

Université de Montréal

**Caractérisation des amas de galaxies avec des
méthodes d'apprentissage automatique**

par

Maria Sadikov

Département de Physique
Faculté des arts et des sciences

Mémoire présenté en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en Astrophysique et Astronomie

17 août 2023

Université de Montréal

Faculté des arts et des sciences

Ce mémoire intitulé

Caractérisation des amas de galaxies avec des méthodes d'apprentissage automatique

présenté par

Maria Sadikov

a été évalué par un jury composé des personnes suivantes :

Paul Charbonneau

(président-rapporteur)

Julie Hlavacek-Larrondo

(directeur de recherche)

Laurence Perreault-Levasseur

(codirecteur)

Yashar Hezavehe

(membre du jury)

Résumé

Les amas de galaxies sont les plus grandes structures gravitationnellement liées de l'Univers. Ils sont communément séparés en trois catégories, basées sur la distribution du gaz intra-amas. Ce gaz peut être très concentré vers le centre de l'amas, il peut être réparti dans l'amas de manière plutôt uniforme, ou encore il peut avoir une distribution légèrement piquée vers le centre dans un cas intermédiaire. Une autre distinction entre les trois catégories est l'interaction entre le trou noir supermassif se trouvant au centre de l'amas de galaxies et le gaz intra-amas environnant. Dans le cas de la première catégorie, lorsque le gaz est concentré au centre de l'amas, le trou noir est dit 'actif'. Il produit alors des jets, qui à leur tour injectent de l'énergie dans le gaz intra-amas sous forme d'ondes sonores, d'ondes de choc et de turbulence. Les amas de galaxies offrent donc une opportunité très intéressante pour étudier ce mécanisme d'échange d'énergie. Afin de mieux caractériser ces processus, il est essentiel d'avoir des méthodes robustes pour classifier les amas de galaxies selon les trois catégories. Il existe plusieurs propriétés pouvant être utilisées comme métriques de classification, mais celles-ci ne sont pas toujours en accord les unes avec les autres. Ces propriétés ont été étudiées pour des petits échantillons d'amas de galaxies, analysés de manière individuelle avec des méthodes traditionnelles. Cependant, avec le développement de puissants instruments d'observation tels que eROSITA, on s'attend à obtenir des échantillons contenant environ 100 000 amas de galaxies. Étant donné la taille de ces ensemble de données, il devient nécessaire d'avoir un moyen rapide, efficace et automatique pour les traiter. On a donc recours à l'apprentissage automatique pour accélérer l'analyse.

Ce mémoire présente une analyse des propriétés du gaz intra-amas avec des méthodes d'apprentissage automatique. On se sert des simulations cosmologiques IllustrisTNG pour obtenir des images en rayons X d'amas de galaxies, à partir desquelles on construit notre ensemble de données. On s'intéresse à cinq propriétés du gaz intra-amas contenu dans les amas de galaxies, qui sont couramment utilisées comme métriques de classification : le temps de refroidissement central, la densité électronique centrale, l'excès d'entropie centrale, le paramètre de concentration de la brillance de surface et le paramètre de courbure du profil

de densité. On explore les relations entre ces différentes métriques, puis on implémente un réseau de neurones qui vise à prédire leur valeur à partir d'une image en rayons X d'un amas de galaxies. Notre réseau atteint un pourcentage d'erreur moyen de 1.8% pour les prédictions de la métrique la plus performante, c'est-à-dire le temps de refroidissement central. Ensuite, afin d'estimer les incertitudes sur les résultats obtenus, on effectue une analyse probabiliste de nos prédictions à l'aide de la méthode de l'inférence sans vraisemblance. On utilise également une méthode de partitionnement de données qui rassemble les images en rayons-X en trois groupes distincts ; on constate que ce regroupement corrèle fortement avec la division des mêmes images en utilisant le paramètre de concentration comme métrique de classification. L'ensemble de ce travail permet de conclure que le temps de refroidissement central et la concentration sont les métriques se prêtant le mieux à une analyse avec des méthodes d'apprentissage automatique, ainsi que de mettre en place les outils qui serviront à caractériser les futurs échantillons d'amas de galaxies.

Mots-clés : amas de galaxies - apprentissage automatique - rayons X - simulations cosmologiques - trous noirs - réseaux de neurones

Abstract

Galaxy clusters are the largest gravitationally bound structures of the universe. They are commonly divided into three categories, based on the distribution of the intracluster gas. In one case, the gas is strongly concentrated towards the center of the cluster. In another case, it is rather uniformly dispersed through the cluster. In a third intermediate case, the distribution is slightly peaked towards the center. The three categories also differ by the interaction between the gas and the supermassive black hole located at the center of the cluster. In the first category, the black hole is said to be ‘active’ and it produces jets that heat up the intracluster gas through shock waves, sound waves and turbulence. The feedback mechanism from the black hole is not entirely understood, and galaxy clusters offer a valuable opportunity to study this energy transfer mechanism in more detail.

Numerous properties can serve as classification metrics, but they are not always consistent with one another. Moreover, traditional methods used to extract those properties are time-consuming and have only been applied to small samples. With the advent of powerful X-ray observatories such as eROSITA, we expect to obtain large galaxy clusters datasets ($\sim 100\,000$). Given the size of the datasets and the number of parameters to consider, machine learning methods are needed to accelerate the data processing.

This thesis presents an analysis of intracluster gas properties with machine learning techniques. We use the galaxy clusters from the IllustrisTNG cosmological simulations to create the X-ray images that make up our dataset. We study five properties of the hot gas in galaxy clusters that are commonly used as classification metrics; the central cooling time, the central electron density, the central entropy excess, the concentration of the surface brightness and the cusiness parameter, which represents the slope of the density profile. We explore the correlations between the different metrics, and implement a neural network that predicts their values from an X-ray image of a galaxy cluster. The network achieves a mean percentage error of 1.8% on the central cooling time predictions, making it the best-performing metric. In order to get uncertainty estimates, we perform a probabilistic analysis of the network predictions using simulation-based inference. We also

use a clustering approach that groups the X-ray images into three separate groups; we note that those groups are consistent with classification based on the concentration parameter. Our results show that the central cooling time and the concentration are the metrics that lend themselves the best to a machine learning analysis of galaxy cluster images. This project aims to lay the groundwork for characterizing future galaxy cluster observations.

Keywords: galaxy clusters - machine learning - X-rays - cosmological simulations - black holes - neural network

Table des matières

Résumé	v
Abstract	vii
Table des figures	xiii
Liste des sigles et des abréviations	xv
Remerciements	xix
Chapitre 1. Mise en contexte théorique	1
1.1. Introduction	1
1.1.1. Description de ce mémoire	2
1.1.2. Déclaration de l'étudiante	3
1.2. Amas de galaxies	3
1.2.1. Milieu intra-amas	4
1.2.2. Trou noir supermassif et noyau actif de galaxie	7
1.2.3. Amas de galaxies à coeur froid et à coeur non-froid	10
1.2.3.1. Courant de refroidissement	11
1.2.3.2. Classification des amas de galaxies	12
1.3. Observations d'amas de galaxies	15
1.3.1. Astronomie rayons X	16
1.3.2. Observatoire de rayons X <i>Chandra</i>	17
1.4. Simulations cosmologiques	19
1.4.1. Notions de cosmologie	19
1.4.2. Illustris TNG	20
1.5. Apprentissage automatique	22
1.5.1. Réseaux de neurones artificiels	22
1.5.1.1. Rétropropagation	24
1.5.1.2. Ensembles d'entraînement, de validation et de test	26

1.5.2.	Réseaux de neurones convolutifs	26
1.5.3.	Réseaux neuronaux résiduels (ResNet)	28
1.5.4.	Simulation-based inference	29
	Bibliographie	30
Chapter 2. Galaxy cluster characterization with machine learning techniques		33
	Abstract	34
2.1.	Introduction	34
2.2.	Simulated Data and Metrics	38
2.2.1.	Simulated Data	38
2.2.1.1.	IllustrisTNG Clusters	38
2.2.1.2.	Center of the cluster	38
2.2.1.3.	Mock <i>Chandra</i> Observations	39
2.2.1.4.	Image preprocessing	40
2.2.2.	Galaxy cluster classification metrics	41
2.2.2.1.	Central cooling time	41
2.2.2.2.	Central electron density	44
2.2.2.3.	Central entropy excess	44
2.2.2.4.	Concentration parameter	46
2.2.2.5.	Cuspiness	46
2.2.2.6.	Comparison of the metrics	47
2.3.	Clustering	47
2.3.1.	Methods	47
2.3.2.	Results	49
2.4.	Regression	50
2.4.1.	Methods	50
2.4.2.	Results	53
2.5.	Discussion	56
2.5.1.	Simulation-based inference	56
2.5.2.	Fraction of cool cores in IllustrisTNG	58
2.5.3.	Metric comparison	62
2.5.4.	Implications for future surveys	63

2.6. Conclusion.....	64
Acknowledgements.....	65
Software.....	66
Bibliography.....	67
Chapitre 3. Conclusion.....	71
Annexe A. Congrès où l'étudiante a présenté ses résultats.....	74
A.1. Rencontre annuelle du Centre de Recherche en Astrophysique du Québec (CRAQ).....	74
A.2. Symposium Annuel de Physique pour un Avenir en Recherche et en Industrie (SAPHARI).....	74
A.3. Octobre Numérique d'IVADO.....	74
A.4. Rencontre annuelle de la Société Canadienne d'Astronomie (CASCA).....	75
A.5. Poster Symposium Targeting Early-career Researchers (PoSTER).....	75
A.6. Symposium Annuel de Physique pour un Avenir en Recherche et en Industrie (SAPHARI).....	75
A.7. Centenaire du Département de Physique de l'Université de Montréal.....	75

Table des figures

1.1	Image de l'amas de galaxies SMACS 0723.....	2
1.2	Image composite de l'amas de Persée.....	5
1.3	Puissance émise par un trou noir selon le taux d'accrétion.....	9
1.4	Jets relativistes de l'AGN.....	10
1.5	Amas à coeur froid et à coeur non-froid.....	11
1.6	Assemblage de miroirs du télescope spatial <i>Chandra</i>	18
1.7	Télescope spatial <i>Chandra</i>	18
1.8	Exemples de fonctions d'activation.....	23
1.9	Exemple de courbe de fonction de coût.....	27
1.10	Saut de connexion dans un réseau neuronal résiduel.....	28
2.1	Example of cluster peoperties for a cool core cluster and a non-cool core cluster .	42
2.2	Distribution of the metrics with mass.....	45
2.3	Comparison of the cluster properties for our sample.....	48
2.4	Results of the clustering algorithm.....	49
2.5	Two types of residual blocks used in the network.....	50
2.6	Neural network architecture.....	51
2.7	Predicted values as a function of true values for all five cluster properties.....	52
2.8	Confusion matrices for the cooling time predictions.....	55
2.9	Confusion matrices for the central electron density predictions.....	55
2.10	Confusion matrices for the central entropy excess predictions.....	55
2.11	Confusion matrices for the concentration predictions.....	56
2.12	Confusion matrices for the cuspiness predictions.....	56
2.13	SBI posteriors for several values of cooling time predictions.....	58
2.14	Example of posteriors for all metrics for a CC cluster.....	59

2.15	Example of posteriors for all metrics for a NCC cluster	60
2.16	Results for the concentration with the retrained network	63

Liste des sigles et des abréviations

Mpc	Mégaparsec
AGN	Noyau actif de galaxie, de l'anglais <i>Active Galactic Nucleus</i>
BCG	Galaxie dominante de l'amas, de l'anglais <i>Brightest Cluster Galaxy</i>
ICM	Milieu intra-amas, de l'anglais <i>Intracluster Medium</i>
CC	Amas à coeur froid, de l'anglais <i>Cool core cluster</i>
NCC	Amas à coeur non-froid, de l'anglais <i>Non-cool core cluster</i>
WCC	Amas à coeur froid modéré, de l'anglais <i>Weak cool core cluster</i>
CRSNG	Conseil de recherche en sciences naturelles et en génie du Canada
NASA	Agence spatiale des États-Unis, de l'anglais <i>National Aeronautics and Space Administration</i>
ESA	Agence spatiale européenne, de l'anglais <i>European Space Agency</i>

CSA	Agence spatiale canadienne, de l'anglais <i>Canadian Space Agency</i>
STScI	Institut scientifique des télescopes spatiaux, de l'anglais <i>Space Telescope Science Institute</i>
NRAO	Observatoire national de radioastronomie, de l'anglais <i>National Radio Astronomy Observatory</i>
SDSS	Relevé du ciel SDSS, de l'anglais <i>Sloan Digital Sky Survey</i>
ESA	Agence spatiale européenne, de l'anglais <i>European Space Agency</i>
HEAO	Série de télescopes pour l'astronomie des hautes énergies, de l'anglais <i>High Energy Astronomy Observatory</i>
CXC	Centre des rayons X de <i>Chandra</i> , de l'anglais <i>Chandra X-Ray Center</i>
HRMA	Assemblage de miroirs à haute résolution de <i>Chandra</i> , de l'anglais <i>High Resolution Mirror Assembly</i>
HRC	Caméra à haute résolution de <i>Chandra</i> , de l'anglais <i>High Resolution Camera</i>
ACIS	Spectromètre imageur de <i>Chandra</i> , de l'anglais <i>Advanced CCD Imaging Spectrometer</i>

CCD	Capteur photographique CCD, de l'anglais <i>Charged-coupled device</i>
NGST	Télescope spatial de prochaine génération, <i>Next Generation Space Telescope</i>
FoF	Algorithme FoF, de l'anglais <i>friends-of-friends</i>
SOXS	Observations simulées de sources en rayons X, de l'anglais <i>Simulated Observations of X-ray Sources</i>
ReLU	Fonction d'activation unitaire linéaire rectifiée, de l'anglais <i>Rectified Linear Unit</i>
ResNet	Réseau neuronal résiduel, de l'anglais <i>Residual Network</i>
SBI	Inférence sans vraisemblance, de l'anglais <i>Simulation-based inference</i>
LFI	Inférence sans vraisemblance, de l'anglais <i>Likelihood-free inference</i>
eROSITA	Télescope eROSITA, de l'anglais <i>extended ROentengen Survey with an Imaging Telescope Array</i>
MPE	Institut Max-Planck de physique extraterrestre, de l'allemand <i>Max-Planck-Institut für extraterrestrische Physik</i>
SRG	Observatoire spatial SRG, de l'anglais <i>Spectrum-Roentgen-Gamma</i>

CNN	Réseau de neurones convolutif, de l'anglais <i>convolutional neural network</i>
BAcc	Précision balancée, de l'anglais <i>Balanced Accuracy</i>
Λ CDM	Modèle cosmologique Λ CDM, de l'anglais <i>Lambda Cold Dark Matter</i>
CIAO	Logiciel d'analyse d'observations <i>Chandra</i> , de l'anglais <i>Chandra Interactive Analysis of Observations</i>
MCMC	Méthode de Monte-Carlo par chaînes de Markov, de l'anglais <i>Markov Chain MonteCarlo</i>
PCA	Analyse en composantes principales, de l'anglais <i>Principal Component Analysis</i>
HIFLUGCS	Catalogue d'amas de galaxies HIFLUGCS, de l'anglais <i>Highest X-ray FLUX Galaxy Cluster Sample</i>
ESZ	Catalogue ESZ, de l'anglais <i>Early Sunyaev-Zel'dovich</i>
FRQNT	Fonds de recherche du Québec - Nature et technologies
IVADO	Institut de valorisation des données
CRAQ	Centre de recherche en astrophysique du Québec

Remerciements

Je tiens tout d'abord à remercier mes superviseuses, Julie Hlavacek-Larrondo et Laurence Perreault-Levasseur, de m'avoir guidé tout au long de ce projet et d'avoir répondu à mes questions. Merci pour votre expertise et tous vos précieux conseils, que ceux-ci portent sur l'astrophysique, sur le monde de la recherche ou sur la vie en général. Vous avez été des mentores extraordinaires.

Merci à Carter Rhea, pour tes réponses à mes innombrables questions, pour ton écoute et tes conseils.

À mes ami.e.s et collègues Benjamin, Carter, Joseph, Ève, Charles, Andi, Alex, David, Myriam, Ronan et tous les autres membres du groupe X-TRA et de l'Institut Ciela, merci pour votre soutien et votre aide.

Merci à Florence, Julie et Nico, mes ami.e.s du baccalauréat. Ces trois années n'auraient pas été les mêmes sans vous.

Merci à tou.te.s les professeur.e.s qui ont marqué mon parcours et qui m'ont encouragée à poursuivre en physique. Merci à Mike McDonald et à son groupe de recherche, de m'avoir partagé votre expertise et de m'avoir accueillie au MIT pour un été. Ça a été une expérience extrêmement enrichissante.

Un merci spécial à Juliette, ma plus fidèle compagne d'étude et ma précieuse amie. Merci de me faire rire, merci d'être toujours là pour moi, merci pour tous les doux moments qu'on passe ensemble.

Un merci spécial à Camille, pour les nombreuses journées de rédaction, pour tes anecdotes sans fin, pour tes analogies impeccables, pour les petits soupers et les grosses discussions en ta compagnie. Merci d'être dans ma vie.

À mon ami Zheng Yu, merci de toujours me comprendre et de m'encourager.

À mes amies du secondaire, Juliette, Kathryn, Séréna, Jeanne, Roxana et Maya : merci pour votre soutien et votre amitié au fil des années, merci de m'inspirer à tous les jours.

Merci à mes ami.e.s (Cassiopée, Jules, Audrey, Mikaël, Maëlle, Olivia, Marie-Laurence, Louis-Olivier et tou.te.s les autres) pour vos encouragements et votre présence dans ma vie.

Merci à ma maman Anna, pour ton amour et ton support inconditionnels depuis le début.

Merci aussi à mes chats, feu Chevalier d'Éon et Tonnerre; il est bien plus agréable d'écrire et de coder avec un de vous à mes côtés.

Finalement, merci à mon copain Achille, pour ta patience, ton support et ton amour. Je suis si heureuse de partager ma vie avec toi.

Chapitre 1

Mise en contexte théorique

1.1. Introduction

Les amas de galaxies sont des objets fascinants qui nous aident à comprendre notre Univers sur plusieurs aspects. Ils peuvent servir à répondre à de nombreuses questions astrophysiques et cosmologiques. Dans le modèle standard de la cosmologie, on assume que les amas de galaxies se sont formés par effondrement gravitationnel suite à une accumulation de matière baryonique et de matière sombre dans les zones de surdensité de l'Univers. De telles zones représentent des perturbations dans la densité de l'Univers, autrement homogène (Allen et al. 2011). En comptant le nombre d'amas de galaxies d'une masse donnée qui existent aujourd'hui, on peut imposer des contraintes sur la valeurs des paramètres cosmologiques qui définissent l'amplitude et l'évolution des perturbations de la densité (Borgani et al. 2001, Pierpaoli et al. 2003). Par ailleurs, les amas de galaxies peuvent jouer le rôle de lentille gravitationnelle, déviant la lumière des objets situés en arrière-plan. Cette déviation peut produire des images déformées, étirées et agrandies des objets cachés en arrière-plan, ce qui nous permet de mieux les étudier (e.g. Kneib & Natarajan 2011, Umetsu 2020). De plus, étant principalement composés de matière sombre, les amas de galaxies sont des excellents sujets d'étude pour tenter d'en apprendre davantage sur celle-ci. Par exemple, les astronomes se tournent souvent vers les amas de galaxies pour tracer la distribution de masse de la matière sombre (e.g. Sand et al. 2004, Jauzac et al. 2018) ou pour faire des tests par rapport à sa nature-même (e.g. Riemer-Sorensen et al. 2007, Bradač et al. 2008). Un exemple d'image d'amas de galaxies est montré à la figure 1.1.

Depuis les premières observations d'amas de galaxies, les astronomes les ont étudiés dans plusieurs longueurs d'onde avec divers instruments terrestres et spatiaux. Les amas de galaxies peuvent être étudiés dans le spectre visible, dans les rayons X, dans le domaine radio ainsi que dans l'infrarouge proche, révélant différentes parties de leur structure selon la longueur d'onde choisie. Dans le cadre de ce mémoire, on s'intéresse à l'émission dans les



Figure 1.1 – Image de l’amas de galaxies SMACS 0723, prise avec le télescope James Webb. Ce champ profond a été pris avec l’instrument NIRC*am* (de l’anglais *Near Infrared Camera*), couvrant les longueurs d’onde de l’infrarouge proche. Cet amas de galaxies très massif agit comme lentille gravitationnelle et déforme les images de galaxies en arrière-plan ; on peut voir les arcs des galaxies étirées. (Source : NASA, ESA, CSA, STScI)

rayons X des amas de galaxies, ayant pour source le gaz chaud qui comble l’espace entre les galaxies. Plusieurs modèles physiques ont été développés pour expliquer les phénomènes qui se produisent dans un amas de galaxies, mais il y en a encore plusieurs éléments qui manquent à notre compréhension. Un de ces éléments est la manière dont les trous noirs supermassifs au centre des amas de galaxies interagissent avec le gaz environnant. Pour répondre à cette question, il est utile d’étudier en plus grand détail certaines propriétés physiques du gaz. Dans le cadre de ce mémoire, j’utilise l’apprentissage automatique pour effectuer une analyse de ces propriétés pour une population d’amas de galaxies extraits de simulations cosmologiques.

1.1.1. Description de ce mémoire

L’objectif de ce mémoire est d’analyser et de comparer cinq propriétés physiques du gaz (le temps de refroidissement, la densité électronique, l’entropie, le paramètre de concentration, le paramètre de courbure du profil de densité) contenu dans les amas de galaxies, puis ensuite de concevoir et entraîner un réseau d’apprentissage automatique qui prédit ces cinq propriétés à partir d’images en rayons X d’amas de galaxies.

La section 1.2 est une introduction théorique aux amas de galaxies, et les propriétés physiques pertinentes sont présentées à la section 1.2.3.2. La section 1.3 détaille les observations des amas de galaxies, puis les simulations cosmologiques utilisées sont présentées à la section 1.4. Les concepts d'apprentissage automatique sont expliqués à la section 1.5. Finalement, le chapitre 2 présente la méthodologie développée dans le cadre de ce mémoire, ainsi que les résultats obtenus.

1.1.2. Déclaration de l'étudiante

Je, Maria Sadikov, déclare que l'entiereté du travail présenté dans ce mémoire est le mien. J'ai effectué la revue de littérature de ce mémoire et rédigé le chapitre 1. Les sources des figures que je n'ai pas produites moi-mêmes sont identifiées dans la description des figures.

Pour l'article rédigé au chapitre 2, le crédit de l'idée scientifique originale revient à Julie Hlavacek-Larrondo et Laurence Perreault-Levasseur. Une partie du travail a été réalisée lors d'un stage au *Massachusetts Institute of Technology* supervisé par Michael McDonald. Les données de simulations proviennent du projet IllustrisTNG. Plusieurs fragments de code servant à extraire et traiter les données de simulations ont été écrits par des membres de l'équipe d'Illustris, et adaptés par Carter Rhea. Les images simulées en rayons X ont été créés avec les logiciels `pyXSIM` et `SOXS`, créés par John Zuhone. L'idée originale pour le type de réseau de neurones utilisé (un ResNet) a été développée par Kaiming He, Xiangyu Zhang, Shaoqing Ren et Jian Sun.

J'ai trié et analysé les données de simulations pour produire les images et extraire les propriétés physiques pertinentes. J'ai implémenté le réseau de neurones ; j'ai effectué l'entraînement et la recherche d'hyperparamètres. J'ai produit et analysé les résultats, puis j'ai rédigé l'article présenté au chapitre 2. J'ai bénéficié des conseils de Julie Hlavacek-Larrondo, Laurence Perreault-Levasseur et Carter Rhea.

1.2. Amas de galaxies

Les amas de galaxies sont les plus grandes structures gravitationnellement liées de l'Univers. Ce sont des ensembles de galaxies liées par la gravité, et ils peuvent contenir des centaines voir des milliers de galaxies ; en-dessous de ce nombre on parle plutôt d'un groupe de galaxies. Ils sont gouvernés par le théorème du viriel, qui stipule que l'énergie cinétique (K) moyenne est égale au négatif de la moitié de l'énergie potentielle (U) moyenne :

$$\langle K \rangle = -\frac{1}{2}\langle U \rangle. \quad (1.2.1)$$

Soumis à un énorme puits gravitationnel, les amas de galaxies contiennent une grande quantité d'énergie, dont une partie est dégagée sous forme d'émission électromagnétique. Ils sont parmi les sources de rayons X les plus lumineuses de l'Univers, ayant une luminosité bolométrique typique entre 10^{43} et 10^{46} erg s⁻¹ (p. ex. Sarazin 1986).

Les amas de galaxies ont un rayon de l'ordre de quelques mégaparsecs (Mpc) et une masse d'environ $10^{14} - 10^{15} M_{\odot}$. Ils sont composés de galaxies, de matière sombre et de gaz intra-amas, qui représentent respectivement $\sim 3\%$, $\sim 84\%$ et $\sim 13\%$ de la masse (Mohr et al. 1999, Vikhlinin et al. 2006, Umetsu et al. 2009). Au centre gravitationnel de l'amas se trouve la galaxie dominante de l'amas (BCG - de l'anglais *Brightest Cluster Galaxy*), qui est plus massive et plus lumineuse que les autres galaxies de l'amas. La figure 1.2 présente une image composite d'un des amas de galaxies les plus étudiés ainsi que la BCG au centre : l'amas de Persée et la galaxie NGC 1275. Le bleu représente l'émission en rayons X captée avec le télescope spatial *Chandra*, qui provient principalement du gaz intra-amas. Souvent appelé milieu intra-amas (ICM, de l'anglais *intracluster medium*), c'est le gaz dans un amas de galaxies qui remplit l'espace entre les galaxies. Le rose représente l'émission dans le domaine radio captée par le Sloan Digital Sky Survey et le Very Large Array. Le blanc représente l'émission dans le domaine du visible. D'une part, on peut voir l'émission captée par le *Sloan Digital Sky Survey* qui provient des galaxies individuelles, et d'une autre part nous avons l'émission de la BCG et de ses filaments, imagés par le télescope Hubble.

Dans le cadre de ce mémoire, les composantes qui nous intéressent le plus sont le gaz intra-amas ainsi que le trou noir supermassif situé au centre de la BCG.

1.2.1. Milieu intra-amas

L'ICM représente la majorité de la masse baryonique (non constituée de matière sombre) de l'amas de galaxies. Il est principalement composé d'hydrogène et d'hélium ionisés, et peut être considéré comme un plasma chaud et diffus. Il possède une température typique de l'ordre de 10^7 K (p. ex. Sarazin 1986), et une densité électronique de l'ordre de 10^{-3} cm⁻³ (p. ex. Sarazin 1986). Grâce à cette haute température, l'ICM émet surtout dans les rayons X. Le processus d'émission dominant est le Bremsstrahlung, un rayonnement électromagnétique produit par le ralentissement d'une particule chargée.

Les amas de galaxies sont souvent définis en fonction de leur masse M_{500} , la masse contenue à l'intérieur d'un rayon R_{500} , qui représente le rayon auquel la densité est 500 fois supérieure à la densité critique de l'univers ρ_c . La densité critique est définie telle que $\rho_c = \frac{3H^2}{8\pi G}$, où H est le paramètre de Hubble caractérisant l'expansion de l'univers et G est



Figure 1.2 – Image reconstituée de l’amas de galaxies de Persée, composée d’une image en rayons X captée par le télescope *Chandra* (en bleu) et d’un ensemble d’images dans le domaine radio (en rose), captées par le Very Large Array. Les galaxies individuelles (en blanc) ont été imagées dans le domaine optique par le *Sloan Digital Sky Survey*. Au centre se trouve la galaxie elliptique géante NGC 1275, dont les filaments ont été imagés dans le visible par le télescope Hubble. Cette image a remporté le *Prix du public Découverte*, parrainé par Radio-Canada, de l’édition 2017 du concours *La preuve par l’image* de l’Acfas et du CRSNG. (Source : M. Gendron-Marsolais, J. Hlavacek-Larrondo, M. Pivin-Lapointe ; NRAO, NASA, SDSS.)

la constante gravitationnelle universelle (e.g. Ryden 2003). On peut déterminer la masse d’un amas de galaxies en considérant l’ICM comme étant en équilibre hydrostatique, par la logique qui suit. Si l’ICM subit une légère perturbation, il faudra un temps t_s pour que les ondes sonores traversent le milieu et que celui-ci se réajuste pour retourner à l’équilibre. Ce temps de propagation du son t_s est donné par :

$$t_s = \frac{D}{c_s}, \quad (1.2.2)$$

où D est le diamètre de l’amas, et c_s est la vitesse du son pour un gaz idéal, qui peut être exprimée comme :

$$c_s = \frac{dP}{d\rho} = \gamma \frac{P}{\rho} = \gamma \frac{k_B T}{\mu m_H}, \quad (1.2.3)$$

où P est la pression, ρ est la densité de masse, k_B est la constante de Boltzmann, T est la température du gaz, μ est le poids moléculaire moyen, m_H est la masse de l'atome d'hydrogène et γ est l'indice adiabatique. On peut adopter une valeur de $5/3$ pour γ (approximation pour un gaz monoatomique idéal), et une valeur de 0.61 pour μ (valeur pour l'abondance élémentaire typique d'un amas). Pour des propriétés typiques d'un amas de galaxies, on obtient un temps de propagation du son t_s de l'ordre de quelques 10^8 années, ce qui est nettement inférieur à l'âge d'un amas de galaxies (de l'ordre de 10^{10} années). Cela veut dire que l'amas aura pleinement le temps de retourner à l'équilibre après une perturbation. On peut donc traiter l'ICM comme étant en équilibre hydrostatique, et appliquer l'équation d'équilibre hydrostatique :

$$\frac{1}{\rho} \frac{dP}{dr} = -\frac{GM(r)}{r^2} \quad (1.2.4)$$

où $M(r)$ est la masse de l'amas contenue à l'intérieur d'un rayon r . La seule dépendance positionnelle est en r , car on considère que les amas ont une symétrie sphérique. En utilisant la substitution $P = \rho k_B T / \mu m_H$, où $\rho = \rho(r)$, et $T = T(r)$, on arrive à

$$M(r) = -\frac{k_B r^2}{G \mu m_H} \left(\rho \frac{dT}{dr} + \frac{T}{\rho} \frac{d\rho}{dr} \right). \quad (1.2.5)$$

Les profils de température et de densité peuvent être obtenus à partir des observations spectroscopiques du gaz en rayons X. Puisqu'on connaît l'énergie des photons recueillis sur chaque pixel du détecteur, on peut obtenir un spectre de la source observée pour chacun des pixels, c'est-à-dire l'intensité de la lumière émise (i.e. le nombre de photons émis) en fonction de l'énergie. Ensuite, en partant du centre de l'amas, on peut tracer des anneaux (assignant chaque pixel à un anneau) et construire un spectre pour chaque anneau. On peut ensuite réaliser un ajustement aux données spectrales en utilisant les modèles connus pour une densité et une température donnée, convolués avec la réponse instrumentale du détecteur. Ceci permet d'extraire les valeurs optimales de densité et de température pour chaque anneau. On obtient ainsi les profils de densité et de température, et on peut donc estimer la masse à l'aide de l'équation 1.2.5.

Une autre méthode indépendante pour estimer la masse des amas de galaxies se sert des lentilles gravitationnelles (Bartelmann & Schneider 2001, Bartelmann 2010). Cette méthode se base sur la théorie de la relativité générale, qui nous dit que la présence d'une masse courbe l'espace-temps. Cette courbure fait dévier la trajectoire des rayons lumineux. Plus la masse d'un objet est grande, plus l'effet de courbure est important. Les amas de galaxies sont des objets extrêmement massifs, et peuvent donc faire dévier de manière significative

la lumière provenant d’une source en arrière-plan. On peut alors voir apparaître une ou plusieurs images déformées de la source d’arrière-plan dans nos observations. La distorsion subie est une mesure du champ gravitationnel de l’amas de galaxies (qui joue le rôle de lentille). En quantifiant la déformation, on peut donc estimer la masse de l’amas (e.g. Broadhurst et al. 1995).

Les estimations de masse obtenues par ces deux méthodes sont généralement du même ordre de grandeur, les valeurs exactes différant par quelques dizaines de pourcents (McNamara & Nulsen 2007).

1.2.2. Trou noir supermassif et noyau actif de galaxie

Au centre de la BCG, se trouve un trou noir supermassif. Les trous noirs sont des objets extrêmement denses et compacts. Ils exercent une énorme force de gravité en raison de leur masse élevée. Ils sont dits supermassifs quand ils ont une masse supérieure à $10^6 M_{\odot}$, et on en trouve au centre de toutes les galaxies massives (p. ex. Merritt 2013). En raison de sa force de gravité, un trou noir peut accréter de la matière. Si cette accréation est suffisamment importante, on dit que la galaxie qui abrite le trou noir est active, et qu’elle contient un noyau actif de galaxie (AGN, de l’anglais *active galactic nucleus*). Des particules relativistes, du gaz, de la poussière et du plasma vont orbiter autour du trou noir et se rassembler pour former un disque d’accréation. Le noyau actif de galaxie réfère alors à l’ensemble du système composé du trou noir et du disque d’accréation. Les particules du disque entrent en collision les unes avec les autres, ce qui tend à diminuer leur moment cinétique. Par conservation du moment cinétique, ces particules vont se rapprocher du trou noir pour diminuer leur rayon orbital. Ceci diminue leur énergie potentielle gravitationnelle, et se traduit par un gain d’énergie thermique (White & Holt 1982); les parties centrales du disque d’accréation sont chauffées davantage et émettent des photons de plusieurs longueurs d’onde.

Les AGNs sont parmi les sources d’émission électromagnétique les plus brillantes de l’univers. Cette émission s’inscrit dans un cycle de refroidissement et de réchauffement du gaz intra-amas, et sert à compenser les pertes d’énergie de l’environnement. On parle donc de rétroaction (*feedback*) de l’AGN, et on distingue deux modes de rétroaction : le mode radio (aussi appelé mode cinétique, mécanique, ou radiativement-inefficace) et le mode radiatif (aussi appelé mode quasar). C’est le taux d’accréation du trou noir qui détermine le mode de rétroaction, et la distinction peut être quantifiée par le ratio d’Eddington η :

$$\eta = \frac{L_{AGN}}{L_{Edd}} \quad (1.2.6)$$

où L_{AGN} est la luminosité bolométrique de l'AGN et L_{Edd} est la luminosité d'Eddington, définie comme la luminosité maximale que peut atteindre un objet sans briser l'équilibre hydrostatique. La luminosité d'Eddington est définie telle que :

$$L_{Edd} = \frac{M4\pi Gm_p c}{\sigma_T} \quad (1.2.7)$$

où M est la masse de l'objet accréant la matière, m_p est la masse du proton, c est la vitesse de la lumière, et σ_T est la section efficace de la diffusion de Thomson ($\sigma_T = \frac{8\pi}{3}(\frac{e^2}{4\pi\epsilon_0 m_e c^2})^2$, où e est la charge élémentaire, ϵ_0 est la permittivité du vide, et m_e est la masse électronique). Pour un faible taux d'accrétion et donc une petite valeur de η , c'est-à-dire jusqu'à environ 0.01-0.1 (Churazov et al. 2005), la rétroaction de l'AGN est en mode radio (cinétique) et l'énergie du trou noir est libérée sous forme de jets de particules. Lorsque le taux d'accrétion et la valeur de η augmentent, la puissance des jets diminue et l'émission radiative prend de l'importance. C'est alors le mode radiatif de rétroaction qui domine. Cette transition du mode radio vers le mode radiatif avec l'augmentation du taux d'accrétion est représentée à la figure 1.3, tirée de Churazov et al. (2005).

La rétroaction de l'AGN en mode radio se produit en présence d'un disque d'accrétion ayant un fort champ magnétique. Suite à une interaction entre le champ magnétique et la matière en rotation formant le disque d'accrétion, un champ magnétique est créé le long de l'axe de rotation du disque. La nature exacte des processus physiques en jeu demeure incertaine (e.g. Chibueze et al. 2021, Osinga et al. 2022). Les lignes de champ magnétique se forment à partir de la surface du disque d'accrétion, et s'étendent à de grandes distances. L'énergie et le moment cinétique du disque d'accrétion peuvent ensuite être extraits de manière électromagnétique (Blandford & Payne 1982, Blandford & Znajek 1977). Du plasma énergétique est alors collimaté puis éjecté à des vitesses relativistes, perpendiculairement au disque d'accrétion (Blandford & Rees 1974). Ces structures étendues sont les jets produits par l'AGN. Ils ont une forte émission synchrotron, émise lorsqu'un champ magnétique dévie la trajectoire d'une particule chargée qui s'y déplace. Cette émission rend les jets principalement visibles dans le domaine radio. En se propageant dans le milieu environnant, les jets seront freinés et à leurs extrémités se formeront des bulles de plasma relativiste. Ces bulles peuvent ensuite injecter de l'énergie dans l'environnement à travers la propagation d'ondes de choc (Fabian et al. 2006), la dissipation d'ondes sonores (Churazov et al. 2001, Fabian et al. 2003), ou à travers des effets de turbulence (Zhuravleva et al. 2014). Dans le cas des amas de galaxies, l'AGN peut ainsi transmettre de l'énergie au milieu intra-amas. La figure 1.4 présente un exemple des jets radio provenant de l'AGN de la galaxie elliptique Hercules A. Le trou noir au centre a une masse de $2.5 \times 10^9 M_\odot$, ce qui est environ 1000 fois plus massif que le trou noir au centre de la Voie Lactée. L'émission radio est tracée en

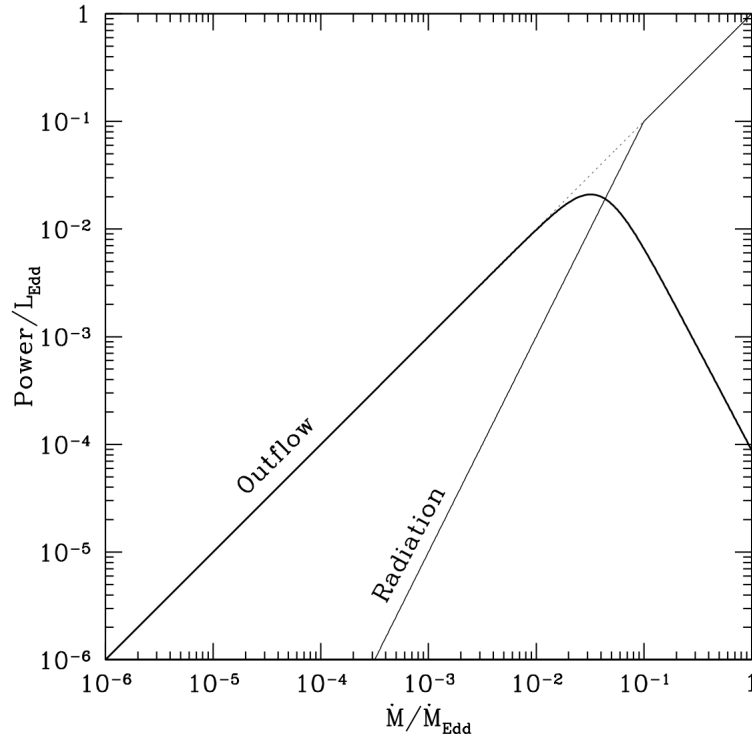


Figure 1.3 – Représentation schématisée de la puissance émise par un trou noir ($Power$), normalisée par la luminosité d’Eddington L_{Edd} , en fonction du taux d’accrétion du trou noir \dot{M} normalisé par le taux d’accrétion d’Eddington \dot{M}_{Edd} . Les lignes pleines représentent la puissance émise sous forme de jets relativistes (*outflows*) ou de radiation. (Source : Figure 1 de Churazov et al. 2005)

rouge ; on peut voir les jets puissants émis à des vitesses relativistes d’un bord et de l’autre de la galaxie. L’émission dans le domaine du visible est en jaune et en blanc ; on peut voir au centre la galaxie Hercules A ainsi que d’autres galaxies voisines.

La rétroaction de l’AGN peut aussi fonctionner dans un mode radiatif : l’énergie est alors redistribuée dans le milieu sous forme de radiation et de vents (Churazov et al. 2005). Ceux-ci sont émis à partir de l’AGN vers l’extérieur de la galaxie, sans direction particulière. Le réchauffement de l’environnement se fait donc de manière plutôt uniforme dans le mode radiatif. Ce mode est également appelé le mode quasar. Les quasars des AGNs extrêmement lumineux, ayant une luminosité bolométrique typique $L > 10^{45}$ erg s $^{-1}$. Ils sont très utiles pour étudier les vents de l’AGN, car ces vents sont souvent observés dans les raies d’absorption des spectres de quasars (Voir par exemple la revue de littérature sur le sujet de Fabian 2012).

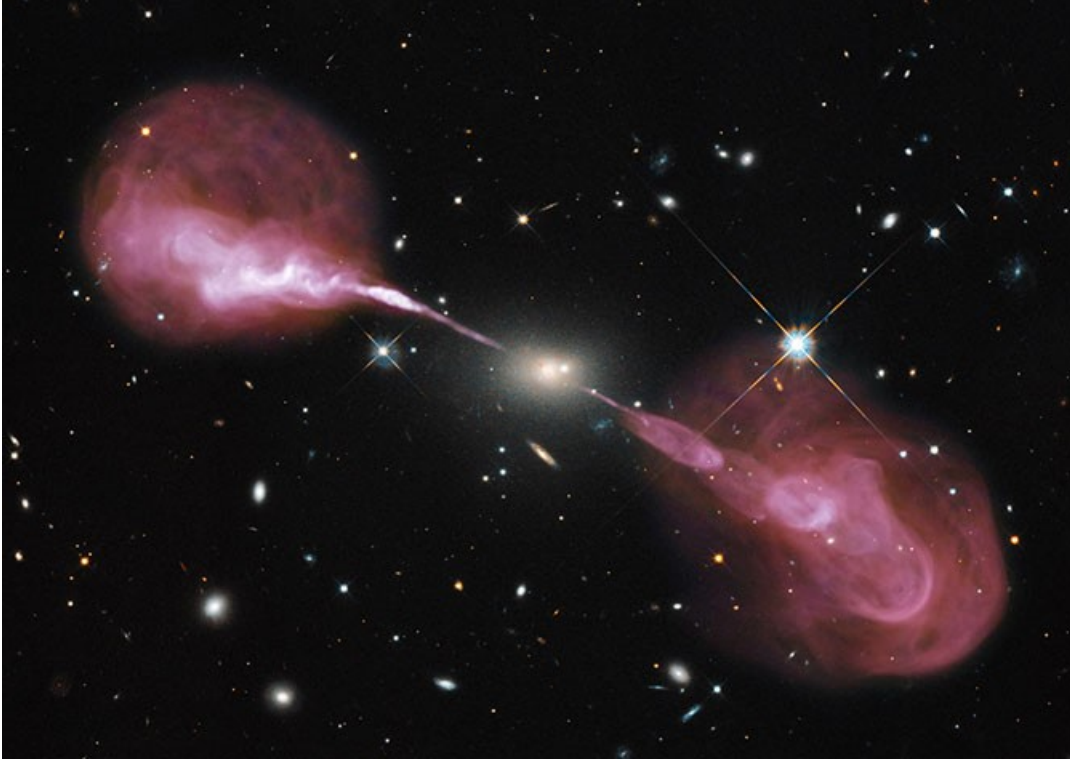


Figure 1.4 – Image composite de la galaxie elliptique Hercules A et des jets radio expulsés par l’AGN de la galaxie. Les éléments en rouge représentent le domaine radio, captés par le télescope terrestre VLA. Les galaxies en jaune et en blanc ont été captés dans le visible par le télescope spatial *Hubble*, alors que . (Source : NASA, ESA, S. Baum et C. O’Dea, R. Perley et W. Cotton, et l’équipe du projet *Hubble Heritage*)

1.2.3. Amas de galaxies à coeur froid et à coeur non-froid

Les observations en rayons X sont le principal moyen d’obtenir de l’information sur les amas de galaxies car elles nous permettent d’étudier les propriétés du ICM. Ces observations nous permettent de voir la brillance de surface, c’est-à-dire la densité de flux par unité d’angle solide du ICM, qui est directement proportionnelle au carré de la densité du gaz (Sarazin 1986). Ces observations nous montrent qu’il existe deux profils communs pour la distribution du ICM dans les amas de galaxies : le gaz peut être concentré au centre pour former un pic d’émissivité, ou bien il peut être distribué de manière plus uniforme (voir figure 1.5). Les amas avec une forte concentration de gaz au centre sont appelés amas à coeur froid (CC, de l’anglais *cool core*), et les amas ayant une distribution de gaz uniforme sont par opposition appelés amas à coeur non-froid (NCC, de l’anglais *non-cool core*). La provenance de cette appellation est expliquée à la section 1.2.3.1. La figure 1.5 présente les amas Abell 2029 (à gauche) et Abell 2319 (à droite), des amas CC et NCC typiques.

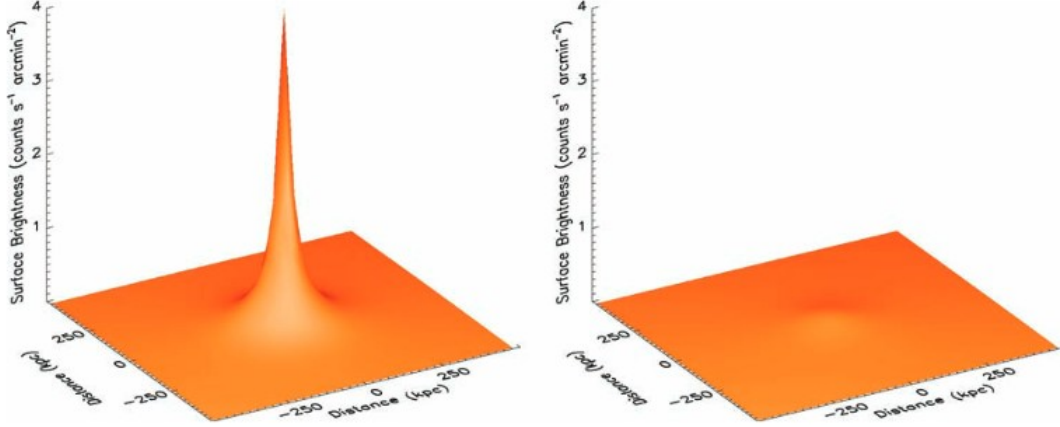


Figure 1.5 – Distribution de la brillance de surface en rayons X pour les deux catégories d’amas de galaxies : CC à gauche et NCC à droite. (Source : Million & Allen 2009)

1.2.3.1. *Courant de refroidissement*

L’émission radiative de l’ICM amène le gaz à perdre de l’énergie et donc à se refroidir. Puisque l’émission est directement reliée à la densité du gaz, le refroidissement est particulièrement marqué dans les coeurs des amas CC. Cette perte d’énergie se traduit par une baisse de pression. Cependant, dans le coeur des amas de galaxies, la pression doit être en mesure de supporter le poids des couches supérieures du gaz. Afin d’éviter l’effondrement de celles-ci, le gaz devient plus dense au centre. On assiste alors à un flux du gaz froid des zones éloignées vers le centre de l’amas ; c’est le courant de refroidissement. Le taux du transfert de la masse \dot{M} est mesuré en unité de masse par unité de temps (souvent M_{\odot}/yr) et s’exprime comme (p. ex. Sarazin 2007) :

$$\dot{M} \approx \frac{2}{5} \frac{L \mu m_H}{k_B T} \quad (1.2.8)$$

où L est la luminosité bolométrique de la région qui subit le refroidissement, μ est le poids moléculaire moyen et m_H est la masse d’un atome d’hydrogène.

Pour les amas CC, ce phénomène mènerait à une grande accumulation de gaz froid au centre de l’amas, d’où leur nom d’amas à coeur froid. Cela prendrait environ 1 Gyr pour les amas CC, ce qui est nettement inférieur à l’âge typique d’un amas de galaxies (environ 10^{10} années). Pour les autres amas, ce processus s’étalerait sur des temps supérieurs à leur âge, et donc on n’observe pas un tel refroidissement (Sarazin 1986). Éventuellement, le gaz froid au centre des amas CC devrait se condenser pour former des nuages moléculaires froids, et la présence de ceux-ci favoriserait un très haut taux de formation stellaire, autour de 100 à 1000 M_{\odot}/an (p. ex. Peterson & Fabian 2006). Cependant, le réel taux de formation stellaire

observé est beaucoup plus bas, autour de 10 à 300 M_{\odot} /an (p. ex. McNamara et al. 2004, McDonald et al. 2015). Suite à ces observations, les astronomes ont conclu qu'il doit y avoir un mécanisme qui empêche l'ICM de refroidir à un rythme aussi rapide que prévu par le modèle du courant de refroidissement. L'hypothèse la plus acceptée à ce sujet est celle de la rétroaction de l'AGN ; le trou noir injecte de l'énergie dans le système principalement sous forme de jets relativistes, ce qui chauffe le gaz.

1.2.3.2. Classification des amas de galaxies

Les amas sont parfois séparés en deux classes (CC et NCC), mais on ajoute souvent une catégorie intermédiaire. On parle alors de CC, de WCC (de l'anglais *weak cool core*, aussi appelés MCC pour *moderate cool core*), et de NCC. Dans le cadre de ce mémoire, nous adoptons une séparation en trois catégories afin de faciliter la comparaison de nos résultats avec des recherches similaires.

Plusieurs propriétés du gaz chaud intra-amas peuvent servir comme métriques de classification pour différencier les catégories d'amas de galaxies. Parmi celles-ci, on retrouve : le temps de refroidissement central, la densité électronique centrale, l'entropie centrale, le paramètre de concentration de la brillance de surface, le paramètre de courbure du profil de densité (*cuspidity*).

Temps de refroidissement central

Le temps de refroidissement est le temps que prendrait le gaz pour dégager l'ensemble de son énergie et se refroidir complètement, en absence des mécanismes de rétroaction. Pour un amas de galaxies, il est défini comme l'enthalpie du gaz divisée par l'énergie perdue par unité de temps, ce qui s'exprime comme :

$$t_{cool} = \frac{3(n_e + n_p)kT}{2n_e n_H \Lambda(T, Z)}, \quad (1.2.9)$$

où n_e est la densité d'électrons, n_p est la densité de protons, n_H est la densité d'atomes d'hydrogène, k est la constante de Boltzmann, T est la température, $\Lambda(T, Z)$ est la fonction de refroidissement (*cooling function*) pour des valeurs de température T et de métallicité Z . On peut simplifier l'expression en prenant $n_p = 0.92n_e$ et $n_H = 0.83n_e$ suivant McDonald et al. (2018), et on se retrouve alors avec :

$$t_{cool} = \frac{3}{2} \frac{2.31kT}{n_e \Lambda(T, Z)}, \quad (1.2.10)$$

où la valeur de $\Lambda(T, Z)$ peut être obtenue en effectuant une interpolation des tableaux fournis dans Sutherland & Dopita (1993), qui décrit les fonctions de refroidissement pour

les plasmas optiquement minces. Si on fait abstraction des mécanismes de rétroaction, les amas qui ont un pic d'émissivité plus prononcé auraient tendance à se débarrasser plus rapidement de leur énergie, et donc auraient un plus petit temps de refroidissement. Les amas CC sont ceux avec un temps de refroidissement central $t_{cool,0}$ inférieur à 1 Gyr, les amas WCC ont un $t_{cool,0}$ entre 1 et 7.7 Gyr, et les amas NCC ont un $t_{cool,0}$ supérieur à 7.7 Gyr (Hudson et al. 2010, McDonald et al. 2013, Barnes et al. 2018). On utilise un seuil de 7.7 Gyrs pour être conforme avec la littérature, mais Hudson et al. (2010) précise que ce temps correspond au temps cosmique pour $z \approx 1$, moment où la plupart des amas auraient bien eu le temps de se stabiliser suffisamment pour former un coeur froid si telle était leur prédisposition.

Densité électronique centrale

Une autre métrique de classification couramment utilisée est la densité électronique centrale $n_{e,0}$, exprimée en cm^{-3} . Les amas avec une $n_{e,0}$ au dessus de $1.5 \times 10^{-2} \text{cm}^{-3}$ sont classifiés comme CC, les amas avec une $n_{e,0}$ entre $0.5 \times 10^{-2} \text{cm}^{-3}$ et $1.5 \times 10^{-2} \text{cm}^{-3}$ sont classifiés comme WCC, et les amas avec une $n_{e,0}$ au-dessous de $0.5 \times 10^{-2} \text{cm}^{-3}$ sont considérés comme NCC (Hudson et al. 2010, Barnes et al. 2018).

Excès d'entropie centrale

L'entropie K peut également servir à différencier les classes d'amas de galaxies, et elle est définie comme :

$$K = \frac{kT}{n_e^{2/3}} \quad (1.2.11)$$

On peut ensuite procéder à un ajustement de la forme d'une loi de puissance (Cavagnolo et al. 2009) :

$$K(r) = K_0 + K_{100} \left(\frac{r}{100 \text{kpc}} \right)^\alpha \quad (1.2.12)$$

où r est le rayon en kpc, K_{100} est un facteur de normalisation pour l'entropie à $r = 100$ kpc, α est l'indice pour la loi de puissance. K_0 n'est pas l'entropie centrale mais plutôt l'excès d'entropie centrale, c'est-à-dire la quantité par laquelle l'entropie du coeur de l'amas surpasse l'ajustement en loi de puissance le mieux adapté aux plus grands rayons. Suivant Cavagnolo et al. (2009), l'ajustement est seulement réalisé à l'intérieur d'un rayon de $0.1r_{500}$. Les amas CC ont un excès d'entropie centrale K_0 inférieur à 30 keV cm^2 , les amas WCC ont un K_0 entre 30 et 60 keV cm^2 , et les amas NCC ont un K_0 supérieur à 60 keV cm^2 (Barnes et al. 2018).

Paramètre de concentration

Introduit par Santos et al. (2008), le paramètre de concentration est une métrique qui se calcule directement à partir des images en rayons X, et non à partir des propriétés physiques intrinsèques des amas de galaxies. En plus de dépendre de la température du gaz, l'émissivité en rayons X est aussi proportionnelle au carré de la densité. Les amas CC ont donc une forte brillance dans les rayons X dans le coeur de l'amas. Le paramètre de concentration vise à comparer le flux en rayons X dans une région centrale de l'amas avec le flux dans une région plus étendue. Cette métrique permet de voir à quel point l'émissivité est concentrée dans le coeur de l'amas. Le paramètre de concentration C_{SB} (l'indice SB signifie *surface brightness*) est défini comme :

$$C_{SB} = \frac{F(r < 40kpc)}{F(r < 400kpc)} \quad (1.2.13)$$

où $F(r < 40kpc)$ est le flux en rayons X (c'est-à-dire la brillance de surface dans une image en rayons X) à l'intérieur d'un rayon de 40 kpc, et $F(r < 400kpc)$ est le flux en rayons X à l'intérieur d'un rayon de 400 kpc. Le paramètre de concentration C_{SB} est un chiffre sans unités entre 0 et 1 (0 s'il y a aucune émissivité au centre, et 1 si toute l'émissivité est concentrée au centre). Les amas avec un C_{SB} supérieur à 0.155 sont classés comme des amas CC, ceux avec un C_{SB} entre 0.075 et 0.155 sont considérés comme WCC, et ceux avec un C_{SB} inférieur à 0.075 sont des NCCs (Santos et al. 2008).

Paramètre de courbure (*cuspiness*)

Finalement, la dernière métrique de classification abordée dans ce travail est le paramètre de courbure du profil de densité (*cuspiness*), α . Il a été introduit par Vikhlinin et al. (2007), et il représente la pente de la courbe de la densité électronique en fonction du rayon, tracée dans un espace log-log. Cette pente sera plus abrupte pour les amas avec un fort pic d'émissivité, et plus lisse pour les amas avec une distribution plus uniforme de gaz. Le paramètre α est défini comme :

$$\alpha = - \left. \frac{d \log n_e(r)}{d \log r} \right|_{r=0.04r_{500}} \quad (1.2.14)$$

La densité électronique est plus élevée au centre de l'amas, et diminue avec une augmentation du rayon. La dérivée sera donc négative, et le signe moins est ajouté pour que α soit défini positif. Les amas CC sont définis comme ayant un α supérieur à 0.70, les amas WCC ont un α entre 0.5 et 0.70, et les amas NCC ont un α inférieur à 0.5 (Vikhlinin et al. 2007).

Métrique	CC	WCC	NCC	Unités
$t_{\text{cool},0}$	$t_{\text{cool},0} < 1$	$1 < t_{\text{cool},0} < 7.7$	$t_{\text{cool},0} > 7.7$	Gyrs
$n_{e,0}$	$n_{e,0} > 1.5$	$1.5 > n_{e,0} > 0.5$	$n_{e,0} < 0.5$	10^{-2} cm^{-3}
K_0	$K_0 < 30$	$30 < K_0 < 60$	$K_0 > 60$	keV cm^2
C_{SB}	$C_{SB} > 0.155$	$0.155 > C_{SB} > 0.075$	$C_{SB} < 0.075$	sans unités
α	$\alpha > 0.7$	$0.7 > \alpha > 0.5$	$\alpha < 0.5$	sans unités

Tableau 1.1 – Résumé des critères pour appartenir aux catégories CC, WCC ou NCC selon les différentes métriques.

Les différentes métriques donnent souvent des résultats différents pour un même amas, et les astronomes ne s’entendent pas toujours pour dire lesquelles sont les plus fiables (p.ex. Hudson et al. 2010). De plus, pour calculer les métriques, il est souvent requis de localiser le centre de l’amas, mais la manière de le définir ne fait pas consensus ; certains utilisent le centre gravitationnel, d’autre optent plutôt pour le pic d’émissivité dans les rayons X (Cui et al. 2016). Dans ce mémoire, nous adoptons la deuxième définition.

Le tableau 1.1 présente un résumé des valeurs de seuil selon chaque métrique pour définir les catégories CC, WCC et NCC.

1.3. Observations d’amas de galaxies

Le contexte de la toute première détection d’un amas de galaxies est inconnu. Suite à des observations dans le domaine du visible réalisées au courant du 18e siècle, un grand nombre de galaxies ont été repertoriées dans les catalogues de Charles Messier et de William Herschel dans les années 1780 (Messier 1781, Herschel 1786). Il est incertain qui a été le premier à suggérer que plusieurs de ces galaxies appartenaient au même système, mais plusieurs astronomes ont joué un rôle clé dans le développement de nos connaissances à propos des amas de galaxies. En 1931, Edwin Hubble et Milton Humason ont publié un article détaillant les distances et les vitesses de plusieurs galaxies et amas de galaxies (Hubble & Humason 1931). Fritz Zwicky a étudiés ces objets, appliquant pour la première fois le théorème du viriel aux amas de galaxies. Il a également examiné les vitesses des galaxies individuelles de l’amas de Coma. Il a conclu que la gravité exercée par la masse totale des galaxies n’était pas suffisante pour expliquer les vitesses observées. Son analyse a mis en évidence que les amas de galaxies étaient composés en grande partie de matière sombre (Zwicky 1933). En 1958, George Abell a publié un catalogue répertoriant 2712 amas de galaxies. Ce catalogue fut une avancée majeure dans le domaine et est utilisé encore aujourd’hui comme ouvrage de référence. Dans les années suivantes, les premières observations en rayons-X ont été réalisées

(Bradt et al. 1967, Byram et al. 1966), révélant que l'espace entre les galaxies est comblé par un gaz diffus avec une très haute température : l'ICM.

1.3.1. Astronomie rayons X

L'atmosphère terrestre est impénétrable pour les rayons X ; toute observation dans ces longueurs d'onde (0.01 à 10 nm ; énergie entre 145 eV à 125 keV) doit donc être faite à partir de l'espace. Le moyen de mesure pour l'astronomie rayons X est la détection de photons individuels qui frappent les détecteurs des télescopes. On vise à identifier quatre de leurs propriétés : leur position (à partir de laquelle on peut créer des images), leur énergie (à partir de laquelle on peut obtenir des spectres), leur temps d'arrivée, et dans certains cas leur polarisation.

Des télescopes conventionnels pour des plus grandes longueurs d'onde (et donc des plus petites énergies) utilisent une série de miroirs pour focaliser les rayons lumineux. Ces rayons frappent la surface de manière quasi perpendiculaire. Ceci ne fonctionnerait pas avec des rayons X car ils sont trop énergétiques pour être réfléchis ; à un tel angle d'incidence ils seraient plutôt transmis ou absorbés. Le télescope Wolter (Wolter 1952) permet de résoudre ce problème en utilisant une série de miroirs que les rayons lumineux viennent frapper avec un très petit angle d'incidence, ce qui permet de légèrement dévier la trajectoire des rayons. En utilisant une combinaison de miroirs paraboliques et hyperboliques, la lumière peut alors être focalisée en un point pour créer une image cohérente.

Le premier satellite lancé spécifiquement pour l'astronomie rayons X est *Uhuru*, qui a pris son envol en 1970 (Giacconi et al. 1971). Il a détecté quelques centaines de sources de rayons X à travers l'ensemble du ciel, dressant ainsi le premier catalogue. À la fin des années 1970, la NASA a lancé une série de trois télescopes dédiés à l'astrophysique des hautes énergies, les HEAOs (Observatoires d'astrophysique des hautes énergies, de l'anglais *High Energy Astronomy Observatories*). Les trois télescopes étaient dédiés à l'astronomie des rayons gamma, des rayons X et des rayons cosmiques. HEAO-2, renommé par la suite *Einstein*, était spécialisé dans les rayons X (Giacconi et al. 1979). Il a permis la découverte et l'étude d'une multitude de nouvelles sources de rayons X : des étoiles jeunes et chaudes, des sources variables, des rémanents de supernova, des pulsars, des puissants jets provenant de centres galactiques, etc. Le prochain développement majeur dans l'astronomie des rayons X s'est fait avec l'arrivée du télescope *Chandra*, dont les paramètres serviront à créer des observations simulées en rayons X utilisées dans le cadre de ce mémoire.

1.3.2. Observatoire de rayons X *Chandra*

Nommé après l'astrophysicien d'origine indienne Subrahmanyan Chandrasekhar, le télescope spatial *Chandra* a été conceptualisé au début des années 1980 pour finalement être lancé en 1999. Le télescope est en orbite elliptique autour de la Terre, avec une période orbitale de 65 heures (Weisskopf et al. 2000). Les données sont ensuite envoyées au CXC (Centre pour les rayons X de *Chandra*, de l'anglais *Chandra X-Ray Center*), qui se trouve au Massachusetts et est géré par Centre d'astrophysique de Harvard et du Smithsonian (*Harvard-Smithsonian Center for Astrophysics*). Même si la mission avait initialement une durée prévue de seulement 5 ans, *Chandra* est toujours en opération 23 ans plus tard.

Chandra utilise un télescope de Wolter composé de plusieurs miroirs (paraboliques et hyperboliques) cylindriques de taille différente imbriqués les uns dans les autres, qui forment le HRMA (*High Resolution Mirror Assembly*). Cet assemblage de miroirs est illustré à la figure 1.6. Au point focal, deux instruments peuvent être placés : la caméra à haute résolution HRC (*High Resolution Camera*) ou le spectromètre imageur ACIS (*Advanced CCD Imaging Spectrometer*). La figure 1.7 est une illustration des composantes du télescope *Chandra*. On peut y voir l'assemblage de miroirs HRMA, ainsi que les deux instruments HRC et ACIS. Les images qui nous intéressent dans le cadre de ce mémoire sont obtenues avec la caméra ACIS, dont les détails sont présentés ci-dessous.

La caméra ACIS est un assemblage de capteurs photographiques CCD (de l'anglais *charged coupled device*). La caméra détecte les photons individuels qui frappent la surface, puis enregistre leur position, leur énergie et leur temps d'arrivée. L'instrument ACIS peut donc être utilisé pour générer des images ainsi que des spectres. Chaque photon qui frappe le détecteur est considéré comme un événement, et les événements sont ensuite analysés par les logiciels de traitement de données. L'instrument contient 10 CCDs, chacun ayant $1024 \text{ pixels} \times 1024 \text{ pixels}$, et chaque pixel mesure $0.492 \text{ arcsecondes} \times 0.492 \text{ arcsecondes}$ sur le ciel. Quatre CCDs sont arrangés en carré 2×2 , formant la configuration ACIS-I et ayant un champ de vue (*field of view*) total de $16.9 \times 16.9 \text{ arcminutes}$. Les six autres CCDs sont arrangés en ligne droite, formant la configuration ACIS-S. Seulement six des dix CCDs peuvent être en fonction simultanément. La configuration ACIS-S contient 2 CCDs rétro-éclairés (*back-illuminated*), alors que les 8 autres CCDs sont éclairés par l'avant. Les CCDs rétro-éclairés sont sensibles à de plus basses énergies, et détectent donc plus de photons. Ceci est avantageux pour des sources peu brillantes, mais introduit également plus de bruit dans les observations. Les amas de galaxies sont généralement observés avec la configuration ACIS-I, et c'est également cette configuration que nous choisissons pour créer

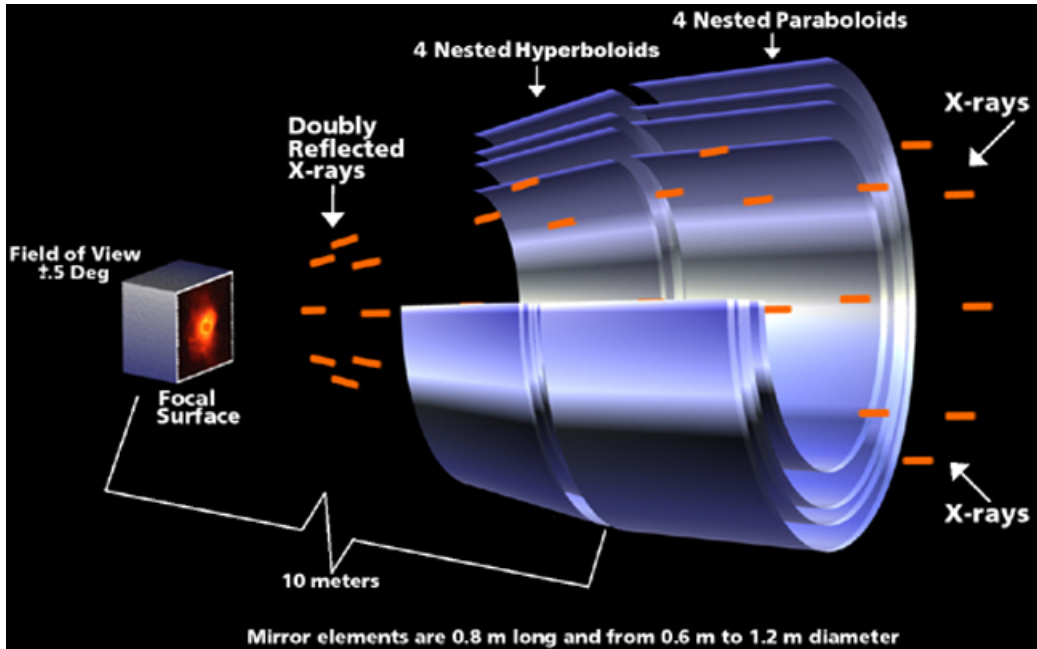


Figure 1.6 – Assemblage des miroirs du télescope *Chandra*. Les rayons X arrivent de la droite (dans l'image), sont réfléchis par un miroir parabolique et ensuite par un miroir hyperbolique, pour enfin arriver au point focal. (Source : NASA/CXC/D.Berry)

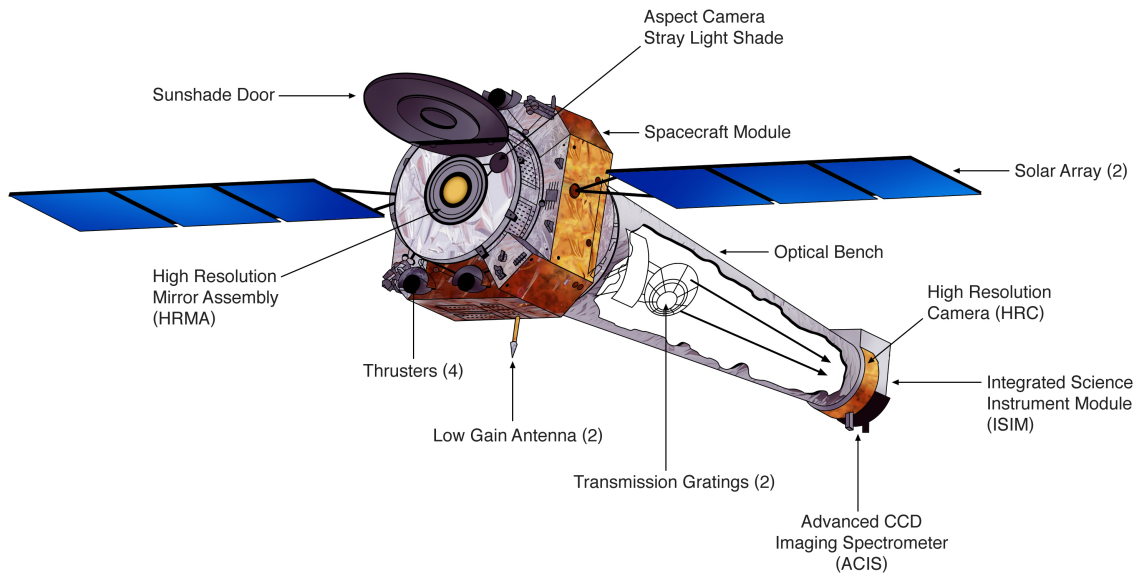


Figure 1.7 – Illustration des différentes composantes du télescope *Chandra*. (Source : NGST & NASA/CXC)

les images simulées en rayons-X.

1.4. Simulations cosmologiques

Les données utilisées dans le cadre de ce travail proviennent de simulations cosmologiques hydrodynamiques. Ces simulations imposent des conditions initiales pour obtenir une réalisation de l'Univers. Celle-ci est ensuite propagée dans le temps en solutionnant les équations de la physique qui gouvernent notre Univers.

1.4.1. Notions de cosmologie

Afin de mieux comprendre le contexte des simulations cosmologiques, nous procédons d'abord à un rappel de quelques notions de cosmologie.

À grande échelle (>100 Mpc), l'Univers est homogène et isotrope. Les travaux de Lemaître (Lemaître 1927) et de Hubble (Hubble 1929) ont permis d'ajouter une troisième caractéristique importante ; l'Univers est aussi en expansion. Ils sont arrivés à cette conclusion en observant que la vitesse des galaxies lointaines est directement proportionnelle à leur distance ; plus un objet est loin, plus il s'éloigne rapidement. La constante de proportionnalité entre la vitesse mesurée et la distance est la constante de Hubble, H_0 :

$$v = H_0 r \tag{1.4.1}$$

où v est la vitesse en km/s et r est la distance en Mpc. On mesure donc H_0 en km/s/Mpc. Plusieurs méthodes existent pour mesurer H_0 . Elles ne sont pas en accord les unes avec les autres, donnant lieu à la tension de Hubble (p. ex. Di Valentino et al. 2021). Dans le cadre de ce mémoire, on adoptera la valeur de la collaboration de Planck : $H_0 = 67,4$ km/s/Mpc (Planck Collaboration et al. 2020).

La vitesse des objets dans l'espace est également liée au concept de *redshift* z (décalage vers le rouge). Le *redshift* exprime la différence entre la longueur d'onde émise par un objet en mouvement $\lambda_{\text{émise}}$ et la longueur d'onde observée λ_{obs} (par effet Doppler) :

$$z = \frac{\lambda_{\text{obs}} - \lambda_{\text{émise}}}{\lambda_{\text{émise}}} \tag{1.4.2}$$

$$1 + z = \frac{\lambda_{\text{obs}}}{\lambda_{\text{émise}}} \tag{1.4.3}$$

À petite échelle, la mesure du *redshift* sera dominée par des effets locaux, comme des mouvements orbitaux. À grande échelle, le décalage mesuré nous indique la vitesse due à l'expansion de l'Univers. Le *redshift* est donc proportionnel à la distance qui nous sépare de l'objet. Puisque la lumière voyage à une vitesse finie, la lumière qui nous parvient des objets éloignés nous permet de les voir tels qu'ils étaient dans le passé, au moment où la lumière a été émise. Plus un objet est loin (et plus son z est grand), plus la lumière qui en provient a voyagé longtemps, plus on le voit loin le passé. Étudier un même type d'objet à différents z nous permet donc d'observer son évolution.

Il existe un nombre de simulations cosmologiques hydrodynamiques ayant un volume suffisamment grand pour produire de larges échantillons statistiques de galaxies. Parmi celles-ci, on retrouve les simulations IllustrisTNG (Nelson et al. 2019, Pillepich et al. 2018, Springel et al. 2018, Nelson et al. 2018, Naiman et al. 2018, Marinacci et al. 2018), ayant une taille de boîte de simulation allant de 50 Mpc comobiles à 300 Mpc comobiles. Il y a aussi les simulations SIMBA (Davé et al. 2019), Horizon-AGN (Dubois et al. 2014) et MassiveBlack-II (Khandai et al. 2015), toutes avec une taille boîte de simulation de 100 Mpc comobiles. Les simulations Romulus (Tremmel et al. 2017) ont une taille de boîte de simulation entre 25 et 50 Mpc comobiles, et les simulations EAGLE (McAlpine et al. 2016) ont une taille de boîte de simulation entre 25 et 100 Mpc comobiles. Ces simulations produisent une population de galaxies pour laquelle les statistiques sont généralement en accord avec les données d'observation (Vogelsberger et al. 2020). D'autres phénomènes physiques (le refroidissement du gaz, la formation stellaire, les effets du champ magnétiques, l'évolution des trous noirs supermassifs, l'interaction avec la matière noire, etc) sont modélisés de manières différentes dans ces simulations, mais ont somme toute une performance similaire lorsqu'on compare les résultats aux observations. Finalement, nous choisissons de travailler avec les simulations IllustrisTNG car le grand volume nous permet d'analyser un plus grand nombre d'amas de galaxies.

1.4.2. Illustris TNG

Le projet Illustris est une série de simulations hydrodynamiques cosmologiques, ayant comme objectif premier d'offrir une modélisation fidèle des processus de formation et évolution des galaxies. Ces simulations offrent simultanément une précision à petite échelle permettant l'étude des galaxies individuelles, mais également une modélisation des structures à une échelle de plusieurs centaines de mégaparsecs. La première version des simulations date de 2014, mais on se concentrera davantage sur la nouvelle génération

rendue disponible en 2018 : IllustrisTNG (TNG pour *the next generation*).

Ces simulations visent à modéliser de manière conjointe l'évolution de la matière noire, du gaz, des étoiles, des trous noirs, des champs magnétiques, de la radiation, etc. IllustrisTNG utilise le code APERO (Springel 2010), qui solutionne les équations magnétohydrodynamiques et gravitationnelles sur une grille mobile dans un volume fini. Des *snapshots* (c'est-à-dire des réalisations de l'Univers à un moment précis) sont disponibles pour différentes valeurs de *redshift* z (et donc différentes valeurs de temps cosmique), allant de $z=127$ jusqu'au présent à $z=0$. Il y a trois volumes disponibles pour les simulations : TNG50, TNG150 et TNG300 ; le nombre fait référence à la taille de la boîte de simulation en unités de Mpc comobiles. Les volumes plus petits offrent une meilleure résolution ; par exemple, la résolution spatiale de TNG100 et TNG300 diffèrent d'un facteur deux (Weinberger et al. 2018). Cependant, TNG300 contient le plus grand nombre d'amas de galaxies et c'est pourquoi on sélectionne ce volume pour réaliser notre analyse.

Plusieurs algorithmes existent pour rassembler des particules de simulation en groupes cohérents. Dans les simulations IllustrisTNG, les structures sont formées avec l'algorithme FoF (de l'anglais *friends-of-friends*, littéralement traduit comme « amis d'amis »). Cet algorithme fonctionne en définissant une longueur de liaison b (*linking length*), puis en imposant que toutes les particules séparées d'une distance inférieure à b se retrouvent dans le même groupe. Chaque particule est donc directement liée aux particules se trouvant à l'intérieur d'un rayon b (ses amis directs), et indirectement liée aux autres particules liées à ces amis directs. Ces particules de lien secondaire sont les amis d'amis - *friends-of-friends*. Dans le cas des simulations IllustrisTNG, cet algorithme est appliqué aux particules de matière noire (car elles jouent le plus grand rôle au niveau de l'attraction gravitationnelle), puis les particules d'autre type (gaz, étoiles, etc) sont placées dans le même groupe que la particule de matière noire la plus proche. Cela permet de rassembler les particules en structures gravitationnellement liées, telles que les amas de galaxies. Dans le cadre de ce mémoire, on utilise seulement les particules de gaz de chaque amas.

Dans ces simulations, les particules de gaz ne sont pas traitées comme des particules individuelles, mais sont plutôt distribuées dans des cellules de gaz ayant un volume fini. On a accès aux propriétés physiques de ces cellules : les coordonnées physiques, la vitesse, la densité, l'énergie thermique, la densité électronique, le taux de formation stellaire, etc. Ces propriétés nous permettent de simuler l'émission électromagnétique du gaz pour un amas de galaxies particulier, puis de générer les photons produits à l'aide de pyXSIM (ZuHone et al. 2014), un logiciel permettant de simuler l'émission en rayons X de sources astrophysiques. Les photons sont simulés à partir d'un modèle cosmologique, d'un modèle d'émission thermique,

ainsi que de certaines propriétés des cellules de gaz, telles que la position, la densité, la température et la métallicité. Ces photons peuvent ensuite être projetés selon un axe et traités comme des événements par un télescope. En précisant un *redshift*, un modèle d'univers, un modèle de télescope, un temps d'exposition et une aire d'exposition, on peut créer des observations simulées du télescope *Chandra* à l'aide du logiciel SOXS (Observations simulées de sources en rayons X, de l'anglais *Simulated Observations of X-ray Sources*; ZuHone et al. 2014). Ces images simulées représentent notre ensemble de données pour l'entraînement de notre modèle d'apprentissage automatique.

1.5. Apprentissage automatique

L'apprentissage automatique est une branche de l'intelligence artificielle. Elle permet à un ordinateur de tirer profit de l'accès à une grande quantité de données pour 'apprendre' à effectuer une tâche particulière sans être explicitement programmé pour celle-ci. Dans un contexte d'apprentissage supervisé, on utilise un ensemble de données étiquetées et l'on souhaite que l'ordinateur apprenne à prédire les étiquettes à partir des données d'entrée. Les données d'entrée peuvent être des images, des vecteurs, des extraits de texte, etc ; et les étiquettes peuvent être des nombres, des indices de classe, des vecteurs.

Il existe plusieurs algorithmes d'apprentissage automatique de complexité variable. Un des systèmes de base de l'apprentissage automatique est le réseau de neurones artificiels.

1.5.1. Réseaux de neurones artificiels

Le concept des réseaux de neurones artificiels est inspiré du fonctionnement des neurones biologiques. Chaque neurone est une unité de traitement, qui peut recevoir de l'information en entrée, effectuer une transformation mathématique et ensuite transmettre un signal de sortie aux prochains neurones. Les neurones sont arrangés en couches, et on considère que chaque neurone reçoit l'information des neurones dans la couche précédente (à sa gauche dans un schéma), et transmet son signal aux neurones de la couche suivante (à sa droite dans un schéma). La configuration la plus simple d'un réseau de neurones est le perceptron multicouche (*multilayer perceptron*), où l'information est traitée de manière unidirectionnelle. Chaque neurone prend en entrée une combinaison linéaire des signaux de sortie des neurones de la couche précédente. Les coefficients dans cette combinaison linéaire sont appelés 'poids', et déterminent l'importance relative du signal de chaque neurone. Soit \mathbf{x} un vecteur de données d'entrée, tel que $\mathbf{x} = [x_1, x_2, \dots, x_d]$. Les éléments de vecteur x_i sont les neurones de

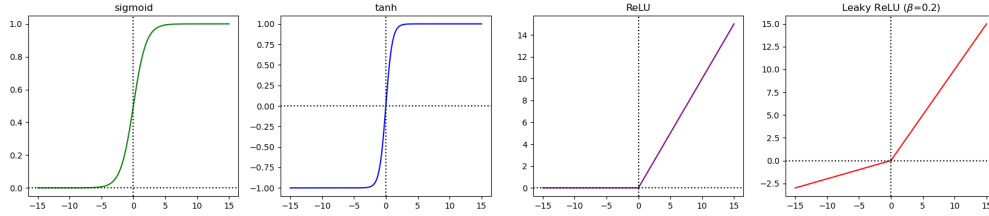


Figure 1.8 – Fonctions d’activation courantes, décrites à l’ensemble d’équations 1.5.3.

la couche d’entrée. Pour les neurones de la première couche cachée, le signal d’entrée sera :

$$a_j = \sum_i w_{ji}^{(1)} x_i + w_{j0}^{(1)}, \quad (1.5.1)$$

où j est l’indice du neurone de la première couche cachée, w_{ji} est le poids de l’élément i de la couche d’entrée vers le neurone j de la première couche cachée. w_{j0} est le biais - un paramètre ajustable qu’on ajoute à la combinaison linéaire pour donner plus de flexibilité au réseau. L’exposant (1) fait référence au fait qu’il s’agit de la première couche cachée du réseau.

Chaque neurone transforme l’information qu’il reçoit en appliquant une fonction d’activation $h(\cdot)$, c’est-à-dire une fonction mathématique non-linéaire et différentiable. Le signal de sortie de chaque neurone de la première couche cachée est donc :

$$g_j = h^{(1)}(a_j) = h^{(1)}\left(\sum_i w_{ji}^{(1)} x_i + w_{j0}^{(1)}\right), \quad (1.5.2)$$

où $h^{(1)}$ est la fonction d’activation pour la première couche cachée. Parmi les fonctions d’activation les plus souvent utilisées, on note : la sigmoïde, la fonction tanh, la ReLU (pour *Rectified Linear Unit*), et la Leaky ReLU (voir figure 1.8) :

$$\begin{aligned} \text{sigmoïde} : h(a) &= \frac{1}{1 + e^{-a}} \\ \text{tanh} : h(a) &= \tanh(a) \\ \text{ReLU} : h(a) &= \max(0, a) \\ \text{Leaky ReLU} : h(a) &= \max(\beta a, a), \text{ où } \beta \text{ est une constante positive} \end{aligned} \quad (1.5.3)$$

La ReLU et la Leaky ReLU ne sont pas différentiables à zéro, mais on peut contourner ce problème en imposant une valeur finie pour la dérivée en ce point. De plus, il est plutôt rare d’avoir exactement $a = 0$.

Pour la deuxième couche cachée, la structure est la même ; le signal de sortie pour chaque neurone est :

$$g_k = h^{(2)}(a_k) = h^{(2)}\left(\sum_j w_{kj}^{(2)} g_j + w_{k0}^{(2)}\right) = h^{(2)}\left(\sum_j w_{kj}^{(2)} h^{(1)}\left(\sum_i w_{ji}^{(1)} x_i + w_{j0}^{(1)}\right) + w_{k0}^{(2)}\right), \quad (1.5.4)$$

où k est l'indice du neurone de la deuxième couche cachée, $h^{(2)}$ est la fonction d'activation pour la deuxième couche cachée, $w_{kj}^{(2)}$ est le poids du neurone j de la première couche cachée vers le neurone k de la deuxième couche cachée, $w_{k0}^{(2)}$ est le biais du neurone k de la deuxième couche cachée. On peut poursuivre le raisonnement pour un nombre de couches de notre choix, où chaque couche peut contenir un nombre différent de neurones, et avoir une fonction d'activation différente, si souhaité.

Finalement, la dernière couche représente la prédiction du réseau. En entraînant le réseau, on souhaite trouver les valeurs des poids qui feront en sorte que cette prédiction se rapproche de l'étiquette véritable. Dans notre cas, les données d'entrée sont les valeurs numériques des pixels d'une image, et l'étiquette à prédire est un vecteur de métriques $y = (y_0, y_1, y_2, y_3, y_4)$. Pour comparer à quel point la prédiction du réseau et l'étiquette véritable sont éloignées les unes des autres, on utilise une fonction de coût \mathcal{L} (L pour *Loss*). Celle-ci peut aussi prendre plusieurs formes différentes, mais elle contient généralement un terme d'erreur de la forme $(y_{véritable} - y_{prédiction})$, où y est l'étiquette.

1.5.1.1. *Rétropropagation*

Les poids \mathbf{w} sont des paramètres dits 'entraînables', c'est-à-dire qu'ils sont modifiés pendant l'entraînement de manière à minimiser le coût (l'erreur). La fonction de coût est différentiable par rapport aux poids du réseau, et la structure du réseau de neurones permet d'utiliser la dérivée en chaîne pour trouver la dérivée de la fonction de coût par rapport à chacun des poids. On peut ensuite ajuster ce poids de manière à minimiser la fonction de coût, où pour un poids quelconque w_{ij} , cet ajustement aura la forme :

$$\Delta w_{ij} = -\alpha w_{ij} \quad (1.5.5)$$

où α est le taux d'apprentissage (*learning rate*), généralement une petite quantité. Si on généralise cela à l'ensemble des poids \mathbf{w} , α est la grandeur du pas d'ajustement effectué dans l'espace des paramètres :

$$\Delta \mathbf{w} = -\alpha \nabla_{\mathbf{w}} \mathcal{L} \quad (1.5.6)$$

où \mathcal{L} est la fonction de coût, et $\nabla_{\mathbf{w}}\mathcal{L}$ est le gradient de la fonction de coût par rapport aux poids \mathbf{w} .

Puisque la fonction de coût n'est souvent pas convexe, il peut arriver qu'une simple descente de gradient nous mène à un minimum local de la fonction de coût qui ne nous convient pas. On peut alors avoir recours à des algorithmes d'optimisation plus complexes qui font usage de certaines astuces. Une astuce possible est d'utiliser la descente de gradient avec momentum. Dans ce formalisme, plutôt que de simplement se déplacer dans la direction opposée au gradient à chaque pas de temps, on garde en mémoire les gradients des pas de temps précédents sous forme de moyenne mobile (*running average*) pour avoir une meilleure estimation de la direction globale dans laquelle l'optimiseur doit se déplacer pour éviter de rester coincé dans des minima locaux. On peut également faire usage d'un taux d'apprentissage adaptatif qui diffère pour chaque poids w .

On dit qu'une époque d'entraînement est complétée quand le réseau a traité tous les exemples de notre ensemble de données. Les exemples sont présentés dans un ordre aléatoire différent d'une époque à l'autre. À chaque époque, les poids peuvent être ajustés une seule ou plusieurs fois. S'il sont ajustés une seule fois, on parle de descente de gradient par lot (*batch*). On calcule alors la moyenne des gradients sur tous les exemples, et on prend un pas d'ajustement à la fin de l'époque. Une autre option serait d'ajuster les poids après chaque exemple individuel. Cette approche est la descente de gradient stochastique. La fonction de coût converge plus rapidement dans ce cas, car pour un ensemble de données contenant n exemples, les poids sont ajustés n fois par époque. Cependant, les exemples ne peuvent pas être traités en parallèle. Une troisième option intermédiaire est la descente de gradient par mini-lot (*minibatch*). Les n exemples sont répartis en sous-ensembles, appelés mini-lots. On prend la moyenne des gradients sur les exemples d'un seul mini-lot pour chaque pas d'ajustement. Le calcul des gradients peut alors être réalisé en parallèle.

En plus des paramètres entraînaibles (les poids), nous avons aussi des hyperparamètres qui déterminent la structure du réseau, mais qui ne sont pas ajustés par rétropropagation pendant l'entraînement. Ces hyperparamètres incluent le nombre de couches, le nombre de neurones par couche, le taux d'apprentissage, la taille des tuiles dans un réseau de neurones convolutif (voir section 1.5.2), la taille de lot, etc. Les valeurs des hyperparamètres peuvent grandement affecter la performance du réseau, il est donc important de les choisir avec soin.

1.5.1.2. *Ensembles d'entraînement, de validation et de test*

La performance de notre réseau de neurones est évaluée selon sa capacité à généraliser ce qu'il a appris, c'est-à-dire à prédire les étiquettes pour des données qui n'ont pas été traitées pendant l'entraînement. On sépare habituellement nos données en trois parties, formant les ensembles d'entraînement, de validation et de test. L'ensemble d'entraînement est utilisé pour entraîner le réseau, c'est-à-dire pour ajuster les poids jusqu'à atteindre une valeur suffisamment petite de la fonction de coût. Les ensembles de validation et de tests ne sont pas utilisés pour ajuster les poids du réseau ; ils servent plutôt à améliorer et évaluer la performance du réseau sur des nouvelles données d'entrée.

D'un côté, l'ensemble de validation sert à surveiller la performance du réseau pendant l'entraînement. Pour cela, on visualise l'évolution de la fonction de coût avec le temps en traçant sa valeur à chaque époque sur l'ensemble d'entraînement ainsi que sur l'ensemble de validation (en prenant la moyenne de la fonction de coût sur l'ensemble des lots). Tant que la fonction de coût diminue avec le temps sur les deux ensembles (entraînement et validation), on dit que le réseau est encore dans la phase d'entraînement. On pourrait ensuite assister à une phase de surapprentissage : la fonction de coût sur l'ensemble d'entraînement diminue, mais celle sur l'ensemble de validation commence à augmenter (voir figure 1.9). Le réseau n'est alors plus en mesure de généraliser ses apprentissages à l'ensemble de validation car il est trop adapté aux particularités de l'ensemble d'entraînement (on dit qu'il 'apprend par coeur'). On souhaite arrêter l'entraînement avant cette phase, au moment où la fonction de coût sur l'ensemble de validation est à son plus bas. D'un autre côté, l'ensemble de validation sert à choisir les meilleurs hyperparamètres. On peut essayer différentes valeurs de ceux-ci, et choisir la combinaison qui maximise la performance sur l'ensemble de validation.

Finalement, l'ensemble de test sert à évaluer la performance de notre réseau sous sa forme finale. Les exemples de l'ensemble de test n'ont pas été utilisés ni lors de l'entraînement ni lors du choix d'hyperparamètres ; c'est donc un ensemble d'exemples complètement nouveaux.

1.5.2. Réseaux de neurones convolutifs

Les réseaux de neurones convolutifs sont particulièrement bien adaptés pour des images, car ils peuvent traiter des données d'entrée hautement dimensionnelles, tout en considérant la dépendance spatiale entre les pixels au sein d'une image.

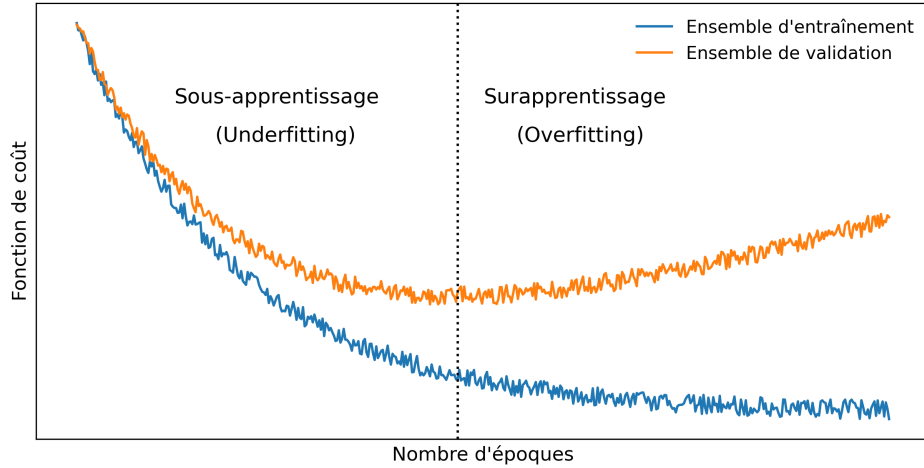


Figure 1.9 – Exemple de courbe de fonction de coût. On y voit la valeur de la fonction de coût en fonction du nombre d'époques pour les ensembles d'entraînement et de test. Les phases d'entraînement (sous-apprentissage et surapprentissage) y sont identifiées.

On peut considérer une image d'entrée comme un tableau ayant deux dimensions. Dans un réseau convolutif, le tableau est séparé en tuiles, et chaque tuile est traitée par un neurone différent. Les neurones d'une même couche partagent les mêmes poids, c'est-à-dire que toutes les tuiles d'un même tableau subissent la même opération, ce qui introduit une symétrie de translation dans le réseau. La combinaison linéaire prend donc la forme d'une convolution en deux dimensions :

$$(x * k)_{i,j} = \sum_{p,q} x_{i+p,j+q} k_{p,q} \quad (1.5.7)$$

où x est l'image d'origine et k représente la matrice contenant les poids du neurone, appelée noyau (*kernel*). p et q sont les indices des éléments de la matrice k , alors que i et j sont les indices des éléments du tableau résultant $(x * k)$.

Dans ce formalisme, les tuiles se chevauchent, et les pixels de chaque tuile sont multipliés par un élément de matrice k_{pq} . Le résultat est un tableau en deux dimensions qui se fait ensuite transformer par une fonction d'activation, avant de servir de tableau d'entrée à la prochaine couche.

Les noyaux peuvent aussi avoir une troisième dimension, et dans ce cas on a plusieurs tableaux de sortie à chaque couche. On parlera d'un nombre de filtres pour indiquer la taille de la troisième dimension.

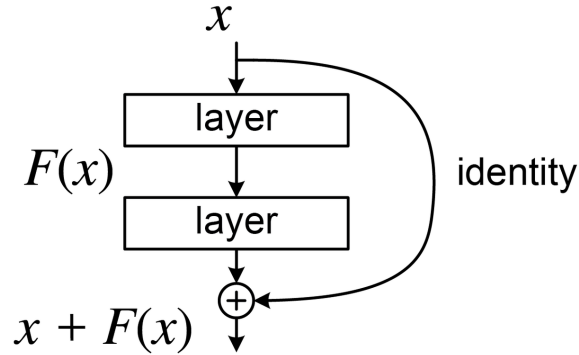


Figure 1.10 – Illustration d’un bloc résiduel contenant un saut de connexion. Les couches du réseau sont identifiées par *layer*, et la fonction identité est identifiée par *identity*. Le fonctionnement est détaillé à la section 1.5.3.

Les couches convolutives sont souvent suivies de couches de sous-échantillonnage (*pooling layers*). Ces couches ne contiennent pas de poids ajustables; on y effectue simplement la moyenne des valeurs de chaque tuile afin d’obtenir les éléments du tableau de sortie. Cela réduit la dimension des tableaux et augmente la taille du champ récepteur pour ces neurones, c’est-à-dire la région de l’image d’entrée qui est connectée à chacun des neurones.

1.5.3. Réseaux neuronaux résiduels (ResNet)

L’architecture choisie dans le cadre de ce projet est un réseau neuronal résiduel (ResNet pour *Residual Network*) comportant des couches convolutives. La particularité d’un ResNet est l’utilisation de sauts de connexion : le signal de sortie de certains neurones va ‘sauter’ des couches, et être acheminé directement vers des neurones se trouvant quelques couches plus loin. Ceci est représenté à la figure 1.10. Le vecteur x représente le signal de sortie d’une couche précédente. Dans un premier temps, x passe par les deux couches suivantes (*layer*) de manière normale. Cette transformation est identifiée par $F(x)$. D’un autre côté, x passe par un deuxième chemin qui saute les deux couches suivantes, et ne subit aucune transformation en passant par ce chemin, ce passage est donc représenté par l’identité. Ensuite, les résultats de ces deux chemins sont additionnés : $x + F(x)$. Cette configuration est appelée bloc résiduel. Le signal de sortie $x + F(x)$ peut ensuite servir d’entrée pour les couches suivantes.

Un des principaux avantages des réseaux résiduels est que les sauts de connexion offrent une plus grande expressivité au réseau sans ajouter davantage de paramètres entraînaibles. Le réseau a donc la capacité de représenter une plus grande famille de fonctions sans être

plus coûteux à entraîner en terme de mémoire.

1.5.4. Simulation-based inference

Une fois que notre réseau est entraîné, on peut l'appliquer sur des nouvelles données pour obtenir des prédictions. Pour utiliser ces résultats à des fins scientifiques, il est essentiel de connaître l'incertitude sur la prédiction obtenue. On veut donc savoir : étant donné la valeur prédite $y_{\text{prédit}}$, quelles sont les possibles valeurs y_{vrai} ? Ou, en d'autres mots, quelle est la distribution de probabilité sous-jacente $p(y_{\text{vrai}}|y_{\text{prédit}})$ pour chaque valeur de y_{vrai} ?

Pour répondre à cette question, il est utile d'avoir recours à la méthode SBI (*simulation-based inference*, également appelée LFI pour *Likelihood-free inference* - inférence sans vraisemblance). Le formalisme de cette méthode est expliqué dans ce qui suit.

Supposons un modèle statistique qui prend en entrée des paramètres θ , à partir desquels il génère des observations x . Pour une observation donnée, on veut savoir quels sont les paramètres qui pourraient l'avoir générée, c'est-à-dire quelle est la distribution de probabilité de θ sachant x : $p(\theta|x)$. Pour cela, on peut avoir recours à la loi de Bayes :

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$
$$p(\theta|x) \propto p(x|\theta)p(\theta),$$

où $p(\theta)$ est la probabilité à priori de θ (la distribution sur toutes les valeurs possibles de nos paramètres), $p(x|\theta)$ est la vraisemblance de l'observation x conditionnée sur les valeurs de θ , et $p(x)$ est l'évidence. Puisque celle-ci ne dépend pas de θ , il est commun de l'omettre si on s'intéresse seulement à la distribution des valeurs de θ .

Dans plusieurs classes de problèmes, on connaît $p(\theta)$ et on dispose d'un simulateur, mais la vraisemblance $p(x|\theta)$ peut être soit très dispendieuse à évaluer explicitement, soit impossible à écrire sous forme fermée. Dans le formalisme de la méthode SBI, on se base sur une approche qui n'évalue pas explicitement la vraisemblance (d'où l'appellation d'inférence sans vraisemblance). On cherche plutôt, dans un premier temps, à utiliser le simulateur pour produire des observations x simulées à partir de paramètres θ tirés de la distribution à priori $p(\theta)$. À partir de ces paires $\{\theta, x\}$, on obtient des échantillons de la distribution jointe $p(\theta, x)$. Dans un second temps, on cherche à modéliser une fonction continue $f = \hat{p}(\theta|x)$ qui approxime $p(\theta|x)$. Dans ce contexte, la fonction f est représentée par un réseau de neurones qui est entraîné sur une banque d'exemples $\{\theta, x\}_i$ pour maximiser $\sum_i \log \hat{p}(\theta_i|x_i)$. Des contraintes sont imposées sur le réseau pour que la fonction modélisée f ait les propriétés

d'une fonction de probabilité, c'est-à-dire qu'elle soit lisse et normalisée. On obtient ainsi notre probabilité à posteriori $\hat{p}(\theta|x)$, et on peut ensuite échantillonner pour construire une distribution en l'évaluant à plusieurs valeurs de x .

Dans notre cas, on cherche $p(y_{\text{vrai}}|y_{\text{prédit}})$, et donc notre paramètre θ correspond à la vraie valeur y_{vrai} , tandis que l'observation x correspond à la valeur prédite $y_{\text{prédit}}$. Notre distribution à priori $p(\theta)$ est une distribution uniforme entre les valeurs minimale et maximale de y . Notre banque d'exemples $\{\theta, x\}_i$ est représentée par les paires de valeurs $(y_{\text{vrai}}, y_{\text{prédit}})$ de l'ensemble d'entraînement. Une fois qu'on a obtenu la fonction $f = \hat{p}(y_{\text{vrai}}|y_{\text{prédit}})$, on peut échantillonner plusieurs valeurs de y_{vrai} et obtenir une distribution de probabilité pour le $y_{\text{prédit}}$ qui nous intéresse.

Bibliographie

- Abell, G. O. 1958, , 3, 211
- Allen, S. W., Evrard, A. E., & Mantz, A. B. 2011, , 49, 409
- Barnes, D. J., Vogelsberger, M., Kannan, R., et al. 2018, Monthly Notices of the Royal Astronomical Society, 481, 1809
- Bartelmann, M. 2010, Classical and Quantum Gravity, 27, 233001
- Bartelmann, M. & Schneider, P. 2001, , 340, 291
- Blandford, R. D. & Payne, D. G. 1982, , 199, 883
- Blandford, R. D. & Rees, M. J. 1974, , 169, 395
- Blandford, R. D. & Znajek, R. L. 1977, , 179, 433
- Borgani, S., Rosati, P., Tozzi, P., et al. 2001, , 561, 13
- Bradač, M., Allen, S. W., Treu, T., et al. 2008, , 687, 959
- Bradt, H., Mayer, W., Naranan, S., Rappaport, S., & Spada, G. 1967, , 150, L199
- Broadhurst, T. J., Taylor, A. N., & Peacock, J. A. 1995, , 438, 49
- Byram, E. T., Chubb, T. A., & Friedman, H. 1966, Science, 152, 66
- Cavagnolo, K. W., Donahue, M., Voit, G. M., & Sun, M. 2009, The Astrophysical Journal Supplement Series, 182, 12
- Chibueze, J. O., Sakemi, H., Ohmura, T., et al. 2021, , 593, 47
- Churazov, E., Brüggén, M., Kaiser, C. R., Böhringer, H., & Forman, W. 2001, , 554, 261
- Churazov, E., Sazonov, S., Sunyaev, R., et al. 2005, Monthly Notices of the Royal Astronomical Society: Letters, 363, L91
- Cui, W., Power, C., Biffi, V., et al. 2016, , 456, 2566
- Davé, R., Anglés-Alcázar, D., Narayanan, D., et al. 2019, , 486, 2827
- Di Valentino, E., Mena, O., Pan, S., et al. 2021, Classical and Quantum Gravity, 38, 153001

- Dubois, Y., Pichon, C., Welker, C., et al. 2014, , 444, 1453
- Fabian, A. 2012, *Annual Review of Astronomy and Astrophysics*, 50, 455
- Fabian, A. C., Sanders, J. S., Crawford, C. S., et al. 2003, , 344, L48
- Fabian, A. C., Sanders, J. S., Taylor, G. B., et al. 2006, , 366, 417
- Giacconi, R., Kellogg, E., Gorenstein, P., Gursky, H., & Tananbaum, H. 1971, , 165, L27
- Giacconi, R., Branduardi, G., Briel, U., et al. 1979, , 230, 540
- Herschel, W. 1786, *Philosophical Transactions of the Royal Society of London Series I*, 76, 457
- Hubble, E. 1929, *Proceedings of the National Academy of Science*, 15, 168
- Hubble, E. & Humason, M. L. 1931, , 74, 43
- Hudson, D. S., Mittal, R., Reiprich, T. H., et al. 2010, , 513, A37
- Jauzac, M., Harvey, D., & Massey, R. 2018, , 477, 4046
- Khandai, N., Di Matteo, T., Croft, R., et al. 2015, , 450, 1349
- Kneib, J.-P. & Natarajan, P. 2011, , 19, 47
- Lemaître, G. 1927, *Annales de la Société Scientifique de Bruxelles*, 47, 49
- Marinacci, F., Vogelsberger, M., Pakmor, R., et al. 2018, , 480, 5113
- McAlpine, S., Helly, J. C., Schaller, M., et al. 2016, *Astronomy and Computing*, 15, 72
- McDonald, M., Gaspari, M., McNamara, B. R., & Tremblay, G. R. 2018, , 858, 45
- McDonald, M., Benson, B. A., Vikhlinin, A., et al. 2013, , 774, 23
- McDonald, M., McNamara, B. R., van Weeren, R. J., et al. 2015, , 811, 111
- McNamara, B. R. & Nulsen, P. E. J. 2007, *ARA&A*, 45, 117
- McNamara, B. R., Wise, M. W., & Murray, S. S. 2004, , 601, 173
- Merritt, D. 2013, *Dynamics and Evolution of Galactic Nuclei*
- Messier, C. 1781, *Catalogue des Nébuleuses et des Amas d'Étoiles (Catalog of Nebulae and Star Clusters)*, *Connaissance des Temps ou des Mouvements Célestes*, for 1784, p. 227-267
- Million, E. T. & Allen, S. W. 2009, , 399, 1307
- Minkowski, R. L. & Abell, G. O. 1963, in *Basic Astronomical Data : Stars and Stellar Systems*, ed. K. A. Strand, 481
- Mohr, J. J., Mathiesen, B., & Evrard, A. E. 1999, , 517, 627
- Naiman, J. P., Pillepich, A., Springel, V., et al. 2018, , 477, 1206
- Nelson, D., Pillepich, A., Springel, V., et al. 2018, , 475, 624
- Nelson, D., Springel, V., Pillepich, A., et al. 2019, *Computational Astrophysics and Cosmology*, 6, 2
- Osinga, E., van Weeren, R. J., Andrade-Santos, F., et al. 2022, , 665, A71
- Peterson, J. R. & Fabian, A. C. 2006, , 427, 1
- Pierpaoli, E. 2011, in *Adventures in Cosmology* (World Scientific)
- Pierpaoli, E., Borgani, S., Scott, D., & White, M. 2003, , 342, 163
- Pillepich, A., Nelson, D., Hernquist, L., et al. 2018, , 475, 648

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, , 641, A6

Riemer-Sorensen, S., Pedersen, K., Hansen, S. H., & Dahle, H. 2007, , 76, 043524

Ryden, B. S. 2003, Introduction to Cosmology (Addison Wesley)

Sand, D. J., Treu, T., Smith, G. P., & Ellis, R. S. 2004, , 604, 88

Santos, J. S., Rosati, P., Tozzi, P., et al. 2008, , 483, 35

Sarazin, C. 2007, in Heating versus Cooling in Galaxies and Clusters of Galaxies, ed. H. Böhringer, G. Pratt, A. Finoguenov, & P. Schuecker (Berlin, Heidelberg : Springer Berlin Heidelberg), 3

Sarazin, C. L. 1986, Reviews of Modern Physics, 58, 1

Springel, V. 2010, , 401, 791

Springel, V., Pakmor, R., Pillepich, A., et al. 2018, , 475, 676

Sutherland, R. S. & Dopita, M. A. 1993, , 88, 253

Tremmel, M., Karcher, M., Governato, F., et al. 2017, , 470, 1121

Umetsu, K. 2020, , 28, 7

Umetsu, K., Birkinshaw, M., Liu, G.-C., et al. 2009, , 694, 1643

Vikhlinin, A., Burenin, R., Forman, W. R., et al. 2007, in Heating versus Cooling in Galaxies and Clusters of Galaxies, ed. H. Böhringer, G. W. Pratt, A. Finoguenov, & P. Schuecker, 48

Vikhlinin, A., Kravtsov, A., Forman, W., et al. 2006, , 640, 691

Vogelsberger, M., Marinacci, F., Torrey, P., & Puchwein, E. 2020, Nature Reviews Physics, 2, 42

Weinberger, R., Springel, V., Pakmor, R., et al. 2018, , 479, 4056

Weisskopf, M. C., Tananbaum, H. D., Van Speybroeck, L. P., & O'Dell, S. L. 2000, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4012, X-Ray Optics, Instruments, and Missions III, ed. J. E. Truemper & B. Aschenbach, 2

White, N. E. & Holt, S. S. 1982, , 257, 318

Wolter, H. 1952, Annalen der Physik, 445, 94

Zhuravleva, I., Churazov, E., Schekochihin, A. A., et al. 2014, , 515, 85

ZuHone, J. A., Biffi, V., Hallman, E. J., et al. 2014, arXiv e-prints, arXiv:1407.1783

Zwicky, F. 1933, Helvetica Physica Acta, 6, 110

Chapter 2

Galaxy cluster characterization with machine learning techniques

M. Sadikov,^{1,2} J. Hlavacek-Larrondo,^{1,2} L. Perreault-Levasseur,^{1,2,3,4} C. L. Rhea,^{1,2} M. McDonald,⁵ M. Ntampaka,^{6,7} J. ZuHone⁸

¹*Département de physique, Université de Montréal, C.P. 6128 Succ. Centre-ville, Montréal H3C 3J7, Canada*

²*Centre de recherche en astrophysique du Québec (CRAQ)*

³*Mila - Québec Artificial Intelligence Institute, Montréal, Canada*

⁴*Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, New York, NY 10010, USA*

⁵*Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

⁶*Center for Astrophysics | Harvard & Smithsonian, Cambridge, MA 02138, USA*

⁷*Harvard Data Science Initiative, Harvard University, Cambridge, MA 02138, USA*

⁸*Smithsonian Astrophysical Observatory, Cambridge, MA 02138 USA*

Cet article est en cours de préparation pour soumission à la revue *The Astrophysical Journal* (ApJ).

Abstract

We present an analysis of the X-ray properties of the galaxy cluster population in the $z = 0$ snapshot of the IllustrisTNG simulations, utilizing machine learning techniques to perform clustering and regression tasks. We study several gas properties which are commonly used as classification metrics to identify cool core (CC), weak cool core (WCC) and non cool core (NCC) clusters of galaxies, and we wish to determine which metrics are most suited for a machine learning analysis of galaxy cluster images. More precisely, we examine five properties of the hot gas: the central cooling time, the central electron density, the central entropy excess, the concentration parameter, and the cuspidity. Using mock *Chandra* X-ray images as inputs, we first explore an unsupervised clustering scheme to see how the resulting groups correlate with the CC/WCC/NCC classification based on the different criteria. We observe that the groups replicate almost exactly the separation of the galaxy cluster images when classifying them based on the concentration parameter. We then move on to a regression task, utilizing a ResNet model to predict the value of all five properties. The network is able to achieve a mean percentage error of 1.8% for the central cooling time, and a balanced accuracy of 0.83 on the concentration parameter, making them the best-performing metrics. Finally, we use simulation-based inference (SBI) to extract posterior distributions for the network predictions. We observe that the predictions for NCC clusters have generally lower uncertainty than for CC clusters. This study demonstrates that machine learning is a viable path to assist in the analysis and classification of the large galaxy cluster datasets that will soon become available with the upcoming X-ray surveys, such as eROSITA.

Keywords: galaxy clusters, machine learning, X-rays, simulations

2.1. Introduction

Galaxy clusters are the largest virialized structures of our universe. They are made up of hundreds to thousands of galaxies, dark matter, and intracluster gas known as the intracluster medium (ICM). The dark matter makes up $\sim 84\%$ of the mass, the galaxies make up $\sim 3\%$ of the mass, and the ICM makes up $\sim 13\%$ of the mass (Mohr et al. 1999, Vikhlinin et al. 2006, Umetsu et al. 2009).

X-ray observations of galaxy clusters probe the hot ICM. The ICM is a diffuse gas, mainly made up of ionized hydrogen and helium, but also containing iron, silicon, sulfur, argon, carbon and nickel (e.g. Loewenstein 2004). Due to its high temperature ($\approx 10^7$ K), the ICM emits strongly at X-ray wavelengths through Bremsstrahlung emission. Galaxy clusters are often separated into three classes based on the distribution of the ICM within

the cluster (e.g. Million & Allen 2009, Hudson et al. 2010): cool core clusters (CC), weak cool core clusters (WCC), and non-cool core clusters (NCC). In cool core clusters, the ICM is strongly concentrated towards the center of the cluster (i.e. the inner 100 kpc; White et al. 1997, Hudson et al. 2010, McDonald et al. 2013), resulting in strong X-ray emission. This leads to a temperature decrease in the core, which results in a pressure decrease as well. However, the gas in the center of the cluster must be able to support the weight of the outer layers, and therefore it becomes denser to prevent the outer layers from collapsing. This creates a cooling flow, where cool gas is infalling towards the center of the cluster. This cool gas should lead to a star formation rate on the order of 100-1000 M_{\odot}/yr (e.g. Peterson & Fabian 2006). However, the observed star formation rate of $\sim 10\text{-}300 M_{\odot}/\text{yr}$ (e.g. McNamara et al. 2004, McDonald et al. 2015) suggests that some mechanism is offsetting the cooling. That mechanism is most likely active galactic nucleus (AGN) feedback from the brightest cluster galaxy (BCG); (see review by Hlavacek-Larrondo et al. 2022). CC clusters therefore give an important insight into the way the supermassive black hole (SMBH) at the center of the BCG interacts with the surrounding gas through AGN feedback processes. On the other hand, clusters with a more uniform distribution of the ICM are non-cool core clusters. NCCs are also more disturbed, while CCs are generally relaxed. Finally, we often add a third category of weak cool core clusters as an intermediate between the two (e.g. Bauer et al. 2005, Sun 2009, Hudson et al. 2010, Bharadwaj et al. 2014).

We expect to learn more about galaxy clusters in the following years from the X-ray telescope eROSITA (extended ROentgen Survey with an Imaging Telescope Array). Built by the Max Planck Institute for Extraterrestrial Physics (MPE), it is the primary instrument on the Russian-German ‘Spectrum-Roentgen-Gamma’ (SRG) mission. Launched in 2019, it is currently performing an all-sky survey within an energy range of 0.2 - 10 keV and it is expected to detect at least 100 000 massive galaxy clusters (Merloni et al. 2012, Pillepich et al. 2012). With this large amount of data to analyze comes the need for an efficient, fast, and automated way to characterize the ICM in galaxy clusters. We aim to do this using neural networks, which are able to process large datasets much faster than traditional methods and have proven to be very effective in extracting meaningful features from astrophysical images (e.g. Tarsitano et al. 2022). Machine learning tools have already been successfully applied to galaxy cluster analysis, for example to estimate galaxy cluster masses (Ntampaka et al. 2019, de Andres et al. 2022, Krippendorff et al. 2023), to infer galaxy cluster mass profiles (Ferragamo et al. 2023), to find galaxy cluster members (Angora et al. 2020), or to classify galaxy clusters into CC, WCC and NCC (Su et al. 2020).

Multiple studies aim to examine the properties of galaxy clusters within the IllustrisTNG cosmological hydrodynamical simulations, notably Barnes et al. (2018) (hereafter B18),

Ntampaka et al. (2019) and Su et al. (2020) (hereafter S20). B18 draw a general portrait of the galaxy cluster population within the IllustrisTNG simulations, without the use of machine learning. They examine several classification criteria for CC clusters and look at correlations between the different properties as well as their evolution with redshift. They find that the different classification criteria do not agree over the entire cluster population, and that the correlations between different criteria are dependent on cluster mass. They also note that at low redshift ($z < 0.25$), the CC fraction in TNG is lower than in observations for 4 of their 6 criteria.

Ntampaka et al. (2019) utilize machine learning to analyze the IllustrisTNG clusters. They use a convolutional neural network (CNN) to estimate galaxy clusters masses from simulated *Chandra* X-ray images. Their results show that a low resolution (128×128 pixels) image can be successfully used to predict cluster mass with a reasonably low scatter (12%). They also perform an interpretability analysis with Google DeepDream tools, allowing to see which parts of the cluster image are most significant for the prediction.

On the other hand, S20 tackle a classification problem. They use a residual network (ResNet; He et al. 2015) with convolutional layers to classify mock *Chandra* images of IllustrisTNG galaxy clusters into CC, WCC and NCC clusters. They use the radiative cooling time to define their class labels, and obtain an average balanced accuracy (BAcc) of 0.85. Using class activation maps, they are able to locate which regions are discriminative for the network. Their results show that an X-ray image of a galaxy cluster contains enough information for a neural network to classify the cluster as CC, WCC or NCC when using the cooling time as the classification criteria.

In this work, we continue the analysis by exploring how other galaxy cluster classification metrics behave under machine learning analysis. We present a deep learning approach for galaxy cluster characterization. We analyze five properties of the hot gas in galaxy clusters, often used as classification metrics: the central cooling time, the central electron density, the central entropy excess, the concentration parameter, and the cusiness parameter. We select these properties as they are the most widely used in the literature for galaxy cluster classification (e.g. Cavagnolo et al. 2009, Santos et al. 2010, Santos et al. 2008, Hudson et al. 2010, Andrade-Santos et al. 2017, Barnes et al. 2018, McDonald et al. 2019). However, as hinted in Barnes et al. (2018), the classification results from the five metrics don't always agree, and the choice of the best metric to use is subject to debate. Our work aims to compare the values of the five metrics, analyzing how informative they are and how well they lend themselves to machine learning analysis. We construct a neural network that simultaneously predicts the values of all five cluster properties from mock X-ray images; we

then analyze the predictions using simulation-based inference. The paper is structured as follows. In Section 2.2.1, we describe the IllustrisTNG simulations, the mock *Chandra* images as well as the input preprocessing steps. We present the five chosen classification metrics in Section 2.2.2. In Section 2.3, we describe our approach for the clustering algorithm and present corresponding results. In Section 2.4, we present our network architecture for the regression task as well as results. In Section 2.5, we present our simulation-based inference (SBI) approach to obtain posterior probability distributions for our network predictions and we discuss our results. We conclude in Section 2.6.

2.2. Simulated Data and Metrics

It is known that training a supervised machine learning algorithm requires large amounts of labeled data. Since observational datasets of galaxy clusters for which the five classification metrics have been calculated are very limited, we turn towards simulations to provide a larger dataset. We select the IllustrisTNG simulations as the large volumes (300 comoving Mpc) provided by the simulations yield a large number of massive galaxy clusters to analyze.

2.2.1. Simulated Data

2.2.1.1. *IllustrisTNG Clusters*

We use galaxy cluster data from the cosmological gravo-magnetohydrodynamical simulations IllustrisTNG (Nelson et al. 2019, Pillepich et al. 2018, Springel et al. 2018, Nelson et al. 2018, Naiman et al. 2018, Marinacci et al. 2018), run with the AREPO (Springel 2010) dynamical mesh code. They use a Λ CDM model, assuming cosmological constants from the Planck collaboration results (Planck Collaboration et al. 2016). The simulations are made up of three volumes: TNG50, TNG100 and TNG300; the number refers to the size of the simulation box in units of comoving Mpc. The large volumes as well as high resolution ($\lesssim 1$ kpc scales; Nelson et al. 2019) provided by these simulations make them suitable for studying the gas in galaxy clusters. Each simulation includes dark matter, gas, stars and supermassive black holes, solving for the coupled evolution of all components through time, with a redshift range of $z = 127$ to $z = 0$.

We select clusters with a mass $M_{500} > 10^{13.57} M_{\odot}$ from the $z = 0$ snapshot of the TNG300 simulation, similar to the work in Ntampaka et al. (2019). We chose to work with a single redshift value because we do not want our network to consider galaxy cluster evolution through time, and $z = 0$ is the snapshot that contains the largest number of galaxy clusters as they have had more time to form. Haloes are identified using the Friends-of-Friends algorithm, run on the dark matter particles. This results in 606 clusters, with an R_{500} range of 517 kpc to 1624 kpc. The value for R_{500} is taken directly from the IllustrisTNG data, using the variable `Group_R_Crit500` and $h = 0.68$.

2.2.1.2. *Center of the cluster*

In order to calculate the gas properties serving as galaxy cluster classification metrics, as well as generate mock X-ray emission, we need to locate the center of the cluster, which in itself can be a complicated task. There are numerous ways of defining the center of a

galaxy cluster, such as the location of the BCG, the minimum of the gravitational potential, the X-ray/optical centroid or the X-ray/optical peak. These definitions result in centers that are offset from one another, in observational data (e.g. McDonald et al. 2014), as well as in simulated data (Cui et al. 2016).

Since we are working with X-ray data and are interested in the morphology of the cluster, we experiment with the X-ray large-scale centroid as well as the X-ray peak. The large-scale centroid is an appropriate choice for NCC clusters, as the X-ray emission is more or less uniform. However, for CC clusters, the core is often offset from the large-scale centroid, as the TNG clusters aren't perfectly spherical. This is also the case for real galaxy clusters (e.g. Lazzati & Chincarini 1998). We're interested in locating the core for these clusters, and therefore define the center of the cluster as the X-ray peak. The calculated metrics (Section 2.2.2) for NCC clusters are not affected by this choice since the gas distribution is more uniform.

To find the X-ray peak, we look at the smoothed 3D distribution of the gas density. We only consider the X-ray emitting gas, i.e. the gas cells with a temperature above 10^7 K. We wish to see where the data most resembles a 3D gaussian peak, and select that point to represent the central peak of the cluster, with a precision of 1 kpc. To do this, we apply a gaussian filter to the data with $\sigma = 5$ kpc, where σ is the standard deviation of the gaussian kernel. Since convolving a 3D kernel with a precision of 1 kpc over the entire cluster would be too time-consuming, we use an iterative approach, starting with a coarse grid and gradually transitioning to a finer grid. We first bin the data into 10 kpc cubes, and convolve the binned data with a gaussian filter that has $\sigma = 250$ kpc. The 500 kpc box around the maximum point of this convolution corresponds to a rough estimate of the central region. We then select the cells of this region and repeat the process: we bin the data into 5 kpc cubes and convolve it with a $\sigma = 50$ kpc gaussian filter. Finally, we select the cells within a 100 kpc box around the new maximum and repeat the process a third time with 1 kpc bins and $\sigma = 5$ kpc. We define the maximum point of the last convolution as the location of the peak of the gas density, corresponding to the 3D cluster center. .

2.2.1.3. *Mock Chandra Observations*

We generate mock X-ray observations of the clusters with the `pyXSIM` (version 3.0.1) and `SOXS` (version 3.0.2) packages (ZuHone et al. 2014). The `pyXSIM` package is an implementation of the `PHOX` algorithm (Biffi et al. 2012, Biffi et al. 2013), and it simulates X-ray emission from astrophysical sources. Using the gas properties extracted from the `IllustrisTNG` data for each cluster, such as density, temperature and metallicity, photons are generated in a 3D space with the `APEC` emission model (Foster et al. 2012). We specify the

calculated 3D center (see Section 2.2.1.2) as the center of the spherical region from which the photons are generated, and we assume a redshift of $z = 0.05$, as done in Ntampaka et al. (2019) and S20. The photons are then projected along the three axes of the simulation box to create three separate images. We chose to only take these 3 projections along the perpendicular axes, as a compromise between data augmentation and ensuring that the different samples remain as independent as possible. We then apply foreground galactic absorption with the `tbabs` model (Wilms et al. 2000), assuming a typical value of $N_H = 0.04 \times 10^{22} \text{ cm}^{-2}$ for the column density (e.g. Ntampaka et al. 2019, S20).

The simulated photons landing on the detector are then convolved with an instrumental response matrix using the `SOXS` package. This package creates simulated X-ray observations of X-ray sources, and has numerous simulated instruments available including realistic instrumental noise. We select *Chandra*'s ACIS-I detector with a configuration corresponding to the Cycle 0, as its large field of view is well-suited for observing low- z clusters. We integrate for a typical exposure time of 100.0 ks, which ensures that we detect enough photons. The final mock images display the soft-band X-ray emission (between 0.5 and 2.0 keV) typical of the ICM in galaxy clusters.

2.2.1.4. Image preprocessing

As a preprocessing step, we use a min-max normalization ($\frac{\text{value}-\text{min}}{\text{max}-\text{min}}$) for every image, constraining the value of the pixels between 0 and 1 so that all the inputs have the same scale. However, the maximum point of the mock X-ray image is sometimes a point source in the background, which is not informative for the cluster morphology. Therefore, we first remove point sources from our mock *Chandra* images, using the CIAO 4.15 software package (Fruscione et al. 2006). We locate the point source regions with `wavdetect`, and then use `dmfilth` to replace the source pixels with values interpolated from background regions. Once the point sources have been removed, we verify manually that the brightest pixel in each image actually belongs to the galaxy cluster in the X-ray image and not a background point source.

In order to reduce the dimension of the input, we downsample the images to a size of 256×256 pixels by applying 8×8 binning. We then apply a min-max normalization to each image. Finally, the dataset is augmented with flips along the vertical and horizontal axes as well as with 90° rotations, which results in 8 different views of the same image. Considering the 2D projections along the three axes, we end up with 24 images per cluster and a total of 14 544 images for our dataset.

2.2.2. Galaxy cluster classification metrics

In order to characterize the galaxy clusters, we have chosen five classification metrics: the central cooling time, the central electron density, the entropy excess, the cuspliness parameter and the concentration parameter. The first four metrics are calculated with information from the IllustrisTNG gas cells, whereas the concentration parameter depends solely on the mock X-ray image.

When observing galaxy clusters in the X-ray range, the pattern on the detector is a projection of all the photons along the line of sight, since the gas is optically thin. In an effort to mimic observations, we use projected 2D values to calculate the metrics of interest. We project our data along the three axes of the simulation box, and compute 2D centers for the three projections of each cluster. We use the same iterative method described in Section 2.2.1.2 to compute the center, but now convolve a 2D gaussian filter over the projected 2D distribution of the gas density. The values for bin width and for σ stay the same as above. We then construct radial profiles for the electron density and the temperature of the gas, using 50 radial logarithmic bins in the range 10^{-3} - $1.5 R_{500}$, as done in B18. For each bin, we take a weighted mean of the observed value (the electron density or the temperature) for the gas cells within that annulus. The value is weighted by the emission measure, which is defined as the square of the electron density. We use the weighted standard deviation as the uncertainty for each bin.

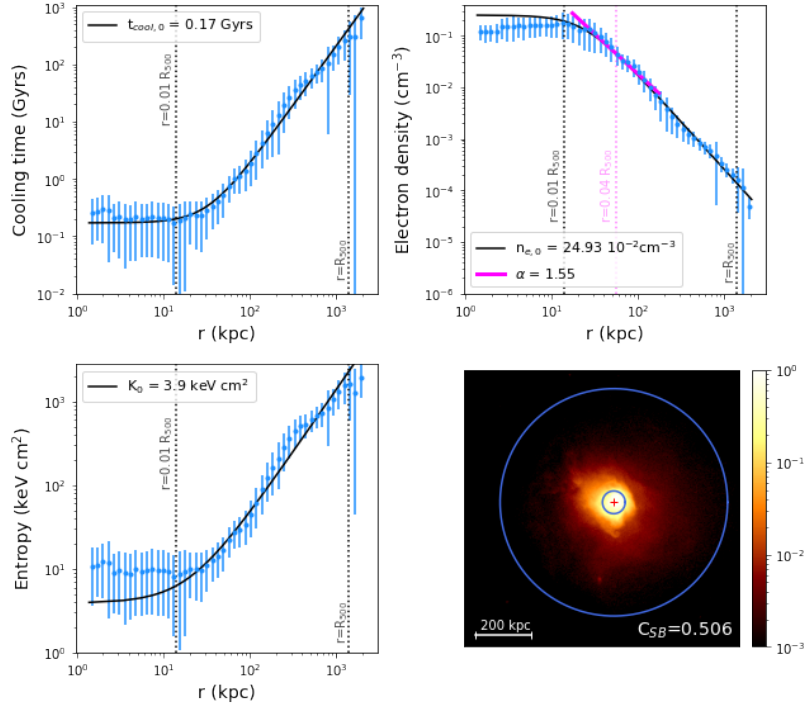
Since we are calculating projected 2D radial profiles, we end up with different observed values for the three different projections of the same cluster, and therefore different values for the metrics. Thus, we refer to 1818 observations (rather than 606 clusters) when talking about the distribution of the metrics, which we will discuss in more detail in the following sections.

2.2.2.1. Central cooling time

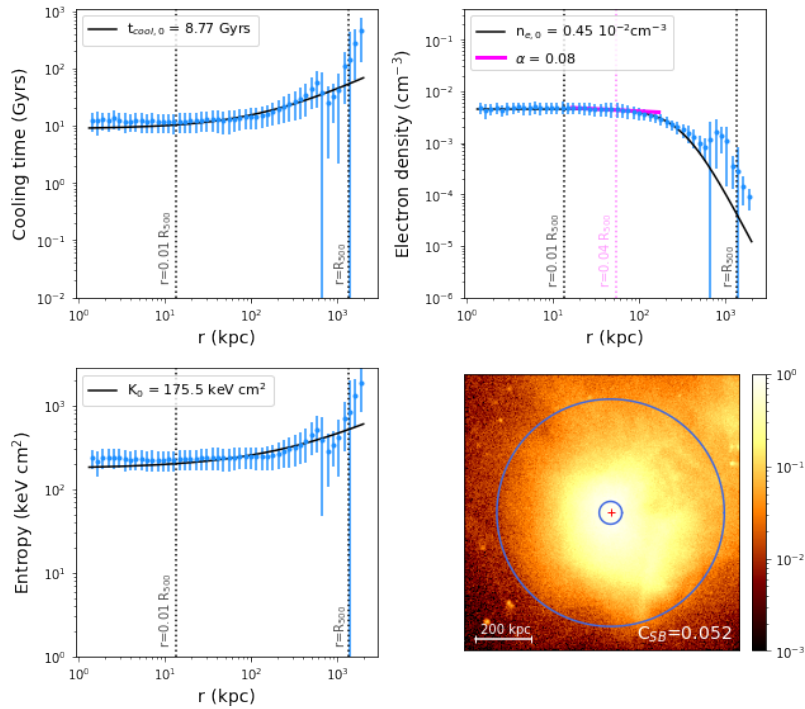
The central cooling time is a widely used metric (Hudson et al. 2010, McDonald et al. 2013) to classify galaxy clusters into CC, WCC and NCC. It is defined as

$$t_{cool} = \frac{3(n_e + n_p)kT}{2n_e n_H \Lambda(T, Z)} \quad (2.2.1)$$

where Λ is the cooling function for an optically thin plasma (Sutherland & Dopita 1993). We assume $n_p = 0.92n_e$ and $n_H = 0.83n_e$, as done in McDonald et al. (2018). We use the cooling function table for a metallicity of $[\text{Fe}/\text{H}] = -0.5$ from Sutherland & Dopita (1993),



(a) Properties for a CC cluster.



(b) Properties for a NCC cluster.

Figure 2.1 – (Continued on the following page)

interpolating $\log(T)$ to obtain $\log(\Lambda)$.

Figure 2.1 – Example of cluster properties for a cool core cluster in (a) and a non-cool core cluster in (b). In the top left, is shown the 2D cooling time profile. The data points are in blue, the best-fit model is in black and the fitted values are shown in the legend. The black dotted vertical lines at $0.01 R_{500}$ and R_{500} represent the boundaries for the fit, as discussed in Sections 2.2.2.1 through 2.2.2.3. In the top right, is shown the 2D electron density profile, as well as the tangent line in pink that is used to calculate the cuspliness at $r = 0.04R_{500}$ (see 2.2.2.5). The fit value for $n_{e,0}$ and the cuspliness α are shown in the legend as well. In the bottom left, is shown the 2D entropy profile, with the fit value for K_0 shown in the legend. For each property, the same y axes are used for both plots (CC cluster and NCC cluster). In the bottom right, is shown the preprocessed (binned and normalized) 2D X-ray image that is used as input for the neural network. The red cross represents the cluster center, identified using the method in Section 2.2.1.2, and the blue circles at $r = 40$ kpc and $r = 400$ kpc represent the regions used in the calculation of the concentration parameter C_{SB} (see 2.2.2.4). The value for C_{SB} is shown in the bottom right. The colorbar for the X-ray image is shown in log scale.

Rather than only considering the gas cells at the center of the cluster to compute the central cooling time, we opt to use the entire cooling time profile and fit for the central value, especially since the TNG data tends to be noisy and often have multiple small overdensities near the center that could bias our estimate. We use our 50-bin density and temperature radial profiles to obtain a cooling time profile within a range of 10^{-3} to $1.5 R_{500}$. We also propagate uncertainties through equation 2.2.1. We follow the method used in B18 and Cavagnolo et al. (2009) to calculate the central entropy excess, and expand the approach to the cooling time. More precisely, we fit the cooling time in the range $0.01 - 1.0R_{500}$ with a power-law of the form:

$$t_{\text{cool}}(r) = t_{\text{cool},0} + t_{\text{cool},100} \left(\frac{r}{100\text{kpc}} \right)^\gamma, \quad (2.2.2)$$

where $t_{\text{cool},0}$ is the central cooling time, $t_{\text{cool},100}$ is a normalization factor at 100 kpc, r is the radial distance from the center and γ is the power-law index (Cavagnolo et al. 2009). We fit this model to our data to get a value for the central cooling time $t_{\text{cool},0}$. We do this with a Markov-Chain Monte Carlo (MCMC) with uniform priors, using the `emcee` python package (Foreman-Mackey et al. 2013). We use 100 walkers and run the MCMC for 10 000 steps. We also perform sigma clipping to remove outliers (e.g. small overdensities or numerical artefacts) outside of 3σ prior to fitting, where σ is the standard deviation. We use the $t_{\text{cool},0}$ parameter as the classification metric, where clusters with $t_{\text{cool},0} < 1$ Gyr are defined as CCs, those with $1 \text{ Gyr} < t_{\text{cool},0} < 7.7 \text{ Gyr}$ are WCCs, and those with $t_{\text{cool},0} > 7.7 \text{ Gyr}$ are NCCs (Hudson et al. 2010, McDonald et al. 2013, B18). Out of our 1818

observations, we have 50 CCs, 908 WCCs and 860 NCCs with these thresholds. The distribution of central cooling time values is shown in the first panel (from the top) of figure 2.2.

2.2.2.2. Central electron density

For the central electron density, we use the same method as for the central cooling time (described in 2.2.2.1) to construct the radial profile that we fit with a MCMC. We use the 50-bin density radial profile with a range of 10^{-3} - $1.5 R_{500}$, and fit it in the range 0.01 - 1.0 R_{500} using the following equation:

$$n_e(r) = n_{e,0} \left(1 + \frac{r^2}{r_c^2} \right)^{-3\beta/2}, \quad (2.2.3)$$

where $n_e(r)$ is the electron density, r is the radial distance from the center, r_c is the core radius, $n_{e,0}$ is the central value of the electron density, and β is a power-law index (Cavaliere & Fusco-Femiano 1976, Cavaliere & Fusco-Femiano 1978, Grego et al. 2001, Reese et al. 2002, Cao et al. 2016). We use a MCMC to fit the model to our data, and use the $n_{e,0}$ parameter as a the metric, where clusters with $n_{e,0} > 1.5 \times 10^{-2} \text{cm}^{-3}$ are defined as CCs, those with $0.5 \times 10^{-2} \text{cm}^{-3} < n_{e,0} < 1.5 \times 10^{-2} \text{cm}^{-3}$ are WCCs, and those with $n_{e,0} < 0.5 \times 10^{-2} \text{cm}^{-3}$ are NCCs (Hudson et al. 2010, B18). Out of our 1818 observations, we have 68 CCs, 175 WCCs and 1575 NCCs with these thresholds. The distribution of central electron density values is shown in the second panel (from the top) of figure 2.2.

2.2.2.3. Central entropy excess

The entropy K is defined as

$$K = \frac{k_B T}{n_e^{2/3}}, \quad (2.2.4)$$

where k_B is the Boltzmann constant, T is the temperature and n_e is the electron density. Once again, following the method described in 2.2.2.1, we use the 50-bin density and temperature radial profiles to obtain an entropy radial profile with a range of 10^{-3} - $1.5 R_{500}$, and we propagate uncertainties through equation 2.2.4. We fit the entropy K in the range 0.01 - 1.0 R_{500} with a power-law of the form:

$$K(r) = K_0 + K_{100} \left(\frac{r}{100 \text{kpc}} \right)^\gamma, \quad (2.2.5)$$

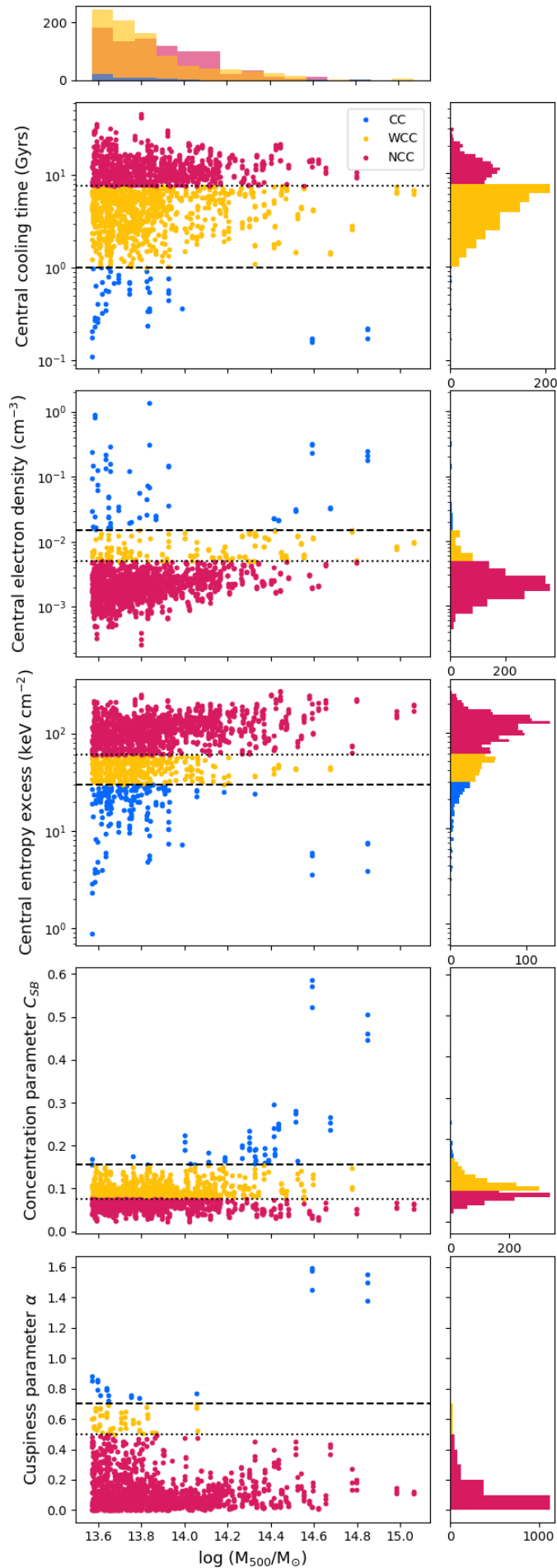


Figure 2.2 – Distribution of cluster properties (central cooling time, central electron density, central entropy excess, concentration and cuspiness) as a function of the cluster mass for the 1818 observations in our sample. The CCs for each metric are plotted in blue, the WCCs are in yellow and the NCCs are in red. The colors in the mass histogram at the top are based on the central cooling time, as it is the most commonly used metric.

where r is the radial distance from the center, K_{100} is a normalization factor at 100 kpc, γ is a power-law index, K_0 is the central entropy excess, defined as the excess entropy above the best-fitting power-law at larger radius (Cavagnolo et al. 2009). We use a MCMC to fit the model to our data, and use the K_0 parameter as the metric. Clusters with $K_0 < 30 \text{ keV cm}^2$ are defined as CCs, those with $30 \text{ keV cm}^2 < K_0 < 60 \text{ keV cm}^2$ are WCCs, and those with $K_0 > 60 \text{ keV cm}^2$ are NCCs (B18). Out of our 1818 observations, we have 185 CCs, 390 WCCs and 1243 NCCs with these thresholds. The distribution of central entropy excess values is shown in the third panel (from the top) of figure 2.2.

2.2.2.4. Concentration parameter

Introduced by Santos et al. (2008), the X-ray concentration parameter is the ratio of the X-ray luminosity within the core compared to the luminosity within a larger radius:

$$C_{SB} = \frac{F(r < 40\text{kpc})}{F(r < 400\text{kpc})}, \quad (2.2.6)$$

where C_{SB} is the concentration parameter (the subscript SB refers to the surface brightness), F is the X-ray flux, and r is the projected radial distance from the center. We compute C_{SB} by taking the ratio of photons within 40 kpc and within 400 kpc in the mock X-ray images, as shown in the bottom right panel of figures 2.1 (a) and (b). C_{SB} is a dimensionless number between 0 and 1, and clusters with $C_{SB} > 0.155$ are defined as CCs, those with $0.075 < C_{SB} < 0.155$ are WCCs, and those with $C_{SB} < 0.075$ are NCCs (Santos et al. 2008). Out of our 1818 observations, we have 57 CCs, 858 WCCs and 903 NCCs with these thresholds. The distribution of concentration values is shown in the fourth panel (from the top) of figure 2.2.

2.2.2.5. Cuspiness

Introduced by Vikhlinin et al. (2007), the cuspiness parameter α aims to quantify the importance of the central cusp in the X-ray brightness distribution, and is defined as:

$$\alpha = - \left. \frac{d \log n_e(r)}{d \log r} \right|_{r=0.04R_{500}}, \quad (2.2.7)$$

where r is the radial distance from the center and n_e is the electron density. Since we have a theoretical profile for the density (equation 2.2.3 with parameters $n_{e,0}$, β and r_c estimated by the MCMC), we can compute the derivative $\frac{d \log n_e(r)}{d \log r}$, evaluated at $r = 0.04R_{500}$. We then use α as the metric, where clusters with $\alpha > 0.7$ are defined as CCs, those with $0.5 < \alpha < 0.7$ are WCCs, and those with $\alpha < 0.5$ are NCCs (Vikhlinin et al. 2007). Out of our

1818 observations, we have 21 CCs, 52 WCCs and 1745 NCCs with these thresholds. The distribution of cuspieness values is shown in the bottom panel of figure 2.2.

2.2.2.6. *Comparison of the metrics*

An example of all five cluster properties is shown in figure 2.1 for a CC and a NCC cluster. These specific clusters were chosen as all five metrics are in agreement for their classification. Since the y axis is the same for the plots on the top (CC) and the bottom (NCC), we can clearly see the difference in the profiles. The value of the cooling time is similar at large radius ($\approx R_{500}$) for both clusters. For the CC cluster, the cooling time drops drastically towards the center of the cluster, then becomes flat at the core. The same can be observed for the entropy profile. For the density, the profile is again similar at large radii. The CC density profile then keeps increasing towards the center while the NCC profile remains flat. This is also seen through the value of the cuspieness α , representing the slope of the density profile. Finally, the X-ray image shows the difference in the distribution of the X-ray brightness: for the CC the emission is concentrated at the center, while for the NCC the emission is more uniform. It is important to note that the X-ray images shown have been normalized (the maximum value is 1 for both), so the overall perceived brightness doesn't reflect the total emission of the cluster.

The distribution of the five cluster properties as a function of mass is shown in figure 2.2. The imbalance between the three classes can be clearly seen; while the exact proportion varies with the properties, there are significantly more NCCs than WCCs and CCs. We present a comparison of the five cluster properties in figure 2.3. We observe a strong correlation between the central cooling time and the central electron density (top left panel), as well as between the central cooling time and the central entropy excess. The correlations between the other metrics are not as significant, with a large proportion of data points not falling into the zones of agreement (shaded regions). As the NCCs are more numerous, we have more data points in the red shaded regions.

2.3. Clustering

2.3.1. Methods

We first explore our data with an unsupervised clustering algorithm. We wish to gather the data into 3 groups, and then see if these 3 groups resemble the CC/WCC/NCC classification for any of the five metrics. To do this, we use the non-augmented version of

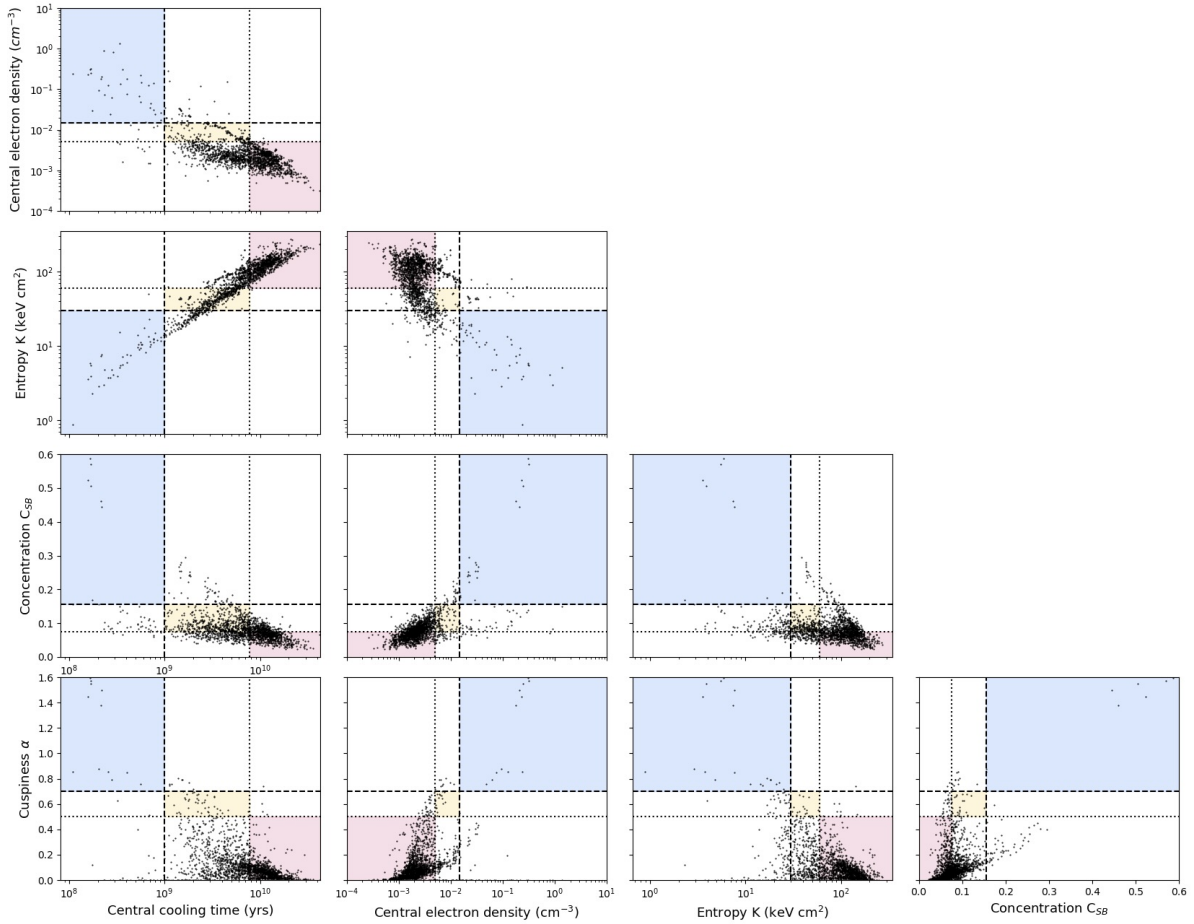


Figure 2.3 – Comparison of the cluster properties for the 1818 observations in our sample. The shaded regions in each plot represent the zones where the clusters would belong to the same class according to both criteria (blue region for CCs, yellow for WCCs, red for NCCs). The dashed line represents the cutoff between CCs and WCCs, whereas the dotted line is the cutoff between WCCs and NCCs.

the dataset, consisting of 1818 images. We start with the normalized images, complete with all the preprocessing steps described in Section 2.2.1.4. However, we want the algorithm to be able to consider the morphology of the cluster as the key factor, rather than just the total amount of light in the image (i.e. the sum of all the pixel values). Therefore, we divide each normalized image by the sum of all pixels in that image. This creates the 256×256 images used as input for the next step. We then perform principal component analysis (PCA) as dimensionality reduction to make the data easier to process for the clustering algorithm. The 256×256 images (65 536 features) serve as input, and are reduced to 1200 components per image. The number of components is chosen to explain 95% of the variance. We then use k-means clustering to gather the data into 3 groups.

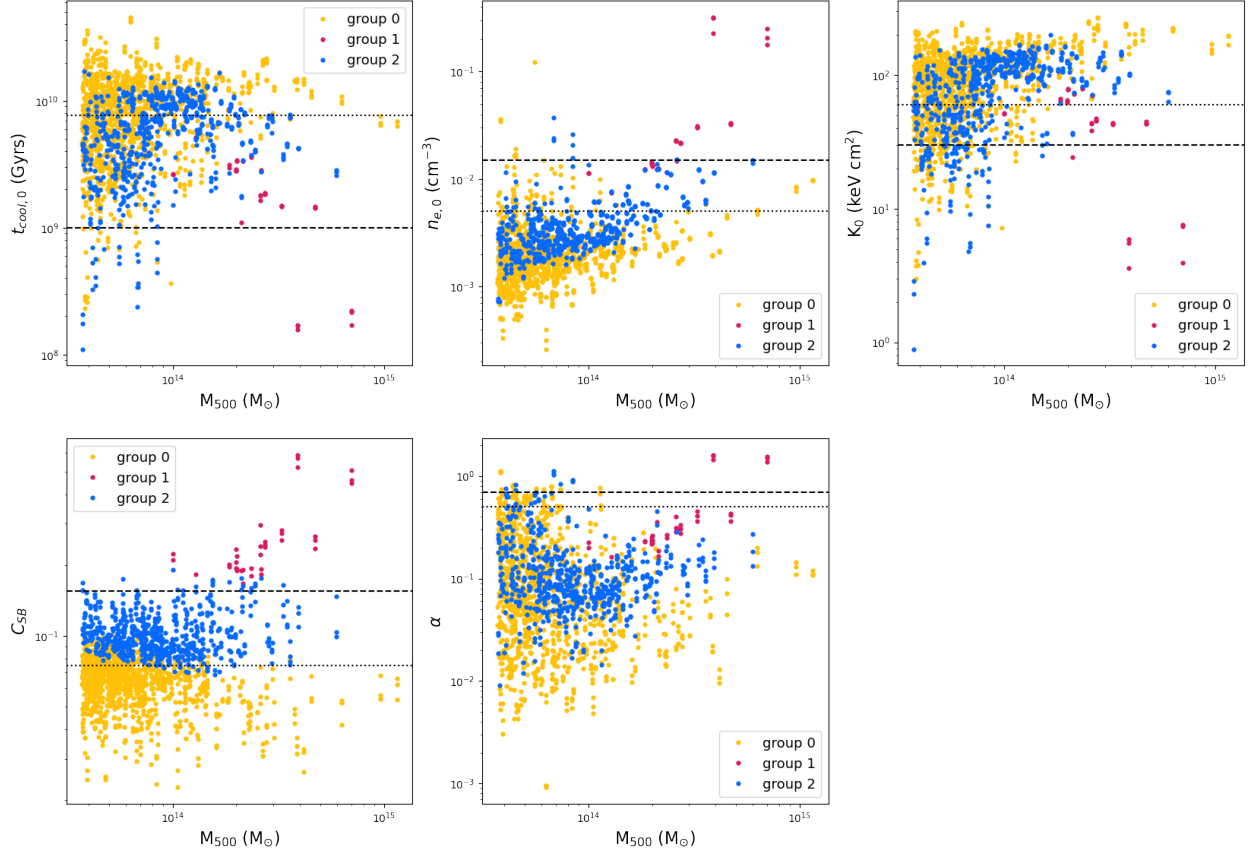


Figure 2.4 – Results of the clustering algorithm. Each galaxy cluster property is plotted as a function of the cluster mass, with clusters belonging to group 0 in yellow, group 1 in red and group 2 in blue. The dashed line represents the cutoff between CCs and WCCs; the dotted line represents the cutoff between WCCs and NCCs.

2.3.2. Results

We present the results of the clustering algorithm in figure 2.4. The galaxy clusters belonging to each of the three groups are shown in different colors. We can very clearly see that there is a strong correlation between the value of the concentration parameter C_{SB} and the group index, with the point of transition between groups corresponding almost exactly to the CC and WCC cutoffs. Since the concentration is the only metric calculated directly from the X-ray image, it is expected to have the strongest correlation out of all the metrics. This result suggests that the PCA and the k-means clustering are able to isolate the shape of the X-ray brightness profile (and therefore the morphology of the cluster) as a feature. The four other metrics do not exhibit such a correlation with the group index.

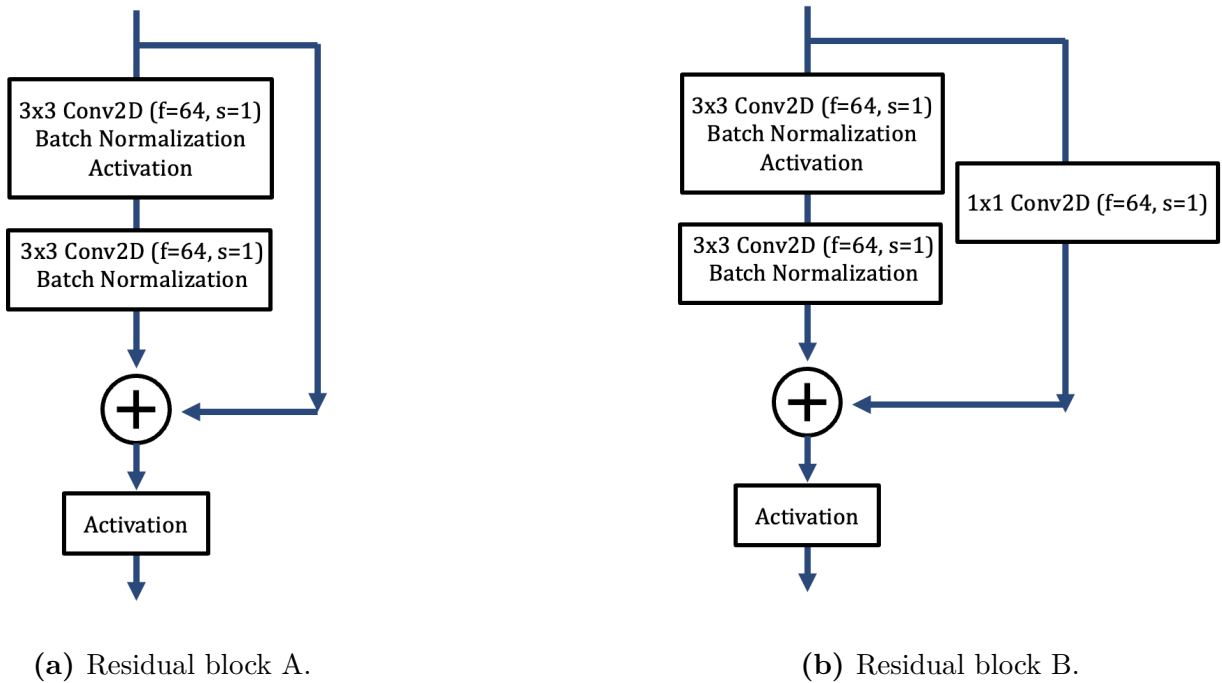


Figure 2.5 – Residual blocks used in the network. Conv2D stands for 2D convolution. f and s refer to the number of filters and the strides.

2.4. Regression

2.4.1. Methods

For the regression task, we have selected a ResNet model as our network architecture. This model utilizes convolutional layers, which are well suited for image analysis as they can extract meaningful spatial features and patterns in a translationally invariant way. ResNet models are made up of residual blocks, which allow the data to pass directly to a superior layer through skip connections. These models have been shown to achieve better results on various computer vision tasks with fewer parameters (He et al. 2015), making them very efficient.

We use two types of residual blocks, as shown in figure 2.5. In one case (block A), the input of the block is simply added two layers later through a skip connection. In the second case (block B), the skip connection involves a 1×1 2D convolution layer. The full architecture is shown in figure 2.6. The input (a 256×256 image) gets passed to a convolution layer, then a pooling layer, then a series of residual blocks A and B. The data is then pooled and flattened before getting passed to a fully-connected network with layers of 128, 32 and 5 neurons each, with 40% dropout between each layer. The final 5-neuron layer is the output of the network, representing the prediction for $\log t_{cool,0}$, $\log n_{e,0}$, \log

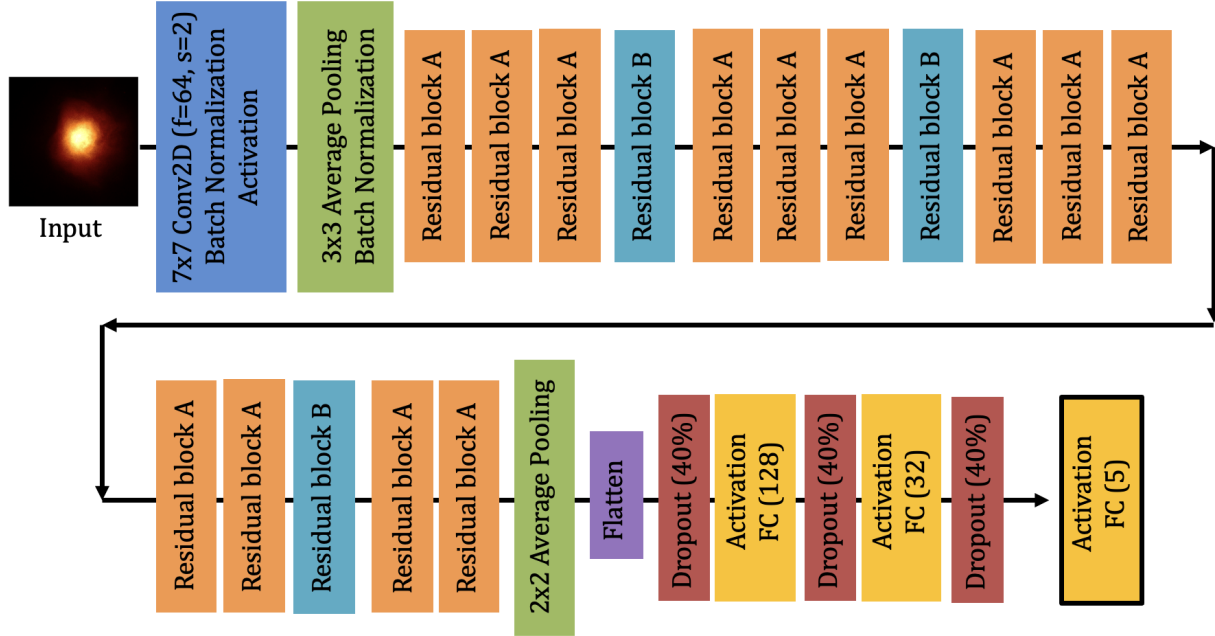


Figure 2.6 – Architecture of our ResNet model. The input is a 256×256 image. The residual blocks A and B are shown in figure 2.5. Conv2D is a 2D convolution, where f and s are the number of filters and the strides. FC is a fully connected layer and the number in brackets corresponds to the number of neurons. The last layer [FC (5)] is the output layer, with the 5 neurons corresponding to the 5 predictions.

K_0 , $\log C_{SB}$ and $\log \alpha$ (here ‘log’ refers to \log_{10}). We chose to predict the log of the values rather than the values themselves because this transformation spreads out the data in a way that amplifies the key differences and results in a range of values of the order of 10^0 or 10^1 , making it easier to work with.

The activation function is a LeakyReLU with a negative slope of $\alpha = 0.1$. We use the ADAM optimizer with an initial learning rate of 10^{-4} , dropping to 10^{-5} at the 400th epoch, and to 10^{-6} at the 700th epoch. The model is trained for 900 epochs. The loss is scaled so that cool core clusters are given more importance, since they’re underrepresented in the sample. For each output we calculate a separate mean squared error loss, scaled by a factor of 1 to 5 depending on how close to a cool core the cluster is. For example, the loss for the smallest $\log t_{cool,0}$ is scaled by 5, and the loss for the largest $t_{cool,0}$ is scaled by 1. The same is done for the other metrics with a factor of 5 corresponding to either the largest or the smallest value (depending on which end represents the CCs). The total loss is then calculated as the mean of the five scaled losses.

Our model takes as input 256×256 images, and is trained to predict a vector ($\log t_{cool,0}$, $\log n_{e,0}$, $\log K_0$, $\log C_{SB}$, $\log \alpha$) from these images. To create training, validation and

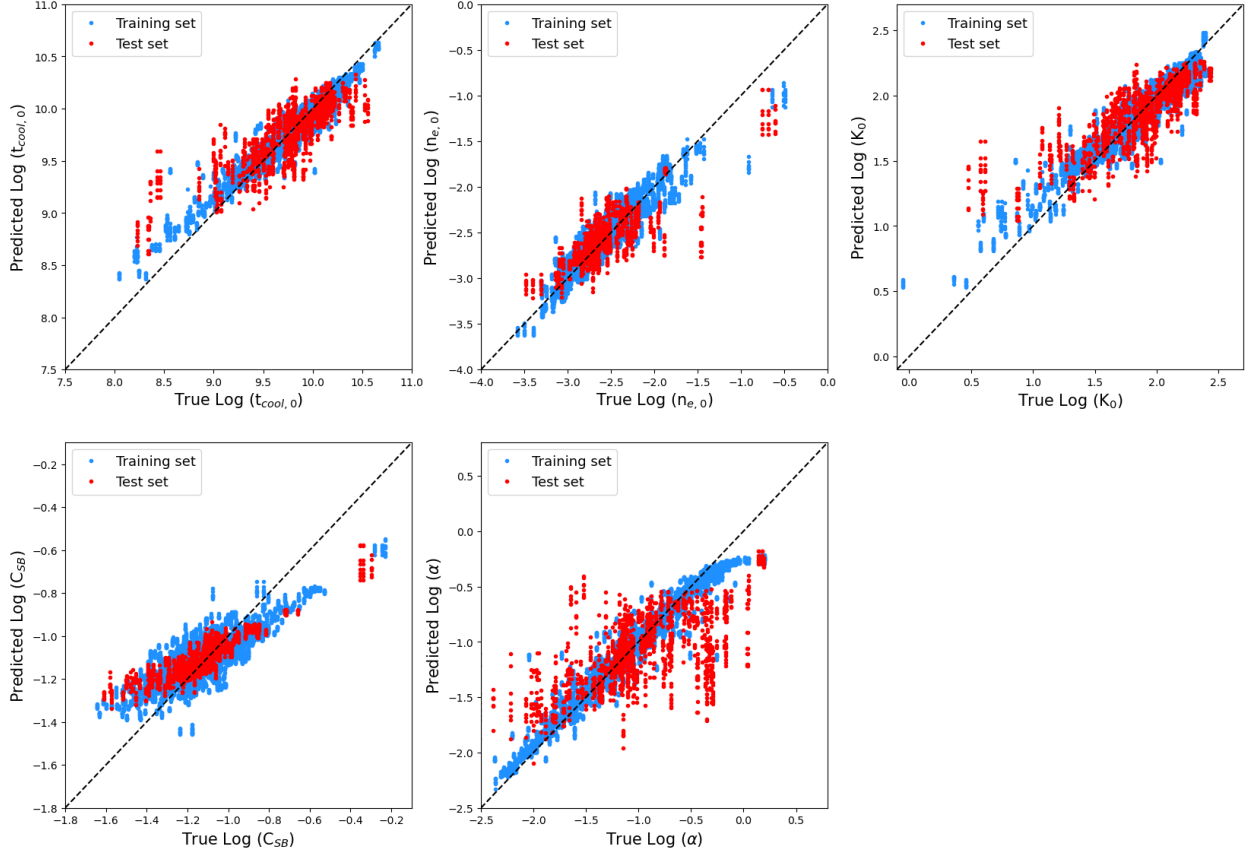


Figure 2.7 – Predicted values as a function of true values for the central cooling time, the central electron density, the central entropy excess, the concentration and the cuspsiness. The training data is shown in blue and the test data is shown in red. The vertical streams show the estimates for each of the 8 views of the same observation.

test sets, the data is divided into folds, making sure that all 24 images of the same cluster always end up in the same fold. We first take 10% of the clusters (61 out of 606) for the test set. To ensure that it is representative, we verify that the CC/WCC/NCC proportion of the entire dataset for every metric is reflected in the test set. We then divide the remaining 545 clusters into 5 folds of 109 clusters each, again making sure that the CC/WCC/NCC proportion for every metric in each fold is about the same as for the entire dataset. We perform a 5-fold cross-validation, with 72% of the data used for training, 18% for validation, and 10% for testing. Each model is trained 5 times with a different fold serving as the validation dataset each time. The hyperparameters that lend the best performance on all five folds are selected and then used to re-train a model using all five folds (90% of the data) as the training dataset. This model is then evaluated on the test set.

2.4.2. Results

The predictions for all five cluster properties are shown in figure 2.7. We observe that the network predictions are biased towards the mean of the distribution. This is commonly observed in neural network predictions, as the networks are trained to minimize a mean error over multiple examples. The plots show that predictions for values corresponding to CCs (low values of $\log t_{cool,0}$, high values of $\log n_{e,0}$, low values of $\log K_0$, high values of $\log C_{SB}$ and high values of $\log \alpha$) are further away from the true values and contain more scatter, which is in part due to the imbalance in the dataset.

For the cooling time (top left panel of figure 2.7), we have a good correspondence between the predictions and true values on the training set as well as the test set. We obtain a mean percentage error of 0.7% on the training set, and 1.8% on the test set.

For the density (second panel of the first row of figure 2.7), we have a good correspondence up to about $\log n_{e,0} = -2.0$, and the network has trouble predicting higher values. We obtain a mean percentage error of 3.9% on the training set and 7.7% on the test set.

For the entropy (top right panel of figure 2.7), we have a larger scatter. The correspondence is good down to about $\log K_0 = 1.3$, and the network struggles predicting lower values. We obtain a mean percentage error of 4.9% on the training set, and 11.4% on the test set.

For the concentration (bottom left panel of figure 2.7), the predictions are very biased towards the mean for both ends of the distribution. This might be because on a log scale, both ends of the concentration distribution have very few clusters. At the center of the distribution, we are able to achieve a good correlation and a relatively small scatter for the test set, when comparing to other metrics. We obtain a mean percentage error of 6.8% on the training set and 8.0% on the test set.

For the cuspidity predictions (second panel in the second row of figure 2.7), we are not able to achieve good performance on the test set. The correlation is good for the training set, but the scatter on the test set is very large, and there doesn't seem to be a clear correlation between predictions and true values for the red data points. We obtain a mean percentage error of 14.3% on the training set and 96.6% on the test set. This is a clear case of overfitting; better performance might be achieved through stricter regularization techniques.

	CC	WCC	NCC	Average
Central cooling time $t_{\text{cool},0}$	0.70	0.81	0.83	0.78
Central electron density $n_{e,0}$	0.75	0.61	0.68	0.68
Central entropy excess K_0	0.74	0.67	0.80	0.74
Concentration C_{SB}	0.75	0.86	0.87	0.83
Cuspiness α	0.50	0.49	0.58	0.52

Table 2.1 – Balanced accuracy (BAcc) calculated on the test set for each class (CC, WCC, NCC) as well as the average, for each classification metric. The BAcc is an accuracy measure defined between 0 and 1, with 1 being the most accurate.

From the results in figure 2.7, we can convert the continuous predictions into class predictions, based on the cutoffs for CC/WCC/NCC clusters for each metric. We can then construct confusion matrices to visualize the performance of this classification. The confusion matrices for each metric are presented in figures 2.8 through 2.12.

For every matrix, we can calculate the balanced accuracy (BAcc) for each class, defined as:

$$\text{BAcc} = \frac{1}{2} \left(\frac{tp}{tp + fn} + \frac{tn}{tn + fp} \right) \quad (2.4.1)$$

where tp , tn , fp and fn are true positives, true negatives, false positives and false negatives. These numbers are defined as numbers of samples, and are not normalized in the calculation of the BAcc. The balanced accuracy for each metric is shown in table 2.1. Based on these results, the best-performing metrics are the central cooling time and the concentration.

We can compare the cooling time balanced accuracy with the results from S20. They tackle a classification problem, predicting the galaxy cluster category (CC/WCC/NCC) for mock Chandra images obtained from the $z = 0$ snapshot of the TNG300 simulations. Their sample contains 318 clusters, and they use the 3D central cooling time to create class labels. They get a balanced accuracy of 0.92 for CCs, 0.81 for WCCs and 0.83 for NCCs. 10% of clusters in their dataset are CCs, 61% are WCCs and 29% are NCCs. Their dataset is therefore more balanced than ours, which tends to improve performance. Finally, their network architecture is similar to ours, but their network itself is pre-trained for a classification task on the ImageNet dataset.

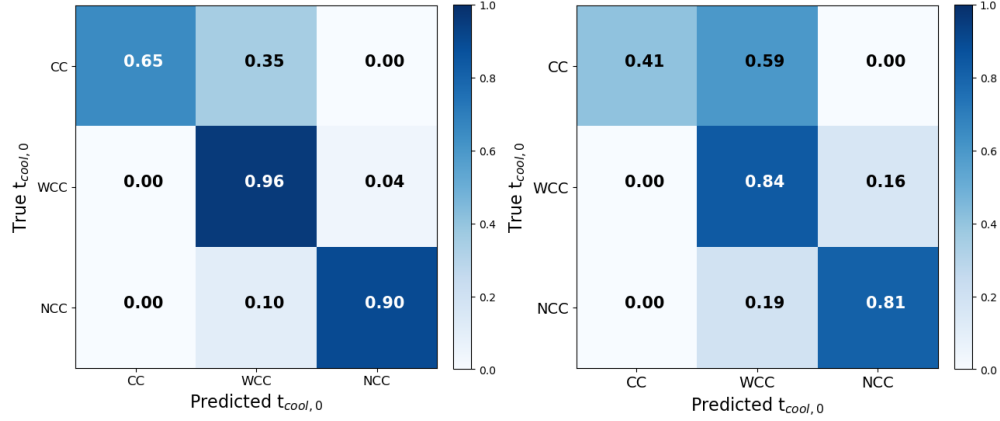


Figure 2.8 – Normalized confusion matrices for the central cooling time predictions, for the training set (left) as well as the test set (right). The matrices are normalized along the true values.

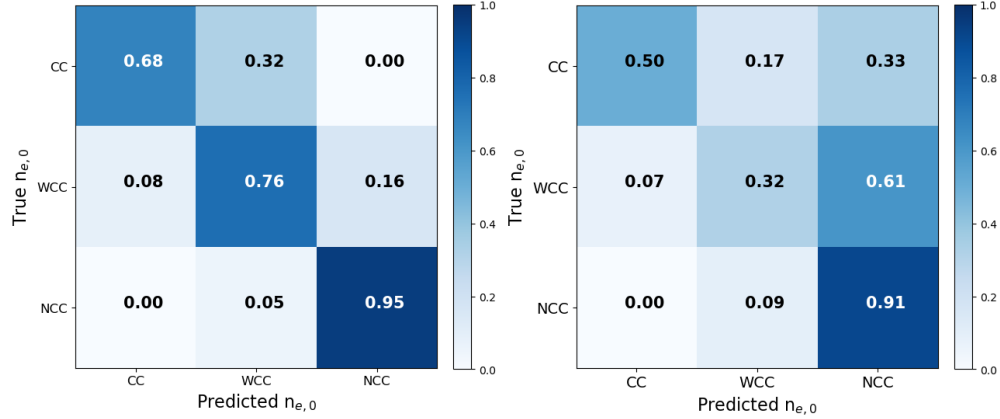


Figure 2.9 – Normalized confusion matrices for the central electron density predictions, for the training set (left) as well as the test set (right). The matrices are normalized along the true values.

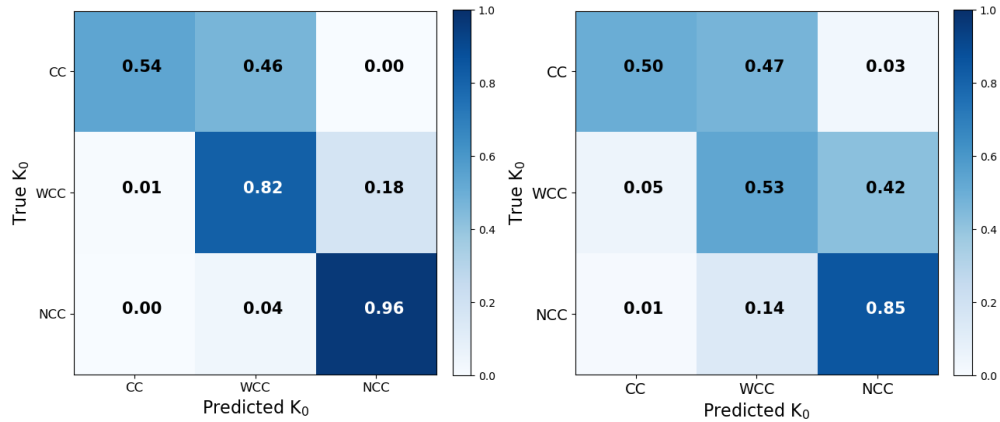


Figure 2.10 – Normalized confusion matrices for the central entropy excess, for the training set (left) as well as the test set (right). The matrices are normalized along the true values.

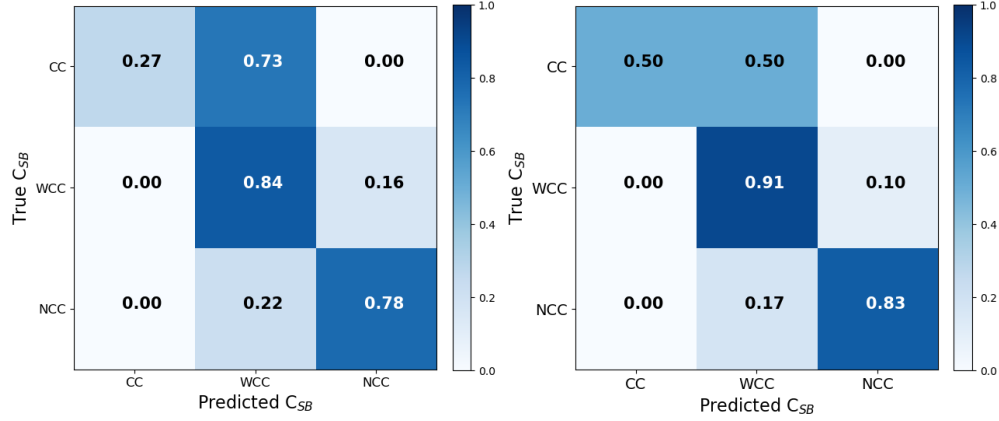


Figure 2.11 – Normalized confusion matrices for the concentration predictions, for the training set (left) as well as the test set (right). The matrices are normalized along the true values.

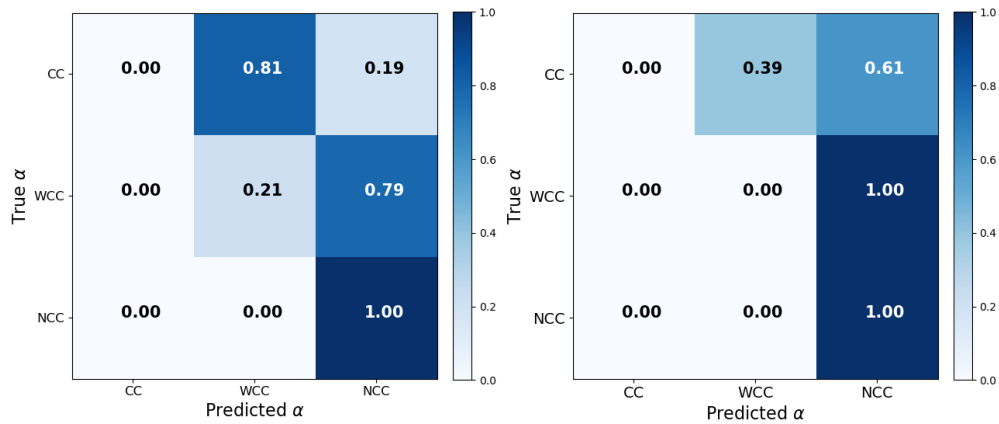


Figure 2.12 – Normalized confusion matrices for the cuspidity predictions, for the training set (left) as well as the test set (right). The matrices are normalized along the true values.

2.5. Discussion

2.5.1. Simulation-based inference

Once the network is trained, we can use it to get predictions on new data. Furthermore, we can obtain a probability distribution over the possible values using the formalism of simulation-based inference (SBI, e.g. Cranmer et al. 2020), also referred to as likelihood-free inference (LFI). This technique allows us to obtain a probability distribution over possible values, rather than just a single point prediction. In this context, we have a model with parameters (latent variables) θ , generating data x . We would like to perform inference on θ given observations $\{x\}_i$ to get $p(\theta|\{x\}_i)$. This can usually be done with Bayes' theorem:

$$p(\theta|x) \propto \frac{p(x|\theta)p(\theta)}{p(x)},$$

where x is the observed data, $p(\theta|x)$ is the posterior distribution of the model parameters given x , $p(x|\theta)$ is the likelihood of the data, $p(\theta)$ is the prior on the model parameters. The likelihood $p(x|\theta)$, however, is often unknown. SBI is useful in those cases, as it does not require an explicit expression for the likelihood. Instead, it only relies on having a model to generate $\{\theta, x\}$ pairs, that are used to fit a continuous function $f = \hat{p}(\theta|x)$ that approximates the distribution $p(\theta|x)$ for those pairs. This is done by first using the dataset $\{\theta, x\}_i$ as training data to model a joint distribution of the parameters and the data $p(x, \theta)$. Then, the posterior for new data examples can be obtained by evaluating the joint distribution at the observed data, yielding $\hat{p}(\theta|x)$. The function f can be represented as a neural network, training on $\{\theta, x\}_i$ to maximize $\sum_i \log p(\theta|x)$. One can use a gaussian model or a normalizing flow to ensure that f is a good fit for a probability distribution, i.e. that it is smooth and normalized. Once f is trained, we can sample from $\hat{p}(\theta|x)$ to build a posterior distribution for a given x .

When dealing with a regression network that predicts point estimates, SBI can be used to obtain posterior distributions for the network predictions (e.g. Legin et al. 2023). In this context, we wish to obtain a probability distribution over possible values of y_{true} , for a given value of y_{pred} . We use the `sbi` python package (Tejero-Cantero et al. 2020) to perform the analysis, where the latent variable θ is y_{true} , x is y_{pred} , and the learned posterior $\hat{p}(\theta|x)$ is given by $\hat{p}(y_{\text{true}}|y_{\text{pred}})$. We use the entire training dataset (all five folds) as observations for training the function $f = \hat{p}(y_{\text{true}}|y_{\text{pred}})$. We train five separate functions for the five cluster properties.

In figure 2.13, we plot a few examples of posterior distributions (in red) for $y = \log(t_{\text{cool}}, 0)$. We observe that the posterior probability distributions exhibit higher precision at the values where there is more training data and where the predictions are closer to the ground truth values, as expected. On the lower end of the distribution, the posteriors are more flat (i.e. the predictions have a higher uncertainty) as the training data points in this region have larger errors, when comparing the predictions to the true values.

Once the posterior distribution is calculated, we can examine the probability distributions for all five metrics for the same example. This is done in figure 2.14 for a cool core cluster and in figure 2.15 for a non-cool core cluster. The clusters used in these figures are drawn from the test set. When comparing figures 2.14 and 2.15, we clearly see that the posterior probability distributions have higher precision for the NCC example. This

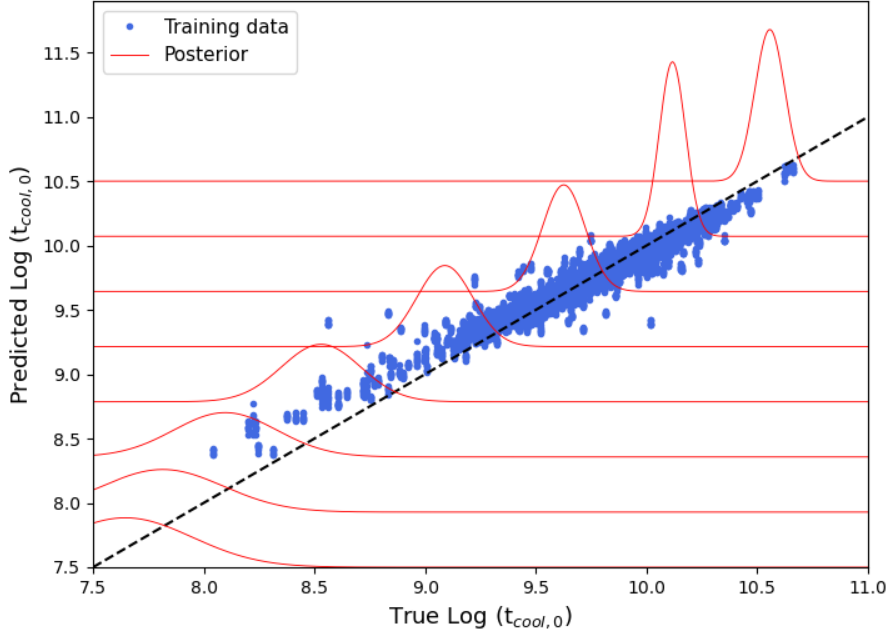


Figure 2.13 – SBI results for the central cooling time predictions. The network predictions for the training data are plotted in blue as a function of the true labels. The black dashed line represents the true values. In red, we overplot the posterior distributions for a few predicted values of the central cooling time.

is expected, as there is more data for NCC clusters and therefore more information to build more constraining posterior distributions. We also see that for the NCC example, the true values generally follow the peak of the posterior distribution. For the density and the cuspiness, the true values have slightly lower probability. For the CC cluster, the posterior distributions are less precise and the true values (in pink) often have low probability. The probability distribution for a single example can also be used to determine what are the chances of that observation belonging to a cluster category (CC/WCC/NCC). The total area under the curve sums to one, and therefore the area under the curve that is within the range for the specified category on the x-axis indicates the probability that the cluster belongs to that category.

2.5.2. Fraction of cool cores in IllustrisTNG

We now proceed to compare our CC fractions to other works that have been based on the IllustrisTNG simulations, as well as on observational data. Our entire sample contains 606 clusters (1818 observations) with $M_{500} > 10^{13.57} M_{\odot}$. Also drawn from IllustrisTNG, the

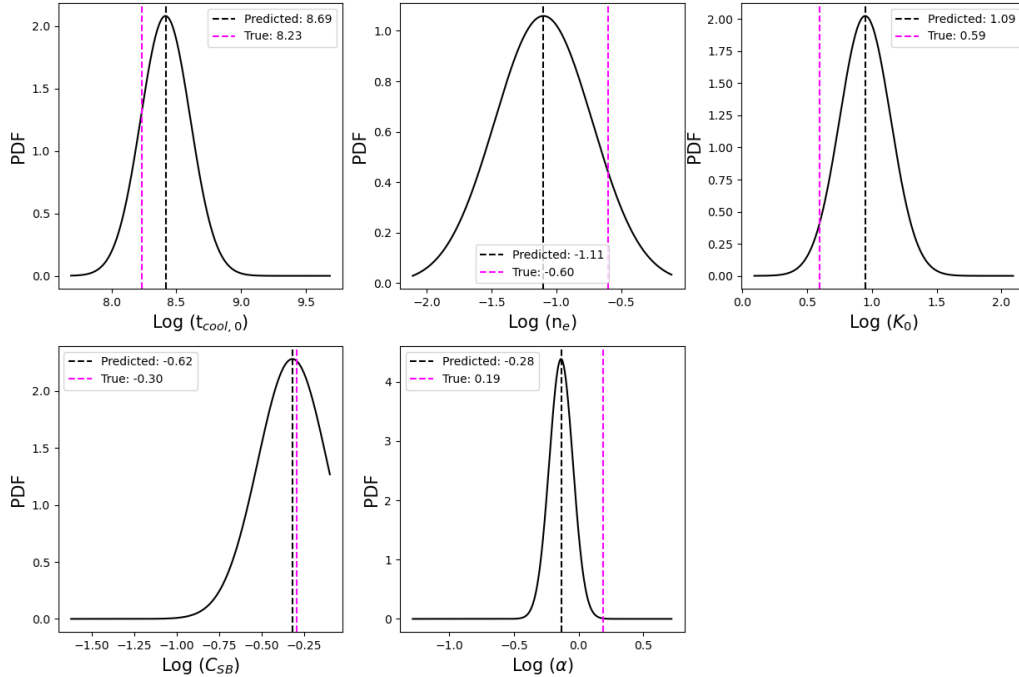


Figure 2.14 – Example of posteriors for the five metrics, for a CC cluster. The example cluster is drawn from the test set. In each plot, the pink dashed line represents the true label and the black dashed line represents the network prediction. The solid black line represents the posterior, i.e. the probability distribution over true values for the given predicted value.

sample in S20 consists of 318 clusters with $M_{500} > 10^{13.75} M_{\odot}$, whereas the sample in B18 contains 370 clusters with $M_{500} > 10^{13.75} M_{\odot}$. Since we are using a different mass cutoff, we present the CC fractions in our entire sample as well as in the subset of 363 clusters with $M_{500} > 10^{13.75} M_{\odot}$ for a closer comparison.

When using the central cooling time as the metric, we obtain a CC fraction of 3% for the entire sample, and 2% for the higher mass subset. S20 obtain a CC fraction of 10%, while B18 obtain a CC fraction of 12% with the same definition ($t_{\text{cool},0} < 1$ Gyr). We compare with observations of the HIGHEST X-ray FLUX Galaxy Cluster Sample (HIFLUGCS) from the ROSAT survey, where 44% of the clusters are CCs (Mittal et al. 2009). Hence, when using the central cooling time as the metric, the CC fraction reported in our work is lower than in S20 and B18, as well as in observational data.

When using the central electron density as the metric, we obtain a CC fraction of 4% for the entire sample, and 3% for the higher mass subset. B18 report a CC fraction

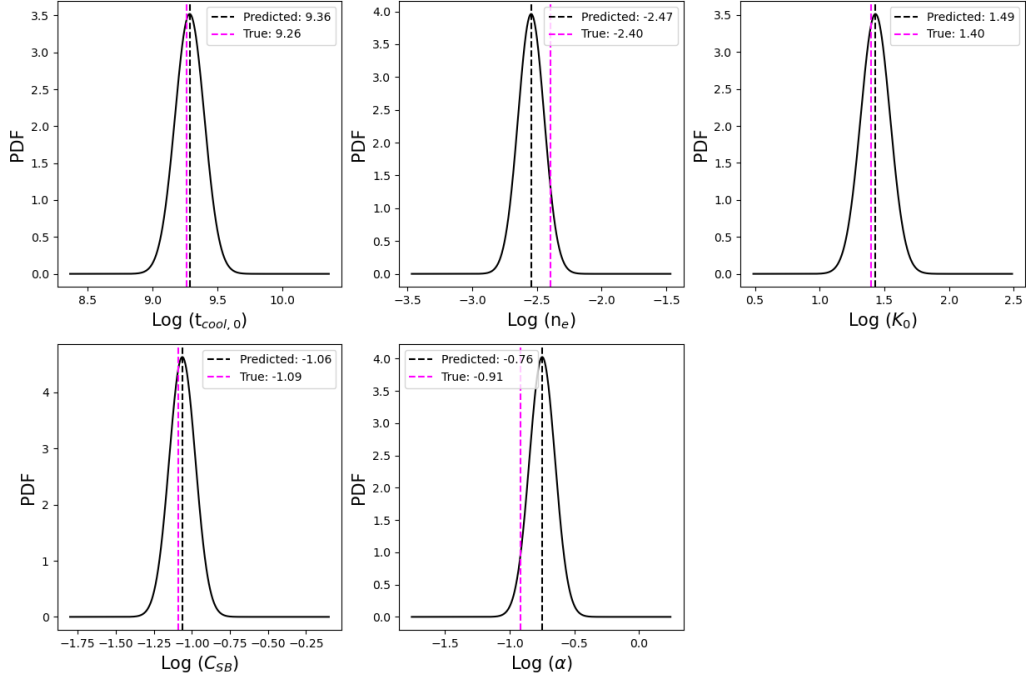


Figure 2.15 – Same as 2.14 but for a NCC cluster. This example cluster is also drawn from the test set.

of 14% with this metric. We compare with *Chandra* observations from the *Planck* Early Sunyaev-Zel’dovich (ESZ) sample (Andrade-Santos et al. 2017) and a flux X-ray limited sample (Andrade-Santos et al. 2017). Selecting clusters with $z < 0.25$, the SZ sample has a CC fraction of 52% and the X-ray sample has a CC fraction of 37%. Hence, when using the central electron density as the metric, the CC fraction reported in our work is lower than in B18 as well as in observational data.

When using the central entropy excess as the metric, we obtain a CC fraction of 10% for the entire sample, and 7% for the higher mass subset. B18 report a CC fraction of 18% with this metric. We compare with a sample of 239 clusters taken from the *Chandra* archive analyzed in Cavagnolo et al. (2009), which yields a CC fraction of 46% for the same definition ($K_0 < 30 \text{ keV cm}^2$). Therefore, when using the central entropy excess as the metric, the CC fraction reported in our work is lower than in B18 as well as in observational data.

When using the concentration parameter as the metric, we obtain a CC fraction of 3% for the entire sample, and 5% for the higher mass subset. B18 report a CC fraction of 1%

with this metric. We compare with the SZ and X-ray samples from Andrade-Santos et al. (2017), selecting one again clusters with $z < 0.25$. The SZ sample has a CC fraction of 34% and the X-ray sample has a CC fraction of 12%. Thus, when using the concentration as the metric, the CC fraction reported in our work is higher than in B18 but lower than in observational data.

When using the cuspsiness parameter as the metric, we obtain a CC fraction of 1% for the entire sample as well as for the higher mass subset. B18 obtain a CC fraction of 21% with this metric. Selecting the clusters with $z < 0.25$ from Andrade-Santos et al. (2017), the SZ sample has a CC fraction of 47% and the X-ray sample has a CC fraction of 28%. Therefore, when using the cuspsiness as the metric, the CC fraction reported in our work is lower than in B18 as well as in observational data.

The CC fractions in this work are overall lower than the ones reported in S20 and B18. While our study, as well as B18 and S20, uses the $z=0$ snapshot of the TNG300 simulation, there are significant differences in the methods, which could explain the discrepancy. First, the methods used to locate the center are different; we use an iterative gaussian filter on the 2D observations to find the X-ray peak while B18 use the cluster’s potential minimum. It is not detailed in S20 how the center is located. Also, B18 and S20 both define the central cooling time as the average within a 3D region with $r < 0.012R_{500}$, while we examine the 2D cooling time profiles and perform a fit for $t_{\text{cool},0}$, in an effort to mimic the procedure used with observations. We also perform a fit for the central electron density based on the 2D profiles, while B18 use the average value within $r < 0.012R_{500}$. For the entropy, our method is the same as in B18; the only difference being them using 3D entropy profiles while we’re working with 2D profiles. For the concentration, the X-ray luminosity is not computed the same way in our work and in B18. We use mock *Chandra* images produced with `pyXSIM` and `SOXS`, whereas B18 first produce mock X-ray spectra for each gas cell and then sum the spectra with an aperture (40 kpc or 400 kpc for C_{SB}) to get the X-ray luminosities. Finally, the cuspsiness is calculated slightly differently; B18 get the cuspsiness from 3D density profiles, while we are once again using 2D profiles. All of these differences in our analysis lead to different values for the gas properties, and therefore different CC fractions.

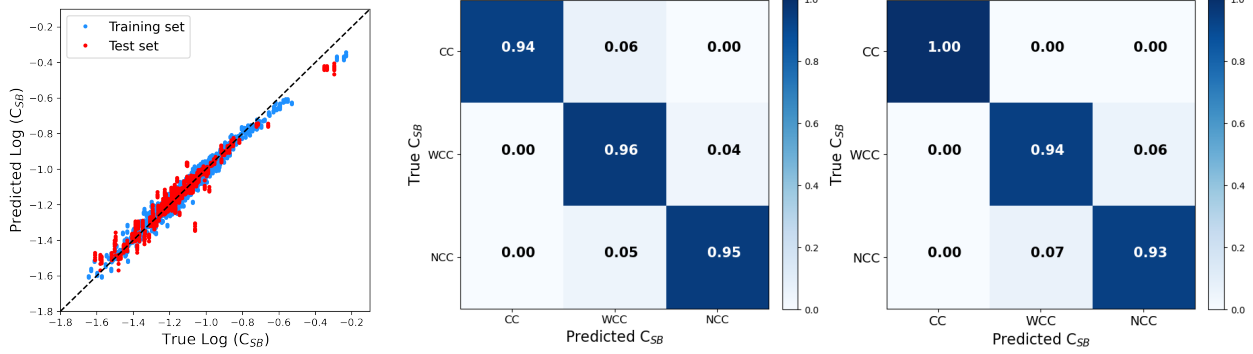
In general, the CC fractions in the $z=0$ snapshot of the TNG300 simulations are significantly lower than in observational data. This might be due to the fact that realistic AGN feedback processes are needed to produce the observed CC fractions, but are challenging to simulate (Morganti 2017). The physical processes happen at very different scales, and large cluster simulations do not have high enough resolution to accurately model the small scale

physics. Therefore, simplified models are employed on larger resolved scales to measure black hole accretion rates and inject the corresponding feedback energy into the surrounding gas (Weinberger et al. 2018). Also, AGN feedback exists in two modes, depending on the value of the accretion rate: radio mode at a low accretion rate, and quasar mode at a high accretion rate. The physics transition between the two is difficult to implement in simulations as it is poorly understood (e.g. Weinberger et al. 2017).

2.5.3. Metric comparison

Based on our analysis, we conclude that the metrics most suited for machine learning analysis of X-ray images of galaxy clusters are the central cooling time and the concentration. The cooling time performs the best out of all the metrics when it comes to the regression task. It has the smallest percentage error on the test set (1.8%) as well as the training set (0.7%), and the second highest balanced accuracy (0.78). The concentration has the highest balanced accuracy (0.83) out of all the metrics, and it has the strongest correlation with the group index obtained with the clustering algorithm. Since the concentration is the only metric calculated directly from the X-ray image, we expect both models (the clustering algorithm and the ResNet model for regression) to perform fairly well when using the concentration as the metric. It is therefore surprising that the ResNet has trouble predicting the C_{SB} values. One reason for this might be that the information needed to predict $t_{cool,0}$, $n_{e,0}$, K_0 and α is significantly different from the information needed to predict the concentration. This would lead the convolutional layers to extract features that are not well-suited to accurately predict the concentration. This can be mitigated by training a separate network for the concentration or by adding more complexity to the existing network. Another path worth exploring would be to replace the input image with its PCA representation, as done for the k-means clustering.

In an effort to improve performance for the concentration, we re-train the network only predicting the concentration values. The network is the same as the one presented in figure 2.6, with the 5 neurons in the output layer being replaced by only one. The results are presented in figure 2.16. The performance is significantly improved; the correspondence between predicted and true values in the left panel of figure 2.16 is drastically better than the one in figure 2.7. The same is true for the values in the confusion matrices in **b)** of figure 2.16: they represent a strong improvement compared to figure 2.11. With the retrained network, we obtain a mean percentage error of 1.6% on the training set, and 2.9% on the test set. We also obtain a balanced accuracy of 1.00 for CCs, 0.94 for WCCs and 0.94 for NCCs resulting in an average of 0.96.



(a) Predicted values as a function of the true values for the concentration
 (b) Normalized confusion matrices for the concentration for the training set (left) and the test set (right)

Figure 2.16 – Results for the concentration with the retrained network.

For the same reason, we also re-train a separate network for the four other properties (the central cooling time, the central electron density, the central entropy excess and the cuspiness) individually. This does not result in significant change (up to 2.8%) for the mean percentage error for these four properties. The balanced accuracy is not greatly affected (up to 0.03) for the central cooling time, the central electron density, the central entropy excess or the cuspiness. The most significant change remains to the concentration, and with the re-trained network the concentration is the metric that is most well-suited for a machine learning analysis when working with X-ray images.

On the other hand, the metric that performs the worst is the cuspiness α . Visually, the predicted values on the test set for this metric do not show any correlation with the true values. It has the highest percentage error on both the training set (14.3%) and the test set (96.6%), and it has the lowest balanced accuracy (0.52). The cuspiness is calculated from the fit obtained on the electron density profile; it is the only metric that is derived in such a secondary way. Also, the cuspiness is a very local measure of the density profile; it is calculated at exactly $0.04R_{500}$. Moreover, when we look at the distribution of the cuspiness values, we only have 21 cool cores out of 1818 observations, making the cuspiness the least balanced metric. The significant imbalance in the dataset, as well as the extremely local calculation for this metric, could explain the poor performance.

2.5.4. Implications for future surveys

There are numerous steps that can be taken in order to improve the performance of the model and make it applicable to observational data. As discussed earlier, the network performance can be improved by training on a larger dataset, by adding complexity to

the network, or by making the input more informative with the addition of spectral X-ray data. A larger training dataset could be sourced from two upcoming projects: TNG-Cluster and MilleniumTNG. TNG-Cluster is a spin-off series of simulations by the IllustrisTNG team, specifically designed for the study of massive galaxy clusters. It is expected to have a simulation box size of 1000 Mpc, compared to the current largest simulation box size of 300 Mpc. Millenium-TNG aims to provide better representation of dark matter structures, while having a simulation box size of 500 Mpc. Both of these simulations are expected to be made publically available in 2024.

A promising path to make the model applicable to observational data is to use transfer learning, which consists of using a pre-trained model on a new problem. In our case, we can use the model trained on simulated data as a starting point and fine-tune it by retraining it on a small dataset of observational data. This has been successfully done in astrophysics (e.g. Ghosh et al. 2022), and could greatly improve the performance on the entire dataset of observational data.

A large dataset of observational X-ray data will become available in the next few years, as the X-ray instrument eROSITA is expected to detect at least 100 000 massive galaxy clusters (Merloni et al. 2012, Pillepich et al. 2012). To analyze these large datasets, galaxy cluster experts will need assistance from machine learning tools, such as the model presented in this work. Our model is a fast and efficient way to characterize the ICM in galaxy clusters. While the model performance is not on-par with expert classification, we have produced a useful tool that provides uncertainty measures. This is a viable path to assist experts in their analysis by isolating the samples that show large uncertainties, helping them target the more complex galaxy clusters and allowing us to make the most out of the upcoming data.

2.6. Conclusion

We have presented a deep learning approach for galaxy cluster characterization, predicting five cluster properties from mock X-ray images. The training dataset is built from 606 clusters with $M_{500} > 10^{13.57} M_{\odot}$, resulting in 1818 observations and augmented to 14 544 images. The network is trained to learn cluster properties from a downsampled 256×256 image. Our method shows a mean percentage error of 1.8% for the central cooling time, and an average balanced accuracy of 0.83 for the concentration. We have performed an SBI analysis to obtain posterior probability distributions for the network predictions. This analysis shows that the predictions for non-cool core clusters have lower uncertainty as well as lower error. Based on the network predictions as well as

the probability analysis, we conclude that the cooling time and the concentration are the metrics that lend themselves the best to machine learning characterization. We also note that the concentration parameter correlates very strongly with the groups resulting from the unsupervised clustering algorithm. This indicates that the classification based on the concentration corresponds to a somewhat natural separation of galaxy cluster X-ray images.

Overall, we have produced a fast and statistically sound model that can be used to accelerate the analysis of large datasets. An important limiting factor in our work is the size and the range of our dataset. This can be improved by including data from other simulations or by pre-training our network on a separate larger dataset of astrophysical images. We could also explore a wider range of neural network models that are better able to capture the discriminative features. In a follow-up study, we would like to apply and train the network on real observations, as well as include spectral information as input, since it is already available with X-ray data. We expect this to significantly improve the performance of the model, making it a powerful tool for extracting key information from galaxy cluster observations.

Acknowledgements

The authors would like to thank the IllustrisTNG team. The IllustrisTNG simulations were undertaken with compute time awarded by the Gauss Centre for Supercomputing (GCS) under GCS Large-Scale Projects GCS-ILLU and GCS-DWAR on the GCS share of the supercomputer Hazel Hen at the High Performance Computing Center Stuttgart (HLRS), as well as on the machines of the Max Planck Computing and Data Facility (MPCDF) in Garching, Germany.

This research was enabled in part by support provided by Calcul Québec (calculquebec.ca) and the Digital Research Alliance of Canada (alliancecan.ca).

This research has made use of software provided by the *Chandra* X-ray Center (CXC) in the application packages CIAO.

M.S. acknowledges financial support from the Fonds de recherche du Québec – Nature et technologies (FRQNT), from IVADO, from the Center for Research in Astrophysics of Quebec (CRAQ), as well as from the physics department of Université de Montréal.

Software

This research made use of `Python` (Van Rossum & Drake Jr 1995), `NumPy` (Harris et al. 2020), `SciPy` (Virtanen et al. 2020), `AstroPy` (Astropy Collaboration et al. 2013, 2018, 2022), `Matplotlib` (Hunter 2007), `TensorFlow` (Abadi et al. 2015), `Keras` (Chollet et al. 2015), `emcee` (Foreman-Mackey et al. 2013), `pyXSIM` (ZuHone & Hallman 2016), `SOXS` (ZuHone et al. 2023), `CIAO` (Fruscione et al. 2006), `sbi` (Tejero-Cantero et al. 2020).

Bibliography

- Abadi, M., Agarwal, A., Barham, P., et al. 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, software available from tensorflow.org
- Andrade-Santos, F., Jones, C., Forman, W. R., et al. 2017, , 843, 76
- Angora, G., Rosati, P., Brescia, M., et al. 2020, , 643, A177
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, , 558, A33
- Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, , 156, 123
- Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, *apj*, 935, 167
- Barnes, D. J., Vogelsberger, M., Kannan, R., et al. 2018, , 481, 1809
- Bauer, F. E., Fabian, A. C., Sanders, J. S., Allen, S. W., & Johnstone, R. M. 2005, , 359, 1481
- Bharadwaj, V., Reiprich, T. H., Schellenberger, G., et al. 2014, , 572, A46
- Biffi, V., Dolag, K., & Böhringer, H. 2013, , 428, 1395
- Biffi, V., Dolag, K., Böhringer, H., & Lemson, G. 2012, , 420, 3545
- Cao, S., Biesiada, M., Zheng, X., & Zhu, Z.-H. 2016, , 457, 281
- Cavagnolo, K. W., Donahue, M., Voit, G. M., & Sun, M. 2009, *The Astrophysical Journal Supplement Series*, 182, 12
- Cavaliere, A. & Fusco-Femiano, R. 1976, , 49, 137
- Cavaliere, A. & Fusco-Femiano, R. 1978, , 70, 677
- Chollet, F. et al. 2015, Keras, <https://keras.io>
- Cranmer, K., Brehmer, J., & Louppe, G. 2020, *Proceedings of the National Academy of Science*, 117, 30055
- Cui, W., Power, C., Biffi, V., et al. 2016, , 456, 2566
- de Andres, D., Yepes, G., Sembolini, F., et al. 2023, , 518, 111
- de Andres, D., Cui, W., Ruppin, F., et al. 2022, in *European Physical Journal Web of Conferences*, Vol. 257, *mm Universe @ NIKA2 - Observing the mm Universe with the NIKA2 Camera*, 00013
- Fabian, A. C. 2012, , 50, 455
- Ferragamo, A., de Andres, D., Sbriglio, A., et al. 2023, , 520, 4000
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, , 125, 306
- Foster, A. R., Ji, L., Smith, R. K., & Brickhouse, N. S. 2012, , 756, 128
- Fruscione, A., McDowell, J. C., Allen, G. E., et al. 2006, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 6270, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, ed. D. R. Silva & R. E. Doxsey, 62701V
- Ghosh, A., Urry, C. M., Mishra, A., et al. 2022, arXiv e-prints, arXiv:2212.00051
- Grego, L., Carlstrom, J. E., Reese, E. D., et al. 2001, , 552, 2

Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Nature*, 585, 357

He, K., Zhang, X., Ren, S., & Sun, J. 2015, arXiv e-prints, arXiv:1512.03385

Hlavacek-Larrondo, J., Li, Y., & Churazov, E. 2022, in *Handbook of X-ray and Gamma-ray Astrophysics*, 5

Hudson, D. S., Mittal, R., Reiprich, T. H., et al. 2010, , 513, A37

Hunter, J. D. 2007, *Computing in Science & Engineering*, 9, 90

Krippendorff, S., Baron Perez, N., Bulbul, E., et al. 2023, arXiv e-prints, arXiv:2305.00016

Lazzati, D. & Chincarini, G. 1998, , 339, 52

Legin, R., Hezaveh, Y., Perreault-Levasseur, L., & Wandelt, B. 2023, , 943, 4

Loewenstein, M. 2004, in *Origin and Evolution of the Elements*, ed. A. McWilliam & M. Rauch, 422

Mantz, A. B., Allen, S. W., Morris, R. G., et al. 2015, , 449, 199

Marinacci, F., Vogelsberger, M., Pakmor, R., et al. 2018, , 480, 5113

McDonald, M., Gaspari, M., McNamara, B. R., & Tremblay, G. R. 2018, , 858, 45

McDonald, M., Benson, B. A., Vikhlinin, A., et al. 2013, , 774, 23

McDonald, M., Benson, B. A., Vikhlinin, A., et al. 2014, , 794, 67

McDonald, M., Allen, S. W., Hlavacek-Larrondo, J., et al. 2019, , 870, 85

McNamara, B. R. & Nulsen, P. E. J. 2007, , 45, 117

Merloni, A., Predehl, P., Becker, W., et al. 2012, arXiv e-prints, arXiv:1209.3114

Million, E. & Allen, S. 2009, *Monthly Notices of the Royal Astronomical Society*, 399, 1307

Mittal, R., Hudson, D. S., Reiprich, T. H., & Clarke, T. 2009, , 501, 835

Mohr, J. J., Mathiesen, B., & Evrard, A. E. 1999, , 517, 627

Morganti, R. 2017, *Frontiers in Astronomy and Space Sciences*, 4, 42

Naiman, J. P., Pillepich, A., Springel, V., et al. 2018, , 477, 1206

Nelson, D., Pillepich, A., Springel, V., et al. 2018, , 475, 624

Nelson, D., Springel, V., Pillepich, A., et al. 2019, *Computational Astrophysics and Cosmology*, 6, 2

Ntampaka, M., ZuHone, J., Eisenstein, D., et al. 2019, , 876, 82

Pillepich, A., Porciani, C., & Reiprich, T. H. 2012, , 422, 44

Pillepich, A., Nelson, D., Hernquist, L., et al. 2018, , 475, 648

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, , 594, A13

Reese, E. D., Carlstrom, J. E., Joy, M., et al. 2002, , 581, 53

Santos, J. S., Tozzi, P., Rosati, P., & Böhringer, H. 2010, , 521, A64

Springel, V. 2010, , 401, 791

Springel, V., Pakmor, R., Pillepich, A., et al. 2018, , 475, 676

Su, Y., Zhang, Y., Liang, G., et al. 2020, , 498, 5620

Sun, M. 2009, , 704, 1586

Tarsitano, F., Bruderer, C., Schawinski, K., & Hartley, W. G. 2022, , 511, 3330

Tejero-Cantero, A., Boelts, J., Deistler, M., et al. 2020, *Journal of Open Source Software*, 5, 2505

Umetsu, K., Birkinshaw, M., Liu, G.-C., et al. 2009, , 694, 1643

Van Rossum, G. & Drake Jr, F. L. 1995, *Python reference manual* (Centrum voor Wiskunde en Informatica Amsterdam)

Vikhlinin, A., Kravtsov, A., Forman, W., et al. 2006, , 640, 691

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nature Methods*, 17, 261

Weinberger, R., Springel, V., Hernquist, L., et al. 2017, , 465, 3291

Weinberger, R., Springel, V., Pakmor, R., et al. 2018, , 479, 4056

White, D. A., Jones, C., & Forman, W. 1997, , 292, 419

Wilms, J., Allen, A., & McCray, R. 2000, , 542, 914

ZuHone, J. A., Biffi, V., Hallman, E. J., et al. 2014, arXiv e-prints, arXiv:1407.1783

ZuHone, J. A. & Hallman, E. J. 2016, pyXSIM: Synthetic X-ray observations generator, *Astrophysics Source Code Library*, record ascl:1608.002

ZuHone, J. A., Vikhlinin, A., Tremblay, G. R., et al. 2023, SOXS: Simulated Observations of X-ray Sources, *Astrophysics Source Code Library*, record ascl:2301.024

Chapitre 3

Conclusion

Dans le cadre de ce mémoire, nous avons présenté une approche pour caractériser les amas de galaxies avec des méthodes d'apprentissage automatique. À partir d'un échantillon de 606 amas de galaxies provenant des simulations cosmologiques IllustrisTNG, nous avons produit 1818 observations simulées en rayons X du télescope *Chandra* à l'aide des logiciels `pyXSIM` et `SOXS`. Nous avons effectué de l'augmentation de données en appliquant des rotations et des réflexions, pour arriver à un ensemble de 14 544 images. Ensuite, nous avons réalisé une analyse de cinq propriétés du gaz chaud intra-amas : le temps de refroidissement central, la densité électronique centrale, l'excès d'entropie centrale, le paramètre de concentration et le paramètre de courbure. Pour obtenir leur valeur, nous avons d'abord construit des profils radiaux du temps de refroidissement, de la densité électronique et de l'entropie pour chaque amas, puis nous avons réalisé un ajustement à ces profils en utilisant les équations théoriques correspondantes. Ceci nous a permis d'obtenir les valeurs des paramètres qui nous intéressent, c'est-à-dire le temps de refroidissement central, la densité électronique centrale, l'excès d'entropie centrale et le paramètre de courbure. Pour le paramètre de concentration, le calcul est a été réalisé directement à partir des pixels de l'image en rayons-X.

Nous avons appliqué une méthode de regroupement de données (*clustering*) sur les images simulées; les groupements résultants démontrent une forte corrélation avec le paramètre de concentration. Ceci indique que la classification selon le paramètre de concentration reflète une séparation naturelle des images en rayons X d'amas de galaxies. Nous avons ensuite implémenté un réseau de neurones résiduel qui prédit simultanément les cinq propriétés étudiées, à partir d'une image simulée en rayons X. Les couches convolutives contenues dans le réseau permettent de traiter efficacement les pixels des images. Nous obtenons un pourcentage d'erreur de 1.8% pour le temps de refroidissement central, et une précision balancée de 0.83 pour le paramètre de concentration. Nos résultats démontrent donc que ces deux propriétés sont celles qui se prêtent le mieux à une analyse d'images

rayons X d’amas de galaxies avec des méthodes d’apprentissage automatique.

Afin d’obtenir des incertitudes sur les prédictions du réseau, nous avons réalisé une analyse probabiliste des prédictions à l’aide de la méthode SBI. Cette méthode permet d’obtenir, pour chaque prédiction, la distribution de probabilité des vraies valeurs possibles. Nos résultats démontrent des plus petites incertitudes sur les prédictions pour les amas NCC que les amas CC.

Finalement, on cherche à voir si le fait de prédire les cinq propriétés de manière simultanée plutôt que de les traiter une à la fois influence la capacité du réseau à faire des prédictions exactes. Nous avons donc re-entraîné le réseau, sans changer son architecture, pour prédire une seule propriété plutôt que les cinq en même temps. On remarque alors que la performance augmente considérablement pour le paramètre de concentration, mais que ce n’est pas le cas pour les autres propriétés. Cela démontre que l’information utilisée par le réseau pour prédire le paramètre de concentration est plutôt différente de l’information utilisée pour prédire les quatre autres métriques.

De manière globale, ce travail présente un outil efficace pour traiter de larges ensembles d’images d’amas de galaxies, fournissant une prédiction pour les cinq métriques différentes ainsi que les incertitudes associées. D’un côté, cet outil sert à efficacement comparer les métriques. D’un autre côté, il peut servir à faciliter la tâche de classification en ciblant les images pour lesquelles les incertitudes sont plus grandes, et permettant donc aux experts de donner plus d’attention aux amas de galaxies plus complexes.

Afin d’améliorer cet outil, un nombre de mesures peuvent être appliquées. Nous pouvons d’abord élargir notre ensemble de données d’entraînement en incluant un plus grand nombre de valeurs de *redshift* des simulations IllustrisTNG, ou bien en ayant recours à d’autres simulations cosmologiques qui contiennent des amas de galaxies. Il est également possible de varier le format sous lequel les données d’entrée sont présentées au réseau. Par exemple, nous pouvons y ajouter de l’information spectrale, puisque celle-ci est souvent déjà disponible pour les observations en rayons X. Nous pouvons également effectuer de la réduction de dimension comme étape de pré-traitement des images afin de présenter l’information sous une forme plus compacte, par exemple en ayant recours à l’analyse en composantes principales. De plus, nous pouvons modifier le réseau de neurones en y ajoutant davantage de flexibilité ou en incorporant une architecture différente. Finalement, une voie à explorer pour appliquer ce réseau sur des vraies données d’observation serait d’utiliser du transfert d’apprentissage. Pour ce faire, on pourrait débiter avec le réseau sous sa forme actuelle, puis le re-entraîner sur un sous-ensemble d’images d’observations d’amas de galaxies, pour

ensuite le tester sur un sous-ensemble différent d'images d'observations. Cela permettrait d'adapter le réseau aux particularités des vraies images d'observations, avant d'être appliqué sur les images d'amas de galaxies fournies dans les prochaines années par eROSITA.

Annexe A

Congrès où l'étudiante a présenté ses résultats

A.1. Rencontre annuelle du Centre de Recherche en Astrophysique du Québec (CRAQ)

Médium : Présentation orale

Titre : Classifier les amas de galaxies avec des méthodes d'apprentissage automatique

Lieu : Auberge Lac-à-l'Eau-Claire, Saint-Alexis-des-Monts, QC, Canada

Date : Mai 2023

Auteurs : M. Sadikov, J. Hlavacek-Larrondo, L. Perreault-Levasseur, C. Rhea, M. McDonald

A.2. Symposium Annuel de Physique pour un Avenir en Recherche et en Industrie (SAPHARI)

Médium : Présentation orale

Titre : Étudier aux cycles supérieurs en astrophysique

Lieu : Campus MIL, Montréal, QC, Canada

Date : Mars 2023

Auteurs : M. Sadikov

A.3. Octobre Numérique d'IVADO

Médium : Présentation orale

Titre : Classifier les amas de galaxies avec des méthodes d'apprentissage automatique

Lieu : Campus MIL, Montréal, QC, Canada

Date : Mars 2023

Auteurs : M. Sadikov, J. Hlavacek-Larrondo, L. Perreault-Levasseur, C. Rhea, M. McDonald

A.4. Rencontre annuelle de la Société Canadienne d’Astronomie (CASCA)

Médium : Affiche

Titre : A game of galaxy cluster classification : using machine learning to understand supermasive black holes

Lieu : En ligne

Date : Mai 2022

Auteurs : M. Sadikov, J. Hlavacek-Larrondo, L. Perreault-Levasseur, C. Rhea, M. McDonald

A.5. Poster Symposium Targeting Early-career Researchers (PoSTER)

Médium : Affiche

Titre : A game of galaxy cluster classification : using machine learning to better understand our universe

Lieu : En ligne

Date : Mai 2022

Auteurs : M. Sadikov, J. Hlavacek-Larrondo, L. Perreault-Levasseur, C. Rhea, M. McDonald

A.6. Symposium Annuel de Physique pour un Avenir en Recherche et en Industrie (SAPHARI)

Médium : Affiche

Titre : Classifier les amas de galaxies avec des méthodes d’apprentissage machine

Lieu : En ligne

Date : Mars 2022

Auteurs : M. Sadikov, J. Hlavacek-Larrondo, L. Perreault-Levasseur, C. Rhea, M. McDonald

A.7. Centenaire du Département de Physique de l’Université de Montréal

Médium : Affiche

Titre : Classifier les amas de galaxies avec des méthodes d’apprentissage machine

Lieu : Campus MIL, Montréal (QC), Canada

Date : Octobre 2021

Auteurs : M. Sadikov, J. Hlavacek-Larrondo, L. Perreault-Levasseur, C. Rhea, M. McDonald