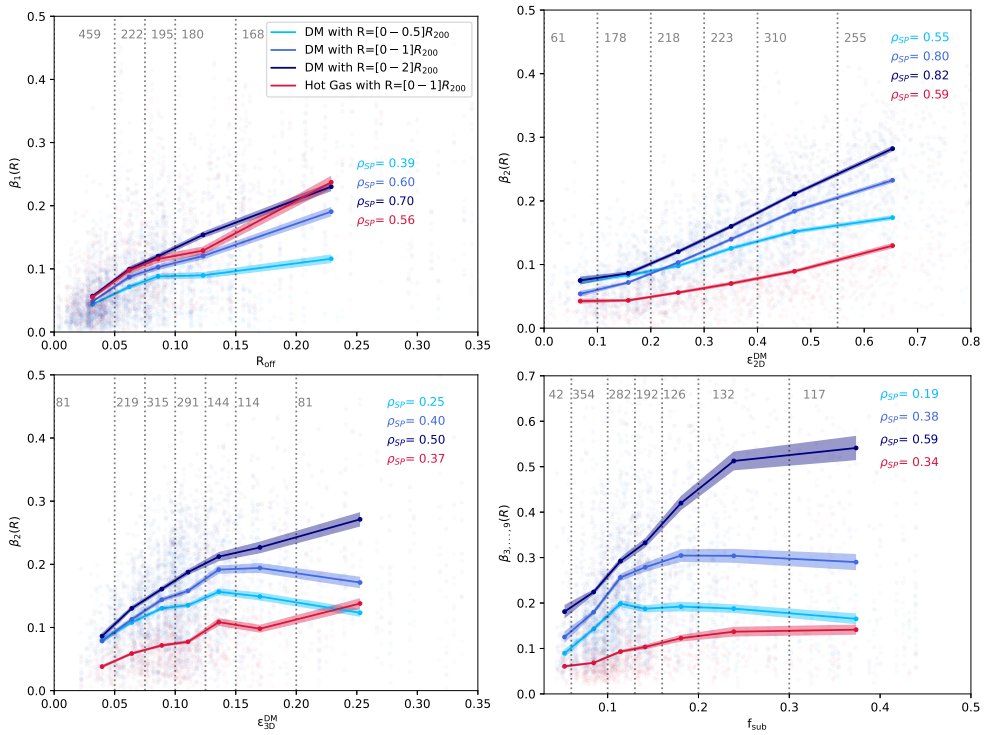

Figure 7.3: Distribution of different multipolar ratios β_m, for different components and different apertures, as a function of the different structural properties of clusters. Top left panel: β_1 (dipole contribution) as a function of the center of mass offset, $R_{\rm off}$. Top right panel: β_2 (quadrupole contribution) as a function of the 2D ellipticity of DM, ϵ_{2D}. Bottom left panel: β_2 (quadrupole contribution) as a function of the 3D ellipticity of DM, ϵ_{3D}. Bottom right panel: sum of high order multipoles β_m contribution (summing contributions from $m = 3, 4, 5, 6, 7, 8, 9$) as a function of the mass fraction of substructures.
The β_m values for the DM distribution with apertures $R < 0.5 \times R_{200}$,$R < 1 \times R_{200}$ and $R < 2 \times R_{200}$, are plotted in light, medium, and dark blue, respectively. The β_m values for hot gas distribution within R_{200} is plotted in red. The mean profiles of β are shown by solid lines, and the errorbars are the errors on the mean computing by bootstrap re-sampling. On each panel, the number of objects used to compute the average in each bin of x-axis (shown in gray dotted lines) is written on the top of the figures in gray.

These are shown in Fig. 7.3. The multipolar ratios have been computed for both dark matter (DM) at various radial apertures ($R < [0.5, 1, 2]R_{200}$) and hot gas within a radial aperture of $R < R_{200}$, depicted in blue and red, respectively.

The top left panel illustrates the potential correlation between the center offset and the dipolar symmetry traced by β_1. The focus is on $m = 1$ because, as mentioned in Sect. 6.1, the dipolar symmetry is expected to reflect the mis-centering of the mass distribution, while the offset center parameter directly quantifies the distance between the peak of the mass distribution and the center of mass. The findings indicate that the dipolar symmetries of DM and hot gas within R_{200} effectively trace the center offset of clusters, with a correlation coefficient between β_1 and $R_{\rm off}$ greater than $\rho_{\rm SP} > 0.5$. We also see that the correlation of the DM dipole asymmetry with $R_{\rm off}$ grows when considering larger apertures, together with the slope of the mean relation.

The top right panel of Fig. 7.3 shows the distribution of the clusters' quadrupole order, β_2, as a function of the two-dimensional ellipticity of projected DM halos. We see that, for DM, the power of the quadrupolar order is strongly correlated with the projected elliptical shape of the halo (consistently with, for example, Clampitt & Jain 2016; Shin et al. 2018), and that increasing the radial aperture beyond R_{200} does not significantly enhance the correlation coefficient, suggesting that most halo shape information is contained within R_{200}. Indeed, enlarging the aperture from 1 to 2 R_{200} increases the overall mean value of β_2, but does not alter significantly the slope nor the correlation. Regarding the hot gas, the quadrupolar order still correlates strongly with the DM ellipticity, although the mean values of β_2 are generally lower than those computed on the DM distribution. This indicates that hot gas inside clusters tends to be more spherically distributed than DM, due to dissipative baryonic processes (as found in Velliscig et al. 2015; Okabe et al. 2018).

The comparison of the β_2 parameter to the 3-dimensional ellipticity of halos reveals the impact of the projection effects on the 2D shape of clusters. In fact, if the main axis of elongation is roughly aligned with the line of sight, even clusters with large 3D ellipticities may appear only mildly elliptical in 2D. This effect of projection has the result of reducing the correlation between the quadrupole β_2 and ϵ_{3D}, as shown in the bottom left panel of Fig. 7.3, especially for the highest ellipticities. This effect appears less strong when considering larger radial apertures; indeed, we see that for the largest aperture, $R < 2R_{200}$, the quadrupolar ratio and 3D ellipticity are quite strongly correlated, with a Spearman coefficient $\rho_{SP} \sim 0.5$.

The bottom right panel of Fig. 7.3 explores the amount of mass within sub-halos via the fraction of substructures. To quantify this structural property, it is best to consider higher harmonic orders, which trace signatures at smaller angular scales. The multipolar ratios β_m from $m = 3$ to 9 are summed

to obtain the overall level of asymmetry for small angular scales, and this is compared to the fraction of substructures. The sum of high-order multipolar ratios and $f_{\rm sub}$ exhibit a correlation for both hot gas and DM, with correlation coefficients around $\rho_{\rm SP} \sim 0.35$, although we see that the mean relation tends to flatten for high $f_{\rm sub}$. This allows for distinguishing between low ($f_{sub} < 0.1$) and high substructure fractions ($f_{sub} > 0.1$), which is an important criterion for separating dynamically relaxed and non-relaxed clusters, as proposed by Cui et al. (2017). We can see again that the mean values of β_m from the hot gas are lower than those from DM, a sign that the latter tends to clump more efficiently in substructures due to its collisionless nature. Finally, we notice that, similarly to the 3D ellipticity, considering large radial apertures ($R < 2R_{200}$) significantly increases the correlation between the fraction of substructures and the level of azimuthal symmetries, reaching $\rho_{SP} \sim 0.6$.

Summarising the results so far, I have shown that the multipolar ratios at different orders are a measure of the relevance of different angular features, and as such are powerful tracers of distinct structural properties of galaxy clusters. In particular, the dipole order, β_1, reflects the miscentering of the cluster mass distribution. Conversely, the quadrupole, β_2, correlates with the elliptical shape of clusters, as also discussed in Gouin et al. (2017, 2020). Higher multipole orders trace angular scales that decrease with increasing m, thus allowing the characterisation of small-scale structures. Therefore, the combined contribution of multipoles from $m = 3$ to 9 is a probe of the substructure fraction. Focusing on the DM results, I showed that the two-dimensional azimuthal matter distribution is strongly correlated with its three-dimensional structural properties, and the correlation increases considering larger radii, beyond R_{200}. Furthermore, the azimuthal distribution of hot plasma appears to follow closely the azimuthal distribution of dark matter, as evidenced by the significant correlation with halo properties. The hot gas plasma distribution remains smoother and more circular than the DM distribution, with lower values of the multipolar ratio β_m for almost all orders.

7.2.2 . Azimuthal distribution related to cluster physical properties

In order to evaluate the correlations between the matter distribution in cluster regions and the cluster physical properties, I decided to focus on a single estimator that traces the overall departure from circular symmetry, in order to combine all the possible sources of asymmetry and have a more comprehensive estimator that might be better correlated with general cluster physical properties. This departure is estimated by the β parameter, defined

as the sum of all the multipolar ratios β_m up to order N,

$$\beta = \sum_{m=1}^{N} \beta_m \,. \tag{7.5}$$

This parameter estimates the total amount of azimuthal symmetries compared to the circular one. The maximum order considered in the sum was $N = 4$. This choice was made after verifying that adding higher orders does not alter the results.

I focus here on two radial apertures: inside clusters ($R < R_{200}$) and in their outskirts (between 1 and $2\,R_{200}$), where the correlations between the cluster asymmetries and their properties are stronger. Indeed, when investigating more distant apertures, only weak correlations were found between 2 and $3\,R_{200}$, and no clear correlations beyond $3\,R_{200}$.

Based on the discussion on the gas phases in Sect. 7.1.1, I decided to use just the hot medium as a probe of the gas asymmetry in the inner aperture, and the WHIM phase in the outer aperture.

Mass and connectivity

I start investigating the relation between the cluster mass and its departure from spherical symmetry. This is shown in Fig. 7.4, in the top panel for the inner cluster aperture, and in the bottom panel for the cluster outskirts.

Looking at the dark matter distribution inside halos ($R < R_{200}$), we see that the β parameter slowly increases with the cluster mass, on average. As shown in Fig. 7.2, the main contribution to the asphericity inside clusters comes from the quadrupole, which is related to the cluster ellipticity (as shown in the Sect. 7.2.1). Therefore, we would expect a larger dependence of β on the cluster mass, as found for the ellipticity in previous studies (e.g. Despali et al. 2014). Instead, we see that the correlation between β and mass is low, with $\rho_{\rm SP} \sim 0.1$.

Focusing on the hot gas, we see that, contrary to the DM, the β parameter decreases at low masses (below $\sim 10^{14} M_\odot/h$), and then increases again for higher masses (with a trend similar to the DM one). On average, hot gas in galaxy groups ($M_{200} < 10^{14} M_\odot/h$) appears more asymmetric than in massive clusters. An explanation of this result can be found in the radial distribution of the hot gas, as shown in detail in Gouin et al. (2022). In fact, while on average the hot gas phase is dominant within $\sim R_{200}$, this is not completely true for low-mass halos, where the hot gas is found to be concentrated in the inner cores (up to $\sim 0.6 R_{200}$). Between ~ 0.6 and $\sim 1 R_{200}$, the main fraction of gas is in a colder, dense phase called warm circumgalactic medium (WCGM). Indeed, if we consider the distribution of the whole gas, instead of just the hot phase, we see a positive correlation similar to that of DM. This means that only

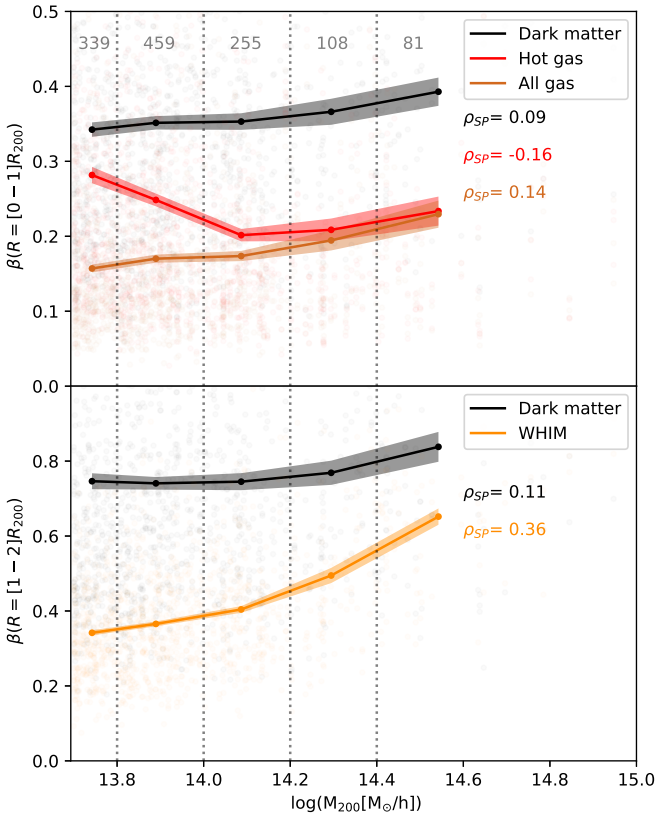

Figure 7.4: Distribution of the azimuthal symmetric excess β (as defined in Eq. 7.5) as a function of the halo mass, inside clusters ($R < R_{200}$) in the top panel, and at cluster peripheries in ($1 < R[R_{200}] < 2$) in the bottom panel. The mean profiles of β and their errors are shown in solid lines. The color of points and lines represent different matter component: dark matter (black), hot gas (red), WHIM (orange), and all gas (light brown). On each panel, the number of objects used to compute the average in each bin of x-axis (shown in gray dotted lines) is written on the top of the figures in gray.

the hot gas is strongly anisotropic inside low-mass halos. One interpretation of this result is that hot gas inside massive clusters is mostly gravitationally heated, whereas in smaller halos the hot gas distribution might be governed by anisotropic accretion processes, thus becoming more asymmetric.

Moving to the bottom panel of Fig. 7.4, we see that, in cluster outskirts (between 1 and $2R_{200}$), the anisotropy of both DM and WHIM distributions is overall higher than inside clusters, and increases with cluster mass. The anisotropy in these regions is expected to be due to the presence of filamentary structures connected to the clusters, and indeed the number of filaments has been found to be positively correlated with the cluster mass (Aragón-Calvo et al. 2010b; Codis et al. 2018; Kraljic et al. 2020; Gouin et al. 2021). Focusing on the WHIM gas, we see a stronger dependence on mass compared to the

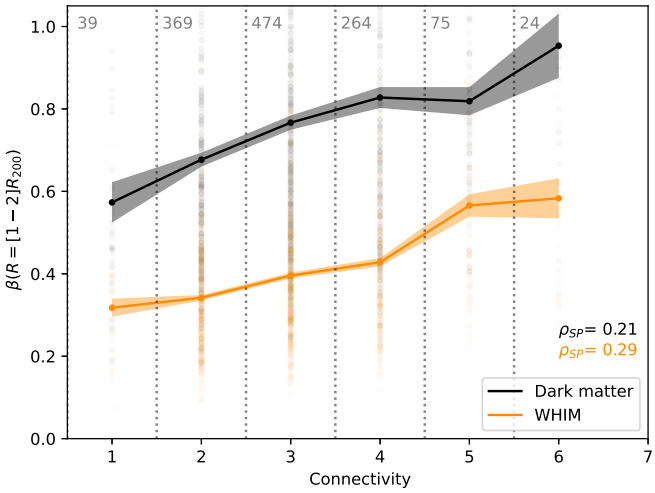

Figure 7.5: Distribution of the azimuthal symmetric excess β (as defined in Eq. 7.5) computed at cluster peripheries in ($1 < R[R_{200}] < 2$) as a function of the halo connectivity, for DM (black) and WHIM (orange). The mean profiles of β and their errors are shown in solid lines. The number of objects used to compute the average in each bin of x-axis (shown in gray dotted lines) is written on the top of the figures in gray.

DM: in general, the WHIM is significantly more asymmetric around massive clusters than low-mass halos. This is consistent with the picture of filaments acting as funnels for the infalling gas, which thus preferentially enters clusters anisotropically through filaments (Rost et al. 2021). Therefore, the more connected massive clusters have less isotropic gas distributions. Furthermore, gas around low-mass groups tends to accumulate more at clusters' peripheries, leading to a more spherical distribution.

To confirm the link between the anisotropies in cluster outskirts and the amount of filamentary structures connected to them, I probed the correlation between β and the connectivity, shown in Fig. 7.5. Here, we see a correlation between the number of filaments and the departure from spherical symmetry of both DM and WHIM gas, with roughly similar slopes and correlation coefficients for the two components. It is important to note that the connectivity shown here accounts for all filaments in 3D, so it is expected that the correlation with the 2D asymmetry is impacted by projection effects, as some filaments might extend along the line of sight. We can nonetheless conclude that the azimuthal distribution of WHIM tends to follow DM in tracing cosmic filamentary structures connected to clusters at their peripheries (in agreement with Galárraga-Espinosa et al. 2021, who found the WHIM to be the dominant phase in cosmic filaments).

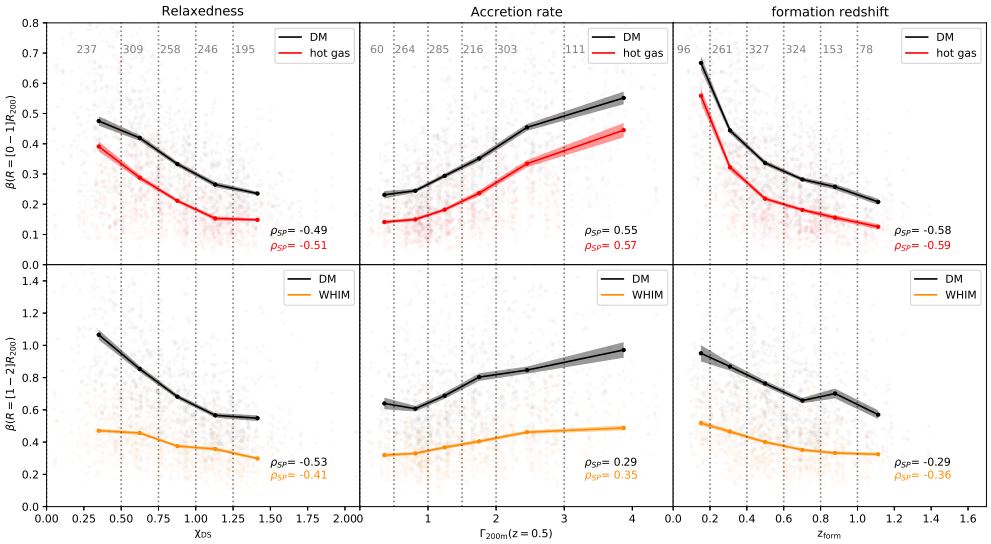

Figure 7.6: Distribution of the azimuthal symmetric excess β (as defined in Eq. 7.5) computed inside clusters ($R < R_{200}$) in top panels, and at cluster peripheries in ($1 < R[R_{200}] < 2$) in bottom panels, as a function of different halo properties: level of relaxation on the left, mass accretion rate in the middle, and formation redshift in the right. The mean profiles of β and their errors are shown in solid lines. The number of objects used to compute the average in each bin of x-axis (shown in gray dotted lines) is written on the top of the figures in gray.

Accretion and dynamical state

So far, I have shown how hot gas is a good tracer of the structural properties of clusters, and that the WHIM azimuthal distribution is correlated with filamentary structures connected to cluster outskirts. Here, I investigate whether these two components, inside and around clusters, also carry information about the cluster dynamical state and its mass assembly history. These properties are represented by three parameters: the relaxedness $\chi_{\rm DS}$, the mass accretion rate Γ, and the formation redshift $z_{\rm form}$, defined in Sect. 7.1.3.

The relations between these three parameters and the β asymmetry parameter are shown in Fig. 7.6, where the value of β is computed on hot gas and DM inside R_{200} (top row), and on WHIM and DM in cluster outskirts ($1 < R[R_{200}] < 2$, bottom row).

The departure from circular symmetry as a function of the relaxation level of clusters is shown in the left column. We see that, the more circular the matter distribution (low values of β), the more dynamically relaxed the cluster ($\chi_{DS} > 1$). Inside clusters, halo relaxedness and asymmetry are strongly correlated for both DM and hot gas, with $\rho_{SP} \sim 0.5$, showing that the azimuthal distribution of hot plasma is a powerful tracer of cluster dynamics. At cluster

peripheries, the azimuthal distribution of WHIM and DM also traces well the cluster relaxation level. This can be explained by the fact that the cluster dynamical state must be affected by the number of cosmic filaments connected to clusters, as shown in Gouin et al. (2021).

The influence of the cluster mass-assembly history on the azimuthal matter distribution is investigated in the middle and right panels, for the mass accretion rate and the formation redshift of clusters, respectively. Both accretion history proxies are correlated with gas and DM azimuthal asymmetry, in particular inside cluster halos, with $\rho_{\rm SP} \sim 0.6$. We find that clusters with faster accretion of matter and of recent formation have larger departures from spherical symmetry in their hot gas distribution. This is coherent with the result of Chen et al. (2020) which showed the ellipticity of cluster gas to be an imprint of the mass assembly history of clusters. In cluster outskirts, we see similar trends, although the correlations are less strong, suggesting that the mass distribution in the inner regions is more sensitive to the cluster past accretion history. Nonetheless, we can deduce that cosmic filaments connected to clusters, traced by the WHIM, have an impact on the cluster assembly history (consistently with the results of Gouin et al. 2021).

7.3 . Discussion and conclusions

In this Chapter, I have shown the application of the aperture multipole decomposition technique to the projected spatial distribution of different gas phases and DM in simulated clusters. For this, I used 3×415 projections of the matter distribution of galaxy cluster environments (out to $5R_{200}$), extracted from the IllustrisTNG simulation. The gas phases were defined by applying cuts in the temperature-density plane, following Martizzi et al. (2019).

The multipolar decomposition was previously used for the study of dark matter (Gouin et al. 2017) and galaxy (Gouin et al. 2020) azimuthal distributions in cluster environments, revealing its ability to probe statistically the shape of clusters and their connections to the cosmic filaments. Here, I focused on the gas distribution, and in particular on its relation to different structural and physical properties of clusters.

The results of this analysis show that:

— The azimuthal features of gas and DM distributions are strongly correlated with the structural properties of the cluster halos. In particular, the strength of the dipolar order reflects the centre offset, the quadrupole describes the halo ellipticity, and larger harmonic decomposition orders trace the amount of halo substructures. The azimuthal distribution of hot plasma appears to follow that of dark matter, reflecting the structural characteristics of the cluster halo. However, the hot gas tends to exhibit a smoother and more circular morphology than the

dark matter distribution (as expected from the ellipsoidal shape of DM and hot gas found in Okabe et al. 2018).

— Studying the relation with the cluster mass, the analysis revealed that the hot gas within low-mass halos exhibits greater asymmetry than that within massive clusters. This is attributed to the fact that the hot gas does not represent the overall gas distribution inside small groups, but is rather concentrated in their cores and must be subject to distortion by anisotropic accretion processes. In contrast, the gas component of massive clusters inside R_{200} is almost exclusively composed of gravitationally heated gas (93% of the total), which explains its more isotropic shape.

— Focusing on cluster outskirts, I showed that the asymmetries of both WHIM and DM increase with cluster mass, in line with the expected increase in the harmonic-space signature of cosmic filaments connected to clusters reported by Gouin et al. (2020). Notice that for low-mass groups, the asymmetries in the WHIM distribution decrease faster than the DM ones. This can be explained with a dynamical analysis of the WHIM gas (Rost et al. 2021; Gouin et al. 2022), which shows that the gas around low-mass halos is both slowly infalling in filaments and outflowing from groups, thus smoothing the filament signatures.

— Probing more directly the relation of asymmetry in cluster outskirts with the connectivity, I found that the asymmetric signatures of WHIM and DM distributions increase with the number of connected filaments, showing that the presence of filaments imprints measurable signatures in the matter azimuthal distribution. Therefore, the WHIM gas phase, as it follows the DM distribution, appears to trace well connections to the cosmic filaments.

— I found that the gas azimuthal distribution is affected by the past assembly history of clusters and that it is a good tracer of its current dynamical state, as good as the reference DM distribution. In detail, it is apparent that dynamically unrelaxed, fast-accreting, and late-formed galaxy clusters exhibit strong departures from spherical symmetry in their hot gas distributions (in agreement with, e.g. Vazza et al. 2011). Furthermore, I showed that clusters' dynamical and accretion-history-related properties are correlated to the gas distribution also in cluster outskirts.

In summary, by using the aperture multipole decomposition technique, I statistically probed the azimuthal distribution of gas in and around clusters, up to the connection to the cosmic filaments in their outskirts. I found that cluster environments strongly affect the gas distribution up to $2R_{200}$ from the

cluster centres, while beyond $\sim 4R_{200}$, the gas distribution is no longer impacted by clusters.

The multipole decomposition has been therefore proven to be a reliable and effective technique to analyse structures traced by gas in and around galaxy clusters. The application of this technique to real observations of clusters is thus the next logical step. In the next Chapter 8, I show how the multipole analysis can be used to describe the gas distribution in the outskirts of a galaxy cluster through its X-ray emission, in the case study of the Abell 2744 cluster.

8 - Tracing gaseous filaments connected to galaxy clusters: the case study of Abell 2744

The content of this Chapter is based on Gallo et al. (2024a).

As mentioned in Chapter 6, the outskirts of galaxy clusters are the regions where the infalling matter starts interacting with the cluster, and thus where the complex physical processes associated with matter accretion, such as mergers, shocks, and turbulence take place (see e.g., Reiprich et al. 2013; Walker et al. 2019; Walker & Lau 2022). In general, matter does not accrete isotropically, but it is funnelled by cosmic filaments, that act as "cosmic highways" towards clusters (Rost et al. 2021; Gouin et al. 2021). Therefore, the matter distribution in cluster outskirts is complex, and characterised by the presence of filamentary structures connected to the clusters. For these reasons, studying these regions is especially interesting: understanding the cluster mass distribution out to large radii is particularly important to estimate more precisely their total mass, and by improving the understanding of the accretion processes one can study their impact on the hydrostatic equilibrium assumption.

Therefore, understanding the distribution and properties of matter falling in clusters through filaments has been the goal of many studies, using hydrodynamical simulations (e.g. Rost et al. 2021; Tuominen et al. 2021; Gouin et al. 2021, 2022, 2023; Galárraga-Espinosa et al. 2021, 2023), stacked observations (e.g. de Graaff et al. 2019; Tanimura et al. 2019, 2020, 2022), between cluster pairs or around single clusters (e.g. Werner et al. 2008; Planck Collaboration et al. 2013b; Eckert et al. 2015; Bulbul et al. 2016; Bonjean et al. 2018; Veronica et al. 2024).

Nevertheless, the detection of these large-scale filaments remains a challenging task, particularly for the gas component, due to their lower density and temperature compared to clusters. This results in fainter signal in both X-rays and the Sunyaev-Zel'dovich (SZ) effect. Therefore, high sensitivity and low noise are required to detect them.

A particular case in which extended emission of X-rays from filaments was observed is in the outskirts of the galaxy cluster Abell 2744 (A2744). Eckert et al. (2015) identified extended structures connected to A2744 from the adaptively-smoothed X-ray surface brightness map, and confirmed the detection using a sample of spectroscopic galaxies, as well as weak lensing observations.

Abell 2744 ($z = 0.306$ Owers et al. 2011) is a particularly massive cluster ($M_{200c} \sim 2 \times 10^{15} M_{\odot}$ Medezinski et al. 2016) and exhibits a highly dis-

turbed dynamical state, with many different massive interacting substructures (Kempner & David 2004; Merten et al. 2011; Owers et al. 2011; Jauzac et al. 2016; Medezinski et al. 2016). It has been extensively observed in X-rays, optical, and radio wavelengths (Eckert et al. 2015; Owers et al. 2011; Merten et al. 2011; Ibaraki et al. 2014; Jauzac et al. 2015, 2016; Boschin et al. 2006; Kempner & David 2004; Rajpurohit et al. 2021; Harvey & Massey 2024; Eckert et al. 2016; Govoni et al. 2001; Hattori et al. 2017; Medezinski et al. 2016; Braglia et al. 2007). In the X-rays, in addition to two main peaks that constitute the main cluster emission, up to four additional cores have been detected (Jauzac et al. 2016); moreover, the presence of density and temperature discontinuities suggests the presence of shocks (Eckert et al. 2016; Hattori et al. 2017). Such a massive and unrelaxed cluster is thus the perfect candidate for being connected to many cosmic filaments, as predicted from simulations (e.g. Darragh Ford et al. 2019; Gouin et al. 2021).

In this Chapter, I present an analysis of the matter distribution in the outskirts of A2744, using as probes the X-ray emission observed by *XMM-Newton* (the same data as Eckert et al. 2015), and a catalogue of spectroscopic galaxies from Owers et al. (2011). These data have been analysed using the two statistical methods presented in Chapter 6, namely the aperture multipole decomposition and the T-REx filament finder, to obtain a blind detection of cosmic filaments connected to a galaxy cluster.

Following Eckert et al. (2015), I considered as cluster radius $R_{\rm vir} = 2.1\,h_{70}^{-1}{\rm Mpc}$, where $h_{70} = H_0/(70\,{\rm km\,s^{-1}\,Mpc^{-1}})$ and $H_0 = 67.4\,{\rm km\,s^{-1}\,Mpc^{-1}}$ (from Planck Collaboration et al. 2020a).

8.1 . Abell 2744: observational data

8.1.1 . X-ray data

The cluster A2744 was observed by *XMM-Newton* X-ray Observatory for 110 ks in 2014 (see Eckert et al. 2015; Jauzac et al. 2016, for a detailed description). The data were re-analysed using the X-COP analysis pipeline, as described in Ghirardini et al. (2019). The raw images were extracted in two energy bands, namely $[0.4-1.2]\,{\rm keV}$ and $[2-7]\,{\rm keV}$, along with the respective background and exposure maps. From these, surface brightness maps were computed (the surface brightness image for the soft band is shown in Fig. 8.1, left panel). Point sources were then identified in each band using the XMMSAS task `ewavelet` and the results cross-matched between the two energy bands. This process yielded a preliminary list of high-reliability point sources. This first list was found to be too limited for the purposes of this study, as it does not include several point-like sources visible in the soft-band image. This is likely due to the differing depths in the soft and hard-band images. Therefore, I decided to add to the list all the point-like sources present in the soft-band

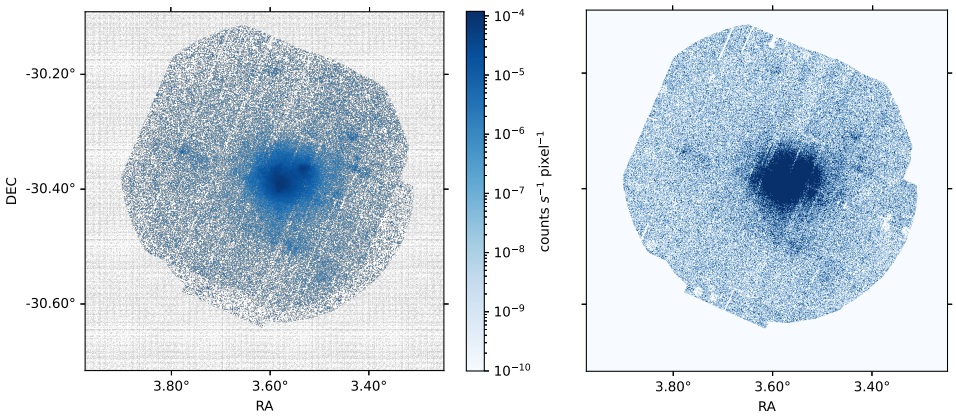

Figure 8.1: X-ray maps of the area around A2744. Left panel: point-source-filtered surface brightness map in the soft energy band, $[0.4 - 1.2]\,\mathrm{keV}$. Right panel: "hit map", obtained from the soft-band surface brightness image setting to 1 all pixels with values > 0.

image. The final source list was employed as a conservative mask to remove any potential signal that was not associated with diffuse emission from the cluster and its outskirts. In practice, the areas covered by the identified point sources were masked and the corresponding pixels were refilled using Poisson realizations of the neighbouring surface brightness, as implemented in the `dmfilth` tool of the `pyproffit` package(Eckert et al. 2020).

For the purpose of this study, I decided to focus only on the structural properties of the X-ray emission. Therefore, I created a binary map obtained setting to 1 every pixel in the (point-source filtered, soft-band) surface brightness map whose value is above 0 (i.e., lit pixels). I call this a "hit map", and it is shown in Fig. 8.1, right panel. Disregarding all information about the signal amplitude, the extended structures are highlighted as spatially concentrated collections of lit pixels, which boosts the low signal structures with respect to higher-signal ones.

In the remainder of this Chapter, this hit map will serve as the data product upon which our analysis will be based. It will be referred to as the X-ray map or X-ray data.

8.1.2 . Spectroscopic galaxies

In addition to the X-ray data, I considered the spectroscopic galaxy catalogue compiled by Owers et al. (2011). It gathers observation of A2744 with the AAOmega multi-object spectrograph on the Anglo-Australian Telescope, and catalogues from the literature (Boschin et al. 2006; Braglia et al. 2009; Couch & Sharples 1987; Couch et al. 1998). This compilation includes the redshifts of 1250 galaxies within $15\,\mathrm{arcmin}$ (~ 4 Mpc) from the cluster centre.

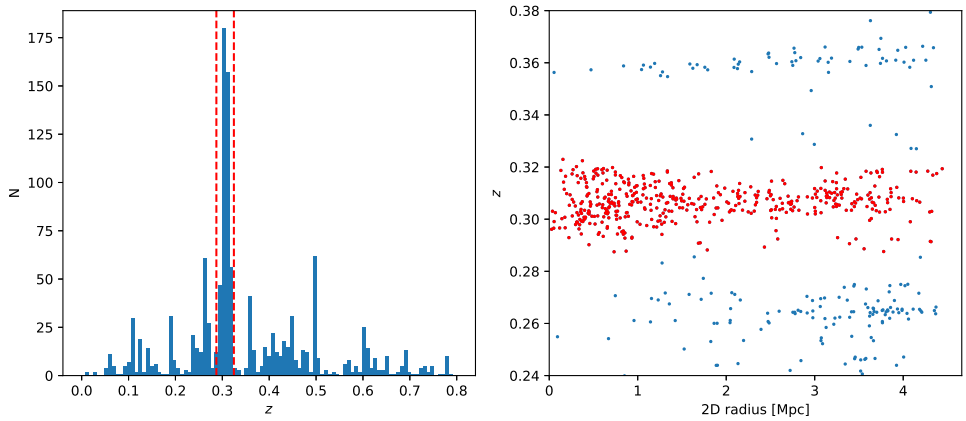

Figure 8.2: Selection of galaxies in the region of A2744. Left panel: redshift histogram of the full catalogue of Owers et al. (2011). The cluster galaxy overdensity is clearly identifiable as the peak around $z \sim 0.3$. The red dashed lines show the redshift selection operated in this work. Right panel: projected 2D radius and redshift distribution of galaxies around A2744. The red points are the galaxies selected to probe the environment of A2744, the blue ones are the other galaxies in the catalogue.

The goal of this analysis is to focus on the cluster outskirts, probing the structures also along the line of sight. Therefore, I select all the galaxies in a rather large volume around the cluster by considering the redshift range $cz_{\text{cluster}} \pm 5600\,\text{km/s}$ (with $z_{\text{cluster}} = 0.306$, Boschin et al. 2006). This redshift selection is shown in Fig. 8.2. For the detection of the filamentary pattern, it is important to ensure spatial uniformity of the completeness so that the excess (lack) of galaxies in a particular region due to selection effects is not mistaken for a real local over(under)density. The spectroscopic completeness (within $\sim 11\,\text{arcmin}$ from the brightest cluster galaxy) was computed by Owers et al. (2011) for different magnitude cuts, and it is shown in their Fig. 9. Based on this result, I chose for the analysis the galaxies with magnitudes $r_F < 20.5$, in order to have the most uniform completeness possible. While there are still some differences in the completeness across regions, in particular towards the west of the cluster, most of the field is complete (with an overall completeness above 90%). The combination of the redshift and magnitude selection provides us with a catalogue of 305 galaxies, which I used to identify the filamentary structures connected to A2744.

In order to study the 3-dimensional structure of the galaxy distribution, I have corrected for the effect of peculiar velocities of galaxies within the cluster on the observed redshifts, in particular the Finger of God (FoG) effect Jackson (1972). To correct for this redshift distortion, I relied on the assumption that a galaxy cluster has a galaxy distribution that is symmetrical along the line

of sight and on the plane of the sky (following Tegmark et al. 2004; Tempel et al. 2012; Hwang et al. 2016). The method we used proceeds in two steps, as in Aghanim et al. (2024). The first step is to identify the affected galaxies thanks to a Friends-of-Friends (FoF) algorithm, using a linking length along the line of sight (LoS) equal to five times the one in the plane of the sky. In our case, I found that this method identifies all the galaxies previously selected as part of the same FoF group[1]. Then in the second step, the LoS distances of the galaxies were compressed according to the group's elongation to remove the FoG distortion. The compression factor is computed as the ratio between the rms of the galaxy positions (w.r.t. the cluster centre) long the LoS and perpendicular to it.

8.2 . The analysis of cluster outskirts

To analyse the distributions of gas and galaxies in the outskirts of A2744 and identify connected filaments, I used the two techniques presented in Chap. 6, namely the aperture multipole decomposition and the T-REx filament finder. In this section I describe how the two methods were applied to the task of analysing galaxy cluster outskirts, and the choice of their free parameters.

8.2.1 . Multipole moments decomposition in cluster outskirts

Galaxy cluster outskirts are characterised by the presence of filamentary structures, which connect to the clusters and funnel the infalling matter (e.g. Reiprich et al. 2013; Walker et al. 2019; Walker & Lau 2022; Rost et al. 2021; Gouin et al. 2021). In these regions, we can expect that matter in the cosmic filaments falls into the gravitational potential well of the cluster in an approximately radial manner. This approximation is expected to be valid within some radial range around the cluster. Therefore, the azimuthal distribution of matter is expected to reveal and identify filamentary structures connected to galaxy clusters.

To study the distribution of gas and galaxies in the outskirts of A2744, I used here the multipolar ratios defined by Eq. 6.2, and the reconstructed map obtained combining the large scales multipoles (Eq. 6.3), both obtained from the aperture multipole moments. When computing the aperture multipole moment decomposition, the two main parameters to consider are the position of the centre and the extent of the aperture $\Delta R = [R_{\rm min}, R_{\rm max}]$. It is therefore important to choose them so that the assumption of radial symme-

[1] I also tested the FoF algorithm on a sample of galaxies with a wider range of velocities around A2744 ($|cz| \lesssim 20000\,km/s$), and found that the galaxies in our selection were in the same group, while the other galaxies were put in different groups in front or behind.

try of the field holds.

Regarding the choice of the aperture centre, in the perfect case of a spherical matter distribution the natural choice would be the minimum of its gravitational potential. However, galaxy clusters are not perfectly spherical and the matter inside them is not homogeneously distributed. Indeed, anisotropic accretion from filaments and merger events disturb their shape and create different offsets between the various matter components and the minimum of the potential. A2744 is a typical example of a significantly disturbed cluster, with many different substructures in the central region and at least two X-ray peaks (e.g. Owers et al. 2011; Jauzac et al. 2016; Medezinski et al. 2016; Harvey & Massey 2024). It is therefore difficult to identify a clear and reliable centre for the azimuthal analysis.

Given that the focus of this work is on the cluster outskirts, I decided to circumvent this issue by examining the isocontours of the X-ray surface brightness map (smoothed with a Gaussian filter of size 7.5 arcsec). I observed that, progressively lowering the signal threshold resulted in the contours evolving from highly disturbed in the cluster innermost region to more regular, and becoming roughly circular at about $1.6 \times 10^{-6}\,\mathrm{counts\ pixel}^{-1}\ s^{-1}$ (with a radius of about $0.6\,R_{\mathrm{vir}} \approx R_{500}$). Lowering the threshold further, the contours started getting more irregular due to the filament emission. Therefore, I decided to centre the aperture ΔR in the middle of this circular contour.

I then chose $0.6\,R_{\mathrm{vir}}$ as the aperture lower radius, R_{min}, in order to exclude the emission from the cluster itself from the analysis, and focus just on the outskirts. The choice of the aperture's upper radius, R_{max}, is restricted by the extent of the data: for the X-ray data, $R_{\mathrm{max}} \leq 1.5\,R_{\mathrm{vir}}$; for the galaxies, $R_{\mathrm{max}} \leq 2.1\,R_{\mathrm{vir}}$. Therefore, for the X-ray case, I chose $R_{\mathrm{max}} = 1.4\,R_{\mathrm{vir}}$, to be symmetric around R_{vir}. For the sparser galaxy data, instead, the need for more statistics motivated the choice of a larger aperture that includes all the available data, $R_{\mathrm{max}} = 2.1\,R_{\mathrm{vir}}$.

A further important parameter to consider in this analysis is the maximum multipolar order included in the computation of the reconstructed map, $m_{\mathrm{max,rec}}$, as it selects the smallest angular scale included in the map. The main interest of this analysis is to identify the signature of large-scale cosmic filaments connected to A2744. It is reasonable to assume that these structures produce an extended signal spanning large angular scales, captured by low multipolar orders. On the other hand, high multipolar orders (that is, small scales) are expected to be dominated by signals from either residual point-like sources (in the case of X-rays) or small concentrations of galaxies. For this reason, I chose $m_{\mathrm{max,rec}} = 7$, which allows me to include all the relevant large-scale contributions in the reconstructed map, omitting the smallest-scale fluctuations (see also Gouin et al. 2020, for details on the choice of multipole order limit).

8.2.2 . Filament detection in cluster outskirts with T-REx

As mentioned above, cosmic filaments connected to clusters are expected to be the most relevant structures in the matter distribution in cluster outskirts. Therefore, the second method I used to characterise the distribution of gas and galaxies around A2744 consists of tracing the ridges of these filamentary structures using the T-REx filament finder (Bonnaire et al. 2020, 2022). One advantage of this method is that T-REx is designed to work with either 2- and 3-dimensional datasets, and thus it can be used to explore the 3D distribution of spectroscopic galaxies and also to trace filamentary structures in the 2D X-ray emission data.

T-REx has already been applied on 2D and 3D galaxy distributions to trace filaments around the Shapley supercluster in Aghanim et al. (2024). Therefore, when applying T-REx to the spectroscopic galaxy distribution around A2744, I used parameter values close to previous studies (Bonnaire et al. 2020; Aghanim et al. 2024): $l = 2$ for the denoising parameter, slightly reduced to account for the lower number of galaxies, and $\lambda = 1$, as suggested in Bonnaire et al. (2020). To compute the probability map, I performed $B = 50$ bootstrap iterations, each time sampling 90% of the dataset, which corresponds to $N_B = 275$.

On the other hand, in this work I employed the T-REx algorithm for the first time to analyse X-ray data, so I needed to adapt the algorithm parameters to the different regime. The X-ray data, as explained in Sect. 8.1, are in the form of a "hit map", where pixels are either 1 or 0 valued. Given that T-REx is designed to work with a discrete set of points, I considered as input for the algorithm the centres of the lit pixels (i.e. equal to 1) in the hit map, whose spatial distribution traces cosmic structures. This results in an input dataset of $N \sim 144\,000$ points. In the central region of the cluster (inside $\sim 0.6\,R_{\rm vir}$), however, the vast majority of pixels in the hit map are lit, resulting in the complete loss of information regarding the inner structure. Therefore, I decided to mask this area and to focus only on the cluster outskirts (beyond $0.6\,R_{\rm vir}$). In these regions outside $0.6\,R_{\rm vir}$, the density of data points is very high, and there is a high noise contamination. Therefore, in order to identify the real signal from filaments, I adapted the values of the algorithm parameters. To reduce the effect of the noise contamination, I used a high value for the denoising parameter, $l = 200$. Then, to encourage the algorithm to focus on the large-scale structures, I imposed a high strength of the topological constraint, $\lambda = 100$, which tends to reduce the total length of the tree and produce smoother branches. For the probability map computation, I used $B = 30$ bootstrap iterations sampling 70% of the data each time, $N_B \sim 101\,000$ points, to allow for more variability between the various realisations and ensure better convergence to the relevant structures.

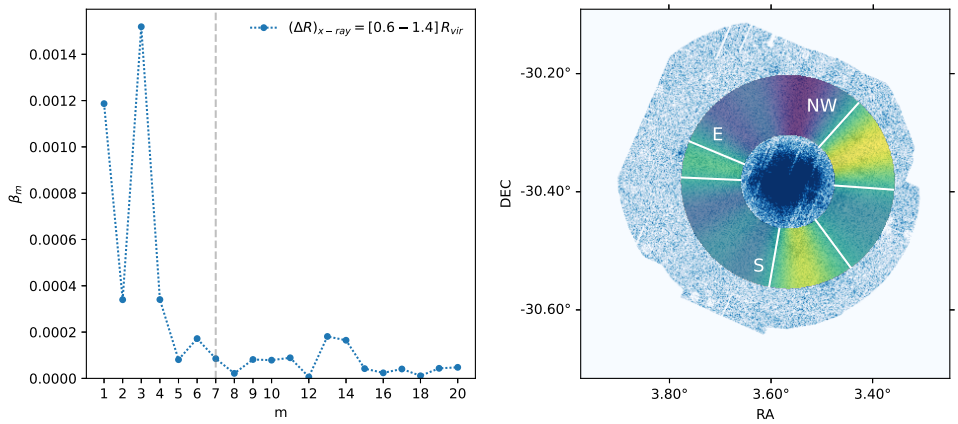

Figure 8.3: Multipolar analysis of the X-ray data. Left panel: Distribution of the multipolar ratio β_m as a function of multipole order m, computed in one radial aperture $(\Delta R)_{\rm x-ray} = [0.6, 1.4]$. The maximum order used in the reconstructed map (right panel), $m_{\rm max,rec} = 7$, is shown as a dashed vertical line. Right panel: The reconstructed map in the aperture $(\Delta R)_{\rm x-ray}$. The white contours represent the threshold of 60% of the map maximum, and identify the relevant filamentary structures, as described in the text. For reference, the X-ray hit-map is shown in the background.

8.3 . The outskirts of Abell 2744

In this Section, I present the results on the outskirts of A2744, from the X-ray and galaxy data, analysed with the aperture multipole moments and T-REx filament finder techniques.

8.3.1 . Azimuthal distribution of matter

I start from the analysis of the azimuthal structures in A2744's outskirts, identified with the use of the aperture multipole moments.

Multipole decomposition of X-ray data

The analysis of the azimuthal distribution of the X-ray data proceeded in two steps. First, I perform the multipolar decomposition in a single large aperture $(\Delta R)_{\rm x-ray} = [0.6, 1.4]\, R_{\rm vir}$, in order to exhibit the large-scale azimuthal behaviour of the X-ray emission in the cluster outskirts. Then, I refine the analysis by splitting the aperture in two smaller apertures inside and outside the virial radius, $(\Delta R)_{\rm in} = [0.6, 1.0]\, R_{\rm vir}$ and $(\Delta R)_{\rm out} = [1.0, 1.4]\, R_{\rm vir}$, in order to probe the radial dependence of the filamentary structures.

Starting from the analysis of the single aperture, we show in the left panel of Fig. 8.3 the multipolar ratios β_m for all orders up to $m = 20$. We can see that the distribution of β_m is not uniform, on the contrary, some orders dominate

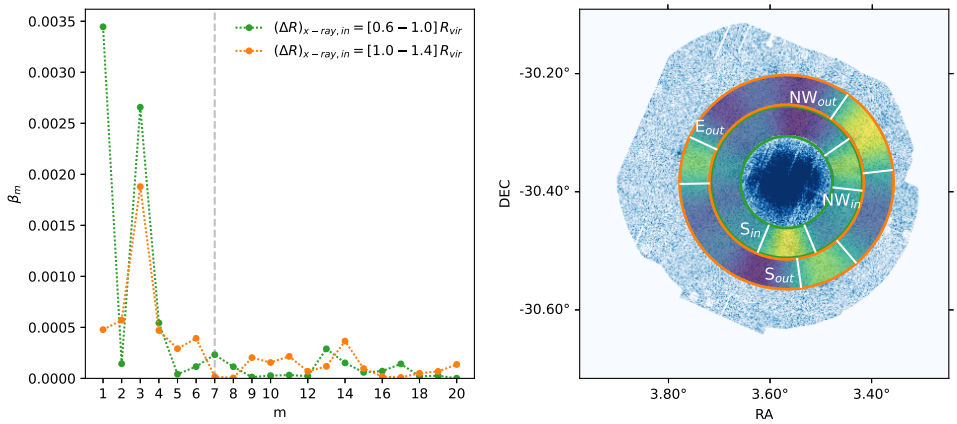

Figure 8.4: Multipolar analysis of the X-ray data. Same as Fig. 8.3 but considering two radial apertures, $(\Delta R)_{\rm in} = [0.6, 1.0]\,R_{\rm vir}$ (in green) and $(\Delta R)_{\rm out} = [1.0, 1.4]\,R_{\rm vir}$ (in orange).

the decomposition, highlighting the angular symmetries of the X-ray field. The most important is the octupolar symmetry (at the order $m = 3$), followed by the dipole ($m = 1$), and then the even orders $m = 2, 4$. These angular symmetries represent structures at large angular scales, as expected for extended emission from cosmic filaments. Indeed, the symmetry at $m = 3$ highlights the presence of three main structures in the outskirts of A2744. In contrast, the dipolar signature shows an asymmetric signal between two halves of the aperture $(\Delta R)_{\rm x-ray}$ (as illustrated in Fig. 6.1). The combination of these two orders suggests that the X-ray data is mostly distributed into three structures, two of which are more prominent than the third (inducing the asymmetry). Finally, the quadrupole order (at $m = 2$) represents an elongated structure. It is often associated with the outer part of the halo elliptical shape (e.g. Gouin et al. 2017, 2020, 2022).

The reconstructed map of the X-ray emission in the aperture $(\Delta R)_{\rm x-ray}$ (obtained by summing the first seven orders of the multipole decomposition) is shown in the right panel of Fig. 8.3. In the map, I identify the relevant structures as the areas where the values are above 60% of the map maximum. These regions are delimited by white contours in the figure. We recognise three structures, which lie approximately in the northwest (NW), south (S) and east (E) of the cluster. We see that, consistently with the expectations from the β_m distribution, the NW and S structures are larger and stronger than the E structure. Additionally, a slight alignment between the E and NW structures can be observed, attributable to the contribution of the even multipole orders $m = 2, 4$.

By considering a large aperture $(\Delta R)_{\rm x-ray} = [0.6, 1.4]\,R_{\rm vir}$, I successfully captured the broad distribution of structures in the outskirts of A2744. How-

ever, inside the aperture the radial information is lost. To recover the radial evolution of the structures, I splitted the aperture in two annuli, $(\Delta R)_{\rm in} = [0.6, 1.0]\, R_{\rm vir}$ and $(\Delta R)_{\rm out} = [1.0, 1.4]\, R_{\rm vir}$. The results of the multipole decomposition in the two sub-apertures are shown in Fig. 8.4.

Focusing on the inner aperture, $(\Delta R)_{\rm in}$, we see that the β_m distribution is dominated by the dipole, with the octupole as second-highest peak (Fig. 8.4, left panel). This inversion of the dominant order (compared to the single-aperture case) suggests a strong difference between two branches of the octupole and the third, which is confirmed by the reconstructed map (right panel). There, we see that only two structures are identified: one in the south ($\rm S_{in}$), and one in the west-northwest ($\rm NW_{in}$) direction. Conversely, no structure is identified in the east of the aperture, where the third branch of the octupolar distribution would be expected. The reason why we cannot identify a structure in the east is that the X-ray signal in that area is weaker and extends over a larger angle compared to the other two regions. This may be related to a shock detected in the northeast region of the cluster (Eckert et al. 2016; Hattori et al. 2017; Rajpurohit et al. 2021).

Looking at the outer aperture, $(\Delta R)_{\rm out}$, we find that β_m is maximum at the order 3 (Fig. 8.4, left panel). All the other orders up to $m = 6$ are weakly contributing to the decomposition, with similar values of β_m. An examination of the reconstructed map of the outer aperture (Fig. 8.4, right panel) reveals the presence of three structures associated with the octupolar order in the northwest ($\rm NW_{out}$), south ($\rm S_{out}$), and east ($\rm E_{out}$) directions. We note that the $\rm NW_{out}$ and $\rm S_{out}$ structures are both shifted counterclockwise with respect to their counterparts in the inner aperture, which highlights a radial dependence of these filaments across the cluster's virial radius. Furthermore, in addition to the angular shift, the $\rm NW_{out}$ structure exhibits a larger angular size than the $\rm NW_{in}$ structure. Conversely, the $\rm S_{\mathit{in}}$ and $\rm S_{\mathit{out}}$ are comparable in size.

Multipole decomposition of galaxy distribution

The low statistics of the galaxy sample (comprising 305 galaxies in total, of which 150 located in the inner cluster region within $0.6\, R_{\rm vir}$) limits the ability of the aperture multipole decomposition, making the results noisy and more sensitive to small galaxy concentrations. Therefore, I consider only one single large aperture, $(\Delta R)_{\rm gal} = [0.6, 2.1]\, R_{\rm vir}$, to analyse the galaxy distribution. It comprises all 155 galaxies beyond $0.6\, R_{\rm vir}$ from the cluster centre.

The left panel of Fig. 8.5 shows the β_m of the galaxy distribution. The most significant order of the decomposition is the octupole ($m = 3$), identically to the X-ray data. We can notice that the quadrupole moment ($m = 2$) is almost as strong as the octupole ($m = 3$). This indicates the presence of three underlying structures, with two aligned on the same axis. Indeed, the reconstructed map (Fig. 8.5, right panel), shows two main structures, identified in the east and west-northwest directions, almost opposite of each other.

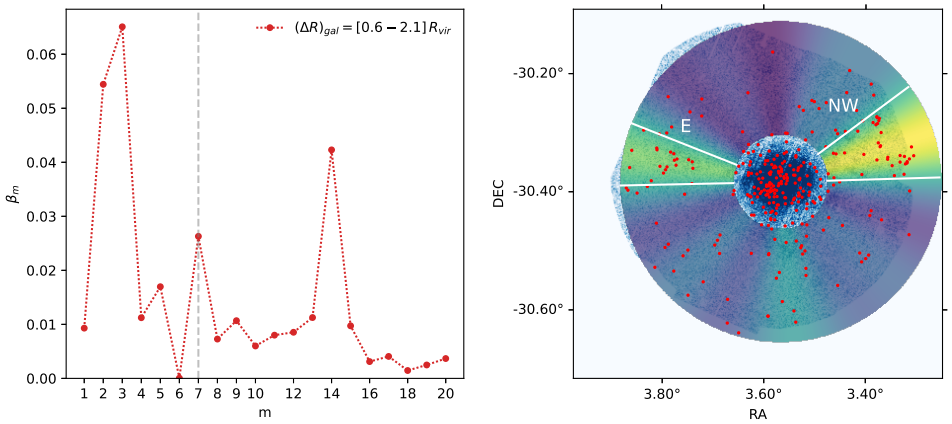

Figure 8.5: Multipolar analysis of the galaxy data. Same as Fig. 8.3, but considering the 2D projected galaxy distribution inside one radial aperture $(\Delta R)_{\rm gal} = [0.6, 2.1]\, R_{\rm vir}$.

A third, less significant structure is hardly visible in the south, consistently with the octupolar symmetry, but it does not cross the threshold of reliable identification.

Another feature visible in the β_m distribution is a high peak at $m = 14$. This order traces very small angular scale structures, incompatible with large-scale cosmic filaments. As discussed in Chapter 7, high multipole orders are correlated with the fraction of substructures. Indeed, by including this order into the reconstructed map, I found that it is driven by small concentrations of galaxies (about $3 - 4$ galaxies with small angular separation). Notice that a similar excess of multipolar power at $m = 13 - 14$ is also weakly significant for X-ray data (in Figs. 8.3 and 8.4). This small-scale feature is not associated with cosmic filaments, but should rather trace sub-clumps of matter, and therefore is not considered further in this analysis.

Comparing the results of the multipolar analysis of galaxies and X-ray signal, there is a very good agreement between the structures identified from the galaxy distribution in the east and northwest of the cluster (Fig. 8.5) and those identified from the X-ray data in the same directions. Conversely, in the south of the cluster the X-ray data reveals a third structure that has no reliable correspondence in the galaxy reconstructed map. A potential explanation for this lack of identification is the non-uniform completeness of the galaxy catalogue. As reported by Owers et al. (2011) (Fig. 9, left panel), in the southern region of the cluster the catalogue is not complete, which might be the reason for the missing S structure in the multipole analysis.

8.3.2 . Filamentary structure around A2744

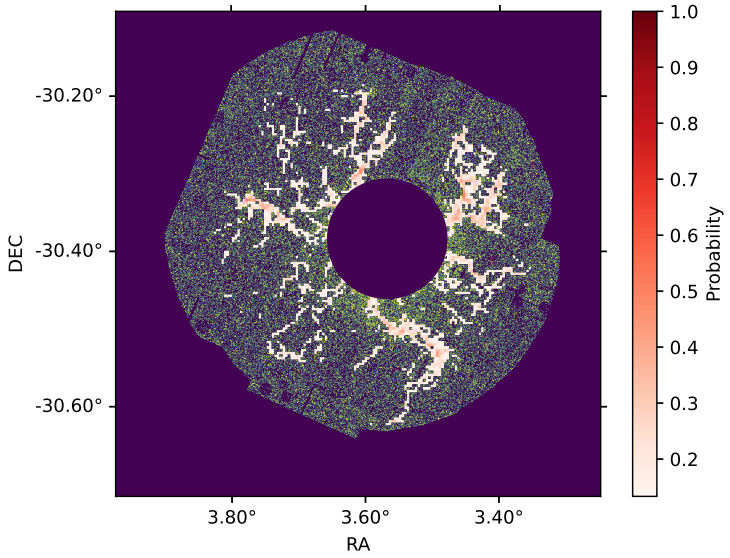

Figure 8.6: Probability map of the filamentary structures from X-ray data, obtained with T-REx . Only pixels with probability larger than 0.1 are shown. For reference, the X-ray hit-map is shown in the background.

A complementary view on the outskirts of A2744 is given by the filamentary structures detected with the filament-finding method T-REx .

Filaments from X-ray data

The result of the T-REx algorithm on the X-ray data is shown in Fig. 8.6, where we see the T-REx probability map (only values above 0.1 are shown), superimposed to the X-ray "hit map". From this map, we can see four large-scale regions of connected, higher-probability pixels, which are identified as filaments. The most prominent of these structures is the filament in the S-SW. From the centre to the periphery, we see it connected to the cluster in the south of the masked area, then it extends to the south-west before exhibiting a bend to the south at RA$\sim 3.5°$. The final section of the structure exhibits a reduction in width and an overall probability, which makes it less reliable as a filament detection. Another structure that can be clearly identified as a filament is the NW one: connected to the cluster in the W-NW, it extends to the NW before spreading and branching out in multiple directions between north and west. When considering the northeast quadrant of the map, the identification of filaments becomes less straightforward. Nevertheless, two structures can be identified due to their higher probability. One extends to the north and splits into two branches of lower reliability. The other extends to the east with relatively high probability, even though its point of connection with the cluster is less clear. Finally, three more sectors show

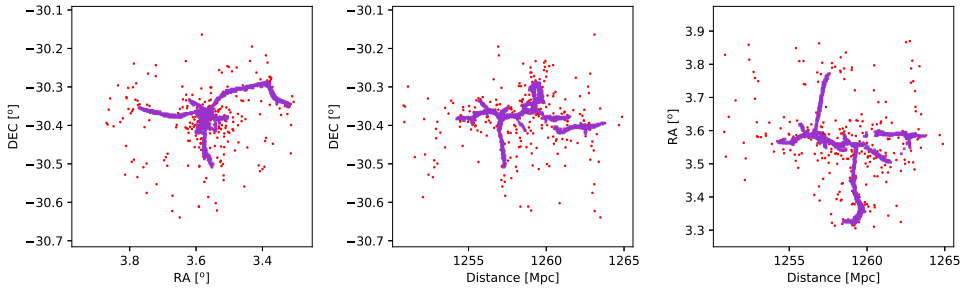

Figure 8.7: Three-dimensional distribution of galaxies (red points) in A2744, superimposed to the 3D probability map of the filamentary structures obtained with T-REx on the galaxy data. Only voxels with probability larger than 0.1 are shown. Left: projection along the line of sight. Middle and right: projections perpendicular to the line of sight, the viewer is on the left in both panels.

loosely-connected regions where the probability is above the threshold of 0.1: west-southwest, southeast, and northeast. While I believe that these regions do not host real filaments, it might be interesting to investigate why some realisations of the T-REx algorithm identify some structures there. It is first necessary to recall that this method was designed to work on much cleaner and sparser data, and that one of its main objectives is to connect overdense regions in a coherent tree structure. This implies that, in a specific bootstrap iteration, if the optimisation identifies a localised overdensity in the noise, it will tend to connect it with the overall tree structure. However, these spurious connections will not be stable across realisations, and will not tend to accumulate into a coherent structure in the probability map. Looking at the surface brightness map, we can notice that in all three regions, there appear to be small overdensities of signal in the vicinity to the cluster, outside the masked region. I am led to believe that these emissions drive the algorithm to link noise structures along those directions, thus generating these broad, noisy probability distributions.

In summary, from the analysis of the T-REx results on X-ray data, four main structures are identified as filaments, extending to the south, northwest, north, and east directions.

Filaments from galaxy distribution

The spectroscopic galaxies allow us to gain access to the full three-dimensional information on the cluster and its environment. This allows the detection of structures along the line of sight, providing a more accurate estimate of the

cluster connectivity. To do that, I applied T-REx to trace the filamentary structures in the 3D distribution of galaxies in and around A2744.

The panels in Fig. 8.7 depict the T-REx probability map thresholded at 0.1 (i.e. regions where the probability is higher than 0.1), superimposed to the galaxy distribution, in three projections. The left panel shows the projection along the line of sight (with axes Right Ascension and Declination), I call this the face-on view. The central panel has the distance along the line of sight as x axis and Declination as y axis, which I call the side view. The right panel has again distance on the x axis but Right Ascension on the y axis, I refer to it as the top view.

Starting from the face-on view, we can see that the T-REx algorithm identifies three filaments connected to the cluster (in the north-west, east, and south directions), consistent with our previous results based on the X-ray data and with the results of Eckert et al. (2015). We notice how the southern filament is considerably shorter than the other two, This is consistent with the results of the multipole analysis, which indicted that in the south region the overdensity of galaxies is less significant than in the other structures. Another noteworthy feature is the rather sharp bend in the NW filament at RA $\sim 3.5°$, connecting a concentration of galaxies in the far west of the field. This may be attributed to a selection effect, being so close to the edge of the observations.

From the side view, we can observe the cluster and its surrounding structures along the line of sight (on the x axis). The first thing we notice is the presence of an elongated structure in the line-of-sight direction (horizontal in the picture), which extends both in front and behind the centre of the galaxy distribution, at relatively the same Declination. I will refer to it as the central structure. Moreover, we observe that T-REx identifies a structure at the rear of the cluster, slightly separated from the main branch in the line of sight. The projected position of this structure along the line of sight (at roughly RA $\sim 3.6°$, Dec $\sim -30.4°$) corresponds to the southern concentration of galaxies visible in the central region of the cluster. Furthermore, the southern filament is visible from an alternative perspective. We can see that it exhibits minimal extension in the line of sight direction, and that it connects to the central structure a little in front of the centre of the galaxy distribution.

Finally, from the top view we can extract some additional information about the E and NW filaments (in the upper and lower parts of the figure, respectively). Indeed, it can be observed that the E filament connects to the central structure in approximately the same position as the S filament. In contrast, the NW filament is connected further back, in a location that, projected on the plane of the sky, can be associated approximately with an X-ray peak in the northwest of the cluster centre. Moreover, we observe that both these filaments (actually all three considering also the S one) do not extend much in the line-of-sight direction, but rather appear to be roughly perpendicular

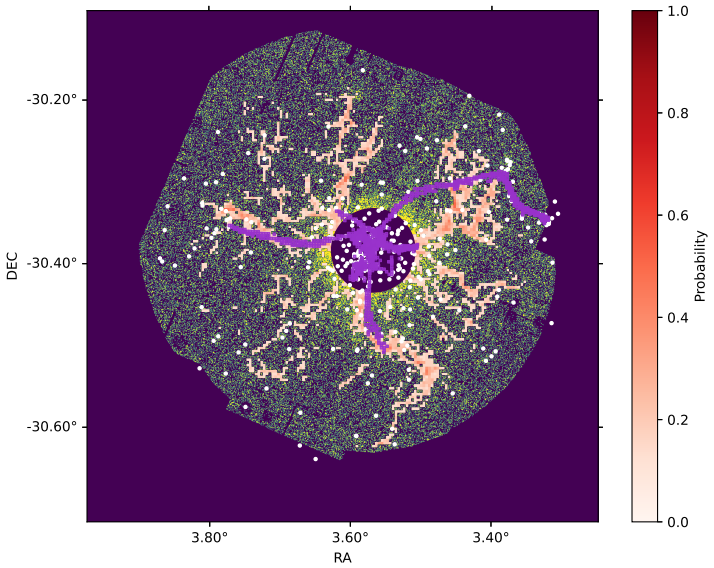

Figure 8.8: Comparison of the T-REx probability map from X-ray data (Fig. 8.6) and the face-on projection of the T-REx probability map from galaxy data (Fig. 8.7, left panel). For reference, both the X-ray hit-map and the galaxy distribution are shown in the background.

to the line of sight. Lastly, we see that the central structure is not exactly parallel to the line of sight, but tends to extend from east to west as the distance increases, with the exception of the front-most part.

From these results, we obtain a picture of the three-dimensional structure of the A2744 cluster environment, as traced by the galaxy distribution. The detected structures are: three filaments almost perpendicular to the line of sight; a long, extended filamentary structure along the line of sight, and a disconnected structure at the back of the cluster (slightly southeast of the centre). Two of the filaments (E and S) are observed connect in the same position in the front-eastern part of the cluster, while the NW filament is connected in the back of the cluster, towards the west. Along the line of sight, the central structure extends beyond the virial radius both in front and behind the cluster. The front branch extends, almost parallel to the line of sight, from the cluster centre. In the back of the cluster, the other branch extends towards the west from the connecting point of the NW filament. Also in the back of the cluster, T-REx identifies another structure not directly connected to the main one, but crossing the virial sphere of the cluster. Located in the southwest of the cluster centre, this back structure can be associated with the southern peak in the X-ray surface brightness map.

A comparison of the probability map derived from the X-ray data with

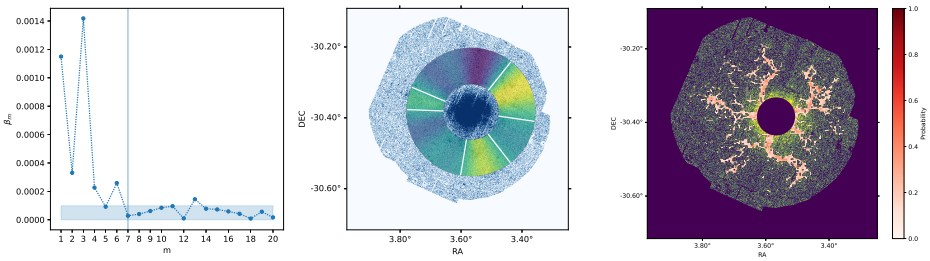

Figure 8.9: Results of the multipole decomposition and T-REx analyses, using the X-ray data with just the high-reliability point sources masked (see text). Left and middle: multipole analysis, same as Fig. 8.3. Right: T-REx probability map, same as Fig. 8.6.

that obtained from the 3D galaxy distribution (Fig. 8.8) reveals a good level of agreement between the filaments identified from the two datasets. The east and south filaments have an excellent overlap, while in the NW the two structures extend in the same broad direction, but present different shapes and cross at larger distance from the cluster centre. The two probability maps reveal also other notable discrepancies. The first one is the absence of a counterpart in the galaxy data for the X-ray-detected N structure. This confirms the results of Eckert et al. (2015), which also identified a northern structure in the X-ray map, but discarded it due to the lack of galaxies in the region, despite the good completeness. Instead, the X-ray emission was attributed to a background galaxy concentration. Another interesting difference is the length of the southern filament. In the X-ray probability map, this filament extends to a considerable length in the SW region, while in the galaxy case only the part close to the cluster is traced, out to $\sim 1\,R_{\rm vir}$. A possible explanation for this difference is again the lack of galaxies in the southwestern region due to the lower completeness (Owers et al. 2011) (Fig. 9). The non-uniform spectroscopic completeness might also explain the different shape of the NW filament in the two T-REx probability maps, since the western region around $R_{\rm vir}$ is also reported to have lower completeness by Owers et al. (2011).

8.4 . Robustness of results

To assess the robustness of the results presented in Sect. 8.3 I performed several tests to investigate the impact of different choices both in the data preprocessing and the methods parameters.

8.4.1 . Robustness to data preprocessing choices

I tested the impact of some of the choices described in Sect. 8.1. For the baseline analysis of the X-ray data, we used a conservative point-source mask, that included all the point-like sources in the soft band image. I tested the im-

pact of this choice by repeating both the multipole analysis and the T-REx filament detection masking only the high-reliability point sources; that is, those obtained by cross-matching the point sources lists detected in the soft and hard energy bands. The results obtained from this set-up (Fig. 8.9) are very similar to those of the baseline set-up (Figs. 8.3 and 8.6). We can notice some small differences in the β_m distribution (Fig. 8.9, left panel), but the reconstructed map (middle panel) does not change appreciably. The T-REx probability map (Fig. 8.9, right panel) highlights the same four filaments as the one obtained using the conservative point-source mask, and it looks somewhat less noisy, probably due to the alignment of some compact emissions with the filament signal, which helps the T-REx algorithm orient in those directions.

Moving on to the 3D galaxy distribution, one assumption in the baseline analysis is the magnitude cut at $r_F < 20.5$, which is based on the spectroscopic completeness. I tested the impact of this choice, by relaxing the magnitude constraint to $r_F < 21$, and repeated the analysis. This updated threshold yields a catalogue of 412 galaxies (35% more than the baseline). As for the X-ray data, the results are consistent with the baseline analysis.

8.4.2 . Robustness to method parameters

I tested the robustness of the analysis techniques, by varying the methods' free parameters. For what concerns the multipole analysis, I focused on the impact of the size of the aperture (i.e., the choice of the radial boundaries of the annulus). The choice of the aperture boundaries, $R_{\rm min}$ and $R_{\rm max}$, clearly influences the results, as demonstrated in Sect. 8.3.1 in the difference between using one or two apertures on the X-ray data (Figs. 8.3-8.4). Nevertheless, I observed that the qualitative results for the X-ray data in the single aperture case are mostly consistent across a wide range of radial limits. Indeed, when I fix the upper boundary to the baseline value, $R_{\rm max} = 1.4\,R_{\rm vir}$ and vary the lower one, $R_{\rm min}$, from 0 to $1.2\,R_{\rm vir}$, I find that the most important order is always the octupole, and the structures identified in the reconstructed map are always three, with only minor differences in position. On the other hand, fixing $R_{\rm min} = 0.6\,R_{\rm vir}$ and varying $R_{\rm max}$, I find that between 0.8 and $1.3\,R_{\rm vir}$ the dipole dominates over the octupole in the β_m distribution, and in the reconstructed maps only the S and NW structures can be identified. This is consistent with the results of the inner aperture in Fig. 8.4, and with the fact that the eastern structure is harder to detect in the vicinity of the cluster. Varying $R_{\rm max}$ to values higher than $1.3\,R_{\rm vir}$, the octupolar order becomes dominant in the decomposition, and I find back the three filaments of the main analysis. These remain mostly unchanged with larger apertures, up to $R_{\rm max} = 1.5\,R_{\rm vir}$, which is the maximum distance allowed by the data.

The impact of the aperture size on the analysis of the galaxy data is more relevant, due to the small number of galaxies in the dataset. In fact, in the

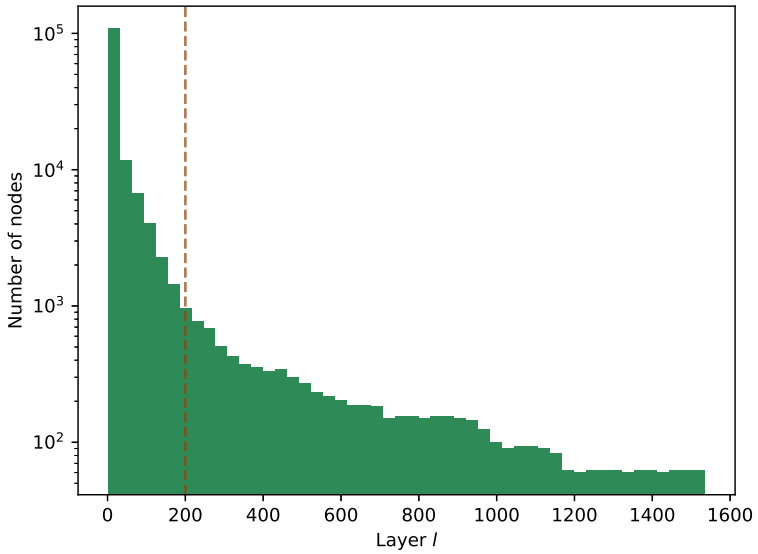

Figure 8.10: Onion decomposition of the minimum spanning tree constructed from the X-ray data. The vertical dashed line represents the denoising level l chosen in the analysis.

sparse data regime, the multipole decomposition becomes particularly noisy, and the inclusion or exclusion of just a few galaxies can have a significant impact on the results. Therefore, it is necessary to use a large enough aperture to accumulate sufficient statistics. In the baseline analysis, I made the conservative decision to include in the aperture all galaxies located beyond $0.6\,R_{\rm vir}$ from the cluster's centre. To test the robustness of the results obtained with this setup, I fixed $R_{\rm min} = 0.6\,R_{\rm vir}$ and progressively increased the aperture size ΔR. Using small apertures, up to $R_{\rm max} \sim 1.5\,R_{\rm vir}$, I identify two structures in the reconstructed map, in the south and northwest directions respectively (as for the X-ray data). Moving the upper boundary further, I find that the southern structure disappears, and later the eastern structure appears (around $1.7\,R_{\rm vir}$) and becomes more prominent as we expand the aperture to larger radii. The identification of a southern structure in the vicinity of the cluster confirms the results obtained from T-REx on the galaxy distribution (Sect. 8.3.2), which found a filament in the south of the cluster, but with shorter length than the ones in the east and northwest.

Regarding the T-REx algorithm, as explained in Sect. 8.2.2, the application on equivalent galaxy data had been already performed (Aghanim et al. 2024), so I needed only small adjustments to the algorithm parameters to obtain the results on the galaxy distribution around A2744, notably reducing the denoising parameter l to account for the lower number of galaxies.

On the other hand, for the X-ray data, I performed several tests to adapt

the parameters to this previously unexplored regime, characterised by a large density of points and high noise level. This was done in order to minimise the noise of the probability maps and ensure the best convergence of the trees. I found that a combination of high denoising, represented by the parameter l, and a high strength of the topological constraint, represented by the parameter λ, helps to reduce the noise in the T-REx probability map. In the tests, iteratively pruning the tree a large number of times proved useful in removing spurious detection. To find a suitable number of pruning iterations, and avoid deleting relevant tree branches, I used the onion decomposition of the initial minimum spanning tree (Hébert-Dufresne et al. 2016), as suggested by Bonnaire et al. (2020). This method identifies the number of nodes in the tree which gets removed by each successive pruning iteration. The resulting onion spectrum for the X-ray data is shown in Fig. 8.10. The decaying number of discarded nodes in the first pruning iterations (small l) is interpreted as the removal of short branches, while when this number becomes almost constant means that the resulting tree is stable in terms of umber of branches. We see that the choice of $l = 200$ used in the analysis is still rather conservative, as it allows for a large number of nodes to survive, thus allowing for more flexibility in the optimisation step of the algorithm. For the choice of λ, I found that high values (of the order of 10^2), help reducing spurious connections and border effects, producing smoother branches and reducing overfitting. Therefore reducing the noise in the probability map.

Nevertheless, I find that the results of T-REx are consistent across a wide range of parameter values, and that the regions identified as filaments tend to have higher probabilities even in very noisy realisations of the probability map. This reduces the need to find the optimal free parameters.

8.5 . Discussion

The outskirts of the galaxy cluster A2744 have been the target of various studies, that probed them with spectroscopic galaxy and X-rays observations (Braglia et al. 2007; Owers et al. 2011; Ibaraki et al. 2014; Eckert et al. 2015; Hattori et al. 2017). Braglia et al. (2007) used a sample of 194 spectroscopic galaxies observed with the VIsible MultiObject Spectrograph at ESO's Very Large Telescope (VLT-VIMOS, all of which were also part of the sample used in this work), while Owers et al. (2011) used the same sample of galaxies I used in this work. Both studies combined position and velocity information, finding two overdense regions in the south and northwest of A2744 which were identified as large-scale filaments. Owers et al. (2011) also highlighted an overdensity of galaxies in the east of the cluster. However, they found that the local velocity distribution does not differ significantly from the overall cluster distribution, so it was not identified as a relevant structure by the authors.

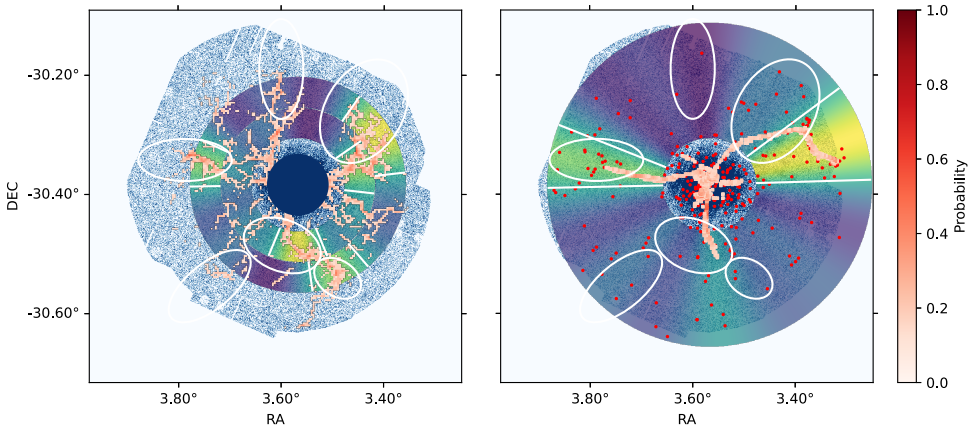

Figure 8.11: Results of the multipole and T-REx analyses, on X-ray data (left) and on galaxy data (right). The reconstructed map from the multipole decomposition is superimposed on the T-REx probability map. The white ellipses correspond to the regions identified in Eckert et al. (2015). For reference, the X-ray hit-map is shown in the background.

Eckert et al. (2015) ($E15$) analysed the adaptively-smoothed X-ray surface brightness map (in the soft energy band) observed by *XMM-Newton*, and identified six regions of extended emission in the cluster outskirts. Four of these filamentary structures were found to coincide with galaxy concentrations: one in the northwest (NW_{E15}), two in the south and southwest ($\mathrm{S{+}SW}_{E15}$), and one in the east (E_{E15}). The first two regions correspond to the findings of Braglia et al. (2007) and Owers et al. (2011), while the east one corresponds to the overdensity identified Owers et al. (2011). The other two X-ray emission regions (in the southeast, SE_{E15}, and north, N_{E15}) were associated with concentrations of galaxies in the background and foreground of A2744, not connected to it. Analysing the X-ray spectra in the detected regions, Eckert et al. (2015) and Hattori et al. (2017) found evidence that the gas originating the X-ray emission is in the form of Warm-Hot Intergalactic Medium (WHIM). This is consistent with the results of Gouin et al. (2023), which showed that, in hydrodynamical simulations, the dominant source of soft X-rays beyond the virial radius is the warm gas in the WHIM and warm circumgalactic medium (WCGM). These pieces of information support the identification of the regions detected from the soft-band X-ray image as cosmic filaments, as several studies showed that, in simulations, the WHIM is the most important gas phase in cosmic filaments, and can be reliably used to trace them (e.g. Martizzi et al. 2019; Galárraga-Espinosa et al. 2021; Tuominen et al. 2021; Gouin et al. 2022, see also Chapter 7).

In this work, I performed the blind detection of filamentary structures connected to A2744, applying two statistical techniques (reconstructed maps

from multipole decomposition, and T-REx filament finder) to the X-ray and spectroscopic galaxy data.

In Fig. 8.11, I compare the structures identified with the two aforementioned methods, for X-ray (left panel) and galaxy (right panel) data, with the X-ray-identified structures (white ellipses) from Eckert et al. (2015) (and thus, implicitly, with the other works mentioned above). From the X-ray data, we see that for the NW, S, and E structures there is a remarkably close correspondence between the results of our two methods. The detected filaments from T-REx match well the region identified in the reconstructed maps, both in terms of position and radial dependence, including the gap between the cluster and the E filament. Furthermore, these structures are well in agreement with the NW$_{E15}$, (S+SW)$_{E15}$ and E$_{E15}$ regions from Eckert et al. (2015). The fact that these regions (the ones with confirmed galaxy counterparts in Eckert et al. (2015)) are the ones where our methods agree highlights the complementarity of the two methods, and the robustness of their combined detections. We also note that the N filament, detected by T-REx , has no significant match in the multipole analysis, but it does match with a structure (N$_{E15}$) initially identified and then discarded by Eckert et al. (2015). Finally, the SE$_{E15}$ ellipse has no detected counterpart in either of our two methods.

For the galaxy data, it should be noted that the T-REx algorithm is run on the 3D distribution of galaxies, while the multipole analysis is performed on the projected 2D distribution. Comparing the results of the two methods along the line of sight (Fig. 8.11, right panel), we see that: in the east, the T-REx -detected filament overlaps well with the E structure in the reconstructed map and with the E$_{E15}$ ellipse. In the northwest, the overlap between the different detections is only partial, but they all agree on the presence of a structure in this area. The reasons for these differences might be selection effects (due to the non-spatially-uniform completeness of the galaxy sample) or border effects. In the south, as mentioned in Sect. 8.3, the detection of a filamentary structure is more difficult: the T-REx algorithm detects a filament, but it only extends to $\sim R_{\rm vir}$, while in the same region the reconstructed map has its third-highest structure, but it falls below the detection threshold (moreover, a structure is detected in this area when using a smaller aperture, e.g. $[0.6 - 1.4]R_{\rm vir}$). This is likely to be attributed to the lower completeness in the southern area beyond $\sim 1\, R_{\rm vir}$ (see Owers et al. 2011, Fig. 9).

By combining the findings from X-ray and galaxy data, I can construct a comprehensive representation of the surrounding environment of the cluster A2744. From the analysis of A2744 on the plane of the sky, I identify three filaments connected to the cluster (NW, S, and E). These three are consistent across almost all the combinations of probe (X-rays or galaxies) and detection method (T-REx or azimuthal analysis), with larger uncertainty for the S filament in galaxy data, and also consistent with previous works. All three

extend out to $\sim 1.5\,R_{\rm vir}$. Furthermore, the application of T-REx to the three-dimensional distribution of galaxies revealed the existence of two additional filamentary structures connected to it, one in front of and one behind the cluster. However, the precise extent of these structures may depend on the specific FoG correction we have applied to the galaxy sample. In total, I thus identify five filamentary structures connected to the core of A2744.

Such a large number of connected structures can help explain the particularly complex internal structure of the cluster. A2744 is known for its exceptional number of massive substructures, which make it a very dynamically disturbed cluster (Merten et al. 2011; Owers et al. 2011; Jauzac et al. 2015, 2016; Medezinski et al. 2016; Bergamini et al. 2023; Harvey & Massey 2024). As mentioned before several times, cosmic filaments act as highways along which gas and galaxies preferentially fall into clusters (Rost et al. 2021), and thus can point to the origin and direction of some substructures. For example, Jauzac et al. (2016) reported the alignment of three substructures in the direction of the NW filament. Another substructure in the north of the central cluster area shows evidence of a northward movement (Owers et al. 2011; Jauzac et al. 2016), suggesting a possible origin in the direction of the S filament. In a similar way, a shock front detected in the southeast of the cluster centre (Owers et al. 2011) suggests motion of substructures coming from the northwest.

These results on the connected filaments of A2744 can be compared with those obtained from statistical studies of larger cluster populations. To do so, I define the cluster connectivity κ as the number of filaments connected to the cluster that cross a sphere of $1.5\,R_{\rm vir}$ radius (following Darragh Ford et al. 2019; Gouin et al. 2021). With the aforementioned uncertainties on the filaments in mind, I estimate the connectivity of A2744 to be in the range $\kappa \sim 3{-}5$. It has been shown from both simulations and observations (Aragón-Calvo et al. 2010b; Codis et al. 2018; Darragh Ford et al. 2019; Sarron et al. 2019; Kraljic et al. 2020; Lee et al. 2021; Gouin et al. 2021; Galárraga-Espinosa et al. 2023) that higher mass clusters tend to have higher connectivity. Moreover, Gouin et al. (2021) showed that, for a given mass, dynamically unrelaxed clusters are typically more connected than relaxed ones. Therefore, a value of $\kappa \sim 5$ for the connectivity of A2744 would be well in agreement with trends from hydrodynamical simulations (Gouin et al. 2021). Comparing to other observations of individual clusters, the connectivity of the Coma cluster ($M_{200} = 5.3 \times 10^{14} M_{\odot}$, Gavazzi et al. 2009) has been determined to be $\kappa = 3$ (Malavasi et al. 2020, 2023), which seems to be more in line with that of relaxed clusters from simulations (Gouin et al. 2021). On the other hand, Einasto et al. (2020, 2021) studied the connectivity of clusters in superclusters, thus in dense regions with many clusters and groups gathered together. In these environments, massive clusters were found to be highly connected. For example, A2142 ($M_{200} = 1.2{\times}10^{15} M_{\odot}$, Munari et al. 2014) has $\kappa = 6{-}7$ (Einasto

et al. 2020), and A2065 ($M_{200} = 2.3 \times 10^{15} M_{\odot}$, Pearson et al. 2014) is even more connected, with $\kappa = 9$.

8.6 . Conclusion

In this Chapter, I presented an analysis of the matter distribution around the galaxy cluster A2744, traced by the X-ray emission and galaxy positions. Using two statistical methods, the aperture multipole decomposition and the filament-finding method T-REx , I identified filamentary structures connected to the cluster. In summary, I presented the following results:

— I reported for the first time the blind detection of filaments connected to a cluster from X-ray data. Both methods applied showed the presence of three filamentary structures in the S, NW and E of the cluster. These findings are consistent with previous visual detection by Eckert et al. (2015).

— Three filaments are also detected from the distribution of galaxies at approximately the same (projected) positions as their X-ray counterparts, although the S filament is only clearly detected by the T-REx method close to the cluster.

— The X-ray-based T-REx probability map showed an additional structure in the north, coincident to one identified by Eckert et al. (2015). However, this result was not confirmed by the multipole analysis nor by galaxy data.

— Through the galaxy distribution, I studied the 3-dimensional filamentary structure extracted with the T-REx method: I found that the three filaments identified in 2D are all almost perpendicular to the line of sight. However, while the S and E filaments lie almost on the same plane, the NW one is found further in the back and connects to a different part of the cluster. The central region of the cluster is strongly elongated in the radial direction, and extends beyond $R_{\rm vir}$ in both the front and back. Furthermore, a loosely connected structure is identified in the back of the cluster, which seems to be aligned with the southern X-ray peak.

— The number and positions of the detected filaments can improve the interpretation of the highly disturbed cluster centre, in particular concerning the origin and direction of motion of its many substructures.

— I estimated the connectivity of A2744 to be in the range $\kappa \sim 3 - 5$, which is in agreement with trends from hydrodynamical simulations for a massive, disturbed cluster.

On the methodology side, I successfully applied T-REx for the first time on X-ray data. By adapting the model's parameters. I proved its applicability to a previously unexplored data configuration and demonstrated its inherent flexibility.

With the combination of different methods for identifying filamentary structures, applied to different probes, I proved the possibility of blind detection of filaments in the outskirts of galaxy clusters. This result could open the way to a systematic search for cosmic filaments connected to clusters, in particular for what concerns the gas component, thanks to large X-ray programs such as eROSITA (Bulbul et al. 2024), CHEX-MATE (CHEX-MATE Collaboration et al. 2021), and XRISM (XRISM Science Team 2020).

9 - A generative model for realistic galaxy cluster images

In order to perform statistical studies on clusters, particularly to test new techniques for cluster detection and analysis, or to explore features that are not easily accessible through observations (like I did in Chapter 7), it is often necessary to obtain large, representative samples of simulated objects that accurately reflect the full variability and range of the cluster population. Therefore, there is the need for cosmological hydrodynamical simulations that are both high-resolution (to capture the cluster physics down to small scales) and encompassing large volumes (to allow for enough high-mass systems to form). But, performing such simulations is extremely computationally expensive (requiring several millions of CPU hours to compute, e.g. Nelson et al. 2018b), and thus the choices of simulated volume and resolution are limited by the available resources. A further complication arises if one wants to study the impact of cosmology on galaxy cluster properties. In that case, not just one simulation is needed, but a series of simulations with varying cosmological parameters (e.g. Villaescusa-Navarro et al. 2021).

One way to obtain samples of simulated galaxy clusters with reduced computational cost is to perform so-called "zoom-in" simulations. In this approach, the starting point is a large dark-matter-only simulation, which is much less computationally intensive than a hydrodynamical one. From this, the regions around clusters are selected and re-simulated adding baryonic physics. This allows the construction of samples of up to a few hundreds of massive clusters with high resolution, without the need to run the hydrodynamical code on the full simulation volume. A comparison of recent hydrodynamical simulations, both full-box and zoom-in, is shown in Fig. 9.1, placed according to their number of massive clusters and baryonic mass resolution.

Another possibility is the use of machine learning methods, trained on simulations, to produce accurate synthetic data that complement or expand existing simulations. The use of machine learning methods in cosmology and astrophysics has grown very rapidly in recent years, with applications in many different aspects, from classification and detection of objects to field-level cosmological analysis. In the specific case of assisting cosmological simulations, some of the main applications concern, for example: the inclusion of baryons in dark-matter-only simulations, either by predicting observational properties from DM halos (Xu et al. 2013; Kamdar et al. 2016; Jo & Kim 2019; de Andres et al. 2023) or directly the baryon distribution (Agarwal et al. 2018; Tröster et al. 2019; Dai & Seljak 2021; Lovell et al. 2022; Delgado et al. 2022); the production of a high-resolution matter distribution starting from a low-resolution one (Kodi

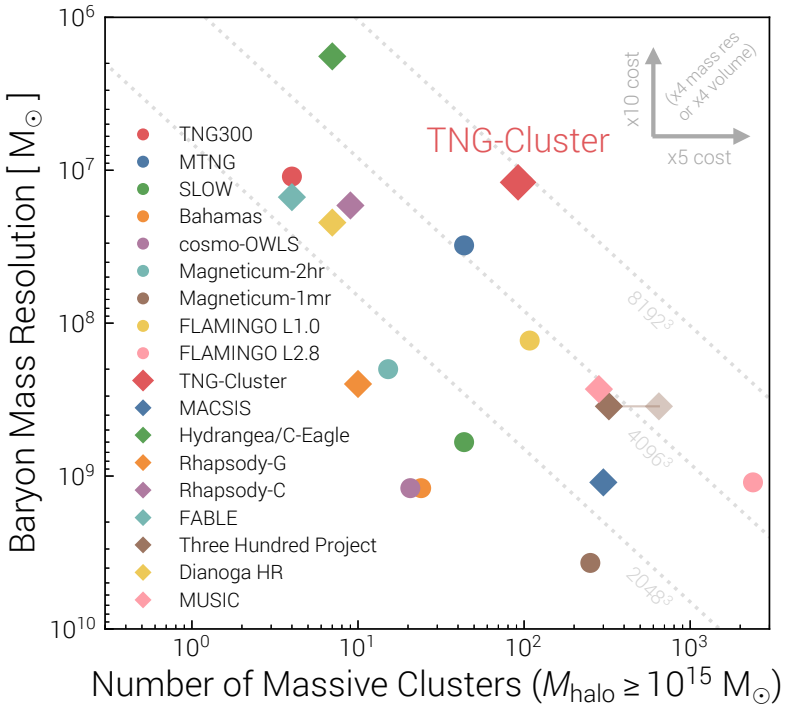

Figure 9.1: Comparison of the resolutions (baryonic particle mass or target cell mass) versus the number of simulated massive halos with $M_{200} \leq 10^{15}\,M_{\odot}$, for different cosmological, hydrodynamical simulations from the literature. Full boxes are shown as circles, and zoom-in suites as diamonds. The grey diagonal lines indicate the number of baryonic resolution elements. The simulations shown are IllustrisTNG (Nelson et al. 2019a), MillenniumTNG (Pakmor et al. 2023), SLOW (Dolag et al. 2023) BAHAMAS (McCarthy et al. 2017), cosmo-OWLS (Le Brun et al. 2014), Magneticum (Dolag et al. 2016), FLAMINGO (Schaye et al. 2023), TNG-Cluster (Nelson et al. 2024), MACSIS (Barnes et al. 2017b), Hydrangea/C-EAGLE (Bahé et al. 2017; Barnes et al. 2017a), Rhapsody-G/C (Hahn et al. 2017; Pellissier et al. 2023), FABLE (Henden et al. 2018), The Three Hundred Project (Cui et al. 2018), DIANOGA (Bassini et al. 2020), and MUSIC (Sembolini et al. 2013) . Figure taken from Nelson et al. (2024).

Ramanah et al. 2020; Li et al. 2021; Rouhiainen et al. 2023; Schanz et al. 2023); or even the direct generation of new synthetic data, bypassing any simulation step (Rodríguez et al. 2018; Mustafa et al. 2019; Ullmo et al. 2021; Han et al. 2021; Jamieson et al. 2023; Boruah et al. 2024; Andrianomena et al. 2024; Ullmo et al. 2024).

Similar to the generation of large-scale structure fields is another approach, which consists of generating images of single objects, instead of full simulations. This was done for galaxy images in the context of *Euclid* by Euclid Collaboration et al. (2022), and for images of the SZ effect from galaxy clusters by

Rothschild et al. (2022). In the context of galaxy cluster, the ability to generate realistic images of cluster observables is particularly interesting in light of the current and upcoming surveys (for example, eROSITA (Bulbul et al. 2024) in X-rays, Euclid (Euclid Collaboration et al. 2024) in optical, and CMB-S4 (Abazajian et al. 2022) at mm wavelengths) to test the effect of cluster morphology on the data analysis pipeline, from the cluster detection and selection function (see also Part III of this Thesis), to the mass estimation and scaling relations, just to cite some examples. Moreover, when generating images of clusters, it is also interesting to have control over some of the physical properties of the clusters, in order to produce different populations of cluster for different purposes Both Euclid Collaboration et al. (2022) and Rothschild et al. (2022) use different techniques to condition their generative ML models to produce images according to some given input parameter.

Inspired by these works, I trained a conditional variational autoencoder (CVAE Sohn et al. 2015) with the goal of generating large samples of realistic high resolution images of the Compton-y parameter of galaxy clusters (proportional to the integrated electron pressure, see Sect. 3), given the cluster mass as input. Such a generative model of cluster images could be a valuable improvement over generation from analytical models (most of which assume spherical symmetry), for example in applications like validation of detection algorithms and selection function characterisation.

9.1 . Generative models

Generative models are a class of machine learning methods, whose goal is to generate new data with similar properties to a target dataset, so that, in the best case scenario, the generated data are indistinguishable from the original one.

This problem is framed in a probabilistic setting, in the following way (Lamb 2021): instances from the considered dataset, x, are treated as samples from a probability distribution, $x \sim p(x)$. On the other hand, we can describe a generative model as an estimating distribution, $q_\theta(x)$, which depends on a set of parameters θ. Therefore, the aim of a generative model is to find the set of parameters θ^*, so that $q_{\theta^*}(x)$ approximates as well as possible $p(x)$. In this way, sampling from $q_{\theta^*}(x)$ gives results with statistical properties similar to the original data.

One common approach to model a generative process is to use neural networks to map a simple probability distribution (for example a Gaussian) into a more complex one. In this way, the parameters of the neural network are the ones defining $q_\theta(x)$, which can be very complex, and at the same time it is easy to generate samples in the space of the simple pdf and feed them to the neural network, which transforms them in samples from $q_\theta(x)$.

While there are many different kinds of generative models, the three main and most popular classes are diffusion models, generative adversarial networks and variational autoencoders:

Diffusion models

Also known as denoising diffusion models (Sohl-Dickstein et al. 2015; Ho et al. 2020), these models learn the mapping between data points (e.g. images) and a random Gaussian field of the same size via a two-step process. The first step consists of adding iteratively small amounts of Gaussian noise to the training data until the noise dominates over the data, while the second step tries to reverse the destructive diffusion process to reconstruct the original data. New data can then be generated by running the reverse process on randomly sampled fields. These methods have proven to produce very high quality generation and good variety of samples, which motivated their recent popularity (e.g. Dhariwal & Nichol 2021; Peebles & Xie 2022), but are quite slow to train and to sample.

Generative Adversarial Networks

Generative Adversarial Network (GANs Goodfellow et al. 2014) are composed of two modules competing against each other. One is the generator, which tries to produce new synthetic data from random noise, and the other is the discriminator, which has to distinguish the generated data from the real training data. The two modules are trained together, to progressively increment their performances. GANs allow fast sampling and high quality generated data (e.g., for cosmic web generation, Rodríguez et al. 2018; Feder et al. 2020; Ullmo et al. 2021), but have proven to be rather difficult to train and prone to overfitting (i.e. learning too precisely the specific training set and reproducing data points that are too similar to the training data, thus lacking the ability to generalise, see e.g. Arora & Zhang 2017; Grover et al. 2017).

In my work, I used a variational autoencoder model to generate images of galaxy clusters. In the next Section, I present these models more in detail, with a focus on their use for image generation.

9.2 . Variational Autoencoders for image generation

Variational autoencoders (VAEs Kingma & Welling 2013) are generative models based on autoencoders, which are composed of two parts, an encoder and a decoder. The goal of the encoder is to compress each input image x into a low-dimensional space, called the *latent space*, so that the relevant information of the input data is stored in a compact form in this latent space, and

similar images are mapped close to each other. This latent representation is then used by the decoder to reconstruct the original data. During training, encoder and decoder are used together to produce reconstructed data that matches as closely as possible with the training data. In this way, they learn to efficiently compress the data information in the latent representation. After training, it is then sufficient to sample points from the latent space and feed them to the decoder to generate new samples. VAEs are good at capturing the full variability of the training data, and are in general faster to train and sample than other methods, but often their outputs are less detailed than those of GANs and diffusion models (Kingma & Welling 2019).

More in detail, the encoder can be represented as a non linear function E_θ, where θ are its trainable parameters. The main feature of a VAE, that distinguishes it from a standard autoencoder, is that while the latter compresses an input image into a single point, z, in the latent space, a VAE encoder outputs a probability distribution for z conditioned on the input x, $q_\theta(z|x)$. This probability is set to be a multivariate Gaussian, so that:

$$\begin{aligned} \boldsymbol{\mu}, \boldsymbol{\sigma} &= E_\theta(x) \\ q_\theta(z|x) &= \mathcal{N}(z; \boldsymbol{\mu}, \mathrm{diag}(\boldsymbol{\sigma})) \,. \end{aligned} \tag{9.1}$$

This approach allows the VAE to learn a continuous latent space representation, that better approximates the underlying distribution of the data and allows meaningful sampling in the regions between the data points, which is crucial for a generative model whose aim is to produce novel images.

The second part of the VAE is the decoder, D_ϕ. It takes as input a sample from $q_\theta(z|x)$, $z \sim q_\theta(z|x)$, and tries to reverse the encoding process, producing a new image $x' = D_\phi(z)$ (and therefore typically has an architecture symmetric to that of the encoder). The goal of the decoder is thus to approximate the true probability of the data conditioned on the latent code, $p(x|z)$, with a learnable parametrised version, $p_\phi(x|z)$. For this reason, during training, the input images are passed through encoder and decoder, and the two are optimised so that the reconstructed output images are as similar as possible to the original ones. A schematic representation of the VAE used for this work is shown in Fig. 9.2.

The amount of information lost in the compression–decompression (i.e. the difference between input and output images) is the first term of the model loss function, $\mathcal{L}$. The loss function is a central aspect of any machine learning model. It represents the performance of the model and is used to adapt its parameters through a gradient descent optimisation. The similarity between a reconstructed image, x', and the original, x, is quantified statistically by the expectation value of the log-likelihood, $\mathbb{E}_{q_\theta(z|x)}[\log p_\phi(x|z)]$, which is usually approximated with the mean square error, $\langle \|x' - x\|^2 \rangle$, and is called

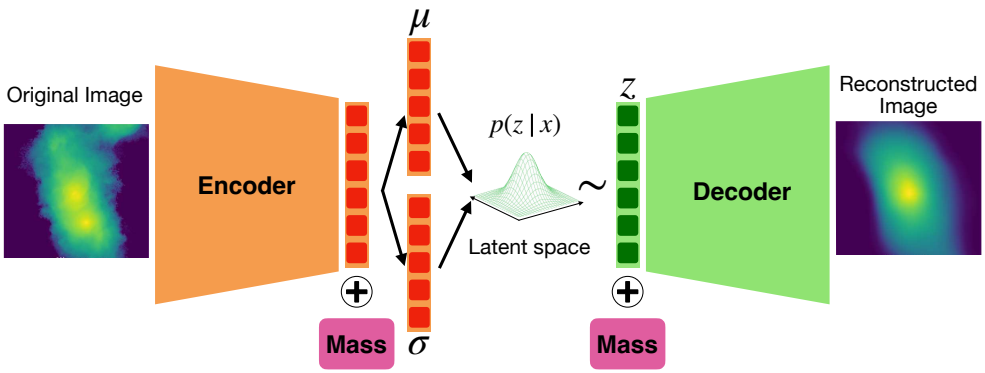

Figure 9.2: Schematic representation of the conditional variational autoencoder used in this work. During training, a cluster image from the simulation is given as input to the encoder (in orange in the picture), which compresses its information into a multivariate Gaussian distribution in the latent space, $\mathcal{N}(z; \boldsymbol{\mu}, \mathrm{diag}(\boldsymbol{\sigma}))$, taking into account the mass of the cluster (concatenated to the last layer of the encoder). A sample from this distribution, together with the cluster mass, is taken as input by the decoder, which produces another cluster image, trying to make it look as similar as possible to the input one. Once the training is completed, only the decoder is used to generate new images, by inputting samples from the unit Gaussian, $\mathcal{N}(0, \mathbb{1})$, and the desired cluster mass.

the reconstruction loss. On top of that, another term is added to the loss function in order to regularise the latent space encoding. This term measures the difference between the learned probability distribution in the latent space, q_θ, and a prior distribution (usually a standard Gaussian, $\mathcal{N}(0, \mathbb{1})$), and takes the form of a Kullback-Leibler divergence (KL Kullback & Leibler 1951), $D_{KL}(q_\theta(z|x)||\mathcal{N}(0, \mathbb{1}))$. This penalises the encoder if it produces latent space distributions that vary too much from the prior, and it is done so that the latent space has a known and meaningful structure that is easy to sample. It also helps in preventing overfitting.

The final loss function for the VAE is[1]:

$$\mathcal{L}(x, \theta, \phi) = \mathbb{E}_{q_\theta(z|x)}[\log p_\phi(x|z)] - \beta_{KL}\, D_{KL}(q_\theta(z|x)||\mathcal{N}(0, \mathbb{1}))\,, \tag{9.2}$$

where β_{KL} is a hyperparameter of the model (i.e. not trainable), and sets the relative strength of the KL divergence with respect to the reconstruction term in the loss. In standard VAEs, $\beta_{KL} = 1$ and the two terms are weighted equally. However, in some applications, the latent space can become too constrained

[1]A more detailed analytical derivation of the VAE loss function can be found in Kingma & Welling (2019).

by the KL loss term, therefore making the model less expressive and incapable of learning more complex features of the data. Tuning the β_{KL} permits a better control over the latent space representation, making it more or less flexible according to the needs (Higgins et al. 2017).

With this loss, we can find the optimal parameters for the encoder and decoder solving the optimisation problem:

$$\theta^*, \phi^* = \text{argmax}_{\theta,\phi} \sum_{i=1}^{N} \mathcal{L}(x_i, \theta, \phi) \,, \tag{9.3}$$

where the index i runs over the training dataset. This optimisation is usually performed with a gradient descent approach; in my work I used the Adam optimiser Kingma & Ba (2014).

Up to now, I treated the two modules of the VAE (encoder and decoder) in full generality, as non-linear parametric functions. In practice, the implementation of these functions depends on the specific application, as some are more adapted to treat specific kinds of data. When dealing with images, for example, it is common to use a convolutional neural network (CNN, for an introduction, see e.g. O'Shea & Nash 2015). Convolution with a filter (also called kernel) is a well known classical technique that has proven very valuable for extracting information from spatially structured data such as images (see for example the discussion of the matched multi-filter technique in Chapt. 10). Given a 2D image $\mathbf{X} \in \mathbb{R}^{\mathbf{n}\times\mathbf{n}}$, the convolution with a kernel $\mathbf{K} \in \mathbb{R}^{\mathbf{m}\times\mathbf{m}}$ (where $m < n$) returns a new image $\mathbf{C} \in \mathbb{R}^{\mathbf{n}\times\mathbf{n}}$ whose elements are:

$$C_{i,j} = \sum_{k=0}^{m} \sum_{l=0}^{m} X_{i-k,j-l} \, K_{k,l} \,. \tag{9.4}$$

This process can be visualised as a sliding window over the input image, which at each step selects an area of the size of the kernel and computes the inner product between the two. The convolution process therefore extracts spatially localised information, since its receptive field is limited by the size of the kernel. Furthermore, it is invariant to translation, in the sense that the response of the kernel to a given structure in the receptive field is the same independently of the position of the field in the original image. Different kernels can highlight different features in the receptive field, such as edges, straight or curved lines, and so on.

The basic idea of a CNN is that the image is analysed with a number of convolutional kernels which are not known a priori. Instead, their weights, $K_{k,l}$, are adapted during training, so that the model learns to recognise the relevant features of the dataset directly from the data. To add non-linearities in the network, thus increasing its expressiveness, the output of each convolution is passed through a non-linear transformation, called the activation

function[2].

To be able to learn more complex information, convolutional layers (each composed of many kernels) are stacked one after the other, so that the output of a layer becomes the input of the next one. On top of that, convolutional layers are alternated with information-compressing layers, that reduce the size of the convolved images (for example, via a max-pooling operation). In this way, convolutional kernels in the next layer will be able to access in their receptive field pixels that carry information about larger areas in the original image, thus allowing the network to learn features at different scales, capturing the full complexity of the image.

This progressive compression of the image information, typical of CNNs, is particularly adapted to the purpose of the VAE encoder, whose aim is exactly to learn the relevant information about the images, and compress it in the latent space. For the decoder, instead, a symmetric architecture is usually applied: convolutional layers are still used to process the information, that in this case goes from a low-dimensional space to a high-dimensional one. Therefore, convolutions are alternated with upsampling operations, in place of the downsampling of the encoder.

9.3 . Generating galaxy cluster images with a conditional VAE

9.3.1 . Training data

The training dataset for the generative conditional VAE consists of images of the Compton-y parameter of galaxy clusters, produced from data taken from the IllustrisTNG simulation (Nelson et al. 2019a, described in Sect. 7.1). The Compton-y parameter is a measure of the strength of the Sunyaev-Zel'dovich (SZ) effect in galaxy clusters (see Chapter 3 for more details).

Similarly to Sect. 7.1, I selected clusters at redshift $z = 0$, with masses larger than $10^{14}\,M_{\odot}$, and more than 3 Mpc away from the edge of the simulation box, for a total of 267 clusters. For each of the selected clusters, I extracted all the gas cells associated with the cluster halos. Starting from the gas quantities for each cell provided by the simulation, namely the density, electron abundance and internal energy, I computed the electron number density, n_e, and temperature, T_e. From these, I computed the electron pressure in each cell as $P_e = k_b n_e T_e$.

I then used the gas pressure information to compute images of the Compton-y signal for the clusters. I took the minimum of the potential energy as the centre of the clusters, and computed for each cluster three projected images (along the axis of the simulation box), of size 4 Mpc and resolution of 128×128

[2]Some common choices include the hyperbolic tangent function, the sigmoid function, and the Rectified Linear Unit (ReLU). For more information, see e.g. Szandała (2020).

pixels. In each pixel, the Compton-y signal was computed as:

$$y = \frac{\sigma_T}{m_e c^2} \int P_e \mathrm{d}l \approx \frac{\sigma_T}{m_e c^2} \frac{\sum_i P_{e,i} V_i}{A_{\mathrm{pix}}} \,, \tag{9.5}$$

where $P_{e,i}$ and V_i are the electron pressure and the volume of the ith gas cell, respectively, and the index i in the sum runs over all gas cells whose centre (provided by the simulation) falls inside the pixel area A_{pix}. This approximation of the line of sight integral is only applicable when the typical size of the gas cells is smaller than the size of the pixels. In principle, this is true only in the centre of the clusters, where density is higher (along with the temperature and therefore pressure). However, in the outer parts of clusters, the cell sizes are larger, but at the same time, the pressure is low, making the error introduced by this approximation negligible. With this procedure, I produced a dataset of 801 different cluster images.

To increase the size of the dataset, and to encourage the neural networks to learn the properties of the images independently of their orientation, I also performed a data augmentation step, which consists of adding to the dataset three copies of each image, obtained rotating the image in 2D by $90°$, $180°$ and $270°$, as well as four other copies, obtained flipping the original image vertically, horizontally, and along the two diagonals, for a total of eight copies for each original image. The final dataset consists then of 6408 images.

Following Rothschild et al. (2022), I also performed a log transformation of the images, in order to compress the range of the image signal and therefore add emphasis on the cluster outskirts, which are otherwise neglected by the neural network. The transformation is of the form:

$$a' = \log_{1/\epsilon} \left(\frac{a + \epsilon}{\epsilon} \right) \tag{9.6}$$

where a is the original pixel value, a' is transformed one, and ϵ is chosen to be the value of the 10th percentile of the overall pixel distribution of all the images, in this case $\epsilon \simeq 10^{-9}$. With this transformation, the range of the pixel distribution changes from roughly $[10^{-11}, 10^{-4}]$ to roughly $[0, 0.6]$.

9.3.2 . Model Architecture and training strategy

The goal of this work, as mentioned in the introduction of this Chapter, is to generate realistic images of galaxy clusters with given input properties, in this case the cluster total mass. For this reason, following Rothschild et al. (2022), I used as architecture for the generative model a modified version of the VAE, called the conditional variational autoencoder (CVAE Sohn et al. 2015). This architecture, represented schematically in Fig. 9.2, is very similar to that of a standard VAE (Kingma & Welling 2013), described in Sect. 9.2. The main difference is that, in the CVAE, the value of the cluster mass is introduced in the network in two places: at the end of the encoder, just before the compressed

image information is placed in the latent space; and at the beginning of the decoder, concatenated to the latent space sample. This encourages both parts of the model to take into account this additional information about the cluster when encoding/decoding the latent space distribution. Therefore, when using the decoder as a generator, the cluster mass can be passed directly as an additional parameter together with the random sample from the latent space, instructing the decoder to produce an image with properties compatible with a that of a cluster of the chosen mass.

As for the specific architecture of the encoder and decoder CNNs, I tested many different configurations in order to improve the results. In this section, I present the parameters of the model that provided the best results, while in the next section I discuss the various tests I performed to improve the quality of the results.

The layer structures of the encoder and decoder are summarised in Table 9.1. The dimension of the latent space is 16. For all convolutional layers, I used kernels of size 3×3 with stride[3] of $(1, 1)$, and the rectified linear unit (ReLU) activation function, defined as $f(x) = \max(0, x)$. Each convolutional layer in the encoder is followed by a max pooling layer of size 2×2, which divides the input images in squares of size 2×2, and takes the maximum value of each square, thus dividing the image size by two. On the other hand, each convolutional layer in the decoder is followed by an upsampling layer of size 2×2, which multiplies the image size by two, simply repeating four times the value of each input pixel. The last layer of the encoder and the first of the decoder are fully connected layers, in which each neuron n_i is connected to all the neurons of the previous layer, m_j, and the output of n_i is given by $n_i = \sum_j w_j m_j + b_i$, where w_j and b_i are trainable parameters.

To train the model, I used the loss function of Eq. 9.2, which combines a pixel-level mean square error as reconstruction loss and the Kullback-Leibler divergence loss, weighted by the β_{KL} parameter. The training was performed using the Adam optimiser (Kingma & Ba 2014), with an initial learning rate of 5×10^{-4} and an exponential decay with a decay rate of 0.9 over 10^4 steps. I also used a batch size of 8.

The value of β_{KL} has been the subject of extensive testing, in order to find the best balance between the two terms of the loss function (see Sect. 9.3.3). Furthermore, to improve the results, I adopted a strategy called β-annealing (Bowman et al. 2015), which consists of progressively increasing the value of β_{KL} during training, starting from 0 and increasing it up to a maximum value, which I set to $\beta_{KL} = 10^{-3}$. This strategy has been proposed to mitigate the

[3]The stride is an hyperparameter of convolutional layers which refers to how much the kernel slides over the input image. When the stride is $(1, 1)$, the convolution happens as in Eq. 9.4; using different values of the stride, (a, b), equates to multiplying the summation indices, $k' = k \times a$ and $l' = l \times b$, in Eq. 9.4.

Table 9.1: Architecture of the encoder and decoder of the CVAE used in this work. Each row represents a layer (or group of layers) in the network, and the information flows from top to bottom. The output dimensions in parenthesis are the dimensions of the convolved images (one for each convolutional kernel), while the numbers without parenthesis represent the full output of the layer. The output of the last layer of the encoder represents the means and standard deviations of the latent space distribution, $\mathcal{N}(z; \boldsymbol{\mu}, \mathrm{diag}(\boldsymbol{\sigma}))$.

	Kind of layer	# of kernels/ neurons	Output dimension
Encoder	Convolution + max-pooling	64	(64×64)
	Convolution + max-pooling	64	(32×32)
	Convolution + max-pooling	128	(16×16)
	Convolution + max-pooling	128	(8×8)
	Convolution + max-pooling	256	(4×4)
	Convolution + max-pooling	256	(2×2)
	Flatten + Concatenate mass	0	1025
	Fully connected	32	$16 + 16$
Decoder	Sampling + Concatenate mass	0	17
	Fully connected	1024	1024
	Reshape	0	(2×2)
	Convolution + upsampling	256	(4×4)
	Convolution + upsampling	256	(8×8)
	Convolution + upsampling	128	(16×16)
	Convolution + upsampling	128	(32×32)
	Convolution + upsampling	64	(64×64)
	Convolution + upsampling	64	(128×128)
	Convolution	1	(128×128)

known problem of VAE optimisation related to undesirable local minima in the loss function, in which the KL divergence goes quickly to 0 and prevents the model from learning to reconstruct the images. Thus, by starting the training with low β_{KL} encourages the model to first learn the relevant features of the training data, and only later to organise the latent space according to the prior. In this model, I used an exponential increase of β_{KL} over 10^4 training steps (corresponding to roughly 13 epochs).

9.3.3 . Results

In order to assess the effectiveness of the model training and thus the quality of the generated images, the first level of evaluation is a visual inspection. In the best case scenario, it should be almost impossible to tell if

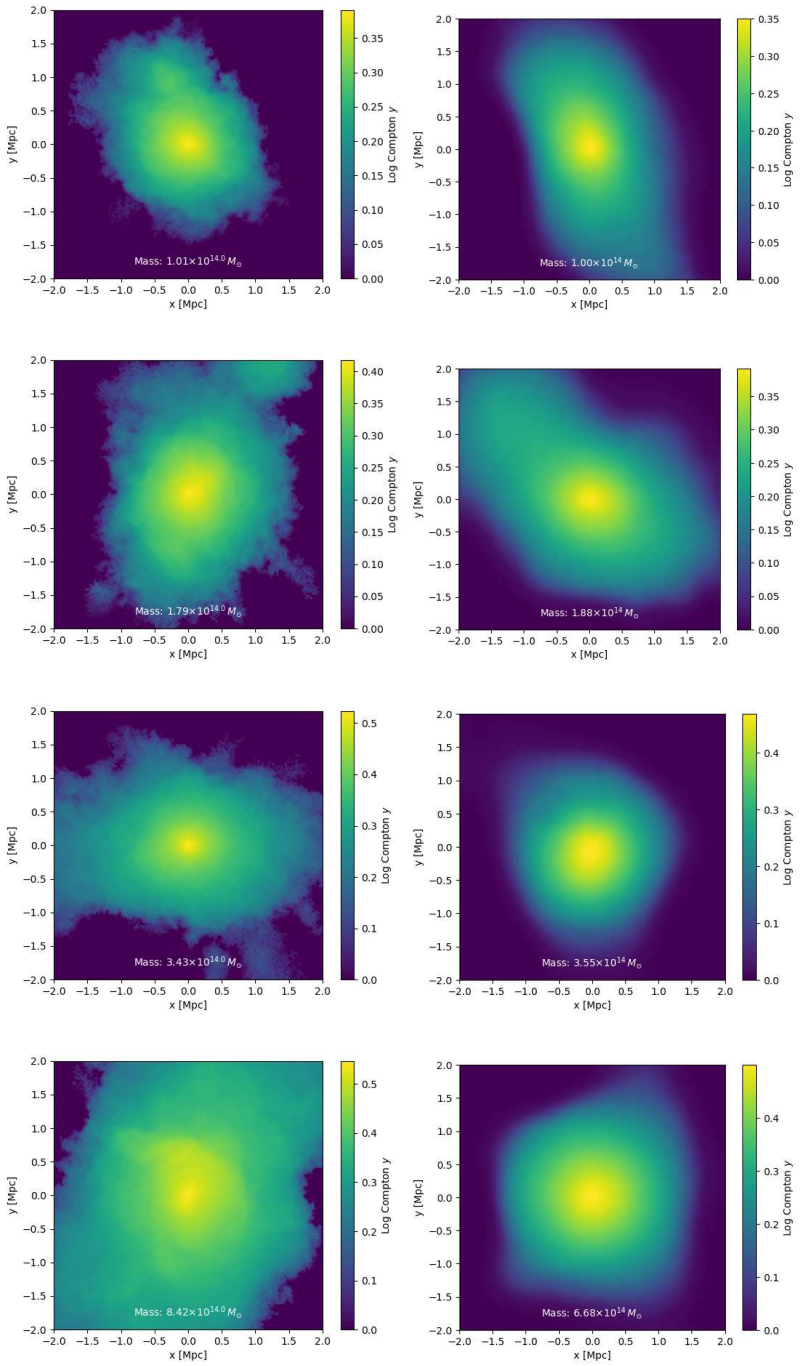

Figure 9.3: Examples of cluster images from the training set (*left*), and generated images from the CVAE, with roughly similar masses (*right*). All images are in logarithmic scale, transformed according to Eq. 9.6.

a given image was taken from the simulation set or if it has been generated by the model. On top of that, the population of the generated images should also have the same statistical properties as the original simulated population. Therefore, I compared statistically the simulation images with a set of generated images with a similar mass distribution, testing the relation of the integrated Compton-y signal with mass, and the morphological properties of the two sets, represented by two estimators: the β parameter (introduced in Sect. 7.1.3), and the concentration parameter.

Visual inspection

In Fig. 9.3, we see some examples of images generated with the CVAE described in the previous Section (on the right), compared with training images of similar mass (on the left). We see clearly that the two sets of images are different. At large scales, we see that the model is able to reproduce some of the anisotropic structures on large scales, and thus go beyond the spherical and even the triaxial approximations. Nonetheless, we notice how the generated images are much smoother than the simulated ones, showing a certain lack of structures at small scales, especially in the outskirts, but also near the cluster cores.

This "blurriness" effect in the generated images is a well-known problem of VAEs (Kingma & Welling 2019), and it is, to some extent, unavoidable. However, similar works (like Rothschild et al. 2022) have been able to obtain results that seem to suffer less from this issue. This points to the general difficulty and lack of robustness in training deep machine learning models, as they can be very sensitive to the details of the training, and often require a high amount of trial and error.

It has been suggested by Kingma & Welling (2019) that a way to counter the "blurriness" issue is to improve the flexibility of the model in order to improve its expressivity. I tested many different possible solutions to this problem, but I was unable to improve the appearance of the generated images beyond what shown in Fig. 9.3.

A first set of tests were focused on the relative strength of the two terms of the loss function (the reconstruction loss ad the KL divergence), controlled by the β_{KL} parameter. As discussed in the previous Section, a possible issue in training a VAE is the so-called vanishing KL problem, that happens when the optimisation of the KL divergence dominates over the reconstruction loss. Since the two terms of the loss function are in competition, a dominating KL divergence makes it less advantageous for the model to improve the quality of the images, favouring the regularisation of the latent space distribution. I found this to be a problem for $\beta_{KL} = 1$, leading to mostly circular images, so I lowered the value of β_{KL} while testing the performance of the model at each step. I found that $\beta_{KL} = 10^{-3}$ produces the best results. For higher

values the KL divergence ended up dominating the loss, producing smooth and elliptical images, while for lower values the reconstruction improved, but at the expense of the "goodness" of the latent space distribution, which was not suitable for generation anymore.

Another path of exploration was directed to the dimension of the latent space. It is known that a low-dimensional latent space might limit the expressivity of the model, since the image information has to be compressed into a smaller set of numbers. On the other hand, a too large number of dimensions might be an obstacle for the generation, since the training images will be sparser, preventing the model from learning to interpolate between then, and consequently generate new samples. I tested different numbers of dimensions, ranging from 2 to 128, finding the values of 16 and 32 to be approximately equal in terms of performance. I chose 16 for the final model.

I also explored modifications in the CNN architectures, in particular regarding the number of convolutional kernels. I modified the number of kernels in each layer, dividing and multiplying it by two, without finding appreciable differences in the results. I also tested a network architecture in which two convolutional layers were stacked before each pooling layer, in the attempt to encourage the model to learn more complex features at each stage, but this test also proved ineffective in achieving sharper images.

Population comparison

In order to test whether the generated images reproduce well the original data at the population level, I generated a sample of images of the same size and with similar mass distribution as the simulated set, and compared the distributions of physical and morphological parameters of the cluster images for both sets.

To obtain a generated population with roughly the same mass distribution as the simulated set, I created a random sample of 800 masses, selected with replacement from the simulation set. Then, to avoid inputting in the generator the same mass values on which the model has been trained on, I added Gaussian noise to the logarithm of the masses (with $\sigma = 0.05$), following (Rothschild et al. 2022).

To test whether the model has been able to learn the physical properties of clusters from the training images, I used the integrated Compton-y parameter:

$$Y = \int y \, dA \propto \int n_e T \, dV \, , \tag{9.7}$$

which is proportional to $M_{\rm gas}T$ and used as a proxy for the total cluster mass (see e.g. Kravtsov & Borgani 2012). Here, I computed the integrated y parameter inside a circle of radius R_{200}, Y_{200}, for both sets of images. By comparing the dependence of Y_{200} on the cluster mass, M_{200}, between the "true" and

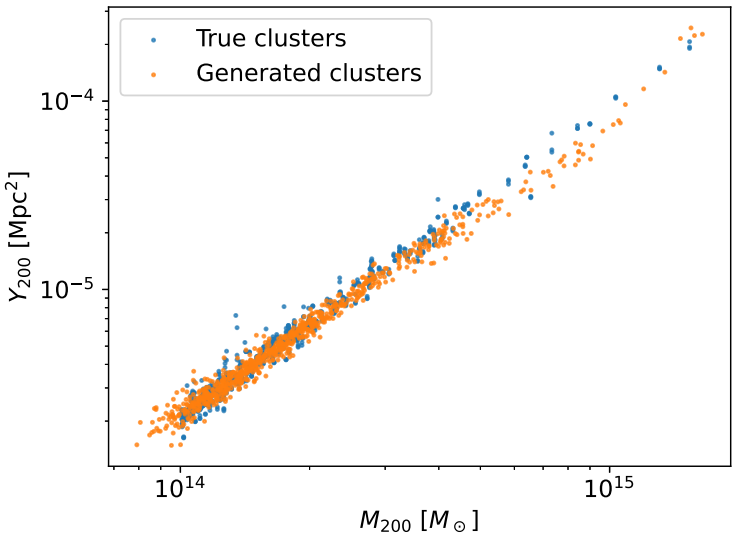

Figure 9.4: Integrated Compton-y parameter, Y_{200}, as a function of cluster mass, M_{200}, for the simulated cluster images (in blue), and the generated cluster images (in orange).

generated sets, we can see whether the model was able to learn this dependence correctly, and thus if the images produced with a given input mass really reproduce clusters with that mass.

The $Y_{200} - M_{200}$ relations for the training and generated datasets are shown in Fig. 9.4. From this Figure, we see that at low mass the Y_{200} distribution of the generated images follows pretty well that of the simulated clusters. Conversely, for masses higher than $\sim 5 \times 10^{14} M_\odot$ the two distributions diverge, and the Y_{200} of the generated images tend to be underestimated compared to the simulated ones. This is probably due to the low number of clusters in this mass range in the training population, which do not constitute a sufficiently large sample for the network to learn to model efficiently this mass range. Indeed, this is a rather common issue in machine learning applications, where samples that are at the edges of the learned distribution tend to be biased towards the bulk values (see e.g. Green et al. 2019; Ntampaka & Vikhlinin 2022).

Beyond reproducing the integrated y signal, one of the main goals of this work is to be able to generate cluster images with realistic morphologies. To test the morphology of generated images compared to the simulated set I used the multipolar ratios presented in Sect. 6.1, in particular the sum of the first 4 orders, β, as an estimator of the overall departure from circular symmetry of the images (as in Sect. 7.2.2). I computed the β parameter from the images selecting the pixels inside R_{200}. In addition, to evaluate the signal dis-

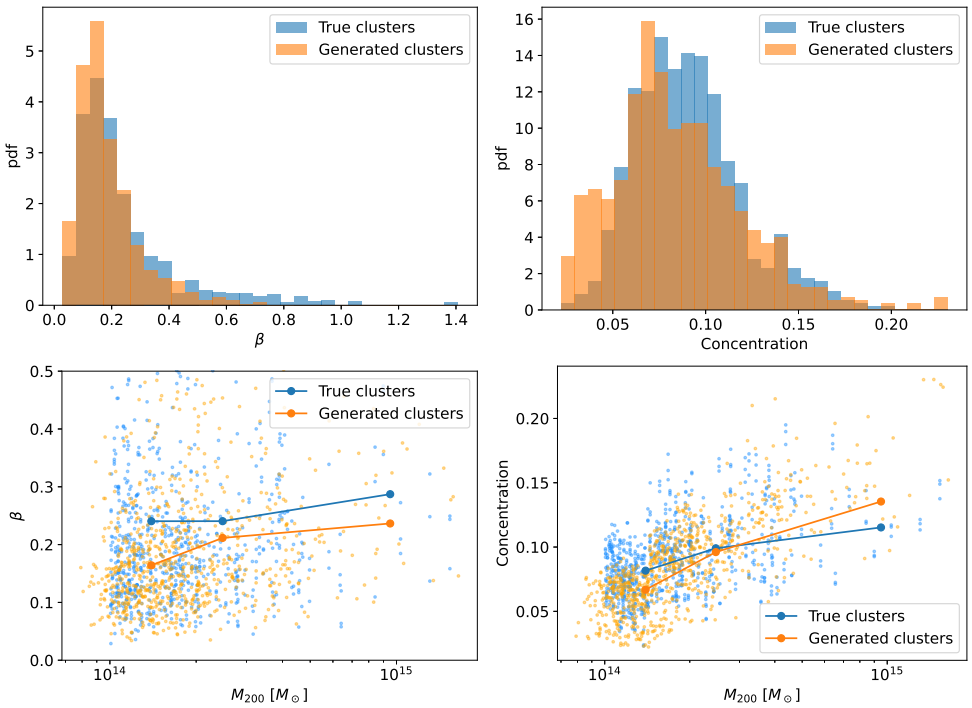

Figure 9.5: Distribution of morphological parameters of cluster images, for the full population (top panels), and as a function of the cluster mass (bottom panels). Left panels: distribution of the β parameter, quantifying the overall level of anisotropy of the images inside R_{200}. Right panels: distribution of the concentration parameter, see definition in the text. In all panels: in blue, simulated images; in orange, generated images.

tribution in the radial direction, I compute the concentration parameter (similar to Rasia et al. 2013; Rothschild et al. 2022), defined as the ratio between the total signal inside $0.1R_{200}$ and inside $1R_{200}$,

$$C = \frac{\sum I(r \leq 0.1R_{200})}{\sum I(r \leq R_{200})} \tag{9.8}$$

where I is the image, and the sum is over the pixel values.

In Fig. 9.5, I show the comparison of the morphological estimators, β and the concentration parameters, between the generated and simulated sets of images. In the top panels, the histograms of the distributions, and in the bottom panels the mass dependence, with the average parameter value in three mass bins.

Starting from the β parameter distribution (top left panel), we see that, although the generated and simulated distributions have similar distributions around the peak, the generated images have in general lower values of β, which means that they are more spherical than the simulated images. This is even more apparent when focusing on the high β values: we see that the sim-

ulated images have a large tail, extending to $\beta \sim 1$ and even beyond, which represents a sizeable population of highly anisotropic images. This is much less present in the generated population, whose β distribution shows a less important high-β tail, with a maximum value around $\beta \sim 0.7$. This is consistent with the visual appearance of the generated images (Fig. 9.3), which are not only smoother and lack small-scale structures, but also appear more spherical than their simulated counterparts. This again points to a failure of the model to learn the more complex structures of the training set, not only at small scales but also at the large angular scales probed by the first 4 orders of the multipolar decomposition. Looking at the dependence of β on the cluster mass (bottom left panel), we see again that the generated and simulated images present distinctive differences. At first, we see that the generated images are on average more spherical (lower β) than the simulated ones at all masses, confirming the overall difficulty of the model to produce complex morphologies. Observing the trend of the average β with mass, we see for the simulated images a mildly increasing trend, consistent with the results of Sect. 7.2.2. For the generated images, we still see an increasing trend, though at low masses the model strongly favours more spherical morphologies compared to higher masses. This shows that the bias in the average departure from spherical symmetry for generated images is mass-dependent, which further compromises their potential use as a representative sample of realistic clusters.

Moving the attention to the concentration parameter (right panels of Fig. 9.5), we again see some differences between the simulated and generated distributions. Looking at the concentration histograms (top right panel), it can be noticed that the distribution of concentrations of the generated set is wider than that of the simulated set, with a more important low-concentration tail and lower average value, but also a high-concentration tail that extends to higher values than the simulated distribution. Even more striking is the correlation between concentration on cluster mass (bottom right panel). It is evident that while there is an increase with mass in signal concentration in the simulation set, this is much more pronounced in the generated set. Generated images of low mass clusters exhibit lower concentrations than the simulated ones, while generated images of high mass clusters demonstrate higher concentrations.

In summary, by studying the distribution of the physical, Y_{200}, and morphological parameters, β and concentration, we see that the statistical properties of the generated images differ noticeably from those of the simulated images. This means that the generative model has learnt only incomplete information about the cluster population and the intrinsic relations between the cluster masses and their appearances, and is therefore unable to gener-

ate realistic cluster images.

9.4 . Conclusions and perspectives

In this Chapter, I introduced generative models as a possible way to accelerate or bypass the computation of cosmological simulations.

I trained a conditional variational autoencoder with the goal of generating realistic images of the Compton-y parameter from galaxy clusters, conditioned on the cluster mass. I compared the generated cluster images with the ones used for training, and found the following results:

— Visual inspection of the generated images reveals that, at large scales, these present anisotropic structures that visually resemble those of the training set. In general, though, generated images are smoother and present less complex structures at small scales, both in the cores and in the outskirts.

— The model is able to reproduce the relation between the integrated y signal, Y_{200}, and the cluster mass, for masses below $\sim 5 \times 10^{14} M_{\odot}$. For higher masses, the model tends to underestimate the Y_{200} signal.

— Looking at the distribution of the morphological estimators, β and concentration, the generated and simulated populations differ significantly. The average departure from spherical symmetry is lower for generated images than for simulated ones, in particular at low masses. The highly-anisotropic tail of the simulated distribution is underrepresented in the generated set. The generated images have a wider concentration distribution, which exhibits a stronger dependence on the cluster mass than the simulated set.

In light of these findings, it can be concluded that although the generated images are not of sufficient quality to be used in place of high-resolution cosmological simulations of galaxy clusters, they still exhibit more realistic morphologies compared to analytically generated images of perfectly spherically symmetric clusters. Therefore, they may be used in low-resolution cases, for example where the smoothing effect due to the instrument's beam is larger than the inherent "blurriness" of the images.

For the future, it might be possible to obtain better results by complexifying the general architecture of the network, which I have not attempted in this work. Some possibilities include: modifying of the structure of the loss function with the *free-bits* method (Kingma et al. 2016), which forces the network to store a given amount of information in each dimension of the latent space; introducing normalising flows to improve the flexibility of the latent

space representation, while at the same time keeping an easy-to-sample overall distribution (Jimenez Rezende & Mohamed 2015; Kingma et al. 2016; Euclid Collaboration et al. 2022); or using multiple latent variables (Kingma & Welling 2019).

Part III

Galaxy cluster detection: The selection function

10 - Galaxy cluster detection with the Planck satellite

In this third part of the Thesis, I focus on the detection of galaxy cluster, in particular on how the assumed cluster model affects the detection process. As outlined in Chapter 5, detecting a galaxy cluster is essentially a process of identification of specific signatures in the data associated with the presence of a cluster. This implies the assumption of a cluster model, to be compared with the data. This model should represent as well as possible the average properties of the galaxy cluster population. If this is not the case, at least for a sizeable sub-class of clusters, it could impact the efficiency of the detection process for such objects, and thus lead to a biased catalogue. This is true, for example, for X-ray surveys, which are known to detect preferentially clusters that are dynamically relaxed, with a bright cooling core (see e.g. Eckert et al. 2011). These biases need to be accurately determined and modelled in the *selection function* when using clusters for cosmological analyses, otherwise the bias can be propagated from the cluster catalogue to the cosmological parameters.

In my work, I focused on the galaxy clusters detected via the Sunyaev-Zel'dovich effect by the *Planck* satellite, in particular taking the *Planck* PSZ2 `MMF3` cosmological catalogue[1] (Planck Collaboration et al. 2016d) as reference. *Planck* (Planck Collaboration et al. 2016a) was a space mission aimed at observing the full sky in nine frequency bands (from 25 to $1000\,\mathrm{GHz}$), primarily to observe the CMB. It observed for approximately 50 months, from 2009 to 2013. In the frequency range covered by the High Frequency Instrument ($100 - 857\,\mathrm{GHz}$, Planck Collaboration et al. 2014f), *Planck* is able to recognise the spectral distortion of the SZ effect produced by the hot gas in galaxy clusters. The cluster detection in *Planck* is done using a matched multi-frequency filter (MMF) algorithm (Herranz et al. 2002; Melin et al. 2006). This method, presented in Sect. 10.1, has been used to detect galaxy clusters not only by *Planck* (Planck Collaboration et al. 2011, 2014c, 2016d), but also by other ground-based experiments like the South Pole Telescope (SPT, Williamson et al. 2011; Bleem et al. 2015, 2024), and the Atacama Cosmology Telescope (ACT, Hilton et al. 2021).

This method is designed to combine the knowledge of the SZ signal, namely its spectral signature, with a model of the galaxy clusters' spatial characteristics, to produce an optimal filter that returns the maximal S/N in the presence of a galaxy cluster. While the spectral shape of the SZ signal is well known

[1] This sample contains 439 clusters, with masses $\in [0.8, 14.7] \times 10^{14}\,M_\odot$ and redshifts $\in [0.01, 0.97]$, detected in a survey area covering 65% of the sky.

and universal (that is, for a thermal gas in the non-relativistic regime), some approximations must be made for the spatial filter. The usual approach is to assume spherical symmetry for clusters, and to model the radial pressure profile according to the average profile of observed clusters. However, as I discussed extensively in the second part of this Thesis, galaxy clusters are generally not spherical, but present complex morphologies due to anisotropic accretion, mergers, and astrophysical processes such as shocks and feedback mechanisms. The variety of morphological features observed in galaxy clusters often prevents a precise match between the cluster SZ signal and the spatial template employed, which, in principle, affects the detection performance. In practice, this effect is reduced by the smoothing induced by the instrument's beam, which tends to symmetrise the signal, especially for clusters with scales comparable to or smaller than the beam size. On top of that, it is clear that not all clusters have exactly the same pressure profile, but there is a certain scatter around the average profile. Furthermore, different studies using different cluster samples found slightly different average profiles (e.g. Arnaud et al. 2010; Planck Collaboration et al. 2013a; Pointecouteau et al. 2021; Tramonte et al. 2023; Melin & Pratt 2023). These aspects impact the efficiency of the detection process, and quantifying their effect on the selection function is the main goal of the analysis presented in Chapter. 11.

10.1 . The Matched Multi-frequency Filter detection algorithm

In this Section, I present briefly the main concepts of the MMF algorithm (Melin et al. 2006), and explain its functioning on the *Planck* data. In particular, I focus on an implementation of the algorithm that I used in my work, which is very similar to the one used to construct the *Planck* cosmological cluster catalogue, called `MMF3` in Planck Collaboration et al. (2011, 2014c, 2016d).

We can model the maps of the cluster emission at each observed frequency ($\nu = 100, 143, 217, 353, 545, 857\,\mathrm{GHz}$) as the vector $\boldsymbol{m}(\boldsymbol{x})$:

$$\boldsymbol{m}(\boldsymbol{x}) = y_0\, \boldsymbol{t}_{\theta_c}(\boldsymbol{x}) + \boldsymbol{n}(\boldsymbol{x})\,, \tag{10.1}$$

where y_0 represents the SZ signal strength, and $\boldsymbol{n}(\boldsymbol{x})$ is the noise (which includes both instrumental noise and astrophysical contaminants). $\boldsymbol{t}_{\theta_c}(\boldsymbol{x})$ represents the distribution of the SZ signal from a cluster with characteristic angular radius θ_c, and it is composed of a spatial template τ_{θ_c}, which is convolved by the instrument's beam at each frequency, $b_i(\boldsymbol{x})$ (Planck Collaboration et al. 2014g), and multiplied by the spectral dependence of the SZ signal, $\boldsymbol{j_\nu}$, to give $\boldsymbol{t}_{\theta_c}(\boldsymbol{x})_i = j_\nu(\nu_i)[b_i * \tau_{\theta_c}]$.

The idea of the MMF method is to build a filter $\boldsymbol{\Psi}_{\theta_c}(\boldsymbol{x})$ that, when convolved with the cluster signal, returns an estimate of y_0:

$$\hat{y}_0 = \int d\boldsymbol{x}\, \boldsymbol{\Psi}^T_{\theta_c}(\boldsymbol{x}) \cdot \boldsymbol{m}(\boldsymbol{x})\,. \tag{10.2}$$

By requiring this estimate to be unbiased $\langle \hat{y}_0 \rangle = y_0$, and the variance to be minimised, one obtains that the filter must be of the form (expressed in Fourier space):

$$\boldsymbol{\Psi}_{\theta_c}(\boldsymbol{k}) = \sigma^2_{\theta_c}\, \boldsymbol{P}^{-1}(\boldsymbol{k}) \cdot \boldsymbol{t}_{\theta_c}(\boldsymbol{k}) \tag{10.3}$$

where $\boldsymbol{P}(\boldsymbol{k})$ is the noise power spectrum matrix, whose components in frequency space are given by $\langle n_{\nu_i}(\boldsymbol{k}) n_{\nu_j}(\boldsymbol{k'}) \rangle = P_{i,j}(\boldsymbol{k})\, \delta(\boldsymbol{k} - \boldsymbol{k'})$, and σ_{θ_c} is the variance of the noise filtered at the scale θ_c,

$$\sigma_{\theta_c} = \left[\int d\boldsymbol{k}\, \boldsymbol{t}_{\theta_c}(\boldsymbol{k})^T \cdot \boldsymbol{P}^{-1} \cdot \boldsymbol{t}_{\theta_c}(\boldsymbol{k}) \right]^{-1/2}. \tag{10.4}$$

In *Planck* , the noise power spectrum $\boldsymbol{P}(\boldsymbol{k})$ is approximated by computing directly the power spectrum of the maps, considering the SZ signal negligible with respect to the noise sources. In a recent approach, Zubeldia et al. (2022) proposed to improve the noise estimation with an iterative approach.

The spatial cluster template, τ_{θ_c}, used in *Planck* is constructed from a spherically symmetric cluster model, whose pressure distribution follows a generalised Navarro-Frenk-White (gNFW) (Navarro et al. 1997; Nagai et al. 2007; Arnaud et al. 2010) profile. The shape of this pressure profile is defined by the parametric form:

$$p(\rho) \propto \frac{1}{(c_{500}\rho)^{\gamma}[1 + (c_{500}\rho)^{\alpha}]^{(\beta-\gamma)/\alpha}} \tag{10.5}$$

where $\rho = r/R_{500}$ is the radius in units of R_{500}, and $[c_{500}, \alpha, \beta, \gamma]$ are the parameters of the profile. The parameters used are those from Arnaud et al. (2010), which combined a sample of 31 clusters observed in X-rays with three sets of numerical simulations, and fitted the average pressure profile with a gNFW model, finding $[c_{500}, \gamma, \alpha, \beta] = [1.177, 0.308, 1.051, 5.491]$. The cluster template is then computed integrating the pressure profile along the line of sight. The only free parameter of the template is the cluster radius, R_{500}, or more properly, its equivalent in angular coordinates, θ_{500}. Therefore, from this template, the MMF filters are constructed varying the cluster size on a grid of 40 logarithmically spaced points from $\theta_{500} = 1.059$ to $41.195\mathrm{arcmin}$.

To detect galaxy clusters, the MMF algorithm takes as input the six full-sky *Planck* HFI frequency maps. These maps are divided into 546 overlapping square patches with side length of 10°. Then, each patch is convolved by the filters, Ψ, at the different cluster scales. The peaks in these filtered maps with a signal-to-noise ratio (S/N) above 3 represent the locations of the candidate cluster detections. The size of a cluster is taken to be the scale of the filter

that maximises the S/N at the detected cluster location. The integrated cluster signal within $5 \times R_{500}$, Y_{5R500}, is extracted from the filtered map at the peak position, and computed as:

$$Y_{5R500} = \int_{\theta<5\theta_{500}} d\theta \, \tau_{\theta_{500}}(\theta) \, . \tag{10.6}$$

Finally, a single full-sky catalogue is constructed, by merging the single-patch detections and discarding the ones with lower S/N that fall within the angular radius, θ_{500}, of another cluster. The final catalogue of detected sources therefore contains the position of the clusters, their S/Ns, and the estimated θ_{500} and Y_{5R500}. In addition, the algorithm returns the noise estimate, $\sigma_{\theta_{500}}$ for each patch and filter size.

10.2 . The selection function

As explained in Chapter 4, one of the key ingredients for cosmological analyses with galaxy cluster number counts is the survey's selection function. This function describes the relationship between the detected objects and the complete underlying population in the survey area, and it thus depends on the characteristics of the survey and of the detection strategy. It can be divided into two separate functions: the purity, which is the probability that a given detection corresponds to a real cluster; and the completeness, the probability of a real cluster to be detected by the survey.

In the context of cosmological analyses with cluster counts, the selection function provides an estimate of the fraction of detected clusters relative to the total number of clusters present in the sky, as a function of the cluster observables. This is crucial information when comparing the observed number of clusters with the predicted number from theoretical models. It is therefore essential to provide an accurate characterisation of this function in order to avoid any potential biases in the constraints placed on cosmological parameters.

The selection function of the *Planck* SZ cluster survey has been investigated by the *Planck* Collaboration, and the results have been detailed in the various catalogue releases (Planck Collaboration et al. 2011, 2014c, 2016d). In particular, it has been shown that the purity of the cosmology catalogue is very high, exceeding 99.8%. For this reason, in this Thesis I focus exclusively on the on the completeness function.

As mentioned, the completeness is the probability that a cluster with given "true" observables (in this case Y_{5R500} and θ_{500}), at position $\boldsymbol{x}$ in the sky, gets detected, given the survey and detection method (*Planck* and MMF, respectively). To estimate this function, a first approximation consists of assuming Gaussian noise on the Compton-y measurements. In this case, the complete-

ness can be determined analytically in each patch i, and takes the form of an error function:

$$\chi(Y_{5R500}, \sigma_{\theta_{500},i}) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{Y_{5R500} - q\,\sigma_{\theta_{500},i}}{\sqrt{2}\,\sigma_{\theta_{500},i}}\right)\right]. \tag{10.7}$$

However, the non-Gaussian nature of the noise (due to astrophysical contaminations), combined with the differences between the real clusters and the detection template, can make the true completeness deviate from this approximation.

To include these effects into the completeness estimation, a more direct Monte Carlo approach is used. This method consists of injecting synthetic SZ signal from clusters directly in the sky maps, and running the MMF algorithm to compare how many of these clusters are detected (Melin et al. 2005; Planck Collaboration et al. 2014c, 2016d). With full knowledge of the true (injected) cluster population, it becomes possible to evaluate the performance of the detection algorithm run on these mock maps. The objective of this strategy is to replicate the conditions of the actual detection task as closely as possible in order to obtain an accurate estimate of completeness, including all possible contaminant effects that may be challenging to model analytically (e.g. algorithmic effects, non-Gaussian noise, cluster morphology, etc.). This method enables the estimation of the completeness by simply taking the ratio of the detected clusters over the injected ones, in bins of Y_{5R500} and θ_{500}.

This approach was used in *Planck* (Planck Collaboration et al. 2016d) to estimate the completeness of the cluster catalogues. For the injection, the clusters were assumed to be spherically symmetric, with a pressure profile similar to the gNFW profile used in the detection template (Arnaud et al. 2010). To account for individual differences in the pressure distribution, the injected clusters' profiles were taken from simulated clusters from the Cosmo-OWLS simulation (Le Brun et al. 2014). The completeness obtained from this method was then compared with the ERF estimate, finding substantial agreement between the two, for the cosmological sample. Therefore, the ERF completeness was incorporated in the baseline cosmological analysis of cluster number counts with the *Planck* SZ catalogue (Planck Collaboration et al. 2014d, 2016f).

11 - Characterising the completeness function of *Planck* clusters

The content of this Chapter is based on Gallo et al. (2024b).

As discussed in the previous Chapter, an accurate characterisation of the selection function, and in particular the completeness, of a galaxy cluster sample is essential to perform a cosmological analysis.

For the *Planck* cluster catalogue, as mentioned above, the completeness was estimated via Monte Carlo injection and compared with the ERF estimate. The two were found to be in agreement with each other (Planck Collaboration et al. 2014c, 2016d), so the latter was used for the cosmological analyses (Planck Collaboration et al. 2014d, 2016f).

In my work, I took these results as a starting point, and analysed the completeness focusing on the case in which the cluster model assumed in the detection algorithm is different from the true cluster signals. This situation is to be expected, to some extent, given that a template is by necessity a simplification constructed to match the 'average' features of a selected cluster sample as closely as possible.

A first clear difference between real clusters and the detection template is cluster morphology. Departure from spherical symmetry can indeed bias cluster detection, and therefore the effect of realistic morphology needs to be tested while characterising the completeness. A second difference may be attributed to the pressure profile. First, not all clusters have exactly the same profile. Profile variations are observed, due for example to the clusters' dynamical state, and these induce a scatter around the average profile of the population. Moreover, the average profile may also differ from the one assumed in the template. This may occur, for instance, if the profile assumed in the detection was derived from a biased sample of clusters.

Some tests were conducted to investigate these aspects, as detailed in Planck Collaboration et al. (2016d). To investigate the effect of cluster morphology, a modest sample of hydrodynamically simulated clusters was used, with a fixed angular scale larger than the *Planck* beam, where the effect of cluster morphology is supposed to be most relevant. No significant difference was found in the completeness when either realistic or spherical morphologies were used. With regard to the profile scatter, the results presented in Planck Collaboration et al. (2016d) demonstrate that the completeness computed from cluster images with pressure profiles scattered around the assumed detection profile is generally consistent with the analytical completeness estimate, with a widening effect in the completeness drop-off.

In this work I studied these effects in a comprehensive way, analysing the impact of having an imperfect cluster model as template for the matched filter detection technique. To do this, I extracted a sample of clusters from the IllustrisTNG hydrodynamical simulation (Sect. 7.1) in a somewhat agnostic manner. I extracted galaxy clusters from the simulation at different redshifts, and produced images of their SZ signal, based on each cluster's redshift and gas distribution. In this manner, when computing the completeness via Monte Carlo injection, any potential redshift dependence of the cluster properties and dynamical states, in addition to the impact of cluster morphologies, were automatically incorporated to increase the realism of the completeness estimation.

11.1 . Cleaned frequency maps

To construct the sky maps into which the cluster images were injected, I started from the six *Planck* HFI frequency maps from the second data release Planck Collaboration et al. (2016b). These maps are given in `HEALPix` pixelisation scheme (Górski et al. 2005), with $N_{\rm side} = 2048$. I chose to use real *Planck* maps (following Planck Collaboration et al. 2014c, 2016d) to ensure the most realistic setting possible for the completeness analysis, including all sources of noise and contaminations present during the real detection process.

Given the fact that the MMF algorithm estimates the noise directly from the input maps, the injection of simulated cluster signals in addition to the real ones already present could potentially alter the properties of the noise, thus affecting the detection. For this reason, I subjected the maps to a cleaning process with the purpose of removing the SZ signal from the real clusters. The process of removing a given cluster from the maps is the following: starting from the cluster's integrated SZ flux and angular scale, $(Y_{5R500}, \theta_{500})$, a circular image of the cluster SZ emission at the frequencies of the different maps is computed using the pressure profile from Arnaud et al. (2010) (in the same way as in Section 11.2.2). This cluster emission is then subtracted from the maps at the position of the original detected cluster. The differences between the circular template and the real cluster signal leave a residual contribution, but its impact on the noise estimation is certainly smaller than that of the original cluster and is therefore considered negligible.

The cleaning process is conducted in two steps: the first step consists in removing all the clusters from the PSZ2 catalogue, obtaining a first cleaned version of the maps. In the second step, the MMF detection is run on the new maps to check for additional signals identified as clusters beyond those contained in the PSZ2 catalogue, with a lower limit of 4.25 in signal-to-noise ratio (S/N). All new detections obtained in this way are then also removed from the frequency maps, which are now, in principle, free from any relevant

SZ source down to S/N ~ 4.5. These are the final cleaned maps we use for the completeness analysis.

Associated with the clean frequency maps, I create a mask that covers the emission from the Galaxy and the Magellanic Cloud, as well as an area of five times the beam size around point sources (from PCCS2 catalogue Planck Collaboration et al. 2016c) and regions of CO emission Planck Collaboration et al. (2014b). The final unmasked sky fraction is about 78%.

11.2 . Cluster SZ images

In this Section, I describe the creation of the galaxy cluster images used to estimate the completeness function. One set of realistic cluster images was computed from the simulated clusters from the IllustrisTNG simulation, described in Sect. 7.1, in a similar way to what described in Sect. 9.3.1. Other sets of images were constructed from spherically-symmetric model with a gNFW pressure profile, where each set is built from a different set of parameters for the profile.

11.2.1 . Simulation images

To create a set of realistic images of the SZ effect from galaxy clusters, as it would be seen by *Planck* , I used the galaxy clusters of the hydrodynamical cosmological simulation TNG300, part of the IllustrisTNG simulation suite (Nelson et al. 2019a, see Sect. 7.1). Snapshots of the simulation are available at different redshifts, each accompanied by a halo catalogue, built with a FoF algorithm (Davis et al. 1985). These catalogues list the positions of the halos (identified by the particle with the minimum gravitational potential energy), their masses M_{500}, and radii R_{500}, among other quantities.

Since the goal of this study is to determine the detection performance of the MMF algorithm, as represented by the completeness function, the sets of clusters used to estimate it need to probe the region in mass and redshift where the detection becomes increasingly more difficult, and the completeness goes from 1 to 0. To get an idea of where is this region, I observed that, for the *Planck* PSZ2 cluster sample (Planck Collaboration et al. 2016d, Fig. 26), the lowest detected cluster mass increases with redshift, from $\sim 10^{14}\,M_\odot$ at $z \sim 0.05$ to $\sim 4 \times 10^{14}\,M_\odot$ at $z \sim 0.4$, and then remains almost constant. Therefore, the cluster sets for this study must extend to masses sufficiently below the limit of the *Planck* detections to be able to study the completeness down to 0. At the same time, it is not advisable to include too many clusters with almost zero probability of being detected, as those would not add any information to the completeness.

Consequently, I selected halos from the simulation with a redshift-dependent lower mass limit of $M_{500} \geq M_{\min} = \frac{6}{5}\left(4z + \frac{7}{15}\right) \times 10^{14}\,M_\odot$ within the redshift

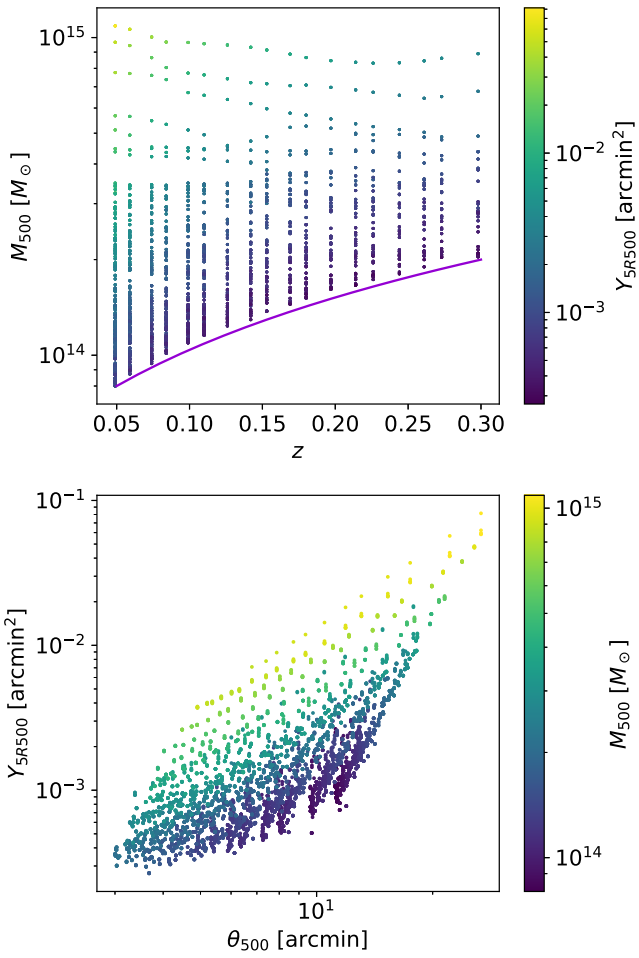

Figure 11.1: Distribution of the galaxy clusters selected from the TNG300 simulation. *Top panel*: Mass and redshift distribution of the galaxy clusters colour coded according to their Y_{5R500}. The violet line shows the lower mass limit imposed in the selection and described in the text. *Bottom panel*: Cluster distribution in integrated y signal, Y_{5R500}, and angular scale, θ_{500}, colour coded according to cluster mass M_{500}.

range $0.05 \leq z \leq 0.3$ (which corresponds to 18 snapshots of the simulation). This selection yields a total of 1487 clusters, whose distribution in mass and redshift is shown in Fig. 11.1.

The redshift range used in this work, which is smaller than that of the *Planck* catalogue ($0.01 < z < 0.97$, Planck Collaboration et al. 2016d), is constrained by the limits of the simulation. Outside of this range, the cluster distribution in radius and SZ flux does not allow the sampling the completeness function properly. In particular, at redshifts greater than 0.3, there is an insufficient number of high-mass halos, while at low redshifts, the spacing between

snapshots results in the presence of empty regions in the domain of the completeness function. A more detailed discussion of this limit can be found in Sect. 11.6. Nevertheless, I verified that about 65% of the *Planck* cosmological sample falls within the mass and redshift range covered by our simulation sample.

Having selected the sample of galaxy clusters, I proceeded by computing the projected images of the SZ effect, which were used for the completeness estimation.

The first part of the procedure is analogous to the one presented in Sect. 9.3.1, where I computed images of the Compton-y parameter from galaxy clusters. Starting from the simulated gas cells associated with each cluster, I computed the electron pressure P_e for all cells. Then, I computed the Compton-y parameter using Eq. 9.5, projecting the pressure distribution along six direction: three along the axes (x, y, z) of the simulation box, and three along the axis rotated by Euler angles $(\alpha, \beta, \gamma) = (45°, 45°, 45°)$. The resulting images, centred on the clusters, are $4\,\mathrm{Mpc}$ wide with a resolution of 256×256 pixels[1]. For each of these images, I computed the integrated y signal within $5 \times R_{500}$, Y_{5R500}.

Prior to the injection of these images into the *Planck* frequency maps, they must first be converted into angular coordinates, as they would be observed on the sky. The images are then rescaled according to their redshift, z, using the following relations:

$$\theta_{img} = \frac{4\mathrm{Mpc}}{d_A(z)}, \qquad \theta_{500} = \frac{R_{500}}{d_A(z)}, \tag{11.1}$$

where $d_A(z)$ is the angular diameter distance, θ_{img} is the angular size of the image, and θ_{500} is the equivalent of R_{500} in angular coordinates. The size of the pixels of the rescaled images $0.5\ \mathrm{arcmin}$. This resolution is thus smaller than that of the *Planck* maps, of about $1.7\ \mathrm{arcmin}$. This is done to avoid having excessively coarse images before convolving them with the *Planck* beams. When the images are injected into the frequency maps, their resolution is adjusted to align with to that of the maps.

To be injected in the frequency maps, the Compton-y images were multiplied by the spectral dependence of the SZ effect at the six frequencies of Planck HFI $(100, 143, 217, 353, 545, 857\mathrm{GHz})$, neglecting relativistic corrections:

$$\frac{\Delta T}{T_{\mathrm{CMB}}} = y \cdot g(\nu), \tag{11.2}$$

where T_{CMB} is the CMB temperature, and $g(\nu)$ is the spectral signature of the tSZ effect integrated over *Planck* frequency bandpasses, taken from Planck

[1]Different pixel resolutions were tested, and no appreciable difference in the average cluster y profile and integrated y signal were found.

Table 11.1: Generalised NFW pressure profile parameters of the different sets of spherical images. In order: Arnaud et al. (2010) profile (Standard), Planck Collaboration et al. (2013a) profile (Planck), Pointecouteau et al. (2021) profile (PACT), Tramonte et al. (2023) profile (Tramonte+23), profile obtained changing the c_{500} of Arnaud et al. (2010) (Peaked), and profile obtained fitting the average profile from the simulation images (SimFit).

Name	c_{500}	α	β	γ
Standard	1.177	1.051	5.4905	0.3081
Planck	1.81	1.33	4.13	0.31
PACT	1.18	1.08	4.30	0.31
Tramonte+23	2.1	2.2	5.3	0.31
Peaked	1.5	1.051	5.4905	0.3081
SimFit	5.1×10^{-3}	0.71	1.33	500

Collaboration et al. (2016h). To make the images consistent with *Planck* observations, at each frequency they were convolved by the corresponding beam, assumed to be circular Gaussian with FWHM taken from Planck Collaboration et al. (2016d, 2014g).

In this way, I obtained a set of 8922 cluster images for each of the *Planck* frequencies, with their angular scale, θ_{500}, and integrated SZ signal, Y_{5R500}.

11.2.2 . Circular images

In addition to the set of images extracted from the simulation, I also constructed different sets of images of spherically symmetric clusters with different profiles, to test the consistency of the results obtained with the simulation set and explore their implications.

In order to obtain a meaningful comparison between the completeness obtained from these sets of images and that obtained from the simulation, their distribution in $(Y_{5R500}, \theta_{500})$ needs to be similar. Therefore, I constructed a catalogue of $(Y_{5R500}, \theta_{500})$, obtained applying a random offset to the $(Y_{5R500}, \theta_{500})$ pairs from the simulation set. The offset was sampled from a Gaussian distribution with standard deviation equals to 5% of the original values. This was done to sample the same region but not exactly the same values. The circular images were constructed based on this new catalogue.

The images in the different sets were all constructed starting from a spherical pressure distribution with a gNFW radial profile. I used six different sets of parameters to build the images, which are detailed in Table 11.1: four come from observational studies (Arnaud et al. 2010; Planck Collaboration et al. 2013a;

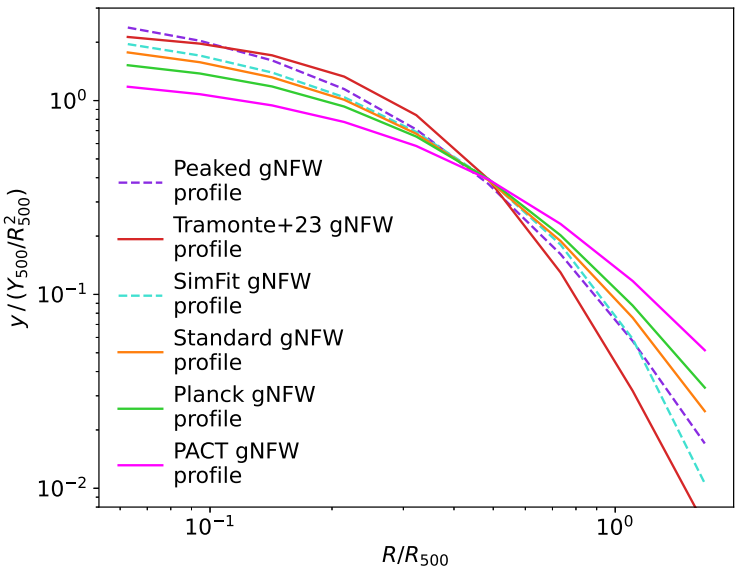

Figure 11.2: Compton-y profiles obtained from the gNFW profiles of Table 11.1, in units of Y_{500}/R^2_{500}, as a function of normalized radius, R/R_{500}. *In orange*: Profile from Arnaud et al. (2010). *In green*: Profile from Planck Collaboration et al. (2013a); *in pink*: profile from Pointecouteau et al. (2021). *In red*: Profile from Tramonte et al. (2023). *In purple*: 'Peaked' profile. *In light blue*: 'SimFit' profile.

Pointecouteau et al. 2021; Tramonte et al. 2023), and two are artificially constructed to approximate the average profile of the simulation images.

To obtain the SZ images for each of the different sets, I first integrated the pressure profile along one direction and transformed into a template y map. I then rescaled this map to match the various $(Y_{5R500}, \theta_{500})$ of the catalogue described above. Hereafter, the y images obtained were treated with the same steps as the simulation images; namely convolution with the *Planck* beams and transformation into frequency images using Eq. 11.2. The Compton-y profiles obtained from the pressure profiles of Table 11.1 are shown in Fig. 11.2.

11.3 . Completeness from Monte Carlo injection

To compute the completeness from the different sets of images, I used the Monte Carlo injection method, presented in Sect. 10.2. The method consists of injecting the synthetic cluster signals into the *Planck* frequency maps, and then running the MMF detection algorithm to determine how well the injected population can be retrieved. The details of the method and the completeness estimation are the following.

In order to obtain a sufficient number of detections to properly sample the completeness in the $(Y_{5R500}, \theta_{500})$ plane, without altering the noise properties of the Planck maps, several realisations of injected sky maps were created, each containing 2000 cluster images. Consequently, the average noise of the injected maps differs from that of the cleaned maps by less than 5%.

A total of 50 mock sky maps were created for the Simulation set and for the Standard set, while 10 maps were created for each of the other sets with different gNFW profiles. For each mock map, a total of 2000 cluster images were randomly selected, along with their associated Y_{5R500} and θ_{500}. The images were injected into randomly selected locations within the cleaned *Planck* maps, distributed uniformly, avoiding the Galactic and point-source masks, as well as overlaps with other injected clusters. This results in an average of approximately four clusters per patch ($10^\circ \times 10^\circ$). The MMF detection algorithm was then applied to the maps, resulting in a catalogue of candidate detections for each, including candidate positions, S/Ns, θ_{500}, and Y_{5R500}. A selection threshold of S/N > 4.5 is imposed on the catalogues. This threshold was selected for all the completeness tests because of the relatively large number of detected clusters, which allows for more robust statistics for the completeness calculation. Selecting a higher S/N threshold (e.g., S/N > 6, as for the *Planck* cosmology sample) yields equivalent results for the completeness, thereby providing a posteriori justification for our initial choice.

11.4 . Completeness results

In this Section, I present the completeness function estimated with the Monte Carlo injection method using the set of cluster images from the IllustrisTNG simulation, and discuss its departure from the analytical ERF completeness with the help of the other sets of circular cluster images with varying profiles.

The completeness function obtained from the simulation images set is shown in Fig. 11.3, as a function of Y_{5R500}, in six θ_{500} bins. The results are compared with the analytical ERF completeness, and with the completeness obtained with the "Standard" image set, which is constructed from the Arnaud et al. (2010) profile (the same profile used to build the detection template). This set of circular images is meant to serve as a benchmark for the injection method and as a test for the ideal case in which all the cluster images match the detection template almost perfectly. In this ideal case, the matched filter is supposed to yield the maximum possible response, and, therefore, the maximum completeness. Consequently, we could expect the simulation images, which exhibit a diverse range of profiles and complex morphologies that deviate from spherical symmetry, will prove more challenging to detect, resulting

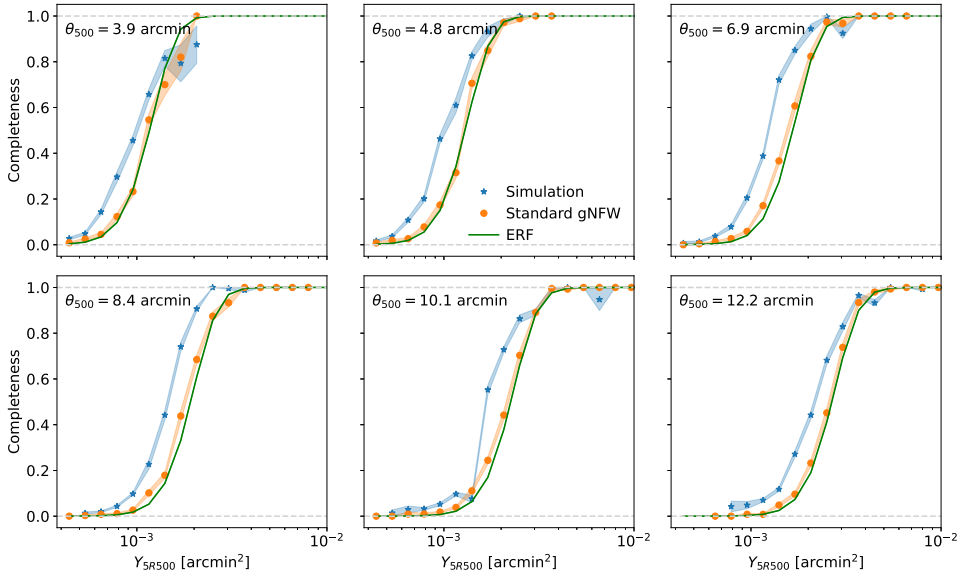

Figure 11.3: Completeness as a function of Y_{5R500} in six θ_{500} bins, computed with Monte Carlo injection for the simulation and circular sets of clusters (in blue and orange, respectively), and estimated with the ERF approximation, in green. The shaded areas represent the uncertainty on the completeness in each bin, computed via bootstrap resampling.

in a lower completeness.

This prediction is not reflected in the results of Fig. 11.3. While the completeness of the Standard images is in good agreement with the analytical ERF (consistent with the results of Planck Collaboration et al. 2014c, 2016d), the simulation images produce a completeness that is almost always higher than both other estimates, with differences of up to 0.4 in some bins. This implies that the simulation cluster images, despite their imperfect match with the detection template, have a higher detection probability than those for which the match is near perfect, for the same $(Y_{5R500}, \theta_{500})$.

In order to gain insight into this result, it is helpful to consider separately the contributions of the simulation images to the completeness: on the one hand, the average y profile, and on the other hand the non-spherical morphology. The following sections investigate the impact of these two aspects on the completeness function.

11.4.1 . Impact of cluster profile

To understand the impact of the cluster y profile on the completeness, I first compared the average profile of the simulation images with the integrated Arnaud et al. (2010) profile, shown in Fig. 11.4. We can see how the average simulation profile tends to be overall steeper than that from Arnaud

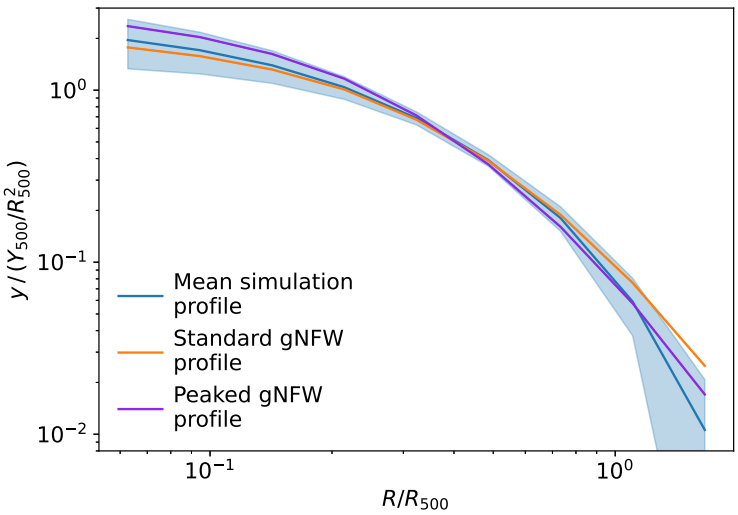

Figure 11.4: Average Compton-y profile of simulation images (in blue, shaded area corresponds to standard deviation of profile sample), compared with the y profile obtained from the Arnaud et al. (2010) pressure profile (in orange) and the 'peaked' profile described in the text (in purple), in units of R_{500} and Y_{500}/R^2_{500}.

et al. (2010), especially in the outer part from $\sim 0.7R_{500}$. Additionally, we see that the simulation profiles are generally more concentrated than the Standard one. This trend may be the cause of the higher completeness of the simulation images compared to the Standard set. That is because the MMF algorithm, in the attempt to find the optimal parameter to fit its flatter profile template to the cluster signal, might favour a smaller radius than the real one. This underestimation of the radius leads to an increase in S/N, given that the MMF noise estimate increases with filter radius. This hypothesis is confirmed by comparing the distribution of the detected cluster radii of the Simulation set over the true injected ones, shown in the bottom left panel of Fig. 11.5. From it, we see that the detected θ_{500} are clearly biased low compared to the true ones, with a median ratio between detected and injected θ_{500} of the order of ~ 0.8. This bias disappears when analysing the detected θ_{500} of the Standard set (bottom right panel) with only about 1% median ratio. This indicates that the performance of the detection process depends on the assumed cluster profile in the template. Comparing the distribution of the detected SZ signal Y_{5R500} with the injected one for the two sets of cluster images (Fig. 11.5, top row) reveals that both tend to be overestimated with respect to the injected quantities. I find a median overestimation of about 9% for the Standard set (which is consistent with the result of Planck Collaboration et al. 2014c). In contrast, the Simulation set has a median overestimation of $\sim 25\%$,

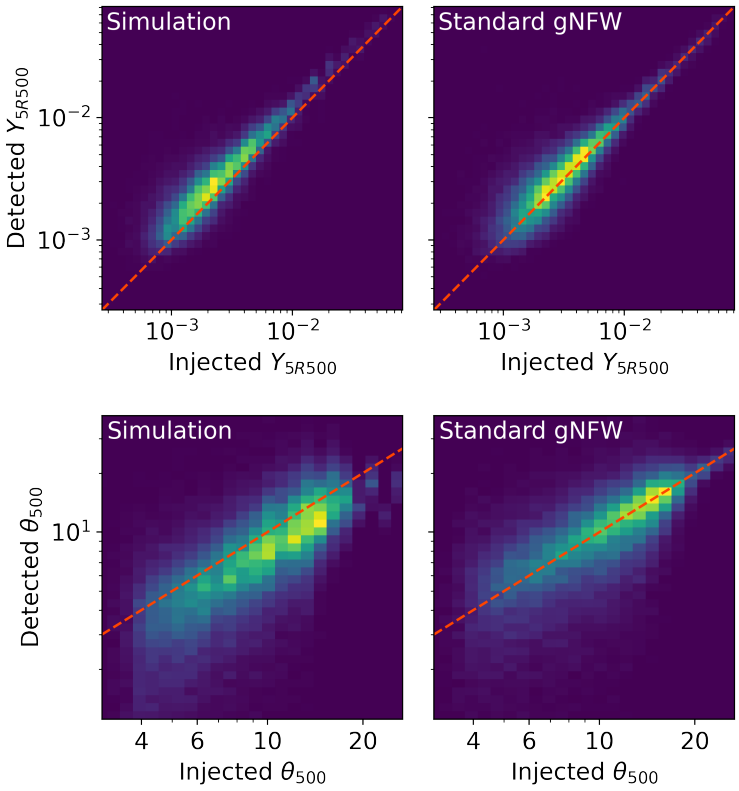

Figure 11.5: Comparison of real cluster properties vs detected ones. The top row shows the total integrated SZ flux, Y_{5R500}, while the bottom row shows the cluster radius, θ_{500}. The columns show two different sets of cluster images: simulation images and circular images with the Arnaud et al. (2010) profile.

which leads to an increased S/N that, in turn, contributes to the increase in the completeness.

To confirm the impact of a different mean profile on the completeness, I used the Peaked set, which is built from a gNFW pressure profile that maintains the same parameters as the Standard profile (Arnaud et al. 2010), except for the concentration parameter, $c_{500} = 1.5$. This higher parameter determines a y profile that is higher in the centre and lower around R_{500} compared to the Standard profile, and roughly reproduces the average profile of the simulated clusters at large radii, as it can be seen in Fig. 11.4. The completeness obtained from this set of images is shown in Fig. 11.6 for two θ_{500} bins as an example. The completeness of the Peaked set appears to agree well with the one obtained with the simulation images, particularly at larger cluster scales. At low θ_{500}, however, the two functions present some differences, and the completeness of the Peaked set appears to be rather in between the simulation and the ERF completeness. This result provides further evidence

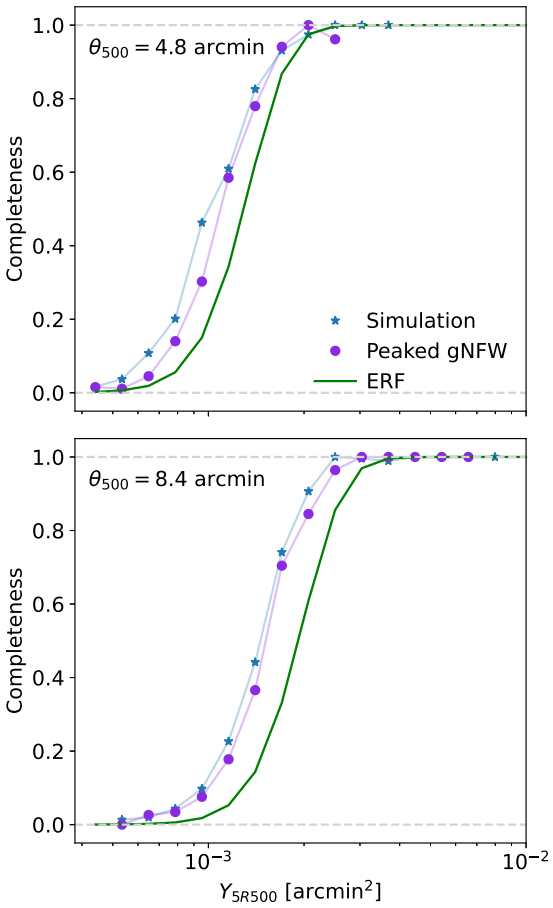

Figure 11.6: Same as in Fig. 11.3, comparing the completeness obtained from the simulation images (*blue*) with that from the Peaked profile images (*purple*) and the ERF completeness (*green*).

that a change in the shape of the profile can lead to an increased detection probability, despite the imperfect match with the filter template.

To probe the dependence of the completeness function on the shape of the cluster profile, I compared the completeness of three sets of circular images, constructed from three different gNFW profiles, obtained fitting the pressure profile of observed cluster samples. The chosen profiles are taken from the following studies: Planck Collaboration et al. (2013a), where the profile was fitted on 62 SZ-selected clusters using Compton-y data from *Planck* and x-ray observations from *XMM-Newton*; Pointecouteau et al. (2021), where the SZ signal from the combined map of *Planck* and ACT (Aghanim et al. 2019) of 31 clusters was used to constrain the gNFW parameters; and finally, Tramonte et al. (2023), where the authors fitted the y profile obtained by stacking the *Planck* SZ signal of a large number of clusters from different surveys, in dif-

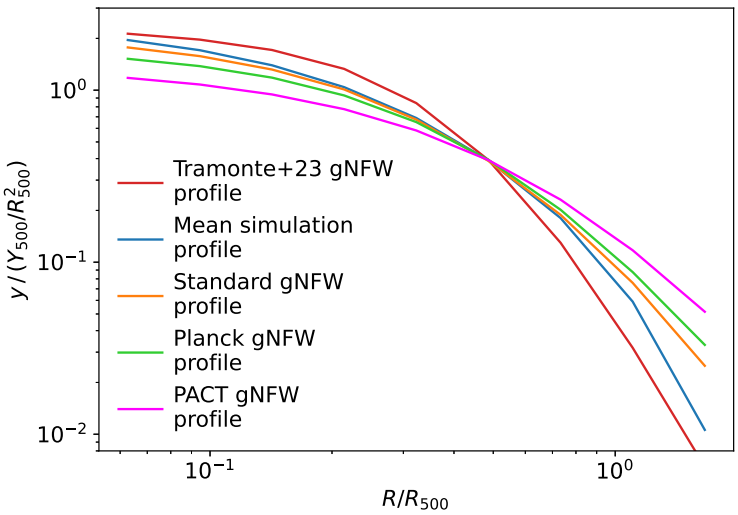

Figure 11.7: Compton-y profiles in units of Y_{500}/R^2_{500} as a function of normalized radius, R/R_{500}. *Blue*: Mean y profile of the simulation images. *Orange*: Profile from Arnaud et al. (2010). *Green*: Profile from Planck Collaboration et al. (2013a). *Pink*: Profile from Pointecouteau et al. (2021). *Red*: Profile from Tramonte et al. (2023).

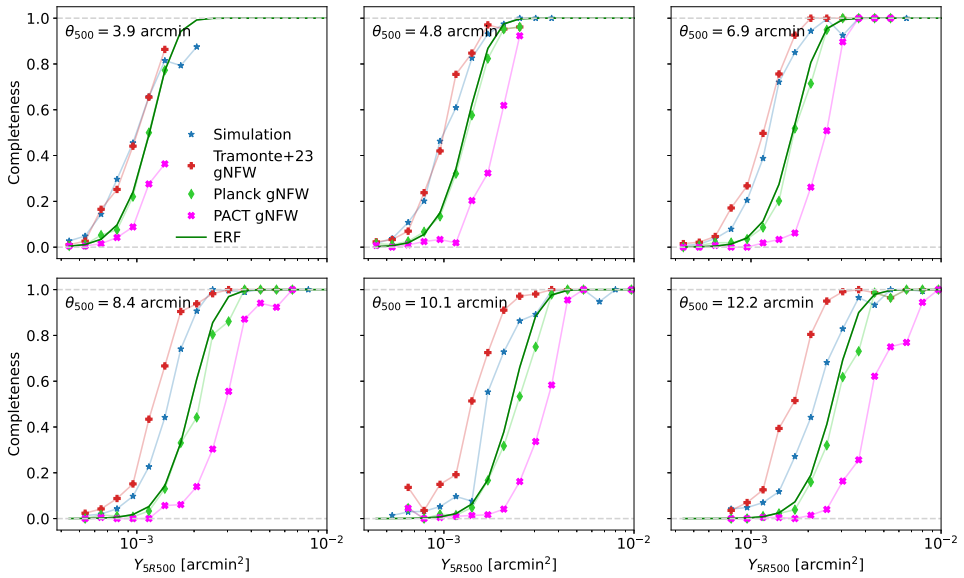

Figure 11.8: Same as in Fig. 11.3, comparing the completeness from images with three different observed profiles. *Green*: Planck set (profile from Planck Collaboration et al. 2013a). *Pink*: PACT set (profile from Pointecouteau et al. 2021). *Red*: Tramonte+23 set (profile from Tramonte et al. 2023).

ferent redshift bins. From this latter study, we take the parameters obtained from the set of 4421 clusters in the $z < 0.35$ redshift bin, which overlaps with our simulation sample. The Compton-y profiles so obtained, shown in Fig 11.7, are named Planck, PACT and Tramonte+23, respectively. The first two profiles are flatter than the Standard profile, with the second showing a greater difference. The third profile is flatter in the inner cluster region (up to $\sim 0.2\, R_{500}$) and then becomes steeper than all the other profiles. With these profiles, I was able to probe the effect on the completeness of having clusters that have either flatter or steeper profiles compared to the detection template.

The completeness functions obtained from these three sets of images, for six cluster scales, are shown in Fig. 11.8, together with the completeness obtained from the simulation set, and the ERF completeness. We see that the shape of the profile has an important effect on the completeness. In particular, we see how clusters with steeper profiles produce higher completeness at all scales compared to the ERF, while flatter profiles result in lower completeness. More specifically, it can be observed that the Planck image set, whose profile most closely resembles the detection template, provides a completeness that is comparable with the ERF estimate, although it tends to be slightly lower at high Y_{5R500}. The PACT set, whose profile is the flattest among the ones considered, produces a completeness that is significantly lower than the ERF curve. On the other hand, the Tramonte+23 set gives a completeness that is overall higher than the ERF, the same effect observed with the simulation set. In all cases, the differences with the ERF completeness seem to increase with the cluster scale.

11.4.2 . Impact of cluster asymmetry

The second aspect of the simulation cluster images that was investigated for its possible impact on the completeness is the cluster morphology, in particular its departure from spherical symmetry. This was quantified using the multipolar ratios introduced in Sect. 6.1. In particular, I focused on the β_2 multipolar ratio (i.e. the ratio between the power of the quadrupolar order and the $0^{\rm th}$ order mutipole). As discussed in Chapter 7, this ratio is the leading non-spherical order inside clusters, and it is correlated with the cluster ellipticity. Therefore, it was chosen as a proxy for the cluster morphology.

I computed the β_2 parameter of the Compton-y images in the Simulation set (before the convolution with the *Planck* beams), in the aperture $\Delta R = [0, \theta_{500}]$. I then selected the images with the top and bottom 25% β_2 values, consisting of 2231 images per group. These two subgroups of images were named the "more elliptical" and "more spherical" subsets, respectively, and represent the two extremes in cluster morphology in the simulation set. I computed the completeness from the two subsets separately, and the results are shown in Fig. 11.9.

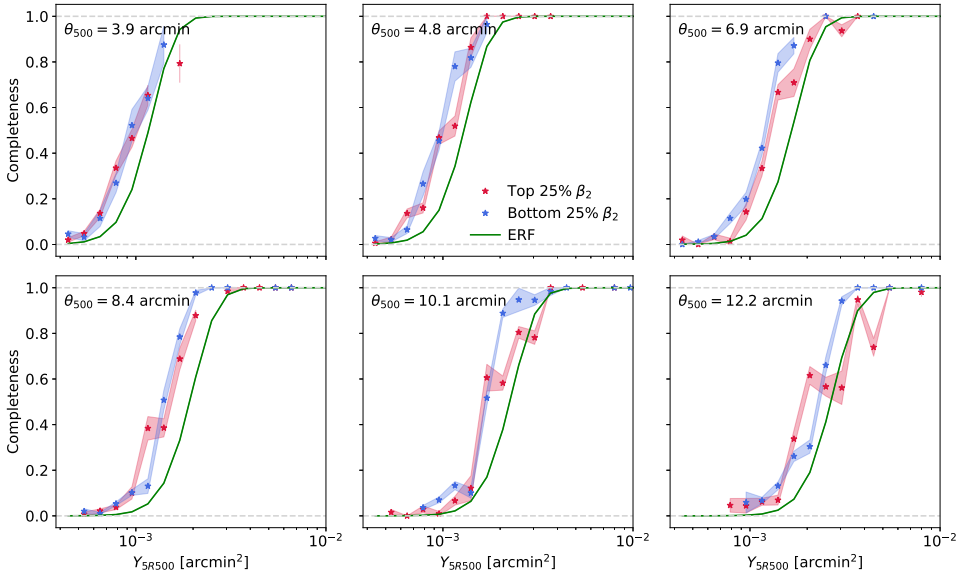

Figure 11.9: Same as in Fig. 11.3, comparing the completeness of two subsets of the simulation images. Images with the 25% highest β_2 (more elliptical) are shown in red, while images with the 25% lowest β_2 (more spherical) are shown in blue.

Comparing the completeness of the more and less spherical images, we see no appreciable difference between the two curves for cluster sizes below ~ 6 arcmin. From Fig. 11.9, we can see that the completeness functions of the two subsets exhibit little differences for small cluster scales (below $\sim 6\mathrm{arcmin}$). This is to be expected, given that the *Planck* beam size is indeed about $6\mathrm{arcmin}$ on average (and reaches $10\mathrm{arcmin}$ for the $100\mathrm{GHz}$ channel), effectively erasing most smaller-scale morphological differences and symmetrising the images. For larger scales, however, a small difference emerges between the two subsets. On average, the more elliptical images exhibit lower completeness than the more spherical ones. This difference appears to become more pronounced with larger cluster radii. This trend can be more clearly visualised in Fig. 11.10, which shows difference between the completeness of the "more spherical" and "more elliptical" subsets, in bins of θ_{500} and Y_{5R500}. From this figure, we indeed see that, for $\theta_{500} \gtrsim 6\mathrm{arcmin}$, the completeness of the "more spherical" subset is overall higher than the "more elliptical" subset completeness, and the difference between the two tends to grow for larger cluster scales. This result indicates that the morphology of resolved clusters has an impact on the detection probability (i.e. the completeness). However, this impact remains relatively modest, at least within the range of scales probed by the simulation cluster sample.

After investigating the completeness of the most and least spherical im-

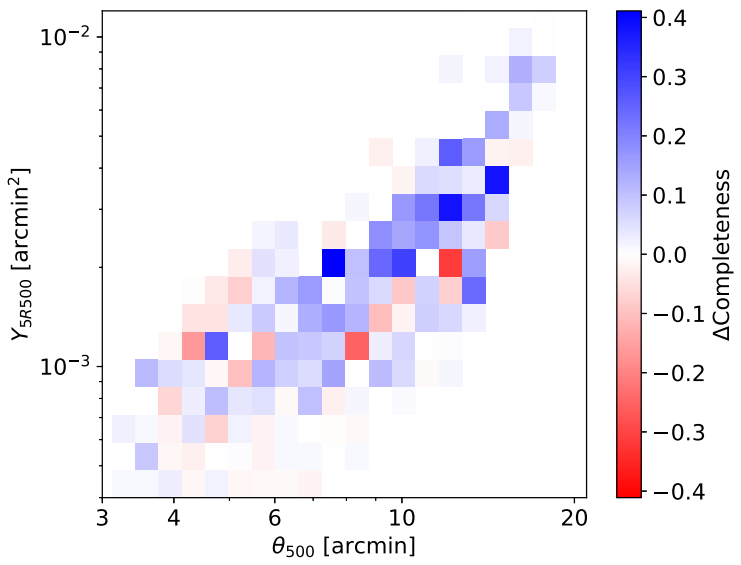

Figure 11.10: Difference in completeness between the 'more spherical' and 'more elliptical' subsets of the simulation images, in bins of (θ_{500}, Y_{5R500}). The bins in blue are the ones in which the completeness from the 'more spherical' subset is higher, while those in red are those in which the completeness of the 'more elliptical' subset dominates.

ages separately, I tested the impact of the full range of cluster morphologies in the Simulation set on the completeness estimation. To this end, I compared the completeness of the Simulation set with the completeness obtained from a set of circular images whose profile was fitted to the average simulation profile (i.e. the SimFit profile in Table 11.1)[2]. The two completeness functions are shown in Fig. 11.11. We can see how the two curves show remarkable agreement for most cluster scales, with small differences visible at high θ_{500}, where the completeness of the Simulation set is slightly smaller. Thus, we see that the effect of cluster morphology is small when considering the full cluster population. Consequently, we may conclude that, in the context of *Planck* and within the range of scales probed by the Simulation set, the departure of clusters from spherical symmetry does not induce a significant bias in the completeness function.

Nevertheless, I demonstrated that the departure from spherical symmetry of the cluster images does produce an effect on the completeness. It is probable that this effect may be more pronounced in a survey where the beam size is smaller.

[2] I also checked for mass and redshift dependence of the y profiles of the Simulation set, and found only minor deviations, much smaller than the scatter of the full sample.

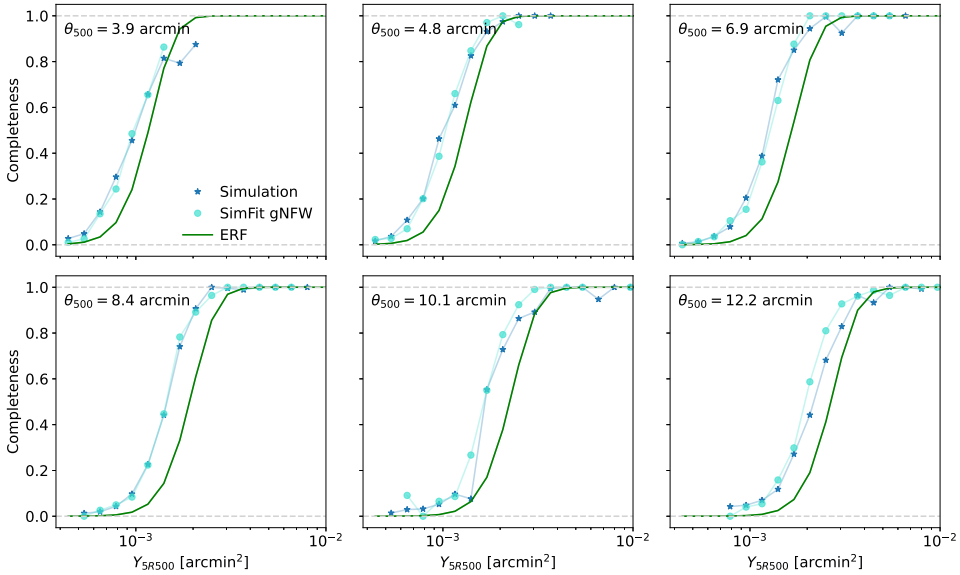

Figure 11.11: Same as in Fig. 11.3, comparing the completeness obtained from the simulation images (in blue) with that from spherical images with the SimFit profile (light blue).

11.5 . Impact on cluster count cosmology

In the previous Section, I analysed the effect on the completeness function of a mismatch between the cluster template assumed in the detection and the "real" clusters to be detected. In this Section, I tested the impact of changing the completeness function in the cosmological analysis. I chose two completeness from the different sets we obtained, one above and one below the ERF estimate: the simulation completeness and the PACT completeness. By modifying the completeness function without altering other elements of the analysis, such as the scaling relations, we are effectively testing a rather extreme scenario in which the true clusters differ from the assumed model in all the steps of the analysis. Consequently, the results presented here should be regarded as indicative of a potential trend rather than as providing definitive results.

As mentioned in the previous Section, the completeness functions shown so far are computed with a selection in S/N equals to S/N > 4.5. In order to perform the cosmological analysis with the *Planck* cosmological sample, I recomputed the completeness with a threshold of S/N > 6.

To simplify the analysis, I chose to approximate the completeness obtained from Monte Carlo injection with an adapted version of the ERF completeness. A simple prescription was employed to approximate the completeness. The functional form of the ERF completeness in Eq. 10.7 was modified

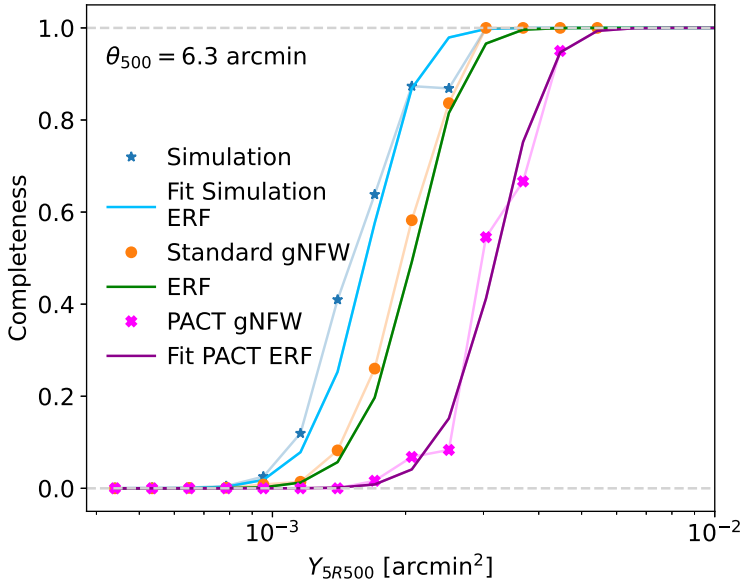

Figure 11.12: Comparison of the three completeness functions used in the cosmological analyses with the functions computed via Monte Carlo injection of simulation images and spherical images with the Arnaud et al. (2010) and Pointecouteau et al. (2021) profiles.

by the addition of a free multiplicative parameter in front of the per-patch noise of the original Planck maps, namely $\sigma_{Yi}^{\rm new}(\theta_{500}) = a_{\rm fit}\,\sigma_{Yi}(\theta_{500})$. This parameter was then optimised to fit the modified ERF completeness to the completeness obtained from Monte Carlo injection. The fitted value of $a_{\rm fit}$ for the simulation completeness is $a_{\rm fit}^{\rm sim} = 1.27$, while for the PACT completeness is $a_{\rm fit}^{\rm PACT} = 0.65$. The fitted ERF completeness, compared to those obtained with Monte Carlo is shown in Fig. 11.12, for one θ_{500} as an example.

I performed a cosmological analysis of cluster number counts, following Planck Collaboration et al. (2016f). I used the *Planck* PSZ2 MMF3 cosmological cluster sample (Planck Collaboration et al. 2016d), adding observational constraints from Big Bang nucleosynthesis (BBN Steigman 2008), $\Omega_b h^2 = 0.022 \pm 0.002$, baryon acoustic oscillation (BAO) measurements from SDSS-III DR12 (Alam et al. 2017), and a prior on $n_s = 0.9624 \pm 0.014$ from Planck Collaboration et al. (2014e). For the mass-observable scaling relations and mass-bias parameters, I used the same priors as the baseline analysis in Planck Collaboration et al. (2016f). The constraints on the cosmological parameters are obtained with a Markov-Chain Monte Carlo (MCMC) implemented in the code `CosmoMC` (Lewis & Bridle 2002). I performed the analysis using the original ERF completeness and the approximated Simulation and PACT completeness described above. The results of the three analyses are compared in Fig. 11.13,

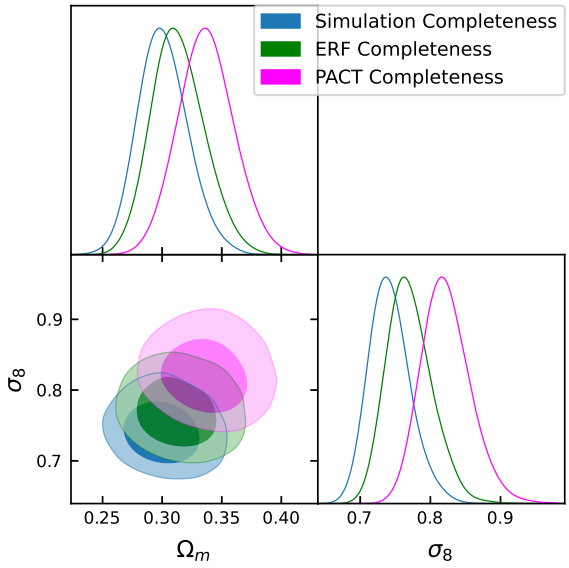

Figure 11.13: Constraints on cosmological parameters (Ω_m, σ_8) from the *Planck* PSZ2 cluster number counts and BAO data, with three different completeness functions: the 'standard' ERF completeness (same as in Planck Collaboration et al. 2016f), and two 'fitted' versions obtained by rescaling the noise per patch by a constant in order to reproduce the completeness derived from the simulation clusters and the PACT (Pointecouteau et al. 2021) profile clusters. The filled contours represent 68% and 95% confidence regions.

for two cosmological parameters: the matter density parameter, Ω_m, and the amplitude of the matter power spectrum, σ_8. The values of the constraints are listed in Table 11.2.

From these results, we note how the change in the completeness shifts the constraints on Ω_m and σ_8. In particular, a higher completeness with respect to the ERF (in this case, the Simulation completeness) favours lower values of the two cosmological parameters. Conversely, a lower completeness (the PACT completeness) favours higher values. For Ω_m, the shift amounts to about 0.6σ for the Simulation completeness, and about 1.1σ for that of PACT. For σ_8, the two shifts are 0.9σ and 1.8σ, respectively. The relation between the completeness function and these two cosmological parameters can be understood in the following terms: as the completeness function informs us about the fraction of detected clusters over the total, keeping the number of detected objects fixed with a higher (respectively, lower) completeness means that the total number of 'true' clusters gets lower (higher). This, in turn, favours a lower (higher) value for both Ω_m and σ_8. It is interesting to note how the direction of the shift in the (Ω_m, σ_8) plane is the same as moving the prior on the hydrostatic mass bias, as shown in Fig. 7 of Planck Collaboration et al.

Table 11.2: Constraints on cosmological parameters obtained from the analysis of *Planck* PSZ2 cluster number counts and BAO, using three different completeness functions, as described in the text.

	Simulation Completeness	ERF Completeness	PACT Completeness
$\mathbf{\Omega_m}$	$0.301^{+0.019}_{-0.023}$	$0.314^{+0.020}_{-0.024}$	0.337 ± 0.024
σ_8	$0.742^{+0.025}_{-0.033}$	$0.770^{+0.027}_{-0.036}$	$0.823^{+0.030}_{-0.039}$

(2016f).

11.6 . Discussion

One limit of this analysis comes from the restricted range of scales of the Simulation set, approximately $[3 - 14]$ arcmin, in comparison to the extent of the Planck catalogue, which is approximately $[1 - 30]$ arcmin. This is due to the limitations of the simulation. First, the restricted volume of the simulation does not permit the formation of a significant number of high-mass clusters. This becomes a limiting factor at high redshift, as there are insufficient high-mass clusters to sample the high-Y_{5R500} part of the completeness. This translates to a lower limit in θ_{500}. The second limitation is associated with the difference in redshift between two successive snapshots, which is approximately 0.01 at low redshifts. When translating the cluster distribution from the mass–redshift plane to the (θ_{500}, Y_{5R500}) plane (as in Fig. 11.1), it becomes evident that including snapshots with $z < 0.05$ produces large empty areas at high θ_{500}. These regions prevent a comprehensive coverage of the completeness in that region and effectively impose a higher limit on the cluster scale.

The limited cluster scale range of the Simulation set precludes the investigation of the impact of realistic cluster shapes in the extremes of the Planck scale range, in particular at large θ_{500}, for which the effect of cluster morphology is expected to be larger. However, this limit does not affect the results concerning the impact of the cluster profile on the completeness. To confirm this, I computed the completeness using images of spherical cluster with different pressure profiles (as in Sect. 11.2.2), but extending the sample to reproduce the extent of the *Planck* catalogue. The results of this test are in agreement with those presented in Sect. 11.4, and show the same differences in the completeness functions obtained from images with different profiles, both at higher and lower θ_{500}.

An aspect of this analysis that could be perceived as overly simplistic is

the use of an identical pressure profile across all circular images within each set. This approach results in the injected cluster signals being merely scaled versions of one another, which is clearly an unrealistic representation, given the scatter around the mean profile observed in the data. The main reason is that a realistic treatment of the detection process is already implemented with the use of the Simulation set. Therefore, I treat the sets of spherically symmetric images as idealised test cases, using them to understand the response of the detection algorithm to different profile shapes. Moreover, the impact of profile scatter among a set of spherical clusters has already been detailed in Planck Collaboration et al. (2014c) and Planck Collaboration et al. (2016d). These works observed that profile variations between clusters create a widening effect in the completeness drop-off region. This effect is qualitatively different from the shift observed when using different mean profiles. Therefore, including a scatter in the profile would not significantly impact the conclusions of this work.

Another potential criticism of the work presented here is the use of a single simulation. Indeed, the specific shape and values of the completeness function identified for the Simulation set are undoubtedly simulation-dependent to some extent. Nonetheless, it can be argued that the chosen simulation is sufficiently realistic in terms of the pressure profile of the cluster gas, as evidenced by the fact that the mean profile of the simulation clusters is not significantly different from that observed in Planck and ACT by Tramonte et al. (2023). Moreover, the primary result of this study is the evidence that the completeness function depends not only on the cluster size, θ_{500}, and total integrated signal, Y_{5R500}, as assumed in the ERF estimate, but on additional cluster characteristics. Notably, as demonstrated in Sect. 11.4, the steepness of the Compton-y profile has a significant impact on the completeness.

11.7 . Conclusions

In this work, I analysed the completeness function for galaxy clusters detected in SZ by *Planck* with the MMF algorithm, focusing on the case where the 'true' clusters differ from the model assumed in the detection method, both in terms of shape and pressure profile. To this end, I performed a Monte Carlo injection of SZ signal from synthetic clusters into cleaned *Planck* sky maps. Comparing the injected sources with those detected by the MMF algorithm, we computed the completeness in bins of injected $(Y_{5R500}, \theta_{500})$ (i.e. total SZ signal and radius of clusters). I used a set of realistic cluster images from the IllustrisTNG simulation, as well as images of spherically symmetric clusters generated from different pressure profiles.

The main result of the analysis presented in this Chapter is the apparent correlation between the completeness function and the shape of the clusters'

pressure profiles compared to the assumed cluster model on the completeness. This effect is observed with both simulation images and circular images. In particular, clusters with pressure profiles steeper than the template one (like for the Simulation set and the Tramonte+23 set) produce completeness functions that are higher than those obtained using clusters with the same profile as in the detection template (the Standard set). Conversely, cluster sets with flatter profiles (e.g., the PACT set) produce lower completeness than the Standard set. The completeness of the Standard set is consistent with the theoretical ERF estimate (based on the assumption of Gaussian noise in the SZ signal), in agreement with Planck Collaboration et al. (2014c, 2016d). A possible explanation for this effect is found observing the estimated properties of detected clusters, compared to their real values. I found that the estimated cluster radius, θ_{500}, is biased on average when the cluster profile differs from the template. In particular, when the cluster profile is steeper (respectively, flatter) than the template, the MMF algorithm tends to underestimate (overestimate) its θ_{500}, which in turn leads to an increase (decrease) in its S/N, given the noise dependence on the scale.

Furthermore, the impact of the cluster morphology on the completeness was also tested. A comparison of the most and least spherical images of the Simulation set revealed that the impact of asphericity is generally small, particularly below the Planck beam size. However, there is a tendency for the "more elliptical subset to be less complete than the "more spherical" subset. This effect becomes increasingly significant as the value of θ_{500} increases, above the beam scale.

Finally, I investigated the effect of modifying the completeness function in the cosmological analysis performed with the *Planck* cosmological cluster sample (Planck Collaboration et al. 2016f). To test this effect, I chose two completeness functions from the ones computed in Sect. 11.4 that deviate substantially from the ERF estimate used in the baseline analysis. I found that the constraints on the cosmological parameters Ω_m and σ_8 are affected by the completeness, exhibiting a shift of approximately one σ. This shift was observed to favour lower values of the cosmological parameters when the completeness was higher, and vice versa. It should be noted that the scenario tested here represents an extreme case in which the true clusters differ from the assumed cluster model in all the steps of the analysis. It is noteworthy that the shift due to changes in the completeness aligns with the shift due to the hydrostatic mass bias reported in Planck Collaboration et al. (2016f). Therefore, I suggest that, in future works, both effects be considered together.

This work demonstrates the necessity of propagating any bias and uncertainty on the completeness into the galaxy cluster number count likelihood for the conduct of robust and accurate cosmological analyses.

Part IV

Conclusions

12 - Conclusions

Galaxy clusters are important cosmological probes. They carry information about the processes of structure formation and the Universe's evolution at low redshift ($z \lesssim 2$), offering a complementary perspective to other probes, such as CMB (e.g. Planck Collaboration et al. 2020a), galaxy clustering (e.g. Elvin-Poole et al. 2018; Abbott et al. 2022), or weak lensing (e.g. Hildebrandt et al. 2017; Hikage et al. 2019; Abbott et al. 2022). In particular, the number of galaxy clusters as a function of mass proxy and redshift has been successfully used to constrain cosmological parameters (e.g. Rozo et al. 2010; Planck Collaboration et al. 2016f; Bocquet et al. 2019; Costanzi et al. 2021; Ghirardini et al. 2024). These successes have also showed that, in order to compare cluster observations with theoretical predictions, it is essential to develop a comprehensive and accurate cluster model that accurately reflects the observed characteristics and the underlying physical relationships between the observable and the hidden quantities, such as the total mass. Indeed, the use of an incorrect cluster model can lead to biased constraints on the cosmological parameters (e.g. Salvati et al. 2020). Therefore, a large effort of the community (e.g. Eckert et al. 2019; Clavico et al. 2019; Campitiello et al. 2022; Wicker et al. 2023; Lebeau et al. 2024; Dupourqué et al. 2024) is put in improving the understanding of the physical processes impacting galaxy clusters (for example accretion, astrophysical feedback processes, and so on), and at the same time ensuring that the simplifying assumptions made in cluster modelling do not affect the cosmological results. In this Thesis, I pursued this line of work, focusing on two main aspects: the characterisation of the matter distribution in cluster environments beyond spherical symmetry, and the impact of assuming an inaccurate cluster model in cluster detection.

For most cosmological purposes, galaxy clusters are approximated as spherically symmetric, isolated objects. In reality, clusters are in general non-spherical (Limousin et al. 2013), and are connected to cosmic filaments, which determine an anisotropic matter accretion (e.g. Gouin et al. 2021; Rost et al. 2021). Therefore, in Part II of this Thesis, I studied the distribution of matter inside and around clusters, in particular focusing on the distribution of the gas component. I performed three studies, which were presented in Chapters 7-9.

In Chapter 7, I investigated statistically the matter distribution in cluster regions, in relation to clusters' structural and physical properties. In particular, I studied the azimuthal distribution of matter in a set of simulated clusters, quantifying their departure from spherical symmetry using the aperture multipole decomposition. I first showed strong correlations between the azimuthal features of gas and DM inside clusters, and the structural properties

of clusters, confirming the ability of the multipole decomposition to trace different structural properties in a comprehensive framework. Then, I established correlations between the level of asymmetry, traced by the β parameter, and not-observable physical cluster properties, such as the cluster's total mass, dynamical state, accretion rate and formation redshift. In cluster outskirts, I showed that the matter asymmetry traces the presence of cosmic filaments connected to the cluster. From these results, I could conclude that matter distribution in and around galaxy clusters is influenced by different cluster properties, and therefore it carries statistical information about the clusters' state and accretion history.

In Chapter 8, I focused on observations of a single cluster, Abell 2744, as a case study for the detectability of filamentary structures in galaxy cluster outskirts with statistical methods. Detecting filaments in cluster outskirts is a challenging task, but a crucial one to study the properties of infalling matter and its complex interactions with the cluster. I analysed the X-ray emission in the outskirts of A2744, combining the analysis of the azimuthal distribution via multipole decomposition with the results of the T-REx filament finder (which I optimised to the use on X-ray data). For the first time in a blind analysis of X-rays alone, I identified three filamentary structures connected to the cluster. I also analysed the distribution of spectroscopic galaxies around A2744, projected in 2D for the multipole analysis, and in 3D with T-REx . Of the three filaments identified in X-rays, two were detected with both methods also from the galaxy catalogue, while the third was less apparent, and it was clearly identified only by T-REx . In addition, T-REx identified two filamentary structures along the line of sight, in the front and in the back of the cluster. In this work, I proved the possibility of detecting cosmic filaments connected to galaxy clusters, in agreement with (Eckert et al. 2015), in a way that could be automated for the use in large X-ray surveys (e.g. XRISM Science Team 2020; CHEX-MATE Collaboration et al. 2021, eROSITA Bulbul et al. 2024).

In Chapter 9, I presented a model to generate realistic galaxy cluster images given the cluster mass. The interest for such a model lies in its ability to generate large samples of cluster images without the need to resort to computationally expensive cosmological simulations. I trained a conditional variational autoencoder based on convolutional neural networks, to generate images of the Compton-y parameter from clusters. The generated images present large-scale anisotropic features, but appear smoother and less complex at small scales compared to the training images. The analysis of the morphological estimators of the generated images reveals that these tend to be, on average, more spherical and less concentrated than training images, with differences also in the correlation with the cluster mass. The relation between cluster mass and integrated y signal is reproduced for the bulk of the sample, but shows deviations at high mass. I concluded that, while the cur-

rent results lack the sufficient quality to be used in place of high-resolution cosmological simulations of galaxy clusters, the model-generated images still exhibit improved realism in their morphological features compared to analytically generated images of perfectly spherically symmetric clusters. Nevertheless, the generative model can be improved in future works, possibly introducing more complex features in the neural network architecture.

In Part III, I focused on studying the impact of modelling assumptions on the cosmological analyses with galaxy clusters. In particular, I investigated the effect of the cluster template in the detection of galaxy clusters with the matched multi-filter method (in its specific application on *Planck* data), and its impact on the cosmological analysis through the selection function. In Chapter 11, I computed the completeness function of galaxy clusters in *Planck*, studying the case in which the real cluster population has properties that differ from the model assumed in the detection template. I showed that the shape of the cluster profile has a strong impact on the completeness: clusters with steeper profiles than the template produce higher completeness functions, while flatter profiles leads to lower completeness. Studying the impact of cluster morphology, I found that the departure from spherical symmetry of the clusters has a moderate impact on the completeness, which tends to increase with the cluster size. I then studied the impact of changing the completeness function obtained with different profiles in the *Planck* cosmological analysis with clusters, and found that the constraints on the cosmological parameters are affected by the change, being shifted by approximately 1σ in my tests. From this study I concluded that the galaxy cluster completeness function is affected by an inaccurate detection template, and that the uncertainties on the completeness can, in turn, impact the cosmological results. Therefore, these uncertainties must be correctly propagated in the cluster number count likelihood to perform robust cosmological analyses with galaxy clusters.

The new generation of large cluster surveys (e.g. Ade et al. 2019; Euclid Collaboration et al. 2019), will provide cluster samples about one order of magnitude larger than current catalogues, which will significantly reduce the statistical uncertainty of cosmological analysis with galaxy clusters. It is therefore becoming increasingly important to account for and reduce the systematic uncertainties associated with inaccurate modelling. This Thesis has its place in this line of work.

List of Figures

List of Tables

Bibliography

Aarseth, S. J., Gott, J. R., I., & Turner, E. L. 1979, ApJ, 228, 664

Abazajian, K., Abdulghafour, A., Addison, G. E., et al. 2022, arXiv e-prints, arXiv:2203.08024

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017, Nature, 551, 85

Abbott, T. M. C., Aguena, M., Alarcon, A., et al. 2020, Phys. Rev. D, 102, 023509

Abbott, T. M. C., Aguena, M., Alarcon, A., et al. 2022, Phys. Rev. D, 105, 023520

Abell, G. O. 1958, ApJS, 3, 211

Ade, P., Aguirre, J., Ahmed, Z., et al. 2019, J. Cosmology Astropart. Phys., 2019, 056

Agarwal, S., Davé, R., & Bassett, B. A. 2018, MNRAS, 478, 3410

Aghanim, N., Douspis, M., Hurier, G., et al. 2019, A&A, 632, A47

Aghanim, N., Tuominen, T., Bonjean, V., et al. 2024, arXiv e-prints, arXiv:2402.18455

Alam, S., Ata, M., Bailey, S., et al. 2017, MNRAS, 470, 2617

Alam, S., Aubert, M., Avila, S., et al. 2021, Phys. Rev. D, 103, 083533

Allen, S. W., Evrard, A. E., & Mantz, A. B. 2011, ARA&A, 49, 409

Allen, S. W., Rapetti, D. A., Schmidt, R. W., et al. 2008, MNRAS, 383, 879

Allgood, B., Flores, R. A., Primack, J. R., et al. 2006, MNRAS, 367, 1781

Andrade-Santos, F., Jones, C., Forman, W. R., et al. 2017, ApJ, 843, 76

Andreon, S. 2015, A&A, 582, A100

Andrianomena, S., Hassan, S., & Villaescusa-Navarro, F. 2024, arXiv e-prints, arXiv:2402.10997

Ansarinejad, B., Raghunathan, S., Abbott, T. M. C., et al. 2024, J. Cosmology Astropart. Phys., 2024, 024

Aragón-Calvo, M. A., Platen, E., van de Weygaert, R., & Szalay, A. S. 2010a, ApJ, 723, 364

Aragón-Calvo, M. A., van de Weygaert, R., & Jones, B. J. T. 2010b, MNRAS, 408, 2163

Arnaud, M., Pratt, G. W., Piffaretti, R., et al. 2010, A&A, 517, A92

Arora, S. & Zhang, Y. 2017, arXiv e-prints, arXiv:1706.08224

Asgari, M., Mead, A. J., & Heymans, C. 2023, The Open Journal of Astrophysics, 6, 39

Bahé, Y. M., Barnes, D. J., Dalla Vecchia, C., et al. 2017, MNRAS, 470, 4186

Balmès, I., Rasera, Y., Corasaniti, P. S., & Alimi, J. M. 2014, MNRAS, 437, 2328

Barnes, D. J., Kay, S. T., Bahé, Y. M., et al. 2017a, MNRAS, 471, 1088

Barnes, D. J., Kay, S. T., Henson, M. A., et al. 2017b, MNRAS, 465, 213

Bartelmann, M. 2010, Classical and Quantum Gravity, 27, 233001

Bassini, L., Rasia, E., Borgani, S., et al. 2020, A&A, 642, A37

Bayes, M. & Price, M. 1763, Philosophical Transactions of the Royal Society of London Series I, 53, 370

Beisbart, C., Valdarnini, R., & Buchert, T. 2001, A&A, 379, 412

Bellagamba, F., Roncarelli, M., Maturi, M., & Moscardini, L. 2018, MNRAS, 473, 5221

Bergamini, P., Acebron, A., Grillo, C., et al. 2023, ApJ, 952, 84

Bhattacharya, S., Heitmann, K., White, M., et al. 2011, ApJ, 732, 122

Biffi, V., Borgani, S., Murante, G., et al. 2016, ApJ, 827, 112

Binggeli, B. 1982, A&A, 107, 338

Binney, J. & Tremaine, S. 1987, Galactic dynamics

Birkinshaw, M. 1999, Phys. Rep., 310, 97

Birrer, S., Shajib, A. J., Galan, A., et al. 2020, A&A, 643, A165

Bleem, L. E., Klein, M., Abbot, T. M. C., et al. 2024, The Open Journal of Astrophysics, 7, 13

Bleem, L. E., Stalder, B., de Haan, T., et al. 2015, ApJS, 216, 27

Bocquet, S., Dietrich, J. P., Schrabback, T., et al. 2019, ApJ, 878, 55

Bocquet, S., Grandis, S., Bleem, L. E., et al. 2023, arXiv e-prints, arXiv:2310.12213

Bocquet, S., Grandis, S., Bleem, L. E., et al. 2024, arXiv e-prints, arXiv:2401.02075

Bocquet, S., Heitmann, K., Habib, S., et al. 2020, ApJ, 901, 5

Böhringer, H., Chon, G., & Fukugita, M. 2017, A&A, 608, A65

Bond, J. R., Kofman, L., & Pogosyan, D. 1996, Nature, 380, 603

Bonjean, V., Aghanim, N., Salomé, P., Douspis, M., & Beelen, A. 2018, A&A, 609, A49

Bonnaire, T., Aghanim, N., Decelle, A., & Douspis, M. 2020, A&A, 637, A18

Bonnaire, T., Decelle, A., & Aghanim, N. 2022, IEEE Transactions on Pattern Analysis and Machine Intelligence, 44, 9119

Bonnet, G., Nezri, E., Kraljic, K., & Schimd, C. 2022, MNRAS, 513, 4929

Borgani, S. & Kravtsov, A. 2011, Advanced Science Letters, 4, 204

Borgani, S., Plionis, M., & Kolokotronis, V. 1999, MNRAS, 305, 866

Boruah, S. S., Fiedorowicz, P., Garcia, R., et al. 2024, arXiv e-prints, arXiv:2406.05867

Boschin, W., Girardi, M., Spolaor, M., & Barrena, R. 2006, A&A, 449, 461

Böss, L. M., Steinwandel, U. P., Dolag, K., & Lesch, H. 2023, MNRAS, 519, 548

Bower, R. G., Lucey, J. R., & Ellis, R. S. 1992, MNRAS, 254, 601

Bowman, S. R., Vilnis, L., Vinyals, O., et al. 2015, arXiv e-prints, arXiv:1511.06349

Boylan-Kolchin, M., Springel, V., White, S. D. M., Jenkins, A., & Lemson, G. 2009, MNRAS, 398, 1150

Braglia, F., Pierini, D., & Böhringer, H. 2007, A&A, 470, 425

Braglia, F. G., Pierini, D., Biviano, A., & Böhringer, H. 2009, A&A, 500, 947

Brunetti, G. & Jones, T. W. 2014, International Journal of Modern Physics D, 23, 1430007

Brunner, H., Liu, T., Lamer, G., et al. 2022, A&A, 661, A1

Buchert, T. 1994, MNRAS, 267, 811

Bulbul, E., Liu, A., Kluge, M., et al. 2024, arXiv e-prints, arXiv:2402.08452

Bulbul, E., Randall, S. W., Bayliss, M., et al. 2016, ApJ, 818, 131

Buote, D. A. & Tsai, J. C. 1995, ApJ, 452, 522

Campitiello, M. G., Ettori, S., Lovisari, L., et al. 2022, A&A, 665, A117

Capalbo, V., De Petris, M., De Luca, F., et al. 2021, MNRAS, 503, 6155

Carlstrom, J. E., Holder, G. P., & Reese, E. D. 2002, ARA&A, 40, 643

Cautun, M., van de Weygaert, R., & Jones, B. J. T. 2013, MNRAS, 429, 1286

Cerini, G., Cappelluti, N., & Natarajan, P. 2023, ApJ, 945, 152

Chen, Y., Mo, H. J., Li, C., et al. 2020, ApJ, 899, 81

CHEX-MATE Collaboration, Arnaud, M., Ettori, S., et al. 2021, A&A, 650, A104

Cialone, G., De Petris, M., Sembolini, F., et al. 2018, MNRAS, 477, 139

Clampitt, J. & Jain, B. 2016, MNRAS, 457, 4135

Clavico, S., De Grandi, S., Ghizzardi, S., et al. 2019, A&A, 632, A27

Clerc, N. & Finoguenov, A. 2023, in Handbook of X-ray and Gamma-ray Astrophysics. Edited by Cosimo Bambi and Andrea Santangelo, 123

Clerc, N., Merloni, A., Zhang, Y. Y., et al. 2016, MNRAS, 463, 4490

Clowe, D., Bradač, M., Gonzalez, A. H., et al. 2006, ApJ, 648, L109

Codis, S., Pogosyan, D., & Pichon, C. 2018, MNRAS, 479, 973

Cole, S. & Lacey, C. 1996, MNRAS, 281, 716

Coles, P. & Lucchin, F. 2002, Cosmology: The Origin and Evolution of Cosmic Structure, Second Edition

Corasaniti, P. S., Ettori, S., Rasera, Y., et al. 2018, ApJ, 862, 40

Corasaniti, P.-S., Sereno, M., & Ettori, S. 2021, ApJ, 911, 82

Costanzi, M., Rozo, E., Simet, M., et al. 2019, MNRAS, 488, 4779

Costanzi, M., Saro, A., Bocquet, S., et al. 2021, Phys. Rev. D, 103, 043522

Couch, W. J., Barger, A. J., Smail, I., Ellis, R. S., & Sharples, R. M. 1998, ApJ, 497, 188

Couch, W. J. & Sharples, R. M. 1987, MNRAS, 229, 423

Cui, W., Knebe, A., Yepes, G., et al. 2018, MNRAS, 480, 2898

Cui, W., Power, C., Borgani, S., et al. 2017, MNRAS, 464, 2502

Dai, B. & Seljak, U. 2021, Proceedings of the National Academy of Science, 118, e2020324118

Dalton, G. B., Maddox, S. J., Sutherland, W. J., & Efstathiou, G. 1997, MNRAS, 289, 263

Darragh Ford, E., Laigle, C., Gozaliasl, G., et al. 2019, MNRAS, 489, 5695

Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, ApJ, 292, 371

de Andres, D., Yepes, G., Sembolini, F., et al. 2023, MNRAS, 518, 111

de Graaff, A., Cai, Y.-C., Heymans, C., & Peacock, J. A. 2019, A&A, 624, A48

de Haan, T., Benson, B. A., Bleem, L. E., et al. 2016, ApJ, 832, 95

De Luca, F., De Petris, M., Yepes, G., et al. 2021, MNRAS, 504, 5383

Del Popolo, A., Pace, F., & Le Delliou, M. 2017, J. Cosmology Astropart. Phys., 2017, 032

Delgado, A. M., Wadekar, D., Hadzhiyska, B., et al. 2022, MNRAS, 515, 2733

Despali, G., Giocoli, C., Angulo, R. E., et al. 2016, MNRAS, 456, 2486

Despali, G., Giocoli, C., & Tormen, G. 2014, MNRAS, 443, 3208

Dhariwal, P. & Nichol, A. 2021, arXiv e-prints, arXiv:2105.05233

Diemer, B. & Kravtsov, A. V. 2014, ApJ, 789, 1

Diemer, B., More, S., & Kravtsov, A. V. 2013, ApJ, 766, 25

Dietrich, J. P., Schneider, P., Clowe, D., Romano-Díaz, E., & Kerp, J. 2005, A&A, 440, 453

Dolag, K., Komatsu, E., & Sunyaev, R. 2016, MNRAS, 463, 1797

Dolag, K. & Schindler, S. 2000, A&A, 364, 491

Dolag, K., Sorce, J. G., Pilipenko, S., et al. 2023, A&A, 677, A169

Doubrawa, L., Cypriano, E. S., Finoguenov, A., et al. 2023, MNRAS, 526, 4285

Dubois, Y., Pichon, C., Welker, C., et al. 2014, MNRAS, 444, 1453

Dupourqué, S., Clerc, N., Pointecouteau, E., et al. 2024, A&A, 687, A58

Eckert, D., Ettori, S., Pointecouteau, E., et al. 2017, Astronomische Nachrichten, 338, 293

Eckert, D., Finoguenov, A., Ghirardini, V., et al. 2020, The Open Journal of Astrophysics, 3, 12

Eckert, D., Ghirardini, V., Ettori, S., et al. 2019, A&A, 621, A40

Eckert, D., Jauzac, M., Shan, H., et al. 2015, Nature, 528, 105

Eckert, D., Jauzac, M., Vazza, F., et al. 2016, MNRAS, 461, 1302

Eckert, D., Molendi, S., & Paltani, S. 2011, A&A, 526, A79

Eckert, D., Vazza, F., Ettori, S., et al. 2012, A&A, 541, A57

Efstathiou, G. 1979, MNRAS, 187, 117

Einasto, J. 1965, Trudy Astrofizicheskogo Instituta Alma-Ata, 5, 87

Einasto, M., Deshev, B., Tenjes, P., et al. 2020, A&A, 641, A172

Einasto, M., Kipper, R., Tenjes, P., et al. 2021, A&A, 649, A51

Einstein, A. 1916, Annalen der Physik, 354, 769

Eisenhardt, P. R. M., Brodwin, M., Gonzalez, A. H., et al. 2008, ApJ, 684, 905

Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al. 2005, ApJ, 633, 560

Elvin-Poole, J., Crocce, M., Ross, A. J., et al. 2018, Phys. Rev. D, 98, 042006

Euclid Collaboration, Adam, R., Vannier, M., et al. 2019, A&A, 627, A23

Euclid Collaboration, Bretonnière, H., Huertas-Company, M., et al. 2022, A&A, 657, A90

Euclid Collaboration, Mellier, Y., Abdurro'uf, et al. 2024, arXiv e-prints, arXiv:2405.13491

Evrard, A. E. 1988, MNRAS, 235, 911

Evrard, A. E. 1997, MNRAS, 292, 289

Feder, R. M., Berger, P., & Stein, G. 2020, Phys. Rev. D, 102, 103504

Finoguenov, A., Rykoff, E., Clerc, N., et al. 2020, A&A, 638, A114

Freedman, W. L., Madore, B. F., Hatt, D., et al. 2019, ApJ, 882, 34

Friedmann, A. 1922, Zeitschrift fur Physik, 10, 377

Fumagalli, A., Costanzi, M., Saro, A., Castro, T., & Borgani, S. 2024, A&A, 682, A148

Galárraga-Espinosa, D., Aghanim, N., Langer, M., Gouin, C., & Malavasi, N. 2020, A&A, 641, A173

Galárraga-Espinosa, D., Aghanim, N., Langer, M., & Tanimura, H. 2021, A&A, 649, A117

Galárraga-Espinosa, D., Cadiou, C., Gouin, C., et al. 2023, arXiv e-prints, arXiv:2309.08659

Gallo, S., Aghanim, N., Gouin, C., et al. 2024a, arXiv e-prints, arXiv:2407.10518

Gallo, S., Douspis, M., Soubrié, E., & Salvati, L. 2024b, A&A, 686, A15

Gavazzi, R., Adami, C., Durret, F., et al. 2009, A&A, 498, L33

Ghirardini, V., Bahar, Y. E., Bulbul, E., et al. 2022, A&A, 661, A12

Ghirardini, V., Bulbul, E., Artis, E., et al. 2024, arXiv e-prints, arXiv:2402.08458

Ghirardini, V., Eckert, D., Ettori, S., et al. 2019, A&A, 621, A41

Gianfagna, G., Rasia, E., Cui, W., De Petris, M., & Yepes, G. 2022, in European Physical Journal Web of Conferences, Vol. 257, mm Universe @ NIKA2 - Observing the mm Universe with the NIKA2 Camera, 00020

Gladders, M. D. & Yee, H. K. C. 2000, AJ, 120, 2148

Gonzalez, A. 2014, in Building the Euclid Cluster Survey - Scientific Program, 7

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., et al. 2014, arXiv e-prints, arXiv:1406.2661

Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759

Gouin, C., Aghanim, N., Bonjean, V., & Douspis, M. 2020, A&A, 635, A195

Gouin, C., Bonamente, M., Galárraga-Espinosa, D., Walker, S., & Mirakhor, M. 2023, A&A, 680, A94

Gouin, C., Bonnaire, T., & Aghanim, N. 2021, A&A, 651, A56

Gouin, C., Gallo, S., & Aghanim, N. 2022, A&A, 664, A198

Gouin, C., Gavazzi, R., Codis, S., et al. 2017, A&A, 605, A27

Govoni, F., Enßlin, T. A., Feretti, L., & Giovannini, G. 2001, A&A, 369, 441

Green, S. B., Ntampaka, M., Nagai, D., et al. 2019, ApJ, 884, 33

Grover, A., Dhar, M., & Ermon, S. 2017, arXiv e-prints, arXiv:1705.08868

Guth, A. H. & Pi, S. Y. 1982, Phys. Rev. Lett., 49, 1110

Haggar, R., Gray, M. E., Pearce, F. R., et al. 2020, MNRAS, 492, 6074

Hahn, O., Martizzi, D., Wu, H.-Y., et al. 2017, MNRAS, 470, 166

Hahn, O., Porciani, C., Carollo, C. M., & Dekel, A. 2007, MNRAS, 375, 489

Hamana, T., Shirasaki, M., Miyazaki, S., et al. 2020, PASJ, 72, 16

Han, D., Sehgal, N., & Villaescusa-Navarro, F. 2021, Phys. Rev. D, 104, 123521

Harvey, D. R. & Massey, R. 2024, MNRAS, 529, 802

Hattori, S., Ota, N., Zhang, Y.-Y., Akamatsu, H., & Finoguenov, A. 2017, PASJ, 69, 39

Heath, D. J. 1977, MNRAS, 179, 351

Hébert-Dufresne, L., Grochow, J. A., & Allard, A. 2016, Scientific Reports, 6, 31708

Henden, N. A., Puchwein, E., Shen, S., & Sijacki, D. 2018, MNRAS, 479, 5385

Herranz, D., Sanz, J. L., Barreiro, R. B., & Martínez-González, E. 2002, ApJ, 580, 610

Higgins, I., Matthey, L., Pal, A., et al. 2017, in International Conference on Learning Representations

Hikage, C., Oguri, M., Hamana, T., et al. 2019, PASJ, 71, 43

Hildebrandt, H., Viola, M., Heymans, C., et al. 2017, MNRAS, 465, 1454

Hilton, M., Sifón, C., Naess, S., et al. 2021, ApJS, 253, 3

Ho, J., Jain, A., & Abbeel, P. 2020, arXiv e-prints, arXiv:2006.11239

Hubble, E. 1929, Proceedings of the National Academy of Science, 15, 168

Hubble, E. P. 1925, ApJ, 62, 409

Hubble, E. P. 1926, ApJ, 64, 321

Huterer, D. 2023, A&A Rev., 31, 2

Hwang, H. S., Geller, M. J., Park, C., et al. 2016, ApJ, 818, 173

Ibaraki, Y., Ota, N., Akamatsu, H., Zhang, Y. Y., & Finoguenov, A. 2014, A&A, 562, A11

Itoh, N., Kohyama, Y., & Nozawa, S. 1998, ApJ, 502, 7

Jackson, J. C. 1972, MNRAS, 156, 1P

Jamieson, D., Li, Y., de Oliveira, R. A., et al. 2023, ApJ, 952, 145

Jauzac, M., Eckert, D., Schwinn, J., et al. 2016, MNRAS, 463, 3876

Jauzac, M., Richard, J., Jullo, E., et al. 2015, MNRAS, 452, 1437

Jaynes, E. T. 1968, IEEE Transactions on Systems Science and Cybernetics, 4, 227

Jeans, J. H. 1902, Philosophical Transactions of the Royal Society of London Series A, 199, 1

Jeffreys, H. 1946, Proceedings of the Royal Society of London Series A, 186, 453

Jimenez Rezende, D. & Mohamed, S. 2015, arXiv e-prints, arXiv:1505.05770

Jo, Y. & Kim, J.-h. 2019, MNRAS, 489, 3565

Kaiser, N. 1986, MNRAS, 222, 323

Kamdar, H. M., Turk, M. J., & Brunner, R. J. 2016, MNRAS, 457, 1162

Katz, N. & White, S. D. M. 1993, ApJ, 412, 455

Kempner, J. C. & David, L. P. 2004, MNRAS, 349, 385

Kingma, D. P. & Ba, J. 2014, arXiv e-prints, arXiv:1412.6980

Kingma, D. P., Salimans, T., Jozefowicz, R., et al. 2016, arXiv e-prints, arXiv:1606.04934

Kingma, D. P. & Welling, M. 2013, arXiv e-prints, arXiv:1312.6114

Kingma, D. P. & Welling, M. 2019, arXiv e-prints, arXiv:1906.02691

Kodi Ramanah, D., Charnock, T., Villaescusa-Navarro, F., & Wandelt, B. D. 2020, MNRAS, 495, 4227

Koester, B. P., McKay, T. A., Annis, J., et al. 2007, ApJ, 660, 221

Komatsu, E. & Seljak, U. 2002, MNRAS, 336, 1256

Kraljic, K., Pichon, C., Codis, S., et al. 2020, MNRAS, 491, 4294

Kravtsov, A. V. & Borgani, S. 2012, ARA&A, 50, 353

Kravtsov, A. V., Vikhlinin, A., & Nagai, D. 2006, ApJ, 650, 128

Kuchner, U., Aragón-Salamanca, A., Pearce, F. R., et al. 2020, MNRAS, 494, 5473

Kullback, S. & Leibler, R. A. 1951, The Annals of Mathematical Statistics, 22, 79

Lamb, A. 2021, arXiv e-prints, arXiv:2103.00265

Le Brun, A. M. C., McCarthy, I. G., Schaye, J., & Ponman, T. J. 2014, MNRAS, 441, 1270

Lebeau, T., Sorce, J. G., Aghanim, N., Hernández-Martínez, E., & Dolag, K. 2024, A&A, 682, A157

Lee, B. E., Le Brun, A. M. C., Haq, M. E., et al. 2018, MNRAS, 479, 890

Lee, J., Shin, J., Snaith, O. N., et al. 2021, ApJ, 908, 11

Lemaître, G. 1931, MNRAS, 91, 483

Lewis, A. & Bridle, S. 2002, Phys. Rev. D, 66, 103511

Li, Y., Ni, Y., Croft, R. A. C., et al. 2021, Proceedings of the National Academy of Science, 118, e2022038118

Libeskind, N. I., van de Weygaert, R., Cautun, M., et al. 2018, MNRAS, 473, 1195

Limousin, M., Morandi, A., Sereno, M., et al. 2013, Space Sci. Rev., 177, 155

Lovell, C. C., Wilkins, S. M., Thomas, P. A., et al. 2022, MNRAS, 509, 5046

Lovisari, L., Forman, W. R., Jones, C., et al. 2017, ApJ, 846, 51

Malavasi, N., Aghanim, N., Tanimura, H., Bonjean, V., & Douspis, M. 2020, A&A, 634, A30

Malavasi, N., Sorce, J. G., Dolag, K., & Aghanim, N. 2023, A&A, 675, A76

Mantz, A., Allen, S. W., Rapetti, D., & Ebeling, H. 2010, MNRAS, 406, 1759

Mantz, A. B., Allen, S. W., Morris, R. G., et al. 2015a, MNRAS, 449, 199

Mantz, A. B., Morris, R. G., Allen, S. W., et al. 2022, MNRAS, 510, 131

Mantz, A. B., von der Linden, A., Allen, S. W., et al. 2015b, MNRAS, 446, 2205

Marinacci, F., Vogelsberger, M., Pakmor, R., et al. 2018, MNRAS, 480, 5113

Markevitch, M., Gonzalez, A. H., David, L., et al. 2002, ApJ, 567, L27

Martizzi, D., Vogelsberger, M., Artale, M. C., et al. 2019, MNRAS, 486, 3766

Marulli, F., Veropalumbo, A., Sereno, M., et al. 2018, A&A, 620, A1

Maturi, M., Angrick, C., Pace, F., & Bartelmann, M. 2010, A&A, 519, A23

McCarthy, I. G., Schaye, J., Bird, S., & Le Brun, A. M. C. 2017, MNRAS, 465, 2936

McClintock, T., Rozo, E., Becker, M. R., et al. 2019, ApJ, 872, 53

Mead, J. M. G., King, L. J., & McCarthy, I. G. 2010, MNRAS, 401, 2257

Medezinski, E., Umetsu, K., Okabe, N., et al. 2016, ApJ, 817, 24

Melin, J. B., Bartlett, J. G., & Delabrouille, J. 2005, A&A, 429, 417

Melin, J. B., Bartlett, J. G., & Delabrouille, J. 2006, A&A, 459, 341

Melin, J. B. & Pratt, G. W. 2023, A&A, 678, A197

Merten, J., Coe, D., Dupke, R., et al. 2011, MNRAS, 417, 333

Mohr, J. J., Evrard, A. E., Fabricant, D. G., & Geller, M. J. 1995, ApJ, 447, 8

Mohr, J. J., Fabricant, D. G., & Geller, M. J. 1993, ApJ, 413, 492

More, S., Diemer, B., & Kravtsov, A. V. 2015, ApJ, 810, 36

Munari, E., Biviano, A., & Mamon, G. A. 2014, A&A, 566, A68

Murray, S. G., Power, C., & Robotham, A. S. G. 2013, Astronomy and Computing, 3, 23

Mustafa, M., Bard, D., Bhimji, W., et al. 2019, Computational Astrophysics and Cosmology, 6, 1

Nagai, D., Kravtsov, A. V., & Vikhlinin, A. 2007, ApJ, 668, 1

Naiman, J. P., Pillepich, A., Springel, V., et al. 2018, MNRAS, 477, 1206

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493

Nelson, D., Pillepich, A., Ayromlou, M., et al. 2024, A&A, 686, A157

Nelson, D., Pillepich, A., Springel, V., et al. 2018a, MNRAS, 475, 624

Nelson, D., Pillepich, A., Springel, V., et al. 2018b, MNRAS, 475, 624

Nelson, D., Springel, V., Pillepich, A., et al. 2019a, Computational Astrophysics and Cosmology, 6, 2

Nelson, D., Springel, V., Pillepich, A., et al. 2019b, Computational Astrophysics and Cosmology, 6, 2

Ntampaka, M. & Vikhlinin, A. 2022, ApJ, 926, 45

Oguri, M., Miyazaki, S., Li, X., et al. 2021, PASJ, 73, 817

O'Hara, T. B., Mohr, J. J., Bialek, J. J., & Evrard, A. E. 2006, ApJ, 639, 64

Okabe, T., Nishimichi, T., Oguri, M., et al. 2018, MNRAS, 478, 1141

O'Shea, K. & Nash, R. 2015, arXiv e-prints, arXiv:1511.08458

Owers, M. S., Randall, S. W., Nulsen, P. E. J., et al. 2011, ApJ, 728, 27

Pacaud, F., Pierre, M., Melin, J. B., et al. 2018, A&A, 620, A10

Pakmor, R., Springel, V., Coles, J. P., et al. 2023, MNRAS, 524, 2539

Parkinson, D., Riemer-Sørensen, S., Blake, C., et al. 2012, Phys. Rev. D, 86, 103518

Pearce, F. A., Kay, S. T., Barnes, D. J., Bower, R. G., & Schaller, M. 2020, MNRAS, 491, 1622

Pearson, D. W., Batiste, M., & Batuski, D. J. 2014, MNRAS, 441, 1601

Peebles, W. & Xie, S. 2022, arXiv e-prints, arXiv:2212.09748

Pellissier, A., Hahn, O., & Ferrari, C. 2023, MNRAS, 522, 721

Pereyra, L. A., Sgró, M. A., Merchán, M. E., Stasyszyn, F. A., & Paz, D. J. 2020, MNRAS, 499, 4876

Pierre, M., Pacaud, F., Adami, C., et al. 2016, A&A, 592, A1

Pierre, M., Pacaud, F., Juin, J. B., et al. 2011, MNRAS, 414, 1732

Pierre, M. & Starck, J. L. 1998, A&A, 330, 801

Piffaretti, R., Arnaud, M., Pratt, G. W., Pointecouteau, E., & Melin, J. B. 2011, A&A, 534, A109

Pillepich, A., Nelson, D., Hernquist, L., et al. 2018, MNRAS, 475, 648

Planck Collaboration, Adam, R., Ade, P. A. R., et al. 2016a, A&A, 594, A1

Planck Collaboration, Adam, R., Ade, P. A. R., et al. 2016b, A&A, 594, A8

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014a, A&A, 571, A1

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014b, A&A, 571, A13

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016c, A&A, 594, A26

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014c, A&A, 571, A29

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014d, A&A, 571, A20

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014e, A&A, 571, A16

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014f, A&A, 571, A6

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014g, A&A, 571, A7

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2013a, A&A, 550, A131

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2013b, A&A, 550, A134

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2011, A&A, 536, A8

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016d, A&A, 594, A27

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016e, A&A, 594, A13

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016f, A&A, 594, A24

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020a, A&A, 641, A6

Planck Collaboration, Aghanim, N., Arnaud, M., et al. 2016g, A&A, 594, A22

Planck Collaboration, Aghanim, N., Arnaud, M., et al. 2016h, A&A, 594, A22

Planck Collaboration, Akrami, Y., Arroja, F., et al. 2020b, A&A, 641, A9

Planelles, S., Schleicher, D. R. G., & Bykov, A. M. 2015, Space Sci. Rev., 188, 93

Pointecouteau, E., Santiago-Bautista, I., Douspis, M., et al. 2021, A&A, 651, A73

Postman, M., Lubin, L. M., Gunn, J. E., et al. 1996, AJ, 111, 615

Power, C., Knebe, A., & Knollmann, S. R. 2012, MNRAS, 419, 1576

Pratt, G. W., Arnaud, M., Biviano, A., et al. 2019, Space Sci. Rev., 215, 25

Press, W. H. & Schechter, P. 1974, ApJ, 187, 425

Rajpurohit, K., Vazza, F., van Weeren, R. J., et al. 2021, A&A, 654, A41

Rasia, E., Meneghetti, M., & Ettori, S. 2013, The Astronomical Review, 8, 40

Rasia, E., Tormen, G., & Moscardini, L. 2004, MNRAS, 351, 237

Reiprich, T. H., Basu, K., Ettori, S., et al. 2013, Space Sci. Rev., 177, 195

Riess, A. G., Casertano, S., Yuan, W., et al. 2021, ApJ, 908, L6

Robertson, H. P. 1935, ApJ, 82, 284

Rodríguez, A. C., Kacprzak, T., Lucchi, A., et al. 2018, Computational Astrophysics and Cosmology, 5, 4

Rodriguez-Gomez, V., Genel, S., Vogelsberger, M., et al. 2015, MNRAS, 449, 49

Rost, A., Kuchner, U., Welker, C., et al. 2021, MNRAS, 502, 714

Rothschild, T., Nagai, D., Aung, H., et al. 2022, MNRAS, 513, 333

Rouhiainen, A., Gira, M., Münchmeyer, M., Lee, K., & Shiu, G. 2023, arXiv e-prints, arXiv:2311.05217

Rowntree, A. R., Singh, A., Vincenzo, F., et al. 2024, MNRAS, 531, 3858

Rozo, E., Wechsler, R. H., Rykoff, E. S., et al. 2010, ApJ, 708, 645

Ruppin, F., Mayet, F., Macías-Pérez, J. F., & Perotto, L. 2019, MNRAS, 490, 784

Rykoff, E. S., Rozo, E., Busha, M. T., et al. 2014, ApJ, 785, 104

Salvati, L., Douspis, M., & Aghanim, N. 2020, A&A, 643, A20

Santos, J. S., Rosati, P., Tozzi, P., et al. 2008, A&A, 483, 35

Sarazin, C. L. 1988, X-ray emission from clusters of galaxies

Sarron, F., Adami, C., Durret, F., & Laigle, C. 2019, A&A, 632, A49

Schade, D., Lilly, S. J., Crampton, D., et al. 1995, ApJ, 451, L1

Schanz, A., List, F., & Hahn, O. 2023, arXiv e-prints, arXiv:2310.06929

Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, MNRAS, 446, 521

Schaye, J., Dalla Vecchia, C., Booth, C. M., et al. 2010, MNRAS, 402, 1536

Schaye, J., Kugel, R., Schaller, M., et al. 2023, MNRAS, 526, 4978

Schneider, P. 1996, MNRAS, 283, 837

Schneider, P. & Bartelmann, M. 1997, MNRAS, 286, 696

Sembolini, F., Yepes, G., De Petris, M., et al. 2013, MNRAS, 429, 323

Sereno, M., Umetsu, K., Ettori, S., et al. 2018, ApJ, 860, L4

Shan, H., Kneib, J.-P., Tao, C., et al. 2012, ApJ, 748, 56

Shin, T.-h., Clampitt, J., Jain, B., et al. 2018, MNRAS, 475, 2421

Silk, J. & White, S. D. M. 1978, ApJ, 226, L103

Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N., & Ganguli, S. 2015, arXiv e-prints, arXiv:1503.03585

Sohn, K., Lee, H., & Yan, X. 2015, in Advances in Neural Information Processing Systems, ed. C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, & R. Garnett, Vol. 28 (Curran Associates, Inc.)

Sousbie, T. 2011, MNRAS, 414, 350

Springel, V. 2005, MNRAS, 364, 1105

Springel, V. 2010, MNRAS, 401, 791

Springel, V., Pakmor, R., Pillepich, A., et al. 2018a, MNRAS, 475, 676

Springel, V., Pakmor, R., Pillepich, A., et al. 2018b, MNRAS, 475, 676

Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629

Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001, MNRAS, 328, 726

Steigman, G. 2008, arXiv e-prints, arXiv:0807.3004

Sunyaev, R. A. & Zeldovich, Y. B. 1970, Ap&SS, 7, 3

Sunyaev, R. A. & Zeldovich, Y. B. 1972, Comments on Astrophysics and Space Physics, 4, 173

Suto, D., Kitayama, T., Nishimichi, T., Sasaki, S., & Suto, Y. 2016, PASJ, 68, 97

Szandała, T. 2020, arXiv e-prints, arXiv:2010.09458

Tanimura, H., Aghanim, N., Douspis, M., & Malavasi, N. 2022, A&A, 667, A161

Tanimura, H., Aghanim, N., Kolodzig, A., Douspis, M., & Malavasi, N. 2020, A&A, 643, L2

Tanimura, H., Hinshaw, G., McCarthy, I. G., et al. 2019, MNRAS, 483, 223

Tegmark, M., Blanton, M. R., Strauss, M. A., et al. 2004, ApJ, 606, 702

Tempel, E., Stoica, R. S., Kipper, R., & Saar, E. 2016, Astronomy and Computing, 16, 17

Tempel, E., Tago, E., & Liivamägi, L. J. 2012, A&A, 540, A106

Teyssier, R. 2002, A&A, 385, 337

Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, ApJ, 688, 709

Tramonte, D., Ma, Y.-Z., Yan, Z., et al. 2023, arXiv e-prints, arXiv:2302.06266

Tröster, T., Ferguson, C., Harnois-Déraps, J., & McCarthy, I. G. 2019, MNRAS, 487, L24

Tuominen, T., Nevalainen, J., Tempel, E., et al. 2021, A&A, 646, A156

Ullmo, M., Aghnim, N., Decelle, A., & Aragon-Calvo, M. 2024, arXiv e-prints, arXiv:2403.02171

Ullmo, M., Decelle, A., & Aghanim, N. 2021, A&A, 651, A46

Umetsu, K. 2020, A&A Rev., 28, 7

Uzan, J.-P., Aghanim, N., & Mellier, Y. 2004, Phys. Rev. D, 70, 083533

Vallés-Pérez, D., Planelles, S., & Quilis, V. 2020, MNRAS, 499, 2303

Vazza, F., Roncarelli, M., Ettori, S., & Dolag, K. 2011, MNRAS, 413, 2305

Vega-Ferrero, J., Yepes, G., & Gottlöber, S. 2017, MNRAS, 467, 3226

Velliscig, M., Cacciato, M., Schaye, J., et al. 2015, MNRAS, 453, 721

Veronica, A., Reiprich, T. H., Pacaud, F., et al. 2024, A&A, 681, A108

Vikhlinin, A., Burenin, R. A., Ebeling, H., et al. 2009, ApJ, 692, 1033

Villaescusa-Navarro, F., Anglés-Alcázar, D., Genel, S., et al. 2021, ApJ, 915, 71

Villaescusa-Navarro, F., Hahn, C., Massara, E., et al. 2020, ApJS, 250, 2

Vogelsberger, M., Marinacci, F., Torrey, P., & Puchwein, E. 2020, Nature Reviews Physics, 2, 42

Walker, A. G. 1937, Proceedings of the London Mathematical Society, 42, 90

Walker, S. & Lau, E. 2022, in Handbook of X-ray and Gamma-ray Astrophysics, 13

Walker, S., Simionescu, A., Nagai, D., et al. 2019, Space Sci. Rev., 215, 7

Wan, J. T., Mantz, A. B., Sayers, J., et al. 2021, MNRAS, 504, 1062

Weinberger, R., Springel, V., & Pakmor, R. 2020, ApJS, 248, 32

Werner, N., Finoguenov, A., Kaastra, J. S., et al. 2008, A&A, 482, L29

White, S. D. M. 1976, MNRAS, 177, 717

White, S. D. M., Navarro, J. F., Evrard, A. E., & Frenk, C. S. 1993, Nature, 366, 429

Wicker, R., Douspis, M., Salvati, L., & Aghanim, N. 2023, A&A, 674, A48

Williamson, R., Benson, B. A., High, F. W., et al. 2011, ApJ, 738, 139

XRISM Science Team. 2020, arXiv e-prints, arXiv:2003.04962

Xu, X., Ho, S., Trac, H., et al. 2013, ApJ, 772, 147

Zaznobin, I. A., Burenin, R. A., Belinski, A. A., et al. 2023, Astronomy Letters, 49, 599

Zel'dovich, Y. B. 1970, A&A, 5, 84

Zubeldia, Í. & Bolliet, B. 2024, arXiv e-prints, arXiv:2403.09589

Zubeldia, Í., Rotti, A., Chluba, J., & Battye, R. 2022, arXiv e-prints, arXiv:2204.13780

Zwicky, F. & Kowal, C. T. 1968, "Catalogue of Galaxies and of Clusters of Galaxies", Volume VI