Matter in the largest structures of the Universe : from galaxies to filaments, observations and data analysis

Victor Bonjean

PSL
UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'Observatoire de Paris

La matière dans les plus grandes structures de l'Univers : des galaxies aux filaments, observations et analyse de données

Soutenue par
Victor BONJEAN
Le 27 Septembre 2019

École doctorale n°127
Astronomie et astrophysique d'Île-de-France

Spécialité
Astronomie et astrophysique

Composition du jury :

Françoise COMBES Professeure au Collège de France	*Présidente*
Stéphane ARNOUTS Chargé de recherce au LAM	*Rapporteur*
Olivier DORE Research scientist au JPL	*Rapporteur*
Sophie MAUROGORDATO Directrice de recherche à l'OCA	*Examinatrice*
Simon, D. WHITE Professeur au MPA	*Examinateur*
Nabila AGHANIM Directrice de recherce à l'IAS	*Directrice de thèse*
Philippe SALOME Astronome au LERMA	*Directeur de thèse*

l'Observatoire de Paris | PSL

"John, the kind of control you're attempting simply is… it's not possible. If there is one thing the history of evolution has taught us it's that life will not be contained. Life breaks free, it expands to new territories and crashes through barriers, painfully, maybe even dangerously, but, uh... well, there it is. [...] I'm simply saying that life, uh... finds a way."

— Dr. Ian Malcolm, *Jurassic Park*

Remerciements

Tout au long de sa vie, l'Homme évolue, essaye, tombe, apprend des ses erreurs, devient plus mature (parfois...), et il y a quantité de choses qui changent. Souvent pour le mieux. Et il y a d'autres choses qui ne changent pas. C'est donc, -no comment-, à quelques heures de la deadline finale pour soumettre ce manuscrit à l'école doctorale que je compose ces quelques mots.

J'aimerais naturellement commencer par remercier Nabila et Philippe, mes deux directeurs, sans qui cette thèse n'aurait pas été possible (et sans qui soit dit en passant je n'aurais même probablement jamais eu de financement). Alors pour avoir cru en moi, m'avoir supporté, et m'avoir donné la liberté que j'ai eu la chance d'avoir durant mes travaux et mes explorations, merci.

Tout naturellement également je souhaite remercier mille fois ma famille: mes parents, pour avoir pu me donner l'opportunité d'en arriver là (je suppose que je n'ai pas toujours été facile...). Mes soeurs, pour m'avoir inspiré et m'avoir donné l'exemple. Mes grands-parents (je vous avais dit de ne pas vous inquiéter non ? :)). Papa, Maman, Aurélie, Léa, Papipaul, Mamini, et tous les autres, merci.

Eux aussi ont été présents tout le long de cette épreuve qu'est la thèse: je ne vais pas me risquer à commencer une liste de nom de peur d'en oublier, mais à tous les gens qui ont été proches de moi, un grand merci. Venant de Dax, Bordeaux, du master, du bureau, les co-LOCOS, et les autres... Un grand merci particulièrement à celle qui a probablement le plus souffert de mon état mental possiblement instable durant la période de rédaction (et pas que ..?) : nenita.

Un grand merci à tous les gens du labo, qui ont permis des discussions florissantes, tant scientifiques, que pas du tout. Donc merci à tous les thésards, post-docs, permanents, admins, et particulièrement aux membres plus ou moins proches de l'équipe ByoPiC.

Pour finir je souhaiterais aussi remercier les membres du jury, d'avoir pris le temps de lire et de comprendre ce manuscrit, et d'avoir pu faire des remarques pertinentes qui m'aideront dans ma potentielle future carrière de chercheur.

Contents

Part I

Cosmological context

Chapter 1

The large-scale structure of the Universe

Contents

In this chapter, I briefly describe the content and the structuration of our Universe, according to the present cosmological model and observations.

Résumé

Dans ce chapitre, je décris brièvement le contenu et la structuration de l'Univers, tel qu'il est compris selon les derniers modèles cosmologiques et les observations les plus récentes.

1.1 Introduction

From various observations of galaxy distribution, it is well established today that the Universe is structured in nodes connected by filaments and walls which surround large void regions : the Cosmic Web. These large-scale structure are very complex, and the physical processes occurring at these large scales are not quite well understood yet. As the structures in the Cosmic Web span a very wide range of densities, their detection, for the less dense ones, is still debated. Therefore, a key step is to detect and characterise the different structures of the Cosmic Web in order to understand their properties and their evolution.

Since the first observations of the night sky with telescopes (e.g., by Galileo), catalogues of stars, planets, and later of galaxies and even galaxy clusters were constructed, extended, and eventually made public. Today, we have built and launched the most advanced telescopes and satellites (like *Planck*, ROSAT, WISE, SDSS, LOFAR, or FERMI), and very promising ones are coming in the near future (like

SKA, Athena, SRG/eROSITA, Euclid, LSST, or WFIRST). These instruments have observed or will observe the entire sky (or at least a large portion of it) in a very wide range of the electromagnetic spectrum (from the radio to the gamma rays). Thanks to all these observations, we can draw a comprehensive picture of our Universe, for example in terms of galaxy distribution with SDSS, and also in terms of gas content with ROSAT or *Planck*.

Gigantic public databases of observations in multi-wavelength are being constructed, making their analysis and their combinations increasingly easy. However, all these available data (and even more in the near-future), opens the era of big data (e.g., 300 PB per year are expected in the case of SKA). The analysis of such quantities of complex data is and will be extremely delicate with the current computational facilities and the current approaches. The development today of new statistical methods to archive and analyse efficiently large data volumes is crucial. Some algorithms, especially machine learning and deep learning algorithms, perform better results than traditional approaches, and most of the time in a smaller amount of time. They are now used or tested in a variety of domains of Astrophysics from classification to emulation or component separation.

In my PhD thesis, I have investigated the properties of the largest scale structures of the Cosmic Web. To this aim, I have used publicly available data in different wavelengths, that I have analysed with statistical methods such as machine learning algorithms.

The following manuscript is organised in three parts: the first one, with chapters 1, 2, and 3, is dedicated to an introduction to the subject, a description of the public data analysed and a presentation of the statistical methods used. The second and the third parts of the manuscript include chapters presenting the different studies on the properties of matter around the large scale structure of the Universe and the data reduction of an ESO large programme. More specifically:

- In chapter 4, I describe a new method to estimate the star formation and the stellar mass of galaxies, that allows us to segregate populations of galaxies inside their host structures. This method is based on a machine learning algorithm, more specifically on random forest. In particular, this method led to the construction of value-added catalogues allowing the comparison of spatial distribution of galaxies (selected by their types) with the spatial distribution of the hot gas properties (derived from the Sunyaev-Zel'dovich (SZ) effect).

- In chapter 5, I present the analysis of an exceptional object: a galaxy cluster pair between the clusters A399 and A401 connected by a bridge of matter. This system has already been characterised in terms of gas via X-rays and the Sunyaev-Zel'dovich effect measured by *Planck*. For the first time I have added a study of the properties of the galaxies lying in between the two clusters.

- In chapter 6, I use the value-added catalogue containing more than 15 million sources constructed with the method described in chapter 4 to statistically characterise the properties of galaxies around cosmic filaments extracted from the SDSS survey. In this chapter, I present the profiles of galaxy density around identified filaments. I then explore the link between the ratio of passive over active galaxies and the profile of the hot gas around filaments. This study opens a path to assess environmental quenching inside the filaments.

- Chapter 7 presents a proof-of-concept study where I use for the first time a state-of-the-art algorithm of deep learning, namely a Convolution Neural Network, in a component separation context to detect the hot and diffuse gas via the SZ effect. This method appears promising for the detection and the characterisation of new individual clusters and complex filamentary structures.

- In chapters 8 and 9, I describe a large ESO programme aimed at confirming galaxy clusters detected via their SZ effect. I have conducted the data reduction of the image and spectra of galaxies in the surrounding regions observed with NTT/EFOSC2 and VLT/FORS2. For one particular object, a giant gravitational arc was detected. In the last chapter, I present the data reduction and the results of a dedicated observation of this object with MUSE.
- Finally, I conclude this manuscript with a summary of the results presented in the different chapters together with the different perspectives that this work opens.

The results presented in this manuscript have led to articles already published or in preparation:

- *Gas and galaxies in filament between clusters of galaxies: The study of A399-A401*, V. Bonjean, N. Aghanim, P. Salomé, M. Douspis, and A. Beelen, 2018, A&A, 609, A49
- *Star formation rate and stellar masses from machine learning*, V. Bonjean, N. Aghanim, P. Salomé, A. Beelen, M. Douspis, and E. Soubrié, 2019, A&A, 622, A137
- *Detection of intercluster gas in superclusters using the thermal Sunyaev-Zel'dovich effect*, H. Tanimura, N. Aghanim, M. Douspis, A. Beelen, and V. Bonjean, 2019, A&A, 625, A67
- *Properties and quenching of WISExSCOS galaxies around SDSS filaments*, V. Bonjean, N. Aghanim, M. Douspis, N. Malavasi, and H. Tanimura, in prep.
- *Extracting the Sunyaev-Zel'dovich effect in Planck with deep learning*, V. Bonjean, in prep.
- *Gas density and temperature in cosmic filaments on scales of tens of megaparsec*, H. Tanimura, N. Aghanim, N. Malavasi, V. Bonjean, A. Kolodzig, and M. Douspis, in prep.
- *Like a spider in its web: a study of the Large Scale Structure around the Coma cluster*, N. Malavasi, N. Aghanim, H. Tanimura, V. Bonjean, and M. Douspis, in prep.
- *Measurement of galaxy correlation with thermal Sunyaev-Zel'dovich emission*, G. Fabbian, F. Bianchini, N. Aghanim, M. Douspis, and V. Bonjean, in prep.

1.2 Composition and structuration of the Universe

1.2.1 Components of the Universe

Today, the Universe is composed of $\sim$ 69.4% of Dark Energy (DE), $\sim$ 25.8% of Dark Matter (DM), and $\sim$ 4.8% of ordinary matter (also called baryons), according to measurements from the Cosmic Microwave Background (CMB) shown in Fig. 1.1 (e.g., Planck Collaboration et al., 2016b). Among these different components, only baryons are directly observable. A budget in their different phases at low redshift has been performed (e.g., Fukugita, Hogan, and Peebles, 1998; Cen and Ostriker, 1999; Shull, Smith, and Danforth, 2012), and according to the latest study by de Graaff et al., 2019 (shown in the right panel of Fig. 1.1), about 14% of the baryons are in stars, cold gas (mostly in galaxies) and in the Circum Galactic Medium (CGM), and about 5% of the baryons are lying in a hot gas at temperature of about 10^8 K in the Intra Cluster Medium (ICM, in the centres of galaxy clusters). The remaining baryons are in the form of diffuse gas, either cold or warm (in the range $10^4 - 10^7$ K), surrounding the inter-galactic medium. Based on numerical simulations, baryons were traced along the epochs and about half of them are expected at low redshift in the form of a warm hot diffuse gas, at temperature of order $10^5 - 10^7$ K (shown in Fig. 1.2):

this phase is called the Warm Hot Inter-galactic Medium (WHIM, Davé et al., 2001; Haider et al., 2016; Martizzi et al., 2019). These baryons have been partially observed in densest environments (e.g., Eckert et al., 2015), with [OVI] absorption lines in quasars line of sights (e.g., Nicastro et al., 2018), or in measurements of the Sunyaev-Zel'dovich (SZ) effect (this effect is described later in Chap. 1.3.2.3) (e.g., Tanimura et al., 2019a; Tanimura et al., 2019b; de Graaff et al., 2019). As a matter of fact, about 20% of the baryons, expected to lie in the WHIM, are still not detected to date (de Graaff et al., 2019).

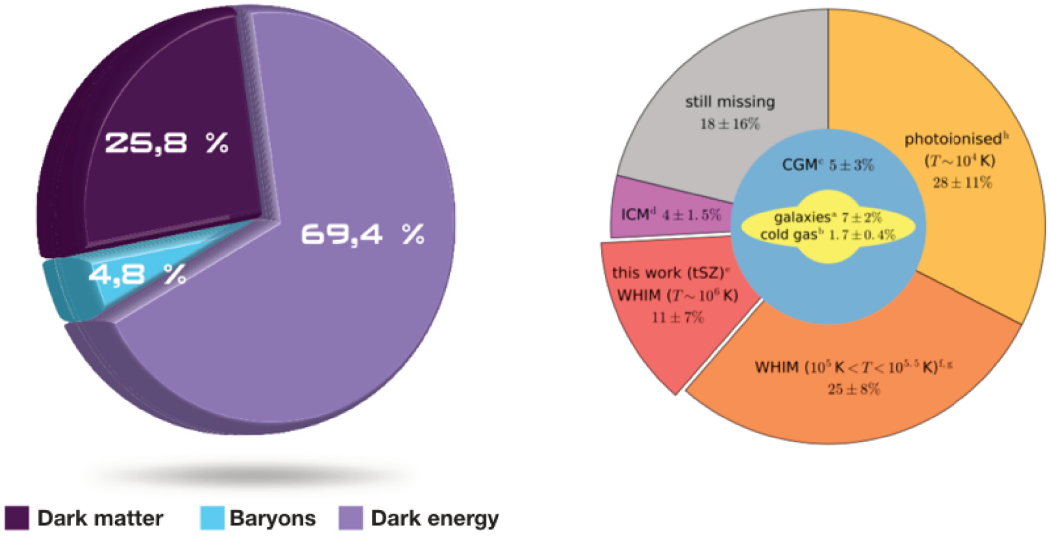

Figure 1.1: Left: content of our Universe according to *Planck* measurements (Planck Collaboration et al., 2016b). Figure adapted from http://public.planck.fr/multimedia/1-photographies. Right: baryon budget at low redshfit. Figure from de Graaff et al., 2019.

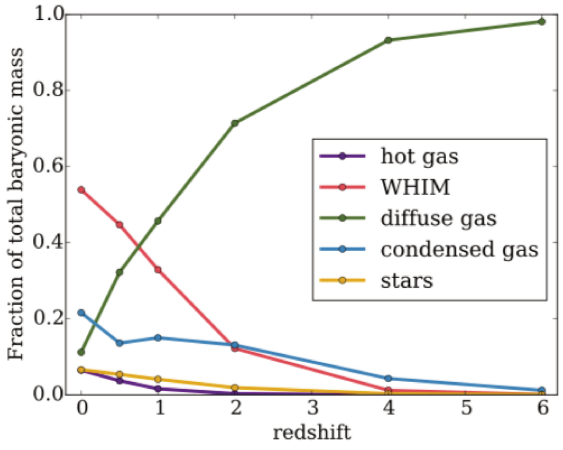

Figure 1.2: Evolution of the baryons in different phases from $z = 6$ to $z = 0$ based on numerical simulations. Half of the baryons are today lying in the WHIM. Figure from Haider et al., 2016.

1.2.2 Structuration into a Cosmic Web

The components of the Universe have evolved together, and eventually formed structures through accretion of matter from the very small fluctuations of the density field due to quantum fluctuations in the primordial Universe. Today, as observed in the distribution of galaxies in large surveys, like the Sloan Digital Sky Survey (SDSS, York et al., 2000), the matter in the Universe is distributed following a highly non-linear density field composed of nodes connecting filaments and walls (also called "pancakes" or "sheets"), themselves surrounding large void regions (shown in Fig. 1.3). This complex network is called the Cosmic Web (e.g., Bond, Kofman, and Pogosyan, 1996). Baryons go from voids to walls, from walls to filaments, and from filaments to nodes, flowing along the skeleton driven by the DM. Matter accreted into nodes eventually virialises and forms the largest gravitationally bound objects in the Universe: the galaxy clusters.

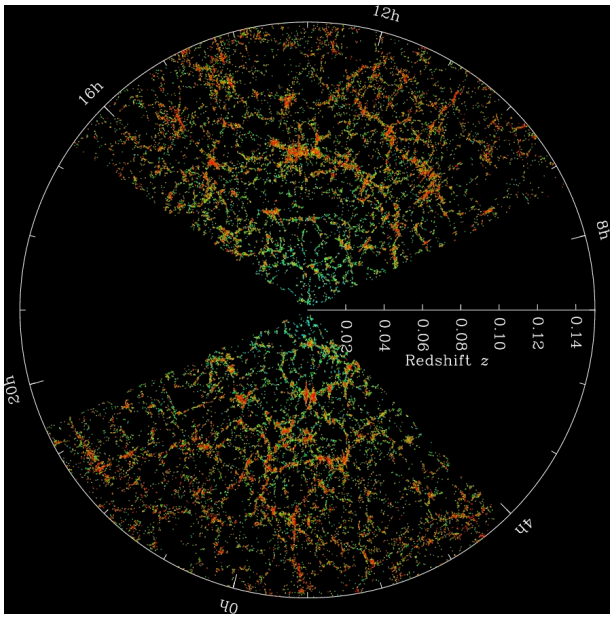

Figure 1.3: Distribution of galaxies in the SDSS survey. Image from https://www.sdss.org/science/. Credits: M. Blanton and SDSS.

This complex structure was also observed in the first N-body numerical simulations (e.g., Zel'Dovich, 1970; Doroshkevich and Shandarin, 1978). These simulations took only gravity into account (neglecting all baryonic effects), and resulted on a network of connected filamentary structures (shown in the left panel of Fig. 1.4). The first observations of the Cosmic Web were made later in the 80's, with the reconstruction of the galaxy distribution around a galaxy cluster in the Center for Astrophysics Redshift Survey (CfA, de Lapparent, Geller, and Huchra, 1986) (shown in the right panel of Fig. 1.4). This observation has demonstrated that galaxies were not randomly distributed in space, but rather assembled around "bubbles". A few decades later, other large galaxy surveys, e.g., the Two degree Field Galaxy Redshift Survey (2dFGRS, Colless et al., 2003), or the SDSS (Adelman-McCarthy et al., 2008), confirmed these typical filamentary structures connected to galaxy clusters and surrounding large voids (see Fig. 1.3).

As it is not possible to directly observe the DM, studying the properties of baryons (mainly in galaxies or hot gas) in the structures of the Cosmic Web is challenging. The most recent hydro-dynamical numerical simulations such as Millennium (Springel, 2005), Horizon-AGN (Dubois et al., 2014), BAHAMAS (McCarthy et al., 2017), or Illustris-TNG (Springel et al., 2018), have made extremely large number of particles (up to 15,625,000,000 DM particles in one of the Illustris-TNG simulation) evolve

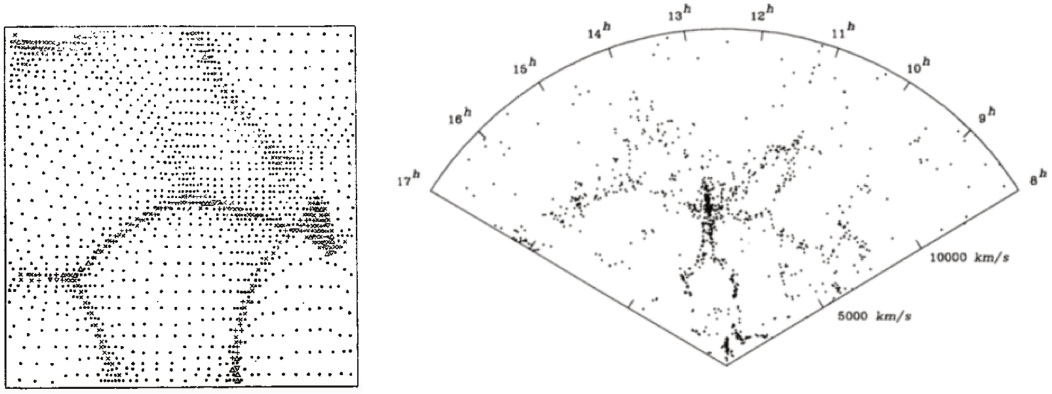

Figure 1.4: First hints of the Cosmic Web structure. Left: one of the first n-body simulations that reproduced the distribution of the matter in the Universe modelling gravity only. Figure from Doroshkevich and Shandarin, 1978. Right: first observation of the large-scale galaxy distributions from the CfA2 survey. Figure from de Lapparent, Geller, and Huchra, 1986.

through gravity and have also included the physical processes related to the interactions of baryons (e.g., hydrodynamics, magnetic fields, cooling, star formation, Active Galactic Nuclei (AGN) feedback, etc.). Hydro-dynamical simulations output properties of baryons around DM structures (as shown in Fig. 1.5), and it is then possible to construct mock images that reproduce observable quantities. Therefore, hydro-dynamical simulations offer the opportunity to study the property of the matter around the structures of the Universe, useful to compare with the observations.

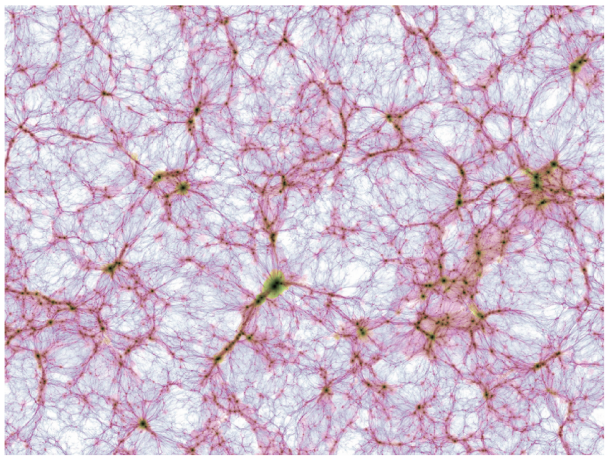

Figure 1.5: Snapshot at $z = 0$ of the TNG300 simulation from Illustris-TNG. It shows the baryonic density field, in a region with sides of about 300 Mpc. Image from the Illustris-TNG website (`http://www.tng-project.org/media/`).

1.2.3 Characterisation of the Cosmic Web

The Cosmic Web exhibits different environments, that mostly depend on the density field. Environments can be either local or global. They can be related to the different elements of the Cosmic Web: the nodes, the filaments, the walls and the voids; each of these environments spanning a wide range of densities and

scales, and drawing complex forms and shapes that follow the gradient of the density field. Environments can therefore be characterised through their over-density, defined as:

$$\delta(x) = \frac{\rho(x) - <\rho>}{<\rho>}, \tag{1.1}$$

where $\rho(x)$ is the density in the spatial position x, and $<\rho>$ is the mean density of the considered field.

1.2.3.1 Identification of structures

The Cosmic Web density field can be reconstructed based on the distribution of galaxies (in the case of observations), or directly on the dark matter particles (in the case of numerical simulations). Several methods have been developed to extract the structures and classify the Cosmic Web into nodes, filaments, walls, and voids, e.g., Bisous (Tempel et al., 2016), DisPerSE (Sousbie, 2011), or NEXUS+ (Cautun, van de Weygaert, and Jones, 2013). These methods are either based on geometrical pattern recognition (by designing specific filters), on topological analysis of the density field, or on the gradient of the density field. Therefore, the resulting detection and classification of the structures heavily depend on the method, making any comparison delicate to perform (see Libeskind et al., 2018 for a detailed review on the detection methods). An illustration of the outputs of nine different methods is shown in Fig. 1.6, where red regions indicate the recovered nodes, blue regions show the filaments, green regions display the walls, and white regions designate the voids. It is clearly seen that all these methods give a large variety of outputs, each one with different proportions and sizes of voids, walls, filaments, and nodes.

1.2.3.2 Over-densities of the Cosmic Web elements

Based on numerical simulations, Cautun et al., 2014 have realised an inventory of the Cosmic Web elements and their evolution. They have detected and classified the structures using their NEXUS+ algorithm. After their identification, they have performed an analysis of the distribution of over-densities associated with each structure. These distributions are shown in Fig. 1.7. Voids, in red, are as expected the lowest density regions, with typical over-densities in the range $0.01 < 1 + \delta < 1$. Voids also dominate the overall over-density distribution (in black) as they occupy most of the volume of the Cosmic Web. On the other side, nodes (in yellow) are the most over-dense objects, as expected, with over-densities starting at around $1 + \delta = 10$ and going up to very extreme values of the order of $1 + \delta = 10,000$. The over-density distributions of both walls and filaments mainly occupy the same range of values. Most of the wall over-densities (in green) are found in the range $0.1 < 1 + \delta < 10$, while filament over-densities (in blue) are found in the range $0.1 < 1 + \delta < 100$. We note that filament over-densities overlap with the nodes' distribution, suggesting that filaments can be rather dense structures. Moreover, the shape of the distribution suggests that there are two categories of filaments: the rather dense ($1 + \delta = 10 - 100$) and the less dense ($1 + \delta = 0.1 - 1$). The densest filaments might be small-sized bridges of matter in dense environments, which may have different properties from large cosmic filaments (e.g., Aragón-Calvo, van de Weygaert, and Jones, 2010).

1.2.3.3 Mass and volume fractions of the Cosmic Web elements

Cautun et al., 2014 have also studied the mass and volume fractions of the detected structures (shown in Fig. 1.8). The two structures that dominate the mass fraction budget are filaments and walls (which may look like filaments when projected in 2D): they account for $\sim 50\%$ and $\sim 24\%$ of the total mass fraction, respectively. Voids represent $\sim 15\%$ of the mass fraction, and nodes, where galaxy clusters lie, account for $\sim 11\%$. In terms of volume, $\sim 77\%$ of the volume fraction is occupied by voids, while filaments and

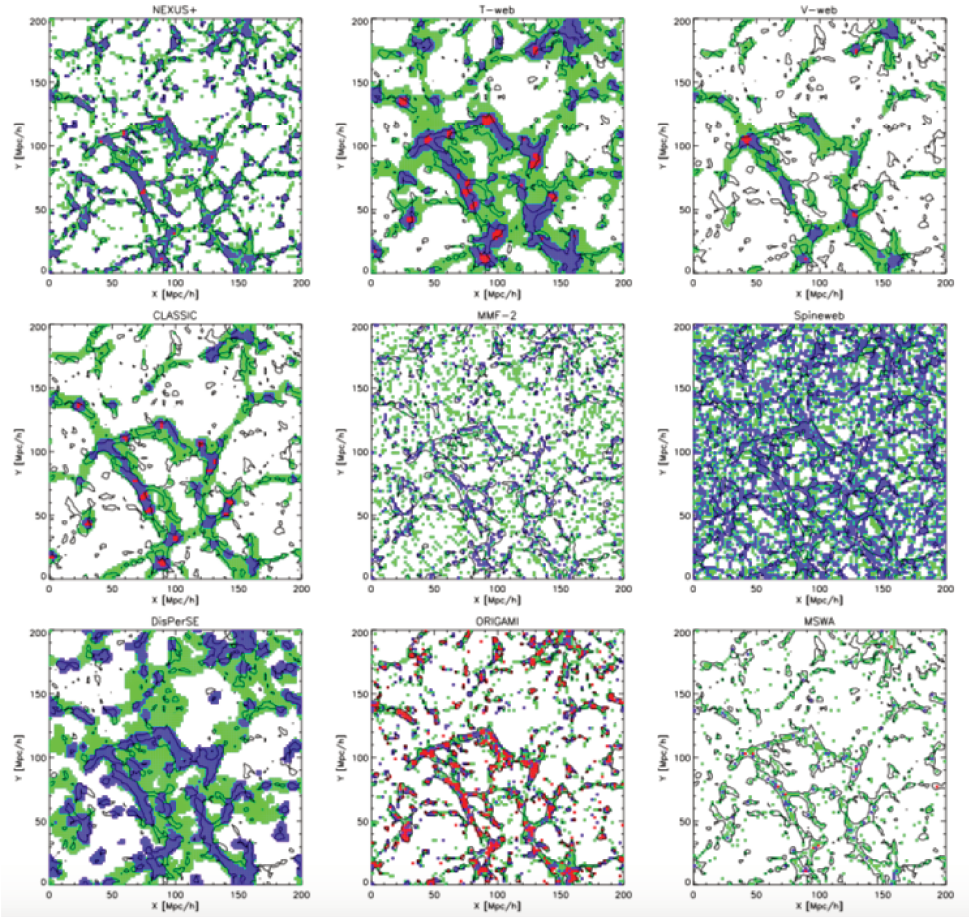

Figure 1.6: Comparison of nine methods to extract structures and classify them into nodes, filaments, walls, and voids. Figure from Libeskind et al., 2018.

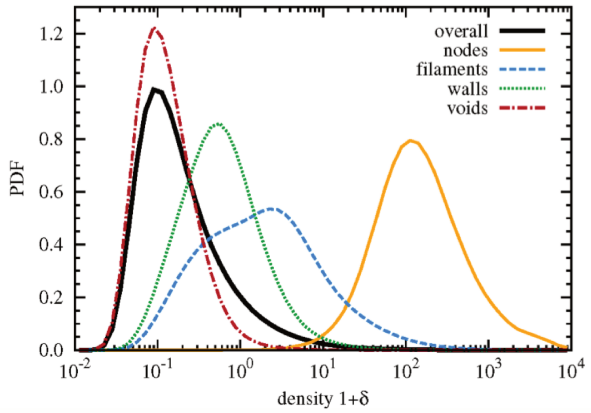

Figure 1.7: Histogram of the density $1 + \delta$ for the four different environments: nodes, filaments, walls, and voids, obtained from numerical simulations. Figure from Cautun et al., 2014.

walls account for $\sim 24\%$, and nodes for only $\lesssim 0.1\%$. Voids, although dominating the volume fraction budget, are one of the least massive structures because of their very low densities ($1 + \delta \sim 0.01$ as shown above). On the other hand, due to their very high densities ($1 + \delta \sim 100 - 1,000$), nodes represent

about the same mass fraction than voids despite their very tiny volume fraction. Filaments and walls occupy both a considerable volume fraction and represent most of the mass in the Universe, making those structures the most suitable to characterise the Cosmic Web.

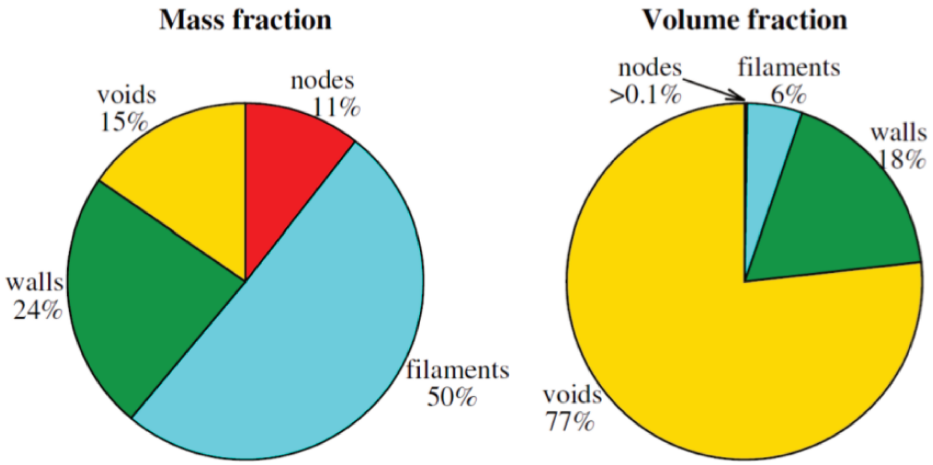

Figure 1.8: Mass and volume fractions occupied by the Cosmic Web structures, which are the nodes, the filaments, the walls, and the voids, from numerical simulations. Figure from Cautun et al., 2014.

1.3 Observation of collapsed objects

Now that the global picture of the Cosmic Web has been presented, I briefly describe here the main astrophysical objects used to trace these large scale structures and to characterise them: these objects are galaxies and galaxy clusters. I present their global properties, and how they are observed.

1.3.1 Galaxies

Galaxies are the easiest extra-galactic collapsed objects to observe since they emit in optical wavelength which is adapted to humans eyes. They were first observed in the 19th century and in the beginning of the 20th century, where they were originally defined as "diffuse nebulae" lying in our own galaxy. In the early 20's, E. Hubble confirmed their origin as extra-galactic, and since then galaxies were rapidly classified as a function of their morphologies and later as a function of their properties (e.g., their observed colours) (e.g., Hubble, 1926; de Vaucouleurs, 1959). Indeed, galaxies span a wide range of morphologies: spheroidal, elliptical, spiral, irregular, etc. They have typical sizes between few parsecs to few kpc, and have typical stellar masses in the range 10^8 – 10^{11} $M_\odot$. They are mainly composed of Dark Matter, that have formed halos which are at the origin of gravitational potential wells. They contain gas, which is the cradle of star formation; it has been accreted inside the potential wells, and it it inside galaxies in form of a cold neutral molecular gas with T $\lesssim$ 300 K, or warm ionised gas with T $\gtrsim$ 5000 K (Kalberla and Kerp, 2009)). Galaxies also contain stars, with populations tracing their ages and thus the age of the galaxies. The last main constituent of galaxies is the dust: it is related to the emission of stars as it absorbs the stellar light emitted in optical, producing a "redenning" (e.g., Trumpler, 1930), and re-emits it in the infra-red wavelengths. Therefore, dust and stars are the main ingredients to measure the stellar population of a galaxy. A central super massive black hole may lie in the centre of some galaxies, producing accretion of matter in their nuclei. The violent physical processes due to the interaction of rel-

ativistic gas and accretion jets emit powerful lights in the whole range of the electro-magnetic spectrum: these galaxies are called Active Galactic Nuclei (AGN) (e.g., Seyfert, 1943).

1.3.1.1 Galaxy bi-modality

Today, it is well established that there are two main populations of galaxies: active and passive galaxies. This is called the "galaxy bi-modality" (e.g., Baldry et al., 2004). The former population, the active one, contains younger galaxies, that are vigorously forming stars fuelled by their molecular gas. They appear bluer in the optical wavelengths, are rather not massive, and have typical spiral morphologies (as shown in the left panel of Fig. 1.9). Galaxies of the later population, the passive ones, have exhausted their molecular gas supplies and therefore no longer form stars; these galaxies are "red and dead", appearing in redder colours in the optical wavelengths. They are more massive, and have elliptical morphologies (shown in the right panel of Fig. 1.9).

Figure 1.9: Difference between passive and active galaxies. Left: the passive galaxy M87 observed with the Hubble Space Telescope. Right: the active galaxy M74 observed with the same telescope. Credits: NASA/ESA.

1.3.1.2 Star formation activity

Generalities

A main quantity allowing to distinguish galaxies from the two populations is the specific star formation rate (sSFR): how many stars they form compared to their stellar mass. This quantity can hence be deduced by the following two quantities: the stellar mass ($M_\star$), and the star formation rate (SFR). Indeed, displaying these two quantities on a diagram makes it possible to segregate passive from active galaxies (as shown in Fig. 1.10). This illustration shows positions in the SFR-$M_\star$ diagram of a sample of galaxies. Star-forming galaxies (blue dots in Fig. 1.10) are aligned along a line called the main sequence (e.g., Brinchmann et al., 2004; Elbaz et al., 2007). Galaxies leave the main sequence when they stop forming stars. This process is called "quenching" and happens when a galaxy looses its cold gas. This process is not well understood yet because it can be a combination of different phenomena, like the interaction of galaxies with an external hotter and denser gas (harassment (e.g., Moore et al., 1996), strangulation (e.g., Peng, Maiolino, and Cochrane, 2015), starvation (e.g., Trussler et al., 2018), ram pressure stripping (e.g., Gunn and Gott, 1972)), tidal interactions due to mergers, or ejection of the gas through AGN jets (e.g., Dubois et al., 2013). In all cases, galaxies stop forming stars and undergo a transitioning stage (green dots in Fig. 1.10, the so-called green-valley (e.g., Alatalo et al., 2014; Moutard et al., 2018), and finally settle in the region of the passive population (red dots in Fig.1.10). Galaxies can also undergo episodes of burst of star formation due to recent accretion of gas (e.g., in mergers): this

is called "starburstiness" (e.g., Elbaz et al., 2011). Such galaxies lie in the upper regions of the main sequence in the SFR-$M_\star$ diagram (shown in purple points in Fig. 1.10).

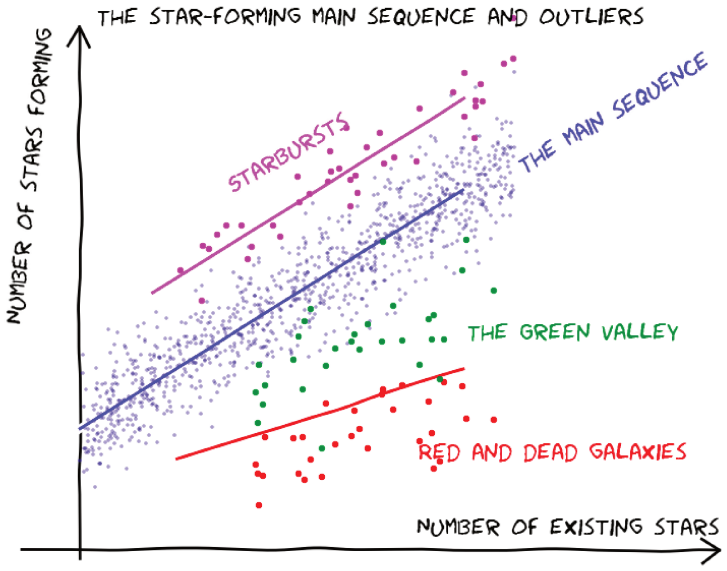

Figure 1.10: Schematic view of the SFR-$M_\star$ diagram. Active galaxies in blue follow a main sequence. Starbursts in purple and passive galaxies in red lie in the upper and in the lower region of the diagram, respectively. Transitioning galaxies in green are populating the green valley. Credits: CANDELS collaboration.

Estimating $M_\star$ and SFR

Estimating the quantities SFR and $M_\star$ is complex (see Kennicutt and Evans, 2012 for a review); they are directly or indirectly related to the observations of stars. But as presented above, the light of stars interacts with other components (e.g., it is absorbed by dust), so a good understanding of the contributions of the different elements interacting with the light from stars is necessary in order to properly estimate the quantities. An example of galaxy spectrum of the galaxy M82 is shown in Fig. 1.11. The different elements contributing to the total light escaping from galaxies (OB stars, non-ionizing stars, HII regions, synchrotron, and dust) have been reconstructed with models of Galliano, Dwek, and Chanial, 2008 and Galliano et al., 2011.

The young and massive O- and B-type stars are the hottest and thus the most energetic stars (e.g., purple line in Fig. 1.11). Their black-body spectra peak in the blue wavelength and they strongly emit in the UV. The UV luminosity of distant galaxies traces these types of stars that in turn directly relate to the SFR, as they represent the youngest stellar populations. However, at these wavelengths the dust absorption is very important and correcting the UV luminosities from the dust attenuation is not trivial (e.g., Lagache, Puget, and Dole, 2005; Kennicutt and Evans, 2012). Multi-wavelength tracers or dust attenuation estimations in the UV/optical are therefore needed to correct UV luminosities and use them as a direct tracer to derive estimations of SFR (e.g., Calzetti, Kinney, and Storchi-Bergmann, 1994; Kennicutt, 1998; Salim et al., 2007; Kennicutt and Evans, 2012; Janowiecki et al., 2017).

The non-ionizing low-mass old stars represent most of the contribution to the galaxy luminosities in the optical (e.g., orange line in Fig. 1.11). As they are the most numerous in a galaxy, the optical luminosity is also directly related to the number of stars, and thus to the stellar mass, given a theoretical model

of star population and an initial mass function (IMF) (e.g. Bruzual and Charlot, 2003). The estimation of stellar masses strongly depends on the assumed IMF. For example, a typical correcting/calibration factor of $\sim$ 1.6 is needed to change from a stellar mass with a Salpeter IMF (Salpeter, 1955) to a stellar mass with a Chabrier (Chabrier, 2003) IMF (Haas and Anders, 2010).

In the near IR (NIR) ($\sim$ 0.8μm$< \lambda <\sim$ 3μm), the old and non-massive stars also represent most of the contribution to the total luminosity. These wavelengths can therefore also trace the stellar mass through the old population, in the same way as optical measurements do (e.g., Wen et al., 2013). In the mid IR (MIR) ($\sim$ 3μm$< \lambda <\sim$ 70μm), the contribution of dust becomes predominant (e.g., red line in Fig. 1.11). Particularly in the 8-12 μm band, the contribution of heated small grains and polycyclic aromatic hydrocarbon (PAH, Leger and Puget, 1984) offers a useful tool to study the composition and the abundance of dust. From $\sim$ 20μm to $\sim$ 70μm, the luminosity is mostly due to thermalised dust and large grains heated by the UV emission of the energetic young O- and B-types stars. The luminosity in the IR is thus indirectly related to the SFR and some relation between these two quantities were built, using for example the 8μm and the 24μm bands from the Spitzer satellite (Werner et al., 2004), or the 12μm and the 22μm from the Wide-Field Infra-red Survey Explorer (WISE, Wright et al., 2010) satellite (e.g., Calzetti et al., 2007; Kennicutt et al., 2009; Jarrett et al., 2013; Cluver et al., 2014; Cluver et al., 2017).

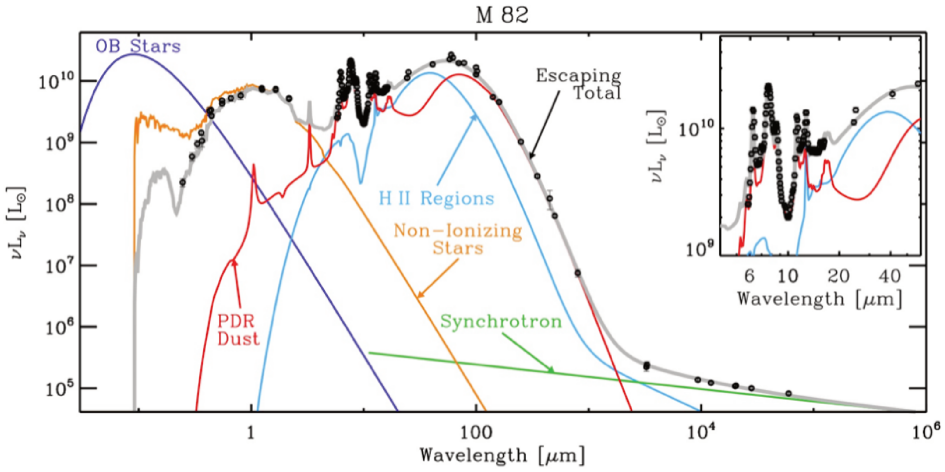

Figure 1.11: Spectrum of the galaxy M82, with observations from Spitzer, IRAS, ISO, WISE, 2MASS and Herschel. The contribution of the different components have been reconstructed with models of Galliano, Dwek, and Chanial, 2008 and Galliano et al., 2011. Image from http://dustpedia.com/science.php. Credits: DustPedia team.

1.3.1.3 Environmental dependency

An environmental dependency on the properties of galaxies has been shown both on observations and on simulations (e.g., Peng et al., 2010; Peng, Maiolino, and Cochrane, 2015; Alpaslan et al., 2016; Malavasi et al., 2017; Laigle et al., 2018; Kraljic et al., 2018; Pintos-Castro et al., 2019; Kraljic et al., 2019; Sarron et al., 2019).

In high density environments, like galaxy clusters, galaxies are observed more passive (e.g., Pintos-Castro et al., 2019). Indeed, galaxy clusters (described in detail the next section) are the "graveyards" of galaxies: galaxies are falling into their potential wells, and due to the interaction with the hot gas inside clusters, stop forming stars and slowly "die". As a result, galaxies in clusters contain mostly populations of old stars, exhibiting similar properties, thus resulting in similar observed spectral energy

distributions (SED). Therefore, clusters can be described as an aggregation of passive galaxies at same redshifts appearing with the same colours. However, as galaxies span a different range of magnitudes (which mostly depend on their sizes or on their stellar masses), they tend to align along a red sequence in colour-magnitude diagrams (e.g., Gladders and Yee, 2000). The red sequence is thus a good indicator of the presence of galaxy clusters: it is even used for their detections in large galaxy surveys (e.g., Rykoff et al., 2014).

Correlations between galaxy properties and less dense environments such as cosmic filaments have also already been observed. For example, spins of galaxies were found to align along the axis of the filaments: this is called "spin-alignment" (e.g., Jones, van de Weygaert, and Aragón-Calvo, 2010; Aragon-Calvo and Yang, 2014; Codis, Pichon, and Pogosyan, 2015; Kraljic, Dave, and Pichon, 2019). It has also been observed that galaxies inside cosmic filaments tend to be more massive and more passive (e.g., Alpaslan et al., 2016; Malavasi et al., 2017; Laigle et al., 2018; Kraljic et al., 2018; Sarron et al., 2019). Figure 1.12 shows the distributions of distances of passive and active galaxies around cosmic filaments detected with DisPerSE on VIPERS galaxies at $z \sim 0.7$. The median distance of most massive galaxies is smaller than the median distance of less massive ones, and the median distance to filaments of passive galaxies is smaller than the median distance of active ones. Recently, some studies have found that Cosmic Web environment may be sufficient to explain the quenching of galaxies (e.g., Gabor and Davé, 2015; Aragon-Calvo, Neyrinck, and Silk, 2016). Aragon-Calvo, Neyrinck, and Silk, 2016 have investigated the effects of the connections of the Cosmic Web in the accretion of the gas inside galaxies and the interactions with the diffuse gas inside filaments as the origin of quenching of galaxies, while Gabor and Davé, 2015 have reproduced the results from Peng et al., 2010 showing the correlation between environment and quenching of galaxies, simply by defining a temperature of quenching inside the host halos. However, the link between the Cosmic Web and the properties of galaxies is still not clear to date, but there may be a correlation with the hot and diffuse gas that is surrounding cosmic filaments and galaxy clusters (e.g., Kukstas et al., 2019). It is thus crucial to investigate both galaxy properties and gas properties around the Cosmic Web structures to understand the mechanisms of quenching.

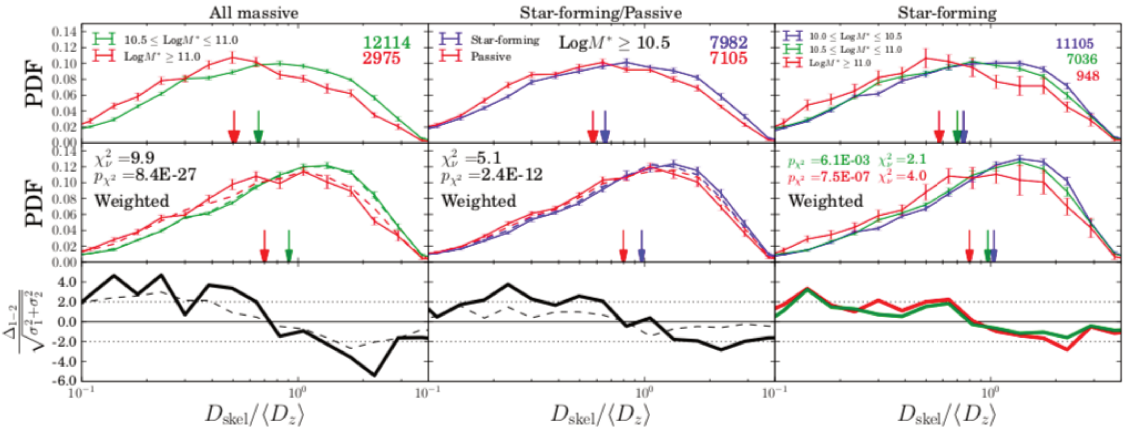

Figure 1.12: Distribution of distance of galaxies to filaments $D_{\rm skel}$ normalised by the mean inter-galaxy separation $< D_z >$ for three selections of galaxies. Left: selection in mass. Middle: selection in galaxy type. Right: selection in mass for active galaxies. Arrows indicate the median values of the distributions. Figure from Malavasi et al., 2017.

1.3.2 Galaxy clusters and large scale structures

1.3.2.1 General properties of clusters

Galaxy clusters are the largest gravitationally bound objects in the Universe. They have typical radius of ~ 1 Mpc and typical masses in the range $10^{13.5} - 10^{15}$ $M_\odot$. Their underlying over-densities can grow, in their centres, up to extreme values (i.e., $1 + \delta \geq 10^4$, as seen in Fig. 1.7).

Due to their high densities, galaxy clusters are relatively easy to observe. However, their extension and their boundaries are not easy to define. To define a cluster's radius, a threshold in density can be set, delimiting a volume in which a certain amount of density is encompassed. Usually, the density is defined as a function of the critical density of the Universe, $\rho_c = \frac{3H(z)^2}{8\pi G}$, and is expressed as follows:

$$\rho_{\Delta_c} = \Delta_c \rho_c. \tag{1.2}$$

Using this definition, a galaxy cluster is defined by the volume where $\rho > \Delta_c \rho_c$, where Δ_c is set arbitrarily. In the literature, three thresholds Δ_c are mainly used to define edges of galaxy clusters, from very central parts to large radii: $\Delta_c = 2500$, $\Delta_c = 500$, and $\Delta_c = 200$. The first two thresholds were initially set to define cores of clusters, in particular when using X-ray observations that are biased to the densest environments. The last threshold $\Delta_c = 200$ was introduced in numerical simulations as a rough proxy to the virial radius. Based on these density thresholds, a radius, and thus a mass can be derived. The radius is defined as the spatial extension up to which the threshold in density is reached: R_{Δ_c}. The mass is defined by the total mass contained inside R_{Δ_c}, and can be estimated with the density:

$$M_{\Delta_c} = \rho_{\Delta_c} V_{\Delta_c}, \tag{1.3}$$

where V_{Δ_c} is the volume embedded in R_{Δ_c}. Assuming spherical symmetry, the volume of a cluster in R_{Δ_c} is:

$$V_{\Delta_c} = \frac{4}{3}\pi {R_{\Delta_c}}^3. \tag{1.4}$$

Mixing Eq. 1.3 with Eq. 1.2, and Eq. 1.4, the mass M_{Δ_c} is thus related to the radius R_{Δ_c}, by:

$$M_{\Delta_c} = \frac{\Delta_c H(z)^2}{2G} {R_{\Delta_c}}^3. \tag{1.5}$$

1.3.2.2 Cluster content

Galaxy clusters are mainly composed of Dark Matter ($\sim 80\%$), hot gas ($\sim 16\%$) at temperature of order 10^8 K, and stars and cold gas in galaxies ($\sim 4\%$) (e.g., Bykov et al., 2015). Dark Matter (originally suggested by Zwicky, 1937) can be indirectly observable via gravitational lensing induced by high densities of clusters: the high dense region disturb space-time and lense background galaxies. This phenomenon produces gravitational arcs around galaxy clusters (e.g., Smail et al., 2007; Wuyts et al., 2010; Dahle et al., 2016), or weak lensing (e.g., Mandelbaum, 2018), allowing the reconstruction of their total masses. Dark Matter in clusters also produces CMB lensing, which also allows the reconstruction of their total masses (e.g., Planck Collaboration et al., 2018a). Studied in numerical simulations, Dark Matter density profiles around galaxy clusters turned out to be universally shaped, following a Navarro, Frenk, & White profile (NFW, Navarro, Frenk, and White, 1996; Navarro, Frenk, and White, 1997). Galaxies, as described in the previous section, are observable in optical and infra-red via the light emitted by stars and dust. Their distributions and their properties in clusters are statistically known thanks to observations in large galaxy surveys like the SDSS (e.g., Baxter et al., 2017; Chang et al., 2018; Adhikari et al., 2019;

Pintos-Castro et al., 2019). Hot gas in galaxy clusters can be observed in several ways. First, it is visible in X-rays via the Bremsstrahlung emission (e.g., Byram, Chubb, and Friedman, 1966; Bradt et al., 1967, for the first detections). This emission is a radiation produced by the deceleration of the hot free electrons by the charged protons lying in the hot ionised gas. Bremsstrahlung emission is proportional to $n_{\rm e}^2$ (where $n_{\rm e}$ is the electron density), making X-rays suitable to trace the densest regions of the hot gas in the ICM. Hot gas can also be detected by the Sunyaev-Zel'dovich effect, that is described in detail in the next section. Statistical properties of the hot gas around galaxy clusters, like pressure or entropy profiles, have been derived thanks to numerical simulations and to observations in X-rays and SZ (e.g., Nagai, Vikhlinin, and Kravtsov, 2007; Arnaud et al., 2010; Planck Collaboration et al., 2013a; Bartalucci et al., 2017; Ghirardini et al., 2019). The distributions of the different components in galaxy clusters are therefore known on average, even up to clusters' outskirts (see the reviews of Kravtsov and Borgani, 2012, Bykov et al., 2015, and Walker et al., 2019).

1.3.2.3 The Sunyaev-Zel'dovich effect

The Sunyaev-Zel'dovich effect (SZ, Sunyaev and Zeldovich, 1970; Sunyaev and Zeldovich, 1972) is a secondary CMB anisotropy, resulting from the interaction of CMB photons with a distribution of high energy electrons. CMB photons interact with the hot ionised electrons via inverse Compton scattering. Therefore, free electrons input energy to CMB photons and the CMB spectrum is slightly distorted with a very peculiar spectral signature: a decrement below 217 GHz, and an increment beyond (as shown in Fig. 1.13). The frequency dependency of the distortion is expressed as a function of the temperature change ΔT as:

$$\frac{\Delta \mathrm{T}}{\mathrm{T_{CMB}}} = f(x)y, \tag{1.6}$$

where $\mathrm{T_{CMB}}$ is the temperature of the CMB, $f(x)$ where $x = \frac{h\nu}{k_{\rm B}\mathrm{T_{CMB}}}$ is the frequency dependency, and y is the Compton parameter that quantifies the amplitude of the SZ effect. The Compton parameter y is given by:

$$y = \frac{\sigma_{\rm T}}{m_{\rm e}c^2} \int n_{\rm e}(l) k_{\rm B} T_{\rm e}(l) \mathrm{d}l, \tag{1.7}$$

where $\sigma_{\rm T}$ is the Thomson cross-section, $m_{\rm e}$ the mass of the electron, c the speed of light, $k_{\rm B}$ the Boltzmann constant, and $n_{\rm e}(l)$ and $T_{\rm e}(l)$ the density and the temperature of the free electrons along the line of sight, respectively.

The frequency dependency of the temperature change, $f(x)$, is given by:

$$f(x) = \left(x \frac{e^x + 1}{e^x - 1} - 4 \right) (1 + \delta_{\rm SZ}(x, \mathrm{T_e}), \tag{1.8}$$

where $\delta_{\rm SZ}(x, \mathrm{T_e})$ is the relativistic correction.

While the effect just described above is called the thermal SZ effect (tSZ), there is another effect: the kinetic SZ effect (kSZ). Here, energy of CMB photons are increased (or decreased) by Doppler effect induced by the global proper motion with respect to the CMB rest frame of a bulk of hot electrons moving towards (or backwards) us. For example, it may happen in galaxy clusters that are moving with respect to the CMB frame. In that case, a change of temperature is expressed as:

$$\frac{\Delta \mathrm{T}}{\mathrm{T_{CMB}}} = -\tau_{\rm e} \left(\frac{v_{\rm pec}}{c} \right), \tag{1.9}$$

where $v_{\rm pec}$ is the velocity of the moving electrons along the line of sight, c the speed of light, and $\tau_{\rm e}$ is the electron opacity:

$$\tau_{\rm e} = \sigma_{\rm T} \int n_{\rm e}(l) {\rm d}l. \tag{1.10}$$

The kSZ effect is another way of detecting the hot and diffuse gas in the Cosmic Web, but as shown in the right panel of Fig. 1.13, it is a very tiny effect, even smaller than the tSZ effect. Therefore its detection is for now only statistical (e.g., Planck Collaboration et al., 2016f; Planck Collaboration et al., 2018b) or in individual structures presenting high velocities observed with high resolutions and high sensitivity telescopes like NIKA2 (e.g., Adam et al., 2017). Hence I only mention the tSZ effect as the "SZ effect" in the following manuscript.

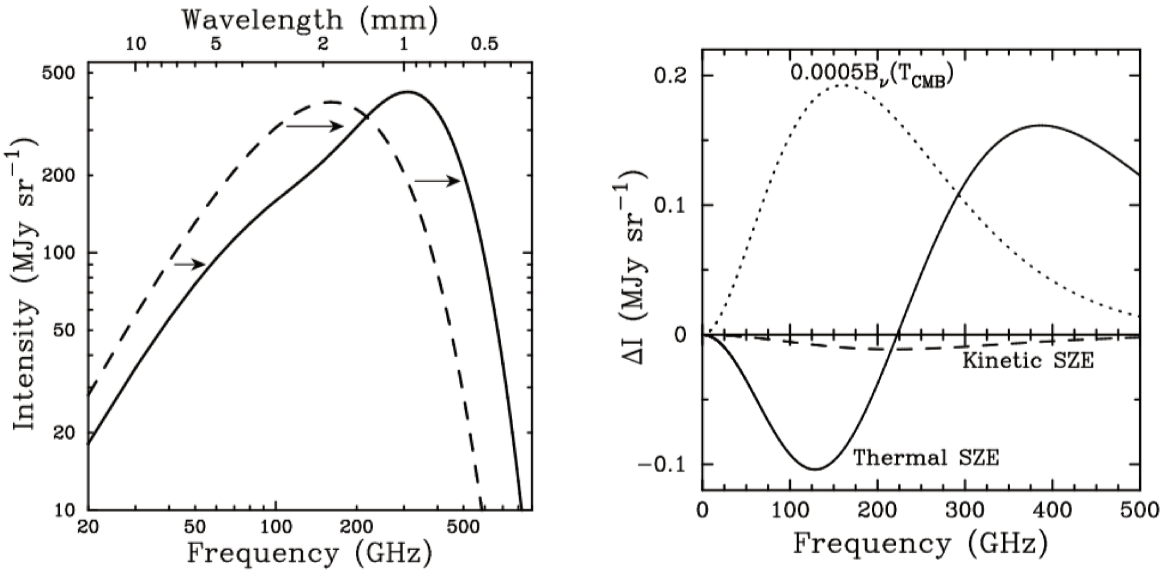

Figure 1.13: Left: CMB spectrum in dotted line, and distorded by the SZ effect in solide line. Here, the effect have been modeled for a galaxy cluster 1000 times more massive than typical galaxy clusters to illustrate the effect. Right: spectral distortion of the SZ effect. A decrement in the CMB itensity is seen below 217 GHz, and an increament is seen beyond. Figure from Carlstrom, Holder, and Reese, 2002.

The probability of a CMB photon to interact with an electron is quite small. Therefore, the SZ effect is preferentially detected in extended and dense regions, where the probability of a CMB photon to be scattered is the highest. Galaxy clusters are very extended objects containing an ionised ICM heated at temperature of about T $\sim 10^8$ K. This hot gas in clusters is an ideal environment to produce SZ effect. For example, this effect has allowed the characterisation and the detections of thousands of galaxy clusters, either in surveys like *Planck*, the Atacama Cosmology Telescope (ACT), or the South Pole Telescope (SPT) (e.g., Planck Collaboration et al., 2016c; Planck Collaboration et al., 2016e; Marriage et al., 2011; Hasselfield et al., 2013; Bleem et al., 2015), or in individual pointings with dedicated instruments like MUSTANG, NIKA2, or ALMA (e.g., Adam et al., 2018; Kitayama et al., 2016). A recent review presenting the astrophysics with the SZ effect have been made in Mroczkowski et al., 2019.

1.3.2.4 Large scale structure detections with stacking and SZ

The SZ effect is proportional to $n_{\rm e} \times T$, making it more sensitive to lower dense region compared to Bremsstrahlung emission seen in X-rays (as seen in Fig. 1.14). The SZ effect is therefore in theory the most suitable tracer of the diffuse gas lying around the Cosmic Web structures. As a matter of fact, the SZ effect has allowed first statistical and individual studies of the hot gas in clusters' outskirts (e.g., Planck

Collaboration et al., 2013c; Planck Collaboration et al., 2013b; Planck Collaboration et al., 2013a; Adam et al., 2018).

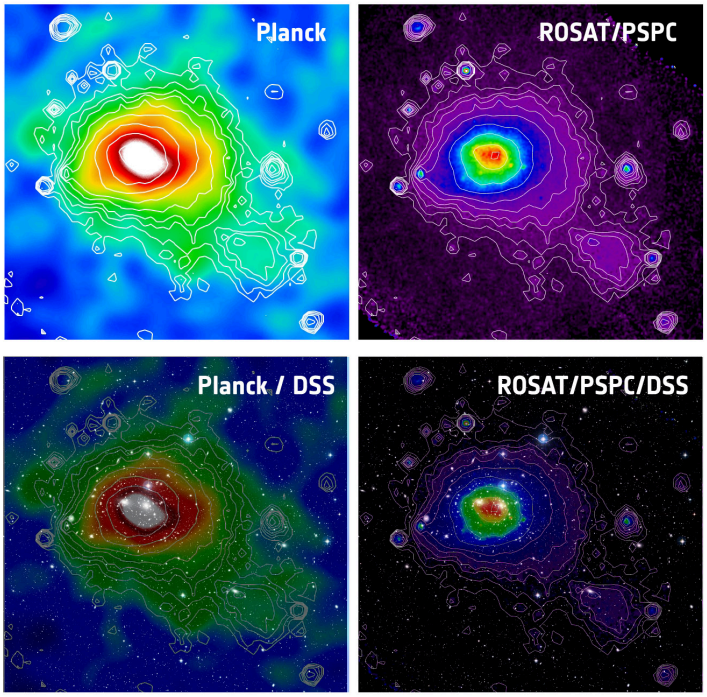

Figure 1.14: The Coma cluster as seen by *Planck* in SZ, and by ROSAT in X-rays. Credits: ESA.

Even outside the region of cluster's outskirts, the SZ effect has also allowed the first detections of the WHIM around different structures using the stacking of the SZ map reconstructed from *Planck*:

- Stacking Luminous Red Galaxies (LRG): the WHIM has first been detected in Tanimura et al., 2019b and de Graaff et al., 2019, where LRG pairs have been stacked in the SZ map and a positive residual emission have detected in between the galaxies.
- Stacking super-clusters: in an other study, Tanimura et al., 2019a have stacked the *Planck* SZ maps at the position of the super-clusters identified in Liivamägi, Tempel, and Saar, 2012, by masking all the known clusters up to $3 \times R_{500}$ lying inside. We have detected in this study an excess of SZ signal due to unbound gas between the clusters in the biggest potential wells generated by the dark matter in super-clusters of $y = (3.5 \pm 1.4) \times 10^{-8}$, that correspond to 17-52% of the missing baryons. The stacked SZ MILCA *Planck* map of the super-clusters masked from the known clusters, and the radial profile with the best fitted model of diffuse gas are shown in Fig. 1.15.
- Stacking filaments: in a more dedicated analysis, Tanimura et al., in prep. have stacked cosmic filaments detected in SDSS galaxies, which allowed the first characterisation in terms of temperature ($T=1.2 \pm 0.4 \times 10^6$ K) and over-density ($\delta \sim 25$) of the gas in cosmic filaments by the SZ detection. The stacked SZ profile of filaments is shown in Fig. 1.16.

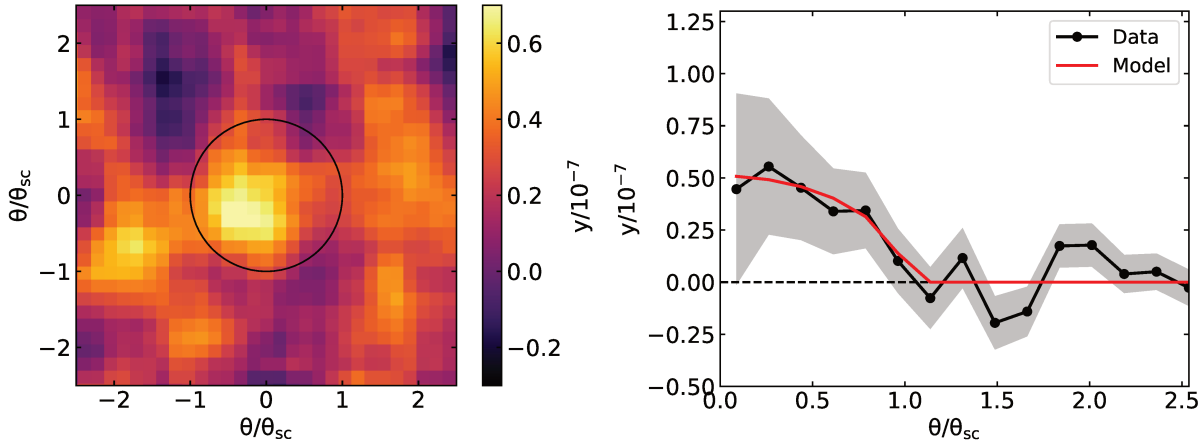

Figure 1.15: Left: MILCA y map stacked at the position of the super-clusters identified in Liivamägi, Tempel, and Saar, 2012. Clusters are masked up to $3\times R_{500}$. Right: radial profile of the stacked y MILCA map. A model of diffuse gas has been fitted, to relate the quantity of unbounded gas to 17-52% of the missing baryons. Figures from Tanimura et al., 2019a.

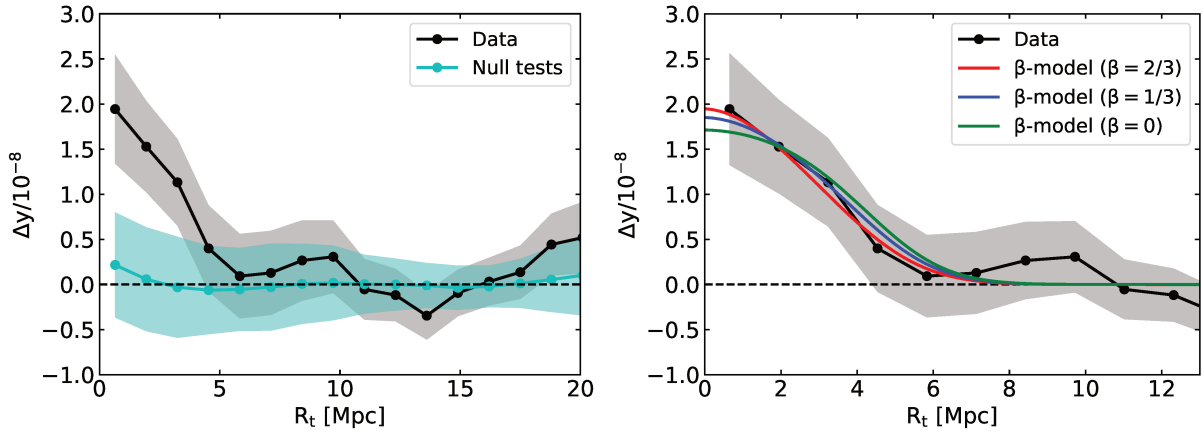

Figure 1.16: Stacked SZ profiles on the *Planck* MILCA SZ map around the cosmic filaments detected with DisPerSE in SDSS galaxies in the range $0.2 < z < 0.6$. Left: stacked profile and null-tests. Right: profile modelled with three β-models: $\beta = 2/3, \beta = 1/3, \beta = 0$. The best fits are for $\beta = 2/3$ and $\beta = 1/3$. Figure from Tanimura et al., in prep.

Chapter 2

Surveys and catalogues in different wavelengths

Contents

In this chapter, I describe the main public data in different wavelengths that I have used for my work. The data analysis aspects, such as for example the selections of the sources in the different catalogues (if any), will be presented in more detail in the different chapters.

Résumé

Dans ce chapitre, je décris les donnée publiques en différentes longueur d'ondes que j'ai utilisée pour mon travail. Les aspects d'analyse de données, par exemple la sélection des sources dans les différents catalogues, seront présentés dans les différents chapitres en détail.

2.1 The Sloan Digital Sky Survey

The Sloan Digital Sky Survey[1] (SDSS, York et al., 2000) is one of the largest available optical survey in astronomy. While other large surveys dedicated to the study of the galaxies cover deeper but smaller

[1] `https://www.sdss.org`

regions in the sky (e.g., GAMA (Driver et al., 2009; Driver et al., 2011), VIPERS (Guzzo et al., 2014), or COSMOS (Scoville et al., 2007)), the SDSS takes advantage of its very large footprint (more than 1/3 of the sky is observed, see Fig. 2.1). The SDSS, started in 2000, has provided images in five optical bands: u, g, r, i and z, and have performed spectroscopic measurements of more than three million astronomical objects (including stars, quasars, galaxies, etc.). It uses the facility of two 2.5m telescopes at the Apache Point Observatory (New Mexico, USA) and at Las Campanas Observatory (Atacama, Chile). The SDSS is a collaboration that aims at mapping the nearby Universe, and at understanding the physical properties of objects in different domains of astrophysics (e.g., stellar physics, cosmology, planetary science). As a matter of fact, SDSS data have been used to probe the existence of dark matter with the distribution of galaxies, to study the BAO, the physics of the galaxies, quasars, stars, and also to study the Milky way, the Solar system and other planetary systems[2].

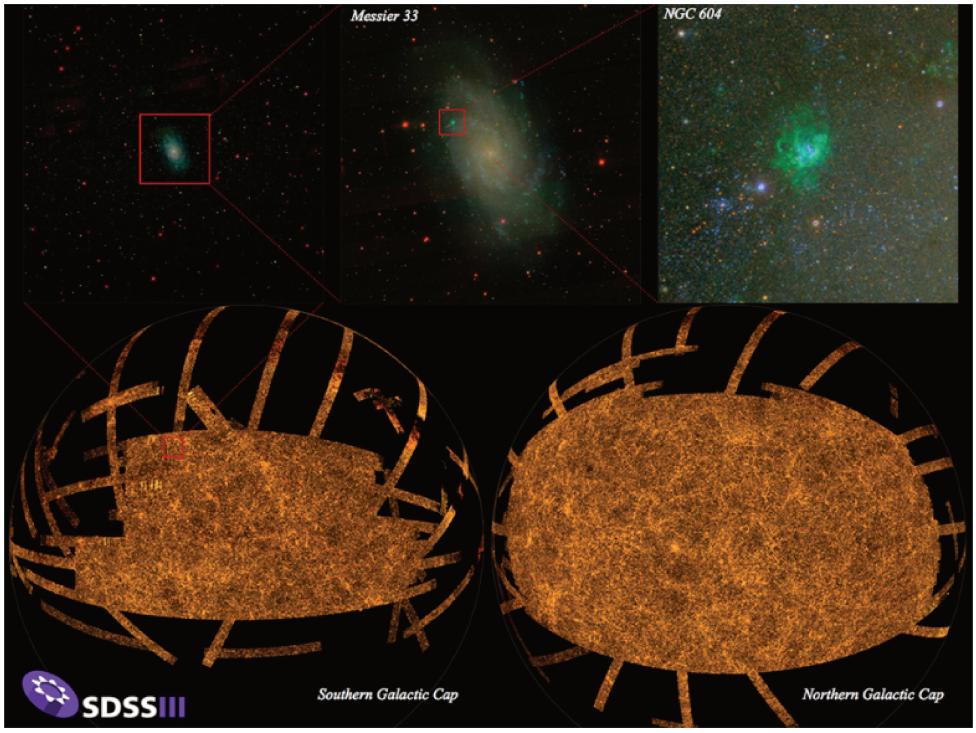

Figure 2.1: Image showing the footprint of the SDSS-III survey, together with observations of the galaxy Messier 33 and its HII region NGC 604. Image from https://www.sdss.org/surveys/. Credits: M. Blanton and SDSS.

The SDSS is currently in its fourth stage of observation, and has made public its 15th data release (SDSS-IV DR15). SDSS has released catalogues of galaxies seen in spectroscopy, i.e., with very reliable redshifts, like the LOWZ and the CMASS selections (both part of the BOSS survey (Dawson et al., 2013)), mapping the low redshifts ($0.15 < z < 0.45$) and intermediate redshift ($0.4 < z < 0.7$) Universe, respectively. I describe these two catalogues in detail in Chap. 6.2.1; the positions and redshifts of these two catalogues have been used by N. Malavasi (IAS, Orsay) to extract a catalogue of cosmic filaments. This catalogue is the base of the study presented in Chap. 6, where I investigate the statistical properties of galaxies around cosmic filaments. From the SDSS catalogue of galaxies observed in spectroscopy, several value-added catalogues have also been generated, based on more detailed study of galaxy spectra, in order to derive their physical properties, like the SFR or the stellar mass (e.g., Brinchmann et al., 2004; Chen et al., 2012; Maraston et al., 2013; Montero-Dorta et al., 2016). I have chosen the SDSS MPA-JHU

[2]https://www.sdss.org/science/

DR8 catalogue (described in detail in Chap. 2.1), as a reference to construct the training catalogue used to train the machine learning algorithm developed to estimate SFR and stellar mass of nearby galaxies, presented in Chap. 4.

The SDSS MPA-JHU DR8 catalogue

On the spectral SDSS galaxy catalogues released at each DR, value-added catalogue were computed by the MPA-JHU groups, from the Max Planck Institute for Astrophysics (MPA) and the Johns Hopkins University (JHU). They have studied in detail the emission lines in the spectra and fitted the continuum to extract different galaxy properties. The catalogue based on the DR8[3] of the SDSS, that is the MPA-JHU DR8 catalogue, is publicly available on the SDSS website[4]. It provides SFR and stellar masses for 1,843,200 galaxies with redshifts up to $z \sim 0.33$, computed following Kauffmann et al., 2003, Brinchmann et al., 2004, and Tremonti et al., 2004. They have used the photometric and the spectroscopic information of the SDSS data to estimate the SFR, leading to a more accurate measurement than using the spectra only. For this reason, I have chosen this catalogue to construct the training set of the machine learning algorithm developed to estimate SFR and $M_\star$ for nearby galaxies (presented in Chap. 4). The distribution of the sources in the SFR-$M_\star$ diagram of the SDSS MPA-JHU DR8 catalogue is shown in Fig. 2.2.

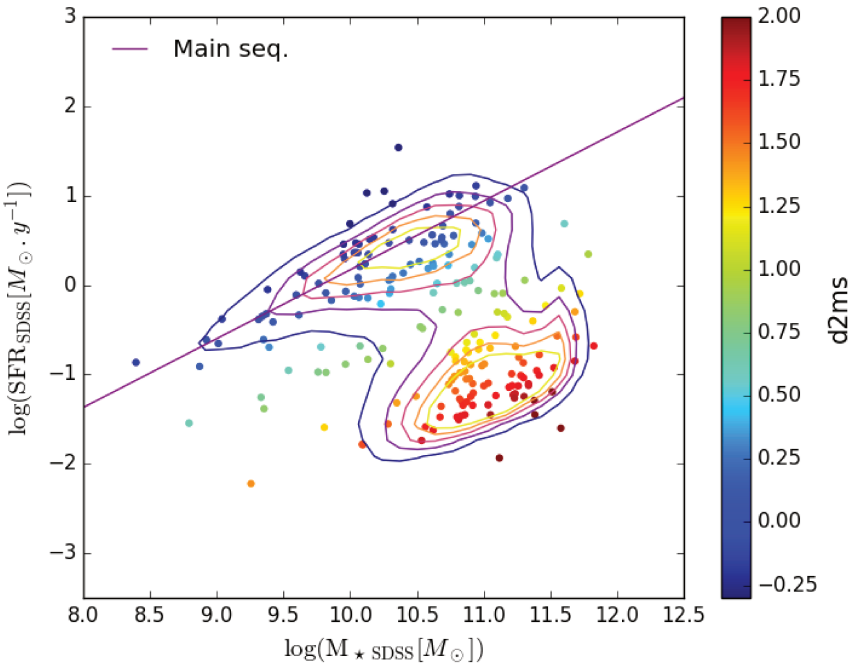

Figure 2.2: SFR-$M_\star$ diagram of SDSS MPA-JHU DR8 galaxies. The lines show the 1σ to 5σ contours. The dots represent 100 randomly selected galaxies from the catalogue. The purple solid line shows the main sequence of star forming galaxies given by (Elbaz et al., 2007). The colours of the galaxies indicate the distance to the main sequence, d2ms, illustrating the passivity.

The median SFR (flagged as SFR_TOT_P50 in the SDSS MPA-JHU DR8 catalogue) are estimated using the nebular emission lines corrected from the dust extinction. For star-forming galaxies, the H_α

[3] http://sdss3.org/dr8/
[4] https://www.sdss.org/dr14/data_access/value-added-catalogs/?vac_id=mpa-jhu-stellar-masses

emission-line and the Balmer decrement H_α/H_β in the fiber aperture are used as proxy of the SFR (Brinchmann et al., 2004), combined with the photometry in the u, g, r, i, and z bands outside the aperture (in order to take into account star forming regions in the entire extension of the galaxies). For AGN or no-emission-line galaxies, the SFR were estimated using a relation between the SFR and the spectral index D_{4000} (Bruzual A., 1983; Balogh et al., 1999; Brinchmann et al., 2004). The median $M_\star$ (flagged as LGM_TOT_P50 in the SDSS MPA-JHU DR8 catalogue) are computed based on Kauffmann et al., 2003. In Kauffmann et al., 2003, they have used Monte Carlo realisations of different star formation histories based on theoretical models of stellar populations of Bruzual and Charlot, 2003. They have assumed a Kroupa initial mass function (IMF) (Kroupa, 2001), and have fitted the models to the magnitudes in the u, g, r, i, and z bands, corrected from nebula emission lines from the spectra.

The SDSS MPA-JHU DR8 catalogue provides a flag (BPTCLASS) indicating the positions of the galaxies in the Baldwin, Phillips & Terlevich (BPT) diagram (Baldwin, Phillips, and Terlevich, 1981). This diagram segregates population of galaxies by comparing the emission-line ratios [OIII] $/H_\beta$ and [NII] $/H_\alpha$. In the classification provided by the MPA-JHU DR8 catalogue, BPTCLASS = 1 corresponds to star-forming galaxies, BPTCLASS = 2 to composite galaxies (transitioning), BPTCLASS = 3 to AGN, BPTCLASS = 4 and BPTCLASS = 5 to low-S/N emission lines galaxies (Brinchmann et al., 2004). The class BPTCLASS = -1 corresponds to galaxies unclassifiable in the BPT diagram: passive galaxies without emission lines (Brinchmann et al., 2004). These classes are efficient to segregate the populations of galaxies, as shown in Fig. 2.3. The BPTCLASS = 1 galaxies are well aligned on the main sequence of star forming galaxies, the BPTCLASS = -1 galaxies are well in the red cloud of passive galaxies, and the BPTCLASS = 2 galaxies are well populating the green valley between the two main populations.

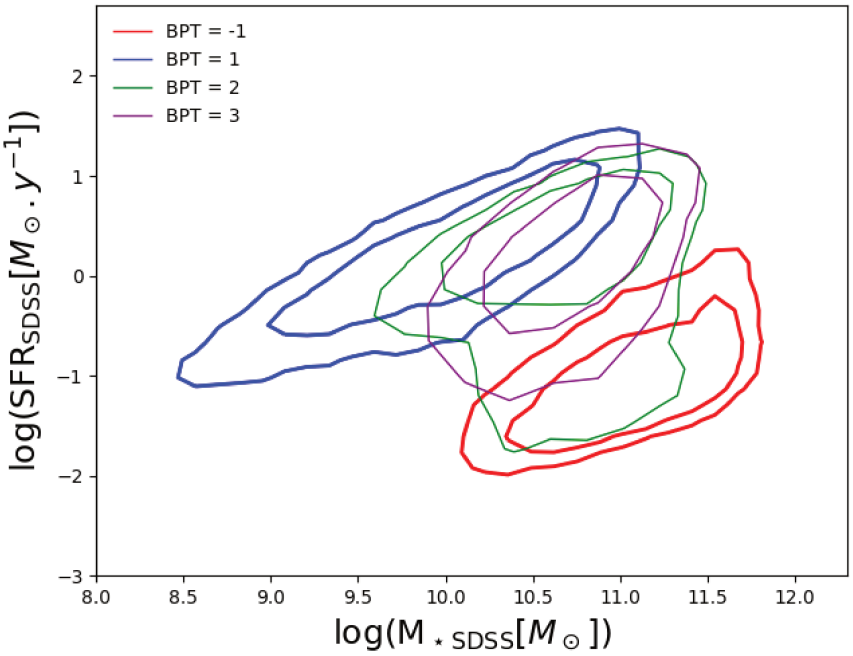

Figure 2.3: BPT classes of the SDSS MPA JHU DR8 catalogue in a SFR-$M_\star$ diagram. The contours show the different BPT classes: BPT = -1 in red show the passive galaxies, BPT = 1 in blue the active galaxies, BPT = 2 in green the transitioning galaxies, and BPT = 3 in purple the AGN.

2.2 The WISE satellite

The Wide-field Infra-red Survey Explorer (WISE) satellite (Wright et al., 2010) has observe the full sky in infra-red with a 40 cm telescope. It was launched on December 2009, and has observed the full-sky in four near- and mid infra-red wavelengths, at 3.4μm (W1), 4.6μm (W2), 12μm (W3), and 22μm (W4), providing maps with spatial resolutions of 6.1", 6.4", 6.5" and 12", respectively. The WISE survey has a sensitivity 1,000 times better than the first survey in infra-red, i.e., the Infra-red Astronomical Satellite (IRAS, Neugebauer et al., 1984). The wavelengths of WISE are ideally placed in the electromagnetic spectrum to study the dust in the galaxies and also the emission of old stars. When combined with other surveys in optical, like the SDSS, an almost complete view of the star formation ongoing in nearby galaxies can be retrieved, as the information of both old stars and dust emitting the light absorbed from the young stars are available. Thus, its wavelengths range and its incredibly large number of detected sources (> 700,000,000), make WISE an ideal instrument to estimate galaxy properties such as SFR and stellar mass. An image combining the full sky observations of WISE in W1, W3, and W4, is shown in Aitoff projection in Fig. 2.4.

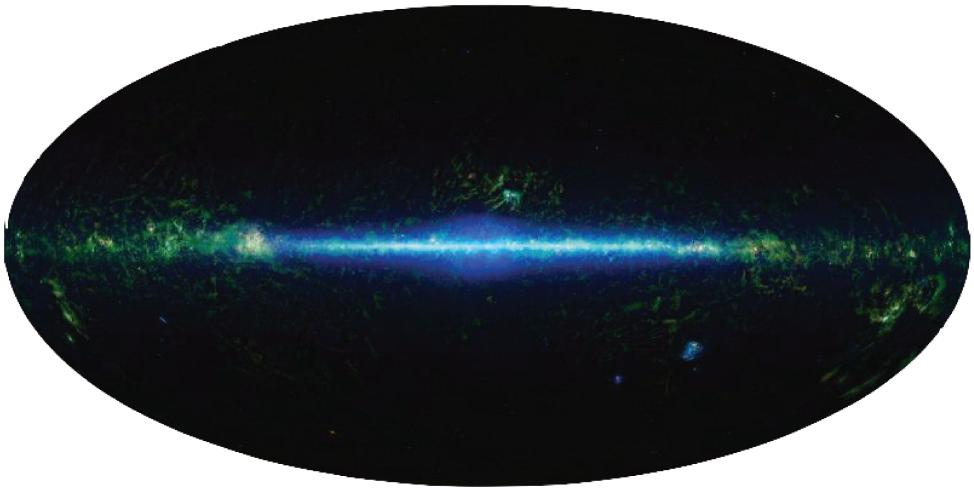

Figure 2.4: Image combining the full sky observations of WISE in the W1 (blue colour), W3 (green colour), and W4 (red colour) bands. Image from the NASA's website: https://www.nasa.gov/mission_pages/WISE/multimedia/pia15481.html. Credits: NASA/JPL-Caltech/UCLA.

2.2.1 The AllWISE Source Catalogue

From the co-added WISE Atlas Images, the AllWISE Source Catalogue was generated and made public[5], with accurate positions, photometry, and ancillary informations for 747,634,026 detected sources (Cutri et al., 2013). The catalogue provides several informations about the sources, such as their magnitude and their signal-to-noise ratio in the four bands, their position on the sky, their ellipticity, etc., together with

[5] Available at `http://wise2.ipac.caltech.edu/docs/release/allwise/expsup/sec1_3.html#src_cat`

the errors of all these quantities. The magnitudes reported in the AllWISE catalogue were computed with different methods (e.g., profile fitting, aperture photometry). For some of these magnitudes, a correction is needed. This is detailed in Chap. 4 where I present the construction of the training set of galaxies used by the machine learning algorithm to estimate SFR and $M_\star$ for nearby galaxies.

2.2.2 WISE-based full sky photometric redshift catalogues

A key information to map the Universe through the distribution of galaxies is to know their redshifts z. It is possible to estimate the redshifts of galaxies very precisely by identifying emission or absorption lines in their spectra. However, spectroscopic observations are expensive in terms of observing time. A parallel (and less expensive) way of estimating the redshifts is the use of the photometric observations in the widest possible range of wavelengths, to fit templates of spectra based on observations or on theoretical models (e.g., `LEPHARE` or `CIGALE` Arnouts et al., 1999; Ilbert et al., 2006; Burgarella, Buat, and Iglesias-Páramo, 2005). Another approach is the use of machine learning algorithms applied on magnitudes and/or colours (e.g., `ANNz` Collister and Lahav, 2004). WISE data have been successfully used to extend to the infra-red wavelengths the range of photometric bands of optical surveys. This has allowed to construct photometric redshift catalogues with increased accuracy. I present here the two full-sky catalogues of photometric redshifts based on WISE that I have used for my analyses: the 2MPZ and the WISExSCOS catalogues.

I have used these two catalogues as their very high statistics and their full-sky coverage enable the study of galaxies in very large fields of view outside of the SDSS footprint, needed to study the large-scale structures. In addition, the machine learning algorithm developed in Chap. 4 can be applied to the WISE measurements and to the photometric redshift estimations available in the catalogues, allowing the estimations of SFR and $M_\star$ of a very high number of sources for both catalogues.

2.2.2.1 The 2MPZ catalogue

The 2MPZ publicly available catalogue[6] (Bilicki et al., 2014) is a cross-match between the WISE infra-red survey, and two near infra-red and optical full-sky surveys: the Two Micron All Sky Survey (2MASS, Skrutskie et al., 2006), and SuperCOSMOS (Hambly et al., 2001a; Hambly, Irwin, and MacGillivray, 2001; Hambly et al., 2001b). 2MASS has observed the J (1.25μm), H (1.65μm), and K (2.17μm) bands in the late 90's with two 1.3m telescopes at the Fred Lawrence Whipple Observatory on Mount Hopkins, Arizona, USA, and at the Cerro Tololo Inter-American Observatory, in Chile. SuperCOSMOS is a digitisation of the sky survey plates taken with the 1.24m UK Schmidt telescope (UKST) in Siding Spring Observatory, Australia, the 1m ESO Schmidt at La Silla, Chile, and the 1.22m Palomar Schmidt, in Mount Palomar, California, USA. The SuperCOSMOS data are publicly available, and provide magnitudes in three optical bands: B, R, and I. The 2MPZ catalogue contains about one million nearby sources, with spectroscopic redshifts for about a third of them, and photometric redshifts for the remaining two thirds, estimated using the library `ANNz` (Collister and Lahav, 2004) trained on SDSS spectral galaxies. The median redshift of the catalogue is $z_{\rm med} \sim 0.08$ and the statistical error on the redshift is $\sigma_z \sim 0.012$. The low sensitivity of 2MASS and the magnitude cut in the 2MASS catalogue $K < 13.9$ (to ensure catalogue uniformity), make this catalogue useful to statistically study the distribution of the brightest and closest galaxies (i.e., the highest mass galaxies). The distribution of the 2MPZ sources as a function of redshift is shown in Fig. 2.5.

[6] `http://ssa.roe.ac.uk/TWOMPZ.html`

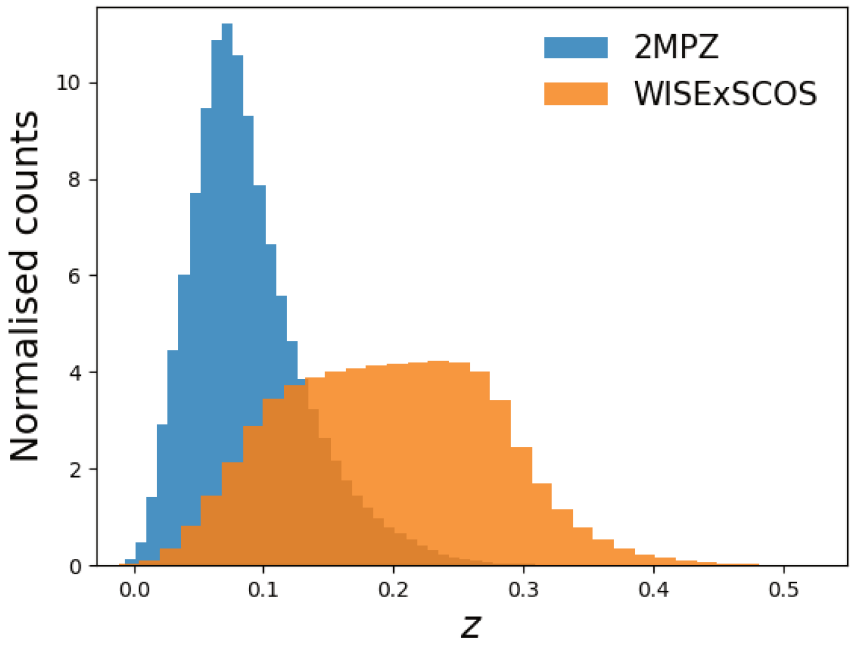

Figure 2.5: Distribution of the 2MPZ and of the WISExSCOS all-sky photometric redshifts catalogues as a function of redshift.

2.2.2.2 The WISExSCOS catalogue

An extension of the 2MPZ was realized by cross-matching WISE and SuperCOSMOS. This catalogue, the WISExSCOS (Bilicki et al., 2016), is also publicly available[7]. For the construction of this catalogue, Bilicki et al., 2016 have no longer made use of the 2MASS survey, that limited the redshift range of the 2MPZ. The WISExSCOS catalogue contains about 20 million of sources and covers about 70% of the sky (outside our galaxy), with a median redshift of $z_{\rm med} = 0.2$, and with a statistical error of $\sigma_z \sim 0.033$. The catalogue provides photometric redshifts for all the sources, estimated with `ANNz` trained on spectroscopic redshifts from the GAMA survey (Driver et al., 2009; Driver et al., 2011). The catalogue has a lower magnitude limit at W1 > 13.8, making it complementary to the 2MPZ catalogue for less bright galaxies. The distribution of the WISExSCOS sources as a function of redshift is shown in Fig. 2.5.

2.3 The *Planck* satellite

2.3.1 *Planck* frequency maps

The *Planck* mission (Tauber et al., 2010) dedicated to cosmology was launched in May 2009. The *Planck* satellite has observed the full sky in nine frequencies, well chosen to encompass the CMB black-body emission and study its anisotropies at an unprecedented precision. *Planck* is the third generation of satellite that aimed at studying the CMB, after COBE (Boggess et al., 1992) and WMAP (Mather et al., 1990). *Planck* comprised two main instruments: the Low Frequency Instrument (LFI, Bersanelli et al., 2010; Mennella et al., 2011), which has observed the sky at 30, 44, and 70 GHz with angular resolutions respectively of 33.29, 27.00, and 13.21 arcmin, and the High Frequency Instrument (HFI, Lamarre et al.,

[7] `http://ssa.roe.ac.uk/WISExSCOS.html`

2010; Planck HFI Core Team et al., 2011) that observed the sky at 100, 143, 217, 353, 545, and 857 GHz with angular resolutions of 9.68, 7.30,5.02, 4.94, 4.83, and 4.64 arcmin, respectively.

The *Planck* collaboration has provided the community with nine full-sky maps of the sky, all of them publicly available on the Planck Legacy Archive[8], in the HEALPIX format[9]. The nine *Planck* frequency maps are shown in Fig. 2.6 in Mollweide projection. The figure is from Planck Collaboration et al., 2016a. I use the *Planck* frequency maps in Chapt 7 to extract the SZ effect with a deep learning algorithm.

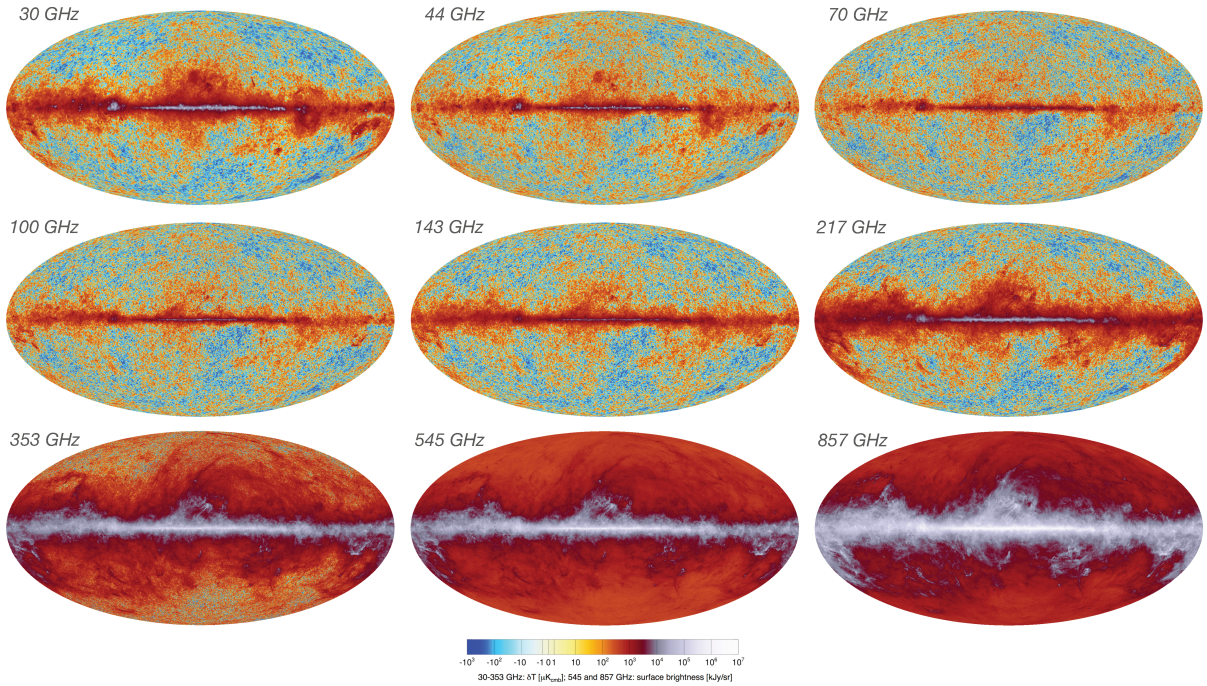

Figure 2.6: Mollweide projections of the sky seen in the nine frequencies of the *Planck* satellite. The figure is from Planck Collaboration et al., 2016a. Credits: ESA and the *Planck* Collaboration.

2.3.2 A *y*-map by *Planck*

The frequencies of *Planck* have been specifically chosen to measure the SZ effect from the hot gas. Based on two component separation techniques, i.e., the Needlet Internal Linear Combination (NILC, Remazeilles, Delabrouille, and Cardoso, 2011) and the Modified Internal Linear Combination Algorithm (MILCA, Hurier, Macías-Pérez, and Hildebrandt, 2013), the *Planck* collaboration has constructed full-sky maps of the *y* SZ Compton parameter at a resolution of 10 arcmin, using the six highest frequencies (Planck Collaboration et al., 2016c). At several times in this manuscrit, i.e., in Chap. 5, Chap. 6, and Chap. 7, I use the latest MILCA *Planck* *y* map from 2015 Planck Collaboration et al., 2016c. This map is publicly available[10] in the HEALPIX format[11], with $n_{\rm side}$=2048 and a pixel size of $\theta_{\rm pix}$=1.7 arcmin. An orthographic projection of the northern and the southern hemispheres of the MILCA *y* map is shown in Fig. 2.7. The figure is from Planck Collaboration et al., 2016c. This map has been extensively used, for example to measure the power spectrum of the SZ emission (e.g., Hurier and Lacasa, 2017), or to derive

[8] `https://pla.esac.esa.int/#home`
[9] Górski et al., 2005.
[10] `http://pla.esac.esa.int/pla/`
[11] Górski et al., 2005.

a universal pressure profile of the gas around galaxy clusters (Planck Collaboration et al., 2013a). The SZ map has also been used for dedicated analysis of individual objects such as the Coma cluster (Planck Collaboration et al., 2013c), or the galaxy cluster pair A399-A401 (Planck Collaboration et al., 2013b).

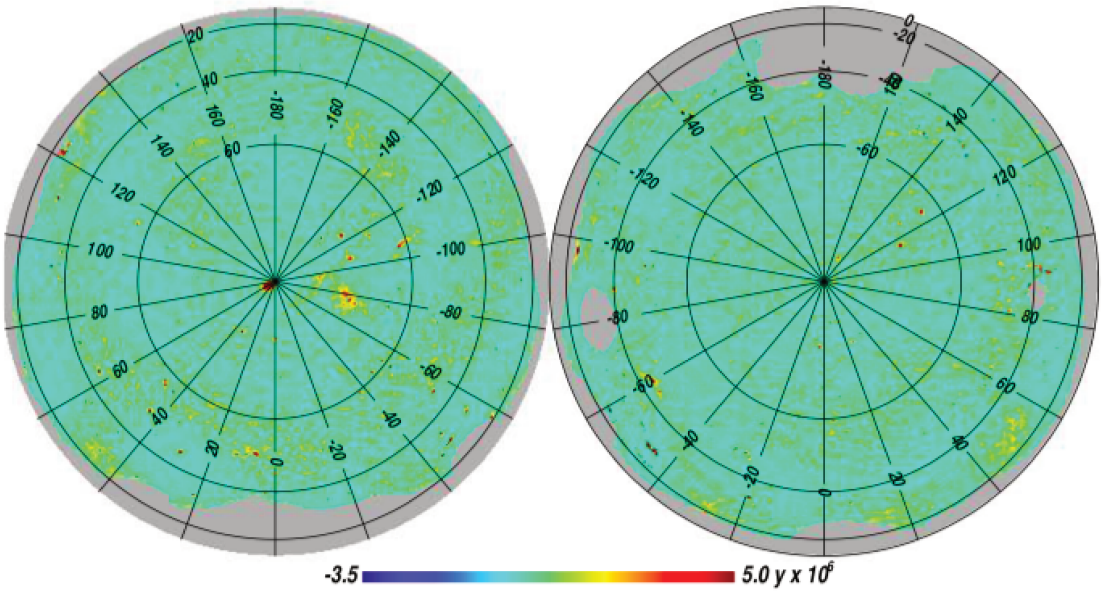

Figure 2.7: Orthographic projection of the *Planck* 2015 MILCA y-map. The northern hemisphere is on the left, with the Coma cluster in its centre, and the southern hemisphere is on the right. The figure is from Planck Collaboration et al., 2016c.

2.3.3 The *Planck* Catalogue of Compact Sources

Planck has provided the community with nine catalogues of compact sources (PCCS, Planck Collaboration et al., 2016d), detected in its nine frequency maps. The catalogues provide positions, flux estimates, and signal-to-noise ratio detections for the detected sources. As the SZ y-maps from *Planck* can be contaminated by infra-red sources or radio sources, these catalogues are very useful to mask regions of possible contamination when studying large-scale structures in the SZ map. I will use these catalogues in this purpose in Chap. 5 and in Chap. 7.

2.4 Galaxy cluster catalogues

In order to study the larger scale structures of the Cosmic Web, galaxy clusters are used as they sit at the intersection of the filaments. Rather easy to detect in different wavelengths due to their high densities, several publicly available catalogues provide samples of galaxy clusters detected in different wavelengths (e.g., in optical, in X-ray, in SZ). I present here some of these catalogues, that I have used for my work.

2.4.1 The RedMaPPer catalogue

The Red-sequence Matched-filter Probabilistic Percolation (RedMaPPer, Rykoff et al., 2014) is an algorithm developed to detect clusters in large galaxy optical surveys, such as the SDSS or the Dark Energy

Survey[12] (DES, The Dark Energy Survey Collaboration, 2005). Based on the detection of red-sequence-galaxy over-densities, the algorithm provides positions and redshift probability distributions for the detected clusters, together with membership probabilities assigned to galaxies, and a richness λ related to the number of galaxies in the clusters. Rykoff et al., 2014 have successfully applied RedMaPPer to the SDSS DR8 spectroscopic galaxies and have detected 25,325 galaxy clusters in the redshift range $0.08 < z < 0.55$, over approximately 10,500 squared degrees on the sky. The RedMaPPer catalogue has been extensively studied in different wavelengths (e.g., Saro et al., 2015; Hurier and Angulo, 2018; Geach and Peacock, 2017), allowing the confirmations of the galaxy clusters and the measurements of their properties (e.g., their masses). I show the resdhift z and the mass M_{500} distributions in Fig. 2.8, compared with other catalogues of galaxy clusters detected in different wavelengths that are presented hereafter. For this figure, I have used the scaling relation between the richness and the stellar mass M_{500} from Saro et al., 2015 to compute the mass estimation of the clusters.

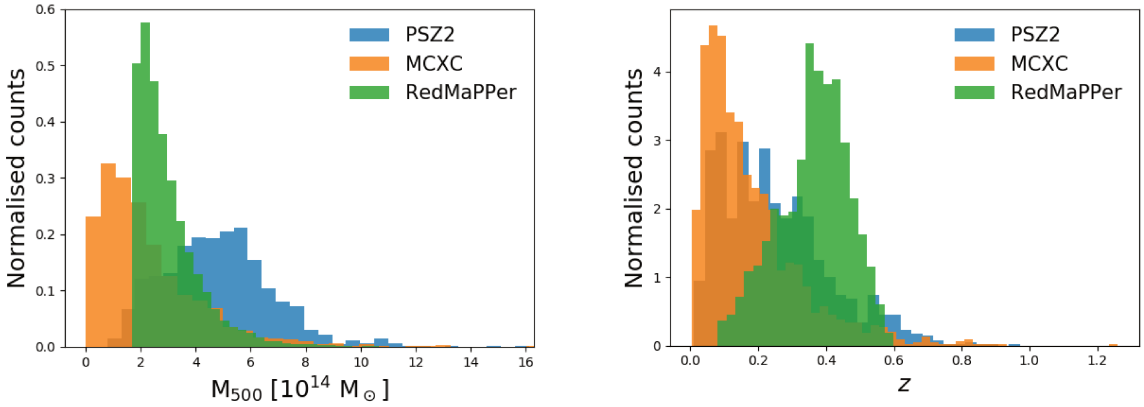

Figure 2.8: Redshift z and mass M_{500} distributions of the PSZ2, the MCXC, and the RedMaPPer catalogue of galaxy clusters.

2.4.2 SZ cluster catalogues

2.4.2.1 *Planck* PSZ2 catalogue

The *Planck* collaboration has used the six *Planck* HFI frequency maps (they have disgarded low frequency maps due to their larger beams (> 10 arcmin) and noise) with multi-match filters, filtering with generalized Navarro, Frenk and White pressure profile model (GNFW, Nagai, Vikhlinin, and Kravtsov, 2007; Arnaud et al., 2010), and taking into account the beam at each frequencies and the spectral dependance of the SZ effect. They have implemented three different algorithms for the cluster detection: two implementations of the Matched Multi-Filter (MMF1 (Herranz et al., 2002) and MMF3 (Melin, Bartlett, and Delabrouille, 2006)), and PowellSnakes (PwS Carvalho, Rocha, and Hobson, 2009; Carvalho et al., 2012). They have detected, with the union of the three methods, 1,653 galaxy cluster candidates with a signal-to-noise ratio greater than 4.5 σ (Planck Collaboration et al., 2016e). The purity of the catalogue is of 83% (Planck Collaboration et al., 2016e), leading to about 300 of false detections, which are infrared or CO residual sources passing through the spatial and spectral filters of the algorithms. For the confirmed galaxy clusters with measured redshifts, mass M_{500} are provided in the catalogue, estimated following Planck Collaboration et al., 2014. The mass and redshift distributions of the PSZ2 clusters with confirmed redshifts are shown in Fig. 2.8.

[12] `https://www.darkenergysurvey.org`

2.4.2.2 The SZ cluster database

A database of clusters detected in SZ has been developed at the Institut d'Astrophysique Spatiale, Orsay, by M. Douspis, N. Aghanim and collaborators. The database gathers clusters detected in SZ by the main CMB experiments: *Planck* (Planck Collaboration et al., 2011; Planck Collaboration et al., 2014; Planck Collaboration et al., 2015a; Planck Collaboration et al., 2016e), the Atacamy Cosmology Telescope (ACT) (Marriage et al., 2011; Hasselfield et al., 2013)), the South Pole Telescope (SPT) (Bleem et al., 2015), and clusters individually observed by the Arcminute Microkelvin Imager (AMI) and the Combined Array for Research in Millimetre-wave Astronomy (CARMA) (Schammel et al., 2013; AMI Consortium et al., 2012; AMI Consortium et al., 2013b; AMI Consortium et al., 2013a). Clusters in the SZ meta-catalogue are thus spanning a large variety of angular sizes, redshifts, and y amplitudes, quantities depending on the resolution and on the sensitivity of the experiments they were detected with. The SZ cluster database is publicly available[13], and contains 2,690 sources. For the 1,681 confirmed clusters, redshifts z and mass M_{500} are provided. The redshift range of the catalogue is $0.011 < z < 1.7$, with a median value $z_{\rm med} \sim 0.31$.

2.4.3 The MCXC catalogue

Galaxy clusters can be detected via the hot gas, detected in X-rays through Bremsstrahlung emission. The ROSAT All-Sky Survey (RASS, Truemper, 1982) is to date the only full-sky survey in X-rays (until the release of SRG/eROSITA that was successfully launched on 13th of July 2019 (Cappelluti et al., 2011)). Galaxy clusters detected based on ROSAT were combined to build a meta-catalogue: the Meta-Catalogue of X-ray detected Clusters (MCXC, Piffaretti et al., 2011). The MCXC combines galaxy clusters from RASS-based catalogues (i.e., the Northern ROSAT All-Sky Survey (NORAS, Böhringer et al., 2000), the ROSAT-ESO Flux Limited X-ray Survey (REFLEX, Böhringer et al., 2004), the ROSAT brightest cluster sample (BCS, Ebeling et al., 1998), galaxy clusters around the South Galactic Pole (SGP, Cruddace et al., 2002), galaxy clusters around the North Ecliptic Pole (NEP, Henry et al., 2006), the Massive Cluster Survey (MACS, Ebeling, Edge, and Henry, 2001), and the Clusters In the Zone of Avoidance (CIZA, Ebeling, Mullis, and Tully, 2002)), and from ROSAT serendipitous catalogues (i.e., the 160 square degree ROSAT Survey catalogue (160SD, Mullis et al., 2003), the 400 square degree ROSAT Cluster Survey catalogue (400SD, Burenin et al., 2007), the bright SHARC survey cluster catalogue (Romer et al., 2000), the Southern SHARC catalogue (Burke et al., 2003), the WARPS survey catalogues (Perlman et al., 2002; Horner et al., 2008), and the Einstein Extended Medium Sensitivity Survey catalogue (EMSS, Gioia et al., 1990)). The MCXC provides a mass M_{500}, a radius R_{500}, and a redshift z for 1,743 galaxy clusters in the all sky. The mass and redshift distributions of MCXC clusters are shown and compared to other SZ and optical catalogues in Fig. 2.8. The MCXC contains mainly lower mass clusters than the PSZ2 or the RedMaPPer cluster catalogues.

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

Chapter 3

Tools and methods

Contents

Statistical tools are nowadays strongly needed in data analysis. For example, the stacking method can be used to enhance the signals of cosmic filaments that can be very faint due to their low densities, Bayesian statistics are useful to constraint any free parameter, the bootstrap method can help to estimate the errors and/or the significances of any measurement, and the fashionable machine learning algorithms are very efficient to estimate any property without the use of a model. In this chapter, I describe some of the statistical tools, codes, or methods that I have used during my PhD thesis. These tools appear at a regular basis in the future chapters. When possible, I illustrate the applications of the methods with simple toy models.

Résumé

Aujourd'hui dans l'ère du "big data", l'astrophysique a plus que jamais besoin d'outils statistiques performants pour analyser le grand nombre de données. Par exemple, la méthode de "stacking" peut être utilisée pour aider à la détection des filaments cosmiques (pouvant présenter de très faibles signaux en raison de leur faible densité), les statistiques bayésiennes peuvent être utiles pour contraindre les paramètre libres d'un modèle, la méthode de "bootstrap" peut aider à estimer les erreurs et/ou les pertinences des mesures, et les algorithmes de "machine learning" peuvent s'avérer être très efficaces pour estimer toute propriété sans utiliser de modèle physique, avec moindre temps de calcul. Dans ce chapitre, je décris

certains des outils (codes ou méthodes statistiques) que j'ai utilisés pendant ma thèse. Ces outils apparaissent régulièrement lors des travaux présentés dans les prochains chapitres. Dans certains cas, j'illustre les applications de ces méthodes à l'aide de modèles simplistes.

3.1 Machine learning algorithms

3.1.1 Introduction

In the last decades, new statistical developments have begun to play an important role in data reduction and in data analysis. Particularly, the studies involving machine learning algorithms have increased exponentially, as such tools are very efficient to identify commonalities in data without resorting to any model. There are two families of machine learning algorithms: the unsupervised and the supervised ones. In the first case, the algorithms are designed to classify the input data. The user must assume the number of classes. This family includes clustering methods, such as k-mean algorithms, and the Self-Organizing Maps (SOM)[1]. In the second case, machine learning algorithms are designed to estimate properties or labels, based on inputs and outputs, both provided by the user. The user must in this case have a perfect knowledge of the labels or of the properties of reference used as output in the training catalogue. This family includes algorithms such as Artificial Neural Networks (ANN, W. White and Rosenblatt, 1963), Random Forests (RF, Ho, 1995), Support Vector Machine (SVM, Hearst, 1998), and algorithm of Deep Learning (DL) such as Convolutional Neural Networks (CNN, Fukushima, 1980). Machine learning algorithms, mostly supervised ones, have already been applied successfully in astronomy, astrophysics, and cosmology (e.g., Baron, 2019, for a review on machine learning algorithms in astrophysics). For example, basic machine learning algorithms, like ANN or RF, have been used to estimate galaxy redshifts or galaxy types (e.g., Bilicki et al., 2014; Bilicki et al., 2016; Krakowski et al., 2016; Siudek et al., 2018; Bonjean et al., 2019), to estimate spectral properties of sources (e.g., Ucci et al., 2018) to classify sources (e.g., Aghanim et al., 2015), to search for variable stars (e.g., Pashchenko, Sokolovsky, and Gavras, 2018), as a very non-exhaustive list of examples of applications. More sophisticated algorithms of machine learning, like DL algorithms (e.g., CNN), widely improve the results compared to results obtained with physical models. In most of the cases, the computation time required to estimate the results is also significantly reduced. For example, CNN algorithms have been already used to estimate galaxy morphologies and redshifts (e.g., Huertas-Company et al., 2015; Pasquet et al., 2019; Boucaud et al., 2019), to fit galaxy surface brightness profiles (e.g., Tuccillo et al., 2018), to compare galaxy surveys (e.g., Domínguez Sánchez et al., 2019), to detect cosmic structures (e.g., Aragon-Calvo, 2019), to learn the structure formation from initial conditions (e.g., Lucie-Smith et al., 2018; He et al., 2018), or to generate fast Cosmic Web simulations (e.g., Rodríguez et al., 2018, Ullmo et al., in prep.).

I describe here three examples of supervised machine learning algorithms that I have used for my analyses: the Random Forests, the Neural Networks, and the Convolutional Neural Networks.

I have also implemented a RF algorithm, together with other algorithms of machine learning (ANN and CNN) in python for a course of machine learning. They are implemented to be applied on a test-case dataset, called the MNIST data. These data are available in the `keras` module in python, and contains $60,000$ 2D 28×28 pixels patches, in which digital numbers are drawn. Each patch is associated with its label, between 0 and 9. These data are known to be the "Hello World !" of machine learning. Implementations, short descriptions, and examples of the different codes applied to the MNIST data are publicly available on my on-line courses of machine learning[2].

[1] `http://scikit-learn.org/`

[2] `https://github.com/vicbonj/cours_ml`

3.1.2 Caveats: bias and error estimates

Machine learning algorithms are very efficient to find correlations in the data. They always output a result, but results can sometimes be very different from the expected ones. This can be due to a bad training of the algorithm, a correlation found between unexpected features in the training catalogue that biases the training, or to an over-fitting of the training catalogue. Therefore, one of the biggest challenges in the machine learning domain is to estimate the performances of the trained algorithms (e.g., estimating the statistical errors, estimating and understand the biases). As errors on individual predictions of machine learning algorithms are usually not estimated by the algorithms, errors are often estimated statistically, over a sub-sample of the training sample. A usual way to estimate the performance and the errors of a machine learning algorithm is to split the training catalogue into three sub-samples, and train, validate, and test the algorithm on these independent samples. For example, a training catalogue is split into 80% / 10% / 10%. The 80%, called training set, are used for the training process, while other first 10%, called validation set, are used to check the results on an independent catalogue during the training process, and thus to prevent over-fitting of the training catalogue. The remaining 10%, called the test set, are independent of the training process, and are thus used to estimate the statistical errors and the biases. Retrospectively, the original sample can be split differently to ensure these percentages do not affect the results.

3.1.3 Random forests

Random Forests are machine learning algorithms based on decision tree learning (e.g., Ho, 1995, for the first implementation). Decision trees split iteratively and optimally the training set into several classes, by associating classes to data points and by reducing the Gini impurity[3]:

$$G = \sum_{i=1}^{C} p(i)(1 - p(i)), \tag{3.1}$$

where C is the number of classes, and $p(i)$) is the probability of picking a data point with class i.

In practice, simple if-else rules on the input features are defined, in order to classify the training set at each splitting, so that each class are equally distributed as a function of the outputs commonalities.

RF algorithms then use the mean estimator of a "forest" of decision trees, trained by bootstrapping the training set. For a training set of n samples, with $X = x_1, ..., x_n$ and $Y = y_1, ..., y_n$ being the inputs and the outputs of the machine learning, respectively, the estimator for an untrained value x' is computed as follows:

$$\tilde{y}\,(x') = \frac{1}{M} \sum_{m=1}^{M} \tilde{y}_m\,(x')\,, \tag{3.2}$$

where M is the number of decision trees, and $\tilde{y}_m$ is the estimator for x' of the decision tree m trained on a random sample with replacement of n elements in the sample of couples (X, Y).

To optimise the training and obtain the best results, some parameters have to be set, such as the number of trees, M, or the maximum depth of the trees (i.e. the maximum number of splitting), $d_{\max}$. These parameters can be set by training the RF on the training set varying the values of M and $d_{\max}$, and by comparing the scores of the RF on the validation set. The best optimised parameters can be set to the ones providing the best score. This method prevents over-fitting.

The RF algorithms are rather easy to understand, and very efficient. In addition, unlike with ANN (described in the next section), no data pre-processing is needed, meaning that the values themselves are

[3] `https://scikit-learn.org/stable/modules/tree.html#classification-criteria`

useful and play a role during the training process. For these reasons, I will use the RF to estimate galaxy properties (SFR and $M_\star$) with machine learning (presented in Chap. 4).

3.1.4 Artificial Neural networks

While very efficient for simple cases, RF are nevertheless limited when the input data are too complex (e.g., learning very high non-linear relations, or using inputs with high dimensions such as images or data cubes). Artificial Neural Networks, also called Multi-Layer Perceptrons (e.g., W. White and Rosenblatt, 1963, for the first implementation) often perform equivalent or better results. ANN are probably the most familiar kind of algorithms used in machine learning, and also the most known. Their principle is inspired by the nature of our brain, composed of "neurons", that are connected and exchanging informations with each other. ANN estimate a prediction (performing classification or regression), by transforming the inputs in successive layers of linear combinations, weighted in the neurons.

In practice, a neuron is a function that linearly combines inputs $\vec{x}$ with weights $\vec{W}$, adds a bias b (different for each neuron to maximise the variety of the information learned), and passes the output into an activation function g, that quantifies the significance of the information learned. It is mathematically written as follows (a schematic view of a neuron is also shown in Fig. 3.1):

$$f(\vec{x}) = g(\vec{W}.\vec{x} + b). \tag{3.3}$$

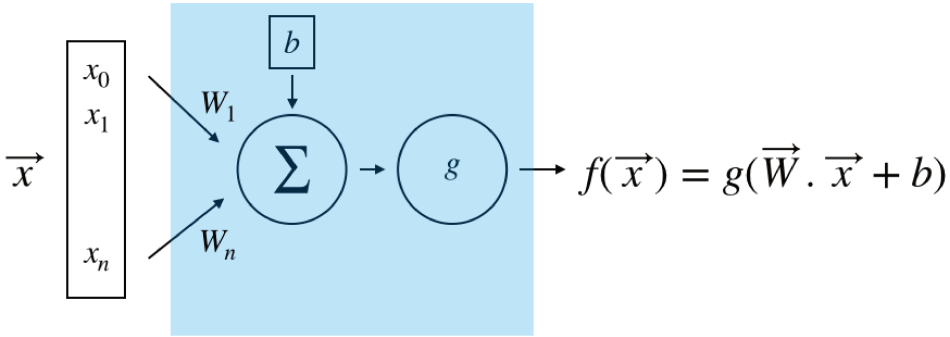

Figure 3.1: Illustration of a neuron in a neural network algorithm. The neuron is the shaded blue region. The x_i are the inputs, the W_i are the weights, b is the bias, and g is the activation function.

There are many different activation functions in the literature (ReLU, Sigmoid, TanH, Softmax, and so on), each of them showing advantages and disadvantages. The choice of the activation function is very dependent of the cases. In DL, the REctified Linear Unit (ReLU), that is $g(x) = \max(0, x)$, is commonly used. The main advantages of the ReLU activation function are the high and non-bounded values that it can reach, allowing a more efficient training for gradient-based learning (contrary to sigmoid activation function where the gradient becomes very small when x increases). Another advantage of the ReLU is a more sparse information learned (allowing the "disconnections" of the neurons in the case where $g(x) < 0$).

During the training process, the weights $\vec{W}$ are learned iteratively so that a loss function that minimises the errors between $f(\vec{x})$ and the outputs $\vec{y}$ is minimised. The errors are back-propagated based on gradient descent. Several functions are used in the literature to optimise the back-propagation of the errors (e.g., Adam (Kingma and Ba, 2014), Stochastic Gradient Descent, Adadelta (Zeiler, 2012)).

The Adam optimiser is commonly use in DL algorithms. It is very efficient as it adapts iteratively the precision of the errors.

One of the main advantages of the neurons is that they can be modulated, to maximise the optimisation of the loss function and learn very high non-linear models. Neurons can be disposed in "layers". In that case, the weights $\vec{W}$ become a matrix $\mathbf{W}$, and the bias b becomes a vector, $\vec{b}$. The biases b_i of the i neurons are different to ensure the independence of the features learned by the neurons. The output of a layer is a vector $f(\vec{x})$, mathematically written as follows:

$$f(\vec{x}) = g(\mathbf{W}.\vec{x} + \vec{b}). \tag{3.4}$$

There is no limit in the number of neurons in a layer, and layers can be themselves superposed to increase the non-linear character of the models in case of very complex data. ANN algorithms with a large number of layers are part of Deep Learning (DL) algorithms. In DL, the outputs of the first layer, $f(\vec{x})$, is taken as inputs of the second layer, and so on, with as many layers of as many neurons as wanted. In some cases (e.g., classifiers or segmentations), when a probability is wanted as output of the network, the last activation function after the last layer is chosen to rescale the value between 0 and 1. The Sigmoid is an example of such activation function:

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}}. \tag{3.5}$$

To summarise mathematically, for an ANN composed of three layers (indexed by 1, 2 and 3) of several neurons each with a Relu activation function, and a Sigmoid last activation function, the outputs of the network that are compared with the outputs $\vec{y}$ are written:

$$f(\vec{x}) = \text{Sigmoid}_3(\mathbf{W_3}.\text{ReLU}_2(\mathbf{W_1}.\text{ReLU}_1(\mathbf{W_2}.\vec{x} + \vec{b_1}) + \vec{b_2}) + \vec{b_3}). \tag{3.6}$$

The deeper the algorithm, the more complex the model is, and the better it learns. However, if the model is too deep for a too simple case, the model may start to learn the statistical noise of the training catalogue (this is called over-fitting). To prevent this effect, results are estimated iteratively on the training set and on the validation set. The training can be stopped when the results on the validation set start to decrease as compared to the results on the training set: this is called the early stopping. Another way of preventing over-fitting is to include dropout layers (Srivastava et al., 2014). Dropout layers randomly ignore a percentage (set by the user) of neurons in the network. This technique approximates a large number of networks with different architectures working together in parallel, improving the results. To increase the performance and the results, parameters can also be tuned. The main ones are the number of layers, and the number of neurons in each layer.

I will use ANN to confirm the results obtained with the RF to estimate SFR and $M_\star$ of nearby galaxies (presented in Chap. 4).

3.1.5 Convolutional Neural Networks

ANN and deep ANN algorithms are ideal to learn very highly non-linear features in very large sized input data, like images. However, when the aim is to analyse spatial coherent objects in input images (e.g., segmentation, detection of sources), the features to capture can be invariant in translation and in rotation. That can be for example the case for the SZ emission generated by the hot gas in galaxy clusters seen in the *Planck* frequency maps. In that case, successive convolutional layers can be applied to the images, to extract the relevant spatial information that is encoded and injected into an ANN (or decoded

with de-convolutional layers in the case of segmentation map reconstructions). These algorithms are called Convolutional Neural Networks (e.g., Fukushima, 1980, for the first implementation).

The principle of the CNN is almost the same than for the NN, except that the neurons are defined differently; they are convolutions (compared to linear combinations). The weights $\vec{W}$ learned during the training are the weights of the convolutional kernels. As for ANN, the outputs of a layer become the inputs of the next layer. By superposing convolutional layers, the typical sizes of the features learned at each layer increase. The encoded signal at the end of the convolutional layers thus contains information at different scales, allowing the learning of very complex objects. Between convolutional layers, MaxPool layers are often added, that reduce by a factor 2 the dimensions of the convolved images by keeping the maximum values of the pixels in (2×2) pixel kernels. The MaxPool layers decrease the number of free parameters, and thus increase efficiently the duration of the training. A schematic view of a CNN is shown in Fig. 3.2, taken from `http://cs231n.github.io/convolutional-networks/`. To optimise the performance and prevent over-fitting, more parameters than in the case of ANN have to be tuned, like the size of the convolutional kernels, or the number of filters.

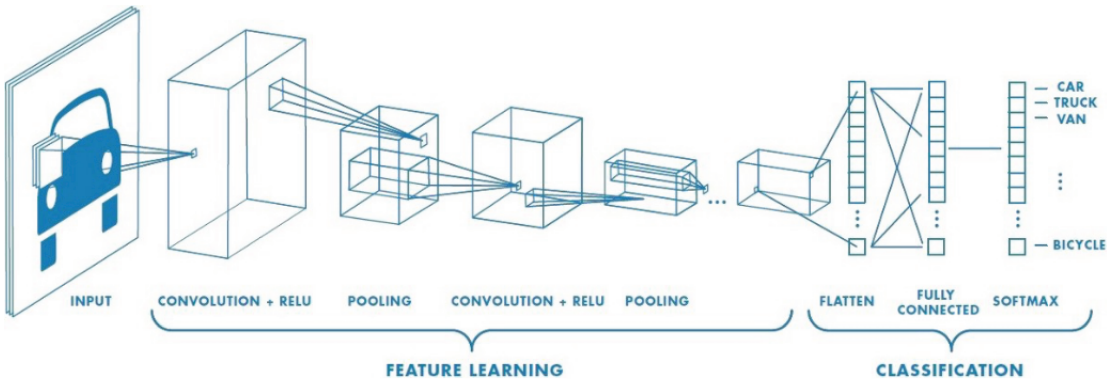

Figure 3.2: Illustration of a CNN classifier, where the inputs is an image of a car. Figure is taken from this website: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53. The convolutional layers encode the information of the spatial features at different scales. The encoded information is injected in an ANN, that learns to classify the images. The last activation function, a sigmoid, gives a probability of belonging to a class (in that case, a car).

I will present how I have used CNN to detect the SZ signal in the *Planck* frequency maps in Chap. 7.

3.2 Statistical methods

3.2.1 MCMC

While machine learning algorithms are very efficient at estimating non-linear properties without resorting to models, these techniques are less efficient at estimating errors on derived quantities. However, when a model is physically motivated to analyse data, another way of estimating (or fitting) properties in a more traditional way is the Monte Carlo Markov Chain (MCMC, Metropolis et al., 1953; Hastings, 1970). Such algorithms are very powerful when computing non-linear models, with numerous parameters with potential degeneracies. In those cases, basic algorithms, like the least square minimisation, only return the parameters that minimise a loss function (in the example of the least square minimisation, the minimisation of the sum of the squared errors). The MCMC goes further and returns a chain of

parameters, allowing the study of the statistics of the derived parameters (e.g., their distributions, their degeneracies). The MCMC samples the probabilities of the parameters as a function of the prior distributions (given by the user), and of a likelihood loss function that is computed based on the differences between the estimated models and the data. MCMC algorithms have already been successfully applied in astrophysics and cosmology, e.g., to estimate the cosmological parameters with *Planck* (e.g., Planck Collaboration et al., 2016b), to compute comet orbits (e.g., Lang and Hogg, 2012), or to compute dust properties in warm debris disks (e.g., Olofsson et al., 2012).

MCMC uses the Metropolis algorithm to generate a chain of parameters. It works as follows. Given any data y, its error σ, and a model $f(x, a)$, where a is the parameter to fit, the required output is the posterior distribution:

$$p(a \mid y, x, \sigma) = p(a) \times p(y \mid x, \sigma, a), \tag{3.7}$$

where $p(y \mid x, \sigma, a)$ is the likelihood function computed based on the model f:

$$\mathcal{L}(a) = p(y \mid x, \sigma, a) \propto -\frac{1}{2} \sum_n \frac{(y_n - f(x_n, a))^2}{\sigma_n^2}, \tag{3.8}$$

and $p(a)$ is the prior distribution of the parameter a, usually a uniform or a normal law, between the fixed boundary conditions $[a_{\min}, a_{\max}]$. Then, the chain of parameters is constructed following these steps:

- Initialise the chain with an arbitrary point a_0.
- For each iteration i, pick up a parameter a_{i+1} that follows the prior distribution $p(a)$.
- Compare the likelihood obtained with a_{i+1} with the one obtained with a_i: $\beta = \mathcal{L}(a_{i+1})/\mathcal{L}(a_i)$.
- Generate a uniform random number $u \in [0, 1]$:
  - if $u \leq \beta$, accept a_{i+1} in the chain,
  - else $u > \beta$, conserve the previous value in the chain: $a_{i+1} = a_i$.

The chain converges to the maximum of likelihood, accepting at each iteration a new parameter in the chain if the likelihood is higher than the previous one, and rejecting with a non zero probability parameters with lower likelihoods. The chain contains the probability distribution of the fitted parameters.

The publicly available `emcee` implementation of the MCMC algorithm in python is very efficient in terms of computation time (Foreman-Mackey et al., 2013). The `emcee` package uses an optimised algorithm, the Affine Invariant MCMC Ensemble Sampler, described in detail in Goodman and Weare, 2010. This tool has been used at a regular basis in astronomy and in cosmology. I will use the `emcee` package in my analysis of the physical properties of the gas in the bridge of matter between the galaxy cluster pairs A399-A401 and A21-PSZ2 G114.90-34.35 (Bonjean et al., 2018), presented in Chap. 5. I will also use it to fit the quiescent fraction profile around cosmic filaments, presented in Chap. 6.5.2.

3.2.2 Stacking

When the detections of individual objects are not possibles due to their very faint signals (e.g., low densities filaments in galaxy density maps, or diffuse gas in SZ *Planck* maps), an averaged characterisation of all the objects is possible by stacking. Stacking is equivalent to averaging signals individually measured on the data. By doing so, the noise of the data is reduced by $\sqrt{n}$ (where n is the number of considered objects), and the average signal is enhanced. This technique has already been succesfully used for example

in Tanimura et al., 2019b, de Graaff et al., 2019, and Tanimura et al., 2019a, to find SZ signal between galaxy pairs and in super-clusters.

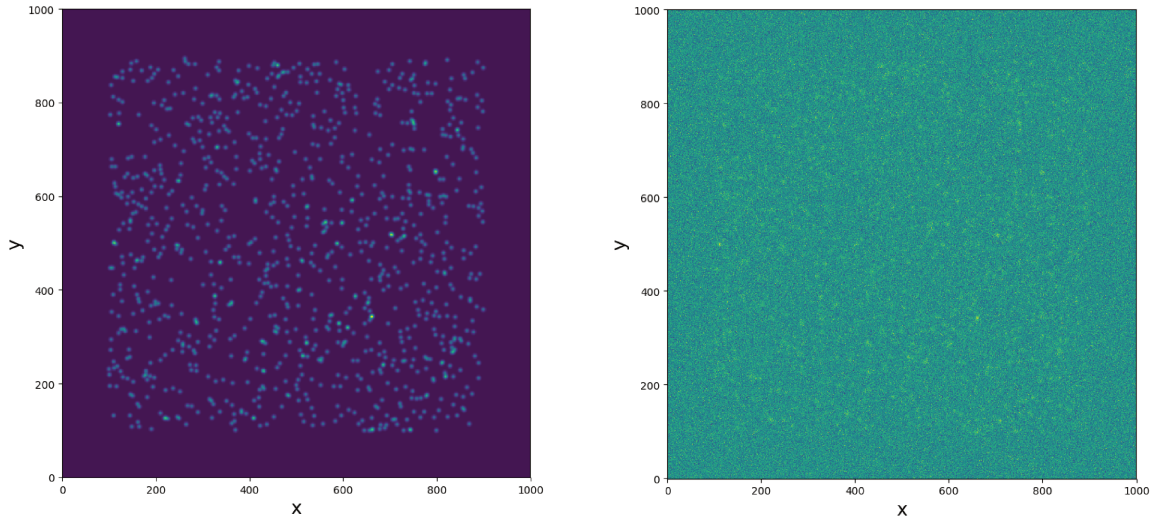

Figure 3.3: Illustration of the stacking method. Left: 1000 Gaussians injected in a 1000×1000 pixels map. 2D Gaussians have the same standard deviations $\sigma = 3$ and amplitudes $A = 1$ (in arbitrary units). Right: a Gaussian noise is added on the map with a standard deviation of $\sigma_{\rm noise} = A$.

To illustrate the "power" of stacking, I present a simple toy model, shown in Fig. 3.3 and Fig. 3.4. I have generated 1000 2D Gaussians uniformly distributed in a 1000×1000 pixels map (in the left panel of Fig. 3.3), with same standard deviations, $\sigma = 3$ pixels, and same amplitudes, $A = 1$ (in arbitrary units). The Gaussians are distributed between 0.1 and 0.9 times the size of the map, to ensure that all the sources are entirely seen in the map. I have injected a Gaussian noise on the map, following a normal law with a mean $m = 0$ and a standard deviation $\sigma_{\rm noise} = A$, so that the 1000 Gaussians are almost completely hidden in the noise (in the right panel of Fig. 3.3). Then, knowing their positions, I have measured the radial profiles around the 1000 2D Gaussians as a function of the distance θ to the centres, and I have stacked all the measured signals by averaging the 1000 profiles. I show in Fig. 3.4 the stacked radial profile and the model of the 2D Gaussians model input in the map with the mean value of the map added. Despite the very high noise input in the map, a very nice agreement is seen between the input model and the stacked measurement. The errorbars of the stacked profile have been computed with another statistical tool, the bootstrap, that I describe in the next section.

I will use the stacking method in Chap. 6 and in Chap. 7, to study galaxy properties around cosmic filaments and to probe the statistical properties of a sample of galaxy clusters detected in the *Planck* frequency maps via the SZ effect with a deep learning algorithm.

3.2.3 Bootstrap

Not only it is important to measure a signal, but it is even more important to estimate the error bar which enables to compute its significance. In statistics, the bootstrap technique is a way of estimating the distribution of any measurement on a sample, when the distribution is a priori not known. It uses random re-sampling with replacement to generate new different samples containing data from the "father" sam-

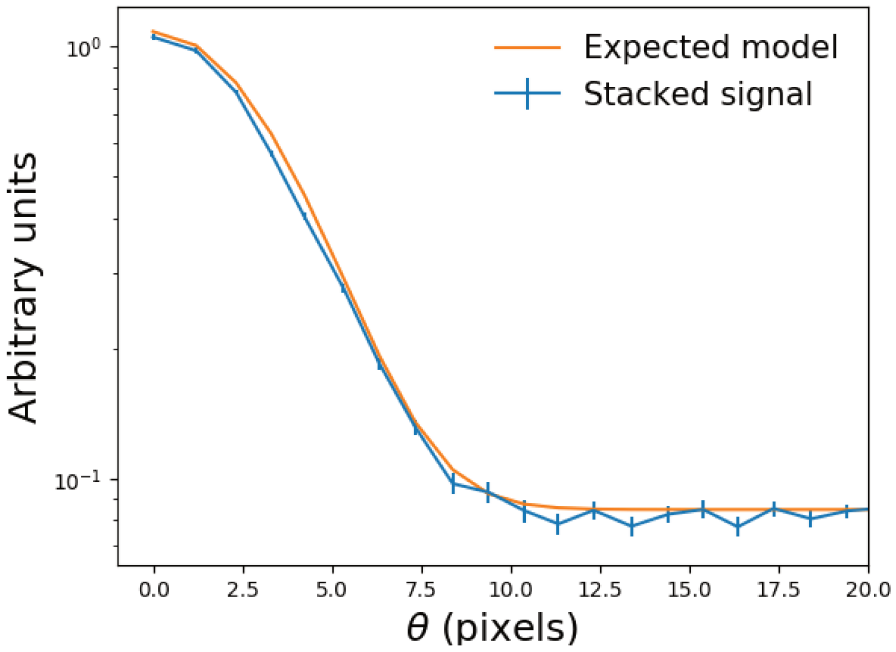

Figure 3.4: Stacked measured profile in blue and input model (where the mean value of the map was added) in orange of the 2D Gaussians shown in Fig. 3.3. Errors on the stacked measured profiles were computed with bootstrap.

ple. Then, measuring a desired quantity on the "re-sampled samples" provide a distribution rather than a single value. That way an error can be estimated. As stacking use a sample of data and provide only one stacked value, the bootstrap method is well suited to estimate the errors on stacked properties.

I illustrate hereafter an application of the bootstrap with a detailed example. I assume that the property of interest is the mean value, and its error bar, of a Gaussian distribution of 1000 points, centred on 0 with a standard deviation $\sigma = 1$. I have generated a sample A of $n = 1000$ points a_i following a Gaussian distribution. The histogram of the points is shown in the left panel of Fig. 3.5. The mean value of this sample can be estimated by using the standard formula: $\bar{A} = \frac{1}{n}\sum_{i=1}^{n} a_i$.

The mean value of the distribution shown in the left panel of Fig. 3.5 is $\bar{A} = -0.00052$, very close to theoretical mean value that is 0. To estimate an error bar on this measure, I have used the bootstrap technique; I have constructed 1000 new samples of 1000 points randomly picked, with replacement, in the sample A (i.e., one point can be picked several times). I have then estimated the mean values of the 1000 samples. Now, a sample B containing the 1000 mean values b_i has been constructed and its distribution is shown in the right panel of Fig. 3.5. The mean and the standard deviation of the mean values can be estimated with this distribution. In this case, I have obtained $\bar{B} = 0.00002$, and $\sigma_B = 0.03078$. The measurement is then $\bar{A} = 0.00002 \pm 0.03078$, fully consistent with the theoretical mean value 0.

In the following, I will use the bootstrap technique to estimate the error bars on the stacked galaxy profiles around cosmic filaments presented in Chap. 6.

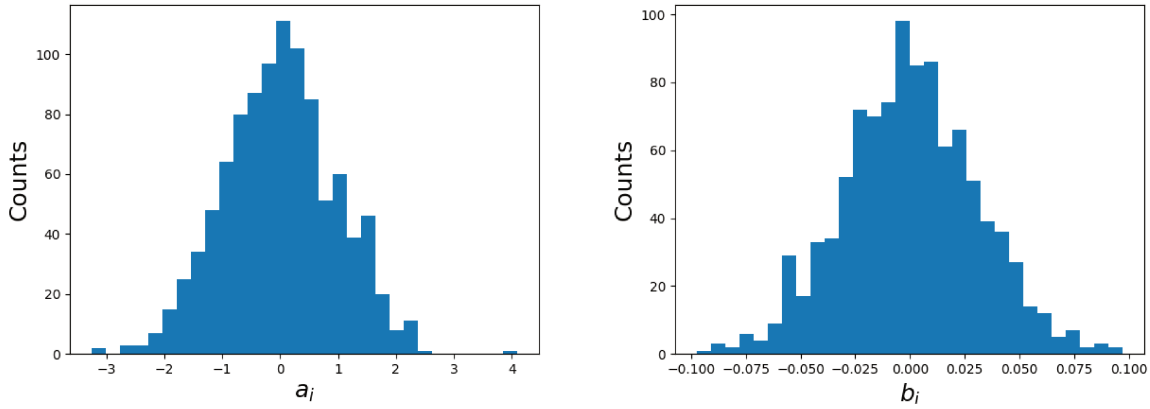

Figure 3.5: Illustration of the bootstrap technique used to compute the mean of a sample A of 1000 points from a normal distribution with a theoretical mean $\bar{A} = 0$. Left: the histogram of the sample A. Right: the distribution of $\bar{B}$ obtained by bootstrapping the sample A 1000 times is shown.

3.3 Developed or adapted codes

3.3.1 pyDTFE

The distribution of galaxies is the best way to capture the large-scale structures in the Universe, and thus the best way to detect them. As presented in Chap. 1.2.2, structures can be described and detected through their shapes and their over-densities. Therefore, estimating the density field in the Universe is a key step for studying the large-scale structures. The Delaunay Tessellation Field Estimator (DTFE, Schaap and van de Weygaert, 2000) has been especially developed to reconstruct the density field in the Cosmic Web. It is widely used to study the Cosmic Web (e.g., Bernardeau and van de Weygaert, 1996; Schaap and van de Weygaert, 2000; Sousbie, 2011; Cautun et al., 2015). The method automatically adapts to large density gradients which is the case for the elements in the Cosmic Web that span a wide range of densities (see Chap. 1.2.2). Furthermore, the method does not depend on any parameters as is the case for other approaches such as the k-nearest neighbours, where the density fields depend on the chosen parameter k. The DTFE bases the computation of the density in each point on the connected cells of the Delaunay Tessellation, that traces triangles (in 2D, or tetrahedra in 3D) between each points and thus takes into account the environment at different scales.

I describe hereafter in detail the different steps of the DTFE, referring step by step to an example of a DTFE application on a sample of 1000 uniformly distributed points (shown in Fig. 3.6). The DTFE is composed of three main steps:

- Based on a sample of discrete points in Cartesian space (upper left panel of Fig. 3.6), a Delaunay Tessellation is computed, i.e., space is divided into triangles (or tetrahedra in 3D) whose vertices are formed by the points of the sample (upper right panel of Fig. 3.6).

- Based on the triangles (or tetrahedra) of the Delaunay Tessellation, an estimate of the density is computed at each point based on the local density. The density is defined as one over the sum of the areas of the n triangles connected to a point k (or of the volumes of the n tetrahedra). To take into account the fact that any area (or volume) is seen $d + 1$ times, where d is the dimension (i.e., 2D or 3D) of the sample, the areas (or volumes) are also divided by $d + 1$:

$$\rho_k = \frac{d+1}{\sum_{i=1}^{n} A_{ki}}, \tag{3.9}$$

where ρ_k is the density at the point k, and A_{ki} are the areas (or volumes) of the triangles (or tetrahedra) connected to the point k.

- The densities estimated with Eq. 3.9 are linearly interpolated on a regular grid, and the density field is reconstructed (bottom panel of Fig. 3.6).

I have developed a python implementation of the DTFE, the `pyDTFE`, publicly available on GitHub[4]. I will use it to compute the full sky galaxy density maps detailed in Chap. 4.6.3. These maps were used to characterise cosmic filaments and clusters (Chap. 6 and Chap. 7). I will also use the `pyDTFE` code to detect galaxy over-densities in between the galaxy cluster pair A399-A401 (Chap. 5.4.3).

3.3.2 RadFil

To derive galaxy properties around cosmic filaments in Chap. 6, I have used the code developed by Zucker and Chen, 2018: RadFil[5]. It is a code that measures radial profiles around filamentary structures, using interpolation of filament' spines and their first derivatives at each point. RadFil was originally developed to study inter-stellar filaments (e.g., Zucker, Battersby, and Goodman, 2018). I have optimised the code to apply it to large-scale structure cosmic filaments. The advantage of the RadFil code is that knowing the positions of the filaments, a profile can be measured in any 2D python array.

The RadFil code can be used in two ways: in the first option (not used here), the user provides the code with a single python 2D array of a given observable, the code detects the skeleton of the filament, if any, and outputs the measured profiles around the filament. In the second option, the user provides the code with two python 2D arrays, one being the observable (the pixel values representing the quantity to measure), and the other one being a mask tracing the spine of the filament around which RadFil will measure the profiles. I use RadFil in Chap. 6 with the second option for two main reasons. First, the cosmic filaments have very low signal-to-noise ratio observables compared to inter-stellar filaments, preventing the RadFil code for detecting them. Second, optimised methods to extract cosmic filaments in the Cosmic Web have already been successfully applied in galaxy surveys (e.g., Libeskind et al., 2018 for a review, Bonnaire et al. in prep.). Positions of filamentary Cosmic Web structures are thus already known so they can be directly input in the code. In my work, I have used the catalogue of cosmic filaments extracted by DisPerSE in a SDSS spectroscopic sample of galaxies that is presented in Chap. 6.2.1.

An illustration of RadFil is shown hereafter on a single toy model.

I have modelled a filament spine in a map of size 1000×1000 pixels, that I smoothed with a Gaussian filter of $\sigma = 30$ pixel width (top line of Fig. 3.7). I have measured the profile of the smoothed filament around the spine of the modelled filament with RadFil (bottom line of Fig. 3.7). I have fitted a Gaussian on the measured average profile to compare the measured and the theoretical filament's width. The standard deviation of the Gaussian measured by the fit is $\sigma_{\rm m} = 30.07 \pm 0.03$ pixels, that is very close to the theoretical one, $\sigma = 30$ pixels.

I will use RadFil to measure the galaxy properties around cosmic filaments based on density maps (presented in Chap. 6). I will also measure SZ profiles in the *Planck* SZ MILCA map, to compare RadFil with other methods based on Healpix projections, presented in Tanimura et al., in prep.

[4] `https://github.com/vicbonj/pydtfe`
[5] `https://github.com/catherinezucker/radfil`

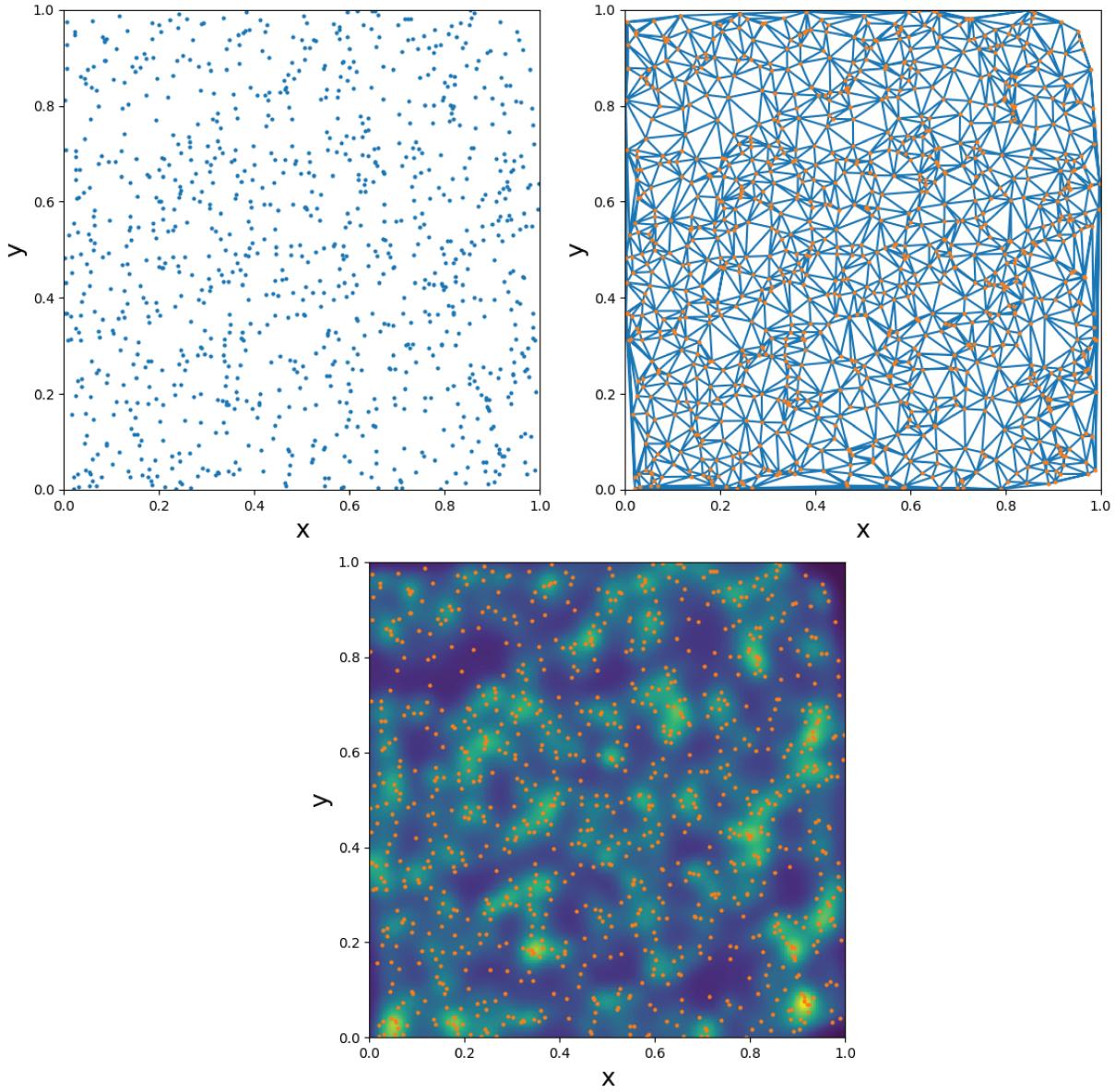

Figure 3.6: Application of the DTFE code to a sample of 1000 uniformly distributed random points. Top left: the spatial distribution of the 1000 random points. Top right: the Delaunay Tessellation of the points. Bottom: densities estimated with Delaunay triangles have been interpolated on a grid. Yellow pixels show over-dense regions while blue pixels show under-dense regions. The field has been smoothed for visualisation. The initial points are overlaid in orange.

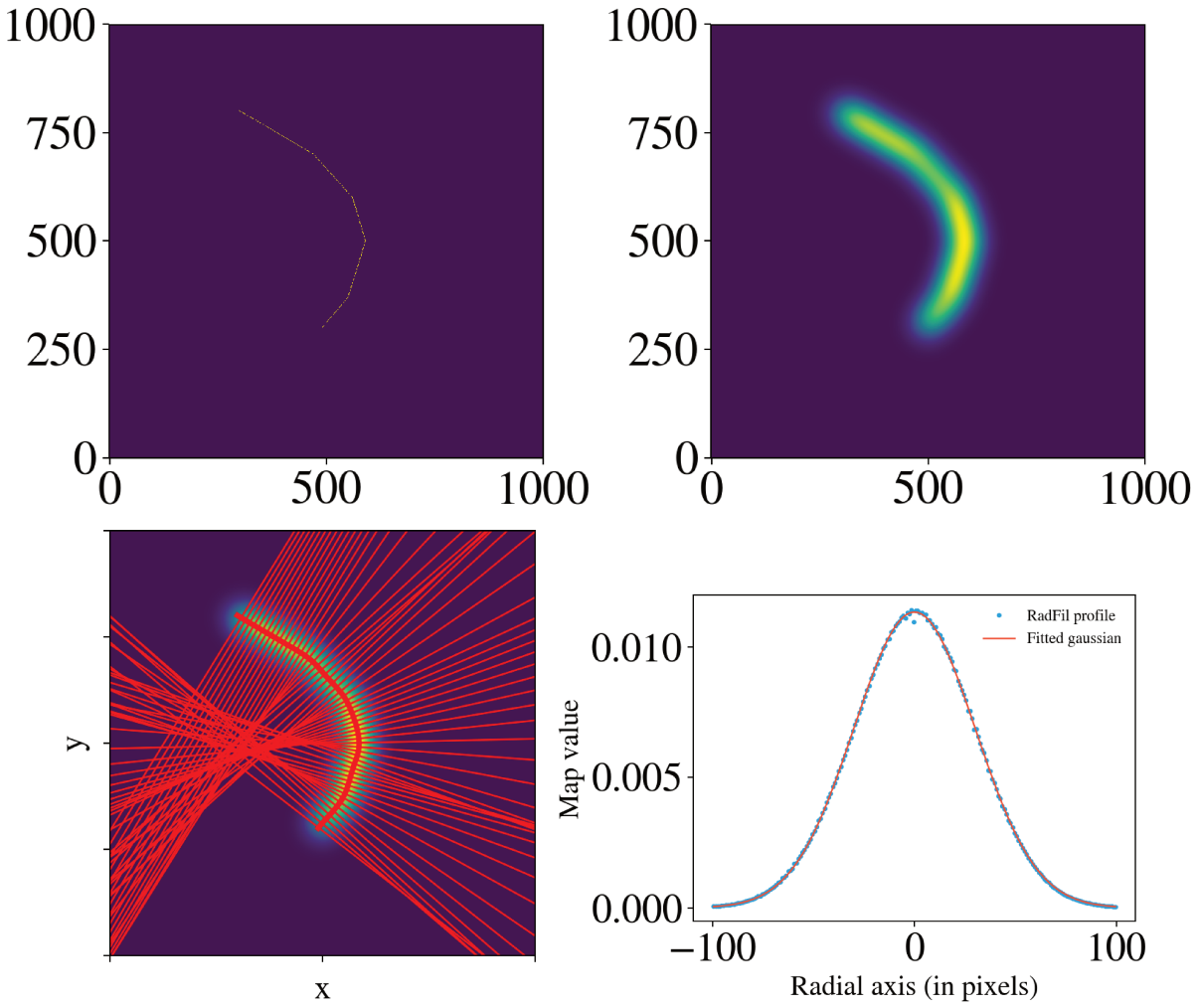

Figure 3.7: Top left: model of filament spine in a 1000×1000 pixels map. Top right: the filament is smoothed with a Gaussian of $\sigma = 30$ pixels width. Bottom left: illustration of RadFIl: the code measures the profiles in each red lines orthogonal to the filament spine. Bottom right: the average radial profiles measured by RadFil. A Gaussian has been fitted to the profile, with a measure width $\sigma_{\rm m} = 30.07 \pm 0.03$, in agreement with the theoretical one, i.e., $\sigma = 30$ pixels.

Part II

Cosmic Web analysis

Chapter 4

Galaxy types with machine learning

Contents

Abstract

Star-formation activity is a key property to probe the structure formation and hence characterise the large-scale structures of the Universe. This information can be deduced from the star formation rate (SFR) and the stellar mass ($M_\star$). Determining these quantities from UV, optical, or IR luminosities relies on complex modelling and on priors on galaxy types. In this chapter, I have developed a method based on the machine-learning algorithm Random Forest to estimate the SFR and the $M_\star$ of galaxies in the redshift range $0.01 < z < 0.3$, independent of their type. The machine-learning algorithm takes as inputs the redshifts, WISE luminosities, and WISE colours in near-IR, and is trained on SFR and $M_\star$ from the SDSS MPA-JHU DR8 catalogue. The trained RF is unbiased with respect to redshift or galaxy type, and it can accurately estimate SFR and $M_\star$ with scatters of $\sigma_{\rm SFR} = 0.38$ dex and $\sigma_{M_\star} = 0.16$ dex for SFR and stellar mass, respectively. The full-sky coverage of the WISE satellite allows us to characterise the star-formation activity of all galaxies outside the Galactic mask with spectroscopic redshifts in the range $0.01 < z < 0.3$. The RF model can also be applied to photometric-redshift catalogues, with best scatters of $\sigma_{\rm SFR} = 0.42$ dex and $\sigma_{M_\star} = 0.24$ dex obtained in the redshift range $0.1 < z < 0.3$. I have thus applied the machine learning algorithm to construct a value-added catalogue based on the WISExSCOS catalogue of photometric redshifts. The value-added catalogue is further used to study the statistical properties of galaxies around cosmic filaments (in Chap. 6), and in other studies discussed in Chap. 4.8. This chapter uses material from *"Star formation rate and stellar masses from machine learning"*, V. Bonjean, N. Aghanim, P. Salomé, A. Beelen, M. Douspis, and E. Soubrié, 2019, A&A, 622, A137.

Résumé

L'activité de formation d'étoiles d'une galaxie est une propriété clé pour sonder la formation des structures, et ainsi caractériser les structures à grande échelle présentes dans l'Univers. Cette information peut être déduite du taux de formation d'étoiles (SFR) et de la masse stellaire ($M_\star$). La détermination de ces deux grandeurs à partir de luminositiés en UV, en optique, ou en IR, repose sur une modélisation complexe et sur des connaissances a priori du types de galaxies. Dans ce chapitre, je présente une méthode que j'ai développé basée sur un algorithme de machine learning (Random Forest) pour estimer SFR et $M_\star$ des galaxies dans la plage de redshift $0,01 < z < 0,3$, indépendamment de leur type. L'algorithme prend en entrée le redshift, les luminosités et les couleurs de WISE en proche IR, et est basé sur les SFR et $M_\star$ du catalogue SDSS MPA-JHU DR8. Le modèle généré est non biasé en ce qui concerne le redshift ou le type de galaxie, et peut estimer SFR et $M_\star$ avec des erreurs statistiques de $\sigma_{\rm SFR} = 0,38$ dex et de $\sigma_{M_\star} = 0,16$ dex respectivement. La couverture complète du ciel du satellite WISE permet la caractériserisation de l'activité de formation d'étoiles de toutes les galaxies à l'extérieur du masque galactique (avec $0,01 < z < 0,3$). Le modèle de RF peut également être appliqué aux catalogues de galaxies avec redshifts photométriques, avec des erreurs statistiques de $\sigma_{\rm SFR} = 0,42$ dex et $\sigma_{M_\star} = 0,24$ dex. J'ai appliqué cette méthode pour construire un catalogue à valeur ajoutée basé sur le catalogue avec redshifts photométriques : le WISExSCOS catalogue. Le catalogue à valeur ajoutée généré est notamment utilisé pour étudier les propriétés statistiques des galaxies autour des filaments cosmiques (présenté au Chap. 6), et dans une autre étude présentée au Chap. 4.8. Ce chapitre utilise du matériel du papier *"Star formation rate and stellar masses from machine learning"*, V. Bonjean, N. Aghanim, P. Salomé, A. Beelen, M. Douspis, et E. Soubrié, 2019, A&A, 622, A137.

4.1 Introduction

Several methods have been developed to estimate SFR and $M_\star$ from infra-red or optical bands (e.g., Calzetti et al., 2007; Kennicutt et al., 2009; Jarrett et al., 2013; Cluver et al., 2014; Cluver et al., 2017). All of these relations are well calibrated. However, as passive and active galaxies have very different

colours in IR, applying these methods to galaxies without having any prior on their types can lead to potential biases. Ideally, optical spectroscopic data are needed to estimate the SFR and $M_\star$ properties, but they are not always available as they are costly in terms of observing time.

In this chapter, an alternative approach is chosen to estimate SFR and $M_\star$ for all galaxies over 70% of the sky (i.e., outside the Galactic plane) with measured redshifts in the range $0 < z < 0.3$, without any priors on galaxy types. To do so, a machine learning algorithm, i.e., a random forest (see Chap. 3.1.3), heve been used. Such a method is able to estimate very non-linear laws based on models trained on reliable given inputs and outputs. In the present case, it allows us to estimate SFR and $M_\star$ independently of any complex model nor any priors on galaxy types. Very recently, Delli Veneri et al., 2019 have used a similar approach to derive SFR based on optical colours from the SDSS photometric survey on a limited portion of the sky. Here, the random forest algorithm is trained on WISE infra-red magnitudes, allowing an application on the full usable sky (masked from our galaxy).

4.2 Constructing a training set

Ensuring good, that is, unbiased, training of the machine learning algorithm, the choice of inputs and reference outputs is essential.

4.2.1 The choice of inputs and outputs

First of all, the inputs have to be defined, that is, the data that are proxy to estimate the SFR and the $M_\star$. Motivated by its full-sky coverage and its very large number of sources (more than 700,000,000 sources), I have chosen the WISE infrared data as inputs, namely the AllWISE Source Catalogue (see Chap. 2.2.1). As the SFR can evolve with redshift, I have also chosen to use the redshift z as input. As a proxy for the stellar mass, I have chosen the WISE luminosity in the W1 band (3.4μm), that traces the old non-ionizing stars (Wen et al., 2013; Jarrett et al., 2013). As a proxy for the SFR, I have chosen the WISE luminosity in the W3 band (12μm), that traces the emission from small grains of dust, thus directly related to the total quantity of dust, itself indirectly related to the SFR (re-emission of the UV from young stars absorbed by the dust) (Jarrett et al., 2013; Cluver et al., 2014; Cluver et al., 2017). I have chosen not to consider the W4 band of WISE although it is a good tracer of the SFR (Jarrett et al., 2013; Cluver et al., 2014; Cluver et al., 2017), as Bilicki et al., 2016 found that its larger beam size (of 12") and its poorer sensitivity could lead to an important incompleteness and a significant bias of source selection with respect to redshift. The aim of my study being to estimate the SFR and the stellar mass for both galaxy types (active and passive), without any prior, I have also chosen as input two WISE colours to segregate the galaxy types: W1-W2 ($3.4 - 4.6\mu$m) and W2-W3 ($4.6 - 12\mu$m). Wright et al., 2010 have found that these two colours are very efficient to segregate morphologies, and thus galaxy types: elliptical E and passive spiral S0 galaxies are mainly located in the regions 0.5<W2-W3<1.5 and -0.1<W1-W2<0.3, while active spiral galaxies are located in the regions 1<W2-W3<4.5 and -0.1<W1-W2<0.7. Alatalo et al., 2014 have also found that WISE colours were a good proxy to segregate transiting galaxies and active galaxies, with a threshold between the two types around $W2 - W3 \sim 2.8$. I show in Fig. 4.1 the colour W2-W3 as a function of the passivity of the galaxies, i.e., the distance to the main sequence d2ms, for a sample of galaxies from the SDSS MPA-JHU DR8 catalogue cross-matched with WISE. It confirms that WISE colours are very efficient to segregate populations.

The second ingredient of the construction of the training catalogue is the choice of the outputs that are used as reliable reference for SFR and $M_\star$. I have chosen the SFR and stellar masses from the SDSS MPA-JHU DR8 catalogue, as their estimations are based on robust methods and on reliable spectra. The catalogue and the details about the estimations of the SFR and the stellar mass are presented in Chap. 2.1.

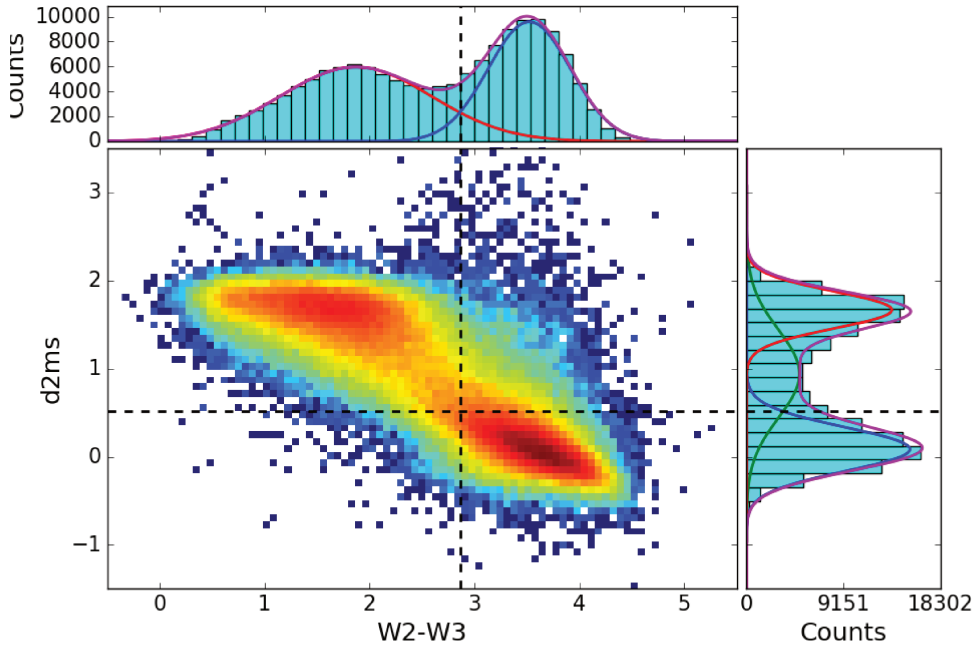

Figure 4.1: 2D histogram of the projected distance to the main sequence in the SFR-M$_\star$ diagram vs. the W2-W3 colour, for 148,685 SDSS galaxies cross-matched with WISE. Top: histogram of the W2-W3 colour. Two Gaussians are fitted for the passive galaxies in red, and for the active galaxies in blue. Right: histogram of the projected distance to the main sequence, d2ms. Three Gaussians are fitted: passive (in red), active (in blue), and transiting (in green) galaxies. The black dotted lines show the intersections of the Gaussians defining active and passive galaxies, in both the colour and the distance to the main sequence.

4.2.2 Selection of WISE sources

On the one hand, I have cleaned the AllWISE Source Catalogue to ensure reliable measurements. I have used the profile-fitted photometry measurements of the W1 (3.4μm), W2 (4.6μm), and W3 (12μm) bands, noted w1mpro, w2mpro, and w3mpro in the AllWISE Source Catalogue. The associated errors and signal-to-noise ratios are noted w1sigmpro, w2sigmpro, w3sigmpro, and w1snr, w2snr, w3snr, respectively. I have rejected all sources with known detection artifacts or measurement artifacts, by adding a constraint cc_flags = 0 for each of the three bands. I have also selected the sources with high-quality photometry measurements, by selecting the higher signal-to-noise ratios. To do so, I have followed Krakowski et al., 2016 and have selected only sources with reliable magnitudes in W1 and W2: w1snr > 2, w2snr > 2, as the WISE magnitudes are upper-values below w*snr < 2 (where * can be 1, 2, or 3) (Krakowski et al., 2016). Around one third of the sources selected in W1 and W2 have w3snr < 2. In this case, following Krakowski et al., 2016, I have corrected the w3mpro magnitudes by adding a correction +0.75. On the cleaned sample, I have also applied a 0^{th}-order k-correction (dependence on redshift) by adding the quantity $-2.5 \times \log(1+z)$ to the magnitudes in each of the three bands (where z is the spectroscopic redshift taken from the SDSS catalogue).

4.2.3 Selection of SDSS sources

On the other hand, I have selected a sub-sample of the SDSS catalogue to keep only sources with reliable estimated properties. To do so, I have set the following flags in the SDSS MPA-JHU DR8 catalogue:

RELIABLE $\neq 0$, Z_WARNING $= 0$, SFR_TOT_P50 $\neq -9999$, LGM_TOT_P50 $\neq -9999$, and $Z > 0$. This cleaned sub-sample contains 794,633 galaxies with reliable SFR and $M_\star$.

4.2.4 Cross-match between WISE and SDSS

Finally, I have constructed the training catalogue for the machine learning algorithm by performing a positional cross-match of the SDSS sub-sample (of 794,633 galaxies) with the cleaned AllWISE Sources Catalogue, within a radius of 6" (the beam of the W1 band of WISE from which the source positions are extracted). I have removed from the cross-matched catalogue all multiple associations to prevent mismatched inputs/outputs properties between WISE and SDSS, and only ensure reliable properties. The final cleaned training catalogue contains 573,582 galaxies. This cross-match catalogue contains spectroscopic redshifts from SDSS, WISE reliable w*mpro magnitudes, and reliable SFR and $M_\star$ measurements from SDSS. These properties are the input and output data of the machine learning algorithm. I show in Fig. 4.2 the range of LW1, LW3, W1-W2 and W2-W3 of the training catalogue. In Fig .4.3, I also show the contours of the SFR-$M_\star$ properties.

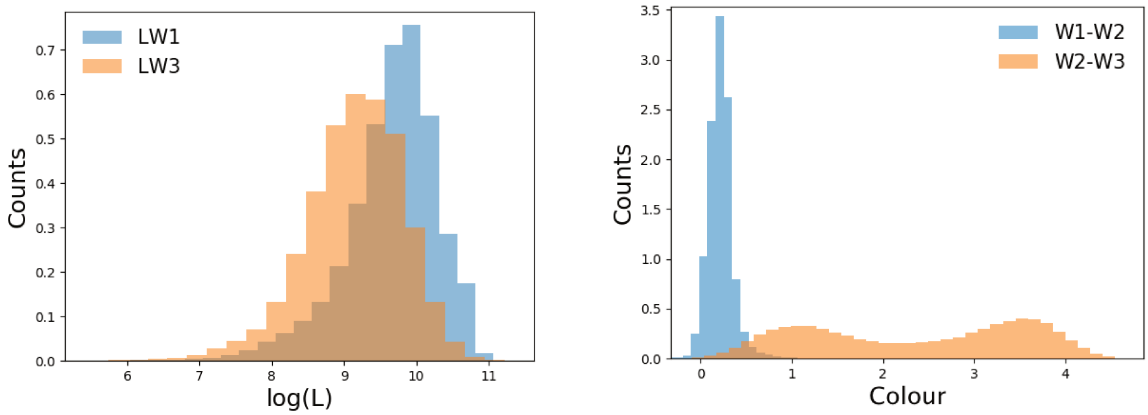

Figure 4.2: Histograms showing the range of the input data. Left: luminosities of the sources of the training catalogue. Right: colours of the same sources.

4.2.5 Training, validation, and test set

To further estimate the errors on the RF, I have split the training catalogue into three sub-samples. Following standard procedures, I have defined three randomised sub-samples: 60% for the training set, 20% for the validation set, and 20% for the test set. I have further checked that changing the sizes of the sub-samples does not affect the results of the RF algorithm.

4.3 Training the machine learning algorithm

4.3.1 Optimisation of the training

To optimise the efficiency of the training of a machine learning algorithm, hyper-parameters have to be tuned. In the case of the random forest, as described in Chap. 3.1.3, the two main hyper-parameters are the number of trees M and the maximum depth of the trees $d_{\rm max}$. To set M and $d_{\rm max}$, I have trained the RF on the training set, varying M and $d_{\rm max}$. For each combination (M, $d_{\rm max}$), I have computed the score of the RF on the validation set using the coefficient of determination $100 \times R^2$, where $R^2 = 1 - \sigma^2_{\rm res}/\sigma^2$.

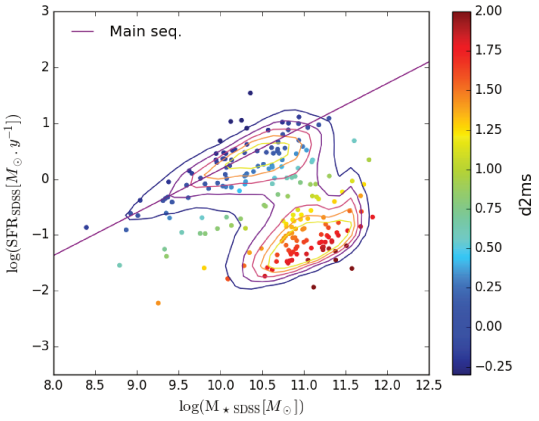

Figure 4.3: SFR vs. $\mathrm{M}_\star$ diagram for the sources of the training catalogue. The lines show the 1σ to 5σ contours. The dots show 100 randomly selected galaxies. The purple solid line traces the main sequence of star forming galaxies given by (Elbaz et al., 2007). The colours of the dots are related to the distance to the main sequence, d2ms, and are thus directly representative of the passivity of the galaxies.

By noting n the number of sources, y_i their output values (from SDSS) and $\tilde{y}_i$ their associated estimated values by the RF, σ^2_{res} can be expressed as $\sigma^2_{\mathrm{res}} = \sum_{i=0}^{n} (\tilde{y}_i - y_i)^2$ and represent the residual sum of squares, and $\sigma^2 = \sum_{i=0}^{n} (y_i - \bar{y})^2$ represent the variance of the output distribution. I have set the two parameters M and d_{max} to the ones giving the highest score on the validation set. Figure 4.4 shows the score of the RF on the validation set, depending on the two parameters M and d_{max}.

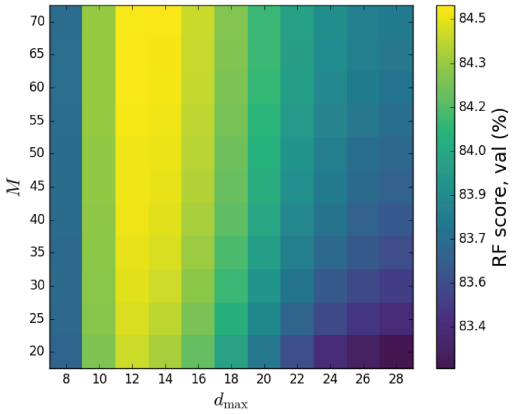

Figure 4.4: Percentage score of the RF results on the validation sample as a function of the RF parameters M and d_{max} (M being the number of trees and d_{max} the maximum depth). Setting $M = 40$ and $d_{\mathrm{max}} = 12$ is enough in this case to optimise the RF.

For the parameter d_{max}, the performance of the RF increases until $d_{\mathrm{max}} = 12$ and starts decreasing beyond, indicating an over-fitting of the training set. Thus, d_{max} is set to $d_{\mathrm{max}} = 12$. The parameter M represents the statistics of the forest (number of trees). Thus, a simple lower limit is enough to optimise the training, meaning that enough statistics was reached to properly estimate the results. Increasing the value of M to higher values does not give significant changes. Here, setting $M = 40$ is sufficient to ensure good statistics. The optimal score on the validation set is 84.5%.

4.3.2 Relevance of the input parameters

One advantage of the RF algorithm is that the relevance of the input parameters during the training process can be estimated and studied. I have studied the relevance of the input parameters, by estimating either SFR alone, $M_\star$ alone, or both SFR and $M_\star$ (left, middle, and right panel in Fig. 4.5, respectively).

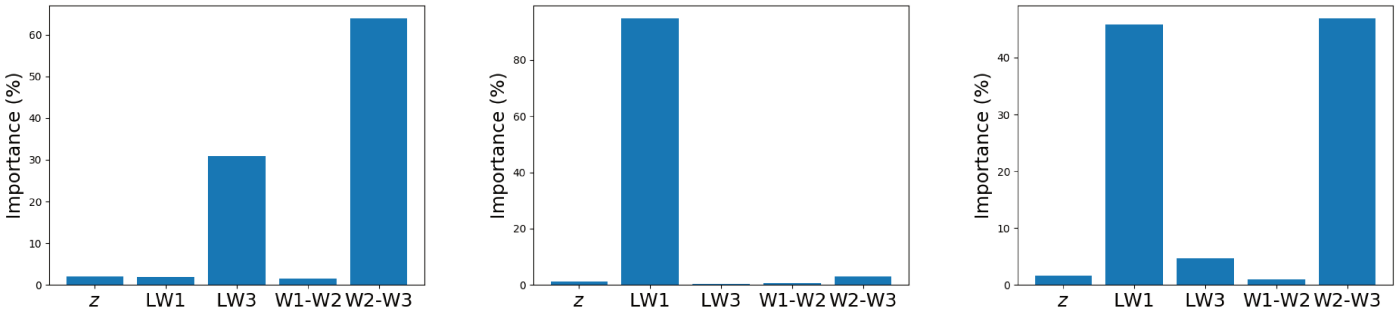

Figure 4.5: Relevance of the input parameters during the training of the RF. Left: RF trained to estimate $M_\star$ only. Middle: RF trained to estimate SFR only. Right: RF trained to estimate both SFR and $M_\star$.

From the five input parameters, the first obvious tendency is the low impact of the redshift (only on the dependence on distance, as the redshift is also hidden in the luminosities LW1 and LW3) and of the colour W1-W2 during the three trainings. For the estimation of $M_\star$ alone, it is clear that the luminosity LW1, as expected, is the more relevant parameter, with only a very slight contribution from the colour W2-W3. For the SFR estimation alone, two more relevant parameters, as expected, are the luminosity LW3 and the colour W2-W3 used to segregate the two main populations of galaxies. The case where the RF is trained to estimate both SFR and $M_\star$ shows that the two more relevant parameters are the luminosity LW1 and the colour W2-W3, with a slight contribution (of about 5%) of the luminosity LW3. This indicates that the two quantities LW1 and W2-W3 are the most efficient to classify and segregate galaxy populations. I have confirmed this results by showing in Fig. 4.6 the two populations of galaxies, i.e., active and passive, very well separated by the two quantities LW1 and W2-W3.

4.3.3 Results

I have trained the RF on the training set, with the two parameters set to $M = 40$ and $d_{\max} = 12$. SFR and $M_\star$ are then estimated on the test set, noted $\mathrm{SFR_{ML}}$ and $\mathrm{M_{\star ML}}$, respectively. I have then compared these results with their reference values from the SDSS catalogue, noted $\mathrm{SFR_{SDSS}}$ and $\mathrm{M_{\star SDSS}}$. The performance of the RF in terms of the errors and biases can hence be estimated by comparing the values from the SDSS and those estimated with the RF. Figure 4.7 shows an overall good agreement between the SDSS values and the values estimated with the RF algorithm, both for SFR and for $M_\star$. This agreement indicates that the RF algorithm is well trained.

For the stellar mass estimated with the RF algorithm (Fig. 4.7 left panel), the scatter between the estimated and reference SDSS values is quantified through the variance: $\sigma^2_{\mathrm{M}_\star} = 0.026$. The associated standard deviation is $\sigma_{\mathrm{M}_\star} = 0.16$ dex, translated into a statistical error of a factor $10^{\sigma_{\mathrm{M}_\star}} = 1.45$. For the SFR (Fig. 4.7 right panel), the scatter is larger and the variance is $\sigma^2_{\mathrm{SFR}} = 0.145$. This gives a standard deviation of $\sigma_{\mathrm{SFR}} = 0.38$ dex, and a statistical error of a factor $10^{\sigma_{\mathrm{SFR}}} = 2.40$.

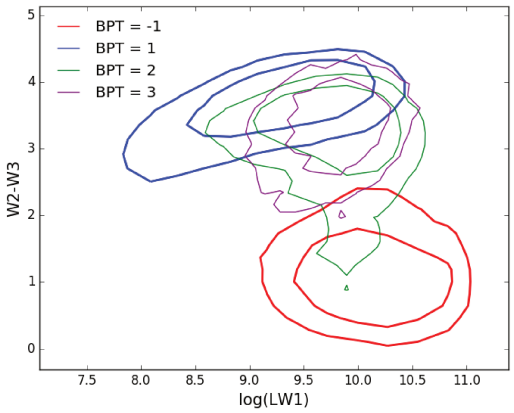

Figure 4.6: 1σ and 3σ contours of galaxy types as a function of the two most relevant parameters during the training of the RF: W2-W3 and LW1. Blue contours represent active galaxies, red contours passive galaxies, green contours transiting galaxies, and purple contours AGN, according to the BPT classification of the SDSS MPA/JHU DR8 catalogue.

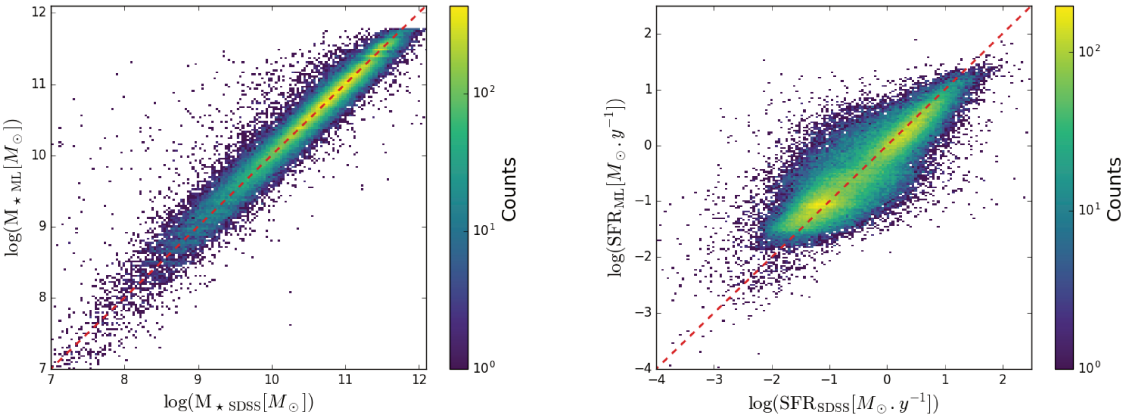

Figure 4.7: Results of the RF on the test sample (20% of the entire sample), with RF parameters set to $M = 40$ and $d_{\max} = 12$. Left: $M_\star$ estimations with the RF compared to $M_\star$ from the SDSS MPA-JHU DR8 catalogue. Right: SFR estimations with the RF compared to SFR from the SDSS MPA-JHU DR8 catalogue.

4.4 Robustness of the algorithm

It is important to have precise results, with error bars estimated from the RF for both SFR and $M_\star$ values, and it is of equal importance to have accurate, that is unbiased, results.

4.4.1 Errors according to redshift

I have first investigated potential bias induced by the redshift dependence of the SFR and stellar mass in the redshift range, $0 < z < 0.3$, of the training catalogue used. I show in Fig. 4.8 the errors (defined as the difference between machine-learning estimated values and SDSS values), for $M_\star$ (left panel) and SFR (right panel) for the galaxies in the test set. No obvious bias on redshift is observed. In the left panel of

Fig. 4.8, a slight increase of the scatter for $M_\star$ is seen at very low redshifts. This is due to biases on the magnitudes for low redshift sources. This is detailed in Sect. 4.4.4.

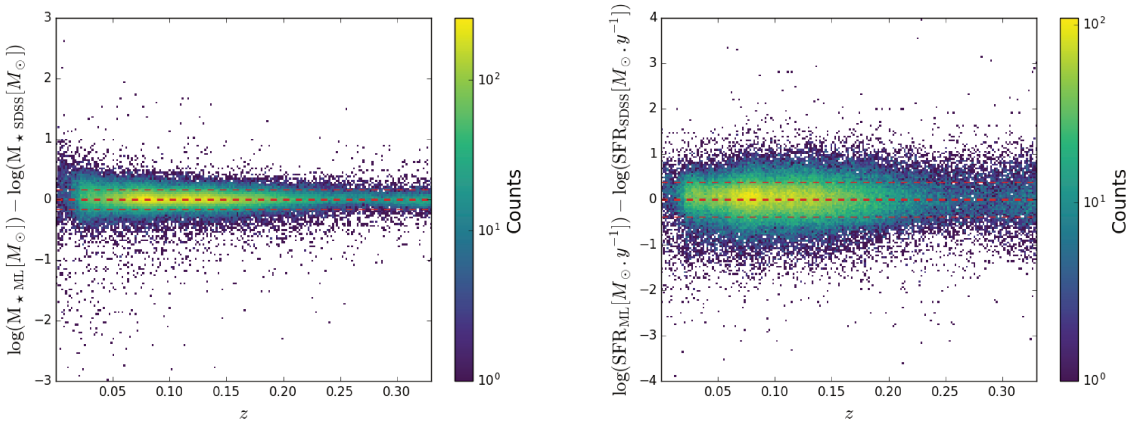

Figure 4.8: Errors of the RF results obtained for sources in the test set (same errors as those presented in Fig. 4.7) as a function of redshift, for $M_\star$ and SFR.

4.4.2 Errors according to galaxy types

Another bias can be potentially induced by the galaxy types. To explore this possibility, I have compared the results of the RF algorithm as a function of the BPT classes provided in the SDSS MPA-JHU DR8 catalogue. The BPT classes are good indicators of the galaxy types (as shown in Fig. 2.3 in Chap. 2.1, where the red contours of passive galaxies with BPT = -1 are well in the cloud of red and dead galaxies, the blue contours of active galaxies with BPT = 1 are well aligned along the main sequence, and the green contours of transiting galaxies with BPT = 2 are well populating the green valley). The positions in the BPT diagram can therefore be reliable proxies of the galaxy type. In Fig. 4.9, I show the results of the RF (same as those displayed in Fig. 4.7), with the contour colours displaying the different BPT classes.

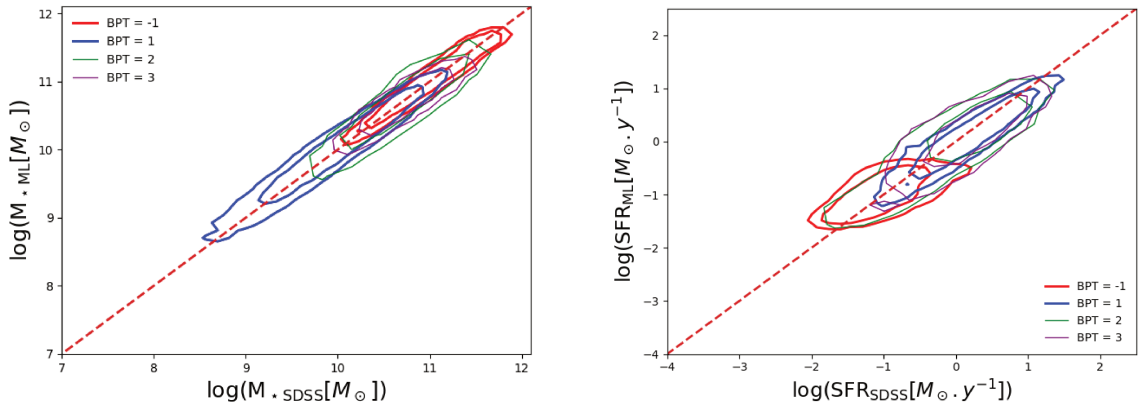

Figure 4.9: SFR and $M_\star$ obtained with the RF algorithm on the test set compared to the SDSS classification based on the BPT diagram. Colour code of the contours is the same as in Fig. 2.3 (in Chap. 2).

The RF performs equally well for any type of galaxy and no strong bias induced by the galaxy type is

	ML with z				ML without z	Analytical
	All	Passive (BPT = -1)	Active (BPT = 1)	Transiting (BPT = 2)	All	All
$\sigma_{M_\star}$	0.16	0.11	0.23	0.13	0.32	0.23
σ_{SFR}	0.38	0.38	0.30	0.39	0.43	0.47 (active only)

Table 4.1: Summary of the different scatters obtained on the test set with different methods.

seen. Moreover, the scatter of the results depends only very slightly on galaxy type. For passive galaxies, with BPT = -1, the scatter on $M_\star$ tends to be reduced: σ_{SFR} = 0.38 dex and $\sigma_{M_\star}$ = 0.11 dex. For active galaxies, with BPT = 1, the inverse trend is seen and the scatter on the SFR tends to be reduced with a small increase of the scatter of $M_\star$: σ_{SFR} = 0.30 dex and $\sigma_{M_\star}$ = 0.23 dex. For transitioning galaxies, with BPT = 2, the scatters are roughly the same as the overall ones obtained on the all the populations: σ_{SFR} = 0.39 dex and $\sigma_{M_\star}$ = 0.13 dex. A summary of the different scatters is shown in Tab. 4.1.

4.4.3 Comparison with other methods

Several studies have derived analytical formulae to determine SFR and $M_\star$, some of them also based on the WISE luminosities (e.g., Wen et al., 2013; Jarrett et al., 2013; Cluver et al., 2014; Cluver et al., 2017). I have compared the SFR and $M_\star$ estimated with the RF algorithm with those derived using different approaches, but based on the same observables (WISE luminosities: LW1 and LW3). I have focused on the $M_\star$ estimated with the relation from Wen et al., 2013, using LW1:

$$\log(M_{\star\,\mathrm{Wen}}) = 1.12 \times \log(\mathrm{LW1}) - 0.04, \tag{4.1}$$

and on the SFR estimated from Cluver et al., 2014 using LW3, (derived for active galaxies only):

$$\log(\mathrm{SFR}_{\mathrm{Cluver}}) = 1.13 \times \log(\mathrm{LW3}) - 10.24. \tag{4.2}$$

I have computed $M_\star$ with Eq. 4.1 for all galaxies. In the left panel of Fig. 4.10, I compare $M_\star$ estimated with Eq. 4.1 with $M_\star$ from the SDSS MPA-JHU DR8 catalogue. I also show the 1, 3, and 5 σ contours of the RF estimations. This comparison shows the smaller scatter of the $M_\star$ estimated with the RF algorithm. This result is not surprising considering that the RF algorithm has five inputs compared to only one for Eq. 4.1. Scatters are $\sigma_{M_{\star\,\mathrm{Wen}}}$ = 0.23 dex and $\sigma_{M_{\star\,\mathrm{ML}}}$ = 0.16 dex (see also Tab. 4.1).

SFR are computed, for star-forming galaxies only with BPT = 1, to satisfy the conditions of Cluver et al., 2014, following Eq. 4.2. They are compared with the SFR from the SDSS MPA-JHU catalogue (right panel of Fig. 4.10). The blue contours show the RF estimations, and the red contours show the SFR computed with the method from Cluver et al., 2014 for passive galaxies (i.e., BPT = -1). A smaller scatter for the SFR estimations from the RF algorithm is found; this is again expected since five inputs are used for the RF algorithm compared to only one for Eq. 4.2. The limitation of the application domain of a relation between LW3 and SFR is also shown in terms of its dependence on galaxy type (huge bias on the red contours showing the passive galaxies). Scatter are $\sigma_{\mathrm{SFR}_{\mathrm{Cluver}}}$ = 0.47 dex and $\sigma_{\mathrm{SFR}_{\mathrm{ML}}}$ = 0.30 dex for active galaxies (see also Tab. 4.1), while for passive galaxies scatters are $\sigma_{\mathrm{SFR}_{\mathrm{Cluver}}}$ = 0.49 dex and $\sigma_{\mathrm{SFR}_{\mathrm{ML}}}$ = 0.38 dex. For passive galaxies computed with Eq. 4.2, a bias, defined as the absolute difference of the means, is found: b_{Cluver} = 0.93 dex (compared with b_{ML} = 0.04 dex for RF estimations).

As a second comparison, I have used a catalogue of galaxies with SFR computed using an alternative method. An example is the extended version of the COLD GASS (CO Legacy Database for GASS) catalogue of nearby galaxies, the xCOLD GASS catalogue[1] (Saintonge et al., 2017). The sample contains 532 nearby galaxies ($z < 0.05$) from SDSS selected in mass ($M_\star > 10^9 M_\odot$). The galaxies span a wide

[1]Publicly avaible at `http://www.star.ucl.ac.uk/xCOLDGASS/index.html`

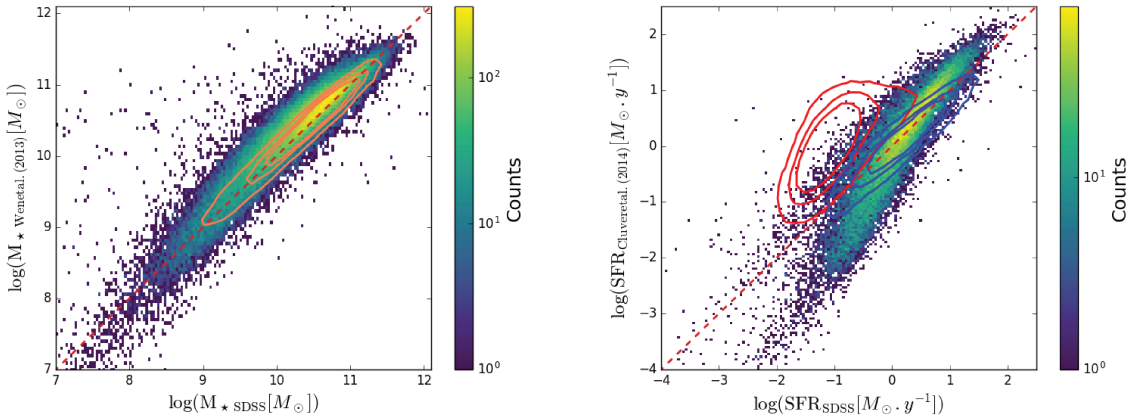

Figure 4.10: Comparisons on the test set. Left: $M_\star$ estimated with the method of Wen et al., 2013, using the LW1, compared with $M_\star$ from SDSS. The lines represent the 1, 3, and 5σ contours of the RF estimations. Right: SFR computed (for star-forming galaxies only) with the method of Cluver et al., 2014, using LW3, compared to SFR from the SDSS catalogue. The blue contours show the results of the RF for active galaxies only (i.e., BPT = 1). The red contours show the SFR estimated with the method of Cluver et al., 2014 for passive galaxies (i.e., BPT = -1). In both panels, the dashed line shows the one-to-one correlation.

range of SFR values and galaxy types, each of them observed IRAM-30m CO(1-0) observations, and as mentioned, *"because the COLD GASS sample is large and unbiased, it serves as the perfect reference for studies of particular galaxy populations"*. SFR and $M_\star$ are computed with the method from Janowiecki et al., 2017, using a combination of UV from the Galaxy Evolution Explorer satellite[2] (GALEX, Milliard et al., 2017) and of infrared from WISE. I show in Fig. 4.11 the results of the RF algorithm for the galaxies of the xCOLDGASS catalogue (with only one association with AllWISE sources within a radius of 6"), compared with their values (noted S17). A good overall agreement is seen, and a small bias is observed in the SFR estimations, especially for passive galaxies. As a bias is also observed between the SDSS MPA-JHU DR8 values and the UV+IR estimation in the right panel of Fig. 4.11, the bias observed in the SFR estimated by the RF algorithm possibly comes from the two different methods for the SFR estimations in Saintonge et al., 2017 and in the SDSS MPA-JHU DR8 catalogue. Indeed, they do not use the same wavelengths to estimate their quantities. This shows that unobscured UV SFR is not seen in WISE or in SDSS. The catalogue chosen as output values for the RF algorithm (the SDSS MPA-JHU DR8) is thus very restrictive and may bias the RF algorithm. Other value-added catalogues of galaxies, with more reliable SFR and $M_\star$ estimations, may be chosen as outputs to construct the training catalogue of the RF algorithm, and hence increase the performance.

4.4.4 Limitation

The results of the RF algorithm agree well with previous works, in particular those of Wen et al., 2013 and Cluver et al., 2014. The domain of application of all three studies in terms of galaxies and more precisely in terms of redshifts is similar (mean redshift at around $z = 0.15$). Comparisons with other available catalogues providing SFR and $M_\star$ are not possible when the sources under consideration are too different from the domain of application of the RF algorithm, that is, the domain of the catalogue chosen as outputs (in this case the SDSS MPA-JHU DR8 catalogue). I have focused on two extreme

[2] `http://www.galex.caltech.edu`

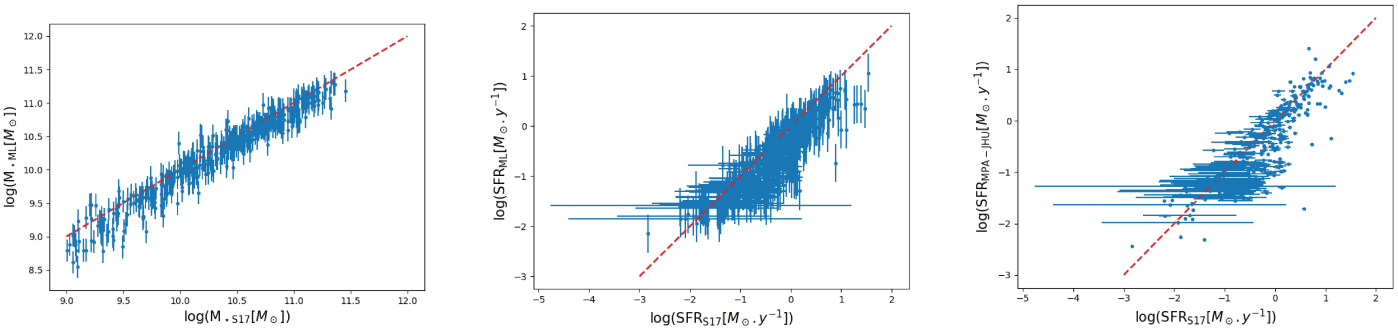

Figure 4.11: Comparison with the xCOLDGAS catalogue from Saintonge et al., 2017 for the sources that match only once with the AllWISE catalogue in a 6" radius. Left: $M_\star$ estimated with the RF compared with the $M_\star$ provided by the SDSS MPA-JHU values catalogue. Middle: SFR estimated with the RF compared with the SFR provided by the xCOLDGAS catalogue, computed with combined UV and IR data. Right: SFR given by the SDSS MPA-JHU DR8 catalogue compared with the SFR from xCOLDGAS.

cases, where the domains of application prevent us from obtaining reasonable estimates. On the one hand, I have considered very nearby galaxies with redshifts $z < 0.01$, and on the other hand, high-redshift galaxies, up to $z = 8$.

For the first case, I have used a sample of very nearby ($z < 0.01$) star-forming galaxies, that was constructed by combining the Spitzer catalogue SINGS (Kennicutt et al., 2003) and the Herschel Space Observatory catalogue KINGFISH (Kennicutt et al., 2011). The SINGS/KINGFISH catalogue of 79 sources was used by Cluver et al., 2017 to successfully estimate the relation between SFR and LW3. In this catalogue, the redshift domain, i.e., $z < 0.01$, implies that the galaxies are resolved, and therefore the use of the WISE Atlas Images is needed to accurately measure the fluxes of the objects (Saintonge et al., 2017; Cluver et al., 2017). The WISE magnitudes may lack flux emission, that is seen in Fig. 4.8, as a higher scatter in $M_\star$ estimations for very-low-redshift galaxies ($z < 0.01$).

For the second case, I have used the COSMOS2015 catalogue[3] (Laigle et al., 2018), which provides apparent magnitudes in 30 bands for approximately half a million objects up to redshifts $z = 8$. It also provides photometric redshifts, SFR, and stellar masses, computed with the code `LEPHARE`[4] (Arnouts and Ilbert, 2011). Here, the application of the RF algorithm is not possible. In this case, the main issue is the resolution of WISE of 6"; too many COSMOS sources are associated with a WISE galaxy inside the WISE beam, and therefore a correct association with the AllWISE catalogue is not possible.

The application domain of the RF algorithm is thus $0.01 < z < 0.3$. Applying it beyond this redshift limit may lead to strong biases, as such sources have not been seen during the training.

4.5 Redshift dependency during the training

4.5.1 Train without redshift

The SFR and stellar mass estimated with the RF rely on the redshift information. Redshifts are used to compute the luminosities and they also impact the evolution of the mean global SFR over time. The need for redshift information to estimate the $M_\star$ and the SFR is very restrictive, as spectroscopic redshifts are hard to obtain, and photometric ones are not accurate. I have tested the performance of the RF method

[3] `http://cosmos.astro.caltech.edu`
[4] `http://www.cfht.hawaii.edu/~arnouts/LEPHARE/lephare.html`

without any redshift information. This implies that the k-correction on the magnitudes have not been applied, and that the luminosities have not been computed. The input parameters are thus only the two WISE magnitudes W1 and W3 and the two WISE colours W1-W2 and W2-W3. In Fig. 4.12, I show the results on the test set. The accuracy of the method is highly degraded: the scatter of $M_\star$ estimations is $\sigma_{M_\star} = 0.32$ dex and the scatter of SFR estimations is $\sigma_{SFR} = 0.43$ dex (compared to $\sigma_{M_\star} = 0.16$ dex and $\sigma_{SFR} = 0.38$ dex with the redshift information in inputs; see also Table 4.1).

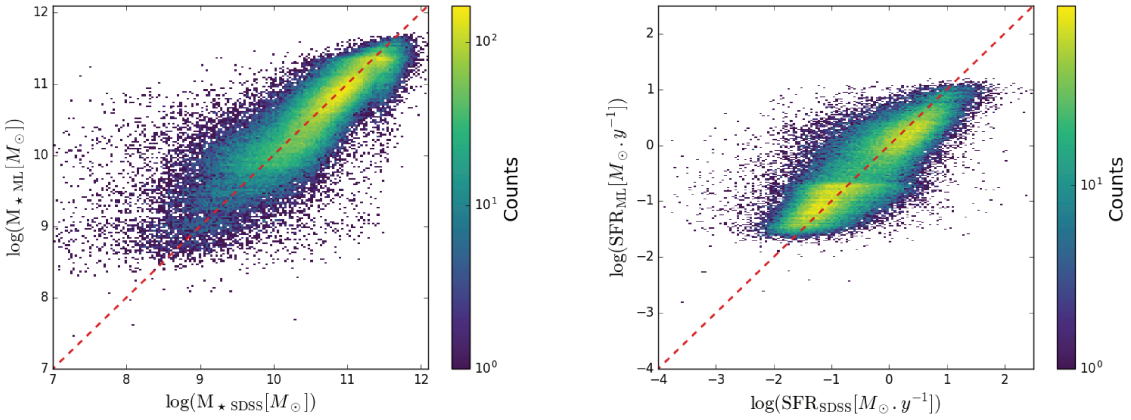

Figure 4.12: Results of the RF on the test sample, with only W1, W3, W1-W2 and W2-W3 in input, i.e., without any information about the redshift z. Left: $M_\star$ estimated with the RF compared with $M_\star$ from the SDSS MPA-JHU DR8 catalogue. Right: SFR estimated with the RF compared with SFR from the SDSS MPA-JHU DR8 catalogue.

4.5.2 Modelling photometric redshifts

In order to investigate if the machine-learning algorithm trained on a spectroscopic-redshift catalogue can also be applied to high-accuracy photometric-redshift catalogues, I have modelled an error as $\sigma_z(1 + z)$ and added it to the spectroscopic redshifts of the test sample. In Fig. 4.13, I show in blue the evolution of the scatter of the SFR and $M_\star$ estimates as a function of $\sigma_z(1 + z)$. A large increase of the scatter is seen for both properties with decreasing redshift accuracy. This trend is due to two effects. On the one hand, the increase in redshift error obviously impacts the SFR and $M_\star$ estimates. On the other hand, an additional bias increases the scatters on SFR and $M_\star$. This is shown in the left panel of Fig. 4.14, for the SFR estimates with $\sigma_z(1 + z) = 0.015$.

In order to correct for this bias, I have arbitrarily modelled it with an exponential $a \times e^{-z/z_0}$ (see Fig. 4.14). As shown in Fig. 4.15, the biases evolve with the photometric redshift errors. In Fig. 4.13, I show in orange the scatter on the bias-corrected properties as a function of redshift. The scatters are significantly reduced, but still rather large: $\sigma_{M_\star} = 0.35$ dex and $\sigma_{SFR} = 0.44$ dex at $\sigma_z(1 + z) = 0.03$. In the range $0.1 < z < 0.3$, the scatters (shown in green in Fig. 4.13) are reduced down to more reasonable values, such as $\sigma_{M_\star} = 0.24$ dex and $\sigma_{SFR} = 0.42$ dex at $\sigma_z(1 + z) = 0.03$ (the accuracy expected for the photometric-redshift catalogue of Euclid). Hence, the RF algorithm could be applied to present or future photometric-redshift catalogues like DES, LSST, Pan-Starrs, Euclid, or WFIRST. In the meantime, I have applied the RF algorithm to a full-sky catalogue of photometric redshifts, the WISExSCOS catalogue.

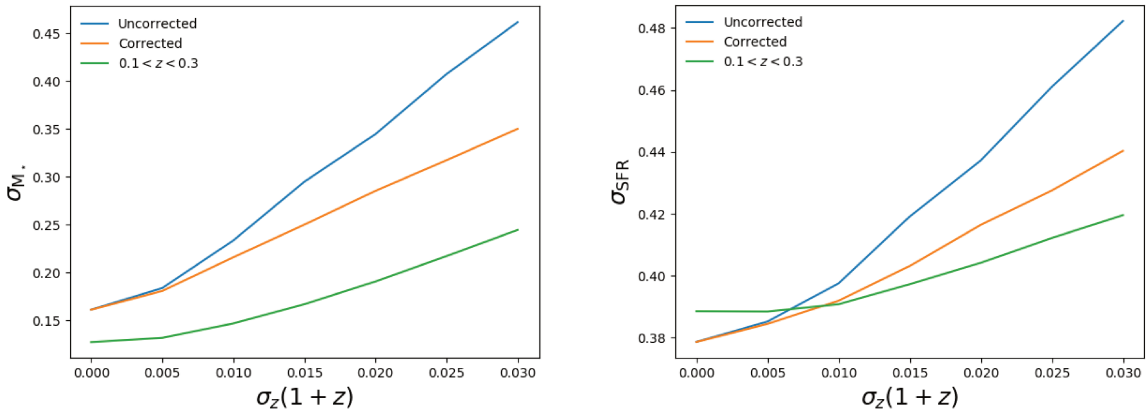

Figure 4.13: Evolution of the scatters of the properties estimated with the RF as a function of the redshift error $\sigma_z(1+z)$. Left: scatter for $M_\star$. Right: scatter for SFR. The blue lines correspond to the scatters of the whole sample, regardless of the induced bias. The orange lines correspond to the scatters of bias-corrected properties from the biases modelled in Fig. 4.15. The green lines correspond to the scatters of sources in the redshift range $0.1 < z < 0.3$ (while blue and orange lines are sources in the range $0 < z < 0.3$).

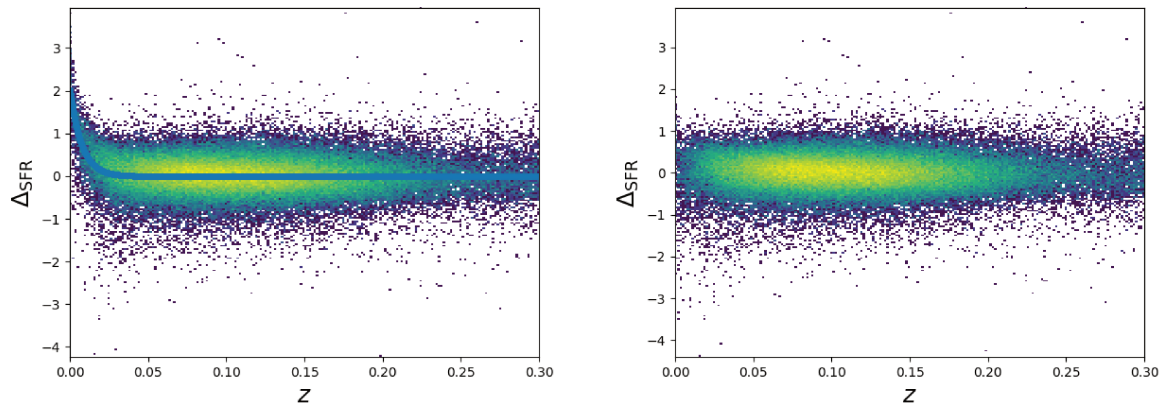

Figure 4.14: Example of bias induced by the additional redshift error. Left: for $\sigma_z(1+z) = 0.015$ the errors on the SFR estimates are shown as a function of redshift. The blue line corresponds to the modelled bias. Right: same errors, corrected for the bias.

4.6 WISExSCOS value-added catalogue and maps

4.6.1 WISExSCOS value-added catalogue

The WISExSCOS photometric redshift catalogue (described in Chap. 2.2.2.2), contains both photometric redshift estimates and WISE magnitudes. Based on these properties, I have estimated SFR and $M_\star$ for 15,765,535 sources in the redshift range $0.1 < z < 0.3$ using the RF algorithm. The range of SFR and of $M_\star$ of the WISExSCOS value-added catalogue is shown in the left panel of Fig. 4.16.

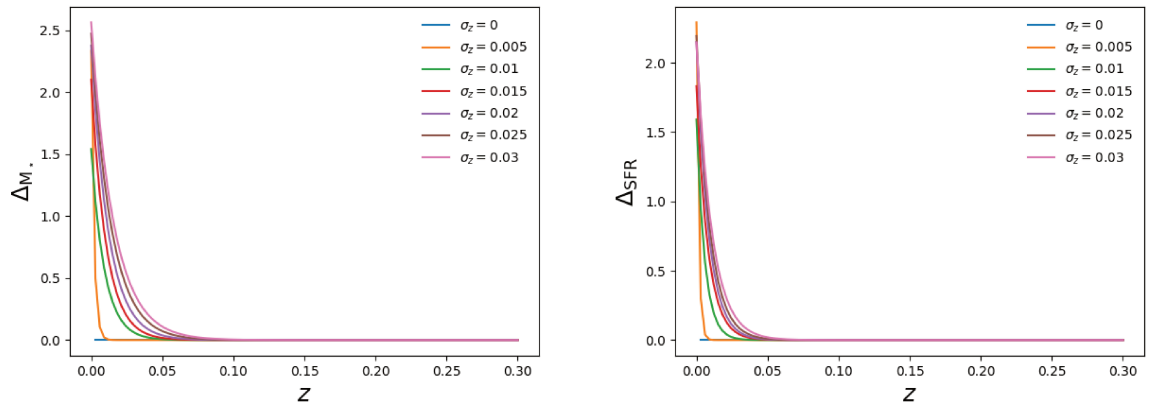

Figure 4.15: Evolution of the bias as a function of redshift, for different redshift errors (indicated by the colours). Left: bias for $M_\star$ estimates. Right: bias for SFR estimates.

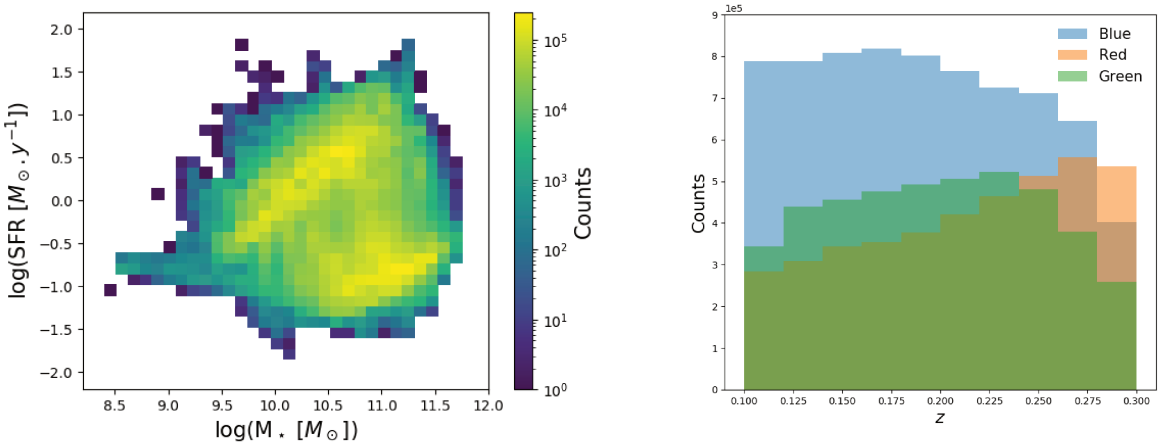

Figure 4.16: Left: range of SFR and of $M_\star$ of the 15,765,535 sources in the WISExSCOS value-added catalogue in the range $0.1 < z < 0.3$. Right: distributions of the active, transitioning, and passive galaxies of the WISExSCOS value-added catalogue as a function of redshift.

4.6.2 Distance to the main sequence estimation

Similarly to estimating the activity of galaxies by computing specific SFR (which illustrates the efficiency of a galaxy in forming stars), the distance to the main sequence on an SFR-$M_\star$ diagram which is translated into a colour in Fig. 4.3 informs us about the star formation activity. Since this distance, noted d2ms, is directly related to the distance of a galaxy from being star-forming, I prefer to use the term passivity rather than activity. The redder the points in Fig. 4.3, the more distant from the main sequence of star-forming galaxies, and the more passive the galaxies are. The quantity d2ms, efficiently estimated by the RF algorithm, can be a very useful property to segregate populations of active, transitioning, or passive galaxies. I have thus split the 15,765,535 sources of the WISExSCOS catalogue in the range $0.1 < z < 0.3$, into the three galaxy populations, i.e., active, transitioning, and passive galaxies, using the d2ms. I have defined as active galaxies sources with d2ms < 0.4, as transiting galaxies sources with 0.4 < d2ms < 1.25, and as passive galaxies sources with d2ms > 1.25. These cuts are defined by the

intersections of the three Gaussians fitted on the distribution of the d2ms on a sample of SDSS galaxies, used to model the three populations of galaxies (shown in the right panel of Fig. 4.1). After the splitting, the catalogue contains 7,249,961 active, 4,353,744 transitioning, and 4,161,830 passive galaxies. The distributions of the three populations of galaxies in redshifts are shown in the right panel of Fig. 4.16.

4.6.3 Galaxy density maps

Based on the WISExSCOS value-added catalogue and on its three sub-samples of galaxy types defined above, I have constructed 3D galaxy density maps, in the redshift range $0.1 < z < 0.3$, using the positions of the sources on the sphere and their redshift information. To do so, I have reconstructed the density field with the pyDTFE code presented in Chap. 3.3.1. Based on 3072 3D density fields in patches of $3.7^\circ \times 3.7^\circ$, I have generated four 3D HEALPIX full-sky maps: one for all galaxies, and one for each of the three populations of galaxies. The number of pixels of the HEALPIX maps are $n_{\rm side} = 2048$, i.e., a pixel resolution of 1.7', and the binning in redshift was arbitrarily set at $\delta_z = 0.01$. An example of a slice at $z = 0.15$ of the 3D passive galaxy density map (smoothed at 30' for visualisation) is shown in Fig. 4.17. The large scale distribution of the galaxies is seen, together with contaminations, i.e., the stripes due to the WISE scanning strategy, the mask of our galaxy and of the Magellanic cloud, and the reddening from dust around our galaxy and the Magellanic cloud. High density concentrations are also seen, which are galaxy clusters. In addition to the four galaxy density maps, I have constructed in the same way 3D maps of SFR and M_* for all galaxies, and for the active, transitioning, and passive populations. The maps are constructed by interpolating at their 3D positions the SFR and M_* estimated with the RF algorithm.

4.7 Towards a WebService tool

I have developed a WebService based on python and on the Django package[5]. It allows to quickly estimate on the fly the SFR and $M_\star$ properties of photometric SDSS sources. The interface is very user friendly, one only has to enter the coordinates of the field and a radius for the cone search. The trained RF model is called and loaded in back-end. HTML SQL queries on the AllWISE catalogue and on the SDSS photometric redshift survey are performed in parallel, and a cross-match between the two catalogues is performed on the fly. The RF model then uses the photometric redshifts from SDSS and the WISE luminosities and colours to estimate SFR and $M_\star$, and display the sources in an interactive 3D fields, together with their associated positions on the SFR-$M_\star$ diagram. The user can easily select with a lasso selection the sources of a desired type with the information on the distance to the main sequence of star-forming galaxies, and see their updated positions in the 3D field. A screen-shot in Fig. 4.18 shows the interface of the WebService, with an example in a field of 30 arcmin around the position (R.A. = 180°, Dec.= 0°). In this example, passive galaxies were selected with the diagram on the right. An over-density of the selected passive galaxies is seen in the 3D density field on the left (light blue circle). The redshift and the position of this over-density of passive galaxies were matched with a galaxy cluster from the RedMaPPer catalogue presented in Chap.2.4.1: RMJ120143.7-001104.2, at $z = 0.16$. Thus, this example also shows the potential of such value-added catalogue and tool to detect galaxy clusters.

4.8 Summary

I have developed a method based on machine learning to estimate the SFR and $M_\star$ of galaxies, in the redshift range $0.01 < z < 0.3$, over the whole usable sky when their redshifts are known.

[5] `https://www.djangoproject.com`

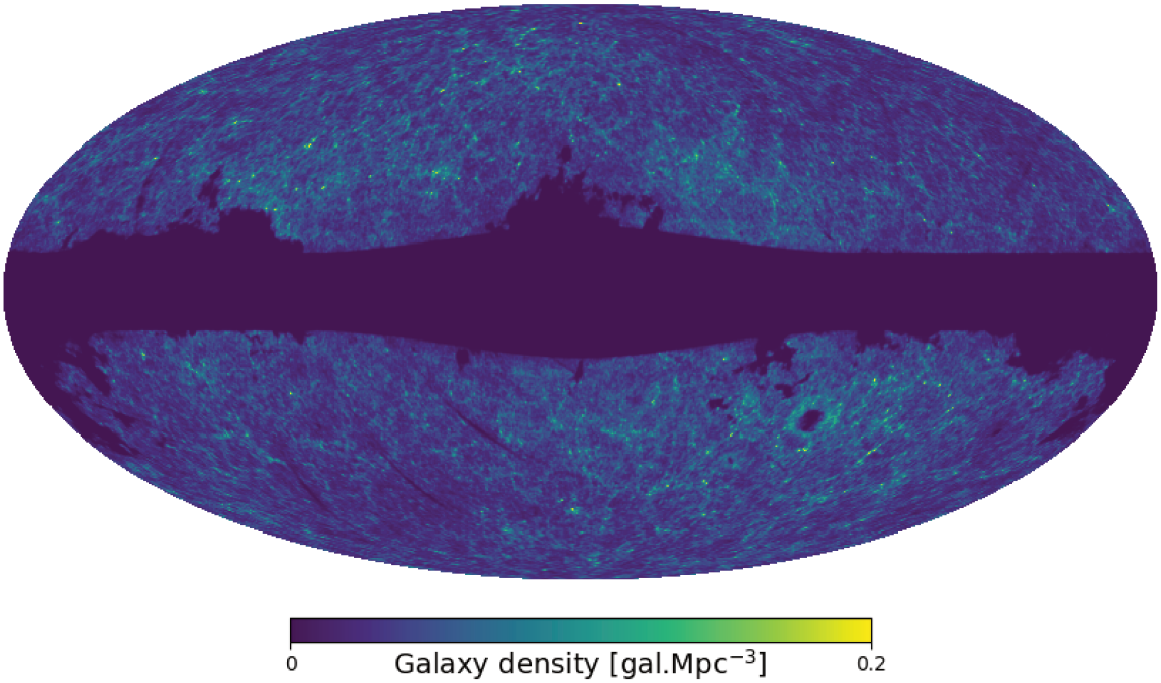

Figure 4.17: Mollweide projection of the slice at $z = 0.15$ of the 3D passive galaxy density map constructed with the pyDTFE code. The map is smoothed at 30' for visualisation. The large-scale distribution of the galaxies is seen, together with artefacts from the WISE scanning strategy, and the masks of our galaxy and of the Magellanic cloud.

The trained Random Forest model gives SFR errors of $\sigma_{\rm SFR} = 0.38$ dex and $M_\star$ errors of $\sigma_{M_\star} = 0.16$ dex, independently of the galaxy types, and unbiased with respect to redshift. I have further checked that the choice of an RF algorithm was not altering the performance by comparing the results to other machine learning algorithms like Artificial Neural Networks.

The RF algorithm is trained on the redshift z, the WISE luminosities LW1 and LW3, and the WISE colours W1-W2 and W2-W3. These input properties allow a very efficient segregation of the different galaxy types, and are good proxies to estimate the stellar mass and the SFR. As outputs to train the algorithm, I have chosen the SFR and $M_\star$ from the SDSS MPA-JHU DR8 catalogue. However, the method can be adapted in the future with any other catalogue with SFR and $M_\star$, to improve the results, and also to increase the redshift range of the application domain.

I have applied the RF algorithm to the WISExSCOS photometric redshift catalogue, computing the SFR and $M_\star$ of 15,765,535 sources in the full usable sky (outside our galaxy), in the range $0.1 < z < 0.3$. This catalogue was used in several projects, such as the cross-correlation of galaxy types and SZ effect (in collaboration with G. Fabbian, F. Bianchini, and N. Aghanim), or the multipolar analysis galaxy cluster outskirts as a function of the galaxy types (in collaboration with C. Gouin and N. Aghanim). I have also used this catalogue to study the properties of the galaxies in the bridge of matter between A399 and A401 (presented in the next chapter), and the derived galaxy density maps to characterise galaxy properties around large-scale cosmic filaments (in Chap. 6).

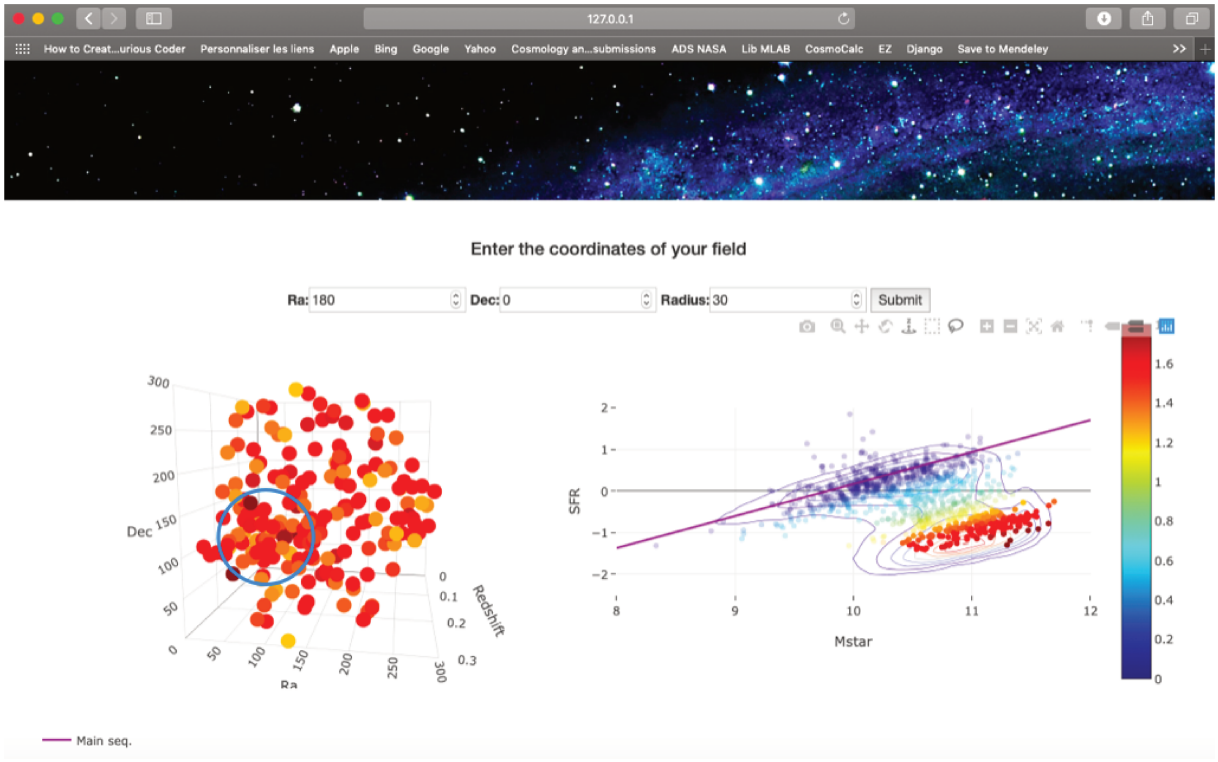

Figure 4.18: Screen-shot of the WebService. Sources are shown in a 30' radius around the arbitrary position (R.A. = 180°, Dec.= 0°). Passive galaxies are selected, showing an over-density (light blue circle). This cluster of passive galaxies corresponds to the RedMaPPer cluster RMJ120143.7-001104.2, at $z = 0.16$.

Chapter 5

Detailed study of the bridge of matter A399-A401

Contents

Abstract

In this chapter, I present a multi-wavelength analysis of two pairs of galaxy clusters selected with the SZ effect. I have focused on one pair of particular interest: A399-A401 at redshift $z \sim 0.073$, that is linked by a bridge of matter of 3 Mpc. I have also performed the first analysis of one lower-significance newly associated pair: A21-PSZ2 G114.09-34.34 at $z \sim 0.094$, separated by 4.2 Mpc. I have characterised the intra-cluster gas using the SZ signal from *Planck* and, when possible, the galaxy optical and infra-red properties based on two photometric redshift catalogues: 2MPZ and WISExSCOS. From the SZ data, I have measured the gas pressure in the clusters and in the inter-cluster bridge. In the case of A399-A401, the results are in perfect agreement with previous studies and, using the temperature measured from the X-rays, I have further estimated the gas density in the bridge, $n_0 = (4.3 \pm 0.7) \times 10^{-4}\mathrm{cm}^{-3}$. The optical and IR colour-colour and colour-magnitude analyses of the galaxies selected in the cluster pair,

together with their star formation activities, show no segregation between the galaxy populations both in the clusters and in the bridge of A399-A401: they are all passive. The gas and galaxy properties of this pair suggest that the whole system has formed at the same time and corresponds to a pre-merger, with a cosmic filament gas heated by the collapse. For the cluster pair A21-PSZ2 G144.90-34.35, I have estimated the pressure in the clusters and in the inter-cluster bridge in SZ. However, as the optical and IR data are limited in this case, concluding on the presence of an actual cosmic filament or proposing a scenario is not possible. This chapter uses material from *"Gas and galaxies in filament between clusters of galaxies: The study of A399-A401"*, V. Bonjean, N. Aghanim, P. Salomé, M. Douspis, and A. Beelen, 2018, A&A, 609, A49.

Résumé

Dans ce chapitre, je présente une analyse multi-longueurs d'onde de deux paires d'amas de galaxies sélectionnées avec l'effet SZ. J'ai mis l'accent sur une paire de particuler intérêt : la paire A399-A401, à redshift $z \sim 0,073$, qui est reliée par un pont de matière long de 3 Mpc. J'ai également effectué la première analyse d'une paire nouvellement associée : la paire A21-PSZ2 G114.09-34.34, à $z \sim 0,094$, séparée par 4,2 Mpc. J'ai caractérisé le gaz intra-amas en utilisant le signal SZ de *Planck*, et spécifiquement pour A399-A401, j'ai étudié les propriétés optiques et infrarouges des galaxies à l'aide de deux catalogues de redshift photométriques : 2MPZ et WISExSCOS. A partir des données SZ, j'ai mesuré la pression du gaz dans les amas et dans le pont. Dans le cas de A399-A401, les résultats sont en accord avec les études précédentes, et en utilisant la température mesurée à partir des observations en X, j'ai estimé la densité du gaz dans le pont comme étant $n_0 = (4,3 \pm 0,7) \times 10^{-4} \mathrm{cm}^{-3}$. Les analyses optiques et infrarouges des couleurs et des magnitude des galaxies sélectionnées comme membre de la paire d'amas, ainsi que leurs activités de formation d'étoiles, ne montrent aucune différence entre les populations de galaxies d'amas et celle du pont : elles sont toutes passives. Les propriétés du gaz et des galaxies de cette paire suggèrent que tout le système s'est formé en même temps et correspond à une pré-fusion, avec un gaz filamentaire cosmique chauffé par l'effondrement des deux amas. Pour la paire d'amas A21-PSZ2 G144.90-34.35, j'ai estimé la pression dans les amas et dans le pont inter-amas avec l'effet SZ. Cependant, comme les données optiques et IR sont limitées dans ce cas, il n'est pas possible de conclure sur la réelle présence d'un filament cosmique ou de proposer un scénario. Ce chapitre utilise du matériel du papier *"Gas and galaxies in filament between clusters of galaxies: The study of A399-A401"*, V. Bonjean, N. Aghanim, P. Salomé, M. Douspis, et A. Beelen, 2018, A&A, 609, A49.

5.1 Introduction

As the galaxy clusters are built up over time from mergers and interactions of smaller systems (e.g., Navarro, Frenk, and White, 1995; Springel et al., 2005), they are naturally connected to the Cosmic Web via the filaments. Strategies to probe the Cosmic Web are thus associated with our ability to probe filamentary structures between clusters or in their outskirts. This is in principle possible via observation of the galaxy distribution (e.g., Durret et al., 2016), the weak gravitational lensing (e.g., Eckert et al., 2015), the X-ray emission from the hot gas (e.g., Eckert et al., 2015), and the thermal SZ effect (e.g., Planck Collaboration et al., 2013b), but the search for filaments linking the clusters to the Cosmic Web is really difficult due to their very low densities. This subject has gained a lot of interest, and focuses mostly on two cases: filaments in the outskirts of individual clusters, and inter-cluster filaments (or bridges) in pairs of clusters. Regarding the former case, Eckert et al., 2015 detected large-scale structures of several Mpc in the outskirts of the galaxy cluster Abell 2744 at redshift $z = 0.306$, from combined observations in the X-ray, of galaxy over-densities, and from weak lensing analysis. For the DAFT/FADA cluster sample (Guennou et al., 2010), Durret et al., 2016 found filaments in clusters' outskirts with

two-dimensional galaxy densities obtained with CFHT and SUBARU observations. Regarding cluster pairs, the inter-cluster filaments or the bridges are expected to be denser, with a hotter gas, and thus in principle easier to detect, in particular in the X-rays and in the SZ effect (Dolag et al., 2006). Cluster pairs are thus good targets and have therefore been the subject of numerous studies. The photometric properties of the galaxies in the inter-cluster filament, their star-formation evolutions, their stacked weak lensing properties, their stacked SZ contributions, and so on, were performed in many selected cluster (or group) pairs (e.g. Fadda et al., 2008; Gallazzi et al., 2009; Edwards et al., 2010; Zhang et al., 2013; Martínez, Muriel, and Coenda, 2016; Epps and Hudson, 2017; Tanimura et al., 2019b; de Graaff et al., 2019). Galaxy clusters may show substructures or evidence of dynamical effect: they merge, interact, and accrete smaller groups. The galaxy properties derived from optical and near-infrared data thus need to be combined, in a multi-wavelength analysis, with the study of cluster gas content. The gas properties of cluster pairs were therefore also investigated mostly using X-ray observations. This is the case of the particular pair A399-A401, thoroughly studied using data from ASCA, ROSAT, Suzaku and XMM-Newton (e.g. Karachentsev and Kopylov, 1980; Ulmer and Cruddace, 1981; Fujita et al., 1996; Fabian, Peres, and White, 1997; Sakelliou and Ponman, 2004; Fujita et al., 2008; Akamatsu et al., 2017). The gas in this galaxy cluster pair was already studied using both the X-ray (with ROSAT) and the SZ effect (with the *Planck* SZ MILCA map of 2013) by Planck Collaboration et al., 2013b. Very recently, this pair was studied in LOFAR (Govoni et al., 2019), where the authors have detected the presence of magnetic fields in the bridge re-accelerating the electrons in the ionised gas between the two clusters.

In this chapter, I present a multi-wavelength analysis of A399-A401, based on the SZ map of 2015 from *Planck*, and optical and near infra-red data from photometric redshift catalogues with physical galaxy properties estimated as described in Chap. 4. I also present a newly interesting pair: A21-PSZ2 G114.90-34.35.

5.2 SZ selection of galaxy cluster pairs

I have first constructed a sample of cluster pairs based on the SZ signal given that it is a priori the most appropriate tracer of the diffuse hot gas. In practice, I have based the selection both on the SZ cluster database (presented in Chap. 2.4.2) and on the signal-to-noise ratio of the SZ signal between the pairs.

5.2.1 SZ selection

Following Planck Collaboration et al., 2013b, I have applied three conditions to select the galaxy-cluster pairs. First, the two clusters need to be at the same redshift, second, the distance between the two clusters needs to be large enough to avoid blending effects, and finally, the significance of the SZ signal in the inter-cluster region needs to be above 2σ. The two first empirical conditions were proposed by Planck Collaboration et al., 2013b: Δz<0.01, and considering the *Planck* MILCA SZ map beam of 10 arcmin, 30 arcmin<$\theta_{\rm sep}$<120 arcmin, where Δz is the redshift difference between the two clusters and $\theta_{\rm sep}$ is the angular distance separating the two clusters. This corresponds to projected distances between 3 and 40 Mpc. I have found a total of 71 cluster pairs satisfying the two conditions (Fig. 5.1) in the SZ cluster database of 2,690 clusters or candidates presented in Chap. 2.4.2.2. This is about three times more pairs than the selection based on clusters from the MCXC catalogue, performed in Planck Collaboration et al., 2013b. About one third of the clusters in these pairs are Abell clusters (Abell, Corwin, and Olowin, 1989), one third are *Planck* newly detected clusters, and the others are X-ray clusters from the MCXC catalogue or SZ clusters detected by SPT.

5.2.2 Projection in patches

Since one of the aim of the analysis is to study and characterise the gas between pairs of clusters, I have considered the SZ signal significance in terms of the signal-to-noise ratio of the SZ emission in the bridge region: $S/N_{\rm fil}$. For each cluster pair, I have projected patches of the *Planck* MILCA SZ map of $s = 90$ pixels aside, centered on the mean of the two cluster positions, with a pixel resolution $\theta_{\rm p}$ depending on the angular distance $\theta_{\rm sep}$ separating the two clusters: $\theta_{\rm p} = (5 \times \theta_{\rm sep})/s$. I have rotated the patches for convenience (the alignment of the clusters is always horizontal), and have masked the SZ signal at the positions of the *Planck* Catalogue of Compact Sources (Planck Collaboration et al., 2016d) presented in Chap. 2.3.3, and at the positions of the other SZ sources in the fields (except in an area defined as the region encompassed within $3 \times r_{500}$ from the cluster centres). In order to define the radii r_{500}, I have used the relation between masses and radii (the Eq. 1.5 in Chap. 1), for the 1,681 SZ clusters for which both redshift and mass estimates $M_{500}^{\rm SZ}$ are available. Figure 5.2 shows two examples of projected SZ maps, for the pair A399-A401 and A21-PSZ2 G144.90-34.35, with the red circles representing the r_{500} radii of the clusters, and the white pixels the masked regions. On these patches, I have defined the area of the potential inter-cluster bridge as the cylinder between the minimum of the two r_{500} in the radial direction and between the two r_{500} in the longitudinal direction. I have estimated the signal-to-noise ratio $S/N_{\rm fil} = (y_{\rm m,fil} - y_{\rm m,bkgd})/\sigma_{\rm bkgd}$, where $y_{\rm m,fil}$ is the mean SZ signal in the area defining the bridge, and $y_{\rm m,bkgd}$ and $\sigma_{\rm bkgd}$ are the SZ mean signal and the standard deviation of the background area beyond the $3 \times r_{500}$ distance from the cluster centres.

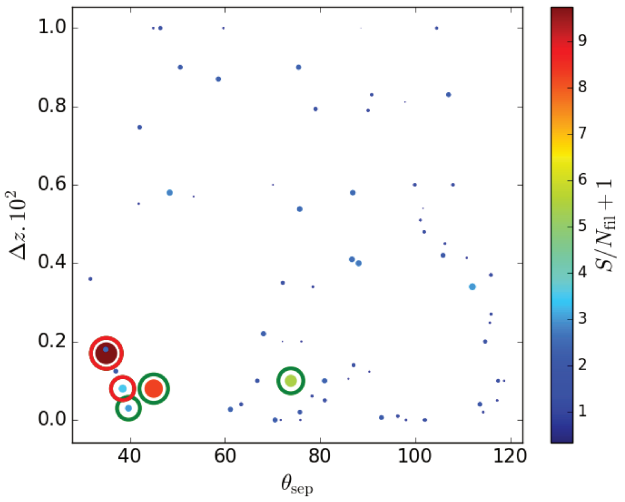

Figure 5.1: Distribution of the 71 cluster pairs in the parameter space Δz-$\theta_{\rm sep}$. The areas and colours of the circles depend on $S/N_{\rm fil}$. The pairs for which $S/N_{\rm fil} > 2$ are shown with green open circles. The red open circles indicate the final selection after exclusion of pairs lying in complex systems.

I show in Fig. 5.1 the 71 cluster pairs that satisfy the two first conditions (redshift and angular separations) in the SZ cluster database. The size and colour of the circles show the estimated $S/N_{\rm fil}$. Green open circles mark the pairs for which the S/N is greater than two. Among these, I have chosen to discard the cluster pairs belonging to larger and more complex systems such as the Shapley super-cluster, or the pair containing the cluster A3395, a system already known to host several groups (Lakhchaura et al., 2011; Planck Collaboration et al., 2013b). I have also finally removed the pair SPT J0655-5234–SPT J0659-5300, at redshift $z = 0.47$, with $S/N_{\rm fil} = 2.03$. The redshift of this system is high, and thus the two catalogues used to study the galaxies in Sect. 5.4, with $z_{\rm med} \sim 0.08$ and $z_{\rm med} \sim 0.2$, lack statistics. Some

of these removed complex systems will conduct to further dedicated analysis (discussed in Chap. 5.6). Eventually, I have focused on two isolated pairs with a significant SZ signal in their inter-cluster region (red open circles in Fig. 5.1): the pair A399-A401 at redshift $z \sim 0.073$ with $S/N_{\rm fil} = 8.74$, and the newly associated pair A21-PSZ2 G114.09-34.34 at redshift $z \sim 0.094$ with $S/N_{\rm fil} = 2.53$. Their main properties are presented in Tab. 5.1, and the projected patches of the *Planck* MILCA SZ map are shown in Fig. 5.2.

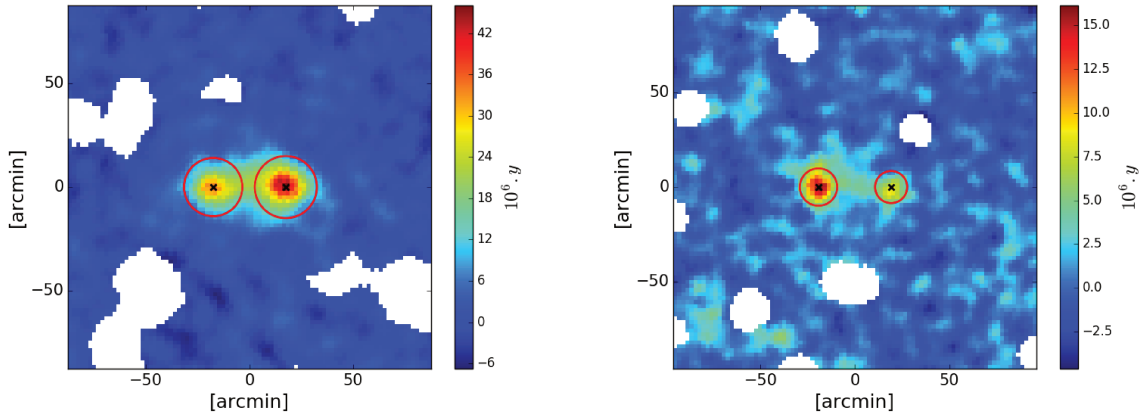

Figure 5.2: Projected patches of the *Planck* MILCA SZ map for the two selected cluster pairs. Left: the pair A399-A401. Right: the pair A21-PSZ2 G114.90-34.35. The red circles show the radii r_{500} of the clusters and the white pixels are masked from the *Planck* Catalog of Compact Sources and from other clusters in the fields. The black crosses show the centres of the SZ clusters.

Cluster	R.A. (deg)	Dec. (deg)	$S/N_{\rm SZ}$	$S/N_{\rm d}$	z	r_{500} (Mpc)	$\theta_{\rm sep}$ (')	$\theta_{\rm sep}$ (Mpc)	$S/N_{\rm fil}$	$S/N_{\rm fil,d}$
A399	44.45	13.05	12.96	8.85	0.072	1.20	35.05	3.0	8.74	7.89
A401	44.73	13.57	19.66	9.74	0.074	1.30				
A21	5.15	28.67	9.36	6.19	0.094	1.07	38.5	4.2	2.53	2.08
PSZ2 G114.9	5.35	28.05	4.78	2.63	0.095	0.93				

Table 5.1: Main properties of the selected pairs: cluster names, SZ positions in R.A. and Dec., S/Ns of the clusters from the SZ catalogue and in galaxy over-density (Sect. 5.4.3), redshifts, estimated radii r_{500}. The last four columns indicate the angular separation in arcmin and in Mpc, and the S/Ns of the bridges in the SZ map (Sect. 5.2) and in galaxy over-density (Sect. 5.4.3).

5.3 Gas properties from SZ analysis

5.3.1 Model

In order to derive the gas properties of the inter-cluster bridge, I have modeled the entire system with four components: two clusters, one bridge, and a planar background (see the schema in Fig. 5.3). To model the galaxy clusters (blue and green components in Fig. 5.3), I have chosen the spherically symmetric Generalised Navarro, Frenk,& White (GNFW) pressure profile (Nagai, Vikhlinin, and Kravtsov, 2007), that was fitted by Arnaud et al., 2010; Planck Collaboration et al., 2013a to best represent the gas pressure profiles around galaxy clusters. This model takes the form:

$$P(\mathbf{r}) = \frac{P_0}{\left(\frac{\mathbf{r}}{r_s}\right)^{\gamma}\left[1+\left(\frac{\mathbf{r}}{r_s}\right)^{\alpha}\right]^{(\beta_{\rm GNFW}-\gamma)/\alpha}}, \tag{5.1}$$

where $\mathbf{r}$ is the radius, P_0 is the central pressure, $r_s = r_{500}/c_{500}$ is the characteristic radius, and γ, α, $\beta_{\rm GNFW}$ are the internal ($\mathbf{r} < r_s$), intermediate ($\mathbf{r} \sim r_s$), and external ($\mathbf{r} > r_s$) slopes, respectively. All the parameters, but two, are fixed to the ones of the universal profile fitted by Planck Collaboration et al., 2013a on stacked SZ clusters. The two parameters let free are $\beta_{\rm GNFW}$ (related to the cluster extension), and P_0 (related to the Compton parameter y amplitude). These two parameters can be degenerated with each other, and with some parameters of the bridge, such as its length or its pressure.

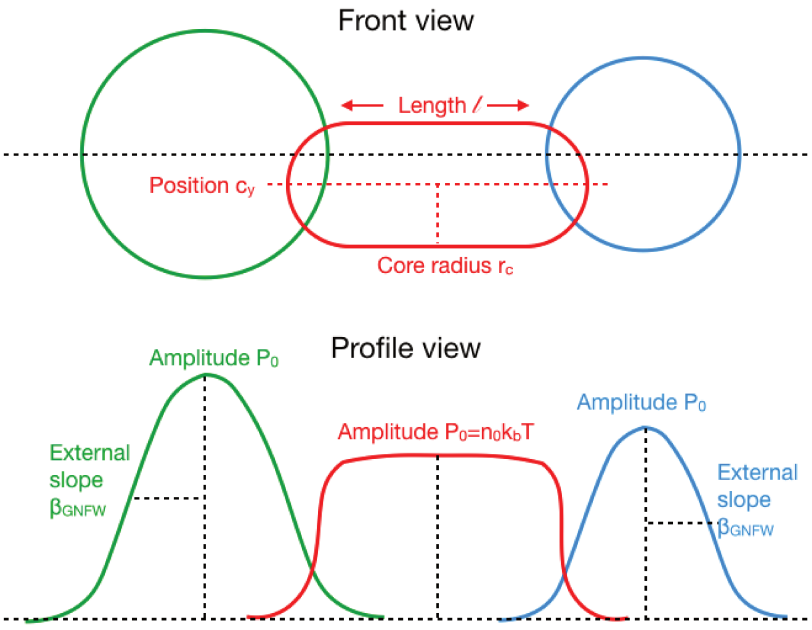

Figure 5.3: Front and profile schematic views of the model: the two clusters in green and in blue with two free parameters each, and the inter-cluster bridge in red with three free parameters. The length of the filament l is fixed to the distance between the two $r_{500}/2$ of each cluster. A planar background with three free parameters is considered.

The GNFW pressure profile was specifically developed to model the pressure distribution in galaxy clusters. As such, it cannot be applied to filaments or bridges. In the absence of any physically motivated or established model of the inter-cluster filamentary gas, I have chosen an isothermal β-model (Cavaliere and Fusco-Femiano, 1978), with a cylindrical symmetry, to describe the pressure distribution in the radial direction of the bridge:

$$n_{\rm e}(\mathbf{r}) = \frac{n_{\rm e,0}}{\left(1+\left(\frac{\mathbf{r}}{r_c}\right)^2\right)^{\frac{3}{2}\beta}}, \tag{5.2}$$

where $\mathbf{r}$ is the radius, $n_{\rm e,0}$ is the central electron density in the bridge, r_c is the core radius, and β the slope, fixed to 4/3 following Planck Collaboration et al., 2013b, to model non-magnetised filament (Ostriker, 1964). Since the recent discovery of magnetic fields in the bridge between the two clusters A399-A401 with LOFAR (Govoni et al., 2019), this non-magnetised model is outdated. Parameters could be improved by setting the slope at a different value, or by choosing a different model. However, results

of my study were published before this discovery. I have thus the slope β to to 4/3 and have derived the other parameters of the model without accounting for a magnetic field. Three parameters are let free: the central pressure in the bridge $P_0 = n_0 \times k_b T$ (where T is the electron gas temperature), the core radius r_c (bridge extension in the radial direction), and the position of the bridge in the radial direction, c_y (expressed as a percentage of the map size). The length of the bridge in the horizontal direction l is fixed to the distance between the two $r_{500}/2$ of each cluster, in order to avoid any degeneracy between the parameters of the bridge and those of the two clusters, and thus avoid any bias in the measurements of the pressures of the clusters.

Finally, a plane ($f(x, y) = ax + by + c$) is chosen to model the background; it corrects from possible residual gradient emission induced by large-scale contamination to the SZ signal, such as galactic dust emission.

5.3.2 Results for the pair A399-A401

I have performed an MCMC analysis using the `python` algorithm `emcee` (Foreman-Mackey et al., 2013) on the pair A399-A401, to fit the ten free parameters to the *Planck* MILCA SZ map. The resulting posterior parameter distributions are shown in Fig. 5.4.

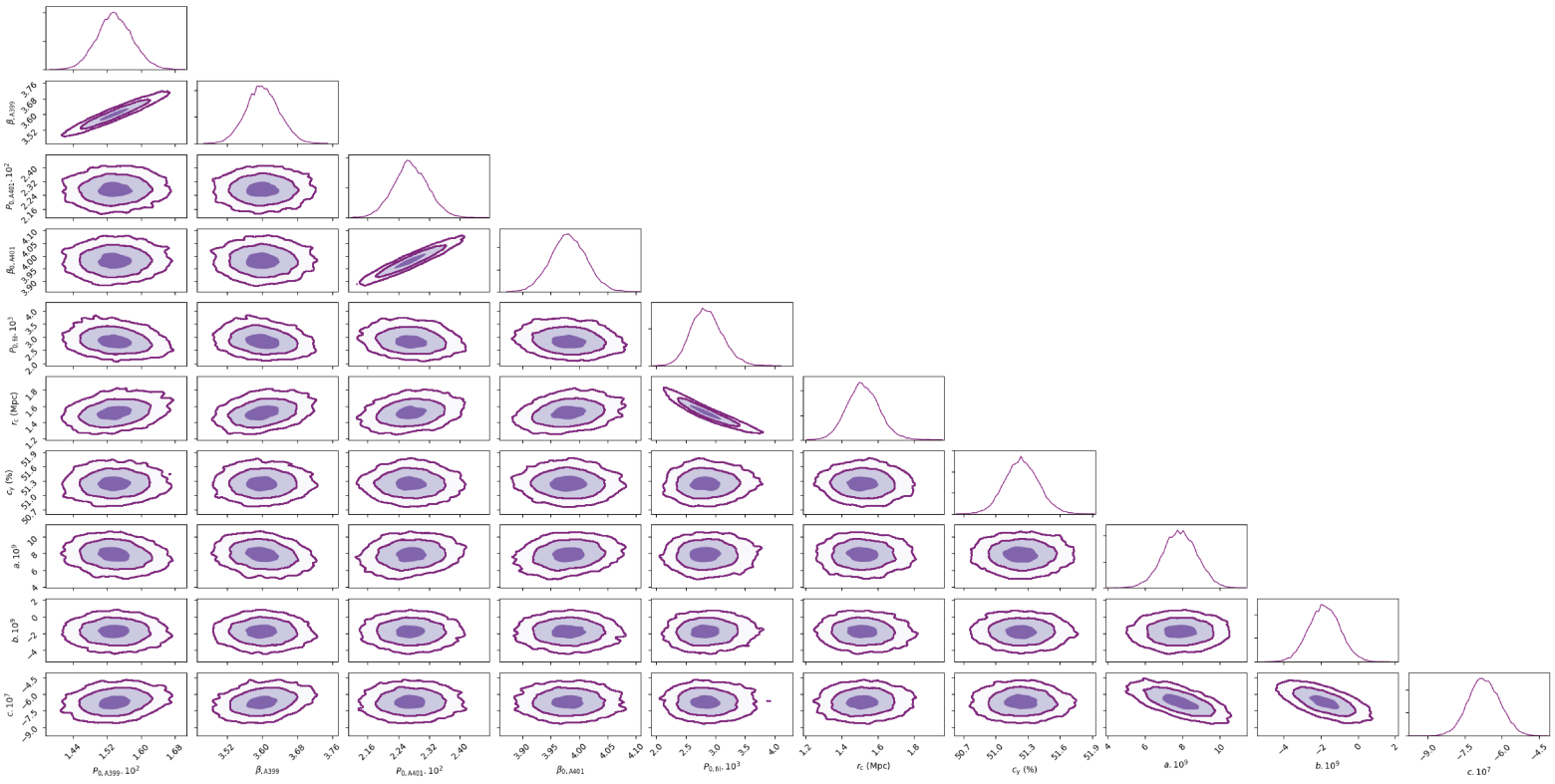

Figure 5.4: Posterior parameter distributions from the MCMC analysis of the system A399-A401. The diagonal plots show the one-dimensional likelihood of the ten parameters.

The Gaussian distributions of the parameters mean that the algorithm have converged. No degeneracies are seen between the three parameters of the background and the physical parameters of the model. This shows that for the cluster pair A399-A401 the measured physical properties, such as the pressures in the clusters or in the bridge, are not biased by a potential large-scale contamination. An interesting point when looking at the correlations between the ten parameters is the degeneracies between the two parameters of the cluster model, i.e., $\beta_{\rm GNFW}$ and P_0. This reflects that a high SZ amplitude can be obtained by

the combination of a high external slope and a small cluster extension. No obvious degeneracies with other parameters are seen. Similarly to the cluster, a degeneracy between the two physical parameters of the bridge, i.e., central pressure and radial extension: $P_0 = n_0 \times k_b T$ and r_c, is seen.

From the MCMC analysis, the best-fit median central pressure in the bridge is $P_0 = (2.84 \pm 0.27) \times 10^{-3}$keV.cm^{-3}, and the best-fit radius is $r_c = 1.52 \pm 0.09$ Mpc, that is, $r_c = 17.6 \pm 1.1$ arcmin. The bridge's pressure obtained from the SZ-only analysis is in agreement with the value obtained with the density and the temperature measurements from the analysis of Planck Collaboration et al., 2013b, that used both SZ and X-rays (ROSAT) data (that is $P_0 = (2.6 \pm 0.5) \times 10^{-3}$keV.cm^{-3}). Assuming the most accurate temperature measured by Fujita et al., 2008 and Akamatsu et al., 2017 with Suzaku in X-rays, $k_b T_X = 6.5 \pm 0.5$keV, I have estimated the central density in the bridge. I have found $n_0 = (4.3 \pm 0.7) \times 10^{-4}$cm^{-3}, which is again in full agreement with the value $n_0 = (3.7 \pm 0.2) \times 10^{-4}$cm^{-3} derived by Planck Collaboration et al., 2013b from their combined analysis of SZ and X-ray signals.

I have simulated a map with the best-fitted model (clusters + bridge) using the parameters derived from the MCMC (see Tab. 5.2). The reduced chi-square value obtained from the residuals between the model and the SZ map is $\chi^2 = 0.97$. I show the residual SZ signal after subtracting the model in the SZ map in Fig. 5.5. It exhibits no significant SZ emission between the clusters. Small stripes at pixel scale are visible at the position of the clusters. This is related to the correlation between the pixels, due to the re-pixelisation of the *Planck* pixels during the projection of the patches, that is not taken into account in the model. Finally, I show the good agreement between the model and the data by showing in the left panel of Fig. 5.6 the model superimposed with the longitudinal cut in the SZ map, across the two clusters, and the residuals after substraction of the model.

A399		A401		Filament		
P_0(keV.cm^{-3})	β_{GNFW}	P_0(keV.cm^{-3})	β_{GNFW}	P_0(keV.cm^{-3})	r_c (Mpc)	c_y (%)
$(1.54 \pm 0.04) \times 10^{-2}$	3.60 ± 0.04	$(2.27 \pm 0.04) \times 10^{-2}$	3.98 ± 0.03	$(2.84 \pm 0.27) \times 10^{-3}$	1.52 ± 0.09	51.24 ± 0.15

Table 5.2: Best-fit parameters of the model derived from the MCMC. The best values are the median of the parameters distributions, and the error-bars are computed with the 16th and the 84th percentiles.

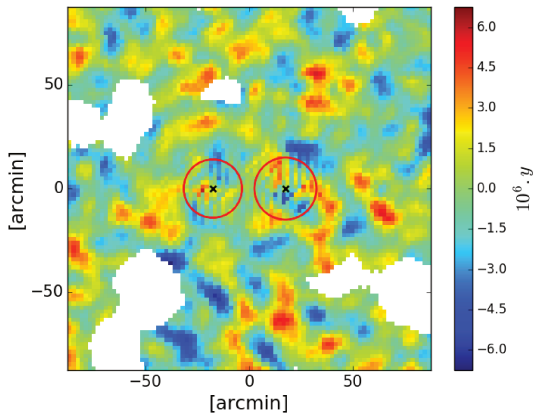

Figure 5.5: Residual *Planck* MILCA SZ map of the pair A399-A401 after subtracting the model (clusters + inter-cluster bridge) with the best-fit parameters from Tab. 5.2. The red circles represent the r_{500} radii of each cluster, and the black crosses their central positions. The small-scale features seen at the position of the two clusters in the residuals are due to the substraction of the model that do not takes into account the re-pixelisation of *Planck* pixels.

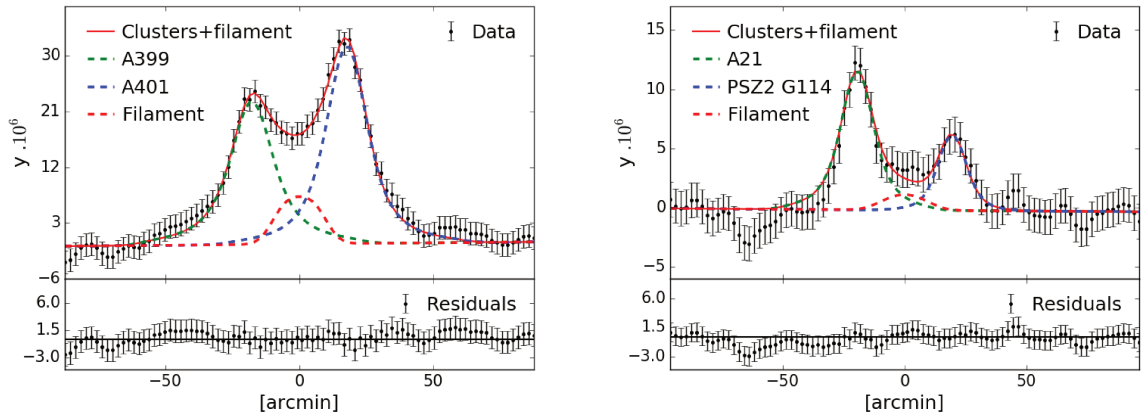

Figure 5.6: Left: longitudinal cut across the A399-A401 pair. Top panel: the fuchsia line shows the *Planck* MILCA SZ data. The red line shows the model (clusters + inter-cluster bridge) with the best-fit parameters from Tab. 5.2. The dotted green, blue, and red lines show the contributions of A399, A401, and of the inter-cluster bridge, respectively. Bottom panel: residuals after subtracting the model (clusters + inter-cluster bridge). Right: longitudinal cut across the A21-PSZ2 G114.90-34.35 pair.

5.3.3 Results for the pair A21-PSZ2 G114.90-34.35

The newly associated pair A21-PSZ2 G114.90-34.35 shows hints of SZ signal at $S/N_{\rm fil} \sim 2.5$, associated with the inter-cluster region. I have performed the same analysis on the SZ map as for A399-A401. Similarly, the parameters P_0 and β_{GNFW} of the two clusters are degenerated. Considering the lower significance of the SZ signal in this case, fitting the parameters of the model with the MCMC is more difficult. Therefore, I have further fixed the extension radius of the bridge $r_{\rm c}$. As this parameter is expected to be lower or equal to the clusters' extensions, I have set it to the smallest r_{500} of the two clusters, here $r_{\rm c}$=0.92 Mpc or $r_{\rm c}$=8.5 arcmin. The gas pressure of the bridge derived by the MCMC is in this case: $P_0 = 1.6^{+0.7}_{-0.3} \times 10^{-3}$keV.cm^{-3}. In the left panel of Fig. 5.6, I show the contributions of the three components of the system (clusters + inter-cluster bridge) computed with the best-fit parameters from the MCMC analysis, in a longitudinal cut across the SZ map. The reduced chi-square is in this case $\chi^2 = 0.96$, indicating a good agreement between the model and the data.

5.4 Galaxy properties in the cluster pairs

The SZ observations have allowed the detection of diffuse gas in between the clusters of the two pairs A399-A401 and A21-PSZ2 G114.90-34.35. They have allowed us to constrain the physical properties of the gas in the bridges of matter between the clusters. Here, I present the study of the galaxies in the three components (clusters and bridges) of the two pairs, and investigate the possible differences between their properties (galaxy types and star formation activities). I focus on the pair A399-A401 and summarise the main results on the pair A21-PSZ2 G114.90-34.35.

5.4.1 Catalogue of galaxies

As the redshift of the pair A399-A401 is quite low, at $z \sim 0.073$, the use of the WISExSCOS value-added catalogue constructed in Chap. 4 alone is not appropriate, as it has a lower flux limit: W1 > 13.8 (the median redshift of the WISExSCOS is $z = 0.2$). This limit cuts the brightest galaxies, that are the higher mass galaxies at low redshift, i.e., galaxies that are expected to be in the pair. However, using

the WISExSCOS catalogue is still useful as it contains lower mass galaxies at low redshift, with W1 > 13.8, potentially also lying in the pair. In order to also take into account the brightest galaxies at low z, I have used the 2MPZ photometric redshift catalogue. The 2MPZ catalogue, with a median redshift of $z = 0.08$, is complementary to the WISExSCOS as there is an upper limit in K < 13.9. Both catalogues are presented in Chap. 2. Following Bilicki et al., 2016, I have unified the two catalogues, assuming that the K band of 2MASS and the W1 band of WISE are equivalent for low redshift galaxies. In that way, the union catalogue is complete up to W1<17, and contains high and lower mass galaxies at the redshift of the galaxy cluster pair, i.e., $z \sim 0.073$. The union catalogue contains photometric redshifts estimations and WISE measurements. I have thus estimated SFR and $M_\star$ with the random forest algorithm presented in Chap. 4 for sources lying in the $3° \times 3°$ patch studied in SZ.

5.4.2 Identification of galaxy members

In order to study the properties of the galaxies in the pair, galaxies have to be associated with the three components (i.e., the two clusters and the bridge). I have first selected the galaxies in the redshift range $0 \lesssim z \lesssim 0.5$ within the $3° \times 3°$ field of the *Planck* MILCA SZ map (seen in Fig. 5.2). This first selection contains 6,487 sources. I have then refined the selection in order to focus on galaxies likely to belong to the A399-A401 pair. To do so, I have selected galaxies in the range $0.068 < z < 0.078$, corresponding to the redshift range between the lowest and the highest redshift of the cluster system plus and minus the velocity dispersion of the two clusters given in Oegerle and Hill, 2001. For simplicity and coherence with the SZ analysis, I have assumed that member galaxies are defined as those within an area of radius r_{500}, centred on the SZ cluster positions. The galaxies belonging to the inter-cluster bridge are defined to be between the positions c_y plus or minus the core radius r_c of the bridge, both fitted with the MCMC in SZ in Sect. 5.3. Galaxies outside these regions are considered as field galaxies. I show in the left panel of Fig. 5.7 the galaxies selected in the redshift range of the pair, i.e., $0.068 < z < 0.078$, identified to the three components of the pair A399-A401. The orange dots represent the field galaxies, whereas the green, blue, and red dots show the galaxies belonging to A399, A401, and to the inter-cluster bridge, respectively.

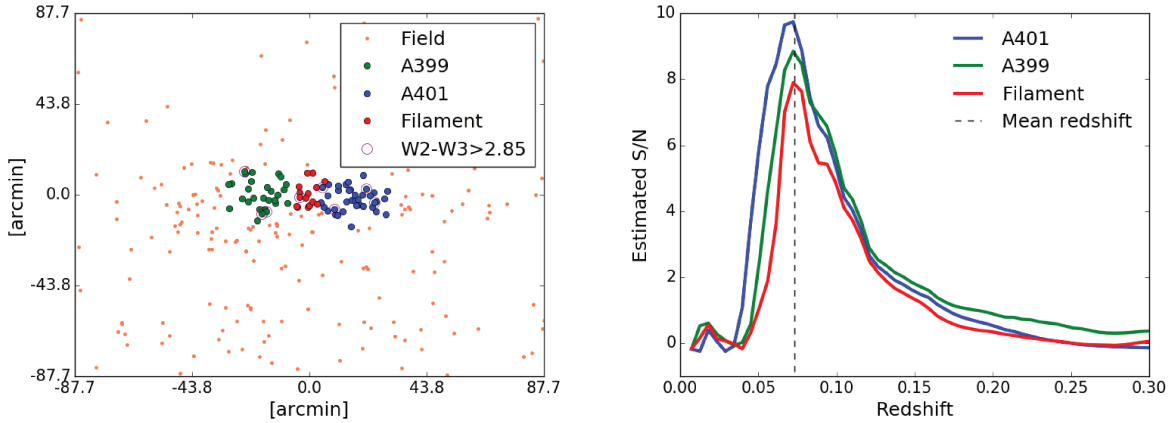

Figure 5.7: Left: Galaxies in the field of the pair A399-A401, in the redshift range $0.068 < z < 0.078$. The orange dots show the field galaxies. The green dots show the galaxies in the galaxy cluster A399, the blue ones the galaxies in A401, and the red dots those in the inter-cluster bridge. The purple circles indicate active galaxies. Right: Distributions of the S/Ns of over-densities for the three components of the pair A399-A401. The blue line corresponds to the cluster A401, the green line to A399, and the red one to the inter-cluster bridge.

5.4.3 Galaxy over-densities

Based on the pyDTFE python code (see Chap. 3.3.1), I have estimated the 3D galaxy density field around the pair A399-A401. I have computed the over-density of the galaxies in the range $0 \lesssim z \lesssim 0.5$, within the field of the *Planck* MILCA SZ map. For each redshift bin of width $\delta z = 0.005$, I have computed the mean over-density in the three regions that define the three components of the pair (i.e., the two clusters and the bridge). I have compared the mean density in the galaxy cluster regions and in the bridge to the mean density of the background and its standard deviation. In the right panel of Fig. 5.7, I show the distribution of the over-densities, noted S/N, as a function of redshift for the three components of the pair. It is clear that the S/Ns in the regions of the two galaxy clusters and the one of the inter-cluster bridge are peaking at the same redshift, that is the mean redshift of the pair. With the galaxies only, the bridge between the two clusters is detected with a high significance, i.e., $S/N_{\rm fil} \sim 8$, that is comparable with the significance of the detection of the gas with the SZ signal.

5.4.4 Galaxy properties

In order to study the different galaxy populations and types in the pair A399-A401, I have focused on several indicators: optical colours, infra-red colours, and SFR-$M_\star$ diagram.

5.4.4.1 Red sequence in optical

I have used the photometric bands of SuperCOSMOS, already provided in the union catalogue, to produce the colour-magnitude diagram B-R versus R, for the galaxies in the pair A399-A401. The resulting plot is shown in the left panel of Fig. 5.8. The galaxies in the three components of the system align in a red sequence. An interesting point is that the galaxies in the bridge are also following the red sequence. I have further investigated this result by studying infra-red properties and the passivity of the sources with their SFR and $M_\star$ estimated with the RF algorithm presented in Chap. 4.

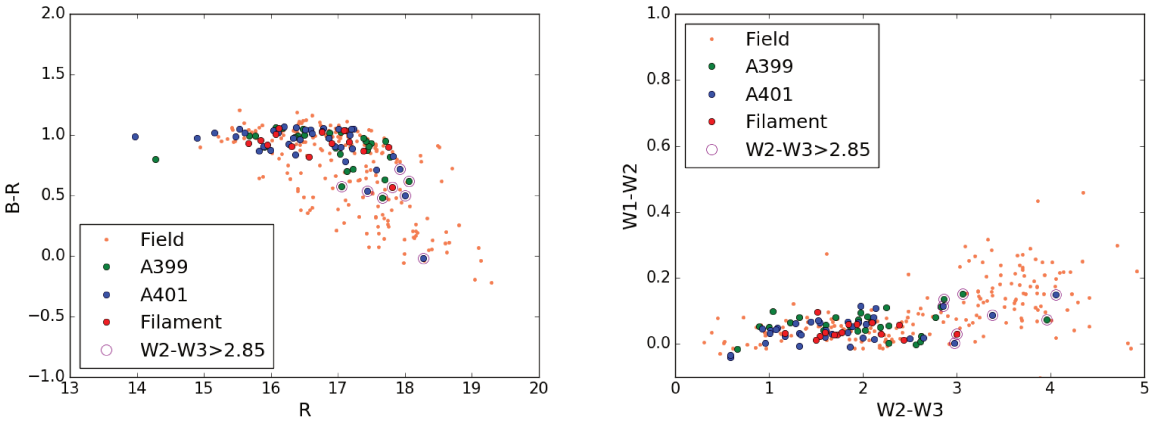

Figure 5.8: Left: Colour-magnitude diagram using B and R photometric bands from SuperCOSMOS. The labels are the same as those in Fig. 5.7. Right: Colour-colour diagram from the photometric bands W1, W2, and W3 of WISE.

5.4.4.2 Morphology in the infra-red

As discussed in Chap. 4.2.1, Wright et al., 2010 have studied the correlation between galaxy morphologies and their positions in the colour-colour diagram based on the W1-W2 and W2-W3 near infra-

red colours from WISE. Elliptical E and spiral S0 galaxies, both passive, are in the range 0.5<W2-W3<1.5 and -0.1<W1-W2<0.3, while active spiral galaxies are in the range 1<W2-W3<4.5 and -0.1<W1-W2<0.7. In a similar way, I show in the right panel of Fig. 5.8 the colour-colour diagram based on the same colours: W1-W2 and W2-W3. The galaxies identified in the three components of the pair A399-A401 are mostly located in the region probing E and S0 galaxies, i.e., passive galaxies.

5.4.4.3 SFR and $M_\star$

I further show in Fig. 5.9 the SFR-$M_\star$ diagram for galaxies in the pair A399-A401. The SDSS main sequence of star-forming galaxies fitted by Brinchmann et al., 2004 is shown in black solid line. The contours show the galaxies of the full SDSS MPA-JHU DR8 catalogue at the same slice of redshift, i.e., $0.068 < z < 0.078$. Almost all the galaxies in the three components of the pair are passive, including the galaxies in the bridge. This result confirms the results obtained both with the red sequence in optical and the morphological identification in infra-red.

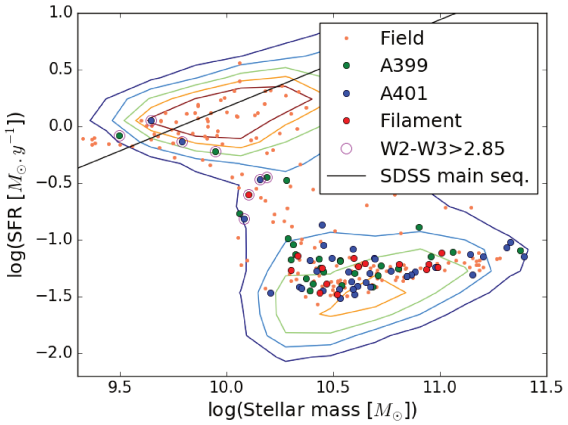

Figure 5.9: SFR-$M_\star$ diagram for the galaxies in the field of A399-A401. The labels are the same as in Fig. 5.7. The black line shows the main sequence fitted by Brinchmann et al., 2004 on the SDSS galaxies. The contours show the galaxies from the SDSS MPA-JHU DR8 catalogue, in the redshift range of the pair, i.e., $0.068 < z < 0.078$.

5.4.5 Results for the pair A21-PSZ2 G114.90-34.35

I have selected sources in the same way as for the pair A399-A401 and computed the 3D density field in the region of the pair A21-PSZ2 G114.90-34.35. I have found an over-density with a S/N of about 2.5 in the inter-cluster region, that is comparable with the significance of the detection of the gas in SZ (see Fig. 5.10). Due to the lack of statistics for this cluster system, a study of the galaxy properties and a conclusion on their types are not possible.

5.5 Discussion

First observations of the pair A399-A401 in the X-rays with the Einstein telescope (Karachentsev and Kopylov, 1980; Ulmer and Cruddace, 1981) did not show any evidence of signal between the two clusters because of the high background noise. It is only with ASCA and ROSAT that Fujita et al., 1996 and Fabian, Peres, and White, 1997 detected an excess of X-ray emission that they associated with a diffuse

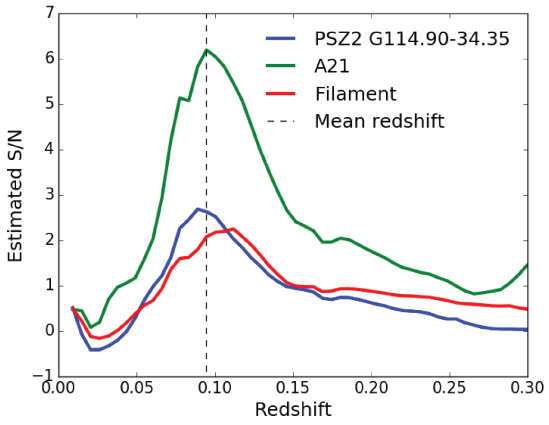

Figure 5.10: Distribution of the S/Ns of over-densities for the three components of the pair A21-PSZG114.90-34.35. The blue line corresponds to PSZ2 G114.90-34.35, the green line to A21 and the red line to the region between the two clusters.

gas between the two clusters. Given that the two clusters, A399 and A401, lacked cooling-flows and contained diffuse radio halos (Murgia et al., 2010), the gas between the two clusters was interpreted as the result of a post-merger (Fabian, Peres, and White, 1997). Alternatively, Fujita et al., 1996 suggested a pre-merger scenario to explain the origin of the X-ray emission, since the temperatures of the gas and the two clusters were of the same order. Another explanation for the presence of a bridge between the two clusters is that the whole system has formed at the same time and that the gas in the inter-cluster region is a relic of a cosmic filament, as proposed by Akamatsu et al., 2017. The higher quality XMM-Newton data from the four pointing observations of A399-A401 have permitted to measure the gas properties of the two clusters and of the bridge between them (Sakelliou and Ponman, 2004). This revealed both the absence of big anomalies in the temperature profiles and the relaxed nature of the clusters, with no significant offset between the positions of their BCGs and the central positions of the X-ray emissions. Sakelliou and Ponman, 2004 further interpreted the radio halos in the two clusters as relics of past minor mergers in each cluster. All these evidences strongly favour a pre-merger scenario that has been reinforced by the Suzaku observations conducted by Fujita et al., 2008, who found a high metallicity in the inter-cluster region, 0.2 $Z_{\odot}$, of the same order of the metallicity as in the clusters, indicative of superwinds that transported metals from the clusters to the inter-cluster region. More recently, Govoni et al., 2019 have observed the pair in radio with LOFAR. They have detected the bridge in radio emission, suggesting the presence of magnetic field, re-accelerating former relativistic electrons.

In order to investigate the different scenarios and discriminate between them, it is necessary to fully characterise the properties of the inter-cluster bridge. In this study, I have thus completed the analysis of the system A399-A401 by exploring the SZ signal and the galaxy properties. I have confirmed the presence of gas between the two clusters A399 and A401 with a significance of approximately 8.5σ from the *Planck* MILCA SZ data alone. I have measured the pressure of the inter-cluster gas, $P_0 = (2.8 \pm 0.27) \times 10^{-3}$keV.cm^{-3}, in agreement with the pressure computed with the density and the temperature fitted from the combined SZ/X-rays analysis by Planck Collaboration et al., 2013b: $P_0 = (2.6 \pm 0.5) \times 10^{-3}$keV.cm^{-3}. Using the most accurate gas temperature of Fujita et al., 2008 and Akamatsu et al., 2017, $k_{\rm b}T_{\rm X} = 6.5 \pm 0.5$keV, I have further estimated a central density of $n_0 = (4.3 \pm 0.7) \times 10^{-4}$cm^{-3}, in agreement with Planck Collaboration et al., 2013b.

Akamatsu et al., 2017 have also studied the geometry of the bridge between A399 and A401. They have considered a uniform density along the line of sight and have estimated a Compton parameter in

the bridge, $y = 14.5 \pm 1.8\left(\frac{l}{\mathrm{Mpc}}\right)^{0.5} \times 10^{-6}$, where l is an effective depth of the bridge along the line of sight. By comparing the obtained Compton parameter with a weight-averaged y parameter in the SZ map (roughly estimated at $y = 14 - 17 \times 10^{-6}$ in Planck Collaboration et al., 2013b), they have deduced an effective density $n_0 = 3.1 \times 10^{-4}\mathrm{cm}^{-3}$ and an effective depth $l = 1.1$ Mpc. Akamatsu et al., 2017 have compared this depth to the size of the bridge in the radial direction, ~ 2.6 Mpc (compatible with the result of my analysis, a size of 3.0 ± 0.2 Mpc), and have concluded that the bridge is flattened. Following their method, I have focused on the very central region of the bridge (within 2' of the longitudinal axis) and have estimated a Compton parameter $y = 17.2 \pm 1.3\left(\frac{l}{\mathrm{Mpc}}\right)^{0.5} \times 10^{-6}$, with $k_\mathrm{b}T_\mathrm{X} = 6.5 \pm 0.5$ keV and $n_0 = 3.3\left(\frac{l}{\mathrm{Mpc}}\right)^{-0.5} \times 10^{-4}\mathrm{cm}^{-3}$ (see their Fig. 7). I have compared this value to the value of the SZ map, $y = 22.2 \pm 1.8 \times 10^{-6}$, and have estimated an effective depth of $l = 1.7 \pm 0.5$ Mpc. This value suggests that the shape of the bridge is consistent with a cylinder, compatible with the hypothesis behind the model chosen for the bridge in this analysis. However, the computations of the bridge depth l strongly depend on the model of the electron density and on the assumed value of y.

The re-analysis of the Suzaku data by Akamatsu et al., 2017 has shown hints of a shock in the direction parallel to the one linking the two clusters. Such a shock would be incompatible with a merger scenario of two clusters only, since numerical simulations predict shocks in the radial direction (Akahori and Yoshikawa, 2008). This suggests a pre-existing cosmic filament, an hypothesis supported by Planck Collaboration et al., 2013b. The analysis presented in this chapter is complementary and brings additional information on the scenario. The SZ emission traces and detects the diffuse gas, with a density one order of magnitude below the mean densities in the clusters. In addition, I have detected a galaxy over-density between the clusters A399 and A401 with a significance of approximately 8σ. The colour-colour, colour-magnitude, and SFR-$\mathrm{M}_\star$ diagram of the galaxies selected in the pair show that they are all passive. No segregation is seen between the galaxies belonging to the three components of the pair. The properties of the galaxies in the pair A399-A401 are those of the typical populations of galaxies in clusters or in dense collapsed structures. This contrasts with the results showing a large fraction of star-forming galaxies in intermediate-density environments such as filaments (Gallazzi et al., 2009; Edwards et al., 2010). It suggests that the mechanisms by which galaxies can undergo a quenching of star formation (e.g. mergers, harassment, ram pressure, etc.) are more efficient in the A399-A401 system which could be related to the hot and pressured gas observed in the pair in SZ, X-rays, and radio.

5.6 Conclusion

I have performed an SZ-based selection of galaxy-cluster pairs showing hints of SZ signal potentially associated with inter-cluster bridges. Among the 71 pairs satisfying the selection criteria on redshift and angular distance separations, I have selected the systems at the highest significance: A399-A401 at redshift $z = 0.073$ with $S/N_\mathrm{fil} = 8.74$, and A21-PSZ2 G114.90-34.35 at redshift $z = 0.094$ with $S/N_\mathrm{fil} = 2.53$. For these two systems, I have performed a multi-wavelength analysis that constrained the gas properties, and I have studied the galaxy properties in the three components of the pair, defined by the two clusters and the inter-cluster bridge.

For the most significant pair, A399-A401, I have measured a gas pressure in the inter-cluster bridge only with the SZ effect, $P_0 = (2.84 \pm 0.27) \times 10^{-3}\mathrm{keV.cm}^{-3}$, that is in agreement with previous studies that used both X-ray and SZ measurements (Planck Collaboration et al., 2013b). The types, morphologies, and star-formation activities of the galaxies in the clusters and in the inter-cluster bridge can not be differenciated: they are all passive. The analysis of the galaxy properties weakens the post-merger hypothesis which was already disfavoured by the lack of a big offset between the positions of the BCGs and the centres of the X-ray emissions. This study, alternatively, favour the scenario in which the gas between the two clusters is associated with a former cosmic filament. The gas is collapsing, smoothly

compressed, and heated by the collision of the two clusters A399 and A401. The current data are however not deep enough to accurately measure the effect of environmental quenching in the bridge that connects the two clusters. However, since the recent observation of this pair in LOFAR by Govoni et al., 2019, a selection of radio sources can be performed in the bridge. A proposal for observing time with ALMA is being prepared in collaboration with G. Castignani (EPFL), P. Salomé (LERMA), F. Combes (Collège de France), N. Aghanim (IAS), and P. Jablonka (EPFL). We will select specific sources based on the SFR in order to study the influence of the hot gas environment on the galaxies.

For the newly associated pair of clusters A21-PSZ2 G114.90-34.35, I have detected an inter-cluster bridge at 2.5σ from the analysis of the *Planck* MILCA SZ map and from the 3D galaxy density field. However, a significant detection of the bridge or a detailed study of galaxy properties around the pair is not possible due to the lack of statistics. Dedicated observations in the X-rays, with a higher-resolution SZ instrument, or optical or near infra-red wide-field spectroscopic data are necessary to confirm the presence of inter-cluster bridge.

More complex systems, super-clusters, with high significance inter-cluster SZ signal were identified in my sample and should be subject to future dedicated analyses: the Shapley super-cluster (in collaboration with N. Aghanim and T. Bonnaire) and the super-cluster A3395-A3391. This system is in the footprint of the PACT *y* map (Aghanim et al., 2019), a high resolution SZ map obtained combining ACT and *Planck* frequency channels. The high resolution of the PACT map should allows to have a more detailed analysis of the gas properties of A3395-A3391.

Chapter 6

Galaxies around cosmic filaments

Contents

Abstract

The role played by the large-scale structures in the galaxy evolution is not quite well understood yet. In this chapter, I investigate the galaxy properties in the range $0.1 < z < 0.3$ from the WISExSCOS value-added catalogue presented in Chap. 4.6.1 around the cosmic filaments detected with DisPerSE. I also link the galaxy properties and the gas content in the Cosmic Web. I have fitted a profile of galaxy over-density around cosmic filaments with a typical radius $r_{\rm m} = 7.4 \pm 0.1$ Mpc. I have measured an excess of passive galaxies near the filament's spine, higher than the excess of transitioning and active galaxies. I have also detected SFR and $M_\star$ gradients pointing towards the filament's spine. I have investigated this result and found an $M_\star$ gradient for each type of galaxies: active, transitioning, and passive, and a positive SFR gradient for passive galaxies. I have investigated the quiescent fraction $f_{\rm Q}$ profile of galaxies around the cosmic filaments. Based on recent studies of the role of the gas and of the Cosmic Web on galaxy properties, I have modelled $f_{\rm Q}$ with a β model of gas pressure. The slope obtained here, $\beta = 0.54 \pm 0.18$, is compatible with the scenario of projected isothermal gas in hydrostatic equilibrium ($\beta = 2/3$), and

with the profiles of gas fitted in SZ ($1/3 < \beta < 2/3$). The material presented in this chapter will be used for a future publication: *"Quenching of WISExSCOS galaxies around SDSS filaments"*, V. Bonjean, N. Aghanim, M. Douspis, N. Malavasi, and H. Tanimura, in prep.

Résumé

Le rôle joué par les structures à grande échelle dans l'évolution des galaxies n'est toujours pas bien compris à ce jour. Dans ce chapitre, j'étudie les propriétés des galaxies dans la gamme de redshift $0,1 < z < 0,3$ du catalogue de valeur ajoutée WISExSCOS présenté au Chap. 4.6.1, autour des filaments cosmiques détectés avec DisPerSE. J'ai étudié le lien entre les propriétés des galaxies et le contenu en gaz dans la toile cosmique. Pour ce faire, j'ai ajusté un profil de surdensité de galaxies autour des filaments cosmiques, et ai obtenu un rayon typique de $r_{\rm m} = 7,4 \pm 0,1$ Mpc. J'ai mesuré un excès de galaxies passives dans les régions internes des filaments. J'ai aussi mesuré des excès de galaxies en transition et actives, mais inférieur à celui des galaxies passives. J'ai aussi détecté des gradients de SFR et de $M_\star$ pointants vers les régions internes des filaments. J'ai investigué ce résultat en séparant les types de galaxies, et j'ai aussi mesuré un gradient de $M_\star$ pour chaque type : active, en transition, et passive, et un gradient de SFR positif pour les galaxies passives. J'ai ensuite étudié le profil de fraction de galaxies éteintes, $f_{\rm Q}$, autour des filaments cosmiques. Basé sur des études récentes de la distribution du gaz autour des filaments de la toile cosmique, j'ai modélisé $f_{\rm Q}$ avec un modèle β de pression du gaz. La pente obtenue ici, $\beta = 0,54 \pm 0,18$, est compatible avec celle d'un profil projeté d'un gaz isotherme en équilibre hydrostatique ($\beta = 2/3$), et compatible avec les profils de gaz détectés en SZ ($1/3 < \beta < 2/3$). Le matériel présenté dans ce chapitre sera utilisé dans le cadre d'une future publication : *"Quenching of WISExSCOS galaxies around SDSS filaments"'*, V. Bonjean, N. Aghanim, M. Douspis, N. Malavasi, et H. Tanimura, en préparation.

6.1 Introduction

While galaxy clusters are relatively easy to study, to characterise, and even to detect, other Cosmic Web structures, like cosmic filaments, are not easily defined because of their low densities. No global picture such as the one drawn for the galaxy clusters has been derived yet, and even the definition of cosmic filaments is still arguable since it depends on the way they are detected. Several methods have been proposed to detect the cosmic filaments, for example Bisous (Tempel et al., 2016), DisPerSE (Sousbie, 2011), NEXUS+ (Cautun, van de Weygaert, and Jones, 2013), or very recently T-ReX (Bonnaire et al., in prep.). Each of them has advantages and disadvandages, making comparison between them really difficult (see Libeskind et al., 2018 for a review or Bonnaire et al., in prep.). Despite this, recent studies have used those methods to investigate the physical properties of the matter (Dark Matter, gas, or galaxies) in the filaments (e.g., Colberg, Krughoff, and Connolly, 2005; Dolag et al., 2006; Aragón-Calvo, van de Weygaert, and Jones, 2010; Cautun et al., 2014; Gheller et al., 2015; Gheller et al., 2016; Martizzi et al., 2019; Gheller and Vazza, 2019, Galárraga et al., in prep.). But studying the gas is only possible in numerical simulations. Indeed, the hot gas around cosmic filaments is very hard to detect either in SZ or in X-rays because of its low density and low temperature. It is only accessible in a few exceptional objects such as the galaxy cluster pair A399-A401 presented in Chap. 5. Alternatively, some studies have used the stacking around the highest density regions (in between galaxy cluster pairs (Tanimura et al., 2019b; de Graaff et al., 2019) or inside super-clusters (Tanimura et al., 2019a)) to detect the densest parts of the gas in the filaments around clusters. A first statistical study of the hot gas using the SZ effect around cosmic filaments is performed in Tanimura et al., in prep.

Galaxies around cosmic filaments, easier to detect, have started to be extensively studied recently in different surveys: in SDSS (e.g., Martínez, Muriel, and Coenda, 2016; Chen et al., 2017; Kuutma,

Tamm, and Tempel, 2017, at $z = 0.1$, $z < 0.7$, and $z < 0.7$ respectively), in GAMA (e.g., Alpaslan et al., 2015; Alpaslan et al., 2016; Kraljic et al., 2018, at $z < 0.2$, $z < 0.2$, and $0.03 < z < 0.25$ respectively), in CFHTLS[1] (e.g., Sarron et al., 2019, at $0.15 < z < 0.7$), in VIPERS (e.g., Malavasi et al., 2017, at $z = 0.7$), and in COSMOS (e.g., Laigle et al., 2018, at $0.5 > z > 0.9$). These studies show evidence of galaxy population segregation inside filaments, hints of pre-processing and quenching process of galaxies while entering the large scale structures, and of a positive stellar mass gradient pointing towards the filament's spines. Although these trends are detected in different studies, the mechanisms responsible of these processes are not understood yet, and the role of the environment on the evolution of the galaxies is still not clear.

In this chapter, I present a statistical study of galaxy properties from the value-added catalogue presented in Chap. 4, around cosmic filaments at low redshift (in the range $0.1 < z < 0.3$) extracted with DisPerSE in a spectroscopic sample of galaxies from the SDSS. I show the statistical distributions of all galaxies, and of passive, transitioning, and active galaxies around the cosmic filaments, together with their stellar mass and SFR profiles. I also investigate the role of the Cosmic Web and of the hot gas on the galaxy quenching around the cosmic filaments, by linking the quiescent fraction profile to a profile of hot gas.

6.2 The catalogue of filaments

6.2.1 Cosmic filaments from LOWZ/CMASS

The BOSS survey is the only extra-galactic survey in the third stage of the SDSS (SDSS-III). It combines the LOWZ and the CMASS catalogues, probing the low redshift ($0.15 < z < 0.45$) and the intermediate redshift ($0.4 < z < 0.7$) Universe, respectively. These two catalogues are extensions of the primary samples of SDSS galaxies to fainter regions and to bluer galaxies, in order to increase the spatial number density by a factor of three[2]. This sample was used by N. Malavasi (IAS, Orsay) to construct a catalogue of cosmic filaments using DisPerSE (Sousbie, 2011). I show in Fig. 6.1 the spatial distribution of the filaments overlaid with the SDSS galaxies density field. The details of the construction of the catalogue are given in Malavasi et al., in prep.

DisPerSE is an algorithm that detects filaments in catalogues of discrete points. The method is based on the density field, reconstructed with the Delaunay Tessellation Field estimator (DTFE, described in Chap. 3.3.1). DisPerSE detects critical points in the density field, that can be either maximum density points, minimum density points, saddle points or bifurcation points. The output of DisPerSE is thus a catalogue of filaments, defined as connecting the saddle points to the maximum density points. Each filament is constructed by several small segments, with given positions and redshifts. DisPerSE also outputs a persistence, that is a quantity related to the significance of the filament's detections, that was set to 3 σ in this case. I have defined the mean positions of the filaments $(\mathrm{R.A.}_{\mathrm{mean}}, \mathrm{Dec.}_{\mathrm{mean}})$ being the mean of the positions $(\mathrm{R.A.}_i, \mathrm{Dec.}_i)$ of the i segments. In the following, I have defined the minimum, mean, and maximum redshifts of the filaments, z_{min}, z_{mean}, and z_{max}, as the minimum, the mean, and the maximum redshifts of the segments composing the filaments.

6.2.2 Selection of DisPerSE filaments

Since I study the properties of the WISExSCOS value-added catalogue, I have thus excluded the filaments for which parts went outside the redshift range of the WISExSCOS catalogue, that is $0.1 < z < 0.3$, to ensure that the filaments of our selection are entirely studied. I have also cut the longest filaments at

[1] https://www.cfht.hawaii.edu/Science/CFHTLS/

[2] https://www.sdss.org/dr15/spectro/extragalactic-observing-programs/

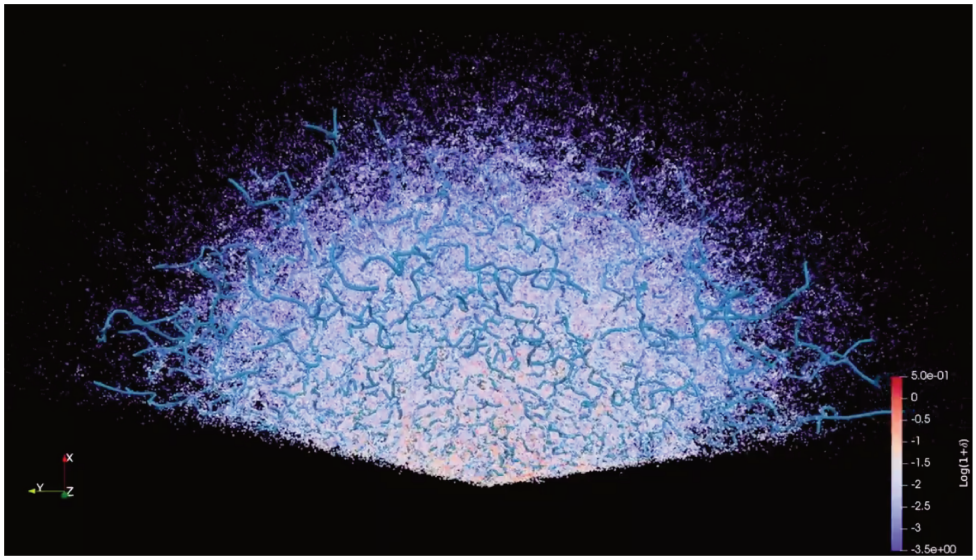

Figure 6.1: Image showing the filaments extracted with DisPerSE (in blue) on SDSS galaxies (in the SDSS DR12 footprint). The density field of the galaxies is overlaid, where the red colour indicates the densest regions. Credits: N. Malavasi.

$l > 100$ Mpc (where l is the 3D length of the filament in Mpc), that may be non reliable filaments. In order to keep the same statistics in each bin of the profiles I compute later, I have also excluded the filaments for which profiles were not complete. After this selection, there are $n = 5559$ cosmic filaments in the sample I use for my analysis. Their length and redshift distributions are shown in Fig. 6.2.

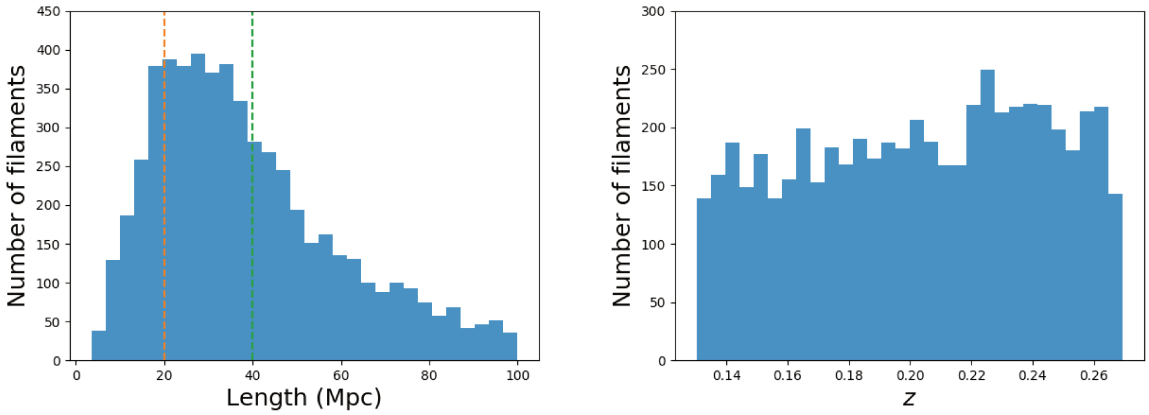

Figure 6.2: Selection of the SDSS-DR12 LOWZ/CMASS DisPerSE filaments used here. Left: 3D length distribution. The orange dotted line shows the limit between the short and the regular filaments, and the green dotted line shows the limit between the regular and the long filaments. Right: redshift distribution.

6.2.3 Separation by length

I have split the 5559 selected DisPerSE filaments over three bins defined by their 3D length l (in Mpc), to later study the effect of the environment on galaxy properties around the filaments. In the following, I define “short”, “regular”, and “long” filaments, as follows:

- $0 < l < 20 : n_{\rm obj} = 1042$,
- $20 < l < 40 : n_{\rm obj} = 2291$,
- $40 < l < 100 : n_{\rm obj} = 2226$.

For the case of short filaments, I have set the upper limit to 20 Mpc based on the information given by the 2-point correlation functions of groups in the 2DF survey (Yang et al., 2005), that infer that filaments below this typical size may be the tiniest and densest filaments, i.e., bridges of matter connecting two clusters, such as the bridge between A399 and A401, presented in Chap. 5. I have set the limit between the regular and the long filaments arbitrarily to 40 Mpc to keep the same statistics in the two categories (about 2250 filaments in each).

6.3 Measuring the profiles

6.3.1 Methodology

In order to measure the radial profiles of galaxy quantities around filaments, I have used the RadFil code (presented in Chap. 3.3.2). To do so, RadFil takes as input two maps: one of the quantity of interest, here the maps of galaxy densities constructed in Chap. 4.6, and one tracing the spine of the filament around which it will perform the measurement, here the maps tracing the spines of the selected DisPerSE filaments. In order to obtain the radial profiles, few steps were needed, explained hereafter.

- Normalisation of the maps

  The galaxies in the WISExSCOS value-added catalogue are not uniformly distributed in redshift, and the three population of galaxies do not follow the same distributions, as it is shown in Fig. 4.16. Thus, the values of the mean galaxy densities in each redshift slice of the galaxy density maps constructed in Chap. 4.6 also follow the same redshift distribution. Measuring the absolute value of the galaxy densities may thus introduce bias. To avoid this, and to measure only the excess of galaxies relative to the mean galaxy density in the field, I have normalised the 3D density maps in order to consider over-densities δ. I have divided each slice of redshift of each 3D map by their mean galaxy density values:

  $$1 + \delta_{\rm gal}(z) = \frac{\rho_{\rm gal}(z)}{< \rho_{\rm gal}(z) >}. \tag{6.1}$$

  In that way, the 3D maps are transformed from biased densities $\rho_{\rm gal}$ to unbiased over-densities $1 + \delta_{\rm gal}$.

- Projection on patches

  For each of the 5559 filaments, I have projected the 3D maps of the obtained galaxy over-density $1 + \delta_{\rm gal}$, on 3D patches centred on the position of the corresponding filament using a tangential projection. The 3D patches have a pixel resolution of $\theta_{\rm pix} = 1.7$ arcmin, a bin in redshift of 0.01 (same as the full-sky maps), and a number of pixels which depends on the length of the filament (computed in arcmin with the mean redshift $z_{\rm mean}$). Doing so, all filaments are entirely seen in their corresponding individual patch and the fields of view of the largest patch is $19^\circ \times 19^\circ$. As 95% of the patches have fields of view below $15^\circ \times 15^\circ$, I have neglected the projection effects and assumed the flat sky approximation.

- Stack along redshift

  Due to the high value of the statistical error on the photometric redshifts of the sources in the WISExSCOS catalogue, $\sigma_z = 0.033$, 3D density profiles around filaments would be biased. Therefore, I have stacked the 3D patches (obtained above) along the redshift axis to remove the uncertainty on the positions of the galaxies in the redshift space. The resulting stacked maps are thus 2D arrays. Before stacking along redshift, in order to minimise the noise due to background and foreground galaxies, I have removed for each filament the regions of the 3D patch that lie outside the redshift range $z_{\rm min} - \sigma_z < z < z_{\rm max} + \sigma_z$, where $z_{\rm min}$ and $z_{\rm max}$ are the minimum and maximum redshifts of the filament. Mathematically, this step translates into:

$$< 1 + \delta_{\rm gal} >= \frac{1}{b_z} \sum_{z=z_{\rm min}-\sigma_z}^{z_{\rm max}+\sigma_z} \Big(1 + \delta_{\rm gal}\Big)(z), \tag{6.2}$$

  where b_z is the number of redshift bins in the range $[z_{\rm min} - \sigma_z, z_{\rm max} + \sigma_z]$.

- Application of RadFil

  I have fed RadFil with the 5559 2D stacked maps obtained above, together with the 5559 associated 2D filament's spine projections, also in the format of 2D arrays. RadFil then measures the radial profiles $< 1 + \delta_{\rm gal} > (r)$ around each of the 5559 filaments.

- Stack the profiles

  Finally, I have stacked the 5559 profiles to get one unique profile, exhibiting statistical trends thanks to the significant reduction of the noise.

- Estimation of the error bars

  To estimate the errorbars on the stacked profiles, I have used the bootstrap method (detailed in Chap. 3.2.3). For the n measured profiles (where n is the number of filaments), I have randomly selected n over n profiles with replacement, and have computed the mean profile. I have repeated this measurement 1000 times and have computed the mean and the standard deviation of the 1000 mean profiles. The mean and standard deviations are taken as final measurements and errors in this study.

6.3.2 Masking the galaxy cluster members

In order to measure galaxy over-density profiles along filaments uncontaminated by the galaxy cluster members, I have removed the regions around known galaxy clusters, by masking the maps at the position of all clusters with $z < 0.4$ from the PSZ2 catalogue (presented in Chap. 2.4.2), the MCXC catalogue (presented in Chap. 2.4.3), the RedMaPPer catalogue (presented in Chap. 2.4.1), and from other catalogues of galaxy clusters detected in the SDSS: AMF9 (Banerjee et al., 2018), WHL2012 (Wen, Han, and Liu, 2012), and WHL2015 (Wen and Han, 2015).

To mask the clusters in an optimal way, I have defined six masks: the first one where the galaxy clusters are not masked, and the five others where the clusters were masked in regions from $1 \times R_{500}$ to $5 \times R_{500}$. The results are shown in Fig. 6.3. The galaxy over-density profiles decrease with increasing radius of the mask, up to $r = 3 \times R_{500}$. Beyond this radius, the profiles are unchanged but the error-bars increase. I have thus chosen to mask the clusters at $r = 3 \times R_{500}$. For clusters without estimated R_{500} (only a handful from the *Planck* catalogue in the SDSS area), I have masked a region with radii increasing up to $r = 10$ arcmin. I show in Fig. 6.4 that masking at $r = 5$ arcmin is enough, as no difference in the profiles is noticed. I have also masked regions around the critical points provided by DisPerSE, namely

the maxima density points, and the bifurcation points, with $z < 0.4$. For these regions, I have masked regions with areas defined by radii varying up to $r = 20$ arcmin (see Fig. 6.5). For the critical points, the profiles are decreasing with the increasing radii up to 20 arcmin. Part of the reason for the decrease of the profiles is due to the loss of the shortest filaments in the selection (i.e., a change in the length distribution seen in the left panels). I have chosen to mask the critical points from DisPerSE at $r = 10$ arcmin, as it is a good compromise to keep the shortest filaments of about $l < 7$ Mpc in the selection of filaments while removing contamination from galaxies in small groups.

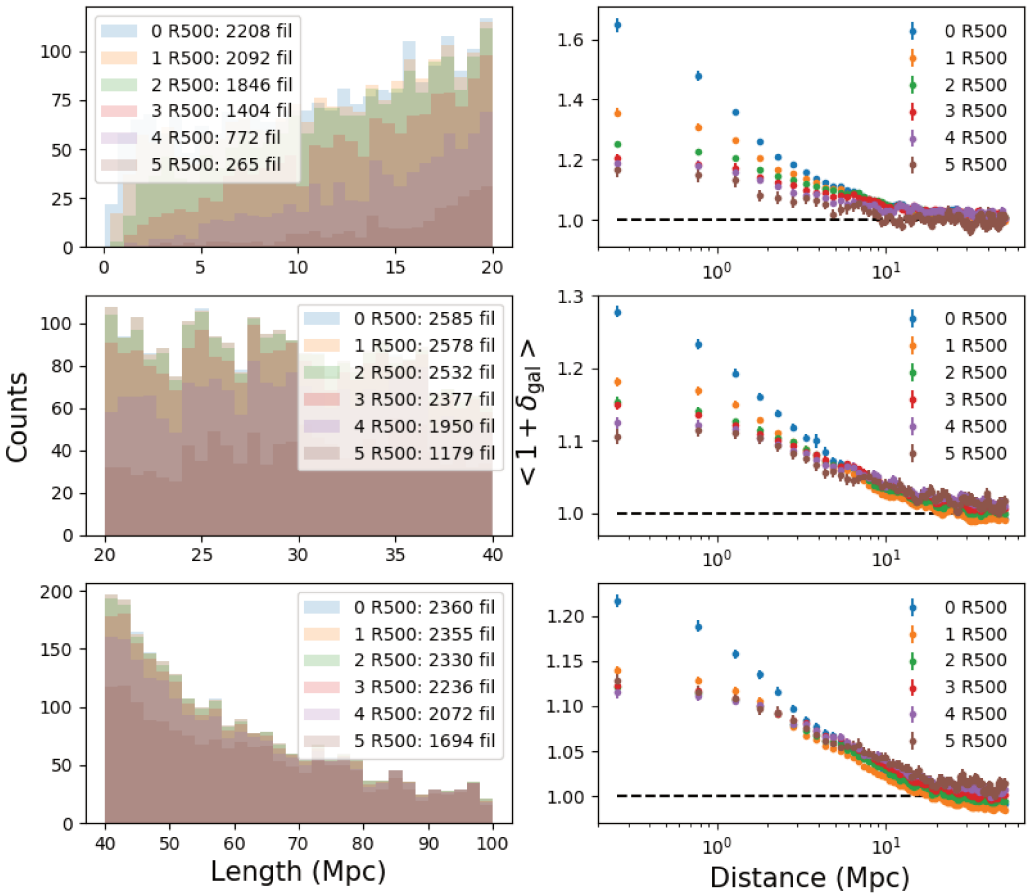

Figure 6.3: Comparison of the effect of different masks. Galaxy clusters with $z < 0.4$ from the PSZ2, MCXC, RedMaPPer, AMF9, WHL12 and WHL15 catalogues are masked with areas of radii from 0 to $5 \times R_{500}$. The left and the right columns show the histogram of filament's lengths and the over-density profiles $< 1 + \delta_{gal} >$ respectively, for the short, regular, and long filaments (from top to bottom). I have chosen the mask at $3 \times R_{500}$.

Finally, I have also masked the area outside the footprint of the SDSS-DR12 area. The union mask is shown in orthographic projection in Fig. 6.6 (only the northern hemisphere is shown, as none of the selected filaments are in the southern hemisphere).

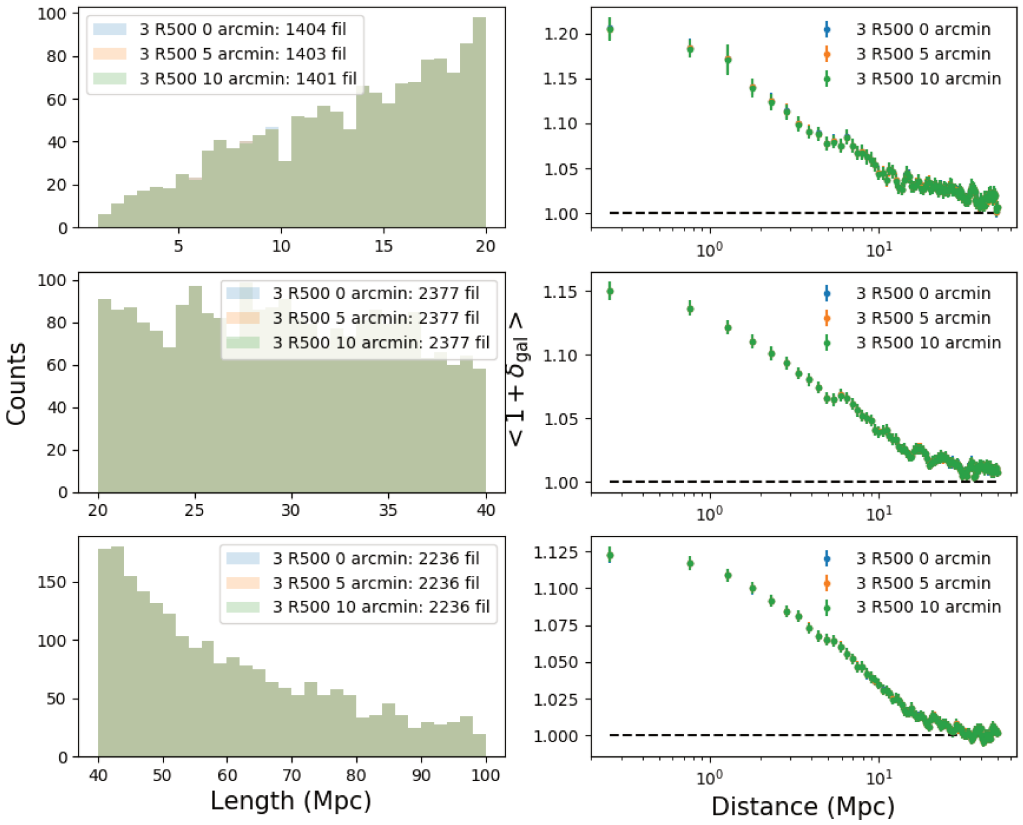

Figure 6.4: Comparison of the effect of different masks. For the galaxy clusters without estimated radius in the *Planck* PSZ2 catalogue with $z < 0.4$, regions defined by areas of 0, 5, and 10 arcmin radii are masked. No difference is noticed and all plots are overlapping. I have therefore chosen to mask at 5 arcmin.

6.3.3 Robustness of the measurement

To ensure the robustness of the measurements and compute their significance, I have performed null-tests. By applying the exact same methodology as the one detailed above, I have computed profiles in patches with the same properties as the ones of the 5559 DisPerSE filaments (number of pixels $\theta_{\rm pix}$ and redshift range $z_{\rm min} - \sigma_z < z < z_{\rm max} + \sigma_z$), but centred on random positions on the sky, and with random rotations. For computational time reasons, I have computed only 100 random position stacked profiles. The profiles are all detected with very high significances, ranging between $\sim 5\,\sigma$ and $32\,\sigma$.

Although the profiles are very significantly detected, there are some effects that may produce biases and trends that may not be explicitly related to galaxies around filaments. I discuss the possible sources of contamination hereafter.

One of the main issues is the error on the photometric redshifts of the sources in the WISExSCOS catalogue, $\sigma_z = 0.033$. This error may induce a bias during the construction of the 3D density maps presented in Chap. 4.6 and make them not reliable in the redshift direction. To mitigate this effect, I have computed the profiles in 2D projected maps stacked in the range $[z_{\rm min} - \sigma_z, z_{\rm max} + \sigma_z]$, where $z_{\rm min}$ and $z_{\rm max}$ are the minimum and maximum redshifts of the filaments.

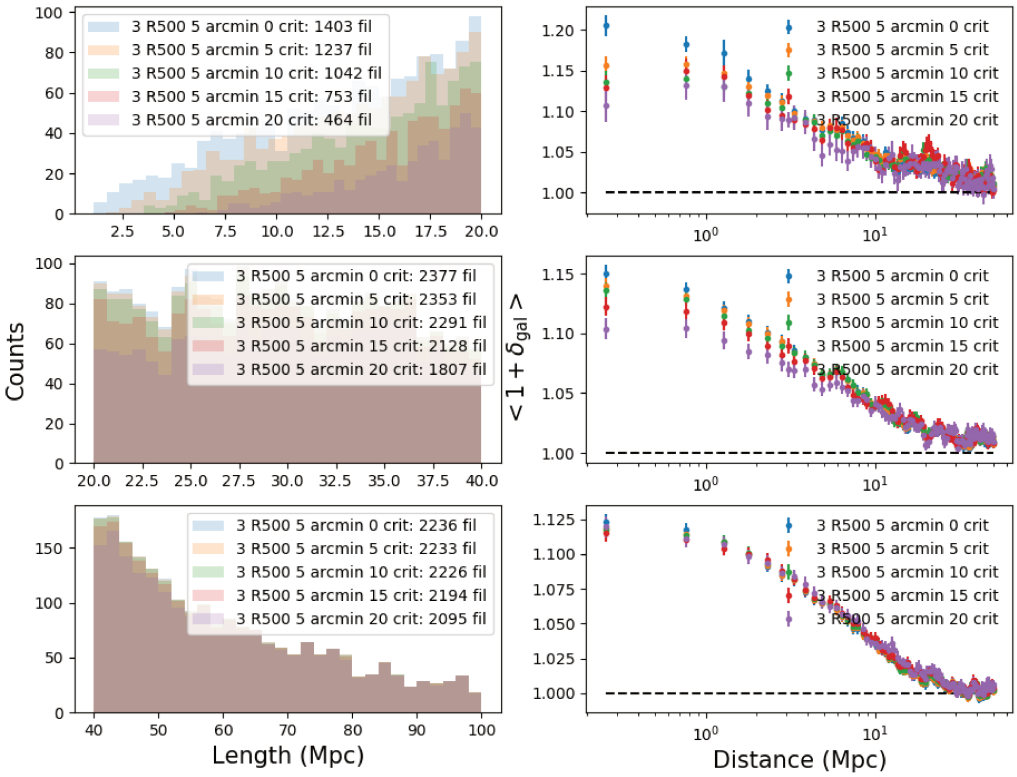

Figure 6.5: Comparison of the effect of different masks. Two type of critical points given by DisPerSE are masked: the maxima density points and the bifurcation points. The regions around the critical points with $z < 0.4$ are masked from 0 to 20 arcmin. Right panel, the signal is decreasing with the size of the masks for short and regular filaments. Left panel, the masks remove the short filaments.

Another possible source of bias may be the 3D to 2D projection of the over-densities. RadFil computes profiles on 2D while the filaments are defined as 3D structures. This may induce a bias in the shape of the profiles, especially when the axis of the filaments are not orthogonal to the line of sight.

Another issue is the reliability of the filaments detected with DisPerSE. Indeed, although the LOWZ/CMASS sample was chosen to minimise these effects, the SDSS galaxies used for the detection may still suffer from holes in the spatial distribution due to foreground bright stars, biased distribution in redshift due to magnitude limit selection, etc. The filament's positions and redshifts may also be affected by the noise of the survey, and by the finger-of-god effect of the galaxies in clusters (e.g., Jackson, 1972), producing shifts of the galaxies in redshift space. The strategy adopted in my study to estimate over-density profiles of galaxies on a catalogue of galaxies (WISExSCOS) independent of the one used to detect the filaments (SDSS) should reduce these biases, since the noises and the biases of the two galaxy catalogues are independent.

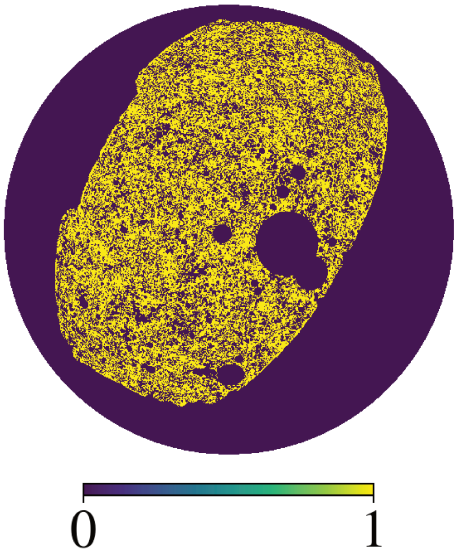

Figure 6.6: Orthographic projection of the union mask of the northern hemisphere used in this analysis.

6.4 Properties around cosmic filaments

6.4.1 Average properties

The stacked profile $< 1 + \delta_{\rm gal} >$ of the 5559 filaments selected from the DisPerSE catalogue are shown in Fig. 6.7. Blue represent the measured profile, and red the null-tests of 100 measurements on random positions and rotations. The binning of the measurement is between $r = 0.25$ and $r = 50$ Mpc. Note that due to the pixel resolution $\theta_{\rm pix}$, no measurement of the very central regions ($r < 0.25$ Mpc) is possible. Profiles presented in this study thus show the behaviour of the galaxies at large distances from the filament's spines. I have modelled the galaxy over-density profile by an exponential law (show in orange in Fig. 6.7), that takes the form:

$$\delta_{\rm gal}(r) = \delta_{\rm gal_0} \times e^{-\frac{r}{r_{\rm m}}} + c_{\rm gal}, \tag{6.3}$$

where $\delta_{\rm gal_0}$ is the mean projected over-density in the spines, $r_{\rm m}$ the typical radius, and $c_{\rm gal}$ an offset that can be due to biased measurements of the background because of the non Gaussianity of the galaxy density field.

The fit of the model to the data with a least square minimisation gives $\delta_{\rm gal_0} = 0.121 \pm 0.002$, $r_{\rm m} = 7.5 \pm 0.2$ Mpc, and a very small offset $c_{\rm gal} = 0.0093 \pm 0.0004$.

Although some studies have found that dark matter halo density profiles around filaments follow a power law with index between -2 and -3 beyond 3 Mpc (Colberg, Krughoff, and Connolly, 2005; Dolag et al., 2006; Aragón-Calvo, van de Weygaert, and Jones, 2010), density profiles of galaxies around filaments have not been studied much. Gheller and Vazza, 2019 have fitted a power law with index -1.84 to galaxy sized halo density profiles around filaments, but as shown in Cautun et al., 2014, filaments are very different in terms of width (with width from 2 Mpc to 10 Mpc for the largest ones), making the average density profiles of galaxies very dependent on the selection of filaments. In addition, the profiles presented here may also contain the correlations between the filaments, that is the probability to cross another filaments up to a certain radius, and the probability to cross the same filaments another time at another radius if the filament is curved. These effects would results in an enlargement of the profile. Hence, a comparison of the fitted typical radius $r_{\rm m} = 7.5 \pm 0.2$ Mpc with other studies is not relevant.

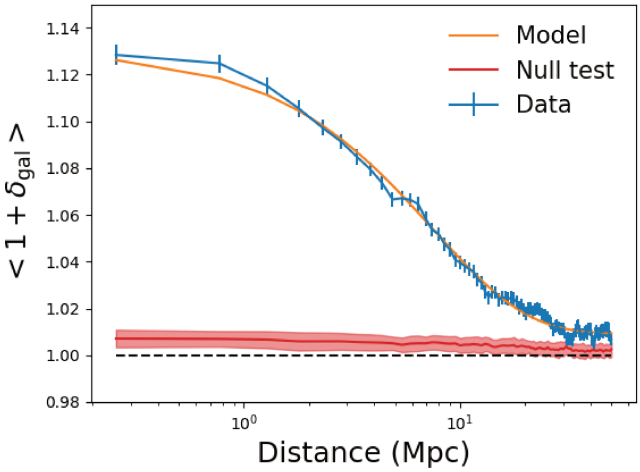

Figure 6.7: In blue, the stacked radial profile of over-density $< 1 + \delta_{\rm gal} >$ of the galaxies from WISExSCOS catalogue around the 5559 filaments. In orange, the best fitted exponential model presented in Eq. 6.3. In red, the 100 measurements on random positions from the null tests.

To study the effect of the global environment, I have averaged the galaxy over-density profiles in the three categories of filaments: the short, the regular, and the long filaments. I show in the left panel of Fig. 6.8 the resulting profiles. A small offset is seen for the regular filaments ($c_{\rm gal} = 0.013$) and for the short filaments ($c_{\rm gal} = 0.018$). However, all three average profiles share the same shape. This trend is confirmed by the computation of the residuals of the three profiles with the exponential model (Eq. 6.3), with the parameters from the average profile ($\delta_{\rm gal_0} = 0.121 \pm 0.002$ and $r_{\rm m} = 7.5 \pm 0.2$ Mpc). The residuals are shown in the right panel of Fig. 6.8. They do not show significant bias outside the error bars, suggesting that the galaxies around the small, regular, and long filaments, are sharing the same shapes. This result is supported by results obtained in numerical simulations (e.g., Colberg, Krughoff, and Connolly, 2005).

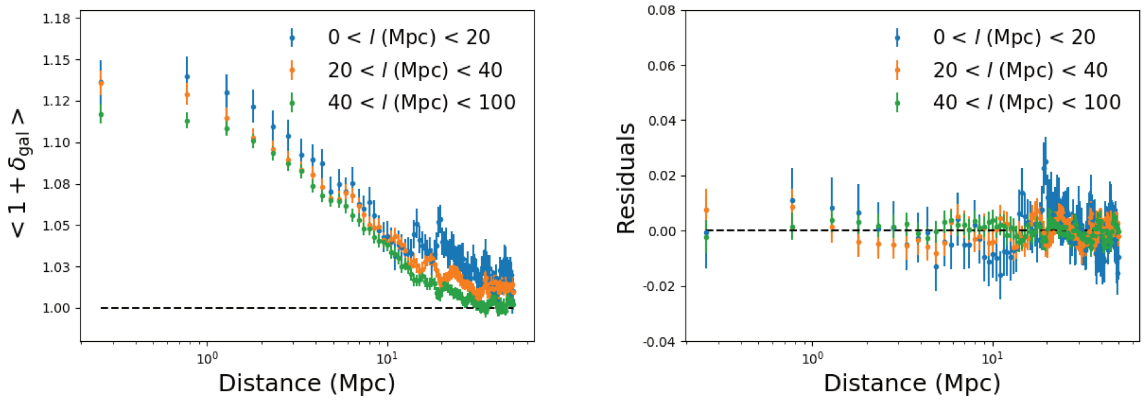

Figure 6.8: Left: radial profiles of galaxy over-densities $< 1 + \delta_{\rm gal} >$ around cosmic filaments, for short, regular, and long filaments. Right: residuals after substracting the exponential model fitted to the average over-density profile, shown in orange in Fig. 6.7, and the tiny offsets.

I have also computed the stacked profiles of excess of SFR and excess of M_*, defined as $< 1 + \delta_{SFR} >$ and $< 1 + \delta_{M_\star} >$ in the same way as for $< 1 + \delta_{gal} >$. I show in Fig. 6.9 the resulting stacked profiles for the short, the regular, and the long filaments. Gradients of SFR and $M_\star$ pointing towards the spines of the filaments are detected with a significance of 6.5 σ and 9.5 σ, respectively. Galaxies are significantly $\sim 0.1\%$ more massive, and $\sim 0.1\%$ less star-forming. This trend is expected since the excess of passive galaxies near the filaments spines was already shown in several studies (e.g., Malavasi et al., 2017; Laigle et al., 2018; Kraljic et al., 2018; Sarron et al., 2019). This implies a different ratio of passive over active galaxies with increasing distance to the filaments' spine, producing these gradients. The $M_\star$ gradient is 60% higher in the case of short filaments. This trend is also expected, as short filaments may represent bridges of matter, i.e., more mature structures lying in the densest environments of the Cosmic Web (e.g., Aragón-Calvo, van de Weygaert, and Jones, 2010) or event resulting from cluster interactions. These environments are already showing evidence of hot diffuse gas potentially inducing quenching of galaxies. For example in the bridge of matter between A399 and A401 (see Chap. 5), most of the galaxies between the two clusters were found to be passive. Therefore, it is hard to identify the origin of the observed gradients at this stage. They may reflect a change in the galaxy populations, or, a real stellar mass gradient of all types of galaxies around filaments.

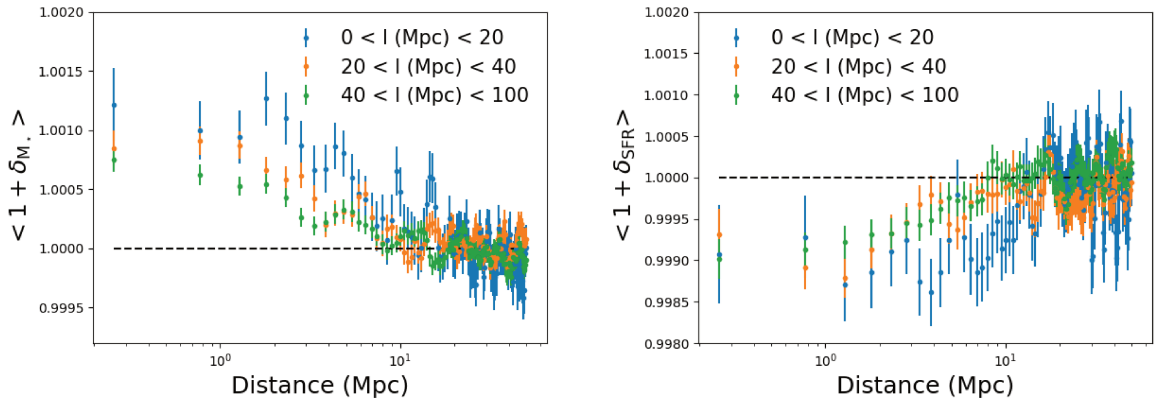

Figure 6.9: Radial stacked profiles of excess of M_* and SFR: $< 1 + \delta_{SFR} >$ and $< 1 + \delta_{M_\star} >$, for the short, the regular, and the long filaments.

6.4.2 Spliting over galaxy types

To explore further the impact of the environment on galaxy properties, I have computed the stacked radial profiles on galaxies, by splitting them in three populations, i.e., active, transitioning, and passive. I have used the 3D maps of galaxy densities constructed for the three types. The resulting stacked profiles $< 1 + \delta_{gal} >$ around the short, regular, and long filaments are shown in Fig. 6.10.

The profiles of galaxy over-density are decreasing with the galaxy types: the excess of passive galaxies around the spines of the filaments is higher than the excess of transitioning galaxies, and the excess of transitioning galaxies is higher than the active galaxies. These results show that whatever the length of the filaments, more passive galaxies are distributed towards the spines of the filament. This indicates a quenching process inside the cosmic filaments, as already shown in, e.g., Martínez, Muriel, and Coenda, 2016; Malavasi et al., 2017; Kraljic et al., 2018; Laigle et al., 2018; Sarron et al., 2019. Another interesting trend to notice is the excess of passive galaxies that is higher around the short filaments than around regular and long ones. This confirms the trend that short filaments may be bridges of matter. However, no difference is seen between the regular and long filaments. These results show that from a

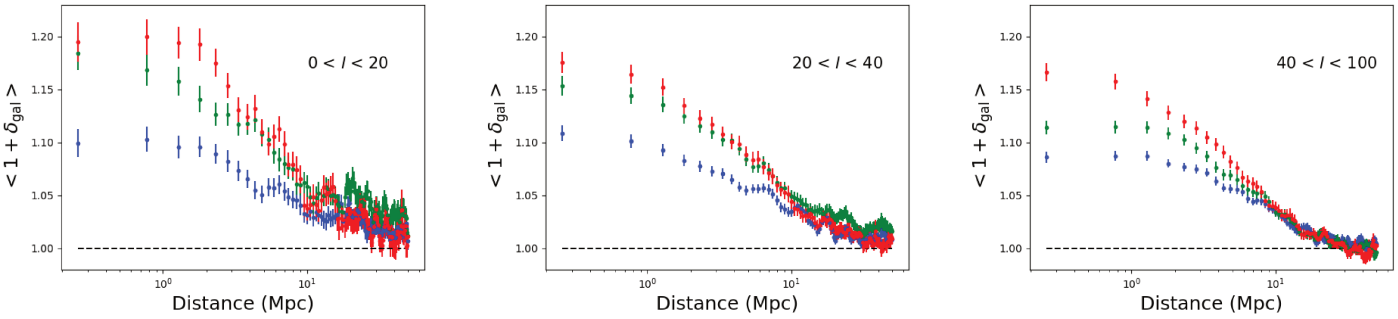

Figure 6.10: Stacked radial over-density profiles of active (in blue), transitioning (in green), and passive (in red) galaxies, as defined with the distance to the main sequence (detailed in Chap. 4.6.1), around the short, the regular, and the long cosmic filaments from left to right.

length of 20 Mpc, cosmic filaments share on average the same properties and the same galaxy population distributions whatever their length.

I have also computed the excess of $M_\star$ and SFR for the three galaxy types. The resulting stacked profiles are shown in Fig. 6.11. In the left panel, $M_\star$ gradient is seen for each types of galaxies: the active, the transitioning, and the passive ones, detected at 2.2 σ, 4.3 σ, and 5.3 σ, respectively. These results show that the galaxies are $\sim 0.05\%$ more massive in the spine of the filaments, whatever the types. An $M_\star$ gradient towards the spines of the filaments in active galaxies was also detected in Malavasi et al., 2017 and in Kraljic et al., 2018. This may be due to mergers of galaxies inside filaments (e.g., Malavasi et al., 2017). In the right panel of Fig. 6.11, a positive SFR gradient is detected at 6.3 σ, only for passive galaxies. This suggests that passive galaxies are more star-forming inside the filaments. This trend could be due to the rejuvenation of the star formation in passive galaxies, due to gas accreted by galaxy mergers.

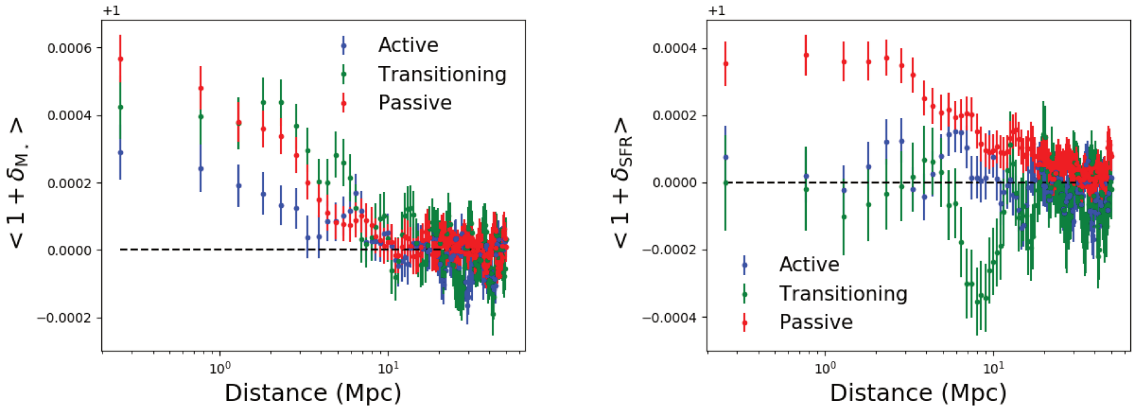

Figure 6.11: Left: $< 1 + \delta_{M_\star} >$ stacked radial profiles for each galaxy types: active galaxies in blue, transitioning galaxies in green, and passive galaxies in red. Right: $< 1 + \delta_{SFR} >$ stacked radial profiles for the same galaxy types.

6.5 A profile derived by the quenching

6.5.1 Quiescent fraction

Motivated by the results of Sect. 6.4.2, that indicate that galaxy populations are differently distributed towards the spines of the filaments, I have computed the fraction of quenched galaxies, which can be interpreted as the quiescent fraction f_Q (e.g., Fontana et al., 2009; Hahn et al., 2015; Martis et al., 2016; Lian et al., 2016). Here the quiescent fraction is defined as the ratio between the excess of passive galaxies $< 1 + \delta_{\mathrm{gal_P}} >$ and the sum of the excess of passive $< 1 + \delta_{\mathrm{gal_P}} >$ and active $< 1 + \delta_{\mathrm{gal_A}} >$ galaxies:

$$f_Q = \frac{< 1 + \delta_{\mathrm{gal_P}} >}{< 1 + \delta_{\mathrm{gal_A}} > + < 1 + \delta_{\mathrm{gal_P}} >}. \tag{6.4}$$

Since the transitioning galaxies are in the process of quenching and can be defined both as star forming, or quenched, I did not take them into account. The quiescent fraction f_Q around cosmic filaments is shown in Fig. 6.12.

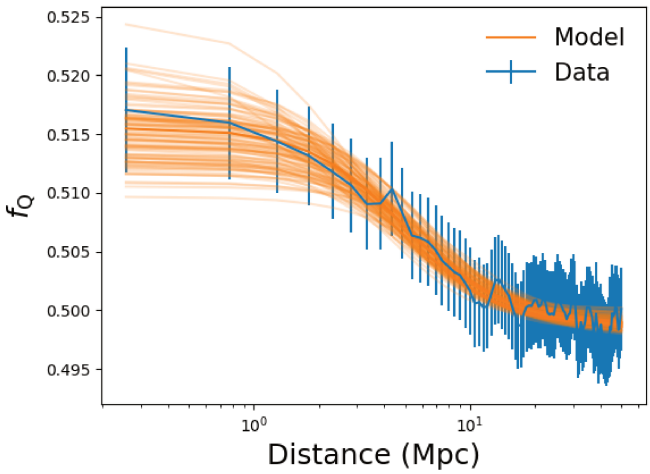

Figure 6.12: In blue, the quiescent fraction profile f_Q, averaged around the 5559 selected filaments. In orange, the β-model (Eq. 6.5) from different realisations of the MCMC.

6.5.2 A link to the gas content?

It is still not clear to date what are the dominant drivers of the quenching process in the galaxies. It has been already shown, and confirmed in the present study, that the Cosmic Web filaments have an impact on galaxy properties (e.g., Malavasi et al., 2017; Laigle et al., 2018; Kraljic et al., 2018; Sarron et al., 2019). Some studies have assumed the quenching being only due to the environment and to the Cosmic Web, via the hot gas in the halos ($T \geq 10^{5.4}$ K) preventing the galaxies from forming stars (Gabor and Davé, 2015) or by Cosmic Web Detachment (Aragon-Calvo, Neyrinck, and Silk, 2016). Peng, Maiolino, and Cochrane, 2015 and Trussler et al., 2018 have studied the effect of the gas of the Cosmic Web on the quenching of the galaxies, and have conclude that starvation and strangulation were the main processes of quenching. In each of these studies, the quenching is thus related to the temperature or the density of the gas, i.e., to the gas pressure around the Cosmic Web. Moreover, as discussed in Chap. 1.3.2.4, in

a study led by H. Tanimura, we have detected the SZ emission around cosmic filaments detected with DisPerSE on SDSS galaxies in the range $0.2 < z < 0.6$. We have estimated the temperature of the gas around the filaments to be $T \sim 10^6$ K, that is larger than the temperature of quenching of the halos found in (Gabor and Davé, 2015): $T \geq 10^{5.4}$ K.

The quiescent fraction profile measured in the previous section and shown in Fig. 6.12 should depend mostly on the processes of quenching, that drives the population changeover. Driven by the results from Tanimura et al., in prep. and by the previous studies discussed above, I have modelled the f_Q profile with a model of distribution of gas density around the cosmic filaments. Considering the results shown in the right panel of Fig. 1.16, where the gas has been successfully fitted in SZ with a β-model, historically used to model the gas density profiles in galaxy clusters (Cavaliere and Fusco-Femiano, 1978), I have used the same model for f_Q. It writes as:

$$f_Q(r) = \frac{f_{Q_0}}{\left(1 + \left(\frac{r}{r_s}\right)^2\right)^{\frac{3}{2}\beta}} + c, \tag{6.5}$$

where f_{Q_0} is the mean ratio of excess of passive galaxies over the sum of the excess of active and passive galaxies in the center of the filaments, r_s is the core radius, β the slope of the profile, and c is the background value.

I have performed an MCMC analysis (described in Chap. 3.2.1) and obtained the posterior distributions of the four parameters of Eq. 6.5. I show their distributions and their correlations in Fig. 6.13. I also display in orange in Fig. 6.12 1000 models randomly picked from the MCMC distribution. The median parameters are: $f_{Q_0} = 0.017 \pm 0.003$, $r_s = 4.4 \pm 1.7$ Mpc, $\beta = 0.54 \pm 0.18$, and the background value $c = 0.498 \pm 0.001$. Assuming the quiescent fraction of galaxies traces the pressure of the gas responsible of the quenching, the values of the slope β of the gas profile and of the quiescent fraction profile should be the same. In this study, the slope β is not well constrained as seen in the distribution in Fig. 6.13: it is fully degenerated with the parameter r_s. However, $\beta = 2/3$ which is the case for projected iso-thermal gas in hydrostatic equilibrium (scenario supported by numerical simulation, e.g., in Gheller and Vazza, 2019), is encompassed in the values allowed by the MCMC. Moreover, the values $1/3 < \beta < 2/3$ that best fit the gas profile in SZ shown in the right panel of Fig. 1.16 (where $\chi^2 = 3.1$, $\chi^2 = 2.7$, and $\chi^2 = 2.5$, for $\beta = 0$, $\beta = 1/3$, and $\beta = 2/3$, respectively) are also encompassed in the allowed values by the MCMC.

6.6 Summary

I have studied in detail the statistical properties of the galaxies from the WISExSCOS value-added catalogue around cosmic filaments detected with DisPerSE in the SDSS. I have measured with a high significance ($\gtrsim 5\sigma$) galaxy over-density radial profiles on the full sample of galaxies, and also on the three populations: active, transitioning, and passive galaxies.

Despite some biases on the measurement due to the methodology or to the data themselves, I have fitted an average profile of galaxy over-density around cosmic filaments with an exponential law. I have obtained a typical radius of $r_m = 7.4 \pm 0.1$ Mpc. I have also pointed out the evidence of a higher excess of passive galaxies than transitioning galaxies, and a higher excess of transitioning galaxies than active galaxies near the filament's spines. This excess of passive galaxies induces an SFR and an M_* gradient pointing towards the filament's spines, that I have also detected. This indicates that there are more passive galaxies near the filament's spines, in agreement with the previous studies (e.g., Malavasi et al., 2017; Kraljic et al., 2018; Laigle et al., 2018; Sarron et al., 2019).

I have also studied the excess of M_* and of SFR for the three galaxy populations, and have pointed out the evidence of a positive M_* gradient for the active, the transitioning, and the passive populations

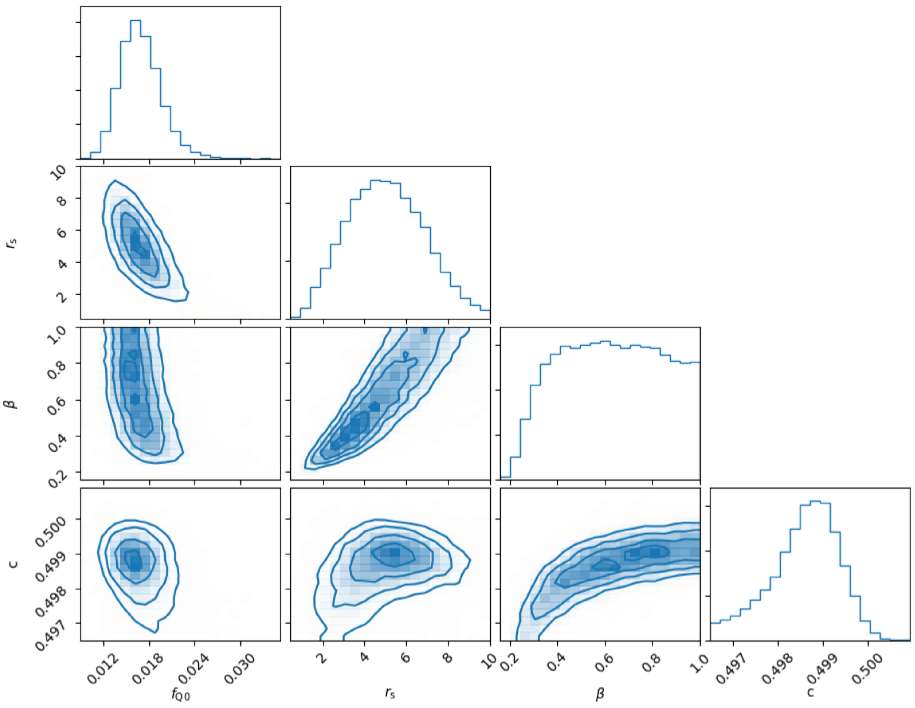

Figure 6.13: Posterior distributions of the four parameters of the β-model (Eq. 6.5), fitted to the quiescent fraction profile shown in Fig. 6.12.

of galaxies. This means that galaxies are more massive in the filaments, whatever their types, in agreement with Kraljic et al., 2018 and Malavasi et al., 2017 who observed a stellar mass gradient for star forming galaxies. I have also detected a positive SFR gradient for passive galaxies, showing that passive galaxies are more star-forming in the filaments. These gradients could be related to galaxy merges inside filaments, that increase the M_* of the galaxies and may rejuvenate the SFR in passive galaxies.

I have investigated how the quiescent fraction of galaxies, f_Q, behaves around cosmic filaments. Following previous studies showing the role of the Cosmic Web and of the gas in the quenching of star formation, and following the recent detection of the diffuse hot gas around cosmic filaments in SZ with T $\sim 10^6$ K, I have modelled the f_Q profile with a model of gas density distribution. I have measured parameters with an MCMC analysis. The obtained slope $\beta = 0.54 \pm 0.18$ is not well constrained. However, the MCMC distribution encompasses the value $\beta = 2/3$ that is the case for projected isothermal gas in hydrostatic equilibirum, that is suggested for example by Gheller and Vazza, 2019. The MCMC distribution of β also encompasses the values of β fitted on gas profile with SZ. The slope of the quiescent fraction profile is not inconsistent with the slope of a gas profile around cosmic filaments. Therefore, it is not to exclude that the gas around the Cosmic Web is the main responsible of quenching of star formation in galaxies.

To capture and understand the connections between the Cosmic Web environment and the galaxy properties, the profiles derived and shown in this study will be compared with numerical simulations in Illustris-TNG (Galárraga et al., in prep.) and with observations of the gas around cosmic filaments (with the SZ effect in *Planck*, or in the X-rays with the future e-Rosita[3] mission). In the context of future

[3] https://www.mpe.mpg.de/eROSITA

very large galaxy surveys like Euclid[4], LSST[5], or WFIRST[6], such studies on galaxy properties will be possible for a broader range of redshift, for a wider field of view, and thus with a higher significance.

[4] https://www.euclid-ec.org
[5] https://www.lsst.org
[6] https://www.nasa.gov/wfirst

Chapter 7

Detecting the SZ effect in *Planck* with deep learning

Contents

Abstract

The *Planck* collaboration have extensively used the six *Planck* HFI maps, each of them with a beam below 10 arcmin, to construct a full sky *y* map and to detect galaxy clusters with dedicated component separation methods. Although powerful, these methods still introduce some bias in the detection of the sources or in the reconstruction of the SZ signal, due to bias priors. For example the cluster detection algorithm use the GNFW profile model as proxy for the shape of the galaxy clusters. This profile is accurate on average, but not in individual clusters. In this chapter, I study the possibility of using deep learning algorithms to detect the SZ signal from the *Planck* HFI frequency maps, without resorting on any physical model. I use U-Net architecture network that shows very good performance, detecting all of the *Planck* clusters in a test area. I have detected more than 13,000 candidate SZ sources in the full sky, and have estimated their reliability by stacking different maps at their positions, i.e., CMB lensing, galaxy over-density, ROSAT. They indicate together that very low pressure SZ sources are detected in *Planck* thanks to the deep learning method. I also show that the diffuse emission is recovered around known large-scale structures such as Shapley, A399-A401, Coma, and Leo. These results are promising for the identification and the characterisation of the gas in the largest structures of the Universe. The material

presented in this chapter will be used for a future publication: *"Extracting the Sunyaev-Zel'dovich effect in Planck with deep learning"*, V. Bonjean, in prep.

Résumé

La collaboration *Planck* a utilisé les six fréquences de *Planck* HFI, chacune d'elles avec un faisceau inférieur à 10 arcmin, pour construire une carte du ciel complet du paramètre y, et pour détecter des amas de galaxies avec des méthodes de séparation de composantes dédiées. Bien que puissantes, ces méthodes introduisent des biais dans la détection des sources ou dans la reconstruction du signal SZ. Par exemple, l'algorithme de détection d'amas utilise le profil GNFW pour modéliser la forme des amas de galaxies. Ce profil représente les amas en moyenne, mais n'est pas optimal pour représenter les amas individuellement. Dans ce chapitre, j'étudie la possibilité d'utiliser des algorithmes de "deep learning" pour détecter le signal SZ dans les cartes de fréquences de *Planck* HFI, sans recourir à aucun modèle physique. J'utilise une architecture de réseau de type U-Net, montrant de très bonnes performances, par exemple en détectant tous les amas *Planck* dans une zone de test. J'ai détecté plus de 13 000 candidats SZ dans le ciel, et j'ai estimé leur fiabilité en utilisant le "stacking" dans différentes cartes à leurs positions, i.e., la carte de lentillage gravitationnel du CMB, des cartes de surdensité de galaxies, et la carte en émission X de ROSAT. Les résultats indiquent que des sources SZ individuelles à très basse pression sont détectées dans *Planck* avec cette nouvelle méthode. Je montre aussi que l'émission SZ diffuse est détectée autour de structures connues comme Shapley, A399-A401, Coma, et Leo. Ces résultats sont prometteurs pour l'identification et la caractérisation du gaz dans les plus grandes structures de l'Univers. Le matériel présenté dans ce chapitre sera utilisé dans le cadre d'une prochaine publication : *"Extracting the Sunyaev-Zel'dovich effect in Planck with deep learning"*, V. Bonjean, en préparation.

7.1 Motivations

The *Planck* satellite (presented in Chap. 2.3), thanks to its spectral coverage between 30 GHz and 857 GHz, have provided data ideal to capture the signature of the SZ effect. Based on two component separation techniques, i.e., NILC and MILCA (presented in Chap. 2.3.2), the *Planck* collaboration has constructed full-sky maps of the y SZ Compton parameter at a resolution of 10 arcmin, using the six frequencies of HFI (Planck Collaboration et al., 2016c).

The *Planck* collaboration has also applied Matched Multi-Filters on the six *Planck* HFI frequency maps (as detailed in Chap. 2.4.2), to detect hundreds of new galaxy clusters via the SZ effect, that were later confirmed in optical (e.g., Planck Collaboration et al., 2016e; Streblyanska et al., 2019). Since then, some studies have shown promising results by increasing the number of *Planck* SZ cluster sources to about 3500 by using new approaches, like artificial neural networks (Hurier, Aghanim, and Douspis, 2017), or by combining with other wavelengths, e.g., in X-ray with ROSAT (Tarrío, Melin, and Arnaud, 2019). Other studies have aimed at producing new higher resolution and lower noise SZ maps by combining *Planck* and ACT data (Aghanim et al., 2019). New detections of individual clusters or of stacked diffuse gas are still ongoing (as discussed in Chap. 1.3.2.4), showing that the full potential of the *Planck* data has not been completely exploited.

Studies deriving SZ catalogues or maps have combined the data with biased knowledge priors (e.g., assuming GNFW profile model), or by degrading the resolutions to the highest angular beam for homogeneity (Planck Collaboration et al., 2016c). Those biased knowledge priors may prevent us from using the full potential of the *Planck* data. In this chapter, I show the possible application of deep learning algorithms on the *Planck* frequency maps to detect the SZ effect. By training on high signal-to-noise SZ sources, i.e., galaxy clusters, deep learning algorithms are able to detect lower signal-to-noise ratio SZ components without depending on any explicit model. This study is promising for the detection and the

characterisation of individual structures with low pressure and diffuse gas in the *Planck* data. This is to date a proof-of-concept study, where the mass or the redshift distributions of the detected SZ candidate sources are still to quantify, and where the limits in signal-to-noise ratios of diffuse gas detections are still to estimate.

7.2 Learning procedure

In this study, I have trained a deep learning algorithm applied on the *Planck* HFI frequency maps to detect low signal-to-noise SZ sources via high signal-to-noise SZ emissions generated by the hot gas inside the galaxy clusters. To do so, I have chosen as inputs of the machine learning algorithm small patches of the *Planck* HFI frequency maps, and as outputs segmentation maps, showing the positions of the clusters in the patches. The trained model thus provides an SZ probability map, between 0 and 1, that can be compared with the known clusters in a test sample, or with the *Planck* MILCA SZ map.

7.2.1 Training catalogue

7.2.1.1 Catalogue of clusters

I have selected three catalogues of galaxy clusters to construct the segmentation maps that are used as output data for the training of the machine learning algorithm. First, the PSZ2 catalogue of clusters, to start with the very simplest case: learning *Planck* with *Planck*. To ensure the purity of the clusters, I have selected the 1,094 PSZ2 sources that are confirmed galaxy clusters, i.e., with measured redshifts. I note this cleaned catalogue the `Planck` z catalogue. I note the catalogue of remaining candidates the Planck no-z catalogue. Second, I have chosen the MCXC catalogue of X-ray clusters, as it contains lower mass clusters. I have selected MCXC clusters that are not included in the Planck catalogue, to merge galaxy clusters detected in different wavelengths. In the following, I note the 1,193 galaxy clusters of the MCXC catalogue that are not included in the PSZ2 catalogue, the `MCXCwP` catalogue. Finally, I have selected the RedMaPPer cluster catalogue from optical data to test the limits of the model and try to detect very low SZ signals, as it contains lower mass and higher redshift clusters. I have used RedMaPPer with different selections in richness λ (relative to the number of galaxies in the clusters), i.e., different selection in mass. I note RM_i the selection of the RedMaPPer clusters with the criterion: $\lambda > i$. In the following, I use the RM_{50} and the RM_{30} cluster catalogues. The three initial catalogues, i.e., PSZ2, MCXC, and RedMaPPer, are presented in detail in Chap. 2.4.

7.2.1.2 Training set and test set

I have used HEALPIX with $n_{\mathrm{side}} = 2$ to split the sky onto 48 tiles of equal-sized area of 860 square degrees each. One of them, the seventh one, is arbitrarily chosen to define a test area. This area is centred on the position ($l = 112.5°, b = 41.81°$). In this area, there are 40 clusters from the `Planck` z catalogue, 18 from the `Planck no-`z catalogue, and 50 from the `MCXCwP` catalogue. This area, and especially the clusters inside, are used as a test area and a test sample, and none of its 860 square degrees is seen by the model during the training. The training set is based on patches (projected as described in the next section), extracted in the remaining 47 tiles of the sky.

7.2.1.3 *Planck* patches and segmentations

I have extracted from the *Planck* HFI frequency maps $n = 100,000$ multi-channel patches, of 64×64 pixels with a resolution of θ_{pix}=1.7 arcmin (giving a field of view of $1.83° \times 1.83°$). These patches are chosen in the sky with random positions and random orientations, but each of the 100,000 patches

contains at least one galaxy cluster of the cluster catalogue chosen as output. I have then constructed their associated segmentation maps, by drawing circles of 5 arcmin diameter at the positions of the clusters. The pixels showing the positions of the clusters are set to 1, while the pixels in any other regions are set to 0. The diameter of the circles showing the positions of the galaxy clusters are set to 5 arcmin, i.e., the size of the smallest beam in the *Planck* HFI frequency maps (at 857 GHz). This fixed size acts like a filter and probably induces a bias in the reconstruction of the SZ sources in the SZ probability maps, preventing the computation of any reliable flux. However, the information up to a resolution of 5 arcmin can be learned by the network. The dimension of the input data is $100,000 \times 64 \times 64 \times 6$ pixels, and the dimension of the output data is $100,000 \times 64 \times 64 \times 1$ pixels. 10% of the 100,000 input patches, randomly chosen, are let selected the validation set, used to compute the performance of the machine learning algorithm during the training process and to stop the training in case of over-fitting. The training set is constructed with the remaining 90% patches.

7.2.2 Data pre-processing

To successfully apply the deep learning algorithm on *Planck* frequency HFI maps, a pre-processing of the data is needed. The mean input data and their standard deviations should be of the order of the unity, as machine learning algorithms perform better results for this range of values. However, in the *Planck* maps, a large variety of sources are detected, that are producing signals with very different spectral responses (e.g., radio sources bright in the low frequencies, dust sources bright in the high frequencies). The shapes of the pixel distributions of the *Planck* HFI frequency maps are thus very non Gaussians, preventing a simple normalisation of the maps to their means and their standard deviations. Here, I have chosen an approach to optimise the capture of the CMB spectrum deviation at the scale of the CMB fluctuation values (i.e., the secondary anisotropies, like the SZ effect).

The CMB fluctuations are themselves Gaussian distributions, but any kind of external sources (other than CMB anisotropies) add positive emissions in the *Planck* frequency maps. This produces an asymmetric Gaussian distribution, extended to the right part of the pixel distributions. I have thus fitted a Gaussian to the left part of the distributions, up to their statistical modes (values that appear most often), as shown in the left panel of Fig. 7.1. This part of the pixel distributions must contain the noise and the CMB fluctuations in each frequencies. I have then normalised each maps by the means and the standard deviations of the fitted Gaussians. This approach optimises the use of deep learning algorithms on the *Planck* frequency maps for the study of the CMB fluctuations at each frequencies (e.g., the SZ effect). I show the pixel distributions of each of the pre-processed HFI frequency maps in the right panel of Fig. 7.1.

7.2.3 U-Net architecture

Convolutional Neural Networks (CNN) are very efficient to encode informations on extended objects that can be invariant in translation or in rotation (detailed in Chap. 3.1.5), such as the case for SZ sources in the *Planck* maps. Moreover, the U-Net architecture, based on CNN, is one of the most efficient to reconstruct segmentation images (Ronneberger, Fischer, and Brox, 2015) (e.g., the U-Net architecture has won the ISBI cell tracking challenge 2015 applied on biomedical image segmentation). For these reasons, I have naturally chosen this network architecture to perform the training. A schema of the original architecture from Ronneberger, Fischer, and Brox, 2015 is shown in Fig. 7.2.

The architecture of the U-Net is symmetric and is composed of two blocks: a contracting part, or encoder (that encodes the spatial relevant informations), and an expansive part, or decoder (that decodes the learned informations to, e.g., reconstruct segmentation maps). This network has already been successfully applied in astrophysics to generate fast numerical simulations, to detect cosmic structures, or to estimate the flux of blended galaxies (e.g., He et al., 2018; Aragon-Calvo, 2019; Boucaud et al., 2019).

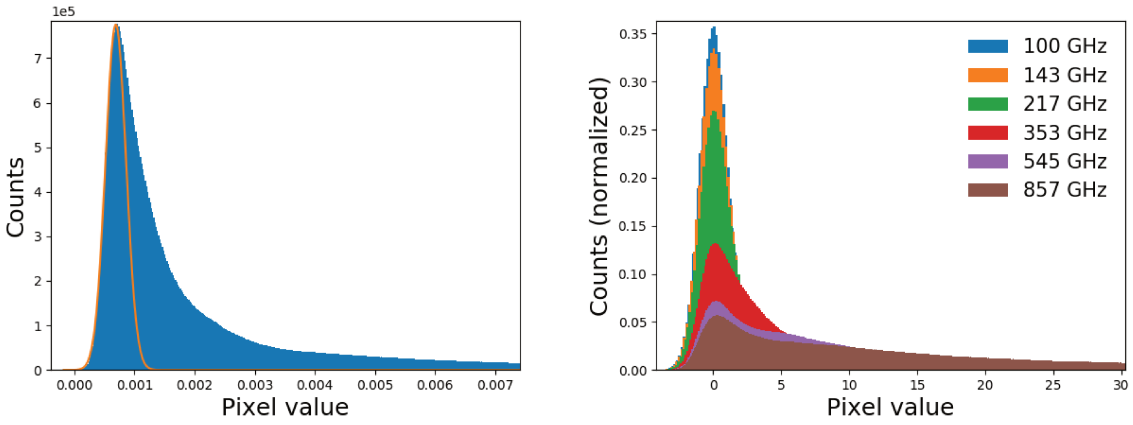

Figure 7.1: Illustration of the data pre-processing. Left: the pixel distribution of the map at 353 GHz. A Gaussian is fitted in orange up to the statistical mode of the distribution. The mean and standard deviation of the fitted Gaussian are used to normalise the data. Right: pixel distribution after normalisation of the six *Planck* HFI frequency maps.

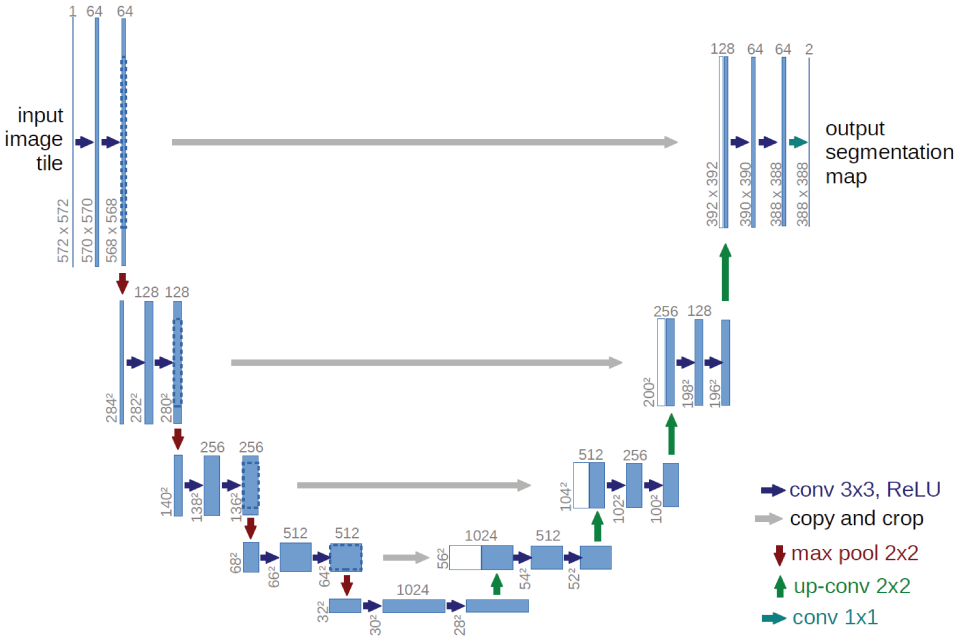

Figure 7.2: Schema of the original U-Net architecture from Ronneberger, Fischer, and Brox, 2015. The architecture consists of a contracting part and an expansive part.

The encoder is based on a typical CNN architecture (CNN architecture and elements are detailed in Chap. 3.1.5), that is a succession of blocks that contain two 3×3 convolutions, with a Rectified Linear Unit (ReLU) activation function, and a 2×2 MaxPooling layer, with the number of filters doubled at each down-sampling. The decoder is also composed of a succession of blocks: a 2×2 up-convolution, with half the number of filters, a concatenation with the corresponding cropped feature maps from the encoder (cropping), followed by two 3×3 convolutions, with a ReLU activation function. A last 1×1 convolution layer is added at the end of the network, for a total of 23 layers. As a last activation function, a sigmoid function is used, that is bounded between 0 and 1, in order to estimate a probability number. I

have also added dropout layers with a rate of 0.2 after each convolutional layers. The number of filters in the beginning of the network is related to the complexity of the recovered different features. As the maps of *Planck* used do not contain much complex spatial features, I have started the number of filters at 8, increasing to 128 in the last layer of the encoder. I have used the Adam optimizer with an initial learning rate set at 0.4, and the binary cross-entropy loss function. All the parameters on which the training depend (i.e., dropout rate, kernel size, number of filters, initial learning rate) were varied in different trainings of the U-Net, until finding the one actually learning the SZ features.

7.3 Cluster detection

I have investigated two cases. First, by training with the `Planck` z clusters, I have studied the possibility of detecting more clusters in the *Planck* maps with deep learning. I show the results hereafter. Second, I have trained three other models based on four different cluster catalogues in the output segmentation maps. I have added successively `MCXCwP`, RM_{50}, and RM_{30}. Those clusters, even though not included in the *Planck* catalogue, show statistical traces of SZ emission (as seen by the positive fluxes in Fig. 7.7). This was performed to study the possibility of detecting very diffuse SZ emissions, and reconstructing an SZ probability map. These results are shown in Chap. 7.4

7.3.1 Results on the test area

To train the algorithm to detect low signal-to-noise SZ sources in *Planck* HFI frequency maps, I have started with the simplest possible case: learn *Planck* with *Planck*. The segmentation maps used as outputs were generated based on the `Planck` z catalogue, to ensure an absolutely pure catalogue of galaxy clusters. The U-net has been successfully trained for about ∼1.5 h on a GPU NVidia Tesla K80, with a batch size of 20, and an early stopping of 12 epochs. Based on the trained model, I have constructed a full sky map of SZ probability p.

To estimate the performance of the model and ensure that the U-Net has learned to detect SZ sources, I have compared the SZ probability map in the test area and the test catalogues (i.e., clusters in the seventh HEALPIX pixel with $n_{\mathrm{side}} = 2$, described in Chap. 7.2.1.2). To detect galaxy clusters, I have simply defined the clusters as areas of probability p greater than a threshold p_{max}. For each area recovered above p_{max}, I have computed the position as the barycentre of the pixels. This detection method is very simplistic and not optimal but it is yet efficient enough to roughly check overall consistency. I have cross-matched the sources detected with this method with the three catalogues: `Planck` z, `Planck no-`z, and `MCXCwP`, and have studied the recovered clusters as a function of the detection threshold p_{max}. I show in Fig. 7.3 the recovered clusters on the three catalogues (left panel) together with the number of new detected sources (right panel). For a threshold $p_{\mathrm{max}} = 0.1$, all the `Planck` z clusters are recovered, together with 89% of `Planck no-`z clusters, and 12% of `MCXCwP` clusters. I have also detected 187 new sources.

To ensure that the newly detected sources do not correspond statistically to point sources that might contaminate the model (e.g., infra-red point sources), I have stacked the 187 new detected sources detected in the test area with the threshold $p_{\mathrm{max}} = 0.1$ in the *Planck* MILCA SZ map and in the *Planck* HFI frequency maps. I show the results of the stack in Fig. 7.4. Presence of SZ sources is suggested by a significant y emission in the *Planck* MILCA SZ map and by a signature of the SZ effect in the centre of the HFI frequency maps are seen (i.e., a negative emission in the 100 and 143 GHz maps, and a positive emission for the frequencies above 217 GHz). SZ sources may be populated by dust, suggested by the excess of signal also seen in the centre of the map at 217 GHz. A bright infra-red source is also seen in the HFI frequency stacked maps (under the centres), with an intensity increasing with the frequencies, together with a complex background in the 100, 143, and 217 GHz stacked maps coming from the CMB.

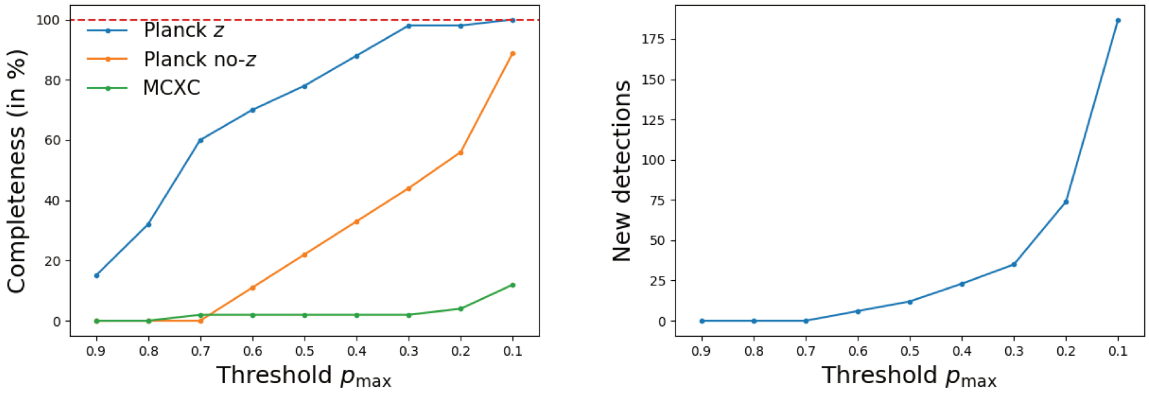

Figure 7.3: Results on the test area (seventh HEALPIX pixel with $n_{side} = 2$), containing 40 `Planck` z, 18 `Planck no-`z, and 50 MCXC galaxy clusters. Left: galaxy clusters recovered with different threshold of detection p_{max}. Right: the number of new detected sources as a function of the threshold p_{max}.

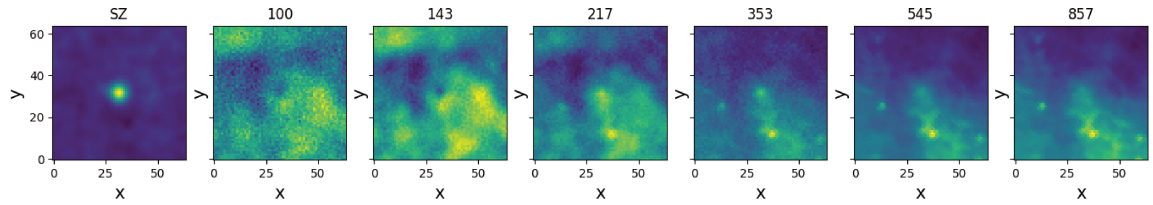

Figure 7.4: Stack of the SZ map and of the six HFI frequency maps at the positions of the 187 new detected sources in the test area. The title in front of the patches indicate the map (the numbers indicate the frequencies of *Planck* HFI maps in GHz). The signature of the SZ effect is seen in the centres of the HFI frequency maps, together with a complex combination of infra-red sources and CMB structures.

7.3.2 Results on the full sky

After checking the results of the U-Net in the test area, I have applied the same detection method to the full-sky SZ probability map, with a detection threshold of $p_{max} = 0.1$ in order to recover the maximum number of `Planck` z clusters. I have detected 14,976 sources in the full catalogue. I have compared the detections with the three catalogues of known galaxy clusters: `Planck` z, `Planck no-`z, and `MCXCwP`. Among the 14,976 detected sources,98.6% of the `Planck` z clusters are recovered, together with 74.1% of `Planck no-`z clusters, and 19.0% of `MCXCwP` clusters. There are also 1 cluster identified by the Atacama Cosmology Telescope (ACT) (Hasselfield et al., 2013) and 85 clusters identified by the South Pole Telescope (SPT) (Bleem et al., 2015) but not included in the *Planck* PSZ2 catalogue. 13,220 sources are thus new detections. I have investigated the purity of the sample of new detected sources, by cross-matching with *Planck* point sources. On the 13,220 newly detected sources, only 5.4% are matched within a cross-match radius of 5 arcmin with the positions of the *Planck* catalogue of galactic cold cores, and only 0.2% are matched with the positions of the *Planck* sources identified at 353 GHz. This shows that the sample of new detected sources is very slightly contaminated. In order to investigate the reliability of the 13,220 new detected sources, I have stacked at their positions 16 maps in different wavelengths, each of them probing different galaxy cluster counterparts. Some of the 16 maps are also based on *Planck* data, and thus are not independents, but some of the maps are independents and show

galaxy cluster counterparts in other wavelengths, i.e., in near infra-red (where galaxies emit) and in X-rays (where the same gas as the one seen in the SZ emits). The 16 maps are the *Planck* SZ MILCA map, the six *Planck* HFI frequency maps, the IRIS map at 100 μm (Miville-Deschênes and Lagache, 2005), the CMB lensing map (based on *Planck*) (Planck Collaboration et al., 2018a), four galaxy over-density maps of all galaxies, passive galaxies, transitioning galaxies, and active galaxies presented in Chap. 4.6, a star formation rate density maps and a stellar mass density maps also presented in Chap. 4.6, and finally, the ROSAT X-ray map (ByoPiC product). The *Planck* maps are masked from the *Planck* Catalogue of Compact Sources (detailed in Chap. 2.3.3, and the ROSAT map is masked from the point sources detected in ROSAT, Chandra, and XMM-Newton (Boller et al., 2016; Evans et al., 2010; Rosen et al., 2016, respectively). I show in Fig. 7.5 the result of the 16 stacks.

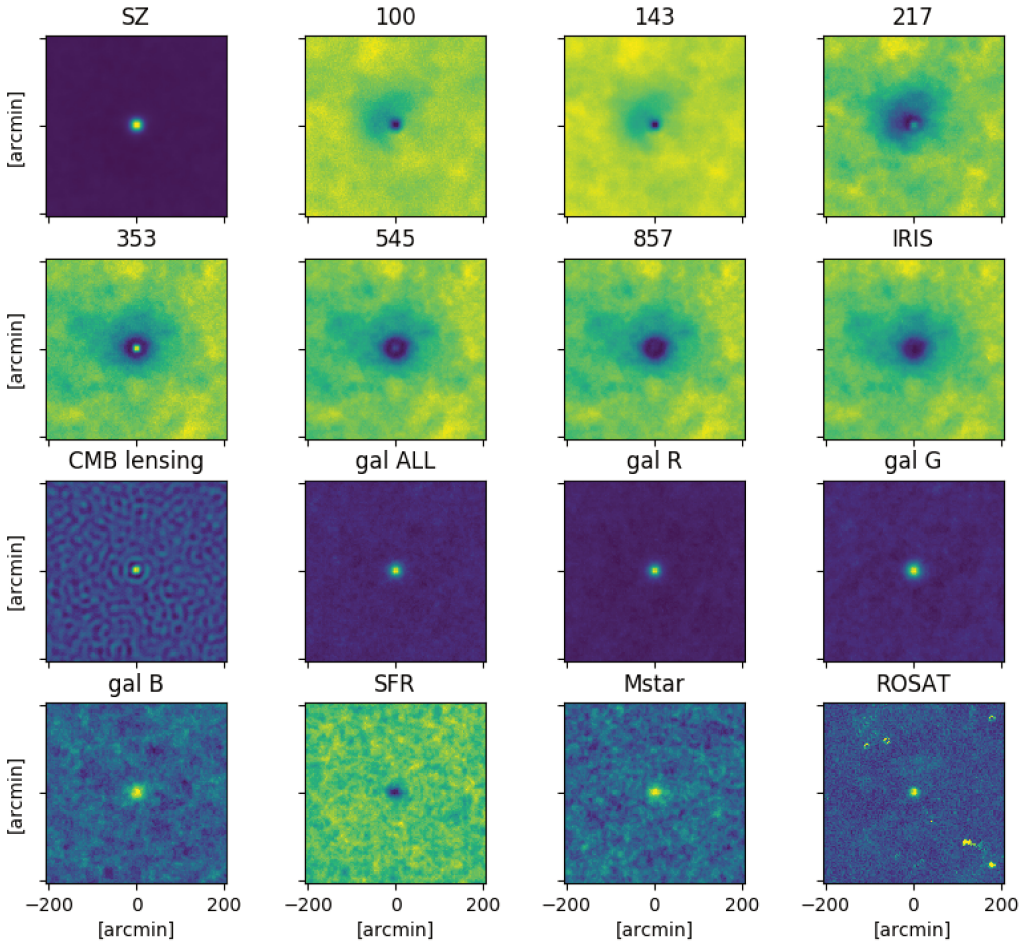

Figure 7.5: Stack of the 13,220 new detected sources in 16 different maps probing counterparts of galaxy cluster components in different wavelengths: the SZ MILCA y map, the six *Planck* HFI frequency maps, the IRIS map at 100 μm, the CMB lensing map, the four galaxy over-density maps based on all, and on red, green, and blue populations of galaxies from the WISExSCOS photometric redshift catalogue (in $0.1 < z < 0.3$), and the ROSAT X-ray map. Each of them show counterparts of galaxy cluster components.

An evidence of dark matter potential is seen in the CMB lensing map. In the over-density maps and galaxy properties density maps, a presence of red galaxy over-density is seen in each of the maps,

suggesting that some of the detected SZ sources are galaxy clusters in the redshift range $0.1 < z < 0.3$ (that is the redshift range of the catalogue of galaxies used to construct the galaxy density maps). In the six *Planck* HFI frequency maps, a complex background is seen, that correlates with the IRIS map at 100μm showing the dust. This result suggests that the background in the *Planck* frequency maps is due to the emission of the dust in our galaxy, and thus that the new detected sources are preferentially in areas without this dust contamination in foreground. In the very centres of the HFI maps, inside the hole produced by the background, a very nice SZ signature is seen, with a decrement at 100 and 143 GHZ, and an increment in the 353 GHZ map, decreasing up to the 857 GHz map. An evidence of Bremsstrahlung in the ROSAT stacked map is also seen. Each of these stacks show that there are evidences of cluster counterparts in different wavelengths, and thus that a large number of the new detected sources are actual clusters. To ensure that the trend seen in the stacks are not due to few outliers that dominate the distributions, and to ensure that setting the threshold $p_{\rm max}$ to 0.1 does not give too many false detections, I have also stacked the new detected sources in bins of their maximum associated probability p, in the X-ray ROSAT map. The bins are chosen so that they contain the same number of sources. I show the radial profiles in the ROSAT map in Fig. 7.6, with error bars computed with the bootstrap method (presented in Chap. 3.2.3) that re-samples the profiles in each bin.

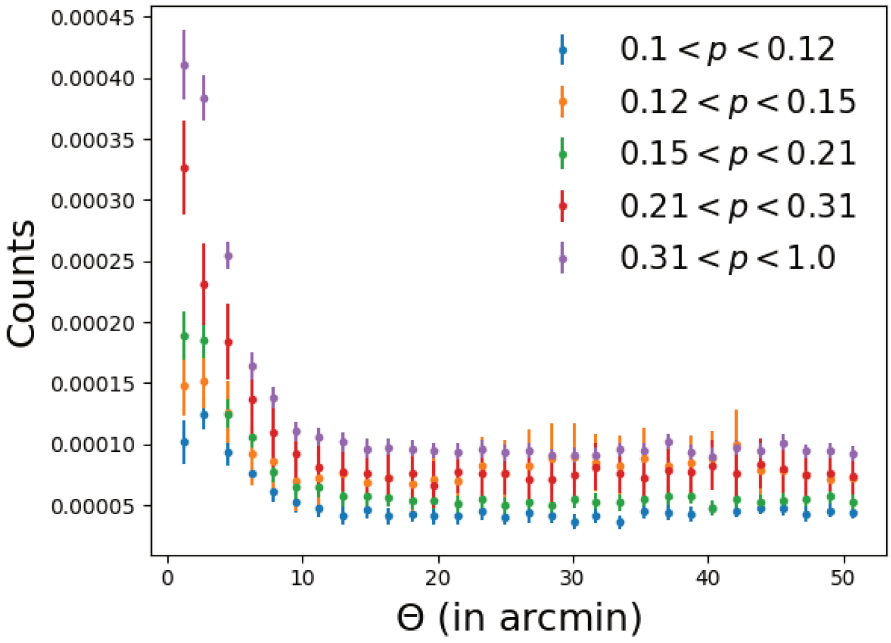

Figure 7.6: Stacked radial profiles of the 13,220 new detected sources in the X-ray ROSAT map as a function of their associated probabilities p. This show evidence of diffuse gas emission in X-ray for the 13,220 detections, that is decreasing as a function of their probability output by the U-Net. Error bars computed with the bootstrap method ensures the significance by re-sampling the profiles.

This result shows evidence of X-ray emission, decreasing with the SZ sources probabilities p output by the U-Net. The bootstrap method ensures the significance of the profiles by re-sampling the profiles in each bin. Even in the lowest probability bin, i.e., $0.1 < p < 0.12$, a significant X-ray emission is observed. This suggests that the low probabilities output by the U-Net indicate low signal-to-noise ratio SZ emissions, e.g., low pressure, low mass, or high z galaxy clusters.

7.3.3 Scaling relations

These results show that the U-Net trained on high signal-to-noise ratio SZ sources (i.e., *Planck* clusters) has learned the frequency-dependency and the spatial features of the SZ effect in the six *Planck* HFI frequency maps, and that a large number of the new detected sources with low associated probability are actual galaxy clusters. If so, the maximum value of probabilities p associated with the detected sources in the SZ probability maps should be proportional to the SZ flux in the MILCA y map. I have analysed the associated maximum probabilities of the 14,976 detected sources in the full-sky, by stacking at their positions the *Planck* SZ MILCA map for different bins of probabilities p. I show in Fig. 7.7 the results, where the blue points are the fluxes in the stacked *Planck* SZ MILCA maps as a function of the mean maximum probability p in the bins. I have added for comparison in orange the stacked flux computed by aperture photometry in the *Planck* SZ MILCA map of the *Planck* PSZ2 clusters, in green the stacked flux of the MCXC clusters, and in red and purple the stacked fluxes of the RedMaPPer clusters selected in richness. All the errors are computed with the bootstrap method. This result suggests that the maximum value of probability p output by the U-Net is related to the integrated flux in the *Planck* SZ MILCA map. Therefore, a scaling law may be estimated to translate the maps from p units to y units.

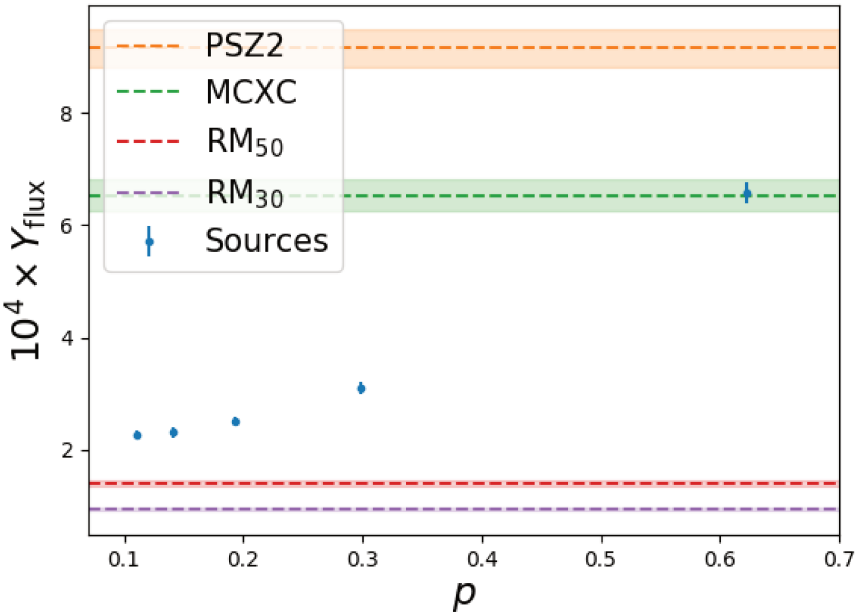

Figure 7.7: Scaling relation between the probability output by the U-Net p and the flux in the MILCA y map. In blue, the fluxes of the 14,976 sources detected on the full sky with the U-Net, stacked on the SZ MILCA y map as a function of their associated probability p. For comparison, the average flux in the MILCA y map of the PSZ2 clusters is shown in orange, the average flux of the MCXC clusters is shown in green, and the average fluxes of the RM_{50} and the RM_{30} clusters are shown in red and purple, respectively. Error-bars are computed with bootstrap.

In Fig. 7.7, I also show that the SZ sources detected with the U-Net have average fluxes between the ones of the MCXC and the RedMaPPer clusters. This suggest that the sources detected with the U-Net contain less massive SZ sources at low redshift, or higher redshift SZ sources, so that the stellar mass and redshift distributions of the SZ detected sources with the U-Net are between the distributions of the

MCXC and the RedMaPPer clusters that are shown in Chap. 2.4. However, as the SZ effect is a projected effect, there is a degeneracy between the redshift and the mass of the individual clusters. A detailed investigation of the range of redshifts and masses of the new detected sources is not performed in the present study, that only shows the potential of the application of deep learning algorithms on *Planck* data. This detailed analysis will be subject of a future study.

7.4 Diffuse SZ emission

Motivated by the construction of a new SZ map more sensitive to lower signal-to-noise SZ emission, I have trained three other U-nets, by choosing lower SZ signal-to-noise ratio galaxy cluster catalogues of reference to construct the segmentation maps of the training catalogue. In practice, I have added the `MCXCwP` clusters, the RM_{50}, and finally the RM_{30} clusters, in addition to the `Planck` z clusters. There are in total four U-Net models, the first one being the one presented in the previous section.

I have generated four full-sky maps of SZ probability based these four U-Net trained models. To illustrate the detection of diffuse gas, I have arbitrarily focused on four regions around large-scale structures already identified as containing diffuse SZ signal: the Shapley super-cluster (Aghanim et al., in prep.), the galaxy cluster pair A399-A401 that is fully described in Chap. 5, the Coma super-cluster (Erler et al., 2015, Malavasi et al. in prep.), and the Leo super-cluster. I show these structures in patches extracted from the SZ probability maps derived from the four models, together with the *Planck* SZ MILCA map for a visual comparison in Fig. 7.8.

When adding the `MCXCwP` clusters, the diffuse gas around the super-clusters and the bridge of matter between A399 and A401 are recovered. For the models with the RedMaPPer clusters, potential indications of large scale structures connecting the structures are seen. The SZ probability maps obtained with the U-Net are visually very close to the MILCA SZ map, but they seem less noisy and better resolved. These preliminary results are promising for the detection and the characterisation of the diffuse gas in the large-scale structures.

7.5 Summary

By training deep learning algorithms such as the U-Net on the *Planck* HFI frequency maps to recognise the spatial and spectral features of high signal-to-noise SZ signatures produced by the hot gas in known galaxy clusters, lower signal-to-noise SZ sources are recovered. In the most conservative case when training with the `Planck` z catalogue of 1,094 sources, I detect about 200 clusters more that are known MCXC clusters, and more than 13,000 new sources above a detection threshold $p_{\rm max} = 0.1$. This is about a factor three more than the numbers of sources found by Hurier, Aghanim, and Douspis, 2017 or Tarrío, Melin, and Arnaud, 2019. Although the detection threshold has been set to a low value, the sample of detected sources does not seem to suffer much from contamination, as shown in the stack of the 13,220 new detected sources in maps probing galaxy cluster counterparts in different wavelengths. The presence of dark matter halos is highlighted by the stacked CMB lensing map, the presence of hot gas by the stacked *Planck* SZ MILCA map and the X-ray ROSAT map, the presence of red galaxy over-density with the stacked galaxy maps constructed with the value-added WISExSCOS photometric redshift catalogue.

By focusing on areas around multiple systems, I have shown that deep learning models can be used to reconstruct an SZ map, more sensitive to lower signal-to-noise ratio SZ emissions. Although a qualitative study is still to perform in a future analysis, this proof-of-concept study shows the potential of applying deep learning algorithms on *Planck* data.

The U-Net and the method presented in this study can in principle also be applied to any components separation in the *Planck* data, e.g., radio emission, dust emission, CO emission, CMB, and so on. I am

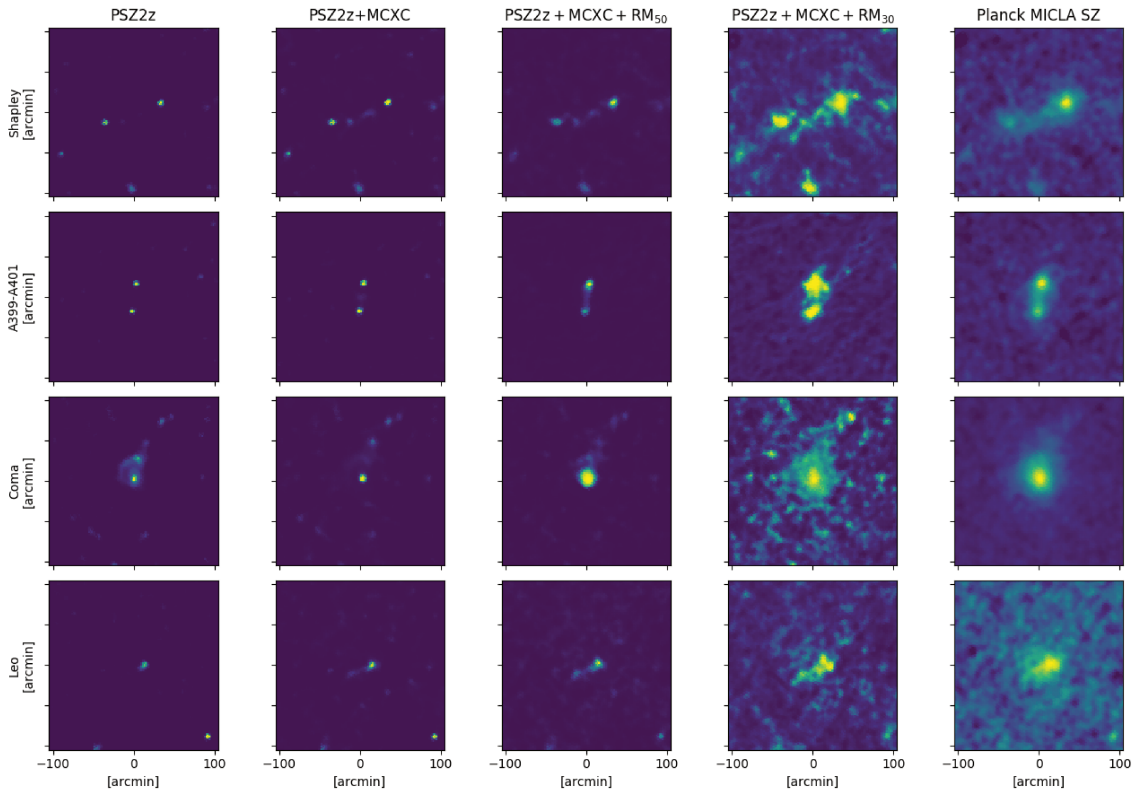

Figure 7.8: Patches of 120 × 120 pixels of size $\theta_{\rm pix}$=1.7 arcmin in four models of learning and in the SZ MILCA y map, for four emblematic large scale structures. From top to bottom: the Shapley super-cluster, the galaxy cluster pair A399-A401, the Coma super-cluster, and the Leo cluster. From left to right: the model PSZ2z, PSZ2Z+MCXC, PSZ2z+MCXC+RM50, PSZ2Z+MCXC+RM30, and the SZ MILCA y map for comparison.

presently investigating the training of more complex deep learning architectures multi-classes classification in collaboration with A. Beelen, A. Decelle, and N. Aghanim, in order to perform multi-component separation.

To improve even more the results, ACT and SPT maps can be included in the training process: results could then be compared to other studies that combined these two missions with the *Planck* data (e.g., Chown et al., 2018; Aghanim et al., 2019).

Part III

Analysis of the ESO follow-up of *Planck* clusters

Chapter 8

Analysis of the follow-up of *Planck* clusters: a *Planck*/ESO legacy sample of the most massive clusters

Contents

Résumé

Dans ce chapitre, je décris l'analyse de données du large programme ESO (PI: N. Aghanim) obtenu pour confirmer par la spectroscopie des galaxies en optique (avec VLT/FORS2) la présence d'amas de galaxies détectés par leur contenu en gaz avec le satellite *Planck*. Due à des problèmes lors de la sélection des sources à observer en spectroscopie, cette analyse requiert finalement une analyse des images en photométrie (NTT/EFOSC2), qui ont du être re-traitées. L'analyse est toujours en cours à ce jour.

8.1 Introduction

Surveying the whole sky with *Planck* has permitted the detection of the most massive clusters and the rarest ones in the exponential tail of the mass function (Planck Collaboration et al., 2016e). Several follow-ups of *Planck* clusters were realised with the European Northern Observatory (ENO) at Las Canarias (Barrena et al., 2018; Streblyanska et al., 2019), the Russian Turkish Telescope (RTT150) (Planck Collaboration et al., 2015b), in the SDSS survey (Streblyanska et al., 2018), and with the Canada-France-Hawaii Telescope (CFHT) (van der Burg et al., 2016), to ensure confirmation and redshift measurement

of Northern *Planck* sources. For Southern sources, a follow-up (P.I. N. Aghanim) that I present in this chapter was realised with ESO facilities. The confirmation of these high redshift and massive *Planck* clusters should provide a complete census of the highest mass and redshift bins, and thus complement nicely the $M - z$ domain occupied by ACT and SPT, lower-masses and higher-redshifts (shown in Fig. 8.1).

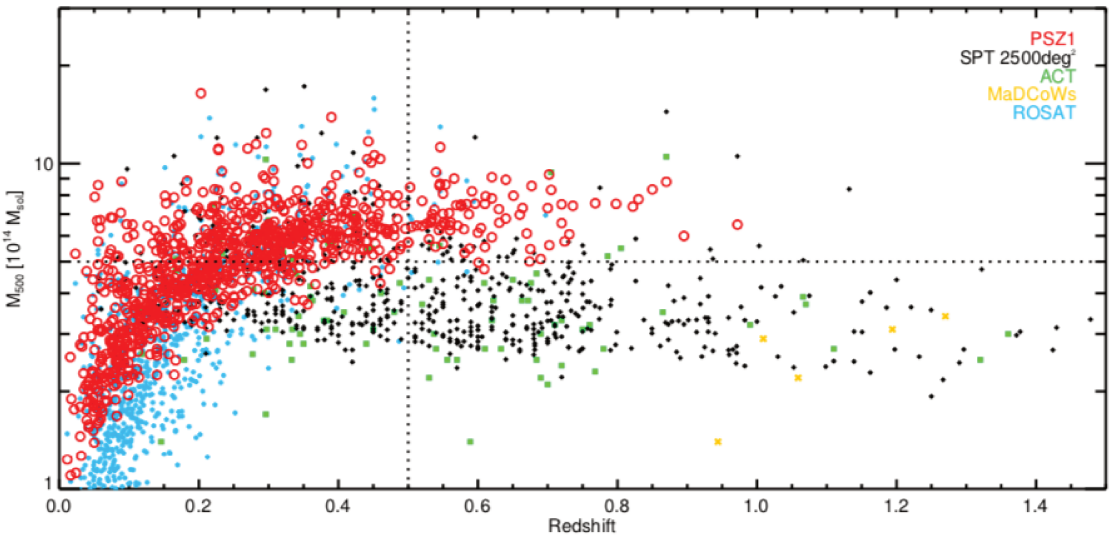

Figure 8.1: Distribution in the M–z plane of SZ clusters from different catalogues from Planck Collaboration et al., 2015a. In red, clusters from the PSZ1 catalogue (Planck Collaboration et al., 2014), compared with those from SPT in black (Bleem et al., 2015), ACT in green (Marriage et al., 2011; Hasselfield et al., 2013)), MaDCoWS (WISE clusters) in yellow (Brodwin et al., 2015), and MCXC in blue (Piffaretti et al., 2011)). As indicated in Planck Collaboration et al., 2015a, the black dotted lines show the *Planck* mass limit for the medium-deep survey zone at 20% completeness (as defined in Planck Collaboration et al., 2014) for a redshift limit of z = 0.5

The abundance of massive clusters is most sensitive to cosmology and their high masses make them a priori easier to model, being less affected by non-gravitational processes. Samples of massive clusters are thus the best cosmological tools already at low/moderate redshifts, $z > 0.3$ (e.g., Allen et al., 2008; Mantz et al., 2010); much more than low-mass clusters which provide comparable results only at much higher redshift. Hence the construction of well selected samples of massive clusters, in particular at intermediate to high redshifts, is crucial both for astrophysical and cosmological studies. For example, neutrino masses can be constrained from the comparison of cluster counts and normalisation derived from CMB measurements (e.g., Salvati, Douspis, and Aghanim, 2018); whereas clusters at high redshifts are needed in order to measure the evolution of the cluster mass function that can be used, e.g., to constrain the dark energy (e.g., Sartoris et al., 2016).

In this context, N. Aghanim and collaborators have set up an ESO large programme, *"a Planck/ESO legacy sample of the most massive clusters"*, that aimed at constructing the largest complete and well understood catalogue of the most massive clusters from the first PSZ1 catalogue of *Planck* (Planck Collaboration et al., 2014), by identifying tens of galaxy cluster members for each cluster. The initial sample of sources selected for observation with the ESO facilities was constructed in 2013-2014 from the PSZ1 catalogue of SZ sources detected down to S/N=4.5, excluding $\sim$ 35% of the sky, dominated by Galactic foregrounds. Within these 65% clean sky, the estimated purity of the sample is, from numerical simulations, 88% at S/N>4.5 (see Planck Collaboration et al., 2014 for details on the construction, statistical

characterization and validation of the PSZ1 catalogue). The sample of selected sources consisted of 142 cluster candidates. These sources comprised 21 high-reliability SZ sources (referred to as class 1 in Planck Collaboration et al., 2014), and 121 good reliability sources (class 2 and 3). In this chapter, I present the ongoing data reduction of both spectroscopic and photometric measurements of the ESO large programme.

8.2 Observations

The observations took place before the beginning of my PhD thesis, between 2013 and 2016, with a total of 233 hours 25 minutes and 33 seconds of observations in photometry and in spectroscopy. In total, 119 cluster candidates were observed with the NTT/EFOSC2 telescope in the R and the z bands (with the filters R#642 and z#623 respectively). These observations allowed the estimation of redshifts, $z_{\rm est}$, based on the red sequence (Gladders and Yee, 2000). The galaxy members of 49 cluster candidates with $z_{\rm est} > 0.5$ were observed in spectroscopy with VLT/FORS2, and some of the remaining candidates with $z_{\rm est} < 0.5$ were additionally observed with a shallow V band (with the V#641 filter in the NTT/EFOSC2).

Figure 8.2: Image of the NTT at La Silla, in the southern part of the Atacama desert of Chile. Image from https://www.eso.org/public/images/ib-la-silla15/. Credits: Iztok Boncina/ESO.

8.2.1 NTT/EFOSC2 imager

The photometric observations were made with the NTT/EFOSC2 in imaging mode, mounted at the Nasmyth B focus of the NTT (3.58m Richey-Chretien telescope, shown in Fig. 8.2). The camera contains 2048×2048 pixels, with sizes of $15 \times 15\mu$m, or $0.12" \times 0.12"$ on the sky, with a reading mode of 2×2 binned pixels corresponding to a field of view of $4.1' \times 4.1'$. For all clusters, three observing blocks were performed, together with corresponding biases and flat fields. A positional offset was applied to the three observing blocks to remove the bad pixels, and to fill the inter-pixel regions.

8.2.2 VLT/FORS2 spectrograph

The spectroscopic observations were made with the VLT/FORS2 instrument in MXU mode, installed on the UT1 telescope (Antu) of the European Southern Observatory (ESO) in Paranal (shown in Fig. 8.3). The camera is composed of a mosaic of two 2k×4k MIT CCDs with 15μm pixels, with a pixel scale of 0.25"/pixel using the Standard Resolution collimator, corresponding to a field size of 6.8'×6.8'. The GRIS300I grism together with the OG590 filter were chosen in order to span a wavelength range of $6000 - 11000$Å, and thus to properly detect the 4000Å break of passive galaxies in the typical range of

Figure 8.3: Image of the VLT in Paranal in Chile. Image from https://www.eso.org/public/images/eso0137k/. Credits: ESO.

redshift: $0.5 \lesssim z \lesssim 1$. For all clusters, two to six observing blocks were taken, together with ten bias, seven flat fields and one arc lamp calibration frames in the end of each observing nights. Standard star frames were observed in MOS mode within one slit of 5", usually one to three nights before or after the MXU observations.

8.3 Data reduction

During my PhD thesis, I have conducted the data reduction of the spectroscopic data of the galaxy members of the 49 clusters observed with VLT/FORS2. During the preparation of the spectroscopic observations in 2013-2015, there was an issue with the selection of the targets. The targets were originally supposed to be selected based on the red sequence sources observed with the NTT/EFOSC2, but the selection was instead done on the pre-images with VLT/FORS2 without information on the colour of galaxies. As as result, the purity of the selected members was very low and only a handful galaxy members were available for each cluster, compared to the 50 to 80 originally expected. Spectroscopic data alone can thus not provide enough information to confirm the redshifts of the clusters: they have to be combined with the information about the red sequence of galaxies, obtained with the photometry with NTT/EFOSC2. Therefore, during my PhD, I have supervised an intern-ship, Raphaël Wicker, to work on the reduction of the photometric NTT/EFOSC2 observations.

8.3.1 Spectroscopy

Following the strategy proposed with the ESO large programme collaboration, I have originally started to reduce the data with the software VIMOS Interactive Pipeline and Graphical Interface (VIPGI) adapted to handle FORS2 spectroscopic data: F-VIPGI (Nastasi et al., 2013). This software was performing the standard data reduction steps: bias substraction, correction from flat-field, spectral calibration, spectro-photometric calibration. The F-VIPGI software, although powerful and interactive, was degrading the spectral resolutions of the VLT/FORS2, and was not providing satisfying results on the spectro-photometric calibration (some sources were quite noisy or contaminated by atmosphere emission lines). I have therefore changed software and used the latest FORS2 pipeline[1], together with the latest `EsoRex` package[2] (v. 3.12.3) to apply the standard reduction steps. The bias frames were combined

[1] available at `http://www.eso.org/sci/software/pipelines/`
[2] available at `http://www.eso.org/sci/software/cpl/esorex.html`

using the `fors_bias` recipe. The flat field frames were combined and corrected from the master bias, the bad pixels, and calibrated in wavelengths with the lamp frame and the lamp line catalog (provided by ESO) using the `fors_calib` recipe. All the observing files (including standard star frames) were corrected from corresponding master bias, master flat field, bad pixels, dispersion and curvature coefficients with the `fors_science` recipe. The spectro-photometric calibration was automatically made by the `fors_science` recipe when a spectro-photometric table computed with the `fors_science` recipe of a standard star frame is given together with a corresponding flux table provided by ESO. ESO reduction recipe `fors_science` outputs calibrated spectra for each observing block. The spectra of each observing block and of each observation night were combined, by averaging them, to reduce the noise.

8.3.2 Photometry

Together with Raphaël Wicker, we have used the latest EFOSC2 pipeline[3], with the `EsoRex` to apply the standard reduction steps to the raw photometric data. `EsoRex` has standardised the recipes for all the instruments. Therefore, the reduction steps were the same as for the spectroscopic observations with VLT/FORS2. The bias frames were combined using the `efosc_bias` recipe. The flat field frames were combined and corrected from the master bias, the bad pixels with the `efosc_calib` recipe. All of the observing files were corrected from corresponding master bias, master flat field, bad pixels, dispersion and curvature coefficients with the `efosc_science` recipe. The photometric calibration was automatically made by the `efosc_science` recipe with the zero points provided by the EFOSC2 pipeline. Images of each observing block and observing night were combined to reduce the noise of the background, and to fill the inter-pixel regions.

The sources were extracted in the combined images, with SExtractor (Bertin and Arnouts, 1996). The resulting source catalogue provides magnitudes in the R, V, and z bands. The S/N of the recovered sources was set at S/N > 1.5 σ for the systematic analysis of the images. Stars from GAIA DR2 in the field of each cluster were removed when they matched SExtractor sources in a radius of 5". False detections due to cosmic rays were also removed from the catalogue based on obvious bias in the magnitudes provided by SExtractor.

8.4 Example of results for PSZ1 G231.05-17.32

I present here one example of the cluster PSZ1 G231.05-17.32, for which both photometric and spectroscopic observations have been performed. This cluster was observed in photometry with NTT/EFOSC2 the night of the 27/28th of March 2014, in the R and Z bands, during 24 min each. Spectroscopy with VLT/FORS2 was observed the night of the 13/14 of September 2015 during 10 min and the night of the 9/10 of December 2015 during 42 min. In the left panel of Fig. 8.4, I show the reduced observation of NTT/EFOSC2, and overlaid the sources extracted with SExtractor. I show in the right panel of Fig. 8.4 the galaxy density maps of the cleaned catalogue of photometric sources, constructed with the pyDTFE code (described in Chap. 3.3.1). It clearly shows a source over-density in the left part of the field of view.

Spectra observed with VLT/FORS2 were also reduced following the reduction steps described in Chap. 8.3.1. I have used the on-line webservice MARZ[4] to fit template spectra and thus measure the redshifts of the sources. The five well fitted spectra in the field of view of the cluster PSZ1 G231.05-17.32 are shown in Fig. 8.5 and in Fig. 8.6

For this cluster, where five spectra were well fitted at the same redshift, and here an over-density of sources is seen in the field-of-view, I confirm the cluster, at $z = 0.644 \pm 0.006$.

[3] available at `http://www.eso.org/sci/software/pipelines/`

[4] `https://samreay.github.io/Marz/`

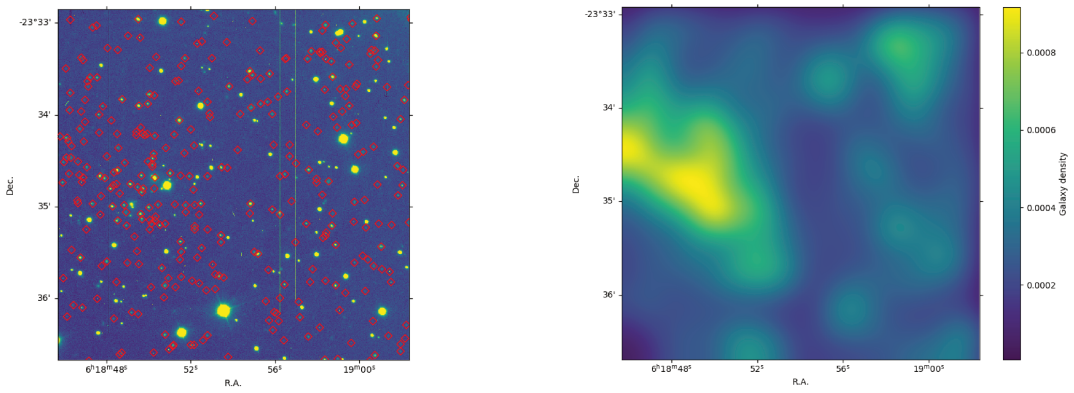

Figure 8.4: Left: combined observations of NTT/EFOSC2 of the cluster PSZ1 G231.05-17.32 in the R band. Red diamonds represent the sources extracted from SExtractor with S/N > 1.5 σ, where I have removed the stars from the GAIA DR2 catalogue that matched a radius of 5", and the fake detections due to cosmic rays. Right: galaxy density map of the cleaned catalogue constructed with the pyDTFE (described in Chap. 3.3.1), and smoothed at 0.1'.

8.5 Summary

I have conducted the data reduction of the spectroscopic follow-up of 49 galaxy clusters observed with VLT/FORS2. Due to an issue during the selection of the spectroscopic targets, the analysis of the confirmation of the galaxy members and the measure of the redshifts have to be combined with the photometry. Photometric data with NTT/EFOSC2 had thus been re-reduced.

The confirmation of the clusters and the analysis of the spectroscopic sources require a very detailed analysis, of about 2500 spectra, and also require confirmation of sources by visual identifications. This analysis is still ongoing to date.

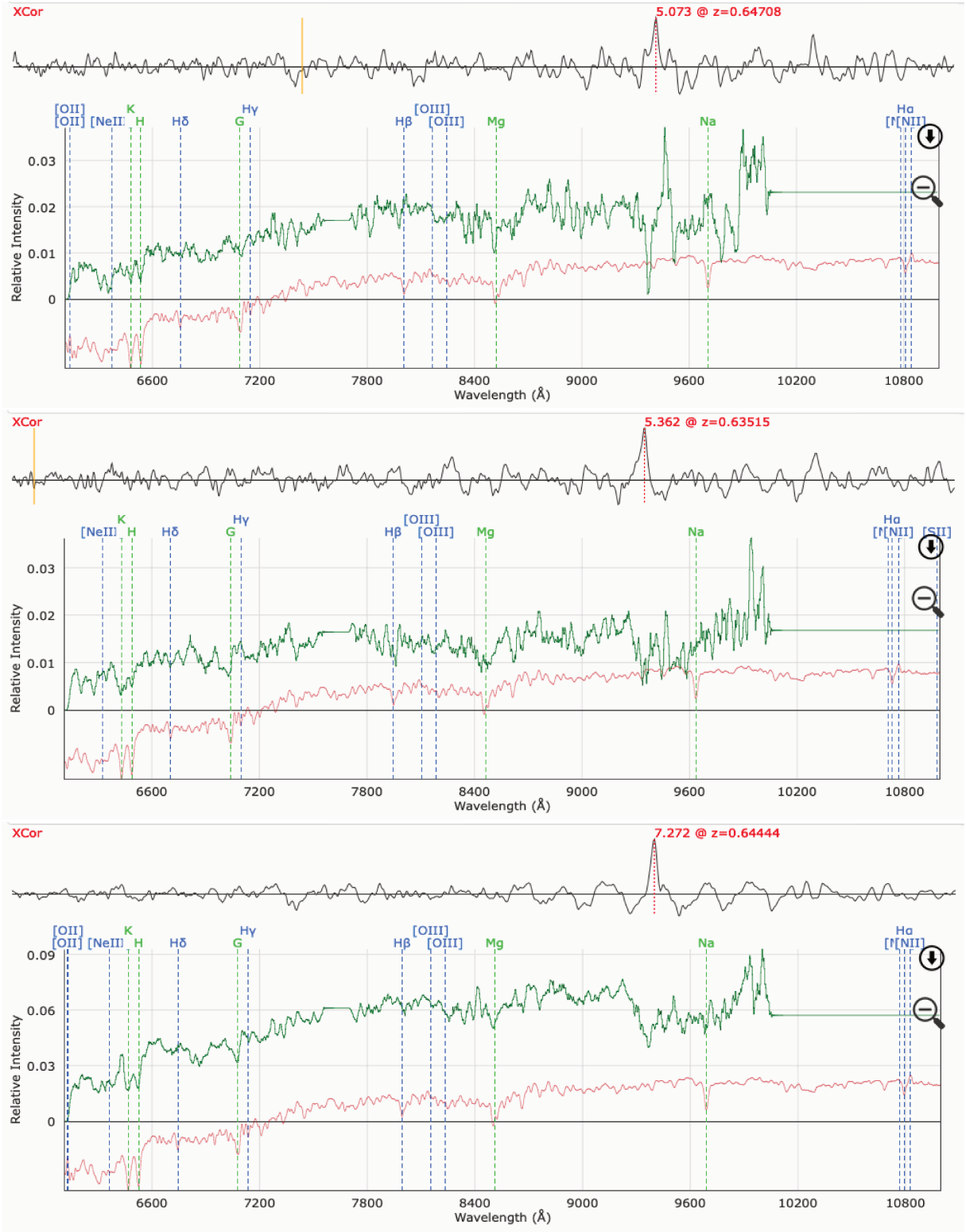

Figure 8.5: The five reduced spectra well fitted from MARZ, all at the same redshifts, $z = 0.644 \pm 0.006$. The three first spectra. The two other are in Fig. 8.6.

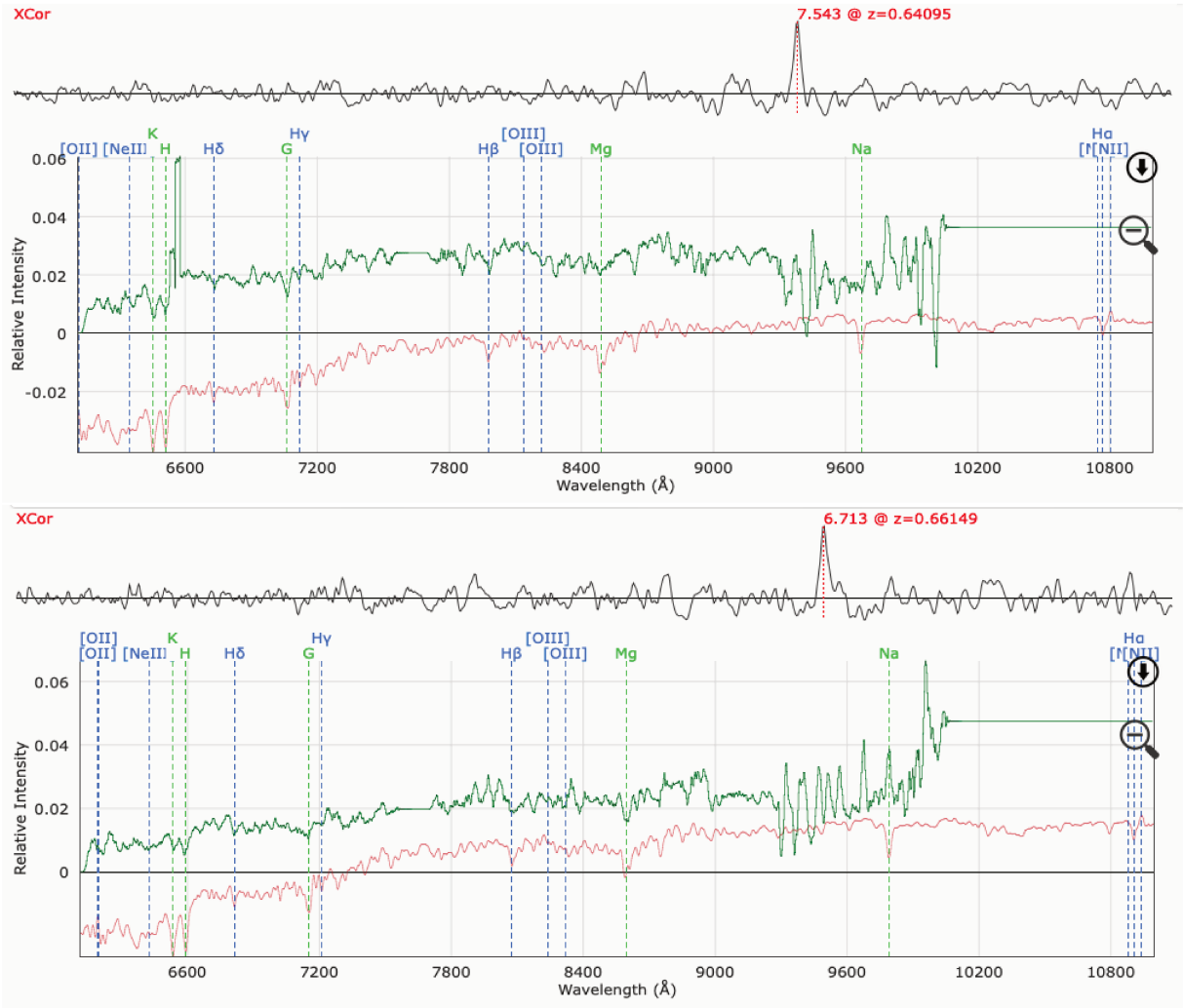

Figure 8.6: The five reduced spectra well fitted from MARZ, all at the same redshifts, $z = 0.644 \pm 0.006$. The two other spectra. The first three ones are in Fig. 8.5.

Chapter 9

MUSE observation of an exceptional giant arc

Contents

Résumé

Lors des analyses des images prises avec le NTT/EFOSC2 dans le cadre du large programme de l'ESO pour confirmer certains amas *Planck*, un arc gravitationnel exceptionellement grand et lumineux a été découvert autour de l'amas PSZ1 G311.65-18.48. Cet objet a été sujet à des observations complémentaires, notamment avec l'instrument MUSE. Je décris ici l'analyse de données de ces observations, qui serviront par la suite à une étude plus poussée dus gaz moléculaire dans la galaxie lentillée à l'aide d'observation en sub-mm, comme avec ALMA.

9.1 Introduction

By construction, the SZ-selected cluster sample seen in spectroscopy is essentially mass limited and contains massive clusters ($> 5 \times 10^{14}\,M_\odot$) beyond $z \sim 0.5$. These clusters are ideal for the observations of lensed background galaxies (LBG). The NTT/EFOSC2 data from the ESO Large Programme have lead to the discovery of the brightest giant arc ever seen in the optical/NIR at redshift $z = 2.37$, with an observed total (AB) magnitudes of $(R, z, J, K_s) = (18.03, 17.92, 17.66, 17.28)$, lensed by the cluster PSZ1 G311.65-18.48 at redshift $z = 0.44$ (Dahle et al., 2016). The combined NTT/EFOSC2 observation in the R band of the field of view of the cluster PSZ1 G311.65-18.48 with the giant arc is shown in the left panel of Fig. 9.1.

This newly discovered arc is much brighter than any previously known cluster-lensed LBG such as the Cosmic Eye (Smail et al., 2007, $z = 3.0$ R=20 AB), or RCSGA 032727-132609 (Wuyts et al., 2010, $z = 1.7$, 38" long arc with R=19 AB). Its exceptional observed length (55"; the width is < 1") indicates extremely high magnification ($\mu \gtrsim 50$). In addition, its redshift ($z = 2.37$) puts it in a fortuitous location where the spectral features ranging from Ly_α to H_α are readily accessible for ground-based

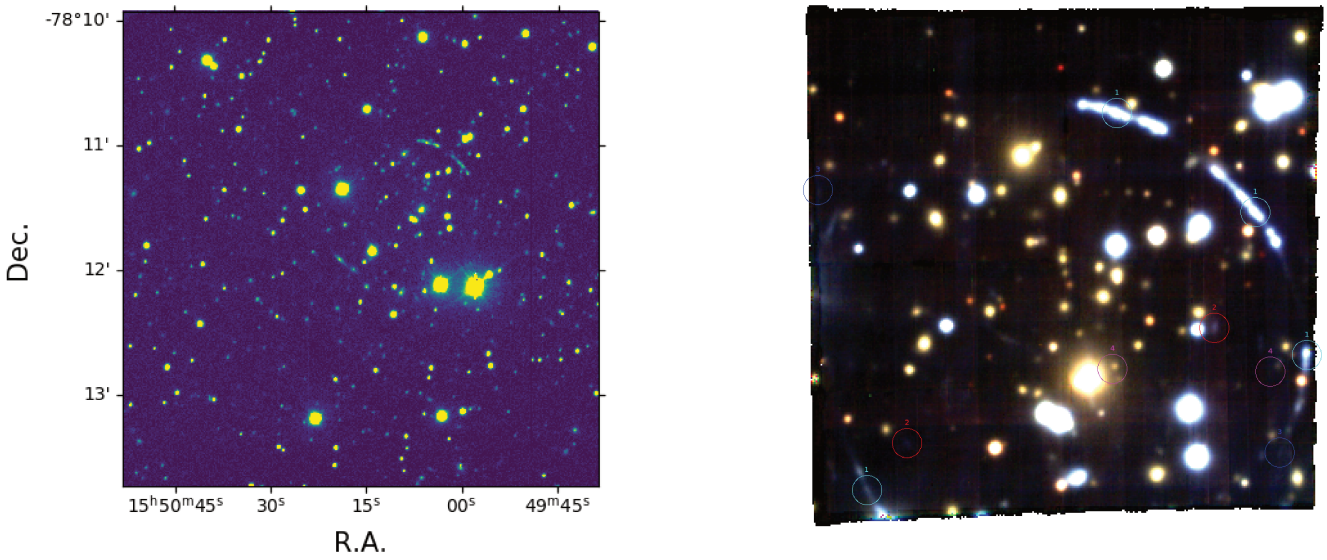

Figure 9.1: Left: NTT/EFOSC2 reduced image of the field of the PSZ1 G311.65-18.48 cluster, where the giant arc has first been discovered. The giant arc is seen on the upper left part. Right: reconstructed RGB image of the giant arc from MUSE. Coloured circles indicate the multiple lensed sources. The light blue circles indicate the four images of the lensed galaxy at $z = 2.37$ producing the giant arc.

observations, making it an ideal target for additional multi-wavelength follow-up observations. Given the average surface brightness of the outstanding giant arc PSZ1-ARC G311.6602-18.4624 of $R \simeq I \simeq z \simeq 22.4$ mag arcsec^{-2} (AB magnitude) and typical peak surface brightnesses of $R \simeq I \simeq z \approx 24$ mag arcsec^{-2} for other lensed background galaxies, observations of the arc needed a 8-10 m class telescope. The *Multi Unit Spectroscopic Explorer* (MUSE, Kelz et al., 2012), installed in the UT4 (Yepun) telescope in the VLT in Paranal, is by far the most efficient ground-based facility and instrument for this observation. The field of view of the VLT/MUSE in Wide Field Mode is very well suited for the observation of the full 55" extent of the giant arc. Moreover, MUSE has the unique capability of providing spectra, over the full 4650-9350Åwavelength domain, of $\sim$ 100 objects at a time.

As part of a DDT program in P97, 297.A-5012(A) (PI: N. Aghanim), shallow observations were made in May 2016 with the VLT/MUSE. A 1×1 arcmin^2 field around the galaxy cluster PSZ1 G311.65-18.48 ($\alpha = 237.531129°$, $\delta = -78.18989°$, $z = 0.444$) and the giant arc was observed for 50 min of time exposure (two pointings of 25 min each).

9.2 Data reduction

During my PhD thesis, I have visited Johan Richard in the Centre de Recherche Astrophysique de Lyon (CRAL), where I have reduced the data cubes of the two pointings. The data cubes were reduced following a standard procedure (bias, flat field, wavelength calibration, sky-subtraction) with `esorex`, and combined using the publicly available MUSE ESO pipeline[1] (Weilbacher et al., 2012; Weilbacher et al., 2014), the Zurich Atmosphere Purge code (ZAP, Soto et al., 2016), and the MUSE Python Data Analysis Framework `mpdaf` (Piqueras et al., 2017).

A reconstructed RGB image of the MUSE data is shown in the right panel of Fig. 9.1. In the image,

[1] `https://www.eso.org/sci/software/pipelines/muse/muse-pipe-recipes.html`

the small circles indicate the multiple images of four lensed sources (one colour for each lensed source). The four images of the galaxy at $z = 2.37$ producing the giant arc are in light blue. The spectrum of the galaxy producing the giant arc, based on the four multiple images of the arc detected in the MUSE image, is shown in Fig. 9.2. Overlaid in red are the positions of the [CIV], [HeII], [OIII], [CIII] and [CII] lines.

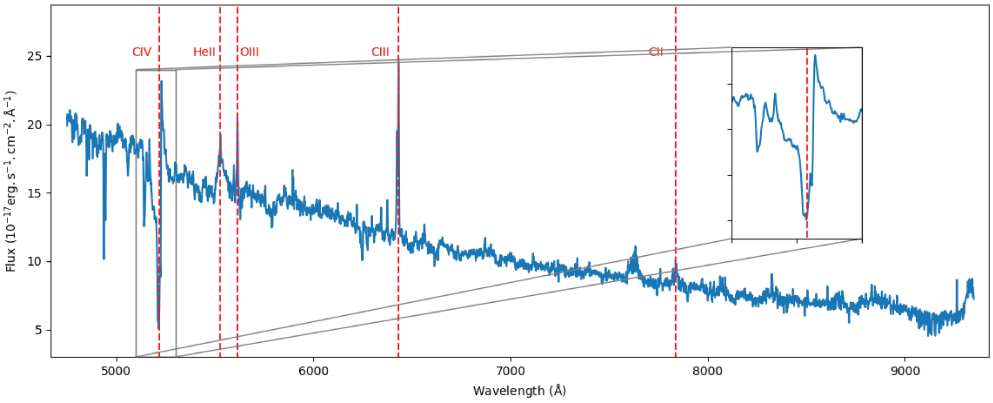

Figure 9.2: Spectrum of the giant arc.

9.3 Catalogue of sources

Over a field of view of $1' \times 1'$, VLT/MUSE provides spectra, over the full 4650 – 9350Å wavelength domain. I have constructed, based on the spectral cube, a white-light image of the field (integration along the spectral axis). Based on this white image, I have run SExtractor to detect the objects with reliable continuum emissions. A total of 180 objects were detected with position, magnitude, and ellipticity measurements. I have then extracted the spectra on the positions of the sources to determine the redshifts. In collaboration with J. Richard (CRAL), I have measured the redshift based on absorption or emission spectral lines, and classified the objects into stars, galaxy cluster members, and foreground or background galaxies. The galaxy cluster members were identified with the 4000Å break around 5800Å. For the background and foreground galaxies, the ones at $z < 1.5$ were identified from [OII] 3727Å emission and the ones at $2.8 < z < 6.6$ were identified by the Ly$_\alpha$ emission, when detected. If none of these emission lines were seen, redshifts have been obtained from atmospheric absorption lines in the rest-frame UV, or from [CIII] 1909Å in emission, which are observable for galaxies in the range $1.4 < z < 3.8$. The spectral resolution of MUSE is also sufficient to resolve the [CIII] 1909Å doublet, helping in the measure of the redshift.

The VLT/MUSE spectral coverage allows us to firmly identify multiple lensed background sources in the FoV and to measure their redshifts at once. The images of the giant arc at $z = 2.369$ were identified with the strong [CIII] doublet around 6425Å, combined with the [HeII] emission around 5520Å, and the [SiII] absorption around 5130Å, and the [CIV] features around 5210Å. These multiply lensed sources with secure redshifts will be used to construct the lensing mass model of PSZ1 G311.65-18.48 from which crucial information on local magnification can be derived.

A catalogue of foreground galaxies was extracted, with reliable positions and spectroscopic redshifts, including one double image of the same source at $z = 1.1864$. In addition, I have run *muselet* (e.g., Richard et al., 2015) to search for faint line emitters not detected by their continuum emission to extract isolated emission line objects on the narrow-band cube. I have found four interesting objects at high

redshift, including one double image of the same source at $z = 3.504$, their index start with 1000 in Tab. 9.1. Four other sources were detected after visual inspection, with also one double image at $z = 4.614$ (with index 2000 in Tab. 9.1). I present the four recovered multiple images that will be used to construct the mass model of the cluster PSZ1 G311.65-18.48 in Tab. 9.1.

In parallel to our own analysis, this giant arc has been studied by an independent team with the Hubble Space Telescope (Rivera-Thorsen et al., 2019), to construct a lens model with a high accuracy.

	Index	RA	DEC	Redshift	A_WORLD	B_WORLD	THETA	MAG_AUTO
3	5	237.563288	-78.197442	2.3700	0.000170	0.000050	62.23	18.8026
23	28	237.522244	-78.184778	2.3695	0.000249	0.000097	19.04	17.0217
40	48	237.499542	-78.188104	2.3694	0.000214	0.000085	46.33	17.6982
117	127	237.491025	-78.192897	2.3697	0.000198	0.000162	-59.13	18.6497
99	109	237.506232	-78.191985	1.1864	0.000113	0.000078	-76.35	20.8396
149	159	237.556697	-78.195855	1.1864	0.000059	0.000037	62.50	22.1555
171	1001	237.571100	-78.187364	3.5040	1.000000	1.000000	0.00	NaN
172	1002	237.495519	-78.196176	3.5040	1.000000	1.000000	0.00	NaN
177	2003	237.497030	-78.193454	4.6140	1.000000	1.000000	0.00	NaN
178	2004	237.522970	-78.193377	4.6140	1.000000	1.000000	0.00	NaN

Table 9.1: Table of the four multiple images detected, used to generate the lensing model. The four first rows are each positions of the four arcs of the giant arc. The second part is the multiple image detected with the continuum and SExtractor. The third part is the multiple image detected by `muselet`. The fourth part is the multiple image detected by visual inspections.

9.4 Summary

I have described the MUSE observation and data reduction of an exceptional giant arc discovered in the *Planck*/ESO large programme. This giant arc is produced by the galaxy cluster PSZ1 G311.65-18.48 at $z = 0.44$, which lenses a star forming galaxy at $z = 2.37$. This analysis is still ongoing to date.

Based on the results obtained in this study, a proposal of observing time with ALMA was performed two times (PI: A. Beelen). The ALMA facilities would allow the detection and the characterisation of gas in resolved clumps inside the lensed galaxy, and therefore study at an unprecedented scale the star formation process in a galaxy at the epoch of the peak of star formation (which occurred at $z \sim 2$).

Part IV

Conclusions

During my PhD, I have explored the properties of the baryonic matter in the Cosmic Web, focusing (i) on the galaxies and (ii) on the hot and diffuse gas, either in the intergalactic medium located along filaments or inside galaxy clusters.

First, I have investigated the potential of using galaxy properties to characterise the largest structures in which they are lying. To do so, I have fully developed a machine learning algorithm to estimate the most relevant galaxy properties that allow their segregation into two types (passive and active) and the measurement of their star formation rates (SFR) and their stellar masses ($M_\star$). This machine learning algorithm is based on random forest, and was trained on infra-red data from the WISE satellite based on SFR and $M_\star$ estimated in optical from the SDSS. The trained model can be applied to WISE counterpart galaxies up to $z = 0.3$. It provides unbiased results with redshift or galaxy type. The model performs good with standard deviation of $\sigma_{\rm SFR} = 0.38$ dex and $\sigma_{M_\star} = 0.16$ dex. This method can be extended to higher redshifts and can be improved by choosing even better quality catalogue as training set for the machine learning algorithm. For example, choosing as training set the SFR and stellar masses from the GAMA galaxy surveys may allow an estimation of these quantities up to higher redshifts (i.e., $z \sim 0.5$). Significant improvements may also be possible with the future large galaxy surveys coming with very high statistics like Euclid or WFIRST.

One product of my PhD was a value-added galaxy catalogue, based on the WISExSCOS photometric redshift catalogue. Using the trained random forest model described above, I have estimated the SFR and the stellar masses for the sources of the WISExSCOS catalogue in the range $0.1 < z < 0.3$. Based on this catalogue, I have investigated the link between galaxy properties and the hot gas and in particular the quenching of the star formation activity of the galaxies. I have focused on an exceptionally dense filament: a bridge of matter between the two merging galaxy clusters A399 and A401. This particular pair was previously known and characterised by its gas content in X-rays and in SZ. In my PhD, I have modelled the gas using only the SZ effect, constrained its properties, and for the first time studied the star formation activity of galaxies lying in the bridge. These galaxies were found to be passive, as the galaxies in the two clusters, confirming the hypothesis of a pre-merger stage where the gas lying in the bridge is heated by the future collapse of the two clusters. This analysis could be completed with (i) observations of the galaxy members in optical and in the mm/sub-mm (SITELLE, ALMA) as well as with (ii) higher resolution in SZ (NIKA2). This bridge of matter is a perfect laboratory to study for the first time in details the relation in a filament between the hot gas and the galaxy properties (like their molecular gas content or their star formation efficiency).

I have extended my work on the bridge between A399 and A401 to a more significant number of cosmic filaments. In order to do so, I have performed a statistical analysis of the galaxy properties in a new sample of larger and less dense filaments. These cosmic filaments were detected thanks to the SDSS galaxy spatial distribution with a method based on the gradient of the galaxy density field named DisPerSE. I have identified 5559 filaments in the redshift range $0.1 < z < 0.3$. I have then studied profiles of galaxy density and of galaxy properties, namely SFR and $M_\star$ around filaments of different sizes. The analysis exhibited an excess of passive, transitioning, and active galaxies around filaments. I have explored the relation between the galaxy properties and the hot gas. More specifically, I have explored the link between the profile of quenching of galaxies, obtained from their passivity, and the profile of the surrounding hot gas, using the outputs of a statistical analysis of the properties of the hot gas around cosmic filaments. No firm conclusion on the relation between both profiles can be drawn for now considering the lack of statistics, but still an MCMC analysis shows commonalities between the allowed value of the shape of the quenching profiles and of the shape of the gas profile. This type of study may

be repeated with more significant conclusions using the future large galaxy surveys with higher statistics like Euclid and WFIRST.

I have then explored the possibility of using deep learning algorithms to improve the detection of the hot and diffuse gas around the structures of the Cosmic Web via the SZ effect. Previous studies have already successfully detect the SZ effect and have constructed full-sky y maps or have detected galaxy clusters using the *Planck* frequency maps. However, the SZ effect is very faint, quite hard to detect, and none of the methods developed for now are optimal for the detection of faint sources. In this context, I have developed a deep learning algorithm to detect the SZ effect in the *Planck* data. In this proof-of-concept, I have detected more than 10,000 galaxy cluster candidates, that together show significant counterparts in other wavelengths by stacking maps at their positions (like presence of dark matter by stacking CMB lensing map or confirmation of diffuse hot gas by stacking ROSAT map in X-rays). This shows promising results for individual detections and characterisation of the hot gas around the largest structures of the Cosmic Web. This method can be improved if one uses higher resolution SZ maps like the ones from the ACT or the SPT surveys. In the future, the galaxy cluster candidates detected thanks to this method may be compared to the next generation of full sky surveys allowing the confirmation of the presence of diffuse gas (i.e., with SRG/eROSITA in the X-rays). This could lead to a relatively high number of new detected galaxy clusters or complex structures in the full sky, mostly expected at intermediate or high redshift (i.e., $z \geq 0.5$).

Finally I worked on the data reduction and analysis of an ESO large programme that aimed at confirming 129 southern SZ-cluster candidates from *Planck* via the detection of their galaxy members in optical bands g, r, and z and in spectroscopy from 4000 to 8000 Angstrom. I have developed pipelines to reduce both photometric data obtained with NTT/EFOSC2 and spectroscopic data obtained with VLT/FORS2. Among the 129 SZ sources, only a handful are false detections. Estimated redshifts can be provided for all the confirmed clusters. A total of 49 clusters with estimated redshifts $z > 0.5$ were observed with VLT/FORS2. The spectroscopic confirmation of about 10 galaxy clusters with at least 10 galaxy members per cluster is secured but the analysis of the full sample of $\sim$ 3000 obtained spectra is still ongoing. Within the sample of 129 SZ sources of this programme, a giant arc was detected around one of the observed clusters. This giant arc has been observed with the MUSE facilities in spectroscopy and in photometry. I present the data reduction, and the final products, i.e., (i) well resolved spectra of the giant arc showing emission lines from regions of active star formation and (ii) an integrated image of the giant arc together with of multiple lensed sources. These data will be the basis for the construction of a lensing mass model to characterise both the lens and the high-redshift background object. We proposed to observe this giant arc with ALMA in order to map its molecular gas content at a high angular resolution. This will permit to study the resolved star formation regions in a galaxy at the epoch of the peak of star formation, i.e., $z \sim 2$.

As a conclusion, the work achieved in this PhD confirmed that using new techniques of machine learning is relevant and very powerful to efficiently measure the physical properties of large numbers of objects (i.e., galaxy surveys) but also to improve the detection and characterisation of faint signals (i.e., hot gas in filaments and clusters via the SZ effect). This is an important step for the study of the interplay between the galaxies and their environment inside the Cosmic Web. This interaction and the possible galaxy quenching induced is an important agent for the evolution of galaxies. During these three years, I have set-up new methods and applied them to a number of different cases that all together shade a new light on the baryonic content of the Cosmic Web (in clusters, bridges of matter, and cosmic filaments). This work can now be improved and extended to new cases. It opens new perspectives and several projects are currently being implemented on this purpose.

U-Net and more generally Convolutional Neural Networks are very powerful, and I intend in the future to keep a close eye on such new statistical developments in data analysis, in order to continue to apply them to various cases in observational cosmology. The increasingly large number of data expected in the near future together with modern data analysis tools are thus very promising for cosmology and in particular for the study of the Cosmic Web. In the context of the big data era in astronomy, well represented for example by the upcoming Square Kilometer Array (SKA) survey, developing computationally and timely efficient tools for data reduction or source extraction in multi-channel data seems to me to be of a highest priority and relevance.

Bibliography

[1] George O. Abell, Jr. Corwin Harold G., and Ronald P. Olowin. "A Catalog of Rich Clusters of Galaxies". In: *The Astrophysical Journal Supplement Series* 70 (1989), p. 1. DOI: 10.1086/191333.

[2] R. Adam et al. "Mapping the kinetic Sunyaev-Zel'dovich effect toward MACS J0717.5+3745 with NIKA". In: *A&A* 598, A115 (2017), A115. DOI: 10.1051/0004-6361/201629182. arXiv: 1606.07721 [astro-ph.CO].

[3] R. Adam et al. "Substructure and merger detection in resolved NIKA Sunyaev-Zel'dovich images of distant clusters". In: *A&A* 614, A118 (2018), A118. DOI: 10.1051/0004-6361/201731950. arXiv: 1712.01836 [astro-ph.CO].

[4] Jennifer K. Adelman-McCarthy et al. "The Sixth Data Release of the Sloan Digital Sky Survey". In: *ApJS* 175.2 (2008), pp. 297–313. DOI: 10.1086/524984. arXiv: 0707.3413 [astro-ph].

[5] Susmita Adhikari et al. "Kinematics of Cluster Galaxies and Their Relation to Galaxy Evolution". In: *ApJ* 878.1, 9 (2019), p. 9. DOI: 10.3847/1538-4357/ab1a39. arXiv: 1806.09672 [astro-ph.GA].

[6] N. Aghanim et al. "PACT I: Combining ACT and Planck data for optimal detection of tSZ signal". In: *arXiv e-prints*, arXiv:1902.00350 (2019), arXiv:1902.00350. arXiv: 1902.00350 [astro-ph.CO].

[7] N. Aghanim et al. "The Good, the Bad, and the Ugly: Statistical quality assessment of SZ detections". In: *A&A* 580, A138 (Aug. 2015), A138. DOI: 10.1051/0004-6361/201424963. arXiv: 1409.6543.

[8] T. Akahori and K. Yoshikawa. "Non-Equilibrium Ionization State and Two-Temperature Structure in the Linked Region of Abell 399 and Abell 401". In: *PASJ* 60 (Aug. 2008), pp. L19–L22. DOI: 10.1093/pasj/60.4.L19. arXiv: 0805.2453.

[9] H. Akamatsu et al. "Properties of the cosmological filament between two clusters: A possible detection of a large-scale accretion shock by Suzaku". In: *A&A* 606, A1 (2017), A1. DOI: 10.1051/0004-6361/201730497. arXiv: 1704.05843 [astro-ph.HE].

[10] Katherine Alatalo et al. "Catching Quenching Galaxies: The Nature of the WISE Infrared Transition Zone". In: *ApJ* 794, L13 (2014), p. L13. DOI: 10.1088/2041-8205/794/1/L13. arXiv: 1409.2489 [astro-ph.GA].

[11] S. W. Allen et al. "Improved constraints on dark energy from Chandra X-ray observations of the largest relaxed galaxy clusters". In: *MNRAS* 383.3 (2008), pp. 879–896. DOI: 10.1111/j.1365-2966.2007.12610.x. arXiv: 0706.0033 [astro-ph].

[12] Mehmet Alpaslan et al. "Galaxy And Mass Assembly (GAMA): stellar mass growth of spiral galaxies in the cosmic web". In: *MNRAS* 457.3 (2016), pp. 2287–2300. DOI: 10.1093/mnras/stw134. arXiv: 1601.03391 [astro-ph.GA].

[13] Mehmet Alpaslan et al. “Galaxy And Mass Assembly (GAMA): trends in galaxy colours, morphology, and stellar populations with large-scale structure, group, and pair environments”. In: *MNRAS* 451.3 (2015), pp. 3249–3268. DOI: 10.1093/mnras/stv1176. arXiv: 1505.05518 [astro-ph.GA].

[14] AMI Consortium et al. “A joint analysis of AMI and CARMA observations of the recently discovered SZ galaxy cluster system AMI-CL J0300+2613”. In: *MNRAS* 433.3 (2013), pp. 2036–2046. DOI: 10.1093/mnras/stt873. arXiv: 1305.6655 [astro-ph.CO].

[15] AMI Consortium et al. “Detailed Sunyaev-Zel’dovich study with AMI of 19 LoCuSS galaxy clusters: masses and temperatures out to the virial radius”. In: *MNRAS* 425.1 (2012), pp. 162–203. DOI: 10.1111/j.1365-2966.2012.21419.x. arXiv: 1205.7067 [astro-ph.CO].

[16] AMI Consortium et al. “Sunyaev-Zel’dovich observations with AMI of the hottest galaxy clusters detected in the XMM-Newton Cluster Survey”. In: *MNRAS* 433.4 (2013), pp. 2920–2937. DOI: 10.1093/mnras/stt932. arXiv: 1305.6654 [astro-ph.CO].

[17] M. A. Aragon-Calvo. “Classifying the large-scale structure of the universe with deep neural networks”. In: *MNRAS* 484 (Apr. 2019), pp. 5771–5784. DOI: 10.1093/mnras/stz393. arXiv: 1804.00816.

[18] M. A. Aragon-Calvo and L. F. Yang. “The hierarchical nature of the spin alignment of dark matter haloes in filaments.” In: *MNRAS* 440 (2014), pp. L46–L50. DOI: 10.1093/mnrasl/slu009. arXiv: 1303.1590 [astro-ph.CO].

[19] Miguel A. Aragon-Calvo, Mark C. Neyrinck, and Joseph Silk. “Galaxy Quenching from Cosmic Web Detachment”. In: *arXiv e-prints*, arXiv:1607.07881 (2016), arXiv:1607.07881. arXiv: 1607.07881 [astro-ph.GA].

[20] Miguel A. Aragón-Calvo, Rien van de Weygaert, and Bernard J. T. Jones. “Multiscale phenomenology of the cosmic web”. In: *MNRAS* 408.4 (2010), pp. 2163–2187. DOI: 10.1111/j.1365-2966.2010.17263.x. arXiv: 1007.0742 [astro-ph.CO].

[21] M. Arnaud et al. “The universal galaxy cluster pressure profile from a representative sample of nearby systems (REXCESS) and the Y_{SZ} - M_{500} relation”. In: *A&A* 517, A92 (2010), A92. DOI: 10.1051/0004-6361/200913416. arXiv: 0910.1234 [astro-ph.CO].

[22] S. Arnouts and O. Ilbert. *LePHARE: Photometric Analysis for Redshift Estimate*. 2011. ascl: 1108.009.

[23] S. Arnouts et al. “Measuring and modelling the redshift evolution of clustering: the Hubble Deep Field North”. In: *MNRAS* 310 (Dec. 1999), pp. 540–556. DOI: 10.1046/j.1365-8711.1999.02978.x. eprint: astro-ph/9902290.

[24] Ivan K. Baldry et al. “Quantifying the Bimodal Color-Magnitude Distribution of Galaxies”. In: *ApJ* 600.2 (2004), pp. 681–694. DOI: 10.1086/380092. arXiv: astro-ph/0309710 [astro-ph].

[25] J. A. Baldwin, M. M. Phillips, and R. Terlevich. “Classification parameters for the emission-line spectra of extragalactic objects”. In: *PASP* 93 (Feb. 1981), pp. 5–19. DOI: 10.1086/130766.

[26] M. L. Balogh et al. “Differential Galaxy Evolution in Cluster and Field Galaxies at z˜0.3”. In: *ApJ* 527 (Dec. 1999), pp. 54–79. DOI: 10.1086/308056. eprint: astro-ph/9906470.

[27] P. Banerjee et al. “An optical catalog of galaxy clusters obtained from an adaptive matched filter finder applied to SDSS DR9 data”. In: *NA* 58 (2018), pp. 61–71. DOI: 10.1016/j.newast.2017.07.008. arXiv: 1704.03529 [astro-ph.CO].

[28] Dalya Baron. "Machine Learning in Astronomy: a practical overview". In: *arXiv e-prints*, arXiv:1904.07248 (2019), arXiv:1904.07248. arXiv: 1904.07248 [astro-ph.IM].

[29] R. Barrena et al. "Optical validation and characterization of Planck PSZ1 sources at the Canary Islands observatories. I. First year of ITP13 observations". In: *A&A* 616, A42 (2018), A42. DOI: 10.1051/0004-6361/201732315. arXiv: 1803.05764 [astro-ph.CO].

[30] I. Bartalucci et al. "Recovering galaxy cluster gas density profiles with XMM-Newton and Chandra". In: *A&A* 608, A88 (2017), A88. DOI: 10.1051/0004-6361/201731689. arXiv: 1709.06570 [astro-ph.CO].

[31] Eric Baxter et al. "The Halo Boundary of Galaxy Clusters in the SDSS". In: *ApJ* 841.1, 18 (2017), p. 18. DOI: 10.3847/1538-4357/aa6ff0. arXiv: 1702.01722 [astro-ph.CO].

[32] Francis Bernardeau and Rien van de Weygaert. "A new method for accurate estimation of velocity field statistics". In: *MNRAS* 279 (1996), p. 693. DOI: 10.1093/mnras/279.2.693. arXiv: astro-ph/9508125 [astro-ph].

[33] M. Bersanelli et al. "Planck pre-launch status: Design and description of the Low Frequency Instrument". In: *A&A* 520, A4 (Sept. 2010), A4. DOI: 10.1051/0004-6361/200912853. arXiv: 1001.3321 [astro-ph.IM].

[34] E. Bertin and S. Arnouts. "SExtractor: Software for source extraction." In: *A&AS* 117 (1996), pp. 393–404. DOI: 10.1051/aas:1996164.

[35] M. Bilicki et al. "Two Micron All Sky Survey Photometric Redshift Catalog: A Comprehensive Three-dimensional Census of the Whole Sky". In: *ApJS* 210, 9 (Jan. 2014), p. 9. DOI: 10.1088/0067-0049/210/1/9. arXiv: 1311.5246 [astro-ph.CO].

[36] M. Bilicki et al. "WISE x SuperCOSMOS Photometric Redshift Catalog: 20 Million Galaxies over 3/pi Steradians". In: *ApJS* 225, 5 (July 2016), p. 5. DOI: 10.3847/0067-0049/225/1/5. arXiv: 1607.01182.

[37] L. E. Bleem et al. "Galaxy Clusters Discovered via the Sunyaev-Zel'dovich Effect in the 2500-Square-Degree SPT-SZ Survey". In: *ApJS* 216.2, 27 (2015), p. 27. DOI: 10.1088/0067-0049/216/2/27. arXiv: 1409.0850 [astro-ph.CO].

[38] N. W. Boggess et al. "The COBE Mission: Its Design and Performance Two Years after Launch". In: *ApJ* 397 (1992), p. 420. DOI: 10.1086/171797.

[39] H. Böhringer et al. "The Northern ROSAT All-Sky (NORAS) Galaxy Cluster Survey. I. X-Ray Properties of Clusters Detected as Extended X-Ray Sources". In: *ApJS* 129.2 (2000), pp. 435–474. DOI: 10.1086/313427. arXiv: astro-ph/0003219 [astro-ph].

[40] H. Böhringer et al. "The ROSAT-ESO Flux Limited X-ray (REFLEX) Galaxy cluster survey. V. The cluster catalogue". In: *A&A* 425 (2004), pp. 367–383. DOI: 10.1051/0004-6361:20034484. arXiv: astro-ph/0405546 [astro-ph].

[41] Th. Boller et al. "Second ROSAT all-sky survey (2RXS) source catalogue". In: *A&A* 588, A103 (2016), A103. DOI: 10.1051/0004-6361/201525648. arXiv: 1609.09244 [astro-ph.HE].

[42] J. Richard Bond, Lev Kofman, and Dmitry Pogosyan. "How filaments of galaxies are woven into the cosmic web". In: *Nature* 380.6575 (1996), pp. 603–606. DOI: 10.1038/380603a0. arXiv: astro-ph/9512141 [astro-ph].

[43] V. Bonjean et al. "Gas and galaxies in filaments between clusters of galaxies. The study of A399-A401". In: *A&A* 609, A49 (2018), A49. DOI: 10.1051/0004-6361/201731699. arXiv: 1710.08699 [astro-ph.CO].

[44] V. Bonjean et al. “Star formation rates and stellar masses from machine learning”. In: *A&A* 622, A137 (Feb. 2019), A137. DOI: 10.1051/0004-6361/201833972. arXiv: 1901.01932.

[45] A. Boucaud et al. “Photometry of high-redshift blended galaxies using deep learning”. In: *arXiv e-prints* (May 2019). arXiv: 1905.01324.

[46] H. Bradt et al. “Evidence for X-Radiation from the Radio Galaxy M87”. In: *ApJL* 150 (1967), p. L199. DOI: 10.1086/180125.

[47] J. Brinchmann et al. “The physical properties of star-forming galaxies in the low-redshift Universe”. In: *MNRAS* 351 (July 2004), pp. 1151–1179. DOI: 10.1111/j.1365-2966.2004.07881.x. eprint: astro-ph/0311060.

[48] M. Brodwin et al. “The Massive and Distant Clusters of WISE Survey. III. Sunyaev-Zel’dovich Masses of Galaxy Clusters at z=1”. In: *ApJ* 806.1, 26 (2015), p. 26. DOI: 10.1088/0004-637X/806/1/26. arXiv: 1410.2355 [astro-ph.CO].

[49] G. Bruzual and S. Charlot. “Stellar population synthesis at the resolution of 2003”. In: *MNRAS* 344 (Oct. 2003), pp. 1000–1028. DOI: 10.1046/j.1365-8711.2003.06897.x. eprint: astro-ph/0309134.

[50] G. Bruzual A. “Spectral evolution of galaxies. I - Early-type systems”. In: *ApJ* 273 (Oct. 1983), pp. 105–127. DOI: 10.1086/161352.

[51] R. A. Burenin et al. “The 400 Square Degree ROSAT PSPC Galaxy Cluster Survey: Catalog and Statistical Calibration”. In: *ApJS* 172.2 (2007), pp. 561–582. DOI: 10.1086/519457. arXiv: astro-ph/0610739 [astro-ph].

[52] D. Burgarella, V. Buat, and J. Iglesias-Páramo. “Star formation and dust attenuation properties in galaxies from a statistical ultraviolet-to-far-infrared analysis”. In: *MNRAS* 360 (July 2005), pp. 1413–1425. DOI: 10.1111/j.1365-2966.2005.09131.x. eprint: astro-ph/0504434.

[53] D. J. Burke et al. “The Southern SHARC catalogue: a ROSAT survey for distant galaxy clusters”. In: *MNRAS* 341.4 (2003), pp. 1093–1108. DOI: 10.1046/j.1365-8711.2003.06378.x.

[54] A. M. Bykov et al. “Structures and Components in Galaxy Clusters: Observations and Models”. In: *SSR* 188.1-4 (2015), pp. 141–185. DOI: 10.1007/s11214-014-0129-4. arXiv: 1512.01456 [astro-ph.CO].

[55] E. T. Byram, T. A. Chubb, and H. Friedman. “Cosmic X-ray Sources, Galactic and Extragalactic”. In: *Science* 152.3718 (1966), pp. 66–71. DOI: 10.1126/science.152.3718.66.

[56] D. Calzetti et al. “The Calibration of Mid-Infrared Star Formation Rate Indicators”. In: *ApJ* 666 (Sept. 2007), pp. 870–895. DOI: 10.1086/520082. arXiv: 0705.3377.

[57] Daniela Calzetti, Anne L. Kinney, and Thaisa Storchi-Bergmann. “Dust Extinction of the Stellar Continua in Starburst Galaxies: The Ultraviolet and Optical Extinction Law”. In: *ApJ* 429 (1994), p. 582. DOI: 10.1086/174346.

[58] N. Cappelluti et al. “eROSITA on SRG. A X-ray all-sky survey mission”. In: *Memorie della Societa Astronomica Italiana Supplementi* 17 (2011), p. 159. arXiv: 1004.5219 [astro-ph.IM].

[59] John E. Carlstrom, Gilbert P. Holder, and Erik D. Reese. “Cosmology with the Sunyaev-Zel’dovich Effect”. In: *ARA&A* 40 (2002), pp. 643–680. DOI: 10.1146/annurev.astro.40.060401.093803. arXiv: astro-ph/0208192 [astro-ph].

[60] Pedro Carvalho, Graça Rocha, and M. P. Hobson. “A fast Bayesian approach to discrete object detection in astronomical data sets - PowellSnakes I”. In: *MNRAS* 393.3 (2009), pp. 681–702. DOI: 10.1111/j.1365-2966.2008.14016.x. arXiv: 0802.3916 [astro-ph].

[61] Pedro Carvalho et al. “PowellSnakes II: a fast Bayesian approach to discrete object detection in multi-frequency astronomical data sets”. In: *MNRAS* 427.2 (2012), pp. 1384–1400. DOI: 10.1111/j.1365-2966.2012.22033.x. arXiv: 1112.4886 [astro-ph.CO].

[62] M. Cautun et al. “Nexus of the Cosmic Web”. In: *Thirteenth Marcel Grossmann Meeting: On Recent Developments in Theoretical and Experimental General Relativity, Astrophysics and Relativistic Field Theories*. Ed. by K. Rosquist. Jan. 2015, pp. 2115–2117. DOI: 10.1142/9789814623995_0371. arXiv: 1211.3574.

[63] Marius Cautun, Rien van de Weygaert, and Bernard J. T. Jones. “NEXUS: tracing the cosmic web connection”. In: *MNRAS* 429.2 (2013), pp. 1286–1308. DOI: 10.1093/mnras/sts416. arXiv: 1209.2043 [astro-ph.CO].

[64] Marius Cautun et al. “Evolution of the cosmic web”. In: *MNRAS* 441.4 (2014), pp. 2923–2973. DOI: 10.1093/mnras/stu768. arXiv: 1401.7866 [astro-ph.CO].

[65] A. Cavaliere and R. Fusco-Femiano. “The Distribution of Hot Gas in Clusters of Galaxies”. In: *A&A* 70 (Nov. 1978), p. 677.

[66] R. Cen and J. P. Ostriker. “Where Are the Baryons?” In: *ApJ* 514 (Mar. 1999), pp. 1–6. DOI: 10.1086/306949. eprint: astro-ph/9806281.

[67] Gilles Chabrier. “Galactic Stellar and Substellar Initial Mass Function”. In: *PASP* 115.809 (2003), pp. 763–795. DOI: 10.1086/376392. arXiv: astro-ph/0304382 [astro-ph].

[68] C. Chang et al. “The Splashback Feature around DES Galaxy Clusters: Galaxy Density and Weak Lensing Profiles”. In: *ApJ* 864.1, 83 (2018), p. 83. DOI: 10.3847/1538-4357/aad5e7. arXiv: 1710.06808 [astro-ph.CO].

[69] Y.-M. Chen et al. “Evolution of the most massive galaxies to z=0.6 - I. A new method for physical parameter estimation”. In: *MNRAS* 421 (Mar. 2012), pp. 314–332. DOI: 10.1111/j.1365-2966.2011.20306.x. arXiv: 1108.4719.

[70] Yen-Chi Chen et al. “Detecting effects of filaments on galaxy properties in the Sloan Digital Sky Survey III”. In: *MNRAS* 466.2 (2017), pp. 1880–1893. DOI: 10.1093/mnras/stw3127. arXiv: 1509.06376 [astro-ph.GA].

[71] R. Chown et al. “Maps of the Southern Millimeter-wave Sky from Combined 2500 deg^2 SPT-SZ and Planck Temperature Data”. In: *ApJS* 239.1, 10 (2018), p. 10. DOI: 10.3847/1538-4365/aae694. arXiv: 1803.10682 [astro-ph.CO].

[72] M. E. Cluver et al. “Calibrating Star Formation in WISE Using Total Infrared Luminosity”. In: *ApJ* 850.1, 68 (2017), p. 68. DOI: 10.3847/1538-4357/aa92c7. arXiv: 1710.03469 [astro-ph.GA].

[73] M. E. Cluver et al. “Galaxy and Mass Assembly (GAMA): Mid-infrared Properties and Empirical Relations from WISE”. In: *ApJ* 782, 90 (Feb. 2014), p. 90. DOI: 10.1088/0004-637X/782/2/90. arXiv: 1401.0837.

[74] Sandrine Codis, Christophe Pichon, and Dmitry Pogosyan. “Spin alignments within the cosmic web: a theory of constrained tidal torques near filaments”. In: *MNRAS* 452.4 (2015), pp. 3369–3393. DOI: 10.1093/mnras/stv1570. arXiv: 1504.06073 [astro-ph.CO].

[75] Jörg M. Colberg, K. Simon Krughoff, and Andrew J. Connolly. “Intercluster filaments in a ΛCDM Universe”. In: *MNRAS* 359.1 (2005), pp. 272–282. DOI: 10.1111/j.1365-2966.2005.08897.x. arXiv: astro-ph/0406665 [astro-ph].

[76] M. Colless et al. “VizieR Online Data Catalog: The 2dF Galaxy Redshift Survey 100k Data Release (2dFGRS Team, 2001)”. In: *VizieR Online Data Catalog*, VII/226 (2003), pp. VII/226.

[77] A. A. Collister and O. Lahav. "ANNz: Estimating Photometric Redshifts Using Artificial Neural Networks". In: *PASP* 116 (Apr. 2004), pp. 345–351. DOI: 10.1086/383254. eprint: astro-ph/0311058.

[78] R. Cruddace et al. "The ROSAT All-Sky Survey: a Catalog of Clusters of Galaxies in a Region of 1 steradian around the South Galactic Pole". In: *ApJS* 140.2 (2002), pp. 239–264. DOI: 10.1086/324519. arXiv: astro-ph/0201069 [astro-ph].

[79] R. M. Cutri et al. *Explanatory Supplement to the AllWISE Data Release Products*. Tech. rep. Nov. 2013.

[80] H. Dahle et al. "Discovery of an exceptionally bright giant arc at z = 2.369, gravitationally lensed by the Planck cluster PSZ1 G311.65-18.48". In: *A&A* 590, L4 (2016), p. L4. DOI: 10.1051/0004-6361/201628297.

[81] Romeel Davé et al. "Baryons in the Warm-Hot Intergalactic Medium". In: *ApJ* 552.2 (2001), pp. 473–483. DOI: 10.1086/320548. arXiv: astro-ph/0007217 [astro-ph].

[82] Kyle S. Dawson et al. "The Baryon Oscillation Spectroscopic Survey of SDSS-III". In: *AJ* 145.1, 10 (2013), p. 10. DOI: 10.1088/0004-6256/145/1/10. arXiv: 1208.0022 [astro-ph.CO].

[83] Anna de Graaff et al. "Probing the missing baryons with the Sunyaev-Zel'dovich effect from filaments". In: *A&A* 624, A48 (2019), A48. DOI: 10.1051/0004-6361/201935159. arXiv: 1709.10378 [astro-ph.CO].

[84] V. de Lapparent, M. J. Geller, and J. P. Huchra. "A slice of the universe". In: *ApJL* 302 (Mar. 1986), pp. L1–L5. DOI: 10.1086/184625.

[85] Gerard de Vaucouleurs. "General Physical Properties of External Galaxies." In: *Handbuch der Physik* 53 (1959), p. 311.

[86] M. Delli Veneri et al. "Stellar Formation Rates for photometric samples of galaxies using machine learning methods". In: *MNRAS* (Mar. 2019). DOI: 10.1093/mnras/stz856. arXiv: 1902.02522 [astro-ph.IM].

[87] K. Dolag et al. "Simulating the physical properties of dark matter and gas inside the cosmic web". In: *MNRAS* 370 (Aug. 2006), pp. 656–672. DOI: 10.1111/j.1365-2966.2006.10511.x. eprint: astro-ph/0511357.

[88] H. Domínguez Sánchez et al. "Transfer learning for galaxy morphology from one survey to another". In: *MNRAS* 484 (Mar. 2019), pp. 93–100. DOI: 10.1093/mnras/sty3497. arXiv: 1807.00807.

[89] A. G. Doroshkevich and S. F. Shandarin. "A statistical approach to the theory of galaxy formation". In: *Sovast* 22 (1978), pp. 653–660.

[90] S. P. Driver et al. "Galaxy and Mass Assembly (GAMA): survey diagnostics and core data release". In: *MNRAS* 413.2 (2011), pp. 971–995. DOI: 10.1111/j.1365-2966.2010.18188.x. arXiv: 1009.0614 [astro-ph.CO].

[91] Simon P. Driver et al. "GAMA: towards a physical understanding of galaxy formation". In: *Astronomy and Geophysics* 50.5 (2009), pp. 5.12–5.19. DOI: 10.1111/j.1468-4004.2009.50512.x. arXiv: 0910.5123 [astro-ph.CO].

[92] Y. Dubois et al. "Dancing in the dark: galactic properties trace spin swings along the cosmic web". In: *MNRAS* 444.2 (2014), pp. 1453–1468. DOI: 10.1093/mnras/stu1227. arXiv: 1402.1165 [astro-ph.CO].

[93] Yohan Dubois et al. “AGN-driven quenching of star formation: morphological and dynamical implications for early-type galaxies”. In: *MNRAS* 433.4 (2013), pp. 3297–3313. DOI: 10.1093/mnras/stt997. arXiv: 1301.3092 [astro-ph.CO].

[94] F. Durret et al. “Searching for filaments and large-scale structure around DAFT/FADA clusters”. In: *A&A* 588, A69 (Apr. 2016), A69. DOI: 10.1051/0004-6361/201527655. arXiv: 1601.07694.

[95] H. Ebeling, A. C. Edge, and J. P. Henry. “MACS: A Quest for the Most Massive Galaxy Clusters in the Universe”. In: *ApJ* 553.2 (2001), pp. 668–676. DOI: 10.1086/320958. arXiv: astro-ph/0009101 [astro-ph].

[96] H. Ebeling et al. “The ROSAT Brightest Cluster Sample - I. The compilation of the sample and the cluster log N-log S distribution”. In: *MNRAS* 301.4 (1998), pp. 881–914. DOI: 10.1046/j.1365-8711.1998.01949.x. arXiv: astro-ph/9812394 [astro-ph].

[97] Harald Ebeling, Christopher R. Mullis, and R. Brent Tully. “A Systematic X-Ray Search for Clusters of Galaxies behind the Milky Way”. In: *ApJ* 580.2 (2002), pp. 774–788. DOI: 10.1086/343790.

[98] D. Eckert et al. “Warm-hot baryons comprise 5-10 per cent of filaments in the cosmic web”. In: *Nature* 528 (Dec. 2015), pp. 105–107. DOI: 10.1038/nature16058. arXiv: 1512.00454.

[99] L. O. V. Edwards et al. “Spitzer Observations of Abell 1763. I. Infrared and Optical Photometry”. In: *AJ* 139 (Feb. 2010), pp. 434–446. DOI: 10.1088/0004-6256/139/2/434. arXiv: 1009.5752.

[100] D. Elbaz et al. “GOODS-Herschel: an infrared main sequence for star-forming galaxies”. In: *A&A* 533, A119 (2011), A119. DOI: 10.1051/0004-6361/201117239. arXiv: 1105.2537 [astro-ph.CO].

[101] D. Elbaz et al. “The reversal of the star formation-density relation in the distant universe”. In: *A&A* 468 (June 2007), pp. 33–48. DOI: 10.1051/0004-6361:20077525. eprint: astro-ph/0703653.

[102] S. D. Epps and M. J. Hudson. “The Weak Lensing Masses of Filaments between Luminous Red Galaxies”. In: *MNRAS* 468 (July 2017), pp. 2605–2613. DOI: 10.1093/mnras/stx517. arXiv: 1702.08485.

[103] J. Erler et al. “Evidence for a pressure discontinuity at the position of the Coma relic from Planck Sunyaev-Zel’dovich effect data”. In: *MNRAS* 447.3 (2015), pp. 2497–2502. DOI: 10.1093/mnras/stu2750. arXiv: 1411.5476 [astro-ph.CO].

[104] Ian N. Evans et al. “The Chandra Source Catalog”. In: *ApJS* 189.1 (2010), pp. 37–82. DOI: 10.1088/0067-0049/189/1/37. arXiv: 1005.4665 [astro-ph.HE].

[105] A. C. Fabian, C. B. Peres, and D. A. White. “A linear X-ray structure in the cluster A399: evidence of a past interaction with A401”. In: *MNRAS* 285 (Mar. 1997), pp. L35–L39. DOI: 10.1093/mnras/285.4.L35.

[106] D. Fadda et al. “Starburst Galaxies in Cluster-feeding Filaments Unveiled by Spitzer”. In: *ApJL* 672 (Jan. 2008), p. L9. DOI: 10.1086/526457. arXiv: 0711.2561.

[107] A. Fontana et al. “The fraction of quiescent massive galaxies in the early Universe”. In: *A&A* 501.1 (2009), pp. 15–20. DOI: 10.1051/0004-6361/200911650. arXiv: 0901.2898 [astro-ph.GA].

[108] D. Foreman-Mackey et al. “emcee: The MCMC Hammer”. In: *PASP* 125 (Mar. 2013), p. 306. DOI: 10.1086/670067. arXiv: 1202.3665 [astro-ph.IM].

[109] Y. Fujita et al. “High Metallicity of the X-Ray Gas Up to the Virial Radius of a Binary Cluster of Galaxies: Evidence of Galactic Superwinds at High-Redshift”. In: *PASJ* 60 (Jan. 2008), S343–S349. DOI: 10.1093/pasj/60.sp1.S343. arXiv: 0705.2017.

[110] Y. Fujita et al. “The X-Ray Structure of A399 and A401: A Pre-Merging Cluster Pair”. In: *PASJ* 48 (Apr. 1996), pp. 191–200. DOI: 10.1093/pasj/48.2.191. eprint: astro-ph/9602059.

[111] M. Fukugita, C. J. Hogan, and P. J. E. Peebles. “The Cosmic Baryon Budget”. In: *ApJ* 503.2 (1998), pp. 518–530. DOI: 10.1086/306025. arXiv: astro-ph/9712020 [astro-ph].

[112] Kunihiko Fukushima. “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position”. In: *Biological Cybernetics* 36.4 (1980), pp. 193–202. ISSN: 1432-0770. DOI: 10.1007/BF00344251. URL: https://doi.org/10.1007/BF00344251.

[113] J. M. Gabor and R. Davé. “Hot gas in massive haloes drives both mass quenching and environment quenching”. In: *MNRAS* 447.1 (2015), pp. 374–391. DOI: 10.1093/mnras/stu2399. arXiv: 1405.1043 [astro-ph.GA].

[114] A. Gallazzi et al. “Obscured Star Formation in Intermediate-Density Environments: A Spitzer Study of the Abell 901/902 Supercluster”. In: *ApJ* 690 (Jan. 2009), pp. 1883–1900. DOI: 10.1088/0004-637X/690/2/1883. arXiv: 0809.2042.

[115] F. Galliano et al. “Non-standard grain properties, dark gas reservoir, and extended submillimeter excess, probed by Herschel in the Large Magellanic Cloud”. In: *A&A* 536, A88 (2011), A88. DOI: 10.1051/0004-6361/201117952. arXiv: 1110.1260 [astro-ph.CO].

[116] Frédéric Galliano, Eli Dwek, and Pierre Chanial. “Stellar Evolutionary Effects on the Abundances of Polycyclic Aromatic Hydrocarbons and Supernova-Condensed Dust in Galaxies”. In: *ApJ* 672.1 (2008), pp. 214–243. DOI: 10.1086/523621. arXiv: 0708.0790 [astro-ph].

[117] James E. Geach and John A. Peacock. “Cluster richness-mass calibration with cosmic microwave background lensing”. In: *Nature Astronomy* 1 (2017), pp. 795–799. DOI: 10.1038/s41550-017-0259-1. arXiv: 1707.09369 [astro-ph.CO].

[118] C. Gheller and F. Vazza. “A survey of the thermal and non-thermal properties of cosmic filaments”. In: *MNRAS* 486.1 (2019), pp. 981–1002. DOI: 10.1093/mnras/stz843. arXiv: 1903.08401 [astro-ph.CO].

[119] C. Gheller et al. “Evolution of cosmic filaments and of their galaxy population from MHD cosmological simulations”. In: *MNRAS* 462.1 (2016), pp. 448–463. DOI: 10.1093/mnras/stw1595. arXiv: 1607.01406 [astro-ph.CO].

[120] C. Gheller et al. “Properties of cosmological filaments extracted from Eulerian simulations”. In: *MNRAS* 453.2 (2015), pp. 1164–1185. DOI: 10.1093/mnras/stv1646. arXiv: 1507.05806 [astro-ph.CO].

[121] V. Ghirardini et al. “Universal thermodynamic properties of the intracluster medium over two decades in radius in the X-COP sample”. In: *A&A* 621, A41 (2019), A41. DOI: 10.1051/0004-6361/201833325. arXiv: 1805.00042 [astro-ph.CO].

[122] I. M. Gioia et al. “The Extended Medium-Sensitivity Survey Distant Cluster Sample: X-Ray Cosmological Evolution”. In: *ApJL* 356 (1990), p. L35. DOI: 10.1086/185744.

[123] M. D. Gladders and H. K. C. Yee. “A New Method For Galaxy Cluster Detection. I. The Algorithm”. In: *AJ* 120 (Oct. 2000), pp. 2148–2162. DOI: 10.1086/301557. eprint: astro-ph/0004092.

[124] J. Goodman and J. Weare. “Ensemble samplers with affine invariance”. In: *Communications in Applied Mathematics and Computational Science, Vol. 5, No. 1, p. 65-80, 2010* 5 (2010), pp. 65–80. DOI: 10.2140/camcos.2010.5.65.

[125] K. M. Górski et al. “HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere”. In: *ApJ* 622 (Apr. 2005), pp. 759–771. DOI: 10.1086/427976. eprint: astro-ph/0409513.

[126] F. Govoni et al. “A radio ridge connecting two galaxy clusters in a filament of the cosmic web”. In: *Science* 364.6444 (2019), pp. 981–984. DOI: 10.1126/science.aat7500. arXiv: 1906.07584 [astro-ph.GA].

[127] L. Guennou et al. “The DAFT/FADA survey. I. Photometric redshifts along lines of sight to clusters in the z = [0.4, 0.9] interval”. In: *A&A* 523, A21 (Nov. 2010), A21. DOI: 10.1051/0004-6361/201015174. arXiv: 1008.5318 [astro-ph.CO].

[128] James E. Gunn and III Gott J. Richard. “On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution”. In: *ApJ* 176 (1972), p. 1. DOI: 10.1086/151605.

[129] L. Guzzo et al. “The VIMOS Public Extragalactic Redshift Survey (VIPERS). An unprecedented view of galaxies and large-scale structure at 0.5 < z < 1.2”. In: *A&A* 566, A108 (2014), A108. DOI: 10.1051/0004-6361/201321489. arXiv: 1303.2623 [astro-ph.CO].

[130] M. R. Haas and P. Anders. “Variations in integrated galactic initial mass functions due to sampling method and cluster mass function”. In: *A&A* 512, A79 (2010), A79. DOI: 10.1051/0004-6361/200912967. arXiv: 1001.2009 [astro-ph.GA].

[131] ChangHoon Hahn et al. “PRIMUS: Effects of Galaxy Environment on the Quiescent Fraction Evolution at z < 0.8”. In: *ApJ* 806.2, 162 (2015), p. 162. DOI: 10.1088/0004-637X/806/2/162. arXiv: 1412.7162 [astro-ph.GA].

[132] M. Haider et al. “Large-scale mass distribution in the Illustris simulation”. In: *MNRAS* 457.3 (2016), pp. 3024–3035. DOI: 10.1093/mnras/stw077. arXiv: 1508.01525 [astro-ph.CO].

[133] N. C. Hambly, M. J. Irwin, and H. T. MacGillivray. “The SuperCOSMOS Sky Survey - II. Image detection, parametrization, classification and photometry”. In: *MNRAS* 326 (Oct. 2001), pp. 1295–1314. DOI: 10.1111/j.1365-2966.2001.04661.x. eprint: astro-ph/0108290.

[134] N. C. Hambly et al. “The SuperCOSMOS Sky Survey - I. Introduction and description”. In: *MNRAS* 326 (Oct. 2001), pp. 1279–1294. DOI: 10.1111/j.1365-2966.2001.04660.x. eprint: astro-ph/0108286.

[135] N. C. Hambly et al. “The SuperCOSMOS Sky Survey - III. Astrometry”. In: *MNRAS* 326 (Oct. 2001), pp. 1315–1327. DOI: 10.1111/j.1365-2966.2001.04662.x. eprint: astro-ph/0108291.

[136] Matthew Hasselfield et al. “The Atacama Cosmology Telescope: Sunyaev-Zel’dovich selected galaxy clusters at 148 GHz from three seasons of data”. In: *Journal of Cosmology and Astro-Particle Physics* 2013.7, 008 (2013), p. 008. DOI: 10.1088/1475-7516/2013/07/008. arXiv: 1301.0816 [astro-ph.CO].

[137] W. K. Hastings. “Monte Carlo sampling methods using Markov chains and their applications”. In: *Biometrika* 57.1 (1970), pp. 97–109. DOI: 10.1093/biomet/57.1.97. eprint: http://biomet.oxfordjournals.org/cgi/reprint/57/1/97.pdf. URL: http://biomet.oxfordjournals.org/cgi/content/abstract/57/1/97.

[138] S. He et al. “Learning to Predict the Cosmological Structure Formation”. In: *arXiv e-prints* (Nov. 2018). arXiv: 1811.06533.

[139] Marti A. Hearst. "Support Vector Machines". In: *IEEE Intelligent Systems* 13.4 (July 1998), pp. 18–28. ISSN: 1541-1672. DOI: 10.1109/5254.708428. URL: http://dx.doi.org/10.1109/5254.708428.

[140] J. Patrick Henry et al. "The ROSAT North Ecliptic Pole Survey: The X-Ray Catalog". In: *ApJS* 162.2 (2006), pp. 304–328. DOI: 10.1086/498749. arXiv: astro-ph/0511195 [astro-ph].

[141] D. Herranz et al. "Filtering techniques for the detection of Sunyaev-Zel'dovich clusters <br/> in multifrequency maps". In: *MNRAS* 336.4 (2002), pp. 1057–1068. DOI: 10.1046/j.1365-8711.2002.05704.x. arXiv: astro-ph/0203486 [astro-ph].

[142] Tin Kam Ho. "Random Decision Forests". In: *Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 1) - Volume 1*. ICDAR '95. Washington, DC, USA: IEEE Computer Society, 1995, pp. 278–. ISBN: 0-8186-7128-9. URL: http://dl.acm.org/citation.cfm?id=844379.844681.

[143] Donald J. Horner et al. "The WARPS Survey. VII. The WARPS-II Cluster Catalog". In: *ApJS* 176.2 (2008), pp. 374–413. DOI: 10.1086/529494.

[144] E. P. Hubble. "Extragalactic nebulae." In: *ApJ* 64 (1926), pp. 321–369. DOI: 10.1086/143018.

[145] M. Huertas-Company et al. "A Catalog of Visual-like Morphologies in the 5 CANDELS Fields Using Deep Learning". In: *ApJS* 221, 8 (Nov. 2015), p. 8. DOI: 10.1088/0067-0049/221/1/8. arXiv: 1509.05429.

[146] G. Hurier, N. Aghanim, and M. Douspis. "MILCANN : A neural network assessed tSZ map for galaxy cluster detection". In: *arXiv e-prints*, arXiv:1702.00075 (2017), arXiv:1702.00075. arXiv: 1702.00075 [astro-ph.CO].

[147] G. Hurier and R. E. Angulo. "Measuring the hydrostatic mass bias in galaxy clusters by combining Sunyaev-Zel'dovich and CMB lensing data". In: *A&A* 610, L4 (2018), p. L4. DOI: 10.1051/0004-6361/201731999. arXiv: 1711.06029 [astro-ph.CO].

[148] G. Hurier and F. Lacasa. "Combined analysis of galaxy cluster number count, thermal Sunyaev-Zel'dovich power spectrum, and bispectrum". In: *A&A* 604, A71 (2017), A71. DOI: 10.1051/0004-6361/201630041. arXiv: 1701.09067 [astro-ph.CO].

[149] G. Hurier, J. F. Macías-Pérez, and S. Hildebrandt. "MILCA, a modified internal linear combination algorithm to extract astrophysical emissions from multifrequency sky maps". In: *A&A* 558, A118 (Oct. 2013), A118. DOI: 10.1051/0004-6361/201321891. arXiv: 1007.1149 [astro-ph.IM].

[150] O. Ilbert et al. "Accurate photometric redshifts for the CFHT legacy survey calibrated using the VIMOS VLT deep survey". In: *A&A* 457 (Oct. 2006), pp. 841–856. DOI: 10.1051/0004-6361:20065138. eprint: astro-ph/0603217.

[151] J. C. Jackson. "A critique of Rees's theory of primordial gravitational radiation". In: *MNRAS* 156 (1972), 1P. DOI: 10.1093/mnras/156.1.1P. arXiv: 0810.3908 [astro-ph].

[152] Steven Janowiecki et al. "xGASS: gas-rich central galaxies in small groups and their connections to cosmic web gas feeding". In: *MNRAS* 466.4 (2017), pp. 4795–4812. DOI: 10.1093/mnras/stx046. arXiv: 1701.01754 [astro-ph.GA].

[153] T. H. Jarrett et al. "Extending the Nearby Galaxy Heritage with WISE: First Results from the WISE Enhanced Resolution Galaxy Atlas". In: *AJ* 145.1, 6 (2013), p. 6. DOI: 10.1088/0004-6256/145/1/6. arXiv: 1210.3628 [astro-ph.CO].

[154] Bernard J. T. Jones, Rien van de Weygaert, and Miguel A. Aragón-Calvo. "Fossil evidence for spin alignment of Sloan Digital Sky Survey galaxies in filaments". In: *MNRAS* 408.2 (2010), pp. 897–918. DOI: 10.1111/j.1365-2966.2010.17202.x. arXiv: 1001.4479 [astro-ph.CO].

[155] Peter M. W. Kalberla and Jürgen Kerp. "The Hi Distribution of the Milky Way". In: *ARA&A* 47.1 (2009), pp. 27–61. DOI: 10.1146/annurev-astro-082708-101823.

[156] I. D. Karachentsev and A. I. Kopylov. "On the mass of the X-ray binary cluster A399/A401". In: *MNRAS* 192 (July 1980), pp. 109–112. DOI: 10.1093/mnras/192.2.109.

[157] G. Kauffmann et al. "Stellar masses and star formation histories for 10^5 galaxies from the Sloan Digital Sky Survey". In: *MNRAS* 341 (May 2003), pp. 33–53. DOI: 10.1046/j.1365-8711.2003.06291.x. eprint: astro-ph/0204055.

[158] Andreas Kelz et al. "Development and performance of the MUSE calibration unit". In: *Ground-based and Airborne Instrumentation for Astronomy IV*. Vol. 8446. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. 2012, 84465T. DOI: 10.1117/12.926003.

[159] Jr. Kennicutt Robert C. "Star Formation in Galaxies Along the Hubble Sequence". In: *ARA&A* 36 (1998), pp. 189–232. DOI: 10.1146/annurev.astro.36.1.189. arXiv: astro-ph/9807187 [astro-ph].

[160] Jr. Kennicutt Robert C. et al. "Dust-corrected Star Formation Rates of Galaxies. I. Combinations of Hα and Infrared Tracers". In: *ApJ* 703.2 (2009), pp. 1672–1695. DOI: 10.1088/0004-637X/703/2/1672. arXiv: 0908.0203 [astro-ph.CO].

[161] Jr. Kennicutt Robert C. et al. "SINGS: The SIRTF Nearby Galaxies Survey". In: *PASP* 115.810 (2003), pp. 928–952. DOI: 10.1086/376941. arXiv: astro-ph/0305437 [astro-ph].

[162] R. C. Kennicutt et al. "KINGFISH—Key Insights on Nearby Galaxies: A Far-Infrared Survey with Herschel: Survey Description and Image Atlas". In: *PASP* 123.910 (2011), p. 1347. DOI: 10.1086/663818. arXiv: 1111.4438 [astro-ph.CO].

[163] Robert C. Kennicutt and Neal J. Evans. "Star Formation in the Milky Way and Nearby Galaxies". In: *ARA&A* 50 (2012), pp. 531–608. DOI: 10.1146/annurev-astro-081811-125610. arXiv: 1204.3552 [astro-ph.GA].

[164] Diederik P. Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: *arXiv e-prints*, arXiv:1412.6980 (2014), arXiv:1412.6980. arXiv: 1412.6980 [cs.LG].

[165] Tetsu Kitayama et al. "The Sunyaev-Zel'dovich effect at 5": RX J1347.5-1145 imaged by ALMA". In: *PASJ* 68.5, 88 (2016), p. 88. DOI: 10.1093/pasj/psw082. arXiv: 1607.08833 [astro-ph.CO].

[166] T. Krakowski et al. "Machine-learning identification of galaxies in the WISExSuperCOSMOS all-sky catalogue". In: *A&A* 596, A39 (Nov. 2016), A39. DOI: 10.1051/0004-6361/201629165. arXiv: 1607.01188.

[167] K. Kraljic et al. "Galaxies flowing in the oriented saddle frame of the cosmic web". In: *MNRAS* 483.3 (2019), pp. 3227–3254. DOI: 10.1093/mnras/sty3216. arXiv: 1810.05211 [astro-ph.GA].

[168] K. Kraljic et al. "Galaxy evolution in the metric of the cosmic web". In: *MNRAS* 474.1 (2018), pp. 547–571. DOI: 10.1093/mnras/stx2638. arXiv: 1710.02676 [astro-ph.GA].

[169] Katarina Kraljic, Romeel Dave, and Christophe Pichon. "And yet it flips: connecting galactic spin and the cosmic web". In: *arXiv e-prints*, arXiv:1906.01623 (2019), arXiv:1906.01623. arXiv: 1906.01623 [astro-ph.GA].

[170] Andrey V. Kravtsov and Stefano Borgani. "Formation of Galaxy Clusters". In: *ARA&A* 50 (2012), pp. 353–409. DOI: 10.1146/annurev-astro-081811-125502. arXiv: 1205.5556 [astro-ph.CO].

[171] P. Kroupa. "On the variation of the initial mass function". In: *MNRAS* 322 (Apr. 2001), pp. 231–246. DOI: 10.1046/j.1365-8711.2001.04022.x. eprint: astro-ph/0009005.

[172] Egidijus Kukstas et al. "Environment from cross-correlations: connecting hot gas and the quenching of galaxies". In: *arXiv e-prints*, arXiv:1907.06645 (2019), arXiv:1907.06645. arXiv: 1907.06645 [astro-ph.GA].

[173] Teet Kuutma, Antti Tamm, and Elmo Tempel. "From voids to filaments: environmental transformations of galaxies in the SDSS". In: *A&A* 600, L6 (2017), p. L6. DOI: 10.1051/0004-6361/201730526. arXiv: 1703.04338 [astro-ph.GA].

[174] Guilaine Lagache, Jean-Loup Puget, and Hervé Dole. "Dusty Infrared Galaxies: Sources of the Cosmic Infrared Background". In: *ARA&A* 43.1 (2005), pp. 727–768. DOI: 10.1146/annurev.astro.43.072103.150606. arXiv: astro-ph/0507298 [astro-ph].

[175] C. Laigle et al. "COSMOS2015 photometric redshifts probe the impact of filaments on galaxy properties". In: *MNRAS* 474.4 (2018), pp. 5437–5458. DOI: 10.1093/mnras/stx3055. arXiv: 1702.08810 [astro-ph.GA].

[176] K. Lakhchaura et al. "Intracluster Medium of the Merging Cluster A3395". In: *ApJ* 743, 78 (Dec. 2011), p. 78. DOI: 10.1088/0004-637X/743/1/78. arXiv: 1109.5688.

[177] J.-M. Lamarre et al. "Planck pre-launch status: The HFI instrument, from specification to actual performance". In: *A&A* 520, A9 (Sept. 2010), A9. DOI: 10.1051/0004-6361/200912975.

[178] Dustin Lang and David W. Hogg. "Searching for Comets on the World Wide Web: The Orbit of 17P/Holmes from the Behavior of Photographers". In: *AJ* 144.2, 46 (2012), p. 46. DOI: 10.1088/0004-6256/144/2/46. arXiv: 1103.6038 [astro-ph.IM].

[179] A. Leger and J. L. Puget. "Identification of the "unidentified" IR emission features of interstellar dust ?" In: *A&A* 500 (1984), pp. 279–282.

[180] Jianhui Lian et al. "The Quenching Timescale and Quenching Rate of Galaxies". In: *ApJ* 832.1, 29 (2016), p. 29. DOI: 10.3847/0004-637X/832/1/29. arXiv: 1609.04805 [astro-ph.GA].

[181] Noam I. Libeskind et al. "Tracing the cosmic web". In: *MNRAS* 473.1 (2018), pp. 1195–1217. DOI: 10.1093/mnras/stx1976. arXiv: 1705.03021 [astro-ph.CO].

[182] L. J. Liivamägi, E. Tempel, and E. Saar. "SDSS DR7 superclusters. The catalogues". In: *A&A* 539, A80 (2012), A80. DOI: 10.1051/0004-6361/201016288. arXiv: 1012.1989 [astro-ph.CO].

[183] L. Lucie-Smith et al. "Machine learning cosmological structure formation". In: *MNRAS* 479 (Sept. 2018), pp. 3405–3414. DOI: 10.1093/mnras/sty1719. arXiv: 1802.04271.

[184] N. Malavasi et al. "The VIMOS Public Extragalactic Redshift Survey (VIPERS): galaxy segregation inside filaments at z = 0.7". In: *MNRAS* 465.4 (2017), pp. 3817–3822. DOI: 10.1093/mnras/stw2864. arXiv: 1611.07045 [astro-ph.GA].

[185] Rachel Mandelbaum. "Weak Lensing for Precision Cosmology". In: *ARA&A* 56 (2018), pp. 393–433. DOI: 10.1146/annurev-astro-081817-051928. arXiv: 1710.03235 [astro-ph.CO].

[186] A. Mantz et al. "The observed growth of massive galaxy clusters - I. Statistical methods and cosmological constraints". In: *MNRAS* 406.3 (2010), pp. 1759–1772. DOI: 10.1111/j.1365-2966.2010.16992.x. arXiv: 0909.3098 [astro-ph.CO].

[187] C. Maraston et al. "Stellar masses of SDSS-III BOSS galaxies at z=0.5 and constraints to galaxy formation models". In: *MNRAS* 435 (Nov. 2013), pp. 2764–2792. DOI: 10.1093/mnras/stt1424. arXiv: 1207.6114.

[188] Tobias A. Marriage et al. "The Atacama Cosmology Telescope: Sunyaev-Zel'dovich-Selected Galaxy Clusters at 148 GHz in the 2008 Survey". In: *ApJ* 737.2, 61 (2011), p. 61. DOI: 10.1088/0004-637X/737/2/61. arXiv: 1010.1065 [astro-ph.CO].

[189] H. J. Martínez, H. Muriel, and V. Coenda. "Galaxies infalling into groups: filaments versus isotropic infall". In: *MNRAS* 455 (Jan. 2016), pp. 127–135. DOI: 10.1093/mnras/stv2295. arXiv: 1510.00390.

[190] Nicholas S. Martis et al. "The Evolution of the Fractions of Quiescent and Star-forming Galaxies as a Function of Stellar Mass Since z = 3: Increasing Importance of Massive, Dusty Star-forming Galaxies in the Early Universe". In: *ApJ* 827.2, L25 (2016), p. L25. DOI: 10.3847/2041-8205/827/2/L25. arXiv: 1606.04090 [astro-ph.GA].

[191] Davide Martizzi et al. "Baryons in the Cosmic Web of IllustrisTNG - I: gas in knots, filaments, sheets, and voids". In: *MNRAS* 486.3 (2019), pp. 3766–3787. DOI: 10.1093/mnras/stz1106. arXiv: 1810.01883 [astro-ph.CO].

[192] J. C. Mather et al. "A Preliminary Measurement of the Cosmic Microwave Background Spectrum by the Cosmic Background Explorer (COBE) Satellite". In: *ApJL* 354 (1990), p. L37. DOI: 10.1086/185717.

[193] Ian G. McCarthy et al. "The BAHAMAS project: calibrated hydrodynamical simulations for large-scale structure cosmology". In: *MNRAS* 465.3 (2017), pp. 2936–2965. DOI: 10.1093/mnras/stw2792. arXiv: 1603.02702 [astro-ph.CO].

[194] J. B. Melin, J. G. Bartlett, and J. Delabrouille. "Catalog extraction in SZ cluster surveys: a matched filter approach". In: *A&A* 459.2 (2006), pp. 341–352. DOI: 10.1051/0004-6361:20065034. arXiv: astro-ph/0602424 [astro-ph].

[195] A. Mennella et al. "Planck early results. III. First assessment of the Low Frequency Instrument in-flight performance". In: *A&A* 536, A3 (Dec. 2011), A3. DOI: 10.1051/0004-6361/201116480. arXiv: 1101.2038.

[196] Nicholas Metropolis et al. "Equation of State Calculations by Fast Computing Machines". In: *The Journal of Chemical Physics* 21.6 (1953), pp. 1087–1092. DOI: 10.1063/1.1699114. URL: http://link.aip.org/link/?JCP/21/1087/1.

[197] Bruno Milliard et al. "GALEX: a UV telescope to map the star formation history of the universe". In: *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. Vol. 10569. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. 2017, p. 1056912. DOI: 10.1117/12.2307874.

[198] Marc-Antoine Miville-Deschênes and Guilaine Lagache. "IRIS: A New Generation of IRAS Maps". In: *ApJS* 157.2 (2005), pp. 302–323. DOI: 10.1086/427938. arXiv: astro-ph/0412216 [astro-ph].

[199] A. D. Montero-Dorta et al. "The high-mass end of the red sequence at z=0.55 from SDSS-III BOSS: completeness, bimodality and luminosity function". In: *MNRAS* 461 (Sept. 2016), pp. 1131–1153. DOI: 10.1093/mnras/stw1352. arXiv: 1410.5854.

[200] Ben Moore et al. "Galaxy harassment and the evolution of clusters of galaxies". In: *Nature* 379.6566 (1996), pp. 613–616. DOI: 10.1038/379613a0. arXiv: astro-ph/9510034 [astro-ph].

[201] Thibaud Moutard et al. "On the fast quenching of young low-mass galaxies up to z=0.6: new spotlight on the lead role of environment". In: *MNRAS* 479.2 (2018), pp. 2147–2160. DOI: 10.1093/mnras/sty1543. arXiv: 1802.07628 [astro-ph.GA].

[202] Tony Mroczkowski et al. “Astrophysics with the Spatially and Spectrally Resolved Sunyaev-Zeldovich Effects. A Millimetre/Submillimetre Probe of the Warm and Hot Universe”. In: *SSR* 215.1, 17 (2019), p. 17. DOI: 10.1007/s11214-019-0581-2. arXiv: 1811.02310 [astro-ph.CO].

[203] C. R. Mullis et al. “The 160 Square Degree ROSAT Survey: The Revised Catalog of 201 Clusters with Spectroscopic Redshifts”. In: *ApJ* 594.1 (2003), pp. 154–171. DOI: 10.1086/376866. arXiv: astro-ph/0305228 [astro-ph].

[204] M. Murgia et al. “A double radio halo in the close pair of galaxy clusters Abell 399 and Abell 401”. In: *A&A* 509, A86 (Jan. 2010), A86. DOI: 10.1051/0004-6361/200913414. arXiv: 0911.3594.

[205] D. Nagai, A. Vikhlinin, and A. V. Kravtsov. “Testing X-Ray Measurements of Galaxy Clusters with Cosmological Simulations”. In: *ApJ* 655 (Jan. 2007), pp. 98–108. DOI: 10.1086/509868. eprint: astro-ph/0609247.

[206] A. Nastasi et al. “F-VIPGI: a new adapted version of VIPGI for FORS2 spectroscopy. Application to a sample of 16 X-ray selected galaxy clusters at $0.6 \leq z \leq 1.2$”. In: *A&A* 550, A9 (2013), A9. DOI: 10.1051/0004-6361/201219862. arXiv: 1301.2327 [astro-ph.IM].

[207] J. F. Navarro, C. S. Frenk, and S. D. M. White. “The assembly of galaxies in a hierarchically clustering universe”. In: *MNRAS* 275 (July 1995), pp. 56–66. DOI: 10.1093/mnras/275.1.56. eprint: astro-ph/9408067.

[208] Julio F. Navarro, Carlos S. Frenk, and Simon D. M. White. “A Universal Density Profile from Hierarchical Clustering”. In: *ApJ* 490.2 (1997), pp. 493–508. DOI: 10.1086/304888. arXiv: astro-ph/9611107 [astro-ph].

[209] Julio F. Navarro, Carlos S. Frenk, and Simon D. M. White. “The Structure of Cold Dark Matter Halos”. In: *ApJ* 462 (1996), p. 563. DOI: 10.1086/177173. arXiv: astro-ph/9508025 [astro-ph].

[210] G. Neugebauer et al. “The Infrared Astronomical Satellite (IRAS) mission.” In: *ApJL* 278 (1984), pp. L1–L6. DOI: 10.1086/184209.

[211] F. Nicastro et al. “Observations of the missing baryons in the warm-hot intergalactic medium”. In: *Nature* 558.7710 (2018), pp. 406–409. DOI: 10.1038/s41586-018-0204-1. arXiv: 1806.08395 [astro-ph.GA].

[212] W. R. Oegerle and J. M. Hill. “Dynamics of cD Clusters of Galaxies. IV. Conclusion of a Survey of 25 Abell Clusters”. In: *AJ* 122 (Dec. 2001), pp. 2858–2873. DOI: 10.1086/323536.

[213] J. Olofsson et al. “Transient dust in warm debris disks. Detection of Fe-rich olivine grains”. In: *A&A* 542, A90 (2012), A90. DOI: 10.1051/0004-6361/201118735. arXiv: 1204.2374 [astro-ph.SR].

[214] J. Ostriker. “The Equilibrium of Polytropic and Isothermal Cylinders.” In: *ApJ* 140 (Oct. 1964), p. 1056. DOI: 10.1086/148005.

[215] I. N. Pashchenko, K. V. Sokolovsky, and P. Gavras. “Machine learning search for variable stars”. In: *MNRAS* 475 (Apr. 2018), pp. 2326–2343. DOI: 10.1093/mnras/stx3222. arXiv: 1710.07290 [astro-ph.IM].

[216] J. Pasquet et al. “Photometric redshifts from SDSS images using a convolutional neural network”. In: *A&A* 621, A26 (Jan. 2019), A26. DOI: 10.1051/0004-6361/201833617. arXiv: 1806.06607 [astro-ph.IM].

[217] Y. Peng, R. Maiolino, and R. Cochrane. "Strangulation as the primary mechanism for shutting down star formation in galaxies". In: *Nature* 521.7551 (2015), pp. 192–195. DOI: 10.1038/nature14439. arXiv: 1505.03143 [astro-ph.GA].

[218] Ying-jie Peng et al. "Mass and Environment as Drivers of Galaxy Evolution in SDSS and zCOSMOS and the Origin of the Schechter Function". In: *ApJ* 721.1 (2010), pp. 193–221. DOI: 10.1088/0004-637X/721/1/193. arXiv: 1003.4747 [astro-ph.CO].

[219] Eric S. Perlman et al. "The WARPS Survey. VI. Galaxy Cluster and Source Identifications from Phase I". In: *ApJS* 140.2 (2002), pp. 265–301. DOI: 10.1086/339685. arXiv: astro-ph/0112190 [astro-ph].

[220] R. Piffaretti et al. "The MCXC: a meta-catalogue of x-ray detected clusters of galaxies". In: *A&A* 534, A109 (Oct. 2011), A109. DOI: 10.1051/0004-6361/201015377. arXiv: 1007.1916.

[221] I. Pintos-Castro et al. "The Evolution of the Quenching of Star Formation in Cluster Galaxies since z $\sim$ 1". In: *ApJ* 876.1, 40 (2019), p. 40. DOI: 10.3847/1538-4357/ab14ee. arXiv: 1904.00023 [astro-ph.GA].

[222] Laure Piqueras et al. "MPDAF - A Python package for the analysis of VLT/MUSE data". In: *arXiv e-prints*, arXiv:1710.03554 (2017), arXiv:1710.03554. arXiv: 1710.03554 [astro-ph.IM].

[223] Planck Collaboration et al. "Planck 2013 results. XXIX. The Planck catalogue of Sunyaev-Zeldovich sources". In: *A&A* 571, A29 (2014), A29. DOI: 10.1051/0004-6361/201321523. arXiv: 1303.5089 [astro-ph.CO].

[224] Planck Collaboration et al. "Planck 2013 results. XXXII. The updated Planck catalogue of Sunyaev-Zeldovich sources". In: *A&A* 581, A14 (2015), A14. DOI: 10.1051/0004-6361/201525787. arXiv: 1502.00543 [astro-ph.CO].

[225] Planck Collaboration et al. "Planck 2015 results. I. Overview of products and scientific results". In: *A&A* 594, A1 (2016), A1. DOI: 10.1051/0004-6361/201527101. arXiv: 1502.01582 [astro-ph.CO].

[226] Planck Collaboration et al. "Planck 2015 results. XIII. Cosmological parameters". In: *A&A* 594, A13 (2016), A13. DOI: 10.1051/0004-6361/201525830. arXiv: 1502.01589 [astro-ph.CO].

[227] Planck Collaboration et al. "Planck 2015 results. XXII. A map of the thermal Sunyaev-Zeldovich effect". In: *A&A* 594, A22 (Sept. 2016), A22. DOI: 10.1051/0004-6361/201525826. arXiv: 1502.01596.

[228] Planck Collaboration et al. "Planck 2015 results. XXVI. The Second Planck Catalogue of Compact Sources". In: *A&A* 594, A26 (2016), A26. DOI: 10.1051/0004-6361/201526914. arXiv: 1507.02058 [astro-ph.CO].

[229] Planck Collaboration et al. "Planck 2015 results. XXVII. The second Planck catalogue of Sunyaev-Zeldovich sources". In: *A&A* 594, A27 (Sept. 2016), A27. DOI: 10.1051/0004-6361/201525823. arXiv: 1502.01598.

[230] Planck Collaboration et al. "Planck 2018 results. VIII. Gravitational lensing". In: *arXiv e-prints*, arXiv:1807.06210 (2018), arXiv:1807.06210. arXiv: 1807.06210 [astro-ph.CO].

[231] Planck Collaboration et al. "Planck early results. VIII. The all-sky early Sunyaev-Zeldovich cluster sample". In: *A&A* 536, A8 (2011), A8. DOI: 10.1051/0004-6361/201116459. arXiv: 1101.2024 [astro-ph.CO].

[232] Planck Collaboration et al. "Planck intermediate results. LIII. Detection of velocity dispersion from the kinetic Sunyaev-Zeldovich effect". In: *A&A* 617, A48 (2018), A48. DOI: 10.1051/0004-6361/201731489. arXiv: 1707.00132 [astro-ph.CO].

[233] Planck Collaboration et al. “Planck intermediate results. V. Pressure profiles of galaxy clusters from the Sunyaev-Zeldovich effect”. In: *A&A* 550, A131 (2013), A131. DOI: 10.1051/0004-6361/201220040. arXiv: 1207.4061 [astro-ph.CO].

[234] Planck Collaboration et al. “Planck intermediate results. VIII. Filaments between interacting clusters”. In: *A&A* 550, A134 (2013), A134. DOI: 10.1051/0004-6361/201220194. arXiv: 1208.5911 [astro-ph.CO].

[235] Planck Collaboration et al. “Planck intermediate results. X. Physics of the hot gas in the Coma cluster”. In: *A&A* 554, A140 (2013), A140. DOI: 10.1051/0004-6361/201220247. arXiv: 1208.3611 [astro-ph.CO].

[236] Planck Collaboration et al. “Planck intermediate results. XXVI. Optical identification and redshifts of Planck clusters with the RTT150 telescope”. In: *A&A* 582, A29 (2015), A29. DOI: 10.1051/0004-6361/201424674. arXiv: 1407.6663 [astro-ph.CO].

[237] Planck Collaboration et al. “Planck intermediate results. XXXVII. Evidence of unbound gas from the kinetic Sunyaev-Zeldovich effect”. In: *A&A* 586, A140 (2016), A140. DOI: 10.1051/0004-6361/201526328. arXiv: 1504.03339 [astro-ph.CO].

[238] Planck HFI Core Team et al. “Planck early results. IV. First assessment of the High Frequency Instrument in-flight performance”. In: *A&A* 536, A4 (Dec. 2011), A4. DOI: 10.1051/0004-6361/201116487. arXiv: 1101.2039 [astro-ph.IM].

[239] Mathieu Remazeilles, Jacques Delabrouille, and Jean-François Cardoso. “CMB and SZ effect separation with constrained Internal Linear Combinations”. In: *MNRAS* 410.4 (2011), pp. 2481–2487. DOI: 10.1111/j.1365-2966.2010.17624.x. arXiv: 1006.5599 [astro-ph.CO].

[240] J. Richard et al. “MUSE observations of the lensing cluster SMACSJ2031.8-4036: new constraints on the mass distribution in the cluster core.” In: *MNRAS* 446 (2015), pp. L16–L20. DOI: 10.1093/mnrasl/slu150. arXiv: 1409.2488 [astro-ph.CO].

[241] T. Emil Rivera-Thorsen et al. “Hubble captures multiply-imaged ionizing radiation from strongly lensed galaxy at z=2.4”. In: *arXiv e-prints*, arXiv:1904.08186 (2019), arXiv:1904.08186. arXiv: 1904.08186 [astro-ph.GA].

[242] A. C. Rodríguez et al. “Fast cosmic web simulations with generative adversarial networks”. In: *Computational Astrophysics and Cosmology* 5, 4 (Nov. 2018), p. 4. DOI: 10.1186/s40668-018-0026-4. arXiv: 1801.09070.

[243] A. K. Romer et al. “The Bright SHARC Survey: The Cluster Catalog”. In: *ApJS* 126.2 (2000), pp. 209–269. DOI: 10.1086/313302. arXiv: astro-ph/9907401 [astro-ph].

[244] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *arXiv e-prints*, arXiv:1505.04597 (2015), arXiv:1505.04597. arXiv: 1505.04597 [cs.CV].

[245] S. R. Rosen et al. “The XMM-Newton serendipitous survey. VII. The third XMM-Newton serendipitous source catalogue”. In: *A&A* 590, A1 (2016), A1. DOI: 10.1051/0004-6361/201526416. arXiv: 1504.07051 [astro-ph.HE].

[246] E. S. Rykoff et al. “redMaPPer. I. Algorithm and SDSS DR8 Catalog”. In: *ApJ* 785, 104 (Apr. 2014), p. 104. DOI: 10.1088/0004-637X/785/2/104. arXiv: 1303.3562.

[247] Amélie Saintonge et al. “xCOLD GASS: The Complete IRAM 30 m Legacy Survey of Molecular Gas for Galaxy Evolution Studies”. In: *ApJS* 233.2, 22 (2017), p. 22. DOI: 10.3847/1538-4365/aa97e0. arXiv: 1710.02157 [astro-ph.GA].

[248] Irini Sakelliou and Trevor J. Ponman. "XMM-Newton observations of the binary cluster system Abell 399/401". In: *MNRAS* 351.4 (2004), pp. 1439–1456. DOI: 10.1111/j.1365-2966.2004.07889.x. arXiv: astro-ph/0403673 [astro-ph].

[249] Samir Salim et al. "UV Star Formation Rates in the Local Universe". In: *ApJS* 173.2 (2007), pp. 267–292. DOI: 10.1086/519218. arXiv: 0704.3611 [astro-ph].

[250] Edwin E. Salpeter. "The Luminosity Function and Stellar Evolution." In: *ApJ* 121 (1955), p. 161. DOI: 10.1086/145971.

[251] Laura Salvati, Marian Douspis, and Nabila Aghanim. "Constraints from thermal Sunyaev-Zel'dovich cluster counts and power spectrum combined with CMB". In: *A&A* 614, A13 (2018), A13. DOI: 10.1051/0004-6361/201731990. arXiv: 1708.00697 [astro-ph.CO].

[252] A. Saro et al. "Constraints on the richness-mass relation and the optical-SZE positional offset distribution for SZE-selected clusters". In: *MNRAS* 454.3 (2015), pp. 2305–2319. DOI: 10.1093/mnras/stv2141. arXiv: 1506.07814 [astro-ph.CO].

[253] Florian Sarron et al. "Pre-processing of galaxies in cosmic filaments around AMASCFI clusters in the CFHTLS". In: *arXiv e-prints*, arXiv:1903.02879 (2019), arXiv:1903.02879. arXiv: 1903.02879 [astro-ph.GA].

[254] B. Sartoris et al. "Next generation cosmology: constraints from the Euclid galaxy cluster survey". In: *MNRAS* 459.2 (2016), pp. 1764–1780. DOI: 10.1093/mnras/stw630. arXiv: 1505.02165 [astro-ph.CO].

[255] W. E. Schaap and R. van de Weygaert. "Continuous fields and discrete samples: reconstruction through Delaunay tessellations". In: *A&A* 363 (Nov. 2000), pp. L29–L32. eprint: astro-ph/0011007.

[256] Michel P. Schammel et al. "AMI SZ observations and Bayesian analysis of a sample of six redshift-one clusters of galaxies". In: *MNRAS* 431.1 (2013), pp. 900–911. DOI: 10.1093/mnras/stt212. arXiv: 1210.7771 [astro-ph.CO].

[257] N. Scoville et al. "The Cosmic Evolution Survey (COSMOS): Overview". In: *ApJS* 172.1 (2007), pp. 1–8. DOI: 10.1086/516585. arXiv: astro-ph/0612305 [astro-ph].

[258] Carl K. Seyfert. "Nuclear Emission in Spiral Nebulae." In: *ApJ* 97 (1943), p. 28. DOI: 10.1086/144488.

[259] J. Michael Shull, Britton D. Smith, and Charles W. Danforth. "The Baryon Census in a Multiphase Intergalactic Medium: 30% of the Baryons May Still be Missing". In: *ApJ* 759.1, 23 (2012), p. 23. DOI: 10.1088/0004-637X/759/1/23. arXiv: 1112.2706 [astro-ph.CO].

[260] M. Siudek et al. "The VIMOS Public Extragalactic Redshift Survey (VIPERS). The complexity of galaxy populations at $0.4 < z < 1.3$ revealed with unsupervised machine-learning algorithms". In: *A&A* 617, A70 (Sept. 2018), A70. DOI: 10.1051/0004-6361/201832784. arXiv: 1805.09904.

[261] M. F. Skrutskie et al. "The Two Micron All Sky Survey (2MASS)". In: *Aj* 131 (Feb. 2006), pp. 1163–1183. DOI: 10.1086/498708.

[262] Ian Smail et al. "A Very Bright, Highly Magnified Lyman Break Galaxy at z = 3.07". In: *ApJ* 654.1 (2007), pp. L33–L36. DOI: 10.1086/510902. arXiv: astro-ph/0611486 [astro-ph].

[263] Kurt T. Soto et al. "ZAP - enhanced PCA sky subtraction for integral field spectroscopy". In: *MNRAS* 458.3 (2016), pp. 3210–3220. DOI: 10.1093/mnras/stw474. arXiv: 1602.08037 [astro-ph.IM].

[264] T. Sousbie. "The persistent cosmic web and its filamentary structure - I. Theory and implementation". In: *MNRAS* 414 (June 2011), pp. 350–383. DOI: 10.1111/j.1365-2966.2011.18394.x. arXiv: 1009.4015.

[265] Volker Springel. "The cosmological simulation code GADGET-2". In: *MNRAS* 364.4 (2005), pp. 1105–1134. DOI: 10.1111/j.1365-2966.2005.09655.x. arXiv: astro-ph/0505010 [astro-ph].

[266] Volker Springel et al. "First results from the IllustrisTNG simulations: matter and galaxy clustering". In: *MNRAS* 475.1 (2018), pp. 676–698. DOI: 10.1093/mnras/stx3304. arXiv: 1707.03397 [astro-ph.GA].

[267] Volker Springel et al. "Simulations of the formation, evolution and clustering of galaxies and quasars". In: *Nature* 435.7042 (2005), pp. 629–636. DOI: 10.1038/nature03597. arXiv: astro-ph/0504097 [astro-ph].

[268] Nitish Srivastava et al. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *Journal of Machine Learning Research* 15 (2014), pp. 1929–1958. URL: http://jmlr.org/papers/v15/srivastava14a.html.

[269] A. Streblyanska et al. "Characterization of a subsample of the Planck SZ source cluster catalogues using optical SDSS DR12 data". In: *A&A* 617, A71 (2018), A71. DOI: 10.1051/0004-6361/201732306. arXiv: 1804.01356 [astro-ph.CO].

[270] A. Streblyanska et al. "Optical validation and characterization of Planck PSZ2 sources at the Canary Islands observatories. I. First year of LP15 observations". In: *arXiv e-prints*, arXiv:1905.13661 (2019), arXiv:1905.13661. arXiv: 1905.13661 [astro-ph.CO].

[271] R. A. Sunyaev and Ya. B. Zeldovich. "The interaction of matter and radiation in the hot model of the Universe, II". In: *ApSS* 7.1 (1970), pp. 20–30. DOI: 10.1007/BF00653472.

[272] R. A. Sunyaev and Ya. B. Zeldovich. "The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies". In: *Comments on Astrophysics and Space Physics* 4 (1972), p. 173.

[273] H. Tanimura et al. "Detection of intercluster gas in superclusters using the thermal Sunyaev-Zel'dovich effect". In: *A&A* 625, A67 (2019), A67. DOI: 10.1051/0004-6361/201833413. arXiv: 1805.04555 [astro-ph.CO].

[274] Hideki Tanimura et al. "A search for warm/hot gas filaments between pairs of SDSS Luminous Red Galaxies". In: *MNRAS* 483.1 (2019), pp. 223–234. DOI: 10.1093/mnras/sty3118. arXiv: 1709.05024 [astro-ph.CO].

[275] P. Tarrío, J. B. Melin, and M. Arnaud. "ComPRASS: a Combined Planck-RASS catalogue of X-ray-SZ clusters". In: *A&A* 626, A7 (2019), A7. DOI: 10.1051/0004-6361/201834979. arXiv: 1901.00873 [astro-ph.CO].

[276] J. A. Tauber et al. "Planck pre-launch status: The Planck mission". In: *A&A* 520, A1 (2010), A1. DOI: 10.1051/0004-6361/200912983.

[277] E. Tempel et al. "Bisous model-Detecting filamentary patterns in point processes". In: *Astronomy and Computing* 16 (2016), pp. 17–25. DOI: 10.1016/j.ascom.2016.03.004. arXiv: 1603.08957 [astro-ph.CO].

[278] The Dark Energy Survey Collaboration. "The Dark Energy Survey". In: *arXiv e-prints*, astro-ph/0510346 (2005), astro–ph/0510346. arXiv: astro-ph/0510346 [astro-ph].

[279] C. A. Tremonti et al. “The Origin of the Mass-Metallicity Relation: Insights from 53,000 Star-forming Galaxies in the Sloan Digital Sky Survey”. In: *ApJ* 613 (Oct. 2004), pp. 898–913. DOI: 10.1086/423264. eprint: astro-ph/0405537.

[280] J. Truemper. “The ROSAT mission”. In: *Advances in Space Research* 2.4 (1982), pp. 241–249. DOI: 10.1016/0273-1177(82)90070-9.

[281] Robert J. Trumpler. “Absorption of Light in the Galactic System”. In: *PASP* 42.248 (1930), p. 214. DOI: 10.1086/124039.

[282] James Trussler et al. “Starvation as the primary quenching mechanism in galaxies”. In: *arXiv e-prints*, arXiv:1811.09283 (2018), arXiv:1811.09283. arXiv: 1811.09283 [astro-ph.GA].

[283] D. Tuccillo et al. “Deep learning for galaxy surface brightness profile fitting”. In: *MNRAS* 475 (Mar. 2018), pp. 894–909. DOI: 10.1093/mnras/stx3186. arXiv: 1711.03108.

[284] G. Ucci et al. “GAME: GAlaxy Machine learning for Emission lines”. In: *MNRAS* 477 (June 2018), pp. 1484–1494. DOI: 10.1093/mnras/sty804. arXiv: 1803.10236.

[285] M. P. Ulmer and R. G. Cruddace. “Einstein observations of the region between the Abell clusters of galaxies A 401 and A 399.” In: *ApJ* 246 (1981), pp. L99–L103. DOI: 10.1086/183563.

[286] R. F. J. van der Burg et al. “Prospects for high-z cluster detections with Planck, based on a follow-up of 28 candidates using MegaCam at CFHT”. In: *A&A* 587, A23 (2016), A23. DOI: 10.1051/0004-6361/201527299. arXiv: 1512.00852 [astro-ph.CO].

[287] B W. White and Frank Rosenblatt. “Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms”. In: *The American Journal of Psychology* 76 (Dec. 1963), p. 705. DOI: 10.2307/1419730.

[288] Stephen Walker et al. “The Physics of Galaxy Cluster Outskirts”. In: *SSR* 215.1, 7 (2019), p. 7. DOI: 10.1007/s11214-018-0572-8. arXiv: 1810.00890 [astro-ph.CO].

[289] P. M. Weilbacher et al. “The MUSE Data Reduction Pipeline: Status after Preliminary Acceptance Europe”. In: *Astronomical Data Analysis Software and Systems XXIII*. Ed. by N. Manset and P. Forshay. Vol. 485. Astronomical Society of the Pacific Conference Series. 2014, p. 451. arXiv: 1507.00034 [astro-ph.IM].

[290] Peter M. Weilbacher et al. “Design and capabilities of the MUSE data reduction software and pipeline”. In: *Software and Cyberinfrastructure for Astronomy II*. Vol. 8451. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. 2012, 84510B. DOI: 10.1117/12.925114.

[291] Xiao-Qing Wen et al. “The stellar masses of galaxies from the 3.4 μm band of the WISE All-Sky Survey”. In: *MNRAS* 433.4 (2013), pp. 2946–2957. DOI: 10.1093/mnras/stt939.

[292] Z. L. Wen and J. L. Han. “Calibration of the Optical Mass Proxy for Clusters of Galaxies and an Update of the WHL12 Cluster Catalog”. In: *ApJ* 807.2, 178 (2015), p. 178. DOI: 10.1088/0004-637X/807/2/178. arXiv: 1506.04503 [astro-ph.GA].

[293] Z. L. Wen, J. L. Han, and F. S. Liu. “A Catalog of 132,684 Clusters of Galaxies Identified from Sloan Digital Sky Survey III”. In: *ApJS* 199.2, 34 (2012), p. 34. DOI: 10.1088/0067-0049/199/2/34. arXiv: 1202.6424 [astro-ph.CO].

[294] M. W. Werner et al. “The Spitzer Space Telescope Mission”. In: *ApJS* 154.1 (2004), pp. 1–9. DOI: 10.1086/422992. arXiv: astro-ph/0406223 [astro-ph].

[295] E. L. Wright et al. “The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance”. In: *Aj* 140 (Dec. 2010), pp. 1868–1881. DOI: 10.1088/0004-6256/140/6/1868. arXiv: 1008.0031 [astro-ph.IM].

[296] Eva Wuyts et al. “A Bright, Spatially Extended Lensed Galaxy at z = 1.7 Behind the Cluster RCS2 032727-132623”. In: *ApJ* 724.2 (2010), pp. 1182–1192. DOI: 10.1088/0004-637X/724/2/1182. arXiv: 1005.2621 [astro-ph.CO].

[297] Xiaohu Yang et al. “The two-point correlation of galaxy groups: probing the clustering of dark matter haloes”. In: *MNRAS* 357.2 (2005), pp. 608–618. DOI: 10.1111/j.1365-2966.2005.08667.x. arXiv: astro-ph/0406593 [astro-ph].

[298] D. G. York et al. “The Sloan Digital Sky Survey: Technical Summary”. In: *AJ* 120 (Sept. 2000), pp. 1579–1587. DOI: 10.1086/301513. eprint: astro-ph/0006396.

[299] Matthew D. Zeiler. “ADADELTA: An Adaptive Learning Rate Method”. In: *arXiv e-prints*, arXiv:1212.5701 (2012), arXiv:1212.5701. arXiv: 1212.5701 [cs.LG].

[300] Y. B. Zel’Dovich. “Reprint of 1970A&A.....5...84Z. Gravitational instability: an approximate theory for large density perturbations.” In: *A&A* 500 (1970), pp. 13–18.

[301] Yuanyuan Zhang et al. “Studying Intercluster Galaxy Filaments through Stacking gmBCG Galaxy Cluster Pairs”. In: *ApJ* 773.2, 115 (2013), p. 115. DOI: 10.1088/0004-637X/773/2/115. arXiv: 1304.6696 [astro-ph.CO].

[302] C. Zucker, C. Battersby, and A. Goodman. “Physical Properties of Large-scale Galactic Filaments”. In: *ApJ* 864, 153 (Sept. 2018), p. 153. DOI: 10.3847/1538-4357/aacc66. arXiv: 1712.09655.

[303] C. Zucker and H. H.-H. Chen. “RadFil: A Python Package for Building and Fitting Radial Profiles for Interstellar Filaments”. In: *ApJ* 864, 152 (Sept. 2018), p. 152. DOI: 10.3847/1538-4357/aad3b5. arXiv: 1807.06567.

[304] F. Zwicky. “On the Masses of Nebulae and of Clusters of Nebulae”. In: *ApJ* 86 (1937), p. 217. DOI: 10.1086/143864.

List of Figures

List of Tables

RÉSUMÉ

L'étude de l'évolution et de la composition des grandes structures de l'Univers, comme les amas de galaxies et les filaments cosmiques, est l'un des sujets de recherche les plus actifs en cosmologie. Au cours de ma thèse, j'ai analysé des relevés publiques multi-longueurs d'onde (SDSS, WISE, *Planck*), en utilisant des techniques d'analyse de données nouvelles comme l'apprentissage automatique ou les méthodes bayésiennes. Le but de mon travail était d'étudier la matière baryonique (gaz ionisé et chaud et galaxies) dans le milieu intra-amas, et dans la toile cosmique (ponts de matière et filaments cosmiques).
Au cours de ma thèse, j'ai notamment développé un algorithme basé sur l'intelligence artificielle afin d'estimer les propriétés des galaxies, telles que le taux de formation des étoiles (SFR) et la masse stellaire dans la plage de redshift 0,05<z<0,3, directement applicable aux données infrarouge WISE. La détermination de ces deux propriétés pour de grands volumes de ciel permet de classer un grand nombre de galaxies en fonction de leurs types (actives, en transition, passives), une étape clé pour étudier l'effet de l'environnement sur l'évolution des galaxies.
En combinant ces propriétés avec les informations sur le gaz chaud et ionisé observé via l'effet Sunyaev-Zel'dovich (SZ), j'ai étudié la paire d'amas de galaxies A399-A401 reliée par un pont de matière de 3Mpc de long, détecté de manière significative sur la carte de *Planck*. J'ai caractérisé le gaz et les galaxies dans ce système double exceptionnel, qui est un cas unique de filament cosmique chauffé par la collision des deux amas.
En estimant le SFR et la masse stellaire du catalogue WISExSCOS (15 millions de sources dans la gamme 0.1<z<0.3), j'ai étendu mon travail sur le pont de matière A399-A401 à une étude statistique des propriétés des galaxies le long des filaments cosmiques, de longueurs comprises entre 10 et 100 Mpc environ. Le catalogue de filaments, extrait de l'échantillon de galaxies spectroscopiques SDSS DR12, contient 10000 objets. J'ai étudié les propriétés statistiques des galaxies actives, en transition, et passives, et de leurs SFRs et masses stellaires autour de ces structures pour explorer le processus d'arrêt de formation d'étoiles.
Un autre aspect de ma thèse était la réduction et l'analyse des données d'un grand programme de l'ESO, qui visait à détecter les galaxies d'amas des sources SZ *Planck* dans les bandes optiques g, r, et z et en spectroscopie de 4000 à 8000 Angstrom. Le but de ce programme était de confirmer ou infirmer la détection de 129 amas massifs identifiés par *Planck* via l'effet SZ dans l'hémisphère sud. J'ai développé des pipelines pour réduire les données photométriques avec NTT/EFOSC2 et les données spectroscopiques avec VLT/FORS2. Des redshifts ont pu être fournis pour tous les amas confirmés. Un total de 49 amas avec des redshifts estimés z>0,5 ont été observés avec VLT/FORS2. L'analyse des 3000 spectres obtenus est en cours et avec les données actuelles, nous confirmerons spectroscopiquement environ 10 amas avec près de 10 galaxies membres par amas.

La quantité de gaz ionisé et chaud de la toile cosmique pourrait représenter environ 40% des baryons à bas redshift. Il est donc important d'utiliser des traceurs d'observation optimisés pour le gaz. Dans ce but, j'ai développé un algorithme d'apprentissage profond utilisant tout le potentiel des données *Planck* pour détecter l'effet SZ des amas de galaxies et du gaz chaud ionisé. J'ai entraîné un réseau neuronal convolutionnel sur les cartes de fréquence de *Planck* à reconnaître les amas sélectionnés en SZ, rayons X et optique. Avec ce modèle, j'ai généré une carte du ciel de SZ à faible bruit, dans laquelle j'ai détecté environ 10 fois plus de candidats que le premier catalogue *Planck* des sources SZ. J'ai montré que l'apprentissage profond semble être une approche prometteuse pour améliorer la carte SZ et pour détecter le gaz à faible pression dans la toile cosmique.

MOTS CLÉS

Cosmologie, toile cosmique, effet SZ, amas de galaxies.

ABSTRACT

Studying the evolution and the composition of the largest structures of the Universe, e.g., galaxy clusters and cosmic filaments, is one of the most challenging research topic in cosmology. During my PhD thesis, I have analysed publicly available multi-wavelength surveys (namely SDSS, WISE, *Planck*), using new techniques in data analysis such as machine learning or Bayesian methods. The purpose of my work was to study the baryonic matter (hot gas and galaxies) in the intra-cluster medium, and in the Cosmic Web (bridges of matter between clusters of galaxies and cosmic filaments).
During my thesis, I have developed an algorithm based on machine learning in order to estimate galaxy properties, such as star formation rate and stellar mass in the redshift range 0.05<z<0.3, directly applicable to WISE infra-red data. Determining rapidly these two properties on large volumes of the sky allows us to classify a large number of galaxies (active, transitioning, passive), which is a key step to probe the effect of the environment on galaxy evolution.
By combining these properties with the information on the hot and diffuse gas in the Cosmic Web observed via the thermal Sunyaev-Zel'dovich effect (SZ), I have studied the galaxy cluster pair A399-A401, a bridge of matter of 3Mpc long, detected with high significance in the *Planck* map. I have characterised the gas and the galaxies in this exceptional double system, which is a unique case of cosmic filament heated by the ongoing collapse of the two clusters.
By estimating SFR and stellar mass of the full-sky photometric redshift catalogue WISExSCOS (15 million sources in the range 0.1<z<0.3), I have extended my work on the bridge of matter A399-A401 to a statistical study of galaxy properties around cosmic filaments, of lengths from about 10 to 100 Mpc. The catalogue of filaments, extracted from the SDSS DR12/BOSS sample of spectroscopic galaxies, contains 10000 objects. I have studied the statistical properties of the active, transitioning, and passive galaxies and of their SFRs and stellar masses in the largest filaments of the Cosmic Web to explore the quenching process.
Another aspect of my PhD thesis was the data reduction and analysis of an ESO large program, that aimed to detect cluster galaxy members in optical bands g, r, and z and in spectroscopy from 4000 to 8000 Angstrom. The goal of this *Planck* follow-up program was to confirm the detection of 129 southern candidate massive clusters identified by *Planck* via the SZ effect. I have developed pipelines to reduce both photometric data with NTT/EFOSC2 and spectroscopic data with VLT/FORS2. Among the 129 SZ sources, only a handful are false detections. Estimated redshifts can be provided for all the confirmed clusters. A total of 49 clusters with estimated redshifts z>0.5 were observed with VLT/FORS2. The analysis of the 3000 obtained spectra is still ongoing and with the actual data, the spectroscopic confirmation of about 10 galaxy clusters with about 10 members per clusters are expected.

The gas content of the Cosmic Web may account for about 40% of the baryons. Therefore, it is important to use optimised observational tracers of the gas. In this purpose, I have developed a deep learning algorithm using the full potential of the *Planck* data to detect the SZ effect tracing galaxy clusters and hot ionised gas. I have trained a convolutional neural network on the *Planck* frequency maps to recognize clusters selected in SZ, X-rays, and optical. With this trained model I have generated a full-sky map of low noise SZ signal, in which I have detected about 10 times more candidates than the first *Planck* catalogue of SZ sources. I have shown that deep learning seems a promising approach to improve the y-map and detect smaller pressure halos and diffuse gas in the Cosmic Web.

KEYWORDS

Cosmology, Cosmic Web, SZ effect, galaxy clusters.